Modelling and Analysing Vague Geographical Places using Fuzzy Set Theory

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

By

Firdos Mohammed Almadani (MSc Leicester)

Department of Geography

University of Leicester

2016

Abstract

__Firdos Mohammed Almadani__

Title: Modelling and analysing vague geographical places using fuzzy set theory

Vagueness is an essential part of how humans perceive and understand the geographical world they occupy. It has now become of increasing important to acknowledge this situation in geographical databases and analyses in the field of Geographical Information Science (GIScience). This research has tackled the wholly original topic of modelling vague geographical places (objects) based on fuzzy set theory with a view to assessing the implications of routing problem around those vague places. The research has focused on the modelling of vague places, for a number of villages and rural settlements, working with national address databases which have numerous ambiguous characteristics which add challenge to the work. It has demonstrated the way in which fuzzy set theory can be used to derive approximate boundaries for vague spatial extents (fuzzy footprint) form sets of precise addresses, reporting rural settlements, recorded in different databases. It has further explored the implications of applying the Travelling Salesman Problem (TSP) in traditional hard village extents versus the modelled fuzzy extents. The introduced methods evaluate the usefulness of fuzzy set theory in modelling and analysing such vague regions. The results imply that the fuzzy model is more efficient than the traditional hard, crisp model of approximating the spatial extent of rural areas. However, the TSP results showed that longer tours were mostly found in the fuzzy model than the traditional crisp model. This is mainly affected by the scale factor of rural areas, considering the relatively small distances between villages. One challenge for the approach outlined here is to incorporate this method applied in other novel analyses of geographical information based on fuzzy representation of geographical phenomena.

Acknowledgments

I am grateful to God for the good health and wellbeing that were necessary to complete this thesis.

I would like to express my deepest appreciation and my sincere thanks to all those who provided me the possibility to complete this thesis. A special gratitude I give to my supervisors, Prof. Lex Comber whose contribution in stimulating suggestions and encouragement, Dr. Claire Jarvis for her continued support and constant supervision, and Prof. Pete Fisher who sadly passed away on May 2014 for his great role in the early days of my research and for his advice and for being a great teacher who taught me the fundamentals of the academic research, I am extremely thankful and indebted to him for sharing expertise, and sincere and valuable guidance and encouragement extended to me while he was supervising my thesis before his sudden death. You will be always memorable for the care and support you have given to me.

I take this opportunity to express my deepest gratitude to all of the geography department staff and research students with whom I interacted, who made this research journey enjoyable, despite the hard times, which I passed through. Furthermore, I would also like to acknowledge with much appreciation the crucial role of all my friends who are sincere friends and encouraged me to finish this research work.

Last but not least, I would like to present my sincere thankfulness to my parents for the unceasing encouragement, support and attention. I am so grateful to them for their support, patience and tolerance through my study. Additionally, I would like to thank my brothers and sisters and my nieces and nephews for their love and care. Finally, I also place on record, my sense of gratitude to one and all, who directly or indirectly, have supported this research.

Table of Contents

ABSTRACT ..... **I**

ACKNOWLEDGMENTS ..... **II**

TABLE OF CONTENTS ..... **III**

LIST OF FIGURES ..... **V**

LIST OF TABLES ..... **VIII**

LIST OF PUBLICATIONS ..... **IX**

ABBREVIATIONS ..... **X**

CHAPTER 1 INTRODUCTION **1**

1.1 OVERVIEW 1
1.2 RESEARCH AIMS & OBJECTIVES 3
1.3 RATIONALE AND SIGNIFICANCE OF THE RESEARCH 4
1.4 THESIS STRUCTURE 5

CHAPTER 2 LITERATURE REVIEW **7**

2.1 INTRODUCTION 7
2.2 VAGUENESS IN GEOGRAPHY 7
2.3 APPROACHES USED TO ADDRESS VAGUENESS 9
2.4 BASIC CONCEPT OF FUZZY SET THEORY 11
2.4.1 Fuzzy Sets 12
2.4.2 Fuzzy Sets Operations 18
2.4.3 Fuzzy Numbers 19
2.4.4 Fuzzy Relations: 20
2.4.5 Fuzzy Logic: 21
2.5 FUZZY SET THEORY IN GISCIENCE 22
2.5.1 Fuzzy Objects: 25
2.6 MODELLING VAGUENESS IN GISCIENCE 26
2.6.1 Higher Order Vagueness 31
2.7 FUZZINESS AND TRAVELLING SALESMAN PROBLEM 32
2.8 SUMMARY AND NEW INSIGHTS 33

CHAPTER 3 VILLAGES' IDENTIFICATION AND DATA SPECIFICATION **35**

3.1 INTRODUCTION 35
3.2 TERMS AND DEFINITIONS OF RURALITY 36
3.3 STUDY AREA 38
3.4 DATA SOURCES 41
3.4.1 Core Data 41
3.4.2 Supplementary Data 45
3.5 SETTLEMENT IDENTIFICATION 48
3.6 DISAMBIGUATION OF PLACE NAME RECORDS 51
3.6.1 Indeterminacy, Incompleteness and Inconsistency 52
3.6.2 Spelling and Punctuation 54
3.6.3 Parish Boundary vs. Village Extent 55
3.6.4 Combined Settlement Names 57
3.7 SUMMARY 58

CHAPTER 4 MODELLING AND ANALYSING VAGUE RURAL SETTLEMENTS **60**

4.1 INTRODUCTION 60
4.2 HARD / FIXED BOUNDARY 61
4.2.1 Convex Hulls 61
4.2.2 Voronoi Tessellation 63
4.2.3 Key Issues: 66

4.3 FUZZY MODELLING 66
4.3.1 Modelling based on Density of Houses 66
4.3.2 Generating α-cuts: 68
4.4 ANALYSING THE SETTLEMENTS' SPATIAL PATTERNS 69
4.4.1 Analysing Settlement Representation in Different Sources 69
4.4.2 Analysing Inclusion 73
4.5 RESULTS AND DISCUSSION 75
4.5.1 Modelling based on the Density of Houses 75
4.5.2 Comparative Analysis of Settlement Representation in Different Sources 77
4.5.3 Analysis of Inclusion 80
4.6 SUMMARY AND GENERAL DISCUSSION 121
4.6.1 Data Used: 121
4.6.2 Fuzzy Model of the Village 122
4.6.3 Using α-cuts 123

CHAPTER 5 IMPLEMENTATION OF THE TRAVELLING SALESMAN PROBLEM IN FUZZY LOCATIONS 124

5.1 INTRODUCTION 124
5.2 BACKGROUND INFORMATION ON TSP 125
5.2.1 General Overview and History: 125
5.2.2 Structure and Formulations of the TSP: 126
5.2.3 Methods to solve the TSP: 127
5.3 IMPLEMENTATION OF THE TSP 132
5.3.1 Determine Locations for TSP 132
5.3.2 Measuring the Network Distance 135
5.3.3 Solve the TSP 135
5.4 RESULTS OF THE TSP 136
5.4.1 Centre Locations for TSP: 136
5.4.2 Solving the TSP (Finding the shortest path – 2-opt) 140
5.5 DISCUSSION 146
5.6 SUMMARY AND FURTHER RESEARCH 148

CHAPTER 6 DISCUSSION 150

6.1 INTRODUCTION 150
6.2 SUMMARY OF RESULTS 151
6.3 REFLECTION ON METHODS 154
6.3.1 Summary of the Methods 154
6.3.2 Possible Alternative Approach 156
6.4 LIMITATIONS 157
6.5 FUTURE RESEARCH 158

CHAPTER 7 CONCLUSION 160

7.1 GENERAL SUMMARY OF RESEARCH CONTRIBUTION 160
7.2 OVERALL RESEARCH OUTLOOK 160

APPENDICES 161

APPENDIX (1): FIELD TRIP AGENDA SEPTEMBER 18, 2013 161
APPENDIX (2): MAPS OF THE NORMALISED KERNEL DENSITY WITH THEIR A-CUTS 176
APPENDIX (3): REGRESSION PLOTS BETWEEN EACH PAIR OF THE DATA SOURCES 230
APPENDIX (4): PLOTS OF THE FUZZY MEMBERSHIP GRADES ALONG EACH THRESHOLDS 233
APPENDIX (5): MAPS OF THE TRAVELLING SALESMAN PROBLEM FOR SOME SUBSETS 245
APPENDIX (6): DESCRIPTION AND PROVENANCE OF R CODES USED IN THIS THESIS 267

BIBLIOGRAPHY 279

List of Figures

Figure 2.1: Fuzzy set over discrete domain, an example of "the sensible number of children". 14
Figure 2.2: Formulations of Membership Function (MF) – (modified after Saggaf and Nebrija, 2003). 15
Figure 2.3: Fuzzy set over continues domain, an example of "distance about 5 Km". 16
Figure 3.1: The new rural definition and area classification modified after Bibby & Shepherd (2004). 37
Figure 3.2: Location of Hinckley and Bosworth District within Leicestershire. (Source: © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.) 38
Figure 3.3: Map of Hinckley and Bosworth Borough contains 24 Civil Parishes (CP). Hinckley is not a parished area (NCP). (Source: © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.) 46
Figure 3.4: Map of Hinckley and Bosworth Borough containing settlement points from the OS strategi layer. (Source: © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.) 48
Figure 3.5: Examples of the disparity in the village names due to using different address databases. The column on the left shows the same settlements with the address being aggregated from all databases. (Source: © Crown copyright and/or database right 2012. All rights reserved.) 53
Figure 3.6: Examples of the disparity in village names resulting from use of a separator. (Source: © Crown copyright and/or database right 2012. All rights reserved.) 54
Figure 3.7: Examples showing the relation between the extents of some village names and their parish boundaries (red polygons). (Source: © Crown copyright and/or database right 2012. All rights reserved.) 55
Figure 3.8: Examples of some parishes (red polygons) associated with more than two settlements. The column on the left shows the addresses belonging to settlements with the same parish names; while the column on the right shows other associated settlements. (Sourcee: © Crown copyright and/or database right 2012. All rights reserved.) 57
Figure 3.9: Examples of parishes with combined settlement names. The first column (on the left) shows the addresses that belong to settlements with the same parish names combined; while the second shows them separated and the last column (on the right) shows other associated settlements. (Source: © Crown copyright and/or database right 2012. All rights reserved.) 58
Figure 4.1: Maps showing the Convex Hulls for The Rural Settlements in The BS76 and POST Datasets; with some examples in large scale on the right. (Source: © Crown copyright and/or database right 2012. All rights reserved.) 63
Figure 4.2: Pseudo-code for generating the Voronoi tessellations for the settlements in the four data sources in the study area. 64
Figure 4.3: Results of Voronoi tessellations for the settlements in the four data sources. A, C, E & G show the mean centres while B, D, F & H show the original address points. 65
Figure 4.4: Pseudo-code for generating the normalised density surfaces and their a-cuts for settlements. 69
Figure 4.5: Pseudo-code to apply the linear regression model in R, full script is available in Appendix (6.3). 72
Figure 4.6: General structure of the R scripts executed to apply the inclusion analysis for all settlements in the different data sources. 74
Figure 4.7: Fuzzy representation for Barwell, Desford & Stoke Golding in each dataset, normalised density in the top left corner, with their nine a-cuts underneath each village. See Appendix (2) for full size images. 76
Figure 4.8: Regression Plots for Barwell, Market Bosworth and Newbold Verdon, which show strong positive linear relationship between the BS76 and POST data and other moderate relation between the other pairs of data (POI vs BS76 & POI vs POST). 79

Figure 4.9: Regression plots for Peckleton, Sutton Cheney and Twycross that showing no apparent pattern in the association between the pairs of the different data sources (not suitable for linear regression). 80
Figure 4.10: Maps of two settlements (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold).where the smallest α-cuts drop before reaching larger thresholds as their points tend to cluster in the middle and scatter further apart. 85
Figure 4.11: Map of two settlements(on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold). These show fairly small α-cuts for all thresholds suggesting the possibility of their addresses being spread out. 86
Figure 4.12: Map of two villages (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold), in which small transition between α-cuts values reflecting the compactness of their addresses....87
Figure 4.13: Map of three settlements (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold), where the smallest α-cuts drop out before reaching larger thresholds which indicate the scatter nature of their points. 92
Figure 4.14: Map of some settlement (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold), in which less disparity between α-cuts values showing that their addresses are compact. 94
Figure 4.15: Map of some villages (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold) showing sudden decline in their α-cut values that relates to points located further and spread away from the main point groups....95
Figure 4.16: Examples of two settlements (on the left, maps showing their address points and fuzzy models, with bar charts on the right indicating the optimal α-cut for each threshold),as a case of less defuse points which have gradual transition between their α-cuts. 96
Figure 4.17: Two examples of villages (maps on the left showing their address points and fuzzy models with bar charts on the right indicating the optimal α-cut for each threshold) that fail to capture 95% of their address points indicating that some of their points are less intense. 101
Figure 4.18: Two examples of small villages (maps on the left showing their address points and fuzzy models, with bar charts on the right indicating the optimal α-cut for each threshold) that have slightly higher value of membership across the different thresholds. 102
Figure 4.19: Two examples of villages (maps on the left showing their address points and fuzzy models with bar charts on the right indicating the optimal α-cut for each threshold) that have gradual decrease in the membership values which show how their points vary in intensity., ...103
Figure 4.20: Maps of two settlements (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold), which show sudden reduction in the membership grades related to the distribution of their point pattern. 104
Figure 4.21: Three examples of settlements (maps on the left showing their address points and fuzzy models with bar charts on the right indicating the optimal α-cut for each threshold) that miss out some thresholds because of their point distribution which contain some diffused point(s). 109
Figure 4.22: Two examples of settlements with overall high-grade of membership and gradual transition between their start and end α-cut values. Maps on the left show their actual address points and fuzzy models with bar charts on the right indicate the optimal α-cut for each threshold. 110
Figure 4.23: The case of some settlements with overall low-grade of membership values across all thresholds indicating the disperse nature of their points. Maps on the left show their actual address points and fuzzy models with bar charts on the right indicate the optimal α-cut for each threshold. 111

FIGURE 4.24: EXAMPLES OF SETTLEMENTS THAT HAVE SOME OF THEIR A-CUT VALUES STAYS STABLE REFLECTS IN THE SPATIAL DISTRIBUTION OF THEIR POINTS. MAPS ON THE LEFT SHOW THEIR ACTUAL ADDRESS POINTS AND FUZZY MODELS WITH BAR CHARTS ON THE RIGHT INDICATE THE OPTIMAL A-CUT FOR EACH THRESHOLD..... 113
FIGURE 4.25: TWO EXAMPLES OF VILLAGES WITH GRADUAL TRANSITION BETWEEN THEIR A-CUTS AS MOST OF THEIR ADDRESS APPEAR MORE CLUSTERED IN THE MIDDLE. MAPS ON THE LEFT SHOW THEIR ACTUAL ADDRESS POINTS AND FUZZY MODELS WITH BAR CHARTS ON THE RIGHT INDICATE THE OPTIMAL A-CUT FOR EACH THRESHOLD........ 114
FIGURE 4.26: TWO EXAMPLES OF VILLAGES WHICH HAVE SUDDEN DECLINE IN THEIR A-CUTS AS PART OF THEIR ADDRESSES ARE QUITE SPREAD OR DEFUSE. MAPS ON THE LEFT SHOW THEIR ACTUAL ADDRESS POINTS AND FUZZY MODELS WITH BAR CHARTS ON THE RIGHT INDICATE THE OPTIMAL A-CUT FOR EACH THRESHOLD..... 115
FIGURE 4.27: PLOTS SHOWING FUZZY MEMBERSHIP VALUES FOR EACH PERCENTAGE IN EACH DATA FOR TWO VILLAGES, WHICH HAVE SIMILAR A-CUTS PROFILE AS A-CUTS ARE ALMOST IDENTICAL IN THE BS76, POST & ALL DATA, BUT SMALLER IN THE POI. FULL-SIZE VERSIONS OF THE PLOTS ARE IN APPENDIX (4). 118
FIGURE 4.28: PLOTS SHOWING FUZZY MEMBERSHIP VALUES FOR EACH PERCENTAGE IN EACH DATA FOR TWO VILLAGES, WHICH HAVE SMALL VARIATIONS BETWEEN THEIR A-CUTS. FULL-SIZE VERSIONS OF THE PLOTS ARE IN APPENDIX (4). 118
FIGURE 4.29: PLOTS OF FUZZY MEMBERSHIP VALUES FOR EACH PERCENTAGE IN EACH DATA FOR TWO SETTLEMENTS THAT HAVE QUITE DIFFERENT A-CUT PROFILES. FULL-SIZE VERSIONS OF THE PLOTS ARE IN APPENDIX (4).119
FIGURE 4.30: PLOTS OF FUZZY MEMBERSHIP VALUES FOR EACH PERCENTAGE IN EACH DATA FOR TWO VILLAGES THAT HAVE DENSITY VALUES NEARLY EQUAL IN THE BS76 & POST DATA AND ALMOST DOUBLED IN ALL DATA COMBINED AND EXTREMELY SMALL IN THE POI DATA. FULL-SIZE VERSIONS OF THE PLOTS ARE IN APPENDIX(4). 120
FIGURE 4.31: PLOTS OF FUZZY MEMBERSHIP VALUES FOR EACH PERCENTAGE IN EACH DATA FOR TWO SETTLEMENTS THAT IRREGULAR PATTERN OF VARIATION IN THEIR DENSITY VALUES. FULL-SIZE VERSIONS OF THE PLOTS ARE IN APPENDIX (4). 121
FIGURE 5.1: PSEUDO-CODE FOR IDENTIFYING THE CENTRE OF THE A-CUT RASTERS........ 134
FIGURE 5.2: PSEUDO-CODE FOR APPLYING THE TRAVELLING SALESMAN PROBLEM ON HARD VERSUS FUZZY LOCATION. 136
FIGURE 5.3: MAP SHOWING THE CENTRE POSITIONS IN THE HARD AND FUZZY APPROACHES........ 137
FIGURE 5.4: MAPS OF TWO CASES WHEN THE FUZZY CENTRES ARE REPLACED TO SMALLER A........ 138
FIGURE 5.5: MAPS SHOWING EXAMPLES OF DISREGARDED VILLAGES WHERE THE CENTRE FALL OUTSIDE THE A-CUT SURFACE. 138
FIGURE 5.6: AN EXAMPLE OF THE TSP RESULTS (TOUR PATHS) OF A SUBSET OF 13 VILLAGES BASED IN THE HARD (UPPER MAP) AND FUZZY (LOWER MAP) LOCATIONS. 142
FIGURE 5.7: AN EXAMPLE OF THE TSP RESULTS (TOUR PATHS) OF A SUBSET OF 18 VILLAGES BASED IN THE HARD (UPPER MAP) AND FUZZY (LOWER MAP) LOCATIONS. 144
FIGURE 5.8: AN EXAMPLE OF THE TSP RESULTS (TOUR PATHS) OF A SUBSET OF 20 VILLAGES BASED IN THE HARD (UPPER MAP) AND FUZZY (LOWER MAP) LOCATIONS. 145

List of Tables

Table 3.1: Background information and history for some settlements........40
Table 3.2: The structure of the British Standard Addresses (BS7666)........43
Table 3.3: The structure of the Postal address type (Royal Mail Delivery Point Address).........44
Table 3.4: Summary of the three addresses theme details........45
Table 3.5: List of the rural settlements in the study area with counts of the addresses identified. 50
Table 3.6: List of the rural settlements that are either misplaced or missed from the data. 51
Table 4.1: Values of the correlation coefficients (r) and the root mean squared error (RMSE) between each pair of the data sources........78
Table 4.2: Results of the containment analysis for the BS76 data. For all considered threshold, the optimum a-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these a-cuts (points/km^2). 83
Table 4.3: Results of the containment analysis for the POST data. For all considered threshold, the optimum a-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these a-cuts (points/km2). 90
Table 4.4: Results of the containment analysis for the POI data. For all considered threshold, the optimum a-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these a-cuts (points/km2). 99
Table 4.5: Results of the containment analysis for all data combined. For all considered threshold, the optimum a-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these a-cuts (points/km2). 107
Table 4.6: Lists of settlement names and numbers of their address points that exist in each datasets. 117
Table 5.1: Table showing four variations of the insertion heuristic and their strategy to choose which city to insert. 131
Table 5.2: Lists of the considered settlements for the TSP with their equivalent a-cuts. 134
Table 5.3: Distance between the centre locations in the two approaches........140
Table 5.4: Comparison of the tour lengths for the possible subsets of settlements in both approach. 141
Table 7.1 List of Village names and ID included in the tours........245
Table 7.2: Comparison of the tour lengths and paths for the possible subsets of settlements in both approach........265

List of Publications

Peer reviewed conference proceedings

FISHER, P.F. & ALMADANI, F., (2011). Fuzzy Geographical Buffers Revisited, *Proceedings of the 19th GIS Research* , 29th April, 2011 2011, University of Portsmouth, pp. 147-152.

ALMADANI, F., FISHER, P. & JARVIS, C., (2012). On the Fuzzy Distance Between Fuzzy Geographical Objects, B. ROWLINGSON and D. WHYATT, eds. In: *The 20th annual GIS Research UK (GISRUK) conference*, 11th - 13th April (2012).

ALMADANI, F., FISHER, P. & JARVIS, C., (2013). Modelling Spatial Extents of Village Territories from Postal Address Records,. In: *The 21st annual GIS Research UK (GISRUK) conference*, 3rd - 5th April (2013).

ALMADANI, F., FISHER, P. & JARVIS, C., (2014). Modelling the Fuzzy Footprints of Villages from Postal Address Records, *Eighth International Conference on Geographic Information Science*, September, 23-2.Vienna 2014.

Abbreviations

OS	Ordnance Survey, Great Britain's national mapping agency
POI	Points of Interest
PAF	Royal Mail’s Postcode Address File
GB	Great Britain
VOA	The Valuation Office Agency
MOWPAs	Multi-occupancies without a postal address
MR	Royal Mail’s new Multi-Residence
OWPAs	Objects without a postal address
BS76	The British Standards Institute form of address
POST	Postal address related to Royal Mail Delivery Point Address
LLPG	The Local Land and Property Gazetteer
ONS	The Office for National Statistics
CP	Civil Parishes
NCP	Non Civil Parish
BS7666Ad_3	BS7666 Address Locality
PostalAd_9	Postal Address Dependent Locality
GIScince	Geographical Information Science
GIS	Geographical Information Systems
TSP	Travelling Salesman Problem
UK	The United Kingdom
r	Correlation coefficient
RMSE	Root mean square error

Chapter 1 Introduction

1.1 Overview

It is believed that uncertainty is indeed part and parcel of human nature and their surrounding world, as they cannot always be sure, consistent and precise in themselves nor in their interaction with the surrounding environment. It is probably very difficult to assume exact knowledge and be very certain about many geographical objects and spatial relations, for instance, if one is directed to go "down the road", "up the hill", or "to city centre", or, to be more specific, to "East Anglia", "the West End" or even "Bloomsbury". In fact, these places are often described as indeterminate, hard to define, ill-defined, or vague; and any two individuals will not completely agree about the exact location of these places or even their extents. However, people are generally tolerant and prepared to conceive of and accommodate the world in terms of that vagueness (Montello *et al.* 2003; Fisher and Robinson, 2014). It should be further notice of that people's conceptions of vagueness are not limited to such vernacular regions, but also extend to other well-defined geographical places. For example, in their social communication they refer to London most likely without being concerned about the exact location (i.e. in the United Kingdom or in Canada) or the precise nature of its boundary (as the capital of the United Kingdom or just the capital of England).

It is not just that people effectively interact with vagueness in their life systems, but they are further participating in a number of ways to provide informative sources about geographical places. This can be illustrated in two broad examples from social media: (1) individuals tend to refer to vague places or use vernacular languages when they are talking about a particular place in Twitter and Facebook messages; and (2) there are some websites, such as Flickr and Geograph, which allow users to upload and locate their photographs on the Earth's surface by latitude and longitude, which are not necessarily well defined places. Goodchild (2007) explains these activities as an explosion of interest in using the Web to create, assemble, and disseminate geographic information provided voluntarily by individuals. He terms this volunteered geographic information (VGI), a special case of the more general Web phenomenon of user-generated content. Despite the debate on VGI quality and reliability, it has recently received a great deal of attention in Geographical Information Science (GIScience). It has the potential to be a significant source of geographers' understanding of the surface of the Earth (Goodchild, 2007). Even more importantly, different types of VGI data sources have been researched to provide insights about the location and extent of vague regions and vernacular place names (e.g. Arampatzis *et al.* 2006; Jones *et al.* 2008; and Twaroch *et al.* 2008 a & b); the next chapter will elaborate on this in more detail.

However, there is a controversial view that relates this vagueness originally to issues related to naming things. Varzi (2001), for instance, considers this as an exclusively semantic problem caused by the naming of things and argues against the existence of vague geographical objects with boundaries that are themselves a matter of degree. To an extent, this cannot be invalidated due to the prominence of vernacular place names and vague terms that widely exist in everyday statements and conversation. From this, it is not possible to clearly define the geographical phenomena (objects, relation and process) in any meaningful way that does not involve an arbitrary cut off (Fisher 2000, Fisher and Robinson, 2014). That consequently indicates the impossibility of conceiving vagueness in geographical phenomena in terms of absolute right versus absolute wrong, and being intolerant of any divergence from that norm. The concepts of vagueness, need not be discussed here, but are comprehensively reviewed in Chapter 2.

In contrast to this wide recognition of uncertainty and vagueness, most analytical functions and conventional modelling techniques developed in GIScience ignore these issues related to the concepts they are storing and analysing. However, researchers recently have endorsed the need to develop modelling techniques that adequately handle vagueness and imprecision in geographical information. The research work in this area is highly informed by: (1) the mathematical and computational formalisation of vague objects stored in Geographical Information Systems (GIS), - (e.g. Cheng *et al.* 2004; Fisher *et al.* 2004; Dilo *et al.* 2007); and (2) gathering information about vaguely defined regions and offering approaches to approximate the regions' extents (e.g. Montello *et al.* 2003; Arampatzis *et al.* 2006; Jones *et al.* 2008). Hence, on the pages that follow, this thesis contributes to the literature by considering one example of a vague region, rural settlement or village. One could argue that a settlement in general is a relatively precise geographical concept in terms of its definite name and precise location. However, rural settlements or villages tend to be vaguely defined, as it is not always easy to consider whether a person is inside the village or outside the village. This is because there are some situations where a person could be partly within the village, partly outside it, or even in some places where it may require some thought and even indecision as to whether the person is out or in.

1.2 Research Aims & Objectives

This research aims to:

- Evaluate and validate the usefulness of using fuzzy set theory in determining the spatial extent (footprint) of vague geographical features in rural areas (villages).

- Explore the implications of applying the travelling salesman problem as a routing and navigation application on the fuzzy model versus the traditional crisp model.

The objectives of this research are:

- Identification of rural areas and villages with insights into the conceptualised nature of vagueness and other possible sources of uncertainty will be provided.

- The spatial extents of these villages will be traced, and models of the fuzzy footprints of these regions will be developed.

- The consequence of the developed methods will be assessed by comparing the traditional (Boolean) and fuzzy model in a routing application –the travelling salesman problem.

1.3 Rationale and Significance of the Research

This research sets out to defend the view that vagueness is fundamental to geographical phenomena and in analysing those phenomena it must be addressed. The aim of this research project has therefore been to develop a method of modelling vague geographical places (i.e. rural settlements in the Hinckley and Bosworth District) with a view to assessing possible application in the resulting fuzzy model of those vague places.

There are several important areas where this study makes a unique and original contribution to the field of GIScience, including:

- It explores how village territories or extents are defined in different data sources and thereby it provides insight into different aspects of uncertainty associated with the concept of individual places. These data include: a formal definition of village boundaries, but these frequently relate to the Parish, a historical ecclesiastical boundary and two address databases (formal and contributed data).

- Most of the research on modelling vague places are based on vernacular geography like empirical data engaging human subjects (Montello *et al.* 2003; Lüscher and Weibel, 2013) or using information from the web - tag points (Goodchild *et al.* 1998, Hollenstein and Purves, 2010) as a source to identify the vague regions. This study employs definite features (geocoded address points) to model vagueness in place names and explore further aspects of uncertainty beyond the vagueness issue. In doing this, it explores the ambiguity and discord within and between data describing villages.

- This research offers a way ahead by considering further analysis once fuzzy regions have been modelled, a view which has been ignored in most of the previous work on modelling vague regions.

1.4 Thesis Structure

The research in this thesis is explained in several self-contained chapters, including this introductory chapter about the aim of the research. This is followed by a discussion of the fundamentals, which describe a particular aspect of the research study, and finally closes by providing some conclusions and recommendations for future work. This section provides a brief description of these chapters, as follows:

- Chapter 2 reviews the literature to provide a conceptual framework of the fundamental aspect of this research. This starts by reviewing the nature of vagueness and its formalisation within GIScience. Then it discusses the approaches used to address the problem of vagueness and elaborates more on the basics of fuzzy set theory. Finally, it explores the previous work in connection with vague regions.

- Chapter 3 sets the scope of the research in terms of the geographical scale of the study area, in addition to a detailed explanation of the data used to identify villages as vague objects.

- Chapter 4 formally describes the proposed methods for approximating the fuzzy footprint for villages. In addition, it further presents and discusses the results obtained from these methods.

- Chapter 5 explains the implementation of applying some heuristic to solve the travelling salesman problem in fixed and indeterminate or fuzzy locations. The results acquired from this approach are also presented and discussed in this chapter.

- Chapter 6 critically reflects on the overall research process along its various stages. It summaries the entire results of the research, assessing whether these finding successfully achieve the research objectives. Then the attention moves to the development and implementation of the suggested approaches, including the limitations and some outstanding methodological issues that require further research.

- Finally, Chapter 7 closes the thesis with some conclusions by summarising the key findings of the research and highlighting the recommendations for future research in the field of GIScience.

Chapter 2 Literature Review

2.1 Introduction

It is becoming increasingly difficult to ignore the uncertainty associated with Geographical Information Science (GIScience). There is rich ground for debate about uncertainty in GIScience (e.g. Fisher, 1999; Goodchild, 2000; Longley *et al.* 2001; Klimesova, 2006; Comber *et al.* 2006; Fisher *et al.* 2006; Vullings *et al.* 2007). The debate, which stems from the large number of different sources and forms of uncertainty, has highlighted problems from measurement errors to issues at every stage in the handling of the geographic data from acquisition through to final use.

Due to the longitudinal discussion about the types of uncertainty, a comprehensive account of this subject is beyond the scope of the present work; here, the consideration is only on one type known as vagueness or fuzziness. Thus, this section and the subsequent sections are devoted to provide an overview of the relevant aspects of vagueness and its involvement in GIScience.

The aim of this chapter is to provide a conceptual framework of the fundamental aspect of this research. Section 2.2 explores the nature of vagueness and its formalisation within GIScience. Then Section 2.3 discusses some of the approaches used to address the problem of vagueness. After that, Section 2.4 presents on theoretical framework of the fuzzy sets theory; the approach used here to address the issue of vagueness. Next, Section 2.5 reviews the role of fuzzy set theory in GIS. Sections 2.6 and 2.7 review further related work in connection with modelling vague regions. Finally, Section 2.8 summarises the covered literature and highlight new insight to the research topic.

2.2 Vagueness in Geography

Although differences of opinion still exist, there appears to be some agreement about exactly what is meant by the term vagueness. According to a definition provided by Worboys (2001), it refers to a particular type of imprecision where it is difficult to decide, in borderline cases, whether a concept or situation applies or not. That means vagueness is a problem about where to draw a demarcation line between cases. Fisher (2000), in his seminal article, was apparently the first to point out the Sorites Paradox, as a foundation of fuzzy set theory, is used to test whether an object or a concept is vague. A well-known example of this argument concerns the cut off point of which a few sand grains make a heap. Other examples, more geographical, involve scales of size and consider issues like when a village becomes a town, when a multi-storey house becomes a skyscraper, when a hill becomes a mountain, or when trees become a forest.

It is now generally recognised that difficulties arise when an attempt is made to identify the category and extent of almost all geographical features. There are some words or expressions used in the literature to describe such features; they are said to be hard to define, ill-defined, indeterminate, or vague. Hollenstein (2008) argues that there are two broad categories of vague geographic entities. The first type relates to the majority of natural geographic phenomena that are spatially ill-defined, such as vegetation zones or soil types. The second type expresses human conceptions about vague places and their extents, for example, aboriginal territories or urban neighbourhoods. This view is supported by Erwig and Schneider (1997), who consider fuzziness as an intrinsic feature of an object itself; each object may either lack a precisely definable border (first type) or by its nature lack the ability to be precisely defined (second type). Furthermore, Mesgari *et al.* (2008) indicate two aspects of fuzziness related to many spatial phenomena: first, either the spatial features or at least their effects are usually without determinate boundaries (fuzzy boundaries); second, every location can be categorized into different classes simultaneously with different degrees of certainty (fuzzy classifications). In a similar way, Fisher *et al.* (2004) use the terms “semantic” and “epistemic” to identify vague features in a philosophical sense. A mountain, for example, is vague object either because of human perception dividing a landscape into features called mountains, or the mountain actually exist as vague objects.

Furthermore, the vast majority of people in everyday life think and communicate about the world in terms of vague concepts. Though they talk about particular geographical areas, they typically use vernacular geographical terms rather than scientific geographical vocabularies when describing regions and spatial relations (Montello *et al.* 2003). In other words, people often refer to geographical contexts without clear definitions of where or what they are. There is also an inconsistency with the interpretation of geographical features because of the vague and imprecise nature of place names. Terms such as "Midwest" in the United States and "Midlands" in the United Kingdom have no formal geometric boundaries and may be interpreted differently by different people (Arampatzis *et al.* 2006). Equally, as Jones *et al.* (2008) state, there are places whose names have been adopted for administrative purposes, but for which the administrative boundary differs from many people's perception of the extent of the place. Indeed, there are many other instances of such vague or vernacular terms used at different levels of geographical scales, such as downtown, city centre, the west end, the rocky mountains, the south (the north), near the park, up the hill, down the road and so forth. As a result, it is an important challenge to develop techniques to approximate the extent of such vague places in a manner that enables them to be interpreted intelligently for both social and scientific purposes.

2.3 Approaches Used to Address Vagueness

A number of methods have been suggested to address the problem of vagueness. According to Williamson (1996) one main approach relies on *many-valued logic*, which replaces the dichotomy of truth and the falsity in classical logic by manifold classification. One of its widely used implementation is fuzzy set theory, which is generally considered the primary method to handle vagueness issues in GIS and spatial databases (Fisher, 2000; Erwig and Schneider, 1997). Section 2.4 elaborates the conceptual details of fuzzy set theory.

There are only a few alternative approaches available that are not based on fuzzy set theory. These might include qualitative approaches, theories of supervaluation and rough sets. Qualitative approaches are accounts of the *egg-yolk model* (Cohn and Gotts, 1996), which is an extension of the RCC model (Region Connection

Calculus) defined by (Randell *et al.* 1992). It describes a vague region as a pair of crisp regions, one enclosing the other. The 'yolk', the inner region, represents the certain part of the vague region, the 'white', the outer region, represents the broad boundary, and the white and yolk together form the egg that is the full extent of the vague region (Dilo *et al.* 2007). Likewise the 9-intersection model (Clementini and di Felice, 1996) is extended to deal with such broad-boundary regions. Both accounts introduce a broad border region that defines the area that partially but not fully belongs to the core region. However, both of them fail to consider the challenge of spatial vagueness like gradual transition.

The basic idea underlying *supervaluation* is that a vague predicate distinguishes entities to which it definitely applies (its positive extension), entities to which it definitely does not (its negative extension) and a penumbra of the predicate (when some entities are indefinite if the predicate applies)(Kulik, 2003). In this theory, a proposition involving a vague term is *supertrue* and *superfalse* if and only if *all* of admissible different ways in which the vague term could be made precise come out true or false, respectively. All other cases fall into a truth-value gap, where propositions are neither true nor false. In other words, the assignment of truth value for all interpretations is a supervaluation, such that anything true in all precisifications is supertrue; anything false is superfalse (Agler, 2010).

Rough sets in contrast can be represented by a pair of classical sets, known as the lower and upper approximation. The lower approximation consists of all elements that certainly belong to the set, whereas the upper approximation consists of all elements that possibly belong to the set. This approach has been discussed in greater detail in Pawlak (1991, 2002). It is worth noting that rough set theory has an overlap with many other theories dealing with imperfect knowledge, such as fuzzy set theory and others. Ahlqvist and co-authors have demonstrated how rough fuzzy sets could be used to integrate different geographical classifications (Ahlqvist *et al.* 2003).

In addition, it should be underlined that methods based on rough set theory have been developed increasingly in GIScience. Worboys (1998a), for instance, has provided a formal framework for reasoning about imprecision in spatial data. As he used rough set to approach the uncertainty associated with information stored at multiple resolutions where the discreteness of entities is progressively improved with greater resolution. This work has further extended to consider both spatial and semantic dimensions in modelling geographical vagueness (Worboys 1998b). Raubal and Worboys (1999) have also applied rough sets theory to model imprecise knowledge in the wayfinding process (in a special scenario of wayfinding in a building that is finding one's way through an airport). Worboys and Clementini (2001) describe a variety of different techniques of three valued logics for handling the integration of imperfect spatial observations. They develop a rough set representation for a region with broad boundary resulting from the observation of a crisp region under conditions of granularity (Imprecise observation of a spatial phenomenon). Indeed, Worboys along with other researchers have demonstrated how rough sets are used widely for modelling spatial and semantic components of geographic information, which commonly assume a variety of vague interpretations, including inaccuracy, imprecision and vagueness (Duckham and Worboys, 2001; Worboys, 2001; and Duckham *et al.* 2001).

2.4 Basic Concept of Fuzzy Set Theory

The principle of fuzzy set theory was first demonstrated extensively by Zadeh (1965). Since his published work, fuzzy set theory and its applications have been well documented in the information science literature in general (Klir *et al.* 1997; Klir and Yuan, 1995; Pedrycz and Gomide, 1998; Zimmermann, 2001; Hayward and Davidson, 2003) and in GIS in particular (Leung, 1983; Altman, 1994; Schneider, 1999, 2000; Robinson, 2003).

Fuzzy set theory has been suggested as alternative to traditional set theory with regard to Boolean logic when dealing with uncertainties associated with many real world phenomena. Consequently, various areas as diverse as engineering, psychology, GIS, artificial intelligence, medicine, ecology, decision theory, pattern recognition, information retrieval, sociology and meteorology produce successful applications based on this theory (Kaymaz, 1995). Hence fuzzy set theory has received more and more recognition as a valid and useful extension of classical set theory. This section aims to provide an in-depth review of the basic concepts of fuzzy set theory.

2.4.1 Fuzzy Sets

As mentioned above, fuzzy set theory is considered as an alternative to the Boolean (crisp) set theory, where members of one set are separated from another with complete certainty and a precise boundary. So, a sharp unambiguous distinction exists between the members and non-members. In other words, each individual is either definitely a member of the set or definitely not a member of it. In daily lives, however, many classification concepts used and expressed in natural languages describe sets that do not exhibit this characteristic. Examples include the sets of *tall people*, *expensive cars*, *close driving distances*, *high salary*, *numbers much greater than four*, *sunny days*, *coast line* and many others. In all these sets there is a gradual transition between their membership and non-membership that leads to imprecise boundaries between them. Therefore, fuzzy set theory is a natural and useful way to characterise such concepts (Klir *et al.* 1997; Klir and Yuan, 1995).

The major motivation behind introducing fuzzy set theory is to represent such vague concepts. An individual element in a fuzzy set might possess some uncertainty, and thus its membership is just a matter of degree. That is to say that the *degree of membership* of an individual element in a fuzzy set shows the *degree of compatibility* or the *degree of truth* of the individual with the concept expressed by the fuzzy set (Zimmermann, 2001). In the instance of the set of “tall people”, a person is a member of the set to the degree to which he or she meets the concept of being tall.

In essence, fuzzy set theory is an extension and generalization of the classical (crisp/ Boolean) sets theory, which assigns a value of either 0 or 1 to each individual element in the universal set. Thereby, it discriminates between members (who have the value of 1) and non-members (who have the value of 0) of the considered set. In contrast, the values that are assigned to each element in the fuzzy set fall within a specified range which determines the membership grade of these elements; the larger the values, the higher the membership grade. This grade can be any real number in the interval [0, 1], where the 0 indicates absence (no membership) and the 1 indicates complete membership. Fuzzy sets are formally defined, in relation to the crisp set, as follows:

> **Definition 1:** Crisp Set versus Fuzzy Set [Zadeh 1965; Schneider 2000]
>
> Let X be a classical (crisp) set of objects, then the membership of the classical subset A of X, denoted by A (x), can be defined by the *characteristics function* as:
>
> $A: X \rightarrow \{0, 1\}$ such that for all $x \in X$: A (x) = 1 if and only if $x \in A$ and
>
> $A\ (x) = 0$ otherwise.
>
> This function can be generalized by assigning to each element x of X a number A (x), sometimes μ (x), in the closed interval [0, 1] that indicates the degree of membership of x in A, and thus the membership function has the form:
>
> $\mu(x): X \rightarrow [0, 1]$.

There are several ways to represent the membership function of fuzzy sets, as each set is uniquely defined. In other words, each member x of the universal set X is assigned to a unique membership degree A (x) in the represented set A. Accordingly, the representations of membership functions differ from each other in the way in which these assignments are expressed. The most common representations could be as graphs, tables, lists, mathematical formulae, or coordinates in the n-dimensional unit cube (Klir *et al.* 1997).

Additionally, fuzzy sets can be outlined over discrete or continuous domains and hence each of these also has different representations (Altman, 1994). In the case of the discrete domain, a fuzzy set has elements that are ordered pairs, denoted by μ/x or (x, μ), where x is the domain value and μ is its degree of membership. For instance, the fuzzy set A representing "*the sensible number of children*" might be defined as:

A = {0.1/0, 0.3/1, 0.7/2, 1/3, 0.6/4, 0.2/5, 0.1/6} or

A = {(0, 0.1), (1, 0.3), (2, 0.7), (3, 1), (4, 0.6), (5, 0.2), (6, 0.1)}

Figure 2.1 describes this set graphically in the conventional way with the Y-axis representing the degree of membership of the corresponding elements along the X-axis.

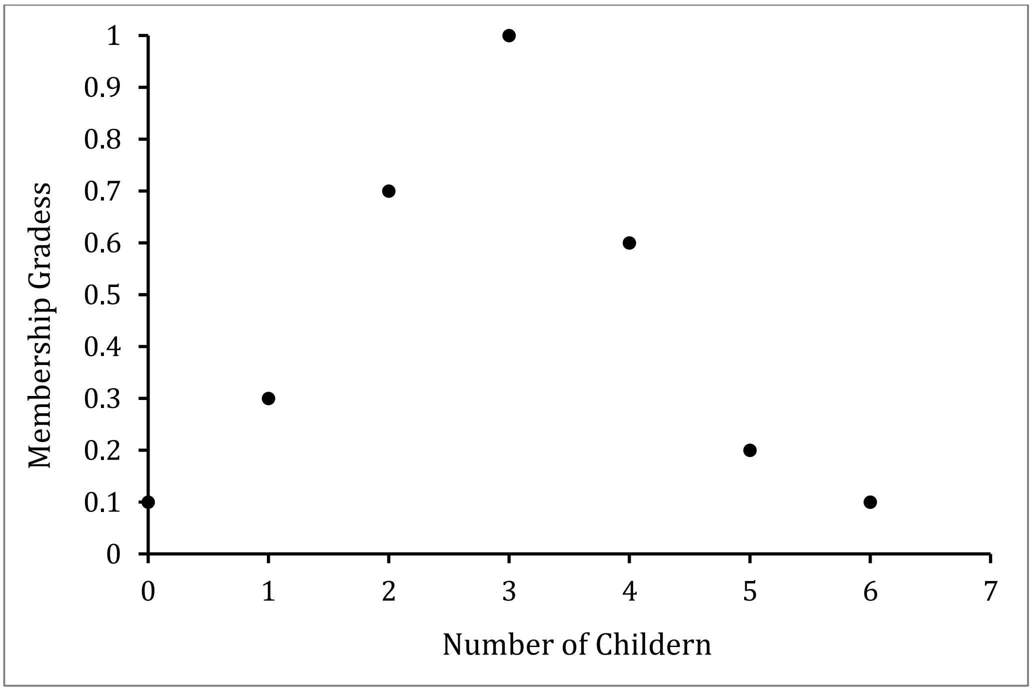


Figure 2.1: Fuzzy set over discrete domain, an example of "the sensible number of children".

When the domain is continuous the fuzzy set has a membership function that identifies a degree of membership for each value in the domain. The membership functions in this case can be manifest by the very different shapes of their graphs according to a particular application (see Figure 2.2). The triangular and trapezoidal shapes however, have gained wide acceptance in many fields.

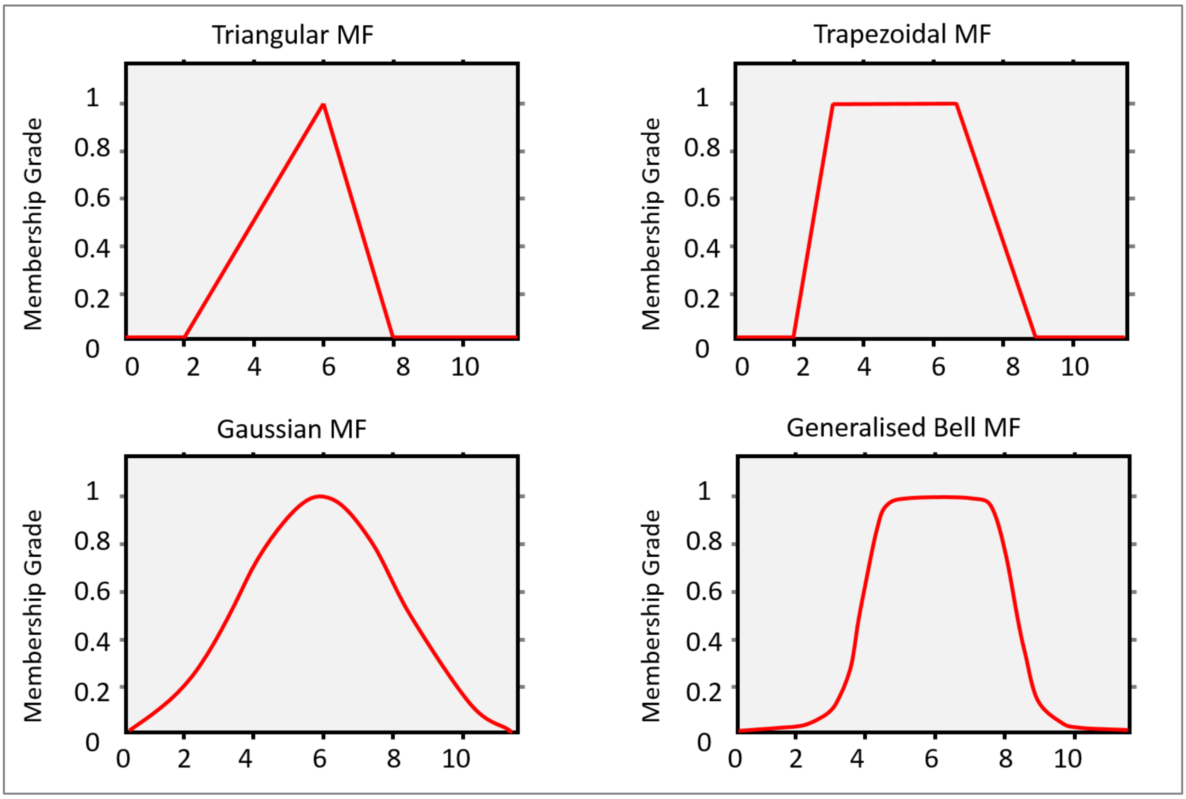


Figure 2.2: Formulations of Membership Function (MF) – (modified after Saggaf and Nebrija, 2003).

Hence the membership function for a fuzzy set A representing "*distance of about 5km*" might be defined as:

$$A = \begin{cases} \dfrac{x-3}{2} & \text{when } 3 \leq x \leq 5 \\ \dfrac{7-x}{2} & \text{when } 5 \leq x \leq 7 \\ 0 & \text{otherwise} \end{cases}$$

Figure 2.3 shows this set graphically in the conventional way with the Y-axis representing the degree of membership of the corresponding elements along the X-axis.

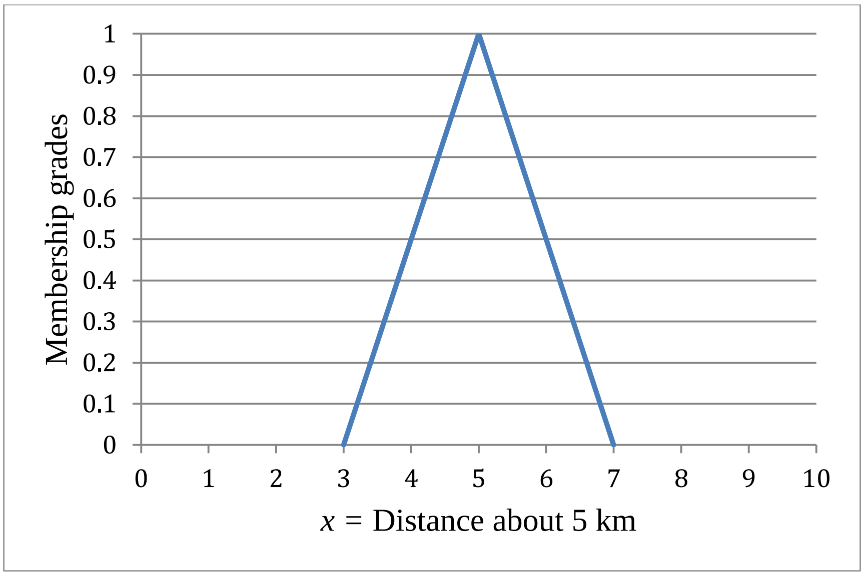


Figure 2.3: Fuzzy set over continues domain, an example of "distance about 5 Km".

Characteristics of a Fuzzy Set

There are some features used in characterizing the membership functions, which usually serve the purpose of identifying fuzzy sets in more detail. These features include the concepts of normality, height, support, convexity, concavity, and cardinality as well as α-cuts of fuzzy sets. All these characteristics are defined formally in the definitions:

Definition 2: Fuzzy Sets Properties [Klir *et al*. 1997; Pedrycz & Gomide, 1998;]

A fuzzy set A is ***normal*** if at least one point of its domain has a membership function that reaches its maximum, the value of 1. This can be seen via the formula

$$\exists\ x|X\ \ and\ \ \mu(x) = 1.$$

In another way, A is such if its membership function attains 1, that is, $\sup_x A(x) = 1$. If the supremum is less than 1, then A is called *subnormal*.

The largest value of the membership function, the supremum above, is usually referred to as the ***height*** of A. Thereby saying that a particular fuzzy set is normal is essentially identical to saying its height is equal to 1.

The ***support*** of a fuzzy set A within a universal set X is the crisp set that contains all elements of X that belong to A to a non-zero degree; formally,

$$Supp(A) = \{x \in X |\ A(x) > 0\}.$$

In contrast, the ***core*** of a fuzzy set *A* is the set of all elements of X for which the degree of membership in *A* is 1; more formally,

$$Core\ (A) = \{x \in X |\ A(x) = 1\}.$$

The ***cardinality*** of a fuzzy set *A* in a finite universe *X*, denoted by Card (*A*) – sometimes $|\tilde{A}|$, is defined as

$$Card\ \ (A) = \sum_{x \in X} A(x).$$

Definition 3: Convexity and Concavity [Klir *et al.* 1997; Pedrycz & Gomide, 1998;]

A fuzzy set *A* is ***convex*** if its membership function is such that

$$\mu_A(\lambda x + (\lambda - 1)y) \geq \min[\mu_A(x), \mu_A(y)] \qquad \forall\ x, y \in X,\ \lambda \in [0, 1].$$

By the same token, a fuzzy set *A* is ***concave*** if the corresponding membership function satisfies the relationship

$$\mu_A(\lambda x + (\lambda - 1)y) \leq \max[\mu_A(x), \mu_A(y)] \qquad \forall\ x, y \in X,\ \lambda \in [0, 1].$$

Definition 4: α-Cuts of Fuzzy Sets [Klir & Yuan, 1995; Klir *et al.* 1997; Pedrycz & Gomide, 1998;]

The *α-cut* of a given fuzzy set *A*, denoted by A_α, is a Boolean set containing all the elements of the universal set X whose membership grades exceed the threshold level α. In other words, it is a restriction of membership degrees that are greater than or equal to (or only greater than in case of a more restricted variant, a *strong α-cut,* denoted by $A_{\alpha+}$) some chosen value α in [0, 1]. Formally,

$$A_\alpha = \{x | A(x) \geq \alpha\}$$

$$A_{\alpha+} = \{x | A(x) > \alpha\}.$$

2.4.2 Fuzzy Sets Operations

As in the traditional sets, there are standard operations which can usually be performed on fuzzy sets. Some of these are complement, intersection, and union. In the context of natural languages, each of these operations reflects discrete meanings, namely: the terms *not*, *and*, and *or* respectively. Generally, the standard operations of complement, intersection, and union of two fuzzy sets A and B can be defined as follows:

Definition 5: Operations on Fuzzy Sets [Klir & Yuan, 1995; Klir *et al.* 1997; Zimmermann, 2001;]

The ***complement*** of a fuzzy set A, denoted by $\bar{A}$, on the universal set X is another fuzzy set on X that inverts, to some extent, the degree of membership associated with A. For each $x \in X$ although $A(x)$ expresses the degree to which x belongs to A, $\bar{A}(x)$ expresses the degree to which x does not belong to A.

$$\bar{A}(x) = 1 - A(x) \qquad \forall\, x \in X$$

Consider two fuzzy sets A and B, defined on the universal set X. Then their ***intersection*** is a new fuzzy set, denoted by $A \cap B$, which is defined by the membership functions through the formula

$$A \cap B(x) = min[A(x), B(x)] \qquad \forall\, x \in X$$

Again consider two fuzzy sets A and B, defined on the universal set X. Then their ***union*** is a new fuzzy set, denoted by $A \cup B$, which is defined by the membership functions using the formula

$$A \cup B(x) = max[A(x), B(x)] \qquad \forall\, x \in X.$$

2.4.3 Fuzzy Numbers

There are many quantifiable expressions used in daily life that do not lend themselves to being characterized in an absolutely precise manner (Klir *et al.* 1997; Schneider, 2000). There are an array of examples including "*about 9 o'clock*", "*below 100*", "*around 2:30*", "*approximately £ 5*", "*nearly 74 years*" and others. Without a shadow of a doubt, these instances have a quantitative meaning (number values) and a linguistic modifier (e.g. approximately, nearly, or around), which can give rise to uncertainty in grasping the exact meaning. Therefore, these can be seen as fuzzy expressions because they include central values and some approximated number values on either side of these central values. As a result of this fact the concept of such statements can be captured by a fuzzy set defined on the set of real numbers. Its membership function should assign the degree of 1 to the central value and degrees to other numbers that reflect their proximity to the central value and thus decrease from 1 to 0 on both sides. These kinds of fuzzy sets are known as *fuzzy numbers* which have attained great prominence in many applications such as decision-making, approximate reasoning, fuzzy control, describing complex systems and statistics with imprecise probabilities (Klir and Yuan, 1995; Pedrycz & Gomide, 1998).

So a fuzzy number is a fuzzy set defined on the domain of real numbers that must be a *convex normalized* fuzzy set (Altman, 1994; Zimmermann, 2001). Definition 6 illustrates the formal notions of fuzzy numbers. In the context of arithmetic and comparative operation, fuzzy numbers are analogous to the computation and manipulation of the ordinary numbers. Although membership functions are of a great variety of shapes that are possible for representing fuzzy numbers, the most common is a trapezoidal shape for its simplicity, which is occasionally referred to as *fuzzy interval* (Klir *et al.* 1997, Zimmermann, 2001). A *triangular fuzzy number* is, of course, a special case of a fuzzy number, which is defined formally as follows:

Definition 6: Fuzzy Number [Klir & Yuan, 1995; Klir *et al.* 1997;Schneider, 2000]

A fuzzy set A on $\Re$, expressed as $A: \Re \rightarrow [0, 1]$, is a fuzzy number if and only if its membership function is such that

$$A(x) = \begin{cases} l(x) & \text{for } x \in [a, b] \\ 1 & \text{for } x \in [b, c] \\ r(x) & \text{for } x \in [c, d] \\ 0 & \text{for } x < a \text{ and } x > b \end{cases}$$

Where $a \leq b \leq c \leq d$; l is a continuous function that is monotonically increased to 1 at the point b; and r is a continuous function that is monotonically decreased from 1 at the point c.

A *triangular fuzzy number* is a fuzzy set which can be expressed in an elementary form as a triple such that: $T = (t_c, t_l, t_r)$
Where t_c is the centre – thus $\mu(t_c) = 1$; t_l and t_r are the left and right spreads – the domain widths of the triangular fuzzy number, and their membership functions decrease from 1 to 0.

Recalling the example of the fuzzy set that represents "*the distance of about 5 km*" (shown in section 2.4.1), this might be represented by the triangular fuzzy number: T_{about_5km} = (5, 2, 2).

2.4.4 Fuzzy Relations:

Fuzzy relations were introduced by Zadeh (1965) in his original paper and they are also extensively covered in the literature (Robinson, 1988; Altman, 1994; Klir *et al.* 1997; Klir and Yuan, 1995; Robinson, 2003). From the traditional perspective, *crisp relations* express solely the *presence* or *absence* of association, interaction or interconnectedness between elements of at least two sets. In contrast, the concept of *fuzzy relations* is appropriate for capturing the *strength* of association (interaction, connection). In other words, in preference to presence/ absence of association, degrees of association can be represented by membership grades in a fuzzy relation to the same extent as degrees of set membership are represented in a fuzzy set. Thus, a *fuzzy relation* is a fuzzy set defined on the universal set which is a subset of the Cartesian product of two or more ordinary sets. What is more is that its membership function indicates the strength of the relation between elements from each component set. The formal definition of fuzzy relation is given bellow.

Definition 5: Fuzzy Relation [Klir & Yuan, 1995; Klir *et al.* 1997;]

A *fuzzy relation R* between variables x and y, whose domains are X and Y, respectively, is defined by a function that maps ordered pairs in $X \times Y$ to their degree in the relation, which is a number between 0 and 1,

$$R: X \times Y \rightarrow [0, 1]$$

More generally, a *fuzzy n-ary relation R* in, $x_1, x_2, \ldots x_n$, whose domains are $X_1, X_2, \ldots X_n$, respectively, is defined by a function that maps an n-tuple $<x_1, x_2, \ldots x_n>$ in $X_1 \times X_2 \times \ldots \times X_n$ to a number in the interval,

$$R: X_1 \times X_2 \times \ldots \times X_n \rightarrow [0, 1]$$

An example is given of a binary fuzzy relation R defined on set X = {red, blue, green} and set Y = {1, 2}:

R (X, Y) = {1.0/(red, 1), 0.8/(red, 2), 0.3/(blue, 1), 0.7/(blue, 2), 0.2/(green, 1), 1.0/(green, 2)}.

As illustrated in the definition above, if the possible values of x and y are discrete, then the fuzzy relation can clearly be expressed in a matrix form. For example, the binary fuzzy relation could be represented as:

$M = [\mu_R(x, y)] \;\; x \in X, y \in Y$

2.4.5 Fuzzy Logic:

Again Zadeh (1965) introduced the concept of fuzzy logic in his seminal paper, for the sake of formalizing a mathematical approach to deal with complex or ill-defined systems. As a result a relatively new mathematical model is consider – fuzzy logic (Hayward and Davidson, 2003). There are two forms in the fuzzy literature for using the term *'logic'*. In the broad sense, it refers to a system of concepts, principles, and methods for dealing with modes of reasoning that are approximate rather than exact, whereas in the narrow sense, it refers to a generalization of the various multiple values, which is related to the area of symbolic logic (Klir *et al.* 1997).

Definition 6: Fuzzy Logic: [Klir & Yuan, 1995; Klir *et al.* 1997]

Consider a fuzzy set A, its membership function A (x), for any element x in the universal set X can be interpreted as the degree of truth of the fuzzy proposition "x is a member of A".

In contrast, an arbitrary proposition "x is F ", where $x \in X$ and F is a fuzzy linguistic expression (such as *low*, *high*, *very far*, *extremely slow*, etc.), its degree of truth may be interpreted as the membership degree A (x) by which a fuzzy set A characterized by the linguistic expression F is defined in a given context.

Davidson and Hayward (2003) present some of the power of fuzzy logic through a simple control example concerning the field of analytical chemistry; fuzzy logic incorporates imprecision and vagueness from measurement noise as well as from linguistic process descriptions to produce operational control systems.

2.5 Fuzzy Set Theory in GIScience

Great efforts have been made to introduce fuzzy concepts into GIS. This section reviews some examples of applications that have adopted the fuzzy set theory in GIS. Moghaddam and Delavar (2007) use some statistical and fuzzy logic operations in order to achieve important criteria for pipelining and generating real cost surface, and their results show that a fuzzy model is a suitable model to generate the cost surface. Lee and Lee (2006) present a probable impact of the representation of geographic boundary for the soil loss model. To do this, the Revised Universal Soil Loss Equation (RUSLE) model was facilitated at a small basin in Korea and then the fuzzy representation of geographic boundary, which is presumably a better description of soil properties in nature, was introduced into the soil factors in the 10 RUSLE. Salski (1999) focuses on two large application areas of the fuzzy set theory in ecological research, namely data analysis (in particular fuzzy cluster analysis and fuzzy kriging) and ecological modeling. Cheng (2002) presents a systematic discussion of the indeterminate nature of geographical entities and how they are represented as fuzzy objects in a GIS. Furthermore, the change detection of fuzzy objects and their uncertainties are investigated. An example of the dynamic changes of sediments along the Dutch coast is applied to illustrate the methodology. The method is also applicable in monitoring geographical entities such as natural vegetation units or land-use areas. Bielefeld (1992) investigates the use of fuzzy numbers in representing uncertainty resulting from data censoring with respect to disease progression in individuals, and in particular the spread of HIV infection and the AIDS disease through a population considered as the application area for the study investigation. Prakash (2003) examines how to address uncertainty in the process of land suitability analysis for agricultural crops by using three approaches: Analytic Hierarchy Process (AHP), Ideal Vector Approach (IVA) and Fuzzy AHP (FAHP). It has been found that the hybrid approach, FAHP, has better performance than the rest as it involves AHP techniques, fuzzy numbers, fuzzy extent analysis, and alpha cut and lambda functions. Tapia (2004) in his study to model vegetation distribution based on remotely sensed data, uses the fuzzy-c-means classifier, which has been found to be a promising tool in multivariate models. Hwang and Thill (2005) discuss how fuzzy set theory can be properly applied in modelling localities. Their result examines whether fuzziness exists in determining the location of a locality. Their study develops a fuzzy set membership function for indeterminate boundaries of localities. Verstrate *et al.* (2005) review two different approaches for representing field-based fuzzy geographic information with respect to their benefits and their drawbacks. These are an extended vector-based method using triangulated irregular networks (TIN) and an extended bitmap model. Liang and Ding (2003) propose a fuzzy multiple criteria decision-making (MCDM) algorithm based on fuzzy set theory and α-cut concept. In their study, they attempt to efficiently grip the representation and comprehension of decision-makers (DMs') opinions and the ambiguity existing in available information by utilizing triangular fuzzy numbers to aggregate the individual opinions used to convey the relevant assessment of all DMs' viewpoints and the monetary/ quantitative terms.

Liu *et al.* (2009) introduce the uncertainty field model based on the conceptual model of spatial assertions to represent the probability distribution of a point locality. Their research extends the research by Guo *et al.* (2008), and aims to develop a general and more complete positioning method that is suitable for a variety of textual descriptions. Fonte and Lodwick (2004) review the computation of areas of fuzzy geographical entities (FGEs) and consider two methods: Crisp – Rosenfeld (1984), which seems to be limited in its applicability, and Fuzzy – New Fuzzy Area Operator, which gives more information about the possible values of the area and enables the propagation of the fuzziness to the spatial extent of the entity. Guesgen and Albrecht (2000) propose a scheme for incorporating imprecise qualitative spatial reasoning with quantitative reasoning in GIS and that is not merely restricted to Euclidean geometry as they have adopted fuzzy sets theory to model qualitative spatial relations among objects. The same idea is also highlighted again by Guesgen (2002); when he looked at qualitative spatial reasoning in GIS and gives focus to the aspect of distance particularly the role of fuzzy set theory in describing the proximity (i.e. how close objects are to each other).

Having mentioned some applications of using fuzzy set theory in GIS, Robinson (1988) identifies two methods of obtaining fuzzy membership in GIS worlds:

1. The *similarity relation model* which mainly concerns data driven and involves searching for patterns within a dataset in much the same way as the traditional clustering and classification methods. The most widely implemented approaches are fuzzy c-means algorithms (or k-means clustering) (Bezdek, 1981) and fuzzy neural networks (Burrough and McDonnell, 1998; Foody, 1996).

2. The *semantic import model*, which in contrast is based on expert knowledge that specifies a formula or formulae to derive it from existing class definition (Altman, 1994; Burrough and McDonnell, 1998).

2.5.1 *Fuzzy Objects:*

The geographical data has two main structures: raster structure (which defines space as an array of equally sized cells arranged in rows and columns, each cell containing an attribute value and a location coordinate), and vector structure (which represents geographic features such as points, lines, and polygons as a set of coordinates). Vagueness is implicit in spatial data and this can be shown in both raster and vector structure.

There are many studies based on the vector representation of spatial vagueness. Dilo *et al.* (2007), for instance, provide a model for a spatial data system that can handle vague objects. This model categorizes the vague spatial objects into vague points, vague lines, vague regions and vague partitions. The authors make an interesting representation of spatial objects that possess thematic vagueness, and argue for vague regional types which cover locational vagueness, although the vague point and line types presume known (crisp) location. Schneider (1999), in contrast, defines a structure of three fuzzy spatial data types for fuzzy points, lines, and regions respectively. In his view a fuzzy point can be defined in two different ways: either as a point in two-dimensional Euclidean space with a membership grade greater than 0, or as a point with a membership function which represents the degree of proximity of this point to another reference point.

On the other hand, there are also many studies that adopt the raster representation for spatial vagueness (Altman, 1994; Verstraete *et al.* 2007). These studies introduce the concept of fuzzy regions that consider a region to be a set of points (locations) rather than being defined by its boundary as given by Hwang and Thill (2005). Each point within the fuzzy region can be interpreted as a degree to which it belongs to that region. In other words, the degree to which that point is inside or part of some features that may have a fuzzy boundary. Alternatively, it can be interpreted as the concentration of some attribute belonging to the feature at the particular point. A more detailed description of this approach (fuzzifying raster maps) can be found elsewhere for example in Guesgen and Hertzberg, (2001), Duff and Guesgen, (2002), or Guesgen *et al.* (2003).

Regardless of the data structures illustrated above, the available models of vague spatial data, particularly those related to vague regions and their topological relations, can be summarized into two groups (Dilo *et al.* 2007; Verstraete *et al.* 2007). The first group is Broad Boundary Regions (sometimes known as the Egg Yolk Models). These consider vague regions to have a homogeneous 2– Dimensional boundary rather than a 1– Dimensional boundary (i.e. inner and outer boundaries). Locations in such models all have the same degree of membership to the region. Furthermore, these models do not take account of gradual transition. The second group is recognized as a Fuzzy Spatial Object, which employs fuzzy set theory for modelling gradual changes.

Another distinction concerning the aspects of spatial vagueness models is made by Schneider (1999). He points out three general design methods: exact model, probabilistic model and fuzzy model. The exact model transfers type systems and concepts for spatial objects with sharp boundaries to objects with unclear boundaries and models both uncertainty and fuzziness in a restricted way. The second, probabilistic, model is described as subjective and depends upon the probability theory, mostly on model positional and measurement uncertainty. The fuzzy model is described as objective, which seems to be preferential, is based on fuzzy set theory and predominantly concerns fuzziness.

2.6 Modelling Vagueness in GIScience

Approaches for modelling vagueness have been proposed in many areas. This section aims to review some of the proposed works, classify them depending on the type of vagueness they convey and provide some generalisations. As indicated above, the issue of vagueness can be discussed under two broad categories; vagueness in determining the location (fuzzy boundary) and vagueness in feature definition (fuzzy class).

The first type of vagueness is the most widely addressed in the current literature, and constitutes the most original part of this report. For example, villages inherently have indeterminate boundaries and can be modelled as fuzzy regions. Crawford (2002) develops and evaluates a novel approach for linking population and environmental data (by transforming discrete village points, with associated demographic and economic attribute data) across a thematic domain, and spatially representing functional regions that partition the landscape into village territories. Mesgari *et al.* (2008) also develop a model that implements overlay functions based on fuzzy set theory to examine the usability of such functions in integrating data related to indeterminate phenomena. They present a case study of determining the suitable locations for building commercial sites for oil production in the west of Iran.

Moreover, there are many works in the fuzzy literature to identify vague regions using qualitative methods engaging human subjects. Montello *et al.* (2003), for example, conduct an empirical study that investigates people's perception about the extent of downtown Santa Barbara. Their method is based on eliciting vagueness in the boundaries in two ways: by comparing variation in boundary locations across participants and by having participants draw different boundaries to indicate their varying confidence in region membership for different parts of the area. Similarly, Mansbridge (2004) examines how people perceive ill-defined geographical space in relation to factors that might be expected to influence their conceptualization of space. The author conducted a survey at three locations in central Sheffield, investigating whether location affects perception of Sheffield City Centre. The survey described in this paper is an exploratory study of where people consider the vague area of Sheffield City Centre to be. She also investigates peoples' perceptions of the Midlands, an imprecise region on a larger scale than Sheffield City Centre.

Recent developments in the qualitative methods have led to a renewed interest in online mapping using web-based questionnaire. Water and Evans (2003 & 2008) produce a new set of tools for capturing fuzzy areas and their associated attributes through a web based mapping system. That contains a spraycan tool allows users to tag information onto diffuse areas of varying density. By way of illustration, they use the system to present the locations considered to be "high crime areas" in Leeds. Recently, a similar study suggested by Rosser and Morley (2010) offers a Web 2.0 mapping system for eliciting people's understanding of places. Their application provides a novel approach to capturing boundaries, focused on users rating areas with a spray can tool. In contrast to the work on capturing fuzzy areas described by Waters and Evans (2008), this study suggests different technologies (JavaScript, Flash and Canvas) in order to prototype and test several interfaces.

Furthermore, there is also a large volume of published studies that employs the web as a source to model the extent of vague places. For instance, Goodchild *et al.* (1998) present methods for searching digital spatial data libraries that extend the case of ill-defined geographic footprints. Such cases are common when information is catalogued or searched for using vernacular place-names, rather than the officially recognized place names which often exist in gazetteers. These methods are implemented within the framework of the Alexandria Digital Library, a project to implement a digital spatial data library for geo-referenced materials. Arampatzis *et al.* (2006) further present several steps in the derivation of boundaries of imprecise regions using the Web as the information source. They explain how to obtain locations that are part of, and locations that are not part of, the region to be delineated and also suggest methods to compute the region algorithmically. In order to evaluate the proposed approach, the authors provide a discussion of experimental results that show how well this approach works in both precise (Wales) and imprecise (East Anglia, Midlands, and South East) regions in the UK.

Another technique to automatically construct a representation of the spatial extent of neighbourhoods is introduced by Schockaert and de Cock (2007). Because of the subjective and vague nature of many neighbourhoods, they do not commit themselves to one single boundary for each neighbourhood. Rather, they represent the extent of a neighbourhood as a fuzzy footprint (set) of locations. Jones *et al.* (2008) also describe and evaluate a method that utilises knowledge acquired from the web pages to model the extent of vague places and generate their approximate boundaries. They argue that vague place names are frequently accompanied in text by the names of more precise co-located places that lie within the extent of the target vague place. Therefore, the density surface modelling of the frequency of co-occurrence of such names provides an effective method of representing the inherent uncertainty of the extent of the vague place. Interestingly this approach generates representations of vague regions that can be stored in digital gazetteers, and provides the ability to process queries that name such vague places by associating them with quantitative geographic regions.

Furthermore, Twaroch *et al.* (2008 a) discuss the acquisition of vernacular use of place names from web sources (such as Flickr and Geograph which facilitate the geo-tagging of personal resources and allow people to mark up photo collections) and their representation as surface models derived by kernel density estimators. They introduce models of vernacular place name geography on a nationwide scale, a specific geographic region on Cardiff, Wales, UK, for which automated methods to acquire the relevant data are required. Additionally, Hall (2010) develops models to represent vague spatial data quantitatively based on text mining among geotagged images and human subject experiments on using some spatial propositions. His system proves the ability to determine images' locations based on the spatial information contained in their captions.

The second type of vagueness, on the other hand, is less treated in the current literature. For examining the vagueness in feature definition Sui (1992) indicates the viability of incorporating fuzzy set theory into GIS modelling, especially in urban applications. This is proposed since the criteria used in the evaluation sometimes cannot be clearly defined, because of the gradual transition of urban land value. His study therefore presents a fuzzy cartographic model for urban land evaluation in Jining City, China. Further work described by Arnot *et al.* (2004) explore the variation of metric values when it is hard to distinguish exactly where one land cover type changes into another; in this case, the ecotone is not an abrupt transition, but has a spatial extent in its own right. The values of metrics are explored in a landscape classification, using satellite imagery and the fuzzy c-means classifier, into fuzzy sets so that every location has a degree of belonging to all classes.

Similarly, Fisher *et al.* (2004) propose an approach based on multiscale analysis, to determine the fuzzy membership of morphometric classes of landscape. They explore this idea by looking at peaks and passes in the Lake District of northwest

England. These are analysed with respect to named features in a place name database for the area, and the application of these results to visibility analysis is explored. In addition, Sicat *et al.* (2005) demonstrate fuzzy modelling of farmers' knowledge (FK-based fuzzy modelling) for agricultural land suitability classification using case spatial data sets from India. Capture of FK was through the rapid rural participatory approach. The authors distinguish two ways for FK-based modelling of pertinent spatial data into fuzzy sets, depending upon the correlation or equivalence between farmers' definitions and scientific classifications of certain land characteristics.

Moreover, Chaudhry and Mackaness (2008) introduce a methodology for automatically discerning mountain ranges in addition to the smaller hills that constitute them. The algorithm used in this study utilizes derivatives of elevation and the density of morphological properties in order to automatically identify individual hills or mountains and ranges, together with their extents. Interestingly enough, this study presents an approach of fiat boundaries for hills and ranges based on their morphological properties and prominence. Although it has not modelled the fuzziness in the output boundaries, it argues that the suggested approach can be utilized in the modelling of fuzziness.

Remarkably, Fisher and Comber along with other colleagues have pay special attention to the issue of fuzziness in land classification. Comber *at al.* (2005) argue for the preference of fuzzy classification, where an object has a membership (however small) to every class, for analysing complex ecosystems where land cover types are heterogeneous or are poorly represented by large pixels. This view is also supported by Fisher *et al*. (2006), who suggest the use of fuzzy models for environmental data, land cover mapping and landscape ecology. They develop fuzzy change matrix (change detection techniques) to accommodate descriptions and measures of sub-pixel changes (fuzzy change), rather than the Boolean approach, which depends on pixel-by-pixel comparison and accepts only binary changes. Moreover, Comber *et al.* (2008-b) employ fuzzy set theory to analyse the conceptual confusion and overlaps, with data and classifications, associated with land cover and land use semantics. They illustrate the differences between the two by analysing their embedded descriptions of land cover and land use definitions of 'forest'. In later studies, Comber *et al*. (2012 a & b) use geographically weighted approaches to describe the spatial variation in the accuracy of Boolean and fuzzy classifications of remotely sensed data. It proposes a portmanteau approach to describe Boolean land cover accuracy and fuzzy difference measures to describe the accuracy of fuzzy land cover. A similar study by (Comber *et al*. 2012 b). Yet it is quite noticeable that fuzzy set theory has proved its success in modelling uncertainty in a variety of remote sensing image classification contexts, as already seen in previous research (Arnot *et al.* 2004; Fisher 2010; Elaalem *et al.* 2011). In addition to latest work by Fisher and Tate (2015) use fuzzy set theory to derive fuzzy class memberships from the results of the UK Output Area Classification (OAC) for the city of Leicester City, UK. This work is further extend and modified to generate type-2 fuzzy sets for each Output Area (Fisher *et al.* 2014).

2.6.1 Higher Order Vagueness

Researchers have paid attention to new types of fuzzy sets – type 2 fuzzy sets (sometimes equated to higher order vagueness). Fisher (2010) and co-authors (Fisher *et al.* 2007, Fisher *et al.* 2014) for example, initiate a powerful and wide-ranging discussion on this scope. Fisher discusses fuzzy classification over a large area (land cover mapping from remotely sensed data) in more depth and proposes alternative methods of reporting fuzzy areas, namely as type 2 fuzzy sets (higher order vagueness) or as both fuzzy areas and fuzzy numbers (2010). Fisher *et al.* (2007) explore the fuzzy representation of higher order vagueness in spatial phenomena with respect to the concept of type *n* fuzzy sets and examine the spatial extent of mountain peaks in Scotland as an instance of exploring the population of geographical Type 2 sets. Cheng *et al.* (2004) also investigate the double vagueness (from space and scale) in identification of coast landscape units. They employ fuzzy set theory to describe the vagueness of geomorphic features due to the continuity in space and use statistic indicators to evaluate the vagueness derived from the scale of measurement. Fisher *et al.* (2004) also propose an approach based on the multiscale analysis of the landscape, and from that derive fuzzy memberships of morphometric classes, and explore the consequences of this by looking at regional scale morphometry in the Lake District of northwest England.

2.7 Fuzziness and Travelling Salesman Problem

Fuzzy set theory has been applied to many disciplines such as computer science, artificial intelligence, expert systems, medicine and human behaviour, pattern recognition, control engineering, operations research, management science, and so forth. Among these, one possible application (considered later in this thesis) focuses in using fuzzy methods in Travelling Salesman Problem (TSP). Traditionally, a map of cities is given to the salesman, who has to visit all the cities only once and return back to the starting point with the minimum travel distance, cost or time. It is typical in such problem to assume that the distance between two cities as a constant, whereas in reality uncertainty always exists. This where fuzzy set comes into play. Lu and Ni (2005) argue that fuzziness and randomness may coexists in traveling salesman problem in real world situations. As the route may change because of the traffic environment (fuzziness), and the speed of travel might be random at different time, weather or traffic circumstance. Therefore, they use a fuzzy random number to represent the travelling time between two cities, and define three concepts of shortest path in fuzzy random situation (expected shortest path, (α, β)-path and chance shortest path according to different optimal desire). Botzheim *et al.* (2009) also offer a novel construction and formulation of the TSP in which they take into consideration the requirements and features of its practical application in road transportation and supply chains. Fereidouni (2011) also develops a fuzzy mathematical programming methodology for solving the TSP in uncertain environments based on a fuzzy multi-objective linear programming (FMOLP) model with piecewise linear membership function for solving a multi-objective TSP in order to simultaneously minimise the cost, distance and time. Some other works have been proposed by Kumar and Gupta (2011 & 2012) for solving fuzzy assignment problems (APs) and fuzzy travelling salesman problems with different membership functions. A new algorithm that is similar to the classical assignment method has also been proposed by Dhanasekar *et al.* (2013) to solve fuzzy TSP considering the ranking method of fuzzy numbers. Similarly, Chandrasekaran *et al.* (2013) introduce a new method for solving travelling salesman problems using transitive fuzzy numbers. The authors present numerical examples to show how the optimal solution as well as the crisp and fuzzy optimal total cost provided from this technique. However, all previously mentioned studies focus mainly on the mathematical and computational aspects of the problem despite their different methods. Besides, no attempt was made to adequately discuss the possible sources of uncertainty and fuzziness associated with the locations themselves. Thus, as part of the main aim of this research is to investigate the implication of using the traditional method of the travelling salesman problem on specific and indeterminate (fuzzy) locations. In particular, unlike previous studies which tend to consider numerical examples to represent fuzziness and imprecision in the travelling salesman problem, this study utilizes the TSP occurring in real life situations proved to be vague and fuzzy model.

2.8 Summary and New Insights

The key point arising from this review is that uncertainty and vagueness exist in many areas of GIScience. Once a geographical phenomenon is identified vague, then it must be addressed. There are several approaches and theories to address vagueness, one of which is fuzzy set theory. Although extensive research has been carried out on modelling geographical features as vague or fuzzy entities (as discussed in the previous sections). This body of work can be classified under two broad divisions: the first set of research have focused on the theoretical and computational formalization of fuzzy objects and the mathematical operation that can be applied (fuzz measures of distance, ranking fuzzy numbers and so on). While other studies focused on modelling vague regions employing the web or other vernacular resources (e.g. fuzzy footprint for downtown, city centre and so forth). There are few research if any that develop the fuzzy model and adequately covers the characteristics of these objects in practical geographical context; and even the operations derived from them have not been identified. This has given an account of the reasons for the need to undertake this research, which sets out to investigate the extent to which fuzzy set theory can be used to model vague geographical entities and then employ the fuzzy model in application analysis.

Chapter 3 Villages' Identification and Data Specification

3.1 Introduction

It is an important challenge for researchers in GIScience to consider aspects of uncertainty that are mainly consequently associated with geographical places as it is generally recognised that people in everyday life think and communicate about the world in terms of vague concepts. Though they talk about particular geographical places, they typically use vernacular geographical terms rather than scientific geographical vocabularies, even when they exist, in describing regions and spatial relations (Montello *et al.* 2003). They often refer to geographical contexts without clear definitions of where or what they are. There is also an inconsistency with the interpretation of geographical features because of the vague and imprecise nature of place names. Terms such as “Midwest” in the United States and “Midlands” in the United Kingdom have no formal geometric boundaries and may be interpreted differently by different people (Arampatzis *et al.* 2006), although they are used in conversation and on road signs.

Equally, as Jones *et al.* (2008) state, there are places whose names have been adopted for administrative purposes, but for which the administrative boundary differs from many people’s perception of the place and its extent. Indeed, there are many instances of vague or vernacular terms used at different levels of geographical scale, one of which is the subject of this chapter: that is, rural settlements or villages.

In this respect, the overall aim of this chapter is to explore how village territories or extents are defined in different data sources and thereby to gain insight into different aspects of uncertainty associated with the concept of individual places, in this case villages. Understanding such qualities associated with spatial information is important to be able to use it effectively and intelligently in both social and physical sciences. The chapter sets the scene for this work by providing conceptual descriptions of rurality in general and its practical definition used in this work (Section 3.2). Then the research study area is introduced in Section 3.3. Detailed descriptions of the acquired data for the research are provided in Section 3.4. The analytical strategy applied in this study to identify villages is explained in Section 3.5. Following from that, Section 3.6 explores the origins of the ambiguity and contradiction in village place names. Finally, Section 3.7 summarises the key points of this chapter.

3.2 Terms and Definitions of Rurality

Definitions of rurality, ruralness and what constitutes a rural area are not consistent or precise. As Flanagan (2007) noted, what is considered to be a rural area in one country is very different from the notions of rurality in another country. As an example, a settlement with a population of 5,000 may be perceived as rural in England but urban in Northern Ireland (Flanagan, 2007). People perceive rurality in a number of different ways. For example, when local people are asked to categorise their own settlement, or whether they identify themselves as being rural residents, their answers will depend on their sense of identity and self (Castleden *et al.* 2010).

As a contested concept, various approaches have been proposed to define rurality based on different factors. Lienau (1973), for example, indicates four aspects of considering settlements. These are the morphological or physiognomic aspect (i.e. shape, form); the functional (as regards its own role as well as its place in the wider set of correlations); the genetic (origin and development up to the present); and the prospective or prognostic aspect (future development by means of planned initiatives or "natural" growth). In later work, Cloke (2006) lists three distinctive categories to identify rural areas. These are based on: firstly, functional perspective, dominated area by extensive land uses, notably agriculture and forestry. Secondly, a political-economic perspective which attempts to position the rural as the product of broader social, economic and political processes. Finally, a perspective in which the area is socially and culturally constructed (engenders a way of life which is characterised by a cohesive identity of cultural and moral values). It has been argued also that attempts at rematerialising the rural have come from three aspects (Woods, 2009) concerning the geographical context of rural locality, the statistical and political definitions of rural space and, finally, conceptualising the rural as a hybrid and networked space.

It is not only the varied functions and meanings attributed to rural space that have made "rurality" into an ambiguous and complex concept, but also a host of settlement terms and terms indicating types or special traits of settlements (Lienau, 1973). Woods (2011) in his book titled "Rural", discusses in depth the meaning of rural settlement from a rural geographer's perspective. Additionally, he elaborates on how these diverse meanings and terminologies have shaped the effects of the social and economic structure of rural settlements on local people`s everyday lives.

One of the more recent statistical or data-driven definitions of rurality is provided by Bibby and Shepherd (2004). They developed a typology for settlements that sought to differentiate between different types of rural and urban settlements. Their study for the UK government describes and illustrates the methodologies used and the key decisions taken in defining a classification of smaller urban areas and rural settlements in England and Wales. Their definition of rurality defines a settlement hierarchy of small villages, hamlets and isolated dwellings, as depicted in Figure 3.1.

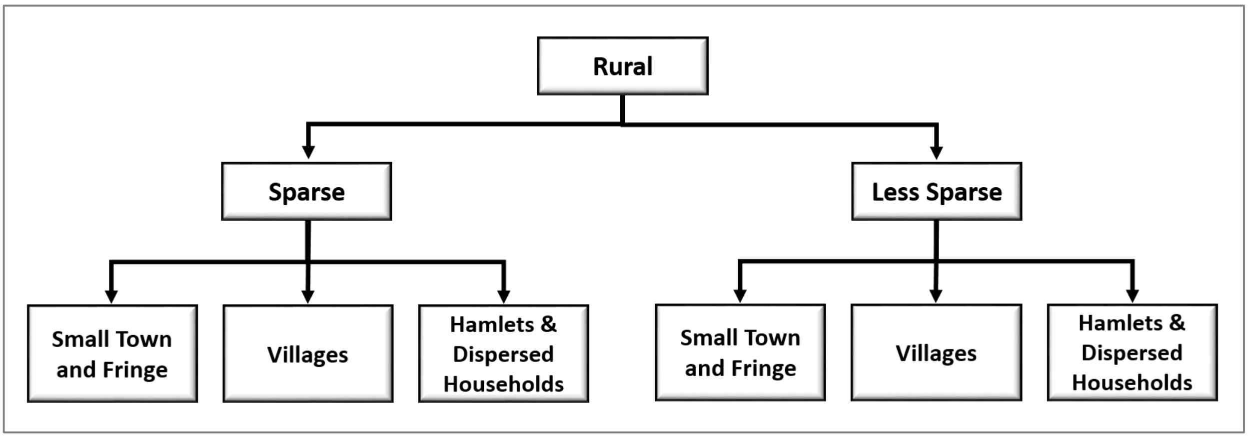


Figure 3.1: The new rural definition and area classification modified after Bibby & Shepherd (2004).

In this thesis the identification of rural settlements is treated from a perspective that considers the settlement as a vague region without knowing where it is actually located or its extent (see Section 3.5). The terms "settlement" and "village" are used interchangeably, despite their meaning within a specific cultural context, to refer to vague rural settlements.

3.3 Study Area

The geographic focus of this research is the rural settlements of Hinckley and Bosworth, a local government district with borough status in southwestern Leicestershire, England (Figure 3.2). It covers an administrative area of 297.35 km^2 and in 2011 had a resident population of 105,078 people (ONS).

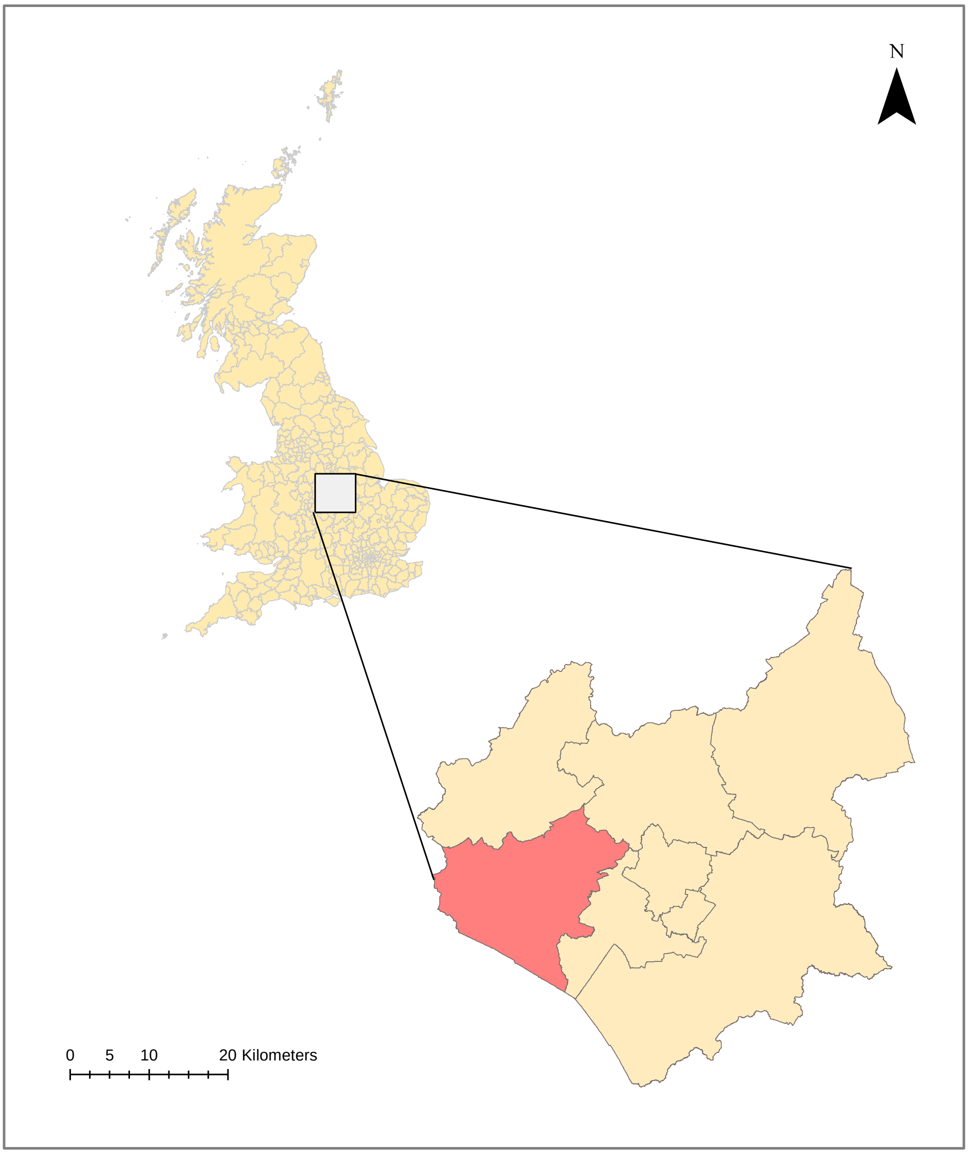


Figure 3.2: Location of Hinckley and Bosworth District within Leicestershire. (Source: © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.)

The borough of Hinckley and Bosworth covers an area of diverse landscape and rolling countryside (Hinckley and Bosworth Borough Council, 2006). That is predominantly rural in nature, with a mixture of enclosed farmland, large cropped fields, intact hedgerow patterns and scattered woodland and hedgerow trees (Hinckley and Bosworth Borough Council, 2014). Hinckley is the district's administrative centre (urban centre) and it is located towards the southeast part of the borough along with Burbage, Earl Shilton and Barwell which constitutes the majority of the built up area.

The Borough was formed (granted borough status) in 1974 under the Local Government Act (1972) by the merger of the Hinckley Urban District and the Market Bosworth Rural District, minus Ibstock. Although most of its settlements dated back to Saxon times. Table 3.1 presents background information and history for some of the settlements, based on their descriptions in the key-to-English-place-names[1] and Hinckley and Bosworth Borough Council websites (Hinckley and Bosworth Borough Council, 2006).

[1] The Institute for Name-Studies (INS) at the University of Nottingham (2015) *Key to English Place-Names.* Available at: http://www.nottingham.ac.uk/ins/key-to-english-place-names.aspx.

Table 3.1: Background information and history for some settlements.

Settlement	Original name	Elements and their meanings	Historical context
Hinckley	Hynca Lēah 'Hynca's wood/clearing'.	**Hynca**: personal name **Lēah**: A forest, wood, glade, clearing; (later) a pasture, meadow.	It is the main urban area, located in the south of the Borough. The town became a centre for the local wool and knitting industries.
Burbage	'Fortification valley/brook'	**burh:** A fortified place. **bǣce:** A beach.	It is located immediately to the southeast of Hinckley town. The original village settlement was developed in linear patterns
Earl Shilton	'Shelf farm/settlement'	**Scelf**: A shelf, a shelving terrain; a pinnacle, battlement. **tūn**: An enclosure; a farmstead; a village; an estate	It is located towards the southeast of the Borough. It developed as a result of the expansion of the hosiery, knitwear and boot and shoe trades.
Market Bosworth	'Bosa's enclosure' 'Market' from the important market here	**pers.n:** Personal name **worð:** An enclosure. **market**: A market, a market-place.	It is an historic market town located in the centre of the Borough. It was originally small nucleated settlement, then it has grown up around the central market place.
Barwell	'Boar spring/stream'	**Bār:** A boar (wild or domestic). **wella:** A spring, a stream	It is located towards the southeast of the Borough. It evolved as a centre for the boot and shoe trade
Desford	'Wild-animal ford' or 'Deor's ford'.	**pers.n:** Personal name **dēor:** An animal, a beast. **Ford:** A ford.	It is predominantly agriculturally based, with small nucleated settlement pattern.
Markfield	'Mercians' open land'.	**Merce**: The Mercians. **Feld:** Open country, unencumbered ground (eg. land without trees as opposed to forest, level ground as opposed to hills, land without buildings); arable land (from late tenth century).	It is located in the north-eastern corner of the Borough, originally was a small linear settlement.
Ratby	Rota's farm/settlement'.	**pers.n.:** Personal name **bȳ**: A farmstead, a village.	It was primarily an agricultural settlement then changed due to the advent of the hosiery industry

3.4 Data Sources

The data used in this research can be categorised into two distinct groups. The first group is core datasets that are used in the analyses to model the village extents. The second one is supplementary data which are of analytical interest for the research problem. All the data are obtained with written permission from Great Britain's national mapping authority, the Ordnance Survey (OS).

3.4.1 Core Data

The main data for this research are records of place names for rural settlements that cover the entire study area. These are of two types: formal (OS MasterMap® Address Layer 2) and contributed data (OS Point of Interest®). These data consist of comprehensive address information for each property and include a settlement name. These are used to identify a particular settlement and then approximate its spatial extent (footprint); this process is discussed fully in Chapter 4.

OS MasterMap® Address Layer2

According to Keith and McLaren (2003), there are two major functions for establishing addresses: firstly, for mail delivery; and secondly, to be used as a reference allowing a property to be found and defined. To respond to this need, there are currently a number of different organisations in Britain that hold and maintain national public sector databases of postal and property-level addresses. The main purpose of this section is to develop an understanding of the Address Point schema. So for this dataset, the source of the address information has been assembled from:

- The Royal Mail's Postcode Address File® (PAF), which supports the national postal service;

- The Ordnance Survey GB Address-Point – created from PAF and providing a national geo-referenced address database; and

- The Valuation Office Agency (VOA) databases that support the tax base for domestic (Council Tax) and non-domestic (National Non-Domestic Rates) properties.

Therefore, the geographical features in this data are broadly classified in terms of their themes, which are:

- Postal address theme – the creation process for an address is the addition of OS identifiers, classification, National Grid coordinates and other metadata to addresses provided in the Royal Mail's PAF. This theme contains residential and commercial premises;

- Multi-occupancies without a postal address (MOWPAs) – these are sourced from the addresses in the Royal Mail's new Multi-Residence (MR) file containing residential premises that fall outside the Royal Mail's definition of a delivery point because the premises' letter boxes are not usually accessible to their delivery person; for example, flats within a converted house that are clearly separate residences but only have one letter box that the postal delivery person can reach (usually the front door of the house); or

- Objects without a postal address (OWPAs) – these are objects that do not have a Royal Mail address, but are still significant enough buildings or structures within the environment that customers may wish to identify, such as churches, halls, car parks, and public conveniences. These are all derived from the OS MasterMap Topography Layer.

Thus, the addresses included in the data are of two types, either a postal address related to a Royal Mail Delivery Point Address or one using the British Standards Institute form of address (BS7666).

British Standard BS7666 Address

This is a spatial dataset for geographical referencing that identifies a national standard for holding details on every street, piece of land and property information (BS 7666-0:2006). This provides a structure for creating and maintaining unique references for any given address. BS7666 breaks down the address into several components, described in Table 3.2, to provide general structures of addresses enabling gazetteers of a range of classes of geographic locations to be created in a consistent way.

Table 3.2: The structure of the British Standard Addresses (BS7666).

Element	Field	Description
Organisation and Premise	Sub-Dwelling	Otherwise known as 'SAO' (Secondary Addressable Object); examples of this might be a flat number, or an apartment name
	Number / Name	Otherwise known as 'PAO' (Primary Addressable Object) or 'Dwelling'; examples of this would be a house number or house name
Thoroughfare	Street	The road on which the building is located
Locality	Locality	The name of a suburb, area or village
	Town	The town in which the address resides
	County	The county or administrative area
Postcode	Postcode	The postcode for the residence

Postal Address

This allows addresses to be written in a simple line format using mandatory fields (postcode, post town and organisation name or PO Box number or building name/ number or sub-building name/number). The full structure of the postal address is presented in Table 3.3 including other optional attributes. Table 3.4 illustrates how the broad categories of themes are obtained with respect to the two basic address types.

Table 3.3: The structure of the Postal address type (Royal Mail Delivery Point Address).

Element	Field	Description
Organisation	Organization Name	The business name given to a delivery point within a building or small group of buildings.
	Department Name	This is indicated in case the mail is received by subdivisions of the main organisation at distinct delivery points.
Premises	PO Box Number	Post Office box identifies the position of the delivery office
	Sub-building Name/Number	These are identifiers for subdivision of properties
	Building Name	A description applied to a single building or a small group of buildings.
	Building Number	A number given to a single building or a small group of buildings, thus identifying it from its neighbours (sometimes called postal number).
Thoroughfare	Dependent Thoroughfare Name	In certain places, for example, town centres, there are named thoroughfares within other named thoroughfares, such as parades of shops on a High Street where different parades have their own identity. For example, KINGS PARADE, HIGH STREET and QUEENS PARADE, HIGH STREET
	Thoroughfare Name	In OS MasterMap Address Layers, a thoroughfare is fundamentally a road, track or named access route on which there are Royal Mail delivery points; for example, HIGH STREET
Locality	Double Dependent Locality	This is used to distinguish between similar or the same thoroughfares within a dependent locality.
	Dependent Locality	This defines an area within a post town.
	Post Town	A town or city in which is located the Royal Mail sorting office from which mail is delivered to its final recipient. There may be more than one, possibly several, sorting offices in a town or city.
Postcode	Postcode	An abbreviated form of address made up of combinations of between five and seven alphanumeric characters.
	Postcode Type	This indicates whether a postcode applies to a single delivery point, which will be indicated with by a large value, or a number of delivery points, indicated by a small value.
	Delivery Point Suffix	A two-character code identifying an individual delivery point within a postcode. Also known as a Premises Code

Table 3.4: Summary of the three addresses theme details.

Theme	Description	Delivered from	Provided in
Postal	Contains residential & commercial premises	Royal Mail's PAF	Postal address & BS7666
Non-postal	Considers the miscellaneous premises, e.g. churches, halls, car parks, etc.	OS MasterMap Topography Layer	BS7666 address only
Multi-occupancy	Contains residential premises in multi-occupied ones, e.g. flats within a house.	Royal Mail's Multi-Residence file	Postal address & BS7666

OS Point of Interest ® (POI)

This is a tabular dataset containing around four million different geographic features that covers all of Great Britain. All features are supplied with location, functional information and addresses where possible. POI is generally considered the main source by which to obtain function information, as it covers the commercial addresses and features of interest. This contrasts with the OS AddressPoint Data which additionally encompasses residential addresses (Lüscher & Weibel, 2013). It has a three-level classification that helps in identifying the features or sets of features required. The top-most level comprises nine Groups, which are: accommodation, eating and drinking; commercial services; attractions; sport and entertainment; education and health; public infrastructure; manufacturing and production; retail; and transport. The second level is broken into 52 Categories for all the groups. The third level of the classification scheme is the most detailed level, which contains more than 600 Classes.

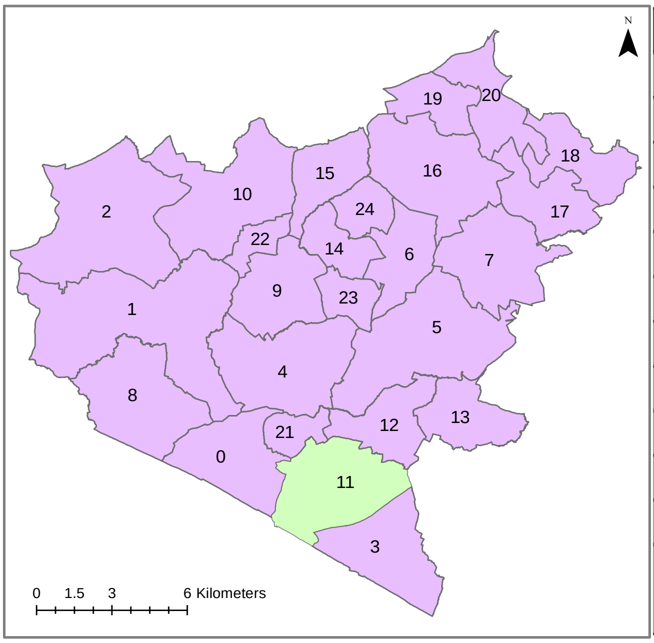
3.4.2 Supplementary Data

Parish Boundary

According to the Department for Communities and Local Government (2010), a parish should reflect a distinctive and recognisable community of place, with its own sense of identity (Page 19, Guidance on community governance reviews). Or to put it differently, the 'name' of a parish in the Local Land and Property Gazetteer

(LLPG) refers to the geographical name of the area concerned, which might be known locally as a town, community, neighbourhood or village, rather than a parish. Historically, a parish is the smallest spatial unit of the administrative areas in England inherited from the administrative structure of the Church of England; the equivalent units in Wales and Scotland are instead known as 'communities' (The Office for National Statistics - ONS). These parishes only existed, for local government purposes, within the boundaries of the former rural district councils (Singh, 2009). As a consequence of this, village names may be expected to be equivalent to parish boundaries and therefore it would be useful to include these data in the research.

The Parish Boundary is part of the OS Boundary-line® dataset that covers all administrative and voting boundaries in the UK. For this study, only the parishes that fall within Hinckley and Bosworth District are considered. There are a total of 24 parishes (also known as Civil Parishes), apart from a parish-free area (formally known as Non Civil Parish) which is part of Hinckley and is superseded by other local government units. A complete list of the included parishes and their locations can be seen in Figure 3.3.



FID	Name	FID	Name
0	Higham on the Hill CP	12	Barwell CP
1	Sheepy CP	13	Earl Shilton CP
2	Twycross CP	14	Osbaston CP
3	Burbage CP	15	Nailstone CP
4	Sutton Cheney CP	16	Bagworth & Thornton CP
5	Peckleton CP	17	Ratby CP
6	Newbold Verdon CP	18	Groby CP
7	Desford CP	19	Stanton-under-Bardon CP
8	Witherley CP	20	Markfield CP
9	Market Bosworth CP	21	Stoke Golding CP
10	Shackerstone CP	22	Carlton CP
11	NCP (Hinckley)	23	Cadeby CP
		24	Barlestone CP

Figure 3.3: Map of Hinckley and Bosworth Borough contains 24 Civil Parishes (CP). Hinckley is not a parished area (NCP). (Source: © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.)

Basemap Background

For visualization and to add more geographical context, two basemaps are used in conjunction with the maps presented in this thesis. These are of two types:

OS MasterMap® Topography Layer

This dataset provides a highly detailed view of surface features on the landscape topography at a scale of 1:1250. It is further subdivided into a number of themes: land area classifications: buildings, roads, tracks and paths, rail, water, terrain and height, heritage and antiquities, structures; and administrative boundaries. For visualization, therefore, it seems useful in mapping the villages to overlay the address point data on the OS MasterMap, as shown in the figures in this chapter.

OpenStreetMap (OSM)

Another background layer that contains data from OpenStreetMap is also used in the next chapter. This layer is one of the data sources that has considerable potential in GIScience, as OSM creates map data that are freely available, providing current digital geographic information without any legal or technical restrictions. In this sense, it is a popular example of Volunteered Geographic Information (VGI), which relies on crowdsourced spatial data (Haklay, 2008; Mooney and Corcoran, 2011). Volunteers in the OSM community, acting as mapping parties, collect geographic information by taking handheld GPS devices (equipped with open source software) with them on journeys or go out specifically to record GPS tracks and then submit this information to the global OSM database (Batty *et al.* 2010).

OS Strategi ®

This is another vector representation of the Ordnance Survey's 1:250,000-scale graphic maps. It provides a generalized overview of many features commonly portrayed on regional-scale topographic maps, such as the road network, railway lines, lakes and watercourses, cities and wooded areas. Of these, only the toponyms of rural areas are used here since this dataset contains an annotation layer in which names are stored as point objects. In this way, a village is represented as single points, as can be seen in Figure 3.4.

Figure 3.4: Map of Hinckley and Bosworth Borough containing settlement points from the OS strategi layer. (Source: © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.)

3.5 Settlement Identification

Having discussed the available data sources that contain location information at property level, such as the OS Address Point and POI datasets, this section explains how these data are treated to represent the settlement pattern. The data contain three different fields associated with village names. These are the locality names in both address systems formally referred to as "BS7666 Address Locality (BS7666Ad_3)" and "Postal Address Dependent Locality (PostalAd_9)"; and the locality field of the POI attribute table "LOCALITY". Each of these has its own list of villages in addition to villages in common. This means that some villages are unique and only identified in one database. A full list of these distinct settlements and those in common across the three different databases with counts of the addresses in each village is shown in Table 3.5. In total, there are 76 rural settlements identified in the study area based on these data. According to an official website for Leicestershire villages[2], some settlements are misplaced or missing, as shown in Table 3.6. By way of illustration, a number of the villages were originally located in neighbouring boroughs (or even counties), but now some addresses fall within the Borough of Hinckley and Bosworth. In this analysis, small settlements (e.g. hamlets and special areas – those with very few houses) are disregarded, as is Hinckley as it is not a particular settlement, but is rather distributed around a borough yet attributed with the same name. The data were combined into a new database that records the settlement names and identifies the database that they are from.

[2] http://www.leicestershirevillages.com/ ; Leicestershire villages an online community for residents of, and visitors to, Leicestershire.

Table 3.5: List of the rural settlements in the study area with counts of the addresses identified.

No.	Village Name	BS76	POST	POI	No.	Village Name	BS76	POST	POI
1	Ambien	4	0	0	39	Market Bosworth	1140	1048	106
2	Appleby Magna	1	10	4	40	Markfield	2291	0	173
3	Aston Flamville	0	8	2	41	Nailstone	260	215	13
4	Atherstone	0	0	8	42	Newbold Verdon	1429	1376	70
5	Atterton	0	15	2	43	Newtown Linford	1	1	0
6	Austrey	3	2	0	44	Newtown Unthank	0	27	10
7	Bagworth	0	604	24	45	Norton Juxta Twycross	0	111	6
8	Bagworth & Thornton	1282	0	0	46	Nuneaton	0	0	23
9	Bardon Hill	0	8	1	47	Odstone	0	30	8
10	Barlestone	1118	1080	30	48	Orton-On-The-Hill	0	81	8
11	Barton In The Beans	0	103	5	49	Osbaston	161	119	10
12	Barwell	4382	4244	221	50	Peckleton	626	95	18
13	Bilstone	0	20	5	51	Pinwall	0	15	5
14	Botcheston	0	198	10	52	Ratby	2010	1904	60
15	Bufton	0	3	0	53	Ratcliffe Culey	0	83	12
16	Burbage	6979	6635	257	54	Sapcote	0	1	0
17	Cadeby	133	83	9	55	Shackerstone	465	68	9
18	Carlton	139	124	8	56	Sharnford	0	2	0
19	Castle	2076	0	0	57	Sheepy	692	0	0
20	Clarendon	5702	0	0	58	Sheepy Magna	0	302	14
21	Coalville	0	0	6	59	Sheepy Parva	0	47	3
22	Coatbridge, Lanarkshire	0	0	1	60	Shenton	0	53	9
23	Congerstone	0	151	9	61	Sibson	0	85	13
24	Copt Oak	0	20	1	62	Snarestone	0	5	0
25	Dadlington	0	119	11	63	Stanton Under Bardon	366	281	23
26	De Montfort	4532	0	0	64	Stapleton	0	195	26
27	Desford	1864	1497	95	65	Stoke Golding	797	793	59
28	Earl Shilton	4884	4763	227	66	Sutton Cheney	339	65	14
29	Ellistown	0	11	5	67	Thornton	0	460	27
30	Elmesthorpe	0	1	0	68	Thurlaston	0	5	1
31	Fenny Drayton	0	222	14	69	Trinity	2988	0	0
32	Gopsall	0	9	4	70	Twycross	508	148	14
33	Groby	3282	2954	113	71	Ulverscroft	0	3	0
34	Higham On The Hill	483	284	15	72	Upton	0	43	8
35	Hinckley	0	0	1346	73	Wellsborough	0	27	4
36	Kirby Muxloe	4	103	3	74	Witherley	767	313	13
37	Kirkby Mallory	0	176	19	75	Wolvey	0	5	3
38	Lindley	0	8	2	76	Wykin	0	25	0

Table 3.6: List of the rural settlements that are either misplaced or missed from the data.

No.	Misplaced Village	District	No.	Missed Village
1	Aston Flamville	Blaby District	1	Brascote
2	Elmesthorpe			
3	Sapcote		2	Bull In Oak
4	Sharnford			
5	Thurlaston		3	Far Coton
6	Newtown Linford	Charnwood Borough		
7	Ulverscroft		4	Field Head
8	Appleby Magna	North West Leicestershire		
9	Bardon Hill		5	Little Orton
10	Coalville			
11	Ellistown		6	Little Twycross
12	Snarestone			
13	Atherstone	County of Warwickshire	7	Merry Lees
14	Austrey			
15	Nuneaton		8	Sketchley
16	Wolvey			
17	Coatbridge, Lanarkshire	Lanarkshire		

3.6 Disambiguation of Place Name Records

Before proceeding to model the rural settlements as vague objects (in the following chapter), this section explains in more detail why these rural areas are considered vague. The data described in Section 3.4 will be used to represent the settlement patterns and to model vagueness in place names, despite the data defining feature locations through geocoded address points, which are assumed to be recorded accurately. In addition, there are further aspects of uncertainty beyond the vagueness issue. Settlement (placial) ambiguity arises from a lack of precise locations, crisp boundaries and universal names for many places (Davies *et al.* 2009), and this sections seeks to provide a starting discussion about the conceptual vagueness and ambiguity reflected in village name records.

3.6.1 *Indeterminacy, Incompleteness and Inconsistency*

As mentioned in the previous section, the variation in the distinct village names in the different databases suggests conceptual inconsistency in the definition of any individual settlement. The combined database allows investigation of these vaguely defined entities. Incompleteness is commonly associated with many spatial data sets (Worboys, 1998 a &b). Bordogna *et al.* (2006) describe this lack of consistent information relative to a spatial sub-region as being due either to the fact that the observation does not cover that specific region or, broadly, to the presence of obstacles between the point of observation and the phenomenon observed. This is evident in the settlement data, in which some village names are recorded differently in different data sources. For instance, village names shared in the three databases are Cadeby and Burbage (Figure 3.5). It is clear from the figure of different point sources that the spatial patterns of the village varies in their distribution and extent. POI tends to have less number of addresses (point features) than the other address point data, whereas the BS7666 have much more points which are basically missing from other database (POI) or named differently in the POST database.

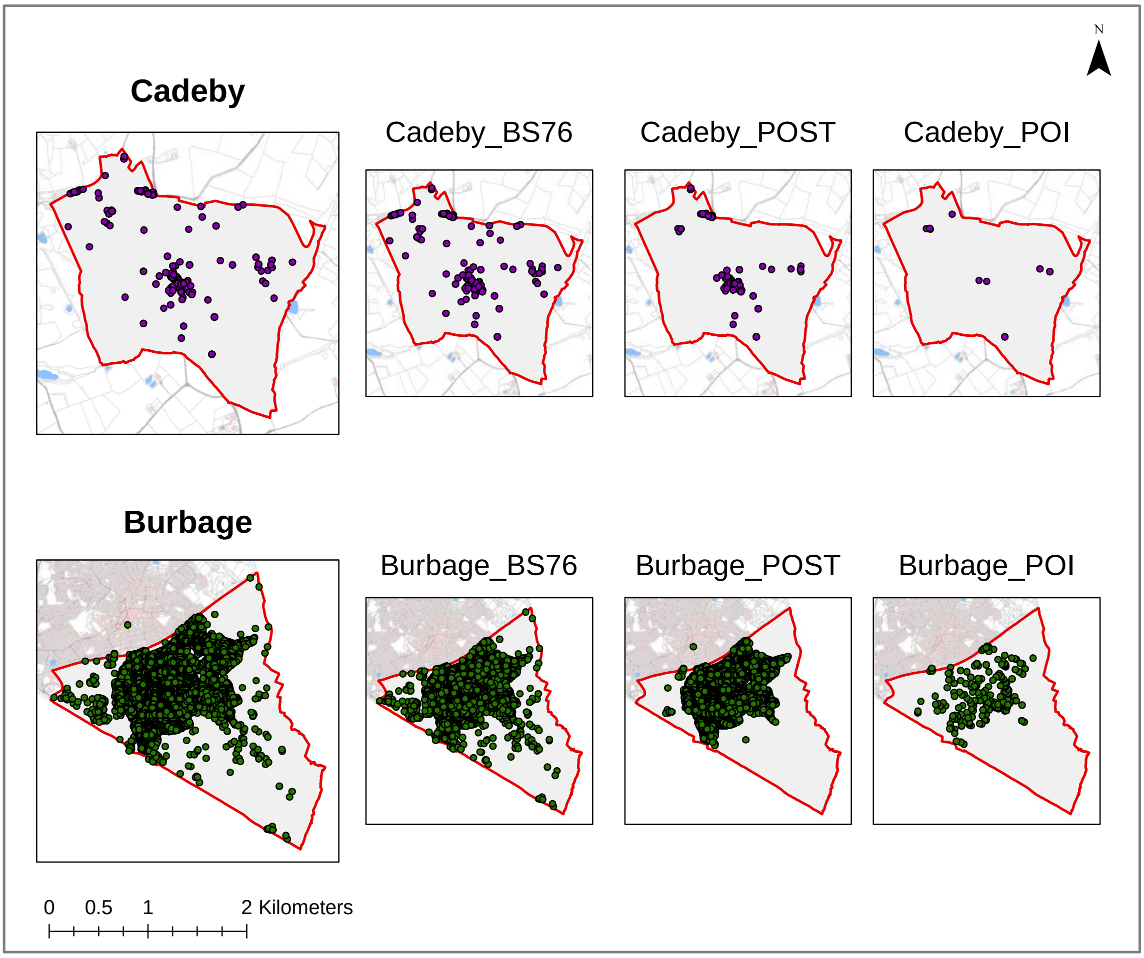


Figure 3.5: Examples of the disparity in the village names due to using different address databases. The column on the left shows the same settlements with the address being aggregated from all databases. (Source: © Crown copyright and/or database right 2012. All rights reserved.)

This indicates a further issue of inconsistency that derives from the availability of multiple observations of the same feature - or in this case village, which often leads to conflicts and contradictions (Bordogna *et al.* 2006), as shown in Figure 3.5. It is clear from the figure that the spatial patterns of the villages vary in the distribution and extent based on the different systems of addresses. The issue of indeterminacy and incompleteness can be further contextualized by mentioning that not all the villages located in Hinckley and Bosworth exist in the OS Strategi dataset (i.e. not represented as single points- see Figure 3.4 and Table 3.5). For example, Wykin, Wolvey, Trinty, Sheepy, Pinwall, Lindley, Gopsall, De Montfort, Clarendon, Castle, Bufton, Austrey, Atterton and Ambien are all disregarded in this dataset.

3.6.2 Spelling and Punctuation

Another source of ambiguity between the different databases is in the spelling and punctuation of village names. Generally, this means that for a certain village perhaps with a well-known name, there are multiple written formats. They should refer to the same place regardless of the differences in spelling or punctuation, but in many cases they have distinct spatial distributions of their address points. This is exemplified by the use of a hyphen or space to separate words in village names, as in the cases of "Higham on the Hill" or "Higham-on-the-Hill" and "Stanton under Bardon" or "Stanton-under-Bardon". Figure 3.6 illustrates these differences. The use of a hyphen or space is not exclusive to a particular database. That is to say that both the BS7666 and Postal address sometimes employ a hyphen in some villages and use a space for others. Further to that, these village names in the POI dataset are even joined with other names (i.e. Nuneaton and Markfield). These indeed add another level of confusion about the place names and therefore more uncertainty.

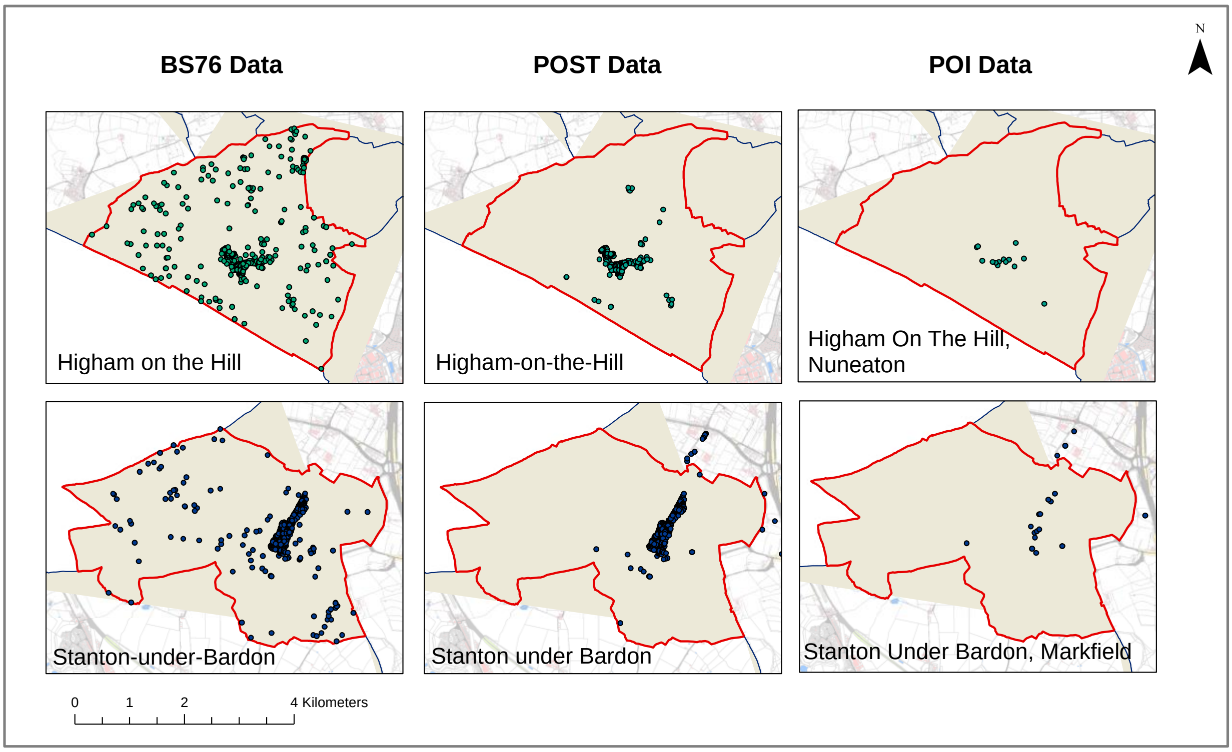


Figure 3.6: Examples of the disparity in village names resulting from use of a separator. (Source: © Crown copyright and/or database right 2012. All rights reserved.)

3.6.3 Parish Boundary vs. Village Extent

Looking at the rural settlements in these data, it is evident that some of these settlements have precisely defined parish boundaries, but the actual extent of the addresses differs. Figure 3.7 shows this point clearly. In the cases of Desford, Earl Shilton and Stoke Golding, all three have clearly defined parish boundaries. However, there are address points that fall beyond or outside the parish boundaries, which form the actual extent of the residences in these villages. It appears that the BS76 data coincide with the parish boundary while the POST and POI data tend to have a different extent. This again illustrates the problem of ambiguity, inconsistency and incompleteness (Worboys, 1998 a&b; Bordogna *et al.* 2006).

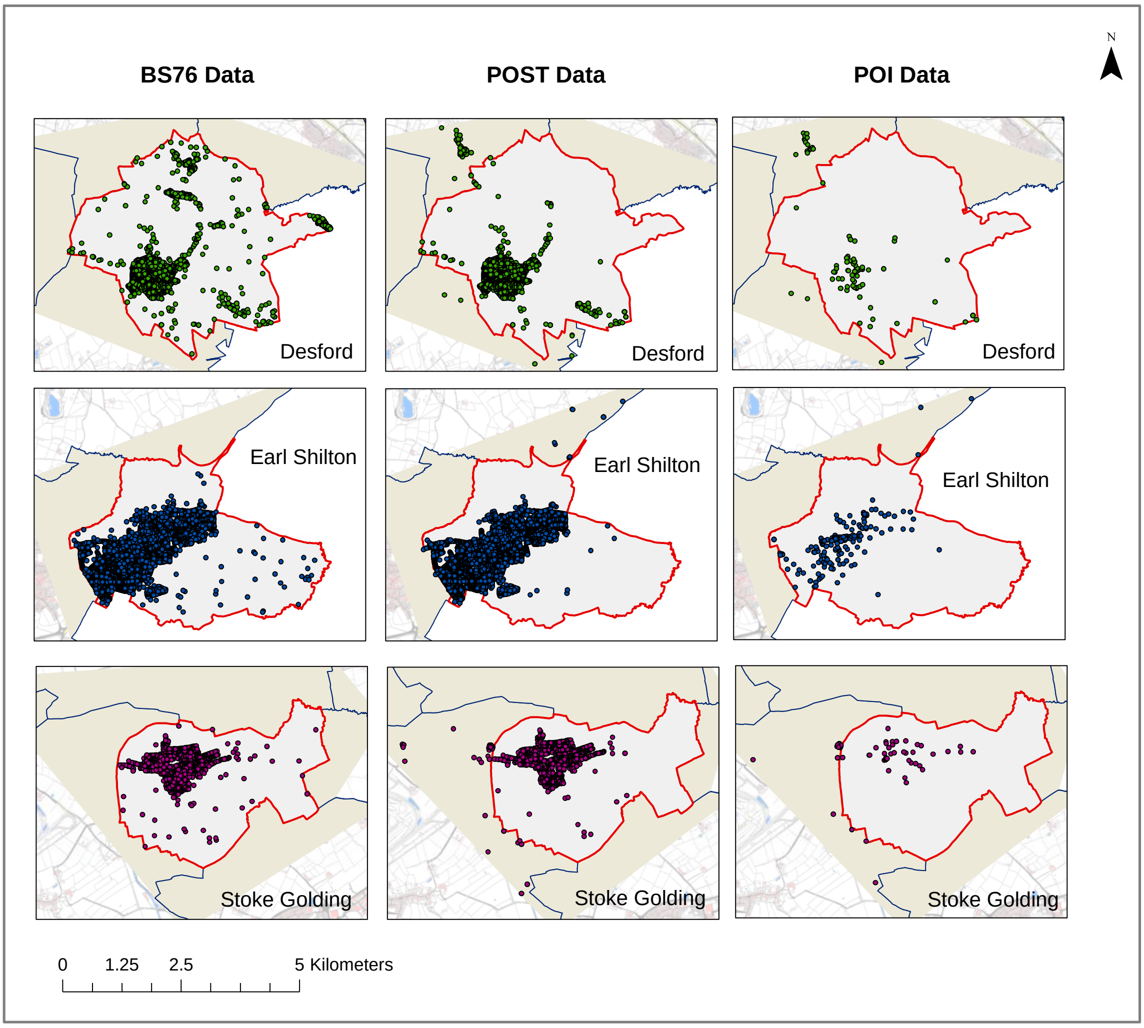


Figure 3.7: Examples showing the relation between the extents of some village names and their parish boundaries (red polygons). (Source: © Crown copyright and/or database right 2012. All rights reserved.)

Another problem is that many parishes contain two or more rural settlements, which often have spatially overlapping addresses. Figure 3.8 shows an example of three parishes - Desford, Sutton Cheney and Witherley - that are associated with more than one locality. This is mainly due to the fact that not every rural settlement has its own parish (with the same name); rather, one parish could be responsible for a number of settlements. Regardless of the data source, many locations have multiple addresses which have different names (Figure 3.8). The first column in Figure 3.8 shows the villages that are equivalent to the parish names, while the second presents other villages within those parishes. This ambiguity could extend to the possibility of having other point addresses from a neighbourhood settlement or even borough, especially in cases of Desford and Witherley as they are located at the eastern and western edges of the borough, respectively.

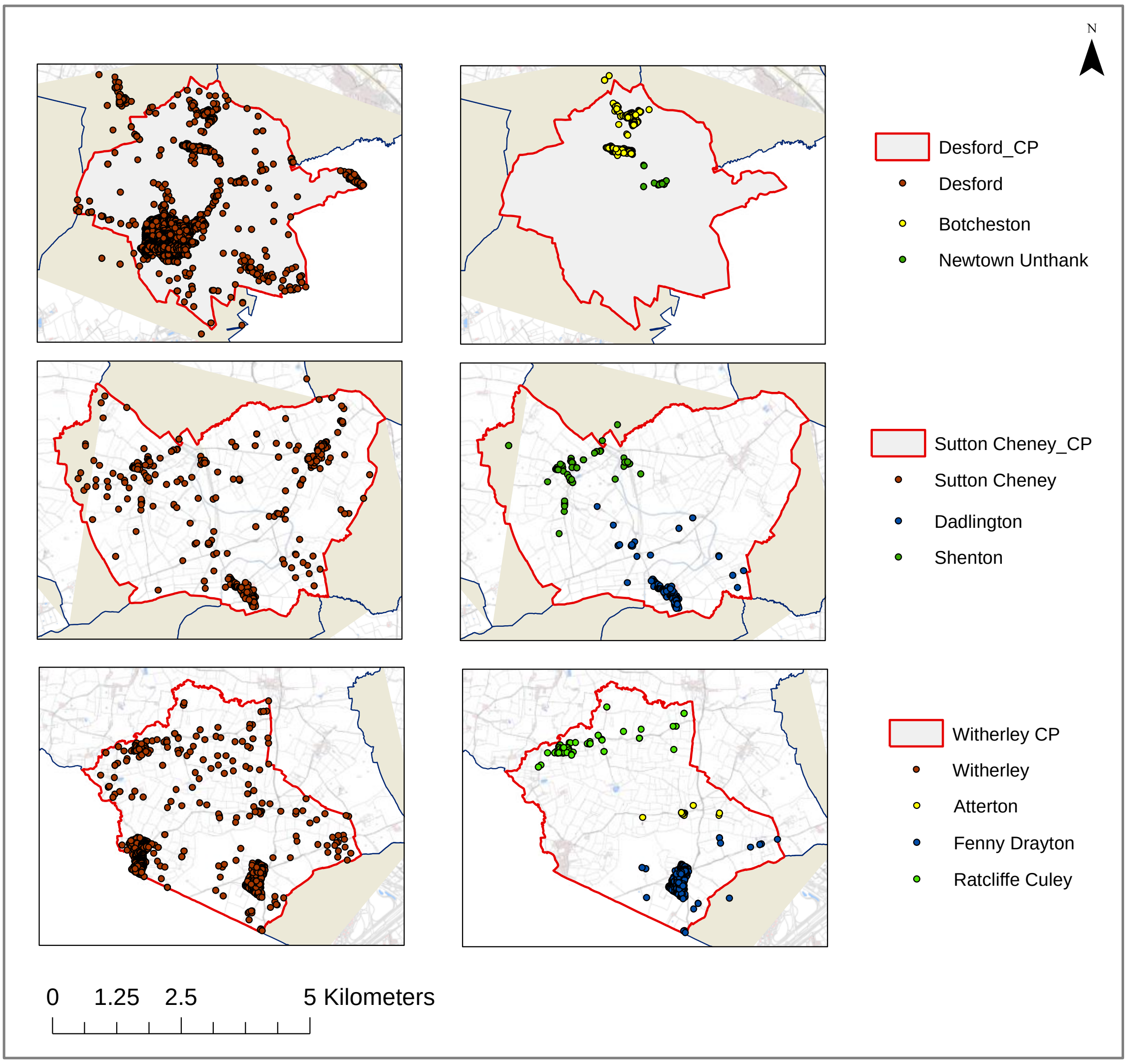


Figure 3.8: Examples of some parishes (red polygons) associated with more than two settlements. The column on the left shows the addresses belonging to settlements with the same parish names; while the column on the right shows other associated settlements. (Sourcee: © Crown copyright and/or database right 2012. All rights reserved.)

3.6.4 Combined Settlement Names

The same principles also apply to combined settlements. From Figure 3.3 and Table 3.4 it is evident that some settlement names appear within the parish boundary data and in the BS76 data, while the other databases have a different semantic view, separating the names. Figure 3.9 illustrates this situation in the case of two or more localities joined together in a combined parish name. The term "combined parish" is generally understood to mean the situation arising when the local government authorities decide to group parishes or change their status or style based on local community needs (Communities and Local Government, 2010). It should also be noted from Figure 3.9 that, in addition to the distinct settlements which comprise the parish names, there are other settlements that belong either partially (e.g. Botcheston, Desford and Sibson) or completely (e.g. Pinwall and Wellsborough) within these parishes.

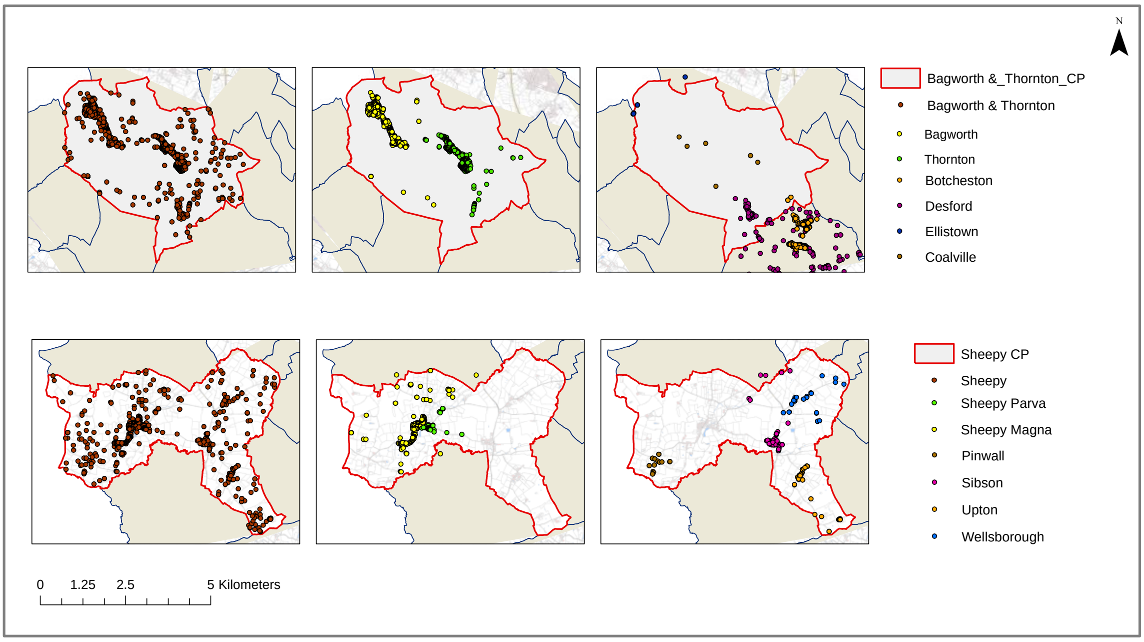


Figure 3.9: Examples of parishes with combined settlement names. The first column (on the left) shows the addresses that belong to settlements with the same parish names combined; while the second shows them separated and the last column (on the right) shows other associated settlements. (Source: © Crown copyright and/or database right 2012. All rights reserved.)

3.7 Summary

This chapter introduces and elaborates the scope of the research in terms of the geographical scale and the study area and provides a detailed explanation of the data sources used to achieve the objectives of the research. It introduces the inherent vagueness and ambiguity in village names as recorded in the different databases, and identifies examples of vague regions as named settlements from address records from different sources (OS address point data – BS76 and POST – and OS POI data). A number of inconsistencies and contradictions have been identified and illustrated in the mappings of address points, with the same apparent villages recorded in multiple datasets and villages with different names at the same location. Moreover, the chapter has shown that the point addresses for certain settlements can differ from the equivalently named settlements' parish boundaries. Taken together this provides evidence that there are a number of different causes and consequences of inherent uncertainty in spatial databases, with important implications for data quality. Overall, this problem reflects the recommendation from other studies (Worboys, 1998 a&b; Fisher 1999; and Davies *et al.* 2009) of the need for a better understanding of the possible sources of uncertainty, and how they may be addressed.

Chapter 4 Modelling and Analysing Vague Rural Settlements

4.1 Introduction

Representing and modelling a 'place' has received considerable critical attention, especially when this place is geographically vague in both definition and extent. As discussed in the previous chapter, rural settlements are a notable example of vague regions, as they often do not have officially defined boundaries or they have formal or informal names connected to past events or landmarks. Their labels or names can be recorded in different ways, causing contradiction – for example, between and within Gazetteers and other databases recording place-based information – and resulting in confusion for users. The spatial footprints of 'places' may vary as a result of differences in individual perception and may depend on social factors as much as physical environmental differences. With these issues in mind, a number of modelling techniques have been developed to handle vagueness and imprecision in GIScience, as described in Chapter 2.

Within this context, this chapter sets out to map the spatial extent or footprint of local rural settlements using records of village names held on national databases and to provide an in-depth analysis of the fuzzy spatial settlement patterns arising from different databases. It is important to note that it has been possible to use a number of different data sources in this thesis to examine vagueness in the spatial nature of rural settlements because of the availability of such data.

Consider the following problem: address datasets include individual addresses that are attributed with a settlement. In rural locations, the settlement attribute describes the village which the address is considered to be within. Different address datasets indicate different addresses and different villages for the same address. The problem being considered in this chapter is how to determine the spatial extent of the village. In Britain, a number of formal data exist that describe villages, their territories and other small areas. However, in developing countries where such infrastructural information may be lacking, and the only address datasets available may be user-generated POIs, settlements are commonly represented as discrete points (Crawford, 2002). The address data may only loosely be attributed. The question is then how such incomplete, sparse, contradictory, poorly attributed and frequently informal or vernacular data could be used to construct the potential areal extent of each village or settlement.

This chapter proceeds as follows: Section 4.2 briefly reviews the issue of applying two approaches to generate traditional hard boundaries from a set of points. Then Section 4.3 illustrates the approach employed for modelling the fuzzy representation of the settlements. Analyses of the resulting models are explained in Section 4.4, before the entire results are discussed in Section 4.5. Finally, Section 4.6 offers a summary and general discussion and recommendations from the proposed modelling and analyses techniques.

4.2 Hard / Fixed Boundary

One of the most essential and critical tasks in GIS is to identify boundaries around sets of points. A number of methods exist for doing this, including convex hull and Voronoi Tessellation described below. These generate fixed or hard boundaries, which may be problematic if the points represent vague locations or regions. This section presents two commonly used methods to generate such boundaries; and highlights some of the key issues that arise from applying them to rural settlement locations, using a small case study around the borough of Hinckley and Bosworth in rural Leicestershire.

4.2.1 Convex Hulls

A convex hull is a fundamental algorithm in computation geometry. It generates a shape that completely encloses a set of points (de Smith *et al.* 2015) and can be thought of as a rubber band that is stretched around the whole set. It is possible, therefore, to create convex hull polygons for tracing the geographical extent of a village based on the distribution of address points attributed to that village name. There are a number of software packages that could be used to obtain the convex hull polygons – Hawth's Tools and Crimestat, to name two. The latter has been chosen for this research because it is freely available and compatible with other GIS software (e.g. ArcGIS).

Given a set of points that represents one of the considered settlements in the address point data a convex hull is computed, and the resultant polygon outlines the boundary. Figure 4.1 presents maps of the convex hulls for the villages in the BS76 and POST data. It is apparent from the maps that, although the convex hulls can be useful for describing the geographical spread of a village, they are markedly affected by extreme values. The two address databases suggest very different settlement areas with large differences in spatial extent. There are gaps between the settlement areas computed from any one of the address databases and some of the areas overlap because of the way that the hull is computed. Simple comparisons can show whether one distribution has a greater extent than another does. For instance, the majority of the settlements in the BS76 data (Figure 4.1 top left map) generally tend to have larger extents than the settlements in the POST datasets (Figure 4.1 bottom left map). In some cases, however, the convex hull does not provide a good representation of the boundaries of a given set of points; it may not represent the actual area occupied by the set of points very well. Consider Sheepy and Sheepy Magna: in this case there are large differences in the village attribution or labelling between the databases and consequently in the settlement areas suggested by the convex hull, which may be vulnerable to extreme values. If one address is a geographical outlier to the settlement, then it is isolated and the hull will be pulled out to that location, as can be seen in Nailstone in the BS76 data and Twycross in the POST data. The spatial extent of the settlement may be dramatically distorted as the areas are defined by the most extreme points. It is important to bear in mind that the settlements that share the same name have different representations in the two datasets. This is evident in the case of Twycross from the two datasets.

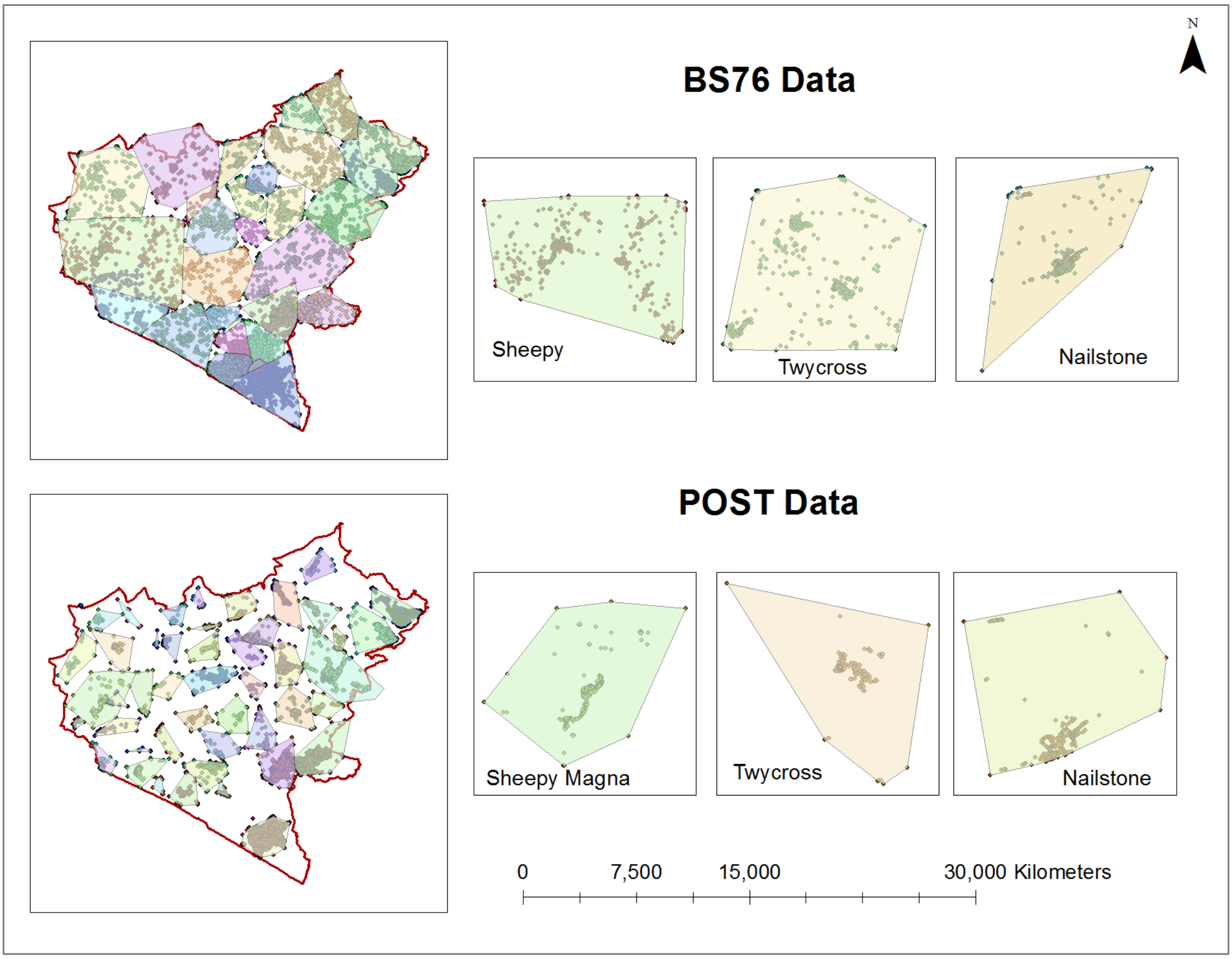


Figure 4.1: Maps showing the Convex Hulls for The Rural Settlements in The BS76 and POST Datasets; with some examples in large scale on the right. (Source: © Crown copyright and/or database right 2012. All rights reserved.)

4.2.2 Voronoi Tessellation

An alternative approach to model the settlement hard boundaries, based on georeferenced addresses is by Voronoi Tessellation. It is a widely held view that the spatial footprints of inherently vague regions intended for use within gazetteers can be delineated by Voronoi Tessellation of point locations known to be located inside or outside the target region for example through their settlement attribution (Alani *et al.* 2001). According to Klein *et al.* (1990), the Voronoi diagram of a set of sites S in the plane partitions the plane into regions, called Voronoi regions, one for each site. The region of site p contains all points of the plane that are closer to p than to any other site S.

Hence, it would be possible to create fixed boundaries around the addresses using Voronoi tessellation to divide the study area (borough) into regions (villages) based on distance to points in a specific location, typically around village centre points (centroid or mean centre). The following pseudo-code (in Figure 4.2) shows how a Voronoi function could be used with the point data to generate the Voronoi tessellation around the village centres as a classic method for generating hard boundaries for the rural settlements . This is done in four different cases, where villages identified in the BS7666 data only, the Postal address data only, the POI data only, and villages exist in all the different sources combined together. Figure 4.3 presents these results for the four different data. A, C, E & G are showing the Voronoi polygons with the village centres; B, D, F & H are showing the Voronoi polygons with the actual address points. Overall, the results show that the address points for a village do not correspond to the Voronoi polygon bounding that village, as some of these addresses fall in outside and also other address points from a different village more likely to be inside this Voronoi polygon.

```
#==========================================================================================
# Script Title:             Code for generating Voronoi Tessellations
#
# 1. Install & Load the required libraries
    library(dismo)
    library(GISTools)
# 2. Read the mean centre points, files for each data sources
      Comp_mc
      BS76_mc
      POST_mc
      POI_mc
# 3. Generate the voronoi tessellation and save them in variables
      Comp_vo <- voronoi(Comp_mc)
      BS76_vo <- voronoi(BS76_mc)
      POST_vo <- voronoi(POST_mc)
      POI_vo  <- voronoi(POI_mc)

# 4. Write the results as a shapefile
      writePolyShape

#==========================================================================================
```

Figure 4.2: Pseudo-code for generating the Voronoi tessellations for the settlements in the four data sources in the study area.

Figure 4.3: Results of Voronoi tessellations for the settlements in the four data sources. A, C, E & G show the mean centres while B, D, F & H show the original address points.

4.2.3 Key Issues:

The two methods are not particularly useful where the actual extent of the village is unknown and is consequently an inherently vague region. For some villages, the output polygons overestimate the area of the village (e.g. by extreme points in the convex hull approach), or underestimate the region (e.g. by the points that fall outside the Voronoi tessellation). Furthermore, these approaches do not allow the address points to have multiple memberships – each village has membership of one and only one village. This clearly reflects the need to develop a better way to approximate the vague extent of the villages by fuzzy models of their boundaries.

4.3 Fuzzy Modelling

4.3.1 Modelling based on Density of Houses

A common approach applied in GIScience to generate approximate boundaries for vague places is based on Kernel Density Estimation (KDE) (Jones *et al.* 2008; Twaroch *et al.* 2008 a; Hollenstein and Purves, 2010; de Berg *et al.* 2011). KDE generates a smooth, continuous surface from point patterns which represents spatial variations of events (Silverman 1986). It is often considered the most potent technique for "hot spot" analysis; as Downs (2010) points out, KDE has been applied widely to identify hot spots of crime, disease, fatalities, traffic accidents and lighting strikes in addition to its usage to characterise the spatial distribution of plants, animals and people.

In this research, however, KDE is used to model the spatial footprint of the "fuzzy extent" of villages. KDE estimates the proportion of total incidents (addresses) that can be expected to occur at any given map location. It works by first overlaying an area of interest with a rectangular two-dimensional grid. It then calculates an estimate of the density of incidents in each grid cell, which is based on a weight function, the kernel (a function of specific shapes and bandwidths or search radius). The general form of the bivariate kernel density estimate f at any point x is:

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \qquad \textbf{Equation 4.1}$$

Where n is the sample size, X_1, X_2, ... X_n are events and *K()* represents the kernel weighting function which operates on the distance between each point and event given a specified bandwidth h (Silverman 1986). This reflects the fact that there is a greater probability of an incident occurring in a given location the closer it is to the location of a known incident. It should be noted that the choice of the surface resolution (grid cell size) and the bandwidth (kernel radius) in this function both influence the shape of the surface in terms of its smoothness or peakness (Silverman 1986; and Bowman and Azzalini, 1997). Caution must be applied in the selection of the surface resolution and the kernel radius (Jones *et al.* 2008). The surface resolution must be sufficient to resolve the boundaries of the region and will vary according to region. The kernel radius should ideally be small enough to represent local variation within the region at a scale commensurate with the size of the region and large enough to capture multiple point locations within the kernel radius.

There is much discussion in the literature about bandwidth selection and there are several varieties of bandwidth selectors for kernel density estimators. In many situations it is adequate to select the bandwidth based on preliminary examination of several density estimates over a range of bandwidths (Keen 2010). Furthermore, many applications require an automatic choice of smoothing parameter when density estimation is to be used routinely on a large number of data sets or as part of a larger procedure (Silverman 1986). In fact, another reason for arguing in favour of objective approach for selectin bandwidth is when no prior knowledge about the shape of the point distribution; and that would be a fair justification for automatic bandwidth selection (Keen 2010). In this research, therefore, the KDE for each village is computed independently for the considered villages in each of the datasets. This is done using the "*kde.points*" function in R, developed by Brunsdon and Chen (2014) based on a rule-of-thumb approach for choosing the bandwidth of a Gaussian kernel density estimator (Venables and Ripley, 2002). In other words, the bandwidth parameter used here, which could be any real number, is unidentified then a rule-of-thumb method is applied by default.

The method was popularised for kernel density estimates by Silverman (1986), who used the normal distribution as the parametric family.

In order to represent the spatial extent of villages as fuzzy objects, a further step is needed to transform the density surface to a fuzzy set-based approach (Zadeh 1965). In other words, there is a need to show the fuzzy memberships from the density surface in order to represent and to quantify the degree to which a particular location is part of the village. Usually, the normalisation process is used for this purpose. This is to change the scale range of the KDE results from 0 to 1. Here, a value of 1 indicates a complete membership and 0 represents classical non-membership; and in between these two extremes, the membership value varies according to how dense the area is.

4.3.2 Generating α-cuts:

As mentioned in the literature review (Chapter 2), recall that an α-cut (alpha-cut) is a fundamental aspect of the fuzzy set theory and plays a key role in this chapter. In essence, the α-cut is a set of objects or locations which have a fuzzy membership greater than or equal to some threshold, α, is defined as being within the set coded as 1, and all other locations outside the set coded 0. Strictly speaking, α-cut is the area delineated by a contour of equal membership, giving a hard, crispened or Boolean version of the fuzzy set; where α can take any value between 0 and 1 (Arnot *et al.* 2004). By generating a set of α-cuts, it is possible, therefore, to model fuzzy regions as a set of crisp regions which in turn allow the use of already developed algorithms (for hard regions) in a fuzzy context (Schmitz and Morris, 2006).

In the work reported here, an R script is used to generate the Boolean maps for the rural settlements for values of α in the range from 0.1 to 0.9, giving nine different Boolean hardenings of the fuzzy representation of the settlements. This is explained using pseudo-code in Figure 4.4, which describes the process to generate normalised density surfaces and their α-cuts for the villages. The full R script, with the looping structure applied through the studied settlements, is provided in Appendix (6.2).

```
#==========================================================================================
# Script Title:  Density Surfaces - generating normalised KDE, and their alpha-cuts
#_________________________________________________________________________________________
# 1. Install & Load the required libraries
    GISTools, spgwr, rgdal, raster & shapefiles
# 2. Set the working directory where the data exist and read in the data
#       a. read the settlements' point data
#       b. read the polygon of H&B district
# 3. Check & investigate the data to identify villages with at least 10 points in any data type
#       a. define lists for each data type and village
            village.list &  data.source.list
# 4. Generate Helper Functions
#     ``````````````````````````
#       a.normalising the data to stretch from 0--> 1
            kde.norm.func
#       b. KDE plot function
# 5. Main Loops
#    ```````````
#       a. Set the working directory where the results to be saved
#       b. Loop 1 generates and plots KDEs
#           - creates a data variable for each village & each data type
#           - loop through the data for each village in each data type
#       c. Loop 2 generates alpha cuts and exports them
#           - loop through the data type list
#           - create alpha cuts for each village in each data type
#           - checking inside the loop:
#               a-cut --> is sequence (0.1 to 1.0)
#               kde.n --> normalised density
#               IF
                    kde.n greater than or equal to a-cut
#               THEN
                    alpha = 1
#               ELSE
                    alpha = 0
#       d. write out the file for each alpha
#==========================================================================================
```

Figure 4.4: Pseudo-code for generating the normalised density surfaces and their α-cuts for settlements.

4.4 Analysing the Settlements' Spatial Patterns

4.4.1 Analysing Settlement Representation in Different Sources

To restate, the data used for modelling the rural settlements comes from different sources: formal data (BS76 and POST) and contributed data (POI). The main cause for the variation between the models is because different databases have different records and attributes, and consequently represent different views of the village. This section focuses on exploring the variation between these data sources under the assumption that each data source represents a variable, and there is a need to gain insight into the relationship between any pairs of these variables. A widely accepted method in statistics for modelling the relationship between two variables is linear regression. It has been applied in this study using the linear model function - "lm" in R (Wilkinson and Rogers 1973; Chambers 1992) for three pairs (POI vs BS76, POI vs POST and BS76 vs POST). The analysis below proceeds as follows. It first creates a scatter plot of the values of the two variables against each other. By doing this, it is possible to observe visually whether there is a linear association between the two variables. Then two statistical measures are used to examine the performance of these models in developing such relationships. These are correlation coefficient (r) and root mean square error (RMSE).

Correlation Coefficient (r)

Correlation (r) is the statistical method (measure) for assessing the association between two quantitative variables, x and y. When they plotted together, how close to a straight line is the scatter of points. It simply measures the degree to which the two vary together (Freeman and Young 2009). The sign of the correlation r indicates the direction of the relationship: $r > 0$ for a positive association (i.e. as the values of one variable increase the values of the other variable increase) and $r < 0$ for a negative association (i.e. as the values of one variable increase the values of the other variable decrease). r is always a number between −1 and 1; perfect correlation occurs when r = ±1 only, which means the points lie exactly on a straight line, as it moves away from 0 towards ±1 as the straight-line relationship gets stronger. The definition of this statistic can be expressed mathematically as:

Give as set of n pairs of observations $(x_1, y_1), (x_1, y_1), \dots (x_n, y_n)$ the formula for the ***correlation confident*** r is given by:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Root Mean Square Error (RMSE).

One of the most widely used statistics in GIS for evaluating the overall quality of a regression model is the root mean square error (RMSE, also known as root mean square deviation). RMSE measures how much error there is between two datasets. RMSE usually compares a predicted value by a hypothetical model and an observed value (GISGeography, 2015). In other words, it measures the quality (goodness) of the fit between the actual data and the predicted model (Salkind, 2010). In the regression analysis, the predicted values are more or less different from the actual observations unless the relationship or correlation is perfect (i.e. $r = \pm 1$). These differences are called prediction errors or residuals. For every data point, these residuals are measured by calculating the distance vertically from the point (actual value) to the corresponding y value on the regression line, and the square of that distance is then quantified. Large distances are indicative of large errors. The sum of the squared values for all data points is added up, and then divided by the total number of points. Finally, the square root is taken. The smaller the RMSE, the closer the fit is to the data. The mathematical formula for computing the RMSE is:

RMSE is the average vertical distance of the actual data points from the fitted. Mathematically, RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}$$

where n denotes the size of the sample or number of observations; x_i represents individual values, and i represents the current predictor.

Since the comparison here is between the same address points in the three different sources, the explanatory variable (or the independent variable which is plotted on the horizontal axis, x) and the response variables (the dependent variable plotted in the vertical axis, y) are not clearly distinguished. However, for consistency reasons, the POI data always represent the dependent variable whenever it is considered. Figure 4.5 explains the pseudo-code for applying this analysis, whereas the real R script is available in Appendix (6.3).

```
#==========================================================================================
# Script Title:  linear regression model - Linear Regression plots 4 the data in the 3 sources
#
# 1. Install & Load the required libraries
    GISTools, spgwr, rgdal, raster & shapefiles
# 2. Set the working directory where the data exist and read in the data
#        a. read the settlements' point data
#        b. read the polygon of H&B district
# 3. Check & investigate the data to identify villages with at least 10 points in any data type
#        a. define lists for each data type and village
            village.list &  data.source.list
# 4. Generate Helper Functions
#     ``````````````````````````
#        a. normalising the data to stretch from 0--> 1
#        b. KDE plot function
#        c. do regression.comparison
#            - get two kde surfaces: kde.1 & kde.2
#            - index to identify cells that are both zero
#            - exclude these from the scatter plots and the regressions
#            - assign them to variables and set the plot parameters
                pt1 = kde.1
                pt2 = kde.2
#            - scatter plots and regression in a variable
#            - return the result
# 5. Main Loops
#    ```````````
#        a. Set the working directory where the results to be saved
#        b. Loop 1 generates and plots KDEs
#        c. Loop 2 generates regressions for each village and each data type
#            - Call in the village and set each data type as a variable
#            - set the plots parameters
                    if statement just to keep the POI in the y-axis always
                        so check and replace if not true
#            - call the regression function
#            - get corrletion and RMSE values from the regression model
#        d. write the results of the regression plots and save (r & RMSE) values in files
#==========================================================================================
```

Figure 4.5: Pseudo-code to apply the linear regression model in R, full script is available in Appendix (6.3).

4.4.2 Analysing Inclusion

It is also important to understand the variation between the different representations of the rural settlements modelled as fuzzy objects across different data sources. From the set of α-cuts, it is possible to parameterise the degree to which a group of address points located in a village are within the fuzzy representation of that village. Point in polygon operation has been used as a method to check whether many points are placed inside or outside a polygon. Thus, it would serve as an ideal tool for use in an inclusion or containment analysis. So for this study, a point in polygon operation is performed on four sets of data: the villages extracted from the BS76 data, POST data, POI data and all the three combined together. For each village with at least 10 addresses, the number of points contained in the different α-cuts are computed. This gives a measure of the proportion of the points that are labelled as being part of a particular village within the associated α-cuts.

In the next phase of this analysis, as a consequence of exploring the interaction between the number of points included and their α-cuts, it is also possible to specify a minimum threshold value of the points to be included in any α-cut. In other words, instead of considering the number of points captured in a village catchment at different α-cuts (where α is the degree of membership), a percentage of the points to be identified by the α-cuts is regarded as a threshold limit of inclusion. Any specific threshold value will yield a particular instance of the α-cut that satisfies this threshold. For example, to determine what α-cut is needed to generate a fuzzy area that includes at least 75% of the original points, several threshold values (i.e. percentage of coverage from 50 to 95%) are examined for each of the respective village in all data sources.

One further consideration is the density of points within the α-cuts that satisfies the threshold values. Different villages may have very different spatial structures – from a dispersed linear village to a highly concentrated and concentric one. This is to investigate whether there is any pattern or structure in the data in terms of the association between the optimum α-cut and the point density. Basically, this could give an idea of how the points are spatially distributed on the village area and thus what membership value (α-cut) can be expected. It is likely that as the point density increases the probability of scattering decreases. Consequently, as a subsequent step in determining the α-cut needed to meet the threshold demand, the density of the points is calculated based on the area of that α-cut. This has been done using R script for every individual village in the considered data (See Appendix 6.4); as the general structure of the code is explained in Figure 4.6.

```
#==========================================================================================
# Script Title: Pseudo-code 4 Points in Polygon Analyses
# general form apply in 4 different R scripts for each data types
#__________________________________________________________________________________________
# 1. Install & Load the required libraries
   GISTools, spgwr, rgdal, raster & shapefiles
# 2. Set the working directory where the data exist and read in the data
#       a. read the settlements' point data
#       b. read the polygon of H&B district
# 3. Check & investigate the data to identify villages with at least 10 points in any data type
#       a. define lists for each data type and village
            village.list &  data.source.list
# 4. Generate Helper Functions
#   ``````````````````````````
#       a. normalising the data to stretch from 0--> 1
#       b. KDE plot function
#       d. calculate density in a-cut area (in km2)
#           - get in the kde surface,
# 5. Main Loops
#   ```````````
#       a. Set the working directory where the results to be saved
#       b. Loop 1 generates and plots KDEs
#       c. Set the parameters for saving the results of the point N poly analysis
            village.list &  data.source.list
            alpha cut     -->   is sequence (0.05 to 0.95 by = 0.05)
            threshold     -->   is sequence (0.5 to 0.95, by = 0.05)
            result.tab    -->   is matrix to save the results

#       d. Loop 2 - Generate the alpha cuts and apply the Point in Polygons
#           - loop through all village in each data type {different file/dataType}
#           - Loop through threshold sequence and set condition to go through alpha cut
                call KDE function to get the density surface for the village
                do an alpha cut until all threshold is met
                transfer the alpha cut raster to polygon
                count number of points in polygon / alpha,  {n.in.poly}
                calculate the proportion %%,    {n.in.poly/total.point}
                ### now do test for Threshold
                Calculate density of points:
                    (1) calculate the area of the polygon {/1000000 to get the area in km2}
                    (2) then divide the total no. of point by the area}
            - call density funtionto get density values at 0.1, 0.3 & 0.5 alpha cuts
            - return all vlaues in dataframe to save out
# 6. write the result out in csv tables
#==========================================================================================
```

Figure 4.6: General structure of the R scripts executed to apply the inclusion analysis for all settlements in the different data sources.

4.5 Results and Discussion

4.5.1 Modelling based on the Density of Houses

The results obtained from applying the KDE analysis to approximate the fuzzy footprints of settlements that exist in all data sources and have a minimum of 10 addresses are presented in Appendix (2). Figure 4.7 presents a subset of the results for three villages: Barwell, Desford and Stoke Golding, evaluated using each data source. In the figures, the normalised density is, in the top left corner, graded from yellow to red in 10 levels representing the values from 0 to 1. As shown in the legend, darker red corresponds to higher densities where the larger intensity of points indicates a larger values of fuzzy membership. The remaining maps in the figure show the Boolean mapping for the nine α-cuts which correspond to areas within the villages that have membership values greater than or equal to α, and zero for areas outside this region. These maps suggest that the largest α-cuts consistently have the smallest area of the examined villages and the smallest α-cut has the largest, as expected. In addition some villages' areas in some α-cuts have discontinuous areas, as appears in many cases in Figure 4.7. This implies that the density of points varied around the settlements, which is logical since some areas are denser than others for many different local reasons and may say something about the nature of the village. The address points are overlaid on top of each map in the figure for visualisation purposes. For comparison purposes, all the maps are shown within the full geographical extent of Hinckley and Bosworth Borough. The complete results for all considered villages are presented in Appendix (2).

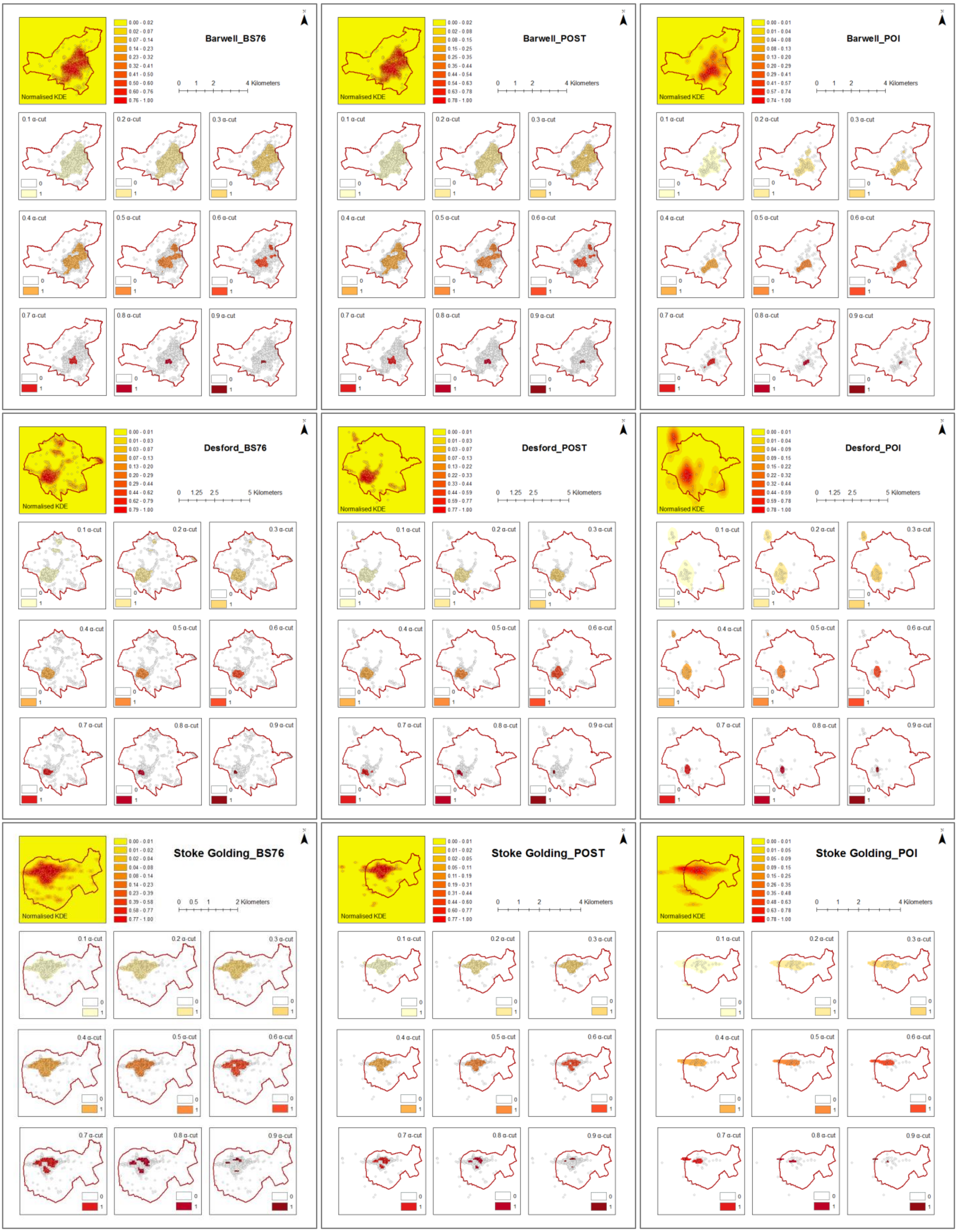


Figure 4.7: Fuzzy representation for Barwell, Desford & Stoke Golding in each dataset, normalised density in the top left corner, with their nine α-cuts underneath each village. See Appendix (2) for full size images.

It is apparent that KDE works better with compact zones, where areas have a high intensity of points, which produce a much better representation than areas with a varying intensity of points (i.e. sparsely populated area). This also accords with Down's (2010) observations, which showed that, despite the wide spread of KDE, it could perform poorly in some situations and for specific types of point patterns. One explanation for the poor performance of KDE in this case could be its sensitivity to bandwidth selection, as different values produce dramatically different results. Given this reason, the selection of bandwidth is generally open to a certain degree of subjectivity.

4.5.2 Comparative Analysis of Settlement Representation in Different Sources

The results of the regression plots are presented in Appendix (3) and Table 4.1 compares the correlation coefficients (*r*) and the root mean squared error (*RMSE*) gained from the linear regression model for the different data sources. The results of the regression analysis can be grouped along two lines. On the one hand, the majority of the settlements (Barlestone, Barwell, Burbage, Desford, Earl Shilton, Groby, Market Bosworth, Nailstone, Newbold Verdon, Ratby, Stanton under Bardon and Stoke Golding), which have a common characteristic in terms of the strong association between the BS76 and POST variables. It can be even generalised that a large number of settlements have a perfect positive linear relationship with the BS76 data versus the POST data except for a few outliers. This is initially expected because these are originally the same dataset, which mostly have many duplicated address points. Thus, the other two plots (POI vs BS76 and POI vs POST) should look identical. This has been typically seen in Barwell, Market Bosworth and Newbold Verdon (Figure 4.8). The squared correlation of the BS76 versus POST data emphasises the perfect relationship between them; and thus the other two pairs (POI vs BS76 and POI vs POST) look very similar even though their *r* values might differ, as is the case in Market Bosworth. The overall patterns of the POI compered to BS76 and POST show a clear direction which moves from lower left to upper right with larger variation in *y* for larger values of *x*. This finding confirms that for large areas of the plots the association between the village representations, where most of the address points overlapped, are similar. Another observation indicates that the RMSE values for this group of settlements are very close to 0. That, again, stresses on the fact that lower values of RMSE indicate better fit, as that reflect how much errors there is between the datasets.

Table 4.1: Values of the correlation coefficients (r) and the root mean squared error (RMSE) between each pair of the data sources.

Village Name	POI vs BS76		POI vs POST		BS76 vs POST	
	r	RMSE	r	RMSE	r	RMSE
Barlestone	0.67	0.02	0.66	0.02	1.00	0.00
Barwell	0.82	0.03	0.82	0.03	1.00	0.00
Burbage	0.88	0.02	0.87	0.03	1.00	0.01
Desford	0.77	0.02	0.79	0.02	0.96	0.01
Earl Shilton	0.66	0.02	0.65	0.02	1.00	0.00
Groby	0.74	0.02	0.74	0.02	0.98	0.02
Higham on the Hill	0.64	0.01	0.69	0.02	0.90	0.01
Market Bosworth	0.67	0.02	0.66	0.02	1.00	0.00
Nailstone	0.88	0.02	0.87	0.02	0.99	0.00
Newbold Verdon	0.84	0.02	0.84	0.01	1.00	0.00
Osbaston	0.69	0.02	0.59	0.02	0.86	0.01
Peckleton	0.14	0.01	0.28	0.03	0.19	0.05
Ratby	0.82	0.02	0.81	0.02	1.00	0.00
Stanton under Bardon	0.83	0.02	0.82	0.02	0.98	0.01
Stoke Golding	0.75	0.03	0.79	0.03	0.99	0.01
Sutton Cheney	0.46	0.03	0.51	0.04	0.28	0.04
Twycross	0.81	0.03	0.48	0.04	0.43	0.04
Witherley	0.40	0.01	0.65	0.02	0.53	0.04

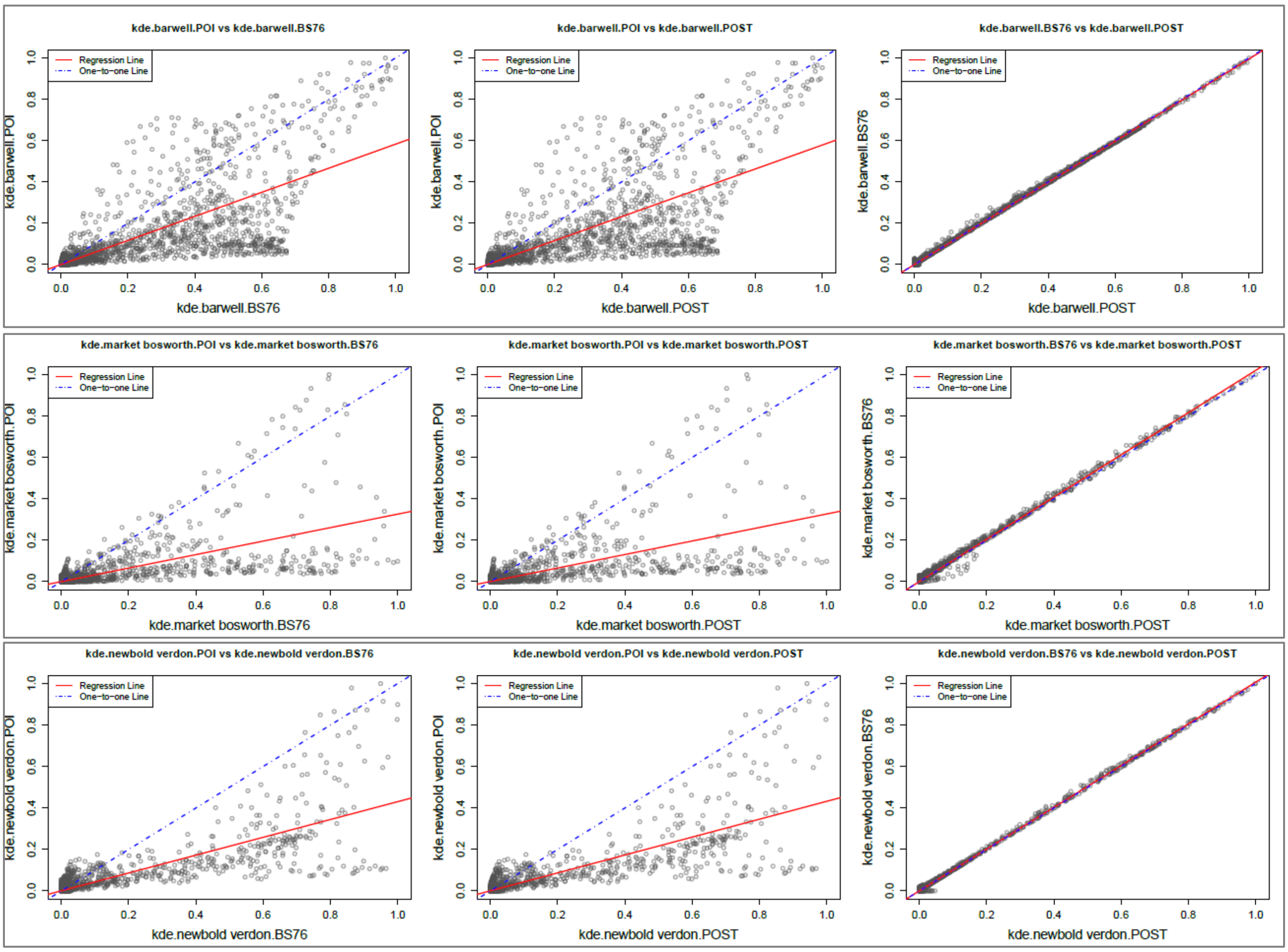


Figure 4.8: Regression Plots for Barwell, Market Bosworth and Newbold Verdon, which show strong positive linear relationship between the BS76 and POST data and other moderate relation between the other pairs of data (POI vs BS76 & POI vs POST).

On the other hand, there are some striking results for the regression plots in the second group of the analysed villages. This include the settlements: Higham on the Hill, Osbaston, Peckelton, Sutton Cheney, Twycross and Witherley. Figure 4.9 shows three examples from this group. It is clear from the plots of these settlements that the fitted models do not follow a linear relationship. Furthermore, the r values in Table 4.1 show weak association between each pair of datasets, especially in Peckleton. However, the *RMSE* values are still relatively small. . This could be explained when looking at the spatial distribution of the original points for these settlements in each source type, in which the actual patterns of the points are different and not responding to each other.

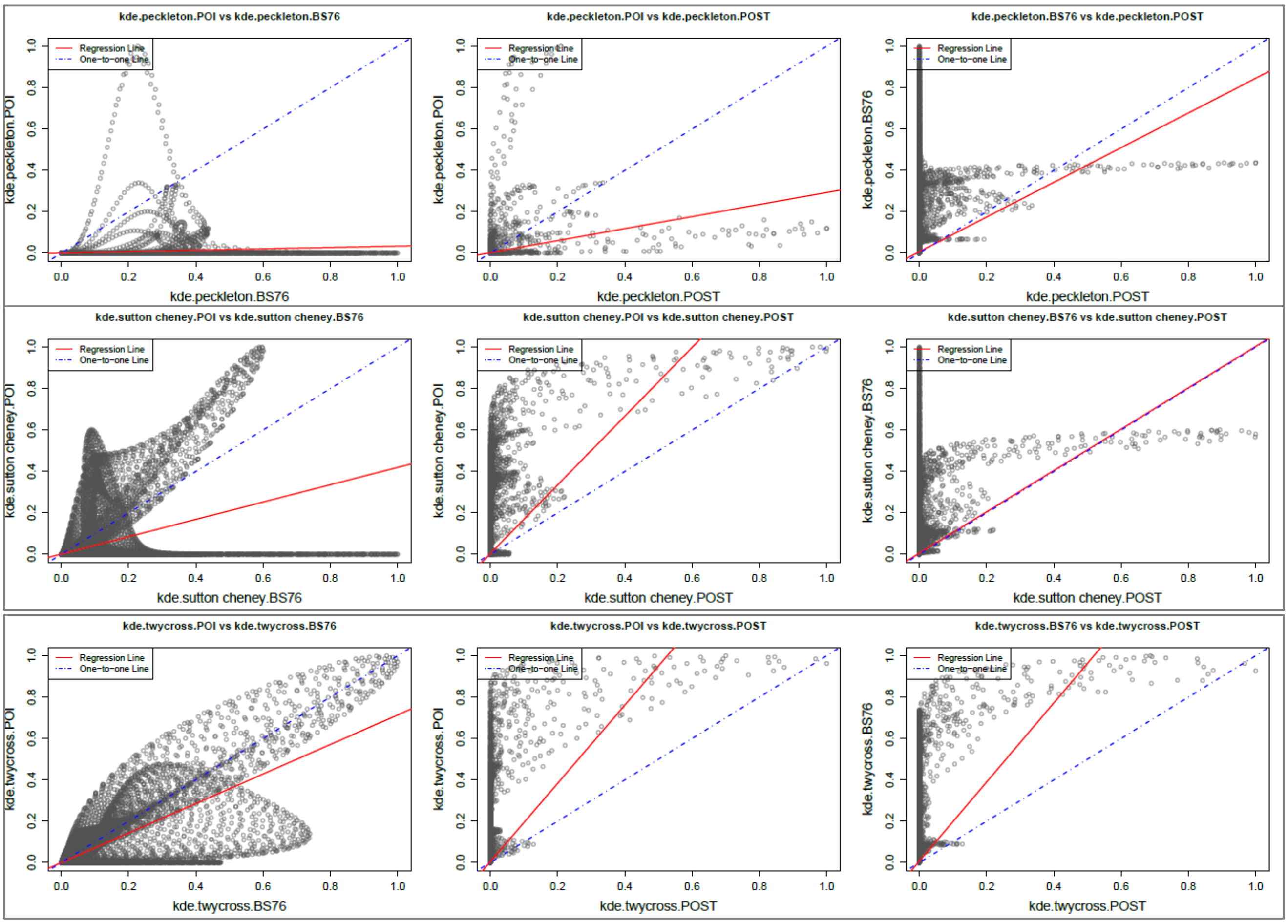


Figure 4.9: Regression plots for Peckleton, Sutton Cheney and Twycross that showing no apparent pattern in the association between the pairs of the different data sources (not suitable for linear regression).

4.5.3 Analysis of Inclusion

The initial purpose of this analysis is to examine the proportion of address points included within the fuzzy model of the villages in each α-cuts. However, this gives rise to an alternative view of the analysis, on which the concern moves to identifying which α-cuts that satisfy a minimum threshold value of the included points. That means which α-cut guarantees the inclusion of a particular percentage of the original address points. It is necessary here to first clarify the difference between the α-cuts and the thresholds in relation to the concept of '*villageness*'. As already mentioned earlier, a village is identified from a set of local points that share a settlement name in their addresses. An *α-cut* then refers to the degree of fuzzy membership of the addresses belonging to that named village. For instance, an α-cut of 0.4 specifies all areas in the village that have a fuzzy membership value greater than or equal to 0.4. A *threshold,* on the other hand, refers to the percentage of the minimum number of the address points to be considered within the village. Hence, it could be generally assumed that, as the threshold value increases, the area captured of the village increases with the possibility of places with low membership.

Following the approach identified above (Section 4.4.2), a set of tables are created to record the results of the containment analysis. These are presented in four tables: Tables 4.2, 4.3, 4.4 & 4.5 for the BS76 data, POST data, POI data and all data combined, respectively. For ease of reading, these tables are structured, in exactly the same way, such that each row records the results from one settlement. Also, the first three columns contain information about the village identification ID in the dataset, its name and the original number of address points. The remaining columns generally identify ten percentages of inclusion (i.e. 50%, 55%, 60% … 95%). Each percentage is further subdivided into three fields to report the results in relation to: (1) the α-cut that assures the threshold values, (2) the density of the address points that fall within this limits based on (3) the area of this α-cut; (as the density is given by dividing the number of points within the α-cut by the area of that α-cut in Km^2). The results given in these tables are discussed individually in a uniform manner below, and then comparisons are made between them.

Results of the BS76 data:

Table 4.2 presents the results obtained from the containment analysis of the BS76 data. Twenty-eight settlements with at least 10 addresses are identified, and these are ranked in descending order by the number of points. In the table, the grades of membership, α-cuts, are displayed in light orange; as the values get smaller the colour becomes lighter. Likewise, point densities are graded in light green whilst lighter green shows low density. The most striking results to emerge from these data are discussed under two main categories in the following paragraphs:

Firstly, exploring α-cut profiles:

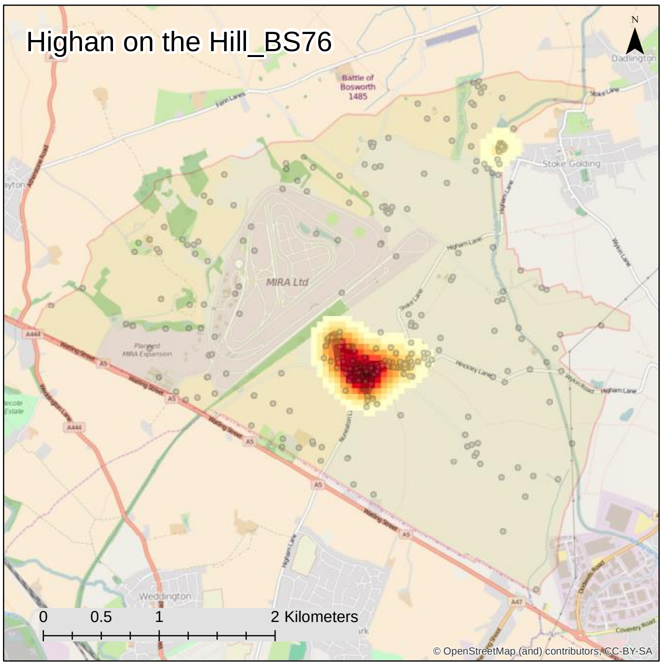
- It can be seen from the table that some villages fail to capture all considered thresholds, as they drop before reaching the maximum threshold value (95%). Examples include Higham on the Hill, Stanton under Bardon, Nailston, Sheepy and Witherley. This can be further illustrated in Figure 4.10, which shows maps of Higham on the Hill and Stanton under Bardon with their address points on top of a series of α-cuts (graded from red to yellow), with bar charts indicating the optimal α-cut for each threshold. It is apparent from this figure that some of the address points are disregarded from the fuzzy model, as the majority of these points have a non-zero membership value less than 0.1. Even the smallest α-cut only captures 65% of the points in the case of Higham on the Hill or 75% in Stanton under Bardon. This is presumably because the fuzzy model is based on the density of points, which appears to be clustered in the middle and scattered further apart.

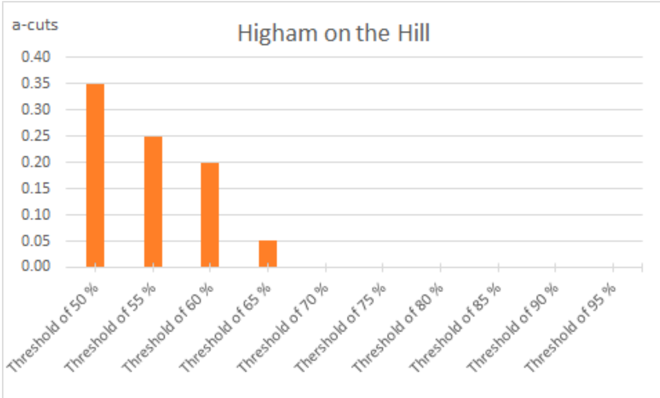
Table 4.2: Results of the containment analysis for the BS76 data. For all considered threshold, the optimum α-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these α-cuts (points/km^2).

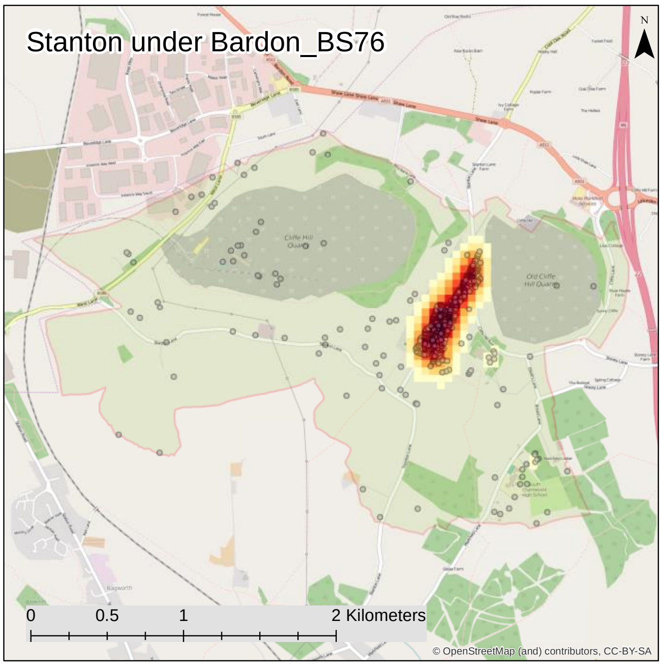
ID	Village Name	No. of Points	Threshold of 50 %			Threshold of 55 %			Threshold of 60 %			Threshold of 65 %			Threshold of 70 %		
			α-cut	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*
5	Burbage	6979	0.60	2478	1.51	0.55	2372	1.84	0.55	2372	1.84	0.50	2237	2.18	0.45	2138	2.54
6	Clarendon	5702	0.50	3237	0.89	0.45	3056	1.15	0.45	3056	1.15	0.40	2856	1.45	0.40	2856	1.45
18	Earl Shilton	4884	0.50	2982	0.94	0.50	2982	0.94	0.45	2893	1.16	0.45	2893	1.16	0.40	2745	1.33
16	De Montfort	4532	0.50	2829	0.82	0.45	2658	1.06	0.45	2658	1.06	0.40	2516	1.26	0.40	2516	1.26
25	Barwell	4382	0.50	3038	0.79	0.45	2907	0.99	0.45	2907	0.99	0.45	2907	0.99	0.40	2765	1.21
23	Groby	3282	0.55	2381	0.73	0.50	2237	0.90	0.50	2237	0.90	0.45	2151	1.03	0.40	2059	1.21
21	Trinity	2988	0.65	3085	0.49	0.60	2821	0.68	0.60	2821	0.68	0.55	2734	0.82	0.55	2734	0.82
13	Markfield	2291	0.60	2629	0.46	0.55	2560	0.52	0.50	2483	0.59	0.45	2411	0.66	0.40	2348	0.72
27	Castle	2076	0.50	4384	0.24	0.45	3955	0.33	0.45	3955	0.33	0.40	3543	0.44	0.40	3543	0.44
19	Ratby	2010	0.70	2664	0.39	0.65	2648	0.46	0.65	2648	0.46	0.60	2510	0.54	0.55	2439	0.61
22	Desford	1864	0.50	1961	0.50	0.45	1924	0.56	0.40	1892	0.60	0.35	1823	0.68	0.30	1729	0.79
28	Newbold Verdon	1429	0.60	2449	0.32	0.55	2411	0.39	0.55	2411	0.39	0.55	2411	0.39	0.50	2346	0.44
24	Bagworth & Thornton	1282	0.60	1009	0.65	0.55	901	0.82	0.50	816	0.99	0.45	728	1.26	0.45	728	1.26
2	Market Bosworth	1140	0.55	2212	0.27	0.50	2094	0.31	0.45	2010	0.36	0.40	1909	0.41	0.35	1769	0.48
14	Barlestone	1118	0.70	2919	0.20	0.65	2745	0.25	0.65	2745	0.25	0.60	2661	0.29	0.55	2601	0.32
1	Stoke Golding	797	0.70	2182	0.20	0.70	2182	0.20	0.65	2082	0.24	0.60	1982	0.27	0.55	1949	0.31
17	Witherley	767	0.60	807	0.50	0.55	701	0.62	0.45	548	0.88	0.35	453	1.20	0.35	453	1.20
10	Sheepy	692	0.30	243	1.53	0.25	195	1.98	0.10	98	4.69	0.10	98	4.69	0.05	54	10.56
20	Peckleton	626	0.75	529	0.59	0.45	174	1.98	0.40	162	2.49	0.35	133	3.16	0.25	92	4.93
7	Twycross	508	0.45	173	1.60	0.40	150	2.03	0.40	150	2.03	0.35	132	2.54	0.30	115	3.15
15	Higham on the Hill	483	0.35	928	0.27	0.25	743	0.36	0.20	680	0.44	0.05	301	1.12	0.00	0	0.00
26	Shackerstone	465	0.55	326	0.77	0.50	288	0.96	0.45	235	1.19	0.40	222	1.48	0.40	222	1.48
12	Stanton under Bardon	366	0.70	2285	0.08	0.60	2058	0.10	0.55	1874	0.12	0.45	1616	0.15	0.20	1027	0.25
3	Sutton Cheney	339	0.40	102	1.76	0.35	86	2.36	0.30	68	3.10	0.25	55	4.03	0.20	47	5.40
9	Nailstone	260	0.55	1502	0.09	0.50	1376	0.11	0.40	1224	0.13	0.35	1161	0.15	0.30	1077	0.17
11	Osbaston	161	0.20	131	0.75	0.20	131	0.75	0.20	131	0.75	0.15	103	1.09	0.10	79	1.70
8	Carlton	139	0.60	881	0.08	0.50	812	0.11	0.50	812	0.11	0.40	697	0.13	0.30	565	0.18
4	Cadeby	133	0.25	246	0.29	0.20	200	0.39	0.15	151	0.67	0.15	151	0.67	0.15	151	0.67

Table 4.2: (**continued**) Results of the containment analysis for the BS76 data. For all considered threshold, the optimum α-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these α-cuts (points/km^2).

ID	Village Name	No. of Points	Threshold of 75 %			Threshold of 80 %			Threshold of 85 %			Threshold of 90 %			Threshold of 95 %		
			α-cut	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*
5	Burbage	6979	0.45	2138	2.54	0.40	2032	2.87	0.35	1959	3.15	0.30	1887	3.40	0.15	1618	4.15
6	Clarendon	5702	0.35	2708	1.73	0.35	2708	1.73	0.30	2538	1.97	0.25	2375	2.20	0.10	1736	3.16
18	Earl Shilton	4884	0.35	2625	1.49	0.35	2625	1.49	0.30	2566	1.67	0.25	2397	1.88	0.20	2223	2.10
16	De Montfort	4532	0.35	2390	1.50	0.30	2162	1.83	0.30	2162	1.83	0.25	2003	2.09	0.20	1840	2.35
25	Barwell	4382	0.40	2765	1.21	0.35	2645	1.38	0.30	2487	1.57	0.25	2362	1.71	0.15	2062	2.04
23	Groby	3282	0.40	2059	1.21	0.35	2019	1.34	0.30	1932	1.51	0.25	1809	1.68	0.15	1501	2.09
21	Trinity	2988	0.50	2665	0.92	0.50	2665	0.92	0.45	2589	0.98	0.35	2396	1.16	0.30	2319	1.23
13	Markfield	2291	0.35	2252	0.79	0.30	2140	0.87	0.20	1833	1.09	0.10	1401	1.50	0.05	1060	2.06
27	Castle	2076	0.35	3267	0.52	0.35	3267	0.52	0.30	2990	0.61	0.25	2791	0.70	0.20	2577	0.79
19	Ratby	2010	0.50	2404	0.66	0.45	2351	0.72	0.40	2241	0.78	0.35	2133	0.85	0.05	1281	1.49
22	Desford	1864	0.25	1607	0.88	0.20	1418	1.06	0.10	1013	1.59	0.05	722	2.35	0.00	0	0.00
28	Newbold Verdon	1429	0.45	2321	0.48	0.40	2265	0.52	0.35	2131	0.58	0.30	2065	0.62	0.10	1436	0.95
24	Bagworth & Thornton	1282	0.40	630	1.56	0.30	474	2.20	0.15	303	3.68	0.05	172	6.80	0.00	0	0.00
2	Market Bosworth	1140	0.30	1610	0.55	0.25	1466	0.64	0.20	1317	0.74	0.05	774	1.36	0.00	0	0.00
14	Barlestone	1118	0.55	2601	0.32	0.45	2371	0.39	0.35	2195	0.44	0.20	1855	0.54	0.10	1408	0.77
1	Stoke Golding	797	0.55	1949	0.31	0.45	1827	0.36	0.40	1739	0.39	0.30	1626	0.44	0.05	1013	0.75
17	Witherley	767	0.30	397	1.48	0.10	186	3.30	0.05	131	5.08	0.00	0	0.00	0.00	0	0.00
10	Sheepy	692	0.05	54	10.56	0.05	54	10.56	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00
20	Peckleton	626	0.20	80	6.02	0.10	53	9.85	0.05	33	17.98	0.05	33	17.98	0.00	0	0.00
7	Twycross	508	0.20	79	4.84	0.05	31	15.14	0.05	31	15.14	0.05	31	15.14	0.00	0	0.00
15	Higham on the Hill	483	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00
26	Shackerstone	465	0.35	185	1.88	0.15	82	4.80	0.15	82	4.80	0.05	40	10.93	0.00	0	0.00
12	Stanton under Bardon	366	0.05	487	0.60	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00
3	Sutton Cheney	339	0.15	38	7.35	0.15	38	7.35	0.10	28	11.20	0.10	28	11.20	0.05	20	16.58
9	Nailstone	260	0.20	871	0.22	0.05	414	0.55	0.05	414	0.55	0.00	0	0.00	0.00	0	0.00
11	Osbaston	161	0.10	79	1.70	0.10	79	1.70	0.05	45	3.33	0.05	45	3.33	0.00	0	0.00
8	Carlton	139	0.20	435	0.25	0.10	314	0.39	0.10	314	0.39	0.05	226	0.56	0.00	0	0.00
4	Cadeby	133	0.15	151	0.67	0.10	100	1.08	0.05	62	1.93	0.05	62	1.93	0.00	0	0.00







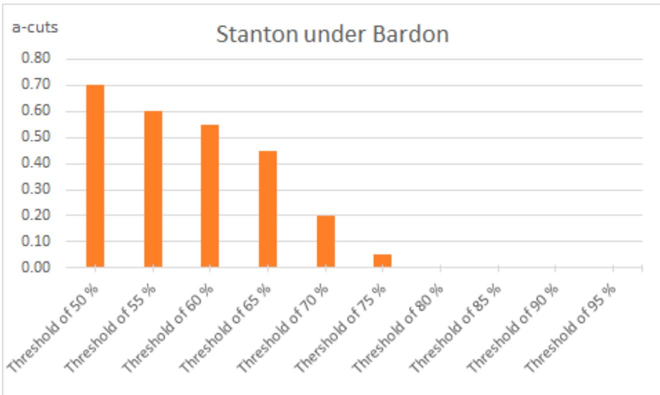


Figure 4.10: Maps of two settlements (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold).where the smallest α-cuts drop before reaching larger thresholds as their points tend to cluster in the middle and scatter further apart.

- Another indication of the addresses being spread out in the area might be when the optimum α-cuts for all thresholds are fairly small as in the case of Cadeby, Osbaston and Sheepy which have an α-cut values mostly just under 2.5. Figure 4.11 presents the instances of Cadeby and Osbaston. Interestingly, the figure also reveals that the fuzzy model of the villages are not always continuous as the points density are varied across the areas. This in turn highlights number of clusters within the villages with some diffuse points, with very low membership values (again greater than zero and less than 0.1), between or around theses clusters.

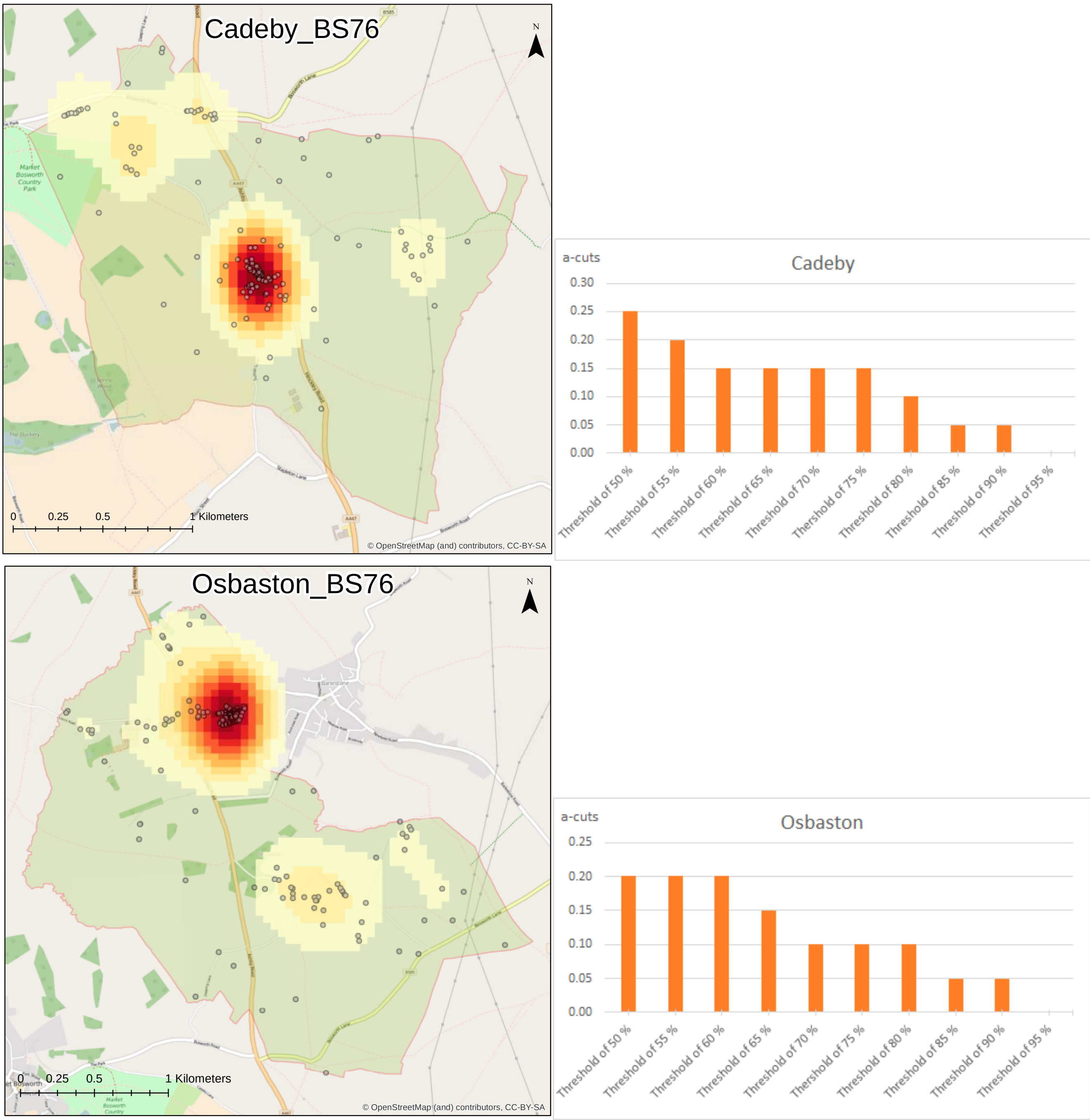


Figure 4.11: Map of two settlements(on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold). These show fairly small α-cuts for all thresholds suggesting the possibility of their addresses being spread out.

- On the contrary, when the α-cut values tend to be large with very little variation across the thresholds, then this implies the compact nature of the settlements. Take, for example, Castle, Trinity, De Montfort and Earl Shilton, which have a gradual decline through the thresholds with α-cut of 0.2 to capture 95% of the points. This can be clearly seen in the maps of Castle and Earl Shilton (with fewer dispersed points in the southeast area of Earl Shilton; $0 < \alpha < 0.1$) in Figure 4.12.

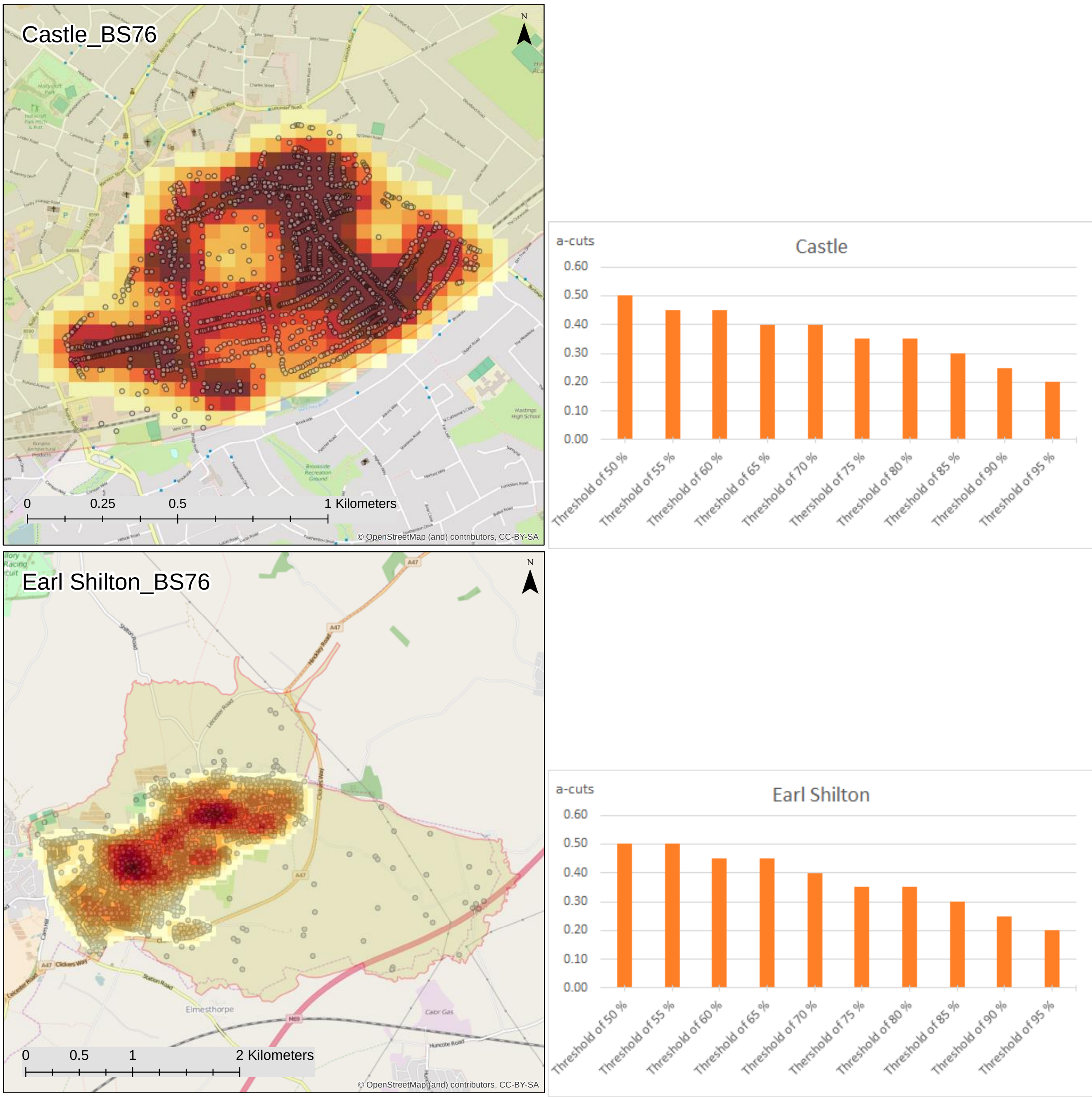


Figure 4.12: Map of two villages (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold), in which small transition between α-cuts values reflecting the compactness of their addresses.

Secondly, exploring relationship with point densities:

- General observation of the density values in Table 4.2 shows that the last/bottom part of the table (last third) appears to have lower density values compared to the upper part of the table. This seems to be in line with number of address points as, for larger numbers, the density tends to be higher (darker green).

- However, there are some exceptions. Looking, for example, at the minimum threshold (50%), it is found that Castle, which has the maximum density (4383.78), has fewer points in a small area than other settlements with lower density like Burbage, Clarendon, and Earl Shilton, which have larger numbers of address points. Moreover, Sutton Cheney with 339 address points has the lowest density (102.18), whereas other settlements with fewer addresses have larger density, such as Carleton and Cadeby, the last two in the table.

- Looking horizontally, density values decline with the increase in the threshold limits. This relation is expected to be applied in all settlements (rows) since a particular threshold indicates the proportion of the points included. Thus, when considering a certain village (one row with the same spatial points structure), half of its points (threshold of 50%) will be more intense than three-quarters of its points (threshold of 75%).

Results of the POST data:

Forty-nine villages are identified in the POST data, and the results of their containment analysis are recorded in Table 4.3 in descending order by number of points. The membership grades (α-cuts) and point densities are also in colour grades of light orange and light green, as presented in the previous results (Table 4.2). Taken together, these results suggest the following:

Firstly, exploring the α-cut profiles:

- The optimum α-cuts for some villages drop out before reaching the maximum threshold, and this reflects the fact that some of their address points are excluded from the fuzzy model (areas with a very small membership value greater than zero and less than 0.1). This in turn attests to the spread characteristics of those points. This can be exemplified in Kirkby Mallory, Atterton and Appleby Magna (Figure 4.13); the smallest α-cut only covers 70 or 85% of those villages. This is even clearer when the total number of address points is remarkably small, as in Atterton (15) and Appleby Magna (10).

Table 4.3: Results of the containment analysis for the POST data. For all considered threshold, the optimum α-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these α-cuts (points/km2).

ID	Village Name	No. of Points	Threshold of 50 %			Threshold of 55 %			Threshold of 60 %			Threshold of 65 %			Threshold of 70 %		
			α-cut	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*
49	Burbage	6635	0.60	2484	1.44	0.55	2363	1.77	0.55	2363	1.77	0.50	2256	2.10	0.50	2256	2.10
48	Earl Shilton	4763	0.50	2907	0.93	0.50	2907	0.93	0.45	2843	1.13	0.45	2843	1.13	0.40	2702	1.32
47	Barwell	4244	0.50	3023	0.79	0.50	3023	0.79	0.45	2880	1.00	0.45	2880	1.00	0.40	2742	1.20
46	Groby	2954	0.55	2377	0.71	0.55	2377	0.71	0.50	2256	0.87	0.50	2256	0.87	0.45	2149	1.00
45	Ratby	1904	0.70	2720	0.36	0.65	2713	0.43	0.65	2713	0.43	0.60	2584	0.49	0.55	2486	0.57
44	Desford	1497	0.60	2288	0.34	0.55	2179	0.39	0.50	2060	0.46	0.45	1948	0.54	0.45	1948	0.54
43	Newbold Verdon	1376	0.60	2426	0.31	0.60	2426	0.31	0.55	2406	0.38	0.55	2406	0.38	0.50	2337	0.43
42	Barlestone	1080	0.70	2893	0.19	0.65	2753	0.24	0.65	2753	0.24	0.60	2645	0.28	0.55	2614	0.32
41	Market Bosworth	1048	0.55	2249	0.25	0.50	2097	0.30	0.50	2097	0.30	0.45	2019	0.36	0.40	1934	0.40
40	Stoke Golding	793	0.65	2073	0.23	0.65	2073	0.23	0.65	2073	0.23	0.60	1939	0.27	0.55	1925	0.30
39	Bagworth	604	0.45	1629	0.21	0.45	1629	0.21	0.40	1532	0.29	0.40	1532	0.29	0.40	1532	0.29
38	Thornton	460	0.60	2101	0.13	0.60	2101	0.13	0.55	2006	0.15	0.55	2006	0.15	0.50	1889	0.19
37	Witherley	313	0.70	2006	0.08	0.65	1865	0.10	0.60	1742	0.11	0.55	1670	0.13	0.50	1586	0.14
36	Sheepy Magna	302	0.55	1247	0.13	0.45	995	0.18	0.45	995	0.18	0.35	855	0.25	0.35	855	0.25
35	Higham on the Hill	284	0.60	2385	0.06	0.55	2319	0.07	0.55	2319	0.07	0.45	1875	0.10	0.40	1646	0.13
34	Stanton under Bardon	281	0.75	2339	0.06	0.70	2207	0.08	0.70	2207	0.08	0.60	2027	0.10	0.60	2027	0.10
33	Fenny Drayton	222	0.65	1515	0.08	0.65	1515	0.08	0.60	1438	0.09	0.50	1255	0.13	0.50	1255	0.13
32	Nailstone	215	0.65	1537	0.07	0.60	1511	0.08	0.55	1520	0.09	0.50	1369	0.11	0.45	1255	0.12
31	Botcheston	198	0.35	1018	0.12	0.35	1018	0.12	0.35	1018	0.12	0.30	831	0.18	0.30	831	0.18
30	Stapleton	195	0.50	1908	0.05	0.35	1577	0.08	0.35	1577	0.08	0.35	1577	0.08	0.25	1350	0.11
29	Kirkby Mallory	176	0.50	1562	0.06	0.45	1461	0.07	0.35	1105	0.10	0.30	1039	0.12	0.25	903	0.14
28	Congerstone	151	0.60	1801	0.04	0.55	1513	0.06	0.55	1513	0.06	0.45	1306	0.08	0.45	1306	0.08
27	Twycross	148	0.55	1280	0.06	0.45	922	0.10	0.45	922	0.10	0.40	894	0.11	0.35	864	0.12
26	Carlton	124	0.65	950	0.07	0.55	820	0.09	0.50	783	0.10	0.50	783	0.10	0.40	690	0.13
24	Dadlington	119	0.75	1207	0.05	0.65	983	0.07	0.60	948	0.08	0.55	908	0.09	0.50	794	0.11
25	Osbaston	119	0.45	667	0.09	0.15	295	0.22	0.10	235	0.37	0.10	235	0.37	0.10	235	0.37
23	Norton Juxta Twycross	111	0.55	886	0.07	0.55	886	0.07	0.50	880	0.08	0.45	865	0.09	0.40	746	0.11
21	Barton in the Beans	103	0.65	1440	0.04	0.55	1266	0.05	0.55	1266	0.05	0.50	1212	0.06	0.40	1011	0.07
22	Kirby Muxloe	103	0.65	3265	0.02	0.55	2950	0.02	0.55	2950	0.02	0.55	2950	0.02	0.45	2833	0.03
20	Peckleton	95	0.55	649	0.08	0.50	595	0.09	0.45	548	0.11	0.30	416	0.15	0.20	326	0.22
19	Sibson	85	0.55	1219	0.04	0.45	1149	0.04	0.35	918	0.06	0.30	817	0.07	0.25	739	0.09
17	Cadeby	83	0.55	459	0.09	0.30	252	0.18	0.25	253	0.22	0.25	253	0.22	0.20	212	0.32
18	Ratcliffe Culey	83	0.60	1075	0.04	0.50	849	0.06	0.45	799	0.06	0.15	376	0.15	0.10	276	0.24
16	Orton-on-the-Hill	81	0.70	379	0.11	0.60	302	0.16	0.55	302	0.18	0.55	302	0.18	0.50	288	0.20
15	Shackerstone	68	0.70	1177	0.03	0.65	1080	0.04	0.55	890	0.05	0.55	890	0.05	0.45	785	0.06
14	Sutton Cheney	65	0.75	571	0.06	0.65	506	0.08	0.65	506	0.08	0.55	406	0.11	0.25	228	0.20
13	Shenton	53	0.50	123	0.24	0.45	103	0.29	0.40	95	0.38	0.40	95	0.38	0.35	88	0.51
12	Sheepy Parva	47	0.50	935	0.03	0.50	935	0.03	0.40	831	0.04	0.35	716	0.04	0.30	686	0.05
11	Upton	43	0.85	187	0.12	0.80	133	0.18	0.55	58	0.46	0.50	55	0.55	0.35	46	0.84
10	Odstone	30	0.80	779	0.02	0.65	471	0.04	0.50	329	0.06	0.45	297	0.07	0.20	156	0.15
8	Newtown Unthank	27	0.95	6648	0.00	0.95	6648	0.00	0.60	2355	0.01	0.20	2078	0.01	0.20	2078	0.01
9	Wellsborough	27	0.60	56	0.25	0.35	27	0.55	0.25	22	0.80	0.25	22	0.80	0.20	20	1.13
7	Wykin	25	0.60	386	0.03	0.50	346	0.04	0.50	346	0.04	0.20	159	0.11	0.20	159	0.11
5	Bilstone	20	0.70	831	0.01	0.55	762	0.01	0.40	772	0.02	0.40	772	0.02	0.35	727	0.02
6	Copt Oak	20	0.80	173	0.06	0.75	161	0.07	0.75	161	0.07	0.45	63	0.21	0.30	43	0.37
3	Atterton	15	0.95	3324	0.00	0.15	1870	0.00	0.15	1870	0.00	0.10	1143	0.01	0.10	1143	0.01
4	Pinwall	15	0.80	92	0.09	0.75	87	0.10	0.75	87	0.10	0.70	74	0.13	0.50	41	0.27
2	Ellistown	11	0.85	21	0.28	0.65	9	0.84	0.65	9	0.84	0.65	9	0.84	0.65	9	0.84
1	Appleby Magna	10	0.50	1247	0.00	0.50	1247	0.00	0.50	1247	0.00	0.10	970	0.01	0.10	970	0.01

Table 4.3: (continued) Results of the containment analysis for the POST data. For all considered threshold, the optimum α-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these α-cuts (points/km2).

ID	Village Name	No. of Points	Threshold of 75 %			Threshold of 80 %			Threshold of 85 %			Threshold of 90 %			Threshold of 95 %		
			α-cut	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*
49	Burbage	6635	0.45	2141	2.46	0.40	2051	2.77	0.40	2051	2.77	0.35	1950	3.09	0.25	1810	3.54
48	Earl Shilton	4763	0.40	2702	1.32	0.35	2610	1.48	0.30	2520	1.67	0.25	2361	1.87	0.20	2198	2.08
47	Barwell	4244	0.40	2742	1.20	0.35	2632	1.37	0.35	2632	1.37	0.30	2476	1.54	0.20	2192	1.86
46	Groby	2954	0.40	2070	1.14	0.40	2070	1.14	0.35	2016	1.26	0.30	1928	1.42	0.25	1816	1.56
45	Ratby	1904	0.50	2428	0.63	0.45	2374	0.69	0.45	2374	0.69	0.40	2295	0.75	0.30	2066	0.88
44	Desford	1497	0.35	1849	0.64	0.30	1793	0.67	0.20	1530	0.84	0.05	943	1.49	0.00	0	0.00
43	Newbold Verdon	1376	0.45	2312	0.46	0.40	2248	0.52	0.35	2109	0.58	0.30	2036	0.62	0.15	1648	0.80
42	Barlestone	1080	0.55	2614	0.32	0.45	2378	0.37	0.40	2262	0.41	0.25	1951	0.51	0.15	1670	0.62
41	Market Bosworth	1048	0.35	1768	0.46	0.30	1610	0.54	0.25	1470	0.62	0.15	1189	0.82	0.10	1015	0.98
40	Stoke Golding	793	0.50	1871	0.32	0.45	1790	0.36	0.35	1679	0.41	0.25	1528	0.48	0.10	1156	0.65
39	Bagworth	604	0.35	1354	0.38	0.35	1354	0.38	0.35	1354	0.38	0.30	1167	0.47	0.15	760	0.76
38	Thornton	460	0.50	1889	0.19	0.45	1716	0.22	0.40	1462	0.27	0.25	1034	0.40	0.05	539	0.82
37	Witherley	313	0.45	1538	0.16	0.40	1457	0.18	0.35	1387	0.20	0.30	1317	0.21	0.10	927	0.32
36	Sheepy Magna	302	0.30	796	0.29	0.20	580	0.43	0.10	405	0.65	0.05	290	0.96	0.00	0	0.00
35	Higham on the Hill	284	0.35	1495	0.15	0.30	1387	0.17	0.25	1289	0.19	0.05	661	0.41	0.00	0	0.00
34	Stanton under Bardon	281	0.55	1958	0.11	0.45	1698	0.14	0.35	1437	0.17	0.15	951	0.27	0.05	588	0.46
33	Fenny Drayton	222	0.45	1178	0.15	0.40	1112	0.16	0.30	1016	0.19	0.25	922	0.22	0.05	562	0.38
32	Nailstone	215	0.35	1183	0.14	0.30	1108	0.16	0.25	1059	0.18	0.05	479	0.43	0.00	0	0.00
31	Botcheston	198	0.25	722	0.23	0.25	722	0.23	0.20	594	0.30	0.20	594	0.30	0.10	356	0.54
30	Stapleton	195	0.20	1230	0.12	0.10	934	0.18	0.05	660	0.27	0.05	660	0.27	0.00	0	0.00
29	Kirkby Mallory	176	0.20	819	0.17	0.15	673	0.21	0.05	471	0.33	0.00	0	0.00	0.00	0	0.00
28	Congerstone	151	0.40	1184	0.10	0.35	991	0.13	0.30	972	0.13	0.25	856	0.16	0.10	572	0.26
27	Twycross	148	0.30	757	0.15	0.20	649	0.19	0.15	548	0.24	0.10	441	0.32	0.05	315	0.46
26	Carlton	124	0.30	570	0.17	0.20	436	0.24	0.20	436	0.24	0.10	323	0.36	0.00	0	0.00
24	Dadlington	119	0.45	779	0.12	0.35	654	0.15	0.15	408	0.26	0.05	247	0.44	0.00	0	0.00
25	Osbaston	119	0.05	115	0.96	0.05	115	0.96	0.05	115	0.96	0.05	115	0.96	0.00	0	0.00
23	Norton Juxta Twycross	111	0.35	688	0.13	0.35	688	0.13	0.25	577	0.17	0.25	577	0.17	0.10	373	0.28
21	Barton in the Beans	103	0.30	863	0.09	0.25	787	0.11	0.25	787	0.11	0.15	570	0.17	0.10	458	0.21
22	Kirby Muxloe	103	0.35	2315	0.03	0.25	1808	0.05	0.20	1662	0.05	0.05	935	0.11	0.05	935	0.11
20	Peckleton	95	0.20	326	0.22	0.15	269	0.29	0.10	202	0.45	0.10	202	0.45	0.05	128	0.72
19	Sibson	85	0.25	739	0.09	0.15	523	0.13	0.10	427	0.18	0.05	305	0.26	0.00	0	0.00
17	Cadeby	83	0.20	212	0.32	0.20	212	0.32	0.05	67	1.22	0.05	67	1.22	0.05	67	1.22
18	Ratcliffe Culey	83	0.10	276	0.24	0.05	184	0.42	0.05	184	0.42	0.05	184	0.42	0.00	0	0.00
16	Orton-on-the-Hill	81	0.45	267	0.24	0.25	181	0.36	0.10	122	0.57	0.05	80	0.98	0.05	80	0.98
15	Shackerstone	68	0.45	785	0.06	0.35	693	0.08	0.15	446	0.13	0.10	368	0.17	0.05	259	0.26
14	Sutton Cheney	65	0.20	225	0.24	0.20	225	0.24	0.10	136	0.43	0.10	136	0.43	0.05	88	0.74
13	Shenton	53	0.35	88	0.51	0.35	88	0.51	0.30	71	0.65	0.20	49	0.99	0.05	24	2.17
12	Sheepy Parva	47	0.15	449	0.09	0.15	449	0.09	0.15	449	0.09	0.10	366	0.12	0.05	232	0.20
11	Upton	43	0.35	46	0.84	0.35	46	0.84	0.35	46	0.84	0.35	46	0.84	0.05	10	4.50
10	Odstone	30	0.20	156	0.15	0.20	156	0.15	0.15	117	0.22	0.10	96	0.30	0.10	96	0.30
8	Newtown Unthank	27	0.10	1524	0.01	0.10	1524	0.01	0.05	799	0.03	0.05	799	0.03	0.00	0	0.00
9	Wellsborough	27	0.20	20	1.13	0.20	20	1.13	0.20	20	1.13	0.15	15	1.74	0.15	15	1.74
7	Wykin	25	0.15	143	0.15	0.15	143	0.15	0.10	105	0.24	0.10	105	0.24	0.10	105	0.24
5	Bilstone	20	0.20	392	0.04	0.20	392	0.04	0.15	329	0.06	0.15	329	0.06	0.15	329	0.06
6	Copt Oak	20	0.30	43	0.37	0.30	43	0.37	0.15	23	0.81	0.15	23	0.81	0.15	23	0.81
3	Atterton	15	0.05	772	0.02	0.05	772	0.02	0.05	772	0.02	0.00	0	0.00	0.00	0	0.00
4	Pinwall	15	0.45	42	0.31	0.45	42	0.31	0.45	42	0.31	0.40	37	0.38	0.35	35	0.43
2	Ellistown	11	0.55	6	1.50	0.55	6	1.50	0.45	4	2.32	0.45	4	2.32	0.40	4	2.80
1	Appleby Magna	10	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00

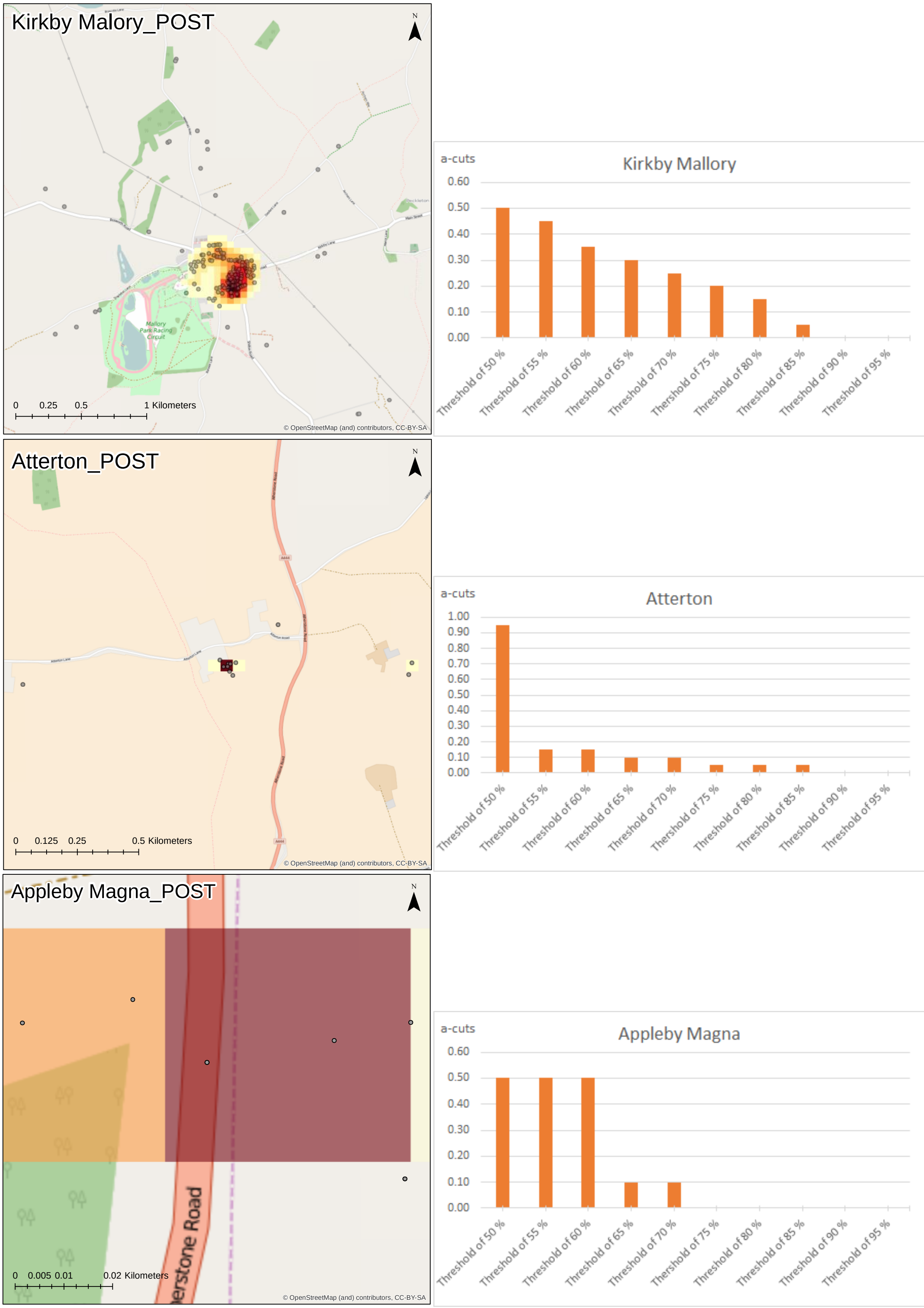


Figure 4.13: Map of three settlements (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold), where the smallest α-cuts drop out before reaching larger thresholds which indicate the scatter nature of their points.

- Another observation shows that villages with larger number of addresses and less difference between their α-cuts imply the compact character of their address points. This is applied in Burbage, Barwell, Groby, Ratby, and Earl Shilton. In addition, villages with few number of address points but still with less difference between their α-cuts, such as Pinwall and Ellistown, also highlight the compactness of their patterns. Figure 4.14 presents three mapping examples for Burbage, Earl Shilton and Pinwall with par charts showing which α-cut is satisfying particular threshold. Mostly the minimum value reached that capture 95% of these villages is about α-cut of 0.2.

Figure 4.14: Map of some settlement (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold), in which less disparity between α-cuts values showing that their addresses are compact.

- Some villages show a sudden decline in their α-cut values, suggesting that their point distributions contain some outliers located further or spread away from the main point groups. This appears in several settlements but is more distinct in one small village, Atterton, and a moderate-sized village, Osbaston (Figure 4.15).

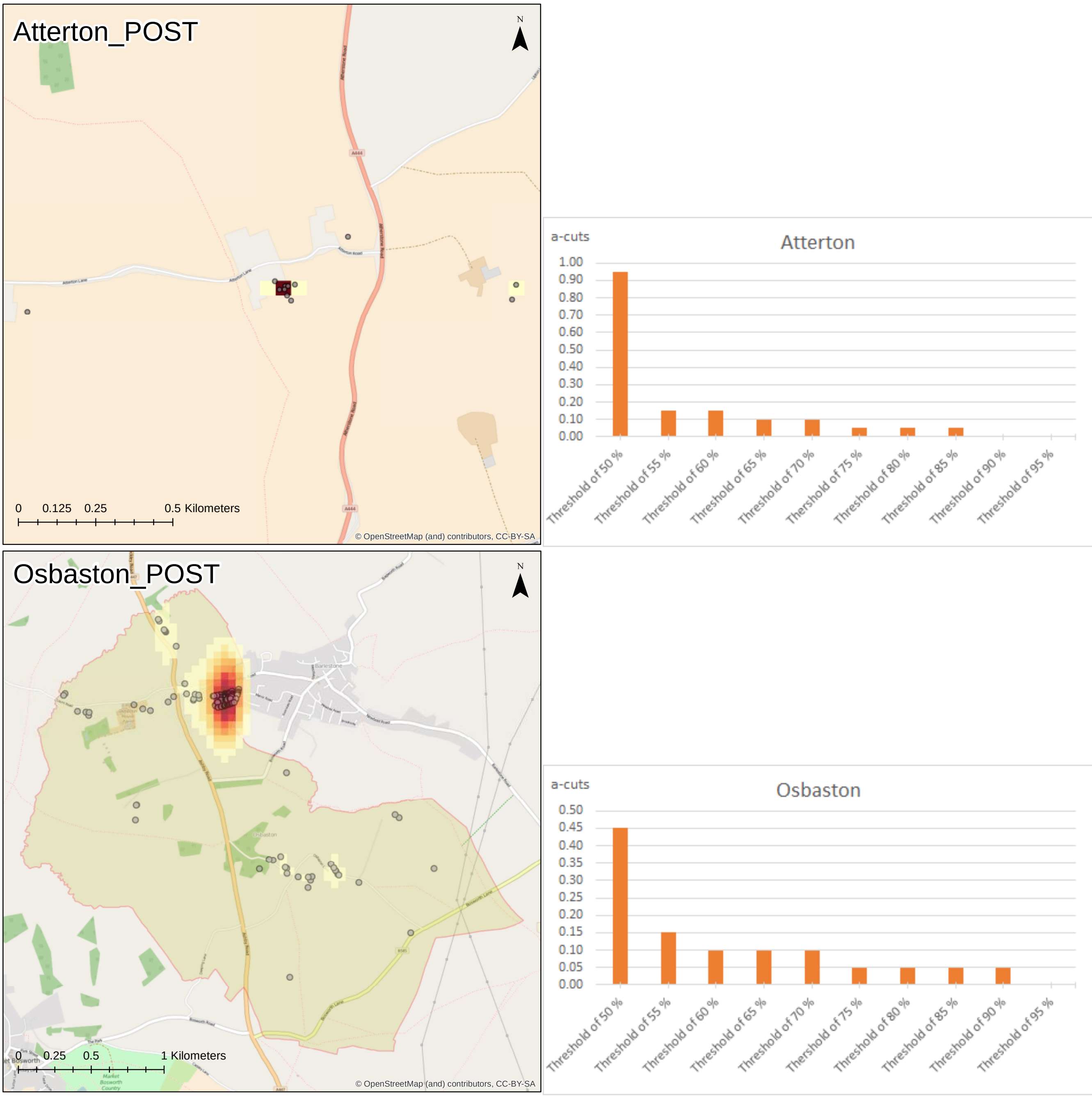


Figure 4.15: Map of some villages (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold) showing sudden decline in their α-cut values that relates to points located further and spread away from the main point groups.

- It should be further noted that some settlements conversely show a gradual decrease of their α-cuts, which signifies that the addresses in these villages are less defuse. Here are some examples: Peckleton, Sibson, Barlestone, Twycross, Congerstone, Norton Juxta Twycross, Barton in the Beans and Botcheston; two of these are mapped in Figure 4.16.

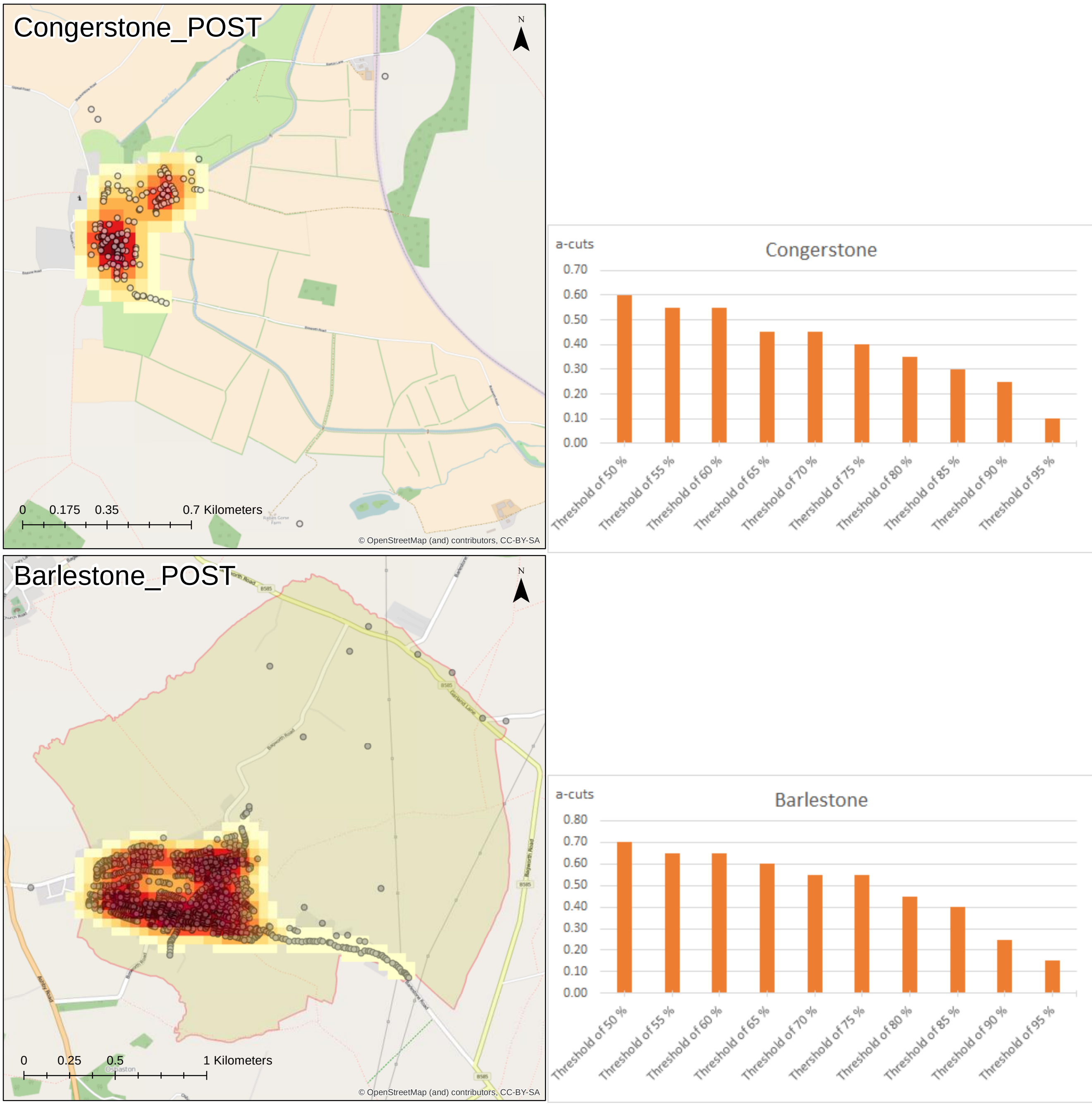


Figure 4.16: Examples of two settlements (on the left, maps showing their address points and fuzzy models, with bar charts on the right indicating the optimal α-cut for each threshold),as a case of less defuse points which have gradual transition between their α-cuts.

<u>Secondly, exploring relationship with point densities:</u>

- Looking at the density of points recorded in Table 4.3, it is apparent that generally there is no clearly discernible pattern between the density and the number of the point features. However, there are some flaws in these results, which appear in darker colour grades.
- Newton Unthank, for example, has the maximum density values in the 50 and 55% with only a few points (27). This is also true for Atterton and Appleby Magna, which have fewer points with quite large density.
- In addition, Kirby Muxloe seems to have larger values of density with 103 points than other settlements which have much more points, such as Burbage, Earl Shilton and even Barwell, which accommodate the maximum addresses.
- Considering the other direction of the table, density values decrease as the proportion of the address points included increase. This has been confirmed in the previous result of BS76 data (Table 4.2).

Results of the POI data

The result of the containment analysis for the POI data are highlighted in Table 4.4. Thirty-one settlements with at least 10 features are identified, and again these are ranked in descending order by the number of points. The same colour grades from the previous Tables, 4.2 & 4.3, for α-cuts (light orange) and densities (light green) are also applied here (Table 4.4). This table is quite revealing in several ways:

<u>Firstly, exploring the α-cut profiles:</u>

- Similar to the previous results from Tables 4.2 and 4.3, some villages fail to capture all considered thresholds as they drop before reaching the maximum threshold value. Desford, Earl Shilton, and Market Bosworth, for example, do not have any α-cut values that cover 95% of their address points. This indicates that in these villages there are some addresses that are spread out (referring to areas in which they have practically a very small value of membership, a non-zero number less than 0.1). Therefore, these are excluded from the fuzzy model, as can be noticed from the maps and the bar charts of Desford and Earl Shilton in Figure 4.17. It is important to notice that firstly in Desford the fuzzy surfaces are not continuous since the addresses have different intensities across the area. Secondly, there is a certain area (north-west of Desford) that appears beyond the parish of Desford although it is usually considered as being part of Desford. Thirdly, the points in Earl Shilton look to be slightly dispersed, which may explain the low values of membership for all thresholds. This can be also realised in Newbold Verdon, which has a maximum membership value of 0.2 to include 50% of the points, which means the majority of them are less intense.

Table 4.4: Results of the containment analysis for the POI data. For all considered threshold, the optimum α-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these α-cuts (points/km2).

ID	Village Name	No. of Points	Threshold of 50 %			Threshold of 55 %			Threshold of 60 %			Threshold of 65 %			Threshold of 70 %		
			α-cut	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*
5	Burbage	257	0.30	105	1.64	0.30	105	1.64	0.30	105	1.64	0.30	105	1.64	0.25	84	2.20
8	Earl Shilton	227	0.20	475	0.25	0.15	333	0.43	0.15	333	0.43	0.10	253	0.70	0.10	253	0.70
3	Barwell	221	0.55	372	0.30	0.45	345	0.40	0.45	345	0.40	0.35	284	0.51	0.25	233	0.73
15	Markfield	173	0.75	156	0.66	0.75	156	0.66	0.70	123	0.86	0.60	89	1.30	0.50	69	1.82
10	Groby	113	0.25	159	0.37	0.20	131	0.54	0.20	131	0.54	0.10	83	1.09	0.10	83	1.09
14	Market Bosworth	106	0.30	506	0.11	0.20	385	0.16	0.15	314	0.24	0.15	314	0.24	0.10	208	0.40
7	Desford	95	0.45	93	0.58	0.45	93	0.58	0.40	85	0.71	0.35	77	0.85	0.30	69	1.03
17	Newbold Verdon	70	0.20	142	0.30	0.20	142	0.30	0.20	142	0.30	0.15	121	0.40	0.10	97	0.64
22	Ratby	60	0.40	94	0.33	0.35	91	0.44	0.35	91	0.44	0.35	91	0.44	0.30	88	0.54
28	Stoke Golding	59	0.60	148	0.24	0.60	148	0.24	0.55	136	0.26	0.40	108	0.38	0.35	101	0.44
2	Barlestone	30	0.35	231	0.06	0.30	208	0.08	0.25	187	0.10	0.20	162	0.13	0.20	162	0.13
30	Thornton	27	0.60	80	0.18	0.55	70	0.21	0.45	54	0.31	0.35	48	0.45	0.35	48	0.45
27	Stapleton	26	0.55	23	0.56	0.50	25	0.73	0.50	25	0.73	0.50	25	0.73	0.40	17	1.19
1	Bagworth	24	0.60	40	0.30	0.45	29	0.55	0.45	29	0.55	0.45	29	0.55	0.40	27	0.71
19	Nuneaton	23	0.50	1	10.98	0.50	1	10.98	0.40	1	18.87	0.35	1	23.88	0.30	1	33.42
26	Stanton under Bardon	23	0.65	95	0.14	0.65	95	0.14	0.60	84	0.17	0.45	58	0.26	0.35	47	0.36
13	Kirkby Mallory	19	0.40	122	0.08	0.35	128	0.09	0.35	128	0.09	0.30	119	0.12	0.30	119	0.12
21	Peckleton	18	0.90	312	0.03	0.30	88	0.16	0.30	88	0.16	0.30	88	0.16	0.30	88	0.16
11	Higham on the Hill	15	0.45	151	0.05	0.35	126	0.08	0.35	126	0.08	0.35	126	0.08	0.30	120	0.09
9	Fenny Drayton	14	0.40	40	0.30	0.40	40	0.30	0.40	40	0.30	0.40	40	0.30	0.40	40	0.30
24	Sheepy Magna	14	0.85	75	0.09	0.55	22	0.41	0.55	22	0.41	0.50	20	0.55	0.50	20	0.55
29	Sutton Cheney	14	0.60	12	0.58	0.55	12	0.77	0.55	12	0.77	0.45	8	1.46	0.45	8	1.46
31	Twycross	14	0.75	18	0.38	0.50	9	0.92	0.45	10	1.15	0.45	10	1.15	0.45	10	1.15
16	Nailstone	13	0.80	121	0.06	0.70	90	0.09	0.70	90	0.09	0.55	61	0.15	0.45	51	0.19
25	Sibson	13	0.60	59	0.13	0.60	59	0.13	0.60	59	0.13	0.50	44	0.20	0.40	38	0.26
32	Witherley	13	0.50	107	0.07	0.50	107	0.07	0.50	107	0.07	0.40	85	0.11	0.35	78	0.13
23	Ratcliffe Culey	12	0.70	27	0.26	0.70	27	0.26	0.65	28	0.36	0.65	28	0.36	0.65	28	0.36
6	Dadlington	11	0.85	104	0.06	0.50	29	0.24	0.50	29	0.24	0.20	13	0.60	0.20	13	0.60
4	Botcheston	10	0.80	32	0.22	0.80	32	0.22	0.80	32	0.22	0.80	32	0.22	0.80	32	0.22
18	Newtown Unthank	10	0.95	2909	0.00	0.95	2909	0.00	0.95	2909	0.00	0.95	2909	0.00	0.95	2909	0.00
20	Osbaston	10	0.60	64	0.09	0.60	64	0.09	0.60	64	0.09	0.45	40	0.18	0.45	40	0.18

Table 4.4: (continued) Results of the containment analysis for the POI data. For all considered threshold, the optimum α-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these α-cuts (points/km2).

ID	Village Name	No. of Points	Threshold of 75 %			Threshold of 80 %			Threshold of 85 %			Threshold of 90 %			Threshold of 95 %		
			α-cut	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*
5	Burbage	257	0.20	73	2.96	0.20	73	2.96	0.15	65	3.65	0.15	65	3.65	0.10	55	4.43
8	Earl Shilton	227	0.10	253	0.70	0.05	165	1.25	0.05	165	1.25	0.05	165	1.25	0.00	0	0.00
3	Barwell	221	0.25	233	0.73	0.15	165	1.12	0.10	135	1.50	0.10	135	1.50	0.05	96	2.22
15	Markfield	173	0.35	48	2.94	0.35	48	2.94	0.30	43	3.48	0.10	25	6.45	0.05	18	9.15
10	Groby	113	0.10	83	1.09	0.10	83	1.09	0.05	52	2.09	0.05	52	2.09	0.05	52	2.09
14	Market Bosworth	106	0.10	208	0.40	0.05	123	0.79	0.05	123	0.79	0.05	123	0.79	0.00	0	0.00
7	Desford	95	0.25	58	1.24	0.10	31	2.55	0.05	21	4.12	0.05	21	4.12	0.00	0	0.00
17	Newbold Verdon	70	0.10	97	0.64	0.10	97	0.64	0.10	97	0.64	0.05	63	1.10	0.05	63	1.10
22	Ratby	60	0.30	88	0.54	0.25	74	0.67	0.20	67	0.79	0.15	60	0.92	0.05	38	1.58
28	Stoke Golding	59	0.30	87	0.52	0.20	73	0.67	0.15	66	0.81	0.10	52	1.05	0.05	38	1.56
2	Barlestone	30	0.15	123	0.19	0.15	123	0.19	0.10	103	0.29	0.10	103	0.29	0.10	103	0.29
30	Thornton	27	0.35	48	0.45	0.35	48	0.45	0.30	44	0.52	0.10	25	1.09	0.10	25	1.09
27	Stapleton	26	0.40	17	1.19	0.35	14	1.49	0.25	10	2.40	0.20	8	2.88	0.10	6	4.16
1	Bagworth	24	0.40	27	0.71	0.35	25	0.85	0.35	25	0.85	0.15	13	1.71	0.15	13	1.71
19	Nuneaton	23	0.25	0	43.32	0.20	0	53.88	0.20	0	53.88	0.20	0	53.88	0.15	0	66.07
26	Stanton under Bardon	23	0.30	39	0.46	0.25	34	0.58	0.25	34	0.58	0.20	30	0.71	0.10	21	1.10
13	Kirkby Mallory	19	0.25	111	0.14	0.25	111	0.14	0.15	74	0.26	0.15	74	0.26	0.15	74	0.26
21	Peckleton	18	0.30	88	0.16	0.15	51	0.33	0.15	51	0.33	0.15	51	0.33	0.10	38	0.47
11	Higham on the Hill	15	0.20	90	0.14	0.20	90	0.14	0.20	90	0.14	0.15	77	0.19	0.15	77	0.19
9	Fenny Drayton	14	0.40	40	0.30	0.40	40	0.30	0.40	40	0.30	0.20	15	0.96	0.20	15	0.96
24	Sheepy Magna	14	0.50	20	0.55	0.10	5	2.74	0.10	5	2.74	0.10	5	2.74	0.10	5	2.74
29	Sutton Cheney	14	0.45	8	1.46	0.45	8	1.46	0.45	8	1.46	0.35	6	2.44	0.35	6	2.44
31	Twycross	14	0.45	10	1.15	0.20	3	3.62	0.20	3	3.62	0.15	3	4.68	0.15	3	4.68
16	Nailstone	13	0.45	51	0.19	0.15	25	0.52	0.15	25	0.52	0.15	25	0.52	0.15	25	0.52
25	Sibson	13	0.40	38	0.26	0.20	26	0.49	0.20	26	0.49	0.20	26	0.49	0.20	26	0.49
32	Witherley	13	0.35	78	0.13	0.20	62	0.21	0.20	62	0.21	0.20	62	0.21	0.20	62	0.21
23	Ratcliffe Culey	12	0.65	28	0.36	0.65	28	0.36	0.20	8	1.52	0.20	8	1.52	0.20	8	1.52
6	Dadlington	11	0.15	11	0.96	0.15	11	0.96	0.15	11	0.96	0.15	11	0.96	0.15	11	0.96
4	Botcheston	10	0.75	26	0.31	0.75	26	0.31	0.65	17	0.52	0.65	17	0.52	0.50	11	0.91
18	Newtown Unthank	10	0.70	1870	0.00	0.70	1870	0.00	0.70	1870	0.00	0.70	1870	0.00	0.15	831	0.01
20	Osbaston	10	0.35	31	0.26	0.35	31	0.26	0.30	34	0.30	0.30	34	0.30	0.30	34	0.30

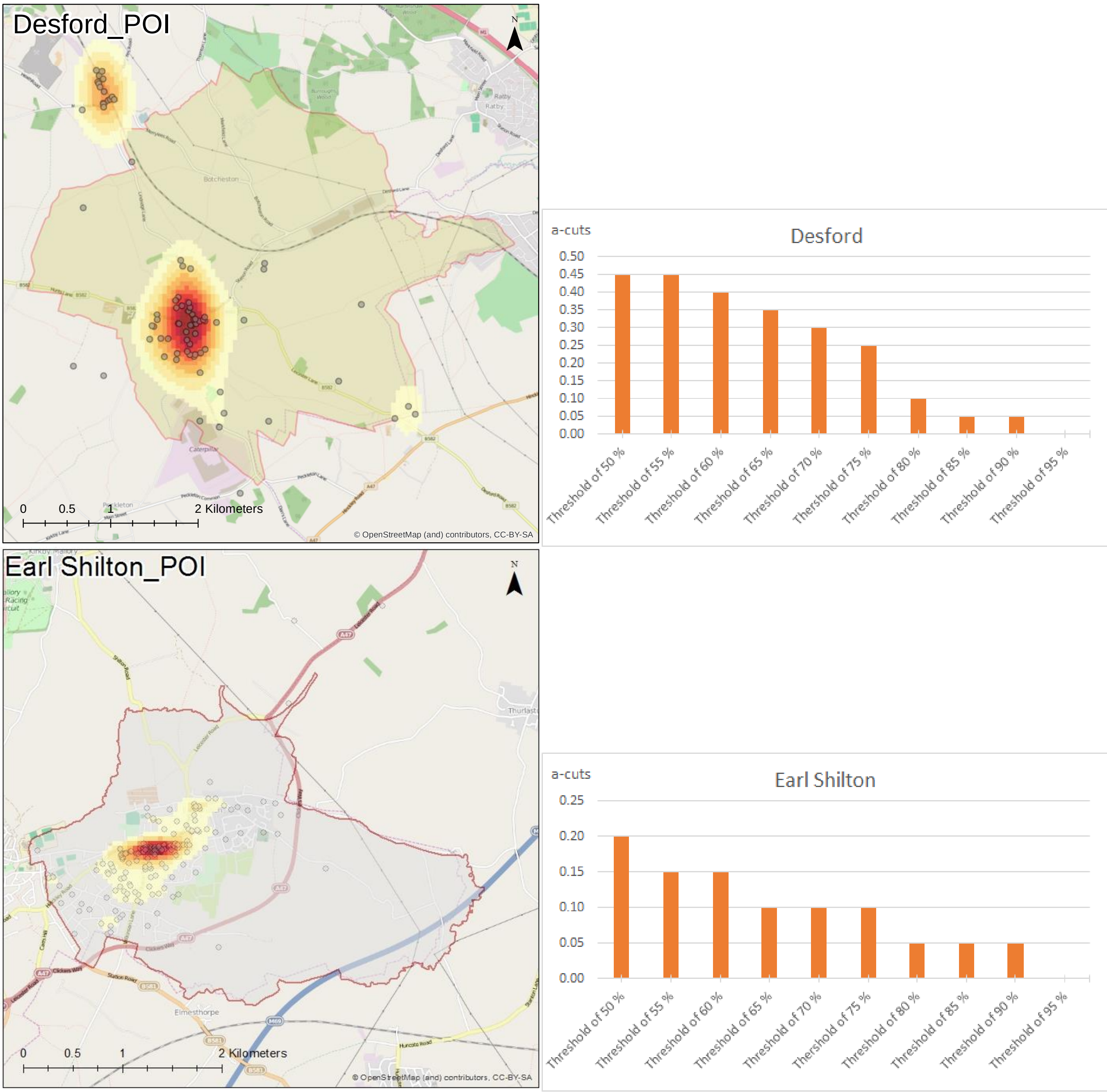


Figure 4.17: Two examples of villages (maps on the left showing their address points and fuzzy models with bar charts on the right indicating the optimal α-cut for each threshold) that fail to capture 95% of their address points indicating that some of their points are less intense.

- There are some settlements that rather have slightly higher value of membership across the different thresholds, which imply the general structure of their points being somehow compact. These involve Botecheston, Newtown Unthank, Osbaston and Ratcliffe Culey. However, it turns out that these villages accommodate few addresses distributed considerably in small area (just about a Kilometre or less) which may affect the results. Figure 4.18 presents maps and plots of α-cuts for Botcheston and Newtown Unthank.

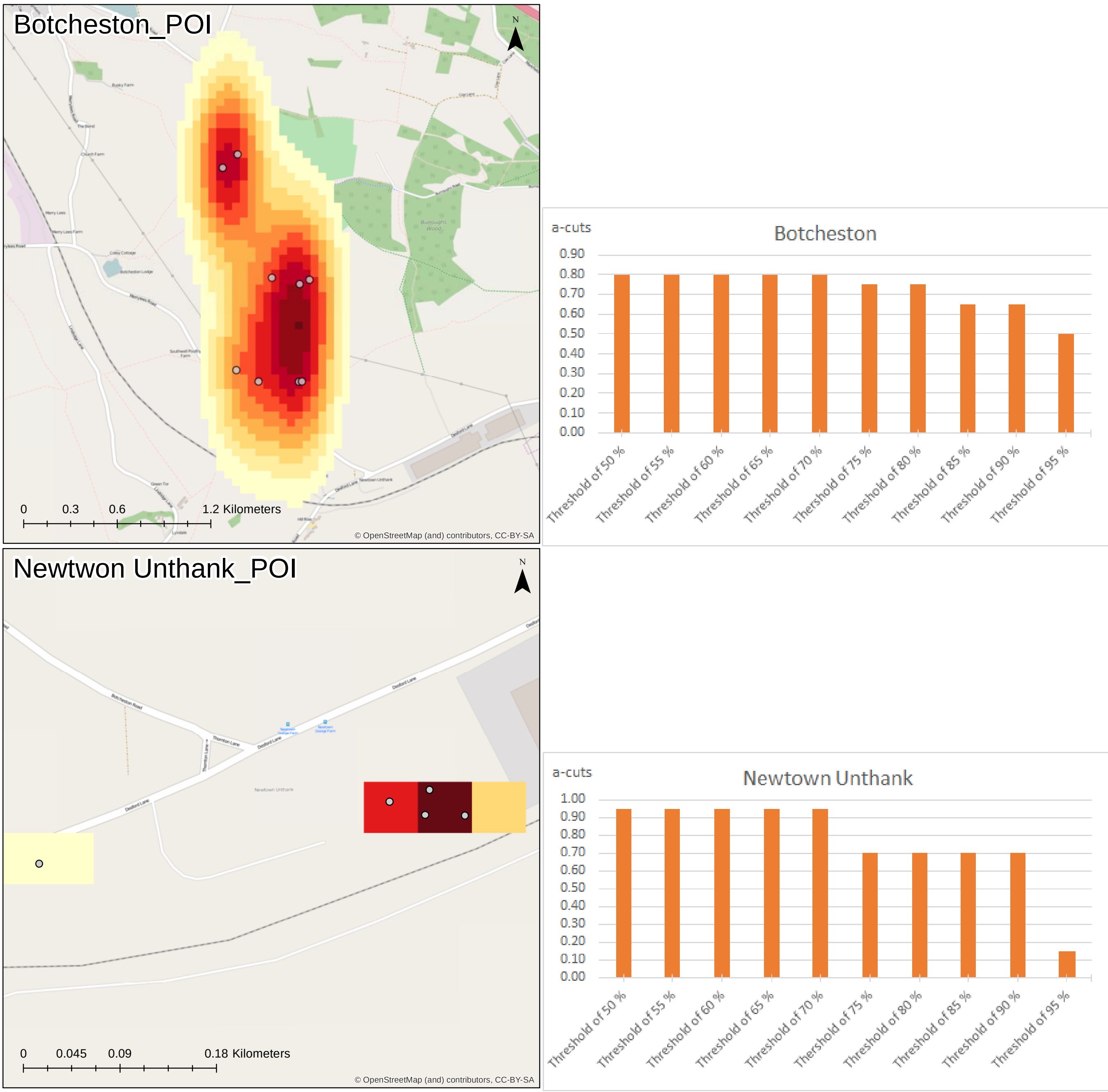


Figure 4.18: Two examples of small villages (maps on the left showing their address points and fuzzy models, with bar charts on the right indicating the optimal α-cut for each threshold) that have slightly higher value of membership across the different thresholds.

- There is a gradual decrease in the membership values for some settlements; Barwell, Stoke Golding and Ratby are examples in this case. Maps and plots of α-cuts for Barwell and Stoke Golding are shown in Figure 4.19. This indicates that the points are more clustered in the inner region of the settlements and more disperse in the outer region. It should also emphasise the fact that the distribution of points in Stoke Golding does not coincide with the equivalent parish name, as some of the points fall nearly outside the parished area, although even so they seem to have a higher membership grade.

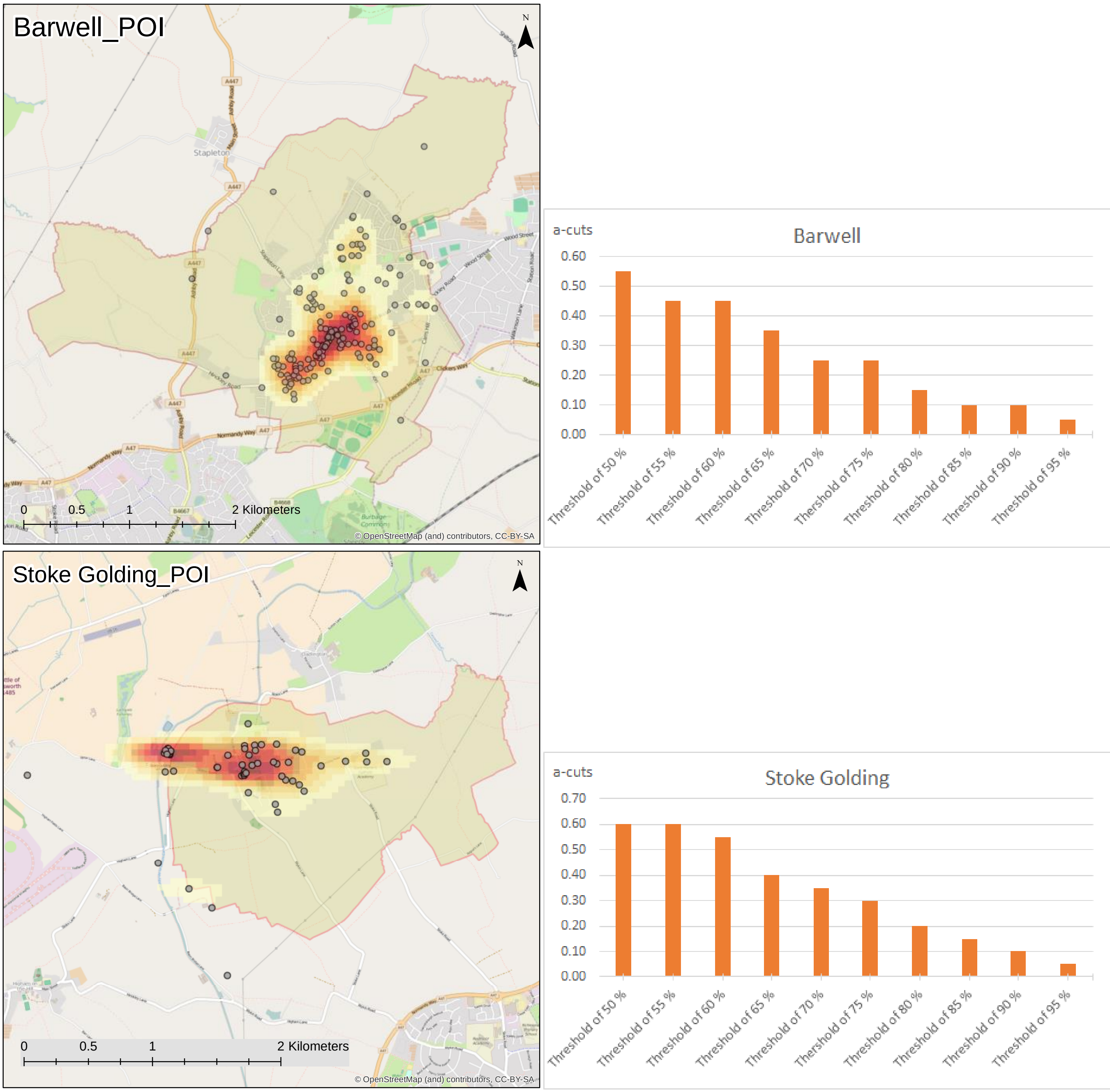


Figure 4.19: Two examples of villages (maps on the left showing their address points and fuzzy models with bar charts on the right indicating the optimal α-cut for each threshold) that have gradual decrease in the membership values which show how their points vary in intensity.,

- In a similar vein, there is a sudden reduction in the membership grades (α-cut values) in some settlements, showing that their point distributions have some outliers a bit further from the main cluster. This can be seen in some of the small-sized villages such as Peckleton, Twycross, Dadlington and Sheepy Magana. Two of those (Peckleton and Dadlington) are displayed with their maps and plots of α-cuts in Figure 4.20.

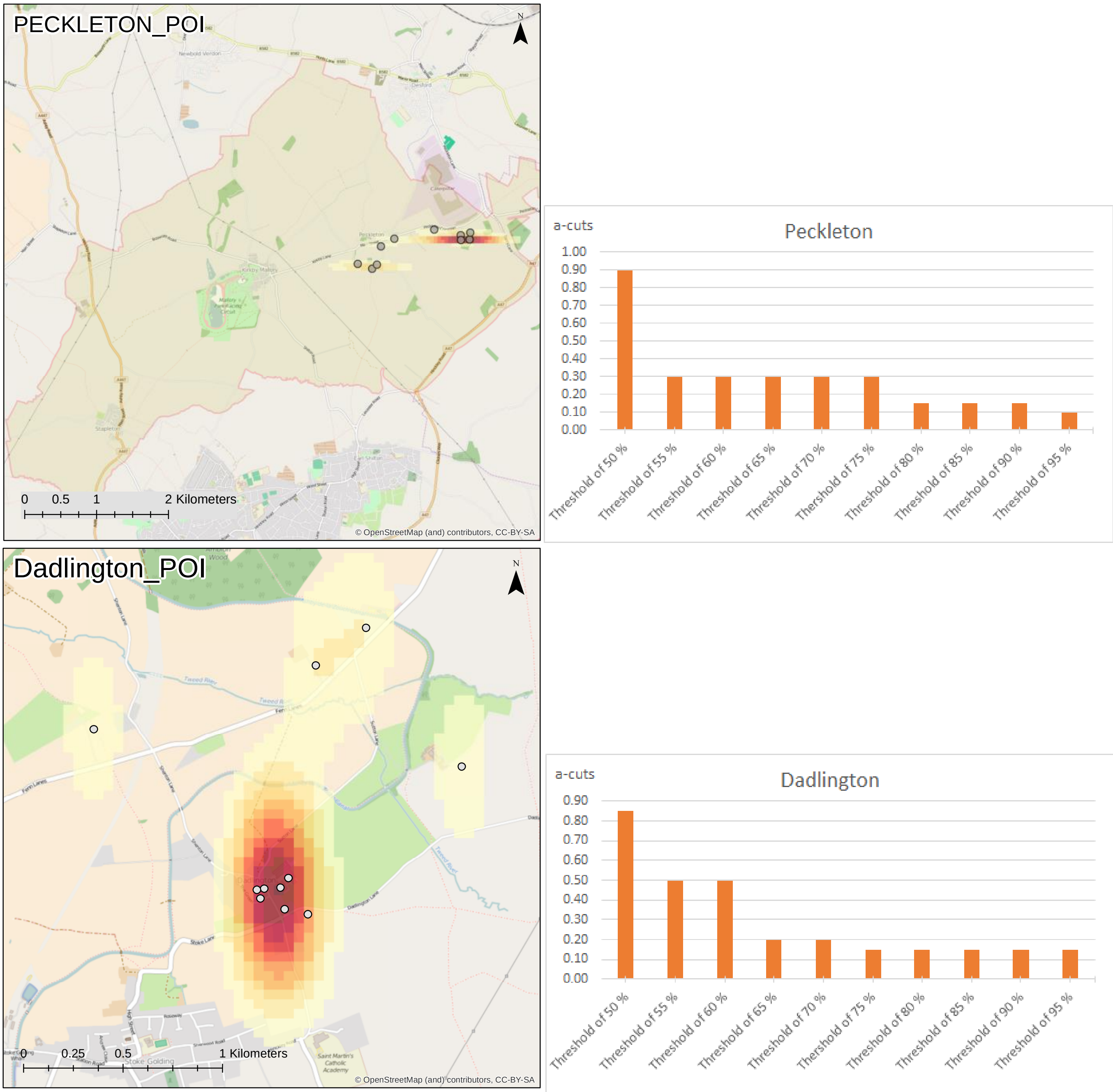


Figure 4.20: Maps of two settlements (on the left showing their address points and fuzzy models) with bar charts (on the right indicating the optimal α-cut for each threshold), which show sudden reduction in the membership grades related to the distribution of their point pattern.

- Other observations found that, in at least the first four examined thresholds (50, 55, 60 and 65%), some villages have a plateau in their α-cuts before declining to their minimums. This is shown in Fenny Drayton, Burbage, Newtown Unthank and Botcheston (the latter two are presented in Figure 4.18). A possible explanation for this might be due to the distribution of their points, which is consistent across these (close) thresholds.

Secondly, exploring the relationship with point densities:

- There is no clear structure for the association between number of points and their density recorded in Table 4.4. There is, however, one striking fact: that Newtown Unthank, with not many points distributed in a tiny area, has the maximum density across all thresholds. Indeed, it has very high density values compared to other settlements in the table.
- Excluding Newtown Unthank from the table, it is found that Market Bosworth, Earl Shilton and Barwell have the top values of density in all thresholds, while Sutton Cheney and Nuneaton have the lowest density across all thresholds. It has to be noted, however, that the order of point density differs in each threshold as the pattern structures for the points are different between settlements. This is illustrated by looking at thresholds of 50 & 55%: Market Bosworth, Earl Shilton, Barwell and Peckleton have the highest density (506, 475, 372 & 312 points/km^2) in the first instance (50%); whereas in the second (55%) the order is different: Barwell (345 points/km^2) is denser than Earl Shilton (333 points/km^2) and Peckleton (88 points/km^2) is replaced by Barlestone (208 points/ km^2). This can be expected to happen in any other instances.
- In the same way as before, the density of points declines horizontally across thresholds. For a particular village, density values decrease as the threshold limits increase.

Results of the all data combined

Table 4.5 presents the results of the containment analysis of settlements that are identified regardless of what their data sources are. This means that for a particular village all the address points which share the same name from any of the three data sources are considered as one single village. Thus, there are 58 villages with more than 10 addresses recognised in this way, and they are ranked in descending order as well. The table has the same format for membership grades and points density as presented in the results from Tables 4.2, 4.3 and 4.4. A number of issues have emerged from Table 4.5, as follows:

Firstly, exploring the α-cut profiles:

Several settlements miss out some thresholds towards the end from the analysis. This is to a large extent due to the distribution of their address points, which are mainly spread or scattered around the village area; leaving out areas with very small grades of membership ($0 < \alpha < 0.1$). This can be seen in Appleby Magna, Higham on the Hill, Twycross, Sheepy and Stanton under Bardon, which drop out at 70, 75 & 80% respectively. Some of these settlements (Higham on the Hill and Stanton under Bardon) contain a single cluster in the middle with a large number of points spread further apart. Also, Appleby Magna has one group of points with an outlier digressing from them. Alternatively, Twycross and Sheepy show more than one cluster with some dispersed points around the rest. Figure 4.21 illustrates these observations in the cases of Stanton under Bardon, Appleby Magna and Sheepy.

Table 4.5: Results of the containment analysis for all data combined. For all considered threshold, the optimum α-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these α-cuts (points/km2).

ID	Village Name	No. of Points	Threshold of 50 %			Threshold of 55 %			Threshold of 60 %			Threshold of65 %			Threshold of 70 %			Threshold of 75 %			Threshold of 80 %			Threshold of 85 %			Threshold of 90 %			Threshold of 95 %		
			α-cut	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*
10	Burbage	13871	0.60	5169	1.44	0.55	4919	1.76	0.55	4919	1.76	0.50	4708	2.06	0.45	4440	2.42	0.45	4440	2.42	0.40	4288	2.73	0.35	4061	3.06	0.30	3902	3.31	0.25	3752	3.52
20	Earl Shilton	9874	0.50	6494	0.78	0.45	6075	1.01	0.45	6075	1.01	0.40	5799	1.22	0.40	5799	1.22	0.35	5537	1.41	0.30	5321	1.61	0.30	5321	1.61	0.25	5057	1.78	0.20	4731	1.99
7	Barwell	8847	0.45	6290	0.85	0.45	6290	0.85	0.45	6290	0.85	0.40	5941	1.06	0.40	5941	1.06	0.35	5620	1.26	0.35	5620	1.26	0.30	5352	1.44	0.25	4999	1.61	0.15	4385	1.94
23	Groby	6349	0.55	4996	0.64	0.50	4800	0.81	0.50	4800	0.81	0.45	4548	0.95	0.40	4310	1.12	0.40	4310	1.12	0.35	4177	1.25	0.30	4023	1.39	0.25	3816	1.51	0.10	2860	2.15
14	Clarendon	5702	0.50	3237	0.89	0.45	3056	1.15	0.45	3056	1.15	0.40	2856	1.45	0.40	2856	1.45	0.35	2708	1.73	0.35	2708	1.73	0.30	2538	1.97	0.25	2375	2.20	0.10	1736	3.16
18	De Montfort	4532	0.50	2829	0.82	0.45	2658	1.06	0.45	2658	1.06	0.40	2516	1.26	0.40	2516	1.26	0.35	2390	1.50	0.30	2162	1.83	0.30	2162	1.83	0.25	2003	2.09	0.20	1840	2.35
40	Ratby	3974	0.65	5426	0.42	0.65	5426	0.42	0.60	5359	0.49	0.60	5359	0.49	0.55	5111	0.58	0.50	4930	0.64	0.50	4930	0.64	0.45	4853	0.70	0.35	4437	0.82	0.20	3785	1.00
19	Desford	3456	0.55	4356	0.40	0.50	4139	0.48	0.45	4012	0.54	0.40	3938	0.59	0.30	3676	0.67	0.20	3131	0.86	0.15	2851	0.98	0.10	2390	1.24	0.05	1685	1.87	0.00	0	0.00
53	Trinity	2988	0.65	3085	0.49	0.60	2821	0.68	0.60	2821	0.68	0.55	2734	0.82	0.55	2734	0.82	0.50	2665	0.92	0.50	2665	0.92	0.45	2589	0.98	0.35	2396	1.16	0.30	2319	1.23
31	Newbold Verdon	2875	0.55	4998	0.33	0.55	4998	0.33	0.50	4918	0.39	0.50	4918	0.39	0.45	4883	0.45	0.45	4883	0.45	0.40	4731	0.50	0.35	4602	0.54	0.25	4087	0.64	0.10	3134	0.88
29	Markfield	2464	0.55	2633	0.50	0.50	2547	0.58	0.45	2482	0.65	0.45	2482	0.65	0.40	2385	0.72	0.35	2322	0.81	0.25	2046	1.02	0.20	1854	1.17	0.10	1387	1.62	0.00	0	0.00
28	Market Bosworth	2294	0.55	4802	0.25	0.50	4536	0.29	0.45	4302	0.35	0.45	4302	0.35	0.40	4077	0.41	0.30	3531	0.53	0.30	3531	0.53	0.20	2903	0.69	0.10	2236	0.94	0.00	0	0.00
5	Barlestone	2228	0.70	6052	0.19	0.65	5791	0.22	0.60	5460	0.27	0.60	5460	0.27	0.55	5392	0.32	0.55	5392	0.32	0.45	4984	0.36	0.40	4672	0.41	0.25	4070	0.49	0.15	3524	0.61
13	Castle	2076	0.50	4384	0.24	0.45	3955	0.33	0.45	3955	0.33	0.40	3543	0.44	0.40	3543	0.44	0.35	3267	0.52	0.35	3267	0.52	0.30	2990	0.61	0.25	2791	0.70	0.20	2577	0.79
50	Stoke Golding	1649	0.70	4510	0.20	0.65	4310	0.23	0.60	4132	0.26	0.60	4132	0.26	0.55	4045	0.30	0.50	3927	0.32	0.45	3748	0.35	0.35	3516	0.40	0.25	3236	0.47	0.05	2112	0.75
4	Bagworth & Thornton	1282	0.60	1009	0.65	0.55	901	0.82	0.50	816	0.99	0.45	728	1.26	0.45	728	1.26	0.40	630	1.56	0.30	474	2.20	0.15	303	3.68	0.05	172	6.80	0.00	0	0.00
57	Witherley	1093	0.50	1493	0.38	0.35	1115	0.58	0.30	1009	0.72	0.30	1009	0.72	0.25	863	0.90	0.15	598	1.48	0.15	598	1.48	0.05	333	2.81	0.00	0	0.00	0.00	0	0.00
24	Higham on the Hill	782	0.45	3471	0.12	0.40	3159	0.14	0.35	2952	0.16	0.25	2359	0.22	0.10	1567	0.36	0.05	1150	0.52	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00
38	Peckleton	739	0.75	563	0.74	0.75	563	0.74	0.70	436	1.03	0.55	230	2.12	0.45	165	3.14	0.35	133	4.29	0.15	75	8.01	0.10	62	10.08	0.05	39	17.83	0.05	39	17.83
43	Sheepy	692	0.30	243	1.53	0.25	195	1.98	0.10	98	4.69	0.10	98	4.69	0.05	54	10.56	0.05	54	10.56	0.05	54	10.56	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00
48	Stanton under Bardon	670	0.70	4649	0.08	0.65	4644	0.08	0.60	4205	0.10	0.55	4184	0.11	0.45	3680	0.13	0.35	3092	0.16	0.05	1327	0.42	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00
54	Twycross	670	0.35	400	0.85	0.30	348	1.10	0.20	228	1.93	0.20	228	1.93	0.15	176	2.72	0.10	133	3.98	0.05	84	6.68	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00
3	Bagworth	628	0.45	1667	0.22	0.45	1667	0.22	0.40	1542	0.30	0.40	1542	0.30	0.40	1542	0.30	0.35	1387	0.39	0.35	1387	0.39	0.35	1387	0.39	0.30	1204	0.47	0.15	780	0.77
42	Shackerstone	542	0.65	461	0.60	0.60	424	0.78	0.60	424	0.78	0.55	354	1.02	0.50	313	1.27	0.45	271	1.50	0.20	118	3.68	0.15	98	4.73	0.05	50	10.27	0.00	0	0.00
30	Nailstone	488	0.55	3174	0.09	0.55	3174	0.09	0.50	2948	0.10	0.40	2540	0.13	0.35	2486	0.14	0.30	2311	0.16	0.20	1991	0.20	0.05	1014	0.43	0.00	0	0.00	0.00	0	0.00
52	Thornton	487	0.65	2171	0.12	0.60	2207	0.14	0.60	2207	0.14	0.55	2035	0.17	0.50	1826	0.21	0.50	1826	0.21	0.45	1699	0.24	0.40	1461	0.29	0.25	1054	0.42	0.05	560	0.84
51	Sutton Cheney	418	0.65	254	0.85	0.35	101	2.40	0.30	88	3.05	0.25	73	3.87	0.20	58	5.05	0.15	45	7.44	0.10	35	10.93	0.10	35	10.93	0.10	35	10.93	0.05	26	15.91
44	Sheepy Magna	316	0.55	1254	0.13	0.45	1044	0.18	0.40	981	0.21	0.35	878	0.26	0.35	878	0.26	0.30	818	0.30	0.25	715	0.35	0.10	423	0.65	0.05	305	0.94	0.00	0	0.00
37	Osbaston	290	0.30	548	0.26	0.15	325	0.61	0.15	325	0.61	0.15	325	0.61	0.10	224	0.99	0.10	224	0.99	0.05	134	1.97	0.05	134	1.97	0.05	134	1.97	0.00	0	0.00
12	Carlton	271	0.55	1903	0.07	0.50	1729	0.09	0.40	1610	0.12	0.40	1610	0.12	0.30	1335	0.15	0.25	1189	0.17	0.20	1044	0.21	0.15	884	0.26	0.05	544	0.45	0.00	0	0.00

Table 4.5: **(continued)** Results of the containment analysis for all data combined. For all considered threshold, the optimum α-cuts that satisfy the threshold limits are identified, with the measure of the density of points fall within these α-cuts (points/km2).

ID	Village Name	No. of Points	Threshold of 50 %			Threshold of 55 %			Threshold of 60 %			Threshold of65 %			Threshold of 70 %			Threshold of 75 %			Threshold of 80 %			Threshold of 85 %			Threshold of 90 %			Threshold of 95 %		
			α-cut	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*	*α-cut*	*Density*	*Area*
22	Fenny Drayton	236	0.65	1528	0.08	0.60	1415	0.10	0.60	1415	0.10	0.50	1239	0.13	0.50	1239	0.13	0.45	1182	0.15	0.35	1069	0.18	0.30	1029	0.20	0.20	859	0.25	0.05	508	0.45
11	Cadeby	225	0.30	615	0.19	0.25	596	0.22	0.20	473	0.32	0.20	473	0.32	0.15	359	0.44	0.10	256	0.71	0.10	256	0.71	0.05	151	1.36	0.05	151	1.36	0.00	0	0.00
49	Stapleton	221	0.45	1862	0.06	0.40	1650	0.08	0.35	1545	0.09	0.35	1545	0.09	0.20	1223	0.13	0.15	1063	0.16	0.10	879	0.21	0.05	612	0.35	0.05	612	0.35	0.05	612	0.35
9	Botcheston	208	0.40	1235	0.08	0.35	993	0.14	0.35	993	0.14	0.35	993	0.14	0.30	805	0.19	0.25	695	0.25	0.25	695	0.25	0.20	577	0.32	0.15	473	0.42	0.15	473	0.42
27	Kirkby Mallory	195	0.45	1354	0.07	0.40	1183	0.09	0.35	1052	0.11	0.30	1066	0.13	0.25	831	0.18	0.20	756	0.20	0.10	561	0.28	0.05	439	0.39	0.00	0	0.00	0.00	0	0.00
15	Congerstone	160	0.60	1837	0.05	0.55	1546	0.06	0.50	1399	0.07	0.45	1258	0.09	0.45	1258	0.09	0.40	1160	0.10	0.35	1043	0.13	0.30	996	0.14	0.25	879	0.17	0.15	687	0.22
17	Dadlington	130	0.70	1316	0.06	0.70	1316	0.06	0.65	1080	0.07	0.60	1016	0.09	0.55	945	0.10	0.40	791	0.13	0.25	559	0.19	0.15	466	0.24	0.10	386	0.31	0.00	0	0.00
33	Norton Juxta Twycross	117	0.50	890	0.08	0.50	890	0.08	0.50	890	0.08	0.45	876	0.09	0.40	757	0.11	0.35	718	0.13	0.35	718	0.13	0.25	611	0.17	0.20	522	0.21	0.05	281	0.41
26	Kirby Muxloe	110	0.65	3064	0.02	0.55	2992	0.02	0.55	2992	0.02	0.55	2992	0.02	0.45	2701	0.03	0.35	2285	0.04	0.35	2285	0.04	0.10	1234	0.08	0.05	937	0.11	0.05	937	0.11
6	Barton in the Beans	108	0.65	1496	0.04	0.55	1265	0.06	0.55	1265	0.06	0.50	1230	0.06	0.40	1053	0.07	0.35	970	0.09	0.25	831	0.11	0.25	831	0.11	0.15	600	0.17	0.10	481	0.21
47	Sibson	98	0.45	1177	0.04	0.40	1020	0.05	0.25	744	0.09	0.25	744	0.09	0.20	649	0.12	0.20	649	0.12	0.15	532	0.15	0.10	406	0.21	0.05	310	0.30	0.05	310	0.30
41	Ratcliffe Culey	95	0.65	1131	0.04	0.50	724	0.07	0.30	455	0.13	0.20	358	0.17	0.15	291	0.23	0.10	226	0.35	0.10	226	0.35	0.05	144	0.60	0.05	144	0.60	0.00	0	0.00
36	Orton-on-the-Hill	89	0.65	354	0.13	0.60	338	0.15	0.55	334	0.17	0.50	304	0.20	0.45	286	0.23	0.40	281	0.25	0.10	135	0.56	0.10	135	0.56	0.05	83	1.05	0.05	83	1.05
46	Shenton	62	0.55	146	0.23	0.50	135	0.28	0.50	135	0.28	0.40	98	0.43	0.35	100	0.53	0.35	100	0.53	0.35	100	0.53	0.35	100	0.53	0.10	38	1.55	0.10	38	1.55
55	Upton	51	0.85	166	0.16	0.60	64	0.47	0.55	60	0.55	0.45	60	0.77	0.45	60	0.77	0.45	60	0.77	0.45	60	0.77	0.45	60	0.77	0.45	60	0.77	0.10	13	3.70
45	Sheepy Parva	50	0.50	1205	0.02	0.50	1205	0.02	0.45	1108	0.03	0.35	942	0.04	0.30	831	0.04	0.15	568	0.07	0.15	568	0.07	0.10	434	0.11	0.10	434	0.11	0.05	273	0.18
35	Odstone	38	0.70	877	0.02	0.35	433	0.06	0.35	433	0.06	0.35	433	0.06	0.25	317	0.09	0.25	317	0.09	0.20	274	0.12	0.20	274	0.12	0.10	189	0.19	0.05	120	0.32
32	Newtown Unthank	37	0.95	9557	0.00	0.95	9557	0.00	0.95	9557	0.00	0.45	5402	0.00	0.45	5402	0.00	0.15	3013	0.01	0.10	2147	0.01	0.05	1959	0.02	0.00	0	0.00	0.00	0	0.00
56	Wellsborough	31	0.60	75	0.21	0.30	32	0.56	0.25	29	0.65	0.20	27	0.93	0.20	27	0.93	0.20	27	0.93	0.20	27	0.93	0.15	21	1.36	0.15	21	1.36	0.10	14	2.11
58	Wykin	28	0.55	416	0.04	0.55	416	0.04	0.55	416	0.04	0.20	208	0.11	0.20	208	0.11	0.20	208	0.11	0.15	151	0.17	0.15	151	0.17	0.10	122	0.23	0.10	122	0.23
8	Bilstone	25	0.50	1080	0.01	0.35	970	0.01	0.25	831	0.02	0.20	789	0.02	0.20	789	0.02	0.20	789	0.02	0.15	513	0.04	0.10	416	0.06	0.10	416	0.06	0.10	416	0.06
34	Nuneaton	23	0.50	1	10.98	0.50	1	10.98	0.40	1	18.87	0.35	1	23.88	0.30	1	33.42	0.25	0	43.32	0.20	0	53.88	0.20	0	53.88	0.20	0	53.88	0.15	0	66.07
16	Copt Oak	21	0.80	199	0.06	0.70	164	0.08	0.70	164	0.08	0.40	68	0.20	0.30	58	0.29	0.30	58	0.29	0.30	58	0.29	0.20	36	0.50	0.10	23	0.92	0.10	23	0.92
39	Pinwall	20	0.50	143	0.08	0.50	143	0.08	0.45	122	0.10	0.40	104	0.13	0.30	78	0.19	0.30	78	0.19	0.25	74	0.26	0.25	74	0.26	0.25	74	0.26	0.25	74	0.26
2	Atterton	17	0.15	1524	0.01	0.15	1524	0.01	0.15	1524	0.01	0.10	1350	0.01	0.10	1350	0.01	0.10	1350	0.01	0.05	890	0.02	0.05	890	0.02	0.00	0	0.00	0.00	0	0.00
21	Ellistown	16	0.90	62	0.14	0.90	62	0.14	0.60	15	0.82	0.60	15	0.82	0.60	15	0.82	0.60	15	0.82	0.50	10	1.37	0.50	10	1.37	0.40	7	2.09	0.30	5	3.01
1	Appleby Magna	15	0.70	2078	0.00	0.70	2078	0.00	0.70	2078	0.00	0.70	2078	0.00	0.15	1524	0.01	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00

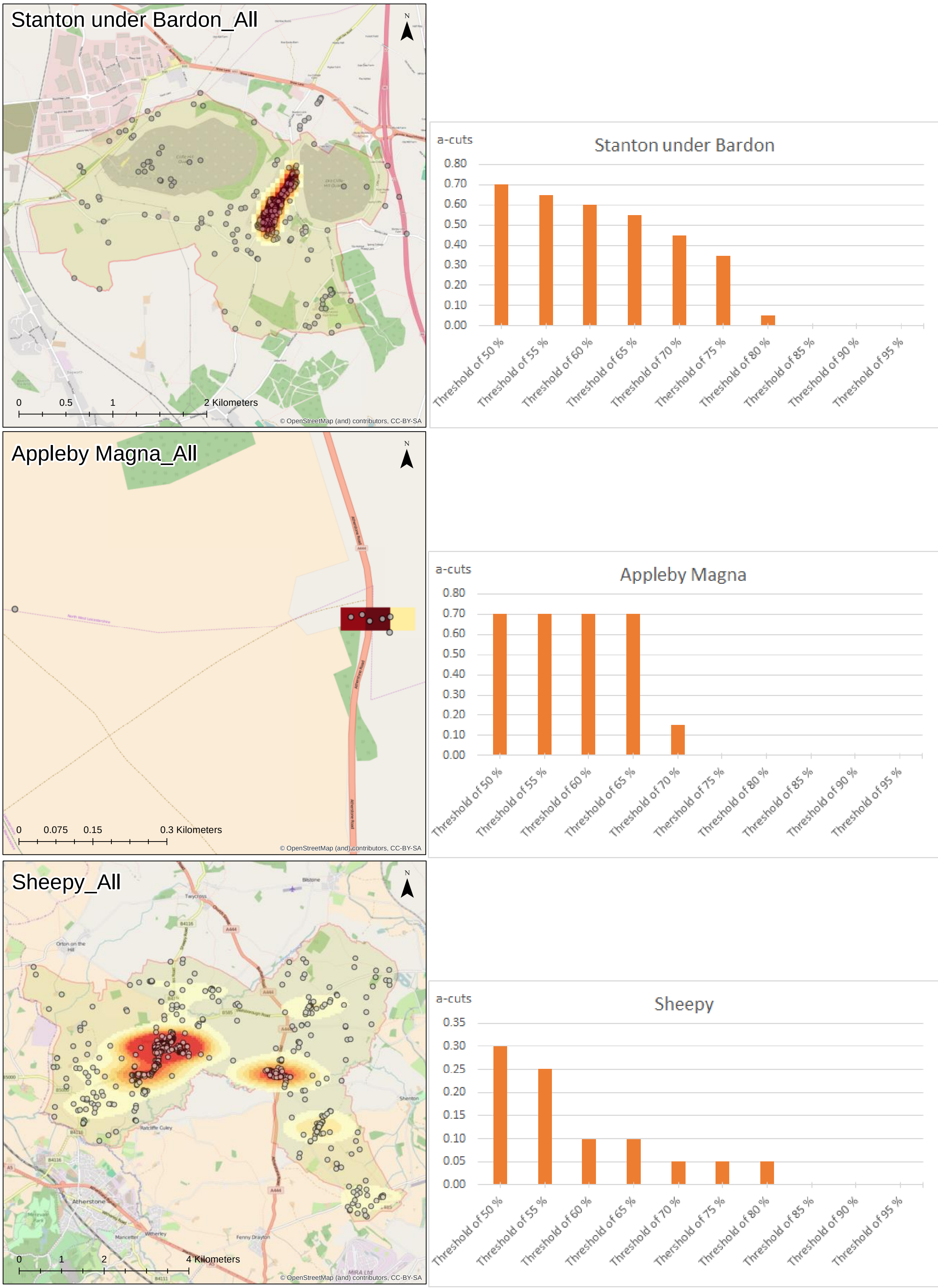


Figure 4.21: Three examples of settlements (maps on the left showing their address points and fuzzy models with bar charts on the right indicating the optimal α-cut for each threshold) that miss out some thresholds because of their point distribution which contain some diffused point(s).

- For some villages, it appears that the optimum α-cuts for all thresholds are quite large with a gradual transition between the start and end values. This is found in some nucleated village or clustered settlements such as Burbage, Earl Shilton, Trinity, Castle and De Montfort. Figure 4.22 presents mapping examples of Earl Shilton and Trinity with their α-cut plots.

Figure 4.22: Two examples of settlements with overall high-grade of membership and gradual transition between their start and end α-cut values. Maps on the left show their actual address points and fuzzy models with bar charts on the right indicate the optimal α-cut for each threshold.

- There are few settlements with low membership values generally not exceeding 0.3 α for any threshold value. These are shown in Atterton, Cadeby and Osbaston. The reason for this is mainly related to the dispersal form of their points, as viewed in the maps and par charts of Cadeby and Osbaston in Figure 4.23.

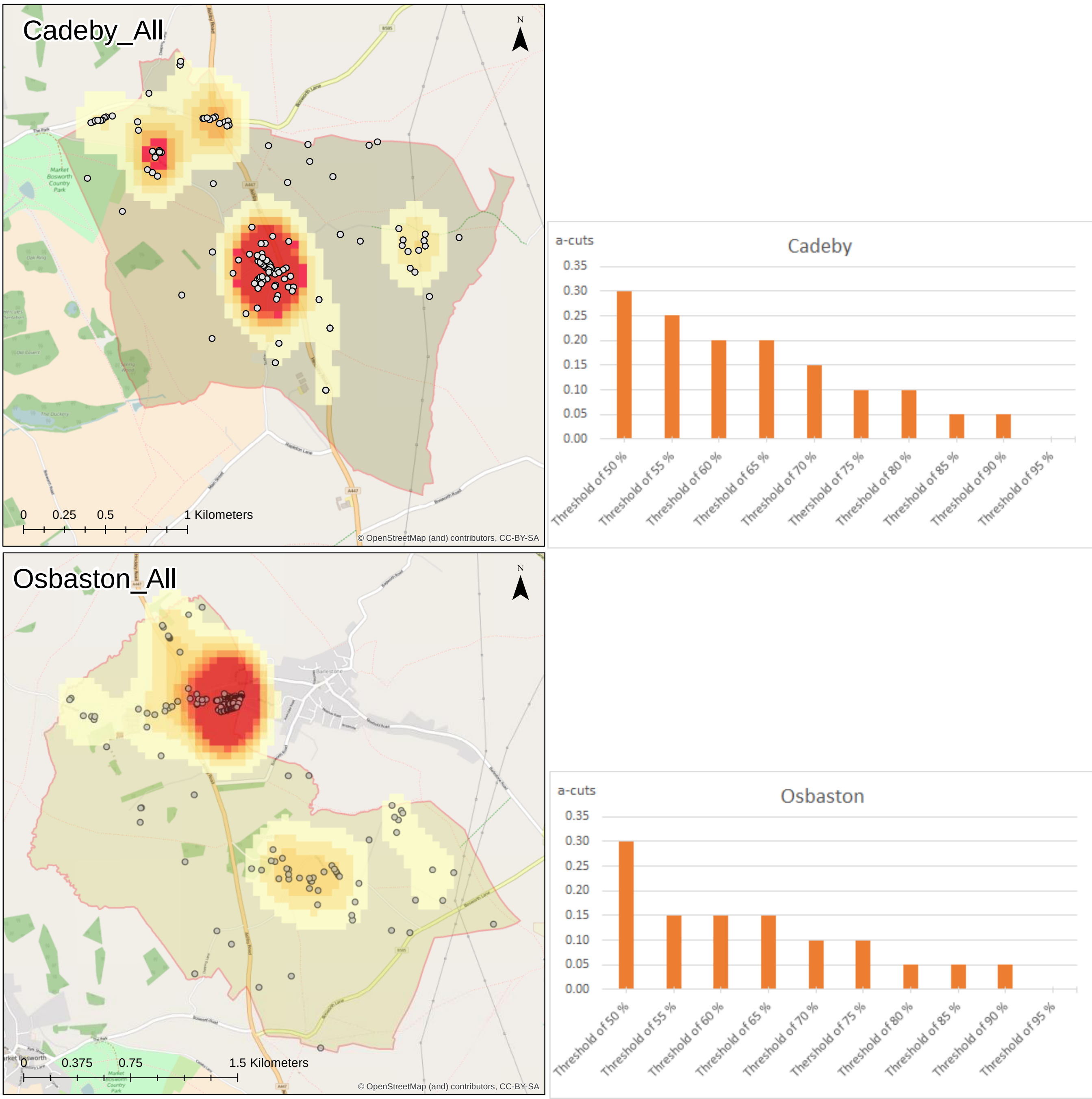


Figure 4.23: The case of some settlements with overall low-grade of membership values across all thresholds indicating the disperse nature of their points. Maps on the left show their actual address points and fuzzy models with bar charts on the right indicate the optimal α-cut for each threshold.

- In some of the settlements, the α-cut value stays stable for number of thresholds (mostly high percentage), such as appeared in Apply Magana, Upton, Shenton, Pinwall and Wellsborough. These settlements appear to be in the last section of the table where the number of points is small. Moreover, their points are distributed in a way that allows them to be captured equally in several thresholds. This can be clearly observed in the cases of Upton and Pinwall in Figure 4.24.

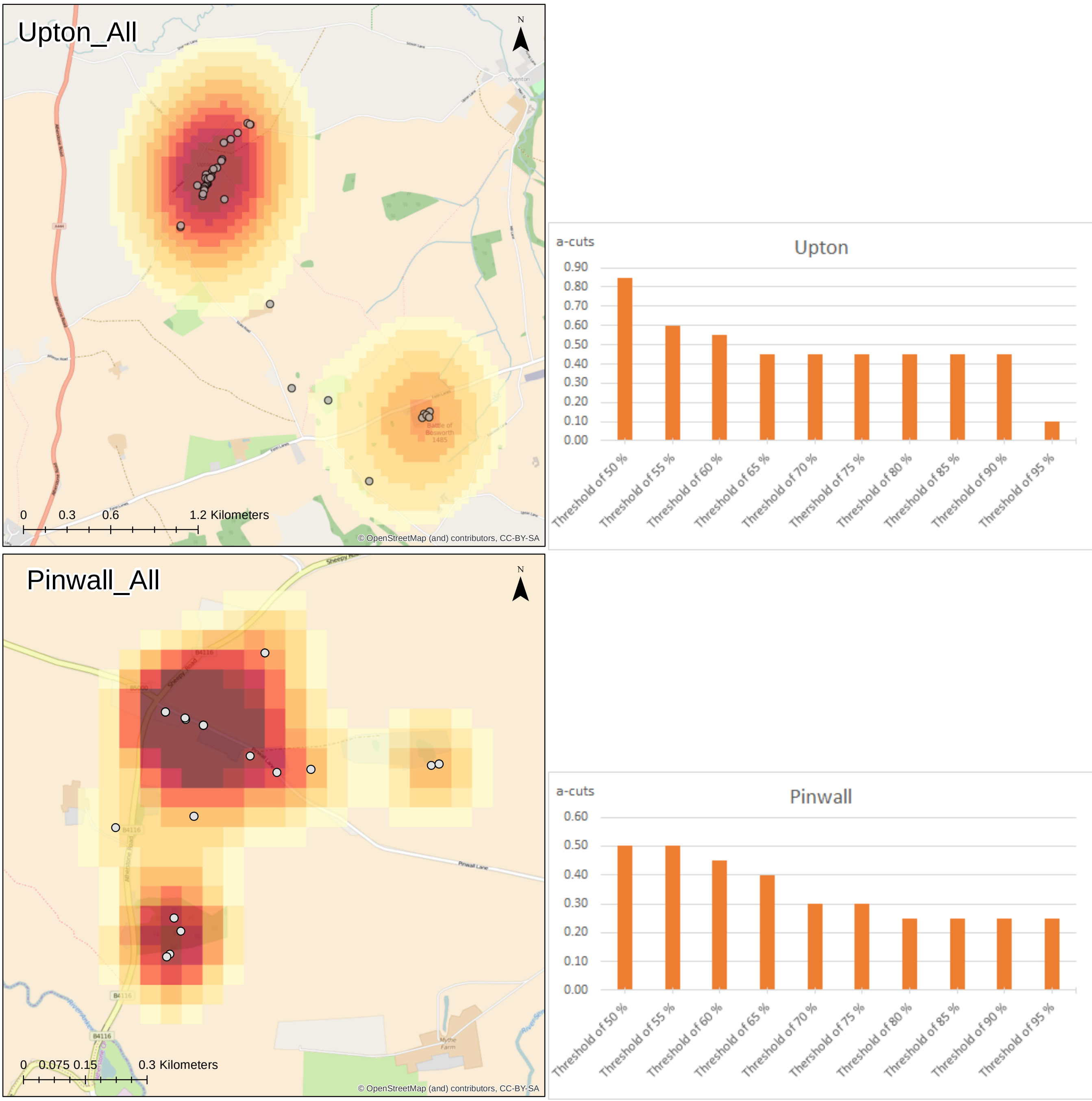


Figure 4.24: Examples of settlements that have some of their α-cut values stays stable reflects in the spatial distribution of their points. Maps on the left show their actual address points and fuzzy models with bar charts on the right indicate the optimal α-cut for each threshold.

- As a general observation, most of the settlements tend to have gradual membership values when the majority of their addresses are grouped around a certain area or areas in multiple clusters with other focal points being scattered further away. This is evidenced by several settlements such as Desford, Market Bosworth, Fenny Drayton, and Barton in the Beans.

Maps and plots for Desford and Barton in the Beans are presented in Figure 4.25.

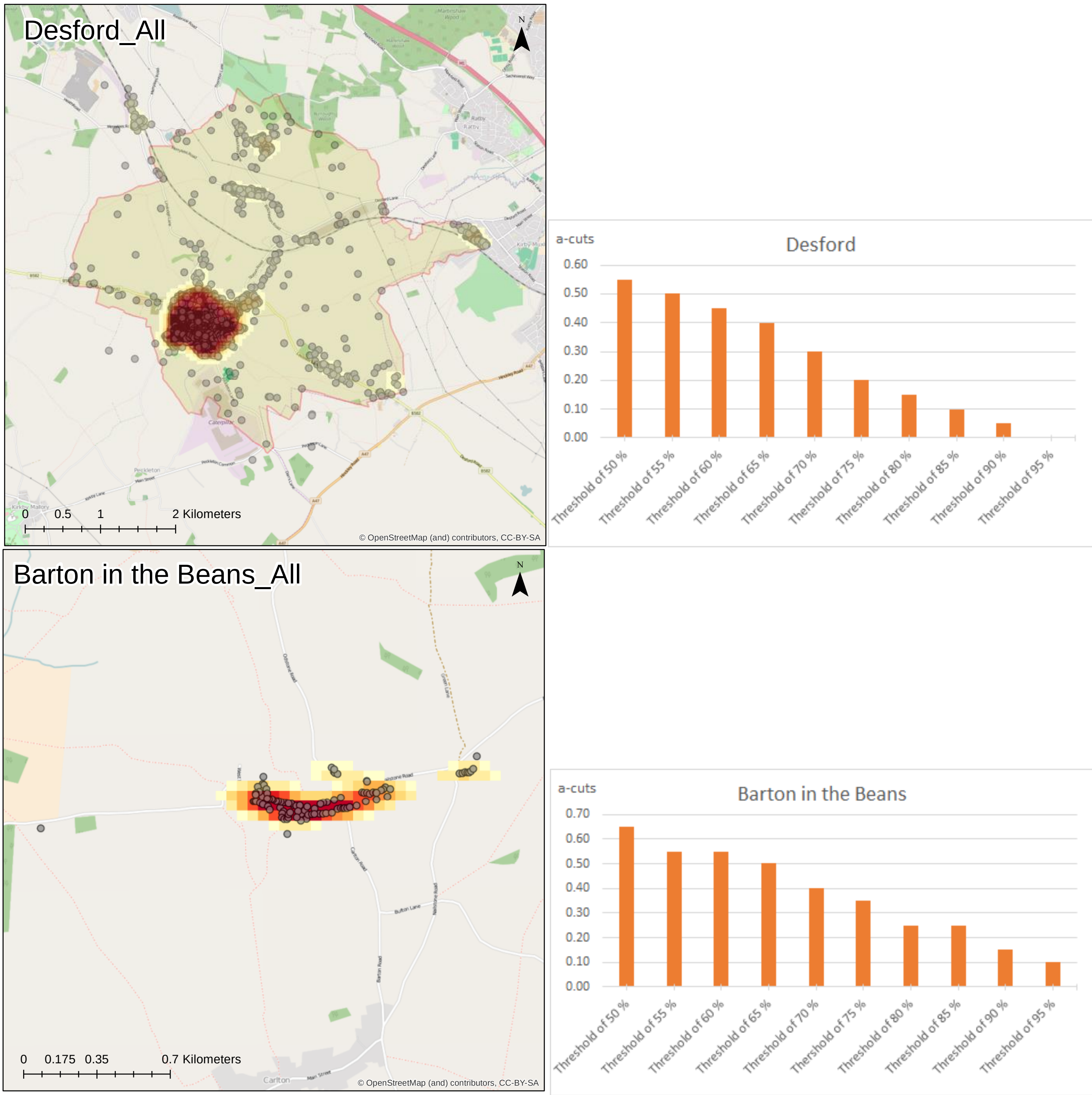


Figure 4.25: Two examples of villages with gradual transition between their α-cuts as most of their address appear more clustered in the middle. Maps on the left show their actual address points and fuzzy models with bar charts on the right indicate the optimal α-cut for each threshold.

- By contrast, as the variance between α-cuts in some villages declines suddenly, this implies the existence of large gaps or obvious outliers that have deviated from the overall pattern of the address points. This can be encountered, for instance, in Sutton Cheney, Newtown Unthank, Odstone and Wykin. Figure 4.26 shows the instances of Sutton Cheney and Odstone.

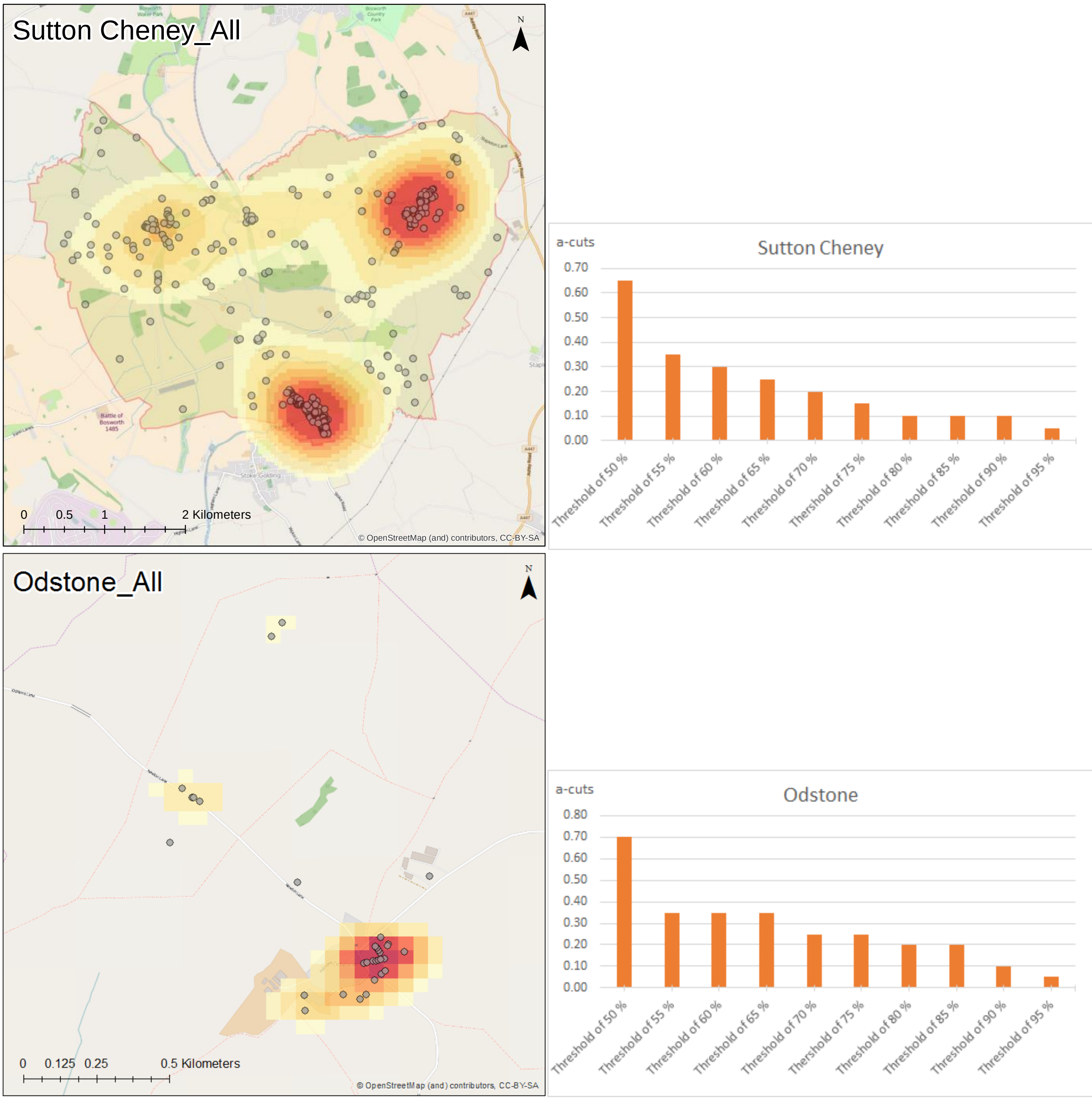


Figure 4.26: Two examples of villages which have sudden decline in their α-cuts as part of their addresses are quite spread or defuse. Maps on the left show their actual address points and fuzzy models with bar charts on the right indicate the optimal α-cut for each threshold.

<u>Secondly, exploring relationship with point densities:</u>

- For point density recorded in Table 4.5, again there is no predictable relationship between number of points and the density. Nevertheless, generally speaking, the top quarter of the table (which has a larger number of addresses and appears in darker green) shows high density values, whereas the highest density is observed in Newtown Unthank, which holds few addresses from the different data sources.

- Another indication of this misleading relation is that Stanton under Bardon, which is located in the second quarter of the table, is denser than other settlements located in the top quarter of the table such as Stoke Golding, Castle and Desford in all thresholds. In addition, Nuneaton, which has the lowest density, still holds the biggest number of addresses compared to other villages presented in the bottom of the table (Copt Oak, Pinwall, Atterton, Ellistown and Appleby Magna).

- By the same token, horizontally, density behaves in exactly the same fashion as viewed in previous results (Tables 4.2, 4.3 & 4.4). It has an inverse relationship with the threshold limits.

Comparison between results/villages in common

To compare between the results obtained from the different data sources (Tables 4.2, 4.3, 4.4 & 4.5), the focus is set only on villages in common. There are several aspects to note here regarding the inclusion analysis through different datasets. For a start, there are 18 settlements identified in the tables. These appear in a different order in each table since each dataset recognises a different number of addresses for a given village. Table 4.6 presents these settlements according to their address numbers.

Table 4.6: Lists of settlement names and numbers of their address points that exist in each datasets.

Order ID	BS76		POST		POI		ALL	
	Settlement Name	*No. of Points*	*Settlement Name*	*No. of Points*	*Settlement Name*	*No. of Points*	*Settlement Name*	*No. of Points*
1	Burbage	6979	Burbage	6635	Burbage	257	Burbage	13871
2	Earl Shilton	4884	Earl Shilton	4763	Earl Shiltoin	227	Earl Shilton	9874
3	Barwell	4382	Barwell	4244	Barwell	221	Barwell	8847
4	Groby	3282	Groby	2954	Groby	113	Groby	6349
5	Ratby	2010	Ratby	1904	Market Bosworth	106	Ratby	3974
6	Desford	1864	Desford	1497	Desford	95	Desford	3456
7	Newbold Verdon	1429	Newbold Verdon	1376	Newbold Verdon	70	Newbold Verdon	2875
8	Market Bosworth	1140	Barlestone	1080	Ratby	60	Market Bosworth	2294
9	Barlestone	1118	Market Bosworth	1048	Stoke Golding	59	Barlestone	2228
10	Stoke Golding	797	Stoke Golding	793	Barlestone	30	Stoke Golding	1649
11	Witherley	767	Witherley	313	Stanton under Bardon	23	Witherley	1093
12	Peckleton	626	Higham on the Hill	284	Peckleton	18	Higham on the Hill	782
13	Twycross	508	Stanton under Bardon	281	Higham on the Hill	15	Peckleton	739
14	Higham on the Hill	483	Nailstone	215	Sutton Cheney	14	Stanton under Bardon	670
15	Stanton under Bardon	366	Twycross	148	Twycross	14	Twycross	670
16	Sutton Cheney	339	Osbaston	119	Nailstone	13	Nailstone	488
17	Nailstone	260	Peckleton	95	Witherley	13	Sutton Cheney	418
18	Osbaston	161	Sutton Cheney	65	Osbaston	10	Osbaston	290

The second point in the comparison concerns α-cuts profiles per a village. Plots of these α-cuts for all examined villages are provided in Appendix (4), based on which three clear distinctions can be made in the way that membership grades of inclusion appear in each dataset:

- Settlements in the first group predominantly have a similar attitude towards the variation in α-cut values, to wit: Barlestone, Barwell, Burbage, Earl Shilton, Groby, Market Bosworth, Newbold Verdon, Ratby and Stoke Golding. Often in these settlements the α-cut values correspond to any particular threshold limit, mostly equivalent in the BS76, POST and all data, and often greater than the α-cuts in the POI data. This emphasises the fact that the BS76 and POST data are in large part synonyms, which originally have very similar distribution of their address points except for a few outliers. 'All data', as its name indicates, is the union of all the three data sources combined, and so matches the BS76 and POST data most times. Figure 4.27 presents plots of two villages that have similar α-cut profiles.

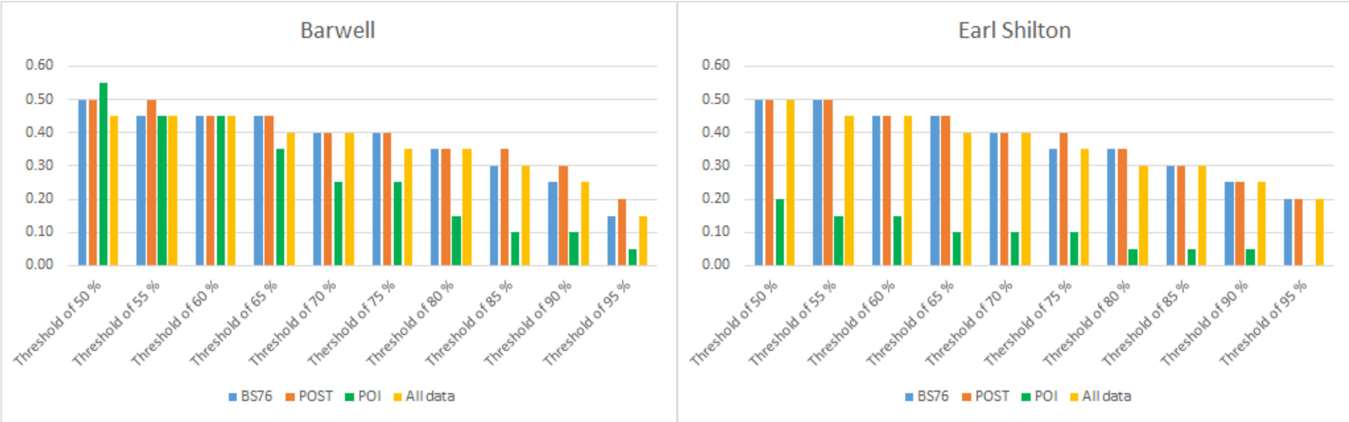


Figure 4.27: Plots showing fuzzy membership values for each percentage in each data for two villages, which have similar α-cuts profile as α-cuts are almost identical in the BS76, POST & All data, but smaller in the POI. Full-size versions of the plots are in Appendix (4).

- The second group represent settlements that have quite small differences between their membership grades among thresholds. These are Desford, Nailstone, Peckleton, Twycross and Witherley. The main common characteristic found in this group is that these villages in the BS76 data have far larger addresses than in the POST and POI data, which mostly have homogeneous distribution. Figure 4.28 depicts the plot of α-cuts for two villages with small differences between their values.

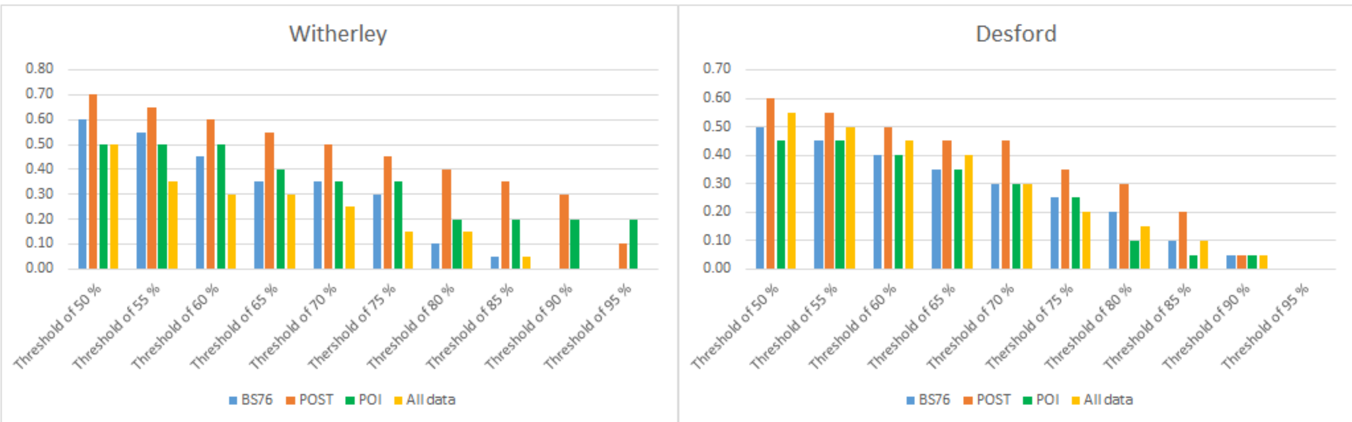


Figure 4.28: Plots showing fuzzy membership values for each percentage in each data for two villages, which have small variations between their α-cuts. Full-size versions of the plots are in Appendix (4).

- Higham on the Hill, Osbaston, Stanton under Bardon and Sutton Cheney represent the final group, which has a quite different level of α-cut profiles. This is most likely caused by the differences between the settlements in terms of the number and structure of the address points that constitute them. Two examples of membership plots are shown in Figure 4.29.

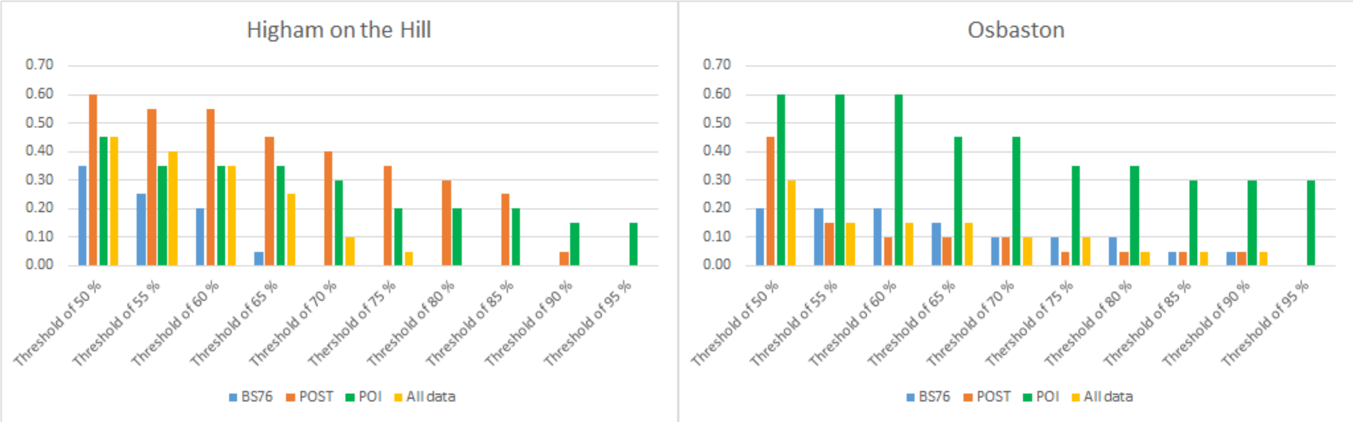


Figure 4.29: Plots of fuzzy membership values for each percentage in each data for two settlements that have quite different α-cut profiles. Full-size versions of the plots are in Appendix (4).

In regards to density of points, the final aspect to analyse in the comparison, the settlements are also classified in two divisions in the way in which density values varied across the sources per village. Plots of the density values for the examined settlements are provided in Appendix (4)

- First group contains more than half of the settlements in which their density values deviate in the same manner. These are Barlestone, Barwell, Burbage, Desford, Earl Shilton, Groby, Market Bosworth, Nailstone, Newbold Verdon, Ratby and Stoke Golding. In these villages, the density values in the BS76 and POST data are so close where values in the POI are extremely smaller; and in the combined data the values are doubled. The explanation of this is that the BS76 and POST data are two sides of the same coin; and thus when the data are combined the possibility to double count the addresses increase. In contrast, the POI data by its nature has fewer number of addresses located in smaller areas. Figure 4.30 presents two settlements in which their density values change in the same manner.

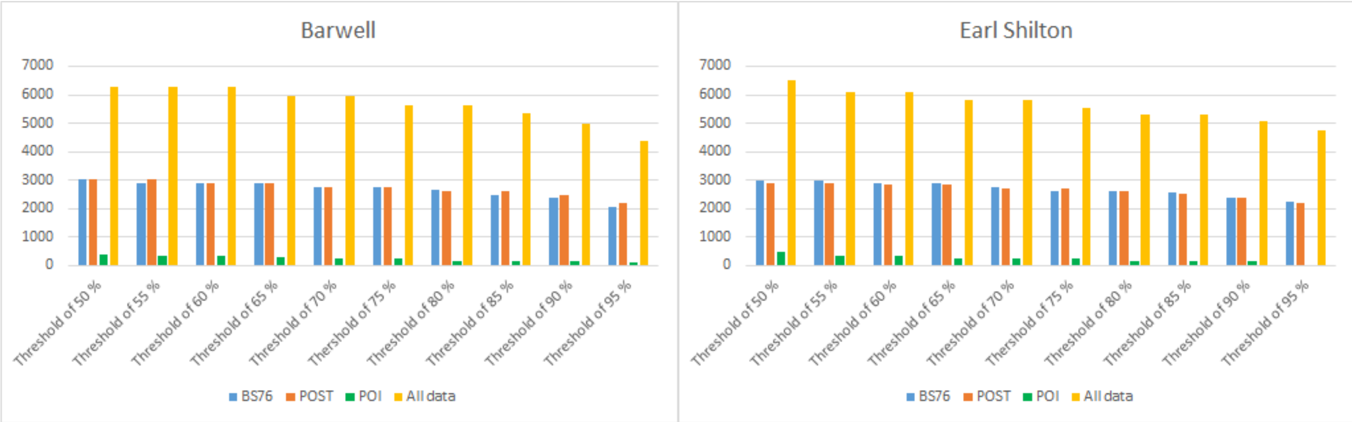


Figure 4.30: Plots of fuzzy membership values for each percentage in each data for two villages that have density values nearly equal in the BS76 & POST data and almost doubled in all data combined and extremely small in the POI data. Full-size versions of the plots are in Appendix(4).

- The second category contains a few settlements with irregular variation in density values. This can be observed in Higham on the Hill, Osbaston, Peckleton, Stanton under Bardon, Sutton Cheney, Twycross and Witherley. It is noted that the density values in the POST data are much higher than the other data and even, in some cases, higher than the combined data. These increments might be explained in two ways:
    - In comparing the values in the BS76 and All data versus the values in the POST data, it is believed that the latter frequently has few addresses clustered and distributed in quite small areas.
    - In contrast, the number of addresses in the POI data is comparable to the addresses in the POST data, but they appear more spread out in the POI data.

Figure 4.31 gives examples of two settlements that have a different pattern of transition in their density values among each data sources.

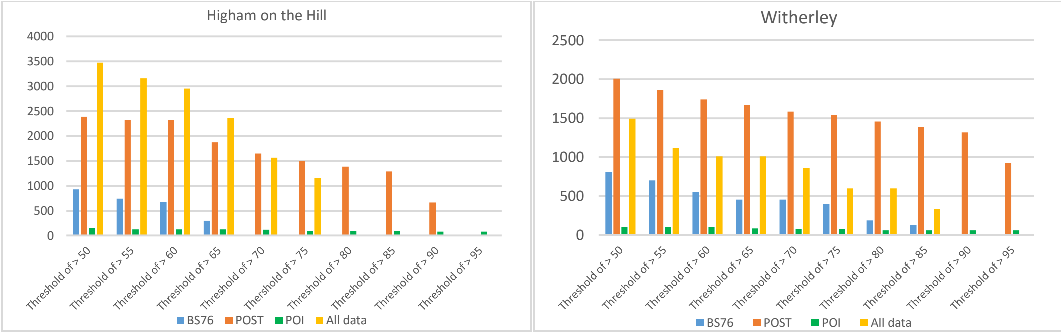


Figure 4.31: Plots of fuzzy membership values for each percentage in each data for two settlements that irregular pattern of variation in their density values. Full-size versions of the plots are in Appendix (4).

4.6 Summary and General Discussion

The present study is designed to determine the effect of using fuzzy set theory in modelling vague regions, focusing on one such example, which is rural settlements. It involves representing these settlements as fuzzy features using spatial density estimation methods. Before concluding, this section reflects on broader issues pertinent to the research approaches and theoretical assumptions presented in this chapter. These are discussed under three headings: Data used, Fuzzy model of the village and Using α-cuts.

4.6.1 Data Used:

As described in the introduction, this chapter uses address data from three sources to model where rural settlements are and approximate their extents. These are empirical data provided from the UK national mapping agency, Ordnance Survey; OS AddressPoint (BS76 & POST) and OS Point of Interest. Details about these data sources and the processes used in identifying village names are available in Chapter 3. However, it is important to point out that these are used individually and only used one time used in conjunction with each other (All data in the analysis of inclusion) to achieve the fuzzy models of villages for comparison purposes. Because the BS76 and POST data are two variants of the same dataset, then in most cases, there is no *a priori* reason to believe that they would yield substantially different results, although differences still exist with the number of villages missing from one dataset (e.g. Castle, De Montfort and Fenny Drayton), or large number of address points missed out or appearing under different village names (as fully discussed in Chapter 3, Section 3.6).

The importance of considering different types of data (informal and formal) is that in Britain, unlike many other places, much data are available for small areas such as rural places with relatively very small populations. However, in the absence or unavailability of such formal data, POI/point of interest data provide an efficient alternative. So from this viewpoint, the OS POI comes to consideration as a third source.

4.6.2 Fuzzy Model of the Village

With references to key scientific publications on modelling vague regions (reviewed in Chapter 2, Section 2.6), the approach suggested in this work advances from using fuzzy set theory to represent rural settlements based on the distribution of addresses derived from their density patterns. In other words, it involves the identification of two main needs: (1) finding the location of a village, and (2) knowing the spatial extent (footprint) of that village. Based on the assumption that the actual location of the village is not clearly defined, all maps presented in this chapter are set to the same geographical spatial extent of the study area. Moreover, it is possible to see that this approach satisfactory yields results which model the reality of rural areas, as a fuzzy geographical entity, better than the traditional Boolean approaches. Such a comparison would in itself represent a valuable contribution, especially when most studies focus on people's perception about vague regions (other than villages) and not on how these places are formally identified. However, it should be acknowledged that the current work implicitly assumes that the fuzzy region is continuous. Unsurprisingly, that is not always the case, as it is commonly seen in villages that intensity of houses varies around the village area (see, for example, Figures 4.11, 20, 23 & 26). This has raised another challenge, which has not yet been considered in this study; for example, the possibility of the existence of obstacles in the region such as a natural or artificial pond or lake used for the storage and regulation of water.

There is abundant room for further investigation, for a situation querying if a house has one village name as its address and it is within a parish of that name. Does it mean that it cannot be in any other village to some (a small) degree, or even quite a large degree? This is an important question, but one that needs asking; and, more to the point, it is important to ask how people living in such places identify their address. This is clearly shown in the fuzzy results arising from the data analysis above.

4.6.3 Using α-cuts

Armed with this model technique, taking α-cuts of the fuzzy objects is extensively exploited in the analysis of inclusion (Section 4.5.3). It is possible to perceive that an α-cut acts as a filter on a fuzzy set to separate high- and low-value elements based on a single threshold (Katinsky, 1994). It has been used in geographical analysis by various researchers (Arnot *et al.* 2004; Fonte and Lodwick, 2004; Fisher, 2010; Schmitz and Morris, 2006). However, in this chapter it is further examined with the pre-specification of the proportion included from the original address points within any particular α-cuts (Tables 4.2, 4.3, 4.4 & 4.5). Perhaps the only generalisation that can be made is that, as the percentage threshold increases, the membership value decreases; which makes sense because this implies including more of the address points, which is more possible with small α-cuts.

Hence, it could conceivably be hypothesised that using α-cuts in this way gives statistical evidence for inclusion of address points within a candidate vague representation of the village area. This corroborates the ideas of Jones *et al.* (2008), who suggested creating a model of a vague region that reflects the variation in confidence of inclusion; and, in doing so, it is possible to generate hard approximations of the region at different levels of confidence as required. This finding has important implications for developing any subsequent analysis. It can thus be suggested to consider number of α-cuts that at least cover 70 or 75% of the address points to be involved, as introduced in the next chapter.

Chapter 5 Implementation of the Travelling Salesman Problem in Fuzzy Locations

5.1 Introduction

The travelling salesman problem (TSP) is well known and researched within both operations research and computer science, and is regularly employed in a wide variety of applications: vehicle routing, computer wiring, cutting wallpaper, and job sequencing or machine sequencing and scheduling in the field of operations research are some examples in the field of operation research. The objective of the TSP is to cover all the towns or cities in a given area using the minimum driving distance, cost or time. Most work, however, is conducted where the visited locations are considered to be known exactly. In real-life situation it may not be possible to get the cost or time as a certain quantity, but also many geographical places have indeterminate or fuzzy locations. As was pointed out previously in Chapter 2, fuzzy set theory was introduced by Zadeh (1965) to directly address the problem of vagueness and imprecision. It has been argued, as a result, that any statement about a vague phenomenon must itself be allowed to be vague (Fisher *et al.* 2007). That simply means if the value of the cost, time or distance is not certain (crisp values), then the travelling salesman problem should become a fuzzy problem as well. Since then, significant progress has been made in developing numerous techniques to address the fuzzy travelling salesman problem (see Section 2.7 for further detail and references). It is noted that, although these studies offer different treatments of the TSP problem, they are mainly focused on the computational aspect either by developing a new algorithm or making modification to an existing one to solve the TSP. More important, there is a paucity of information, if any, on the uncertainty and fuzziness of the locations themselves. This chapter, in accordance with the main goal of this thesis, takes a slightly different view by considering the implication of using the traditional method of the travelling salesman problem (not fuzzy) on fuzzy locations.

This chapter is structured in the following manner: Section 5.2 begins by laying out the theoretical dimensions of the chapter's concepts, and looks at how these are applied in the field. Section 5.3 outlines the details of the implementation of the method of analyses, and the relevant preparation of the data. Sections 5.4 and 5.5 present and discuss the results obtained from applying the travelling salesman problem. Finally, a concluding summary and suggestions for future work are given in Section 5.6.

5.2 Background Information on TSP

5.2.1 General Overview and History:

The travelling salesman problem (also known as travelling salesperson problem or TSP for short) is a well-known and important combinatorial optimisation problem. The TSP has mathematical origins and has been identified as an element of graph theory (Curtin *et al.* 2013). It was studied in the middle of the 18th century by an Irish mathematician named Sir William Rowam Hamilton and a British mathematician named Thomas Penyngton Kirkman.

Theoretically, the TSP is a classical 'tour' problem in which a hypothetical salesman needs to find the most efficient sequence of destinations in a territory, stopping only once at each, while returning at the end of the tour to the initial starting location (Curtin *et al.* 2013). Effectively, in a TSP, it does not matter from which city or which node the travelling salesman starts. The only concern is, although the person can start from any node, he has to visit every other node once and only once and come back to the staring node in as short a route as possible. If n is the number of cities to be visited, then the total number of possible routes covering all cities can be given as a set of feasible solutions of the TSP and is given as $(n - 1)!/2$. The question is then to find among these feasible solutions the one with the best value or the minimum distance travelled. As n increases the computational time one would need to evaluate all possible tours, this results in a problem which is extremely hard to solve. In fact, the TSP has been proven to be an NP-complete combinatorial optimisation problem, which means that no polynomial-time algorithm is known for solving it (Curtin *et al.* 2013; Maredia 2010).

5.2.2 *Structure and Formulations of the TSP:*

There are many structures and formulations for variants of the TSP, employing a variety of constraints that enforce the requirements of the problem (Curtin *et al.* 2013), although only two of the common mathematical formulations are discussed in this research. One of these comes from a graph theoretic problem – a graph is a collection of nodes or vertices representing cities and arcs or edges representing distances between the given cities. Here the TSP is formulated by means of a complete graph in which each node is connected to each of the others with one edge between each pair of nodes (Caldwell 1995). The goal is then to find out whether the graph is Hamiltonian; or, more specifically, finding a Hamiltonian cycle that visits each node in the graph exactly once with the least weight in the graph. There is another way of rephrasing the TSP issue, which naturally amounts to finding the minimum spanning trees, the connection between all nodes with least weight, for tour construction or edge exchanges to improve existing tours (Hahsler and Hornik 2007).

The traveling salesman problem can be described as follows:
TSP = {(G, f, t): G = (V, E) a complete graph,
f is a function V×V Z, →
t ∈ Z,
G is a graph that contains a traveling salesman tour with cost that does not exceed t}.

The other way of representing the TSP issue can also be structured as integer and linear programming problems. The integer programming (IP) formulation is based on the assignment problem with an additional constraint of no sub-tours (Márquez and Nieto, 2013). Suppose the solution matrix X $=(x_{ij})$ of the assignment problem represents a tour or a collection of sub-tours (several unconnected cycles) where x_{ij} is equal to1, if the person goes immediately from i to j and the objective function is to minimise the total distance travelled. This can be explained by the following notations:

Minimize $\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}x_{ij}$

Subject to $\sum_{i=1}^{n} x_{ij} = 1, \quad \forall j$

$\sum_{j=1}^{n} x_{ij} = 1, \quad \forall i$

$x_{ij} = \{ 0 , 1 \}$

no sub tours allowed

5.2.3 Methods to solve the TSP:

There are ongoing efforts to develop solution procedures for the TSPs. However, some of them are discussed in this research. These approaches can be handled under two categories: (1) exact solution procedures; and (2) approximate approaches.

Exact solution approaches:

This approach guarantees providing optimal solutions of the TSP, but – due to the combinatorial complexity of the problem – they can generally only be successfully used for modestly sized problem instances. When one thinks of solving the TSP, the first method that might come to mind is to evaluate all possible combinations of network elements (tours) and to choose the set that performs best (Curtin *et al.* 2013). This simply means: generate all possible tours and compute their distances, and then the shortest tour is thus the optimal solution. This method is termed complete enumeration (Curtin *et al.* 2013) or the brute-force method (Maredia 2010, Dasgupta *et al.* 2006) Naive Solution (Goyal, 2010). This strategy definitely works well as long as the number of nodes (n) is relatively small. This is because solving a TSP with a large number of n can be frustrating and can even take years or centuries, as it is highly unlikely to be solvable in polynomial time (Maredia 2010, Dasgupta *et al.* 2006).

A much faster approach was proposed by Held and Karp (1962) to find the optimal solution for the TSP based on the dynamic programming algorithm. The idea of this technique is simply to break a problem down into sub-problems to get partial solutions by solving a smaller problem. Having solved all these sub-problems, the answer to the main question (the original problem) would correspond to the biggest sub-problem (Awuni 2014). In solving the TSP in this case, the most obvious partial solution is the initial portion of a tour, which seeks to solve a problem by first solving smaller instances of the same problem by looking at a slightly smaller-sized problem. Suppose that there are n cities or destinations and a salesman who has started the tour at city 1, then the question is: what is the best order to visit just one of the destinations?

Here, then, is an appropriate sub-problem:

The TSP is to find a permutation $P = (1, i_1, i_2, \dots i_n)$ that minimise the distance $a_{\alpha\beta}$ between each pair of cities in the tour.
Given a subset of city indices (excluding the first city) $S \subset \{2, 3, \dots, n\}$ and $l \in S$, let $D(S, l)$ denote the minimum length of a path from city *1* to city l that visits precisely all cities or vertices in the set S exactly once each. Then

a. $(n(S) = 1): \quad D(\{l\}, l) = a_{1l} \qquad \forall l \in S$

b. $(n(S) > 1): \quad D(S, l) = min_{m \in S-l}[D(S - l, m) + a_{ml}]$

where m is the city that immediately proceeds the final destination in the tour city l
Then the minimum length of the path for a complete tour can be obtained by recursively compute the quantities in equation b; which take the form:

$$D^{`} = min_{l \in \{2,3,\dots,n\}}[D(\{2,3, \dots n\}, l) + a_{l1}]$$

It should point out the fact that a permutation P can only be optimal if, and only if,

$$D^{`} = D(\{2,3, \dots n\}, i_n) + a_{i_n 1}$$

and, for $2 \leq p \leq n - 1$,

$$D(\{i_2, i_3, \dots, i_p, i_{p+1}\}, i_{p+1}) = D(\{i_2, i_3, \dots, i_p\}, i_p) + a_{i_p i_{p+1}}$$

A different method that can deal with larger instances is based on the *branch-and-bound algorithm*. This strategy is similar to the previous *dynamic programming* technique in dividing a problem to be solved into a number of sub-problems. When solving a sequence of sub-problems in the branch and bound approach, each sub-problem may have multiple possible solutions and the solution chosen for one sub-problem may influence the possible solutions of later sub-problems (Pothineni 2013). This means that, according to Hahsler and Hornik (2007), branching is done iteratively, yielding a binary tree of sub-problems each of which is either solved without further branching or is found to be irrelevant because it produces a solution with a longer path than a solution of another sub-problem. The general structure of the algorithm is as follows:

Given a subset of solutions S and
let $L(S)$ denotes a lower bound on the cost of any solution belonging to S, and
let C be the cost of the optimum solution found so far; then
if $C \leq L(S)$ → there is no need to explore S
else $C > L(S)$ → explore S further as it may contain a better solution

Approximate solution procedures – heuristics:

Heuristic solution procedures are approximate approaches that never guarantee an optimal solution but give a near optimal solution by obtaining feasible solutions within a reasonable amount of computing time. Curtin *et al.* (2013) point out two key criteria to evaluate heuristics, which are: (1) the total computational time (speed or the number of iterations required to reach a solution); and (2) performance with respect to the optimal solution. Although there is a wide variety of heuristics that can be applied to the TSP, these mainly fall into two categories: one in which tours are created from scratch – tour construction heuristics; and the other which uses simple local search heuristics to improve existing tours – tour improvement heuristics. Not all of these heuristics can be reviewed here; however, the following are discussed in this section: the nearest neighbour algorithm and the insertion algorithm as examples of a construction heuristic; and the k-Opt heuristics as a generic description of improvement heuristics.

The Nearest Neighbour heuristic (NN) (or so-called greedy algorithm) is the simplest and the most straightforward type of heuristic for solving the TSP. This NN algorithm allows the salesman to repeatedly choose the least-cost edge to cities not already in the tour, as the next move in the tour, and add that edge to the tour until all cities are reached. This algorithm quickly yields an effectively short route, but it does not provide particularly good solutions for even modest-sized problems (Curtin *et al.* 2013). An extension to this algorithm is to repeat it with each city as the starting point and then return to the best tour found; this heuristic is called *repetitive nearest neighbour* (Hahsler and Hornik 2007).

Many variations of the *insertion heuristic* are in common use: *nearest*, *furthest*, *cheapest* and *arbitrary* insertions. All of them follow the same strategy, which is to construct the approximation tour by a sequence of steps in which tours are constructed for progressively larger subsets of the nodes (Rosenkrantz *et al.* 1977; Matai *et al.* 2010). Or, simply, this means that the insertion methods start with a tour of a small set of cities, and then increase the tour by inserting the remaining cities one at a time until all the cities are visited. In this algorithm, a path of the tour is constructed as follows:

1. Start with an arbitrary city
2. Choose city k not yet in the tour, having the shortest distance to any one of the cities
   this city is inserted into the existing tour between two consecutive cities i and j, such that
   the insertion cost, the increase in the tour's length,
   $d(i,k) + d(k,j) - d(i,j)$ is minimized
3. Finally stop when all cities are on the tour

The insertion methods differ in the way the city to be inserted next is chosen; Table 5.1 explains these variations (adapted from Hahsler and Hornik 2007).

Table 5.1: Table showing four variations of the insertion heuristic and their strategy to choose which city to insert.

Variations	**Selection criterion** *The city chosen in each step as the city which is*:
Nearest Insertion	closest to nodes in the tour
Farthest Insertion	farthest from any of the cities on the tour
Cheapest Insertion	the cost of inserting the new city is minimal
Arbitrary insertion	chosen randomly from all cities not yet on the tour

Having constructed an optimal tour with any type of tour construction heuristics, a small modification can be made to the tour repeatedly stepwise in order to reduce its cost, and that is when the tour improvement heuristics come into consideration. Tour improvement heuristics are simple local search procedures which try to improve an initial tour by making local modifications or an iterative improvement process until a short tour is found (Karkory and Abudalmola 2013; Fischer 2014). A simple example of this type of algorithms is the k-Opt heuristics.

The idea behind the k-opt algorithm is to specify a tour and then perturb it in some way to check if an improved tour can be obtained by deleting k edges and replacing them with a set of different feasible edges (Matai *et al.* 2010; Fischer 2014). This is known as a k-Opt move and the resulting tour represents a local optimum which is referred to as the k-optimal (Hahsler and Hornik 2007; Matai *et al.* 2010). This heuristic is usually applied for k = 2 or k = 3. Empirical evidence suggests that the computational time increases as the number of k grows rapidly.

5.3 Implementation of the TSP

To apply the TSP in this research, three main stages are required. First, determine the locations to be visited in each settlement. Second, measure the road network distance between these locations. Finally, solve the Travelling Salesman Problem (TSP). These three stages are applied in two ways: Boolean, Crisp or Hard approach (original) and Fuzzy or Soft approach (new).

5.3.1 Determine Locations for TSP

Initially, it is helpful to focus the analysis here on the POI data and only consider the settlements with at least 10 addresses (as discussed in Chapter 4). It is further assumed that one way of deciding where the village is located is to take the centroid of the modelled village (Comber *et al.* 2008-a). It is best, thus, to regard the centre of the settlement as the location to be visited. On the *Boolean, Crisp or Hard approach*, the settlement is originally represented as a set of addresses (point features), and thus the centroids of each village (middle point or median) are identified. This has been done in ArcMap using the Spatial Statistics Tools to determine the Median Centre.

On the other hand, the settlements in the fuzzy approach are modelled as a set of α-cut (surfaces) that quantify the degree of uncertainty or vagueness related to the villages. Indeed, it is possible to take into account all levels of α-cuts per village or even take one specific α-cut and there is no reason to be identical (the same) in all settlements. Nevertheless, it seems difficult to choose one α-cut that satisfies all the villages, as discussed in Chapter 4 (Section 4.5.3). For this reason, in this stage the option is to pick those α-cuts that capture 75% of the village address points, and these are listed in Table 5.2. Having identified the optimum α-cut per village, it is vital then to determine the location of that α-cut to be used in the subsequent analysis. Figure 5.1 presents the general structure of the script that is being executed to identify the centre of the α-cuts – the raster surface (see Appendix 6.5 for the complete script).

Table 5.2: Lists of the considered settlements for the TSP with their equivalent α-cuts.

ID	Village Name	No. of Points	α-cut for 75%	Continuous Surface	Centre Position	Decision taken
0	Bagworth	24	0.4	✓	correct	-
1	Barlestone	30	0.2	×	ok	-
2	Barwell	221	0.3	×	ok	-
3	Botcheston	10	0.8	×	replaced	new α-cut = 0.5
4	Burbage	257	0.2	✓	correct	-
5	Dadlington	11	0.2	×	ok	-
6	Desford	95	0.3	×	outside	disregarded
7	Earl Shilton	227	0.1	×	ok	-
8	Fenny Drayton	14	0.4	×	ok	-
9	Groby	113	0.1	×	ok	-
10	Higham on the Hill	15	0.2	×	outside	disregarded
11	Kirkby Mallory	19	0.3	×	outside	disregarded
12	Market Bosworth	106	0.1	×	outside	disregarded
13	Markfield	173	0.4	×	ok	-
14	Nailstone	13	0.5	✓	correct	-
15	Newbold Verdon	70	0.1	×	outside	disregarded
16	Newtown Unthank	10	0.7	✓	correct	-
17	Osbaston	10	0.4	×	outside	disregarded
18	Peckleton	18	0.3	×	outside	disregarded
19	Ratby	60	0.3	✓	correct	-
20	Ratcliffe Culey	12	0.7	✓	correct	-
21	Sheepy Magna	14	0.5	✓	correct	-
22	Sibson	13	0.4	✓	correct	-
23	Stanton under Bardon	23	0.3	×	ok	-
24	Stapleton	26	0.4	×	replaced	new α-cut = 0.3
25	Stoke Golding	59	0.3	✓	correct	-
26	Sutton Cheney	14	0.5	×	ok	-
27	Thornton	27	0.4	✓	correct	-
28	Twycross	14	0.5	✓	correct	-
29	Witherley	13	0.4	✓	correct	-

```
#==========================================================================================
#                    Script Title:  Centroid extraction from alpha cuts
#
# 1. Install & Load the required libraries
        GISTools, spgwr, rgdal, raster & shapefiles
# 2. Set the working directory where the data exist and read in the data
#       a. Make a list of things to read in
#       b. in DOS window create a list of files and edit the file name
            to include the villages from POI data only
# 3. Loop through each settlement and calculate its centres
#       a. Read the file
#       b. Convert raster to points
#       c. Calculate the centroid for X & Y
#       d. Return the values of (x,y) & keep record of the village name
# 4. Write the results out in a csv table
#==========================================================================================
```

Figure 5.1: Pseudo-code for identifying the centre of the α-cut rasters.

5.3.2 Measuring the Network Distance

Initially, a distance measure is needed in order to apply the TSP. Here the road network distances between each pair of settlements' centres is considered rather than simple Euclidean distances. For this reason, a road data for the study area is required to generate an OD cost matrix (distance matrix). The road data used is the OS MasterMap – Integrated Network Layer (ITN), which is a vector data which comprises up-to-date information on roads in the UK downloaded under academic licence from the EDINA Digimap[3] collection. Given that, to generate the OD cost matrix (distance matrix), it essential to determine the supply (origin) and demand (destination) of some resources prior to running the analysis. Herein nodes representing the origin and destination, essentially identical, which are basically the location of settlement centres. By doing this, these nodes are included in the network dataset, which makes it now ready to run the network analysis. It should be clarified that the network analysis is applied twofold for both approaches. That generates a hard OD matrix (for the original centres) and a fuzzy OD matrix (for the new centres based on α-cuts). From these matrices, the line layers are exported in a readable format (shapefiles) to be used as the main inputs for applying the TSP in the next stage.

5.3.3 Solve the TSP

In this section, the TSP could serve as an implication of comparative analysis between hard and fuzzy approaches of identifying village locations. Recall that the TSP goal is to find the shortest tour that visits each city in a given list exactly once and then returns to the starting city. On this basis, the method of 2-Opt improvement heuristic algorithm is adopted to apply the TSP in this study, using the TSP package in R developed by Hahsler and Hornik (2015).

The purpose of this section analysis is to do more than merely point out some solutions of possible tours. It is rather to examine the inference of employing the TSP in different centre locations for every individual settlement. Thus, it could be wise to apply this analysis repeatedly for different subsets of villages. The smallest

[3] Road data obtained from http://edina.ac.uk/digimap

subset contains three villages to make a tour. Figure 5.2 illustrates the pseudo-code of the implementation of the TSP in hard and fuzzy locations, the full script is presented in Appendix (6.6).

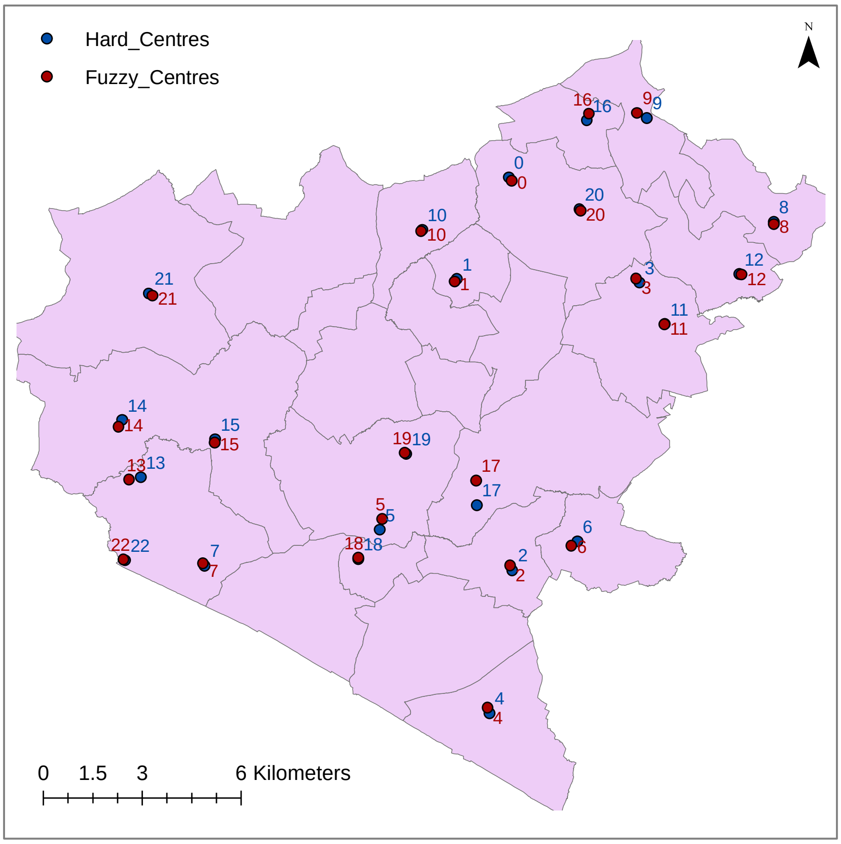
```
#==========================================================================================
#                Script Title:  Applying the Travelling Salesman Problem (TSP)
#
# 1. Install & Load the required libraries
       GISTools, shapefiles & TSP
# 2. Set the working directory where the data exist and read in the data
       csv files: 1. Hard OD Matrix
                  2. Fuzzy OD Matrix
# 3. Set up the OD cost matrices and rearrange their structure to be readable in the TSP
# 4. Analysis with all villages and compare between methods for TSP -(optional)
       "nearest_insertion", "farthest_insertion","cheapest_insertion",
       "arbitrary_insertion", "nn", "repetitive_nn", "2-opt"
# 5. Analyse the differences in the centre locations in  both approaches (hard and fuzzy centres)
#      a. calculate the distance between centre locations
#      b. save the result in table
# 6. Select subset from the data to apply TSP based on tge following functions:
#      a. generate subset from villages (3- 22) randomly
#      b. apply TSP for hard and fuzzy locations
#      c. save the tour path and length
#      d. plot the result, map showing the route with visited village
#      e. keep records of the  tour length, path
# 7. Write the results out in a csv table
#==========================================================================================
```

Figure 5.2: Pseudo-code for applying the travelling salesman problem on hard versus fuzzy location.

5.4 Results of the TSP

5.4.1 Centre Locations for TSP:

The first stage of the analysis was to identify a set of locations that represents the villages in order to be utilised in the TSP analysis. These locations have been identified in two approaches hard and fuzzy, as mapped in Figure 5.3. This map generally shows that for the majority of villages, the two centres are not much away from each other and in some cases they overlap (e.g. Groby, Naileston and Newtown Unthank).



ID	Name
0	Bagworth
1	Barlestone
2	Barwell
3	Botcheston
4	Burbage
5	Dadlington
6	Earl Shilton
7	Fenny Drayton
8	Groby
9	Markfield
10	Nailstone
11	Newtown Unthank
12	Ratby
13	Ratcliffe Culey
14	Sheepy Magna
15	Sibson
16	Stanton under Bardon
17	Stapleton
18	Stoke Golding
19	Sutton Cheney
20	Thornton
21	Twycross
22	Witherley

Figure 5.3: Map showing the centre positions in the hard and fuzzy approaches.

It should be pointed out that, in the fuzzy approach, there are some decisive factors that affect the process of defining the centre location. The intention was to select the centre of α-cut that covers 75% of the addresses for each settlement. However, in some villages the α-cut surface is not always continuous (i.e. non-continuous α-cut that has an area with zero membership in the middle) and may have two or more surfaces. This is primarily due to the nature or structure of the address points that represent the village, which have varying density. As a result, the centre may be placed outside the α-cut surface. The total number of considered settlements in this analysis is 30; over half of them have non-continuous surfaces and nine of those suffer from misplaced centres. Table 5.2 summarises these results, identifying the decisions taken to overcome this issue, one in which is to choose the next smaller α-cut that covers the location of its centre, as is the case in Botcheston and Stapleton (Figure 5.4). Nonetheless, seven villages – Desford, Higham on the Hill, Kirkby Mallory and four others – still do not account for this solution, as even the smallest available α-cut (0.1α) does not cover the centroid area (Figure 5.5). These villages are excluded from the analysis since there are barely enough number of eligible settlements to undertake the analysis.

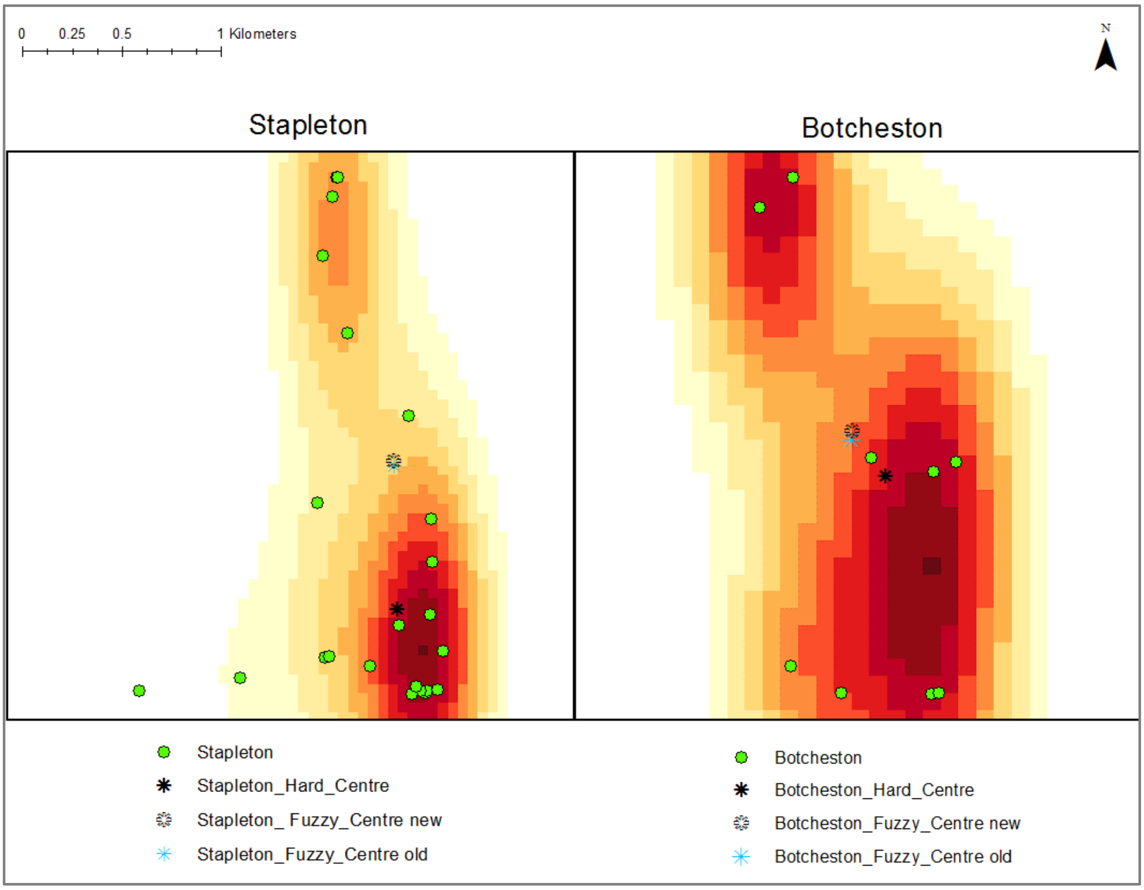


Figure 5.4: Maps of two cases when the fuzzy centres are replaced to smaller α.

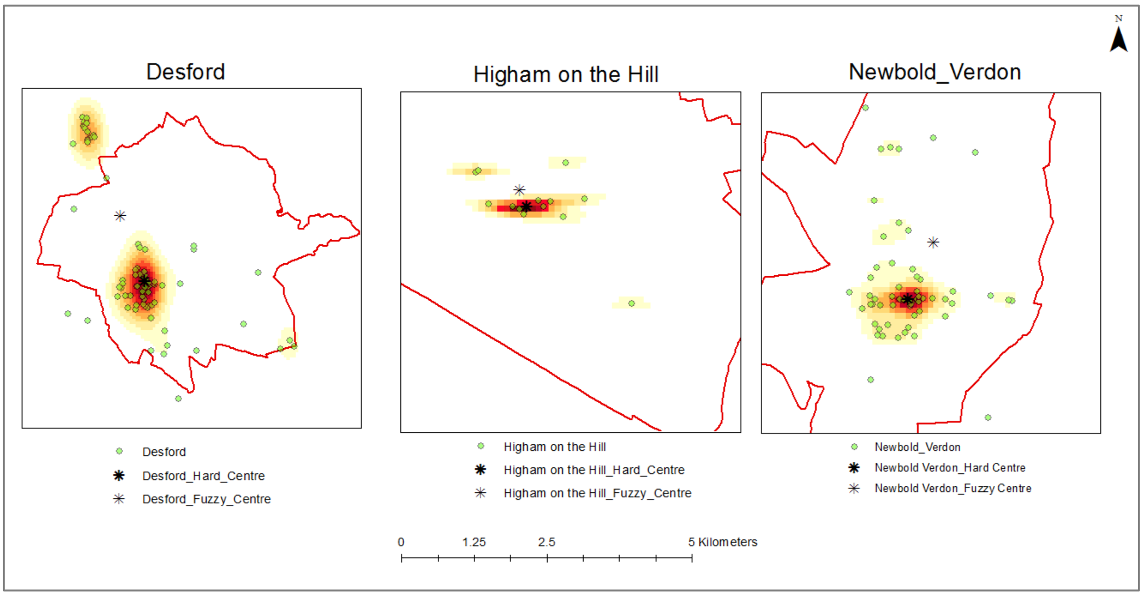


Figure 5.5: Maps showing examples of disregarded villages where the centre fall outside the α-cut surface.

Having decided in the fuzzy approach which centre should be adopted, Table 5.3 presents the distance in metres between the old and new centres for each examined settlement (hard and fuzzy approaches respectively). In fact, not many of them are even a quarter of a kilometre apart and the largest distance is about only three-quarters of a kilometre, which is in Stapleton. The reason for this slightly large distance can be traced back to the fact that the centre location is replaced in the fuzzy model due to the spatial pattern of this village, as depicted in Figure 5.4.

Table 5.3: Distance between the centre locations in the two approaches.

Old Centre		New Centre		Distance in Meters between Old & New
ID	*Name*	*ID*	*Name*	
0	Bagworth	0	Bagworth	137.81
1	Barlestone	1	Barlestone	110.83
2	Barwell	2	Barwell	159.64
3	Botcheston	3	Botcheston	158.96
4	Burbage	4	Burbage	176.98
5	Dadlington	5	Dadlington	338.43
6	Earl Shilton	6	Earl Shilton	219.54
7	Fenny Drayton	7	Fenny Drayton	88.77
8	Groby	8	Groby	75.00
9	Markfield	9	Markfield	335.68
10	Nailstone	10	Nailstone	49.60
11	Newtown Unthank	11	Newtown Unthank	13.95
12	Ratby	12	Ratby	71.03
13	Ratcliffe Culey	13	Ratcliffe Culey	357.04
14	Sheepy Magna	14	Sheepy Magna	232.58
15	Sibson	15	Sibson	101.49
16	Stanton under Bardon	16	Stanton under Bardon	212.92
17	Stapleton	17	Stapleton	746.85
18	Stoke Golding	18	Stoke Golding	37.10
19	Sutton Cheney	19	Sutton Cheney	58.57
20	Thornton	20	Thornton	54.37
21	Twycross	21	Twycross	126.13
22	Witherley	22	Witherley	60.53

5.4.2 Solving the TSP (Finding the shortest path – 2-opt)

It was decided here to adopt the k-Opt heuristic algorithm, particularly 2-Opt, as it is one of the improvement heuristic algorithms that may improve on the TSP results in both approaches. In this stage, the TSP analysis is applied repeatedly to some subsets of settlements to figure out how the differences in centre locations affect the TSP results in terms of the number of villages involved, the tour lengths and the path orders for all tours. Table 5.4 displays the distance of the tour lengths in kilometres for each subset. It has to be mentioned again that the subset ranges from 3 to 22 villages and these are selected randomly depending upon the average of all possible choices. The full details of the villages selected and the actual tours for each subset are presented in Appendix (5).

Table 5.4: Comparison of the tour lengths for the possible subsets of settlements in both approach.

Subset of	Tour Length in KM (Old centre)	Tour Length in KM (New centre)	Difference in Tour lengths
3	23.15	23.70	0.55
4	30.97	31.80	0.83
5	41.06	42.01	0.95
6	59.20	59.93	0.73
7	60.46	61.42	0.96
8	64.78	65.88	1.10
9	66.32	72.32	6.00
10	68.42	69.89	1.47
11	69.79	72.76	2.97
12	74.56	81.89	7.33
13	77.96	79.83	1.86
14	79.99	87.93	7.94
15	86.70	87.30	0.60
16	83.35	93.89	10.54
17	87.39	89.08	1.69
18	97.89	98.18	0.29
19	98.23	98.42	0.18
20	101.52	102.09	0.57
21	101.36	102.60	1.23
22	98.93	103.25	4.32

A general observation is that, although the distances between centre locations are slightly small, the results of the TSP are quite variant. It should be underlined that the fuzzy centres seem to have longer tours than the original centres. Moreover, even the order of the villages included in the routes is rather inconsistent. For the subsets of three to eight villages, the tours are comparable in the two approaches explicitly when the starting points are the same. This is also true for the subsets of 10, 11, 15 and 17 villages. By contrast, for the remaining subsets (9, 12, 13, 14, 16, 18, 19, 20, 21 & 22), the variations in the paths are apparent in more instances. This can be further demonstrated in a subset of 13 settlements, shown in Figure 5.6, in which the paths are different in going south from Twycross (coded as 21). This discrepancy could be attributed to the choice of the village to be visited next to Twycross downward, which is mainly influenced by the heuristic nature of the process. From Twycross, the path in the original centres goes to Sibson then to

Ratcliffe Culey (i.e. 21 → 15 → 13). Whereas the path in the new centres moves to Ratcliffe Culey instead then goes to Sibson (i.e. 21 → 13 → 15).

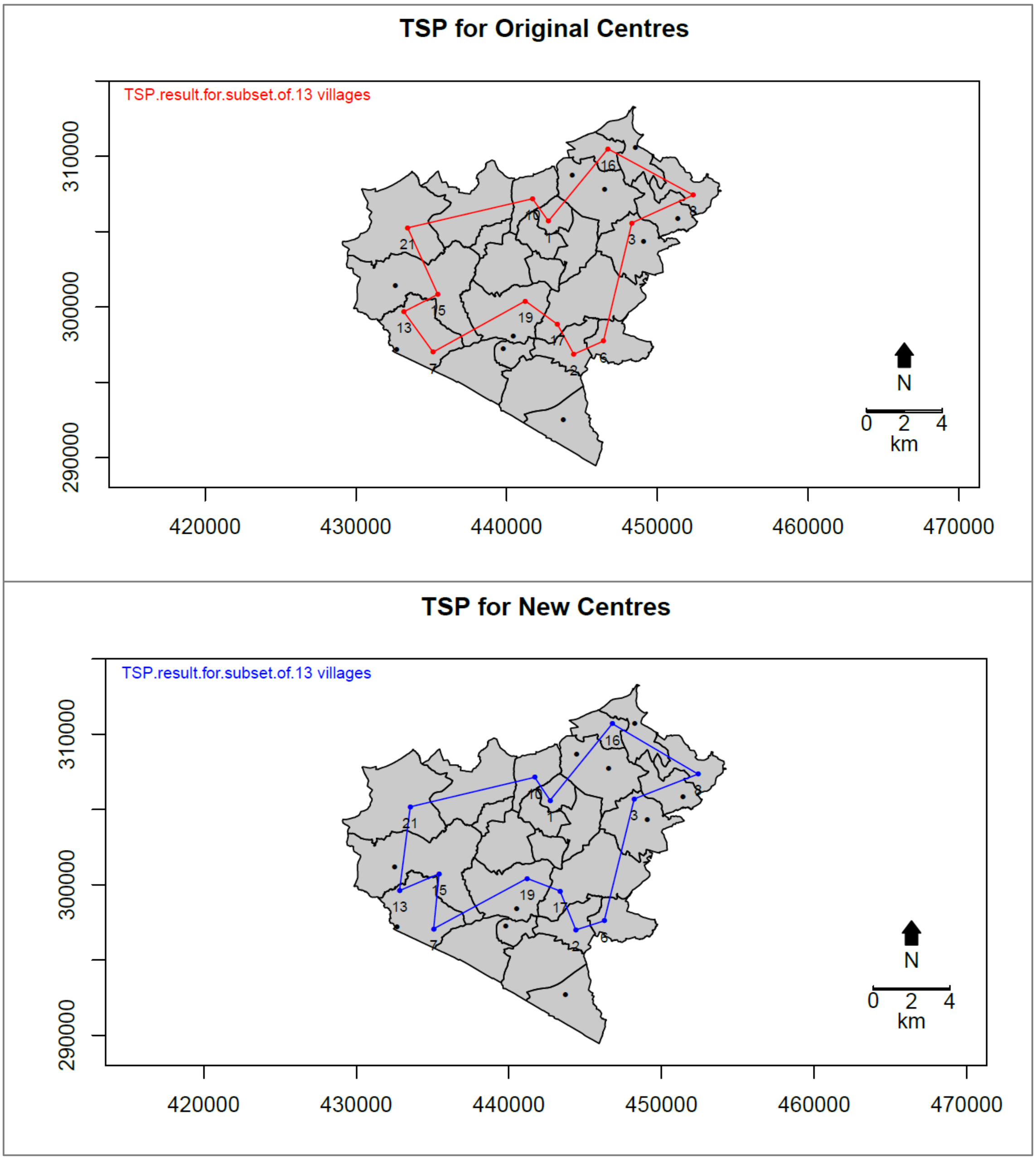


Figure 5.6: An example of the TSP results (tour paths) of a subset of 13 villages based in the hard (upper map) and fuzzy (lower map) locations.

In a similar case, in Figure 5.7 a subset of 18 villages presents the agreement in the two approaches only in the path going from Nailstone (coded as 10) to Sutton Cheney (coded as 19) through Burbage (coded as 4). The other direction of the paths connecting these two however are quite different. As for the original centres, the path goes southeast from Nailstone to Barlestone then goes up to Bagworth then Stanton under Bardon (i.e. 10 $\rightarrow$ 1 $\rightarrow$ 0 $\rightarrow$ 16). However, on the new centres, instead the path goes straight from Nailstone to Bagworth then Stanton under Bardon (i.e. 10 $\rightarrow$ 0 $\rightarrow$ 16). In addition, the path go directly from Botcheston to Sutton Cheney in the hard approach (i.e. 3 $\rightarrow$ 19), but it goes from Botcheston to Barleston then to Sutton Cheney in the fuzzy approach (i.e. 3 $\rightarrow$ 1 $\rightarrow$ 19), Besides, when 20 villages are selected to apply the TSP (Figure 5.8), only short parts of the routes are identical. These are the one connecting Botcheston, Newtown Unthank, Groby and Stanton under Bardon (i.e. 3 $\rightarrow$ 11 $\rightarrow$ 8 $\rightarrow$ 16); and the other which link between Ratcliffe Culey, Witherley and Fenny Drayton (i.e. 13 $\rightarrow$ 22 $\rightarrow$ 7).Whilst the rest of the routes are extremely varied in their paths among the original centres and the new centres.

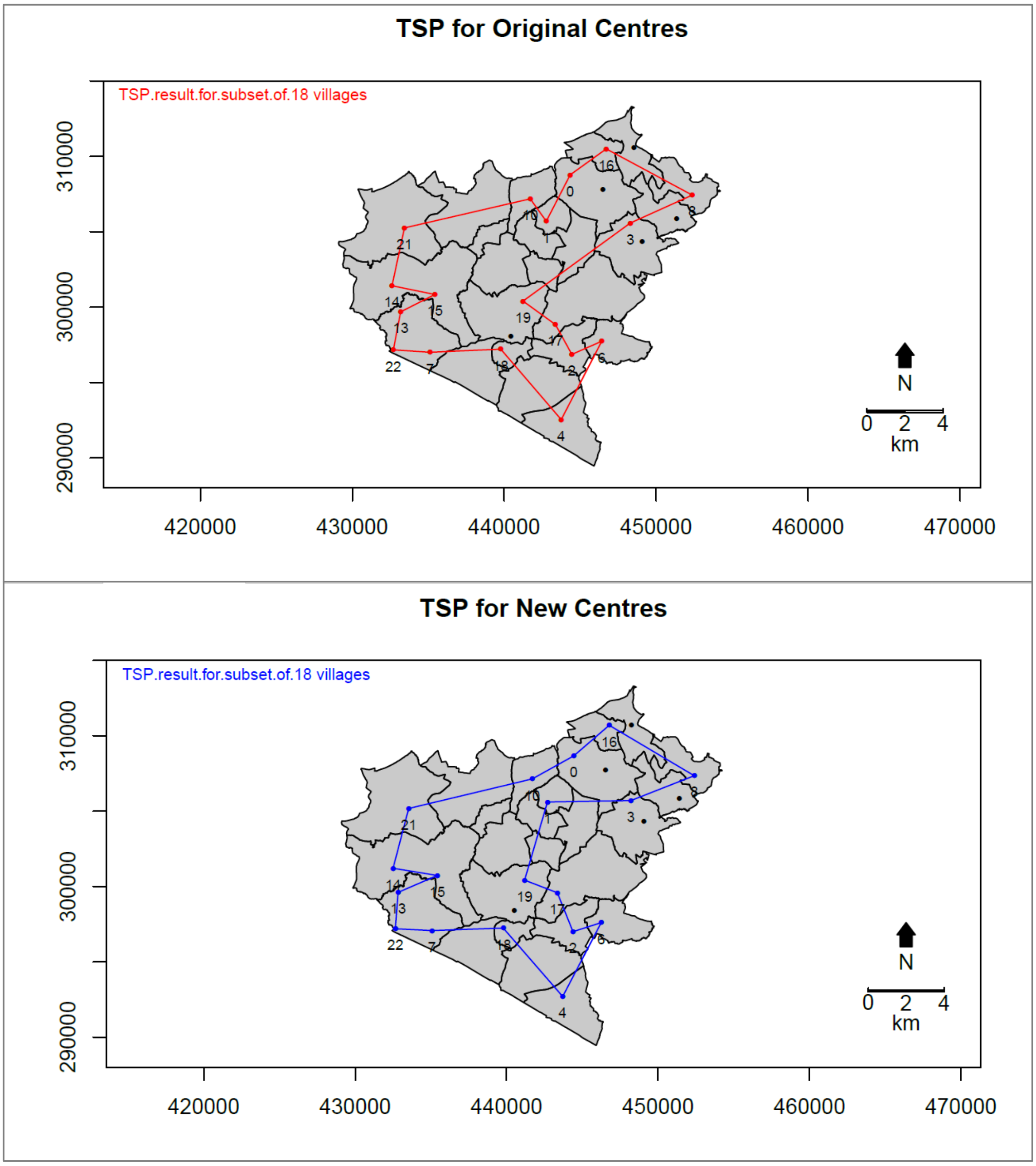


Figure 5.7: An example of the TSP results (tour paths) of a subset of 18 villages based in the hard (upper map) and fuzzy (lower map) locations.

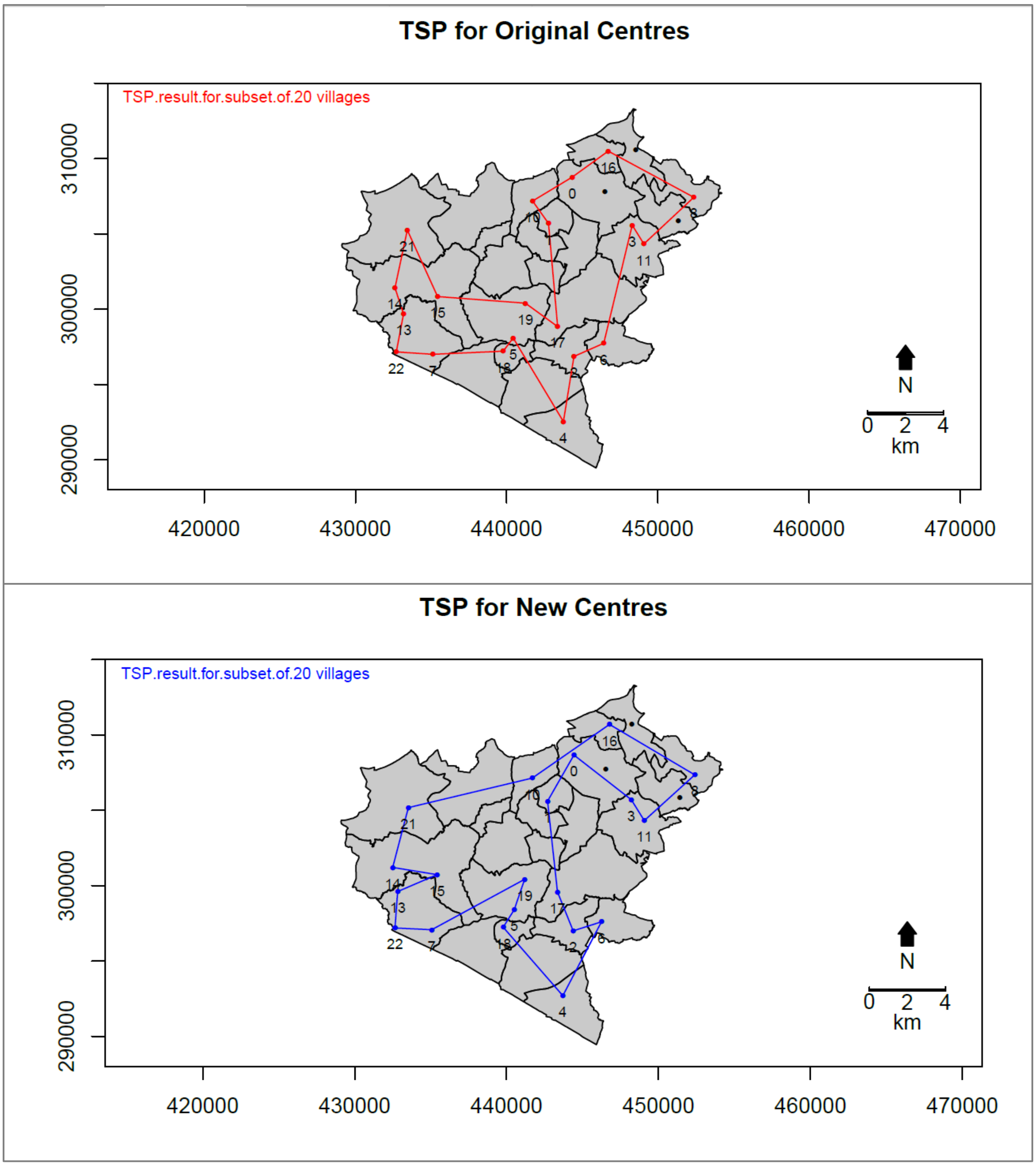


Figure 5.8: An example of the TSP results (tour paths) of a subset of 20 villages based in the hard (upper map) and fuzzy (lower map) locations.

5.5 Discussion

The work presented in this chapter is an implication of using the TSP on indeterminate or fuzzy locations (rural settlements in Hinckley and Bosworth). Although there is a substantial amount of research currently available that shows the integration between fuzzy set theory and TSP, there is a relatively small amount of research considering the fuzziness of the locations themselves. Even though this body of research (e.g. Botzheim *et al.* 2009; Fereidouni, 2011; Kumar and Gupta 2011 & 2012; Dhanasekar *et al.* 2013) is diverse in terms of the adapted methods, it all focuses on either developing a new algorithm or making a vital reconstruction of an existing one to address the fuzzy TSP. Looking from a slightly different angle, and building on the results from Chapter 4, modelling fuzzy objects, this chapter entirely focuses on application examples exploring the quantitative differences between using the TSP on specific and indeterminate (fuzzy) locations.

In constructing and applying the TSP analyses, some important simplifications and assumptions have been made that warrant discussion. First, among the available data identifying the rural settlements in this thesis, only the OS POI data are used in this chapter; exclusively the villages with at least 10 addresses. That is because POIs are originally contributed data, as opposed to the OS AddressPoints, which contain information from over 150 authoritative suppliers (Ordnance Survey 2014). This in turn makes the POI increasingly used in the research community as an important source of information, especially in the case of unavailability of formal infrastructure data (Burgoine and Harrison 2013; Neis and Zielstra 2014). Besides, POI features take another major role in less-developed countries, with the growth of outdoor tracking available to phones and other mobile devices, which allows contributors to share their geographic information on a number of selected online portals. Indeed, this incorporates to a new research platform, coined by Goodchild (2007): volunteered geographic information (VGI).

Second, it was recommended previously (from Chapter 4) that an optimal α-cut could be selected to represent the fuzzy model of a village, instead of considering the entire set of α-cuts, to conduct any subsequent analysis. Insofar as there is no single α-cut value that is recognised as best for all different settlements, it seems wise to opt for a mix of different α-cuts for each village as long as they capture 75% of the village address points (Table 5.2). One part of the analyses in this chapter (network analysis) is concerned with identifying centre locations, or centroids, as the locations to be visited and connected in the TSP. This is in line with other studies indicating that centroids are often used for GIS analyses to relate polygon-based objects or surfaces to linear networks (Comber *et al.* 2008-a). The results of this analysis in the fuzzy approach have confirmed the fact that the process of selecting the centre of an optimum α-cut is to a large extent affected by the shape or structure of the village (see Table 5.2 and Figure 5.4).

Third, it is generally expected that some places are much closer while others are further away. That could also be assumed when comparing the centre locations for the hard and fuzzy approaches. For some settlements, the exact centre and the fuzzy centre are much the same, while for others they are quite distinct. This is highly related to the shape and structure of the settlement patterns with varying density. For this reason, it was decided to consider the TSP analysis on a number of settlements that are selected randomly regardless the distances between their centres. As the observed differences between the centres in Table 5.3 are not significant.

The results of the TSP analyses (presented in Section 5.4.2) show different solution methods and heuristics to solve the TSP based on specific and fuzzy locations. There are two possible reasons for the differences in the results. The first one, as should be expected, stems from the varied nature of those heuristics, as different implementations to the same set of villages do not, however, always yield the same tours. This is in agreement with previous studies such as Johnson and McGeoch (2002) and Curtin *et al.* (2013). The second reason, less obvious but perhaps more appealing, goes back to how the small differences in centre locations affect the options made in constructing or improving the tour. This is not relevant just for the starting point of the tour, but more so for the village to be visited next.

The method applied in this chapter has a number of distinct advantages including: first, this approach employs the road network distance to measure the actual distances between settlements, rather than the implicit ones using the Euclidean or straight line distances. Second, this work acknowledges the uncertainty about the location of the settlements to be visited, simply by using the α-cut approach (optimal α-cuts). Third, the infrastructure of the analyses to handle and solve TSP in hard and fuzzy locations is based on standard GIS and statistical software widely available in the GIS community. However, one of the key constraints of this method is that the procedures of solving the TSP are accordingly heuristic and not guaranteed to determine the optimal solution. The method precisely adopts the approximation algorithms, which very likely provide a suboptimal solution for the TSP. Also, it has not been possible to identify the best solution using the exact algorithms, which could limit the validity of the comparative results of the hard and fuzzy approaches.

5.6 Summary and Further Research

This chapter has explored the implications of applying some heuristics to solve the travelling salesman problem in fixed and indeterminate or fuzzy locations. Although the differences between the exact and the fuzzy centres for the settlements are quite small, the resultant tours still varied in both the tour length and the actual routes. These findings are mainly attributed to the heuristic techniques in the first place, and that is ascribed to differences in the visited centres. It is hoped that the work presented in this chapter will encourage GIS researchers to at least acknowledge the uncertainty and vagueness inherent in the geographical phenomena and also their analyses.

This research has thrown up many questions in need of further investigation. It might be reasonable, for example, to bring into consideration the Concorde TSP solver (Applegate *et al.* 2006) or other advanced methods which efficiently compute exact solutions to validate the compression among the different heuristics in the traditional and fuzzy approaches. More broadly, it would be interesting to engage participants in the research, first to identify where they think the settlements centre locations, and, second, to explore their sense about the distance of the tours between the settlements.

Chapter 6 Discussion

6.1 Introduction

Many geographical and spatial phenomena are subject to vagueness, and perhaps to other kinds of uncertainty, for a number of reasons. Various models have been developed to address these issues, one of which is based on fuzzy set theory. Throughout this thesis, villages have been conceptualised as vague and so have been considered to be suitable for fuzzy models of uncertainty. This research has shown how sets of precise addresses, reporting rural settlements, recorded in different databases are geographically vague, both in their definition of any given village and in their description of its spatial extent. This research has sought therefore to assess the utility of fuzzy set theory in modelling and analysing such vague regions. Additionally, it has explored the application of other analytical approaches to describe and reason with vagueness measures generated from the fuzzy model. The potential impacts of spatial vagueness were illustrated through a comparison of the results generated using formal (crisp) village extents compared to using modelled fuzzy extents using from the Travelling Salesman Problem as applied to informal, contributed data (POI data). The methods and results provide a strong indication of how such an approach could be suitable in regions for which no formal infrastructure is available, such as developing countries.

This chapter brings together all of the previously discussed material. It explains how this material contributes to the field and provides a critical reflection of the thesis including its merits, assumptions and shortcomings. The next section provides an overall summary of results from the previous chapters. Section 6.3 reflects on the methods employed and their assumptions, and some possible modifications are suggested. Section 6.4 discusses the limitations and Section 6.5 suggests some areas for future studies. Finally, a summary is given in Section 6.6.

6.2 Summary of Results

With respect to modelling the fuzzy footprints of rural settlements, the results in Chapter 4 present a fuzzy representation (model) of villages derived from discrete address points as recorded in different data sources (Section 4.5.1). These were analysed to determine the degree to which individual points, nominally associated with a particular village, were contained within the crisp and fuzzy village models (Section 4.5.3). This research has shown that the vagueness associated with rural settlements can be adequately described by considering two types of vagueness they exhibit. The first concerns the locational vagueness and is associated with the settlement location and its extent (fuzzy boundary), and the second relates to vagueness in feature definition (fuzzy class). Section 3.6 in Chapter 3 discussed some evidential problems in the address point datasets and showed that villages are typically not represented as formal administrative boundaries and historically often have been defined by the parish boundaries. However, it is found that this is not always the case, since some settlements do not have equivalent parishes, or even if they do exist the address points belonging to these villages spill over the parished areas. Another indication of vagueness and imprecision is found in the extraction of address points that are located in a particular area but could have different definitions (settlement name) in different databases. This is generally true for many, if not most, places used for any application in GIS that would be represented differently in different systems.

This study applied spatial density estimation methods (KDEs) to quantify the villages' fuzzy entity and Boolean mappings of settlements using different α-cuts were presented (Figure 4.7 and Appendix 2). These fuzzy models can be viewed as a set of locations where each point has a membership in the range [0, 1]. The most obvious finding was that these results model the reality of rural areas, as a fuzzy geographical entity. Consider for instance the three villages Barwell, Desford and Stoke Golding (shown in miniature in Figure 4.7; and the same holds for all data sources and also any other villages identified in this thesis). Although a settlement may have an unequivocal parish boundary associated with it, the exact spatial extent of the village area is different and hard to define. This is because it is difficult to decide consistently where one definitely is within the village becomes in an area that is definitely outside that village. These results formalised this lack of definition in assigning membership grades to any location (or point) matching the definition of the village depending upon the density of houses (address points). In other words, the presence of houses indicates spatial extent (or extents – when there are two or more distinct clusters within the village) but moving towards the edge of that village the quantity of houses thins out (it does not remain at a high density) and eventually drops to nothing. This fuzziness is observed in the villages identified in this thesis from the different data sources (the BS76, POST and POI datasets) and can be seen in any towns or even cities around the world (Fisher and Wood 1998).

Another major benefit of fuzzy set theory is the concept of alpha-cut (α-cut). That relates to the second main group of results concerning the inclusion of a group of addresses that share the same village name within the fuzzy models. This shows that the spatial patterns of each rural settlement vary enormously and thus affect the fuzzy model. As been noticed in an exploratory trip (Appendix 1), some villages are clustered housing whilst others are a ribbon development along a main road. Others function as service centres for other neighbouring villages, such as main school, medical practice, and general store and post office. Also, many villages grew up around agriculture and forestry, others developed around fishing, and still others around industries such as mining. These different patterns typically indicate the varying density of houses within villages, which in turn influences the fuzzy model of the settlements. This study explored this effect by showing the degree to which a group of address points located in a village are within the fuzzy representation of that village. The results of these containment analyses were presented and discussed in Section 4.5.3 in Chapter 4. The overall conclusion about these results was that using α-cuts develops a series of alternative hard (Boolean or crisp) versions of the fuzzy set by varying the value of α. It is then possible to use a combination of different Boolean hardenings of the fuzzy model or, rather, select an optimum one for particular reasons. This study, unlike others in this respect (Arnot *et al.* 2004; Fonte and Lodwick, 2004; Fisher *et al.* 2004; Schmitz and Morris, 2006; Fisher, 2010), settled upon adopting those α-cuts that ensure a

75% level of inclusion for address points to apply the analysis of TSP. There are two reasons for this decision. The first is practical and arises from the fact that it was difficult to select a single α-cut that equally provides reasonable extent through all villages. The second is theoretical and implies the need for the salesman to visit the location that serves at least 75% of the houses in a village.

The results in Chapter 5 relate to exploration of the implication of using the fuzzy representations within the travelling salesman problem as a routing and navigation application. A comparison was made there between solving the TSP based on hard and fuzzy locations. These locations are considered to be in the centre of the villages and are assumed to be in position close to a number of houses. However, in comparing the hard and fuzzy centre locations for the considered villages, it was found that their locations in most villages are very close or almost identical, while only a few have slight distance between the two (see Figure 5.2 and Table 5.2 in Chapter 5). These results indicated that the distances between the hard and fuzzy centres, for all considered villages, are only a few metres. Not many of them are even a quarter of a kilometre apart, and the largest distance is only about three-quarters of a kilometre, which is in Stapleton. In fact, this is not surprising as one could expect this situation in rural areas where the villages are quite small and not very far from each other. That may make it difficult, sometimes, to recognise the shift from one village to another (Fisher and Wood, 1998).

Apparently, although the distances between centre locations are slightly small, the results of the TSP are quite varied (See Figures 5.6 to 5.8 and Appendix 5). These results showed that the fuzzy (new) centres seem to have longer tours than the hard (original) centres. Moreover, even the order of the villages included in the routes is rather inconsistent. These results generally found that, while the distances between the centres in the hard and fuzzy models are slightly small, the resulting TSP tours are varied. That is not to say the fuzzy approach is better than the hard one, nor would it be true in all circumstances, but simply that they are divergent approaches in the conceptualisation of the villages, which may be more valid in acknowledging their reality. Furthermore, these results could have been influenced by the scale factor of rural areas, as it is anticipated that other vague regions in a larger scale would provide more variation between the two approaches.

6.3 Reflection on Methods

6.3.1 Summary of the Methods

The approach used in this thesis was to approximate the spatial extents (fuzzy footprints) of rural settlements based on normalised kernel density estimation of addresses, as illustrated in Chapter 4 (Section 4.3.1). This technique generates a probability surface, and it has been used widely in modelling vague regions (Montello *et al.* 2003; Jones *et al.* 2008; Twaroch *et al.* 2008 a & b ; Hollenstein and Purves 2010; de Berg *et al.* 2011). In this thesis, it was used under fuzzy set theory. The assumption was that a settlement name recorded in any address databases is more likely giving reference to a typical member or good part of the settlement (i.e. where it is located). Density surface modelling methods can then be used to identify regions corresponding to the most frequently co-occurring addresses to define the extent of the settlement as a vague place. This is partially inspired by similar studies on identifying settlement extents. For example, Chaudhry in his thesis also (2007) defines 'citiness' from a prototypical and functional point of view based on the area and density of buildings.

This technique has again raised the question of the relationship between fuzzy sets and probability, which has been, and continues to be, an object of controversy (Zadeh, 1995 & 2002). Models of probability and fuzzy membership are both used to model uncertainty in spatial phenomena (Fisher, 1994), and both allocate modelled objects a value between 0 and 1. However, as argued by Fisher (1994), in determining the visibility from a position in the landscape, a viewshed operation, the primary distinction is mainly conceptual. That is to say, the line-of-sight to a location may be uninterrupted, making the location visible (an issue of probability), but it may not be possible to clearly discern an object which may be at that location (fuzziness). This means that a village which may have a precise (crisp) boundary at any given location within it or not is considered an issue of probability (i.e. the accuracy to be either inside or outside the village). However, it may not be possible to clearly discern the degree to which an object may be at that location, which represents an issue of fuzziness (i.e. relates to vagueness that the object is definitely outside the village and occasionally within it). This controversial view is not restricted to geographers, as Zadeh (2003) stated "Fuzzy logic is probability theory in disguise" (page 1) and "…that most of the information relevant to probabilistic analysis is intrinsically imprecise, and that there is imprecision and fuzziness not only in probabilities, but also in events, relations and properties such as independence"(page 143). It is therefore possible to adopt this approach to model vagueness in rural settlements or, more precisely, the fuzzy footprint for villages.

At the heart of fuzzy modelling is the fuzzy membership and its basic related concept of generating alpha cuts. This has been used in combination with the point-in-polygon strategy to analyse the spatial pattern of villages in relation to the inclusion of address points within a candidate settlement, as demonstrated in Chapter 4 (Section 4.3.2). This is similar to the idea of generating crisp approximations of the boundary of the vague region at different levels of confidence if they are required, as suggested by Jones *et al.* (2008). These methods have shown potential in providing information and help in selecting an α-cut level to be employed in any subsequent analysis. For example, in Chapter 5, Section 5.3 has contributed an implication of using the TSP on indeterminate or fuzzy locations at a particular confidence level . On that point, it was recommended to consider the α-cuts that guarantee 75% of the address points to be included in a single village (see Table 5.2 for details). The reason for doing this is because no single α-cut value was recognised as best for all different settlements. It was possible therefore to consider the 75% a valid choice to implement the TSP. Other options for capturing different percentages of the points are also possible. However, they probably will not provide much difference in the results because the spatial patterns of the points in any village are consistent and only the proportions of the points captured decrease with the increase of membership values. In addition to that, since α-cuts are always nested, then the centres' positions will not change dramatically.

As mentioned before, the integration between fuzzy set theory and TSP is not new and has been suggested before by other research (e.g. Botzheim *et al.* 2009; Fereidouni, 2011; Kumar and Gupta 2011 & 2012; Dhanasekar *et al.* 2013). These studies differ in terms of the adapted methods, but they are all similar in focusing on either developing a new algorithm or making vital reconstruction to an existing one to address the fuzzy TSP. This study, however, did not consider the combined methods as suggested by these researchers; it was rather aimed at comparing the performance of applying the traditional method of travelling salesman problem on hard and fuzzy locations.

6.3.2 Possible Alternative Approach

An alternative method for modelling the fuzzy footprint for rural settlements could be applied based on interpolating distance (Almadani *et al.* 2014). The main idea is inspired by the conceptualisation of fuzzy c-mean classification. That does classify a dataset into groups (clusters) such that the points that belong to the same group are more similar than the points belonging to different groups. Besides, it is fuzzy because it is a generalisation of a hard clustering partitioning method, which makes a clear-cut decision for each object (i.e. each object of the dataset is assigned to one and only one cluster). In contrast, a fuzzy clustering method allows for some ambiguity in the data, which often occurs. So, each object is spread out over the various clusters by means of degree of belonging, which is quantified as membership coefficients that range from 0 to 1 (Kaufman and Rousseeuw 1990). As it is assumed that a set of address points that share a village name represent one single cluster, it is then possible to consider the centroid as the core area of that village. So from the address points in a village the centroid is identified, and for every house the distance away from the village centre is calculated. Interpolation from the address points extends the memberships to a continuous surface over the entire area within that village. Following Hall *et al.* (2011), ordinary kriging is used to transform these point measurements into the continuous field representation. Again, normalisation is necessary to transform values to fuzzy memberships. This approach might be promising in regard to representing how fuzzy memberships are used to represent fuzziness in vague regions. However, specifically in this research, it comes out that due to the spatial distribution of the address points two main shortcomings are explicit: extrapolating and artefact (Almadani *et al.* 2014).

There are other more mathematical and computational methods that could provide a good representation of the vague region based on graph theory. An example of this is the work of de Berg *et al.* (2011), which proposes a method to delineate imprecise regions from a set of points using shortest-path graphs based on the squared Euclidean distance. Their work is successful in dealing with regions containing holes. Other approaches consider a geometric notion of 'shape' (or, more precisely, shape generation algorithms), such as the work of Reinbacher *et al.* (2008), which suggests a method based on α-shapes to determine a reasonable boundary for an imprecise region. Additionally, there is the study of Duckham *et al.* (2008), which argues for the potential applications of the characteristic shapes or simply χ-shapes in generation of geographic "footprints" for vague and imprecise spatial concepts. Although these approaches may be applicable to the problem at hand, they have not been considered in this research due to the high computational demands required compared to the relatively intuitive and straightforward (conceptually elegant) approach suggested.

6.4 Limitations

Although several key findings have emerged from the research, these are subject to a number of limitations. Initially, the issue of ambiguity is probably the biggest limitation inherent in spatial datasets. This prominently occurs when there is doubt as to how a phenomenon should be classified because of differing perceptions of it (Fisher 1999). This typifies the situation existing here in which most villages are indicated either in different ways or by different names (see Section 3.6 in Chapter 3). This indeed extends to the description of villages and rural places in other informal data such as tweets, geotags to Flickr and so forth, given that people normally have different perceptions about a place and cannot always agree on a given precise description about it either.

This leads to a limitation: the work described in this study did not consider modelling areas with multiple membership, although there is evidences that they exist. The fuzzy model was based on using the data separately in order to get a discrete view about the village representation in each dataset (Section 4.5.1 in Chapter 4). This suggests the need for further research to develop modelling techniques that address the issue of multiple membership.

The implementation of the TSP (in Chapter 5, Section 5.4) is not free of problems. Probably the most important one was that this study was not able to compare all exact solutions for the TSP for the entire settlements in the two approaches, since this requires an advanced tool under different licence and with particular computational speed.

6.5 Future Research

In reviewing the relevant literature, the issue of identifying the geographical distance between two places modelled as fuzzy objects was raised (Almadani *et al.* 2012). This is to argue that, within a Boolean concept of space, the distance between two locations is simple and well understood (can be found with the Pythagorean equation), but if the model of the objects is changed from Boolean to fuzzy then the problem of identifying the distance between them is vastly complicated. The issue of fuzzy distance is not particularly new (see, for example, Rosenfeld, 1985; Altman, 1994; Voxman, 1998; Bloch, 1999; and Guha and Chakrborty, 2010). Most previous research extends the concept of distance to subsets of a metric space, and argues for the representation of fuzzy distance as a fuzzy number. These studies tend to suggest many potential applications for different areas, including pattern recognition, image processing, robotics, computer graphics and engineering. However, few if any make use of geographical distance or link fuzzy distance to real world phenomena, although Guesgen and Hertzberg (2001), Guesgen *et al.* (2003) and Fisher and Almadani (2010) have all looked at the related topic of fuzzy buffering.

There are other possible directions for future work based on the findings and limitations discussed throughout this thesis. Further study could look into the applicability of this method in other places. It is expected that for some countries such formal data are not an issue and similar results could be obtained. However, this is not the case in many other places with emergent spatial data infrastructures. It would be possible, then, to adapt the method to fully exploit the potentials of other crowdsourced data (e.g. Twitter, Flickr, and so on) to generate fuzzy extents of different geographical regions.

A variety of alternative methods are currently available for modelling vague regions (Section 6.3.2) which aim to address the issue of vagueness, but do not consider fuzzy set theory. A direct extension then is to explore possible integration between these approaches and fuzzy set theory.

The remaining challenge involves the development of other novel analyses of geographical information based on fuzzy representation of geographical phenomena. Buffer operations, for example, are typically used in site selection procedures; Fisher (2009) shows the advantage of using fuzzy buffers around Boolean entities in site selection. One challenge for the approach outlined in this thesis should therefore concentrate on the investigation of possible implementation of a fuzzy buffer method as applied in methods such as site selection analysis, location allocation or perhaps gravity model application.

Chapter 7 Conclusion

7.1 General Summary of research contribution

This research has demonstrated the way in which fuzzy set theory can be used to derive approximate boundaries (fuzzy spatial extents - footprints) for vague regions such as villages. These were approximated from named settlements recorded in different address databases. It has further discussed how such regions are typically not represented as formal administrative boundaries and are often considered to be vague and imprecise. The methods introduced evaluate the usefulness of fuzzy set theory in modelling and analysing such vague regions. It has further explored the implications of applying the Travelling Salesman Problem using informal, contributed data (POI data) in formal, crisp village extents versus the modelled fuzzy extents. In short, the results imply that the fuzzy model is more efficient than the traditional Boolean, crisp model of approximating the spatial extent of rural areas. However, the TSP results showed that to a large extent longer tours were found in the fuzzy model than the traditional crisp model. This was mainly affected by the scale factor of rural areas, considering the relatively small distances between centres' locations in the two approaches.

7.2 Overall Research Outlook

It is indicated that further work needs to incorporate other novel analyses of geographical information based on fuzzy representation of other geographical phenomena with varied attributes in terms of scale or location (e.g. urban areas in other countries). Overall, this study strengthens the idea that vagueness and uncertainty in general are fundamental to geographical phenomena. It is therefore necessary to acknowledge these issues in geographical databases and analyses. The subsequent question then becomes, how do such problem with vague regions, specifically informally named in places where no infrastructure data available or under severe restriction for worldwide application, be solved. This in turn highlight the valuable role of the general public who act voluntarily to create a global patchwork of geographic information sources.

Appendices

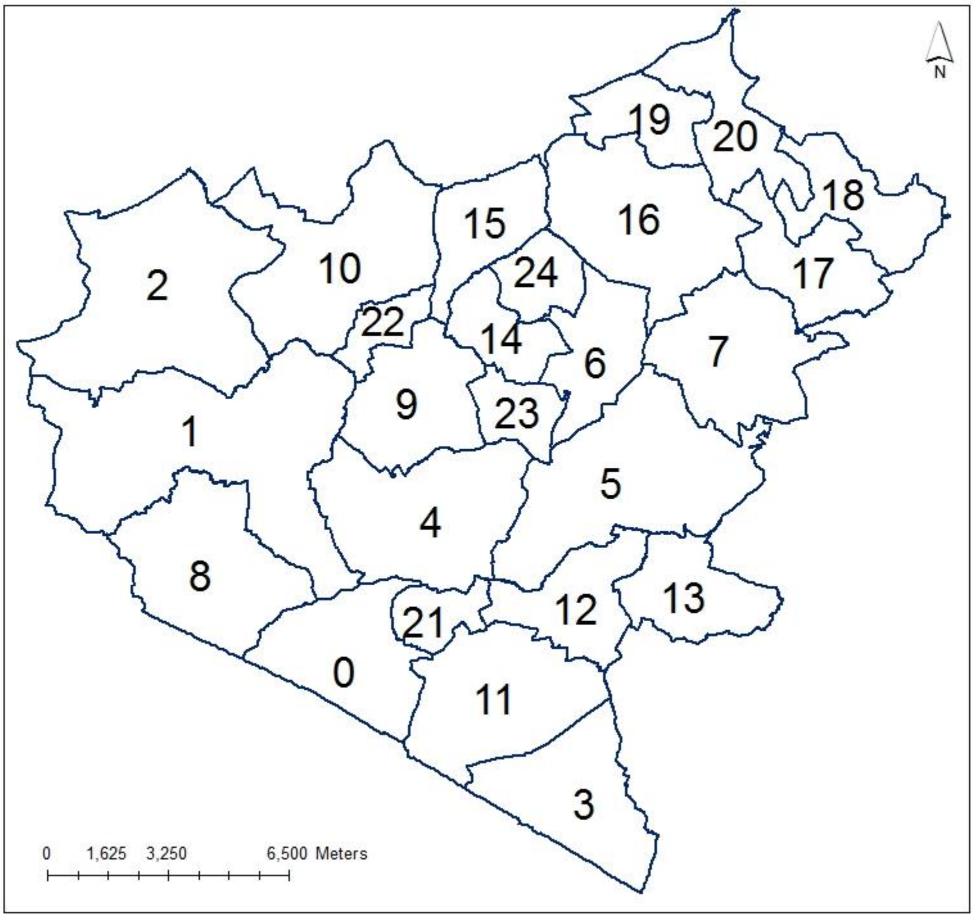
Appendix (1): Field trip agenda September 18, 2013

Some notes

- Some of the villages which have equivalent parish boundaries and either
  - Their extents are limited with the boundaries [Twycross, Peckleton, Osbaston, Nailstone, Cadeby,], or
  - Exceed the parish boundaries [Desford, Earl Shilton, Stanton-under-Bardon, and Stoke Golding].
- Villages that exist only in one address types:
  - BS7666 only
  - Postal only
  - Villages in common but different distribution// can't be seen on the field just looking at the maps///
- Compound names and punctuation
  - Bagworth and Thornotn
  - "Higham on the Hill" OR "Stanton under Bardon"
- Missing places from my data

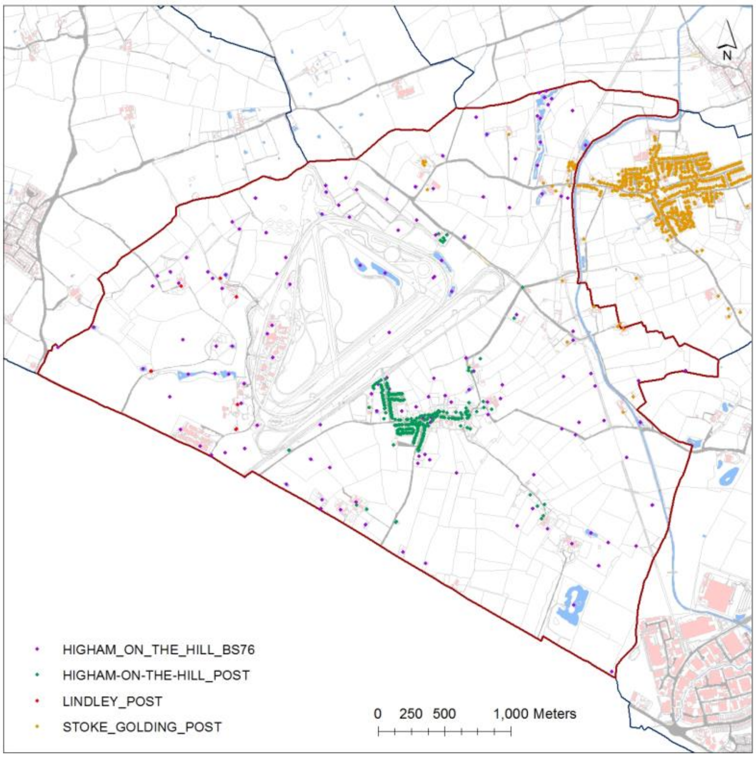
  There are some places missing from my data, which are listed below:
  - Brascote – [Newbold verdon]
  - Bull in Oak – [Cadeby]
  - Coton– [Market Bosworth]
  - Far Coton – [Market Bosworth]
  - Field Head – [Groby]
  - Hinckley – [Hinckley]
  - Hollycroft – [Hinckley]
  - Little Orton – [Twycross]
  - Little Twycross – [Twycross]
  - Merry Lees – [Bagworth]
  - Peckleton Ind Est [4]– [Desford]
  - Sketchley –[Hinckley or Burbage]

[4] May be stands for "Industrial Estate"

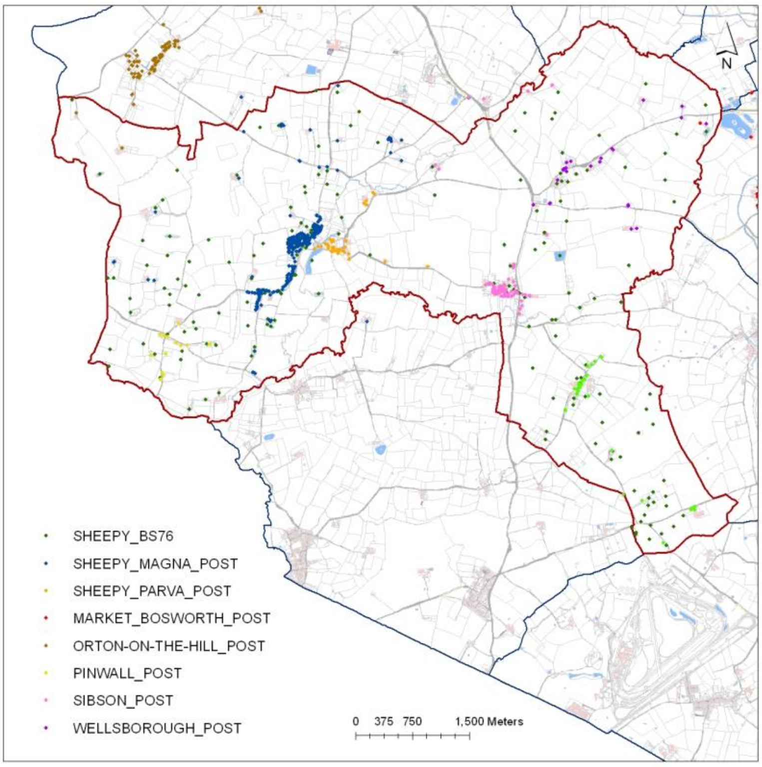


#	Parish Name	#	Parish Name
0	Higham on the Hill CP	12	Barwell CP
1	Sheepy CP	13	Earl Shilton CP
2	Twycross CP	14	Osbaston CP
3	Burbage CP	15	Nailstone CP
4	Sutton Cheney CP	16	Bagworth & Thornton CP
5	Peckleton CP	17	Ratby CP
6	Newbold Verdon CP	18	Groby CP
7	Desford CP	19	Stanton-under-Bardon CP
8	Witherley CP	20	Markfield CP
9	Market Bosworth CP	21	Stoke Golding CP
10	Shackerstone CP	22	Carlton CP
11	NCP * (Hinckley)	23	Cadeby CP
		24	Barlestone CP

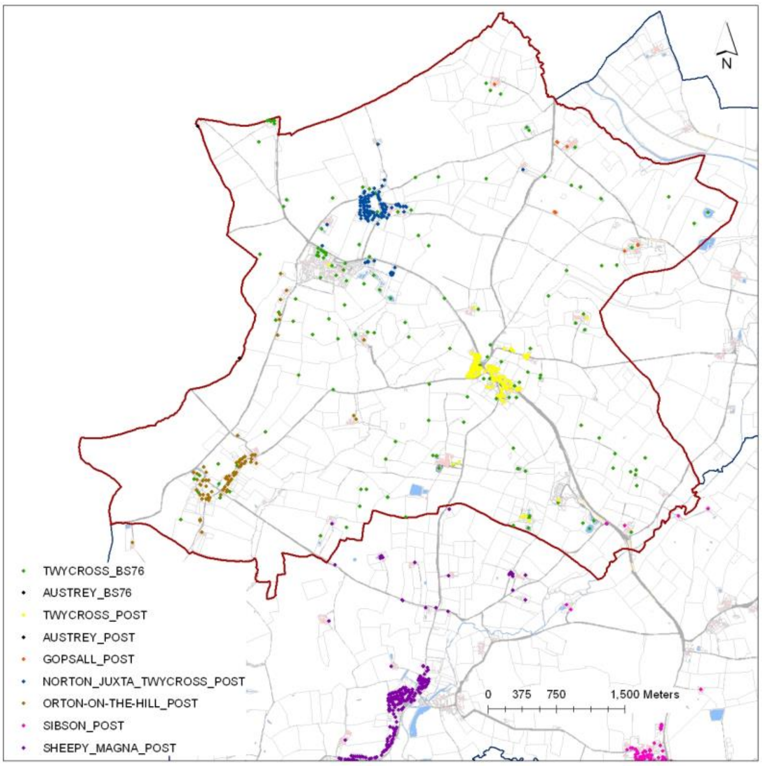
Higham on the Hill [no. 0]



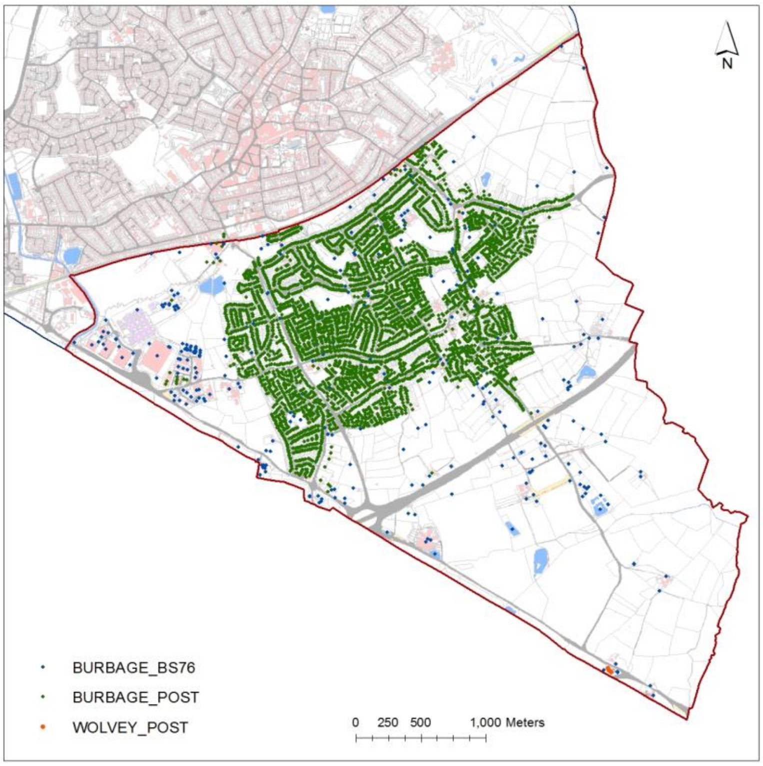
1- Sheepy [no.1]



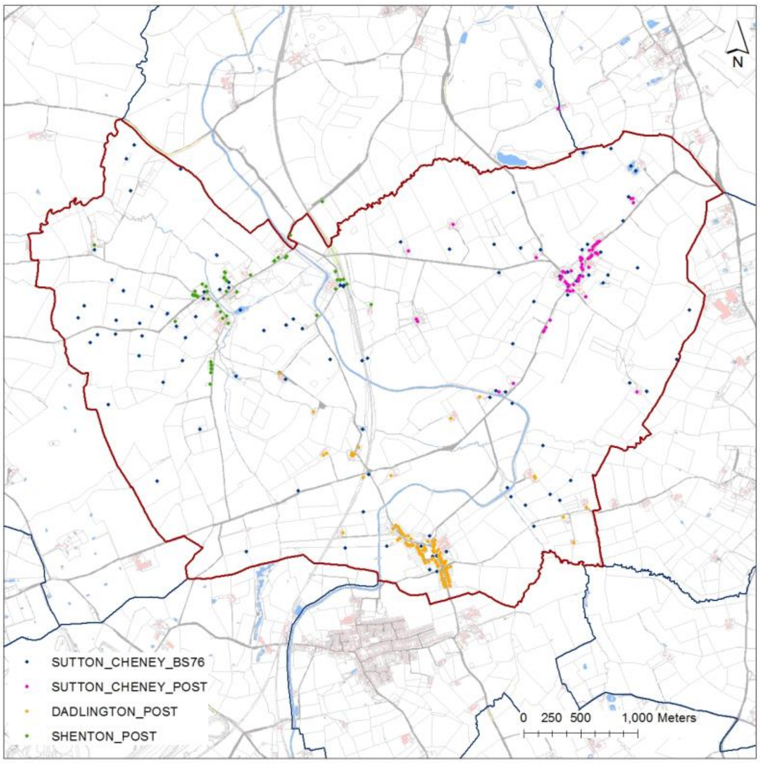
2- Twycross[no. 2]



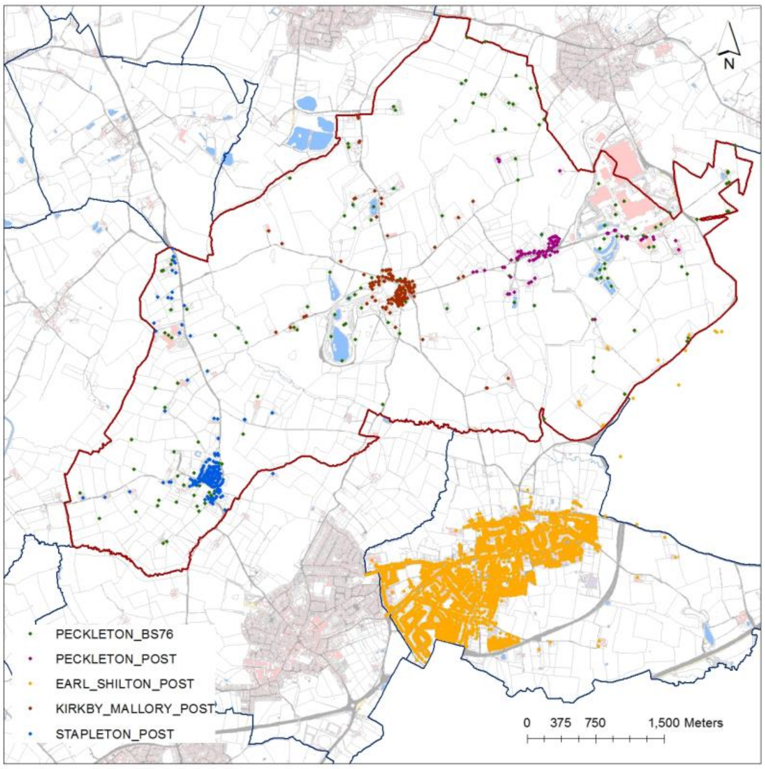
3- Burbage [no.3]



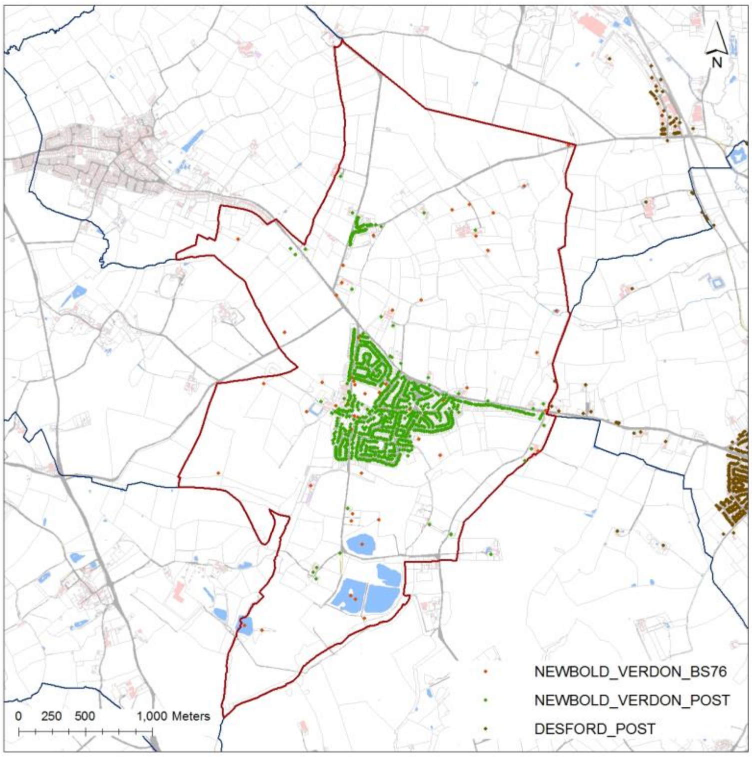
4- Sutton Cheney [no.4]



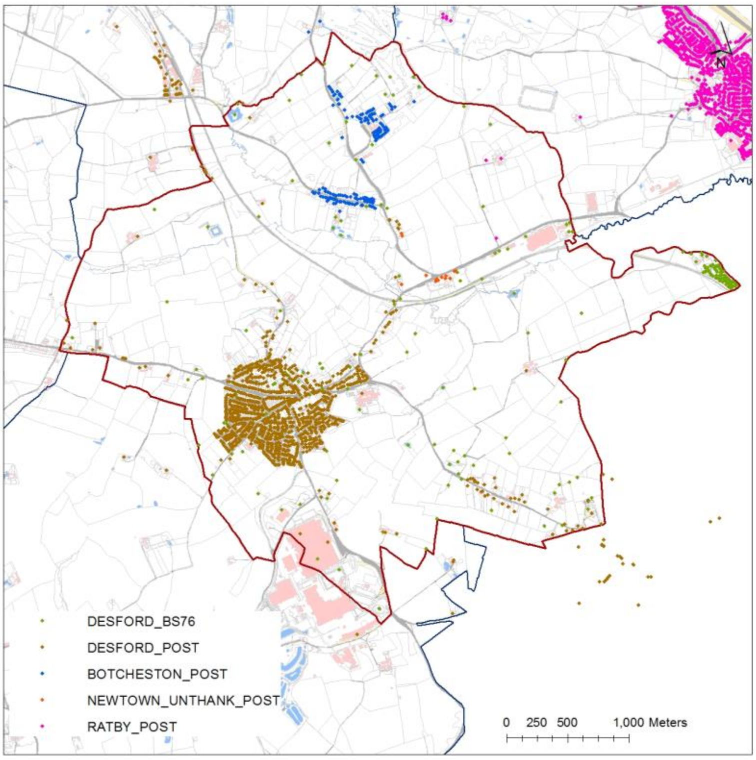
5- Peckleton [no.5]



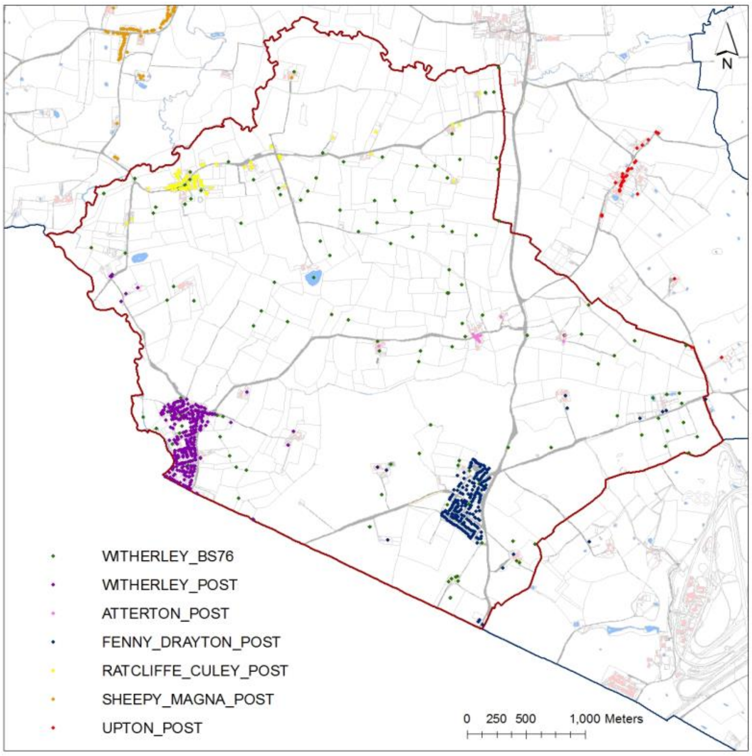
6- Newbold Verdon [no.6]



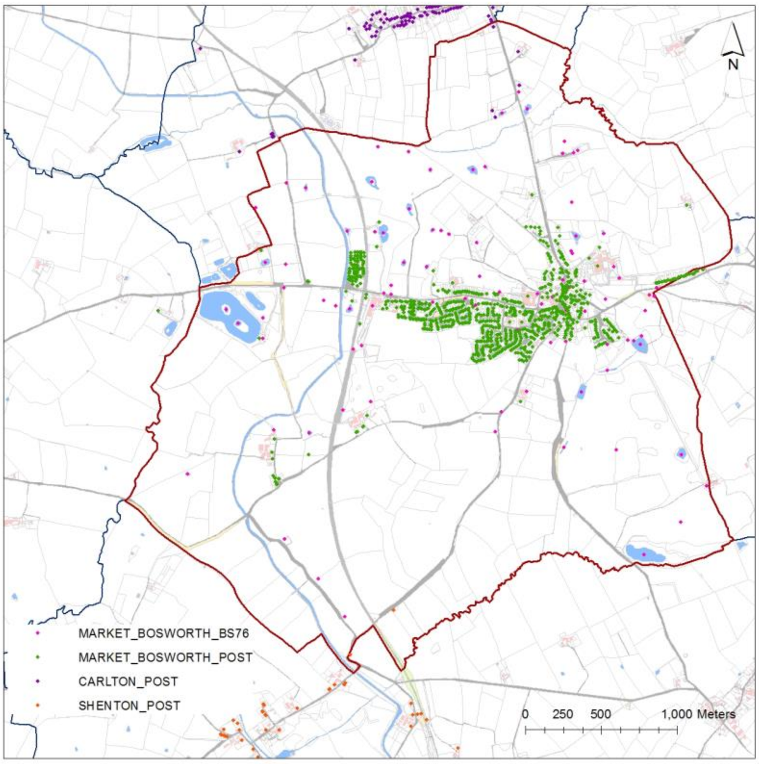
7- Desford [no.7]



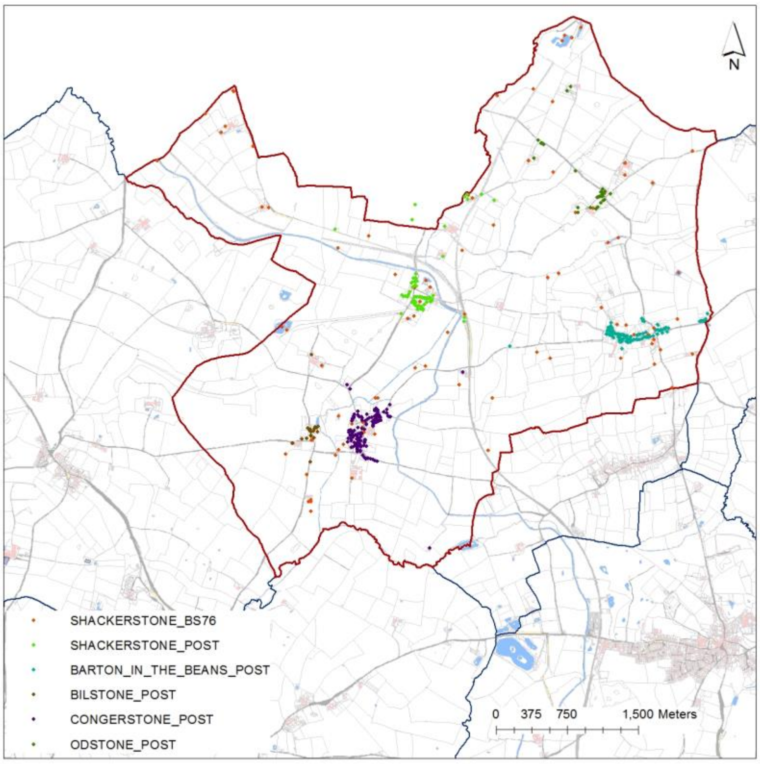
8- Witherley [no.8]



9- Market Bosworth [no.9]

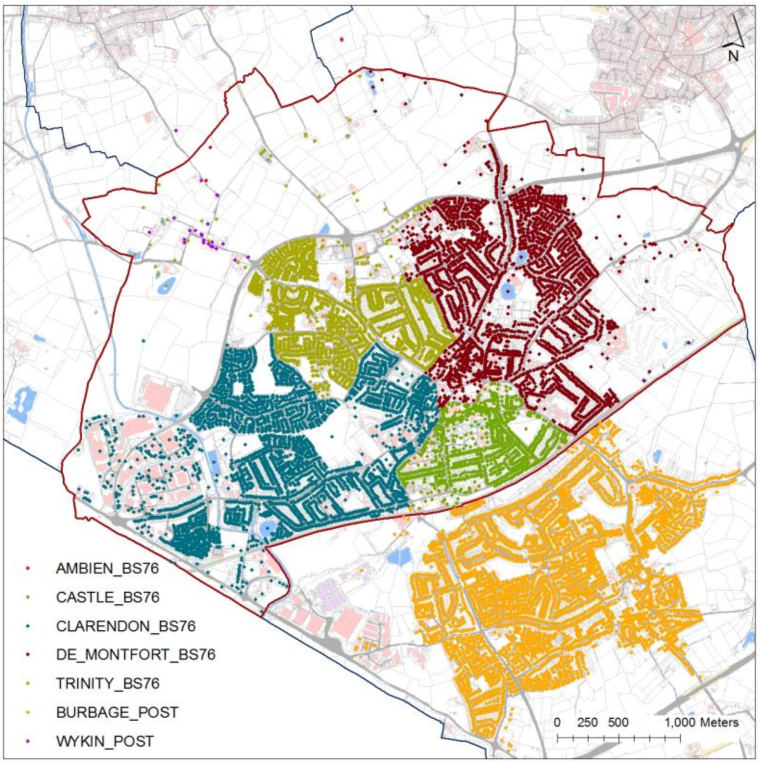


10- Shackerston [no.10]

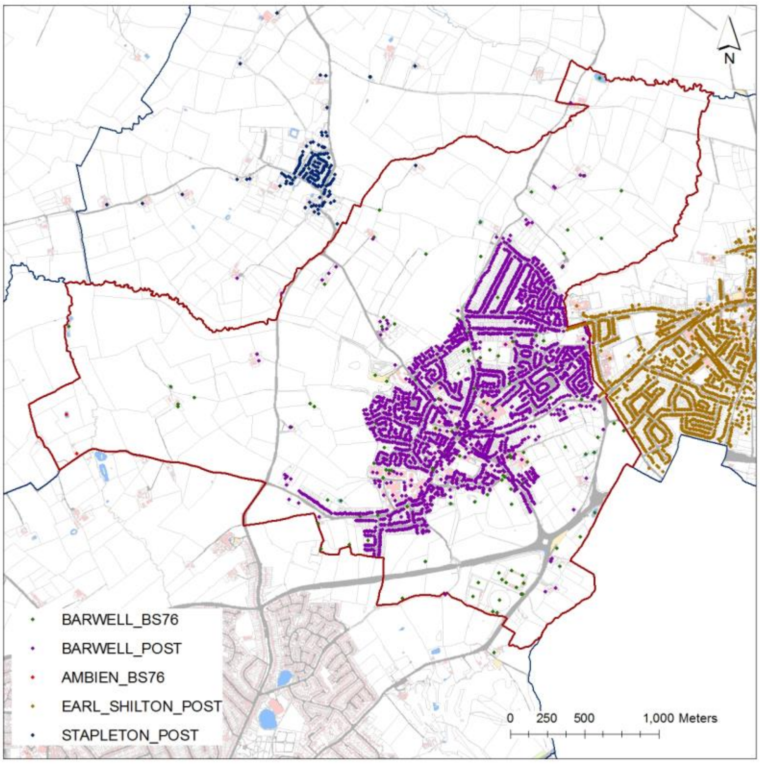


11- ....NCP [no.11]

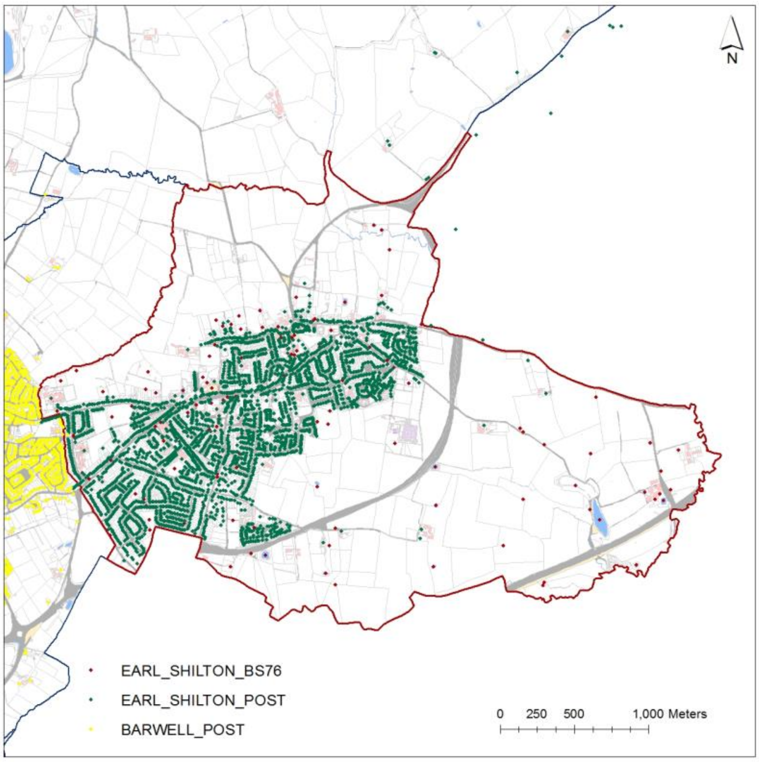
This parish is Non-Civil Parish or Community; that is way it doesn't have a name.



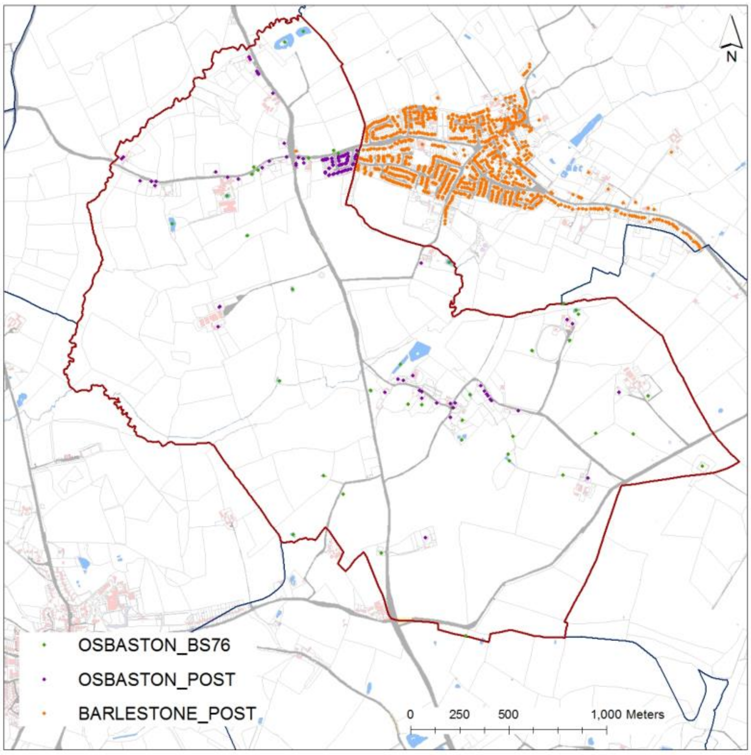
12- Barwell [no.12]



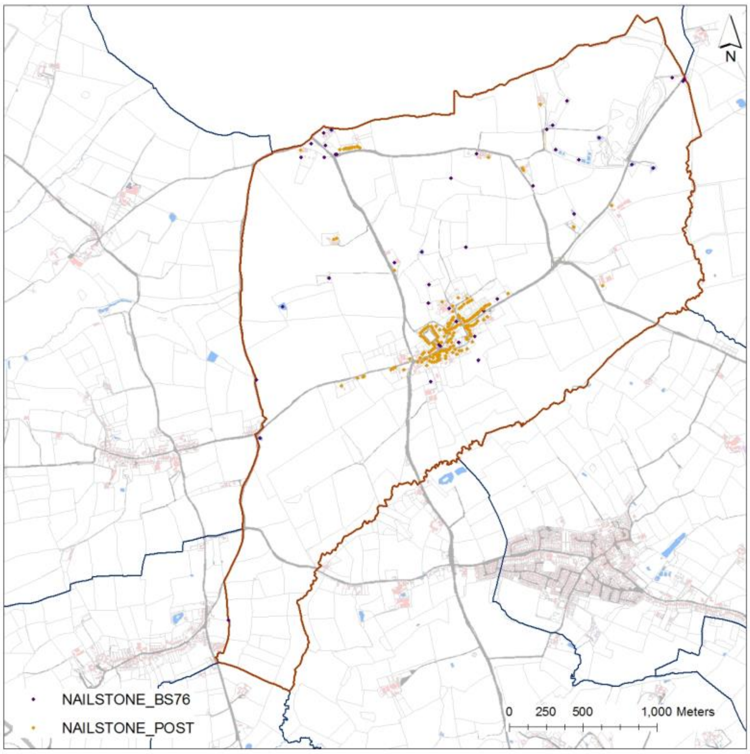
13- Earl Shilton [no.13]



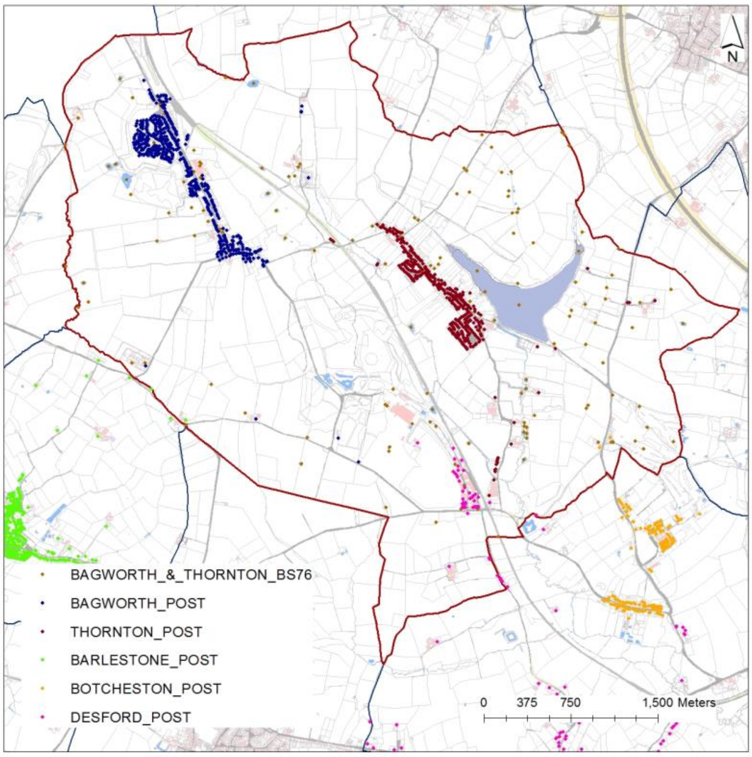
14- Osbaston [no.14]



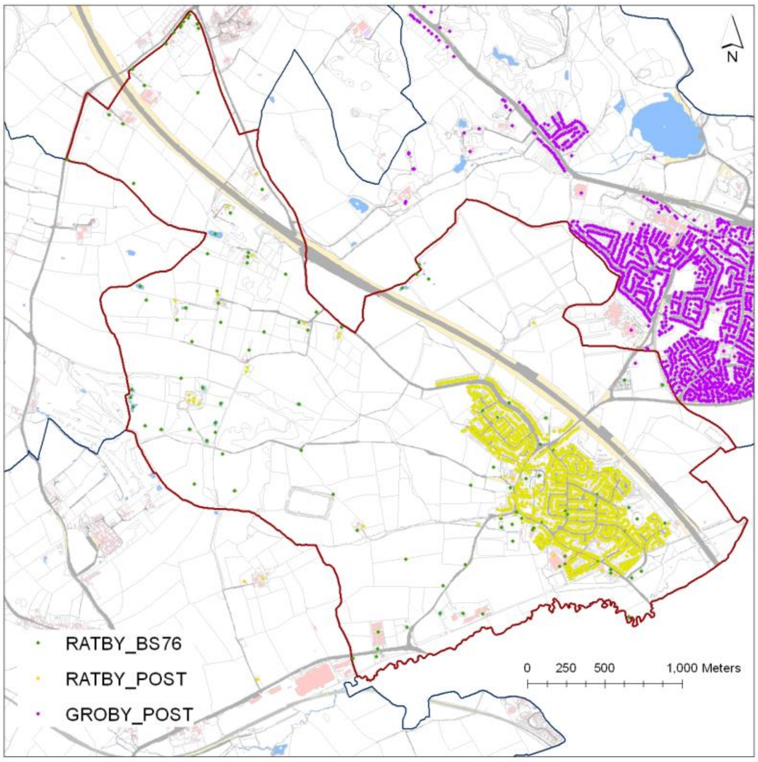
15- Nailstone [no.15]



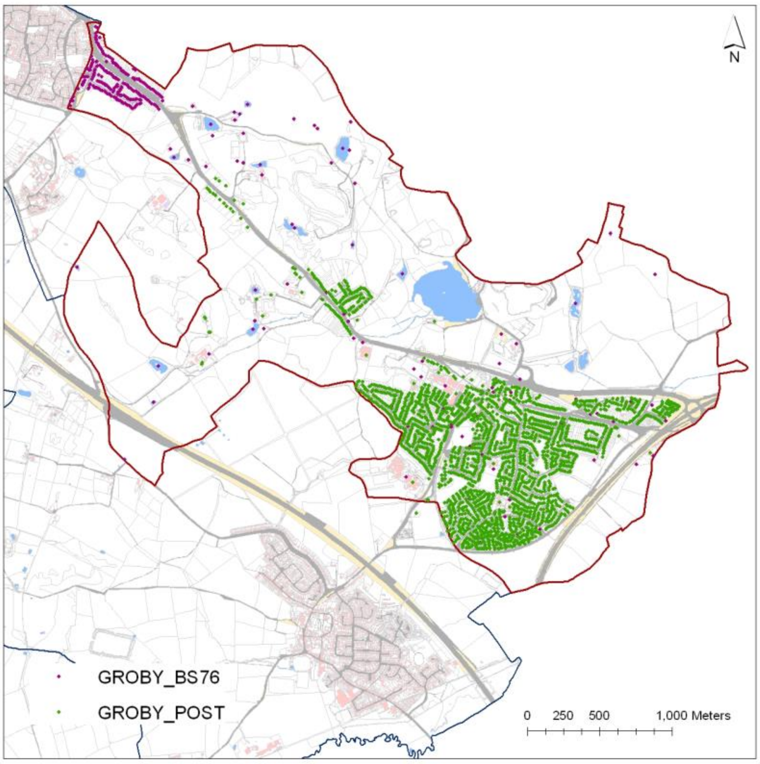
16- Bagworth & Thornton [no. 16]



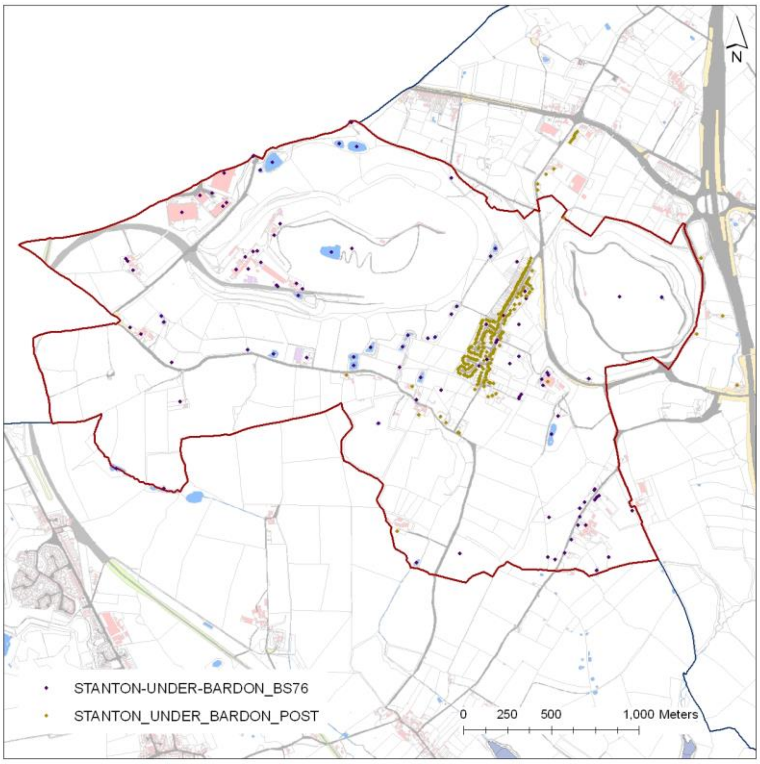
17- Ratby [no.17]



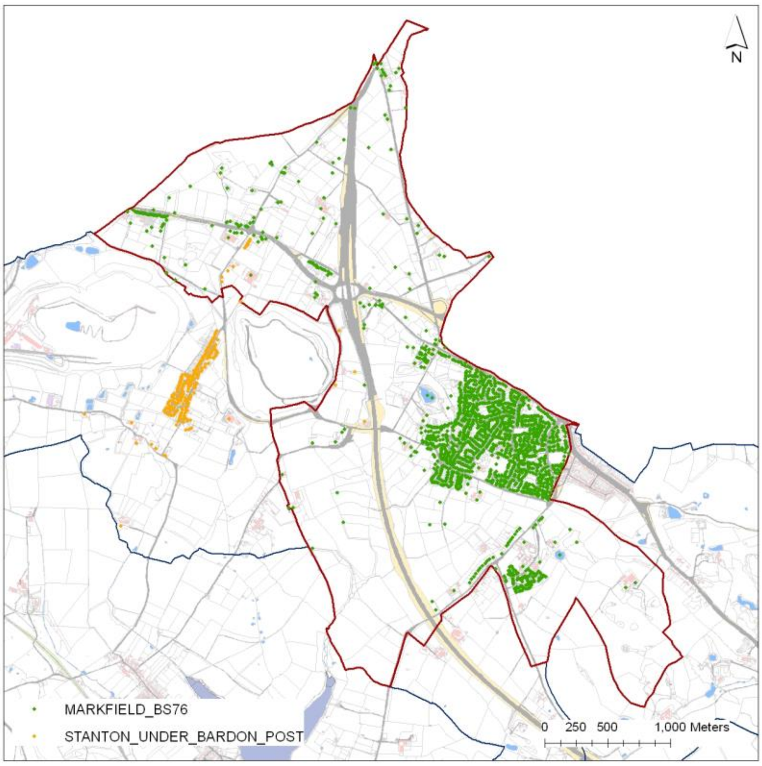
18- Groby [no.18]



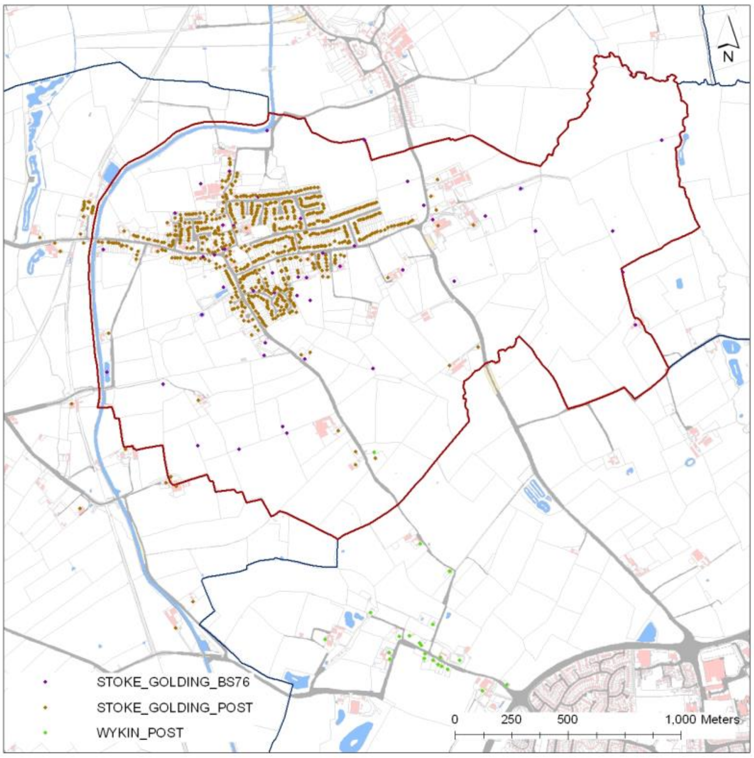
19- Stanton-under-Bardon [no.19]



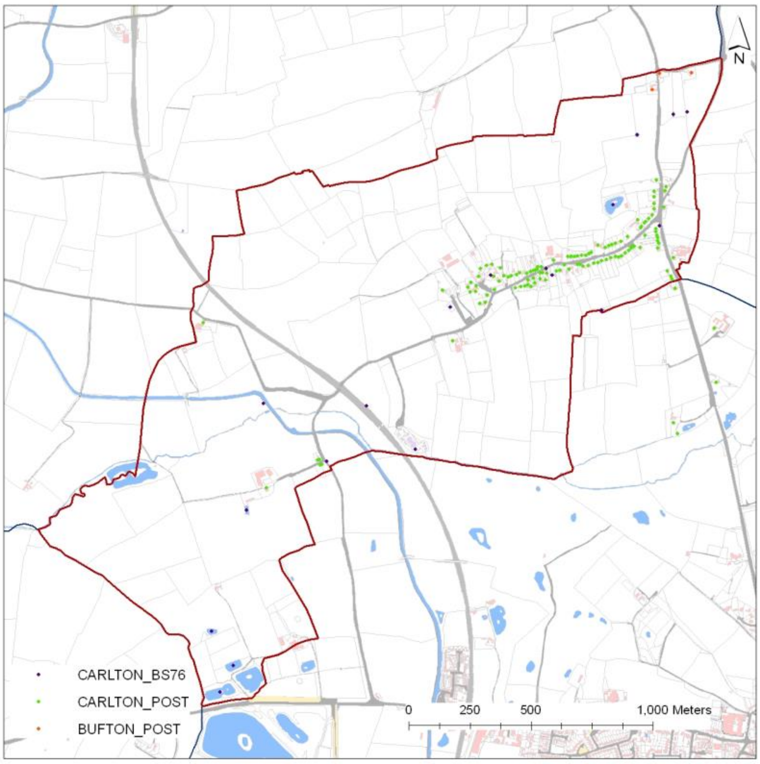
20- Markfield [no.20]



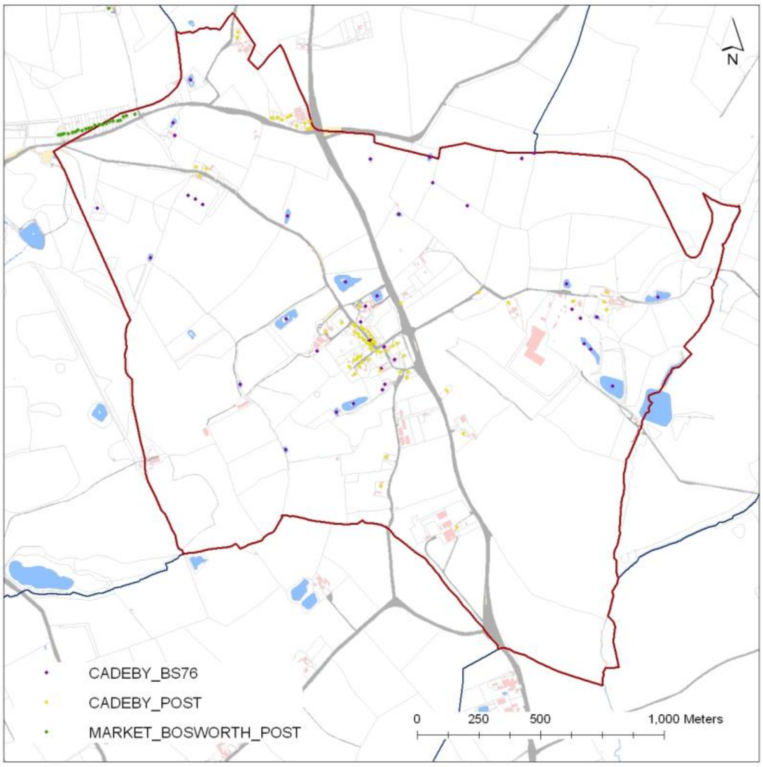
21- Stoke Golding [no.21]



22- Carlton [no.22]

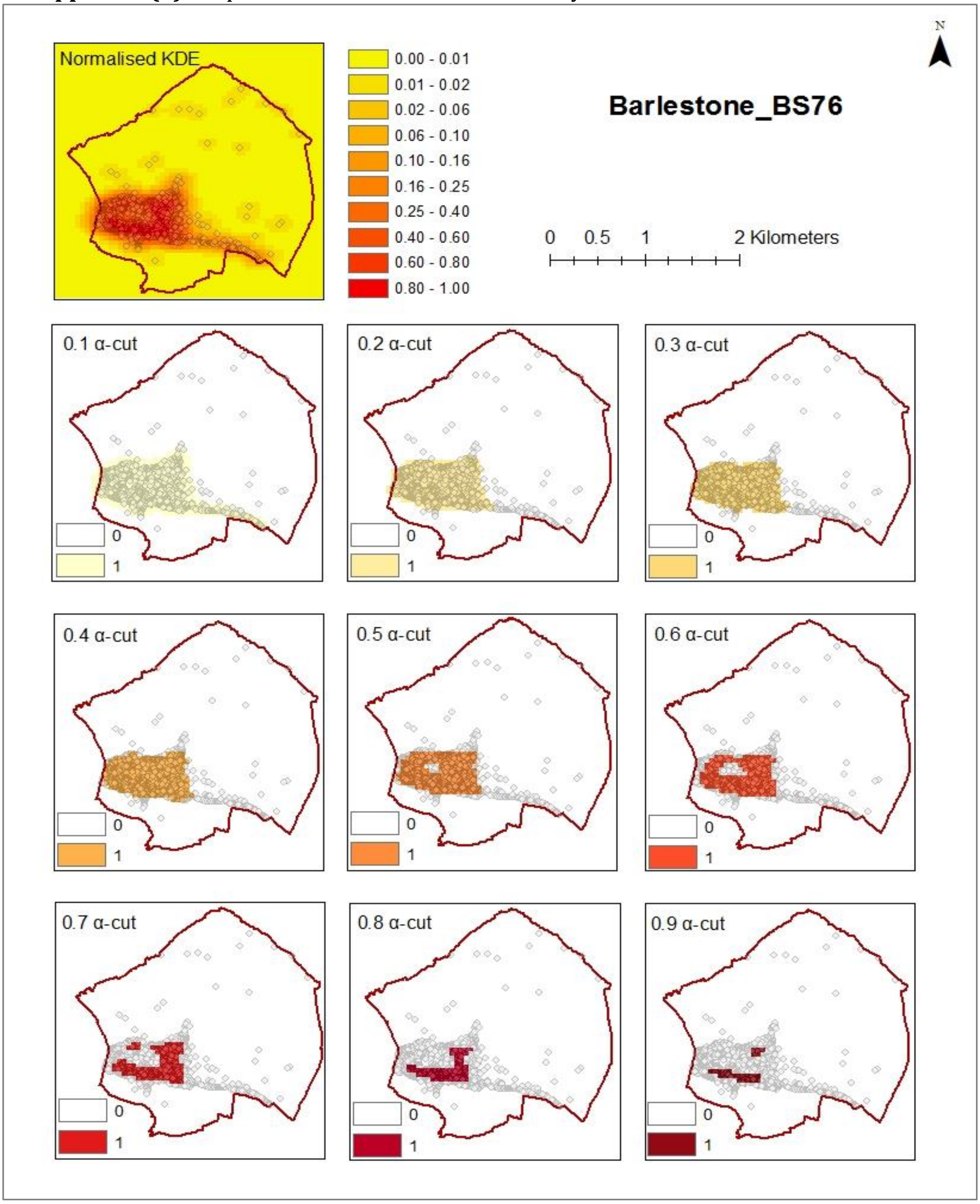


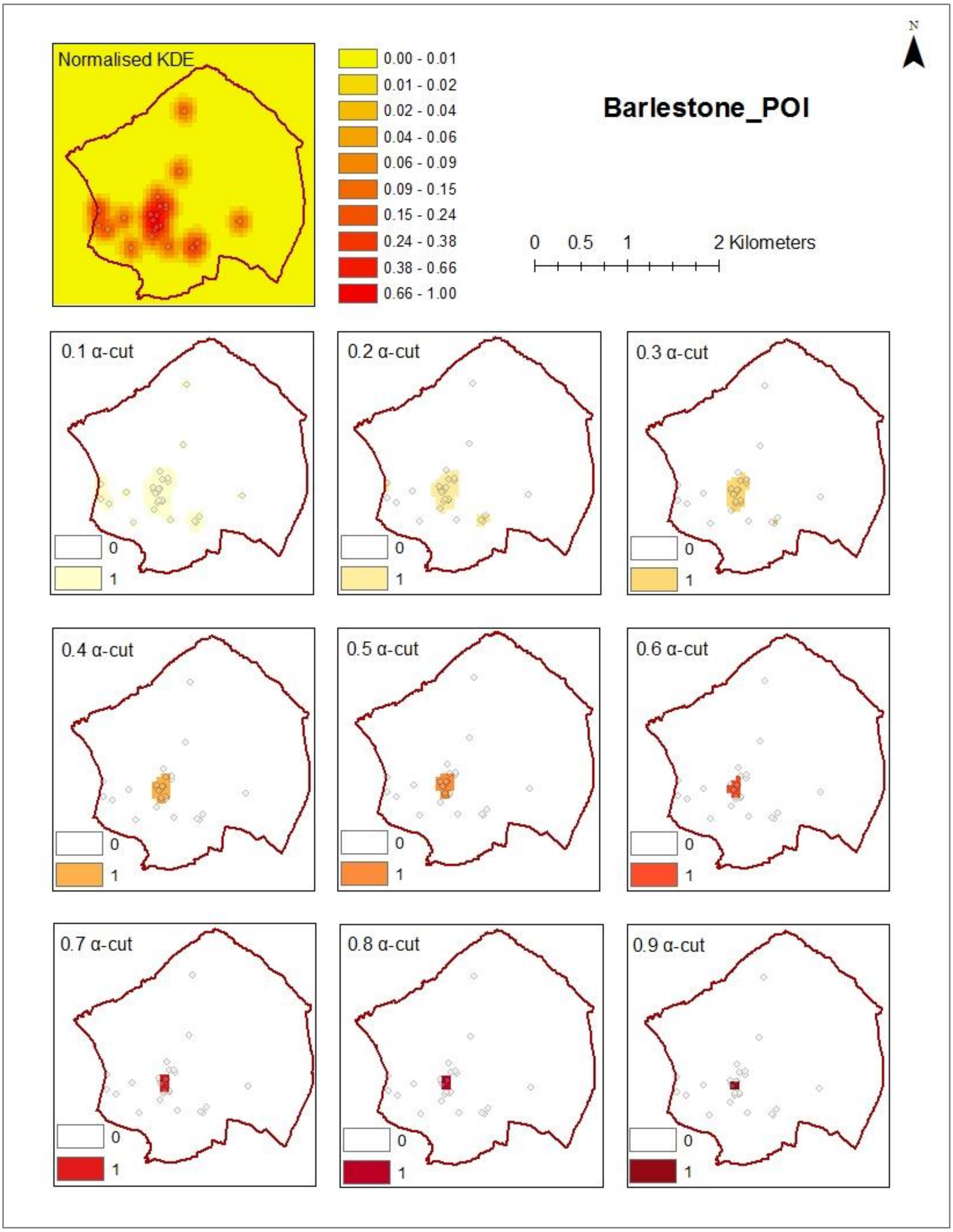
23- Cadeby [no.23]



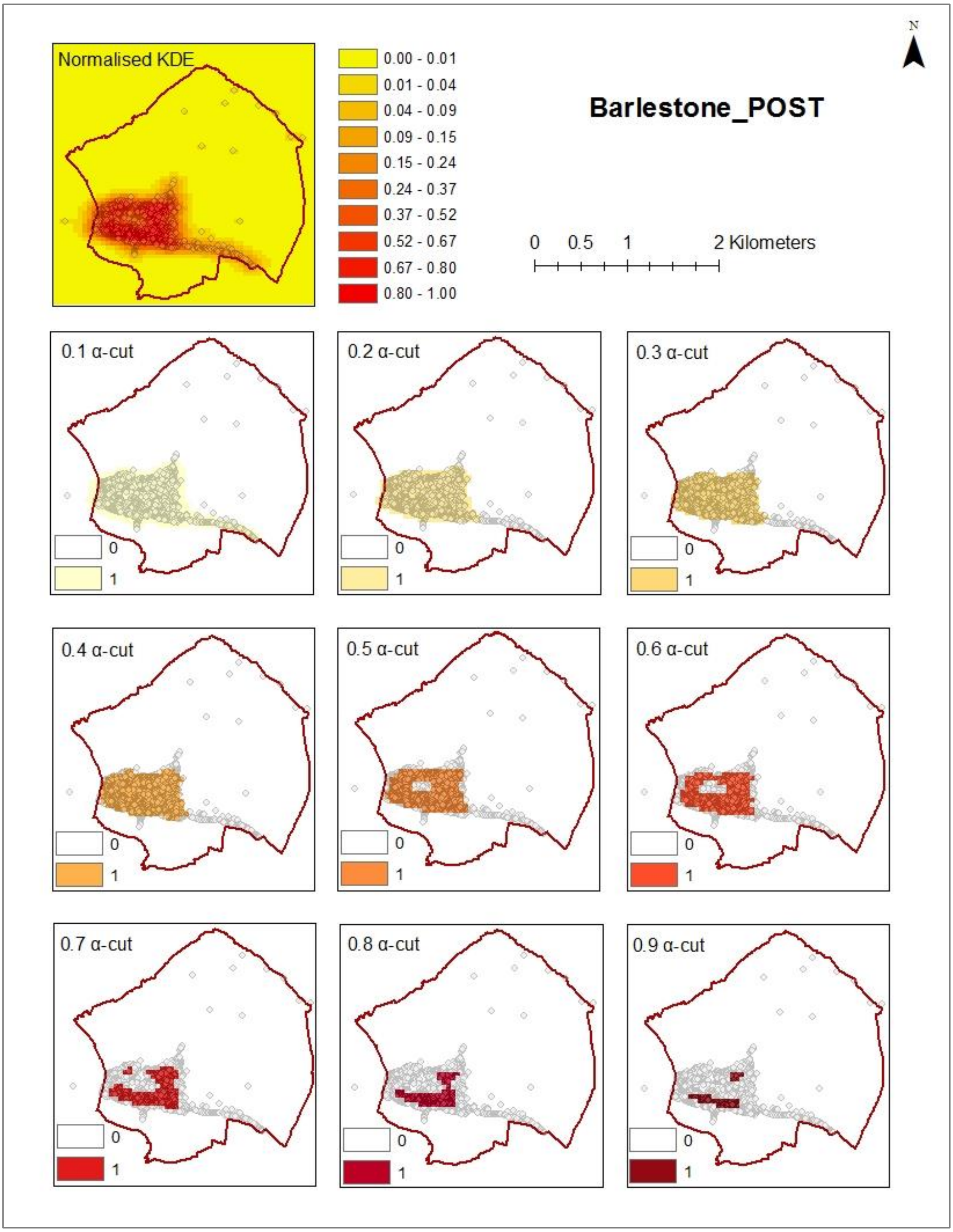
24- Barlestone [no.24]

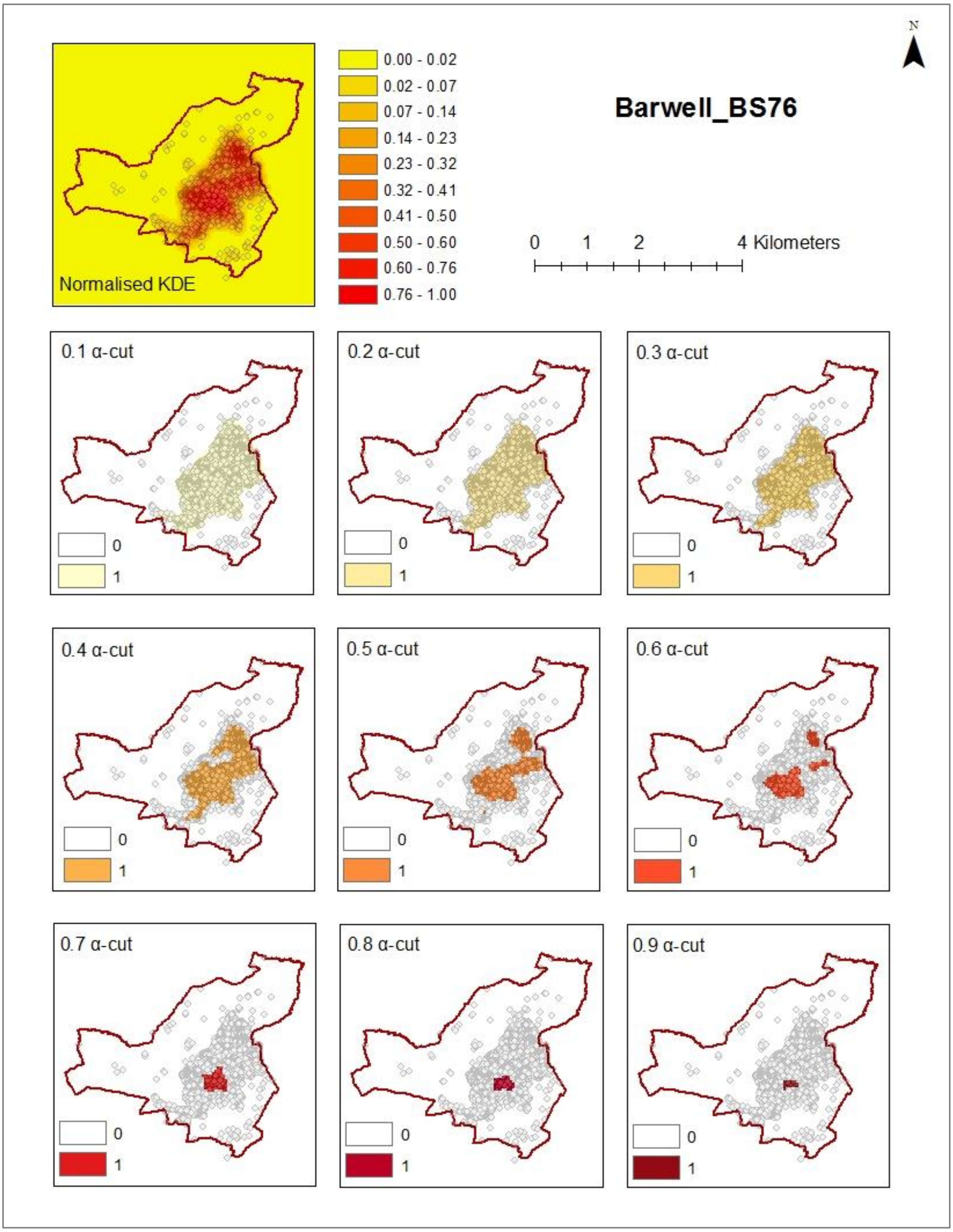
Appendix (2): Maps of the normalised kernel density with their α-cuts

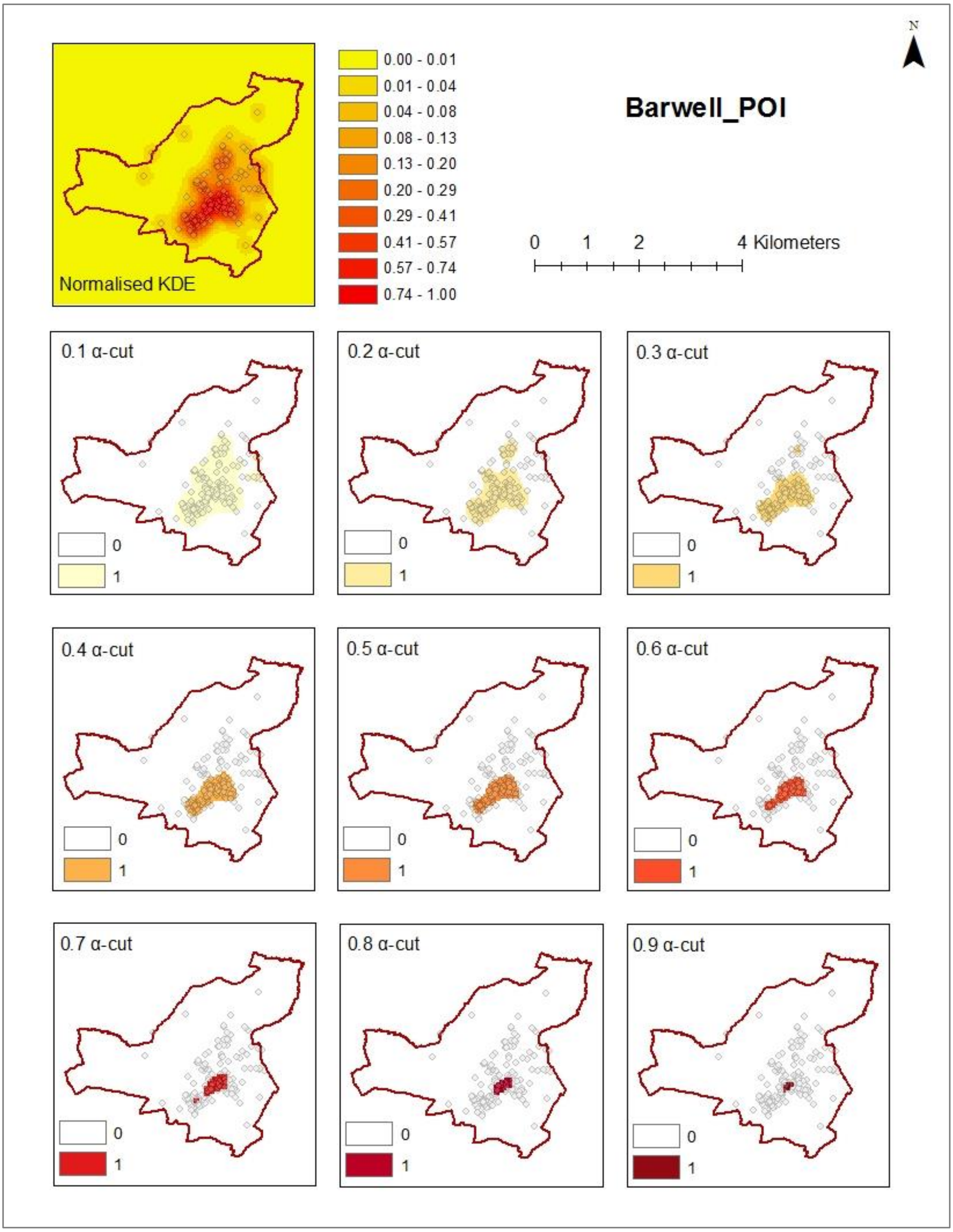


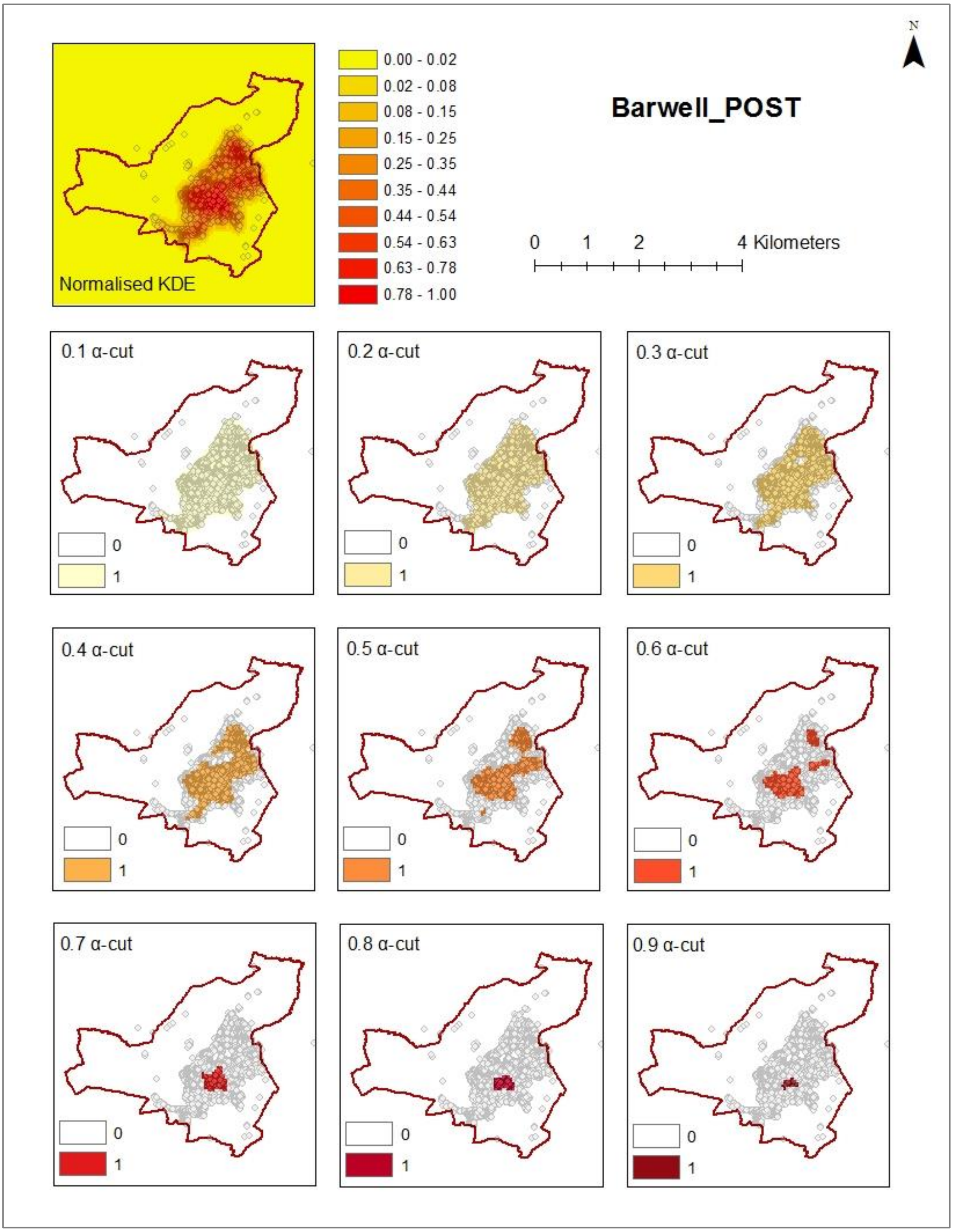


Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.06
0.06 - 0.09
0.09 - 0.15
0.15 - 0.24
0.24 - 0.38
0.38 - 0.66
0.66 - 1.00
Barlestone_POI
N
0 0.5 1 2 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

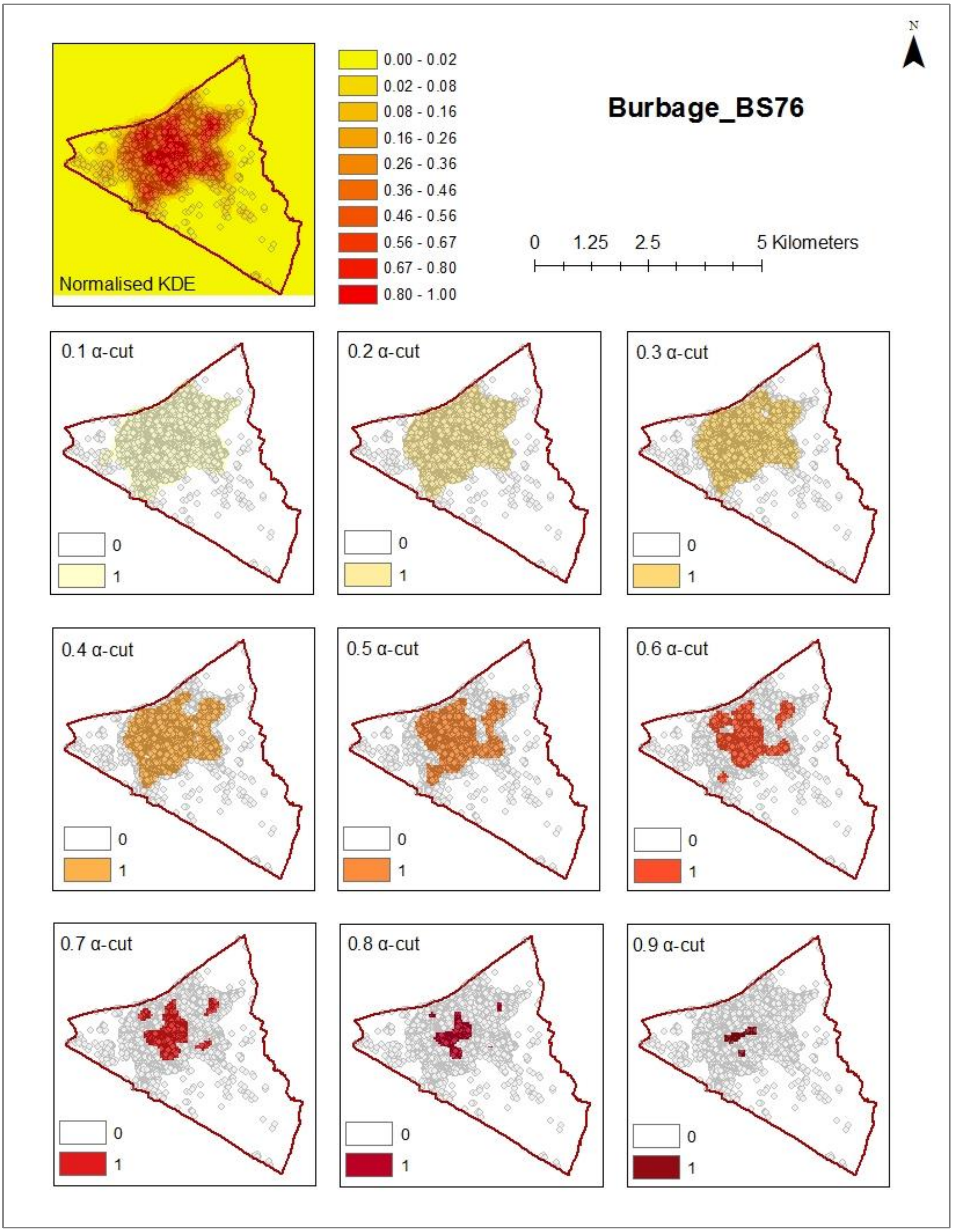

Barlestone_POST
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.09
0.09 - 0.15
0.15 - 0.24
0.24 - 0.37
0.37 - 0.52
0.52 - 0.67
0.67 - 0.80
0.80 - 1.00
0 0.5 1 2 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

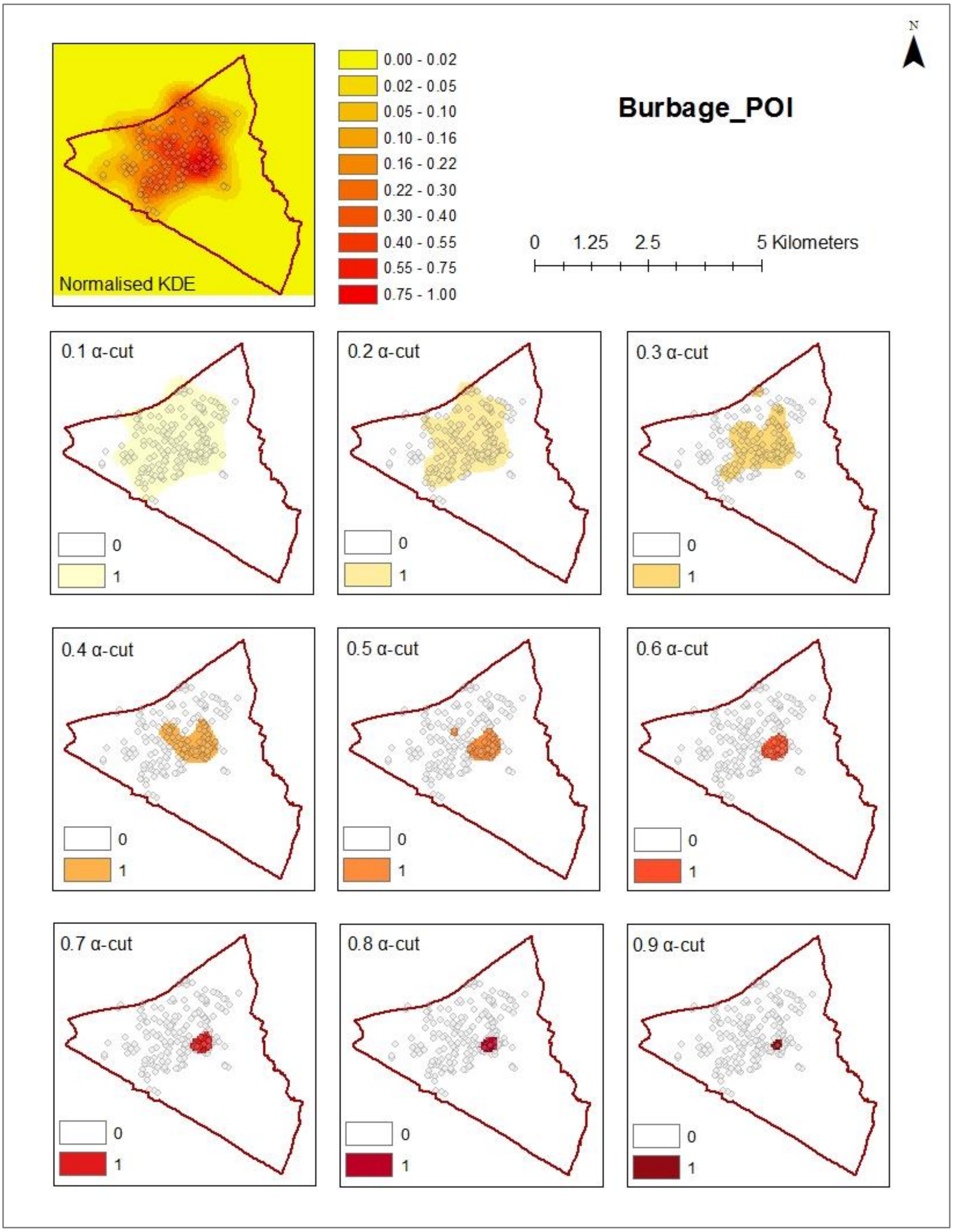

Barwell_BS76
Normalised KDE
0.00 - 0.02
0.02 - 0.07
0.07 - 0.14
0.14 - 0.23
0.23 - 0.32
0.32 - 0.41
0.41 - 0.50
0.50 - 0.60
0.60 - 0.76
0.76 - 1.00
0 1 2 4 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

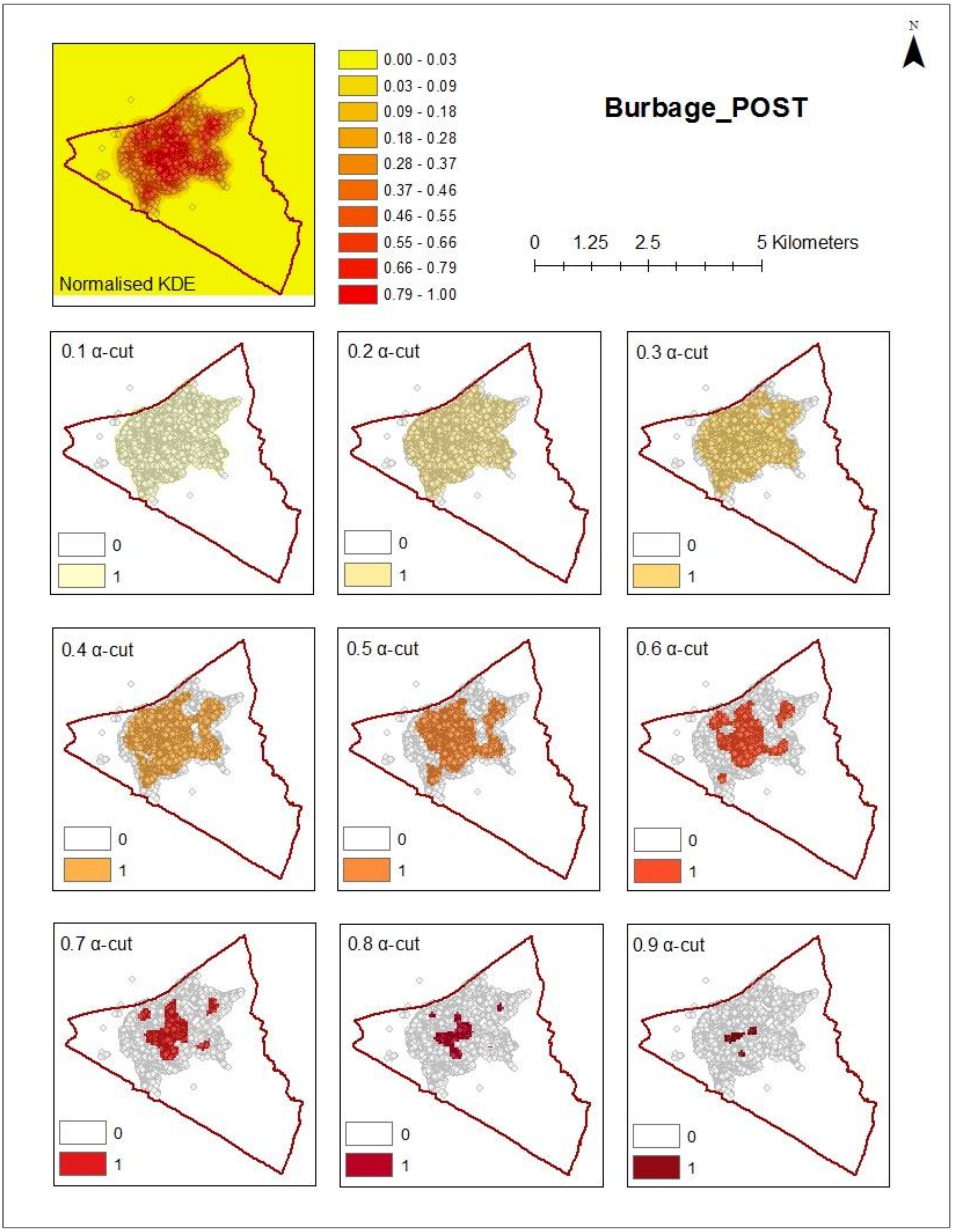

Barwell_POI
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.08
0.08 - 0.13
0.13 - 0.20
0.20 - 0.29
0.29 - 0.41
0.41 - 0.57
0.57 - 0.74
0.74 - 1.00
0 1 2 4 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

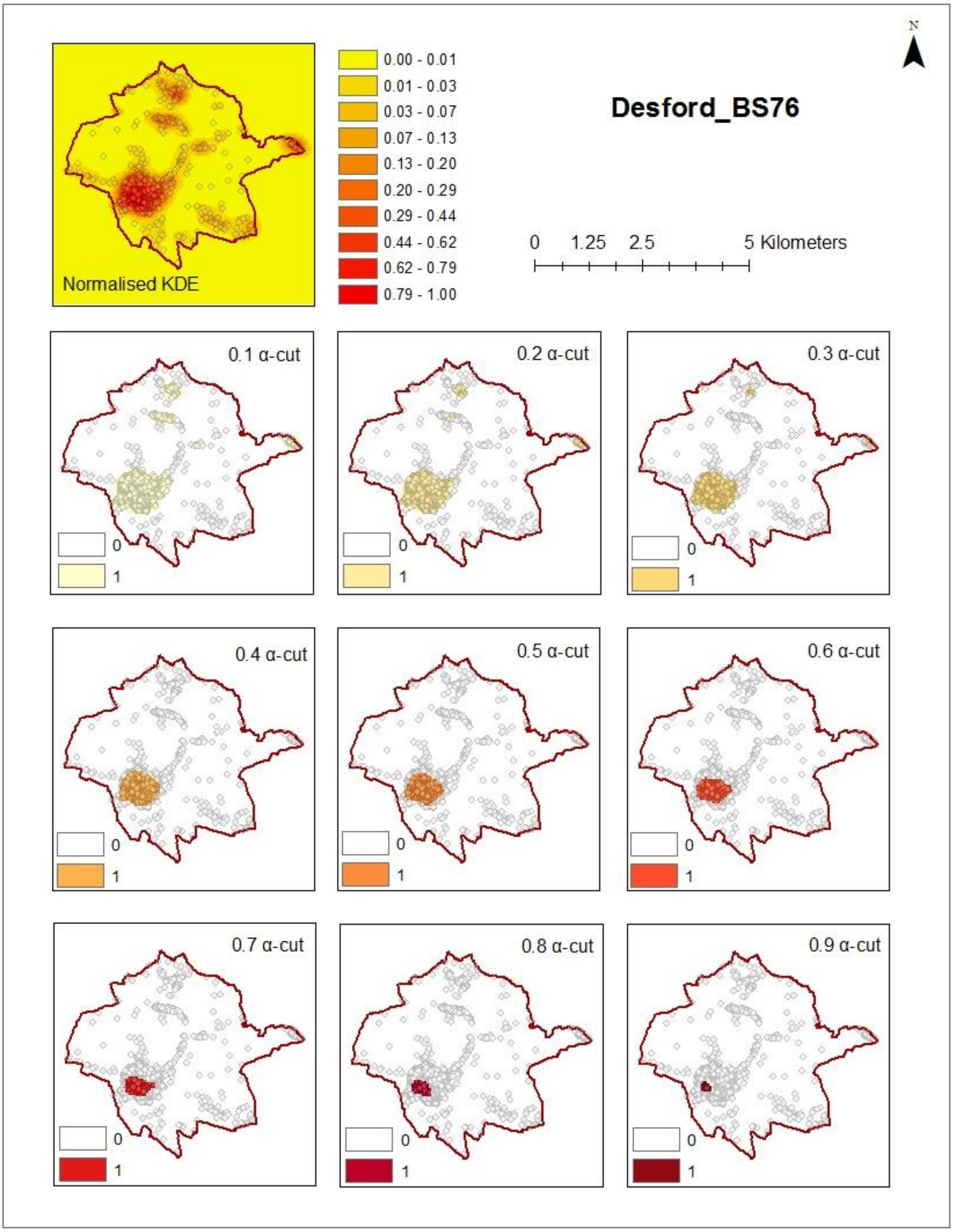


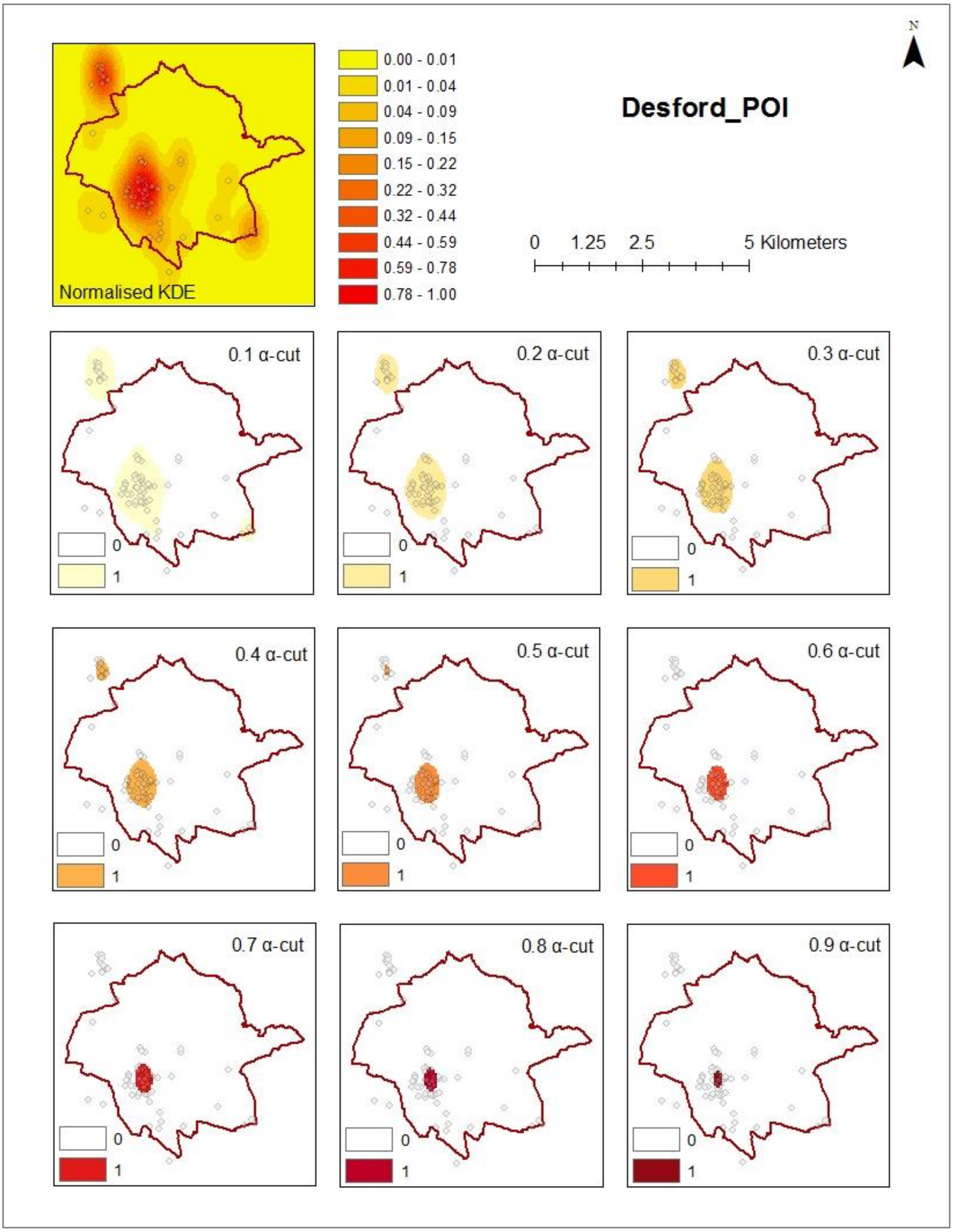
Barwell_POST
Normalised KDE
0.00 - 0.02
0.02 - 0.08
0.08 - 0.15
0.15 - 0.25
0.25 - 0.35
0.35 - 0.44
0.44 - 0.54
0.54 - 0.63
0.63 - 0.78
0.78 - 1.00
0 1 2 4 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


Burbage_BS76
N
Normalised KDE
0.00 - 0.02
0.02 - 0.08
0.08 - 0.16
0.16 - 0.26
0.26 - 0.36
0.36 - 0.46
0.46 - 0.56
0.56 - 0.67
0.67 - 0.80
0.80 - 1.00
0 1.25 2.5 5 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

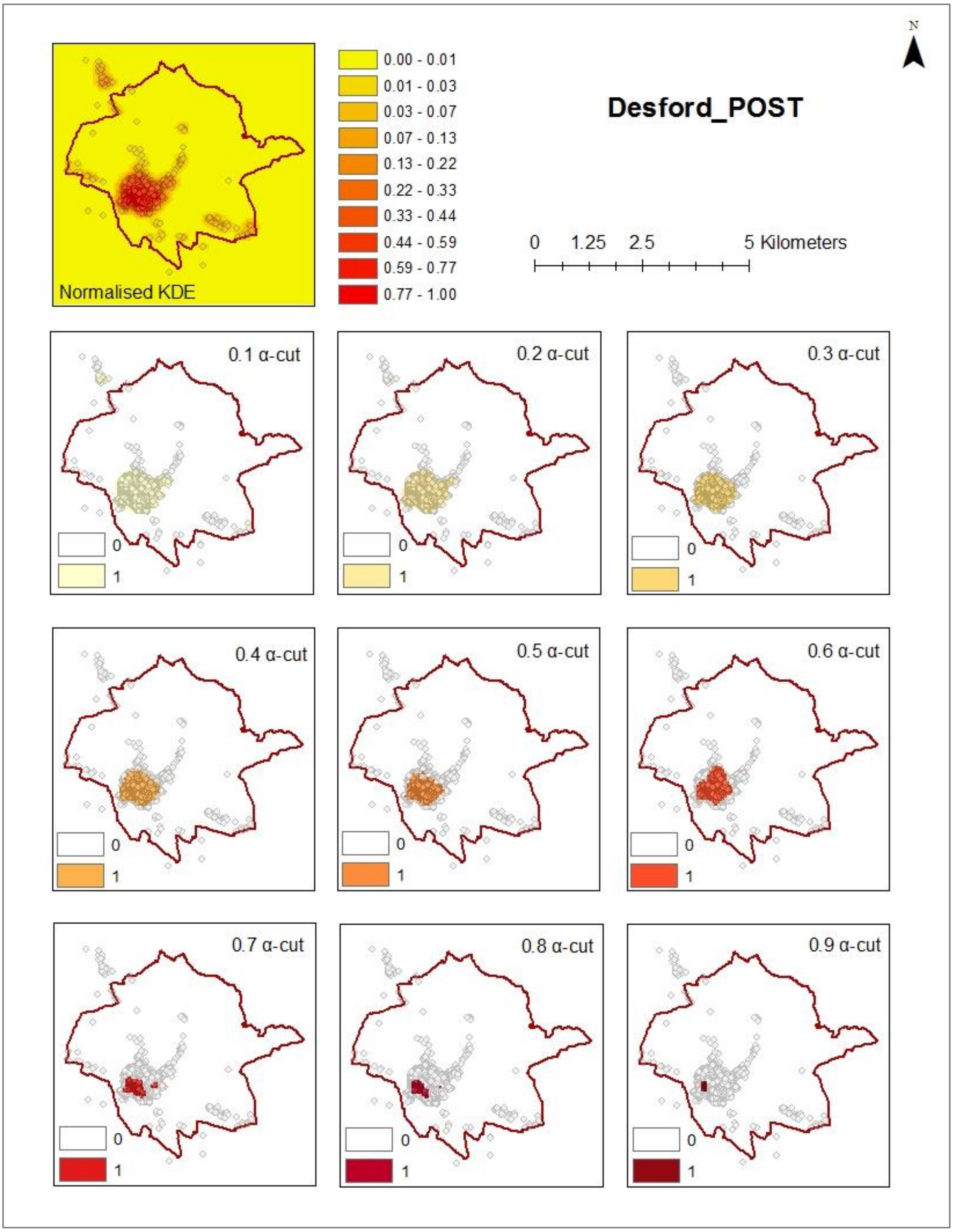

Burbage_POI
Normalised KDE
0.00 - 0.02
0.02 - 0.05
0.05 - 0.10
0.10 - 0.16
0.16 - 0.22
0.22 - 0.30
0.30 - 0.40
0.40 - 0.55
0.55 - 0.75
0.75 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


Burbage_POST
Normalised KDE
0.00 - 0.03
0.03 - 0.09
0.09 - 0.18
0.18 - 0.28
0.28 - 0.37
0.37 - 0.46
0.46 - 0.55
0.55 - 0.66
0.66 - 0.79
0.79 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

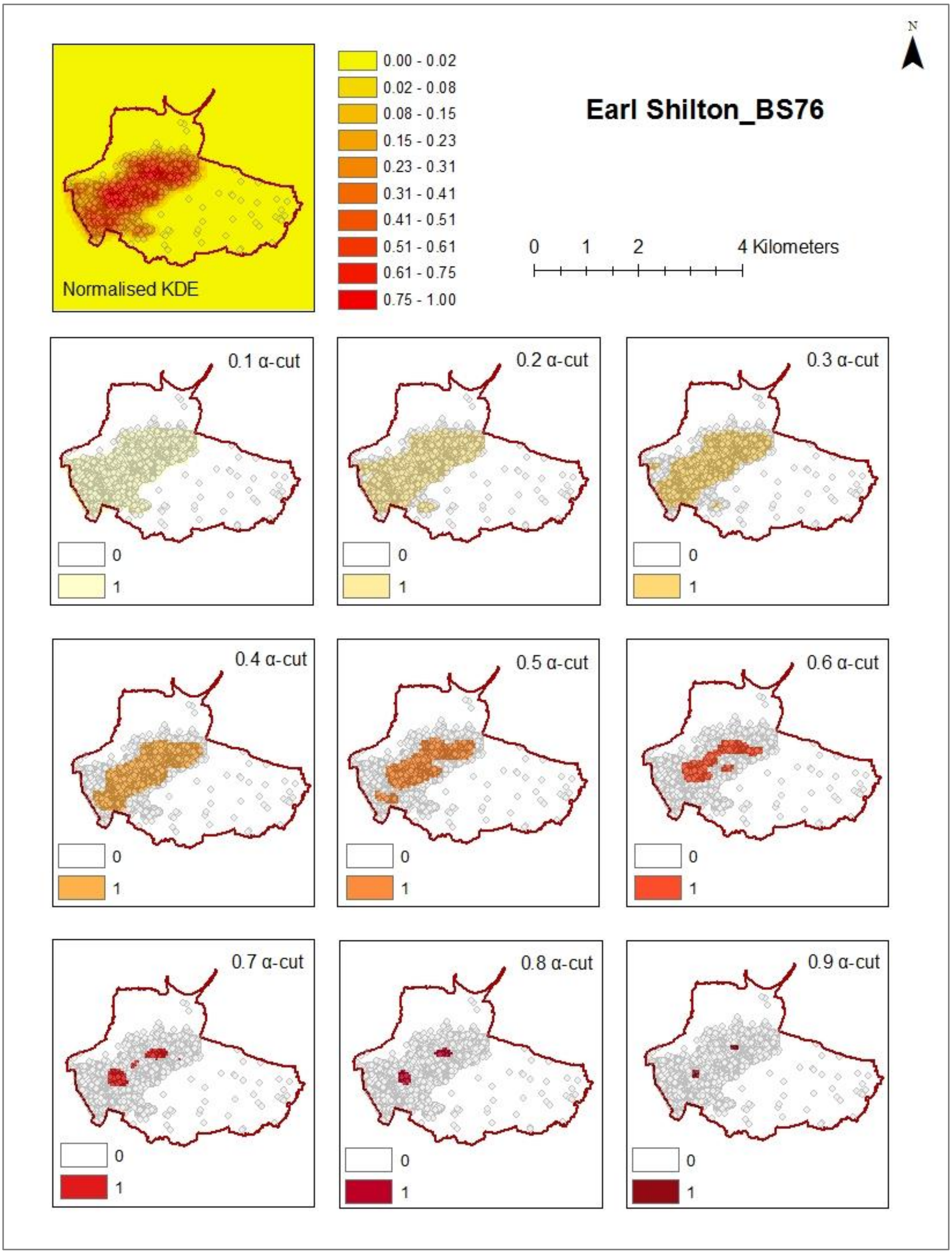

Desford_BS76
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.07
0.07 - 0.13
0.13 - 0.20
0.20 - 0.29
0.29 - 0.44
0.44 - 0.62
0.62 - 0.79
0.79 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

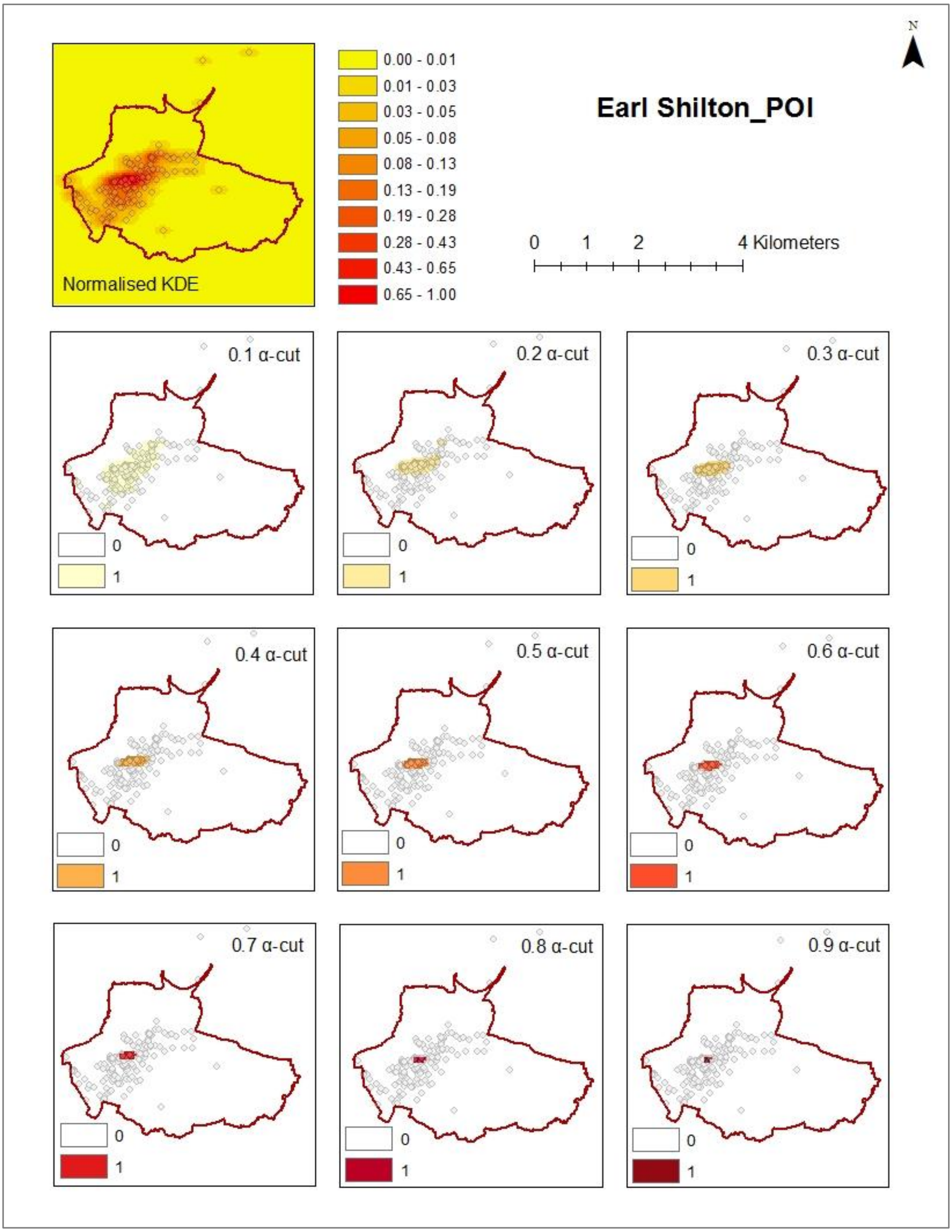


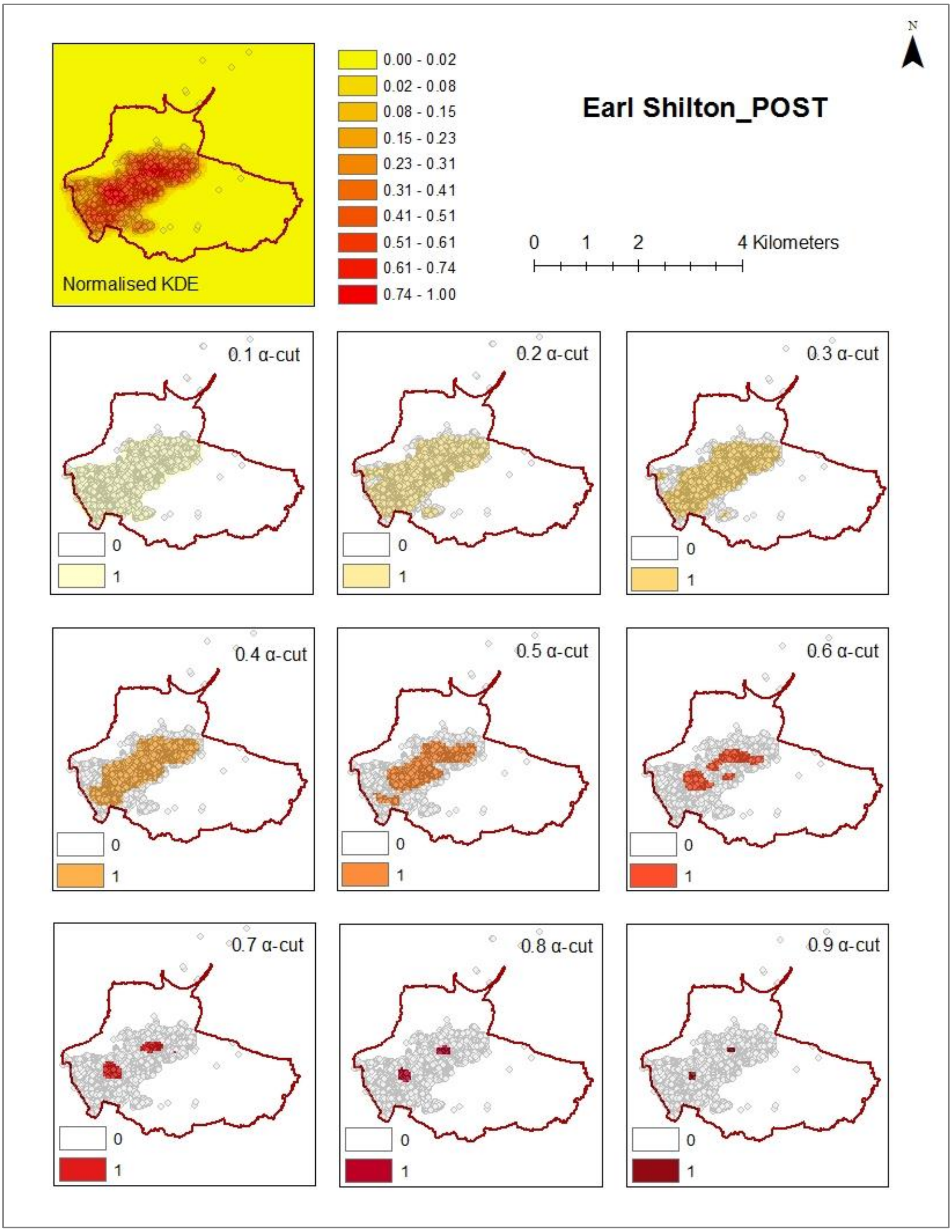
Desford_POI
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.09
0.09 - 0.15
0.15 - 0.22
0.22 - 0.32
0.32 - 0.44
0.44 - 0.59
0.59 - 0.78
0.78 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

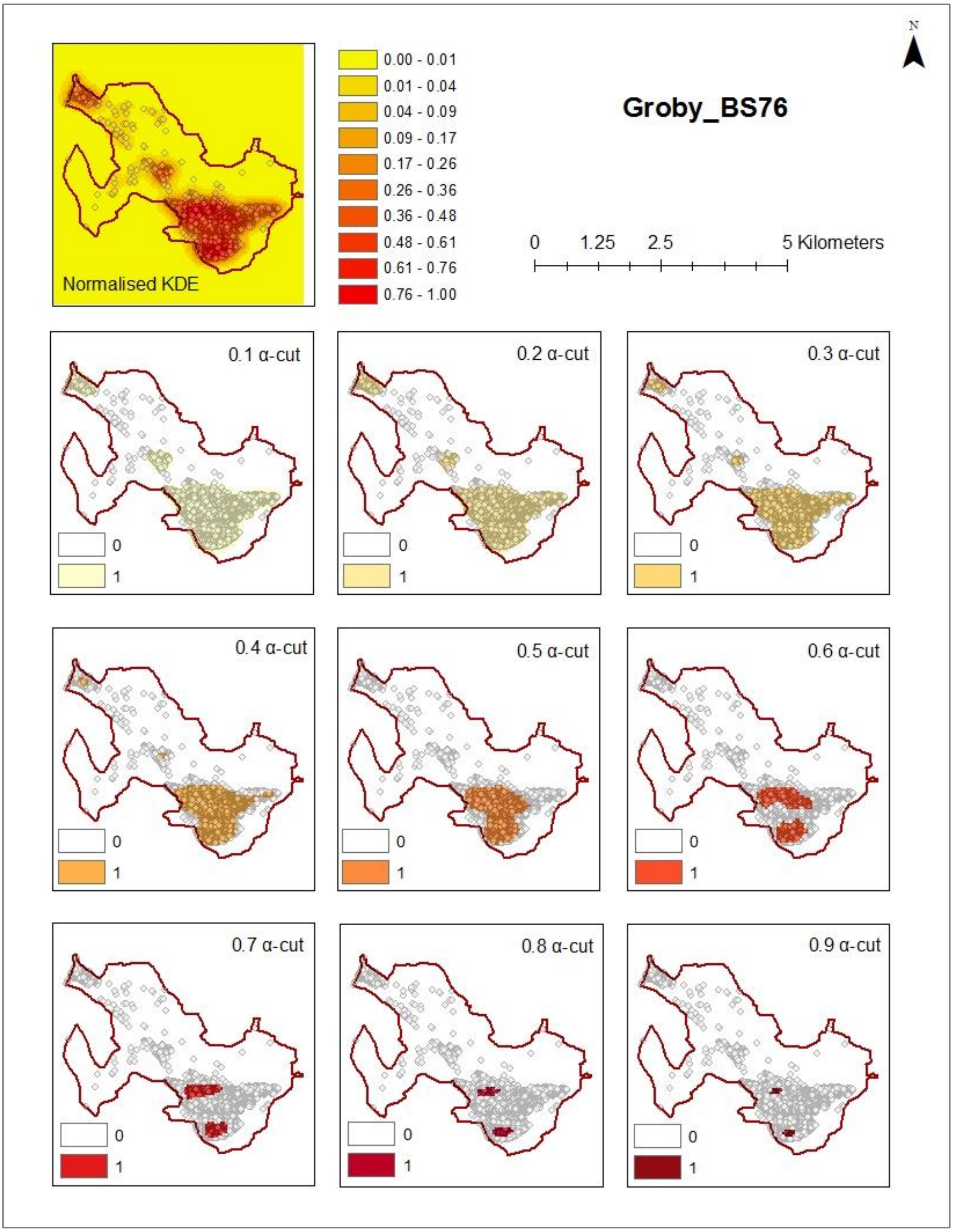


Desford_POST
N
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.07
0.07 - 0.13
0.13 - 0.22
0.22 - 0.33
0.33 - 0.44
0.44 - 0.59
0.59 - 0.77
0.77 - 1.00
0 1.25 2.5 5 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

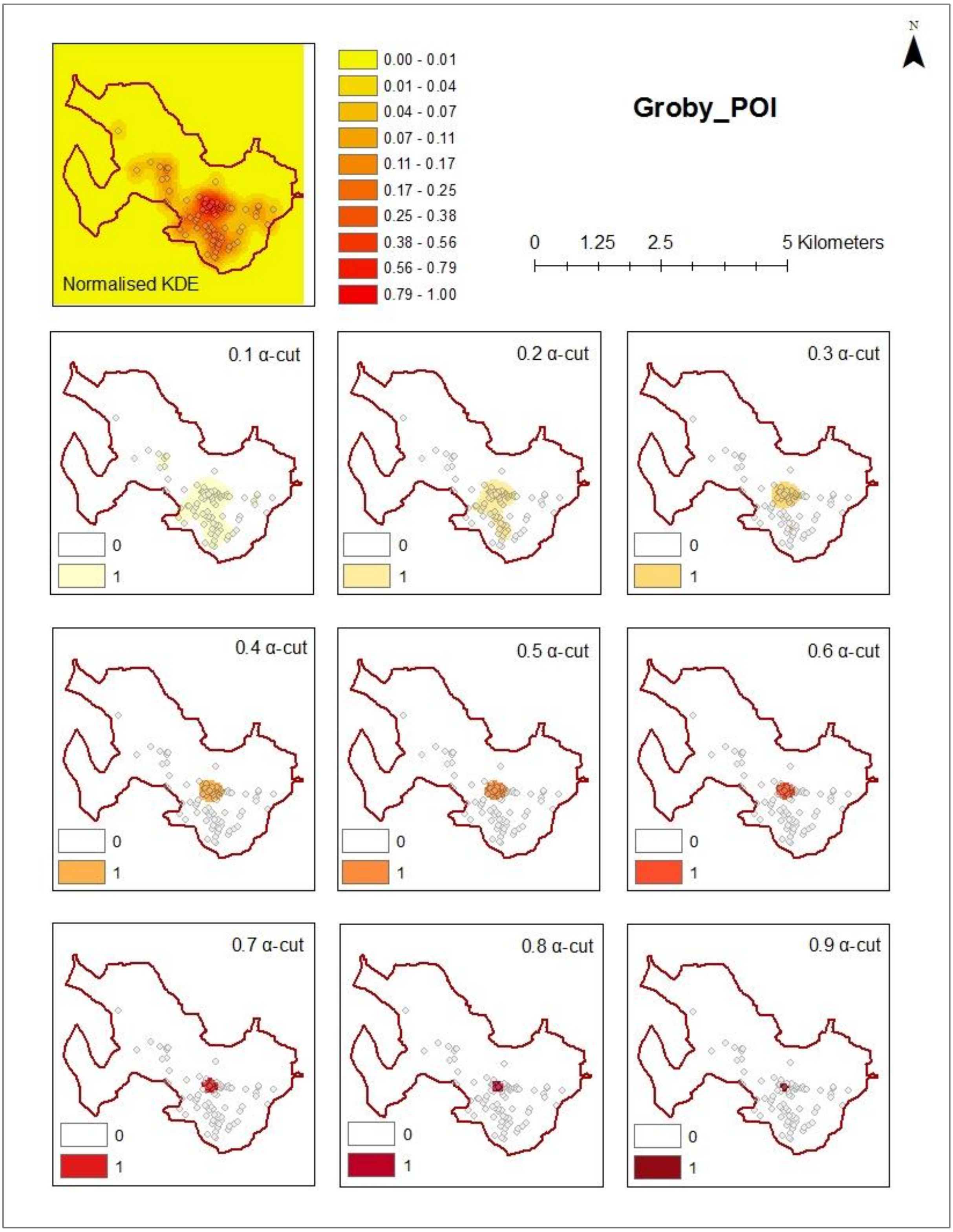

Earl Shilton_BS76
N
Normalised KDE
0.00 - 0.02
0.02 - 0.08
0.08 - 0.15
0.15 - 0.23
0.23 - 0.31
0.31 - 0.41
0.41 - 0.51
0.51 - 0.61
0.61 - 0.75
0.75 - 1.00
0 1 2 4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


Earl Shilton_POI
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.05
0.05 - 0.08
0.08 - 0.13
0.13 - 0.19
0.19 - 0.28
0.28 - 0.43
0.43 - 0.65
0.65 - 1.00
0 1 2 4 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

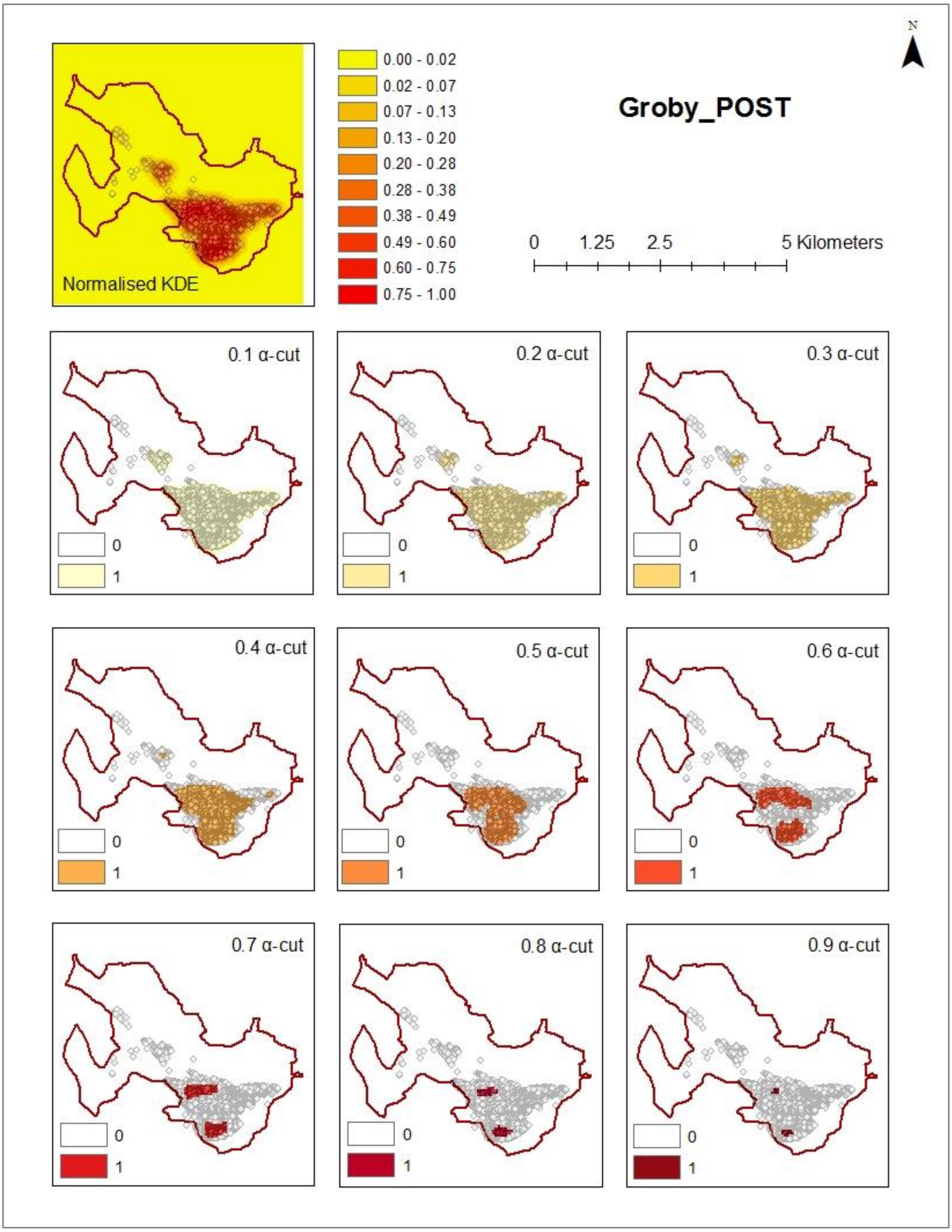

Earl Shilton_POST
Normalised KDE
0.00 - 0.02
0.02 - 0.08
0.08 - 0.15
0.15 - 0.23
0.23 - 0.31
0.31 - 0.41
0.41 - 0.51
0.51 - 0.61
0.61 - 0.74
0.74 - 1.00
0 1 2 4 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1



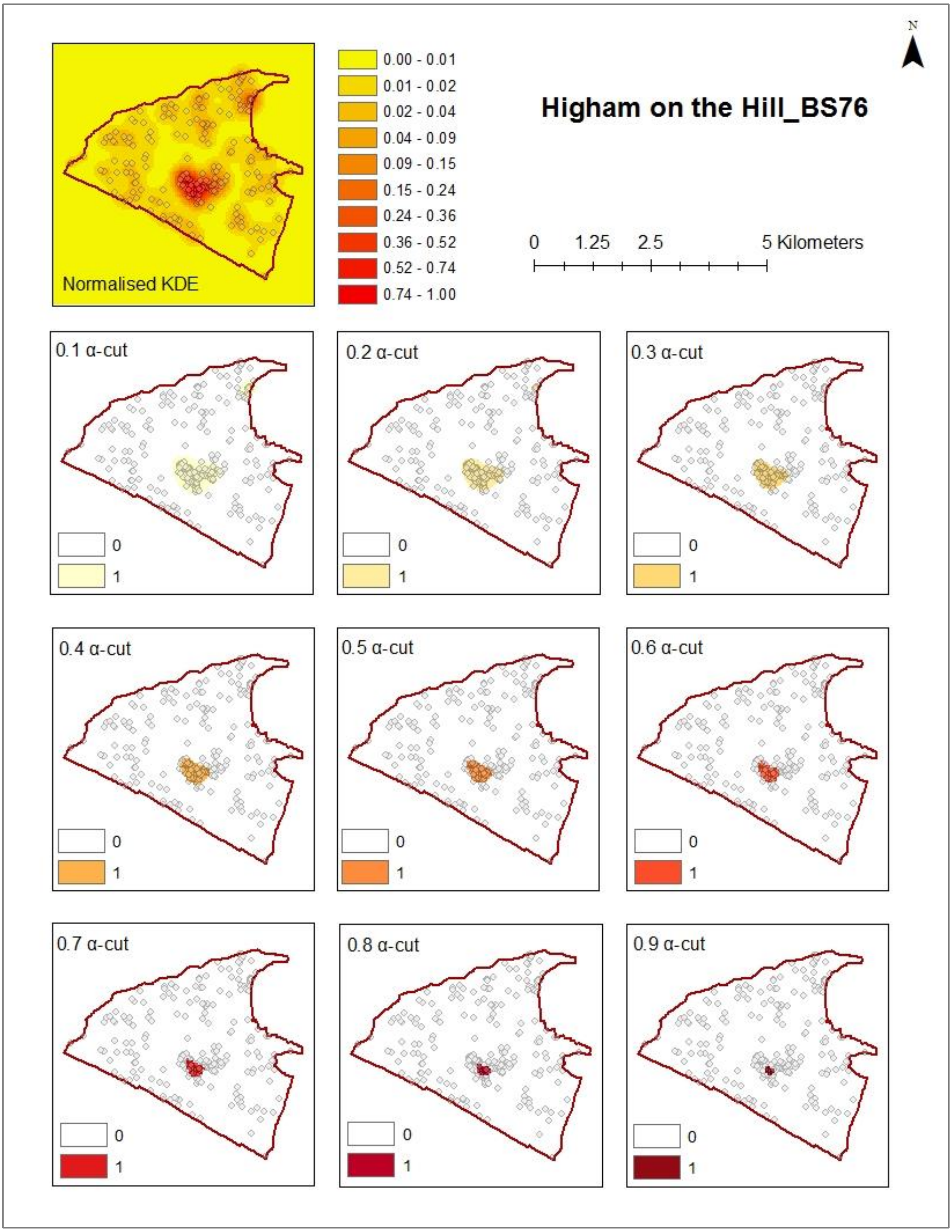
Groby_BS76
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.09
0.09 - 0.17
0.17 - 0.26
0.26 - 0.36
0.36 - 0.48
0.48 - 0.61
0.61 - 0.76
0.76 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1



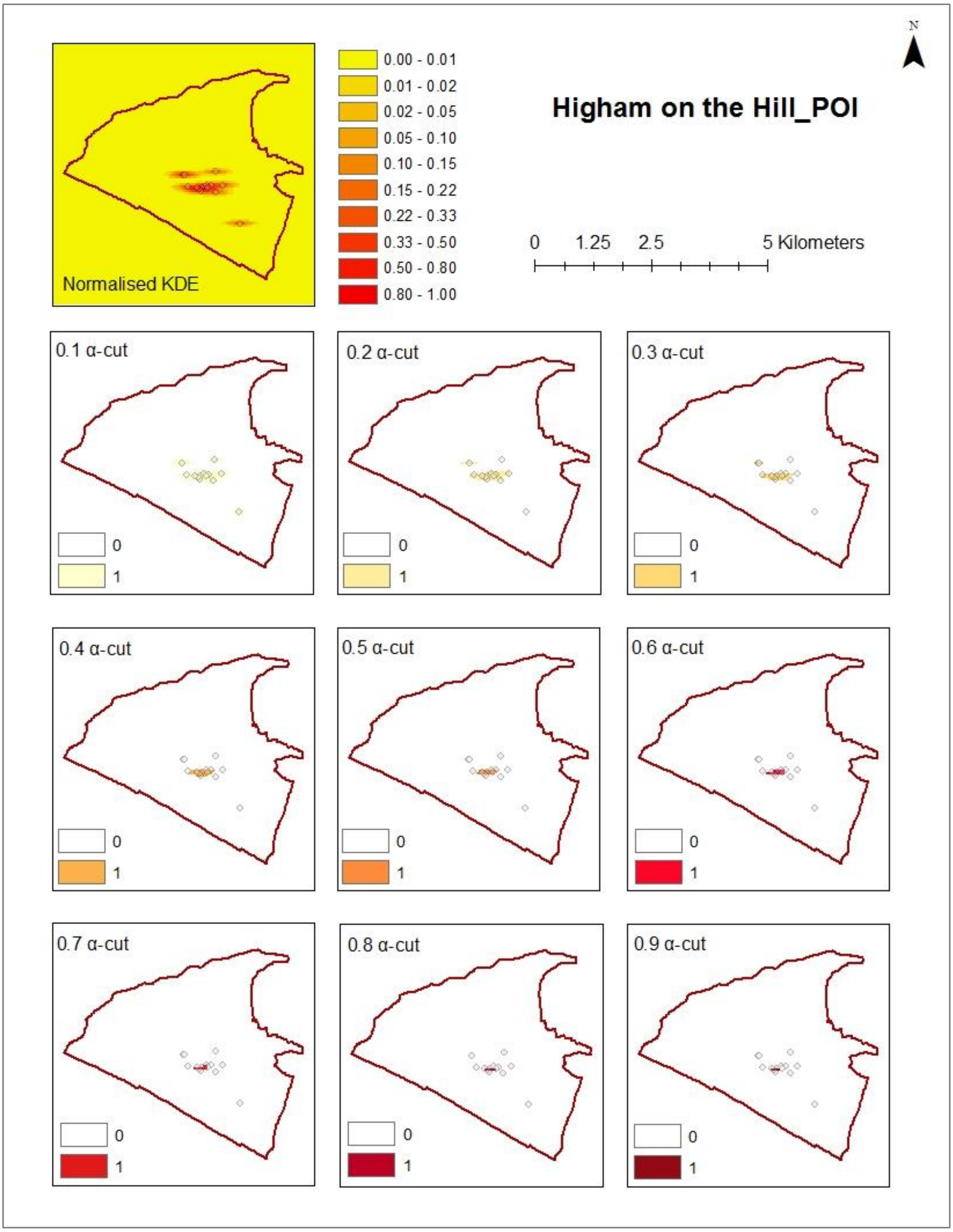
Groby_POI
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.07
0.07 - 0.11
0.11 - 0.17
0.17 - 0.25
0.25 - 0.38
0.38 - 0.56
0.56 - 0.79
0.79 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1



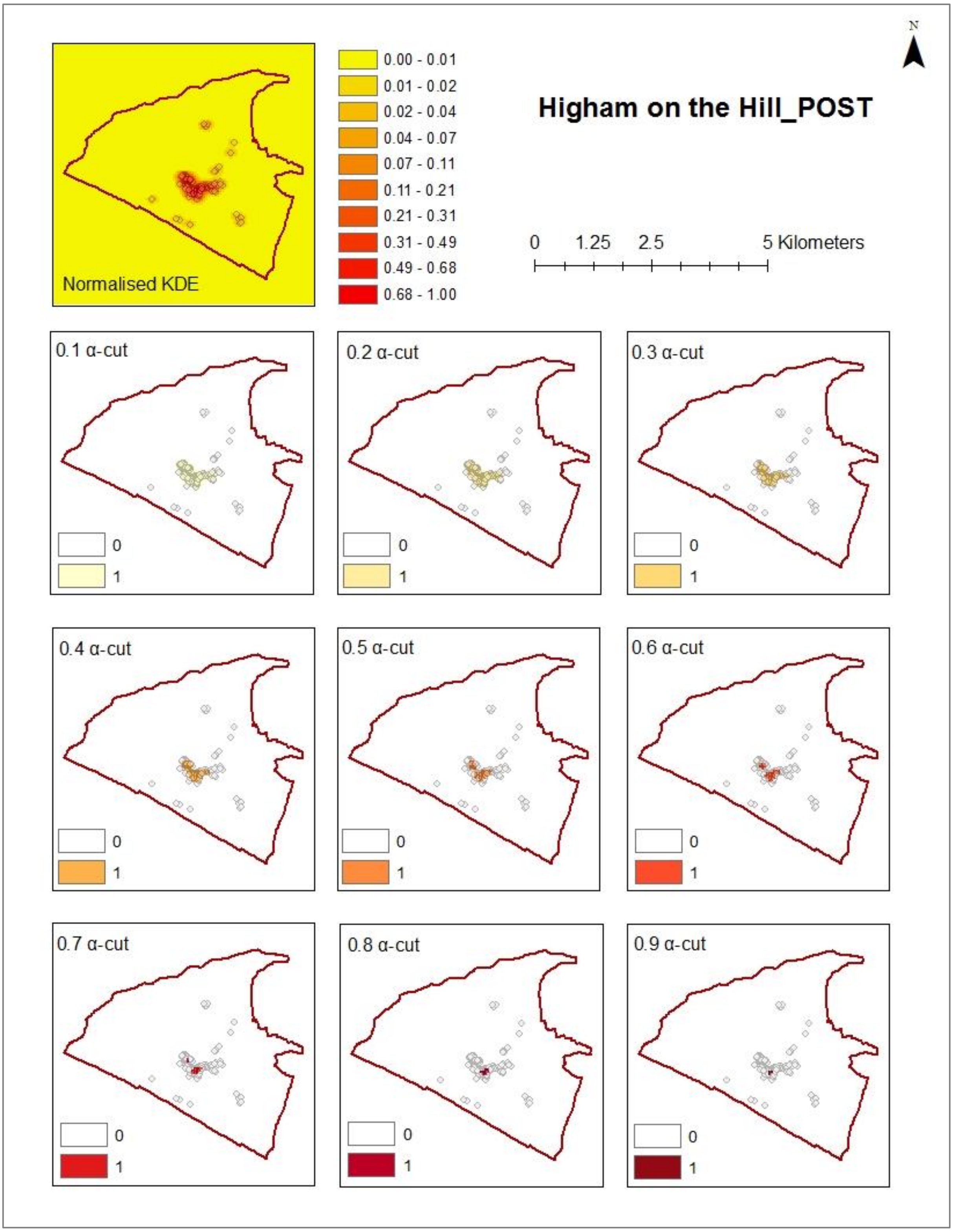
Groby_POST
Normalised KDE
0.00 - 0.02
0.02 - 0.07
0.07 - 0.13
0.13 - 0.20
0.20 - 0.28
0.28 - 0.38
0.38 - 0.49
0.49 - 0.60
0.60 - 0.75
0.75 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

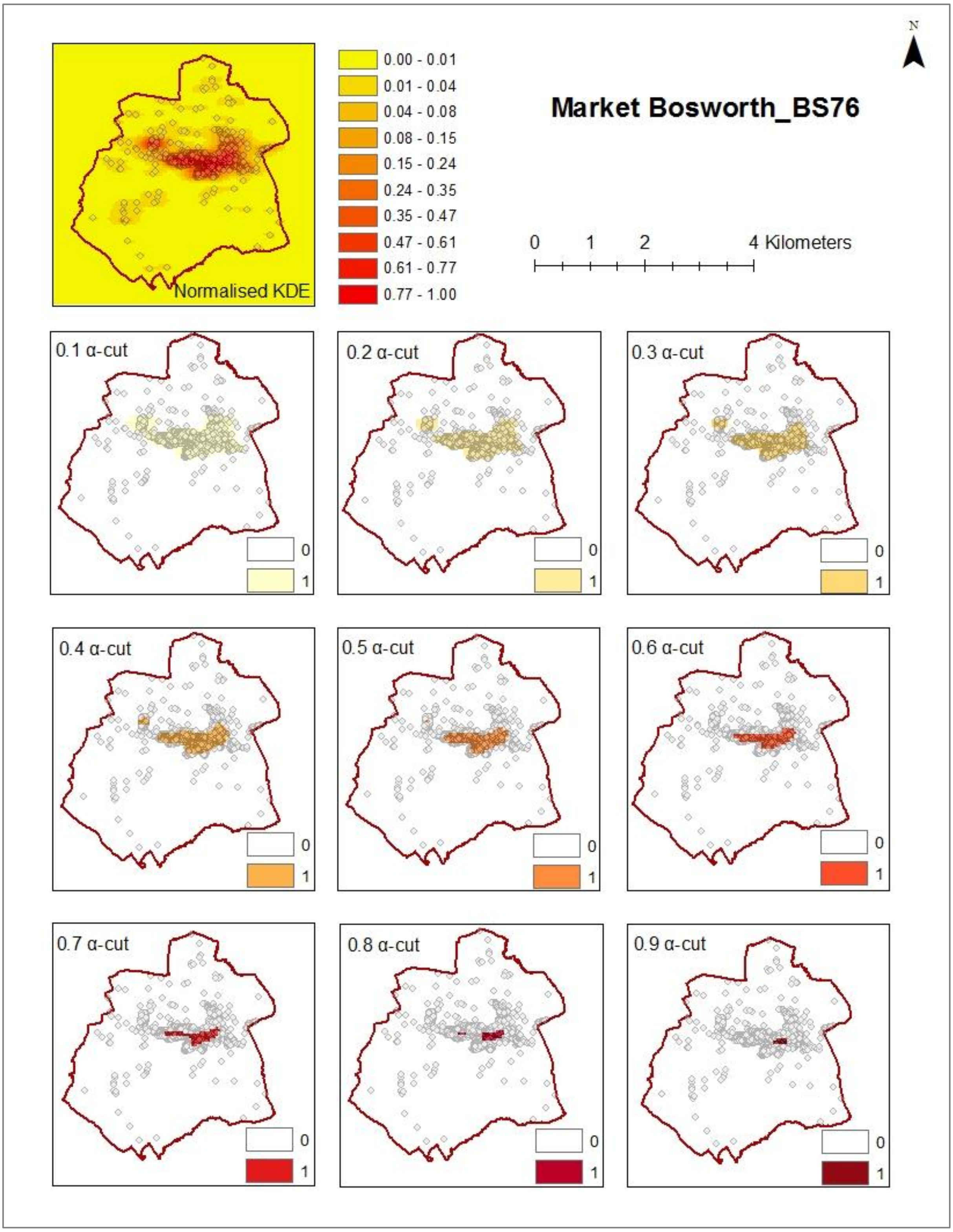


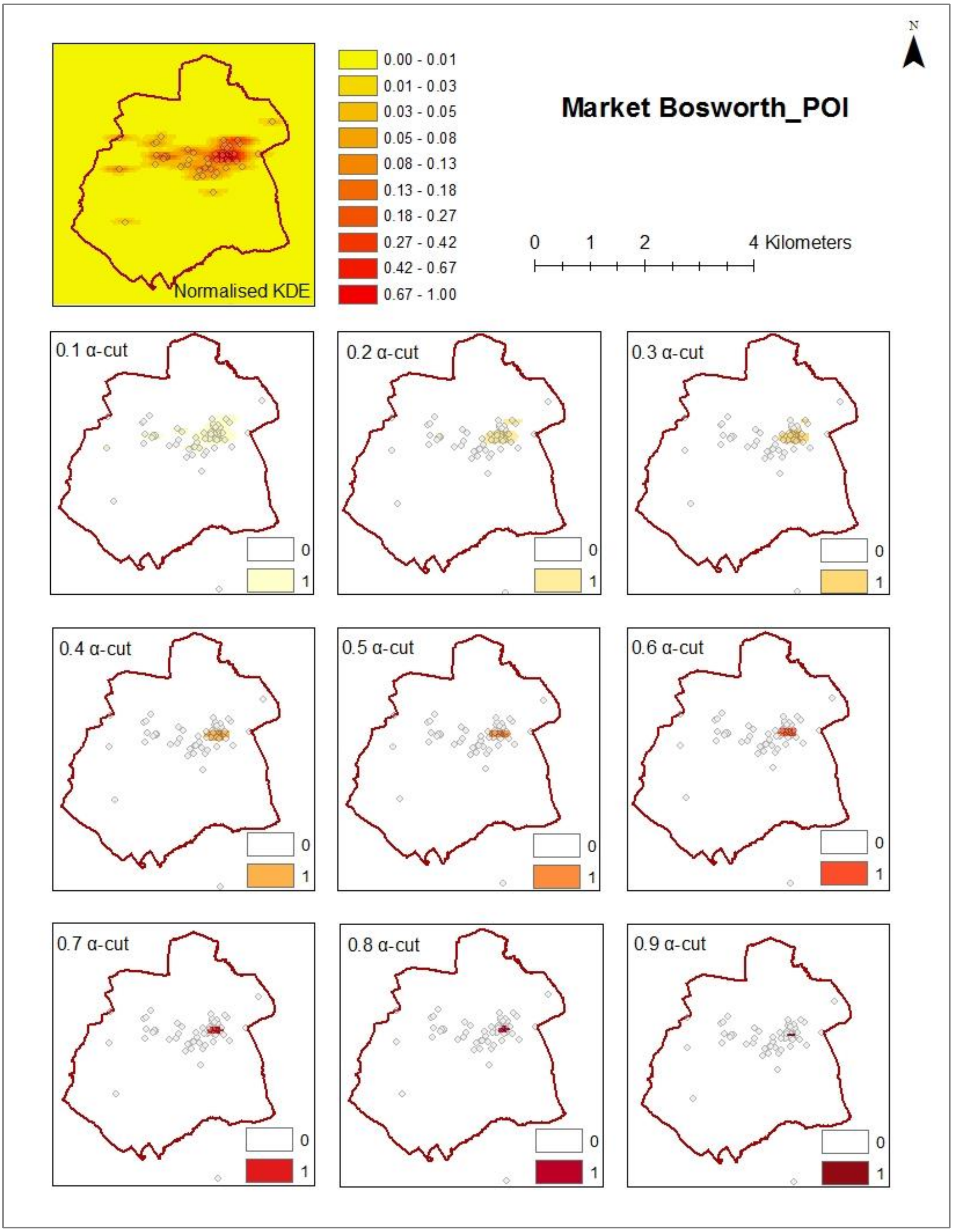
Higham on the Hill_BS76
N
0 1.25 2.5 5 Kilometers
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.09
0.09 - 0.15
0.15 - 0.24
0.24 - 0.36
0.36 - 0.52
0.52 - 0.74
0.74 - 1.00
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

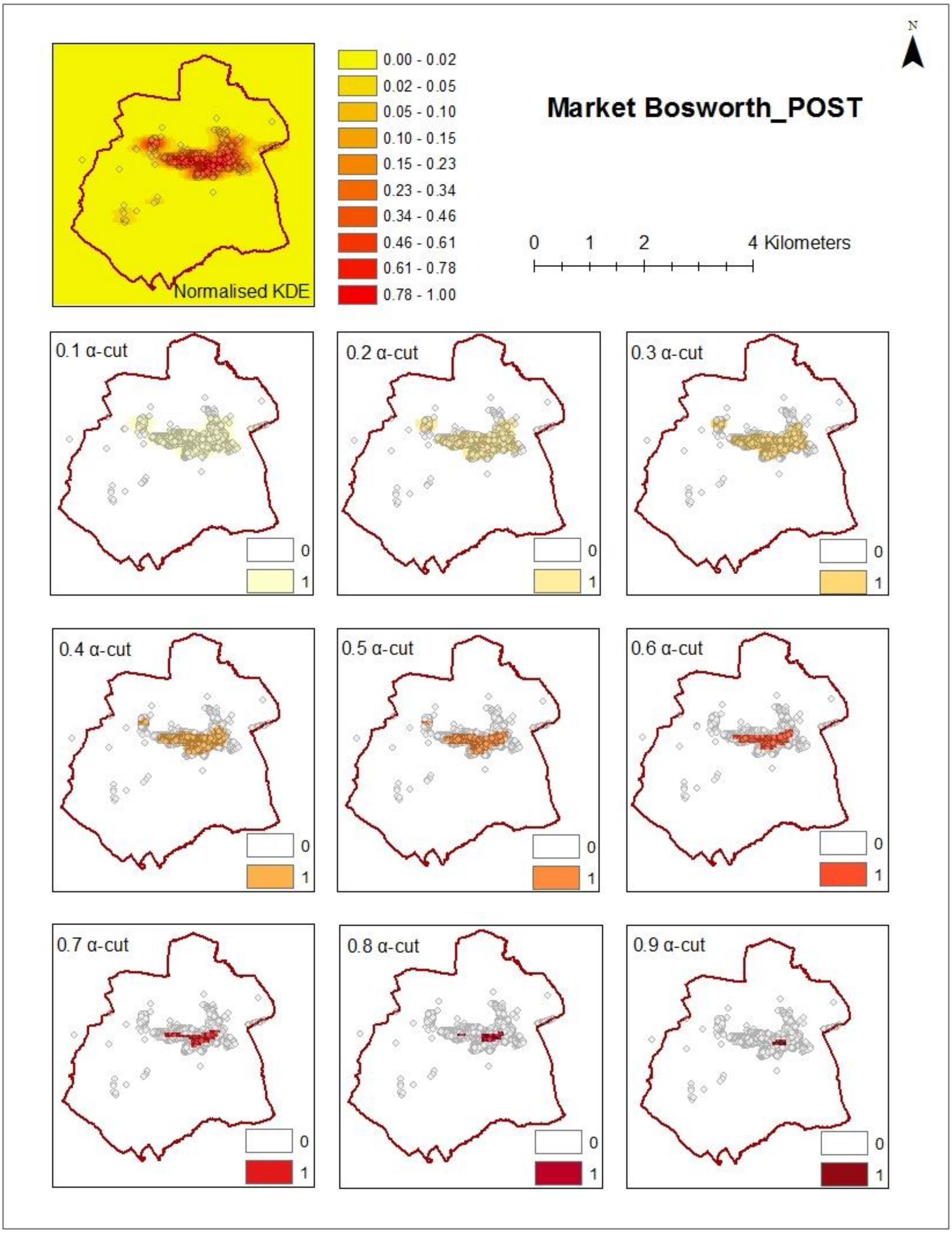


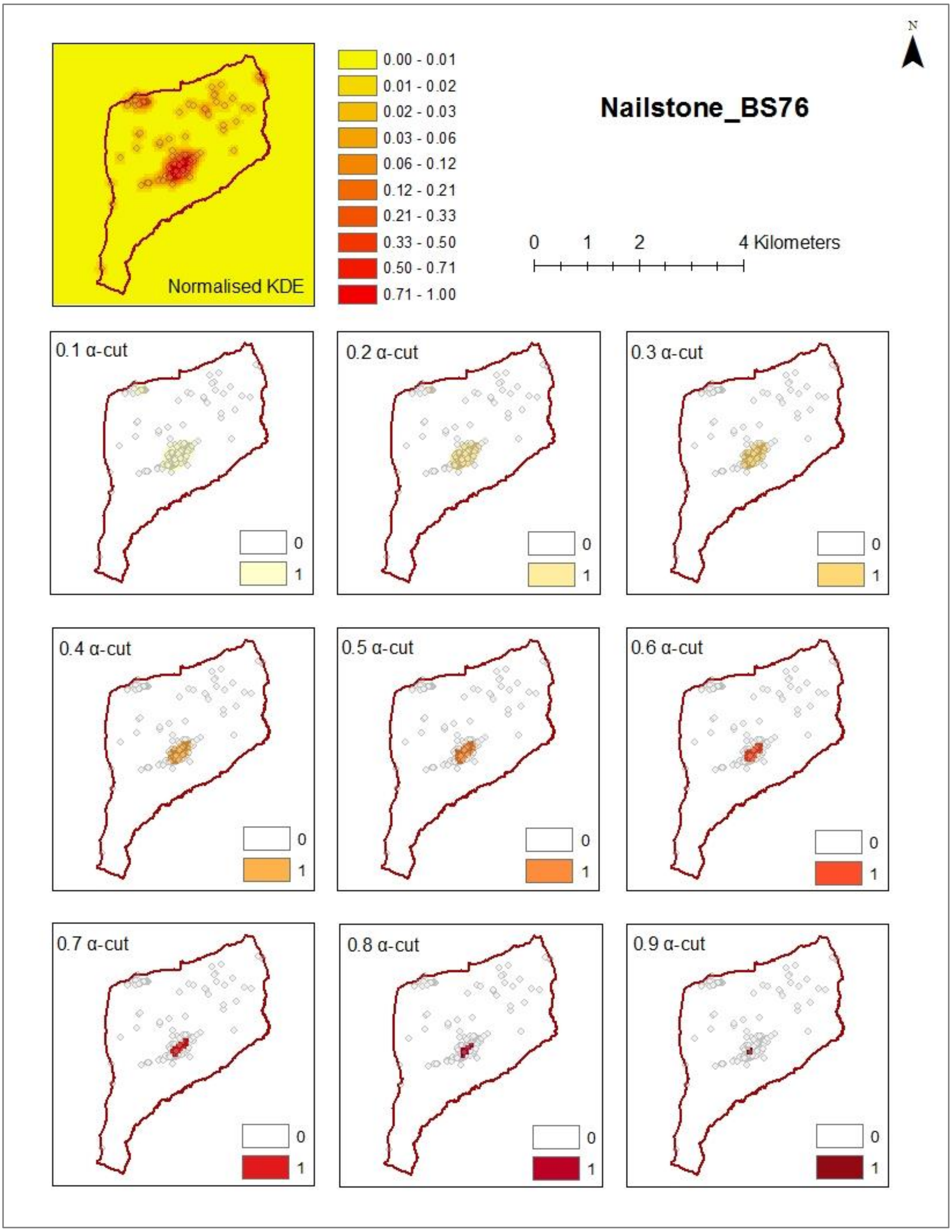
Higham on the Hill_POI
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.05
0.05 - 0.10
0.10 - 0.15
0.15 - 0.22
0.22 - 0.33
0.33 - 0.50
0.50 - 0.80
0.80 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


Higham on the Hill_POST
N
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.11
0.11 - 0.21
0.21 - 0.31
0.31 - 0.49
0.49 - 0.68
0.68 - 1.00
0 1.25 2.5 5 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

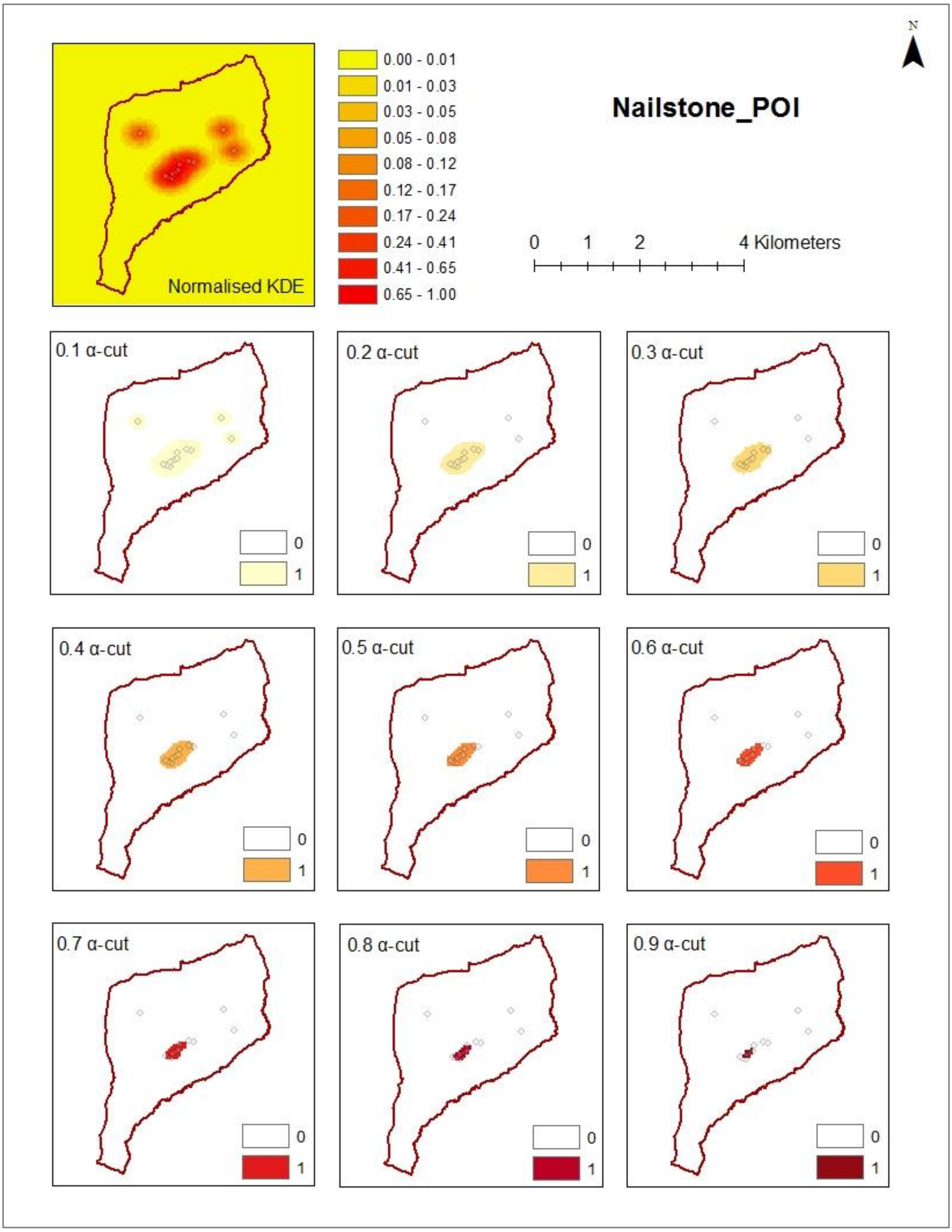

Market Bosworth_BS76
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.08
0.08 - 0.15
0.15 - 0.24
0.24 - 0.35
0.35 - 0.47
0.47 - 0.61
0.61 - 0.77
0.77 - 1.00
0 1 2 4 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

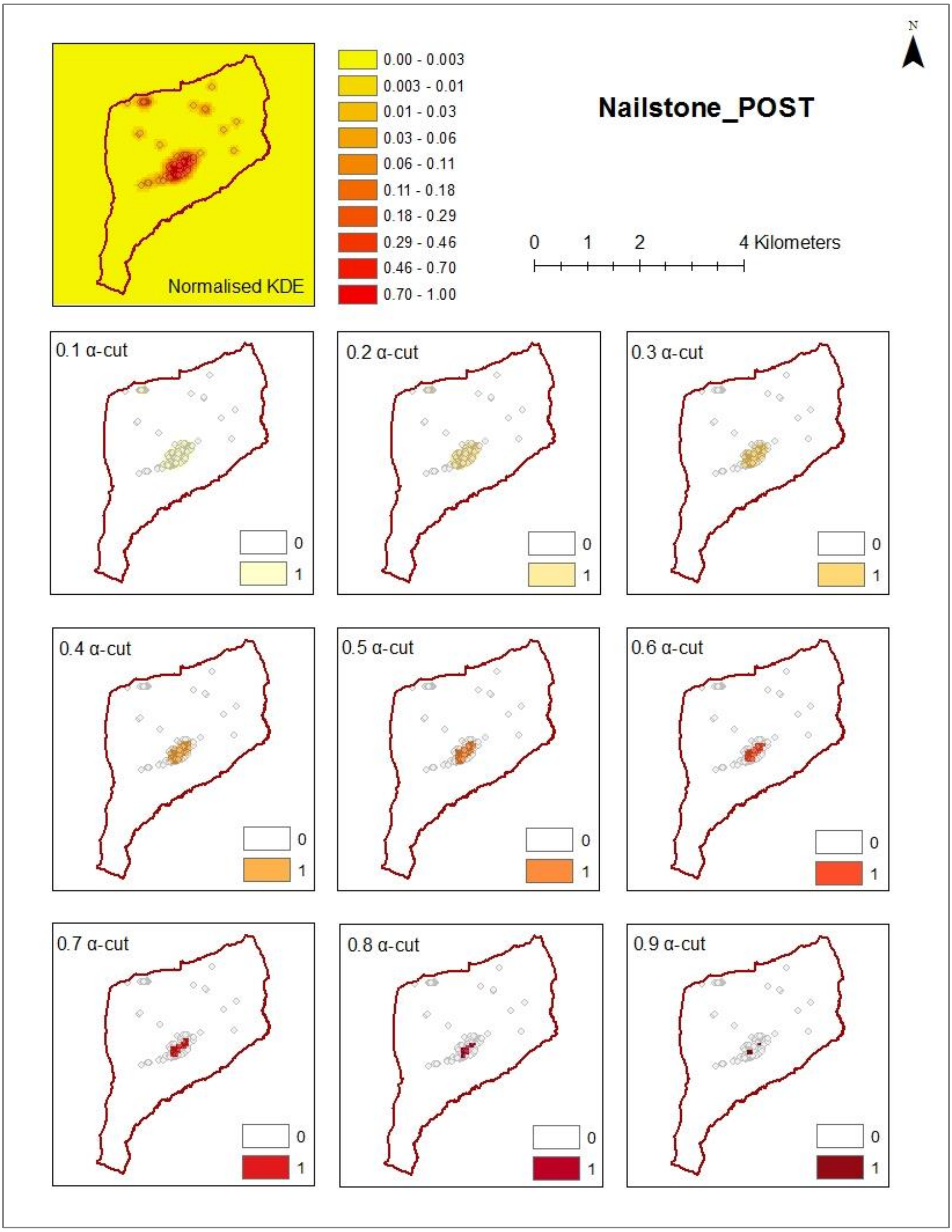

Market Bosworth_POI
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.05
0.05 - 0.08
0.08 - 0.13
0.13 - 0.18
0.18 - 0.27
0.27 - 0.42
0.42 - 0.67
0.67 - 1.00
0 1 2 4 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

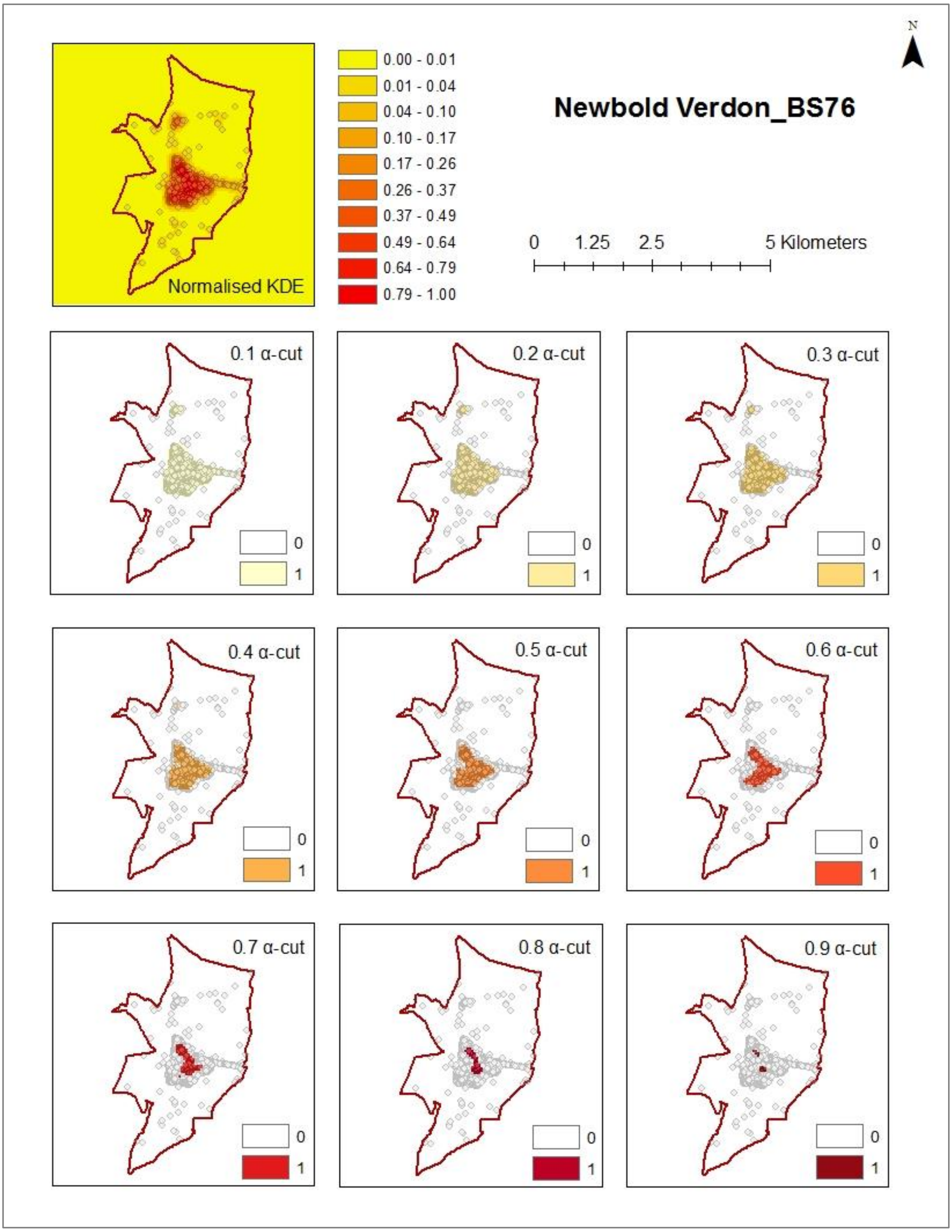

Market Bosworth_POST
Normalised KDE
0.00 - 0.02
0.02 - 0.05
0.05 - 0.10
0.10 - 0.15
0.15 - 0.23
0.23 - 0.34
0.34 - 0.46
0.46 - 0.61
0.61 - 0.78
0.78 - 1.00
0 1 2 4 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1



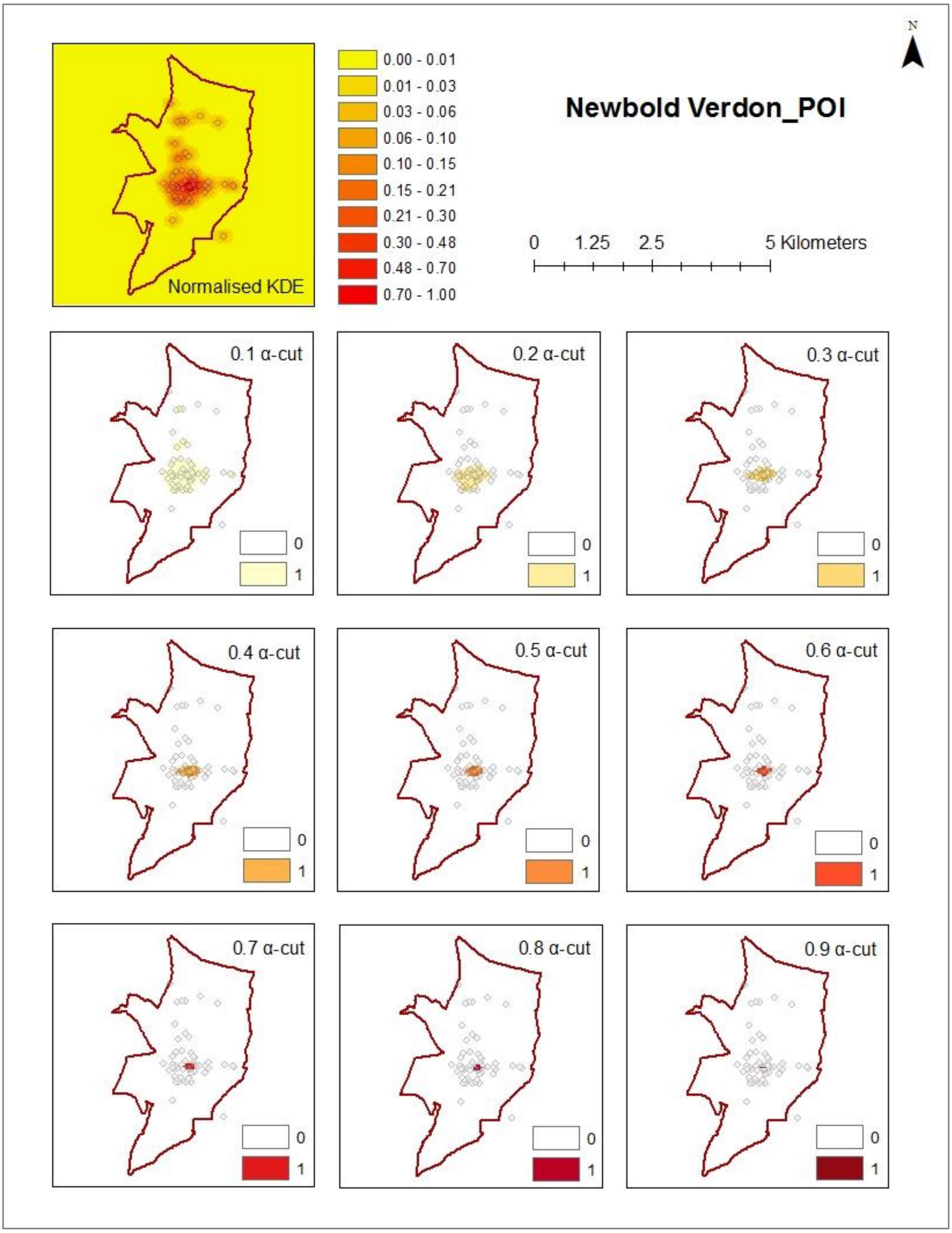
Nailstone_BS76
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.03
0.03 - 0.06
0.06 - 0.12
0.12 - 0.21
0.21 - 0.33
0.33 - 0.50
0.50 - 0.71
0.71 - 1.00
N
0 1 2 4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


Nailstone_POI
N
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.05
0.05 - 0.08
0.08 - 0.12
0.12 - 0.17
0.17 - 0.24
0.24 - 0.41
0.41 - 0.65
0.65 - 1.00
0 1 2 4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

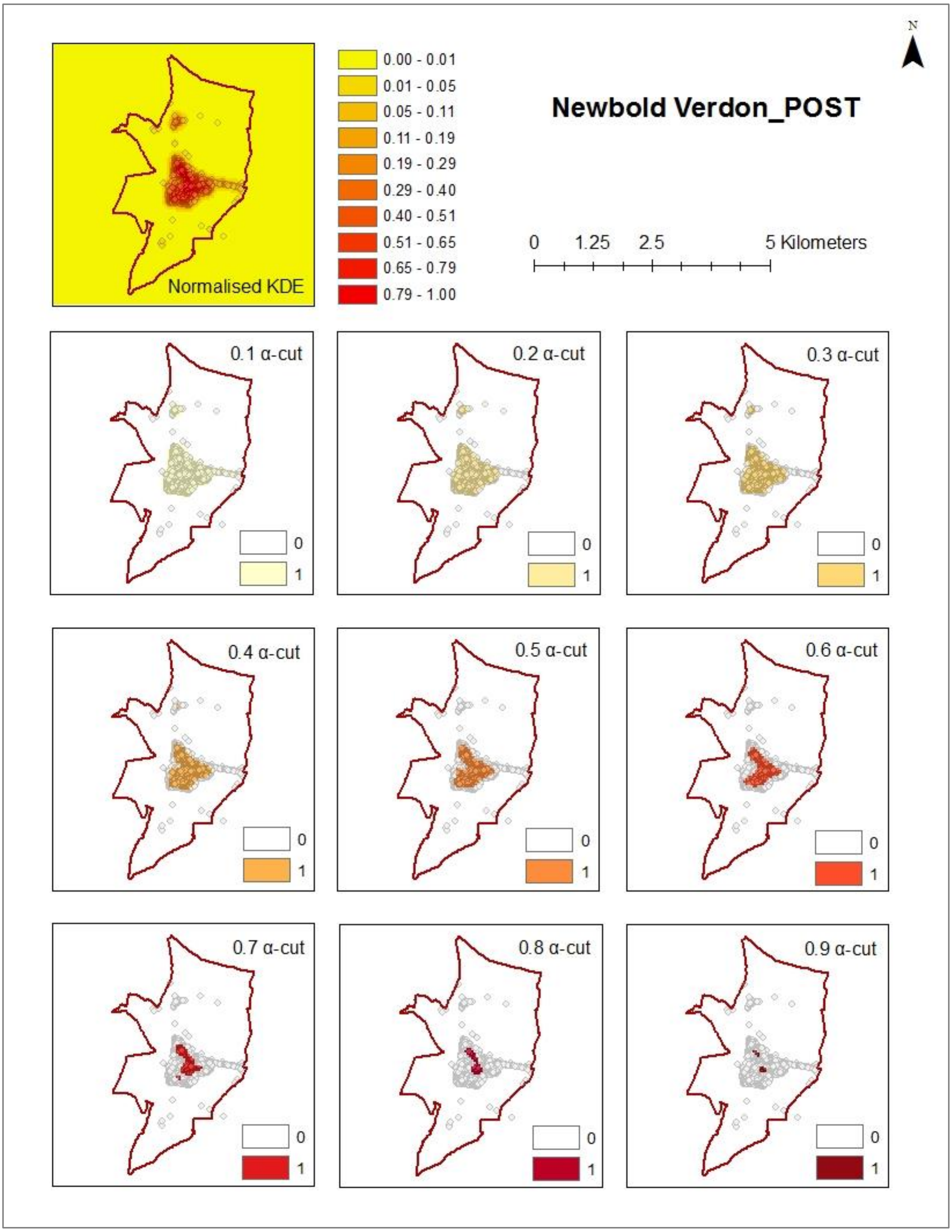

Nailstone_POST
N
Normalised KDE
0.00 - 0.003
0.003 - 0.01
0.01 - 0.03
0.03 - 0.06
0.06 - 0.11
0.11 - 0.18
0.18 - 0.29
0.29 - 0.46
0.46 - 0.70
0.70 - 1.00
0 1 2 4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

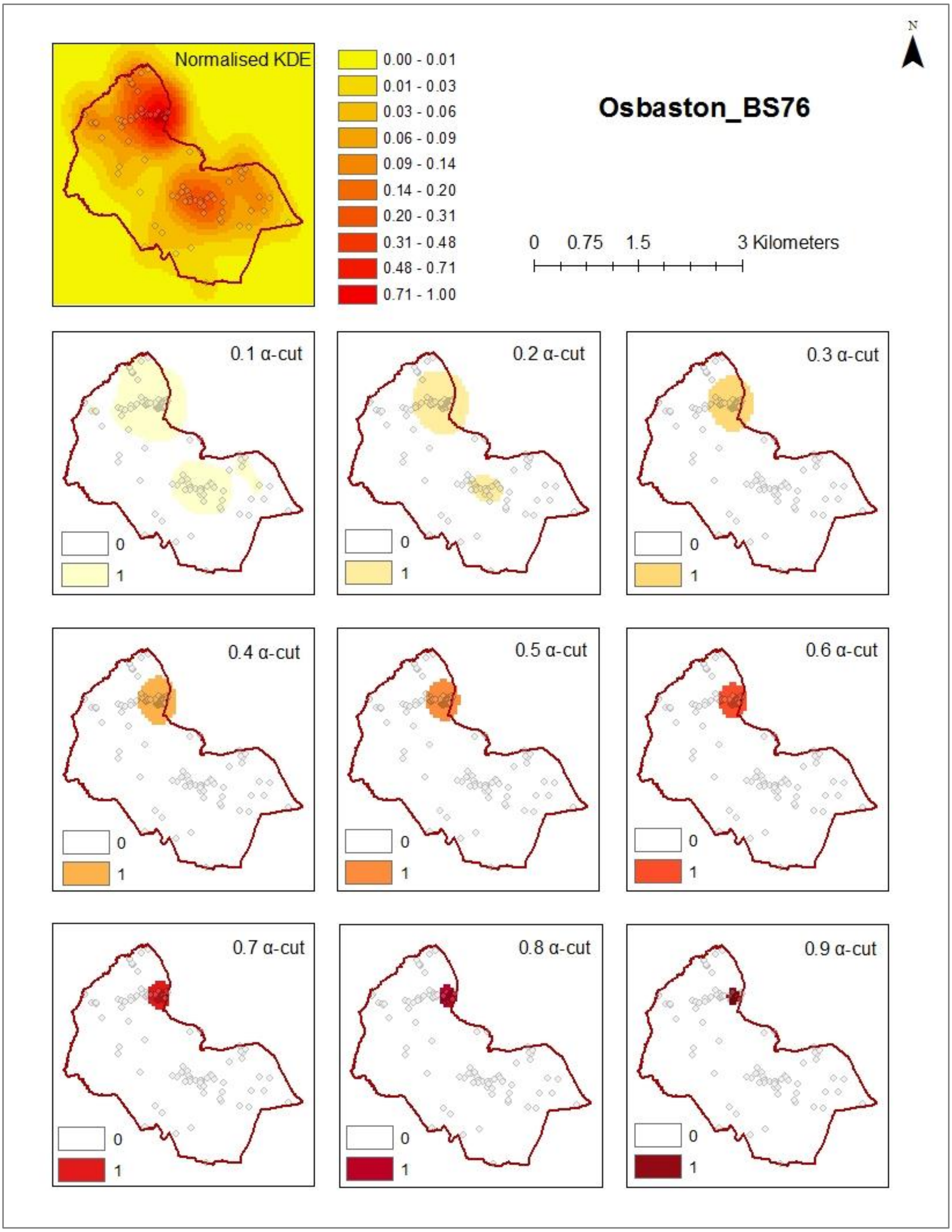


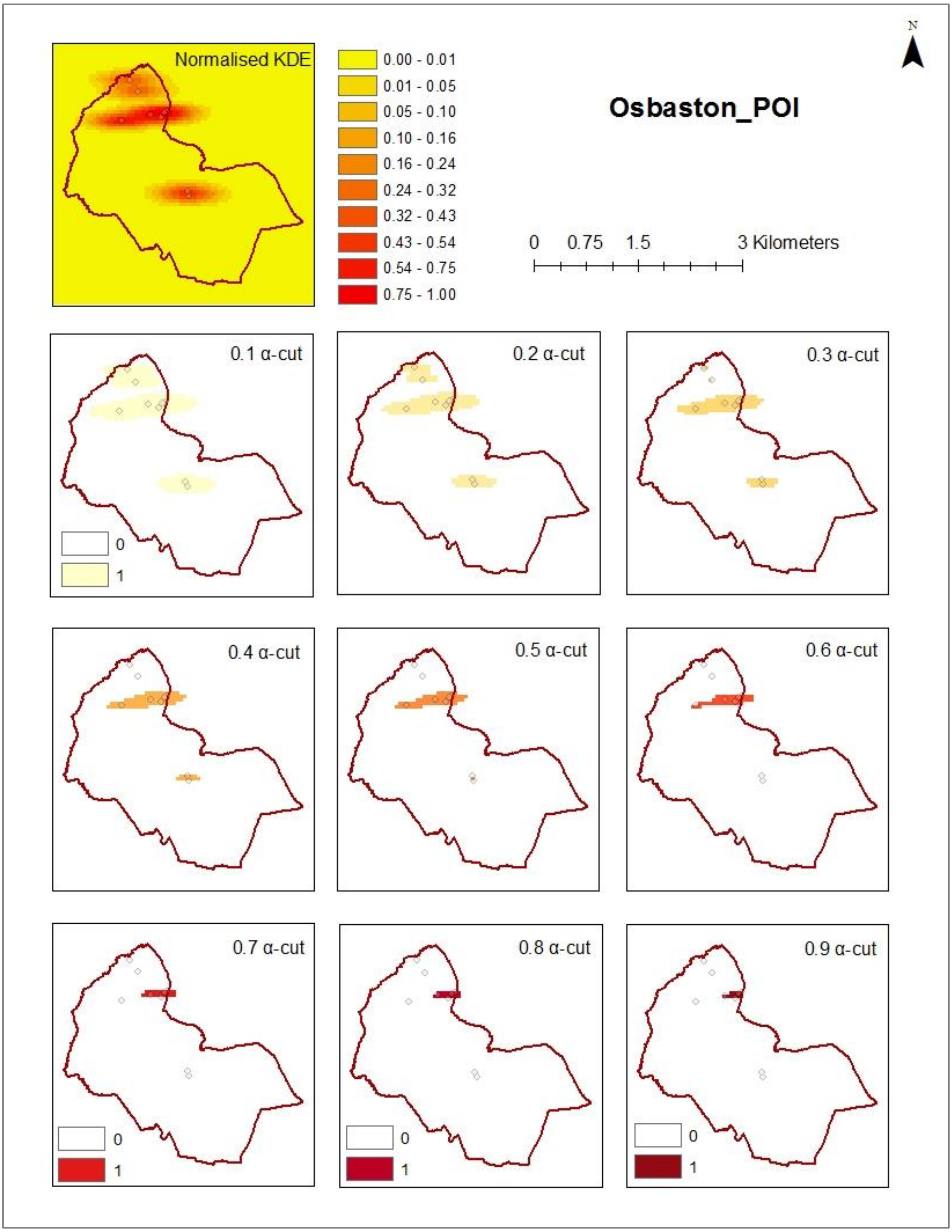
Newbold Verdon_BS76
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.10
0.10 - 0.17
0.17 - 0.26
0.26 - 0.37
0.37 - 0.49
0.49 - 0.64
0.64 - 0.79
0.79 - 1.00
0
1.25
2.5
5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

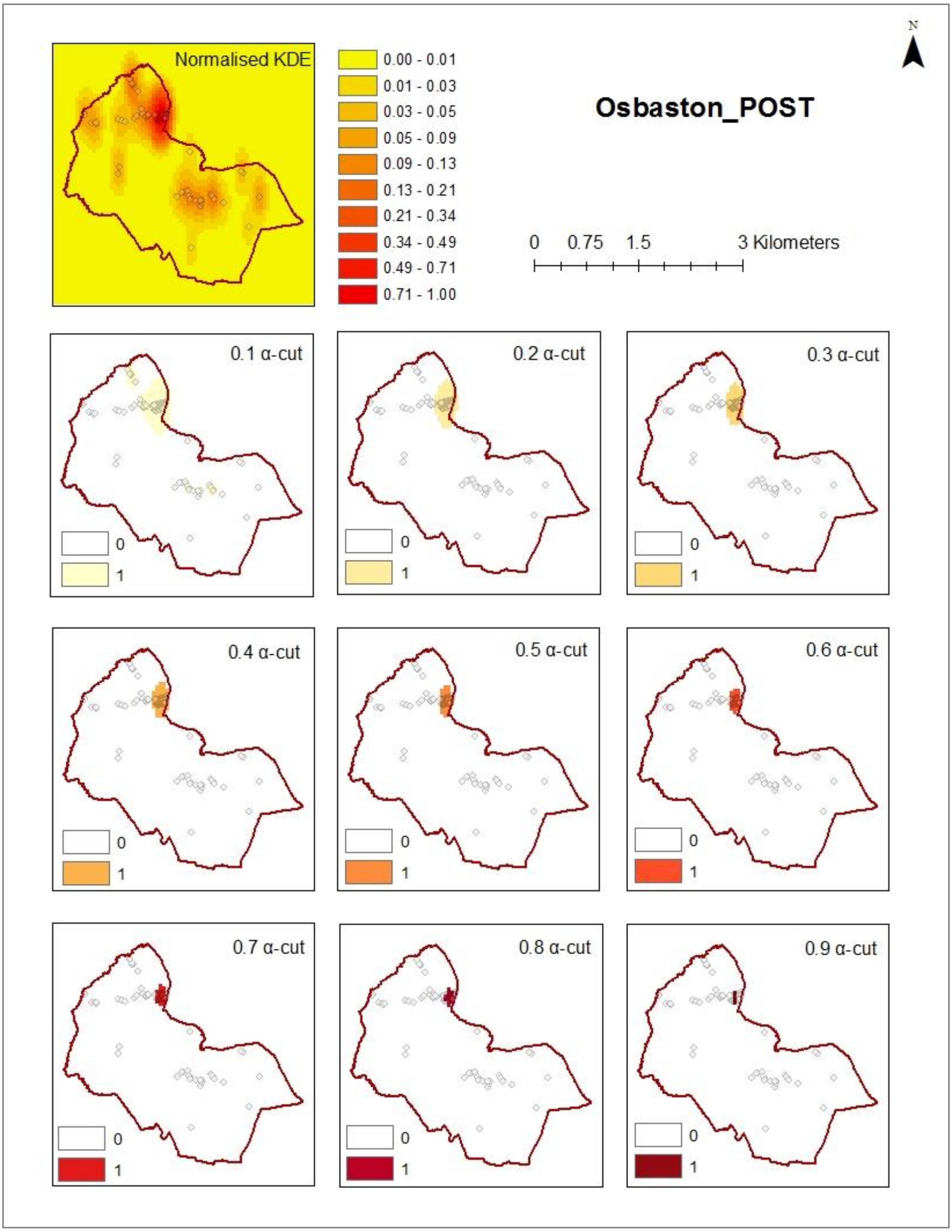


Newbold Verdon_POI
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.06
0.06 - 0.10
0.10 - 0.15
0.15 - 0.21
0.21 - 0.30
0.30 - 0.48
0.48 - 0.70
0.70 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

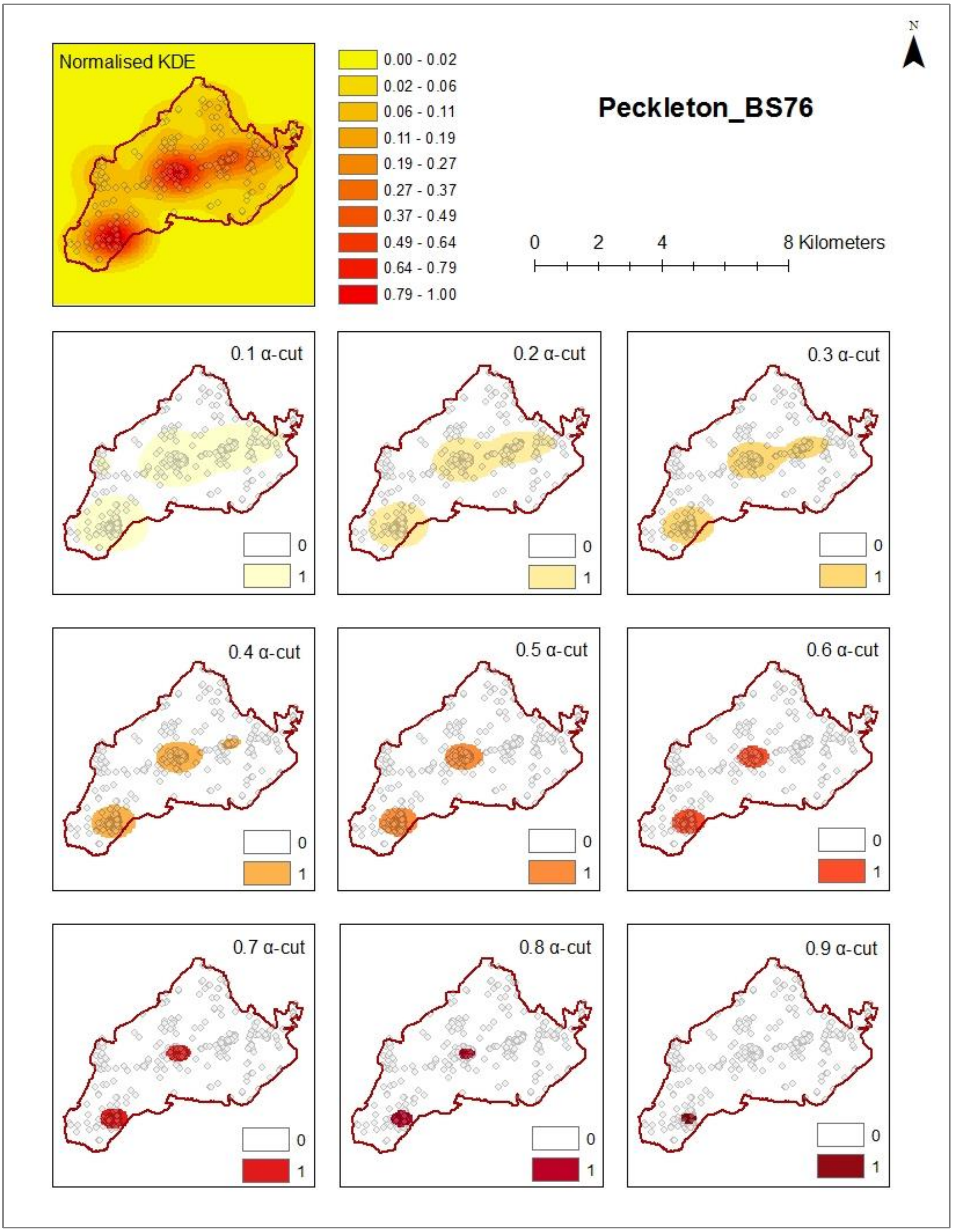

Newbold Verdon_POST
Normalised KDE
0.00 - 0.01
0.01 - 0.05
0.05 - 0.11
0.11 - 0.19
0.19 - 0.29
0.29 - 0.40
0.40 - 0.51
0.51 - 0.65
0.65 - 0.79
0.79 - 1.00
0 1.25 2.5 5 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

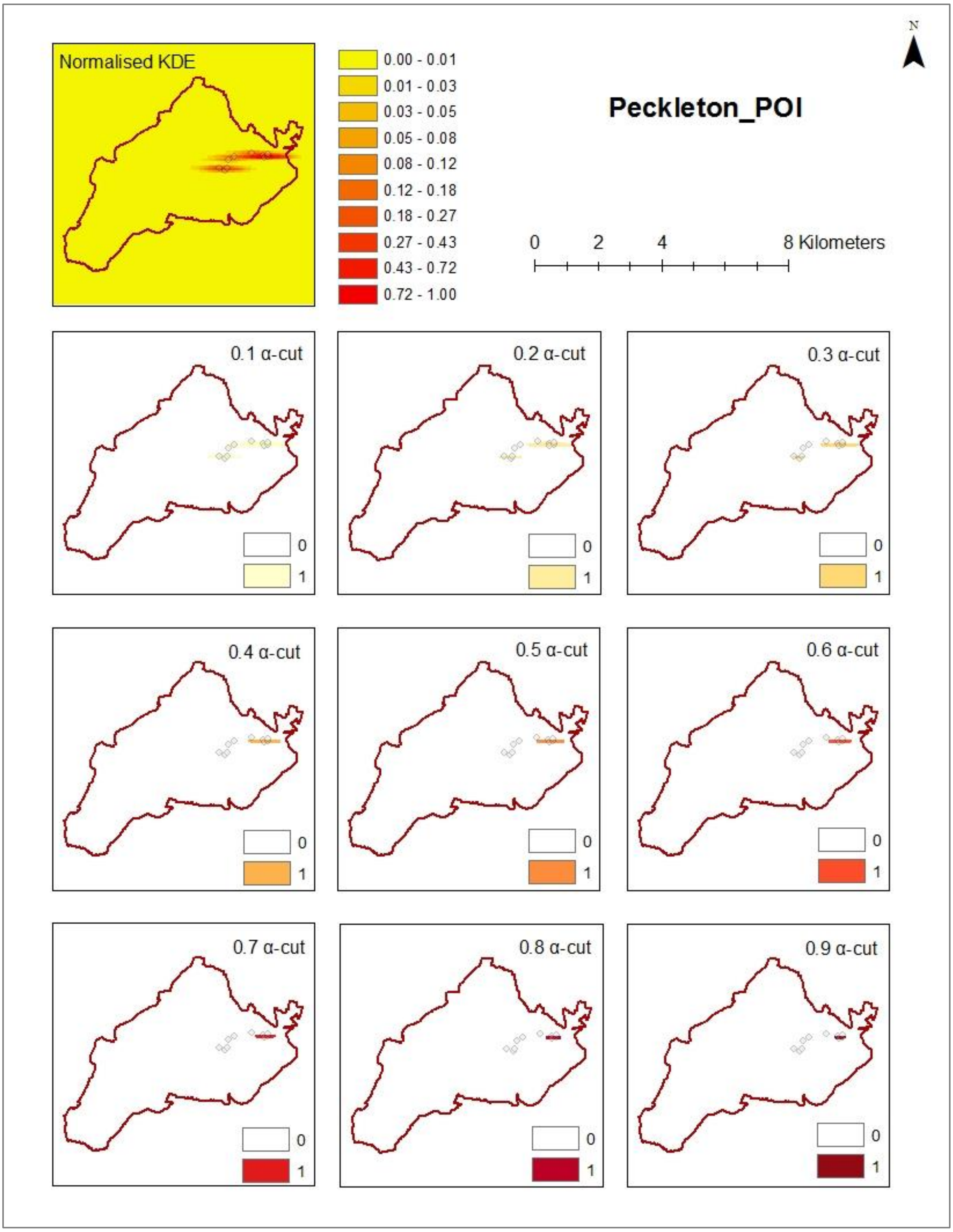

Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.06
0.06 - 0.09
0.09 - 0.14
0.14 - 0.20
0.20 - 0.31
0.31 - 0.48
0.48 - 0.71
0.71 - 1.00
Osbaston_BS76
0 0.75 1.5 3 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

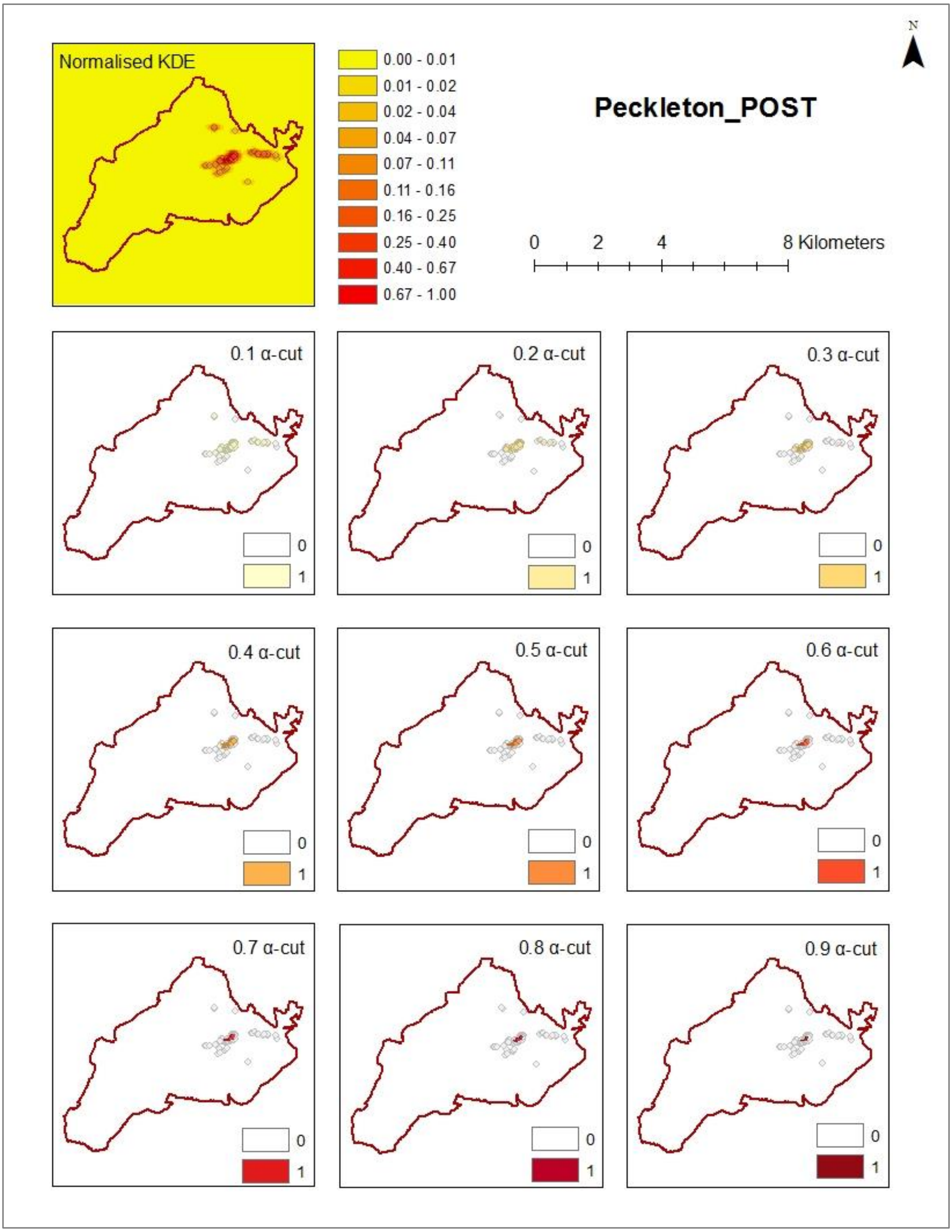

N
Normalised KDE
0.00 - 0.01
0.01 - 0.05
0.05 - 0.10
0.10 - 0.16
0.16 - 0.24
0.24 - 0.32
0.32 - 0.43
0.43 - 0.54
0.54 - 0.75
0.75 - 1.00
Osbaston_POI
0 0.75 1.5 3 Kilometers
0.1 α-cut
0
1
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0
1
0.8 α-cut
0
1
0.9 α-cut
0
1



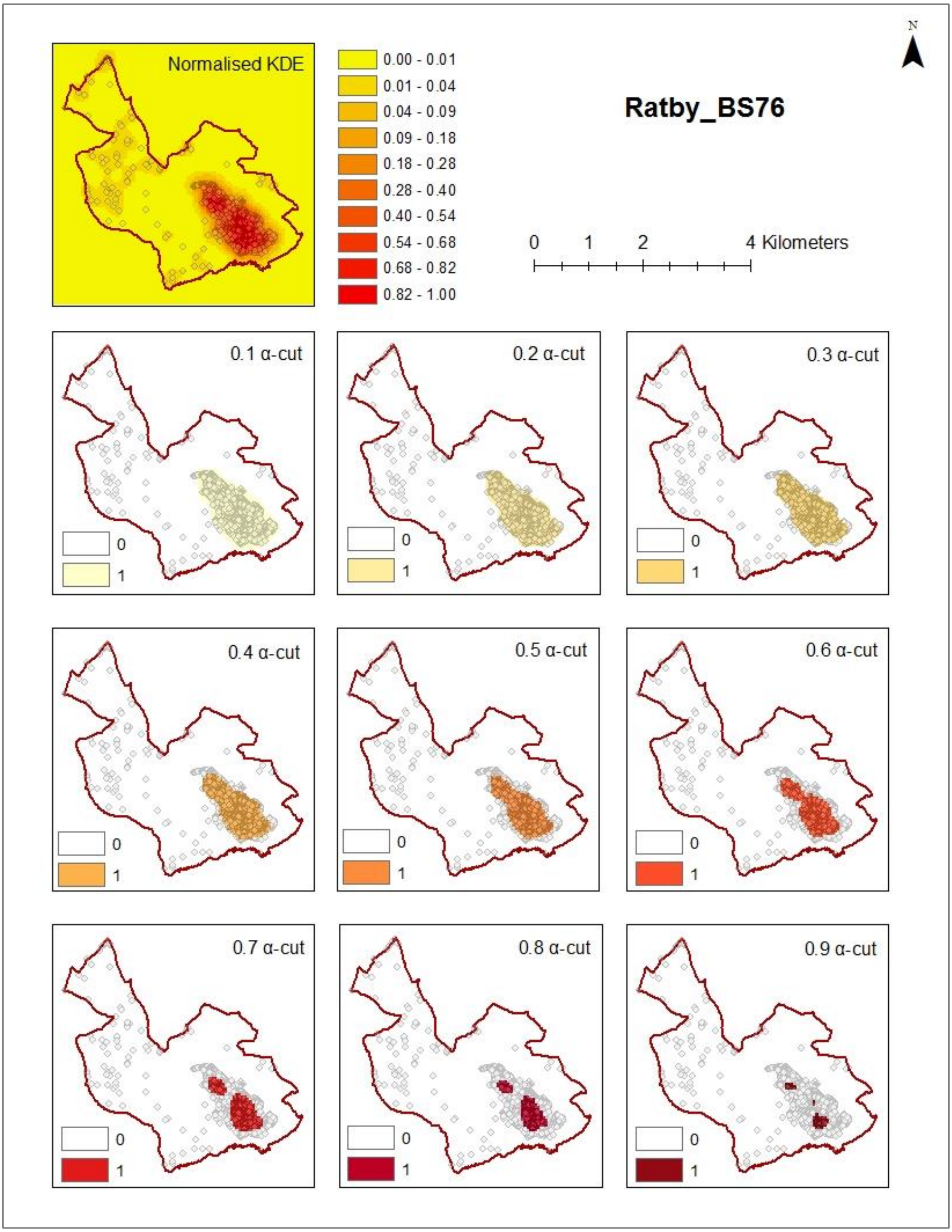
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.05
0.05 - 0.09
0.09 - 0.13
0.13 - 0.21
0.21 - 0.34
0.34 - 0.49
0.49 - 0.71
0.71 - 1.00
N
Osbaston_POST
0 0.75 1.5 3 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

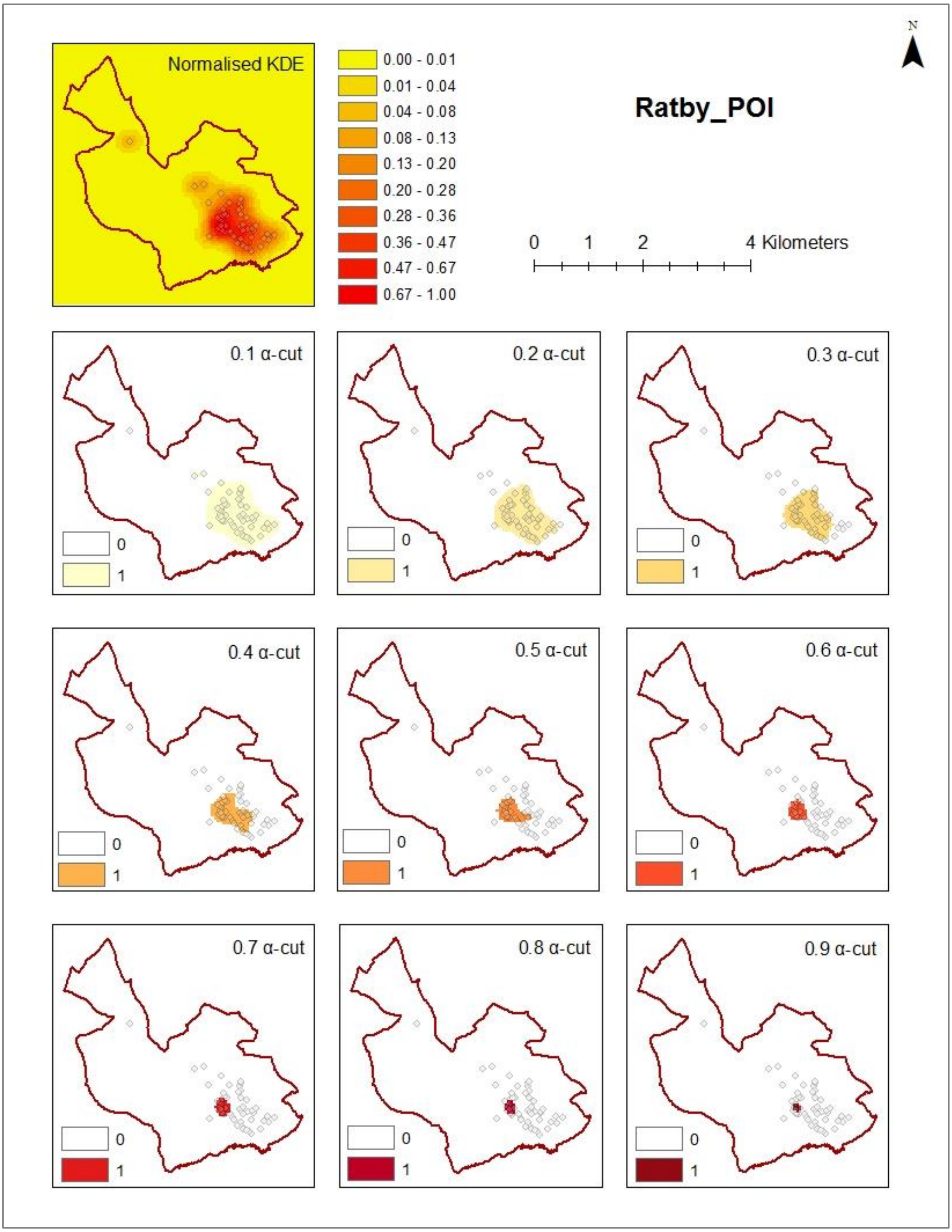

Peckleton_BS76
Normalised KDE
0.00 - 0.02
0.02 - 0.06
0.06 - 0.11
0.11 - 0.19
0.19 - 0.27
0.27 - 0.37
0.37 - 0.49
0.49 - 0.64
0.64 - 0.79
0.79 - 1.00
0
2
4
8 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.05
0.05 - 0.08
0.08 - 0.12
0.12 - 0.18
0.18 - 0.27
0.27 - 0.43
0.43 - 0.72
0.72 - 1.00
Peckleton_POI
N
0 2 4 8 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

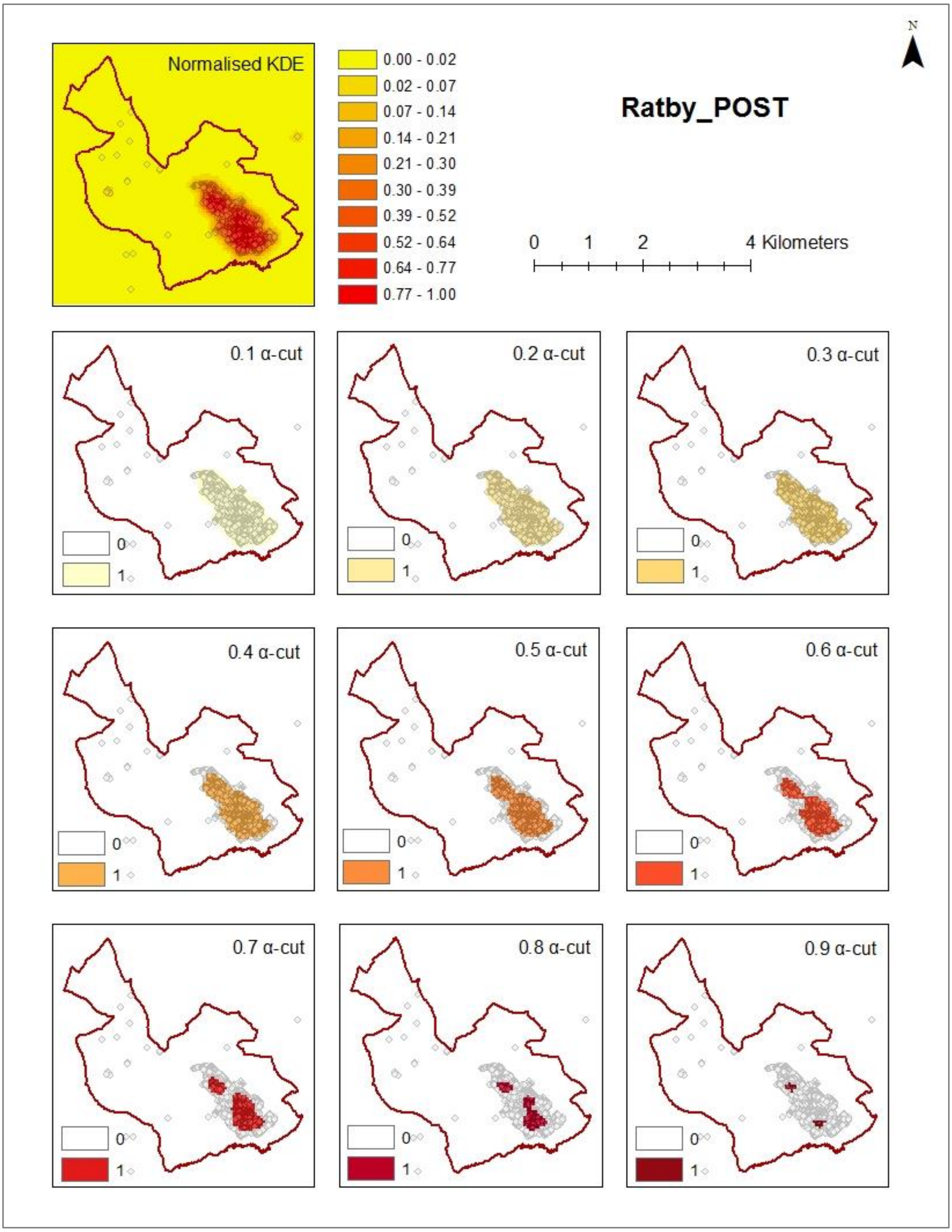


Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.11
0.11 - 0.16
0.16 - 0.25
0.25 - 0.40
0.40 - 0.67
0.67 - 1.00
N
Peckleton_POST
0
2
4
8 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

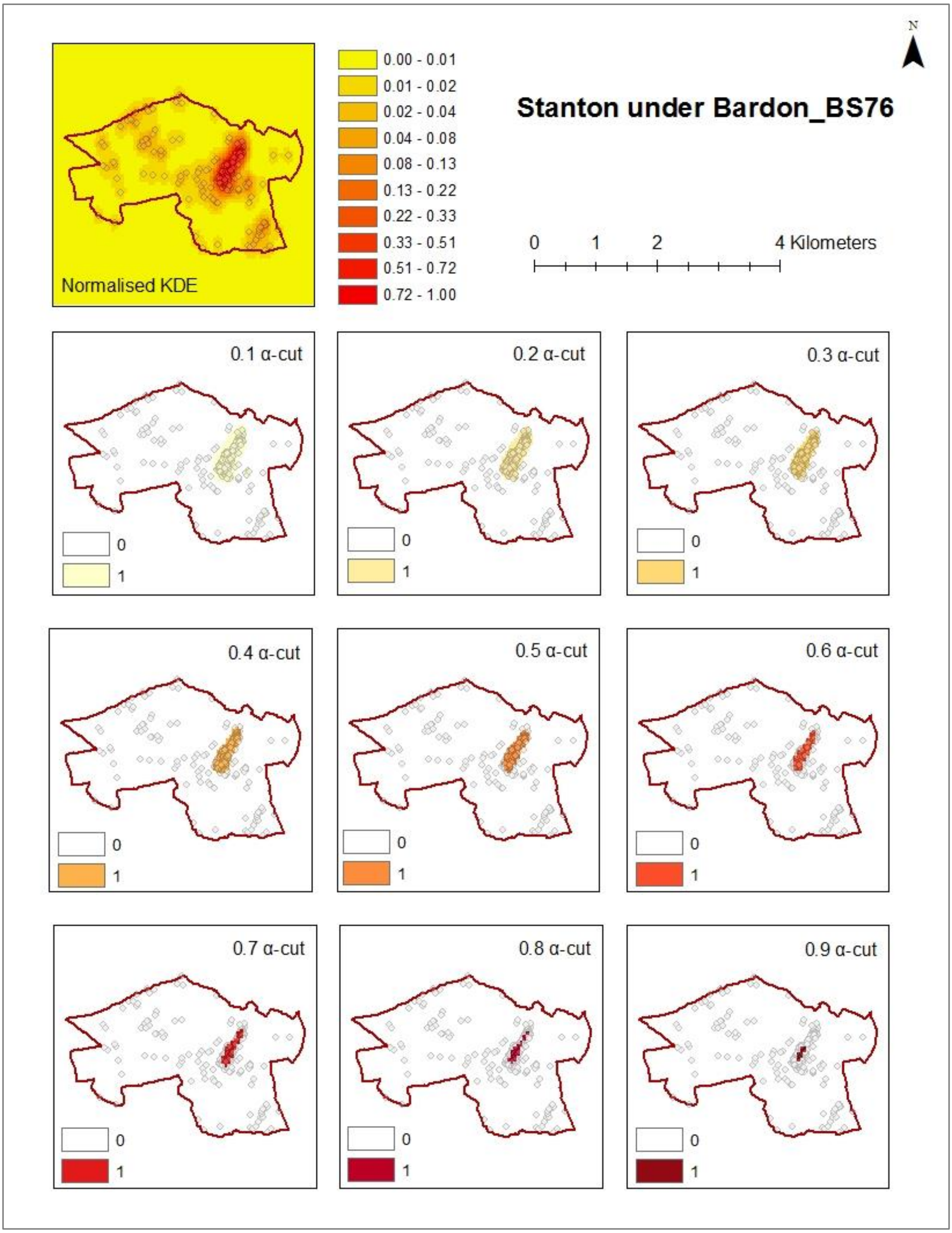

Ratby_BS76
N
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.09
0.09 - 0.18
0.18 - 0.28
0.28 - 0.40
0.40 - 0.54
0.54 - 0.68
0.68 - 0.82
0.82 - 1.00
0
1
2
4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

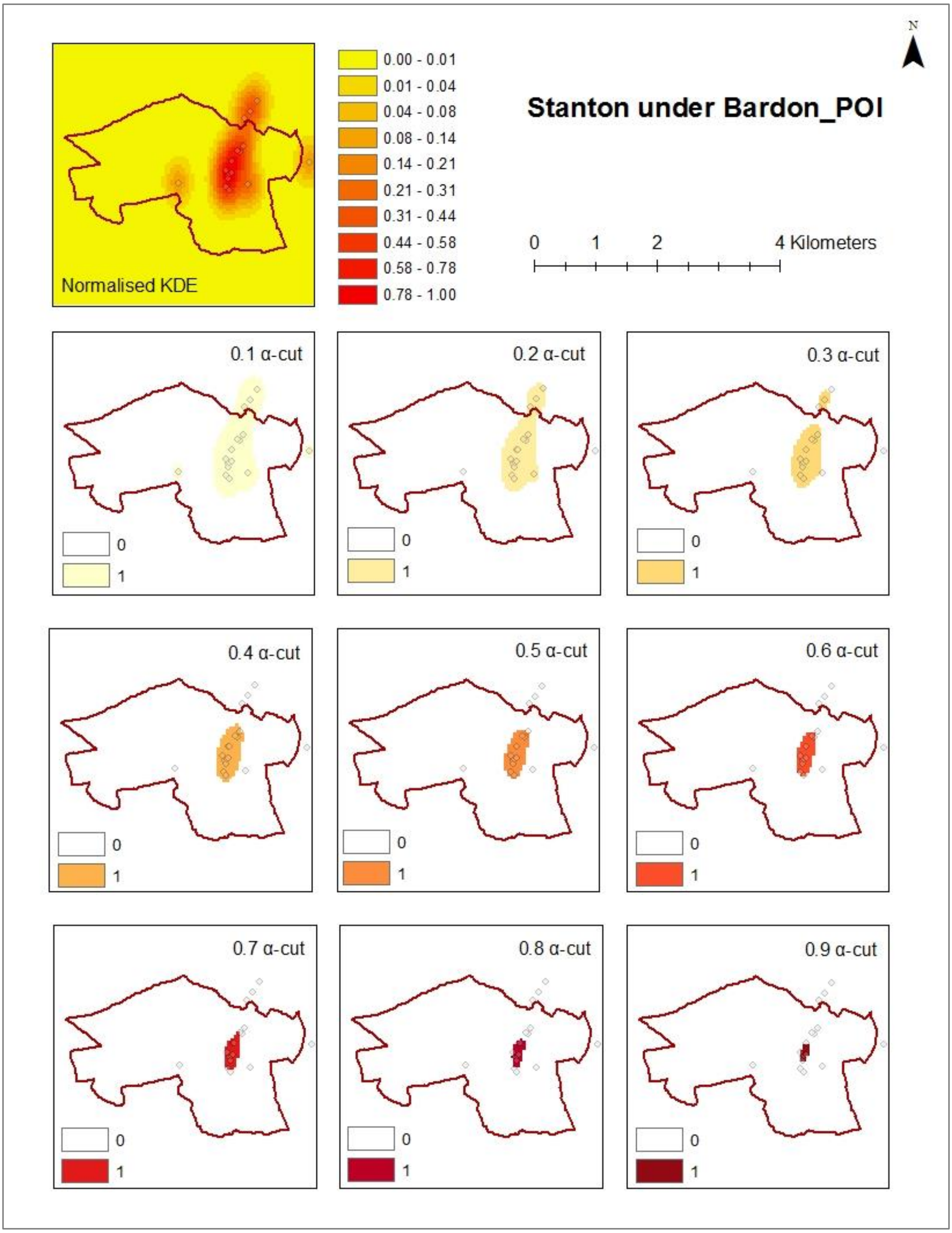


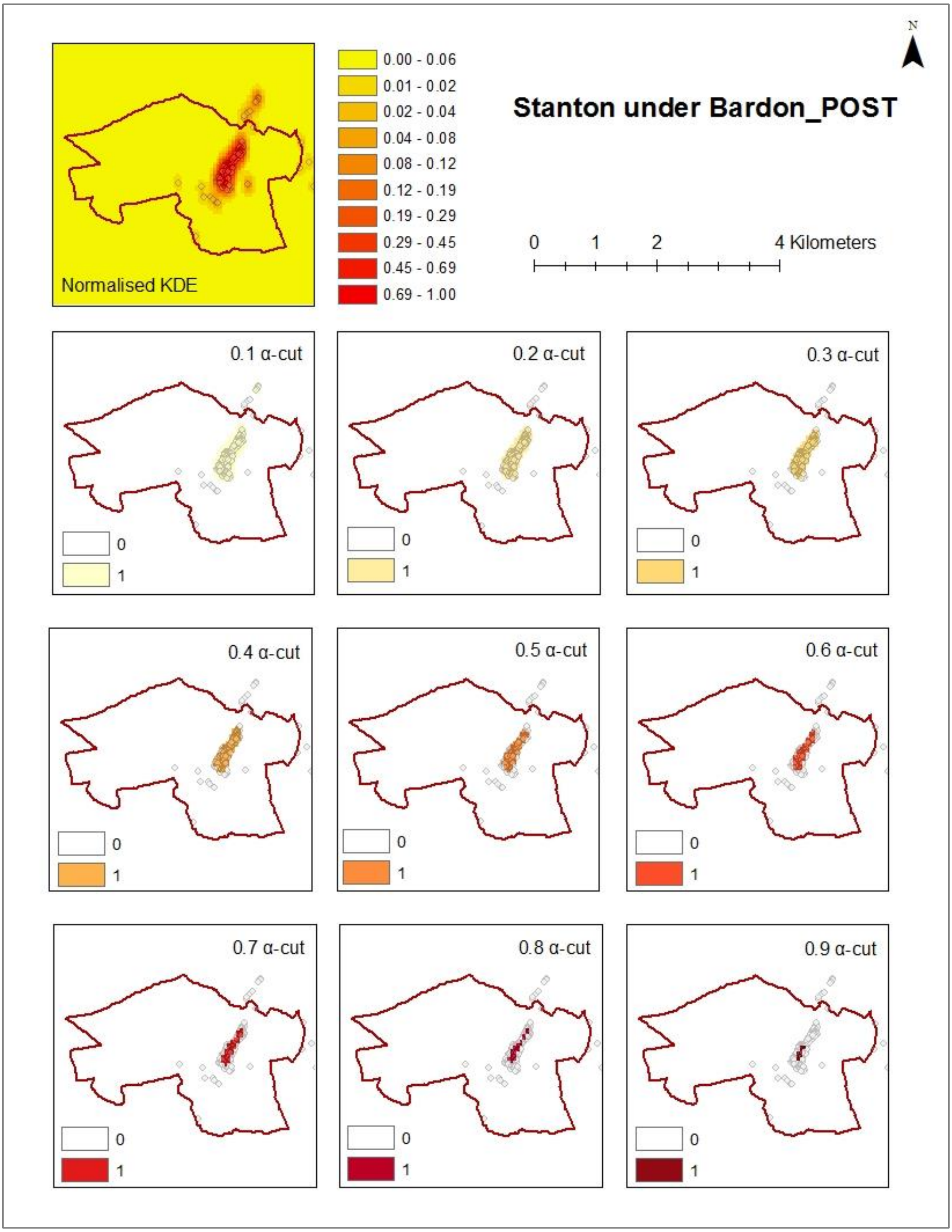
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.08
0.08 - 0.13
0.13 - 0.20
0.20 - 0.28
0.28 - 0.36
0.36 - 0.47
0.47 - 0.67
0.67 - 1.00
Ratby_POI
N
0 1 2 4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

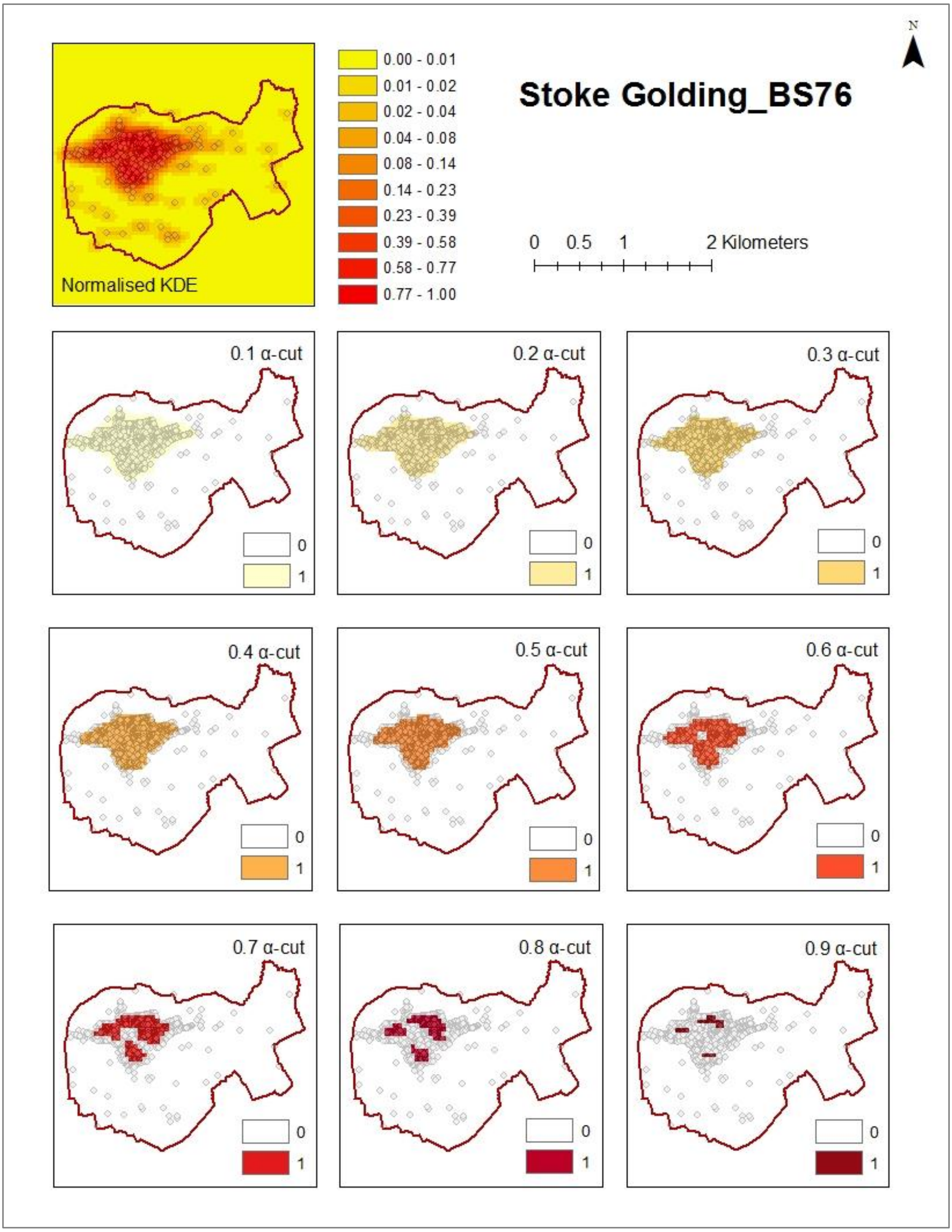


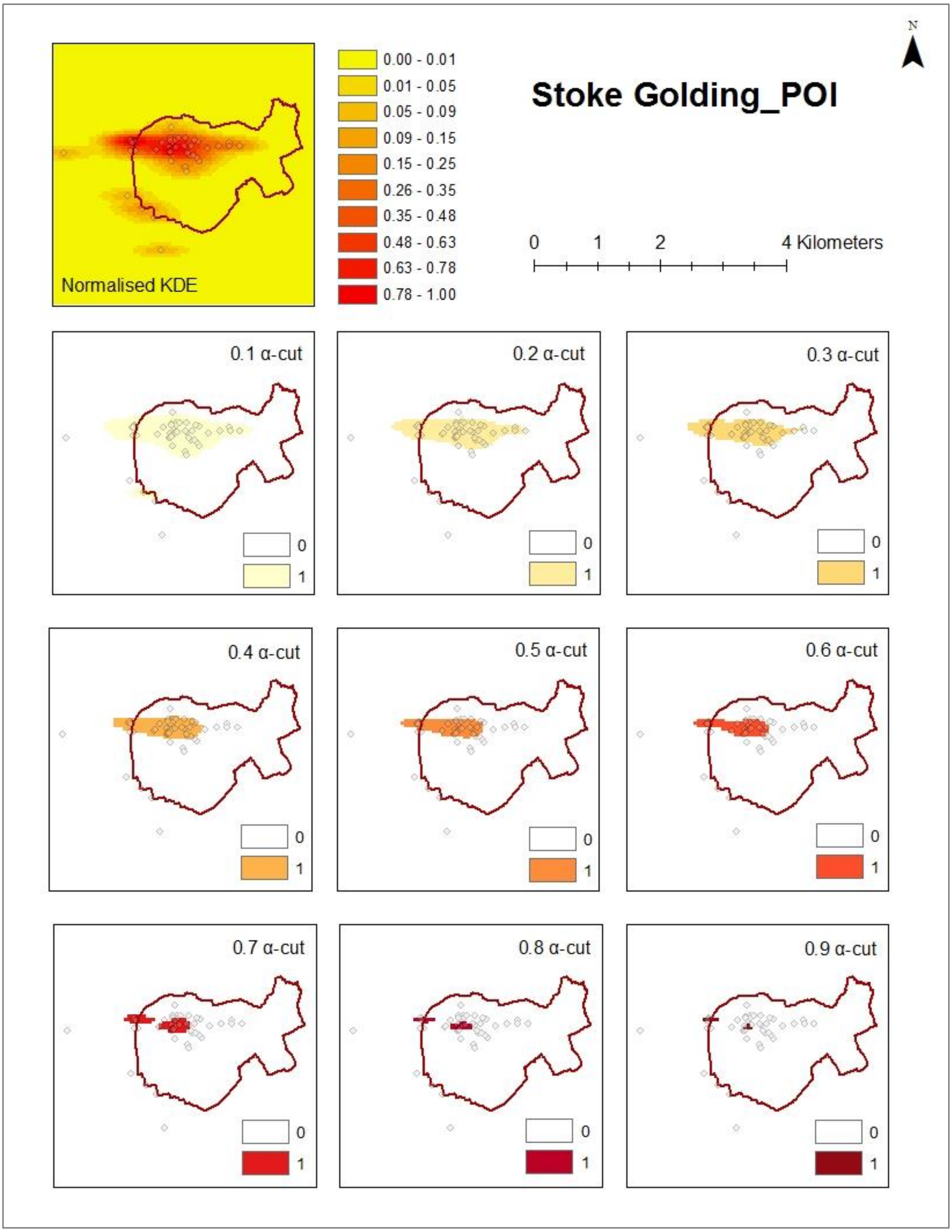
Normalised KDE
0.00 - 0.02
0.02 - 0.07
0.07 - 0.14
0.14 - 0.21
0.21 - 0.30
0.30 - 0.39
0.39 - 0.52
0.52 - 0.64
0.64 - 0.77
0.77 - 1.00
Ratby_POST
N
0 1 2 4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


N
Stanton under Bardon_BS76
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.08
0.08 - 0.13
0.13 - 0.22
0.22 - 0.33
0.33 - 0.51
0.51 - 0.72
0.72 - 1.00
0
1
2
4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

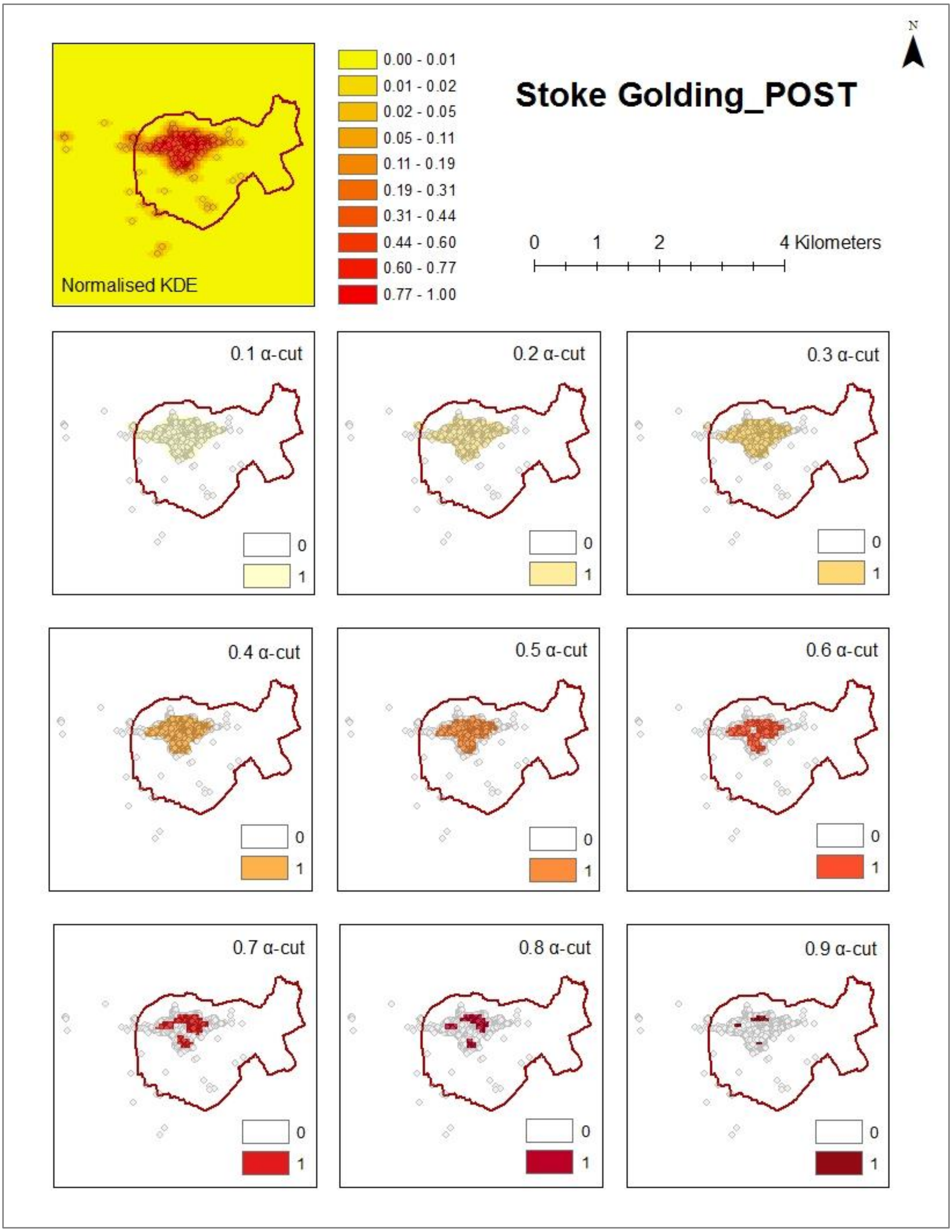

Stanton under Bardon_POI
Normalised KDE
0.00 - 0.01
0.01 - 0.04
0.04 - 0.08
0.08 - 0.14
0.14 - 0.21
0.21 - 0.31
0.31 - 0.44
0.44 - 0.58
0.58 - 0.78
0.78 - 1.00
0 1 2 4 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


Stanton under Bardon_POST
Normalised KDE
0.00 - 0.06
0.01 - 0.02
0.02 - 0.04
0.04 - 0.08
0.08 - 0.12
0.12 - 0.19
0.19 - 0.29
0.29 - 0.45
0.45 - 0.69
0.69 - 1.00
0
1
2
4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

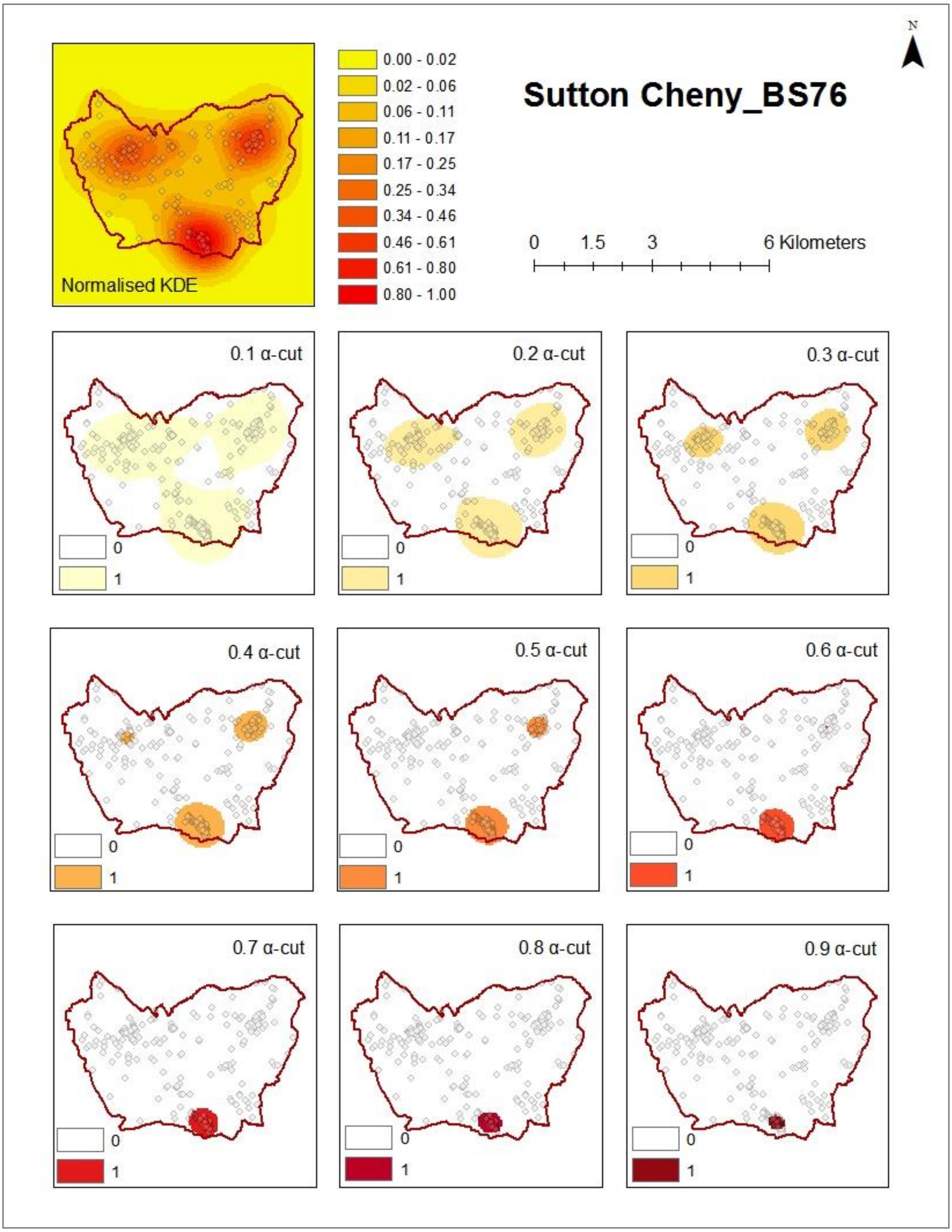

N
Stoke Golding_BS76
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.08
0.08 - 0.14
0.14 - 0.23
0.23 - 0.39
0.39 - 0.58
0.58 - 0.77
0.77 - 1.00
0 0.5 1 2 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

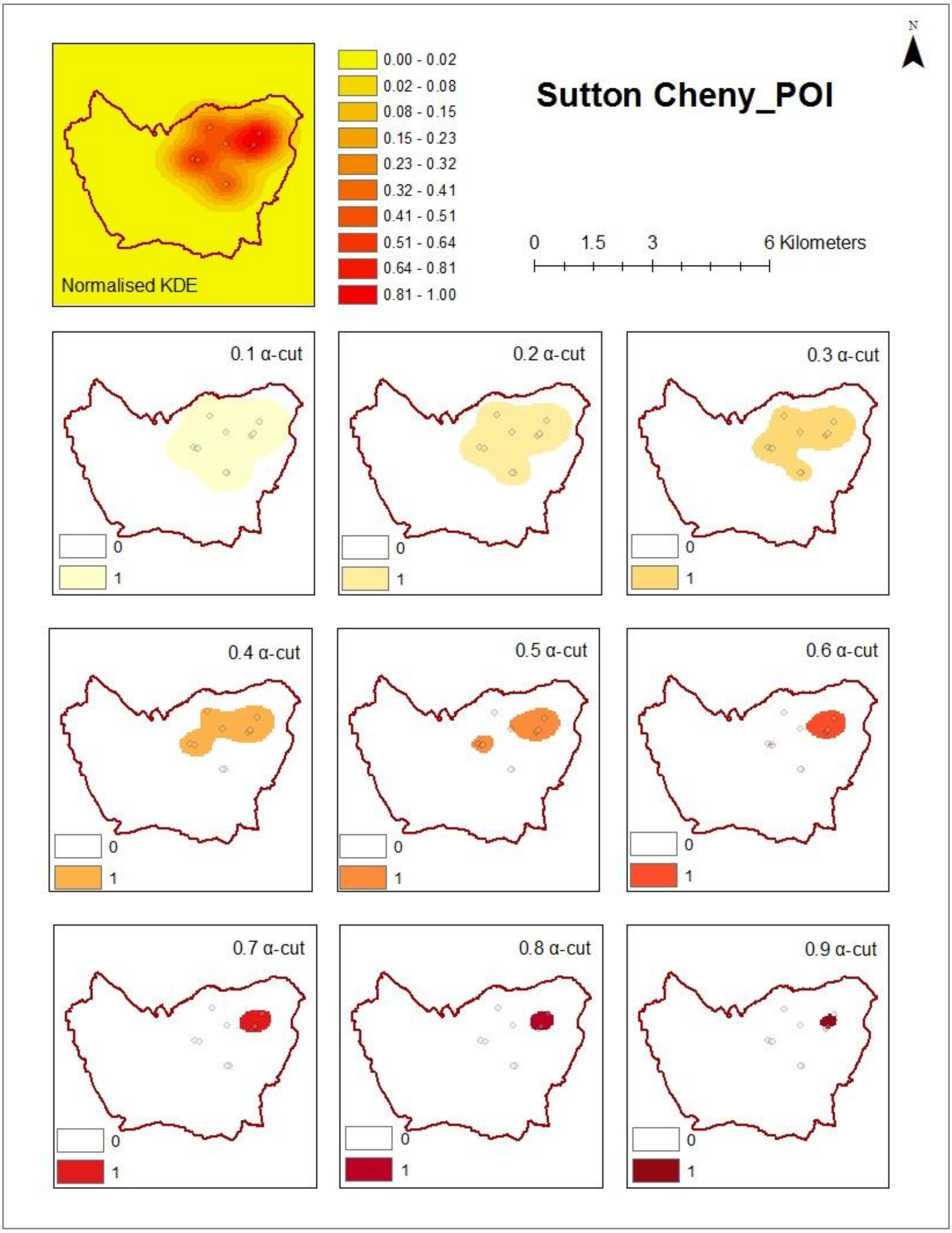


Stoke Golding_POI
N
Normalised KDE
0.00 - 0.01
0.01 - 0.05
0.05 - 0.09
0.09 - 0.15
0.15 - 0.25
0.26 - 0.35
0.35 - 0.48
0.48 - 0.63
0.63 - 0.78
0.78 - 1.00
0
1
2
4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut

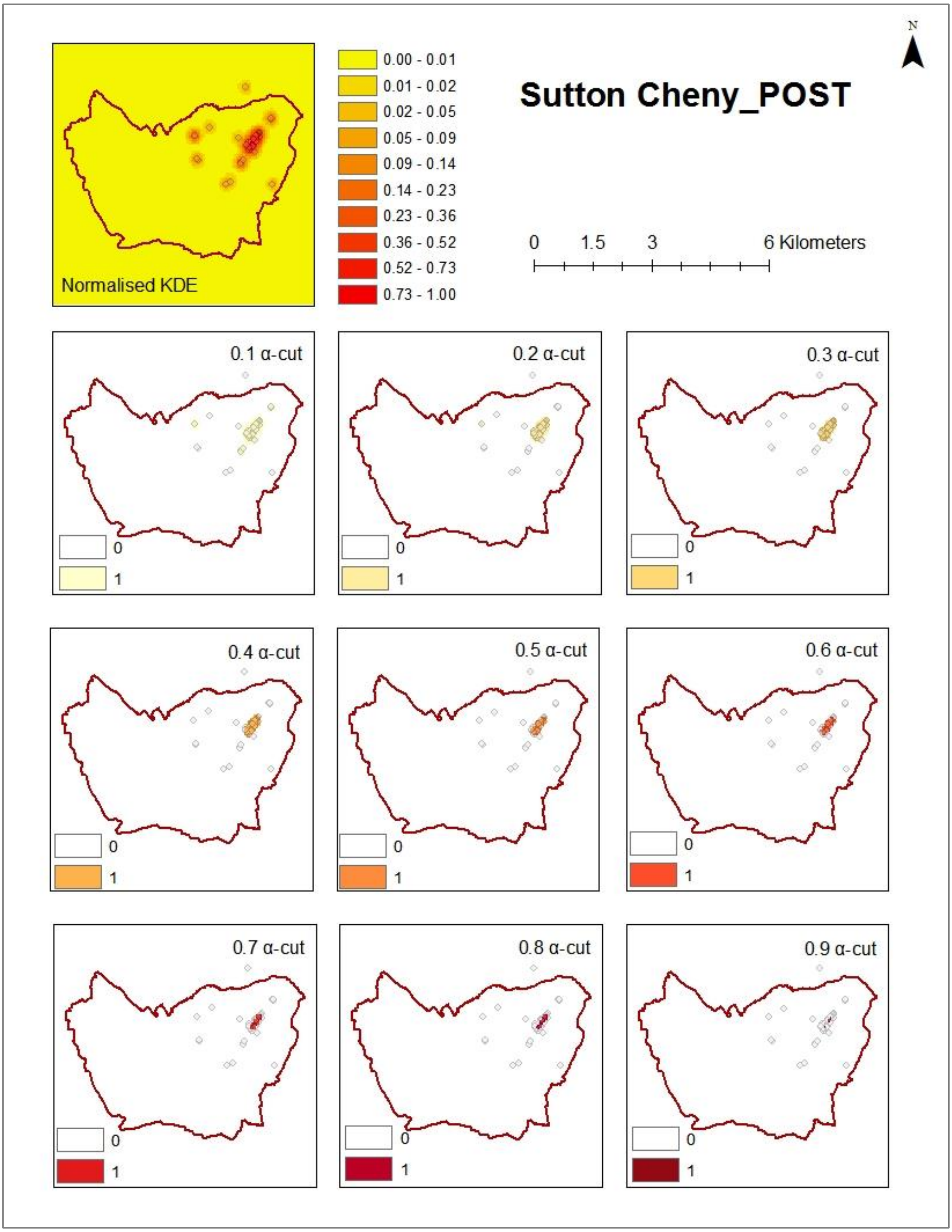


Stoke Golding_POST
N
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.05
0.05 - 0.11
0.11 - 0.19
0.19 - 0.31
0.31 - 0.44
0.44 - 0.60
0.60 - 0.77
0.77 - 1.00
0 1 2 4 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

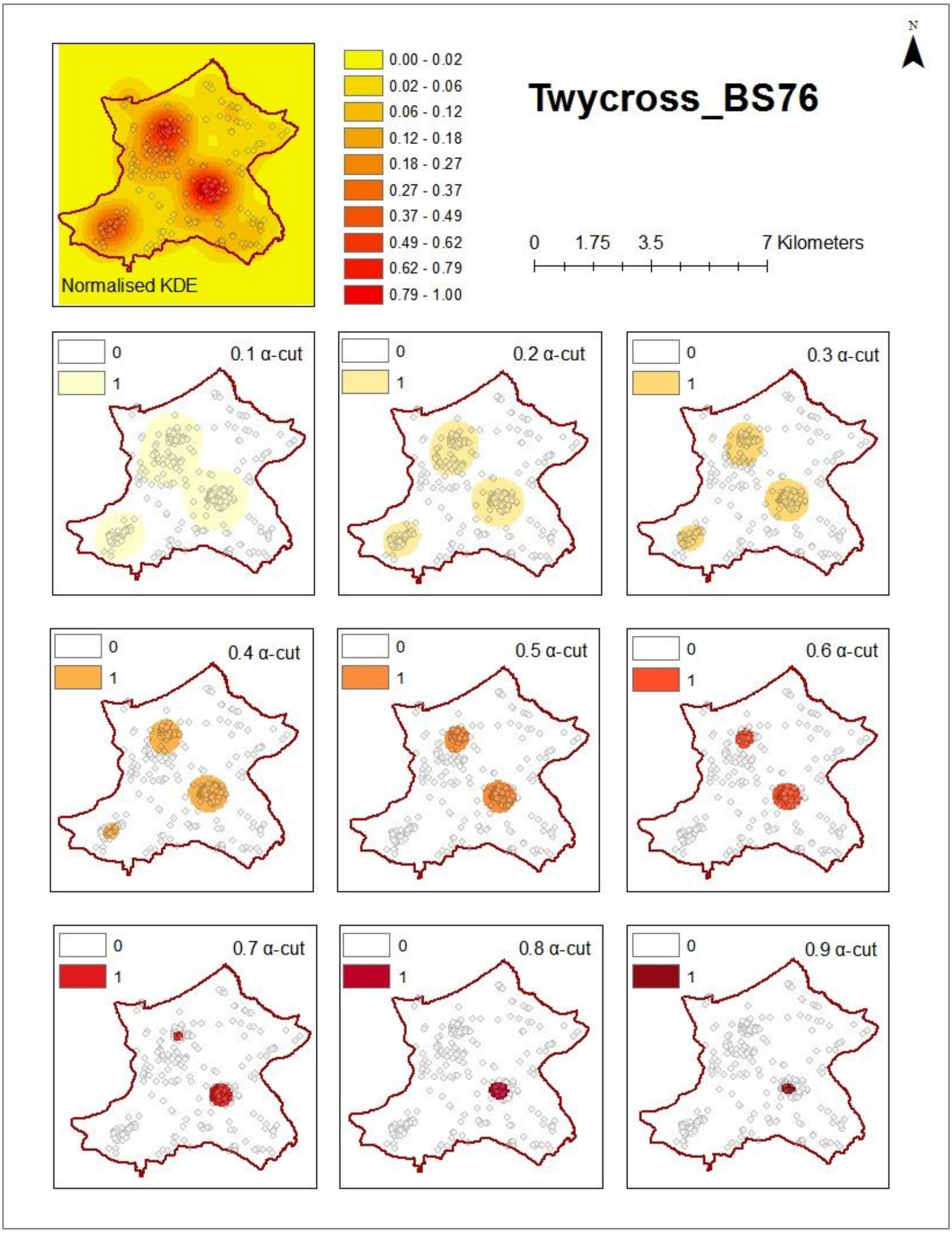

Sutton Cheny_BS76
N
0.00 - 0.02
0.02 - 0.06
0.06 - 0.11
0.11 - 0.17
0.17 - 0.25
0.25 - 0.34
0.34 - 0.46
0.46 - 0.61
0.61 - 0.80
0.80 - 1.00
Normalised KDE
0 1.5 3 6 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

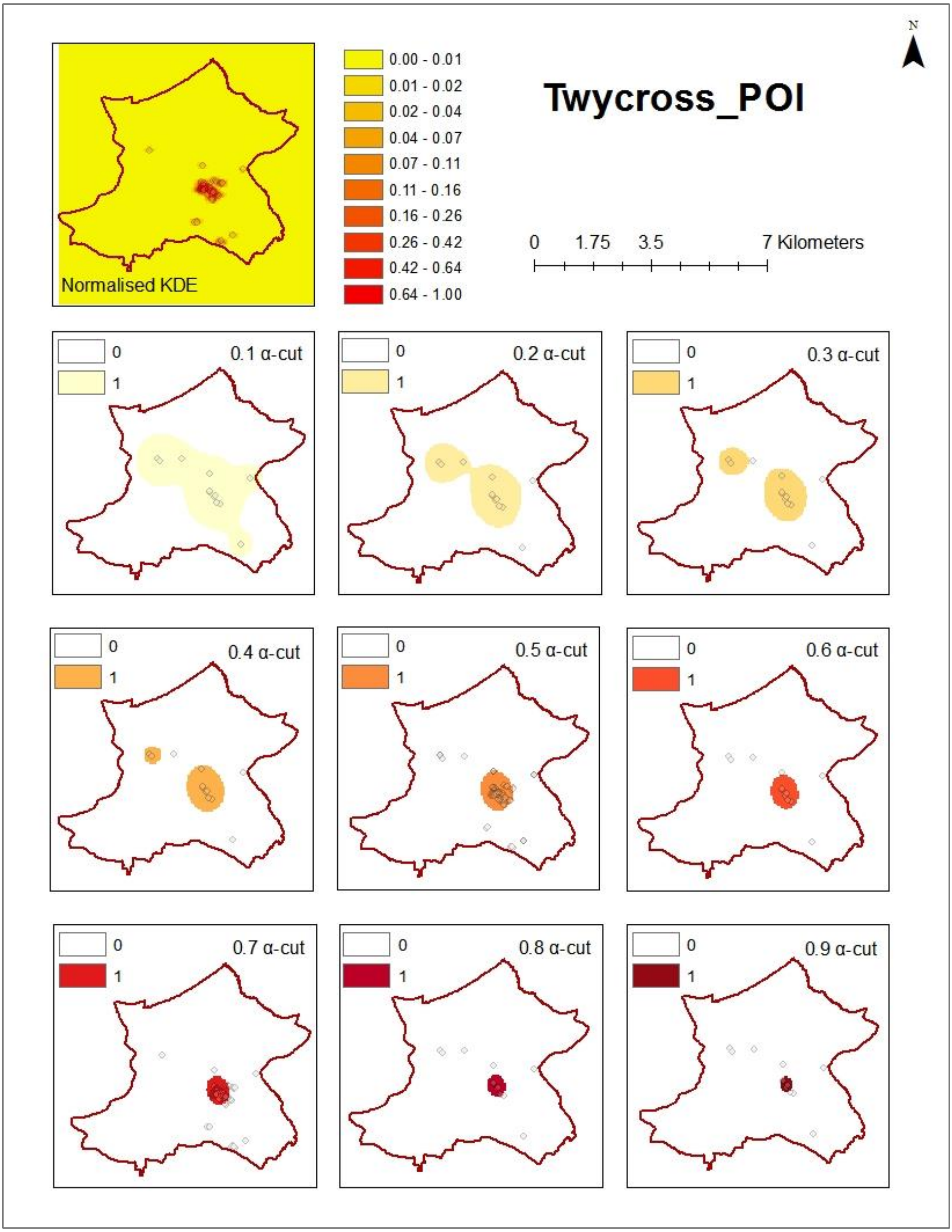


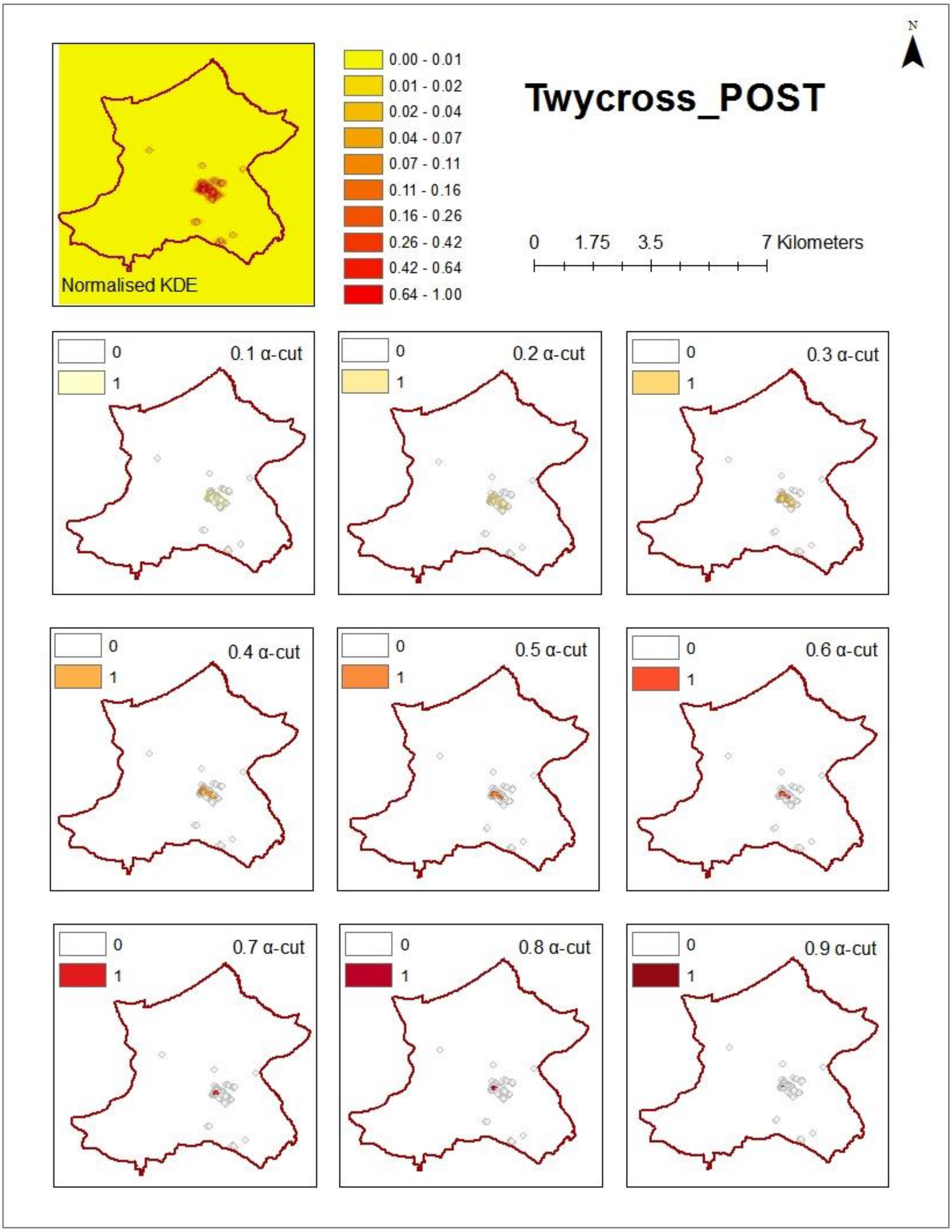
Sutton Cheny_POI
N
Normalised KDE
0.00 - 0.02
0.02 - 0.08
0.08 - 0.15
0.15 - 0.23
0.23 - 0.32
0.32 - 0.41
0.41 - 0.51
0.51 - 0.64
0.64 - 0.81
0.81 - 1.00
0 1.5 3 6 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

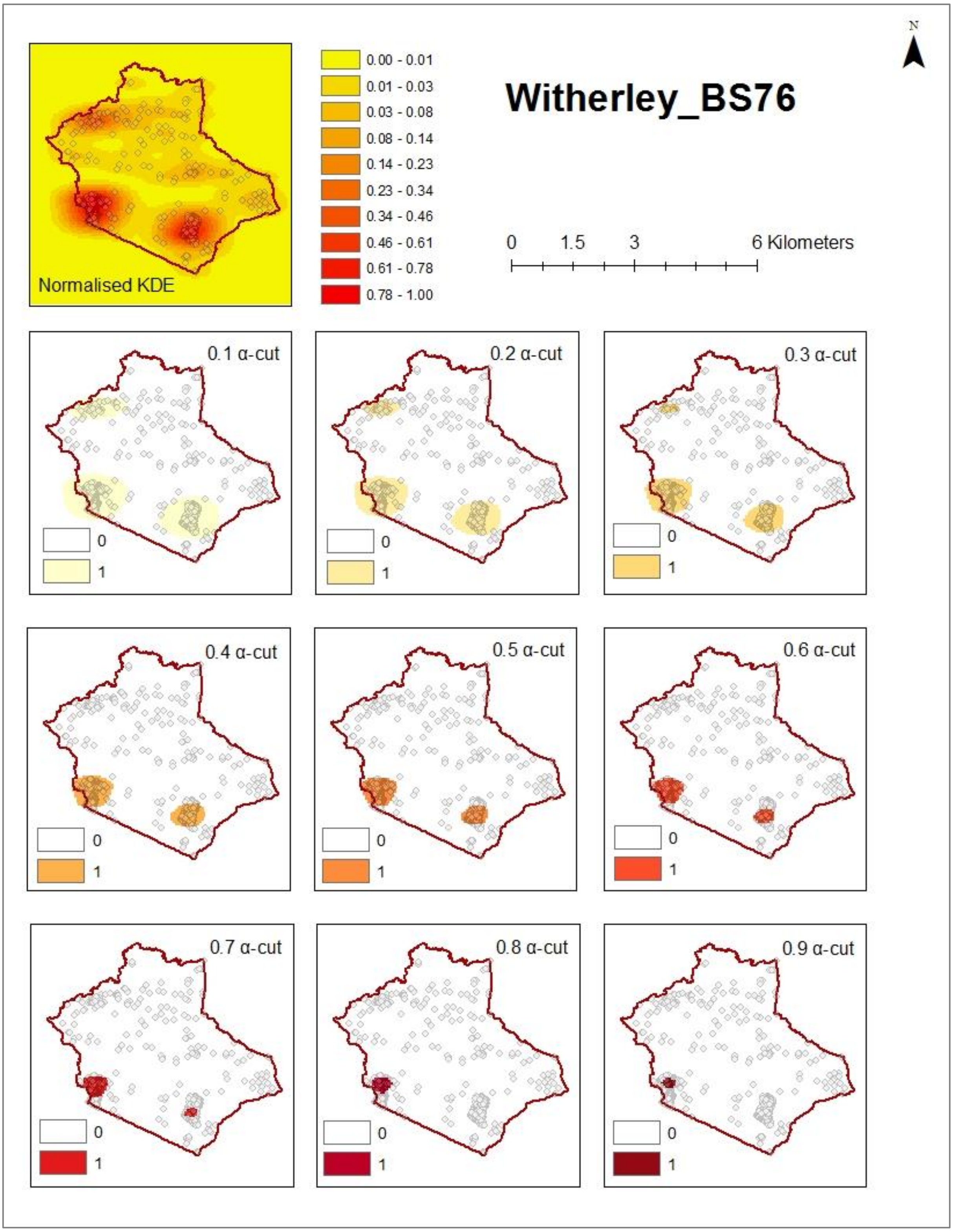


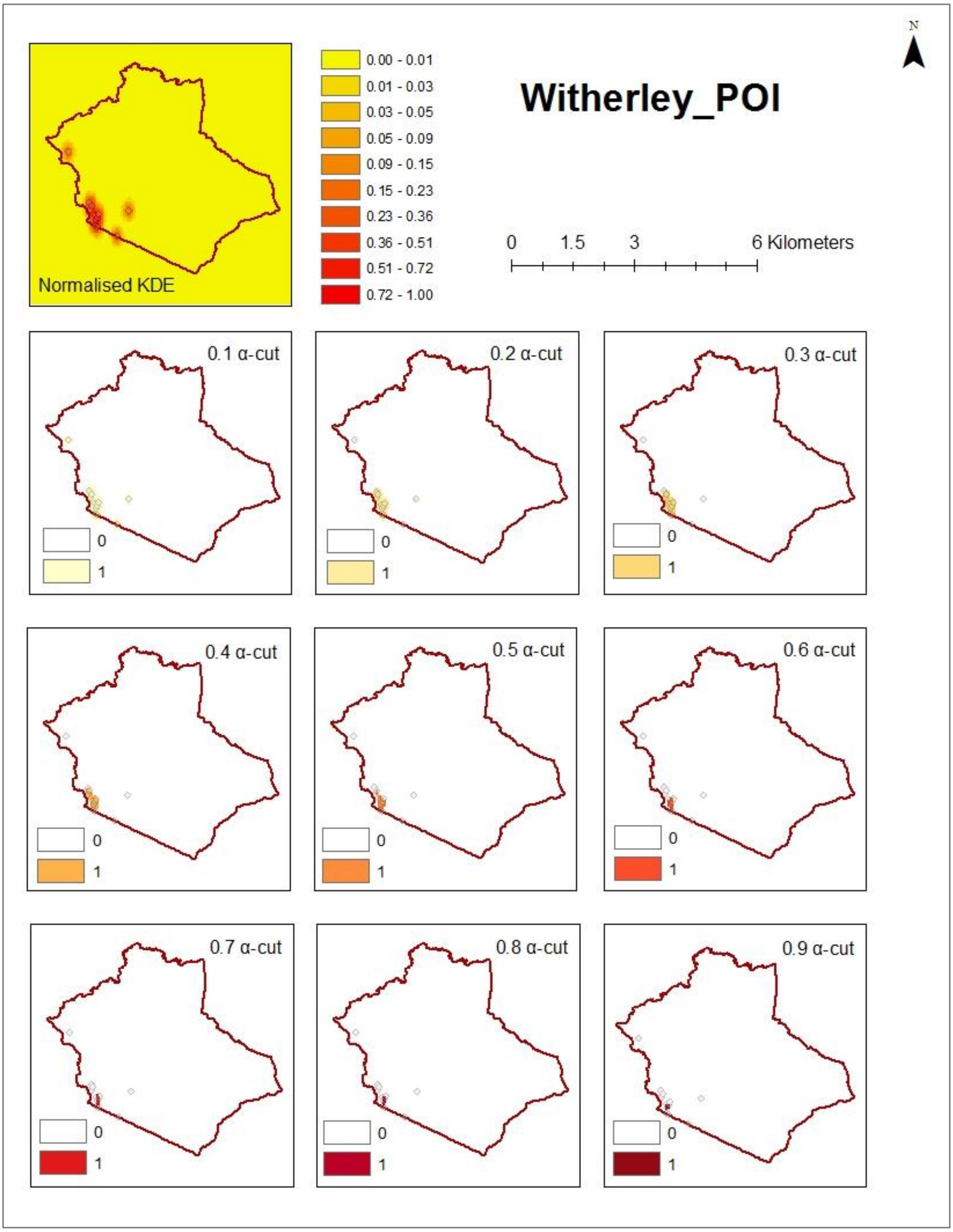
Sutton Cheny_POST
N
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.05
0.05 - 0.09
0.09 - 0.14
0.14 - 0.23
0.23 - 0.36
0.36 - 0.52
0.52 - 0.73
0.73 - 1.00
0 1.5 3 6 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


Twycross_BS76
N
Normalised KDE
0.00 - 0.02
0.02 - 0.06
0.06 - 0.12
0.12 - 0.18
0.18 - 0.27
0.27 - 0.37
0.37 - 0.49
0.49 - 0.62
0.62 - 0.79
0.79 - 1.00
0 1.75 3.5 7 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

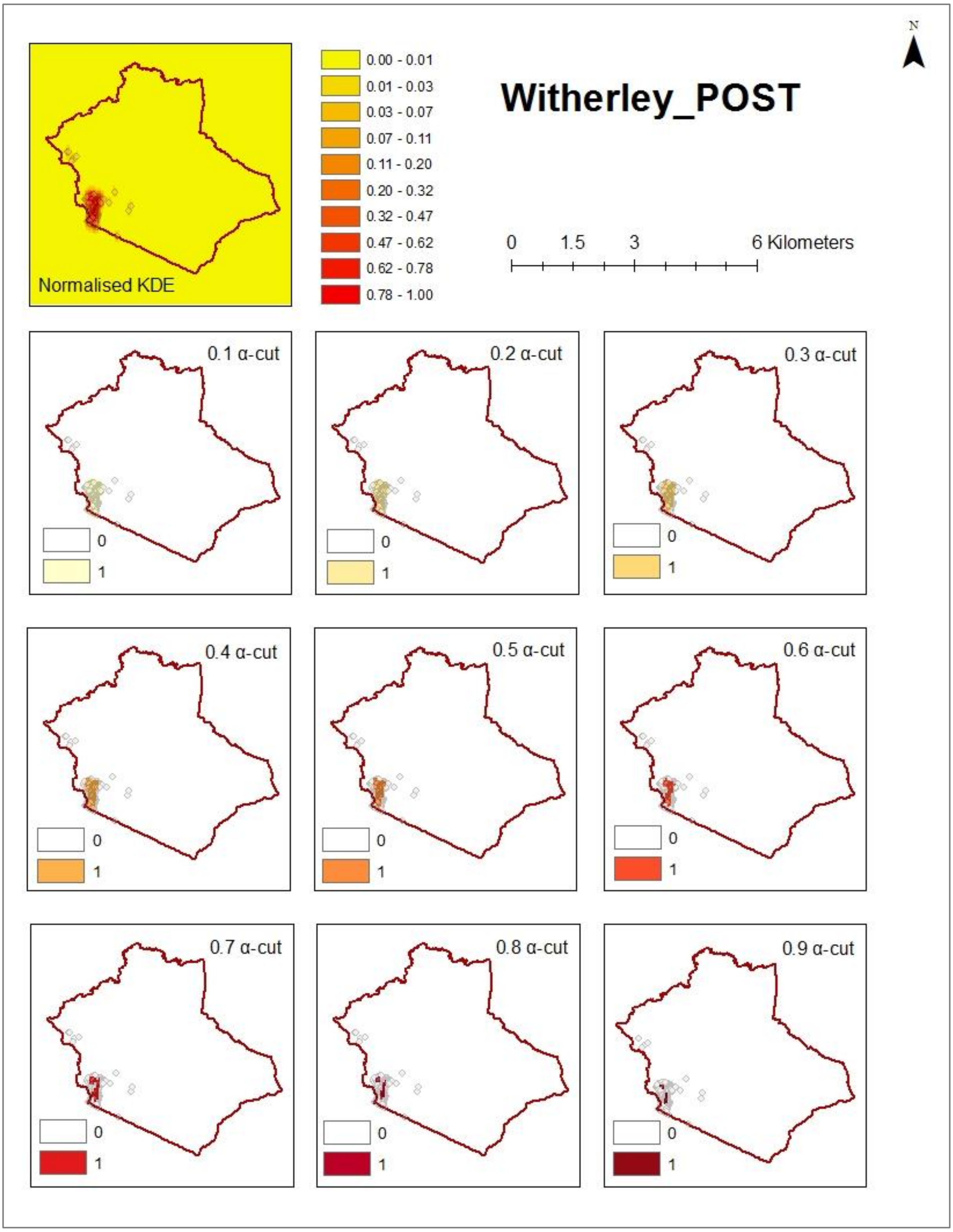

Twycross_POI
N
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.11
0.11 - 0.16
0.16 - 0.26
0.26 - 0.42
0.42 - 0.64
0.64 - 1.00
0 1.75 3.5 7 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1


Twycross_POST
Normalised KDE
0.00 - 0.01
0.01 - 0.02
0.02 - 0.04
0.04 - 0.07
0.07 - 0.11
0.11 - 0.16
0.16 - 0.26
0.26 - 0.42
0.42 - 0.64
0.64 - 1.00
0 1.75 3.5 7 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

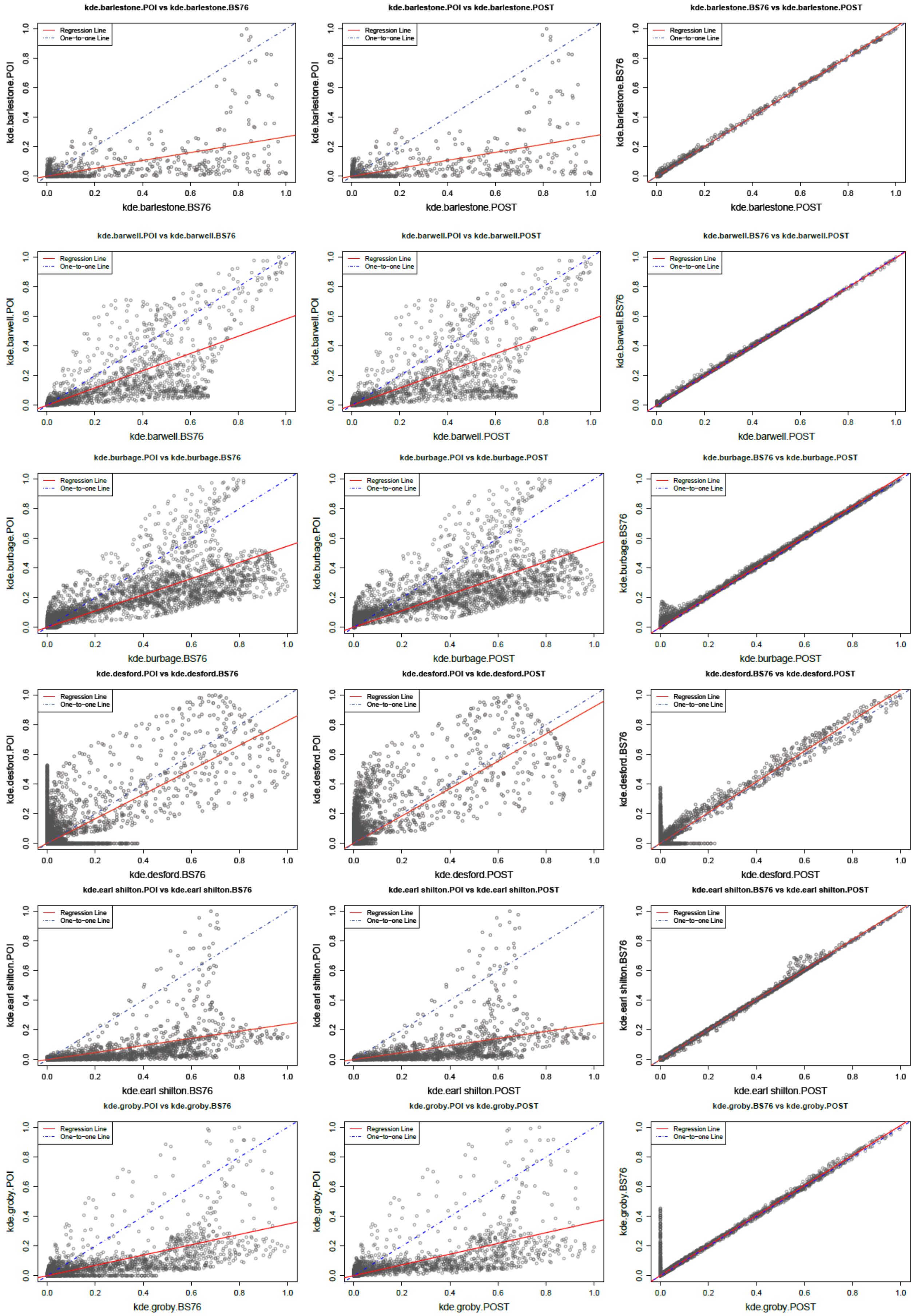

Witherley_BS76
N
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.08
0.08 - 0.14
0.14 - 0.23
0.23 - 0.34
0.34 - 0.46
0.46 - 0.61
0.61 - 0.78
0.78 - 1.00
0 1.5 3 6 Kilometers
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

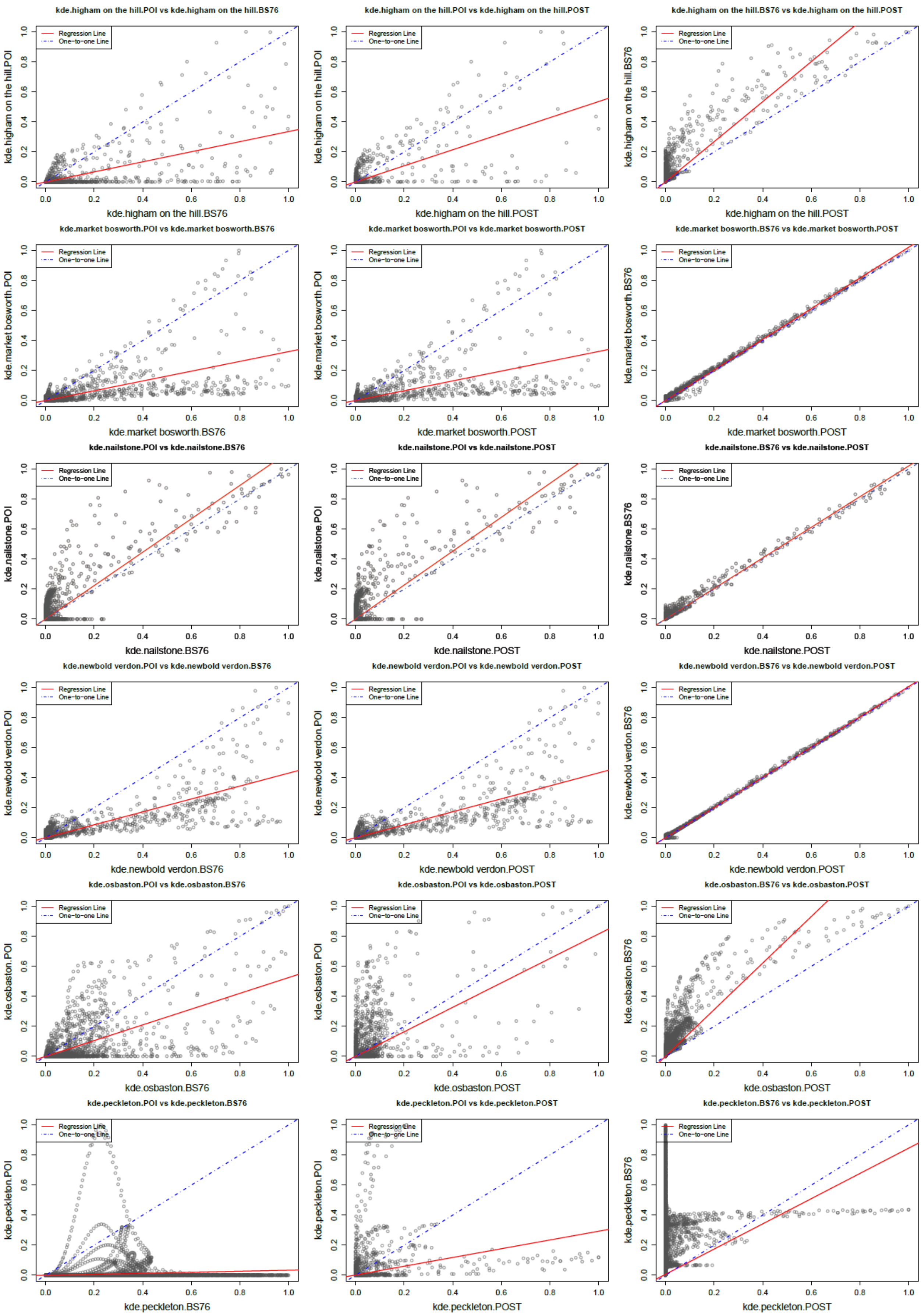


Witherley_POI
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.05
0.05 - 0.09
0.09 - 0.15
0.15 - 0.23
0.23 - 0.36
0.36 - 0.51
0.51 - 0.72
0.72 - 1.00
0 1.5 3 6 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

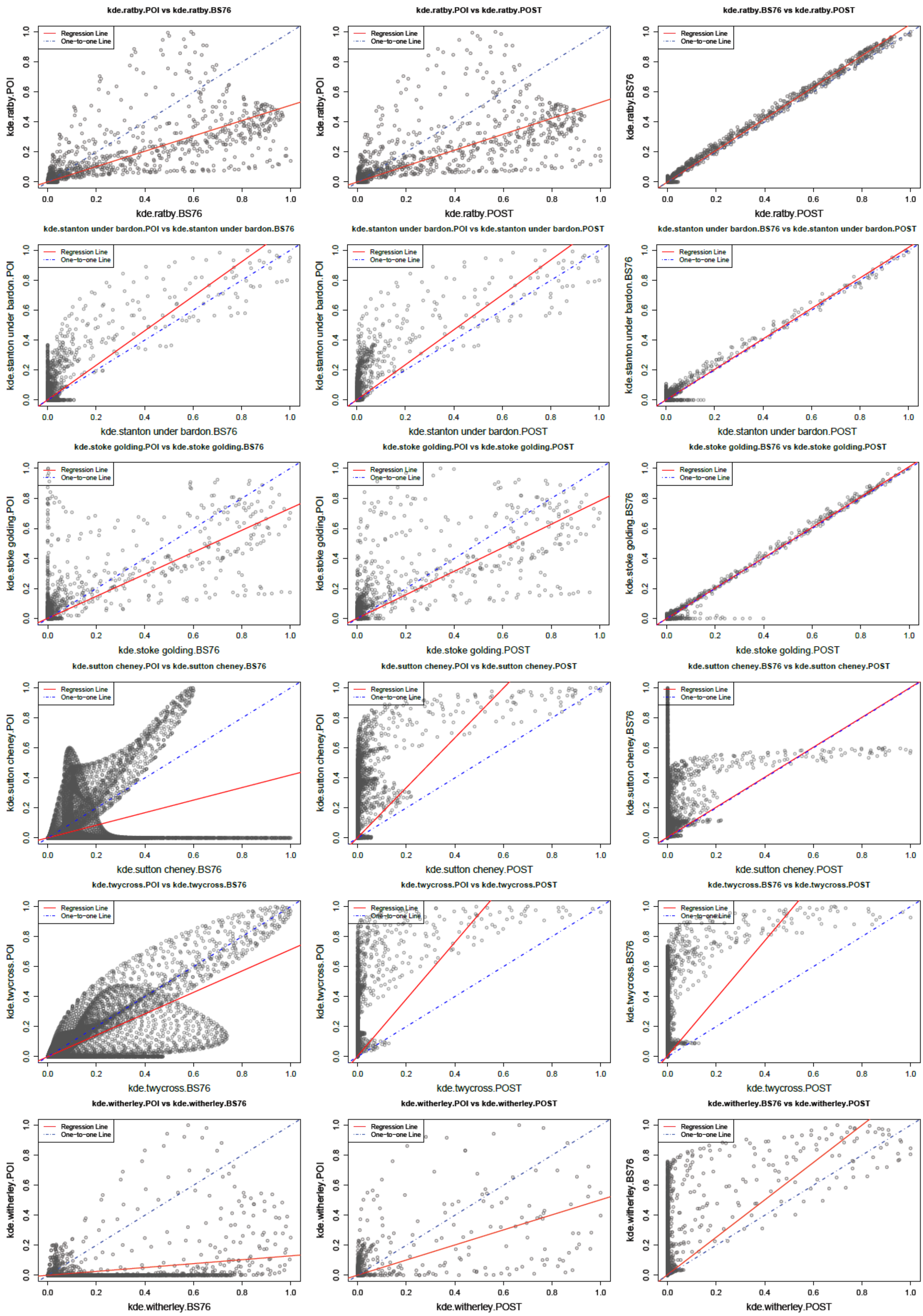

Witherley_POST
Normalised KDE
0.00 - 0.01
0.01 - 0.03
0.03 - 0.07
0.07 - 0.11
0.11 - 0.20
0.20 - 0.32
0.32 - 0.47
0.47 - 0.62
0.62 - 0.78
0.78 - 1.00
0 1.5 3 6 Kilometers
N
0.1 α-cut
0.2 α-cut
0.3 α-cut
0.4 α-cut
0.5 α-cut
0.6 α-cut
0.7 α-cut
0.8 α-cut
0.9 α-cut
0
1

Appendix (3): Regression plots between each pair of the data sources

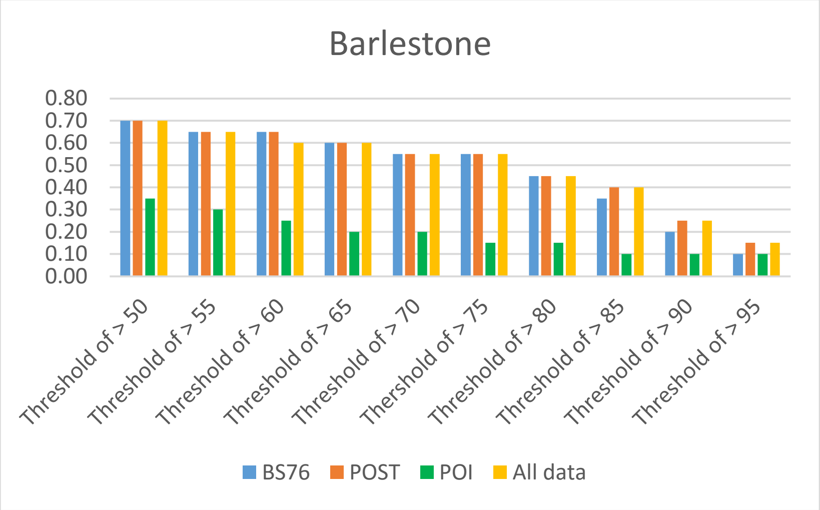


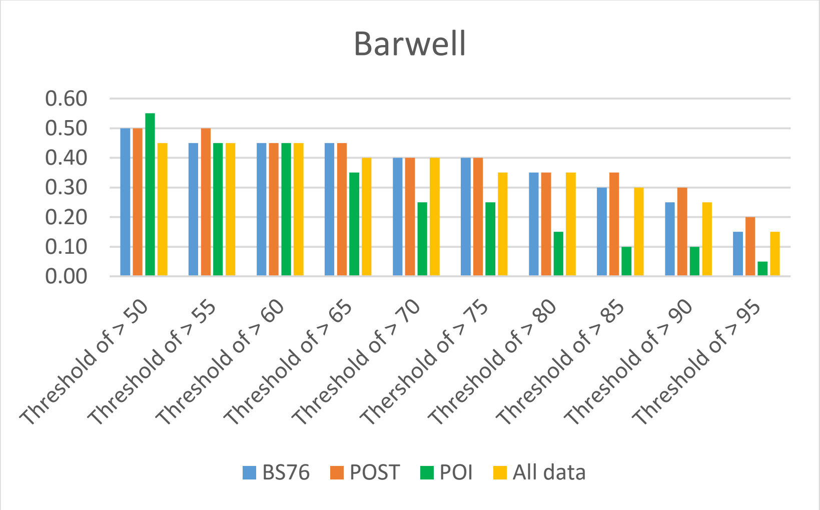


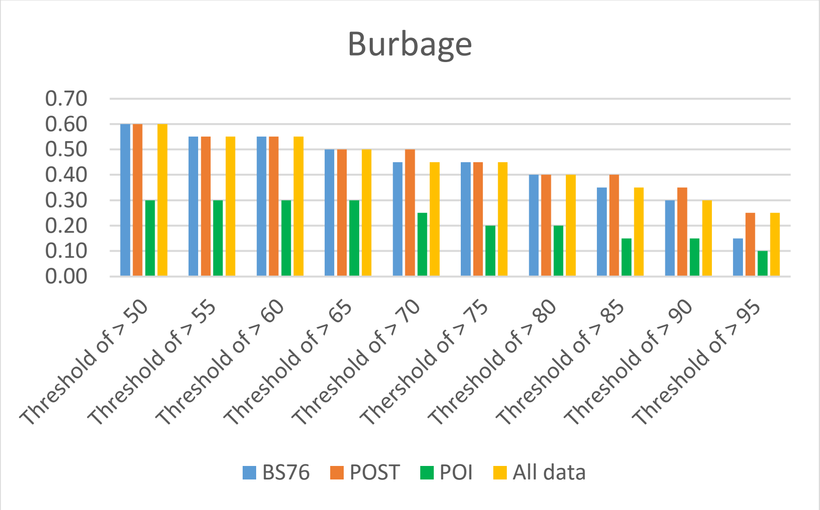
kde.higham on the hill.POI vs kde.higham on the hill.BS76
kde.higham on the hill.POI vs kde.higham on the hill.POST
kde.higham on the hill.BS76 vs kde.higham on the hill.POST
kde.market bosworth.POI vs kde.market bosworth.BS76
kde.market bosworth.POI vs kde.market bosworth.POST
kde.market bosworth.BS76 vs kde.market bosworth.POST
kde.nailstone.POI vs kde.nailstone.BS76
kde.nailstone.POI vs kde.nailstone.POST
kde.nailstone.BS76 vs kde.nailstone.POST
kde.newbold verdon.POI vs kde.newbold verdon.BS76
kde.newbold verdon.POI vs kde.newbold verdon.POST
kde.newbold verdon.BS76 vs kde.newbold verdon.POST
kde.osbaston.POI vs kde.osbaston.BS76
kde.osbaston.POI vs kde.osbaston.POST
kde.osbaston.BS76 vs kde.osbaston.POST
kde.peckleton.POI vs kde.peckleton.BS76
kde.peckleton.POI vs kde.peckleton.POST
kde.peckleton.BS76 vs kde.peckleton.POST
Regression Line
One-to-one Line
kde.higham on the hill.POI
kde.higham on the hill.BS76
kde.higham on the hill.POST
kde.market bosworth.POI
kde.market bosworth.BS76
kde.market bosworth.POST
kde.nailstone.POI
kde.nailstone.BS76
kde.nailstone.POST
kde.newbold verdon.POI
kde.newbold verdon.BS76
kde.newbold verdon.POST
kde.osbaston.POI
kde.osbaston.BS76
kde.osbaston.POST
kde.peckleton.POI
kde.peckleton.BS76
kde.peckleton.POST
0.0
0.2
0.4
0.6
0.8
1.0

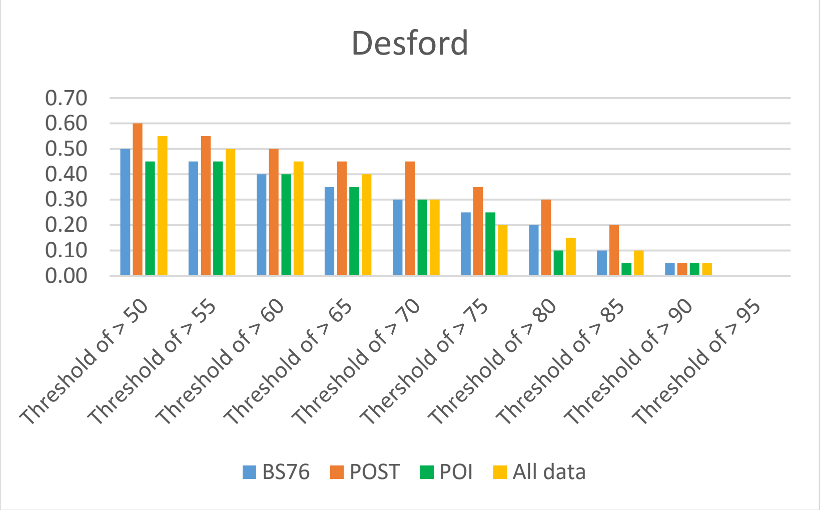

kde.ratby.POI vs kde.ratby.BS76
kde.ratby.POI vs kde.ratby.POST
kde.ratby.BS76 vs kde.ratby.POST
Regression Line
One-to-one Line
kde.ratby.POI
kde.ratby.BS76
kde.ratby.POST
kde.stanton under bardon.POI vs kde.stanton under bardon.BS76
kde.stanton under bardon.POI vs kde.stanton under bardon.POST
kde.stanton under bardon.BS76 vs kde.stanton under bardon.POST
kde.stanton under bardon.POI
kde.stanton under bardon.BS76
kde.stanton under bardon.POST
kde.stoke golding.POI vs kde.stoke golding.BS76
kde.stoke golding.POI vs kde.stoke golding.POST
kde.stoke golding.BS76 vs kde.stoke golding.POST
kde.stoke golding.POI
kde.stoke golding.BS76
kde.stoke golding.POST
kde.sutton cheney.POI vs kde.sutton cheney.BS76
kde.sutton cheney.POI vs kde.sutton cheney.POST
kde.sutton cheney.BS76 vs kde.sutton cheney.POST
kde.sutton cheney.POI
kde.sutton cheney.BS76
kde.sutton cheney.POST
kde.twycross.POI vs kde.twycross.BS76
kde.twycross.POI vs kde.twycross.POST
kde.twycross.BS76 vs kde.twycross.POST
kde.twycross.POI
kde.twycross.BS76
kde.twycross.POST
kde.witherley.POI vs kde.witherley.BS76
kde.witherley.POI vs kde.witherley.POST
kde.witherley.BS76 vs kde.witherley.POST
kde.witherley.POI
kde.witherley.BS76
kde.witherley.POST

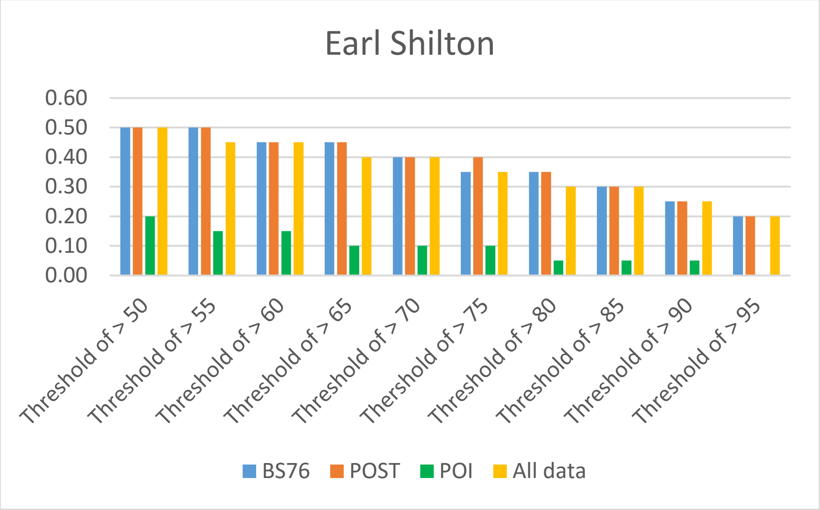
Appendix (4): Plots of the fuzzy membership grades along each thresholds

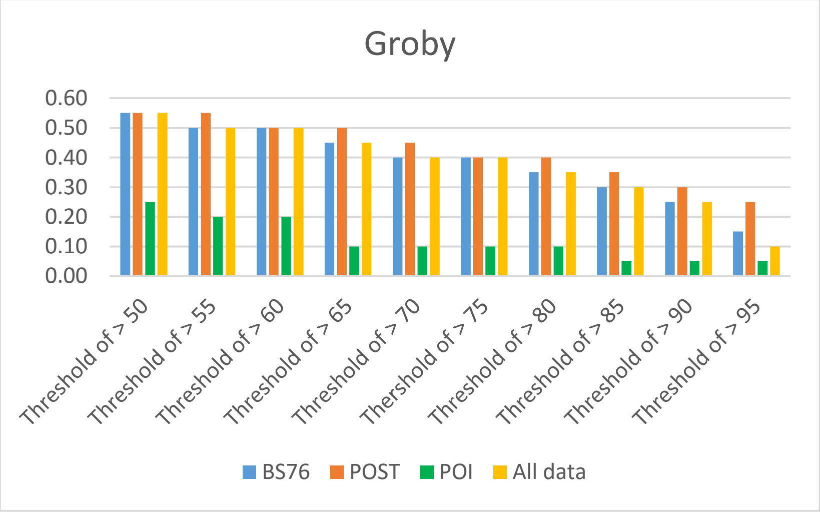


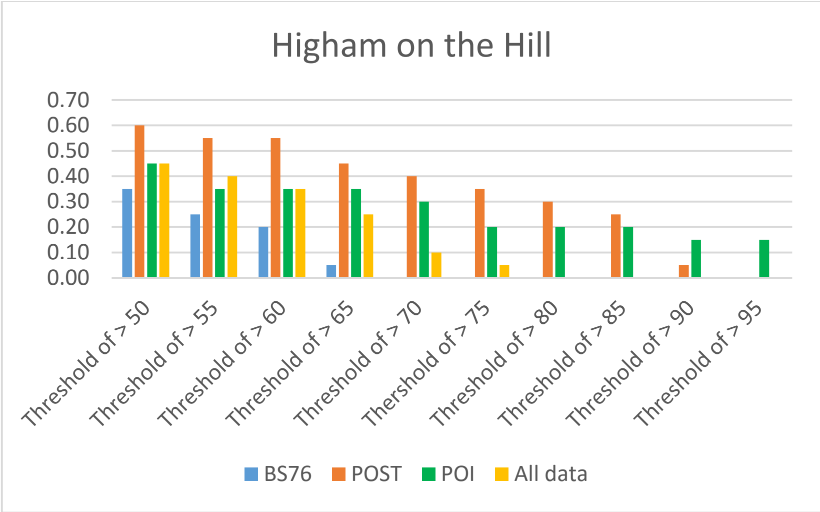


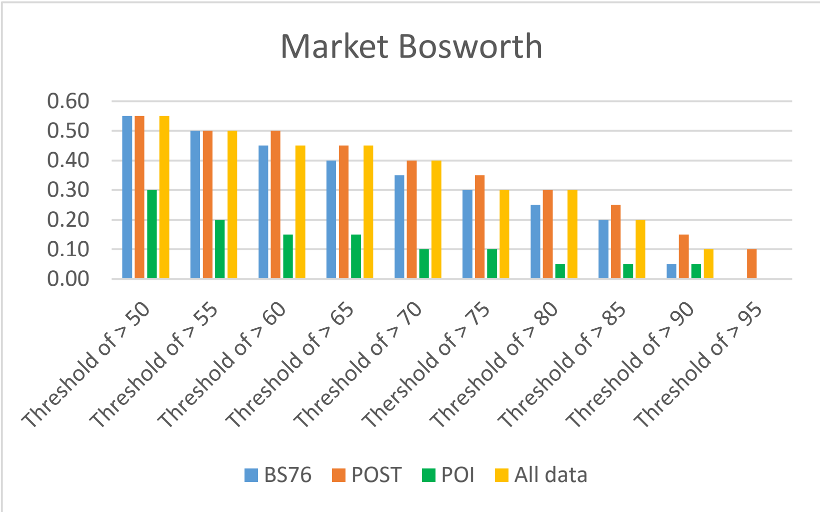


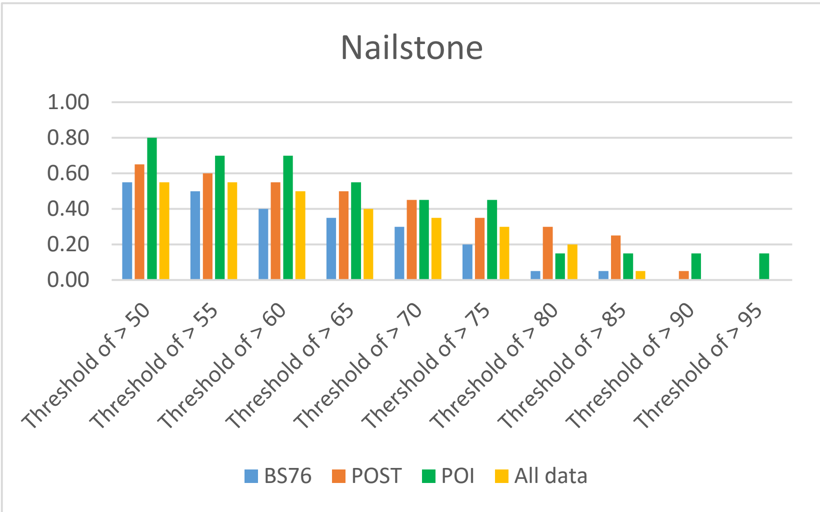

Desford
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

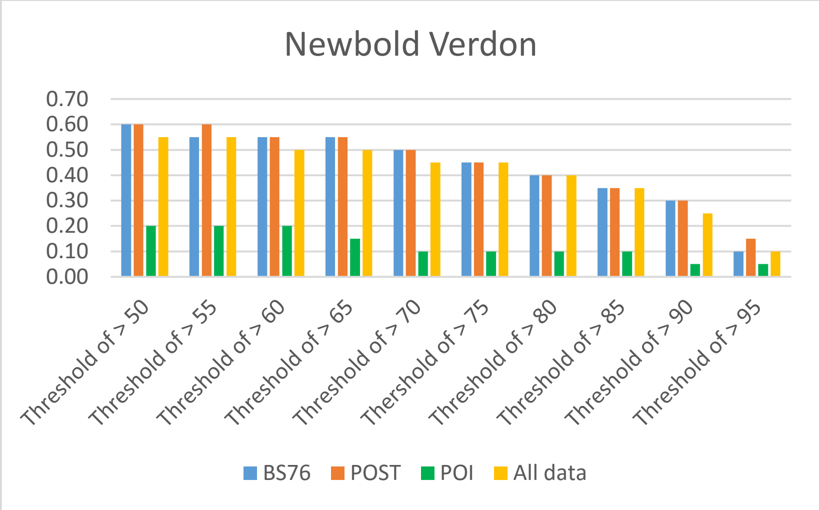

Earl Shilton
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

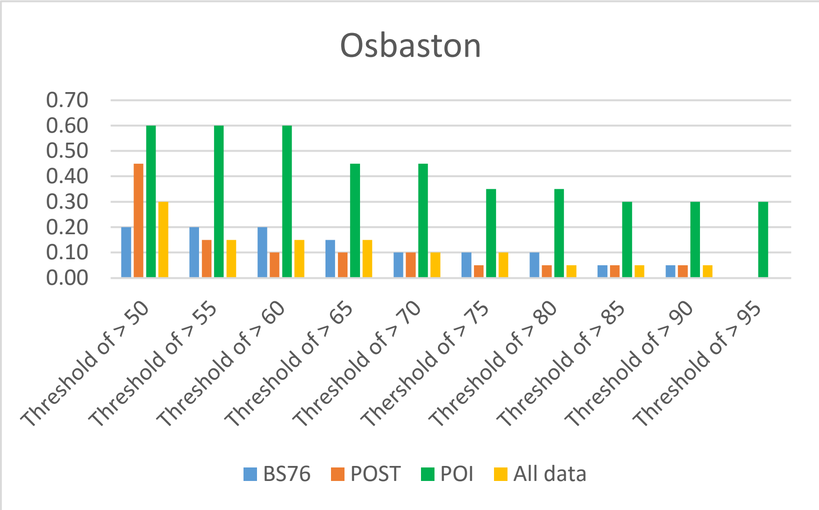

Groby
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

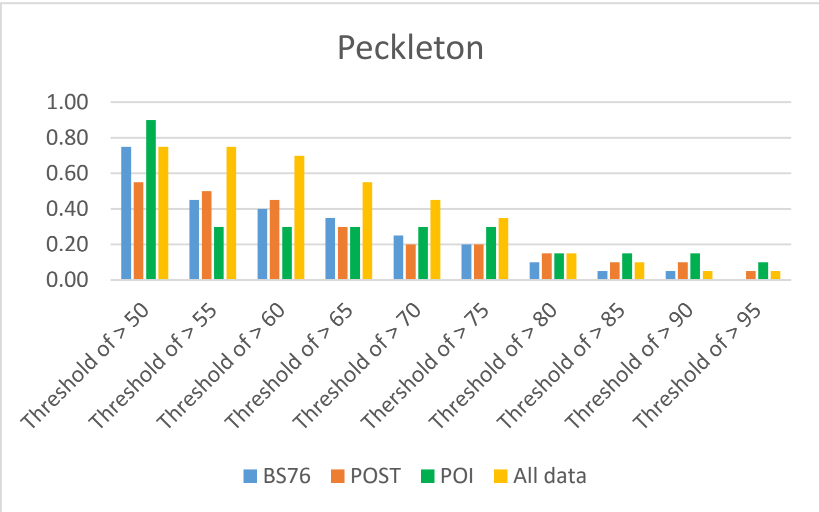

Higham on the Hill
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

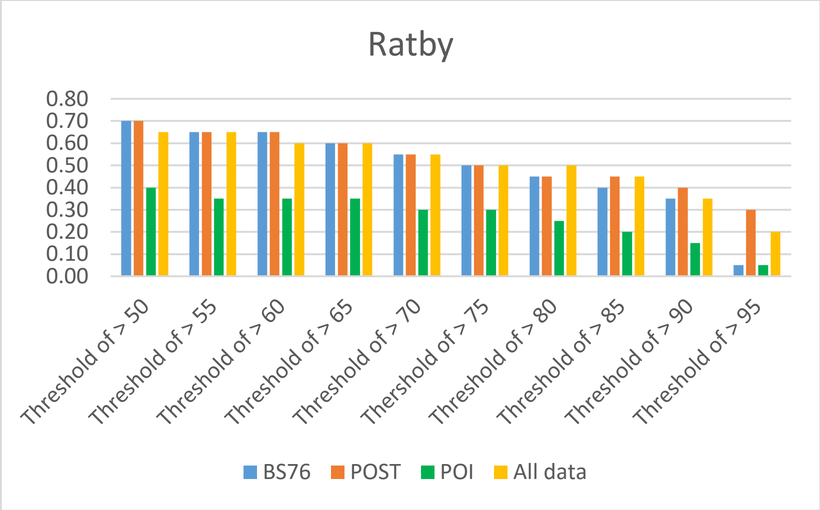

Market Bosworth
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

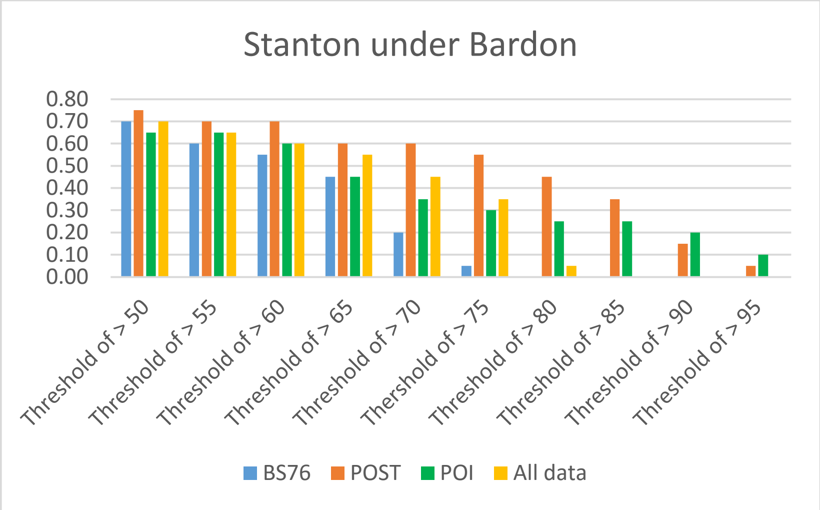

Nailstone
1.00
0.80
0.60
0.40
0.20
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

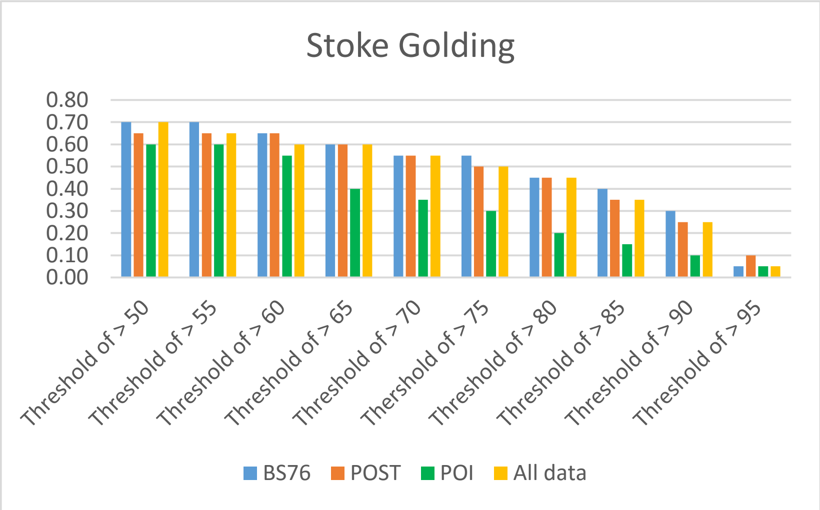

Newbold Verdon
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

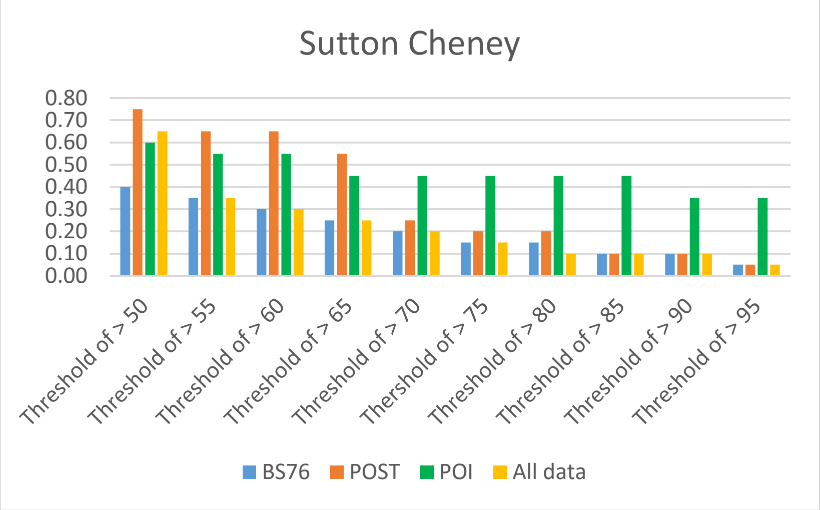

Osbaston
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

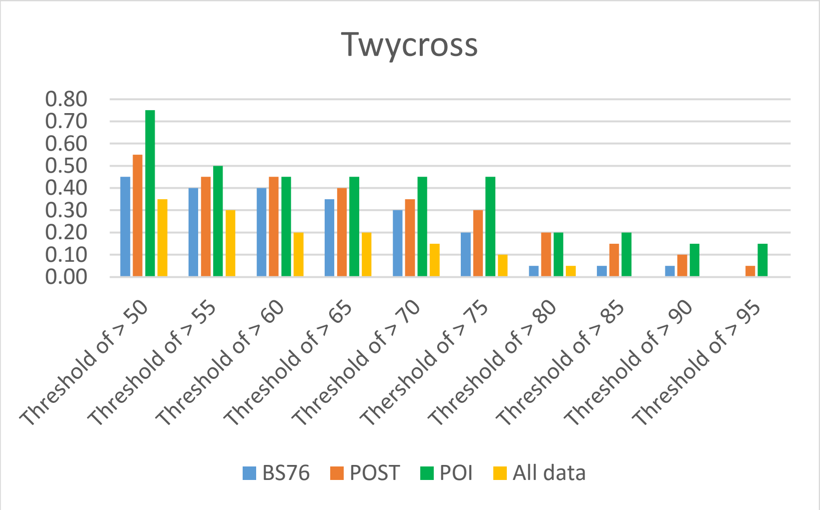

Peckleton
1.00
0.80
0.60
0.40
0.20
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

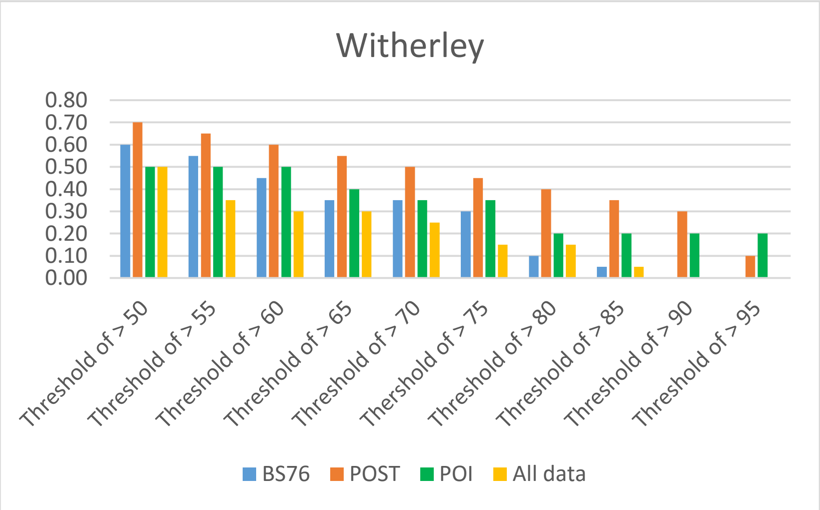

Ratby
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data


Stanton under Bardon
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

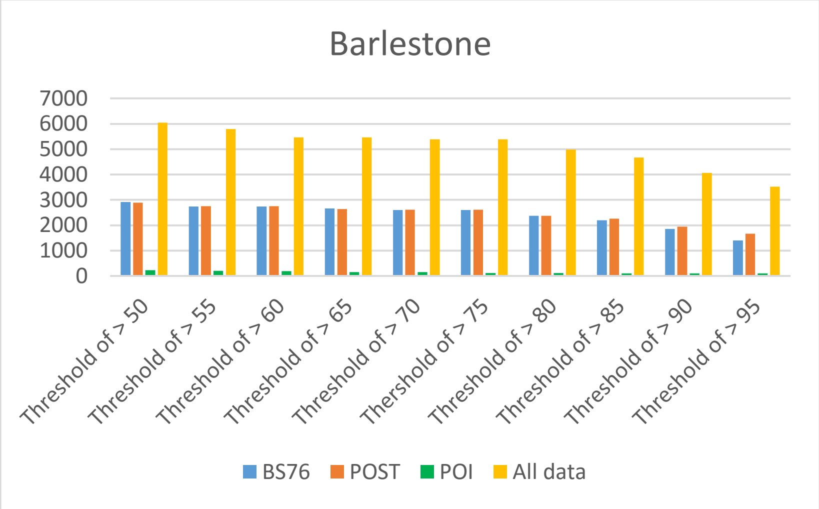

Stoke Golding
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

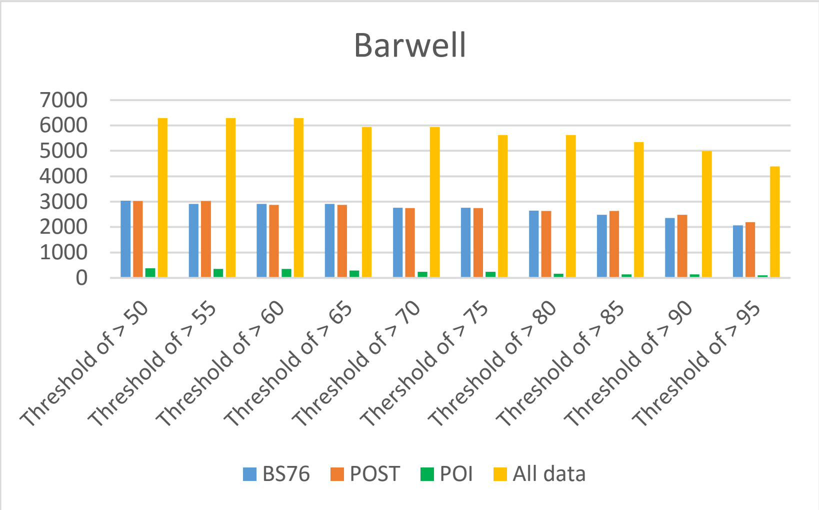

Sutton Cheney
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

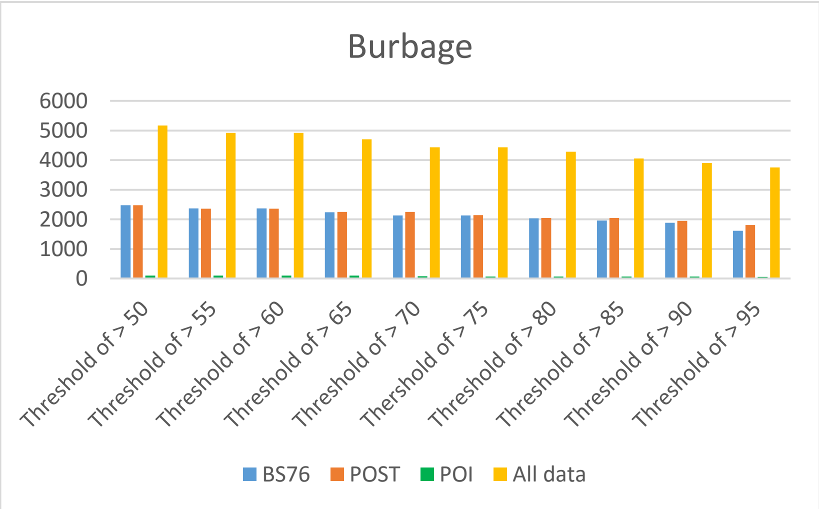

Twycross
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

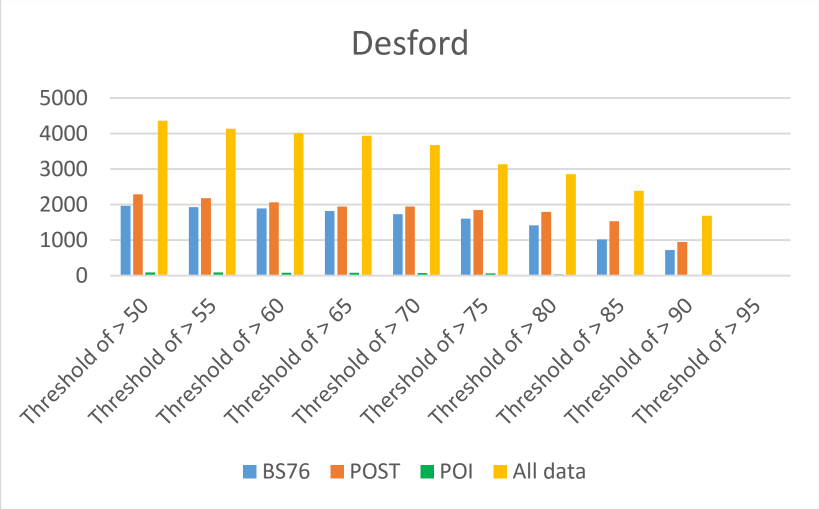

Witherley
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

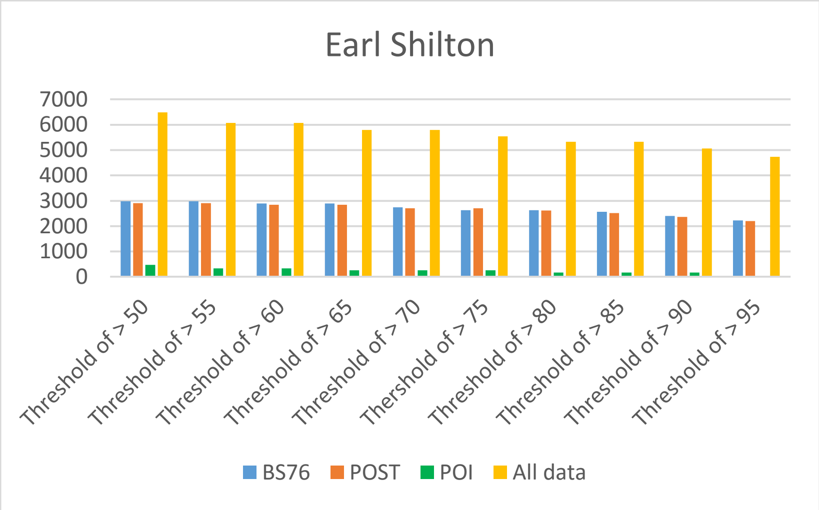
Plots of the density values along each thresholds

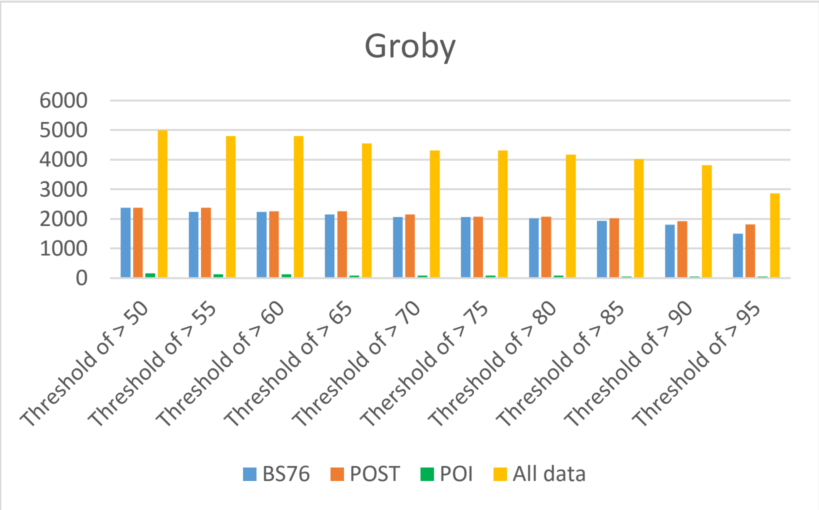


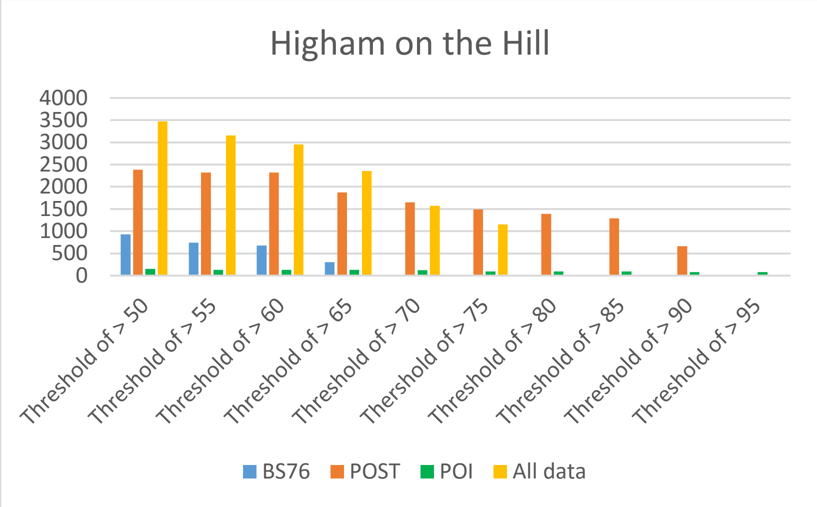


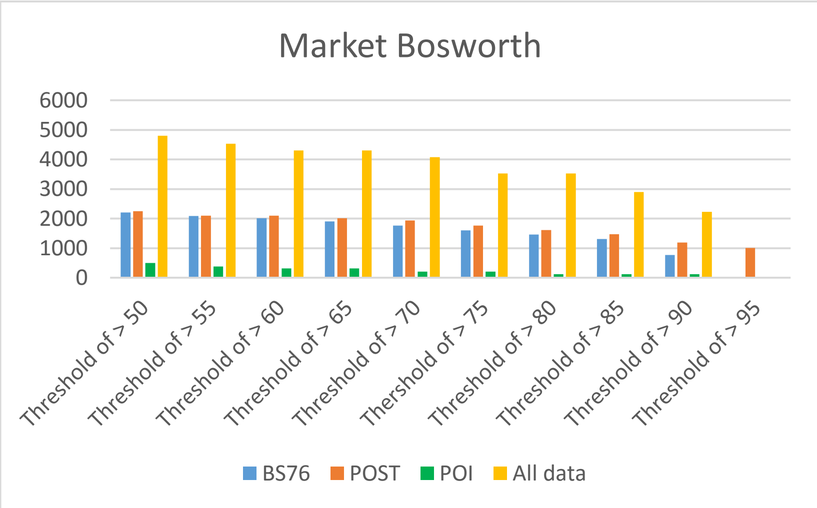


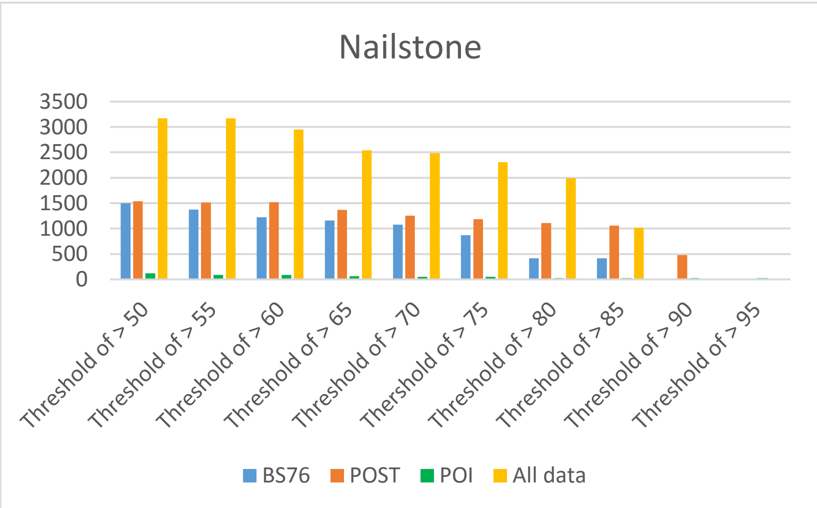

Desford
5000
4000
3000
2000
1000
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

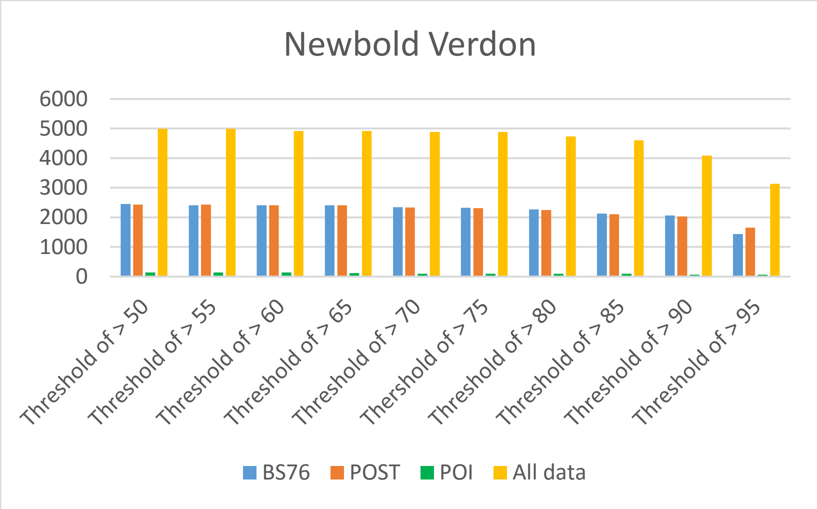

Earl Shilton
7000
6000
5000
4000
3000
2000
1000
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

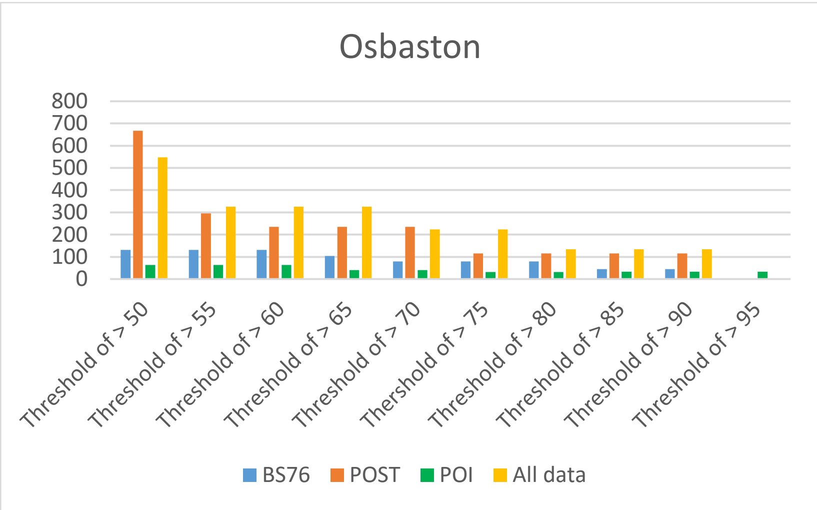

Groby
6000
5000
4000
3000
2000
1000
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

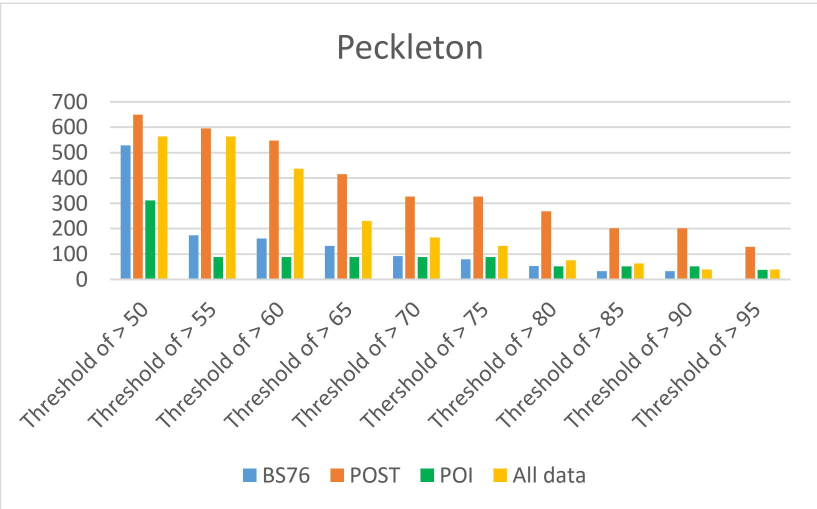

Higham on the Hill
4000
3500
3000
2500
2000
1500
1000
500
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

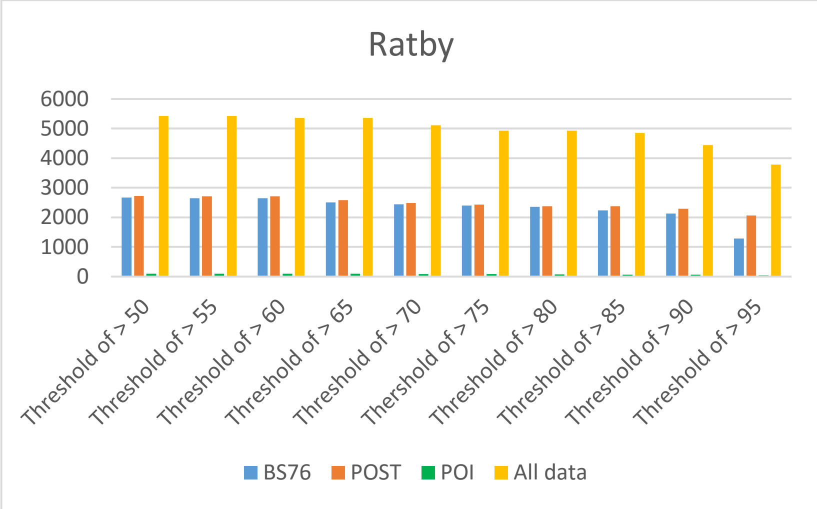

Market Bosworth
6000
5000
4000
3000
2000
1000
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

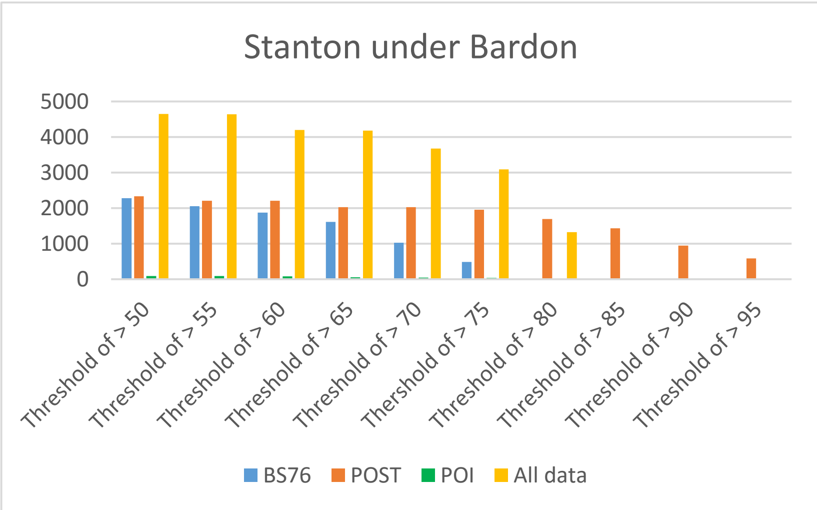

Nailstone
3500
3000
2500
2000
1500
1000
500
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

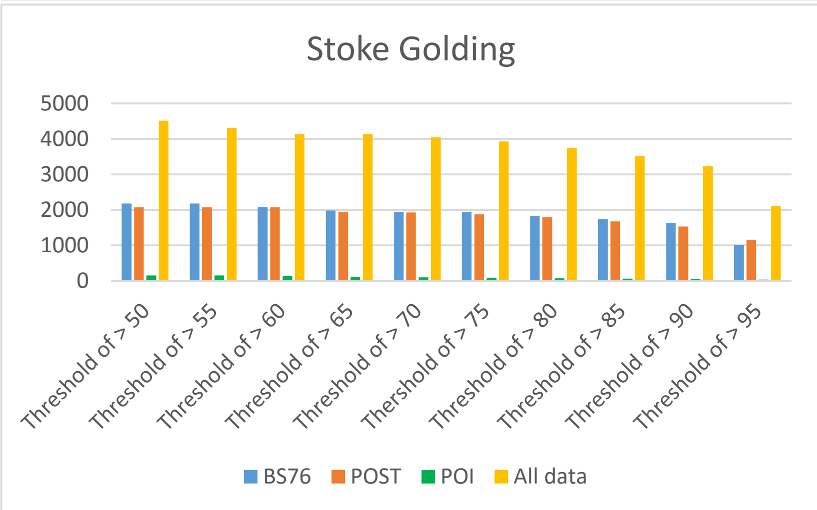

Newbold Verdon
6000
5000
4000
3000
2000
1000
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

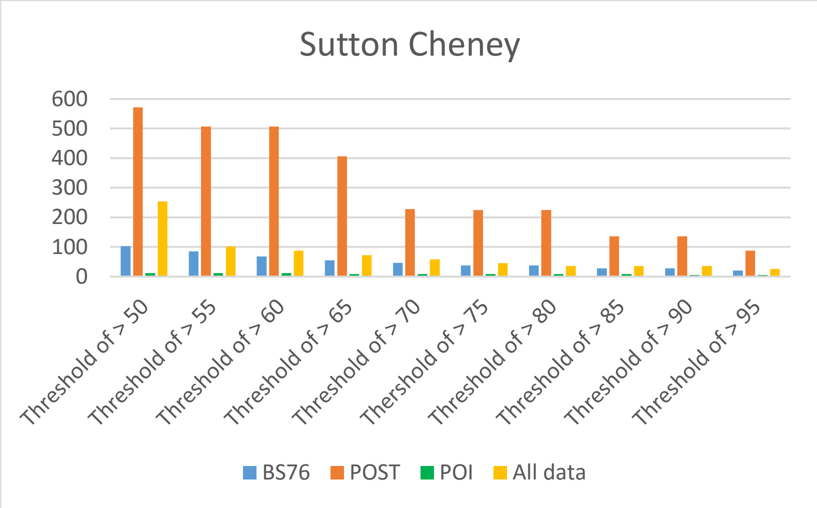

Osbaston
800
700
600
500
400
300
200
100
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

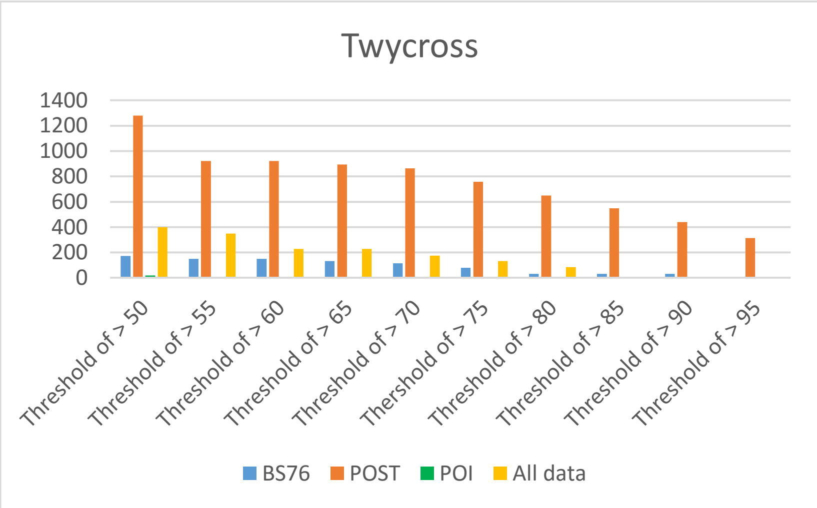

Peckleton
700
600
500
400
300
200
100
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

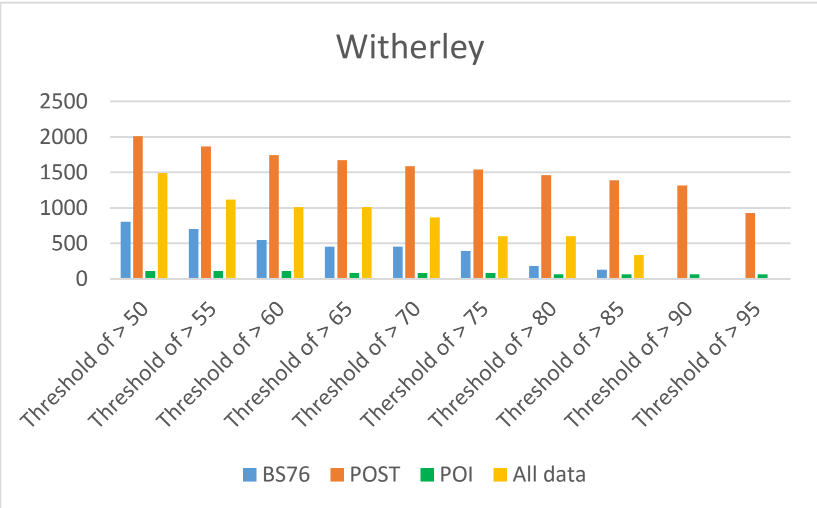

Ratby
6000
5000
4000
3000
2000
1000
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data


Stanton under Bardon
5000
4000
3000
2000
1000
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data


Stoke Golding
5000
4000
3000
2000
1000
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data


Sutton Cheney
600
500
400
300
200
100
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

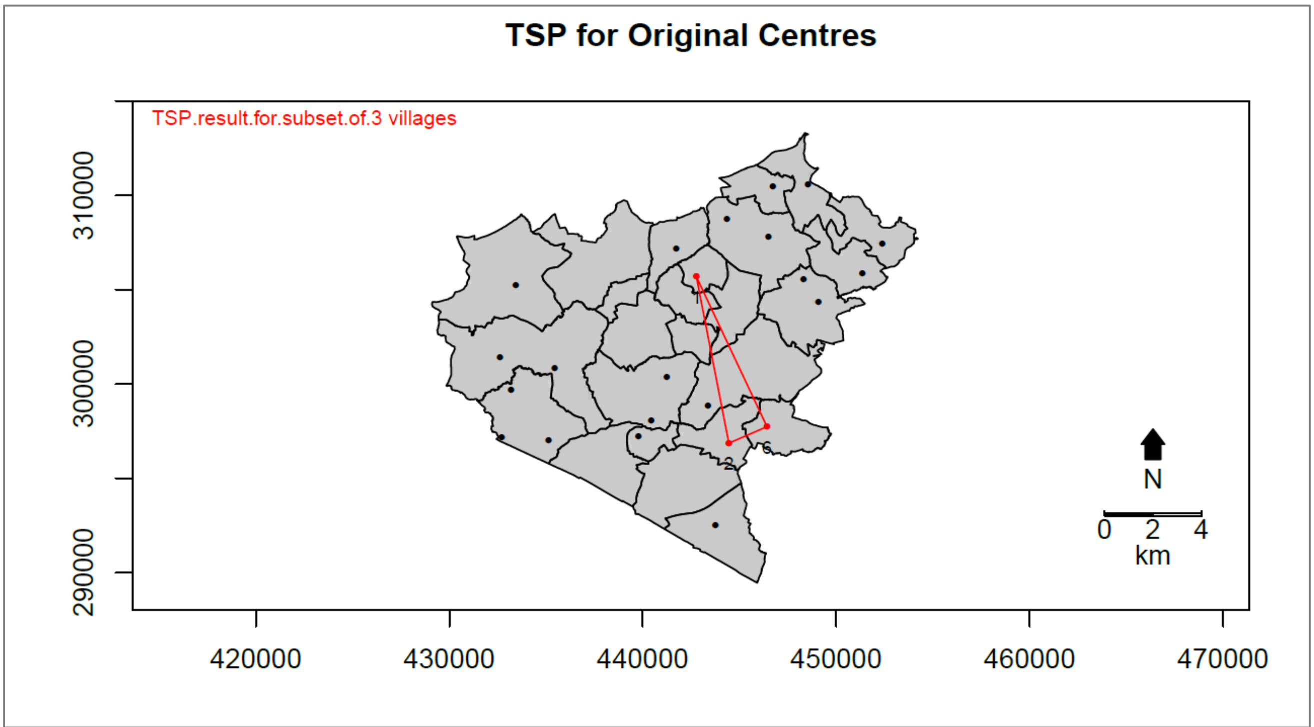

Twycross
1400
1200
1000
800
600
400
200
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

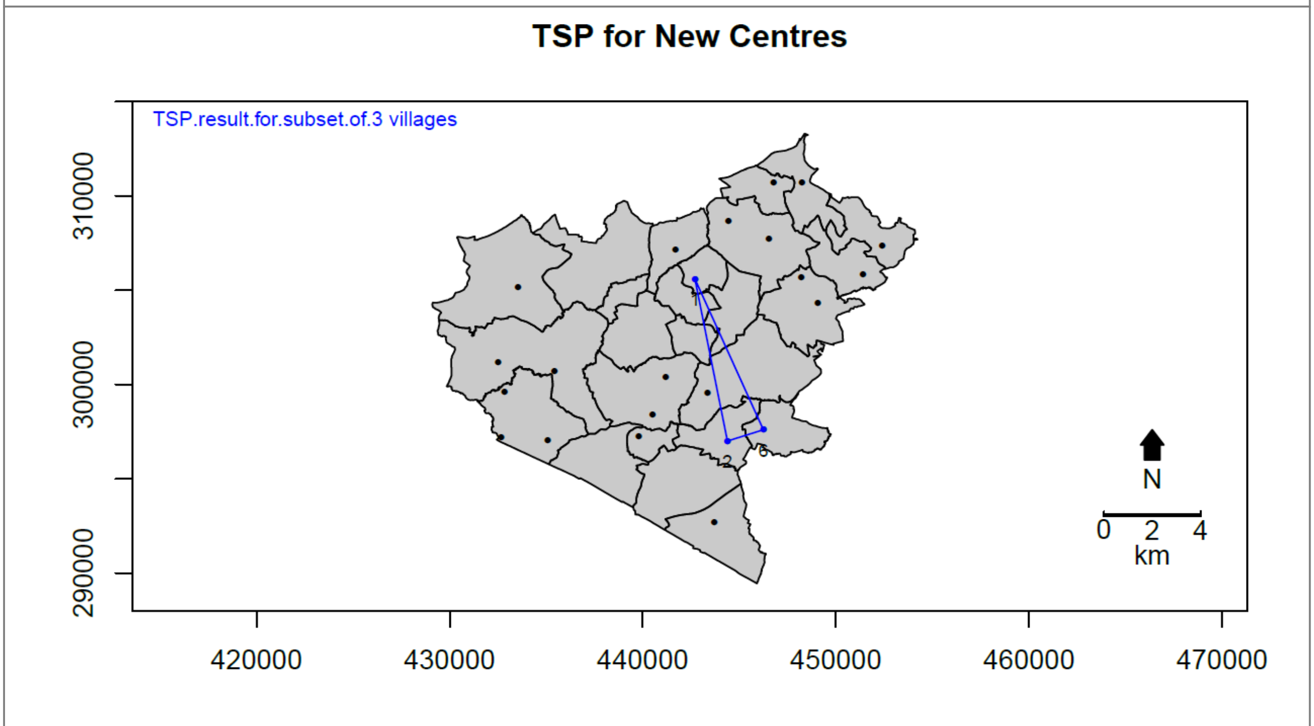

Witherley
2500
2000
1500
1000
500
0
Threshold of > 50
Threshold of > 55
Threshold of > 60
Threshold of > 65
Threshold of > 70
Thershold of > 75
Threshold of > 80
Threshold of > 85
Threshold of > 90
Threshold of > 95
BS76
POST
POI
All data

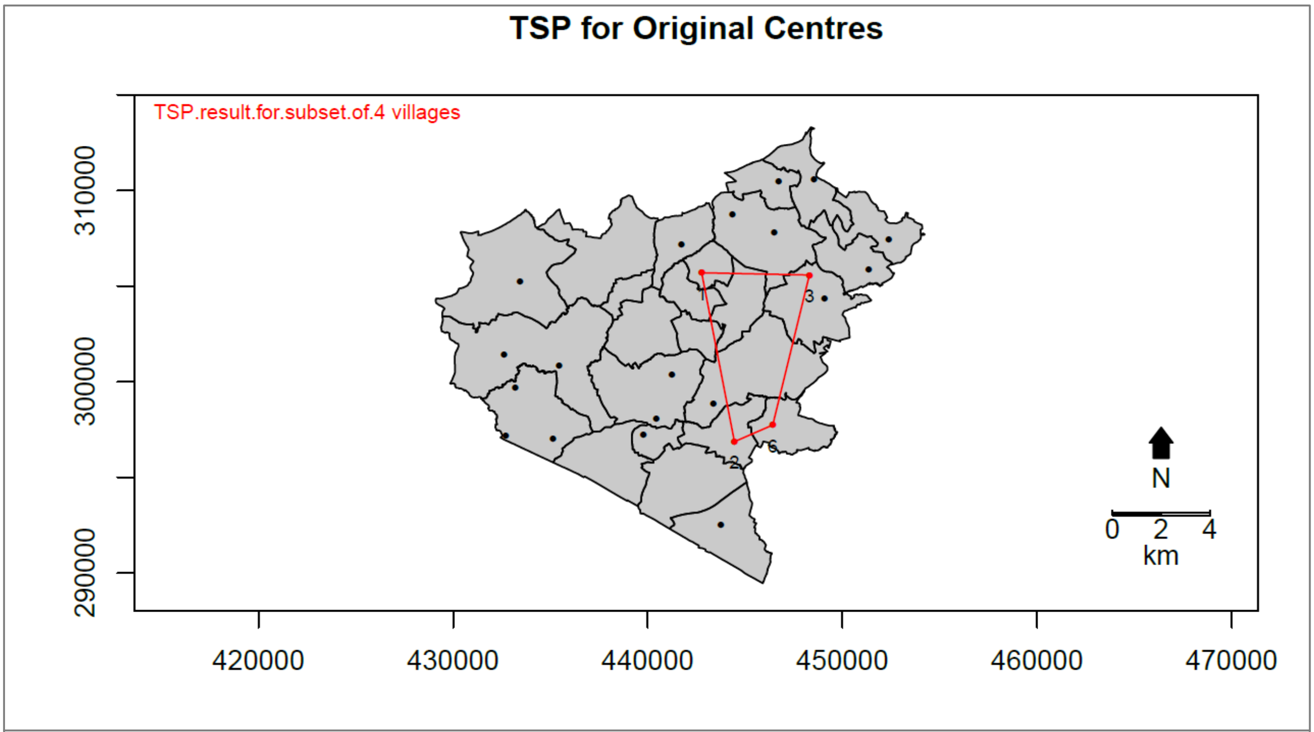
Appendix (5): Maps of the Travelling Salesman Problem for some subsets

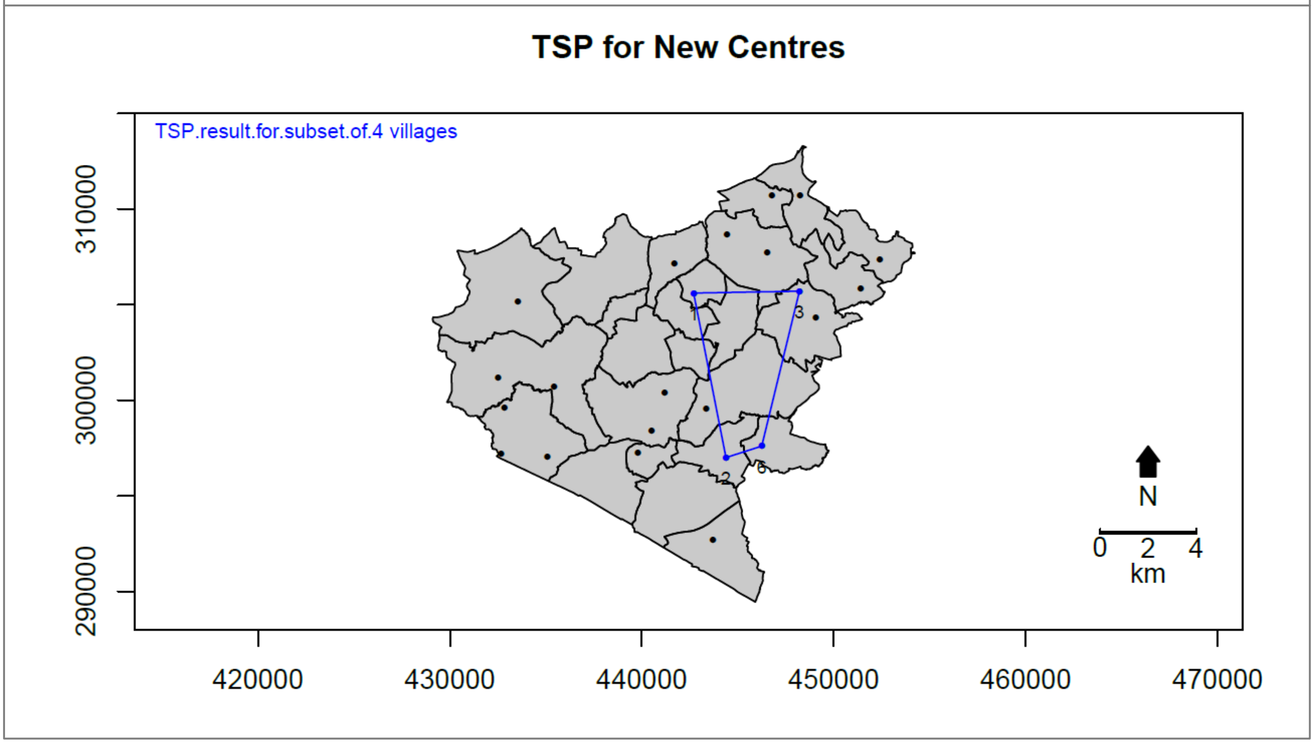
Table 7.1 List of Village names and ID included in the tours

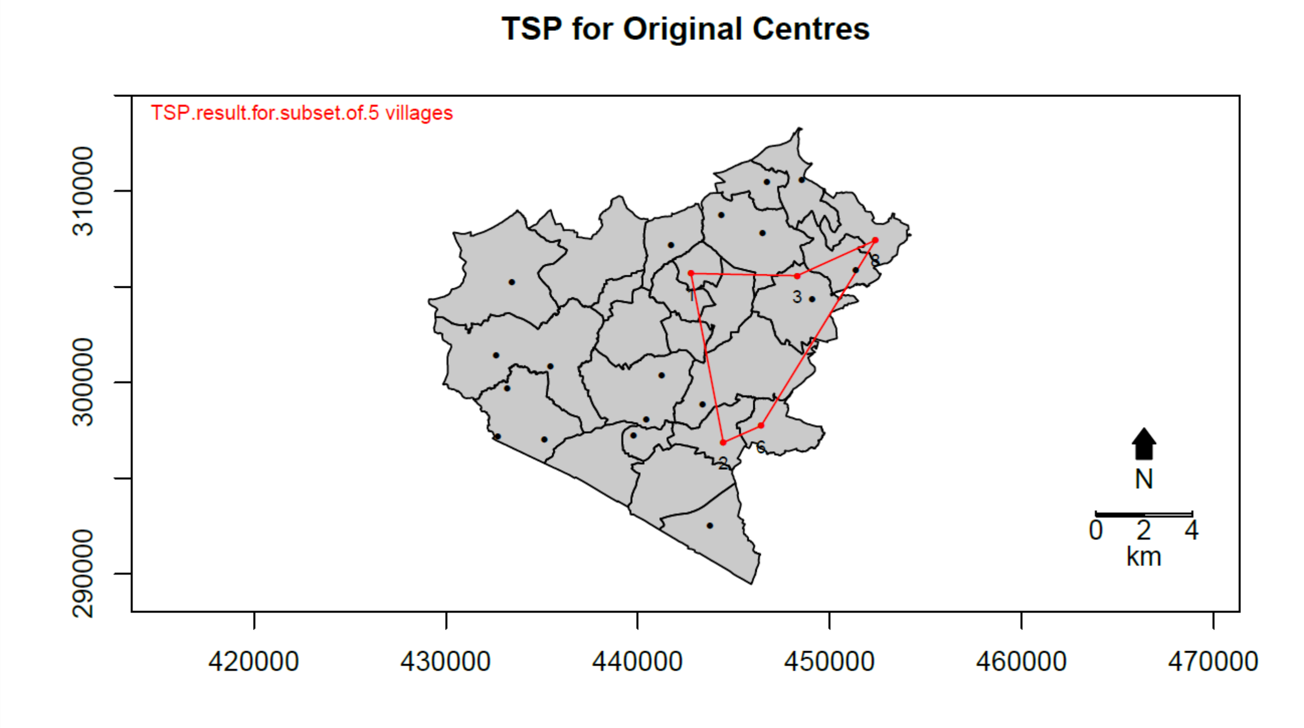
ID	Name	ID	Name	ID	Name	ID	Name
0	Bagworth	6	Earl Shilton	12	Ratby	17	Stapleton
1	Barlestone	7	Fenny Drayton	13	Ratcliffe Culey	18	Stoke Golding
2	Barwell	8	Groby	14	Sheepy Magna	19	Sutton Cheney
3	Botcheston	9	Markfield	15	Sibson	20	Thornton
4	Burbage	10	Nailstone	16	Stanton under Bardon	21	Twycross
5	Dadlington	11	Newtown Unthank			22	Witherley

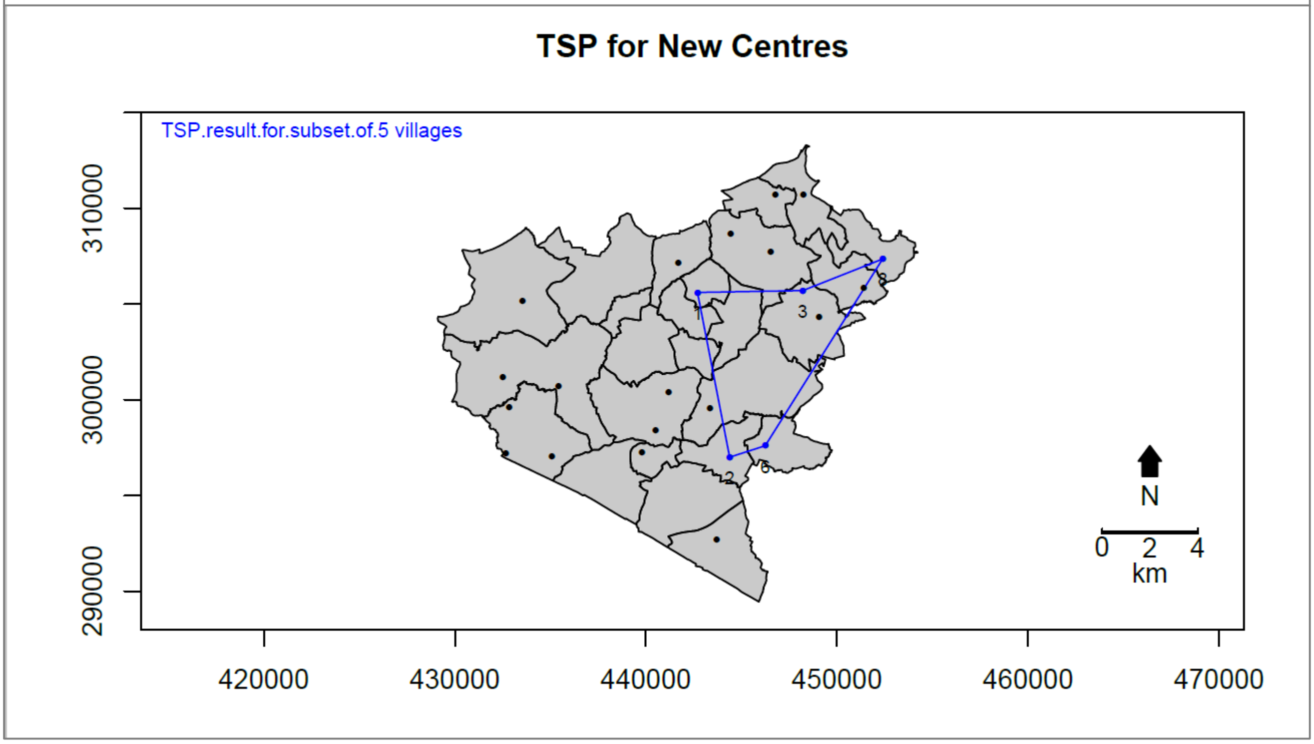


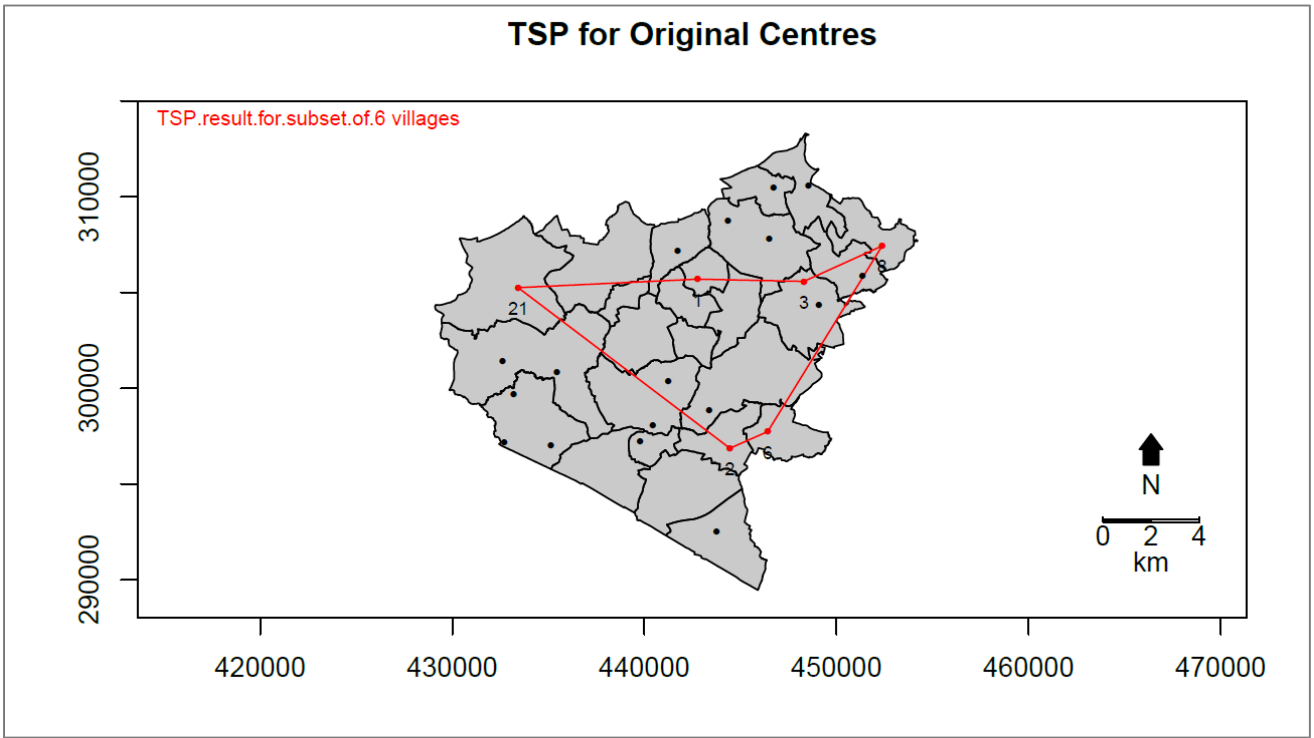


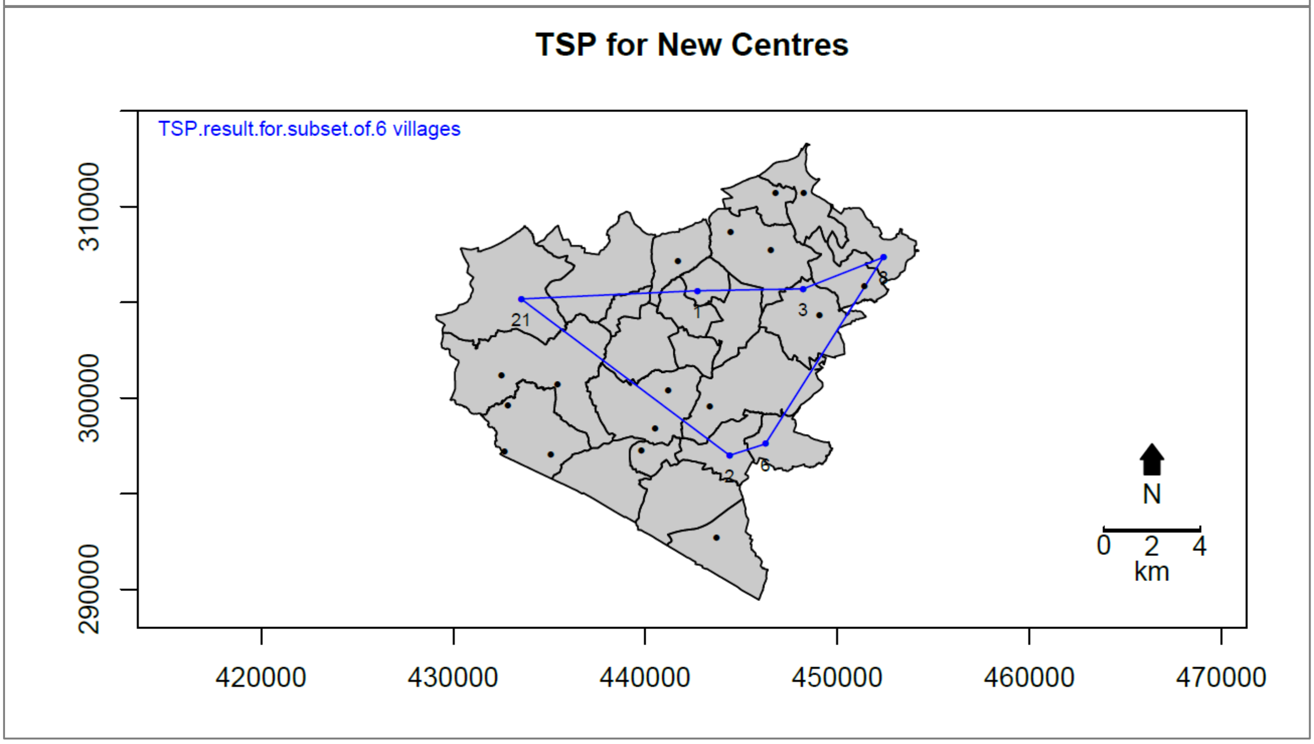

TSP for Original Centres
TSP.result.for.subset.of.4 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

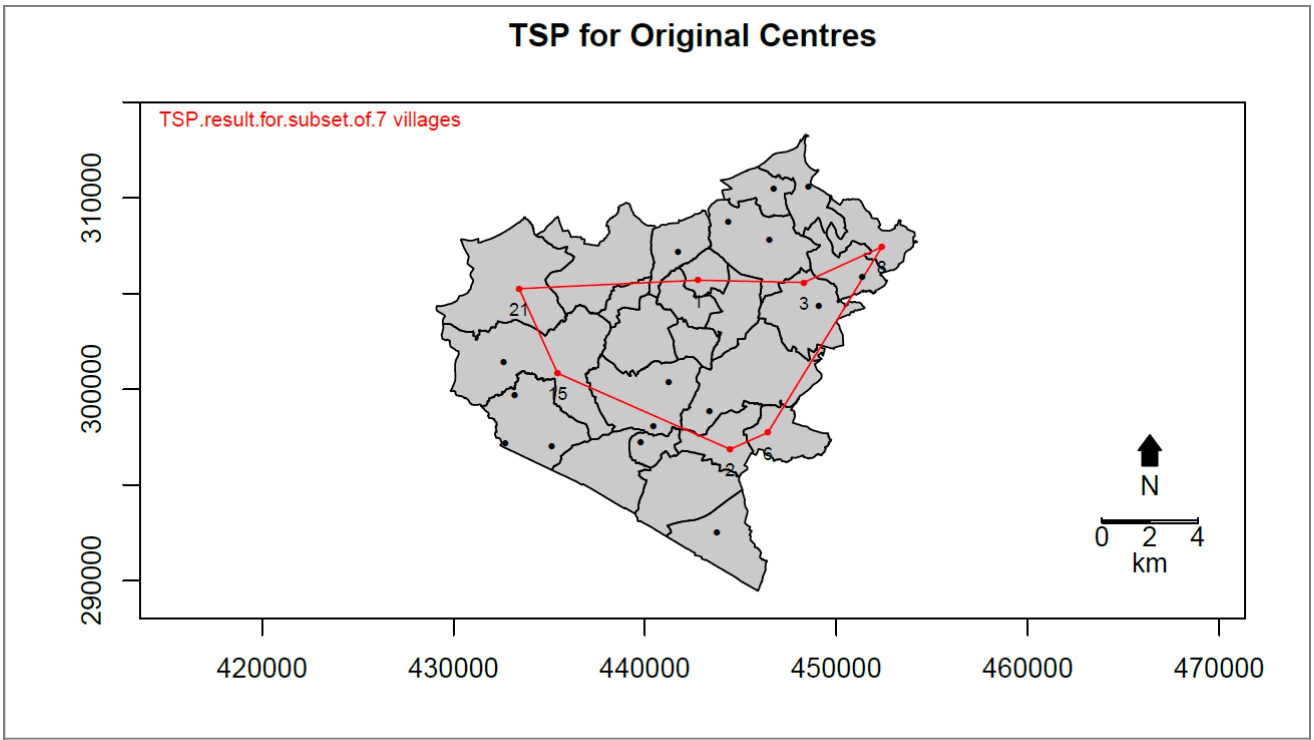

TSP for New Centres
TSP.result.for.subset.of.4 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

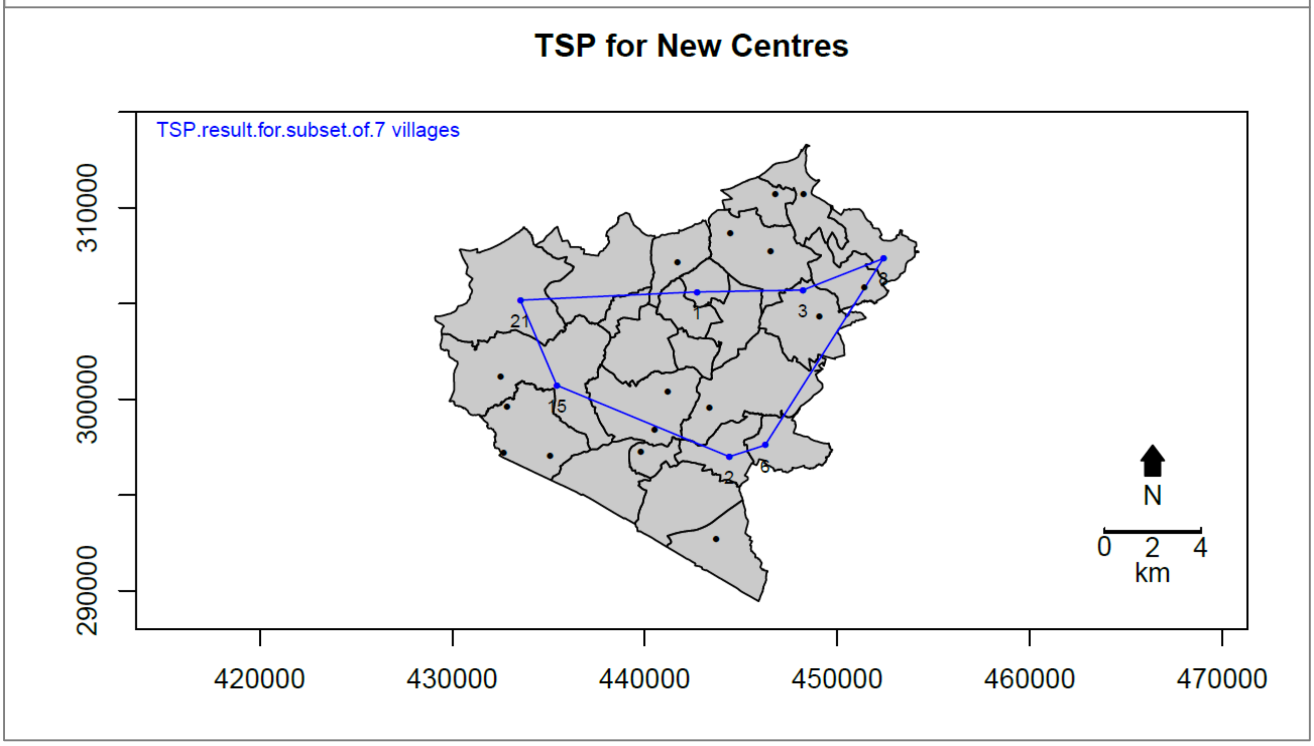

TSP for Original Centres
TSP.result.for.subset.of.5 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

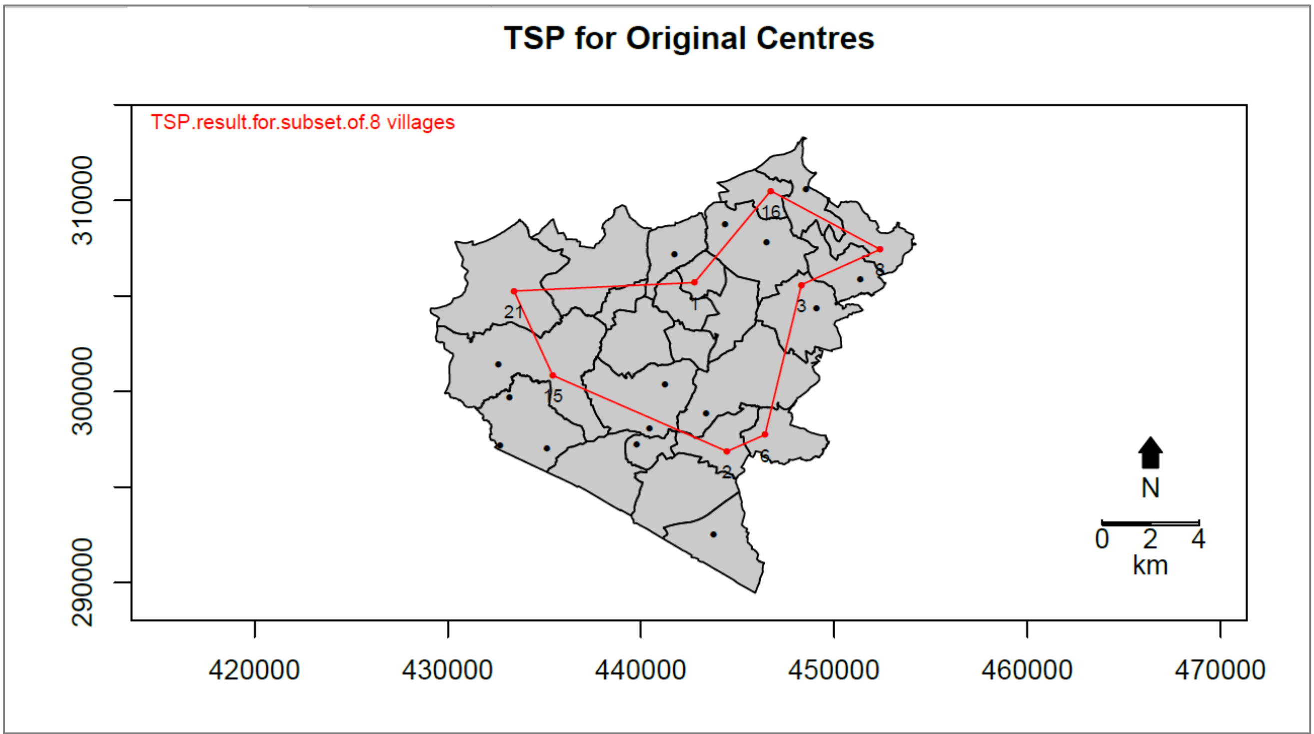

TSP for New Centres
TSP.result.for.subset.of.5 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

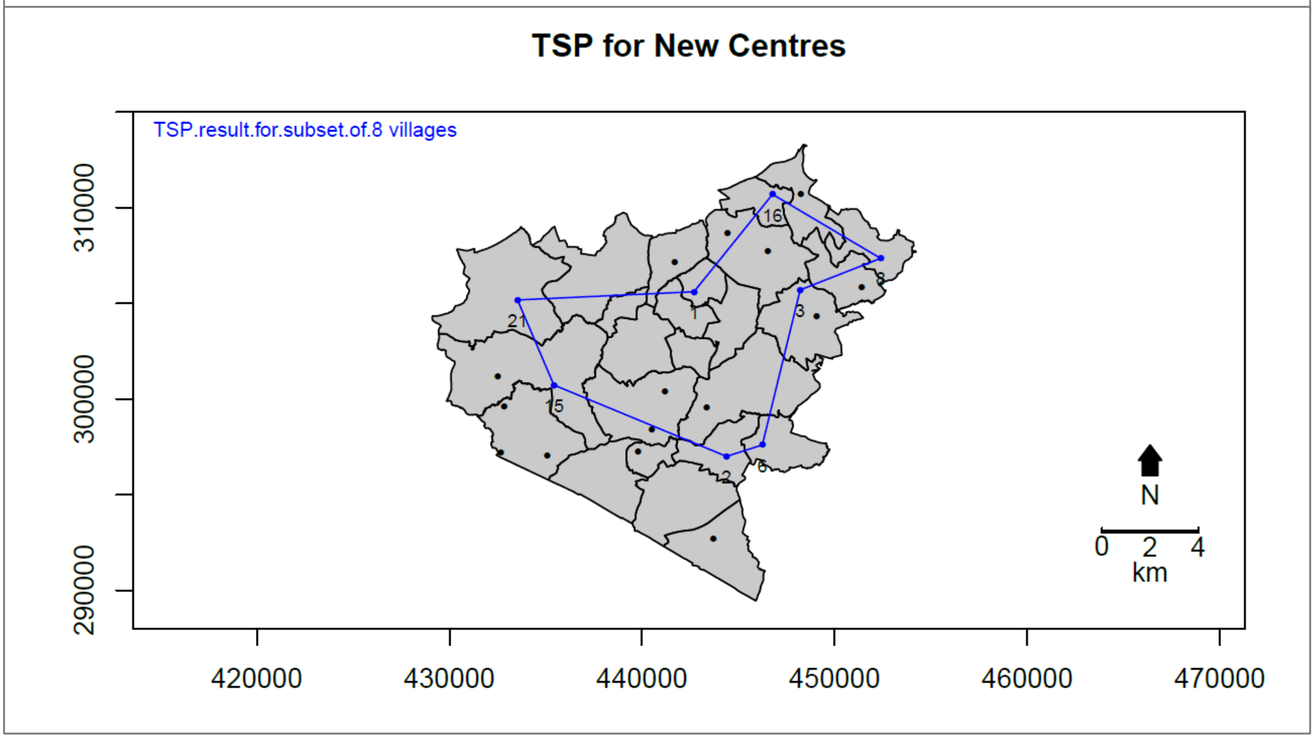

TSP for Original Centres
TSP.result.for.subset.of.6 villages
21
1
3
8
2
6
N
0 2 4
km
310000
300000
290000
420000
430000
440000
450000
460000
470000

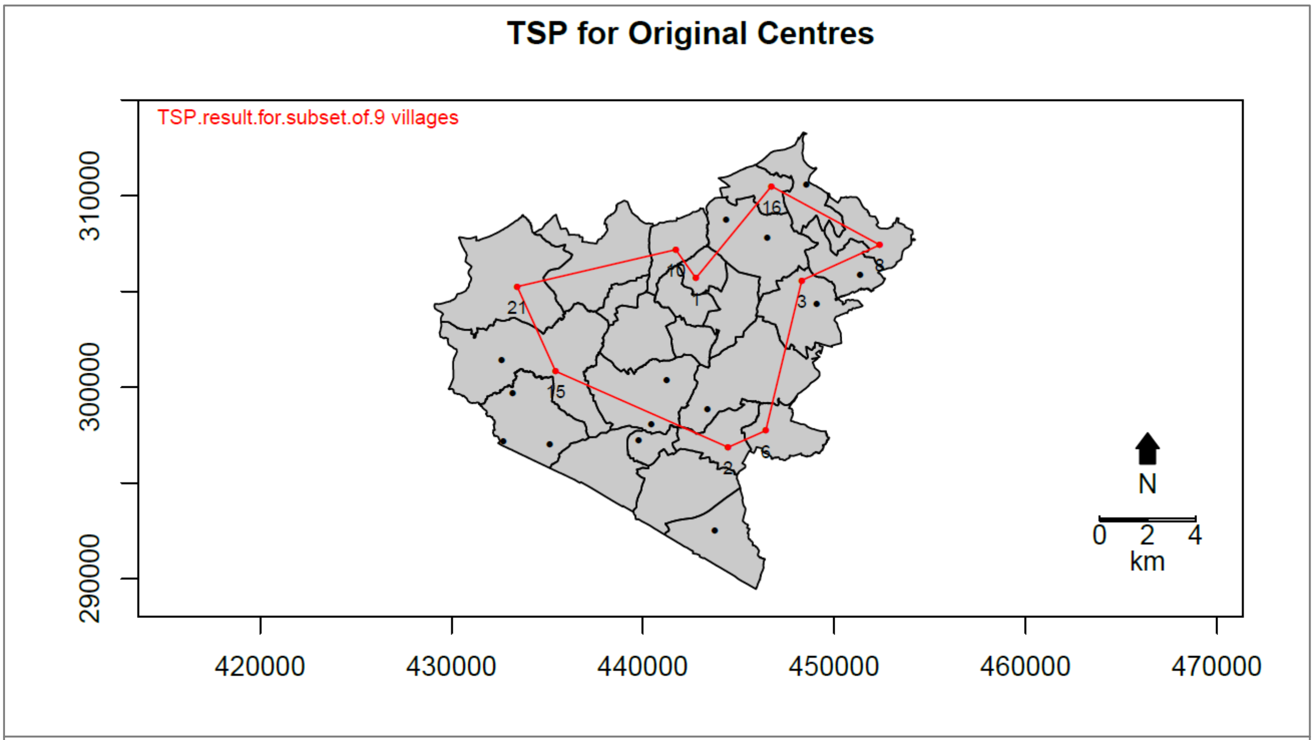

TSP for New Centres
TSP.result.for.subset.of.6 villages
21
1
3
8
2
6
N
0 2 4
km
310000
300000
290000
420000
430000
440000
450000
460000
470000


TSP for Original Centres
TSP.result.for.subset.of.7 villages
21
1
3
8
15
2
6
N
0 2 4
km
310000
300000
290000
420000
430000
440000
450000
460000
470000

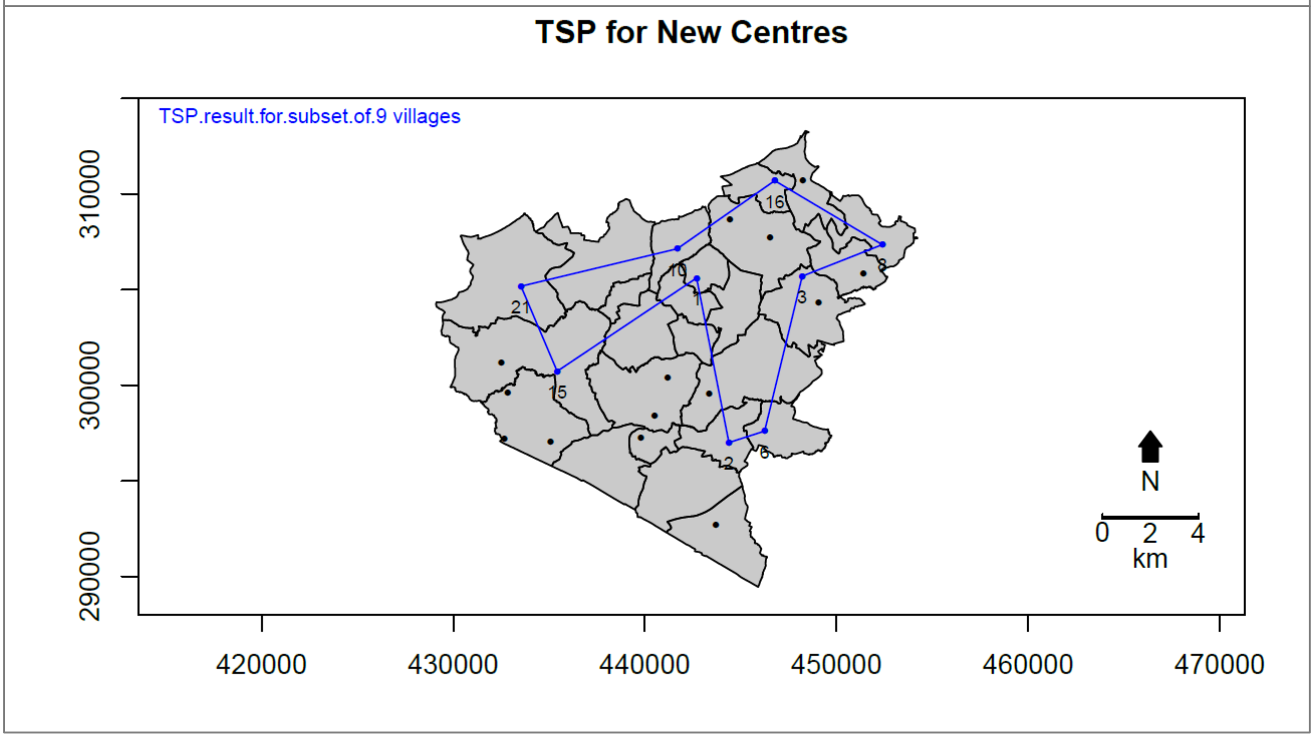

TSP for New Centres
TSP.result.for.subset.of.7 villages
21
1
3
8
15
2
6
N
0 2 4
km
310000
300000
290000
420000
430000
440000
450000
460000
470000


TSP for Original Centres
TSP.result.for.subset.of.8 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
1
21
15
2
6
3
8
N
0 2 4
km

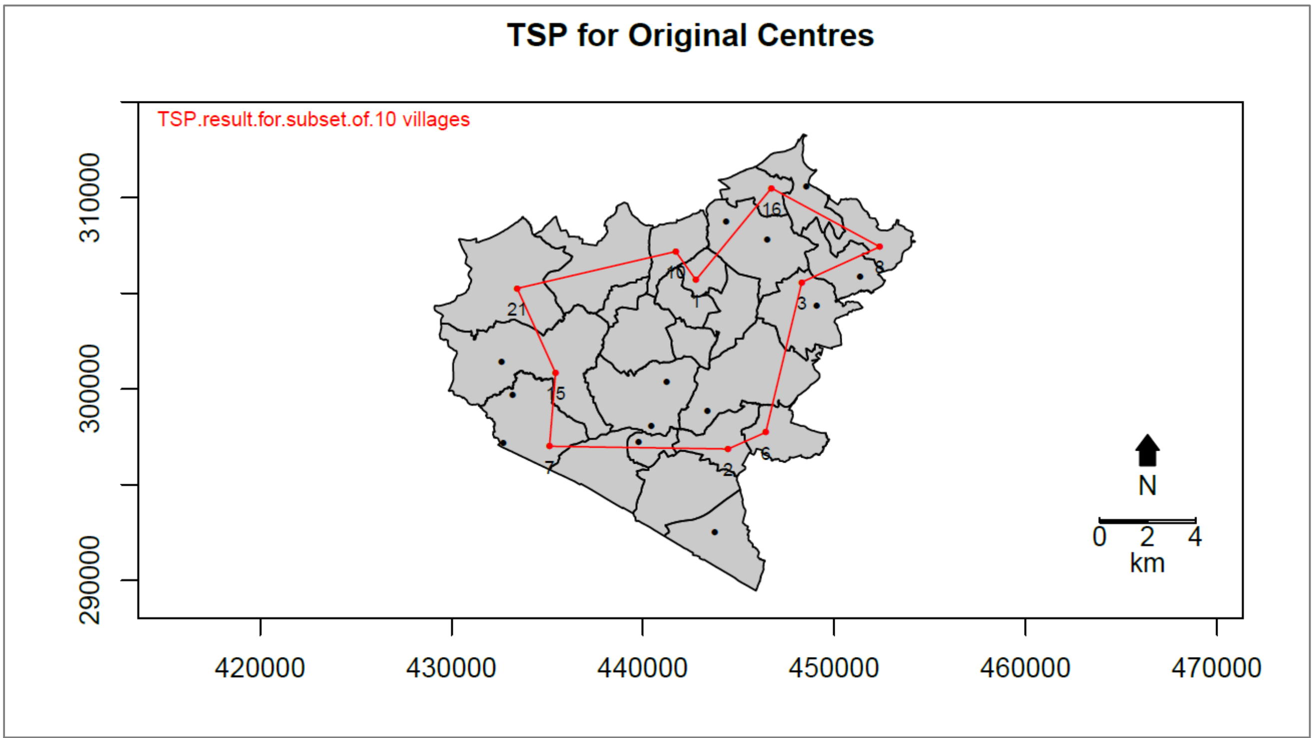

TSP for New Centres
TSP.result.for.subset.of.8 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
1
21
15
2
6
3
8
N
0 2 4
km

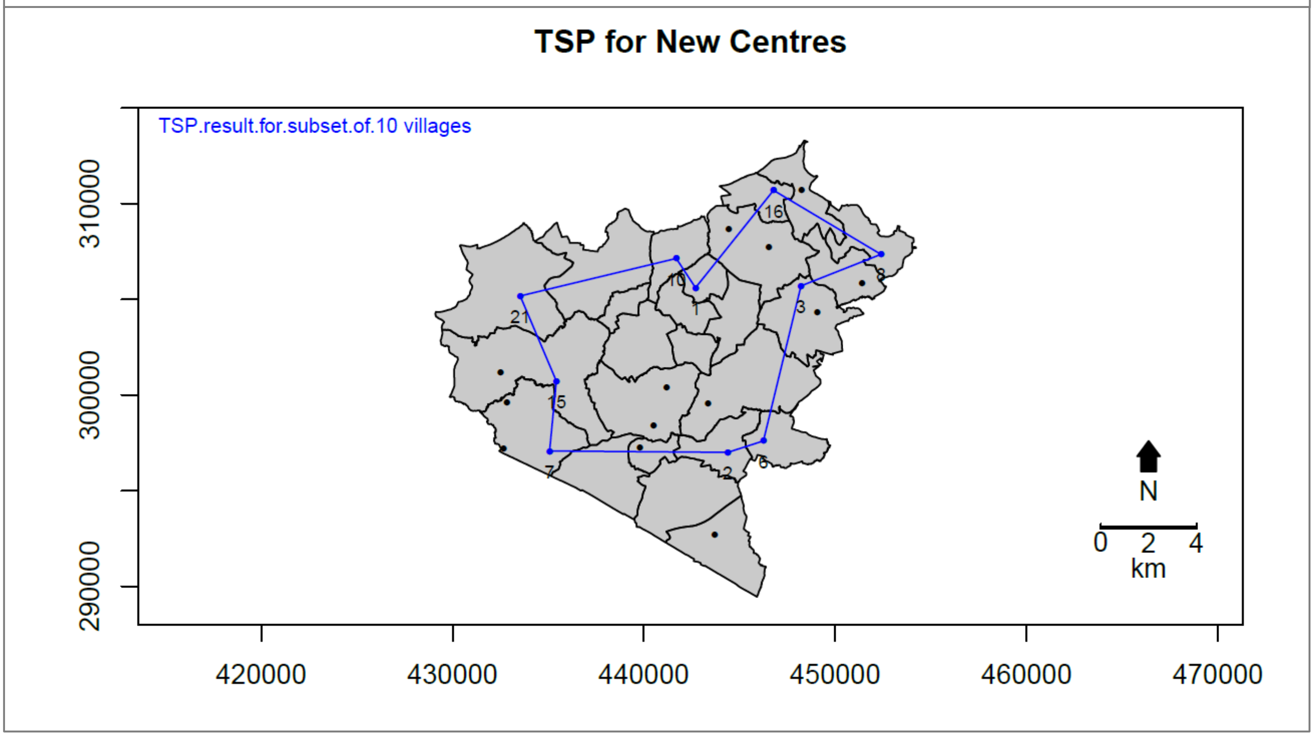


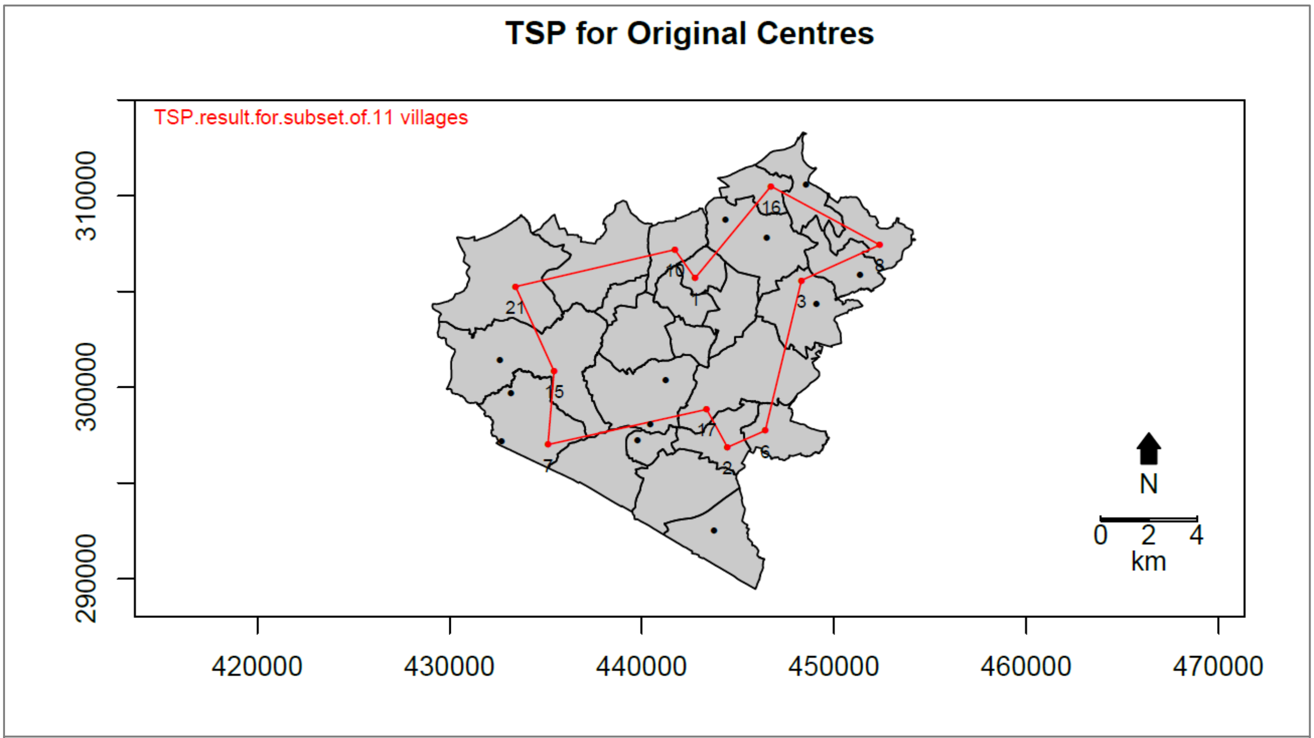
TSP for Original Centres
TSP.result.for.subset.of.9 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
10
1
21
3
8
15
2
6
N
0 2 4
km

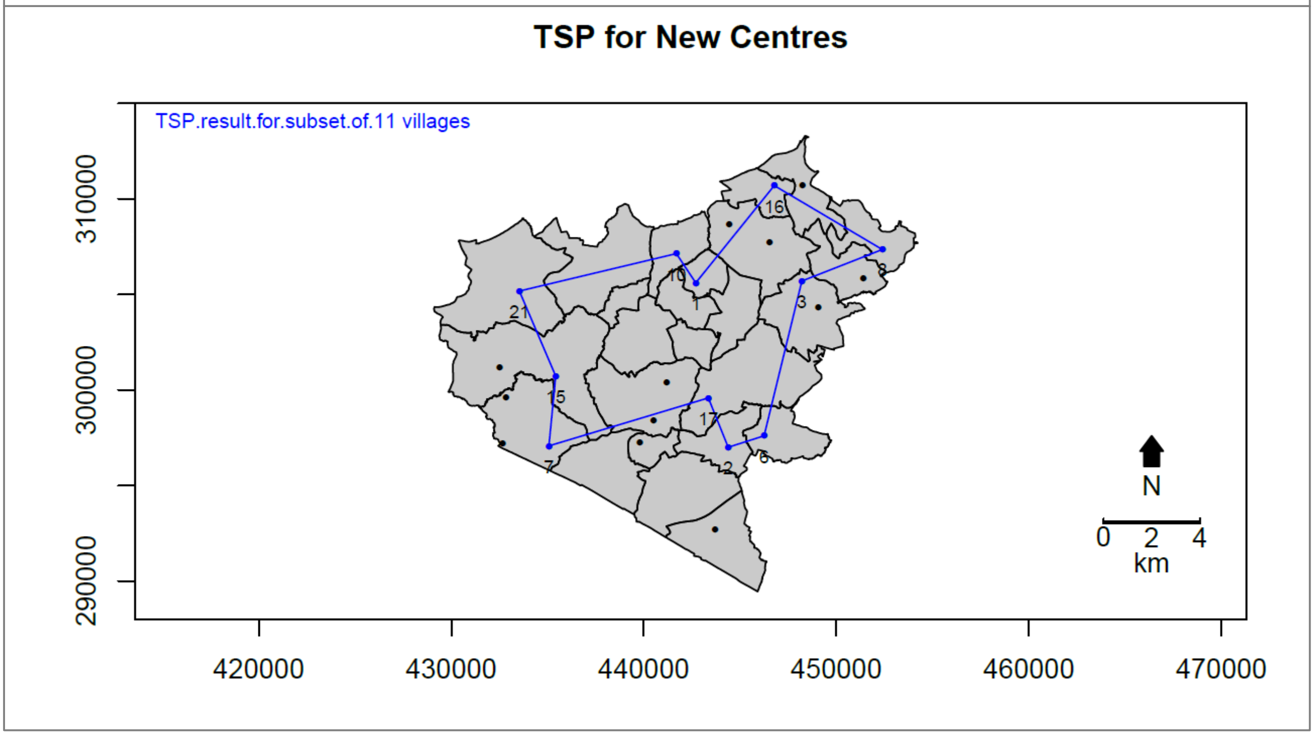


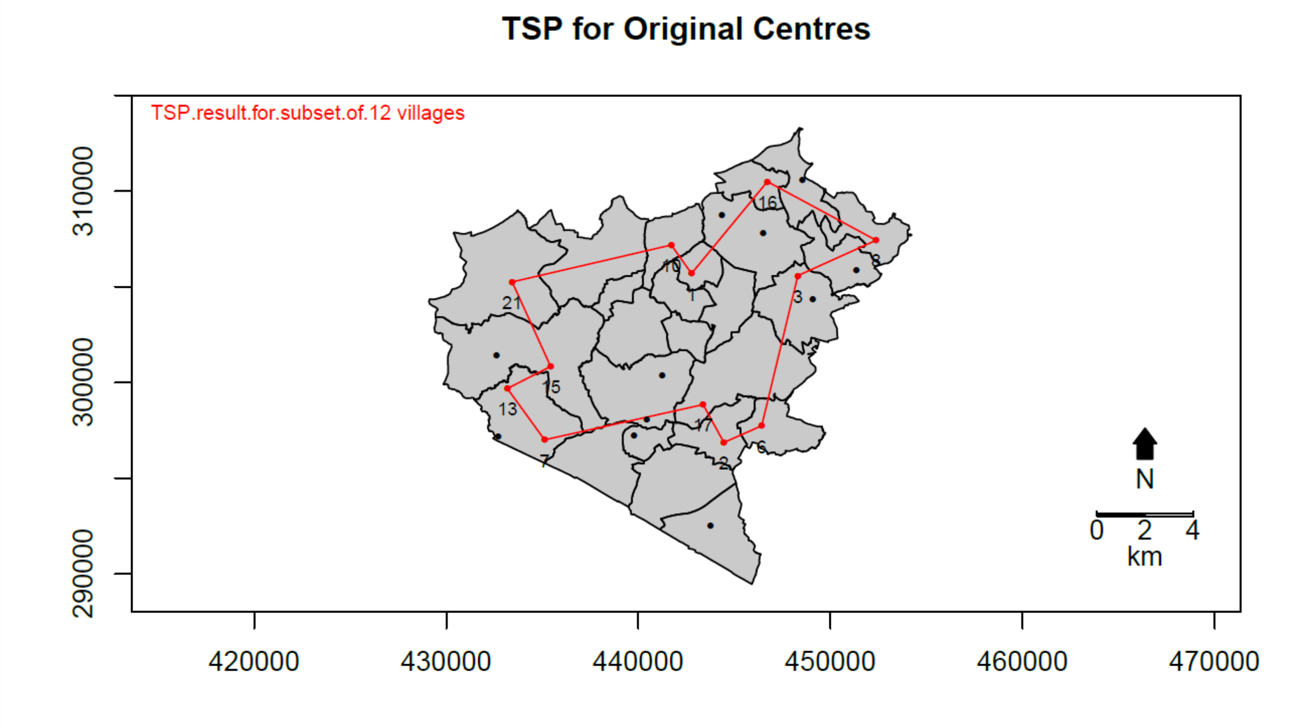
TSP for New Centres
TSP.result.for.subset.of.9 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
10
1
21
3
8
15
2
6
N
0 2 4
km

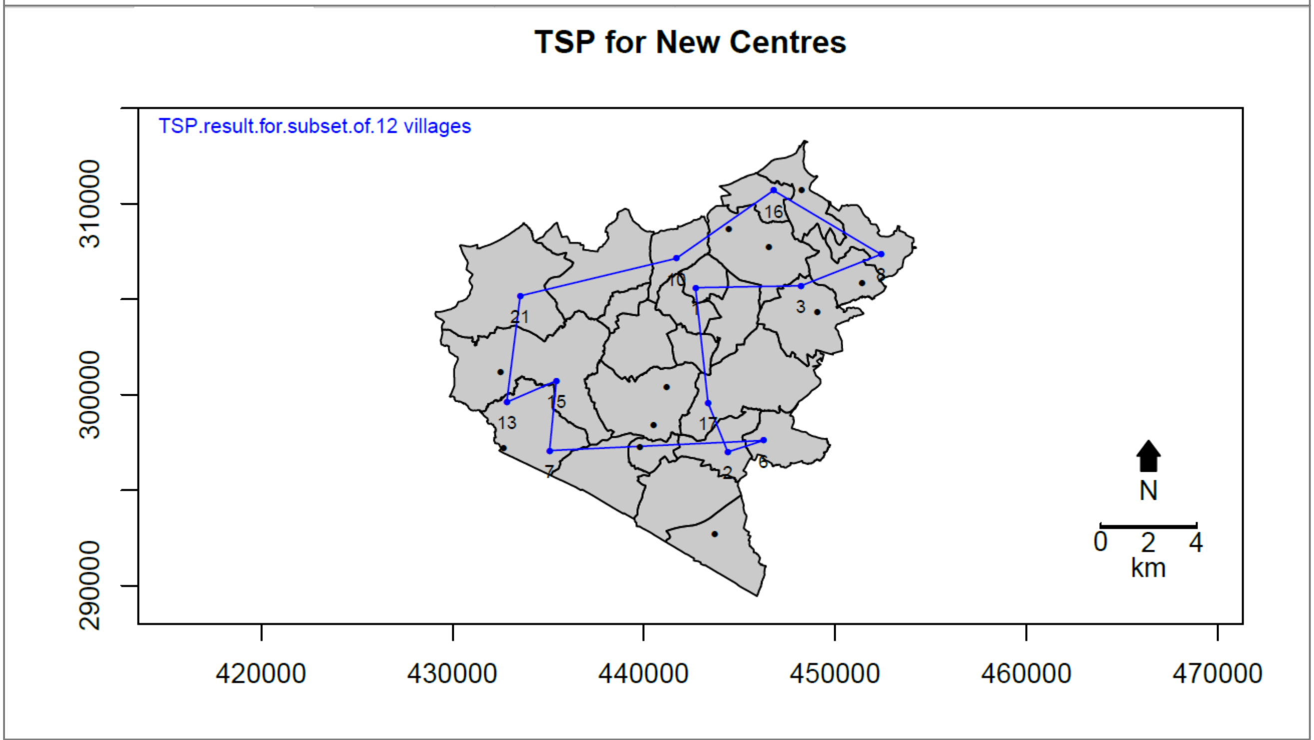

TSP for Original Centres
TSP.result.for.subset.of.10 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km


TSP for New Centres
TSP.result.for.subset.of.10 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

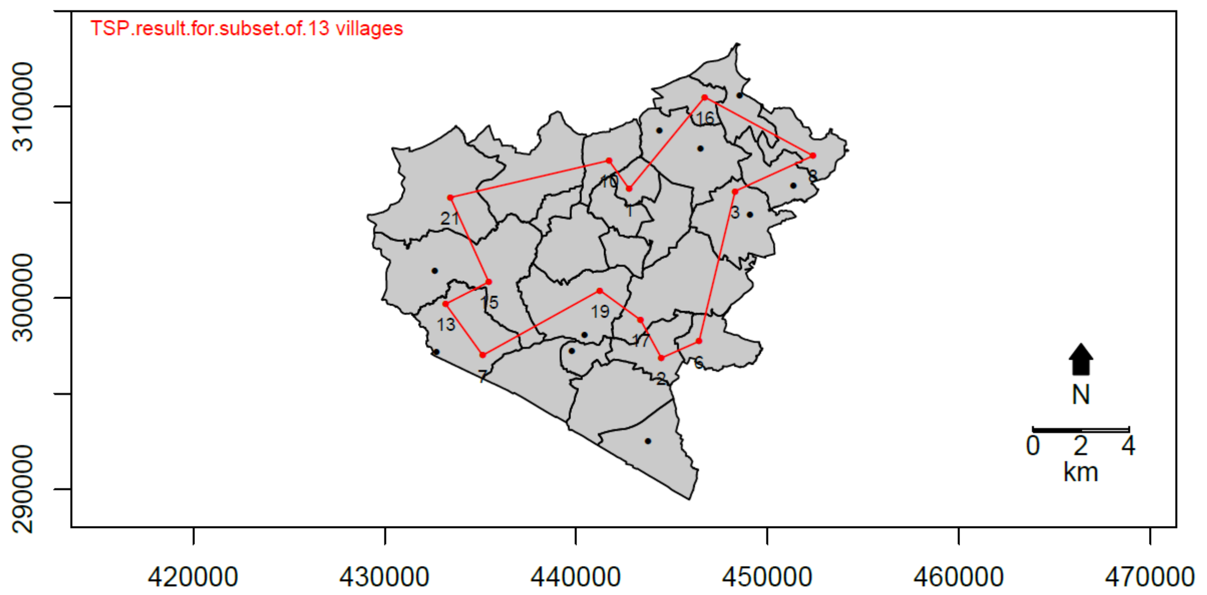

TSP for Original Centres
TSP.result.for.subset.of.11 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
10
21
1
3
8
15
17
7
2
6
N
0 2 4
km


TSP for New Centres
TSP.result.for.subset.of.11 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
10
21
1
3
8
15
17
7
2
6
N
0 2 4
km

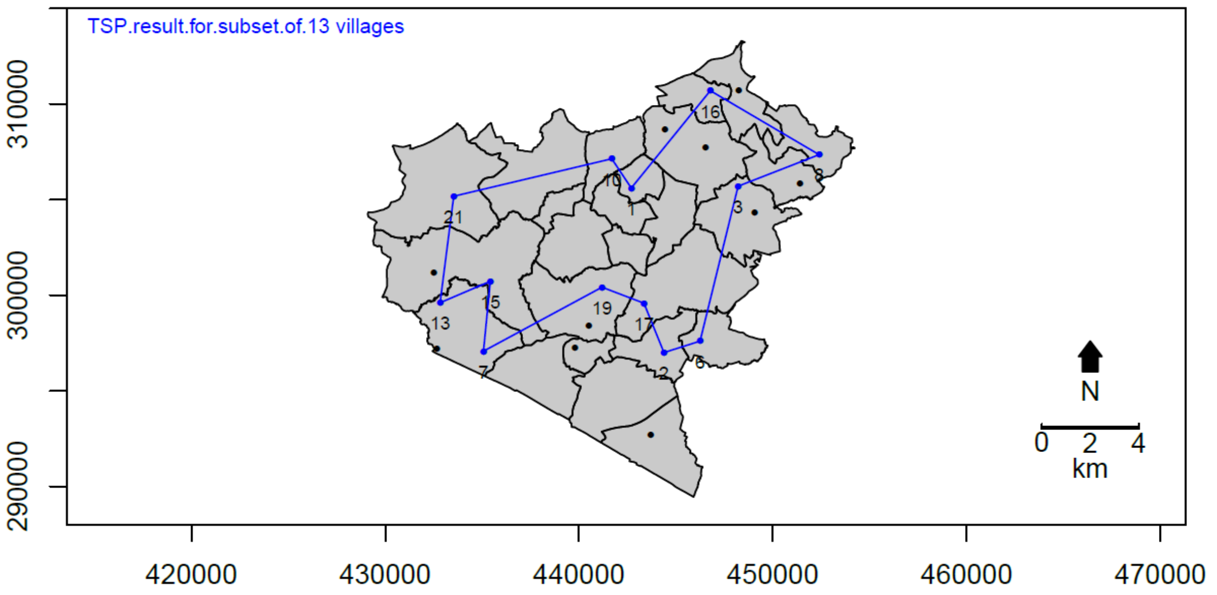

TSP for Original Centres
TSP.result.for.subset.of.12 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
10
21
3
15
13
17
7
2
6
N
0 2 4
km

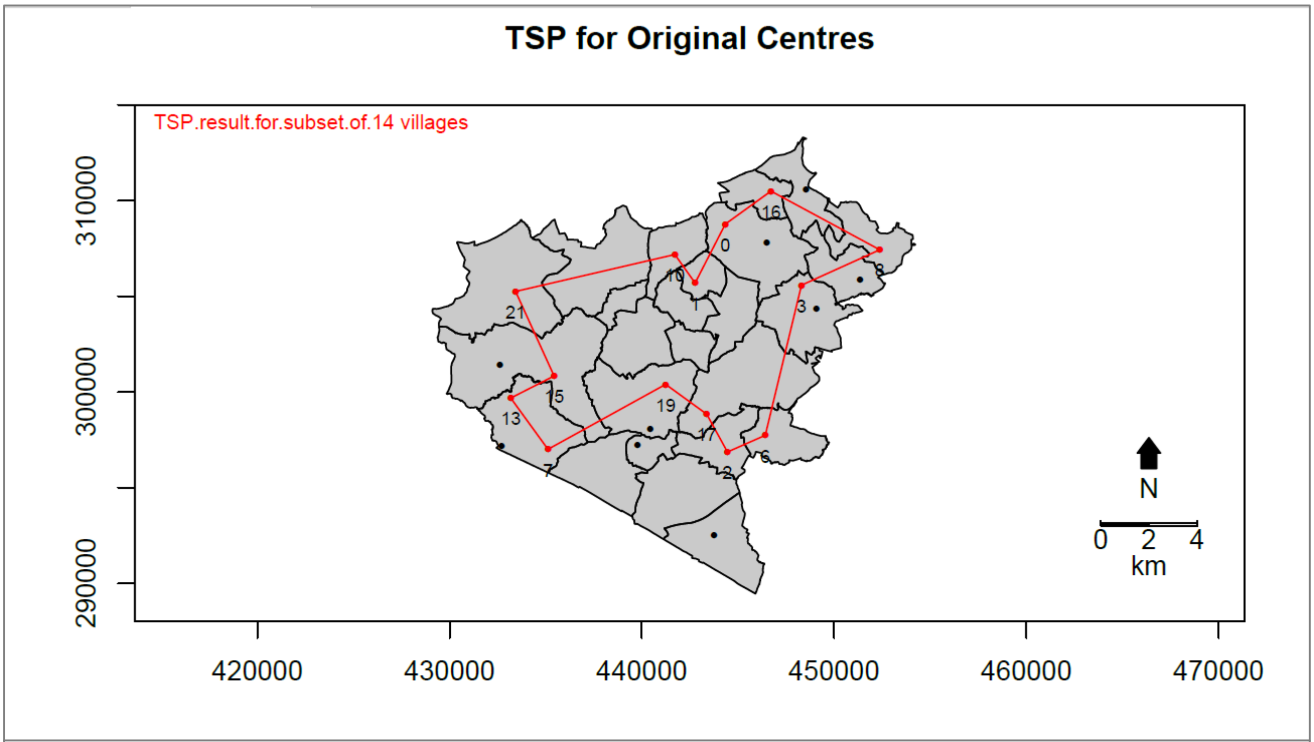

TSP for New Centres
TSP.result.for.subset.of.12 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
10
21
3
15
13
17
7
2
6
N
0 2 4
km

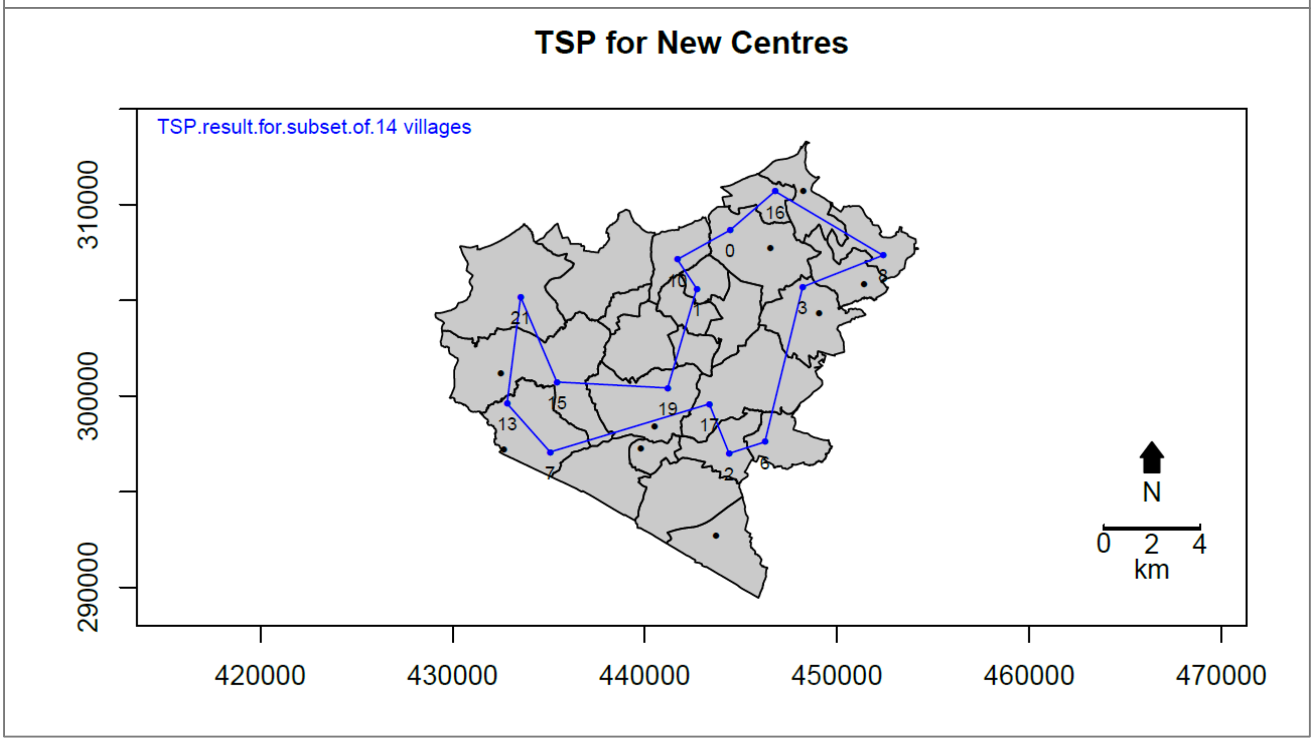
TSP for Original Centres

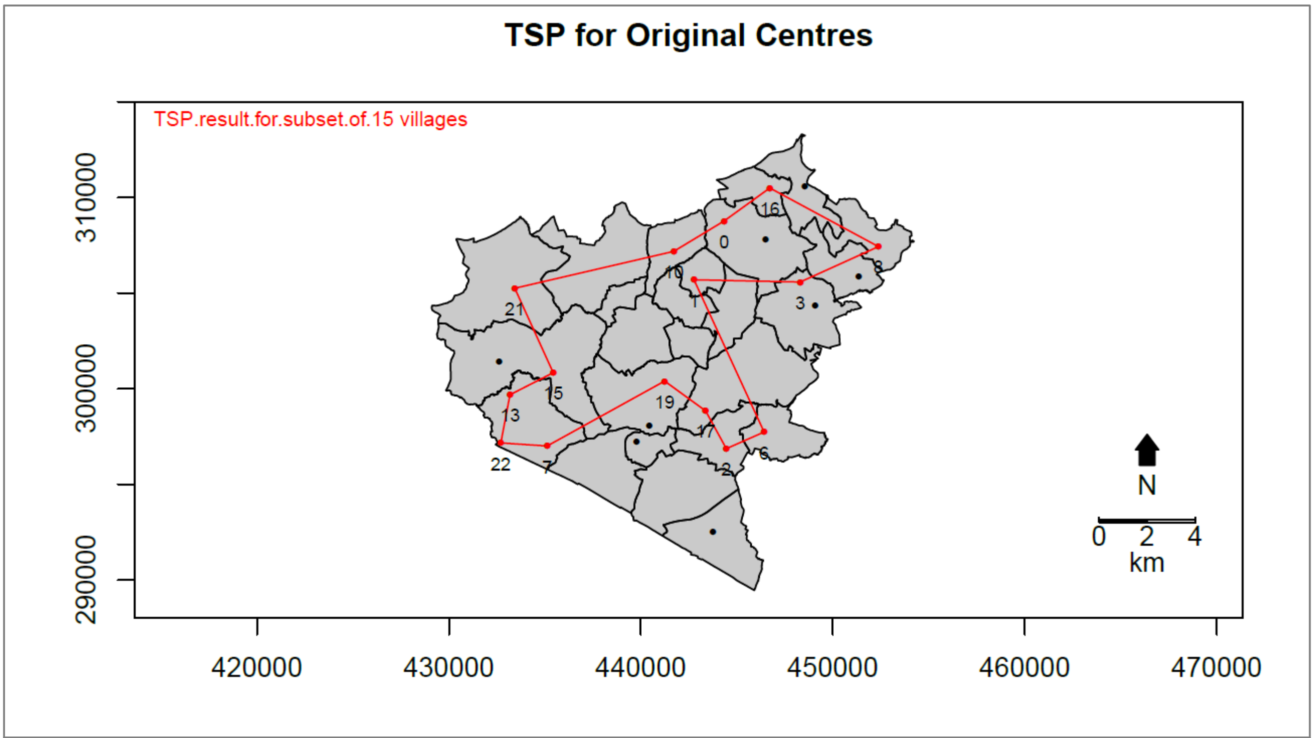


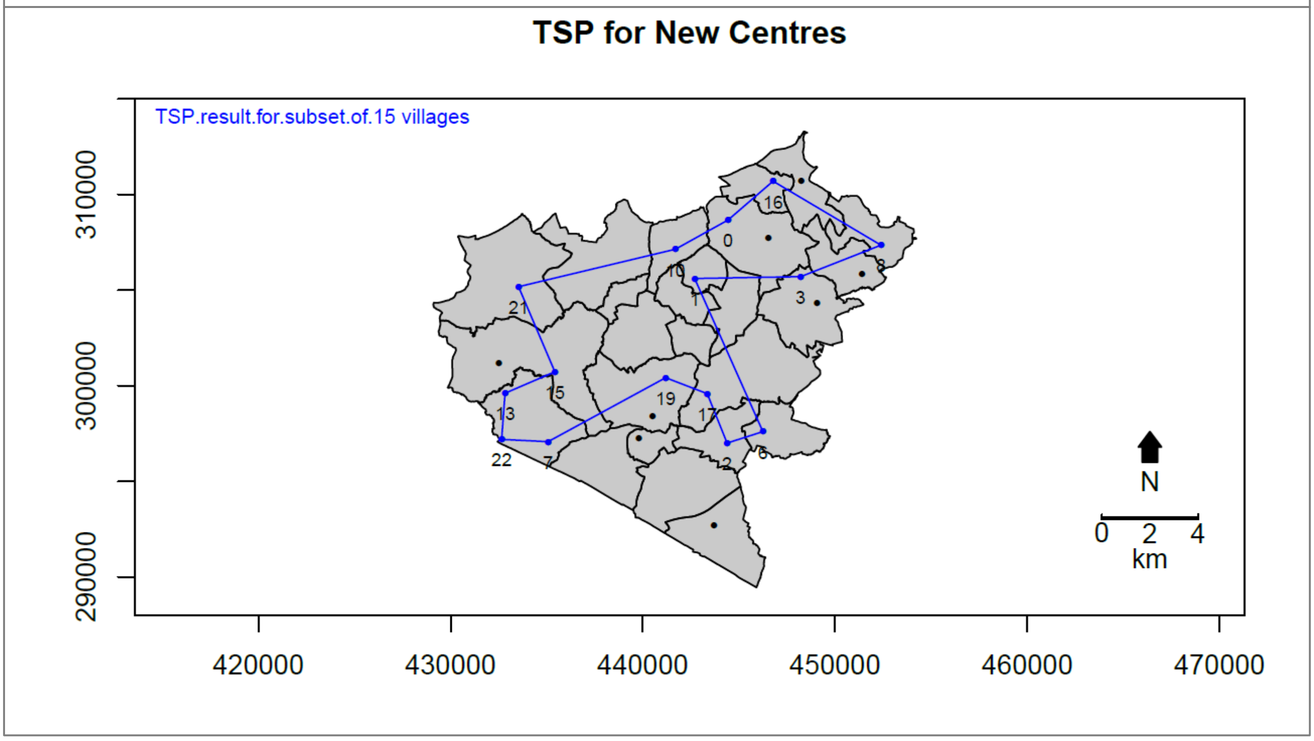
TSP for New Centres

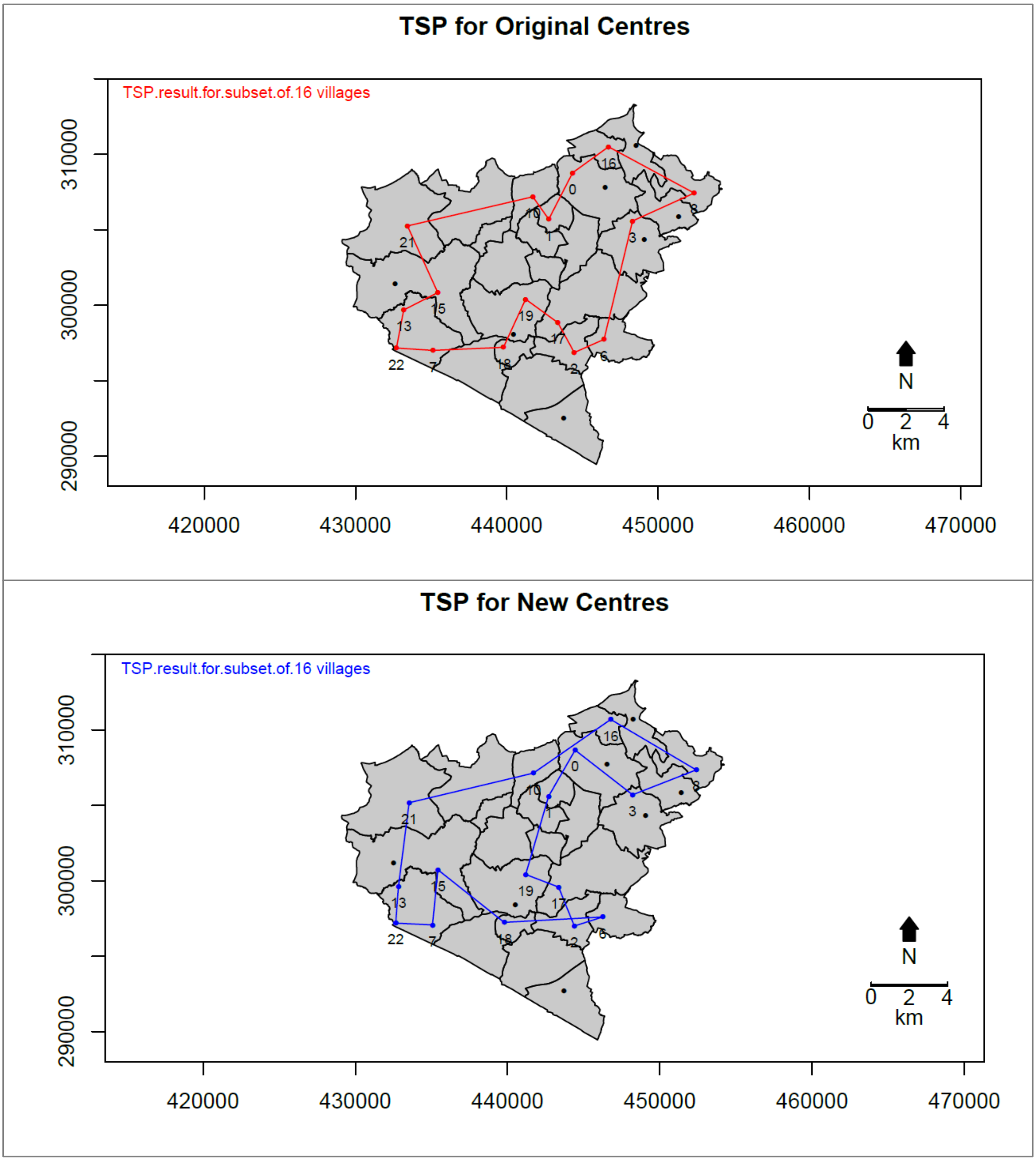


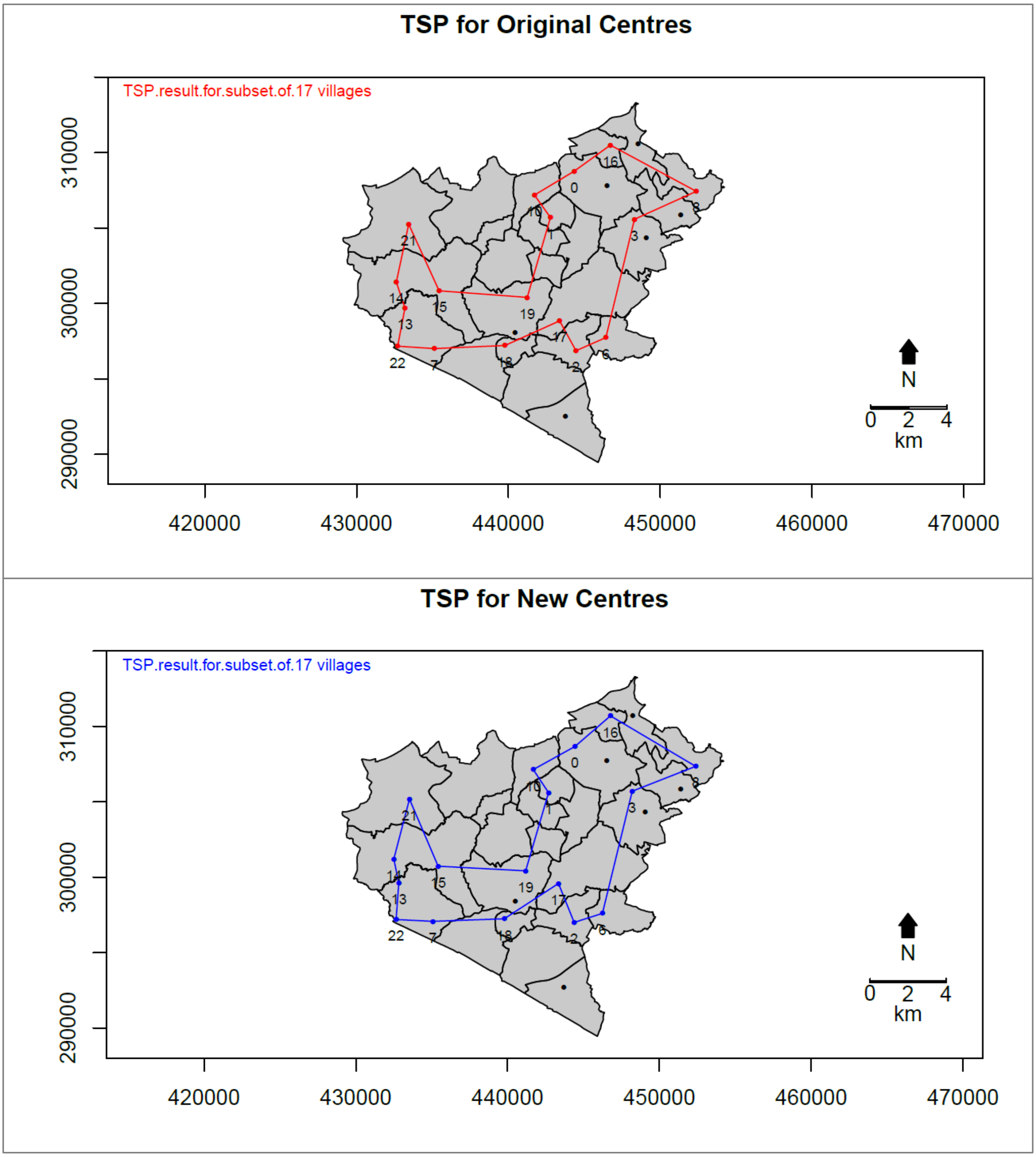

TSP for Original Centres
TSP.result.for.subset.of.14 villages
N
0 2 4
km
420000
430000
440000
450000
460000
470000
290000
300000
310000

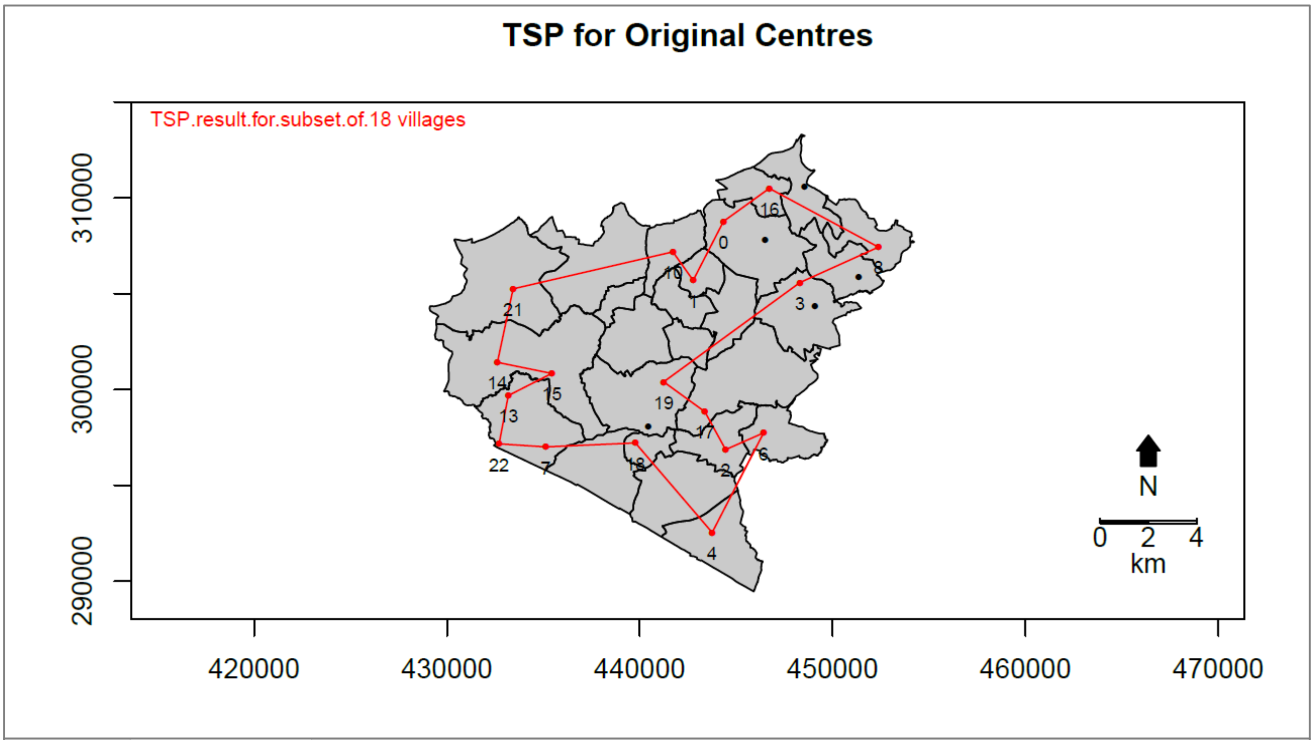

TSP for New Centres
TSP.result.for.subset.of.14 villages
N
0 2 4
km
420000
430000
440000
450000
460000
470000
290000
300000
310000

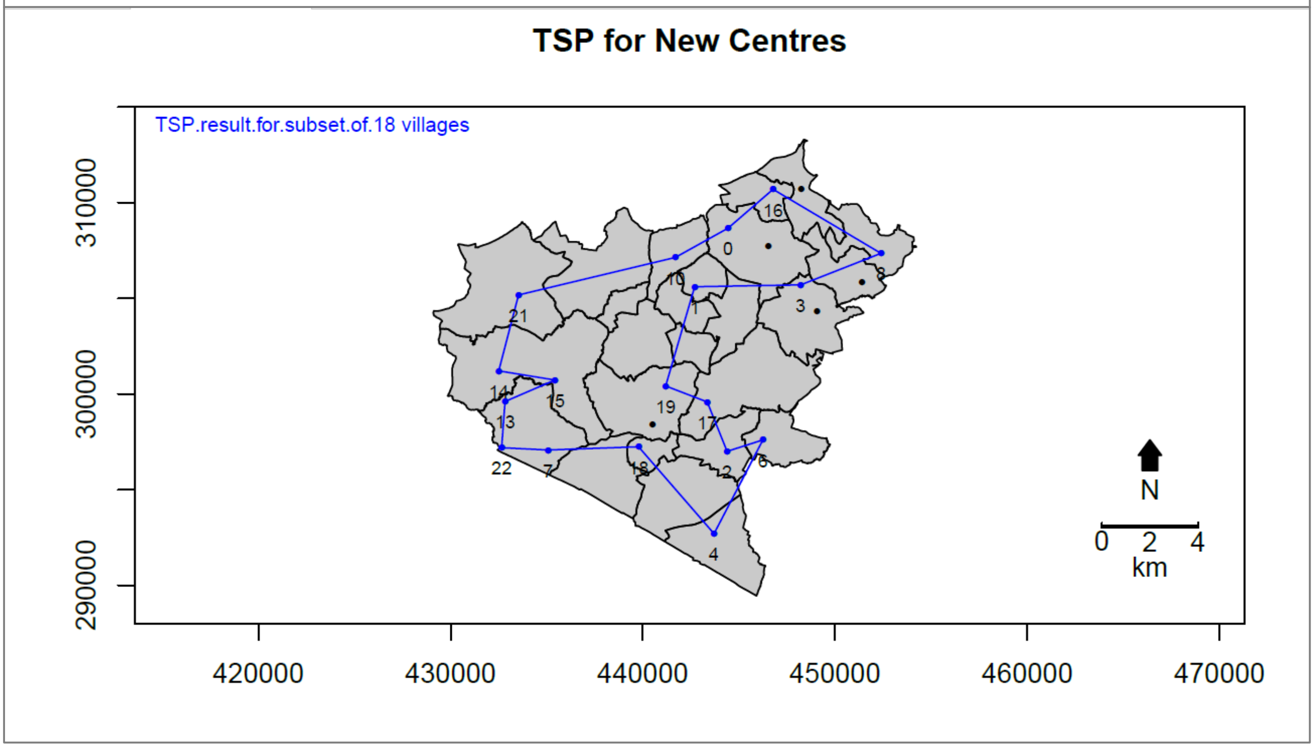

TSP for Original Centres
TSP.result.for.subset.of.15 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
0
10
21
1
3
15
13
19
17
22
7
2
6
N
0 2 4
km

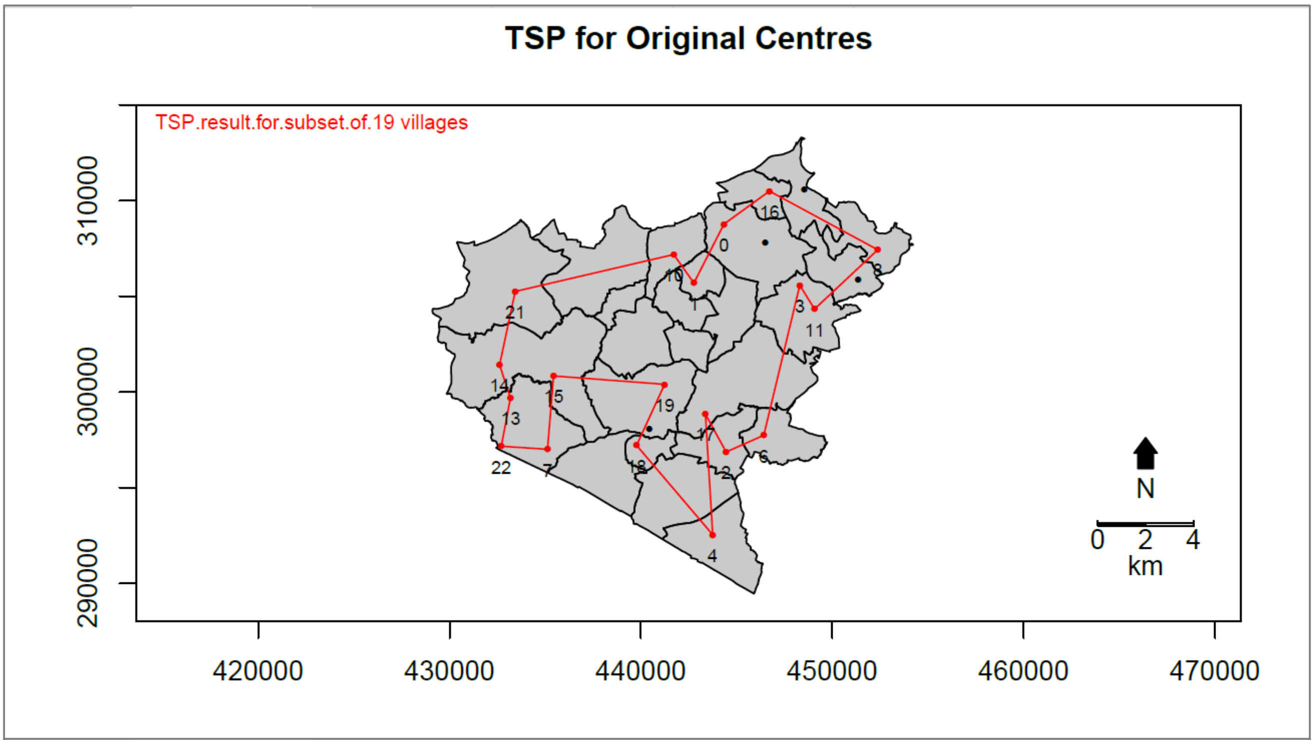

TSP for New Centres
TSP.result.for.subset.of.15 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
16
0
10
21
1
3
15
13
19
17
22
7
2
6
N
0 2 4
km

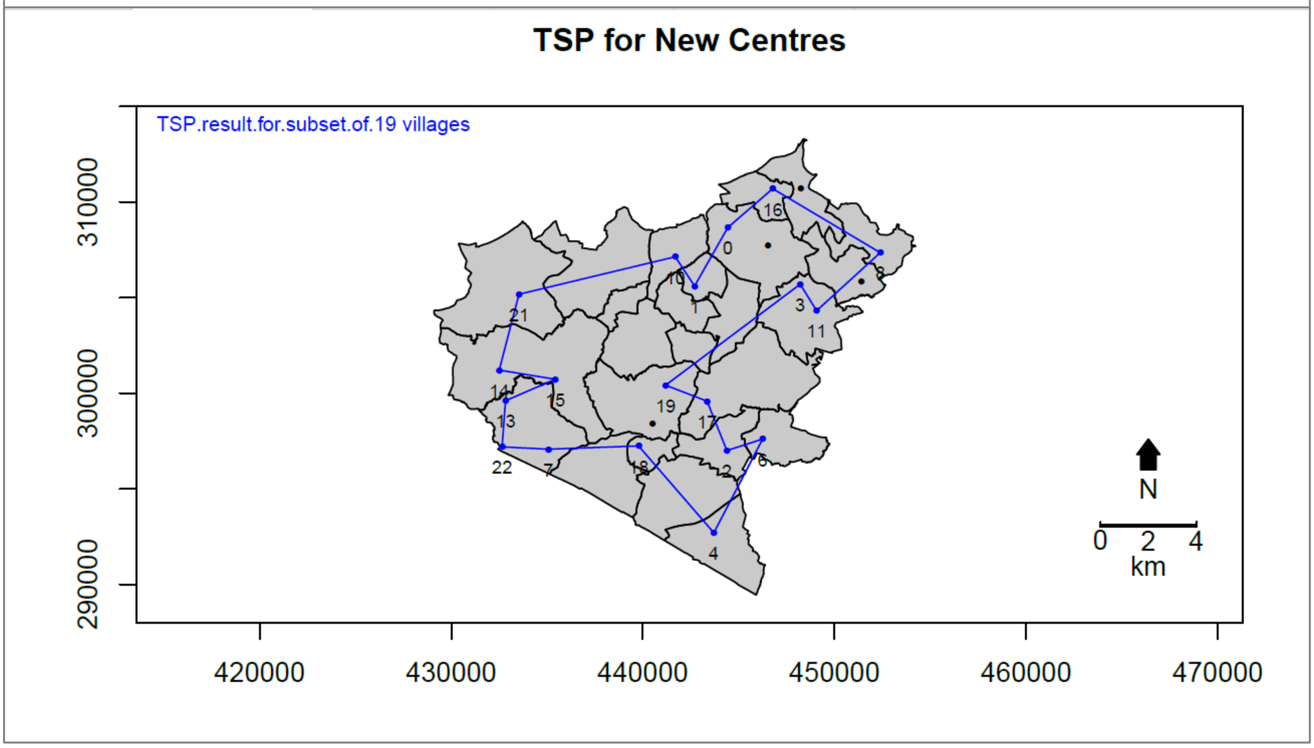

TSP for Original Centres
TSP.result.for.subset.of.16 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km
TSP for New Centres
TSP.result.for.subset.of.16 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

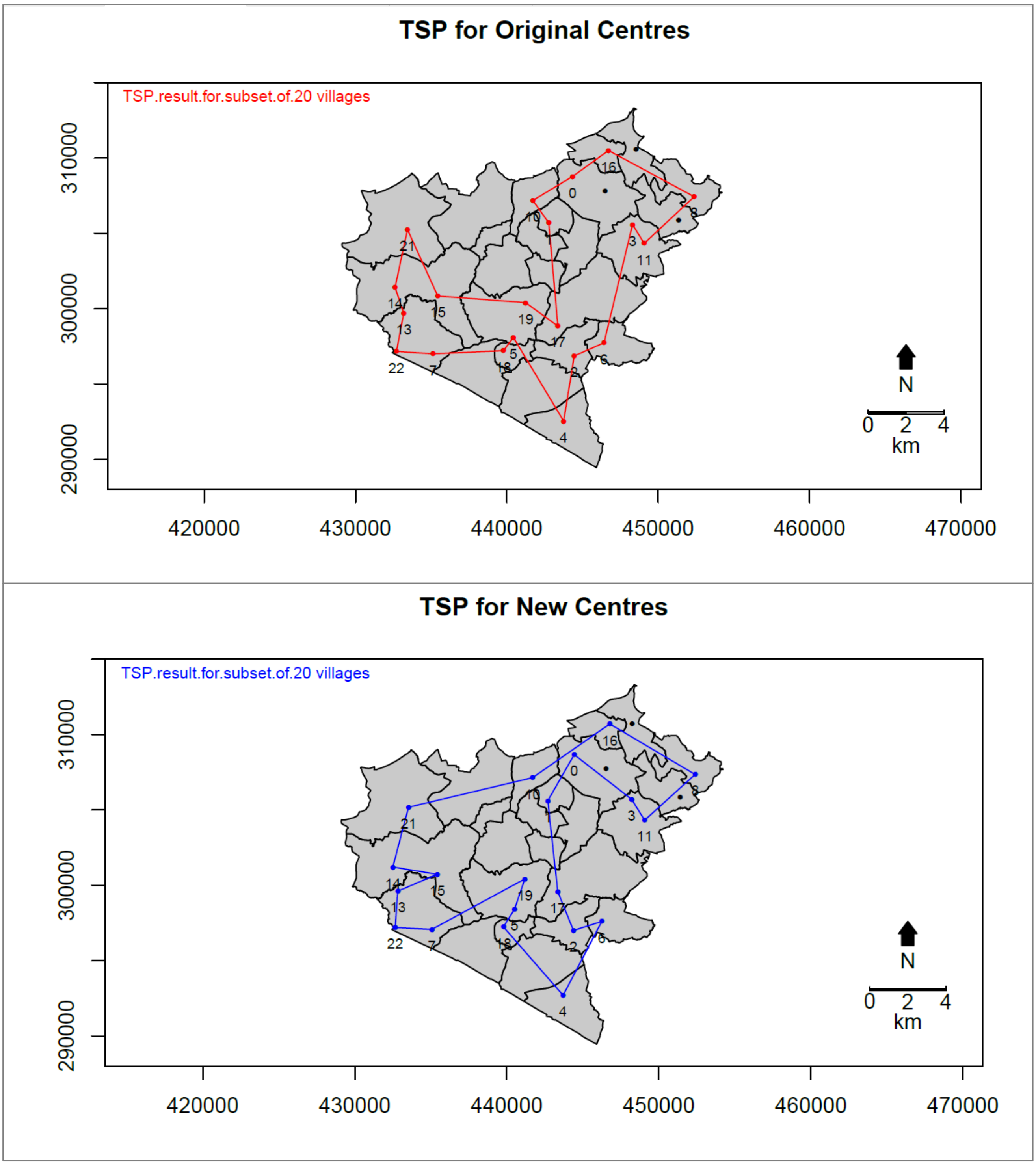

TSP for Original Centres
TSP.result.for.subset.of.17 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km
TSP for New Centres
TSP.result.for.subset.of.17 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

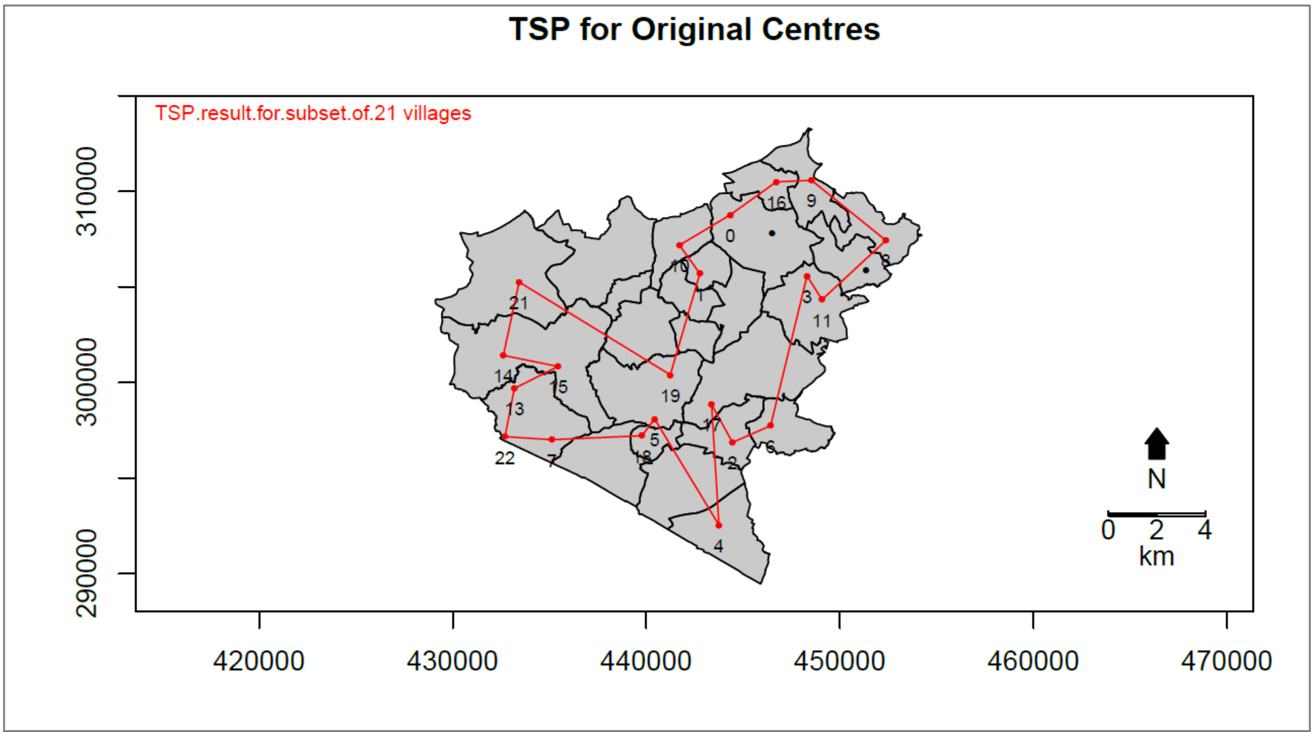

TSP for Original Centres
TSP.result.for.subset.of.18 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

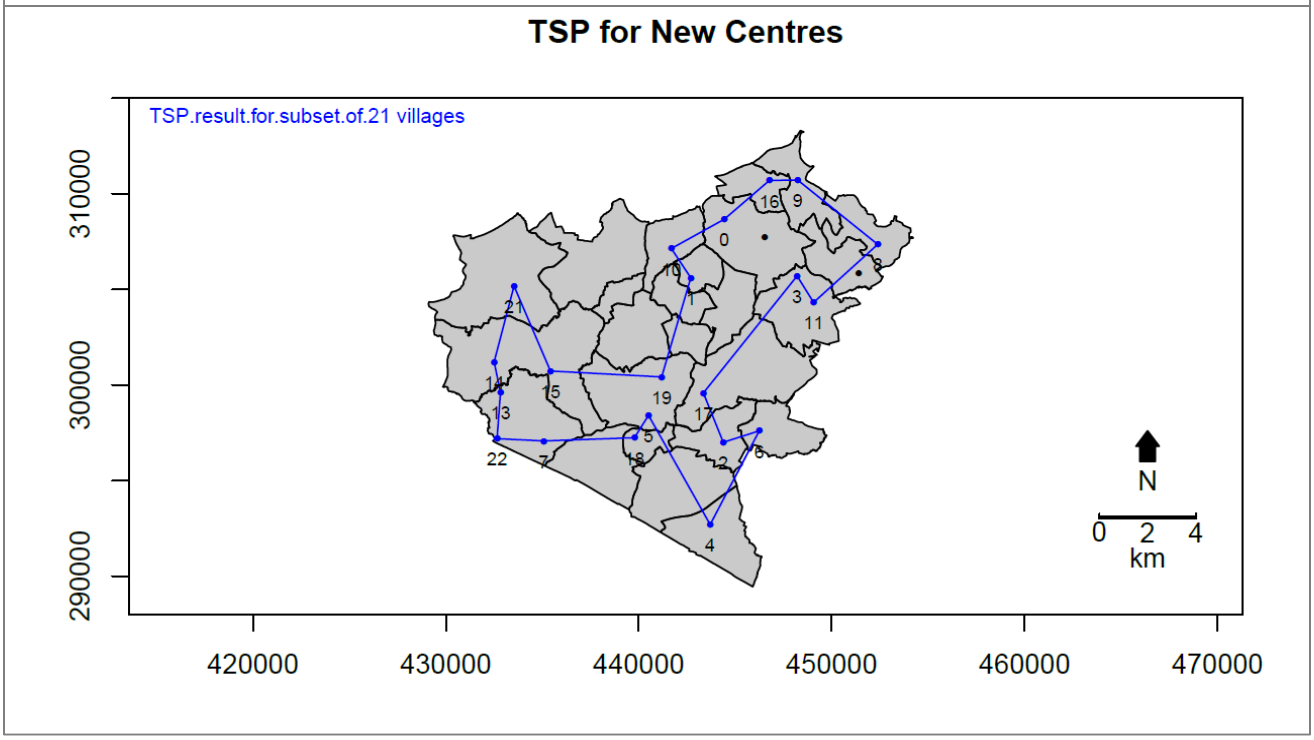

TSP for New Centres
TSP.result.for.subset.of.18 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

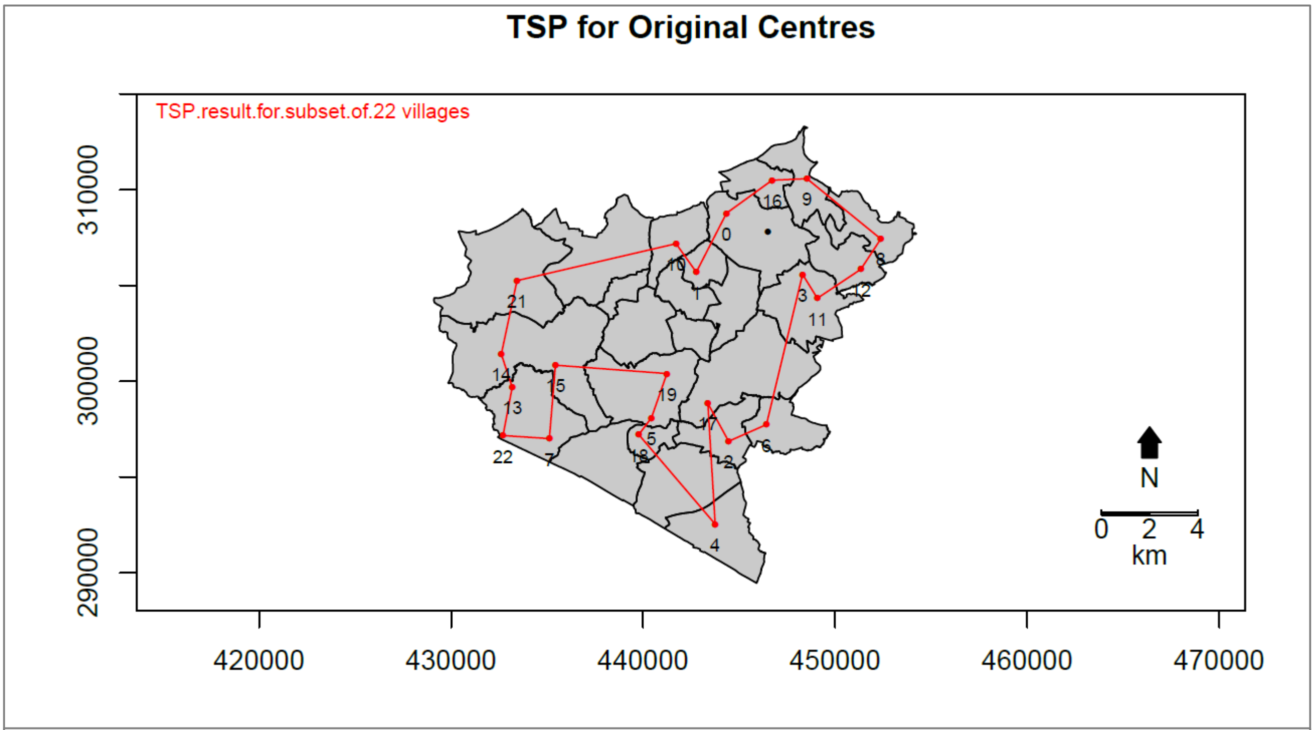

TSP for Original Centres
TSP.result.for.subset.of.19 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km


TSP for New Centres
TSP.result.for.subset.of.19 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km


TSP for Original Centres
TSP.result.for.subset.of.20 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km
TSP for New Centres
TSP.result.for.subset.of.20 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km


TSP for Original Centres
TSP.result.for.subset.of.21 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km


TSP for New Centres
TSP.result.for.subset.of.21 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km


TSP for Original Centres
TSP.result.for.subset.of.22 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

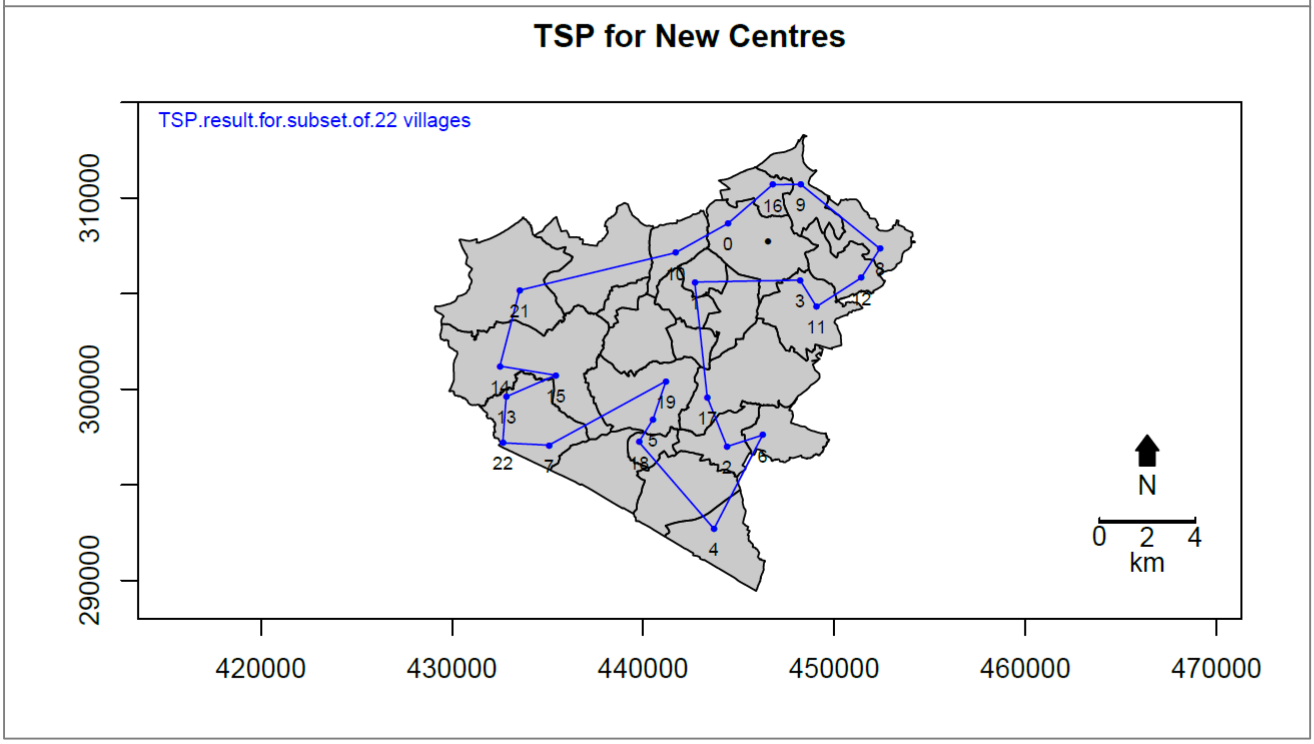

TSP for New Centres
TSP.result.for.subset.of.22 villages
310000
300000
290000
420000
430000
440000
450000
460000
470000
N
0 2 4
km

Table 7.2: Comparison of the tour lengths and paths for the possible subsets of settlements in both approach.

Subset of	Tour Length (O) KM	Tour Length (N) KM	Difference KM	Tour Path (O)	Tour Path (N)
3	23.15	23.70	0.55	Barlestone ,Barwell ,Earl Shilton	Barlestone ,Earl Shilton ,Barwell
4	30.97	31.80	0.83	Earl Shilton ,Botcheston ,Barlestone ,Barwell	Earl Shilton ,Botcheston ,Barlestone ,Barwell
5	41.06	42.01	0.95	Barlestone ,Barwell ,Earl Shilton ,Groby ,Botcheston	Barlestone ,Barwell ,Earl Shilton ,Groby ,Botcheston
6	59.20	59.93	0.73	Barlestone ,Twycross ,Barwell ,Earl Shilton ,Groby ,Botcheston	Earl Shilton ,Groby ,Botcheston ,Barlestone ,Twycross ,Barwell
7	60.46	61.42	0.96	Earl Shilton ,Groby ,Botcheston ,Barlestone ,Twycross ,Sibson ,Barwell	Twycross ,Barlestone ,Botcheston ,Groby ,Earl Shilton ,Barwell ,Sibson
8	64.78	65.88	1.10	Botcheston ,Groby ,Stanton under Bardon ,Barlestone ,Twycross ,Sibson ,Barwell ,Earl Shilton	Earl Shilton ,Botcheston ,Groby ,Stanton under Bardon ,Barlestone ,Twycross ,Sibson ,Barwell
9	66.32	72.32	6.00	Twycross ,Nailstone ,Barlestone ,Stanton under Bardon ,Groby ,Botcheston ,Earl Shilton ,Barwell ,Sibson	Sibson ,Twycross ,Nailstone ,Stanton under Bardon ,Groby ,Botcheston ,Earl Shilton ,Barwell ,Barlestone
10	68.42	69.89	1.47	Botcheston ,Earl Shilton ,Barwell ,Fenny Drayton ,Sibson ,Twycross ,Nailstone ,Barlestone ,Stanton under Bardon ,Groby	Nailstone ,Barlestone ,Stanton under Bardon ,Groby ,Botcheston ,Earl Shilton ,Barwell ,Fenny Drayton ,Sibson ,Twycross
11	69.79	72.76	2.97	Stanton under Bardon ,Groby ,Botcheston ,Earl Shilton ,Barwell ,Stapleton ,Fenny Drayton ,Sibson ,Twycross ,Nailstone ,Barlestone	Barwell ,Earl Shilton ,Botcheston ,Groby ,Stanton under Bardon ,Barlestone ,Nailstone ,Twycross ,Sibson ,Fenny Drayton ,Stapleton
12	74.56	81.89	7.33	Earl Shilton ,Botcheston ,Groby ,Stanton under Bardon ,Barlestone ,Nailstone ,Twycross ,Sibson ,Ratcliffe Culey ,Fenny Drayton ,Stapleton ,Barwell	Fenny Drayton ,Earl Shilton ,Barwell ,Stapleton ,Barlestone ,Botcheston ,Groby ,Stanton under Bardon ,Nailstone ,Twycross ,Ratcliffe Culey ,Sibson
13	77.96	79.83	1.86	Ratcliffe Culey ,Fenny Drayton ,Sutton Cheney ,Stapleton ,Barwell ,Earl Shilton ,Botcheston ,Groby ,Stanton under Bardon ,Barlestone ,Nailstone ,Twycross ,Sibson	Groby ,Stanton under Bardon ,Barlestone ,Nailstone ,Twycross ,Ratcliffe Culey ,Sibson ,Fenny Drayton ,Sutton Cheney ,Stapleton ,Barwell ,Earl Shilton ,Botcheston
14	79.99	87.93	7.94	Bagworth ,Barlestone ,Nailstone ,Twycross ,Sibson ,Ratcliffe Culey ,Fenny Drayton ,Sutton Cheney ,Stapleton ,Barwell ,Earl Shilton ,Botcheston ,Groby ,Stanton under Bardon	Nailstone ,Bagworth ,Stanton under Bardon ,Groby ,Botcheston ,Earl Shilton ,Barwell ,Stapleton ,Fenny Drayton ,Ratcliffe Culey ,Twycross ,Sibson ,Sutton Cheney ,Barlestone
15	86.70	87.30	0.60	Witherley ,Ratcliffe Culey ,Sibson ,Twycross ,Nailstone ,Bagworth ,Stanton under Bardon ,Groby ,Botcheston ,Barlestone ,Earl Shilton ,Barwell ,Stapleton ,Sutton Cheney ,Fenny Drayton	Barlestone ,Earl Shilton ,Barwell ,Stapleton ,Sutton Cheney ,Fenny Drayton ,Witherley ,Ratcliffe Culey ,Sibson ,Twycross ,Nailstone ,Bagworth ,Stanton under Bardon ,Groby ,Botcheston
16	83.35	93.89	10.54	Nailstone ,Twycross ,Sibson ,Ratcliffe Culey ,Witherley ,Fenny Drayton ,Stoke Golding ,Sutton Cheney ,Stapleton ,Barwell ,Earl Shilton ,Botcheston ,Groby ,Stanton under Bardon ,Bagworth ,Barlestone	Bagworth ,Botcheston ,Groby ,Stanton under Bardon ,Nailstone ,Twycross ,Ratcliffe Culey ,Witherley ,Fenny Drayton ,Sibson ,Stoke Golding ,Earl Shilton ,Barwell ,Stapleton ,Sutton Cheney ,Barlestone
17	87.39	89.08	1.69	Barlestone ,Nailstone ,Bagworth ,Stanton under Bardon ,Groby ,Botcheston ,Earl Shilton ,Barwell ,Stapleton ,Stoke Golding ,Fenny Drayton ,Witherley ,Ratcliffe Culey ,Sheepy Magna ,Twycross ,Sibson ,Sutton Cheney	Barlestone ,Nailstone ,Bagworth ,Stanton under Bardon ,Groby ,Botcheston ,Earl Shilton ,Barwell ,Stapleton ,Stoke Golding ,Fenny Drayton ,Witherley ,Ratcliffe Culey ,Sheepy Magna ,Twycross ,Sibson ,Sutton Cheney

Subset of	Tour Length (O) KM	Tour Length (N) KM	Difference KM	Tour Path (O)	Tour Path (N)
18	97.89	98.18	0.29	Witherley ,Fenny Drayton ,Stoke Golding ,Burbage ,Earl Shilton ,Barwell ,Stapleton ,Sutton Cheney ,Botcheston ,Groby ,Stanton under Bardon ,Bagworth ,Barlestone ,Nailstone ,Twycross ,Sheepy Magna ,Sibson ,Ratcliffe Culey	Bagworth ,Stanton under Bardon ,Groby ,Botcheston ,Barlestone ,Sutton Cheney ,Stapleton ,Barwell ,Earl Shilton ,Burbage ,Stoke Golding ,Fenny Drayton ,Witherley ,Ratcliffe Culey ,Sibson ,Sheepy Magna ,Twycross ,Nailstone
19	98.23	98.42	0.18	Sutton Cheney ,Sibson ,Fenny Drayton ,Witherley ,Ratcliffe Culey ,Sheepy Magna ,Twycross ,Nailstone ,Barlestone ,Bagworth ,Stanton under Bardon ,Groby ,Newtown Unthank ,Botcheston ,Earl Shilton ,Barwell ,Stapleton ,Burbage ,Stoke Golding	Sheepy Magna ,Sibson ,Ratcliffe Culey ,Witherley ,Fenny Drayton ,Stoke Golding ,Burbage ,Earl Shilton ,Barwell ,Stapleton ,Sutton Cheney ,Botcheston ,Newtown Unthank ,Groby ,Stanton under Bardon ,Bagworth ,Barlestone ,Nailstone ,Twycross
20	101.52	102.09	0.57	Dadlington ,Burbage ,Barwell ,Earl Shilton ,Botcheston ,Newtown Unthank ,Groby ,Stanton under Bardon ,Bagworth ,Nailstone ,Barlestone ,Stapleton ,Sutton Cheney ,Sibson ,Twycross ,Sheepy Magna ,Ratcliffe Culey ,Witherley ,Fenny Drayton ,Stoke Golding	Fenny Drayton ,Sutton Cheney ,Dadlington ,Stoke Golding ,Burbage ,Earl Shilton ,Barwell ,Stapleton ,Barlestone ,Bagworth ,Botcheston ,Newtown Unthank ,Groby ,Stanton under Bardon ,Nailstone ,Twycross ,Sheepy Magna ,Sibson ,Ratcliffe Culey ,Witherley
21	101.36	102.60	1.23	Groby ,Markfield ,Stanton under Bardon ,Bagworth ,Nailstone ,Barlestone ,Sutton Cheney ,Twycross ,Sheepy Magna ,Sibson ,Ratcliffe Culey ,Witherley ,Fenny Drayton ,Stoke Golding ,Dadlington ,Burbage ,Stapleton ,Barwell ,Earl Shilton ,Botcheston ,Newtown Unthank	Stoke Golding ,Dadlington ,Burbage ,Earl Shilton ,Barwell ,Stapleton ,Botcheston ,Newtown Unthank ,Groby ,Markfield ,Stanton under Bardon ,Bagworth ,Nailstone ,Barlestone ,Sutton Cheney ,Sibson ,Twycross ,Sheepy Magna ,Ratcliffe Culey ,Witherley ,Fenny Drayton
22	98.93	103.25	4.32	Witherley ,Ratcliffe Culey ,Sheepy Magna ,Twycross ,Nailstone ,Barlestone ,Bagworth ,Stanton under Bardon ,Markfield ,Groby ,Ratby ,Newtown Unthank ,Botcheston ,Earl Shilton ,Barwell ,Stapleton ,Burbage ,Stoke Golding ,Dadlington ,Sutton Cheney ,Sibson ,Fenny Drayton	Sibson ,Sheepy Magna ,Twycross ,Nailstone ,Bagworth ,Stanton under Bardon ,Markfield ,Groby ,Ratby ,Newtown Unthank ,Botcheston ,Barlestone ,Stapleton ,Barwell ,Earl Shilton ,Burbage ,Stoke Golding ,Dadlington ,Sutton Cheney ,Fenny Drayton ,Witherley ,Ratcliffe Culey

Appendix (6): Description and provenance of R codes used in this Thesis

1. R script for generating the Voronoi tessellations for the settlements in the four data sources in the study area:

```
#=============================================================================================
#                          Code for generating Voronoi Tessellations
#
# Load the required libraries
#install.packages("dismo", dep = T)
library(dismo)
library(GISTools)
# read the mean centre points
Comp_mc <- readShapePoints("F:/Firdos_PhD/Entire_Data/MeanCenters/Comp_MC.shp")
BS76_mc <- readShapePoints("F:/Firdos_PhD/Entire_Data/MeanCenters/BS76_MC.shp")
POST_mc <- readShapePoints("F:/Firdos_PhD/Entire_Data/MeanCenters/POST_MC.shp")
POI_mc  <- readShapePoints("F:/Firdos_PhD/Entire_Data/MeanCenters/POI_MC.shp")
# read the polygon of H&B district
SExtent <- readShapePoly("F:/Firdos_PhD/Entire_Data/Hinckley&Bosworth_District.shp")

# generate the voronoi tessellation
Comp_vo <- voronoi(Comp_mc)
BS76_vo <- voronoi(BS76_mc)
POST_vo <- voronoi(POST_mc)
POI_vo  <- voronoi(POI_mc)

# write the results as  a shapefile
writePolyShape(Comp_vo, "F:/Firdos_PhD/Entire_Results/Voronoi/Comp_vo.shp")
writePolyShape(BS76_vo, "F:/Firdos_PhD/Entire_Results/Voronoi/BS76_vo.shp")
writePolyShape(POST_vo, "F:/Firdos_PhD/Entire_Results/Voronoi/POST_vo.shp")
writePolyShape(POI_vo, "F:/Firdos_PhD/Entire_Results/Voronoi/POI_vo.shp")

#=============================================================================================
```

2. R script for generating the normalised density surfaces and their α-cuts for settlements.

```r
#==================================================================================================
# Script Title:  Density Surfaces - generating normalised KDE, and their alpha-cuts
#
# Load the required libraries
library(GISTools)
library(spgwr)
library(rgdal)
library(raster)
library(shapefiles)
#------------------------------------------------------------ part one-----------------------------
# Set the working directory where the data exist
setwd("F:/Firdos_PhD/Entire_Data/")
# read the settlements' point data
data <- readShapePoints("All_HB_Data.shp")
# read the polygon of H&B district
lim.ex <- readShapePoly("Hinckley&Bosworth_District.shp")

# Check & investigate the data
plot(lim.ex)# H&B border
plot(data, pch = 1, add = T, cex = 0.5, col = "#FB6A4A4C")#plot the point data combined together
head(data.frame(data))
mean(table(data$Settlement, data$SourceType)[,1])
NAdd <- 10 # could be changed to 200 for test with fewer villages
index <- table(data$Settlement, data$SourceType) >= NAdd
summary(index)
which(index)
table(tolower(data$Settlement), data$SourceType)
colSums(index)
summary(rowSums(index[,c(1,2)]))

# based on investigation select 3 kinds of SourceType data
index <- data$SourceType == "BS76" | data$SourceType == "POST" | data$SourceType ==  "POI"
dat <- data[which(index), ]
dat$Settlement <- tolower(dat$Settlement)
# now identify villages with at least 10 points in any data type
tab <- table(dat$Settlement, dat$SourceType)
index <- tab >= NAdd
index <- index[,1] == "TRUE" | index[,2] == "TRUE" | index[,3] == "TRUE"
#index <-  index[,2] == "TRUE"
which(index)
# define lists for each data type and village
village.list <- names(which(index))
data.source.list <- colnames(tab[,1:3])

#------------------------------------------------------------ part two-----------------------------
##### Helper Functions
#`````````````````````
# normalising the data to stretch from 0--> 1
kde.norm.func <- function(kde){
  kde$kde <- data.frame(kde)[,1] / max(data.frame(kde)[,1])
  return(kde)
}

# KDE plot function
do.kde.plot <- function(kde, points) {
  shades = shading(breaks=c(0.2,0.4,0.6, 0.8),cols=brewer.pal(5,'YlOrRd'))
  level.plot(kde, shades = shades)
  plot(points, add = T)
  title(points)
}
```

```

#------------------------------------------------------------ part three---------------------------
##### Main Loops
#```````````````
# Set the working directory where the results to be saved
setwd("F:/Firdos_PhD/Entire_Results/Density&alpha")
### Loop 1 generates and plots KDEs
# creates a dat variable for each village & each data type
for (i in 1:length(village.list)) {
  vill = village.list[i]
  index <- dat$Settlement == vill
  dat.tmp <- dat[index,]

  # loop through the data type list and create a kde for each village for data type
  for (j in 1:length(data.source.list)) {
    index <- dat.tmp$SourceType == data.source.list[j]
    dat.j <- dat.tmp[index,]
    kde.j <- kde.norm.func(kde.points(dat.j, n = 500, lims = lim.ex))
    # assign the kde a unique name
    # test by running

    # sprintf("kde.%s.%s", vill, j)
    assign(sprintf("kde.%s.%s", vill, data.source.list[j]), kde.j )
    assign(sprintf("dat.%s.%s", vill, data.source.list[j]), dat.j)
    file.name = sprintf("KDE.%s.%s", vill, data.source.list[j])
    writeGDAL(kde.j, fname = file.name, drivername="GTiff")
  }
}

#------------------------------------------------------------ part four----------------------------
# Loop 2 generates alpha cuts and exports them
## this generates 100s of files so unceck the comment below by the writeGDAL
for (i in 1:length(village.list)) {
  vill = village.list[i]
  # loop through the data type list and create alpha cuts for each village for data type
  for (j in 1:length(data.source.list)) {
    kde.1.name <-sprintf("kde.%s.%s", vill, data.source.list[j])
    kde.1 <- get(kde.1.name)
    for (k in seq(0.1, 1, by = 0.1) ) {
      kde.copy <- kde.1
      index = kde.copy$kde >= k
      kde.copy$kde[index] = 1
      # set those for which alpha cut is false to 0
      kde.copy$kde[!index] = 0
      file.name = sprintf("%s.%s.alpha.%s.gtiff", vill, data.source.list[j], k)
      image(kde.copy)
      writeGDAL(kde.copy, fname = file.name, drivername="GTiff")

    }
  }
}
#==========================================================================================
```

3. R script to apply the linear regression model for settlements in each data types:

```r
#=================================================================================================
# Script Title:  linear regression model - Linear Regression plots 4 the data in the 3 sources
#
# Load the required libraries
library(GISTools)
library(spgwr)
library(rgdal)
library(raster)
library(shapefiles)
#------------------------------------------------------------ part one----------------------------
# Set the working directory where the data exist
setwd("F:/Firdos_PhD/Entire_Data/")
# read the settlements' point data
data <- readShapePoints("All_HB_Data.shp")
# read the polygon of H&B district
lim.ex <- readShapePoly("Hinckley&Bosworth_District.shp")

# Check & investigate the data
plot(lim.ex)# H&B border
plot(data, pch = 1, add = T, cex = 0.5, col = "#FB6A4A4C")#plot the point data combined together
head(data.frame(data))
mean(table(data$Settlement, data$SourceType)[,1])
NAdd <- 200 # could be changed to 200 for test with fewer villages
index <- table(data$Settlement, data$SourceType) >= NAdd
summary(index)
which(index)
table(tolower(data$Settlement), data$SourceType)
colSums(index)
summary(rowSums(index[,c(1,2)]))

# based on investigation select 3 kinds of SourceType data
index <- data$SourceType == "BS76" | data$SourceType == "POST" | data$SourceType ==  "POI"
dat <- data[which(index), ]
dat$Settlement <- tolower(dat$Settlement)
# now identify villages with at least 10 points in any data type
tab <- table(dat$Settlement, dat$SourceType)
index <- tab >= NAdd
index <- index[,1] == "TRUE" & index[,2] == "TRUE" & index[,3] == "TRUE"
which(index)

# define lists for each data type and village
village.list <- names(which(index))
data.source.list <- colnames(tab[,1:3])

#------------------------------------------------------------ part two----------------------------
##### Helper Functions
#``````````````````````
# normalising the data to stretch from 0--> 1
kde.norm.func <- function(kde){
    kde$kde <- data.frame(kde)[,1] / max(data.frame(kde)[,1])
    return(kde)
}

# KDE plot function
do.kde.plot <- function(kde, points) {
    shades = shading(breaks=c(0.2,0.4,0.6, 0.8),cols=brewer.pal(5,'YlOrRd'))
    level.plot(kde, shades = shades)
    plot(points, add = T)
    title(points)
}
# Plot Points function
do.regression.comparison <- function(kde.1, kde.2, kde.1.name, kde.2.name) {
    index = kde.1$kde == 0 & kde.2$kde == 0
    pt1 <- kde.1$kde[!index]
    pt2 <- kde.2$kde[!index]
    #regression
    mod.1 <- lm(pt2~pt1)
    plot(pt1, pt2, col = "#52525280", xlab = kde.1.name, ylab = kde.2.name)
    abline(mod.1, col=2, lwd=1.5)
    tit <- sprintf("%s vs %s", kde.2.name, kde.1.name)
    title(tit)
    return(mod.1)
}
```

```
#------------------------------------------------------------- part three--------------------------
##### Main Loops
#```````````````
# Set the working directory where the results to be saved
#setwd("F:/Firdos_PhD/Entire_Results/Regression_Plots")
setwd("F:/Firdos_PhD/Entire_Results/Test/GWR")
### Loop 1 generates and plots KDEs
# creates a dat variable for each village & each data type
for (i in 1:length(village.list)) {
    vill = village.list[i]
    index <- dat$Settlement == vill
    dat.tmp <- dat[index,]

    # loop through the data type list and create a kde for each village for data type
    for (j in 1:length(data.source.list)) {
        index <- dat.tmp$SourceType == data.source.list[j]
        dat.j <- dat.tmp[index,]
        kde.j <- kde.norm.func(kde.points(dat.j, n = 500, lims = lim.ex))
    # assign the kde a unique name
    # test by running

    # sprintf("kde.%s.%s", vill, j)
    assign(sprintf("kde.%s.%s", vill, data.source.list[j]), kde.j )
    assign(sprintf("dat.%s.%s", vill, data.source.list[j]), dat.j)
    file.name = sprintf("KDE.%s.gtiff", vill, data.source.list[j], kde.j)
    writeGDAL(kde.j, fname = file.name, drivername="GTiff")
    }

#### Plot to PDF
    # now plot the data and the surfaces
    name.tmp <-sprintf("plots.kde.%s.pdf", vill)
    name.tmp <- gsub(" ", "", name.tmp)
    # may need to change height and width with more data types
    pdf(file = name.tmp, width= 12,height= 8)
    #windows(h = 8, w = 16)
    par(mfrow = c(2,length(data.source.list))) # how many rows and columns in the plot
    par(mar = c(1,1,2,1))

# plot kdes
    for (j in 1:length(data.source.list)) {
        name.tmp <-sprintf("kde.%s.%s", vill,data.source.list[j])
        kde.tmp <- get(name.tmp)
        # a white plot to set the plot area - so it is the same
        shades = shading(breaks=c(0.2,0.4,0.6, 0.8),cols=brewer.pal(5,'YlOrRd'))
        image(kde.tmp, col = "white")
        level.plot(kde.tmp, shades = shades, add = T)
        plot(lim.ex, add = T)
        choro.legend(430159, 295016, shades, cex = 0.9)
        name.tmp <-sprintf("kde.%s.%s", vill,data.source.list[j])
        title(name.tmp)
    }

    # Plot points
    for (j in 1:length(data.source.list)) {
        name.tmp <-sprintf("dat.%s.%s", vill, data.source.list[j])
        dat.tmp <- get(name.tmp)
        # a white plot to set the plot area - so it is the same
        image(kde.tmp, col = "white")
        plot(dat.tmp, pch = i, add = T,col = "#96969680" )
        plot(lim.ex, add = T)
        name.tmp <-sprintf("dat.%s.%s", vill, data.source.list[j])
        title(name.tmp)
    }
    dev.off()    # close PDF
}
```

```
#------------------------------------------------------------- part four---------------------------
# Loop 2 generates regressions for each village and each data type
res.tab <- matrix(data=0, ncol=(length(data.source.list)),nrow=(length(village.list)))
for (i in 1:length(village.list)) {
    vill = village.list[i]

    #### Do comparitive analysis
    # 1. scatterplots
    # 2. regression
    # set the plot parameters
    # index to identify cells that are both zero
    # exclude these from the scatter plots and the regressions

    name.tmp <-sprintf("plots.regression.%s.pdf", vill)

    # may need to change height and width with more data types
    pdf(file = name.tmp, width= 16,height= 4)
    #windows(h = 8, w = 16)
    par(mfrow = c(1, length(data.source.list) ))

    for (j in 1: (length(data.source.list)) ) {
        kde.1.name <-sprintf("kde.%s.%s", vill, data.source.list[j])
        kde.1 <- get(kde.1.name)
        cat(kde.1.name)
        if(j != length(data.source.list)) {
            kde.2.name <-sprintf("kde.%s.%s", vill, data.source.list[j+1])
            kde.2 <- get(kde.2.name)
            cat(kde.2.name)
            # the following if statement just to keep the POI in the y-axis
            # so check and replace
            if(data.source.list[j+1]!='POI'){
                kde.poi.tem <- kde.1.name
                kde.1.name  <- kde.2.name
                kde.2.name  <- kde.poi.tem
                kde.1 <- get(kde.1.name)
                cat(kde.1.name)
                kde.2 <- get(kde.2.name)
                cat(kde.2.name)
            }
            regression.mod <- do.regression.comparison(kde.1, kde.2, kde.1.name, kde.2.name)
            assign(sprintf("mod.%s.vs.%s", kde.1.name, kde.2.name), regression.mod)
        } else{
            kde.2.name <-sprintf("kde.%s.%s", vill, data.source.list[j-2])
            kde.2 <- get(kde.2.name)
            cat(kde.2.name)
            regression.mod <- do.regression.comparison(kde.1, kde.2, kde.1.name, kde.2.name)
            assign(sprintf("mod.%s.vs.%s", kde.1.name, kde.2.name), regression.mod)
        }
        r.sq <- summary(regression.mod)$r.squared
        res.tab[i,j] <- r.sq
    }
    dev.off()
}
rownames(res.tab) <- village.list
write.csv(res.tab, "MyOutFile.csv")
#===========================================================================================
```

4. R script applying points in polygon analyses and tables for inclusion:

```r
#==========================================================================================================
# Script Title:  Applying Points in Polygon Analyses
# apply pointNpolygon counts 4 all data and generate tables
# ________________________________________________________________________________________________________
# Load the required libraries
library(GISTools)
library(spgwr)
library(rgdal)
library(raster)
library(shapefiles)

# Set the working directory where the data exist
setwd("F:/Firdos_PhD/Entire_Data/")
# read the settlements' point data
data <- readShapePoints("All_HB_Data.shp")
# read the polygon of H&B district
lim.ex <- readShapePoly("Hinckley&Bosworth_District.shp")
# parish <- readShapePoly("H&B_Parishes.shp")
# Check & investigate the data
# plot(lim.ex)# H&B border
# plot(parish, add= T)
# plot(data, pch = 1, add = T, cex = 0.5, col = "#FB6A4A4C")#plot the point data combined together
head(data.frame(data))
mean(table(data$Settlement, data$SourceType)[,1])
NAdd <- 10
index <- table(data$Settlement, data$SourceType) >= NAdd
summary(index)
which(index)
table(tolower(data$Settlement), data$SourceType)
colSums(index)
summary(rowSums(index[,c(1,2)]))

# based on investigation select 3 kinds of SourceType data
index <- data$SourceType == "BS76" | data$SourceType == "POST" | data$SourceType ==  "POI"
#index <- data$SourceType =="BS76"
dat <- data[which(index), ]
dat$Settlement <- tolower(dat$Settlement)
# now identify villages with at least 10 points in any data type// NAdd= 100
tab <- table(dat$Settlement, dat$SourceType)
```

```r
tab <- table(dat$Settlement, dat$SourceType)
index <- tab >= NAdd
index <- index[,1] == "TRUE" | index[,2] == "TRUE" | index[,3] == "TRUE" # FA: may need to change & to |
#index <-  index[,1] == "TRUE"
which(index)
# define lists for each data type and village
village.list <- names(which(index))
data.source.list <- colnames(tab[,1:3])

#---------------------------------------------------------- part two------------
##### Helper Functions
#``````````````````````
# normalising the data to strecth from 0--> 1
kde.norm.func <- function(kde){
    kde$kde <- data.frame(kde)[,1] / max(data.frame(kde)[,1])
    return(kde)
}
# make a local polygon around the study area
make.local.poly <- function(dat.tmp, fac) {
    x.min <- bbox(dat.tmp)[1,1] - fac
    x.max <- bbox(dat.tmp)[1,2] + fac
    y.min <- bbox(dat.tmp)[2,1] - fac
    y.max <- bbox(dat.tmp)[2,2] + fac
    p <- Polygon(cbind(c(x.min, x.min, x.max, x.max, x.min), c(y.min, y.max, y.max, y.min, y.min)))
    p <- Polygons(list(p), "S1")
    p <- SpatialPolygons(list(p))
    return(p)
}
# KDE plot function
do.kde.plot <- function(kde, points) {
    shades = shading(breaks=c(0.2,0.4,0.6, 0.8),cols=brewer.pal(5,'YlOrRd'))
    level.plot(kde, shades = shades)
    plot(points, add = T)
    title(points)
}
# Plot Points function
do.regression.comparison <- function(kde.1, kde.2, kde.1.name, kde.2.name) {
    index = kde.1$kde == 0 & kde.2$kde == 0
```

```r
    pt1 <- kde.1$kde[!index]
    pt2 <- kde.2$kde[!index]
    #regression
    mod.1 <- lm(pt2~pt1)
    plot(pt1, pt2, col = "#52525280", xlab = kde.1.name, ylab = kde.2.name)
    abline(mod.1)
    tit <- sprintf("%s vs %s", kde.1.name, kde.2.name)
    title(tit)
    return(mod.1)
}

# Density stuff

den.val <- function(thresh.j, kde.1, alpha.j, dat.tmp){
  kde.copy <- kde.1
  alpha.j <- thresh.j
  index = kde.copy$kde >= alpha.j
  kde.copy$kde[index] = 1
  kde.copy$kde[!index] = 0
    r = raster(nrow = 500, ncol = 500, ext = extent(kde.copy))
    r <- rasterize(kde.copy, r, "kde")
    pol <- rasterToPolygons(r, fun=function(x){x>0}, dissolve = T)
    n.in.poly <- poly.counts(dat.tmp, pol)
  polArea <- poly.areas(pol)/1000000 # to get the area in km2
    d.1 <- n.in.poly/polArea
  return(d.1)
}

#------------------------------------------------------------ part three------------
##### Main Loops
#```````````````
# Set the working directory where the results to be saved
setwd("F:/Firdos_PhD/Entire_Results/PointsNPolygon/AllD_results")
### Loop 1 generates and plots KDEs
# creates a dat variable for each village & each data type
for (i in 1:length(village.list)) {
    vill = village.list[i]
    index <- dat$Settlement == vill
    dat.tmp <- dat[index,]

    #dat.i <- dat.tmp[index,]
    kde.i <- kde.norm.func(kde.points(dat.tmp, n = 500, lims = lim.ex))
    # assign the kde a unique name
    # test by running

    # sprintf("kde.%s.%s", vill, i)
    assign(sprintf("kde.%s.%s", vill, i), kde.i )
    assign(sprintf("dat.%s.%s", vill, i), dat.tmp)


 #### Plot to PDF
    # now plot the data and the surfaces
    name.tmp <-sprintf("plots.kde.%s.pdf", vill)
    name.tmp <- gsub(" ", "", name.tmp)
    # may need to change height and width with more data types
    pdf(file = name.tmp, width= 12,height= 8)
    #windows(h = 8, w = 16)
    #par(mfrow = c(2,length(data.source.list))) # how many rows and columns in the plot
    par(mar = c(1,1,2,1))

 # plot kdes
    name.tmp <-sprintf("kde.%s.%s", vill,i)
    kde.tmp <- get(name.tmp)
    # a white plot to set the plot area - so it is the same
    shades = shading(breaks=c(0.2,0.4,0.6, 0.8),cols=brewer.pal(5,'YlOrRd'))
    image(kde.tmp, col = "white")
    level.plot(kde.tmp, shades = shades, add = T)
    plot(lim.ex, add = T)
    choro.legend(430159, 295016, shades, cex = 0.9)
    name.tmp <-sprintf("kde.%s.%s", vill,i)
    title(name.tmp)

    # Plot points
    #for (j in 1:length(data.source.list)) {
        name.tmp <-sprintf("dat.%s.%s", vill,i)
        dat.tmp <- get(name.tmp)
        # a white plot to set the plot area - so it is the same
        image(kde.tmp, col = "white")
        plot(dat.tmp, pch = i, add = T,col = "#96969680" )
        plot(lim.ex, add = T)
        name.tmp <-sprintf("dat.%s.%s", vill,i)
        title(name.tmp)

    dev.off()   # close PDF
}
#---------------------------------------------------------- part four------------
Ndt <- length(unique(data$SourceType))
#Nvil <- length(unique(data$Settlement)) # change this when you want to run through the 76 villages
Nvil <- length(village.list)
my.alpha <- seq(0.05, 0.95, by = 0.05)
my.thresh <- seq(0.5, 0.95, by = 0.05)
```

```
res.tab <- matrix(data =10, ncol = (length(my.thresh) *3)+4, nrow = 0)#for point in polygon counts
colnames(res.tab) <- c("N", "Den 0.1", "Den 0.3", "Den 0.5", "T50", "T50D", "T50A",
  "T55", "T55D", "T55A",
  "T60", "T60D", "T60A",
  "T65", "T65D", "T65A",
  "T70", "T70D", "T70A",
  "T75", "T75D", "T75A",
  "T80", "T80D", "T80A",
  "T85", "T85D", "T85A",
  "T90", "T90D", "T90A",
  "T95", "T95D", "T95A")

# Loop 3 - Generate the alpha cuts and apply the Point in Polygons
for (i in 1:Nvil) {
    vill = village.list[i]
    index <- dat$Settlement == vill
  n.points <- sum(index)
    dat.tmp <- dat[index,]
  kde.1.name <-sprintf("kde.%s.%s", vill, i)
    kde.1 <- get (kde.1.name)
  # what I want to do now is to do an alpha cut until the T is met
  temp.res.line <- vector()
  for(j in 1: length(my.thresh)) {
    thresh.j <- my.thresh[j]
    cond.not.met = TRUE
    k = 1
    while(cond.not.met == TRUE) {
      kde.copy <- kde.1
      alpha.j <- my.alpha[k]
      #kde.copy <- kde.1
            index = kde.copy$kde >= alpha.j
            kde.copy$kde[index] = 1
            # set those for which alpha cut is false to 0
            kde.copy$kde[!index] = 0
            r = raster(nrow = 500, ncol = 500, ext = extent(kde.copy))
            # £££ I add this to write out the alpha cut START HERE
            file.name = sprintf("%s.%s.allD.alpha.%s.gtiff", vill, i, alpha.j)
            writeGDAL(kde.copy, fname = file.name, drivername="GTiff")
            # ££££ my edit ends here!!!
            r <- rasterize(kde.copy, r, "kde")
            pol <- rasterToPolygons(r, fun=function(x){x>0}, dissolve = T)

            ### Do point in poly to determine how many points in polygon / alpha
            n.in.poly <- poly.counts(dat.tmp, pol)
            ### To calculate the proportion %%, get the total no. of points first
            total.point <- length(dat.tmp)
            p.in.poly  <- (n.in.poly/total.point)
      ### now do test for Threshold
      # To Calculate density of points:
        # (1) calculate the area of the polygon
            # (2) then divide the total no. of point by the area
            polArea <- poly.areas(pol)/1000000 # to get the area in km2
            Density.d <- n.in.poly/polArea
      if( p.in.poly < thresh.j) {
        if(exists("old.alpha.j")) {
          temp.res.part.line <- data.frame(old.alpha.j, old.Density.d, old.polArea)
          cond.not.met <- FALSE
          break
        }
        if(!exists("old.alpha.j")) {
          temp.res.part.line <- data.frame(c(0,0,0))
          cond.not.met <- FALSE
          break
        }

      }

      cat(alpha.j, "\n")
      old.alpha.j <- alpha.j
      old.Density.d <- Density.d
      old.polArea <- polArea
      k = k + 1
    }
    temp.res.line <- append(temp.res.line, temp.res.part.line)
    rm(list = c("old.alpha.j", "old.Density.d", "old.polArea", "alpha.j", "Density.d", "polArea"))

    #length(unlist(temp.res.line))


    }
  d.1 <- den.val(0.1, kde.1, 0.7, dat.tmp)
  d.3 <- den.val(0.3, kde.1, 0.7, dat.tmp)
  d.5 <- den.val(0.7, kde.1, 0.7, dat.tmp)

  vill.line.start <- data.frame(n.points, d.1, d.3, d.5)
  temp.res.line2 <- unlist(append(vill.line.start, temp.res.line))
  names(temp.res.line2) <- colnames(res.tab)
  res.tab <- rbind(res.tab, unlist(temp.res.line2))


}
rownames(res.tab) <- village.list
write.csv(res.tab, "MyOutFile_4all.csv")

#End here!!
#===========================================================================================
```

5. R script for centroid extraction from the fuzzy extents of the villages:

```
#==========================================================================================
# Script Title:  Centroid extraction from alpha cuts
#
# Load the required libraries
library(GISTools)
library(spgwr)
library(rgdal)
library(raster)
library(shapefiles)
#------------------------------------------------------------ part one--------------------------
# Set the working directory where the data exist
#setwd("F:/Firdos_PhD/Entire_Data/")
#setwd("F:/Firdos_PhD/Entire_Results/KDE&alpha")
setwd("F:/Firdos_PhD/Entire_Results/Density&alpha")
## 1. Make a list of things to read in
## in DOS window create a list of files and edit
##> dir *POI*.* /b > my.POI.list
my.list <- read.delim("my.POI.list", sep = "\n")

listchar = ".POI"

calc.rast.cents <- function(my.list, listchar) {
  res.vec.x <- vector()
  res.vec.y <- vector()
  names.vec <- character()
  alpha.vec <- character()
  for (i in 1:dim(my.list)[1]) {
    f.name <- as.character(my.list[i,])
    r <- readGDAL(f.name)
    r <- raster(r)
    p <- rasterToPoints(r, dissolve = F)
    index <- p[,3] == 1
    p <- p[index,]
    if( length(p) == 3) {
      res.vec.x <- append(res.vec.x, p[1])
      res.vec.y <- append(res.vec.y, p[2])
    }
    if( length(p) != 3) {
      b.box <- bbox(p)
      X <- ((b.box[1,2] - b.box[1,1]) / 2) + b.box[1,1]
      res.vec.x <- append(res.vec.x, X)
      Y <- ((b.box[2,2] - b.box[2,1]) / 2) + b.box[2,1]
      res.vec.y <- append(res.vec.y, Y)
    }
    f.name1 <- strsplit(f.name, listchar)[[1]][1]
    names.vec <- append(names.vec, f.name1)
    f.name2 <- strsplit(f.name, "alpha.")[[1]][2]
    f.name2 <- gsub(".gtiff", "", f.name2)
    alpha.vec <- append(alpha.vec, f.name2)
  }
  res.out <- data.frame(res.vec.x, res.vec.y, names.vec, alpha.vec)
  names(res.out) <- c("X", "Y", "Village", "Alpha")
  res.out <- SpatialPointsDataFrame(res.out[,1:2], data = data.frame(res.out[,3:4]))
  return(res.out)
}

t.1 <- calc.rast.cents(my.list, listchar)
plot(t.1)

write.csv(t.1, "F:/Firdos_PhD/Entire_Data/4TSP/Fuzzy_POI_centres.csv")
writePointsShape(t.1, "F:/Firdos_PhD/Entire_Results/Density&alpha/list.centre.shp")

write.shapefile(t.1, "F:/Firdos_PhD/Entire_Results/Density&alpha/list.centre.shp")
```

6. R script for applying the travelling salesman problem on the hard extent versus the fuzzy extents of villages:

```
#==================================================================================================
# Script Title:  Applying the Travelling Salesman Problem (TSP)
#
# Install & Load the required libraries
#install.packages("TSP")
library(GISTools)
library(shapefiles)
library(TSP)

#------------------------------------------------------------ part one-----------------------------
##### Set up the OD cost matrix for the hard and fuzzy centres
### Set the working directory to read the files from
setwd("F:/Firdos_PhD/Entire_Data/4TSP")
### 1. Original, Boolean or Hard centres
### read in data from ARCGIS
data.tmp = read.csv("Hard_OD_Matrix.csv")
# I have edited the file by renaming the column name from (Destination) to (DestID).
## create distance matrix
attach(data.tmp)
d.mat.o = tapply(Total_Leng,list(OriginID, DestID),function (x) x)
detach(data.tmp)
## test it!
isSymmetric(d.mat.o)

### 2. New, soft or Fuzzy centres
### read in data from ARCGIS
data.tmp = read.csv("Fuzzy_OD_Matrix.csv")
# I have edited the file by renaming the column name from (Destination) to (DestID).
## create distance matrix
attach(data.tmp)
d.mat.n = tapply(Total_Leng,list(OriginID, DestID),function (x) x)
detach(data.tmp)
## test it!
isSymmetric(d.mat.n)
# read in other data
pt.o <- readShapePoints("Hard_Centres.shp")
pt.n <- readShapePoints("Fuzzy_Centres.shp")
parish <- readShapePoly("F:/Firdos_PhD/Entire_Data/H&B_Parishes.shp")

#------------------------------------------------------------ part two-----------------------------
### 3.  Analysis with all
# other way to compare between the methods of TSP
## Set the working directory to save the result in
setwd("F:/Firdos_PhD/Entire_Results/TSP/NewRes")
# Original / Hard
tour.o.dist <- 0
dist.tsp <- TSP(d.mat.o)
### Methods used to solve TSP - algorithms
methods <- c("nearest_insertion", "farthest_insertion","cheapest_insertion",
             "arbitrary_insertion", "nn", "repetitive_nn", "2-opt")
tours.o <- lapply(methods, FUN = function(m) solve_TSP(dist.tsp, method = m))
names(tours.o) <- methods
tours.o#[[1]]
# 7 in the loop refers to number of arguments methods - algorithms in TSP
for(i in 1:7){
  tour.o.dist[i] <- attr(tours.o[[i]], "tour_length")
}

# New / Fuzzy
tour.n.dist<- 0
dist.tsp <- TSP(d.mat.n)
tours.n <- lapply(methods, FUN = function(m) solve_TSP(dist.tsp, method = m))
names(tours.n) <- methods
tours.n#[[1]]
for(i in 1:7){
  tour.n.dist[i] <- attr(tours.n[[i]], "tour_length")
}

## plot map
name.tmp <- "Plots.comparing.some.TSP.METHODS.pdf"
name.tmp <- gsub(" ", "", name.tmp)
pdf(file = name.tmp, width= 8,height= 10)
par(mfrow = c(2,1))
dotchart(c(sapply(tours.o, FUN = attr, "tour_length")), xlab = "tour length",
         xlim = c(0, 150000))
title(sprintf("TSP for Original Centres"))
dotchart(c(sapply(tours.n, FUN = attr, "tour_length")), xlab = "tour length",
```

```
            xlim = c(0, 150000))
title(sprintf("TSP for New Centres"))
dev.off()
## save tour length in the two cases
res.tab.1 <- matrix(data=0, ncol = 7, nrow =2)
tsp.tmp <- rbind(tour.o.dist, tour.n.dist)
res.tab.1 <- tsp.tmp
colnames(res.tab.1) <- c("nearest_insertion", "farthest_insertion","cheapest_insertion",
                         "arbitrary_insertion", "nn", "repetitive_nn", "2-opt")
write.csv(res.tab.1, file = "TSP.out.res1.csv")


#------------------------------------------------------------ part three---------------------------
### 4. Subset based on largest distances between new and old centres

n = 22 # this is the number of maximum subset(total no. of villages -1)
## calculate distance between hard and fuzzy centres
dists.n.o <-  diag(spDists(pt.o, pt.n))
res.tab <- matrix(data=0, ncol=3, nrow=length(pt.o))
colnames(res.tab) <- c("Old_Centre","New_Centre","Distance")

res.tab[,1] <- pt.o$Name
res.tab[,2] <- pt.n$Village
res.tab[,3] <- dists.n.o
write.csv(res.tab, file = "TSP.dists.n.o.csv")

## find the largest distances
ord.index <- order(dists.n.o, decreasing = T)

## subset the d.mat for the n best differences
res.tab.2 <- matrix(data=0, ncol = 3, nrow =n)
colnames(res.tab.2) <- c("tour_length.o", "tour_length.n","diff.n.o")
my.tsp.function <- function(n, ord.index, d.mat.n, d.mat.o ) {

  index <- ord.index[1:n]
  d.mat.subset.o <- d.mat.o[index, index]
  d.mat.subset.n <- d.mat.n[index, index]
  # Original
  dist.tsp <- TSP(d.mat.subset.o)
  tour.o <- solve_TSP(dist.tsp, method = "2-opt")
  print(tour.o)
  tour.o.dist <- attr(tour.o, "tour_length")
  tour.o.path <- labels(tour.o)

  # New
  dist.tsp <- TSP(d.mat.subset.n)
  tour.n <- solve_TSP(dist.tsp, method = "2-opt")
  print(tour.n)
  tour.n.dist <- attr(tour.n, "tour_length")
  tour.n.path <- labels(tour.n)

  ## plot map
  name.tmp <- sprintf("Plots.of.TSP.results.for.subset.of.%s.pdf",n)
  name.tmp <- gsub(" ", "", name.tmp)
  pdf(file = name.tmp, width= 8,height= 10)
  par(mfrow = c(2,1))
  plot(as(pt.o, "Spatial"), axes=TRUE, ylim = c(289000, 314000))
  plot(parish, add=TRUE, col = "#96969680")# neeed resizing
  ## plot tour and add cities
  tour_line <- SpatialLines(list(Lines(list(
    Line(pt.o[c(tour.o, tour.o[1]),])), ID="1")))
  plot(tour_line, add=TRUE, col = "red")
  points(pt.o, pch=16, cex=0.5, col="black")
  title(sprintf("TSP for Original Centres"))
  ##  par( "usr" ) returns a vector containing xleft=1, xright=2, ybottom=3, ytop=4.
  usr <- par( "usr" )
  text( usr[ 1 ]+1000, usr[ 4 ]-500,list(tour.o.path), side=2, adj = c( 0, 1 ),cex=.75,col="red")

  plot(as(pt.n, "Spatial"), axes=TRUE, ylim = c(289000, 314000))
  plot(parish, add=TRUE, col = "#96969680")# neeed resizing
  ## plot tour and add cities
  tour_line <- SpatialLines(list(Lines(list(
    Line(pt.n[c(tour.n, tour.n[1]),])), ID="1")))
  plot(tour_line, add=TRUE, col = "blue")
  points(pt.n, pch=16, cex=0.5, col="black")
  title(sprintf("TSP for New Centres"))
  # usr <- par( "usr" )
  text( usr[ 1 ]+1000, usr[ 4 ]-500,list(tour.n.path), side=2, adj = c( 0, 1 ),cex=.75,col="blue")
  dev.off()

 ## returning the tour length as results
 tsp.res <- (c(tour.o.dist, tour.n.dist))
 return(tsp.res)

  dev.off()  # close PDF

}

# loop in to
for (i in 3: n) {
  tsp.tmp <- my.tsp.function(i, ord.index, d.mat.n, d.mat.o )

  res.tab.2[i,1] <- tsp.tmp[1]
  res.tab.2[i,2] <- tsp.tmp[2]
  res.tab.2[i,3] <- tsp.tmp[2] - tsp.tmp[1]

}

write.csv(res.tab.2, file = "TSP.out.res2.csv")
#------------------------------------------------------------ ENDS HERE----------------------------
```

Bibliography

Agler, D.W. (2010) *Vagueness and Its Boundaries: A Peircean Theory of Vagueness.* Master of Arts in the Department of Philosophy, Indiana University.

Ahlqvist, O., Keukelaar, J. and Oukbir, K. (2003) 'Rough and Fuzzy Geographical Data Integration', *International Journal of Geographical Information Science,* 17(3), pp. 223-234.

Almadani, F. (2014) 'Modelling the Fuzzy Footprints of Villages from Postal Address Records', *The 8th International Conference on Geographic Information Science (GIScience)*, 23-26 SeptemberVienna.

Almadani, F., Fisher, P. & Jarvis, C. (2012) 'On the Fuzzy Distance Between Fuzzy Geographical Objects', Rowlingson,B. & Whyatt,D. (eds.) *The 20th Annual GIS Research UK (GISRUK) Conference*, 11-13 April.

Altman, D. (1994) 'Fuzzy Set Theoretic Approaches for Handling Imprecision in Spatial Analysis', *International Journal of Geographical Information Systems,* 8(3), pp. 271-289.

Applegate, D., Bixby, R., Chvatal, V. and Cook, W. (eds.) (2006) *Concorde TSP Solver.* htpp://www.tsp.gatech.edu/concorde/.

Arampatzis, A., Van Kreveld, M., Reinbacher, I., Jones, C.B., Vaid, S., Clough, P., Joho, H. and Sanderson, M. (2006) 'Web-based Delineation of Imprecise Regions', *Computers, Environment and Urban Systems,* 30(4), pp. 436-459.

Arnot, C., Fisher, P.F., Wadsworth, R. and Wellens, J. (2004) 'Landscape Metrics with Ecotones: Pattern Under Uncertainty', *Landscape Ecology,* 19(2), pp. 181-195.

Awuni, E. 2014 *Design and analysis of algorithm Documentation of Dynamic programming (brute force) for TSP for O-Mopsi Project IMPIT 2012* , University of Eastern Finland Faculty of Science and Forestry, FINLAND.

Batty, M., Crooks, A., Hudson-Smith, A., Milton, R., Anand, S., Jackson, M. and Morley, J.2010'Data Mash-ups and the Future of Mapping', JISC, Bristol.

Bezdek, J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms.* New York: Plenum Press.

Bibby, P. and Shepherd, J. (2004) 'Developing a New Classification of Urban and Rural Areas for Policy Purposes – the Methodology', *London: DEFRA,* .

Bielefeld, R.A. (1992) *Fuzzy Representation of Uncertainty in Disease Progression, PhD Thesis.* Case Western Reserve University (Health Science).

Bloch, I. (1999) 'On Fuzzy Distances and Their Use in Image Processing Under Imprecision', *Pattern Recognition Letters,* 32(11), pp. 1873-1895.

Bordogna, G., Chiesa, S. and Geneletti, D. (2006) 'Linguistic Modelling of Imperfect Spatial Information As A Basis for Simplifying Spatial Analysis', *Information Sciences,* 176(4), pp. 366-389.

Botzheim, J., Földesi, P. & Kóczy, L.T. (2009) 'Solution for Fuzzy Road Transport Traveling Salesman Problem Using Eugenic Bacterial Memetic Algorithm', Carvalho,J.P., Dubois,D., Kaymak,U., et al (eds.) *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009 Proceedings)*, 20-24 JulyLisbon, Portugal, pp. 1667.

Bowman, A.W. and Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis*
The Kernel Approach with S-Plus Illustrations. Oxford: Clarendon Press.

Brunsdon, C. (1995) 'Estimating Probability Surfaces for Geographical Point Data An Adaptive Kernel Algorithm', *Computers & Geosciences,* 21(7), pp. 877-894.

Brunsdon, C. and Chen, H. (eds.) (2014) *GISTools: Some further GIS capabilities for R.* 0.7-4 edn.

Burgoine, T. and Harrison, F. (2013) 'Comparing the Accuracy of Two Secondary Food Environment Data Sources in the UK Across Socio-economic and Urban/Rural Divides', *International Journal of Health Geographics,* 12(2).

Burrough, P.A. and McDonnell, R.A. (1998) *Principles of Geographical Information Systems.* New York: Oxford University Press.

Caldwell, C. (1995) *Graph Theory Glossary.* Available at: http://primes.utm.edu/graph/glossary.html.

Castleden, H., Crooks, V.A., Schuurman, N. and Hanlon, N. (2010) '"It's not necessarily the distance on the map…": Using place as an analytic tool to elucidate geographic issues central to rural palliative care', *Health & place,* 16(2), pp. 284-290.

Chambers, J.M. (1992) 'Linear Models', in Chambers, J.M. and Hastie, T.J. (eds.) *Chapter 4 of Statistical Models in S.* Wadsworth & Brooks/Cole.

Chandrasekaran, S., Kokila, G. and Saju, J. (2013) 'A New Approach to Solve Fuzzy Travelling Salesman Problems by using Ranking Functions', *International Journal of Science and Research (IJSR),* 4(5), pp. 2258-2260.

Chaudhry, O. (2008) *Modelling Geographic Phenomena at Multiple Levels of Detail: A Model Generalisation Approach based on Aggregation.* The University of Edinburgh.

Chaudhry, O. and Mackaness, W. (2008) 'Automatic Identification of Urban Settlement Boundaries for Multiple Representation Databases', *Computers, Environment and Urban Systems,* 32(2), pp. 95-109.

Chaudhry, O. and Mackaness, W. (2008) 'Creating Mountains out of Mole Hills: Automatic Identification of Hills and Ranges Using Morphometric Analysis', *Transactions in GIS,* 12(5), pp. 567-589.

Cheng, T. (2002) 'Fuzzy Objects: Their Changes and Uncertainties', *Photogrammetric Engineering & Remote Sensing,* 68(1), pp. 41-49.

Cheng, T., Fisher, P. and Zhilin, L. (2004) 'Double Vagueness: Uncertainty in Multi-scale Fuzzy Assignment of Duneness', *Geo-spatial Information Science (Quarterly),* 7(1), pp. 58-66.

Clementini, E. and di Felice, P. (1996) 'An Algebraic Model for Spatial Objects with Indeterminate Boundaries', in Burrough, P.A. and Frank, A.U. (eds.) *Geographic Objects with Indeterminate Boundaries.* Number 2 in GISDATA edn. London: Taylor & Francis, pp. 155-169.

Cloke, P. (2006) 'Conceptualizing Rurality', in Anonymous *Handbook of Rural Studies.* SAGE, pp. 18-29.

Cohn, A.G. and Gotts, N.M. (1996) 'The ‘egg-yolk’ Representation of Regions with Indeterminate Boundaries', in Burrough, P.A. and Frank, A.U. (eds.) *Geographic Objects with Indeterminate Boundaries.* Number 2 in GISDATA edn. London: Taylor & Francis, pp. 171-187.

Comber , A.J., Brunsdon, C. and Green, E. (2008) 'Using a GIS-based Network Analysis to Determine Urban Greenspace Accessibility for Different Ethnic and Religious Groups', *Landscape and Urban Planning,* 86(1), pp. 103-114.

Comber, A., Fisher, P., Brunsdon, C. and Khmag, A. (2012 a) 'Spatial Analysis of Remote Sensing Image Classification Accuracy', *Remote Sensing of Environment,* 127, pp. 237-246.

Comber, A., Fisher, P. and Wadsworth, R. (2005) 'What is land cover?', *Environment and Planning B: Planning and Design,* 32(2), pp. 199-209.

Comber, A.J., Wadsworth, R.A. and Fisher, P.F. (2008) 'Using Semantics to Clarify the Conceptual Confusion between Land Cover and Land Use: the Example of 'Forest'', *Journal of Land Use Science,* 3(2-3), pp. 185-198.

Comber, A., Fisher, P., Brunsdon, C. & Khmag, A. (2012 b) 'Geographically Weighted Methods for Examining the Spatial Variation in Land Cover Accuracy', *Proceeding of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 10- 13 JulyFlorianopolis-SC, Brazil.

Communities and Local Government 2010 *The Local Government Boundary Commission for England. Guidance on community governance reviews*, Crown Copyright, London.

Coombes, M.G. (2000) 'Defining Locality Boundaries with Synthetic Data', *Environment and Planning A,* 32(8), pp. 1499-1518.

Crawford, T.W. (2002) 'Spatial Modelling of Village Functional Territories to Support Population-Environment Linkages', in Walsh, S.J. and Crews-Meyer, K.A. (eds.) *Linking People, Place, and Policy: A GIScience Approach.* Boston: Kluwer Academic Publishers, pp. 91-111.

Curtin, K., Voicu, G., Rice, M. and Stefanidis, A. (2013) 'A Comparative Analysis of Traveling Salesman Solutions from Geographic Information Systems', *Transactions in GIS,* 18(2), pp. 286-301.

Dasgupta, S., Papadimitriou, C.H. and Vazirani, U. (2006) *Algorithms.* New York: McGraw-Hill.

Davies, C. (2009) 'Are Places Concepts? Familarity and Expertise Effects in Neighborhood Cognition', in Anonymous *Spatial Information Theory Proceedings of the 9th International Conference, COSIT.* Volume 5756 of the series Lecture Notes in Computer Science edn. France: Springer Berlin / Heidelberg, pp. 36-50.

Davies, C., Holt, I., Green, J., Harding, J. and Diamond, L. (2009) 'User Needs and Implications for Modelling Vague Named Places', *Spatial Cognition & Computation,* 9(3), pp. 174-194.

de Berg, M., Meulemans, W. & Specjmann, B. (2011) 'Delineating Imprecise Regions via Shortest-Path Graphs', *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 1-4 NovemberACM, pp. 271.

de Smith, M.J., Goodchild, M.F. and Longley, P. (2015) *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools.* 5th edn. Troubador Pub.

Dhanasekar, S., Hariharan, S. and Sekar, P. (2013) 'Classical Travelling Salesman Problem (TSP) based Approach to Solve Fuzzy TSP using Yager's Ranking', *International Journal of Computer Applications,* 74(13), pp. 1-4.

Dilo, A., De By, R.A. and Stein, A. (2007) 'A System of Types and Operators for Handling Vague Spatial Objects', *International Journal of Geographical Information Science,* 21(4), pp. 397-426.

Downs, J.A. (2010) 'Time-Geographic Density Estimation for Moving Point', in Fabrikant, S., Reichenbacher, T., Kreveld, M. and Schlieder, C. (eds.) *Geographic Information Science*
Lecture Notes in Computer Science Volume 6292. Springer Berlin Heidelberg, pp. 16-26.

Duckham, M., Mason, K., Stell, J. and Worboys, M. (2001) 'A Formal Approach to Imperfection in Geographic Information', *Computers, Environment and Urban Systems,* 25(1), pp. 89-103.

Duckham, M. & Worboys, M. (2001) 'Computational Structure in Three-valued Nearness Relations', Montello,D. (ed.) *Spatial Information Theory: COSIT 01 vol. 2205 of Lecture Notes in Computer Science*, 2001Springer-Verlag, Berlin, pp. 76.

Duckham, M., Kulik, L., Worboys, M. and Galton, A. (2008) 'Efficient Generation of Simple Polygons for Characterizing the Shape of a Set of Points in the Plane', *Pattern Recognition,* 41(10), pp. 3224-3236.

Duff, D. and Guesgen, H.W. (2002) 'An Evaluation of Buffering Algorithms in Fuzzy GISs', in Egenhofer, M.J. and Mark, D.M. (eds.) *Geographic Information Science, LNCS.* Heidelberg: Springer - Verlag Berlin, pp. 80-92.

Elaalem, M., Comber, A. and Fisher, P. (2011) 'A Comparison of Fuzzy AHP and Ideal Point Methods for Evaluating Land Suitability', *Transactions in GIS,* 15(3), pp. 329-346.

Erwig, M. & Schneider, M. (1997) 'Vague Regions', Scholl,M. & Voisard,A. (eds.) *Advances in Spatial Databases, 5th International Symposium, SSD '97 Conference Proceedings*, 15-18 JulyBerlin: Springer, pp. 298.

Evans, A.J. and Waters, T. (2008) 'Mapping Vernacular Geography: Web-baed GIS Tools for Capturing "Fuzzy" or "Vague" Entities', *International Journal of Technology, Policy and Management,* 7(2), pp. 1468-4322.

Fereidouni, S. (2011) 'Solving Traveling Salesman Problem by Using a Fuzzy Multi-objective Linear Programming', *African Journal of Mathematics and Computer Science Research,* 4(11), pp. 339-349.

Fischer, F.2014'Mathematics of Operational Research', .

Fisher, P. (2010) 'Remote Sensing of Land Cover Classes as Type-2 Fuzzy Sets', *Remote Sensing of Environment,* 114(2), pp. 309-321.

Fisher, P. (2000) 'Sorites Paradox and Vague Geographies', *Fuzzy Sets and Systems,* 113(1), pp. 7-18.

Fisher, P., Arnot, C., Wadsworth, R. and Wellens, J. (2006) 'Detecting Change in Vague Interpretations of Landscapes', *Ecological Informatics,* 1(2), pp. 163-178.

Fisher, P., Cheng, T. and Wood, J. (2007) 'Higher Order Vagueness in Geographical Information: Empirical Geographical Population of Type n Fuzzy Sets', *Geoinformatica,* 11(3), pp. 311-330.

Fisher, P. and Tate, N. (2015) 'Modelling Class Uncertainty in the Geodemographic Output Area Classification', *Environment and Planning B: Planning and Design,* 42(3), pp. 541-563.

Fisher, P., Tate, N. & Slingsby, A. (2014) 'Type-2 Fuzzy Sets Applied to Geodemographic Classification', *the 8th International Conference on Geographic Information Science (GIScience 2014)*, 23-26 September.

Fisher, P., Wood, J. and Cheng, T. (2004) 'Where is Helvellyn? Fuzziness of Multi-scale Landscape Morphometry', *Transactions of the Institute of British Geographers,* 29(1), pp. 106-128.

Fisher, P.F. (2009) 'The Representation of Uncertain Geographical Information', in Marguerite Madden (ed.) *The Manual of Geographical Information Systems.* Bethesda, MD., pp. 235-264.

Fisher, P.F. (1999) 'Models of Uncertainty in Spatial Data', in Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (eds.) *Geographical Information*

Systems: Principles and Technical Issues. 2 nd edn. The United States of America: Joh Wiley & Sons, Inc., pp. 191-205.

Fisher, P.F. (1994) 'Probable and Fuzzy Concepts of the Uncertain Viewshed', in Worboys, M. (ed.) *Innovations in GIS.* London: Taylor & Francis, pp. 161-175.

Fisher, P.F. & Almadani, F.M. (2011) 'Fuzzy Geographical Buffers ', *Proceedings of the Geographic Information Science Research UK - the19th Annual Conference (GISRUK 2011)*, 27- 29 AprilUniversity of Portsmouth, pp. 147.

Fisher, P. and Robinson, V.B. (2014) 'Fuzzy Modelling', in Abrahart, R.J. and See, L.M. (eds.) *GeoComputation.* 2 nd edn. CRC Press, pp. 283-306.

Fisher, P. and Wood, J. (1998) 'What is a Mountain? Or The Englishman who went up a Boolean Geographical Concept but Realised it was Fuzzy', *Geographical Association,* 83(3), pp. 247-256.

Flanagan, D.2007'What is Rural? Rural Ministries Discussion Paper', *Rural Ministries Discussion Paper,* .

Fonte, C.C. and Lodwick, W.A. (2004) 'Areas of Fuzzy Geographical Entities', *International Journal of Geographical Information Science,* 18(2), pp. 127-150.

Foody, G.M. (1996) 'Approaches for The Production and Evaluation of Fuzzy Land Cover Classification from Remotely-Sensed Data', *Ineternational Journal of Remote Sensing,* 17(7), pp. 1317-1340.

Freeman, J.V. and Young, T. (2009) 'Correlation Coefficient: Association between two continuous variables', in Freeman, J.V. (ed.) *Scope Tutorial Manual.* Scope Papers (collection of statistical papers first published in Scope), pp. 31-33.

GISGeography (2015) *Root Mean Square Error RMSE in GIS.* Available at: http://gisgeography.com/root-mean-square-error-rmse-gis/.

Goodchild, M.F. (2007) 'Citizens as Sensors: the World of Volunteered Geography', *GeoJournal,* 69(4), pp. 211-221.

Goodchild, M.F. (2000) 'Introduction: special issue on 'Uncertainty in Geographic Information Systems'', *Fuzzy Sets and Systems,* 113(1), pp. 3-5.

Goodchild, M.F., Montello, D.R., Fohl, P. & Gottsegen, J. (1998) 'Fuzzy Spatial Queries in Digital Spatial Data Libraries', *Fuzzy Systems Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference*, 4-9 MayIEEE, pp. 205.

Goyal, S.2010'A Survey on Travelling Salesman Problem', *http://www.cs.uwec.edu/MICS/papers/mics2010_submission_51.pdf*, .

Guesgen, H. and Albrecht, J. (2000) 'Imprecise Reasoning in Geographic Information Systems', *Fuzzy Sets and Systems,* 113(1), pp. 121-131.

Guesgen, H.W. (2002) 'Reasoning about Distance Based on Fuzzy Sets', *Applied Intelligence,* 17(3), pp. 265-270.

Guesgen, H.W., Hertzberg, J., Lobb, R. and Mantler, A. (2003) 'Buffering Fuzzy Maps in GIS', *Spatial Cognition & Computation,* 3(2 - 3), pp. 207-222.

Guha, D. and Chakraborty, D. (2010) 'A New Approach to Fuzzy Distance Measure and Similarity Measure between Two Generalized Fuzzy Numbers', *Applied Soft Computing,* 10(1), pp. 90-99.

Guo, Q., Liu, Y. and Wieczorek, J. (2008) 'Georeferencing Locality Descriptions and Computing Associated Uncertainty Using A Probabilistic Approach', *International Journal of Geographical Information Science,* 22(10), pp. 1067-1090.

Hahsler, M. and Hornik, K. (2007) 'TSP—Infrastructure for the Traveling Salesperson Problem', *Journal of Statistical Software,* 23(2), pp. 1-21.

Hahsler, M. and Hornik, K. (eds.) (2015) *TSP: Traveling Salesperson Problem (TSP). R package version 1.1-3.* http://CRAN.R-project.org/package=TSP.

Haklay, M. (2008) 'How Good is OpenStreetMap Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets for London and the Rest of England', *Environment & Planning,* 37, pp. 682-703.

Hall, M. (2010) *Modelling and Reasoning with Quantitative Representations of Vague Spatial Language used in Photographic Image Captions.* PhD Thesis. Cardiff University.

Hall, M., Smart, P.D. and Jones, C.B. (2011) 'Interpreting Spatial Language in Image Captions', *Cognitive processing,* 12(1), pp. 67-94.

Hayward, G. and Davidson, V. (2003) 'Fuzzy Logic Applications', *The Royal Society of Chemistry,* 128, pp. 1304-1306.

Held, M. and Karp, R. (1962) 'A Dynamic Programmin Approach to Sequencing Problems', *Journal of Society for Industrial and Applid Mathematics,* 10(1), pp. 196-210.

Hinckley and Bosworth Borough Council 2014 *Sustainability Appraisal of Development Management and Site Allocations Development Plan Document Sustainability Appraisal Report & Non Technical Summary*.

Hinckley and Bosworth Borough Council 2006 *Landscape Character Assessment Hinckley and Bosworth Borough*.

Hollenstein, L. (2008) *Capturing Vernacular Geography from Georeferenced Tags.* University of Zurich.

Hollenstein, L. and Purves, R.S. (2010) 'Exploring Place through User-Generated Content: Using Flickr Tags to Describe City Cores', *Journal of Spatial Information Science,* 1, pp. 21-48.

Hwang, S. and Thill, J.C. (2005) 'Modeling Localities with Fuzzy Sets and GIS', in Petry, F.E., Robinson, V.B. and Cobb, M.A. (eds.) *Fuzzy Modeling with Spatial Information for Geographic Problems.* Springer - Verlag Berlin, pp. 71-104.

JOHNSON, D. and ›□œŸ□›□œŸ□McGeoch, L. (2002) 'Expiremental Analysis Heuristics for STSP', in GUTIN, G. and PUNNEN, A. (eds.) *The Traveling Salesman Problem and Its Variations.* Springer US: Kluwer Academic Publishers, pp. 1-80.

Jones, C.B., Purves, R.S., Clough, P.D. and Joho, H. (2008) 'Modelling Vague Places with Knowledge from the Web', *International Journal of Geographical Information Science,* 22(10), pp. 1045-1065.

Karkory, F.A. and Abudalmola, A.A. (2013) 'Implementation of Heuristics for Solving Travelling Salesman Problem Using Nearest Neighbour and Minimum Spanning Tree Algorithms', *International Journal of Mathematical, Computational,Natural and Physical Engineering,* 7(10), pp. 987-997.

Kaufman, L. and Rousseeuw, P.J. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis.* New York: John Wiley and Sons.

Kaymaz, E. (1995) *Identification and Fuzzy Logic Control of Nonlinear Dynamic Systems, PhD Thesis.* Electrical Engineering in Texas Tech University.

Keen, K. (2010) *Graphics for Statistics and Data Analysis with R.* Chapman and Hall/CRC.

Keith, G. & McLaren, R. (2003) 'Addressing for Britain is not as simple as going from A to B', *The AGI Conference at GeoSolutions*, 16-18 September.

Klein, R. (1990) 'On the Construction of Abstract Voronoi Diagrams, Il '', in Anonymous ***Algorithms: International Symposium SIGAL**.* Springer Science & Business Media.

Klimesova, D. (2006) 'GIS and Uncertainty Management', *Remote Sensing: From Pixels to Processes*, 8-11 MayISPRS Commssion VII Mid-term Symposium, pp. 480.

Klir, G.J., Clair, U.S. and Yuan, B. (1997) *Fuzzy Set Theory: Foundations and Applications.* London: Prentice Hall.

Klir, G.J. and Yuan, B. (1995) *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice-Hall of India: Ghosh, A.K.;.

Kolog, E. (2012) *Design and Analysis of Algorithm Documentation of Dynamic Programming (Brute Force) for TSP.* Faculty of Science and Forestry, University of Eastern Finland.

Kulik, L. (2003) 'Spatial Vagueness and Second-Order Vagueness', *Spatial Cognition & Computation,* 3(2-3), pp. 157-183.

Kumar, A. and Gupta, A. (2012) 'Assignment and Travelling Salesman Problems with Coefficients as LR Fuzzy Parameters', *International Journal of Applied Science and Engineering,* 10(3).

Kumar, A. and Gupta, A. (2011) 'Methods for Solving Fuzzy Assignment Problems and Fuzzy Travelling Salesman Problems with Different Membership Functions', *Fuzzy Information and Engineering, Springer-Verlag,* 3(1), pp. 3-21.

Lee, G.S. and Lee, K.H. (2006) 'Application of Fuzzy Representation of Geographic Boundary to The Soil Loss Model', *Hydrology and Earth System Sciences Discussions,* 3(1), pp. 115-133.

Leung, Y. (1983) 'Fuzzy Sets Approach to Spatial Analysis and Planning - A Nontechnical Evaluation', *Geografiska Annaler. Series B, Human Geography,* 65(2), pp. 65-75.

Liang, G.S. and Ding, J.F. (2003) 'Fuzzy MCDM based on the concept of α-cut', *Journal of Multi-Criteria Decision Analysis,* 12(6), pp. 299-310.

Lienau, C. (1973) 'A Matter of Terminology: Rural Settlements', *Geoforum,* 4(4), pp. 92-94.

Liu, Y., Guo, Q.H., Wieczorek, J. and Goodchild, M.F. (2009) 'Positioning Localities Based on Spatial Assertions', *International Journal of Geographical Information Science,* 23(11), pp. 1471-1501.

Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (2001) *Geographic Information Systems and Science.* 3 rd edn. New York, USA: Wiley, J.; Sons, Inc.

Lu, Y. & Ni, Y. (2005) 'Fuzzy Random Traveling Salesman Problem', *Proceedings of the Fifth International Conference on Electronic Business, Hong Kong*, December 5-9, pp. 901.

Lüscher, P. and Weibel, R. (2013) 'Exploiting Empirical Knowledge for Automatic Delineation of City Centres from Large-Scale Topographic Databases', *Computers, Environment and Urban Systems,* 37, pp. 18-34.

Mansbridge, L. (2005) *Preceptions of Imprecise in Relation to Geographical Infrormation Retrieval.* MSc thesis, University of Sheffield.

Maredia, A. (2010) *History, Analysis and Implementation of Traveling Salesman Problem (TSP) and Related Problems.* University of Houston-Downtown.

Márquez, F.P.G. and Nieto, M.R.M. (2013) 'Heuristic Approaches for a Dual Optimization Problem', in Márquez, F.P.G. and Lev, B. (eds.) *Engineering Management.* INTECH Open Access Publisher.

Matai, R., Singh, S. and Mittal, M. (2010) 'Traveling Salesman Problem: An Overview of Applications, Formulations, and Solution Approaches', in Davendra, D. (ed.) *Traveling Salesman Problem, Theory and Applications.* InTech, pp. 1-25.

Mesgari, M.S., Pirmoradi, A. and Fallahi, G.R. (2008) 'Implementation of Overlay Function Based on Fuzzy Logic in Spatial Decision Support System', *World Applied Sciences Journal,* 3(1), pp. 60-65.

Moghaddam, H.K. and Delavar, M.R. (2007) 'A GIS - based Pipelining Using Fuzzy Logic and Statistical Models', *International Journal of Computer Science and Network Security,* 7(2), pp. 117-123.

Montello, D.R., Goodchild, M.F., Gottsegen, J. and Fohl, P. (2003) 'Where's Downtown? Behavioral Methods for Determinig Referents of Vague Spatial Queries', *Spatial Cognition & Computation,* 3(2 - 3), pp. 185-204.

Mooney, P. & Corcoran, P. (2011) 'Accessing the History of Objects in OpenStreetMap.', *The 14th AGILE International Conference on Geographic Information Science*Geertman, S.; Reinhardt, W.; Toppen, F.;, pp. 155.

Moore, D.S. (2010) *The Basic Practice of Statistics.* Fifth Edition edn. New York: W.H. Freeman and Company.

Neis, P. and Zielstra, D. (2014) 'Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap.', *Future Internet,* 6, pp. 76-106.

Ordnance Survey 2014 *Points of Interest User Guide v3.3*, © Crown copyright.

Ordnance Survey 2011 *OS MasterMap Address Layer & Address Layer 2 User Guide*, Crown copyright.

Ordnance Survey 2011 *OS MasterMap Address Layer 2 Technical Specification*, Crown Copyright.

Osborne, B. (2011) *Mission and Discipleship in a Rural Context.* Available at: http://www.ruralmissionsolutions.org.uk/app/download/5784283968/Mission+and+Discipleship+in+a+Rural+Context.pdf.

Osborne, B. (2010) *Defining Rural.* Available at: http://www.ruralmissionsolutions.org.uk/app/download/5784283939/Defining+Rural.pdf.

Pawlak, Z. (2002) 'Rough Set Theory and Its Applications', *Journal of Telecommunications and Information Technology,* 3, pp. 7-10.

Pawlak, Z. (1991) *Rough Sets: Theoretical Aspects of Reasoning about Data.* London : Kluwer: Dordrecht.

Pedrycz, W. and Gomide, F. (1998) *An Introduction to Fuzzy Sets: Analysis and Design.* London: MIT Press.

Pothineni, C.2013'Travelling Salesman Problem using Branch and Bound Approach', .

Prakash, T.N. (2003) *Land Suitability Analysis for Agricultural Crops: A fuzzy Multicriteria Decision Making Approach*. Internationsl Institute for Geo-information Science an Earth Observation inpartial fulfilment of the requirements for the degree of Msc.

Randell, D.A., Cui, Z. & Cohn, A.G. (1992) 'A Spatial Logic Based on Regions and Connection', *Principles of Knowledge Reasoning: Proceeding of the 3rd International confercnce*Morgan Kaufmann, pp. 165.

Raubal, M. and Worboys, M. (1999) 'A Formal Model of the Process of Wayfinding in Built Environments', in Anonymous *Spatial information theory. Cognitive and computational foundations of geographic information science.* Berlin Heidelberg: Springer, pp. 381-399.

Reinbacher, I., Benkert, M., van Kreveld, M., Mitchell, J.S., Snoeyink, J. and Wolff, A. (2008) 'Delineating Boundaries for Imprecise Regions', *Algorithmica,* 50(3), pp. 386-414.

Reinbacher, I., Benkert, M., Van Kreveld, M., Mitchell, J.S. and Wolff, A. (2005) 'Delineating Boundaries for Imprecise Regions', in Anonymous *Algorithms.* ESA edn. Berlin Heidelberg: Springer, pp. 143-154.

Robinson, V.B. (2003) 'A Perspective on the Fundamentals of Fuzzy Sets and their Use in Geographic Information Systems', *Transactions in GIS,* 7(1), pp. 3-30.

Robinson, V.B. (1988) 'Some Implications of Fuzzy Set Theory Applied to Geographic Databases', *Computers, Environment and Urabn Systems,* 12, pp. 89-98.

Rosenfeld, A. (1985) 'Distance between Fuzzy Sets', *Pattern Recognition Letters,* 3, pp. 229-233.

Rosenfeld, A. (1984) 'The Diameter of A Fuzzy Set', *Fuzzy Sets and Systems,* 13(3), pp. 241-246.

Rosenkrantz, D., STEARNS, R. and LEWIS II, P. (1977) 'An Analysis of Several Heuristics for the Travelling Salesman Probalem', *SIAM J. Computing,* 6(3), pp. 563-581.

Rosser, J. & Morley, J. (2010) 'Rate My Place: A Social Network Application for Crowd-Sourcing Vernacular Geographic Areas', Haklay,M., Morley,J. & Rahemtulla,H. (eds.) *Proceedings of the GIS Research UK 18th Annual Conference*, 14 - 16 AprilUniversity College London, pp. 155.

Saggaf, M.M. and Nebrija, E.L. (2003) 'A Fuzzy Logic Approach for the Estimation of Facies from Wire-Line Logs', *AAPG Bulletin,* 87(7), pp. 1223-1240.

Salkind, N.J. (2010) *Encyclopedia of Research Design.* Thousand Oaks, CA, USA: SAGE Publications, Inc.

Salski, A. (1999) 'Fuzzy Logic Approach to Data Analysis and Ecological Modelling', *Proceeding of European Symposium on Intelligent Techniques' (EUFIT99)*Orthodox Academy of Crete, Greece.

Schmitz, A. and Morris, A. (2006) 'Modeling and Manipulating Fuzzy Regions: Strategies to Define the Topological Relation between Two Fuzzy Regions', *Control and Cybernetics,* 35(1), pp. 37.

Schneider, M. (2000) 'Metric Operations on Fuzzy Spatial Objects in Databases', *Proceeding of the 8th ACM International Symposium on Advances in Geographic Information Systems*ACM New York, USA, pp. 21.

Schneider, M. (1999) 'Uncertainty Management for Spatial Data in Databases: Fuzzy Spatial Data Types', *Lecture Notes in Computer Science: Advances in Spatial Databases,* 1651, pp. 330-351.

Schockaert, S. & de Cock, M. (2007) 'Neighborhood Restrictions in Geographic IR', *The 30th Annual International ACM SIGIR Conference*, 23-27 July, pp. 167.

Sicat, R.S., Carranza, E.J.M. and Nidumola, U.B. (2005) 'Fuzzy Modeling of Farmers' Knowledge for Land Suitability Classification', *Agricultural Systems,* 83, pp. 49-75.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* New York, USA: Chapman and Hall.

Singh, U.B. (2009) *Decentralized Democratic Governance in New Millennium.* 1st edn. Concept Publishing.

Sui, D.Z. (1992) 'A Fuzzy GIS Modeling Approach for Urban Land Evaluation', *Computers, Environment and Urabn Systems,* 16, pp. 101-115.

Tapia, R. (2004) *Optimization of Sampling Schemes for Vegetation Mapping Using Fuzzy Classification.* Internationsl Institute for Geo-information Science an Earth Observation, Msc Thesis.

The Institute for Name-Studies (INS) at the University of Nottingham (2015) *Key to English Place-Names.* Available at: http://www.nottingham.ac.uk/ins/key-to-english-place-names.aspx.

Twaroch, F.A., Jones, C.B. & Abdelmoty, A.I. (2008 a) 'Acquisition of A Vernacular Gazetteer from Web Sources', *Proceedings of the 1st international workshop on Location and the web*, 21-25 AprilACM New York, USA, pp. 61.

Twaroch, F.A., Smart, P.D. & Jones, C.B. (2008 b) 'Minning the Web to Detect Place Names', *Proceeding of the 2nd International Workshop on Geographic Information Retrieval (GIR '08)*ACM, pp. 43.

Varzi, A.C. (2001) 'Vagueness in Geography', *Philosophy and Geography,* 4(1), pp. 49-65.

Venables, W.N. and Ripley, B.D. (2002) *Modern Applied Statistics with S.* 4th edn. New York: Springer-Verlag.

Verstrate, J., de Tre, G., de Caluwe, R. and Hallez, A. (2005) 'Field Based Methods for the Modeling of Fuzzy Spatial Data', in Petry, F.E., Robinson, V.B. and Cobb, M.A. (eds.) *Fuzzy Modeling with Spatial Information for Geographic Problems.* Germany: Springer, pp. 41-69.

Verstrate, J., Hallez, A. and De Tre, G. (2007) 'Fuzzy Regions: Theory and Applications', in Morris, A. and Kokhan, S. (eds.) *Geographic Uncertainty in Environmental Security.* Amesterdam: Springer, pp. 1-17.

Voxman, W. (1998) 'Some Remarks on Distances between Fuzzy Numbers', *Fuzzy Sets and Systems,* 100, pp. 353-365.

Vullings, W., de Veries, M. & de Borman, L. (2007) 'Dealing with Uncertainty in Spatial Planning', Aalborg University,G. (ed.) *The 10th AGILE International Conference on Geographic Information Science*, 8-11 MaySpringer, pp. 1.

Waters, T. & Evans, A.J. (2003) 'Tools for Web-based GIS Mapping of A "Fuzzy" Vernacular Geography', *Proceeding of 7th International Conference on GeoComputation, Southampton* Southampton, 8-10 September.

Wilkinson, G.N. and Rogers, C.E. (1973) 'Symbolic descriptions of factorial models for analysis of variance', *Applied Statistics,* 22, pp. 392-399.

Williamson, T. (1996) *Vagueness.* New York: Routledge.

Woods, M. (2011) 'Grounding global uncertainties and the relational politics of the rural', *AAG Annual Meeting, Seattle, Washington*, 12 - 16 Apri.

Woods, M. (2011) *Rural.* Illustrated edn. Taylor and Francis.

Worboys, M. (1998 b) 'Computation with imprecise geospatial data', *Computers, Environment and Urban Systems,* 22(2), pp. 85-106.

Worboys, M. (1998 a) 'Imprecision in Finite Resolution Spatial Data', *Geoinformatica,* 2(3), pp. 257-279.

Worboys, M. and Clementini, E. (2001) 'Integration of Imperfect Spatial Information', *Journal of Visual Languages and Computing,* 12, pp. 61-80.

Worboys, M.F. ((2001)) 'Nearness Relations in Environmental Space', *International Journal of Geographical Information Science,* 15(7), pp. 651.

Zadeh, L.A. (1965) 'Fuzzy Sets', *Information and Control,* 8, pp. 338-353.

Zadeh, L. (2003) 'Probability Theory & Fuzzy Logic', *Los Alamos National Laboratory – Sponsored by ESA Div*, April 24, 2003Computer Science Division ,Department of EECS, UC Berkeley.

Zadeh, L. (1995) 'Discussion: Probability Theory and Fuzzy Logic are Complementary rather than Competitive.', *Technometrics,* 37(3).

Zimmermann, H.J. (2001) *Fuzzy Set Theory and its Applications.* Fourth Edition edn. The United States of America: Kluwer Academic Publishers.