Pergamon

Computers & Geosciences Vol. 23, No. 4, pp. 387–395, 1997

Printed in Great Britain
PII: S0098-3004(97)00005-8
0098-3004/97 $17.00 + 0.00

VISUALIZING SPATIAL DATA UNCERTAINTY USING ANIMATION

CHARLES R. EHLSCHLAEGER,[1] ASHTON M. SHORTRIDGE,[2] and MICHAEL F. GOODCHILD[2]

[1]Department of Geography, Hunter College, New York, NY 10021, and [2]NCGIA, University of California, Santa Barbara, CA 93106
(*e-mail*: chuck@everest.hunter.cuny.edu)

(Received 4 October 1996; accepted 15 November 1996)

Abstract—This paper examines methodologies for dynamically displaying information about uncertainty. Modeling uncertainty in elevation data results in the generation of dozens or hundreds of realizations of the elevation surface. Producing animations of these surfaces is an approach to exploratory data visualization that may assist the researcher in understanding the effect of uncertainty on spatial applications as well as in communicating the results of the research to a wider audience. A nonlinear method for interpolation between the surface realizations is introduced which allows for smooth animation while maintaining the surface characteristics prescribed by the uncertainty model. © 1997 Elsevier Science Ltd

Key Words: Animation, Uncertainty, Spatial data, Digital elevation model, Optimal route, Random fields.

INTRODUCTION

Recent research has resulted in several models of the uncertainty for spatial data and their applications (Dettinger and Wilson, 1981; Heuvelink, Burrough, and Stein, 1989; Theobald, 1989; Goodchild, 1992; Goodchild, Sun, and Yang, 1992; Hootsman and van der Wel, 1993; Ehlschlaeger and Goodchild, 1994b). A model of uncertainty is often needed when the data available are too coarse or generalized for the application. In the example used in this paper, the problem to be solved uses digital elevation data with a sampling interval of 30 m, but the only data available have a sampling interval of 3 arc sec, approximately three times more coarse. In such situations, it is useful to know the uncertainty that has been introduced by using data that is too coarse for the application. A model of uncertainty can provide the answer, if it is capable of simulating the missing variation; in other words, the range of possible 30 m DEMs that would be consistent with the available 3 arc sec DEM. The parameters of this model would come, as they do in this paper, from analysis of areas for which both 3 arc sec and 30 m DEMs are available for comparison. The situation is not only theoretical; whereas complete 3 arc sec DEM coverage is available for the 48 contiguous United States, many areas lack 30 m coverage (for current United States 30 m DEM coverage, see USGS, 1996).

Stochastic approaches to modeling spatial uncertainty result in the creation of many potential realizations for the spatial dataset of interest (for examples, see Openshaw, 1979; Goodchild, Sun, and Yang, 1992; Fisher, 1993a; Ehlschlaeger and Goodchild, 1994b; Ehlschlaeger and Shortridge, 1996). Examining these realizations (which may number in the hundreds) and exploring the similarities and differences between them can be a major challenge. The authors' efforts to accomplish this prompted an exploration of non-traditional forms of cartographic representation, including animation. A particular problem in the generation of smooth animations from a series of "stills" is the creation of intermediate images to blend from one still to the next. These intermediate images are critical for a smooth blending from one realization to the next, but if they lack the statistical characteristics of the actual realizations then the animation will misrepresent the data and the form is inappropriate for analysis or communication.

The following section discusses the goals and methodology involved in the creation of the simulated surface realizations. The third section covers the application of surface realizations to a least-cost path algorithm, which provides a measure for the expected distribution of path costs. The fourth section describes the role that animation can play in visualizing uncertainty: the user may develop a better understanding for the impact of generalized spatial data on the outcome of the least-cost path algorithm. Then, the fifth section examines conceptual and practical issues for the interpolation

between realizations necessary for generating an animated sequence. Finally, the paper addresses the contribution animations may make in the analysis and communication of uncertainty in spatial data. Animation, if correctly produced, may offer an alternative to the usual examination of tens or hundreds of static maps during the exploratory phase of data analysis, or to the usual slides and transparencies developed for a presentation. It is worth investigating whether animation could provide any additional benefits for uncertainty analysis beyond simply keeping the audience awake during a presentation.

POTENTIAL REALIZATIONS OF THE LANDSCAPE

The research which resulted in this paper was primarily concerned with examining the impact of spatial uncertainty in elevation data upon a corridor location algorithm (Church, Loban, and Lombard, 1992). The outcome of this analysis was a large number of potential realizations of the elevation surface and the cost surface. This paper concentrates on issues of representation, so the theory and methodology employed to generate the realizations are covered only briefly here. The interested reader is referred to Ehlschlaeger and Shortridge (1996) for a more detailed discussion. The following paragraphs provide a concise description of the elevation uncertainty model.

For the purposes of this paper, we assume that the corridor location problem to be solved requires a DEM of 30 m sampling interval. However, such data are not always available, and thus we examine the effects of replacing them with coarser 3 arc sec data, using an error model calibrated in regions for which both are available. We assume that the differences between the two resolutions are consistent between both areas: the area where the problem must be solved, and the area where both datasets are available. The relationship between 30 m and the coarser 3 arc sec data may be characterized by examining the distribution of differences between the two datasets at a large number, J, of randomly selected spatially uncorrelated locations on the surface. The difference is modeled using the mean and standard deviation of 96 sets of randomly drawn, spatially independent points scattered across the surface. Unconditional stochastic simulation can then be employed to define a probability density function (p.d.f). The p.d.f. generates random surfaces with a Gaussian distribution matching the mean and standard deviation observed in the difference maps and is represented by:

$$E(\mathbf{u}) = m(\mathbf{u}) + m(m(T)) + (m(s^2(T))\epsilon)Z(\mathbf{u}) \quad (1)$$

where $E(\mathbf{u})$ is a realization of the higher quality 30 m elevation data using the more generalized 3 arc sec data $m(\mathbf{u})$, T is a group of sets of spatially uncorrelated sample points, ϵ is a random variable with mean 0.0 and variance 1.0 perturbing the realizations' standard deviation, and $Z(\mathbf{u})$ is a random field perturbing all points u within the realization. The following expressions define the remaining terms:

$$m(m(T)) = \sum_j \frac{m(T_j)}{J},\ j \subset T, \quad (2)$$

$$m(s^2(T)) = \sum_j \frac{s^2(T_j)}{J}, \quad (3)$$

$$s^2(s^2(T)) = \sum_j \frac{s^2(T_j) - m(s^2(T_j)))^2}{J-1}, \quad (4)$$

and

$$Z(\mathbf{u}) = \frac{\sum_v w_{u,v}\epsilon_v}{\sqrt{\sum_v w_{u,v}^2}}, w_{u,v}$$

$$= \begin{cases} 1 & : d_{u,v} \le F \\ \left(1 - \dfrac{d_{u,v} - F}{D - F}\right)^E & : F < d_{u,v} < D,\ u \subset \mathbf{u},\ v \subset \mathbf{v} \\ 0 & : d_{u,v} \ge D \end{cases} \quad (5)$$

where $m(m(T))$ is the average mean for all sets; $m(s^2(T))$ is the average variance for all sets; $s^2(s^2(T))$ is the variance of the variances for all sets; and $Z(\mathbf{u})$ specifies the random field with spatial dependence parameters where u is a point on the random field with a theoretical mean of 0.0 and theoretical variance of 1.0, $\mathbf{v}$ is the set of points affecting $\mathbf{u}$, v is the spatial autocorrelative effect between points u and v, ϵ_v is a random variable with a mean of 0.0 and variance of 1.0, $d_{u,v}$ is the distance between u and v, D is the minimum distance of spatial independence, E is the distance decay exponent, and F is a parameter that adds flexibility to the probability distribution function model fitting process. Matching the spatial autocorrelation using F is important in order to capture the terrain texture observed in the higher resolution dataset (Goodchild, 1986; Theobald, 1989), and was implemented as the GRASS command r.random.surface (Ehlschlaeger and Goodchild, 1994a). The result of performing this analysis is a set of parameters defining the p.d.f. for modeling the uncertainty of the 3 arc sec elevation surface:

$$E(\mathbf{u}) = m(\mathbf{u}) - 6.57 + (23.53 + 5.06\epsilon)Z(\mathbf{u}) \quad (6)$$

where $D = 4600$, $E = 0.07$, and $F = 200$ for the parameters of $Z(\mathbf{u})$. Each surface created from these parameters is a potential realization of what the actual difference surface might be. By adding each difference field to the 3 arc sec surface, a large collection of alternative, equally probable models of the elevation surface is created.

The digital elevation models cover a large study area around Santa Barbara, California, as depicted in *CD-Figure 1* in the digital version of the paper. The 3 arc sec dataset, USGS quad Los_Angeles-w, extends from 119–120°W and from 34–35°W. The terrain is characterized by mountainous topography extending inland from the Pacific Ocean, punctuated by river valleys and narrow coastal plains. A test area several kilometers on each side was defined within this DEM. It extends from the coastal plain at the town of Goleta in the south across the Santa Ynez mountains to the Santa Ynez river valley in the north. For the purposes of this paper, we assumed that no 30 m data existed for this test area, and that we would have to determine the location of a path within this area using 30 m quality data.

Six 30 m datasets outside of the test area were compared to collocated data from the Los_Angeles-w DEM to develop parameters for the p.d.f. The p.d.f. was used to generate 250 realizations simulating 30 m quality elevation data for the test area. These elevation surfaces were processed to create 250 corridor realizations. The next section describes how the corridor realizations may be used to demonstrate uncertainty in the route location application caused by the coarseness of the 3 arc sec elevation data.

CORRIDOR LOCATION ANALYSIS

Generating accumulated cost surfaces is a two-step procedure. First, a cost surface for each of the 250 realizations of the test area elevation datasets was produced. Cost is a function of horizontal distance, slope (Horn, 1981), and absolute elevation, as calculated for each cell by the following:

$$Cost = 30 + 300\tan(Slope) + \max(0,\ Elev - 400). \quad (7)$$

Second, two locations in the test area were chosen to be the endpoints for a hypothetical path. For each of the 250 realizations created in the first step, an accumulated cost surface was generated. The value in each cell of the accumulated cost surface represents the accumulated cost to travel to this cell from both of the endpoints across the cost surface produced in step one (Church, Loban, and Lombard, 1992). The resulting cost pattern in any single realization gives a visual indication of the degree to which the character of the terrain restricts the corridor of the least-cost path.

The authors have explored two methods for displaying realization results. The first method, used in an earlier study, condenses optimal routes from all realizations into a single, static map. The second method presents route cost information in animations of cost surface realizations.

Using method one, as visualized in *CD-Figure 2*, routes were generated connecting the two white spheres. This figure portrays the terrain of the study area (at a vertical exaggeration of 1.5) draped with a representation of the 250 optimal paths. Because the density of paths across certain portions of the test area was so great, displaying each path individually on the same image was impossible. Instead, we have colored the raster based on the number of optimal routes passing through each cell. Out of the 250 realizations, 40 or more optimal paths traveled through blue cells, 20 or more optimal paths through red cells, five or more paths through green cells, one path through yellow cells, and no optimal paths through gray cells. The white line on the surface outlines the optimal route traced on the original 3 arc sec dataset, which had a cost of 56,561 units. Using the actual 30 m data for the test area, the optimal path had a cost of 61,368 and followed the red line in *CD-Figure 2*. The distribution, as seen in the histogram in Figure 1 was unimodal; the mean cost of the 250 optimal paths was 64,034 units with a standard deviation of 2991 units.

Although this information provided a good measure of expected costs, it left several questions unanswered. Since a major goal of spatial data uncertainty research is to understand impacts of generalized map data on applications, these questions include: Under what conditions does a realized optimal path deviate from the optimal path on the 3 arc sec dataset? Whereas it is obvious that the quality of DEM necessarily depends on the choice of start and stop locations, does the large spatial variability of optimal paths indicate that 3 arc sec elevation data are inadequate for determining the optimal path between these two locations? Method two attempts to answer these questions through the generation and dynamic visualization of accumu-

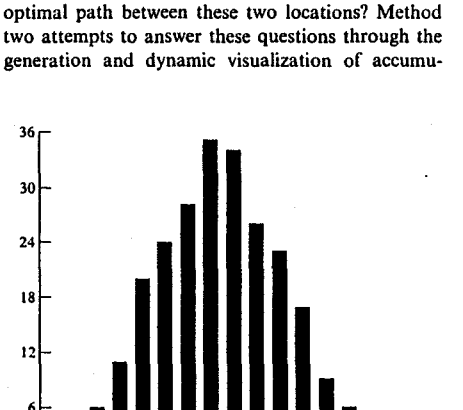

Figure 1. Histogram of potential optimal paths.

ated cost surface realizations, rather than individual optimal paths.

We refer to each of these realizations as accumulated cost surfaces, because they portray the aggregate cost to travel to any particular cell on the surface. These surfaces may be preferred to a simple calculation of the optimal route for a given elevation realization. In addition to showing a single specific optimal path, accumulated cost surfaces provide a better idea of the general optimal path corridor. By comparing the different patterns from all of these realizations, the researcher can gain insight into how uncertainty in the spatial data affects the result of the path algorithm. Although 250 realizations were used to calculate the statistics of DEM uncertainty for the optimal route problem, a maximum of 35 were used in these animations.

The colors in both static and dynamic depictions of these accumulated cost surfaces (see the animations in "Optimal Route Movie", as well as *CD-Fig. 4*; both are in the digital version of this paper) attempt to maximize the contrast between the least expensive optimal route and the most expensive optimal route of the accumulated cost surfaces. The color scheme employed is different from the one used for the path image. The least expensive route (or routes if multiple paths have the same cost) within a realization are represented by white cells. Black cells indicate the cells that fall within one percent of the least expensive optimal route cost in the 250 realizations. The colors then ramp from black through blue and green to yellow (yellow cells have accumulated costs 10% greater than the least expensive optimal path in the 250 realizations). Gray cells have costs greater than 10% of that for the optimal path. The red line encompasses the set of cells with costs below that of the most expensive optimal path, 83,809 units, of any realization.

We were interested in determining the effectiveness of non-traditional methods of visualization for illustrating the impact of uncertainty on the application. The following section discusses the methods we employed to generate a smooth animation to accomplish this.

ROLE OF ANIMATION IN VISUALIZING UNCERTAINTY IN SPATIAL DATA

Data visualization may be categorized along a continuum that stretches throughout the duration of a research project (DiBiase, 1990). During the early stages of the work, animation of uncertainty can be an invaluable aid for exploratory analysis of the data. The methodology employed here to generate animation sequences did not consume a large amount of time, so incorporating animations into this phase is technically feasible. At the opposite end of the continuum, graphic representations of the data and analysis can assist in communicating results and clarifying important points to the scientific community.

Spatial autocorrelative characteristics can play a significant role in understanding the impact of uncertainty during the research process. This information can be especially critical in a spatial application for which the relative locations of objects are important (e.g. optimal path routes and viewshed analysis). The video "Random Fields and their use in Representing Spatial Autocorrelation" (Ehlschlaeger, 1994) communicates the importance of spatial autocorrelation in representing spatial uncertainty. The video includes two animations, both of which show the impact of potential sea-level rise on the shoreline of Boston Harbor. The first animation, "Ignoring Autocorrelation Movie", assumes the uncertainty term has a constant value for each realization. Although the magnitude of error is represented correctly at every cell, the shoreline shapes do not represent potential results. The second animation, "Spatial Autocorrelation Included Movie", incorporates the spatial autocorrelative characteristics of uncertainty. In both animations, the amount of time a section of land remains underwater represents the probability of submergence given a 1.9 m rise in sea level. However, the second animation details the effect of ocean level rising on contiguous regions (e.g. "What is the probability that this road will not be covered with water?").

The spatial autocorrelative characteristics described for the second animation movie are used in the realization for the animation developed in this study. Each realization possesses different landscape elevation values. As a result of the changes to each elevation surface, the accumulated cost surface is also unique for each realization. In an attempt to portray both simultaneously, the main approach adopted in this paper employed a 2.5-dimensional perspective view of the test area with the accumulated cost surface draped over the elevation surface. Although visually appealing, the use of the perspective view is not immune from criticism because of the problems of displaying three-dimensional data on a flat, two-dimensional screen. People may have trouble perceiving perspective models (due to distortion and obscured sections) and gathering useful information from them (Dorling and Openshaw, 1992). Therefore, for comparative purposes, a 2-D animation using elevation contours to portray the terrain was produced; it is viewable at the main animation page. In either instance, however, the goal of the visualization is not to provide precise rendering of the detail of a single scene, but to promote understanding of the magnitude of change between images.

The spatial pattern of any single realization is not of particular interest, for no single realization is more likely than any other to approximate the actual elevation surface. Instead, more interesting

information may be gleaned from the details of the relationships between the realizations. The changes in the width and route of the corridor illustrate the impact of uncertainty. If there is little change from one realization to another, one can be fairly certain that the least-cost path lies along a well-defined corridor. On the other hand, if change is dramatic, and a variety of differing corridors are suggested, then error in the spatial data has translated into a great deal of uncertainty in the application. Much can be gleaned from comparing static images of multiple realizations, but developing an implicit understanding of the changes between each image can be time consuming. Researchers may benefit from the ability of animation to depict dynamically the range of uncertainty inherent in their data; they may spot relationships in the data, identify errors in their assumptions, or consider new directions for research (Fisher, 1993b; van der Wel, Hootsmans, and Ormeling, 1994). The particular animation method employed may affect how the viewer perceives the data and the relationships between variables. Although this work does not address the impact of various animation techniques on viewer perception, it does present a method of interpolation which avoids smoothing the intermediate images.

Another role for animation during the exploratory phase of research is the generation of additional information about the uncertainty of data and how data uncertainty affects the application. The process of interpolation employed here creates a large number of statistically valid dependent realizations. Animating these images provides a natural way to view this massive influx of data in a time-effective manner (Dorling, 1992). Viewers also see the simultaneous movement of elevation and accumulated cost surface, providing greater understanding for that relationship. As mentioned in the reviews section, there were two questions we wished to answer with this animation: Under what conditions does a realized optimal path deviate from the 3 arc sec optimal path? And, does the large spatial variability of optimal paths indicate that 3 arc sec datasets are inadequate for determining the optimal path between these two locations? The animation in this study, "Optimal Route Movie", helps to answer these questions.

Observing the animation, the viewer may notice several factors. The most obvious is that optimal paths often change location and cost for reasons not easily perceived simply by viewing the elevation surface. Although the overall shape of the 3 arc sec dataset does not change, many smaller ridges and valleys appear and disappear within the realizations. By comparing the images of the 3 arc sec dataset and the 30 m dataset side-by-side, one notices that the 3 arc sec dataset is missing many ridges and valleys apparent in the 30 m data. This missing terrain texture is apparently why optimal routes on the realizations of the fine resolution 30 m DEM are approximately 12% more expensive than the optimal route on the coarse resolution 3 arc sec DEM. The viewer will also notice realizations for which the optimal path shifts to a dramatically new location. The animation makes it clear that there is no simple relationship between the optimal routes on different realizations, or between the optimal routes computed at different spatial resolutions. Some-fine resolution optimal paths are similar to the coarse-resolution optimum, but some are different. Clearly, the uncertainty introduced by resorting to coarse resolution data is propagated and amplified in the optimal routes. To answer the two questions posed earlier, the animations demonstrate that there are no simple relationships between optimal paths at the two resolutions; and coarse-resolution data are indeed inadequate for the original purpose of finding an optimal path between the two endpoints originally chosen.

Although the animations in this study show a particular situation for which uncertainty in coarse data renders that data inadequate, conditions in other situations may be such that uncertainty does not invalidate the same coarse dataset. For example, *CD-Figure 5* in the digital version of the paper illustrates optimal paths between two different endpoints using the same cost equation and the same 250 surface realizations (graphic representation is the same as for Fig. 2). The spatial distribution of these optimal paths demonstrates that there is little difference between corridor solutions for the second set of endpoints. In this instance, 3 arc sec data is adequate.

PRODUCING ANIMATIONS OF UNCERTAINTY

The production of animations is not technically difficult. By stringing together a sequence of realizations and smoothing the transitions between them, one can readily create an animated sequence of images using current technology and public domain software. Significant theoretical issues arise, however, in developing the interpolation method and calibrating the frame sequencing. This section concentrates on these factors.

A central issue for the production of a smooth animation is the generation of intermediate images to ensure that the transition between images is gentle and cohesive (MacEachren and DiBiase, 1991). Allowing for a transition permits the viewer to see the magnitude and pattern of the differences in elevation and cost surface between the realizations. These intermediate images are interpolations between the original realizations; we generated eight interpolated images between each of the 35 realizations to develop a smooth transition. Nearly 90 percent of the frames in the animation, then, are not original realizations; they are interpolated

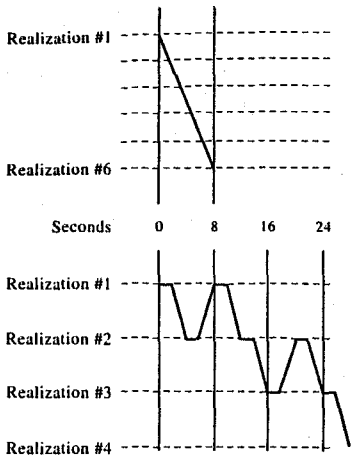

Figure 2. Timelines of animations: upper from "Optimal Route Movie", lower from "Spatial Autocorrelation Included".

images. In Figure 2 the upper timeline represents the animation sequence for "Optimal Route Movie". The horizontal lines represent independent realizations. The dark line represents the flow of "Optimal Route Movie". Although there are only 35 independent realizations, the 61 sec animation portrays 306 different realizations of the optimal path.

For this study of uncertainty, the statistical and spatial characteristics of the interpolated surfaces must match the error model, or the resulting animation will become a misleading visualization tool, or at least it will be much more difficult to achieve desired results. For example, "Spatial Autocorrelation Included Movie" used a linear interpolation for the transitions. In order for the viewer to best view actual realizations, the authors developed the lower timeline in Figure 2. The dark line represents the flow of the "SAI Movie", with the horizontal sections representing pauses in the animation flow for 2 sec at each independent realization. Between each independent realization, the diagonal lines represent intermediate frames morphing from one realization to the next. Because these intermediate frames were not independent realizations, the goal of the "SAI Movie" was to allow half the time to be spent showing actual realizations, while the other half created a transition between realizations. Therefore, the animation was only able to show 29 realizations in 113 sec. In addition to only half of the time being spent viewing realizations, terrain motion did not provide accurate visual clues of uncertainty in terrain estimation.

The interpolation method is clearly important for animations of stochastic surfaces. Linear interpolation results in intermediate images with different characteristics than the independent realizations they tie together. Mean uncertainty is modeled correctly, but variance of uncertainty is lower for the linearly interpolated images. And, the spatial autocorrelative characteristics of uncertainty are also not representative. A nonlinear interpolation presented here solves this problem. The following equations are used to interpolate between independent realizations in "Optimal Route Movie":

$$\epsilon^{i}_{x,y} = \epsilon_x \cos\left(\frac{i\pi}{2}\right) + \epsilon_y \sin\left(\frac{i\pi}{2}\right), \quad 0 \le i \le 1 \qquad (8)$$

where $\epsilon^{i}_{x,y}$ is an "interpolated" random value between random values ϵ_x and ϵ_y with a mean of 0.0 and standard deviation of 1.0; and:

$$Z^{i}_{x,y}(\mathbf{u}) = Z_x(\mathbf{u}) \cos\left(\frac{i\pi}{2}\right) + Z_y(\mathbf{u}) \sin\left(\frac{i\pi}{2}\right), \quad 0 \le i \le 1 \qquad (9)$$

where $Z^{i}_{x,y}(\mathbf{u})$ is an "interpolated" surface between surfaces $Z_x(\mathbf{u})$ and $Z_y(\mathbf{u})$, with every point having a mean of 0.0 and a standard deviation of 1.0. We draw attention to the word "interpolated" because we are trying to create values of $Z^{i}_{x,y}(\mathbf{u})$ and $\epsilon^{i}_{x,y}$ that are appropriate for our p.d.f. and have values similar to nearby values of i. In place of a formal proof, imagine the two endpoints x and y as orthogonal unit vectors. Since each vector in $\epsilon^{i}_{x,y}$ and $Z^{i}_{x,y}(\mathbf{u})$ has a mean of 0.0 and standard deviation of 1.0, we can retain their statistical properties by locating intermediate realizations along a circle between them centered at the origin (which explains the sin() and cos() functions). By using the functions for $\epsilon^{i}_{x,y}$ and $Z^{i}_{x,y}(\mathbf{u})$, interpolations of independent realizations are also (dependent) realizations in their own right. Since these interpolated images are valid representations of the surface, the employment of the nonlinear function allows one to generate visually accurate animations.

CONCLUSION

This paper presents a method for developing animations from realizations of a surface. By viewing the dynamic transformations of the surface, the viewer can gain an understanding for the role that uncertainty plays in the spatial outcome of the analysis. The nonlinear interpolation method presented here maintains the equivalence of the intermediate images to the distribution from which the independent realizations are drawn. The resulting frames exhibit the proper statistical characteristics and are a valid means for visualizing uncertainty. The utility of these animations is perhaps greatest for exploratory analysis. The images are visually complex, and more general audiences may find them most useful for a quick qualitative impression of the magnitude of uncertainty.

Viewing animations of spatial data uncertainty as they affect applications also provides a good mechanism to facilitate visual perception of the probabilistic nature of uncertainty (Beard, Buttenfield, and Clapham, 1991). In a deterministic world, we expect answers to questions such as: "How long will the optimal route be?" or "Where are the possible locations of optimal routes?". Probabilistically, we may never know what the actual answers will be with generalized datasets, but we may learn what factors will determine the actual answers and how the actual answers relate to the information that is unavailable. And, on occasions, we will learn that generalized data may be useful for meeting the needs of specific applications requesting precise data.

Several directions present themselves for further exploration. This example is simple, as complex interactions go. The researcher is confronted with only one independent variable, that being elevation, and one dependent variable, that being the cost surface. Many spatial applications involve considerably more inputs. Extending the methodology presented here may produce a method for visualizing the role uncertainty plays for each input layer individually upon the outcome of the analysis. This research focused primarily upon the interpolation method, but frame rate and duration are other key technical factors which affect perception of the animation. These were examined in earlier work on the Boston Harbor data, but relationships between all three factors merit continued exploration. Additional research is also warranted to assess the relative ease (or difficulty) people have in making sense of animations of abstract concepts like uncertainty (see Evans, 1996 for current research into viewers' perception of simultaneous displays of spatial data and their associated reliability). This direction is especially important for understanding the role for animations of uncertainty to communicate results to the community. Understanding how people perceive spatial user animation representations should provide a stronger basis for developing more effective animations.

The implementation of the nonlinear interpolation method allows for the portrayal of uncertainty as the image shifts smoothly through a series of realizations. The resulting animation is a complex visualization tool for perception of a complex spatial phenomenon. Dynamic visualization may prove valuable for developing understanding and appreciation for the role data uncertainty plays in spatial analysis.

Acknowledgments—We thank Alan MacEachren and Menno-Jan Kraak for their suggestions on this paper and for the hard work they put in to make this special visualization issue. In addition, the paper benefited from the constructive and thoughtful comments of an anonymous reviewer.

REFERENCES

Beard, K., Buttenfield, B., and Clapham, S., 1991, NCGIA research initiative 7: visualization of spatial data quality: NCGIA Technical Paper 91-26, Santa Barbara, California, 222 p.

Brown, B., 1992, SG3d, supporting information: U.S. Army Corps of Engineers, Construction Engineering Research Lab. Technical Paper, Champaign, Illinois, 14 p.

Church, R. L., Loban, S. R., and Lombard, K., 1992, An interface for exploring spatial alternatives for a corridor location problem: Computers & Geosciences, v. 18, no. 8, p. 1095–1105.

Dettinger, M. D., and Wilson, J. L., 1981, First order analysis of uncertainty in numerical models of groundwater flow: part 1: mathematical development: Water Resources Research, v. 17, no. 1, p. 149–161.

DiBiase, D., 1990, Visualization in the earth sciences: Bull. College of Earth and Mineral Sciences, Pennsylvania State Univ., v. 59, no. 2, p. 13–18.

Dorling, D., 1992, Stretching space and splicing time: from cartographic animation to interactive visualization: Cartography and Geographic Information Systems, v. 19, no. 4, p. 215–227.

Dorling, D., and Openshaw, S., 1992, Using computer animation to visualize space–time patterns: Environment & Planning B, v. 19, p. 639–650.

Ehlschlaeger, C. R., 1994, Random fields and their use in representing spatial autocorrelation: Unpubl. Videotape, NCGIA, Santa Barbara, California.

Ehlschlaeger, C. R., and Goodchild, M. F., 1994a, Uncertainty in spatial data: defining, visualizing, and managing data errors: Proc. GIS/LIS '94, Phoenix, Arizona, p. 246–253, LINK: <http://everest.hunter.cuny.edu/~chuck/gislis/gislis.html>.

Ehlschlaeger, C. R., and Goodchild, M. F., 1994b, Dealing with uncertainty in categorical coverage maps: defining, visualizing, and managing data errors: Proc., Workshop on GIS, Conference on Information and Knowledge Management, Gaithersburg, p. 86–91, LINK: <http://everest.hunter.cuny.edu/~chuck/acm/paper.html>.

Ehlschlaeger, C., and Shortridge, A., 1996, Modeling elevation uncertainty in geographical analysis: Proc. Spatial Data Handling '96, Delft, The Netherlands, v. 2, p. 9B.15–9B.25, LINK: <http://everest.hunter.cuny.edu/~chuck/SDH96/paper.html>.

Evans, B. J., 1996, Cartographic display of data-reliability: does it benefit the map user?: ICA Comission on Visualization Working Papers, LINK: <http://www.gis.psu.edu/ica/icavis/Evans/Evans.html>. (also see this issue).

Fisher, P. F., 1993a, Algorithm and implementation uncertainty in viewshed analysis: Intern. Jour. Geographical Information Systems, v. 7, no. 4, p. 331–347.

Fisher, P. F., 1993b, Visualizing uncertainty in soil maps by animation: Cartographica, v. 30, no. 2, p. 20–27.

Goodchild, M. F., 1986, Spatial autocorrelation: CATMOG 47, Geo Books, Norwich, 57 p.

Goodchild, M. F., 1992, Geographical data modeling: Computers & Geosciences, v. 18, no. 4, p. 401–408.

Goodchild, M. F., Sun, G., and Yang, S., 1992, Development and test of an error model for categorical data: Intern. Jour. Geographical Information Systems, v. 6, no. 2, p. 87–104.

Heuvelink, G. B. M., Burrough, P. A., and Stein, A., 1989, Propagation of errors in spatial modelling with GIS: Intern. Jour. Geographical Information Systems, v. 3, no. 4, p. 303–322.

Hootsman, R., and van der Wel, F., 1993, Detection and visualization of ambiguity and fuzziness in composite

spatial datasets: Proc. EGIS (1993) Genoa, Italy, p. 1035–1046.
Horn, B. K. P., 1981, Hill shading and the reflectance map: Proc. IEEE, v. 69, no. 1, p. 14–47.
MacEachren, A. M., and DiBiase, D., 1991, Animated maps of aggregate data: conceptual and practical problems: Cartography and Geographic Information Systems, v. 18, no. 4, p. 221–229.
Openshaw, S., 1979, A methodology for using models for planning purposes: Environment and Planning, v. 11, p. 879–896.
Theobald, D. M., 1989, Accuracy and bias issues in surface representation, *in* Goodchild, M. F., and Gopal, S., eds., Accuracy of Spatial Databases: Taylor & Francis, London, p. 99–106.
USGS (1996) USGS 7.5 Minute DEM Coverage, LINK: <http://www-nmd.er.usgs.gov/metadata/7.5dem.html>.
van der Wel, F. J., Hootsmans, R. M., and Ormeling, F., 1994, Visualization of data quality, *in* MacEachren, A. M., and Taylor, D. R. F., eds., Visualization in Modern Cartography: Pergamon, New York, p. 313–331.

APPENDIX I

Movies From: Visualizing Spatial Data Uncertainty Using Animation

Four different animations illustrate many of the issues brought up in "Visualizing Spatial Data Uncertainty Using Animation". They are available in the digital version of this paper. The first two animations are from the video: "Random Fields and their use in Representing Spatial Autocorrelation" (Ehlschlaeger, 1994). They are designed to communicate the importance of spatial autocorrelation. The third animation (a set of animations) shows different realizations of 1-Degree DEM (3 arc sec resolution) at 7.5' DEM (30 m resolution) level quality 2.5-D surfaces with Optimal Path solutions for that realization draped over them. Vertical resolution is exaggerated 1.5 times in all animations. See the main text for a complete description of each movie. An alternative to perspective terrain views is provided by the fourth animation. It uses dynamically changing contour lines to portray alterations in elevation between realizations.

In order to facilitate the viewing of the movies, we have generated mpeg and quicktime versions of the animations. There are various options depending on whether the user wishes to view a glimpse of each movie, or a longer version. We have also created a version of the third animation with only two interpolations between realizations instead of eight. This version shows three times as many independent realizations in the same time period, but it is more difficult to see uncertainty effects in the DEM clearly, since the transition between images is less smooth.

In the digital version, users may click on the file size (e.g. 2.4 megs) to retrieve the animation. Files should work well using the appropriate Unix workstation software. On PCs, large .mov files will not load properly in QT for Windows 2.1 using a 486 at 66 mHz w/ 16 megs of RAM running Windows for Workgroups 3.11. The same computer can show .mpg files using Net Toob (downloadable at <http://www.nettoob.com/>).

	Quick Peek		Long Version	
	Mpeg (.mpg)	Quicktime (.mov)	Mpeg (.mpg)	Quicktime (.mov)
Ignoring Spatial Autocorrelation	1.2 megs	0.4 megs	4.6 megs	
Spatial Autocorrelation Included	1.3 megs 10 Realizations	0.5 megs	3.0 megs 35 Realizations	1.1 megs
Optimal Route, 8 interpolations between realizations	2.4 megs	1.9 megs	8.9 megs	7.2 megs
Optimal Route, 2 interpolations between realizations	1.4 megs	0.7 megs	5.0 megs	2.4 megs
Optimal Route, 8 interpolations between realizations, contour lines representing elevation		8.9 megs		

APPENDIX II

Procedure

Here is a step-by-step procedure to generate the potential realizations and animation in this paper. A Silicon Graphics workstation running IRIS 5.3 and GRASS 4.1 with SG3d (Brown, 1992) were used for this project. Shell scripts are available at the link: <http://everest.hunter.cuny/~chuck/CGFinal/paper.htm>.

1. Randomly determine 250 samples of independent random points from the 3 arc sec dataset using *makeInd.csh*.

2. Develop statistics from these points by comparing the difference between the 3 arc sec dataset and available 30 m data.

3. Find the parameters that best fit the random surface model by fitting the difference of 3 arc sec data to available 30 m data. So far, every time the random surface model parameters D, E, and F were used to describe the p.d.f., the 3-D solution space was valley shaped. The following shell scripts test various combinations of model parameters, moving closer to the optimal combination: *mC1.csh*, *mC2.csh*, *mC3.csh*, *mC4.csh*. The goal of each shell script was to test 27 locations surrounding the best solution identified by that stage in the analysis. If a better solution was found, the next shell script would choose locations around the new optimal parameters. Otherwise, the next shell script would test 27 locations closer to the previous optimal solution.

4. Employ the spatial statistical parameters from step 3 to generate 250 realizations using *makeVis.csh*.

5. Build nonlinear interpolations between realizations. Building an interpolation between realizations takes less time than building the random surface. *makeInterp.csh*

6. (Optional) Check to see that interpolations of random surfaces have similar statistical and spatial statistical characteristics as the original realizations. *checkInterp.csh*, *InterpSS.txt*.

7. Generate optimal path for 3 arc sec data and render it. Generate optimal path for 30 m data in the study area and render it. In an actual application, this data would not be available. But, it is important to check and determine whether results meet expectations. *visDted.csh*, *visDEM.csh*.

8. Determine range of optimum path values. Load data into spreadsheet for histogram. Develop color scheme to demonstrate this range, and the color scheme to all the realizations. *describeLCP.csh*, *makeColor.csh*.

9. Generate the script that runs the animation creation software. Finally, run the animation creation software script. *mMovieCsh.csh*, *makeMovie.csh*