Quantifying Resolution sensitivity of spatial autocorrelation: A Resolution Correlogram approach

Pradeep Mohan*, Xun Zhou*, and Shashi Shekhar

University of Minnesota, Minneapolis, USA
{mohan,xun,shekhar}@cs.umn.edu

Abstract. Raster spatial datasets are often analyzed at multiple spatial resolutions to understand natural phenomena such as global climate and land cover patterns. Given such datasets, a collection of user defined resolutions and a neighborhood definition, resolution sensitivity analysis (RSA) quantifies the sensitivity of spatial autocorrelation across different resolutions. RSA is important due to applications such as land cover assessment where it may help to identify appropriate aggregations levels to detect patch sizes of different land cover types. However, Quantifying resolution sensitivity of spatial autocorrelation is challenging for two important reasons: (a) absence of a multi-resolution definition for spatial autocorrelation and (b) possible non-monotone sensitivity of spatial autocorrelation across resolutions. Existing work in spatial analysis (e.g. distance based correlograms) focuses on purely graphical methods and analyzes the distance-sensitivity of spatial autocorrelation. In contrast, this paper explores quantitative methods in addition to graphical methods for RSA. Specifically, we formalize the notion of resolution correlograms(RCs) and present new tools for RSA, namely, rapid change resolution (RCR) detection and stable resolution interval (SRI) detection. We propose a new RSA algorithm that computes RCs, discovers interesting RCRs and SRIs. A case study using a vegetation cover dataset from Africa demonstrates the real world applicability of the proposed algorithm.

Keywords: Resolution correlograms, descriptive correlogram statistics, rapid change resolution, stable resolution intervals, resolution sensitivity analysis

1 Introduction

Many raster spatial datasets exhibit spatial autocorrelation [8, 5, 11, 9, 6, 16, 15], a unique property recognized by Tobler’s famous observation that “all things are related, but nearby things are more related” [16]. GIScience research tradition has christened this observation the *First law of geography* and echoed the need to honor spatial autocorrelation while formulating analytical methods for spatial data analysis [9, 6]. A common application of spatial autocorrelation is land

* Corresponding Authors

cover assessment [12, 10, 3, 4, 19], where it is useful in analyzing raster datasets to determine the natural patch size of the underlying land cover (e.g., patches of plant growth in mountainous regions [12]). Here, spatial autocorrelation analysis provides insights into similarities and differences between different land cover types within and across eco-regions [12, 3, 4]. This analysis helps scientists incorporate spatial dependence structures into multiple regression models that may help predict land cover dynamics [3, 4].

However, spatial autocorrelation has been observed to be sensitive to spatial parameters such as neighborhood definition, e.g., used to define spatial weights matrix [12, 3, 4, 1, 2]. To illustrate, we consider a simple example from land cover assessment. Figure 1(a) shows a sample land cover image dataset where individual cell values is the mean of three bands (e.g. Red, Green and Blue). Figure 1(b)

(a) Input Image at 512×512 Resolution (Mean of the three bands)

(b) Distance sensitivity of spatial autocorrelation (fixed resolution)

Fig. 1: Sample Land cover dataset (Best Viewed in Color)

shows the sensitivity of spatial autocorrelation value as measured by the Moran's I spatial autocorrelation measure. There are three different spatial neighborhood sensitivity plots in this figure, each corresponding to a different spatial resolution of the input image (Figure 1(a)). Each of the sensitivity plots in Figure 1(b) is a *spatial correlogram* [5]. Since these plots reveal the variation in spatial autocorrelation across several neighborhood sizes, these plots can be termed as *distance based correlograms*.

Even though Figure 1(b) reveals the distance based sensitivity of spatial autocorrelation, many applications, particularly land cover assessment require tools that can quantify the sensitivity of spatial autocorrelation (e.g. by detecting useful patterns of variation in spatial autocorrelation) with respect to the resolution (cell size) of the spatial dataset. Such an analysis is particularly useful to produce datasets where the presence or absence of certain patterns in natural phenomena may be dictated by the cell size in the dataset (as observed in existing literature [14, 1, 12, 3, 4]). In this application context, quantifying resolution sensitivity becomes important to provide additional insights into the patterns of variation in spatial autocorrelation [19]. Some important questions posed by this application include, *Are there resolutions where spatial autocorrelation changes rapidly ? Are there resolution ranges where spatial autocorrelation is stable?*. Performing resolution sensitivity analysis (RSA) on raster datasets generated

from applications like land cover assessment and climate science may help answer such questions, allowing users to make informed decisions about the extent of cell aggregation.

In typical application scenarios, users provide a raster spatial dataset similar to the one shown in Figure 1(a) at different spatial resolutions or cell sizes. In addition, information pertaining to a fixed spatial neighborhood and a threshold on the sensitivity of spatial autocorrelation are also provided. Given these inputs, the goal of RSA is to quantify the variability in spatial autocorrelation by reporting interesting patterns in this variability. For example, given a raster dataset such as Figure 1, RSA can identify rapid change resolutions (RCRs) and discover stable resolution intervals (SRIs) by computing *resolution correlograms* (RCs) as opposed to distance correlograms. By identifying these interesting patterns, RSA can provide additional insights for choosing appropriate aggregation levels for raster datasets.

However, performing RSA is challenging for two key reasons: (a) the absence of a multi-resolution definition of spatial autocorrelation and, (b) possible non-monotone variation in spatial autocorrelation across resolutions. For example, the value of Moran's I can change in either direction (positive or negative), implying that overall decreasing trends in autocorrelation may also include increasing trend subsets.

Related Work: Approaches to scale sensitivity analysis of spatial autocorrelation fall broadly into two categories: (a) distance sensitivity analysis (DSA) [12, 3–5] and (b) Resolution sensitivity analysis (RSA). Figure 2 shows the classification of different approaches to scale sensitivity analysis of spatial autocorrelation. The left category, DSA is based on well known methods such as distance based correlograms that are primarily graphical methods. A graphical method of sensitivity analysis may be helpful to un-

Fig. 2: Classification of Related Work

derstand the overall trend in the variability of spatial autocorrelation. However, such methods are limited in quantitative reasoning and do not reveal interesting insights regarding different patterns of change in spatial autocorrelation. Our work addresses this gap by focusing on a quantitative methodology that provides new insights by discovering patterns of change in spatial autocorrelation in addition to providing a basic graphical representation of resolution sensitivity.

Our Contributions: Specifically, this paper makes the following contributions: (a) We define the Resolution sensitivity analysis problem; (b) We formalize the notion of resolution correlograms (RCs) for two popular spatial autocorrelation measures, namely, Moran's I [11] and Geary's C [8]; (c) We provide simple examples via descriptive resolution correlogram statistics that can describe simple trends in spatial autocorrelation across resolutions; (d) We propose a novel RSA algorithm that can compute resolution correlograms, descriptive correlo-

gram statistics and discover interesting patterns, including rapid change resolutions(RCRs) and stable resolution intervals(SRIs);(d)Finally, we provide a case study using a GIMMS vegetation cover dataset from Africa to validate the usefulness of RSA in a real world application setting.

Scope: While the notion of resolution sensitivity can be applied to vector datasets as well, this paper focuses primarily on raster data. The notion of scale can have multiple meanings and definitions. This paper primarily focuses on spatial resolution as a form of scale. Detailed performance evaluation of the RSA algorithm is beyond the scope of this paper. Also, the aim of the case study here is to demonstrate the real world applicability of proposed approaches. A detailed domain interpretation of discovered patterns or insights is beyond the scope of the current work. Finally, the RSA problem described in this paper relies on the user to input data pertaining to different resolutions or to specify a meaningful aggregation scheme for pixels. For simplicity, we make use of a pixel aggregation based mean of neighboring pixels. Other aggregation schemes are beyond the scope of this paper.

Convention: Resolution in raster datasets is usually referred to as the inverse of cell size. In this paper, references to change in resolution implies any change in cell size. Spatial neighborhood can be specified via topological notions such as queen connectivity or via a fixed spatial lag. However, examining differences between use of either spatial neighborhood definition is beyond the scope of this paper.

Outline: The paper is organized as follows: (a) Section 2 formalizes the notion of resolution correlograms and formulates the RSA problem; (b) Section 3 outlines the general layout of the RSA algorithm and describes specific details of RCR detection and SRI discovery; (c) Section 4 evaluates the real world applicability of RSA on a vegetation cover dataset from Africa and reports potentially interesting trends in spatial autocorrelation for this dataset; (d) Section 5 discusses several issues relevant to RSA, including its relationship with other forms of multi-resolution analysis; (f) Finally, Section 6 concludes the paper.

2 Basic Concepts and Problem Statement

This section reviews several basic concepts, formalizes the notion of resolution correlograms, presents descriptive correlogram statistics and formulates the RSA problem.

2.1 Basic Concepts

A **raster spatial dataset** is a sample of continuous natural phenomena as observed by a data collection system such as a sensor or a satellite. Raster datasets consist of simple units called **pixels or picture cells**. Many raster datasets are characterized as matrices of cells and are labeled by the number of cells contained in the rows and columns of this matrix. Figure 1(a) shows a raster dataset with rows and columns containing 512 cells each.

Each cell within a raster dataset may contain some information about a phenomenon that was observed by a data capturing system. This information may

sometimes be distributed across several layers or bands. For example, the raster dataset in Figure 1(b) contains aggregate information, that is, the mean of three individual color bands namely, red, green and blue.

The **Resolution** of a raster dataset is determined based on the physical size of cells, e.g., $1m \times 1m$ or $1km \times 1km$ or $1^\circ \times 1^\circ$ latitude, longitude. Due to the availability of multiple data layers, scientists may choose to create raster datasets at different spatial resolutions of a spatial dataset to understand a natural phenomenon such as land cover. For example, Figure 3 shows the dataset in Figure 1(a) at two different resolutions. The aggregation process usually in-

(a) Aggregated image (cell size = 5) (b) Aggregated image (cell size = 9)

Fig. 3: A raster dataset at multiple resolutions(Best Viewed in Color)
 involves combining neighboring cells to form cells of larger size. For example, in Figure 1(a) the aggregation of 5 neighboring cells yields an aggregated dataset (Figure 3(a)) where cells sizes are 5 times those in the original dataset. Similarly, combining 9 neighboring cells yields a coarser dataset as shown in Figure 3(b). It is apparent from Figure 3 that as the cell size increases, the cell level information changes. An important consequence of this is the possible change in cell neighborhood information.

In spatial analysis, neighborhood information is usually represented as a **spatial weights matrix** denoted as **W** [1]. Cell aggregation in raster datasets essentially changes the original dataset and disturbs the W-Matrix, making it sensitive to a change in spatial resolution. This change in the structure of the W-Matrix makes all spatial autocorrelation analysis techniques sensitive to changes in spatial resolution. Traditional spatial autocorrelation measures such as Moran's I and Geary's C are particularly sensitive to changes in data resolution as well as change in the W-Matrix, especially the row normalized W-Matrix.

When the resolution of a raster dataset is varied via aggregation, the sensitivity in Moran's I and Geary's C with respect to resolution can be graphically represented as a **Resolution Correlogram** as illustrated in Figure 4. Figure 4 (a) and (b) show the Moran resolution correlogram (MRC) and the Geary resolution correlogram (GRC) respectively derived from the image dataset of Figure 1(b). The X-axis in Figure 4(a) and (b) shows different cell sizes corresponding to different resolution levels. The Y-axis shows the spatial autocorrelation level at the corresponding resolution measured using either Moran's I or Geary's C. In addition to a resolution correlogram, many users may be interested in descriptive statistics that can provide insights on the distribution of spatial autocorrelation

Fig. 4: Resolution correlograms for Image dataset in Figure 1(b)

across resolution.

Descriptive resolution correlogram statistics provide a quantitative summary of spatial autocorrelation sensitivity across resolutions. For example,

Fig. 5: Descriptive resolution correlogram statistics

Figure 5(a) and (b) show the histogram with descriptive resolution statistics for the MRC and the GRC respectively. In Figure 5(a) the sample shows a negative skew with the median autocorrelation at 0.469. Also, the sample variance is significantly low (e.g., 0.001) indicating a possible positive spatial autocorrelation across all resolutions. A similar trend is observed in Figure 5(b) for the descriptive GRC statistics. However, descriptive statistics may provide only a summary view of resolution sensitivity. For deeper insight, one may want to detect interesting, useful and non-trivial patterns from resolution correlograms. Two such patterns are: (a) rapid change resolution and (b) stable resolution interval .

A **rapid change resolution (RCR)** represents a pattern of rapid increase or decrease in spatial autocorrelation across resolutions. RCR can be a **change point** or **change interval**. For example, Figure 6(a) shows RCRs computed from the MRC highlighted as red ellipses. RCRs can be quantified using several schemes, including the CUSUM statistic [13] and other measures based on the rate of change in spatial autocorrelation across resolutions [20]. In this figure, the curve with a thick black line and points represented as squares is the MRC.

Fig. 6: Resolution change patterns in spatial autocorrelation

The second curve with dotted lines and data points represented by circles is the CUSUM statistic for the MRC. Figure 6(a) shows dotted red ellipses highlighting RCRs corresponding to intervals and points respectively. The CUSUM value corresponding to the point RCR is also highlighted. While RCRs can help identify resolutions of unstable spatial autocorrelation, discovering resolutions of stable autocorrelation may help scientists make informed decisions to appropriately choose the correct spatial resolution for analysis. A **stable resolution interval** (SRI) can be defined as a collection of resolutions at which spatial autocorrelation is relatively unchanging. For example, Figure 6(b) shows SRIs for the MRC highlighted by dotted blue ellipses.

Problem Statement : Based on the above concepts, we define the RSA problem as follows:

Input: (a) a raster spatial dataset at multiple resolutions; (b) a sensitivity threshold and ; (c) a fixed spatial neighborhood definition.

Output: (a) resolution correlograms (e.g. MRC and GRC); (b) descriptive resolution correlogram statistics; (c) rapid change resolutions; (d) stable resolution intervals.

Constraints: Correctness and Completeness.

Example: In land cover assessment, the input raster spatial dataset may be available at different resolutions as shown in Figure 3. The sensitivity threshold can be defined as a standard deviation threshold (e.g. 0.005). The fixed spatial neighborhood can be defined as a topological neighbor (e.g. queen connectivity) or lag based neighbor (e.g. 1000 meters). Based on these inputs, the goal of the RSA problem is to find the following outputs: (a) Resolution correlograms, MRC and GRC as shown in Figure 4; (b) descriptive resolution correlogram statistics as shown in Figure 5; (c) rapid change resolution as shown in Figure 6(a); and (d) stable resolution intervals, as shown in Figure 6(b).

3 Resolution sensitivity Analysis Algorithm

In this section we describe the general structure of our resolution sensitivity analysis (RSA) algorithm, including steps for detecting intervals of stable resolution and rapid change. The RSA algorithm has four important steps: (a)

Resolution correlogram computation, (b) Descriptive correlogram statistics computation, (c) rapid change resolution detection and, (d) stable resolution interval discovery.

Detailed explanation of steps in RSA algorithm:

Algorithm 1 RSA Algorithm

Input: (a) *Raster Spatial Dataset at different resolutions.*
(b) *Spatial neighborhood size.*
(c) *A sensitivity threshold.*

Output: (a) *Resolution Correlogram for Moran's I and Geary's C.*
(b) *Descriptive correlogram statistics.*
(c) *Abrupt change resolutions.*
(d) *Stable resolution intervals.*

- 1: *Initialize MRC, GRC, MRW*
- 2: **for** $r := 1 \rightarrow \text{maxResolution}$ **do**
- 3: *Compute W – Matrix, $W(r)$*
- 4: $MRW \leftarrow MRW \cup W(r)$
- 5: *Compute Moran's I, $I(r)$ using $W(r)$*
- 6: $MRC \leftarrow MRC \cup I(r)$
- 7: *Compute Geary's C, $C(r)$ using $W(r)$*
- 8: $GRC \leftarrow GRC \cup C(r)$
- 9: **end for**
- 10: *Compute Descriptive Correlogram Statistics*
- 11: *Compute Rapid Changes (RCR)*
- 12: *Compute Stable Resolution Intervals(SRI)*

Steps 2-9 of the algorithm compute the Moran Resolution Correlogram (MRC) and the Geary Resolution Correlogram (GRC). To do this, the algorithm computes the W-Matrix, $W(r)$ corresponding to a resolution r . For example, given a dataset such as the one in Figure 3 and suitable thresholds, these steps compute MRC and GRC as shown in Figure 4.

Step 10 computes different descriptive correlogram statistics, including, the sample correlogram mean, sample variance, sample median, sample skewness, and the sample mode. This step also computes a histogram to represent the population of spatial autocorrelation values.

Step 11 discovers interesting patterns of change in spatial autocorrelation, namely, rapid change. In this step, these are points and intervals where a spatial autocorrelation value that undergoes a sharp increase or decrease is reported.

Step 12 discovers other interesting trends in spatial autocorrelation, namely, resolution intervals where spatial autocorrelation is stable within a sensitivity threshold.

Steps 11 and 12 of the RSA algorithm report interesting, useful and non-trivial trends in spatial autocorrelation based on different resolution correlograms. We provide additional details of these steps in the next section.

3.1 Discovering spatial autocorrelation trends

The RSA algorithm reports two types of patterns in spatial autocorrelation: (a) rapid change resolution (RCR) and (b) stable resolution intervals (SRI)

Fig. 7: Discovering interesting spatial autocorrelation trends(Best viewed in Color)

based on the computed resolution correlograms and user specified sensitivity thresholds. In step 11, the algorithm computes RCRs that include both points and intervals. To detect point RCRs, RSA computes the CUMulative SUM (CUSUM) [13] statistic and reports any resolution that may show a rapid change in spatial autocorrelation value with respect to the mean level. To compute interval RCRs, the RSA algorithm evaluates the persistence of any rapid change within a resolution interval. Since it is also likely that the autocorrelation is non-monotone within a resolution interval, the RSA algorithm makes use of a statistic that is based on the average change in autocorrelation across resolutions to quantify the rate of change. The algorithm picks the RCRs that exceed the top α percentile in slope as “rapid change units.” The score of the statistics for each candidate interval RCR is computed as follows: $score = \frac{AVG(all \ \Delta I)}{AVG(\Delta I \text{ of rapid change units})}$. It can be proved that the value is between 0 and 1, where a larger value indicates that the overall change trend is more likely to be rapid. For a given sensitivity threshold, the algorithm finds all the resolution intervals that have a score that is larger than this threshold, and eliminates shorter intervals that are subsets of any RCR. In the running example of Figure 6(b), we discovered a number of such resolution intervals, including the one from pixel size 33 to 57. A detailed explanation of different enumeration schemes to find the RCR efficiently was explored in our earlier paper [20].

Figure 7 illustrates the RCR computation step. The boxes represent all the resolutions and corresponding intervals. The diagonal line represents individual resolutions. To find a resolution change point, RCR analysis examines all diagonal candidates sequentially and keeps a score until the change point is discovered. To find resolution intervals, the RCR discovery step traverses the entire space of resolution intervals, and computes the score for each one. The figure illustrates

(a) Input data (original resolution) (b) Aggregated data (resolution at 0.35 degrees) (c) Aggregated data (resolution at 0.7 degrees)

Fig. 8: Input data at different resolutions(Best in Color)

a row-wise enumeration strategy where all the intervals ending with the same resolution are examined from longer ones to shorter ones. Detailed pseudo code can be found in our related paper [20].

In step 12, the RSA algorithm computes other interesting resolution ranges (e.g. stable resolution intervals) by making use of an altered score function that can be written as follows: $score = SQRT(AVG(X_i^2) - (AVG(X))^2)$, where $AVG(X_i^2)$ and $(AVG(X))^2$ can be computed using simple functions such as $SUM()$ and $COUNT()$. Techniques such as building lookup tables can be used to further accelerate the computations. Details can be found in our related paper [20]. As shown in Figure 6, we found that the Moran's I value is stable at resolutions ranging from 13 to 25. Also, Figure 7 shows that stable resolution intervals (SRIs) can also be computed in a manner similar to RCRs. The SRIs enumerated are shown as red colored boxes in Figure 7.

4 Case Study

We illustrate the usefulness of RSA through a real world case study on a GIMMS vegetation cover dataset in Africa [17]. Figure 8(a) shows one snapshot of this dataset in August, 1981. Data values are the Normalized Difference Vegetation Index (NDVI) measured between 0 and 1. A larger value indicates more vegetation cover on the ground. Ocean and areas outside the study region are marked with invalid value, and were ignored in the analysis. The dimension of this dataset is 1152 by 1152 pixels, where each original pixel represents about 0.07 degree on the earth surface. Figure 8 shows one snapshot of this dataset at three different resolutions. We applied the RSA algorithm on four snapshots in the dataset, namely, August 1981, November 1981, February 1982 and May 1982, and present the results. These four snapshots of vegetation cover correspond to four different seasons in Africa. This analysis was performed using Matlab 2010b on a 4 Core 2.53G Workstation with a Ubuntu Linux system.

4.1 Method and Results

For this study, we ran the RSA algorithm to compute both the MRC and GRC for 40 different resolutions of the GIMMS vegetation dataset, with cell sizes

ranging from 1 to 40 times the original size. Aggregated images for resolutions 10 and 20 are shown in Figure 8(b) and Figure 8(c). After computing the resolution correlograms, the algorithm also identified RCR intervals and SRIs from these correlograms. Figure 9 shows the results of applying the RSA on the four different seasons of GIMMS vegetation dataset at 40 different resolutions. Figure 9(a) and Figure 9(c) show the MRC and the GRC respectively for the four seasons. The general trend observed in these figures is that the spatial autocorrelation measured by Moran’s I increases at very fine resolutions, reaching a peak and then slowly drops to a lower level. This suggests that the data contains a certain level of local heterogeneity. The turning point of the curve shows the resolution at which autocorrelation or heterogeneity vanish.

This analysis helps in data preprocessing and noise smoothing. However, for different seasons, the turning points vary. As shown in Figure 9(a), in summer (August), the data is more locally heterogeneous as the Moran’s I value reaches its maximum at a coarser resolution. Analogous trends are also observed for the Geary’s C measure in Figure 9(c) where the value of Geary’s C decreases and then increases. This can be interpreted as an increase and then subsequent decrease in spatial autocorrelation with respect to resolution.

Figure 9(b) and 9(d) show different patterns in spatial autocorrelation sensitivity detected by RSA. The thick red lines in Figure 9(b) and 9(d) show the RCR intervals and the dotted blue lines show the SRIs. Among the four seasons, the ones for February (winter) and August (summer) show interesting results. With the sensitivity threshold set to $|\Delta I| \geq 0.005$ and a score threshold of 0.5, we find that the curve for February has hardly any RCR intervals while the August curve has a steadily decreasing interval from resolution 17 to 40 (as shown in Figure 9(b)). With the same sensitivity threshold value of 0.005, the Moran’s I value for the February curve stays stable from resolution 2 to 21 and from 23 to 37, while there is no stable interval found in the August curve using the same threshold. As for the Geary’s C measure, even though the trends are opposite, the interpretation is analogous.

These results are likely due to the fact that the vegetation in the rain forest and grassland are less irrigated by precipitation in the dry winter, which brings down the spatial heterogeneity at large scales. By contrast, the dense rain forest and grassland in summer (August) makes the land cover quite different at large scales, compared to the large area of deserts in the north.

5 Discussion

In this section we briefly discuss several issues relevant to RSA including, first steps to new research directions and multi-resolution tools that do not directly analyze spatial autocorrelation sensitivity.

The RSA algorithm proposed in this paper does not explore any computationally efficient schemes to guarantee better computational performance. Exploring new computational approaches that can speed up the performance of RSA may be another interesting direction for research. For example, in the current formulation of the algorithm, the W-Matrix is computed repeatedly. However, similar

(a) The Moran Correlogram of Africa in August, November 1981 and February, May 1982.(Best Viewed in Color)

(b) Stable and rapid change resolution intervals of the August and February Moran Correlograms(Best Viewed in Color)

(c) The Geary Correlogram of Africa in August, November 1981 and February, May 1982.(Best Viewed in Color)

(d) Stable and rapid change resolution intervals of the August and February Geary Correlograms(Best Viewed in Color)

Fig. 9: Results of RSA on GIMMS vegetation cover dataset from Africa

to data transformation via pixel aggregation, it may also be possible to define new W-transformation techniques that are cheaper than re-computing W itself. This might be a promising direction, particularly in cases where the spatial neighborhood sizes are large. Also, from a spatial database perspective, W computation can be viewed as a spatial join [15]. The spatial database literature has explored a vast family of spatial indices (e.g. Quad Tree, R-Tree) that may be useful for W computation.

While this paper focuses on multi-resolution sensitivity of spatial autocorrelation, multi-resolution analysis itself is a well studied area, particularly dominated by a family of mathematical structures called Wavelets [18, 7]. Wavelets are primarily parametric methods which make use of a fixed set of basis functions to model observations of a natural phenomena that may exist across multiple resolutions. In addition, wavelet based methods assume a fixed aggregation hierarchy for pixels in powers of two, which may or may not represent the reality. In contrast, this paper explored non-parametric methods (e.g., SRI discovery) to discover interesting patterns in resolution sensitivity of spatial autocorrelation.

6 Conclusion

This paper explored the resolution sensitivity of spatial autocorrelation in the context of land cover assessment. This paper formalized the notion of resolution correlograms based on the popular Moran's I and Geary's C measures of spatial autocorrelation. We introduced a new resolution sensitivity analysis algorithm that computes these correlograms, descriptive correlogram statistics and reports interesting patterns of change in spatial autocorrelation. Finally, a case study using the GIMMS vegetation cover dataset from Africa validated the real world applicability of resolution sensitivity analysis. In resolution sensitivity analysis, sometimes it may be useful to aggregate pixels via clusters. In future work, we hope to explore the effect of different aggregation schemes including clustering to generate datasets of coarser resolution. The approach for resolution sensitivity analysis proposed here utilizes global autocorrelation statistics. However, local autocorrelation statistics might provide an enhanced view of variability in spatial autocorrelation levels across resolutions. Hence, in future work, we plan to explore analysis of local resolution sensitivity.

Acknowledgement: We thank the members of the spatial database and data mining group for their feedback on the initial versions of this paper. We also thank Kim Koffolt for helping to improve the readability of this paper. This work was supported by grants from the US Army (W9132V-09-C-0009) and NSF Expeditions in Computing (No. 1029711).

References

1. L. Anselin. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3):247–267, 2002.
2. P. M. Atkinson and N. J. Tate. Spatial scale problems and geostatistical solutions: A review. *The Professional Geographer*, 52(4):607–623, 2000.
3. G. de Koning, A. Veldkamp, and L. Fresco. Land use in ecuador: a statistical analysis at different aggregation levels. *Agriculture, Ecosystems and Environment*, 70(2-3):231 – 247, 1998.
4. G. de Koning, P. Verburg, A. Veldkamp, and L. Fresco. Multi-scale modelling of land use change dynamics in ecuador. *Agricultural Systems*, 61(2):77 – 93, 1999.
5. D. Ebdon. *Statistics in geography*. Blackwell Publisher, 1985.
6. M. Fischer and A. Getis. *Handbook of applied spatial analysis: software tools, methods and applications*. Springer Verlag, 2010.
7. E. Foufoula-Georgiou and P. Kumar. *Wavelets in geophysics*. Number v. 4 in Wavelet analysis and its applications. Academic Press, 1994.
8. R. C. Geary. The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5(3):pp. 115–127+129–146, 1954.
9. M. F. Goodchild. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers*, 94(2):300–303, 2004.
10. J. Ju, S. Gopal, and E. Kolaczyk. On the choice of spatial and categorical scale in remote sensing land cover classification. *Remote Sensing of Environment*, 96(1):62–77, 2005.
11. P. A. P. Moran. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):pp. 243–251, 1948.

12. K. Overmars, G. de Koning, and A. Veldkamp. Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 164(23):257 – 270, 2003.
13. E. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.
14. D. Quattrochi and M. Goodchild. *Scale in remote sensing and GIS*. Mapping Sciences Series. Lewis Publishers, 1997.
15. S. Shekhar and H. Xiong. *Encyclopedia of GIS*. Springer Reference. Springer, 2008.
16. W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:pp. 234–240, 1970.
17. C. Tucker, J. Pinzon, and M. Brown. Global inventory modeling and mapping studies (gimms) satellite drift corrected and noaa-16 incorporated normalized difference vegetation index (ndvi), monthly 1981-2002. *Global Land Cover Facility, University of Maryland*, 2004.
18. A. Willsky. Multiresolution markov models for signal and image processing. *Proceedings of the IEEE*, 90(8):1396–1458, 2002.
19. C. E. Woodcock and A. H. Strahler. The factor of scale in remote sensing. *Remote Sens. Environ.*, 21(3):311–332, Apr. 1987.
20. X. Zhou, S. Shekhar, P. Mohan, S. Liess, and P. K. Snyder. Discovering interesting sub-paths in spatiotemporal datasets: a summary of results. In *GIS*, pages 44–53, 2011.