Luke Spadavecchia

Thinking Spatially

When people ask me what I do, I generally tell them I'm an Ecologist. This is not untrue; I did my first degree here at Edinburgh in Ecology. But I have a guilty secret; I spend most of my time working with spatial statistics, remote sensing and ecosystem modelling. Try telling people that you are interested in statistics, and they switch off pretty quickly. I'm going to try to convince you that they are both interesting and important.

Spatial statistics offers a set of methods to work with data at known locations in space an time. I have always found this exciting, especially in ecology, where we work with spatially distributed processes all the time. I once read that ecology concerns itself with the question of why the Earth isn't covered in a thin layer of green slime: That is, why is biomass clustered? Why are species and communities distributed as they are? Does it seem strange to anyone else that in a subject entirely concerned with spatial phenomena, we choose to ignore the coordinates of our data?

The reasons for this lie in our dependence on the seminal contributions Fisher made to agricultural statistics, where spatial effects were viewed as a nuisance. That and the fact that spatial stats tend to look a little scarier at first glance, with an abundance of i's and j's to worry about. This is because we now have to worry about at least three values per datum; the location and the sample value. So is it worth all the extra fuss to think spatially?

While Fisher was undoubtedly a genius, and frequentist statistics remain a cornerstone of ecological science, there are alternative ways to deal with our data. Simply plotting the locations of our data increases our understanding of the underlying processes which created them. Are there gradients across the study site? Are the data spatially aggregated? These effects can be judged by eye, or more formally using statistical tests.

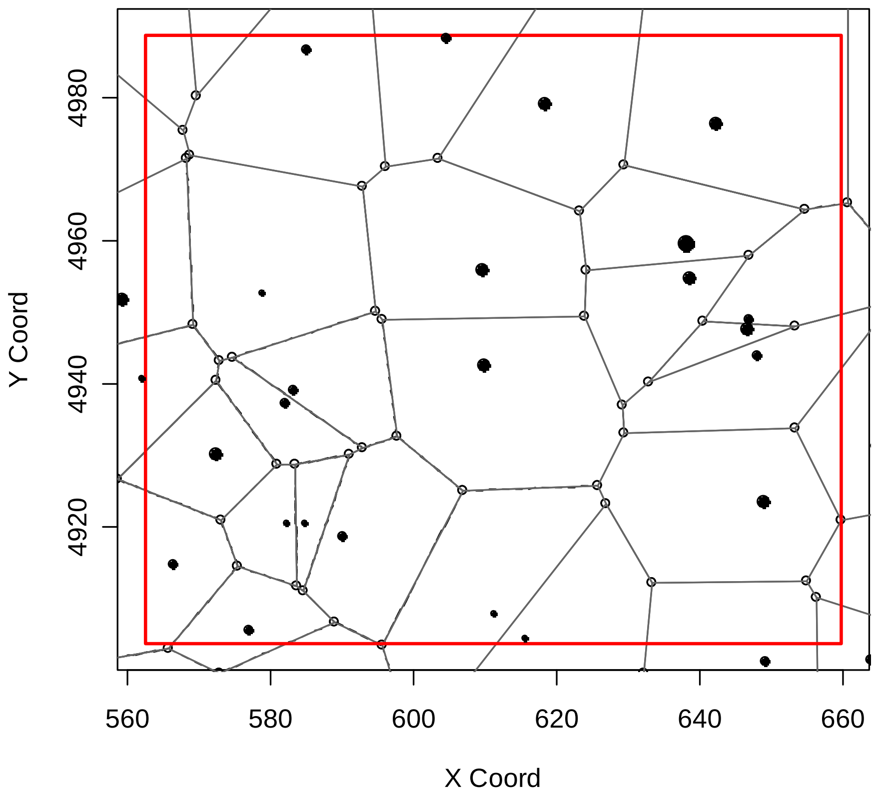

Fig 1: Data plot: Point sizes indicate the magnitude of the data value, i.e. larger points have high values. Gray polygons are the nearest neighbour areas. The non-spatial mean is -1.337, whilst the declustered mean is -1.194.

Even something as simple as taking an average from spatial data is a little more complicated than we might first imagine: As a good ecologist, our first thought is to sample a set of randomly distributed points (an attempt to minimise any spatial effects) and take the arithmetic mean. But what happens if our randomly distributed points are clustered in an area of high data values? The mean of this data will not be representative of the whole region because we have oversampled an abnormal area. Incidentally, the best way to avoid this effect is to sample on a regular grid. To correct for the effect we can do a bit of spatial stats. First we draw a set of polygons around the locations, which delimit the area where each datum is the closest value (nearest neighbour polygons). If we calculate the areas of these polygons, we can use this information to generate a weighted average of the data, which corrects for clustering. This is called a polygon declustering (Fig 1).

Knowing the mean is useful, but we are still throwing away our spatial information. What would be nice is if we could predict values at unsampled locations somehow. During the development of spatial statistics, a family of regression techniques, known as Kriging, were developed to do just this.

We can predict a datum value from a weighted linear combination of the observed data set. How do we decide how much weight to put on each sample location? Logically, we might think it is a good idea to weight data points closest to the prediction location more heavily than those further away, having correctly observed that things close together tend to be similar. This allows us to map the spatial variation across our study area (Fig 2). The technique also allows us to plot the uncertainty of our prediction, although this is a little more involved.

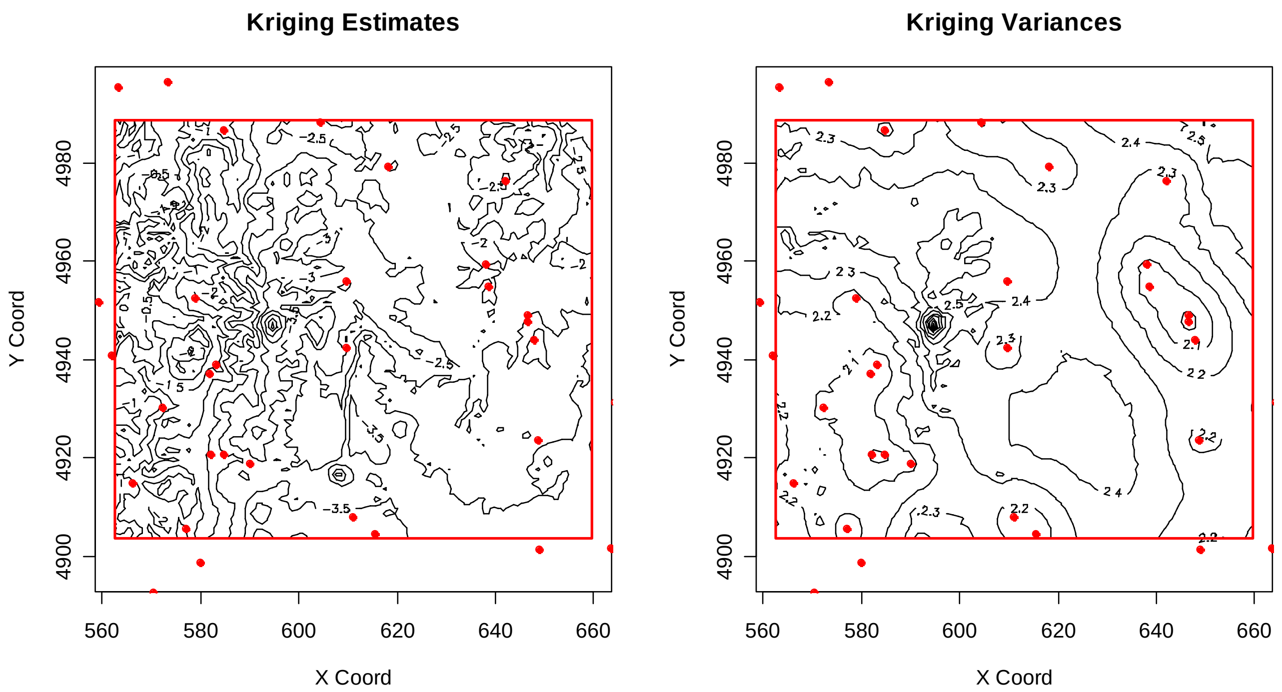

Fig 2: Maps of Kriging estimates and variances for the data shown in Fig 1.

I hope that this short article has intrigued you. Thinking spatially allows us to make more of our data, by making trends and patterns explicit in one of the most powerful ways possible, as a map: As they say, a picture tells a thousand words. For more information visit my website at www.geos.ed.ac.uk/homes/s0198247.