ALTERNATE DISTANCE METRICS IN SPATIAL STATISTICS: RADIAL ADJUSTMENT OF VECTORS IN EUCLIDEAN NETWORKS

ALFRED T. POWER

Alternate distance metrics in spatial statistics: Radial adjustment of vectors in Euclidean networks

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the

requirements for the degree of

Master of Arts

Department of Geography

Memorial University of Newfoundland

September 19, 2010

St. John's Newfoundland

Table of Contents

Abstract iv

Acknowledgements v

List of Figures vi

1. Introduction 1

  1.1 Introduction 1

  1.2 The objective 4

  1.3 Thesis organization 6

2. Literature review 8

  2.1 Introduction 8

  2.2 Alternate metrics 9

  2.3 Distance 10

  2.4 Alternative distances 13

  2.5 Manifolds 15

  2.6 Connectivity 18

  2.7 Spatial statistics 20

    2.7.1 Fundamental global statistics indices 21

      2.7.1.1 *Standard distance* 21

      2.7.1.2 *Average nearest neighbour* 22

    2.7.2 Spatial autocorrelation indices 22

      2.7.2.1 Various measures of spatial autocorrelation 25

3. Methodology 27

  3.1 The effects of alternate distance metrics on spatial statistics measures 27

  3.2 Spatial context of the analysis area 27

  3.3 Procedure 32

  3.4 Transformation of the Cartesian grid 36

3.5 Procedure classification 37

4. Results and discussion 39

4.1 Comparative results 39

4.2 Overview of transformation results 41

4.3 Selected spatial statistics 43

4.3.1 Distance only statistics 43

4.3.1.1 *Standard distance* 43

4.3.1.2 *Average nearest neighbour* 45

4.3.2 Distance and count statistics 47

4.3.2.1 Spatial autocorrelation (Moran's I) 47

4.3.2.2 High/Low clustering (Getis and Ord General G) 50

4.3.3 Distance and ratio value statistics 52

4.3.3.1 Cluster and outlier analysis - Anselin's Local Moran's I. Total dependency ratio 2006. 53

4.3.3.2 Getis and Ord Gi* (Hot spot) analysis rendering. Total dependency ratio, 2006. 57

4.3.4 Comparative results: Dependency ratios 61

4.3.4.1 Getis and Ord Gi* results 61

4.3.5 Results conclusion 65

5. Conclusions 66

References 69

Appendices 73

Appendix 1 – Selected standard distance results. 74

Appendix 2 – Selected average nearest neighbour results. 76

Abstract

The propensity of utilizing Euclidean distance metrics when calculating spatial statistics generally ignores the underlying connectivity between the features under analysis. A procedure is developed to compensate for the distance discrepancies inherent in spatial statistics algorithms by temporarily transforming the model features into an alternate distance metric space that more realistically represents the functional connectivity distance between spatial elements.

Comparative statistical analysis results between the adjusted and un-adjusted spatial arrangements suggest that statistical measures that are strictly distance based can display dramatic differences in the magnitude of these results. Global autocorrelation measures display much less variation while local autocorrelation measures can result in regions of expanded spatial clustering.

Acknowledgements

I would like to thank my supervisor, Dr. Alvin Simms, for his guidance and support as well as for the freedom to satisfy my curiosity. Thank you to committee members Dr. David Freshwater (University of Kentucky) and Dr. Josh Lepawsky for their expertise that helped to shape the human and academic elements of this paper. I am grateful to Dr. Elizabeth L. Simms for the extensive structural review and to Dr. Paul Plummer (University of Calgary) for his helpful comments and suggestions. I would like to offer a special thank-you to Carole-Anne Coffey for steering me through the thesis process.

I wish to express my appreciation to the Newfoundland and Labrador Statistics Agency for its financial support and to my colleagues for providing a motivating environment.

Last but not least I wish to thank my wife and closest friend Regina for her endless support and dedication. A final thank you to my children Amanda and Colin for your encouragement. I hope that I have instilled an interest in life-long learning to you both.

List of Figures

Figure 1: Hidden metric spaces influence the structure and function of complex networks (Source: Boguñá et al.) 4

Figure 2: Lines and points in Euclidean (1), elliptical (2), and hyperbolic (3) geometries (Source: Pedersen 2005). 12

Figure 3: Avalon Peninsula spatial model features.... 12

Figure 4: Topologically equivalent torus manifolds (Source: Lee, p. 5). 16

Figure 5: Grid transformations (Source: Klinkenberg, 2009).... 18

Figure 6: Euclidean distance (A-B) vs. network distance (A-C-D-E-B). 19

Figure 7: Insular Newfoundland. Settlement distribution and road connectivity. 28

Figure 8: Example of Euclidean vs. road network distance.... 30

Figure 9: Sample radial vector adjustment. 32

Figure 10: Transformation geometry. 33

Figure 11: One dimensional distance transformation. 36

Figure 12: Admiral's Beach origin. Distance-compensated Cartesian grid. 37

Figure 13: Admiral's Beach origin. Distance-compensated grid and landmass. 37

Figure 14: Alternate distance metrics for selected communities (common scale). 42

Figure 15: Standard distance results (common scale).... 44

Figure 16: *Average nearest neighbour* results (common scale). 46

Figure 17: *Average nearest neighbour* exaggerated clustering, Placentia (common scale). 47

Figure 18: Moran's I results, 2001 Population, (common scale). 49

Figure 19: Getis and Ord General G results, 2001 population (common scale). 51

Figure 20: Cluster and outlier analysis – Anselin's Local Moran's I. Total dependency ratio 2006. Whitbourne transformation (common scale). 54

Figure 21: Cluster and outlier analysis – Anselin's Local Moran's I. Total dependency ratio 2006. Bay de Verde transformation (common scale). 55

Figure 22: Cluster and outlier analysis – Anselin's Local Moran's I. Total dependency ratio 2006. Wabana transformation (common scale). 56

Figure 23: Getis and Ord Gi* analysis rendering. Total dependency ratios 2006. Whitbourne transformation (common scale). 58

Figure 24: Getis and Ord Gi* analysis rendering. Total dependency ratios 2006. Bay de Verde transformation (common scale). 59

Figure 25: Getis and Ord Gi* analysis rendering. Total dependency ratios 2006. Wabana transformation (common scale). 60

Figure 26: Getis and Ord Gi*. Aged dependency ratio. Normal metric values. 63

Figure 27: Getis and Ord Gi*. Aged dependency ratio. Alternate metric values. 63

Figure 28: Getis and Ord Gi*. Youth dependency ratio. Normal metric values. 64

Figure 29: Getis andOrd Gi*. Youth dependency ratio. Alternate metric values. 64

Figure 30: Getis and Ord Gi*. Total dependency ratio. Normal metric values. 65

Figure 31: Getis and Ord Gi*. Total dependency ratio. Alternate metric values. 65

1. Introduction

1.1 Introduction

Tobler (2001, para.1) asserts, "The earth is shrivelling as it shrinks", encapsulating the idea that "in contemplating relations between places on the earth … it is often not the geodetic distance that is most important but rather the time or cost which must be overcome. Some places are now closer but others are relatively further away". Massey (1991, p. 24) notes a similar "time-space compression" and the "spatial disruption" that globalization has on the local sense of place. These observations suggest that the concept of human occupied space is amorphous and influenced by available transportation options.

Physical interaction between communities is reflected in the transportation linkages between them. Settlement implicitly suggests a semi-permanent location in which infrastructure develops to facilitate efficacy and convenience, which in turn encourages continued social and economic expansion opportunities. Connectivity within and between communities is fashioned by transportation infrastructure whose forms reflect their historical development context. At sufficient scale, these connective frameworks exhibit topological and network characteristics. Planned communities typically display regular geometric forms while unrestricted development trends towards more composite forms, particularly in areas with complex geomorphology. In older communities, original footpaths or *desire lines* (Bachelard, 1969) often determine subsequent street configurations. The morphology of the resulting connectivity framework between communities affects the logistics of physical human interaction between them. Radical shifts in prevailing transportation technology can render traditional pathways obsolete while imposing new ones. Interconnectivity frameworks between communities can play an influential role in social, political, and economic relationships.

The implementation of regional socio-economic policy is often a nebulous process driven by competing economic, social, and political factors operating largely outside quantitative decision frameworks. Recently however "governments are increasingly being called upon to be more accountable for results" (Sivagnanasothy, 2010). Such developments have given rise to the concept of evidence informed management decision processes that go beyond the mere tallies of inputs, activities, and outputs towards measures of outcomes and impacts (ibid). These processes can change the internal culture of policy decision-making agencies by imposing transparency and accountability obligations. Thus, quantitative methods are increasingly used to justify policy decisions

Spatial analysis is becoming pervasive as a suite of powerful analytical techniques that offer quantifiable support for decision makers. During its development, spatial analysis has addressed several complex issues (such as spatial autocorrelation) that emerge during model abstraction processes. Spatial statistics have their origins in the application of traditional statistical techniques to the attributes of spatially distributed phenomenon. One concern of this approach is based on the idea that spatial statistics contravene geography's only tenet; that of Tobler's 'first law of geography', in which nearer objects tend to be more similar than those further away (Tobler, 1970). Mainstream statistical theory is fundamentally based on the assumption of complete spatial randomness. Despite being initially restricted by non-random aspects of spatially distributed phenomenon, spatial statistics has endeavoured to develop techniques such as geographically weighted regression (Fotheringham et al., 2002) that are compliant with classical statistical theory; a basis that has proven to be essential for academic respectability. The casual adoption of techniques from other research areas such as geostatistical kriging (Matheron, 1965) cannot simply be applied to social issues without a considered evaluation of the assumptions inherent in the physical model. Model parameterizations often do not transfer well from physical to social models.

The following research addresses a minutia common in the implementation of spatial statistics … that of using straight-line distance between discrete spatial features rather than a more realistic connectivity route distance between those features. Pidwirny (2006) contends that "geographers generally conceptualize two types of space: (1) concrete space represents the real world or environment, and (2) abstract space models reality in a way that distils much of the spatial information contained in the real world". The latter abstraction process, while often necessary, can also eliminate critically detrimental components of the environment.

Figure 1 highlights the abstraction between concrete (hidden metric) space and the topological representation (observable network topology) typically used by spatial statistics. Terrain is rendered isotropic and actual physical connectivity is ignored. Undue attention to the spatial location of the features themselves, while ignoring their spatial context, contributes to an over-simplification of the spatial model. The rationale of substituting a Pythagorean distance as a proxy for actual network connection distance is valid only when the path represents the actual physical connectivity. Such circumstances arise when techniques are adopted without a thorough evaluation of the empirical implications. This particular issue stems from the origins of distance calculations in raster data sets. Early work in interpolation was concerned with developing methods to impute field values from (random) samples. The resulting techniques are routinely applied to population (not sample) data sets in vector environments.

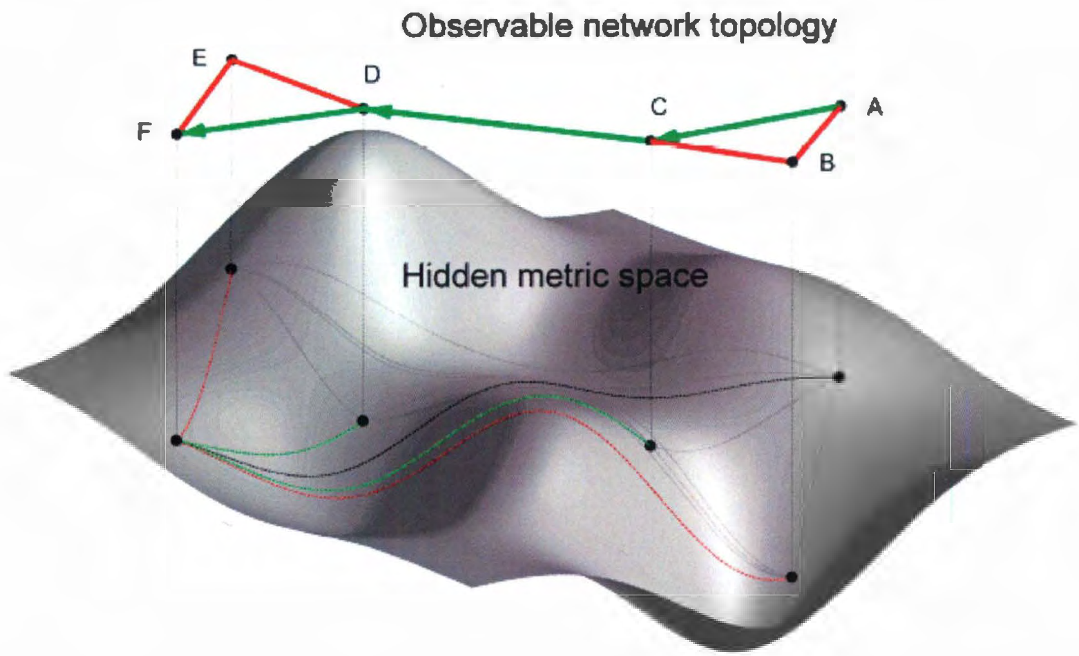


Figure 1: Hidden metric spaces influence the structure and function of complex networks (Source: Boguñá et al.)

It is thus proposed that a relative spatial adjustment of the intervening distance metrics between features should more accurately reflect the true spatial nature between them.

1.2 The objective

The primary objective is to evaluate whether selected spatial statistic measures derived using suitable alternate distance metrics, will result in more intuitive outcomes for community spatial model relationships when they are based on road network connectivity. Non-intuitive results can lead stakeholders to question the merits of spatial analysis due to logical inconsistencies brought about by using unsuitable distance metrics. Mitigating this issue would encourage the increased utilization of spatial statistics within network models.

This research examines the effects of utilizing alternate distance metrics on selected spatial statistics by proposing a method to transform spatial point feature within a Euclidean space into an adjusted spatial metric that compensates for the unacknowledged additional distance burden imposed by transportation networks. Comparative analysis of the results from both metrics is examined to assess whether the differences are of any consequence in the interpretation of the selected spatial indices.

The selected statistics represent three category types based upon the nature of attribute information under consideration. The first type simply examines the spatial distance relationship between features. The second type uses count data associated with the point locations, while the third category considers ratio value attributes.

The initial step involves the development of a systematic means of adjusting the spatial distance metric relationship between point features given the unique spatial association each feature has to other points within the spatial network. Each point feature will serve as the origin of its own particular distance metric, adjusting the relative distance of its neighbours along existing direction vectors. The following approach is utilized:

- Calculation of the Euclidean distance and direction between all point features within a transportation network structure.

- Determination of alternate (vector equivalent) coordinates (from each point of origin) to each of the other points based on travel distance along an existing transportation network.
- Application of a unique polynomial transformation (*rubber-sheeting*) to the Cartesian grid and polygon base map to visualize the associated relative displacement effects of the transformation

The second phase of the research examines the effects of this transformation methodology on a selection of spatial statistics within two broad categories:

- Proximal distance based measures; and
- Spatial autocorrelation measures that examine attribute (non-spatial) numerical characteristics associated with point locations.

In short, selected spatial statistics results are compared under two spatial metrics to highlight the effects that an alternate road distance metric have on these measures.

1.3 Thesis organization

This research paper is divided into four sections. The following section begins with an acknowledgement of methodological issues of applying classical statistical theory to spatial differentiation. A case is subsequently put forth by way of literature review to support the rationale of utilizing an alternate distance metric as a means of mitigating the effects of ignoring spatial context within spatial statistic analyses.

The methodology section develops a procedural technique to compensate for the additional (but often unacknowledged) distance inherent in road network vs. straight-line distance. The proposed technique temporarily transforms a set of points to reflect the actual magnitude of the road distance between points rather than the straight-line distance between them.

The results and discussion section contrasts the results of selected spatial statistics utilizing a normal arrangement of points with the distance compensated spatial arrangement. A simple comparison between results obtained using the two distance metrics is intended to highlight the differing conclusions that may be drawn by considering a compensated distance metric.

The conclusion reiterates the suggestion that location based analyses should consider pragmatic distances between point features and not merely the minimum spatial distance between them. Available transportation options largely determine the physical interaction between communities. Consideration of physical connectivity can enhance the reliability of spatial modeling.

2. Literature review

2.1 Introduction

Recent attempts towards the bridging of cultural and physical geography represent a return to geography's more holistic roots. In antiquity, geography was an all-encompassing endeavour, literally "to describe or write about the earth", via cartography, philosophy, mathematics, and literature. The academic discipline specialization that following the scientific revolution (i.e., physical sciences, biological sciences, social sciences, and humanities), avoided the study of reality in its totality, thus orphaning geography precisely because of its broad focus. Geography has survived as a discipline by bridging the human and physical sciences largely by reclaiming "location" as a centrally defining concept.

As a multi-disciplinary field, geography has often applied techniques developed in other areas to spatially distributed phenomena. Individuals such as Von Thünen sought generalized (nomothetic) econometric insights into specific (ideographic) phenomenon by way of idealized distance decay models. While such idealized models suffer from self-imposed restrictions, location and distance were reinforced as major determining factors in spatial interaction.

Geography's quantitative response to criticisms of its research validity during the 1950s resulted in a new paradigm, heavily dependant on analytical methods of classical statistical theory. This reinvigorated legitimacy revalidated distance and proximity as a primary means of differentiating spatially distributed phenomena. Attempts at implementing aspatial techniques from other fields encountered difficulties when geography's spatial aspect was introduced into the adopted techniques. Such is the case with spatial statistics where the theoretical basis of statistical theory assumes complete (spatial) randomness, an assumption often at odds with the reality of many spatially distributed phenomena. Openshaw (1984, p. 6) is critical of the Faustian dependence on 'plagiarized' techniques that have 'blinkered' geography over recent decades "at the expense of geographical considerations". For example, notwithstanding the dismissal of environmental determinism, there are instances where environmentally deterministic constraints are relevant, such as transportation networks, often constructed under fiscal and engineering restraints. Underlying environmental constraints can influence the context of spatial distributions and should be of concern to geographers rather than ignoring them for the expediency of adopted analysis techniques.

2.2 Alternate metrics

Despite widespread acknowledgement of the importance of spatial differentiation in geospatial analysis, the intervening distance metric between features is rarely considered. The propensity for utilizing the universal Euclidean distance metric over all analysis scales can superimpose the distance metric of the Cartesian plane onto features in the environment that actually have intrinsic alternate distance metrics. Such is the case when a road network determines the distance between communities but spatial analysis is conducted on the spatial distribution of communities as if the road network did not exist, using instead the direct line distance between them. This distance assumption has been inappropriately coined "as the crow flies" (as anyone who has observed foraging crows will attest).

Modest literature exists on alternate distance metrics within the field of geography. Other fields, less indoctrinated to rectilinear grids, appear more inclined to consider alternate distance metrics. Ecological sciences in particular utilize water (river) distances along the central channel of streams and rivers, similar to the familiar highway mile system (Ganio et al., 2005; Gardner et al., 2003; Lyon et al., 2008; and Curriero, 2006). Linear referencing systems (LRS) such as dynamic segmentation are familiar distance metrics in transportation and utility network applications (Puu & Beckmann, 2003). Genetic researchers utilize non-Euclidean distances for separation of chromosomes and genes within the three dimensional structure of the DNA double helix (Bozkaya & Ozsoyoglu, 1997). Alternate distance metrics are central to multi-dimensional scaling, popular in marketing and data mining research (Borg & Groenen, 1997).

2.3 Distance

Distance is perhaps the most central and intuitive attribute of spatial relationships. Despite an innate sense of distance, human distance cognition is often distorted by spatial and temporal perception. More specifically, distance is a quantifiable description of the space that separates objects. The measurement units are typically either a physical length, a period of time, or in relation to an arbitrary criteria, (e.g. "two doors down"). These latter types of distances are typically used during normal human interaction where a more relevant reference frame is warranted. The long-standing debate as to whether geographic space is quantifiably distance based or intuitively topological is perhaps dependent on the type and scale of the application. Jiang (1998, p. 54) states that "Human thinking is not metric based" but "maps are the most efficient and effective way of communicating metric properties of large scale space". Thus, as in most debates, both sides have valid arguments for their positions. The two opinions are currently being reconciled in the concept of *naive geography* (Egenhofer & Mark, 1995) where more human-based techniques are under development to facilitate more 'natural' information transferral.

It is in the scientific realm where the concept of distance expands into multi-dimensional domains and where it is subject to definitions that are more rigorous. Mathematically, a distance function (metric) defines the distance between elements of a set, which defines the metric space. Space is simply a "set" that contains elements (points). A metric space requires a "metric" to numerically quantify two points in the space and map them to a number. A distance metric is required and defined in such a way that the shortest distance between any two points is a straight line and must obey three mathematical properties (Table 1).

Table 1: Mathematical conditions required for distance metrics.

Function	Property	Explanation
d(*a*,*b*)>=0	Non-negativity	Distance is always positive or zero.
d(*a*,*b*)=d(*b*,*a*)	Symmetry	The distance between *a* and *b* is the same in either direction.
d(*a*,*b*)+d(*b*,*c*)>=d(*a*,*c*)	Triangle inequality	The shortest distance between two points is a straight line.

Distance calculations within a homogeneous Euclidean metric are easily determined by Pythagorean methods but when the analysis space becomes non-Euclidean such as hyperbolic, elliptical, or otherwise non-planar, Euclid's parallel postulate must be modified, resulting in what is referred to as absolute or neutral geometry. Such non-Euclidean relationships allow for topological associations that facilitate alternate linkages between objects within the topological space. See Figure 2.

Figure 2: Lines and points in Euclidean (1), elliptical (2), and hyperbolic (3) geometries (Source: Pedersen 2005).

Irregular network connectivity imposes a variable distance metric contrary to that assumed by most categories of spatial statistics. Most spatial statistic processes assume a barrier-free Euclidean separation between features that, while over-emphasizing the strength of local relationships, can also result in forcing a neighbour relationship where none exists. Figure 3 presents the typical spatial model primitives for the Avalon Peninsula of Newfoundland. Most spatial statistic calculations will utilize attributes of the point features (Figure 3a) outside the context of the restricting environment imposed by the road connectivity matrix (Figure 3b) and the landmass configuration (Figure 3c).

Figure 3a: Point features (locations)

Figure 3b: Linear features (constraints)

Figure 3c: Areal features (barriers)

Figure 3: Avalon Peninsula spatial model features.

Ignoring the contextual determinants that establish topological relationships, spatial statistics essentially discounts the actual distance metric by substituting the minimum Euclidean (as the crow flies) distances between point features. It is more realistic to study spatial statistics at the human scale within a network rather than a Cartesian framework. Consequently, this paper proposes to evaluate the concept of transforming the underlying metric space between point features (on a non-regular transportation matrix) to compensate for the effects that actual travel distance have on measures of spatial dependency.

2.4 Alternative distances

Shu et al. (2001) contend, "In geographic space, it is well known that spatial behaviors of humans are directly driven by their spatial cognition, rather than by the physical or geometrical reality". Furthermore, "In the past work, the physical or Euclidean distances are used very often. In practice, many inconsistencies are found between the cognitive distance and the physical distance". Cognitive distance is the perceptually estimated distance between two locations. Individual experience shapes cognitive distance estimations so that "physical distance is mostly overestimated or underestimated in the process of human spatial cognition and spatial behaviors" (Qi et al., 2006, p. 408). An intuitive response to this discrepancy is to use travel time as a surrogate measure for travel distance. People often reply in time units in response to questions of travel distance.

On a more philosophical level, the seminal work 'The Production of Space" (Lefebvre, 1974) has influenced current urban studies into 'socially produced spaces', as opposed to natural or absolute space. Lefebvre's contention that space is a social construction based on values and social production of meanings and thus affecting spatial practices and perceptions has provided researchers such as Harvey (1990) and Löw (2006) a means of conceptualizing space and distance outside the norms of quantifiable space. Advances in communication and transportation technologies are changing modern concepts of distance, space, and time (Massey, op. cit.), a development that lends itself to analysis by network-topology methodologies rather than geometries of absolute locations.

Interactive events typically start with a conscious determination of a destination from an implicit origin. Familiarity with the regional context determines the degree of conscious route planning involved. A routine journey, such as a daily commute, requires less planning than travelling to an unfamiliar destination. Once established, the network rules and connectivity will determine the particulars of the route traversed. Perception of the planned route at an overview scale is linear, but at local human scales, the network details introduce changes in course direction while simultaneously accumulating distance. This exemplifies the issue at hand. Analysis techniques that ignore the functional extra distance between origin and destination events will overstate the degree of association between them by assuming the points are nearer than they functionally are, introducing erroneous topology, as well as confounding aspects of proximity and adjacency. Essentially, the orthogonal Cartesian metric does not reflect the reality of the true travel distance and network connectivity between locations.

A conceptual model is essential to visualize and manipulate the approach to this research topic. Physically analogous models can help to visualize the dynamics of a changing distance metric. The basic components of the proposed model are a set of spatially arranged point features, a connective network, and a terrain surface wherein each set of primitives require spatial adjustment with respect to inherent restrictions imposed by either of the other feature types.

Several techniques offer avenues for investigating a means of compensating for irregular geometric transformation between alternate metric spaces. Mathematical approaches, (including algebraic topology, differential geometry, matrix algebra, and geometric transformation) offer computational procedures for spatial transformations however; most transformation techniques employ a linear or other simplistic uniform alteration to the distance metric during transformation. The aim of the proposed approach is to allow the distance metric to vary while maintaining relative direction. The intention is to develop a method of transforming a point set into an alternate metric space to reflect road network connectivity between communities.

2.5 Manifolds

The following will examine whether the concept of manifolds can provide a conceptual basis for examining points in Euclidean and alternate spaces.

> A manifold is a topological space that is locally Euclidean (i.e., around every point, there is a neighbourhood that is topologically the same as the open unit ball in Rn). To illustrate this idea, consider the ancient belief that the Earth was flat as contrasted with the modern evidence that it is round. The discrepancy arises essentially from the fact that on the small scales [sic] that we see, the Earth does indeed look flat. In general, any object that is nearly "flat" on small scales is a manifold, and so manifolds constitute a generalization of objects we could live on in which we would encounter the round/flat Earth problem, as first codified by Poincaré. More concisely, any object that can be "charted" is a manifold. (Rowland, 2010, para.1)

It is useful to situate the observer's point of view with respect to manifold characteristics. A manifold is a type of space wherein near every point of space there is a coordinate system like that in Euclidean space. An individual traveling in a straight line on the surface of the earth experiences the environment as if it were a normal Euclidean space. To an orbiting external onlooker however, the individual is observed on a spherical surface traveling in an arc. These two viewpoints are considered intrinsic and extrinsic respectively. In addition to a manifold's intrinsic geometry, they also have a geometry inside other spaces, an extrinsic geometry, that depends on how they are mapped into another space. This premise will form the basis of the transformation technique developed in the following section.

Conceptually, this model is suitable for comparing spatial relationships between two different distance metrics. Theoretically, any point on the left hand torus in Figure 4 can be mapped by way of a function to an equivalent corresponding location on the right hand torus (i.e. they are topologically equivalent) but the transformation would require unique transformation parameters for each point. Unfortunately, the complexity involved in the differential calculus necessary for such a continuous transformation precludes the utility of this approach. DeCarlo (1998) determined that approximately seventy shape parameters are required to describe a piecemeal torus to mug transformation. However, a similar approach using a two-dimensional reference grid can greatly simplify the process.

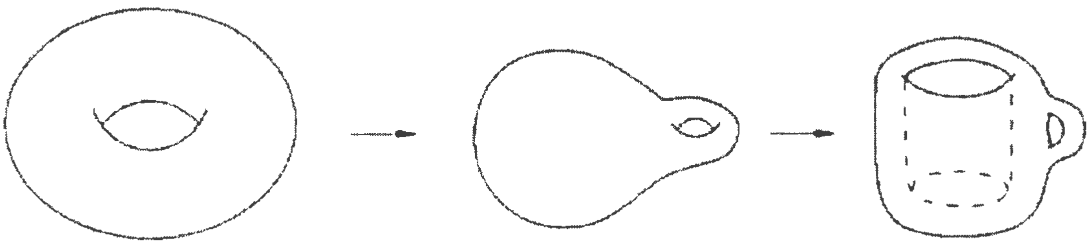


Figure 4: Topologically equivalent torus manifolds **(Source: Lee, p. 5).**

Caveat lector - a coordinate transformation is a conversion from one system to another *to describe the same space*. The technique proposed for this research is to transform to an alternate metric space to facilitate a (geostatistical) calculation and then return to the original (metric) space. The topological ‘space’ remains the same but the change in the distance metric results in an alternate spatial arrangement of the features (similar to deflating a beach ball). Notwithstanding this ‘same space’ criterion, the following will utilize the alternate space merely as a surrogate for (geostatistical) calculation purposes.

Standard linear transformation techniques that employ coordinate shift, scaling, rotation, or skew are ill suitcd to situations where the transformation varies along the vector for each set of points on the surface (Figure 5). Similarly, a first order affine transformation, (which scales differently along different orthogonal axes) is inadequate, because the displacement is of constant distance and direction. Higher order polynomial transformations (warps or *n-order* transformations), will vary to the degree of the number of points under consideration resulting in complex surfaces. Proper transformation techniques require that the differences between the coordinate systems be mathematically systematic. ‘*Rubber-sheeting’* is a suitable *n-order* transformation technique for current purposes due to its local, rather than global manifold transformation properties. “*Rubber-sheeting* is a procedure for adjusting the coordinates of all the data points in a dataset to allow a more accurate match between known locations and a few data points within the dataset. *Rubber-sheeting* preserves the interconnectivity between points and objects through stretching, shrinking, or reorienting their interconnecting lines” (de Smith et al., 2008: p. 18).

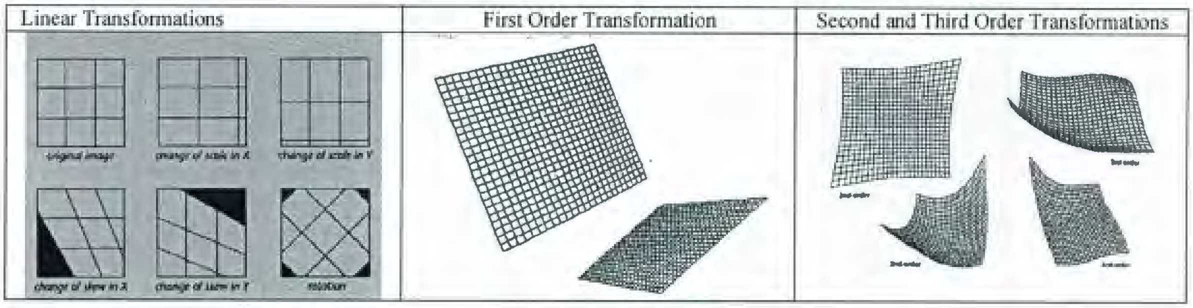


Figure 5: Grid transformations (Source: Klinkenberg, 2009).

It is important to note that the alternate metric space will be unique for each point serving as the origin to all other points. This process is more easily visualised by the transformation of a ruled surface that changes non-uniformly across the surface. Non-linear transformation of the grid allows the retention of relative point displacement.

2.6 Connectivity

Connectivity between features can utilize alternative distance metrics such as rectilinear Manhattan distance, spherical distance, or travel time. Often spatial statistics tacitly assumed that features displaying Euclidean adjacency are proximate neighbours as well. In circumstances such as point locations on a dense regular grid, this proximation assumption can often be ignored without serious consequence, but there are situations where adjacent features are not necessarily neighbours. Natural and artificial barriers influence the magnitude of the distance metric as well as restrict the connectivity between spatial features.

Figure 6 illustrates a model of a series of network nodes on an irregular shaped surface representing a series of communities connected by a road network in a coastal environment. Any number of algebraic, geometrical, or statistical methods can analyze the spatial arrangement of the points in such a model but most methods will disregard the environmental restrictions imposed on the connectivity between the spatial features.

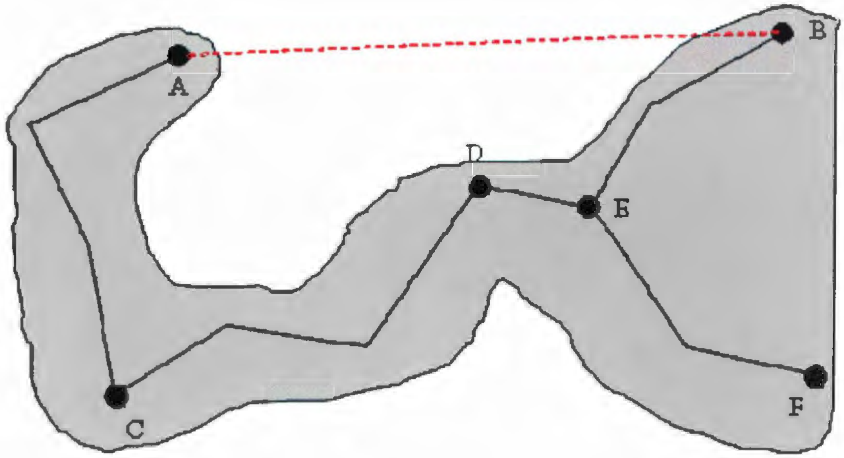


Figure 6: Euclidean distance (A-B) vs. network distance (A-C-D-E-B).

The isotropic Euclidean distance A-B does not reflect the transportation network path A-C-D-E-B necessary to travel between them. Neither are features A and D adjacent neighbours in a network context given the restrictions imposed by the model's environment. When such situations arise, spatial associations should be adjusted to maintain logical consistency with the reality of the underlying spatial model. The additional burden of this manual intervention process is often neglected when local knowledge of the underlying spatial context is incomplete. Current Geographical Information Systems (GIS) can easily produce spatial statistics but the simplicity of using these tools belies the complexity of the processes they utilize. These methods can be lax in their assumptions regarding the functional associations between features. Many spatial analysis techniques assume a Euclidean metric that facilitates Pythagorean methods to calculate relative position and distance. When spatial features are functionally located at greater distances, statistical results are exaggerated because the distance between features is minimized.

The plausibility of using a *spatial weights matrix* (W) to address this issue is intuitive but was determined to be of limited utility. The focus of the spatial weights matrix is adjacency and relative proximity between point or polygon features, determined by the linear Euclidean distances between points (or centroids) (de Smith et al., 2008). Such a matrix could compensate for distance discrepancies ***if*** distance is the only transformation variable under consideration. The counter-argument maintains that only a subset of spatial statistics is concerned solely with distances between features. More often, it is a particular attribute of the feature set that is the subject of investigation. Typically, spatial weights matrices adjust the relative weight of a particular feature attribute and not the spatial metric between features.

Distance between communities is rarely a matter of "as the crow flies". Even in the regular urban grid environment, methods such as rectilinear or Manhattan geometry are often employed to more accurately simulate real-world conditions. In this vein, the following analysis will endeavour to more realistically simulate the reality of transportation connectivity between irregularly distributed rural communities.

2.7 Spatial statistics

Many spatial statistics are global in nature meaning that the resulting indices are relevant to the entire selection of the spatial elements under examination. The more fundamental spatial statistic indices such as *standard distance*, *average nearest neighbour*, and (global) *spatial autocorrelation* indices are essentially determined by geometrically derived methods, wholly determined by the distance metric of the spatial model. More complex statistical measures such as *Local Indicators of Spatial Autocorrelation [LISA]*, (Anselin, 1995) and *Geographically Weighted Regression*

(GWR), (Fotheringham et al., 2002), conduct systematic localized examination across the analysis space and are thus intuitively more suited to real world contexts where discrete features are unique with respect to their relationship to their neighbours. The following sections will use this categorical distinction between global and local statistics to examine the relative effects of alternate distance metrics on both.

2.7.1 Fundamental global statistics indices

Spatial analysis of the distribution of point features must address methodological issues that arise due to the spatial dimension of their location. Often there are unobserved (or unacknowledged) environmental factors, both discrete and ubiquitous, that play determining roles in the relationship between spatial features. *Point Pattern Analysis* is a suite of investigative methodologies used to uncover underlying patterns within an analysis region. Generally, these patterns are determined by whether or not they vary from an assumed random spatial distribution. Spatial indices such as *standard distance* provide measures for the aggregate (global) dataset. Such measures reinforce the assumption of uniformity within the analysis area.

2.7.1.1 *Standard distance*

Bachi (1963) characterizes *standard distance* as "a simple, intuitive measure of the dispersion … obtained by averaging all distances between all possible pairs of cases" (p. 86). *Standard distance* measures the degree to which features are concentrated or dispersed around the points by way of: $\text{standard distance} = \frac{\sqrt{(\sum d^2)}}{n}$ where d is the distance to a given point (coordinates x, y) from the mean centre ($\bar{x}, \bar{y}$) and n is the total number of points. The *standard distance* can also consider a weighting attribute such as population that will result in a population-weighted distance. Thus, alternate distance metric should produce considerably different results particularly in weighted cases due to the increased leveraging effect of the weighting value.

2.7.1.2 *Average nearest neighbour*

The theoretical basis of the *average nearest neighbour* statistic is much more than the averaging of nearest neighbour distances as its label implies. It is actually, "a measure of the manner and degree to which the distribution of individuals in a population on a given area departs from that of a random distribution" (Clark & Evans, 1954, p. 445). The *average nearest neighbour* statistic calculates a nearest neighbour index based on the average distance from each feature to its nearest neighbouring feature:

$d(NN) = \sum_{i=1}^{N} [\frac{Min(d_{ij})}{N}]$ where $Min(d_{ij})$ is the distance between each point and its nearest neighbour, and N is the number of points in the distribution. The process is sensitive to area manipulation and boundary effects as acknowledged by Clark and Evans (1954) and reiterated by Pinder et al. (1979 p. 430-31).

2.7.2 Spatial autocorrelation indices

Spatial autocorrelation arose from the concept of temporal autocorrelation where time series data are not independent of their own historical values. Spatial autocorrelation extends the concept from the temporal to the spatial dimension.

The origins of spatial autocorrelation analysis were global in nature but recent trends are towards localized examination of spatial dependency that tacitly acknowledges Tobler's law regarding spatial association and focuses on the unique local nature of spatial relationships. Thus, spatial autocorrelation is relevant to both global and local analysis and may serve as a bridging template for localized consideration of other statistical measures. By shifting the focus of global statistics to more local levels, the unique circumstances of the local area's contribution to the overall spatial patterns can be quantified. Notwithstanding the potential benefits of the global approach, the fact remains that distance is the major contributing factor in autocorrelation analyses.

Distance between discrete features in spatial autocorrelation analyses is calculated as the minimum Euclidean distance between the features, essentially ignoring the actual connectivity through the intervening space, which is determined by the reality of the underlying environment.

> Spatial autocorrelation is one of the relatively small set of techniques which deals with both locational and attribute information … A pair of spatial features, for example two cities, may be similar or dissimilar in attributes, and their proximity will determine how similar they are in spatial location. In its broadest sense, spatial autocorrelation compares the two sets of similarities. If features which are similar in location also tend to be similar in attributes, then the pattern as a whole is said to show positive spatial autocorrelation. Conversely, negative spatial autocorrelation exists when features which are close together in space tend to be more dissimilar in attributes than features which are further apart. And finally the case of zero autocorrelation occurs when attributes are independent of location. (Goodchild, 1988, p.4)

In its simplest form, spatial autocorrelation is based on grid adjacency measures where 'neighbourliness' is determined by whether entities share a common border such as on the squares of a chessboard. The game board analogy used in the theoretical development of this concept, has resulted in retained vestigial chess terminology such as rook, king, and queens-case versions of adjacency contingent on whether conditions for inclusion are met by the movement rules of the chess pieces. It is from this grid-based theoretical development that distance calculations between features often utilize simple row and column differences to determine the geometric Pythagorean distance. Such an assumption minimizes the Euclidean space between elements and presumes uniformity across the intervening space between two points. While this assumption is reasonable for situations where the intervening space between points are indeed homogeneous, it is at odds with the reality of applications of spatial autocorrelation involving non-regular networks such as in the case of discrete community points along a road network. While communities may be spatially close, physical barriers and difficult topography can often complicate their connectivity and effectively lengthen the distance between them.

Haining (1990) remarks on spatial stationarity considerations between lattice and non-lattice case models, noting "the concept of stationarity is of questionable value for processes operating on irregularly scattered point sites or across continuous space … even if there exists an underlying continuous space process which is stationary …inter-area relationships for such irregular systems …have implications for how spatial dependency is represented" (p. 69). The representations are various graph structures (edges and weights) used to examine non-lattice proximity and interaction including Gabriel graphs, Delaunay triangulation, central place system of sites, and directed neighbourhood system (p. 72). Haining notes, 'the use of proximity criteria seems most appropriate where inter-site connections in terms of flow of information and material are not limited to special transportation networks … The use of interaction data, on the other hand, often reflects the presence of distinct transportation networks by which information and material flow between sites' (pp. 70-71).

Aniostropy, or directional bias is a fundamental consideration when determining distances between communities given that limited connectivity and a host of complicating travel factors combine to extend the functional association distance between communities.

The seminal work of Cliff and Ord (1973; 1981) extended the work of Geary (1954), Moran (1950), and others by providing a means of testing for departure from a random spatial pattern in attribute values (Getis, 1995, p. 247).

2.7.2.1 Various measures of spatial autocorrelation

Getis (2007) lists the following techniques for measuring *spatial autocorrelation*:

- Moran's *I* (a global covariance representation)
- Geary's *c* (a global differences representation)
- R (a cross product representation)
- Getis and Ord's *G* (a global multiplicative representation)
- Ripley's *K* (a cumulative pairs over distance representation)
- ρλ (autoregressive coefficients in various regression representations)
- Getis and Ord's G_i and G^*_i (local cluster representations)
- Anselin's I_i and c_i (local indicators of spatial association (LISA) statistics)
- Ord and Getis's *O* (a local representation taking into account global autocorrelation)
- 1/λ (the inverse of the semivariogram; i.e., the correlogram

(adapted from Getis, 2007, p. 494)

Such a listing suggests the extensive research interest that conducted on this topic and the potential importance it is anticipated to have for the future of spatial analysis within geography.

The prevalence of the global vs. local distinction is important. Within geostatistical parlance, global measures calculate all distance pairings simultaneously while local measures calculate distance pairings within a finite distance constraint. Getis and Ord (1996, p.262) caution that utilizing global (geostatistical) measures over large areas "contributes little meaning in such situations" while "any global statistic measure at a large scale of analysis provides little useful information" (ibid., p. 261-262). Local measures however, determine dependence in localized areas of the study area, typically up to predetermined distance from each feature. The focus of the current analysis on distance will examine whether the effects are similar for each category (global and local) given the fundamental role distance has on autocorrelation statistics.

3. Methodology

3.1 The effects of alternate distance metrics on spatial statistics measures

The following will develop and utilize a non-linear transformation technique to facilitate a more realistic determination of functional distances between network point features. The process develops a temporary alternative spatial distance metric that will allow spatial statistics to more pragmatically consider the associations between spatially distributed features. Euclidean proximity is adjusted to compensate for actual travel distance between communities resulting in a truer representation of neighbours within a connective network while avoiding potential topological errors from consideration in spatial statistic calculations.

3.2 Spatial context of the analysis area

The island of Newfoundland is the easternmost part of North America. It currently has some 750 identifiable communities. Most are adjacent to the rugged coastline and connected by some 7000 kilometres of road transportation infrastructure (Figure 7). The community distribution, clearly influenced by access to the shoreline, reflects the historical development prior to the construction of the road system.

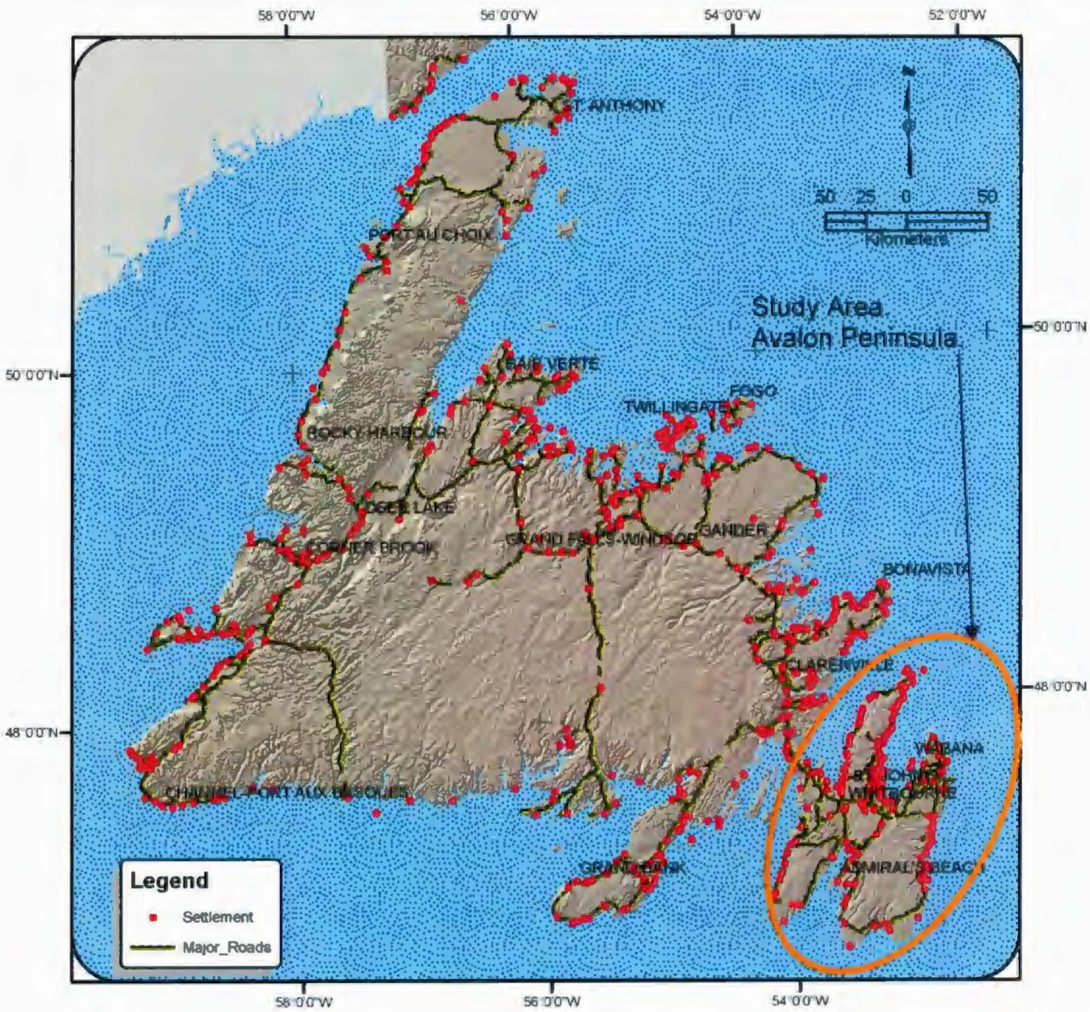


Figure 7: Insular Newfoundland. Settlement distribution and road connectivity.

Insular settlement is historically tied to prevailing transportation technologies. The birch canoe technology of the indigenous, semi-nomadic Beothuk allowed their seasonal exploitation of the riverine, estuarial, and littoral environments. The marine interests of the colonizing Europeans, in contrast, restricted their exploitation to the coastal and inshore areas thus constraining their settlement expansion along the coastline. Eventual attempts during the nineteenth and twentieth centuries to "open up" the interior by way of a rail line proved financially disastrous and eventually futile. The extensive road building projects that followed confederation with Canada was a concerted effort at modernization by facilitating the establishment of regional growth centres and discouraging the administrative challenges of servicing hundreds of scattered coastal settlements.

The present configuration of the road network, largely determined by the location of the pre-existing settlements, reflects the ubiquitous reliance on access to marine resources. Consequently, the irregular coastline contains numerous incidences where two settlements are separated by a short distance across a body of water, but which require a much longer distance to realize by road travel. The spatial relationship between such communities can be misrepresented in Euclidean models due to unacknowledged topographical and hydrological barriers. An additional source of distance error is the disregard of additional vertical travel distance imposed by often-rugged terrain. In Figure 8 the distance between communities A and B is a planimetric Euclidean distance of approximately two kilometres via the 1000 meter grid but the road network travel distance is approximately ten kilometres and likely more if elevation distance is considered. Road network connectivity between A and B can only be realized via the intermediate community C. Communities A and B are not topologically adjacent neighbours within a road network context.

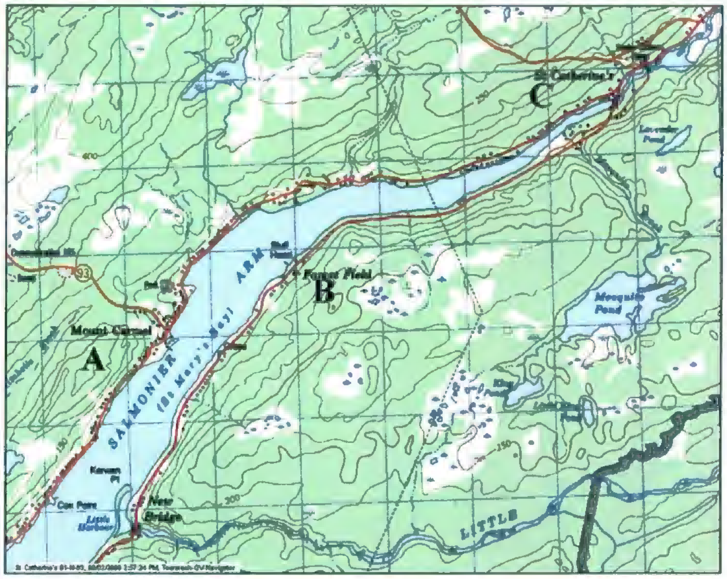


Figure 8: Example of Euclidean vs. road network distance.

Within living memory, the effect of changes in the prevailing transportation technology has altered community interaction dynamics. The supplanting of the traditional maritime connectivity by an overland option has fostered new connections in some cases, while diminishing them in others.

In order to compensate for distance issues inherent in the recently imposed connectivity framework, spatial relationships between communities should be transformed into a geometric space that translates the meandering road network into a vector representation. In effect, straighten the connecting road between two settlements into a single vector equivalent. The approach will resolve the intrinsic distance presumption issue of certain spatial statistics analysis. To reiterate, the proposed non-regular alternative metric is only relevant to single point of origin at any given time.

Transformations using alternate spatial metrics can result in changes in the topological relationships between features. The Euclidean distance between points is inherently monotonic given the intervening metric space is isotropic. When the spatial metric becomes non-isotropic (as is proposed here), non-monotonic circumstances will arise with respect to the original Euclidean arrangement but which will remain monotonic in the alternate transformation. From a Euclidean perspective B is between A and C in Figure 8. After the proposed transformation however, C will be between A and B, being restricted to the imposed road network. While both remain monotonic within their perspective metrics, each appears non-monotonic to the other. This is a transitory issue because the transformation is reversed after analysis calculations are conducted.

Figure 9 highlights the discrepancy between the two distance measures of Euclidean and vector equivalent under consideration. The local area of Figure 8 is expanded to a regional context to illustrate the increase in the Euclidean distance vectors from this area to all other insular communities. Figure 9b illustrates the vector equivalent of Figure 9a compensated for actual road distance.

The summary statistics show a near twofold increase in distance measures suggesting that the choice of metric will have similar considerable effect on other distance related statistics.

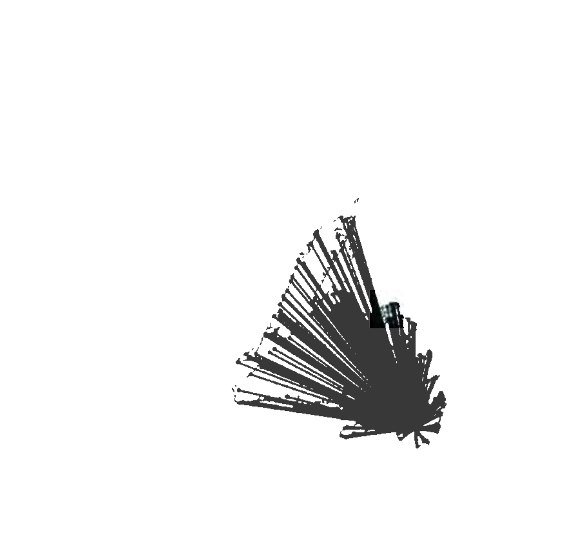
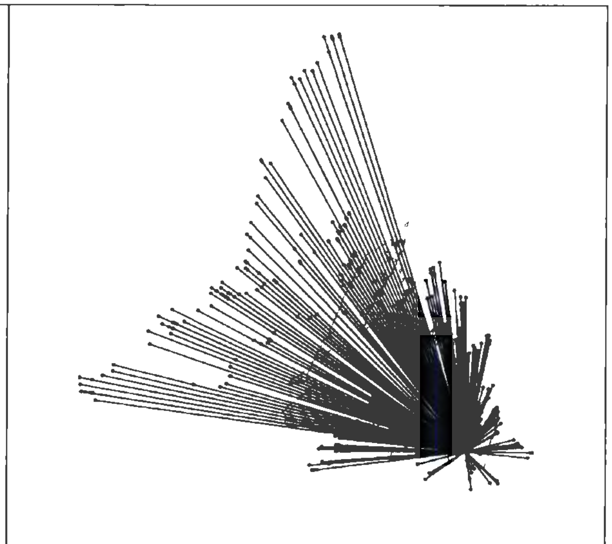
	
Total vector length: 79,095 km. Mean: 233 km. Standard Deviation: 14 km.	Total vector length: 139,550 km. Mean: 412 km. Standard Deviation.: 27 km.
Figure 9a: Sample community Euclidean distance vectors (Cartesian grid metric).	**Figure 9b: Sample community displacement vectors equivalent (road distance metric).**

Figure 9: Sample radial vector adjustment.

3.3 Procedure

The basis of this analysis is the transformation of a set of feature points within a Euclidean metric to an alternate, spatially adjusted metric that is determined by the actual travel distances between points. The process requires two versions of a point feature dataset (Figure 9), one in a standard Euclidean metric (Figure 9a) and a second transformed metric to straighten the connecting paths into an equivalent vector (Figure 9b). Spatial adjustments must be calculated separately for each point serving as the origin because each point is unique in the configuration of its distance relationship to all other points.

The initial task is to create a network dataset incorporating the point features with a topological road network. A network origin-to-destination model produces a line feature layer that calculates the travel distance between all point sets on the road network. Note that the graphical display of the results (Figure 9b) is a straight-line representation of the travel distance and not the actual network path between the points. The calculated travel distance value is stored in the feature attribute table; a favourable result because the geometry of the graphical representation is essentially the equivalent linear (Euclidean) distance between points that, when transformed, will make the distances between features equivalent to the network distance between them.

GIS maintains feature attribute information in tabular form. In addition to non-spatial information, this table can contain relevant spatial information as well. By editing the original point features attribute table, it is a relatively simple matter to determine the alternate point coordinates via simple Pythagorean and trigonometric methods.

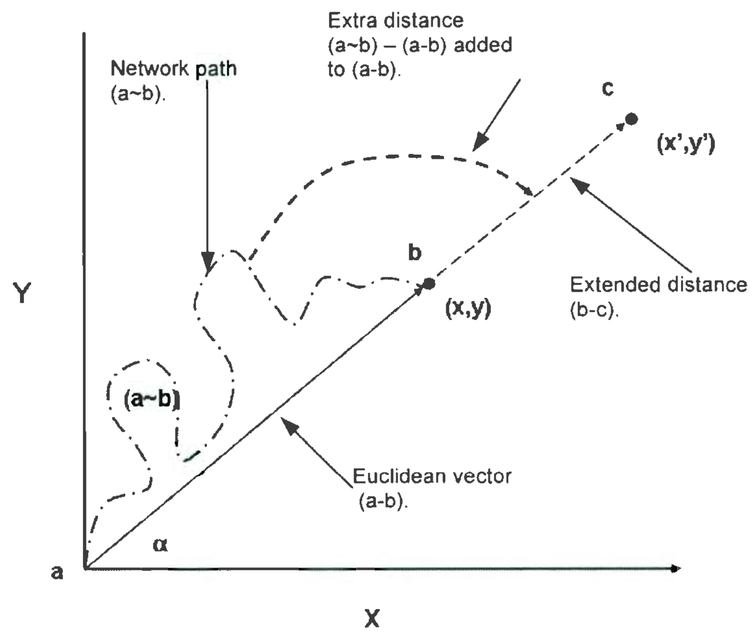


Figure 10: Transformation geometry.

The proposed point displacement procedure requires that the vector (a-b) between the existing Euclidean points *a* and *b* in Figure 10 be extended along the same vector by the additional distance inherent via the actual road path (a~b) and repositioned at point *c*. For each scenario neighbouring points are extended from each point of origin (a) to reflect the path of the actual road distance (a~b). The position of the displaced points is determined for each record by calculating new endpoint coordinate values in the feature attribute table. The existing coordinates (x,y) as well as the vector displacement (a-b), and direction (α) are geometrically determined for each feature in the Euclidean feature attribute table. The origin-destination network analysis results provide the alternate vector displacement (a~b) facilitating the calculation of the new point coordinates of the displaced destinations (x',y'). After the calculation, the feature attribute table contains value fields for the original point coordinates with the original distance and bearing, as well as the coordinates of the displaced points with the adjusted distances (a-c). Each record now has two sets of alternate coordinates (x,y and x',y') which will form the basis of a transformation matrix using the coordinates of the point features as control points.

Rubber-sheeting is an *n-order* transformation technique developed to merge spatial information of inaccurate or unknown map parameters into one of known parameters. Typical applications include the integration of maps produced under differing standards, especially historical maps of varying positional accuracy. The key to this technique is identifying and determining the coordinates of 'control points' in both map projections. The control points serve as anchors while allowing the relative metric space between features to vary by shrinking and stretching the metric space between them accordingly.

Rubber-sheeting requires a transformation matrix containing pairs of original and displaced coordinates. The coordinate value pairs for the study area (previously calculated as indicated) is simply exported as a tab-delimited text file for use in the spatial adjustment process. For each point of origin, the resulting table matrix provides two sets of coordinates for all other points; the first in the original space and the second for the alternate space adjusted for the road distance. Conceptually, this may be visualized by an elastic grid or web where the intersections are stretched relative to each other.

The transformation procedure employed the following methodology:

- A topological network matrix is constructed and utilized to calculate travel distances between all communities represented as points on the matrix.
- Directions - calculated from all points of origin to their respective destination communities.
- Vectors - calculated from all points of origin to all their destinations by direction (calculated direction) and magnitude (calculated distance).

Thus, for example the Euclidian distance for Admiral's Beach to St. Anthony is 507 kilometres at azimuth 342 degrees. The road distance straight-line equivalent maintains the same bearing (to maintain relative point arrangements) but substitutes the actual road distance of 1038 kilometres, in effect doubling the actual distance that would be normally be used in adjacency calculations.

Given that distance between point features is the primary focus, an alternate analysis path was briefly considered in the early stages of this research, the merits of which may be of passing interest here. While traversing a road network, direction may seem irrelevant considering it is predetermined by the topologic properties of the network itself. If direction is rendered immaterial, a conceptual model could be reduced to a one-dimensional representation along the positive number line. Figure 11 displays the concept of transferring the magnitude of the direction vectors to the positive number line.

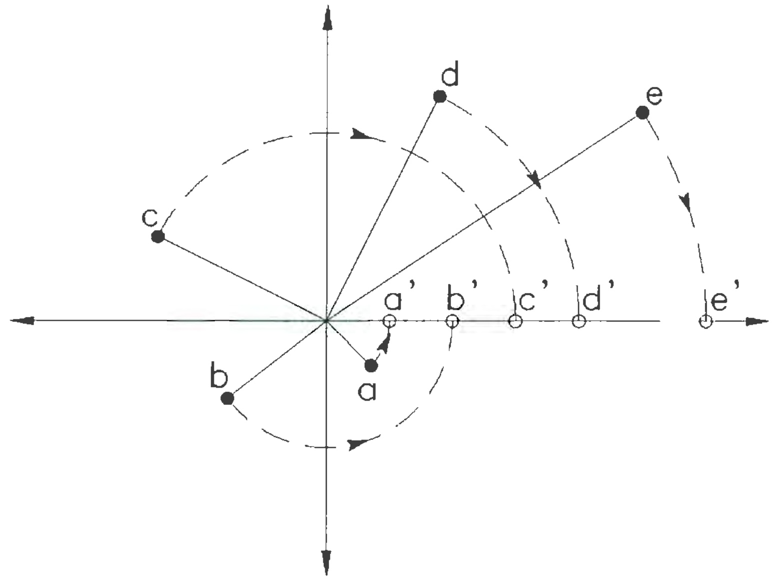


Figure 11: One dimensional distance transformation.

A one dimensional approach may be of limited value in visualizing distance to weighted centres but most spatial statistics consider direction and enclosing area during calculation, which restricts the utility of this one dimensional approach.

3.4 Transformation of the Cartesian grid

While the points representing the communities have been transformed, the underlying geography does not yet reflect this new relationship. An accompanying similar transformation of the underlying Cartesian grid is warranted to highlight the degree of the transformation. The result in a non-linear projection producing a novel visualization that reflects the varying effects of road network travel distance from a particular community to all other communities. Figure 12 for example shows the accelerating warping effect that actual road travel distance has on a Cartesian reference grid centered on Admiral's Beach. The warping effect is at a minimum near the origin but is progressively more distorted as distant accumulates due to the non-linear increase in distance to more remote points. Figure 13 includes the underlying geography.

Figure 12: Admiral's Beach origin. Distance-compensated Cartesian grid.

Figure 13: Admiral's Beach origin. Distance-compensated grid and landmass.

Such transformation measures are of limited utility over large geographical areas. Distortion effects increase as distance increases from the point of origin. As a result, spatial analyses over large geographic areas rarely produce meaningful results. A local spatial focus will increase the relevance of the transformation process.

3.5 Procedure classification

The proposed transformative visualization process is a synthesis of various techniques employed to spatially adjust a relative network relationship. As such, it is ambiguous as to how to categorize this geo-visualization process. Fundamentally, it is a spatial transformation from an orthogonal matrix grid (mesh) to a varying non-orthogonal matrix. From this basis, it suggests a similar tact in recent cartogram research where mesh transformation methods adjust the vertices of a polygon mesh while retaining relative topological integrity of adjacent areas (Keim et al., 2005; Andrieu et al. 2007). The focus of this paper however is on relative displacement of point features within a variable metric space, more suggestive of a linear or distance cartogram classification than a continuous area cartogram. Linear cartograms are pseudo-schematic with relaxed topological characteristics. The model proposed here actually enforces topological integrity as utilized by spatial statistics algorithms while maintaining accurate distances. Cartograms are typically qualitative visual aids and not vehicles for quantitative analysis.

In a multi-disciplinary approach, this proposed procedure borrows concepts from a variety of fields (mathematics, ecology, and physics) to address a geospatial issue. A broad description such as radial adjustment of vectors in Euclidean networks succinctly captures the procedure of compensating for alternate distances by straightening travel distances into vector equivalents.

4. Results and discussion

4.1 Comparative results

Section 4 compares the results of selected spatial statistics for a sampling of communities whose spatial proximity measures have been modified to reflect travel distances between them. Communities were chosen to represent features from a variety of relative locations (central and peripheral) within the analysis region. The choice of spatial statistics is somewhat arbitrary but those chosen were selected to highlight the effects on the analysis of certain population cohorts that are often of particular concern to social policy initiatives. Broad measures of population characteristics such as dependency ratios are often employed in the early stages of policy development to obtain a comparative measure of community viability from which to develop equitable policy strategies.

Regional policy analyses often utilize spatial frameworks imposed by *a priori* data collection processes that aggregate information into convenient collection units. Such restriction precludes any alternate differentiation based on the original un-aggregated information. This much-maligned circumstance is most evident in census enumerations where information is aggregated into seemingly random areas for confidentiality reasons. For population researchers, this is an example of the "tyranny of an artificially imposed and fixed set of census geographies" (Openshaw & Rao, 1995) that dilutes the quality of the original information. Still, administrative units are considered by policy personnel to be an equitable means of resource allocation despite the trouble they create for spatial analysts. Nevertheless, conveniently available information is frequently used without consideration of scale and aggregation manipulation issues such as those inherent in the Modifiable Area Unit Problem (Openshaw, 1984). Despite the quasi-arbitrary nature surrounding the establishment of administrative areas, there is often an underlying rationale influenced by factors other than efficient spatial analysis. Agencies that establish administrative areas do so to facilitate their own internal mandate. Third party agencies that attempt to leverage pre-existing administrative geographies often find them ill-suited to their own particular needs. Despite any shortcomings of imposing administrative areas within a region, they can foster the perception of a certain degree of homogeneity within the regions and can lead to the eventual adoption of the externally imposed defining characteristics. Communities exist within a myriad of these administrative structures, all of which function pragmatically, if not practically for their originally intended purposes. Unfortunately, efficacy often requires the utilization of pre-existing administrative structures.

The subsequent examples will examine the effects of using alternate distance metrics for communities in the Avalon Peninsula on the Island of Newfoundland. The Avalon area is perhaps the most consistently distinct regional administrative area of the Province given its peninsular geography. For current purposes, the Avalon is considered a discrete, self-contained administrative area. The particulars of the physical model are:

- Landmass: 9000 km^2
- Coastline: 1640 km.
- Road network: 4800 km.
- Number of community points: 81
- 5000 metre reference grid: 21,600 km^2

The following results focus on two categories of spatial statistics. The first category considers purely distance-based measures. These techniques are applied to the spatial distribution of the community point features only. The second category is those measures that consider non-spatial attributes with respect to distance. The principal attributes derive from aggregate census returns, primarily age cohorts, which are associated with the point representations of the communities. For demonstration purposes, the selected attributes are those that are pertinent to dependency ratios (i.e. relative youth, adult, and aged age cohorts). While it is possible to analyse the results for all possible community permutations, it is more effective to examine the effects of the proposed technique by individual case. The utility of this research is anticipated to be in determining whether policy decisions arrived at via geo-spatial analysis is equally valid when the analysis space is adjusted for distance effects. It can aid in resource allocation scenarios or in situations where distance decay is important. It also allows a community specific focus for examining the relative effects of distance on community interaction and relationships.

4.2 Overview of transformation results

Figure 14 displays an overview of the relative change in metric space using the foregoing methodology. Figure 14a shows the spatial arrangement of the communities in a normal Cartesian metric space. The transformation results (Figures 14b, 14c, and 14d) show the relative change in metric space from Cartesian to vector road distance for selected communities. The communities of origin were chosen to examine the effects of relative location within the spatial analysis area. The choices include a community near the centre of the spatial configuration, Whitbourne, (Figure 14b), and two communities near the extremities, Bay de Verde, (Figure 14c) and Wabana, (Figure 14d).

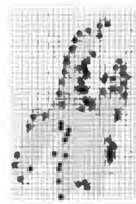


Figure 14a. Community distribution in normal Euclidean metric space.

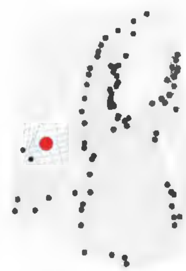


Figure 14b. Transformation result of central community (Whitbourne).

Figure 14c. Transformation result of peripheral community (Bay de Verde).

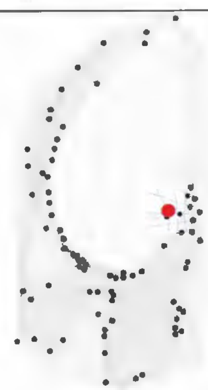


Figure 14d. Transformation result of peripheral community (Wabana).

Figure 14: Alternate distance metrics for selected communities (common scale).

The relative degree of transformation varies considerably across the analysis space. The central feature transformation Figure 14b displays relatively less overall shape distortion than the peripheral community points due to the extra distances required to connect to all other communities. Peripheral points accumulate greater overall connection distances due to their relative location with respect to other communities. Centralized locations result in a radial distortion pattern while peripheral locations display more linear distortion trends. These general observations are manifest in the distance statistics as well.

The following baseline case (in a normal Euclidean space) and three alternate scenarios will form the basis of a comparative spatial analysis into the effects of utilizing the proposed alternate transformation procedure. Each of the transformations is subjected to the noted analytical techniques followed by a general commentary of the results.

4.3 Selected spatial statistics

4.3.1 Distance only statistics

4.3.1.1 *Standard distance*

Figure 15 shows the relative increase in *standard distance* calculations for communities that result from the transformation process. The *standard distance* of the points in a normal projection space is 48.25 km., a value that increases by between 40 and 100 percent (68 to102 km.) over the various point sets (Table 2, Appendix). These higher distance values contain implications for analyses that utilize the *standard distance* measure. Distances are always greater in transformed metrics under the present scenarios due to the positive cumulative distance increases introduced by the transformation.

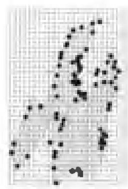
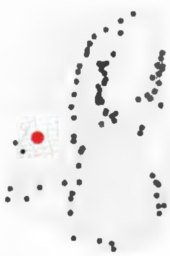
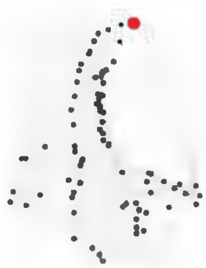
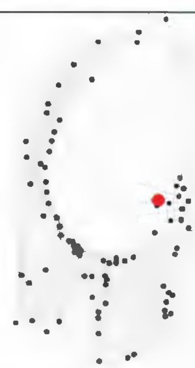
Community of origin	Standard distance results
 **Figure 15a: Normal projection (baseline).**	Standard distance: 48.25 km. Increase from Cartesian: 0.00 km. Increase percent: 0.0%
 **Figure 15b: Whitbourne transformation.**	Standard distance: 71.30 km. Increase from Cartesian: 23.05 km. Increase percent: 47.8%
 **Figure 15c: Bay de Verde transformation.**	Standard distance: 75.03 km. Increase from Cartesian: 26.78 km. Increase percent: 55.5%
 **Figure 15d: Wabana transformation.**	Standard distance: 97.82 km. Increase from Cartesian: 49.57 km. Increase percent: 102.7%

Figure 15: Standard distance results (common scale).

4.3.1.2 *Average nearest neighbour*

Figure 16 displays the *average nearest neighbour* results for the selected transformed communities. Unsurprisingly, observed and expected mean distances increase significantly due to the extra distances imposed between the community points.

Interestingly, the clustering significance confidence interval increases from 95% to 99% for all the transformed results despite the fact the points are moving apart relative to each other. The arrangement of the points in normal Euclidean space is relatively compact and denser at that scale. By expanding the distance metric, the clustering is accentuated as indicated by the change in z-score and the clustering significance from <5% to <1% for all scenarios. The dynamics of this process may be conceptualized by considering the points to be on an elastic surface such as a balloon. Initially the points appear either random or clustered, but inflating (stretching) the surface can result in a reversal of the initial observation. The effect is further enhanced if the deformation is not directionally uniform as is the case under study. See Figure 17 for an example of enhanced clustering in the alternate metric.

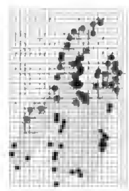
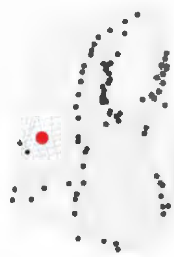
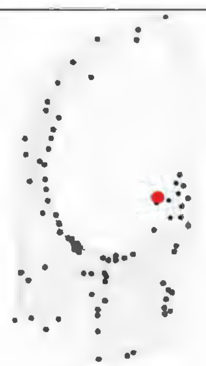
Community of origin	Average nearest neighbour results
 **Figure 16a: Normal projection**	Observed Mean Distance: 5.85 Expected Mean Distance: 6.65 Nearest Neighbour Ratio: 0.8800 Z Score: -2.065 p-value: 0.0389 Significance:<5% clustered Percent change from baseline: 0%
 **Figure 16b: Whitbourne transformation.**	Observed Mean Distance: 7.45 Expected Mean Distance: 10.14 Nearest Neighbour Ratio: 0.7349 Z Score: -4.563 p-value: 0.000005 Significance: <1% clustered Percent change from baseline: 27.35%
 **Figure 16c: Bay de Verde transformation.**	Observed Mean Distance: 7.85 Expected Mean Distance: 10.98 Nearest Neighbour Ratio: 0.7150 Z Score: -4.905 p-value: 0.000001 Significance: <1% clustered Percent change from baseline: 34.19%
 **Figure 16d: Wabana transformation.**	Observed Mean Distance: 10.49 Expected Mean Distance: 13.23 Nearest Neighbour Ratio: 0.7916 Z Score: -3.587 p-value:0.000334 Significance: <1% clustered Percent change from baseline: 79.32%

Figure 16: *Average nearest neighbour* results (common scale).

Figure 17a: 95% probability of clustering in normal space

Figure 17b: 99% probability of clustering in transformed space.

Figure 17: ***Average nearest neighbour*** **exaggerated clustering, Placentia (common scale).**

4.3.2 Distance and count statistics

Spatial analysis is more often interested in the distribution of attribute characteristic than merely the spatial location of features. Determining patterns in the spatial distribution of quantifiable characteristics is the principal concern of spatial analysis. Clustering analysis is a suite of fundamental exploratory spatial data analysis techniques that evaluate the spatial distribution of attributes and determines whether the patterns are statistically significant from a random distribution. The primary autocorrelation techniques examine attribute values while considering their spatial relationships to the other features (Wulder n.d.). Since all measures use Euclidean distances, it is reasonable to expect that all autocorrelation measures will overstate the degree of adjacency. The remainder of section 4.3.2 will compare various autocorrelation results for the selected communities in both Cartesian and transformed metric spaces using 2001 census population counts as the attribute of interest. The first row of each table shows the results of the unadjusted spatial metric.

4.3.2.1 Spatial autocorrelation (Moran's I)

Figure 18 shows the results for the global Moran Statistic I for the 2001 population for the selected communities. As expected, there is a significant increase in the search threshold for the transformed arrangements. Under the normal space scenario there is a less than 1% expectation that the distribution of the population values is due to chance. Similar results are evident for the central location (Figure 18b Whitbourne) and one of the peripheral communities (Figure 18d Wabana) which both show higher Moran values and z-scores. The exception is the peripheral community of Bay de Verde (Figure 18c), which was expected to have similar values to the other peripheral community of Wabana (Figure 18d). On reflection, the choice of inverse distance weighting in the 'conceptualization of spatial relationship' option may explain this result. The two peripheral communities differ greatly with respect to their relationship to the major population centre of St. John's. Wabana is adjacent while Bay de Verde is furthest away.

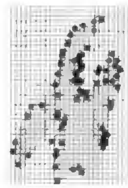
Community of origin	Global Moran's I results
 **Figure 18a: Normal projection**	Significance: <1% clustered Moran's Index: 0.1124 Expected Index: -0.0125 Variance: 0.0015 Z Score: 3.133 p-value: 0.0017 search threshold: 14.8km
 **Figure 18b: Whitbourne transformation.**	Significance: <1% clustered Moran's Index: 0.1479 Expected Index: -0.0125 Variance: 0.0013 Z Score: 4.383 p-value: 0.00001 search threshold: 26.05km
 **Figure 18c: Bay de Verde transformation.**	Significance: somewhat clustered – random Moran's Index: 0.0477 Expected Index: -0.0125 Variance: 0.0027 Z Score: 1.148 p-value: 0.2508 search threshold: 26.70km
 **Figure 18d: Wabana transformation.**	Significance: <1% clustered Moran's Index: 0.1263 Expected Index: -0.0125 Variance: 0.0011 Z Score: 4.129 p-value: 0.00003 search threshold: 33.68km

Figure 18: Moran's I results, 2001 Population, (common scale).

4.3.2.2 High/Low clustering (Getis and Ord General G)

Figure 19 show the results for the Getis and Ord General G spatial autocorrelation statistic. Under the normal metric, (Figure 19a) the results suggest a 95% probability of high value clustering. The probability increases to 99% for the results of the representative central feature case (Figure 19b, Whitbourne). The two extremity cases (Bay de Verde, Figure 19c, and Wabana, Figure 19d) offer conflicting results similar to the Moran index. Considering both are extreme case scenarios, similar results were expected. It is speculated that this result is due to the multiplicative nature of this statistic. The Wabana case (Figure 19d) is near the highest concentration of larger communities while the Bay de Verde case (Figure 19c) is furthest from the major population centre of St. John's. The choice of 'inverse distance' as the most appropriate option for the 'conceptualization of the spatial relationship' within the GIS module leverages closer communities while diminishing those further away. It is proposed that this resulted in the case near the high population area (Figure 19d, Wabana) being more affected than was the case at the most extreme distance from it (Figure 19c, Bay de Verde). In any event, the results confirm that changing the distance metric between a set of points can produce various results for the General G statistic.

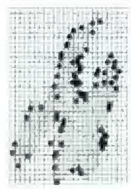
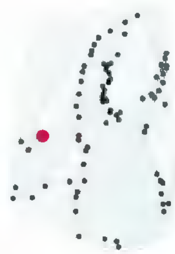
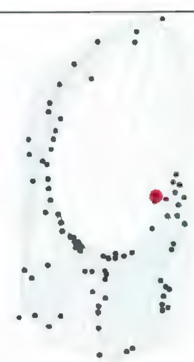
Community of origin	Getis and Ord G results
 **Figure 19a: Normal projection**	Significance: <5% high clustered Observed General G: 0.00003 Expected General G: 0.00001 Variance: 0 Z Score: 2.359 p-value: 0.0183 search threshold: 14.84 km.
 **Figure 19b: Whitbourne transformation.**	Significance: <1% high clustering Observed General G: 0.00003 Expected General G: 0.000009 Variance: 0 Z Score: 3.548 p-value: 0.0003 search threshold: 26.05 km.
 **Figure 19c: Bay de Verde transformation.**	Significance: no apparent clustering Observed General G: 0.00001 Expected General G: 0.000009 Variance: 0 Z Score: 0.5061 p-value: 0.6127 search threshold: 26.70 km.
 **Figure 19d: Wabana transformation.**	Significance: <1% high clustering Observed General G: 0.00002 Expected General G: 0.000006 Variance: 0 Z Score: 4.372 p-value: 0.00001 search threshold: 33.68 km.

Figure 19: Getis and Ord General G results, 2001 population (common scale).

4.3.3 Distance and ratio value statistics

Policy analysis often attempts to determine whether localized areas display significant variance for a derived attribute value. Comparative analysis requires standardized values to counteract the effect of larger centres overwhelming smaller ones. Comparison between populations of varying size is commonly achieved by calculating proportional values rather than absolute counts.

Changing demographics often requires a change in focus for service delivery depending on the demographic mix of a population. Growth in the relative youth component of a population is a harbinger of developing educational and recreational needs while an increasing aged population may suggest a pending demand for support structures that cater to this group. Clusters of communities that display similar pending social needs will rank higher in the delivery of regionalized services under fiscal cost benefit considerations.

The urban rural divide has always been a contentious issue in social policy. The expectation of minimum levels of social services regardless of population density requires a normalizing approach to compensate for population weighting effects. Comparison of populations that differ greatly in size is accomplished by examining relative proportions (or incidence) of population characteristics to highlight relevant areas and by subsequently taking into account the absolute numbers when considering policy implementation. Large population centres will inevitably overwhelm smaller settlements when count data is utilized but both are on a more equitable footing when ratio values are used.

Dependency ratios are useful statistics to determine the overall status of a population. In general, it is the ratio of the dependent component of a population to the (economic) supporting component. Closer examinations may use either the youth or aged components as a sub grouping. Temporal analysis of these ratios can reveal trends with respect to anticipated change in dependent groups.

The remainder of this section will examine the dependency ratios for selected communities within the study area. The analysis will focus on whether clustering results are markedly changed by using the alternate distance metric procedure previously outlined.

4.3.3.1 Cluster and outlier analysis - Anselin's Local Moran's I. Total dependency ratio 2006.

Figures 20 to 22 are the rendering results[1] of the cluster and outlier analysis - Local Moran's I, (LISA) of the total dependency ratio results from the 2006 census. Each figure displays the normal Euclidean results on the left with a comparison to each of the alternate community transformation result on the right.

---

[1] The following results uses a Z-Score rendering technique designed for use with Cluster and Outlier as well as Hot Spot analyses. The Z renderer creates a layer file with z scores rendered in the following manner:

- Z scores below –2 standard deviations are rendered dark blue.
- Z scores between –2 and –1 standard deviations are cyan.
- Z scores between –1 and +1 standard deviations are neutral.
- Z scores between 1 and 2 standard deviations are orange.
- Z scores above 2 standard deviations are bright red.

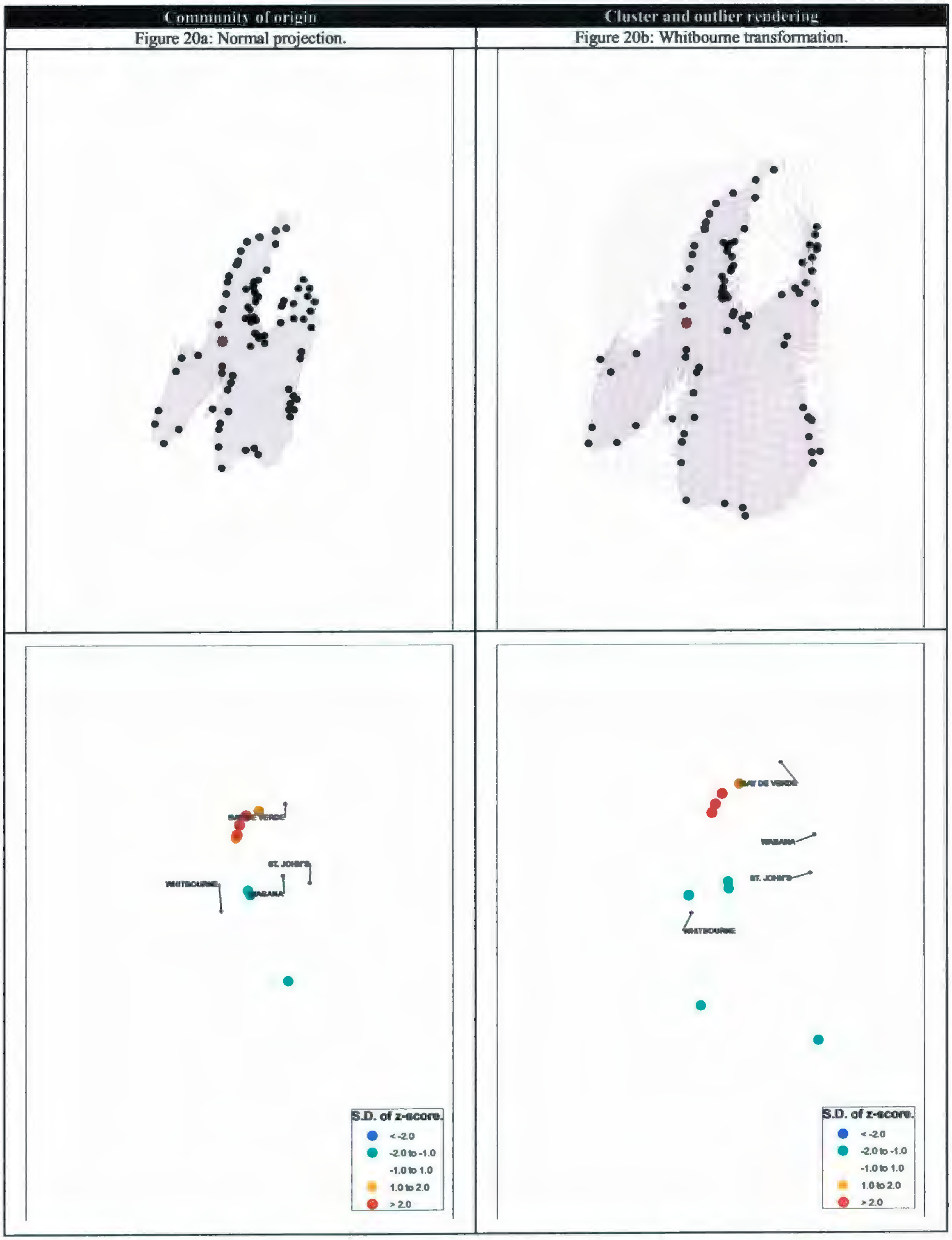


Figure 20: Cluster and outlier analysis – Anselin's Local Moran's I. Total dependency ratio 2006. Whitbourne transformation (common scale).

Community of origin	Cluster and outlier rendering
Figure 21a: Normal projection.	Figure 21b: Bay de Verde transformation.

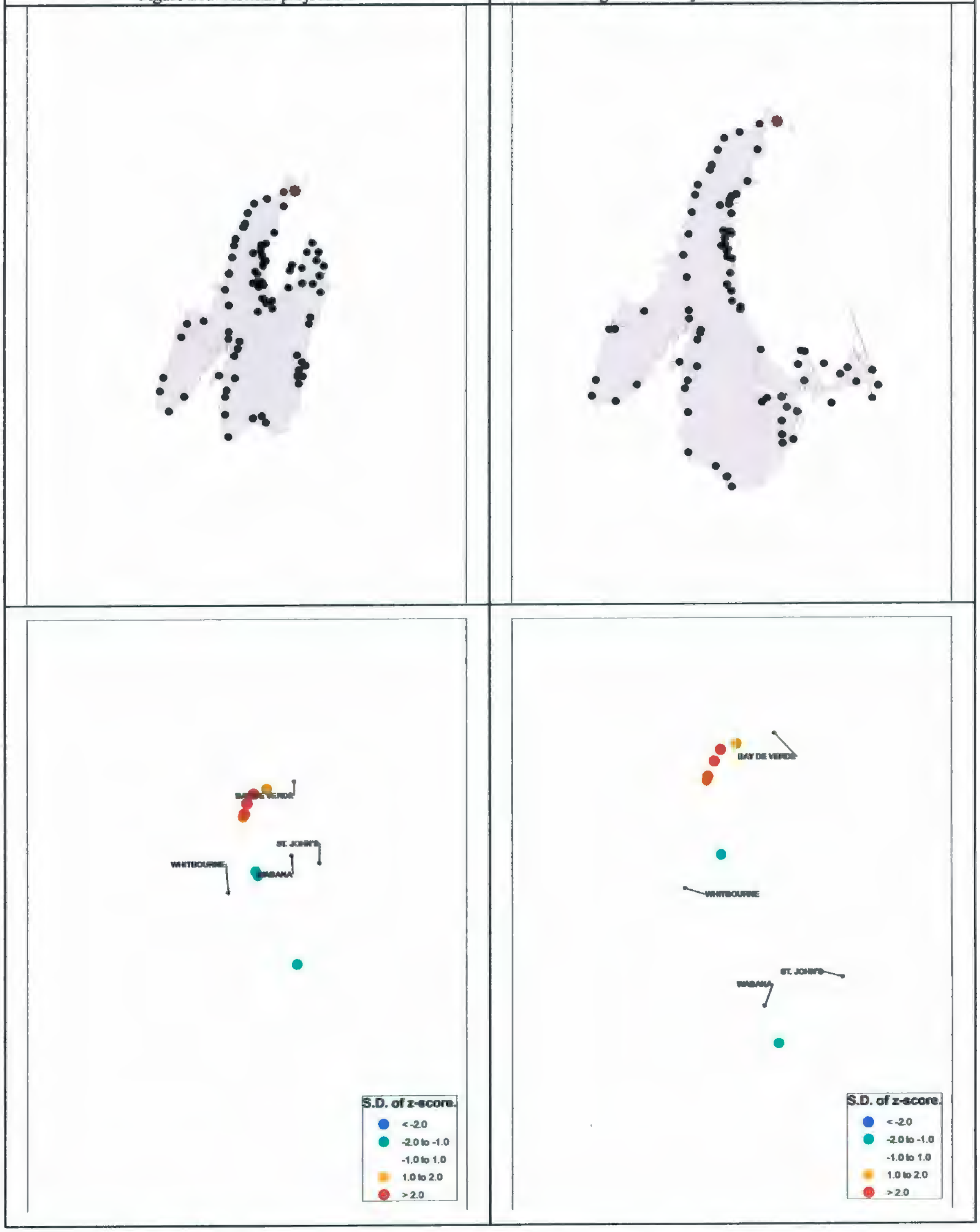


Figure 21: Cluster and outlier analysis – Anselin's Local Moran's I. Total dependency ratio 2006. Bay de Verde transformation (common scale).

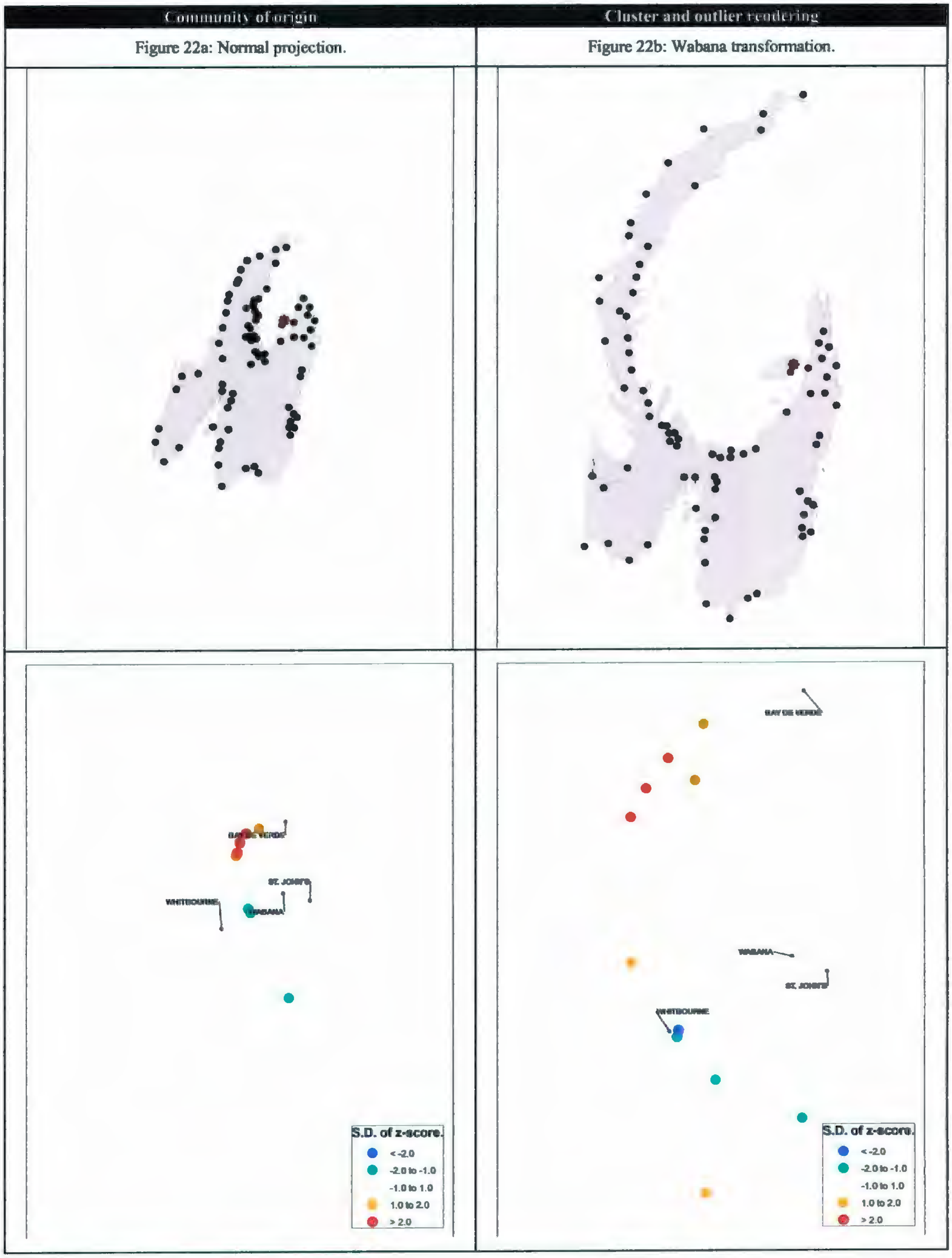


Figure 22: Cluster and outlier analysis – Anselin's Local Moran's I. Total dependency ratio 2006. Wabana transformation (common scale).

Overall, the alternate metric results are indistinguishable from those calculated under the Euclidean metric. The communities for which Local Moran's I is significant ($z>1.96$) are consistently determined under all alternate metrics. The result is somewhat surprising in light of the observations for the global Moran statistic that suggest clustering is accentuated due to the relative expansion of the intervening metric space. The outcome suggests that Local Moran statistic is less sensitive to changes in the alternate metric.

4.3.3.2 Getis and Ord Gi* (Hot spot) analysis rendering. Total dependency ratio, 2006.

Figures 23 to 25 show the results of the Getis and Ord Gi* statistic for the total dependency ratio from the 2006 census. The Getis and Ord Gi* measure identifies clusters of high and low values unlike the LISA statistic which does not differentiate in this manner.

Figure 23: Getis and Ord Gi* analysis rendering. Total dependency ratios 2006. Whitbourne transformation (common scale).

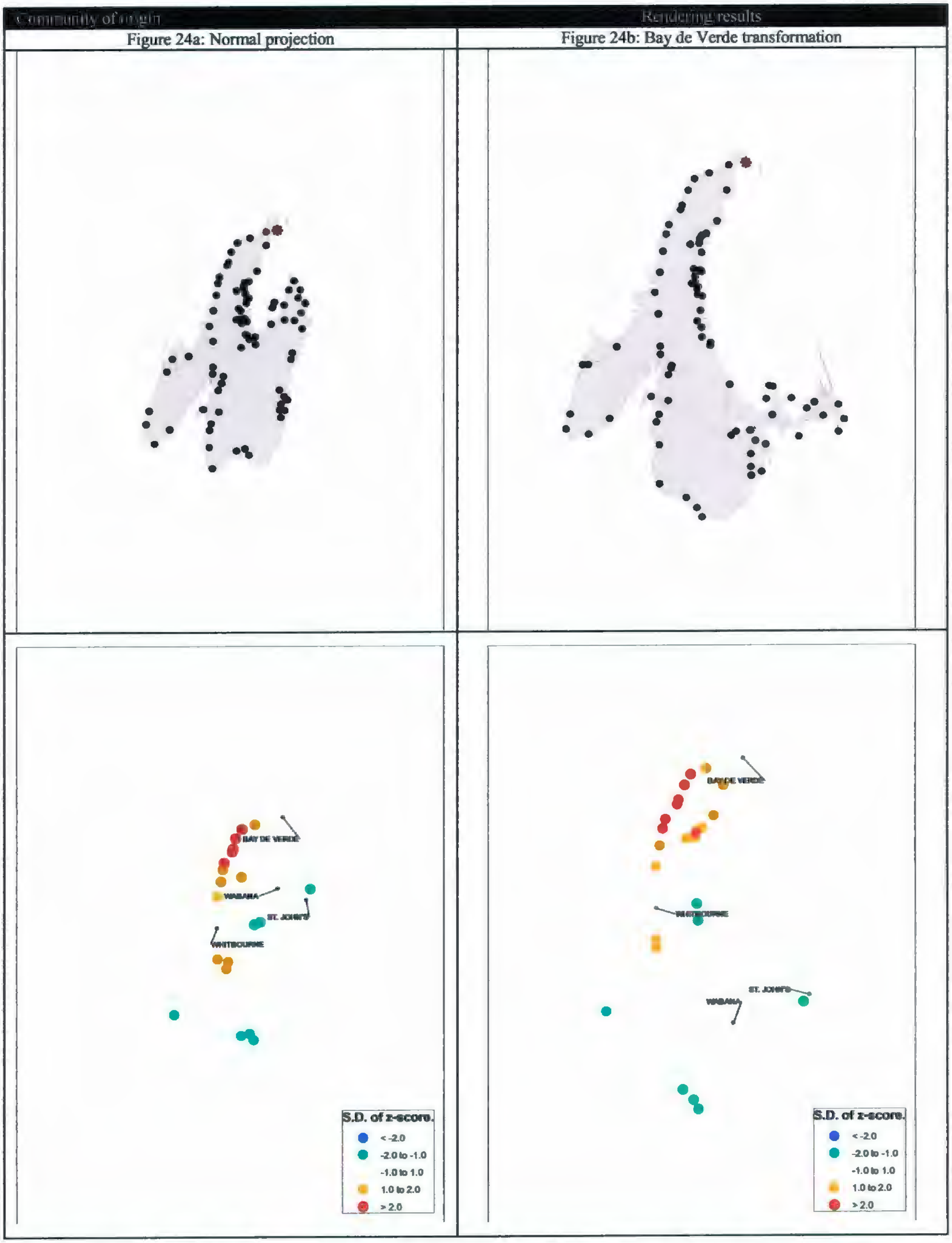


Figure 24: Getis and Ord Gi* analysis rendering. Total dependency ratios 2006. Bay de Verde transformation (common scale).

Community of origin	Rendering results
Figure 25a: Normal projection.	Figure 25b: Wabana transformation.

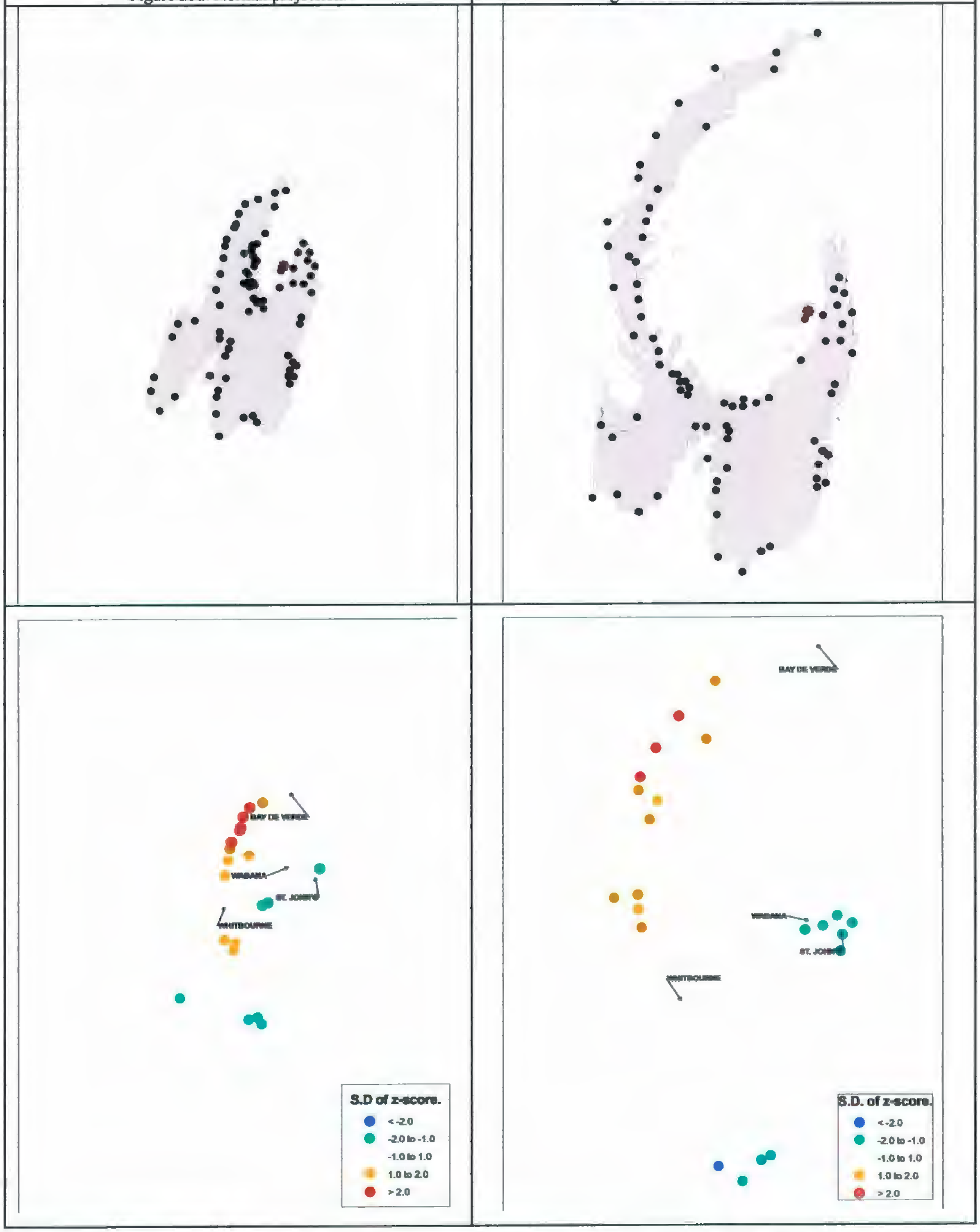


Figure 25: Getis and Ord Gi* analysis rendering. Total dependency ratios 2006. Wabana transformation (common scale).

The most noticeable difference between these two measures is the marked increase in the extent of clustering under the Gi* statistic near the point of origin of each transformation. Although these two statistical measures are not comparable, both respond similarly to changes in distance metrics.

4.3.4 Comparative results: Dependency ratios

Analysis using alternate visualizations is cumbersome and potentially confusing. Section 4.3.4 will address this issue by avoiding the alternate metric display but substituting the alternate values in the normal metric visualization. The result will present an alternate version of analysis results calculated within the alternate metric. The examples will examine dependency ratios under the two metric spaces but will only visualize them in normal space, essentially highlighting the differences that result from the two metrics.

Dependency ratios are useful indices that allow comparisons of populations with respect to the relative proportions of general age groups. Broadly, it is measure of the relative ratio between the economically dependent proportions (i.e. the youth and aged) of a population to the economically productive (working age) component. Often the dependent components are determined separately when focusing on either the youth or aged age groups.

4.3.4.1 Getis and Ord Gi* results

The following example will offer a descriptive comparison of Getis and Ord Gi* analysis for total, youth, and aged dependency ratios within the study area. The alternate distance metric is centred on the provincial capital city of St. John's, i.e. distances are adjusted using St. John's as the central point. The capital city location was chosen to examine a scenario where the major population centre served as the origin of the transformation procedure rather than examining a relative position within a network (i.e. central, peripheral) considered previously.

4.3.4.1.1 Aged dependency ratio

Figure 26 shows the rendered results of the aged dependency ratio under the normal metric case. A comparison with the alternative metric space values in Figure 27 show a strengthening of the significant low value clustering in the St. John's area. Several peripheral but marginally significant communities (those within an index values range between 1 and 2) under the normal metric become highly significant (greater than 2) under the alternate metric. At the same time, several normally insignificant values (0 to 1) near the periphery of the significant clusters (west of Wabana) become marginally significant (1 to 2). These results suggest that the extent of the clustering increases due to utilization of the alternate distance metric.

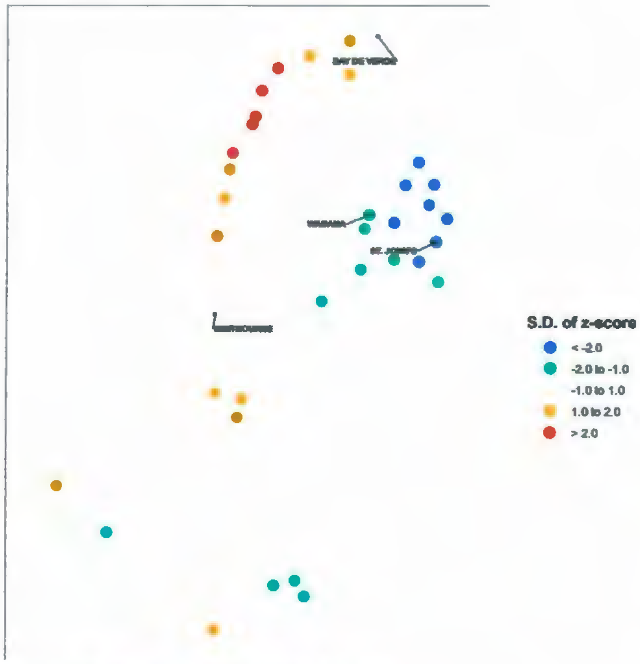


Figure 26: Getis and Ord Gi*. Aged dependency ratio. Normal metric values.

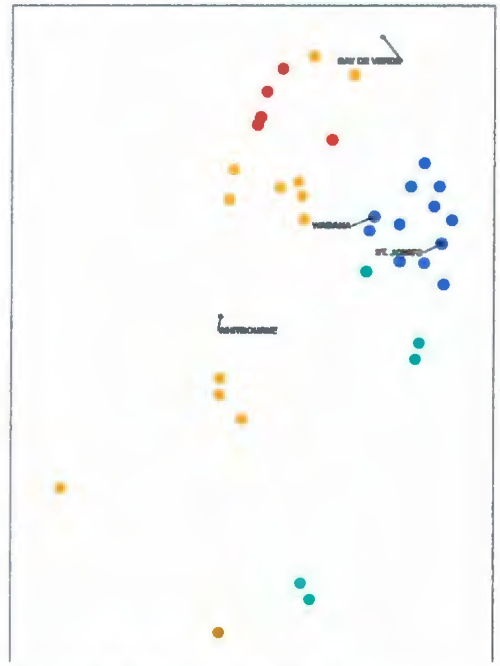


Figure 27: Getis and Ord Gi*. Aged dependency ratio. Alternate metric values.

4.3.4.1.2 Youth dependency ratio

Youth dependency results (Figure 28 and Figure 29) show similar but more localized results. The extension of the high value clustering (near St. John's) reinforces the low value clustering of the previous aged dependency ratio results since both proportions of the population are inversely related. There is some expansion of localized significance where low significant neighbours become marginally significant after adjustment.

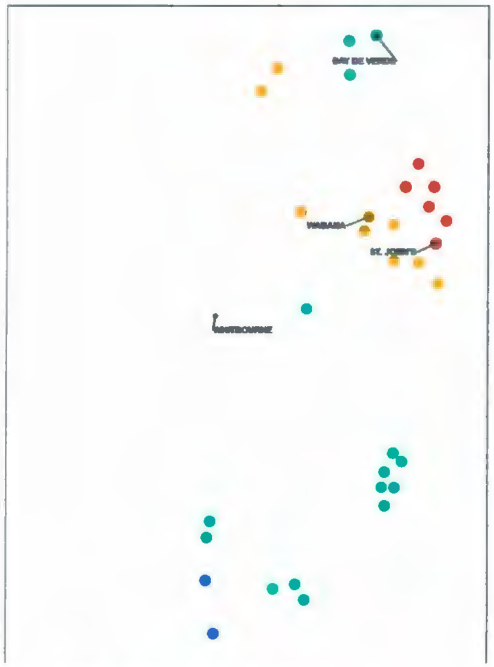


Figure 28: Getis and Ord Gi*. Youth dependency ratio. Normal metric values.

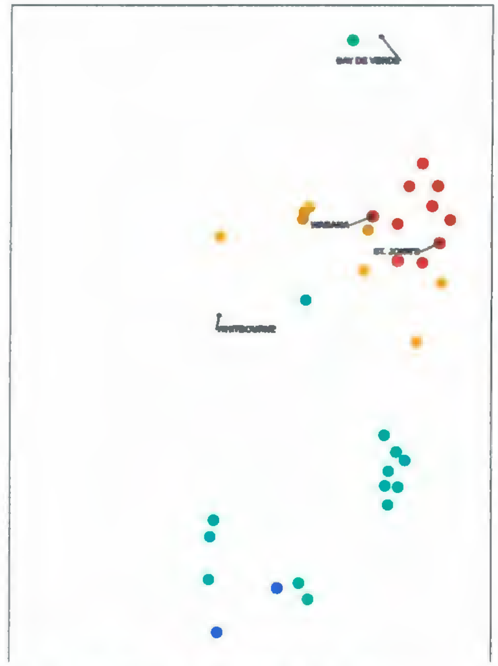


Figure 29: Getis andOrd Gi*. Youth dependency ratio. Alternate metric values.

4.3.4.1.3 Total dependency ratio

Total dependency ratio combines both dependent groups and as such is a more general comparative measure of economic dependency in the population. The results (Figure 30 and Figure 31) show most of the change is in the upper centre (west of Wabana) where there is a noticeable increase in marginally significant high value neighbours. The lack of any high significance in the upper right (near St. John's) where youth and aged were formerly high, suggests a balancing effect between the youth and aged components. The overall dependency ratio is less variable than the individual dependent components at this scale.

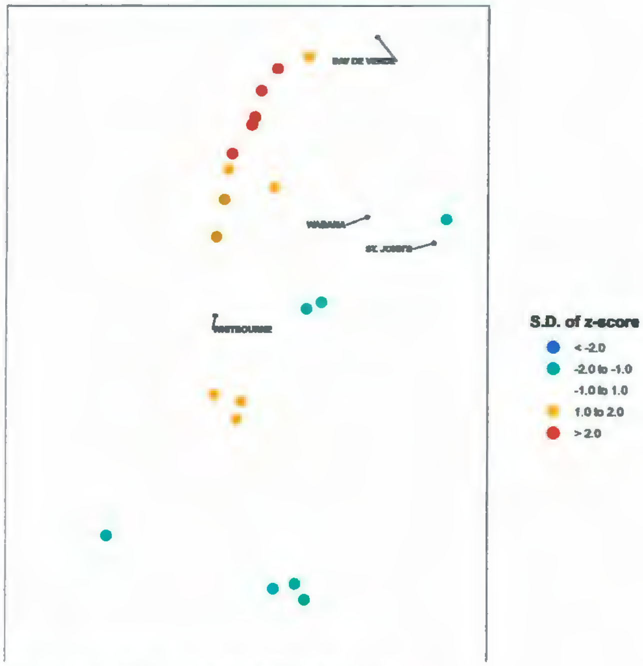


Figure 30: Getis and Ord Gi*. Total dependency ratio. Normal metric values.

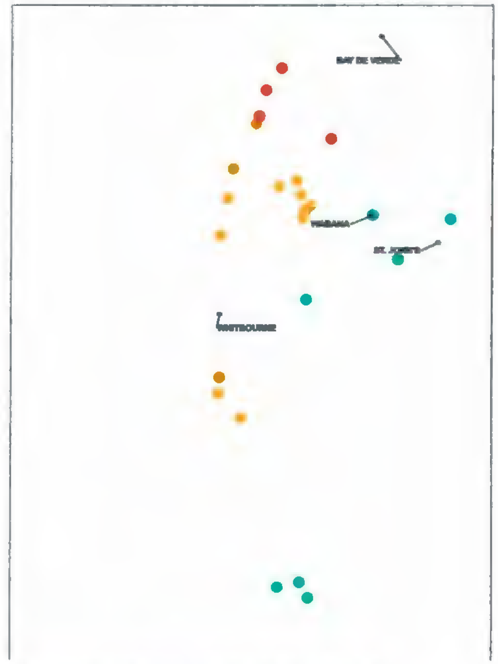


Figure 31: Getis and Ord Gi*. Total dependency ratio. Alternate metric values.

4.3.5 Results conclusion

The consideration of alternate distance metrics can enhance spatial patterns that may be understated using a normal distance metric. The previous example has shown that an alternate distance metric can augment the degree of clustering observed for Getis and Ord Gi* analysis. It is not intended to conclude that the alternate metric is in any way superior to established methodology or that it provides better results. What is suggested is that under circumstances where interaction between communities is determined by road network connectivity, the incorporation of an alternate distance metric can mitigate concerns regarding straight line vs. actual travel distances.

5. Conclusions

Spatial statistics have enabled geographers to begin modeling the processes that underlie spatial phenomenon. Many of the statistical techniques developed are relevant to field phenomenon (geostatistics) or to discrete population phenomenon (point pattern analysis) where the intervening distance metric between observations is uniformly isotropic. The connectivity among features is often restricted to pathways that mitigate the effort required to traverse a given space. When the assumption of underlying homogeneity is invalid, measures should be taken to compensate for the distance measures used in spatial statistics analysis. The geometry of the connectivity network often determines the distance between features, not the minimum linear distance.

Harvey Miller's forum address to the Annals of the Association of American Geographers (2004) notes that shortest path relations between all pairings of features in geo-space are the minimum-cost routes for physical movement or virtual interaction between objects. Furthermore:

> In most of the geographic and related literature, nearness is typically defined based on the straight-line segment connecting two locations, that is, the Euclidean distance for the location pair. This is only one possibility. There are an infinite number of shortest-path relations that obey the metric space conditions of symmetry, non-negativity, and triangular inequality … If we are willing to relax these metric requirements so that only the triangle inequality condition holds, the resulting space is a quasi-metric. This can still support measurement and spatial analysis …Nearness is a central organizing principle of geo-space, but it is not required to be a function of Euclidean, metric, or even an empty space. There is a wide range of analytical and computational techniques for representing and analyzing these spaces and no reason in principle why they should not be part of a standard GIS toolkit. (Miller, 2004, para.1)

The Abel Prize is the Nobel Prize equivalent for mathematics, (there being no Nobel Prize in mathematics). Since the inaugural Abel Prize in 2003, five of the eight laureate citations include reference to advances in spatial research areas, in particular, manifolds, topology, and geometry. Mikhail Gromov was awarded the 2009 Prize for, among other things, "the notion of distance which he has introduced in completely surprising situations and exploited with elegance" (Hansen, 2010). Renewed mathematical interest in distance and alternate metrics is an area that holds potential interest for spatial statistics as well.

The research objective was to evaluate whether the substitution of a variable road distance metric into spatial statistics calculations would render results that are more meaningful. The foregoing research has shown that the radial transformation of point sets to reflect actual linear distances between features can have varying effects on spatial statistic calculations, depending on the type of data under analysis. Spatial statistics that are primarily distance based, such as *standard distance* and *average nearest neighbour*, show greater impacts due to the proposed transformation technique than other statistical subgroups such as the various global spatial autocorrelation measures, primarily due to the normalizing tendency of the latter. Certain autocorrelation measure such as Getis and Ord Gi* were shown to enhance the extent and significance of spatial clustering.

The foregoing has noted a lacuna of investigation into the analysis of alternate distance metrics within the field of geography and has outlined a pragmatic procedure to compensate for alternate distances when calculating spatial statistics. The procedure utilizes a variable distance metric that enforces the utilization of network distance measures as well as precludes the consideration of topological inconsistencies that result from conducting spatial analyses without regard for underlying topological determinants.

While the initial heuristic intent was an interim transformation to calculate certain distance-dependent spatial statistics, visual aspects of the resulting transformations may also offer insight into perception of travel distance. The approach may also hold potential for other alternate distance metrics such as interactive social distance where physical proximity may be irrelevant.

References

Andrieu, D., Kaiser C., Ourednik A., and Lévy J. (2007). Advanced cartogram construction using a constraint based framework. In: *Geocomputation 2007*, National University of Ireland, Maynooth, 3-5 September.

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

Bachelard, G. (1969). *The poetics of space (originally La poétique de l'espace)*. Trans. Maria Jolas. Beacon Press, Boston.

Bachi, R. (1963). Standard distance measures and related methods for spatial analysis. *Papers in Regional Science, 10*(1), 83-132.

Barnes, T. (2004). A paper related to everything but more related to local things, *Annals of the Association of American Geographers, 94*(2), 278-283.

Boguñá, M., Krioukov, D. and Claffy, K. C. (2008). Navigability of complex networks. *Nature Physics, 5,* 74-80. doi:10.1038/nphys1130.

Borg, I. and Groenen, P. (1997). *Modern multidimensional scaling. Theory and applications.* New York: Springer.

Bozkaya, T. and Ozsoyoglu, M. (1997). Distancc-based indexing for high-dimcnsional metric spaces. Proceedings of the 1997 ACM SIGMOD International Conference on Management of data, Tucson. ACM SIGMOD Record, 26(2), 357-368. doi:10.1145/253260.253345]

Clark, P. J. and Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology, 35*(4), 445-453.

Cliff, A.D. and Ord, J.K. (1973). *Spatial autocorrelation*. London: Pion.

Cliff, A. D. and Ord, J. K. (1981). *Spatial processes: models and applications*. London: Pion.

Curriero, F. (2006). On the use of non-Euclidean distance measures in geostatistics. *Mathematical Geology, 38*(8), 907-926.

de Smith, M. J., Goodchild, M. F. and Longley, P. A.(2008). *Geospatial analysis: A comprehensive guide to principles, techniques and software tools*, Smith, M. J., Goodchild, M. F., Longley, P. A. (Ed.) Leicester, UK: Troubador.

DeCarlo, D. and Dimitris M. (1998). Shape evolution with structural and topological changes using blending. In: *IEEE PAMI, 20*(11), 1186-1205.

Egenhofer, M. and Mark D. M. (1995). Naïve geography, in: Frank A.U. and Kuhn W. (Eds.) *Spatial information theory: A theoretical basis for GIS*, pp. 1-15. Springer

Fotheringham, A. S., Brunsdon C. and Charlton M. (2002). *Geographically weighted regression - the analysis of spatially varying relationships*. John Wiley & Sons Inc.: New York.

Ganio, L. M., Torgersen, C. E. and Gresswell R. E. (2005). A geostatistical approach for describing spatial pattern in stream networks. *Frontiers in Ecology and the Environment, 3*(3), 138-144.

Gardner, B., Sullivan, P. J. and Lembo, A. J. (2003). Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. *Canadian Journal of Fisheries and Aquatic Sciences, 60*, 344–351.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician, 5*(3), 115–145. doi:10.2307/2986645.

Getis, A. (1995). Classics in human geography revisited. *Progress in Human Geography, 19*(2), 245-249.

Getis, A. (1999). "Spatial statistics" Chapter 16. In P. Longley, M. Goodchild, D. Maguire, and D. Rhind, (Eds.).*GIS: Principles, techniques, management, and applications.* John Wiley and Sons, Inc., New York.

Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics, 37*(4), 491-496.

Getis, A. and Ord, K. (1996). Local spatial statistics: an overview. In: P. Longley and M. Batty (Eds.) *Spatial analysis: Modelling in a GIS environment*. John Wiley and Sons.

Goodchild, M. (1988). Spatial autocorrelation. In: *Concepts and techniques in modern geography (CATMOG)* Number 47. Norwich: Geobooks.

Haining, R. P. (1990). *Spatial data analysis in the social and environmental sciences*. Cambridge University Press.

Hansen, V. L. (2010). *The mathematics of Mikhal Gromov*. Retrieved from http://www.abelprisen.no/en/prisvinnere/2009/hansen/index.html.

Harvey, D. (1990). Between space and time: Reflections on the geographical imagination. *Annals of the Association of American Geographers, 80*(3), 418–434.

Jiang, B. (1998). A space syntax approach to spatial cognition in urban environments. In Hirtle, S., MacEachren, A. M., (Eds.), *Cognitive Models of Dynamic Geographic Phenomena and Their Representations,* pp. 48-58. Pittsburgh, PA.

Keim, D. A., Panse, C. and North, S. C. (2005). Medial-axis-based cartograms. *IEEE Computer Graphics and Applications, 25*(3), 60-68.

Klinkenberg B., (2009). Image available from: http://www.geog.ubc.ca/courses/klink/g470/notes/coordinates/coordinates.html. Accessed August 18, 2009.

Lee, J. M., (2000). Introduction to topological manifolds. *Graduate texts in mathematics*: Vol. 202. Springer.

Lefebvre, H. (1974). *La production de l'espace (The production of space)*. Paris: Anthropos.

Löw, M. (2006). The Social Construction of Space and Gender. *European Journal of Women's Studies, 13*(2), 119-133. doi: 10.1177/1350506806062751

Lyon, S.W., Seibert, J., Lembo, A. J., Steenhuis, T. S., and Walter, M. T. (2008). Incorporating landscape characteristics in a distance metric for interpolating between observations of stream water chemistry. *Hydrology and Earth System Sciences, 12*, 1229-1239.

Massey, D. (1991). A global sense of place. *Marxism Today,* June: p. 24-29.

Matheron, G. (1965). *Les variables regionalisees et leur estimation*, Masson, Paris.

Miller, H. J. (2004). "Tobler's first law and spatial analysis," *Annals of the Association of American Geographers, 94*(2), 284-289.

Moran, P. A. P. (1950). "Notes on continuous stochastic phenomena," *Biometrika, 37*, 17–33.

Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and techniques in modern geography (CATMOG)* Number 38. Norwich: Geobooks

Openshaw, S. and Rao, L. (1995). Algorithms for reengineering 1991 census geography. *Environment and Planning A, 27*(3), 425-446.

Pederson, S. P. (2005). Surfaces demonstrating Euclidean, elliptical, and hyperbolic geometry (Euclidian_and_non_euclidian_geometry.png). Retrieved from http://upload.wikimedia.org/wikipedia/commons/5/5e/Euclidian_and_non_euclidian_geometry.png. Accessed November 25, 2009.

Pidwirny, M. (2006). Concepts of time and space in physical geography. In *Fundamentals of Physical Geography*, 2nd Edition. Retrieved from http://www.physicalgeography.net/fundamentals/3c.html

Pinder, D., Shimada, I., and David G. (1979). The nearest-neighbor statistic: Archaeological application and new developments. *American Antiquity, 44*(3), 430-445.

Puu, T. and Beckmann M. J. (1999). Continuous space modelling In: R.W. Hall (Ed.) *Handbook of transportation science* (pp. 269-310). Kluwer Academic, Norwell, Mass.

Qi, C., Shu, H., and Xu, A. (2006). "Formal properties of cognitive distance in geographical space," icat, pp.408-412, 16th International Conference on Artificial Reality and Telexistence--Workshops (ICAT'06).

Rowland, T. (2010). Manifold. Retrieved from from: http://mathworld.wolfram.com/Manifold.html

Shu, H., Edwards, G., and Qi, C. (2001) Cognitive distance. In: Sharatchandra Pankanti and Runsheng Wang (Eds.) *Object detection, classification, and tracking technologies. Proceedings of SPIE. Volume 4554* (pp. 290-296).

Sivagnanasothy, V. (2010). Evidence based decision making in government: Strategies to enhance the use of data in development planning, management, monitoring, and evaluation: A Sri Lankan perspective. Keynote address to: *Using data in evidence-based decision-making - Launch of Dev Info version 6.0 in Asia and the Pacific"*, February 16, 2010. Retrieved from http://www.unescap.org/Stat/di6launch/Key-note-Sri-Lanka.pdf

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234–40.

Tobler, W. (2001). *The world is shriveling as it shrinks.* Presentation at Texas A&M University, 26 October.

Wulder, M. (n.d.). *Spatial autocorrelation*. [Online]. Retrieved from: http://cfs.nrcan.gc.ca/subsite/wulder/spatial. Accessed August 18, 2009.

Appendices

Appendix 1 – Selected standard distance results.

Table 2: Standard distances.

Community Of Origin	Standard distance (km.)	Increase (km.)[2]	pct
St. Vincent's-St. Stephen's-Peter's River	68.25	20.00	41.5%
Point Lance	68.46	20.21	41.9%
St. Shott's	69.05	20.80	43.1%
Colinet	70.48	22.23	46.1%
Gaskiers-Point La Haye	70.72	22.47	46.6%
Witless Bay	71.01	22.76	47.2%
Branch	71.23	22.98	47.6%
Whitbourne	71.30	23.05	47.8%
St. Mary's	71.72	23.47	48.6%
Bay Bulls	71.77	23.52	48.7%
Holyrood	71.94	23.69	49.1%
Mount Carmel-Mitchell's Brook-St. Catherines	71.99	23.74	49.2%
St. Joseph's	72.27	24.02	49.8%
Riverhead, St. Mary's Bay	73.26	25.01	51.8%
Old Perlican	73.83	25.58	53.0%
St. Bride's	74.08	25.83	53.5%
Hant's Harbour	74.55	26.30	54.5%
South River	74.76	26.51	54.9%
Colliers	74.94	26.69	55.3%
Winterton	74.94	26.69	55.3%
Bay De Verde	75.03	26.78	55.5%
Clarke's Beach	75.14	26.89	55.7%
Harbour Main-Chapel Cove-Lakeview	75.27	27.02	56.0%
Heart's Content	75.63	27.38	56.7%
Conception Harbour	76.00	27.75	57.5%
Whiteway	76.07	27.82	57.7%
Avondale	76.15	27.90	57.8%
Spaniard's Bay	76.27	28.02	58.1%
New Perlican	76.27	28.02	58.1%

[2] Standard distance for unadjusted points is 48.25 km.

Community Of Origin	Standard distance (km.)	Increase (km.)^2	pct
Heart's Delight-Islington	76.76	28.51	59.1%
Heart's Desire	76.89	28.64	59.4%
Bay Roberts	77.02	28.77	59.6%
Carbonear	77.02	28.77	59.6%
North River	77.06	28.81	59.7%
St. John's	77.29	29.04	60.2%
Victoria	77.36	29.11	60.3%
Mount Pearl	77.76	29.51	61.2%
Cupids	77.94	29.69	61.5%
Admiral's Beach	78.29	30.04	62.3%
Logy Bay-Middle Cove-Outer Cove	78.31	30.06	62.3%
Placentia	78.37	30.12	62.4%
Brigus	78.44	30.19	62.6%
Salmon Cove	78.49	30.24	62.7%
Harbour Grace	79.09	30.84	63.9%
Petty Harbour-Maddox Cove	79.19	30.94	64.1%
Trepassey	79.27	31.02	64.3%
Small Point-Broad Cove-Blackhead-Adam's Cove	79.30	31.05	64.3%
Bishop's Cove	79.54	31.29	64.8%
Portugal Cove South	79.80	31.55	65.4%
Upper Island Cove	80.02	31.77	65.8%
Biscay Bay	80.46	32.21	66.8%
Bryant's Cove	80.72	32.47	67.3%
Paradise, Conception Bay	81.37	33.12	68.6%
Ferryland	81.69	33.44	69.3%
Torbay	82.55	34.30	71.1%
Cape Broyle	83.28	35.03	72.6%
Flatrock	83.42	35.17	72.9%
Conception Bay South	83.52	35.27	73.1%
Fox Harbour, Placentia Bay	84.10	35.85	74.3%
Aquaforte	84.78	36.53	75.7%
Renews-Cappahayden	85.67	37.42	77.6%
Portugal Cove-St. Philips	86.44	38.19	79.1%
Fermeuse	86.54	38.29	79.4%
Port Kirwan	86.80	38.55	79.9%
Pouch Cove	87.11	38.86	80.5%
Bauline	90.33	42.08	87.2%
Wabana	97.82	49.57	102.7%
Ccs 1r	102.45	54.20	112.3%

Appendix 2 – Selected average nearest neighbour results.

Table 3: Average nearest neighbour.

	Observed mean distance:	Expected mean distance:	Nearest neighbour ratio:	z-score:	p-value:	significance	pct change from baseline
Baseline–Euclidean	5.85	6.65	0.880055	-2.06517	0.038907	<5% clustered	0%
Admiral's Beach	8.68	12.87	0.674734	-5.600311	0	<1% clustered	62.30%
Bay Roberts	7.85	10.03	0.782445	-3.745789	0.00018	<1% clustered	59.60%
Branch	8.76	11.70	0.748402	-4.331923	0.000015	<1% clustered	47.60%
CCSIR	10.65	14.06	0.757404	-4.176922	0.00003	<1% clustered	112.30%
Flatrock	8.82	12.90	0.683892	-5.442637	0	<1% clustered	72.90%
Old Perlican	7.61	10.86	0.70076	-5.152204	0	<1% clustered	53.00%
Placentia	7.80	11.72	0.665364	-5.761638	0	<1% clustered	62.40%
Portugal Cove South	8.96	12.71	0.705401	-5.072303	0	<1% clustered	65.40%
St. Vincent's	8.06	11.64	0.692402	-5.296117	0	<1% clustered	41.50%
Whitbourne	7.45	10.14	0.734976	-4.56309	0.000005	<1% clustered	47.80%