The sensitivity of pattern analysis due to cartographic scale and map complexity

Elizabeth A. Wentz
School of Geographical Sciences
Arizona State University
Tempe, AZ 85287-0104
wentz@asu.edu

Introduction

The results of pattern analysis techniques such as Moran's I, LISA and other indicators of spatial autocorrelation is known to vary at different geographic scales (McGarigal et al. 2002). When scale is defined by geographic extent, varying scale can mean that pattern results vary from negative spatial autocorrelation, no spatial autocorrelation, to positive. Other pattern indices, such as measures of composition or complexity, also vary when scale changes. Attribute composition, when measured by the total number of attribute classes, increases when cartographic scale increases. That is, when maps become more detailed (larger scale), the number of categories also increases. The importance of scale and the relationship to pattern analysis are seemingly well understood.

There are two components on the relationship between pattern and scale that remain unclear. The first is which pattern indices are the most sensitive to scale changes and to what extent. An important component of this analysis is what aspects of pattern are measured by the indices that are most and least sensitive to scale. The second is how variations of pattern differ when the density of geographic entities varies. The inherent relationship between scale and object density is due to increased generalization associated with decreased scale. The goal of this research is to report on the sensitivity of pattern metrics to the scale changes and geographic object density. Geographic object density can vary due to scale and the processes that influence them. While the question of pattern sensitivity to scale is purely descriptive by design, it will enhance our understanding of the strengths and limits to these measures.

The method of assessment will be to empirically evaluate six different pattern analysis approaches in the context of the three different density contexts. One of the datasets used will be simulated so that there will be known controls over accuracy, scale, and the spatial distribution of geographic objects. The pattern analysis techniques that will be evaluated will reflect measures of composition and configuration.

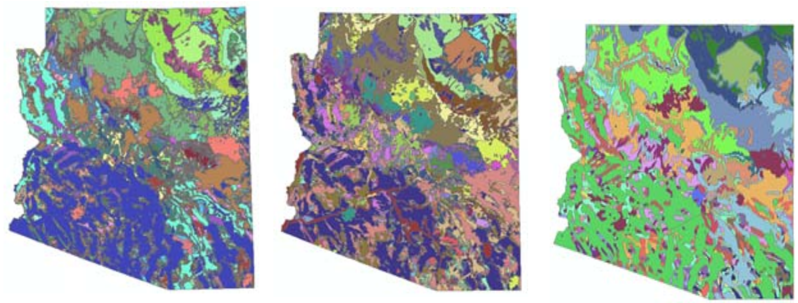
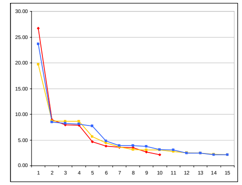


Figure 2. Boundaries of Arizona geology from data sources at three different scales; a) 1:500,000; b) 1:1,000,000; c) 1:2,500,000.



This graphic shows the percent of the total area for the geology types that are greater than 2% of the total study area (Table 3). Red line shows the data from the 1:500,000 scale; gold line shows the 1:1,000,000; blue line shows the 1:2,500,000.

Geology Type	Total Area (km2)	Mean patch size (km2)	Percent of Total Area	Number of patches
Arizona Geology 1:500,000 (all types greater than 2% of the total area)				
Qs	78503.05	251.06	26.71	313
QTs	26342.67	69.32	8.95	380
Pst	23367.04	120.45	7.94	194
QTb	23069.90	21.20	7.84	1088
JTRgc	13735.44	429.23	4.67	32
pCgr	11280.45	18.61	3.83	606
TRc	10614.40	74.75	3.61	142
Qb	10304.14	20.82	3.50	495
TRm	7905.52	35.61	2.69	222
pCgn	6361.50	29.05	2.16	219
Arizona Geology 1:1,000,000 (all types greater than 2% of the total area)				
Q	57967.39	325.66	19.71	178
Tsy	25529.41	99.34	8.68	257
Tv	25442.67	75.96	8.65	335
P	25440.26	438.63	8.65	58
Qs	16723.33	72.40	5.69	231
Jgc	12920.43	717.80	4.39	18
TrC	10704.59	281.70	3.64	38
QTb	9219.12	200.42	3.13	46
PP	8957.78	105.39	3.05	85
Qy	8927.72	247.99	3.04	36
Xg	8194.09	46.82	2.79	175
TrM	7496.58	144.17	2.55	52
Tby	7242.26	127.06	2.46	57
Tb	6608.41	35.72	2.25	185
MC	6250.83	61.89	2.13	101
Arizona Geology 1:2,500,000 (all types greater than 2% of the total area)				
Q	69827.46	2586.20	23.69	27
Ts	24941.64	692.82	8.46	36
Tpv	24240.77	306.95	8.23	79
Tpc	23835.71	567.52	8.09	42
P2b	22684.43	1193.92	7.70	19
JTr	14071.78	1563.53	4.77	9
P2a	11517.77	426.58	3.91	27
Xg	11383.87	132.37	3.86	86
Qv	11092.32	252.10	3.76	44
Xm	9310.51	141.07	3.16	66
uPz	9105.80	130.08	3.09	70
uK2	7212.39	1030.34	2.45	7
lTv	7193.17	133.21	2.44	54
J	6305.82	262.74	2.14	24
X	6208.39	117.14	2.11	53

Table 2. Total area by category type, mean patch size, percent total area, and number of patches for the three Arizona geology layers.

Methods

Data

Two different data sources were used in this study. The first is a simulated data source so that I have complete control over the size, shape, configuration, and composition of the entities within the geographic display (Figure 1). I used a vector representation of square grid regions with three different sizes of grid squares. It appears as a raster display would except the boundaries of the regions are defined by vector lines. The number of categories per density of areas will increase proportional to the number of vector regions. The categories will be assigned to regions based on a random number generator. The random numbers will be assigned to the unique vector id of each polygon.

The second data source is digital data of geology in the State of Arizona (Figure 2). The original analog maps were at three different scales so the differences in pattern techniques can be compared on real world data rather than the simulated data. The three scales are 1:500,000, 1:1,000,000, and 1:2,500,000. The 1:500,000 of Arizona geology was published in 1983 by the U.S. Geological Survey by Wilson, Moore, and Coope (Hershberg and Pitts 2000). The geology data at 1:1,000,000 was published by the Arizona Geologic Survey in the Lambert Conformal Conic projection published in 1988, and obtained from the United States Bureau of Land Management. The third data source was from the Geology of the Conterminous United States at the scale of 1;2,500,000 obtained from the U.S. Geological Survey Digital Data Series DDS-11, U.S. Geological Survey, Reston, VA (Schruben et.al. 1994). The portion representing only Arizona was extracted for use in this study.

Analysis

To measure composition I will calculate total area, number of categories, and mean patch size per category. I will use Moran's I, nearest neighbor analysis, and fractal analysis to measure configuration of pattern. Moran's I is a global measure of spatial similarity and nearest neighbor analysis measures the clustering of objects. Fractal analysis can be used to measure map complexity. Emerson et al. (1999) reported that the fractal dimension indicates great complexity as pixel size increases and varies depending on the landscape type.

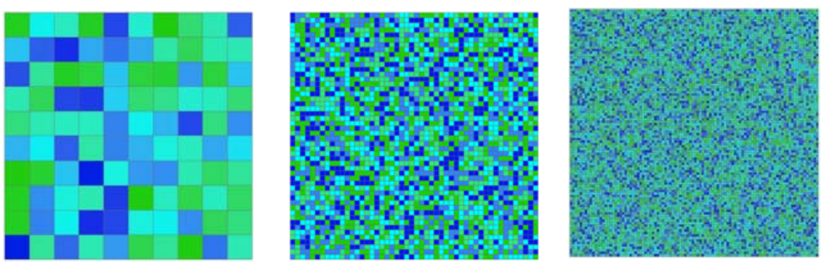


Figure 1. Three different simulated vector data with square grid polygons; a) 10 rows and columns; b) 50 rows and columns; c) 100 rows and columns.

Results

Using the regression relationship (r2 = 0.988) between the total number of categories and the total number of regions in the geology data, I determined the total number of categories needed for the simulated data. The results, along with the summary descriptions of pattern composition, are reported in Table 1. The total area of each category type and mean patch size by category for each of the geology data are reported in Table 2. The results in Table 2 however display only those where the percent study area is greater than 2%. The pattern configuration results are displayed in Table 3. They show that the overall pattern of the geology data are have a random dispersion. The square grid polygon data have a uniform distribution. There appears to be little difference between the data scales.

	Total Area (km2)	Total Number of Categories	Total number of regions (polygons)
Simulated square grid vector data			
10 x 10	10000	35	100
50 x 50	10000	44	2500
100 x 100	10000	73	10000
Arizona geology			
1:500,000	294197.15	77	10806
1:1,000,000	294097.00	46	3415
1:2,500,000	367113.13	41	1001

Table 1. Composition pattern metrics

	Fractal Analysis	Nearest Neighbor Analysis	Moran's I
10 x 10	0.35	2.22*	n/a **
50 x 50	0.23	2.04*	n/a
100 x 100	0.14	2.02*	n/a
1:500,000	1.507	0.66*	0.000 (random)
1:1,000,000	1.463	0.88*	0.001 (random)*
1:2,500,000	1.294	0.72*	0.008 (random)

* denotes a statistically significant z-score
** not enough variation in the attribute value

Table 3. Pattern configuration described by fractal analysis, nearest neighbor analysis, and Moran's I.

Discussion and Conclusions

Pattern composition reported by the total number of classes will vary depending on scale. The results here agree with that assumption and empirical evidence (Table 1). In contrast to that expectation, when analyzing the percent area of different classes and examining only those classes with a percent of total area greater than 2%, the three different scales are represented by approximately the same number of classes (Table 2). In fact, the class type with the greatest percent area was the same geology type (or subtype in the case of the 1:500,000 scale). This is encouraging because it tells us that the results from metrics for pattern composition do not vary much between different data scales. The majority of the class types are represented similarly. Pattern composition metrics are less sensitive to scale changes when examining the data on a per class basis.

A basic result is the total area – of individual classes and the complete study area. The official area of the state of Arizona, as reported by the UC Census, is 294312.06 km2. This value is close to the values reported for the 1:500,000 scale and the 1:1,000,000 scale data. The smallest scale, however, was higher by a much larger margin of error.

As expected, as scale increased, configuration of pattern changed. This conclusion is more notable in the geology data. The fractal dimension, and therefore the geometric complexity, is highest with the smallest scale data. The 1:500,000 scale data also had the smallest nearest neighbor score, showing a random distribution. The Moran's I score also suggested a random distribution but this score was statistically significant for the 1:1,000,000 scale only. The nearest neighbor score showed a uniform distribution for the square grid polygon data, which is what we expected. As the density of the data increased, however, the statistic suggested a less uniform distribution, which could lead to misleading conclusions.

The study here was limited because in the total number of datasets analyzed and the number of pattern metrics applied. A more comprehensive study would have much more variability in the types of data analyzed to explore the range of attribute composition and geographic distributions. The second limitation is the pattern metrics selected. I selected what I assume are commonly applied metrics that represent how attribute composition and geometric configuration are assessed. These particular measures, however, fail to explore local patterns (e.g., LISA measures would be a possible choice) and are limited in the ability to analyze the distribution of nominal data.

References

McGarigal, K., S. A. Cushman, M. C. Neel, and E. Ene. 2002. FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: www.umass.edu/landeco/research/fragstats/fragstats.html

Emerson, C. W., N. S. N Lam, and D. A. Quattrochi 1999. Multi-scale fractal analysis of image texture and pattern. *Photogrammetric Engineering and Remote Sensing* 65(1): 51-61.