Spatial Forecasting of Disease Risk and Uncertainty

Lee De Cola

ABSTRACT: Because maps typically represent the value of a single variable over 2-dimensional space, cartographers must simplify the display of multiscale complexity, temporal dynamics, and underlying uncertainty. A choropleth disease risk map based on data for polygonal regions might depict incidence (cases per 100,000 people) within each polygon for a year but ignore the uncertainty that results from finer-scale variation, generalization, misreporting, small numbers, and future unknowns. In response to such limitations, this paper reports on the bivariate mapping of data "quantity" and "quality" of Lyme disease forecasts for states of the United States. Historical state data for 1990-2000 are used in an autoregressive model to forecast 2001-2010 disease incidence and a probability index of confidence, each of which is then kriged to provide two spatial grids representing continuous values over the nation. A single bivariate map is produced from the combination of the incidence grid (using a blue-to-red hue spectrum), and a probabilistic confidence grid (used to control the saturation of the hue at each grid cell). The resultant maps are easily interpretable, and the approach may be applied to such problems as detecting unusual disease occurrences, visualizing past and future incidence, and assembling a consistent regional disease atlas showing patterns of forecasted risks in light of probabilistic confidence.

KEYWORDS: Risk maps, kriging, choropleth maps, uncertainty, disease, forecasting, ARIMA

Not only is it easy to lie with maps, it's essential. To portray meaningful relationships for a complex three-dimensional world on a flat sheet of paper or a video screen, a map must distort reality [Monmonier 1991].

Abstraction is necessary to cope with the complexity of the real world. All abstraction, however, introduces uncertainty into our analysis and policy formation [MacEachren 1994].

No map of any phenomenon can fully represent multi-scale complexity, temporal dynamics, or the uncertainty that is always associated with a value at a mapped point. Indeed, the above quotations remind us that maps inevitably lie because, like all models, they are abstractions. Conscientious geographers try to overcome the limitations of abstraction by striving to balance the conflicting goals of intelligibility and accuracy. They may also add sufficient information to a map to convey whatever additional complexity they feel is necessary to help the viewer more fully understand the problem at hand. In the case of uncertain information, they may even provide multiple maps of what Schweizer and Goodchild (1992) call data "quantity" and "quality." To address the problem of understanding disease risk, this paper presents a way of forecasting expected Lyme disease cases over time and space. The methods of bivariate mapping are used to portray distinct measures of risk and confidence so that data values and their quality are simultaneously represented.

The techniques presented here build upon recent cartographic developments. Monmonier (1992b) for example, uses two maps shown side by side, one portraying quantity, and the other quality, while in the area of disease mapping Pickle and Herrmann (1994) use colors to represent mortality and hatching to signify lack of confidence in data values. MacEachren et al. (1998) explore alternative symbology to facilitate the perception of data quantity and quality in maps depicting several variables manifesting varying degrees of spatial autocorrelation across scales. They explore three strategies: two distinct maps (one showing data values, the other showing data reliability); the use of color alone in a single map; and the simultane-

Lee De Cola, Research Physical Scientist, U.S. Geological Survey, 521 National Center, Reston, VA 20192, USA. E-mail: <ldecola@usgs.gov>. http://www.usgs.gov.

ous use of color and texture. Finally, Nelson (1999) demonstrates that map viewers are capable of distinguishing two dimensions of information within a single map; she also elaborates the important dimensions of representation—objects and their colors (Nelson 2000). This paper develops a straightforward variation on these approaches to bivariate mapping.

The paper is divided as follows. A formal treatment of spatial forecasting is presented first and, in the second section, Lyme disease and three key components of disease risk are introduced. The third section demonstrates how these components are analyzed for the United States as a whole and forecasted from the 1990s into the 21st century for each of the states. Section four illustrates the use of kriging to interpolate the forecasts over incidence and probability surfaces, which are then combined to make a color map in which hue signifies future incidence and saturation signifies confidence. The paper concludes with ideas for improvements and applications.

Spatial and Temporal Forecasting

Because the concept of health is so complicated and yet so fundamental to human welfare, it is important to communicate information about the spatial and temporal distribution of diseases. We therefore begin with a formal exposition of the problem of forecasting, which is the problem of modeling temporal and spatial patterns of future events; our expectation of the occurrence of the events; and our confidence in those expectations. The relationship between temporal and spatial forecasting is illustrated in the following discussion.

Let there be a set of locations and times $A = \{(x, t)\}$, and let Φ be a field having values everywhere in A. Let there also be a set of points $A_0 = \{(i, j): i = 1, \ldots, N; j = 1, \ldots, T\}$ at N locations and T time instants, and let $Z_0(i, j)$ be an $N \times T$ data matrix of values recorded at $A_0 \subset A$ (Knorr-Held and Besag 1998; De Cola 1994). In the case of a disease, for example, a fine-scale representation of the data might be an animated multifractal "dust" of infection events (x, t) in space-time (Mandelbrot 1983) or a detailed geographic timeline tracing the continuous location of each infected organism in space (Haggett 2000). Forecasting in general is the problem of assembling a structure of relationships and parameters into a predictive model whose forecasts $Z(x, t)$ (Lawson et al. 1999) should:

1. Behave much as the field Φ is expected to do: for example, be defined everywhere, bounded, continuous, and relate (although perhaps nonlinearly) to other fields (Earn et al. 2000);
2. Be "close" to the data Z_0 so that the errors $|Z_0(i, j) - Z(i, j)|$, which measure uncertainty (low confidence), will be small; and
3. Be visualizable in the sense that subsets of $Z(A)$ can be represented in a table, graphic, map, or animation.

Temporal forecasting in particular is a type of modeling that provides predictions Z_1 about Φ, particularly for times $t > T$, although the analysis will also examine, and may even predict, historical values as well (Makridakis and Wheelwright 1997):

$$Z_1(i, t) = F(Z_0(i, j)) \quad (1)$$

where Z_1 is a model encompassing not only future values at time t but also information about the quality of those values, such as standard errors (Venables and Ripley 1999). The related techniques of spatial forecasting develop models Z_2 of Φ over space, particularly at points (or regions as sets of points) where $x \neq i$:

$$Z_2(x, t) = G(Z_1(i, t)) = G(F(Z_0(i, j)) \quad (2)$$

This formal discussion, which is meant to add rigor to the analysis that follows, leads to several comments. The forecasting of spatial fields is intended to produce robust, continuous and comprehensible models $Z(A)$ that represent not only what are likely to be the values of $\Phi(A)$ but also our confidence in those forecasts. Risk analysis must therefore deal with what in point #2 above is represented by the difference between observed data and corresponding forecasts. From a practical standpoint, however, it is convenient to deal with temporal forecasting F( ) before spatial modeling G( ) because:

- Time is measured along a single dimension, whereas space has three dimensions;
- Temporal measurements are generally more regular because they are typically recorded at equal intervals; and
- Usually relatively fewer temporal than spatial measurements are made.

Nevertheless, because the fundamental principles of modeling autocorrelated measurements in time and space are similar (see Figure 3 below), it is possible to interpolate measurements over a region in space and then forecast each value (Haggett et al. 1977).

Analyzing Risk

Considerable effort is devoted to the development of geographical information systems that

increase understanding of public health problems, and in particular to collaborative efforts among clinicians, epidemiologists, ecologists, and geographers to map and forecast disease risk (Croner et al. 1996). But forecasting is a complicated process that involves keen judgment, especially when applied to problems of health. Since its origins in the 19th century, biomedical science has provided rapidly increasing information about disease (Foucault 1994), so that epidemiologists may use information about past disease incidence to forecast future illness. Less attention, however, is given to how little is known about future uncertainties, the effects of human/environment interactions, and especially the way complicated systems respond to interventions (Illich 1995). The increasing affluence of Western society and the proliferation of fragmented information may be creating what Beck (1999) calls a "risk society" in which people are increasingly preoccupied with threats to their well-being, many of which are new or merely newly apparent. Monmonier (1997), for example, provides insight into this idea by providing guidance in the mapping of "cartographies of danger," while the critical social scientist Mike Davis (1998) systematically studied a southern Californian "ecology of fear" arising from a growing preoccupation with tectonic, climatic, social, and epidemiological threats.

The Components of Risk

Since its founding in 1879, the U.S. Geological Survey (USGS) has acquired considerable expertise in the spatial analysis of risk. In 1998, USGS scientists began working with epidemiologists of the U.S. Centers for Disease Control and Prevention (CDC) to bring this expertise to bear on the problem of understanding the role of vector-borne diseases (illnesses carried by arthropods, such as plague, LaCrossse encephalitis and West Nile virus) in the health of regions and their organisms (Peterson and Rochrig 2001). My role in this research is to forecast disease case reports by modeling a space-time field Φ representing the spatial patterns of disease risk.

It is essential to untangle multiple components of health risk, as follows. Beginning with Kasperson's (1992) definition of risk as "threats to people and the things they value," I define our first risk component as the threatened event or hazard itself (Bernstein 1996). In the case of Lyme disease (described below), the operational definition of hazard is the event of presenting a "clinical case" to a medical professional who then reports it to a public health agency (CDC 1997). For the United States during the historical period 1990-2000, the magnitude of the Lyme disease hazard was represented by 140,381 officially reported cases during the 11 years of record (CDC 2002).

Kasperson goes on to discuss "the probability of experiencing harm," and so we need to consider as well the expectation of an event's occurring to someone at some time and place. For example, Lyme disease incidence is officially defined by the CDC as the probability that a randomly chosen U.S. resident was reported as being infected. During the 11-year period, this ratio was 140,381 cases / 262,089,000 people = 0.000536, and for any year within the 11-year period this expectation was 1/11 as much or 0.0000487 (Timmreck 1994).[1] Epidemiologists usually convert these small ratios into larger numbers by multiplying by 10^5 (or some other power of 10), which, in the present case, would be expressed as a yearly expectation of "4.87 cases per 100,000," but I prefer to work with the base-10 logarithm:

$$\text{INCIDENCE}_{US} = \log_{10}(\text{CASES}_{US}/\text{POPULATION}_{US}) = -4.31 \quad (3)$$

This index, which I shall simply call INCIDENCE and whose negative sign distinguishes it from more conventional measures (e.g., cases per 100,000), is a convenient representation of disease expectation, particularly in situations where, as with Lyme disease, there is a wide variation in values: INCIDENCE = –3, for example, means a 1-in-1000 expectation, while INCIDENCE = –7 means 1-in-10,000,000. This is the range we shall encounter in the analysis below; in comparison, U.S. motor vehicle occupant crash fatalities in 2000 had a comparable INCIDENCE of –3.89 (U.S. Department of Transportation 2002).

Finally, because we never confidently know the expectation of an event, we must add the third concept of confidence to the concept of risk. In the case of our basic example of national Lyme disease, although mean INCIDENCE for the decade was –4.31, during 1990-2000 this index had a standard deviation of 0.115 and ranged from –4.50 to –4.19. Thus, considering that this standard deviation represents a factor of 2 in the actual (unlogged) incidence, we can see that a considerable amount of quantifiable uncertainty adheres to the notion of Lyme disease risk during the decade, say, as represented by the standard error of yearly incidence $= {}^{0.115}\!/\!_{\sqrt{11}} = 0.0347$).

[1] Note that this ratio is not the expectation of actually acquiring the Lyme disease infection—which cannot be known precisely—but rather is based on specific diagnostic criteria.

It should be borne in mind that the uncertainty I analyze here is due only to temporal and spatial variation in the reported data and is therefore strictly empirical; these particular types of error would not exist were the data themselves to conform, for example, to perfectly linear trends over time and space. Beyond the temporal variation discussed here there are many other sources of uncertainty, just a few of which are feedback, seasonality, errors in diagnosis and reporting, and small number problems.

To sum up, the examination of data for the United States as a whole provides us with three risk components: the hazard itself (reports of Lyme disease); the incidence or expectation of the hazard ($10^{-4.31}$ for the U.S. as a whole); and whatever confidence may be associated with the event. This multidimensionality of risk highlights a major challenge of geographic information science—that of dealing with what Goodchild (1998) calls "uncertainty, the Achilles heel of GIS." Even after the first component is defined, any spatial representation of Lyme disease must take into account both "data quantity and quality" (Schweizer and Goodchild 1992), and each component must be treated separately before any synthesis is possible.

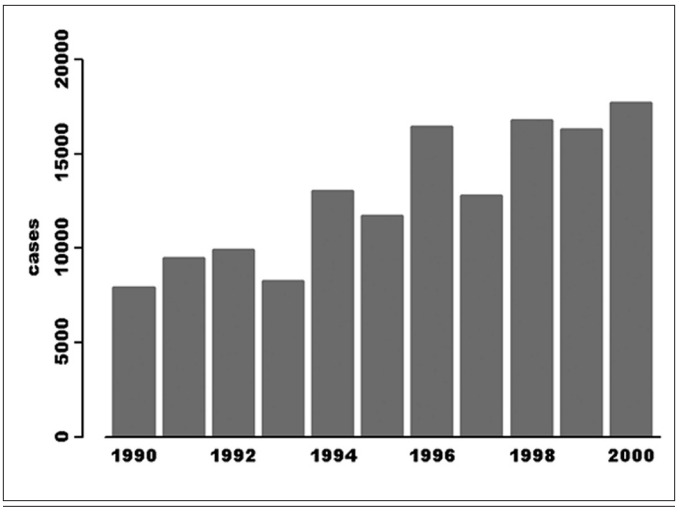

Figure 1. Lyme disease cases in the United States 1990-2000.

Lyme Disease

The application of the above ideas to Lyme disease is particularly appropriate because the illness is controversial, cases are difficult to characterize, and epidemiological information is fraught with the problems of misreporting and small numbers. There is considerable speculation about the epidemiological future of the disease; the only consensus is that incidence is rapidly growing (Rahn and Evans 1998) and that cures are difficult (Eppes 2001). The illness, which is caused by infection with the spirochete bacterium *Borrelia burgdorferi*, is the most commonly reported vector-borne disease in the United States (CDC 2001), where its "landscape epidemiology" is quite complicated. A person may be infected with the spirochete if they are bitten by an infected *Ixodes* tick, which itself acquires the spirochete from mice, but the ticks usually require deer as part of their life cycle. Given that all of the organisms in this process—bacteria, ticks, other invertebrates, humans, other mammals, and even reptiles—each require specific environmental conditions for living and breeding, it is obvious that mapping the spatial epidemiology of Lyme disease is challenging.

Figure 1 illustrates that Lyme disease cases in the United States show a generally increasing trend of about 1000 cases/year, but yearly changes in the data manifest a negative temporal autocorrelation that may be due to a cycling mechanism (since 1991, each yearly change has been followed by a change in the opposite direction). There is speculation that, in the absence of preventive interventions, incidence is likely to continue to increase because of widespread human/vector interaction, avian dispersal (Glavanakov et al. 2001), reforestation (Frank et al. 1998), and habitat fragmentation (Ostfeld et al. 1998).

The legend of Figure 2 shows that the statistical distribution of mean yearly INCIDENCE during 1990-2000 for the 49 contiguous states and the District of Columbia is reasonably symmetric, and so I have mapped the data using equally spaced interval breaks centered on each of the integers from –7 to –3 as shown. All spatial data have been transformed from geographic to kilometer coordinates using Albers equal-area conic projection with standard latitudes N 29.5° and N 45.5° and central longitude W 96° (Becker and Wilks 1993), which is used in the U.S. National Atlas (U.S. Department of the Interior 2002). The map reflects the fact that official Lyme disease cases are widespread, as all states save Montana reported at least one case during the period, and in no year did fewer than 44 states report cases. Yet INCIDENCE is obviously not spatially uniform but is, rather, concentrated in the Northeast and the Upper Midwest (Kitron 1998) with low levels in the South and Far West (Orloski et al. 2000).

Because the notion of risk as represented by INCIDENCE is arrayed on a continuum, and because the empirical data are symmetric as shown by the histogram of Figure 2, an appropriate color scheme is Brewer's (1994) concept of a "diverging" scale (her Figure 7.5c) in which two hues are used

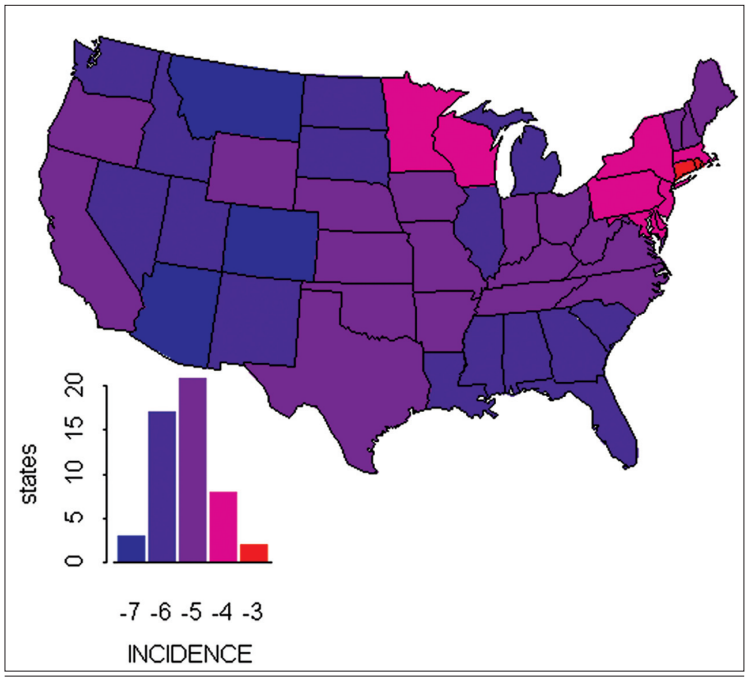

Figure 2. Mean state Lyme disease INCIDENCE 1990-2000.

to represent the endpoints. A fully saturated hue gamut is used to portray the low-to-high INCIDENCE continuum; blue values reassure (1-in-10,000,000) and red values alarm (1-in-1000), spanning four orders of magnitude. The color scheme for this map is presented in Table 1 as a hue spectrum from blue to red, where the choices have been adjusted to suit what I saw on a ViewSonic PT813 color computer monitor. Although this choice of complete saturation appears to ignore the modern cartographic canon recommending more subdued colors (Tufte 1997), saturation is here reserved as an independent color space dimension to be modified in the section below on mapping risk and confidence.

Forecasting Risk

National Forecasts

Although Figures 1 and 2 are useful visualizations of the temporal and spatial disease patterns for empirical data, our objective is the spatial forecast of a space-time risk field Φ represented by a surface providing INCIDENCE values at times beyond the empirical year 2000 and continuous across the United States (Webster et al. 1994). As was done above for the national decade data, INCIDENCE was computed for each year (Equation (3) above) and, because national data show a trend as well as negative autocorrelation, an autoregressive model was used to estimate the linear and yearly change in case counts (Makridakis and Wheelwright 1997). The forecasting system used to do this is outlined in Figure 3: a linear model was used to generate, for each year, a vector of trend predictions, residuals, and standard errors. For the U.S. CASES the linear equation was:

$$Z(t) = 7790 + 994t, \quad R^2 = 0.832 \qquad (4)$$

where the independent time variable t = (year – 1990), so that t = (0, 1, ..., 11) for ease of interpretation. This model shows that there was an increase of about 1000 cases/year during the period, and that a purely linear temporal model would forecast about 23,000 cases by the year 2005.

Although Equation (4) accounts for much of the year-to-year variation in the case data, analysis of residuals from the model shows a significant negative lag-1 autocorrelation = –0.698 (Makridakis and Wheelwright 1997), also seen in Figure 1. In order for unbiased forecasts to be generated, an autoregressive model was estimated on the residuals to provide predictions (added to the linear trend) as well as a further set of standard errors (added to the first set of standard errors, as shown in Figure 3). The autoregressive integrated system (ARI) that was estimated on the residuals from the linear equation uses first-order differences (degree I = 1) to predict current change from the changes of two prior time periods (autoregressive order AR = 2):

$$Z(t) - Z(t-1) = -1.08\,(Z(t-1) - Z(t-2)) - 0.279\,(Z(t-2) - Z(t-3)) \qquad (5)$$

The signs of the coefficients are due to negative autocorrelation among the residuals. Although the second term is not significant (standard error = 0.339), both terms are retained because this model often works well for the highly variable state data analyzed below. Note that Equation (5) may be simplified to:

$$Z(t) = -0.083\,Z(t-1) + 0.804\,Z(t-2) + 0.279\,Z(t-3) \qquad (6)$$

showing that the terms sum to 1 and that values two periods ago have the strongest influence on the current year.

Figure 3 shows how the linear and autoregressive models of Equations 4 and 6 are combined to give the forecasts of Table 2, which suggest that the numbers of U.S. Lyme cases may more than

INCIDENCE	RED	GREEN	BLUE	HUE
–3	1	0	0	red
–4	1	0	0.8	
–5	0.7	0	1	magenta
–6	0.5	0	1	
–7	0	0	1	blue

Table 1. Hue values for the maps.

double in the period 2000-2010. Moreover, the standard errors associated with these forecasts increase from 14 percent of the 2000 cases to over 18 percent of the 2010 cases, illustrating how confidence inevitably declines in the future. Moreover, by the central limit theorem, we can expect that relative standard errors for smaller regions (such as states) will be even larger than for the nation as a whole (Griffith and Amrhein 1991), making it even more critical to account for confidence in fine-scale forecasts. Adding the temporal linear and autoregressive case forecasts and dividing by U.S. population and projections (U.S. Bureau of the Census 2000), Equation (3) provides the yearly INCIDENCE results visualized in Figure 4, clearly indicating future increases, year-to-year fluctuations, and decreasing confidence.

Modeling Confidence

The forecasting system of linear and autoregressive models provides not only expected numbers of CASES, but also for each year, a STANDARD ERROR that we can use to evaluate the forecasts. Table 2, for example, reports that in 2005, the national cases are expected to be 22,633 with a standard error of 3,794. Given these two statistics, consider the (one-tailed) null hypothesis Ho that the nation will have zero Lyme cases in 2005. We can use the Student t-statistic (not to be confused with the time index above):

$$t_{US} = \text{CASES}_{US} / \text{STANDARD ERROR}_{US} = 5.97 \qquad (7)$$

with 7 (= 11 years – 4 coefficients) degrees of freedom to test this hypothesis (Griffith and Amrhein 1991). For example, given a 1 percent significance level, the fact that:

$$Pr(t > 5.97) = 0.0003 \qquad (8)$$

means that we can reject the hypothesis that the United States will have no Lyme cases in the year 2005. Now consider the probabilistic complement of the outcome shown in Equation (8):

$$Pr(t \leq 5.97) = 1 - 0.0003 = 0.9997 \approx 1 \qquad (9)$$

Year	Cases	Standard Error
1990	7869	
1991	9567	
1992	9896	
1993	8318	
1994	13016	
1995	11717	
1996	16332	
1997	12798	
1998	16639	
1999	16406	
2000	17735	
2001	18647	2563
2002	19595	2703
2003	20660	3285
2004	21590	3437
2005	22633	3794
2006	23591	3982
2007	24610	4261
2008	25587	4471
2009	26592	4718
2010	27578	4936

Table 2. U.S. Lyme disease cases, actual (1990-2000) and forecast (2001-2010).

which reflects our confidence in the alternative hypothesis that the nation will continue to have Lyme disease in 2005.

Although this probability argument lends strong confidence to INCIDENCE forecasts for the United States, for smaller regions we should not expect that numbers of cases will always be such large multiples of standard errors. Consider the typical state of Virginia (VA), whose 2005 combined (linear and autoregressive) forecast is CASES_{VA} = 133 which, with a projected 2005 population of 7,320,000 gives INCIDENCE_{VA} = -4.81 according to Equation (3). The forecasting model also provides a total $\text{STANDARD ERROR}_{VA}$ = 113 cases, which, when used in Equation (9) gives:

$$\text{PROBABILITY}_{VA} = Pr(t_{VA} \leq \text{CASES}_{VA} / \text{STANDARD ERROR}_{VA}) = 0.67 \qquad (10)$$

Equation (10) shows that for states (as opposed to the nation as a whole), the greater temporal variability of case data will likely produce forecasts in which we can have less confidence. Then, given the t-distribution with seven degrees of freedom, Equation (10) provides a confidence index operationally defined as:

$$\text{PROBABILITY} = Pr(t \leq \text{CASES} / \text{STANDARD ERROR}) \qquad (11)$$

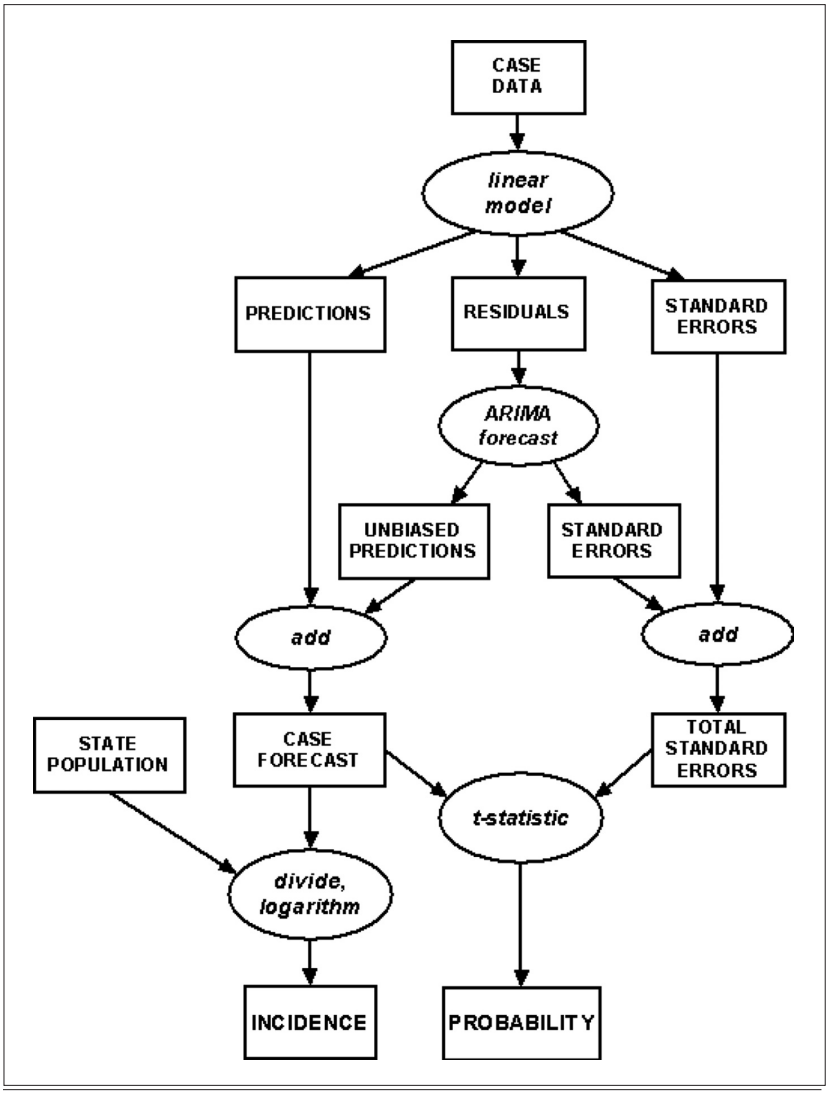

Figure 3. Simplified flowchart of the temporal forecasting process (data shown in rectangles, operations in ellipses).

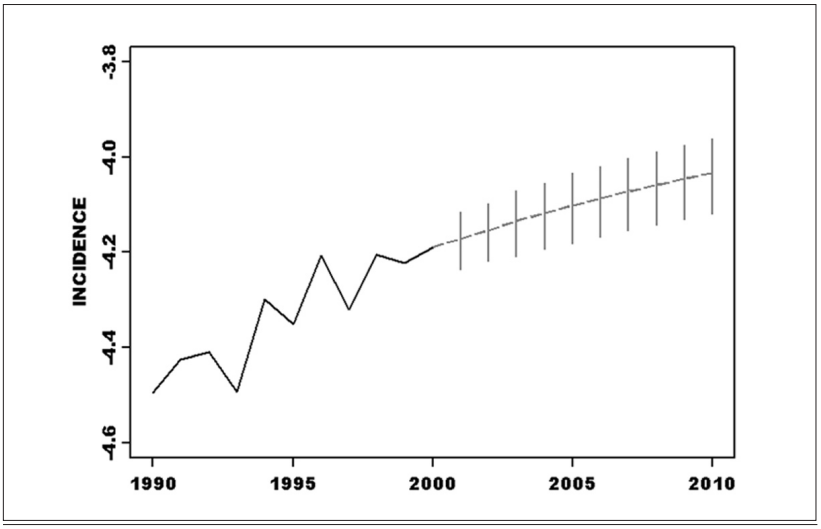

Figure 4. Lyme disease INCIDENCE1990-2000 actual and 2001-2010 forecast with ±1 standard error shown as vertical lines.

This confidence index ranges between 0 (no confidence) and 1 (virtual certainty), and we can associate it with the forecasted INCIDENCE for any region. If the forecasted number of cases is very much larger than the associated STANDARD ERROR, then we shall obtain a PROBABILITY value close to 1 and may, therefore, associate a high degree of confidence with the result. This single number summarizes not only temporal variability but also (because it is based on the analysis of the raw case numbers) the size of the region as a sample.

Before proceeding with the application of this argument to smaller regions, I note that the language of uncertainty is itself somewhat confusing. In what follows I use the word "confidence" to refer to a high expectation not only that the results from a model are close to real-world data (predictions have a low standard error), but also that future forecasts will be close as well. As our information becomes more reliable—or our forecasts more near-term—then the probability of rejecting a null hypothesis becomes more nearly certain, so it is intuitive to think of our confidence as increasing. Such a formulation reflects a positive correlation between expectations and language (as well as cartographic symbology—see below). Nevertheless, the term "uncertainty" has become common in geographic language, and my use of the term "more confidence" can always be translated into "less uncertain." MacEachren et al (1998), for example, uses the coefficient of variation (STANDARD ERROR/MEAN) as an uncertainty measure, whereas I prefer a related index that ranges between 0 and 1 to use in mapping forecasts.

The full case data matrix for the U.S. (states plus the District of

Columbia) consists of (N × T) = 51 regions × 11 years. When the temporal forecasting model of Figure 3 is applied to each of the regions we obtain 51 × 2 forecasts (INCIDENCE and PROBABILITY) for any year in the future, using the results of Equations (3) and (11). These indices are given in Figure 5 for each region in 2005, with our typical state of Virginia near the middle of the scatter.

Figure 5 shows how each of the PROBABILITY values (a function of the relative sizes of the case forecasts and standard errors) relates to each of the INCIDENCE values. The scatter confirms the logical expectation that higher PROBABILITY levels should be associated with higher INCIDENCE values (in 2005 R^2 = 0.807). Consider for example Connecticut, which has the highest forecasted INCIDENCE = –2.8 as well as the highest PROBABILITY = 0.9998. At the other end of this scatter is California, which—although it had the 10th largest number of cases during the 1990-2000—is forecasted in 2005 to have both a low forecasted INCIDENCE = –7.1 (because of its very large population) as well as a low PROBABILITY = 0.0 (because of the wide fluctuations in the historical data). Between these extremes are states such as Arizona, which are forecasted to have low INCIDENCE and high PROBABILITY values, as well as those like Delaware, which are likely to have high INCIDENCE but low PROBABILITY values.

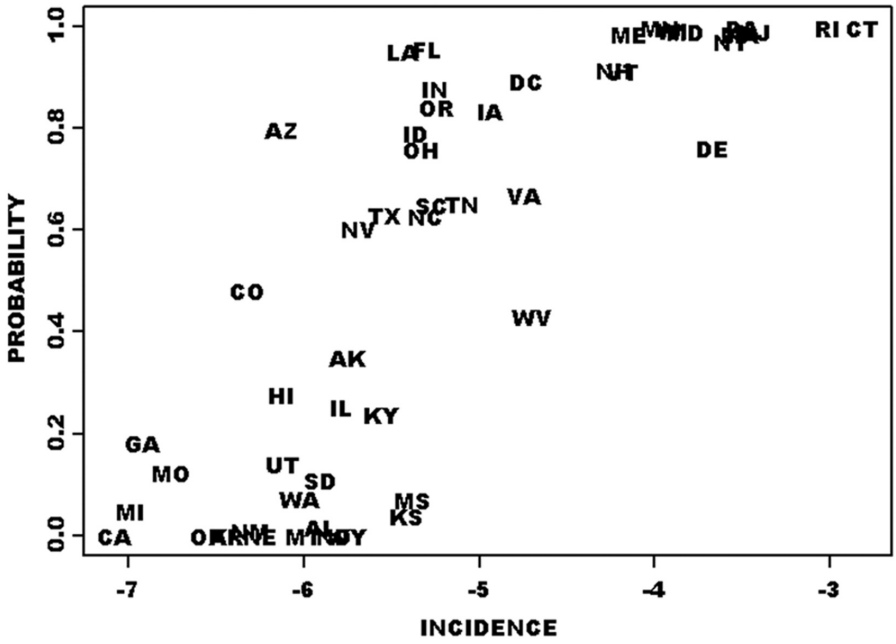

Figure 5. Scatter diagram of 2005 PROBABILITY v. INCIDENCE. Data points are labeled with state abbreviations; this enables the location of isolated values and outliers, but closely valued states will obscure one another.

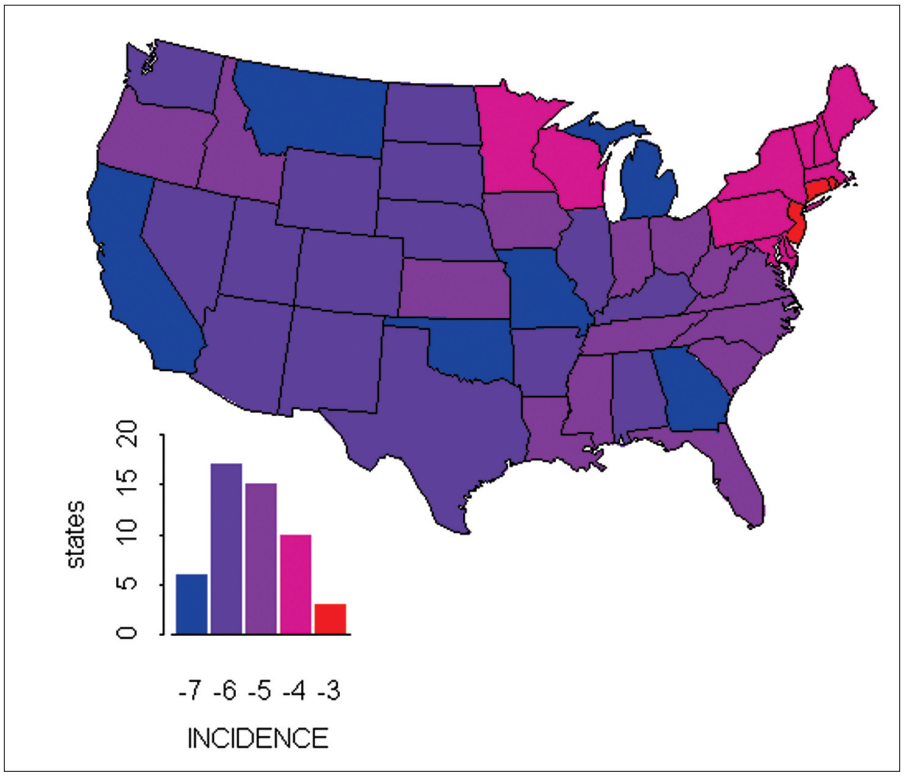

Figure 6. Forecasted state Lyme disease INCIDENCE 2005.

Four regions require further explanation before we proceed with the spatial analysis. Alaska and Hawai'i have been excluded from all the maps as well as from the subsequent spatial analysis because of their distance from the contiguous United States. The District of Columbia is regarded by the CDC for many purposes as a reporting unit, so

I too shall regard DC as a "state." Finally, Montana was the only state for which there were no cases during the historical period, and although linear regression will give predictions in such a situation, the autoregressive model will not, so that one Lyme disease case was introduced into the state in 1995 simply to force the model to work for all states. This resulted in a 2005 Montana forecast of INCIDENCE = –6 and PROBABILITY = 0.

Mapping Risk and Uncertainty

We now have two correlated but distinct components of risk that can be used to model where Lyme disease is likely to be found in the future. An obvious way to portray the values for states is a choropleth map such as Figure 6, which reveals similar patterns to those of the 1990-2000 data shown in Figure 2 ($R^2 = 0.850$). A comparison with the legend histogram of Figure 2 shows two changes. First, there are more states in the high-INCIDENCE classes as the disease intensifies in those regions where it is already known to be endemic. Second, there are more states in the low-INCIDENCE classes as reporting standards continue to become more stringent. Both of these changes reflect the fact that the mean standard deviation for INCIDENCE among states during the 1990-2000 period was 0.823 while the standard deviation for the forecasted year 2005 has risen to 1.09. Moreover, a comparison of the maps themselves appears to show a lower overall INCIDENCE level as areally large Western states show a general decline in the number of expected cases. We must keep in mind that these forecasts are based on official CDC data that reflect not only growing national INCIDENCE (Figure 4), but also increasingly stringent case definitions as well as widely varying state trends that result from spatiotemporal variation in the Lyme disease epidemic, modeled by the risk surface itself. It is variations in the latter two factors—as well as increases in standard errors that inevitably result from future forecasting—that largely account for wide differences among state INCIDENCE values as well as declining state PROBABILITY values.

Cartographic research suggests that well designed computerized choropleth maps such as Figures 2 and 6—particularly if carefully studied according to clear instructions (Slocum and Egbert 1993, Appendix C)—can communicate information about spatial patterns. But as visualizations of fields representing the quantity and quality of our forecasts, such maps have at least three shortcomings. First, choropleth maps may mask the spatial patterns of the "risk surface" with the details of the state boundaries (Kitron 2000). Second, although unclassed choropleth maps are possible, it becomes difficult to discern differences among regions if too many colors are used (Eicher and Brewer 2001). Third, a bivariate representation of INCIDENCE and PROBABILITY would require additional symbology, such as hatching (MacEachren et al. 1998), which is more appropriate for data exploration and finer-scale cluster detection than for the identification of broad patterns.

Modeling a Risk Surface

An alternative approach to bivariate mapping is to represent the changing INCIDENCE field using a surface-fitting technique. One way to distribute the INCIDENCE values over the 49 regions (contiguous states and DC) is a rather crude prism representation (SAS 1990) or an iterative pyncnophylactic interpolation (Tobler 1979), which is not widely implemented. Because we seek to model a risk field Φ, an alternative is to associate empirical surface values $Z(i)$ with centroids i (Equation (2) above), which can be done with any one of a number of surface-fitting techniques (polynomial, loess, triangular interpolation, spline), each with their own advantages and disadvantages (Venables and Ripley 1999).

The method used here is kriging (Carrat and Valleron 1992), in which spatial autocorrelation is modeled by first estimating a large-scale surface (trend, spline, polynomial), and then a kriged surface is generated from residuals from the first surface.[2] The kriging process generates a surface of predictions, as well as an error surface that estimates the extent to which the predictions can be expected to be close to the empirical measurements. The technique also provides explicit control over the nature of the fit: the so-called "range" parameter (beyond which spatial autocorrelation is assumed to no longer occur) determines local-scale fit, while "nugget" controls the influence of the large-scale global model (Kaluzny et al. 1998).

Once the centroids are determined (Becker and Wilks 1993), the next step in kriging is to examine the distribution of the variables to be mapped. Although the 2005 INCIDENCE forecasts fail a Kolmogorov-Smirnov goodness-of-fit test for nor-

[2] The overall approach is conceptually similar to the spatial autocorrelation modeling outlined in Figure 3 in which residuals from a linear model are used to estimate an ARIMA model for forecasting.

mality ($p = 0.006$) (Taylor 1977), the legend for Figure 6 shows that taking the logarithms of the data (Equation 3) provides 2005 forecasts that have neither outliers nor skewness. Kriging then entails the examination of spatial autocorrelation among observations within distance bands (Ripley 1981). Figure 7a is a correlogram of the INCIDENCE data with a 95 percent confidence level, indicating significant spatial autocorrelation among values up to about 600 kilometers. The figure also shows the exponential curve used to model this autocorrelation, using the empirically determined range = 600 km (Atkinson and Tate 2000). A plane (order-1 polynomial surface) was then used to account for this large-scale variation, with the following coefficients:

$$\text{INCIDENCE} = 0.104 + 0.596\,\text{EAST} + 0.767\,\text{NORTH}, \; R^2 = 0.57 \qquad (12)$$

where the directional variables are in units of 1000 km. The model predicts a statistically significant increase in disease risk to the northeast, as we expect from the maps. When this trend model is used to generate residuals we get a reduction in autocorrelation at all distance bands, although significant autocorrelation remains at 100 km (Figure 7b).

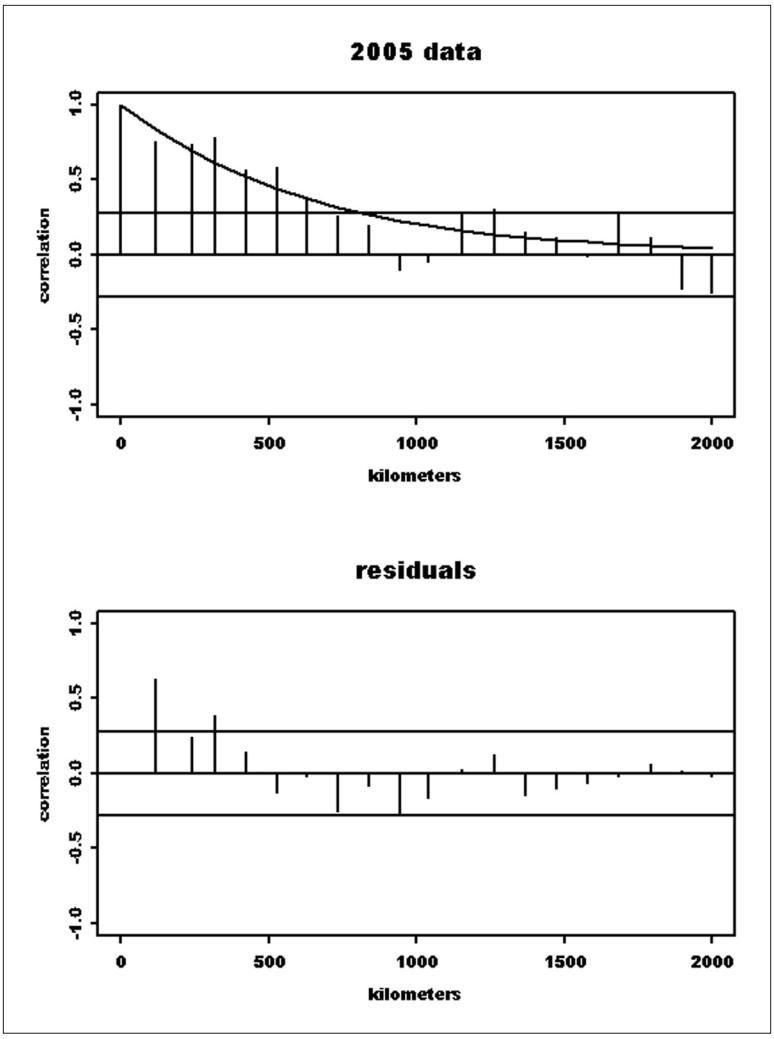

Figure 7. 2005 spatial correlograms; 7a. raw data, 7b. residuals from an order-1 surface fit.

Because of its flexibility and rigor, kriging requires a number of parametric choices. I have chosen a range of 600 km (based on the correlogram) and a nugget of 1/16 (based on visual results) but the selection of resolution of the resultant surface presents interesting opportunities. The S-PLUS kriging system was modified to generate surface predictions at multiple grid resolutions while maintaining approximately square grid cells. The advantage of this configuration is that fast results can be tested and visualized at a low resolution (26 rows × 39 columns of cells, each of which is 118 km square, see Figure 8) and then map-quality graphics can be produced at a higher resolution (209 rows × 314 columns of 15 km cells).

Three surfaces from the kriging analysis are shown in Figure 8 for a grid of 26 × 39 cells. Figure 8a is the fitted INCIDENCE surface clearly showing the influence of the trend model (values rising to the northeast), the spatial complexity introduced by fitting the covariance model to the data, and the overall smoothing effect of the technique. Figure 8b shows the standard errors of the INCIDENCE fit, illustrating, as expected, that the model fits closest near the state centroids and diverges beyond the hull of the dataset (i.e., the national boundary). Although the kriging could be performed within the national boundaries, I have chosen to model the rectangular study region and then later mask the extra-national results. Finally, Figure 8c is the krige model applied to the forecasted state PROBABILITY values, suggesting that confidence is highest where forecasted INCIDENCE is highest, but, as was shown in Figure 5, the association is not perfect.

Bivariate Mapping

For any year t and location x in A we now have forecasts $Z(x, t)$ of expected Lyme disease INCIDENCE and PROBABILITY values representing our confidence in those expectations (Figures 8a and

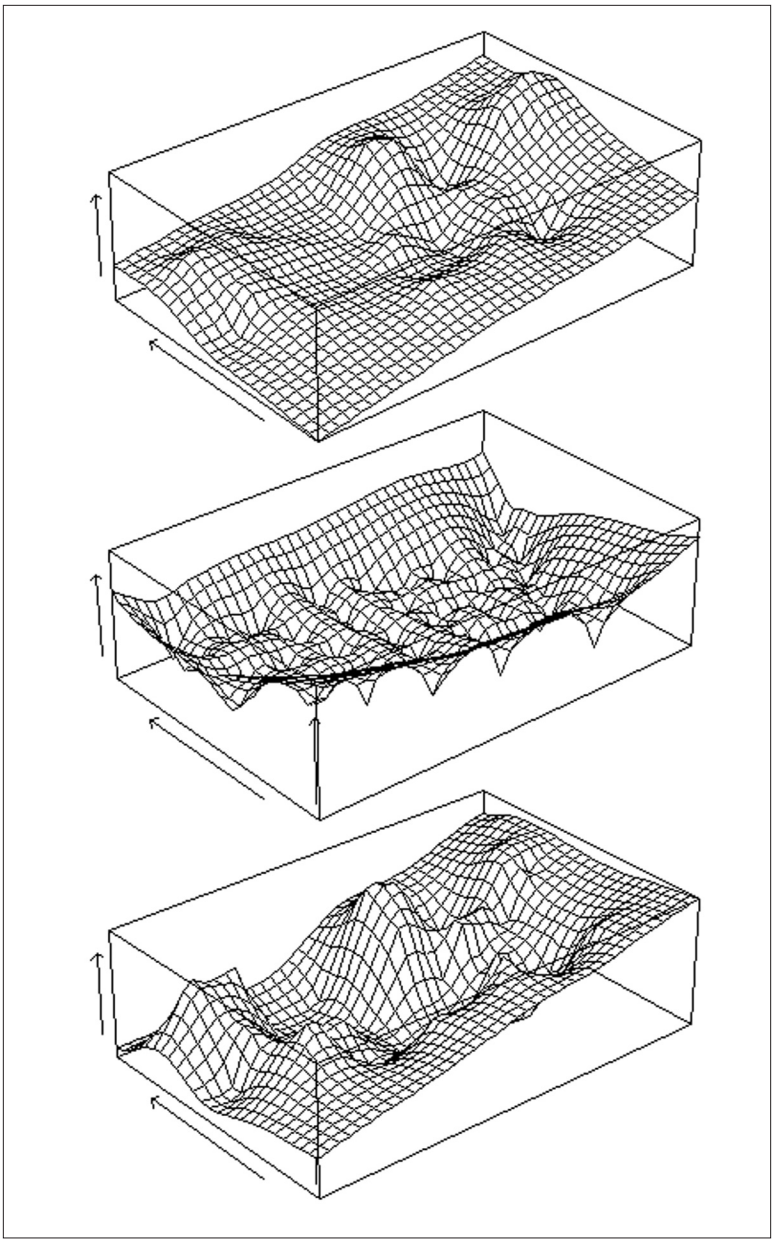

Figure 8. Kriged surfaces for 118-km cells in 2005: 8a. INCIDENCE index, 8b. standard errors from INCIDENCE model, 8c. PROBABILITY index.

c, respectively). These two indices vary in time and space, and surfaces representing their fields can be generated for any desired spatial resolution. Several approaches to the bivariate mapping of data quantity and quality were discussed in the introduction. Color is often regarded as having three degrees of freedom: for example, red/green/blue for computer graphics; yellow/cyan/magenta in printing; and hue/saturation/value in artistic rendering. I have chosen to map using the independent hue and saturation dimensions of the third color space (Foley et al. 1990). The five discrete hue values from the choropleth maps (see Table 1) are used in Figure 9a to map INCIDENCE within the 14,000,000 km^2 rectangular kriged region *A* and the smaller 7,800,000 km^2 of the conterminous states.

Figure 9b is a greyscale rendering of the PROBABILITY surface, which ranges in value between 0 and 1. In this illustration, smaller values of PROBABILITY (less confidence in INCIDENCE values) are rendered in white and larger values (more confidence) are rendered in black. The values of these grid cells will be used to drive the saturation dimension of the finished map; when green is added to the blue/red values in inverse proportion to PROBABILITY, the grid cells become more nearly white, conveying an impression of uncertainty in the values at a location. This approach may be related to the findings of Schweizer and Goodchild (1992), who attempted to communicate data quantity and quality using other dimensions of color space. Their results were ambiguous for a number of reasons, including subjects' lack of familiarity with confidence concepts; difficulties in intuiting relationships among quantities, qualities, and color space dimension; as well as the inherent subtlety of the bivariate problem. Nevertheless, the research appeared to reinforce the feeling among map viewers that "dark" colors convey the notion that there is "more" of something. In response to this finding, I represent confidence using saturation, in which darker colors denote more confidence in forecasts.

Figure 10 outlines a technique that weights the fully saturated blue-red hue of INCIDENCE with the values of PROBABILITY, using a "double negative" approach that also adds a national mask and state boundaries (Jasc Software 1998). The final map is shown in Figure 11 in which the two dimensions of INCIDENCE and PROBABILITY are rendered simultaneously, allowing the viewer to understand not only where Lyme incidence is low or high (along the blue-to-red hue dimension) but how much confidence we may place in a spatial forecast (along the unsaturated-to-saturated dimension).

The map may be interpreted by examining patterns along a scale hierarchy, beginning with national trends and ending with patterns evident from the

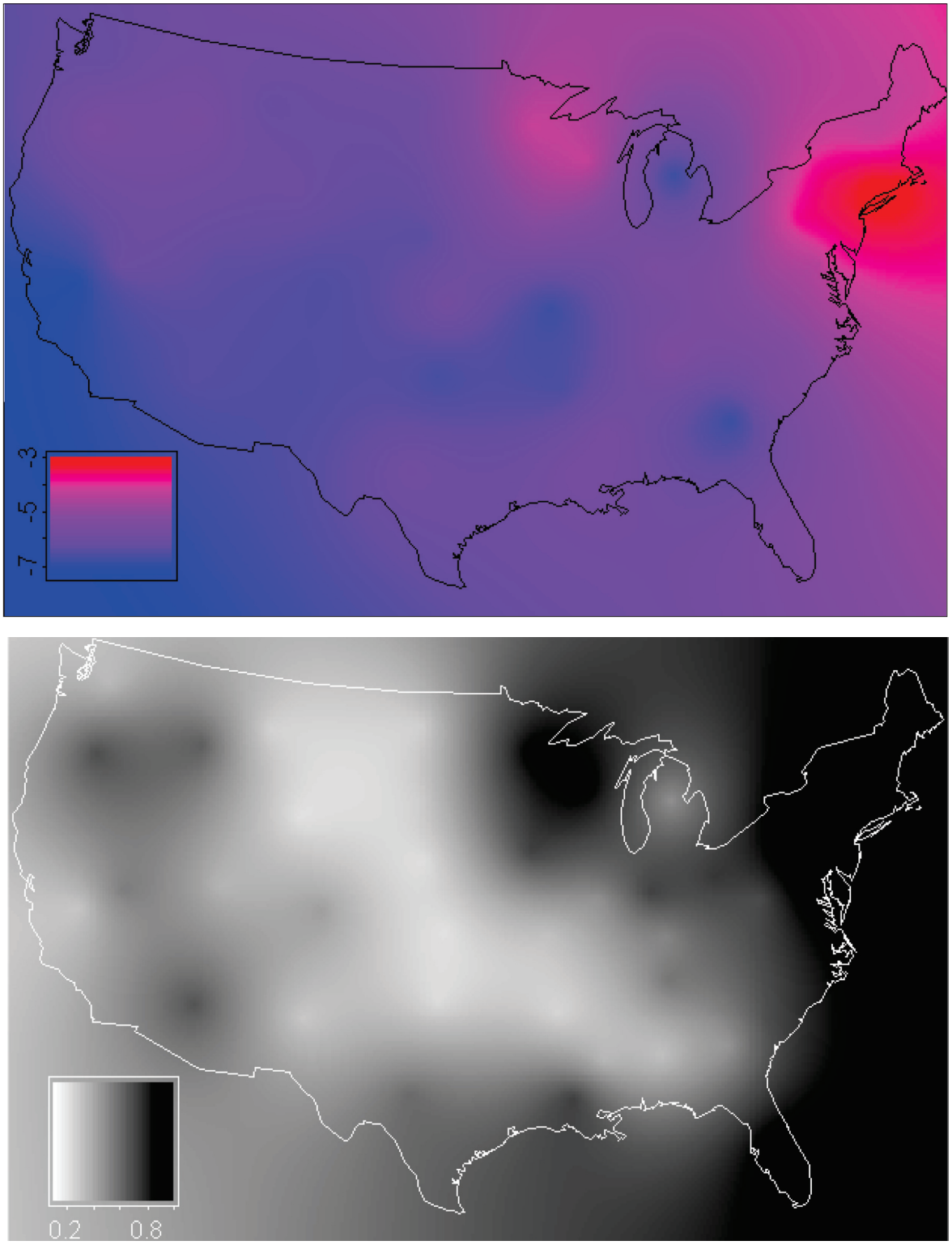

Figure 9. Kriged surfaces for 2005: 9a. INCIDENCE using a red-blue hue spectrum, 9b. PROBABILITY using a white-black grey scale.

values generated by states within smaller multistate regions. Some of the patterns we see are:

- A northeast INCIDENCE trend, represented by the order-1 trend model;
- High values of INCIDENCE in New England and the Upper Midwest and low values in the Southwest;
- A band of low PROBABILITY from Montana/Dakotas to the Deep South; and

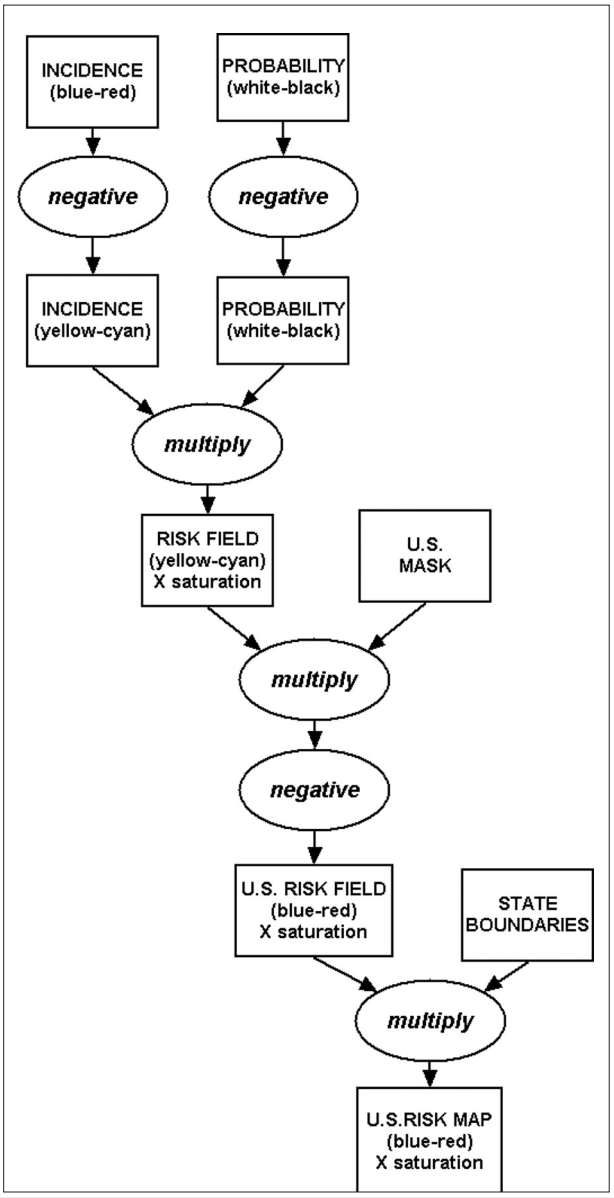

Figure 10. Flowchart outlining the steps from two grids to a bivariate risk map. The "negative" operation reverses greyscale (white to black) or hue: (blue to yellow), etc.

- Spatial outliers highlighted by low PROBABILITY values in Michigan, Georgia, and California.

Figure 11 is a significantly different approach to risk visualization from the choropleth map in Figure 6. Drawing attention to extreme values and areas of low confidence, the continuous map does not portray misleadingly sharp differences between regions (such as that between the Dakotas and Minnesota, an artifact of state boundaries) but renders the smooth increase in expected INCIDENCE that would likely be encountered in a transect from west to east across the United States. This continuous visualization is a more faithful representation of the underlying national risk field modeled on the basis of state-level resolution, and although we have no data outside of the U.S., the surface representation also suggests that high values of Lyme disease could be expected to be found in eastern Canada (Estrada-Pena 1999).

Even with its limitations, the addition of a data quality measure improves our probabilistic understanding of the INCIDENCE field. For example, although the map of Figure 9a implies low expected disease INCIDENCE in the Plains states, the addition of PROBABILITY reduces our confidence in those values.

In sum, a bivariate surface map allows us to perceive national and regional patterns without the distractions of political boundaries between uniform values and discrete classifications among values. Most importantly, I introduce a risk field Φ as a continuous surface whose height reflects INCIDENCE forecasts. I then represent this field as a 2-dimensional map whose color is also controlled by a confidence measure that reflects temporal variation. Building upon this idea, a more elaborate analysis would have this confidence index include other factors not analyzed here, such as data quality, environmental influences, and geostatistical error (Figure 8b).

Discussion

The spatial forecasting system described in this paper is based on the sequential modeling of temporal and spatial autocorrelation of a hazard in order to map both the value of expected incidence as well as the probability associated with each value. Obviously a map such as Figure 11 cannot substitute for the original data table or for the explicit association that a choropleth map provides between a state and a range of values. As a single representation of disease risk, however, this spatial forecast is useful for making rapid interpretations of patterns, and particularly in making comparisons:

- Among diseases, as in an atlas contrasting, say, cancers, respiratory illness, and infectious diseases (Devesa et al. 1999);
- Among time periods, as in a surveillance system seeking unusual patterns, or in an animation of past and future risk surfaces (Peterson 1999);
- Among regions with extremely wide variations in risk levels characterized by disparate levels of forecast confidence, as was done here.

This paper has presented detailed results for Lyme disease, but the system can be generalized to map multiple levels of a disease hier-

archy (National Library of Medicine 2000), from broadly defined bacterial infections, through arthropod-borne illnesses, to narrow disease categories such as Lyme or West Nile Virus (for a broader discussion of these issues see Mennis et al. (2000)). Moreover, because the system provides forecasts for any historical or future year, it is possible to animate these maps, so that a viewer might select a disease and a period to examine. The viewer of such an interactive atlas could move through time, across space, or among diseases, seeking comparative foci, unusual patterns (Kuldorff 2001), and even regions requiring improved data or resolution (Monmonier 1992a). Lyme disease is an excellent test of the approach, given the wide range of INCIDENCE (4 orders of magnitude among states) and the degree of controversy surrounding its diagnosis, treatment, and forecasting (Talleklint-Eisen and Lane 1999). This study shows not only that incidence is obviously widespread (Figure 2) and increasing (Figures 4 and 6), but also that the state trends vary greatly (Figure 5), so that the national pattern is obscured by areas of low confidence (Figure 11).

Data or Parameters	Choices
Data matrix	49 states x 11 years
Population projection model	loess
INCIDENCE TRANSFORMATION	log(base 10)
Temporal trend	linear
ARIMA model	lag 2, difference 1
Projection parameters	Albers, latitudes, longitude
Trend surface order	1 (plane)
Krige parameters	Nugget, range, resolution
Hue	Endpoints (blue, red), equalization
Saturation	Endpoints (white, black)

Table 3. Models, parameters and choices.

Improvements

There are, of course, several avenues for improvement in the system. First, it must always be borne in mind that any disease analysis based solely on U.S. data is likely to be incomplete; consequently, Lyme disease data from Canada (and possibly Mexico) will be added to the research. Second, a focus on counties (Glass et al. 1995), census tracts, and even point locations will certainly show finer-scale patterns, but this entails a loss of temporal extent because county Lyme disease data have not been compiled for the entire 1990-2000 period studied here, and in any case are not public information. Third, more detailed patterns will reveal the operation of various spatial dynamics, such as anisotropic spread or diffusion, land cover fragmentation, and population migration. Finally, this effort to understand spatio-temporal patterns (Rogerson 2001) will profit as well from the incorporation of additional geographic influences, including environmental factors (weather, climate, vegetation), host and vector distributions (Guerra et al. 2002), the spatial distribution of genetic patterns (Qiu et al. 2002), and rapidly changing human influences on the region at multiple scales.

Yet there is a wider range of cartographic issues that remain to be explored. Two of the three dimensions of color space have been used to portray data quantity and quality, and many other choices can be made to link less/more qualities in the data with up/down and left/right orientations in the legends and with blue/red or black/white or dark/bright symbolization in the map (Schweizer and Goodchild 1992). A central issue unexplored in this paper is linking model results with maps on the computer screen as well as on paper—an extremely complicated issue subject to infinite variation (Foley et al. 1990).

Although this paper has been preoccupied with the development of the final map, the approach is quite flexible, as shown by Table 3, which lists most of the major elements of the system as well as the choices made to perform the Lyme analysis. For the analysis of a decade of yearly, state-level Lyme disease case data for the U.S., I have attempted to make parsimonious choices. A more complex phenomenon (finer-resolution spatial or temporal data, seasonality, other explicit sources of uncertainty) would obviously require a more elaborate approach. Some of the model and parameter selections shown in the table may be made either more interactive or less subject to inspection and choice; and the system must more smoothly integrate data management, analysis (multivariate, temporal, and spatial), GIS/mapping, and cartography/symbology/graphics. Working from a data matrix to a finished bivariate map required the use of Excel and S-PLUS (with its sophisticated analytical spatial tools coupled to a crude GIS capability (Becker and Wilks 1993)), as well as the low-level image algebra tools (multiplication, mask) of a computer graphics package. Although this first research effort focused on the analysis of error due to temporal change, error estimated in the spatial modeling (Figure 8b) may also be incorporated into the final map.

Envisioning Risk

Cartographic improvements are also possible. Although I present a complete approach to the visualization of spatial forecasts of disease risk, the bivariate mapping technique developed here needs to be assessed for its communication effectiveness within the broad framework established by MacEachren et al. (1998), who suggest that an integrated approach to color alone, even when deconstructed into hue, saturation, and intensity, may not be the optimal way to communicate quantity and quality, at least when multiscale patterns of spatial autocorrelation are being described. Alternatives to the scheme presented here must be tested on the basis of more rigorous semiotic principles.

Finally, the analysis also needs to investigate the always critical dimension of scale. Fully multiscale approaches to analysis and visualization are resulting in fundamental improvements in geography (Atkinson and Tate 2000). Autoregressive kriging analysis, however, is based on such mid-20th century mineral exploration preoccupations as range, nugget, sill and trend analysis—concepts that need to be translated into the new language of multiscale complexity that increasingly informs contemporary geographic information science (Quattrochi and Goodchild 1996). The flexibility of this spatial forecasting system inherently lends itself to a broad understanding of what was described above as the first dimension of risk, that of hazard itself. The problems studied here can in theory be examined as realizations of a multifractal process unfolding at many biological scales (Margulis and Sagan 1995): from the level of organs (spirochetes in tissue) through that of organisms (ticks on deer) to that of organization (populations experiencing (re-)emerging diseases in habitats). The maps presented here reveal just a few aspects of this process, modeled as a scaling risk field.

This research may be broadly considered as a contribution to medical imaging, which is making rapid progress in its use of graphical and cartographic techniques (Hall 1993). Just as computer-aided tomography (CAT scanning) aids the diagnosis of a brain exposed to chemicals, so may risk maps describe the changing health of a region under varying anthropogenic stress. Tools are therefore needed to allow analysts to smoothly zoom in on finer-scale patterns and unknowns, while minimizing the artifacts of political boundaries, optimizing interpolation parameters, and visualizing uncertainty (Young 1998). It is critical that we manage and visualize space-time information with due regard for when and where risk is well known so that decisions may be based on quality information and uncertainty is reduced where it is highest (Kelmelis 2000). Current tools for mapping are often too cumbersome to support productive dialogue with data.

The results presented in this paper advance the geographic management of spatiotemporal uncertainty. When data are subject to errors of measurement and under- or over-reporting, when small numbers obscure important patterns, or when rapid change (as in biodefence) make the detection of sensitive events important (Henderson 1999), spatial forecasting can be used not only to detect emerging patterns but also to warn us where Type I errors (the imputation of pattern where none may exist) are likely. We must also keep in mind that on the scale of global risk, Lyme disease—however painful its consequences for individual sufferers—pales in comparison with ancient and new global illnesses (Kleinschmidt et al. 2000) and environmental problems (Meade and Earickson 2000). As humanity increasingly overwhelms the global environment, the need to understand risk and uncertainty in time and space has never been more urgent.

ACKNOWLEDGMENTS

Among the many people who have assisted me during the very long period that this research was produced, I thank colleagues at the U.S. Centers for Disease Control and Prevention (especially Ned Hayes and Chuck Croner) and from the Geographic Analysis of Disease project at the U.S. Geological Survey.

REFERENCES

Atkinson, Peter M., and Nicholas J. Tate. 2000. Spatial scale problems and geostatistical solutions: A review. *Professional Geographer* 52(4): 607-23.

Beck, Ulrich. 1999. *World risk society*. Oxford, England: Blackwell.

Becker, Richard A., and Allan R. Wilks. 1993. *Maps in S*. AT&T Bell Laboratories, Murray Hill, New Jersey.

Bernstein, Peter L. 1996. *Against the Gods: The remarkable story of risk*. New York, New York: Wiley.

Brewer, Cynthia. 1994. Color use guidelines for mapping and visualization. In: Alan M. MacEachren and D. R. F. Taylor (eds), *Visualization in modern cartography*. Oxford, England: Elsevier Science Ltd. pp. 123-47.

Carrat, Fabrice, and Alain-Jacques Valleron. 1992. Epidemiologic mapping using the "kriging" method: Application to an influenze-like illness epidemic in France. *American Journal of Epidemiology* 135(11): 1293-300.

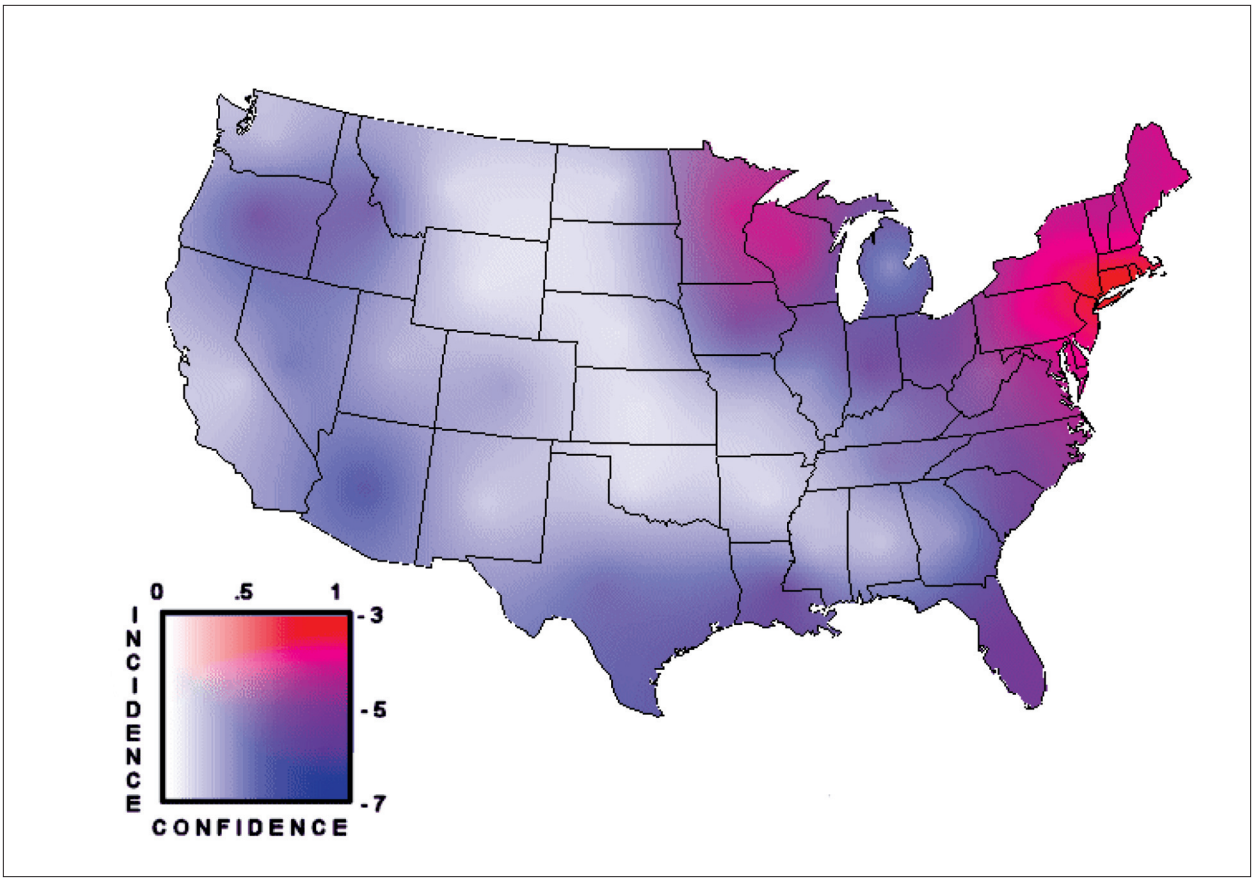

Figure 11. Lyme disease in 2005; INCIDENCE shown on a blue-red hue spectrum and PROBABILITY on a saturation gradient.

CDC. 1997. Case definitions for infectious conditions under public health surveillance. *Morbidity and Mortality Weekly Report* 46(RR10):1-55.

CDC. 2001. Lyme disease—United States, 1999. *Morbidity and Mortality Weekly Report* 50(10): 181-5.

CDC. 2002. Lyme disease—United States, 2000. *Morbidity and Mortality Weekly Report* 51(02): 29-31.

Croner, Charles M., Jonathan Sperling, and Frederick R. Broome. 1996. Geographic information systems (GIS): New perspectives in understanding human health and environmental relationships. *Statistics in Medicine* 15:1961-77.

Davis, Mike. 1998. *Ecology of fear*. New York, New York: Henry Hold.

De Cola, Lee. 1994. Simulating and mapping spatial complexity using multiscale techniques. *International Journal of Geographical Information Systems* 8(5): 411-27.

Devesa S. S., D. G. Grauman, W. J. Blot, G. Pennello, R. N. Hoover, and F. Fraumeni Jr. 1999. *Atlas of cancer mortality in the United States, 1950-94*. National Institute of Health, Washington, D.C.

Earn, David J. D., Pejman Rohani, Benjamin M. Bolker, and Bryan T. Grenfell. 2000. A simple model for complex dynamical transitions in epidemics. *Science* 287: 667-70.

Eicher, Cory L., and Cynthia A. Brewer. 2001. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Systems* 28(2): 125-38.

Eppes, Stephen C. 2001. Lyme disease: Current therapies and prevention. *Infections in Medicine* 18(8): 388-95.

Estrada-Pena, Agustin. 1999. Geostatistics and remote sensing using NOAA-AVHRR satellite imagery as predictive tools in tick distribution and habitat suitability estimations for *Boophilus microplus* (Acari: Isodidae) in South America. *Veterinary Parasitology* 81: 73-82.

Foley, James D., Andries van Dam, Steven K. Feiner, and John F. Hughes. 1990. *Computer graphics: Principles and practice,* 2nd ed. Reading, Massachusetts: Addison-Wesley.

Foucault, Michel. 1994. *The birth of the clinic: An archaeology of medical perception*. New York, New York: Random House.

Frank, Denise, Durland Fish, and Fred H. Moy. 1998. Landscape features associated with Lyme disease risk in a suburban residential environment. *Landscape Ecology* 13: 27-36.

Glass, Gregory E., B. S. Schwartz, J. M. I. Morgan, D. T. Johnson, P. M. Noy, and E. Israel. 1995. Environmental risk factors for Lyme disease identified with geographic information systems. *American Journal of Public Health* 85(7): 944-8.

Glavanakov, Stephan, Dennis J. White, Thomas Caraco, Andrei Lapenis, Geogre R. Robinson, Boleslaw K. Szymanski, and William A. Maniatty. 2001. Lyme disease in New York State: Spatial pattern at a regional scale. *American Journal of Tropical Medicine and Hygiene* 65(5): 538-45.

Goodchild, M. F. 1998. Uncertainty: The Achilles heel of GIS? *Geo Info Systems* 8(11): 50-2.

Griffith, Daniel A., and Carl G. Amrhein. 1991. *Statistical analysis for geographers*. Englewood Cliffs, New Jersey: Prentice Hall.

Guerra, Marta, Edward Walker, Carl Jones, Susan Paskewitz, M. Roberto Cortinas, Ashley Stancil, Louisa Beck, Matthew Bobo, and Uriel Kitron. 2002. Predicting the risk of Lyme disease: Habitat suitability for *Ixodes scapularis* in the north central United States. *Emerging Infectious Diseases* 8(3).

Haggett, Peter. 2000. *The geographical structure of epidemics*. Oxford, U.K.: Oxford University Press.

Haggett, Peter, Andrew Cliff, and Allan Frey. 1977. *Locational analysis in human geography*. London, U.K.: Edward Arnold.

Hall, Stephen S. 1993. *Mapping the next millennium: How computer-driven cartography is revolutionizing the face of science*. New York, New York: Vintage Books.

Henderson, Donald A. 1999. The looming threat of bioterrorism. *Science* 283: 1279-82.

Illich, Ivan. 1995. *Limits to medicine*. London, U.K.: Marion Boyars.

Jasc Software. 1998. *Paint Shop Pro 5, user's guide*. Minnetonka, Minnesota: Jasc Software, Inc.

Kaluzny, Stephen P., Silvia C. Vega, Tamre P. Cardozo, and Alice A. Shelly. 1998. *S+Spatial Stats user's manual for Windows and UNIX*. New York, New York: Springer.

Kasperson, Roger E. 1992. The social amplification of risk: Progress in developing an integrative framework. In: Sheldon Krimsky and Dominic Golding, *Social Theories of Risk*. Westport, Connecticut: Praeger, p. 153.

Kelmelis, John A. 2000. Flood damage, risk, and levees in a changing environment. *Technology* 7: 21-31.

Kitron, Uriel. 1998. Landscape ecology and epidemiology of vector-borne diseases: Tools for spatial analysis. *Journal of Medical Entomology* 35(4): 435-45.

Kitron, Uriel. 2000. Risk maps: Mapping transmission and burden of vector-borne diseases (Comment). *Trends in Parasitology* 16(8):325.

Kleinschmidt, I., M. Bagayoko, G. P. Y. Clarke, M. Craig, and D. Le Sueur. 2000. A spatial statistical approach to malaria mapping. *International Journal of Epidemiology* 29: 355-61.

Knorr-Held, Leonhard, and Julian Besag. 1998. Modelling risk from a disease in time and space. *Statistics in Medicine* 17: 2045-60.

Kuldorff, Martin. 2001. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society, A* 164(1): 61-72.

Lawson, Andrew B., Annibale Biggeri, Dankmar Bohning, Emmanuel Lesaffre, Jean-Francois Viel, and Roberto Bertollini. 1999. *Disease mapping and Risk Assessment for Public Health*, New York, New York.

MacEachren, Alan M. 1994. *Some truth with maps*. American Association of Geographers, Washington D.C.

MacEachren, Alan M., Cynthia A. Brewer, and Linda W. Pickle. 1998. Visualizing georeferenced data: Representing reliability of health statistics. *Environment and Planning: A* 30:1547-61.

Makridakis, Spyros G., and Steven C. Wheelwright. 1997. *Forecasting: Methods and applications*.New York, New York: John Wiley & Sons.

Mandelbrot, Benoit B. 1983. *The fractal geometry of nature*. New York, New York: Freeman.

Margulis, Lynn, and Dorion Sagan. 1995. *What is life?* Berkeley, California: University of California.

Meade, Melinda S., and Robert J. Earickson. 2000. *Medical geography*. New York, New York: Guildford.

Mennis, Jeremy L., Donna J. Peuquet, and Liujian Qian. 2000. A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* 14(6): 501-20.

Monmonier, Mark. 1991. *How to lie with maps*. Chicago, Illinois: University of Chicago Press.

Monmonier, Mark. 1992a. Authoring graphic scripts: Experiences and principles. *Cartography and Geographic Information Systems* 19(4): 247-60.

Monmonier, Mark. 1992b. Summary graphics for integrated visualization in dynamic cartography. *Cartography and Geographic Information Systems* 19(1): 23-36.

Monmonier, Mark. 1997. *Cartographies of danger: Mapping hazards in America*. Chicago, Illinois: University of Chicago.

National Library of Medicine. Medical subject headings. [http://www.nlm.nih.gov/mesh/meshhome.html]. Accessed August 22, 2000.

Nelson, Elisabeth S. 1999. Using selective attention theory to design bivariate point symbols. *Cartographic Perspectives* 32: 6-28.

Nelson, Elisabeth S. 2000. Designing effective bivariate symbols: The influence of perceptual grouping processes. *Cartography and Geographic Information Systems* 27(4): 261-78.

Orloski, Kathleen A., Edward B. Hayes, Grant L. Campbell, and David T. Dennis. 2000. Surveillance for Lyme disease—United States, 1992-1998. *CDC Morbidity and Mortality Weekly Report* 49(SS03): 1-11.

Ostfeld, Richard S., Felicia Keesing, Clive G. Jones, Charles D. Canham, and Gary M. Lovett. 1998. Integrative ecology and the dynamics of species in oak forests. *Integrative Biology* 1:178-186.

Peterson, Lyle R., and John T. Rochrig. 2001. West Nile virus: A reemerging global pathogen. *Emerging Infectious Diseases* 7(4).

Peterson, Michael P. 1999. Active legends for interactive cartographic animation. *International Journal of Geographical Information Science* 13(4): 375-83.

Pickle, Linda W., and Douglas J. Herrmann. 1994. The process of reading statistical maps: The effect

of color. *StatLib—Statistical Computing and Graphics Newsletter* 5(1).

Qiu, Wwi-Gang, Daniel E. Dykhuizen, Michael S. Acosta, and Benjamin J. Luft. 2002. Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Isodes scapularis*) in the northeastern United States. *Genetics* 160: 833-49.

Quattrochi, Dale A., and Michael F. Goodchild. 1996. *Scale in remote sensing and GIS*. Boca Raton, Florida: Lewis.

Rahn, D.W., and J. Evans (eds). 1998. *Lyme disease*. Philadelphia, Pennsylvania: American College of Physicians.

Ripley, Brian D. 1981. *Spatial statistics*. New York, New York: John Wiley.

Rogerson, Peter A. 2001. Monitoring point patterns for the development of space-time clusters. *Journal of the Royal Statistical Society, A* 164(1): 87-96.

SAS, Inc. 1990. *SAS/GRAPH Software*. Cary, North Carolina: SAS, Inc.

Schweizer, Diane M., and Michael F. Goodchild. 1992. Data quality and chropleth maps: An experiment with the use of color. In: *GIS/LIS 1999 Conference Proceedings*, San Jose, California. pp. 686-99.

Slocum, Terry, and Stephen L. Egbert. 1993. Knowledge acquisition from choropleth maps. *Cartography and Geographic Information Systems* 20(2): 83-95.

Talleklint-Eisen, L., and Robert S. Lane. 1999. Variation in the density of questing *Ixodes pacificus* (Acari: Ixodidae) nymphs infected with Borrelia burgdorferi at different spatial scales in California. *Journal of Parasitology* 85(4): 824-31.

Taylor, Peter J. 1977. *Quantitative methods in geography*. Prospect Heights, Illinois: Waveland Press.

Timmreck, Thomas C. 1994. *An introduction to epidemiology*. Boston, Massachusetts: Jones and Bartlett.

Tobler, Waldo R. 1979. Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association* 74(367): 519-30.

Tufte, Edward R. 1997. *Visual explanations: Images and quantities, evidence and narrative*. Cheshire, Connecticut: Graphics Press.

U.S. Bureau of the Census. 2002. Projections of the total population of states: 1995 to 2025. [http://www.census.gov/population/projections/state/stpjpop.txt]. Accessed July 2002.

U.S. Department of the Interior. 2002. *National atlas of the United States*. [http://nationalatlas.gov/].

U.S. Department of Transportation. 2002. *Fatality Analysis Reporting System (FARS)*. [http://www-fars.nhtsa.dot.gov/]. Accessed January, 2002.

Venables, W. N., and B. D. Ripley. 1999. *Modern applied atatistics with S-PLUS*. New York, New York: Springer-Verlag.

Webster, R., M. A. Oliver, K. R. Muir, and J. R. Mann. 1994. Kriging the Logal risk of a rare disease from a register of diagnoses. *Geographical Analysis* 26(2): 168-85.

Young, John D. 1998. Underreporting of Lyme disease (Letter). New *England Journal of Medicine* 338(22): 1629.