

ESTIMATING POPULATION SURFACES IN AREAS WHERE  
ACTUAL DISTRIBUTIONS ARE UNKNOWN: DASYMETRIC  
MAPPING AND PYCNOHYLACTIC INTERPOLATION ACROSS  
DIFFERENT SPATIAL SCALES

Thesis submitted for the degree of  
Doctor of Philosophy  
at the University of Leicester

by

Idris Jega Mohammed MSc (Leicester)  
Department of Geography  
University of Leicester

June 2015

## **Idris Jega Mohammed**

**Title:** Estimating population surfaces in areas where actual distributions are unknown: dasymetric mapping and pycnophylactic interpolation across different spatial scales.

### **ABSTRACT**

Spatially distributed estimates of population provide commonly used demand surfaces in support of spatial planning. In many countries, spatially detailed population summaries are not available. For such cases a number of interpolation methods have been proposed to redistribute summary population totals over small areas. Population allocations to small areas are commonly validated by comparing the estimates with some known values for those areas. In areas where spatially detailed estimates of the population do not exist, that is where the actual population in small areas is unknown, such as Nigeria validation is problematic. This research explores different interpolation methods applied at different scales in areas where the actual population distribution is known and where validation is possible. It then applies the parameters developed from these results to areas where the distribution is unknown. The binary dasymetric method using land cover data derived from a classified 30m spatial resolution satellite imagery as the ancillary data input and with disaggregation over 30m support grids, was found to provide the best target zones estimates of the population. The demand surfaces were then used to evaluate current health facility locations and then to suggest alternative spatial arrangements for health centres in Port-Harcourt, Nigeria. The average distance from each demand point to the nearest healthcare centre was found to be 1204m. When alternative locations for the current health centres were identified, the results suggest 13 service provision points would provide almost the same demand coverage as the 17 current PHCCs. This research develops methods that can be used to support informed decision making in spatial planning and policy development.

## **DEDICATION**

To my late father, Alhaji Muhammadu Jega (1<sup>st</sup> Dan Lawan Gwandu).

## **ACKNOWLEDGEMENT**

I would like to express my special appreciation and thanks to my supervisors Professor Alexis J. Comber and Dr. Nicholas J. Tate who patiently provided the vision, encouragement and advise necessary for me to continuously progress through the stages of my thesis. I am also deeply grateful to Professor Chris Brunsdon for his priceless contributions during the first year of my research as my supervisor.

I would like to express my gratitude to Petroleum Technology Development Fund (PTDF) for fully sponsoring my PhD. To the data providers; Ordnance Survey UK, Astrium Services UK and Geo-technics Services Limited, I am grateful.

Special thanks to my family and friends. Words cannot express how grateful I am to all for the sacrifices you made on my behalf. Your prayer for me was what sustained me thus far. At the end, I would like to express my appreciation to my beloved wife Hafsat and my children (Asma'u and Ahmed) for their endurance during the time I had been away from home. I have no suitable word that can fully describe their everlasting love for me.

## LIST OF ABBREVIATIONS AND ACRONYMS

ACLP	Anti-Covering Location Problem
CASWEB	Census Area Statistics on the Web
CEDS	Cadastral-based Expert Dasymetric System
CoV	Coefficient of Variance
EIA	Energy Information Administration
ELP	Expropriation Location Problem
EM	Expectation-Maximization
EMS	Emergency Medical Services
ETM+	Enhanced Thematic Mapper Plus
EU	European Union
GGA	Grouping Genetic Algorithm
GIS	Geographical Information Science
GOR	Government Office Regions
GP	General Practitioner
GWEM	Geographically Weighted Expectation Maximization
GWR	Geographically Weighted Regressions
HUP	Homogeneous Urban Patches
INCRA	National Institute of Colonization and Agrarian Reform
IQR	Interquartile Range
ISODATA	Iterative Self-Organising Data Analysis Technique Algorithm
LGAs	Local Government Areas
LiDAR	Light Detection and Ranging
LILP	Limited Impact Location Problem
LSCP	Location Set Covering Problem
LSOA	Lower Super Output Area
LV	Limiting Variable
MCLP	Maximal Covering Location Problem
MCLPDC	Minimum Covering Location Problem with Distance Constraints
MDGs	Millennium Development Goals
MEXCLP	Maximal Expected Covering Location Problem
MLMCD	Multi-Layer Multi-Class Dasymetric

MS	Multiple Sclerosis
MSOA	Middle Super Output Area
NASRDA	National Space Research and Development Agency
NHP	National Health Policy
NNPC	Nigerian National Petroleum Corporation
NPC	National Population Commission
NPHCDA	National Primary Health Care Development Agency
OA	Output Area
OD	Origin and Destination
OLS	Ordinary Least Squares
ONS	Office of National Statistics
PHCC	Primary Health Care Centre
PHCCs	Primary Health Care Centres
PTDF	Petroleum Technology Development Fund
PUMS	Public Use Microdata Sample
RMSE	Root Mean Squared Error
RNB	Road Network Buffer
RNHP	Revised National Health Policy
SANET	Spatial Analysis Network Tools
SW	Street Weighting
TIGER	Topologically Integrated Geographic Encoding and Referencing
TIN	Triangulated Irregular Networks
UA	Unitary Authority
UAVs	Unmanned aerial vehicles
UK	United Kingdom
UPCs	Unit Postcodes
US	United States
USGS	United States Geological Survey

## LIST OF PUBLICATIONS

### Peer reviewed conference proceedings

**Mohammed, I. J.**, Comber, A. and Tate, N, 2013. Effects of Land Cover resolution on spatially distributed demand surfaces: the binary dasymetric approach. GISRUUK 2013, *Proceedings of the Geographical Information Science UK Conference*. University of Liverpool 3rd-5th April 2013. University of Liverpool, Liverpool

**Mohammed, I. J.**, Comber, A. and Brunson, C., 2012. Population estimation in small areas: combining dasymetric mapping with pycnophylactic interpolation. GISRUUK 2012, *Proceedings of the GIS Research UK 20th Annual Conference, Volume 1*, pp.79-87 (eds. Barry Rowlingson and Duncan Whyatt), 11-13th April 2012, Lancaster.

### Poster presentation at festival of postgraduate research

**Mohammed, I. J.**, 2013. Effects of satellite image resolution on population estimates, *University of Leicester Festival of Postgraduate Research 2013*.

# TABLE OF CONTENTS

ABSTRACT.....	ii
DEDICATION.....	iii
ACKNOWLEDGEMENT .....	iv
LIST OF ABBREVIATIONS AND ACRONYMS .....	v
LIST OF PUBLICATIONS .....	vii
TABLE OF CONTENTS.....	viii
LIST OF FIGURES .....	xiii
LIST OF TABLES.....	xix
Chapter 1.....	1
1. INTRODUCTION .....	1
1.1 Problem description .....	1
1.2 Motivation.....	4
1.3 Aim and objectives of the study .....	5
1.4 Research questions.....	6
1.5 Thesis Structure .....	7
Chapter 2.....	9
2. LITERATURE REVIEW .....	9
2.1 Introduction.....	9
2.2 Areal Interpolation Techniques .....	9
2.2.1 The Pycnophylactic Interpolation Technique .....	12
2.2.2 The Dasymetric Mapping Method.....	14
2.3 Influence of demand population on spatial accessibility .....	18
2.4 GIS and Geographical Access to Healthcare .....	19
2.4.1 Review of spatial access .....	20
2.5 Location-allocation models.....	22
2.5.1 The p-Median problem .....	23

2.6 Summary .....	27
Chapter 3 .....	28
3. METHODS .....	28
3.1 Introduction.....	28
3.2 Study areas .....	30
3.2.1 Leicester.....	30
3.2.2 Port-Harcourt .....	32
3.3 Data.....	34
3.3.1 Leicester.....	34
3.3.2 Port-Harcourt .....	43
3.4 Implementation of areal interpolation methods .....	50
3.4.1 Supervised Classification.....	50
3.4.2 The binary dasymetric method .....	56
3.4.3 Pycnophylactic Interpolation Method.....	58
3.5 Evaluation of surfaces.....	62
3.6 Summary .....	65
Chapter 4.....	67
4. RESULTS: LEICESTER.....	67
4.1 Introduction.....	67
4.2 Supervised classification.....	67
4.2.1 Classified images .....	68
4.2.2 Binary classified images .....	72
4.3 Areal interpolation methods.....	76
4.3.1 The binary dasymetric method .....	76
4.3.2 Pycnophylactic interpolation method .....	80
4.4 Evaluation of Surfaces .....	83
4.4.1 Residual Maps.....	87
4.5 Summary of results for Leicester.....	92
Chapter 5.....	93

5. RESULTS: PORT-HARCOURT .....	93
5.1 Introduction.....	93
5.2 Supervised classification.....	93
5.2.1 Classified images .....	96
5.2.2 Binary classified images .....	100
5.3 Areal interpolation methods.....	104
5.3.1 The binary dasymetric method .....	104
5.4 Visual inspection of demand surfaces .....	108
5.5 What the errors from Leicester mean in Port-Harcourt .....	112
5.6 Summary.....	113
Chapter 6.....	114
6. APPLICATION: Location-allocation in Port-Harcourt .....	114
6.1 Case study: Primary Health Care Centres (PHCCs) in Port-Harcourt.....	114
6.2 Data and preparation.....	116
6.2.1 Network distances.....	117
6.3 Methods and models .....	121
6.4 Results.....	124
6.4.1 Current locations of PHCCs .....	124
6.4.2 Potential locations of PHCCs .....	125
6.4.3 Adjust current locations .....	126
6.4.4 Adjust potential locations .....	131
6.5 Summary.....	134
Chapter 7.....	135
7. DISCUSSION .....	135
7.1 Introduction.....	135
7.2 Leicester: surface generation .....	136
7.3 Port-Harcourt: surfaces and location-allocation .....	139
7.4 Methods .....	141
7.5 Reflections on the methods and results.....	144

7.6 Limitations and suggestions for future work .....	147
Chapter 8.....	152
8. CONCLUSIONS .....	152
8.1 Introduction.....	152
8.2 Research findings.....	152
8.2.1 Research objective one: The application of dasymetric and pycnophylactic interpolations across different spatial scales to redistribute aggregate population census data for Leicester over small areas.....	153
8.2.2 Research objective two: A comparison of the estimated populations from the interpolations for three different census units and the known census counts in each case, to test the performance of the interpolation methods.....	153
8.2.3 Research objective three: Adapt the model with the most appropriate parameters obtained from Leicester to estimate demand population values in Port- Harcourt, Nigeria. ....	154
8.2.4 Research objective four: Evaluate the public health facility locations currently in place in Port-Harcourt. ....	154
8.2.5 Research objective five: Suggest alternative spatial arrangement of public health facilities using heuristic location-allocation modelling approaches. ....	155
8.3 Contributions .....	155
8.4 Policy recommendation .....	156
9. BIBLIOGRAPHY.....	157
10. APPENDICES .....	184
Appendix 1: Metadata file information for Leicester image downloaded from USGS .....	184
Appendix 2: Metadata file information for Port-Harcourt image downloaded from USGS .....	188
Appendix 3: Port-Harcourt Spot5 imagery information from Astrium Services.....	191
Appendix 4: The mean plot window for resampled aerial photo of 10m and 3m spatial resolution image of Leicester. ....	192
Appendix 5: The signature editor and the signature mean plot for the combined signatures for resampled aerial photo of 10m and 3m spatial resolution image of Leicester.....	194

Appendix 6: Classification accuracy report for Leicester .....	196
Appendix 7: Residual maps .....	200
Appendix 8: Signature Mean Plot, Mean plot and Signature editor for the combined signatures for Port-Harcourt .....	224
Appendix 9: Classification accuracy report for Port-Harcourt.....	228
Appendix 10: The results of visual inspection of surfaces .....	233
Appendix 11: Codes for allocating demand to all 17 current health centres .....	248
Appendix 12: Codes for generating optimal locations using 85 potential locations	250
Appendix 13: Spatial distributions of 5 to 16 PHCCs selected from current locations .....	252
Appendix 14: Spatial distributions of 5 to 20 PHCCs selected from potential locations. ....	264
Appendix 15: Codes for locating specified number of sites for current and potential locations. ....	271
Appendix 16: Permission to use pycno illustration from Uwe Deichmann .....	273
Appendix 17: Request for Spot5 (colour) 10m spatial resolution .....	276

## LIST OF FIGURES

Figure 3.1 - The research design highlighting the first stage in blue dash and the second stage in red long dash dots.....	29
Figure 3.2 - The map of (a) England showing location of Leicestershire County; (b) Leicestershire County with location of Leicester UA; (c) Leicester UA. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.....	32
Figure 3.3 - The map of (a) Nigeria showing location of Rivers State; (b) Rivers State showing location of Port-Harcourt; (c) Port-Harcourt City Local Government Area. The digital boundaries are Copyright for Geotechnics Services 2011.....	33
Figure 3.4 - The Landsat7 (ETM) image 30m spatial resolution of the Leicester area (black polygon) with image band combination 4:3:2. ....	35
Figure 3.5 - One of the 98 tiles of 25cm ortho-rectified aerial photograph with image band combination 1:2:3. The tile is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. ....	36
Figure 3.6 - 25cm ortho-rectified aerial photograph of Leicester area resampled to 3m with image band combination 1:2:3.....	37
Figure 3.7 - 25cm ortho-rectified aerial photograph of Leicester area resampled to 10m with image band combination 1:2:3.....	38
Figure 3.8 – The boundary data for; (a) Leicester UA, (b) 100m square grids, and (c) 30m square grids. The digital boundary is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. ....	39
Figure 3.9 – MSOA used as the testing zones. The digital boundary is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. ....	40
Figure 3.10 - LSOA used as the testing zones. The digital boundary is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. ....	41
Figure 3.11 - OA used as testing zones. The digital boundary is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.....	42
Figure 3.12 - The Landsat7 (ETM+) 30m spatial resolution covering Port-Harcourt area (shaded in grey), Rivers State, Nigeria with image band combination 4:3:2. ....	44
Figure 3.13 - Subset for Port-Harcourt acquired from the Landsat (ETM+) 30m spatial resolution. Port-Harcourt city covers an area of about 109 square kilometres.....	45

Figure 3.14 – Spot5 colour image 10m spatial resolution for Port-Harcourt obtained from Astrium Services SpotCatalog ( <a href="http://catalog.spotimage.com/PageSearch.aspx">http://catalog.spotimage.com/PageSearch.aspx</a> ). 46	
Figure 3.15 – Quickbird satellite image 60cm spatial resolution covering Port-Harcourt (red polygon) with image band combination 1:2:3. The image was obtained from Geotechnics Services Limited, Port-Harcourt. ....	47
Figure 3.16 - Quickbird satellite imagery 60cm spatial resolution covering Port-Harcourt resampled to 3m with image band combination 1:2:3.....	48
Figure 3.17 - The boundary data for; (a) Port-Harcourt LGA (source zone), and (b) 30m square grids used to derive the modelled population surface. The digital boundary is Copyright for Geotechnics Services 2011. ....	49
Figure 3.18 - Signature mean plot evaluating signatures for vegetation, builtup and water from Landsat7 (ETM) 30m spatial resolution image.....	52
Figure 3.19 - Signature editor for the combined signatures from Landsat7 (ETM) 30m resolution image.....	52
Figure 3.20 - Signature mean plot for the combined signatures from Landsat7 (ETM) 30m resolution image.....	53
Figure 3.21 – Flowchart showing steps involved in supervised classification.....	55
Figure 3.22 - Implementation steps for the binary dasymmetric method (vector mode) ..	57
Figure 3.23 - Source zones used for the pycnophylactic interpolation method with the City of Leicester (study area) shaded in grey. ....	59
Figure 3.24 – The pycnophylactic interpolation method .....	60
Figure 3.25 - A flow chart for the implementation steps for the pycnophylactic interpolation method. ....	61
Figure 3.26 - Leicester LSOAs intersect interpolated gridded pycnophylactic surfaces at resolutions of 100m postings .....	64
Figure 4.1 – The classified Leicester image derived from Landsat7 (ETM) 30m spatial resolution. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. ....	69
Figure 4.2 - The classified Leicester image derived from resampled aerial photo data of 10m spatial resolution. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.....	70
Figure 4.3 - The classified Leicester image derived from resampled aerial photo data of 3m spatial resolution. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.....	71

Figure 4.4 – A binary mask derived from land cover data derived from classified Landsat7 (ETM) 30m spatial resolution. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.....	73
Figure 4.5 – A binary mask derived from land cover data of 10m spatial resolution derived from classified resampled aerial photo data. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. ....	74
Figure 4.6 - A binary mask derived from land cover data of 3m spatial resolution derived from classified resampled aerial photo data. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. ....	75
Figure 4.7 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified Landsat7 (ETM) 30m spatial resolution ancillary data input.....	77
Figure 4.8 - The dasymetric map of population surface at 100m posting created using land cover data derived from classified resampled aerial photo data of 10m spatial resolution ancillary data input.....	78
Figure 4.9 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified resampled aerial photo data of 3m spatial resolution ancillary data input.....	79
Figure 4.10 – The change in population density at 30m resolution output grid. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.....	81
Figure 4.11 - The change in population density at 100m resolution output grid. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.....	82
Figure 4.12 - The spatial distribution of residuals at MSOA from a 100m gridded pycnophylactic population surface. The mean count error is 0 and a standard deviation of 3975. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. ....	89
Figure 4.13 - The spatial distribution of residuals at LSOA from a 30m gridded dasymetric population surface using land cover data derived from classified resampled aerial photo data of 3m spatial resolutions as the ancillary data input. The mean count	

error is -233 and a standard deviation of 1151. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. 90

Figure 4.14 - The spatial distribution of residuals at OA from a 30m gridded dasymetric population surface using land cover data derived from classified resampled aerial photo data of 10m spatial resolutions as the ancillary data input. The mean count error is -107 and a standard deviation of 434. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service. .... 91

Figure 5.1 - Signature mean plot evaluating signatures for vegetation, built-up and water from Landsat7 (ETM+) 30m spatial resolution image. .... 94

Figure 5.2 - Signature mean plot for the combined signatures from Landsat7 (ETM) 30m resolution image..... 95

Figure 5.3 - Signature editor for the combined signatures from Landsat7 (ETM) 30m resolution image..... 95

Figure 5.4 - The classified Port-Harcourt image derived from Landsat7 (ETM+) 30m spatial resolution. The digital boundary is Copyright for Geotechnics Services 2011... 97

Figure 5.5 - The classified Port-Harcourt image derived from Spot5 (colour) 10m spatial resolution. The digital boundary is Copyright for Geotechnics Services 2011..... 98

Figure 5.6 - The classified Port-Harcourt image derived from resampled Quickbird (60cm) image data of 3m spatial resolution. The digital boundary is Copyright for Geotechnics Services 2011. .... 99

Figure 5.7 - A binary mask derived from land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution image data. The digital boundary is Copyright for Geotechnics Services 2011. .... 101

Figure 5.8 - A binary mask derived from land cover data of 10m spatial resolution derived from classified Spot5 (colour) image data. The digital boundary is Copyright for Geotechnics Services 2011. .... 102

Figure 5.9 - A binary mask derived from land cover data of 3m spatial resolution derived from classified resampled Quickbird (60cm) image data. The digital boundary is Copyright for Geotechnics Services 2011..... 103

Figure 5.10 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution ancillary data input. The digital boundary is Copyright for Geotechnics Services 2011. .... 105

Figure 5.11 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified Spot5 (colour) 10m spatial resolution ancillary data input. The digital boundary is Copyright for Geotechnics Services 2011. ....	106
Figure 5.12 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified resampled quickbird (60cm) image of 3m spatial resolution ancillary data input. The digital boundary is Copyright for Geotechnics Services 2011. ....	107
Figure 5.13 - The demand surfaces for Port-Harcourt with 200 random points generated within the boundary of Port-Harcourt. The digital boundary is Copyright for Geotechnics Services 2011. ....	108
Figure 5.14 - An overlay of study area boundary with random points generated within the boundary on Google Earth 7.1. ....	109
Figure 5.15 - Random point on the populated surface corresponds to an unpopulated area in Google Earth reference. ....	110
Figure 5.16 - Random point on the unpopulated surface is identified in a populated area. ....	111
Figure 5.17 - Random point on the populated surface is identified in an industrial area. ....	111
Figure 6.1 - Source zone with roads (in grey) and the locations of current PHCCs in Port-Harcourt. The digital boundary is Copyright for Geotechnics Services 2011. ....	115
Figure 6.2 - The road network dataset showing roads (in grey colour) and road junctions in red dots. ....	117
Figure 6.3 – The spatial distribution of 17 current PHCCs (red dots) and the demand points generated from areal interpolation (grey dots). The digital boundary is Copyright for Geotechnics Services 2011. ....	118
Figure 6.4 - Lines representing shortest distances from each demand to each health centre. ....	119
Figure 6.5 – The distribution of lines on the road network ....	120
Figure 6.6 - The spatial distribution of 85 potential new locations of PHCCs (red dots) and the demand points generated from areal interpolation (grey dots). The digital boundary is Copyright for Geotechnics Services 2011. ....	121
Figure 6.7 - Spatial distribution of 17 optimal locations selected (red circles) from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geotechnics Services 2011. ....	127

Figure 6.8 - Average distances to the nearest current PHCCs plotted against the number of current PHCCs in a subset.....	128
Figure 6.9 - The spatial distributions of selected locations of 5 PHCCs from current locations. The digital boundary is Copyright for Geotechnics Services 2011.....	129
Figure 6.10 - The spatial distributions of selected locations of 10 PHCCs from current locations. The digital boundary is Copyright for Geotechnics Services 2011.....	130
Figure 6.11 - The spatial distributions of selected locations of 15 PHCCs from current locations. The digital boundary is Copyright for Geotechnics Services 2011.....	131
Figure 6.12 - Variation in average distance travelled from demand to: (⊕) nearest location of current PHCCs (current distance); and (•) nearest potential location selected from 85 potential new locations.....	133
Figure 6.13 - The spatial distribution of 13 optimal locations (red circles) selected from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geo-technics Services 2011. ....	134

## LIST OF TABLES

Table 2.1 Characteristics of areal interpolation methods .....	11
Table 2.2 Target zones used in previous literature whose population is known .....	18
Table 3.1 Percentage change in population for Leicester from 1951 to 2011 .....	31
Table 3.2 Data: Leicester .....	34
Table 3.3 Data: Port-Harcourt .....	43
Table 3.4 Classified land cover categories for Leicester .....	51
Table 3.5 Population totals for the source zones used to implement the pycnophylactic interpolation method. ....	58
Table 4.1 - Comparison of the overall accuracy and Kappa statistic between the different land cover data .....	68
Table 4.2 - Comparison of the total built-up area in the source zone between the different land cover data .....	72
Table 4.3 - Population density per 10,000 m <sup>2</sup> for binary dasymetric maps of population .....	76
Table 4.4 - Interpolation results from Leicester UA to 30m square grids postings, aggregated at MSOA .....	84
Table 4.5 - Interpolation results from Leicester UA to 100m square grids postings, aggregated at MSOA .....	84
Table 4.6 - Interpolation results from Leicester UA to 30m square grids postings, aggregated at LSOA.....	85
Table 4.7 - Interpolation results from Leicester UA to 100m square grids postings, aggregated at LSOA.....	85
Table 4.8 - Interpolation results from Leicester UA to 30m square grids postings, aggregated at OA .....	86
Table 4.9 - Interpolation results from Leicester UA to 100m square grids postings, aggregated at OA .....	86
Table 5.1- Comparison of the overall accuracy and Kappa statistic between the different land cover data .....	96
Table 5.2 - Comparison of the total built-up area in the source zone between the different land cover data .....	100

Table 5.3 - Population density per 10,000 m <sup>2</sup> for binary dasymetric maps of population .....	105
Table 5.4 – Summary of visual inspection of surfaces .....	110
Table 6.1 Demand allocated to current PHCCs in Port-Harcourt.....	125
Table 6.2 Demand allocated to potential PHCCs in Port-Harcourt.....	126
Table 6.3 Current and modelled average distances for facilities .....	132

# Chapter 1

## 1. INTRODUCTION

### 1.1 Problem description

Population estimates for small areas are important for many types of spatial data analysis. They are especially important for accessibility studies that are commonly used to support spatial planning and policy development, and facility location-allocation analyses. Population censuses provide a reliable record of socioeconomic characteristics and the spatial distribution of residential population (Langford 2013) and thereby support geodemographic analyses (Harris and Longley 2002). In the U.K. census, population counts are collected for each household and published as aggregate counts and statistics for fixed pre-defined spatial units with Output Areas (OA) being the most detailed. The OA is similar to a U.S. census block. The OA was designed to be as homogenous as possible and to have a similar population size (Martin 1997; 1998). The target size of an OA is 125 households or approximately 300 people (Martin 1997). The main reason for aggregating population census counts is to reduce data volume and maintain confidentiality and respondent anonymity. In some countries census data are spatially aggregated to very coarse summaries that limit their use in further spatial analysis. For example, in Nigeria, simple population totals are provided for each state and local government area (LGA), with LGA being the most detailed. A LGA is similar to the size of a Unitary Authority (UA) district in the U.K. This level of aggregation makes accessibility studies and many types of spatial data analysis difficult as smaller area population estimates are often required than those provided by the census (Leyk et al. 2013).

Areal interpolation is the process of transforming values of interest from source zones to provide estimates over a set of target zones with unknown values (Goodchild and Lam 1980). A number of areal interpolation techniques have been developed but their performance has been found to depend on specific characteristics of the original data such as its known errors, its extent and its spatial properties (Zandbergen and Ignizio

2010; Wu et al. 2005) as well as the characteristics of any ancillary data used, for example, to constrain the disaggregation (Langford 2013).

One of the simplest areal interpolation techniques is the areal weighting method that maintains total data volume and assumes that population is uniformly distributed within the source zones (Goodchild and Lam 1980). In reality, population distribution is not typically uniform within a source zone and assigning population density to every location is not a realistic representation due to the existence of water bodies, forests etc. where people are unlikely to reside. Point-based areal interpolation methods (Lam 1983) have been used to overcome some of the errors associated with the assumption of uniform densities within source zones. These methods assign census zone populations to the centroid of each source zone, and then population counts are estimated by summing all points within the target zone. The major shortcoming of this method is that the polygon centroid is used to represent the total population within the polygon and by completely allocating (or not) to the target zone estimate the total population depending on its intersection status (Langford and Higgs 2006). This can cause aggregation errors when the centroids are used to measure accessibility to service facilities (Hewko et al. 2002). Tobler (1979) proposed the pycnophylactic interpolation technique to overcome the shortcoming of point based approaches. Pycnophylactic interpolation generates a spatially varying but smooth surface from polygon data, whilst preserving the total data volumes. It assigns a non-zero population density value to every location within the study area. In reality, the areas of interest often have sudden changes in population density that coincide with rivers, roads and other uninhabitable areas. Thus approaches that make use of ancillary data to constrain areas within source zones over which data are disaggregated have been suggested (Langford et al. 1991; Eicher and Brewer 2001; Mennis 2003). One common approach for the interpolation of population data has been to include ancillary data on urban extent to drive this constrained allocation technique.

Remotely sensed data such as aerial photographs have been used by researchers since the 1950s to visually interpret, analyse and estimate population (Green 1956). Lo (2008) describes three main approaches used to visually interpret aerial photographs for population mapping: first, counting individual dwelling units from the photographs; second, extracting the extent of urban settlement; and third, measuring areas of different land use. The increasing ability to process digital images has led to the development of

automated digital image classification based on spectral features of satellite imagery (Lillesand and Kiefer 1987; Lo 2006), and land cover derived from classified satellite imagery has been used by many researchers as ancillary data in spatial interpolation (Langford and Unwin 1994; Mennis 2003; Langford 2013). This is because the remotely sensed imagery provides spatial and spectral information that can be related to different land uses including residential areas.

The dasymetric mapping method is an areal interpolation technique that incorporates additional data sources as control variables (e.g. such as can be derived from remote sensing data) to identify zones having different population densities (Langford and Unwin 1994; Cromley et al. 2011). In this way it is able to constrain the disaggregation of values of interest (e.g. population counts) from source zones to provide estimates over a set of target zones with unknown values (Langford et al. 1991; Langford and Unwin 1994; Mennis 2003, 2009; Eicher and Brewer 2001; Tapp 2010). The binary dasymetric method (Langford and Unwin 1994) which simply divides source zones into populated and unpopulated areas is the most common application of dasymetric mapping for population cross-areal interpolation (Mennis 2003). The binary dasymetric method uses only the populated area to calculate the population density, instead of using the total area of the source zone. This idea has been extended by Su et al. (2010) by further dividing the populated area into multiple classes depending on the availability of additional data. They applied transportation layers, topography and land use zoning to estimate population distribution in Taipei, Taiwan. An evaluation of a 3-class dasymetric model has been carried out (Eicher and Brewer 2001; Langford 2006), and although more complex than the binary dasymetric method, the results do not appear to show any strong benefit over the simple binary dasymetric method. Recent research has shown how improvements in the accuracy of areal interpolation can be advanced by using different statistical methods, varying from simple proportions to more sophisticated procedures (Qiu and Cromley 2013), a quantile regression approach (Cromley et al. 2011) and by utilising different data sources; for example, three dimensional LiDAR data (Sridharan and Qiu 2013), open access vector map data (Langford 2013) and household sample data (Leyk et al. 2013).

It is a common practice to redistribute population census totals from an initial census area, as the source zone (e.g. middle super output area (MSOA) in the U.K.) to smaller

zones such as OA and to compare the result with the actual population counts of the lower level census area (e.g. OA) for validation. Previous research has used multispectral imagery mainly of 30m spatial resolutions to redistribute aggregate census data to a lower level census unit as the target zones for which true populations are known (Mennis 2003; Langford 2006; Lo 2008; Su et al. 2010). Langford (2013) draws attention to the implications of this practice: first, the performance of the most spatially detailed census data are not often measured because they are reserved for testing the performance of the interpolation methods; and second, it is difficult to evaluate the performance of target zones smaller than the lowest level census spatial unit because their true values are not known. He demonstrates the possibility of using unit postcodes (UPCs) in the UK as the target zones with an acceptable precision. The UPCs are smaller than the finest census zone division, the OA in the U.K. The population totals of the UPCs are not reported in the U.K. hierarchy of census units but are known and available at the Office of National Statistics (ONS) U.K. All these examples are drawn from a relatively “data rich” environment with census statistics available at a variety of scales extending down to fine divisions from Government Office Regions (GOR) of England to Unitary Authority districts, Census Wards, MSOA, Lower Super Output Areas (LSOA) and to OA. There remains an important question which is: How can a population surface be determined in areas of unknown distributions and with no validation data? This is an important issue as the spatial distribution of demand population is a critical input to determining demand and to evaluate access distance to facilities in location-allocation modelling. This is the primary motivation for the research reported here: to explore the methods to support location-allocation in Nigeria where census data are aggregated to very coarse summaries that limit their use in further spatial analysis.

## **1.2 Motivation**

Census data at the small area level (e.g. MSOA, LSOA, OA, etc.) are unavailable in most parts of the world, especially in developing countries. In such places the need to estimate aggregate population counts to small areas to represent demand population (Cromley et al. 2012; Tomintz et al. 2013) is important in order to support better spatial planning and policy making. More specifically, calibrated solutions for estimating populations over small areas are needed to support repeatable and transparent facility

location analyses. Dasymetric interpolation using land cover and population data of the area under consideration potentially offers one such approach.

In Nigeria population counts are collected for each household. Population censuses were conducted in 1963, 1973, 1991 and 2006; although it was to be conducted every 10 years (NPC 2009). Census data are published only as spatially aggregated figures for States and LGAs. The aggregation of census data to LGA level, purportedly for confidentiality reasons, poses serious problems for effective spatial planning, such as for healthcare planning objectives. To date there appears to be no published research reporting the use of areal interpolation techniques to estimate aggregate census data over small areas in any part of Nigeria. However, population data for small areas have long been more generally found to provide information on local population characteristics that assist in coordinating, monitoring and evaluating service delivery (Curtis and Taket 1989). Reliable, spatially detailed population estimates are essential to support economic development, management decisions, disaster management, and urban and regional planning (Mennis 2009). Providing information describing the weight and distribution of demand over the region of interest also allows analyses of facility locations and population allocation.

### **1.3 Aim and objectives of the study**

The main aim of this research was to generate population surfaces in areas where their actual distributions are largely unknown. The thesis evaluates dasymetric mapping and pycnophylactic interpolation applied across different spatial scales and using land cover data derived from classified satellite imagery of differing spatial resolutions as the ancillary data input. Initial work explored a U.K. case study to test the performance of the interpolation methods and the inputs in order to determine how well the populations reported in census small areas were estimated by interpolations with different input parameters. The best performing model and parameters developed from the U.K. case study were then adapted and applied to the Nigerian case study to estimate summary population totals over small areas of unknown distributions. The demand population established in this way was subsequently used to provide inputs to a location-allocation model in order to evaluate the current distribution of health facilities and to assess the

efficiency of their current locations and possible future spatial optimisation of these facilities.

This research has three key aims:

- To estimate population surfaces across different spatial scales in areas where actual distributions are unknown.
- To transfer the ‘best’ solution found during this phase so as to create population surfaces and use these surfaces to evaluate current health facility locations.
- To suggest alternative spatial arrangements of health facilities so as to improve their overall spatial accessibility.

The specific objectives are to:

- Apply dasymetric and pycnophylactic interpolation approaches across different spatial scales to redistribute aggregate population census data for Leicester over small areas.
- Compare the estimated populations from the interpolations for three different census units and the known census counts in each case, to test the performance of the interpolation methods.
- Adapt the model with the most appropriate interpolation method, grid size and ancillary data input to estimate demand population in Port-Harcourt, Nigeria.
- Evaluate the public health facility locations currently in place in Port-Harcourt.
- Suggest alternative spatial arrangement of public health facilities using heuristic location-allocation modelling approaches.

#### **1.4 Research questions**

The following questions were identified in order to achieve the overall objectives of this research.

- Broad research question:  
How can estimates of the population for small areas be determined where detailed local mapping of census do not exist?
- Specific research questions:

- I. What is the relationship between estimated populations from different interpolations and the known census counts?
- II. Which is the most appropriate interpolation method, scale of disaggregation, and resolution of ancillary input data to apply to areas with no validation data?
- III. How well do current public health facilities serve the current population distribution?
- IV. What improvements in accessibility arise when public health facilities are optimally located?

## **1.5 Thesis Structure**

The thesis comprises of eight chapters beginning with an introductory chapter, a general discussion and a conclusion chapter at the end. The main body of the thesis is divided into five chapters. This section provides a brief description of these chapters as follows:

- Chapter One presents a brief introduction to the thesis providing background to small area population estimation and spells out the problem area, objectives, research questions and the motivation for the study.
- Chapter Two reviews literature on areal interpolation techniques, GIS and geographical analyses of access to health care and location-allocation models. The literature review forms the framework against which the analyses in subsequent chapters are undertaken.
- Chapter Three presents the methodology. It describes the study area, the data and the implementation of areal interpolation methods. The chapter also highlights the evaluation of surfaces.
- Chapter Four presents the results of areal interpolation for Leicester, the surfaces generated and the results from evaluation of surfaces. It highlights the most appropriate parameters from the results of the analyses in Leicester to be used to standardise the model for the analysis in Port-Harcourt.
- Chapter Five presents the results of areal interpolation for Port-Harcourt, the inspection of surfaces generated using Google Earth and a brief description of what the errors from Leicester study mean for surfaces generated in Port-Harcourt.

- Chapter Six presents case study analyses, methods and findings on location-allocation in Port-Harcourt. The current public health facilities in Port-Harcourt were evaluated and alternative spatial arrangement of public health facilities were suggested.
- Chapter Seven links the objectives with the key findings from the research to present a general discussion of the demand population values generated for Leicester and Port-Harcourt, the methods used and their assumptions. The chapter also reflects on the methods used and the result obtained, discusses limitations of the research and suggests areas for future work.
- Chapter Eight links the findings from the literature, interpolations and the case study to present the conclusions from this research.

## Chapter 2

### 2. LITERATURE REVIEW

#### 2.1 Introduction

In this chapter, the literature on the concepts and methods relating to the main aims of the research are reviewed. The objective is to review different techniques for areal interpolation and some of the more widely used location-allocation models.

Section 2.2 reviews areal interpolation techniques. Section 2.3 considers the notion of geographic access measure more generally and in the specific context of access to health care, and considering facility location in health planning. Section 2.4 introduces location-allocation models and describes the  $p$ -median problem. The last section presents a summary of the review.

#### 2.2 Areal Interpolation Techniques

Areal interpolation is the process of spatially disaggregating attributes (such as summary of counts) of some phenomenon such as population across incongruent boundaries where areas may be smaller, almost the same or bigger. Values are estimated for target zones (small areas) from source zones over which the data are summarised (Markoff and Shapiro 1973). A number of areal interpolation techniques have been proposed and used in order to improve estimation accuracy and to provide spatially distributed estimates of population over small areas (e.g. Markoff and Shapiro 1973; Tobler 1979; Goodchild and Lam 1980; Lam 1983; Flowerdew and Green 1991; Langford et al. 1991; Goodchild et al. 1993; Burrough and McDonnell 1998; Eicher and Brewer 2001; Mennis 2003; Cromley et al 2011; Langford 2013; Schroeder and Van Riper 2013). One reason for this is that many problems relating to spatially disaggregating attributes of some phenomenon largely depend on the spatial distribution of demand (Cromley et al. 2011; Tomintz et al. 2013). For these reasons, some of the proposed techniques use ancillary masks to spatially constrain the re-allocation (e.g.

Langford et al. 1991; Eicher and Brewer 2001) while others do not (e.g. Goodchild and Lam 1980; Tobler 1979).

Table 2.1 shows a selection of previous literature reporting different areal interpolation methods. It illustrates the diversity and the duration of areal interpolation as a research topic. The table compares the characteristics of thirteen methods used by different authors in terms of assumptions made about their distribution of population, their use of ancillary data (or not) and their functions. The contents presented in the table are arranged alphabetically by the underlying methodology, starting first with methods that do not use ancillary data, then the intelligent methods. From Table 2.1, the distribution of population refers to assumptions made by each method whether the population is homogeneously or heterogeneously distributed within the source zone. The use of ancillary data characteristic describes the type of ancillary data used (or none) by each method to constrain the re-allocation. The Functions column describes the major spatial operation(s) used by each method to estimate population. The Authors column lists authors that have used each method.

The review of areal interpolation methods focuses on the binary dasymetric method (Langford and Unwin 1994) and the pycnophylactic interpolation technique (Tobler 1979). A combination of these two methods have been shown to produce better results than that obtained by either of the methods individually (Comber et al. 2008b; Kim and Yao 2010; Kim and Choi 2011) but it has not been tested in this research. This is because the primary purpose of this research is to evaluate dasymetric and pycnophylactic interpolations across different spatial scales. These techniques were chosen because of the variable of interest (population), data available and prior knowledge of the study areas. The basic difference between the two techniques is that the pycnophylactic interpolations do not make use of ancillary data and generate a smooth surface, while binary dasymetric surfaces use ancillary masks to constrain the re-allocation to only areas identified as populated and produces a non-smooth spatially discontinuous surface with sharp density transitions.

Table 2.1 Characteristics of areal interpolation methods

S/NO	Method	Authors	Ancillary data	Distribution of population	Functions
1	Areal Weighting	Goodchild & Lam, 1980; Poulsen & Kennedy 2004; Cromley et al. 2009	None	Homogeneous	Spatial Overlay
2	Point in Polygon	Okabe & Sadahiro, 1997	None	Homogeneous	Point- in- Polygon
3	Pycnophylactic Interpolation	Tobler 1979; Rase 2001; Comber et al. 2008; Kim & Yao 2010	None	Heterogeneous	Smoothing, Zoned Statistics
4	Address Weighting (AW) & Parcel Distribution (PD)	Harris and Longley 2000; Tap 2010; Zhang and Qiu 2011	Address points & cadastral data (parcels)	Heterogeneous	Algorithm
5	Binary Dasymetric Method	Langford & Unwin 1994; Eicher & Brewer 2001; Mennis 2003	Land use/ Land cover	Heterogeneous	Spatial Overlay, Classification
6	Cadastral-Based Expert Dasymetric System (CEDS)	Maantay et al. 2007; Bentley et al. 2013	Cadastral data	Heterogeneous	Spatial Overlay, Classification, Expert allocation system
7	Control Zones	Goodchild et al. 1993	User defined zones	Heterogeneous	Digitizing, Spatial Overlay
8	Expectation Maximisation (EM)	Flowerdew & Green 1989; Flowerdew & Green 1991; Dempster et al. 1977; Schroeder and Van Riper 2013	Various variables	Heterogeneous	Statistical Algorithm
9	Heuristic Sampling	Mennis & Hultgren 2006; Sleeter & Gould 2007	Land cover/ Land use	Heterogeneous	Classification, Empirical sampling
10	Neural Networks	Turner & Openshaw 2001	Various variables such as distance, elevation etc.	Heterogeneous	Heuristic Algorithm
11	Regression Analysis	Langford et al. 1991; Yuan et al. 1997; Cromley et al. 2011.	Land use/ Land cover	Heterogeneous	Regression, Spatial Overlay, Classification
12	Road Network Method	Xie 1995; Mrozinski & Cromley 1999; Reibel & Bufalino 2005	Road network data	Heterogeneous	Spatial Overlay
13	Smart Interpolation Method	Diechmann 1996	Location of rivers, transport structures etc.	Heterogeneous	Heuristic Algorithm

### 2.2.1 The Pycnophylactic Interpolation Technique

Tobler (1979) proposed the pycnophylactic interpolation technique. The objective of the technique was to create a smooth surface of interpolated values from polygon data (choropleth to isopleth) with no sudden change at the polygon boundaries and to preserve the total source volume. That is, the total value of target zones equal that for source zones. The term “pycnophylactic” is derived from Greek *pyknos* for mass, density and *phylax* for guard, which means volume preserving (Rase 2001). The technique assumes heterogeneous population distribution within the source zone but does not draw on information about the underlying population distribution in the source zone, thereby assigning a non-zero population density value to every location (Kim and Yao 2010; Amaral et al. 2012). The lack of information about source zone internal structure is a major drawback as in reality some places do have abrupt barriers in the landscape such as elevated highways, rivers etc. (Reibel 2007). Such areas are uninhabitable and should have zero population density value. The pycnophylactic interpolation technique will be described formulaically in the methods chapter.

The pycnophylactic interpolation technique has been used to improve the population density distribution of Seoul (Chang 2003). The results showed that population of Seoul were assigned to urban areas instead of residential areas. The assignment of population to urban areas was due to lack of information about source zone internal structure as the technique does not make use of ancillary data to spatially constrain the re-allocation. Some authors have attempted to improve the pycnophylactic interpolation technique by using attribute information of the population weighted centroid associated with each census tract (e.g. Martin 1989; Bracken and Martin 1989) to generate surfaces from points, and not areas. This surface generation technique was applied by Bracken and Martin (1989) to create spatial population distributions from census centroid data in South Wales, U.K. using the summary of the intra zone distribution of the population and the spatial configuration of centroid locations to generate a population surface independent of a particular zonal geography. Bracken and Martin found the method to identify populated areas based on population weighted centroid and to preserve sharp gradients at the source zones boundaries. The method also revealed the presence of unpopulated areas based on distance decay from population weighted centroid. Similarly, Harris and Longley (2000) created a discrete surface from points using

multiple population points based on address code records for each source zone instead of using population-weighted centroids. Zhang and Qiu (2011) suggest the use of high value points (school locations) that are assumed to provide a reasonable alternative for population-weighted centroids within the source zones. The school locations were used as control points in Collin County, Texas and the results show the method achieved a comparable accuracy with street weighting method. The major shortcoming of creating surfaces from points is the assumption about the nature of the population weighted centroid and lack of actual evidence of distribution (Langford and Higgs 2006). This causes aggregation error when the centroids are used to measure accessibility to service facilities (Hewko et al. 2002).

The pycnophylactic interpolation technique was enhanced to model the distribution of population in Germany for regional planning (Rase 2001). Rase used triangulated irregular networks (TIN) instead of regular grids to improve visualisation and to maintain the original boundaries of the polygons. Another advantage of using a TIN over a traditional rectangular grid is that it uses boundary points as vertices in TIN model. The use of TIN is seen to be better than using regular grids as it allows the preservation of the original data points in the model of a surface (Peucker et al. 1978; Rase 2001). The disadvantage of the TIN model is that it requires additional processing time and large computer storage. Other example applications of the pycnophylactic technique include: point in polygon analysis (Okabe and Sadahiro 1997) and a geostatistical method of kriging (Kyriakidis 2004).

Some authors have considered the limitations of pycnophylactic interpolation and combined the technique with the dasymetric method in order to draw the strengths and improve on the weaknesses of both techniques, which are perfectly complementary to each other (Kim and Yao 2010; Comber et al. 2008b). Kim and Yao (2010) developed a hybrid method that combines pycnophylactic interpolation with dasymetric method and compared the estimation accuracy of the hybrid method with areal weighting, binary dasymetric and the pycnophylactic interpolation methods. They found the combined technique to perform better than all others tested. Comber et al. (2008b) considered the benefits of pycnophylactic interpolation and combined the technique with dasymetric method to create the national agricultural land use dataset for England and Wales at

1km resolution. They found the technique to overcome differences in areal reporting units such as Parish (or Parish Groups), District, County etc.

Many researchers acknowledged Tobler's smoothing algorithm is cartographically suitable but not appropriate for advanced spatial analysis (Yuan et al. 1997; Eicher and Brewer 2001; Mennis 2003; Langford 2007; Tapp 2010). This is because the technique does not make use of ancillary information to reveal the underlying distribution of population in the source zones. It is simply a smoothed representation of a conventional choropleth map (Kim and Yao 2010). The concept of volume preserving remains very important in recent population estimation methods so as people are neither created nor destroyed. In view of this, further ancillary information is required to understand how population is distributed within the source zone to improve the interpolation process.

### **2.2.2 The Dasymetric Mapping Method**

The history of dasymetric map dates back to the year 1833. The concept of dasymetric map were proposed by the Russian geographer Benjamin (Veniamin Petrovich) Semenov-Tian-Shansky (1870–1942) in his report to the Russian Geographic Society in 1911 where he defined dasymetric map as *“population density, irrespective of any administrative boundaries, is shown as it is distributed in reality, that is, by natural spots of concentration and rarefaction”* (Petrov 2008, 134). The term “dasymetric” is derived from Greek words for density and measurement. This indicates the technique focused on the density within the underlying (population) surface, which can be recalculated using ancillary data that provide relevant additional information (Leyk et al. 2013). The dasymetric mapping technique when applied to cross-area population modelling makes use of ancillary information (e.g. remote sensing data) to infer underlying population distribution (Wright 1936) and provides a clear understanding of population distribution within a source unit (Langford and Unwin 1994; Langford and Higgs 2006).

The basic principle of dasymetric mapping is to use additional control variables to identify zones having different population densities (Wright 1936). Many authors have used different types of control variables such as remotely sensed land cover data (e.g. Langford and Unwin 1994; Eicher and Brewer 2001; Mennis 2003), road network data

(e.g. Xie 1995; Mrozinski and Cromley 1999), cadastral data (e.g. Maantay et al. 2007), address points (e.g. Zandbergen and Ignizio 2010), and parcel data (e.g. Tapp 2010). The advantage of the technique is that it eliminates sharp differences at source zone boundaries and reduces errors of within-zone uniformity (Langford et al. 2008). In terms of population estimation, the model gives better information about the distribution of population (Cai et al. 2006). A possible limitation of dasymetric mapping is where no part of the source zone is identified as populated due to classification error (Martin et al. 2000). This means population will not be assigned to that source zone or could be uniform in this case. The dasymetric mapping technique has been applied in different research areas such as environmental justice (Higgs and Langford 2009), public service accessibility (Langford et al. 2008), environmental health (Maantay et al. 2008), creation of National Agricultural Land Use Dataset (Comber et al. 2008b) and crime analysis (Poulsen and Kennedy 2004; Herrmann and Maroko 2006). The dasymetric mapping method will be described formulaically in the methods chapter.

The ancillary data most commonly used in dasymetric mapping research to date has been land cover information derived from classified satellite imagery (Langford 2013; Kim and Yao 2010), and such data commonly have errors. Fisher and Langford (1996) analysed the sensitivity of the population estimates to error in the classified imagery with the assumption that the error was randomly spatially distributed and not related to specific classes. They found dasymetric mapping to be simple and robust to classification error. The technique when compared with four others provided better areal interpolation at classification error rates as high as 40%.

A review of dasymetric mapping literature shows the development of several techniques using different types of ancillary data all with the aim of improving accuracy and possibly simplifying the method (Mennis 2003). The diversity of approaches is as a result of the absence of a universally accepted dasymetric mapping technique (Mennis 2009). Some of the commonly used ancillary data are: topographic sheet (Wright 1936), aerial photographs (Green 1956), satellite imagery (Langford and Unwin 1994; Wu et al. 2005), land cover (Mennis 2003; Ryznar and Wagner 2001; Ward et al. 2000), image pixels (Harvey 2002; Holt et al. 2004), image texture (Chen 2002; Liu et al. 2006), slope (Schumacher et al. 2000), raster pixel maps (Langford 2007), vector GIS (Eicher and Brewer 2001), control zone (Goodchild et al. 1993), night-time lights (Pozzi et al.

2003; Briggs et al. 2007), cadastral data (Maantay et al. 2007; Bentley et al. 2013), housing distribution (Moon and Farmer 2001; Poulsen and Kennedy 2004; Leyk et al. 2013), parcel data (Tapp 2010), OS VectorMap District (Langford 2013) and three dimensional LiDAR data (Sridharan and Qiu 2013).

The most common approach to implementing dasymetric mapping is via the binary dasymetric method (Langford and Unwin 1994). This divides the source zone into populated and unpopulated regions, and the source zone population is then allocated to only the populated regions (Eicher and Brewer 2001; Mennis 2003; Langford 2007). However, the binary dasymetric approach still assumes that population density is uniformly distributed across the populated regions, within a source zone which is not likely to be true (Mennis and Hultgren 2006; Maantay et al. 2007). For this reason, a three-class dasymetric model has been evaluated by assigning population densities subjectively (Eicher and Brewer 2001) or by using a heuristic method based on empirical sampling (Langford 2006), but the results tested so far do not appear to show any significant benefits in accuracy over the simpler binary dasymetric method.

Some authors have shown dasymetric mapping to consistently provide better target zone estimates than other areal interpolation techniques when they are compared (Langford 2013; Poulsen and Kennedy 2004; Eicher and Brewer 2001; Martin et al. 2000; Cockings et al. 1997; Fisher and Langford 1995; Langford et al. 1993). Poulsen and Kennedy (2004) applied areal weighting and binary dasymetric to show the distribution of residential burglaries in central Massachusetts with the rate expressed as the number of residential burglaries per number of housing units. They found the results obtained from areal weighting method to be misleading since the housing units in the census are not evenly distributed. They applied dasymetric mapping and masked out non-residential areas using a residential land use layer as the ancillary input data. This gave better result when compared with those obtained from areal weighting. Similarly, Eicher and Brewer (2001) applied five areal interpolation techniques to model population densities across 159 counties in Pennsylvania, West Virginia, Maryland, Virginia, and the District of Columbia. They used U.S. County zones as the source units and U.S. block group as the target unit to test the model. They found the binary dasymetric method to be the most straightforward and assigned 100% of the population to only urban and agricultural land use types. Langford (2013) investigated the accuracy of

different techniques of areal interpolation in redistributing aggregated census data to areas smaller than the finest census division (UK postcode units). The study compared nine different interpolation methods and found the dasymetric mapping method to perform better than street weighting, a population density surface method and areal weighting.

Other researchers have used the dasymetric method to mask out areas of environmental restriction to human presence in order to generate a potential surface of population occurrence in the municipality of Maraba, Para State, in the Brazilian Amazon (Amaral et al. 2012). The aggregated population counts were then redistributed into cells. They found that the spatial patterns were compatible with the occupation process described in the literature and verified by field work. The possibility of redistributing aggregate census data to target zones smaller than the finest census zone division (OA in the U.K.) has been demonstrated (Langford 2013). Langford applied areal interpolation across two spatial resolutions. First, using lower super output areas (LSOAs) as the source zones and output areas (OAs) as the target zones. Second, output areas were used as the source zones and unit postcodes (UPCs) as the target zones. The UPCs are smaller than the OAs. The population totals of the UPCs are not reported in the U.K. hierarchy of census units but are known and are available at the Office of National Statistics (ONS) U.K. The results show that aggregate census data at OAs were successfully redistributed to UPCs with an acceptable precision. Previous research in areal interpolation literature used target zones whose actual values are known and are reported (see Table 2.2). This was done to test the performance of the interpolation techniques and to allow comparison of different interpolation techniques (Langford 2013). The implication of this is that small areas were referred to as not only “geographic target units” but those areas whose actual population is known.

Table 2.2 Target zones used in previous literature whose population is known

Author/year	Study area	Title	Source zone	Target zone
Fisher & Langford 1995	Leicestershire County, UK	Modelling areal interpolation errors	U.K. ward	U.K. ED
Eicher and Brewer 2001	Four U.S. States	Areal interpolation using dasymetric mapping	U.S. county	U.S. block group
Mennis 2003	Southeastern Pennsylvania, US	Population modelling using dasymetric mapping	U.S. census tract	U.S. block group
C.P. Lo 2008	Atlanta, Georgia, US	Population estimation using GWR	U.S. census tract	U.S. census tract
Su et al. 2010	Taipei, Taiwan	Multilayer/multiclass dasymetric mapping	County	Chinese Li
Langford 2013	Cardiff, South Wales, UK	Small area population using open access data	LSOA/ OA	OA/UPC
Sridharan and Qiu 2013	Round Rock, Texas	Areal interpolation using LiDAR	U.S. block group	U.S. block group

### 2.3 Influence of demand population on spatial accessibility

Small area population estimates are important for accessibility studies that are commonly used to support spatial planning and policy development, and facility location-allocation analyses. In some countries (e.g. Nigeria) where small area population estimates are often required than those provided by the census, aggregate population census totals are redistributed to smaller areas to represent demand population values. The small area population estimates generated largely depend on the underlying population distribution model used.

Previous studies have examined the influence of alternative population distribution models on GIS-based spatial accessibility analyses using the two-step floating catchment analysis technique (Langford et al. 2008). They applied areal weighting and dasymetric mapping to redistribute aggregate population for the city of Cardiff UA in South Wales, U.K. The population estimates were used to represent spatially distributed

demand values to evaluate accessibility to five services (General Practice, Dentist, Primary School, Post Office and Pharmacy). The study found that the underlying population distribution model used in generating demand surfaces has significant influence on the computed accessibility scores to such services. Similarly, Langford and Higgs (2006) used population estimates from dasymetric mapping to represent demand. They applied the two-step floating catchment technique to measure potential access to primary health services in three unitary authorities within Wales, U.K. These results showed that the dasymetric model returned lower accessibility scores than those from a standard pro rata model. Other authors have also used population estimates from dasymetric model to evaluate representation and scale error in the maximal covering location problem (Cromley et al. 2012). They found a significant difference when compared with estimates from point-based and areal weighting methods on a small scale data. This indicates an improvement in the spatial distribution of demand, thereby suggesting maximal coverage for all demand points.

## **2.4 GIS and Geographical Access to Healthcare**

Geographical accessibility to health facilities is the relative ease with which individuals may travel a reasonable distance from their homes to the nearest health facility (Guagliardo 2004; Owen et al. 2010). Presently, health care needs of the population and how these needs can be met equitably are of particular interest to policy makers and health system planners (Cromley and McLafferty 2002). Researchers all over the world are concerned with “*the relationship between territory and health*” (Benigeri 2007, p.3) and Geographical Information Systems (GIS) is a diverse, powerful and effective planning tool that is increasingly being used in providing such information (Higgs 2005; Carr and Addyson 2010). Health system planners require evidence to support unbiased decision making in resource location-allocation and the analyses of access to health facilities using GIS may help to provide this.

Some of the recent applications of GIS in health planning include; the analysis of access to Multiple Sclerosis (MS) specialty care in Veterans Health Administration (Culpepper et al. 2010). Highly developed GIS mapping techniques were used with an existing database to generate travel times and to develop an approach to assess irregularities in access to specialty care. This is useful in identifying underserved areas for all conditions

and diseases affecting the population. Comber et al. (2010) applied a modified grouping genetic algorithm to existing emergency medical services (EMS) data to select alternative sites for locating ambulances in Niigata, north-western Japan. The result is likely to improve average EMS response time by 1 min 14 s. The importance of this cannot be overemphasized as ambulance response time is a critical determinant for patients' continued existence and the technique provides strong evidence in support of spatial planning. Dulin et al. (2010) applied GIS techniques to available patient data of a community health centre to create a model that demonstrates the health care needs of the community and could also predict its future needs. The technique is called Multiple Attribute Primary Care Targeting Strategy. This is very useful for implementing changes and planning in order to improve access to health services in the community. The impact of community-based outreach immunisation services in Zambia were analysed and optimal locations of the Growth Monitoring Program Plus (service points) were identified using GIS (Sasaki et al. 2010). The result shows that access distance to service points were reduced by about 30 per cent, with more people having access to the service points thereby increasing immunisation coverage within the communities. This provides strong evidence in support of spatial planning.

#### **2.4.1 Review of spatial access**

In terms of health care provision, GIS allow “*access distance to health services to be measured and provides a practical way to assess the geographic accessibility of said services*” (Sasaki et al. 2010, p.1). In recent times, researchers have used GIS to measure access to services in a number of ways: Euclidean distance approach (McGregor et al. 2005; Apparicio et al. 2008; Jones et al. 2010), cost-path analysis (Brabyn and Skelly 2002), travel time (Martin et al. 2002; Delamater et al. 2012), isochrones approach (Tanser et al. 2006; O’Sullivan et al. 2000), standard deviational ellipse (Sherman et al. 2005; Nemet and Bailey 2000), floating catchment method (Luo 2004; Langford and Higgs 2006; McGrail and Humphreys 2009; McGrail 2012), Gravity model (Rosero-Bixby 2004; Guagliardo 2004), Space-time prism (Lee and McNally 2002), Spatial Analysis Network Tools (SANET) (Okabe et al. 2006), and GIS network analysis method (Sasaki et al. 2010).

Some authors used the Euclidean distance approach to measure straight line distances (as the crow flies) from home to health facility (McGregor et al. 2005). The disadvantage of using this technique is that it does not consider topology and assumes people only use facilities that are closest to their homes. Other authors have developed an application that built isochrones (lines of equal travel time) for travel using different modes of public transport in Glasgow (O'Sullivan et al. 2000). There are other authors that have applied a space-time prism approach to measuring access to health facilities in Portland, Oregon (Lee and McNally 2002) and highlighted its usefulness in measuring access at both unit and zonal scales. Brabyn and Skelly (2002) in their cost-path analysis used street network distances to consider routes that are more likely to be used from homes to health facility. Some other authors enhanced the accuracy of distances measured on street network by including travel times (Martin et al. 2002). The travel time on a network is the time it takes to travel over a distance between two points on the network. There are other authors who considered individual's activity space to measure access to primary health care in Carolina using a 1-km road network buffer (Sherman et al. 2005). A buffer is a circle drawn around a facility with population closest to the facility weighing more than the distant ones. The major limitation of the technique is that it does not account for travel routes (roads) and assumes the population within the circle all have the same access to the facility.

Some authors have proposed spatial analysis network tools (SANET) which denotes real world events by using points on a network instead of considering events as a set of points on a plane (Okabe et al. 2006). The technique involves determining supply and demand and is likely to address questions relating to linear networks, utilities and facilities, finding the nearest facility, estimating distance to service areas, route planning etc. (Comber et al. 2008a). Gibson et al. (2011) applied a GIS network analysis of shortest travel distance to calculate distance travelled along the road network to find the most distant households from health centre in rural China. They mentioned the technique offers a more reasonable estimate of accessibility than other methods that disregard the terrain. Sasaki et al. (2010) studied the relationship between immunisation coverage and distances to immunisation service points in a peri-urban area in Zambia. They used a GIS network analysis method to measure distances from households to George health centre and the nearest Growth Monitoring Program Plus sites and found

that residents are less likely to access and utilise health facilities as their access distance to the facilities increases.

These studies have demonstrated the utility of GIS and why it is being increasingly used to address key policy concerns such as health inequalities, accessibility and resource allocation. It also highlights the role GIS techniques can play in providing information needed to ensure health services are delivered closer to where people live, and at the same time to provide evidence in support of spatial planning. The flexibility and ease of obtaining information from GIS have shown health system planners' interest in using GIS research findings as evidence to support their decisions on resource location and allocation. Optimal locations of health facilities is the key feature in healthcare planning and location-allocation models assist health system planners in determining and assigning users to the new facility (Bowerman et al. 1999; Cromley and McLafferty 2002).

In summary, this review shows most studies to date in Health Care planning have used planar based approaches to measure accessibility, but analysis of access to health services requires a more realistic estimate of accessibility and GIS network analysis technique provides this (Gibson et al. 2011). The measure of the shortest distance travelled on road network between each demand and each facility is used as an input into location-allocation models together with the demand surfaces generated from the areal interpolations of population data and facility locations to evaluate location decisions. The review also highlights the importance of facility location and allocation of resources in determining geographical accessibility to health services.

## **2.5 Location-allocation models**

The location-allocation problem was first introduced by Cooper (1963). Location-allocation models are spatial analyses model that determine optimal facility locations based on set criteria by minimising the total average distance travelled from demand to all facility locations, or by increasing demand coverage (Hakimi 1964; Teitz and Bart 1968). These models are used in GIS to select the best possible location for a facility to best serve a demand area, or at the same time, to allocate demand that the facility will serve best (Abernathy and Hershey 1972; Oda and Yamamura 1987; Bashiri and Fotuhi

2009). The main objective of location-allocation models is to determine the optimal locations of facilities (Kumar 2000). They require three critical inputs: spatially distributed demand surfaces, facility locations to serve as the supply points, and shortest network distance travelled between demand and supply. A number of location-allocation models have been applied to various fields of study (e.g. GIS and spatial planning, health facilities, engineering, industry etc.) and are also implemented with various constraints, such as minimum distance, maximum coverage etc. (Jia et al. 2007).

Some authors classify location-allocation models based on certain criteria, the most common of these are: the objective function of the model (Rosing et al. 1979; Chen and Handler 1993), input parameters used (Marianov and ReVelle 1992), methods used in obtaining results (Brandeau and Larson 1986), topological characteristics (Campbell 1994; Daskin 1995), features of the facilities (Pirkul and Schilling 1991), locating facilities at different time periods (Erlenkotter 1981), type of supply chain (Gao and Robinson 1994), demand patterns (Plastria 1997) and single/multi product model or push/pull models (Klose and Drexler 2004). Jia et al. (2007) considers the objective function and classifies location models into three types namely; p-median, p-centre and Covering models. The p-median approach aims to minimise total weighted distances (Teitz and Bart 1968). The covering models maximises demand coverage (Toregas et al. 1971; Church and ReVelle 1974). The p-centre approach attempts to find the smallest radius that covers all points on a circle (Sylvester 1857).

### **2.5.1 The p-Median problem**

The p-median problem was introduced by Hakimi (1964) with an objective function of selecting  $p$  facilities (among the total facilities) that minimises the total weighted distance travelled (or time) between facilities and demand points (Teitz and Bart 1968; ReVelle and Swain 1970), thereby maximising accessibility. There are two basic approaches to solve the p-median problem: optimal and heuristic (Church and Sorensen 1994). The most robust of the optimisation techniques is the Lagrangian Relaxation with sub-gradient optimisation developed by Narula et al. (1977) but this technique requires too much computation time to solve (Church and Sorensen 1994). A number of heuristic procedures have been developed to help solve the p-median problem. They include: Maranzana (1964), Teitz and Bart (1968), Genetic Algorithm (Holland 1975),

Simulated Annealing, TABU search (Glover 1990), GRASP and a combination of two or more procedures to form the hybrid approach (Church and Sorensen 1994). The general principles of these heuristic procedures are discussed below.

A warehouse location heuristic was developed by Maranzana (1964), as an alternative strategy for the  $p$ -median problem. The Maranzana heuristic starts with a configuration of  $p$  facility locations, selecting a subset ( $p$ ) of, for example 20 out of 85 potential locations. The heuristic first assign each demand to the closest facility location creating service areas around each facility. It then takes each service area separately and relocating the facility within that service area to the place that best serves the demand. The heuristic repeats these two steps until there is no change in the service areas and the locations of facilities within each service area.

An alternative approach to solve the  $p$ -median problem by adding an interchange heuristic to the model was suggested by Teitz and Bart (1968). The interchange heuristic controls the selection of locations that are more likely to reduce the average weighted distance (or time) from demand to all locations. The process starts with a configuration of  $p$  facility locations. The selected locations in the subset are called the  $p$ -facility locations while the potential locations not in the subset are the candidates. The process selects a candidate site and swaps this candidate for each of the current  $p$ -facility locations. Any swap that improves weighted distance, replaces the  $p$ -facility location. The process then continues by selecting another candidate site, testing swaps and replacing the  $p$ -facility location where there is an improvement in weighted distance. The heuristic stops when no swap between candidate and  $p$ -facility location improves weighted distance.

Genetic Algorithms are heuristic search and optimisation algorithms, introduced by Holland (1975) that “simulates the process of genetic mutation and selection in biological evolution” (Sasaki et al. 2010). The optimisation process proceeds by first initialising a population of chromosomes. These chromosomes represent individual sites in a multi-site optimisation (Comber et al. 2010). Then evaluating each chromosome in the population based on some fitness criteria to assess its performance. The criteria involve creating new chromosomes by mating chromosomes in the current population, using crossover and mutation. Then deleting members of the existing population to

make way for new members and evaluating the new members to add them to the population. The evaluation stages are repeated until some convergence criteria are met or usually based on the number of populations produced. The heuristic algorithm then returns the best chromosome as the solution.

Tabu search is an advanced interchange heuristic that has the capability to arrive at a local optimum, back away from a local optimum, and attempt to re-optimize using interchange (Glover 1986, 1990). The process can use both short and long term memory. It uses short term memory (TABU List) to manage interplay between what enters and exits the tabu list and when. It also uses long term memory to help in identifying either productive or untried areas for search. Tabu search algorithm starts by choosing an initial solution (Tabu list). It then creates a candidate list of solutions. The algorithm evaluates solutions and chooses the best admissible solution. If the stopping conditions are satisfied then it is a final solution, else, it will update the tabu list as a new initial solution and create a new candidate list of solutions.

The first Simulated Annealing algorithm was developed by Metropolis et al. (1953). The algorithm begins with creating an initial random placement (location). It then makes a defined move to modify the initial placement and create a new placement. It evaluates placements by calculating the change in placement due to creation of new placement. Depending on the change in placement, the algorithm either accepts or rejects the new placement. It then updates the acceptance or rejection and repeat by making another new placement. The process is done until there is no change in placement with a new placement.

Some authors have used the p-median method of location-allocation to measure the efficiency of a facility location (Rahman and Smith 2000; Jia et al. 2007). They found an increase in the efficiency of a facility location when the total average distance from demand to all facilities reduces. This is often used in research relating to health geography to select optimal locations of emergency medical services (EMS) and other health facilities (Tomintz et al. 2013). The p-median model and all algorithms used to solve the p-median problem assume that minimising the total distance between demand and supply is more likely to provide optimal accessibility to facility locations. The p-median problem also assumes the facilities to be located have the same capacity and

provide the same services. This is the major shortcoming of the p-median problem. In situations where the hierarchical nature of the facilities to be located is incorporated in the model, it was found to produce a less appropriate result (Hodgson 1990). This is because the assumption made by the model does not fit the facilities to be located. The advantage of the p-median problem is that it analyses the network as it minimises the total weighted distance (ReVelle and Eiselt 2005). The p-median problem was first described mathematically by ReVelle and Swain (1970) as in Equation 2.1 subject to three conditions outlined in Equation 2.2 to 2.4:

$$\text{Minimise } Z \quad \sum_{i=1}^n \sum_{j=1}^n a_i d_{ij} x_{ij} \quad [2.1]$$

Where  $Z$  is the objective function,  $i$  is a set of demand areas (residential areas),  $j$  is a set of supply areas (health facility locations),  $d_{ij}$  is the shortest distance travelled between demand and supply,  $a_i$  is the volume of demand (the population),  $n$  is the number of points (demand/supply),  $P$  is the number of facilities to be located, and  $x_{ij}$  equals 1 if demand  $i$  is assigned to facility  $j$ , otherwise 0.

Each demand point must be assigned to at least one facility as in Equation 2.2. This means that each demand point is assigned to its closest facility location, thereby creating service areas around each facility location to suggest a catchment for the facilities and indicate the likely geographical coverage for each facility.

$$\sum_{j=1}^n x_{ij} = 1 \quad \text{for all } i \quad [2.2]$$

Only the  $p$  facilities are to be located as in Equation 2.3. This means out of the total number of potential locations (e.g. 85), only a subset (e.g. 32 optimal sites) is to be located. Also, assignment can be made only to a facility as in Equation 2.4. This means a demand can only be allocated to its closest facility.

$$\sum_{j=1}^n x_{jj} = P \quad [2.3]$$

$$x_{ij} \leq x_{jj} ; x_{ij} = 0, 1 \text{ for all } i, j \quad [2.4]$$

Some authors compared Teitz and Bart's heuristic with optimal solutions derived by the linear searches (Rosing et al. 1979). They found Teitz and Bart's (1968) heuristic to perform better than linear searches when there are large numbers of facilities to locate. Other authors compared Teitz and Bart's heuristic with Ardalán's p-median problem, with and without maximum distance constraints (Rahman and Smith 1991). They found Teitz and Bart's heuristic to perform better than Ardalán's p-median. The advantage of using Teitz and Bart over Ardalán's method is that the latter seeks to minimise weighted distances but does not substitute an existing facility by a potential facility as it is done by Teitz and Bart's heuristic. Another shortcoming of the Ardalán's method is that it "*does not guarantee optimality at termination*" (Rahman and Smith 1991). Teitz and Bart heuristic is seen to be superior to other heuristic searches because it is insensitive to starting configuration as its application for specific problems need to be started only once (Rosing et al. 1979; Fotheringham et al. 1993).

The choice of heuristic procedure to solve the p-median problem in a GIS system must be based on its robustness, speed, simplicity and ease of integrating into existing data structure and software. Teitz and Bart's (1968) heuristic is a "*proven approach, easy to program, relatively fast, easy to explain, and produces good results*" (Church and Sorensen 1994).

## 2.6 Summary

The key points arising from this review are that the pycnophylactic interpolation technique assumes a heterogeneous population distribution and assigns a non-zero population density value to every location across the study area. The binary dasymetric method uses ancillary data to understand the underlying population distribution and only allow values to be redistributed over areas identified as populated. The GIS network analyses technique is more likely to provide a realistic estimate of accessibility to health services than the planar based approaches. The Teitz and Bart's heuristic for the p-median model have been found to be robust, relatively fast and produce generally better results compared to other search approaches.

## Chapter 3

### 3. METHODS

#### 3.1 Introduction

This chapter presents the implementation of areal interpolation methods for Leicester, U.K. This research aims to determine population surfaces in areas of unknown distributions. The research design applies areal interpolation techniques at different spatial scales to redistribute summary population totals over small areas in a location where small area census data is available for validation (Leicester, U.K.), in order to adapt the most appropriate method, target grid sizes, and ancillary data for areas where small area census data is not available (Port-Harcourt, Nigeria). The idea is to develop and validate models for Leicester so that the most appropriate model and parameters can be calibrated with land cover and census data for Port-Harcourt. The methods used and the stages involved in the research design are illustrated in Figure 3.1.

The research design was implemented in two stages: The first stage, highlighted in blue dash in Figure 3.1, shows that the dasymetric method was applied to census totals for Leicester UA, the source zone while the pycnophylactic interpolation method was applied to census totals for Leicester UA together with all the surrounding UAs (Harborough, Blaby, Charnwood and Oadby and Wigston) with which it shares a common boundary, as the source zones to derive an interpolated gridded population surface at resolutions of 100m and 30m postings, to then be aggregated to MSOA, LSOA and OA target units. The estimated populations were then compared with the known census counts in each case, for validation. This chapter reports the implementation of the first stage and the results are reported in Chapter Four. The second stage, highlighted in red long dash dots in Figure 3.1, used the results from the first stage to adapt the most appropriate interpolation method, target grid sizes and ancillary data and applied these to the Port-Harcourt case study and the results are presented in Chapter Five.

The next section of the chapter introduces the two study areas. Section 3.3 describes the data acquired for the research. Section 3.4 describes the implementation of areal interpolation methods. Section 3.5 describes the evaluation of surfaces generated from areal interpolations. The last section presents a summary of the chapter.

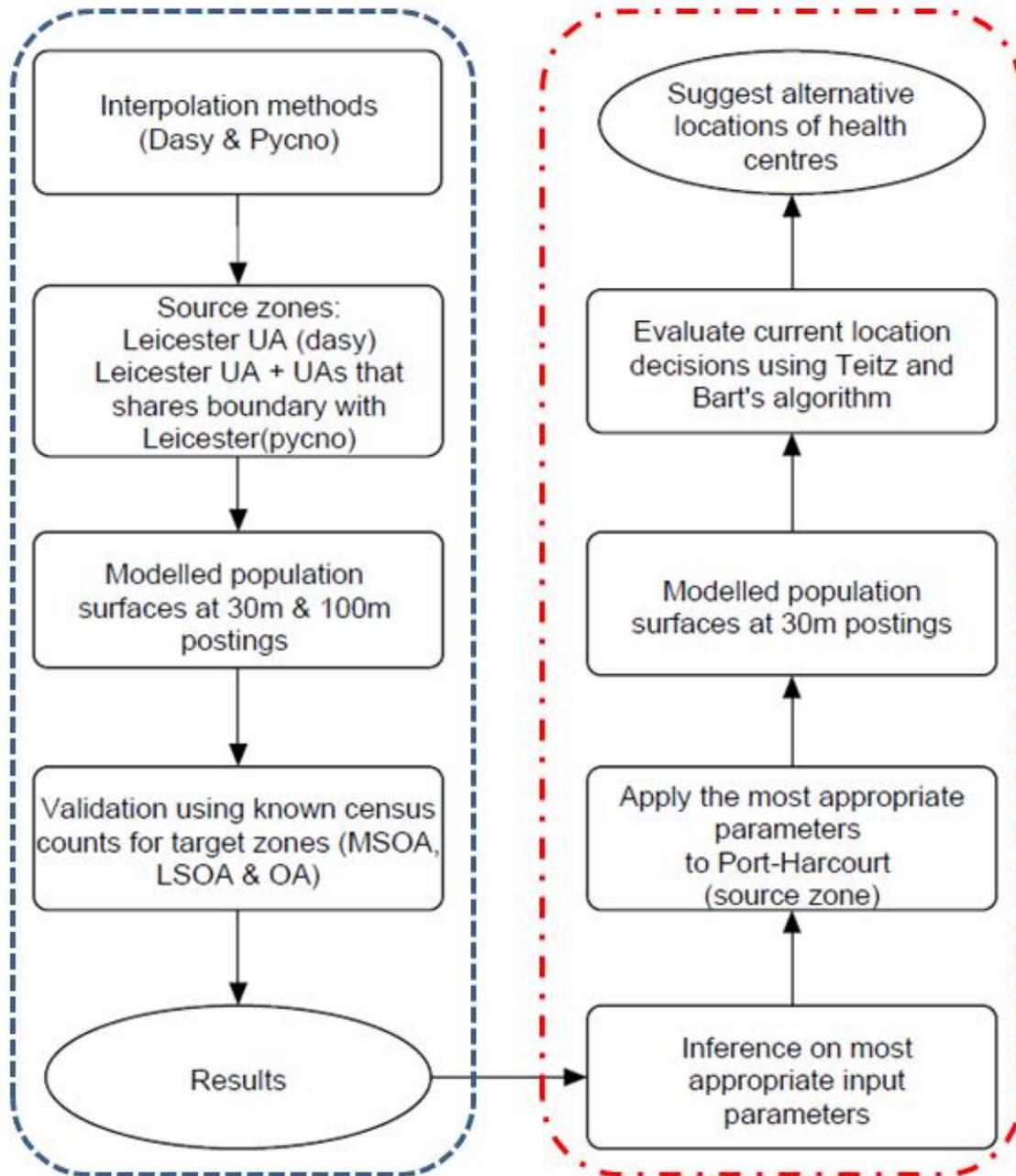


Figure 3.1 - The research design highlighting the first stage in blue dash and the second stage in red long dash dots.

## **3.2 Study areas**

This research aims to redistribute summary population totals to small areas where the population distribution is unknown. The research design involves selecting an area where the actual population distribution is known and where validation is possible, to develop models and evaluate their performance. The most appropriate model will then be calibrated using land cover and population data for an area where the actual population distribution is unknown to estimate its aggregate population to small areas. The methodology was first developed for Leicester. This is because of the analysts' knowledge of the area, the ease of validation and data availability. The model was then calibrated using ancillary information and 2006 population census data for Port-Harcourt, Rivers State, Nigeria to redistribute summary population totals to small areas. Given the diversity of settings between Leicester and Port-Harcourt, one would not necessarily expect models developed for Leicester to capture accurately the relationship between population density and land cover in Port-Harcourt. However, satellite images for the two study areas were not classified the same way. The satellite images for the city of Leicester show industrial zones to be very different from residential areas whereas in Nigeria, although there is a distinction between industrial zones and residential areas, it is not very clear. To minimise the effect of this, the study assumes all built-up areas are residential areas because it will be difficult to extract only residential areas for Port-Harcourt.

### **3.2.1 Leicester**

The city of Leicester is the largest in East Midlands and the tenth largest in the U.K. Leicester covers an area of about 73 km<sup>2</sup> with a total population of 279921 in 111148 resident households as at 2001 census (<http://www.leicester.gov.uk/about-leicester/>). Leicester is one of the U.K.'s most ethnically and culturally diverse places with residents of the city originally from over 50 countries from across the globe. The diversity is as a result of a number of reasons such as economic, family and the fleeing of persecution. The city of Leicester witnessed migration of Ugandan Asians in early 1970s although significant in Leicester's history but does not have much effect on the population. This is because the census figures show a decrease in total population by 1% from 1971 to 1981 (see Table 3.1). The city of Leicester witnessed intense

migratory change in the 1990s when asylum seekers and refugees from Balkans, Iraq, Afghanistan, Iran and Kurdish area of Turkey began to arrive into the U.K. The migration of Somali into the U.K. since 2000, and the arrival of migrants from European Union (EU) countries due part to the accession of 10 countries into the EU in 2004 were major migration events in Leicester's history ([www.leicester.gov.uk/research](http://www.leicester.gov.uk/research)).

Table 3.1 shows total population for the city of Leicester from 1951 - 2011. It also shows the percentage increase/decrease in the population for the period from 1951 - 1961 up to 2001 - 2011. There was a significant increase in total population by 6% and 17% during the periods 1991 - 2001 and 2001 - 2011 respectively. This is important because the variable of interest is population and the two study areas (Leicester and Port-Harcourt) witnessed an increase in population from 1990 to 2011.

Table 3.1 Percentage change in population for Leicester from 1951 to 2011

Census Year	Population	From Census Year	To Census Year	Percentage Increase/Decrease
1951	285200			
1961	288100	1951	1961	1
1971	284200	1961	1971	-1
1981	280300	1971	1981	-1
1991	272133	1981	1991	-3
2001	279921	1991	2001	3
2011	329839	2001	2011	17

The map of England highlighting the location of Leicestershire County in English midlands is shown in Figure 3.2a. Figure 3.2b shows map of Leicestershire County with the city of Leicester in dark shade at the middle of the County. Figure 3.2c shows the boundary map of Leicester.

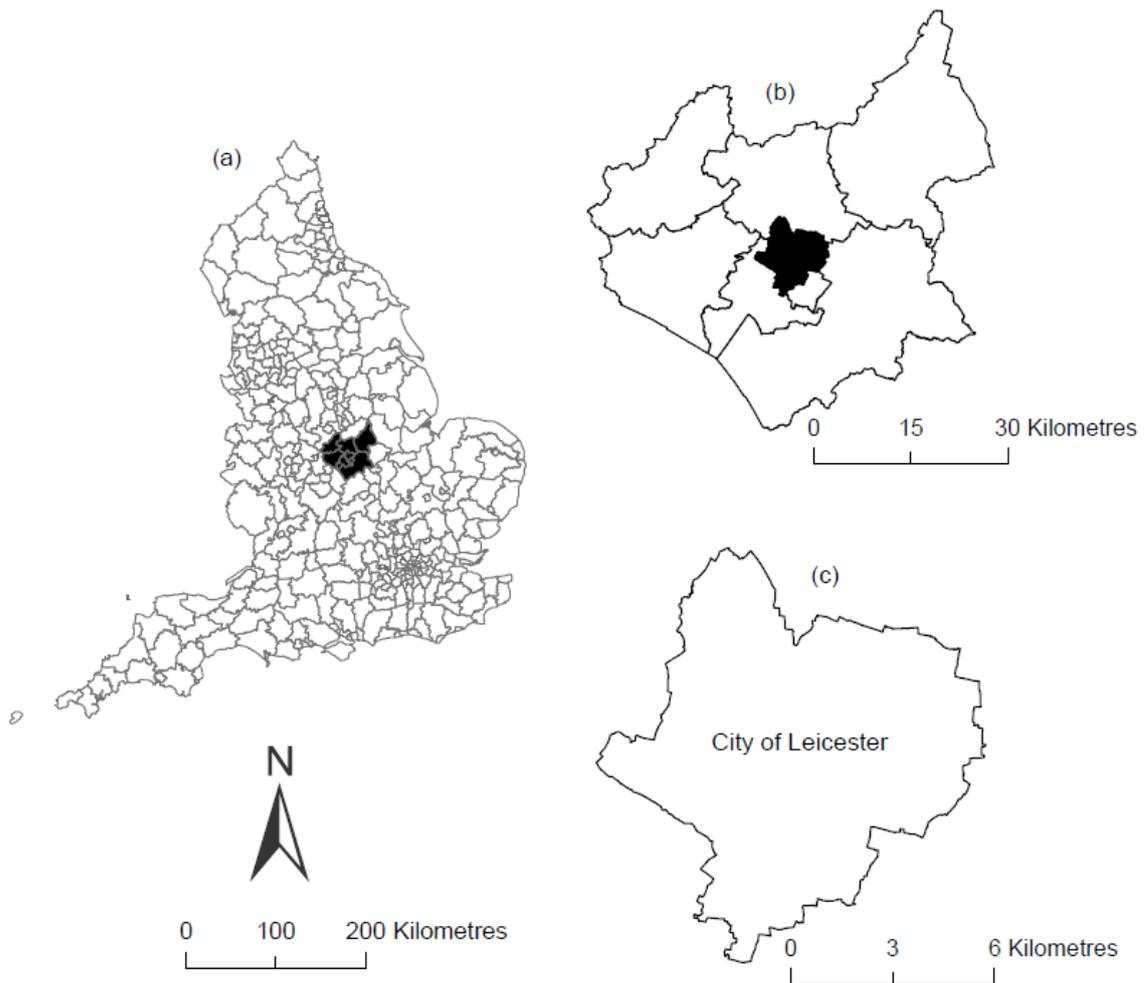


Figure 3.2 - The map of (a) England showing location of Leicestershire County; (b) Leicestershire County with location of Leicester UA; (c) Leicester UA. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

### 3.2.2 Port-Harcourt

Port-Harcourt is the capital of Rivers State from the South-south zone of Nigeria. It is the country's second largest commercial centre. Port-Harcourt city local government covers an area of about 109 square kilometres with a population of 541115 in 126010 households (NPC 2006a). It lies on the coastal plain of eastern Niger Delta along Bonny river located in Nigeria's oil rich Niger Delta (Amnesty International 2010). Crude oil in commercial quantity was discovered in Oloibiri town, Rivers State in 1956 and was first exported in 1958 (NNPC 2011). The Delta has an estimated reserve of 37.2 billion barrels (as at December 30 2013) of crude oil with over 14 major exploration and

production companies having their offices in Port-Harcourt (EIA 2010). Oil and Gas sector generated approximately \$600 billion from 1960 to 2009 with three-quarters of all rural communities in the Delta lacking access to fertile arable land, safe water source and health care (Amnesty International 2009a). Port-Harcourt city experienced intense migratory change from other states of Nigeria because of oil and gas activities, and from rural areas of Rivers State where farmlands have been destroyed due to oil spills in the 90s (Nna and Pabon 2012).

The map of Nigeria highlighting the location of Rivers State in the South-south zone is shown in Figure 3.3a. Figure 3.3b shows map of Rivers State with Port-Harcourt LGA in dark shade around the middle of the map. Figure 3.3c shows the boundary map of Port-Harcourt LGA.

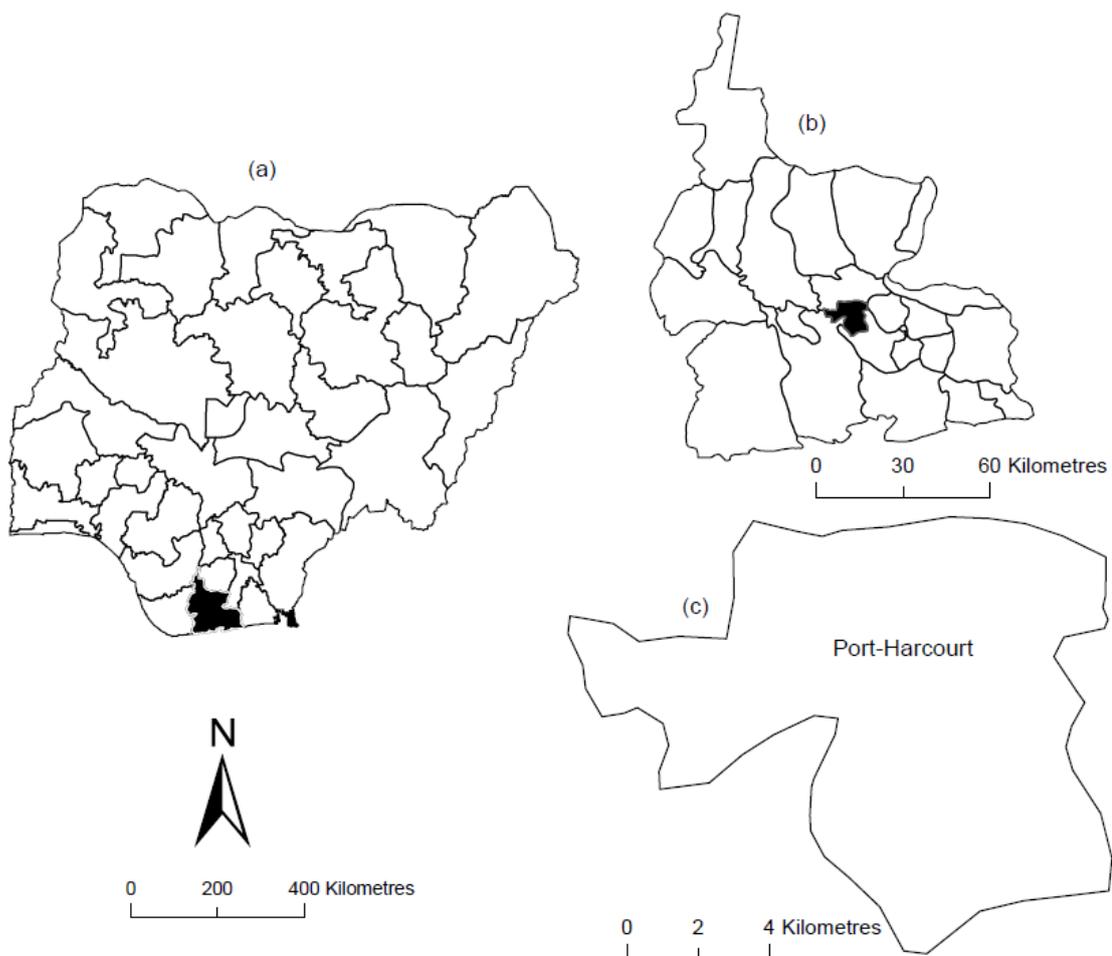


Figure 3.3 - The map of (a) Nigeria showing location of Rivers State; (b) Rivers State showing location of Port-Harcourt; (c) Port-Harcourt City Local Government Area. The digital boundaries are Copyright for Geotechnics Services 2011.

### 3.3 Data

The aim of this research is to redistribute summary population totals available for relatively large UA region for reasons of confidentiality to predict those in much smaller areas such as LSOAs and OAs. A possible solution is to use population census data for Leicester UA with digital boundaries as the source zones, a set of square grids to derive the modelled population surface and digital boundaries for MSOAs, LSOAs and OAs as target zones. Satellite imagery covering the study area is also required because one of the interpolation techniques to be evaluated uses land cover data extracted from a classified satellite image as the ancillary data input. The research also aims to use the population surfaces that are generated to evaluate current health facility locations and to suggest alternative spatial arrangement of public health facilities using heuristic location-allocation approaches. This requires the locations of health centres and a road network dataset. The datasets acquired for implementing areal interpolation for Leicester and Port-Harcourt are presented in Table 3.2 and Table 3.3 respectively.

#### 3.3.1 Leicester

Table 3.2 shows the data acquired for implementing areal interpolation for Leicester. The table also shows the format and source of the data.

Table 3.2 Data: Leicester

<b>Data</b>	<b>Format</b>	<b>Source</b>
*Landsat7 (ETM) 30m spatial resolution	Image	United States Geological Survey (USGS) website ( <a href="http://www.usgs.gov/">http://www.usgs.gov/</a> )
**Ortho-rectified aerial photograph 25cm spatial resolution	Image	Ordnance Survey, U.K. © Crown copyright and/or database right 2013. All rights reserved.
Census data (U.K. 2001) with boundaries of OAs, LSOAs and MSOAs	Shapefile	Census Area Statistics on the Web (casweb) ( <a href="http://casweb.mimas.ac.uk/2001/start.cfm">http://casweb.mimas.ac.uk/2001/start.cfm</a> ).

\*Landsat7 (ETM) acquired 16th April 2003, WRS\_PATH=202 and WRS\_ROW=023.

\*\* Ortho-rectified aerial photograph acquired 22nd May 2010

Figure 3.4 shows a Landsat7 (ETM) image (30m spatial resolution) covering part of England with Leicester area highlighted with a black polygon with image band combination 4:3:2 with WRS\_PATH = 202 and WRS\_ROW = 023. The city of Leicester covers an area of about 73 Square Kilometres. The metadata file is available in Appendix 1.



Figure 3.4 - The Landsat7 (ETM) image 30m spatial resolution of the Leicester area (black polygon) with image band combination 4:3:2.

Figure 3.5 shows one of the 98 tiles of 25cm ortho-rectified aerial photographs that covers the city of Leicester with image band combination 1:2:3.



Figure 3.5 - One of the 98 tiles of 25cm ortho-rectified aerial photograph with image band combination 1:2:3. The tile is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

### *Image Resample*

The 25cm ortho-rectified aerial photograph was used to resample image pixels to different pixel resolutions (3m and 10m) without altering the projected coordinate system. The cubic convolution resample method that uses the closest 4 x 4 block of input cells to fit a smooth curve through the cell centres to find the value, was used to compute each output cell value. The weighting factors for the average of the input cells are computed using a cubic function of distance. The cubic convolution resample method was used because it reduces blurring and produces a smoother output image when compared with the output of nearest neighbour or bilinear method. The resampled

images for Leicester area with 3m and 10m spatial resolutions are shown in Figures 3.6 and 3.7 respectively.



Figure 3.6 - 25cm ortho-rectified aerial photograph of Leicester area resampled to 3m with image band combination 1:2:3.



Figure 3.7 - 25cm ortho-rectified aerial photograph of Leicester area resampled to 10m with image band combination 1:2:3.

The boundary of Leicester area and the support grids (30m and 100m square grids) used to derive the modelled population surfaces are shown in Figures 3.8.

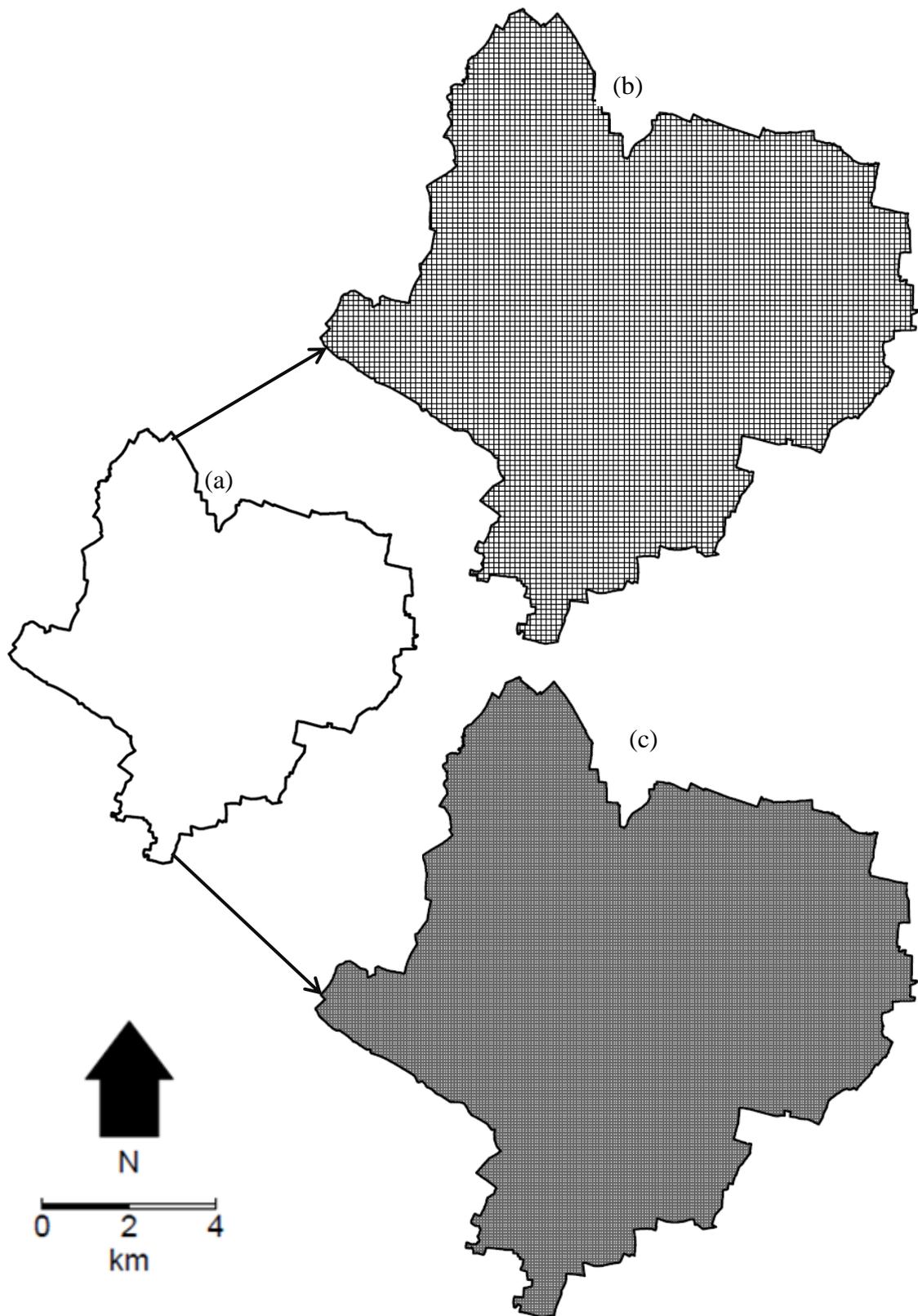


Figure 3.8 – The boundary data for; (a) Leicester UA, (b) 100m square grids, and (c) 30m square grids. The digital boundary is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

The boundaries of MSOA, LSOA and OA for Leicester UA used as target zones to test the performance of the interpolation methods are shown in Figures 3.9 to 3.11 respectively.

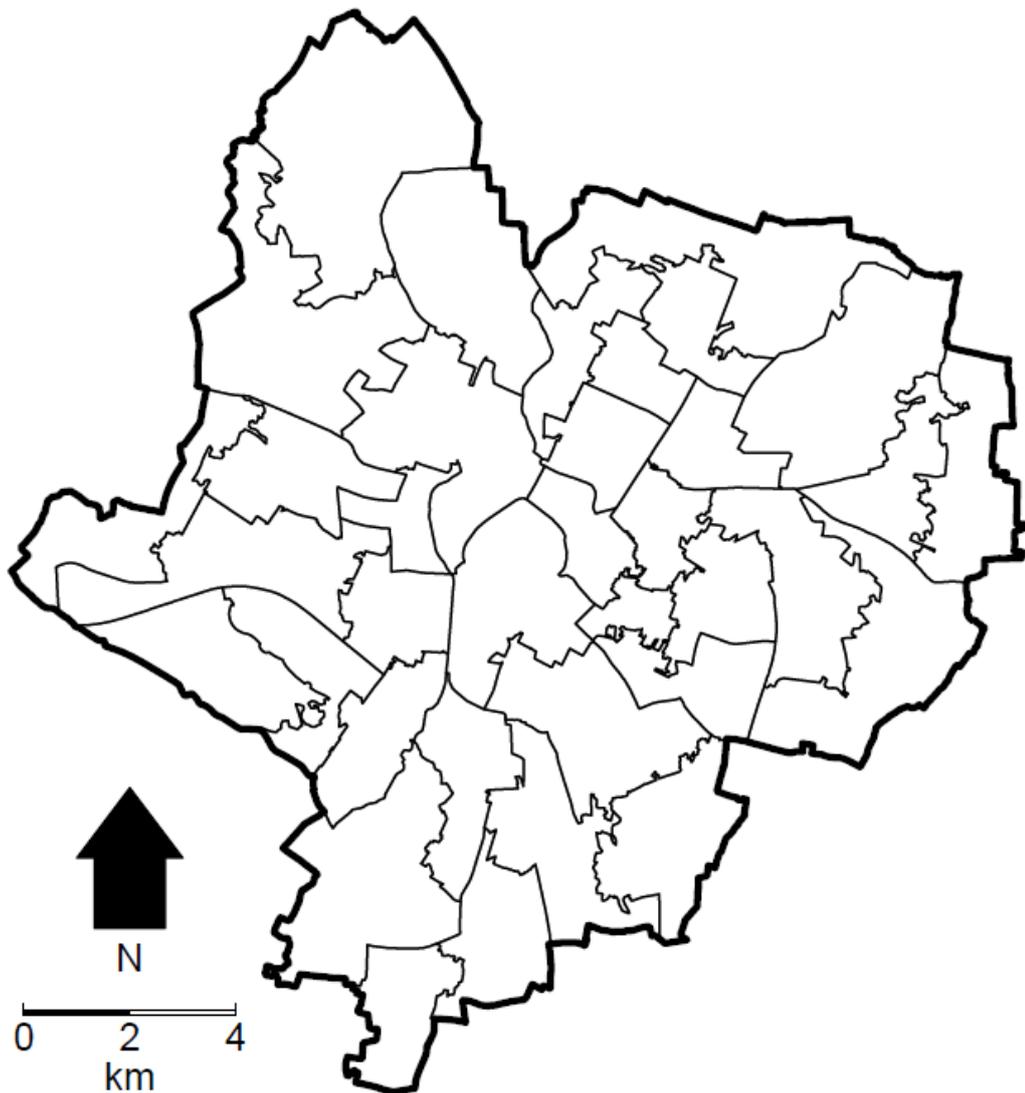


Figure 3.9 – MSOA used as the testing zones. The digital boundary is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

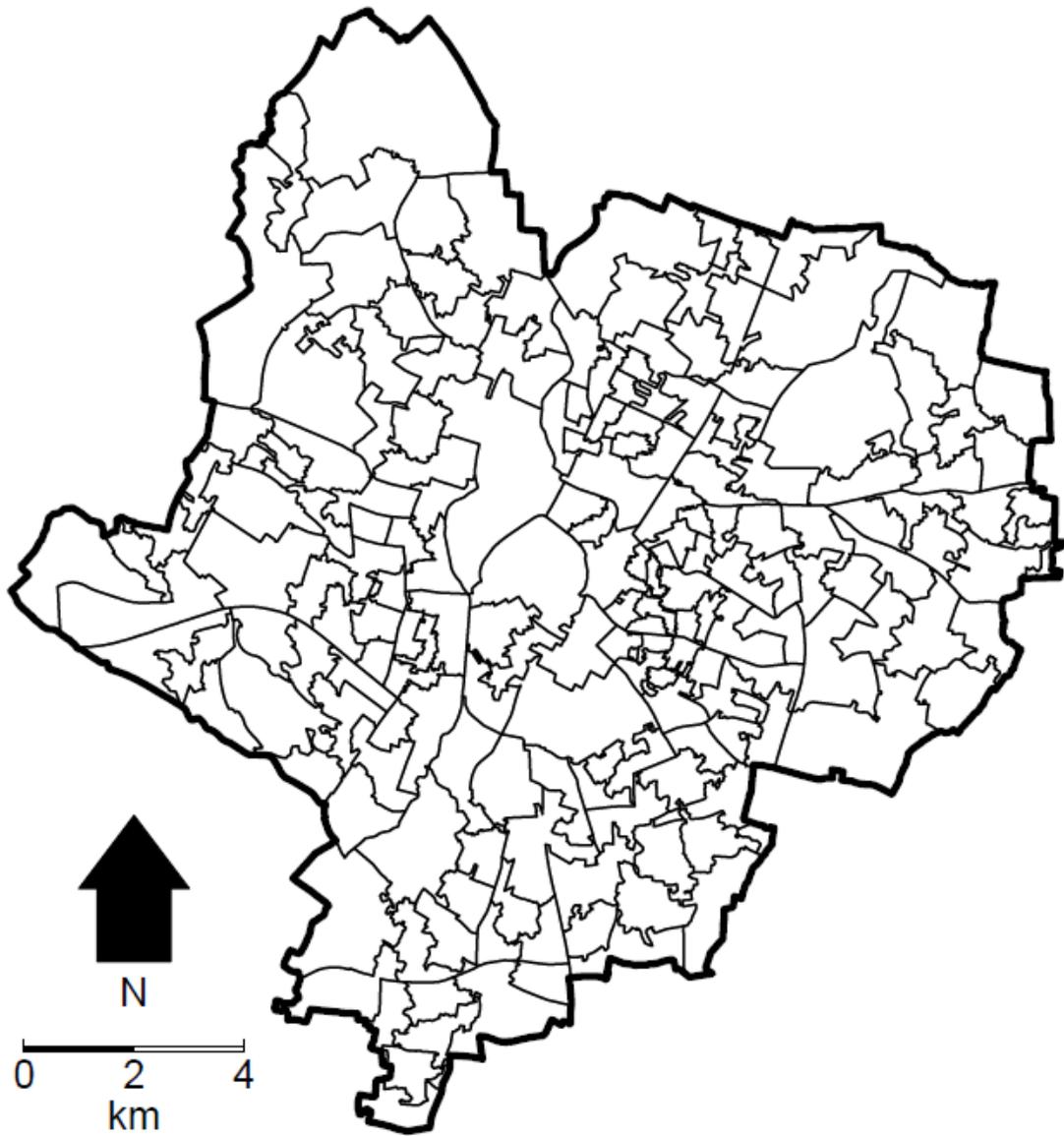


Figure 3.10 - LSOA used as the testing zones. The digital boundary is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

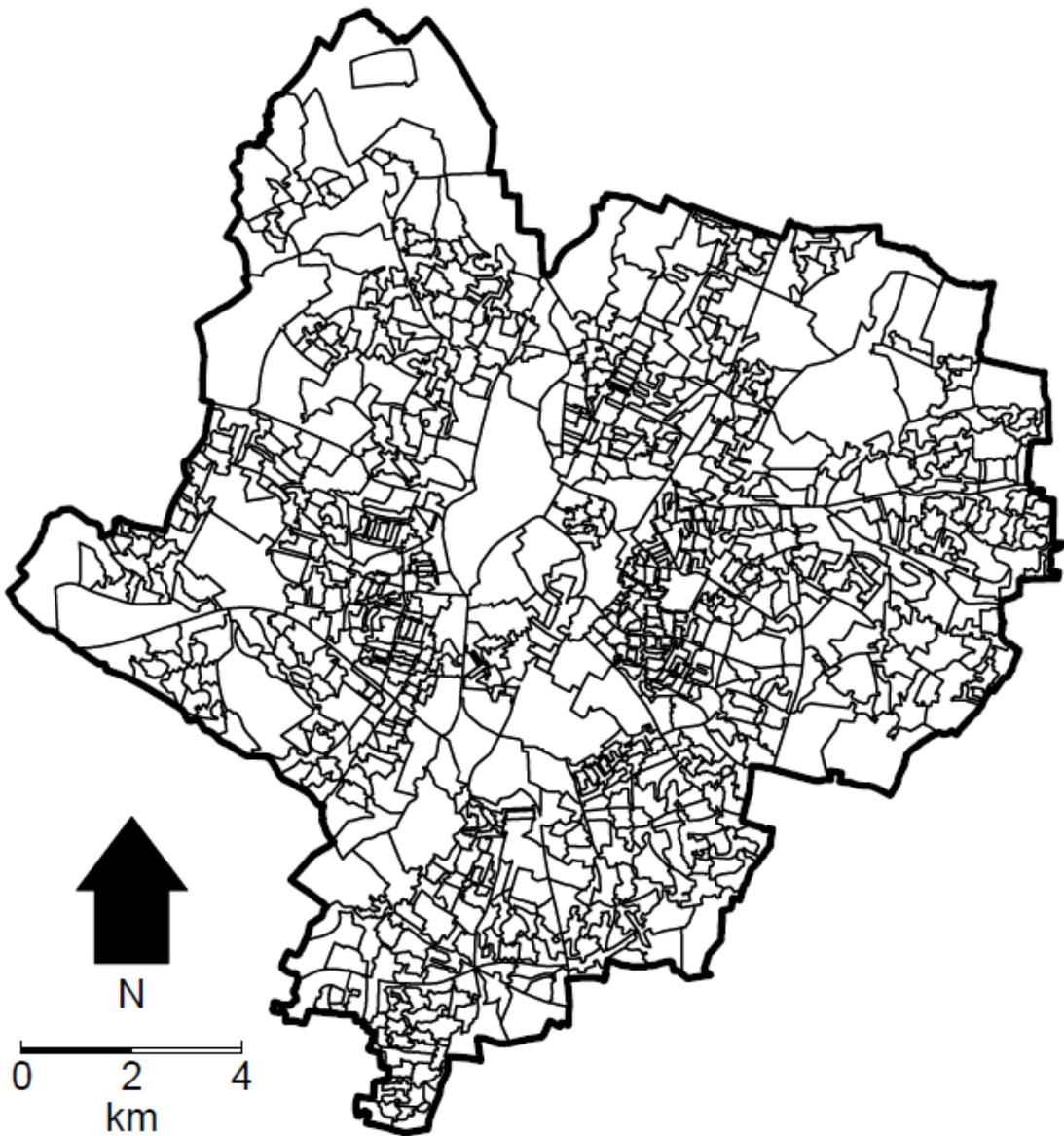


Figure 3.11 - OA used as testing zones. The digital boundary is © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

### 3.3.2 Port-Harcourt

Table 3.3 shows the data acquired for implementing areal interpolation and location-allocation for Port-Harcourt. The table also shows the format and source of the data.

Table 3.3 Data: Port-Harcourt

Data	Format	Source
*Landsat7 (ETM+) 30m spatial resolution	Image	United States Geological Survey (USGS) website ( <a href="http://www.usgs.gov/">http://www.usgs.gov/</a> )
**Spot5 (colour) 10m spatial resolution	Image	Astrium Services U.K. ( <a href="http://www.astrium-geo.com">www.astrium-geo.com</a> )
***Quickbird 60cm spatial resolution	Image	Geo-technics Services Limited, Port-Harcourt.
Population census (2006) with priority tables	Excel	National Bureau of Statistics' website <a href="http://www.nigerianstat.gov.ng/">http://www.nigerianstat.gov.ng/</a>
Primary Health Care Centres (PHCCs)	MS word	Planning, research & statistics dept., Ministry of Health, Port-Harcourt.
Road network data	Shapefile	Geo-technics Services Limited, Port-Harcourt
States and LGAs boundary	Shapefile	Geo-technics Services Limited, Port-Harcourt

\*Landsat7 (ETM+) acquired 8th January 2003 WRS\_PATH=188 and WRS\_ROW=057.

\*\*Spot5 acquired 10th January 2007

\*\*\* Quickbird image acquired 2006

Figure 3.12 shows a Landsat7 ETM+ image (30m spatial resolution) covering Port-Harcourt area, Rivers State, Nigeria with image band combination 4:3:2, and with WRS\_PATH = 188 and WRS\_ROW = 057. A subset of the image for Port-Harcourt was obtained as shown in Figure 3.13. The satellite image was used to derive land cover data for Port-Harcourt at 30m spatial resolution that was used as the ancillary data input for the binary dasymmetric method. Port-Harcourt city covers an area of about 109 Square Kilometres. The metadata file information is available in Appendix 2.

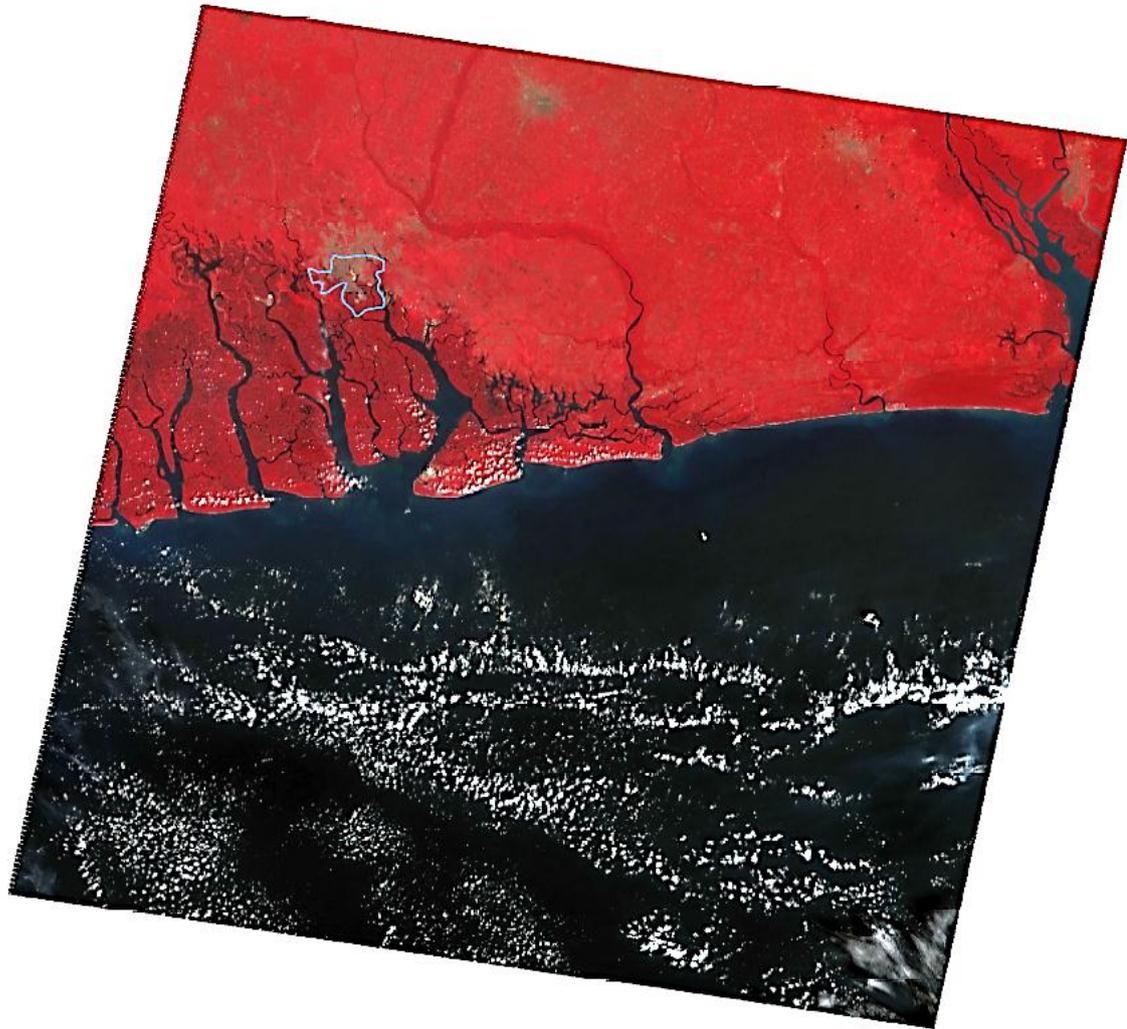


Figure 3.12 - The Landsat7 (ETM+) 30m spatial resolution covering Port-Harcourt area (shaded in grey), Rivers State, Nigeria with image band combination 4:3:2.

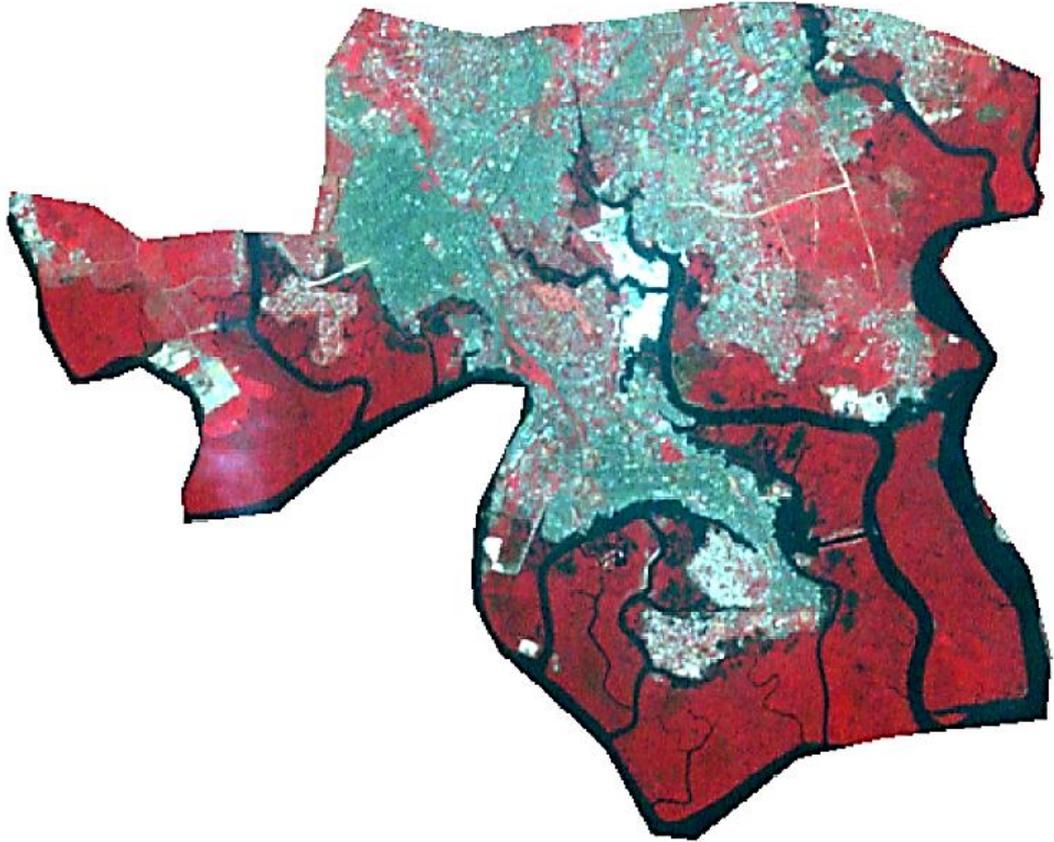


Figure 3.13 - Subset for Port-Harcourt acquired from the Landsat (ETM+) 30m spatial resolution. Port-Harcourt city covers an area of about 109 square kilometres.

Figure 3.14 shows Spot5 colour image (10m spatial resolution) covering Port-Harcourt area with image band combination 4:3:2. The satellite image was used to derive land cover data for Port-Harcourt at 10m spatial resolution that was used as the ancillary data input for the binary dasymetric method. The metadata file information is available in Appendix 3.

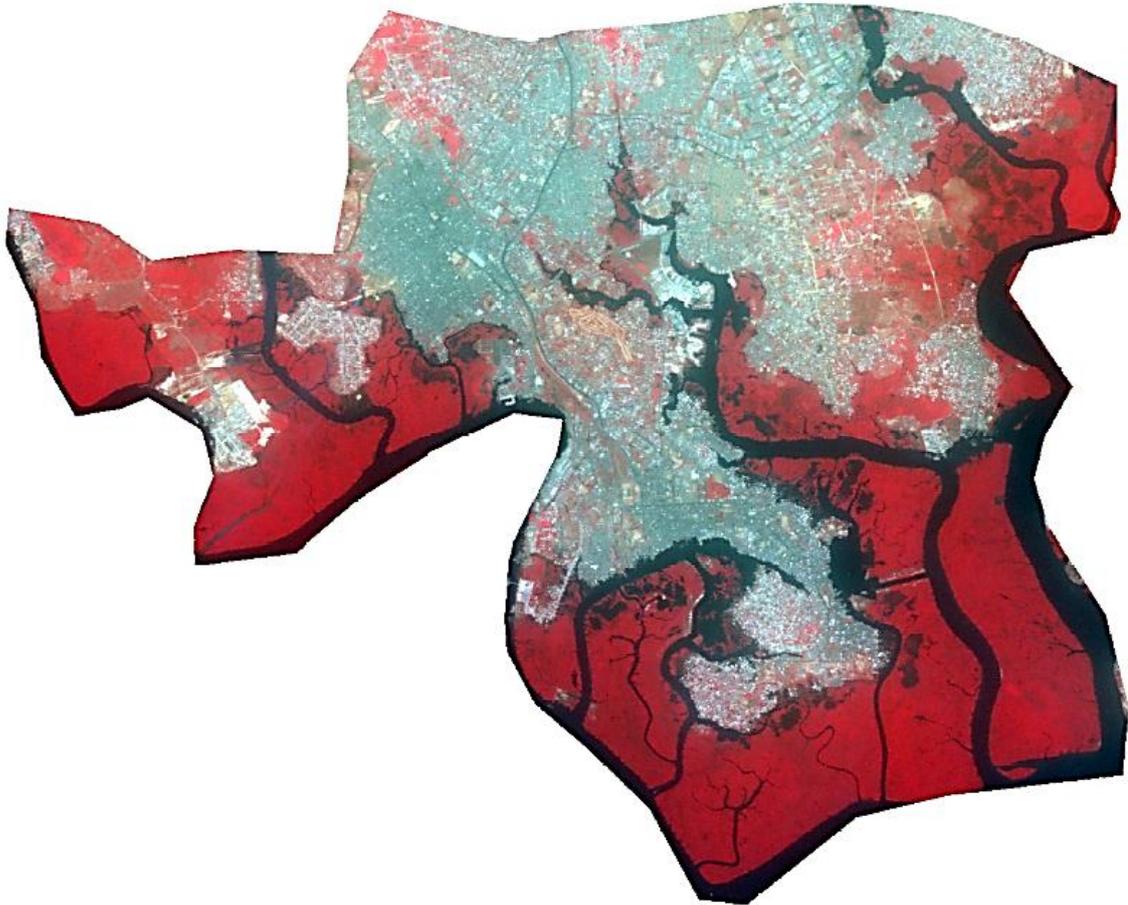


Figure 3.14 – Spot5 colour image 10m spatial resolution for Port-Harcourt obtained from Astrium Services SpotCatalog (<http://catalog.spotimage.com/PageSearch.aspx>).

Figure 3.15 shows a Quickbird satellite image 60cm spatial resolution covering Port-Harcourt (red polygon) with image band combination 1:2:3. The image was obtained from Geo-technics Services Limited, Port-Harcourt, Nigeria in ecw format. The image was resampled to 3m spatial resolution using cubic convolution resample method as described earlier. The resampled image for Port-Harcourt area with 3m spatial resolution is shown in Figures 3.16.



Figure 3.15 – Quickbird satellite image 60cm spatial resolution covering Port-Harcourt (red polygon) with image band combination 1:2:3. The image was obtained from Geo-technics Services Limited, Port-Harcourt.

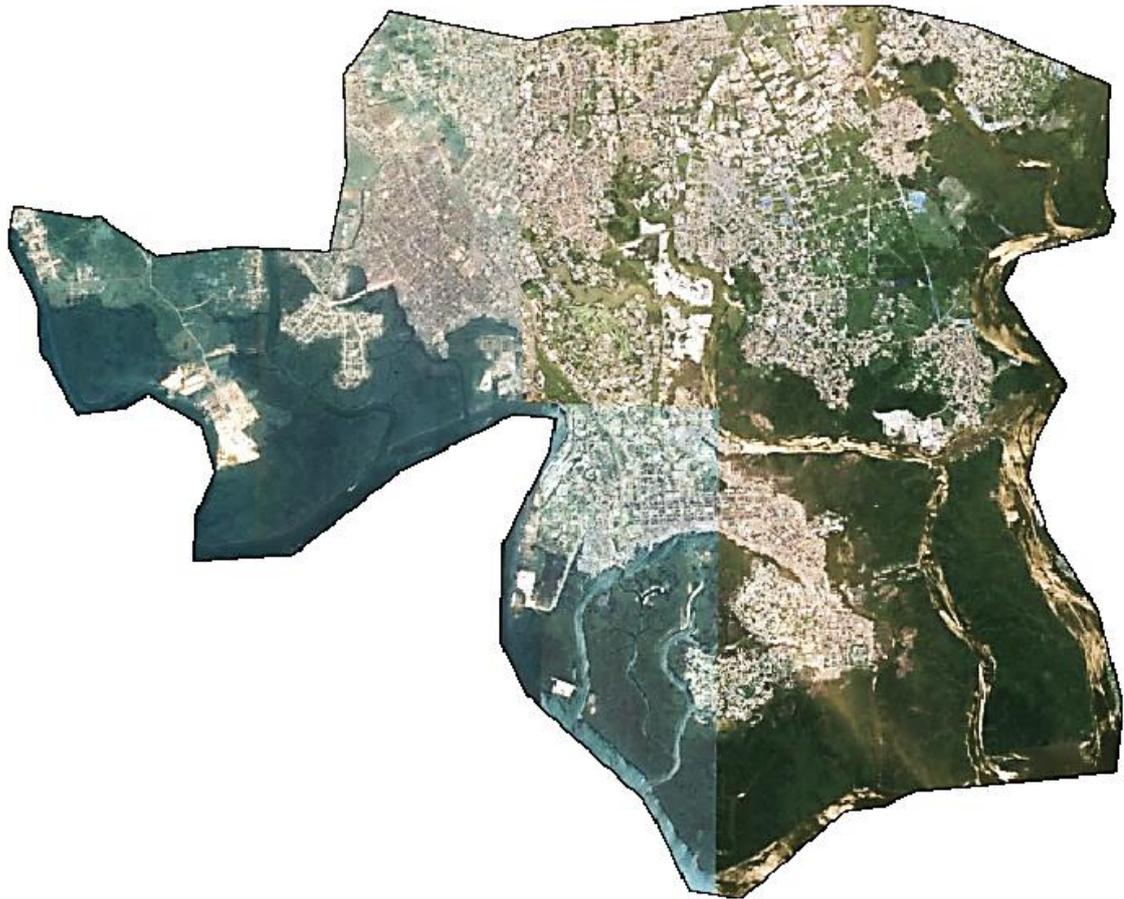


Figure 3.16 - Quickbird satellite imagery 60cm spatial resolution covering Port-Harcourt resampled to 3m with image band combination 1:2:3.

The source zone and the support grids (30m) used for the implementation of areal interpolation for Port-Harcourt are shown in Figures 3.17.

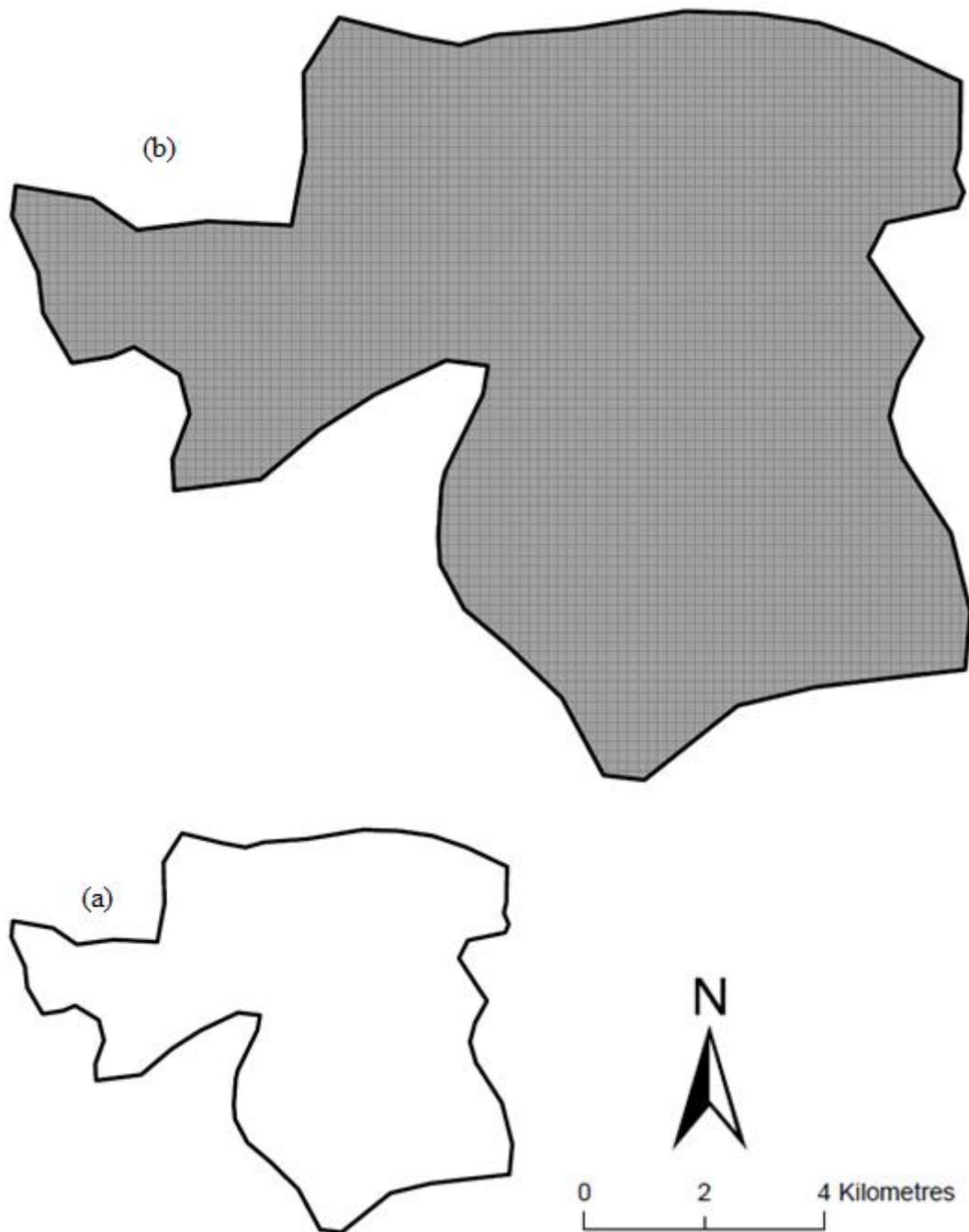


Figure 3.17 - The boundary data for; (a) Port-Harcourt LGA (source zone), and (b) 30m square grids used to derive the modelled population surface. The digital boundary is Copyright for Geotechnics Services 2011.

### **3.4 Implementation of areal interpolation methods**

This section describes the implementation of the two areal interpolation methods: the binary dasymetric and pycnophylactic interpolation technique. The data described in section 3.3.1 was used to apply areal interpolation methods across different spatial scales over Leicester area. Leicester UA boundary was used as the source zone for the binary dasymetric method while Leicester UA together with all the surrounding UAs with which it shares a common boundary were used as source zones for the pycnophylactic interpolation method, to derive an interpolated gridded population surface at resolutions of 100m and 30m postings, to then be aggregated to MSOA, LSOA and OA target units and assessed for accuracy using known census counts in each case. An essential component for dasymetric mapping method is an ancillary data (e.g. land use) that provides a binary divide between populated and unpopulated areas (Kim and Choi 2011; Qiu et al. 2012). The satellite images described in section 3.3.1 were classified into land cover types and reclassified to create a binary divide between built-up and non-built-up areas.

#### **3.4.1 Supervised Classification**

The supervised classification technique was employed to process the remotely sensed images (described in section 3.3.1) and identify the extent of built-up areas so that it can be used as ancillary data in the implementation of the binary dasymetric method in order to redistribute aggregate population of the source zone to only those areas. This was done in Erdas Imagine 2013. Landsat7 (ETM) 30m spatial resolution and aerial photo 25cm spatial resolution resampled to 10m and 3m were used to derive the land cover data. Areas of homogenous land cover in the image were selected as the training sets and circumscribed by polygon boundaries with each representing known land cover category. Spectral signatures that represent the mean digital numbers of those pixels selected in the training sites were defined. Fourty, thirty-four and thirty-two spectral signatures were selected for each of 30m, 10m and 3m spatial resolution image respectively. These spectral signatures were evaluated using the *display mean plot window* to ensure the signatures for each land cover type are closely related and to generate as little confusion as possible, suggesting a clear separation of land cover classes before classification. The mean plot window for the Landsat7 (ETM) 30m

spatial resolution image is shown in Figure 3.18 while those of 10m and 3m from resampled aerial photo of Leicester are shown in Appendix 4. After the signatures were evaluated and are determined to be satisfactory, the signatures of each land cover type were combined using the *merged selected signatures* utility in the tool bar creating a new weighted signature for each land cover type. This was done to ensure the signatures of each land cover type include each training site selected. The names and colour symbols for the new combined signatures were changed to identify the land cover type that each signature represents. Three land cover types were identified for the 30m spatial resolution image (vegetation, water and built-up) and five land cover types (water, bare ground, vegetation, thick vegetation and built-up) for each of 10m and 3m image were identified. This is because the 10m and 3m images appear to offer greater spatial precision in the depiction of different land cover types compared to the 30m image. These land cover types were selected with special interest on built-up areas because they are more likely to be associated with residential land use. Category names (vegetation, thick vegetation, built-up, bare ground and water) were assigned to these classes and are described in Table 3.4. These were used in the supervised classification. The signature editor and the signature mean plot for the combined signatures for Landsat (ETM) 30m spatial resolution image are shown in Figures 3.19 and 3.20 respectively, and those for 10m and 3m from resampled aerial photo of Leicester are shown in Appendix 5.

Table 3.4 Classified land cover categories for Leicester

<b>Category name</b>	<b>Definition</b>
Built-up	All built-up areas. This includes residential areas, commercial areas, urban recreational areas, industrial facilities, cemeteries, campus-like institutions, roads, schools etc.
Vegetation	All agricultural land uses, parks, gardens, shrub, grass, crops, tree cover etc.
Thick vegetation	All forests, thick tree covers, woods etc.
Water	Open water such as rivers, lakes, ponds, reservoirs, streams, canals etc.
Bare ground	All open fields and non-built-up areas, exposed soil.

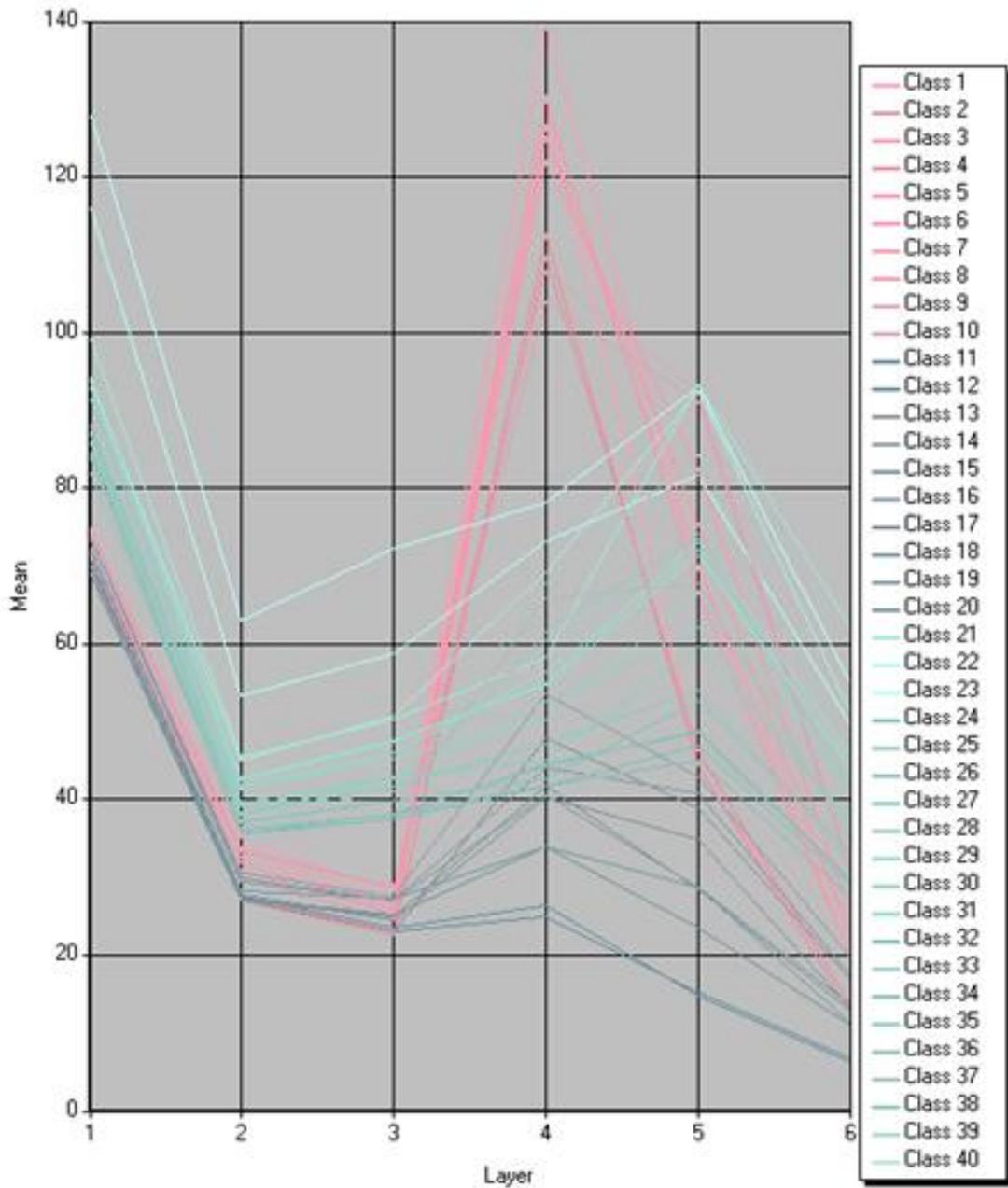


Figure 3.18 - Signature mean plot evaluating signatures for vegetation, builtup and water from Landsat7 (ETM) 30m spatial resolution image.

Class #	>	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1	>	Vegetation	<span style="color: green;">■</span>	0.000	1.000	0.000	41	41	1224	1.000	✓	✓	✓	✓	
2	>	Water	<span style="color: blue;">■</span>	0.000	0.000	1.000	42	42	16	1.000	✓	✓	✓	✓	
3	▶	builtup	<span style="color: brown;">■</span>	0.627	0.322	0.176	43	43	1069	1.000	✓	✓	✓	✓	

Figure 3.19 - Signature editor for the combined signatures from Landsat7 (ETM) 30m resolution image

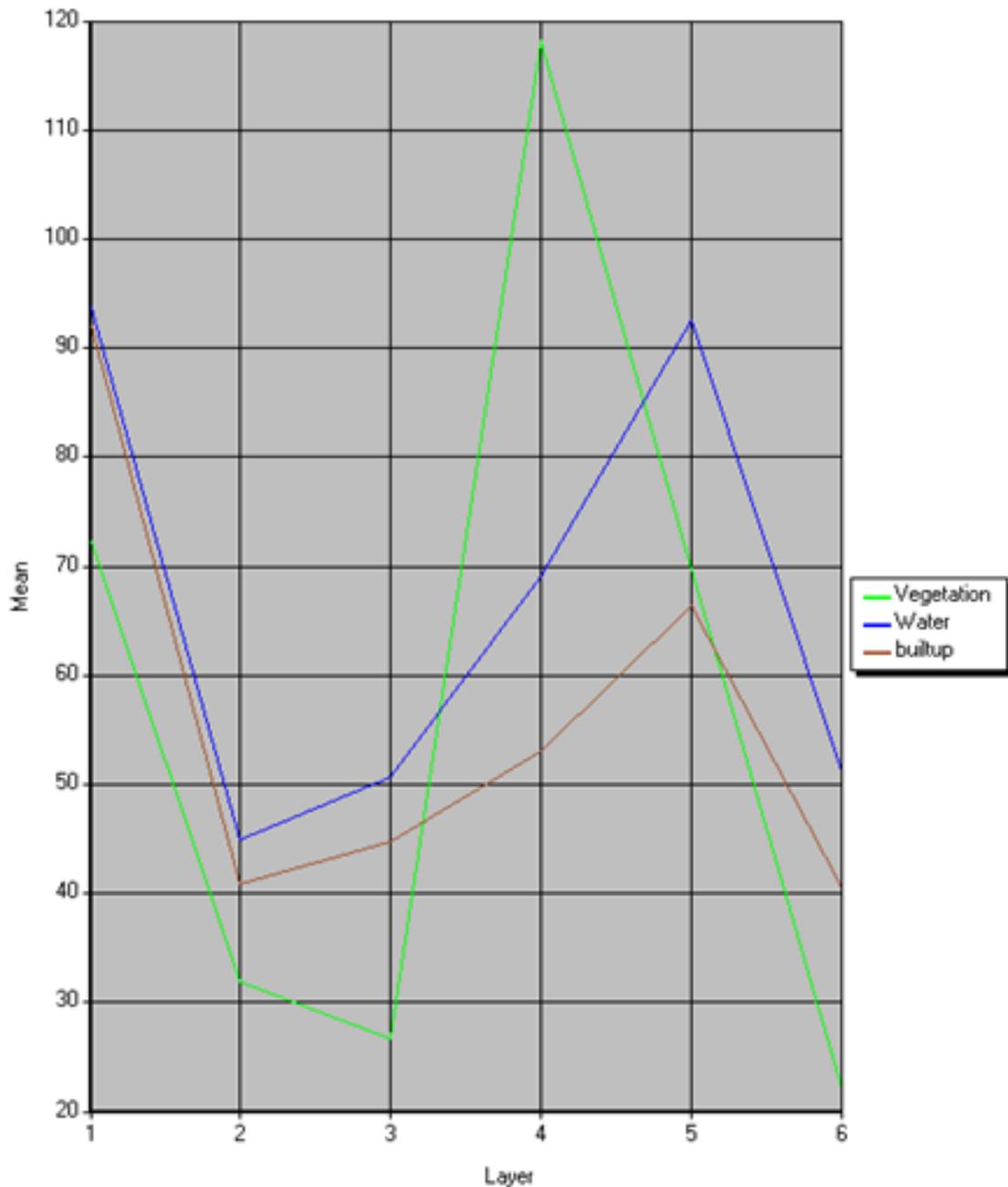


Figure 3.20 - Signature mean plot for the combined signatures from Landsat7 (ETM) 30m resolution image.

The next step applies the maximum likelihood classification algorithm. The algorithm compares every pixel in the training samples with various spectral signatures and assigns all the pixels in an image into land cover types based on their most likely statistical relationship as described by Lillesand et al. (2008). Figure 3.21 shows the steps involved in supervised classification. It is common practise to use maximum likelihood classifier to classify land cover based on spectral signatures at per pixel level,

while ignoring spatial features in an image. Maximum likelihood classification algorithm can provide reasonably good classification results for Landsat imagery (Lu et al. 2012; Blaschke 2010). However, there are a number of issues related to using maximum likelihood classifier for medium and high resolution imagery. This is because a significant proportion of medium and high spatial resolution imagery in urban areas can be affected by shadows (Zhou et al. 2009). In this study, extracting urban land cover from resampled aerial photo data was more difficult compared to using the Landsat (ETM) source. Lu et al. (2010) have shown how the use of spatial features improves land cover classification, especially when high spatial resolution images are used. Object-based classification provides an alternative for classifying remotely-sensed images into thematic map. Lu et al. (2012) compared object-based classification with maximum likelihood and found object-based classification to be especially valuable for higher spatial resolution images. The object-based classification algorithm was not applied in this study. The supervised classification procedure was repeated several times using a consistent criterion throughout the classification process until the best possible results with global classification accuracy of 87.89%, 83.20% and 82.03% was achieved using 30m, 10m and 3m spatial resolution satellite image respectively. The accuracy of classification was assessed by comparing 256 random points identifying certain pixels in the classified image to reference pixels for which the class is known. The accuracy reports are available in Appendix 6. The results of the supervised classification using 30m, 10m and 3m spatial resolution images are presented in Chapter 4. The classified images were reclassified into a simple binary division of built-up and non-built-up areas. The built-up areas were those classified as built-up while the non-residential areas were those classified as vegetation, thick vegetation, bare ground and water. A weighting factor of 1 was assigned for the built-up areas and 0 for the non-built-up areas. The binary masks derived from land cover data derived from classified satellite images of 30m spatial resolution, and those of 10m and 3m spatial resolutions derived from classified resampled aerial photo data which represent the underlying spatial distribution of population within the source zone, are presented in Chapter 4.

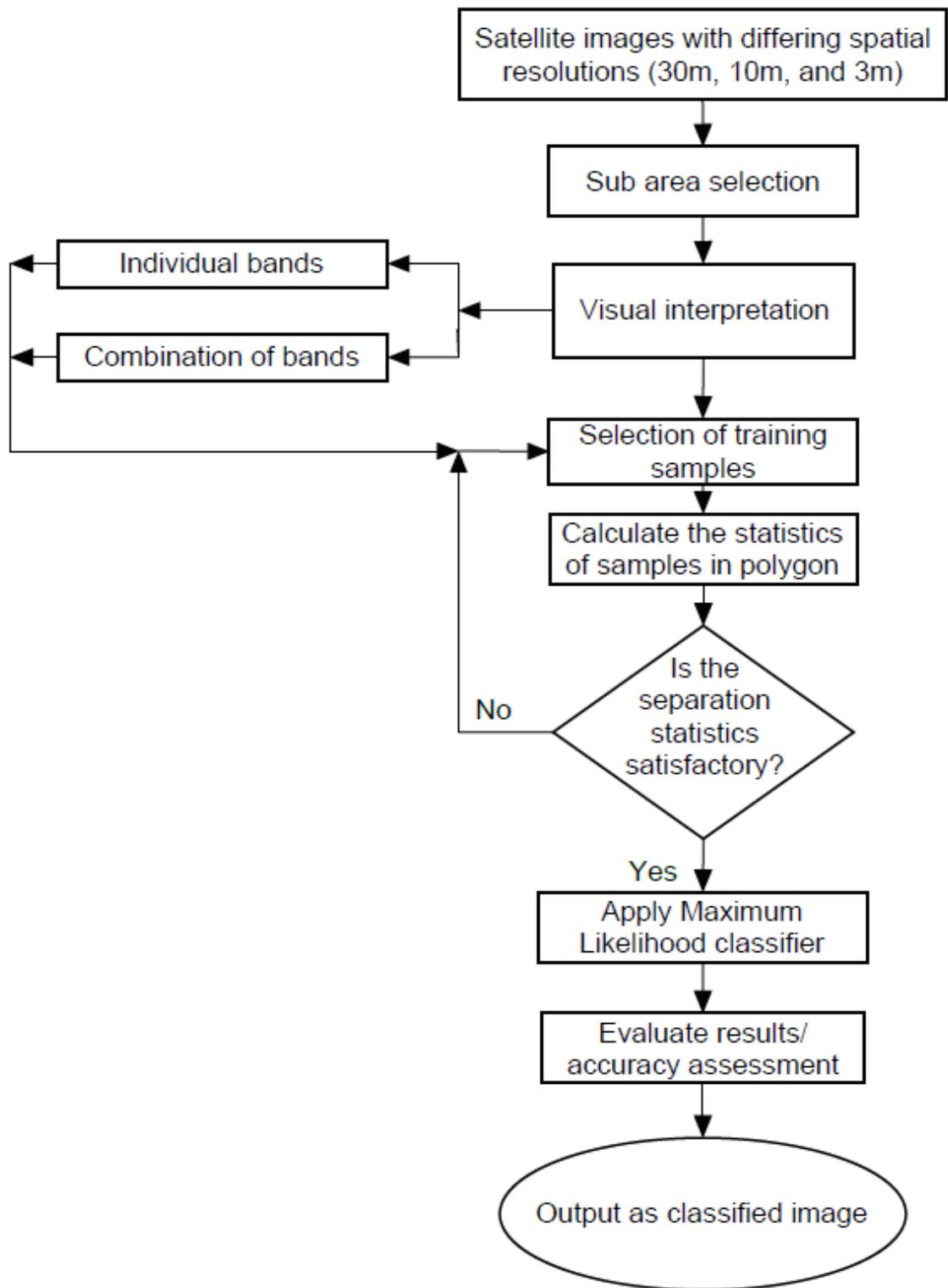


Figure 3.21 – Flowchart showing steps involved in supervised classification

### 3.4.2 The binary dasymetric method

The binary dasymetric method is based on using additional geographic information (e.g. land use) that provides a binary divide between built-up and non-built-up areas within the source zone (Langford et al. 1991; Langford and Unwin 1994; Eicher and Brewer 2001; Mennis 2003; Qiu et al. 2012). The use of this additional geographic information (ancillary data) as control units allows the source zone population to be redistributed only to built-up areas and assumes a constant population density within the source zone (Langford and Unwin 1994; Mennis 2003; Qiu et al. 2012). The population density is calculated by dividing the population count of a source zone by the total size of all built-up areas within the source zone. The population density is multiplied by the area of overlap between areas identified as built-up and the target zone to obtain an estimate of the population, and then all population estimates that falls within a target zone are summed up to obtain the target zone population. The binary dasymetric method can be implemented in both raster and vector modes, and may produce similar interpolation results (Qiu et al. 2012). In this study, the binary dasymetric method was implemented in vector mode. Land cover data derived from classified satellite imagery of the Leicester area with differing spatial resolutions (30m, 10m and 3m) were each converted to vector and used as the ancillary input data to implement the vector-based binary dasymetric method.

The binary dasymetric method (vector mode) was implemented in nine steps and it involves: (1) overlaying the vector ancillary data representing populated areas within the source zone with the boundary of the source zone using the intersect tool in ArcGIS 10.2.2; (2) calculating the area of overlap using calculate geometry; (3) summing up all the areas of overlap to obtain the total area for the source zone; (4) calculating the density of the populated areas by dividing the population of each source zone by the total area for that source zone. This is expressed mathematically in Equation 3.1; (5) overlaying each of 30m and 100m grids of cells with the areas of intersect derived in (1) above; (6) calculating the areas of the newly overlaid grids using calculate geometry; (7) calculating population estimate for each overlaid grid by multiplying the area with the density. This is expressed mathematically in Equation 3.2; (8) overlaying the target zones with the population for each overlaid grid using the intersect tool; and (9) obtaining the interpolated population of each target zone by summing all the

populations for each overlaid grid within the target zone. A flowchart describing these steps is shown in Figure 3.22

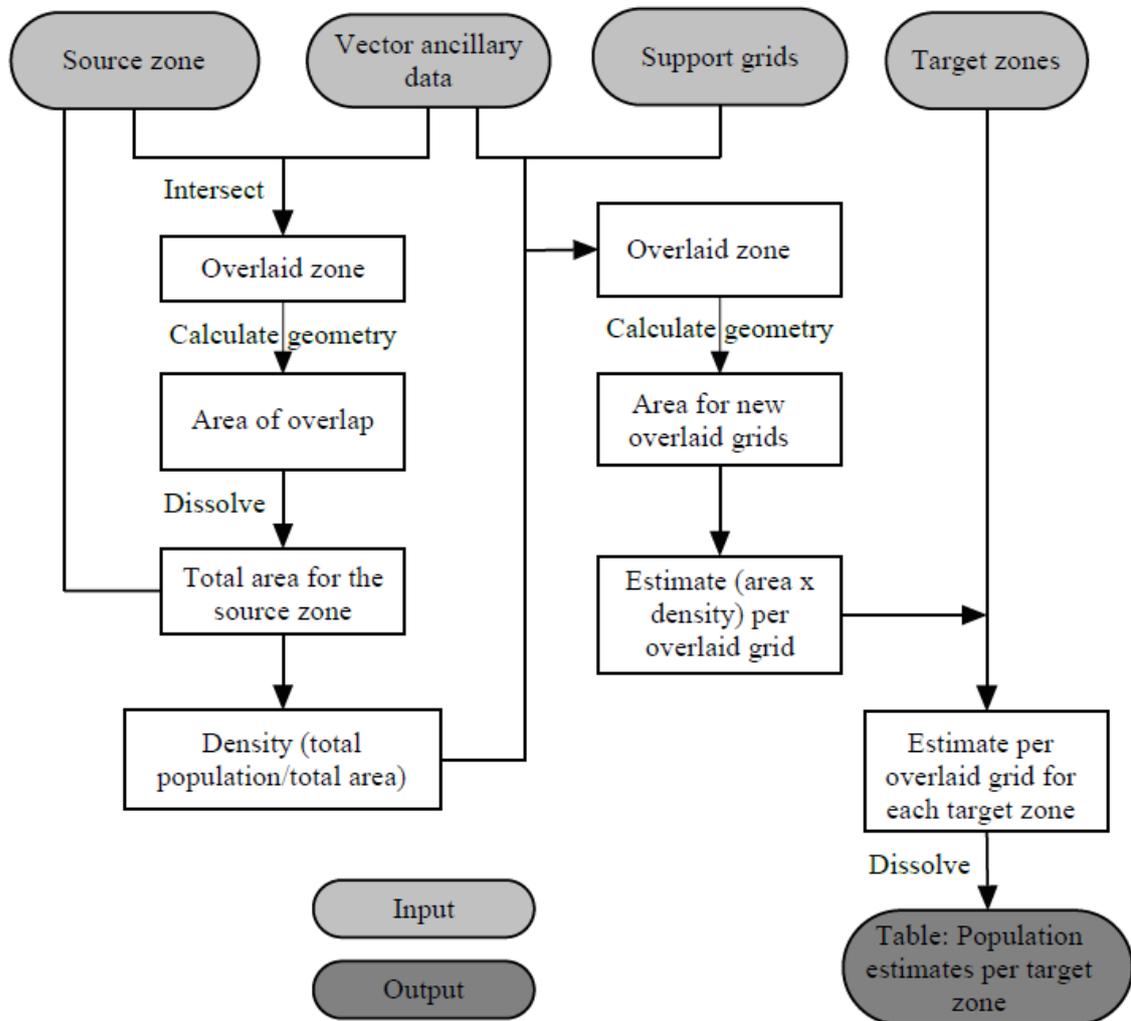


Figure 3.22 - Implementation steps for the binary dasymetric method (vector mode)

The population density of the source zone is expressed mathematically as:

$$d_{sp} = \frac{P_s}{A_{sp}} \quad [3.1]$$

Where  $d_{sp}$  is the population density of the source zone  $s$ ,  $P_s$  is the total population of source zone  $s$  and  $A_{sp}$  is the area of source zone  $s$  having land cover identified as populated.

The estimated population for the overlaid grid is expressed mathematically as:

$$\hat{P}_t = \sum_{s=1}^s A_{tsp} d_{sp} \quad [3.2]$$

Where  $\hat{P}_t$  is the estimated populations of overlaid grid,  $t$ ;  $s$  is the number of source zones,  $A_{tsp}$  is the area of intersection between overlaid grid  $t$  and source zone  $s$  and  $d_{sp}$  as defined in Equation 3.1.

### 3.4.3 Pycnophylactic Interpolation Method

The pycnophylactic interpolation of the population count for Leicester UA was used to derive an interpolated gridded population surface at resolutions of 100m and 30m postings, to then be aggregated to MSOA, LSOA and OA target units and assessed for accuracy using known census counts in each case. The unitary authority of Leicester is a “single polygon” and the technique can only be applied to two or more polygons. Polygons of districts that share a common boundary with Leicester UA (Harborough, Blaby, Charnwood and Oadby and Wigston) were included to generate the pycnophylactic surface with five source zones (as in Figure 3.23). The total population for each source zone is shown in Table 3.5 below.

Table 3.5 Population totals for the source zones used to implement the pycnophylactic interpolation method.

<b>Unitary Authority</b>	<b>Total Population</b>
Blaby District	90252
Charnwood District	153462
City of Leicester	279921
Harborough District	76559
Oadby and Wigston District	55795

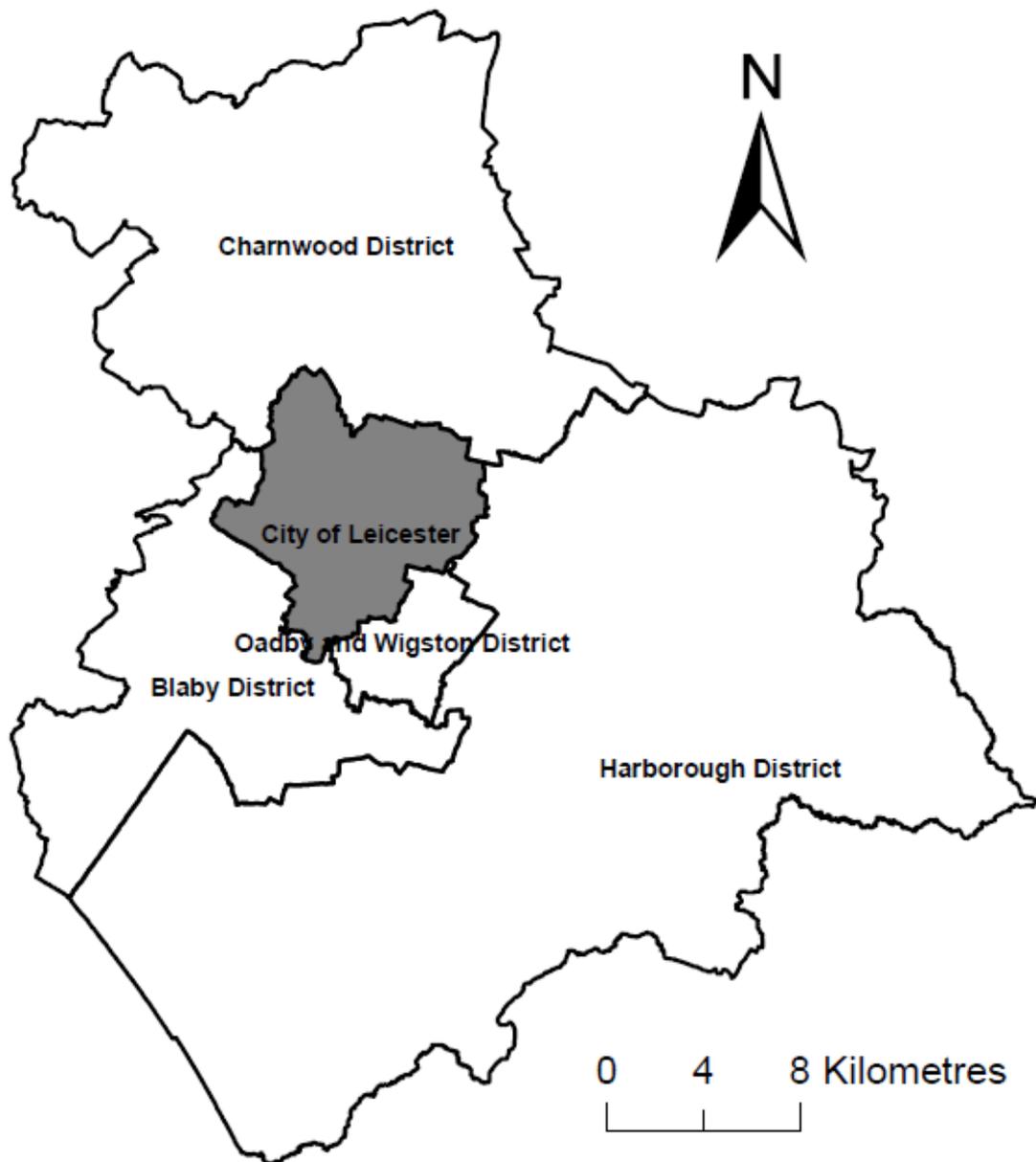


Figure 3.23 - Source zones used for the pycnophylactic interpolation method with the City of Leicester (study area) shaded in grey.

The basic principle of the pycnophylactic interpolation is to create a smooth surface across the study area with no sudden change at polygon boundaries and the total value of target polygons must equal that for source polygons (Tobler 1979) and for each source zone population to be unchanged. Figure 3.24 illustrates the general concept of the pycnophylactic interpolation. Given source zones with total populations (0 iterations), and with each source zone being represented by a different colour as shown in Figure 3.24. The source zone populations are reassigned by mass preserving

reallocation to remove abrupt changes in source zones boundaries. The technique computes continuous population density (per cell) in each source zone. The population density per cell is then smoothed repeatedly by replacing the value of each cell with the weighted average of its neighbours. The volume of the attributes within each source zone remains unchanged but varies smoothly at the boundaries. It is assumed that with 25 iterations as in Figure 3.24, a better representation of the variation is a smooth surface (Tobler 1979).

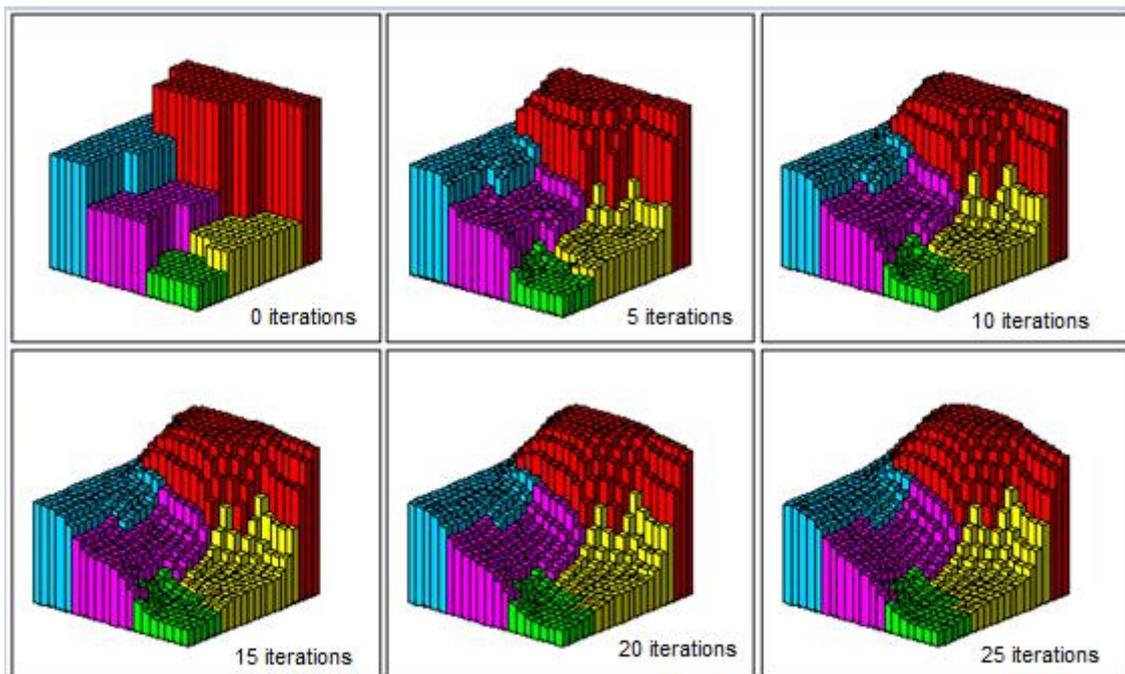


Figure 3.24 – The pycnophylactic interpolation method

(Source: Tobler, W. R., 1992)

The procedure for generating the pycnophylactic surface for the total populations of the source zone has been described by Qiu et al. (2012) and it involves: (1) converting the source zone data (vector) to raster using the *feature-to-raster* tool; (2) joining the value attribute table (VAT) of the raster source zone to the vector source zone to access source zone populations; (3) computing the population density per cell in the attribute table of the source zone (using each of 30m and 100m support grids) by dividing the population by the number of cells in each source zone using the field calculator; (4) calculating a new density by replacing the value of each cell with the weighted average of its neighbours using the *focal analysis tool*; (5) estimating the density for each source zone using the new per cell density (for each of 30m and 100m support grids) with the

*zonal statistics function*; (6) adjusting the new density by multiplying each cell value with the ratio between the original population and the estimated total population density of each source zone; (7) repeat steps 3-6 until no more adjustment is required (e.g. maximum change in any pixel density values between iterations falls below a threshold level), such that zone total equals original value (pyncnophylactic condition); (8) obtaining the estimated interpolated gridded population of each target zone by summing the adjusted population density of each cell falling within each target zone using the *zonal statistics tool*. The implementation steps described above are illustrated using a flowchart in Figure 3.25.

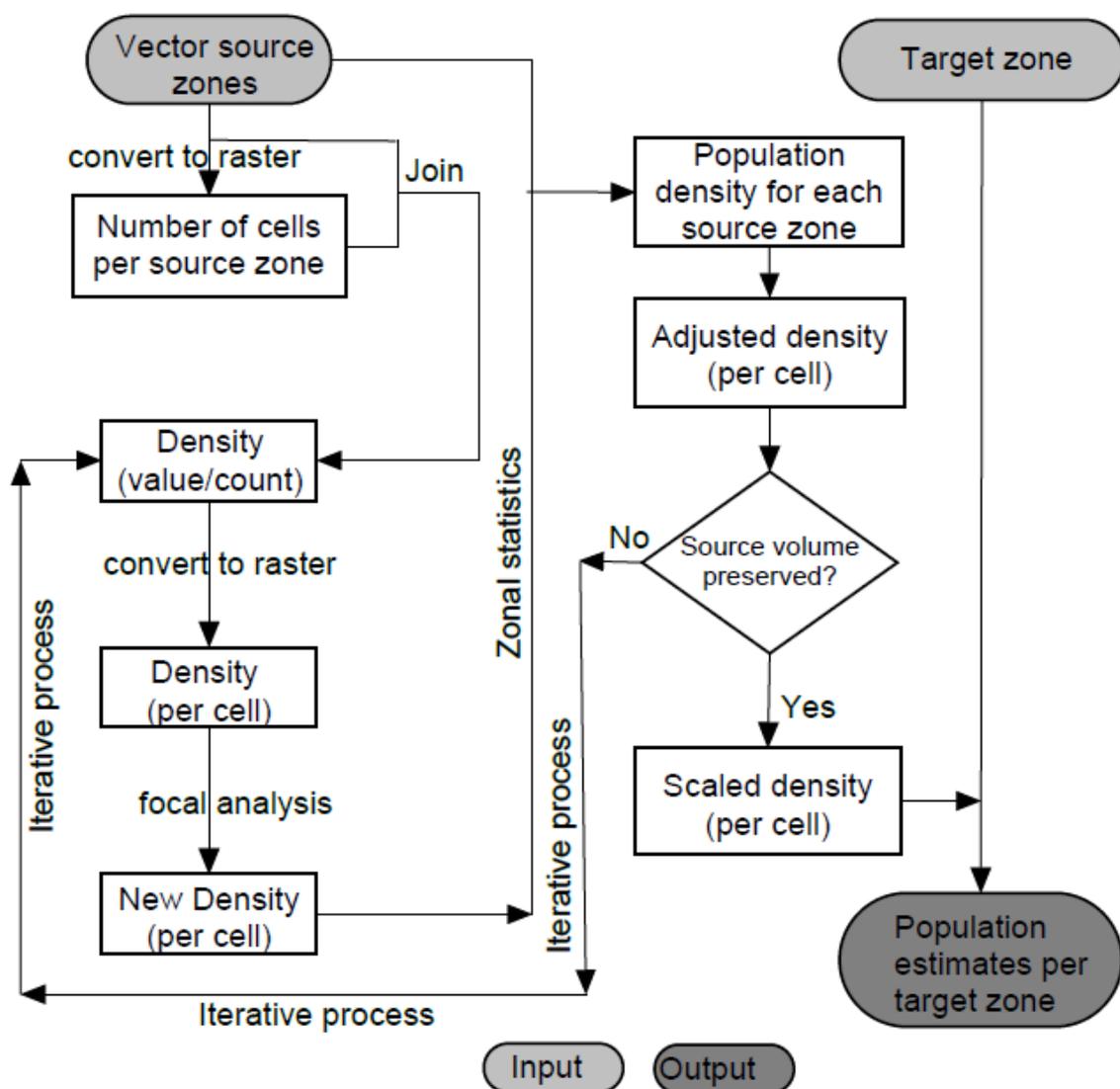


Figure 3.25 - A flow chart for the implementation steps for the pyncnophylactic interpolation method.

### 3.5 Evaluation of surfaces

The interpolated gridded pycnophylactic surfaces and the dasymetric population surfaces at 100m and 30m resolutions of the output grid (created in sections 3.4.1 and 3.4.2) were overlaid with the boundaries of MSOA, LSOA and OA target zones for Leicester and then aggregated to obtain estimates of the populations for these target zones, and assessed for accuracy using known census counts in each case. The digital boundaries for the target zones were overlaid with the interpolated gridded population surfaces based on spatial location and each polygon in the target zones was assigned a summary of all counts in the interpolated gridded population surfaces that intersect the polygon. The estimated populations in each target zone were then obtained and compared with known census counts for each target zone. Figure 3.26 shows an example where the boundaries of Leicester LSAOs were intersected with the interpolated gridded pycnophylactic surfaces at 100m resolution of the output grid and the estimates of the population at LSOA target zones were obtained. The same procedure was repeated to obtain population estimates for the three U.K. census units (MSOAs, LSOAs and OAs) for Leicester that were used as the target zones from the interpolated gridded pycnophylactic surfaces and the dasymetric population surfaces.

The residuals were calculated and mapped to visually explore the nature of the error as has been done in previous research (Langford 2013; Qiu et al. 2012; Mennis and Hultgren 2006; Eicher and Brewer 2001). The residual is calculated as the estimated population subtracted from the actual populations of each census unit. The accuracy of the interpolation is measured using the root mean squared error (RMSE) metric described by Fisher and Langford (1995) as has been done in previous research (Langford 2013; Tapp 2010; Mennis and Hultgren 2006; Gregory 2002; Eicher and Brewer 2001). The RMSE metric gives a summary of the error within census units. The RMSE metric was used to be able to compare between alternative methods applied to a common set of source and target units. The error within a given source zone (RMSE) uses absolute difference between estimated populations and the populations reported for the census units within each of the target zones and is calculated as in Equation 3.6.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2} \quad [3.6]$$

Where:  $X_i$  is the known census count at zone  $i$ ,  $Y_i$  is the estimated population from the interpolation at zone  $i$ , and  $n$  is the number of target zones.

The RMSE metric is “*less useful for comparing between different sets of source and target units*” (Langford 2013), particularly where resolution change is involved. This is because the RMSE metric is affected by the absolute size of estimated values (e.g. MSOA counts are as expected larger than LSOA counts and would have a larger RMSE values). Previous research (Eicher and Brewer 2001; Mennis and Hultgren 2006; Langford 2013) considers the variation in actual population of the target zones (e.g. MSOA and OA), and to account for these variations, the RMSE score is divided by the average known population of each target zone to obtain the coefficient of variance (CoV). The CoV provides a relative error metric suitable for comparing values across the target zones. This is a useful metric as this research seeks to test performance over census areas of differing resolutions and CoV is more appropriate for cross-resolution comparisons. The CoV is calculated as in Equation 3.7.

$$\text{CoV} = \frac{\text{RMSE}}{\bar{x}} \quad [3.7]$$

Where:  $\bar{x}$  is the mean population of the known census count for each target zone.

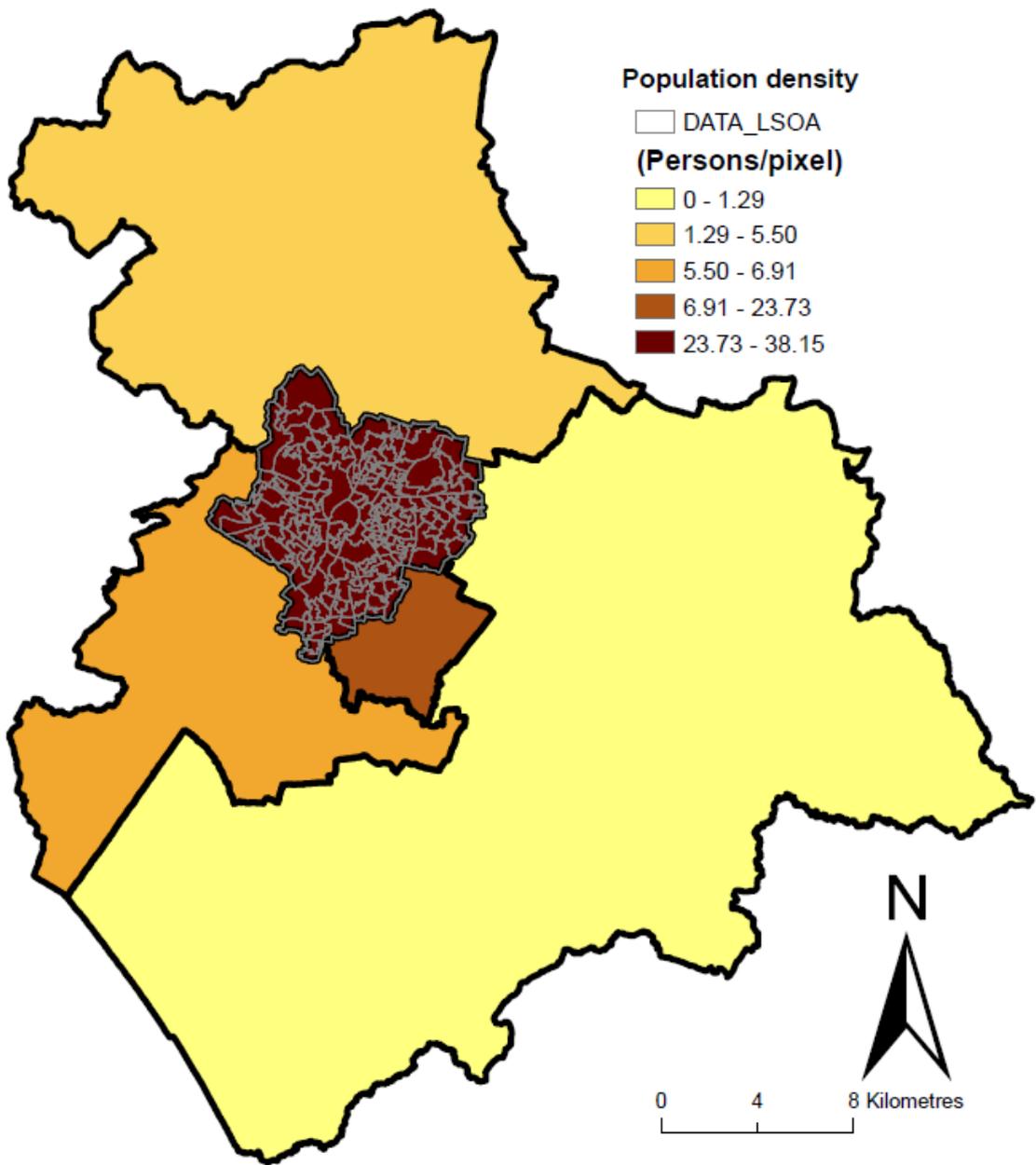
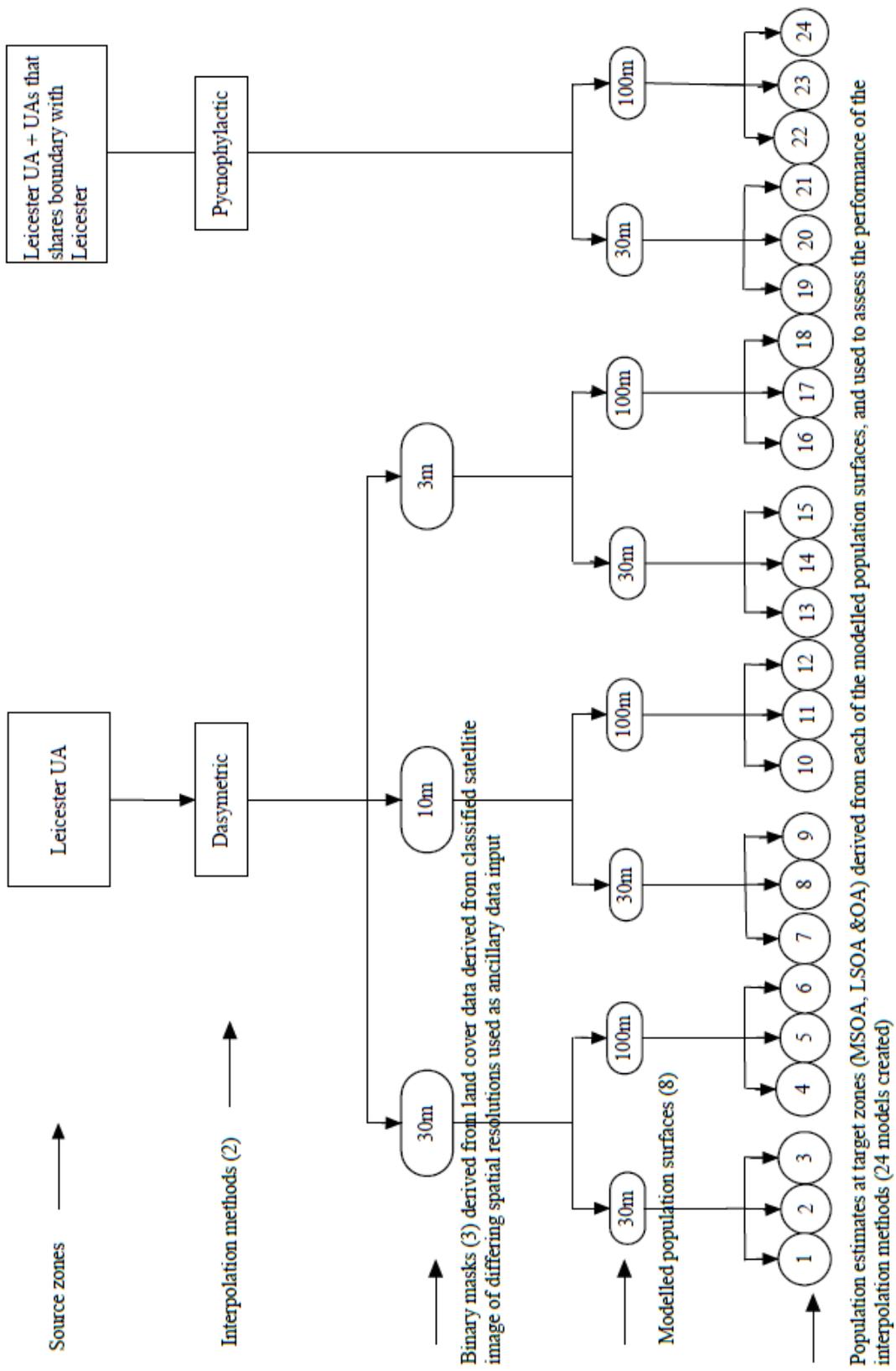


Figure 3.26 - Leicester LSOAs intersect interpolated gridded pycnophylactic surfaces at resolutions of 100m postings

### 3.6 Summary

This chapter has presented the methodology used to redistribute aggregate population census data for Leicester over small areas. Figure 3.27 presents a summary chart that provides a visual framework for all of the various models created and interpolations undertaken for Leicester study area. Eight modelled population surfaces for Leicester were generated using different combinations of land cover derived from classified imagery and source and target zone sizes. Two of these are pycnophylactic surfaces created using 30m and 100m support grids. The remaining six are dasymetric surfaces created using; a binary mask derived from land cover data derived from classified satellite image of 30m spatial resolution used as the ancillary data input at 30m and 100m support grids postings; and a binary mask derived from land cover data of 10m and 3m spatial resolution each derived from classified resampled aerial photograph data used as the ancillary data input at 30m and 100m support grids postings. This was done to explore if different land cover data of both different sources and resolutions of the same source would reveal spatial variations within the study area, and also to explore whether a more detailed ancillary data input could reduce the estimation error (or not), as suggested by Zandbergen and Ignizio (2010). The estimated interpolated gridded population of each target zone (MSOA, LSOA and OA) were summed to obtain the estimated populations for each of the target zones. The estimated populations for these target zones were compared with known census counts of the same target zones to assess the performance of the interpolation methods. The results of the 24 experiments are presented in Chapter Four together with the error values for the simple areal weighting interpolation method. The simple areal weighting method defines the ‘lowest common denominator in terms of methodological sophistication and acts as a useful benchmark against which other techniques may be measured’ (Langford 2013). The most appropriate parameters obtained from Leicester study were applied to the Port-Harcourt case study.



Population estimates at target zones (MSOA, LSOA & OA) derived from each of the modelled population surfaces, and used to assess the performance of the interpolation methods (24 models created)

Figure 3.27 - A visual framework for all of the various models created and interpolations undertaken for Leicester study area.

## Chapter 4

### 4. RESULTS: LEICESTER

#### 4.1 Introduction

This chapter presents the results of areal interpolation for Leicester implemented in Chapter Three. The summary population totals for Leicester (from U.K. 2001 census) were disaggregated using two approaches: binary dasymetric and pycnophylactic interpolation. The binary dasymetric method used the Leicester UA as the source zone while the pycnophylactic interpolation method used the Leicester UA together with all the surrounding UAs with which it shares a common boundary as source zones. The binary dasymetric method is based on ancillary data (such as land use) that controls the redistribution to only built-up areas. A binary mask derived from land cover data derived from classified satellite image of 30m spatial resolution and a binary mask derived from land cover data derived from classified resampled aerial photo of 10m and 3m spatial resolutions was each used as the ancillary data input. These source zones were used to create modelled population surfaces at resolutions of 100m and 30m, and then aggregated to MSOA, LSOA and OA target units and assessed for accuracy using known census counts in each case. The accuracy of the interpolation techniques used was measured using RMSE and CoV as suggested by Langford (2013).

The next section presents the results of supervised classification. Section 4.3 presents the results of areal interpolations at different spatial scales for Leicester. Section 4.4 presents results of evaluation of surfaces. The last section (4.5) provides a summary of the results for Leicester indicating the most appropriate parameters to apply to the Port-Harcourt case study.

#### 4.2 Supervised classification

The remotely sensed images acquired for Leicester and described in section 3.3.1 were processed to identify the extent of built-up areas that were used as the ancillary data input for the binary dasymetric method. This section presents the classified images

produced and the binary classified images created from the classified images, to show only built-up and non-built-up areas.

#### 4.2.1 Classified images

The classified images obtained from supervised classification described in section 3.4.1 are shown in Figures 4.1 to 4.3. Figure 4.1 shows a classified image using Landsat7 (ETM) 30m spatial resolution. The classified images using resampled aerial photo data of 10m and 3m spatial resolutions are shown in Figures 4.2 and 4.3 respectively. The accuracy of classification was assessed by comparing 256 random points comparing certain pixels in the classified image to reference pixels for which the class is known.

Table 4.1 presents the comparison of the overall accuracy and Kappa statistic between the different land cover data of both different sources and resolutions of the same source. The accuracy reports indicate a good agreement between thematic maps generated from image and the reference data when Landsat7 (ETM) 30m spatial resolution was used. The classification accuracy is above the minimum standard of digital image classification for optical remote sensing data (85%) recommended by Paul (1991) and Jansen et al. (2008) while the classification accuracy recorded from resampled aerial photo data of both 10m and 3m spatial resolution imagery were slightly below the minimum standard of digital image classification for optical remote sensing data (see Table 4.1).

Table 4.1 - Comparison of the overall accuracy and Kappa statistic between the different land cover data

Land cover data	Overall accuracy (%)	Kappa accuracy (%)	Kappa coefficient
30m spatial resolution	87.89	81.08	0.81
10m spatial resolution	83.20	73.07	0.73
3m spatial resolution	82.03	71.64	0.72

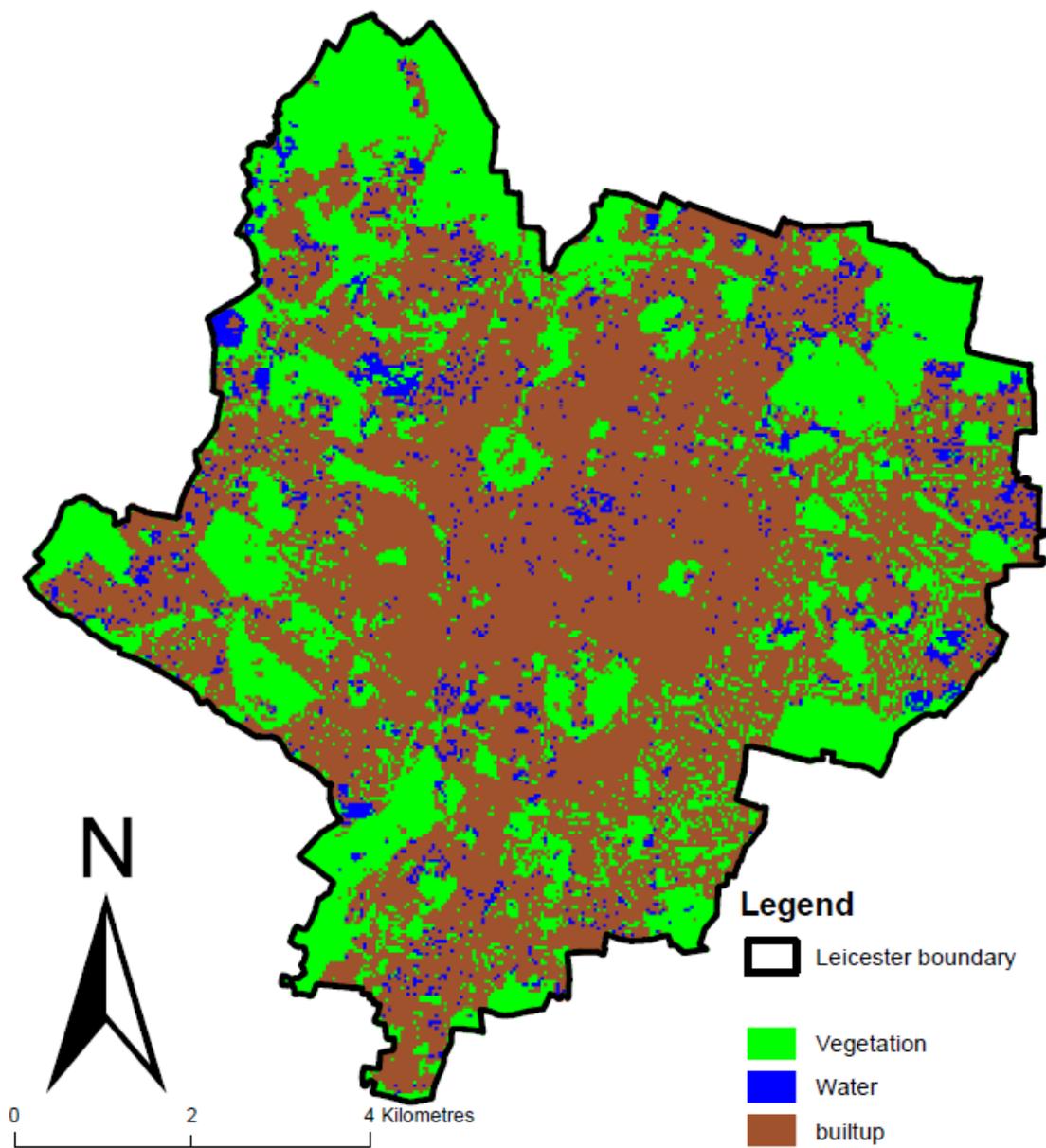


Figure 4.1 – The classified Leicester image derived from Landsat7 (ETM) 30m spatial resolution. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

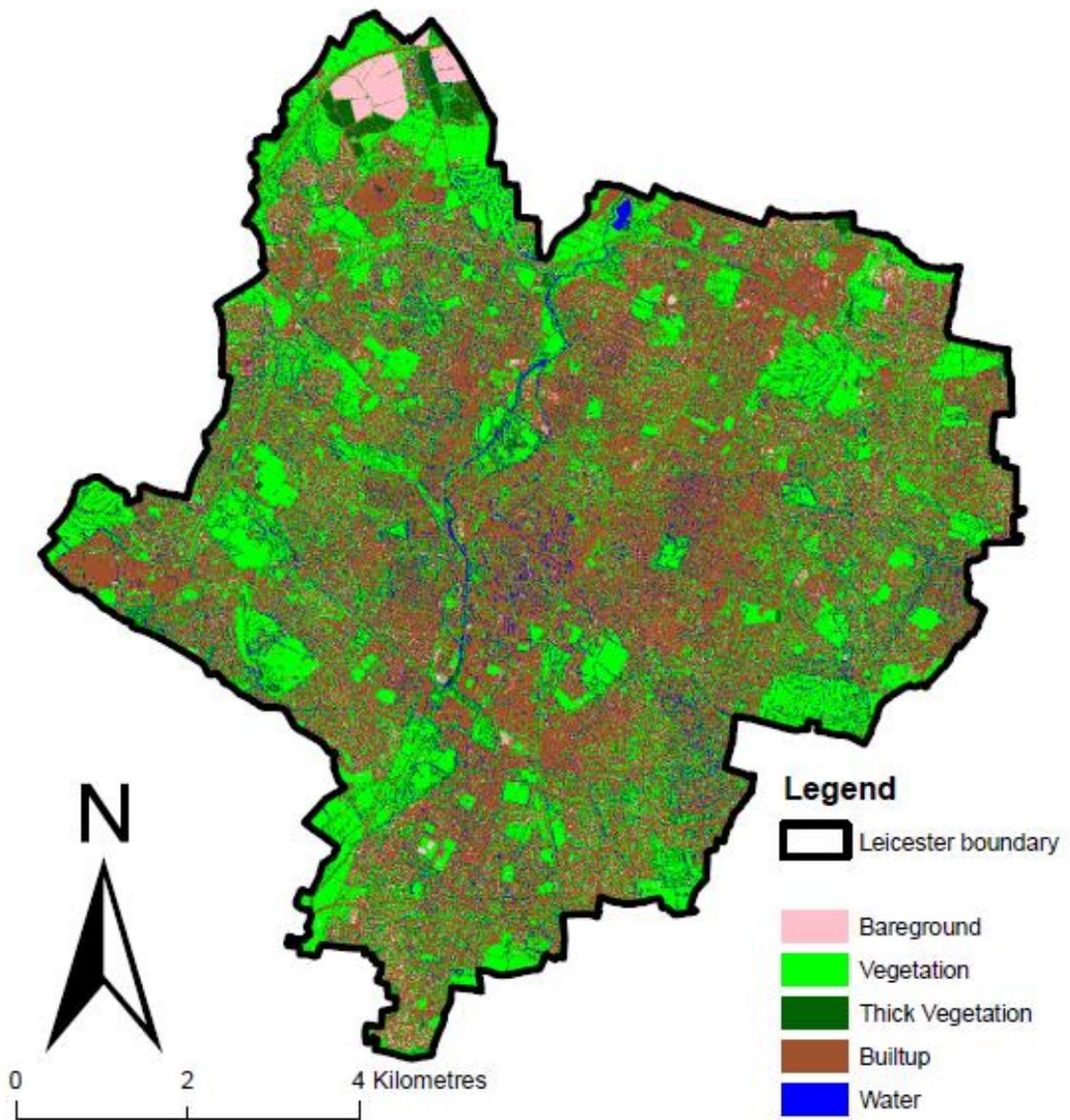


Figure 4.2 - The classified Leicester image derived from resampled aerial photo data of 10m spatial resolution. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

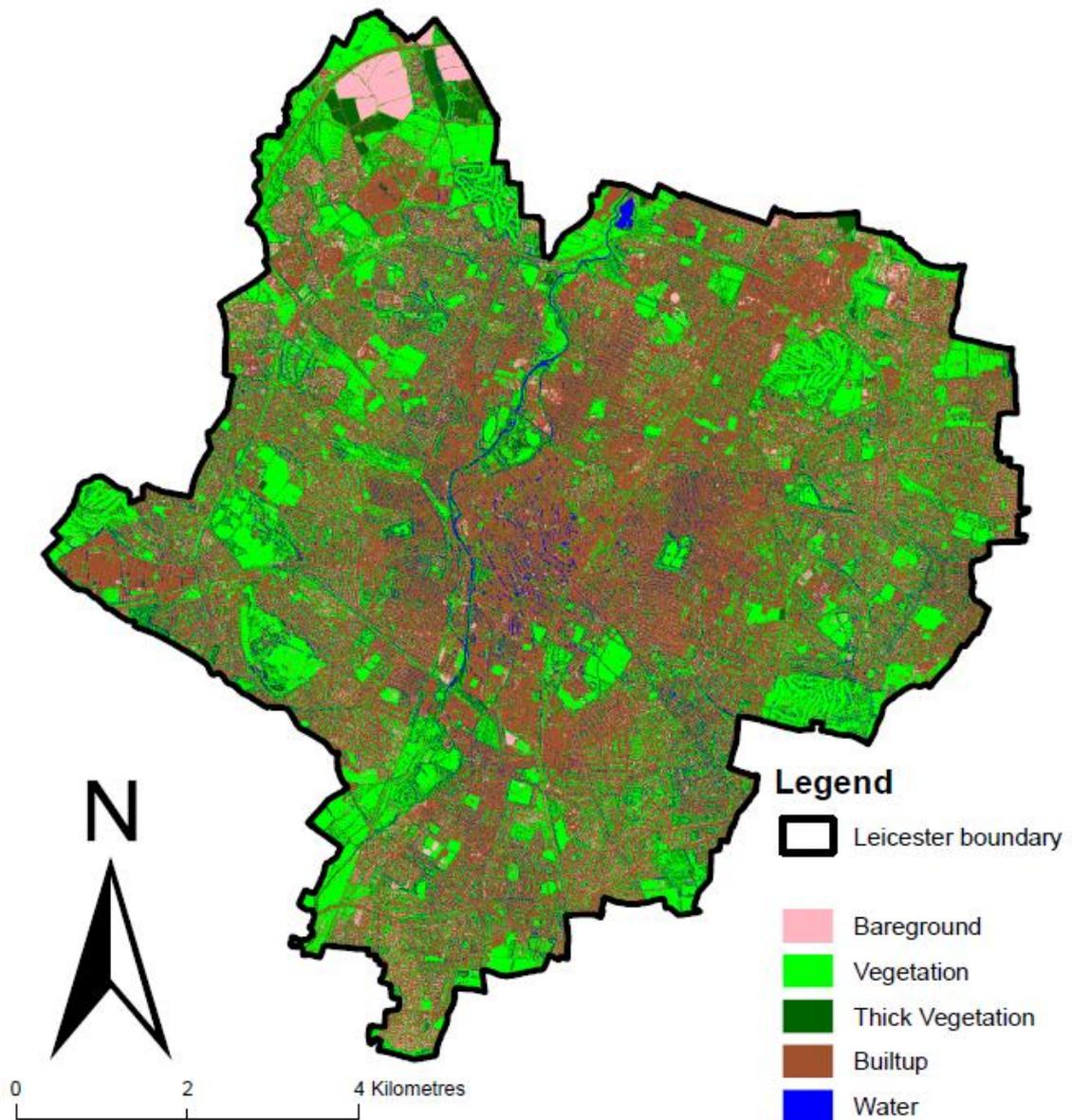


Figure 4.3 - The classified Leicester image derived from resampled aerial photo data of 3m spatial resolution. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

#### 4.2.2 Binary classified images

The binary dasymetric method uses additional geographic information (e.g. land use) to divide the source zone into built-up and non-built-up areas. The classified images presented in section 4.2.1 were reclassified into a simple binary division by assigning a weighting factor of 1 to all areas classified as built-up areas and 0 to all other areas (as described in section 3.4.1). The binary masks derived from land cover data derived from classified satellite images of both different sources and resolutions of the same source which represent the underlying spatial distribution of population within the source zone, are presented in Figures 4.4 to 4.6. Figure 4.4 shows a binary mask derived from land cover data derived from classified Landsat7 (ETM) 30m spatial resolution image data. Figures 4.5 and 4.6 show binary masks derived from land cover data derived from classified resampled aerial photo data of 10m and 3m spatial resolutions respectively.

Table 4.2 compares the sizes of the total built-up areas in the source zone derived from different land cover data. The binary classified images showed that built-up areas derived from land cover data derived from classified Landsat7 (ETM) 30m spatial resolution has a larger area size compared to those derived from classified resampled aerial photo data of 10m and 3m spatial resolutions. The results showed that the area classified as built-up reduces in size as the resolution of land cover data increases. This is because 10m and 3m spatial resolution images appear to offer greater spatial precision in the depiction of different land cover types compared to the 30m image. The results do not show much difference between the 10m and 3m spatial resolutions resampled aerial photo data. One possible reason for this could be because they are both from the same source.

Table 4.2 - Comparison of the total built-up area in the source zone between the different land cover data

Built-up areas	Total source zone area (km <sup>2</sup> )
Landsat7 (ETM) 30m spatial resolution	45.31
Resampled aerial photo data 10m spatial resolution	35.94
Resampled aerial photo data 3m spatial resolution	35.57

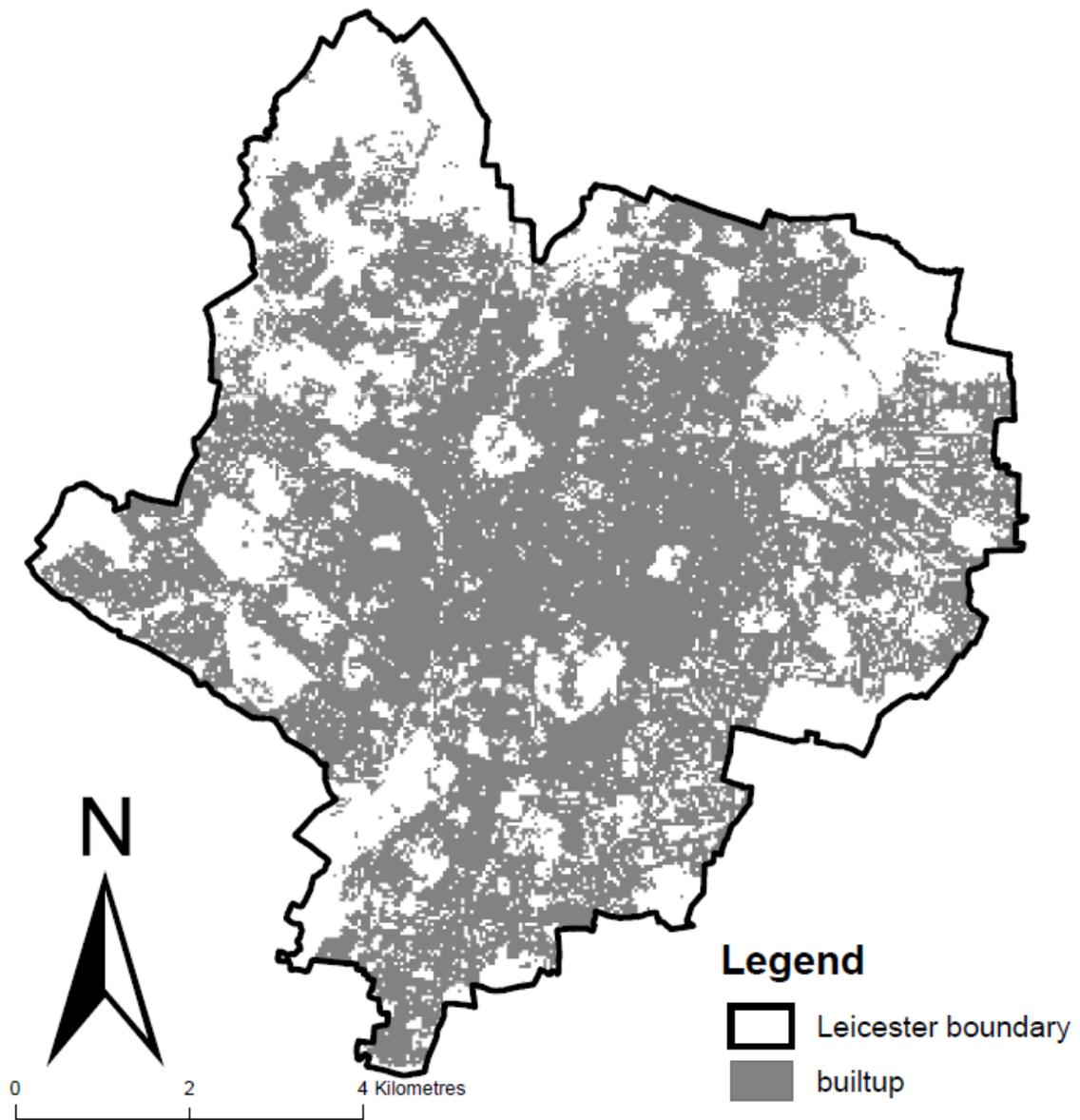


Figure 4.4 – A binary mask derived from land cover data derived from classified Landsat7 (ETM) 30m spatial resolution. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

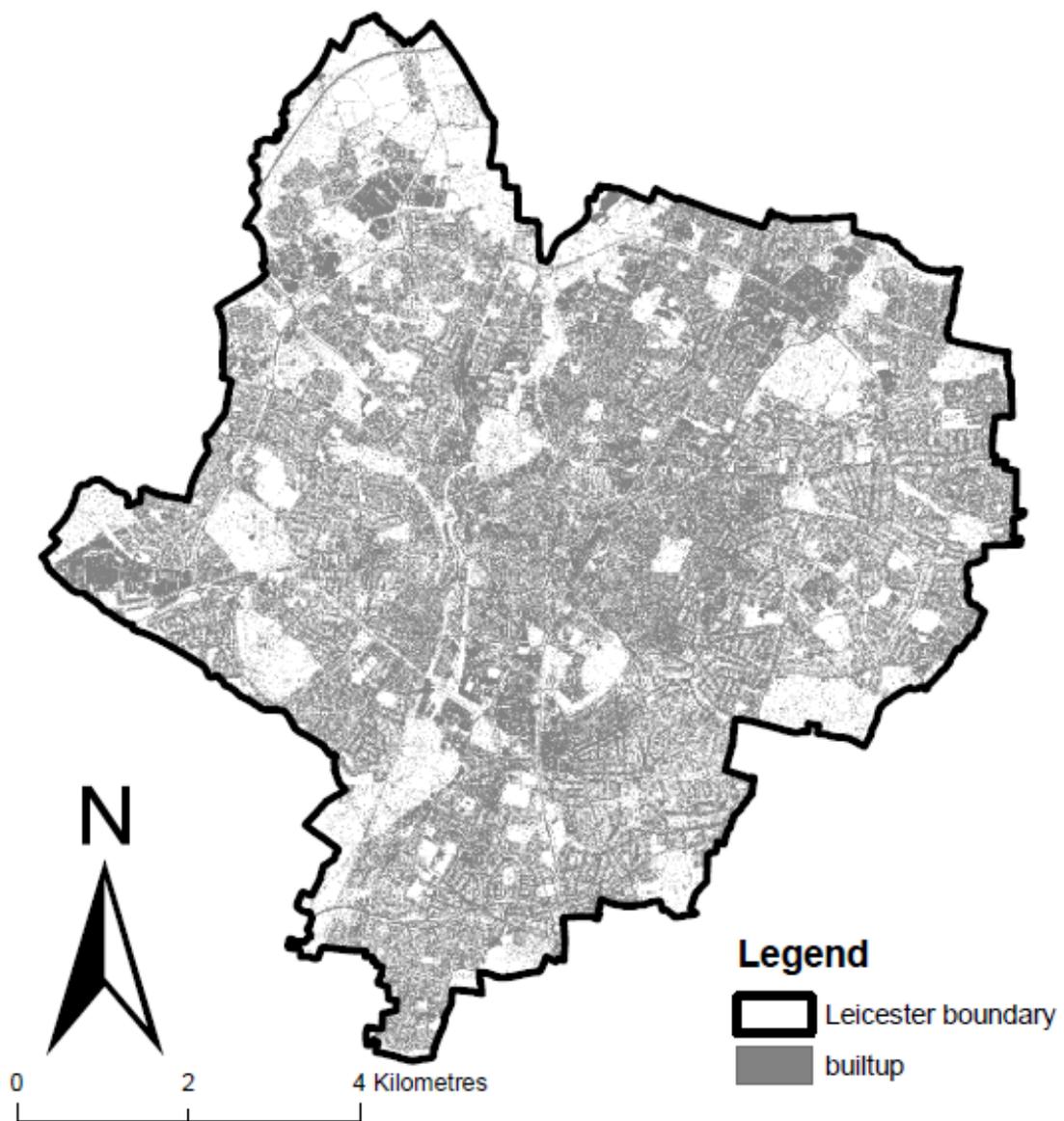


Figure 4.5 – A binary mask derived from land cover data of 10m spatial resolution derived from classified resampled aerial photo data. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

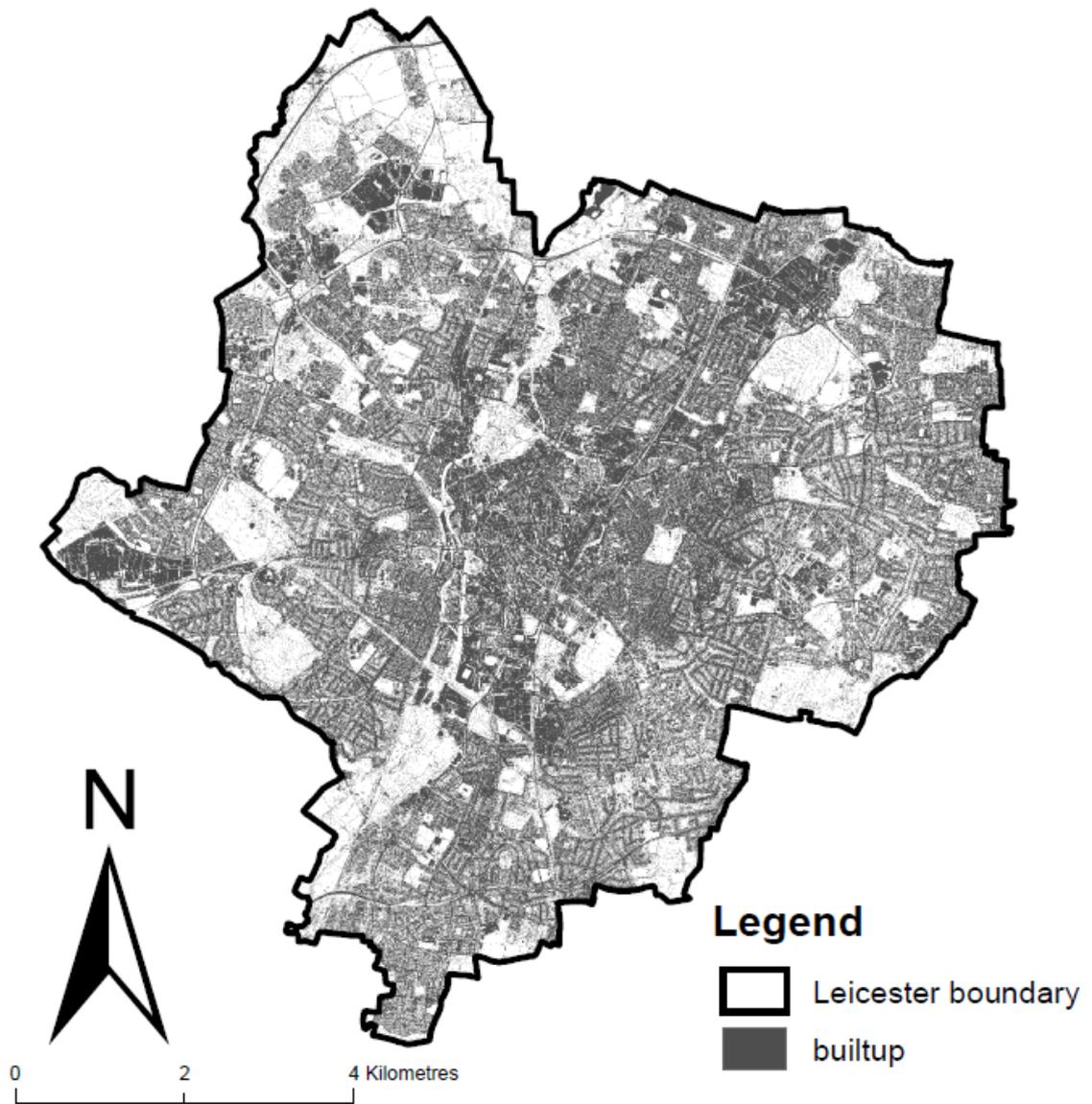


Figure 4.6 - A binary mask derived from land cover data of 3m spatial resolution derived from classified resampled aerial photo data. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

### 4.3 Areal interpolation methods

The binary dasymetric and the pycnophylactic interpolation methods were tested across different spatial scales. The differences in the assumptions made by the two methods and the scales of disaggregation used in the implementation of the methods generated different target zone estimates of the population.

#### 4.3.1 The binary dasymetric method

The binary dasymetric method described in section 3.4.2 uses land cover data derived from classified satellite imagery as ancillary masks to spatially constrain the re-allocation, as has been done by many authors. In this study, three different land cover data of 30m, 10m and 3m spatial resolutions were used as the ancillary data input for the binary dasymetric method and square grids of 30m and 100m were each used to create the modelled population surfaces. Leicester UA was used as the single source zone for the binary dasymetric method. The population density for each of 30m, 10m and 3m spatial resolution land cover data used as the ancillary data input was calculated as a single uniform density estimate across Leicester.

Table 4.3 shows the population densities (persons/10,000 m<sup>2</sup>) for the different ancillary data input used for the binary dasymetric method. The results show that land cover data derived from classified Landsat7 (ETM) 30m spatial resolution when used as the ancillary data input provided a lower population density compared to 10m and 3m spatial resolution land cover data derived from classified resampled aerial photo data. The 10m and 3m spatial resolution land cover data derived from classified resampled aerial photo data show small difference in population density across Leicester. This is not surprising because they are both from the same source.

Table 4.3 - Population density per 10,000 m<sup>2</sup> for binary dasymetric maps of population

Ancillary data input	Population density (persons/10,000 m <sup>2</sup> )
Landsat7 (ETM) 30m spatial resolution	6.178
Resampled aerial photo data 10m spatial resolution	7.788
Resampled aerial photo data 3m spatial resolution	7.870

The population densities for each of the support grids (30m and 100m) were aggregated to create the final population surface layers across Leicester. The dasymetric population surfaces created using 30m, 10m and 3m ancillary data input are shown in Figures 4.7, 4.8 and 4.9 respectively as representative example maps of the dasymetric population surfaces.

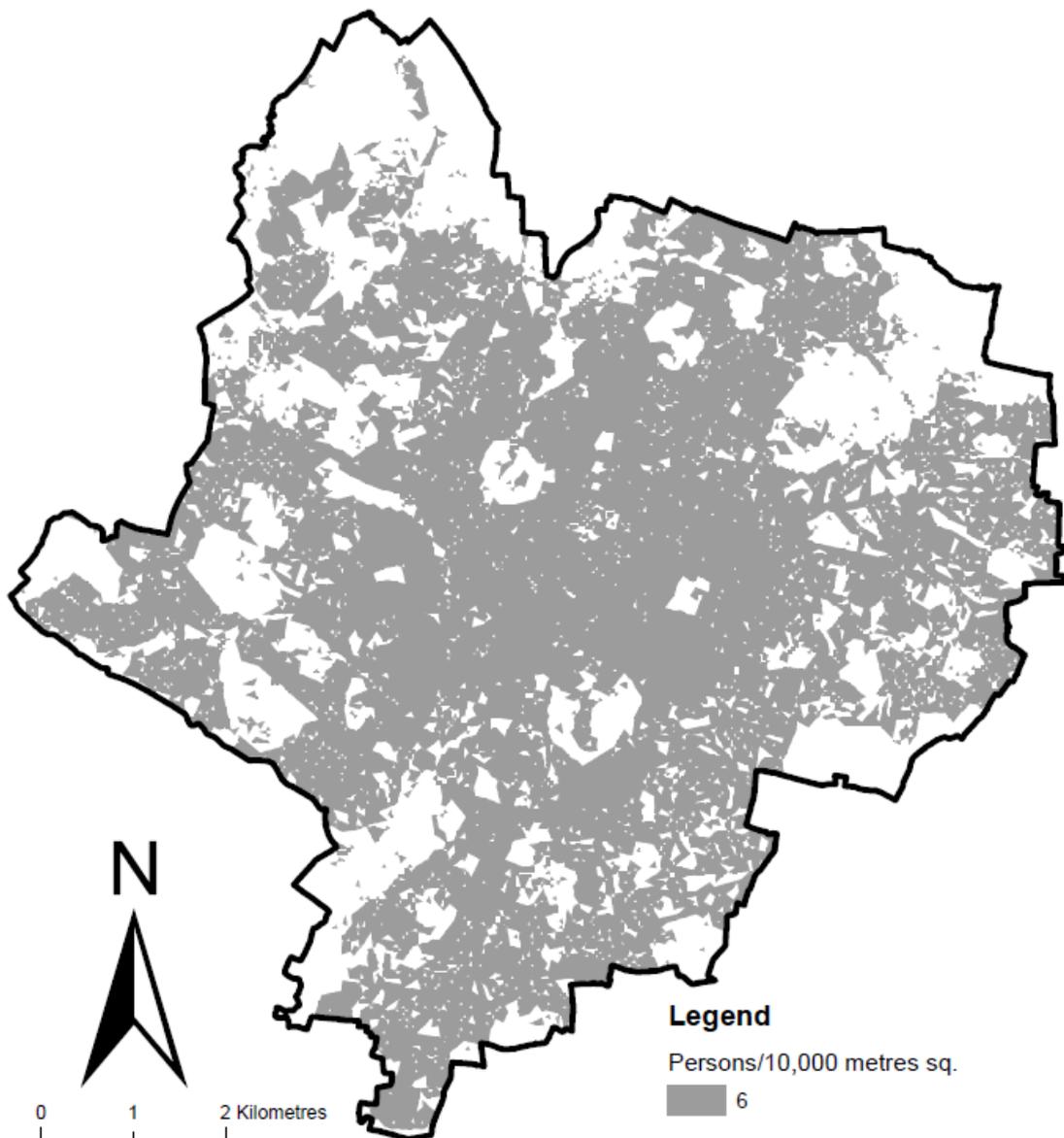


Figure 4.7 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified Landsat7 (ETM) 30m spatial resolution ancillary data input.

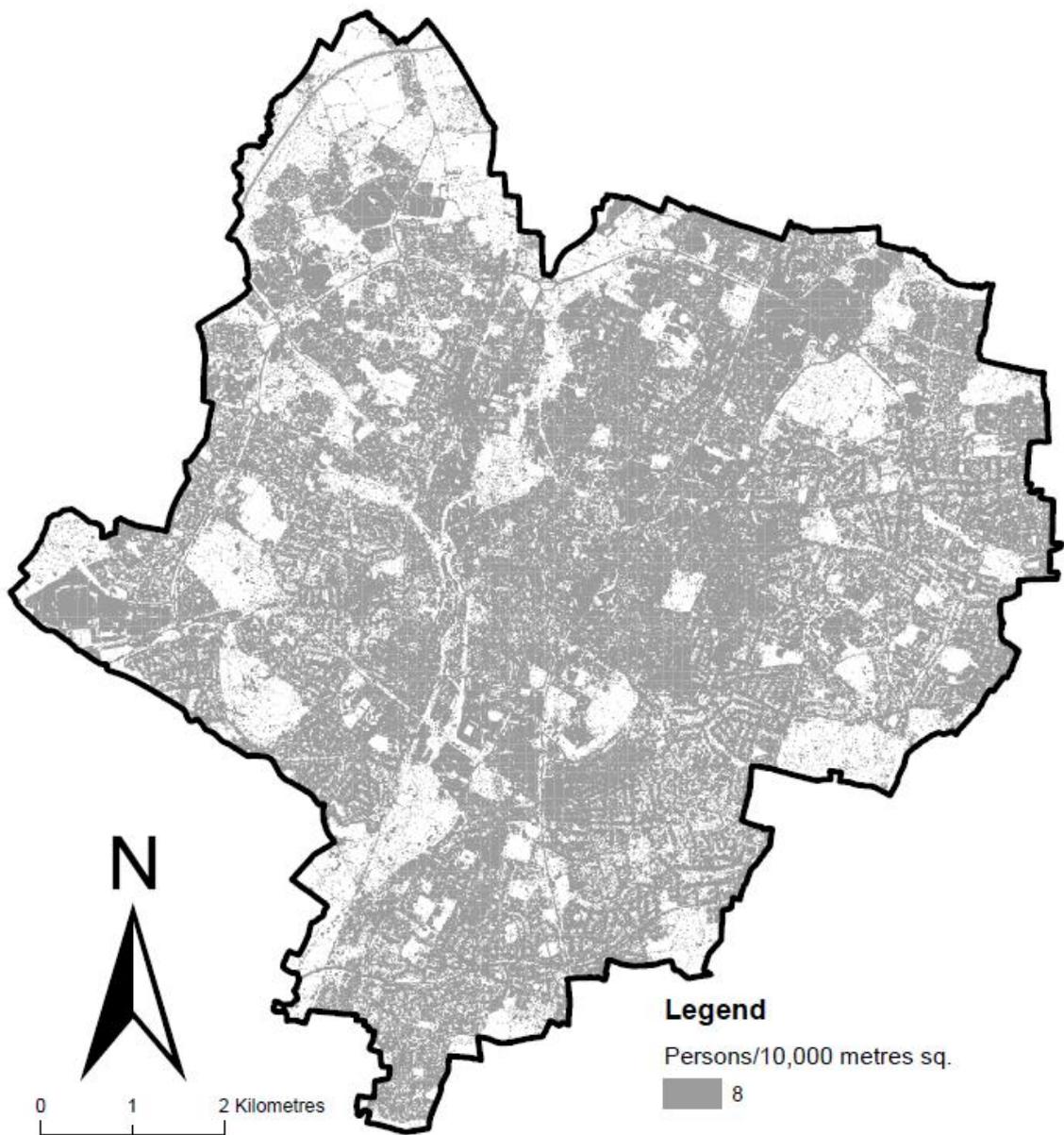


Figure 4.8 - The dasymetric map of population surface at 100m posting created using land cover data derived from classified resampled aerial photo data of 10m spatial resolution ancillary data input.

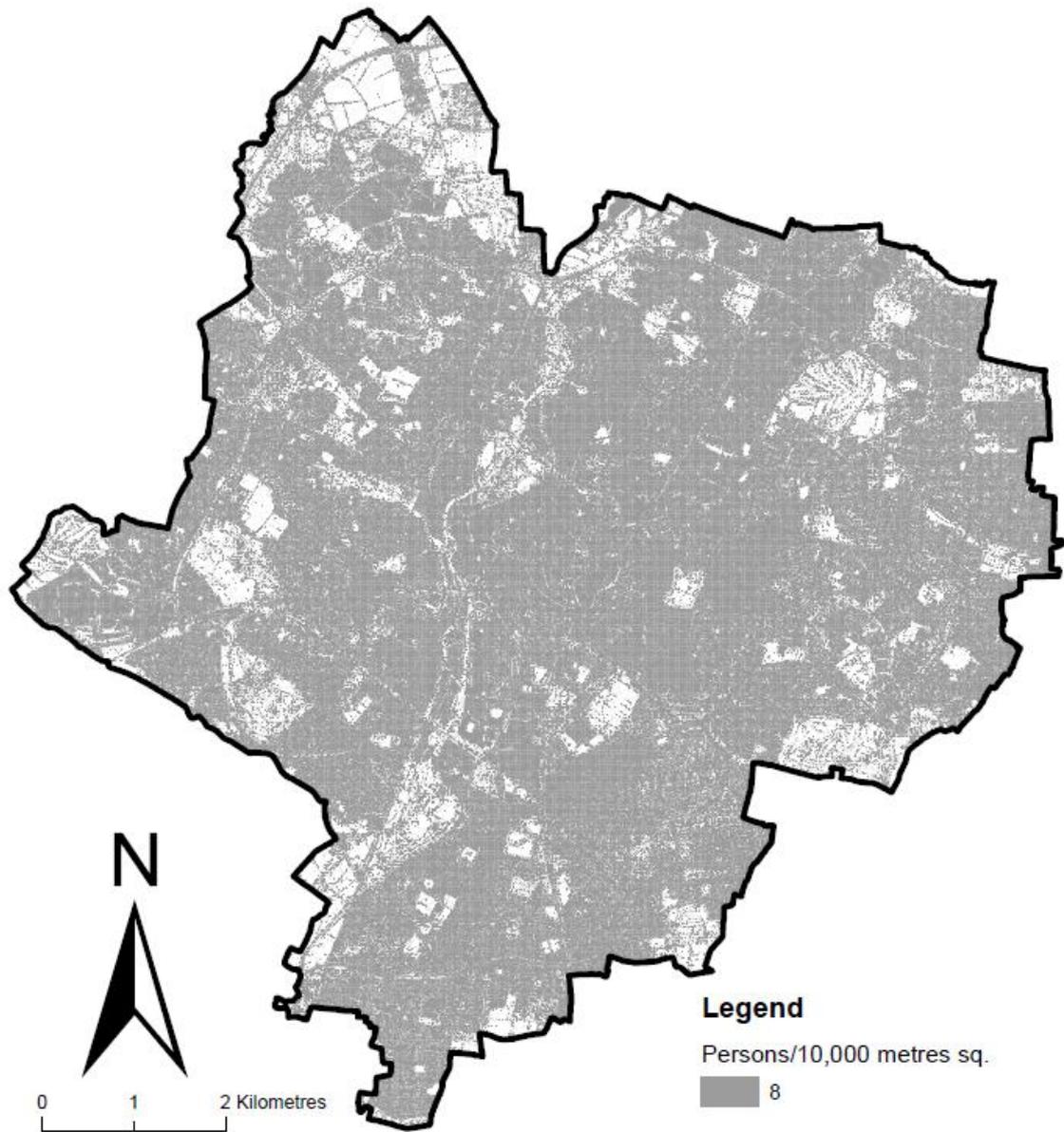


Figure 4.9 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified resampled aerial photo data of 3m spatial resolution ancillary data input.

### 4.3.2 Pycnophylactic interpolation method

The modelled population surfaces generated for Leicester are raster grids of 30m and 100m resolutions as shown in Figures 4.10 and 4.11 respectively. The results show gradual density change around the source zone boundaries by applying a smooth density function across boundaries. The value of each of the cells is shown as a continuous surface in five classes ranging from low population density (shown in green) to high population density (shown in red). The modelled population surfaces do not represent a more realistic smooth distribution of population density. This is because the pycnophylactic method does not draw on further ancillary information about the underlying population distribution in the source zone.

Figure 4.10 shows the change in population density at 30m resolution output grid. The figure shows a smooth and gradual density for the population density in the interpolated gridded population surface at 30m resolution of the output grid. High population density values are seen across the city of Leicester and most parts of Oadby and Wigston district, and gradually decreased to other source zones with low populations. The effect of source zone area size on interpolated gridded population surface is seen in the Harborough district. Low population density values are seen across the Harborough district. This is because it is the largest source zone in terms of area size and the second in terms of low population counts.

Figure 4.11 shows the change in population density at 100m resolution output grid. The figure shows a smooth and gradual density for the population density in the interpolated gridded population surface at 100m resolution of the output grid. High population density values are seen in most parts of the city of Leicester and gradually decreased to other source zones with low populations. Similar to interpolated gridded population surface at resolution of 30m postings, Harborough district has the lowest population density values amongst the five districts used as the source zones.

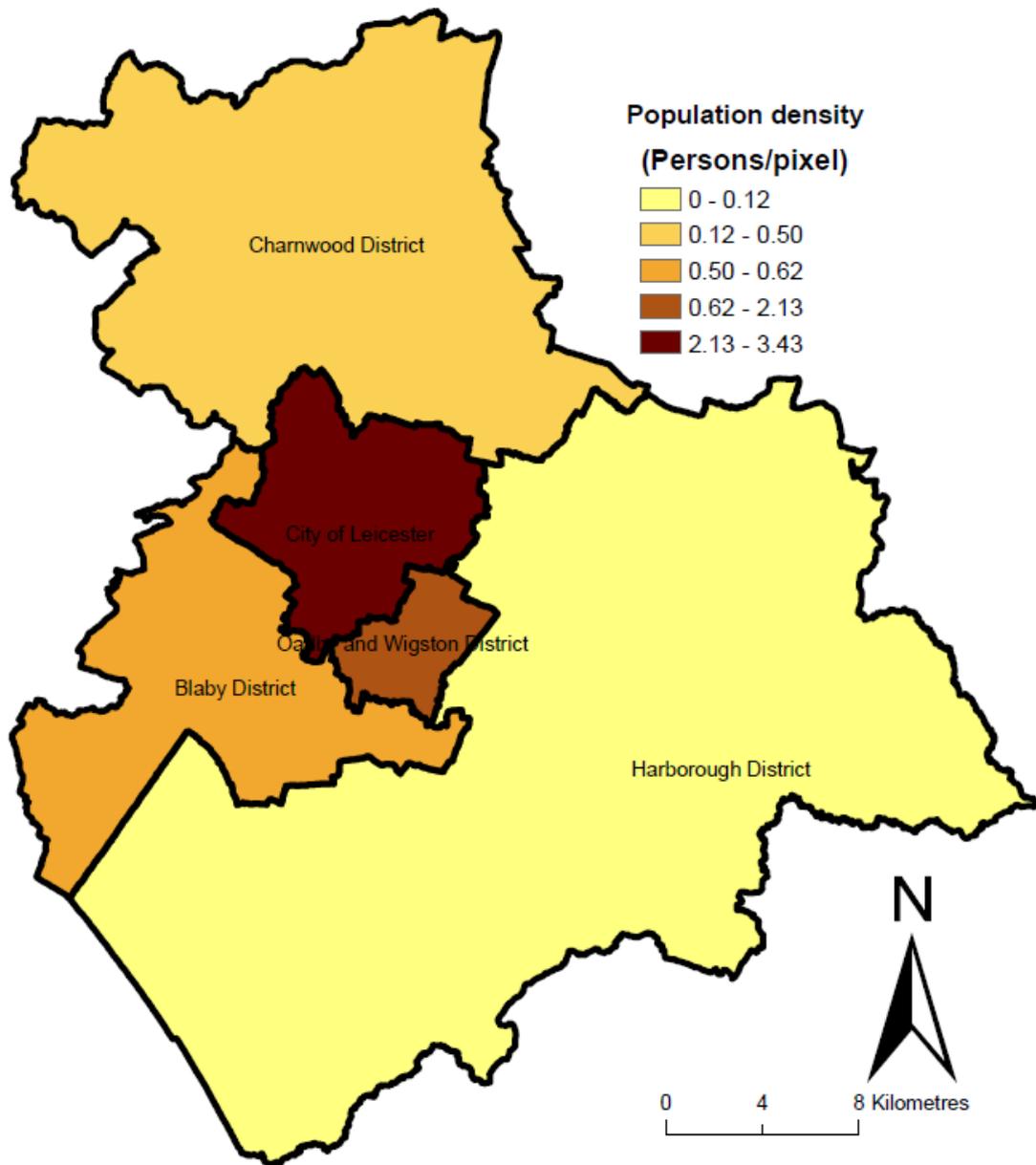


Figure 4.10 – The change in population density at 30m resolution output grid. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

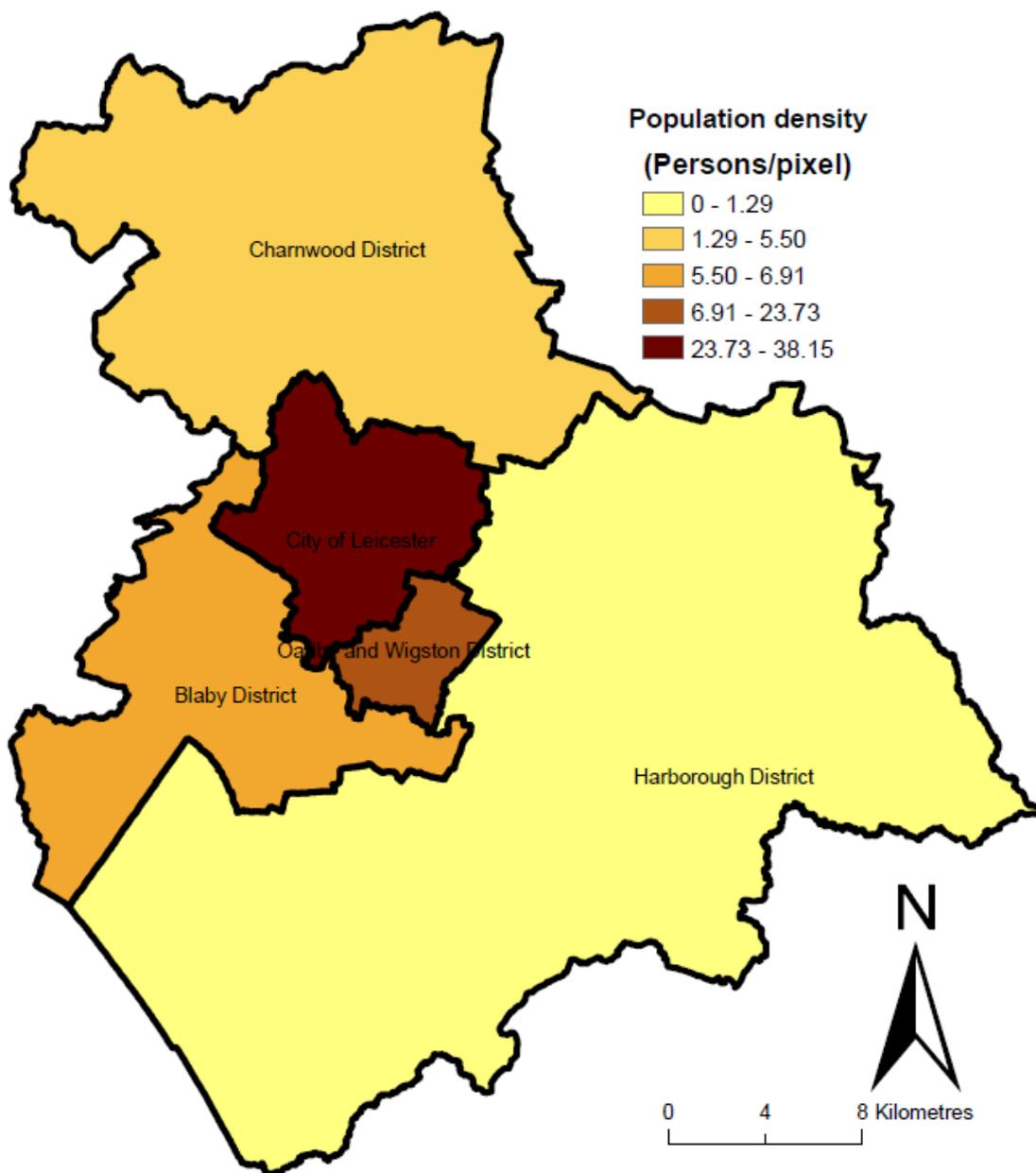


Figure 4.11 - The change in population density at 100m resolution output grid. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

#### 4.4 Evaluation of Surfaces

The estimated populations for the target zones were compared with known census counts of the same target zones to assess the performance of the interpolation methods. Performance measures for the various experiments undertaken are summarized in six tables. The error values for the simple areal weighting interpolation method which makes the assumption that source zone population is evenly spatially distributed within the zone boundary were also added to be able to see improvements made by the dasymetric mapping and the pycnophylactic interpolation methods. The simple areal weighting method defines the ‘lowest common denominator in terms of methodological sophistication and acts as a useful benchmark against which other techniques may be measured’ (Langford 2013). The results are presented in Tables 4.4 to 4.9 in order of increasing accuracy, as expected, areal weighting performs least well in all the models tested, setting a benchmark against which dasymetric mapping and pycnophylactic interpolations may be assessed. Tables 4.4 and 4.5 show interpolation results from Leicester UA to 30m and 100m square grids postings respectively, aggregated at MSOA. Tables 4.6 and 4.7 show interpolation results from Leicester UA to 30m and 100m square grids postings respectively, aggregated at LSOA. Tables 4.8 and 4.9 show interpolation results from Leicester UA to 30m and 100m square grids postings respectively, aggregated at OA. The accuracy of the interpolations is measured using the RMSE metric and CoV.

Table 4.4 shows the interpolation results from Leicester UA to 30m square grids postings, aggregated at MSOAs in order of increasing accuracy. The pycnophylactic method performed least well with RMSE score of 4233.4 and a CoV of 0.544. Interpolations using binary dasymetric with classified land cover data used as the ancillary data input returns significant improvement compared to that recorded using the pycnophylactic method. At this resolution of interpolation, the binary dasymetric methods tested using ancillary data input of differing spatial resolutions recorded slightly different CoV scores. The binary dasymetric model using land cover data derived from classified Landsat7 (ETM) 30m spatial resolution as the ancillary data input provided the best estimates among the methods tested with the lowest recorded RMSE of 2943.8 and a CoV of 0.379. The most striking feature in Table 4.4 is that the binary dasymetric model using land cover data derived from classified resampled aerial

photo data of 10m spatial resolutions as the ancillary data input achieved a RMSE values of 3314.4 which marginally improves to 3304.9 compared to a binary dasymetric model using land cover data derived from classified resampled aerial photo data of 3m spatial resolutions as the ancillary data input. This is surprising because higher resolution land cover data that offer greater spatial precision in the depiction of building locations does not automatically improves interpolation performance. One possible reason for this could be because they are both from the same source.

Table 4.4 - Interpolation results from Leicester UA to 30m square grids postings, aggregated at MSOA

Interpolation method	RMSE	CoV
Areal weighting using zone boundaries only	4486.9	0.577
Pycnophylactic interpolation	4233.4	0.544
Binary dasymetric using 10m resolution classified land cover	3314.4	0.426
Binary dasymetric using 3m resolution classified land cover	3304.9	0.425
Binary dasymetric using 30m resolution classified land cover	2943.8	0.379

*Note:* Mean population of target units is 7776.

Table 4.5 - Interpolation results from Leicester UA to 100m square grids postings, aggregated at MSOA

Interpolation method	RMSE	CoV
Areal weighting using zone boundaries only	4934.1	0.635
Pycnophylactic interpolation	3974.9	0.511
Binary dasymetric using 3m resolution classified land cover	3668.7	0.472
Binary dasymetric using 10m resolution classified land cover	3661.5	0.471
Binary dasymetric using 30m resolution classified land cover	3579.7	0.460

*Note:* Mean population of target units is 7776.

Table 4.5 shows the interpolation results from Leicester UA to 100m square grids postings, aggregated at MSOAs in order of increasing accuracy. The results are similar to that reported in Table 4.4 with the pycnophylactic method performing least well and the binary dasymetric model using land cover data derived from classified Landsat7 (ETM) 30m spatial resolution as the ancillary data input providing better target zone

estimates at this resolution of interpolation. At MSOA, interpolations to 30m square grids are better compared to 100m square grids. Also, different from Table 4.4, the RMSE value and CoV recorded for the binary dasymetric model using land cover data derived from classified resampled aerial photo data of 10m spatial resolutions as the ancillary data input marginally improves those recorded for the binary dasymetric model using land cover data derived from classified resampled aerial photo data of 3m spatial resolutions as the ancillary data input.

Table 4.6 - Interpolation results from Leicester UA to 30m square grids postings, aggregated at LSOA

Interpolation method	RMSE	CoV
Areal weighting using zone boundaries only	1497.8	1.001
Pycnophylactic interpolation	1368.5	0.914
Binary dasymetric using 3m resolution classified land cover	1173.9	0.784
Binary dasymetric using 10m resolution classified land cover	1155.6	0.772
Binary dasymetric using 30m resolution classified land cover	1087.5	0.726

*Note:* Mean population of target units is 1497.

Table 4.7 - Interpolation results from Leicester UA to 100m square grids postings, aggregated at LSOA

Interpolation method	RMSE	CoV
Areal weighting using zone boundaries only	1805.3	1.206
Pycnophylactic interpolation	1517.7	1.014
Binary dasymetric using 3m resolution classified land cover	1467.5	0.980
Binary dasymetric using 10m resolution classified land cover	1436.1	0.959
Binary dasymetric using 30m resolution classified land cover	1309.4	0.875

*Note:* Mean population of target units is 1497.

Tables 4.6 and 4.7 show the interpolation results from Leicester UA to 30m and 100m square grids postings respectively, aggregated at LSOAs in order of increasing accuracy. The results recorded have similar pattern to that recorded in Table 4.5. The pycnophylactic method performed least well with RMSE of 1368.5 and 1517.7 (for 30m and 100m postings respectively) and CoV of 0.914 and 1.014 (for 30m and 100m

postings respectively). The binary dasymetric model using land cover data derived from classified Landsat7 (ETM) 30m spatial resolution as the ancillary data input provided the best target zone estimates at this resolution of interpolation with a recorder RMSE of 1087.5 and 1309.4 (for 30m and 100m postings respectively) and CoV of 0.726 and 0.875 (for 30m and 100m postings respectively).

Table 4.8 - Interpolation results from Leicester UA to 30m square grids postings, aggregated at OA

Interpolation method	RMSE	CoV
Areal weighting using zone boundaries only	586.2	1.861
Pycnophylactic interpolation	516.8	1.641
Binary dasymetric using 3m resolution classified land cover	458.1	1.454
Binary dasymetric using 10m resolution classified land cover	447.4	1.420
Binary dasymetric using 30m resolution classified land cover	429.6	1.364

*Note:* Mean population of target units is 315.

Table 4.9 - Interpolation results from Leicester UA to 100m square grids postings, aggregated at OA

Interpolation method	RMSE	CoV
Areal weighting using zone boundaries only	761.9	2.419
Pycnophylactic interpolation	664.3	2.109
Binary dasymetric using 3m resolution classified land cover	630.1	2.000
Binary dasymetric using 10m resolution classified land cover	614.4	1.950
Binary dasymetric using 30m resolution classified land cover	503.5	1.598

*Note:* Mean population of target units is 315.

Tables 4.8 and 4.9 show the interpolation results from Leicester UA to 30m and 100m square grids postings respectively, aggregated at OAs in order of increasing accuracy. The results recorded have similar pattern to those recorded at LSOAs. The pycnophylactic method performed least well with RMSE of 516.8 and 664.3 (for 30m and 100m postings respectively) and CoV of 1.641 and 2.109 (for 30m and 100m postings respectively). The binary dasymetric model using land cover data derived from classified Landsat7 (ETM) 30m spatial resolution as the ancillary data input provided

the best target zone estimates at this resolution of interpolation with a recorder RMSE of 429.6 and 503.5 (for 30m and 100m postings respectively) and CoV of 1.364 and 1.598 (for 30m and 100m postings respectively).

It can be seen from the results presented in Tables 4.4 to 4.9 that:

1. The interpolation results aggregated at MSOAs have larger RMSE values than those aggregated at LSOAs, which are also larger than those aggregated at OAs. This is expected because the RMSE metric is affected by the absolute size of estimated values and the target size population for an OA is less than that of an LSOA, which is also less than that of an MSOA.
2. Interpolations using binary dasymetric method with land cover data derived from classified satellite image used as the ancillary data input returns significant improvement when compared with the interpolation results recorded from the pycnophylactic method.
3. The binary dasymetric model using land cover data derived from classified Landsat7 (ETM) 30m spatial resolution as the ancillary data input provided the lowest recorded RMSE score for all the models tested, for the three target zones.
4. The interpolation results from Leicester UA to 30m square grids postings provided lower RMSE score compared to 100m square grids postings, for all the models tested.
5. The CoV scores, which are appropriate for comparison across target zones, show interpolation results from Leicester UA to 30m square grids postings, aggregated at MSOAs provided the lowest CoV score among the solutions tested and for the three census areas used as the target zones.

#### **4.4.1 Residual Maps**

The residuals in all the census units tested were calculated and mapped. The residual maps are presented to provide a view of the absolute error present across the study area. Figures 4.12 to 4.14 show the target zones representative example residual maps for MSOAs, LSOAs and OAs respectively. The remaining residual maps are available in Appendix 7. The class intervals are shown by standard deviation from the mean error for each target zone. Standard deviations are the best way to symbolise normally

distributed quantitative data on maps making classes easy to interpret. From the residual maps presented in Figures 4.12 to 4.14, some pattern persists across scales. It becomes increasingly clear that a degree of spatial ‘smoothing’ is present in the estimates. That is, the very densely populated inner city OAs are underestimated, the less dense band running alongside the river north-south through Leicester is overestimated, and many large rural OAs are overestimated. Evidently, the residual maps show more census areas are subject to overestimation, as compared to underestimation, at greater than one standard deviation. The residual maps show that relatively large rural census units tend to be overestimated while relatively small urban census units tend to be underestimated. This is because they are designed to have a common target population count (Martin 1997). Similar patterns have been found by other researchers (e.g. Mennis and Hultgren 2006; Eicher and Brewer 2001), where relatively large rural blocks tend to be overestimated while relatively small rural blocks tend to be underestimated. In this study, the underestimated census units are mainly the smaller census units in the more densely populated areas such as the city centre while the overestimated census units are the larger spatial units in the less densely populated areas away from the city centre. A possible reason for this is that the satellite data being used as the ancillary data input is more likely to show houses and other built-up areas but not how many people live inside them. It is also likely in some areas there may be socioeconomic or cultural reasons why some houses have greater occupancy rates than the others.

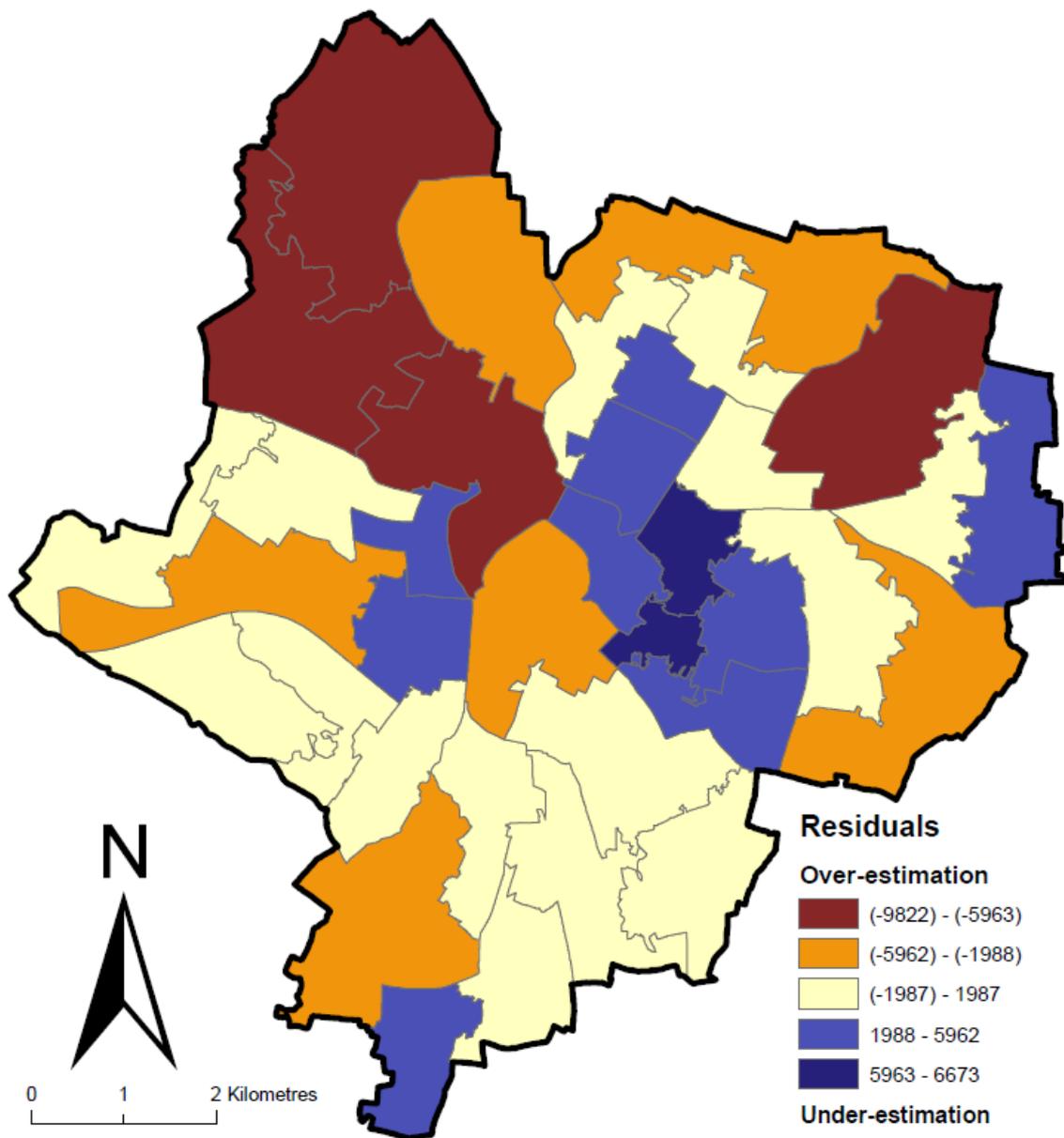


Figure 4.12 - The spatial distribution of residuals at MSOA from a 100m gridded pycnophylactic population surface. The mean count error is 0 and a standard deviation of 3975. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

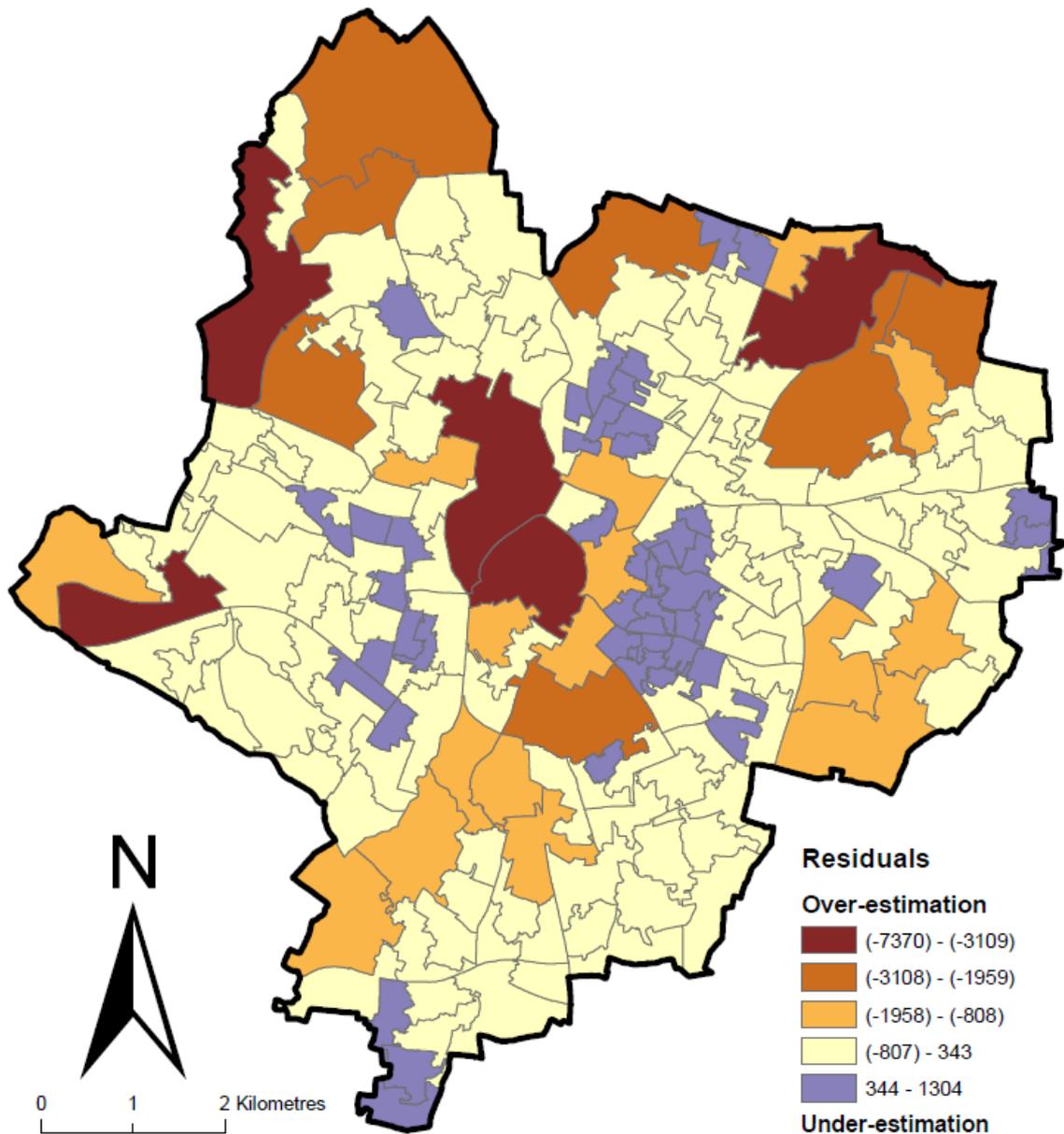


Figure 4.13 - The spatial distribution of residuals at LSOA from a 30m gridded dasymetric population surface using land cover data derived from classified resampled aerial photo data of 3m spatial resolutions as the ancillary data input. The mean count error is -233 and a standard deviation of 1151. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

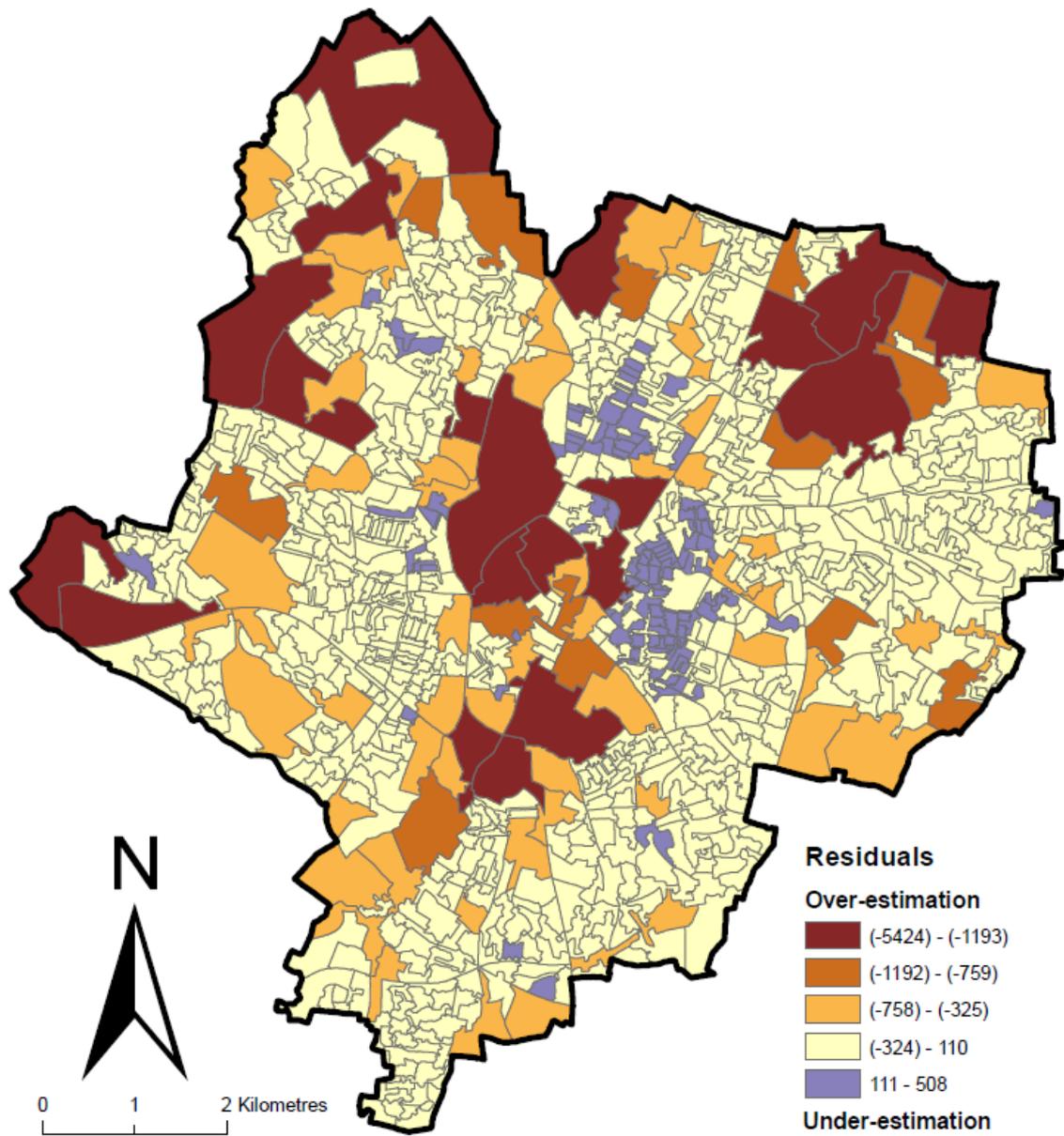


Figure 4.14 - The spatial distribution of residuals at OA from a 30m gridded dasymetric population surface using land cover data derived from classified resampled aerial photo data of 10m spatial resolutions as the ancillary data input. The mean count error is -107 and a standard deviation of 434. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

#### **4.5 Summary of results for Leicester**

This chapter compares the results of two areal interpolation methods used to estimate the population of Leicester UA at different spatial scales over small areas of unknown distributions. This was done in order to select the most appropriate interpolation method, support grid and ancillary data input to disaggregate 2006 population census totals for Port-Harcourt, Nigeria. The results of the error analyses show the dasymetric method provides the best fitting target zone estimates. Land cover information derived from classified Landsat7 (ETM) 30m spatial resolution when used as the ancillary data input for the binary dasymetric method was found to provide better estimates in the target zones compared to land cover data derived from classified resampled aerial photo data of 10m and 3m spatial resolutions used as the ancillary data input. The results also show that disaggregation over 30m square grids is more likely to provide better estimates in the target zones compared to when 100m square grids are used.

The binary dasymetric method using land cover data derived from classified Landsat7 (ETM) 30m spatial resolution when used as the ancillary data input and with a modelled population surface created using 30m support grids are seen to be the most appropriate parameters to apply for Port-Harcourt study.

## Chapter 5

### 5. RESULTS: PORT-HARCOURT

#### 5.1 Introduction

This chapter presents the results of areal interpolation for Port-Harcourt using the binary dasymetric method. It derives land cover information from classified Landsat7 (ETM+) 30m spatial resolution covering Port-Harcourt to identify built-up and non-built-up areas, and redistributes 2006 population census totals to create a modelled dasymetric population surface using a 30m square grids. The input parameters were determined in Chapter 4. The binary dasymetric method was also applied using land cover data derived from classified Spot5 (colour) 10m spatial resolution and from classified resampled Quickbird (60cm) image data of 3m spatial resolution, to create modelled dasymetric population surfaces over a 30m square grids. This was done to verify if the most appropriate parameters determined for Leicester study (presented in Chapter 4) would provide the best fitting target zone estimates for Port-Harcourt. In the absence of known small area census counts for Port-Harcourt to assess the performance of the interpolations, a random sample of 200 locations was taken for each surface to visually inspect the results in order to assess the accuracy of the population distribution.

The next section of the chapter presents the results of supervised classification. Section 5.3 presents the results of areal interpolations. Section 5.4 discusses the visual inspection of demand surfaces. Section 5.5 highlights what the errors from Leicester mean in Port-Harcourt. The last section (5.6) provides a summary of the results for Port-Harcourt.

#### 5.2 Supervised classification

The remotely sensed images for Port-Harcourt described in section 3.3.2 were processed in the same way as the data for Leicester described in section 3.4.1 to identify the extent of built-up areas that were used as the ancillary data input for the binary dasymetric method. Different from the Leicester study, four land cover types were identified for

each of Spot5 (colour) 10m spatial resolution and resampled Quickbird (60cm) image data of 3m spatial resolution. Category names (vegetation, thick vegetation, built-up and water) were assigned to these four classes and are described in Table 3.4. These were used in the supervised classification. The signature mean plot for all signatures, the mean plot and signature editor for the combined signatures for Landsat (ETM+) 30m spatial resolution image are shown in Figures 5.1, 5.2 and 5.3 respectively, and those for 10m and 3m images are shown in Appendix 8.

The next step applied the maximum likelihood classification algorithm as described in section 3.4.1 and a global classification accuracy of 81.64%, 79.30% and 78.13% was achieved using 30m, 10m and 3m spatial resolution satellite image respectively. The accuracy reports are available in Appendix 9. This section presents the classified images produced and the binary classified images created from the classified images, to show only built-up and non-built-up areas.

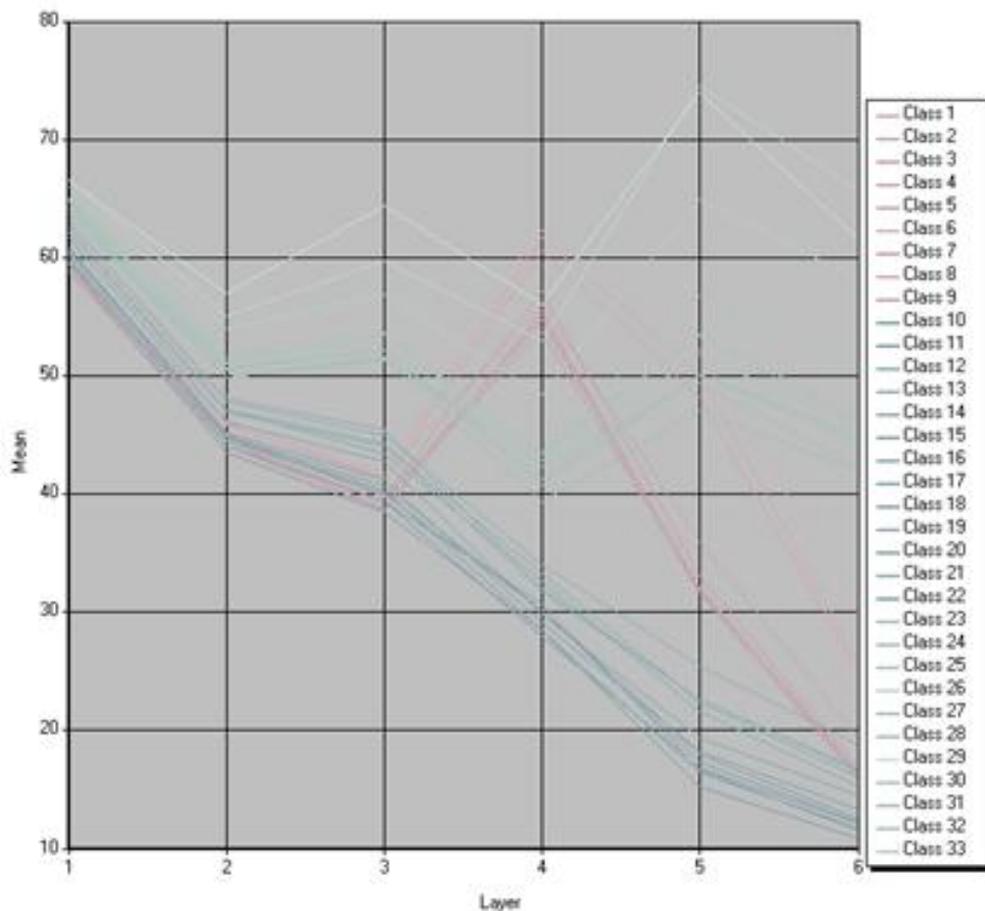


Figure 5.1 - Signature mean plot evaluating signatures for vegetation, built-up and water from Landsat7 (ETM+) 30m spatial resolution image.

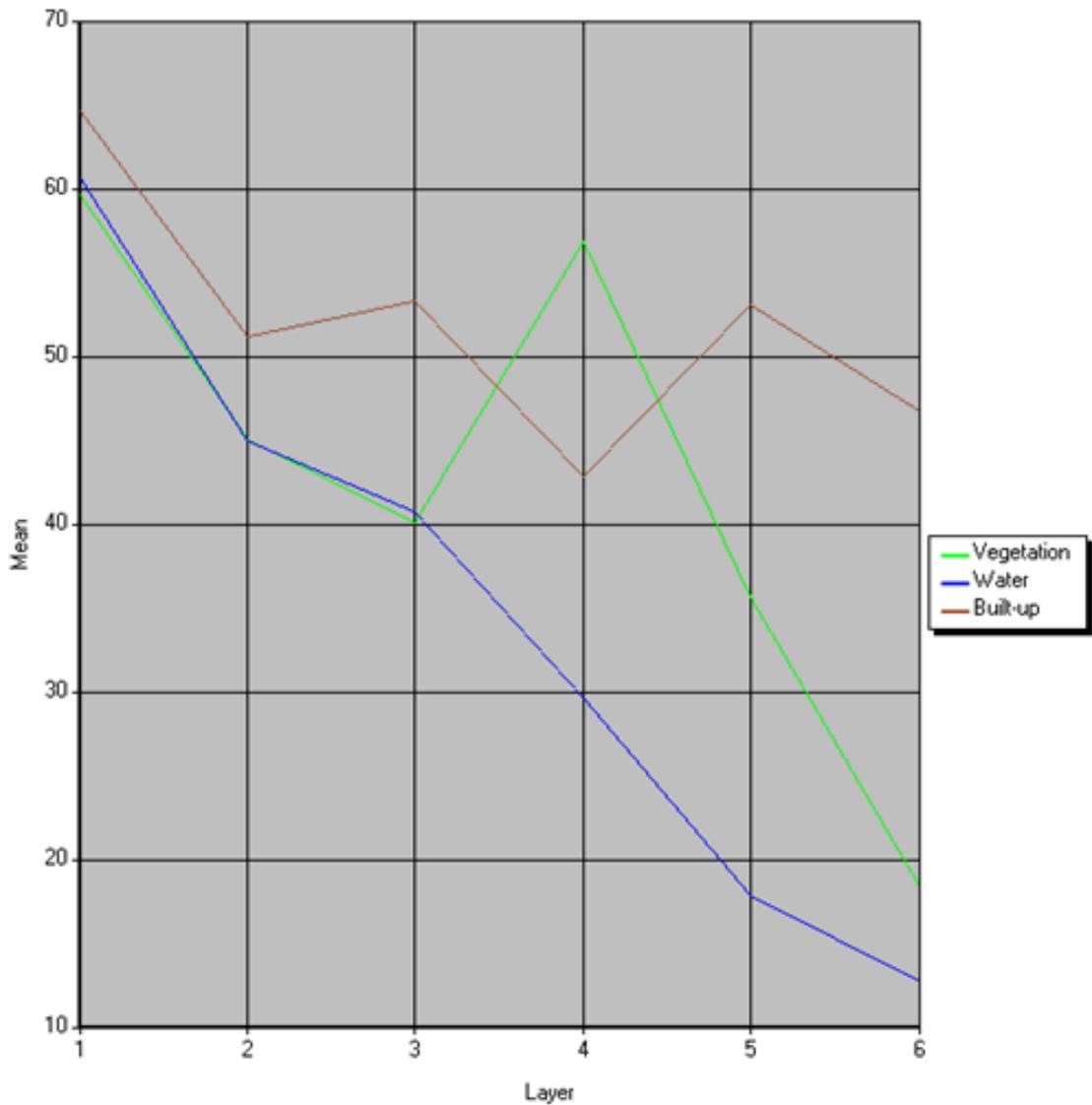


Figure 5.2 - Signature mean plot for the combined signatures from Landsat7 (ETM) 30m resolution image.

Class #	>	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1		Vegetation	<span style="color: green;">■</span>	0.000	1.000	0.000	34	34	780	1.000	✓	✓	✓	✓	
2		Water	<span style="color: blue;">■</span>	0.000	0.000	1.000	1	35	2107	1.000	✓	✓	✓	✓	
3	▶	Built-up	<span style="color: brown;">■</span>	0.627	0.322	0.176	2	36	791	1.000	✓	✓	✓	✓	

Figure 5.3 - Signature editor for the combined signatures from Landsat7 (ETM) 30m resolution image.

### 5.2.1 Classified images

The classified images created from supervised classification described in section 3.4.1 are shown in Figures 5.4 to 5.6. Figure 5.4 shows a classified image using Landsat7 (ETM+) 30m spatial resolution image data. Figure 5.5 shows the classified image using Spot5 (colour) 10m spatial resolution image data. Figure 5.6 shows classified image using resampled Quickbird (60cm) image data of 3m spatial resolution. The accuracy of classification was assessed in the same way as the data for Leicester described in section 4.2.1.

Table 5.1 presents the comparison of the overall accuracy and Kappa statistic between the different land cover data of different sources and resolutions. The accuracy reports indicate a good agreement between thematic maps generated from image and the reference data although the three satellite images used recorded a little below the minimum standard of digital image classification for optical remote sensing data (85%) recommended by Paul (1991) and Jansen et al. (2008). However, the classification accuracy recorded for Landsat7 (ETM+) 30m spatial resolution image data was better than those recorded for Spot5 (colour) 10m spatial resolution image data and resampled Quickbird (60cm) image data of 3m spatial resolution (see Table 5.1).

Table 5.1- Comparison of the overall accuracy and Kappa statistic between the different land cover data

Land cover data	Overall accuracy (%)	Kappa accuracy (%)	Kappa coefficient
30m spatial resolution	81.64	73.72	0.74
10m spatial resolution	79.30	71.58	0.72
3m spatial resolution	78.13	69.55	0.70

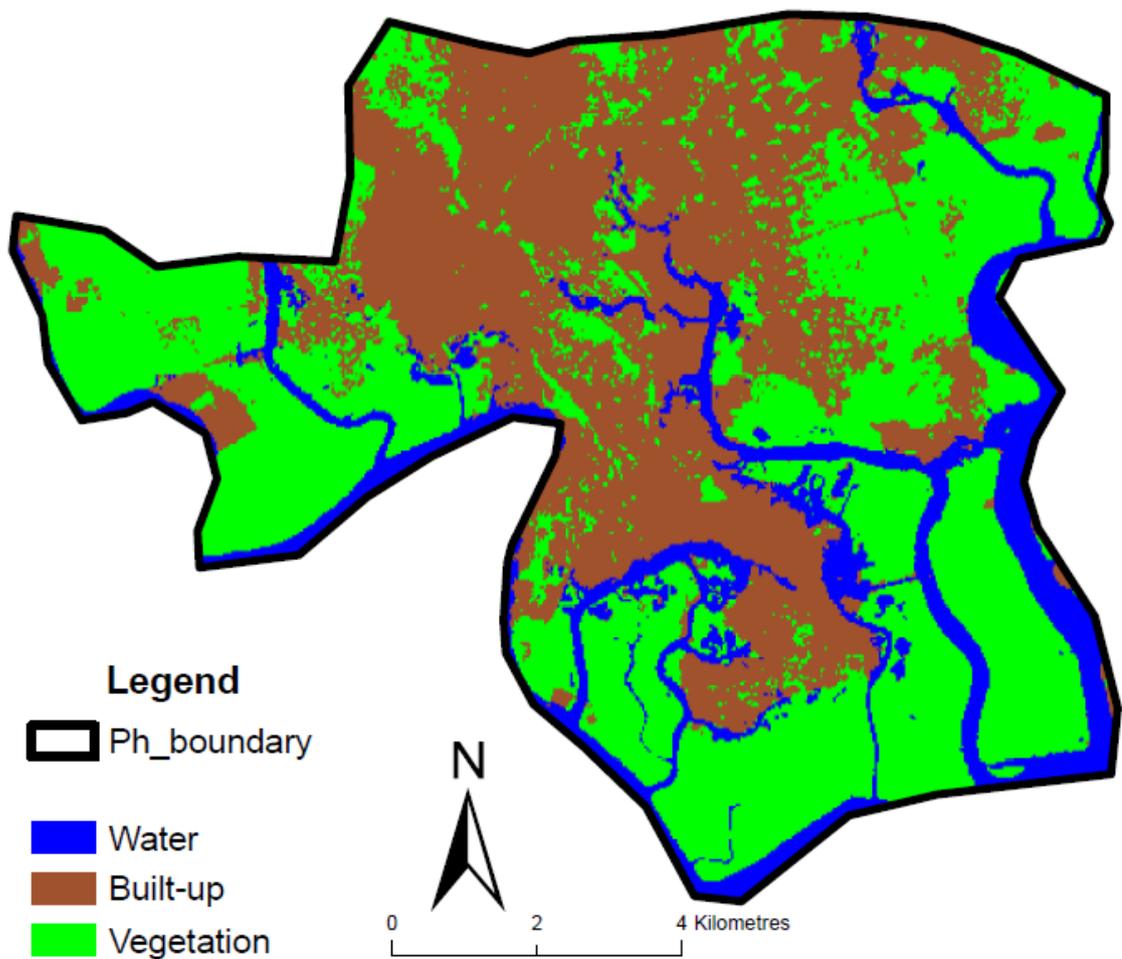


Figure 5.4 - The classified Port-Harcourt image derived from Landsat7 (ETM+) 30m spatial resolution. The digital boundary is Copyright for Geotechnics Services 2011.

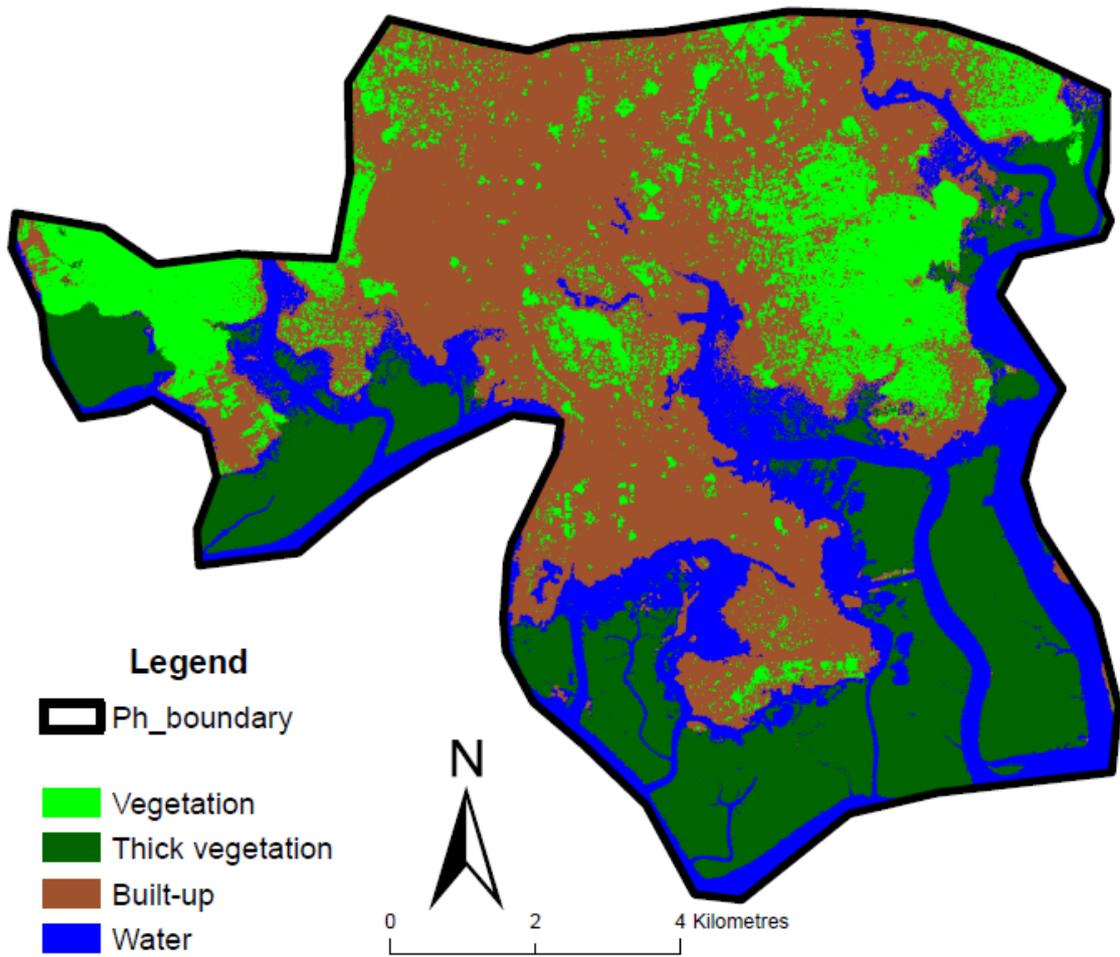


Figure 5.5 - The classified Port-Harcourt image derived from Spot5 (colour) 10m spatial resolution. The digital boundary is Copyright for Geotechnics Services 2011.

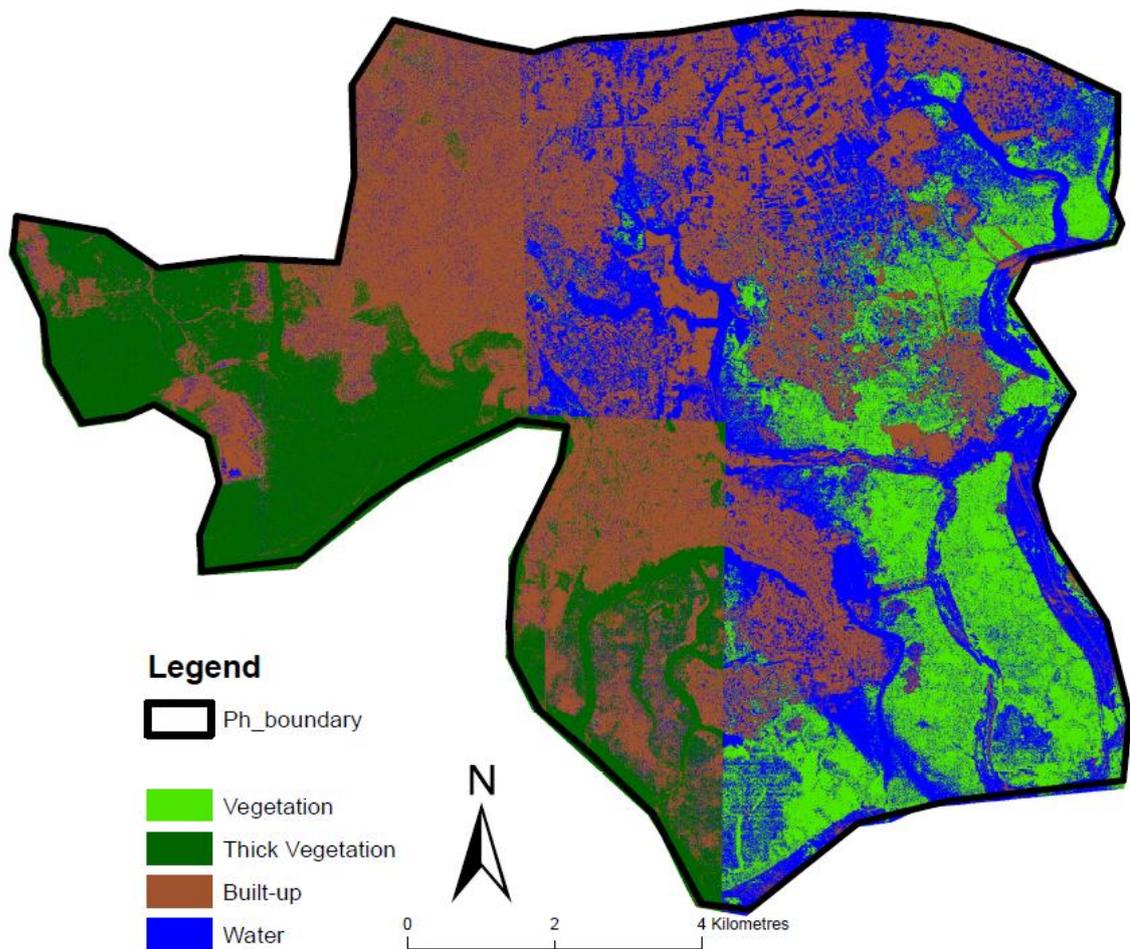


Figure 5.6 - The classified Port-Harcourt image derived from resampled Quickbird (60cm) image data of 3m spatial resolution. The digital boundary is Copyright for Geotechnics Services 2011.

## 5.2.2 Binary classified images

The binary dasymetric method is based on additional geographic information (e.g. land use) that provides a binary divide between built-up and non-built-up areas. The classified images presented in section 5.2.1 were reclassified into a simple binary division by assigning a weighting factor of 1 to all areas classified as built-up areas and 0 to all other areas (as described in section 3.4.1). The binary masks derived from different land cover data derived from classified satellite images of different resolutions and sources which represent the underlying spatial distribution of population within the source zone, are presented in Figures 5.7 to 5.9.

Figure 5.7 shows a binary mask derived from land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution image data. Figure 5.8 shows a binary mask derived from land cover data derived from classified Spot5 (colour) 10m spatial resolution image data. Figure 5.9 shows a binary mask derived from land cover data derived from classified resampled Quickbird (60cm) image data of 3m spatial resolution.

Table 5.2 compares the sizes of the total built-up areas in the source zone derived from different land cover data. The binary classified images showed that built-up areas derived from land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution and that derived from classified resampled Quickbird (60cm) image data of 3m spatial resolution have the same area size. The area size slightly reduced for the binary classified image derived from land cover data derived from classified Spot5 (colour) 10m spatial resolution.

Table 5.2 - Comparison of the total built-up area in the source zone between the different land cover data

Built-up areas	Total source zone area (km <sup>2</sup> )
Landsat7 (ETM+) 30m spatial resolution	43.69
Spot5 (colour) 10m spatial resolution	43.64
Resampled Quickbird image data 3m spatial resolution	43.69

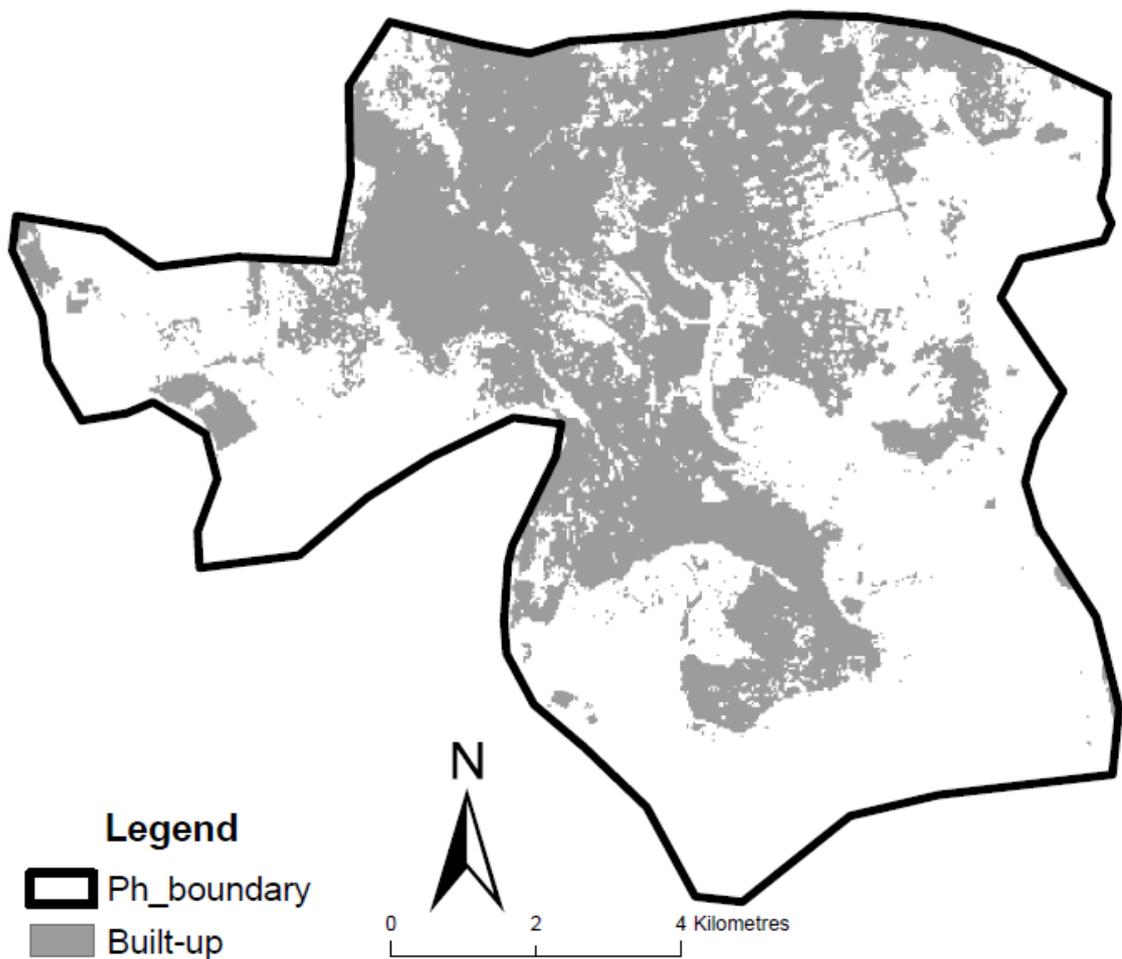


Figure 5.7 - A binary mask derived from land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution image data. The digital boundary is Copyright for Geotechnics Services 2011.

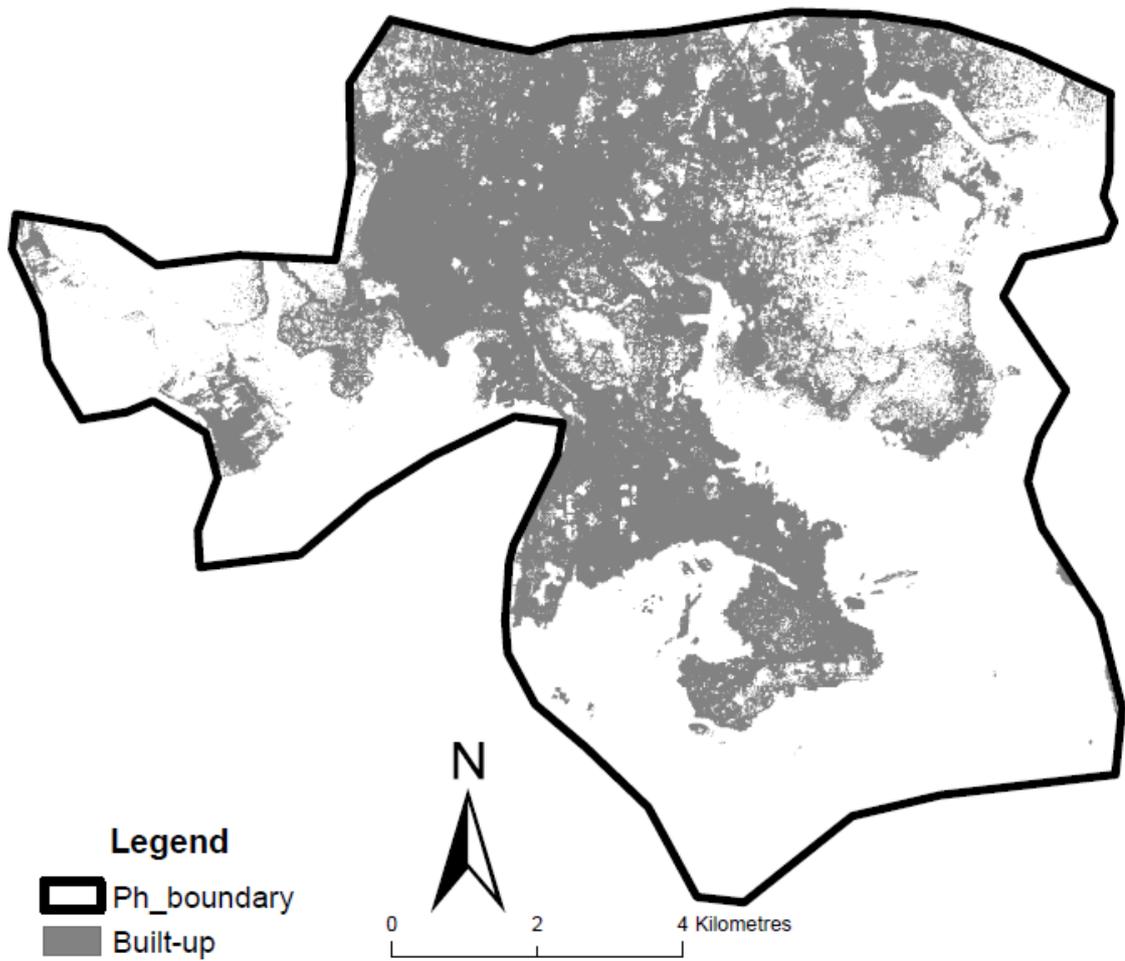


Figure 5.8 - A binary mask derived from land cover data of 10m spatial resolution derived from classified Spot5 (colour) image data. The digital boundary is Copyright for Geotechnics Services 2011.

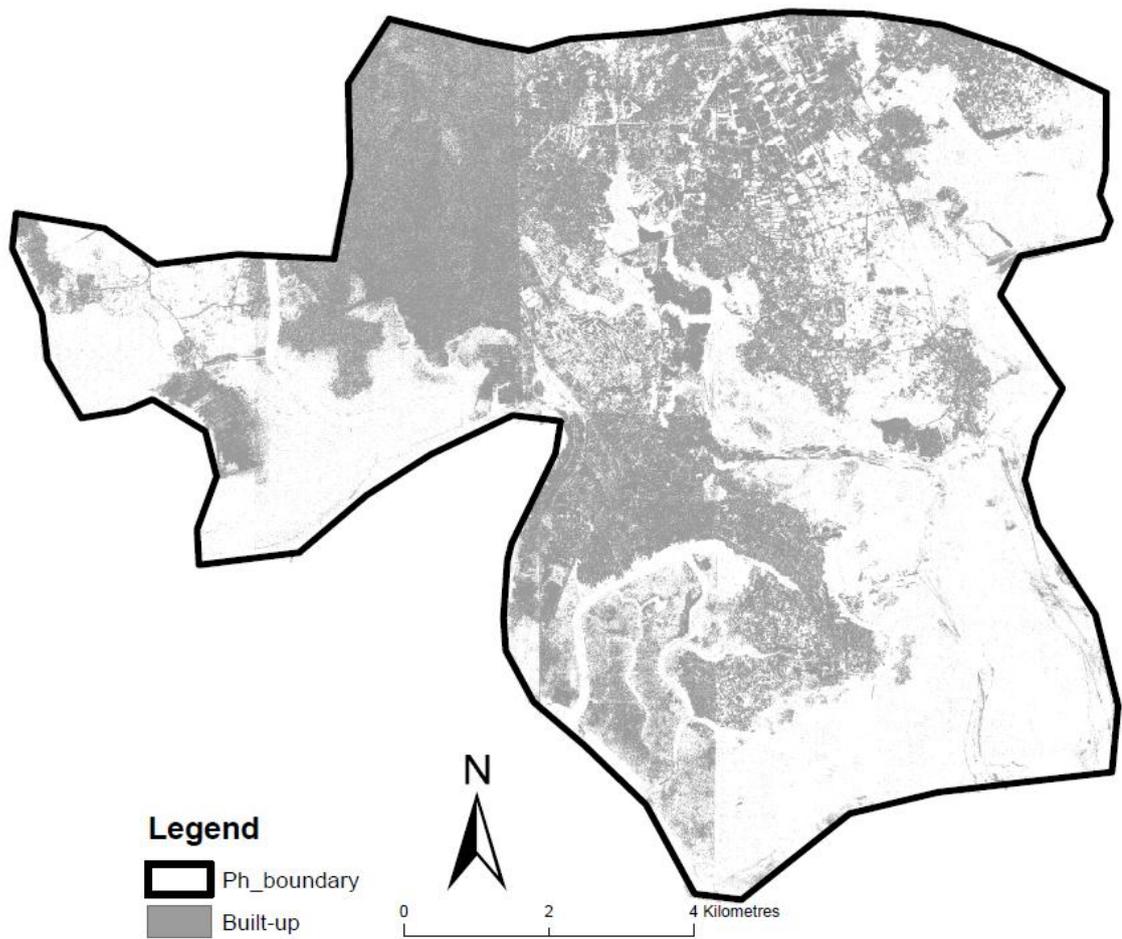


Figure 5.9 - A binary mask derived from land cover data of 3m spatial resolution derived from classified resampled Quickbird (60cm) image data. The digital boundary is Copyright for Geotechnics Services 2011.

### **5.3 Areal interpolation methods**

The binary dasymetric method was implemented using land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution for Port-Harcourt as the ancillary data input to create a modelled dasymetric population surface using a 30m square grid. The input parameters were determined from the results of Leicester study, in Chapter 4. The binary dasymetric method was also applied using land cover data derived from classified Spot5 (colour) 10m spatial resolution and from classified resampled Quickbird (60cm) image data of 3m spatial resolution, to create modelled dasymetric population surfaces over a 30m square grids to verify if the most appropriate parameters determined for Leicester study (presented in Chapter 4) would provide the best fitting target zone estimates for Port-Harcourt. The scales of disaggregation used in the implementation generated different target zone estimates of the population.

#### **5.3.1 The binary dasymetric method**

The binary masks derived from land cover data of different sources and resolutions (presented in section 5.2.2) were converted to vector and used together with 30m square grids to create dasymetric population surfaces for Port-Harcourt. The population density for each of 30m, 10m and 3m spatial resolution land cover data used as the ancillary data input was calculated and mapped as a single uniform density estimate across Port-Harcourt as shown in Figures 5.10, 5.11 and 5.12 respectively. Figure 5.10 shows the dasymetric map of population surface at 30m posting created using land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution ancillary data input. Figure 5.11 shows the dasymetric map of population surface at 30m posting created using land cover data derived from classified Spot5 (colour) 10m spatial resolution ancillary data input and Figure 5.12 shows the dasymetric map of population surface at 30m posting created using land cover data derived from classified resampled quickbird (60cm) image of 3m spatial resolution ancillary data input.

Table 5.3 shows the population densities (persons/10,000 m<sup>2</sup>) for the different ancillary data input used for the binary dasymetric method. The results show that land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution, land cover data derived from classified Spot5 (colour) 10m spatial resolution and land cover data

derived from classified resampled Quickbird (60cm) image data of 3m spatial resolution all provided the same population density across Port-Harcourt.

Table 5.3 - Population density per 10,000 m<sup>2</sup> for binary dasymetric maps of population

Ancillary data input	Population density (persons/10,000 m <sup>2</sup> )
Landsat7 (ETM+) 30m spatial resolution	12.385
Spot5 (colour) 10m spatial resolution	12.398
Resampled Quickbird (60cm) data of 3m spatial resolution	12.386

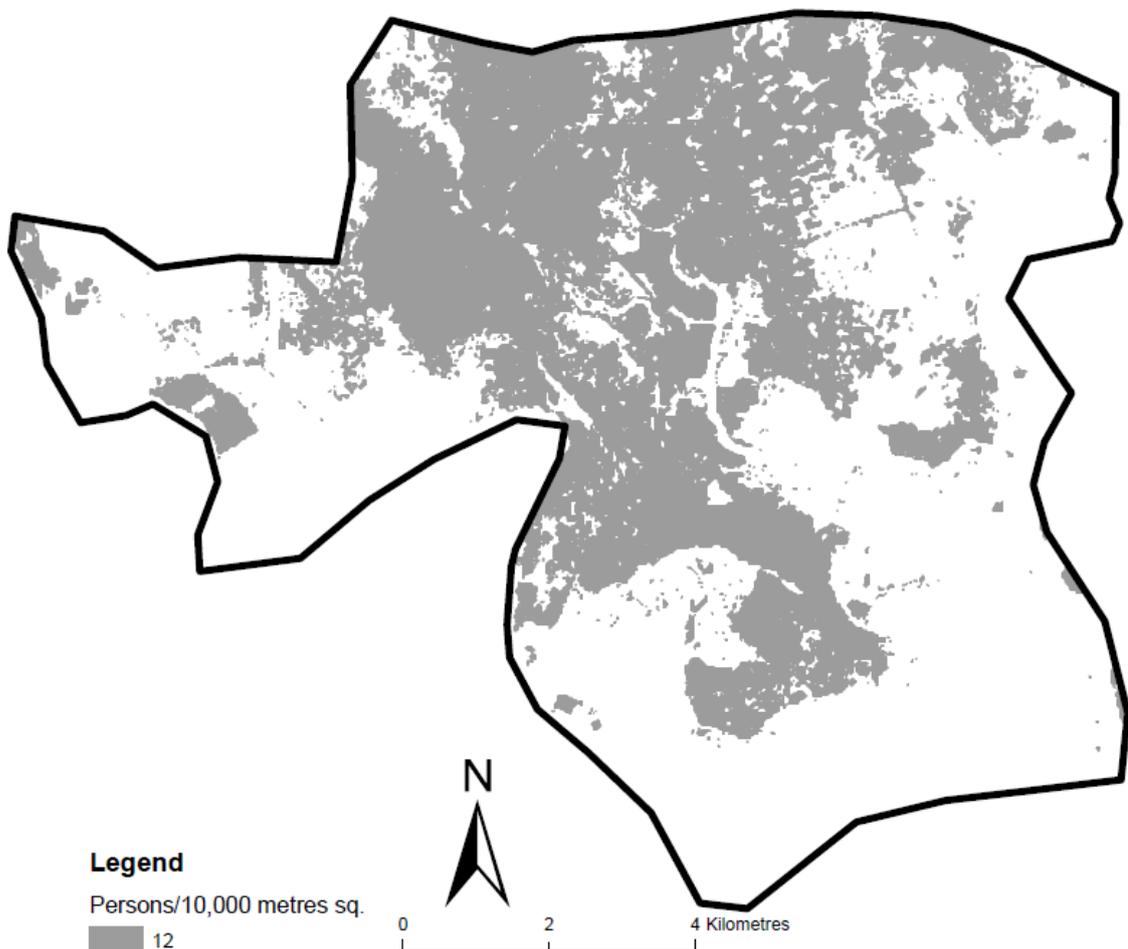


Figure 5.10 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution ancillary data input. The digital boundary is Copyright for Geotechnics Services 2011.

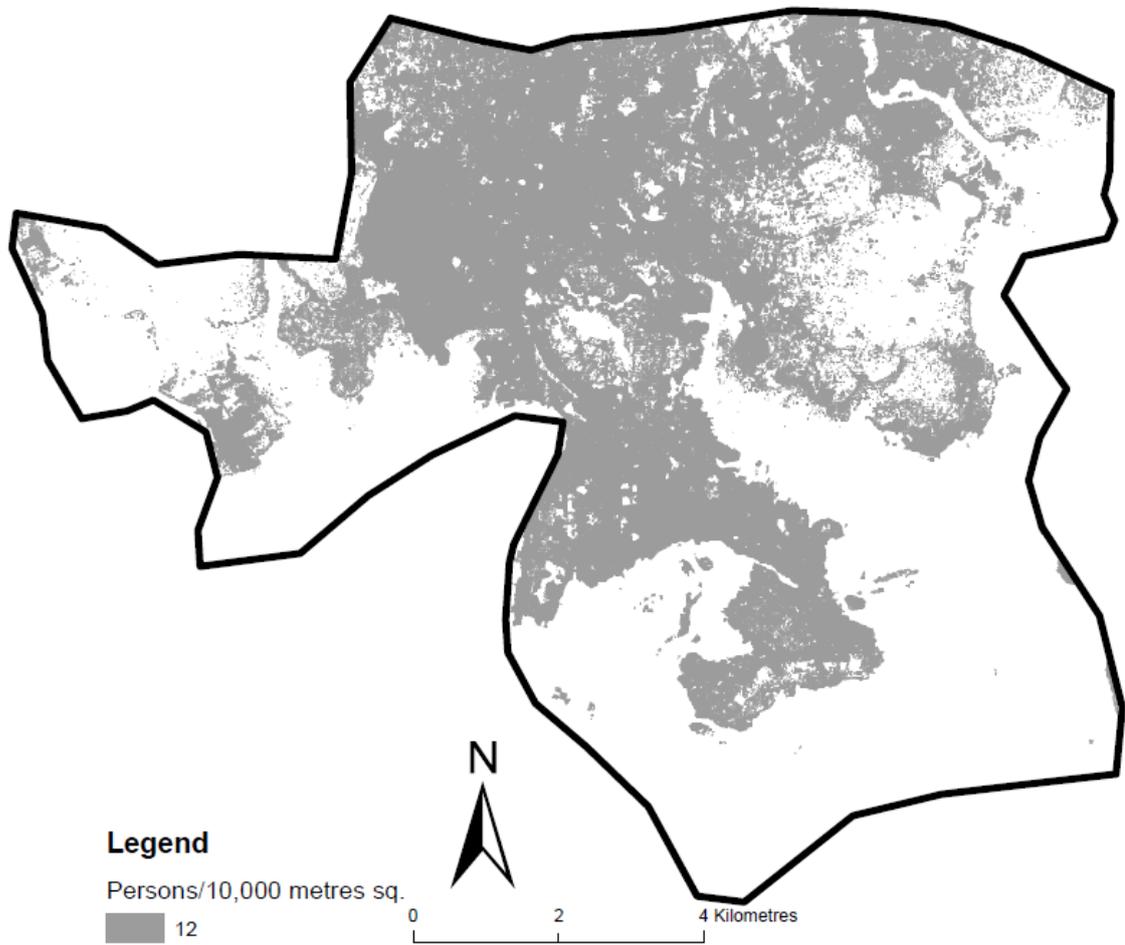


Figure 5.11 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified Spot5 (colour) 10m spatial resolution ancillary data input. The digital boundary is Copyright for Geotechnics Services 2011.

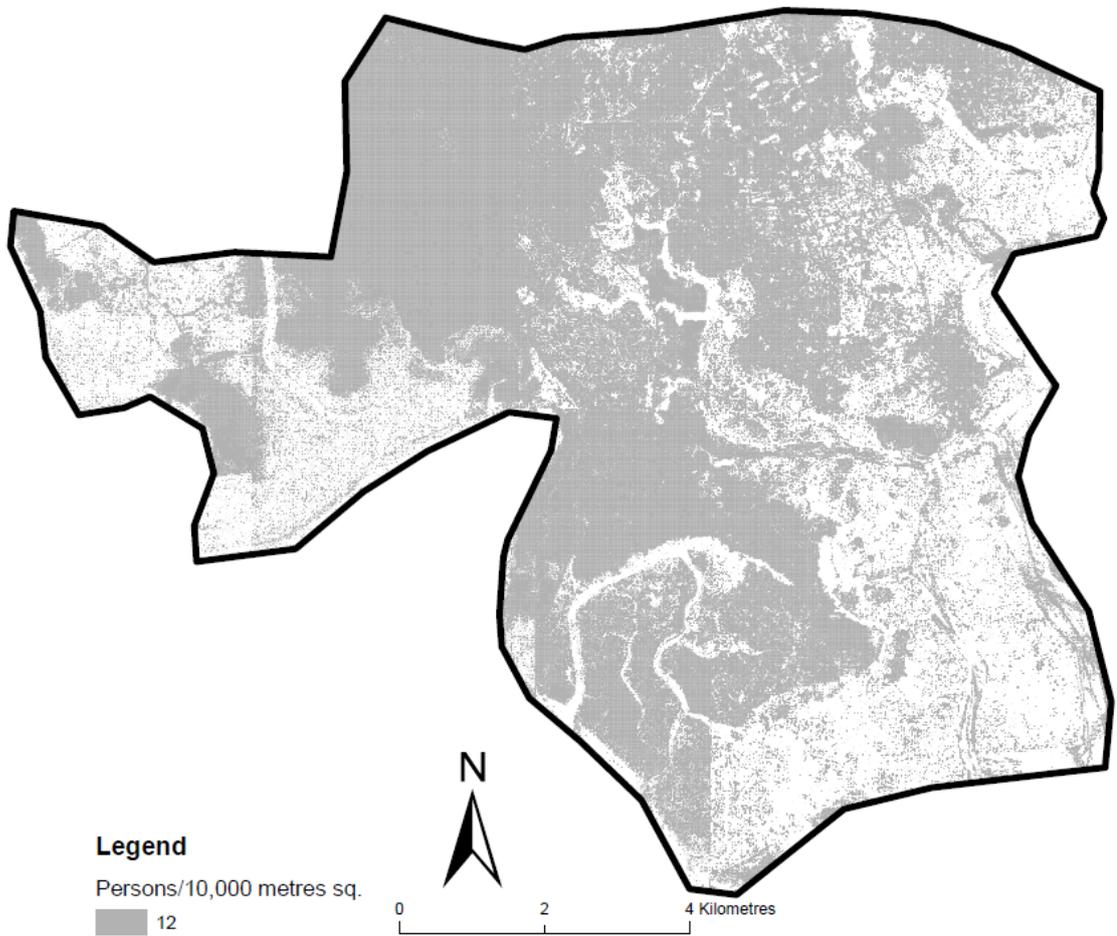


Figure 5.12 - The dasymetric map of population surface at 30m posting created using land cover data derived from classified resampled quickbird (60cm) image of 3m spatial resolution ancillary data input. The digital boundary is Copyright for Geotechnics Services 2011.

#### 5.4 Visual inspection of demand surfaces

In the absence of field information (survey data) or another data source to validate the modelled population surfaces, a random sample of 200 locations was taken to visually inspect the results in order to assess the accuracy of the population distribution. The randomly placed points were generated to have a shortest distance of 200m between any two random points (see Figure 5.13). Google Earth 7.1 was used as a reference to inspect the surfaces as shown in Figure 5.14. The results of visual inspection of surfaces are available in Appendix 10.

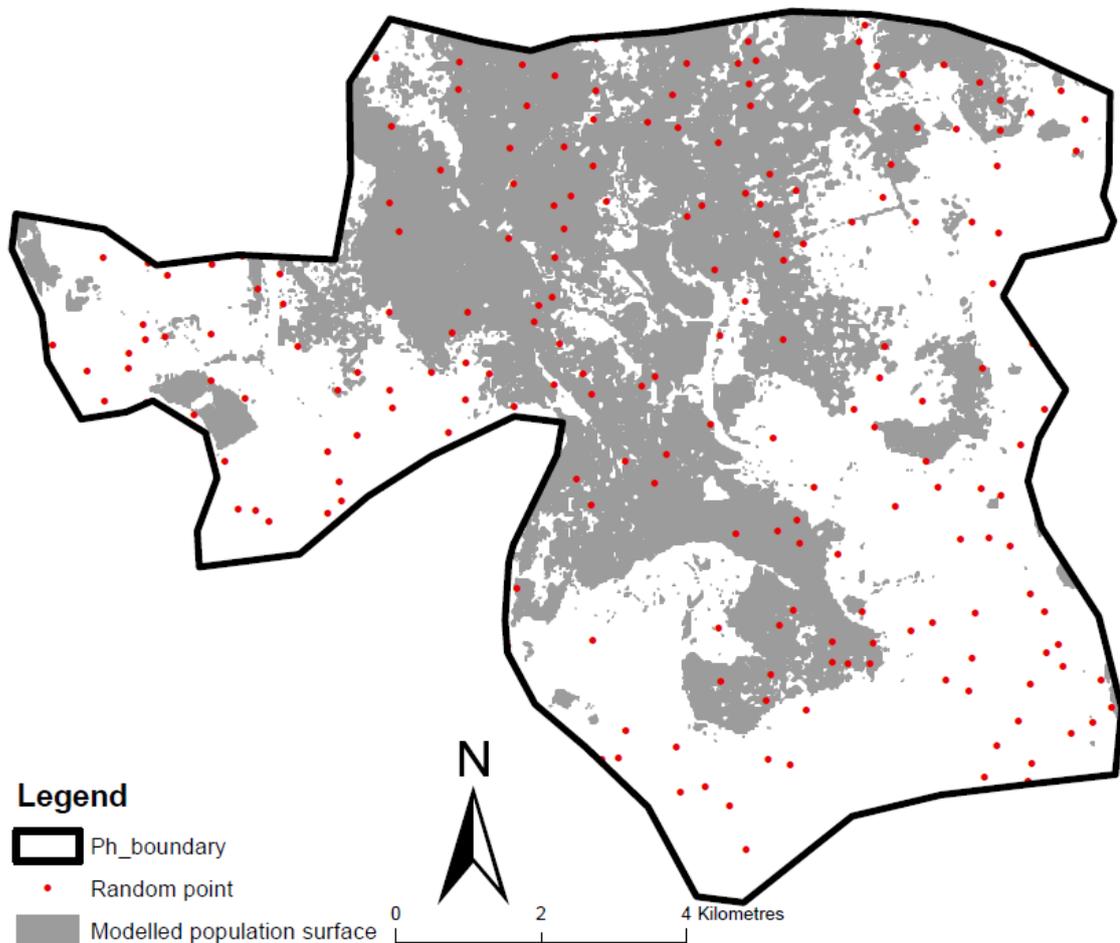


Figure 5.13 - The demand surfaces for Port-Harcourt with 200 random points generated within the boundary of Port-Harcourt. The digital boundary is Copyright for Geotechnics Services 2011.

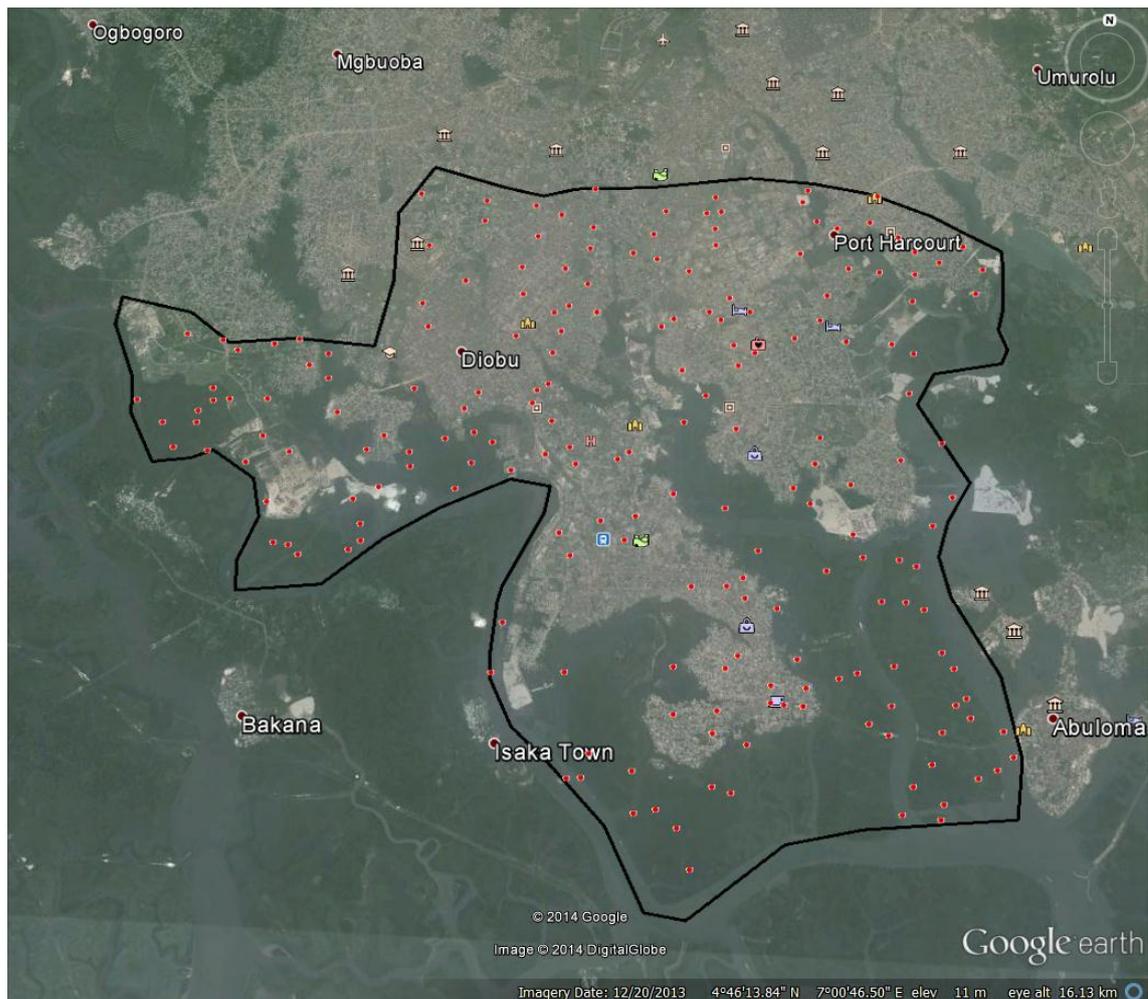


Figure 5.14 - An overlay of study area boundary with random points generated within the boundary on Google Earth 7.1.

Table 5.4 summarises the results of the visual inspection for the dasymetric population surfaces created for Port-Harcourt using land cover data derived from classified; Landsat7 (ETM+) 30m spatial resolution, Spot5 (colour) 10m spatial resolution, and resampled quickbird (60cm) image 3m spatial resolution as the ancillary data input. The results reveal some of the random points on the populated surfaces do not correspond to a populated area when viewed using Google Earth. This is likely due to land cover changes between the date of the Google Earth reference (20<sup>th</sup> December 2013) and the dates the imagery were acquired (see Table 3.3). Figure 5.15 shows an example of a random point selected on a populated surface located in an unpopulated area. Some other random points on the unpopulated surfaces were identified on Google Earth in populated areas (see Figure 5.16). The effect of using built-up areas to represent populated areas is seen where some of the random points on the populated surface were

identified in an industrial area as in Figure 5.17. The results of the inspection show sixteen out of two hundred random points selected (eight per cent of the random points selected) identified as populated on the surface were actually in unpopulated areas. Five per cent of the random points not on the populated surface correspond to populated areas on Google Earth reference. A total of twenty-six out of two hundred random points selected (thirteen per cent) on the interpolated surface do not correspond with Google Earth reference. When 10m and 3m spatial resolution land cover data were used, 14.5 and 22 per cent respectively of the random points selected were not correctly redistributed.

Table 5.4 – Summary of visual inspection of surfaces

		Dasymetric Population Surfaces					
		Populated			Unpopulated		
Image resolution		30m	10m	3m	30m	10m	3m
Google Reference	Populated	75	55	45	10	10	15
	Unpopulated	16	19	29	99	116	111

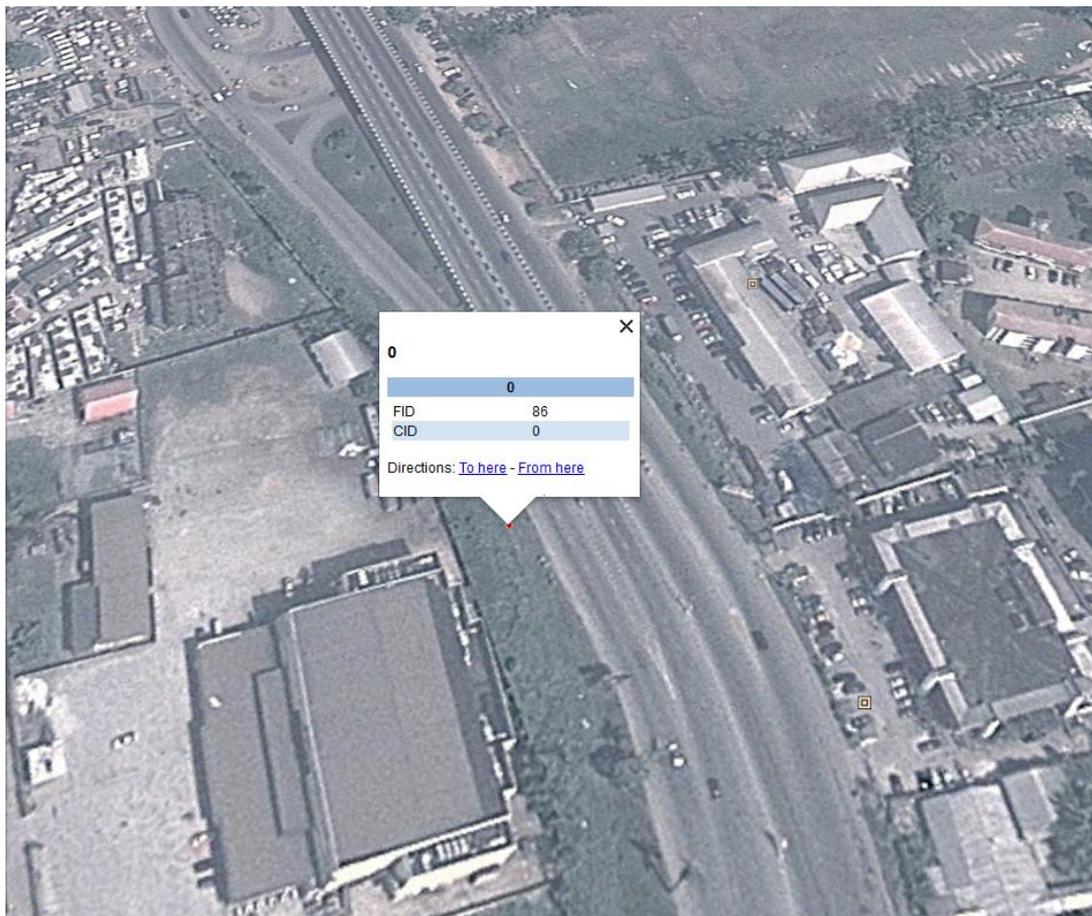


Figure 5.15 - Random point on the populated surface corresponds to an unpopulated area in Google Earth reference.

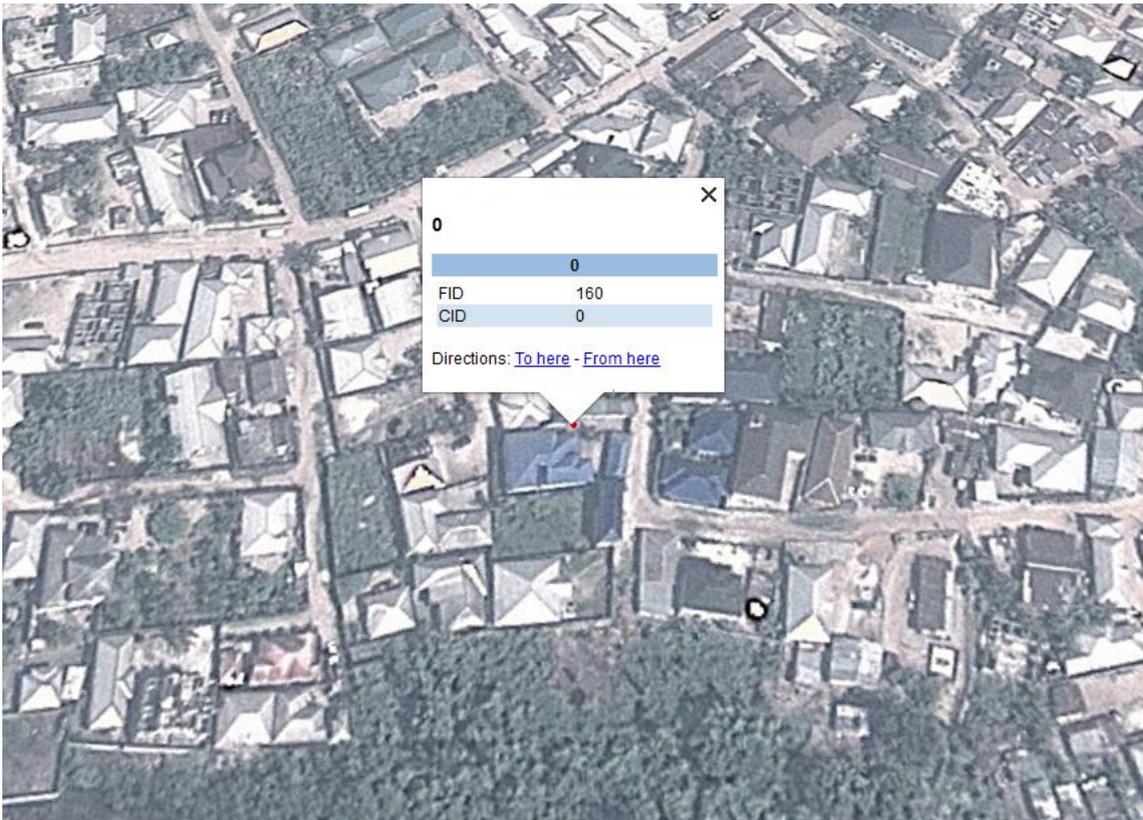


Figure 5.16 - Random point on the unpopulated surface is identified in a populated area.

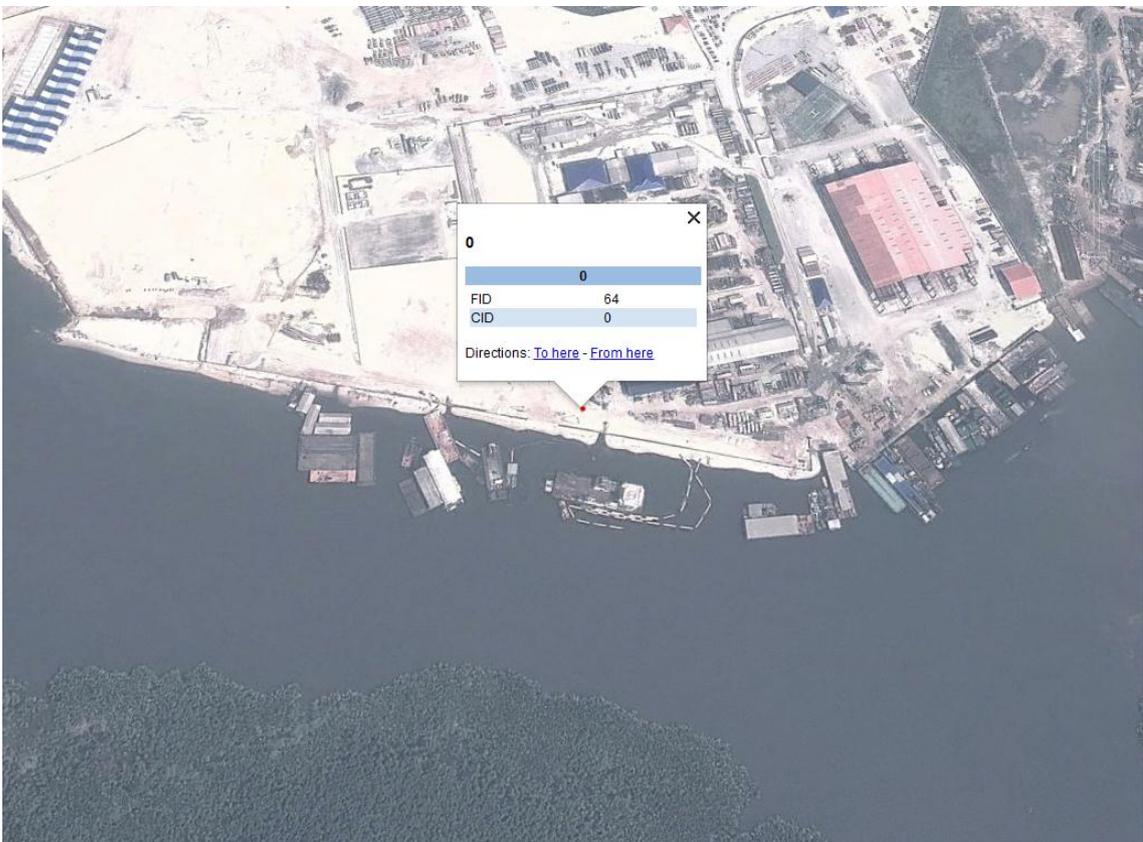


Figure 5.17 - Random point on the populated surface is identified in an industrial area.

## **5.5 What the errors from Leicester mean in Port-Harcourt**

Population census totals for Leicester UA were redistributed to small areas of unknown distributions at different spatial scales. This was done to develop models using land cover and census data for Leicester where the population distribution is known and validation is possible in order to select the best performing model and parameters to apply to Port-Harcourt. The results show differences in demand surfaces generated by each method and for each support grid. These differences are largely due to different assumptions made by the two methods, the size of census areas used for testing the performance of the interpolations and the support grid sizes. The patterns of errors from Leicester study were carefully considered before selecting the input parameters for Port-Harcourt.

There are similarities between the population density, structure and demography for Leicester and that of Port-Harcourt. The spatial distribution of housing units in Leicester is assumed to be similar to that of Port-Harcourt (e.g. similar levels of overcrowding, lone occupancy etc.) with dense housing units in and around the city centre. This cultural assumption was made due to limited spatial data infrastructure in Port-Harcourt. There are also large tracks of unpopulated land within the urban fabric in both cities which makes it easier to depict the underlying distribution of the population with the use of ancillary information derived from classified satellite imagery. The errors from the Leicester study suggest the population estimates for Port-Harcourt are likely to be underestimated. This is because the error maps for Leicester suggest more urban census units with large populations were underestimated as compared to overestimated. For the pycnophylactic method, the assumption of continuous population surface having a non-zero population density value at every location does not reflect the underlying population distribution in Port-Harcourt. This is because the information obtained from classified imagery shows about one third of Port-Harcourt is covered by water and other uninhabitable areas. This means if the pycnophylactic model were to be used in Port-Harcourt, a large number of population estimates would be redistributed in uninhabitable areas thereby underestimating the actual population in Port-Harcourt.

## **5.6 Summary**

Population census totals for Port-Harcourt aggregated for reasons of confidentiality and to reduce data volume has been successfully redistributed over small areas of unknown distributions. The spatial distribution of population estimates reflects the underlying distribution of population in Port-Harcourt because ancillary data input was used to constrain the re-allocation of population counts to only areas identified as populated. The visual inspections of surfaces revealed about eighty per cent of the demand population values created correspond to areas identified as built-up in the Google Earth reference. The spatially distributed demand population values created for Port-Harcourt can be used as input to the location analyses described in the next chapter.

## Chapter 6

### 6. APPLICATION: Location-allocation in Port-Harcourt

#### 6.1 Case study: Primary Health Care Centres (PHCCs) in Port-Harcourt

The case study is to evaluate the location of the 17 current primary health care centres (PHCCs) in Port-Harcourt, Rivers State, Nigeria. The PHCCs are the first level of health care service provision and are critical to the survival of children below the age of five. These services include; immunisation against infectious diseases, control of preventable diseases (Such as Malaria, Diarrhoea etc.), child survival and nutrition, maternal and new born care, health education and community mobilisation, antenatal care, family planning services, basic surgical services and general well-being of the people (Policy Project 2002). Rivers State government made known its commitment on the 7<sup>th</sup> of April 2009 to provide quality healthcare that is affordable and accessible to the people in all areas of the state. The government is constructing 160 PHCCs across the 23 local government areas (LGAs) in the state. As at May 2014, a total of 130 PHCCs have been completed and are in use, with the remaining 30 PHCCs at various stages of completion (Parker 2014).

The National Health Policy (NHP) recommends at least one PHCC in each electoral ward. The electoral ward is the smallest constituency represented by a councillor with an estimated population of less than thirty thousand people (NPHCDA 2007). There are twenty wards in Port-Harcourt (Independent National Electoral Commission 2011). On the locations of health facilities, the Revised National Health Policy states that “*Ministries of health shall review the distribution and establish standards to regulate the locations of health care facilities*” (RNHP 2004, P.23). This gives government officials (in ministries of health) the mandate to decide on suitable location of health facilities. The final decision on the locations and number of PHCCs to construct in each of the 23 LGAs in the state was believed to have been made by government officials, politicians and local planners in the state. A total of 17 out of 160 PHCCs were sited within Port-Harcourt. The residents of Port-Harcourt have expressed concern over the process by which decision on the number of PHCCs to construct and their chosen

locations were finalised. This is because the residents alleged the decisions were made without any formal analysis and generation of alternatives as the government did not provide any evidence to support its decision.

Geographical Information Systems (GIS) is an effective planning tool that is increasingly being used to obtain reliable health information that improves unbiased decision making in public health policies and health system planning (Higgs 2005). The case study is a p-median problem because the government had already sited 17 PHCCs in Port-Harcourt. The aim was to evaluate current health facility locations and suggest alternative spatial arrangements of facilities to improve spatial accessibility.

The source zone, the boundary of Port-Harcourt local government area with the distribution of road network within Port-Harcourt and the locations of current PHCCs in Port-Harcourt are shown in Figure 6.1. The locations of 17 current PHCCs were directly observed during field trip.

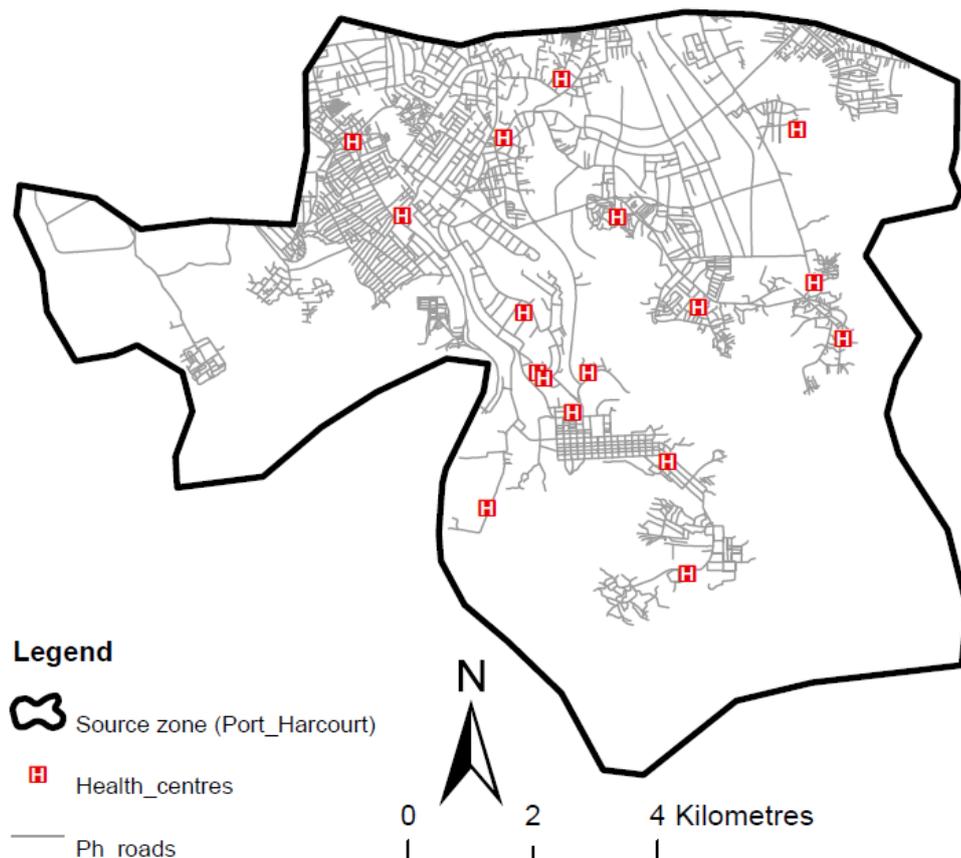


Figure 6.1 - Source zone with roads (in grey) and the locations of current PHCCs in Port-Harcourt. The digital boundary is Copyright for Geotechnics Services 2011.

This analysis evaluates the distribution of 17 current PHCCs in Port-Harcourt against the demand population values described in Chapter 5. The spatial distribution for the demand needs is an important factor in the location-allocation problems and the p-median method was developed to support facility location planning using point representation for the demand surfaces (Cromley et al. 2012). Teitz and Bart's (1968) algorithm was used to solve the p-median problem. This is because the aim of the case study analysis was to optimise the overall shortest distance travelled from each demand point to each health centre. The critical inputs into Teitz and Bart's algorithm are the number of facilities to locate, the demand population values created from areal interpolation and the distance matrix. The demand population values were interpolated down to a very fine spatial detail (30m grid postings). It is therefore expected that the error rate (RMSE/CoV) associated with these values would be significantly higher than those reported earlier for OAs in Leicester, although the actual error is unknown. This is because of the size of the target units (30m grid postings) used for the interpolation. Findings from review of spatial access (section 2.4.1) show GIS network analysis method to be an ideal technique for evaluating access distance to facilities. A number of different approaches could be used to obtain access distance depending on the software used and the analysis to be performed. ArcGIS Desktop 10.2.2 was used to calculate the shortest travelled distance between each demand point and each health centre.

The next section of the chapter describes the data processing. Section 6.3 describes the implementation of Teitz and Bart's heuristics to solve the p-median problem. Section 6.4 presents the results. The last section presents a summary of the chapter.

## **6.2 Data and preparation**

The dasymmetric map of population surface at 30m postings created using land cover data derived from classified Landsat7 (ETM+) 30m spatial resolution covering Port-Harcourt area as the ancillary data input was converted to points using *feature-to-point function* in ArcGIS 10.2.2. There are 56457 grid points spaced at 30m apart with each point representing estimates of the population redistributed from 2006 census totals for Port-Harcourt. The spatially distributed surfaces representing demand in Port-Harcourt was used as input to the location analyses with roads data and point locations of current PHCCs (shown in Figure 6.1) to obtain the shortest travelled distance from each

demand point to each PHCC. The Network Analyst extension in ArcGIS was used. This extension requires a specific type of road network to be able to generate the shortest travelled distance on the road network from each demand point to each PHCC. These specifications include roads that: keep turns in, are connected at nodes, have known attributes (e.g. using metres to calculate distance) etc. The roads network dataset was built in ArcCatalog 10.2.2 using the roads layer (see Figure 6.2).

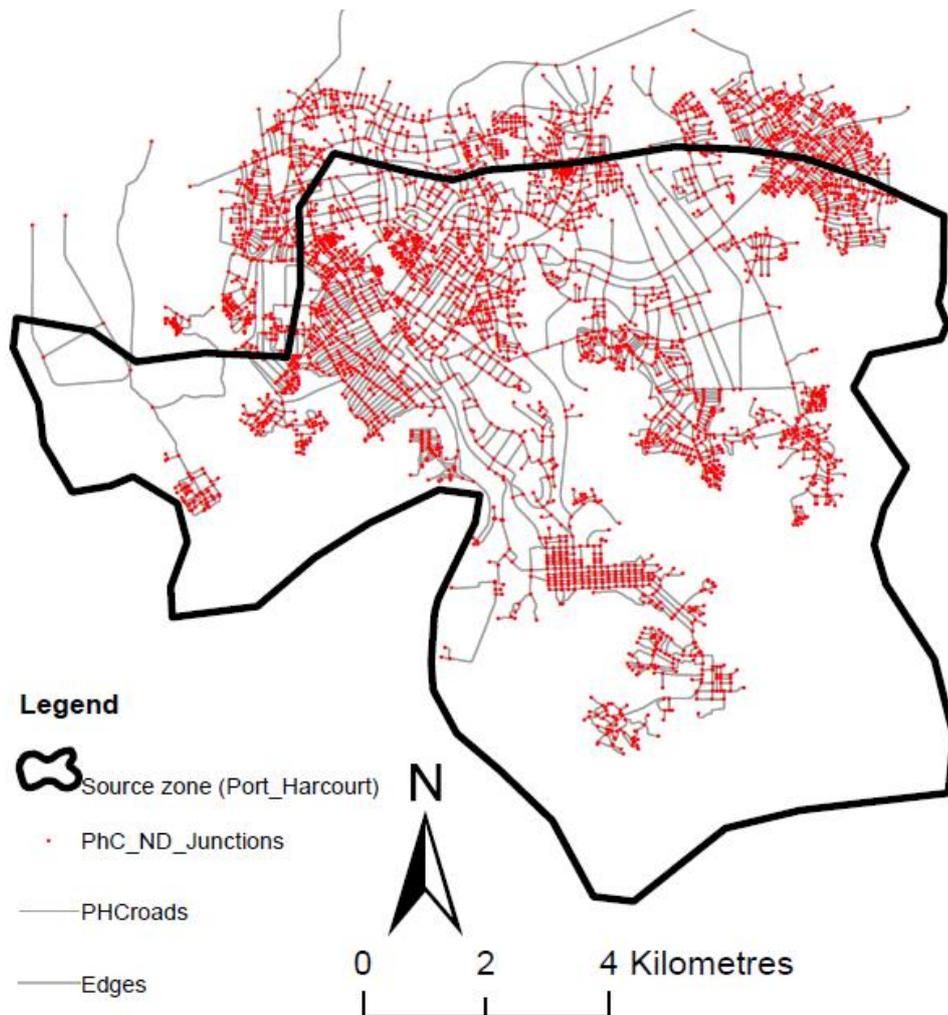


Figure 6.2 - The road network dataset showing roads (in grey colour) and road junctions in red dots.

### 6.2.1 Network distances

The Origin and Destination (OD) cost matrix was run in ArcGIS 10.2.2 to generate the network distances. The grid points representing estimates of the population created from

areal interpolation were used as the Origins and the point locations of current PHCCs were used as Destinations. Figure 6.3 shows the spatial distribution of 17 current PHCCs (Destinations) in Port-Harcourt over demand points generated from areal interpolation (Origins). The analysis calculates the distances between each Origin and each Destination and lists them in the *Line feature*. When displayed the *Line feature* maps a line between each Origin and each Destination (see Figure 6.4). The lines indicate the links between points and their length does not represent the actual distance on the road network between points. Figure 6.5 shows the distribution of lines on the road network. The attributes of the Line data are: distances between each Origin and each Destination point, ID of each Origin and each Destination and ranking with the shortest distance first. The accessibility to current PHCCs was measured based on the minimum distance between each demand and each PHCC on road network as has been done by Owen et al. (2010). The record of the shortest travelled distances on the road network between each Origin and each Destination point created from a GIS network analysis were selected and a layer was created.

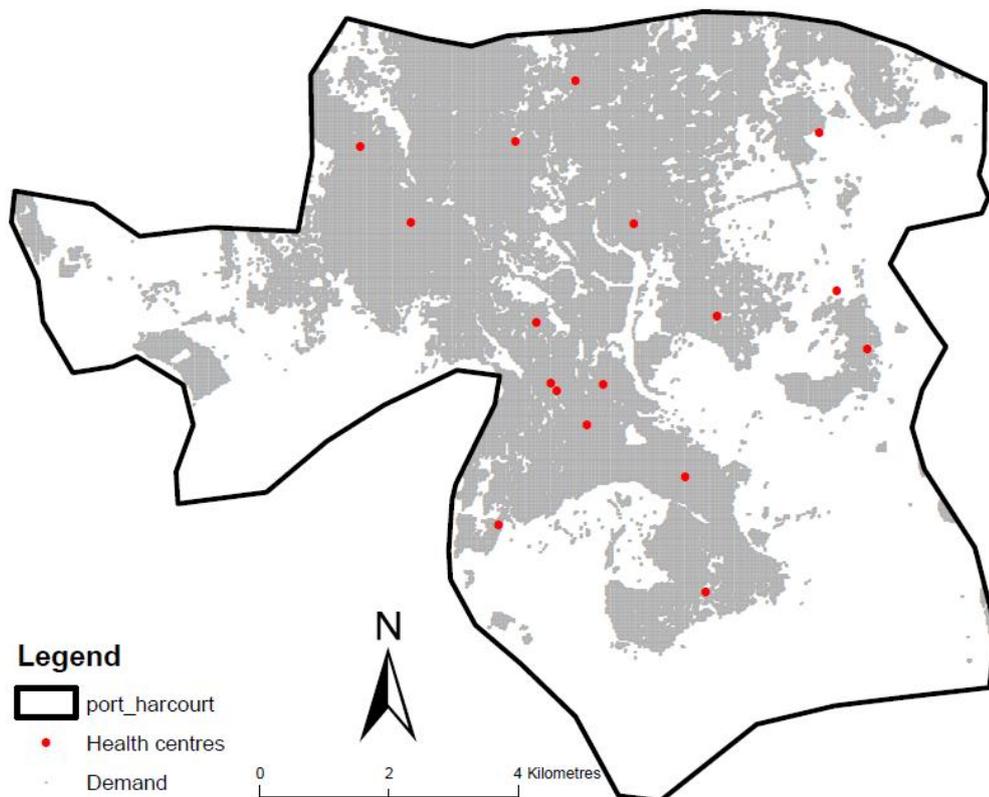


Figure 6.3 – The spatial distribution of 17 current PHCCs (red dots) and the demand points generated from areal interpolation (grey dots). The digital boundary is Copyright for Geotechnics Services 2011.

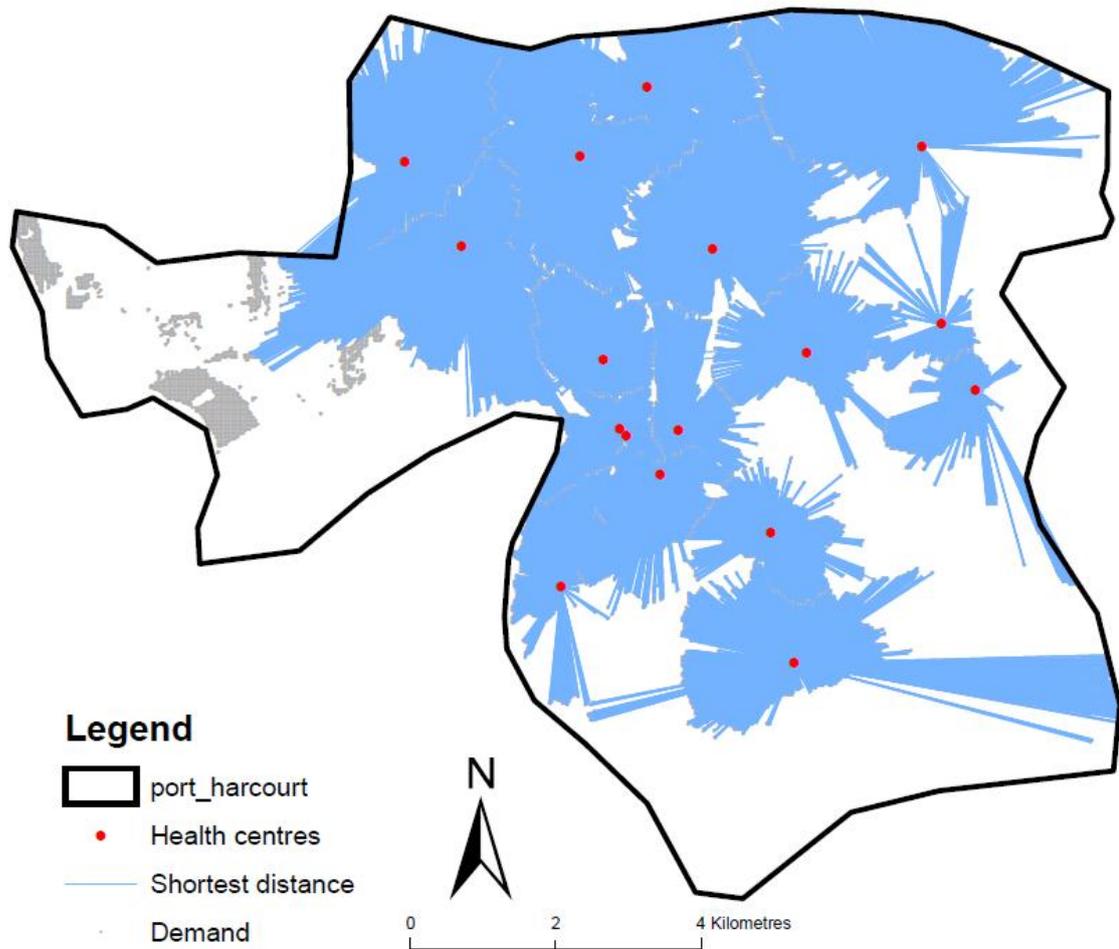


Figure 6.4 - Lines representing shortest distances from each demand to each health centre

A second distance matrix was created of distances between 85 grid points representing potential new locations of PHCCs and the demand points generated from areal interpolation. This was done to examine the potential of new locations of PHCCs and to demonstrate the selection of alternative locations of PHCCs. The grid points generated across the study area for potential new locations were spaced at 500m apart and chosen to be within 30m of an existing road. This was done for two reasons: first, to select future locations that are accessible by road; second, to be able to obtain the shortest travel distance on road network from each demand point to each potential new location. Figure 6.6 shows the spatial distribution of 85 potential new locations of PHCCs (red dots) and the demand points generated from areal interpolation (grey dots).

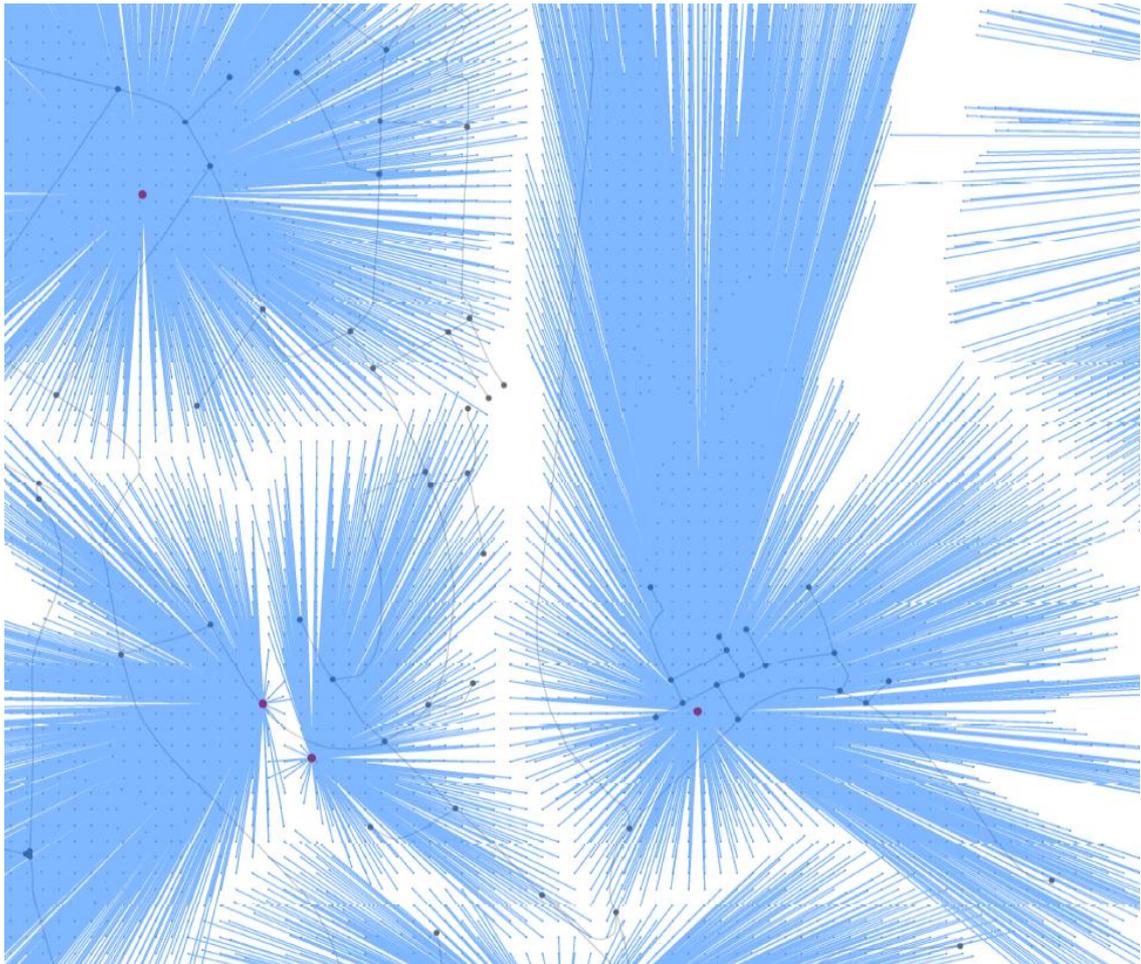


Figure 6.5 – The distribution of lines on the road network

The layers for the shortest travelled distances on the road network created using locations of 17 current PHCCs and 85 potential new locations of PHCCs were exported and used together with the demand points from areal interpolation as the input into Teitz and Bart's algorithm in R statistical software in order to allocate demand to PHCCs, evaluate 17 current locations of PHCCs, suggest alternative locations of PHCCs and optimise fewer locations of current and potential PHCCs to provide the same services.

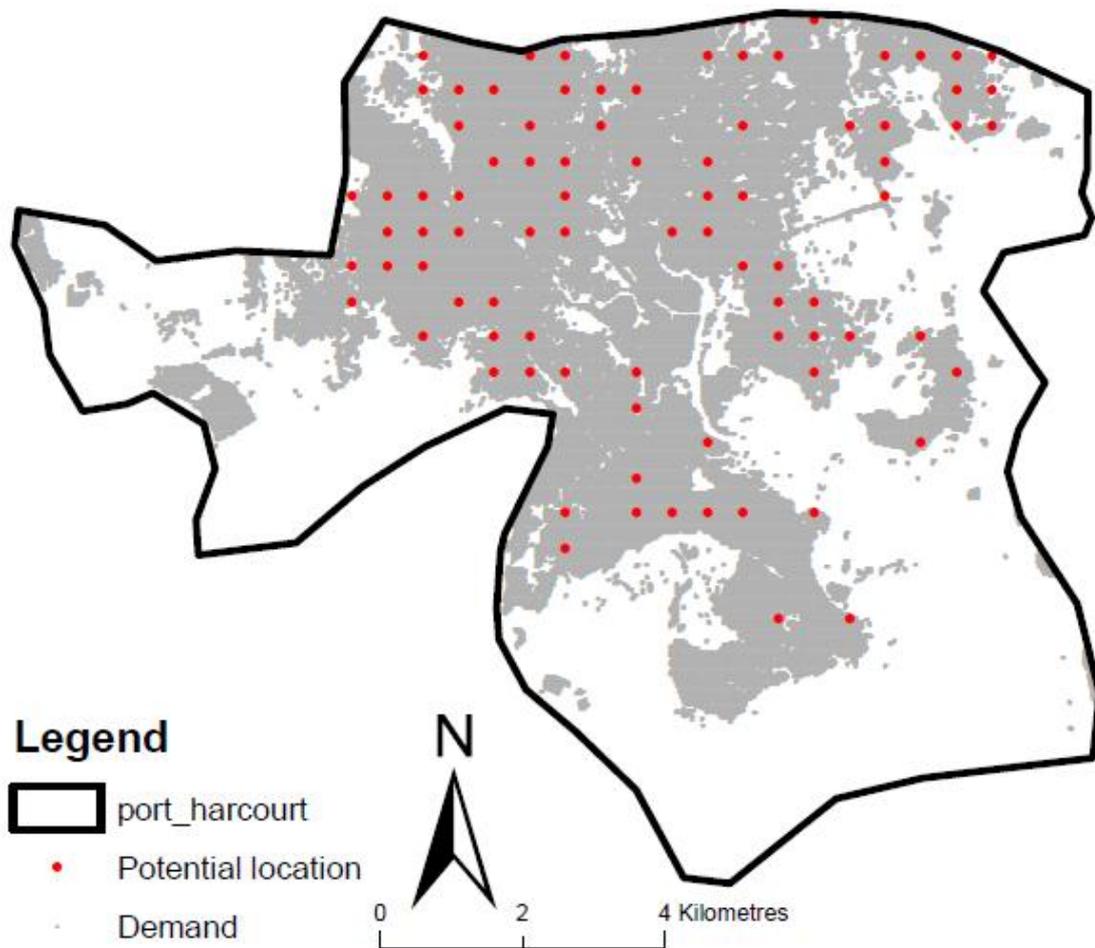


Figure 6.6 - The spatial distribution of 85 potential new locations of PHCCs (red dots) and the demand points generated from areal interpolation (grey dots). The digital boundary is Copyright for Geotechnics Services 2011.

### 6.3 Methods and models

The objective function of the p-median method is to select the required number of facilities among the total potential facilities that minimises the total weighted distance travelled between facilities and demand points. The Teitz and Bart's heuristic search algorithm requires the number of facilities to locate, the demand population values generated from areal interpolation and the distance matrix. The algorithm first randomly search for the required number of facilities, then substitute one of the selected with one not selected and tests to verify if the average weighted travel distance from demand to all locations is minimised. This is repeated until no distance is minimised by the substitution then the heuristic stops and the selected locations are assumed to be the

optimal locations. The use of interchange heuristic by Teitz and Bart's algorithm controls the selection of locations that are more likely to reduce the average weighted travel distance from demand to all locations.

Some assumptions were made by the Teitz and Bart's heuristic algorithm both on the demand and supply used. On the demand side, the heuristic assumes the majority of the population attends the nearest PHCC on a road network, and residents do not attend PHCCs outside the boundary of Port-Harcourt. Although in reality, PHCCs located outside the boundary of Port-Harcourt are likely to be a potential service centre and some residents are likely to attend PHCCs that are closer to their places of work and not residential homes. On the supply side, the heuristic assumes PHCCs have the same capacity and provide the same services but in reality they do not provide the same services.

Three analyses were run on the data to assess the locations of PHCCs in different situations: first, to evaluate current numbers of PHCCs; second, to suggest new potential locations; and third, changing numbers of PHCCs in current and new locations.

#### *Evaluate current numbers*

The distance matrix generated in section 6.2.1 for current numbers of PHCCs was used to classify PHCCs in terms of distances by identifying health centres nearest to each demand based on the demand weighted distance. The total demand allocated to each PHCC (catchment) is generated. This is important as it allocates each PHCC to each demand to suggest its catchment and assess its location by indicating the likely geographical coverage for each location. A location that minimises the total travel distance between a PHCC and its allocated demand is an optimal location. The R code used is available in Appendix 11.

#### *Suggest new potential locations*

The Teitz and Bart's algorithm was used to select a subset of 17 optimal locations from 85 potential locations. The suitability of optimal locations was assessed using the mean

distance between demand and the nearest selected new location. The algorithm allocates demand to each location and the mean distance between demand and the nearest optimal location was obtained. This was done to explore the potential of new locations and to demonstrate how the method could be used to select alternative future locations. The importance of this is to determine whether re-allocating some of the current PHCCs to new locations could improve overall accessibility in terms of reduced average distance travelled from demand to the nearest PHCC. The spatial distributions of the 17 optimal locations selected were mapped together with the location of 17 current PHCCs (see Figure 6.7 later). The R code used is available in Appendix 12.

### *Adjust the number of locations*

Policy makers and health system planners need to be able to develop methods for optimising current facility locations. One way of checking this is to change the number of locations and compare the mean distance of the changed locations with that of the current locations. If the changed locations minimise the total travelled distance of the current locations, then the changed locations are better than the current locations. The importance of this is to determine whether the current locations of 17 PHCCs could be improved in terms of minimising the average distance from demand to the nearest PHCC, thereby evaluating service delivery with fewer PHCCs based on improved locations. This will inform decision makers of the need to close some of the current PHCCs or add new ones.

The distance matrix for the current PHCCs and the interpolated populations were used as the input into Teitz and Bart's algorithm in R statistical software to generate an optimal distribution of subsets of different number of locations (from 5 to 16 PHCCs) from current locations (see Appendix 13). The spatial distributions of some of the selected locations from current PHCCs are shown later in Figures 6.9 to 6.11. The procedure was repeated using a distance matrix for potential future locations, in this case to select subsets of 5 to 20 locations from 85 potential sites (see Appendix 14). This was done to determine the benefits of increased or reduced numbers of PHCCs. A number of outputs were produced for each subset. These include: the total demand allocated to each potential location in the subset and the mean distances between each potential location in the subset and each demand within its catchment. The variation in

the average distance travelled from demand to the nearest location of current PHCCs was compared with that of potential locations of PHCCs for each subset (see Figure 6.12). The R code used to adjust the number of locations for both current and potential PHCCs is available in Appendix 15.

## **6.4 Results**

This section presents results of location-allocation for current PHCCs, potential locations of PHCCs and changed locations for both current and potential PHCCs.

### **6.4.1 Current locations of PHCCs**

Table 6.1 shows the total demand allocated to each of the 17 current locations of PHCCs and the mean distances between each PHCC and each demand within its catchment. The average mean distance is 1204m. The data in Table 6.1 show the first five PHCCs were allocated about 54% of the total demand with the remaining twelve PHCCs having about 46% of the total demand. Considering the assumption that these PHCCs have the same capacity and provide the same services, it is expected that the demand allocation should be equal for an optimal location. This means the current locations of PHCCs are not optimal. The implication of this result is that PHCCs with high demand allocation will be overstretched while those with low demand allocation will be underutilised. This result provides evidence for informed decision making in spatial planning and policy development.

#### **Key message: current PHCCs**

A person would have to travel an average distance of 1204m to access the closest PHCC in Port-Harcourt.

Table 6.1 Demand allocated to current PHCCs in Port-Harcourt

S/No	Health_centres	Demand	Demand (%)	Mean.dist (m)
1	OKIJA STREET	65001	12.68	1672
2	AZUABIE	57179	11.16	2717
3	AMADI-AMA	56645	11.05	1383
4	OROGBUM	53543	10.45	1263
5	MILE3 FSP CLINIC	41784	8.15	1343
6	BORIKIRI	41133	8.03	1306
7	ELEKAHIA	32059	6.26	1151
8	OZUBOKO	28393	5.54	946
9	POTTS JOHNSON	24911	4.86	1005
10	CHURCHILL STREET	23115	4.51	813
11	BMH IMMUN. POST	20683	4.04	1132
12	MARINE BASE	18337	3.58	1156
13	BUNDU AMA	17263	3.37	968
14	ABULOMA	12838	2.51	784
15	BANK ROAD	10203	1.99	1472
16	OKURU-AMA	5273	1.03	897
17	CITY COUNCIL	4090	0.80	465

#### 6.4.2 Potential locations of PHCCs

Table 6.2 shows the total demand allocated to each of the 17 optimal locations of PHCCs selected from 85 potential new locations and the mean distances between each optimal location and each demand within its catchment. The average mean distance for the 17 optimal locations was found to be 1074m. This indicates a reduction in the average distance travelled when compared with the current locations of PHCCs by 130m. This suggests an improvement in accessibility could be achieved with some re-allocations to potential locations. The spatial distribution of the optimal locations (red circles) were mapped together with the current locations of 17 PHCCs (blue crosses) as shown in Figure 6.7.

Table 6.2 Demand allocated to potential PHCCs in Port-Harcourt

S/No	Health_centres	Demand	Demand (%)	Mean.dist (m)
1	2	43625	8.51	1704
2	41	43542	8.50	1115
3	10	40839	7.97	1141
4	62	38070	7.43	1256
5	30	37194	7.26	1053
6	68	35797	6.99	1080
7	56	35755	6.98	1001
8	33	34986	6.83	1113
9	71	31854	6.22	1045
10	46	31305	6.11	1038
11	16	23883	4.66	1129
12	8	21508	4.20	784
13	42	21501	4.20	795
14	4	19265	3.76	869
15	15	18668	3.64	1173
16	19	17498	3.41	1008
17	72	17160	3.35	947

**Key message: potential locations**

Re-allocating current PHCCs to potential locations reduces the average distance travelled from demand to nearest health centre by 130m.

**6.4.3 Adjust current locations**

The Teitz and Bart’s heuristic algorithm was applied to select sets of 5 to 16 PHCCs from current locations to determine the benefits of reduced numbers of PHCCs. A number of outputs were produced for each subset and the mean distances to the nearest PHCC (in metres) were plotted against the number of PHCCs in each subset (see Figure 6.8).

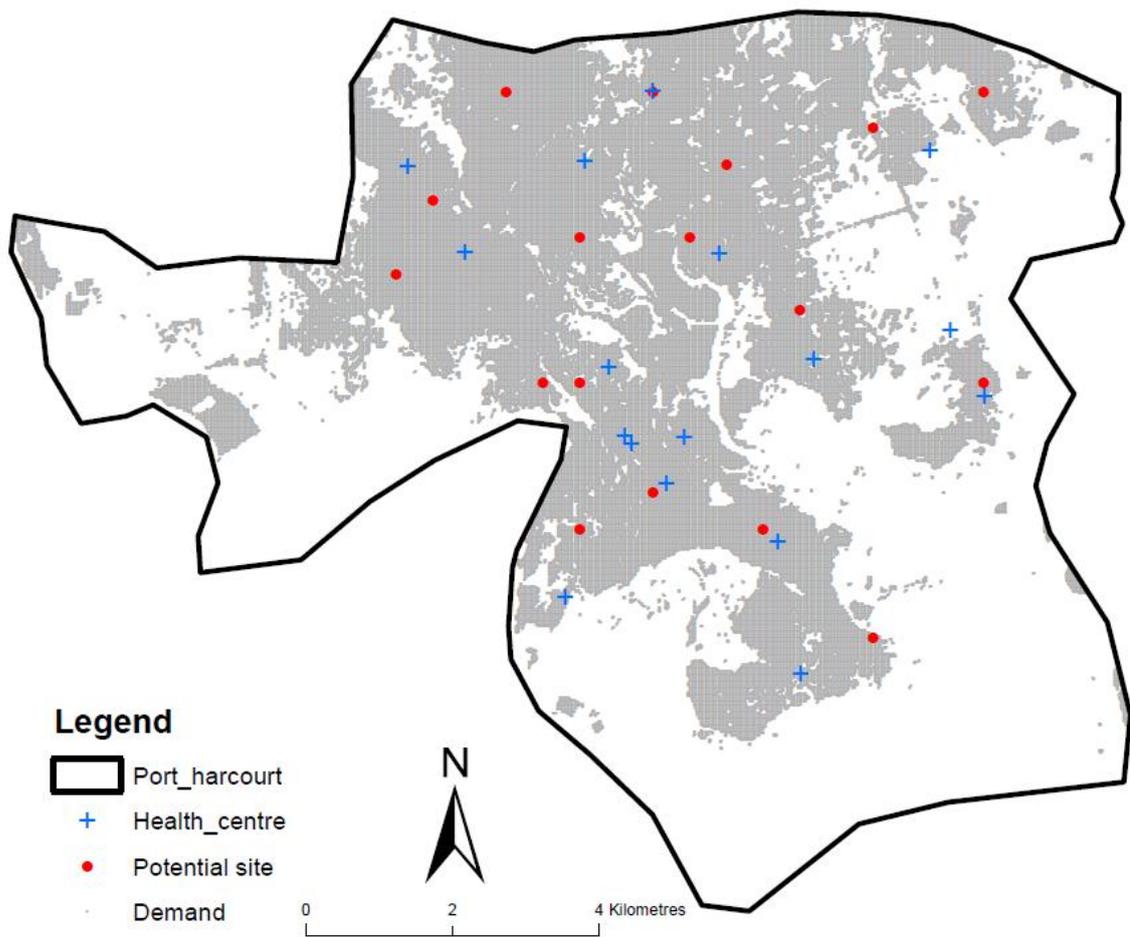


Figure 6.7 - Spatial distribution of 17 optimal locations selected (red circles) from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geotechnics Services 2011.

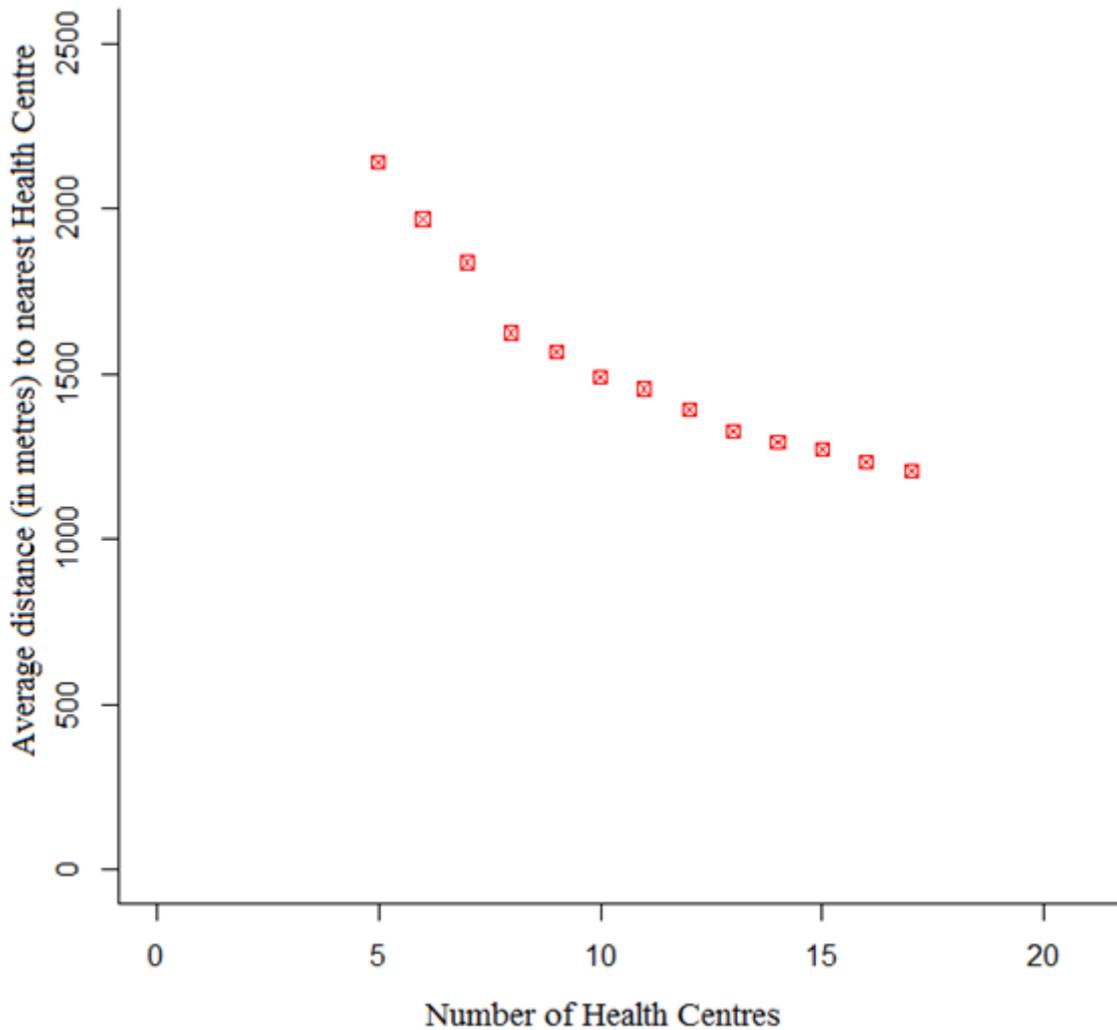


Figure 6.8 - Average distances to the nearest current PHCCs plotted against the number of current PHCCs in a subset.

Figure 6.8 shows a decrease in average distance to the nearest current PHCC as the number of PHCCs selected from 17 current locations in a subset increases. The graph shows a clear direction with the overall pattern moving from upper left to lower right. That is, an increase in the average distances to the nearest PHCC with a reduction in the number of locations in a subset. This suggests no improvement in the changed locations from current PHCCs was achieved in terms of reduction in the average travelled distance. This is not surprising as the average persons' distance to the nearest PHCC is expected to reduce when a new PHCC is added to the current ones. It is worth noting that the total average travelled distance for all demand for the 17 current PHCCs increases by about 120m when 13 optimal locations were selected from the 17 current PHCCs. This means that 13 optimal locations selected from the 17 current locations of

PHCCs are almost as good as the 17 current PHCCs. This suggest almost the same level of accessibility as is currently achieved could be matched by using fewer but more optimally sited service provision points from the current locations of PHCCs, thereby potentially reducing the operating cost of these PHCCs. In terms of coverage, the model allocates all demand to all facilities in the subset. The spatial distributions of some of the selected locations of current PHCCs are shown in Figures 6.9 to 6.11 with the locations of selected PHCCs shown in red dots while those not selected are shown in blue dots.

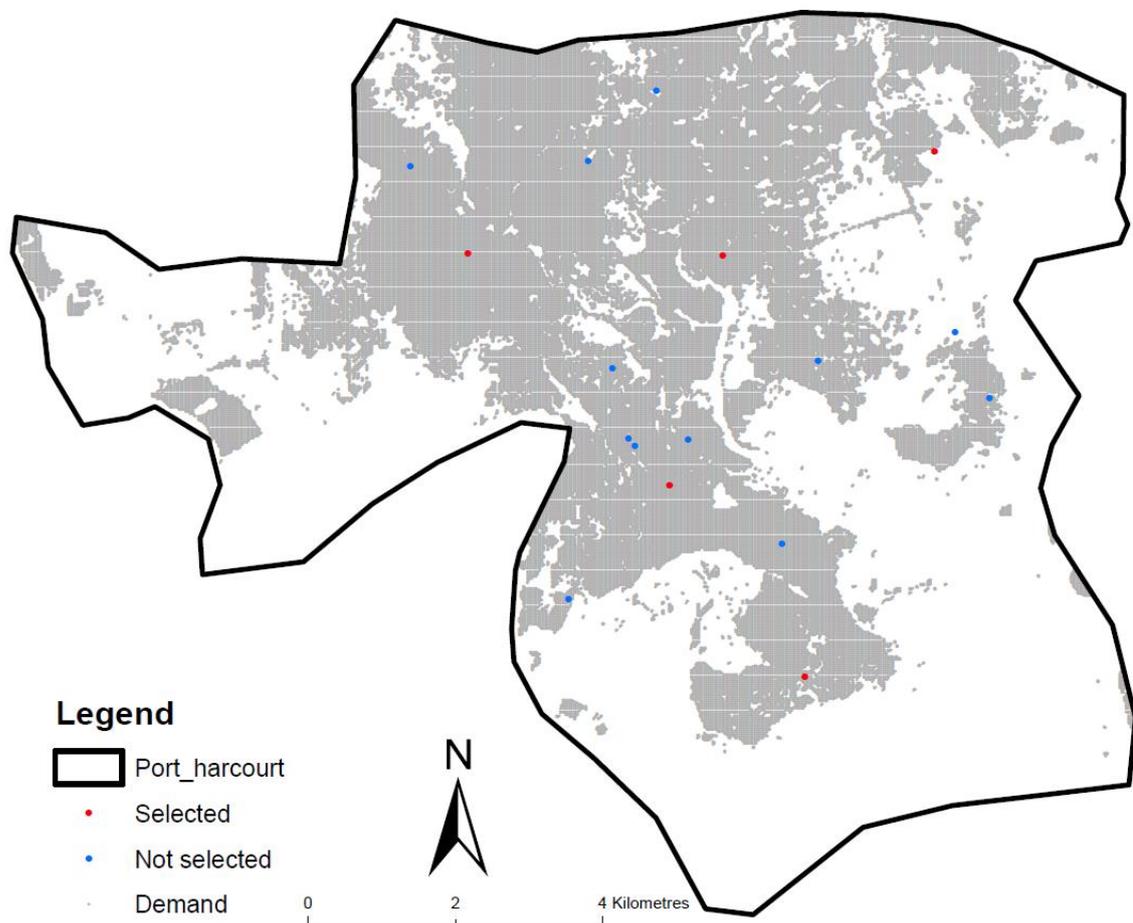


Figure 6.9 - The spatial distributions of selected locations of 5 PHCCs from current locations. The digital boundary is Copyright for Geotechnics Services 2011.

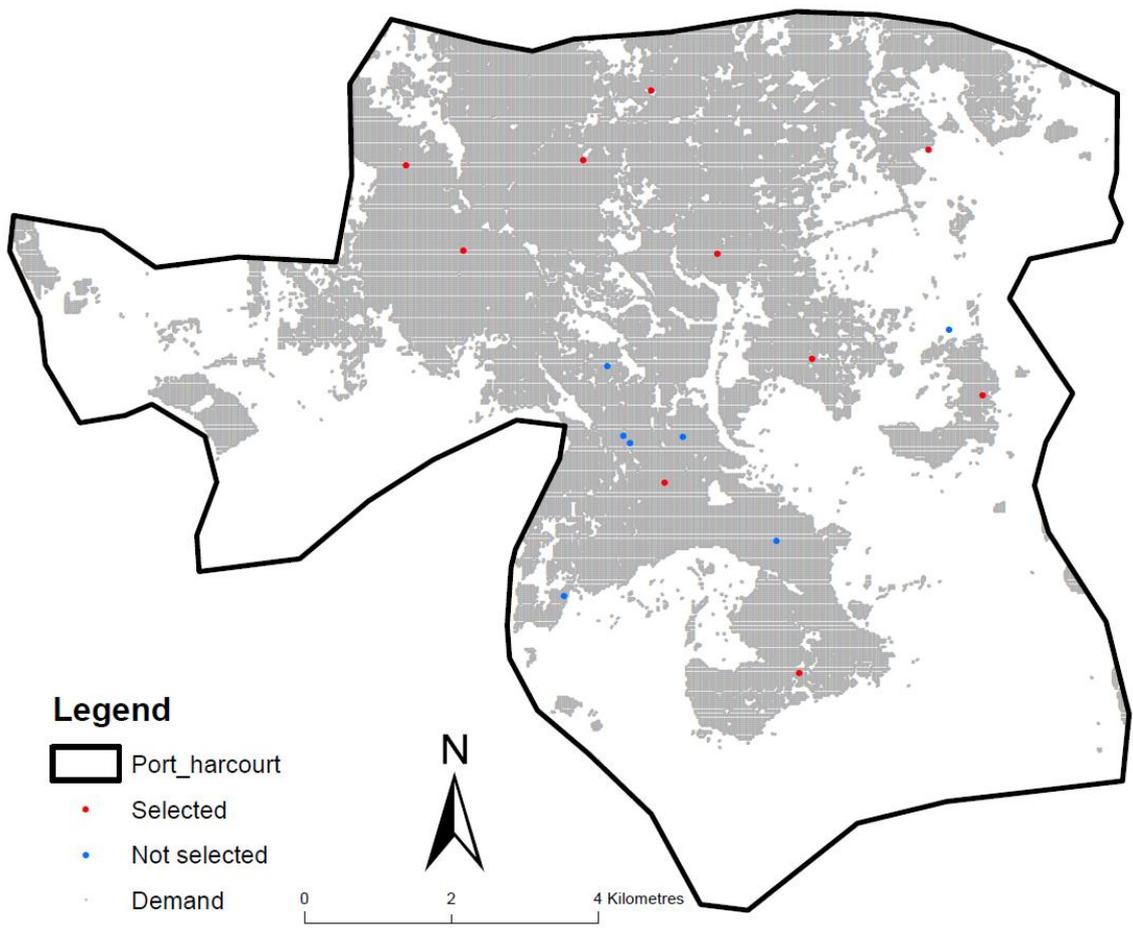


Figure 6.10 - The spatial distributions of selected locations of 10 PHCCs from current locations. The digital boundary is Copyright for Geotechnics Services 2011.

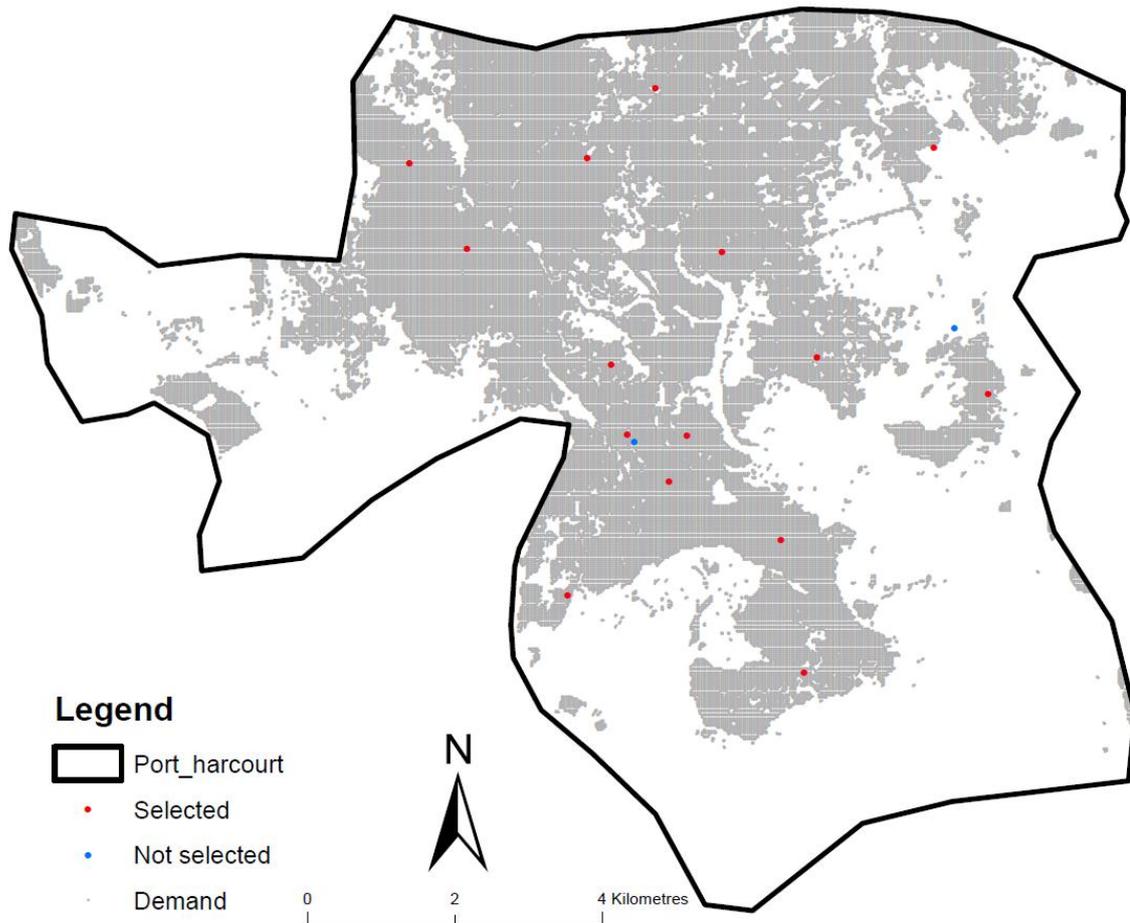


Figure 6.11 - The spatial distributions of selected locations of 15 PHCCs from current locations. The digital boundary is Copyright for Geotechnics Services 2011.

#### 6.4.4 Adjust potential locations

The Teitz and Bart's algorithm was also applied to select sets of 5 to 20 locations from 85 potential new locations to determine the benefits of increased or reduced numbers of PHCCs. A maximum number of twenty locations were selected. This is based on the recommendation of the National Health Policy that at least one PHCC should be sited in each electoral ward (NPHCDA 2007) and there are twenty electoral wards in Port-Harcourt. Table 6.3 shows the current distances, modelled distances and the difference between current and modelled distances for specific number of PHCCs used as the subsets. The average distances (in metres) to the nearest current locations (current distances) and optimal locations selected from 85 potential new locations (modelled distances) were plotted against the number of PHCCs in a subset (see Figure 6.12).

Table 6.3 Current and modelled average distances for facilities

PHCCs	Current_distances	Modelled_distances	Difference_metres
5	2141	2017	124
6	1968	1839	129
7	1837	1716	121
8	1624	1604	20
9	1565	1476	89
10	1488	1348	140
11	1454	1307	147
12	1392	1255	137
13	1324	1206	118
14	1292	1162	130
15	1270	1130	140
16	1235	1105	130
17	1204	1074	130
18	-	1046	1046
19	-	1033	1033
20	-	996	996

The graph shows a clear direction with the overall pattern moving from upper left to lower right. That is, the average travelled distance for both current and modelled distances reduces as the number of locations in the subset increases. This is expected as the average distance to facilities from homes is more likely to reduce as the number of facilities increases. The graph also shows the modelled distances minimised the current distance travelled for all the subsets. The Teitz and Bart's algorithm allocates total demand to all facilities and in terms of the total average travelled distance for all demand, the results suggest that 13 optimally sited service provision points selected from 85 potential new locations are as good as the 17 current locations of PHCCs. This suggests the potential locations are better than the current locations because with fewer resources (13 service provision points) the same demand coverage would be achieved, thereby potentially saving money, or alternatively that spending the same money on the same number of provision points but with these more optimally sited would offer a useful improvement in net accessibility to the service as it reduces the average travelled

distance from demand to nearest health centre by 130m. The result is similar to findings of Comber et al. (2010) who applied Grouping Genetic Algorithm (GGA) to identify sets of optimal ambulance locations to ascertain the advantages of having fewer ambulance site locations in Niigata city in north-western Japan. They found 27 optimal sites for ambulances to be located, that is, four new ambulance locations were added to existing 23 locations, significantly improves the average emergency medical services (EMS) response time by 1 minute 14 seconds compared to using 35 current locations. This is important as it provides spatial evidence to support planning and allocation of future resources.

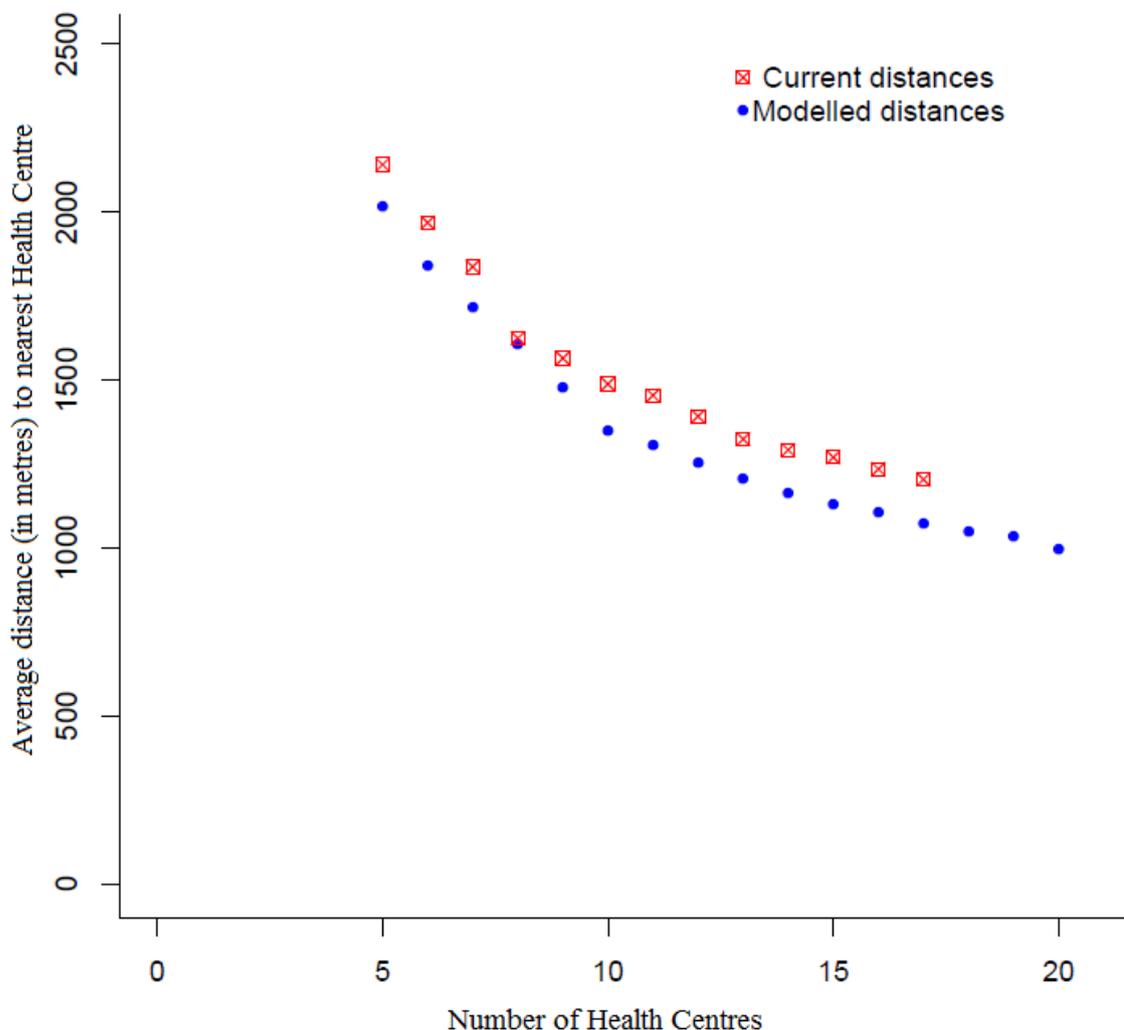


Figure 6.12 - Variation in average distance travelled from demand to: (⊕) nearest location of current PHCCs (current distance); and (•) nearest potential location selected from 85 potential new locations.

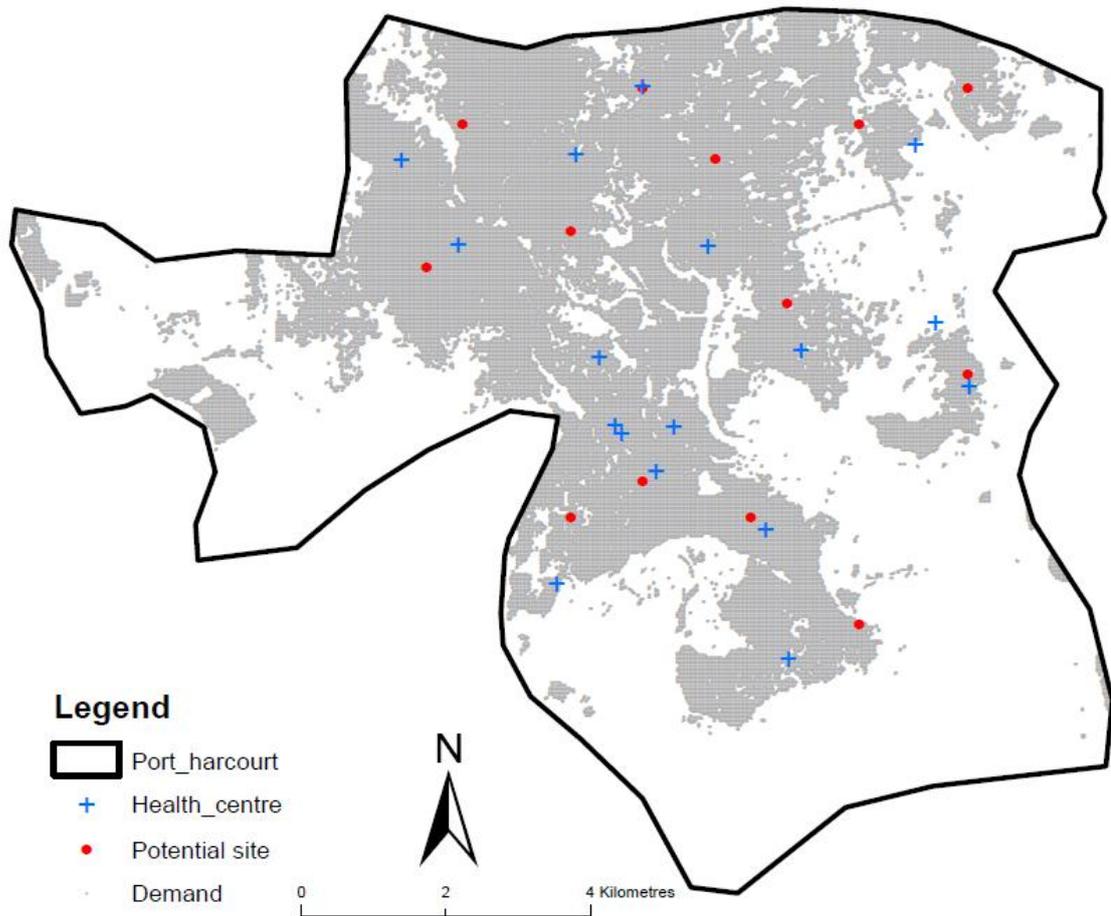


Figure 6.13 - The spatial distribution of 13 optimal locations (red circles) selected from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geo-technics Services 2011.

## 6.5 Summary

This chapter evaluates the distribution of current locations of PHCCs in Port-Harcourt using the demand population values created from areal interpolation. The results show a person would have to travel an average distance of 1204m to access the closest PHCC in Port-Harcourt. The results also show re-allocating current PHCCs to potential locations would reduce the operating cost as fewer resources would achieve almost as good as the current demand coverage with the same average travelled distance. Redistributing summary population totals over small areas, and using the spatially distributed demand points created to evaluate location decisions provides health policy planners with evidence to support policy decisions.

## Chapter 7

### 7. DISCUSSION

#### 7.1 Introduction

Census data at the small area level are unavailable in most parts of the world, especially in developing countries. In such places the need to estimate aggregate population counts over small areas to represent demand (Cromley et al. 2012; Tomintz et al. 2013) is critical in order to support spatial planning and policy development. More specifically calibrated solutions for estimating populations over small areas are needed to support facility location analyses. Many studies have applied different areal interpolation techniques in order to generate spatially distributed estimates of the population over small areas (Markoff and Shapiro 1973; Tobler 1979; Goodchild and Lam 1980; Lam 1983; Flowerdew and Green 1991; Langford et al. 1991; Goodchild et al. 1993; Burrough and McDonnell 1998; Eicher and Brewer 2001; Mennis 2003; Kim and Yao 2010; Cromley et al. 2011; Qiu et al. 2012; Langford 2013). Recent studies have evaluated the use of areal interpolation techniques for the estimation of summary population totals over areas smaller than the U.K. finest census division (Langford 2013). However, to date, little research has estimated population surfaces in areas where the actual distributions are unknown and there is no published research reporting the use of areal interpolation techniques to disaggregate census data over small areas in any part of Nigeria. One possible reason is the difficulty in validating population allocations for small areas in such places by comparison with target units whose populations are known, as is common practice elsewhere.

Some research has considered the validation of spatially allocated small area estimates from census data. This includes the use of historical census data for which a full population count is available at a fine spatial resolution (Ruther et al. 2013). Other work by Amaral et al. (2012) used dasymetric interpolation of population data in the Brazilian Amazon and the methodology was first developed for the municipality of Maraba, Para State and then adapted with local parameters for thirteen municipalities of the

Sustainable Forests District (SFD) generating population distribution surfaces for 2000 and 2007.

The research reported in this thesis estimated population surfaces in areas where actual distributions were unknown. It explored the different interpolation methods applied at different scales in areas where the actual population distribution is known and where validation is possible. It then applied the parameters developed from these results to areas where the distribution was unknown and used the population surfaces generated in this way to evaluate current health facility locations and to suggest alternative spatial arrangements of facilities to improve spatial accessibility. The key finding in this research is for locations where spatially detailed estimates of the population do not exist, that is, where the actual population in small areas is unknown, summary population totals can be redistributed over small areas by adapting a model that was developed for areas where validation is possible with land cover and census data for that area as a ‘best bet’ solution, given the circumstances. Addressing problems in spatially detailed estimates of the population in locations where small area population summaries are not available and using these estimates to evaluate facility location provides evidence to support spatial decision-making and planning.

The next section discusses surface generation, the results of areal interpolation techniques for Leicester. Section 7.3 discusses the results of areal interpolation and location-allocation for Port Harcourt. Section 7.4 discusses the methods employed and their assumptions. Section 7.5 reflects on the results relating to methods. The last section discusses the limitations and suggestions for future work.

## **7.2 Leicester: surface generation**

This study applied the binary dasymetric and pycnophylactic interpolation methods across different spatial scales over Leicester area. Leicester UA boundary was used as the source zone for the binary dasymetric method while Leicester UA together with all the surrounding UAs with which it shares a common boundary were used as source zones for the pycnophylactic interpolation method, to derive an interpolated gridded population surface at resolutions of 100m and 30m postings, to then be aggregated to

MSOA, LSOA and OA target units and assessed for accuracy using known census counts in each case.

The interpolated gridded pycnophylactic population surfaces created for Leicester are raster grids of 30m and 100m resolutions as shown in Figures 4.10 and 4.11 respectively. The analysis over the source zones showed that the pycnophylactic method locally overestimated actual population in larger census units within each source zone and underestimated the actual populations in small census units within the same source zones. The spatial distribution of residuals from the gridded pycnophylactic population surfaces also showed that relatively large source zones tend to be overestimated while relatively small source zones tend to be underestimated. Similar patterns have been found by other researchers (e.g. Mennis and Hultgren 2006; Eicher and Brewer 2001), where relatively large rural blocks tend to be overestimated while relatively small urban blocks tend to be underestimated. In this study, the underestimated census units are mainly the smaller census units in the more densely populated areas such as the city centre while the overestimated census units are the larger spatial units in the less densely populated areas away from the city centre. Hawley (2005) reported similar patterns when he used the pycnophylactic interpolation.

Some authors have reported that the pycnophylactic method provides more accurate results than other areal interpolation techniques that do not make use of ancillary data (such as areal weighting), when compared (Hawley 2005; Kim and Yao 2010). Other research assigned a non-zero population density value in every location using the pycnophylactic interpolation in combination with dasymetric mapping to provide the best fitting target zone estimates (Comber et al. 2008b; Kim and Yao 2010). This study did not consider the combined method as it aims to compare the performance of the binary dasymetric and pycnophylactic methods evaluated at different spatial scales.

The dasymetric surfaces generated over Leicester (see Figures 4.7 to 4.9) showed that the use of ancillary information to constrain the population distribution improves the accuracy of areal interpolation from that provided by the pycnophylactic method. The dasymetric models using ancillary data to constrain the re-allocations generally performed better than the pycnophylactic approach. This is because the pycnophylactic method has no detailed information that correlates to likely true distribution of the

population in the source zone. This result is similar to findings of other researchers that compared the results of dasymetric method with other areal interpolation techniques and found the technique consistently provided better target zone estimates (Langford 2013; Mennis and Hultgren 2006; Poulsen and Kennedy 2004; Eicher and Brewer 2001; Martin et al. 2000; Cockings et al. 1997; Fisher and Langford 1995; Langford et al. 1993). The results also indicate that larger census units are more likely to produce better results as it shows improvements in RMSE and the values of CoV as the size of the spatial aggregation increases. This result is not surprising as one would expect higher accuracies when values are disaggregated over coarser spatial units. This result is similar to findings of Comber et al. (2008) where a combination of pycnophylactic interpolation with the dasymetric method was used to create the National Agricultural Land Use Dataset. They reported improvement in  $R^2$  and RMSE values for Arable and Grass land uses for Kent, U.K. as the size of the spatial aggregation increases by the plots from 1 km<sup>2</sup> to 25 km<sup>2</sup>.

Most dasymetric research has used land cover information derived from classified satellite imagery as the ancillary data input (Mennis 2003; Zandbergen and Ignizio 2010; Langford 2013). However, deriving such information from classified satellite imagery requires specialized skills. Also, identifying residential land use and whether they are truly occupied from classified satellite imagery is another error associated with using land cover data derived from classified satellite imagery as the ancillary data input. This study evaluated land cover data classified from satellite imagery and resampled aerial photograph of differing spatial resolutions (30m, 10m and 3m) as ancillary data input to the dasymetric method. The results from the error analyses showed that the land cover data derived from classified Landsat7 (ETM) 30m spatial resolution ancillary data input provided the lowest values of RMSE and CoV for all the models tested. These results, along with the free availability of 30m spatial resolution remote sensing data (Landsat etc.) and the ease with which it can be classified into urban and non-urban areas suggests its suitability as input for the dasymetric method. This study found that land cover data derived from classified resampled aerial photo data of 10m and 3m spatial resolution when used as the ancillary data input for the dasymetric method does not improve the results as well as the additional processing time and storage compared to land cover data derived from classified Landsat7 (ETM) 30m spatial resolution. The results indicate that the quality of the land cover data is not

as important as its ability to predict the population estimates. This is likely due to the ability of multispectral satellite imagery (Landsat7 ETM) to differentiate land cover classes. Also, the performance of 10m and 3m resampled aerial photo data can be attributed to using land cover information of different resolutions of the same source. Nagle et al. (2014) developed the penalized maximum entropy dasymetric model (P-MEDM) and found that the land cover data derived from high-resolution imagery provided less accurate information about the population than lower-resolution image data. Similarly, Langford (2013) found OS VectorMap District (raster-mode processing) with 5m resolution recorded RMSE value of 153 compared to Enhanced Thematic Mapper Plus (ETM+) of 30m resolution with a recorded value of 145.

### **7.3 Port-Harcourt: surfaces and location-allocation**

The dasymetric method was used to redistribute 2006 summary population totals for Port-Harcourt with ancillary land cover information derived from classified Landsat7 (ETM+) 30m spatial resolution, classified Spot5 (colour) 10m spatial resolution and classified resampled quickbird (60cm) image of 3m spatial resolution, over a 30m square grids. The parameters used to generate these surfaces for Port-Harcourt were based on the areal interpolation for Leicester where the actual population distribution is known and the results were validated. In the absence of field information (survey data) or another data source to validate these population values, a random sample of 200 locations was taken to visually inspect the results in order to assess the accuracy of the population distribution. Google Earth 7.1 was used as a reference, to inspect the surfaces. The results of the inspection show sixteen out of two hundred random points selected (eight per cent of the random points selected) identified as populated on the surface were actually in unpopulated areas. There were only ten out of two hundred random points selected (five per cent of the random points selected) identified as unpopulated on the surface but were actually in populated areas. A total of twenty-six out of two hundred random points selected (thirteen per cent) on the interpolated surface do not correspond with Google Earth reference. When 10m and 3m spatial resolution land cover data were used, 14.5 and 22 per cent respectively of the random points selected were not correctly redistributed. This difference may be partly due to classification error but it is more likely due to differences in spatial limits and temporal differences between the date of the Google Earth reference (20<sup>th</sup> December 2013) and

the date the Landsat7 (ETM+) imagery used to derive the land cover data was acquired (8<sup>th</sup> January 2003). This is a potential source of error which may have affected the result due to possible land cover changes between the periods. However, it was useful in inspecting the population surfaces. Other advantages of using Google Earth as a validation tool are that it is easily and freely accessible and it allows visualization of large datasets of high resolution images. This approach is similar to a recent study by Amaral et al. (2012) where the dasymetric method was applied over the municipality of Maraba, Para State and then adapted using local parameters for thirteen other municipalities, and used the population data for the official settlement projects in 2003 from the National Institute of Colonization and Agrarian Reform (INCRA) as a reference for validation.

The spatially distributed demand population values generated for Port Harcourt were used as input into location-allocation models with road network data to analyse the accessibility of current and alternative future locations of health centres. Specifically, the demand population values were used to allocate catchment areas for the PHCCs representing the expected geographical coverage for each PHCC and to analyse travel distance which gives information about the benefits of locating different numbers of PHCCs. Cromley et al. (2009) and Langford and Higgs (2006) have discussed the significant contribution of the spatial distributions of demand to potential measure of access to health facilities in the absence of small area census data. This study evaluated the locations of 17 current PHCCs in Port-Harcourt. The results showed the average distance to nearest PHCC to be 1204m. This distance is within the international criteria of 2km maximum distance a patient is expected to travel to an urban health centre that provides a comprehensive primary care as outlined by the Centre for Health Policy (1993) and discussed in Rispel et al. (1995) and Doherty et al. (1996).

When alternative locations for the current PHCCs were identified, the results showed the average distance to the closest facility to be reduced by 130m. It can be argued that for this set of experiment, the results show a little improvement in the current locations of PHCCs but the study suggests some improvement in accessibility to PHCCs could be achieved when alternative locations are used. Other analyses quantified the impacts of locating different numbers of PHCCs in terms of average travel distances for both 17 current PHCCs and 85 potential new locations. The results showed modelled distances

from potential new locations minimised the current distances for all the subsets, suggesting the methods applied in this study optimised current facility locations.

Travel distance to health facilities from homes has been shown to be a barrier to accessing health centres (Payne et al. 2000; Brewer et al. 2012) and correlates with the rate of child mortality (Frankenberg 1995; Tanser et al. 2006). Many studies have reported reduced childhood vaccination coverage with increased travel distance from homes (Jamil et al. 1999; Acharya and Cleland 2000; Tanser et al. 2006). Some other studies have reported strong association between routine childhood vaccinations and survival among infants in developing countries (Kristensen et al. 2000; Arevshatian et al. 2007) as these vaccinations guard against life-threatening diseases and reduce the spread of these diseases (Gu et al. 2010). Potential beneficiaries are more likely to complete routine childhood immunization if the services are provided within a reasonable travel distance from their homes. The implication of lack of this is that more children remain unvaccinated and are exposed to risk of death from preventable diseases. Therefore, optimal location of PHCCs is of critical importance to provision of adequate care as it is more likely to increase patients' attendance to PHCCs and reduce outbreak of diseases, thereby reducing the death of children below the age of five (Jamil et al. 1999; Hemat et al. 2009).

#### **7.4 Methods**

This section discusses areal interpolation and location-allocation methods used in this research. The dasymetric method was used to redistribute summary population totals over small areas for Leicester and Port-Harcourt while the pycnophylactic method was used for only Leicester area. For the location-allocation case study in Port-Harcourt, the p-median method implemented in Teitz and Bart's heuristic algorithm (Teitz and Bart 1968) was used. The differences in population estimates obtained were due to the adoption of different areal interpolation methods.

The dasymetric representation provided detailed and more realistic understanding of population distribution within the source zone, and therefore should provide better target zone estimates than the pycnophylactic method. The dasymetric method used ancillary input data on urban extent that relates to spatial distribution of the population

in the source zone which could strongly support its output. Also, the assignment of population totals to only areas identified as populated would seem to provide a more likely representation of where people truly live. This study has shown that the choice of areal interpolation method has a substantial impact on the estimates in the target zones. Some authors that compared population estimates from the dasymetric method with other areal interpolation techniques have reported similar findings (Langford 2013; Kim and Yao 2010; Eicher and Brewer 2001; Martin et al. 2000; Langford et al. 1993; Fisher and Langford 1995).

This study applied the binary dasymetric method which assumes population is most likely to be found in only urban areas, and therefore assigned a weighting factor of 1 for the residential areas and 0 for non-residential areas. The population surfaces produced in this way depict the underlying population distribution in the source zone. However, the binary dasymetric approach assume population density is uniformly distributed across the populated regions within each source zone, which is not likely to be true (Mennis and Hultgren 2006; Maantay et al. 2007). For this reason, there is the possibility that a multiclass dasymetric model may likely provide some insights since “*a range of residential densities will be present in most census reporting zones*” (Langford 2006, p.167). The multiclass dasymetric approach was not used in this study because previous research by Eicher and Brewer (2001) and Langford (2006) evaluated three-class dasymetric model but the results when compared do not show any significant benefit over the binary dasymetric method.

The use of other ancillary data such as street segments (Xie 1995; Riebel and Bufalino 2005; Tapp 2010) could provide more accurate target zones estimates in Leicester where streets are regularly spaced and houses are located at regular intervals on roads. This is because the approach assumes residential population density gradient is constant at a given distance from the nearest road or street. However, this method is not likely to work well in Port-Harcourt because a number of housing units are not located close to the road network. For this reason, local circumstance and available resources were considered prior to the decision on the ancillary data input to use.

This study has shown that the resolution of land cover data used to constrain the re-allocation could affect the population surfaces. The evaluation of land cover classified

from satellite imagery of differing resolutions was influenced by its ability to reveal spatial variations within the study area and to explore whether a more detailed ancillary data input could reduce the estimation error (or not), as suggested by Zandbergen and Ignizio (2010). The expectation is that high resolution satellite image, which appears to offer greater spatial precision in identifying urban extent, could lead to reduction in land cover classification error. This is because the spectral signatures for each land cover type are likely to generate as little confusion as possible with a clear separation of land cover classes before classification. This is more likely to improve the redistribution to only areas identified as populated. But the results presented in this research showed that land cover data derived from classified resampled aerial photo data of 10m and 3m spatial resolution when used as the ancillary data input does not improve the results of the redistribution compared to those derived from classified Landsat7 (ETM) 30m spatial resolution. This is likely due to the ability of multispectral satellite imagery (Landsat7 ETM) to differentiate land cover classes. Also, the performance of 10m and 3m resampled aerial photo data can be attributed to using land cover information of different resolutions of the same source.

This study has also shown that the size of the support grids could have an effect on the population estimates generated in the target zones. This study redistributed summary population totals to two different support grids (30m and 100m) for each method. A 30m support grid was chosen to correspond to the resolution of the Landsat (ETM) image used, as has been done by Su et al. (2010). Also, multiples of 10m and 3m image resolutions will correctly match the 30m grids of cells used. On the other hand, a 100m grid was chosen although multiples of 10m image resolutions can fit into 100m grids but not that of 3m and 30m. This was chosen to examine the benefit or otherwise of redistributing aggregate population to target areas whose resolution does not correspond with the resolution of the image. The results of this study showed using smaller support grids provided better population estimates in the target zones compared to using larger support grids.

There are other heuristic searches that could have an effect on optimal locations of facilities when used to solve the p-median problem. The use of Genetic Algorithm (GA) that employs natural selection similar to that in biological evolution is likely to provide different results, but GA is more suitable where there are large numbers of facilities to

locate (Comber et al. 2010). For this reason, GA was not used in this study as there were only 17 PHCCs to locate. This study assessed the suitability of optimal locations using the mean distances between demand and the nearest location of PHCCs as has been done by other researchers that have used Teitz and Bart's algorithm to solve the p-median problem (Rosing et al. 1979; Church and Sorensen 1994; Rahman and Smith 2000; Jia et al. 2007; Tomintz et al. 2013).

## **7.5 Reflections on the methods and results**

Geographic data are often aggregated into areal units for a number of reasons. For example, population census counts are aggregated to reduce data volume and maintain confidentiality and respondent anonymity. When data come from different sources for the same geographic area, they incorporate alternative spatial aggregations resulting in different sets of areal units. This changes the scale and the aggregated boundary, and to a large extent would affect the results of spatial analysis (Openshaw 1984). There is the need to disaggregate the data for a common geographic area to ease spatial analysis. The ability of GIS to integrate different data sources into a common database has increased the need for areal interpolation (Cromley and Qiu 2013). Areal interpolation is the process of spatially disaggregating summary of counts of some phenomenon such as population over smaller areas (Goodchild and Lam 1980). In areal interpolation research, particularly those based on dasymetric modelling, ancillary data input that are related to the spatial distribution of the population in the source zone has been used to constrain the re-allocation (Langford 2013; Mennis 2009, 2003; Langford and Unwin 1994). One common approach for the interpolation of population data has been to redistribute summary population totals for a census area and compare the results with actual population counts of a lower census area for validation.

This study used land cover data derived from classified satellite imagery as the ancillary data input for the dasymetric models as has been used by many researchers (Langford and Unwin 1994; Eicher and Brewer 2001; Mennis 2003; Wu et al. 2005; Langford 2013). There are a number of concerns relating to the use of satellite imagery to derive land cover data to be used as the ancillary data input for dasymetric method. First, classifying satellite image requires specialized skills. Second, the selection of training sets, generation of spectral signatures and interpretation of results depends on analysts'

decision and knowledge of the study area. Third, identifying residential land use and whether they are truly occupied. The image classification techniques allow identification of pixels relating to different land cover making it easier to reclassify land use types into residential and non-residential areas. In reality, it is likely that some of the pixels assigned to residential areas are not actually occupied and those assigned to non-residential (such as farmlands) do have populations. There are other issues with high resolution image data in pixel based urban classification such as misclassification due to shadows, mixing of vegetation etc. The classification algorithm applied in this study, the maximum likelihood classifier, classifies land cover based on spectral signatures at per pixel level, while ignoring spatial features in an image. This is likely to increase misclassification due to shadows and uncertainty in the position of object borders in satellite images. For these reasons, this study found a lower resolution land cover data derived from classified Landsat7 (ETM) 30m spatial resolution easier to detect a residential class compared to using land cover data of 10m or 3m spatial resolution. However, some researchers have used other ancillary data sources to improve the accuracy of the interpolations such as road data (Mrozinski and Cromley 1999; Xie 1995), cadastral data (Maantay et al. 2007; Bentley et al. 2013), address points (Zandbergen and Ignizio 2010), household sample data (Leyk et al. 2013), open access data (Langford 2013) and three dimensional LiDAR data (Sridharan and Qiu 2013). The choice of ancillary data is important as the performance of the interpolation methods have been found to depend on specific characteristics of the original data such as its known errors, spatial properties and its extent (Zandbergen and Ignizio 2010; Wu et al. 2005) as well as the characteristics of any ancillary data used to constrain the disaggregation (Langford 2013).

The binary mask separating populated and unpopulated areas used in the dasymetric model may have affected the result because it considered all areas classified as built-up, as the populated areas. This is based on the assumption that people live only within built-up areas. This is a major source of error because it fails to differentiate residential built-up and non-residential built-up areas (e.g. industrial, commercial etc.). Using built-up areas as residential areas has effects on error. Some of these commercial/industrial areas can be differentiated from residential areas as they look spectrally distinct (especially in Leicester area), but most cannot be differentiated for the Port-Harcourt area. The residential areas for Leicester could have been extracted from the built-up

areas using the attributes of unit postcode data to select only areas that are directly associated with residential addresses but this cannot be done objectively in Port-Harcourt. For such places (e.g. Port-Harcourt) where there are no reliable data that clearly identify residential and non-residential areas, an analyst with knowledge of the study area may likely exclude non-residential areas from the dasymetric map leaving only the residential areas. This is subjective as it depends on individuals' knowledge and cannot be easily replicated. The built-up areas for both study areas were used as residential areas because the aim of this study is to develop a model for disaggregating population census data to small areas where validation data is available, and that are easily calibrated with land cover and census data for areas with no validation data.

The temporal difference between the dates on which the satellite imagery was acquired and the census year may have affected the result due to possible land cover changes between the periods. The Landsat imagery used for Leicester is more current than the census data by two years while the aerial photograph is more current by nine years than the census data used. For Port-Harcourt study, the Landsat imagery predates the census data used by three years and the Spot5 imagery is more current than the census data by one year. This is a likely source of error and could affect the result of this study. More so in Nigeria, perhaps where urban growth and land cover change is more dynamic.

The difficulty in identifying high-rise buildings may have affected the result especially in Leicester. The use of building volume to constrain the redistribution allow the modelling of both horizontal and vertical population distribution and is more likely to offer the most reasonable representation of where people actually live. This is more of an issue in Leicester area compared to Port-Harcourt area because there are relatively few high-rise buildings in Leicester around city centre location compared to low-rise buildings in Port- Harcourt, which reduced the potential estimation error. Sridharan and Qiu (2013) used LiDAR-derived building volumes as ancillary data input to estimate the population of Round Rock, Texas in the US and found the technique provided more detailed and realistic estimates than using area-based ancillary input data. The conversion of classified raster into vector may have affected the result as it is likely to introduce some distortions due to pixilation effect. This has however been reduced when a 10m and 3m satellite imagery were used to derive land cover data that was used as the ancillary data input for the dasymetric method.

The adoption of parameters developed for one area, to another area may have affected the result. It can be argued that the best performing parameters for an area may not likely provide the best fitting target zone estimates for another different area, as has been found by Zandbergen and Ignizio (2010) who compared dasymetric mapping techniques in four US States, and no single technique performed best in more than one state. The spatially distributed population surfaces generated in this study by the dasymetric method showed estimates were allocated to only areas identified as populated and the surfaces depict the underlying population distribution in the source zone. This is because the method used ancillary input data that are related to the spatial distribution of the population in the source zone to constrain the redistribution. The sensitivity of the population estimates to error in the classified imagery was examined by Fisher and Langford (1996) with the assumption that errors are spatially random and the results show the dasymetric method to be robust to classification error. A robust technique is more likely to “*perform well for different study areas across a range of different conditions*” (Zandbergen and Ignizio 2010, p.212). Despite this assumption, the surfaces generated have a potentially useful degree of precision based on the results of the visual inspection of some random points using google earth reference.

## **7.6 Limitations and suggestions for future work**

There are a number of limitations in the methods adopted in this research. A major limitation is the lack of validation for the population estimates generated for Port-Harcourt and a need to make inferences from places where it is known. One approach to validate the findings is to compare the results of a section of the study area with estimates obtained using manually digitised data for the selected area. The concept of dasymetric mapping was useful in Port-Harcourt since about one-third of the area within its boundary is covered by water, but caution must be exercised in generalising the result since the parameters adapted were only based on the analysis conducted in Leicester. While census data at small area level (for example OAs) are available in Leicester for validation, the parameters developed from Leicester study alone are not sufficient to conclude they are the best fitting parameters to apply for Port-Harcourt. There is the need to apply these techniques to other locations where the results can be validated to be able to make a firm conclusion on the parameters to be adapted for Port-Harcourt study. Another limitation is that the dasymetric method assumes uniform

density within the populated areas in each source zone, but in reality, population density may vary greatly within a single land cover class. Also, boundary alignment issues exist as is the case for most settings that overlay spatial data from different sources. The polygon boundary is noticeably out of alignment with the extent of the satellite imagery used but the effect of such discrepancy was corrected as the polygon boundary was used to subset the imagery, and to generate grids of cells that were used as the target zones.

Population data are likely to be uncertain, so also the intersection of census boundary with land covers. The dasymetric modelling approach applied in this research does not account for uncertainties in the population data, the dasymetric outputs and the relationship between ancillary and target variables. The difficulty in identifying residential, commercial or industrial areas from the built-up areas is another limitation of this research. One way of resolving this is to digitize and update the building attribute information which among others will describe the building usage. Also, it would be useful to obtain LiDAR data for the area to differentiate the heights of the buildings and then determine the volume of individual residential buildings. Another major concern is that residential housing units are not located at regular intervals, and a number of these residential housing units are not close to road network. The government will be encouraged to put in place appropriate policy and strategies such that new housing units will be located on the road network to ease evaluation of access distance. Some broad cultural assumptions were made with regards to population in adapting parameters developed for Leicester area in Port-Harcourt which may have affected the result. These include assumptions of similar levels of overcrowding, lone occupancy, different social mixes, feeling of place and neighbourhood, bus routes, etc. For the location-allocation analyses, it only considered PHCCs within Port Harcourt and it is likely that to some residents the nearest health centre is outside the boundary of Port Harcourt. Traditionally, most studies on location-allocation assume that the demand for services is static (e.g. Hakimi 1964; Church and ReVelle, 1974). This study assumes demand only comprises residential-based night time population as obtained from census estimates and not proximity to work places which may have affected the results of the analyses. This is because daily activities such as travel to work that characterise peoples' behaviour were not considered when modelling locations for services.

Distance plays a vital role in evaluating spatial access to health services and analysis of distance from patients' homes to the nearest health centre is an objective indicator of geographic accessibility to health services (Seidel et al. 2006). In such situations, the most appropriate method is to measure actual travel distance (or travel time) on a road network (Kohli et al. 1995). Distance metrics such as Euclidean, Manhattan, Road Network distances etc. are used in spatial analytical modelling to estimate distances from patient residence to hospital with each metric providing a single model of travel over a given network. An important aspect is the fundamental role of time in accessing health care services. The travel time measurements are more relevant and the most accurate estimates of accessibility (Schuurman et al. 2006). However, actual travel time on a road network depends on spatial and temporal conditions such as weekdays vs. weekends, traffic congestion, lane closures, rush vs. non-rush hours etc. Ignoring time-dependent variations in travel may result in sub-optimal location choices (Schmid and Doerner 2010) or underestimation of actual travel times (Eglese et al. 2006). This study assumes all roads are just as easily travelled as each other with no traffic concentrations because access distance was evaluated based on the assumption that people travel to health facilities on the road network in Port Harcourt. In the future, footpath data would be obtained and incorporated into the distance matrix to evaluate access. This is because in reality some residents travel on foot from home to PHCCs, which most times travelling on the footpath is shorter than the road network data used. The study also considered a decrease in average travel distance, although travel time may be a better choice, is more likely to improve overall accessibility to PHCCs.

Despite these limitations, this research is one of the first to apply dasymetric mapping techniques to redistribute census data over small areas in any part of Nigeria. One of the strengths of this research is that the datasets used are freely available and the technique can easily be replicated. The generation of demand population values using these methods in Port-Harcourt provides evidence to support spatial planning and location-allocation decisions. Specifically, the use of these demand population values to obtain optimal locations of health centres has the potential to reduce travel distance to PHCCs, and more children are likely to attend routine childhood immunisation that is more likely to reduce outbreak of diseases and thereby, reduce the death of children below the age of five. This is vital because the Millennium Development Goals (MDGs) declaration (target-four) as described by UN (2000) is committed to reducing “*child*

*mortality rate among children under 5 by two-third by the year 2015*". It is therefore important for government officials, politicians and local planners charged with decision-making on the numbers and locations of health facilities to be aware of methods applied in this research that provides evidence to support decisions.

There are a number of areas for possible future work based on findings and limitations of this research. Future work will consider conducting small area survey to collect data from individual households or another data source at small area level, for use in validation. The built-up areas in Port-Harcourt will be digitized and the attribute information will be updated to clearly identify residential built-up areas instead of using built-up areas. It is expected that more detailed census data for Nigeria would be available (at least at Ward level) as the office of the Surveyor General of Nigeria and the National Boundary Commission are jointly working on demarcating ward boundaries to allow creation of digital boundaries for wards. This will make it easier for future population census counts to be published as aggregate counts and statistics at wards level. Despite the current limitations of operating unmanned aerial vehicles (UAVs) in Nigeria, in the future, high resolution data and sophisticated object-based image classification using UAVs would be considered. Proxy datasets such as mobile phone records would be obtained from Nigerian Communications Commission (NCC) and used to represent the spatial distribution of the population. There are a number of concerns with these proxy datasets such as the different time of the day the data was recorded, the data may not represent all population as the children and elderly are not likely to use mobile phones, duplication of records as some residents may have more than one mobile phone etc.

The suitability of parameters developed from the areal interpolation for Leicester to generate surfaces for Port-Harcourt was partly due to the similarities between the two cities, an analysis using two or more study areas is more likely to provide different insights. The performance of interpolation methods was tested using known values of OAs, LSOAs and MSOAs. In the future, known values of UPC for Leicester could be used to test the performance of the methods and compare the results with the census units reported. This study used U.K. 2001 census total for Leicester, which has increased by about 17% as at 2011 census. In the future, U.K. 2011 population totals would be used to compare with the results obtained using 2001 population totals for

Leicester. The data for all the 160 newly constructed PHCCs in the 21 LGAs of Rivers State would be obtained and incorporated into the analyses to evaluate the boundary effect on PHCCs within Port-Harcourt. This will allow the application of genetic algorithm (GA) to solve the p-median problem because the number of facilities to locate has increased. The results of using GA can be compared with that of Teitz and Bart's algorithm. The locational analyses would be extended to include other public services (such as schools, petrol service stations, banks etc.) to evaluate accessibility.

## Chapter 8

### 8. CONCLUSIONS

#### 8.1 Introduction

This chapter draws together the findings from the literature, interpolations and the case study to explore the practical and theoretical implications of the research. The key aims of the research were extensively discussed in the previous chapters. The first aim was the determination of population surfaces in areas of unknown distributions to support spatial planning. The second was the evaluation of current health facility locations using the demand population values generated from the interpolations. The third was the suggestion of alternative spatial arrangements of health facilities to improve spatial accessibility. These three aims have been achieved through areal interpolation of population counts, allocation of catchment areas for the health centres and generation of potential new locations to evaluate accessibility from demand to potential new locations.

The next section summarises the key findings of this study relating to the aims of the research. Section 8.3 highlights the contributions of the research. The last section recommends policy options for health facility location planning and allocation of future resources.

#### 8.2 Research findings

The key finding in this research is for locations where detailed local mapping of census do not exist, that is, where the actual population in small areas is unknown, summary population totals can be redistributed over small areas by adapting a model that was developed for areas where validation is possible with land cover and census data for that area as a best bet solution given the circumstances. Other findings from this research are presented based on the five research objectives that addressed the aims of the research outlined in section 8.1.

**8.2.1 Research objective one:** The application of dasymetric and pycnophylactic interpolations across different spatial scales to redistribute aggregate population census data for Leicester over small areas.

The aim of this objective was to draw from a review of previous techniques of areal interpolations to select methods that best redistributes aggregate population counts to target zones with unknown distributions and evaluate the methods across different spatial scales. From the results presented in Chapter Four, this research concludes that the dasymetric method provided the best fitting target zone estimates. The research has shown that land cover information derived from classified Landsat7 (ETM) 30m spatial resolution when used as the ancillary data input for the dasymetric method provided the best estimates in the target zones compared to when 10m or 3m spatial resolution land cover data derived from classified resampled aerial photo data was used. This research also concludes that redistribution over 30m square grids is more likely to provide better estimates in the target zones compared to when 100m square grids are used.

**8.2.2 Research objective two:** A comparison of the estimated populations from the interpolations for three different census units and the known census counts in each case, to test the performance of the interpolation methods.

The aim of this objective was to test interpolation performance and to compare alternative areal interpolation algorithms. This objective also aims to draw from the population values compared to select the best solution found during this phase so as to adapt the model with the most appropriate interpolation method, support grid and ancillary data input to estimate demand population values in Port-Harcourt, Nigeria. This research concludes that the dasymetric model using land cover data derived from classified Landsat7 (ETM) 30m spatial resolution as the ancillary data input with redistribution over 30m square grids was the best solution found at this resolution of interpolation.

**8.2.3 Research objective three:** Adapt the model with the most appropriate parameters obtained from Leicester to estimate demand population values in Port-Harcourt, Nigeria.

The aim of this objective was to transfer the ‘best’ solution found for Leicester study based on the most appropriate interpolation method, support grid and ancillary data input so as to create demand population values in Port-Harcourt, Nigeria and use these surfaces to evaluate current health facility locations. The spatial distribution of population estimates reflects the underlying distribution of population in Port-Harcourt because ancillary data input was used to constrain the re-allocation of population counts to only areas identified as populated. The visual inspections of surfaces revealed about eighty per cent of the demand population values created correspond to areas identified as built-up in the Google Earth reference.

**8.2.4 Research objective four:** Evaluate the public health facility locations currently in place in Port-Harcourt.

The aim of this objective was to use the spatially distributed demand population values created for Port-Harcourt as input into location-allocation models with road network data to analyse the accessibility of current locations of health centres. The demand population values were used to allocate catchment areas for the health centres representing the expected geographical coverage for each health centre in order to analyse travel distances from each demand to each health centre. The results presented in Chapter Six show that a person would have to travel an average distance of 1204m to access the closest health centre. This distance is within the international criteria of 2km maximum distance a patient is expected to travel to an urban health centre that provides a comprehensive primary care as outlined by the Centre for Health Policy (1993) and discussed in Rispel et al. (1995) and Doherty et al. (1996). However, identifying optimally sited service provision points is more likely to reduce the average travel distance from each demand to each health centre, thereby improving overall accessibility to health centres.

**8.2.5 Research objective five:** Suggest alternative spatial arrangement of public health facilities using heuristic location-allocation modelling approaches.

The aim of this objective was to quantify the improvements in accessibility when the current health centres are optimally located. The results show the modelled distances from demand to the nearest potential new location minimised the average travelled distance from demand to the nearest health centre by 130m. This suggests alternative locations are optimal. This research concludes that some improvement in the overall spatial accessibility to health centres could be achieved when optimal locations are used.

**Key message: findings**

The key message to other researchers is that the dasymetric model with land cover data derived from a classified low resolution (e.g. 30m) satellite image as the ancillary data input with disaggregation over 30m square grids is more likely to be the solution for small area estimates in unknown distributions and without validation data. The model will provide better estimates in the target zones that will effectively depict the spatial heterogeneity of the population distribution in the source zone.

**8.3 Contributions**

This thesis provides an important contribution to knowledge, with respect to estimating population surfaces. Fine scale estimates of spatial population have relevance for a broad range of applications, and therefore the findings of this research are of value beyond the field of Geographical Information Science. This is demonstrated, to some extent, by the case study presented in Chapter 6.

The challenge of applying areal interpolation techniques to a region where population data are less readily available is a key contribution of this research. The production of accurate population maps for cities such as Port-Harcourt has important implications for policy development in these areas.

#### **8.4 Policy recommendation**

This study highlights the use of areal interpolation techniques and location-allocation models as a policy tool in identifying underserved areas. More specifically, this research draws the attention of policy makers, service providers and researchers into the need to collect and update small area census data in places where detailed census data are not currently reported. Analyses of this nature provide documentary evidence to support health facility location planning and allocation of future resource. The allocation of demand to current PHCCs provides evidence for informed decision making in spatial planning and policy development. The results of this research provide strong evidence to support spatial planning as against decisions taken by local planners and government officials on political grounds, without any formal analysis and generation of alternatives.

## 9. BIBLIOGRAPHY

Abernathy, W. J. and Hershey, J. C., 1972. A Spatial Allocation Model for Regional Health Services Planning, *Operations Research*, 20 (3), pp.629-42.

Acharya, L. B. and Cleland, J., 2000. Maternal and child health services in rural Nepal: Does access or quality matter more? *Health Policy and Planning*, 15(2), pp.223–29.

Alexandris, G. and Giannikos, I., 2010. A new model for maximal coverage exploiting GIS capabilities, *European Journal of Operation Research*, 202, pp.328-38.

Amaral, S., Gavlak, A. A., Escada, M. I. S. and Monteiro, A. M. V., 2012. Using remote sensing and census tract data to improve representation of population spatial distribution: case studies in the Brazilian Amazon, *Journal of Population Environment*, pp.1-29.

Amnesty International, 2009. *Nigeria: Petroleum, Pollution and poverty in the Niger Delta*. London: Amnesty international publications.

Amnesty International, 2010. *Forced eviction in Port-Harcourt, Nigeria, "Just move them"*, London: Amnesty international publications.

Apparicio, P., Abdelmajid, M., Riva, M. and Shearmur, R., 2008. Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues, *International Journal of Health Geographics*, 7(7), pp.1-14.

Arevshatian, L., Clements, C., Lwanga, S., Misore, A. O., Ndumbe, P., Seward, J. F. and Taylor, P., 2007. An evaluation of infant immunization in Africa: is a transformation in progress? *Bulletin of the World Health Organisation*, 85, pp.449-57.

- Ayeni, M. A., Rushton, G. and McNulty, M. L., 1987. "Improving the geographical accessibility of health care in rural areas: A Nigerian case study", *Social Science and Medicine*, 25, pp.1083–94.
- Barnard, D. K. and Hu, W., 2005. The Population Health Approach: health GIS as a bridge from theory to practice, *International Journal of Health Geographics*, 4(23), pp.1-9.
- Bashiri, M. and Fotuhi, F., 2009. A Cost-Based Set-Covering Location-Allocation Problem with Unknown Covering Radius, *Proceedings of the 2009 IEEE IEEM*, pp.1-5.
- Benigeri, M., 2007. Geographical Information Systems (GIS) in the Health Field, An Opportunity to Bridge the Gap Between Researchers and Administrators, *Canadian Journal of Public Health*, S75, pp.1-3.
- Bennett, W. D., 1981. "A location-allocation approach to health care facility location: A study of the undoctored population in Lansing, Michigan, *Social Science and Medicine*, 15D, pp.305–12.
- Bentley, G. C., Cromley, R. G. and Atkinson-Palombo, C., 2013. "The Network Interpolation of Population for Flow Modeling Using Dasymetric Mapping," *Geographical Analysis* 45, pp.307–23.
- Berman, O. and Krass, D., 2002. The generalized maximal covering location problem, *Computers and Operations Research*, 29, pp.563-81.
- Berman, O., Drezner, Z. and Wesolowsky, G. O., 2003. The expropriation location problem, *Journal of Operational Research Society*, 54, pp.769-76.
- Berman, O. and Huang, R., 2008. The minimum weighted covering location problem with distance constraints, *Computer and Operations Research* 35, pp.356-72.

Black, M., Ebener, S., Aguilar, P. N., Vidaurre, M. and El Morjani, Z., 2004. Using GIS to Measure Physical Accessibility to Health Care, *International HealthCare Users Conference*, 2004: Washington DC.

Blaschke, T., 2010. Object based image analysis for remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), pp.2-16.

Bowerman, R. L., Calamai, P. H. and Hall, G. B., 1999. The demand partitioning method for reducing aggregation errors in p-median problems, *Computer and Operations Research*, pp.1097-1111.

Brabyn, L. and Skelly, C., 2002. Modelling population access to New Zealand public hospitals, *International Journal of Health Geographics*, 1, pp.1-9.

Bracken, I. and Martin, D., 1989. The generation of spatial population distributions from census centroid data, *Environment and Planning A*, 21, pp.537-43.

Brandeau, M. L. and Larson, R. C., 1986. Extending and applying the hypercube queuing model to deploy ambulances in Boston, in *Management Science and the Delivery of Urban Service*, Ignall, E. and Swersey, A. J., Eds., TIMS Studies in the Management Sciences Series, North-Holland/Elsevier.

Brewer, N., Pearce, N., Day, P. and Borman, B., 2012. Travel time and distance to health care only partially account for the ethnic inequalities in cervical cancer stage at diagnosis and mortality in New Zealand, *Australian and New Zealand Journal of Public Health*, 36(4), pp.335-42.

Briggs, D. J., Gulliver, J., Fecht, D. and Vienneau, D. M., 2007. Dasymetric modelling of small-area population distribution using land cover and light emissions data, *Remote Sensing of Environment*, 108, pp.451-66.

Brinegar, S. J. and Popick, S. J., 2010. A comparative analysis of small area population estimation methods, *Cartography and Geographic Information Science*, 37(4), pp.273-84.

Brunsdon, C. F., Fotheringham, A. S. and Charlton, M., 1996. Geographically weighted regression - a method for exploring spatial non-stationarity, *Geographical Analysis*, 28, pp.281-298.

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

Cai, Q., Rushton, G., Bhaduri, B., Bright, E. and Coleman, P., 2006. Estimating Small-Area Populations by Age and Sex Using Spatial Interpolation and Statistical Inference Methods, *Transactions in GIS*, 10(4), pp.577-598.

Campbell, J. F., 1994. Integer programming formulations of discrete hub location problems, *European Journal of Operational Research*, 72, pp.387-405.

Carr, B. G. and Addyson, D. K., 2010. Geographic Information Systems and Emergency Care Planning, *Academic Emergency Medicine*, 17(12), pp.1274-78.

Centre for Health Policy, 1993. The determination of need norms for health services, Part 2. A summary of norms in other countries, Document submitted to the Department of National Health and Population Development.

Charlton, M. and Fotheringham, A. S., 2009. Geographically Weighted Regression White Paper 2009. *National Centre for Geocomputation*, National University of Ireland Maynooth, Maynooth, Co Kildare, Ireland, pp.1-17.

Chen, R. and Handler, G.Y., 1993. The Conditional p-Center Problem in the Plane, *Naval Research, Log*, 40, pp.117-28.

Chen, K., 2002. An approach to linking remotely sensed data and areal census data. *International Journal of Remote Sensing*, 23(1), pp.37-48.

Church, R. L. and Gerrard, R. A., 2003. The Multi-level Location Set Covering Model, *Geographical Analysis*, Vol.35 (4), pp.277-289.

Church, R. L. and ReVelle, C. S., 1974. The maximal covering location problem papers, *Regional Science Association* 32, pp.101-108.

Church, R. L. and Sorensen, P., 1994. Integrating Normative Location Models into GIS: Problems and Prospects with the p-median Model, *National Centre for Geographic Information and Analysis*, Technical Report 94(5), pp.1-16.

Cockings, S., Fisher, P. and Langford, M., 1997. Parameterization and visualization of the errors in areal interpolation, *Geographical Analysis*, 29 (4), pp.314-328.

Comber, A. J., Brunsdon, C. and Green, E., 2008a. Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups, *Landscape and Urban Planning*, 86, pp.103-14.

Comber, A., Proctor, C. and Anthony, S., 2008b. The Creation of a National Agricultural Land Use Dataset: Combining Pycnophylactic Interpolation with Dasymetric mapping technique, *Transactions in GIS*, 12(6), pp.775-91.

Comber, A. J., Brunsdon, C., Hardy, J. and Radburn, R., 2009. Using a GIS-Based Network Analysis and Optimisation Routines to Evaluate Service Provision: A Case Study of the UK Post Office, *Journal of Applied Spatial Analysis*, 2 pp.47-64.

Comber, A. J., Sasaki, S., Suzuki, H. and Brunsdon, C., 2010. A modified grouping genetic algorithm to select ambulance site locations, *International Journal of Geographical Information Science*, *iFirst*, 2010, pp. 1-17.

Comber, A., Fisher, P.F., Brunsdon, C. and Khmag, A., 2012. Spatial analysis of remote sensing image classification accuracy, *Remote Sensing of Environment*, 127, pp.237-46.

Comber, A. J., 2013. Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies, *Remote Sensing Letters*, 4(4), pp.373-80.

Comber, A., See, L., Fritz, S., Van Der Velde, M., Perger, C. and Foody, G. M., 2013. Using control data to determine the reliability of volunteered geographic information

about land cover, *International Journal of Applied Earth Observation and Geoinformation*, 23, pp.37-48.

Cooper, L., 1963. Location-allocation problems, *Operations Research*, 11, pp.331-43  
Cromley E. and McLafferty S. 2002, *GIS and Public Health*, New York: Guilford Press, pp.35-37.

Cromley, R. G., Ebenstein, A. Y. and Hanink, D. M., 2009. Estimating Components of Population Change from Census Data for Incongruent Spatial/Temporal Units and Attributes, *Spatial Science*, 54(2), pp.89-100.

Cromley, R. G., Lin, J. and Merwin, D. A., 2012. Evaluating representation and scale error in the maximal covering location problem using GIS and intelligent areal interpolation, *International Journal of Geographical Information Science*, 26:3, pp 495-517.

Cromley, R. G., Hanink, D. M. and Bentley, G. C., 2011. A Quantile Regression Approach to Areal Interpolation, *Annals of the Association of American Geographers*, 102(4), pp.763-77.

Culpepper, W. J., Cowper Ripley, D., Litt, E. R., McDowell, T. and Hoffman, P. M., 2010. Using geographic information system tools to improve access to MS specialty care in Veterans Health Administration, *Journal of Rehabilitation Research and Development*, 47(6), pp.583-92.

Curtis, S. E. and Taket, A. R., 1989. The development of Geographical Information Systems for Locality Planning in Health Care, *Royal Geographical Society with Institute of British Geographers*, Area 21(4), pp.391-99.

Daskin, M. S., 1983. A maximal expected covering location model: formulation, properties and heuristic solution, *Transportation Science*, 17, pp.48–70.

Daskin, M. S., 1995. Network and discrete location: models, algorithms, and applications. New York: John Wiley and Sons.

D'Earth, M., Sixsmith, J., Cannon, R. and Kelly, L., 2005. The experience of people with disabilities in accessing health services in Ireland: Do inequalities exist? *Report to the National Disability Authority*, pp. 1-127.

Deichmann, U. and Eklundh, L., 1991. Global digital datasets for land degradation studies: A GIS approach, United Nations Environment Programme, *Global Resource Information Database*, Case Study No. 4, Nairobi, Kenya.

Deichmann, U., 1996. A Review of Spatial Population Database Design and Modelling, *Technical Report 96-3 FOR National Center for Geographic Information and Analysis*, pp.1-62.

Delamater, P. L., Messina, J. P., Shortridge, A. M. and Grady, S. C., 2012. Measuring geographic access to health care; raster and network-based methods, *International Journal of Health Geographics*, 11(15), pp.1-18.

Dempster, A. P., Laird, N. M., and Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, B 39, pp. 1–38.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C. and Worley, B.A., 2000. LandScan: a global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, 66(7), pp. 849–57.

Doherty, J., Rispel, L. and Webb, N., 1996. Developing a plan for primary health care facilities in Soweto, South Africa, Part II: Applying locational criteria, *Health Policy and Planning*, 11(4), pp.394-405.

Dray, S., Pettorelli, N. and Chessel, D., 2003. Multivariate analysis of incomplete mapped data, *Transactions in GIS*, 7(3), pp.411-22.

Dulin, M. F., Ludden, T. M., Tapp, H., Blackwell, J., Smith, H. A. and Furuseth, O. J., 2010. Using Geographic Information Systems (GIS) to Understand a Community's Primary Care Needs, *Journal of American Board of Family Medicine*, 23, pp.13-21.

Eastman, R. J., 2006. *IDRISI Andes: Guide to GIS and image processing*. Worcester, MA: Clark Labs, Clark University.

Eaton, D. J., Hector, M. L., Sanchez, U., Ricardo, R. L. and Morgan, J., 1986. Determining ambulance deployment in Santa Domingo, Dominican Republic, *Journal of Operational Research Society*, 37, pp.113-26.

Eglese, R., Maden, W. and Slater, A., 2006. A road timetable to aid vehicle routing and scheduling, *Computers and Operations Research*, 33(12), pp.3508-19.

Eicher, C. and Brewer, C., 2001. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, 28, pp.125-38.

Erlenkotter, D., 1981. A comparative study of approaches to dynamic location problems, *European Journal of Operational Research*, 6(2), pp.133-43.

Energy Information Administration, 2014. *Country Analysis Brief: Nigeria Energy Data, Statistics and Analysis – Oil, Gas, Electricity, and Coal*. U.S. EIA 2014.

Fisher, P. F. and Langford, M. 1995. Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation, *Environment & Planning A*, 27, pp.211-24.

Fisher, P. F. and Langford, M. 1996. Modelling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping, *The Professional Geographer* 48(3), pp.299-309.

Flowerdew, R. and Lovett, A., 1988. Fitting Constrained Poisson Regression Models to Inter-urban Migration Flows, *Geographical Analysis*, 20 (4), pp.297-307.

Flowerdew, R. and Green, M. 1989. Statistical methods for inference between incompatible zonal systems In: Goodchild, M., and Gopal, S. (eds) *Handling geographical information: Methodology and potential applications*. New York, NY: Longman, pp. 239–247.

Flowerdew, R., Green, M. and Kehris, E., 1991. Using areal interpolation methods in GIS, *Papers in Regional Science: The Journal of the RSAI* 70 (3), pp. 303–315.

Flowerdew, R. and Green, M., 1991. Data integration: Statistical methods for transferring data between zonal systems. In *Handling geographical information*. Edited by I. Masser and M. B. Blakemore, Longman Scientific and Technical, pp.38-53.

Flowerdew, R. and Green, M., 1992. Developments in Areal Interpolation Methods and GIS, *Annals of Regional Science*, 26, pp.67-78.

Fotheringham, A. S., Curtis, A. and Densham, P. J., 1993. The zone definition problem in location-allocation modelling. Unpublished manuscript. NCGIA, State University New York, Buffalo, NY.

Frankenberg, E., 1995. The effects of access to health care on infant mortality in Indonesia, *Health Transition Review*, 5(2), pp.143–63.

Gao, L. L. and Robinson, E. P., 1994. Uncapacitated facility location: general solution procedure and computational experience, *European Journal of Operational Research*, 76, pp.410-27.

Gao, J., Tang, S., Tolhurst, R. and Rao, K., 2001. Changing access to health services in urban China: implications for equity, *Health Policy and Planning*, 16(3), pp.302-12.

Gao, J. and Liu, Y., 2010. Determination of land degradation causes in Tongyu County, Northeast China via land cover change detection. *International Journal of Applied Earth Observation and Geo-information*, 12(1), pp.9-16.

Geanuracos, C. G., Cunningham, S. D., Weiss, G., Forte, D., Henry Reid, L. M. and Ellen, J. M., 2007. Use of Geographic Information Systems for Planning HIV Prevention Interventions for High-Risk Youth, *American Journal of Public Health*, 97(11), pp.1974-81.

Gibson, J., Deng, X., Boe-Gibson, G., Rozelle, S. and Huang, J., 2011. Which households are most distant from health centres in rural China? Evidence from a GIS network analysis, *GeoJournal*, 76, pp.245-55.

Glover, F., 1986. Future Paths for Integer Programming and Links to Artificial Intelligence. *Computers and Operations Research*, 13, pp. 533-49.

Glover, F., 1990. Tabu search: a tutorial. *Interfaces*, 20, pp.74-94.

Goodchild, M. F. and Lam, N., 1980. Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing*, 1, pp.297–312.

Goodchild, M. F., Anselin, L. and Deichmann U., 1993. A framework for the areal interpolation of socioeconomic data, *Environment and Planning A* 25, pp.383–97.

Green, N. E., 1956. Aerial photographic analysis of residential neighbourhoods': An evaluation of data accuracy. *Social Forces*, 35(2), pp.142-147.

Griffith, D. A., 2013. "Estimating Missing Data Values for Georeferenced Poisson Counts," *Geographical Analysis* 45, pp.259–84.

Gu, W., Wang, X. and McGregor, S. E., 2010. Optimization of preventive health care facility locations, *International Journal of Health Geographics*, 9(17), pp.1-16.

Guagliardo, M. F., 2004. Spatial accessibility of primary care: concepts, methods and challenges, *International Journal of Health Geographics*, 3(3), pp.1-13.

Hakimi, S. L., 1964. Optimal locations of switching centres and absolute centres and medians of a graph, *Operations Research*, 12, pp.450-59.

Harris, R. J. and Longley, P. A., 2000. "New Data and Approaches for Urban Analysis: Modelling Residential Densities." *Transactions in GIS* 4, pp.217-34.

Harris, R. J. and Longley, P. A., 2002. Creating small area measures of urban deprivation, *Environment and Planning A*, 34(6), pp.1073–1093.

- Harvey, J. T., 2002. "Population Estimation Models Based on Individual TM Pixels", *Photogrammetric Engineering and Remote Sensing*, 68(11), pp.1181-1192.
- Hawley, K. and Moellering, H., 2005. A comparative analysis of areal interpolation methods, *Cartography and Geographic Information Science*, 32(4), pp.411-23.
- Hay, S. I., Noor, A. M., Nelson, A. and Tatem, A. J., 2005. The accuracy of human population maps for public health application, *Tropical Medicine and International Health*, 10, pp.1073–1086.
- Hemat, S., Takano, T., Kizuki, M. and Mashal, T., 2009. Health-care provision factors associated with child immunization coverage in a city centre and a rural area in Kabul, Afghanistan. *Vaccine*, 27, pp.2823-9.
- Herrmann, c. and Maroko, A. R., 2006. Crime pattern analysis: Exploring Bronx aut thefts using GIS. In: Maantay, J. A. and Ziegler, J., *GIS for the urban environment*. Redlands, California: ESRI Press.
- Hewko, J., Smoyer-Tomic, K. E. and Hodgson, M. J., 2002. Measuring neighbourhood spatial accessibility to urban amenities: does aggregation error matter? *Environment and Planning A*, 34, pp.1185-1206.
- Higgs, G., 2005. A Literature Review of the Use of GIS-Based Measures of Access to Health Care Services, *Health Services and Outcomes Research Methodology*, 5, pp.119-139.
- Higgs, G. and Gould, M., 2001. Is there a role for GIS in the 'New NHS'? *Health and Place*, 7(3), pp.247-59.
- Hodgson, M. J., 1990. A Flow-Capturing Location-Allocation Model, *Geographical Analysis*, 22(3), pp.270-79.
- Holland, J. H., 1975. *Adaptation in natural and artificial systems*. Michigan: University of Michigan Press.

Holt, J., Lo, C. P. and Hodler, T. W., 2004. Dasymetric estimation of population density and areal interpolation of census data, *Cartography and Geographic Information Science* 31, pp.103–121.

Hosage, C. M. and Goodchild, M. F., 1986. Discrete space location-allocation solutions from genetic algorithms, *Annals of Operations Research*, 6, pp.35-46

Independent National Electoral Commission, 2011, Information for Voters [online] Available at: <http://www.inecnigeria.org/information-for-voters/> [Accessed 17 June 2011].

Jamil, K., Bhuiya, A., Streatfield, K. and Chakrabarty, N., 1999. The immunization programme in Bangladesh: Impressive gains in coverage, but gaps remain, *Health Policy and Planning*, 14(1), pp.49–58.

Jansen, L. J. M., Bagnoli, M. and Focacci, M., 2008. Analysis of land-cover/use change dynamic in Manica Province in Mozambique in a period of transition (1990-2004), *Forest Ecology and Management*, 254, pp.308-26.

Jia, H., Ordonez, F. and Dessouky, M., 2007. A modelling framework for facility location of medical services for large-scale emergencies, *IIE Transactions* 39, pp.41-45.

Jones, S. G., Ashby, A. J., Momin, S. R. and Naidoo, A., 2010. Spatial Implications Associated with Using Euclidean Distance Measurements and Geographic Centroid Imputation in Health Care Research, *Health Services Research*, 45(1), pp.316-27.

Jordan, H., Roderick, P., Martin, D. and Barnett, S., 2004. Distance, rurality and the need for care: access to health services in South West England, *International Journal of Health Geographics*, 3(21), pp.1-9.

Khumawala, B. M., 1973. An efficient algorithm for  $p$ -median problem with maximum distance constraints, *Geographical Analysis*, 5, pp.309-21.

Kim, H. and Yao, X., 2010. Pycnophylactic interpolation revisited: integration with the dasymetric-mapping method, *International Journal of Remote Sensing*, 31(21), pp.5657-71.

Klose, A. and Drexler, A., 2004. Facility location models for distribution system design, *European Journal of Operational Research*, 162, pp.4–29.

Kohli, S., Sahlen, K., Sivertun, A., Lofman, E. T. and Wigertz, O., 1995. Distance from the Primary Health Centre: a GIS method to study geographical access to health care, *Journal of Medical Systems*, 19(6), pp.425-36.

Kristensen, I., Aaby, P. and Jensen, H., 2000. Routine vaccinations and child survival: follow up study in Guinea-Bissau, West Africa. *British Medical Journal*, 321, pp.1-8.

Kruger, D. J., Brady, J. S. and Shirey, L. A., 2008. Using GIS to Facilitate Community-Based Public Health Planning of Diabetes Intervention Efforts, *Health Promotion Practice*, 9(76), pp.76-81.

Kumar, N., 2000. "Locational Analysis of Public and Private Health Services in Rohtak and Bhiwani Districts, India, 1981-1996", *Proceedings of Map-2000*.

Kyriakidis, P. C., 2004. A geostatistical framework for area-to-point spatial interpolation, *Geographical Analysis*, 36, pp.259-89.

Lam, N. S., 1983. Spatial Interpolation Methods: A Review, *The American Cartographer*, 10 (2), pp.129-49.

Langford, M., Maguire, D. J. and Unwin, D.J., 1991. The area transform problem: Estimating population using remote sensing in a GIS framework. In: Masser, I. and Blakemore, M. (eds) *Handling Geographical Information: Methodology and Potential Applications*. London: Longman, pp.55–77.

Langford, M., Fisher, P. F. and Troughear, D., 1993. Comparative accuracy measurements of the crossareal interpolation of population, *In Proceedings of European*

*conference on geographical information systems (EGIS93)*, pp.663–674 Utrecht: EGIS Foundation.

Langford, M. and Unwin, D., 1994. Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal* 31, pp.21–26.

Langford, M., 2003. “Refining Methods for Dasymetric Mapping Using Satellite Remote Sensing,” In *Remotely Sensed Cities*, ed. V. Mesev. London: *Taylor and Francis*, pp.137–56.

Langford, M., 2006. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method, *Computers, Environment and Urban Systems*, 30, pp.161–180.

Langford, M. and Higgs, G., 2006. Measuring Potential Access to Primary Healthcare Services: The Influence of Alternative Spatial Representations of Population, *The Professional Geographer*, 58(3), pp.294-306.

Langford, M., 2007. Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems*, 31, pp.19-32.

Langford M, Higgs G, Radcliffe J. and White S., 2008. Urban population distribution models and service accessibility estimation. *Computers, Environment and Urban Systems*, 32, pp.66–80.

Langford, M., 2013. “An Evaluation of Small Area Population Estimation Techniques Using Open Access Ancillary Data.” *Geographical Analysis*, 45, pp.324–44.

Lee, M. S. and McNally M. G., 2002. Measuring Physical Accessibility with Space-Time Prisms in a GIS: A Case Study of Access to Health Care Facilities: Irvine, *Institute of Transport studies*.

- Leyk, S., Buttenfield, B. P., Nagle, N. N. and Stum, A. K., 2013. Establishing relationships between parcel data and land cover for demographic small area estimation, *Cartography and Geographic Information Science*, 40(4), pp.305-15.
- Leyk, S., Nagle, N. N. and Buttenfield, B. P., 2013. Maximum Entropy Dasymmetric Modeling for Demographic Small Area Estimation, *Geographical Analysis* 45, pp.285–306.
- Lillesand, T. M. and Kiefer, R. W., 1987. *Remote sensing and image interpretation*, 2nd ed. New York: John Wiley and Sons.
- Lillesand T.M., Kiefer R. W. and Chipman J. W., 2008. *Remote sensing and Image Interpretation* 6th ed. John wiley & sons. Inc. USA.
- Lin, J., Cromley, R. and Zhang, C., 2011. Using Geographically Weighted Regression to solve the Areal Interpolation Problem, *Annals of GIS*, 17, pp.1-14.
- Lindsey, G., Maraj, M. and Kuan, S. C., 2001. Access, equity, and urban greenways: An exploratory investigation, *The Professional Geographer*, Volume 53, pp.332-47.
- Liu, X., Clarke, K. and Herold, M., 2006. Population density and image texture: a comparison study, *Photogrammetric Engineering and Remote Sensing* 72 (2), pp.187–96.
- Lo, C. P., 2006. “Estimating Population and Census Data,” in *Remote Sensing of Human Settlements, Manual of Remote Sensing*, third ed., Vol.5, Ridd, M. K., and Hipple, J. D. (Eds.), Bethesda, MD: American Society for Photogrammetry and Remote Sensing, pp.337-77.
- Lo, C. P., 2008. Population Estimation Using Geographically Weighted Regression, *GIScience and Remote Sensing*, 45 (2), pp.131-48.

- Lu, D., Hetrick, S. and Moran, E., 2010. Land cover classification in a complex urban-rural landscape with QuickBird imagery, *Photogrammetric Engineering and Remote Sensing*, 76(10), pp.1159-68.
- Lu, D., Li, G., Moran, E., Freitas, C. C., Dutra, L. and Anna, S. J. S., 2012. A Comparison of Maximum Likelihood Classifier and Object-Based Method Based on Multiple Sensor Datasets for Land-Use/Cover Classification in the Brazilian Amazon, *Proceedings of the 4<sup>th</sup> GEOBIA*, May 7-9, Rio de Janeiro, Brazil, pp. 20-24.
- Luo, W., 2004. Using a GIS-based floating catchment method to assess areas with shortage of physician, *Health and Place*, 10(1), pp.1-11.
- Maantay, J. A., Maroko, A. R. and Herrmann, C., 2007. Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System (CEDS), *Cartography and Geographic Information Science*, 34(2), pp.77-102.
- Maantay, J. A., Maroko, A. R. and Porter-Morgan, H., 2008. "Research Note—a New Method for Mapping Population and Understanding the Spatial Dynamics of Disease in Urban Areas: Asthma in the Bronx, New York." *Urban Geography* 29 (7), pp.724–38.
- Maranzana, F., 1964. On the location of supply points to minimize transport costs, *Operational Research Quaterly*, 15, pp.261-70.
- Marianov, V. and ReVelle, C. S., 1992. A probabilistic fire-protection siting model with joint vehicle reliability requirements, *Papers in Regional Science*, 71, pp.217-41.
- Marianov, V. and ReVelle, C. S., 1994. The queuing probabilistic location set covering problem and some extensions. *Socio-Economic Planning Sciences*, 28, pp.167-78.
- Markoff, J. and Shapiro, G., 1973. The Linkage of Data Describing Overlapping Geographical Units, *Historical Methods Newsletter*, 7, pp.34-46.
- Martin, D., 1989. Mapping population data from zone centroid locations, *Transactions of the Institute of British Geographers*, 14, pp.90–97.

- Martin, D., 1997. From enumeration districts to output areas: Experiments in the automated creation of a census output geography. *Population Trends*, 88, pp.36–42.
- Martin, D. 1998. Optimizing census geography: The separation of collection and output geographies, *International Journal of Geographical Information Science*, 12, pp.673–85.
- Martin, D., Tate, N. J. and Langford, M., 2000. Refining population surface models; experiments with Northern Ireland census data, *Transactions in Geographical Information Systems*, 4, pp.342–60.
- Martin, D., Wrigley, H., Barnett, S. and Roderick, P., 2002. Increasing the sophistication of access measurement in a rural healthcare study, *Health and Place*, 8, pp.3–13.
- Martin, D., 2010. Understanding the social geography of census undercount, *Environment and Planning*, 42, pp.2753 –70.
- Mathys, T. and Kamel Boulos, M. N., 2011. Geospatial resources for supporting data standards, guidance and best practice in health informatics, *British Medical Council Research Notes*, 4(19), pp.1-18.
- McGrail, M. R., 2012. Spatial accessibility of primary health care utilising the two step floating catchment area method: an assessment of recent improvements, *International Journal of Health Geographics*, 11(50), pp.1-12.
- McGrail, M. R. and Humphreys, J. S., 2009. Measuring spatial accessibility to primary care in rural areas: Improving the effectiveness of the two-step floating catchment area method, *Applied Geography*, pp.533-41.
- McGregor, J., Hanlon, N., Emmons, S., Voaklander, D. and Kelly, K., 2005. If all ambulances could fly: putting provincial standards of emergency care access to the test in Northern British Columbia, *Canadian Journal of Rural Medicine*, 10, pp.163–8.

McLafferty, S. L., 2003. GIS and Health Care, *Annual Reviews Public Health*, 24, pp.25-42.

McLafferty, S. L. and Grady, S., 2004. Prenatal Care Need and Access: A GIS analysis, *Journal of medical systems*, 28(3), pp.321-333.

Mehretu, A., Wittick, R. I. and Pigozzi, B. W., 1983. Spatial design for basic needs in eastern Upper Volta, *The Journal of Development Area*, 7, pp.383-94.

Mennis, J., 2003. Generating surface models of population using dasymetric mapping, *The Professional Geographer*, 55, pp.31-42.

Mennis, J., and Hultgren, T., 2006. Intelligent dasymetric mapping and its application to areal interpolation, *Cartography and Geographic Information Science* 33 (3), pp.179-94.

Mennis, J., 2009. Dasymetric Mapping for Estimating Population in Small Areas, *Geography Compass*, 3(2), pp.727-45.

Mitropoulos, P., Mitropoulos, I., Giannikos, I. and Sissouras, A., 2006. A biobjective model for the locational planning of hospitals and health centers, *Health Care Management Science*, 9(2), pp.171-79.

Moon, I. D. and Chaudhry, S., 1984. An analysis of network location problems with distance constraints, *Management Science*, 30, pp.290-307.

Moon, Z. K. and Farmer, F. L., 2001. Population density surface: A new approach to an old problem, *Society & Natural Resources*, 14, pp.39-49.

Moore, D. S., 2010. *The Basic Practice of Statistics*. 5<sup>th</sup> ed. New York: W. H. Freeman and Company.

Mrozinski, R. Jr. and Cromley, R., 1999. Singly and doubly-constrained methods of areal interpolation for vector-based GIS, *Transactions in GIS* 3, pp.285-301.

Mugglin, A. S. and Carlin, B. P., 1998. Hierarchical Modelling in Geographical Information Systems: Population Interpolation Over Incompatible Zones, *Journal of Agricultural, Biological, and Environmental Statistics*, 3(2), pp.111-30.

Mugglin, A. S., Carlin, B., Zhu, L. and Conlon, E., 1999. Bayesian areal interpolation, estimation, and smoothing: an inferential approach for geographic information systems. *Environment and Planning A*, 31, pp.1337–52.

Mulatu, D. W., Van Der Veen, A., Becht, R., Van Oel, P. R., and Bekalo, D. J., 2013. Accounting for Spatial Non-Stationarity to Estimate Population Distribution Using Land Use/Cover, Case Study: the Lake Naivasha Basin, Kenya, *Journal of Settlements and Spatial Planning*, 4(1), pp.33-44.

Murray, A. T., Church, R. L., Gerrard, R. A. and Tsui, W. S., 1998. Impact models for siting undesirable facilities, *Papers in Regional Science*, 77, pp.19-36.

Nagle, N. N., Battenfield, B. P., Leyk, S. and Spielman, S., 2014. Dasymetric Modelling and Uncertainty, *Annals of the Association of American Geographers*, 104(1), pp.80-95.

Narula, S. C., Ogbu, U. I. and Samuelsson, H. M., 1977. An algorithm for the p-median problem, *Operations Research*, 25, pp.709-13.

Narula, S. C., 1984. Hierarchical location-allocation problems: A classification scheme, *European Journal of Operation Research* 15, pp.93-99.

National Population Commission (NPC) [Nigeria] and ORC Macro. 2004. *Nigeria Demographic and Health Survey 2003*. Abuja, Nigeria: National Population Commission and ORC Macro.

National Population Commission (NPC), 2006a. Nigeria: National Population Census Priority tables Vol. 1.

National Population Commission (NPC) [Nigeria] and ICF Macro 2009. *Nigeria Demographic and Health Survey 2008*, Abuja, Nigeria: National Population Commission.

National Primary Health Care Development Agency (NPHCDA), 2007. Ward minimum health care package 2007-2012.

Nemet, G. F. and Bailey, A. J., 2000. Distance and health care utilization among the rural elderly, *Social Science and Medicine*, 50, pp.1197-1208.

Nigerian National Petroleum Corporation, 2011. History of the Nigerian Petroleum Industry.

Nna, N. J. and Pabon, B. G., 2012. Population, Environment and Security in Port-Harcourt, *Journal of Humanities and Social Science*, 2(1), pp.1-7.

NRC (National Research Council), 2007. *Tools and methods for estimating populations at risk from natural disasters and complex humanitarian crises*. Washington, DC: The National Academies Press.

Oda, T. and Yamamura, E., 1987. A location-allocation model for health care services planning, *Environmental Science*, Hokkaido, 10 (1), pp.37-51.

Okabe, A. and Sadahiro, Y., 1997. Variation in Count Data Transferred from a Set of Irregular Zones to a Set of Regular Zones through the Point-in-Polygon Method, *Geographical Information Science*, 11(1), pp.93-106.

Okabe, A., Okunuki, K. and Shoide, S., 2006. SANET: A Tool for spatial analysis on a network, *Geographical Analysis*, 38, pp.57-66.

Openshaw, S., 1977. Algorithm 3: A procedure to generate pseudo-random aggregations of N zones into M zones, where M is less than N, *Environment and Planning A*, 9, pp.1420-26.

Openshaw, S., 1984. The modifiable areal unit problem, *Concepts and Techniques in Modern Geography*, 28, pp.38-41.

Oppong, J. R. and Hodgson, M. J., 1994. Spatial accessibility to health facilities in Suhum district, Ghana, *Professional Geographer*, 46, pp.199-209.

O'Sullivan, D., Morrison, A. and Shearer, J., 2000. Using desktop GIS for the investigation of accessibility by public transport: an isochrones approach, *International Journal of Geographic Information Science*, 14 (1), pp.85-104.

Owen, K. K., Obregon, E. J. and Jacobsen, K. H., 2010. A geographical analysis of access to health services in rural Guatemala. *International Health*, 2, pp.143-49.

Parker, S., 2009. World Health Day Celebration, Address presented by Hon Commissioner for Health, Rivers State.

Parker, T. S., 2014. Achieving results in health, A publication of the ministry of health, Rivers State.

Patel, A. and Waters, N., 2012. Using Geographical Information Systems for Health Research, *Application of Geographical Information Systems*, 16, pp.303-20.

Paul, M., 1991. *Computer processing of remotely sensed images: An introduction*, p.352. Biddley Limited Publication, UK.

Payne, S., Jarrett, N. and Jeffs, D., 2000. The impact of travel on cancer patients' experiences of treatment: a literature review, *European Journal of Cancer Care*, 9, pp.197-203.

Petrov A. N., 2008. Setting the Record Straight: On the Russian Origins of Dasymmetric Mapping. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 43 (2), pp.133–36.

Pirkul, H. and Schilling, D. A., 1991. The maximal covering location problem with capacities on total workload, *Management Science*, 37, pp.233–48.

Plastria, F., 1997. Profit maximising single competitive facility location in the plane, *Studies of Locational Analysis*, 11, pp.115–26.

Policy Project (Nigeria), 2002. *Child Survival in Nigeria: Situation, Response and Prospects*.

Poulsen, E., and Kennedy, L. W., 2004. Using dasymetric mapping for spatially aggregated crime data, *Journal of Quantitative Criminology* 20 (3), pp.243–62.

Pozzi, F., Small, C. and Yetman, G., 2003. Modeling the Distribution Of Human Population with Night-Time Satellite Imagery and Gridded Population of the World, *American Society for Photogrammetry and Remote Sensing*, pp.1-9.

Qiu, F., Zhang, C. and Zhou, Y., 2012. The Development of an Areal Interpolation ArcGIS Extension and a Comparative Study. *GIScience and Remote Sensing*, 49(5), pp.644-63.

Qiu, F. and Cromley, R., 2013. Areal Interpolation and Dasymetric Modeling, *Geographical Analysis*, 45, pp.213-15.

Rahman, S. and Smith, D. K., 1991. "A Comparison of Two Heuristic Methods for the P-median Problem With and Without Maximum Distance Constraints", *International Journal of Operations and Production Management*, 11(6), pp.76-84.

Rahman, S. and Smith, D. K., 1996. Locating health facilities in rural Bangladesh, In: Rosenhead, J. and Tripathy, A. (eds). *Operational Research for Development*, New Age Publication: New Delhi, pp.184-96.

Rahman, S. and Smith, D. K., 1999. Deployment of rural health facilities in a developing country, *Journal of the Operational Research Society*, 50, pp.892-902.

Rahman, S. and Smith, D. K., 2000. Use of location allocation models in health services development planning in developing nations, *European Journal of Operation Research*, 123, pp.437-452.

- Rase, W. D., 2001. Volume-preserving Interpolation of a Smooth Surface from Polygon Related Data, *Journal of Geographical Systems*, 3, pp.199–203.
- Reibel, M., 2007. Geographic Information Systems and Spatial Data Processing in Demography: a review, *Population Research and Policy Review*, 26, pp.601-18.
- Reibel, M. and Bufalino, M. E., 2005. Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems, *Environment and Planning A*, 37, pp.127–139.
- Reibel, M. and Agrawal, A., 2007. Areal interpolation of population counts using preclassified land cover data, *Population Research Policy Review* 26, pp.619–633.
- ReVelle, C. S. and Eiselt, H. A., 2005. Location analysis: A synthesis and survey, *European Journal of Operational Research*, 165, pp.1-19.
- ReVelle, C. S. and Swain, R. W., 1970. Central facilities location, *Geographical Analysis*, 2, pp.30-42.
- Revised National Health Policy, 2004. Federal Republic of Nigeria, Federal Ministry of health Abuja.
- Ricketts, T. C., 2003. Geographic Information Systems and Public Health, *Annual review public health*, 24, pp.1-6.
- Rispel, L., Beattie, A., Xaba, M., Fonn, S., Cabral, J. and Marawa, N., 1995. A description and evaluation of primary health care services delivered by the Alexandra Health Centre and University Clinic. Johannesburg: *Centre for Health Policy*.
- Rosero-Bixby, L., 2004. Spatial access to health care in Costa Rica and its equity: a GIS-based study, *Social Science and Medicine*, 58, pp.1271-84.

- Rosing, K., ReVelle, C. and Rosing-Vogelaar, H., 1979. The p-median and its linear programming relaxation: an approach to large problems, *Journal of the Operational Research Society*, 30, pp.815-23.
- Ruther, M., Maclaurin, G., Leyk, S., Battenfield, B. and Nagle, N., 2013. Validation of spatially allocated small area estimates for 1880 census demography, *demographic research*, 29(22), pp.579-616.
- Ryznar, R. M. and Wagner, T. W., 2001. Using remotely sensed imagery to detect urban change: Viewing Detroit from space, *Journal of the American Planning Association*, 67, pp.327–36.
- Sadahiro, Y., 1999. Accuracy of areal interpolation: A comparison of alternative methods. *Journal of Geographical Systems*, 1, pp.323–46.
- Salvati, N., Tzavidis, N., Pratesi, M. and Chambers, R., 2012. Small area estimation via M-quantile geographically weighted regression, *Test* 21, pp.1-28.
- Sasaki, S., Igarashi, K., Fujino, Y., Comber, A. J., Brunson, C., Muleya, C. M. and Suzuki, H., 2010. The impact of community-based outreach immunisation services on immunisation coverage with GIS network accessibility analysis in Peru-urban areas, Zambia, *Journal of Epidemiol Community Health*, pp.1-8.
- Schmid, C.F. and MacCannell, E.H., 1955. Basic problems, techniques, and theory of isopleths mapping, *Journal of the American Statistical Association* 50(269), pp.220-39.
- Schmid, V. and Doerner, K. F., 2010. Ambulance location and relocation problems with time-dependent travel times, *European Journal of Operational Research*, 207(3), pp.1293-1303.
- Schroeder, J. P. and Van Riper, D. C., 2013. Because Muncie's Densities Are Not Manhattan's: Using Geographical Weighting in the Expectation-Maximization Algorithm for Areal Interpolation. *Geographical Analysis*, 45, pp.216–37.

Schumacher, J. V., Redmond, R. L., Hart, M. M. and Jensen, M. E., 2000. Mapping patterns of human use and potential resource conflicts on public lands, *Environmental Monitoring and Assessment*, 64 (1), pp.127–37.

Schuurman, N., Fiedler, R. S., Grzybowski, S. C.W. and Grund, D., 2006. Defining rational hospital catchments for non-urban areas based on travel-time, *International Journal of Health Geographics*, 5(43), pp.1-11.

Seidel, J. E., Beck, C. A., Poccobelli, G., Lemaire, J. B., Bugar, J. M., Quan, H., Ghali, W. A., 2006. Location of residence associated with the likelihood of patient visit to the preparative assessment clinic, *BMC Health Services Research*, 6(13), pp.1-9.

Sherman, J. E., Spencer, J., Preisser, J. S., Gesler, W. M. and Arcury, T. A., 2005. A suites of methods for representing activity space in a health care accessibility, *International Journal of Health Geographic*, 4 (24), pp.1-21.

Sleeter, R. and Gould, M., 2007. Geographic information system software to remodel population data using dasymetric mapping methods, *U.S. Geological Survey Techniques and Methods* 11-C2, pp.1-15.

Sridharan, H. and Qiu, F., 2013. A Spatially Disaggregated Areal Interpolation Model Using Light Detection and Ranging-Derived Building Volumes, *Geographical Analysis*, 45, pp.238-58.

Su, M., Lin, M., Hsieh, H., Tsai, B. and Lin, C., 2010. Multi-Layer Multi-class Dasymetric Mapping to Estimate Population Distribution, *Science of the Total Environment* 408, p.4807–16.

Sylvester, J. J., 1857. A question in the geometry of situation, *Quarterly Journal of Pure and Applied Mathematics*, 1, p.79.

Talen, E., 2003. Neighbourhoods as service providers: a methodology for evaluating pedestrian access, *Environment and Planning B: Planning and Design*, Volume 30, pp.181-200.

Tanser, F., Gijsbertsen, B. and Herbst, K., 2006. Modelling and understanding primary health care accessibility and utilisation in rural South Africa; An exploration using a geographical information system, *Social Science and Medicine*, 63, pp.691-705.

Tapp, A. F., 2010. Areal Interpolation and Dasymetric Mapping Methods Using Local Ancillary Data Sources, *Cartography and Geographic Information Science*, 37(3), pp.215-28.

Teitz, M. B. and Bart, P., 1968. Heuristic Methods for Estimating the Generalised Vertex Median of a Weighted Graph, *Journal of the Operational Research Society of America* 16 (5), pp. 955-61.

Tobler, W., 1979. Smooth Pycnophylactic Interpolation for Geographical Regions, *Journal of the American Statistical Association*, 74 (367), pp.519-30.

Tobler, W. R., 1992. Preliminary representation of World population by spherical harmonics, *Proceedings of the National Academy of Sciences of the United States of America*, 89, 14:6262-6264.

Tomintz, M., Clarke, G., Rigby, J. and Green, J., 2013. Optimising the location of antenatal classes, *Midwifery*, 29, pp.33-43.

Toregas, C., Swain, R., ReVelle, C. S. and Bergman, L., 1971. The location of emergency service facilities, *Operations Research* 19, pp.1363-73.

Turner, A. and Openshaw, S., 2001. Disaggregative spatial interpolation, *Paper presented at GISRUUK 2001 in Glamorgan Wales*.

United Nations Development Programme, 2000. Millennium Development Goal (UN 2000) [Online] Available at: <http://www.undp.org/mdg/goal4.shtml> [Accessed 30 December 2010].

Wang, Q., Ni, J. and Tenhunen, J., 2005. Application of geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems, *Global Ecology and Biogeography*, 14, pp.379–39.

Ward, D., Phinn, S. R. and Murray, A. T., 2000. Monitoring growth in rapidly urbanizing areas using remotely sensed data, *Professional Geographer*, 52, pp.371-86.

Wikina, B., 2011, Governor Amaechi to meet Bill Gates on health investment in Rivers State, Nigeria.

World Health Organisation, 1987. *Indicators for Primary Health Care*, World Health Organisation Regional Office for Europe, Copenhagen.

Wright, J. K., 1936. A method of mapping densities of population, *The Geographical Review* 26, pp.103–10.

Wu, S., Qiu, X. and Wang, L., 2005. Population estimation methods in GIS and remote sensing: a review. *Geographic Information Science and Remote Sensing*, 42, pp.80–96.

Xie, Y., 1995. The Overlaid Network Algorithms for the Areal Interpolation Problem, *Computer, Environment and Urban Systems*, 19(4), pp.287–306.

Yuan, Y., Smith, R. M. and Limp, W. F., 1997. Remodelling census population with spatial information from Landsat TM imagery, *Computers Environment and Urban Systems*, 21, pp.245–58.

Zandbergen, P. A. and Ignizio, D. A., 2010. Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates, *Cartography and Geographic Information Science*, 37(3), pp.199-214.

Zhang, C. and Qiu, F., 2011. A Point-based Intelligent Approach to Areal Interpolation, *The Professional Geographer* 63, pp.262–76.

Zhou, W., Huang, G., Troy, A. and Cadenasso, M. L., 2009. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study, *Remote Sensing of Environment*, 113, pp.1769-77.

## 10. APPENDICES

### Appendix 1: Metadata file information for Leicester image downloaded from USGS

```
GROUP = L1_METADATA_FILE
GROUP = METADATA_FILE_INFO
  ORIGIN = "Image courtesy of the U.S. Geological Survey"
  REQUEST_ID = "0101410048359_00002"
  LANDSAT_SCENE_ID = "LE72020232003106EDC00"
  FILE_DATE = 2014-10-05T00:09:14Z
  STATION_ID = "EDC"
  PROCESSING_SOFTWARE_VERSION = "LPGS_12.5.0"
  DATA_CATEGORY = "NOMINAL"
END_GROUP = METADATA_FILE_INFO
GROUP = PRODUCT_METADATA
  DATA_TYPE = "L1T"
  ELEVATION_SOURCE = "GLS2000"
  OUTPUT_FORMAT = "GEOTIFF"
  EPHEMERIS_TYPE = "DEFINITIVE"
  SPACECRAFT_ID = "LANDSAT_7"
  SENSOR_ID = "ETM"
  SENSOR_MODE = "SAM"
  WRS_PATH = 202
  WRS_ROW = 023
  DATE_ACQUIRED = 2003-04-16
  SCENE_CENTER_TIME = 10:46:55.9703464Z
  CORNER_UL_LAT_PRODUCT = 54.11875
  CORNER_UL_LON_PRODUCT = -2.24115
  CORNER_UR_LAT_PRODUCT = 54.04025
  CORNER_UR_LON_PRODUCT = 1.40991
  CORNER_LL_LAT_PRODUCT = 52.13949
  CORNER_LL_LON_PRODUCT = -2.27524
  CORNER_LR_LAT_PRODUCT = 52.06639
  CORNER_LR_LON_PRODUCT = 1.21252
  CORNER_UL_PROJECTION_X_PRODUCT = 549600.000
  CORNER_UL_PROJECTION_Y_PRODUCT = 5997000.000
  CORNER_UR_PROJECTION_X_PRODUCT = 788700.000
  CORNER_UR_PROJECTION_Y_PRODUCT = 5997000.000
  CORNER_LL_PROJECTION_X_PRODUCT = 549600.000
  CORNER_LL_PROJECTION_Y_PRODUCT = 5776800.000
  CORNER_LR_PROJECTION_X_PRODUCT = 788700.000
  CORNER_LR_PROJECTION_Y_PRODUCT = 5776800.000
  PANCHROMATIC_LINES = 14681
  PANCHROMATIC_SAMPLES = 15941
  REFLECTIVE_LINES = 7341
  REFLECTIVE_SAMPLES = 7971
  THERMAL_LINES = 7341
  THERMAL_SAMPLES = 7971
  FILE_NAME_BAND_1 = "LE72020232003106EDC00_B1.TIF"
  FILE_NAME_BAND_2 = "LE72020232003106EDC00_B2.TIF"
  FILE_NAME_BAND_3 = "LE72020232003106EDC00_B3.TIF"
  FILE_NAME_BAND_4 = "LE72020232003106EDC00_B4.TIF"
  FILE_NAME_BAND_5 = "LE72020232003106EDC00_B5.TIF"
  FILE_NAME_BAND_6_VCID_1 = "LE72020232003106EDC00_B6_VCID_1.TIF"
  FILE_NAME_BAND_6_VCID_2 = "LE72020232003106EDC00_B6_VCID_2.TIF"
  FILE_NAME_BAND_7 = "LE72020232003106EDC00_B7.TIF"
  FILE_NAME_BAND_8 = "LE72020232003106EDC00_B8.TIF"
```

GROUND\_CONTROL\_POINT\_FILE\_NAME = "LE72020232003106EDC00\_GCP.txt"  
METADATA\_FILE\_NAME = "LE72020232003106EDC00\_MTL.txt"  
CPF\_NAME = "L7CPF20030401\_20030531.09"  
END\_GROUP = PRODUCT\_METADATA  
GROUP = IMAGE\_ATTRIBUTES  
CLOUD\_COVER = 0.00  
IMAGE\_QUALITY = 9  
SUN\_AZIMUTH = 153.69006941  
SUN\_ELEVATION = 44.40193007  
GROUND\_CONTROL\_POINTS\_VERSION = 1  
GROUND\_CONTROL\_POINTS\_MODEL = 183  
GEOMETRIC\_RMSE\_MODEL = 4.027  
GEOMETRIC\_RMSE\_MODEL\_Y = 3.480  
GEOMETRIC\_RMSE\_MODEL\_X = 2.027  
END\_GROUP = IMAGE\_ATTRIBUTES  
GROUP = MIN\_MAX\_RADIANCE  
RADIANCE\_MAXIMUM\_BAND\_1 = 191.600  
RADIANCE\_MINIMUM\_BAND\_1 = -6.200  
RADIANCE\_MAXIMUM\_BAND\_2 = 196.500  
RADIANCE\_MINIMUM\_BAND\_2 = -6.400  
RADIANCE\_MAXIMUM\_BAND\_3 = 152.900  
RADIANCE\_MINIMUM\_BAND\_3 = -5.000  
RADIANCE\_MAXIMUM\_BAND\_4 = 241.100  
RADIANCE\_MINIMUM\_BAND\_4 = -5.100  
RADIANCE\_MAXIMUM\_BAND\_5 = 31.060  
RADIANCE\_MINIMUM\_BAND\_5 = -1.000  
RADIANCE\_MAXIMUM\_BAND\_6\_VCID\_1 = 17.040  
RADIANCE\_MINIMUM\_BAND\_6\_VCID\_1 = 0.000  
RADIANCE\_MAXIMUM\_BAND\_6\_VCID\_2 = 12.650  
RADIANCE\_MINIMUM\_BAND\_6\_VCID\_2 = 3.200  
RADIANCE\_MAXIMUM\_BAND\_7 = 10.800  
RADIANCE\_MINIMUM\_BAND\_7 = -0.350  
RADIANCE\_MAXIMUM\_BAND\_8 = 243.100  
RADIANCE\_MINIMUM\_BAND\_8 = -4.700  
END\_GROUP = MIN\_MAX\_RADIANCE  
GROUP = MIN\_MAX\_PIXEL\_VALUE  
QUANTIZE\_CAL\_MAX\_BAND\_1 = 255  
QUANTIZE\_CAL\_MIN\_BAND\_1 = 1  
QUANTIZE\_CAL\_MAX\_BAND\_2 = 255  
QUANTIZE\_CAL\_MIN\_BAND\_2 = 1  
QUANTIZE\_CAL\_MAX\_BAND\_3 = 255  
QUANTIZE\_CAL\_MIN\_BAND\_3 = 1  
QUANTIZE\_CAL\_MAX\_BAND\_4 = 255  
QUANTIZE\_CAL\_MIN\_BAND\_4 = 1  
QUANTIZE\_CAL\_MAX\_BAND\_5 = 255  
QUANTIZE\_CAL\_MIN\_BAND\_5 = 1  
QUANTIZE\_CAL\_MAX\_BAND\_6\_VCID\_1 = 255  
QUANTIZE\_CAL\_MIN\_BAND\_6\_VCID\_1 = 1  
QUANTIZE\_CAL\_MAX\_BAND\_6\_VCID\_2 = 255  
QUANTIZE\_CAL\_MIN\_BAND\_6\_VCID\_2 = 1  
QUANTIZE\_CAL\_MAX\_BAND\_7 = 255  
QUANTIZE\_CAL\_MIN\_BAND\_7 = 1  
QUANTIZE\_CAL\_MAX\_BAND\_8 = 255  
QUANTIZE\_CAL\_MIN\_BAND\_8 = 1  
END\_GROUP = MIN\_MAX\_PIXEL\_VALUE  
GROUP = PRODUCT\_PARAMETERS  
CORRECTION\_GAIN\_BAND\_1 = "CPF"  
CORRECTION\_GAIN\_BAND\_2 = "CPF"  
CORRECTION\_GAIN\_BAND\_3 = "CPF"  
CORRECTION\_GAIN\_BAND\_4 = "CPF"

CORRECTION\_GAIN\_BAND\_5 = "CPF"  
CORRECTION\_GAIN\_BAND\_6\_VCID\_1 = "CPF"  
CORRECTION\_GAIN\_BAND\_6\_VCID\_2 = "CPF"  
CORRECTION\_GAIN\_BAND\_7 = "CPF"  
CORRECTION\_GAIN\_BAND\_8 = "CPF"  
CORRECTION\_BIAS\_BAND\_1 = "INTERNAL\_CALIBRATION"  
CORRECTION\_BIAS\_BAND\_2 = "INTERNAL\_CALIBRATION"  
CORRECTION\_BIAS\_BAND\_3 = "INTERNAL\_CALIBRATION"  
CORRECTION\_BIAS\_BAND\_4 = "INTERNAL\_CALIBRATION"  
CORRECTION\_BIAS\_BAND\_5 = "INTERNAL\_CALIBRATION"  
CORRECTION\_BIAS\_BAND\_6\_VCID\_1 = "INTERNAL\_CALIBRATION"  
CORRECTION\_BIAS\_BAND\_6\_VCID\_2 = "INTERNAL\_CALIBRATION"  
CORRECTION\_BIAS\_BAND\_7 = "INTERNAL\_CALIBRATION"  
CORRECTION\_BIAS\_BAND\_8 = "INTERNAL\_CALIBRATION"  
GAIN\_BAND\_1 = "H"  
GAIN\_BAND\_2 = "H"  
GAIN\_BAND\_3 = "H"  
GAIN\_BAND\_4 = "L"  
GAIN\_BAND\_5 = "H"  
GAIN\_BAND\_6\_VCID\_1 = "L"  
GAIN\_BAND\_6\_VCID\_2 = "H"  
GAIN\_BAND\_7 = "H"  
GAIN\_BAND\_8 = "L"  
GAIN\_CHANGE\_BAND\_1 = "HH"  
GAIN\_CHANGE\_BAND\_2 = "HH"  
GAIN\_CHANGE\_BAND\_3 = "HH"  
GAIN\_CHANGE\_BAND\_4 = "LL"  
GAIN\_CHANGE\_BAND\_5 = "HH"  
GAIN\_CHANGE\_BAND\_6\_VCID\_1 = "LL"  
GAIN\_CHANGE\_BAND\_6\_VCID\_2 = "HH"  
GAIN\_CHANGE\_BAND\_7 = "HH"  
GAIN\_CHANGE\_BAND\_8 = "LL"  
GAIN\_CHANGE\_SCAN\_BAND\_1 = 0  
GAIN\_CHANGE\_SCAN\_BAND\_2 = 0  
GAIN\_CHANGE\_SCAN\_BAND\_3 = 0  
GAIN\_CHANGE\_SCAN\_BAND\_4 = 0  
GAIN\_CHANGE\_SCAN\_BAND\_5 = 0  
GAIN\_CHANGE\_SCAN\_BAND\_6\_VCID\_1 = 0  
GAIN\_CHANGE\_SCAN\_BAND\_6\_VCID\_2 = 0  
GAIN\_CHANGE\_SCAN\_BAND\_7 = 0  
GAIN\_CHANGE\_SCAN\_BAND\_8 = 0  
END\_GROUP = PRODUCT\_PARAMETERS  
GROUP = RADIOMETRIC\_RESCALING  
RADIANCE\_MULT\_BAND\_1 = 0.779  
RADIANCE\_MULT\_BAND\_2 = 0.799  
RADIANCE\_MULT\_BAND\_3 = 0.622  
RADIANCE\_MULT\_BAND\_4 = 0.969  
RADIANCE\_MULT\_BAND\_5 = 0.126  
RADIANCE\_MULT\_BAND\_6\_VCID\_1 = 0.067  
RADIANCE\_MULT\_BAND\_6\_VCID\_2 = 0.037  
RADIANCE\_MULT\_BAND\_7 = 0.044  
RADIANCE\_MULT\_BAND\_8 = 0.976  
RADIANCE\_ADD\_BAND\_1 = -6.97874  
RADIANCE\_ADD\_BAND\_2 = -7.19882  
RADIANCE\_ADD\_BAND\_3 = -5.62165  
RADIANCE\_ADD\_BAND\_4 = -6.06929  
RADIANCE\_ADD\_BAND\_5 = -1.12622  
RADIANCE\_ADD\_BAND\_6\_VCID\_1 = -0.06709  
RADIANCE\_ADD\_BAND\_6\_VCID\_2 = 3.16280  
RADIANCE\_ADD\_BAND\_7 = -0.39390

```
RADIANCE_ADD_BAND_8 = -5.67559
END_GROUP = RADIOMETRIC_RESCALING
GROUP = PROJECTION_PARAMETERS
MAP_PROJECTION = "UTM"
DATUM = "WGS84"
ELLIPSOID = "WGS84"
UTM_ZONE = 30
GRID_CELL_SIZE_PANCHROMATIC = 15.00
GRID_CELL_SIZE_REFLECTIVE = 30.00
GRID_CELL_SIZE_THERMAL = 30.00
ORIENTATION = "NORTH_UP"
RESAMPLING_OPTION = "CUBIC_CONVOLUTION"
END_GROUP = PROJECTION_PARAMETERS
END_GROUP = L1_METADATA_FILE
END
```

## Appendix 2: Metadata file information for Port-Harcourt image downloaded from

### USGS

GROUP = L1\_METADATA\_FILE  
GROUP = METADATA\_FILE\_INFO  
ORIGIN = "Image courtesy of the U.S. Geological Survey"  
REQUEST\_ID = "0101206084812\_00004"  
PRODUCT\_CREATION\_TIME = 2012-06-10T04:17:59Z  
STATION\_ID = "EDC"  
LANDSAT7\_XBAND = "2"  
GROUND\_STATION = "SGS"  
LPS\_PROCESSOR\_NUMBER = 1  
DATEHOUR\_CONTACT\_PERIOD = "0300814"  
SUBINTERVAL\_NUMBER = "04"  
END\_GROUP = METADATA\_FILE\_INFO  
GROUP = PRODUCT\_METADATA  
PRODUCT\_TYPE = "L1T"  
ELEVATION\_SOURCE = "GLS2000"  
PROCESSING\_SOFTWARE = "LPGS\_12.0.2"  
EPHEMERIS\_TYPE = "DEFINITIVE"  
SPACECRAFT\_ID = "Landsat7"  
SENSOR\_ID = "ETM+"  
SENSOR\_MODE = "SAM"  
ACQUISITION\_DATE = 2003-01-08  
SCENE\_CENTER\_SCAN\_TIME = 09:33:30.6905415Z  
WRS\_PATH = 188  
STARTING\_ROW = 57  
ENDING\_ROW = 57  
BAND\_COMBINATION = "123456678"  
PRODUCT\_UL\_CORNER\_LAT = 5.2896247  
PRODUCT\_UL\_CORNER\_LON = 6.3620346  
PRODUCT\_UR\_CORNER\_LAT = 5.2950219  
PRODUCT\_UR\_CORNER\_LON = 8.4783619  
PRODUCT\_LL\_CORNER\_LAT = 3.3918211  
PRODUCT\_LL\_CORNER\_LON = 6.3686200  
PRODUCT\_LR\_CORNER\_LAT = 3.3952763  
PRODUCT\_LR\_CORNER\_LON = 8.4796659  
PRODUCT\_UL\_CORNER\_MAPX = 207600.000  
PRODUCT\_UL\_CORNER\_MAPY = 585300.000  
PRODUCT\_UR\_CORNER\_MAPX = 442200.000  
PRODUCT\_UR\_CORNER\_MAPY = 585300.000  
PRODUCT\_LL\_CORNER\_MAPX = 207600.000  
PRODUCT\_LL\_CORNER\_MAPY = 375300.000  
PRODUCT\_LR\_CORNER\_MAPX = 442200.000  
PRODUCT\_LR\_CORNER\_MAPY = 375300.000  
PRODUCT\_SAMPLES\_PAN = 15641  
PRODUCT\_LINES\_PAN = 14001  
PRODUCT\_SAMPLES\_REF = 7821  
PRODUCT\_LINES\_REF = 7001  
PRODUCT\_SAMPLES\_THM = 7821  
PRODUCT\_LINES\_THM = 7001  
BAND1\_FILE\_NAME = "L71188057\_05720030108\_B10.TIF"  
BAND2\_FILE\_NAME = "L71188057\_05720030108\_B20.TIF"  
BAND3\_FILE\_NAME = "L71188057\_05720030108\_B30.TIF"  
BAND4\_FILE\_NAME = "L71188057\_05720030108\_B40.TIF"  
BAND5\_FILE\_NAME = "L71188057\_05720030108\_B50.TIF"  
BAND61\_FILE\_NAME = "L71188057\_05720030108\_B61.TIF"  
BAND62\_FILE\_NAME = "L72188057\_05720030108\_B62.TIF"  
BAND7\_FILE\_NAME = "L72188057\_05720030108\_B70.TIF"  
BAND8\_FILE\_NAME = "L72188057\_05720030108\_B80.TIF"

GCP\_FILE\_NAME = "L71188057\_05720030108\_GCP.txt"  
METADATA\_L1\_FILE\_NAME = "L71188057\_05720030108\_MTL.txt"  
CPF\_FILE\_NAME = "L7CPF20030101\_20030331\_06"  
END\_GROUP = PRODUCT\_METADATA  
GROUP = MIN\_MAX\_RADIANCE  
LMAX\_BAND1 = 293.700  
LMIN\_BAND1 = -6.200  
LMAX\_BAND2 = 300.900  
LMIN\_BAND2 = -6.400  
LMAX\_BAND3 = 234.400  
LMIN\_BAND3 = -5.000  
LMAX\_BAND4 = 241.100  
LMIN\_BAND4 = -5.100  
LMAX\_BAND5 = 47.570  
LMIN\_BAND5 = -1.000  
LMAX\_BAND61 = 17.040  
LMIN\_BAND61 = 0.000  
LMAX\_BAND62 = 12.650  
LMIN\_BAND62 = 3.200  
LMAX\_BAND7 = 16.540  
LMIN\_BAND7 = -0.350  
LMAX\_BAND8 = 243.100  
LMIN\_BAND8 = -4.700  
END\_GROUP = MIN\_MAX\_RADIANCE  
GROUP = MIN\_MAX\_PIXEL\_VALUE  
QCALMAX\_BAND1 = 255.0  
QCALMIN\_BAND1 = 1.0  
QCALMAX\_BAND2 = 255.0  
QCALMIN\_BAND2 = 1.0  
QCALMAX\_BAND3 = 255.0  
QCALMIN\_BAND3 = 1.0  
QCALMAX\_BAND4 = 255.0  
QCALMIN\_BAND4 = 1.0  
QCALMAX\_BAND5 = 255.0  
QCALMIN\_BAND5 = 1.0  
QCALMAX\_BAND61 = 255.0  
QCALMIN\_BAND61 = 1.0  
QCALMAX\_BAND62 = 255.0  
QCALMIN\_BAND62 = 1.0  
QCALMAX\_BAND7 = 255.0  
QCALMIN\_BAND7 = 1.0  
QCALMAX\_BAND8 = 255.0  
QCALMIN\_BAND8 = 1.0  
END\_GROUP = MIN\_MAX\_PIXEL\_VALUE  
GROUP = PRODUCT\_PARAMETERS  
CORRECTION\_METHOD\_GAIN\_BAND1 = "CPF"  
CORRECTION\_METHOD\_GAIN\_BAND2 = "CPF"  
CORRECTION\_METHOD\_GAIN\_BAND3 = "CPF"  
CORRECTION\_METHOD\_GAIN\_BAND4 = "CPF"  
CORRECTION\_METHOD\_GAIN\_BAND5 = "CPF"  
CORRECTION\_METHOD\_GAIN\_BAND61 = "CPF"  
CORRECTION\_METHOD\_GAIN\_BAND62 = "CPF"  
CORRECTION\_METHOD\_GAIN\_BAND7 = "CPF"  
CORRECTION\_METHOD\_GAIN\_BAND8 = "CPF"  
CORRECTION\_METHOD\_BIAS = "IC"  
BAND1\_GAIN = "L"  
BAND2\_GAIN = "L"  
BAND3\_GAIN = "L"  
BAND4\_GAIN = "L"  
BAND5\_GAIN = "L"

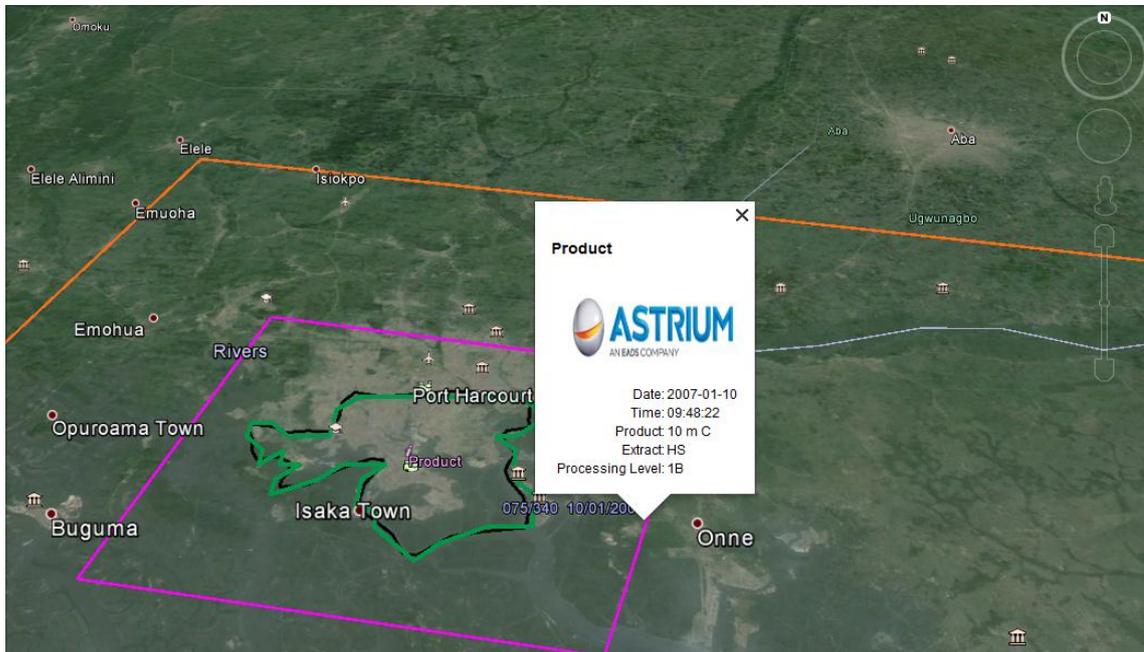
```

BAND6_GAIN1 = "L"
BAND6_GAIN2 = "H"
BAND7_GAIN = "L"
BAND8_GAIN = "L"
BAND1_GAIN_CHANGE = "0"
BAND2_GAIN_CHANGE = "0"
BAND3_GAIN_CHANGE = "0"
BAND4_GAIN_CHANGE = "0"
BAND5_GAIN_CHANGE = "0"
BAND6_GAIN_CHANGE1 = "0"
BAND6_GAIN_CHANGE2 = "0"
BAND7_GAIN_CHANGE = "0"
BAND8_GAIN_CHANGE = "0"
BAND1_SL_GAIN_CHANGE = 0
BAND2_SL_GAIN_CHANGE = 0
BAND3_SL_GAIN_CHANGE = 0
BAND4_SL_GAIN_CHANGE = 0
BAND5_SL_GAIN_CHANGE = 0
BAND6_SL_GAIN_CHANGE1 = 0
BAND6_SL_GAIN_CHANGE2 = 0
BAND7_SL_GAIN_CHANGE = 0
BAND8_SL_GAIN_CHANGE = 0
SUN_AZIMUTH = 132.6855179
SUN_ELEVATION = 49.7604205
OUTPUT_FORMAT = "GEOTIFF"
END_GROUP = PRODUCT_PARAMETERS
GROUP = CORRECTIONS_APPLIED
STRIPING_BAND1 = "NONE"
STRIPING_BAND2 = "NONE"
STRIPING_BAND3 = "NONE"
STRIPING_BAND4 = "NONE"
STRIPING_BAND5 = "NONE"
STRIPING_BAND61 = "NONE"
STRIPING_BAND62 = "NONE"
STRIPING_BAND7 = "NONE"
STRIPING_BAND8 = "NONE"
BANDING = "N"
COHERENT_NOISE = "Y"
MEMORY_EFFECT = "N"
SCAN_CORRELATED_SHIFT = "N"
INOPERABLE_DETECTORS = "N"
DROPPED_LINES = "N"
END_GROUP = CORRECTIONS_APPLIED
GROUP = PROJECTION_PARAMETERS
REFERENCE_DATUM = "WGS84"
REFERENCE_ELLIPSOID = "WGS84"
GRID_CELL_SIZE_PAN = 15.000
GRID_CELL_SIZE_THM = 30.000
GRID_CELL_SIZE_REF = 30.000
ORIENTATION = "NUP"
RESAMPLING_OPTION = "CC"
MAP_PROJECTION = "UTM"
END_GROUP = PROJECTION_PARAMETERS
GROUP = UTM_PARAMETERS
ZONE_NUMBER = 32
END_GROUP = UTM_PARAMETERS
END_GROUP = L1_METADATA_FILE
END

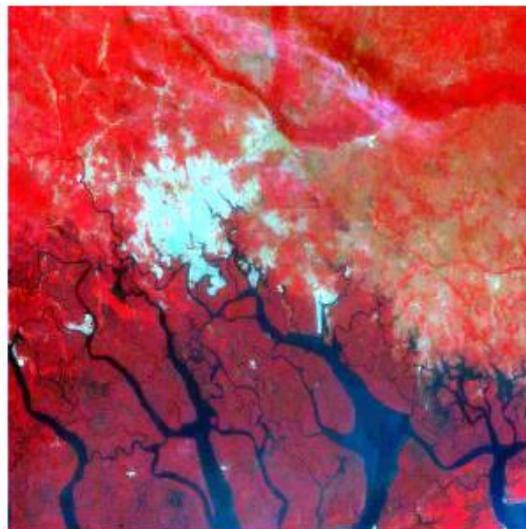
```

### Appendix 3: Port-Harcourt Spot5 imagery information from Astrium Services

Spot5 imagery colour 10m spatial resolution taken 10<sup>th</sup> January 2007



075/340 10/01/2007



Satellite: Spot5  
 Center Latitude: 4.75  
 Center Longitude: 7.10  
 Date: 10/01/2007  
 Time: 09:48:22  
 Mode: J  
 Angle: -25.81°  
 SAT: Shift5  
 Quality: EEEE

[Click here for more information](#)

**Appendix 4: The mean plot window for resampled aerial photo of 10m and 3m spatial resolution image of Leicester.**

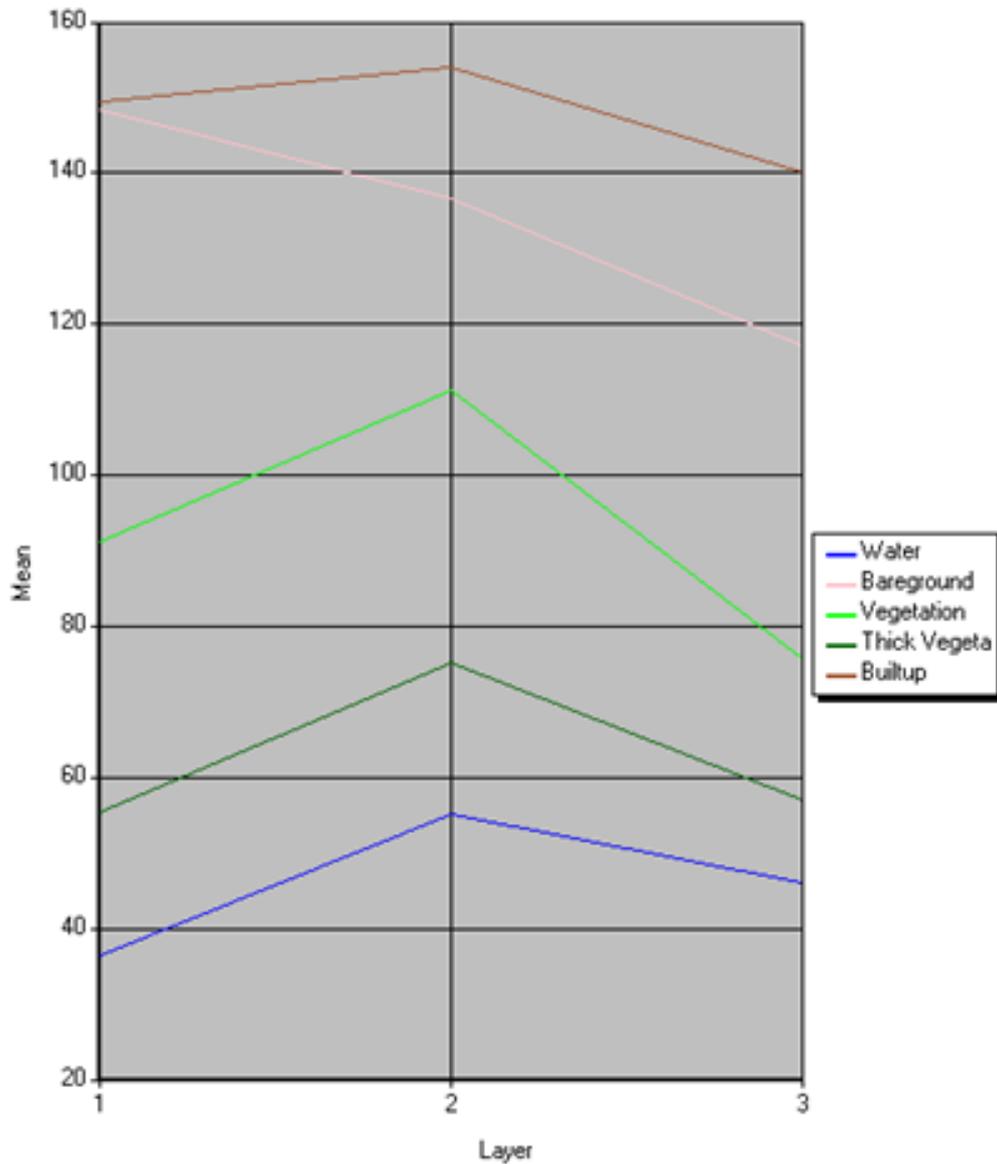


Figure A4.1 - Signature mean plot for the combined signatures from 10m resolution image

Class #	>	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1		Water	<span style="color: blue;">■</span>	0.000	0.000	1.000	35	36	349	1.000	✓	✓	✓	✓	
2		Bareground	<span style="color: pink;">■</span>	1.000	0.753	0.796	1	37	2791	1.000	✓	✓	✓	✓	
3		Vegetation	<span style="color: green;">■</span>	0.000	1.000	0.000	2	38	2090	1.000	✓	✓	✓	✓	
4		Thick Vegetation	<span style="color: darkgreen;">■</span>	0.000	0.392	0.000	3	39	1120	1.000	✓	✓	✓	✓	
5	▶	Builtup	<span style="color: brown;">■</span>	0.627	0.322	0.176	4	40	1537	1.000	✓	✓	✓	✓	

Figure A4.2 - Signature editor for the combined signatures from 10m resolution image

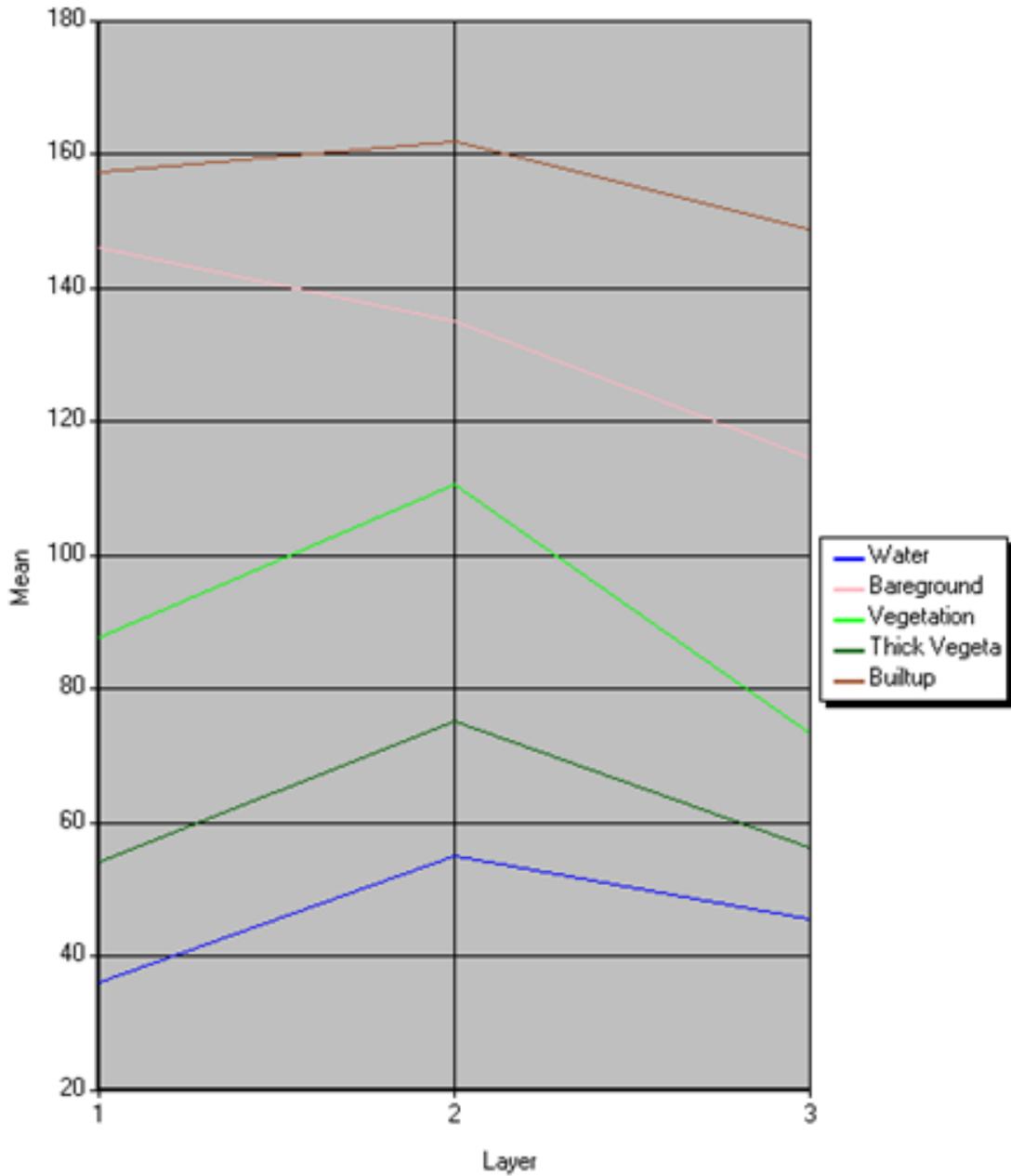


Figure A4.3 - Signature mean plot for the combined signatures from 3m resolution image

Class #	>	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1	▶	Water	<span style="color: blue;">■</span>	0.000	0.000	1.000	33	33	2726	1.000	✓	✓	✓	✓	
2		Bareground	<span style="color: pink;">■</span>	1.000	0.714	0.757	1	34	27164	1.000	✓	✓	✓	✓	
3		Vegetation	<span style="color: green;">■</span>	0.000	1.000	0.000	2	35	16704	1.000	✓	✓	✓	✓	
4		Thick Vegetation	<span style="color: darkgreen;">■</span>	0.000	0.392	0.000	3	36	14115	1.000	✓	✓	✓	✓	
5		Builtup	<span style="color: brown;">■</span>	0.627	0.322	0.176	4	37	14623	1.000	✓	✓	✓	✓	

Figure A4.4 - Signature editor for the combined signatures from 3m resolution image

**Appendix 5: The signature editor and the signature mean plot for the combined signatures for resampled aerial photo of 10m and 3m spatial resolution image of Leicester**

Class #	>	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1		Water		0.000	0.000	1.000	35	36	349	1.000	✓	✓	✓	✓	
2		Bareground		1.000	0.753	0.796	1	37	2791	1.000	✓	✓	✓	✓	
3		Vegetation		0.000	1.000	0.000	2	38	2090	1.000	✓	✓	✓	✓	
4		Thick Vegetation		0.000	0.392	0.000	3	39	1120	1.000	✓	✓	✓	✓	
5	▶	Builtup		0.627	0.322	0.176	4	40	1537	1.000	✓	✓	✓	✓	

Figure A5.1 - Signature editor for the combined signatures from resampled aerial photo of 10m spatial resolution.

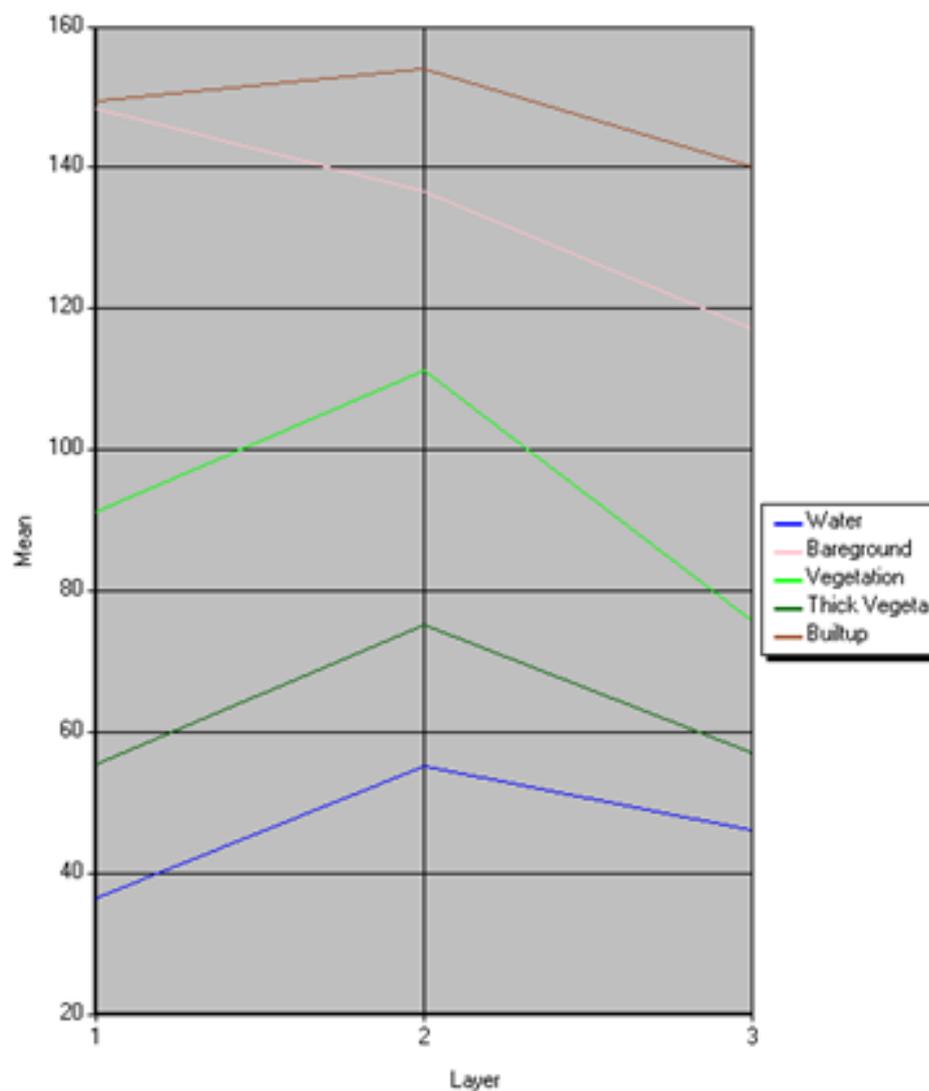


Figure A5.2 - Signature mean plot for the combined signatures from resampled aerial photo of 10m spatial resolution.

Class #	>	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1	▶	Water		0.000	0.000	1.000	33	33	2726	1.000	✓	✓	✓	✓	
2		Bareground		1.000	0.714	0.757	1	34	27164	1.000	✓	✓	✓	✓	
3		Vegetation		0.000	1.000	0.000	2	35	16704	1.000	✓	✓	✓	✓	
4		Thick Vegetation		0.000	0.392	0.000	3	36	14115	1.000	✓	✓	✓	✓	
5		Builtup		0.627	0.322	0.176	4	37	14623	1.000	✓	✓	✓	✓	

Figure A5.3 - Signature editor for the combined signatures from resampled aerial photo of 3m spatial resolution.

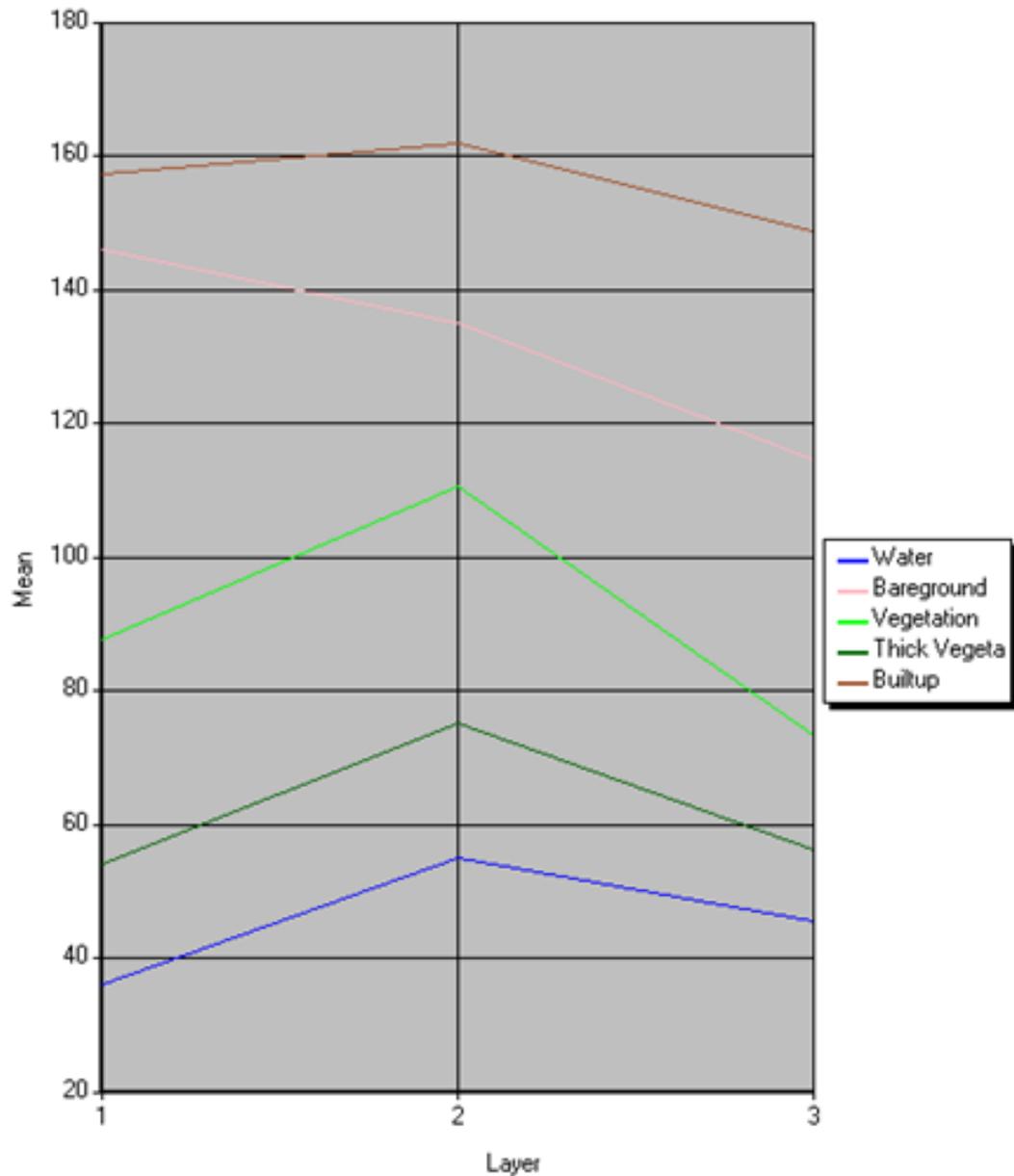


Figure A5.4 - Signature mean plot for the combined signatures from resampled aerial photo of 3m spatial resolution.

## Appendix 6: Classification accuracy report for Leicester

Image File: z:/thesis\_correction/leicester\_30m/leic30m\_sup.img

User Name: ijm14

Date: Mon Nov 24 11:22:35 2014

### ERROR MATRIX

Classified Data	Reference Data				Row Total
	Unclassified	Vegetation	Water	Builtup	
-----	-----	-----	-----	-----	-----
Unclassified	130	0	0	0	130
Vegetation	0	26	23	0	49
Water	0	0	2	0	2
Builtup	0	7	1	67	75
Column Total	130	33	26	67	256

----- End of Error Matrix -----

### ACCURACY TOTALS

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
-----	-----	-----	-----	-----	-----
Unclassified	130	130	130	---	---
Vegetation	33	49	26	78.79%	53.06%
Water	26	2	2	7.69%	100.00%
Builtup	67	75	67	100.00%	89.33%
Totals	256	256	225		

Overall Classification Accuracy = 87.89%

----- End of Accuracy Totals -----

### KAPPA (K<sup>^</sup>) STATISTICS

Overall Kappa Statistics = 0.8108

#### Conditional Kappa for each Category

Class Name	Kappa
Unclassified	1.0000
Vegetation	0.4612
Water	1.0000
Builtup	0.8555

----- End of Kappa Statistics -----

**Accuracy report for 10m image**

Image File: z:/thesis\_correction/leicester\_10m/leic\_10m\_sup.img

User Name: ijm14

Date: Mon Nov 24 23:03:09 2014

**ERROR MATRIX**

Reference Data							Row
Classified Data	Unclassifi	Bareground	Vegetation	Thick Veg	Builtup	Water	Total
Unclassified	147	0	0	0	0	0	147
Bareground	0	3	0	0	0	0	3
Vegetation	0	0	34	2	0	14	50
Thick Veget.	0	0	0	4	0	0	4
Builtup	0	22	0	3	26	0	51
Water	0	0	0	0	0	1	1
Column Total	147	25	34	9	26	15	256

----- End of Error Matrix -----

**ACCURACY TOTALS**

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Unclassified	147	147	147	---	---
Bareground	25	3	3	12.00%	100.00%
Vegetation	34	50	32	100.00%	64.00%
Thick Veget.	9	4	4	44.44%	100.00%
Builtup	26	51	26	100.00%	50.98%
Water	15	1	1	6.67%	100.00%
Totals	256	256	213		

Overall Classification Accuracy = 83.20%

----- End of Accuracy Totals -----

KAPPA (K^) STATISTICS

Overall Kappa Statistics = 0.7307

Conditional Kappa for each Category

Class Name	Kappa
Unclassified	1.0000
Bareground	1.0000
Vegetation	0.5886
Thick Vegetat.	1.0000
Builtup	0.4544
Water	1.0000

----- End of Kappa Statistics -----

**Accuracy report for 3m image**

Image File: z:/thesis\_correction/leicester\_3m/leic\_3m\_sup.img

User Name: ijm14

Date: Wed Nov 26 13:26:36 2014

ERROR MATRIX

Reference Data

Classified Data	Unclassifi	Bareground	Vegetation	Thick Veg	Builtup	Water	Row Total
Unclassified	143	0	0	0	0	0	143
Bareground	0	0	0	0	3	0	3
Vegetation	0	0	29	8	0	6	43
Thick Veget.	0	0	0	3	0	4	7
Builtup	0	21	1	0	34	0	56
Water	0	0	0	3	0	1	4
Column Total	143	21	30	14	37	11	256

----- End of Error Matrix -----

ACCURACY TOTALS

---

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
-----	-----	-----	-----	-----	-----
Unclassified	143	143	143	---	---
Bareground	21	3	0	0.00%	0.00%
Vegetation	30	43	29	96.67%	67.44%
Thick Veget.	14	7	3	21.43%	42.86%
Builtup	37	56	34	91.89%	60.71%
Water	11	4	1	9.09%	25.00%
Totals	256	256	210		

---

Overall Classification Accuracy = 82.03%

----- End of Accuracy Totals -----

KAPPA (K<sup>^</sup>) STATISTICS

Overall Kappa Statistics = 0.7164

Conditional Kappa for each Category

Class Name	Kappa
Unclassified	1
Bareground	0.0894
Vegetation	0.6312
Thick Vegetat.	0.3955
Builtup	0.5408
Water	0.2163

----- End of Kappa Statistics -----

## Appendix 7: Residual maps

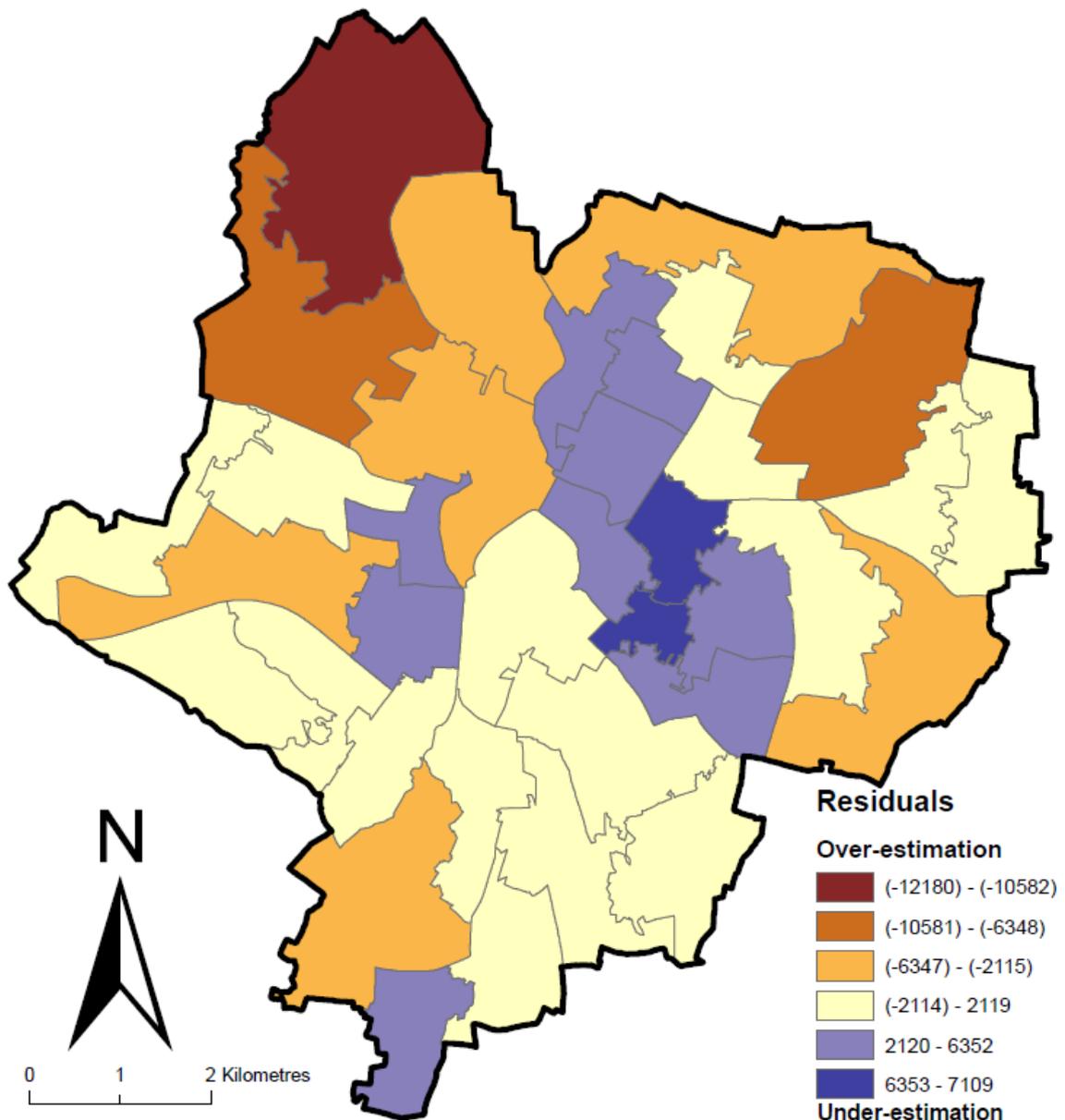


Figure A7.1 - The distribution of residuals using a 30m gridded pycnophylactic population surface at MSOA. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

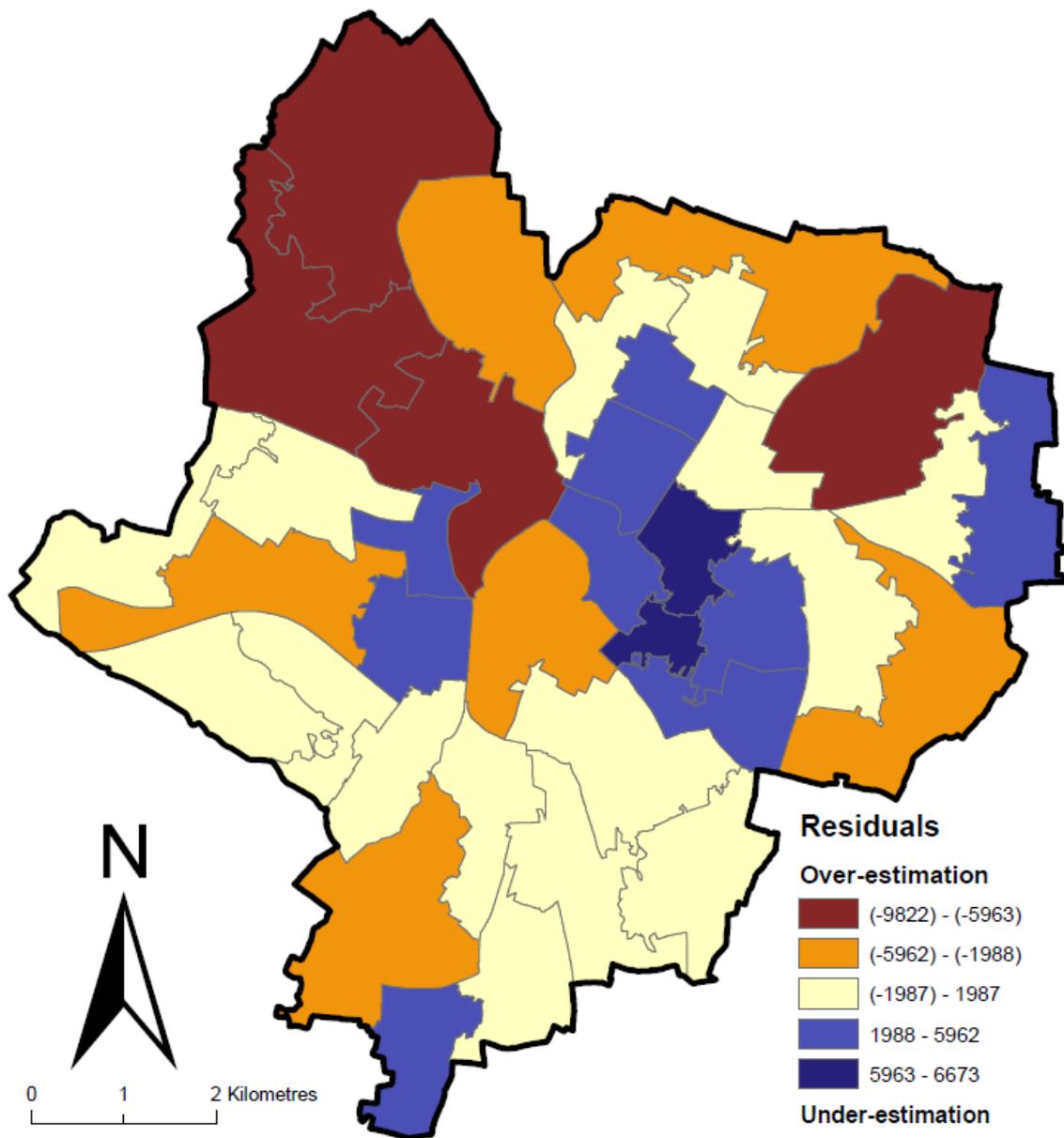


Figure A7.2 - The distribution of residuals using a 100m gridded pycnophylactic population surface at MSOA. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

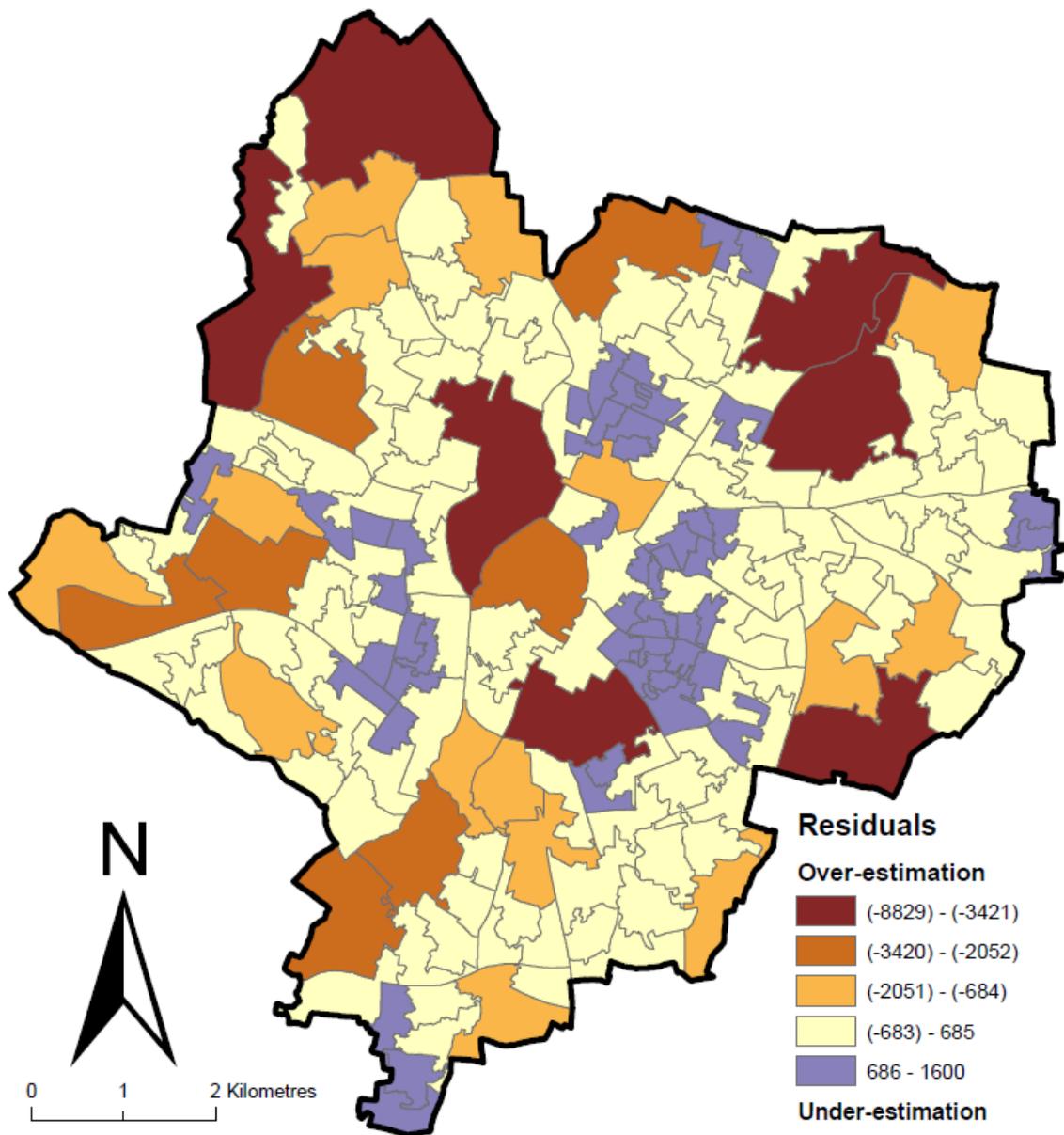


Figure A7.3 - The distribution of residuals using a 30m gridded pycnophylactic population surface at LSOA. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

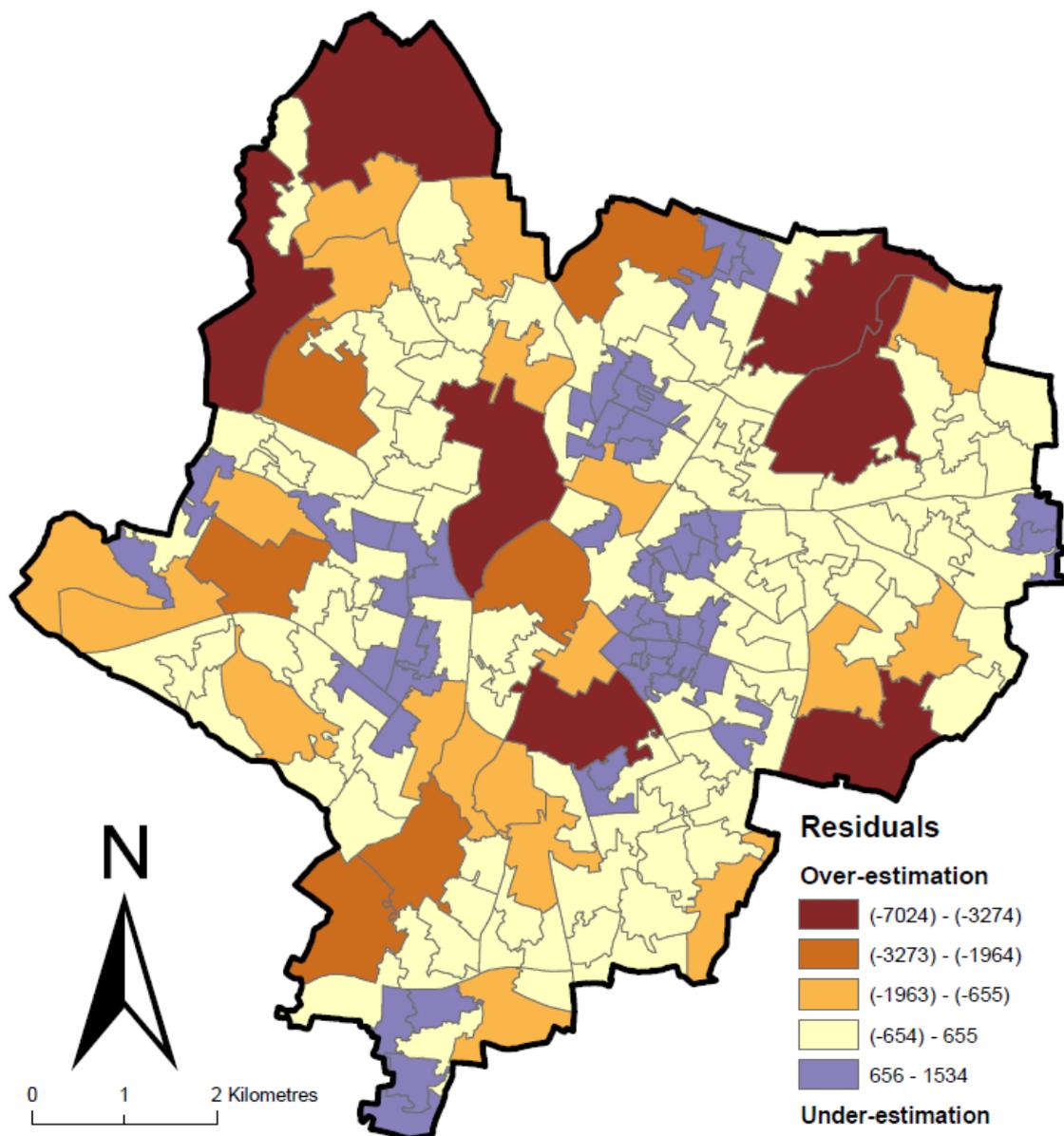


Figure A7.4 - The distribution of residuals using a 100m gridded pycnophylactic population surface at LSOA. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

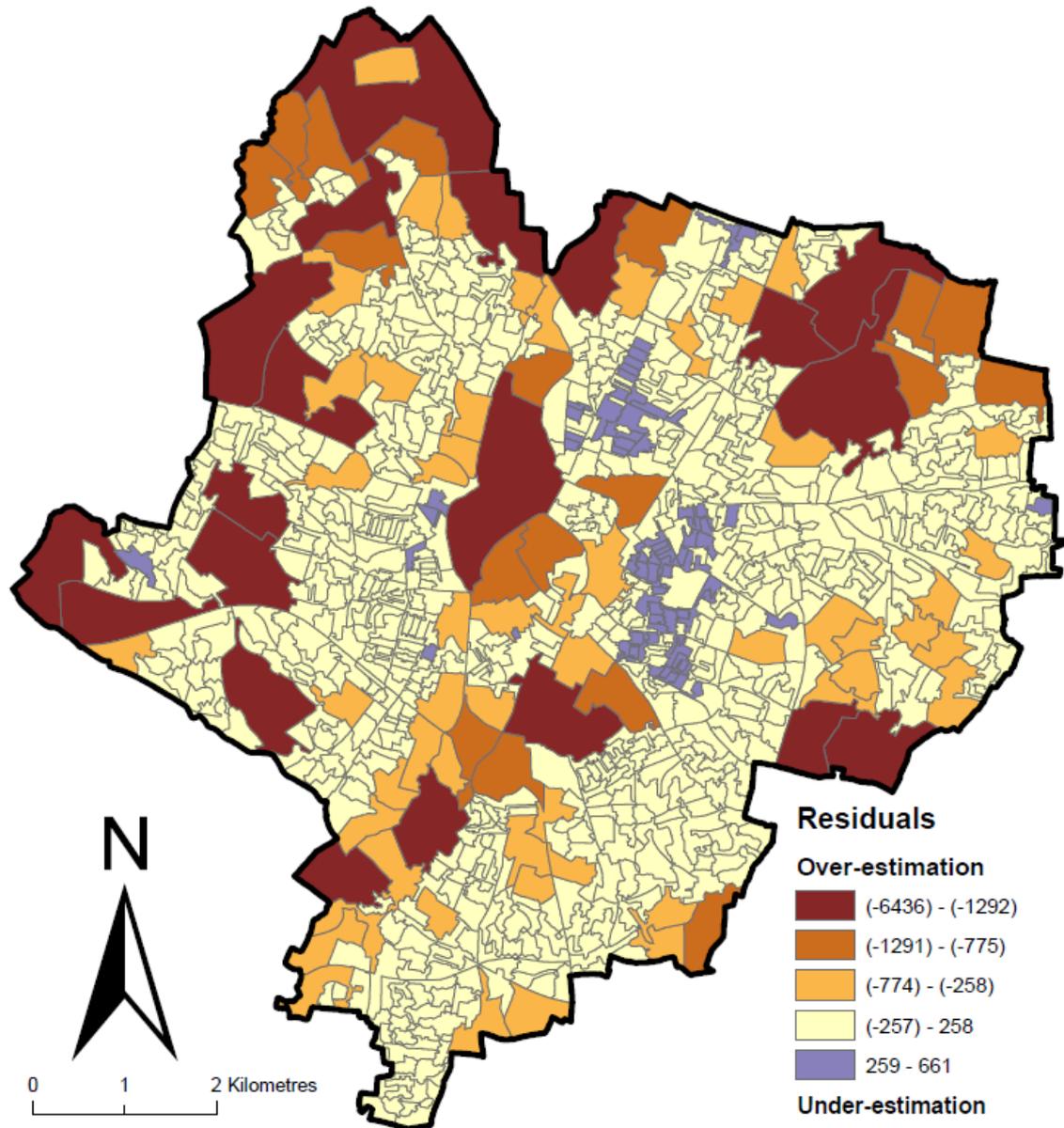


Figure A7.5 - The distribution of residuals using a 30m gridded pycnophylactic population surface at OA. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

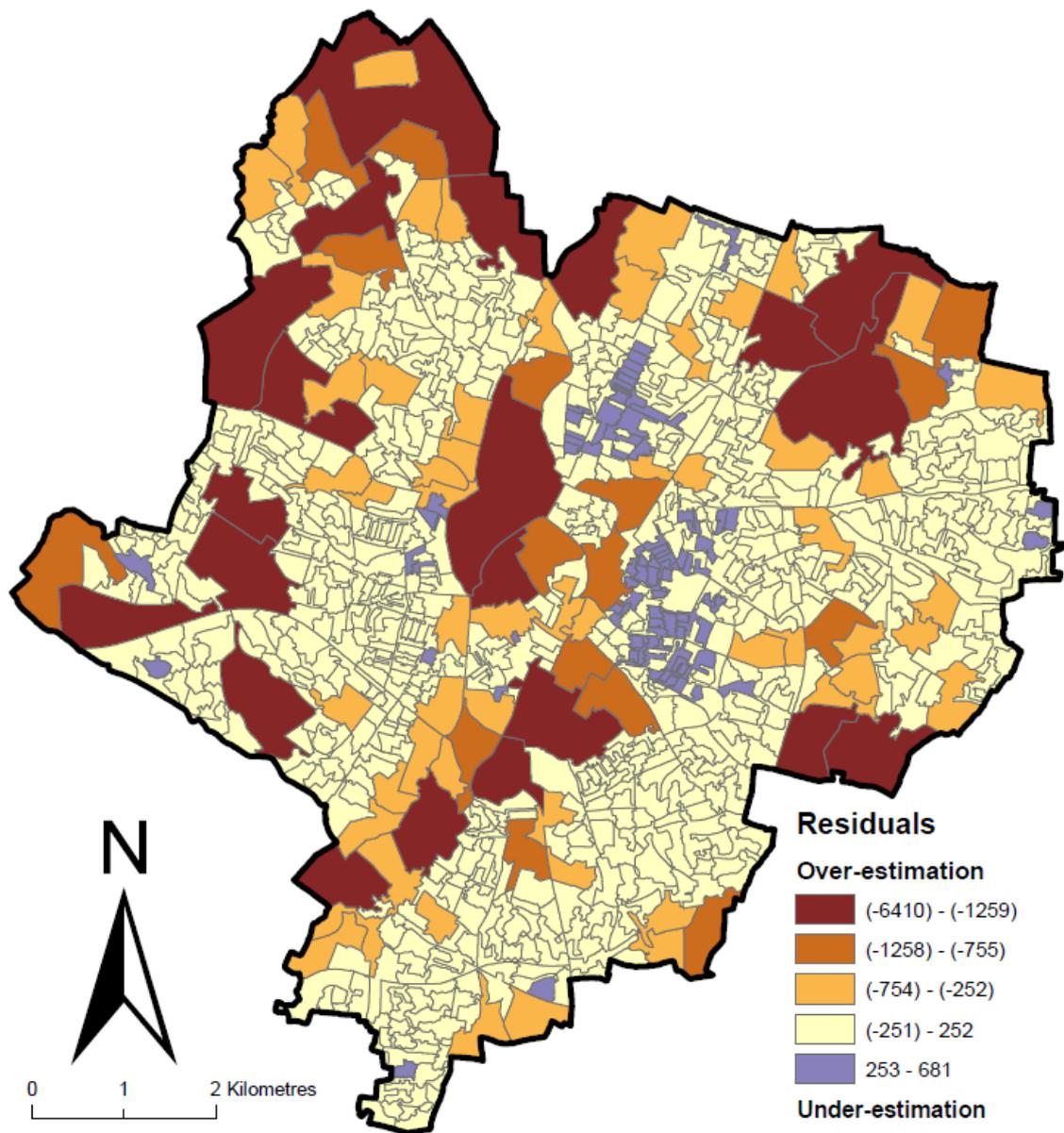


Figure A7.6 - The distribution of residuals using a 100m gridded pycnophylactic population surface at OA. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

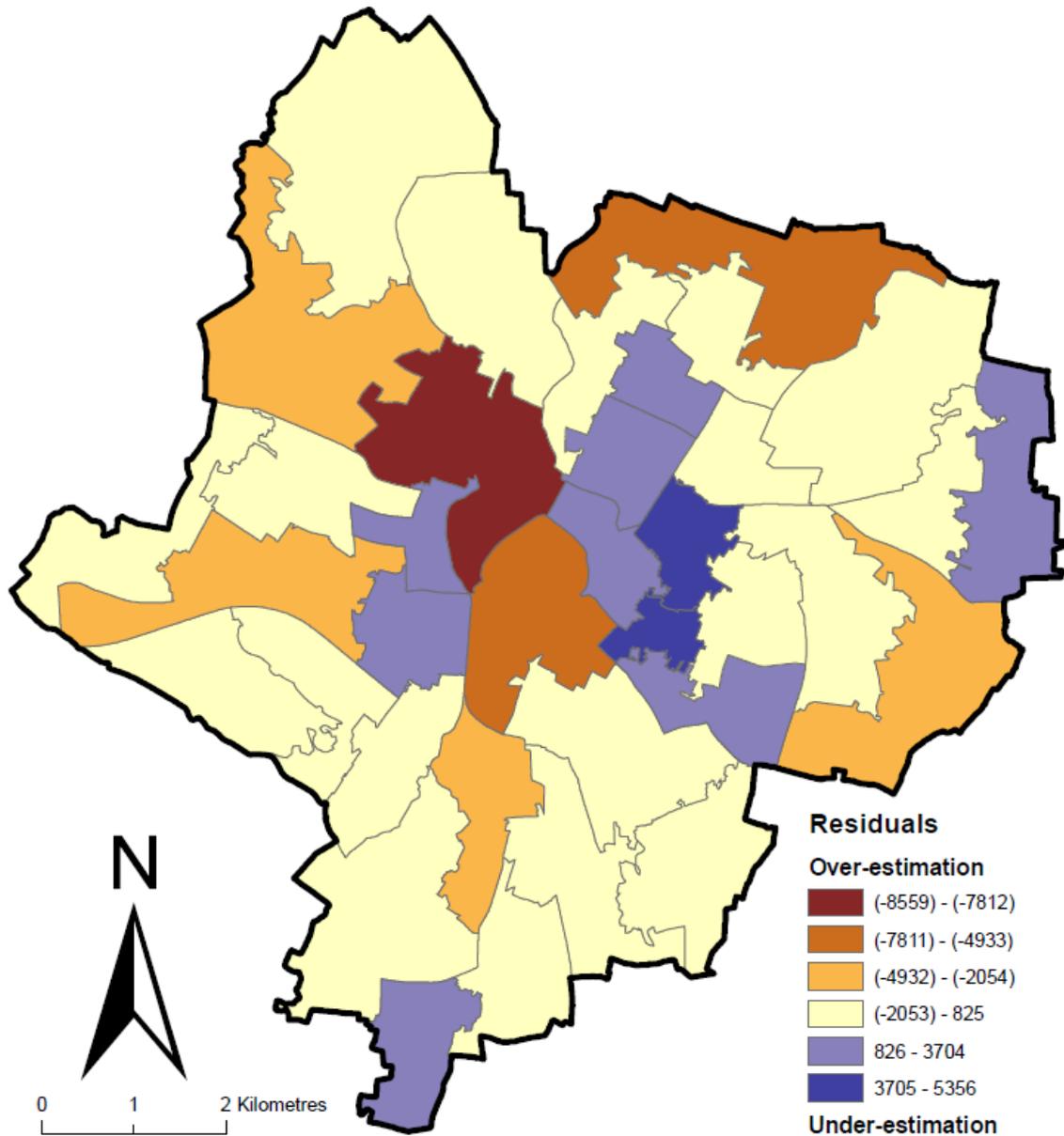


Figure A7.7 - The distribution of residuals using a 30m gridded dasymetric population surface at MSOA using 30m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

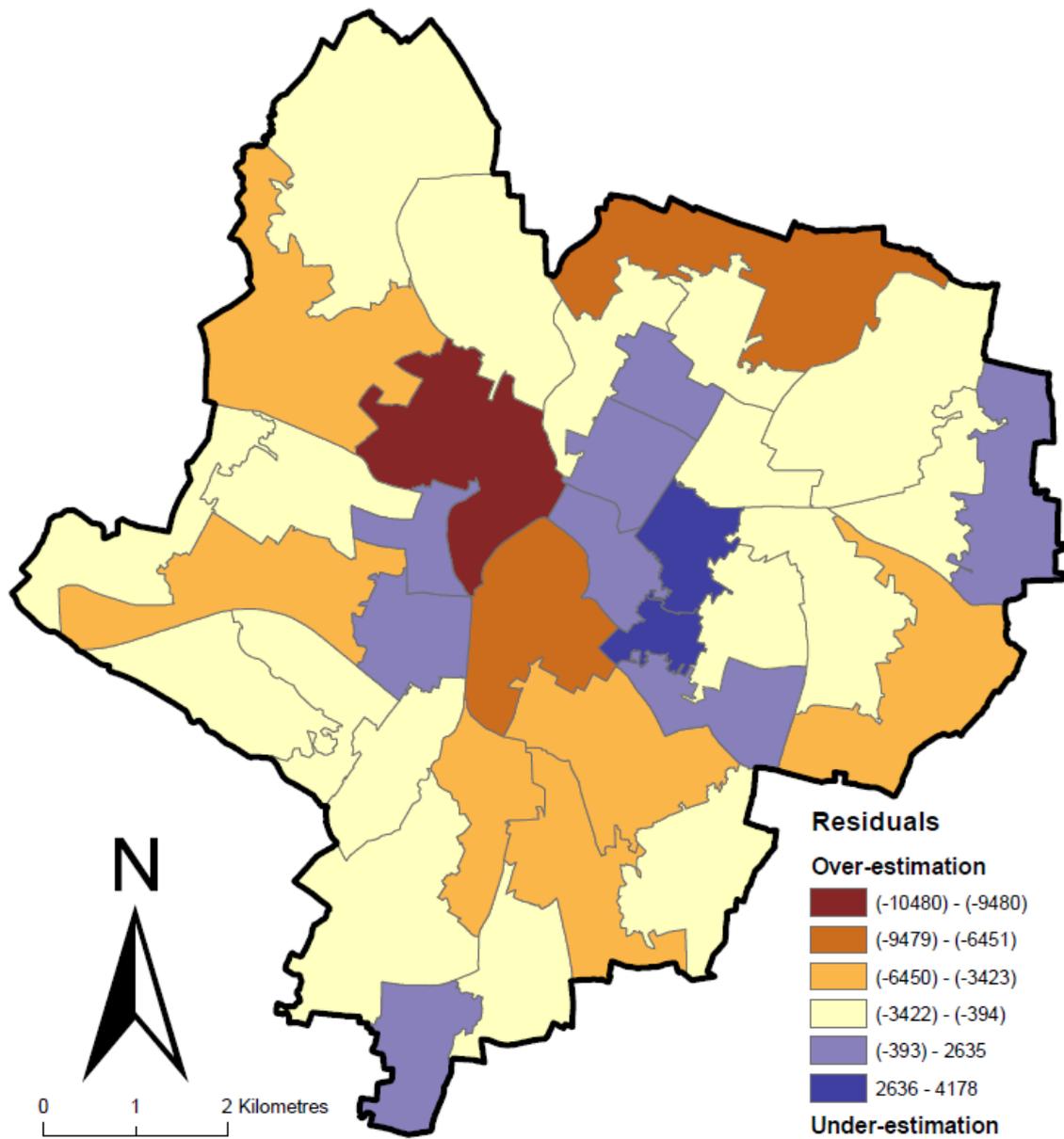


Figure A7.8 - The distribution of residuals using a 100m gridded dasymmetric population surface at MSOA using 30m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

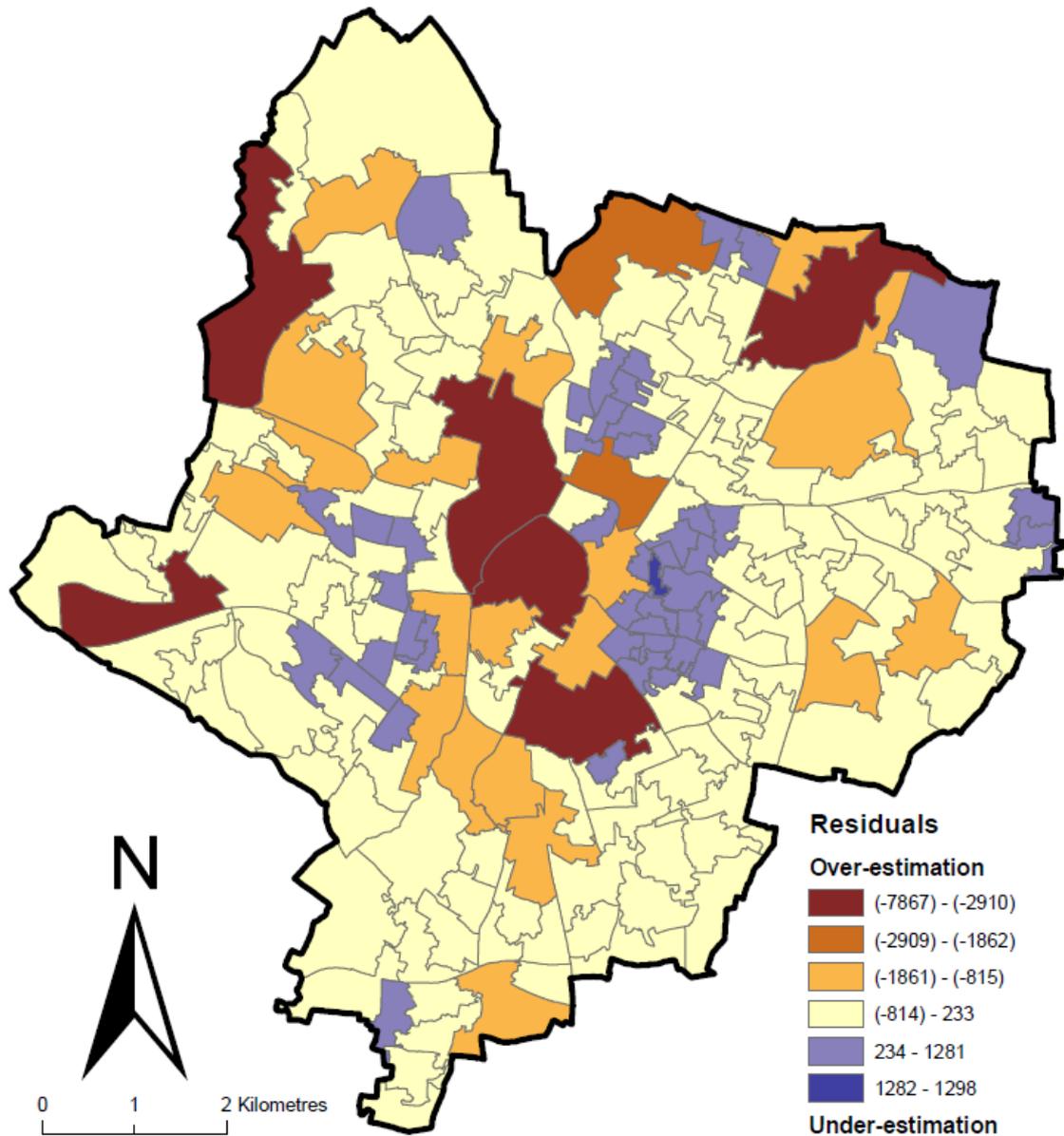


Figure A7.9 - The distribution of residuals using a 30m gridded dasymetric population surface at LSOA using 30m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

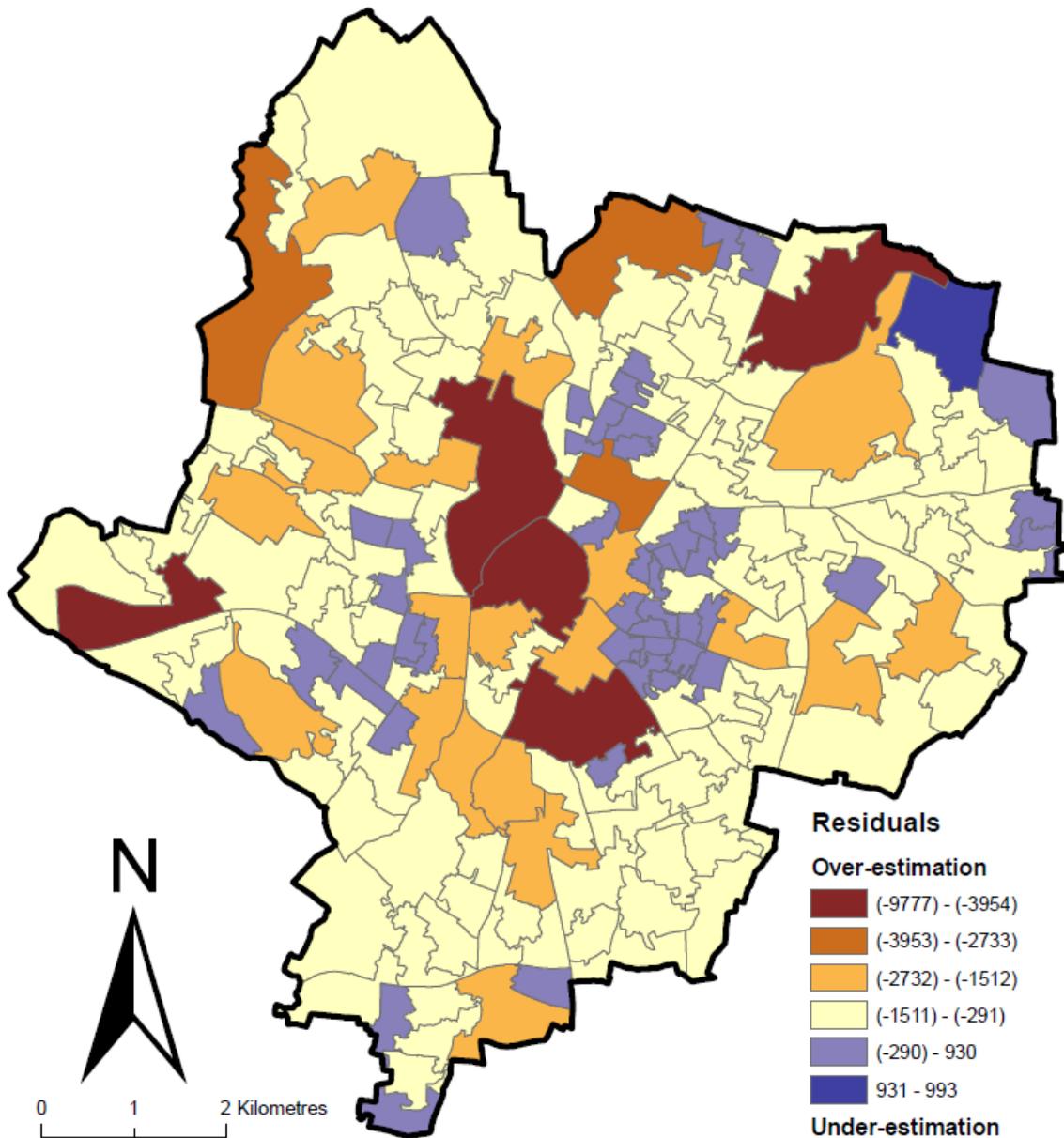


Figure A7.10 - The distribution of residuals using a 100m gridded dasymmetric population surface at LSOA using 30m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

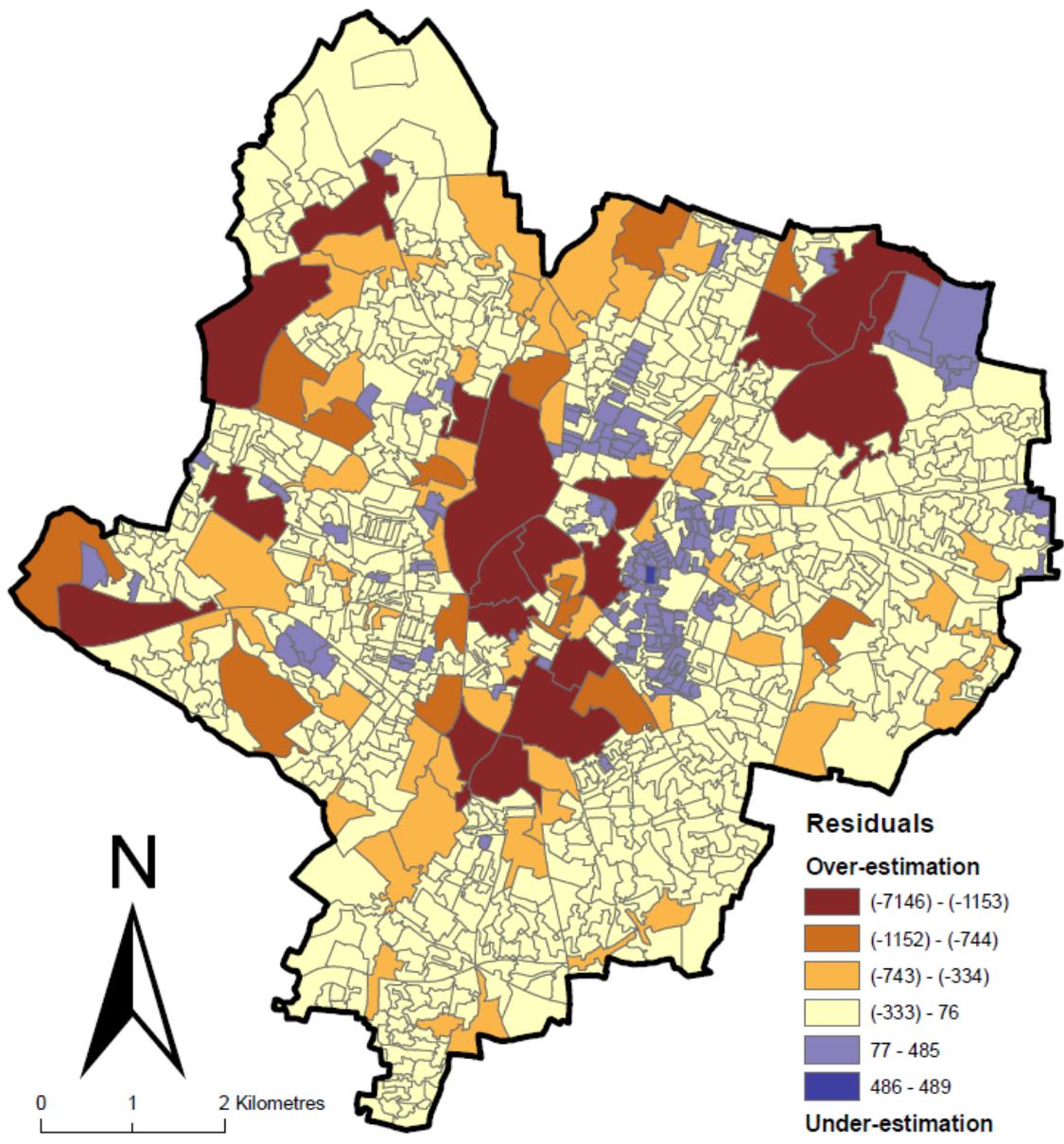


Figure A7.11 - The distribution of residuals using a 30m gridded dasymetric population surface at OA using 30m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

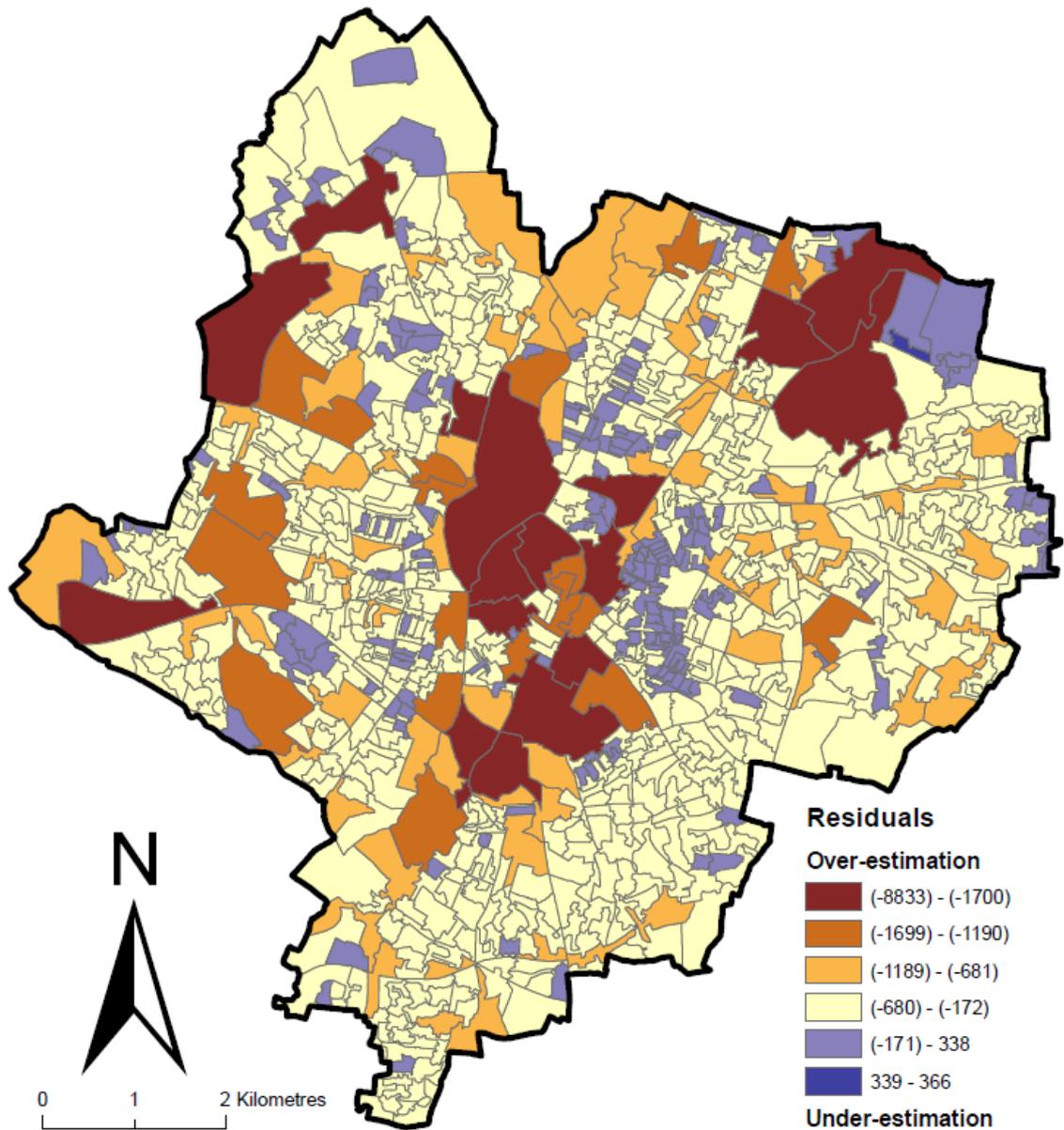


Figure A7.12 - The distribution of residuals using a 100m gridded dasymetric population surface at OA using 30m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

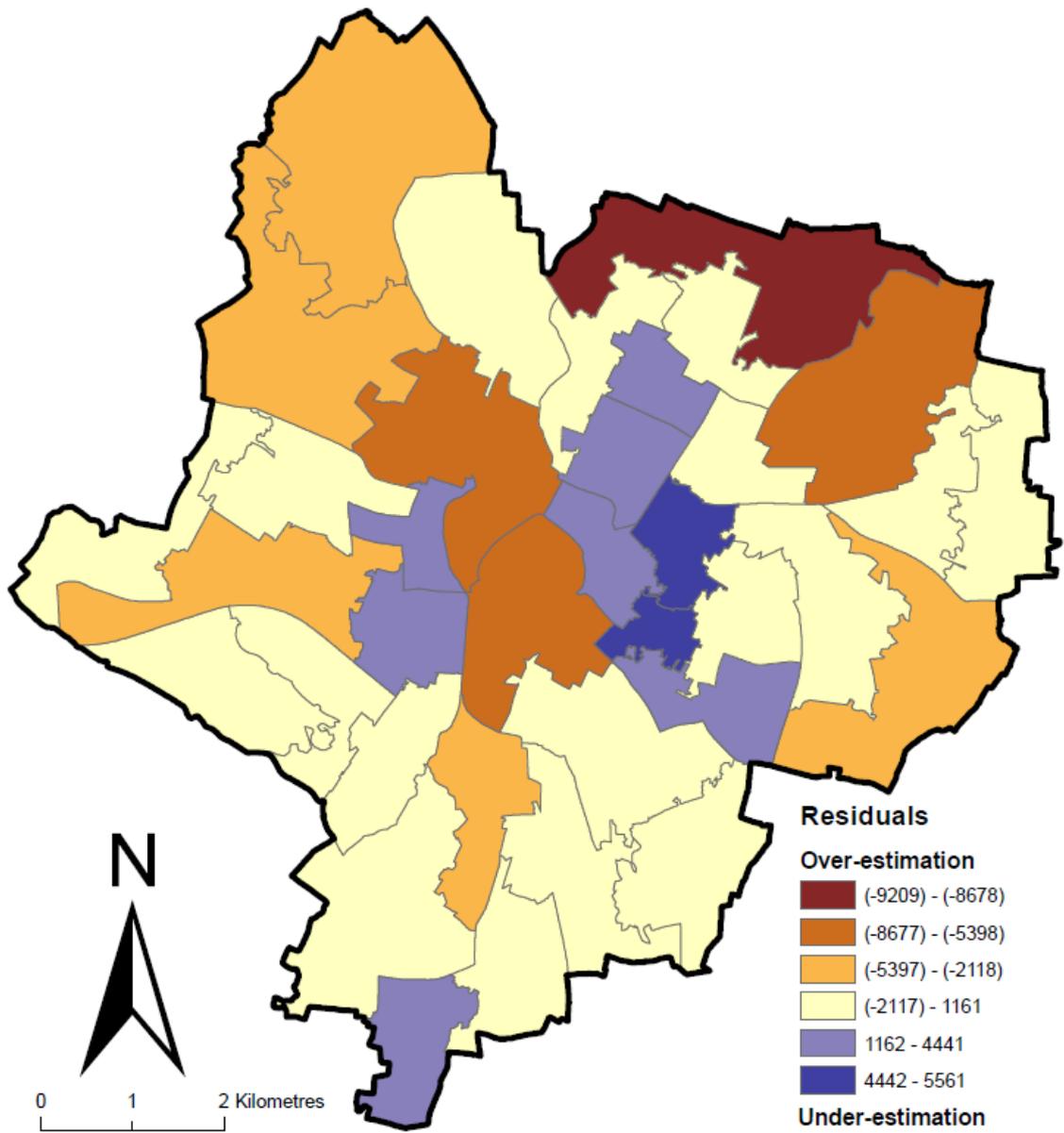


Figure A7.13 - The distribution of residuals using a 30m gridded dasymetric population surface at MSOA using 10m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

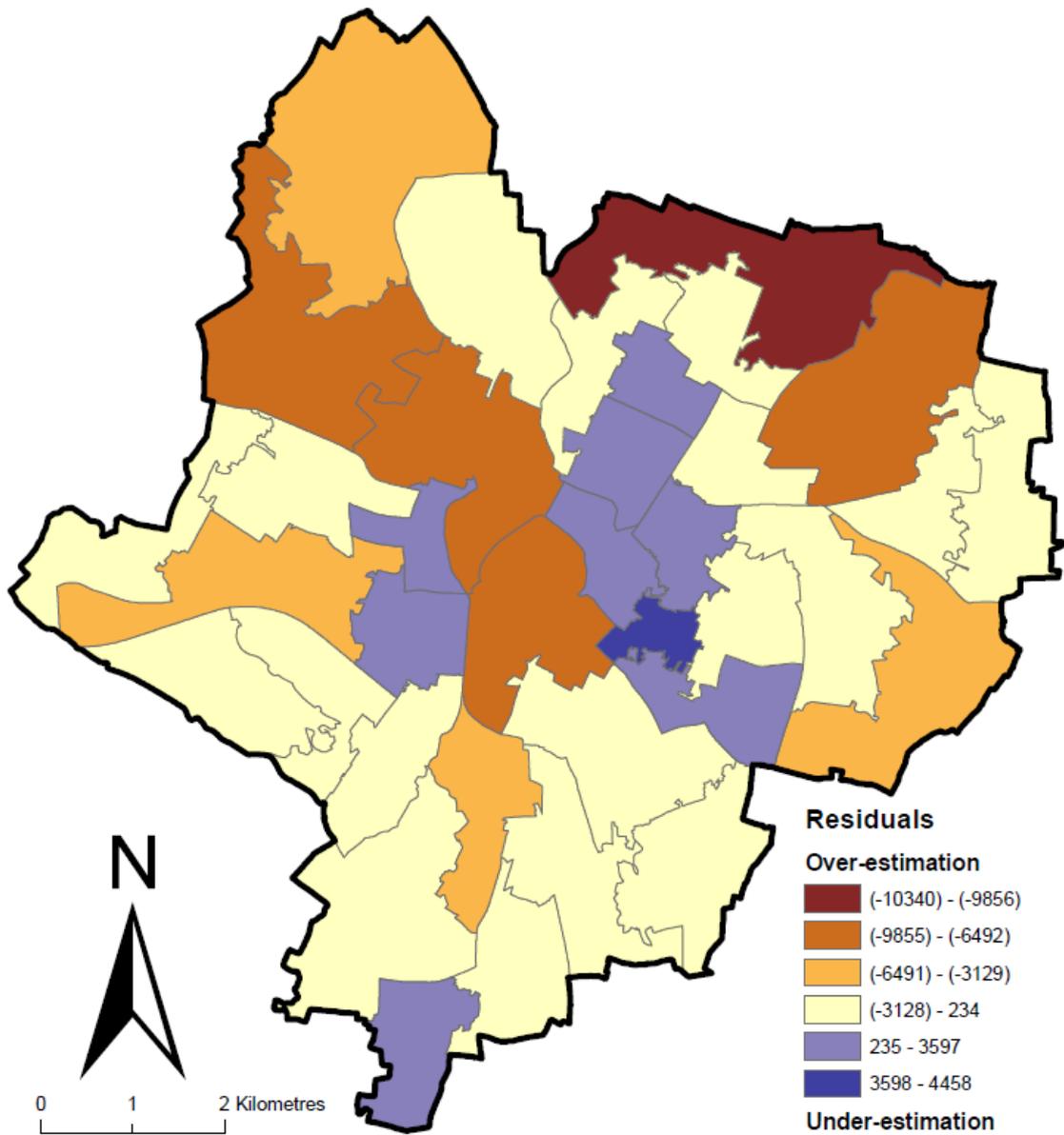


Figure A7.14 - The distribution of residuals using a 100m gridded dasymmetric population surface at MSOA using 10m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

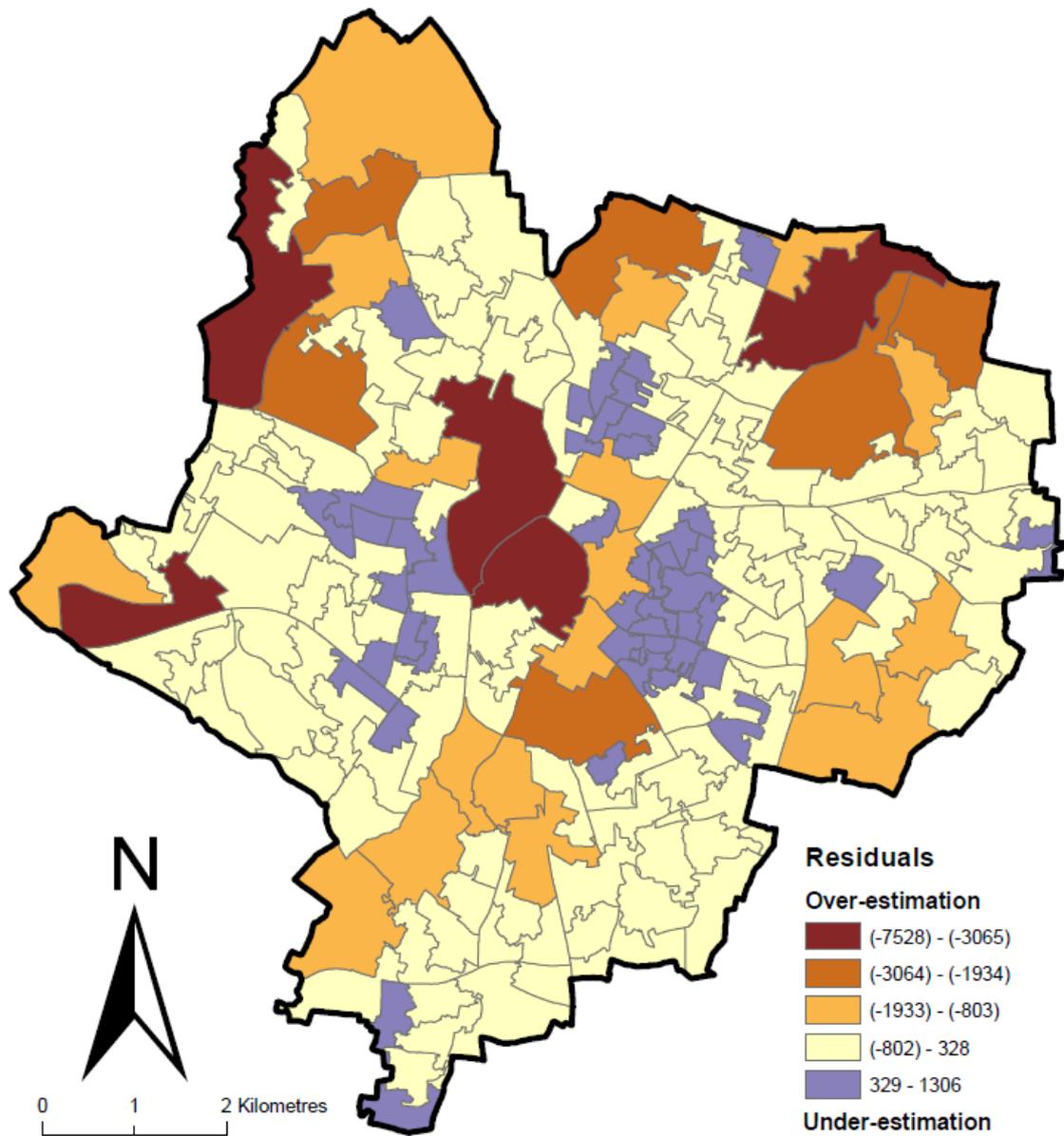


Figure A7.15 - The distribution of residuals using a 30m gridded dasymetric population surface at LSOA using 10m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

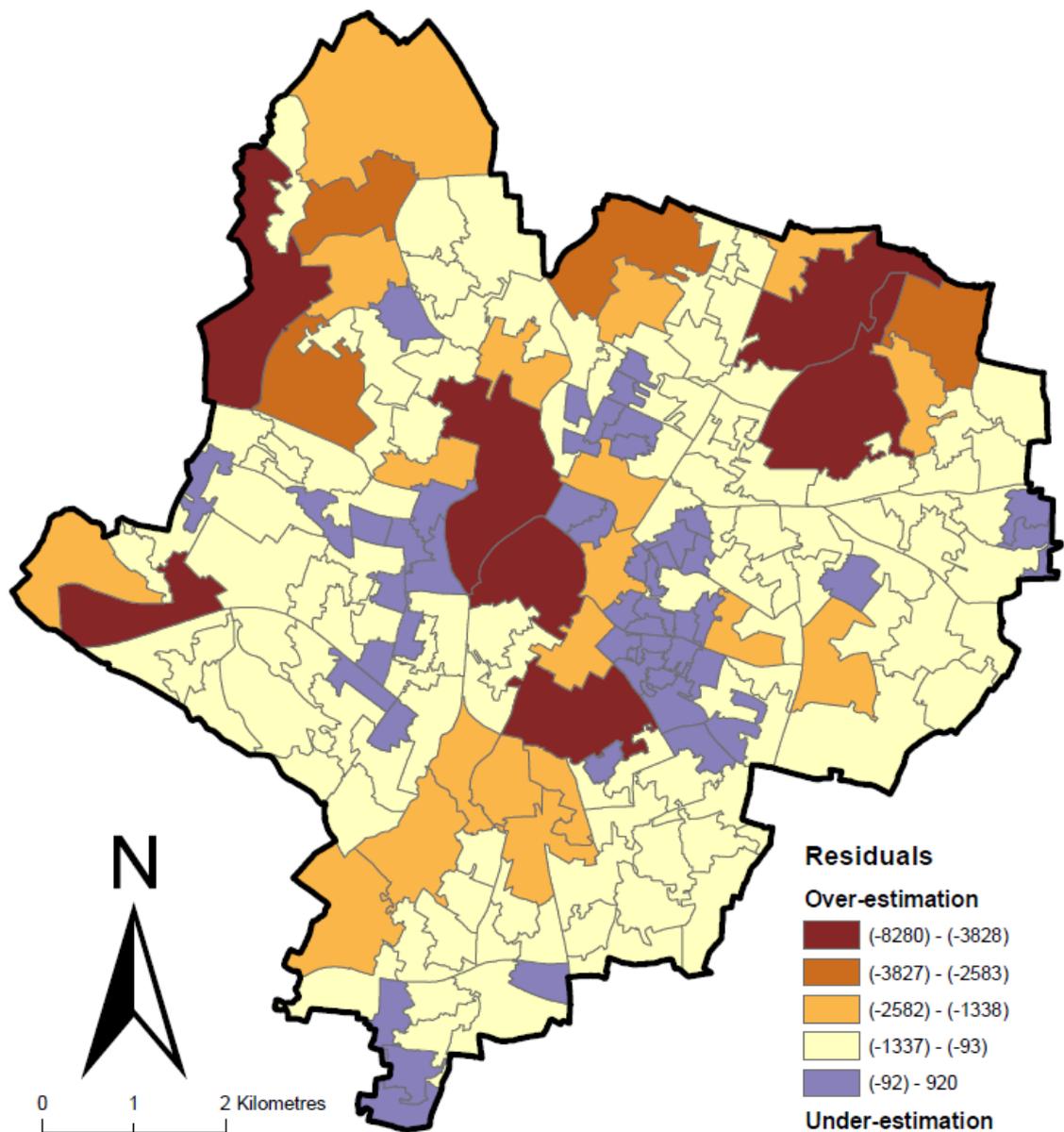


Figure A7.16 - The distribution of residuals using a 100m gridded dasymetric population surface at LSOA using 10m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

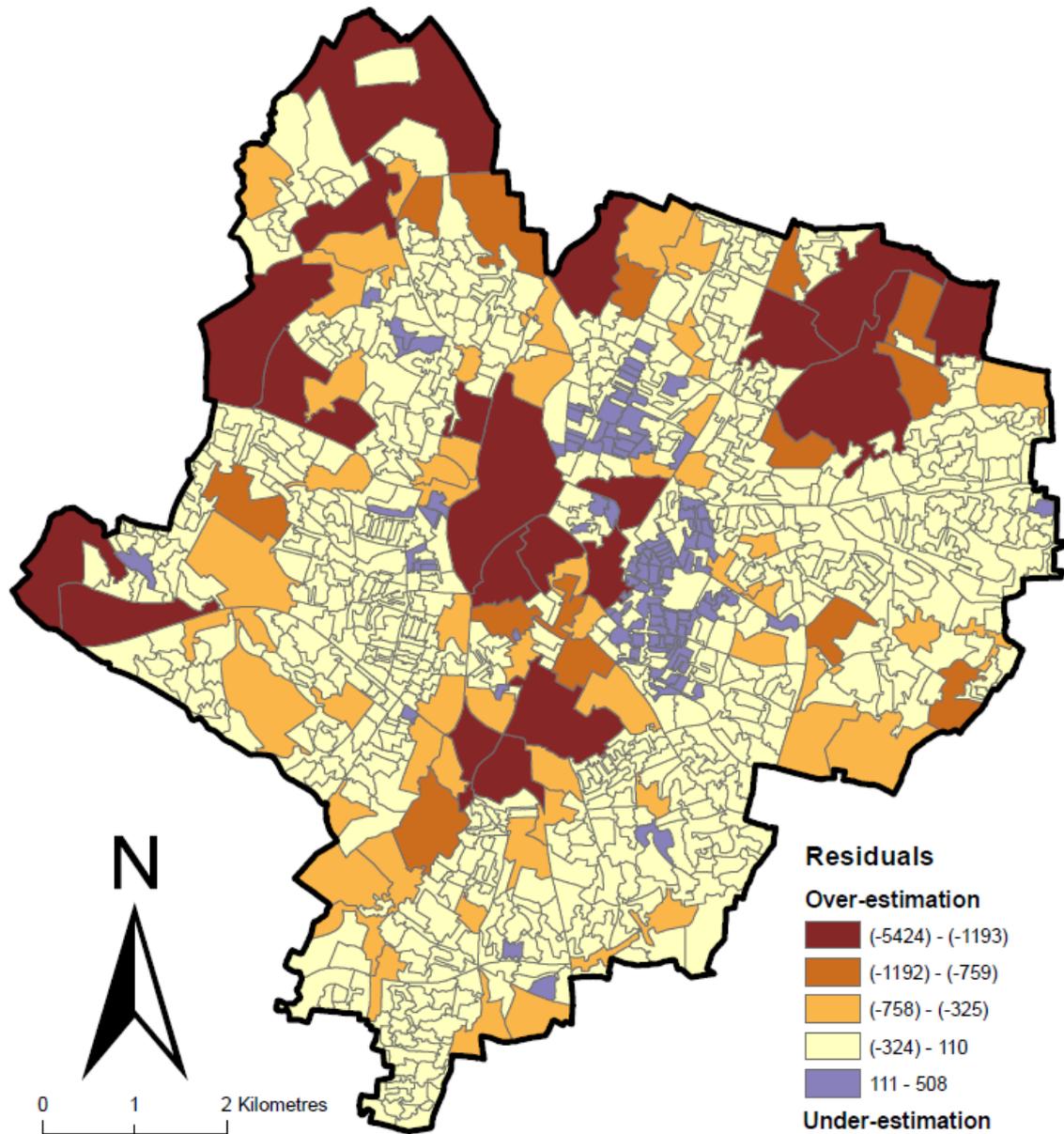


Figure A7.17 - The distribution of residuals using a 30m gridded dasymetric population surface at OA using 10m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

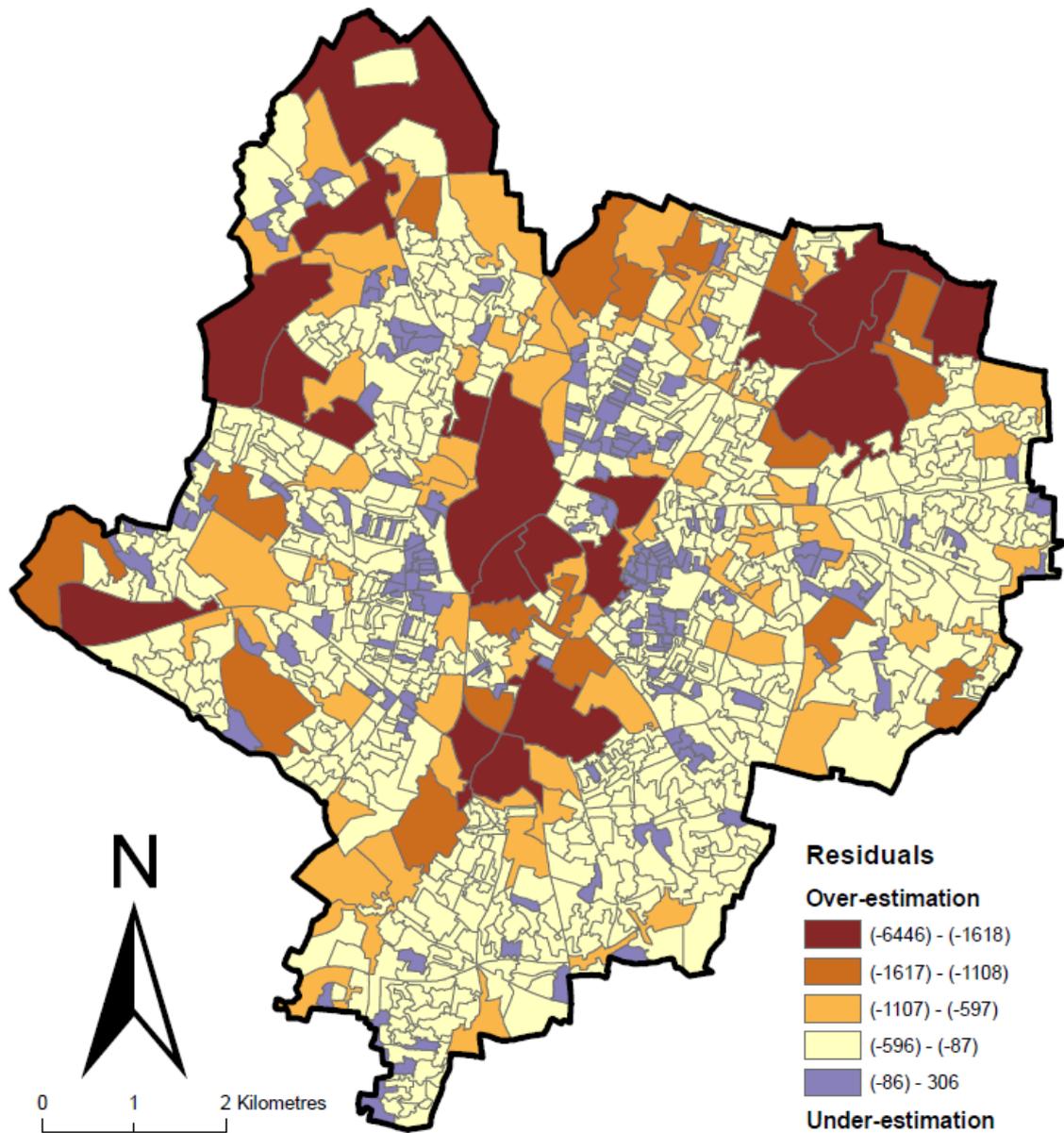


Figure A7.18 - The distribution of residuals using a 100m gridded dasymmetric population surface at OA using 10m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

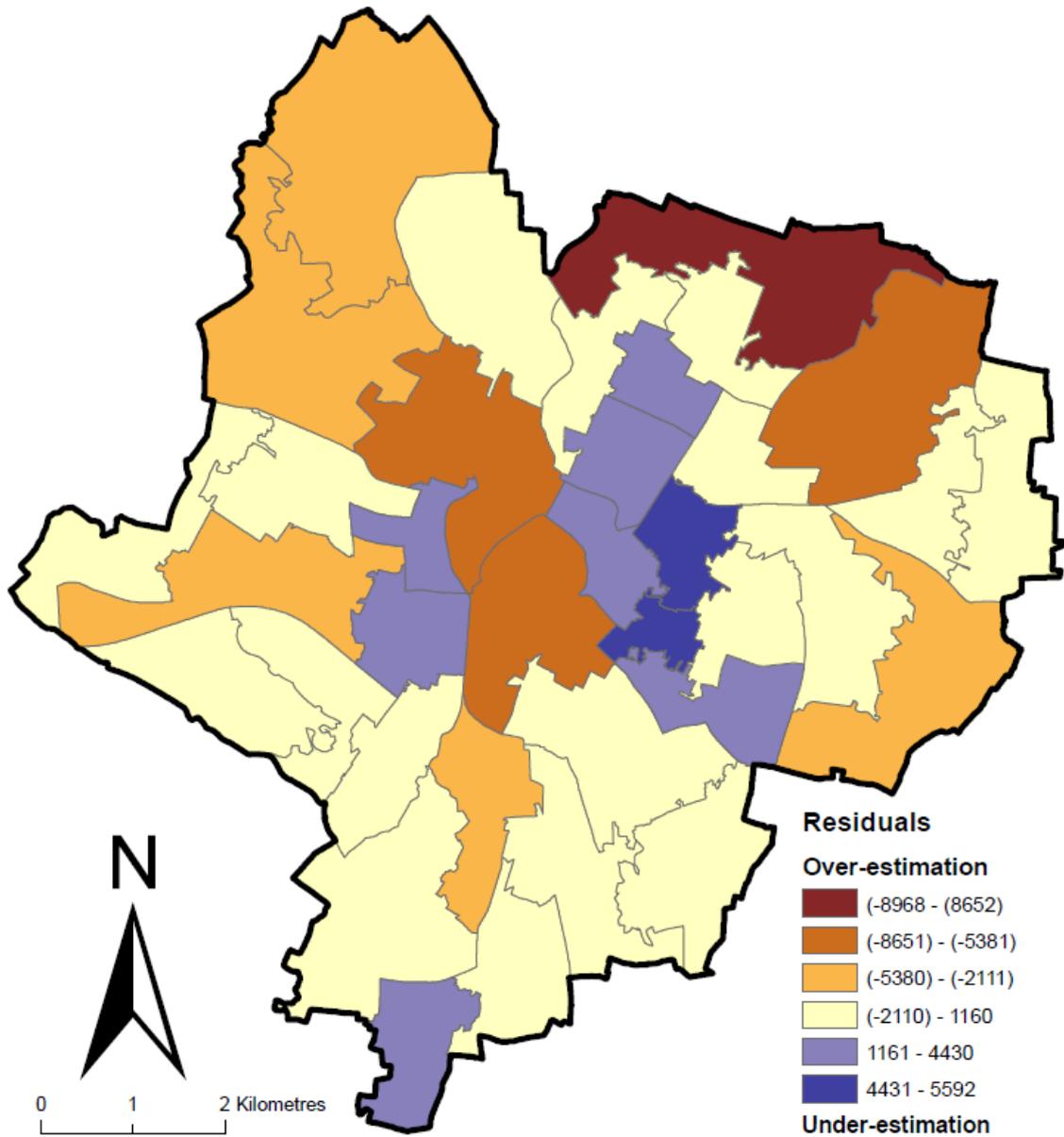


Figure A7.19 - The distribution of residuals using a 30m gridded dasymmetric population surface at MSOA using 3m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

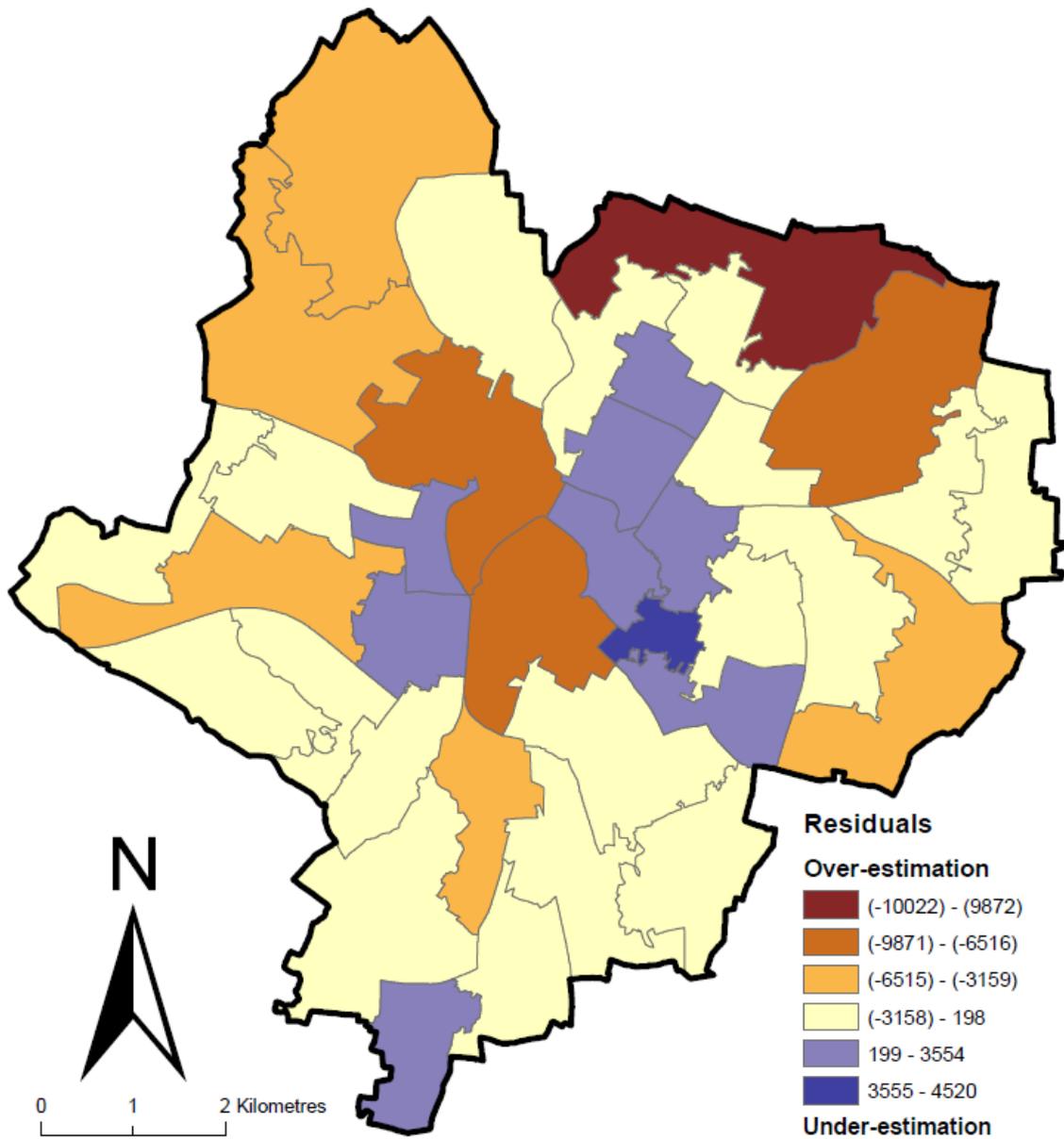


Figure A7.20 - The distribution of residuals using a 100m gridded dasymetric population surface at MSOA using 3m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

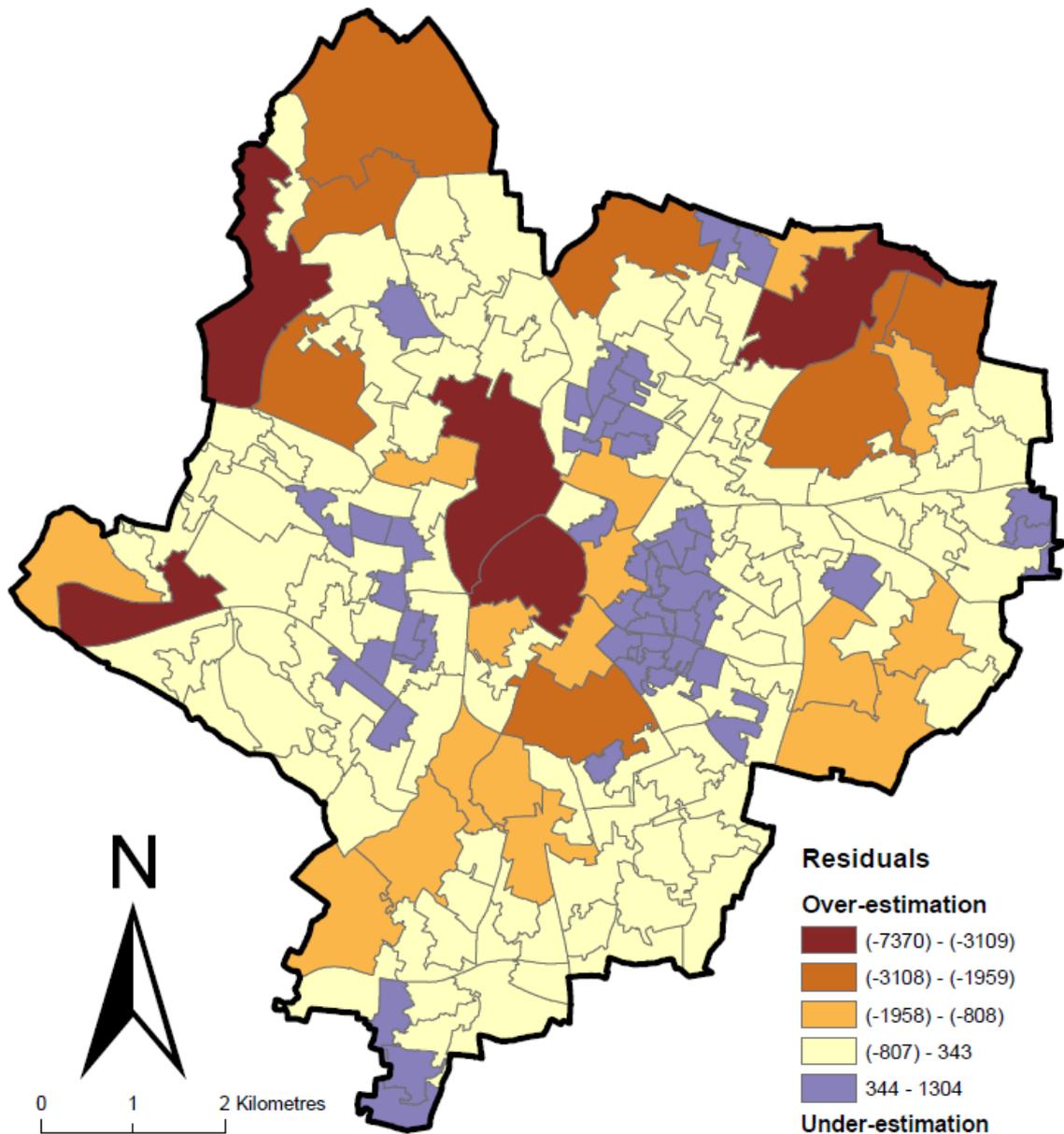


Figure A7.21 - The distribution of residuals using a 30m gridded dasymetric population surface at LSOA using 3m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

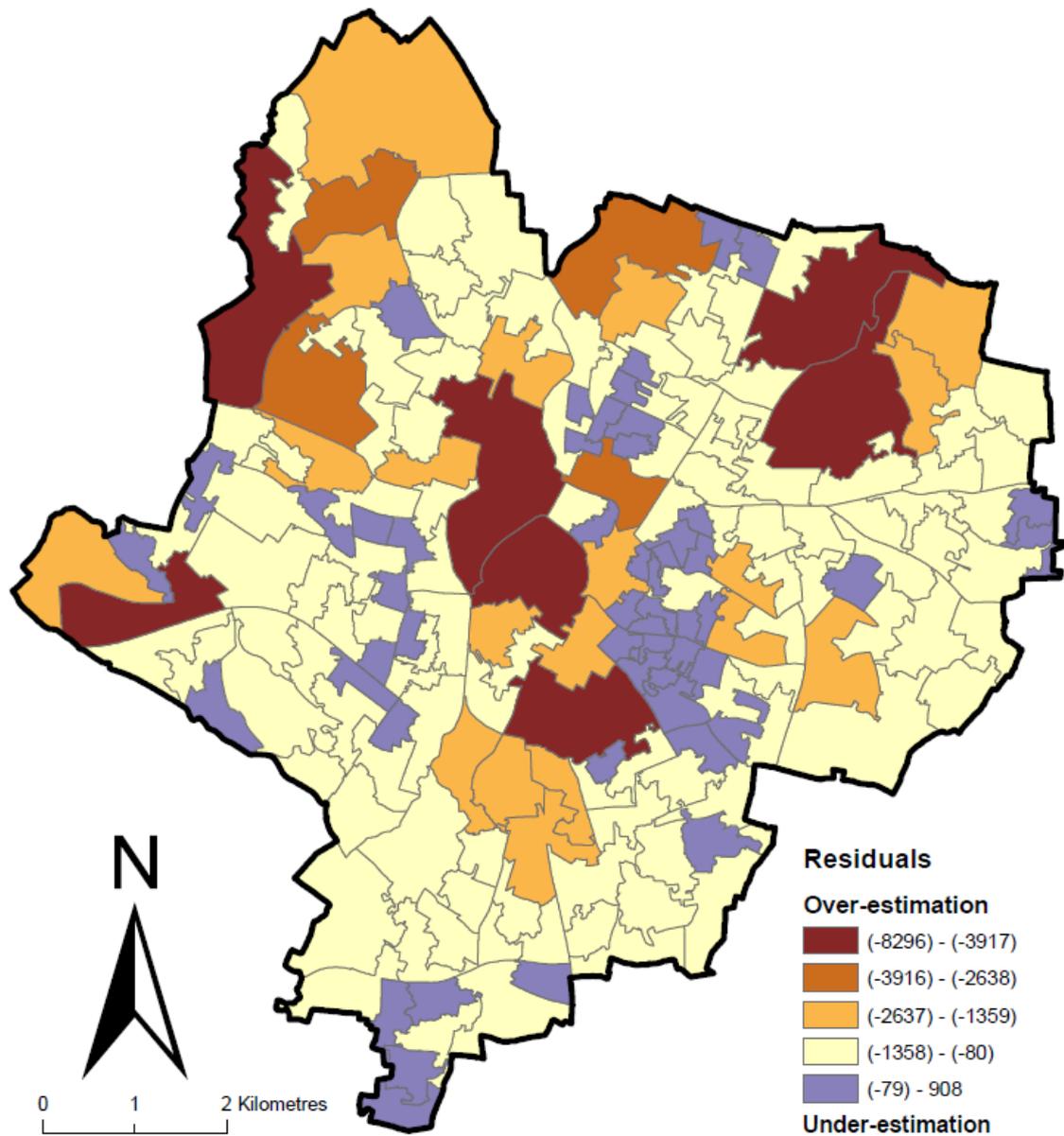


Figure A7.22 - The distribution of residuals using a 100m gridded dasymmetric population surface at LSOA using 3m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

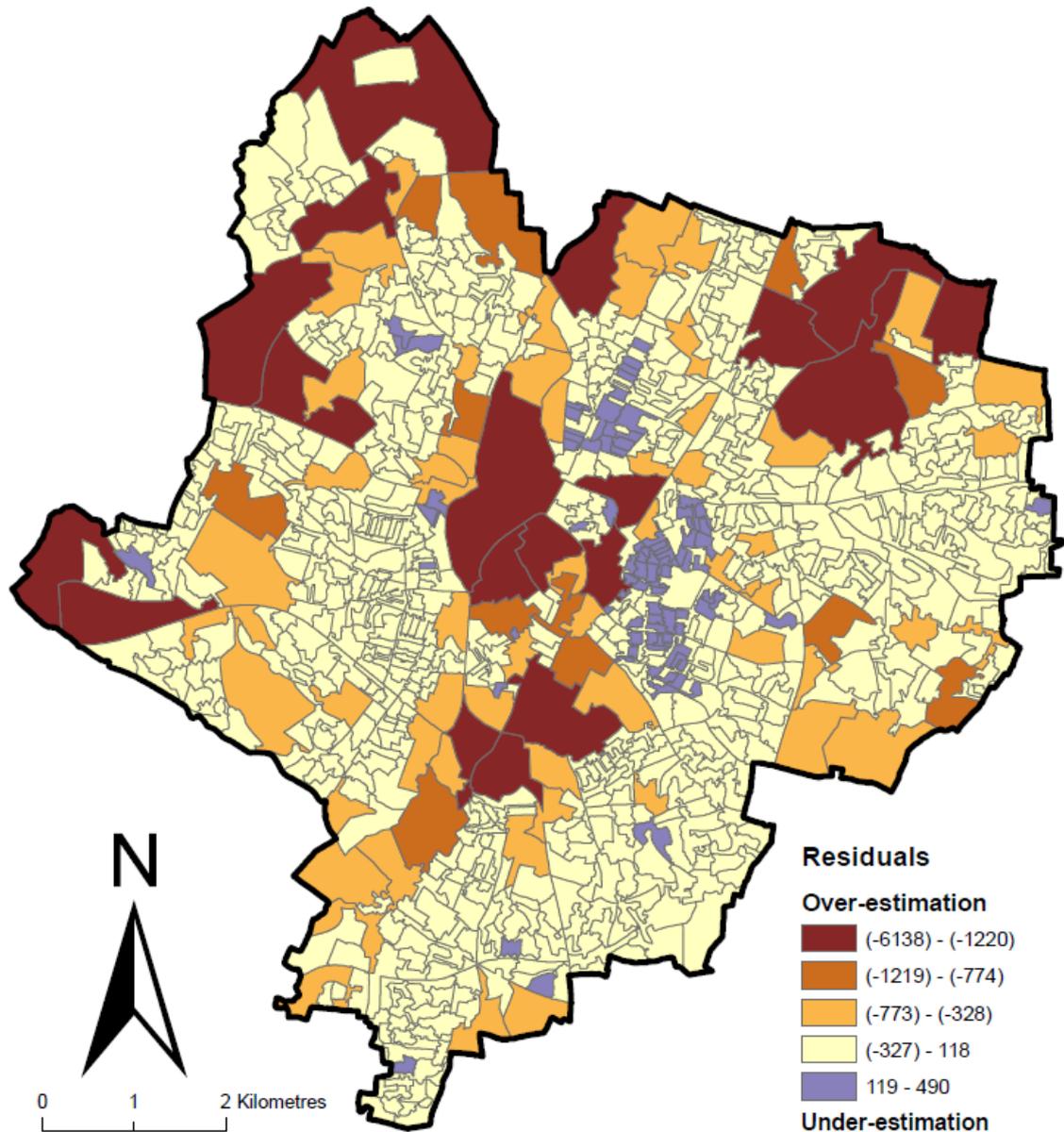


Figure A7.23 - The distribution of residuals using a 30m gridded dasymmetric population surface at OA using 3m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

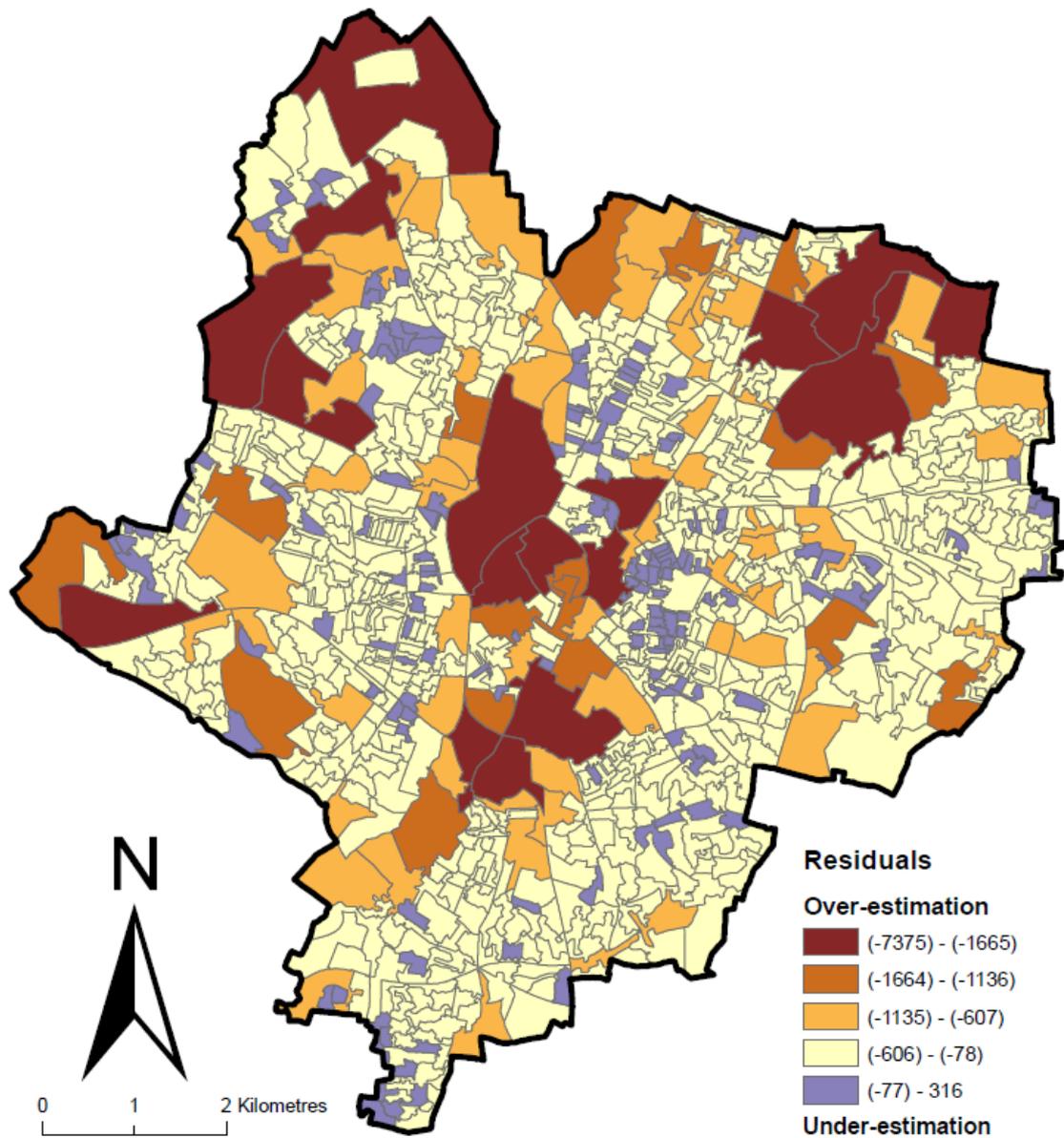


Figure A7.24 - The distribution of residuals using a 100m gridded dasymmetric population surface at OA using 3m spatial resolution land cover data as the ancillary data input. The digital boundaries are © Crown Copyright and/or database right 2013. An Ordnance Survey/EDINA supplied service.

**Appendix 8: Signature Mean Plot, Mean plot and Signature editor for the combined signatures for Port-Harcourt**

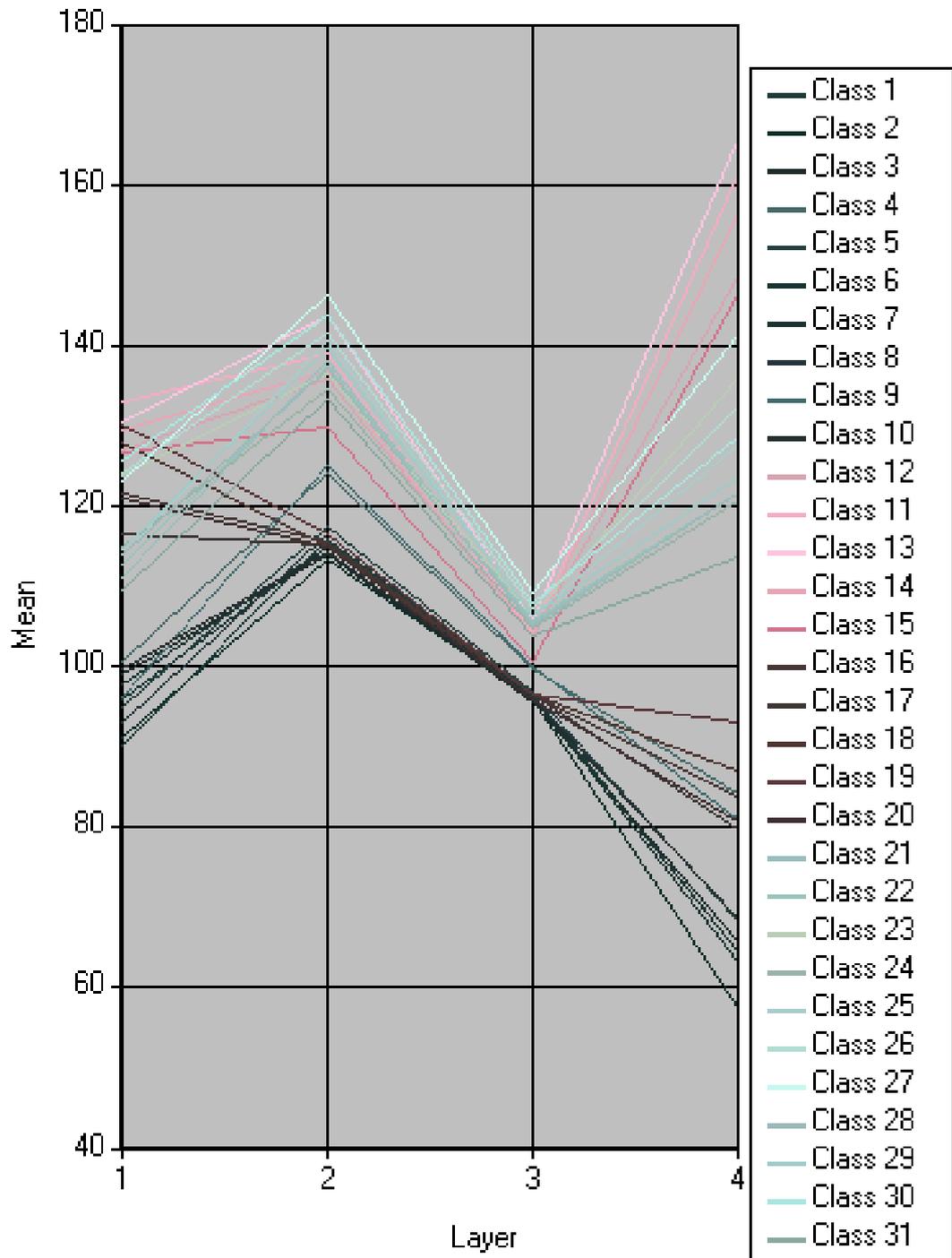


Figure A8.1 - Signature mean plot evaluating signatures for vegetation, built-up and water from Spot5 (colour) 10m spatial resolution image.

Class #	>	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1		Water		0.000	0.000	1.000	32	33	17119	1.000	✓	✓	✓	✓	
2		Vegetation		0.000	1.000	0.000	1	34	962	1.000	✓	✓	✓	✓	
3		Thick vegetation		0.000	0.392	0.000	2	35	2749	1.000	✓	✓	✓	✓	
4	▶	Built-up		0.627	0.322	0.176	3	36	391	1.000	✓	✓	✓	✓	

Figure A8.2 - Signature editor for the combined signatures from Spot5 (colour) 10m spatial resolution image.

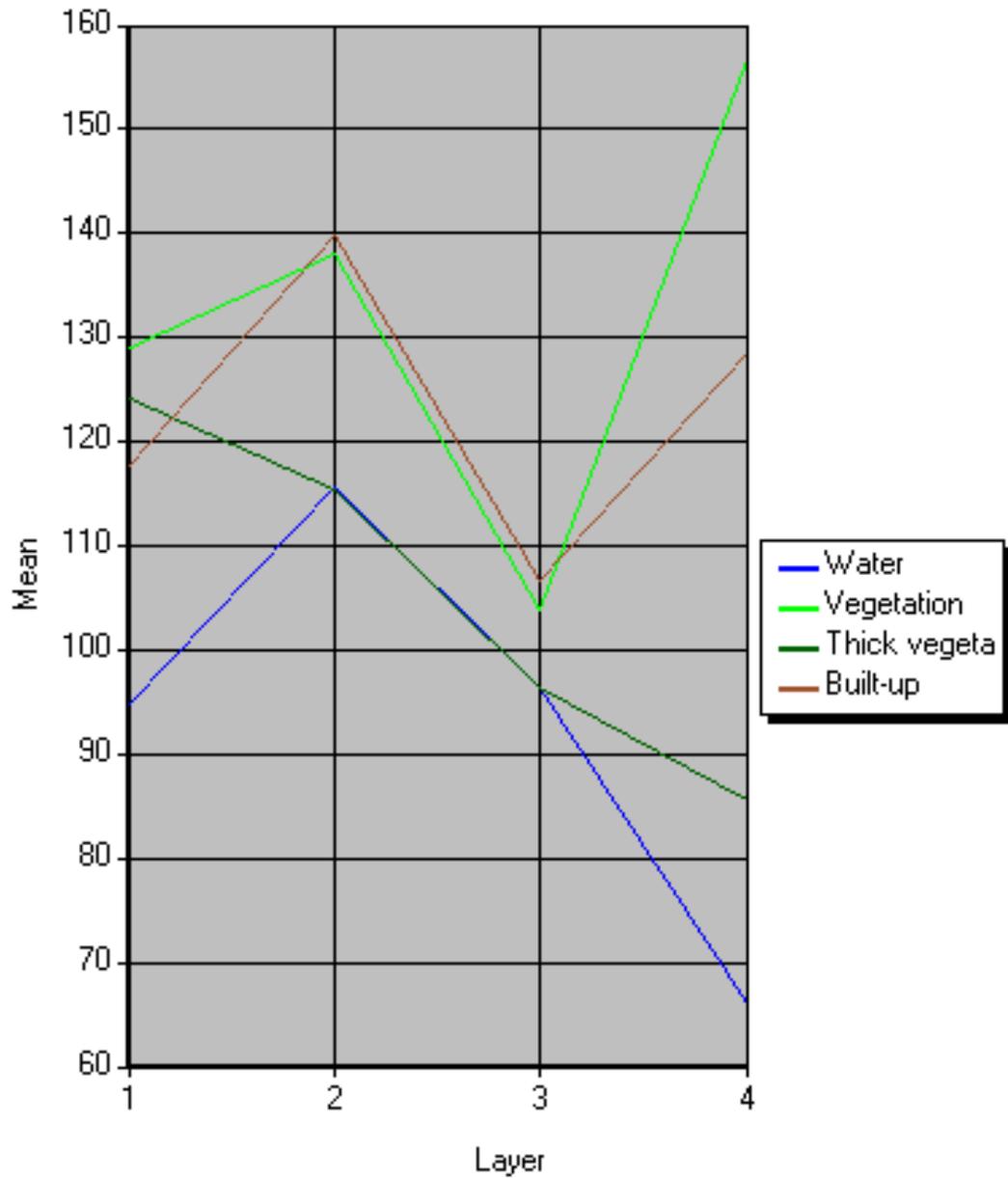


Figure A8.3 - Signature mean plot for the combined signatures from Spot5 (colour) 10m spatial resolution image.

### 3m Image

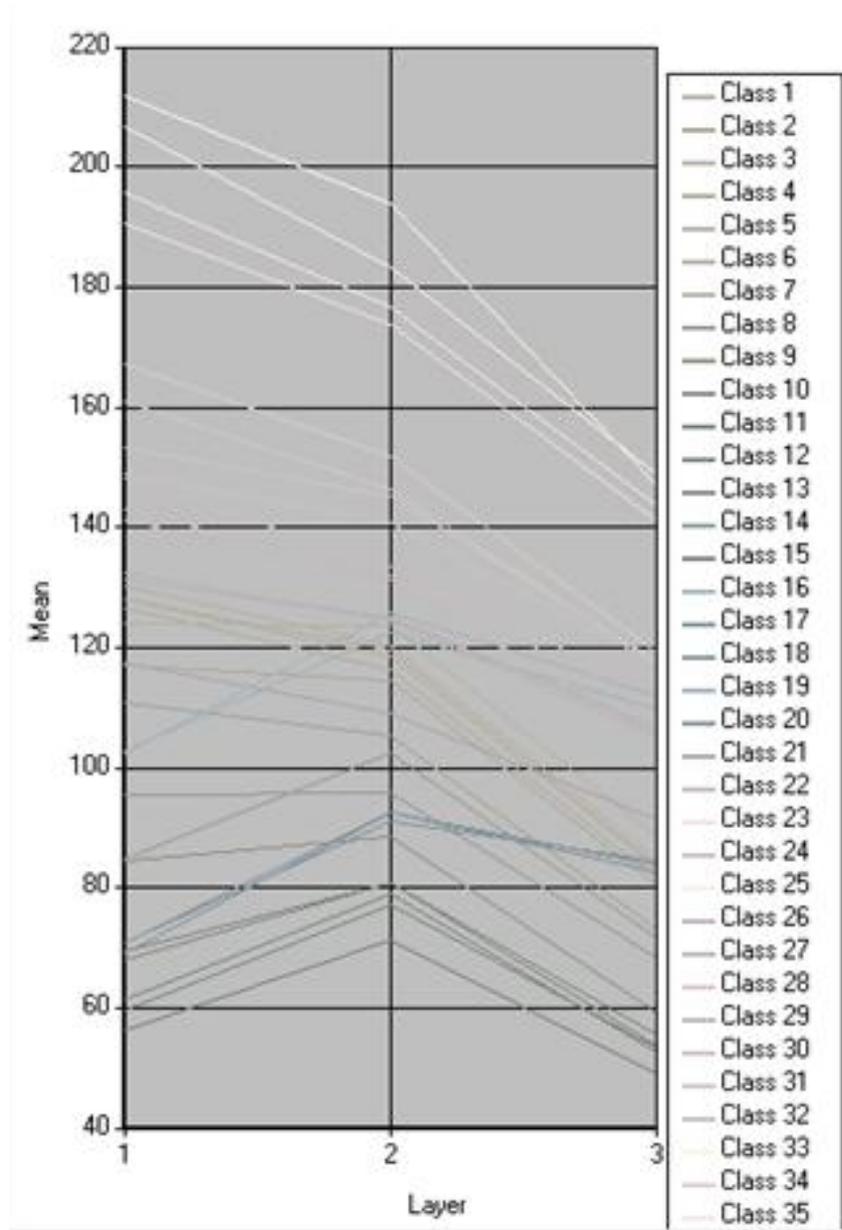


Figure A8.4 - Signature mean plot evaluating signatures for vegetation, thick vegetation, built-up and water from resampled quickbird image of 3m spatial resolution.

Class #	>	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1		Water	Blue	0.000	0.000	1.000	36	36	119322	1.000	✓	✓	✓	✓	
2		Vegetation	Green	0.000	1.000	0.000	1	37	6373	1.000	✓	✓	✓	✓	
3		Thick Vegetation	Dark Green	0.000	0.392	0.000	2	38	38041	1.000	✓	✓	✓	✓	
4	▶	Built-up	Brown	0.627	0.322	0.176	3	39	5441	1.000	✓	✓	✓	✓	

Figure A8.5 - Signature editor for the combined signatures from resampled quickbird image of 3m spatial resolution

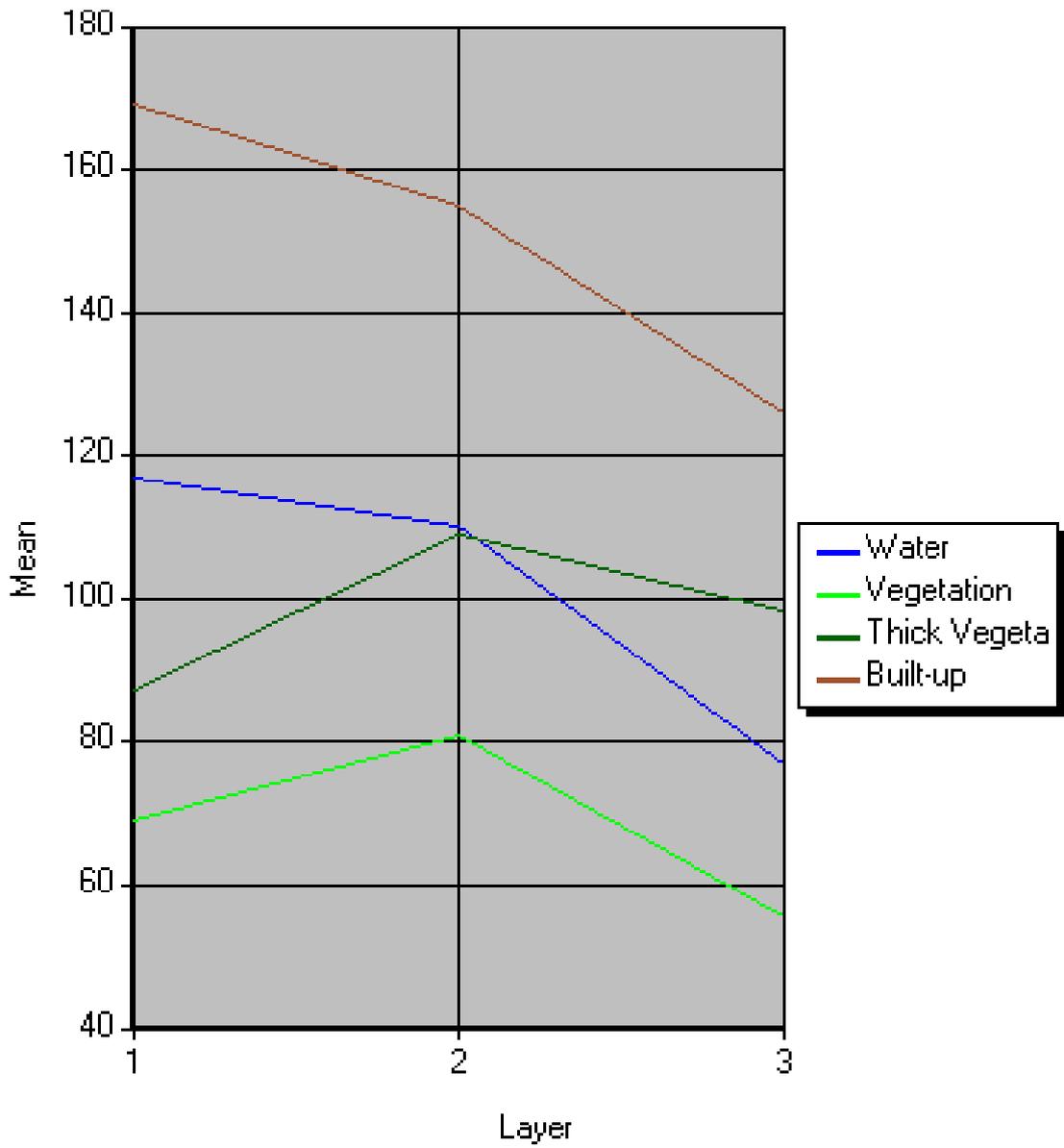


Figure A8.6 - Signature mean plot for the combined signatures from resampled Quickbird image of 3m spatial resolution image.

## Appendix 9: Classification accuracy report for Port-Harcourt

Image File: z:/thesis\_correction/port-harcourt/30m/sup\_ph.img

User Name: ijm14

Date: Mon Dec 22 11:14:10 2014

### ERROR MATRIX

Reference Data					
Classified Data	Unclassified	Water	Built-up	Vegetation	Row Total
Unclassified	112	0	0	0	112
Water	0	15	0	10	25
Built-up	0	2	42	11	55
Vegetation	0	24	0	40	64
Column Total	112	41	42	61	256

----- End of Error Matrix -----

### ACCURACY TOTALS

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Unclassified	112	112	112	---	---
Water	41	25	15	36.59%	60.00%
Built-up	42	55	42	100.00%	76.36%
Vegetation	61	64	40	65.57%	62.50%
Totals	256	256	209		

Overall Classification Accuracy = 81.64%

----- End of Accuracy Totals -----

KAPPA (K<sup>^</sup>) STATISTICS

Overall Kappa Statistics = 0.7372

Conditional Kappa for each Category.

Class Name	Kappa
Unclassified	1
Water	0.5259
Built-up	0.7172
Vegetation	0.5077

----- End of Kappa Statistics -----

**10m**

Image File: z:/thesis\_correction/port-harcourt/10m/sup\_ph10m\_spot.img

User Name : ijm14

Date: Tue Dec 23 13:29:43 2014

ERROR MATRIX

-----

Reference Data

Classified Data	Unclassified	Vegetation	Thick vegetation	Built-up	Water	Row Total
Unclassified	111	0	0	0	0	111
Vegetation	0	18	2	3	5	28
Thick vegetation	0	0	24	0	9	33
Built-up	0	20	0	39	0	59
Water	0	2	12	0	11	25
Column Total	111	40	38	42	25	256

----- End of Error Matrix -----

ACCURACY TOTALS

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Unclassified	111	111	111	---	---
Vegetation	40	28	18	45.00%	64.29%
Thick vegetation	38	33	24	63.16%	72.73%
Built-up	42	59	39	92.86%	66.10%
Water	25	25	11	44.00%	44.00%
Totals	256	256	203		

Overall Classification Accuracy = 79.30%

----- End of Accuracy Totals -----

KAPPA (K<sup>^</sup>) STATISTICS

Overall Kappa Statistics = 0.7158

Conditional Kappa for each Category.

Class Name	Kappa
Unclassified	1
Vegetation	0.5767
Thick vegetation	0.6797
Built-up	0.5945
Water	0.3794

----- End of Kappa Statistics -----

### 3m

Image File: z:/thesis\_correction/port-harcourt/3m/sup\_ph3m.img

User Name : ijm14

Date: Wed Feb 04 22:57:35 2015

#### ERROR MATRIX

-----

#### Reference Data

-----

Classified Data	Unclassified	Vegetation	Thick vegetation	Built-up	Water	Row Total
Unclassified	114	0	0	0	0	114
Vegetation	0	8	0	0	13	21
Thick Vegetation	0	0	19	0	7	26
Built-up	0	15	0	48	0	63
Water	0	17	4	0	11	32
Column Total	114	40	23	48	31	256

----- End of Error Matrix -----

#### ACCURACY TOTALS

-----

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Unclassified	114	114	114	---	---
Vegetation	40	21	8	20.00%	38.10%
Thick Vegetation	23	26	19	82.61%	73.08%
Built-up	48	63	48	100.00%	76.19%
Water	31	32	11	35.48%	34.38%
Totals	256	256	200		

Overall Classification Accuracy = 78.13%

----- End of Accuracy Totals -----

KAPPA (K<sup>^</sup>) STATISTICS

-----

Overall Kappa Statistics = 0.6955

Conditional Kappa for each Category.

-----

Class Name	Kappa
-----	-----
Unclassified	1
Vegetation	0.2663
Thick vegetation	0.7042
Built-up	0.7070
Water	0.2533

----- End of Kappa Statistics -----

## Appendix 10: The results of visual inspection of surfaces

Table A10.1 - Results of the visual inspection of surfaces using Landsat7 (ETM+) 30m spatial resolution image.

FID	CID	X	Y	Surface	G_Earth	Correct
0	0	283,149.56	530,560.14	Populated	Populated	Y
1	0	286,464.28	523,625.92	Unpopulated	UnPopulated	Y
2	0	273,405.58	529,135.30	Unpopulated	Unpopulated	Y
3	0	278,618.83	532,734.37	Populated	Populated	Y
4	0	279,362.93	527,003.66	Populated	Populated	Y
5	0	277,370.41	528,482.54	Unpopulated	UnPopulated	Y
6	0	284,802.84	524,522.67	Unpopulated	UnPopulated	Y
7	0	285,600.83	524,157.77	Unpopulated	Unpopulated	Y
8	0	282,881.97	524,466.88	Populated	Populated	Y
9	0	283,329.30	533,283.51	Populated	Unpopulated	N
10	0	273,742.66	529,822.50	Unpopulated	UnPopulated	Y
11	0	281,681.10	529,470.53	Unpopulated	Unpopulated	Y
12	0	279,056.52	528,302.86	Unpopulated	UnPopulated	Y
13	0	274,806.36	528,113.45	Unpopulated	UnPopulated	Y
14	0	282,381.23	530,988.09	Populated	Populated	Y
15	0	282,294.98	523,045.87	Unpopulated	Unpopulated	Y
16	0	278,164.57	528,464.40	Unpopulated	UnPopulated	Y
17	0	284,254.73	525,006.07	Unpopulated	Unpopulated	Y
18	0	278,506.12	528,005.80	Unpopulated	Populated	N
19	0	274,955.23	526,557.26	Unpopulated	Unpopulated	Y
20	0	284,906.87	532,496.74	Populated	Populated	Y
21	0	279,713.27	523,122.76	Unpopulated	Unpopulated	Y
22	0	284,022.15	530,551.89	Unpopulated	UnPopulated	Y
23	0	278,846.78	529,403.30	Populated	UnPopulated	N
24	0	279,028.35	529,524.92	Populated	Populated	Y
25	0	281,996.97	523,121.61	Unpopulated	Unpopulated	Y
26	0	285,142.92	523,299.22	Unpopulated	Unpopulated	Y
27	0	280,734.30	523,293.55	Unpopulated	Unpopulated	Y
28	0	281,314.67	531,668.75	Populated	Populated	Y
29	0	273,203.85	528,730.62	Unpopulated	Unpopulated	Y
30	0	283,744.02	526,623.07	Unpopulated	Unpopulated	Y
31	0	283,601.93	528,846.17	Populated	UnPopulated	N
32	0	278,415.02	524,689.09	Unpopulated	Unpopulated	Y
33	0	281,725.93	533,060.17	Populated	UnPopulated	N
34	0	281,470.62	522,471.02	Unpopulated	Unpopulated	Y
35	0	282,207.88	530,024.15	Populated	Populated	Y
36	0	274,337.21	528,358.20	Populated	Unpopulated	N
37	0	276,795.11	530,833.04	Populated	Populated	Y

38	0	272,157.33	528,862.64	Unpopulated	Unpopulated	Y
39	0	279,293.97	530,923.60	Populated	Populated	Y
40	0	280,257.74	528,292.28	Populated	UnPopulated	N
41	0	282,121.08	530,378.37	Populated	Populated	Y
42	0	280,437.49	526,944.84	Populated	Populated	Y
43	0	281,694.28	521,866.63	Unpopulated	Unpopulated	Y
44	0	276,818.62	531,893.66	Populated	Populated	Y
45	0	282,624.15	526,891.17	Unpopulated	Unpopulated	Y
46	0	285,606.49	525,407.67	Unpopulated	Unpopulated	Y
47	0	276,926.33	530,420.00	Populated	Populated	Y
48	0	277,745.97	532,403.63	Populated	Populated	Y
49	0	278,686.39	532,162.75	Populated	Populated	Y
50	0	276,083.07	528,221.48	Unpopulated	Populated	N
51	0	279,585.31	524,761.25	Unpopulated	UnPopulated	Y
52	0	281,759.72	532,173.59	Populated	Populated	Y
53	0	284,754.68	524,064.75	Unpopulated	Unpopulated	Y
54	0	277,498.94	531,285.58	Populated	Populated	Y
55	0	285,573.69	522,823.64	Unpopulated	Unpopulated	Y
56	0	283,217.28	532,100.73	Populated	Populated	Y
57	0	273,471.69	529,985.80	Unpopulated	Unpopulated	Y
58	0	284,544.50	533,241.52	Populated	Populated	Y
59	0	286,720.39	523,839.62	Unpopulated	Unpopulated	Y
60	0	284,925.70	526,867.65	Unpopulated	Unpopulated	Y
61	0	280,876.57	532,755.57	Populated	Populated	Y
62	0	279,564.76	526,641.63	Populated	Populated	Y
63	0	274,983.25	529,627.02	Populated	Populated	Y
64	0	284,168.69	527,247.58	Populated	Populated	Y
65	0	281,346.24	524,198.03	Populated	Populated	Y
66	0	279,070.61	532,584.64	Populated	Populated	Y
67	0	281,090.80	530,793.83	Populated	Unpopulated	N
68	0	283,691.18	531,361.69	Populated	Populated	Y
69	0	274,706.68	526,581.04	Unpopulated	UnPopulated	Y
70	0	280,031.24	527,239.97	Populated	Populated	Y
71	0	283,575.70	530,907.81	Unpopulated	UnPopulated	Y
72	0	274,525.05	527,251.28	Unpopulated	Unpopulated	Y
73	0	282,952.10	525,966.30	Unpopulated	Unpopulated	Y
74	0	283,443.71	524,724.23	Populated	Populated	Y
75	0	283,964.34	524,902.77	Unpopulated	Unpopulated	Y
76	0	281,316.51	524,943.71	Populated	Unpopulated	N
77	0	284,797.19	530,551.40	Unpopulated	UnPopulated	Y
78	0	282,343.51	525,177.46	Populated	Populated	Y
79	0	281,736.44	532,474.06	Populated	Populated	Y
80	0	282,480.89	530,263.55	Unpopulated	UnPopulated	Y
81	0	276,131.16	526,690.60	Unpopulated	Unpopulated	Y
82	0	279,775.25	530,839.60	Unpopulated	Populated	N
83	0	277,865.36	529,307.21	Populated	Populated	Y

84	0	285,436.41	523,653.52	Unpopulated	Unpopulated	Y
85	0	282,026.29	531,217.99	Populated	Populated	Y
86	0	278,782.10	529,173.91	Populated	UnPopulated	N
87	0	285,819.34	524,591.15	Unpopulated	Unpopulated	Y
88	0	286,051.74	524,403.25	Unpopulated	Unpopulated	Y
89	0	281,265.54	529,889.59	Populated	Populated	Y
90	0	283,461.74	527,721.69	Unpopulated	UnPopulated	Y
91	0	273,444.33	528,056.29	Unpopulated	Unpopulated	Y
92	0	281,832.84	532,796.82	Populated	Populated	Y
93	0	280,759.27	531,862.96	Populated	UnPopulated	N
94	0	281,332.64	528,987.62	Unpopulated	Unpopulated	Y
95	0	283,292.61	525,172.47	Populated	Unpopulated	N
96	0	284,053.71	531,873.62	Populated	Populated	Y
97	0	283,244.97	533,060.96	Unpopulated	Unpopulated	Y
98	0	285,470.54	527,463.12	Unpopulated	Unpopulated	Y
99	0	277,652.20	529,018.16	Populated	Populated	Y
100	0	279,190.54	530,462.89	Populated	Populated	Y
101	0	285,798.88	525,164.62	Unpopulated	Unpopulated	Y
102	0	276,797.69	528,219.90	Unpopulated	UnPopulated	Y
103	0	285,192.60	531,830.31	Populated	Populated	Y
104	0	277,844.08	528,614.52	Unpopulated	UnPopulated	Y
105	0	280,442.81	528,425.69	Populated	Populated	Y
106	0	284,840.71	525,148.95	Unpopulated	Unpopulated	Y
107	0	285,089.70	529,699.28	Unpopulated	UnPopulated	Y
108	0	273,707.53	528,962.53	Unpopulated	Unpopulated	Y
109	0	285,145.82	531,341.88	Unpopulated	UnPopulated	Y
110	0	276,343.56	527,600.70	Unpopulated	UnPopulated	Y
111	0	279,068.79	530,073.21	Populated	Populated	Y
112	0	281,888.72	530,818.02	Populated	Populated	Y
113	0	272,634.00	528,500.06	Unpopulated	Unpopulated	Y
114	0	278,543.25	525,495.17	Populated	UnPopulated	N
115	0	282,429.86	526,103.40	Populated	Populated	Y
116	0	280,343.88	531,952.26	Populated	Populated	Y
117	0	282,204.06	528,923.45	Populated	Populated	Y
118	0	285,631.90	528,869.09	Unpopulated	UnPopulated	Y
119	0	280,682.70	532,318.51	Populated	Populated	Y
120	0	281,556.00	526,242.37	Populated	Populated	Y
121	0	282,064.26	527,566.55	Unpopulated	Unpopulated	Y
122	0	284,415.74	532,746.51	Populated	Populated	Y
123	0	275,326.44	529,415.68	Unpopulated	Unpopulated	Y
124	0	280,884.07	530,644.59	Populated	UnPopulated	N
125	0	279,127.94	528,882.88	Unpopulated	Populated	N
126	0	286,571.60	524,226.15	Unpopulated	Unpopulated	Y
127	0	283,528.87	528,395.24	Unpopulated	UnPopulated	Y
128	0	274,767.17	530,077.71	Unpopulated	UnPopulated	Y
129	0	285,800.12	527,957.73	Unpopulated	Unpopulated	Y

130	0	279,632.67	532,385.64	Populated	Populated	Y
131	0	283,176.15	527,965.19	Unpopulated	Populated	N
132	0	286,023.01	532,386.32	Unpopulated	Unpopulated	Y
133	0	279,629.54	533,104.57	Unpopulated	Populated	N
134	0	274,352.69	529,968.44	Unpopulated	Populated	N
135	0	274,103.03	527,897.39	Unpopulated	Unpopulated	Y
136	0	285,988.57	524,706.02	Unpopulated	Unpopulated	Y
137	0	281,131.46	522,731.21	Unpopulated	Unpopulated	Y
138	0	281,204.78	527,757.77	Populated	Populated	Y
139	0	286,163.20	523,481.19	Unpopulated	Unpopulated	Y
140	0	282,035.39	524,294.66	Populated	UnPopulated	N
141	0	275,132.20	526,408.41	Unpopulated	Unpopulated	Y
142	0	282,522.14	523,794.31	Unpopulated	Unpopulated	Y
143	0	284,639.40	526,173.91	Unpopulated	Unpopulated	Y
144	0	279,568.26	528,169.59	Populated	Populated	Y
145	0	280,045.18	523,520.84	Unpopulated	Unpopulated	Y
146	0	285,165.41	530,404.06	Unpopulated	UnPopulated	Y
147	0	275,283.09	529,843.06	Unpopulated	Unpopulated	Y
148	0	279,452.92	528,450.61	Populated	Populated	Y
149	0	283,094.45	524,441.92	Populated	Populated	Y
150	0	284,972.67	522,863.67	Unpopulated	Unpopulated	Y
151	0	285,324.75	526,080.55	Unpopulated	Unpopulated	Y
152	0	282,879.87	524,737.06	Populated	Populated	Y
153	0	282,129.55	526,285.60	Populated	Populated	Y
154	0	285,030.94	526,182.60	Unpopulated	Unpopulated	Y
155	0	279,050.42	530,791.81	Populated	Populated	Y
156	0	278,448.99	531,586.12	Populated	Populated	Y
157	0	284,584.12	531,838.38	Populated	UnPopulated	N
158	0	280,599.57	527,348.84	Populated	Populated	Y
159	0	285,609.66	532,073.26	Unpopulated	Populated	N
160	0	284,120.96	528,074.30	Unpopulated	Populated	N
161	0	285,194.64	532,244.08	Populated	Populated	Y
162	0	284,443.81	524,222.99	Unpopulated	Unpopulated	Y
163	0	279,599.78	531,992.21	Unpopulated	UnPopulated	Y
164	0	277,840.14	528,093.35	Unpopulated	UnPopulated	Y
165	0	286,232.43	531,540.77	Unpopulated	Populated	N
166	0	279,939.69	523,146.44	Unpopulated	Unpopulated	Y
167	0	284,332.68	526,876.95	Unpopulated	Unpopulated	Y
168	0	277,604.06	527,644.64	Unpopulated	Unpopulated	Y
169	0	281,977.16	523,939.78	Populated	Populated	Y
170	0	272,869.86	528,083.78	Unpopulated	Unpopulated	Y
171	0	280,790.42	522,654.81	Unpopulated	Unpopulated	Y
172	0	279,190.85	531,598.47	Populated	Populated	Y
173	0	275,533.06	528,835.04	Populated	Populated	Y
174	0	273,435.22	528,922.12	Unpopulated	Unpopulated	Y
175	0	283,849.25	532,606.71	Populated	Populated	Y

176	0	283,495.30	532,714.44	Unpopulated	Unpopulated	Y
177	0	278,428.76	530,338.66	Populated	Populated	Y
178	0	284,948.11	528,538.47	Populated	Populated	Y
179	0	279,593.50	531,341.02	Populated	Populated	Y
180	0	283,400.52	524,439.14	Populated	Populated	Y
181	0	277,752.84	532,779.95	Populated	Populated	Y
182	0	276,783.17	529,317.02	Unpopulated	Unpopulated	Y
183	0	272,855.24	530,063.37	Unpopulated	UnPopulated	Y
184	0	274,341.36	529,001.68	Unpopulated	UnPopulated	Y
185	0	282,392.27	526,431.46	Populated	Populated	Y
186	0	282,150.38	524,964.24	Populated	Populated	Y
187	0	276,602.78	532,842.30	Populated	Populated	Y
188	0	285,200.87	526,777.89	Unpopulated	Unpopulated	Y
189	0	286,354.12	531,985.38	Unpopulated	Unpopulated	Y
190	0	273,202.72	528,529.29	Unpopulated	Unpopulated	Y
191	0	278,499.33	531,096.22	Populated	Populated	Y
192	0	276,099.51	526,968.40	Unpopulated	Unpopulated	Y
193	0	275,944.85	526,532.71	Unpopulated	Unpopulated	Y
194	0	281,582.88	532,756.06	Populated	Unpopulated	N
195	0	281,687.91	530,949.80	Populated	Populated	Y
196	0	276,351.45	528,477.48	Populated	Populated	Y
197	0	285,621.93	523,057.96	Unpopulated	Unpopulated	Y
198	0	276,831.06	527,973.53	Unpopulated	Unpopulated	Y
199	0	275,939.60	527,376.61	Unpopulated	Unpopulated	Y

Populated surface on google earth 16/200 = 8%

Unpopulated surface on google earth 10/200 = 5%

Total 26/200 = 13%

Table A10.2 - Results of the visual inspection of surfaces using Spot5 (colour) 10m spatial resolution image.

FID	CID	X	Y	Surface	G_Earth	Correct
0	0	279,247.79	526,153.52	Populated	Populated	Y
1	0	281,799.19	523,814.76	Populated	unPopulated	N
2	0	278,899.36	524,692.87	Unpopulated	UnPopulated	Y
3	0	286,600.56	531,784.78	Unpopulated	Unpopulated	Y
4	0	278,629.61	528,178.54	Populated	Populated	Y
5	0	273,094.83	530,165.86	Unpopulated	Unpopulated	Y
6	0	281,292.25	530,353.47	Populated	UnPopulated	N
7	0	273,777.16	527,964.22	Unpopulated	Unpopulated	Y
8	0	285,298.90	524,827.69	Unpopulated	Unpopulated	Y
9	0	280,556.27	526,928.00	Populated	Populated	Y
10	0	285,418.86	531,570.99	Unpopulated	Unpopulated	Y
11	0	284,337.04	525,672.53	Unpopulated	Unpopulated	Y
12	0	280,033.99	527,764.31	Populated	Populated	Y
13	0	280,375.69	523,120.55	Unpopulated	Unpopulated	Y
14	0	276,864.31	531,389.63	Populated	Populated	Y
15	0	281,112.88	531,331.93	Populated	Populated	Y
16	0	283,478.13	526,249.17	Unpopulated	Unpopulated	Y
17	0	283,323.17	530,121.74	Unpopulated	UnPopulated	Y
18	0	281,978.52	530,856.52	Unpopulated	Unpopulated	Y
19	0	283,145.17	533,042.30	Populated	Populated	Y
20	0	285,026.12	529,445.34	Unpopulated	Unpopulated	Y
21	0	273,169.44	529,489.29	Unpopulated	Unpopulated	Y
22	0	272,479.34	529,641.08	Unpopulated	Unpopulated	Y
23	0	285,698.32	532,460.38	Unpopulated	Populated	N
24	0	284,058.84	531,573.94	Populated	Populated	Y
25	0	283,797.64	522,486.62	Unpopulated	Unpopulated	Y
26	0	279,607.15	526,177.13	Populated	Populated	Y
27	0	283,318.90	526,861.57	Unpopulated	Unpopulated	Y
28	0	280,570.74	529,915.59	Populated	Unpopulated	N
29	0	283,349.81	524,040.60	Unpopulated	Unpopulated	Y
30	0	281,761.74	522,127.20	Unpopulated	Unpopulated	Y
31	0	284,050.61	524,755.51	Unpopulated	Unpopulated	Y
32	0	274,060.14	528,835.69	Unpopulated	Unpopulated	Y
33	0	281,675.50	531,499.48	Populated	Populated	Y
34	0	280,572.02	531,699.27	Populated	Populated	Y
35	0	286,239.33	525,028.89	Unpopulated	Unpopulated	Y
36	0	282,817.11	523,570.56	Unpopulated	Unpopulated	Y
37	0	280,031.96	526,834.98	Populated	Populated	Y
38	0	274,569.95	526,880.40	Unpopulated	Unpopulated	Y
39	0	283,358.84	530,487.12	Unpopulated	unPopulated	Y
40	0	282,352.73	526,229.98	Populated	Populated	Y

41	0	280,435.80	531,526.70	Unpopulated	Populated	N
42	0	273,037.49	528,804.05	Unpopulated	Unpopulated	Y
43	0	282,784.82	527,831.37	Unpopulated	Unpopulated	Y
44	0	273,326.66	528,620.03	Unpopulated	Unpopulated	Y
45	0	282,816.63	525,795.56	Unpopulated	Populated	N
46	0	281,463.98	527,487.19	Unpopulated	Unpopulated	Y
47	0	277,323.87	528,158.70	Unpopulated	Unpopulated	Y
48	0	279,209.15	525,358.70	Unpopulated	Populated	N
49	0	282,753.51	531,715.95	Unpopulated	Populated	N
50	0	281,338.05	525,583.86	Unpopulated	Unpopulated	Y
51	0	278,615.39	530,602.02	Populated	Populated	Y
52	0	286,384.25	531,068.16	Unpopulated	Unpopulated	Y
53	0	280,338.51	531,894.13	Populated	Populated	Y
54	0	279,747.92	524,242.88	Unpopulated	Unpopulated	Y
55	0	280,094.19	526,064.14	Populated	Populated	Y
56	0	278,215.42	527,780.73	Unpopulated	Unpopulated	Y
57	0	278,509.31	531,812.56	Populated	UnPopulated	N
58	0	285,426.15	531,079.64	Unpopulated	Unpopulated	Y
59	0	275,415.30	528,585.56	Unpopulated	Unpopulated	Y
60	0	283,546.30	523,604.49	Unpopulated	Unpopulated	Y
61	0	284,328.62	526,949.19	Unpopulated	Unpopulated	Y
62	0	280,505.18	524,827.01	Unpopulated	Unpopulated	Y
63	0	278,425.07	532,257.60	Populated	Populated	Y
64	0	284,349.96	531,734.92	Populated	Populated	Y
65	0	282,984.30	523,207.75	Unpopulated	Unpopulated	Y
66	0	277,614.22	531,039.59	Populated	Populated	Y
67	0	285,575.00	528,311.47	Unpopulated	Unpopulated	Y
68	0	280,186.14	526,677.27	Populated	Populated	Y
69	0	280,938.56	531,500.63	Populated	Populated	Y
70	0	278,719.40	530,827.82	Populated	Populated	Y
71	0	284,426.90	531,122.40	Populated	Unpopulated	N
72	0	274,433.22	525,977.28	Unpopulated	Unpopulated	Y
73	0	285,196.49	532,899.26	Unpopulated	Populated	N
74	0	276,845.36	532,672.66	Populated	Unpopulated	N
75	0	279,140.31	526,415.97	Populated	Populated	Y
76	0	279,429.24	529,009.03	Unpopulated	unPopulated	Y
77	0	274,837.86	529,374.70	Unpopulated	UnPopulated	Y
78	0	282,458.03	523,324.33	Unpopulated	Unpopulated	Y
79	0	276,002.49	528,170.28	Unpopulated	Unpopulated	Y
80	0	281,662.32	528,261.83	Populated	Unpopulated	N
81	0	278,786.91	528,841.04	Unpopulated	UnPopulated	Y
82	0	278,308.80	529,732.39	Populated	Populated	Y
83	0	279,366.92	527,737.96	Populated	Populated	Y
84	0	281,509.70	521,430.91	Unpopulated	Unpopulated	Y
85	0	285,789.05	525,843.50	Unpopulated	Unpopulated	Y
86	0	278,498.67	529,044.58	Populated	Unpopulated	N

87	0	284,236.08	526,712.93	Unpopulated	Unpopulated	Y
88	0	280,316.27	530,333.69	Populated	Populated	Y
89	0	286,332.08	524,274.27	Unpopulated	Unpopulated	Y
90	0	272,712.10	528,100.41	Unpopulated	UnPopulated	Y
91	0	286,508.86	525,084.88	Unpopulated	Unpopulated	Y
92	0	275,802.15	526,673.97	Unpopulated	Unpopulated	Y
93	0	283,534.29	530,126.88	Unpopulated	unPopulated	Y
94	0	281,327.13	527,989.25	Unpopulated	Unpopulated	Y
95	0	285,183.08	523,405.42	Unpopulated	Unpopulated	Y
96	0	280,710.08	522,224.04	Unpopulated	Unpopulated	Y
97	0	283,493.49	522,776.53	Unpopulated	Unpopulated	Y
98	0	283,674.43	525,934.36	Unpopulated	Unpopulated	Y
99	0	282,524.53	524,604.30	Populated	Populated	Y
100	0	280,825.66	521,611.64	Unpopulated	Unpopulated	Y
101	0	283,864.74	531,423.84	Populated	Populated	Y
102	0	283,857.45	530,392.62	Unpopulated	unPopulated	Y
103	0	276,728.82	530,092.89	Populated	Populated	Y
104	0	283,320.09	530,928.92	Unpopulated	UnPopulated	Y
105	0	281,201.54	528,533.00	Unpopulated	Unpopulated	Y
106	0	279,131.15	528,210.91	Populated	Populated	Y
107	0	276,924.11	528,866.51	Unpopulated	Unpopulated	Y
108	0	284,807.76	526,838.12	Unpopulated	Unpopulated	Y
109	0	275,884.09	529,918.19	Unpopulated	Unpopulated	Y
110	0	281,122.70	529,921.45	Populated	Populated	Y
111	0	281,035.52	522,268.43	Unpopulated	Unpopulated	Y
112	0	284,651.10	530,454.21	Unpopulated	Unpopulated	Y
113	0	281,200.82	522,402.05	Unpopulated	Unpopulated	Y
114	0	277,828.52	532,083.57	Populated	Populated	Y
115	0	286,598.73	523,347.56	Unpopulated	Unpopulated	Y
116	0	278,306.35	532,039.69	Populated	Populated	Y
117	0	274,653.94	528,346.67	Populated	Unpopulated	N
118	0	280,186.67	524,889.83	Unpopulated	Unpopulated	Y
119	0	283,999.12	523,489.57	Unpopulated	Unpopulated	Y
120	0	280,705.51	532,581.04	Populated	Populated	Y
121	0	278,287.21	528,974.95	Unpopulated	UnPopulated	Y
122	0	286,249.91	531,265.66	Unpopulated	Unpopulated	Y
123	0	276,777.06	527,034.88	Unpopulated	Unpopulated	Y
124	0	280,805.60	521,885.38	Unpopulated	Unpopulated	Y
125	0	282,060.83	532,654.75	Populated	Populated	Y
126	0	279,087.78	528,588.49	Populated	Populated	Y
127	0	279,681.17	528,846.13	Populated	Populated	Y
128	0	279,217.01	532,289.86	Populated	Populated	Y
129	0	279,272.21	524,984.77	Unpopulated	UnPopulated	Y
130	0	282,237.71	524,733.43	Populated	Populated	Y
131	0	283,868.20	526,951.24	Unpopulated	Unpopulated	Y
132	0	278,148.50	528,581.98	Populated	UnPopulated	N

133	0	286,115.43	524,173.40	Unpopulated	Unpopulated	Y
134	0	284,030.13	529,261.67	Unpopulated	Unpopulated	Y
135	0	282,029.57	525,024.90	Populated	Populated	Y
136	0	285,657.80	523,477.06	Unpopulated	Unpopulated	Y
137	0	283,754.47	522,994.70	Unpopulated	Unpopulated	Y
138	0	285,951.75	525,719.23	Populated	Unpopulated	N
139	0	275,169.26	527,655.42	Populated	Populated	Y
140	0	282,209.56	531,801.78	Populated	Populated	Y
141	0	280,471.40	525,097.30	Unpopulated	Unpopulated	Y
142	0	285,498.99	532,816.00	Unpopulated	Populated	N
143	0	283,025.47	531,884.85	Populated	Populated	Y
144	0	278,122.72	532,426.83	Populated	Populated	Y
145	0	281,192.73	524,325.41	Populated	Populated	Y
146	0	282,398.11	527,647.31	Unpopulated	Unpopulated	Y
147	0	279,006.64	524,299.79	Unpopulated	UnPopulated	Y
148	0	279,258.76	531,342.87	Populated	Populated	Y
149	0	276,458.56	530,565.29	Populated	UnPopulated	N
150	0	285,810.03	525,419.18	Unpopulated	Unpopulated	Y
151	0	283,887.69	528,949.74	Unpopulated	Unpopulated	Y
152	0	285,184.34	529,729.71	Unpopulated	Unpopulated	Y
153	0	280,189.25	528,664.72	Unpopulated	UnPopulated	Y
154	0	275,381.88	529,526.29	Unpopulated	Unpopulated	Y
155	0	279,829.28	524,857.74	Unpopulated	Unpopulated	Y
156	0	281,183.22	523,674.78	Populated	Populated	Y
157	0	282,947.91	530,253.08	Populated	Populated	Y
158	0	280,218.38	523,708.29	Unpopulated	Unpopulated	Y
159	0	285,583.82	527,130.79	Unpopulated	Unpopulated	Y
160	0	282,874.23	532,894.35	Populated	Unpopulated	N
161	0	283,553.49	533,023.43	Populated	Unpopulated	N
162	0	281,875.37	529,745.15	Populated	Populated	Y
163	0	274,947.67	528,405.73	Unpopulated	Unpopulated	Y
164	0	276,695.02	531,135.54	Populated	UnPopulated	N
165	0	277,818.37	529,467.33	Populated	Populated	Y
166	0	283,912.45	525,583.73	Unpopulated	Unpopulated	Y
167	0	278,468.19	528,545.57	Populated	Populated	Y
168	0	279,448.00	531,518.39	Populated	UnPopulated	N
169	0	275,337.28	529,930.84	Populated	Unpopulated	N
170	0	282,475.78	527,880.91	Unpopulated	Unpopulated	Y
171	0	281,397.38	523,960.52	Populated	Populated	Y
172	0	280,542.25	532,798.10	Unpopulated	Populated	N
173	0	282,747.22	527,194.03	Unpopulated	Unpopulated	Y
174	0	284,041.79	532,813.44	Populated	Unpopulated	N
175	0	281,258.13	523,307.71	Unpopulated	Unpopulated	Y
176	0	275,404.35	527,379.74	Unpopulated	Unpopulated	Y
177	0	278,448.17	528,062.26	Populated	Populated	Y
178	0	275,200.29	528,484.18	Unpopulated	Unpopulated	Y

179	0	285,825.41	524,088.24	Unpopulated	Unpopulated	Y
180	0	283,286.90	526,482.09	Unpopulated	Unpopulated	Y
181	0	274,240.63	528,264.88	Unpopulated	Unpopulated	Y
182	0	276,492.71	529,448.48	Populated	Unpopulated	N
183	0	282,823.38	523,979.17	Unpopulated	Unpopulated	Y
184	0	283,969.36	523,971.39	Unpopulated	Unpopulated	Y
185	0	285,145.54	524,962.29	Unpopulated	Unpopulated	Y
186	0	275,820.81	527,405.68	Unpopulated	Unpopulated	Y
187	0	282,010.35	527,719.81	Unpopulated	Unpopulated	Y
188	0	283,703.74	525,450.02	Unpopulated	Unpopulated	Y
189	0	278,075.71	532,728.99	Populated	Populated	Y
190	0	272,721.24	529,590.14	Unpopulated	Unpopulated	Y
191	0	285,926.50	523,353.16	Unpopulated	Unpopulated	Y
192	0	276,526.65	528,383.85	Unpopulated	Populated	N
193	0	284,243.09	528,132.99	Unpopulated	Populated	N
194	0	281,272.18	532,849.60	Populated	Populated	Y
195	0	282,214.27	530,133.34	Unpopulated	UnPopulated	Y
196	0	281,443.24	522,087.13	Unpopulated	Unpopulated	Y
197	0	285,553.40	526,819.04	Unpopulated	Unpopulated	Y
198	0	277,778.56	531,397.38	Unpopulated	Unpopulated	Y
199	0	282,958.69	526,669.13	Unpopulated	Unpopulated	Y

Populated surface on google earth 19/200 = 9.5%

Unpopulated surface on google earth 10/200 = 5%

Total 29/200 = 14.5%

Table A10.3 - Results of the visual inspection of surfaces using Quickbird (60cm) image resampled to 3m spatial resolution.

FID	CID	X	Y	Surface	G_Earth	Correct
0	0	284,523.98	523,932.77	Unpopulated	Unpopulated	Y
1	0	285,498.01	524,206.11	Unpopulated	Unpopulated	Y
2	0	283,030.24	522,360.57	Unpopulated	Unpopulated	Y
3	0	279,262.77	532,584.52	Populated	Unpopulated	N
4	0	281,039.39	521,412.38	Unpopulated	Unpopulated	Y
5	0	281,902.47	531,585.61	Populated	populated	Y
6	0	279,588.65	527,344.57	Unpopulated	populated	N
7	0	279,987.87	530,512.30	Unpopulated	Unpopulated	Y
8	0	279,763.72	530,661.25	Populated	populated	Y
9	0	284,872.15	532,669.55	Unpopulated	Populated	N
10	0	282,866.56	528,018.28	Populated	Populated	Y
11	0	283,421.34	527,633.76	Unpopulated	Unpopulated	Y
12	0	286,140.31	531,646.62	Populated	Unpopulated	N
13	0	283,136.85	529,174.84	Unpopulated	Populated	N
14	0	272,898.66	530,158.62	Unpopulated	Unpopulated	Y
15	0	279,435.74	533,040.69	Populated	populated	Y
16	0	278,637.72	532,653.03	Populated	populated	Y
17	0	284,925.70	529,625.60	Unpopulated	Unpopulated	Y
18	0	285,368.43	530,078.73	Unpopulated	Unpopulated	Y
19	0	274,722.90	529,846.80	Unpopulated	Unpopulated	Y
20	0	279,334.52	527,987.10	Unpopulated	populated	N
21	0	282,648.37	530,547.46	Unpopulated	Unpopulated	Y
22	0	283,337.34	531,026.96	Populated	populated	Y
23	0	282,201.37	523,545.04	Unpopulated	Unpopulated	Y
24	0	281,331.98	521,508.03	Unpopulated	Unpopulated	Y
25	0	283,815.43	527,555.64	Populated	Populated	Y
26	0	280,099.59	524,548.23	Unpopulated	Unpopulated	Y
27	0	279,372.49	529,966.89	Unpopulated	Unpopulated	Y
28	0	280,173.88	532,297.66	Populated	Populated	Y
29	0	271,952.19	530,419.57	Populated	Unpopulated	N
30	0	285,391.12	528,860.37	Unpopulated	Unpopulated	Y
31	0	273,945.94	528,979.61	Unpopulated	Unpopulated	Y
32	0	281,247.90	524,296.72	Populated	Populated	Y
33	0	282,550.25	527,038.14	Unpopulated	Unpopulated	Y
34	0	283,622.11	528,939.74	Unpopulated	Unpopulated	Y
35	0	283,675.63	524,284.16	Unpopulated	Unpopulated	Y
36	0	277,811.70	530,295.89	Populated	populated	Y
37	0	283,521.79	528,507.29	Unpopulated	Unpopulated	Y
38	0	281,382.40	523,702.06	Populated	Populated	Y
39	0	283,587.30	526,607.60	Unpopulated	Unpopulated	Y
40	0	274,443.10	527,465.12	Populated	Unpopulated	N

41	0	285,916.54	532,489.16	Populated	Unpopulated	N
42	0	280,045.41	528,365.46	Populated	Unpopulated	N
43	0	280,939.40	531,773.61	Populated	populated	Y
44	0	283,276.90	532,059.19	Unpopulated	Unpopulated	Y
45	0	283,972.40	529,411.88	Unpopulated	Unpopulated	Y
46	0	279,570.22	528,058.29	Populated	Unpopulated	N
47	0	274,753.36	529,216.81	Unpopulated	Unpopulated	Y
48	0	281,002.98	526,405.71	Populated	populated	Y
49	0	285,503.65	524,809.15	Unpopulated	Unpopulated	Y
50	0	278,442.49	527,863.78	Populated	Unpopulated	N
51	0	278,796.24	529,467.22	Unpopulated	Unpopulated	Y
52	0	282,298.04	531,595.46	Populated	populated	Y
53	0	281,203.20	530,852.80	Populated	populated	Y
54	0	284,738.27	527,607.39	Unpopulated	Populated	N
55	0	282,635.48	527,883.89	Unpopulated	Unpopulated	Y
56	0	283,187.81	532,689.96	Unpopulated	Populated	N
57	0	279,953.98	528,886.05	Populated	Unpopulated	N
58	0	279,804.93	527,964.44	Unpopulated	Unpopulated	Y
59	0	279,196.35	525,308.86	Unpopulated	populated	N
60	0	282,813.02	531,330.09	Unpopulated	Unpopulated	Y
61	0	285,598.53	524,540.20	Unpopulated	Unpopulated	Y
62	0	283,328.41	523,109.70	Unpopulated	Unpopulated	Y
63	0	275,971.71	528,452.00	Unpopulated	Unpopulated	Y
64	0	284,304.37	524,171.90	Unpopulated	Unpopulated	Y
65	0	283,350.82	522,527.93	Populated	Unpopulated	N
66	0	278,092.05	528,887.97	Unpopulated	Unpopulated	Y
67	0	284,765.87	526,484.20	Unpopulated	Unpopulated	Y
68	0	279,862.22	523,869.89	Unpopulated	Unpopulated	Y
69	0	275,182.23	526,028.91	Unpopulated	Unpopulated	Y
70	0	275,016.40	529,089.17	Unpopulated	Unpopulated	Y
71	0	278,185.10	532,761.23	Populated	populated	Y
72	0	279,881.39	529,352.87	Unpopulated	Unpopulated	Y
73	0	281,634.35	530,087.98	Populated	Populated	Y
74	0	280,950.73	528,504.72	Populated	Unpopulated	N
75	0	281,121.41	532,533.86	Populated	Populated	Y
76	0	285,011.80	524,143.52	Unpopulated	Unpopulated	Y
77	0	285,012.09	526,688.42	Unpopulated	Unpopulated	Y
78	0	282,184.29	526,438.12	Populated	Populated	Y
79	0	284,137.35	527,796.18	Unpopulated	Unpopulated	Y
80	0	279,884.52	525,836.21	Populated	populated	Y
81	0	285,807.51	531,940.26	Unpopulated	Populated	N
82	0	285,727.85	524,979.80	Unpopulated	Unpopulated	Y
83	0	272,142.06	530,317.91	Unpopulated	Unpopulated	Y
84	0	282,792.68	528,231.54	Populated	Unpopulated	N
85	0	280,991.45	533,176.64	Populated	Populated	Y
86	0	284,455.33	529,396.83	Unpopulated	Unpopulated	Y

87	0	276,223.18	529,732.99	Populated	Unpopulated	N
88	0	277,715.00	530,877.34	Populated	populated	Y
89	0	283,341.82	524,444.29	Unpopulated	Populated	N
90	0	275,067.80	528,517.74	Unpopulated	Unpopulated	Y
91	0	281,531.22	523,178.90	Unpopulated	Unpopulated	Y
92	0	272,509.60	529,877.15	Unpopulated	Unpopulated	Y
93	0	281,832.86	530,975.17	Unpopulated	Unpopulated	Y
94	0	284,999.21	526,354.78	Unpopulated	Unpopulated	Y
95	0	280,004.19	526,059.55	Populated	populated	Y
96	0	276,453.95	529,152.85	Populated	Unpopulated	N
97	0	282,556.96	529,951.55	Unpopulated	Populated	N
98	0	278,635.54	528,389.10	Populated	Unpopulated	N
99	0	285,080.24	522,994.06	Unpopulated	Unpopulated	Y
100	0	276,172.49	527,319.64	Unpopulated	Unpopulated	Y
101	0	284,553.97	530,590.70	Unpopulated	Unpopulated	Y
102	0	284,467.51	527,051.34	Unpopulated	Unpopulated	Y
103	0	284,956.26	530,522.89	Unpopulated	Unpopulated	Y
104	0	282,332.73	532,729.33	Unpopulated	Unpopulated	Y
105	0	285,350.75	526,648.14	Unpopulated	Unpopulated	Y
106	0	281,444.47	528,852.99	Unpopulated	Unpopulated	Y
107	0	273,264.29	528,349.05	Unpopulated	Unpopulated	Y
108	0	284,143.72	529,862.21	Unpopulated	Populated	N
109	0	276,966.48	527,916.74	Unpopulated	Unpopulated	Y
110	0	282,790.46	529,343.20	Unpopulated	Populated	N
111	0	278,918.41	530,850.03	Populated	populated	Y
112	0	283,615.39	525,313.46	Populated	Unpopulated	N
113	0	285,182.17	527,635.51	Unpopulated	Unpopulated	Y
114	0	274,675.65	526,909.09	Unpopulated	Unpopulated	Y
115	0	280,742.09	527,453.98	Populated	populated	Y
116	0	281,260.52	525,700.93	Unpopulated	Unpopulated	Y
117	0	275,468.87	526,297.27	Unpopulated	Unpopulated	Y
118	0	285,938.69	523,132.02	Unpopulated	Unpopulated	Y
119	0	281,757.38	525,255.54	Populated	Unpopulated	N
120	0	281,938.37	530,147.01	Unpopulated	Unpopulated	Y
121	0	277,609.17	529,041.83	Populated	populated	Y
122	0	280,786.41	530,336.35	Populated	unpopulated	N
123	0	277,041.70	532,371.07	Populated	Unpopulated	N
124	0	276,359.68	528,388.03	Populated	populated	Y
125	0	282,458.10	523,248.95	Unpopulated	Unpopulated	Y
126	0	276,924.03	531,321.39	Populated	populated	Y
127	0	280,761.79	523,274.71	Unpopulated	Unpopulated	Y
128	0	277,380.51	527,843.40	Unpopulated	Unpopulated	Y
129	0	275,433.94	529,331.47	Unpopulated	Unpopulated	Y
130	0	279,989.60	524,192.63	Populated	Unpopulated	N
131	0	273,114.98	529,539.55	Unpopulated	Unpopulated	Y
132	0	276,164.55	528,105.54	Populated	populated	Y

133	0	274,742.38	528,541.56	Unpopulated	Unpopulated	Y
134	0	279,442.82	531,890.08	Unpopulated	populated	N
135	0	280,300.68	527,409.85	Populated	populated	Y
136	0	282,682.53	531,496.90	Unpopulated	Unpopulated	Y
137	0	281,626.30	533,197.67	Unpopulated	Unpopulated	Y
138	0	281,970.11	522,044.60	Unpopulated	Unpopulated	Y
139	0	283,604.48	533,290.99	Unpopulated	Populated	N
140	0	272,053.14	528,978.14	Unpopulated	Unpopulated	Y
141	0	274,613.83	527,738.93	Populated	Unpopulated	N
142	0	285,563.28	525,970.74	Unpopulated	Unpopulated	Y
143	0	276,432.56	526,932.27	Unpopulated	Unpopulated	Y
144	0	283,298.20	530,205.73	Unpopulated	Unpopulated	Y
145	0	275,269.37	528,869.41	Unpopulated	Unpopulated	Y
146	0	276,475.57	532,722.47	Populated	Unpopulated	N
147	0	279,578.46	529,589.12	Populated	Unpopulated	N
148	0	284,364.77	529,686.90	Unpopulated	Unpopulated	Y
149	0	276,234.19	527,606.60	Unpopulated	Unpopulated	Y
150	0	286,277.62	523,086.91	Unpopulated	Unpopulated	Y
151	0	285,589.97	528,321.04	Unpopulated	Unpopulated	Y
152	0	275,298.84	529,575.11	Populated	Unpopulated	N
153	0	275,683.86	528,157.21	Unpopulated	Unpopulated	Y
154	0	280,418.05	529,684.01	Unpopulated	Unpopulated	Y
155	0	278,714.99	529,870.70	Populated	populated	Y
156	0	277,076.22	528,485.63	Unpopulated	Unpopulated	Y
157	0	280,433.55	527,804.69	Populated	populated	Y
158	0	276,754.45	527,513.96	Unpopulated	Unpopulated	Y
159	0	280,744.06	531,040.21	Populated	populated	Y
160	0	284,266.02	530,173.62	Unpopulated	Unpopulated	Y
161	0	281,775.25	521,684.90	Unpopulated	Unpopulated	Y
162	0	280,851.01	529,804.93	Unpopulated	Unpopulated	Y
163	0	284,516.55	531,709.33	Unpopulated	Unpopulated	Y
164	0	286,670.98	524,010.72	Unpopulated	Unpopulated	Y
165	0	284,912.67	530,888.83	Unpopulated	Unpopulated	Y
166	0	279,095.14	526,996.25	Unpopulated	populated	N
167	0	283,405.69	526,303.59	Unpopulated	Unpopulated	Y
168	0	286,033.19	531,450.08	Unpopulated	Unpopulated	Y
169	0	278,954.47	530,060.19	Populated	populated	Y
170	0	275,954.39	528,098.62	Unpopulated	Unpopulated	Y
171	0	284,307.60	527,578.91	Populated	Populated	Y
172	0	274,900.37	528,020.09	Populated	populated	Y
173	0	281,508.95	525,821.75	Unpopulated	Unpopulated	Y
174	0	284,648.38	530,108.65	Unpopulated	Unpopulated	Y
175	0	283,392.45	533,209.95	Unpopulated	Unpopulated	Y
176	0	285,380.58	524,008.20	Unpopulated	Unpopulated	Y
177	0	283,618.70	531,012.48	Unpopulated	Unpopulated	Y
178	0	279,283.34	530,269.02	Populated	populated	Y

179	0	280,394.91	524,657.01	Unpopulated	Unpopulated	Y
180	0	278,477.56	531,403.82	Populated	populated	Y
181	0	281,728.08	528,705.24	Populated	populated	Y
182	0	274,989.24	526,155.02	Unpopulated	Unpopulated	Y
183	0	281,187.40	522,195.56	Unpopulated	Unpopulated	Y
184	0	284,413.75	528,075.07	Populated	Populated	Y
185	0	284,588.33	528,613.27	Populated	Populated	Y
186	0	278,085.93	528,116.17	Populated	Unpopulated	N
187	0	276,355.18	531,404.69	Populated	Unpopulated	N
188	0	280,649.51	523,764.45	Populated	Unpopulated	N
189	0	281,679.90	523,334.48	Populated	Unpopulated	N
190	0	277,440.98	529,246.20	Populated	populated	Y
191	0	285,233.92	523,559.32	Unpopulated	Unpopulated	Y
192	0	274,281.61	529,472.84	Unpopulated	Unpopulated	Y
193	0	272,408.15	528,177.58	Unpopulated	Unpopulated	Y
194	0	278,697.53	525,802.91	Populated	Unpopulated	N
195	0	276,631.72	529,058.84	Populated	populated	Y
196	0	285,148.30	524,752.47	Unpopulated	Unpopulated	Y
197	0	281,472.61	532,416.43	Populated	Populated	Y
198	0	285,585.20	530,681.91	Unpopulated	Unpopulated	Y
199	0	277,396.57	528,711.99	Populated	populated	Y

Populated surface on google earth 29/200 = 14.5%

Unpopulated surface on google earth 15/200 = 7.5%

Total 44/200 = 22%

## Appendix 11: Codes for allocating demand to all 17 current health centres

Some of the functions were obtained from Professor Alexis Comber and modified to suit the objectives of this research.

```
#load the libraries needed for the analyses
library(GISTools)
#load the source code and support functions for the GGA
source("permute.r")
source("jegafunctions.r")
#Load the source code for the Tietz-Bart algorithm
source('pmedians.R')
#set working directory to the right folder and read in the data to be
used
demand <- readShapePoints("Ph30m_30mgridpoints.shp")
supply <- readShapePoints("Health_centres.shp")
ph <- readShapePoly('port_harcourt.shp')
roads <- readShapeLines('PHCRoads.shp')
#read in the OD Matrix csv file from ArcGIS
access <- read.csv(current_access.csv',as.is=T)
head(access)
#attach the dataset (access) to make it available to memory and refer
to each variable by name
attach(access)
#calculate the distance matrix by grouping items together from the
first category, converting the list to an array such that when the
file is attached column names becomes variables using an identify
function (function (x) x) that returns what is given
d.mat = tapply(Total_Length,list(OriginID, DestinationID),function (x)
x)
#check the dimension of the distance matrix
dim(d.mat)
#detach the dataset (access)
detach(access)
#select the estimated populations and call it pops
pops <- demand$estimates
#check the length of pops
length(pops)
#use only the lines that were used by the OD Matrix
pops <- pops[1:54072]
#select the health centres for which demand is to be allocated
tb.best.loc <- 1:17
selected <- tb.best.loc

#classify the health centres interms of distance - that is the minimum
distance to each demand
dist <- classify(selected, type="dists")
#name the health centre allocated to each demand
names <- classify(selected)
#put the result in a data frame
results <- data.frame(healthcentre=names, Distance=dist, demand=pops)
#obtain the total population (demand) allocated to each health centre
- rowDemand
ans1 <- tapply(results$demand, results$healthcentre, sum)
#obtain the mean distance from demand to health centre within the
catchment
ans2 <- tapply(results$Distance, results$healthcentre, mean)
```

```

#obtain the maximum distance from demand to health centre within the
catchment
ans3 <- tapply(results$Distance,results$healthcentre,max)
#put the results in a data frame
ans4 <-
data.frame(Health_centres=rownames(ans3),Demand=ans1,mean.dist=ans2,
maximum=ans3)
#Show the demand allocation in terms of percentage of the total demand
#divide the demand by its sum and call it newDemand
newDemand <- ans4$Demand/sum(ans4$Demand)
#multiply newDemand by 100 to show each demand as a percentage of
total demand
Demand_percent <- newDemand * 100
#add Demand_percent to ans4
ans4_results <- cbind(ans4,Demand_percent)
#write it out as a csv file
write.csv(ans4_results,'Demandallfacilities.csv')

```

## Appendix 12: Codes for generating optimal locations using 85 potential locations

Some of the functions were obtained from Professor Alexis Comber and modified to suit the objectives of this research.

```
#load the libraries needed for the analyses
library(GISTools)
#load the source code and support functions for the GGA
source("permute.r")
source("jegafunctions.r")
#Load the source code for the Tietz-Bart algorithm
source('pmedians.R')
#set working directory to the right folder and read in the data to be
used
demand <- readShapePoints("Ph30m_30mgrids_points.shp")
ph <- readShapePoly('port_harcourt.shp')
roads <- readShapeLines('PHCRoads.shp')
supply <- readShapePoints("85potential.shp")
#read in the OD Matrix csv file from ArcGIS
access2 <- read.csv('potential_access.csv',as.is=T)
head(access2)
#attach the dataset (access2) to make it available to memory and refer
to each variable by name
attach(access2)
#Remove the comma's in OriginID and Total length from access2 file and
name it access
access <- data.frame(ObjectID = ObjectID, Name = Name,
OriginID=as.numeric(gsub(",","",OriginID)),DestinationID =
DestinationID, DestinationRank = DestinationRank,
Total_Length=as.numeric(gsub(",","",Total_Length)))
detach(access2)
attach(access)
#calculate the distance matrix by grouping items together from the
first category, converting the list to an array such that when the
file is attached column names becomes variables using an identify
function (function (x) x) that returns what is given
d.mat = tapply(Total_Length,list(OriginID,DestinationID),function (x)
x)
#check the dimension of the distance matrix
dim(d.mat)
#detach the dataset (access)
detach(access)
#select the estimated populations and call it pops
pops <- demand$estimates
#check the length of pops
length(pops)
#use only the lines that were used by the OD Matrix
pops <- pops[1:54072]
#check the quality of the data for holes
holes <- spot.holes(d.mat)
holes
#run the Teitz-Bart algorithm with 17 health facilities to allocate
from 85 potential sites
tb.best.loc <- tb(1:17,d.mat,pops)
#select the best subset and call it selected
selected <- tb.best.loc
#display the best subset from the supply data and call it optimal
point
```

```

optimal.points <- supply[selected,]
#display the non optimal points
not.optimal.points <- supply[-selected,]
#classify the health centres interms of distance - that is the minimum
distance to each demand
dist <- classify(selected, type = "dists")
#name the health centre allocated to each demand
names <- classify(selected)
#put the result in a data frame
results <- data.frame(healthcentre=names, Distance=dist, demand=pops)
#obtain the total population (demand) allocated to each health centre
- rowDemand
ans1 <- tapply(results$demand,results$healthcentre,sum)
#obtain the mean distance from demand to health centre within the
catchment
ans2 <- tapply(results$Distance,results$healthcentre,mean)
#obtain the maximum distance from demand to health centre within the
catchment
ans3 <- tapply(results$Distance,results$healthcentre,max)
#put the results in a data frame
ans4 <-
data.frame(Health_centres=rownames(ans3),Demand=ans1,mean.dist=ans2,
maximum=ans3)
#Show the demand allocation in terms of percentage of the total demand
#divide the demand by its sum and call it newDemand
newDemand <- ans4$Demand/sum(ans4$Demand)
#multiply newDemand by 100 to show each demand as a percentage of
total demand
Demand_percent <- newDemand * 100
#add Demand_percent to ans4
ans4_results <- cbind(ans4,Demand_percent)
#write it out as a csv file
write.csv(ans4_results,'17potentialsites.csv')

```

**Appendix 13: Spatial distributions of 5 to 16 PHCCs selected from current locations**

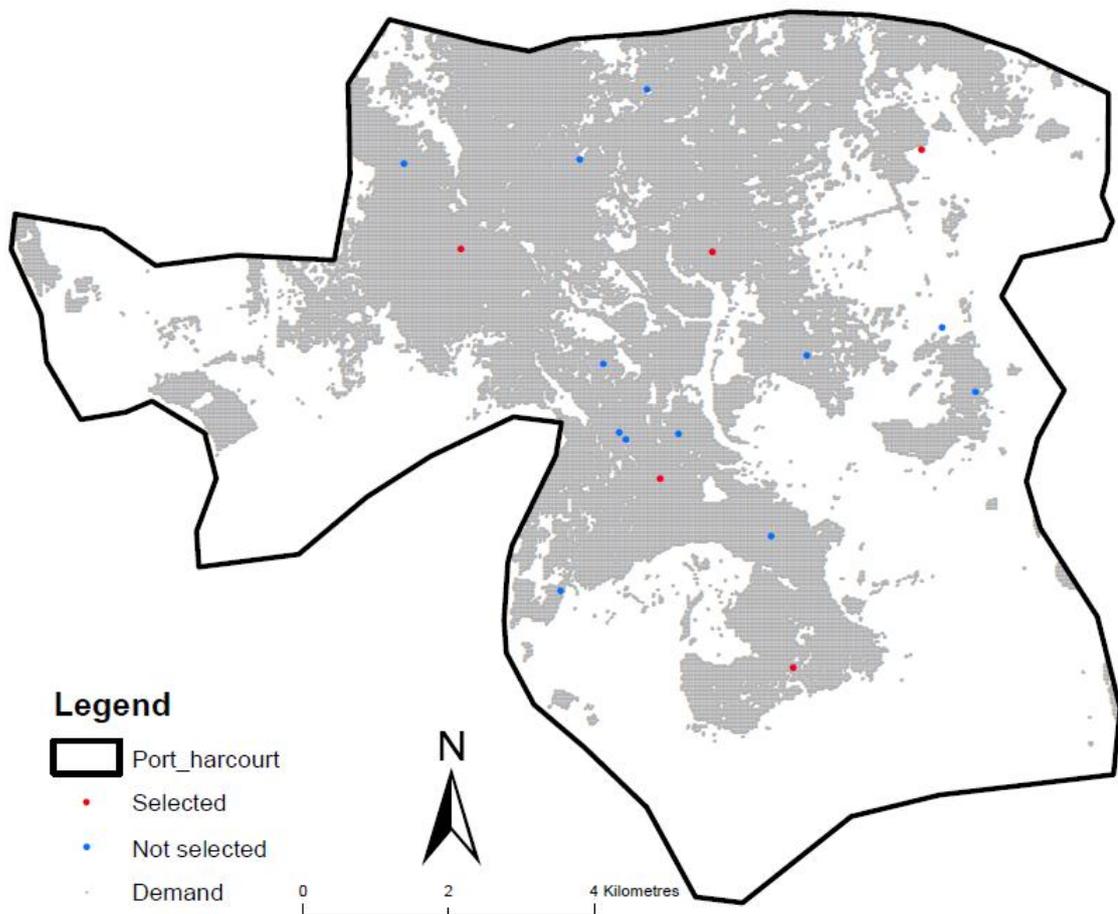


Figure A13.1 - The spatial distributions of selected locations of 5 PHCCs from current locations. The digital boundary is copyright for Geotechnics Services 2011.

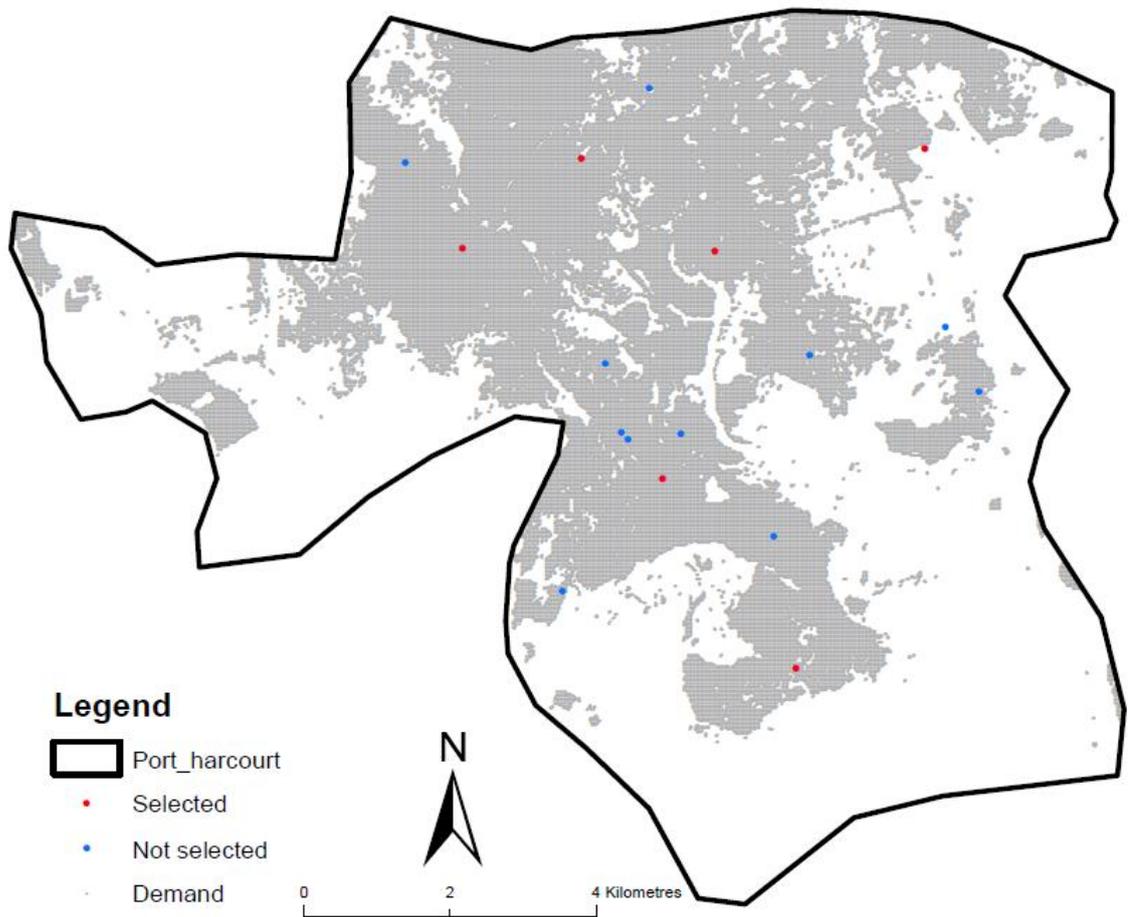


Figure A13.2 - The spatial distributions of selected locations of 6 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

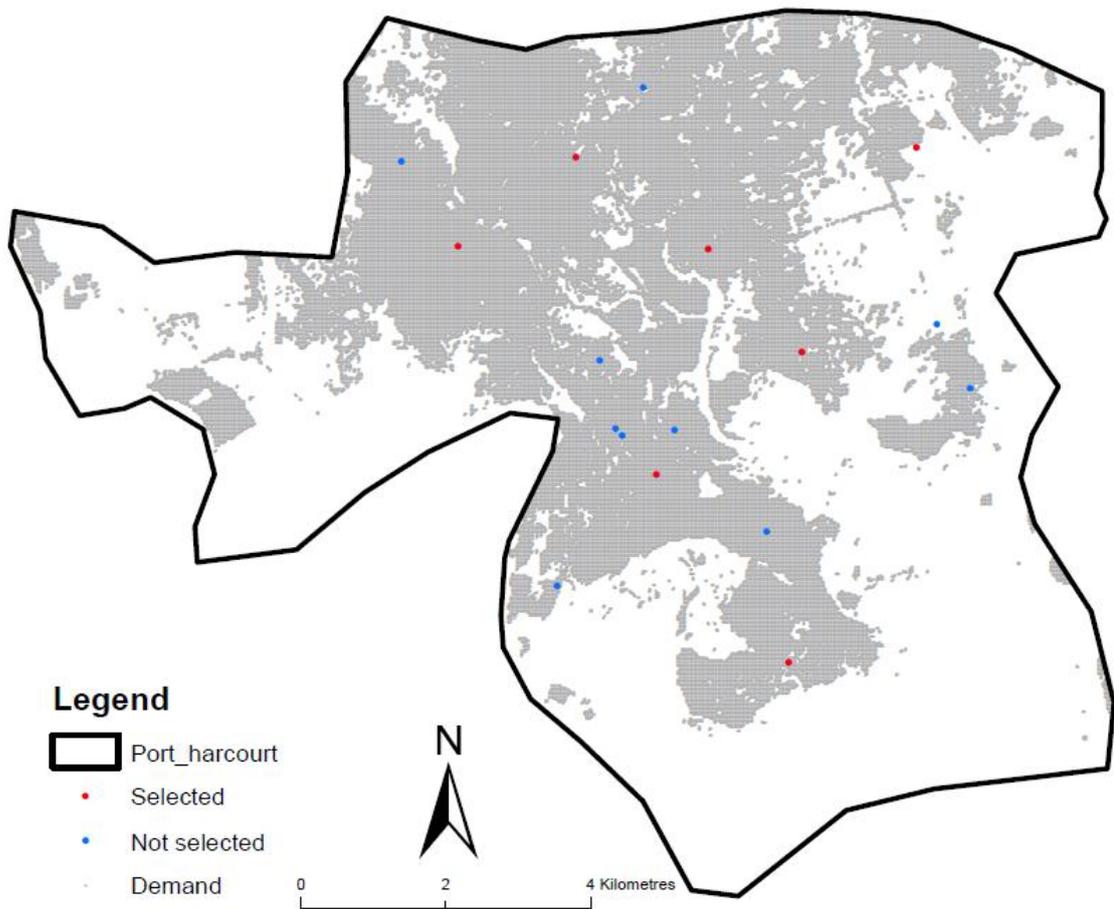


Figure A13.3 - The spatial distributions of selected locations of 7 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

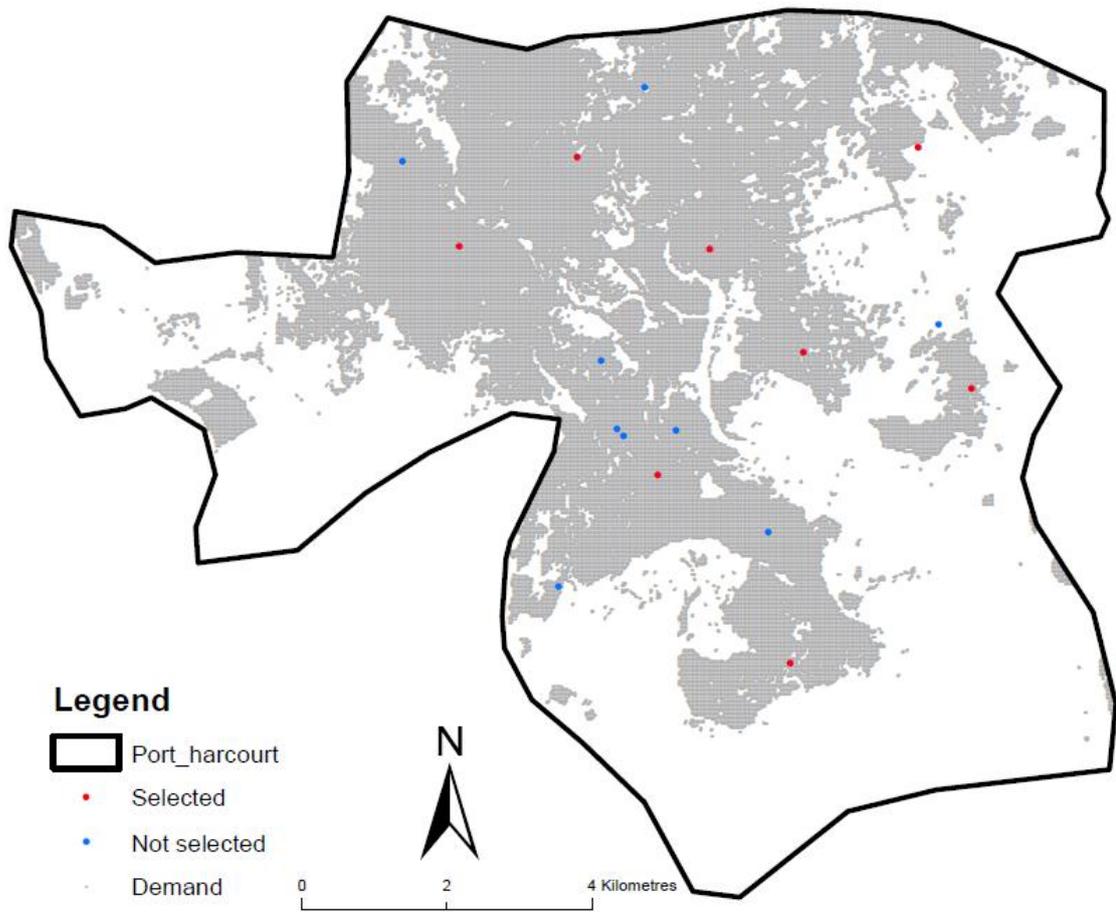


Figure A13.4 - The spatial distributions of selected locations of 8 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

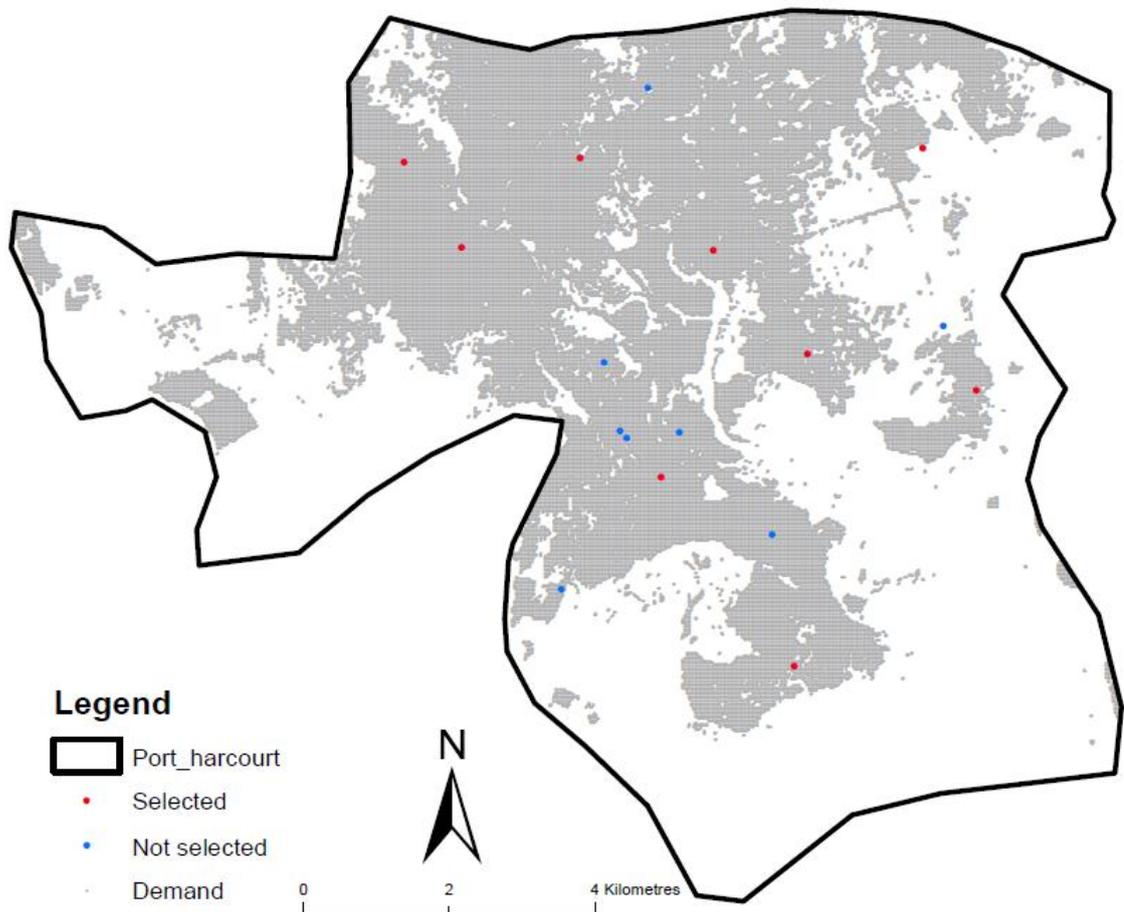


Figure A13.5 - The spatial distributions of selected locations of 9 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

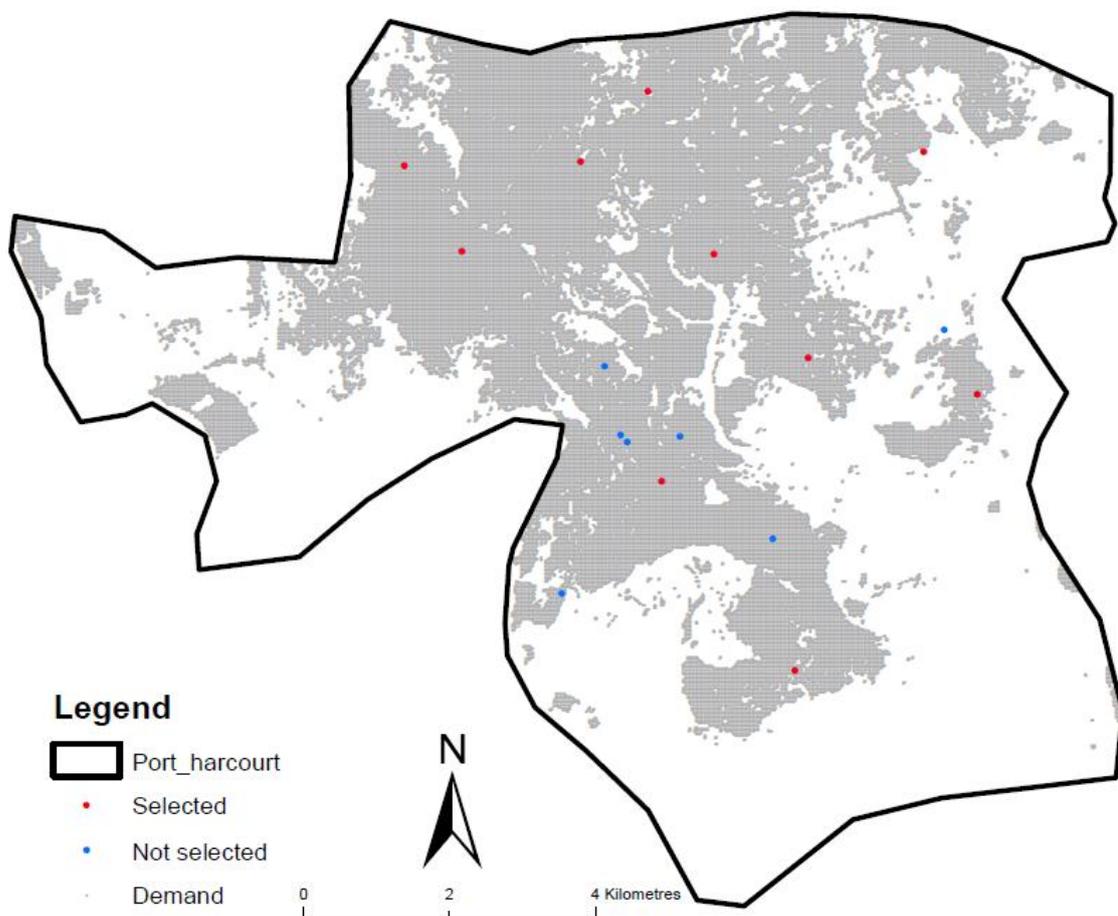


Figure A13.6 - The spatial distributions of selected locations of 10 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

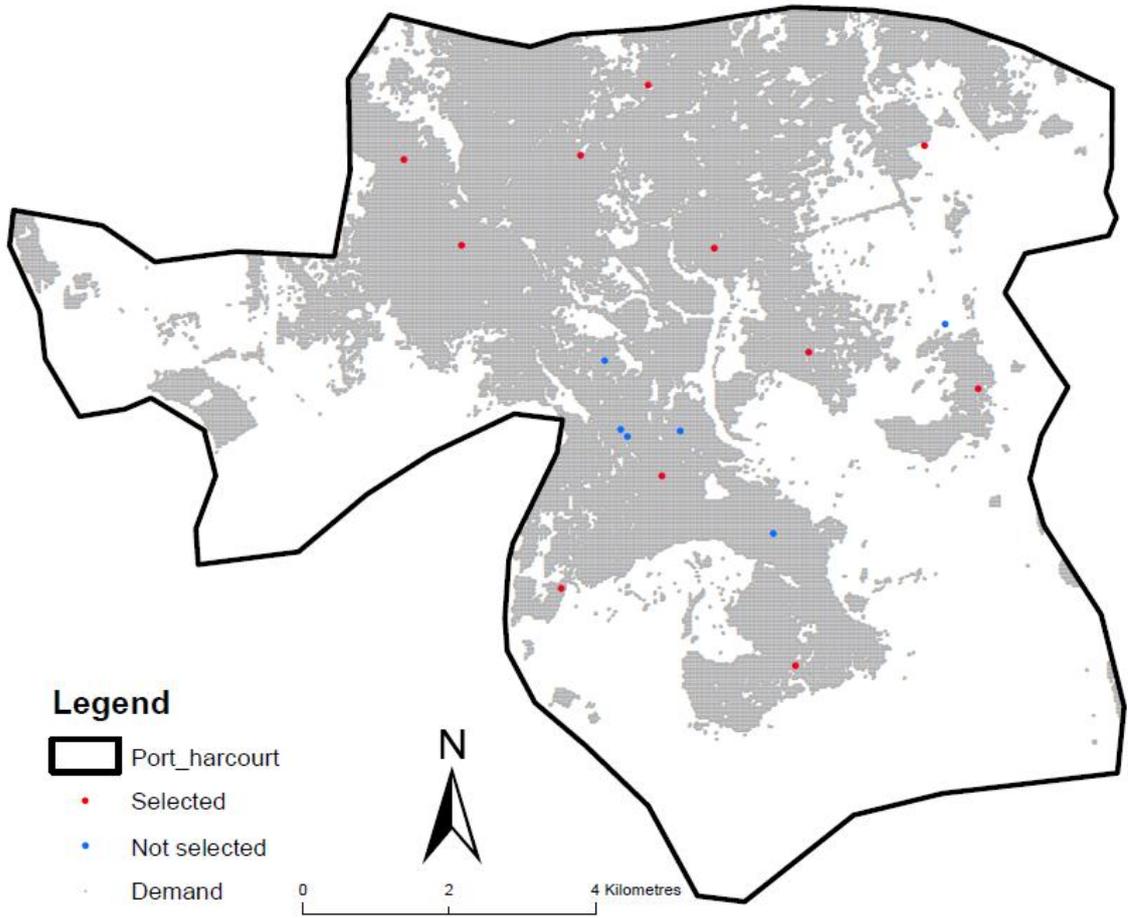


Figure A13.7 - The spatial distributions of selected locations of 11 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

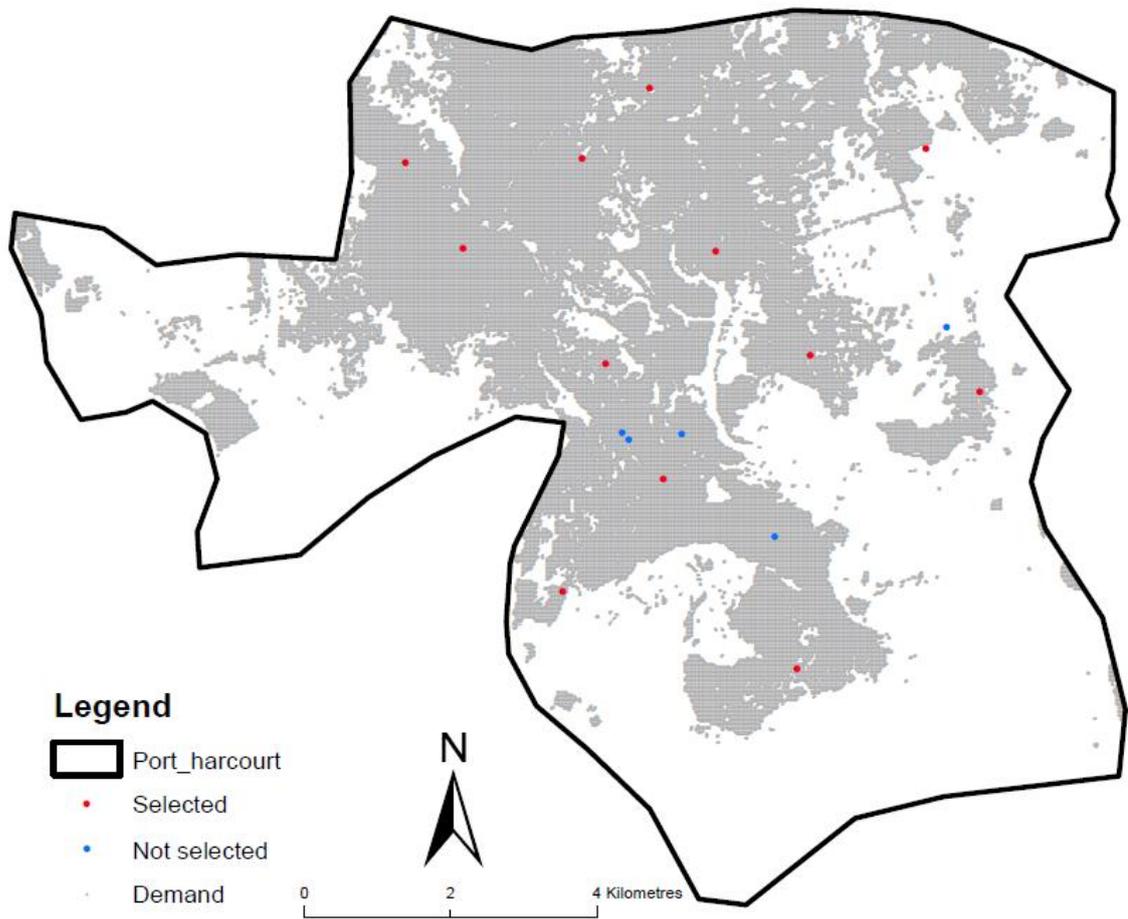


Figure A13.8 - The spatial distributions of selected locations of 12 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

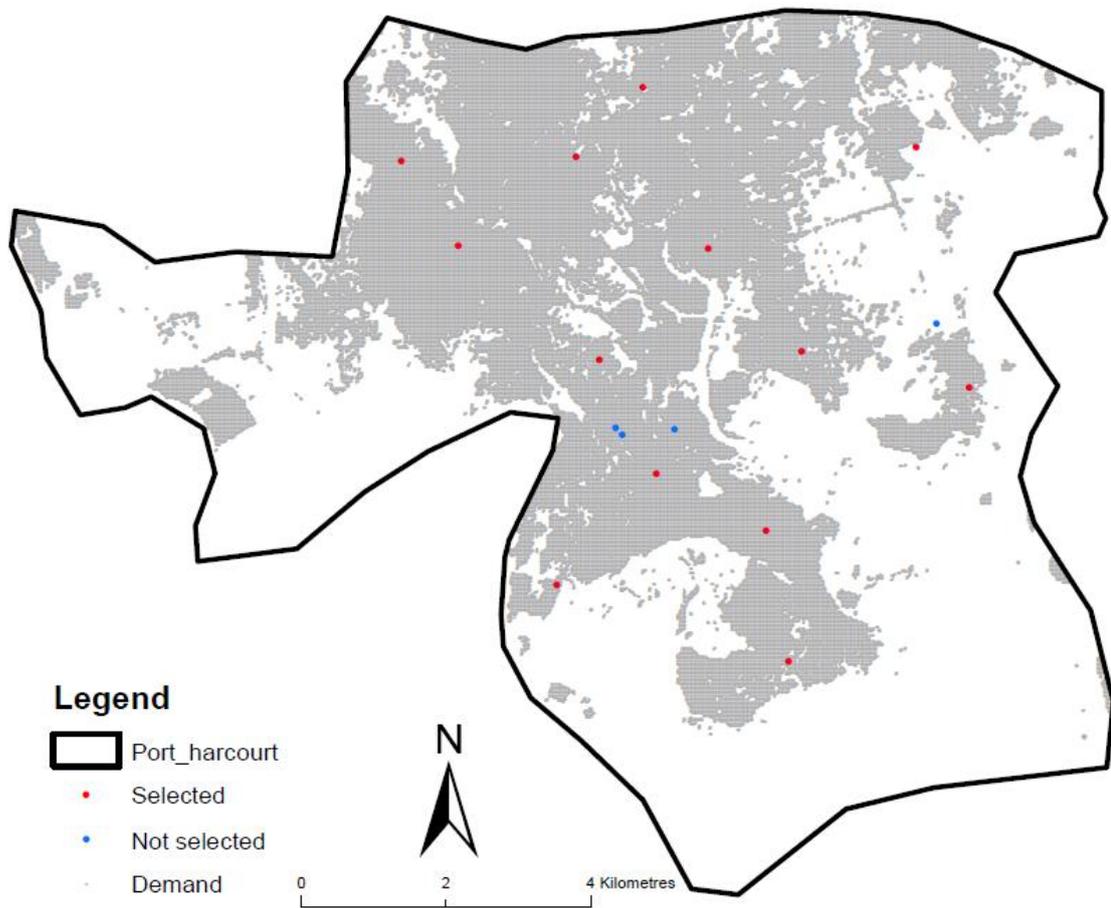


Figure A13.9 - The spatial distributions of selected locations of 13 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

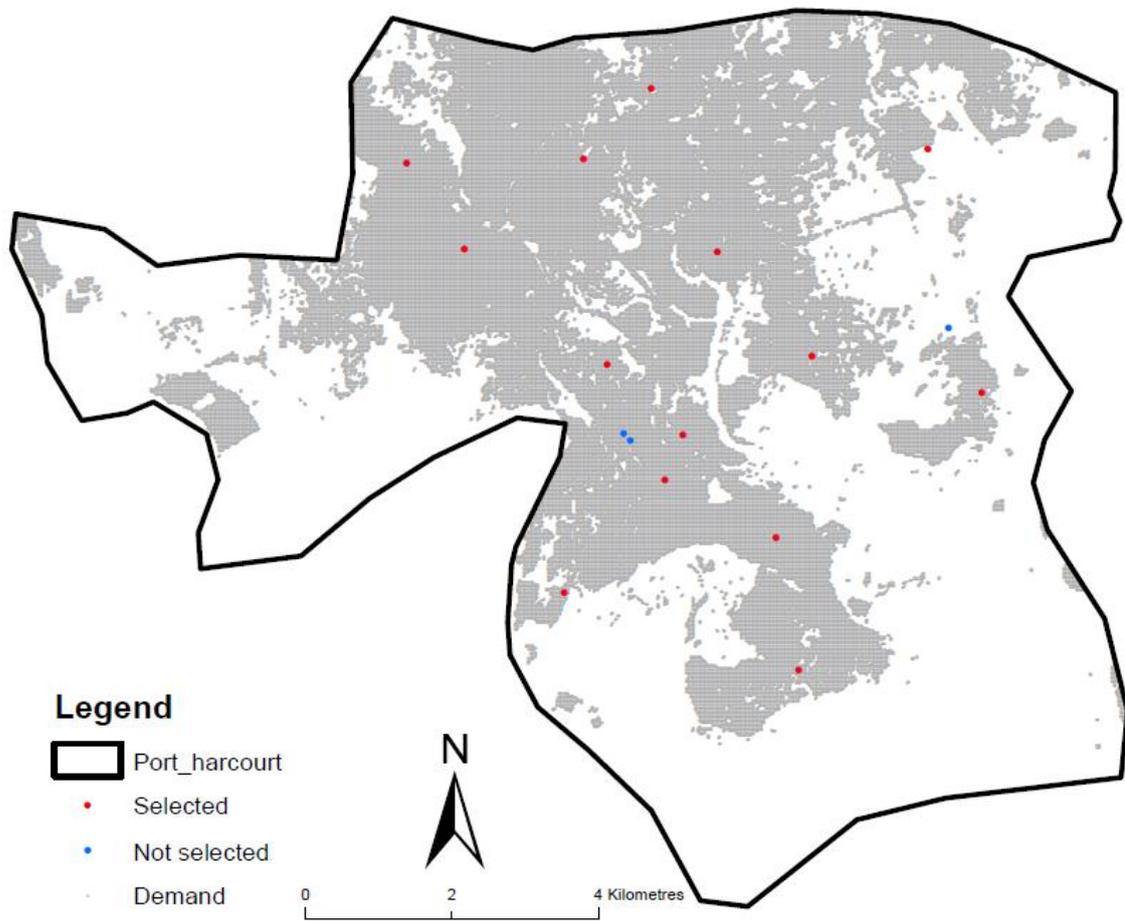


Figure A13.10 - The spatial distributions of selected locations of 14 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

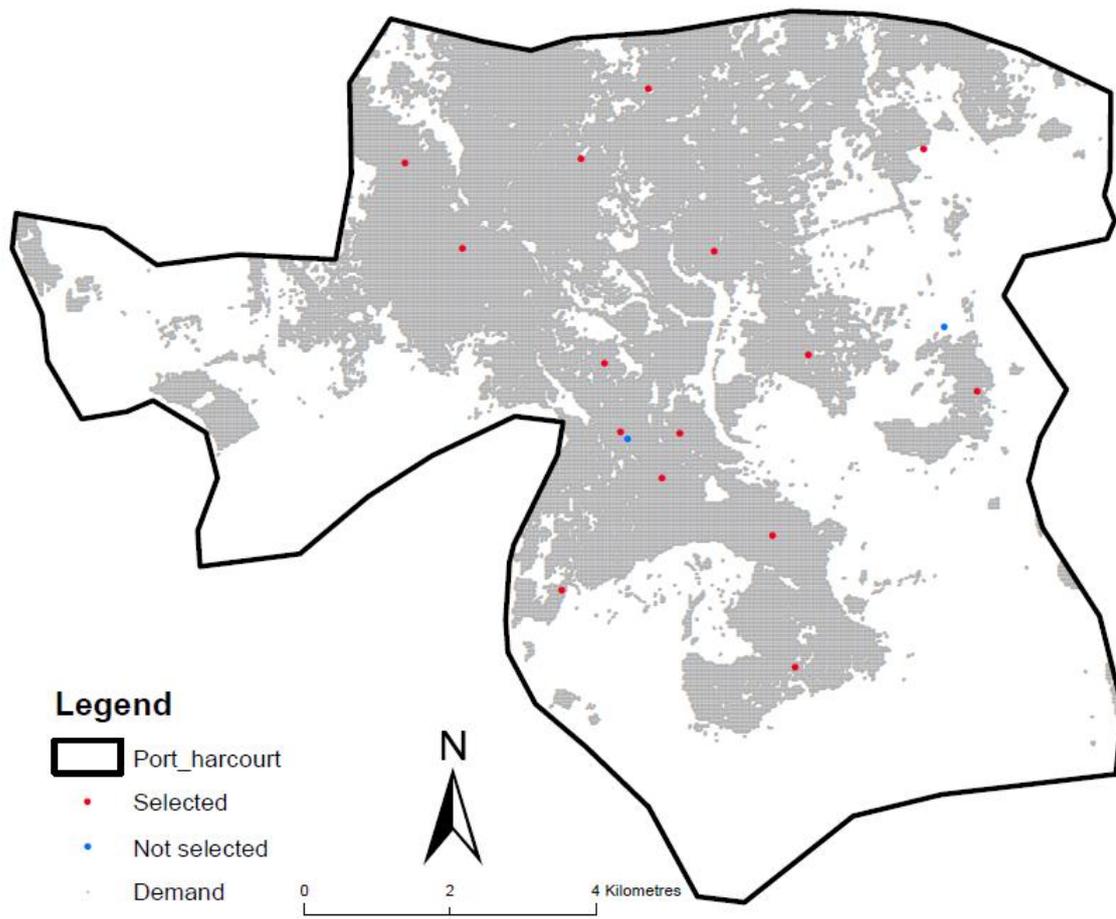


Figure A13.11 - The spatial distributions of selected locations of 15 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

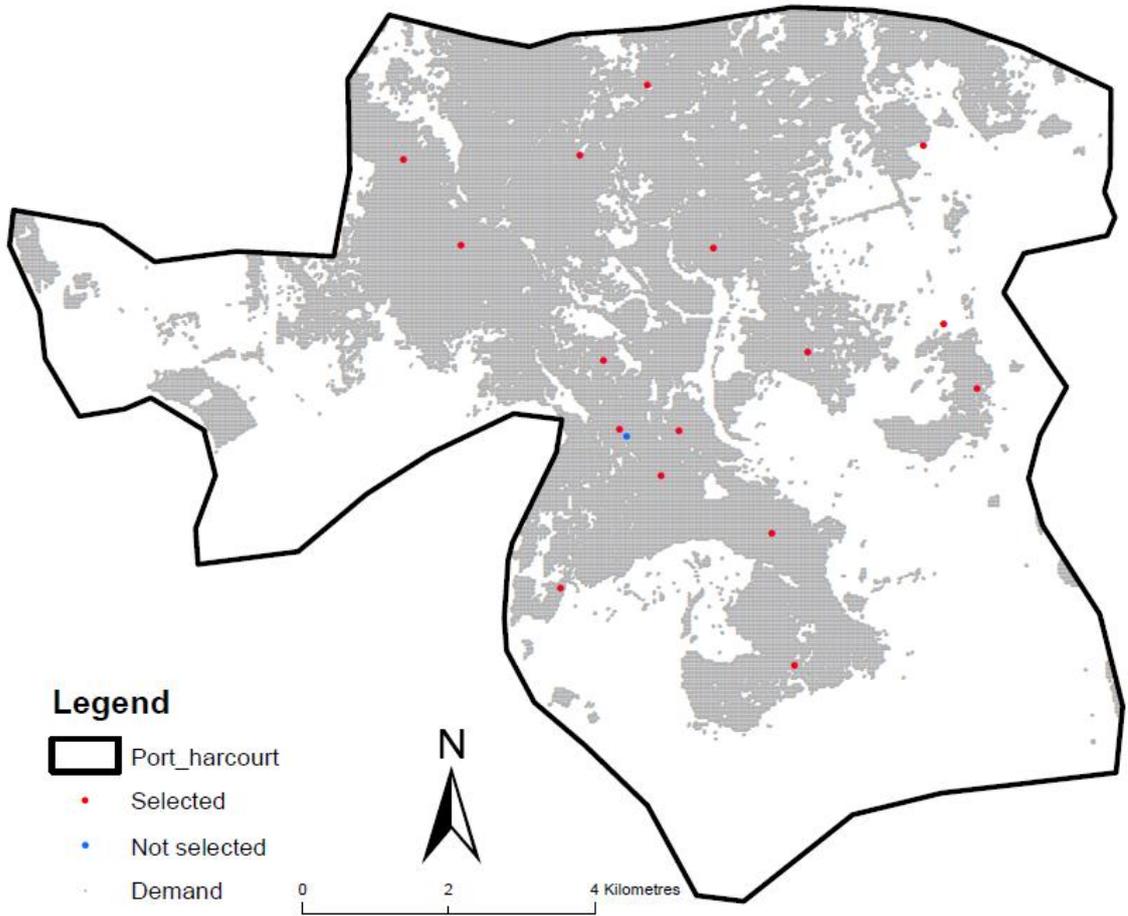


Figure A13.12 - The spatial distributions of selected locations of 16 PHCCs from current locations. The digital boundary is Copyright for Geo-technics Services 2011.

**Appendix 14: Spatial distributions of 5 to 20 PHCCs selected from potential locations.**

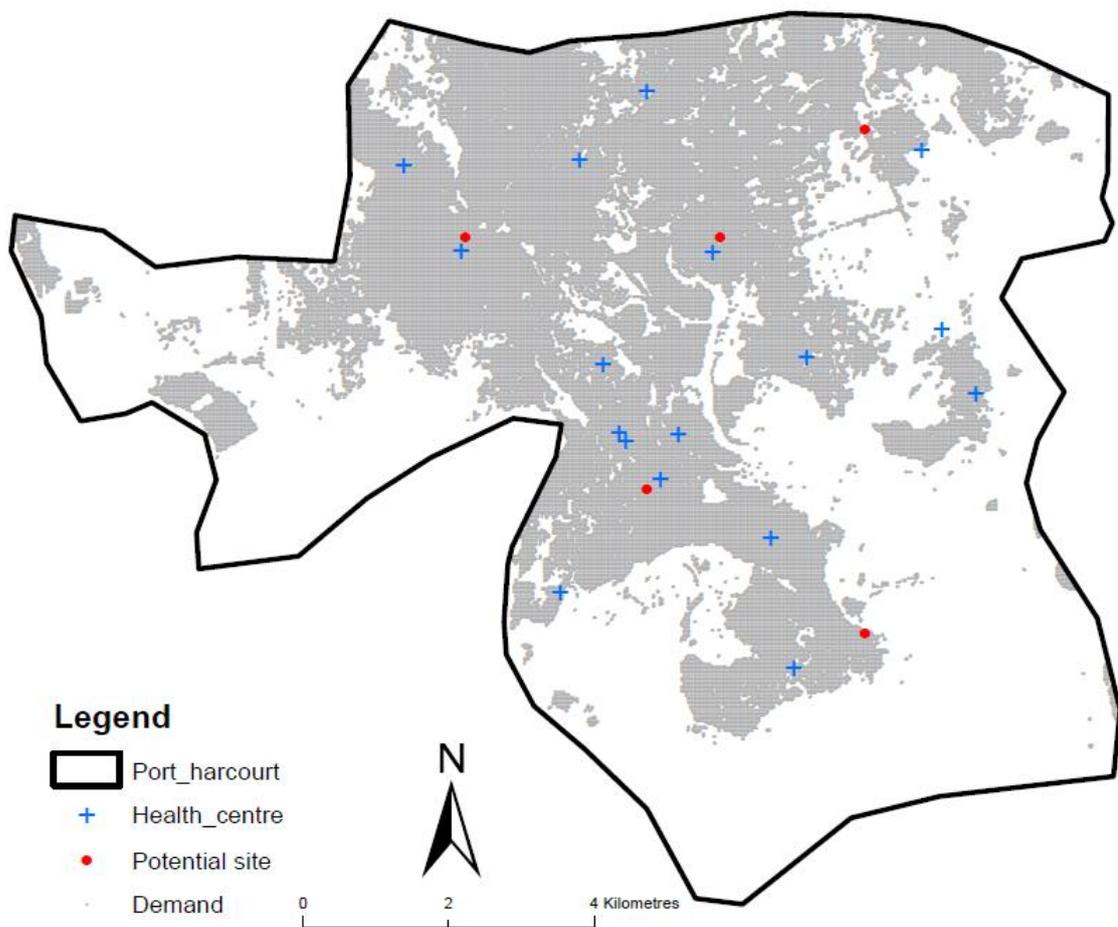


Figure A14.1 - The spatial distribution of 5 optimal locations (red circles) selected from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geo-technics Services 2011.

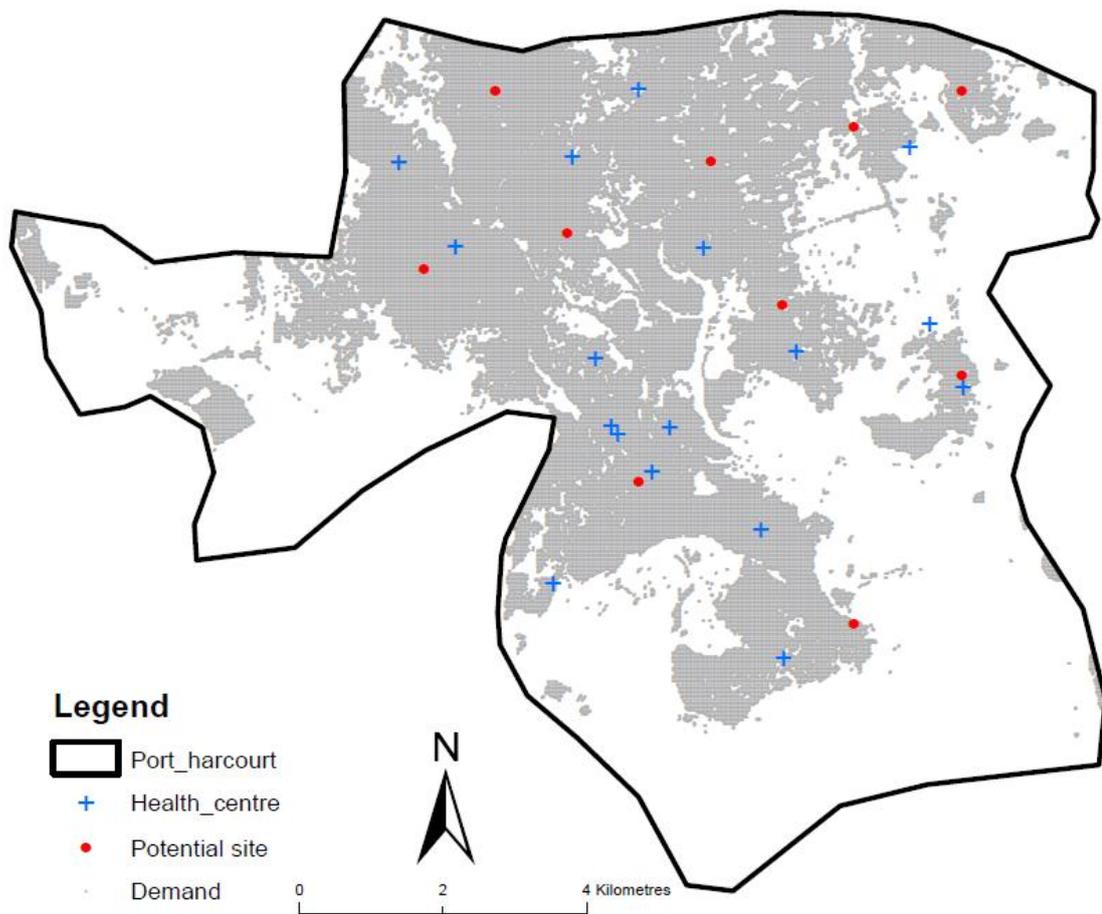


Figure A14.2 - The spatial distribution of 10 optimal locations (red circles) selected from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geo-technics Services 2011.

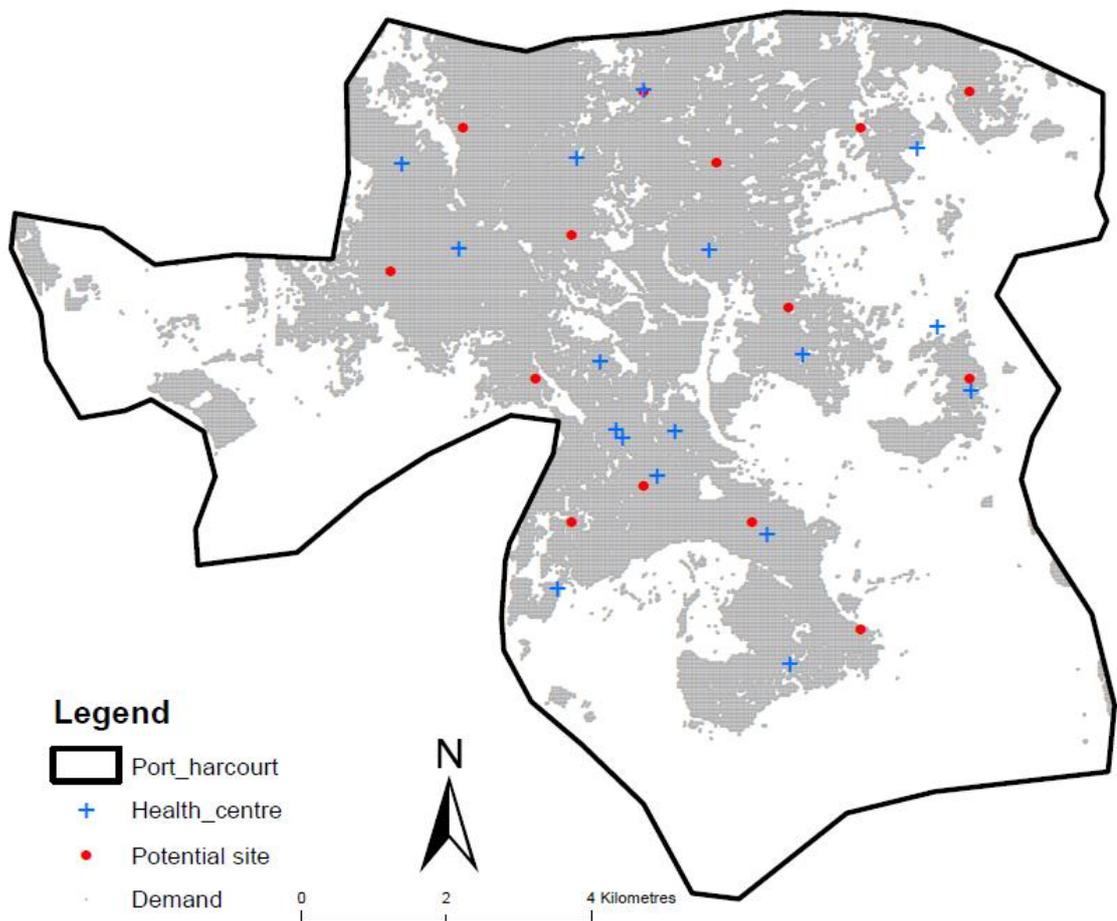


Figure A14.3 - The spatial distribution of 14 optimal locations (red circles) selected from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geo-technics Services 2011.

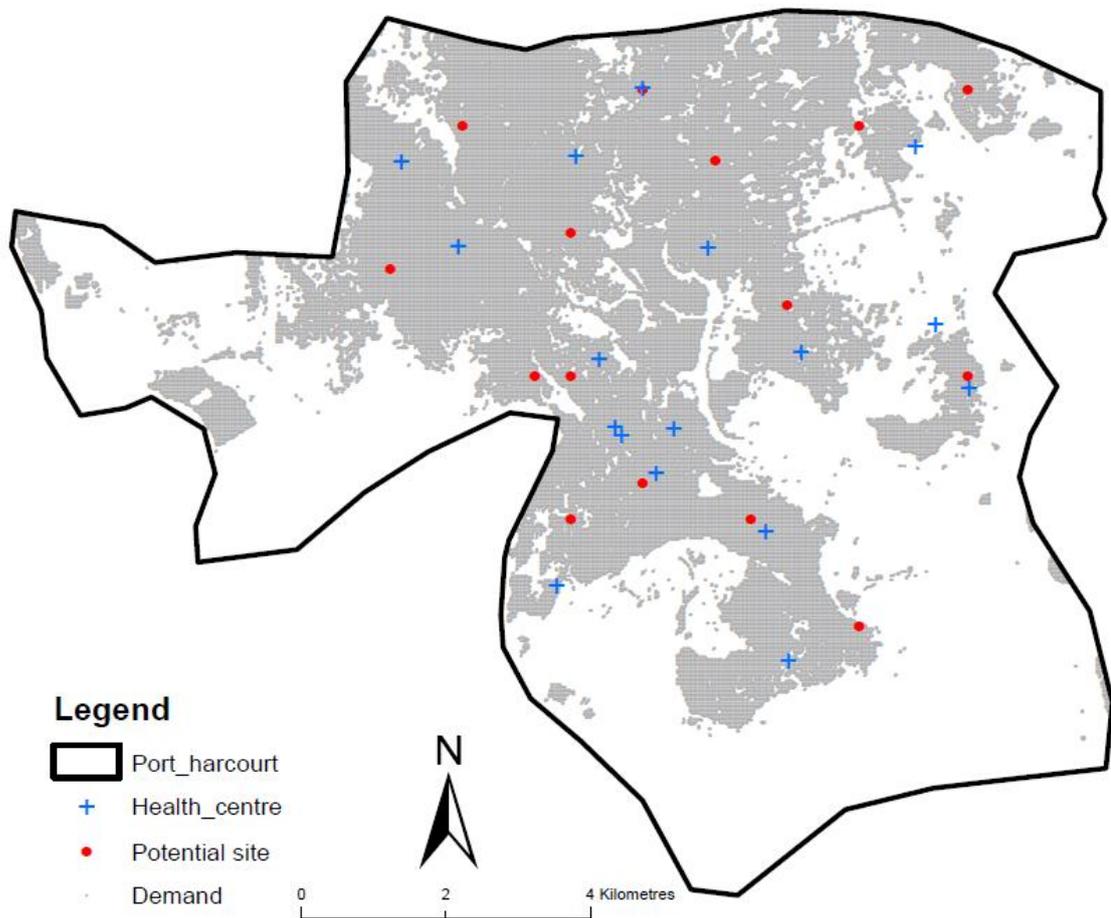


Figure A14.4 - The spatial distribution of 15 optimal locations (red circles) selected from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geo-technics Services 2011.

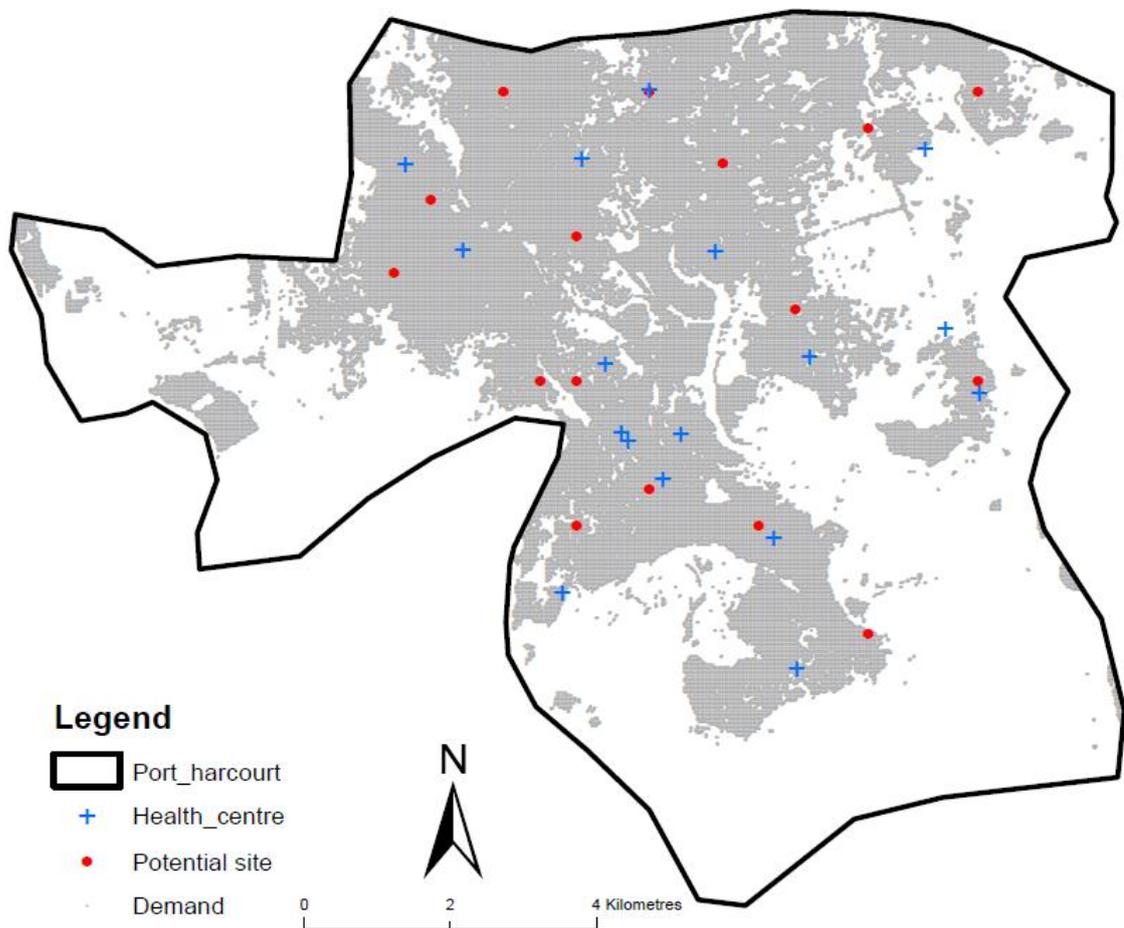


Figure A14.5 - The spatial distribution of 16 optimal locations (red circles) selected from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geo-technics Services 2011.

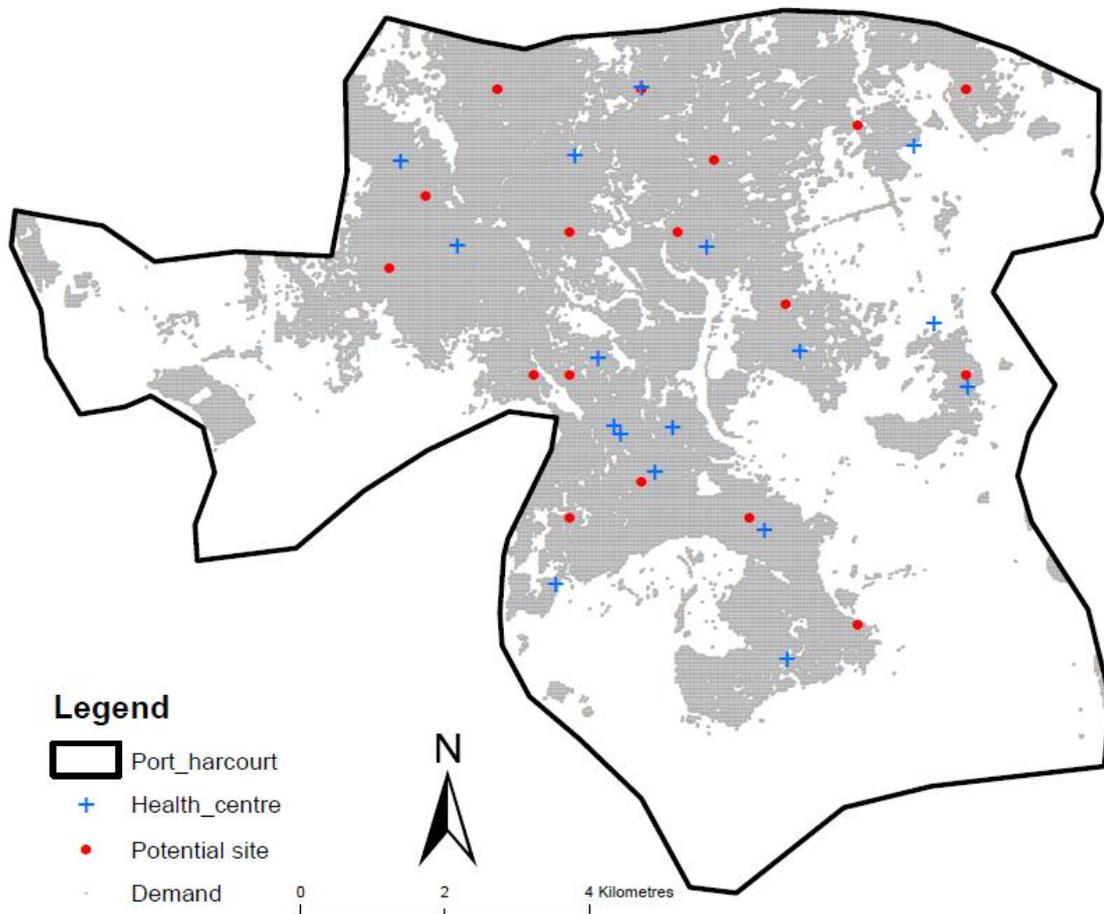


Figure A14.6 - The spatial distribution of 17 optimal locations (red circles) selected from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geo-technics Services 2011.

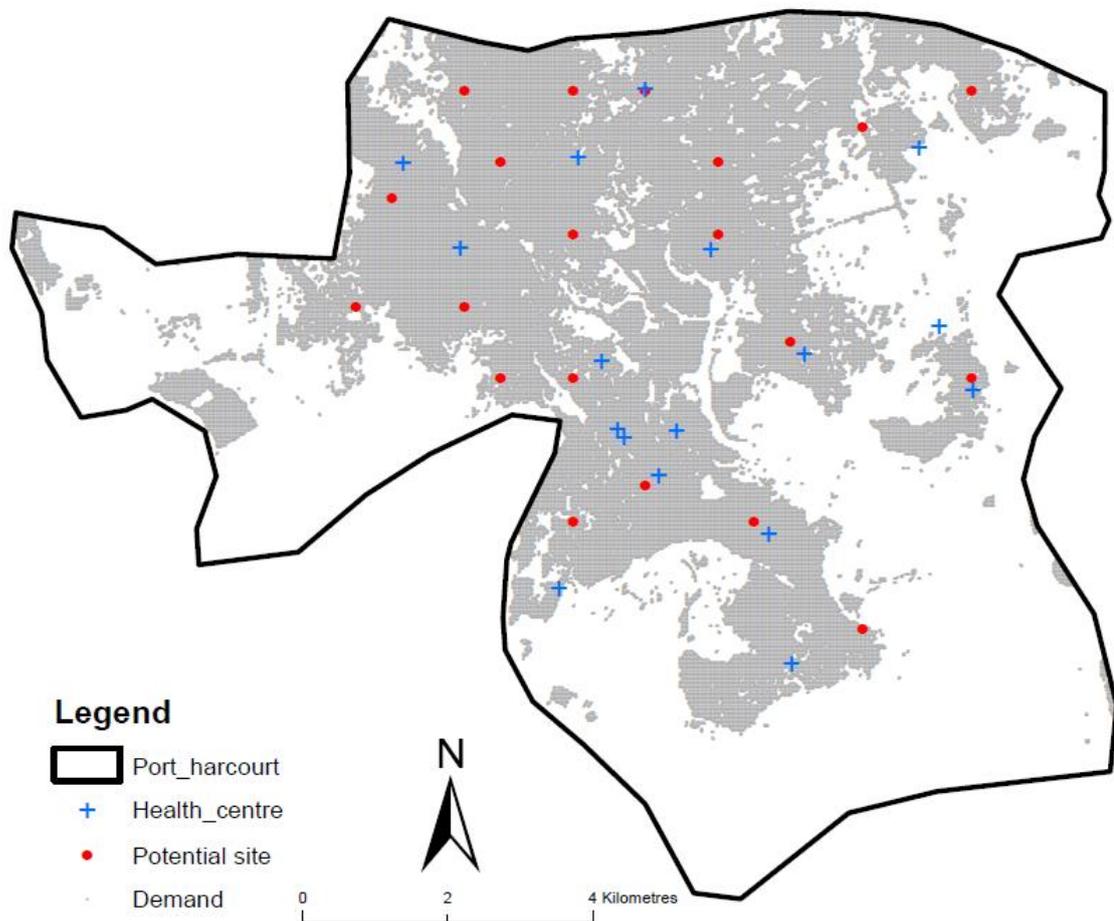


Figure A14.7 - The spatial distribution of 20 optimal locations (red circles) selected from 85 potential locations and 17 current locations of PHCCs (blue crosses). The digital boundary is Copyright for Geo-technics Services 2011.

## Appendix 15: Codes for locating specified number of sites for current and potential locations.

Some of the functions were obtained from Professor Alexis Comber and modified to suit the objectives of this research.

```
#load the libraries needed for the analyses
library(GISTools)
#load the source code and support functions for the GGA
source("permute.r")
source("jegafunctions.r")
#Load the source code for the Tietz-Bart algorithm
source('pmedians.R')
#set working directory to the right folder and read in the data to be
used
demand <- readShapePoints("Ph30m_30mgrids_points.shp")
ph <- readShapePoly('port_harcourt.shp')
roads <- readShapeLines('PHCRoads.shp')
supply <- readShapePoints("500m_points.shp")
#read in the OD Matrix csv file from ArcGIS
access2 <- read.csv('potential_access.csv',as.is=T)
head(access2)
#attach the dataset (access2) to make it available to memory and refer
to each variable by name
attach(access2)
#Remove the comma's in OriginID and Total length from access2 file and
name it access
access <- data.frame(ObjectID = ObjectID, Name = Name,
OriginID=as.numeric(gsub(",","",OriginID)),DestinationID =
DestinationID, DestinationRank = DestinationRank,
Total_Length=as.numeric(gsub(",","",Total_Length)))
detach(access2)
head(access)
attach(access)
#calculate the distance matrix by grouping items together from the
first category, converting the list to an array such that when the
file is attached column names becomes variables using an identify
function (function(x) x) that returns what is given
d.mat = tapply(Total_Length,list(OriginID,DestinationID),function (x)
x)
#check the dimension of the distance matrix
dim(d.mat)
#detach the dataset (access)
detach(access)
#select the estimated populations and call it pops
pops <- demand$estimates
#check the length of pops
length(pops)
#use only the lines that were used by the OD Matrix
pops <- pops[1:54072]
#check the quality of the data for holes
holes <- spot.holes(d.mat)
holes
#run the Teitz-Bart algorithm with 17 health facilities to allocate
from 85 potential sites
tb.best.loc <- tb(1:12,d.mat,pops)
```

```

#select the best subset and call it selected
selected <- tb.best.loc
length(selected)
#display the best subset from the supply data and call it optimal
point
optimal.points <- supply[selected,]
#display the non optimal points
not.optimal.points <- supply[-selected,]
#classify the health centres interms of distance - that is the minimum
distance to each demand
dist <- classify(selected, type = "dists")
#name the health centre allocated to each demand
names <- classify(selected)
#put the result in a data frame
results <- data.frame(healthcentre=names, Distance=dist, demand=pops)
#obtain the total population (demand) allocated to each health centre
- rowDemand
ans1 <- tapply(results$demand,results$healthcentre,sum)
#obtain the mean distance from demand to health centre within the
catchment
ans2 <- tapply(results$Distance,results$healthcentre,mean)
#obtain the maximum distance from demand to health centre within the
catchment
ans3 <- tapply(results$Distance,results$healthcentre,max)
#put the results in a data frame
ans4 <-
data.frame(Health_centres=rownames(ans3),Demand=ans1,mean.dist=ans2,
maximum=ans3)
#Show the demand allocation in terms of percentage of the total demand
#divide the demand by its sum and call it newDemand
newDemand <- ans4$Demand/sum(ans4$Demand)
#multiply newDemand by 100 to show each demand as a percentage of
total demand
Demand_percent <- newDemand * 100
#add Demand_percent to ans4
ans4_results <- cbind(ans4,Demand_percent)
#write it out as a csv file
write.csv(ans4_results,'12potentialsites.csv')

```

## Appendix 16: Permission to use pycno illustration from Uwe Deichmann

**From:** Uwe Deichmann [mailto:udeichmann@worldbank.org]  
**Sent:** 29 January 2015 01:15  
**To:** Jega, Idris M.  
**Subject:** Re: Pycnophylactic Interpolation - Request for permission  
Dear Idris,

Please feel free to use the figure. I'm very glad to hear that you find it useful.

Best,  
Uwe

**From:** Jega, Idris M. <[ijm14@leicester.ac.uk](mailto:ijm14@leicester.ac.uk)>  
**Sent:** Wednesday, January 28, 2015 4:22 PM  
**To:** Uwe Deichmann  
**Subject:** Pycnophylactic Interpolation - Request for permission

Dear Uwe Deichmann,

I am Idris Jega, a PhD candidate at the Department of Geography, University of Leicester, UK.

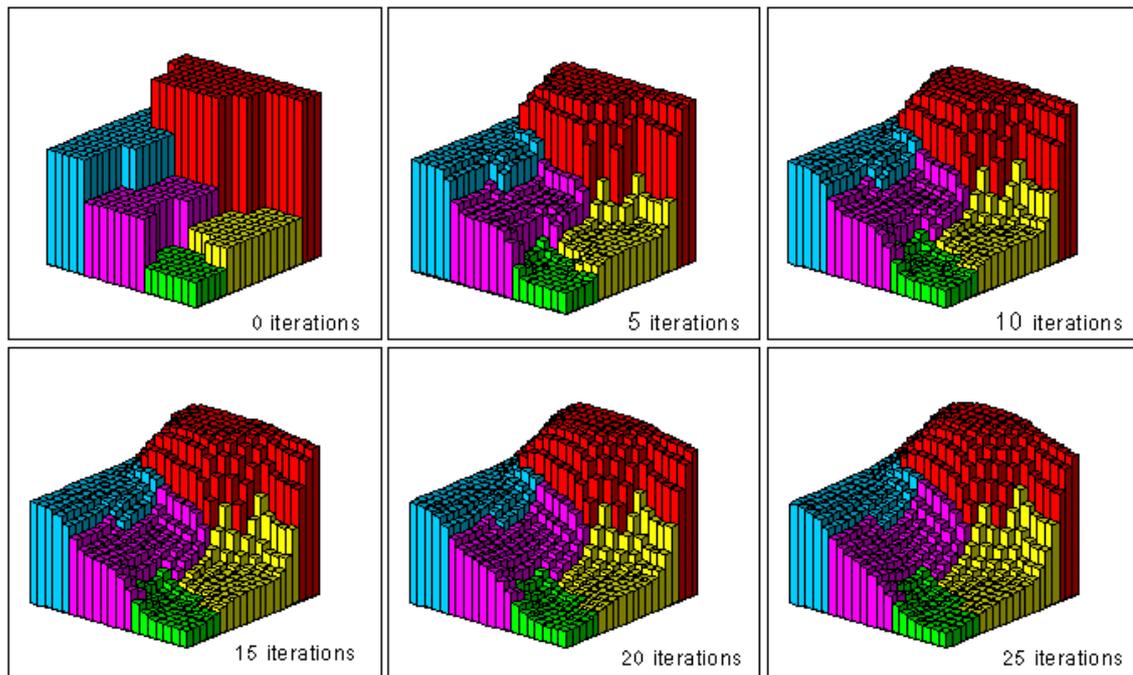
My thesis is looking at the challenges of applying areal interpolation techniques to a region where population data are less readily available (such as Nigeria). It involves creating population surfaces using different interpolation methods at different spatial scales.

Please I sent a request for permission to use the image below to describe the basic principles of pycnophylactic interpolation to Waldo Tobler who mentioned that the illustration below was likely produced by you.

I defended my PhD thesis on the 30<sup>th</sup> October 2014 (Mitch Langford was my examiner) and part of the recommended amends to my thesis is to use a figure to illustrate the pycnophylactic interpolation method. Langford suggests I use this illustration.

Please I am requesting for your permission to use the illustration below in my PhD thesis.

Thank you,  
Idris.



Idris

**From:** Waldo Tobler [<mailto:wtobler@earthlink.net>]  
**Sent:** 28 January 2015 19:51  
**To:** Jega, Idris M.  
**Subject:** Re: Pycnophylactic Interpolation - Request for permission

Idris,  
 Thank you for your query. Your citations seem correct. Unfortunately the illustration shown, although it gives a very nice interpretation of what goes on, was not produced by me but rather a student. I think it was Uwe Deichmann, who worked on the global demography project with me. I don't think that he copyrighted it. I think your version came from one of my power point presentations. Did you get it from my CD sent to one of your professors?  
 Waldo

-----Original Message-----

From: "Jega, Idris M."  
 Sent: Jan 27, 2015 4:01 PM  
 To: "tobler@geog.ucsb.edu"  
 Subject: Pycnophylactic Interpolation - Request for permission

Dear Waldo Tobler

I am Idris Jega, a PhD candidate at the Department of Geography, University of Leicester, UK.

My thesis is looking at the challenges of applying areal interpolation techniques to a region where population data are less readily available (such as Nigeria). It involves creating population surfaces using different interpolation methods at different spatial scales.

Please I am requesting for your permission to use the image below to describe the basic principles of pycnophylactic interpolation. May I please ask which reference best suits the image, the original reference (1979) or the 1992 proceeding as below?

Thank you for your time and consideration. I look forward to hearing from you.

Sincerely,  
Idris.

The original reference is

Tobler, Waldo R. (1979), Smooth pycnophylactic interpolation for geographical regions, *Journal of the American Statistical Association*, 74, 367:519-530.  
an adjustment for interpolation on the sphere is described in the NCGIA Technical Report TR95-6.

See also

Tobler, Waldo R. (1992), Preliminary representation of World population by spherical harmonics, *Proceedings of the National Academy of Sciences of the United States of America*, 89, 14:6262-6264.

Idris

## Appendix 17: Request for Spot5 (colour) 10m spatial resolution

**From:** White, Sheena [mailto:sheena.white@infoterra-global.com]  
**Sent:** 19 December 2012 09:03  
**To:** Comber, Alexis (Dr.)  
**Cc:** Mohammed, Idris J.  
**Subject:** RE: Satellite imagery request

Hi Lex

The data is ready to copy over from the ftp site for you.

You need to go to

<ftp://ftp.infoterra-global.com>

Then login as

Login leicsuniv  
Password dfkewogfj

Please let me know when you have successfully downloaded the data.

Many thanks and best regards

Sheena

**From:** Comber, Alexis (Dr.) [mailto:ajc36@leicester.ac.uk]  
**Sent:** 14 December 2012 12:07  
**To:** White, Sheena  
**Cc:** Mohammed, Idris J.  
**Subject:** RE: Satellite imagery request

Dear Sheena

Thank you for this. We are no longer requesting the imagery from Andrew Tewkesbury – we were pushing in to Astrium in 2 places – so apologies to all for that.

Of the KML you sent you we would like the subset of the 10<sup>th</sup> January 2007. We would like this at 1B Processing Level if possible.

Many thanks, Lex

Dr Alexis Comber  
Reader in Geographic Information  
Department of Geography  
University of Leicester  
Leicester, LE1 7RH, UK  
Tel +44(0)116 252 3812 / 3823  
Fax +44(0)116 252 3854  
Email [ajc36@le.ac.uk](mailto:ajc36@le.ac.uk)  
<http://www2.le.ac.uk/departments/geography/people/ajc36>

**From:** White, Sheena [mailto:sheena.white@infoterra-global.com]  
**Sent:** 14 December 2012 11:58  
**To:** [ajc36@leicester.ac.uk](mailto:ajc36@leicester.ac.uk)  
**Subject:** FW: Satellite imagery request

Hi

Thanks for your call. Please find the attached file which details the processing levels. I only have approval to supply a Level 1A, 1B or 2A scene.

Best regards

Sheena

**Sheena White**

Data Sales Manager | United Kingdom  
GEO-Information Services

**Astrium Services**

Europa House, The Crescent | Farnborough GU14 0NL | United Kingdom  
Tel +44 (0)1252 362080 | Fax +44 (0)1252 362012 | Mob +44 (0) 7799 437 122

[sheena.white@infoterra-global.com](mailto:sheena.white@infoterra-global.com) | [www.astrium-geo.com](http://www.astrium-geo.com)

>> Spot Image & Infoterra are teaming up. The **GEO-Information Services division of Astrium** will bring a consolidated portfolio of services & products to customers worldwide.



**From:** Comber, Alexis (Dr.) [<mailto:ajc36@leicester.ac.uk>]

**Sent:** 28 November 2012 09:44

**To:** Lamb, Alistair

**Cc:** Mohammed, Idris J.

**Subject:** Re: Satellite imagery request

Dear Alistair

we have finally narrowed down what we are looking for - 1/8th of a SPOT scene for Port Harcourt in Nigeria (kml files are in the attached).

My student was quoted ~£800 for this (see the attached correspondence between him and Astrium), however, we would be most grateful if you were able to get this scene for him.

Many thanks, Lex

Dr Alexis Comber  
Reader in Geographic Information  
Department of Geography  
University of Leicester  
Leicester, LE1 7RH, UK  
Tel +44(0)116 252 3812 / 3823  
Fax +44(0)116 252 3854  
Email [ajc36@le.ac.uk](mailto:ajc36@le.ac.uk)  
<http://www2.le.ac.uk/departments/geography/people/ajc36>