Spatial Statistics When Locations Are Uncertain

Geoffrey M. Jacquez

BioMedware, Inc.
516 North State Street,
Ann Arbor, MI 48104-1236 USA.

Abstract

Spatial statistics quantify spatial pattern and identify local and global departures from null spatial models. As part of Exploratory Spatial Data Analysis they play a critical role in the evaluation of spatial pattern, and in the formulation of hypotheses to explain spatial pattern. While all spatial data have imprecise locations, the magnitude of this imprecision can vary dramatically from one measuring instrument to another and from one study to another. When does location uncertainty impede our ability to quantify spatial pattern? This paper describes credibility-based spatial randomization tests that propagate location uncertainty through proximity metrics and into spatial statistics. Credibility is a flexible new approach to spatial randomization tests, but is not a panacea. It applies to spatial statistics that incorporate measures of geographic proximity (*e.g.* spatial adjacency, weight, nearest neighbor relationship, distance *etc.*). It uses Monte Carlo sampling to generate the null distribution, and not distribution theory, as classical statistics do. It is a technique for testing hypotheses regarding spatial pattern, and is best described as a method for Exploratory Spatial Data Analysis. It is meant to complement, not replace, traditional spatial statistics that use P-values and alpha levels. In conjunction with these techniques it forms a quantitative basis for evaluating the likely impact of location uncertainty on one's ability to make statistical decisions with spatial data.

I. INTRODUCTION

Webster's dictionary defines uncertainty as 'lack of certainty, doubt'. Uncertainty ranges from a simple lack of sureness regarding a precise value (e.g. uncertainty about the location of a place of residence) to inherent vagueness (e.g. who knows what the future may bring). Uncertainty can occur both in locations (location uncertainty) and in observations (attribute uncertainty). Research to date has dealt primarily with uncertain attributes (Goodchild and Gopal 1989, Heuvelink, Burrough et al. 1989) and error propagation through map operations (Haining and Arbia 1993). Techniques for dealing with attribute uncertainty in statistical analyses are well developed (e.g. Viertl 1996). In contrast, methods for assessing location uncertainty and its impact on outcomes such as spatial statistics have received little attention. Some research has dealt with location uncertainty and the calculation of lengths and areas (Keefer, Smith et al. 1991). Altman (1994) presents a fuzzy theoretic approach for representing location uncertainty, and Jacquez (1996) used fuzzy set theory to develop disease cluster tests for imprecise locations. Using a probabilistic approach, Kiiveri (1997) presents a technique for assessing uncertainty in the locations of points and lines. Such probabilistic approaches can result in many potential realizations of a spatial data surface (see Goodchild, Sun et al. 1992, Ehlschlaeger and Shortridge 1996, and references therein). This has motivated research on techniques for visualizing uncertainty, including animation (Ehlschlaeger, Shortridge et al. 1997). To date little, if any, research has dealt with the issue of propagating location uncertainty through spatial statistical analyses, and the consequences of location uncertainty on statistical inference. Jacquez (2000) addresses this issue in detail, and portions of his work are presented here to give an overview of this new approach. In a related paper, Jacquez and Jacquez (1999) detail the mathematical forms of several location models, and how they may be used to propagate location uncertainty in tests of disease clustering. Jacquez and Jacquez (1999) also describe how location models are sampled within Monte Carlo algorithms.

Statistical inference for uncertain locations is the topic of this paper. It introduces spatial data and issues of statistical inference. It describes the components of statistical inference from spatial data—the test statistic, null spatial model, reference distribution, and alternative spatial model. These are placed in the context of traditional statistics, the sampling space, and randomization tests. The use of a general form, called the gamma product, for representing spatial statistics is presented, along with the concept of location uncertainty. Finally, the notion of credibility-based statistics is introduced.

II. SPATIAL DATA

Spatial data may be represented for convenience as the matrix

1082-4006/99/0502-77$5.00

$$\mathbf{Z} = \begin{bmatrix} x_1 & y_1 & z_{11} & \cdots & z_{p1} \\ & & \cdot & & \\ & & \cdot & & \\ & & \cdot & & \\ x_n & y_n & z_{1n} & \cdots & z_{pn} \end{bmatrix}. \quad (1)$$

Here x_i,y_i is the coordinate of location i, and $z_{1i},\ldots,z_{pi}$ are observations p variables at that location. There are n locations and p variables.

Example 1: Childhood leukemia in North Humberside

62 cases of childhood leukemia were observed in North Humberside, England, between 1974 and 1986. 141 matched controls were sampled from a population registry for the corresponding period (Cuzick and Edwards 1990). The first four rows of the **Z** matrix are:

Easting	Northing	Case/control identity
4882	4420	1
5153	4300	1
5088	4318	1
5147	4654	1

The x, y coordinates are decimeters east (Easting) and north (Northing) of an artificial origin in the south of England. Case/control identity is coded as a '1' if the observation is a case, and as a '0' if it is a control. There are n=203 locations, and p=1 variables.

Characteristics of Spatial Data

By their very nature spatial systems are usually large and the phenomena being investigated often take place on relatively long time scales. For example, changes in forest composition can take decades or even centuries to evolve. In addition, spatial systems are often difficult to manipulate. For these reasons designed experiments on spatial systems are often difficult to accomplish, and spatial data from these systems tend to be observational, uncertain, taken from a limited sampling space, and autocorrelated. What do these characteristics imply?

The scientist's control over a system varies, from laboratory settings that are carefully controlled, to natural systems that are not. Experimental data are collected from a system that is manipulated in order to control covariates and/or to perturb the system by introducing materials or energy into the system. *Observational* data are collected by passive observation rather than by designed experiments. In general, spatial data are observational because they are collected from natural systems that are not manipulated or controlled by the observer.

Uncertainty implies partial ignorance of a measurement's true value. This contrasts with variability, which represents a system's inherent heterogeneity. While uncertainty may be reduced by more careful measurement, variability is characteristic of the system under study and is not reducible. Spatial data are uncertain in two ways: locations may be uncertain (location uncertainty), and the values observed at those locations may be uncertain (attribute uncertainty). Although uncertainty may be reduced by refined measurement techniques, it is always present because our measurement instruments are imprecise.

Sampling space refers to the population or hypothetical universe from which samples are drawn, and is a critical concept for two reasons. First, it influences how one designs a scheme to sample the population of interest. Second, inferences drawn from statistical analyses apply to this sampling space. An overly restricted spatial sampling space will cause us to unnecessarily limit the scope of our statistical inferences. The spatial sampling space for geographic systems is over-restricted whenever information describing the spatial distribution of the study population is limited or is not incorporated into statistical tests.

Spatially autocorrelated means the observed values are not independent of one another, and that this lack of independence is a function of geographic proximity between sample locations. Positive spatial autocorrelation occurs when nearby locations tend to have similar values, and may be caused by common history, causal relationships with other variables that are themselves autocorrelated, and interaction (*e.g.* exchange of material) among adjacent locations. When not accounted for spatial autocorrelation can bias statistical tests that assume independent observations. It also can provide clues to the underlying space-time processes that produced the observed spatial pattern.

As with most generalizations, these characteristics apply to a greater or lesser extent depending on the application. Controlled agricultural field trials can be replicated and are more experimental than observational. Studies of short-lived phenomena in small spatial systems are more easily replicated. Uncertainty is always present but is decreased by more accurate and precise measuring instruments. The underlying population in some studies may be rigorously defined and carefully sampled. Spatial autocorrelation is almost always present, but its strength varies considerably from one kind of variable to another. The key lesson is that these characteristics influence our ability to conduct statistical inference with spatial data.

Space-time Processes

Almost all spatial data are the result of space-time processes. Examples include forest composition, which is the result of succession; spatial disease patterns, which are the result of epidemics; and pollutant plumes, which are the result of geochemistry, sub-surface flow and diffusion. For many systems change in a variable's spatial pattern occurs on a longer time frame than our ability to observe, and spatial data often represent 'snap shots' in time of a single realization of a space-time process. A realization is defined as a specific instance of a process. This recognizes the role of natural variability in the evolution of spatial pattern. For example, take two identical patches of barren ground and record vegetation composition through time. Although both initial conditions and successional sequences are essentially the same, the observed spatial patterns in the two plots will differ because of natural variability in intra- and inter-specific interactions (competition, predation, symbiosis), deposition of wind-borne seeds and so on. Species composition in the two plots is said to be two realizations of a common successional process. The key concepts are first, that spatial data are realizations of space-time processes, and second, that spatial patterns can provide clues to their generating processes.

III. STATISTICAL INFERENCE FOR SPATIAL DATA

Models of process, models of data, and ESDA

Models of process, models of data, and Exploratory Spatial Data Analysis (ESDA) require different amounts of knowledge and have different goals.

Models of process describe systems in terms of their basic processes or mechanisms. Parameters of models of process quantify biological and/or physical attributes of the system and are readily interpretable in terms of the underlying generating process. One example is compartmental models (Jacquez 1996). These models require a detailed understanding of the space-time processes giving rise to the observed spatial pattern.

Models of data are constructed to fit a particular data set without explicit reference to the system's basic mechanisms. They usually are fitted directly to an observed spatial pattern. Econometrics, kriging and regression models are examples of this genre. Models of data are useful for prediction, but have limited application beyond the particular data set. In addition, their parameters often are not directly interpretable in terms of the underlying generating process. Of course models of process and models of data are not exclusive, and a given model may blend them.

ESDA seeks to identify spatial pattern with the objectives of (1) quantifying spatial pattern, and (2) suggesting hypotheses regarding the underlying processes. Our knowledge of the system under study is often limited, and sample sizes may be too small to support construction of models of data. Examples include spatial autocorrelation analysis, and methods of spatial point pattern analysis. ESDA may be a first step in constructing a model of data, and the hypotheses generated may eventually form the basis of a model of process. This paper is concerned with statistical inference to support ESDA. Its contribution is in testing hypotheses regarding spatial pattern, with two principle goals. First, to quantify spatial pattern, and second to infer past generating processes that gave rise to the pattern.

A distinction is made between statistical hypotheses (e.g. the null hypothesis, the alternative hypothesis), that are statements regarding a variable's sampling distribution, and scientific hypotheses, that have to do with the fundamental questions we wish to answer. When formulating a statistical test it is essential to begin with the scientific hypotheses, and to then specify the corresponding statistical hypotheses. This assures an investigation is driven by questions related to our conceptualization of the world around us, rather than by a method's statistical hypothesis. In general, our scientific hypotheses have to do with the underlying space-time processes, while statistical hypotheses are concerned with spatial patterns. Thus there is not a direct correspondence between scientific and statistical null hypotheses, and careful attention is required to assure a statistical test is appropriate to the scientific query.

Components of spatial statistical inference

Spatial data usually represent partial knowledge of a spatial system at one or a few points in time. Because of this it is difficult or impossible to directly evaluate hypotheses regarding space-time processes. Instead, we explore spatial pattern in the hopes of gaining insights into the processes that produced the pattern. How does one construct a test to quantify spatial pattern? Six components were given by Waller and Jacquez (1995).

1. The null spatial model, describing the spatial distribution of the variables expected in the absence of the alternative spatial model (see below). The null model defines the null distribution of any proposed test statistic.
2. The null hypothesis, usually expressed in terms of parameters of the null spatial model.
3. The test statistic, a data summary whose

distribution under the null spatial model can be computed or found from tables (such as chi-squared tables).

4. The null distribution of the test statistic. This distribution is obtained either theoretically or empirically through Monte Carlo simulation. Both the theoretical derivation and the simulation procedure must be consistent with the null spatial model. Probability values (P-values) under the null hypothesis are obtained by comparing the value of the test statistic with that of the null distribution.
5. The alternative hypothesis, stated in terms of parameters of the null spatial model or in terms of additional parameters used to model the alternative spatial pattern. The distribution of the test statistic under the alternative hypothesis is different than its distribution under the null hypothesis, which enables probabilistic assessments.
6. The alternative spatial model, which may be an omnibus "not the null spatial model" or a more specific model describing the alternative spatial pattern. An example of the latter would be a model where persons near a hazardous waste site experience an elevated disease rate.

These components constitute an explicit framework for specifying a test for spatial pattern. The alternative spatial model may be poorly specified (e.g. 'not the null hypothesis') or it may be chosen to correspond to a more specific spatial pattern (e.g. high values near a specific location). I emphasize that it describes an alternative *spatial pattern*, and is neither a model of data nor a model of process. By evaluating alternatives describing spatial pattern, it may be possible to refine one's thinking and generate plausible hypotheses regarding underlying space-time processes.

Classical statistical inference

Spatial data analysis and classical statistical inference differ to some extent in their theoretical backgrounds. Haining (1990) observed that classical statistics assume data from designed experiments that can be replicated, and samples drawn from a hypothetical universe defined by a sampling space. The inference framework is based on comparison of a test statistic calculated for a sample to the distribution of the statistic under the null hypothesis for this sample space (the reference distribution). A distribution of the test statistic can be obtained by replicating the experiment. Within this framework type I error (α) is the probability of rejecting the null hypothesis when it actually is true, and type II error (β) is the probability of accepting the null hypothesis when it actually is false. Statistical power—the probability of correctly rejecting the null hypothesis—is 1-β.

Randomization tests and statistical inference

Methods of classical statistical inference assume experimental data and are not strictly appropriate for observational spatial data. As a result, randomization tests have gained currency, but at a substantial lost of robustness: inference applies only to the sample, as now described.

Manly (1991) provides a succinct description of randomization tests and their relationship to classical statistical theory. The data are from only one sample, the concept of a 'designed experiment that can be replicated' does not apply, and classical statistical inference therefore is inappropriate. A commonly used alternative is randomization tests, which determines whether pattern exists in a *sample*. The null hypothesis is that any pattern is a chance occurrence, and the alternative hypothesis is that 'true' pattern exists. Some statistic, Γ, is selected that quantifies the pattern of interest. The value, Γ^*, from the observed data is then compared to a reference distribution obtained by repeatedly reordering the data at random, and by calculating Γ for each repetition. The significance level of Γ^* is the proportion of the reference distribution that is as large or larger than Γ^*. Interpretation of this significance level is similar to conventional tests based on the classical model: if less than or equal to the α level (usually 5%) the null hypothesis of 'no pattern' is rejected. Manly further observed that randomization tests have two principle strengths: They are valid even without random samples, and non-standard test statistics may be used. These advantages have led to the wide use of randomization tests for the analysis of spatial data. However, results pertain only to the sample, and this single sample is the sampling space upon which the reference distribution is based.

Spatial statistics in randomization tests

The gamma product of two *nxn* matrices, **A** and **B** is:

$$\Gamma = \mathbf{A} \otimes \mathbf{B} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij} . \qquad (2)$$

For spatial data we rewrite the gamma product as

$$\Gamma = m\Delta \otimes \mathbf{D} = C \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij} d_{ij} . \qquad (3)$$

Here m is a constant (scalar), n is the number of locations, δ is a measure of geographic proximity measure (*e.g.* adjacency, nearest neighbor measure, distance or weight) and d is calculated from the observations on z (as defined in Eqn. 1). Several authors have shown that many commonly used spatial statistics are special cases of the Γ product (Haining 1990 pg 230, Marshall 1991, Wartenberg and Greenberg 1990, Getis 1992, Jacquez 1996, Jacquez

and Jacquez 1999). Mantel's test (1967) for space-time interaction results when δ_{ij} and d_{ij} are elements of distance matrices. Cuzick & Edwards test (1990) results when $\delta_{ij} = 1$ if location j is a nearest neighbor of location i (otherwise it is 0), and $d_{ij} = 1$ when both observations i and j are disease cases (if 1 or both of them is a control $d_{ij} = 0$). Moran's I (1950) results when $d_{ij} = (z_i - \bar{z})(z_j - \bar{z})$ and δ_{ij} corresponds to elements of a weight matrix. The join-count statistic (Cliff and Ord 1981) obtains when $d_{ij} = (z_i z_j)$ and δ_{ij} is the adjacency between areas i and j. Here variable z is binary, with a '1' indicating a labeled area. Pearson product-moment correlation and multiple regression may be written in Γ form (Smouse, Long et al. 1986), as can local autocorrelation statistics (Anselin 1995, Getis and Ord 1996). These examples illustrate the flexibility of the gamma product in quantifying a broad spectrum of statistical tests.

One can use a normal approximation for the randomization distribution of gamma (see Mantel 1967 and Haining 1990 for moments of this distribution) to assess statistical significance of an observed value. This approach has been criticized (Mielke 1978, Faust and Romney 1985) and it is better to calculate the distribution under randomization, and then compare the observed value to this distribution (Manly 1991). This is accomplished under a statistical null hypothesis of independence (Cressie 1991 terms this Complete Spatial Randomness or CSR) between the d_{ij} and the δ_{ij} using a randomization equivalent to a relabeling so the zi are sprinkled at random across the locations. Given $z=(z_1,...,z_n)$ values on a map, spatial randomization tests permute the z values over the sample locations. There are two limitations of randomization tests conducted in this fashion. First, the spatial sampling space is defined to consist solely and entirely of the sample locations, and second, inference applies only to the sample.

The spatial sampling space and statistical inference

To summarize, when spatial data are observational inference is often undertaken within the framework of an exploratory data analysis whose purpose is to detect structure and pattern. In these instances randomization tests are frequently used because the assumptions of classical statistical inference no longer apply.

While randomization tests may be appropriate when the experimental design justifies randomization testing, they can be problematic for spatial data because they take the sampling space to be the locations at which the observations were made. That is, spatial randomization tests erroneously assume the universe of locations to consist entirely and solely of the sample locations. In most situations we could have sampled other locations in the study area, but spatial randomization tests based only on the sample locations ignore this fact. This means the sampling space is incorrectly specified, and the reference distribution pertains only to the sample, and not to the population extant within the study area.

Example 2: The spatial sampling space and Mantel's test

Imagine we conduct two replicates of an experiment that introduces an infectious disease into a population, and we record case locations and times as the epidemic evolves. Figure 1 shows two realizations of this disease process. Locations of place of residence of the at-risk population are shown as filled circles, squares around circles indicate case locations. The locations of the at-risk population is the same in both realizations, but case locations change because of inherent variability in the contagious process.

Now consider how Mantel's test is typically applied to each realization. Mantel's test is sensitive to space-time interaction that arises when nearby cases occur at about the same time, a pattern that may be caused by a contagious agent or common exposure. The test regresses the waiting times on the corresponding spatial distances between pairs of cases, and the standardized test statistic is the correlation between spatial distance and waiting time. The reference distribution is generated under randomization by sprinkling the times of case occurrence over the case locations, and by then calculating the test statistic for each randomization. Notice this randomization test uses only the cases in each realization—the locations indicated by squares—as the sampling space for the reference distribution. Because of this any statistical inference applies only to the sample, and not to the population.

How would the statistic have been implemented under the classical statistical paradigm? The reference distribution under the null hypothesis would be generated by repeating the experiment many times under the corresponding null model, calculating the test statistic for each realization, with each realization

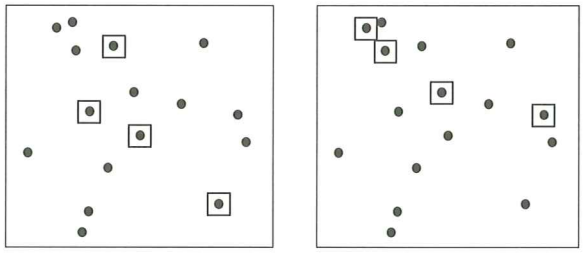

Figure 1. Two realizations of a spatial epidemic.

taken from the at-risk population. The classical approach treats all of the at-risk population as potential cases, and because of this any statistical inference applies to the entire population, and not just the cases in one realization.

Example 2 illustrates how spatial randomization tests can overly restrict the spatial sampling space, leading to unnecessarily weak inference structures. In epidemiology, spatial randomization tests, when structured as illustrated in the example, make little sense because they assume the at-risk population consists only of the sample. Until recently this issue has largely been ignored because of difficulties in specifying the underlying populations spatial distribution, and, in incorporating such information into the null spatial model and reference distribution. Recent advances in spatial knowledge (such as GIS) and computing power have eased these difficulties. The approach described later in this paper makes use of these advances to specify the spatial sampling space so that it corresponds more closely to the at-risk population, rather than to just the sample.

Consider the at-risk population in example 2. Until recently data describing the geographic distribution of human populations have not been available, precluding specification of the universe of possible sample locations. This is no longer the case. Spatially referenced data are now available describing the global population density distribution within 5' quadrilaterals (Tobler, Deichmann et al. 1995), and census data coupled with address matching software can locate street addresses within an accuracy of 100m (Rushton and Lolonis 1996). Our approach uses location models and such spatial population data to specify the spatial sampling space. This can be viewed as a step towards the stronger inference model of classical statistics because it recognizes that samples could have been taken at other locations in the study area (a sampling experiment). This effectively specifies the study's spatial sampling space and, not surprisingly, improves our ability to correctly detect spatial pattern.

Location uncertainty and statistical inference

The previous sections described some of the implications of the observational characteristic of spatial data, and how this can result in a restricted sampling space in spatial randomization tests. Recall the third characteristic of spatial data is *uncertainty*. Here we concern ourselves with the notion of location uncertainty. For our present discussion it is sufficient to distinguish between uncertain locations, which are the uncertain coordinates (denoted x_U, y_U obtained using an imperfect measurement instrument, and precise locations, which are the precise coordinates (denoted x_P, y_P) that would have been obtained using a perfect measurement instrument. We can only observe uncertain locations, precise locations exist in theory only. Of course, the amount and importance of location uncertainty varies depending on our measurement instrument and the spatial scale of the study. The key lesson is that in practice our point locations are *always* uncertain to some degree. We wish to account for location uncertainty in spatial statistical tests.

How does statistical inference work when locations are precise? Consider some test statistic (Γ) and suppose we have access to a perfect measurement instrument. The test statistic is denoted Γ_P, and is the test statistic based on precise locations. The reference distribution under the null hypothesis, g_P, is obtained by sprinkling observations over the sample locations in a manner consistent with the null hypothesis. A P-value for the test is the probability, under the null hypothesis, of obtaining a value of the statistic as large or larger than the observed, written $P(\Gamma_P \geq \Gamma_P^*)$ (Figure 2). Here Γ_P^* denotes the observed value of the test statistic, while Γ_P is the test statistic under the null hypothesis. A decision criteria based on a type I error level (usually $\alpha = 0.05$, so there is a 5% chance of rejecting the null hypothesis when it actually is true) is used to evaluate the test. When $P(\Gamma_P \geq \Gamma_P^*) \leq \alpha$ the null hypothesis is rejected and the alternative is accepted, otherwise the null hypothesis is accepted. In practice all measurement instruments are imprecise, locations are uncertain, and the inference mechanism for precise locations shown in Figure 2 does not apply. In particular, it fails to account for location uncertainty within the statistical inference structure.

What happens when this approach is applied to uncertain locations? As noted by Jacquez and Jacquez (1999), location uncertainty has several sources. Errors in georeferencing may be represented as $x_u = x_P + m_x$. Here m_x represents measurement error in the x ordinate. In GIS a common source of location uncertainty is the use of centroids locations. Centroid locations arbitrarily assign values associated with an

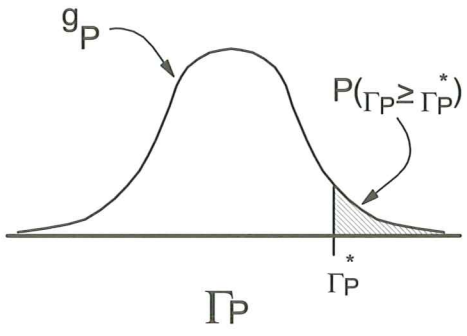

Figure 2. Statistical inference for precise locations

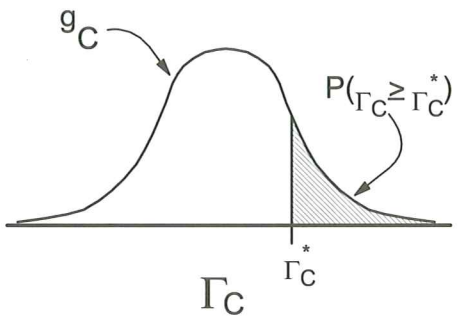

Figure 3. Statistical inference for centroid locations using the classical model

area to that area centroid. An example in epidemiology is the use of census tract centroids instead of actual place of residence. A common approach when working with such data is to ignore the uncertainty terms and to analyze the data as though they were precise. In these instances the test statistic is denoted Γ_C, and is called the test statistic based on centroid locations. The reference distribution under the null hypothesis, g_C, is obtained by sprinkling the observations over the centroid locations in a manner consistent with the null hypothesis. A P-value for the test is the probability, under the null hypothesis, of obtaining a value of the statistic as large or larger than the observed, written $P(\Gamma_C \geq \Gamma_C^*)$ (Figure 3). Here Γ_C^* is the test statistic based on the observed data at the centroid locations, while Γ_C is the test statistic under the null hypothesis. Statistical inference is evaluated in a manner similar to that used for precise locations: When $P(\Gamma_C \geq \Gamma_C^*) \leq \alpha$ the null hypothesis is rejected and the alternative is accepted, otherwise the null hypothesis is accepted. Unfortunately this approach is often used with spatial data; location uncertainty is ignored, and statistical inference is conducted in the same manner as if the locations had been precise.

So what is wrong with this approach to the spatial statistical analysis of uncertain locations? First, location uncertainty is not represented in the statistical results; it should be propagated through the proximity metric and represented in the test statistic. The inference mechanism illustrated in Figure 3 doesn't do this.

Second, P-values for uncertain locations can differ markedly from those based on precise locations (Jacquez and Waller, 2000). In general, uncertain locations tend to be *hyperdispersed*, so they are more uniform than expected under a random spatial point process. This is attributable to the resolution of the measuring instrument, which results in a 'graininess' beyond which locations cannot be resolved. Hyperdispersed spatial point distributions are not consistent with statistics that assume a random (Poisson) spatial null model. In a simulation study, Jacquez and Waller (2000) demonstrated that P-values calculated from centroids can differ markedly from those calculated from precise locations.

Third, as demonstrated earlier, sample-based randomization tests overly restrict the spatial sampling space. This means inference pertains only to sample, and, because the reference distribution is based only on the sample, statistical power may be reduced.

IV. CREDIBILITY AND STATISTICAL INFERENCE

Figures 2, 3 and their discussion illustrate the weakness of spatial randomization tests as they are commonly implemented. What is needed is an inference mechanism that:

1. Accounts for location uncertainty.
2. Specifies the spatial sampling space to correspond to the underlying population.
3. Incorporates spatial autocorrelation under the null hypothesis.
4. Makes a statistical inference about the underlying population, rather than just the sample.

Credibility-based statistics accomplish this using location models, spatial Monte Carlo methods, and spatially restricted randomization to maintain spatial autocorrelation under the null hypothesis. For now, consider the advantages conveyed (Figure 4).

Location models used in conjunction with spatial Monte Carlo techniques builds up the reference distribution, g_L, by repeatedly sampling from the underlying population, whose geographic distribution is specified using an appropriate location model. Unlike the examples in Figures 2 and 3, this reference distribution is thus population-based, rather than sample-based. Location models are also used to model uncertainty in the observed locations, resulting in a distribution of the test statistic g_T. This distribution quantifies how location uncertainty impacts the test statistic. It represents possible values of the test statistic, premised on a model of location uncertainty (See Jacquez and Jacquez, 1999, for a detailed presentation of location models). Inference is conducted based on *credibility*, which is the proportion of g_T greater than or equal to the α critical value of the reference distribution $C = P(\Gamma_T \geq \Gamma_\alpha)$. When the distribution of the test statistic is far to the right of the distribution under the null hypothesis, credibility is large, and the null hypothesis is rejected. When these distributions are similar, credibility is small and the null hypothesis is accepted. One can

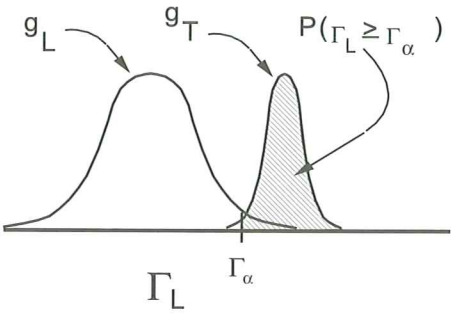

Figure 4. Statistical inference using credibility

define a *critical credibility*, C', to use as a trigger point in decision making. When $C \geq C'$ the null hypothesis is rejected, otherwise it is accepted. For now, consider the advantages of the credibility approach. Credibility

(a) Uses location models to account for location uncertainty, and propagates this uncertainty into the distribution of the test statistic.
(b) Specifies the spatial sampling space to correspond to the underlying population through spatial Monte Carlo techniques and location models.
(c) Incorporates spatial autocorrelation under the null hypothesis by spatially restricting the randomization procedure.
(d) Makes a statistical inference about the underlying population, rather than just the sample.

V. EXAMPLES

Consider an example to illustrate some of these concepts. All calculations were accomplished using the Gamma software (http://www.biomedware.com). The data are real, but are georeferenced to hypothetical management units in order to illustrate statistical inference using credibility. The locations and years of occurrence of 299 fires in a Northern Quebec forest were recorded from 1920 to 1983[1]. It is hypothesized that insect infestations impose a space-time pattern in fire occurrence caused by a cycle of forest growth, infestation, accumulation of dead wood, and combustion. If true, this hypothesis would result in positive spatial autocorrelation in times of fire occurrence, such that nearby fires tend to occur in the same or temporally adjacent years. The locations of the fires were recorded within 20 management districts, whose centroids are shown in Figure 5 (Top Left). The management districts partition the forest

[1] Data kindly provided by M. J. Fortin.

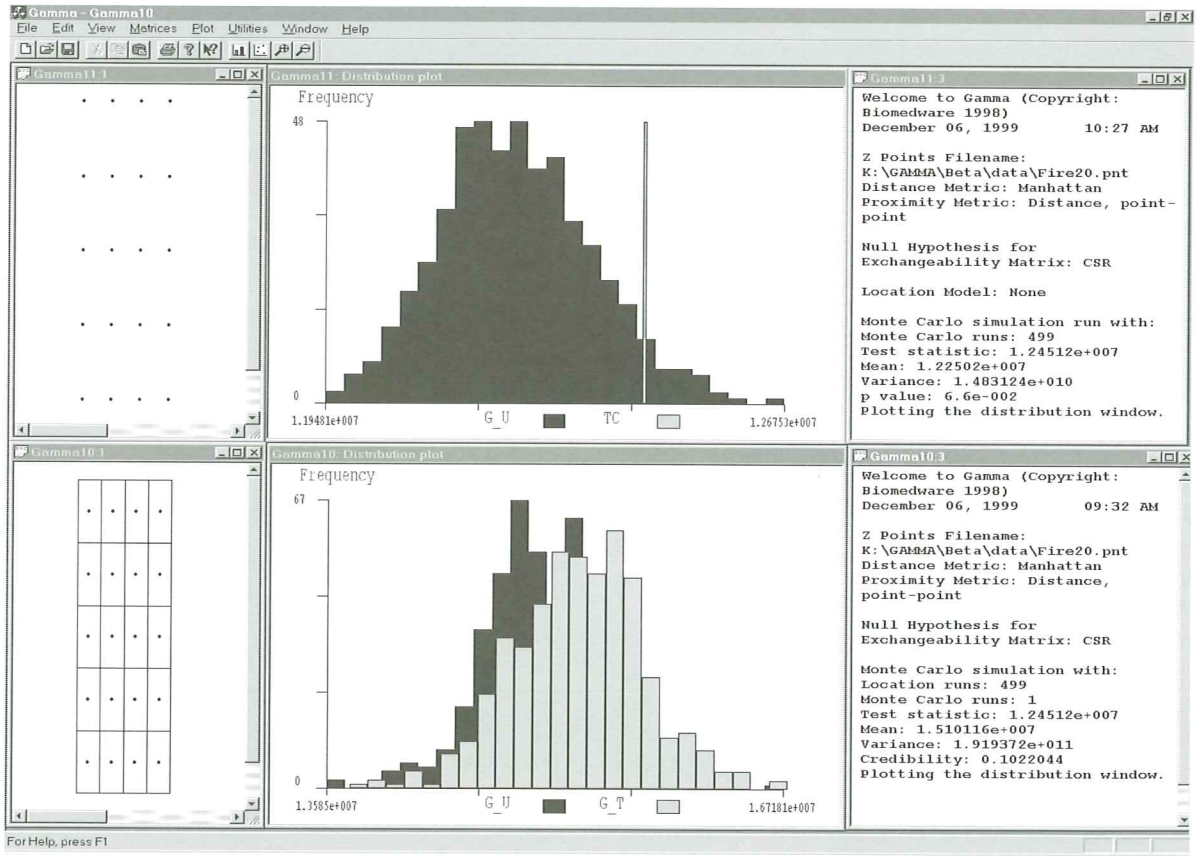

Figure 5. Analysis of fire data georeferenced to 20 region centroids using traditional and credibility approaches.

into 20 areas of equal size (Figure 5 bottom left). Mantel's test was used to determine whether there is association between fire locations and times of occurrence. Hence the geographic proximity metric (δ_{ij} in equation 3) was distance between pairs of fires, while the data metric (d_{ij} in equation 3) was the waiting time between pairs of fires. As noted earlier, fire location is georeferenced to the centroids of the 20 management districts, and a statistical analysis does not find a significant spatial association (Figure 5, top). The test statistic is well within the reference distribution, as shown by the upper histogram in Figure 5. This analysis ignores location uncertainty introduced by georeferencing to centroids, and a second analysis is conducted that represents location uncertainty and propagates it through the proximity metric to generate a distribution of the test statistic, shown by the histogram in light gray (Figure 5, bottom). The distribution under location uncertainty is shifted to the right of its reference distribution (shown in black). Credibility is 0.1022, meaning there is a greater than 10% chance of a statistically significant association, once our lack of knowledge of actual fire location is taken into account. Perhaps of greater import, the analysis clearly demonstrates that the impact of location uncertainty is large – the distribution of the test statistic accounting for uncertainty is as broad as its distribution under the null hypothesis! One concludes the amount of uncertainty introduced by georeferencing to management district centroids has a severe impact on spatial analyses.

Of course the 20 management districts are a construct I used to illustrate how location uncertainty may be incorporated into spatial statistics. In reality the centers of the fires were known with high resolution, and the analysis of these "exact" data is shown in Figure 6. The map of the locations of the fires is on the left of Figure 6, and the value of the test statistic (vertical line on the histogram) is superimposed on the reference distribution (black histogram). The P-value is 0.012, and the observed space-time association between fire locations and times is unlikely to be due to chance alone. There indeed is positive spatial autocorrelation in the times when fires occur.

This example illustrates how credibility and location models may be used to determine the impact of location uncertainty on spatial analyses.

VI. CONCLUSION

Credibility is a flexible new approach to spatial randomization tests, but is not a panacea. It applies

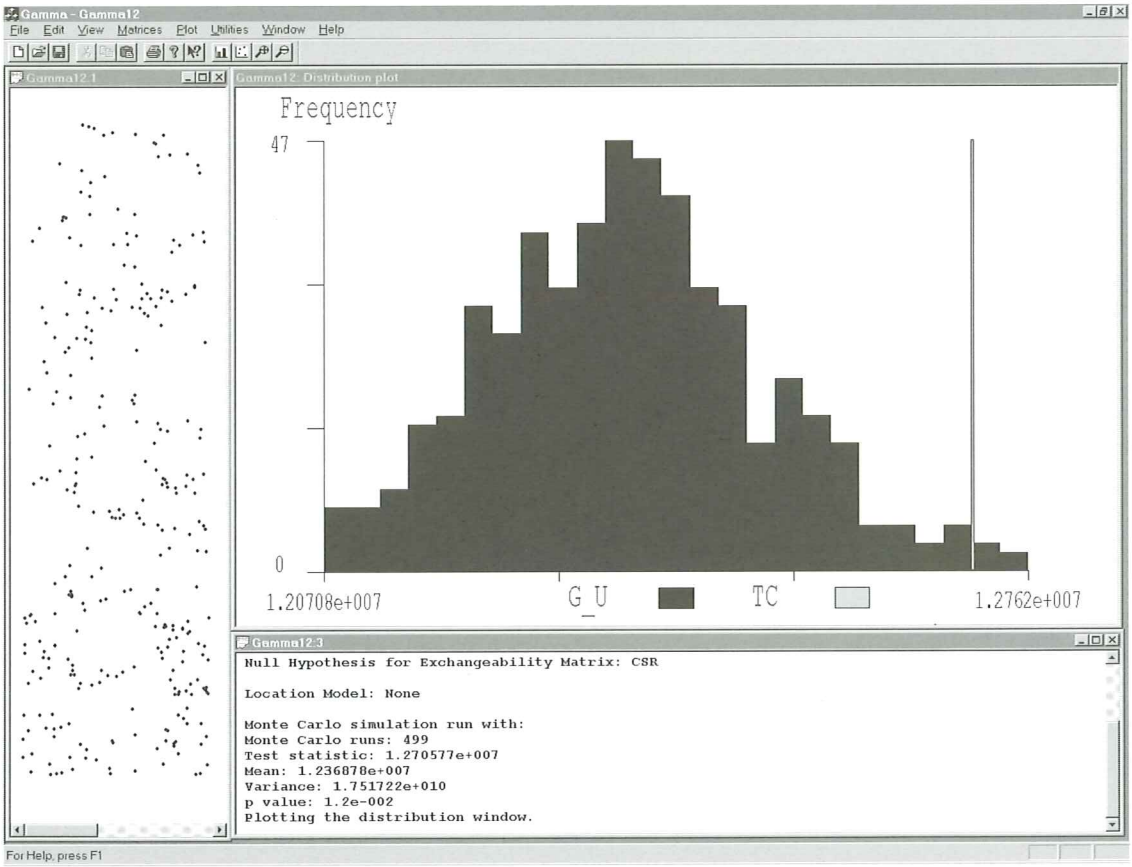

Figure 6. Analysis of fire data georeferenced to locations of fires.

only to spatially referenced data, and to spatial statistics that incorporate measures of geographic proximity or distance. It uses Monte Carlo sampling to generate the null distribution, and not distribution theory, as classical statistics do. It is a technique for testing hypotheses regarding spatial pattern, and is best described as a method for Exploratory Spatial Data Analysis. It is meant to complement, not replace, traditional spatial statistics that use P-values and alpha levels. In conjunction with these techniques it forms a quantitative basis for evaluating the likely impact of location uncertainty on one's ability to make statistical decisions with spatial data. Finally, it is appropriate for spatial data that are observational, autocorrelated, and uncertain.

How do P-values relate to credibility? Suppose we have a measuring instrument for determining geographic coordinates. Location uncertainty is present in sample locations because of measurement error in the instrument. We use an appropriate location model for modeling the measurement error, and propagate this error through the proximity metric into the distribution, g_T, of the test statistic. Now suppose we improve the measuring instrument, reducing the variance in g_T. As the measuring instrument becomes increasingly precise, variance in g_T decreases. When the instrument is perfect, locations are known precisely and g_T condenses to a point mass at the value of the test statistic, Γ^*. Credibility then is 1 when Γ^* is greater than or equal to that value of the reference distribution corresponding to α, the Type I error. For the imaginary situation where locations are measured without error, C=1 when the P-value is less than or equal to α, otherwise C=0. The mechanism for statistical inference using P-values is the special case of credibility for precise locations. This is clearly shown in Figure 7, which shows an analysis of the fire data using 1500 artificial management units. Location uncertainty is substantially less than in Figure 5 (1500 polygons vs. 20 polygons) and the distribution of the test statistic is much narrower and entirely outside the reference distribution. The corresponding credibility is 1.0.

How does credibility relate to statistical power? Recall, under the classical statistical paradigm, power is the probability of correctly rejecting the null hypothesis. This assumes precise locations and a distribution under the alternative hypothesis that is constructed by repeating the appropriate experiment. In contrast, credibility is the proportion of possible realizations of the uncertain locations that are statistically unusual.

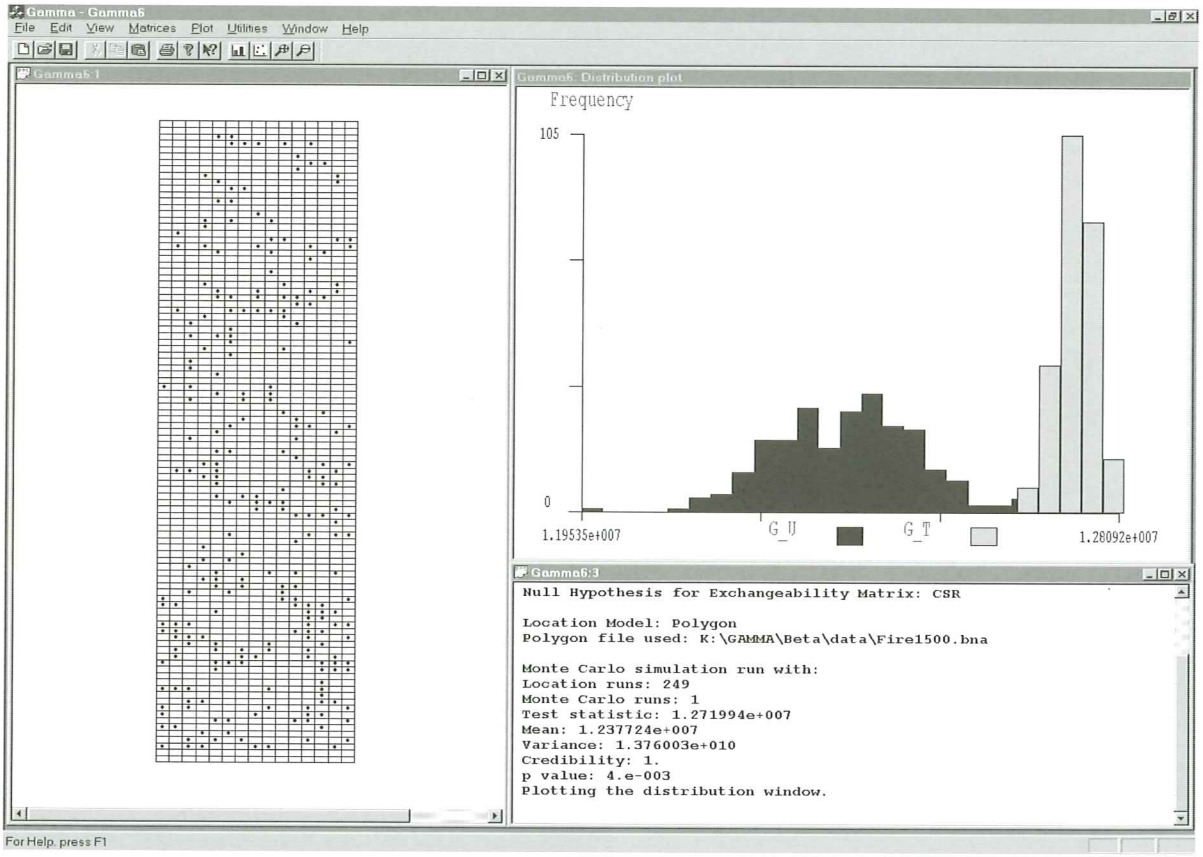

Figure 7. Analysis of fire data georeferenced to 1,500 region centroids.

ACKNOWLEDGMENTS

This research was funded by Small Business Innovation Research grant R43 CA65366 from the National Cancer Institute. Its contents are solely the responsibility of the author and do not necessarily represent the official views of the NCI.

REFERENCES

[1] Altman, D., 1994. Fuzzy set theoretic approaches for handling imprecision in spatial analysis, *International Journal of Geographical Information Systems,* 8:270-289.

[2] Anselin, L., 1995. Local indicators of spatial association – LISA, *Geographical Analysis,*27(2):93-115.

[3] Cliff, A. D. and J. K. Ord .1981. Spatial Processes : Model and Applications. London, Pion.

[4] Cressie, N., 1991. Statistics for Spatial Data, New York, John Wiley and Sons.

[5] Cuzick, J. and R. Edwards, 1990. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society*, 52:73-104.

[6] Ehlschlaeger, C. and A. Shortridge. 1996. Modeling elevation uncertainty in geographical analysis, Spatial Data Handling '96, Delft, The Netherlands.

[7] Ehlschlaeger, C.R., A. M. Shortridge, and M. F. Goodchild. 1997. Visualizing spatial data uncertainty using animation, *Computers & Geosciences*, 23(4):387-395.

[8] Faust, K. and A. K. Romney. 1985. The effect of skewed distributions on matrix permutation tests. *British Journal of Mathematical and Statistical Psychology,* 38:152-160.

[9] Getis, A., 1992. Spatial interaction and spatial autocorrelation: A cross-product approach. *Environment and Planning A* , 23:1269-1277.

[10] Getis, A. and J. K. Ord. 1996. Local spatial autocorrelation statistics: An overview. Spatial analysis: Modelling in a GIS environment, P. Longley and M. Batty. Cambridge, Geoinformation International.

[11] Goodchild, M. and S. Gopal. 1989. Accuracy of Spatial Data Bases. New York, Taylor and Francis.

[12] Goodchild, M. F., G. Sun, and S. Yang. 1992. Development and test of an error model for categorical data. *International Journal of Geographical Information Systems* , 6(2): 87-104.

[13] Haining, R., 1990. Spatial Data Analysis in the Social and Environmental Sciences. Cambridge, Cambridge University Press.

[14] Haining, R. P. and G. Arbia .1993. Error propagation through map operations, *Technometrics*, 35:293-305.

[15] Heuvelink, G.B.M., P.A. Burrough, and A. Stein.1989. Propagation of errors in spatial modeling with GIS, *International Journal of Geographical Information Systems*, 7:231-246.

[16] Jacquez, G.M., 1996. Disease cluster statistics for imprecise space-time locations, *Statistics in Medicine*, 15:873-885.

[17] Jacquez, G. M. and L. A. Waller, 2000. The effect of uncertain locations on disease cluster statistics. *Quantifying Spatial Uncertainty in Natural Resources.* H. T. Mowrer and R. G. Congalton. Ann Arbor Press, Chelsea, Michigan.

[18] Jacquez, G.M., 2000. Spatial Randomization and Statistical Inference. Ann Arbor, MI, BioMedware Press

[19] Jacquez, J.A., 1996. Compartmental analysis in biology and medicine. Ann Arbor, MI, BioMedware Press.

[20] Jacquez, G.M. and J.A. Jacquez. 1999. Disease clustering for uncertain locations, In Disease Mapping and Risk Assessment for Public Health. Eds. A. Lawson et al. John Wiley and Sons, Chichester.

[21] Keefer, B.J., J.L. Smith, and T.G. Gregoire.1991. Modeling and evaluating the effects of stream mode digitizing errors on map variables, *Photogrammetric Engineering and Remote Sensing* , 57:957-963.

[22] Kiiveri, H.T., 1997. Assessing, representing and transmitting positional uncertainty in maps, *International Journal of Geographical Information Science,*11(1):33-52.

[23] Manly, B.F.J., 1991. Randomization and Monte Carlo Methods in Biology. London, Chapman and Hall.

[24] Mantel, N., 1967.The detection of disease clustering and a generalized regression approach, *Cancer Research*, 27:201-218.

[25] Marshall, R.J., 1991. A review of methods for the statistical analysis of spatial patterns of disease, *Journal of the Royal Statistical Society, Series A,* 154:421-441.

[26] Mielke, P.W., 1978. Clarification and appropriate inferences for Mantel and Valand's nonparametric multivariate analysis technique, *Biometrics*, 34:277-282.

[27] Moran, P.A., 1950. Notes on continuous stochastic phenomena, *Biometrica*, 37:17-23.

[28] Rushton, G. and P. Lolonis. 1996. Exploratory spatial analysis of birth defects in an urban population. *Statistics in Medicine*, 15: 717-726.

[29] Smouse, P.E., J.C. Long, and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test for matrix correspondence. *Systematic Zoology* , 35:627-632.

[30] Tobler, W., U. Deichmann, J. Gottsegen and K. Maloy. 1995. The global demography project. Santa Barbara, National Center for Geographic Information and Analysis.

[31] Viertl, R., 1996. Statistical methods for non-precise data. New York, CRC Press.

[32] Waller, L.A. and G.M. Jacquez. 1995. Disease models implicit in statistical tests of disease clustering, *Epidemiology* , 6(6):584-590.

[33] Wartenberg, D. and M. Greenberg. 1990. Space-time models for the detection of clusters of disease, *Spatial Epidemiology*. R.W. Thomas. London, Pion. 21.