Measuring the spatial structure of biodiversity

by

Brian McGill

School of Natural Resources
Bioscience East Rm 325
University of Arizona
Tucson, AZ 85721
(520) 621-9389 mail@brianmcgill.org

Chapter 9 for "Frontiers in Measuring Biodiversity"

9.1. Introduction

The world is increasingly becoming a mapped world. Google and Yahoo are now highly spatially aware and are quick to represent searches on a map when possible. The study of biodiversity is becoming increasingly a mapped discipline as well. The 19th century approach to biodiversity across space was a "checklist" – a list of species found within a geographic entity like a park, county/shire, state/province or country. But it is clear that the 21st century approach to biodiversity is to map it. This is in no small part due to the advancement of technology. The ready availability of GIS programs, Google Earth, etc have made us less willing to live without maps. But this propensity to map also arises from the fact that our sampling of biodiversity has intensified to the point where we now have many measures repeated across space which is a prerequisite for mapping biodiversity.

Drawing a map consists of taking data, locating it in space and then plotting the data with an accurate (scale) representation of the spatial location of data points relative to the other data points. Usually other "layers" of information are added such as political boundaries, rivers or elevation. Such direct acts of drawing maps are best left to a GIS course or textbook (and are fairly simple these days with the software tools available) and are therefore not covered here. Here I explore the more detailed question of how we can measure and use the spatial structure of biodiversity to ask rigorous questions. Questions about the fundamental processes that drive biodiversity. And questions about the implications of the spatial structure of biodiversity for management and conservation.

9.1.1. Questions

Tobler (1970) coined the first law of geography: "everything is related to everything else, but near things are more related than distant things". The contrapositive of the second half of Tobler's law is "far things are less related". This implies that a law of geography is that everything varies across space. One would be hard press to dispute this from the study of biodiversity. The distribution of diversity across space varies at all spatial scales (See Chapter ??? – Jetz Global). The abundance of an individual species varies across space (Brown *et al.* 1995); even whether a species is present or absent varies across its range (See Chapter ??? – Occupancy). Of course this should come as no surprise because the underlying environmental variables such as elevation, temperature, and soil properties vary across space.

Modern spatial statistics has converged around the model:

(1) $z=\mu(s)+\varepsilon(s)+\eta$

where z is the variable of interest (e.g. biodiversity, abundance of a target species), $\mu(s)$ is the average value of z in the area of spatial location s, η is noise or measurement error which does not depend on spatial position and is independent (uncorrelated) across space, and $\varepsilon(s)$ is variability about the mean that is spatially structured (i.e. correlated across space). It is common to refer to μ as a first-order effect because it looks only at one point in space and ε as a second order effect because it by definition depends on two points in space at once (or conceivably 3 or more points at once but such higher effects are usually ignored). I extend this semantics to refer to η as a zeroth order effect since it does not depend on space at all. Figure 1 explores what the variation in z across space likes with different combinations of zeroth, first, and second order effects. Conveniently, and not coincidentally, this mathematical organization matches nicely with the two most common questions one asks about the spatial structure of biodiversity.

Question 1 – first order effects – is there variation in abundance across space

The most basic and most useful question is whether there is systematic change in the variable of interest across space. For example, are there regions where diversity or the abundance of a target species is particularly high or particularly low? From a basic research point of view this is interesting because it raises the question of why? Questions of what controls diversity and abundance are central to ecology and exploring how this varies across space is one of the easiest ways to study these questions. From a management point of view, identify specific areas of high abundance or diversity can lead immediately to management decisions such as the locations of reserves in biodiversity hotspots (Prendergast *et al.* 1993; Reid 1998) or minimizing human impact from roads and buildings in regions of high abundance for an endangered species. Statistically, these questions all map back to questions about first-order processes – changes in the intensity of a process (the average of the variable per unit area) across space.

Question 2 – second order effects - are there interactions across space

Second order effects explore the interactions between the organisms under study. Given that the target species is found at some location, will the species be more likely or less likely to be found 20 meters away? 100 meters away? 10 km away? or is there no effect? This is generally demonstrated in patterns that are clumped, random or regular (overdispersed) (see Figure 3). In the early days of the study of this question, it was generally assumed that individuals of a species would have a deterrent effect on other members of the same species nearby due to competition. A few such examples have been found. For example adult (but not juvenile) creosote bushes appear to be overdispersed, presumably due to either competition for water or possibly allelopathy (Phillips & MacMahon 1981). But the vast majority of studied organisms appear clumped(He & LaFrankie 1997; Condit *et al.* 2000). Increasingly sophisticated analysis techniques allow one to explore how these patterns change with scale (e.g. species might be clumped at small scales due to limited seed dispersal but random at larger scales).

9.1.2. Types of data

Data on biodiversity across space can take many forms (Figure 1, Table 1). This is one of the sources of the proliferation of techniques and resulting confusion. But there are just a few key variations. The single most important question is the nature of the spatial structuring of the data and is described by the dichotomous key below:

a) the locations recorded are controlled by natural events → **individual point process data** Examples include the locations of tree trunks in forest, bird nests, and lightning strikes.

b) the locations are chosen by the scientist in the design of the sampling protocol

   i. the locations are spaced regularly across space → **quadrat data** Examples include quadrats arranged across a line (transects) and quadrats arranged across a rectangular area. In addition to the regularity of the data there is also an assumption that some area (line or rectangle) has been exhaustively covered.

   ii. the locations spaced irregularly, perhaps apparently randomly across space → **geospatial data** Examples include counts of abundance of a target species in quadrats where the quadrats are scattered across a large area or true point measurements such as soil moisture at irregular spacing across an area. The location may be literally random (a randomization was used in the protocol) or may be just irregular with the actual locations driven in part by accessibility to roads and trails, etc.

Note that there is a gradient as one proceeds from the top (a – individual point proceses) to the bottom (b-ii – geospatial data). As one moves down, the amount of information decreases, but

the ease of obtaining the information increases. Type a data can be converted to type b-i data which can be converted to type b-ii data but not in the other direction.

A second key question about the data is what the statistical nature of the measurement is at each location – is it binary (e.g. presence/absence) or count data (e.g. abundance of a species) or continuous (e.g. soil pH). Most commonly, point processes record only binary information (tree trunk here) while the scientist controlled data typically takes a numerical measurement. Some point processes record information such as the diameter of each tree or the species of each tree and these are known as marked point processes. In general, analysis methods exist for any combination of location structure and measurement variable type. I summarize this classification of different types of spatial data given by three spatial structures and two types of variables measured in Table 1.

Although it will not affect the type of analysis done, many readers are aware that the spatial scale at which the data is measured can have profound implications for the results found and the types of questions that can be asked. The scale of the data is usually described by two terms. The grain is the scale of a single measurement. Thus if a single measurement is a 1m X 1m quadrat (as is common in grasslands), then the grain is 1 m (or 1 m^2 depending on if linear or areal measurements are used). The extent is basically the distance between the farthest apart measurements. This can range from meters to 1000s of kilometres. The number of sampling units, N is another key attribute of any survey. These three factors, if grain and extent are measured in areal units, can be combined to describe the coverage of study (coverage=N*grain/extent – giving the % of the study area actually measured). In point process and quadrat data (types a & b-i above) it is often assumed that coverage is 100% while the geospatial data survey method is often chosen when the amount of resources available for survey is inadequate to provide 100% coverage of the area of interest. The results of analyzing data can depend heavily on the scales involved (see Figure 2e described later for an example). There is no such thing as a best scale for grain or extent. The goal is that the scales should be chosen to match the question at hand. Thus if one is studying bird dispersal, then the grain should be small relative to the dispersal distance but the extent should be large relative to the dispersal distance. If one is studying tree mortality which is a rare event, then the total area covered (N*grain) needs to be large while the extent should match the question (is one making assertion about mortality across the geographic range of a species or in a specific park?).

9.1.3. – Number of variables recorded – pattern or association?

The final fundamental question is whether one is recording only one target variable of interest (such as species richness or abundance of a target species) or if one is recording multiple variables (such as other species or environmental variables) and attempting to find associations between the variables. With multiple variables, finding associations can inform about the causes of or lead to predictive models about a variable of interest. Thus one can build a spatially informed model of which environmental variables increase species richness. Or one can build a predictive model about which habitat factors are associated with high abundance for an endangered species. Such associational analyses have traditionally been done in a non-spatial context, treating measurements across space merely as replicates and using methods such as regression to find the patterns. There is however growing awareness that if the replicates are across space there are a number of special statistical challenges and opportunities. These are discussed in section 9.2.4 below.

In summary the combination of: 1) three aspects of spatial structuring (section 9.1.2), 2) two types of question (9.1.1); and 3) two options for number of variables recorded (9.1.3) combine in a 3x2x2 grid to determine what types of spatial analysis tools can and should be

performed (Table 2). If one has a spatial dataset but doesn't know what types of analyses to perform, one should start by determining where in this 3x2x2 grid the data and the questions of interest fall.

9.2. Methods

There has been a great proliferation of spatial statistical methods in the last few decades, and starting out in spatial statistics can be overwhelming. Books on spatial statistics, whether targeted at statisticians (Cressie 1992) or ecologists (Perry *et al.* 2002; Fortin & Dale 2005b), list dozens of different techniques. There have been a few attempts to find links between the methods (e.g. Dale *et al.* 2002), but there has been distressingly little attempt to stress the underlying unity of methods to ease the learning curve for beginners. In this chapter I will adopt the attitude that details of calculations can be left to software and emphasize the interpretation of results and the conceptual unity of different methods.

9.2.1. Null models

The simplest null model is that the variable of interest is constant across space (Figure 1A), but this is not particularly useful. To be useful, a null model must make some specification of the variability expected under a null model. The simplest such null model goes under varying names including "complete spatial randomness (CSR)" and "homogenous Poisson process". In this null model, there are no second-order effects – every location is spatially independent of every other location. And the intensity (expected mean value of the target variable) is constant across space (i.e. homogeneous or $\mu(s)=\mu$). This can be modelled simply by treating each point as a sample from the Poisson distribution with the parameter λ held constant (Figure 1B). Conveniently, this can also be modelled as a point process (where we model the location of individuals) as a Poisson point process (Figure 2A). Each point is placed in space randomly (independently of other points) with an intensity λ. (Computer simulation of a Poisson-point process is trivial – if we rescale spatial coordinates to range from 0-1 on both axes and then sample two random uniform numbers over the interval 0-1 (something which nearly every programming language provides) and treat these as the x-y coordinates of a point we will be creating a Poisson-point process where λ=N/Area (N=total # of points simulated)). If we draw a grid on this process and count # of individuals per grid-cell we recapture the first null model of a Poisson distributed count per cell with independence between cells. So at the null model level the two fundamental types of data (nature controlled point processes and human controlled measurements of intensity) converge into a single null model derived from the Poisson distribution. In a Poisson distribution with intensity (average rate λ), the probability of observing n points in an area A is given by $P(k=n|\lambda A)=(\lambda A)^k e^{-\lambda A}/k!$. A particularly useful special case of this is the probability that at least one point is observed $P(k>0|\lambda A)= 1-P(0|\lambda A)=1-e^{-\lambda A}$. Often one omits A for convenience and it is assumed k is a density (i.e. per unit area). More complicated null models become appropriate when several variables are measured and these will be discussed in section 9.2.4.

9.2.2. Estimating intensity (first order effects)

One central goal of spatial analysis (question #1) is simply to create a map of intensity across space. This then allows exploration of what factors cause high vs. low intensity and management decisions based on areas that are favourable (e.g. high intensity of the target organism). Two issues arise in creating such maps. First, if coverage is less than 100% (i.e. the area sampled is a fraction of the area of interest, typically the geospatial data case), then

interpolation is needed to make a prediction about the areas that are not sampled. Second, whether the coverage is 100% or not, there is a goal to remove the zeroth order noise and find the underlying "true" signal. One would hate to locate a critical management resource (e.g. extra food) at a site to find out it appeared to have a high density due to chance on the day the survey was done with no unusually high abundance ever observed since. This process of removing noise is called **smoothing**. Usually one technique accomplishes both smoothing and, if needed, interpolation. We discuss four broad classes of techniques that can be used.

- **Local surfaces (smoothing)** – Many techniques can be lumped together as smoothing. For any point where a prediction is needed, the predicted value is a weighted average of nearby observed points. If the spot is where an observation occurs, then only smoothing is involved, but if the spot is where no observation occurred then interpolation is also occurring. Typically nearer points are given a greater weight. This can be captured in the equation:

$$z_p = \frac{\sum_{i \in S} w_i z_i}{\sum_{i \in S} w_i}$$

The predicted value z at point p is a weighted average of the observed values z_i over some subset, S, of all the observed values. The simplest case is nearest neighbour interpolation where $w_i=1$ if the point i is the closest point to p and $w_i=0$ otherwise. Moving average is another technique on gridded data where $w_i=1$ for the cells within h cells of the target cell p. A similar approach used on geospatial data is the moving window where a box with sides of length h are drawn around each point and $w_i=1$ if the point i is in the box and zero otherwise. Similar techniques can be used where points are differentially down-weighted the further away they are. Exponential smoothing, where $w_i=\exp(-hd_{ip})$ where d_{ip} is the distance between i and p is a common choice. The most modern and probably best technique is known as kernel smoothing which weights the points according to a symmetric probability distribution centered around the point p and where $w_i=k[(z_p-z_i)/h]$. A 2-dimensional Guassian bell curve is a common choice where $w_i=\exp(-d_{ip}^2/h^2)$ as is the quartic kernel $w_i=3(1-d^2/h^2)^2/\pi$ when $d<h$ and 0 if $d>h$. In the quartic case any point further than bandwidth h away has no effect whereas in the Guassian it has an increasingly small effect. Notice that in all of these scenarios a specification of a smoothing parameter h is required where h is a measure of the distance at which effects are still important. As h increases more smoothing occurs. In the limit when h is very large a complete flat surface (figure 1A) will be produced.

- **Global surfaces (trend surfaces) –** In contrast to smoothing approaches which are local in nature, a trend surface is a global prediction that can be summarized by relatively few parameters. A trend surface is described by a function:
    - $z_p = f(x(z_i), y(z_i)|\beta)$

where $x(z_i)$ and $y(z_i)$ give the x and y coordinates of the point and β is a set of parameters. The function f can vary from simple to quite complicated. The simplest trend surface to fit is a linear model where f is a plane (i.e. $z_p=a*x+b*y+c$). Such fitting is usually done by a least-squares criterion and is basically just a two-variable regression with the independent variables being spatial coordinates. The next step up is to use a polynomial. A quadratic polynomial would be $z_p=ax+bx^2+cy+dy^2+exy+f$. A quadratic polynomial yields a 3-D parabola and as such can only represent a unimodal (one-peaked) surface. More commonly cubic or quartic polynomials are used. The number of coefficients that need to be estimated goes up quickly with degree (linear=3, quadratic=6, cubic=10 and generally a polynomial

of degree p has (p+1)(p+2)/2 coefficients). Care must be used. Fitting a quartic (p=4) polynomial with only 20 data points would leave only 5 (20-15) degrees of freedom which is probably too low to be reliable. A variety of functions with f more complex than polynomials can also be used. For example if one is modelling a surface that varies periodically sine or cosine functions may be appropriate. The most general form is to use splines for the function f. This then begins to come around and be not too different in practice from the local surface/smoothing approach.

- **Kriging –** The first two techniques (local and global surfaces) deal only with first order effects. However, second order effects can also impact the predicted values z_p. Over very large scales the second order effects should decay to zero and a surface based on first order effects will be unbiased (i.e. be accurate on average). However, on smaller scales second order effects can significantly affect the predicted values. Knowing that the variable of interest (e.g. abundance) is unusually high at a point p implies that it will usually be unusually high in a neighbourhood around p. First order processes take this into account to some degree (small bandwidths h and highly flexible global surfaces f, such as splines, are most influenced by local conditions). However, a second order effect is probably a more accurate way to incorporate this information and allows us to make inferences about the nature and distances of interactions. The means to do this is called kriging Kriging is the attempt to produce more accurate predictions of intensity by combining first-order and second-order analyses of the data. As such, I postpone further discussion of kriging to section 9.2.3 where second order effects are covered in detail.
- "**Let it all hang out**" – One approach is to not smooth or interpolate but simply to report the data as it exists. Press and colleagues argue for just this approach(2007). The human mind is naturally good at interpolation. It is only moderately good at smoothing, but certainly it can detect and dismiss outliers.

In comparing these techniques, three things are worth noting. First, the global trend surface approach is effectively making a prediction for all points in the area of interest. The smoothing and kriging techniques only make predictions for a specified set of points $p \in P$. This set can either be a set of target points of interest (e.g. locations under consideration for management decisions) or it can just be a regularly spaced grid of fairly fine resolution placed across the area of interest. Once a trend surface or a grid of predicted points is in hand, the data can be plotted a number of ways. Three-dimensional surface plots are most commonly used for trend surfaces. Contour maps are probably the most common way to plot smoothed data. In either case color (or grayscale) intensity can be overlayed.

Secondly, of the four techniques for estimating intensity surfaces, the fourth is the only one directly applicable to point process data. To estimate intensity surfaces across space from point process data, two choices are available. The first is to lay a grid over the data, and then count the number of points in each grid cell, thereby moving from type i data to type ii-a. Then any of the four techniques can be used. This entails a loss of information and works best when the points are fairly dense in space. Secondly the local surface method can be modified where instead of using z, the observed intensity at a point, one uses counts of points (replace z_i with n_i) and adjust the weights by area so $\lambda_p = \sum_{i=1}^{n} \frac{1}{h^2} k(\frac{p-p_i}{h}) / \int_A \frac{1}{h^2} k(\frac{p-u}{h}) du$ with the intensity at a point p basically being a density weighted average of neighboring points. Some of the earliest attempts to estimate density of trees was based on measuring nearest neighbour distances (Konig 1835), but they are accurate only if complete spatial randomness with no second order effects holds and not generally recommended today (see section 9.2.3 for a brief description of one of these methods).

Finally, the first two techniques require the input of a smoothing parameter (degree of smoothing). The smoothing parameter is implicit as the bandwidth, h, in the local smoothing case and implicit in the flexibility of the function f (e.g. degree of polynomial) in the global surface case. If too much smoothing occurs then true variation will be eliminated, but if not enough smoothing occurs then noise will be represented as pattern. Unfortunately, there is no way to determine the mathematically correct smoothing parameter, despite the existence of various guidelines and rules of thumb. The smoothing parameter is ultimately a subjective, human-chosen parameter. The third approach (kriging) essentially derives the degree of smoothing from the data itself and must be considered to have an advantage for this reason. The final approach also avoids a smoothing parameter. Overall, kriging has often been avoided due to its perceivedcomplexity, but with software readily available this argument has lost weight. Kriging also doesn't work well on discrete space (gridded data), although as the number of grid cells becomes large, the difference between discrete and continuous space begins to disappear. On the whole I would recommend the kriging unless the data is on a fairly small grid. And I would recommend always plotting the "let it all hang out" approach for comparison.

9.2.3. Studying effects at a distance (second order effects)

Early approaches

There are a couple of reasons one might want to study second order effects. Most simply, one might wish to make more accurate predictions of intensity. A more involved reason would be to test whether there are effects over distance and whether these effects are attractive or repulsive in nature. An even more informative question is to look at how these effects vary with distance, possibly allowing one to identify critical distances beyond which effects are especially strong or weak. From a basic research point of view, this then immediately leads into research questions about what processes cause these effects (dispersal limitation, species interactions, resource competition). From a management point of view these might give information about whether to colocate two species or what spatial scales are optimal for a reserve.

The oldest methods for exploring second order effects is the variance-to-mean-ratio used on counts taken in quadrats. This test derives from the fact that the variance of the Poisson distribution is equal to the mean of the distribution, λ. This is an unusual and strong property (compare to the normal distribution where the mean and variance are completely unrelated). Thus we can examine the index of dispersion ID=var(z_i)/average(z_i)=$[\Sigma_{i=1}^{n}(z_i - \overline{z_i})^2/(n-1)]/\ [\Sigma_{i=1}^{n} z_i/n]$. Under CSR/Poisson, ID=1. When ID>1, then there are more quadrats with either very low (near zero) or very high (well above average) abundance which is a sign of a clumped distribution (positive or attractive second order interactions). Conversely, when ID<1 then most quadrats have z_i very close to $\overline{z_i}$ which is a sign of regularity or overdispersion. Conveniently, if n>6 and $\overline{z_i}$>1 then (n-1)*ID is distributed approximately as a chi-square distribution with (n-1) degrees of freedom which allows for tests of whether ID is significantly larger or smaller than 1. There has been a great deal of criticism levelled at the ID (Hurlbert 1990) and it is true that ID=1 does not imply a Poisson process, but ID > 1 or ID<1 does imply clumping or dispersion. A more telling criticism is that this measure has no spatial structure (two quadrats 1km apart are compared in the same way that two quadrats 10m apart are), but this is such a quick and dirty measure it belongs in the toolbox of spatial analysis.

The first attempts to incorporate spatially explicit measures of second-order interactions (where quadrats far apart are treated differently than those close together) apply only to point-process data and are based on distance to the nearest point. Diggle's function G(d) is defined as the fraction of points p_i where the nearest neighbour of p_i is at a distance less than or equal to d.

Under the null hypothesis of CSR, $G(d) = 1 - e^{-\lambda\pi d^2}$ (see the formulas in section 9.2.1) and an empirically observed G(d) can be plotted against the null G(d). First order information (intensity) can also be derived from nearest neighbour information under (and only under!) the assumption of CSR, where a maximum likelihood estimate gives $\lambda = 4/(\pi\overline{d^2})$ (Diggle 1983). The related function F(d) gives the probability that the distance from a randomly chosen location in the area (which will usually not coincide with an observed point) to the nearest observed point is less than d. The use of these functions and related ideas based on nearest neighbour distances such as the Clark-Evans ratio, Hopkins ratio, and Blyth-Ripley ratio, are losing favour because they are poor estimates of second order effects under some scenarios. Imagine the scenario where every point in a random process is replaced by two points a constant, very small distance apart – the nearest neighbour distance will imply perfect regularity (nearest neighbour at a constant distance) even though the data is strongly clumped at a slightly larger scale and generally gives a poor picture of overall second order effects at greater distances.

As a result, other than the quick-and-dirty use of the ID, second order effects should be estimated using estimates that incorporate all spatial distances, not just nearest neighbours. The exact nature of these estimations depends on whether the data points are nature-controlled (point-process) or human-controlled (quadrat and geospatial).

"grams" and kriging

We cover the quadrat/geospatial approach first. There are a very of tightly related ways to describe second-order effects. The most direct is the idea of covariance $Cov(z_i, z_j) = \sigma_{ij} = \frac{1}{n}\sum_{k=1}^{n}(z_{i,k} - \overline{z})(z_{j,k} - \overline{z})$. Note that $Cov(z_i,z_i)=Var(z_i)=\sigma_i^2$. Such a definition assumes that we have multiple measures of z (k=1…n) at each spatial location which is rarely the case (although it could happen if for example we take repeated measures over time). This definition also implies the possibility of dependence on the specific locations i and j. A generalization is $C(d) = \sigma_{ij}(d) = \frac{1}{n_d}\sum_{\|i-j\|=d}(z_i - \overline{z})(z_j - \overline{z})$. Here we assume covariance depends only on the distance between two points ||i-j|| and calculate the covariance between all n_d pairs of points that are distance d apart. On gridded data and for discrete distances there are likely to be many such points (e.g. 5 cells apart). On geospatial data where d is continuous, it is unlikely to find two pairs of points that are exactly the same distance apart. In this case it is common to group data into bins and take all points that are between d and d+δd units apart. This binning is similar to what occurs in the creation of a histogram. A rule of thumb (Rossi *et al.* 1992) is that each bin should have 30-50 data points. Moreover d should typically only go up to half the maximum distance between points (when d is close to the maximum observed distance one is only comparing edges to edges which may not be representative).

A plot of C(d) vs. d is called a covariogram. According to the first law of geography, the covariance should decrease with increasing distance, eventually reaching zero and this is in fact what is observed. Typically $C(0)=\sigma^2$ (the variance in the data) and this decreases in an exponential- or hyperbolic-like fashion to $C(\infty)=0$. Certain circumstances can also cause negative covariances. For example if the data is a peak (e.g. Figure 1C) then distances that match the distance from peak to the valley around it will actually have a negative covariance. Several closely related plots can also be calculated (See Figure 4). A correlogram ρ(d) is simply the covariogram C(d) rescaled to vary between -1 and 1, just as a correlation coefficient does (the rescaling is based on the fact that $r=\sigma_{ij}/(\sigma_i\sigma_j)$. A variogram analyzes not the covariance of two points but the variance of the differences of two points as

$V(d) = E\left[\left(z_i - z_j\right)^2\right] = \frac{1}{n_d}\sum_{||i-j||=d} (z_i - z_j)^2$ with the same notation as for the covariogram. A factor of two appears because the pair i,j is counted twice in z_i-z_j and z_j-z_i. Thus most often the semivariogram defined as $\gamma(d)=V(d)/2$ is more convenient to analyze (for example $\gamma(\infty)=\sigma^2$ while $V(\infty)=2\sigma^2$) but is often confusingly just called the variogram. The semivariogram is related to the covariogram as $\gamma(d)=\sigma^2-C(d)$ (although in practice estimates of $\gamma(d)$ and $C(d)$ do not fit this relationship – for example $\gamma(0)$ is rarely 0 as this formula would require). In general, if the data violates the assumptions of stationarity (discussed below) the equivalences begin to breakdown. Thus all four "grams" are related linearly (I use "grams" to refer to the collect set of the four different types of plots: $C(d)$,$\rho(d)$, $V(d)$, $\gamma(d)$).

The main difference between the "grams" is that the covariogram and correlogram isolate interactions only at distance d, while the (semi)variogram describes an accumulation of effects up to distance d. In this sense the difference is conceptually similar (although not mathematically equivalent to) the difference in probability between a probability density function (PDF) and a cumulative density function (CDF). The strengths and weaknesses of the different approaches come from this fact. I find the correlogram the easiest to interpret. If the line is above 0 at distance d then there are positive (attractive) second-order effects at distance d and similarly for negative effects. This is less obvious from a (semi)variogram (a region d of the variogram is always still positive but where the slope is high one would have high correlation at d and low correlation at d if the slope of $V(d)$ is low). In other words, a high variance at distance d,$V(d)$, may be due to processes causing high variability near d=0 or to processes causing high variability near to d or at any distance in between. Conversely though, estimates of the variogram are generally considered more robust to outliers and hence more accurate. This is largely because the first order process for a variogram, $E(z_i-z_j)$ must equal zero, but the first-order process for covariogram $E(z_i)$ is not generally zero and must be estimated.

The use of "grams" has come from the world of geospatial data where points are random, continuous distances apart. However, they carry over with little modification to gridded data. Indeed, gridded data can make their application easier in practice. For continuous data one must find an appropriate bin size δ so that enough points occur in each bin. In gridded data, this is not a problem. One can use the natural distance of the grid where one is guaranteed to have many pairs of points that are one cell length apart, two cell lengths apart, etc. Early analyses of gridded data ignored most diagonal pairs (e.g. a cell 3 cells North and 2 cells East), but if one locates each grid cell at its center these distances can also be calculated and the bins can be centered around discrete cell lengths as in 0.5-1.5, 1.5-2.5, 2.5-3.5 etc cell lengths. Prior to the dominance of the "gram" format, the Moran's I and Geary's C statistics were used. These were typically applied to gridded data and used only on non-diagonal distances. Interestingly, aside from the grid-derived nature of the pairings, Moran's I corresponds exactly to the correlogram and Geary's C corresponds to the variogram divided by σ^2 (Figure 4). Moran's I and Geary's C are also often commonly applied to irregular polygon data such as when a variable is known for each county or shire. In these cases an adjacency matrix W is developed that gives the distance between each pair of polygons. Moran's I and Geary's C can then be generalized to use W instead of the classic discrete cell-distances. However, in the end all these methods directly relate to the correlogram and variogram. I treat them here as conceptually identical with the difference occurring only in the practicality of the calculations (which can be relegated to software). One minor difference that is important is that, unlike the correlogram, Moran's I can go outside of the interval (-1,1) and for no correlation (i.e. CSR) has $I=-1/(n-1)$ instead of $I=0$ where n is the number of observations. As n gets large the difference disappears.

Aside from visual inspection of a correlogram or semivariogram, several analyses can be done. Variograms are often described by three measures: $\gamma(0)$ is called the nugget (this is theoretically zero but empirically usually not), $\gamma(\infty)$, i.e. the asymptote or σ^2 is called the sill, and the distance d at which the sill is hit is called the range. A second type of analysis that can be done on "grams" is tests of statistical significance (most commonly whether a correlogram is statistically significantly different from zero). For Moran's I, the statistic is asymptotically normally distributed with a known mean and variance (Legendre & Legendre 1998) allowing for a simple analytical test of significance. Alternatively, a randomization test can be performed; keeping the same spatial locations randomly reshuffle the values observed. Calculate the correlogram that results. Repeat this for say 999 times and draw the 95th perecentile envelope of these randomizations. Parts of the correlogram outside the envelope are significant. Any significance test on a correlogram faces a problem of "repeated measures"; if there are 20 bins there are 20 tests of significance and the probability of making a Type II error (believing the null model is rejected when it should not be) is high. All the classic approaches to multiple tests apply. The Bonferroni test (use significance levels of 0.05/n instead of 0.05) is best known but overly conservative (Garcia 2004 and see solutions therein). In a final type of analysis, a model of the form $\gamma(d)=f(d|\theta)$ or $\rho(d)=g(d|\theta)$ where θ is a set of parameters can be fit to the empirical estimates. This can smooth out some of the noise in estimation. It can also allow comparison of the parameters between different "grams". The semivariogram is an increasing, decelerating, asymptoting function and is typically fit by the spherical model $f(d|r)=\sigma^2[3d/2r-d^3/(2r^3)]$ if $d \leq r$ or σ^2 if $d>r$; by the exponential $f(d|r)=\sigma^2(1-e^{-3d/r})$; or the Gaussian $f(d|r) = \sigma^2\left(1 - e^{-\frac{3d^2}{r^2}}\right)$. All three models pass through (0,0) and asymptote at (∞,σ^2) with one scale parameter r. Offsets can be added to allow for a sill (e.g. $f(d|r)=a+(\sigma^2-a)(1-e^{-3d/r})$ for the exponential. The correlogram is typically fit by simple exponential decay $g(d|r)=\exp(-d/r)$ and the covariogram by $G(d|r,\sigma^2)=\sigma^2\exp(-d/r)$.

Kriging, mentioned in section 9.2.3 as a means of making predictions of intensity that incorporate both first and second order effects, builds directly on the fact that the covariogram can be described by only the two parameters σ^2 and scale r. In the local smoothing methods of section 9.2.3 the question arose of how much weight to give to points at a distance d away and the only answer was to choose an arbitrary smoothing or bandwidth parameter h. The covariogram and a fitted model $G(d|r,\sigma^2)$ provides an obvious answer. Weight points that are a distance d away by the amount $G(d|r,\sigma^2)$! The empirical covariogram curve contains many degrees of freedom and could not be practically estimated, but if we fit a model and boil it down to two parameters, we only lose two degrees of freedom and now get an empirical estimate of how best to incorporate second-order effects.

We can now return to equation 1 ($z=\mu(s)+\varepsilon(s)+\eta$) where μ is some surface (e.g. linear) of spatial location and ε is a function only of the distance between two points and is described completely by the covariogram and its two parameters (η is assumed to be distributed normally with mean zero and variance τ^2 just as in traditional regression). Equation 1 essentially now fits the GLS (generalized least squares) and mixed model forms of regression. Although not well known, these are standard statistical techniques and methods for fitting them are readily available (typically involving some iterative solutions and likelihood methods). And in practice the semivariogram (which is more stable) is used rather than the covariogram. But conceptually, this is all there is to kriging. I leave the details to the software. Several flavours of kriging exist. Simple kriging assumes $\mu(s)=0$ (or that μ is known instead of estimated) and is not too common. Ordinary kriging assumes $\mu(s)=c$ with c to be estimated and requires that the user ensure that

there are no trends in the data (or that the trends be removed and residuals analyzed – see next paragraph). Universal kriging models the full equation 1 with first-order and second-order effects treated simultaneously.

The assumption made throughout this section on "grams" that second order effects (covariance) depends only on d is in fact an assumption that may be untrue. One common violation is when covariance depends on direction. This could occur if dispersal limitation is a major driver of second order effects and dispersal depends on prevailing winds or stream currents. When covariance is independent of direction we call it isotropic and we call it anisotropic if covariance is a function of distance. Another violation is when the covariance depends on location. This could occur in the previous scenario if the strength of prevailing winds or currents varies significantly across space. A full definition of stationarity (the usual assumption in spatial statistics) is that the mean is constant (expected value $E(z_i)=\mu$ independent of location) and that the covariance σ_{ij} depends only on the distance between i and j. The condition on the mean is probably more commonly violated than the condition on the covariance. For example a simple gradient or trend in z_i across space violates stationarity. It is common to fit a trend surface to the data and then to analyze the second-order structure (covariance) on the residuals, thereby removing the trend. On gridded data, differencing $z_i'=z_{i+1}-z_i$ will also remove a trend. With the exception of some analyses that deal with anisotropy, the covariance condition for stationarity is almost always assumed true. If one is worried, one can check this using local autocorrelation statistics. Local autocorrelation statistics (Anselin 1995) calculate a measure of correlation for each point just with nearby points and do not lump this together with the correlations of other points with their neighbours (as is done in a global correlogram). The result can be plotted as an intensity of local correlation. If this varies drastically across space, then the assumption of stationarity may not be justified. If one has a particular process in mind (such as dispersal limitation) underlying the covariance structure, then local autocorrelation statistics can be an indicator of the varying strength of that process across space.

In summary, correlograms and its relatives, provide a simple graph that enables one to determine at what spatial scales the variable of interest is interacting (the regions of d where the highest or lowest regions of the graph occur).They also allow for determination of whether the effects are positive or negative (the sign of the graph). Kriging which is a spatial form of smoothing makes use of "grams" to determine how to weight the points.

Mantel tests

Another test which is focused on statistical significance but which does not produce graphs nor identify scales of importance is the Mantel test. A Mantel test starts with two distance matrices and performs a randomization test that returns a correlation coefficient r and a significance value p on whether the two measures of distance are correlated or not. In this application, each matrix would be square with one row and column for each point. One matrix would hold the physical distance $||i-j||$ between each pair of points in the appropriate cells of the matrix. The other matrix would hold the difference z_i-z_j. The Mantel test can be run and if $p<0.05$ then distance between points is significantly correlated to the value observed at those points with a strength and sign given by r. Mantel tests work by a special sort of permutation. Because a point is represented by both a row and a column, one can not just randomly reshuffle the entries in a matrix, but must simultaneously shuffle rows and columns. Under this constraint a large number (e.g. 999) permutations of one matrix is done and then a simple Pearson's r correlation is calculated between the entries in matrix A and matrix B. Pearson r correlation is also calculated on the uncorrelated matrix. This is returned as the r value, and its significance is

determined from the percentile of this unpermuted r amongst all of the r's calculated from permuted matrices.

2nd order processes in point process data

Early on I set out a fundamental distinction between data where the point locations were driven by nature and where they were chosen by the human. The "gram" analyses in the previous section are the dominant tool for studying second-order effects in human-driven data. They look at differences in z (z_i-z_j) as a function of distance between the points (d_{ij}). In point-process data, we look instead at number of points within a neighbourhood of distance d. The F and G statistics mentioned earlier are examples of this but in their focus on nearest neighbours can be dominated by smaller scale second-order effects. The superior approach is known as Ripley's K. This method gives the average number of points with a neighbourhood of radius d around the points in the point process. This is then normalized by dividing by the intensity. By exploring various radii around a point, it avoids the problems of nearest-neighbor statistics. K(d) is simple to estimate. If $N(d,p_i)$ is the number of points in a radius d around point p_i (not including the point p_i), then

$$K(d) = \frac{1}{n\lambda}\sum_{i=1}^{n} N(d, p_i) \qquad L(d) = r - \sqrt{\frac{K(d)}{\pi}}$$

Under the Poisson case, the average number of points in an area A is λA and for a circle of radius d it is $\lambda\pi r^2$. , K(d)=average # points/$\lambda=\pi r^2$. This fact is used to create the closely related statistic L where L(d)=0 for all values of d under CSR. When L(d)>0 then the points are overdispersed at scale d, and clumped if L(d)<0 (note the simple menomic; over 0 implies overdispersed, under 0 = underdispersed=clumped). Unfortunately L(d) is not entirely standardized and some people use L*(d)= -L(d), so one most read carefully when looking at L(d) diagrams.

Conceptually, K(d) and L(d) are similar to the semivariogram and the correlogram. K(d) and semivariograms are cumulative monotonically increasing. Correlograms and L(d) are referenced relative to zero (with CSR giving a flat line at zero). And locations where the correlogram or L(d) are noticeably above or below the line (equivalently K(d) and semivariogram change slope sharply) indicate clumping or overdispersion at those scales.

I prefer an ecologically less-well-known alternative to L(d) that is common in physics and known as the pair correlation function g(d)=K'(d)/(2πd) (Wiegand & Moloney 2004). Thus K(d) is a cumulative density function, and g(d), being the rescaled derivative to it is related to the probability density function. The rescaling ensures that g(d)=1 under CSR. Thus from an interpretation point of g(d) plays a very similar role to L(d) and correlograms but it is shifted up to g(d)=1 instead of L(d)=0 or ρ(d)=0.So g(d)>1 equates to clustered patterns and g(d)<1 relates to overdispersed/regular patterns. I prefer g(d) because L(d) is derived from K(d) under the assumption of CSR while g(d) is derived from K(d) more generally with no assumption of pattern. Perhaps the best-known application of g(d) to ecology is by Condit and colleagues (2000) who looked at the spatial aggregation of tropical trees; they called their function omega, Ω(d), but it is identical to g(d) which has much longer precedence outside of ecology.

As with, "grams", the most widely accepted test of significant deviation from CSR is based on randomization. In this case, if there are n points, then some number, say 99, simulations of a Poisson process in the area studied with n points is run. The statistic of interest (say g(d)) is calculated on each simulation and the 95th percentile envelopes can be drawn. Places where the statistic goes outside the envelope are significant. Issues of multiple tests arise as in the tests of correlograms.

9.2.4. Looking for causes – other species and environment

Sections 9.2.2 dealt with first order effects with one variable and section 9.2.3 dealt with second order-effects with one variable. Of course it is very common to have more than one variable. For example one might wish to explore how temperature, soil moisture, landcover or other environmental variables interacts with a dependent variable of interest such as abundance or diversity. Nearly all of the techniques already identified can be applied to two variable problems. So, for example, where a covariogram looks at the average value of $(x_i-\bar{x})(x_j-\bar{x})$, the cross-covariogram looks at the average value of $(y_i-\bar{y})(x_j-\bar{x})$ (where y and x are two different variables and and i and j are still two different points. This is called a crosscovariogram. Crosscorrelograms, crosssemivariograms, and also cross-K(d) and cross-g(d) statistics can also be calculated. These statistics reveal the second order interactions between two variables – e.g. does having high soil moisture nearby make it more (positive correlogram) or less (negative correlogram) likely to have high species richness. Crosscorrelograms can also be used to leverage information from variables that are measured at fewer locations than the target variable of interest in the form of crosskriging. Mantel tests can likewise be used in this context with the two matrices representing distance in two different variables (x_i-x_j) vs (y_i-y_j). These models are all analogous to traditional correlation in that there is symmetry between the variables involved.

If one wants a more quantitatively predictive model and has a presumption about which variables are dependent and independent, then regression techniques are more appropriate. Imagine one is trying to predict species richness (S) as a function of temperature (t) and tree height(h) and has measurements of all three variables at many sites across space. The simplest approach is a classic multivariate regression $S_i=\beta_0+\beta_1*t_i+\beta_2*h_i+\epsilon_i$ for i=1..n fit via ordinary least squares (OLS). The problem with this approach is that a key assumption is that the errors are independent ($cov(\epsilon_i,\epsilon_j)=0$). This is almost never true in a spatial context and would technically count as pseudoreplication (Hurlbert 1984). Aside from this worry, we might actively wish to incorporate first order effects or second order effects. Further complicating the picture is that the second-order effects might come in as an effect of S_j on S_i (spillover effects possibly due to dispersal), t_j on t_i (temperature is nearly always autocorrelated), or by effects of ϵ_j on ϵ_i (autocorrelation of errors). At the extreme we could imagine setting up a regression such as:

$$S_i= \beta_0+\beta_1*t_i+\beta_2*h_i+\gamma WS+\alpha Vt+ [x,y]\zeta+\eta+\epsilon_i \quad (2)$$

where W is a matrix giving spatial distance derived weights (and with 0s on the diagonal so that S_i does not depend on S_i) and γ indicates the strength of such effects (i.e. is a coefficient to be estimated), similarly for V as a spatial weight matrix and α as a coefficient, [x,y] is a matrix of the spatial coordinates of the points with ζ being coefficients (i.e. fitting a linear trend surface) and η is a random term with $E(\eta)=0$ and $Var(\eta)$ being a covariance matrix of second-order effects (possibly derived from a model of a covariogram). The term $[x,y]\zeta$ incoroporates first-order effects (and presumably is capturing effects of environmental variables that were not measured and used in the model). The terms γWS and αVt capture the second-order effects from the dependent and independent variables and are collectively known as "lag" (derived from timeseries lags) or autocovariate terms while η captures second order effects as an "error" term and by depending on a covariance matrix is closely related to the "gram" techniques above. To my knowledge no one has ever been crazy enough to try and fit the entire model give above, but a variety of techniques have been proposed which fit some subset of the model. Such regression techniques that use some subset of equation 2 are known as spatial regression and, being explicitly aware of spatial relationships, are designed to address the shortcomings of OLS on spatial data. A good review of various spatial regression models is by Dormann and colleagues (2007).

Very often when a field sees a new technique introduced, there is a period where methods proliferate and become increasingly confusing and then finally a consolidation phase is reached where the strengths and weaknesses are identified and the toolkit is narrowed back down to one or two techniques. The study of spatial regression (at least in ecology) appears to be just entering this consolidation phase with a robust debate occurring (Lennon 2000; Jetz & Rahbek 2002; Diniz *et al.* 2003; Dormann 2007b, 2007a; Hawkins *et al.* 2007; Kissling & Carl 2008). Although it is still early, I will provide my personal recommendation. First, there is growing evidence that the original non-spatial OLS may not be so bad. A regression produces three results, estimates of the coefficients β, a measure of fit (r^2), and a measure of the significance (chance of the null hypothesis of β=0 being true). Theory predicts, and studies have shown (Dormann 2007b; Hawkins *et al.* 2007; Kissling & Carl 2008) that even in the face of spatial autocorrelation the estimates of β are unbiased (on average correct), almost as efficient (i.e. standard errors of β only slightly larger) as spatial regression estimates, and the r^2 for OLS is lower than in spatial regression but then spatial regression has more parameters. The only real problem with OLS is that the p-values are very wrong (much more Type I error than reported). If one only cares about prediction and not testing then this is not a problem. If one cares about testing, then one has two choices. Simulations have shown that the p-value can be very accurately corrected using Dutelliel's method, which adjusts the degrees of freedom in the model downwards depending on the amount of autocorrelation (more degrees of freedom lost with high autocorrelation) (Dutilleul *et al.* 1993; Dale & Fortin 2002; Fortin & Dale 2005a). The other choice to get accurate p-values is to use a spatial regression model. Which one to use? There is growing evidence that lag-based models can be biased and should not be used (Dormann 2007a). This leaves the error approach where η is incorporated. This has the added advantage of boiling down to the already known GLS (General Least Squares) or the slightly more general LMM (Linear Mixed Model) approach and the use of covariograms to describe η. A further advantage is that these techniques readily generalize to binomial (logistic) and Poisson regression. So my recommendation is to: 1) run an OLS; 2) calculate a correlogram on the residuals y_i-$\hat{y}_i$; 3) if the correlogram shows significant autocorrelation in the residuals then do not use the p-values from the OLS; 4) if you need p-values then a) use spatial regression based on errors (GLS/LMM) if and only if second order effects are of interest to you but otherwise b) just use Dutelliel's correction to the OLS p-values. This recommendation will be enormously controversial but it is my best read of current evidence combined with an inclination to stick with tried and true methods unless there is a compelling reason.

As one might imagine, interpreting all of the coefficients in the regression of equation 2, even if only a subset of it is used, can be overwhelming. An interesting tool is the use of variance partitioning (Borcard *et al.* 1992). Variables can be lumped into groups (environmental=β and spatial=ζ is the most common application), and then the amount of variation explained by each group can be assessed. Of course the environmental and spatial variables are collinear (correlated with each other) and regression does not have the ability to resolve this. The result is that a % variance is assigned to each of four categories: environmental only, spatial only, environmental and spatial combined, and unexplained. Together these percentages sum to one (the sum of the first three is the r^2 of the environmental+spatial model). Calculating these numbers is easy. Run three regressions (env+spatial, env only, and spatial only) and note the r^2. Then %environment only=($r^2_{env+spat}$-r^2_{spat}), %spatial only=($r^2_{env+spat}$-r^2_{env}), %environment/spatial combined=$r^2_{env+spat}$-%environment only-%spatial only=r^2_{env}+r^2_{spat}-$r^2_{env+psat}$ and %unexplained=1-$r^2_{env+spat}$. Very often the combined environment/spatial variance explained is much larger than either factor alone, which is disappointing as this is the least informative category. It is not surprising but worth noting that the proportions assigned to each category can depend on how many variables are

used in each category; dozens of environmental variables and just x-y spatial coordinates biases toward environment explaining a higher proportion, and using a few environment variables and a complex representation of space (e.g. PCNM Dray *et al.* 2006) can tilt things the other way (Jones *et al.* 2008). The variance partitioning approach just described treats each factor as conceptually equal. An alternative approach can treat one factor as having precedence. In this case if we give environment logical primacy over spatial, then %environment=r^2_{env} and %spat=$r^2_{env+spat}$-r^2_{spat}. This was done for example by Lichstein and colleagues (Lichstein *et al.* 2003) and also done implicitly when regression on residuals is used (e.g. Wilcox 1978). There is considerable controversy again over which is the right approach. Ultimately though it is not question of math, it is a question of assumptions and appropriateness for the question at hand.

9.2.5. Software available

I have consciously chosen to emphasize the conceptual unity of spatial statistics. However, in doing this I have swept under the rug a large number of issues. The practical calculations can depend heavily on the exact type of data used. Moreover, I have completely ignored the issue of edge effects (the fact that some points are near the "edge of the earth" or at least the edge of the quadrat raises complications for many methods). There are several methods for dealing with edge effects but they complicate the calculations. Finally statistical significance test are most often done by randomization methods which is not conceptually difficult but requires additional computer code. For all of these reasons I strongly recommend using off-the-shelf software for spatial analysis rather than implementing your own methods. Two excellent and free pieces of software are available for immediate download that handle all of the methods discussed herein. The first is SAM (Spatial Analysis of Macroecological data)(Rangel *et al.* 2006). This is a custom built software package targeted at spatial analysis of ecological data. It contains an easy to use interface and an impressive list of spatial methods. The more generic (and harder to use package) is R, a general purpose statistical package (R Development Core Team 2005). With the addition of readily available packages (libraries) in R including spatial,spatstat and splanc (for point processes) gstat (for kriging) and spdep (for gridded data) (Bivand *et al.* 2008).

9.3. Prospectus

At this point in time spatial statistics probably needs to enter a consolidation phase where the emphasis is on simplification, rejection of outdated techniques, highlighting the underlying unity of methods, and an effort to streamline communication of these methods. In practice, the scientific community is not well incented to do this and it may not happen. Additional development in some areas is needed such as improved methods of testing statistical significance and better understanding of the strengths of different forms of spatial regression. But these are relatively minor.

There are three new techniques that I believe deserve highlighting. First a technique common in the soils literature but that I have seen applied only once in ecology (Kendrick *et al.* 2008) is based on nested models of variograms (also called coregionalisation). Basically the variogram is fit using piecewise regression. This automated breakout of scales immediately suggests (but does not require) distinct processes at these different scales and assigns variance components to make statements about which scales are most important. Unfortunately I am not aware of software commonly used in ecology that performs this test. Another promising technique is geographically weighted regression (GWR). This model is basically a regression on spatial points but the coefficients (i.e.the relative importance of different independent variables)

is allowed to change across space (Fotheringham *et al.* 2002; Wimberly *et al.* 2008). In one example, Wimberly and colleagues found that climate limited tick abundance in the Eastern US but landscape structure was limiting in the Western portions of the range. Instead of treating nonstationarity as a nuisance, this method embraces and measures nonstationarity. Finally, an increasingly promising alternative to spatial regression and the exploration of causal factors of autocorrelation is the development of process-based models. Houchmandzadeh (2008) recently developed a model that predicts the pair correlation function g(d) under the assumption of simple neutral (diffusive) dispersal. Stronger models of how species distribution is affected by the environmental context will ultimately allow for the teasing apart of causality.

For the actual application of spatial statistics to ecological data, I believe we are entering an exciting time where software tools can bury the details and let the users focus on interpreting and learning from their data without having to climb a mountain of technical details first.

9.4. Key points

1. Spatial statistics is broadly organized around zeroth-order, first-order and second-order effects.
2. There is a fundamental distinction between data in which the spatial locations are human-chosen versus nature-chosen. Differences within the human-chosen category (e.g. quadrats vs. transects vs. geospatial data) have important implications for how calculations are performed but conceptually are of little importance and have been exaggerated.
3. Many methods that have been quite popular are now outdated (e.g. F and G functions, variance-to-mean-ratio/ID method, moving average and exponential smoothing) with better alternatives available.
4. The most modern methods can be extremely difficult to calculate by hand and naïve implementations in software are likely to be wrong, but fortunately good software is readily available and should be used.
5. Covariograms/kriging/correlograms are probably the single central concept today and users should familiarize themselves with their use and interpretation (although not their precise calculations). Moran's I and Geary's C are actually members of this set.

Acknowledgements. I greatfully acknowledge Marty Lechowicz, Marcia Waterway and Graham Bell as well as for the sharing of their data on species richness used in figure 2B and figure 4. I also appreciate McGill University for its stewardship of the Gault Nature Reserve from which the data came and for the many enjoyable hours I personally spent there.

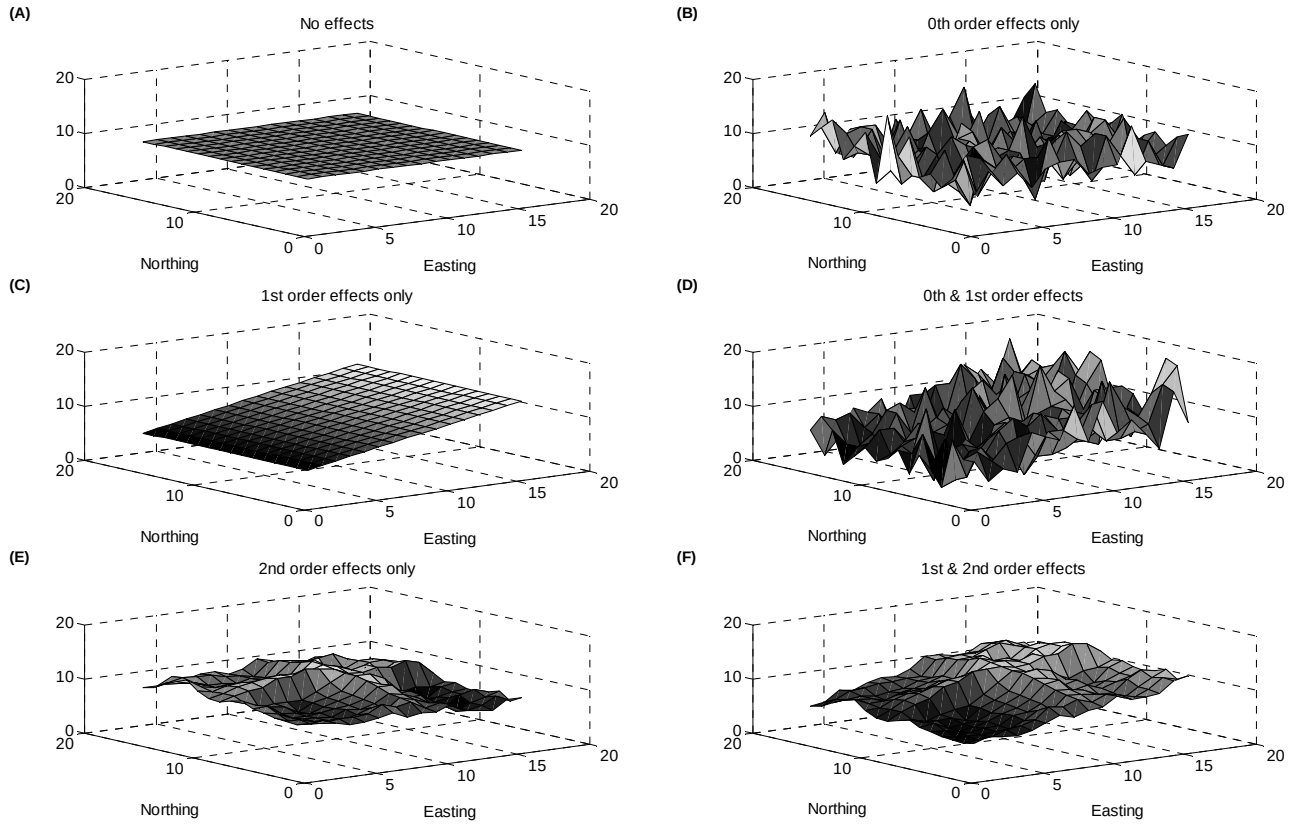
Spatial sampling design		Binary Variable (Present/Absent)	Continuous Variable (Amount/Abundance)
Sampler controlled	Regular across space	"Presence/absence quadrat"	Quadrat
	Random across space	"Presence/absence geospatial"	Geospatial
Biology controlled		Point-process	Marked point-process

750

Table 1 – Major categories of spatial data. The

		Point locations	Quadrat counts	Geospatial
One variable only	**What is intensity (interpolation & smoothing)**	Grid, kernel smoothing	Local smoothing Trend surfaces Kriging	Local smoothing (with interpolation) Trend surfaces Kriging
	What is effect at distance? Are they aggregated?	Ripley's K (# of events within radius)	Moran's I Mantel	Covariograms etc Spectral Mantel
Two variables (dependent, independent)	**What controls intensity**	Intensity regression	spatial regression	spatial regression
	Interactions occur at distance	Cross K	Mantel	Cross variogram Mantel

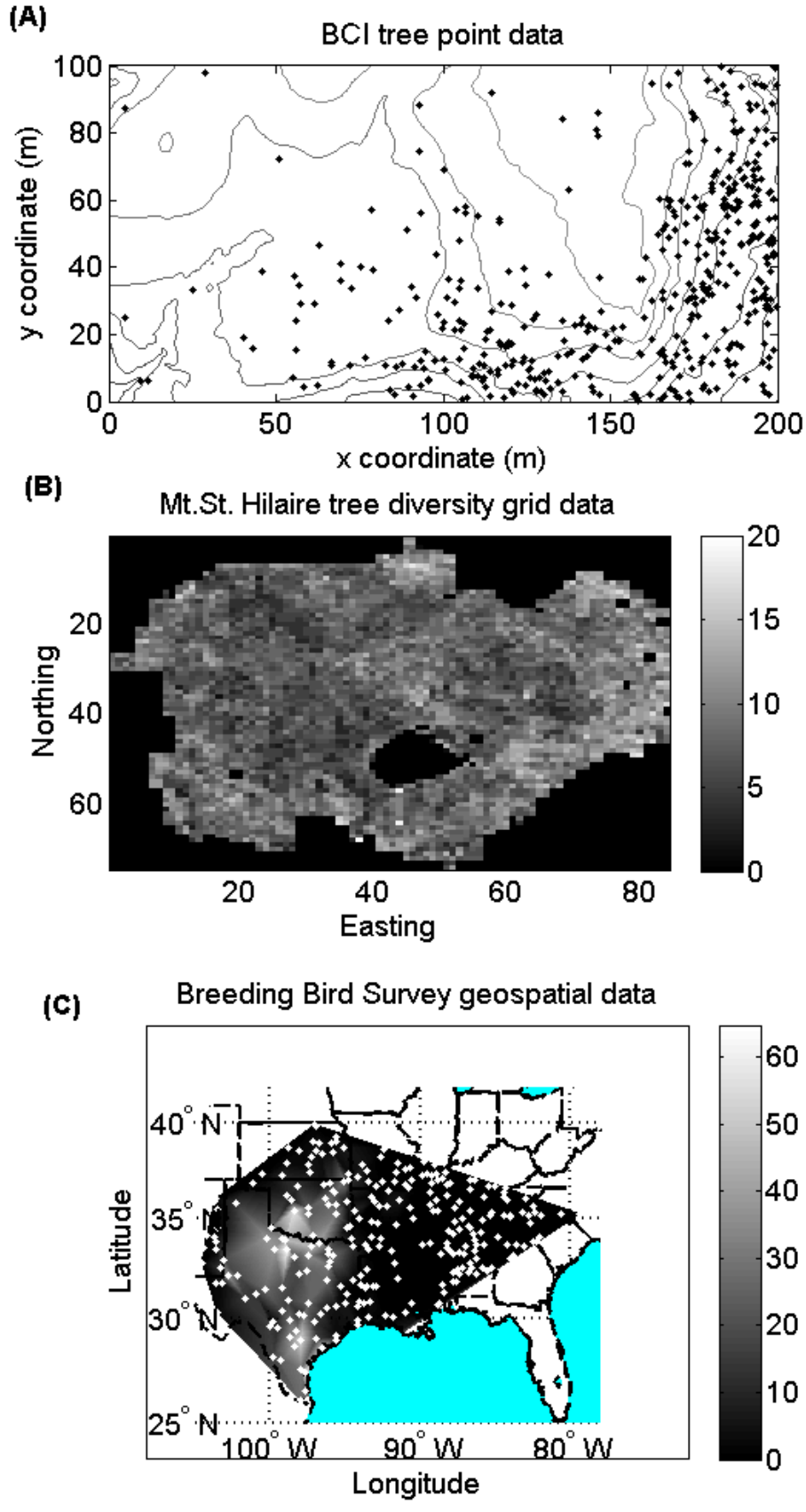
Table 2 – common spatial statistics applied depending on the type of spatial data available, the question asked, and the number of variables measured.



765

Figure 1 – Different order effects in spatial modelling. a) No effects – the measurement variable (possibly abundance of a species, possibly species diversity, possibly soil moisture) is entirely constant. b) Only zeroth-order (measurement error, innate variability or noise) effects.
770 The variable is on average constant but there is variability. The variability is completely independent, even between adjacent points. c) 1st order effect – a systematic change in the mean of the variable across space. Here a simple linear trend is modeled, but of course if the mean is tracking some underlying variable like soil depth the system can look rugged and irregular even with only 1st order effects. d) Combined 0th and 1st order effects. Here there is variability due to
775 0th order effects overlayed on top of the same 1st order trend in the mean found in figure (c). e) 2nd order effects only. The mean is constant across space and no measurement error is modeled.

However there is autocorrelated variability. When one point is above the mean, adjacent points are more likely (although not guaranteed) to also be above the mean. This starts to give rise to coherent features that are spread across several points, like the peak in the back corner or the trough running across the front corner. f) 2^{nd} order effects (figure e) overlayed on top of first order effects (figure c).


(A)
BCI tree point data
y coordinate (m)
x coordinate (m)
0
20
40
60
80
100
0
50
100
150
200
(B)
Mt.St. Hilaire tree diversity grid data
Northing
Easting
20
40
60
80
0
5
10
15
20
(C)
Breeding Bird Survey geospatial data
Latitude
Longitude
40° N
35° N
30° N
25° N
100° W
90° W
80° W
0
10
20
30
40
50
60

785

Figure 2 – different types of spatial data. a) Point data from Barro Colorado Island's tropical tree study. This map shows the location of individual tree of the species *Licania platypus* (with a diameter at breast height of 1cm or greater). Elevation contours are plotted as well, showing the plant is biased towards slopes (Condit *et al.* 2000; Harms *et al.* 2001). b) This plot shows how the species richness of
790 trees varies across the Mt. St. Hilaire nature reserve. The data is collected on a regular grid of 50mX50m cells. c) This plot shows how the abundance of an individual species of a bird (the Scissor-tailed Flycatcher) varies across space (notice that the highest abundances are in Texas and Oklahoma but the range extends all the way east to the Carolinas). The white dots represent routes were the birds were surveyed and are placed irregularly in space. Data from the North American Breeding Bird Survey
795 (Robbins *et al.* 1986; Patuxent Wildlife Research Center 2001).

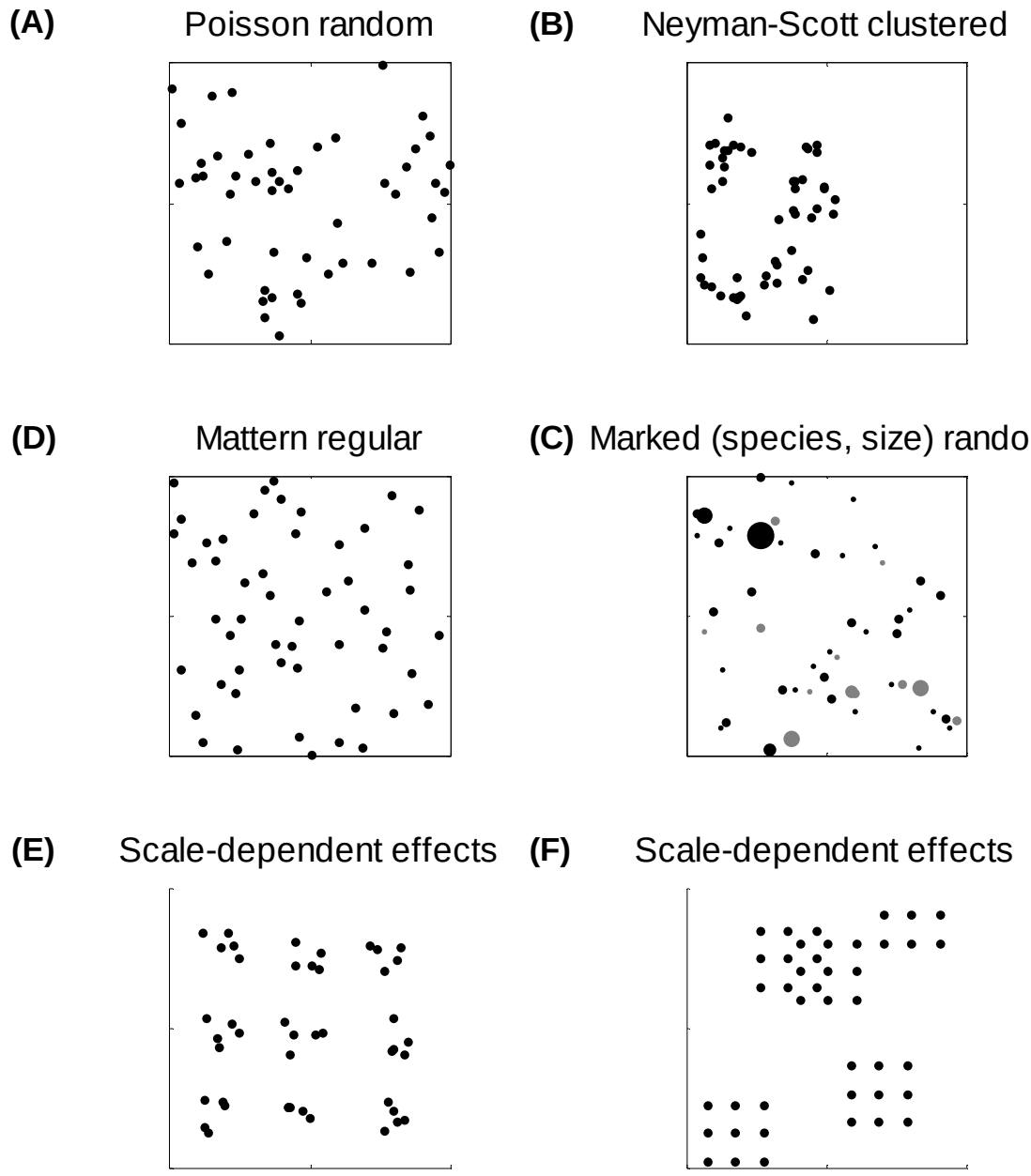


Figure 3 – different kinds of second order (point interaction) processes give rise to different spatial patterns. A) A Poisson random process. B) An example of clustering (attraction between points) simulated by the Neyman-Scott process. C) An example of regular or over-dispersed (repulsion between points) simulated by the Mattern hard-core process. D) A process where the locations are Poisson

random, but the points are marked by two features: size (e.g. tree diameter) and black/gray (e.g. species). E) An example of scale-dependent effects – appears random at small scales, clumped at larger scales, and overdispersed at very large scales. F) Another example scale-dependent effects which appear regular at small scales but random at large scales.

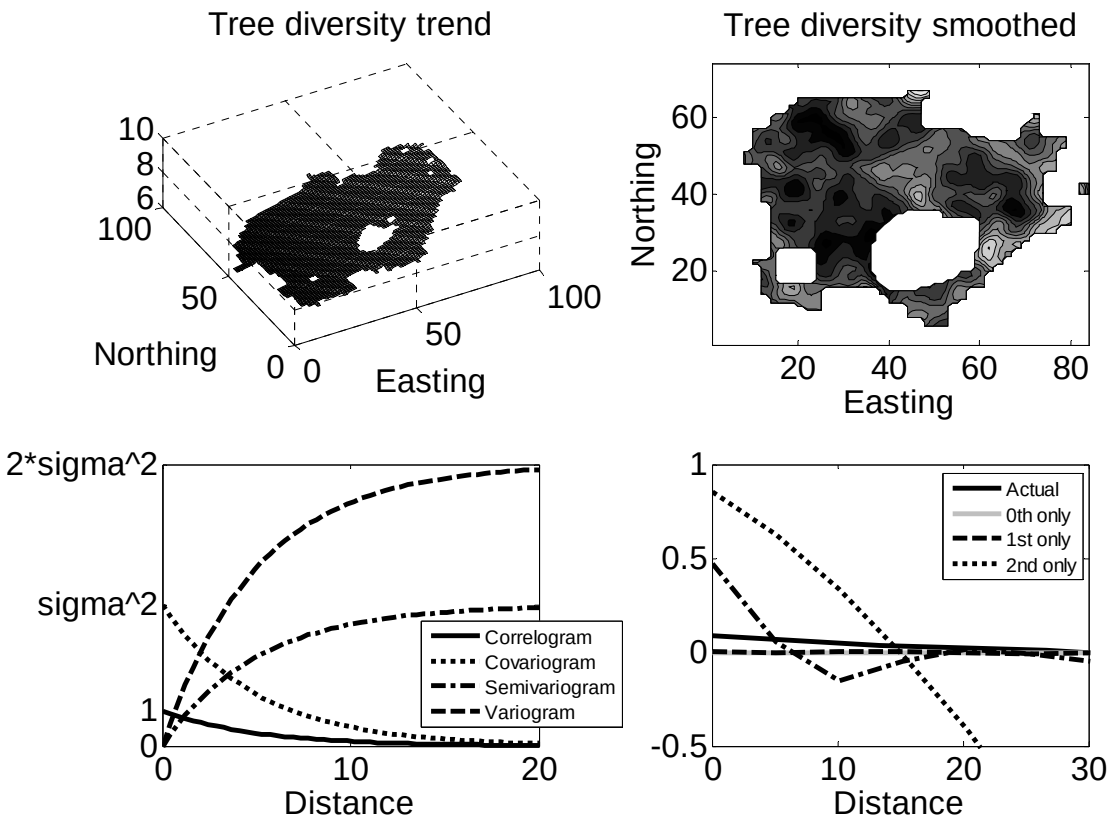


Figure 4 – Example output from analyses of data discussed in the text. A) A trend surface (global smoothing) fitted to the data of Figure 2B. Generally, diversity increases from West to East which also matches a transition from active human use to a protected area. B) The same data kernel-smoothed (Gaussian kernel) and plotted in a colored contour plot. Considerable variation occurs with peaks on the North and East edges of the lake (large empty region in the center). C) 4 types of "grams" on a theoretical data set showing the rescalings that occur. D) Correlograms on theoretical data (Figure 1) with zeroth, firtst and 2nd order effects, and then actual data from Figure 2B. Note that the zeroth order effect shows 0 correlation at all distances – there is no spatial interaction. The 1st order effect starts with a very high correlation at short distances and goes to a very negative correlation at long distances due to the trend surface. The 2nd order effects start with a fairly high correlation in nearby sites, which drops to a negative correlation at

820 the distance at which peaks to valleys are compared and then fades to zero correlation at long distances, indicating no interactions. Thus this analysis gives a clear indication of the scales of interaction. The actual data shows a low, but positive, correlation at short distances fading out to zero correlation at longer distances.

825

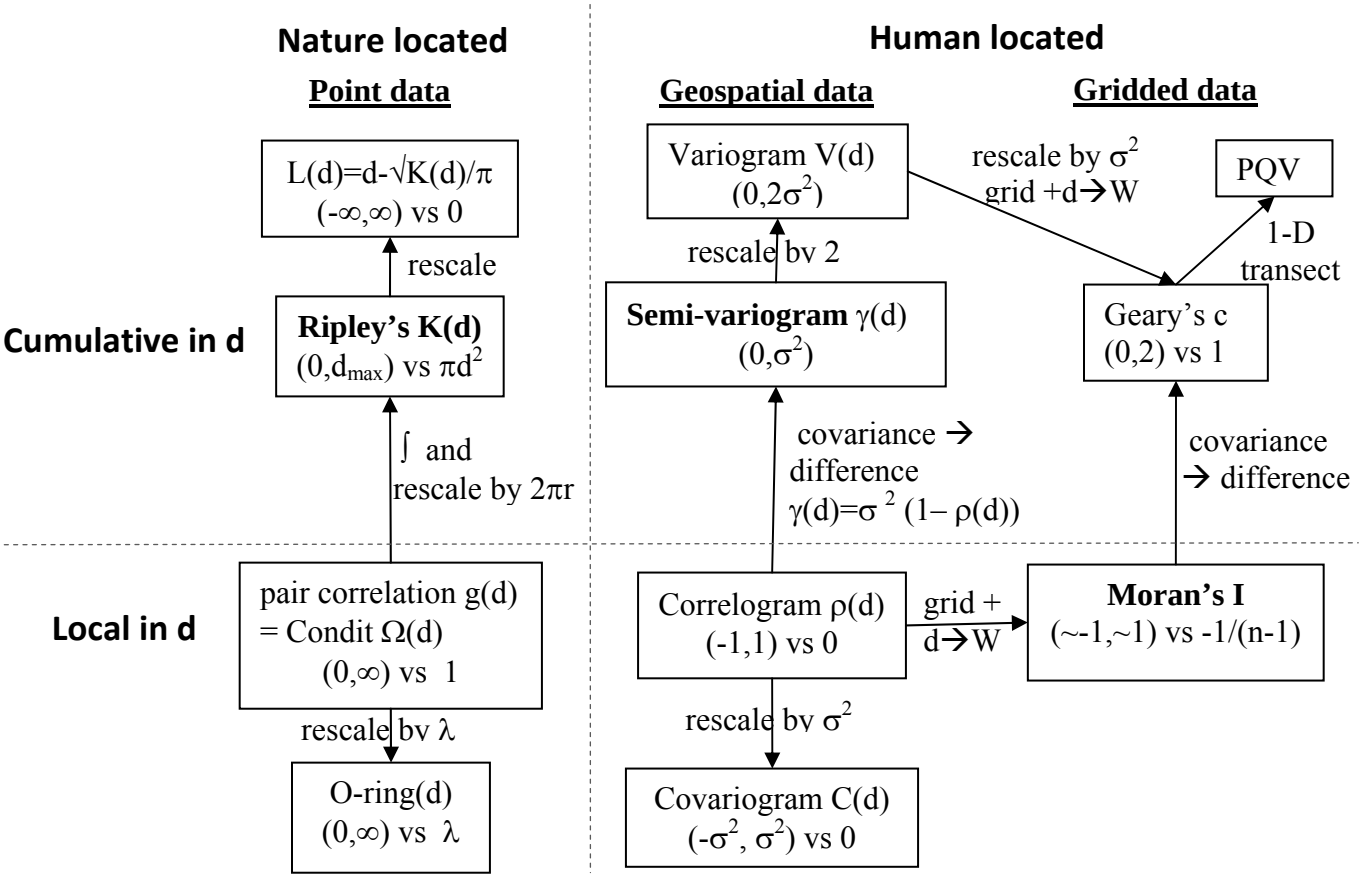


Figure 5 – The relationships between the most common second-order spatial statistics. In almost all case, the conversions are simply a matter of linear rescaling (g(d)=a*f(d)+b) or moving between a cumulative vs. a local form. Minor adaptations are also needed when moving from continuous space (geospatial data) to discrete space (gridded data) or from 2 dimensions to 1-dimension. In the above figure, d=distance between two points, λ=intensity (average # of events per area), W=a connection matrix (on gridded data or on areal data such as a map of states), $\sigma^2=\mathrm{Var}(z_i)$ the variance of the z value across all points in space. The most commonly used technique for each type of data is in bold print. The phrase "grid +d→W" indicates that the data, instead of being located continuously in space, is located on a grid and consequently the idea of a distance, d, between two observations is replaced by a notion of adjacency, summarized in the matrix W. Note that there are no arrows connecting point data to geospatial data, indicating that the match is conceptual rather than mathematical.

Anselin L. (1995) Local indicators of spatial association-LISA. *Geographical analysis*, 27, 93-115

Bivand R.S., Pebesma E.J. & Gómez-Rubio V. (2008) *Applied spatial data analysis with R*. Springer Verlag.

Borcard D., Legendre P. & Drapeau P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, 1045-1055

Brown J.H., Mehlman D.H. & Stevens G.C. (1995) Spatial variation in abundance. *Ecology*, 76, 2028-2043

Condit R., Ashton P.S., Baker P., Bunyavejchewin S., Gunatilleke S., Gunatilleke N., Hubbell S.P., Foster R.B., Itoh A., LaFrankie J.V., Lee H.S., Losos E., Manokaran N., Sukumar R. & Yamakura T. (2000) Spatial patterns in the distribution of tropical tree species. *Science*, 288, 1414-1418

Cressie N. (1992) *Statistics for spatial data*. Wiley Interscience.

Dale M.R.T., Dixon P., Fortin M.-J., Legendre P., Myers D.E. & Rosenberg M.S. (2002) Conceptual and mathematical relationships among methods for spatial analysis. *Ecography*, 25, 558-577

Dale M.R.T. & Fortin M.J. (2002) Spatial autocorrelation and statistical tests in ecology. *Ecoscience*, 9, 162-167

Diggle P.J. (1983) *Statistical analysis of spatial point patterns*. Academic Press London.

Diniz J.A.F., Bini L.M. & Hawkins B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, 12, 53-64

Dormann C.F. (2007a) Assessing the validity of autologistic regression. *Ecological Modelling*, 207, 234-242

Dormann C.F. (2007b) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16, 129-138

Dormann C.F., McPherson J.M., Araújo M.B., Bivand R., Bolliger J., Carl G., Davies R.G., Hirzel A., Jetz W. & Kissling W.D. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609

Dray S., Legendre P. & Peres-Neto P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196, 483-493

Dutilleul P., Clifford P., Richardson S. & Hemon D. (1993) Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, 305-314

Fortin M.J. & Dale M.R.T. (2005a) *Spatial analysis*. Cambridge University Press.

Fortin M.J. & Dale M.R.T. (2005b) *Spatial analysis: a guide for ecologists*. Cambridge University Press.

Fotheringham A.S., Brunsdon C. & Charlton M. (2002) *Geographically weighted regression: the analysis of spatially varying relationships*. Wiley.

Garcia L.V. (2004) Escaping the Bonferroni iron claw in ecological studies. *Oikos*, 105, 657

Harms K.E., Condit R., Hubbell S.P. & Foster R.B. (2001) Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *Journal of Ecology*, 947-959

Hawkins B.A., Diniz-Filho J.A.F., Bini L.M., De Marco P. & Blackburn T.M. (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography*, 30, 375

He F. & LaFrankie J.V. (1997) Distribution patterns of tree species in a Malaysia tropica rain forest. *Journal of Vegetation Science*, 8, 105-114

Houchmandzadeh B. (2008) Neutral clustering in a simple experimental ecological community. *Physical review letters*, 101, 078103

Hurlbert S.H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecological monographs*, 54, 187-211
Hurlbert S.H. (1990) Spatial distribution of the montane unicorn. *Oikos*, 257-271
Jetz W. & Rahbek C. (2002) Geographic Range Size and Determinants of Avian Species Richness. *Science*, 297, 1548-1551
Jones M.M., Tuomisto H., Borcard D., Legendre P., Clark D.B. & Olivas P.C. (2008) Explaining variation in tropical plant community composition: influence of environmental and spatial data quality. *Oecologia*, 155, 593-604
Kendrick G.A., Holmes K.W. & Van Niel K.P. (2008) Multi-scale spatial patterns of three seagrass species with different growth dynamics. *Ecography*, 31, 191
Kissling W.D. & Carl G. (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, 17, 59-71
Konig G. (1835) Die Forst-Mathematik. *Beckersche Buchhandlung, Gotha*
Legendre P. & Legendre L. (1998) *Numerical Ecology*. 2nd edn. Elsevier, Amsterdam.
Lennon J.J. (2000) Red-shifts and red herrings in geographical ecology. *Ecography*, 101-113
Lichstein J.W., Simons T.R., Shriner S.A. & Franzreb K.E. (2003) Spatial Autocorrelation And Autoregressive Models In Ecology. *Ecological Monographs*, 72, 445-463
Patuxent Wildlife Research Center (2001) Breeding Bird Survey FTP site. URL ftp://www.mp2-pwrc.usgs.gov/pub/bbs/Datafiles/
Perry J.N., Liebhold A.M., Rosenberg M.S., Dungan J., Miriti M., Jakomulska A. & Citron-Pousty S. (2002) Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. *Ecography*, 25, 578
Phillips D.L. & MacMahon J.A. (1981) Competition and spacing patterns in desert shrubs. *The Journal of Ecology*, 97-115
Prendergast J.R., Quinn R.M., Lawton J.H., Eversham B.C. & Gibbons D.W. (1993) Rare Species, the Coincidence of Diversity Hotspots and Conservation Strategies. *Nature*, 365, 335-337
Press W.H., Teukolsky S.A., Vetterling W.T. & Flannery B.P. (2007) *Numerical recipes: the art of scientific computing*. Cambridge university press.
R Development Core Team (2005) *R: A language and enviornment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
Rangel T., Diniz-Filho J.A.F. & Bini L.M. (2006) Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecology and Biogeography*, 15, 321-327
Reid W.V. (1998) Biodiversity hotspots. *Trends in Ecology & Evolution*, 13, 275-280
Robbins C.S., Bystrak D. & Geissler P.H. (1986) *The breeding bird survey: its first fifteen years, 1965-1979*. US Dept of the Interior, Fish and Wildlife Service, Washington, DC.
Rossi R.E., Mulla D.J., Journel A.G. & Franz E.H. (1992) Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, 277-314
Tobler W.R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic geography*, 234-240
Wiegand T. & Moloney K.A. (2004) Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, 104, 209
Wilcox B.A. (1978) Supersaturated island faunas: a species-age relationship for lizards on post-pleistocene land-bridge islands. *Science*, 199, 996-998
Wimberly M.C., Yabsley M.J., Baer A.D., Dugan V.G. & Davidson W.R. (2008) Spatial heterogeneity of climate and land-cover constraints on distributions of tick-borne pathogens. *Global Ecology and Biogeography*, 17, 189-202