Essays on Space-Time Interaction Tests

by

Nicholas Malizia

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2013 by the
Graduate Supervisory Committee:

Luc Anselin, Chair
Alan Murray
Sergio Rey

ARIZONA STATE UNIVERSITY
May 2013
Researchers across a variety of fields are often interested in determining if data are of a random nature or if they exhibit patterning which may be the result of some alternative and potentially more interesting process. This dissertation explores a family of statistical methods, i.e. space-time interaction tests, designed to detect structure within three-dimensional event data. These tests, widely employed in the fields of spatial epidemiology, criminology, ecology and beyond, are used to identify synergistic interaction across the spatial and temporal dimensions of a series of events. Exploration is needed to better understand these methods and determine how their results may be affected by data quality problems commonly encountered in their implementation; specifically, how inaccuracy and/or uncertainty in the input data analyzed by the methods may impact subsequent results. Additionally, known shortcomings of the methods must be ameliorated.

The contributions of this dissertation are twofold: it develops a more complete understanding of how input data quality problems impact the results of a number of global and local tests of space-time interaction and it formulates an improved version of one global test which accounts for the previously identified problem of population shift bias. A series of simulation experiments reveal the global tests of space-time interaction explored here to be dramatically affected by the aforementioned deficiencies in the quality of the input data. It is shown that in some cases, a conservative degree of these common data problems can completely obscure evidence of space-time interaction and in others create it where it does not exist. Conversely, a local metric of space-time interaction examined here demonstrates a surprising robustness in the face of these same deficiencies. This local metric is revealed to be only minimally affected by the inaccuracies and incompleteness introduced in these experiments. Finally, enhancements to one of the global tests are presented which solve the problem of population shift bias associated with the test and better contextualize and visualize its results, thereby enhancing its utility for practitioners.
To Mom and Dad
ACKNOWLEDGEMENTS

This dissertation could not have been written without the support of a number of people and organizations. Most immediately thanks are due to my committee members, Luc Anselin, Alan Murray and Serge Rey, for helping guide me through this process. Their instruction and insights over the past four years have helped forge my identity as an academic. Although not formally on my committee, Julia Koschinsky played a large role in the development of this document as well. Her assistance and guidance during my tenure at the GeoDa Center was invaluable. Her inspiring attitude and enthusiasm were contagious. Not enough good things can be said about Julia and the investment of time she has made in me and countless other students. I also wish to thank Billie Turner and Gil Pontius for their mentoring and friendship over the years. The inspiration for me to pursue a doctoral degree came from my time as student of Gil and Billie at Clark University. Since my first year at Clark they have invested liberally in my personal and professional development and opened up countless doors for me. I will be forever grateful to them.

I am extremely grateful and appreciative of the support provided to me by the National Science Foundation and their Graduate Research Fellowship Program. This award afforded me complete intellectual freedom during my tenure at ASU and allowed me to focus on the research presented herein. Additionally, the work was supported by Award No. 2009-SQ-B9-K101 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect those of the Department of Justice. My time at Arizona State was also supported by a fellowship from the GeoDa Center for Geospatial Analysis and Computation.
My time in Tempe would not have been the same without the company of my friends and colleagues at the School of Geographical Sciences, especially Daniel Arribas-Bel, Pedro Amaral, Joanna Merson, Shai Kaplan, Kelly Turner, Dave Folch, Phil Stephens, Steph Kleinschmidt, John Connors, Yin Liu and Ellen Schwaller. There was never a shortage of good conversation or delicious food in their company.

My largest debt of gratitude, however, is owed to my family. My parents, Darlene and David, were unwavering in their support of my efforts. They provided me with constant inspiration. Additionally, they instilled in me the strength of character needed to see this process through. Mom and Dad, could you have imagined that the little boy crying while doing his homework at the kitchen table all those years ago would have gone on to write a doctoral dissertation? Your love and nurturing made it possible. My brothers, Geoff and Evan, have kept me young at heart and made sure my sense of humor made it through this process intact. My friend Viktor has brought me up from many a low point (and shared many a high point) and generally been the very best kind of friend a person could ever hope to have. And Elizabeth, my dear, my love, your support and affection have kept me sane during the culmination of this process. I’m so happy we found each other. I’m looking forward to enjoying the fruits of this labor with you.
TABLE OF CONTENTS

LIST OF TABLES .......................................................... viii
LIST OF FIGURES ....................................................... x
CHAPTER ................................................................. 1
1 INTRODUCTION ......................................................... 1
2 THE EFFECT OF DATA INACCURACY ON TESTS OF SPATIO-TEMPORAL
   INTERACTION ......................................................... 9
   2.1 Introduction ................................................... 9
   2.2 Background ................................................... 11
      Locational Inaccuracy ......................................... 13
      Temporal Inaccuracy .......................................... 17
   2.3 Tests of Space-Time Interaction ............................. 18
      Knox Test ....................................................... 18
      Mantel Test ................................................... 19
      Jacquez Test .................................................. 21
   2.4 Methods ......................................................... 22
      Experiment 1 ................................................... 23
      Experiment 2 ................................................... 26
      Inaccuracy Parameters ....................................... 27
   2.5 Results .......................................................... 30
      Experiment 1 ................................................... 30
      Experiment 2 ................................................... 34
   2.6 Discussion and Conclusion .................................. 36
3 INACCURACY, UNCERTAINTY AND THE SPACE-TIME PERMUTATION
   SCAN STATISTIC ...................................................... 40
   3.1 Introduction .................................................. 40
## Chapter 3.2 Background
- Space-Time Permutation Scan Statistic ............................................... 42
- Data Quality Deficiencies ................................................................. 44
  - Spatial Inaccuracies ........................................................................ 45
  - Temporal Inaccuracies ..................................................................... 47

## Chapter 3.3 Methods
- Synthetic Data .................................................................................. 49
- Mesa, AZ Burglary Data ....................................................................... 53

## Chapter 3.4 Results
- Synthetic Data .................................................................................. 55
- Empirical Data ................................................................................... 63

## Chapter 3.5 Discussion and Conclusion ............................................. 67

## Chapter 4 Enhancing the Jacquez K Nearest Neighbor Test for Space-Time Interaction
- Introduction ....................................................................................... 71
- Background on Space-Time Interaction Tests .................................... 73
- Jacquez Test ...................................................................................... 77
  - Calculation ...................................................................................... 77
  - Combined Test ................................................................................ 79
  - Issues and Proposed Solutions ....................................................... 80
- Data .................................................................................................. 82
- Methods ............................................................................................ 85
  - Establishing the Spatial and Temporal Scale for k ......................... 85
    - Example Using Burkitt’s Lymphoma Data ...................................... 87
    - Comparison with Knox Results .................................................... 88
  - Visualization .................................................................................... 90
  - Space-Time Cube ............................................................................. 90
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison with Knox Visuals</td>
<td>93</td>
</tr>
<tr>
<td>Incorporating Population Shift</td>
<td>94</td>
</tr>
<tr>
<td>4.6 Discussion and Conclusions</td>
<td>99</td>
</tr>
<tr>
<td>5 CONCLUSION</td>
<td>103</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>108</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>2.1</td>
<td>Rejection rates associated with perturbed versions of original patterns exhibiting differing degrees of interaction: high (2.1a), medium (2.1b) and low (2.1c).</td>
</tr>
<tr>
<td>2.2</td>
<td>Values of Kendall’s coefficient of concordance ($W$) using perturbations as judges and tests as objects for the results found in Table 2.1. Symbols correspond to significance at the following $\alpha$ levels: ** = 0.01, * = 0.05</td>
</tr>
<tr>
<td>2.3</td>
<td>Values of Kendall’s coefficient of concordance ($W$) using tests as judges and perturbations as objects for the results found in Table 2.1. Symbols correspond to significance at the following $\alpha$ levels: ** = 0.01, * = 0.05</td>
</tr>
<tr>
<td>2.4</td>
<td>Values of Kendall’s coefficient of concordance ($W$) using tests as judges and perturbations as objects for the results found in Table 2.1. Symbols correspond to significance at the following $\alpha$ levels: ** = 0.01, * = 0.05</td>
</tr>
<tr>
<td>2.5</td>
<td>Observed rejection rates where $\alpha = 0.01$ (2.5a), $\alpha = 0.05$ (2.5b) and $\alpha = 0.10$ (2.5c) for perturbed versions of completely random patterns. Asterisks indicate that the proportion of rejections is within two standard deviations of the nominal $\alpha$ used to establish the rejection rate (i.e. 0.01, 0.05 or 0.10) according to a test of proportions.</td>
</tr>
<tr>
<td>3.1</td>
<td>Pseudo $p$-values as calculated by the STPSS associated with Clusters 1 and 2 for the hotspots of varying intensity.</td>
</tr>
<tr>
<td>3.2</td>
<td>Proportion of perturbed patterns whose MLC overlaps Cluster 1 in both space and time.</td>
</tr>
<tr>
<td>3.3</td>
<td>Proportion of perturbed patterns overlapping the original Cluster 1 in both space and time with $p$-values below 0.05.</td>
</tr>
<tr>
<td>3.4</td>
<td>Average pseudo $p$-values for MLCs observed in perturbed datasets and percent of patterns rejecting the null hypothesis (at $\alpha = 0.05$) for various levels of perturbation.</td>
</tr>
</tbody>
</table>
### Table 4.1 Population trends and cases of Burkitt’s Lymphoma in the West Nile District derived from Williams et al. (1978).

Page 84

### Table 4.2 Results for the combined Jacquez test ($J_{10}$) across the study time periods based on 999 permutations.

Page 87
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Map of the hypothetical metropolitan area within which burglary and theft events are simulated. Darker areas indicate regions with a higher likelihood of observing a criminal event.</td>
<td>24</td>
</tr>
<tr>
<td>2.2 Profile showing the probability of burglary and theft events over the course of one week, within the different risk regions of the study area. Note, the risk regions here correspond to those identified in the study area map.</td>
<td>25</td>
</tr>
<tr>
<td>2.3 Distribution of the range of temporal uncertainty associated with burglaries and thefts in Mesa, AZ for the period 2004–2009.</td>
<td>29</td>
</tr>
<tr>
<td>3.1 Intensity of the three simulated event patterns. Each space-time cube portrays the intensity of events generated across space and time for the three patterns. The left most cube corresponds to the high intensity cluster pattern (where cluster events are concentrated in a smaller area) and the right-most cube corresponds to the lowest intensity pattern. The top of each cube shows an areal view of the study area (i.e. a conventional map), while the sides show the intensity through time (represented as the height of the cube) and across a particular spatial dimension (either x or y). Lighter areas indicate a higher intensity of events.</td>
<td>51</td>
</tr>
<tr>
<td>3.2 Cumulative distribution functions showing the distribution of offsets associated with each level of perturbation.</td>
<td>52</td>
</tr>
<tr>
<td>3.3 Sample of burglary events occurring in Mesa, Arizona during 2008 employed in the analysis. Additional geographic identifiers have been omitted from the map to preserve privacy.</td>
<td>54</td>
</tr>
</tbody>
</table>
3.4 Plots of MLCs identified by the STPSS. The spatial footprint of the MLCs for the original datasets are shown in red and the secondary cluster with the next lowest $p$-value is shown in green. MLCs from perturbed versions of the same dataset are shown in black. The intensity of the original clusters decreases from the top down while the intensity of perturbation increases from the left to the right. This layout is followed in subsequent graphics. 56

3.5 Plots of the duration of MLCs identified with the STPSS. The duration of the MLCs for the original datasets are denoted using horizontal red bars, secondary clusters are shown using green bars. MLCs from perturbed versions of the same dataset are shown as thin black vertical lines. 58

3.6 Plots of the spatial distribution of clusters identified by the STPSS which overlap the MLC identified in the original pattern. 60

3.7 Plots of the temporal distribution of clusters identified by the STPSS which overlap the MLC identified in the original pattern. 61

3.8 Plots showing the distributions of spatial and temporal distances between the center of Cluster 1 (identified in the original data) and the center of clusters identified in the perturbed datasets which overlap Cluster 1. Each plot corresponds to a different initial intensity and the distributions across different perturbation levels are represented using different colors in each plot. 62

3.9 Plots of the spatial area and temporal durations of clusters identified in the perturbed datasets which overlap the original Cluster 1. The area and duration of the original Cluster 1 is identified as a hashed vertical black line in the figures. 63

3.10 Plots of MLCs identified within the Mesa crime data using the STPSS. The spatial footprint of the MLC for the original dataset is shown in red. MLCs from perturbed versions of the same dataset are shown in black. 64
<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.11 Plots of the duration of MLCs identified within the Mesa crime data using the STPSS. The duration of the MLC for the original dataset is denoted using horizontal red lines. MLCs from perturbed versions of the same dataset are shown as black vertical lines.</td>
<td>65</td>
</tr>
<tr>
<td>3.12 Plots showing the distributions of spatial and temporal distances between the center of the MLC identified in the original data and the center of MLCs identified in the perturbed versions of the Mesa dataset. The different perturbation levels are represented using different colors in each plot.</td>
<td>66</td>
</tr>
<tr>
<td>3.13 Plots of the spatial area and temporal durations of clusters identified in the perturbed datasets which overlap the original MLC in the Mesa data. The area and duration of the original MLC is identified as a hashed vertical black line in the figures.</td>
<td>66</td>
</tr>
<tr>
<td>4.1 West Nile district of Uganda. Cases of Burkitt’s Lymphoma between 1961 and 1975 are shown as grey dots.</td>
<td>83</td>
</tr>
<tr>
<td>4.2 A collection of plots to contextualize the $J_k$ results for the 1961-1965 and 1972-1973 time periods. The pseudo-significance of test statistics is determined via the traditional permutation approach (using 999 permutations) and is denoted by filled circles. The dashed lines denote the approximate scale of observed space-time interaction.</td>
<td>89</td>
</tr>
<tr>
<td>4.3 Space-time cube of cases for 1961-65 and connections for $J_5$.</td>
<td>91</td>
</tr>
<tr>
<td>4.4 A set of plots to visualize space-time interaction as detected by the Jacquez test ($J_5$): (a) Conventional map of nearest neighbor connections. (b) Plot of nearest neighbor connections with latitude on the $x$-axis and time on the $y$-axis. (c) Plot of nearest neighbor connections with longitude on the $x$-axis and time on the $y$-axis. (d) Space-time cube of nearest neighbor connections.</td>
<td>93</td>
</tr>
<tr>
<td>4.5 A comparison of the locations of significant linkages identified by the Jacquez $(k = 5)$ and Knox (space = 13 km, time = 90 days) tests.</td>
<td>95</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>4.6</td>
<td>A comparison of the corrected and uncorrected combined Jacquez test results</td>
</tr>
</tbody>
</table>
Researchers across a variety of fields are often interested in determining if patterns observed within their data are of a random nature or if they may have been caused by some alternative and potentially more meaningful process. Although direct, causal links between patterns and processes are elusive due to the problem of equifinality (Franklin, 2010; Anselin and Rey, 2010), identification of structured patterns often suggests the presence of an interesting data generating process or combinations thereof. As such, a variety of methodologies have been put forth to identify structure within patterns. While the diversity of such measures is as expansive as the types of patterns and data they analyze, this dissertation explores one family of methods designed specifically to identify structure within patterns of three-dimensional space-time event data. The work endeavors to better understand how these metrics to identify so-called “space-time interaction” are affected by uncertainty and inaccuracies commonly encountered in their input data and, additionally, offers a series of improvements to enhance the utility of one of the methods.

The space-time event data analyzed by these methods are discrete instances of some phenomenon marked with spatial and temporal coordinates. Examples of such data, commonly analyzed using these methods, include cases of disease or criminal incidents. A pattern of these events exhibits space-time interaction when, generally speaking, pairs of events which are close to each other in space are also close to each other in time, more so than would be expected due to random chance. In practice, this means events which occur in close proximity to each other in space, also occur near one another in time, relative to the other events in the pattern. Although groups of events within the same pattern may be closer to each other in space and time than would be expected due to randomness (i.e. the events may exhibit both spatial and temporal clustering), space-time
interaction occurs only in instances where there is a significant positive relationship between the spatial and temporal distances between pairs of events (Kulldorff, 1998; Tango, 2010). Such interaction can be global, occurring across the entire dataset, or localized, present only in discrete hotspots. Establishing the presence of either type of interaction is important because it may indicate a proximate underlying causal process. For example, in the context of disease cases, interaction may be linked to an infectious or viral etiology or the presence of transient localized hazard exposure (Marshall, 1991; Jacquez, 1996; McNally, 2010), while in relation to criminal events it may point to foraging, near-repeat victimization or other similar spatio-temporal trends (Knox, 2002; Townsley et al., 2003; Grubesic and Mack, 2008; Johnson et al., 2009).

Given the specific nature of space-time interaction, methods to establish its presence are necessarily distinct from conventional methods for detecting spatial or temporal clustering. Tests of interaction are designed to “detect space-time clustering above and beyond any purely spatial or purely temporal clustering;” meaning, the tests determine if event pairs that are close in space are also close in time (Kulldorff, 1998, pg. 58). The null hypothesis of these tests is that there is no relationship between the spatial and temporal distances separating pairs of events. The alternative hypothesis is that events which are near to each other in space also tend to be near to each other in time. Distinct methods exist to identify global and local space-time interaction (Kulldorff, 2010; Tango, 2010). Global tests are focused on establishing if interaction is present across an entire space-time event pattern while local tests are employed to identify hotspots in space and time where the concentration of events is higher than would be expected due to spatio-temporal randomness. The global tests of space-time interaction most frequently employed in practice and in the literature include the Knox test (1964), Mantel test (1967) and Jacquez test (1996). Local tests of space-time interaction include the cylindrical space-time scan (Kulldorff et al., 1998), the flexibly shaped space-time scan statistic.
(Takahashi et al., 2008) and the space-time permutation scan statistic (Kulldorff et al., 2005). Both global and local metrics are commonly employed within the fields of public health (Kulldorff, 1998; Ward and Carpenter, 2000b; McNally and Colver, 2008; Meliker, 2009; Rogerson and Yamada, 2009; Tango, 2010) and criminology (Knox, 2002; Johnson and Bowers, 2004; Grubesic and Mack, 2008; Johnson, 2010) and have also been employed within fields such as ecology (Legendre and Fortin, 1989; Fortin and Gurevitch, 1993; Michener, 1997; Legendre and Fortin, 2010), forestry (Jacquez, 1996) and others (Johnson and Braithwaite, 2009; Braithwaite and Johnson, 2012).

Although these tests are widely employed, more exploration is needed to determine how their reported results may be affected by data quality problems commonly encountered in their implementation. In addition, known inadequacies, such as susceptibility to population shift bias (Kulldorff and Hjalmars, 1999; Mack et al., 2012), must be ameliorated. When employing these methods in an applied context, problems are often encountered (and just as often overlooked) surrounding uncertainty, inaccuracy and completeness in the spatio-temporal data being analyzed. In a survey of research needs within the field of spatio-temporal epidemiology, Meliker and Sloan (2011) underscore the importance of quantifying the effect of these deficiencies on epidemiological analyses. They state, “methods for quantifying locational, attribute, and temporal uncertainty [and inaccuracy], and propagating these uncertainty estimates into spatial pattern analyses, exposure assessments, or epidemiologic regression analyses are important avenues of research” (pg. 7). Quantitative analyses of spatio-temporal event data in the fields of criminology and spatial epidemiology are often affected by data quality problems resulting from inaccurate geocoding, incomplete data reporting and privacy concerns (Ratcliffe and McCullagh, 1998; Jacquez, 2004; Meliker and Sloan, 2011). As a rule, these datasets are imperfect representations of the phenomena of interest (crimes or cases of disease), and subject to different forms of error (Goodchild, 2000; Jacquez, 2004). This
error, inherent in the data, may be propagated throughout the analyses and have a considerable effect on the results of analyses (Arbia et al., 1998). This work explores the effect of these problems on the results of space-time interaction tests. Although previous work has partially investigated the impact of locational inaccuracy on the results of global tests (i.e. Jacquez and Waller, 2000) and shown a considerable effect, more work remains to better understand the effect of this and other forms of inaccuracy. For example, the effects of temporal inaccuracy or combinations of spatial and temporal inaccuracies on the results of these tests remain completely unexplored.

This dearth of research is extremely troubling and I address it here directly. The major contributions of this dissertation are twofold: (1) it generates a better understanding of how common data quality problems may impact the results of both global and local tests of space-time interaction and (2) it formulates an improved version of the Jacquez $k$-nearest neighbor test which accounts for the problem of population shift bias. The studies carried out here are an important contribution to the scientific corpus because these tests have been widely advertised in the instructive literature of spatial epidemiology (e.g. Lawson, 2006; Pfeiffer et al., 2008; Rogerson and Yamada, 2009; Tango, 2010) and are increasingly employed in the field of criminology (Johnson, 2010) without consideration of how sensitive their results may be to known or suspected errors in the input data. This investigation aims to explore this sensitivity and educate users about the impact these errors may have on subsequent results. When interaction is detected in these contexts, more work is often warranted at a finer scale to establish the specific process responsible (Pfeiffer et al., 2008; McNally, 2010). Alternatively, further investigation may be abandoned if no interaction is detected and the pattern of events is perceived to be random. By generating a better understanding of these methods and their performance in the face of common data deficiencies, the proposed work will help users to avoid false positive or false negative results and, by extension, misallocation of resources and missed
opportunities. Reducing these errors will help practitioners focus efforts on instances where true interaction is occurring. With this goal in mind, the work also formulates an enhanced version of the Jacquez $k$-nearest neighbor test which aims to ameliorate the known problem of population shift bias (Kulldorff and Hjalmars, 1999; Mack et al., 2012) in its results. The work provides an alternative version of this powerful test for global space-time interaction that is less vulnerable to the issue hampering the original.

This dissertation progresses as follows. First, Chapter 2 investigates the effect of these deficiencies in the input data on results of three of the most commonly employed tests of global space-time interaction: the Knox (1964), Mantel (1967) and Jacquez (1996) tests. Specifically, the effects of locational and temporal inaccuracy present in the input data on the results of each test are explored. While previous work has partially investigated the impact of locational inaccuracy on the results of these tests (i.e. Jacquez and Waller, 2000), the impact of temporal inaccuracy remains unexplored. Here I examine the sensitivity of the results of these tests to these problems, individually and collectively, using a series of simulation experiments. Essentially, the experiments attempt to understand the effect of introducing these problems on results both when interaction was originally present and when it was not, the two situations that may be encountered by practitioners. Results of these experiments indicate that in some cases, these common data problems can completely obscure evidence of space-time interaction (Type II errors) and in others create it where it does not exist (Type I errors). Although the findings are to some degree data- and design-specific (i.e. the quantity of these errors introduced), the take-home message remains clear: the possibility exists that even a moderate degree of these common problems can drastically alter the results of these tests. Estimates of confidence in their results that fail to consider the potential impact of these problems must not be taken at face value.
Next, Chapter 3 extends the investigation carried out in Chapter 2 and examines the potential impact of inaccurate and uncertain input data on results of analyses for a local metric of space-time interaction, the space-time permutation scan statistic (STPSS). The STPSS is designed to identify hot (and cool) spots of space-time interaction within patterns of spatio-temporal events. While the method has been adopted widely in practice, there has been little consideration of the effect inaccurate input data may have on its results. Again, given the pervasiveness of inaccuracy and uncertainty within spatio-temporal datasets and the popularity of the method, this issue warrants further investigation. Here, a series of simulation experiments using both synthetic and real-world data are carried out to better understand how deficiencies in the spatial and temporal accuracy of the input data may affect the results of the STPSS. The parameters for the deficiencies introduced in these experiments are identical to those used in exploring the global tests in Chapter 2. The findings reveal a surprising robustness of the method’s results in the face of these deficiencies. As expected, the experiments demonstrate that greater degradation of input data quality leads to greater variability in results. Additionally, they show that weaker signals of space-time interaction are those most affected by the introduced deficiencies. However, in stark contrast to previous investigations into the impact of these input data problems on global tests of space-time interaction, this local metric is revealed to be only minimally affected by the degree of inaccuracy and incompleteness introduced in these experiments, with most signals of space-time interaction being detected by the method even in the face of the introduced deficiencies.

The first two chapters aim to understand the direct impacts of spatial and temporal uncertainty and inaccuracy on the results of global and local tests, helping researchers to understand when confidence in results is warranted and when it is not, given the quality of data that were input into the methods. Chapter 4, however, shifts focus of the work and
offers suggestions for improving one of the global metrics of space-time interaction, the Jacquez $k$ nearest neighbor test. The Jacquez test, originally developed to improve upon shortcomings of existing tests for space-time interaction, has been shown to be a robust and powerful method of detecting interaction. Despite its flexibility and power, the test has three main shortcomings: (1) it discards important information regarding the spatial and temporal scale at which detected interaction takes place; (2) the results of the test are not visualized; (3) research has demonstrated the test to be susceptible to population shift bias. This study presents enhancements to the Jacquez $k$ nearest neighbors test with the goal of addressing each of these three shortcomings and improving the utility of the test. Spatio-temporal data for cases of Burkitt’s lymphoma in Uganda between 1961-1975 are employed to illustrate the modifications and enhanced visual output of the test. Output from the enhanced test is compared to that provided by alternative tests of space-time interaction. Results show the enhancements presented in this study transform the Jacquez test into a complete, descriptive, and informative metric that can be used as a stand alone measure of global space-time interaction with results less affected by the problem of population shift bias than its original formulation.

Finally, Chapter 5 concludes the dissertation with a summary of key findings and outlines directions for future research. The exploration conducted here provides a first step at educating end-users about problems to be cognizant of when employing these metrics and offers methodologists interested in furthering work on these or similar methods insight into areas where future attention should be directed. These contributions will help practitioners and researchers in the fields of spatial epidemiology, criminology, and beyond to utilize these tests more appropriately and judiciously, helping them to minimizing errors due to data quality problems and population shift bias. Additionally, the work identifies the need and sets the stage for the development of more robust global
methods for identifying space-time interaction able to better account for the ubiquitous challenges presented by uncertainty and inaccuracy in spatio-temporal event data.
Tests of space-time interaction are used to examine patterns of spatio-temporal events and determine if, generally speaking, events which are close to each other in space are also close to each other in time, and vice versa (Jacquez, 1996; Kulldorff, 1998). These tests are primarily employed in the fields of spatial epidemiology (McNally, 2010; Tango, 2010), ecology (Legendre and Fortin, 1989; Fortin and Gurevitch, 1993; Jacquez, 1996; Michener, 1997; Legendre and Fortin, 2010) and criminology (Johnson, 2010; Leitner and Helbich, 2011) to help link observed event patterns to certain processes which may be responsible for their development. For example, in epidemiological studies, the detection of interaction within a pattern of disease cases may indicate the presence of a contagious process or exposure to transient, localized hazards (Marshall, 1991; Jacquez, 1996), while the presence of interaction among criminal incidents has been linked to sprees (Knox, 2002), near repeat victimization (Townsley et al., 2003) and criminal foraging (Johnson et al., 2009).

In spite of the widespread application of these tests in these contexts, there is only a limited understanding of how common deficiencies found in their input data may subsequently impact results (Jacquez and Waller, 2000; Jacquez, 2004; Meliker and Sloan, 2011). This work addresses this paucity by examining how tests of space-time interaction are affected by locational and temporal inaccuracy associated with their input data. Given that it is often impossible to avoid inaccuracy in the datasets under analysis, a better understanding of the effects of these common data problems is essential in order to gauge the amount of confidence that users should place in the results of these tests. Often, in the
applied contexts mentioned above, when interaction is detected by these tests, more work is required at a finer scale to establish the specific process responsible (Pfeiffer et al., 2008; McNally, 2010). Alternatively, further investigation may be abandoned if no interaction is detected and the pattern of events is perceived to be random. By generating a better understanding of these methods and their performance in the face of common data problems, this work will help practitioners to better identify situations where further investigation may be fruitful and help them to avoid allocating resources in situations where the tests may be reporting erroneous or misleading results.

This study investigates the effect of deficiencies in the input data on the results of three of the most commonly employed tests of global space-time interaction: the Knox (1964), Mantel (1967) and Jacquez (1996) tests. Simulation experiments are carried out to determine the sensitivity of these tests to inaccuracies in the input data. The results of these experiments indicate that in some cases, these common data problems completely obscure evidence of space-time interaction, while in others they may create it where it did not originally exist. Although the findings are to some degree data-specific, the take-home message remains clear: the possibility exists that even a conservative degree of these common problems (such as that modeled here) can alter the results of these tests. Estimates of confidence in their results that fail to consider the potential impact of these problems must not be taken at face value.

The chapter proceeds as follows. Section 2.2 provides a brief survey of the literature related to inaccuracy and uncertainty in spatio-temporal data and establishes a common working vocabulary for the study. Next, Section 2.3 reviews the tests explored in this chapter. Section 2.4 then describes the methodology for the experiments carried out as part of this study. The results are reported in Section 2.5. The results are then discussed and final conclusions are offered in Section 2.6.
It is well recognized within the field of geographic information science that spatial and spatio-temporal data can offer only imperfect representations of real-world phenomena. Zhang and Goodchild (2002, pg. 4) note this “inevitable discrepancy between the modeled and real worlds [...] may turn spatial decision-making sour.” Unwin (1995, pg. 549) goes so far as to claim that a consequence of this discrepancy is “an evident inability of GIS technology to adequately represent the real world in a form that is able to inform much geographical theory.” The challenge of addressing this gap between model and reality is daunting. Considerable research has endeavored to raise awareness of this problem, better understand its effects and develop methods which inherently account for it (e.g. Goodchild and Gopal, 1989; Unwin, 1995; Fisher, 1999; Congalton, 2000; Foody and Atkinson, 2002; Shi et al., 2002; Zhang and Goodchild, 2002). From such studies, a language has evolved to describe the various facets of the problem. Drawing heavily from Unwin (1995), the definitions of major terms are reviewed below to establish a common vocabulary for the purposes of this study.

To begin broadly, error is defined by Heuvelink (1993, pg. 27) and subsequently Unwin (1995, pg. 550) as “the difference between reality and our representation of reality.” Aside from broader deficiencies related to the specific model chosen to mimic real-world phenomena, there are a number of common components of error in spatial and spatio-temporal data, including (1) inaccuracy, (2) imprecision and (3) incompleteness (Unwin, 1995). Although accuracy and precision are often conflated in lay conversation, they have distinct technical definitions. In a classical sense, precision of measurements refers to their degree of repeatability. That is, if numerous measurements were taken of the same quantity they would be considered precise if there were little variation in the reported values and imprecise if this were not the case. In a more modern sense, precision
can also refer to the exactness with which values are stored in a database (Dutton, 1989). In the context of geographic information science this often translates to the number of significant digits associated with measurements (Unwin, 1995). Meanwhile, accuracy refers to a measure of how close the measured value is to the true value (Dutton, 1989; Unwin, 1995). By definition, quantifying inaccuracy requires knowledge of the true value. In the context of spatio-temporal event data, this may refer to the actual location and instant an event occurred, rather than the place or time serving as its proxy in a database. The distance in space and time between the measured coordinates and the actual coordinates would constitute the spatial and temporal inaccuracy associated with the event. These components of error may manifest themselves differently across dimensions. For example, the locational precision and accuracy of measurements may be very high but the accuracy or precision of the measurement in time may be lacking. It should also be noted that precision does not guarantee accuracy, and vice versa. Error can also result when the data fail to completely describe the phenomena they claim to represent. Missing or omitted observations result in an incomplete, and therefore flawed, dataset, based on the definition of error provided above.

Although the quality of data ultimately depends upon context and purpose (Unwin, 1995), generally speaking, a high degree of inaccuracy, imprecision and/or incompleteness will negatively impact quality. As many studies have shown, decreased quality in the input data can then, in turn, propagate to uncertainty (doubt and distrust) in the results (Arbia et al., 1998; Kiiveri, 1997; Ratcliffe, 2004). Of course, uncertainty may also manifest in the individual components as well. Goodchild (1998) states, “uncertainty has emerged as the preferred term for all that the database does not capture about the real world, or the difference between what the database indicates and what actually exists.” Consider, how certain can one really be that their data are accurate, precise and complete?
This work will focus on evaluating how known or suspected inaccuracies in spatio-temporal event patterns, specifically in their locational and temporal coordinates, impact the results of analyses testing for space-time interaction. In the sections that follow, existing literature on these and related topics is reviewed, exploring both the sources and characteristics of inaccuracy and uncertainty in spatio-temporal data and the impacts of these problems on subsequent analyses.

**Locational Inaccuracy**

The presence of locational inaccuracy in spatial data and its associated consequences on cross-sectional analyses are well documented. In the context of point or event data, much of this literature is concerned with incorrect geocoding. According to Goldberg et al. (2007, pg. 33), geocoding is “the act of turning descriptive locational data such as a postal address or a named place into an absolute geographic reference.” Commonly, descriptive data are either linked to a geographic area or a unique point in space (Rushton et al., 2006). Geocoded data are often assumed to provide an accurate and complete representation of the phenomena examined; however, this is virtually never the case. A number of studies have explored errors in geocoded data, considering both locational accuracy and completeness or “match rate” and concluded that errors are essentially unavoidable (Goldberg et al., 2007; Zandbergen and Hart, 2009). In conducting analyses based on geocoded data, the appropriate question then seems to be not are errors present?, but how much error is acceptable? and, what methods are robust to these problems?

Studies examining positional accuracy in geocoded point data have found it to vary largely based on the geocoding method used and the quality of the underlying spatial data upon which it is based (Rushton et al., 2006; Whitsel et al., 2006; Zhan et al., 2006;)

---

1In addition to locational and temporal inaccuracies, an incomplete dataset is also of serious concern to the integrity of spatial and spatio-temporal analyses. However, this problem is not considered here as my focus is on issues pertaining to the spatial and temporal coordinates of the input data. Interested readers are referred to work conducted and reviewed by Malizia (2013).
Zandbergen, 2008; Zandbergen et al., 2011; Zandbergen, 2011) as well as the density of settlement within a study area (Bonner et al., 2003; Cayo and Talbot, 2003; Ward et al., 2005; Kravets and Hadden, 2007). Other studies have suggested (Cayo and Talbot, 2003) and demonstrated (Zimmerman et al., 2010) instances where the locational errors are spatially autocorrelated. Additional research has explored the statistical distribution associated with the locational inaccuracy of geocoding errors. In the case of Zimmerman et al. (2007), the distribution of distances between true and observed values were found to be modeled most appropriately with mixtures of bivariate $t$ distributions while the errors reported by Cayo and Talbot (2003) appear to follow an exponential distribution. A number of other studies have shown the distribution of locational errors to be log-normally distributed (Whitsel et al., 2006; Zandbergen, 2007; Zandbergen and Hart, 2009).

The effect of this positional inaccuracy on subsequent analyses is explored in a number of studies. Burra et al. (2002) explore the effect of geocoding errors on the results of global and local metrics of spatial autocorrelation, including Moran’s $I$, Anselin’s local Moran and Getis and Ord’s $G_i$ and $G^*_i$ statistics. Their investigation revealed that geocoding errors affecting only 1% of the original data were enough to distort the results of the local statistics while the global statistics appeared unaffected by this degree of inaccuracy. Further work by DeLuca and Kanaroglou (2008) also showed discrepancies in results of analyses conducted using different geocoding methods. In their study, they examine the effect on Kulldorff’s spatial scan statistic, kernel density estimation and bivariate $K$ functions, noting the results of the $K$ functions were least affected by the problems in the input data. Further work by Mazumdar et al. (2008) shows that the relationships between environmental exposure data and health outcomes are weakened when geocoding accuracy declines. Similarly, Jacquez and Rommel (2009) examined the effect of this positional inaccuracy on the formation of spatial weights matrices and found considerable effects suggesting that knowledge of geocoding accuracy is necessary for
understanding the validity of spatial weights derived from geocoded data and any subsequent analyses which employ them. Zinszer et al. (2010) also looks at the effect of inaccurate geocoding (both completeness and positional accuracy) and show a considerable effect on kernel density estimates of disease distribution. However, to counter this flood of troublesome findings, Zimmerman et al. (2010) demonstrate the positional errors in geocoded data to be spatially autocorrelated and illustrate that this actually mitigates the negative effect of these errors in certain tests of clustering.

While the locational inaccuracy explored in many of the preceding studies was unintended (or simulated to be so), this is not always the case. Occasionally, these errors are introduced intentionally, typically to preserve confidentiality and mask individual locations (Armstrong et al., 1999; Fefferman et al., 2005; VanWey et al., 2005). This is often the case for health data where patient confidentiality is mandated by law (Cox, 1996), or occasionally in crime data where it is necessary to preserve the identity of victims and/or offenders.2 A great deal of work, especially by health researchers, has focused on exploring methods to preserve confidentiality in reported data, while allowing it to retain characteristics that will provide useful information for spatial analysts (Leitner and Curtis, 2004; Boulos et al., 2006; Leitner and Curtis, 2006; Olson et al., 2006; Wieland et al., 2008). In practice, one common approach is to simply aggregate point data to larger areal units (Armstrong et al., 1999; Fefferman et al., 2005). However, this approach introduces the complexities of the modifiable areal unit problem (Openshaw, 1984; Ozonoff et al., 2007; Jeffery et al., 2009). Additionally, underlying inaccuracies in the original coordinates may result in observations being allocated to the wrong areal unit (Krieger et al., 2001; Ratcliffe, 2001; Kravets and Hadden, 2007). An alternative approach to preserving confidentiality involves obscuring point data by assigning a new location

---

2One additional point to raise is that locational error may also be introduced even prior to geocoding based on the level of precision with which the addresses of events are recorded. For example, some crime data are recorded or distributed by police departments aggregated to points within blocks or a nearest intersection.
randomly within some distance threshold (Armstrong et al., 1999; Wieland et al., 2008). Kwan et al. (2004) examined the difference in results between $K$ functions of an original dataset and versions masked using these random perturbations. Their study observed a tradeoff between the level of privacy and accuracy of results, but suggested an optimal threshold may exist in certain circumstances where both privacy and accuracy of results may be maintained.

It should be noted that the studies examined in this review thus far are concerned entirely with cross-sectional data. The only study to be found on the subject of spatial inaccuracy and its effect on spatio-temporal analyses was carried out by Jacquez and Waller (2000). Their study explored the difference between results of analyses conducted using the actual locations of cases of a disease and those where the locations had been moved to centroids of counties, bicounties and regions. The effect of this geographic masking was examined in the context of the Knox, Mantel, and Jacquez tests of space-time interaction. For each test, the study showed a positive relationship between the level of aggregation and the percentage of type II errors. The authors observed drops in statistical power ranging from 3.67% to 36.63% at the county level, from 6.58% to 44.1% at the bicounty level and from 22.60% to 59.83% at the regional level. In each case, the lowest decrease in power was observed in the Knox results, while the greatest was observed in the Jacquez results, suggesting the Jacquez test is more susceptible to the problem. Although this is an interesting study, the work needs to be extended and improved. The primary problem with the study is that it assumes locational inaccuracy in the data will be a direct result of moving observations to the centroid of a larger polygon to mask the individual locations. Any practical study employing such a methodology would meet severe criticism from reviewers. The interaction tests are specifically designed for use with individual event data. By moving all the events within a particular geography to the centroid of that geography the value of analyzing the data in this manner goes down
considerably. It would be more appropriate at that point to change the level of geographic support and use methods specifically designed for analyzing count data within a lattice. That being said, their paper does effectively illustrate a potential worst-case scenario of the effects of spatial inaccuracy and manipulation. This study will explore what effect a more reasonable level of locational inaccuracy may introduce into the analysis. For example, inaccuracy introduced by errors associated with geocoding as expressed above. Additionally, this study will also consider the effect of inaccuracy in the temporal data as discussed in the next section.

Temporal Inaccuracy

In contrast to the well-studied problem of locational inaccuracy, the problem of temporal inaccuracy and uncertainty in spatio-temporal data is seriously neglected. Given the pertinence of the problem in the context of spatio-temporal analyses and the considerable associated literature described above for strictly spatial data, one would expect some of the knowledge and logic to have filtered through to research on spatio-temporal data; however, as alluded to before, this has not been the case. There are no studies (which the author is aware of) that explore inaccuracy in the temporal dimension. This may be explained, at least within the field of spatial epidemiology, by the fact that temporal data often correspond to dates of diagnosis or onset of symptoms, which serve as proxies for infection date or the point when an illness began. It would seem that it is tacitly assumed that these data serve only as a rough representation of the phenomena of interest. For criminal data however, it is possible for the data to correspond to the exact date and time the offense occurred. While this can often be determined with considerable precision and accuracy for certain types of crimes (i.e. murders, assaults, etc.) for others, the crime is often reported after the event has occurred and the victim or reporting party may have only an approximate estimate of when the crime occurred, as in the case of property crimes such as burglary or theft. While such inaccuracies may be easily conceptualized in the
context of criminological data, in terms of health data, the entire logic of linking a case of
disease to a discrete time (or space, for that matter) seems dubious at best, especially in
the context of diseases with a long-latency period (Jacquez, 2004). However, this is the
prevailing approach to analyzing individual health data in the absence of the methods and
data supporting the analysis of personal exposure histories such as those suggested by the
work of Hägerstrand (1970) and Miller (2007).\textsuperscript{3} As such, the impact of the assumptions
associated with this approach need to be assessed.

\subsection*{2.3 Tests of Space-Time Interaction}

Here, the tests of space-time interaction considered in this study are examined in greater
detail. These methods are designed to “detect space-time clustering above and beyond any
purely spatial or purely temporal clustering;” meaning, the tests determine if event pairs
that are close in space are also close in time (Kulldorff, 1998, pg. 58). The null hypothesis
of these tests is that there is no relationship between the spatial and temporal distances
separating pairs of events. The alternative hypothesis is that events which are near to each
other in space also tend to be near to each other in time. There are distinct methods to
identify global and local space-time interaction (Tango, 2010); however, the focus of this.chapter is the former. Three of the most popular global tests are considered in this study:
the Knox, Mantel and Jacquez tests. Each are described in the subsections below.\textsuperscript{4}

\subsubsection*{Knox Test}

The concept of space-time interaction was introduced by Knox (1964). In his paper, Knox
defines the phenomenon of interaction and formulates the first metric to test for its
presence. To calculate the test, critical space and time distance thresholds ($\delta$ and $\tau$,
respectively) are specified by the user, defining adjacency between events. The test

\textsuperscript{3}Work by Jacquez et al. (2005) and Sabel et al. (2009) approaches this goal.
\textsuperscript{4}For an extended discussion of applied work employing these tests the reader is referred to reviews by
statistic is then calculated as the count of unique event pairs that are adjacent in both time and space. Formally, the test statistic is specified in Equation 2.1, where $n = \text{number of events}$, $a^s = \text{adjacency in space}$, $a^t = \text{adjacency in time}$, $d^s = \text{distance in space}$, and $d^t = \text{distance in time}$.

$$X = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a^s_{ij} a^t_{ij}$$  \hspace{1cm} (2.1)

$$a^s_{ij} = \begin{cases} 
1, & \text{if } d^s_{ij} < \delta \text{ and } i \neq j \\
0, & \text{otherwise} 
\end{cases}$$

$$a^t_{ij} = \begin{cases} 
1, & \text{if } d^t_{ij} < \tau \text{ and } i \neq j \\
0, & \text{otherwise} 
\end{cases}$$

In practice a permutation approach is often employed to generate an estimate of the variance of the statistic and test the null hypothesis of no space-time interaction (Mantel, 1967; Kulldorff and Hjalmars, 1999). Under this approach, the spatial coordinates of events are fixed, while the temporal coordinates are permuted. The test statistic is recalculated for each permutation and a distribution of results is generated. The observed statistic and its relative rank within the permuted distribution is used to assess the pseudo-significance of the original observed value.

**Mantel Test**

A modification of the Knox test was proposed by Mantel (1967). Mantel formulated two versions (unstandardized and standardized) of a test which considers the spatial and temporal distances between all pairs of events in the pattern. Following closely the form
of the Knox test, the unstandardized Mantel test statistic is the sum of the products of the spatial and temporal distances between all event pairs in the dataset. This statistic is specified in Equation 2.2.

\[ Z = \sum_{i=1}^{n} \sum_{j=1}^{n} (d_{ij}^s + c_s)^{p_s} (d_{ij}^t + c_t)^{p_t} \]  \hspace{1cm} (2.2)

Again, \(d^s\) and \(d^t\) denote distance in space and time, respectively. Mantel advocated the addition of constants (\(c_s\) and \(c_t\) in space and time, again, respectively) to the raw distances to prevent multiplication by zero. Additionally, he prescribed a reciprocal transformation of the resulting distance to temper the effect of larger distances between events. This transformation can be achieved by assigning a value of -1 to the spatial and temporal exponent terms, \(p_s\) and \(p_t\). In addition to the unstandardized version, Mantel also formulated a standardized version which amounts to a measure of correlation between the spatial and temporal distance matrices. The standardized test statistic is formulated in Equation 2.3, where \(\bar{d}^s\) refers to the average distance in space, and \(\bar{d}^t\) the average distance in time between all pairs of events. For notational convenience \(\sigma_{d^s}\) and \(\sigma_{d^t}\) refer to the sample (not population) standard deviations, for distance in space and time, respectively. This standardized test statistic is confined to the range \(-1 < r < 1\).

\[ r = \frac{1}{(n^2 - n - 1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[ \frac{d_{ij}^s - \bar{d}^s}{\sigma_{d^s}} \right] \left[ \frac{d_{ij}^t - \bar{d}^t}{\sigma_{d^t}} \right] \]  \hspace{1cm} (2.3)

The significance of both test statistics is assessed using the permutation approach suggested above for use with the Knox test. The standardized version of the test will be employed in the analyses conducted in this study.
Jacquez Test

In an effort to address shortcomings of the Knox and Mantel methods, Jacquez (1996) developed a test of space-time interaction based on nearest neighbor distances. Jacquez proposed two statistics: a cumulative measure of interaction, \( J_k \), and a \( k \)-specific measure, \( \Delta J_k \). The cumulative measure locates the \( k \) nearest neighbors to a point in both space and time and then tabulates the number of events that are nearest neighbors in both dimensions. The value for \( k \) is specified by the user. This is expressed formally in Equation 4.5, where \( n \) = number of cases; \( a_s \) = adjacency in space; \( a_t \) = adjacency in time.

\[
J_k = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ijk}^s a_{ijk}^t
\]  

\[
a_{ijk}^s = \begin{cases} 
1, & \text{if event } j \text{ is a } k \text{ nearest neighbor of event } i \text{ in space and } i \neq j \\
0, & \text{otherwise}
\end{cases}
\]

\[
a_{ijk}^t = \begin{cases} 
1, & \text{if event } j \text{ is a } k \text{ nearest neighbor of event } i \text{ in time and } i \neq j \\
0, & \text{otherwise}
\end{cases}
\]

The \( k \)-specific statistic, \( \Delta J_k \), is a measure of space-time interaction in excess of that observed for \( J_{k-1} \). This additional metric was formulated because values for the cumulative test statistic (\( J_k \)) include counts of proximate events also accounted for in tests using smaller values of \( k \) (1996). The \( \Delta J_k \) statistic, however, is independent of other levels of \( k \). The formulation of this statistic is given in Equation 4.6.

\[
\Delta J_k = J_k - J_{k-1}
\]  

21
The significance of both the cumulative and $k$-specific statistics is assessed using the permutation method described above for the other tests. This study will only consider the cumulative Jacquez test.

2.4 Methods

Two experiments were run to better understand how the results of these three tests may be affected by the previously described data problems. The first considers patterns demonstrated to exhibit space-time interaction and examines how perturbing the spatial and temporal coordinates of their constituent events may affect the ability of these tests to detect the interaction originally present in the pattern. This is done by perturbing the patterns and examining the frequency with which the tests reject the null hypothesis of no space-time interaction when analyzing the perturbed patterns. In the second, a set of random space-time patterns are generated. The spatial and temporal coordinates of the patterns are perturbed and the patterns are then tested for interaction. The percentage of patterns for which the null hypothesis is rejected (for each test and various values of $\alpha$) are recorded. The goal of the first experiment is to establish the the likelihood of correctly rejecting the null hypothesis of no space-time interaction in the face of these perturbations. Essentially, the experiment explores the potential contribution of these data problems to committing a Type II error, i.e. not rejecting the null hypothesis when in fact the alternative hypothesis is true for the original data. Meanwhile, the second experiment seeks to determine if the presence of spatial and temporal perturbations may introduce interaction into test results, thereby increasing the likelihood of committing a Type I error. The specifics of these experiments are discussed below.
Experiment 1

Here, the effects of introducing spatial and temporal inaccuracies into patterns which exhibit space-time interaction are examined. The purpose of this experiment is to gain a better understanding of how these data problems may affect the ability of these tests to correctly reject the null hypothesis of no space time interaction.

To begin, a series of patterns are generated which exhibit space-time interaction. To create these patterns, a simplistic model of interacting spatio-temporal events was created. Developed to mimic a pattern of burglary and theft data, the model documents the probability of events, or “crimes” in this case, occurring at locations and times within a hypothetical metropolitan area over the course of one month. A map of the metropolitan area is shown in Figure 2.1. Event patterns were simulated by assigning events to locations and times within this study area over the course of a 30 day period based on a probability distribution identifying the likelihood of these criminal events throughout space and time. Darker areas in the map indicate regions where the risk of a criminal event is greater. The probability distribution underlying the event generation and its change over time and space (the latter represented by the different risk regions documented in the map) is shown in Figure 2.2. This distribution was designed to mimic findings that burglary and theft events are known to peak at certain points during the day and during the week (Sorensen, 2004; Townsley, 2008). Here, probabilities are elevated during evenings and weekends to coincide with the increased frequency of property crimes during these periods. While this model is minimalistic, this study is not intended to be an exercise in modeling crime, but rather an exploration of how uncertainty and inaccuracy in spatio-temporal data may affect the results of these tests. The model presented here, although simple, is adequate for these purposes. When events are simulated following the

---

5 The same hypothetical metropolitan area was employed in a study by Mack et al. (2012) in their study of population shift bias in these tests.
Figure 2.1: Map of the hypothetical metropolitan area within which burglary and theft events are simulated. Darker areas indicate regions with a higher likelihood of observing a criminal event.

probabilities expressed in this distribution, hot spots of burglaries and thefts in space and time will result more frequently than would be expected under complete spatio-temporal randomness.

Using this model, 1,000 event patterns were simulated which serve as the basis for selecting patterns to which spatial and temporal perturbations will be introduced. The intensity ($\lambda$) of the event patterns was kept constant at 100 events to ensure comparability of results across tests and perturbation parameters. These patterns were then tested for
space-time interaction using the three tests. Pseudo-significance of the test statistics were
established using a Monte Carlo permutation approach employing 999 permutations. The
generation of these events, as well as the rest of the experiments conducted here, were
coded using the Python programming language (van Rossum and Drake, 2009). Code to
run the various tests of space-time interaction has been made available as part of PySAL,
an open-source spatial analysis library (Rey and Anselin, 2010).

From this series, three space-time event patterns are selected for use with each test
(i.e. Knox, Mantel and Jacquez). Three patterns for each test (nine overall) were selected
because each exhibited a different degree of space time interaction: high, medium, and
low according to their corresponding test. These degrees of interaction correspond to the pattern being pseudo-significant at \( \alpha \) levels of 0.01, 0.05 and 0.10, respectively, according to the test of interest.\(^6\) These original patterns are then subjected to the deficiencies described above (i.e. locational and temporal inaccuracy). The process employed here to degrade the quality of the datasets is described below after the description of Experiment 2. The deficiencies are introduced both individually and collectively to the datasets. Each dataset is subjected to these deficiencies multiple times to generate a series of perturbed patterns generated from each original pattern. For each original pattern, seven separate series of perturbed patterns are created (three each subject to locational inaccuracies, one subject to temporal inaccuracies, and three a combination of the two deficiencies). Each series contains 1,000 perturbed versions of the original pattern. The perturbed patterns in these series are then tested for interaction. For each series the proportion of results for which the null hypothesis is rejected is recorded. These proportions are explored in the results.

**Experiment 2**

In the second experiment, 2,000 space-time event patterns are generated. Here, rather than employ the probability model described above, random \( x, y \) and \( t \) coordinates are drawn from the interval \([0,10]\) assuming a uniform probability distribution across all dimensions.\(^7\) Each pattern is then tested for interaction using the three tests and the pseudo \( p \)-values are recorded for each observed test/pattern combination. From these, a series of 1,000 patterns are selected which do not exhibit space-time interaction according to any of the tests (i.e. they have pseudo \( p \)-values greater than 0.10 according each of the

---

\(^6\)Individual patterns were used for each test and \( \alpha \) combination because there is not uniform agreement across the three tests regarding the degree of interaction (as measured by their \( p \)-value) for individual patterns. As such it is difficult to identify individual patterns that exhibit the same degree of interaction according to the various tests. To cope with this, three patterns were picked for each test which demonstrated pseudo \( p \)-values just below the three different \( \alpha \) levels. This ensures that the degree of interaction (as measured by \( p \)-value) is similar across the tests, prior to perturbation.

\(^7\)This corresponds to a study area 10 km\(^2\) and ten days in length.
tests). This selection took place to ensure that none of the patterns in the series exhibited interaction. If this step had not been taken and only 1,000 random patterns were generated, a percentage of those patterns (equivalent to the nominal $\alpha$) would have exhibited space-time interaction. These non-interacting patterns are then distorted using the same parameters as above in Experiment 1, generating seven perturbed versions of each original pattern. The resulting seven series of 1,000 perturbed datasets are then also tested for interaction. Again, the rates at which the null hypothesis is rejected for various $\alpha$ values are recorded and examined. Whereas the first experiment focuses on examining the consequences of inaccuracies in patterns exhibiting space-time interaction, this experiment looks at the consequences across a series of patterns which do not originally exhibit space-time interaction. Instead of focusing on the impact of these inaccuracies on Type II errors and subsequently the statistical power associated with the tests, this experiment examines whether or not these problems introduce interaction into patterns where it was originally not found. This, in turn, speaks to the possibility that these problems may contribute to Type I errors.

Inaccuracy Parameters

In the experiments above, common data deficiencies (i.e. locational and temporal inaccuracies) are introduced into the simulated data. Because the data have been simulated, the spatial and temporal coordinates are known accurately, precisely and completely for each original pattern. The coordinates associated with a single event $i$ in a pattern can be represented formally by the tuple $(x_i, y_i, t_i)$. However, as discussed in Section 2.2, these data are often obscured by inaccuracies and uncertainty, such that the coordinates are offset by values represented by the tuple $(\Delta x_i, \Delta y_i, \Delta t_i)$. The parameters employed in this study to model these deficiencies and their empirical roots in the literature are described here.
Random locational inaccuracies (i.e. $\Delta x$ and $\Delta y$) were introduced into the data to model the errors commonly encountered in geocoded data. To subject a pattern of events to these inaccuracies, the events in the patterns are offset from their original locations by a distance that was randomly drawn from an exponential distribution. Three levels of spatial inaccuracy were introduced in the experiments: low, medium and high. The means of the exponential distributions used for these levels of spatial offset are 50 m, 100 m and 200 m, respectively. Direction was assigned randomly. The choice to employ exponential distributions to model the locational inaccuracy was based on the work conducted by Cayo and Talbot (2003) reviewed above. The mean values for the distribution of inaccuracies at the various levels of perturbation were also derived from their study. In the study, the authors note that the mean distance associated with the locational inaccuracy is dependent upon the type of geocoding and the density of settlement. For geocoding based on an underlying street network (a common approach) they report a mean value of 58 meters for urban areas, 143 meters for suburban areas and 614 meters for rural areas. The values employed here were chosen to provide a low to middling estimate of these positional inaccuracies. Consequently, the impact of this inaccuracy on the test results should provide a similarly conservative estimate of the possible associated problems.

To perturb the original data with temporal uncertainties (i.e. $\Delta t$), a similar approach was taken to that used for the locational inaccuracies. However, due to the aforementioned lack of literature on temporal inaccuracies within spatio-temporal data, parameters for the perturbations could not be drawn from the literature. Instead, raw data on the suspected range of temporal inaccuracies associated with burglaries and thefts were used. These data were acquired from the Mesa, Arizona Police Department. The database, comprised of 50,203 entries, describes burglaries and thefts in Mesa, Arizona occurring during the period of 2004 to 2009. A kernel density estimate of the distribution of ranges within the empirical data is shown in Figure 2.3. For the crimes in this dataset, the exact
time and date of occurrence not available, so the department has recorded a range within which the crime could have occurred. In practice, this is a common challenge confronting end-users given that most crimes of this nature occur while the victim is away from home, either at work, running errands, or on vacation. To conduct any analyses on the data however, the methods here require an analyst to select a single date and time to serve as a proxy for the true time. This proxy could be when the crime was reported or the start or end date of the range when the crime is suspected to have occurred. To model the temporal inaccuracies in the context of the simulations here, the time stamp associated with each event is offset randomly within a range drawn from this empirical distribution. Note, because the purpose is to explore the effect of uncertainty, any entries without this range (i.e. zeros) have been omitted. The offset timestamp is then used as the reported temporal coordinates associated with the event.
2.5 Results

*Experiment 1*

The effects of introducing perturbations to patterns which demonstrate space-time interaction are explored here. Original patterns, exhibiting space-time interaction were perturbed according to the above-described parameters and the resulting series of perturbed patterns were then tested for interaction using the various tests. Rejection frequencies were recorded across combinations of the various parameters (i.e. test, perturbation and degree of original interaction) and the results are presented below in Table 2.1. Given that the original patterns exhibit significant interaction at the respective \( \alpha \) levels, in the absence of any effect from the perturbations one would expect the perturbed versions to continue to reject the null hypothesis of no space-time interaction, that is report rejection frequencies of close to 100%. Inspection of the rejection frequencies listed in Table 2.1 show that this is often not the case. It appears that these perturbations affect the ability of the tests to identify the interaction present in the original patterns, in some cases severely (e.g. the numerous instances rejection frequency falls below 10%).

Given the apparent impact of these perturbations on the detection of interaction, the question then becomes, are there any trends within these observed effects with respect to the impact on different tests of the various types and degrees of perturbations? To investigate this, Kendall’s coefficient of concordance is employed. Also referred to as Kendall’s \( W \), this metric quantifies the degree of agreement in the ranking of \( n \) objects amongst a group of \( m \) judges (Kendall and Smith, 1939). Values for \( W \) range from 0, indicating complete randomness amongst the rankings of the judges, to 1, signaling unanimous agreement among them. The method is used here to evaluate the following questions:
Table 2.1: Rejection rates associated with perturbed versions of original patterns exhibiting differing degrees of interaction: high (2.1a), medium (2.1b) and low (2.1c).

1. Is there order in the degree to which the three tests are affected by the perturbations?

For example, is one test more consistently affected than another? In applying Kendall’s $\mathcal{W}$ to answer this question, the seven different perturbations will act as judges and rank the three tests, giving the highest rank to the test with the highest rejection frequency (meaning it was least impacted by the perturbations). The null
hypothesis here assumes there is no order to the degree the tests are affected (using the perturbations as judges) beyond what would be expected under a condition of randomness.

2. Is there order in the impacts of the various perturbation types on rejection rates across tests? In calculating $W$ here, the tests will act as judges and rank the different perturbation types (temporal, spatial and combinations thereof). The tests will give their highest rank to the perturbation type with the highest rejection frequency (i.e. that which is affected less by the perturbations). The specific null hypothesis here is that there is no order to the rankings of the different perturbation types across the tests (i.e. the tests serve as judges) beyond what would be expected under a condition of randomness.

3. Is there order in the impacts of the different degrees of spatial perturbation on rejection rates across tests? Here, again, the tests act as judges and rank the different levels of spatial perturbation (i.e. high, medium and low). Again, the highest rank will be awarded by each test to the perturbation level with the highest rejection frequency. Here the null hypothesis is that there is no order to the rankings of the different degrees of spatial perturbation across the tests (again, the tests are serving as judges) beyond what would be expected under a condition of randomness.

These explorations and calculations of $W$ are conducted for each of the degrees of original space-time interaction: low (where the $p$-value associated with the original pattern is approximately 0.10), moderate (original $p$-value $\approx$ 0.05) and high (original $p$-value $\approx$ 0.01). Here, rather than assess the significance of the test statistics assuming the values are $\chi^2$ distributed with $n − 1$ degrees of freedom, a permutation based approach is employed because of the small number of objects being judged. The permutation approach was
shown by Legendre (2005) to more accurately assess the significance of the statistic for situations such as the one here where \( n \) (the number of objects being rated) is small.

For the first question, the values for \( W \) (shown in Table 2.2) indicate that although there is not unanimous agreement between the tests, there is a relationship in the rankings of the differing effects of the perturbations on the tests beyond what would be expected due to randomness. Notably the Jacquez is the least affected, the Mantel the most, and the Knox results somewhere in between. The values of \( W \) show significantly more agreement than a random ranking when a high and low degree of space-time interaction was present in the original patterns. However, the value for \( W \) in this case was not statistically significant across the perturbed versions of patterns that exhibited a moderate degree of space-time interaction in the original pattern.

<table>
<thead>
<tr>
<th>original p-value</th>
<th>( W )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.633**</td>
</tr>
<tr>
<td>0.05</td>
<td>0.387</td>
</tr>
<tr>
<td>0.10</td>
<td>0.633**</td>
</tr>
</tbody>
</table>

Table 2.2: Values of Kendall’s coefficient of concordance (\( W \)) using perturbations as judges and tests as objects for the results found in Table 2.1. Symbols correspond to significance at the following \( \alpha \) levels: ** = 0.01, * = 0.05

For the second question, values for \( W \) (shown in Table 2.3) again indicate less than unanimous agreement between the tests as judges in their rankings of the effects of the various perturbations on rejection rates. While the agreement is imperfect, trends are again present. For the original patterns which exhibited high and moderate space-time interaction, there is a significant agreement among the tests that the temporal and combined perturbations most severely impact the rejection rates of the tests. The spatial perturbations lead to less of a decrease in rejection rates and are consistently ranked higher across the tests. This trend is also observed for the original pattern exhibiting the lower degree of space-time interaction although it is not determined to be significant.
Finally, values for $W$ are shown in Table 2.4 in response to the third question, i.e. whether the varying degrees of spatial perturbation affect rejection rates regularly across the tests. The values for $W$ here show discordance. Their associated $p$-values show this is to a higher degree than observed in the previous two questions as they fail to reject the null hypothesis that the results are more ordered than would be expected due to randomness. This indicates there is little order in the effect of the different degrees of spatial perturbation across the tests.

<table>
<thead>
<tr>
<th>original $p$-value</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.111</td>
</tr>
<tr>
<td>0.05</td>
<td>0.778</td>
</tr>
<tr>
<td>0.10</td>
<td>0.111</td>
</tr>
</tbody>
</table>

Table 2.4: Values of Kendall’s coefficient of concordance ($W$) using tests as judges and perturbations as objects for the results found in Table 2.1. Symbols correspond to significance at the following $\alpha$ levels: ** = 0.01, * = 0.05

**Experiment 2**

The results of the second experiment are presented in Table 2.5. Here, 1,000 completely random space-time event patterns (i.e. they fail to reject the null hypothesis of space-time interaction across all the tests) were perturbed using the same seven different schemes employed in Experiment 1. The perturbed versions of the patterns were then tested for
interaction using the three tests. Rejection rates were recorded for \( \alpha \) values of 0.01, 0.05 and 0.10. These results are presented below.

<table>
<thead>
<tr>
<th></th>
<th>Knox</th>
<th>Mantel</th>
<th>Jacquez</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 m</td>
<td>0.005*</td>
<td>0.008*</td>
<td>0.003</td>
</tr>
<tr>
<td>100 m</td>
<td>0.002</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>200 m</td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) \( \alpha = 0.01, \sigma = 0.003 \)

<table>
<thead>
<tr>
<th></th>
<th>Knox</th>
<th>Mantel</th>
<th>Jacquez</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 m</td>
<td>0.027</td>
<td>0.041*</td>
<td>0.030</td>
</tr>
<tr>
<td>100 m</td>
<td>0.022</td>
<td>0.046*</td>
<td>0.022</td>
</tr>
<tr>
<td>200 m</td>
<td>0.028</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) \( \alpha = 0.05, \sigma = 0.007 \)

<table>
<thead>
<tr>
<th></th>
<th>Knox</th>
<th>Mantel</th>
<th>Jacquez</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 m</td>
<td>0.060</td>
<td>0.096*</td>
<td>0.059</td>
</tr>
<tr>
<td>100 m</td>
<td>0.050</td>
<td>0.090*</td>
<td>0.057</td>
</tr>
<tr>
<td>200 m</td>
<td>0.056</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) \( \alpha = 0.10, \sigma = 0.009 \)

Table 2.5: Observed rejection rates where \( \alpha = 0.01 \) (2.5a), \( \alpha = 0.05 \) (2.5b) and \( \alpha = 0.10 \) (2.5c) for perturbed versions of completely random patterns. Asterisks indicate that the proportion of rejections is within two standard deviations of the nominal \( \alpha \) used to establish the rejection rate (i.e. 0.01, 0.05 or 0.10) according to a test of proportions.

Inspection of these results show that rejection rates for the perturbed versions of these data were often higher than those for the original data (i.e. 0). Perturbations affecting the temporal dimension of the patterns appear to yield a greater frequency of rejection. The results of the Mantel test are especially affected by these temporal
perturbations. In these instances, the rejection rates for the Mantel test yield rejection rates that are not significantly different from the nominal $\alpha$. Overall these rejection rates show the perturbations introduced to the original non-interacting data led to an increase in the number of patterns which rejected the null hypothesis of space-time interaction. Effectively, this finding shows that there are instances where the introduction of perturbations can result in the generation of artificial space-time interaction where previously none existed.

2.6 Discussion and Conclusion

The results of these experiments illustrate that when tested for space-time interaction, datasets of degraded quality are likely to yield results different from those obtained using more accurate versions of the same data. Experiment 2 showed that when perturbations are introduced to non-interacting patterns, the patterns may subsequently demonstrate evidence of significant interaction according to these tests. The results of Experiment 1, meanwhile, revealed that these tests, when applied to perturbed versions of datasets which originally exhibited interaction, often failed to reject the null. This indicates a decrease in the ability of the tests to identify the presence of interaction under these conditions. Given the extent of inaccuracy and uncertainty present in data commonly employed in applied studies and the relatively conservative degree imposed here these findings should be a consideration for practitioners.

Surveying the results of these experiments, several additional points should be made. Generally speaking, it appears from the results of both experiments that temporal inaccuracies are more problematic than spatial in terms of obscuring the presence of space-time interaction across all the tests. This, however, is most certainly the result of choosing a more conservative estimate of the spatial inaccuracies relative to the temporal inaccuracies. Consider that for Experiment 1 the means of the spatial perturbations (i.e.

---

8Significance was established using an $\alpha$ of 0.05.
50 m, 100 m, and 200 m) correspond to offsets approximately 0.3%, 0.6% and 1.2% of the average diameter of the study area (approximately 17 km) while the average range drawn from the empirical data for use in creating the temporal perturbations is 1.16 days or 3.9% of the temporal span of the study. This discrepancy in the relative sizes of these offsets is the reason that the temporal inaccuracies appear to have an outsize effect on the rejection rates of the tests. This can be said with certainty due to the mathematically identical roles played by the spatial and temporal adjacency or distance matrices employed in the calculation of these metrics. What varies here is the relative size of the spatial and temporal perturbations to the extent of the study area and period of interest across the respective dimensions. On that note, it is surprising that concordance was not observed in the results of Experiment 1 between the degree of spatial perturbation introduced and the rejection rates. This may be a result of the small size of all of the spatial perturbations relative to the larger pattern. Comparing across tests, it appears that the results of the Mantel test are the most susceptible to being distorted by these inaccuracies, most notably those involving the temporal dimension. In fact, across most of the perturbations where the temporal dimension was affected (i.e. for straight temporal perturbations or combinations of spatial and temporal) the Mantel results essentially devolve to randomness (across both experiments) and show no indication of the original space-time interaction in Experiment 1. The rejection rates are almost in line with the user-defined values for $\alpha$.

Linking these observations back to the formulation of the tests, remember the Knox and Jacquez tests are simply element-wise products of binary adjacency matrices. As long as the perturbations introduced do not upset the contents of these matrices too drastically, the resulting product will be similar to the original. Of course, the question is, how can you quantify what is “too drastically?” This is, of course, a characteristic of the initial arrangement of spatio-temporal distribution of events on the original landscape and
within the time period of interest. For example, for patterns where events are similar
distances away from each other and close to the thresholds defining adjacency for each of
the respective tests, perturbations are likely to have a greater effect than they would if
there was greater distance between the events across the dimensions. Consider two events,
A and B, which neighbor a third, X, in space. Events A and B are 499 and 501 meters away
from X, respectively. If a Knox test were employed here with a spatial threshold distance
of 500 m, only one of the events should fall within the threshold. However, if they are
perturbed spatially, both could fall within the threshold, both could be excluded, things
could stay as they are, or B could be included but A excluded. This arrangement is quite
vulnerable to perturbation. Whereas if the distance between the points and the threshold
were greater, it would be less likely that they would be pushed into our out of the range of
the threshold. In the latter case, the pattern itself would be less vulnerable to
perturbations. The same situation could be imagined for a pattern analyzed by the Jacquez
test, in that instance however, rather than be close to a specific distance threshold if events
A and B were a similar distance away from event X, perturbations may confuse which
order nearest neighbor each event is. Owing to these reasons it is impossible to
mathematically determine the exact vulnerability of these methods to a predetermined
degree of inaccuracy or uncertainty due to the fact that it depends as much on the original
pattern as it does the method used to test for interaction.

Although the results presented here are dependent upon the data employed and the
subjective choices made in modeling the inaccuracies, the take-home message, that these
perturbations impact the ability of these tests to correctly identify interaction, should not
be affected. In challenging the conclusion presented here, one may find fault with the
choice of parameters for modeling the inaccuracies introduced here; however, care was
taken to introduce a conservative degree of these inaccuracies based on levels reported in
the literature. The exception to this, of course, was the choice to base temporal
inaccuracies on an empirical distribution due to a lack of alternative sources. However, few should find fault in this approach given that it is grounded in empirical data drawn from a large sample ($n > 50,000$). Further investigations into the effects of these data deficiencies on these tests could explore the effect that alternative perturbation choices may have on the results (i.e. how different error distributions or parameter values may affect the results). Additionally, work should continue investigating the sensitivity of the results to differing degrees of perturbations. It may be useful to conduct this work in a relative context: that is, what effect do spatial or temporal perturbations amounting to 1%, 5%, 10% etc. of the extent of the study area and time period of interest have on the results? Ultimately, this will be of greater relevance than the absolute metrics of perturbation employed here.

The findings of this study, while specific to global tests of space-time interaction, also beg the question, how are local metrics of space-time interaction affected by similar problems? This is an area ripe for further research. Given the findings of Burra et al. (2002) that local measures of spatial clustering are more susceptible to positional inaccuracies than global measures, further work needs to be undertaken to explore the effect of these problems on the various flavors of the space-time scan statistic (i.e. Kulldorff et al., 1998, 2005; Takahashi et al., 2008) as the impact may be more severe than on the global tests explored here. This is a topic taken up in the next chapter.
The space-time permutation scan statistic, introduced by Kulldorff et al. (2005), is used to identify clusters, or hotspots, of space-time interaction within patterns of spatio-temporal events. In certain contexts (e.g., when analyzing cases of disease or incidents of crime), such clusters are important to identify as they may indicate certain data generating processes or point to emergent trends (Tango, 2010). A variety of metrics have been put forth to identify space-time interaction both globally (e.g., Knox, 1964; Mantel, 1967; Diggle et al., 1995; Jacquez, 1996) and locally (e.g., Kulldorff et al., 1998; Takahashi et al., 2008). The space-time permutation scan statistic (henceforth, STPSS) is among the latter and is most relevant for identifying such patterning when information pertaining to the distribution and dynamics of the underlying background population from which events are drawn is unavailable. The method has been utilized widely in practice, thanks, in part, to its implementation within the SaTScan software (Kulldorff, 2010). It has been employed to investigate spatio-temporal distributions of disease both prospectively (Kulldorff et al., 2005; Haas et al., 2011; Hyder et al., 2011) and retrospectively (Gaudart et al., 2006; Recuenco et al., 2007; Ward et al., 2008; Cooper et al., 2008; Fischer et al., 2008; McNally et al., 2009; Jin-feng et al., 2011; Poljak et al., 2010; Ducheyne et al., 2011) and has also been used retrospectively to analyze distributions of wildlife sightings (Webb et al., 2008; Duffy, 2010), wildfires (Tuia et al., 2008) and violent events (O’Loughlin et al., 2010; O’Loughlin and Witmer, 2010).

In spite of growing use of the STPSS, there has been no consideration of the impact inaccurate or uncertain input data may have on its results. This absence is
troubling given the pervasiveness of such data deficiencies, especially in the context of geographic information (Goodchild and Gopal, 1989; Unwin, 1995; Zhang and Goodchild, 2002) where a variety of studies which have demonstrated these deficiencies to have a disconcerting impact on the results of spatial (Burra et al., 2002; Kwan et al., 2004; Ozonoff et al., 2007; DeLuca and Kanaroglou, 2008; Mazumdar et al., 2008; Zinszer et al., 2010) and spatio-temporal analyses (Jacquez and Waller, 2000). This study explores the possible consequences of deficiencies in the spatial and temporal accuracy of the input data on results of the STPSS. Specifically, this study endeavors to determine if a commonly encountered degree of these deficiencies is enough to prevent the method from successfully identifying hotspots of space-time interaction. Or, alternatively, from a practical perspective, will practitioners employing this method be misled by results affected by less than perfect input data?

A series of simulation experiments are employed in this pursuit, using both synthetic and real-world data. These experiments reveal the results of the STPSS to be relatively robust in the presence of the introduced inaccuracies. While the method is still affected by the deficiencies, their impact on results is less than expected based on the findings of previous research into the effect of such problems on global metrics of space-time interaction (i.e. Chapter 2). The results of this work suggest the STPSS to be a highly versatile tool for investigations concerned with identifying local space-time interaction, even in the face of common data deficiencies.

The paper proceeds as follows. Section 3.2 provides technical background on the STPSS as well as a brief overview of data quality deficiencies commonly encountered in spatio-temporal datasets. Section 3.3 then describes the simulation experiments carried out as part of this study. Section 3.4 reports the results of those experiments while Section 3.5 discusses the findings and offers concluding remarks.
3.2 Background

*Space-Time Permutation Scan Statistic*

Part of a broader family of spatial and space-time scan statistics (see Kulldorff, 1997; Kulldorff et al., 1998, 2007; Takahashi et al., 2008), the STPSS identifies the location and size of likely hotspots (or coolspots) of events in space and time and tests the significance of those concentrations using a Monte Carlo permutation approach. To calculate the statistic, the study area and time period of interest is first subdivided into areas \((s)\) and time periods \((t)\) within which the observed number of events of interest is tallied. The total number of observed events \((N)\) can be calculated as the sum of events observed in each of these areas across all times as shown in Equation 3.1.

\[
N = \sum_s \sum_t n_{st}
\]  

The expected number of cases in each area and time period (i.e. \(\mu_{st}\)) is calculated by conditioning on the observed marginals as shown in Equation 3.2. The STPSS assumes the function responsible for the generation of events operates uniformly across all time periods and areal subdivisions (Kulldorff et al., 2005). This is in contrast to other similar methods such as the cylindrical and flexibly shaped space-time scan statistics which assume spatial and temporal heterogeneity in the data generating process.

\[
\mu_{st} = \frac{1}{N} \left( \frac{\sum_s n_{st}}{\sum t} \right) \left( \frac{\sum t n_{st}}{\sum t} \right)
\]  

Local concentrations of space-time interaction are identified using a cylindrical search window that moves methodically throughout the study area and time period of interest. The radius and height of the cylinder, which correspond to distances in space and
time, respectively, vary as the cylinder moves across the study area and time period of interest. The number of events observed within the cylinder for all size/location/time combinations is compared to the number expected. The space-time permutation scan then maximizes the Poisson likelihood function described in Equation 3.3 across all cylinder radii, heights and starting locations to identify a most likely cluster (MLC) and possible secondary clusters. Pseudo-significance of the identified clusters is established using Monte Carlo hypothesis testing. For this, permuted versions of the original dataset are created by shuffling the temporal coordinates of the pattern and reassigning them to the original spatial locations. Most likely clusters are identified in these permuted patterns in the same manner as the original data and their associated likelihood ratios are recorded. Pseudo significance of clusters identified in the observed data is established by ranking their associated likelihood ratios within the distribution of likelihood ratios generated from analyzing the permuted datasets.

\[
\left( \frac{c}{E[c]} \right)^c \left( \frac{C-c}{C-E[c]} \right)^{C-c} I \tag{3.3}
\]

When calculating the likelihood function, \( C \) is the total count of cases, \( c \) is the count of observed cases within the scanning cylinder, and \( E[c] \) is expected number of observed cases within the cylinder based on the expectation of spatio-temporal randomness. Meanwhile, \( I \) is an indicator function denoting a higher or lower than expected number of cases within the scanning window. When searching for areas of high concentration, this assumes a value of 1 when the cylinder has a greater number of cases than expected and 0 otherwise. The opposite is true when the method is employed to search for areas and times with a lower than expected number of cases (i.e. cool spots). Due to its inability to incorporate information on the dynamics of the background population, users must be aware that the method may erroneously identify clusters due to
spatial and temporal variation in the underlying population from which events are drawn (Kulldorff et al., 2005). Where this is a potential problem and the necessary data are available, the more relevant cylindrical (Kulldorff, 1997; Kulldorff et al., 1998) and flexible (Takahashi et al., 2008) space-time scans should be employed as they incorporate this knowledge directly.

As implemented in the SaTScan software (Kulldorff, 2010), the results of the STPSS consist of a set of identified likely clusters and their associated parameters. For each cluster these parameters include the spatial coordinates of its center, its radius and temporal duration, a list of events included in the cluster, as well as the associated test statistic (generated using Equation 3.3) and a pseudo \( p \)-value. A most likely cluster (MLC) is identified as the cluster with the lowest pseudo \( p \)-value. In addition, a series of possible secondary clusters are also identified.

**Data Quality Deficiencies**

While the specific nature of any inaccuracies or uncertainties associated with the input data analyzed by the STPSS depends on the field of study in which it is applied, generally speaking, such problems are related to the geographic coordinates (i.e. the \( x \) and \( y \) coordinates of events), their associated time stamps (i.e. \( t \)) and the completeness of the dataset. Common problems encountered in spatio-temporal data include inaccurate or imprecise recording of the locations and times of events as well as under-reporting of the events. Additionally, uncertainty may result when the true locations and/or times of events are unknown and/or the completeness of the dataset under examination is questionable.

Individually and collectively, such deficiencies in the quality of input data have been shown to degrade the integrity of results for spatial and spatio-temporal analyses (Jacquez and Waller, 2000; Burra et al., 2002; Ratcliffe, 2004; DeLuca and Kanaroglou, 1

1See Chapter 4 of this dissertation provides an extended discussion of this phenomenon.
Spatial Inaccuracies

Common sources of deficiencies in the location information associated with spatio-temporal event data include inaccurate geocoding, the application of privacy masks (i.e. aggregation to coarser scales or shuffling of locations), and uncertainty pertaining to latency and mobility. The consequences and extent of these problems on spatial analyses are well documented and the relevant literature is discussed below.\(^2\) The effect of these problems on spatio-temporal analyses have been investigated to a far lesser degree; however, existing studies on this topic are covered here as well.

Inaccuracies in spatial event data due to the geocoding process (i.e. matching an address or other locational description to absolute geographic coordinates) are understood to be widespread in data created in this manner (Goldberg et al., 2007; Zandbergen and Hart, 2009). The severity of the inaccuracies in geocoded data varies based on the quality of the underlying spatial data used in the geocoding process (Rushton et al., 2006; Whitsel et al., 2006; Zhan et al., 2006; Zandbergen, 2008; Zandbergen et al., 2011; Zandbergen, 2011) as well as the density of addresses in the vicinity of the geocoded locations (Bonner et al., 2003; Cayo and Talbot, 2003; Ward et al., 2005; Kravets and Hadden, 2007). The detrimental impact of inaccurate geocoding on subsequent spatial analyses has been

\(^2\)This review is based on the more extensive treatment of these topics provided in Chapter 2.

\(^3\)Due to the relevance of uncertainty stemming from latency and mobility to the temporal dimension of the data, this topic will be discussed in Section 3.2.
demonstrated by a number of studies. For example, Burra et al. (2002) showed that geocoding errors affecting even a small number of observations (in their study, only 1% of the original data) impacted the results of analyses for local metrics of spatial autocorrelation. DeLuca and Kanaroglou (2008) observed variation in results of Kulldorff’s spatial scan statistic, kernel density estimation and bivariate K functions when different geocoding methods were employed to generate the raw data analyzed by the metrics. Mazumdar et al. (2008) demonstrated a decreased ability to recover relationships between environmental exposures and health outcome data as geocoding accuracy declined. Zinszer et al. (2010) illustrated that moderate amounts of geocoding errors (affecting only 10% of records) were enough to modify disease distribution maps created using kernel density estimation. In a spatio-temporal context, Chapter 2 of this dissertation showed that a conservative degree of spatial inaccuracy in the form of simulated geocoding errors was capable of severely affecting the results of global tests of space-time interaction.

In addition to those introduced unintentionally via the geocoding process, spatial inaccuracies may also be introduced into spatial data intentionally to mask identity and preserve individual privacy (Armstrong et al., 1999; Fefferman et al., 2005; VanWey et al., 2005). Such inaccuracies are common in the context of health and crime data where the confidentiality of patients and victims (or offenders) is required. A common approach to the masking of locations is to aggregate the data to larger areal units (Armstrong et al., 1999; Fefferman et al., 2005). This approach, however, can yield different results than would be observed if the data were analyzed at the original level of spatial support (Jacquez and Waller, 2000; Ozonoff et al., 2007; Jeffery et al., 2009). Additionally, errors

---

4This practice brings to the fore a concern often confronted in the practice of spatial analysis: the modifiable areal unit problem (MAUP). The MAUP is composed of two distinct problems: (1) the scale problem and (2) the aggregation problem (Openshaw, 1984). The scale problem, of primary concern in this instance, “is the variation in results that can often be obtained when data for one set of areal units are progressively aggregated into fewer and larger units for analysis” (Openshaw, 1984, pg. 8). The aggregation problem yields a similar variation in results, however the number of areal units analyzed is held constant, but alternative
in the original spatial coordinates may result in the observations being aggregated to the wrong areal unit, further exacerbating such problems (Krieger et al., 2001; Ratcliffe, 2001; Kravets and Hadden, 2007). As an alternative to data aggregation, the privacy of individual events may be protected by assigning events to a new randomly generated location that falls within some specified radius of the original location (Armstrong et al., 1999; Wieland et al., 2008). This perturbation approach has also been demonstrated to negatively affect the results of subsequent analyses in a manner proportional to the size of the radius (Kwan et al., 2004).

Temporal Inaccuracies

In spite of being equally relevant in terms of spatio-temporal analyses, inaccuracies in the temporal dimension of spatio-temporal data have received far less attention in the literature than their spatial counterparts. Such temporal inaccuracies encountered in event data commonly stem from the problems of latency and uncertainty.

The former is especially relevant to studies exploring the distribution of health and disease (Jacquez, 2004). In this context, the period of time between an initial infection or exposure and the onset of symptoms or eventual diagnosis can, for certain diseases, be on the order of years or decades. However, most methods for analyzing spatio-temporal patterns (including the STPSS) require the specification of a single time (and place) where the event occurred, rather than incorporate the information available in a space-time path (Hägerstrand, 1970; Miller, 2007) or employ an aoristic approach (Ratcliffe and McCullagh, 1998). This, of course, relates to the discussion above on spatial inaccuracies, as during this time individuals may be mobile and it may be virtually impossible to assign a single discrete location to the disease case. This forced discretization in turn introduces definitions for the areal units are employed. As numerous studies (including those mentioned above) have shown, the consequences of the MAUP for analysis are potentially profound and the nature of the findings across levels of aggregation can be completely contradictory (Openshaw, 1984).
errors into the analysis as the phenomenon cannot be accurately represented using a single point in space or time.

There is also the more general problem of uncertainty surrounding when an event that can be represented as a discrete event actually happened. A classic example, often offered, is that of a burglary event that occurs while the victim is away (Ratcliffe, 2002). For all practical purposes, the burglary can be represented as a discrete event in space and time, however, given that the victim was away, it is often unknown exactly when the crime occurred. The question remains: what should be used as the temporal coordinate of the burglary for analytical purposes? Should it correspond to the date and time the victim left and their home was untouched? Should it correspond to the date they discovered and reported the burglary? Or should it be some average of the two? This question is addressed by Ratcliffe’s work on aoristic analysis (2000; 2002) who advocates that the entire time span should be used. This of course, is often not the approach employed in practical analyses. The only study which explicitly investigates the consequences of this forced discretization in the context of spatio-temporal analysis is Chapter 2 of this dissertation which examined the effect of temporal uncertainty on tests of global space-time interaction. The study demonstrated that uncertainty in the temporal dimension of the input data can greatly distort the results of analyses, in some cases completely obscuring patterns of space-time interaction where they existed and in others creating them where they did not exist.

3.3 Methods

To explore the effect of these commonly encountered data deficiencies on the results of the STPSS, this study employs two approaches. First, an experiment is undertaken in which a series of synthetic event patterns, exhibiting space-time interaction, are generated on a hypothetical landscape. These patterns are then perturbed to varying degrees by
introducing spatial and temporal inaccuracies to the data. The parameters associated with these perturbations are in line with what practitioners may encounter using real-world data and are based on estimates found in the existing literature or empirical observations. The effect of these perturbations on STPSS analyses are then assessed. The second approach, rather than relying on synthetic data, employs an observed pattern of criminal events for the analysis. The pattern of criminal events is perturbed in a manner similar to the simulated patterns above and the effect on the results of the STPSS is then assessed. These experiments provide a window into understanding how commonly encountered levels of spatial and temporal uncertainty and inaccuracy may affect the results of STPSS analyses. The question of primary interest addressed by these experiments is will the STPSS continue to identify the hotspots of interaction that were found in the original datasets even after the events in those datasets are perturbed? The specifics of the different approaches to investigating this question are described in greater depth below.

*Synthetic Data*

For the first experiment, three synthetic patterns are generated on a hypothetical landscape. The study area measures 10 km square and the duration of the study period of interest is 100 days. Each of the original patterns generated within this space-time window include a background population of 200 events randomly distributed in space and time and two spatio-temporal hotspots: Cluster 1, in the northeast quadrant, late in the study period (seeded with 30 events) and Cluster 2, a smaller concentration in the southeast quadrant, early in the study period (seeded with 20 events). The background event population was designed to mimic a random distribution of events across a landscape, for example crimes or cases of disease, amongst a uniformly distributed population susceptible to the events. The hotspots, meanwhile, simulate concentrations of those events in space and time higher than expected under conditions of randomness and uniform probability. These could be due to some data generating process of interest (in the context of disease cases it could
mean a localized outbreak of an infectious or viral condition or exposure to some transient hazard, while in terms of criminal events it could indicate a spree) or they may simply be the result of a higher than expected concentration of individuals susceptible to the particular condition or event in space and time (due to population movement to a particular location at a certain time, for example).

The hotspots here are simulated independently of the background population by generating events surrounding two seed locations in space and time. The seed point for Cluster 1 is located at coordinates \((7.5, 7.5)\) in space and at day 55 of the study period while the seed point for Cluster 2 is located at \((2.5, 2.5)\) in space and day 10 of the study period. The events composing the clusters are generated by drawing coordinates randomly from normal distributions with a mean corresponding to the coordinates of the seed point in the respective dimension. The choice to draw cluster cases from a normal distribution was made based on a similar approach employed by Maciejewski et al. (2009) as a default setting in their software tool used to generate synthetic syndromic surveillance data. The spatial intensity of the simulated space-time hotspots is varied in each of the three original patterns by adjusting the standard deviation associated with the distributions. The standard deviations used to generate the spatial coordinates for the events in hotspots of the three patterns are 500 m, 1,000 m, and 1,500 m, respectively. By generating patterns with clusters of various sizes, the experiment will provide insight into the impact of perturbations on clusters of various sizes. The standard deviation associated with the temporal dimension is held constant across all three patterns at 10 days. The intensities of events across space and time for the three patterns is illustrated using space-time cubes in Figure 3.1. The different perspectives of the pattern offered by the cubes illustrate the locations in space and time of the two simulated spatio-temporal hotspots.

Viewing these cubes, the change in the size, shape and intensity of the hotspots is apparent when the different values are employed for the spatial standard deviation \((\sigma)\).
\( \sigma = 500 \text{ m} \)
\( \sigma = 1000 \text{ m} \)
\( \sigma = 1500 \text{ m} \)

Figure 3.1: Intensity of the three simulated event patterns. Each space-time cube portrays the intensity of events generated across space and time for the three patterns. The left most cube corresponds to the high intensity cluster pattern (where cluster events are concentrated in a smaller area) and the right-most cube corresponds to the lowest intensity pattern. The top of each cube shows an areal view of the study area (i.e. a conventional map), while the sides show the intensity through time (represented as the height of the cube) and across a particular spatial dimension (either \( x \) or \( y \)). Lighter areas indicate a higher intensity of events.

The change is most notable when looking at the top of the cubes (i.e. a conventional map perspective). As the value for \( \sigma \) increases, the radius of the clusters increases. The height of the clusters is maintained (because the temporal standard deviation remains the same) so they become more disc-like rather than spherical in shape when the cube is viewed from the side. The simulated event patterns were then analyzed using the STPSS as implemented in SaTScan. For all the generated patterns, the scan identified Cluster 1 as the MLC and Cluster 2 as a secondary cluster with a highly significant \( p \)-value. The specifics of these results are discussed further below in Section 3.4.

With the original patterns simulated and analyzed, the accuracy of the datasets was then degraded based on quality estimates found in the literature (i.e. Cayo and Talbot, 2003). Three degrees of spatial inaccuracies were introduced into each of the datasets. These inaccuracies were introduced by randomly drawing an offset distance from exponential distributions with means of 50, 100, and 200 m, corresponding to low,
medium, and high levels of spatial perturbation. These offsets were designed to provide conservative estimates of the empirically observed positional accuracy rates for geocoded data reported by Cayo and Talbot (2003). A plot of the cumulative distribution functions for offsets associated with these three perturbation schemes are shown in Figure 3.2. The direction associated with the spatial offset was established using a random draw.

![Cumulative distribution functions showing the distribution of offsets associated with each level of perturbation.](image)

Figure 3.2: Cumulative distribution functions showing the distribution of offsets associated with each level of perturbation.

Temporal inaccuracies were introduced by offsetting the temporal coordinates based on a random draw from an empirical distribution of suspected temporal inaccuracies for burglaries and thefts occurring in Mesa, Arizona. This distribution, also employed in Chapter 2, is composed of over 50,000 entries. A kernel density estimation of the suspected ranges of inaccuracy is shown in Figure 2.3. To offset the temporal coordinates, a range is randomly selected from this empirical distribution. The range is then multiplied by a random value drawn from a uniform random distribution on the interval [-1,1] and
the product is added to the original timestamp. The last step is taken to ensure that the resulting offset for the temporal coordinate occurs at a random point within the possible range, rather than consistently at the beginning or end of the possible period. Additionally, any events moved out of the study area or period during the perturbation process were omitted from subsequent analyses.

This methodology was used to create three series of 1,000 degraded alternative versions (one each using means of 50, 100, and 200 m for the spatial offset distances) for each of the original three patterns (nine series total). The perturbed patterns were then individually analyzed using the STPSS implemented in SaTScan. The results reported for the original patterns and the patterns of degraded quality are compared in the Section 3.4.

*Mesa, AZ Burglary Data*

Rather than rely solely on the synthetic data to explore the effect of data quality deficiencies on the STPSS, a second experiment was also carried out employing real-world data. Following a form similar to the one described above, this second experiment differs in that it employs a pattern of burglary events observed in Mesa, Arizona during 2008 as the original event dataset for the experiment. The pattern is a sample of 200 burglaries drawn from the database kept by the Mesa Police Department. The raw data are shown in Figure 3.3. Spatial reference information have been omitted to preserve privacy.

Again, the data were analyzed using SaTScan and the STPSS. The data were then perturbed in a manner identical to the synthetic data so that the spatial and temporal coordinates of the data were affected. Given that these data are empirical, variability in the spatial intensity of the clusters was not used as a parameter in this experiment; however, the degree of perturbation was still varied as in the synthetic datasets. The results of analyses for the original and perturbed data are explored and compared in the following
Figure 3.3: Sample of burglary events occurring in Mesa, Arizona during 2008 employed in the analysis. Additional geographic identifiers have been omitted from the map to preserve privacy.

3.4 Results

The results from these experiments demonstrate the STPSS to be robust to the moderate amount of perturbations introduced into the data here. While trends were observed indicating that more perturbation led to greater variability in results and in turn a greater likelihood of misidentifying the most likely cluster (MLC, i.e. the cluster with the largest likelihood ratio) in the patterns, the STPSS was still often successful in identifying the statistically significant clusters present in the original patterns. The results for both the synthetic and empirical data are explored in greater detail in the sections below.
In exploring the results from the experiments employing the simulated data, the first step is to examine the spatial and temporal distributions of the MLCs identified by the STPSS for both the original and perturbed datasets. These distributions are shown in Figures 3.4 and 3.5, respectively.

Figure 3.4 shows the spatial distributions of MLCs identified in the three original datasets and compares them to the locations of MLCs identified in their perturbed counterparts. Rows in the figure correspond to the three different initial spatial intensities used to generate the simulated hotspots: the top row shows the results for the patterns constructed using a standard deviation ($\sigma$) of 500 m, for the middle row $\sigma = 1,000$ m and for the bottom row $\sigma = 1,500$ m. The three columns, meanwhile, correspond to the different levels of spatial perturbation these original patterns were subjected to. The results in the left-most column are based on data whose spatial coordinates were perturbed based on a draw from an exponential distribution with a mean ($\mu$) of 50 m, for the middle column $\mu = 100$ m and for the right column $\mu = 200$ m. The MLCs identified in the original (unperturbed) datasets are shown as red circles in the figures while the second most likely cluster in each of the original datasets is shown as a green circle. The radius of the circle corresponds to the spatial extent of the identified cluster. Note that these original MLC locations and sizes are different for each row, but identical across the columns in a row, because they show the results for the original datasets of varying spatial intensity. In each of these three original patterns, the MLC identified by the STPSS was in the vicinity of Cluster 1 (in the northeast of the study area) and the second most likely cluster was in the vicinity of Cluster 2 (in the southwest of the study area). Meanwhile, in each of the plots, black circles identify the size and locations of the MLCs identified in the 1,000
perturbed versions of the original patterns. The maps reveal there are only limited instances where these MLCs did not overlap either Cluster 1 or Cluster 2.

Figure 3.4: Plots of MLCs identified by the STPSS. The spatial footprint of the MLCs for the original datasets are shown in red and the secondary cluster with the next lowest p-value is shown in green. MLCs from perturbed versions of the same dataset are shown in black. The intensity of the original clusters decreases from the top down while the intensity of perturbation increases from the left to the right. This layout is followed in subsequent graphics.
These results are viewed from a temporal perspective in Figure 3.5. Again, it is important to note the row and column structure of the subfigures is identical to that of Figure 3.4 where the different rows and columns correspond to the various spatial intensity and perturbation parameters. Here, the temporal extent of the study period serves as the y-axis of the subfigures while the x-axis indexes the perturbed datasets. The time span of the MLC identified for each of the perturbed datasets is plotted as a vertical black line. The height of these lines show the duration of the identified clusters and their associated y-coordinates show the position in time within the study period. Additionally, in each subplot, a horizontal red bar shows the duration and temporal position of the MLC identified for the corresponding observed dataset (i.e. Cluster 1) while a horizontal green bar notes the duration and temporal position of the secondary cluster (Cluster 2). For many of the subplots the red bar is obscured by the vertical black lines, indicating that the MLCs identified in the perturbed patterns overlap quite frequently with the MLC from the original patterns.

Together, the spatial and temporal perspectives of these results show that for the majority of the perturbed patterns, the STPSS identified Cluster 1 as the MLC in spite of the perturbations; although, Cluster 2 was also frequently identified as the MLC even though it was seeded with less events and had a larger initial $p$-value (as can be seen in Table 3.1). Only occasionally were MLCs identified in the perturbed patterns which did not overlap the original Cluster 1 or 2. From these figures, it also appears that MLCs found in the perturbed data were more likely to overlap the original MLC (i.e. Cluster 1) when the level of spatial perturbation was low. This finding is confirmed in Table 3.2, which denotes the percentage of perturbed patterns whose MLC overlaps Cluster 1 in both space and time. Across all initial intensities, this percentage decreases as the level of perturbation increases. Unsurprisingly, in Figures 3.4 and 3.5 there also appears to be a
Figure 3.5: Plots of the duration of MLCs identified with the STPSS. The duration of the MLCs for the original datasets are denoted using horizontal red bars, secondary clusters are shown using green bars. MLCs from perturbed versions of the same dataset are shown as thin black vertical lines.

positive relationship between variability in the spatial and temporal locations of the MLCs in the perturbed data and the degree of perturbation introduced.

Examination of these figures and this table shows that the perturbations appear to hinder the ability of the STPSS to identify Cluster 1 as the MLC (i.e. it does not consistently identify the original MLC as the most likely cluster in the perturbed patterns). The question then becomes, despite this, does the STPSS still identify Cluster 1 as a statistically significant cluster within the perturbed patterns? To investigate this, all
Table 3.1: Pseudo $p$-values as calculated by the STPSS associated with Clusters 1 and 2 for the hotspots of varying intensity.

<table>
<thead>
<tr>
<th>Spatial Offset ($\mu$)</th>
<th>Intensity ($\sigma$)</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500 m</td>
<td>0.000037</td>
<td>0.00011</td>
</tr>
<tr>
<td></td>
<td>1000 m</td>
<td>0.00032</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>1500 m</td>
<td>0.0036</td>
<td>0.063</td>
</tr>
</tbody>
</table>

Table 3.2: Proportion of perturbed patterns whose MLC overlaps Cluster 1 in both space and time.

<table>
<thead>
<tr>
<th>Spatial Offset ($\mu$)</th>
<th>Intensity ($\sigma$)</th>
<th>50 m</th>
<th>100 m</th>
<th>200 m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500 m</td>
<td>0.780</td>
<td>0.611</td>
<td>0.549</td>
</tr>
<tr>
<td></td>
<td>1000 m</td>
<td>0.952</td>
<td>0.902</td>
<td>0.817</td>
</tr>
<tr>
<td></td>
<td>1500 m</td>
<td>0.937</td>
<td>0.842</td>
<td>0.649</td>
</tr>
</tbody>
</table>

clusters identified by the STPSS in each of the perturbed versions of the original patterns were examined (not just the MLC, as above). From the results for each perturbed pattern, the cluster identified by the STPSS which overlaps Cluster 1 the most (across space and time, collectively) was identified. The spatial and temporal distribution of these clusters are plotted in Figures 3.6 and 3.7.

The $p$-values associated with these clusters identified in the perturbed patterns which overlap the original Cluster 1 were also examined. The proportion of clusters identified by the STPSS in the vicinity of Cluster 1 which reject the null hypothesis of no space-time interaction at an $\alpha$ of 0.05 are recorded in Table 3.4. The results show that the majority of the clusters identified in the perturbed data which overlap the original Cluster 1 are, in fact, statistically significant. Although, the percentage of overlapping clusters which are significant decreases with both increased perturbation and increased variance of the initial cluster seed.
Figure 3.6: Plots of the spatial distribution of clusters identified by the STPSS which overlap the MLC identified in the original pattern.

While these clusters do overlap the original Cluster 1 in space and time and for the most part identify significant localized space-time interaction, there is clearly variability in their location and size after perturbation. Although this variability is not entirely unexpected, it is still important to examine the degree of this variability. The location of these clusters is investigated further in Figures 3.8a and 3.8b which plot the distributions of spatial and temporal distances, respectively, from the center of the clusters identified in
the perturbed datasets which overlap the original MLC to the original MLC itself.
Collectively, the graphics show that as the level of spatial perturbation increases, the average distance between the MLC identified in the original data and the clusters in the perturbed data which overlap it increases. The size of the clusters identified in the perturbed datasets was also examined. Figures 3.9a and 3.9b show the size of the perturbed clusters across the spatial and temporal dimensions and compare them to the size of the original Cluster 1. The graphics show that as the degree of spatial perturbation increases, the variability in cluster size also increases.
| Intensity (σ) | 50 m | 100 m | 200 m | 500 m | 1.000 | 1.000 | 1.000 | 1000 m | 1.000 | 0.996 | 0.994 | 1500 m | 0.980 | 0.960 | 0.902 |
|--------------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|

Table 3.3: Proportion of perturbed patterns overlapping the original Cluster 1 in both space and time with p-values below 0.05.

Figure 3.8: Plots showing the distributions of spatial and temporal distances between the center of Cluster 1 (identified in the original data) and the center of clusters identified in the perturbed datasets which overlap Cluster 1. Each plot corresponds to a different initial intensity and the distributions across different perturbation levels are represented using different colors in each plot.

Overall, for the synthetic data examined here, these investigations show the STPSS to be highly effective at identifying the original MLC within perturbed versions of the original datasets. While there is some variability to the location and size of the corresponding clusters identified within the perturbed datasets and a decrease in their
(a) Distributions of cluster areas.

(b) Distributions of cluster durations.

Figure 3.9: Plots of the spatial area and temporal durations of clusters identified in the perturbed datasets which overlap the original Cluster 1. The area and duration of the original Cluster 1 is identified as a hashed vertical black line in the figures.

$p$-values, the method was generally successful at identifying the simulated pockets of local space-time interaction in a vast majority of the perturbed patterns.

**Empirical Data**

The results for the simulation experiments based on the Mesa crime data are now explored. Analysis of the original data using the STPSS revealed a single, statistically significant space-time hotspot within the dataset. As such, the impact of perturbations on clusters of different spatial intensities were not explored in this experiment. However, the effect of varying degrees of spatial perturbation and the effect of temporal inaccuracy on the detection of this hotspot were explored. To examine the impact of these introduced
data problems, the spatial and temporal distributions of the MLCs within the original and perturbed data are examined in Figure 3.10 and 3.11, respectively.

\[ \mu = 50 \text{ m} \quad \mu = 100 \text{ m} \quad \mu = 200 \text{ m} \]

Figure 3.10: Plots of MLCs identified within the Mesa crime data using the STPSS. The spatial footprint of the MLC for the original dataset is shown in red. MLCs from perturbed versions of the same dataset are shown in black.

The MLC for the original Mesa dataset are shown in red in Figures 3.10 and 3.11, no other statistically significant (at \( \alpha = 0.05 \)) secondary clusters were identified in the original data. The MLCs identified within the perturbed versions of these datasets are shown on the same figures in black. In contrast to the results for the synthetic data, these initial explorations into the spatial and temporal distribution of the identified MLCs show remarkable stability in both dimensions across the various levels of perturbation: all of the
MLCs identified within the perturbed datasets overlap the MLC identified in the original dataset in both the spatial and temporal dimensions. This stability is likely due to the fact that no other significant hotspots were identified in the original data which could be misconstrued as MLCs by the STPSS in the perturbed data.

While the identified MLCs overlap the cluster identified in the original data, there is still variability in their spatial and temporal coordinates as well as their size and the associated $p$-values. The distribution of spatial and temporal distances between the center of the MLCs identified in the perturbed data and the center of the MLC identified in the
original data are shown in Figure 3.12 while the spatial extent and durations of the MLCs identified in the perturbed data are shown in Figure 3.13. In examining these figures, there is not a pronounced trend, regarding the effect of perturbation, as there was upon further inspection of the results of experiments based on the synthetic data. Although, the figures illustrate clearly that as perturbations are introduced the location and dimensions of the clusters identified in the data will vary from their original forms.

Figure 3.12: Plots showing the distributions of spatial and temporal distances between the center of the MLC identified in the original data and the center of MLCs identified in the perturbed versions of the Mesa dataset. The different perturbation levels are represented using different colors in each plot.

Figure 3.13: Plots of the spatial area and temporal durations of clusters identified in the perturbed datasets which overlap the original MLC in the Mesa data. The area and duration of the original MLC is identified as a hashed vertical black line in the figures.
Turning to the $p$-values associated with these clusters, the MLC identified in the original data was determined to be highly significant (with a $p$-value of 0.000063). When the $p$-values for the MLCs identified in the perturbed data were examined, it was found that they too were, for the most part, highly significant. Table 3.4 below lists the mean of the $p$-values associated with these MLCs and shows the percent of patterns which reject the null hypothesis of no space-time interaction at an $\alpha$ of 0.05. There does not appear to be a strong relationship between level of perturbation and $p$-values for these data. Also, it is noteworthy that only one of the perturbed patterns examined here failed to identify a significant cluster (although it did have a $p$-value of less than 0.10) overlapping the original MLC.

<table>
<thead>
<tr>
<th>Perturbation ($\mu$)</th>
<th>Average $p$-Value</th>
<th>$% &lt; 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 m</td>
<td>0.0008</td>
<td>100.0</td>
</tr>
<tr>
<td>100 m</td>
<td>0.0008</td>
<td>99.9</td>
</tr>
<tr>
<td>200 m</td>
<td>0.0011</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 3.4: Average pseudo $p$-values for MLCs observed in perturbed datasets and percent of patterns rejecting the null hypothesis (at $\alpha = 0.05$) for various levels of perturbation.

3.5 Discussion and Conclusion

While prior studies (i.e. Jacquez and Waller, 2000, and Chapter 2) have shown global tests of space-time interaction to be highly volatile in the face of similar data deficiencies, collectively the findings from these experiments demonstrate a marked departure from this precedent in the case of this local method. Based on the experiments conducted here, the results of the STPSS appear to be quite robust to the moderate degree of the common data problems introduced. While there is an observed negative trend between degree of perturbation and ability to locate the correct MLC, especially within the simulated patterns with multiple significant hotspots, the method was generally successful in identifying the
original MLC as a significant cluster in the results, even though it may not have been determined to have the lowest p-value among clusters found in the perturbed datasets.

Given the findings of Chapter 2, it is surprising how well the STPSS performs in the face of these common perturbations as compared to the global methods for detecting space-time interaction. Previous studies (i.e. Burra et al., 2002) have shown local tests of spatial dependence (specifically the local Moran) to be more susceptible to spatial inaccuracies than their global counterparts. One might have expected a similar relationship between the global and local tests of space-time interaction examined here. However, Chapter 2 of this work employed identical parameters to perturb data in an exploration of the effect of data inaccuracy on global tests of space-time interaction and found the results of those tests were far less likely to identify the presence of the original interaction after these perturbations than the local test considered here. In some cases, the results indicated the patterns were essentially random rather than significantly clustered. How, then, is this local metric of space-time interaction still able to successfully identify the approximate location of hotspots present in the original data when global tests are unsuccessful in identifying interaction? One hypothesis, is that the global tests are far more sensitive to small changes in the arrangement of the spatio-temporal event pattern because of the fact that they are looking for global relationships between the spatial and temporal distances separating pairs of events. As such, their focus is on the pattern as a whole which makes them susceptible to the entire collection of perturbations.

This local metric, however, assesses a large number of subsets of the pattern using scanning windows of various sizes. The use of these variable scanning windows provides a more flexible framework for searching for the local interaction. Connecting to the formal specification of the method, the value for the likelihood ratio for each scanning window is only affected by the perturbations to the events within the window. Or, more specifically, it is only affected when events are moved out of or into its bounds (i.e. when
the value for $c$ is altered in Equation 3.3). For the experiments carried out here, given that the approach of the STPSS is to use a large number of these windows, of varying sizes, it seems likely that scanning the perturbed data will identify at least one window close to the location of the MLC in the original data, in spite of introduced perturbations, with a count of events similar to that seen in the original MLC. This in turn will yield a likelihood ratio identical (or similar) to that of the original cluster. As the results here show, though, it is often in slightly different location in space and time owing to the perturbations. In the face of more dramatic perturbations (relative to the size of the pattern), this may not be the case however. Even if a similar likelihood ratio is reported, however, problems may still arise given that the perturbations may have also affected other subsets within the pattern by pushing events closer together in space and time and thereby increasing their likelihood ratios as well. This may subsequently upset the relative rankings of the likely clusters and subsequently distort the pseudo $p$-values reported by the Monte Carlo permutation process. It is most certainly a result of both of these processes that produces the variability in results that were observed here due to the perturbations.

While this initial study suggests robustness of the STPSS, subsequent work will be needed to further explore this topic and these results in greater depth. It should be noted that while all facets of uncertainty and inaccuracy discussed in the literature review were incorporated into the experiments here (i.e. the data were both spatially and temporally perturbed) only in the case of the spatial perturbation was any sensitivity really explored. In the case of the temporal dimension, changing the perturbation systematically (as in the case of the spatial perturbations) was not an option given the lack of empirical research in this area on which to ground the sensitivity analysis. Further work is needed in this area to assess the accuracy of temporal coordinates in a variety of applied contexts. Alternatively, as suggested in the conclusion of Chapter 2, further work to assess sensitivity of the results of the STPSS to varying degrees of temporal (or spatial) perturbation could employ
a relative approach (for example introducing perturbations of varying size relative to the extent of the pattern or clusters) to defining those perturbations.

Although the results presented here cast a favorable light on the STPSS, care should be taken not to overstate their significance or overestimate the ability of this method to handle inaccuracies and uncertainty. The perturbations imposed on the data employed here were of a conservative nature. It is likely that far less favorable results would be observed if stronger degrees of inaccuracy or uncertainty encountered. Of particular concern may be the use of this method to identify patterns in cases of diseases with long latencies (Jacquez, 2004). Additionally, I would caution against the extension of these findings to other local tests of space-time interaction such as the cylindrical and flexible space-time scans as these have the added parameter of background population to account for. In the case of those methods, potential inaccuracy in accounting for spatially and temporally heterogeneous background populations offers an additional dimension of concern that may warrant further investigation.

In spite of these caveats, this research has shown that in contexts where researchers have reasonable confidence in the spatial and temporal accuracy and precision of their data they should also have confidence in the integrity of the reported results of the STPSS.
A pattern of events exhibits space-time interaction when, generally speaking, pairs of events in that pattern which are close to each other in space are also close to each other in time. Although events within the same pattern may be closer to each other in space or time than would be expected under the null hypothesis of spatio-temporal randomness, space-time interaction occurs only in instances where, generally speaking, there is a positive relationship between the spatial and temporal distances between pairs of events (Kulldorff, 1998; Tango, 2010). Given the specific nature of space-time interaction, methods to establish its presence are necessarily distinct from conventional methods for detecting purely spatial or temporal clustering. Tests of interaction are designed to “detect space-time clustering above and beyond any purely spatial or purely temporal clustering;” meaning, the tests determine if event pairs that are close in space are also close in time (Kulldorff, 1998, pg. 58).

Originally developed within the field of spatial epidemiology, tests of space-time interaction remain popular tools to analyze patterns of disease cases (Kulldorff, 1998; Ward and Carpenter, 2000b; McNally and Colver, 2008; Meliker, 2009; Rogerson and Yamada, 2009; Tango, 2010). In this context, the identification of space-time interaction may indicate an infectious or viral etiology, or the presence of transient, localized hazard exposure (Marshall, 1991; Jacquez, 1996). The Knox test for space-time interaction, for example, has been used extensively to provide evidence in support of a viral etiology of leukemia (Alexander, 1992; Petridou et al., 1996; Kulldorff and Hjalmars, 1999; Bosch and Muñoz, 2002; Zur Hausen, 2009). In addition to the field of spatial epidemiology,
tests of space-time interaction have also been used increasingly in the field of criminology (Knox, 2002; Johnson and Bowers, 2004; Grubesic and Mack, 2008) and ecology (Legendre and Fortin, 1989; Fortin and Gurevitch, 1993; Michener, 1997; Legendre and Fortin, 2010).

One of the most commonly used tests of space-time interaction is the Jacquez k nearest neighbor test (Jacquez, 1996). The test has a number of positive attributes that contribute to its utility and prevalence in the literature. First, as the name implies, it employs a nearest neighbor approach to establishing proximity in space and time instead of defining explicit distance- or time-based thresholds. This unique approach to detecting space-time interaction yields a test that is robust to non-linear associations in space and time and has been demonstrated to be quite powerful (Jacquez, 1996; Tango, 2010). It also reduces the subjectivity associated with parameter selection common to alternative tests (Jacquez, 1996; Grubesic and Mack, 2008). Second, the test and its significance can be computed quickly relative to some other tests. Independent of its flexibility, speed and power, however, the Jacquez test does have three key shortcomings. First, the k nearest neighbor approach discards important information regarding the spatial and temporal scale at which the detected interaction takes place (O’Sullivan and Unwin, 2003). Second, the test is not accompanied by any explicitly spatial or temporal output that can be visualized. Third, recent research has shown the test to be susceptible to population shift bias (Mack et al., 2012). The goal of this chapter is to enhance the Jacquez test by addressing each of these shortcomings, thereby increasing the utility of this test for researchers.

The first enhancement to the test offered here provides supplementary information about the spatial and temporal scales at which events are interacting. This information links the abstract concept of nearest neighbor interaction to tangible spatial and temporal information for each of the events of interest. Second, several methods for visualizing the interaction identified by the nearest neighbor test are presented. Finally, a version of the
Jacquez test is formulated and implemented to account for potential bias resulting from a shift in the underlying population over time. Throughout the chapter, the utility of the modifications and visualization techniques are illustrated by comparing the results of the enhanced Jacquez test to those produced by the Knox test for space-time interaction. These comparisons are demonstrated using a dataset compiled by Williams et al. (1978), which provides location and time of onset for 188 cases of Burkitt’s lymphoma in the West Nile District of Uganda during the period 1961-1975.

4.2 Background on Space-Time Interaction Tests

Before proceeding to a discussion of the enhancements to the Jacquez test developed in this chapter, this section will provide a brief overview of the three other commonly used tests for space-time interaction: the Knox test, Mantel test and space-time $K$ function. For a more complete technical review of these methods, readers are referred to the original citations noted in the text or to the extensive review provided by Tango (2010).¹

The first method formulated to detect global space-time interaction was the Knox test (1964). Developed within the context of epidemiology, the Knox test compares all possible pairs of events within a pattern and evaluates whether or not they fall within critical thresholds for distance in space and in time of each other (Knox, 1964). The Knox test statistic is a count of the number of event pairs that are within both thresholds simultaneously. A formalization of the Knox test is given in Equation 2.1. While there has been work focused on deriving the exact distribution of this test statistic (e.g. Knox, 1964; David and Barton, 1966), most implementations of the test today determine the significance of the statistic using a permutation approach (Aldstadt, 2007). Although widely used in the literature, concerns identified surrounding the Knox test include: the

¹Also, it should be noted that scan-based tests including Kulldorff et al.’s space-time scan statistic (Kulldorff et al., 2005) and Takahashi et al.’s flexible space-time scan statistic (Takahashi et al., 2008) which are concerned with the detection of localized clusters of events in three-dimensional space are not considered here. Tango (2010) distinguishes these from tests of space-time interaction in that the latter have a global focus whereas the scan statistics are interested in finding significant local clusters in space and time.
impact of edge effects (De Smith et al., 2006; Tango, 2010), an inability to detect non-linear interaction (Jacquez, 1996), the subjectivity introduced by the selection of critical distances (Jacquez, 1996; Ward and Carpenter, 2000a; Grubesic and Mack, 2008) and a loss of power associated with a high concentration of points across the study area (De Smith et al., 2006). Numerous alternative formulations have been suggested to deal with one or more of these problems or to adapt the test to different applications (Baker, 1996; Kulldorff and Hjalmars, 1999; Rogerson, 2001; Baker, 2004). For example, Kulldorff and Hjalmars (1999) formulated an unbiased version of the Knox test to account for the issue of population shift bias (i.e., the propensity of the test to identify spurious interaction due to shifts in the underlying population over time and not true interaction resulting from the data generating process responsible for producing the events of interest). This phenomenon will be discussed in greater detail below, in the context of the Jacquez test.

A more generalized version of the Knox test was proposed by Mantel (1967). Although mathematically related to the Knox, this test removes the subjectivity associated with the selection of critical space and time thresholds required by the Knox test, and instead detects interaction by considering the spatial and temporal distances between all pairs of events (Tango, 2010). There are two versions of the Mantel test statistic: unstandardized and standardized. The unstandardized statistic is calculated by computing spatial and temporal distance matrices for the event pattern, multiplying the two matrices together in an element-wise fashion and then summing the elements of the resulting matrix of products. The standardized test statistic is calculated by measuring the correlation between the elements of the spatial and temporal distance matrices. Formalizations of the unstandardized and standardized Mantel test are given in Equations 2.2 and 2.3, respectively. Mantel advocated the addition of a constant to the raw distance matrices to prevent multiplication by zero. Additionally, he prescribed a reciprocal
transformation of the resulting distances to temper the effect of events distant from each other in space and time (Mantel, 1967). Irrespective of the version of the test statistic computed, the resulting statistic is highly dependent upon the selection of these parameters (Jacquez, 1996; Tango, 2010).

A third global test for space-time interaction is the space-time $K$ function proposed by Diggle et al. (1995). The method was formulated as an extension of the spatial $K$ function (Ripley, 1976; Diggle and Chetwynd, 1991). Calculation of the $K$ function in space and time is shown in Equation 4.1. Where $s$ and $t$ are thresholds in space and time within which events are counted. These parameters take on a range of values specified by the user. $R$ is the area of the region studied and $T$ is the overall timespan of the study. $I_s$ and $I_t$ are indicator functions that are equal to 1 when the distance and time between $i$ and $j$ are under the thresholds of $s$ and $t$, respectively, otherwise these values are 0. The terms $w_{ij}$ and $v_{ij}$ are edge correction mechanisms in space and time. Tango (2010) argues, however, that to test the null hypothesis of no space-time interaction such corrections are unnecessary.

\[
\hat{K}(s,t) = \frac{RT}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{I_s(s_{ij})I_t(t_{ij})}{w_{ij}v_{ij}}
\]  

\[
I_s = \begin{cases} 
1, & \text{if } s_{ij} < s \text{ and } i \neq j \\
0, & \text{otherwise}
\end{cases}
\]

\[
I_t = \begin{cases} 
1, & \text{if } t_{ij} < t \text{ and } i \neq j \\
0, & \text{otherwise}
\end{cases}
\]

To use the space-time $K$ function to test for interaction, individual $K$ functions in both space and time, $\hat{K}(s)$ and $\hat{K}(t)$, respectively, must also be calculated. Using these
individual $K$ functions and $\hat{K}(s,t)$, Diggle et al. (1995) outline a variety of metrics to describe possible space-time interaction. The first, $\hat{D}(s,t)$, shown in Equation 4.2, calculates the difference between the product of the individual $K$ functions for space and time and the combined space-time $K$ function. In the absence of space-time interaction the expected value of this difference is zero. Another function, shown in Equation 4.3, standardizes these differences.

\[
\hat{D}(s,t) = \hat{K}(s,t) - \hat{K}(s)\hat{K}(t)
\]

\[
\hat{R}(s,t) = \frac{\hat{D}(s,t)}{\sqrt{\text{Var}(\hat{D}(s,t))}}
\]

These functions are then used in a formal test of space-time interaction. In this test, shown in Equation 4.4, the standardized residuals are summed over a set of spatial ($p$) and temporal ($q$) distances.

\[
U = \sum_p \sum_q R(s_p, t_q)
\]

The permutation method described above for the Knox and Mantel methods is used to derive pseudo-significance for this value. While the space-time $K$ function is recognized as a distinct method to test for space-time interaction, it is closely related to the Knox test. Essentially, the analysis provided by the space-time $K$ function amounts to a collection of Knox tests run across varying spatial and temporal thresholds (Bhopal et al., 1992; Tango, 2010). As noted by Kulldorff and Hjalmars (1999), its power resides somewhere between the Knox and Mantel tests. One of the major issues associated with this test, however, which has perhaps prevented more widespread use of the method, is that it is relatively more computationally burdensome than the Knox and Mantel tests.
4.3 Jacquez Test

As discussed above, each of the aforementioned tests have drawbacks associated with them. In response to these drawbacks, Jacquez (1996) proposed a test based on a nearest-neighbors distance calculation. The test offers two distinct improvements over other tests of space-time interaction: (1) it eliminates the need for the user to specify an absolute threshold distance within which interaction will be detected; (2) it inherently accounts for geographic variation in population density by identifying interaction based on nearest neighbor relationships rather than absolute distance.

Calculation

The test is composed of two statistics: a cumulative measure of interaction, $J_k$, and a $k$-specific measure of interaction, $\Delta J_k$. The cumulative measure locates the $k$ nearest neighbors to a point in both space and time and then tabulates the number of events that are nearest neighbors in both dimensions. This is expressed mathematically in Equation 4.5, where $n = \text{number of cases}; a^s = \text{adjacency in space}; a^t = \text{adjacency in time}.$

$$J_k = \sum_{i=1}^{n} \sum_{j=1}^{n} a^s_{ijk} a^t_{ijk}$$

$$a^s_{ijk} = \begin{cases} 
1, & \text{if event } j \text{ is a } k \text{ nearest neighbor of event } i \text{ in space} \\
0, & \text{otherwise}
\end{cases}$$

$$a^t_{ijk} = \begin{cases} 
1, & \text{if event } j \text{ is a } k \text{ nearest neighbor of event } i \text{ in time} \\
0, & \text{otherwise}
\end{cases}$$
To determine if event $j$ is a $k$ nearest neighbor of event $i$ in a particular dimension (either space or time) a distance matrix, $D$, for that dimension must be calculated. Entries in the distance matrix, $d_{ij}$, denote the distance between all pairs of events in the event pattern. The first nearest neighbor of an event is the closest neighboring event. For values of $k$ larger than 1, the $k$th nearest neighbor is the $k$th closest neighboring event. The set of $k$ nearest neighbors for an event, however, includes the $k$th nearest neighbor and the nearest neighbors associated with all lower orders of $k$ (Jacquez, 1996). For example, the third nearest neighbor of an event of interest $i$ is the third closest event to $i$ while the set of $k = 3$ nearest neighbors of an event include $i$’s first, second and third nearest neighbors.\footnote{For situations where there are numerous $k^{th}$ neighbors of an event, due to the neighbors being equidistant from event $i$, ties are broken selecting one (or more) of the tied events to ensure that the total number of neighbors considered by the statistics remains at $k$.}

The $k$-specific statistic, $\Delta J_k$, is a measure of space-time interaction for $J_k$ in excess of that observed for $J_{k-1}$. This additional metric was formulated because values for the cumulative test statistic ($J_k$) are not independent of one another. This dependence is due to the fact that pairs of events included in smaller values of $k$ are also included in larger values of $k$. Therefore, larger values of $k$ will exhibit increased space-time interaction by virtue of the fact that more space-time case pairs are included in the calculation of the test statistic (Jacquez, 1996). The $\Delta J_k$ statistic, however, is completely independent of other levels of $k$. The formulation of this statistic is given in Equation 4.6.

$$\Delta J_k = J_k - J_{k-1}$$ \hspace{1cm} (4.6)

Jacquez (1996) advocated that significance of both statistics be assessed using a Monte Carlo permutation method similar to that originally advocated for the Mantel test (1967), where the spatial coordinates are fixed, but the temporal coordinates of the data
are permuted and the observed statistics are then compared to the distribution of statistics generated by running the test on the permuted data to generate a pseudo $p$-value.

**Combined Test**

As formulated, the Jacquez test statistics resolve several of the issues mentioned associated with other tests for space-time interaction, however, a remaining problem is the subjectivity related to the selection of an appropriate value for $k$. This shortcoming is overcome by constructing a combined test based on multiple values of $k$ as proposed by Jacquez (1996). The combined test evaluates whether significant interaction between events exists over a range of values for $k$ instead of just at a single value of $k$. The mechanics of the combined test involve calculating either of the test statistics ($J_k$ or $\Delta J_k$) at multiple levels of $k$ and then assessing the combined probability of the results.

This assessment of combined probability needs to account for the problem of multiple testing (Sainani, 2009). Jacquez noted however that the commonly used Bonferroni and Simes adjustments both result in an excessively conservative assessment of combined significance (Jacquez, 1996). To resolve this issue, he proposed a centroid distance method for combining probabilities across multiple levels of $k$. This approach involves converting the results of the test across multiple levels of $k$ to a $1 \times m$ vector $\mathbf{j}$ where the result at each level of $k$ occupies an element in the vector (Jacquez, 1996). These vectors are saved for each of the $N$ permutations in the Monte Carlo procedure. The vectors can then be conceptualized as a cloud of points in $m$-dimensional space. The centroid of this cloud is determined and the distances from the centroid to the points composing the cloud are calculated. Pseudo-significance of the observed $\mathbf{j}$ vector is assessed by tallying the number of points in the cloud that are closer to the centroid than the observed vector. This value, $c$, is then inserted into Equation 4.7 to calculate the pseudo-significance of the combined test results.
\[ p = \frac{c + 1}{N + 1} \]  

(4.7)

Issues and Proposed Solutions

Despite the improvements offered by the Jacquez test relative to the other tests of space-time interaction outlined above, the test has problems of its own. First, although the nearest neighbors approach to detecting interaction has been found effective, it does not provide any indication as to the spatial and temporal scales at which the detected interaction occurs. Second, there have been no efforts to visualize the results in a spatio-temporal context. Finally, the test has been shown to be highly susceptible to population shift bias (Mack et al., 2012). These points are discussed in greater detail in this section. The enhancements outlined by this work address each of these shortcomings.

While there are issues of subjectivity associated with computing space-time interaction tests based on absolute distance thresholds in space and time (Jacquez, 1996; Ward and Carpenter, 2000a; Grubesic and Mack, 2008), completely abandoning real-world linkages in exchange for the relativistic approach offered by nearest neighbors proximity creates different problems. Although the nearest neighbor based Jacquez test effectively detects interaction between events, at what spatial and temporal scales does the interaction occur? As designed, the test does not indicate to the user the real-world spatial and temporal scales at which the interaction is observed. Consider, what does it actually mean in terms of real-world distances to be a third nearest neighbor of an event in time and space? Inherent in the Knox test and the space-time \( K \) test are ways to determine the scale of interaction. When a spatial and temporal scale is specified with the Knox test and significant interaction is identified, the results indicate the scale of the interaction in metrics of both space and time that are familiar to the user. With the Jacquez however,
translating the detection of significant interaction for a specific value of $k$ to familiar measures of distance and time is less intuitive.

Another question of interest left unanswered by the Jacquez test is, where and when within the study area and period does the identified interaction occur? Although locating and assessing the significance of local event clusters pertains to the realm of space-time scan statistics (i.e. Kulldorff et al., 2005; Takahashi et al., 2008), efforts have been made to explore the constituent parts of global space-time interaction tests visually, most notably in the case of the Knox test. This is done by mapping the links between pairs of events that are within the specified critical spatial and temporal distances of each other. Although not suggested in Knox’s original paper, over time, it has become a conventional method of displaying the results graphically. An early example of this can be found in Williams et al. (1978) and a more modern example in Grubesic and Mack (2008). Given the similar structure of the Knox and Jacquez tests with respect to identifying spatial and temporal adjacency, it seems logical to extend a similar visualization approach to the Jacquez test. In the context of this test, the links between events would identify instances where one event is both a spatial and temporal $k$ nearest neighbor of another event. The problem with this approach to visualizing results, as currently implemented for the Knox test, is that it remains essentially atemporal and therefore, it is not possible to assess how close in time the links are relative to one another on the map. There has been considerable work in terms of the visualization of space-time paths within a cube or aquarium (Hägerstrand, 1970; Miller, 2003; Kwan, 2004; Kraak and Koussoulakou, 2004); however, visualizing individual events in a cube is perhaps more challenging as it is difficult to plot discrete points in time and space and contextualize their positions relative to one another because of their lack of dimensionality. This study demonstrates examples for visualizing the spatio-temporal results of the Jacquez test employing the space-time cube.

The third and final issue regarding the Jacquez test addressed in this chapter is
population shift bias. This phenomenon was first identified by Mantel (1967) in the results of the Knox test and explored in greater detail by Kulldorff and Hjalmars (1999). It was shown to affect other tests of space-time interaction by Mack et al. (2012). Collectively, these studies demonstrate that spatially heterogenous change in the distribution of the underlying population from which space-time events are drawn is capable of biasing the results of these tests. This stems from the fact that traditional methods of assessing the significance of these statistics assume the events are drawn randomly from a probability distribution that is static across time and space. Instead, the significance of the tests must be determined using a probability distribution which accounts for the dynamic nature of the underlying population from which events are drawn. Failing to do so leads the tests to detect interaction due to clustering of the underlying population in space and time, unrelated to interaction stemming from the data generating process of interest. This results in more rejections of the null hypothesis than warranted for a given $\alpha$ level, yielding an increase in Type I errors and thereby biasing the test results. Here, we demonstrate the construction and implementation of a version of the Jacquez test which accounts for this population shift bias. The significance of the statistics calculated by this modified test are also determined using a Monte Carlo procedure, however, in this case, the reference distribution does not come from a permutation of the observed data. Instead, the reference distribution is simulated based on knowledge of the underlying population and its dynamics over time. Use of this simulated distribution results in a more accurate estimation of the significance of the test results.

4.4 Data

The data used in this chapter to illustrate the proposed enhancements come from a study conducted by Williams et al. (1978), investigating the spatio-temporal patterns of Burkitt’s lymphoma in the West Nile district of Uganda during the period 1961 to 1975. A data appendix accompanying their study provides spatial and temporal coordinates for the
onset of 188 cases of the disease throughout the region over the 15 year study period. The locations of Burkitt’s lymphoma cases are shown in Figure 4.1 along with the counties in the West Nile district of Uganda. A shapefile of the counties in the West Nile district was created by georeferencing and digitizing the study area map from the original publication (Williams et al., 1978). Although the map from the publication is clearly crude, it provides a reasonable basis for approximating the spatial area and extent of the West Nile district and its composite counties as they were demarcated at the time the study was conducted.

This dataset was chosen to illustrate the enhancements presented in this study for two reasons. First, these data are freely available both online and in print so interested readers may replicate the results and explore the methods proposed in this chapter simply...
by downloading the data and the software used to produce the results. Second, it is a
dataset that has been thoroughly explored in the literature, and as a result, there is a solid
understanding of both the space-time pattern of Burkitt’s lymphoma (Williams et al.,
1969; Morrow et al., 1971; Williams et al., 1978) and the etiological processes that are
primarily responsible for this disease (Ferry, 2006). This knowledge will help in the
interpretation of the results of the methods proposed in this chapter. It will also provide a
means of comparing the results generated in this work to those provided by other studies.
For the analyses carried out here, the data were divided into the same time periods used by
Williams et al. (1978): three five year periods (1961-65, 1966-70, 1971-75) and one two
year period (1972-3).³

<table>
<thead>
<tr>
<th>County</th>
<th>Cases</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1961</td>
</tr>
<tr>
<td>Koboko</td>
<td>9</td>
<td>23,081</td>
</tr>
<tr>
<td>Aringa</td>
<td>31</td>
<td>42,265</td>
</tr>
<tr>
<td>Maracha</td>
<td>25</td>
<td>48,545</td>
</tr>
<tr>
<td>Terego</td>
<td>34</td>
<td>44,134</td>
</tr>
<tr>
<td>Ayivu</td>
<td>36</td>
<td>57,108</td>
</tr>
<tr>
<td>Vurra</td>
<td>14</td>
<td>28,114</td>
</tr>
<tr>
<td>Okoro</td>
<td>4</td>
<td>50,839</td>
</tr>
<tr>
<td>Padyere</td>
<td>12</td>
<td>48,747</td>
</tr>
<tr>
<td>Jonam</td>
<td>13</td>
<td>27,138</td>
</tr>
<tr>
<td>Madi</td>
<td>15</td>
<td>29,628</td>
</tr>
</tbody>
</table>

Table 4.1: Population trends and cases of Burkitt’s Lymphoma in the West Nile District
derived from Williams et al. (1978).

In addition to the locations and times for disease cases, the calculation of the
Jacquez test which accounts for population dynamics also requires an estimate of the
density of the susceptible population and its change throughout time. Due to a paucity of
³This two year period was examined by Williams et al. (1978) in spite of the fact that it is a part of the
1971-75 period because the events contained within exhibited space-time interaction even though those from
the more expansive 1971-75 period did not. For the sake of consistency with their work I examine this
time period as well.
digital spatial or demographic data for the West Nile district of Uganda for the time the
case data were collected, we rely on the maps and demographic statistics published in
Williams et al. (1978) to generate these estimates. Estimated population data based on the
1968 Ugandan census were published with the original paper along with the estimated
change from 1959 to 1969 for each county in the West Nile district. These data were used
to estimate the compound annual growth rate during the study period which was then
extrapolated to estimate the populations in each year during the study for each county,
following the methodology of Kulldorff and Hjalmars (1999). These populations, along
with the lymphoma cases observed in each county are shown in Table 4.1. These data are
used in the subsequent sections to illustrate the enhancements offered by our version of
the Jacquez test.

4.5 Methods

In this section, methods are described which address the problems in the Jacquez test
outlined above. The methods developed have been implemented in Python and in some
cases R. Some of these methods have been packaged as part of the open-source
spatio-temporal analysis software, PySAL (Rey and Anselin, 2010). Where appropriate,
the enhancements to the Jacquez test presented here are compared to results from existing
tests of space-time interaction using the Burkitt’s Lymphoma data.

Establishing the Spatial and Temporal Scale for k

As mentioned previously, one drawback associated with the Jacquez test’s nearest
neighbor approach to testing for space-time interaction is the ambiguity surrounding how
values of k relate to real-world metrics of spatial and temporal distance. This section
illustrates a simple, yet effective technique for linking each value of k to a spatial and
temporal scale. Given that each value of k corresponds to a set of nearest neighbors for
each observation in the dataset, assigning a single value for the spatial and temporal scales
associated with each $k$ requires a distillation of these distributions. The technique proposed here to do this is a variation on the mean nearest neighbor distance proposed by Clark and Evans (1954). The approach involves computing the average spatial and temporal distances across all pairs of events that comprise the set of $k$ nearest neighbors for all events in the dataset. This is expressed mathematically in Equations 4.8 and 4.9.

$$\hat{S}_k = \frac{\sum_i^n \sum_j^n (a_{ij}^s d_{ij}^s)}{nk}$$

(4.8)

$$\hat{T}_k = \frac{\sum_i^n \sum_j^n (a_{ij}^t d_{ij}^t)}{nk}$$

(4.9)

The spatial average for level $k$ is shown in Equation 4.8 and the temporal average in Equation 4.9. Here $a_{ij}^s$ and $a_{ij}^t$ refers to adjacency in space and time, respectively, at level $k$ between events $i$ and $j$; defined previously in Equation 4.5. Terms $d_{ij}^s$ and $d_{ij}^t$ refer to the distance between events $i$ and $j$ in space and absolute value of the difference between the events in time, respectively. Essentially, these equations average the times and distances for the sets of $k$ nearest neighbors associated with each observation in the dataset. The average spatial and temporal distance associated specifically with a value for $\Delta k$ (i.e. the additional neighbors considered when moving between $k - 1$ and $k$), can be determined using Equations 4.10 and 4.11.

$$\hat{S}_{\Delta k} = \frac{\sum_i^n \sum_j^n (a_{ij}^s d_{ij}^s - a_{ijk-1}^s d_{ij}^s)}{n}$$

(4.10)

$$\hat{T}_{\Delta k} = \frac{\sum_i^n \sum_j^n (a_{ij}^t d_{ij}^t - a_{ijk-1}^t d_{ij}^t)}{n}$$

(4.11)
Example Using Burkitt’s Lymphoma Data

To establish spatial and temporal scales associated with space-time interaction for the Jacquez test, significant interaction must first be detected. To test for interaction in the Burkitt’s Lymphoma data, a combined $J_k$ test was run for each of the four time periods specified in the Williams et al. (1978) study: 1961-65, 1966-70, 1971-75, and 1972-73. The combined $J_k$ test is employed to establish if space-time interaction is present in the data across a range of values for $k$ because the exact scale of the interaction is not known. As described previously, when assessing the significance of results across a range of values for $k$, the combined test must be employed to account for the problems introduced by multiple testing. In the combined tests run here, $k$ assumed all values in the range from 1 to 10. This range has been used in the literature previously to establish significance using the Jacquez test (Jacquez, 1996; Ward and Carpenter, 2000a). The pseudo-significance of the tests for each period are shown in Table 4.2. The results from the combined Jacquez test mimic the conclusions reported by Williams et al. (1978) generated using the Knox test. The table shows that two periods, 1961-65 and 1972-73, exhibit significant space-time interaction whereas the other two periods do not.

<table>
<thead>
<tr>
<th>Period</th>
<th>Cases</th>
<th>Statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>61-65</td>
<td>35</td>
<td>10.484</td>
<td>0.001</td>
</tr>
<tr>
<td>66-70</td>
<td>72</td>
<td>2.513</td>
<td>0.571</td>
</tr>
<tr>
<td>71-75</td>
<td>81</td>
<td>3.004</td>
<td>0.498</td>
</tr>
<tr>
<td>72-73</td>
<td>37</td>
<td>7.452</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 4.2: Results for the combined Jacquez test ($J_{10}$) across the study time periods based on 999 permutations.

Given that significant space-time interaction was observed for the periods 1961-65 and 1972-73 up to and including the scale of $k = 10$, it was necessary to establish the
values of $k$, beyond $k=10$, for which interaction was no longer detected. This value was determined by increasing the value of $k$ iteratively in steps of 1. For each increase in the value of $k$ the significance of the cumulative $J_k$ was calculated. This process was repeated until the test statistics for individual levels of $k$ were consistently no longer significant. By examining the spatial and temporal distances associated with the highest value for $k$ which proved to be significant (determined using Equations 4.8 and 4.9) the approximate spatial and temporal scales at which the interaction occurs can be established. The results of this analysis are shown in Figure 4.2 for 1961-65 (top) and 1972-73 (bottom). The figures on the left link the different levels of $k$ to particular distances in space (kilometers), while those on the right link the different levels of $k$ to particular times (in days). Approximate boundaries for the scale of interaction are denoted by dashed lines in each of the figures.

Comparison with Knox Results

To corroborate the scales of interaction diagnosed by this enhancement to the Jacquez test the results were compared to the original results reported by Williams et al. (1978) established via the Knox test. While the results of the two tests will not align exactly because the distances and times reported by the enhanced Jacquez are averages across all $k$th nearest neighbors and the Knox test was only calculated at set intervals, similarity in the results provides a method of verifying the utility of the proposed enhancement.4

In the original analysis by Williams et al. (1978), the Knox test detected significant interaction for the 1961-65 period for most combinations of threshold distances and times less than the critical values of 360 days and 40 kilometers. Similar results were found with the combined Jacquez test. Values for the cumulative $J_k$ statistic for this subset of the data remain significant up to $k = 19$. As Figures 4.2a and 4.2b show, the average spatial

---

4In their original exploratory analysis of the Burkitt’s lymphoma dataset Williams et al. (1978) employed critical spatial distances of 2.5, 5, 10, 20, and 40 kilometers, and critical temporal distances of 30, 60, 90, 120, 180 and 360 days for the Knox test.

88
distance between cases associated with this level of $k$ is $\approx 31$ kilometers and the average temporal distance is $\approx 307$ days. Examination of the 1972-73 period reveals similar agreement of the enhanced Jacquez with the results reported by Williams et al. (1978). The Knox revealed significant interaction for critical spatial distances between 5 and 40 kilometers, and critical temporal distances between 90 and 180 days. The cumulative Jacquez test produced significant clustering for values of $k$ between 6 and 31, which correspond to a spatial scale between 14 and 34 kilometers and a temporal scale between
40 and 174 days using the enhancement described above. The relative agreement between the results of the two tests is an indication of the effectiveness of this proposed enhancement to the Jacquez test.

Visualization

As mentioned previously, one of the shortcomings of the Jacquez test is the lack of visual output associated with the results. This issue is largely a by-product of the global nature of the test statistic. However, by visualizing the links between events and their $k$ nearest neighbors common in both space and time, the user can gain a better understanding of when and where these links which comprise the Jacquez test statistic occur within a dataset. This section explores different methods of visualizing this information.

Space-Time Cube

To start, the links produced by the enhanced Jacquez statistic are visualized in a three-dimensional space-time cube. The cube was implemented using the R environment for statistical computing (R Development Core Team, 2011) and the Scatterplot3D library (Ligges and Mächler, 2003). An illustration of this cube with the Burkitt’s Lymphoma cases from 1961-65 is shown in Figure 4.3. The $x$- and $y$-axes of the cube correspond to the spatial coordinates of the cases while the $z$-axis or height of the cube corresponds to the temporal dimension. Events are plotted in space and time along with the mutual nearest neighbor linkages comprising the $J_5$ statistic, visualized as bold black lines.

The user is able to rotate the cube to explore the cases and their interaction with one another. In examining these cases in the space-time cube, a number of centers of interaction are apparent. However, it is difficult to gauge the proximity of cases in this implementation of the cube because the user has no sense of perspective. This lack of
perspective is recognized as one of the issues with this visualization approach (Kraak, 2003; Andrienko et al., 2003). The disorientation may be mitigated somewhat by utilizing the cube in conjunction with maps and other graphical displays to best highlight trends in data (Gatalsky et al., 2004).

In exploring the results as presented in the space-time cube, it became apparent that most information was actually gleaned from a quasi-areal perspective (seen mainly as a map) or from a side view where either the x- or y-axis is placed horizontally in front of the user while the z-axis remains vertical. Consequently, the best static method to display this information is in a multi-paned graphic comprised of four elements: the space-time

Figure 4.3: Space-time cube of cases for 1961-65 and connections for $J_5$. 

In exploring the results as presented in the space-time cube, it became apparent that most information was actually gleaned from a quasi-areal perspective (seen mainly as a map) or from a side view where either the x- or y-axis is placed horizontally in front of the user while the z-axis remains vertical. Consequently, the best static method to display this information is in a multi-paned graphic comprised of four elements: the space-time
cube, a conventional map of space-time linkages, a slice of the space-time cube with the 
$y$-axis of the plot in the traditional horizontal location of the $x$-axis and time in the vertical 
location of the $y$-axis, and a slice of the space-time cube with the $x$-axis in its traditional 
location and time in the location of the $y$-axis. These elements are shown in Figure 4.4.

An examination of the linkages identified as Collection 1 in the space-time cube 
(Figure 4.4d) reveals the utility of this multiple perspective approach. If the user were to 
consider only the static space-time cube presented in Figure 4.4d, the projection effect 
resulting from viewing three-dimensional phenomena on a two-dimensional surface (i.e. 
computer screen or paper) (Gatalsky et al., 2004), may lead the user to believe that this 
collection of linkages is close in space to Collections 2 and 3. However, when the 
collections in the cube are considered in conjunction with the map (Figure 4.4a), $x$-axis 
profile (Figure 4.4c), and $y$-axis profile (Figure 4.4b), it becomes obvious that this 
collection is quite distant from Collections 2 and 3. This multi-paned perspective also 
helps to identify that Collection 1 is located between Collection 2 and 3 in time. Based 
solely on the view afforded by the static space-time cube however, it may appear 
Collection 1 occurs after the other two collections in time.

This example clearly demonstrates that multiple perspectives are needed to fully 
understand the distribution of the links generated by the Jacquez test. This visualization 
approach is similar to the dynamic linking techniques advocated by Andrienko et al. 
(2003), which enables users to translate the three-dimensional visualization of 
phenomenon in the space-time cube to a two-dimensional map, or similar frame of 
reference, without losing their orientation. Although more interactive and dynamic 
approaches, such as that developed by Andrienko et al. (2003), may prove more useful for 
exploring the Jacquez results, a key advantage of the technique presented in this chapter is 
that it effectively visualizes three-dimensional phenomena on a two-dimensional surface 
and is thus more relevant for print or static digital media.
Figure 4.4: A set of plots to visualize space-time interaction as detected by the Jacquez test ($J_5$): (a) Conventional map of nearest neighbor connections. (b) Plot of nearest neighbor connections with latitude on the $x$-axis and time on the $y$-axis. (c) Plot of nearest neighbor connections with longitude on the $x$-axis and time on the $y$-axis. (d) Space-time cube of nearest neighbor connections.

Comparison with Knox Visuals

Finally, as part of the work to visualize the results of Jacquez test, the results are compared to those of the Knox test. Given that the Knox results are not conventionally
visualized in a space-time cube, the comparison was made using the linkages as they are represented on a conventional map. To visualize the results of both tests, the data from 1961-1965 are used and the Jacquez linkages are shown for $J_5$. For the Knox test, critical thresholds of 13 kilometers (space) and 90 days (time) were used to approximate the distance and time associated with a average spatial and temporal distance associated with the links for $J_5$ according to Figure 4.2. These maps are shown in Figure 4.5.

The comparison reveals general consistency between the location of links identified by the two tests: both identify numerous linkages in the northwest portion of the study area and less in the southern portion. It is apparent from these visuals that the Jacquez test identifies more space-time links than the Knox test. This is due to the more robust nearest neighbor approach of the Jacquez, which is unconstrained by the set thresholds of the Knox test and thereby adjusts the definition of adjacency based on event concentration. Its nonlinear nature also allows it to detect a greater number of links in areas with a more dispersed concentration of events (i.e. the southern part of the study area). Generally though, there is visual agreement between the two tests, which corroborates the findings from Section 4.5. After exploring the results using the different perspectives offered by the space-time cube, however, it is apparent that plotting the results using only a map as a visual aid tells only part of the interaction story.

*Incorporating Population Shift*

Having addressed the issues of ambiguous spatial and temporal scales of interaction, and visualization of the Jacquez test results, this section focuses on the last of the three shortcomings associated with this test: population shift bias. As discussed previously, spatially heterogenous changes in the underlying population from which events are drawn can lead to an increase in Type I errors if the changes are not reflected in the probability distribution used to assess the significance of the results (Kulldorff and Hjalmars, 1999).
Figure 4.5: A comparison of the locations of significant linkages identified by the Jacquez $(k = 5)$ and Knox (space = 13 km, time = 90 days) tests.
Here, a method is demonstrated that illustrates how to conduct statistical inference for the Jacquez test in study areas that experience such heterogeneous population growth (or contraction). The method follows the general framework provided by Kulldorff and Hjalmars (1999), but differs in that it does not base inference on a standard probability distribution (e.g. Poisson or Normal as in the case of the Knox test). Instead, a Monte Carlo method is employed to account for the fact that the exact probability distribution for the Jacquez test is unknown (Jacquez, 1996). Unlike standard Monte Carlo approaches to significance testing for space-time interaction however, the reference distribution does not come from a permutation of the observed data. Instead the reference data are simulated based on knowledge of the underlying population and its dynamics through time. The steps for conducting this test of space-time interaction which corrects for heterogenous population dynamics, originally outlined by Kulldorff and Hjalmars (1999), are described below as adapted for the Jacquez test.

**Step 1:** Generate $N$ random event datasets such that each contains the same number of events, $\lambda$, as the observed data. The events in the simulated datasets must be distributed randomly throughout the study area and time period of interest based on probabilities proportional to the population at all given location and time combinations (for an extended explanation see Kulldorff and Hjalmars (1999)). To achieve this, there must be an estimate of the spatial and temporal distribution of the underlying population across the study area throughout the time period of interest. Although an exact and continuous measure of the underlying population throughout time is more than likely unavailable, reasonable discrete estimates can be made based on population information available through time for defined spatial units. For our example, population estimates were derived for the counties in the West Nile district for the periods 1961-65 and 1972-73 based on the methodology described in Section 4.4.
Step 2: Calculate the test statistics for the observed and $N$ simulated datasets. Here, the $J_k$ statistics for $k = 1$ through 25 for the period 1961-1965 and $k = 1$ through 35 for the period 1972-1973 were calculated. A combined test across multiple values of $k$ could also be specified and either formulation of the test, the cumulative or the $k$-specific measure ($J_k$ or $\Delta J_k$) could be used.

Step 3: Finally, the pseudo-significance of the Jacquez test statistic for the observed data is determined by ranking it within the distribution of test statistics for the $N$ simulated datasets. The number of statistics greater than the observed value is tallied. This number, $c$, is then inserted into Equation 4.7 to get the pseudo $p$-value for the combined Jacquez test, corrected for spatially heterogenous population dynamics.

The result of this process for the Burkitt’s Lymphoma data for the 1961-1965 and 1972-1973 periods are presented in Figure 4.6. The pseudo $p$-values generated by this permutation approach which corrects for heterogeneity in the population dynamics are compared to the pseudo $p$-values generated by the assessment approach originally advocated by Jacquez (for both, $N = 999$). The results show that for these data, there is only a small difference in the results generated by the two methods of assessing pseudo-significance. However, as anticipated from the results provided by Kulldorff and Hjalmars (1999) and Mack et al. (2012), the corrected values are slightly higher than those provided by the original estimation across most values of $k$ where significant interaction was observed.

The difference between the $p$-values for the two versions of the test shown in Figure 4.6 reflect the heterogeneous growth of the underlying susceptible population. By not accounting for this growth, the original test for space-time interaction appears to yield

---

5These values of $k$ correspond to the range of values over which significant interaction was identified in Section 4.5.
Figure 4.6: A comparison of the corrected and uncorrected combined Jacquez test results
Even after accounting for this shift, though, by adjusting for the heterogenous growth in the underlying population using the method described here, the statistic still remains significant at an $\alpha$ level of 0.05 for most values of $k$. In this example, the differences in results produced by the two methods are minimal because of the small population shifts observed during the time periods examined here. The difference in annual rate of population change between the slowest (2.33%) and fastest (6.91%) growing regions manifest only in slight heterogeneity of the populations of the regions over the 2 and 5 year periods examined here. Population shift bias becomes more of a problem when more dramatic heterogenous population changes (e.g. when growth and contraction happen within different regions of the same study area) occur over longer time periods (Kulldorff and Hjalmars, 1999) or when such population changes occur in short time periods examined using a high number of temporal intervals (Mack et al., 2012). In these situations, the user should take care to assess significance of the Jacquez results using the suggested approach to avoid increasing the likelihood of Type I errors.

4.6 Discussion and Conclusions

The Jacquez test for space-time interaction has been shown to be quite powerful and has demonstrated greater flexibility over other tests of space-time interaction (Jacquez, 1996). The test is particularly relevant for studies where the suspected interaction is nonlinear or does not conform well to the explicit thresholds presumed by other tests. Despite the advantages associated with this test however, it has been utilized in practice less frequently than other tests of space-time interaction. Potential reasons for this limited application are the ambiguity associated with diagnosing the spatial and temporal scale at which interaction occurs for a $k$ nearest neighbor based test statistic and the limited visual output of the test.
However, with the three enhancements developed in this study the utility of this test is greatly improved. By keeping track of the average distances and times associated with specific levels of $k$ the work helps to better contextualize the test’s results in terms of real world distances and times. This is done by bringing the distances into the adjacency matrix, thereby preserving valuable information that would previously be lost in the calculation of the Jacquez statistics. Additionally, modifying the way the pseudo-significance of the statistics is calculated helps reduce the impact of population shift bias. Rather than relying the traditional approach of permuting the pairings between spatial and temporal coordinates, the enhanced method simulates entirely new patterns by incorporating estimates of spatially heterogenous changes in the underlying population from which events are drawn. Although the process is far more computationally intensive than the original, it provides a more accurate estimate of the true unusualness of a particular spatiotemporal pattern. Finally, the work provides tools to visualize the results by visualizing the affirmative links in the adjacency matrix. The additional information gleaned from these enhancements combined with the visualization tools help make the Jacquez test results more relevant and easily interpreted. Additionally, by accounting for population shift bias, the specificity of the test has been increased. In spite of these contributions however, future research is necessary to expand upon the visualization of the enhanced Jacquez test results presented in this study.

While this work presents a key first step in visualizing the results, additional efforts could be directed at visualizing the results in a more intelligent dynamic space-time cube. Specifically, the cube could be made to interact with the map and perspective views presented here, through dynamic linking and brushing techniques (Andrienko et al., 2003; Gatalsky et al., 2004). Implementation of these tools would mitigate the abstraction of the cube. Further work also needs to be invested in developing methods for the space-time cube to allow for the display of a greater number of
observations. Although the dataset employed in this study was small, orientation issues are still present when rotating the cube. This problem will only be compounded with the use of larger datasets or examining linkages generated by a larger value of \( k \). However, the inclusion of more ancillary data within the cube itself may assist with this problem.

Along with improvements in visualization, further work remains on the decomposition of the Jacquez results in terms of the spatial and temporal scales associated with each level of \( k \). While this work presents an average of the distances and times associated with different levels of nearest neighbor linkages, additional techniques for summarizing these distributions of spatial and temporal data should be explored. Simple extensions might include the use of the median distance and time between nearest neighbor events, in place of the mean. Additional work might also explore the variability within the distribution of distances associated with each level of \( k \) in space and examine how it changes through time (and vice versa). This may yield a more complete picture of the scale of interaction.

However, the manner in which the spatial and temporal scale of interaction is diagnosed in this study certainly provides useful information that can be used to approximate the scale of space-time interaction. This information can be used in a comparative context to evaluate the results of other tests for space-time interaction. Alternatively, this scale information might also be used to better inform the selection of critical space and time thresholds for the computation of the Knox test, or the space and time sub-intervals necessary for the computation of the space-time \( K \) function. Such information obtained from the enhanced Jacquez test would be particularly beneficial if the researcher has no prior knowledge of the scale of interaction, and would thus reduce the subjectivity associated with the selection of spatial and temporal thresholds needed by other tests.
Although additional work remains, the enhancements presented in this study represent a series of useful techniques that transform the Jacquez test into a complete, descriptive, informative metric that can be used as a stand alone measure of global space-time interaction. Not only do these enhancements address some of the suggestions for the Jacquez test made by prior studies, such as the development of a version of the test accounting for heterogenous population dynamics (Kulldorff and Hjalmars, 1999), but these enhancements open the door to increasingly sophisticated evaluations and visualizations of space-time interaction.
Chapter 5

CONCLUSION

The development of methods to identify interaction within patterns of spatio-temporal events began almost fifty years ago with an attempt by Knox (1964) to better understand and quantitatively define the spatio-temporal footprint of contagion and epidemicity. In the half century since the publication of Knox’s paper, which coined the term “space-time interaction,” researchers have put forth a variety of methods to identify the phenomenon from both a global and local perspective. Although these methods were initially purposed with identifying interaction in the context of epidemiological datasets, new applications for the methods have been identified, expanding their utility far beyond this initial context. Despite the considerable effort to advance these methods in this time, one important facet of their development was left unexplored: how the tests are affected by the data inaccuracy and uncertainty that is ubiquitous in spatial and spatio-temporal datasets (Zhang and Goodchild, 2002; Meliker and Sloan, 2011). Chapters 2 and 3 of this dissertation address this paucity directly and bridge the gap between existing studies on the prevalence of uncertainty and inaccuracy in spatial and spatio-temporal data and the body of work surrounding methods to detect space-time interaction. These studies help to better understand how a selection of global and local space-time interaction tests are affected by these complications common to real-world data. The experiments carried out in this regard, while simulations, are rooted in empiricism, both in terms of the nature of the patterns and the inaccuracies introduced to them as part of the experiments. Additionally, Chapter 4 of this work enhanced the functionality of a global test of space-time interaction (the Jacquez test) by ameliorating some of its known shortcomings. Collectively, these three studies help to better understand the strengths and weaknesses of these methods and clarify how they can be used most effectively and appropriately. Although this document
The experiments conducted in Chapter 2 showed the results of global tests of space-time interaction to be negatively affected by the degree of inaccuracy introduced here. The series of simulation experiments conducted in this chapter demonstrated that in some cases, common data problems affecting the integrity of spatial and temporal coordinates of input data (i.e. inaccuracy and uncertainty) can completely obscure evidence of space-time interaction in the results of these tests, while in others they may create it where it did not originally exist. The take away message from this work is that estimates of confidence in the results of these tests that fail to consider the potential impact of these problems must not be taken at face value. Although the examined global tests were severely affected by the introduced perturbations, the local test explored in Chapter 3 (i.e. the space-time permutation scan statistic) was shown to be more robust, relatively speaking. While the data problems introduced in the experiments of Chapter 3 were identical to those employed in Chapter 2 for the global tests, the space-time permutation scan statistic (STPSS) results were, on the whole, more likely to correctly identify the presence of localized interaction. In spite of its superior performance relative to the global tests, the STPSS was still afflicted by the inaccuracy and uncertainty. Specifically, as the degree of these data problems increased the ability of the method to correctly identify the true most likely cluster attenuated, especially when multiple significant hotspots were present.
While these two studies clearly illustrated that problems can result when inaccurate or uncertain data are analyzed by these tests, they only scratch the surface of the work that is required to develop a complete understanding of the impact of such problems on these tests. Moving forward, more work needs to be devoted to generating a more generalized picture of how uncertainty and inaccuracy in the spatial and temporal input data affect these methods. In Chapters 2 and 3 efforts were taken to introduce a conservative degree of these data problems, based on the extent of these problems reported in the literature. While these findings are important, as they indicate how the methods can be affected by even conservative estimates of these problems, more research needs to be undertaken to better understand how sensitive these results are to varying degrees of these problems in the input data. One approach to conducting a study of this nature would be to introduce differing levels of spatial and temporal perturbations (as was done here) but vary the degree of perturbation systematically and record the degree of perturbations relative to the overall size and duration of the study area of interest. This approach would be valuable because it ensures the resulting insights would transcend units of any particular study area thus be more universally relevant than those presented here. This appears to be the most logical starting point for follow on work from this research. As presented, the results for the first two chapters are largely scale dependent. While this work shows that commonly encountered levels of these effects can be problematic, a more abstract and scale-independent study needs to follow on this theme and develop a more generalized understanding of how much of an effect a given level of uncertainty or inaccuracy will have on the results of these tests. In spite of the scale-dependence of these results, I thought it important to first establish the need for a more abstracted study using empirically derived parameters for the perturbations. In the absence of examples like those I’ve provided here, a study comprised of purely synthetic examples and user-defined perturbations may have seemed out of context.
To pull back and examine this research from a broader perspective, the common problem undergirding each of these studies and the tests examined here (and many more which were not) is the fact that the methods assume that researchers can precisely and accurately represent the complex and dynamic world in which we live in a static spatial or spatio-temporal database: this is an illusion that we have become experts at propagating (Koch and Denike, 2007). As shown here, conducting analyses based on this assumption can yield misleading results when it is violated. To truly bring about value from the work conducted here, it is important to recognize that more robust methodologies (generally and in the context of space-time interaction tests) which account for these inaccuracies and uncertainties need to be developed and brought into practice by researchers. The results of these investigations offer further fodder to the call for the development of methods which inherently account for the presence of inaccuracy or uncertainty in their input data and adjust findings accordingly (Wei and Murray, 2012). Such methods are especially important in the context of these tests given that they are most frequently employed in the field of epidemiology where the consequences of false positive and false negative results are especially meaningful. False negatives introduced by data uncertainty may lead researchers to abandon efforts to investigate event patterns further, potentially losing out on valuable inference. Alternatively, false positive results would lead researchers to devote time, efforts, and funds in areas where little knowledge can be gained.

Keeping with this theme of enhancing the utility of these tests for practitioners, Chapter 4 of this work ventured beyond exploring the impact of spatio-temporal uncertainty and inaccuracy on the results of these tests into developing tangible improvements for one of the methods. Here methods were developed and implemented which address some additional shortcomings (outside of those identified here) of the Jacquez test, one of the global test of space-time interaction examined in Chapter 2. By adjusting for the problem of population shift bias in the results of the Jacquez tests,
visualizing its results, and better contextualizing them in terms of real-world distances and times, the chapter offers a series of improvements to the test that make it a more useful metric for practitioners. However, as previously discussed, further work obviously needs to be dedicated to improving the performance of this test in the face of uncertainty and inaccuracy as well.
REFERENCES


