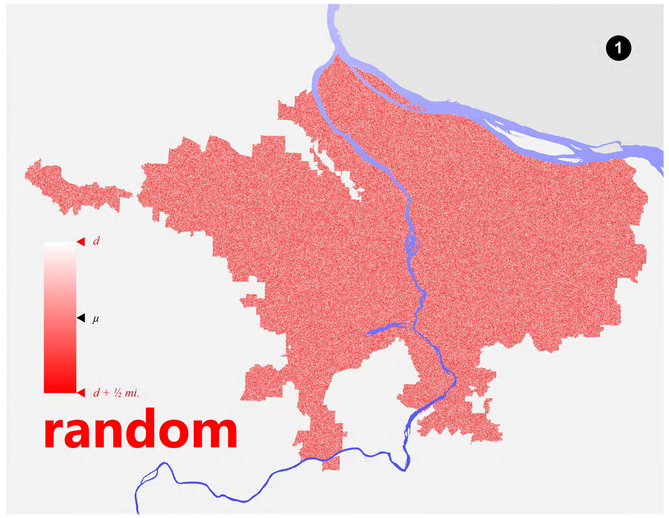
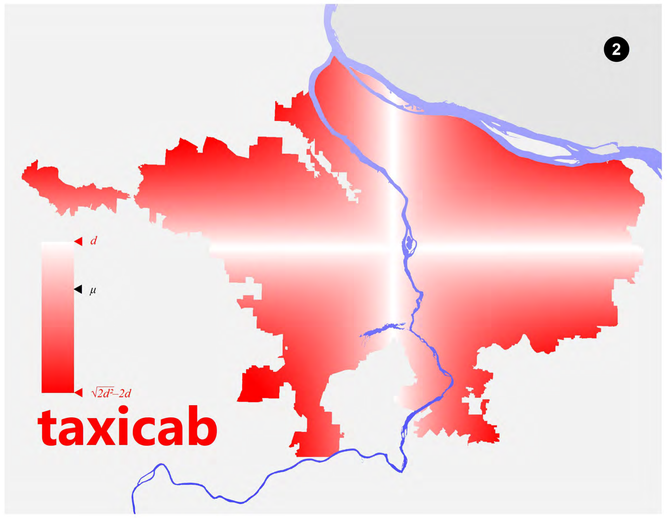
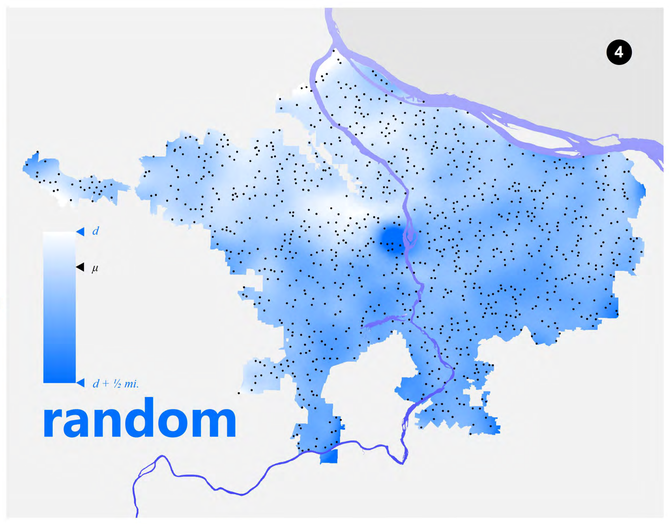
Comparing empirical and interpolated data of random and systematic errors in ArcMap

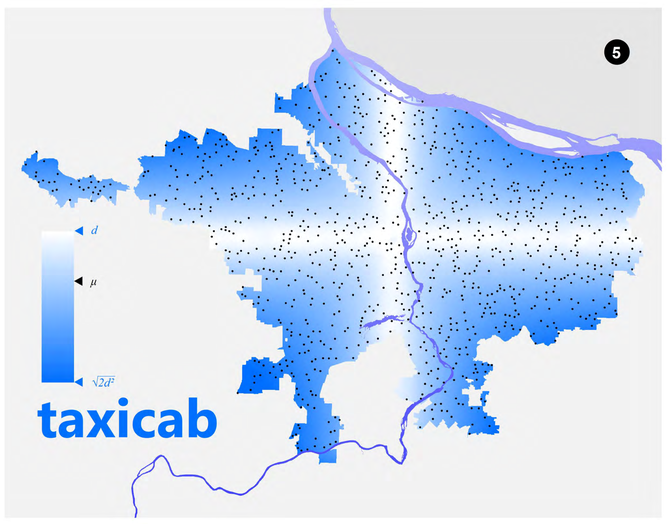
empirical

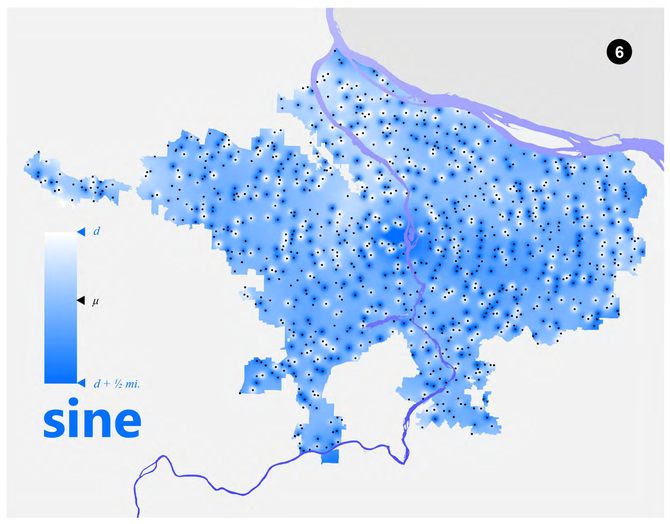




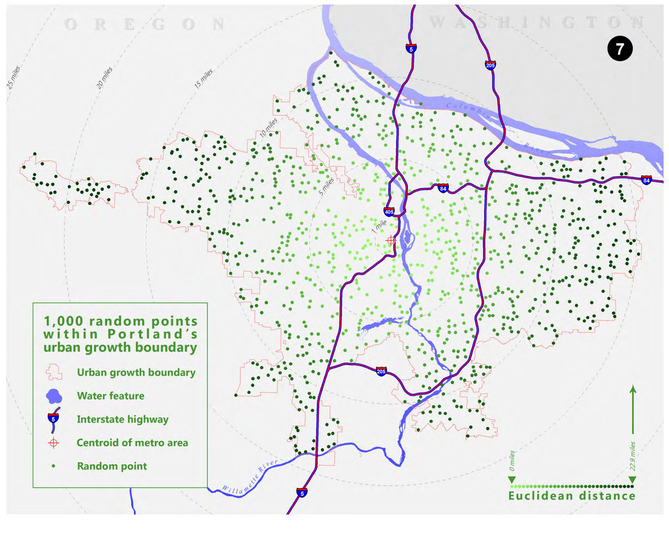
interpolated

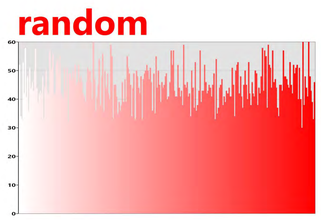


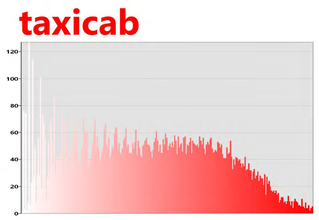


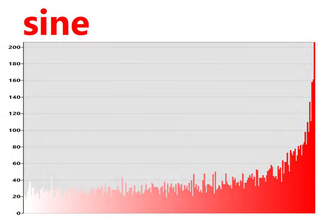


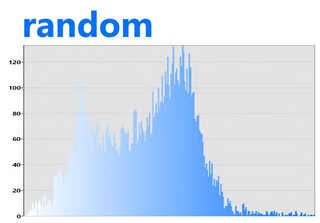
the model

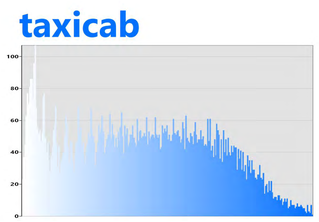


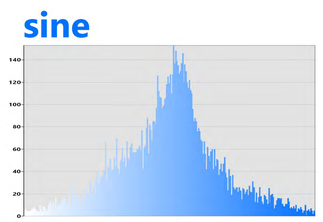












Introduction

As a result of being continuous and well-ordered, distance is the simplest concept that shows spatial autocorrelation. Within the Portland area, we can see the distance from the urban growth boundary's centroid encoded within the points shown in ❼. As we move farther away from the centroid, distance increases uniformly in all directions. Interpolation, a process for computing unknown values based on known measurements, works well on concepts such as distance, and ArcGIS supplies the geographer with several interpolation tools.

To explore the limitations of interpolation within ArcGIS, we begin with the simple concept of distance. To that distance, small quantites of error will be added. Finally, samples will be taken and fed back into the model to see if any kinds of error are well tolerated for reconstructing the original conditions.

Data & Methods

The location for this task was unimportant; Portland was circumstantially chosen due to data availability and familiarity. The city's urban growth boundary was utilized as a processing extent and its centroid was calculated. The empirical rasters were created in Python. Each one adds some quantity to the Euclidean distance from the centroid to each cell. ❶ adds a random number from 0 to 2,640 ft. to the cell. ❷ replaces the Euclidean distance function with the non-Euclidean function used in a taxicab geometry. ❸ adds the absolute value of a sine function to each cell, with a maximum addition of 2,640 ft.

From these previous maps, the model chose 1,000 random points at varying distances from the centroid (see ❼). From those points with the values obtained above, the Geostatistical Wizard within ArcMap's Geostatistical Analyst add-on was run. In each case, the data required minor declustering and correction for directionality. The models were adjusted (as recommended) several times until a best-fit condition was obtained, resulting in the maps ❹, ❺, and ❻

Results

In all three cases (❹, ❺, and ❻), the minimum and maximum error values remained true within one percent. Of the three maps, ❺ performed the best, recreating the structure of the error faithfully, having a comparable mean, and showing a similar structure within the histogram. Although ❹ and ❻ performed equally poorly, ❻ would have to be ranked in the middle, if only for its visual resemblance to the empirical model. Despite the visual resemblance, the hisograms are non-comparable and the means are dissimilar. Lastly, with the random maps, there is no visual similarity, and there are differing distributions as well as means.

Conclusion

Kriging, the method used here for interpolation, makes several assumptions about the data. First, it assumes the data is spatially continuous. Although distance itself is a continuous phenomenon, the errors added here are not necessarily continuous (such as with the random model). The taxicab model performed especially well because of continuity. Another assumption of kriging is that the data is normally distributed. The random and sine distributions are clearly not normal, and as the kriging model tried to normalize them, the original data pattern became lost.

Distance lends itself well to interpolation methods like kriging. Other variables's interaction with distance can greatly influence how well interpolation can reveal the empirical pattern behind the data.

Sources

DEM: NOAA National Centers for Environmental Information, GLOBE (Global Land One-km Base Elevation) Project, May 1999.

Starfield: Donald R. Pettit, NASA, ISS-E-40544, March 2003.

World Countries: ArcWorld, US Census, CIA Factbook.

Portland Area Features: Oregon Metro RLIS.

Cartography

Map by Todd Albertson
March 22, 2017