

**IMPROVING POPULATION ESTIMATION MODELS USING
REMOTELY SENSED AND ORDNANCE SURVEY DATASETS**

**Thesis submitted for the degree of
Doctor of Philosophy
At the University of Leicester**

by

**Mustafa Kose
Department of Geography
University of Leicester**

May 2015

Improving Population Estimates Using Remotely Sensed and Ordnance Survey Datasets

Mustafa Kose

ABSTRACT

The accuracy of population data is of critical importance in supporting the design of public and private-sector facilities. Demographic data are usually supplied by national census organisations at pre-defined census output levels. However, demographic datasets may be required at user-defined spatial units that can be different from the initial census output levels. A number of population estimation techniques have been developed to address these problems. This thesis is one of those attempts aimed at improving small-area population estimates by using spatial disaggregation models of: 1) binary mapping, 2) address-weighted dasymetric and 3) volumetric estimation models. These interpolation approaches employ high-resolution aerial imagery, LiDAR-derived building volumes and the integration of building address points and occupancy information sourced from the Ordnance Survey © and Airbus Defence and Space. Census wards and output areas were used as source zones and target zones respectively, to estimate population counts in Leicester City and the Borough of Kensington and Chelsea, London where the population is distributed both horizontally and vertically. The predicted population values were compared with 2011 census of actual population datasets. Each method employed in the study generated different population estimates depending on their assumptions and required datasets. The accuracy appears to be mainly influenced by the type and quality of the ancillary datasets and also the interpolation method adopted. Based on the disaggregation models adopted in this study, the address-weighted model produced the best population estimates with Root Mean Square Error (RMSE) value of 0.64 and R^2 score of 0.998 for the City of Leicester and RMSE value of 0.236 and R^2 score of 0.997 for the Borough of Kensington and Chelsea. This estimation is an indication that building address point datasets that contain information on occupancy can be used within Dasymetric mapping approaches to improve population estimates over a range of urban areas.

DEDICATION

This is dedicated to my mother Emine Kose and father Mehmet Kose who have supported me throughout the period of this study. I will always appreciate all they have done. I could not have made it so far without their constant support and dedication.

ACKNOWLEDGEMENT

My great thanks go to Allah, who gave me the knowledge, wisdom, ability and patience to complete this research.

It is a pleasure to thank those who made this thesis possible. My great thanks go to my supervisors, Dr Kevin Tansey and Dr Nicholas Tate for their supervision, outstanding support, extremely valuable comments, recommendations and the many useful contributions they made from the very infant stage of this thesis to completion. I would not have been able to achieve this milestone without your scientific guidance and countless help. I would like to thank my thesis committee for their great advice and constructive criticism. My PhD studies in the United Kingdom could not have been possible without the financial support from the Republic of Turkey Ministry of Education. I am highly grateful to this government for funding my studentship. My word of gratitude also goes to Andrew Tewkesbury, who assisted and facilitated in sourcing data from digital elevation models from Airbus Defence and Space (Astrium Ltd. 2014). I would like to thank Dr Murtala Chindo and Karl R. Sears for proofreading of the entire thesis. In addition, I would like to acknowledge the following:

- Ordnance Survey for providing Aerial imagery, address layer 2 data and the boundary of census geographies.
- Airbus Defence and Space (Astrium Ltd. 2014) for providing LiDAR derived digital elevation models
- Office for National Statistics for providing UK 2011 census data and Occupancy information.

My unconditional love, great appreciation and gratitude also goes to my family especially my mother, father, sisters and brother for their prayers and support for my success. Furthermore, I want to thank Aykan, Mehmet Veysi, Emre, Islam, Yahya, Mustafa Sukru and Ali Yavuz for their friendship. My sincere thanks go to Mohammed Idris, Bashir, Yahaya, Zamzam, Firdos, Claire and Paul for exchanging information and for been there in times of search of light relief. Finally, my other friends and all the staff in the Department of Geography deserve special praise for being there for me during my studies.

LIST OF ABBREVIATIONS AND ACRONMYS

CAS	Census Area Statistics
CEDS	Cadastral-based Expert Dasymetric System
DEM	Digital Elevation Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
ED	Enumeration Districts
EM	Expectation Maximization
ESRI	Environmental Systems Research Institute
ETM+	Enhanced Thematic Mapper Plus
GIScience	Geographical Information Science
GIS	Geographical Information System
IDM	Intelligent Dasymetric Mapping
IDW	Inverse Distance Weighting
JAC	June Agricultural Census
LIDAR	Light Detection and Ranging
LSOA	Lower Super Output Area
MSA	Metropolitan Statistical Area
MSOA	Middle Super Output Area
NeSS	The Neighbourhood Statistics
OA	Output Area
ONS	Office for National Statistics
OS	Ordnance Survey
PD	Parcel Distribution
PE	Percentage Error
PPH	Persons per Household
RBKC	Royal Borough of Kensington and Chelsea
RE	Relative Error
RGB	Red, Green, Blue
RMSE	Root Mean Square Error
ROI	Regent of Interest
S	the Standard Error of Regression

UK	United Kingdom
UPC	Unit Postcodes
US	United States
WHO	World Health Organisation

LIST OF PUBLICATIONS

1. Kose, M., Tansey, K. and Tate, N. (2014) Refining existing Dasymetric population Estimation Models. In: *the Remote sensing and photogrammetry society UK conference*. University of Aberystwyth 2nd and 5th September 2014. University of Aberystwyth, Aberystwyth.
2. Kose, M. Estimating population using high resolution building attribute information Poster session presented at: 10th University of Leicester Festival of Postgraduate Research; 2014 June 26; Leicester, UK.
3. Kose, M. Using high-resolution aerial imagery and LiDAR-Derived building volume information for population estimation. Poster presented at: the RSPSoc Annual Student Conference; 2012 March 14; University of Exeter, Falmouth, Cornwall.

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	xii
LIST OF TABLES	xx
CHAPTER 1	1
1 Introduction.....	1
1.1 Counting the Population.....	2
1.2 Study Aim	3
1.3 Thesis structure	4
CHAPTER 2	5
2 Population Estimation Methods: A Review	5
2.1 Introduction	5
2.2 Population Estimation and Population Distribution Mapping	5
2.3 Small-area Population Estimates.....	6
2.4 Census and Definitive Estimation Methods.....	7
2.5 Areal Interpolation Methods	8
2.5.1 Basic areal interpolation methods.....	9
2.5.2 Dasymetric mapping methods	14
2.5.3 Statistical dasymetric models.....	26
2.6 Research Questions	38
2.7 Summary	39
CHAPTER 3	41
3 Research Methodology and Datasets	41
3.1 Introduction	41
3.2 Population Estimation Algorithms.....	44

3.2.1	The area-weighted technique	45
3.2.2	The binary dasymetric mapping technique	46
3.2.3	The volumetric estimation technique.....	48
3.2.4	The Address-weighted disaggregation technique	50
3.3	Study Area.....	53
3.3.1	Case study area 1: the City of Leicester	54
3.3.2	Case study area 2: the Borough of Kensington and Chelsea, London.....	55
3.4	Datasets	58
3.4.1	Source and target zones, 2011 census of population data.....	59
3.4.2	Aerial imagery data.....	60
3.4.3	Digital elevation models	60
3.4.4	OS MasterMap® VectorMap district data.....	61
3.4.5	OS MasterMap® Address Layer 2 data	61
3.4.6	Occupancy information.....	62
3.5	The Binary Dasymetric Mapping Data Implementation.....	63
3.5.1	Remote sensing image pre-processing.....	63
3.5.2	Land cover classification –Supervised classification	67
3.5.3	Accuracy assesmmnt of land cover classifcation.....	70
3.5.4	The creation of binary zones -Reclassify land cover map-.....	71
3.6	The Generation of Building Volumes	74
3.6.1	Building height information	74
3.6.2	The creation of residential building volumes	79
3.6.3	The generation of population weighting factor of residential building volumes 80	
3.7	Building Address Point Datasets.....	81
3.7.1	Separation of residential buildings and non-residential structures	81
3.7.2	The population weighting factor of residential address points	82
3.8	Software Used for the Research	83
3.9	Accuracy Measurements for Estimation Results	84
3.10	Summary	85
CHAPTER 4.....		87
4	Results of areal interpolations for Leicester City	87
4.1	Introduction	87

4.2	The Results of the Areal Interpolation Approaches	88
4.2.1	Areal weighting interpolation method	88
4.2.2	Binary dasymetric mapping method	92
4.2.3	The volumetric estimation method	98
4.2.4	The address-weighted dasymetric method.....	103
4.2.5	The occupancy-adjusted address-weighted dasymetric method.....	107
4.3	Comparison of Areal Interpolation Approaches	111
4.3.1	Comparison of the areal weighting and the binary dasymetric method .	111
4.3.2	Comparison of volumetric estimation and the address-weighted methods	117
4.4	Summary	124
CHAPTER 5.....		126
5	Areal interpolation results for the Borough of Kensington and Chelsea	126
5.1	Introduction	126
5.2	The Results of the Areal Interpolation Approaches.....	127
5.2.1	The areal weighting interpolation method	127
5.2.2	The binary dasymetric mapping method	132
5.2.3	The building volume based dasymetric method	137
5.2.4	The address points based dasymetric mapping method.....	141
5.2.5	The address-weighted method with occupancy information	145
5.3	The Comparison of Areal Interpolation Approaches	149
5.3.1	Comparison of areal weighting and the binary dasymetric methods.....	150
5.3.2	Comparison of the volumetric and the address-weighted methods	155
5.4	Summary	162
CHAPTER 6.....		164
6	Discussion	164
6.1	Introduction	164
6.2	Discussion of the Results of Existing Approaches	165
6.3	Discussion of the Results of Novel Approaches	169
6.4	Summary	173
CHAPTER 7.....		175

7	Conclusion	175
7.1	The key Findings of the Disaggregation Process.....	176
7.2	Research question re-visited	179
7.3	Contribution of This Research	181
7.4	Limitations of the Study.....	182
7.5	Recommendations for Future Work.....	182
	Appendices	199

LIST OF FIGURES

Figure 2.1: The comparison of UK census Dissemination Units by showing 1991 Enumeration Districts (ED), 2001 Output areas, and 2011 Output areas by showing a sample from the City of Leicester. © Crown Copyright / database right 2013, An Ordnance Survey/EDINA supplied service.	8
Figure 2.2: An illustration of population distribution models: a) in a source area population distributed uniformly, b) a source zone classified as populated (grey) and unpopulated (white) areas, c) population distributed within populated areas of a source zone (the binary dasymetric population mapping model (modified from Langford, 2007)).	20
Figure 2.3: An illustration of three class population distribution model: a) an administrative area assuming population totals shows uniform pattern and a) the distribution of population according to the weighting factor of each urban, suburban and rural areas.....	22
Figure 2.4: The concept of image texture dasymetric method: a) Source zone and b) Raster dasymetric model (Red pixels represent populated area and green pixels represent unpopulated areas).	25
Figure 2.5: Methodological differences and potential improvement of population estimation of the CEDS method (c), over both filtered areal weighting (b), simple areal weighting (a) (modified from Maantay et al., 2007).	32
Figure 2.6: Schematic example of the address-weighted interpolation method. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	34
Figure 2.7: Graphical illustration of the building volume based population estimation model; a) represents all building footprints and b) represents all high-rise and low-rise building blocks for creating building volume.....	37
Figure 3.1: The graphical representation of methodological sequences of the interpolation techniques.....	43

Figure 3.2: The steps of the binary dasymetric interpolation technique.....	48
Figure 3.3: Implementation steps for the volumetric areal interpolation technique.....	50
Figure 3.4: Implementation steps for the address-weighted dasymetric interpolation technique.....	53
Figure 3.5: Study area showing: (a) the location map of Leicestershire County within the UK, (b) local authority of Leicester City and (c) the source zones (census wards) in Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.	56
Figure 3.6: Study area showing: (a) the location map of Greater London Administrative area within the UK, (b) the Borough of Kensington and Chelsea, and (c) the source zones (census wards) in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	57
Figure 3.7: Ancillary datasets of aerial photography that used to undertake binary dasymetric mapping methods. (a) is one of the 94 tiles of 0.25 meter resolution aerial photograph of Leicester City cover 1 km ² area and (b) is one of the 27 tiles of 0.25 meter resolution aerial photograph of the Borough Kensington and Chelsea. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.....	64
Figure 3.8: The subset imagery of 25 cm spatial resolution aerial photography of study area, Leicester City with image band combination of Red, Green and Blue. The enhanced imagery used to employ the binary dasymetric mapping method. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.....	65
Figure 3.9: The subset imagery of 25 cm spatial resolution aerial photography of study area; the Borough of Kensington and Chelsea with image band combination of Red, Green, Blue. The enhanced imagery used to employ the binary dasymetric mapping method. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.	66
Figure 3.10: Results of classification. The classified image of aerial photography by using maximum likelihood supervised classification. The case site: Leicester City. ©	

Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.	68
Figure 3.11: Results of classification. The classified image of aerial photography by using maximum likelihood supervised classification. The case site: the Borough of Kensington and Chelsea. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.....	69
Figure 3.12: The binary dasymetric zones created from classified aerial imagery, case site: Leicester City. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.....	72
Figure 3.13: The binary dasymetric zones created from classified aerial imagery, case site: The Borough of Kensington and Chelsea. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.	73
Figure 3.14: 1 meter resolution of digital surface model and digital terrain model of Leicester City, 1km ² sample area, All LiDAR data ©Airbus Defence and Space (Astrium Ltd. 2014).	75
Figure 3.15: 1 meter resolution of digital surface model and digital terrain model of the Borough of Kensington and Chelsea, 1km ² sample area, All LiDAR data ©Airbus Defence and Space (Astrium Ltd. 2014).	76
Figure 3.16: Normalised Digital Surface Model (nDSM) for the Borough of Kensington and Chelsea (DSM was subtracted from DTM by using spatial analyst tools in ArcGIS). All LiDAR data ©Airbus Defence and Space (Astrium Ltd. 2014).....	77
Figure 3.17: Normalised Digital Surface Model (nDSM) for Leicester City (the DSM covers 952 output areas of 969, digital elevation models are not available for 27 output areas on the northern part of the city. All LiDAR data © Airbus Defence and Space (Astrium Ltd. 2014)	78
Figure 3.18: The residential building footprints with height information. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service and All LiDAR data © Airbus Defence and Space (Astrium Ltd. 2014).	80

Figure 3.19: The extracted residential building points © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.....	82
Figure 4.1: The population estimation results of areal weighting interpolation method in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	89
Figure 4.2: Regression model for the areal weighting interpolation method. The estimated population totals of the City of Leicester at output area level versus the actual population of output areas released by census.....	91
Figure 4.3: Dasymetric surface of Leicester City obtained by using the binary dasymetric mapping method. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	94
Figure 4.4: Population estimation results of the binary dasymetric mapping model in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	95
Figure 4.5: Regression model for the binary dasymetric mapping interpolation method. The estimated population totals of the City of Leicester at output area level versus the actual population of output areas released by census.	97
Figure 4.6: Map of building volumes by output areas for the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	100
Figure 4.7: Population estimation results of building volume based dasymetric mapping method in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	101
Figure 4.8: Regression model for the building volume based dasymetric mapping interpolation method. The estimated population totals of the City of Leicester at output area level versus the actual population of output areas released by census.....	102
Figure 4.9: Population estimation results of address-weighted model in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.	104

Figure 4.10: Regression model for the address-weighted dasymetric mapping interpolation method. The estimated population totals of the City of Leicester at output area level versus the actual population of output areas released by census.....	106
Figure 4.11: Population estimation results of refined address-weighted dasymetric mapping method in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.	108
Figure 4.12: Regression model for the refined address-weighted dasymetric mapping method. The estimated population totals Leicester City at output areas level were plotted against the census released population of output areas.	110
Figure 4.13: Residual map of areal weighting method obtained population estimates in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	113
Figure 4.14: Residual map of the binary dasymetric mapping method obtained population estimates in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service	114
Figure 4.15: Spatial autocorrelation of the areal weighting method provided population residuals in Leicester City.....	116
Figure 4.16: Spatial autocorrelation of the binary dasymetric mapping provided population residuals in Leicester City.	116
Figure 4.17: Residual map of volume based dasymetric mapping method obtained population estimates in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service	118
Figure 4.18: Residual map of the address-weighted dasymetric mapping method obtained population estimates in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	119
Figure 4.19: Residual map of the occupancy information used address-weighted dasymetric mapping method obtained population estimates in the City of Leicester. ©	

Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service	120
Figure 4.20: Spatial autocorrelation of the volumetric estimation model obtained population residuals in Leicester City.	122
Figure 4.21: Spatial autocorrelation of the address-weighted dasymetric model provided population residuals in Leicester City.	122
Figure 4.22: Spatial autocorrelation of the address-weighted method using occupancy data obtained population residuals in Leicester City.	123
Figure 5.1: The population estimation results of areal weighting interpolation method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.	129
Figure 5.2: The regression model for the areal weighting interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.	130
Figure 5.3: Dasymetric surface of the Borough of Kensington and Chelsea obtained by using the binary dasymetric mapping method. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.	133
Figure 5.4: Population estimation results of the binary dasymetric mapping method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.	134
Figure 5.5: Regression model for the binary dasymetric mapping interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.....	136
Figure 5.6: Map of building volumes by output areas for the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.	138

Figure 5.7: Population estimation results of building volume based dasymetric mapping method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.	139
Figure 5.8: Regression model for the building volume based dasymetric mapping interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.....	140
Figure 5.9: Population estimation results of the address-weighted dasymetric mapping method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.	143
Figure 5.10: Regression model for the address-weighted dasymetric mapping interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.....	144
Figure 5.11: Population estimation results of the refined address-weighted dasymetric mapping method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	147
Figure 5.12: Regression model for the refined address-weighted dasymetric mapping interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.....	148
Figure 5.13: Residual map of areal weighting method obtained population estimates in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service	152
Figure 5.14: Residual map of the binary dasymetric mapping method obtained population estimates in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.....	153

Figure 5.15: Spatial autocorrelation of the areal weighting method provided population residuals in the Borough of Kensington and Chelsea	154
Figure 5.16: Spatial autocorrelation of the binary dasymetric mapping provided population residuals in the Borough of Kensington and Chelsea	155
Figure 5.17: Residual map of volume based dasymetric mapping method obtained population estimates in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service	157
Figure 5.18: Residual map of the address-weighted dasymetric mapping method obtained population estimates in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service	158
Figure 5.19: Residual map of the address-weighted dasymetric mapping method obtained population estimates using occupancy datasets in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service	159
Figure 5.20: Spatial autocorrelation of the volumetric estimation model provided population residuals in the Borough of Kensington and Chelsea	160
Figure 5.21: Spatial autocorrelation of the address-weighted dasymetric method provided population residuals in the Borough of Kensington and Chelsea	161
Figure 5.22: Spatial autocorrelation of the address-weighted method using occupancy data obtained population residuals in the Borough of Kensington and Chelsea.	161
Figure 6.1: The histogram of errors of estimated population for both study areas. a) Areal weighting, b) the binary mapping, c) the volumetric estimation, d) the address-weighted, and e) the address-weighted interpolation method with occupancy data inputs.	168

LIST OF TABLES

Table 2.1: Selection of previously implemented basic areal interpolation methods	10
Table 2.2: Selection of previously used traditional dasymetric mapping methods	17
Table 2.3: The sample ancillary classes of three-class interpolation method.....	21
Table 3.1: Additional datasets showing sources and characteristics	58
Table 3.2: Datasets used in implemented areal interpolation models.....	59
Table 3.3: The land cover classes of study area with four different groups	67
Table 3.4: Accuracy measurement of land cover classification of aerial photography (Case study area: Leicester City).	70
Table 3.5: Accuracy measurement of land cover classification of aerial photography (Case study area: The Borough of Kensington and Chelsea).	70
Table 4.1: Accuracy measurements for the areal weighting interpolation method obtained population estimates for the City of Leicester.	92
Table 4.2: The accuracy measurements for the binary dasymetric mapping method obtained population estimates for the City of Leicester.	98
Table 4.3: The accuracy measurements for the volumetric estimation method obtained population estimates for the City of Leicester.	103
Table 4.4: Accuracy measurements for the address-weighted dasymetric method obtained population estimates for the City of Leicester.	106
Table 4.5: Accuracy measurements for the improved address-weighted dasymetric method obtained population estimates for the City of Leicester.	110
Table 5.1: The accuracy measurements for the areal weighting interpolation obtained population estimates for the Borough of Kensington and Chelsea.....	131

Table 5.2: Accuracy measurements for the binary dasymetric mapping method obtained population estimates for the Borough of Kensington and Chelsea.....	136
Table 5.3: Accuracy measurements for the volumetric estimation technique obtained population estimates for the Borough of Kensington and Chelsea.....	141
Table 5.4: Accuracy measurements for the address-weighted dasymetric method obtained population estimates for the Borough of Kensington and Chelsea.....	145
Table 5.5: Accuracy measurements for the improved address-weighted dasymetric method obtained population estimates for the Borough of Kensington and Chelsea...	149
Table 6.1: The measurements of accuracy for population disaggregation results from census wards to output areas for the City of Leicester	167
Table 6.2: The measurements of accuracy for population disaggregation results from census wards to output areas for the Borough of Kensington and Chelsea.....	168

CHAPTER 1

Introduction

Since the 18th century Industrial Revolution, global population has been growing at a higher rate than the previous years (Li and Weng, 2005, Chen, 2002). The United Nations population statistics estimate shows that the world's population living in towns, cities and urban areas will continue to grow: it was 50% of total population in 2008 and has increased to over 54% (3.9 billion) in 2014 (World Health Organization - WHO). Population distribution changes rapidly because of population movements. Population totals have very strong interactions between the natural and social environment due to its movement and dynamism. According to Gerland et al. (2014) and Lu et al. (2006), increasing population in urban areas and rapid urban growth is not only potentially damaging to natural resources and environmentally sensitive areas, but can also create several environmental, social, economic and governance issues. Li and Weng (2005) observed that as population growth continues, it can cause severe environment pressures. For example, as urban areas are expanding, agricultural lands and forest areas have been decreasing at the same rate. Also, Lu et al. (2006) points out that population increase is closely associated with urban extension, the decrease of forest area, damaged agricultural land and the deformation of environmental conditions in the neighbourhood of urban area. As a way of reducing the severity of these problems, population estimation and its distribution are important for providing applicable land use planning and for analysing the relationship between human beings and the natural environment in a timely and accurate manner (Deng et al., 2010, Dong et al., 2010, Lu et al., 2010, Silván-Cárdenas et al., 2010, Lu et al., 2006).

In order to allocate resources, an understanding of the size and distribution of the population is generally essential for state and local governments (Deng et al., 2010, Smith et al., 2002). Consequently, accurate population data have a primary importance in supporting planning processes, including the design of public facilities and private sector facilities, such as site-location identification, sewage treatment plants, public transit route design, resource allocation, urban infrastructure planning, customer-profile analysis, health planning, social service allocation, social securities, industrial planning and water resource management (Deng et al., 2010, Silván-Cárdenas et al., 2010,

Hardin et al., 2008, Wu and Murray, 2007, Wu et al., 2005). All these suggest that the wealth of population totals and their distribution are required in order to address those aforementioned and other challenging problems.

1.1 Counting the Population

The main purpose of a population census is to accurately and authoritatively gain information on the number of people and to determine where they usually live (Valente, 2010, Wu et al., 2008, Smith and Mandell, 1984). Population information is usually derived from census data. On one hand, the census gives the fullest and most reliable characteristics of the population; on the other hand, census data collection is labour intensive, time-consuming and substantially expensive (Dong et al., 2010, Lu et al., 2010, Li and Weng, 2005). Moreover, census data becomes outdated a few years after the census enumeration because of the long intervals between censuses (Erener and Düzgün, 2009). As stated by Li and Weng (2005), if the gap between the conduct of two censuses is too long (5 or 10 years), population related planning is difficult to perform due to the lack of knowledge about the population counts in non-census years. Additionally, the census values are reported in pre-defined spatial output units due to confidentiality, privacy issues and nondisclosure requirements (Langford, 2013, Sridharan and Qiu, 2013, Su et al., 2010, Gregory, 2002). These census dissemination units may change from census to census and often the user defined study areas can be different from census report zones (Sridharan and Qiu, 2013). For instance, UK census output units have changed in the last three censuses and different administrative units have also been used. Therefore, the long-term (census to census) comparison of population counts may be crucially problematic because of the census geography alteration (Gregory et al., 2010, Gregory, 2002). For these reasons, researchers have started searching for new estimation methods to obtain an efficient means of population prediction.

In inter-censal years, traditional estimation methods use demographic records to estimate population totals at different scales (nation, region and city) in terms of the proposed target zones. For relatively small areas (e.g., census tracts at USA and census ward level, or local authorities in the UK), the method of counting housing units has been gradually applied for estimating population counts. Various population estimation methods have been described and implemented in the GIS and remote sensing literature

(Deng et al., 2010, Dong et al., 2010, Silván-Cárdenas et al., 2010, Wu et al., 2005). Li and Weng (2005) reported that the geographic study of population distribution and population estimation started in the 1930s. For example, Wright (1936) generated a population distribution map in Cape Cod, Massachusetts, US. Similarly, remotely sensed datasets have been used since the 1950s in order to provide population estimates (Silván-Cárdenas et al., 2010). Before the 1950s, aerial photographs were employed for counting individual housing units (Silván-Cárdenas et al., 2010, Lu et al., 2006). Later on, satellite images have become significant resources for obtaining spatially distributed population surfaces. Recently, both remote sensing images and GIS-based datasets have been combined and used as a source of ancillary data for generating small-area population estimates (Deng et al., 2010, Dong et al., 2010, Silván-Cárdenas et al., 2010). Remote sensing and GIS technology have been proven to provide accurate and inexpensive data and have shown promising results (Lu et al., 2011a, Li and Weng, 2005). However, Wu et al. (2005) emphasise that in the content of population estimation, various techniques are being improved based on remotely sensed image datasets. The effect of these approaches, as Li and Weng (2005) admitted, is changing the potential for the integration of satellite image data with census data for estimating population density. From the literature, explored in Chapter 2, it becomes clear that there is a need to resolve the new and improved ancillary datasets that can be used within interpolation and statistical modelling techniques. These methods include areal interpolation, dasymetric and spatial disaggregation. Traditionally, in the housing unit based population estimation, the numbers of houses were counted from aerial photographs for generating population estimates in non-census years (Dong et al., 2010, Silván-Cárdenas et al., 2010). This method was mainly based on visual detection and counting the individual housing units for estimating population totals (Smith and Cody, 1994). Qiu et al. (2010) claim that remotely sensed data involves extensive analysis of information on the dispersion of human settlements, which can be utilised as a possible indicator of population.

1.2 Study Aim

Given the projected future increase in human population, the importance of cities and the pressure on governments to provide adequate housing for populations to occupy it is absolutely critical that there are ways to estimate population over well characterised

areas in inter-census periods. The knowledge of accurate population counts and inter-censal population estimates are also important for planners, policy makers and researchers for the purpose of successful planning processes in inter-census years. Due to the lack of accurate census data and the non-coterminous census output units, the population estimation approaches and projections becomes more of an issue to generate population estimates in an intended scale. Therefore, the aim of the study is to use novel ancillary datasets for the disaggregation of population totals in order to estimate population more accurately. The proposed interpolation models require different sources of ancillary datasets for producing population estimates. The research also examines the performance of the variety of novel geographical datasets obtained from remote sensing, Ordnance Survey, Airbus Defence and Space, and local councils as control variables in dasymetric population modelling, by comparing the predicted results with census derived population figures. The methods are applied over two diverse and complex urban areas, namely Leicester City and the London Borough of Kensington and Chelsea. Results obtained are compared with census data for the two regions.

1.3 Thesis Structure

This thesis consists of seven chapters. The current chapter introduces the thesis and deals with the main research issues. It introduces the problems being investigated and the aim of the study, including how the thesis is organised. Chapter 2 provides a review of literature relevant to this study including population estimation methods, GIS and remote sensing based estimation methods, and a review regarding areal interpolation techniques that have been used to disaggregate population counts was also presented. The research questions that emerged from the literature are also contained in this chapter. Chapter 3 describes population disaggregation approaches, the study areas, the datasets used in the study, and it provides the results of remotely sensed data analysis to derive land use. In addition to this, the process of building volume calculation and the extraction of residential housing units are detailed in Chapter 3. The results from the application of these estimation methods to the region of Leicester City and the Borough of Kensington and Chelsea are presented in Chapter 4 and Chapter 5 respectively. The results of interpolation based population estimates presented in the previous chapters are discussed in Chapter 6. General conclusions drawn from the research questions are considered in Chapter 7 along with recommendations for future research.

CHAPTER 2

Population Estimation Methods: A Review

2.1 Introduction

In this chapter, the processes of population estimation and areal interpolation techniques are reviewed. These interpolation approaches have been used in order to spatially disaggregate the aggregated population totals from one spatial unit to other areal units. The disaggregation frameworks may vary based on the model functioning and the available ancillary input datasets. In this chapter, the process of population interpolation and the key issues of population estimation are highlighted. Furthermore, the chapter focuses on previous studies where interpolation techniques were implemented for providing small-area population estimates. This chapter is structured as follows: the first section (Section 2.2) emphasises why population distribution modelling and population estimation is important. The importance of disaggregation of population totals from larger areal units to smaller spatial units is provided in Section 2.3. The UK census is investigated and the census based population distribution approaches are explained in Section 2.4. In Section 2.5, areal interpolation techniques that have been used in previous studies are reviewed. The questions developed from the review of previous sections (2.2, 2.3, 2.4 and 2.5), used in meeting the aims of the study are outlined in Section 2.6. In the final section (Section 2.7), a summary of spatially population disaggregation models is provided.

2.2 Population Estimation and Population Distribution Mapping

Accurate population information is important for a variety of socio-economic applications involving urban planning, management of natural resources and environmental risk assessment (Lu et al., 2010). The information of characteristics of population can be a primary base for accurate resource allocation in areas such as community infrastructure development, the provision of recreational opportunities, transportation and environmental facilities (Maantay et al., 2007). Additionally, precise urban population distribution is one of the main components of a wide variety of

planning processes to locate private and public facilities (Lu et al., 2010). Moreover, population distribution at a finer scale is extremely important for the analysis of human beings' interaction with their regional socioeconomic and physical environments (Ural et al., 2011). Traditionally, population information is mostly procured throughout a census (Dong et al., 2010). These census-derived datasets store various types of demographic information and population-related information (Lu et al., 2010). In most countries, such as the United States and United Kingdom, census data is collected every decade. In Japan, Canada and Australia, for instance, censuses are conducted once every five years. Regardless of the differences between countries' frequency of census, Lu et al., (2010) noted that accurate population information does not exist within the inter-censal times (Maantay et al., 2007). Therefore, new population estimation models are needed that provide complete, informative and accurate inter-censal data to achieve the above objectives. To this effect, many methods have been introduced for population estimation using remote sensing and geographical information systems (GIS) (Wu et al, 2005). These approaches are explained further in the following section (2.5).

2.3 Small-area Population Estimates

Small-area population estimates and other demographic variables are necessary for the analysis of different datasets and efficient integration of geographic information system (GIS) models (Leyk et al., 2013b, Langford, 2013, Schmid and Münnich, 2013, Deng et al., 2010, Leyk et al., 2010). The national census data represent the “gold standard”, portraying the characteristics and spatial distribution of a country's population (Langford, 2013, page 324). Population counts are collected for non-modifiable entities (housing units) (Openshaw, 1984), and counted population values are released for a predefined set of geographic units instead of individual census records due to the statutory obligations, administrative convenience and privacy issues (Langford, 2013, Sridharan and Qiu, 2013). These areal zones are arbitrary and modifiable. As a result, census derived population dissemination units may be different from the user's desired geographic units. For instance, environmental researchers may require population datasets for watersheds, land cover parcels and vegetation zones. Socioeconomic research may need demographic data for tax zones, postal delivery zones, and facility service areas. In nature, boundaries of natural phenomena may seldom overlap with census dissemination geographies (Sridharan and Qiu, 2013). Therefore, obtaining

population data for the user-defined areal units may be highly problematic. Also, the process of population disaggregation is most likely subject to the modifiable areal unit problem (MAUP). This is because the results of statistical analysis usually depends on the scale and the pattern of areal zones used (Flowerdew, 2011, Openshaw, 1984). When geographical data is integrated to another zone, different results may be obtained depending on the difference in interpolation methods. Consequently, population estimation and redistribution techniques have been used to acquire population values in desired areal zones.

2.4 Census and Definitive Estimation Methods

Census population data are one of the fundamental datasets for most population related studies (Martin, 2004, Robinson, 2004). The census is based on the counting of total population and involves gathering information on household statistics (Office for National Statistics- ONS in the UK). The main purpose of the census is to accurately and authoritatively count the number of people and to determine where they usually live (Martin et al., 2013, Martin, 2011, Valente, 2010, Wu et al., 2008, Smith and Mandell, 1984). Planning and conducting a census is a large-scale operation aimed at providing an accurate population count on Census Night (Martin, 1996). In order to make population counting as easy as possible in the UK, the country is divided into hierarchical census geographies (Martin, 2004). The largest geographical level is a county as a whole, whilst the output area is the lowest geographical level at which population is counted and census results are released. The remaining four levels which are the most useful include those of census wards, Middle Layer Super Output Area (MSOA), Lower Layer Super Output Area (LSOA) and Output area (Martin, 1996).

Since 1801 a census has been collected in Britain, and the detailed population attribute data are reported at district-level since 1851 (Gregory, 2002). It can be critically difficult to compare national censuses with each other due to the changes in the boundary of census collection geographies (Langford, 2007, Gregory 2002, Flowerdew and Green, 1992) (see, Figure 2.1). Censuses can be published using different administrative units over the long term. For instance, although national census data have been collected in every decade since 1801 in Britain, in some cases (i.e., after 1911 and 1974) a completely different system was used (Gregory and Ell, 2005, Gregory, 2002). As Gregory (2002) states, the hierarchical census geography generally remains the

same, however, there have been multiple boundary alterations in the intervening decades. The 2001 United Kingdom census used a completely different collection and dissemination of geographic units based on enumeration districts and output areas respectively (Langford, 2007).

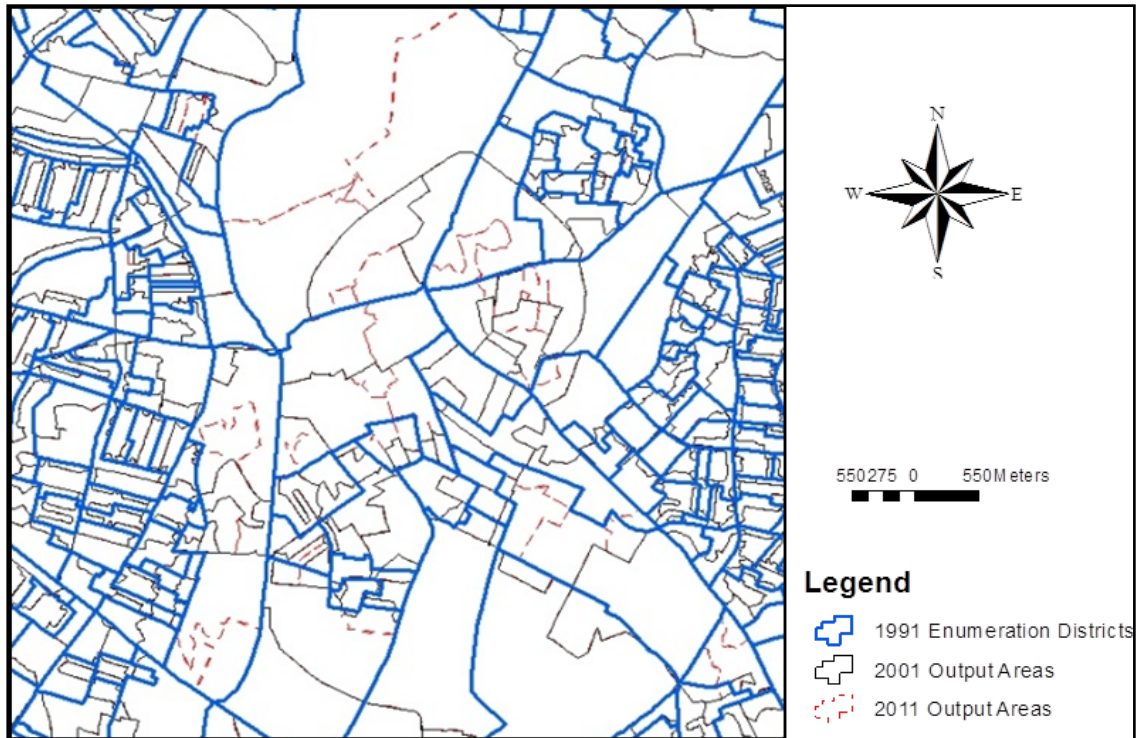


Figure 2.1: The comparison of UK census Dissemination Units by showing 1991 Enumeration Districts (ED), 2001 Output areas, and 2011 Output areas by showing a sample from the City of Leicester. © Crown Copyright / database right 2013, An Ordnance Survey/EDINA supplied service.

2.5 Areal Interpolation Methods

The areal interpolation is a geographical data transformation from source zones to target zones (Mennis, 2003). The process of areal interpolation can provide population estimates for target zones based on the known population of source zones (Langford, 2007, Mennis and Hultgren, 2006b, Goodchild et al., 1993). Studies of interpolation methods have shown that population distribution methods have been generally classified into three categories (Langford, 2013, Kim and Yao, 2010, Mennis, 2009, Maantay et al., 2007, Hawley and Moellering, 2005, Wu et al., 2005, Eicher and Brewer, 2001): basic areal interpolation methods (do not require ancillary data), dasymetric mapping techniques (do make use of external data inputs) and statistical

modelling. How these methods work is presented in Sections 2.5.1, 2.5.2 and 2.5.3 respectively. There are, however, other interpolation techniques aimed at improving the accuracy of population estimation (Mennis, 2009, Goodchild et al., 1993, Flowerdew, 1991, Flowerdew and Green, 1989).

2.5.1 Basic areal interpolation methods

This section presents two interpolation methods that do not make use of ancillary data: the areal weighting and the pycnophylactic interpolation frameworks. The previous articles using these approaches are reviewed in Table 2.1. The method of areal interpolation is a common technique for quantifying disaggregated population values (Maantay et al., 2007). Initially, the technique has been defined as the process of transferring the spatial data of interest from one set of areal units (source areas) to a different set of areal units (target areas) (Gregory, 2002, Eicher and Brewer, 2001, Bloom et al., 1996, Langford and Unwin, 1994). In essence, this approach was improved as a statistical method (e.g. the re-distribution population in a geographical unit) rather than for map production (Eicher and Brewer, 2001). All interpolation methods proposed to interpolate statistical variables, such as population counts, from a fixed output geography to the other finer areal geographies are based on specific assumptions (Sridharan and Qiu, 2013). The interpolation process is required when user-defined spatial units are incompatible with the set of census output geographies (Su et al., 2010, Flowerdew and Green, 1992). Theoretically, areal interpolation can compare multiple geographical datasets that are derived from distinct enumeration zones (Mennis, 2003, Eicher and Brewer, 2001). The approach enables the transfer of an individual dataset to a common set of enumeration units (e.g. census tract) that allows effective analysis and comparison (Eicher and Brewer, 2001).

Both areal weighting and pycnophylactic frameworks are volume preserving methods (Kim and Yao, 2010, Lam, 1983) which means that the original variables of interested source zones and the total estimated values of target zones are equal. This framework is closely related to dasymetric mapping of population densities (Maantay et al., 2007, Holt et al., 2004). The major difference between dasymetric mapping techniques and basic areal interpolation methods is that dasymetric mapping excludes the final step in areal interpolation of re-aggregating to a desired enumeration unit type (Eicher and Brewer, 2001). The areal interpolation method re-aggregates population data into a

preferred enumeration unit; however, the dasymetric mapping approach re-distributes population data into defined land use/cover groups (Maantay et al., 2007). Areal weighting methods and pycnophylactic interpolation methods are reviewed in the following sub-sections.

Table 2.1: Selection of previously implemented basic areal interpolation methods

Method	Ancillary Variable	Distribution of Variables	Discussed and applied by:
1. Areal Weighting	None	Homogenous	Briggs, Gulliver, Fecht & Vinneau, 2007 Brinegar & Popick, 2010 Eicher & Brewer, 2001 Fisher & Langford, 1996 Gregory, 2002 Hawley & Moellering, 2005 Langford, 2013 Maantay, Maroko & Hermann, 2007 Reibel & Bufalino, 2005
2. Pycno-phylactic	None	Heterogeneous	Comber, Proctor & Anthony, 2008 Hawley & Moellering, 2005 Kim & Yao, 2010

2.5.1.1 The areal weighting method

The areal weighting approach is one of the most common types of interpolation technique to have been used for population distribution (Sridharan and Qiu, 2013, Su et al., 2010, Eicher and Brewer, 2001, Cockings et al., 1997, Fisher and Langford, 1996, Goodchild et al., 1993, Goodchild and Lam, 1980). This method is referred to as the ‘area-weighted interpolation technique’ by Sridharan and Qiu (2013). The technique is the most basic form of interpolation methodology to quantify the values of interested an variable by a ratio obtained from the relative areal measurements of the source and target zones (Maantay et al., 2007, Mennis and Hultgren, 2006b, Fisher and Langford, 1996, Goodchild and Lam, 1980). This methodology is simply based on the geometric intersection of the source and target zones and the method does not need ancillary data for interpolation process (Langford, 2006, Goodchild et al., 1993). The areal interpolation method is mainly based on the presumption that variables (i.e. population) are distributed evenly through the source zone (the original geographical unit of data aggregation) (Maantay et al., 2007). Therefore the total values of target zones which fall into one or more source zones can be estimated by the ratio of a predefined target zone to source zone. This approach has been commonly applied to map the distribution of

population in target zones (Liao et al., 2010). The amounts of population are assigned to the target zones according to the proportion of targets zones which fall into source zones (Langford, 2007, Maantay et al., 2007, Langford, 2006).

Goodchild et al. (1993) used areal interpolation to determine the socioeconomic variables (such as population, employment and income) of predefined target areas in 58 counties of California assuming that variables were distributed homogeneously in the source areas (the counties). When the authors compared the result of the areal weighting technique with other techniques applying statistical methods, they found that the areal weighting method had a higher mean percentage error than the other statistical techniques (residential units and cadastral-based expert dasymetric methods). Similarly Hawley and Moellering (2005) used the areal weighting interpolation method in order to distribute population totals from Census Tract level to Census Block Group level in the counties of Franklin, Hamilton and Jefferson, USA. The authors compared the implemented disaggregation methods and they found that the areal weighting method consistently produced the least accurate estimation results in terms of accuracy measurements. Very recently, Langford (2013) used the areal weighting method to disaggregate the aggregated population counts from lower super output areas (LSOAs) to output areas, and from output areas to UK unit postcodes (UPCs), in order to evaluate the performance of interpolation methods in Cardiff City, Wales. Considering the distribution of population totals from LSOAs to output areas, the areal weighting method provides the worst population estimates with the applied approaches. Regarding the distribution of population values from outputs to UPCs, it shows that the area-weighted method performed better than the street weighted and surface model (used primary schools and bus stops) interpolation approaches.

Maantay et al. (2007) conducted research on areal weighting applications and revealed that one of the main limitations of the areal weighting technique is that, naturally, population is not distributed homogeneously through a geographical unit. In the source zones, large areas could be unpopulated either because of the existence of water bodies, green spaces and industrial areas or due to the different housing types in the zone - one part of the zone may have high-rise building units and the other may contain low-rise single family houses (Maantay et al., 2007). Thus, separating the uninhabited areas from inhabited areas - rather than assuming homogeneity - and adding housing volume information may help to estimate population counts of the target areas. Eicher and

Brewer (2001) suggest that the areal weighting method may not be employed to generate spatially distributed population surfaces because it does not use ancillary data inputs that aid to separate built-up areas from non-residential uses.

The areal weighting interpolation technique is a long-held and straightforward algorithm for areal interpolation (Tapp, 2010, Goodchild and Lam, 1980) and it may remain an acceptable interpolation technique, especially when additional data is not available for the distribution of population totals (Zhou et al., 2012, Qiu et al., 2010, Fisher and Langford, 1996, Xie, 1995, Flowerdew and Green, 1992). Residential buildings are rarely homogenously distributed across the landscape, but generally they are clustered in residential areas and are surrounded by non-residential land use and cover types (Fisher and Langford, 1996). The areal weighting interpolation method based population estimates may contain important errors due to these geographical differences (Langford, 2013, Hawley and Moellering, 2005, Flowerdew and Green, 1992). Commonly, the area based population estimates may be inaccurate. Owing to the homogenous distribution, this interpolation approach is the least demanding interpolation technique in estimating population (Langford, 2007).

2.5.1.2 The pycnophylactic interpolation method

This interpolation method was proposed by Tobler (1979) in order to obtain smooth population density data from spatially aggregated population datasets. The approach depends on the assumption of a smooth density function to provide heterogeneous distribution within the target zones by considering the effect of an adjacent source zone (the weighted nearest neighbours) (Kim and Yao, 2010, Hay et al., 2005, Tobler, 1979) while preserving the volume of interested variable in each zone (Kim and Choi, 2011, Kim and Yao, 2010, Tobler, 1979). This approach allocates the interested values to each grid cell by dividing the total values of the source area with the number of grid cells within the source areas (Hawley and Moellering, 2005). A value of the interested variable (Z value) is calculated for individual grid cells by the average value of each of the cell's four neighbours. The following equation is described by Hawley and Moellering (2005) for computing the Z value.

$$Z_{i,j} = \frac{1}{4} (z_{i,j+1} + z_{i,j-1} + z_{i+1,j} + z_{i-1,j}) \quad (2.1)$$

Where \mathbf{Z} is the value of the interested variable in each grid and cell and \mathbf{z}_{ij} is the density in cell ij .

In order to provide the pycnophylactic condition, the estimated value of each source unit is compared with the true values of source units (Hawley and Moellering, 2005). The simplified description of implementation of this method is defined by Hawley (2005) as follows:

$$\iint_{R_i} \mathbf{Z}(x, y) dx dy = H_i \quad (2.2)$$

where R_i denotes the i th region, $\mathbf{Z}(x, y)$ represents the density function and H_i represents the value of the interested variable in region i .

The pycnophylactic method, which does not require any ancillary data inputs, has been used by several studies due to its ability to generate heterogeneous distribution of the interested variable (Kim and Yao, 2010, Hawley and Moellering, 2005). Hawley and Moellering (2005) applied the pycnophylactic interpolation technique in order to distribute population totals from Census Tract level to Census Block Group level in the counties of Franklin, Hamilton and Jefferson, USA. When the authors compared the results of population distribution obtained from the interpolation methods, they found that pycnophylactic interpolation obtained the second least accurate population estimates in terms of accuracy. Hay et al. (2005) used this approach in order to create human population distribution maps for epidemiological studies in Kenya. The risk of malaria to the human population was determined using census figures. Earlier studies suggest that the main superiority of the pycnophylactic method over areal weighting methods is the ability to generate heterogeneous distribution of spatial variables within the target zones (Kim and Yao, 2010). In reality, the assumption of homogeneity is extremely unrealistic because of uninhabited zones in the study areas where population density is zero. In order to improve the accuracy of interpolation obtained population estimates the use of external data inputs is needed. A variety of differing, more precise interpolation techniques have been developed to disaggregate the aggregated population counts within the source zones (Gregory, 2002).

Some studies have used a combination of methods to improve the accuracy of their population estimate method. For example, Kim and Yao (2010) combined pycnophylactic interpolation with the dasymetric mapping method to develop a hybrid approach. In order to evaluate the performance of the developed hybrid technique it was applied to the Atlanta metropolitan statistical area (MSA), which consists of 28 counties, by distributing population totals from census tracts to census block groups. The hybrid method obtained better population estimates from all used areal weighting, pycnophylactic and binary mapping methods. Earlier than the Kim and Yao (2010) study, Comber et al. (2008) combined dasymetric and volume preserving methods in order to provide a national agricultural land use data of 1 km² resolution. The authors used aggregated June Agricultural Census (JAC) as source zones and 1 km² grids as target zones. The study shows that the coarse levels of geographical variables can be reliably disaggregated into finer target areas.

2.5.2 Dasymetric mapping methods

In this section, existing dasymetric mapping techniques have been reviewed. In order to distribute population, numerous techniques have been used in the field of geographic information systems (GIS) and remote sensing. By applying geographical information science theory, many distribution techniques have been improved to map smaller geographical areas of population distribution according to aggregated values and ancillary datasets, usually known as ‘Dasymetric Mapping’ (Lwin and Murayama, 2010). The choropleth map and the dasymetric map are widespread cartographic forms of population distribution mapping (Lwin and Murayama, 2010). Furthermore, choropleth maps cannot represent statistical differences within the administrative areal units (choropleth map zones) such as population density differences (Lwin and Murayama, 2010, Mennis, 2009, Mennis and Hultgren, 2006a). In order to avoid this limitation, the dasymetric mapping technique can be used to transfer the administrative units into smaller, relevant map zones (Lwin and Murayama, 2010, Bielecka, 2005). The dasymetric mapping approach (one of the cartographic techniques) is the process of disaggregating an aggregated spatial dataset into smaller units (Eicher and Brewer, 2001) by using ancillary data inputs to refine location of the interested geographical variables (Zandbergen, 2011, Mennis, 2009, Maantay et al., 2007, Mennis, 2003). Petrov (2012) describes dasymetric mapping methodology, as an interpolation framework that produces a population density map using external datasets. Reibel and

Bufalino (2005) call the dasymetric mapping method as an ‘ancillary weighting method’ due to the obligatory requirement of external inputs. This technique basically separates a geographical unit into more meaningful homogenous geographical zones (Mennis, 2009, Bielecka, 2005, Eicher and Brewer, 2001). The process of the subdivision of source areas into finer spatial units, which possess a superior internal consistency of variable being mapped, is the fundamental principle of dasymetric mapping (Bajat et al., 2013, Petrov, 2012). For this reason, previous articles (Petrov, 2012, Mennis, 2009, Langford, 2007) show that the quality and reliability of ancillary datasets are significant for obtaining more accurate estimations in dasymetric mapping methodology. Dasymetric maps are usually considered to be more accurate than choropleth maps in visualising population distribution and population densities (Petrov, 2012, Mennis and Hultgren, 2006b, Wu et al., 2005).

The concept of dasymetric mapping, and the term itself, were first explored about a century ago and improved upon by developing innovative forms of mapping models (Petrov, 2012, Petrov, 2008). Even though Mennis (2009) claims that there are many uncertainties concerning the source of dasymetric mapping methodology in the literature, Petrov (2012) indicates that the term ‘dasymetric mapping’ was first seen in Semenov-Tian-Shansky’s publication in 1911. Before this publication, a dasymetric map of the population density of Ireland for the Second Report of the Railway Commissioners was created by Henry Drury Harnessin in 1837 (Robinson, 1955). These two maps employed shading to indicate the extent of population density where the constant boundaries of shaded regions do not match constantly with the boundaries of administrative areas (Mennis, 2009). Therefore, both of these maps would be conceived as dasymetric maps, even though the dasymetric mapping approach had not yet been discovered (Petrov, 2012).

The Russian geographer, Benjamin Semenov-Tian-Shansky, proposed to use the term ‘dasymetric’ to mean ‘density measurement’ (Petrov, 2008) in a 1911 report to the Russian Geographic Society (Mennis, 2009). After Semenov-Tian-Shansky’s 1923 work titled ‘The Population Density Mapping Project of European Russia’, the dasymetric mapping method became extensively recognised in Russia (Mennis, 2009, Petrov, 2008). It is generally considered that the American geographer, John Kirtland Wright (1891-1969), was the first to publish a paper on dasymetric mapping in the

English-language (Petrov, 2012, Maantay et al., 2007). Several scholars who used dasymetric mapping techniques have cited John K. Wright's publication titled 'A Method of Mapping Densities of Population', as the basis in the improvement of dasymetric mapping methods (Mennis, 2009). Subsequently, researchers in spatial analysis (Wu et al., 2005, Goodchild et al., 1993) have made substantial contributions to dasymetric mapping with the help of developments in environmental remote sensing and GIS (Mennis and Hultgren, 2006b). Since the rise of computational geography in the 1960s, coupled with improvements in analytical and geo-processing capacities, the interest in dasymetric mapping and areal interpolation techniques have led to the formulation of complex interpolation processes (Petrov, 2012, Mennis, 2009, Tobler, 1979). The recent popularisation of dasymetric methodology is caused by a considerable growth in the application of dasymetric mapping techniques to generate high resolution\quality population distribution map and estimate products for many purposes (Petrov, 2012). Petrov (2012) stresses that dasymetric mapping was traditionally used for visualising population distribution, and this approach has been employed for population estimation (Mennis, 2009) and spatial population forecast (Deng et al., 2010).

Remote sensing imagery has been widely used as a source of ancillary data in dasymetric mapping applications (Alahmadi et al., 2013, Bajat et al., 2013, Petrov, 2012). New techniques and approaches of dasymetric mapping utilise a variety of different ancillary data and processing approaches (Lu et al., 2011b, Mennis, 2003, Eicher and Brewer, 2001), and 3D dasymetric mapping (Lwin and Murayama, 2010, Qiu et al., 2010). Recently GIS and remote sensing image based datasets such as land use and land cover datasets (Mennis, 2009, Eicher and Brewer, 2001, Gallego and Peedell, 2001), light emission data (Briggs et al., 2007), road network data (Zandbergen and Ignizio, 2010, Holt et al., 2004), address point dataset (Zandbergen, 2011), building footprint dataset and building volume information (Lwin and Murayama, 2009) have been widely used as ancillary data in dasymetric population distribution studies. The selection of articles, which have used the different forms of dasymetric mapping approaches in population distribution, are shown in Table 2.2.

Table 2.2: Selection of previously used traditional dasymetric mapping methods

Method	Ancillary Variable	Distribution of Variables	Discussed and applied by:
1. Binary Method	Land use Land Cover data, OS Vector Map District buildings	Heterogeneous	Bielecka, 2005 de Geer, 1926 Eicher & Brewer, 2001 Hawley & Moellering, 2005 Holt, Lo, & Hodler, 2004 Langford, 2013 Leyk, Battenfield, Nagle & Stum, 2012 Mennis, 2009
2. Three Class Method	Land use zoning, topography, Public use micro data sample of US Census Bureau	Heterogeneous	Bajat , Krubnic & Kilibarda, 2011 Eicher & Brewer, 2001 Langford, 2007 Leyk, Nagle & Battenfield, 2013 Maantay, Maroko & Hermann, 2007 Mennis, 2009 Su, Lin, Hsieh, Tsai & Lin, 2010
3. Limiting Variable Method	Land Use and Land Cover, Satellite Imagery	Heterogeneous	Eicher & Brewer, 2001 Gregory, 2002 Maantay, Maroko & Hermann, 2007 McCleary, 1969
4. Image Texture Method	Satellite imagery data, Geospatial data	Heterogeneous	Azar, Engstrom, Graesser & Comenetz, 2013, Langford, 2007 Langford, Higgs, Radcliffe & White, 2008 Maantay, Maroko & Hermann, 2007

As Lwin and Murayama (2010) and Maantay et al. (2007) state, data transformation from one set of geographic zones to other set of zones, which are not coincident, is often required in spatial analysis. For example, the analyst may have human population data on an administrative unit level but needs to estimate the population of a smaller area within the given administrative units or the estimation area that includes only part of an administrative unit and some part of the other administrative units (Maantay et al., 2007). Mennis (2009) claims that GIS and satellite remote sensing are the two associated technologies advanced enough to change dasymetric mapping from a cartographic technique to a more well-known and well-defined topic of current investigation. The superiorities of the dasymetric mapping method over the choropleth map approach have been well documented in a variety of earlier studies (Lwin and

Murayama, 2010). The smaller geographical unit of population data is important for effective and efficient micro-scale spatial analyses such as disaster management, emergency preparedness, retail market competition and other population related analysis (Lwin and Murayama, 2010). The following approaches are some of the improved techniques and methods in the dasymetric mapping concept: the binary dasymetric mapping (Eicher and Brewer, 2001, Gallego and Peedell, 2001), the three-class (Langford, 2006), the limiting variable (Maantay et al., 2007) and the image texture based dasymetric model (Azar et al., 2013). These methods are discussed as follows.

2.5.2.1 The binary dasymetric method

The binary method was originally developed to be used as a mapping technique (Langford and Unwin, 1994) but was eventually applied in areal interpolation problems (Cockings et al., 1997, Fisher and Langford, 1995). The method uses ancillary data within the process of dasymetric mapping to provide the spatial distribution of population variables in target zones (Langford, 2007, Eicher and Brewer, 2001). Standard areal weighting method assumes that the values of the variable of interest are uniformly distributed within the source area (Flowerdew and Green, 1992). However, dasymetric mapping employs external data inputs of study regions to distinguish residential areas from non-residential areas for achieving more accurate and efficient target unit estimates (Mennis, 2009, Langford, 2007, Eicher and Brewer, 2001). Figure 2.2 presents an illustration of the binary dasymetric mapping approach. The figure shows that population totals are distributed using external datasets to refined target areas. The method is one of the most widely applied dasymetric mapping methods in the study of population estimates (Mennis, 2009, Bielecka, 2005). Commonly, remote sensing provided land use and land cover datasets that have been used as control variables to date (Mennis, 2009, Bielecka, 2005). The land use/cover datasets are mainly grouped into several classes such as water bodies, tree cover, and urban area and so on. Later on, the classified imagery is reclassified into two binary masking zones, or classes, which are inhabited and uninhabited classes.

The binary masking zones approach has been theoretically investigated to redistribute statistical population data exclusively to areas classified as residential areas (Mennis, 2009). In the binary method, an area-class map defining populated and unpopulated

areas is employed to redistribute data from choropleth map areas, where the data are concentrated homogenously in the inhabited areas of each single zone (Mennis, 2009). In the generated dasymetric map, the unpopulated parts of the map are dedicated zero data, and in populated areas the density is raised as compared to the data density in the choropleth zone (Mennis, 2009, Mennis and Hultgren, 2006b). For instance, people would be eliminated from uninhabited areas such as wetlands, bare grounds, forest or water and accumulated into the other land cover.

The main advantage of the binary method is its theoretical simplicity and easy implementation (Mennis, 2009, Bielecka, 2005, Eicher and Brewer, 2001, Langford and Harvey, 2001). This technique also performs well compared to other forms of areal interpolation methods that do not make use of ancillary data (Langford, 2013, Eicher and Brewer, 2001). For example, Kim and Yao (2010), Mennis and Hultgren (2006b), and Hawley and Moellering (2005) used the areal weighting, the pycnophylactic and the binary dasymetric mapping methods to distribute population totals. All the above research used remotely sensed products as control variables in the process of binary mapping. Furthermore, the authors compared the results to test the performance of applied approaches. They found that the binary mapping method provides better estimation results consistently. Additionally, Langford (2013) applied the areal weighting, the street weighting and the binary mapping methods in the City of Cardiff in South Wales. The author used Landsat ETM+ obtained land cover and OS VectorMap District buildings datasets in both raster and vector format as external inputs in the dasymetric model. The article shows that the binary mapping process substantially improved the accuracy of population estimates over other used interpolation methods. However, the method has several shortcomings. The main limitation of this technique is the subjectivity of reclassification when populated and unpopulated regions are determined (Bielecka, 2005). The other disadvantage is that this approach cannot differentiate areas according to population density within populated areas due to the complexities in population density (Maantay et al., 2007, Langford, 2006).

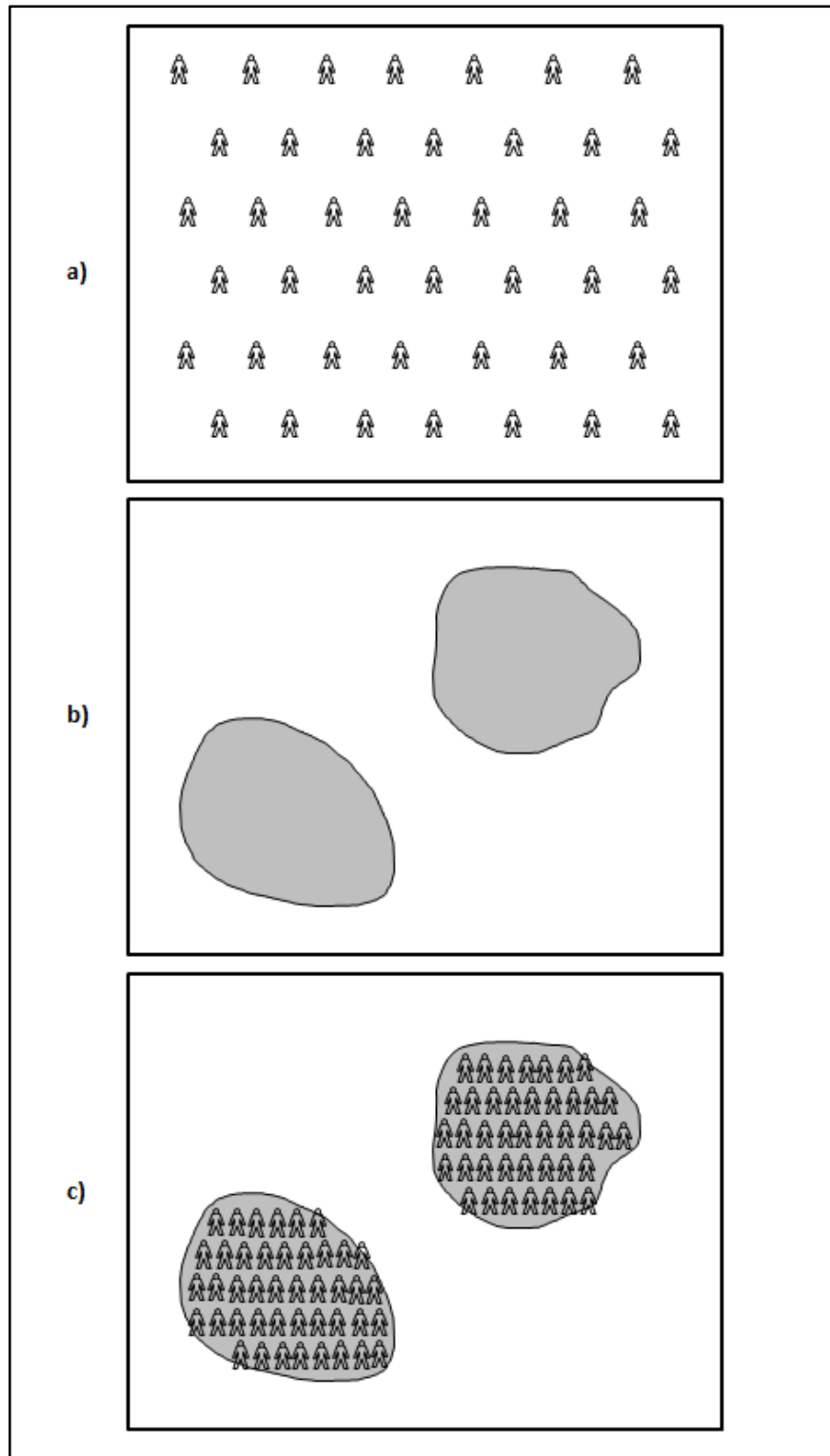


Figure 2.2: An illustration of population distribution models: a) in a source area population distributed uniformly, b) a source zone classified as populated (grey) and unpopulated (white) areas, c) population distributed within populated areas of a source zone (the binary dasymetric population mapping model (modified from Langford, 2007)).

2.5.2.2 Three-class method

While the binary dasymetric method quantifies the functional relationship between the external data and the statistical surface, the three-class technique quantifies the functional correlation between ancillary classes and the statistical surface on a percentage basis (Mennis, 2009, Maantay et al., 2007, Langford, 2006, Eicher and Brewer, 2001). At first, land use/cover data are divided into three or more classes in order to distribute population. The next step is to consider variations in population density within different classes. Percentages are applied to all the three (or more) main land use classes for that area, representing the proportion of population that is probably included within that land use per district (Maantay et al., 2007). Figure 2.3 shows the process of this interpolation model. Eicher and Brewer (2001) adopted the following weighting scheme to assign population to three different land uses classes:

Table 2.3: The sample ancillary classes of three-class interpolation method

Urban area	60%
Agriculture/woodland/exurban area	25%
Forested area	15%

Single categories in the area-class map are apportioned a percentage, thereby the percentage of all categories sum to 100% (Mennis, 2009). Based on the percentage of each area classes, population is redistributed from choropleth zones to different dasymetric zones assuming population is distributed evenly within every zone (Mennis, 2009). For instance, suppose a 50km² source area which consists of 20km² of inhabited and 30km² of uninhabited area includes 1,000 people. Further, the method separates populated area into class such as high density urban or low density urban area, agricultural area and forested area, for re-distributing total population. Using the scheme given above, 60% of people reside in urban areas, 25% of people reside in agricultural area and 15% of people reside in forested area. Su et al. (2010) used this interpolation method to better capture population distribution by using remote sensing imagery, land value and transportation networks as ancillary datasets, separately in the Taipei metropolitan area in Taiwan. The study region was grouped into four categories: non-populated, rural, suburban and urban areas in order to distribute population totals according to the weighting factor of each class. The authors conclude that the

performance of population redistribution improved by more spatial discrimination into the method using further external data.

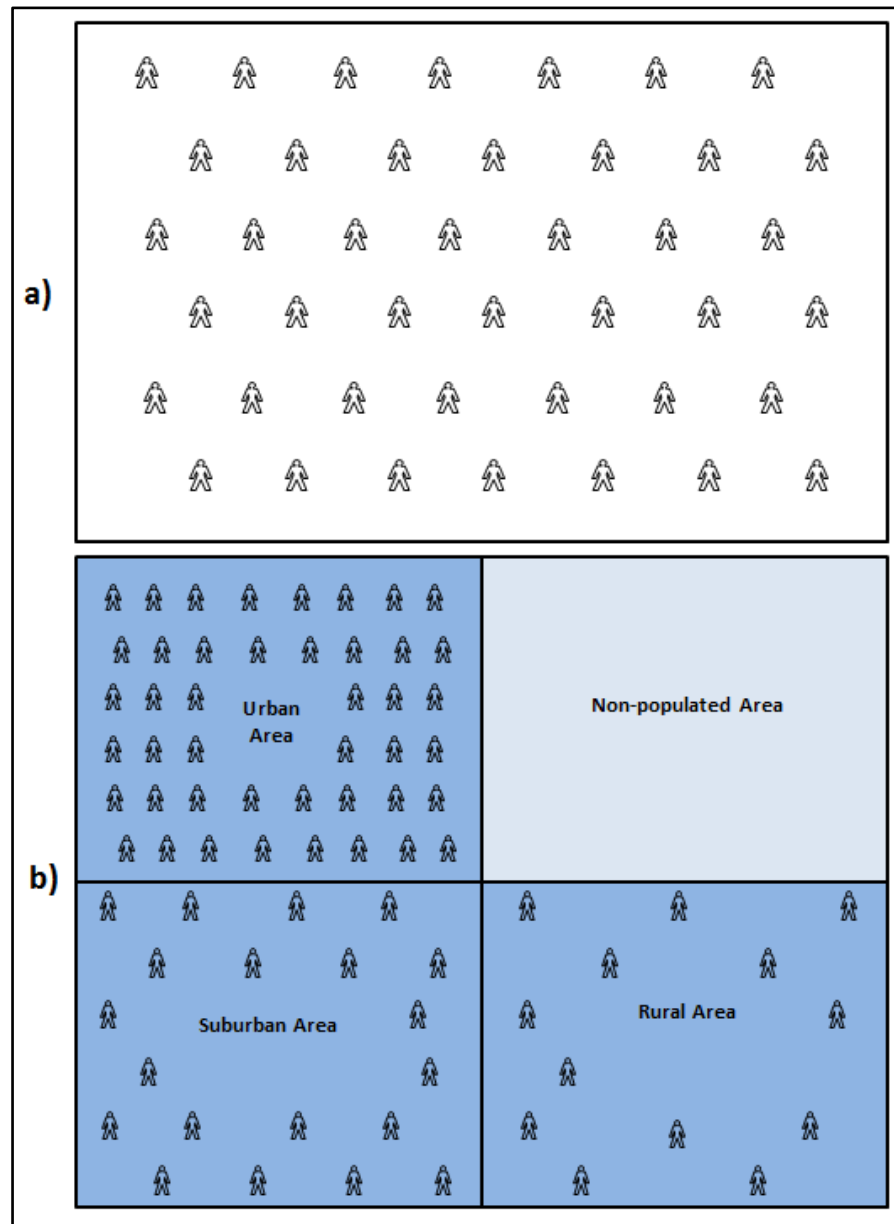


Figure 2.3: An illustration of three class population distribution model: a) an administrative area assuming population totals show uniform pattern and a) the distribution of population according to the weighting factor of each urban, suburban and rural area.

Although this technique is called the three area-class technique, it can be employed to any number of categories (Mennis, 2009). The following formula is algebraically described by Langford (2006).

$$\hat{P}_t = \sum_{s=1}^S \sum_{c=1}^C \frac{A_{tsc} P_s}{A_{sc}} = \sum_{s=1}^S \sum_{c=1}^C A_{tsc} d_{sc} \quad (2.3)$$

where \hat{P}_t is the predicted population of target unit t , A_{tsc} is the area of overlap between target unit t and source unit s , and having land cover determined as inhabited class c , P_s is the total population of source unit s , A_{sc} is the area of source unit s having land cover determined as inhabited class c , S is the total number of source units, C is the number of populated land cover classes, and d_{sc} is the dasymetric density of inhabited class c in source unit s .

The main limitation of the three-class method is that population density may differ within the same classes (Langford, 2006, Eicher and Brewer, 2001). However, the method assumes that the same land use polygons have a characteristic population density. Although this method recognises the differences between land use categories, Maantay et al. (2007) argue that it does not address those intra-land use class differences. Because of the “presence of different building densities and different physical housing types” (Langford, 2006, page 167), all residential areas have different population densities (Maantay et al., 2007, Liu et al., 2006). With the advent of available datasets, the number of classes may be extended to improve population estimates.

2.5.2.3 Limiting variable method

Although the “limiting variable” method was described by McCleary (1969) (Mennis, 2009, Eicher and Brewer, 2001), Wright (1936) and Robinson (1955) applied the original form of this method in order to redistribute population to urbanised regions by limiting population density in rural regions (Mennis, 2009). This method “expands upon the three-class method by setting threshold density limits for population assigned to the various categories of land use polygons” (Maantay et al., 2007, page 82). The process begins with the classification of a source area into populated (urban, agricultural/ woodland and forested areas) and unpopulated (water or empty land cover) zones.

Thereafter, total population is re-distributed throughout each dasymetric zone of the study area by using the simple areal weighting method, and later the limiting thresholds are applied to each land use classes (Maantay et al., 2007, Eicher and Brewer, 2001). For instance, agricultural/woodland areas are limited to 40 people per km² and forested areas limited to 10 people per km². After these thresholds, the remaining population from agricultural/woodland and forested areas are assigned to urban areas. Basically, this approach defines a maximum population density of the determined area-class categories and it allocates maximum population density limits to the land use map classes (Mennis, 2009).

As a final step, these threshold values are used to arrange to the statistical data within each study area (Eicher and Brewer, 2001). If an area-class polygon density exceeded its threshold, the exceeded population is removed and reassigned to the other remaining zones within that geographic unit (Maantay et al., 2007, Eicher and Brewer, 2001). The main weakness of the limiting variable method, like the three-class method, is that the method significantly recognises the differences between land use classes but ignores the difference within a land use class (Maantay et al., 2007).

2.5.2.4 Image texture method

The image texture method is another population estimation method that mainly uses very high spatial resolution satellite imagery such as QuickBird and GeoEye products to distribute population totals (Maantay et al., 2007). In order to quantify population density, this approach examines the relationship between population density and image texture using satellite images instead of land use classes (Maantay et al., 2007). Population is located to pixels which are determined as residential uses within the given jurisdiction. In dasymetric mapping, the significant source of errors are classification mistakes in the ancillary dataset of remote sensing images (Langford, 2007). Therefore, raster pixel maps can be used instead of classified satellite images to drive the dasymetric mapping methodology (Langford, 2007). Based on the spectral signature of pixels and RGB bands, populated pixels can be discriminated from unpopulated pixels. However, pixel maps (raster scan maps) and aerial images use only RGB bands to separate built-up and non-built-up pixels with the help of colour code information (Langford, 2007). Figure 2.4 presents the process of image texture based dasymetric mapping method. Population can be easily redistributed to the target zones, based on

determined populated pixels. The image texture technique reaches better results compared to other dasymetric interpolation techniques, which employ land use/cover datasets, especially in rural areas (Langford, 2007, Maantay et al., 2007).

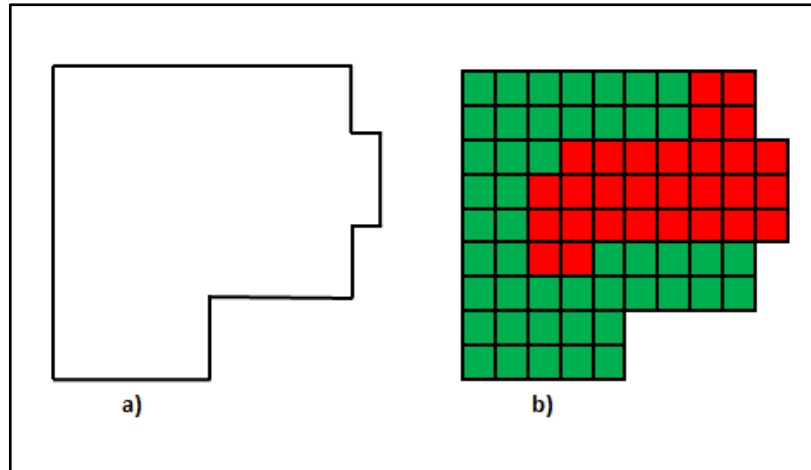


Figure 2.4: The concept of image texture dasymetric method: a) Source zone and b) Raster dasymetric model (Red pixels represent populated area and green pixels represent unpopulated areas).

This method has several weaknesses. One of the main limitations of the image texture method is that it does not distinguish industrial and commercial areas from residential urban zones. Thus, in order to produce even better estimation results, Langford (2007) suggests using the combination of the remote sensing images with other geographical datasets to mark residential uses and non-residential uses (commercial, industrial and institutional areas). Additionally, while the relationship between image texture and census population density can be examined, it may not be enough to produce reliable estimates of population distribution (Maantay et al., 2007). Pixel-based results show that remote sensing images can improve the accuracy of a population distribution map (Liu et al., 2006 as cited in Maantay et al., 2007). As an example, Azar et al. (2013) produced a gridded population data for Pakistan by developing a population distribution mapping model which uses high and medium-resolution satellite imagery as external data inputs to disaggregate population counts. They claim that the population distribution results are limited by the quality of used external data inputs.

2.5.3 Statistical dasymetric models

The statistical methods intend to quantify the functional relationship between control variables and areal units by extracting patterns embedded in external datasets (Mennis, 2009). These models appear to be more robust and reliable compared to the traditional cartographic approach because it does not require any background knowledge by the researcher (Mennis, 2009). It is also capable of revealing the inherent relationships between variables that are otherwise concealed within the datasets. In order to compute the functional relationship between the external data inputs and the population distribution map in a more advanced way compared to traditional cartographic methods, scholars in areal interpolation have developed a statistical perspective (point of view) in dasymetric mapping methodology (Mennis, 2009, Mennis and Hultgren, 2006b). Statistical methods desire to determine the correlation from patterns concealed in the dataset rather than utilising a priori knowledge to measure this correlation (Murakami and Tsutsumi, 2011). As Mennis (2009) states, the main process of this method has been explained in the content of areal interpolation, where the data transformation from the choropleth map areas to the other target zones is facilitated by utilising a third set of ‘control’ zones, which play an equal role with the area-class map in traditional dasymetric mapping methodology. In contrast to traditional dasymetric mapping, the boundaries of target zones do not essentially overlap with the boundaries of the choropleth map and area-class map, however, these target zones are an independent zonal scheme (Mennis, 2009).

A variety of statistical dasymetric mapping models have been developed and includes regression based approach, expectation and maximisation algorithm, street weighting method (Bentley et al., 2013, Reibel and Bufalino, 2005), Cadastral-based expert dasymetric system (CEDS) (Maantay et al., 2007), Intelligent Dasymetric Method (IDM) (Mennis, 2009, Mennis and Hultgren, 2006b), address-weighted (AW) (Zandbergen, 2011) and parcel distribution (PD) methods (Tapp, 2010). A sample of studies that used these statistical models are summarised in Table 2.4.

Table 2.4: Selection of previously developed Statistical Modelling Techniques

Method	Ancillary Variable	Distribution of Variables	Discussed and applied by:
1. Street Weighting Method,	OS VectorMap District roads, Road network data	Heterogeneous	Bentley, Cromley & Atkinson-Palombo, 2013 Brinegar & Popick, 2010. Hawley & Moellering, 2005 Langford, 2013 Maantay, Maroko & Hermann, 2007 Mennis & Hultgren, 2006 Reibel & Bufalino, 2005
2. Intelligent Dasymetric Mapping	Nigh-time Lights, Imperviousness Road network	Heterogeneous	Mennis & Hultgren, 2006 Mennis, 2009 Zandbergen & Ignizio, 2010
3. Expectation Maximization Algorithm,	The GIRAS-based land use, Land cover classes	Heterogeneous	Schroeder & Riper, 2013 Mennis, 2009 Langford, 2007 Mennis & Hultgren, 2006 Gregory, 2002
4. Inverse Distance Weighting,	Genetic Programming and Genetic Algorithms, Land cover data	Heterogeneous	Liao, Wang, Meng & Li, 2008 Mennis, 2003
5. CEDS	Cadastral data	Heterogeneous	Maantay, Maroko & Hermann, 2007 Maantay & Maroko, 2009
6. The address-weighted	Building address point data	Heterogeneous	Lwin & Murayama, 2009 Tapp, 2010 Zandbergen, 2011
7. The volumetric estimation	Building height, Building volume data	Heterogeneous	Lwin & Murayama, 2010 Qiu, Sridharan & Chun, 2010 Sridharan & Qui, 2013

2.5.3.1 Street weighting method

Despite various dasymetric techniques that are widely used in land use and land cover data, the street weighting method is an interpolation approach that uses a street network data as ancillary data to derive population estimation (Maantay et al., 2007, Reibel and Bufalino, 2005). The method assumes that population counts are distributed equally along the streets in each areal unit (Xie, 1995). Basically, the complexity and density of the street network is used as a population indicator. On the one hand, it assumes that

denser and more complex street networks present a higher population density, but on the other hand less complex and scattered street networks show a low population density (Reibel and Bufalino, 2005). In this method, all vector layers that are source areas, target areas and street network layers are required to allocate population totals to street segments (Bentley et al., 2013, Reibel and Bufalino, 2005, Xie, 1995). The method firstly overlays the layers of source units and the street networks to define the length of road segments within the individual source unit and then assigns the population totals in source units based on the length of street segments (Xie, 1995). Later, the weights of street segments within each source zone are computed. The results obtained are finally summed to get population values of each target area (Reibel and Bufalino, 2005, Xie, 1995). The description of implementation of this approach is described by Langford (2013) as follows:

$$\hat{P}_t = \sum \frac{L_{ts}}{L_s} \cdot P_s \quad (2.4)$$

Where P_s is the actual population of source unit s , \hat{P}_t is the predicted population of target unit t , L_{ts} is the length of each street vector in that intersection area between source unit s and target units t and L_s is the total length of street vectors in the source units.

Despite the boundary changes of the enumeration districts during the trend interval, the street weighted method can be applied to provide relatively reliable estimates of population trends for complete sets of local areas across a region (Reibel and Bufalino, 2005). Xie (1995) used the street weighting method to distribute population totals in the City of Buffalo and its nearby suburbs. Reibel and Bufalino (2005) statistically tested the street weighted interpolation method in order to obtain varying density weights for small areas within the original areal units, employing a digital map layer that represents streets and roads. They claim that this method is much easier to apply compared to other interpolation methods and can also reduce errors in estimation when compared to the area-based weighting methods (Maantay et al., 2007, Reibel and Bufalino, 2005). However, similar to the image texture technique, the street weighting technique has difficulties in distinguishing urbanised areas from industrial areas (Maantay et al., 2007).

2.5.3.2 Intelligent dasymetric method (IDM)

Intelligent Dasymetric Mapping (IDM) was improved by Mennis and Hultgren (2006a) as an automated technique to create a dasymetric map of the interested variables (Slocum, 2009). IDM combines the area-weighted interpolation with the population density of ancillary classes in order to disaggregate the aggregated population totals from choropleth map zones to dasymetrically divided zones (Mennis and Hultgren, 2005, Mennis, 2003). This model was called “Intelligent” dasymetric mapping by Mennis and Hultgren (2006) because of its ability to explore the correlation between ancillary dataset and statistical surface in three ways: 1) using the researchers’ domain knowledge; 2) using an innovative empirical sampling model and 3) integrating the empirical and subjective models. The model was used by Mennis and Hultgren (2006a) to dasymetrically interpolate population totals from the census tract level to sub-tract areas in the Front Range of Colorado, USA. They state that with the help of appropriate parameter settings in the process of population disaggregation, the IDM model reaches better results than areal weighting and binary dasymetric mapping techniques. IDM has several superiorities over traditional interpolation models. At first, previous methods establish a straight relationship between external data input and the interested statistical variable. IDM further enables the analyst to integrate domain knowledge with statistical surface to specify functional correlation (Mennis and Hultgren, 2006a). Lastly, the IDM provides relatively better disaggregation of the statistical variable because of its sampling and parameterisation flexibility compared to previous dasymetric mapping models (Mennis and Hultgren, 2006a).

2.5.3.3 Expectation Maximization (EM) algorithm

The Expectation-Maximization (EM) algorithm was originally developed to deal with incomplete and missing data problems (Dempster et al., 1977). When population data are assigned to new target units it may be assumed a missing data problem in dasymetric mapping process (Flowerdew and Green, 1991). Therefore, the Expectation-Maximization (EM) algorithm has been used to deal with missing data in the dasymetric mapping process (Mennis, 2009, Gregory, 2002, Flowerdew and Green, 1992). This method generally begins by employing basic areal weighting to distribute geographical datasets from the zones of choropleth maps to dasymetric intersection areas (Sridharan and Qiu, 2013, Mennis, 2009). In order to predict the density of a particular area-class

map category, the maximum likelihood method is the next step (Mennis, 2009). Furthermore, a particular dasymetric zone's population is re-predicted utilising the new information about the estimated density of a particular area-class category. Maximum likelihood is used next to predict the density of data for a particular area-class map category and so on until the algorithm converges (Mennis, 2009).

2.5.3.4 Inverse distance weighting

One of the simple interpolation approaches is the Inverse Distance Weighting (IDW) method. It was first described and used by Bracken and Martin (1989) as a sophisticated technique to develop surface representation of population information for census enumeration districts (EDs) in the United Kingdom (Liao et al., 2010, Mennis, 2003). It is important to note that knowing how to assign population data to grid cells is a key point to building demographic data surface models in this method (Liao et al., 2010). In the interpolation process, the centroid points of the original areal units are defined and population counts are assigned to a set of summary points (Liao et al., 2010, Mennis, 2003). Basically, the process of creating population surface models in IDW methodology contains three stages. First of all, a grid is generated using the centroid points of original areal units (Liao et al., 2010, Mennis, 2003). In the second stage, population counts are estimated at the grid points based on the distance of centroids by assuming that population density decreases away from the centroid of a given administrative area according to some distance decay function (Mennis, 2003). Finally, the predicted population of a grid is disaggregated to new areas moving from centroids (Liao et al., 2010, Mennis, 2003). Therefore, some raster surfaces which are closer to centroids have greater density and vice versa (Liao et al., 2010). However, Liao et al. (2010, page 49) argue that “this approach is over-simplified, and its accuracy needs to be improved” in order to generate dasymetric maps. The following algorithm is used by Martin (1989, page 96) to distribute the centroid datasets.

$$\hat{P} = \sum_{j=1}^C P_j \cdot W_{ij} \quad (2.5)$$

Where \hat{P} is the predicted population to fall within grid cell i of the matrix, P_j is the empirical population associated with centroid j , C is the total number of centroids within the area to be mapped and W_{ij} is the unique weighting based on the distance from i to j and the clustering of other local centroids.

Earlier studies have used the IDW method to distribute population totals to target zones. For instance, Mennis (2003) generated a 100 metres resolution regular raster grid system from the urbanisation data in the southeast of Pennsylvania to derive population surfaces. The actual population counts of block groups are disaggregated to pre-defined grid cells based on the difference of population densities within the urbanisation classes and the total intersection area of urbanisation classes and block groups. Liao et al (2010) produced one kilometre square grid cells to estimate population distribution for Heshun, north of China. The study indicates that the usage of ancillary variables (slope, land use/cover data, the influence of neighbouring villages) helps to generate a more precise distribution of human population. Geographical reasons do impact on the nature of population distribution in specific areal units.

2.5.3.5 Cadastral-based expert dasymetric system

Maantay et al. (2007) introduced a new dasymetric method that uses census population data together with cadastral-based data in order to more accurately calculate population distribution/location. This method uses specific cadastral data as the ancillary dataset in order to distribute the census population information (Maantay and Maroko, 2009). “Cadastral data is property tax lot data, and is used in recording property boundaries, property ownership, property valuation and for property tax collection” (Maantay and Maroko, 2009, page 4). Cadastral-based datasets mostly consist of population-related information such as zoning designation, land use, residential area, residential units and lot size (Maantay and Maroko, 2009). These types of data are mostly available for urbanised areas in the US, Western Europe and other developed countries (Maantay et al., 2007). A recently-developed Cadastral-based Expert Dasymetric System (CEDS) leads to better estimation of population counts in hyper-heterogeneous urban areas compared to traditional dasymetric mapping methods (Maantay and Maroko, 2009, Maantay et al., 2007). This technique is mainly suitable for urban areas where detailed parcel data are available (Maantay et al., 2007). This CEDS method uses the number of residential units and residential areas as control variables in order to disaggregate the aggregated population totals. Basically, the method assumes that higher populations are concentrated in areas where there are more potential living accommodations (Maantay et al., 2007). First, the CEDS redistributes the population among the cadastres (tax lots) using the residential area and the number of residential units (Maantay et al., 2007). The following equation is described by Maantay et al. (2007).

(2.6)

$$POP_1 = POP_c * U_1 / U_c$$

Where POP_1 is dasymetrically derived tax lot-level population, POP_c is census population, U_1 is the number of proxy units at the tax lot level, and U_c is the number of proxy units at the census level.

The second stage is applying the expert system that is designed to define a variable – the number of residential units, residential area - that more accurately predicts the population distribution (Maantay et al., 2007). Two administrative areas are employed and compared to assess the variables. Tract population data are redistributed to the tax lots and subsequently re-aggregated to the block groups (Maantay et al., 2007). The following formula was used by Maantay et al. (2007) to define the absolute value of the difference between known population and estimated population.

(2.7)

$$POP_{diff} = |POP_{BG} - POP_{est}|$$

where POP_{diff} is the difference between census and predicted population per block group, POP_{BG} is census block group population, and POP_{est} is estimated population derived from the census tract.

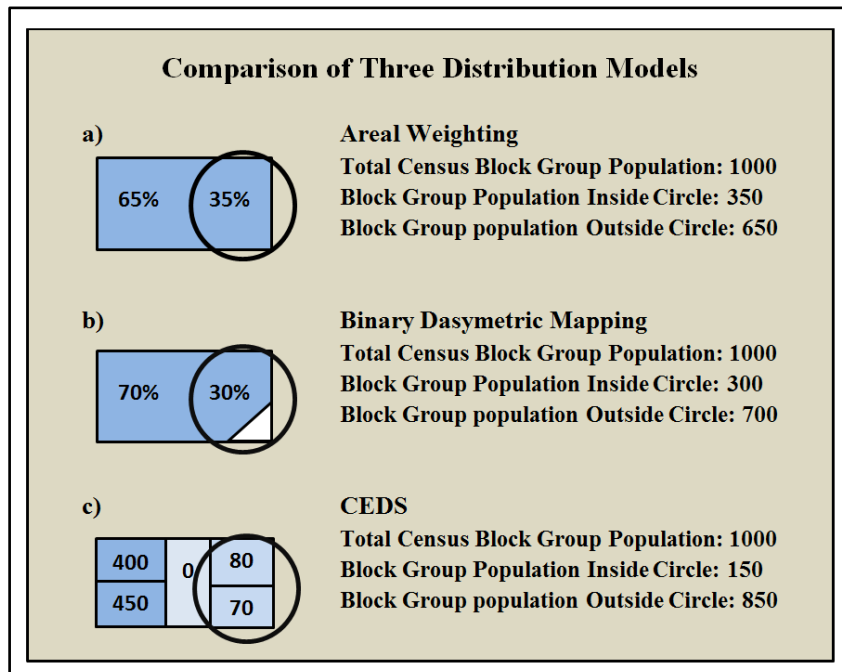


Figure 2.5: Methodological differences and the potential improvement of population estimation of the CEDS method (c), over both filtered areal weighting (b), simple areal weighting (a) (modified from Maantay et al., 2007).

2.5.3.6 The address-weighted method

The address-weighted dasymetric model is one of the population distribution methods which use point data inputs. The approach uses individual building address points instead of land use/cover data, pixel maps, road networks or cadastral-based data. This model explores the functional relationship between address point data and a statistical surface (Zandbergen, 2011, Tapp, 2010). The individual residential housing units are the finest scale that represent where people are actually residing (Sridharan and Qiu, 2013, Zandbergen, 2011, Qiu et al., 2010). Address point data inputs have been recently used as ancillary data in dasymetric population distribution (Zandbergen, 2011, Tapp, 2010). Theoretically, the address points of housing units supply accurate datasets on the location of individual housing units and every non-residential unit (Zandbergen, 2011, Tapp, 2010, Maantay et al., 2007). Therefore, it can be considered that the density of housing unit address points is closely related to population density (Zandbergen, 2011). As Zandbergen (2011) states, this method theoretically presumes a linear correlation between residential address point density (the numbers of address points) and population density. For instance, for a ward with 4,000 people and 2,000 address points, each address point is assigned 2 people. The performance of dasymetric mapping method uses address points data as a source of ancillary data (Tapp, 2010). Figure 2.6 presents an illustration of the address-weighted dasymetric model. As shown in the figure, the number of housing units are falling into target zones that aid to estimate the population of each target unit.

The housing unit based estimation method is established by counting dwelling units at the proposed geographical level (town, city or county). Two fundamental components of the housing unit method are housing unit counts and persons per household (PPH) (Deng et al., 2010). On the one hand, the number of residential buildings can be estimated from “building permit and electrical-costumer information” (Deng et al., 2010, page 5676). On the other hand, residential buildings can be counted from remotely sensed images to generate population estimation. Manually, housing units are counted from finer resolution remotely sensed data, and subsequently the number of houses are multiplied by household size to achieve population estimates (Smith and Mandell, 1984). As stated by Deng et al. (2010), this manual method is not commonly preferred by state and local agencies because a large amount of time and labour is involved. For addressing this problem, national building address point datasets have

been integrated into the dasymetric mapping model (Tapp, 2010). Zandbergen (2011) used the address-weighted method with several other approaches including areal weighting, road network based method and land cover based models in order to determine the performance of address points in the process of dasymetric population distribution. The census tracts were used as source zones and the population totals were estimates at census block groups in 16 counties in Ohio, USA. The author found that building address points performed significantly better when compared with other data inputs.



Figure 2.6: Schematic example of the address-weighted interpolation method. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

One of the main advantages of address point data is the simplicity of the algorithm employed in dasymetric mapping methodology (Zandbergen, 2011). As referred to earlier, each point is weighted equally and population is spatially allocated using the point density in a given location. As Zandbergen (2011) remarks, a calibration and statistical model fitting is not needed, but land use and land cover data are required when employing them as ancillary sources in dasymetric mapping. Therefore, this approach can easily be employed to other study areas which have a national address point data, and can also be compared and combined with different methods in order to generate dasymetric population mapping. Additionally, the individual residential address points substantially show where people actually reside. The address-weighted dasymetric method assumes a uniform population distribution through each address point in source zones.

Previous studies show that address point datasets perform substantially better as ancillary data in dasymetric population mapping when compared with other types of ancillary dataset (e.g. land use/cover dataset, street network dataset, night-time lights, road density maps, and satellite images) (Zandbergen, 2011, Tapp, 2010, Lwin and Murayama, 2009). However, some underlying issues should be addressed in using the address-weighted method in population mapping. At first, the method can perform better in rural areas because single-family housing units are mostly dominant, but may not perform in urban areas due to the complexities of residential buildings (Zandbergen, 2011). For example, low-rise buildings such as single-family housing units and high-rise buildings such as multi-family housing units and apartments co-exist together. Each of the address points may represent more than one single housing unit in complex urban areas and, as such, the average population size may vary considerably between rural and urban areas (Zandbergen, 2011, Lwin and Murayama, 2010). Therefore, as Zandbergen (2011) concludes, the address point data performs well in single-family housing unit dominant areas and performs relatively poorly in complex urban areas. Secondly, an equal population distribution is not possible due to the different type of residential housing units, the diversity of occupied households and the vacant homes (Zandbergen, 2011). Therefore, further information on the type and size of residential units can be used to develop population estimates using address point data (Lwin and Murayama, 2009). The limitations may be addressed by incorporating building use type (i.e. residential buildings mixed with commercial buildings) and building status (i.e. newly

constructed or abandoned buildings) (Lwin and Murayama, 2009). The importance of occupancy rate has been discussed but not used in the small-area dasymetric population distribution. The third issue is the detection of vacant homes in address point datasets for small-area population estimation. As Tapp (2010) argued, local ancillary data (address point and parcel data) are unable to define the information of vacancy rates, making empty residences one of the main error sources in the address-weighted and parcel distribution techniques. These methods can be improved with the inclusion of occupancy rate into the method and incorporating metrics for the type of unit (e.g. single family or multi-family).

2.5.3.7 The volumetric estimation method

Dasymetric interpolation techniques have been improved to create smaller geographical areas of population distribution based on aggregated demographic variables with the aid of external data inputs (Langford, 2013, Sridharan and Qiu, 2013, Lwin and Murayama, 2010). Recently, the use of GIS and remote sensing data has seen the growth in small-area population studies due to the availability of higher spatial resolution images and the availability of fine-scale GIS data with detailed attribute information (Lu et al., 2010, Lwin and Murayama, 2010, Tapp, 2010). In the midst of advancement in the above fields, the volumetric estimation technique has been designed to disaggregate the population of the source zones into smaller geographical units (target zones), based on the building volumes (Sridharan and Qiu, 2013, Qiu et al., 2010). LiDAR (Light detection and ranging) derived residential building volume is presently employed as ancillary data to provide spatially distributed population surfaces (Sridharan and Qiu, 2013, Lwin and Murayama 2010). Using LiDAR remote sensing products as ancillary data in dasymetric mapping has provided possibilities to generate population estimates at the building level using the 3D volume information (Sridharan and Qiu, 2013, Lwin and Murayama, 2010, Qiu et al., 2010, Lwin and Murayama, 2009). LiDAR data provides three-dimensional (3D) information of buildings, which increases the accuracy of population estimates (Lu et al., 2010). The volumetric dasymetric model measures both vertical and horizontal distribution of population, where one-dimensional and two-dimensional data may not quantify vertical distribution (Sridharan and Qiu, 2013). The building footprint datasets and building height information are used to differentiate high-rise and low-rise buildings. Figure 2.7 shows a graphical illustration of the building volumes based on population estimation model.

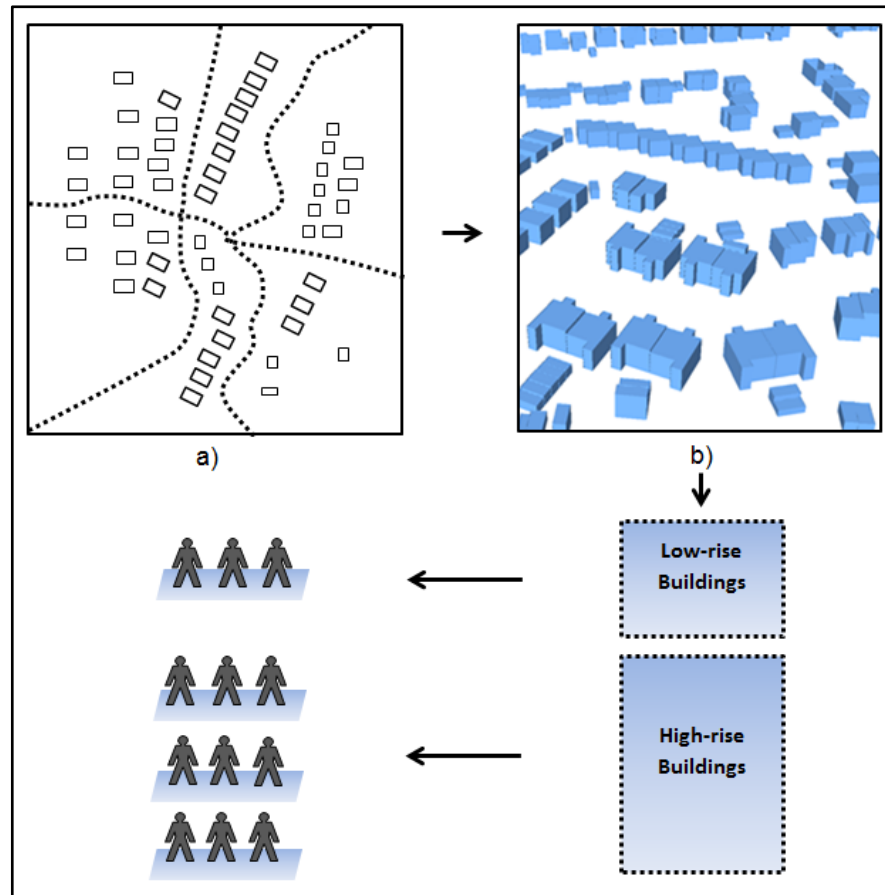


Figure 2.7: Graphical illustration of the building volume based population estimation model; a) represents all building footprints and b) represents all high-rise and low-rise building blocks for creating building volume.

The building volume-based dasymetric model may have several shortcomings. First of all, the building footprint datasets and building height information used to differentiate high-rise and low-rise buildings may not be enough to determine occupied housing units, making non-residential buildings a main source of error in the volumetric estimation technique. Estimations can improve by incorporating building use type (i.e. residential buildings mixed with commercial buildings) and building status information (i.e. newly constructed or abandoned buildings) into the volumetric method (Lwin and Murayama, 2009). Moreover, the volumetric dataset has only been used in the process of population distribution very recently (Sridharan and Qiu, 2013, Lwin and Murayama, 2009) and, as such, the application of LiDAR technology in extracting the characteristics of building units and using them as indicator variables to generate population estimates, has not yet been established for modelling population at a fine spatial scale (Qiu et al., 2010).

2.6 Research Questions

The literature reviewed above shows that building volumes and building address datasets have only very recently been considered in spatial mapping and their robustness in population estimates are yet to be investigated. As stated in Chapter 1, the aim of this research is to make better population estimates by using novel geographical datasets in dasymetric models. The work undertaken involves three stages: (1) the analyses of remote sensing products and Ordnance Survey based external datasets used as control variables in the process of population disaggregation, (2) the implementation of differing areal interpolation methods to provide spatially distributed population surfaces and to estimate population totals at output level, and (3) the comparison of the results of the deployed interpolation techniques to identify the best method and ancillary datasets within the proposed models in both study regions. To demonstrate accuracy improvements made to the population estimation, it is necessary to test the power of the new methods of estimate population against existing approaches in a number of complex and diverse urban and semi-urban regions inhabited by humans. In addressing the aim of the study, a number of research questions that are raised by the authors of underpinning research are forwarded in this study. They can be stated as follows:

1. Do high resolution aerial photography-derived land use/cover datasets and LiDAR-derived building volumes used as external datasets in binary dasymetric mapping and volumetric estimation methods increase the accuracy of population estimates?
2. How well do the datasets of residential housing units perform as control variables in the process of existing dasymetric mapping models in order to obtain small area population estimates?
3. In the address-weighted dasymetric model, how much does the availability of an occupied housing unit's datasets improve the estimate of population?
4. How well do the models of areal interpolation with innovative ancillary datasets perform in high density and low density areas of population, and what are the limiting factors in the approaches developed?

2.7 Summary

The review of areal interpolation methods has shown that basic areal interpolation approaches and many complex population disaggregation models (based on the availability of external data inputs) can be used to estimate population totals in intended geography. In addition, the techniques that use external data inputs as control variables significantly improve population estimation results. This is because some places may be non-residential areas and therefore the population density is zero because they are uninhabited areas. Ancillary datasets help to distribute population totals to inhabitable zones. The chapter has also shown that area-weighted and binary dasymetric mapping techniques are the most commonly employed frameworks in previous research (Petrov, 2012, Mennis, 2009, Eicher and Brewer, 2001, Flowerdew and Green, 1992). Among them, address-weighted and volumetric estimation methods are the most recently developed methods. These interpolation models are different due to their assumptions and the requirement of distinct ancillary datasets.

The literature indicates that dasymetric mapping methods and statistical population disaggregation models generate more accurate small-area population estimates than other areal interpolation techniques that do not utilise ancillary datasets. These models may also provide more accurate population distributions within the spatial units when the relation between population totals and ancillary datasets is utilised. Traditional and statistical dasymetric mapping methods have been introduced, examined, evaluated and applied for the estimation of population distribution mapping to date. The dasymetric approach is preferable to the choropleth mapping approach in depicting the distribution of irregular phenomena in space. In the dasymetric concept, ancillary data plays an important role to generate precise population distribution maps. However, various elements can affect estimation accuracy, but the main driver is the quality of ancillary dataset (Kim and Yao, 2010). Thus, accurate ancillary datasets are required to differentiate the inhabited areas from uninhabited areas. The classified remote-sensing images of land cover and land use data are the most commonly used ancillary data source for dasymetric mapping. Also, land cover, imperviousness, road networks, night-time lights, and residential address points have been used as ancillary datasets in dasymetric maps. Residential address points datasets have been used as ancillary data in dasymetric mapping in very recent studies (Zandbergen, 2011, Owens et al., 2010, Tapp, 2010). For example, Zandbergen (2011) stressed that the simple algorithm

utilised in dasymetric mapping is one of the advantages of address points. It is considered that all residential points are weighted equally and population is spatially distributed, employing the spatial density of points. When a land cover dataset is used as ancillary data, statistical model fitting or calibration is necessary, but when point dataset is employed as additional data, it does not require any statistical output. That is why this method can be replicated in other areas and can be compared with different study areas. Based on the extensive literature reviewed in this chapter, the interpolation models and ancillary datasets adopted and used in this research are presented in the next chapter.

CHAPTER 3

Research Methodology and Datasets

3.1 Introduction

This research aims to generate spatially distributed population surfaces by integrating the interpolation processes of building address point data obtained from Ordnance Survey, occupancy information, LiDAR-derived building volume information, and aerial photography obtained binary masks of land cover classes as ancillary datasets. Although, a range of remote sensing imagery has been used as ancillary data in dasymetric mapping models to date (Langford, 2006, Mennis and Hultgren, 2006b, Bielecka, 2005), building address and height information are only very recently used in the US (Sridharan and Qiu, 2013, Zandbergen, 2011, Qiu et al., 2010). The existing interpolation methods were refined using these external data inputs. The proposed disaggregation models are mathematically formulated in the following sections. This methodology chapter describes the use of five different forms of areal interpolation methods. Figure 3.1 describes the implementation of areal interpolation techniques in the thesis. Additionally, this chapter presents the processing of aerial photography and LiDAR-derived building height data in raster format, building address and building footprints in vector file format in order to generate small-area population estimates. The pre-defined external data inputs have been utilised as control variables in order to generate population estimates for both the City of Leicester and the Borough of Kensington and Chelsea. The City of Leicester is a more sparsely densely populated urban area compared to the Borough of Kensington and Chelsea. The other aim of this research is to evaluate the performance of chosen interpolation techniques in different densely populated regions.

The study intends to estimate population totals in small-areas and to achieve this aim, population counts have been transferred from larger geographies to smaller areal units. Census wards have been used as source zones and output areas have been used as target zones. The designed hierarchy of UK census dissemination units have been used to report 2001 and 2011 census results (The Neighbourhood Statistics - NeSS). The main

level of census output geographies is explained in Section 2.4. These levels of geographies consist of similar sized population and so can be used to compare population changes over time (ONS). Census Area Statistics (CAS) wards were created for reporting census results at a smaller scale than local authorities. The local administrative authority areas have been divided into census wards, all of which nest within local authority boundaries (ONS). MSOAs and LSOAs could be chosen as source zones to distribute population totals to smaller areal units. Census wards were chosen as source units because local authorities use census wards to report census data and inter-censal estimation results. The output areas are the finest geographical units of UK census hierarchy and the variation of population in output areas is far less than variation in ward population (NeSS). The latest census figures are available for all census dissemination units to validate the predicted population values. Due to these reasons, census wards are used as source zones and output areas are used as target zones.

The chapter is organised into 10 sections. Section 3.2 covers population disaggregation algorithms. Section 3.3 presents the study areas of Leicester City and the Borough of Kensington and Chelsea, respectively. Section 3.4 provides ancillary datasets that were used as control variables in dasymetric mapping. The processes of image classification are illustrated in Section 3.5. A subset of aerial image was first created for both study areas from one square km aerial photography tiles, followed by aerial images classified to obtain land cover classes. Finally, the created land cover classes were reclassified into binary classes as populated and non-populated areas. Section 3.6 describes the generation of building volumes that are used as external control variables in the volumetric estimation model. Digital elevation models and building footprints were used to obtain the volume of individual building units within the target zones. Section 3.7 explains the preparation of building address points for the address-weighted dasymetric method including the extractions of residential address points from non-residential structures. The extracted residential housing units were used as a main indicator to distribute population totals through only residential buildings. Section 3.9 is a brief description of the software packages used to perform the analysis mentioned in the previous sections. Finally, a summary of research methodology and ancillary datasets is presented in Section 3.10.

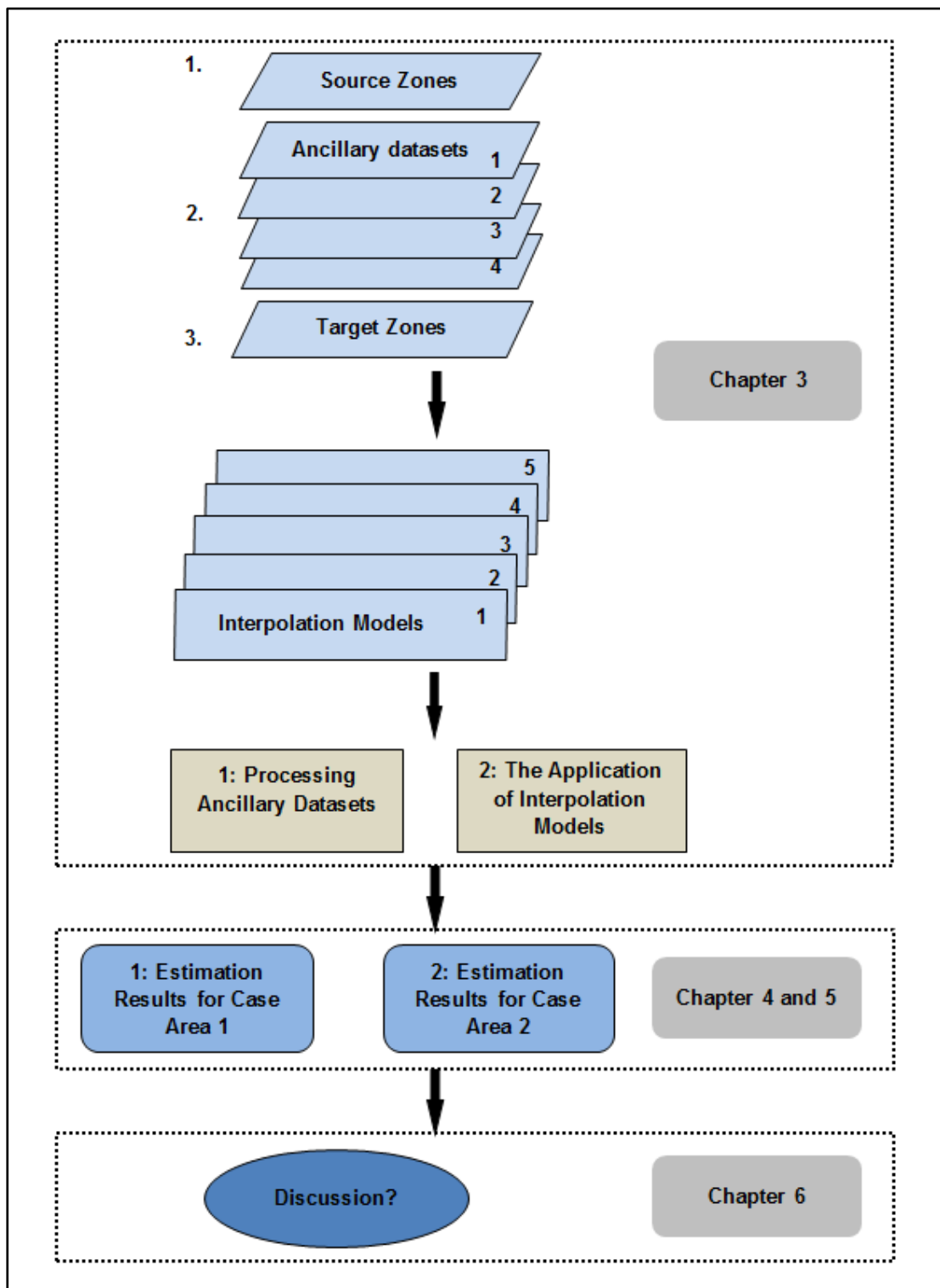


Figure 3.1: The graphical representation of methodological sequences of the interpolation techniques.

3.2 Population Estimation Algorithms

The literature review in Chapter 2 shows that many forms of areal interpolation approaches have been used to disaggregate the aggregated population totals to smaller areas (Bentley et al., 2013, Langford, 2013, Leyk et al., 2013a, Leyk et al., 2012, Zhang and Qiu, 2011, Mennis, 2009, Maantay et al., 2007, Hawley and Moellering, 2005, Eicher and Brewer, 2001). These disaggregation frameworks have been developed according to the availability of ancillary data inputs, spatial characteristics of study areas and the intended resolution of population distribution. As highlighted in the review, the areal weighting method is the simplest form of interpolation processes by allocating the variable of interest (population totals) without using ancillary data except the boundaries of source and target zones. However, in order to estimate the distribution of population, the other forms of dasymetric and statistical interpolation models require specific external inputs based on their mathematical functioning.

This study generated spatially distributed population surfaces by employing selected disaggregation models. Three existing (the areal weighting, the binary and the address-weighted techniques) and two novel interpolation algorithms (such as, in volumetric estimation and address-weighted approaches) were developed and compared based on available ancillary data on case study regions. These particular interpolation models have been chosen for several reasons. First of all, earlier research indicates that the selected dasymetric methods have the potential to redistribute population totals to smaller areal units precisely with the advent of external geographical datasets. Secondly, the proposed approaches are data driven techniques and the datasets used have the potential to differentiate residential areas to generate dasymetric population surfaces. The selected external datasets help to differentiate residential uses from non-residential areas. Very high-resolution aerial images aid to classify study areas as built-up and non-build up areas. LiDAR-derived building height information differentiates low-rise and high-rise housing units. The data of address points helps to define each housing unit within the areal units. Additionally, these interpolation approaches basically redistribute population counts based on external control variables and they preserve the original volume of each census ward in the process of population distribution. The UK national mapping agency (Ordnance Survey) and academic centre (Edina) holds several products that provide a very good opportunity to use them as

control variables in population disaggregation process (Langford, 2013, Gregory, 2002) in the chosen study regions (explained in section 3.3). Secondly, the usability of building height data was tested by using building volumes as control variables in innovative disaggregation processes.

3.2.1 Area-weighted technique

Areal weighting method is a simple interpolation technique which assumes that population totals are distributed homogeneously within a given areal unit (Goodchild and Lam, 1980, Flowerdew and Green, 1992). The method is classified as a basic interpolation technique because it does not require any ancillary datasets besides the geography of source and target zones, and the variable to be interpolated (Langford, 2006, Hawley and Moellering, 2005, Fisher and Langford, 1996, Flowerdew and Green, 1992, Lam, 1983). The population totals are assigned to the target zones according to the proportion of targets zones that fall into source zones (Langford, 2006, Maantay et al., 2007). This process is based on the assumption that variables of interest (population totals) are uniformly distributed within source zones (original data collected area) (Langford, 2013, Maantay et al., 2007, Hawley and Moellering, 2005, Flowerdew and Green, 1992).

The implementation stages of the areal weighting technique to redistribute population totals to target zones were described here. First of all, a 90-m sized grid map of local authorities was created to redistribute population totals to each grid cell within the study areas. In the second stage the layers of source units, target units and grid cells were overlaid to intersect source and target areas with grid cells. The source units consist of a polygon layer, which divides both study areas into a number of census wards and target units include a layer that separates each study area into a number of output areas. This process aims to define the target units of each source unit. The third stage was the estimation of population totals for each grid cell by using Equation 3.2. The final stage was the calculation of the sum of the estimated population totals for each output area. The approach measures a constant population density for each source unit and divides the size of total area by total population counts (Weng, 2012, Zhou et al., 2012). The assumption of uniform population distribution, according to Langford (2013) and Gregory (2002), is extremely seldom in reality. For example, population distribution can be affected by many factors such as, slope, elevation, and land use types (Higgs et

al., 2012, Flowerdew and Green, 1992). The areal weighting method was chosen because of ease of use and the wider applicability in comparing the results of population distribution with any other dasymetrically generated population surface (Langford, 2013, Zandbergen and Ignizio, 2010, Hawley and Moellering, 2005). The intended approach utilised the equation below from Fisher and Langford (1996) and Langford (2006):

$$\hat{P}_t = \sum_{s=1}^S \frac{A_{ts}P_s}{A_s} \quad (3.1)$$

Where \hat{P}_t is the predicted population of target unit t (the estimated population of output areas), A_{ts} is the total area of overlap between target zone t (output area) and source zones s (census ward) P_s is the total population of source unit s (the total population of census wards), A_s is the total area of source zone s (the area of census wards), S is the total number of source zones.

3.2.2 Binary dasymetric mapping technique

The binary mapping method is a simplistic and widely used technique of dasymetric mapping methodology (Mennis, 2009, Maantay et al., 2007, Bielecka, 2005, Mennis, 2003, Eicher and Brewer, 2001). Functionally, this method evaluates the relationship between statistical surface and land use/cover classes subjectively utilising a priori knowledge (Mennis, 2009, Eicher and Brewer, 2001, Fisher and Langford, 1996). This current approach uses the aerial photography obtained land cover information to distribute population totals to only built-up areas over the target areas.

The implementation stages of the binary mapping model to redistribute population totals to target zones were described here. The first stage was to classify the aerial image to define populated areas. Aerial images of both study areas were classified using a supervised classification technique in ENVI image processing software by selecting four land use/cover categories (vegetation, tree cover, water bodies and urban area). In the second stage, land use/land cover classes were regrouped into two classes (populated and unpopulated areas) in order to distinguish populated zones from unpopulated zones such as, vegetation, water bodies, tree cover. In the third stage, the

layers of source zones, target zones and the binary dasymetric zones were overlaid in ArcMap to create intersections of dasymetric zones. In the fourth stage, total population is allocated to the urban dasymetric zones and then population counts of each output area were calculated by using Equation 3.2. The method assumes that population is evenly distributed within all dasymetric zones. The binary dasymetric technique is applied to the study area using different levels of census dissemination geographies and the binary masking of inhabited and uninhabited zones. Figure 3.2 shows the implementation steps of the binary dasymetric mapping process.

The intended binary method employed the equation described by Fisher and Langford (1995) and Langford (2006) is below:

$$\hat{P}_t = \sum_{s=1}^S \frac{A_{tsp} P_s}{A_{sp}} = \sum_{s=1}^S A_{tsp} d_{sp} \quad (3.2)$$

Where \hat{P}_t is the estimated population of target unit t (the predicted population of output area), A_{tsp} is the total area of overlap between target zone t (output area) and source zone s (census ward) and having land cover identified as populated, P_s is the population of source units, A_{sp} is the area of source units having land cover identified as populated (the total built-up area of census ward), S is the total number of source zones (the number of census wards), and $d_{sp} = \frac{P_s}{A_{sp}}$ is the dasymetric density of the populated class in source zone s .

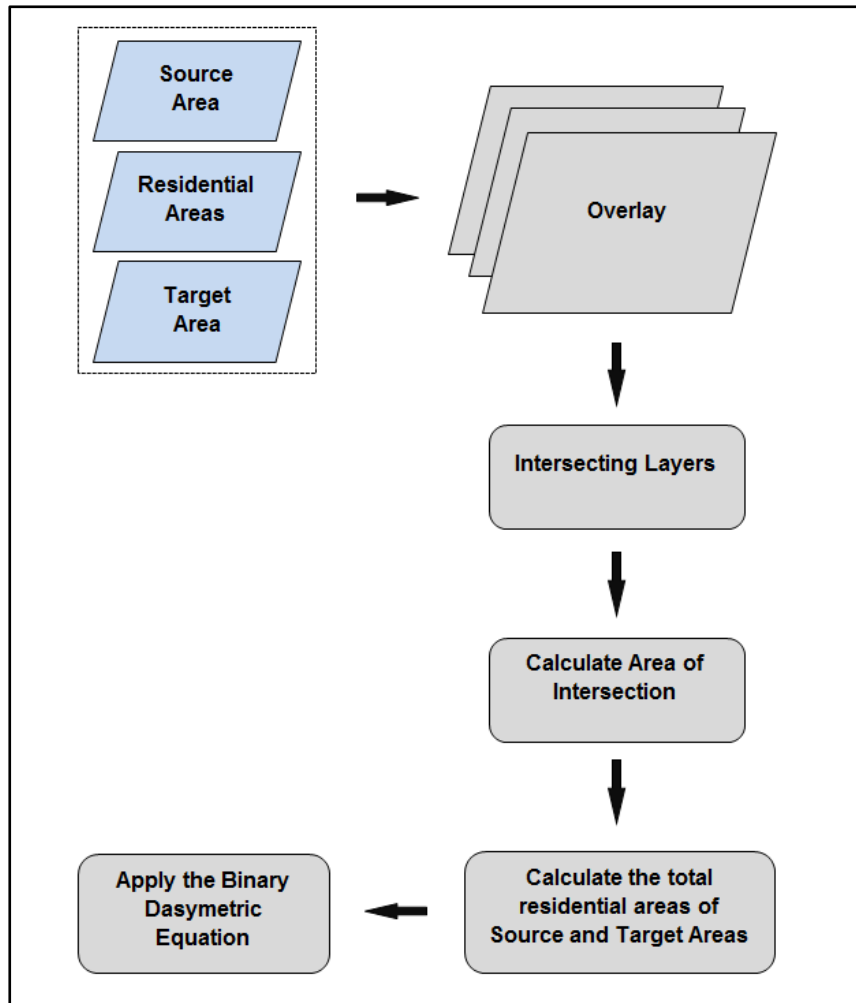


Figure 3.2: The steps of the binary dasymetric interpolation technique.

3.2.3 The volumetric estimation technique

The volumetric estimation is one of the dasymetric areal interpolations that the variable of interest is distributed from one spatial unit to another with the aid of ancillary information (Langford, 2013, Sridharan and Qiu, 2013, Lwin and Murayama, 2010). The approach is one of the very recent dasymetric interpolation methods (Qiu et al., 2010). This method uses the data of building volume as control variables in order to disaggregate population totals in population mapping process. LiDAR data provides three-dimensional information of buildings, and this information may increase the accuracy of population estimates (Lu et al., 2010). The building footprint and LiDAR-derived building volume information can be used as the source of ancillary data in dasymetric mapping. LiDAR-derived residential building volume has been presently employed as ancillary data to spatial disaggregate population information by Sridharan and Qiu (2013), Lwin and Murayama (2010) and Qiu et al. (2010).

The building volume based dasymetric interpolation technique has been proposed to disaggregate the population of the source zones into smaller geographical units (target zones). Figure 3.3 presents the implementation steps of the volumetric estimation method. When the volume of residential building is accessible, the volumetric dasymetric interpolation technique may be used to disaggregate population counts. The LiDAR-obtained digital elevation models have been processed to generate building volumes. The building footprint datasets and building height information are used to differentiate high-rise and low-rise buildings. These datasets may not be enough to determine residential or non-residential housing units, making non-residential buildings a main source of error in the volumetric estimation technique (Sridharan and Qiu, 2013). The volumetric estimation method can be improved to incorporate building use type (i.e. residential buildings mixed with commercial buildings) and building status information (i.e. newly constructed or abandoned buildings) (Lwin and Murayama, 2010) into the volumetric method. The Landmap obtained building block data has been used to define residential building blocks.

The implementation stages to create building volumes for target areas were described here. At first, normalised digital surface models were converted to point features to calculate the height of each building structure. Secondly, the layers of height data of points and building footprints are overlaid to determine the average height of each building structure within each source zone. Then the average height of each housing unit is calculated and the total volume of building units was calculated by multiplying the average height and area of each single housing unit. Furthermore, the layer of target zones is overlaid with building footprints to sum building volume of each target zone. The summed volume of each target unit has been further used as control variables to distribute population totals within target zones. The following equation introduced by Lwin and Murayama (2010) to estimate population totals by using building volume information. This equation was used to assign population counts to each building block and then total population of each output area was calculated.

$$BP_i = \left(\frac{CP}{\sum_{k=1}^n BV} \right) BV_i \quad (3.3)$$

Where BP_i is the population of building i , CP is census ward population, BV_i is total volume of building i , n is number of building blocks that meet user defined criteria and fall inside the CP polygon and i, k is summation indices.

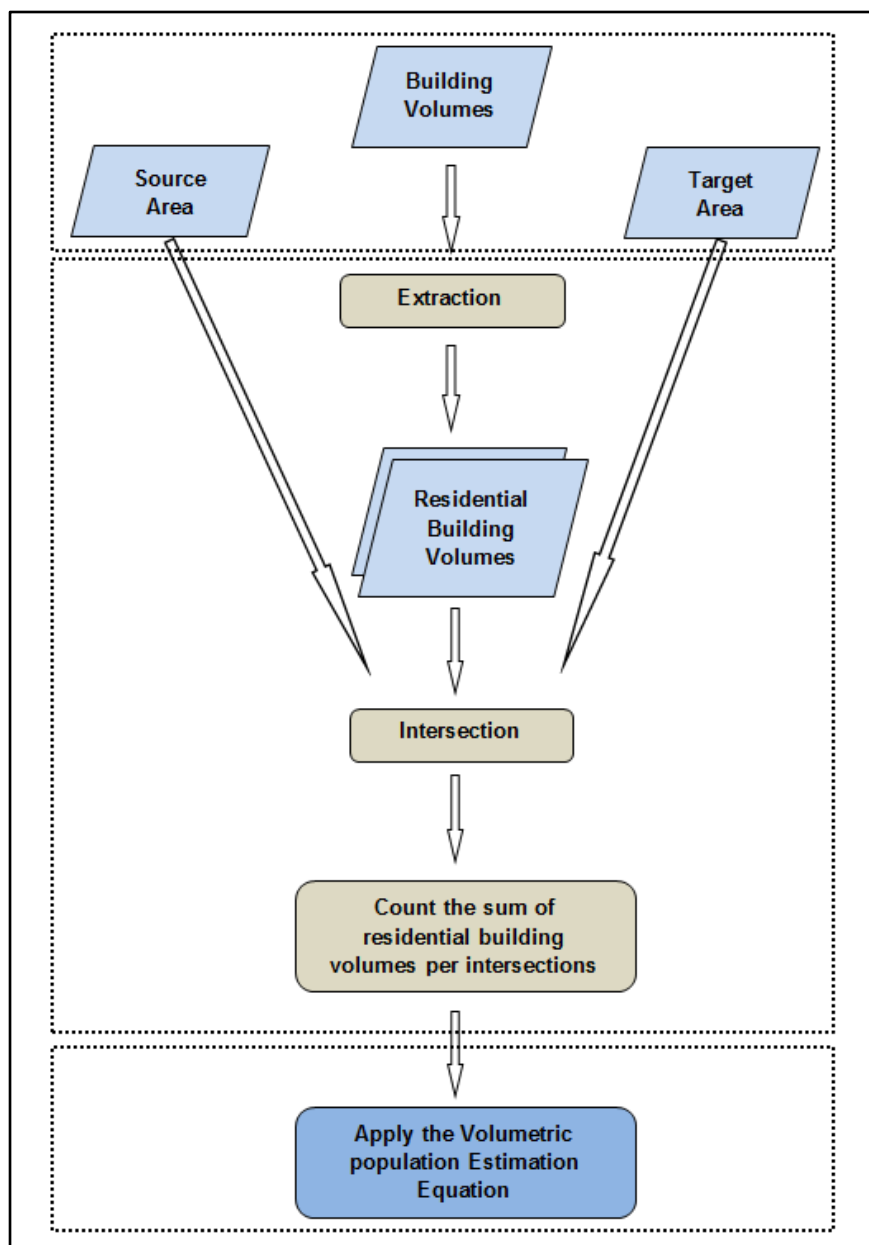


Figure 3.3: Implementation steps for the volumetric areal interpolation technique.

3.2.4 Address-weighted disaggregation technique

The address-weighted method is one of the spatial disaggregation processes. This interpolation technique refines the spatial distribution of population characteristics within source units using the ancillary data of building address points as spatial control

variables (Zandbergen, 2011). The data of address points are a better determiner of population distribution when compared to land use/cover data, pixel maps, street centrelines or cadastral-based data, and other external data points such as schools and bus stops. The method explores the functional relationship between address point data and a statistical surface. Theoretically, housing unit address points datasets supply accurately the location of individual housing units and every non-residential unit within an authority (Tapp, 2010). Therefore, it can be argued that the density of housing unit address points is closely related to population density (Zandbergen, 2011). The popularity of this interpolation technique has recently increased with the availability of address point data (Zandbergen, 2011, Tapp, 2010 and Zandbergen and Ignizio, 2010). The address-weighted method was chosen for this research because it has been used very recently as control variables, but the usability of this method has not been widely evaluated. The approach was applied two times for each study area by using different external inputs. First of all, only building address points were used as control variables and furthermore the building occupancy information and building address points were used as external data inputs in disaggregation process. As Zandbergen (2011) states, this method theoretically presumes a linear correlation between residential address point density (the numbers of address points) and population density. The implementation steps of address-weighted techniques are shown in Figure 3.4. The method was applied using the following equation from Zandbergen (2011).

$$W_c = 1/N_a \quad (3.4)$$

$$P_a = W_c * P_t$$

$$P_b = N_{ab} * P_a$$

Where W_c is the weight of addresses in a census ward, N_{ab} is the total number of addresses in the output areas, P_a is the average population of each address points, P_t is the total population of the census ward, and P_b is the estimated population of an output area.

The weight of individual addresses in a census ward are equal to one the other. The average population of individual addresses (P_a) is the product of the address-weight and the total population of the census ward (P_t). Finally, the total population of a target

zone is calculated by multiplying the total building points with the average population at each address.

One of the main advantages of address point data is that a very simple algorithm is employed in dasymetric mapping methodology (Zandbergen, 2011). Each building point is weighted equally and population is spatially allocated using point density in the given location. As Zandbergen (2011) remarks, a calibration and statistical model fitting is not needed, but land use and land cover data are required when employing them as ancillary sources in dasymetric mapping. Therefore, this approach can be easily employed in other study regions where the data of address points are available. The method can also be compared with other population disaggregation methods, even though it assumes an even population distribution in each address point. Additionally, individual residential address points are substantially related to population count.

The importance of occupancy rate has been discussed but not used in the process of dasymetric population mapping (Qiu et al, 2010). In the second implementation, the address-weighted method has been performed by using both the data of building address points and occupancy information as control variables. The occupancy data was obtained from local city councils for each study area and were cross-checked with population figures from the Neighbourhood Statistics. The datasets represent how many households are resided at the time they were reported at output area level. The equation adopted for this approach has changed slightly by adding the information of occupancy rate as shown below.

$$P_b = N_{ab} * P_a * O_b \quad (3.5)$$

Where W_c is the weight of addresses in a source zones, N_{ab} is the total number of addresses in the target zones (output areas), P_a is the average population of each building address, P_t is the total population of the study region, P_b is the population of a target zones, and O_b is occupancy rate of the target areas.

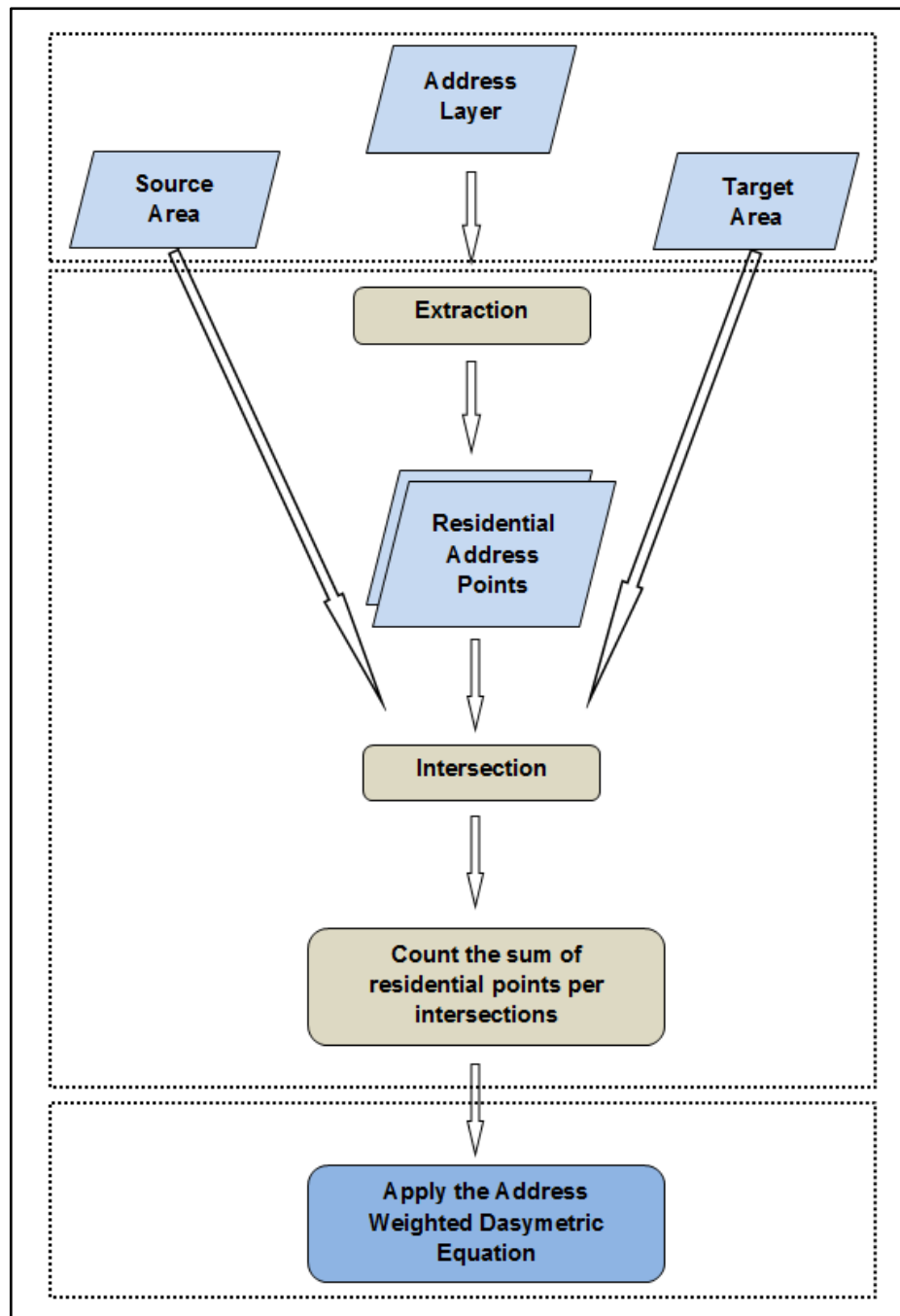


Figure 3.4: Implementation steps for the address-weighted dasymetric interpolation technique.

3.3 Study Area

This research has selected two study areas in order to obtain spatially disaggregated population totals at small-area level. These study regions are the local authorities of Leicester in Leicestershire County and the royal Borough of Kensington and Chelsea in Greater London. Due to the overlapping boundary of these spatial units with each other, the boundary of census wards were used as source zones and the boundary of output

areas were used as target zones to generate population estimates. According to the 2011 UK census reports, the population density of Leicester is 45 persons per hectare; however, the population density of the Borough of Kensington and Chelsea is 130.8 persons per hectare. The intended areal interpolation frameworks were applied where population density is different in these study regions. The Borough of Kensington and Chelsea has one of the most dense building blocks in the UK. Additionally, housing units are more dispersed horizontally in Leicester than in Kensington and Chelsea where high-rise buildings are the dominant buildings. On the one hand, Kensington and Chelsea has a mixture of housing units and it can be a very expensive place to live, but on the other hand, both lower and higher income groups are mixed in Leicester City. While the Borough of Kensington and Chelsea has fewer residents from other ethnicities (such as Mixed, and Asian), Leicester is a multi-cultural city (ONS). These differences indicate that the characteristics of population distribution vary considerably in these areas. For example, ethnic differences may impact on average household size. The methodologies of interpolation models were initially developed for Leicester City and subsequently employed to the Borough of Kensington and Chelsea. The availability of validation data at the scale of output area provides a good opportunity to evaluate the performance of each interpolation models. In this way, the effect of different disaggregation models and a range of external data inputs on the estimation results will be observed in different environments. As Gregory (2002) states, the performance of interpolation techniques depends on the variables used, its relationship with population totals and the choice of target geographies.

3.3.1 Case study area 1: the City of Leicester

The City of Leicester, consisting of the unitary authority area, has been chosen as the first study area. The 2011 UK census shows that Leicester is predominantly an urban region, consisting of a total population of 329,839 and total resident households of 127,383 (ONS, <http://www.neighbourhood.statistics.gov.uk/>). Leicester covers an area of 73.32 square kilometres containing 22 census wards and 969 output areas (<http://census.ukdataservice.ac.uk/get-data/boundary-data.aspx>). Figure 3.5 shows the location of Leicester City in Leicestershire County. The output area is the lowest geographical level at which a census data is released. The UK 2011 Census is published with a variety of geographical resolutions, from the finest spatial units (output areas) to

the larger geographical area (country). Output areas contain approximately 300 persons and 125 households. In order to report demographic figures, the 2011 UK Census used the census output geographies of 2001 with some alterations on areal units in England and Wales (Cockings et al., 2011). Therefore, the 2001 output geographies were maintained by splitting and merging the existing output geographies to make them appropriate for the publication of 2011 UK Census data (Cockings et al., 2011). For instance, the number of wards does not change, but the number of output areas has increased from 890 to 969 in Leicester authority area. There are two main factors that account for the selection of the City of Leicester as the main case area to apply population distribution models. Leicester City was chosen as the initial case study because of logistics and familiarity reasons: there is ease of access to the city and researcher's familiarity with the location. The other main reason has to do with the availability of data. The planned external data types (areal imagery, the models of digital elevations, address point, building footprints) and the updated validation data of population information are available for the smallest census dissemination unit.

3.3.2 Case study area 2: the Borough of Kensington and Chelsea, London

The Royal Borough of Kensington and Chelsea has been chosen as the second study region. The Royal Borough covers an area of 12.13 square kilometres consisting of 18 census wards and 631 output areas (<http://census.ukdataservice.ac.uk/getdata/boundary-data.aspx>). This study region is one of the local authority areas of Greater London, containing a total population of 158,649 and total households of 87,705 (ONS <http://www.neighbourhood.statistics.gov.uk/>). This local authority is one of the most densely populated areas in the United Kingdom (ONS, 2011 UK census reports). The population is distributed densely compared with the City of Leicester as explained in Section 3.3. However, both case areas are located in England and ONS use the same census hierarchy to release census results. Furthermore, both areas have the same required ancillary datasets. One of the objectives of this work is to test the applicability of interpolation models in different locations in order to determine their accuracy for small-area population estimation. The intended interpolation techniques were applied to the Borough of Kensington and Chelsea in order to provide precise population estimates in different regions where population is more or less densely distributed. Figure 3.6 shows the location of the London Borough of Kensington and Chelsea.

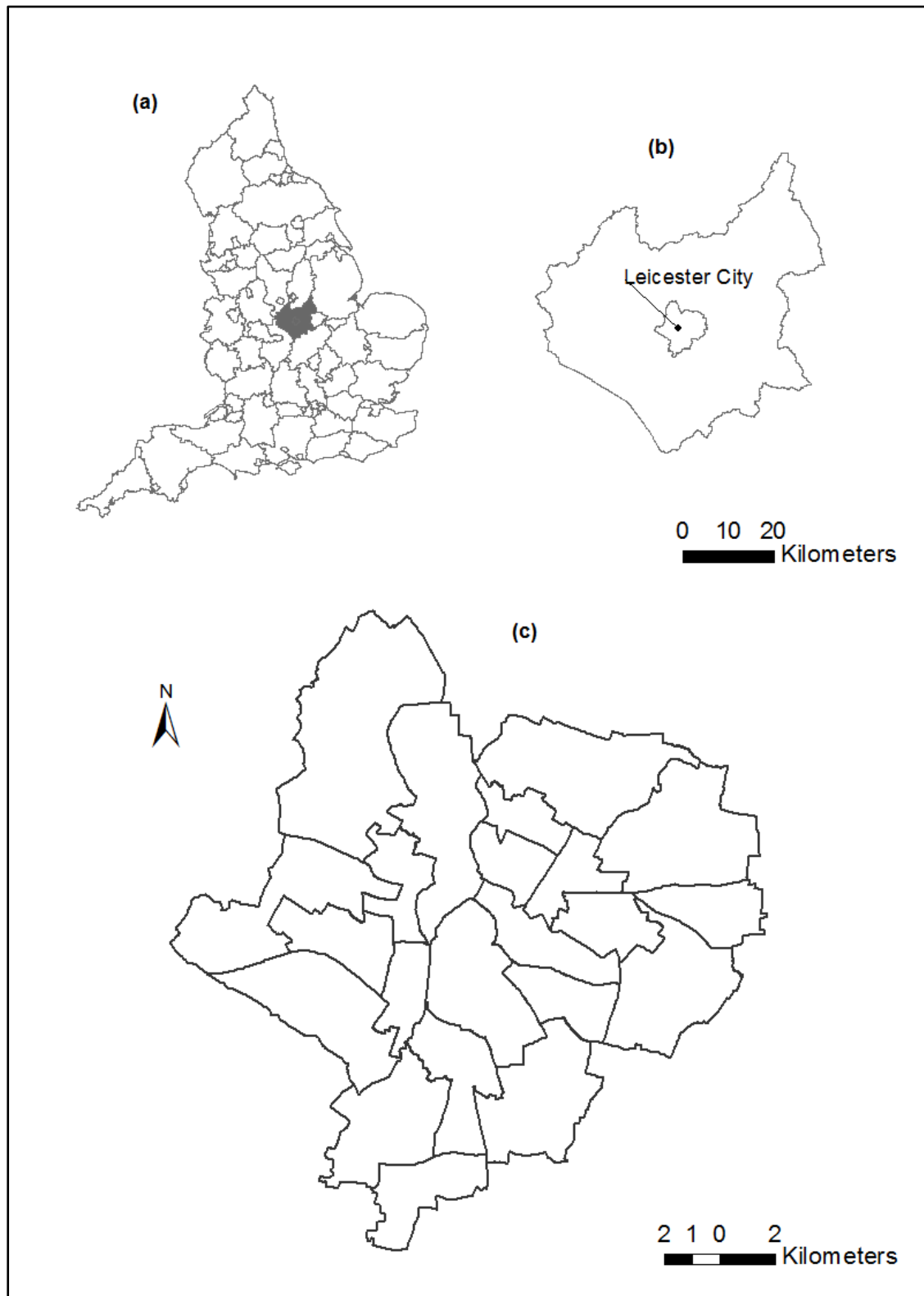


Figure 3.5: Study area showing: (a) the location map of Leicestershire County within the UK, (b) local authority of Leicester City and (c) the source zones (census wards) in Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

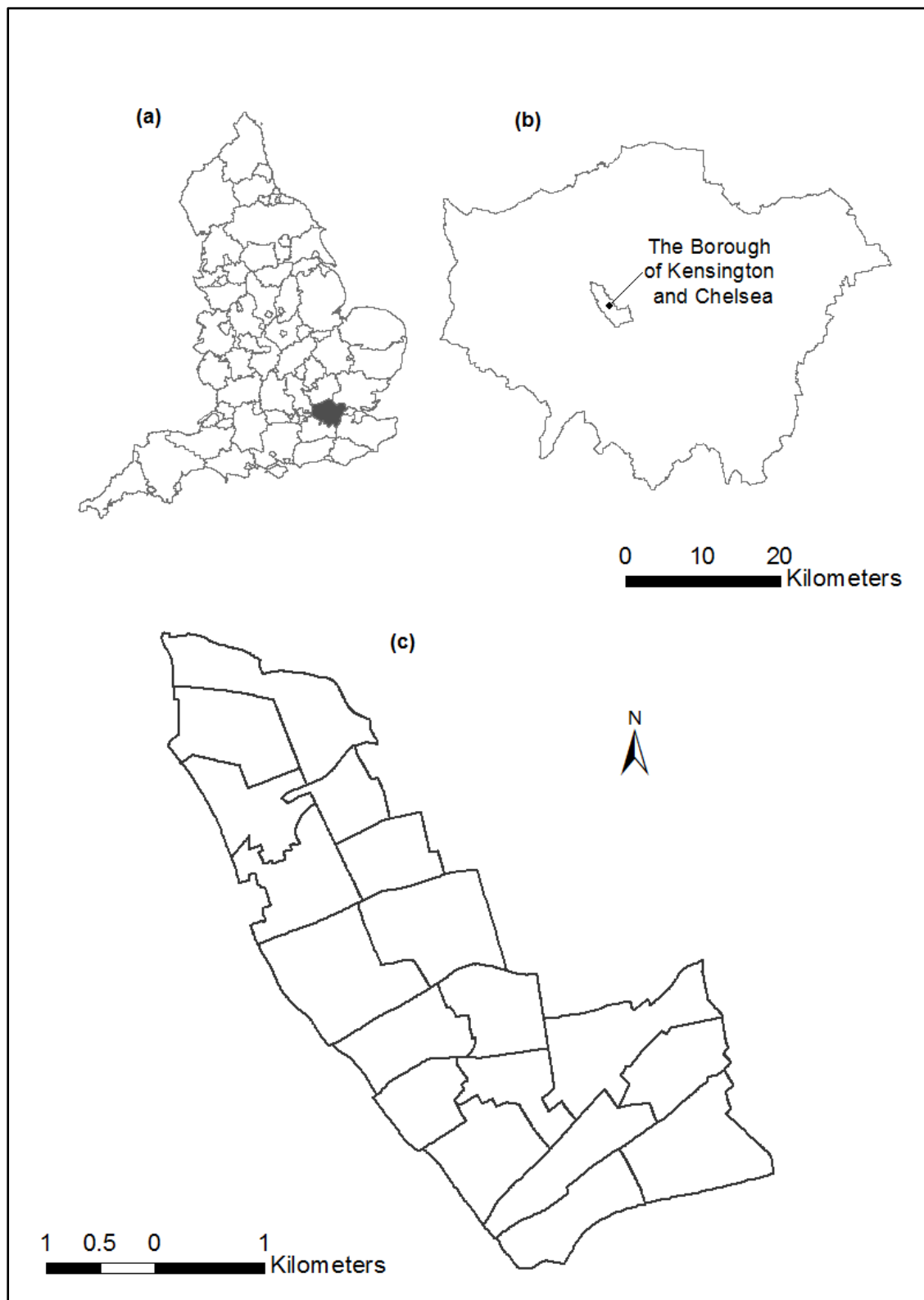


Figure 3.6: Study area showing: (a) the location map of Greater London Administrative area within the UK, (b) the Borough of Kensington and Chelsea, and (c) the source zones (census wards) in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

3.4 Datasets

The selection of the ancillary datasets and the quality of datasets are significantly important because the total population of an administrative area is redistributed to dasymetric zones, and ancillary data and results directly affects estimation results of dasymetric mapping models. Spatial disaggregation processes have been used as control variables in different forms of ancillary input data. Remotely sensed land cover data are mostly used as control variables to distribute population totals. This study aimed to use high resolution imagery data and a range of other external data to provide population estimates. These datasets, obtained in digital formats (vector and raster), are convenient to open and analyse in GIS spatial analysis tools and remote sensing image processing software (see Table 3.1). Table 3.2 below summarises the data requirements for interpolation modelling and are briefly discussed in the following sections.

Table 3.1: Additional datasets showing sources and characteristics

Ancillary Datasets	Sources	Year	Spatial Resolution
DEMs	Airbus Defence and Space	2012-2002	Raster 1 m ²
Aerial Photography	Ordnance Survey – MasterMap [®]	2010-2008	Raster 0.25 m ²
Building Layer 2	Ordnance Survey – MasterMap [®]	2012	Vector - points
Building Footprints	Ordnance Survey – Edina	2012	Vector - polygons
Census Wards	Ordnance Survey – InFuse	2012	Vector - polygons
Output Areas	Ordnance Survey – InFuse	2012	Vector - polygons
StreetView Data	Ordnance Survey – Edina	2013	Vector - polygons
Landmap Data	Mimas – UKMap	2013	Vector - polygons
Average Household Size	Local City Councils – ONS	2013	N/A

Table 3.2: Datasets used in implemented areal interpolation models

Data (raster/ vector)	Spatial unit	Data source and processing
Census geographies (vector)	Census wards (source zones) and output areas (OAs) (target units)	National census 2011 datasets registered to the census dissemination geographies
Aerial photography (raster)	0.25m pixel Image classification The binary classification	Aerial image tiles mosaicked in order to create aerial photography of both study regions
Land cover/use map (raster)	The mosaicked areal image classified (such as, urban areas, green areas and water bodies)	Classified Aerial photography with 4 classes: urban space, grass and tree cover, bare surface and water surface. Generated by Isodata supervised classification using ENVI 4.7
The binary class map (raster)	Land cover/use map is reclassified into two classes (build-up area and non-build-up area).	Land cover/use map reclassified into two classes: urban space and non-urban space and overlaid with the boundaries of census geographies
Intersection map (vector)	Census-intersected-binary class zone (target unit)	The binary class map intersect by census geographies
Building volume data (raster)	LiDAR-derived Digital Elevation Models	The DSMs and DTMs were used to generate building volume information
Building address points data (vector)	Coded address points in vector format	The address points were grouped into two classes as residential and non-residential housing units

3.4.1 Source and target zones, 2011 census of population data

The boundaries of whole local authorities, source and target zones were downloaded from Ordnance Survey Edina Supplier (<http://borders.edina.ac.uk/ukborders/>). The data provides the hierarchy of 2011 UK census dissemination units including the boundaries of the local authority of Leicester City and the local authority of Kensington and Chelsea, census wards, and output areas. A grid map of local authority districts was created with a cell size measuring 60 metres to allocate population totals to each cell. This is because the output areas are the lowest level of geography of UK census hierarchy that are target units of current research, and the size of output areas are varying based on the residential household and population density. These census report geographies were created based on the average number of population and households. In order to determine non-residential uses of target zones a grid map was created. The

cell size was agreed as 60 metres in order to create target cells of local authority districts for representing housing unit density.

In addition to this, the census reported actual population counts of each unit were obtained from the UK national statistics data centre, Mimas was the supplier for both study regions (ONS, <http://infuse2011.mimas.ac.uk/>). These census datasets were provided in excel format by supplying a unique code of each spatial unit. The population information in excel was joined with a vector layer to transfer census values to vector data of source units and target units. These datasets were used in all proposed areal interpolation methods.

3.4.2 Aerial imagery data

The aerial imagery data used with the binary dasymetric mapping model were obtained from the UK national mapping agency (Ordnance Survey). Ninety four tiles of 0.25m resolution ortho-rectified aerial photography for Leicester City, dated 22nd May 2010, were provided. Twenty seven tiles of imagery for the Borough of Kensington and Chelsea, dated 6th May 2008, were obtained. These imagery tiles cover 1 square kilometre and the coordinate system of image tiles is British national grids. They were mosaicked to create aerial imagery of whole study regions and, furthermore, the binary map of built-up and non-built-up area classes was created for operation as ancillary data in the binary mapping process.

3.4.3 Digital elevation models

LiDAR-derived digital terrain models (DTMs) and digital surface models (DSMs) of the study regions were obtained from Airbus Defence and Space (Astrium Ltd. 2014). A Digital Elevation Model is a terrain model based on grid or raster cells. Any cells in the DEM have an elevation value defining the height of the area. Each cell in remotely sensed images has an x and y value representing length and width of the area. In addition to these, the DEM has z value, which represents the height of the area. These elevation models have generated building volume information that was employed as control variables in the volumetric disaggregation model.

3.4.4 OS MasterMap® VectorMap district data

The Ordnance Survey (National Mapping Agency of Great Britain) MasterMap is a constant and maintained Digital National Framework (Holland, 2002). In the OS MasterMap, each single geographical feature is assigned a unique identifier, which is a Topographic Identifier (TOID), offering to analysts making a determined reference to each particular building, land parcel, road segment or other topographic feature in the datasets (Holland, 2002). The framework consists of four discrete layers, which are Address Layer 2, Imagery Layer, Integrated Transport Network™ (ITN) Layer and Topography Layer. The topography layer, for example, provides topographic, cartographic, and road network features (OS MasterMap User Guide). The building polygon features have been obtained from MasterMap® Topography Layer in vector file format from Edina Supplier (<http://digimap.edina.ac.uk/digimap/home>), which is one of the main datasets of the process of population disaggregation. One of the primary properties of OS MasterMap is that each single layer can be integrated with each other layer (OS MasterMap User Guide, 2007). For that reason, various ranges of selections, visualisations and queries can be performed within a GIS by means of OS MasterMap provided attribution information is available. Secondly, building blocks were downloaded from Landmap Supplier (<http://www.landmap.ac.uk/Datasets/Building-Blocks/classification>) in vector file format in order to classify building footprints as residential and non-residential uses. Landmap is a Mimas (a nationally designed academic data centre) service based at the University of Manchester. Landmap-derived datasets consist of building usage information used to determine residential uses.

3.4.5 OS MasterMap® Address Layer 2 data

OS MasterMap® Address Layer 2 (ADDRESS-POINT) is one of the four layers of the Ordnance Survey Digital National Framework. The datasets of ADDRESS-POINT are provided with a National Grid coordinate and a unique reference for all postal addresses in the United Kingdom (ADDRESS-POINT User Guide, 2010). The address point is the first address based dataset in digital format launched in the early 1990s and the current spatial address dataset in Great Britain is the Ordnance Survey MasterMap Address layer 2 (OS User Guide, 2010). When constructing the national address dataset, each different type of structure comprising residential units (e.g. house, apartment, trailer, and duplex) and non-residential structures (e.g. secondary structure, utility, commercial,

institutional and industrial) is recorded for every address point (Zandbergen, 2011). Address points represent the location of all residential buildings and non-residential structures within a particular jurisdiction. Zandbergen (2011) states that several developed countries have national address point datasets, including the United Kingdom. The development of a national address point database is very challenging; however, it is a crucial element in implementing a population census (Zandbergen, 2011). This product was obtained in vector format (each housing unit coded as point) to cover both study regions. The point features are supplied with a unique identifier (TOID), the same as building footprints on the National Grids (MasterMap User Guide, 2008). According to the feature unique identifier, the address layer table interface includes various ranges of tabular information, such as a series of cross-reference identifiers, one or more addresses, features usage information, features structure information, feature post code information, and the classification of features whether they have been changed or not. Therefore, this data allows users to select, visualise and make wide ranging queries in terms of the specified details of geographical features. For instance, geographical features can be categorised according to their residential and commercial postal address in the intended scale. On that sense, the data of OS MasterMap[®] Address Layer 2 has been used as control variables in the address-weighted dasymetric mapping process.

3.4.6 Occupancy information

The knowledge of occupied houses is important in generating housing unit level population estimates. In the data of 'Address Layer 2', all occupied and unoccupied housing units are recorded. The address-weighted method assumes a uniform population distribution per address point. Thus defining the total number of unpopulated housing units is the main issue in the address-weighted dasymetric mapping method. This information can be obtained from local city councils or national statistical databases. The Office for National Statistics (ONS) provides Housing Stock information from the smallest geographical scale to the biggest geographical scale in Britain. In the Housing Stock dataset of Neighbourhood Statistics, all household spaces are grouped in terms of whether they are occupied or unoccupied units, and that dataset is continuously updating in non-census years. According to the ONS Neighbourhood Statistics, the total housing units of 123,125 (96.7%) are occupied and 4,258 (3.3%) are non-occupied in

the City of Leicester local authority. Additionally, total housing units of 78,536 (89.5%) are occupied and of 9,169 (10.5%) are non-occupied in the Borough of Kensington and Chelsea (NeSS, 2014). These numbers of non-occupied housing units may play an important role in decreasing the accuracy of population estimation results if they are considered as occupied. Non-occupied houses could also be a main error source where the data of building occupancy are not available.

3.5 The Binary Dasymetric Mapping Data Implementation

The binary dasymetric mapping model requires land cover datasets to create binary masking zones as populated and non-populated areas. The inhabited areas are used as control variables in order to distribute the aggregated population totals through only residential areas in the binary dasymetric mapping. High resolution aerial photography was used to create land cover classification. This classification was reclassified to obtain binary masking of land cover classes. The following subsection presents remote sensing analyses to process aerial photography for the creation of binary map zones.

3.5.1 Remote sensing image pre-processing

According to Langford (2006) and Eicher and Brewer (2001), remotely sensed imagery is the commonest ancillary data set the in binary dasymetric mapping method. In order to estimate small-area population totals, aerial photography was used as ancillary data in the binary dasymetric mapping. Remote sensing data were acquired as 1 square km image tiles from Ordnance Survey (see Figure 3.7 a, b). The image tiles were mosaicked to define residential uses within the study area, and a subset for the City of Leicester and the Borough of Kensington and Chelsea was generated (see Figure 3.8 and Figure 3.9 respectively). The subsets of aerial photographs were used to select region of interests (ROIs) that were used as training areas in maximum likelihood supervised classifiers. The aerial photography has a 25cm spatial resolution which is suitable to classify different land cover/use of the intended scale. Aerial photography derived land cover information was used to generate a binary map for the population estimates of residential areas. The classification stages will be detailed in the following section.



(a)



(b)

Figure 3.7: Ancillary datasets of aerial photography that used to undertake binary dasymetric mapping methods. (a) is one of the 94 tiles of 0.25 meter resolution aerial photograph of Leicester City cover 1 km² area and (b) is one of the 27 tiles of 0.25 meter resolution aerial photograph of the Borough Kensington and Chelsea. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.

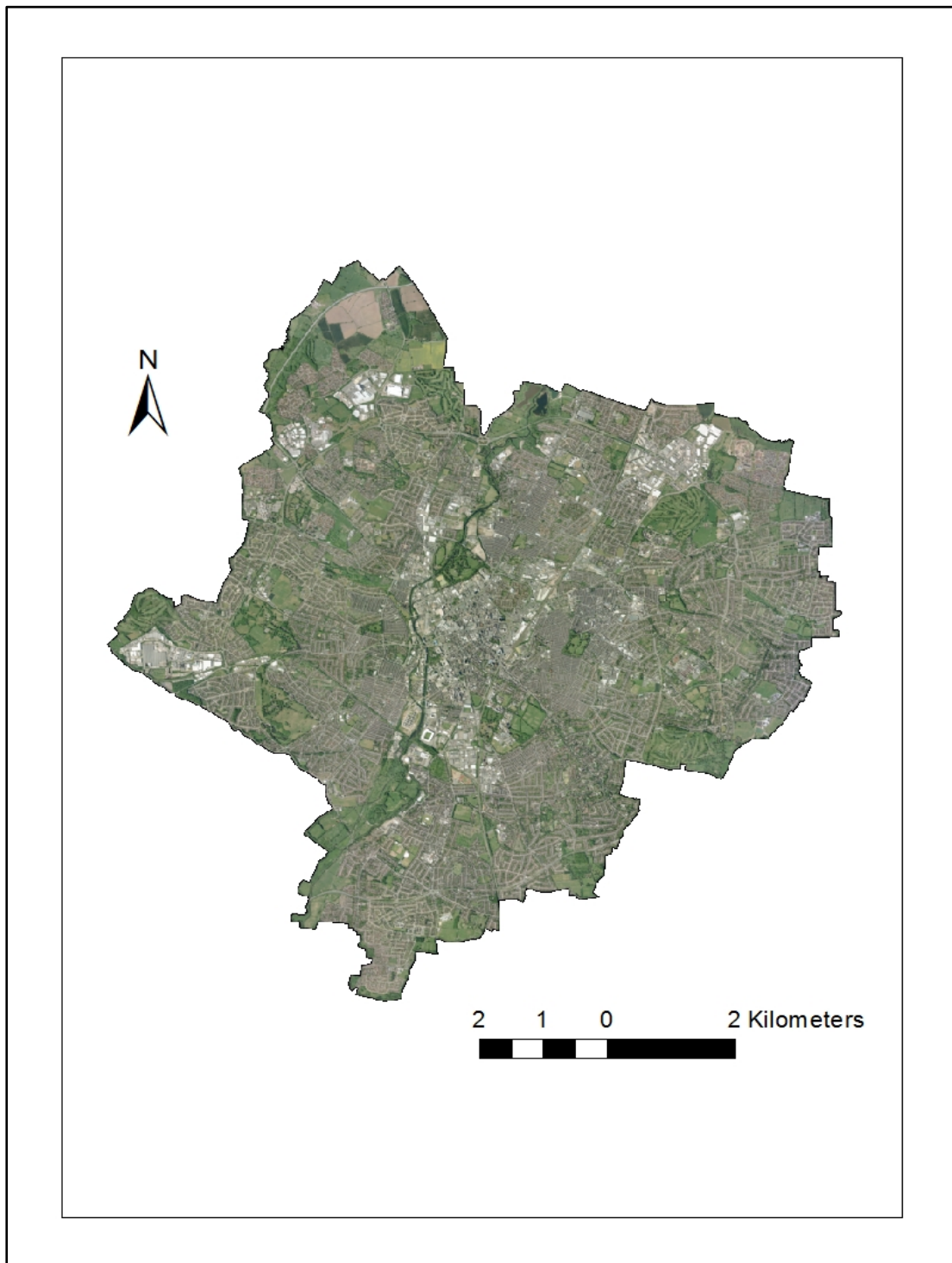


Figure 3.8: The subset imagery of 25cm spatial resolution aerial photography of study area, Leicester City with image band combination of Red, Green and Blue. The enhanced imagery used to employ the binary dasymetric mapping method. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.



Figure 3.9: The subset imagery of 25 cm spatial resolution aerial photography of study area; the Borough of Kensington and Chelsea with image band combination of Red, Green, Blue. The enhanced imagery used to employ the binary dasymetric mapping method. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.

3.5.2 Land cover classification –Supervised classification

This subsection details the steps involved in the creation of a land cover map from the subset imagery under supervised classification technique. The mosaicked images of both case sites were classified to identify different land cover/uses within the study area by using supervised maximum likelihood classifier. A supervised classification process (algorithm) classifies the imagery based on the predefined training samples of similar land cover. The first stage was to define the main land cover classes for the enhanced imagery. Green space, tree cover, water bodies and urban areas are the four land cover classes identified (see Table 3.3). Urban land uses were classified as one class due to the similarities of building blocks. An aerial image can differentiate between different land use structures but may not between residential, institutional and commercial. It might be difficult to separate residential buildings from aerial images than other structures because of changes property uses.

Table 3.3: The land cover classes of study area with four different groups

Land cover class	Description
Vegetation	Parks, golf courses, agricultural land uses and open grasses
Tree cover	Garden trees, park trees, and other vegetation
Water bodies	Streams and lakes
Urban areas	Predominantly residential areas and built-up area

In the second stage, various training sites were selected for each of the defined land cover classes based on the researcher's knowledge of the study areas. A large number of samples for each class were chosen to minimise classification error. These training samples helped to group pixels based on the training selections. Images were classified based on these training selections and the classified images were saved for both case sites in raster format. In the third and last stage, a maximum likelihood supervised classification technique was conducted to obtain land cover map. The classified aerial photography is a raster map with each pixel labelled by one of the four defined classes. Figure 3.10 and 3.11 show the created land cover maps. The pixels of the same land cover type are aggregated to form predefined different land cover types. The classified image was reclassified into two classes that are populated areas and non-populated areas (see section 3.5.5). The accuracy assessment was carried out in the following section.

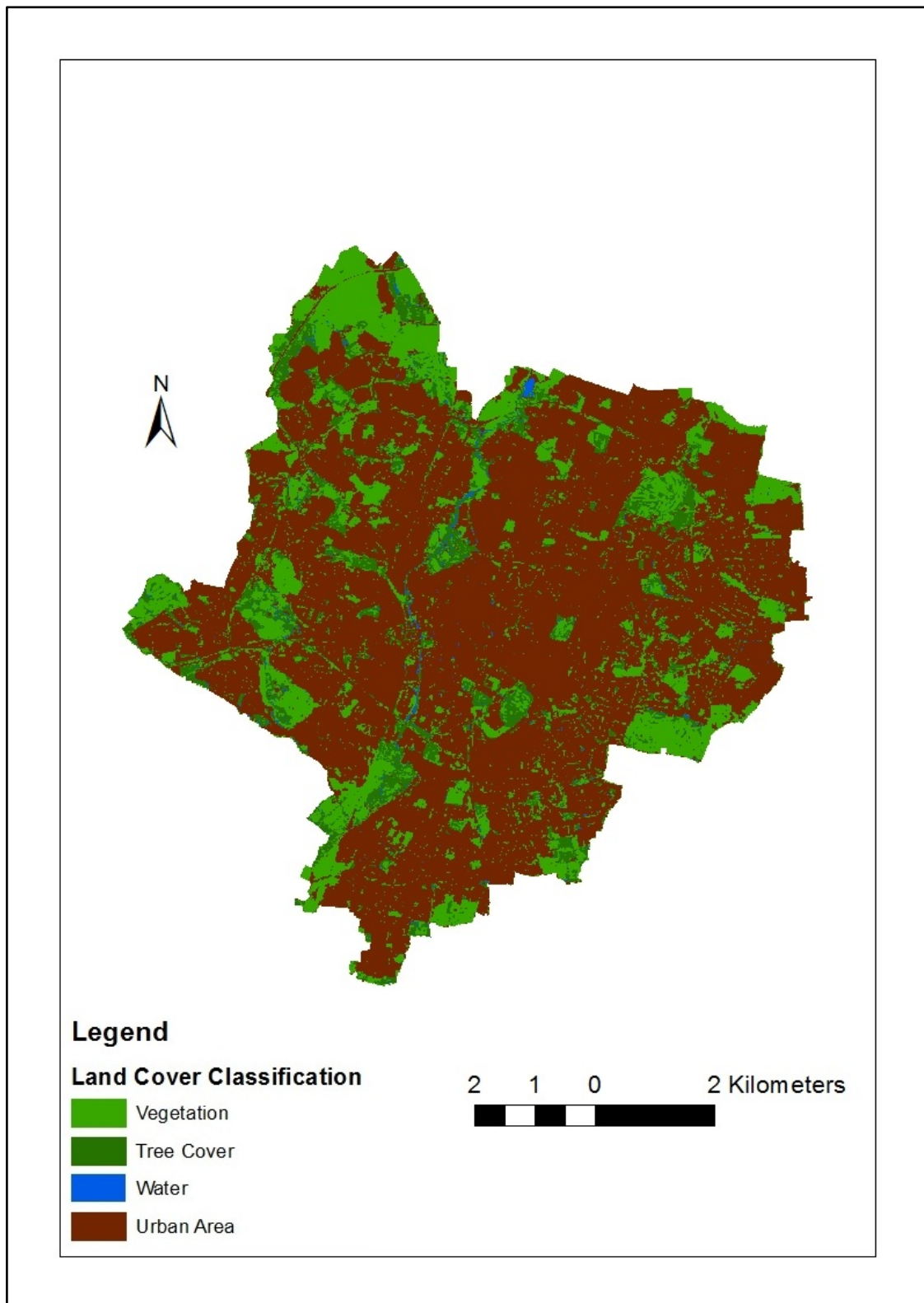


Figure 3.10: Results of classification. The classified image of aerial photography by using maximum likelihood supervised classification. The case site: Leicester City. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.

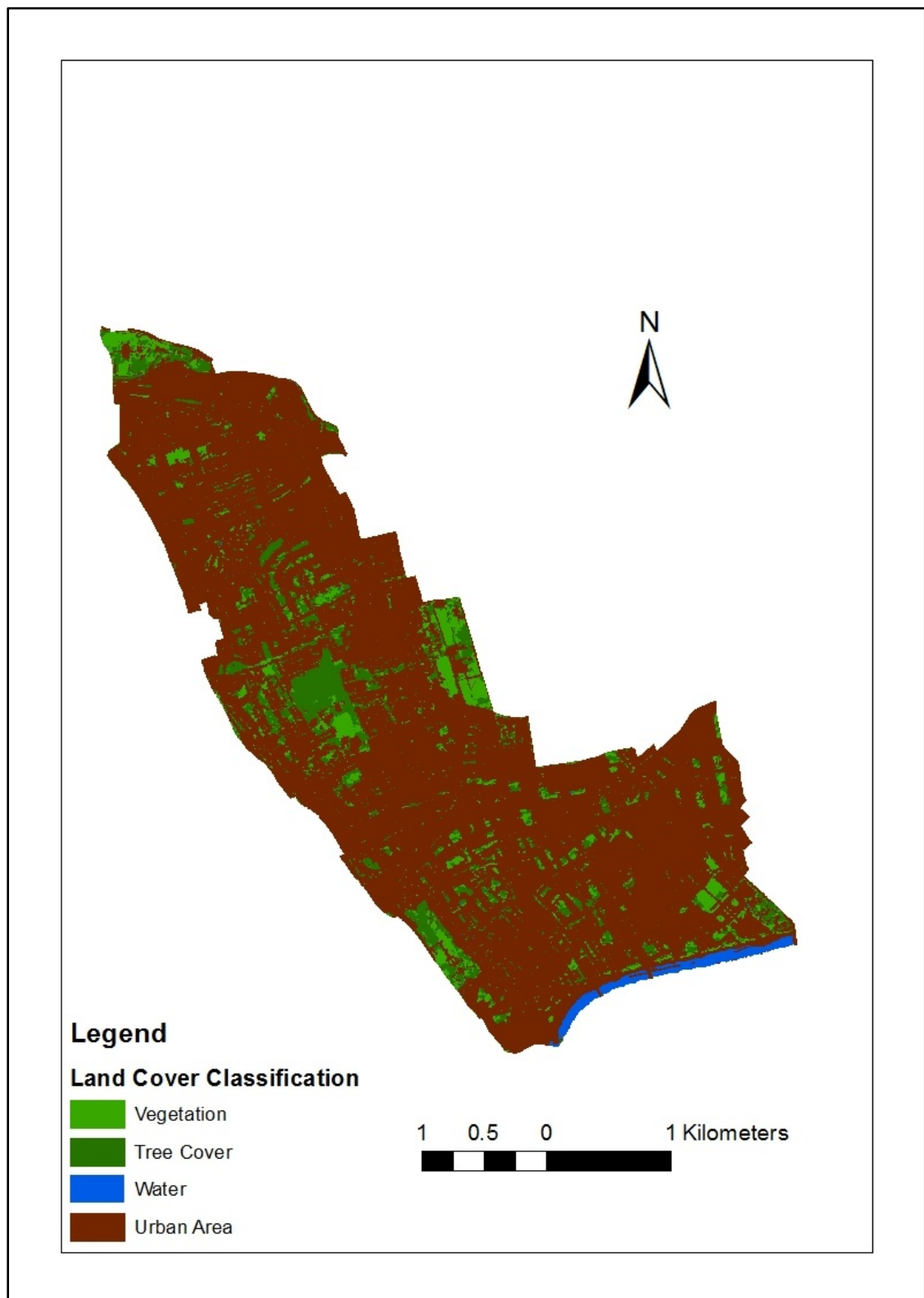


Figure 3.11: Results of classification. The classified image of aerial photography by using maximum likelihood supervised classification. The case site: the Borough of Kensington and Chelsea. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.

3.5.3 Accuracy assessment of land cover classification

The accuracy assessment is an important stage that defines the reliability of classified imagery. The accuracy of classification results were assessed using the ‘Confusion matrix’ tool in ENVI by comparing a classification result with ground truth of region of interests (ROIs). The producers and user accuracy are shown in Table 3.3 and 3.4. According to ENVI User’s Guide (2004, page 639): “Producers accuracy is the probability that a pixel in the classification image is put into class X given the ground truth class is X and user accuracy is the probability that the ground truth class is X given a pixel is put into class X in the classification image”. Photographs were taken to validate land cover classes from randomly selected areas in the study region of Leicester City (Appendix 3). All ground truth ROIs were used as a reference data for accuracy assessment of land cover classification by conducting confusion matrix. An overall accuracy of 89% was gathered for the image of the Borough of Kensington and Chelsea, and 89 % was also obtained for the image of Leicester City. The following tables show that there are some confusion between tree cover and green spaces owing to their spatial similarities (Appendices 1 and 2). This problem may not affect the accuracy of the binary division map because the green space and tree cover were reclassified as non-populated areas and urban classes were reclassified as populated areas.

Table 3.4: Accuracy measurement of land cover classification of aerial photography (Case study area: Leicester City).

Class	Producer’s Accuracy %	User Accuracy %
Green Areas	90.04	97.12
Tree Cover	72.36	45.77
Water Bodies	94.56	94.37
Urban Area	92.91	40.41

Overall Accuracy = (900679/1007593) 89.3 % Kappa Coefficient = 0.7797

Table 3.5: Accuracy measurement of land cover classification of aerial photography (Case study area: The Borough of Kensington and Chelsea).

Class	Producer’s Accuracy %	User Accuracy%
Green Areas	90.24	69.86
Tree Cover	68.63	88.53
Water Bodies	99.50	97.02
Urban Area	94.55	99.80

Overall Accuracy = (318443/355172) 89.6 %Kappa Coefficient = 0.8548

3.5.4 The creation of binary zones -Reclassify land cover map-

This subsection presents the creation of the binary division of land cover map as built-up and non-built-up areas. The binary dasymetric model assumes that total population is concentrated in residential areas (Mennis, 2009, Eicher and Brewer, 2001, Langford and Harvey, 2001). Due to this reason, residential uses of the land cover map were extracted from non-populated land cover. As explained in the previous section, residential uses were classified as urban area, and non-residential areas were classified as green space, tree cover and water bodies. The classified land cover image was reclassified into two classes: (1) built-up areas consist of urban classes and (2) non-built-up areas include water bodies, green areas, tree cover and other vegetation. Figure 3.10 and Figure 3.11 present residential land cover obtained from aerial photography that is used to obtain population density of residential areas. Figure 3.12 and Figure 3.13 show the binary division of the classified images of both study areas. As a first stage, built-up areas were assigned “1” showing inhabited areas, and the non-built-up areas were assigned “0” showing uninhabited areas in the binary map. The population counts were homogeneously distributed into residential land cover classes.

In the second stage, the binary land cover map was converted into a vector file from a raster file and then, residential land use types were extracted to create a layer of built-up areas in the source and target areas. The source areas (census wards) were overlaid with target areas (output areas) using ArcGIS 10.1 analysis tools in order to create intersection layer of target areas. The sum of the built-up areas of source areas and target areas was calculated, respectively. This operation allows the identification of total built-up areas within each source areas. Also, the intersected built-up polygons were overlaid with target areas. In the third stage, the population density of source areas is measured by dividing population totals of each census ward by its total built-up area. This process distributes population counts to all areas of intersection within the study area. The population total of target areas was estimated by multiplying population density of intersection area by target areas. Chapter 4 and 5 presents the binary dasymetric representation of population estimates.

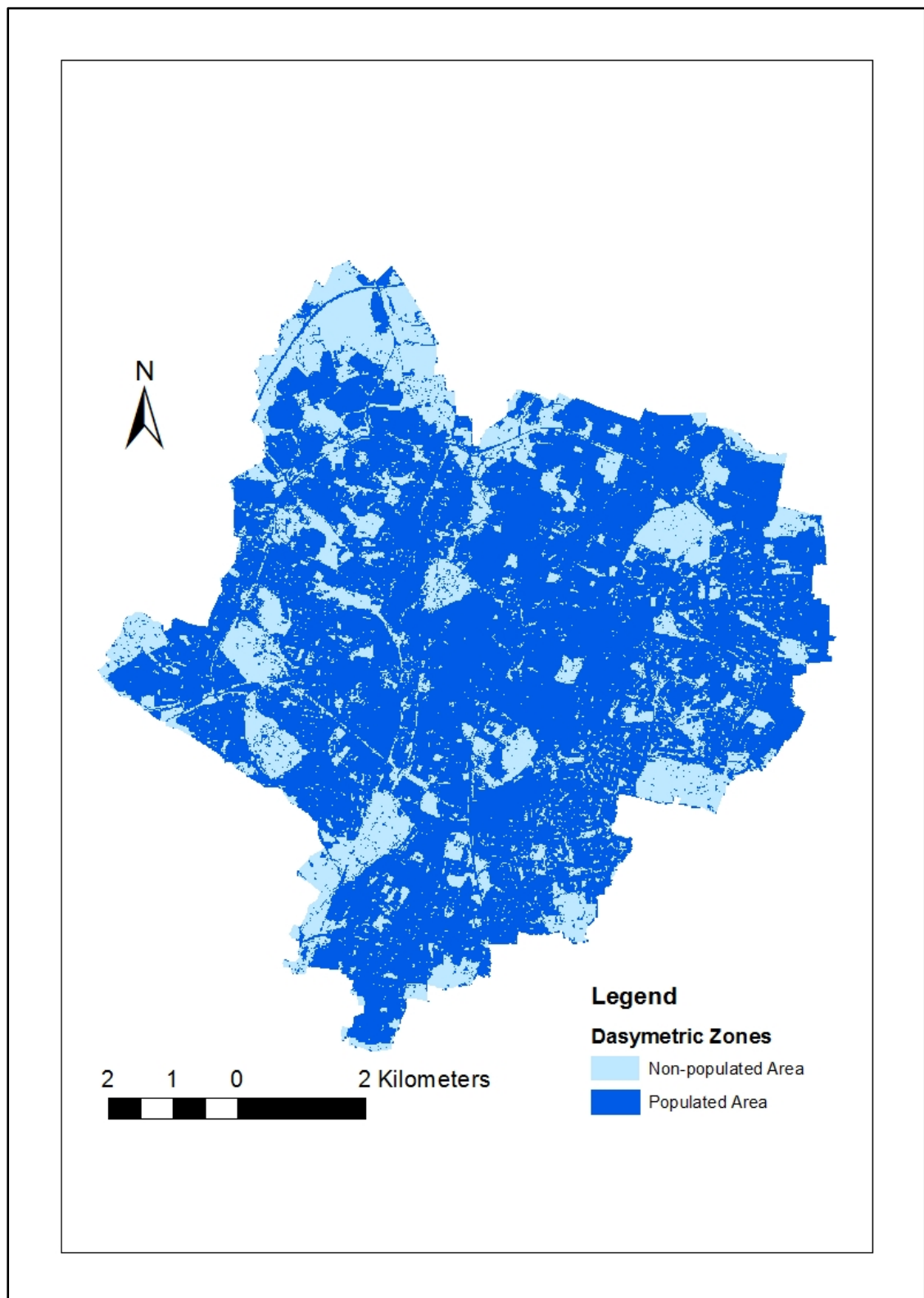


Figure 3.12: The binary dasymetric zones created from classified aerial imagery, case site: Leicester City. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.

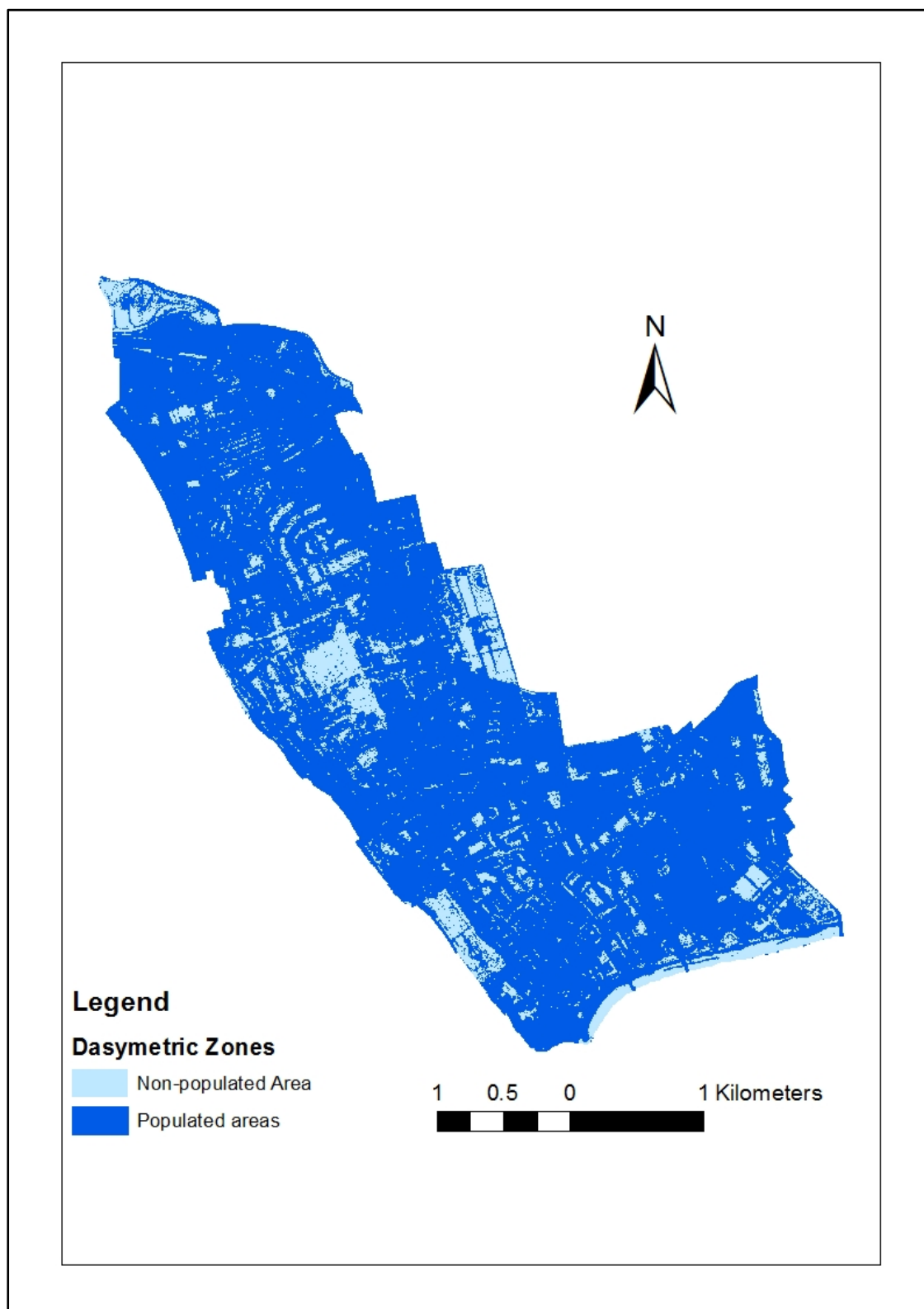


Figure 3.13: The binary dasymetric zones created from classified aerial imagery, case site: The Borough of Kensington and Chelsea. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.

3.6 The Generation of Building Volumes

The volumetric estimation method uses ancillary datasets of building volumes as control variables in order to distribute population totals over built-up areas. As criticised by Sridharan and Qiu (2013), building volumes have been recently used to predict population but have not been utilised for interpolation methods. Basically, the volumetric estimation method does require building height, footprint and usage type information to identify the volume of residential buildings. Building structures may be commercial, institutional and industrial uses. The additional datasets do need several processes (GIS analyses) to identify the volume of residential building blocks within the study areas. These operations, to create residential building volume of each source and target zone, are presented in the following subsections.

3.6.1 Building height information

A digital elevation model (DEM) is one of the main inputs to obtain the height of each building unit. The external inputs of LiDAR data were obtained as digital surface models (DSMs) and digital terrain models (DTMs). Figure 3.14 and 3.15 represent samples of digital elevation models from each study region. LiDAR-derived building heights were used to obtain building volumes that were utilised as control variables in the process of dasymetric population estimation. The DSMs were obtained as 1 square km image tiles. First of all, these tiles were mosaicked to provide the coverage of DSM and DTM for both study regions. There are 98 of 1 square km image tiles of DSMs and DTMs for the City of Leicester and 26 of 1 square km image tiles of DSMs and DTMs for the Borough of the Kensington and Chelsea combined. The DSM represents the surface of earth including all manmade and natural objects on the ground; however, the DTM presents the surface of earth without any determined objects on the ground. The mosaicked DSMs were subtracted from DTMs in order to identify the height of any objects, like buildings and trees on the ground surface. Finally, normalised digital surface models (nDSM) were generated. Figure 3.16 and Figure 3.17 show nDSMs for the Borough of Kensington and the City of Leicester, respectively. The nDSMs were converted from raster pixels to vector points preserving height information of each pixel for the purpose of the creation of building heights. The layer of vector points and the OS layer of building footprints of both study regions were loaded into ArcGIS 10.1 to determine the points which fall within each building footprint. The average height of

points was then calculated and assigned to each building footprints by using spatial analysis tools in ArcGIS 10.1. Later on, the heights of each building block within the source and target zones were determined separately. In the completion step, the heights of each building were multiplied by the area of each polygon to calculate the volume of buildings in source and target zones.

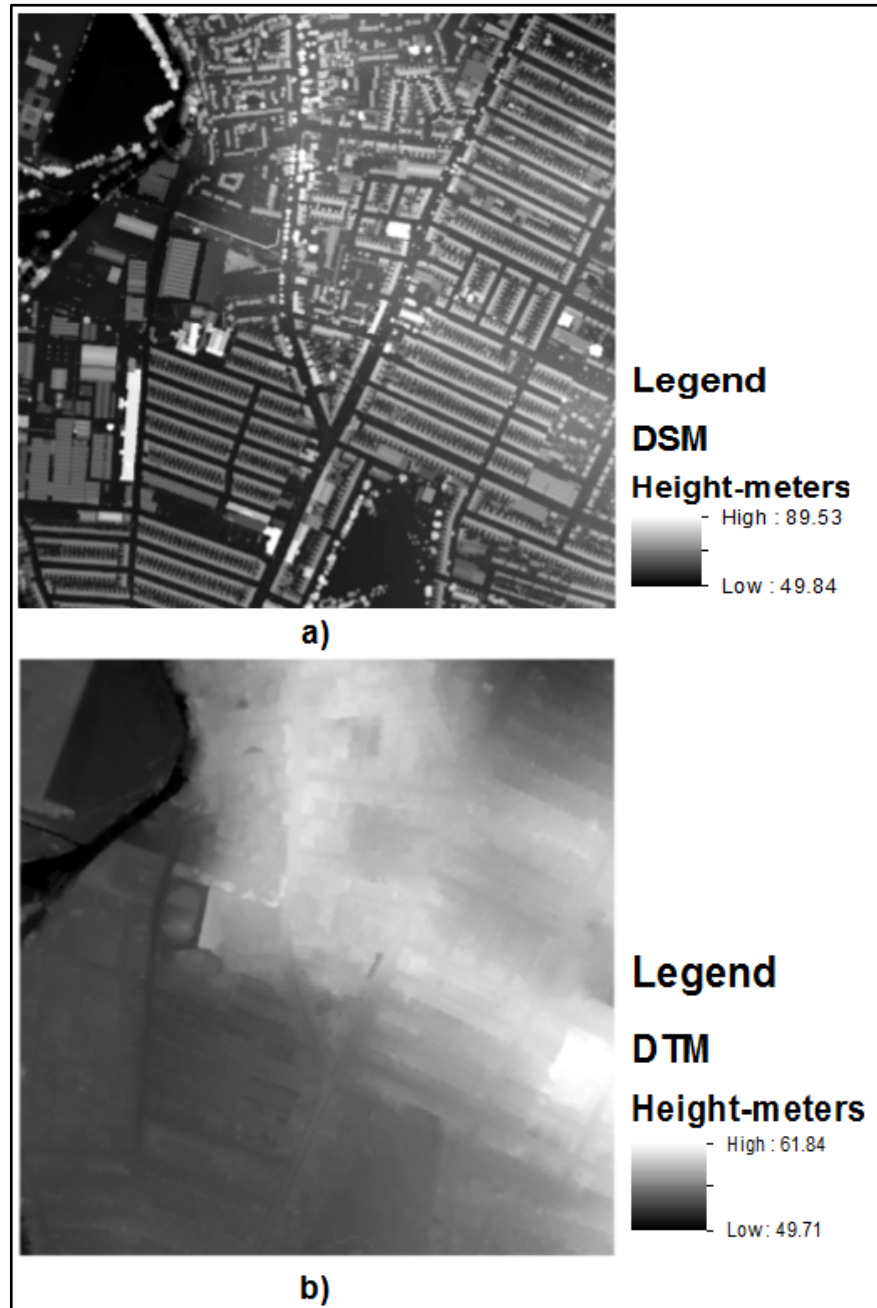


Figure 3.14: 1 meter resolution of digital surface model and digital terrain model of Leicester City, 1km² sample area, All LiDAR data ©Airbus Defence and Space (Astrium Ltd. 2014).

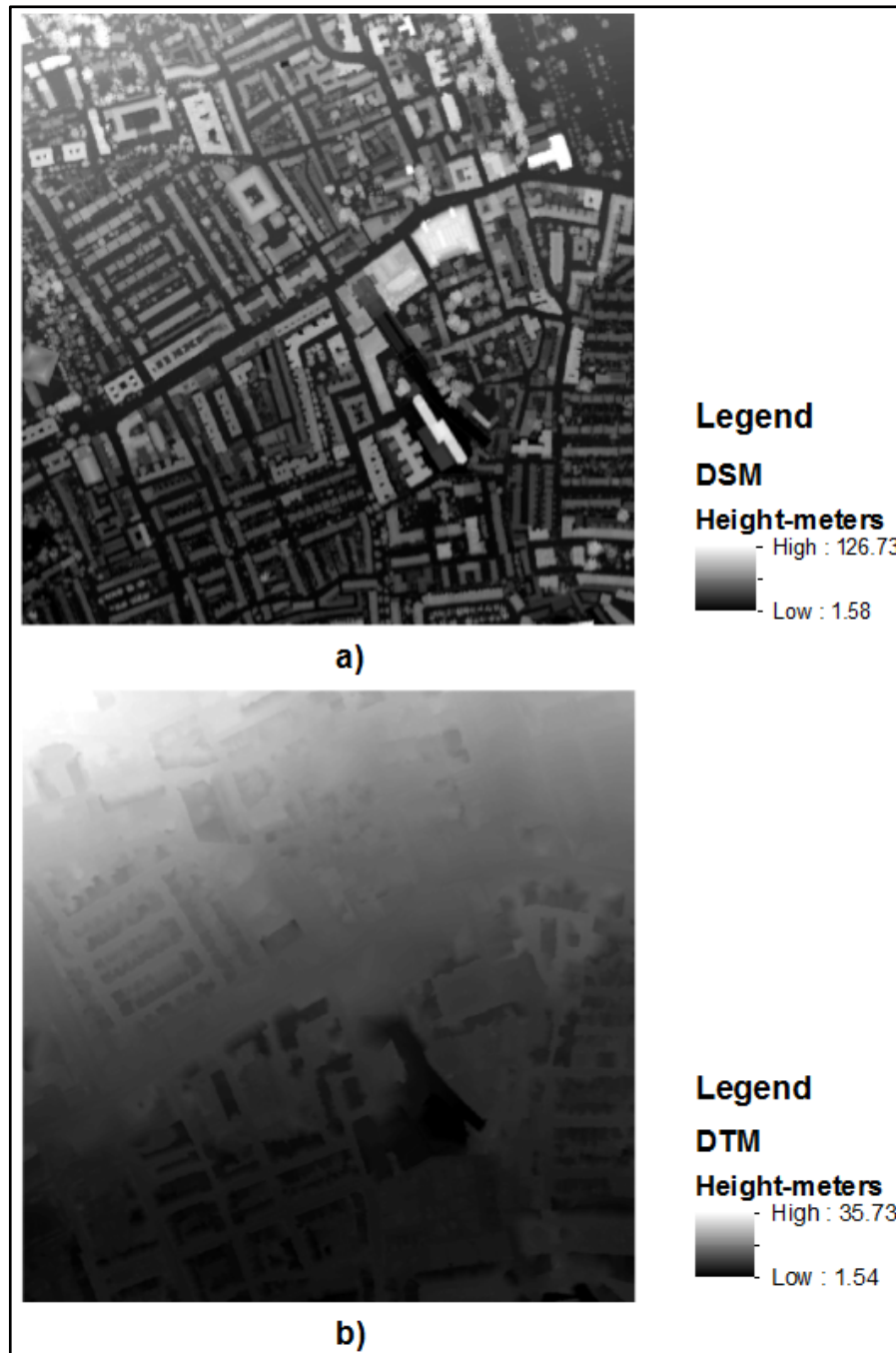


Figure 3.15: 1 meter resolution of digital surface model and digital terrain model of the Borough of Kensington and Chelsea, 1km² sample area, All LiDAR data ©Airbus Defence and Space (Astrium Ltd. 2014).

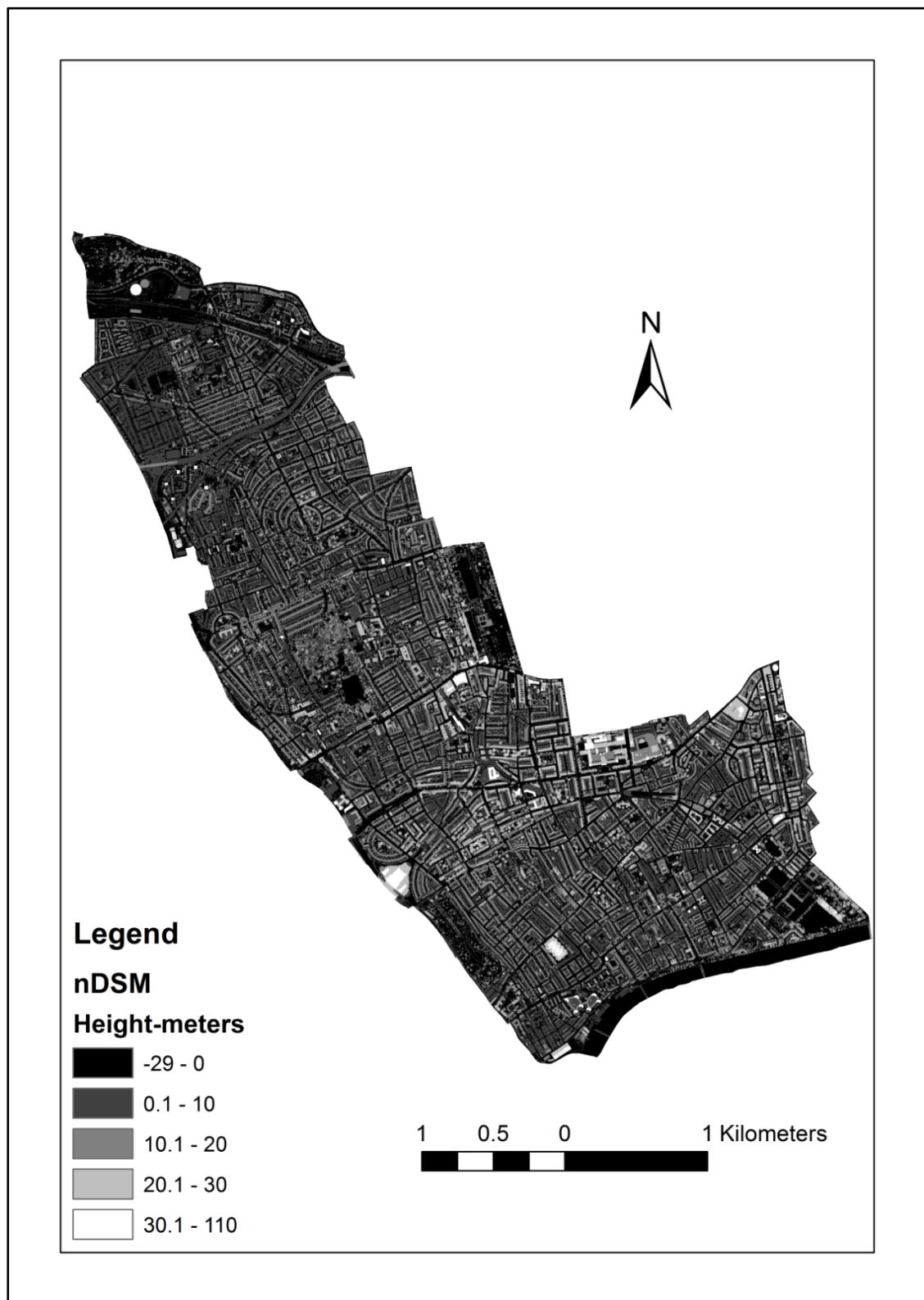


Figure 3.16: Normalised Digital Surface Model (nDSM) for the Borough of Kensington and Chelsea (DSM was subtracted from DTM by using spatial analyst tools in ArcGIS). All LiDAR data ©Airbus Defence and Space (Astrium Ltd. 2014)

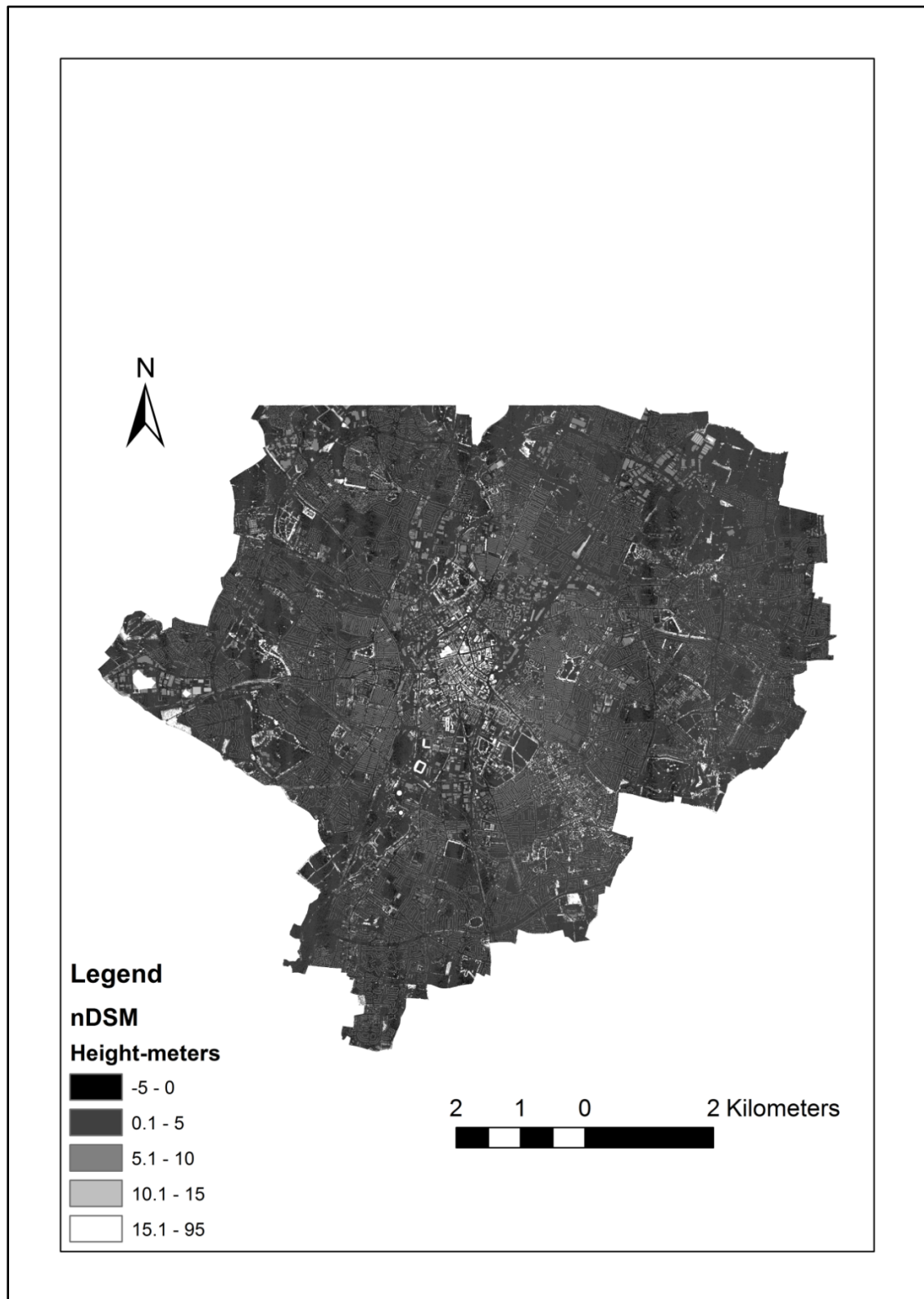


Figure 3.17: Normalised Digital Surface Model (nDSM) for Leicester City (the DSM covers 952 output areas of 969, digital elevation models are not available for 27 output areas on the northern part of the city. All LiDAR data © Airbus Defence and Space (Astrium Ltd. 2014)

3.6.2 The creation of residential building volumes

Building footprints and usage datasets covering the study areas were obtained from the Ordnance Survey © as vector maps (polygons) and building heights were obtained from Airbus Defence and Space as DEMs. In order to create residential building volumes, several analyses were carried out as explained below. The first stage was to identify the height of individual polygons. The polygons of building footprints were overlaid with the points of height information. The height of individual building units was assigned by using the spatial join tool in ArcGIS 10.1 toolbox. After the operation of join, each polygon does have z value representing the height of building footprints. Figure 3.18 shows individual building footprints and their heights. In the second stage, residential buildings were identified using the intersection tool. Building footprint layer was overlaid with building usage type data. These datasets contain detailed information of each building structure such as building usage type (residential, general commercial, institutional, industrial etc.) (Ordnance Survey MasterMap® User Guide, 2012). The accuracy of building usage datasets were validated by comparing OS Street View derived building usage data with OS MasterMap ©. The labelled residential polygons were extracted from non-residential structures within the study areas and saved as new residential building polygons. In the third stage, some filter operations were carried out to purify residential building blocks. The layers were filtered by footprint height and footprint size. Very small polygons lower than 2 metres high were extracted from the layer. These polygons may be building parts where people do not reside. In the last stage, the volume of the individual residential polygon was measured by multiplying the area of each polygon with its height. The obtained building volumes were used as control variables for the generation of estimate population totals as output area level. The resultant building volumes were used as ancillary data in order to improve the accuracy of population estimates.

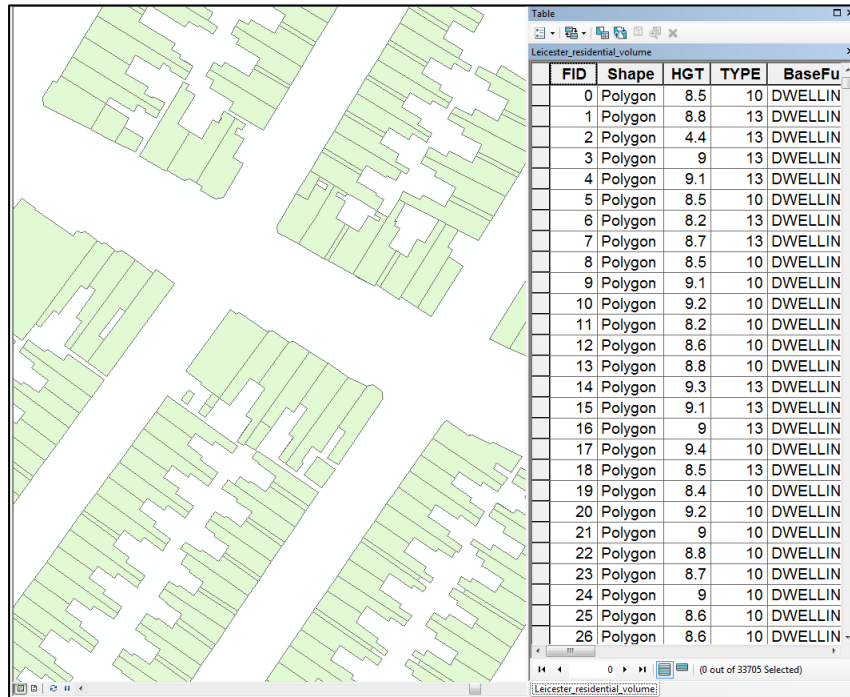


Figure 3.18: The residential building footprints with height information. © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service and All LiDAR data © Airbus Defence and Space (Astrium Ltd. 2014).

3.6.3 The generation of population weighting factor of residential building volumes

This subsection presents the identification of individual residential building volumes of source zones and target zones. Instead of imagery derived residential areas, residential building volumes were used as control variables in order to predict population totals in a volume based estimation model. First of all, the boundary of source zones and residential buildings footprints layer were overlaid using analysis tools in ArcGIS 10.1. The residential buildings were divided through individual sources units (census wards). Secondly, population totals were distributed from source zones to residential buildings (i.e. dividing total population of source zones by their total residential volume) as proposed by Sridharan and Qiu (2013). Thirdly, the boundaries of target and source zones were overlaid with the populated residential building units in order to define intersection between source and target zones respectively. In the final stage, the total residential buildings volumes were used as weighting factors to predict the population of individual target areas. The final residential building volumes were extracted in Excel format to apply volumetric areal interpolation equations.

3.7 Building Address Point Datasets

This section provides a description of the building address point datasets and GIS analyses to make address point data ready to use as ancillary data in the address-weighted dasymetric mapping method. Previously, address point data has been used in dasymetric mapping by Zandbergen (2011) and Tapp (2010). The United Kingdom has a robust national address point database (this dataset is detailed in the section of datasets). Ordnance Survey datasets of building address layer (called as Address Layer 2 or ADDRESS-POINT datasets) contain building features as point map. The data of address points provide comprehensive information of each building unit such as address, usage type, age etc. (Ordnance Survey MasterMap® User Guide, 2012). Point datasets are available from the largest census geographical extent to the finest census outputs. Owing to the availability of residential buildings, the address point data was intended to be made as the main ancillary data inputs for estimating population of study areas in output area level. The following subsections explain how GIS analysis identifies residential address points used as a control variable in dasymetric population mapping.

3.7.1 Separation of residential buildings and non-residential structures

OS MasterMap® Address point data was obtained as a vector map (points) from the Ordnance Survey© that cover both study areas. Address Layer 2 provides the location of each addressable property with a unique reference number in UK (OS User Guide). The frameworks of OS layers have been updated to create a consistent and standard geographical database. These points contain residential and non-residential building usage type information (such as commercial, industrial, institutional etc.). Because of this, building usage data is critical in deriving the total numbers of housing units within the given study area. The separation of residential building from non-residential structure is the main step in the residential building points based population estimation. The data require several analyses to obtain residential address points for source zones and target zones within the study areas.

The implementation stages to derive building usage information were provided here. In the first stage, the address point layer was loaded into ArcMap. The address points which are labelled as dwellings (that means residential housing units) were chosen by

using the selection tool in ArcGIS Analysis toolboxes. The selected features were then extracted as a new layer that contains only residential address points. The second stage was to count the number of address points that intersect within the source and target zones. The extracted residential point map was overlaid with both the boundaries of source zones and the target zones in ArcMap. The sums of the residential units within each source and target area were counted, by utilising intersection and spatial join analysis tools in ArcGIS 10.1. Finally, the source and target layer maps (polygon) output consist of the total number of residential points within each polygon (source and target units). For validation, residential housing units were cross-checked with the category of land use data from the building database in OS StreetView and household data from the NeSS.

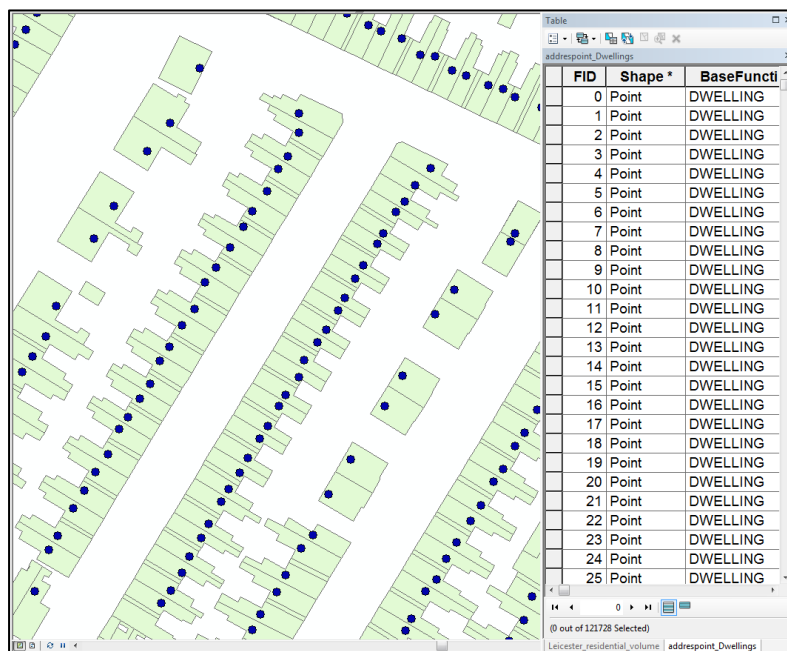


Figure 3.19: The extracted residential building points © Crown Copyright/database right 2012. An Ordnance Survey/EDINA supplied service.

3.7.2 The population weighting factor of residential address points

The average household size from ONS and the occupancy information derived from city councils were used with address layer data to generate population estimates. Residential address point data was obtained in vector format, while average household size and occupancy information were acquired in Excel file based on the census wards and

output areas. To identify the weight of each housing unit within the output areas, several implementations are required.

First of all, the census wards and output areas were exported from vector files into an Excel file. The Excel software was used to calculate the weighting factor of the residential units in different output areas. The data contains the name of individual output and the number of residential points which fall in each polygon. Secondly, the occupancy information of each census ward and output areas were combined. The numbers of unoccupied housing units of output areas were subtracted from the total number of residential housing units. Thirdly, the average household size of each source unit and target unit was multiplied by the numbers of total occupied housing units in order to find the population weighting factor. Fourthly, population totals of output areas were estimated based on the address point data, average household size and occupancy information. Lastly, the obtained population estimates in the Excel tables were combined with vector file layers to visualise the estimated population counts of output areas.

3.8 Software Used for the Research

A variety of software packages are required to perform the various analyses involved in this research. ENVI 5.0 image analysis software, ArcGIS Desktop 10.1, Minitab 16 statistical software and R statistical programming software comprise the major packages used.

1. The ENVI image processing package was utilised for the image processing of aerial photography. Mosaicking the tiles of aerial photographs and sub-setting the study areas and the classification of the mosaicked final image into 1 and 0 maps.
2. ArcGIS analysis tools were used to convert the classified image from raster to vector in order to calculate the extent of built-up areas within the source and target areas. The analytical tools in the software were also used to measure dasymetric population estimates.
3. The R statistic package was used to create scatterplots of estimated population versus actual population totals. Also, statistical analyses, including coefficients

of the results, root mean square error and R squared were calculated using this software.

4. Minitab 16 Statistical Software was utilised to generate scatterplots for the comparison of interpolation estimated population totals with census reported actual population values. Additionally, the standard error of regression was calculated in Minitab.

3.9 Accuracy Measurements for Estimation Results

In previous research, a variety of error measurements have been popularly used to measure the magnitude of estimation error for areal interpolation models population estimates (Su et al., 2010, Hawley and Moellering, 2005, Gregory, 2002, Cockings et al., 1997, Fisher and Langford, 1995). In order to assess the overall accuracy of the deployed population disaggregation model, a range of measurements including R squared (R^2), root mean square error (RMSE), absolute relative error (RE), percentage error (PE) and standard error of regression have been measured in this study. The regression coefficient and the standard error of estimated values calculate precision showing the extent to which predicted population of an area is close to true population of target areas. The description of RMSE given by Fisher and Langford (1995) is defined in Equation 3.5. The RMSE employs absolute values of the difference between known census population and the estimated population within every target area (Gregory, 2002).

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - y_i)^2} \quad (3.5)$$

Where;

x_i is the know census population of target zone,

y_i is the estimated population of target zone,

m is the number of target zones.

Additionally, Relative Error (RE) compares the estimated results of disaggregation models with census reported actual values to examine the goodness of the interpolation methods. RE is formulated as shown in equation 3.6;

$$RE = \frac{(P_e - P_g)}{P_g} * 100 \quad (3.6)$$

Where;

RE is Relative Error,

P_e and **P_g** are the estimated and actual population values, respectively.

It helps to evaluate error distribution in relation to population density (Lu et al., 2006).

The results of relative error were multiplied by 100 to calculate percentage error.

3.10 Summary

This chapter has presented the external data inputs that were utilised as control variables in the dasymetric process. These ancillary datasets are in both vector and raster format and they were processed in spatial image processing software and GIS programs for use in interpolation methods. Additionally, spatial analysis of remote sensing and GIS datasets have been used as control variables in dasymetric mapping process. Ancillary data used as control variables to predict population totals were processed in the previous sections (3.5, 3.6 and 3.7). In the first section, a land cover map was created from the aerial photography. The classified imagery was reclassified to define the extent of residential area uses for handling as auxiliary data in the binary dasymetric mapping method. In the second section, the LiDAR-derived building height data and building footprints were processed to obtain building volumes as ancillary data in the volumetric estimation method. In the third section, Ordnance Survey acquired address point data were processed in ArcGIS 10.1 in order to separate residential building units and non-residential structures to use as control variables in the address-weighted dasymetric interpolation model.

Furthermore, the proposed spatial population disaggregation algorithms deployed to estimate population totals at output area level for the City of Leicester and the Royal Borough of Kensington and Chelsea were presented. Each areal interpolation model used census wards as source zones and output areas as target zones to general spatial population distribution surfaces. Population counts were simply redistributed to smaller areal units depending on external datasets. The applied population estimation models are volume preserving approaches. The total population of each census ward is the same with the disaggregated total population of output areas which fall within the same

census ward. The actual population values of target zones were utilised to assess the goodness of the interpolation algorithms. These interpolation techniques and available data sources were selected to compare performances of interpolation methods for obtaining precise estimation results in both different densely populated areas. These areal interpolation techniques were chosen to explore whether a relation exists between the selected ancillary variables and residential population density. In order to test correlation, the estimated population values were compared with the actual population figures reported by the census. This also shows the reliability of ancillary data. Finally, the performance of each of the proposed dasymetric mapping models was observed by looking at the results. The results obtained from areal interpolation models of each of the case sites are represented in Chapter 4 and Chapter 5.

CHAPTER 4

Areal interpolation Results for Leicester City

4.1 Introduction

A number of areal interpolation techniques have been previously improved and used in many studies in different environments (explained in Chapter 2). Three existing and two novel spatial disaggregation methods were used to predict population totals in the study region. These five models are areal weighting, the binary, the volumetric estimation and the address-weighted dasymetric mapping and occupancy information based interpolation approach. These distinct disaggregation techniques have their own assumptions to predict population totals in target areas. The UK 2011 census population counts were distributed from larger spatial units to smaller units. In each method, census wards were used as source zones, the output areas used as target zones and additional information was utilised as control variables in order to estimate population totals for unitary authority of Leicester City. Therefore, the aim of this chapter is to present the results of population estimation for the City of Leicester. In this research, the address-weighted method disaggregated population totals based on the number of residential housing units (derived from OS, 2012) within the target areas. The binary dasymetric technique makes use of additional information derived from supervised classification for aerial images (obtained from OS, 2012) in order to spatially disaggregate population counts. Whereas the volumetric estimation method used residential building volumes (produced from datasets from OS, 2012 and Airbus Defence and Space, 2014) as external input to obtain population estimates at intended level, areal weighting approach used only source zones and target zones to distribute population totals.

In order to assess the accuracy of estimated population totals, several error measurements were used, including root mean square error (RMSE), the standard error of regression (S), and R squared (R^2) in a manner similar to Gregory (2002), Cockings et al. (1997) and Fisher and Langford (1995). These tests present the errors between estimated variables and actual population totals. Additionally, these accuracy measurements display the relationship between true and predicted population values.

The presentation of the results of areal interpolation methods applied to predict the population totals of small areas is organised as follows. The results of the areal interpolation approaches are presented in Section 4.2. Section 4.3 compares the areal interpolation processes deployed to predict population totals. This section also includes the comparison of the area-weighted and binary mapping interpolation techniques; in addition to the comparison of the volume based estimation model and address-weighted dasymetric method. Finally, Section 4.4 provides a summary of population estimation results for the City of Leicester.

4.2 The Results of the Areal Interpolation Approaches

Five areal interpolation methods were employed for obtaining population estimates and comparing their results: the area-weighted, binary dasymetric mapping, building volume based estimation models, the address-weighted and occupancy information used address-weighted interpolation techniques. The estimation results from each of these methods is detailed in the sections that follow.

4.2.1 Areal weighting interpolation method

Areal weighting interpolation was the first technique utilised to disaggregate census population totals from one areal unit (source zone) to other, different areal units (target zones). This interpolation approach was applied in order to estimate the population totals for Leicester City at the target units (output areas). Figure 4.1 demonstrates the results of the areal weighting technique with regards to the all census wards in Leicester City. The method assumes that residential buildings and population totals are distributed homogenously through census geographies. The areal weighting method does not consider different land uses within the source zones. The distribution of population counts appears mostly dissimilar in different land use types. For instance, the population density in urban areas is expected to be higher than that of rural areas. Therefore, the size of output areas and the population densities are the main factors that affect the weight of the population estimates within individual target units.

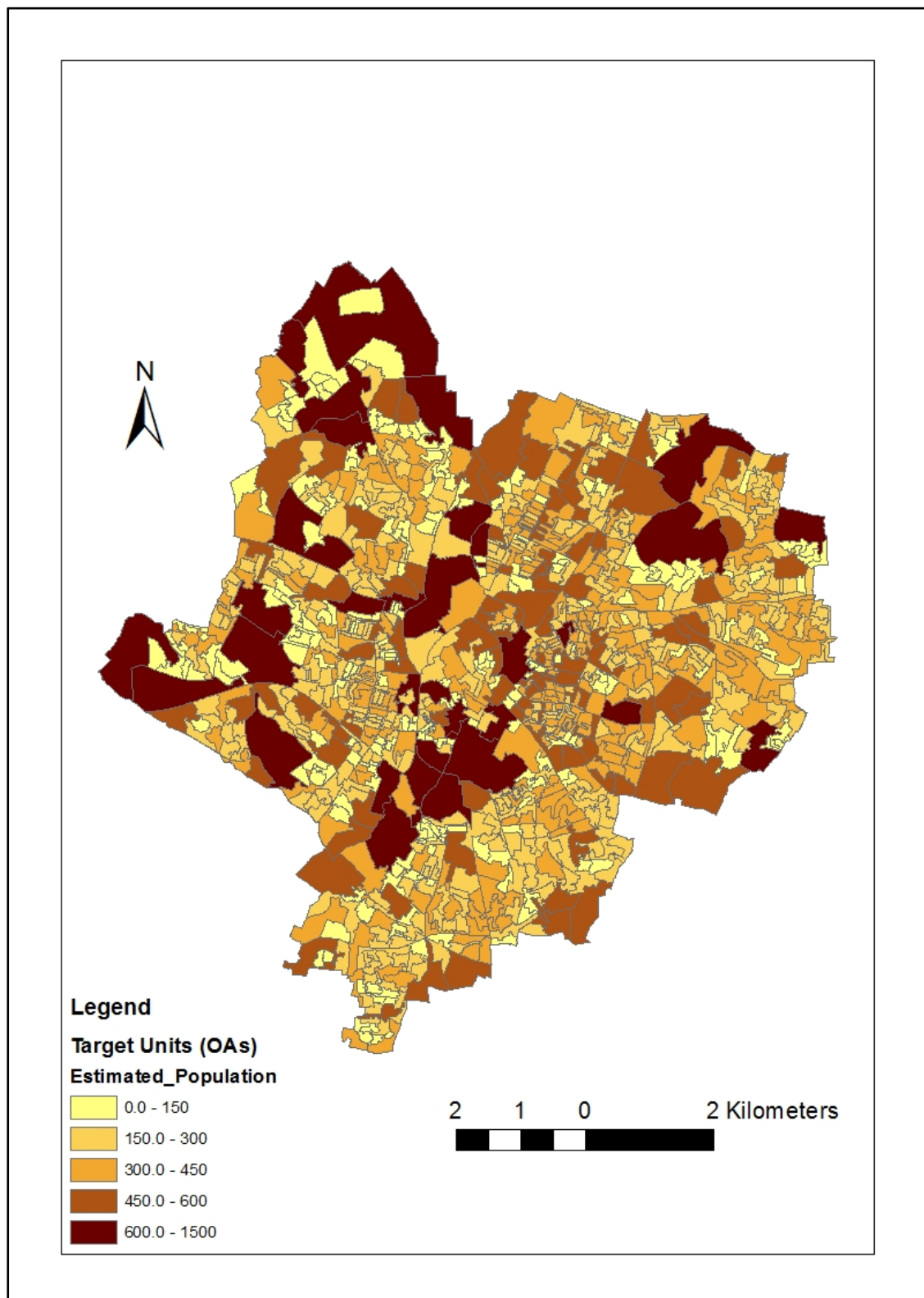


Figure 4.1: The population estimation results of areal weighting interpolation method in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

The 2011 census counts were disaggregated from census wards to output areas using areal weighting interpolation approach. The population density of each source zone was calculated by dividing the area size of census wards with their total population. The population density of the individual census ward was used to as a control variable to reassign the population counts to each target zone. The area-based population estimates were mapped at output area level. Figure 4.1 above shows a map of estimated population totals for the City of Leicester. The legend of the map demonstrates the range of estimated population values in five classes with colour changing from yellow to dark brown as the population totals increases. In the figure, output areas with low population were labelled with a yellow colour and output areas of those with high population are in a dark brown colour.

The visualisation of the results of the area-weighted interpolation explicitly shows that more people are located in the large-sized output areas (brown-coloured spatial units) and there are less people located in small-sized output areas (yellow-coloured spatial units). As can be clearly seen from the aerial photography (Figure 3.8 in Chapter 3, page 65), those larger-sized output areas overlap with non-residential uses such as green spaces, parks, golf courses and agricultural uses. Non-residential areas such as the pattern of commercial, industrial and shopping centre areas cover the most part of the study region. These zones are mostly close to the boundary of local authority of Leicester City. Also, the small-sized spatial units overlap with predominantly residential areas, which are scattered within the study area, and they are largely concentrated around the city centre and towards the southern and eastern parts of the city. The area-weighted interpolation technique assumes uniform population distribution within the source zone without considering these non-residential zones (Zandbergen and Ignizio, 2010, Maantay et al., 2007, Flowerdew and Green, 1992). Thus, population totals assign the areas where people do not live in reality as displayed in aerial photography of the city. Therefore, the area-weighted method overestimates large-sized output areas and underestimates in small-sized output areas.

A regression model was produced to establish the relationship between predicted and actual values (see Figure 4.2). When the estimated population totals were plotted against the known population counts, patterns of the points were widely scattered randomly around the regression line. This is an indication to the not strong relationship

in the regression model, and there are various numbers of outliers. This observation shows that the performance of the areal weighting model fully depends on the size of the source and target zones and the assumption of the homogenous population distribution within the study region. It is clearly noticeable from the experiments that the larger output areas are subject to overestimation on the one hand, and relatively smaller output areas are subject to underestimation on the other hand.

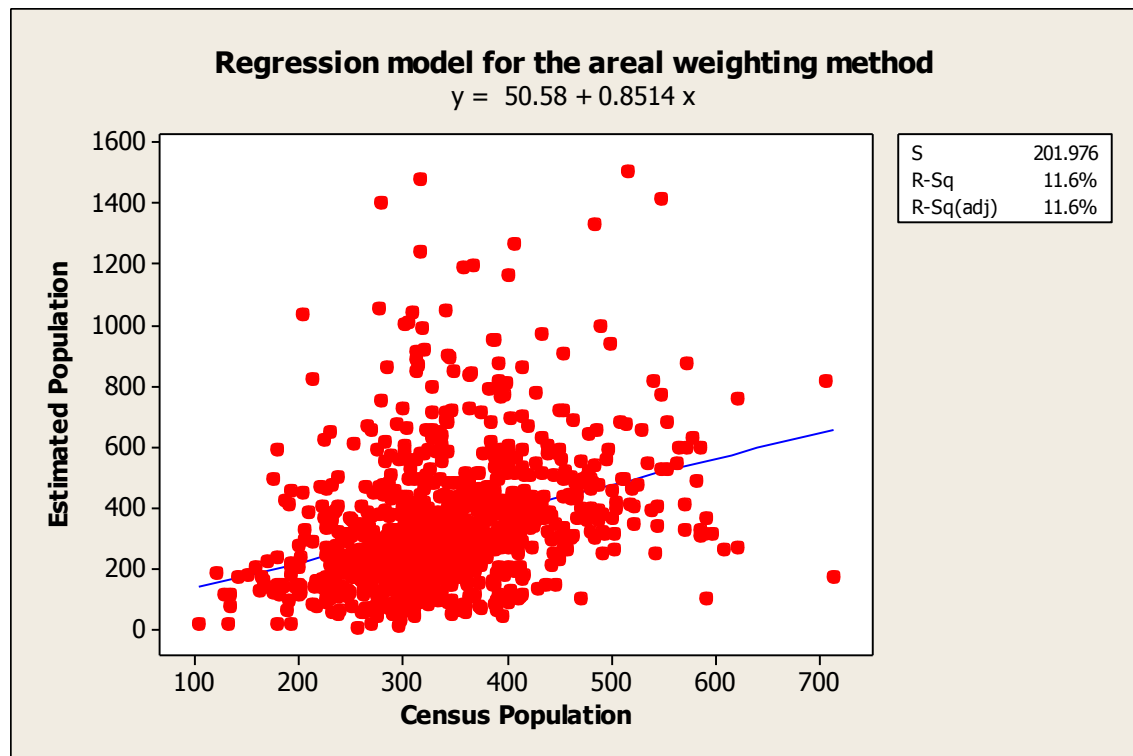


Figure 4.2: Regression model for the areal weighting interpolation method. The estimated population totals of the City of Leicester at output area level versus the actual population of output areas released by census.

The estimated population of the output areas were compared with known census population information to calculate the degree of error introduced. Accuracy is measured using the root mean square error (RMSE), and the square root of the mean square error (S). The way Fisher and Langford (1995) define and describe RMSE metric was already presented in Chapter 3 (page 84). The RMSE employs absolute values of the difference between known census population and the estimated population within every target area (Gregory, 2002). The area-weighted model has the correlation R^2 is 0.11, which suggests a very low relationship between the predicted population values and census reported actual population counts. Table 4.1 demonstrates the accuracy

measurements of regression coefficients, R squared and RMSEs for each applied disaggregation model. The Areal weighting method was also compared with actual population totals in the following section (4.3) to compare the result of areal weighting based population estimates.

Table 4.1: Accuracy measurements for the areal weighting interpolation method obtained population estimates for the City of Leicester.

Accuracy Measurements	Accuracy Results
Regression Coefficients	0.851
R Squared	0.116
Standard Error	201.976
RMSE	9.123
Relative Error (RE)	0.4101
Percentage Error (PE)	41.01%

Overall, all these findings suggest that the area-weighted model can be employed in environments where population density is homogenous; however, the model may not be used in areas where population totals are distributed heterogeneously. Under normal situations, population totals are distributed heterogeneously but the areal weighting method obtained very poor results, similar to Hawley and Moellering (2005) and Zandbergen and Ignizio (2010). The population results are mostly overestimated in the edge of city where non-residential areas cover large places and the estimation results are commonly underestimated in the densely populated output areas which are close to the centre of the city. This method was used as a benchmark, and further population estimates are improved using ancillary data to population counts as heterogeneously within the source zones, as in the following dasymetric mapping model.

4.2.2 Binary dasymetric mapping method

Binary dasymetric mapping was the second interpolation method considered to generate population estimates at output area level. This dasymetric method is relatively simple, using ancillary information to define the extent of built-up areas in order to disaggregate population counts through spatial units (Eicher and Brewer, 2001). While the approach assumes that population totals are heterogeneously distributed considering non-

residential uses within the urban areas, the assumption of the area-weighted method is that of a uniform distribution. Land use and land cover maps have been commonly used to estimate population (Liao et al., 2010, Mennis and Hultgren, 2006a, Tian et al., 2005). The implemented binary dasymetric mapping technique used census wards as source zones and the classified land cover data as ancillary input data to redistribute population totals to small areas (output areas) in the City of Leicester.

The 2011 census of population counts were disaggregated from census wards to output areas based on the classified built-up areas. The population density of each intersection built-up area was measured by dividing the intersection area of census wards with their population counts. This calculated population density was used as a control variable to generate population estimates in each target unit. The land cover based population estimates were mapped at output area scale. First of all, Figure 4.3 demonstrates the binary dasymetric surface that population totals distributed to only built-up areas. Population counts were allocated to grid cells. The binary dasymetric zones and the 60 metre cell-sized grid maps were overlaid to measure the weight of cells to allocate population totals to each cell. In the dasymetric surface, grid cells, which mostly overlap with unpopulated land cover, are labelled with a white colour and populated land uses are labelled with yellow and brown colours. Secondly, the total population of output areas were measured by summing population of cells which falls within the output areas. Figure 4.4 shows the estimated total population of output areas by using the binary dasymetric mapping. The legend of the estimation map shows the range of predicted population totals in five classes with colour changing from yellow to dark brown as the population totals increase. In the estimation map, output areas with low population are labelled with a yellow colour and output areas with high population are in a dark brown colour.

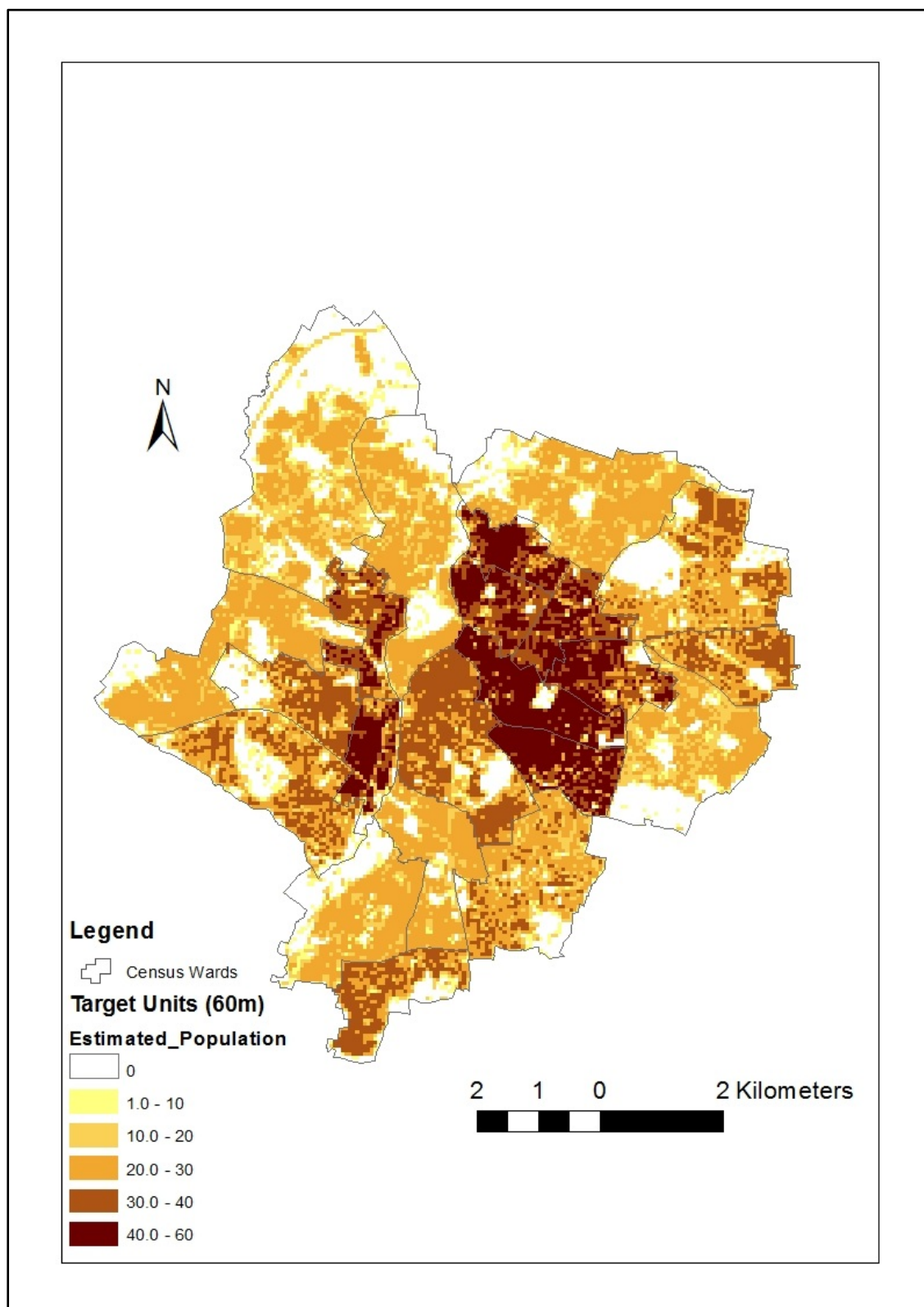


Figure 4.3: Dasymetric surface of Leicester City obtained by using the binary dasymetric mapping method. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

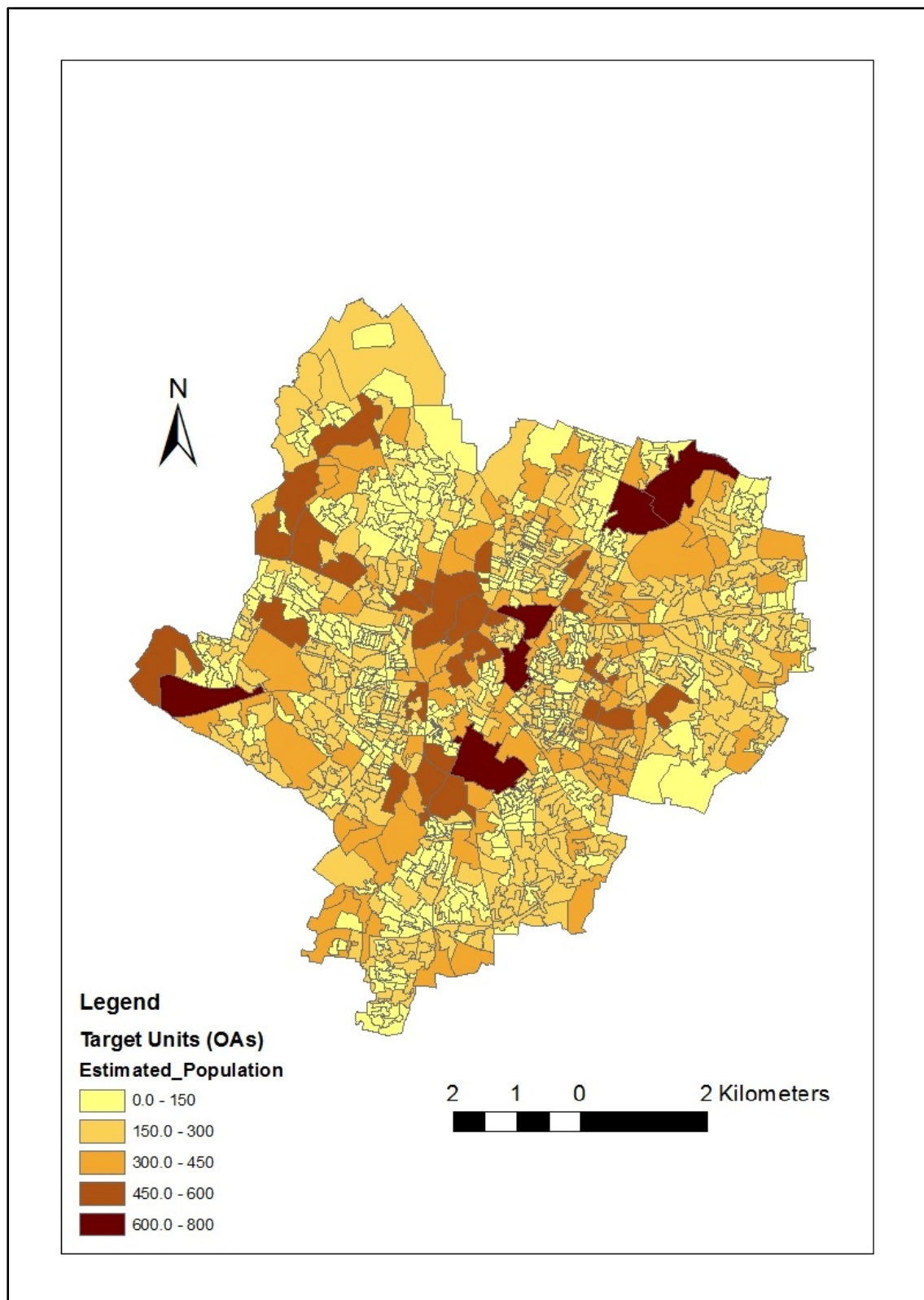


Figure 4.4: Population estimation results of the binary dasymetric mapping model in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

Figure 4.4 shows high population values located in output areas where commercial and industrial areas cover considerably bigger areas and it is subject to over estimation. The main reason is that residential and non-residential uses in urban areas are not differentiated. The dasymetric map shows that population allocation is proportional to the size of built-up areas of target zones. This is because the binary mapping technique allocates population homogenously within the built-up area of target areas. As a result, more people are located to the large-sized built-up areas (dark brown-coloured zones) and less population totals distributed the small-sized built-up areas (yellow-coloured zones). It can be seen from aerial photography of the study area (see the aerial photography of Leicester City in Figure 3.8, page 68). The mostly populated areas overlap with predominantly residential uses and the zero populated and less populated areas coincide with non-residential uses. The binary interpolation resulted in heterogeneous population distribution within the study area. Similar results were obtained by Langford and Unwin (1994). The less populated areas are close to the boundary of the administrative area (see Figure 4.3. and Figure 4.4). The parks, agricultural areas, golf courses and water surfaces are mostly zero-populated areas. This makes the classified images relatively accurate to use as control variables. In addition to this, population totals are mostly cumulated in the central area of Leicester City, in particular, the eastern part of the city is more populated than the northern and southern parts (see Figure 4.3). This observation shows that dasymetric model performance depends on the quality and precise classification of ancillary input data as residential and non-residential areas. This is because the binary model assigns the population totals to only built-up areas. If the classification of external imagery data were not relatively accurate, population totals may be located in uninhabited zones, which affect the accuracy of the predicted population counts.

The regression model was created to evaluate the relationship between estimated population by the binary model and census released actual population counts (see Figure 4.5). The predicted values were plotted against the known population totals. The pattern of the points were scattered close to the regression line and there are a few numbers of outliers. Thus, the scatterplot suggests that the relationship is much stronger than the results of the area-weighted interpolation. This test demonstrates that the performance of the binary dasymetric mapping method mostly depends on the accuracy of the external imagery data. It is notable from the experiments that the built-up areas

mostly covered by non-residential building structures are subject to overestimation and the misclassified residential uses are subject to underestimation.

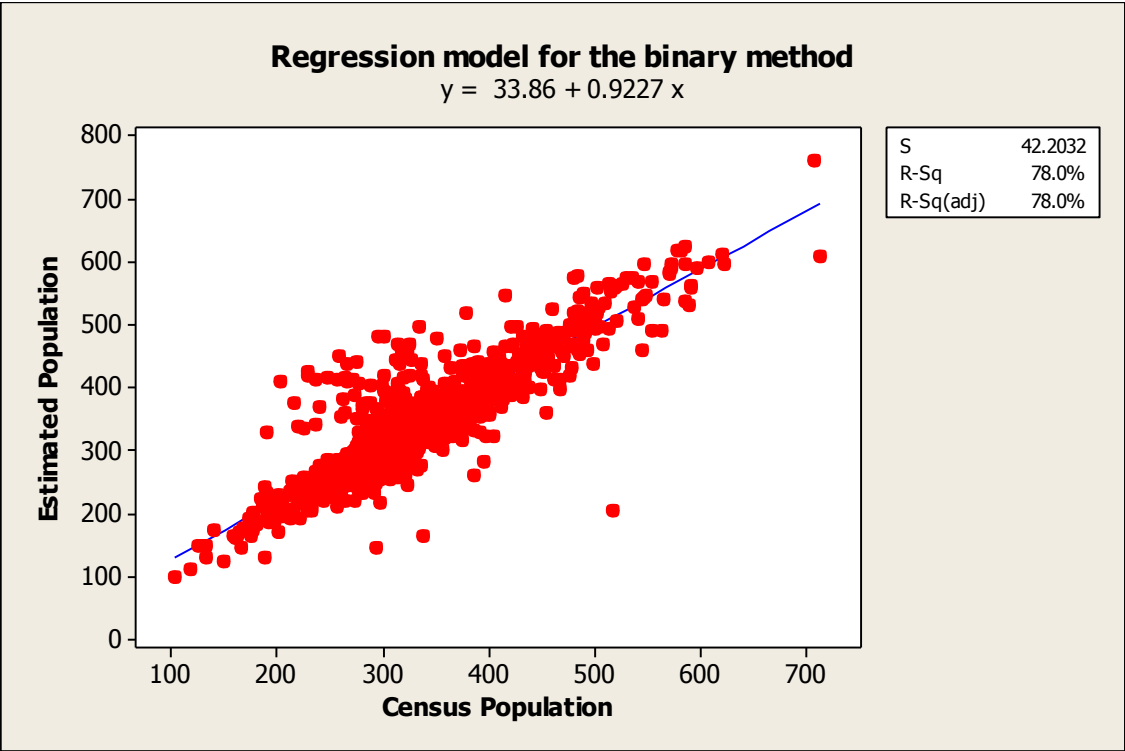


Figure 4.5: Regression model for the binary dasymetric mapping interpolation method. The estimated population totals of the City of Leicester at output area level versus the actual population of output areas released by census.

The estimated population totals of each output area were compared with known census population information in order to assess the accuracy of interpolation model and to calculate the degree of error introduced. The accuracy is assessed by using the same measurements used in the estimation results of the area-weighted interpolation. The binary model has the correlation R^2 is 0.78, which suggests that the correlation is substantially strong between the estimated population and the census released population totals. Table 4.2 demonstrates the regression coefficients, R squared and RMSEs for each applied disaggregation models. It can be seen that there are significant differences between the result of areal weighting technique and the binary dasymetric technique.

Table 4.2: The accuracy measurements for the binary dasymetric mapping method obtained population estimates for the City of Leicester.

Accuracy Measurements	Accuracy Results
Regression Coefficients	0.922
R Squared	0.780
Standard Error	42.203
RMSE	0.642
Relative Error (RE)	0.0825
Percentage Error (PE)	8.25%

In summary, the overestimated population totals were substantially concentrated in the industrial and commercial areas of Leicester City. These regions are mostly classified as built-up areas and considered as residential uses. These uninhabited built-up areas are primary error sources in the binary dasymetric process. On the other hand, the underestimated population counts were cumulated in areas where the population density is higher and the sizes of output areas are smaller than the other target zones; this can be clearly noted in Figure 4.4. These findings suggest that the binary dasymetric mapping technique can be utilised with the availability of ancillary imagery data in order to generate relatively accurate population estimates.

4.2.3 Volumetric estimation method

The volumetric dasymetric mapping method was the third technique employed to disaggregate population counts from census wards to output areas in the same manner previously used in interpolation model. In this approach, building volume data were used to identify residential buildings in order to estimate the population of target units for Leicester City. The volume of each building was used as the main external dataset to distribute population totals. The total building volumes of output areas were calculated by summing the volumes of high-rise and low-rise buildings which fall to each output area. The total volumes of output areas are mostly higher in areas where high-rise buildings are dominant; in contrary, the total volumes of output areas are smaller in areas where low-rise buildings are dominant. Population totals were then spatially disaggregated based on the total building volume of each output area. LiDAR-derived DEMs cover 942 out of 969 output areas in the study areas. Population totals were

estimated for output areas which consist of LiDAR-derived building height information. Figure 4.6 shows the map of building volumes by output areas that are used as control variables to estimate population totals. The volume based estimation method assumes that population totals reside within the residential housing units. The extraction of residential buildings from non-residential structures is substantially important in the population estimation process because the accuracy of estimation results depends on the building usage, such as residential, commercial, industrial or institutional (Sridharan and Qiu, 2013, Lwin and Murayama, 2010). Residential building volumes were used as ancillary input data to disaggregate census population totals to small areal units within the study region.

The 2011 census counts were disaggregated from census wards to output areas, employing building volume based interpolation process. The volume of each source zone was provided by intersecting residential buildings which fall in each source zones. Additionally the population of individual housing units was determined by dividing the total population with the volume of each housing unit. The population of housing units within each intersection area was summed to obtain the population of output areas. The volume based population estimates were mapped at output level. Figure 4.7 displays a map of predicted population values for the City of Leicester. In the legend of the estimation map, the same range of colours was used in the five classes as with previous estimation maps. The results in the figures show the estimated population counts for all output areas. High-populated output areas were shown by a dark brown colour and the low-populated output areas were shown by a yellow colour. The volumetric estimation model considers the presence of both low and high-rise buildings within each target unit. The map of volumetric model shows that population totals are distributed heterogeneously in the output areas of each census ward. Population counts are concentrated in central part of the city. It is obvious from the estimation that the northern and the southern part of the city is the less populated (see Figure 4.7). The performance of the volumetric model depends on the determination of building volume from the external building height and footprint datasets. This is because the volumetric estimation model locates the population counts of residential building units based on their volumes. Building volume information is important to define residential information in the study area. If the residential building units were not differentiated

accurately, population counts can be assigned non-residential structures. This shows the accuracy of the obtained population by volumetric method.

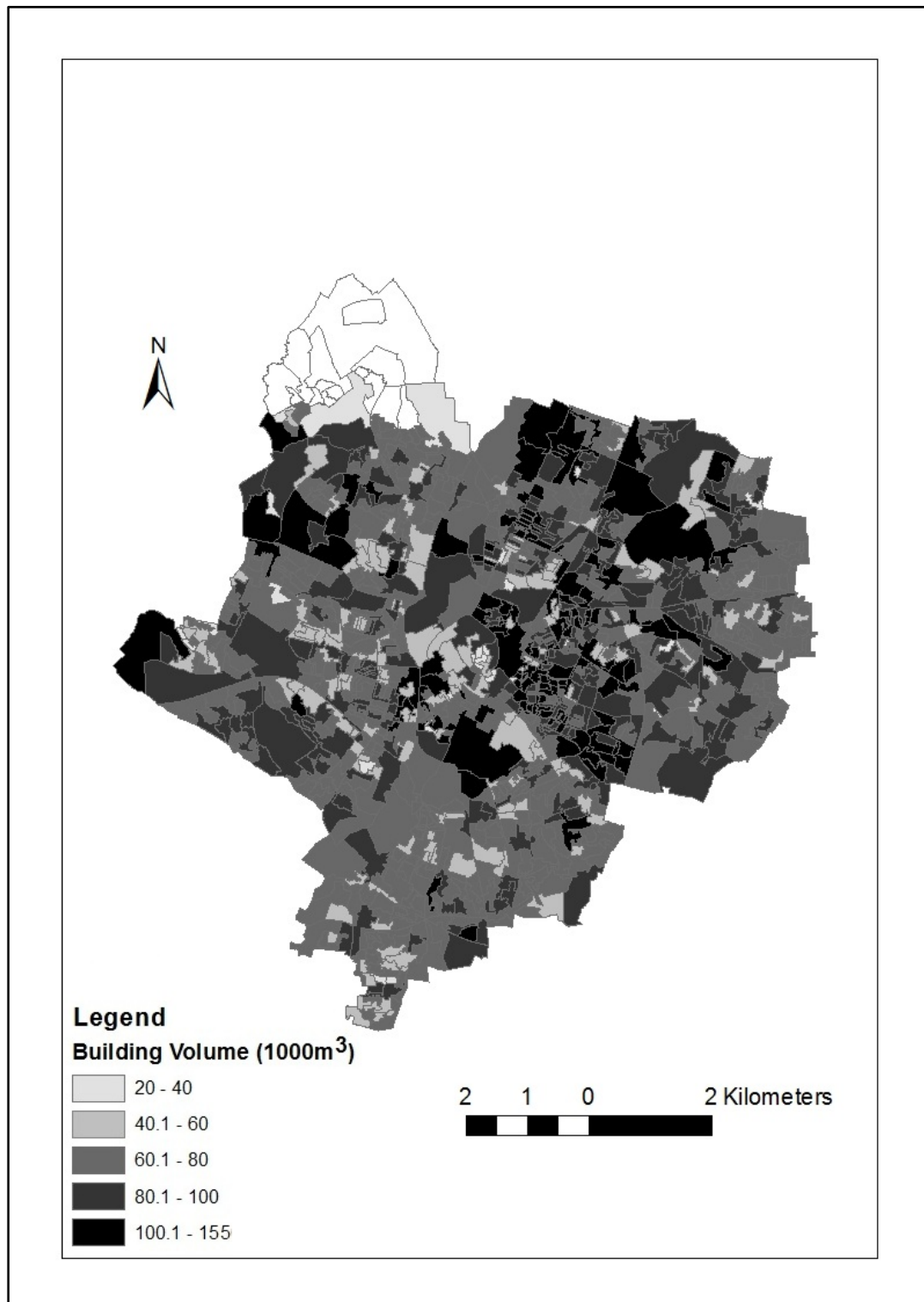


Figure 4.6: Map of building volumes by output areas for the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service and All LiDAR data © Airbus Defence and Space (Astrium Ltd. 2014).

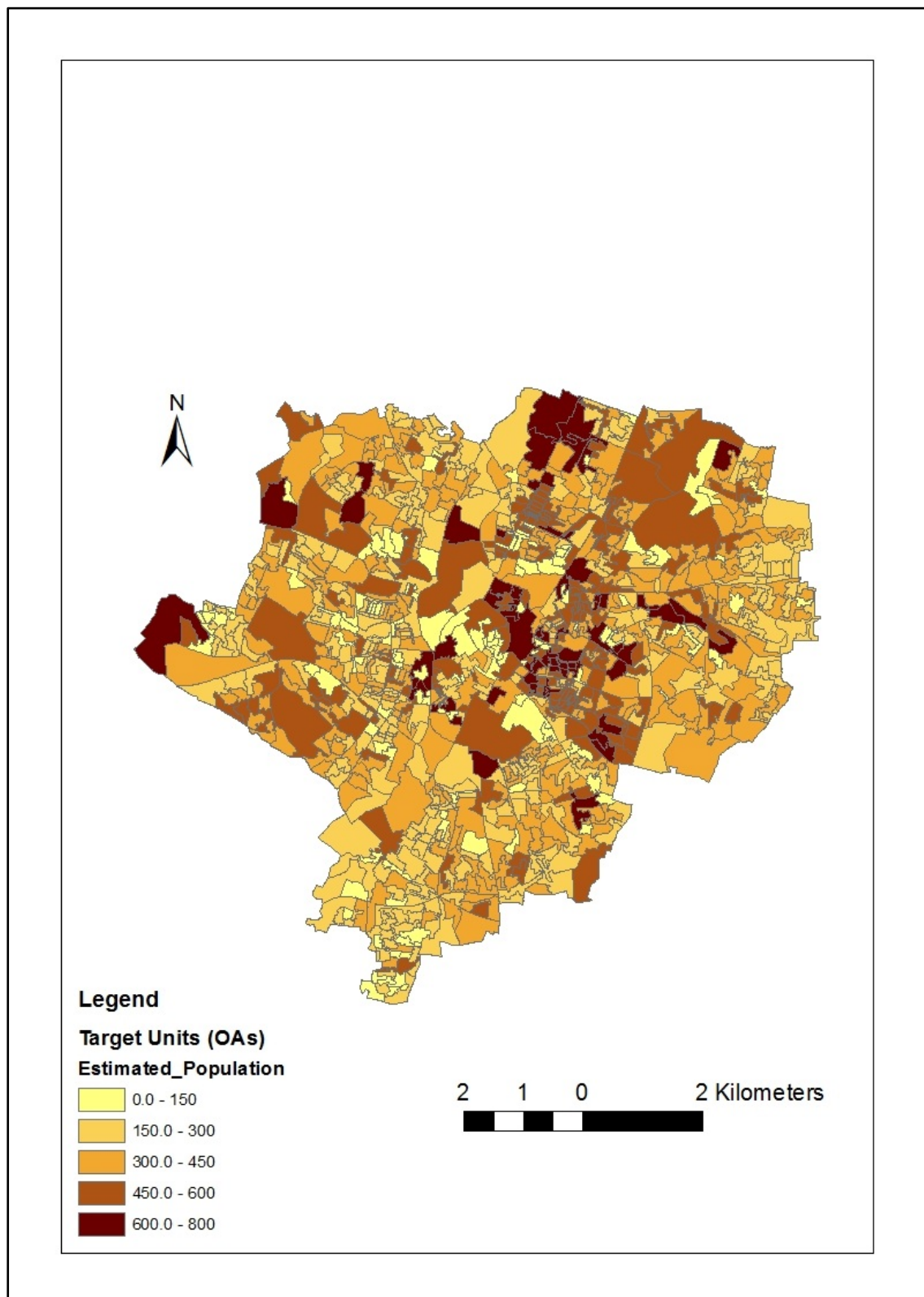


Figure 4.7: Population estimation results of building volume based dasymetric mapping method in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

As done in the previous methodologies, a regression model was generated to detect the relationship between the volumetric model obtained population and the census reported actual population values (see Figure 4.8). The estimated population totals versus the actual population counts were plotted on a scatterplot. The point patterns were scattered around the regression line, suggesting that the correlation between the estimated and the actual population values is relatively good with an R^2 of 0.73 in the volumetric model. There are, however, a number of outliers in the scatterplot. This experiment illustrates that the performance of the volumetric estimation model largely depends on the determined building volumes from ancillary inputs. It is obvious that the high-volume buildings are subject to overestimation in output areas.

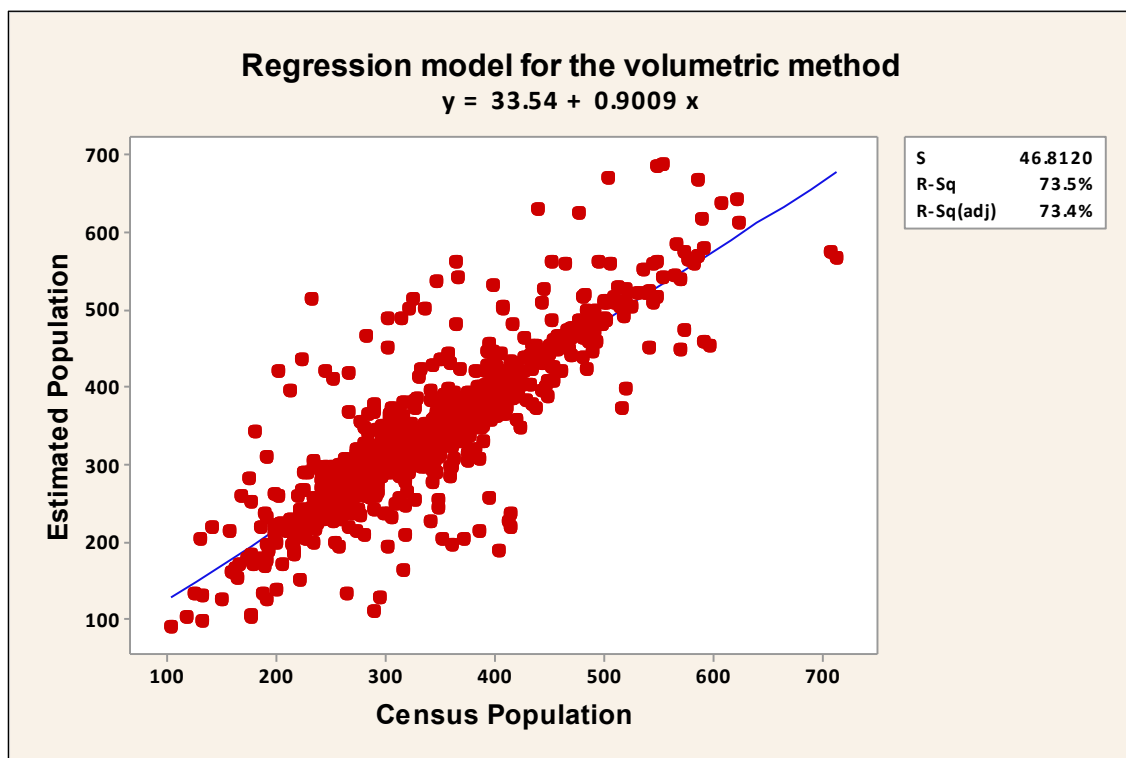


Figure 4.8: Regression model for the building volume based dasymetric mapping interpolation method. The estimated population totals of the City of Leicester at output area level versus the actual population of output areas released by census.

Table 4.3: The accuracy measurements for the volumetric estimation method obtained population estimates for the City of Leicester.

Accuracy Measurements	Accuracy Results
Regression Coefficients	0.901
R Squared	0.734
Standard Error	46.812
RMSE	0.773
Relative Error (RE)	0.0719
Percentage Error (PE)	07.19%

The predicted population of target units were compared with census true population values to measure the accuracy of the volumetric estimation model. The accuracy is calculated by utilising the same tests used in the previous interpolation models. This volume based dasymetric interpolation has an R^2 of 0.73 which suggests the correlation is reasonably high between the estimated population totals and the census released population counts. Table 4.3 demonstrates the regression coefficients, R squared and RMSEs for each proposed interpolation model. It can be seen that there are significant differences between the result of each applied interpolation model based on their assumption and the external input data. The implementation of the volumetric estimation approach using building volume information helped to improve accurate population estimate in target areas.

4.2.4 Address-weighted dasymetric method

The address-weighted dasymetric process was the fourth interpolation technique applied to produce population estimates at small spatial units. This approach assumes that people reside in residential building units within the residential uses. Thus, the model predicts population totals by multiplying the number of residential housing units with average household size in target areas. In order to obtain the number of residential units, OS Address Layer 2 datasets were used in the dasymetric population disaggregation process in a similar manner to Zandbergen (2011) and Tapp (2010). The employed address-weighted model used census wards as source zones, the output areas as target zones, and the numbers of residential housing units as control variables to estimate the population of each output area in Leicester City. The residential building data are much

more detailed than residential land use and road networks to identify housing units where people actually reside (Sridharan and Qiu, 2013, Zandbergen, 2011).

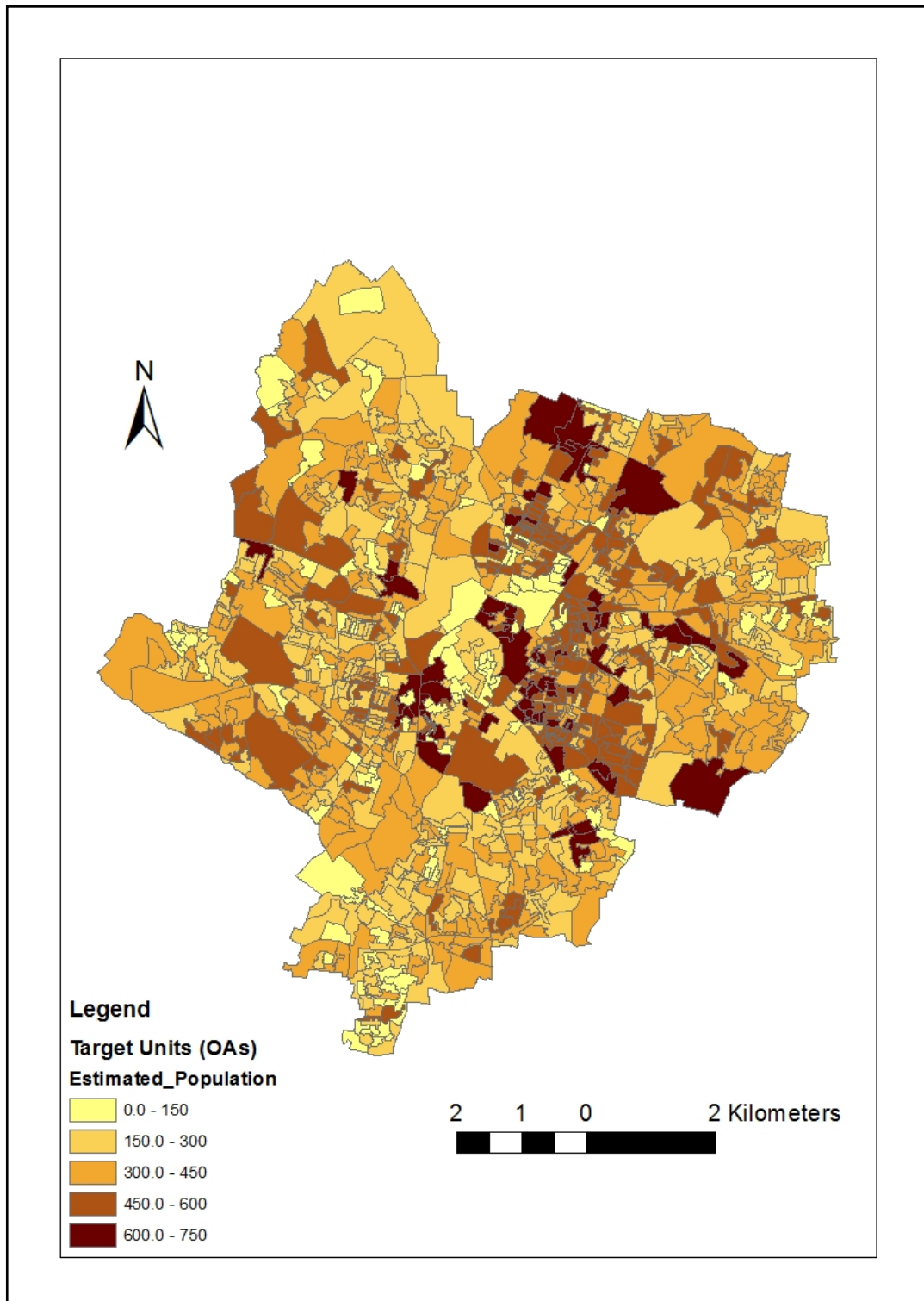


Figure 4.9: Population estimation results of address-weighted model in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

The 2011 census population totals were assigned to output areas based on the number of residential housing units and the population per household. The estimated population totals of each housing unit was aggregated to individual output areas. Figure 4.9 shows the obtained population totals by the address-weighted dasymetric mapping model. On the map, the estimated population totals were grouped into five classes. The legend on the estimation map shows the range of predicted population totals, with colours changing from yellow to dark brown as the population totals increase. In the estimation map, output areas with low populations are labelled in a yellow colour and output areas with high population are in a dark brown colour. The estimation map shows that more people are located in the output areas which consist of more residential housing units. Also, less population was obtained in areas where less housing units exist in output areas.

The estimated population of building units were aggregated to the output areas and shown in Figure 4.9. The number of people may be allocated incorrectly due to the assumption that all building units are residential within the geographical units. Because of this reason, this model overestimated the population totals in output areas where there are non-occupied housing units. The address-weighted method shows substantial improvement over the areal interpolation and the volume based dasymetric mapping method. The results from this interpolation may be improved by utilising building occupancy rate datasets, which provides the information of occupied housing units within the source areas.

Similar to the previously used methods, the regression model was produced for the address-weighted estimation results to test the relationship between the estimated and the known population values (see Figure 4.10). The predicted population values were plotted against the census reported values. Most of the point patterns were scattered along the regression line. This plotting result suggests that there is a very strong relationship between those estimated and true population values. There are a few numbers of outliers in the scatterplot. This measurement demonstrates that the performance of the address-weighted dasymetric model directly depends on differentiating the number of residential housing units from the ancillary input of Address Layer 2. The non-occupied housing units within the target units are primary error sources and they are subject to overestimating in study regions.

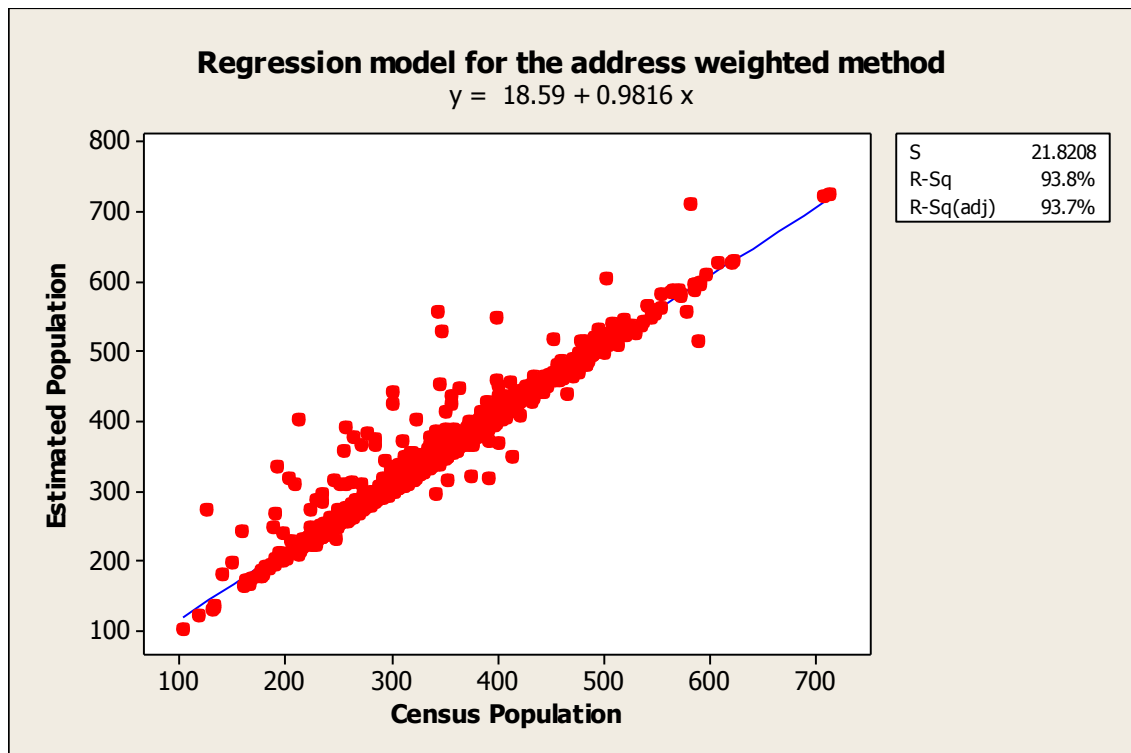


Figure 4.10: Regression model for the address-weighted dasymetric mapping interpolation method. The estimated population totals of the City of Leicester at output area level versus the actual population of output areas released by census.

Table 4.4: Accuracy measurements for the address-weighted dasymetric method obtained population estimates for the City of Leicester.

Accuracy Measurements	Accuracy Results
Regression Coefficients	0.981
R Squared	0.937
Standard Error	21.820
RMSE	0.160
Relative Error (RE)	0.0407
Percentage Error (PE)	4.07%

The estimated population counts were compared with actual census population totals in order to assess the accuracy of the dasymetric method and to calculate the degree of error introduced. The accuracy of estimated results was measured using the same tests used in the areal weighting and the binary dasymetric mapping approach. Table 4.4 demonstrates the regression coefficients, R squared and RMSEs for each applied

dasymetric mapping models. The modelling strategies performed very well with R^2 of 0.93. It can be seen that there are significant differences between the results of the address-weighted method and other interpolation techniques. This result suggests that the relationship between estimated and true values is very strong. This affirms that building point datasets can be used as the control variables to disaggregate population totals from source zones to target areas.

4.2.5 Occupancy-adjusted address-weighted dasymetric method

In this section, the address-weighted model was used to generate the estimation of population totals at output area level, utilising the external datasets of building address points, building occupancy data and the average household size of target zones. In the address layer, all building units are labelled (coded) as residential or non-residential housing units. These high resolution building attribute information are superior to define residential housing units in source and target areas, respectively. The second external data are building occupancy information that provides the number of occupied housing units within each output area. In this model, building occupancy rate information was first integrated to building address point data to determine the number of occupied housing units within each target zone. Census wards were used as source zones and output areas were used as target zones, the same as in earlier approaches. However, the combination of address points and occupancy information was employed as control variables to predict population totals of each output areas. This dasymetric model was performed in the City of Leicester, similar to the procedure explained in Section 4.2.4. The address of individual residential housing units provides more precise information of population distribution than generalised (generated) land cover data and other datasets. The estimated population totals are fully based on the total number of occupied residential housing units and average household sizes in each target unit. As a result, the address-weighted dasymetric mapping model, using building address points and building occupancy information, obtains the highest dasymetric interpolation accuracy over the other tested methods.

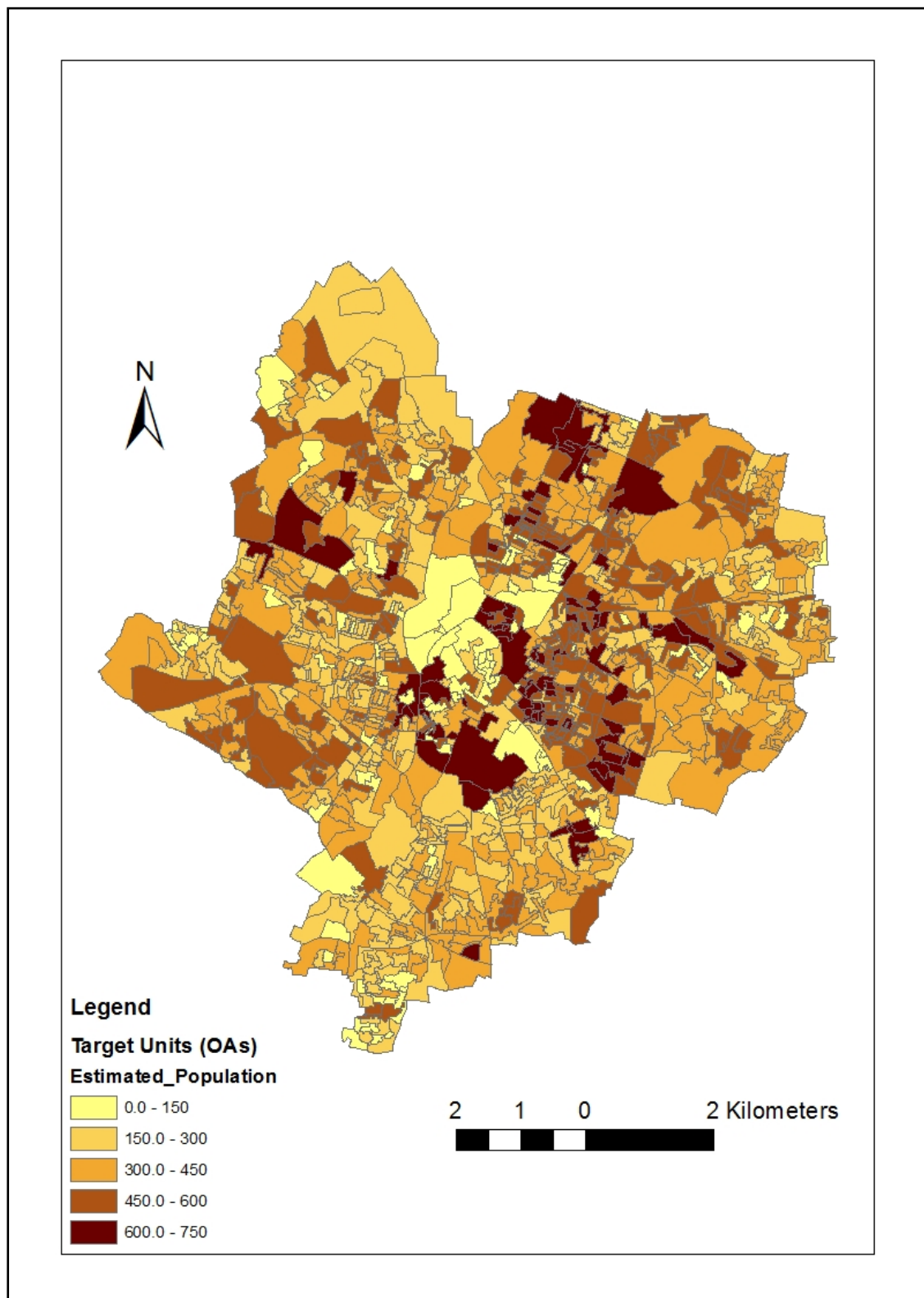


Figure 4.11: Population estimation results of refined address-weighted dasymetric mapping method in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

The area weighting and the binary method need the most current census population to create a population distribution map, whereas the address-weighted dasymetric mapping method does not use the census population. The population estimation maps reveals that this dasymetric estimation model generates the closest population totals to the known accurate population figures, due to extraction of empty housing units from building address datasets. Figure 4.11 shows the distribution of estimated population based on the occupancy data and building address points in the process of the address-weighted model.

A regression model was generated for the residential building address points and occupancy information based population estimates in order to evaluate the correlation between predicted and the census reported population totals (see Figure 4.12). The estimated population totals versus the actual population counts were plotted. The point patterns were scattered along with the regression line. Most of the points were scattered from lower left to upper right suggesting a very strong correlation between predicted and true population values. There are no significant outliers in the scatterplot. This measurement demonstrates that the performance of the address-weighted dasymetric model generally depends on the accurate number of residential housing units from the ancillary input of Address Layer 2 and the realistic occupancy information. The inclusion of occupancy information has addressed overestimation in target areas where there is a significant number of non-occupied housing units. This can be clearly seen from comparing the scatterplots of the address-weighted (Section 4.2.4) and the refined address-weighted (Section 4.2.5) dasymetric mapping models.

The predicted population of target areas was compared with the known population in order to measure the accuracy of the dasymetric method and to calculate the degree of error introduced. The accuracy was assessed by calculating the same measurements utilised in the other applied areal interpolation techniques. Table 4.5 shows the measured accuracy metrics. The best results were obtained by using the address-weighting dasymetric approach with the ancillary variable of building occupancy information. A correlation coefficient of 0.99 was obtained when all 969 output areas in the study area were used in the analysis. When occupancy information was added to the address-weighted model, the dasymetric method generated the highest R^2 of 0.998. This estimation result suggests that the relationship between predicted population counts and

actual population values is very high. It confirms that control variables of residential address points and occupancy information can be use in the dasymetric mapping method to estimate population totals.

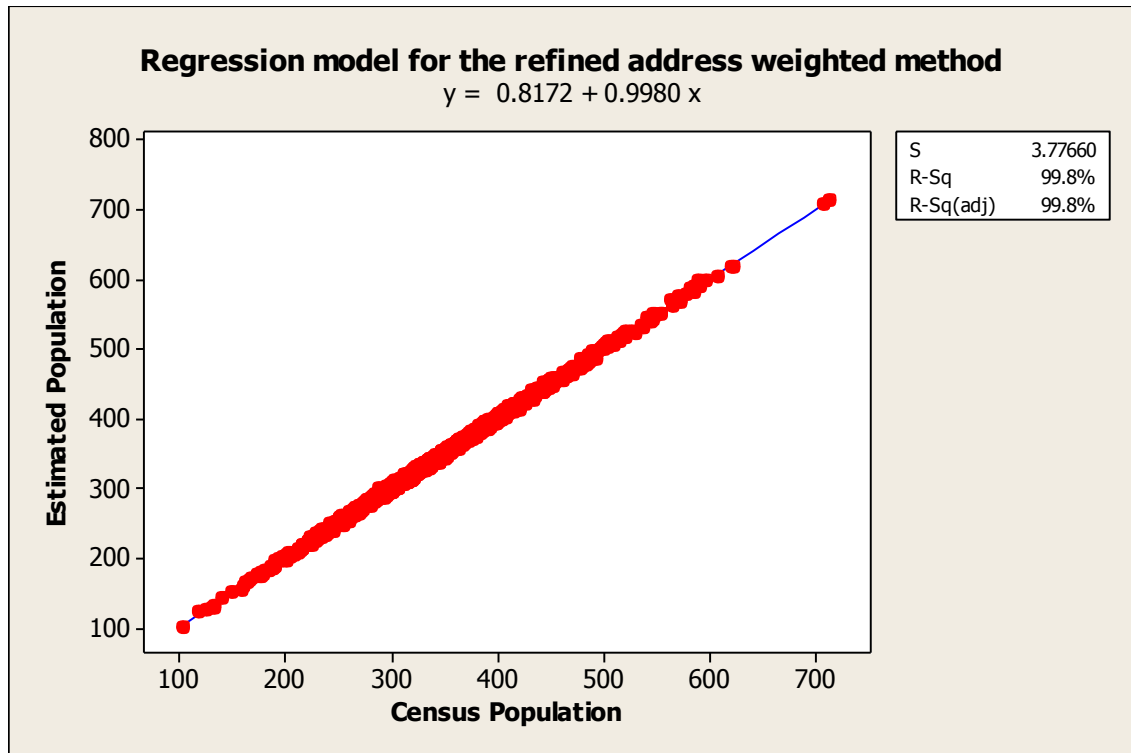


Figure 4.12: Regression model for the refined address-weighted dasymetric mapping method. The estimated population totals Leicester City at output areas level were plotted against the census released population of output areas.

Table 4.5: Accuracy measurements for the improved address-weighted dasymetric method obtained population estimates for the City of Leicester.

Accuracy Measurements	Accuracy Results
Regression Coefficients	0.999
R Squared	0.998
Standard Error	3.778
RMSE	0.064
Relative Error (RE)	0.0108
Percentage Error (PE)	1.08%

Overall, there are less overestimated and underestimated population values in target zones compared to earlier interpolation results. This is because people are mostly placed

correctly to where population resides. So, the performance of the address-weighted method is better than the previously evaluated interpolation model. The results suggest that the availability of high resolution building attribute information and the occupancy information help to generate the most accurate population estimates in the intended scale. The dasymetric model obtained the most accurate estimation results and was substantially improved by using occupancy information.

4.3 Comparison of Areal Interpolation Approaches

In this section, the results of population disaggregation approaches were compared in order to evaluate the performance of interpolation models. First of all, the differences between the areal weighting and the binary dasymetric model were presented. Secondly, the volumetric estimation and the address-weighted dasymetric process were compared for measuring the accuracy of models. Lastly, the comparison of the volumetric estimation and the address-weighted dasymetric method were presented. Comparison results show how well the proposed interpolation approaches and the selected ancillary input datasets perform in order to distribute population totals in the study region. These comparison results are detailed in the following sub-sections.

4.3.1 Comparison of the areal weighting and the binary dasymetric method

These two areal information processes were used in order to distribute population totals through the study region. These methods generated notably different estimation results due to their functioning differences. The areal weighting method places people everywhere and the binary mapping method places people to predefined residential zones. The results of these models were compared in order to evaluate the performance of these approaches in this section. The residual maps were generated by subtracting the interpolation obtained predicted population from the census reported actual population totals. Figure 4.13 shows the residual map of the areal weighting based population estimates, and Figure 4.14 illustrates the residual map of the binary mapping model at output areal level. These residual maps show to what extent the people were placed inaccurately. The legend demonstrates the range of population residual values in five classes with the colour changing from blue to red as the population residuals increase. The yellow colour presents the output areas where estimation residuals are between -25 and +25. The results of mapping the differences between these methods in the case of

Leicester City shows that population totals disaggregated according to model assumption.

The main difference between these methods is that the area-weighted method does not require any external data and the binary method requires a land cover map to separate residential uses from non-residential areas. The usage of additional datasets in the binary dasymetric mapping process provides more realistic built-up areas to disaggregate population totals through residential uses. As a result of this, there were considerable differences between the results, particularly in the northern and western parts of the city. The differences were, however, less in the central regions and eastern parts of Leicester (see Figure 4.13 and Figure 4.14). The large area sized target units are subject to positive estimation error and the small area sized target zones are subject to the negative estimation error in the area-weighted results. Misclassified land cover types are also subject to estimation error in the binary dasymetric mapping. Therefore, the binary model generates different population estimates for output areas.

In the area-weighted based estimation, bigger positive errors occurred in the large-sized output areas. The size of output areas may be larger due to the transition from urban to rural land uses (Mennis and Hultgren, 2006a). They are mostly located close to the boundary of the study region. However, bigger negative errors occurred in the small-sized output areas and they are sparsely distributed within the administrative area. These small census units are subject to underestimation (see Figure 4.13). In contrast to the areal weighting, more people were incorrectly placed on the non-residential built-up areas by using the binary mapping method (see Figure 3.8, in page 65). The estimation errors mostly arise from the misclassification of ancillary imagery data and the heterogeneous population distribution within the residential uses. This is because of the assumption of uniform distribution within residential areas. It can be clearly seen from the residual map of the binary method that residual errors arise in the city centre where non-residential buildings, such as shopping centres, are mostly dominant and in the northern, western and north-eastern parts of the study area where commercial centres and factories are dominant. The binary dasymetric map overestimated the large agricultural areas in the northern part of the Leicester. This is because areal images were taken in a season when there was no agricultural production, and so agricultural areas show similar reflectance to built-up areas. Therefore, these land uses were most probably misclassified as urban areas and subjected the classification to overestimation.

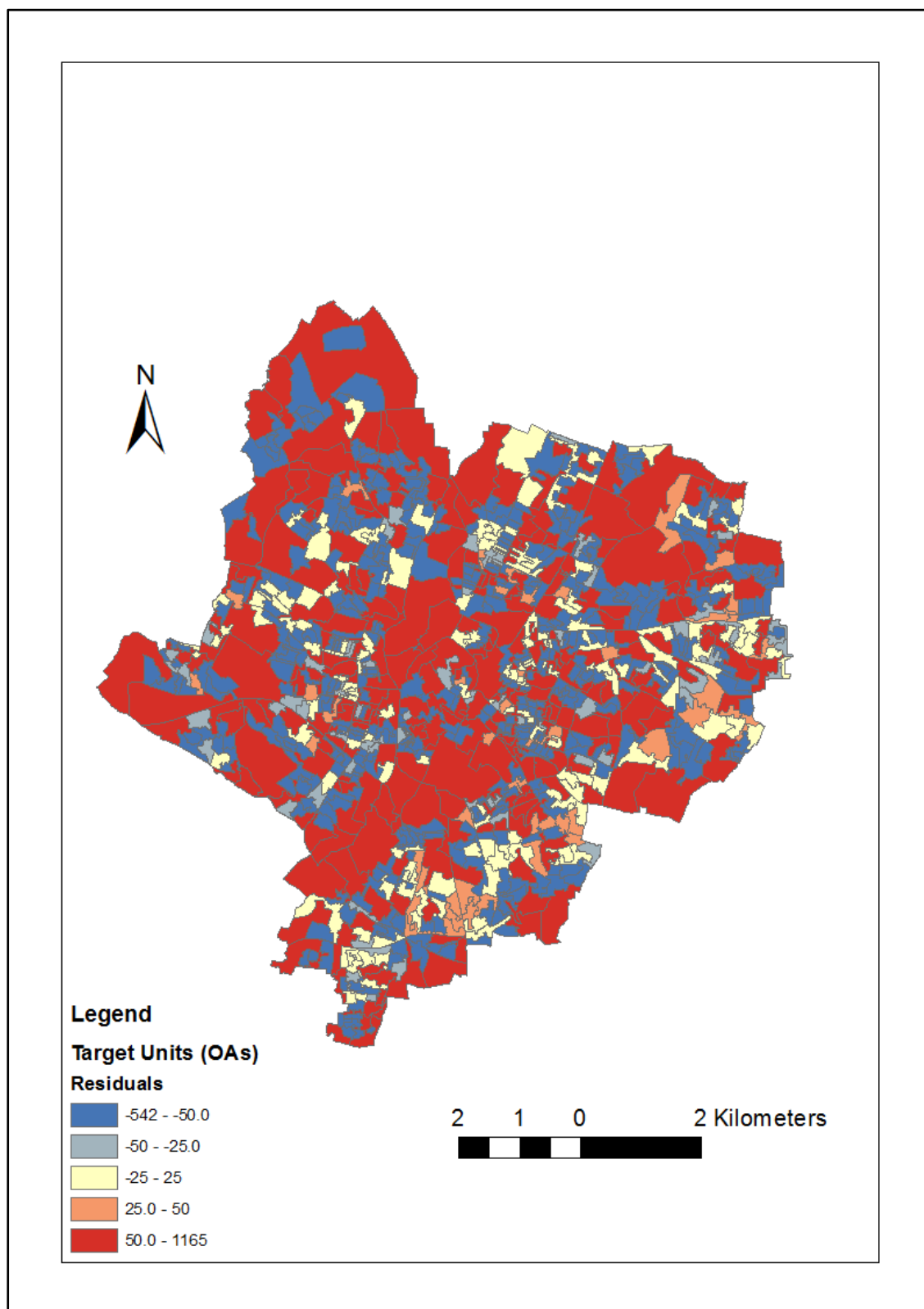


Figure 4.13: Residual map of areal weighting method obtained population estimates in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

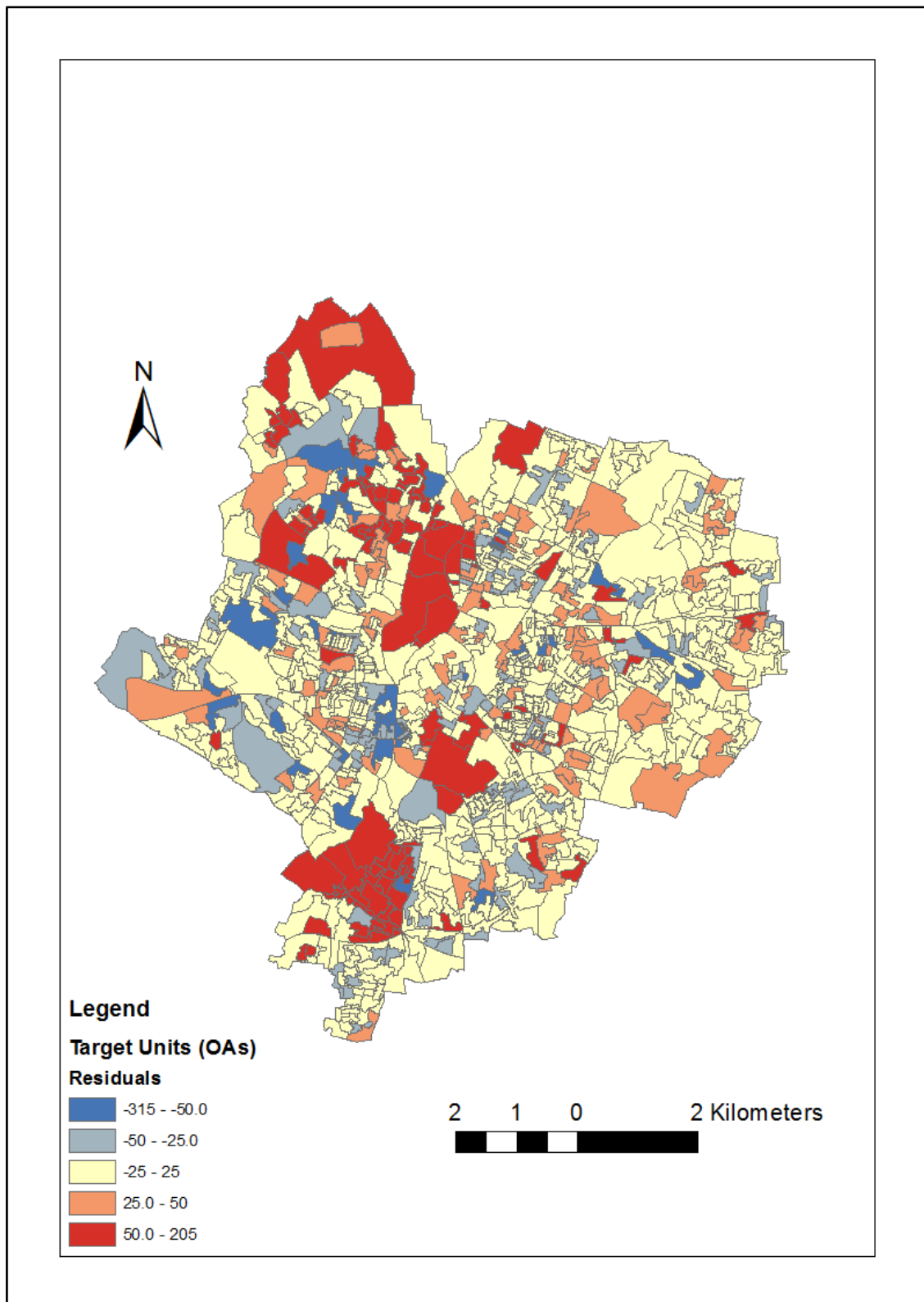


Figure 4.14: Residual map of the binary dasymetric mapping method obtained population estimates in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

In order to test the relationship between population estimation residuals and spatial units, the spatial autocorrelation function was applied. The analysis of spatial autocorrelation examines the similarity of the observed value of interested variables over space and test neighbouring effects (Goodchild, 1986). One of the common spatial autocorrelation indices is Moran's I (Anselin et al., 2010). The Moran's I index could be between -1 and +1: negative results show correlation and positive values suggest positive correlation between the interested objects and local neighbourhood. If the value is close to zero it shows there is no considerable spatial relation. The residuals of population totals were spatially autocorrelated by using Moran's I to observe whether residuals are clustered or dispersed randomly. Figure 4.15 and Figure 4.16 show the spatial autocorrelation results of the areal weighting and the binary dasymetric mapping methods, respectively. The tested interpolation methods yielded different spatial autocorrelation results. The residuals of the areal interpolation based population estimates has a Moran's I value of -0.005069 and the residuals of binary mapping method has a value of 0.176294. The results of the autocorrelation analysis display the residual of areal weighted method based estimations were randomly distributed over target zones in Leicester City. In contrast, the results of autocorrelation analysis indicated a clustered distribution of the binary mapping yielded population residuals. The results of spatial analysis may be affected by the accuracy of land cover classification.

To conclude, the binary dasymetric method with ancillary data of land use/cover has generated a better population distribution in areas where residential building structures are dominant and a less accurate estimate in areas where non-residential building blocks are mixed with residential blocks. The results of each estimation model indicate that the assumption of different interpolation models based on the size of source areas and utilised ancillary datasets played a substantial role. The areal weighting model produces the least accurate estimation results, with a lower coefficient and much larger RMSE values in the case of Leicester (see Table 4.1). In contrast, a binary model generates better results with the help of auxiliary datasets (see Table 4.2). The proposed binary dasymetric interpolation method performs better than the applied areal interpolation approach. This implies that results from binary dasymetric interpolation can improve when the land parcel data, which provide the information of residential and non-residential areas, are used.

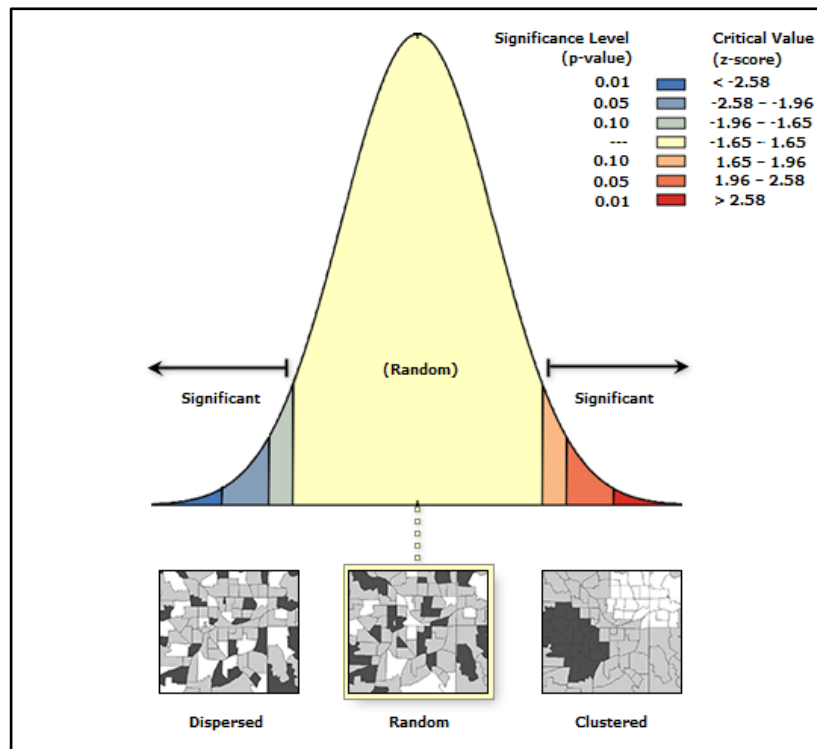


Figure 4.15: Spatial autocorrelation of the areal weighting method provided population residuals in Leicester City.

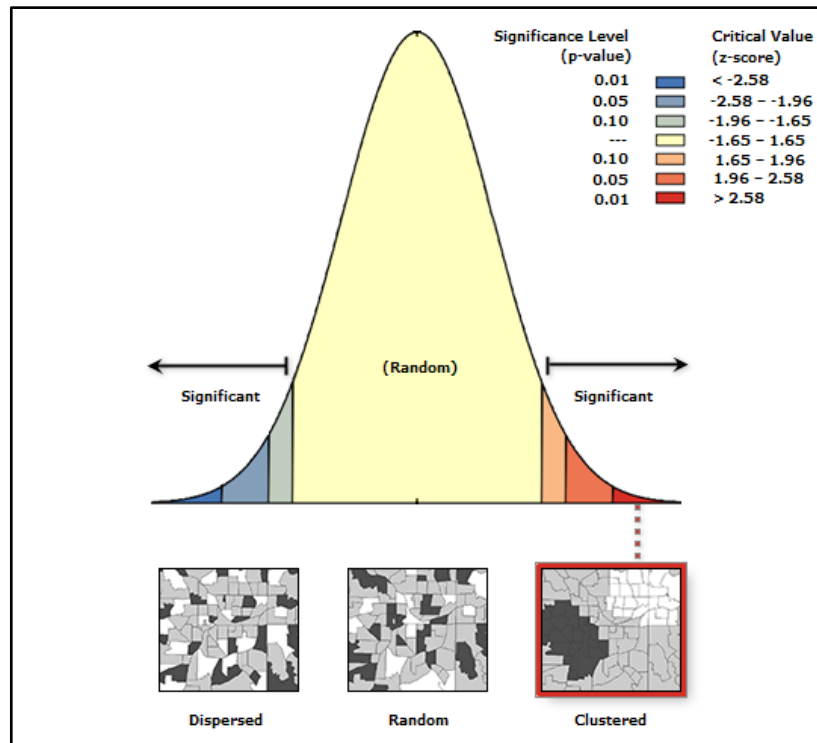


Figure 4.16: Spatial autocorrelation of the binary dasymetric mapping provided population residuals in Leicester City.

4.3.2 Comparison of volumetric estimation and the address-weighted methods

In this section, the results of building volume based population estimates were compared with the address-weighted dasymetric obtained population estimates. To achieve this, estimated population results were subtracted from the actual census values in the first stage. Residual results of these interpolation models indicate that there were notable differences between both models when estimating population totals in the City of Leicester (see Figure 4.17 and 4.18 and 4.19). The residuals were grouped into five classes to define the overestimated and underestimated output areas. The negative residuals were shown by a blue colour and the positive residuals were presented by a red colour in the legend of the maps. The yellow colour shows smaller differences between actual and estimated population values. These residuals were compared to test the performance of the volume and address based dasymetric mapping processes. Due to the assumption in population distribution, the methods generated remarkably different interpolation results. These differences can be seen from the residual maps. The legend of the residual map of the occupancy data based address weighted method was re-structured into three classes to show negative and positive residuals (see Figure 4.19).

The estimated population totals were compared with the true population totals in order to evaluate the performance of each model. Residual maps present that there are significant differences between the building volume and building address based population estimates. In most parts of the city, both address-weighted methods obtained population estimates that are the most close to actual values. This was due to the integration of residential buildings, average household size and occupancy data. Ancillary input data assigns people much more accurately to where they live in reality. The mixed-use building blocks have a considerable impact in both estimation results. In vertically stacked building blocks, each floor is coded based on its usage. Therefore, differentiating the flats which are used for residential purpose is paramount as it directly affects estimation results. The volumetric estimation method uses LiDAR-derived building height information and land parcel data to disaggregate population totals to target areas. However, building usage information does not determine mixed-use building structures, and this directly causes poor population estimation compared to the building address point based population estimates.

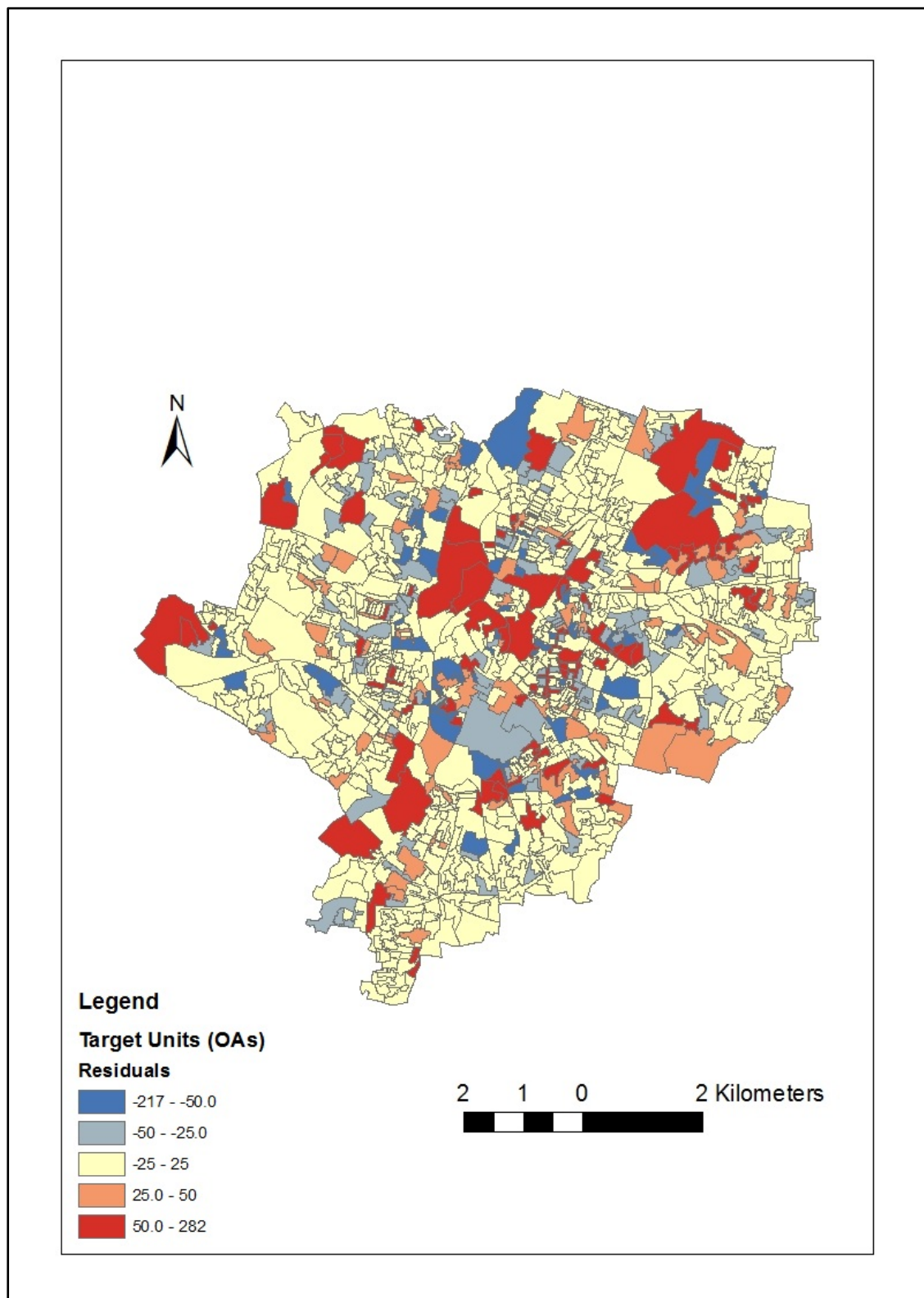


Figure 4.17: Residual map of volume based dasymetric mapping method obtained population estimates in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

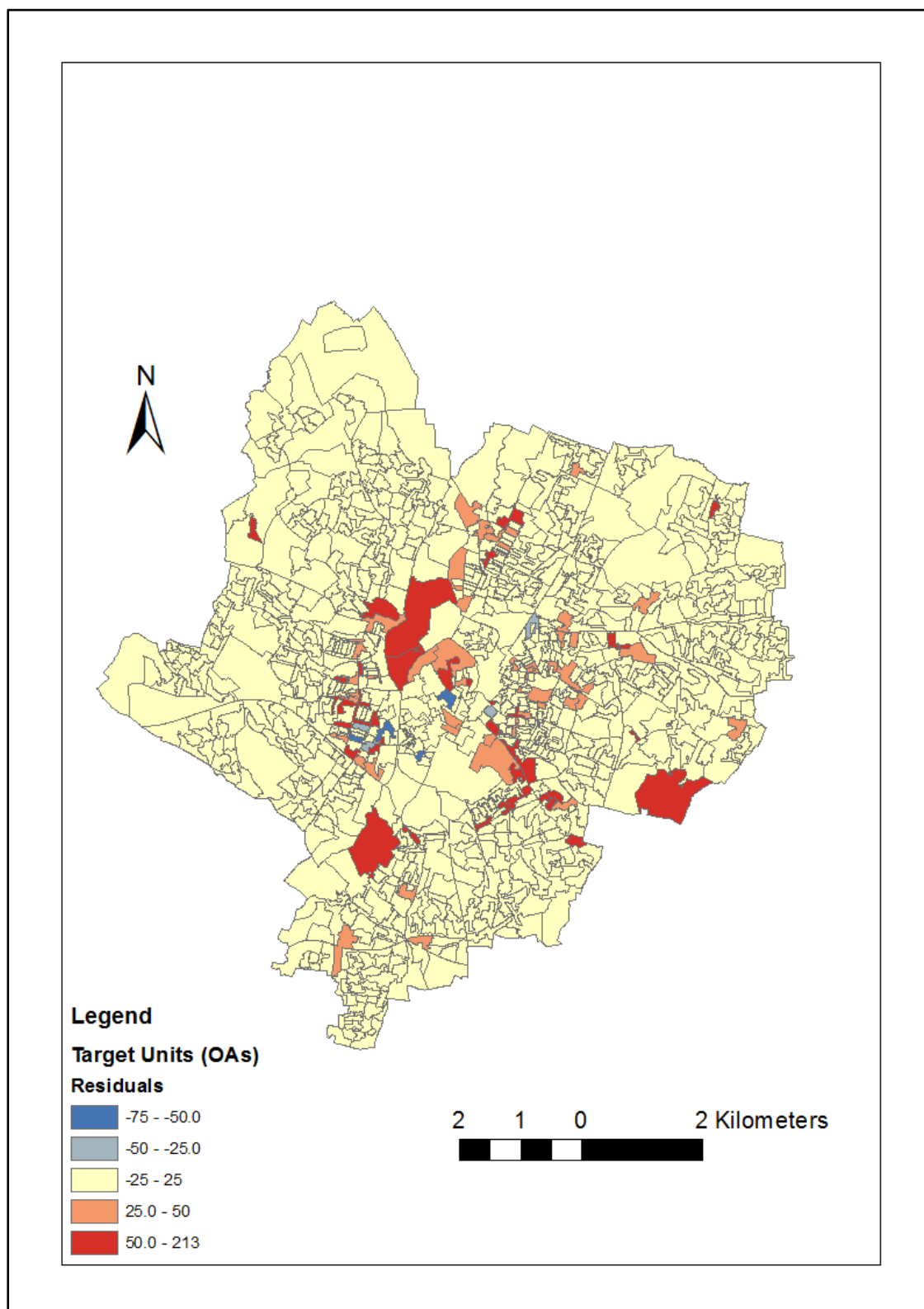


Figure 4.18: Residual map of the address-weighted dasymetric mapping method obtained population estimates in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

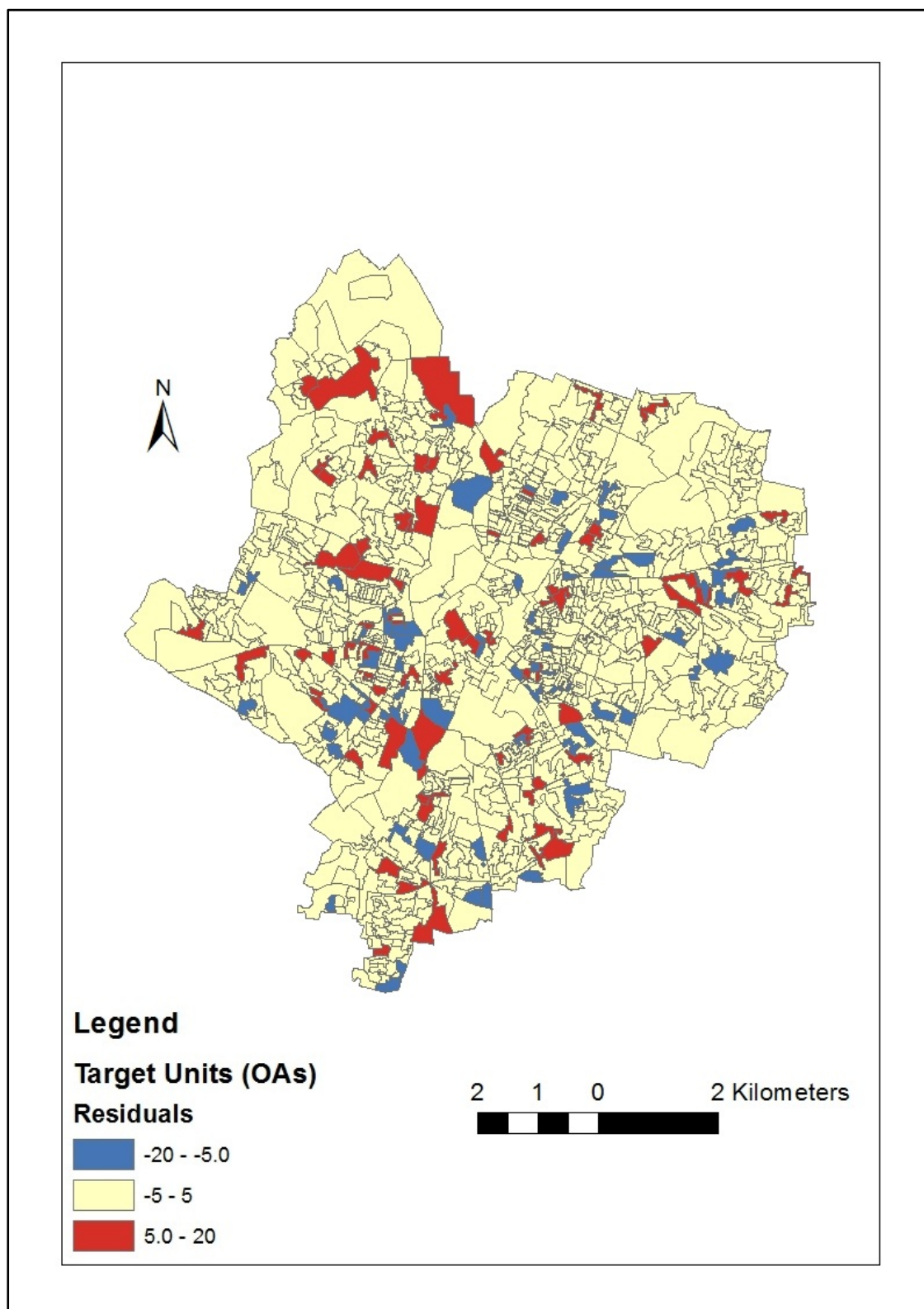


Figure 4.19: Residual map of the occupancy information used address-weighted dasymetric mapping method obtained population estimates in the City of Leicester. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

It can be clearly seen from the residual maps that the results of the existing model are mostly overestimated due to a missing count on unoccupied buildings within the target areas. Conversely, the building occupancy based estimation model does not count the non-occupied housing units. For this reason, the refined dasymetric model based population estimates become more close to the actual population of target areas. The other result is that there are substantial differences when mapping the estimation residual in Leicester City. This was due to the high probability of having more empty housing units in areas close to the city centre and central areas and less empty homes in the settlement of suburban areas. With the lack of occupancy information, the address based model counts all housing units within the source areas. Therefore, population estimates are likely to be overestimated in bigger urban areas. Finally, the estimation results show that the occupancy information is useful to generate small-area population totals. The address-weighted dasymetric mapping model using building address points and building occupancy information obtains the highest dasymetric interpolation accuracy over the other tested methods. These datasets may have a potential to define residential buildings and determine different housing heights and characteristics. The comparison of dasymetric areal interpolation techniques shows that building address point data serves better than building volume data as ancillary input data.

Spatial autocorrelation analysis of Moran's I was also applied for the results of volumetric estimation and the address-weighted models. While Figure 4.20 and 4.21 shows spatial autocorrelation of the results of volumetric estimation and the address-weighted dasymetric model, Figure 4.22 displays spatial autocorrelation of the results of occupancy data based interpolation method. The residuals of building volumes based population estimates has a Moran's I value of -0.003600, the building addresses based estimations have a value of 0.066480, the occupancy datasets based estimation has a value of -0.015153. On the one hand, the results of autocorrelation analysis display that the residual of the volumetric method and occupancy datasets based estimations were randomly distributed over target zones in the local authority of Leicester City. On the other hand, the results of autocorrelation analysis indicated a clustered distribution of the address-weighted method obtained population residuals. The application result indicates a random distribution of population residuals in the building volumes and building address points used to estimate population totals. Therefore, there is not any negative and positive correlation between the observed value and neighbouring

locations. However, the residuals of the address-weighted model are spatially clustered when the combination of building addresses and occupancy information are used as control variables to disaggregate population values.

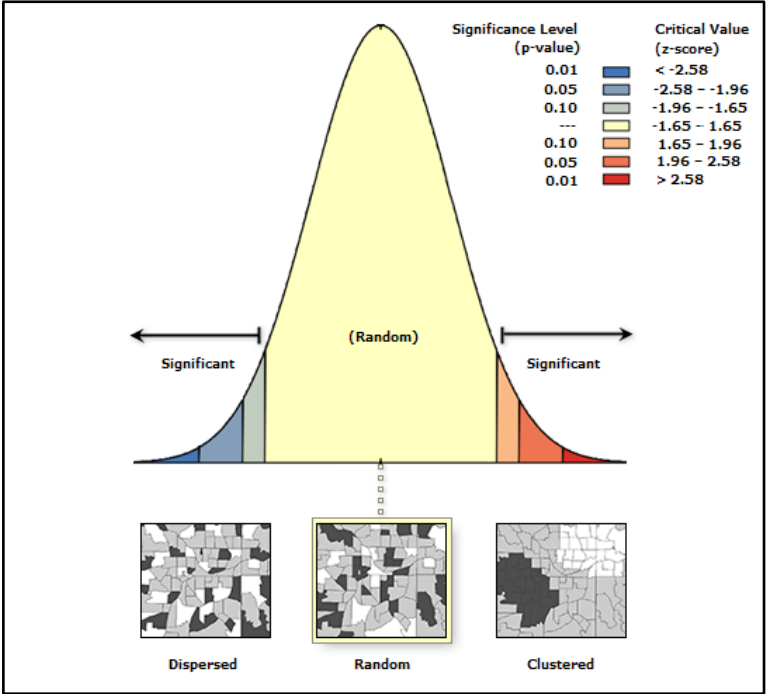


Figure 4.20: Spatial autocorrelation of the volumetric estimation model obtained population residuals in Leicester City.

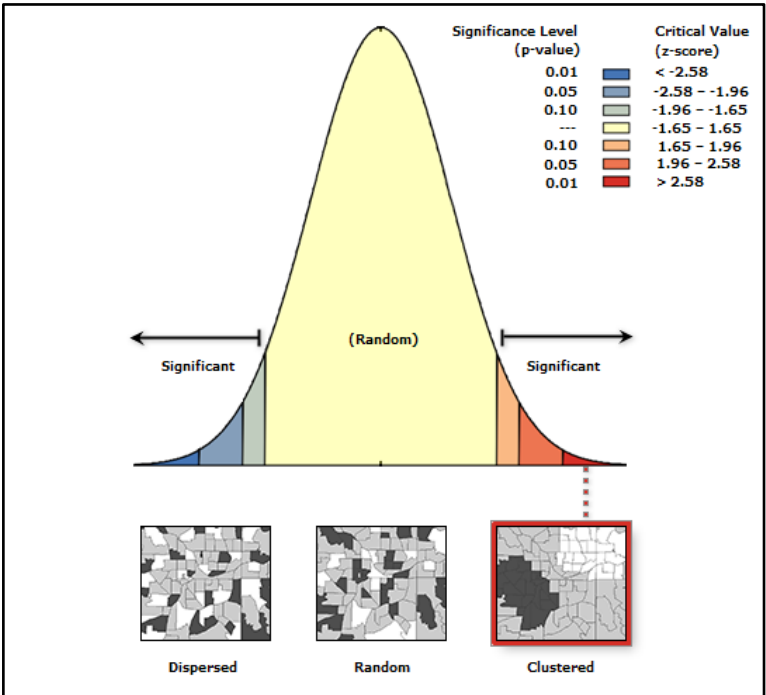


Figure 4.21: Spatial autocorrelation of the address-weighted dasymetric model provided population residuals in Leicester City.

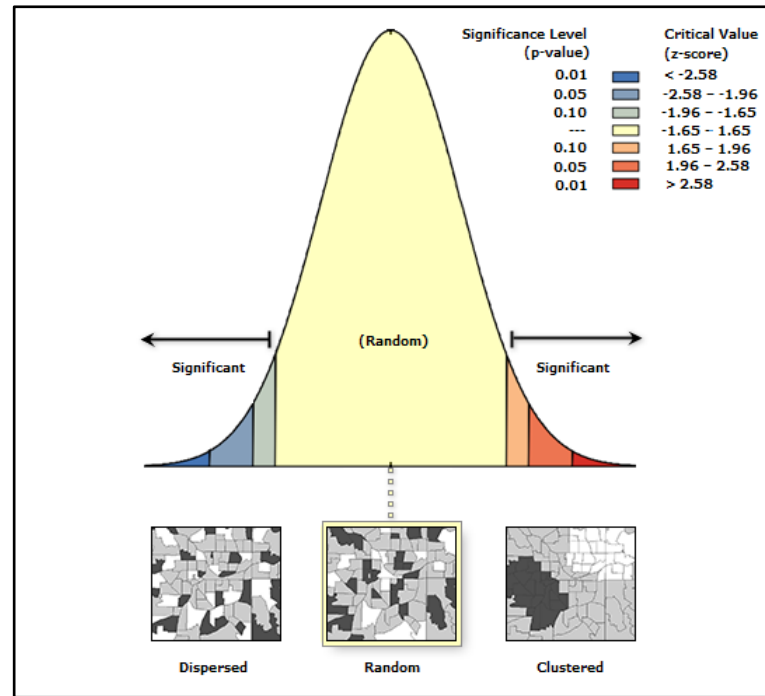


Figure 4.22: Spatial autocorrelation of the address-weighted method using occupancy data obtained population residuals in Leicester City.

In conclusion, the estimation results reveal that all interpolation models usually generate different population totals due to the different functioning of areal interpolation processes. The address-weighted dasymetric method yields better estimation results with higher coefficients and the smaller RMSE values than other used areal weighting, binary dasymetric and volume based interpolation approaches (see Table 4.5). In contrast, the volumetric estimation model was obtained with less precise results than the address-weighted dasymetric method in the City of Leicester. The volumetric estimation uses LiDAR-derived building and LandMap-derived building data to estimate population totals. Building datasets do allow differentiating building structures but not mixed-use building blocks and they directly cause poor estimation compared to building address points. The maps of estimation residuals indicate that the differences in estimation will be reduced using the realistic information of residential building units and occupancy rate information of source and target zones. The address-weighted approach provided reliable population estimates in different urban environments. Finally, realistic information of residential uses as external input data help to obtain precise population totals by using areal interpolation frameworks.

4.4 Summary

Five different areal interpolation approaches – areal weighting and the binary, the volumetric estimation, the address-weighted and the refined address-weighted based on the ancillary occupancy data interpolation techniques – were applied to the case site of Leicester City. These techniques predict small-area population based on the actual census population values of source zones. The interpolation is conducted from the larger geographical units (UK, census wards level) to target areas (output areas) of lower spatial units in the whole case sites. The interpolation results show that each model produced different population distribution for the target areas due to the different assumption of applied areal interpolation models.

The estimated population totals were compared with the actual census population counts of the target units for model validation and for accuracy assessment of population estimates. It is possible to validate these interpolation results owing to the availability of actual population counts of output areas. First of all, the areal weighting method was applied. This model is the least accurate within interpolated population results. This may have occurred due to the extent of small residential areas within the larger output areas. The areal weighting method distributes population counts to all areas without considering non-residential areas within the target areas. Secondly, the binary dasymetric method was applied. Binary dasymetric interpolation methods utilised additional datasets to distribute spatial variables, albeit, the areal weighting method does not employ auxiliary datasets. The use of a land cover map as additional information achieved good results in the study region. The volumetric estimation model used building volume information and obtained relatively acceptable population totals. Finally, the address-weighted interpolation approach performs most consistently across study regions within the conducted areal interpolation frameworks.

The relationship between the result of population estimates and the additional datasets is statistically evaluated by comparing their accuracies. Also, correlation coefficients (R^2) and root mean square errors (RMSEs) were measured. Interpolation results shows that residential building address point data provided the lowest RMSE values and the highest R^2 values in Leicester City. The binary dasymetric mapping method provided the second lowest RMSE values and acceptable R^2 values. In addition to this, the highest RMSE and the highest R^2 values were provided by the areal weighting method.

The best estimated population values were obtained by making use of the address-weighted dasymetric approach where R^2 values are greater than 0.93. All the results of the error measurements strongly suggested that the accurate residential area and residential building delineation is required for reliable spatial variable interpolation. Scatterplots of areal interpolation results were created to visualise the difference between actual population reported by census and predicted population values. This study indicates that the address-weighted dasymetric model produced the most accurate population estimates, followed by the binary dasymetric mapping method and the volumetric estimation technique. Conversely, the simple areal weighting technique achieves the least accurate population estimation results in terms of regression coefficient.

Overall, the proposed dasymetric mapping model indicates considerable improvement over the areal weighting interpolation technique in the instance of Leicester City. Differences in population estimates show that the predicted results fully rely on the applied interpolation approach and the used control variables. The study shows that the address-weighted dasymetric model performs relatively better than both areal weighting and the binary dasymetric method in the City of Leicester. Consequently, if building address point datasets are available, they can be used to generate precise population estimates at any defined spatial level. These population disaggregation models were applied to the Borough of Kensington and Chelsea, London, in order to measure the performance of the proposed method in different environments. The Borough of Kensington and Chelsea has more complex building structures, more dense residential areas and more high-rise building blocks than Leicester City. As presented in the Leicester case, dasymetric methods have produced precise population estimates in sparsely dense urban areas. The performance of the dasymetric methods and selected ancillary datasets has been examined in more densely urban areas. The results of the population estimates are detailed in the following chapter.

CHAPTER 5

Areal interpolation results for the Borough of Kensington and Chelsea

5.1 Introduction

This chapter contains the results of areal interpolation approaches for the Borough of Kensington and Chelsea, London. The study area is characterised by more high-rise buildings compared to the City of Leicester and the authority contains a wide range of buildings and structures. Three existing and two novel areal interpolation processes were used to estimate population at small-area level in more dense residential areas. The UK 2011 census counts were disaggregated using the same areal interpolation approaches used for the City of Leicester. This is to allow the performance of each interpolation method to be compared by applying these methods in densely and sparsely populated environments. The census wards were used as source zones, output areas as target zones and external input data as control variables, as indicated in case site one. A 25cm spatial resolution aerial photography was used as ancillary input for the binary dasymetric approach. The building height derived building volumes and residential address point datasets were used as ancillary inputs to estimate population at the finest output level. The address-weighted dasymetric model was refined using building occupancy information as ancillary control variables. The main objective is to apply these interpolation methods with several external inputs in order to measure the performance of alternative interpolation techniques to obtain precise population estimates at intended geographical scale. All the dasymetric mapping methods obtain relatively precise estimation results in the implemented areas.

The accuracy measurements explained in Chapter 3 were utilised in order to assess the accuracy of the results of dasymetrically provided population estimates. They are RMSE, S and R squared measurements. Due to the availability of actual output area population, it is possible to compare the performance of each population estimation model for the accuracy of results and the validation of areal interpolation techniques. The estimation maps, regression models and residual maps were observed for each areal

interpolation based result in order to detect the relationship between the estimated population values and the census released true population counts. Estimation results were plotted against actual population in order to observe how well the points are spread around the regression line.

The results of the areal interpolation processes are represented in Section 5.2. Section 5.3 presents the comparisons of the areal interpolation processes used to generate population estimates. This comparison includes that of the areal weighting method and the binary dasymetric mapping, the volumetric estimation model and the address-weighted dasymetric method. As a consequence, Section 5.4 is a summary of population estimation results for the study region.

5.2 The Results of the Areal Interpolation Approaches

The same interpolation methods were employed in the Borough of Kensington and Chelsea. The proposed dasymetric mapping frameworks were performed in the case of Leicester City (see Chapter 4) and the interpolation results were compared. In the second study case, these interpolation processes were applied in a different environment to examine the performance of the disaggregation process and the chosen external input data. In the following sections, the results of each method are presented.

5.2.1 The areal weighting interpolation method

The areal-weighted process was used to estimate population totals at the output area level in the Borough of Kensington and Chelsea as it was done to Leicester City. The population totals were distributed from census wards to output areas. Figure 5.1 shows the results of the area-weighted technique obtained population estimates of output area for the Borough of Kensington and Chelsea. This approach estimates population totals concerning uniform distribution within the given administrative area. The population density of source zones and the size of target zones are the main parameters used to estimate population counts of each target. The assumption of homogenous distribution can be seen from the created estimation of the study region (see Figure 5.1). Population is mostly concentrated in areas of residential use. Non-residential area uses are the main error sources in the area-weighted interpolation technique. Interpolation results show population totals are located everywhere without considering non-residential uses.

The true population of census wards reported by the census were disaggregated to output areas using the area-weighted method. The population density of the source units was measured by dividing the total area of census wards with their population. The measured population densities of the census ward were employed as control variables in order to estimate the total population total of each target zone. The area-based population estimates were mapped at output area level. The predicted population values were grouped into five classes in order to visualise overestimated and underestimated target zones. The legend of the estimation map shows that high-populated output areas are coloured dark brown and low-populated output areas are shown in a yellow colour. The map of the area-weighted population estimates shows that a large number of people are assigned to the large-sized target units (output areas with a brown colour) and a small number of people are placed to the small-sized target units (output areas with yellow colour). The population was overestimated in the areas where non-residential uses cover most of the study region. These high population estimates of target zones comprise large-sized output areas such as cover parks, trees and shopping centres. This can be seen clearly by comparing aerial imagery of the study area (see Figure 3.9) with the map of area-weighted obtained population estimates (Figure 5.1). However, population totals were underestimated in small-sized output areas. These outputs are widely cumulated in the central part of the study region. In reality, their actual population is higher and residential buildings are concentrated, but due to the nature of their area, less people are located incorrectly and they become subject to underestimation.



Figure 5.1: The population estimation results of the areal weighting interpolation method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

In order to test the relationship between estimated and actual population of each output area, a regression model was created by plotting estimated population against the census reported population values (see Figure 5.2). The point patterns were scattered randomly and do not show any consistency along regression line. The scatterplot suggests that the correlation is not strong between plotted values, and numerous outliers are obvious. This process demonstrates that the performance of the area-weighted interpolation approach completely depends on the uniform distribution of population within the given spatial units. It can be seen from the result of areal interpolation that the large-sized target zones are subject to overestimation, while the small-sized target zones are subject to underestimation. Similar estimation results were obtained for the City of Leicester.

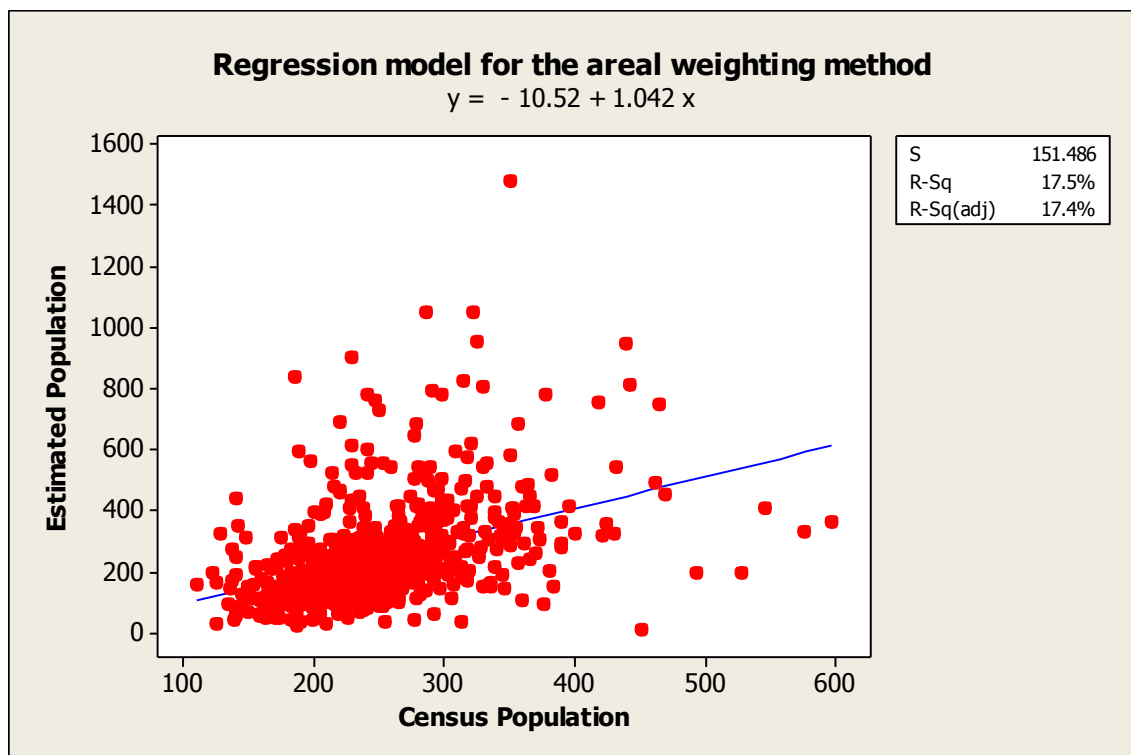


Figure 5.2: The regression model for the areal weighting interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.

In order to measure the accuracy of the predicted population values, the areal weighting obtained population totals were compared with the census derived population totals. Accuracy was assessed using the same measurements utilised for the results of areal interpolation in Leicester City; these are regression coefficient, the root mean square error (RMSE), and the square root of the mean square error (S). The area-weighted model obtained the correlation R^2 is 0.17, which suggests that the relationship is very low and insignificant between the estimated population totals and the actual population counts. Table 5.1 shows the results of accuracy measurements of regression coefficients, R squared and RMSEs for the implemented interpolation techniques. These poor results can be expected in areal weighting due to its uniform assumption. The area-weighted interpolation technique assumes uniform population distribution within the source zone. Therefore, the method overestimates in large output areas and underestimates in smaller output areas. RMSE allows the researcher to compare the magnitude of population estimation errors of alternative interpolation models that used common sets of spatial units within the study region (Langford, 2013, Reibel and Bufalino, 2005).

Table 5.1: The accuracy measurements for the areal weighting interpolation obtained population estimates for the Borough of Kensington and Chelsea.

Accuracy Measurements	Accuracy Results
Regression Coefficients	1.041
R Squared	0.175
Standard Error	151.5
RMSE	4.510
Relative Error (RE)	0.4054
Percentage Error (PE)	40.54%

As a conclusion, the results indicate enormous differences between predicted values and actual population figures. In a normal situation, population totals are distributed heterogeneously and the homogenous distribution is generally not possible. Therefore the results of the area-based population estimates suggest that areal weighting process may not be a good choice to estimate population totals if the ancillary input data is available to distribute population.

5.2.2 The binary dasymetric mapping method

This interpolation method estimates the population of target zones using the external input data to locate population totals to only residential area uses. An aerial photography derived land cover map was used to predict population counts at output areal level in the Borough of Kensington and Chelsea as it was processed for the City of Leicester. Figure 5.3 shows that binary mapping technique provided a dasymetrically population distribution surface. Census wards were used as source zones and land cover classification was used as a control variable to disaggregate population counts to target zones. The legend of the map demonstrates a range of population residual values in five classes with the colour changing from yellow to dark brown as the predicted population totals increase. As can be seen in the map of dasymetric surface (Figure 5.3), population totals were distributed to pre-defined residential areas within the source zones. A large number of people were placed in residential built-up areas (labelled by a brown colour) and the small numbers were placed to non-residential uses (labelled by white and yellow colours). It can be clearly seen from the classified imagery of the study area that highly populated areas overlap with predominantly residential areas and the zero and less populated areas coincide with non-residential areas (see Figure 3.9 in page 66 and Figure 3.11 in page 69). The land cover based interpolation process resulted in heterogeneous population distribution - similar estimation results were obtained in the case of Leicester City. While parks, tree cover areas and open areas are mostly zero-populated areas, population is concentrated in areas where building structures are dominant (see Figure 5.3). These results indicate that when the binary masking of land cover data is used as control variables in the process of dasymetric mapping, population totals are disaggregated to built-up areas. As a second stage, the disaggregated population totals were aggregated to outputs to estimated population totals at output area level. Figure 5.4 demonstrates the binary dasymetric approach obtained from total population estimates of output areas. This map of output areas with population totals shows output areas with low population counts in a yellow colour and those with high population in brown.

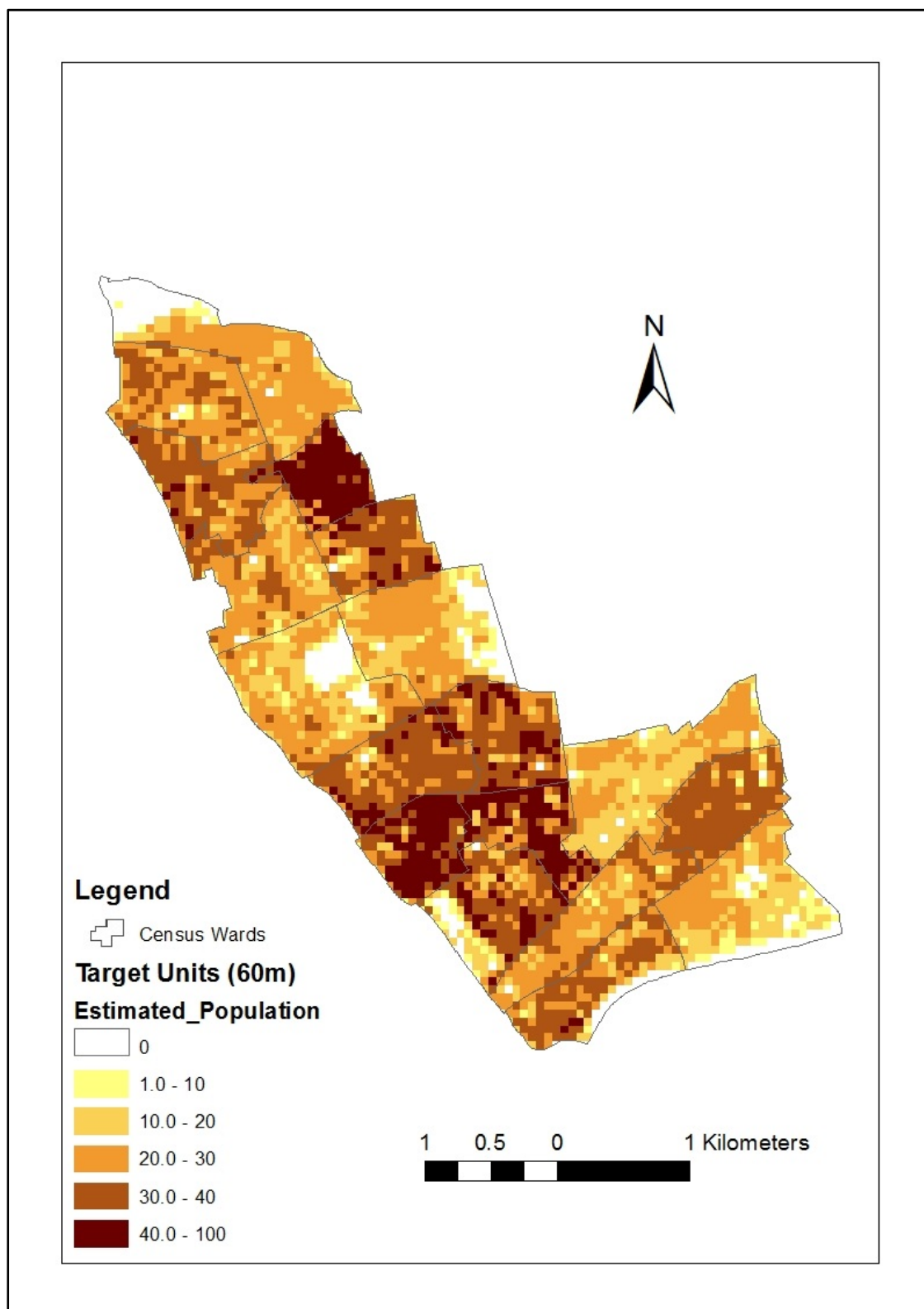


Figure 5.3: Dasymetric surface of the Borough of Kensington and Chelsea obtained by using the binary dasymetric mapping method. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

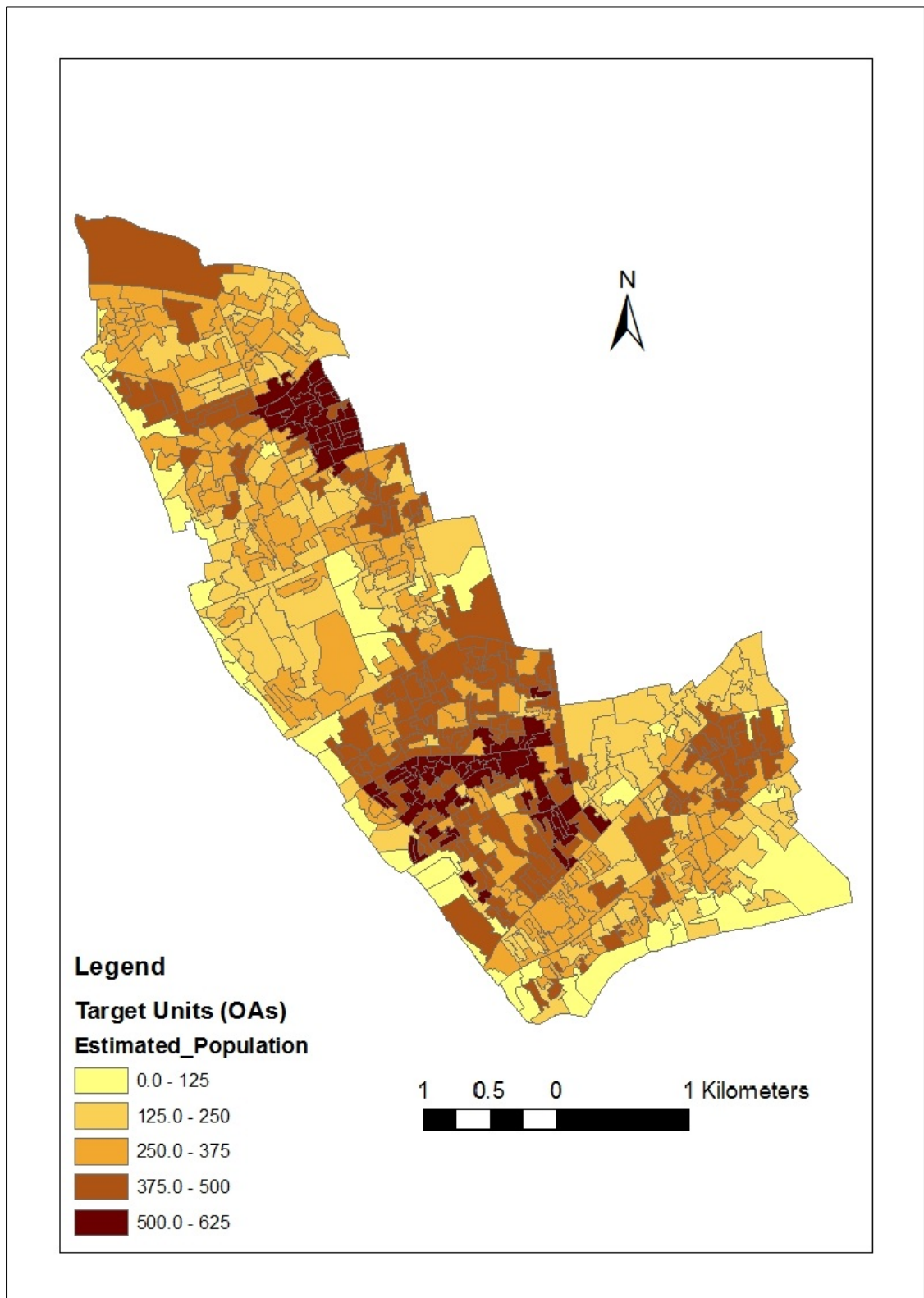


Figure 5.4: Population estimation results of the binary dasymetric mapping method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

In order to measure the accuracy of the binary mapping model, the predicted and census reported population totals were compared. Accuracy is measured by using the same accuracy analysis used for the areal weighting obtained population estimates. The binary dasymetric technique has the correlation R^2 is 0.68, which suggests the correlation is stronger than the area-weighted method between the estimated population and the census released population totals. Table 5.2 presents the regression coefficients, R squared and RMSEs for each applied disaggregation model. The use of ancillary imagery data in binary the dasymetric method obtained more accurate population estimation results than the areal weighting interpolation technique which does not use ancillary inputs. This is because inhabited areas are not taken into consideration in the areal weighting model and, as such, are subject to estimation errors in areas where non-populated areas largely cover the given study area. Estimation results reveal that the performance of the binary interpolation process widely depends on the realistic classification of a land cover map. This is because the land cover based interpolation process disaggregates population counts through only built-up areas. If the classification of the imagery data were not relatively accurate, people may be incorrectly placed to non-residential uses. This misclassification may affect accuracy of the estimated population totals.

Regression analysis was conducted for the land cover based population estimates in order to analyse the relationship between predicted and actual population values as created in the previous chapter. Figure 5.5 shows a scatterplot of the binary mapping method obtained estimation results. The y axis represents estimated values and x axis represents actual values. Most of the point patterns were spread around the regression line moving from the lower left to upper right, yet, there is large number of outliers. This scatterplot suggests that the correlation is stronger between the binary dasymetric method derived population values than the areal weighting obtained estimation results. These estimation results and the accuracy measurements show that the performance of the binary dasymetric mapping process directly depends on the accuracy of the classified imagery product. In land cover based population estimates, misclassified non-residential areas and non-residential building structures within the built-up areas are the main error sources. The binary model mostly begins to fail in areas residential builds area mixed with non-residential manmade structures. People do not actually reside on

the non-residential structures. It is notable from the experiments that these areas are subject to overestimation and underestimation in the study region.

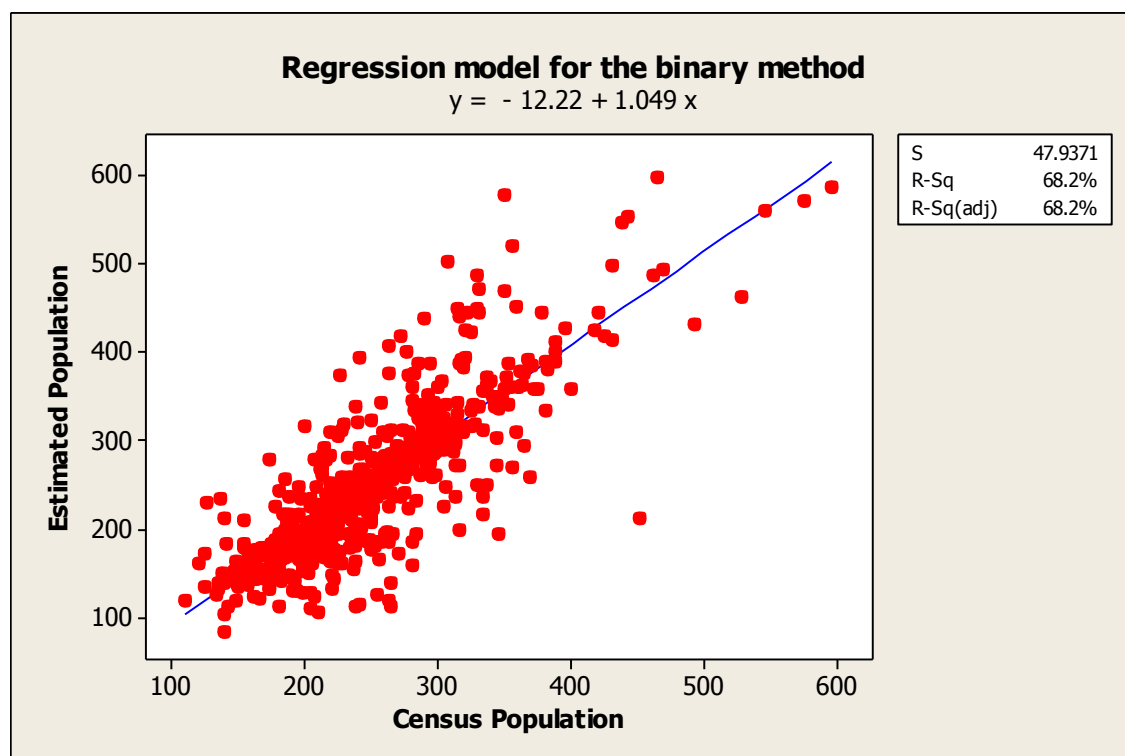


Figure 5.5: Regression model for the binary dasymetric mapping interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.

Table 5.2: Accuracy measurements for the binary dasymetric mapping method obtained population estimates for the Borough of Kensington and Chelsea.

Accuracy Measurements	Accuracy Results
Regression Coefficients	1.048
R Squared	0.682
Standard Error	47.94
RMSE	0.899
Relative Error (RE)	0.1260
Percentage Error (PE)	12.60%

Overall, these results suggest that external input data have substantial importance to distributed population totals throughout the study areas. When compared with the area-weighted interpolation method, population values were located heterogeneously through

built-up areas within the target zones. However, non-residential building structures are the main error sources in the binary dasymetric process. Lastly, the accuracy assessment of the binary method reveals that external imagery data has improved the accuracy of population estimates. Remote sensing data can be used as control variables to disaggregate population totals from source zones to target zones. The result of binary dasymetric disaggregation for Kensington and Chelsea is slightly less accurate than for Leicester City. The reason can be that Kensington and Chelsea has more mixed-use building blocks. Also, classified images show that Leicester City has a more segregated land use compared to the Borough of Kensington and Chelsea (see, figures 3.10 and 3.11). The segregation of land use may affect population estimation results.

5.2.3 Building volume based dasymetric method

The building volume based dasymetric approach is the third process utilised to disaggregate population counts from source zones to target zones. This approach used building volume data (see, Figure 5.6) as control variables in order to identify residential buildings for obtaining a dasymetric population estimation of the Borough of Kensington and Chelsea. Census wards were used as the source zones, and the output areas were used as target zones in order to obtain an estimation of population totals for each output area within the study region. Results of the population estimate obtained by the volumetric estimation method is illustrated in Figure 5.7. The volumetric method assumes that people reside only in residential housing units and population was assigned through the buildings according to their volumes. Therefore, realistic information of residential buildings is significant to distribute population counts.

First of all, the population counts of source zones were disaggregated to individual residential buildings based on their volumes. Secondly, the population of buildings, which are intersecting target units, were summed to estimate the population at output area level. The estimated population totals were mapped showing the population of each output area in the Borough of Kensington and Chelsea. The predicted population totals were grouped into five classes to present lower estimated and higher estimated output areas. The legend of the estimation maps shows low-populated target units in a yellow colour and high-populated target units in a dark brown colour. Volume based technique considers the presence of both low and high-rise buildings within the study region. Building volume information is important to define residential information in the study

area. The map of estimation results demonstrates that population counts are distributed heterogeneously through the target units based on their volumes (see Figure 5.7).

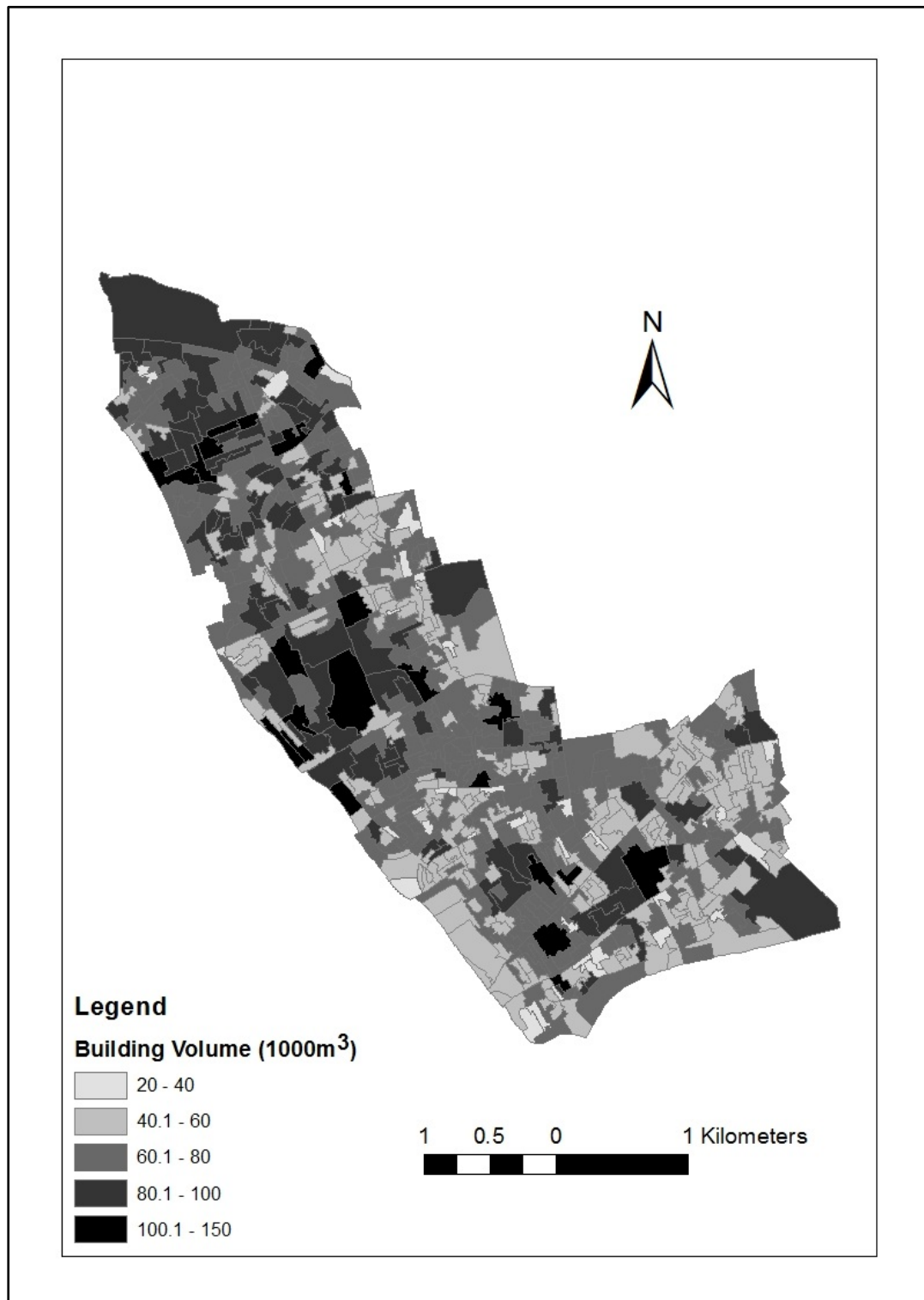


Figure 5.6: Map of building volumes by output areas for the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service and All LiDAR data © Airbus Defence and Space (Astrium Ltd. 2014).



Figure 5.7: Population estimation results of building volume based dasymetric mapping method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

In order to evaluate the correlation between building volumes based population totals and actual population counts, regression analysis was carried out similar to previously used interpolation models (see Figure 5.8). The predicted values were plotted in the y axis and census reported actual values were plotted in the x axis. The point patterns showed linear distribution around the regression line. There are several outliers of point patterns in the scatterplot. This regression analysis suggests that there is a very strong relation between volume based estimations with correlation value R^2 of 0.81. Accuracy analyses suggest that there is a strong relation between volume based estimates and census reported actual population totals. This analysis demonstrates that a volumetric estimation method can achieve population estimates with higher accuracy that is close to actual population values by using realistic information of residential building volumes. The results also show that high-volume buildings are subject to overestimation and low-volume buildings are subject to underestimation.

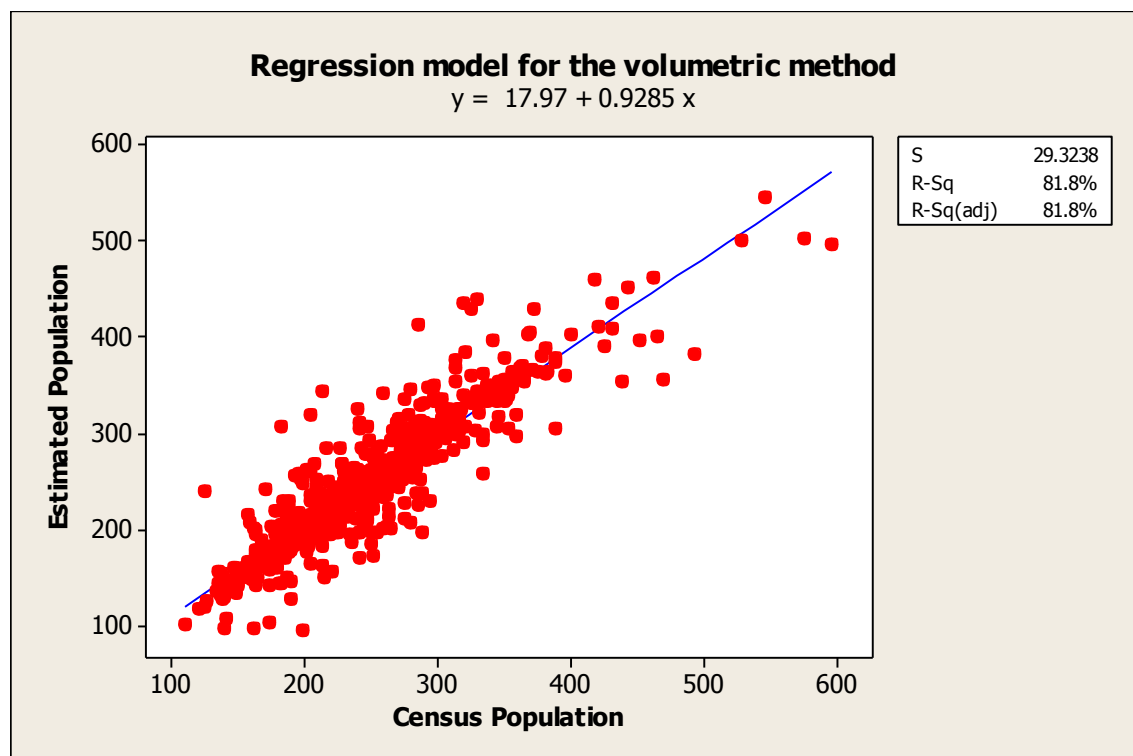


Figure 5.8: Regression model for the building volume based dasymetric mapping interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.

Table 5.3: Accuracy measurements for the volumetric estimation technique obtained population estimates for the Borough of Kensington and Chelsea.

Accuracy Measurements	Accuracy Results
Regression Coefficients	0.928
R Squared	0.818
Standard Error	29.32
RMSE	0.746
Relative Error (RE)	0.0773
Percentage Error (PE)	7.73%

Estimated and actual population values were compared to assess the accuracy of the volumetric dasymetric method. Accuracy was measured by using the same methodologies utilised in the previously implemented interpolation frameworks. Table 5.3 presents the measurements of the regression coefficients, R squared and RMSEs for each of the implemented interpolation techniques. A comparison of the volumetric results with land cover based results reveals that residential building volume information is superior to estimate population totals in the case of the Borough of Kensington and Chelsea. The residential building data are much more detailed than residential land use and road networks to identify housing units where people actually reside (Sridharan and Qiu, 2013, Zandbergen, 2011). Therefore, population totals are placed to building footprint based on their volumes rather than residential uses. However, the social status of residents may affect building volume and resident relationship. Residential building volumes are likely to be higher in areas where people with higher incomes reside than areas where people with lower incomes live (Sridharan and Qiu, 2013). The volumetric estimation model has achieved a comparable estimation accuracy with a land cover based model deploying residential building volumes as control variables in the study region. As a result, population estimates may be more accurate using the volume of individual housing units. That makes the volume based model the method of choice when volumetric ancillary input data is available.

5.2.4 The address points based dasymetric mapping method

The address-weighted dasymetric methodology uses building point data in order to estimate population totals of target areas. The residential population totals were

disaggregated from source zones to target zones by using residential address points as ancillary inputs. Due to this reason, the number of residential housing units are primary control variables for the estimation process. This point based interpolation method was performed to estimate population totals at output area level in the Borough of Kensington and Chelsea. This study region is located in Greater London where high-rise buildings are mostly dominant and it is a densely populated urban area compared to Leicester City. The UK 2011 census of population values were disaggregated from census wards to output areas by using the building address points and the average household size as control variables. Figure 5.9 shows the population totals obtained by the address-weighted dasymetric model. Estimated population totals were grouped into five classes. The legend of the map demonstrates these classes in different colours: the yellow colour presents the lower estimated population totals and the dark brown colour presents the higher estimated population totals. It can be seen from the estimation map that more people were placed in target zones which consist of more housing units and less people were located to target zones which consist of less housing units. This can be clearly seen by comparing Figure 3.19 (on page 82), that shows the total numbers of residential housing units, with Figure 5.9 that presents address based population estimates for output areas. The predicted population totals were compared with the known census population figures to assess the performance of the building address based on estimation model and the measure the errors introduced. Estimation accuracy results in Table 5.4 illustrate the regression coefficients, RMSEs and R^2 for dasymetric interpolation techniques. These calibration tests revealed that the address-weighted approach in densely populated areas produce less accurate population estimates than sparsely populated areas. These findings reveal that the address-weighted model produces a relatively precise estimation of population totals in areas where building address point datasets area is available. However, the chances of empty residential housing units are the main errors in this approach. The larger errors occurred in densely populated urban areas. These results illustrate that an address-weighted model performance fully depends on accurate address point information.

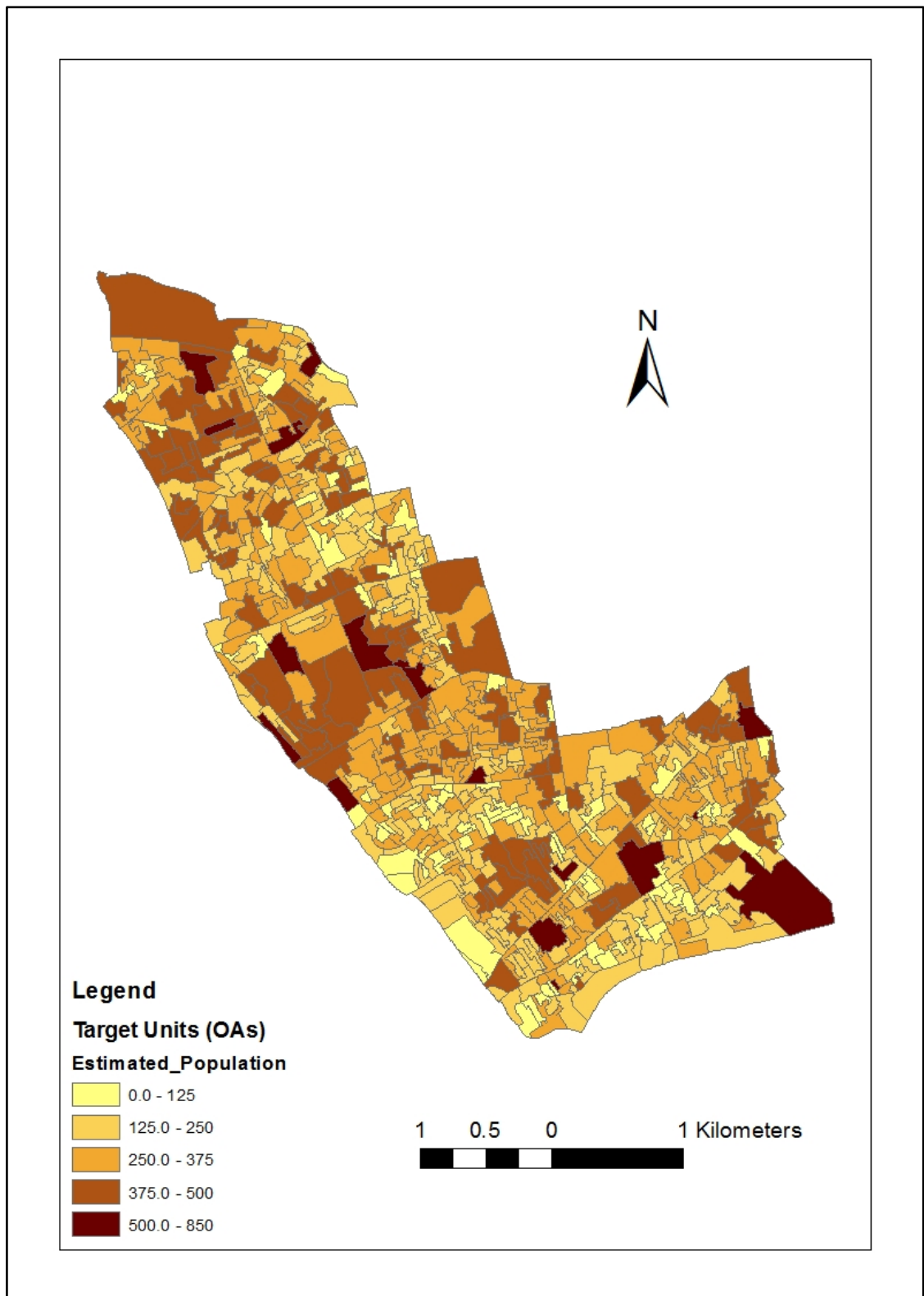


Figure 5.9: Population estimation results of the address-weighted dasymetric mapping method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

Regression analysis was carried out to observe if there is any variation of points from the regression line in the scatterplot. This scatterplot demonstrates the relationship between areal weighting interpolations obtained population totals and the actual population figures. The estimated population totals were plotted in the y axis and the true population counts was plotted in the x axis (see Figure 5.10). All point patterns are mostly fitted along a regression line by showing linear distribution, but few outliers can be seen in the scatterplot of address-weighted based estimates. This was done in order to underlay the relationship between single housing units and population totals to obtain accurate population totals for each area. This analysis shows that the performance of building an address based dasymetric process immediately depends on the number of residential housing units. The non-occupied housing units within the target zones are potential sources of error and they are subject to overestimation in the study region. The analyses provide evidence that population totals can successfully be estimated with the availability of high resolution building attribute information at intended scale.

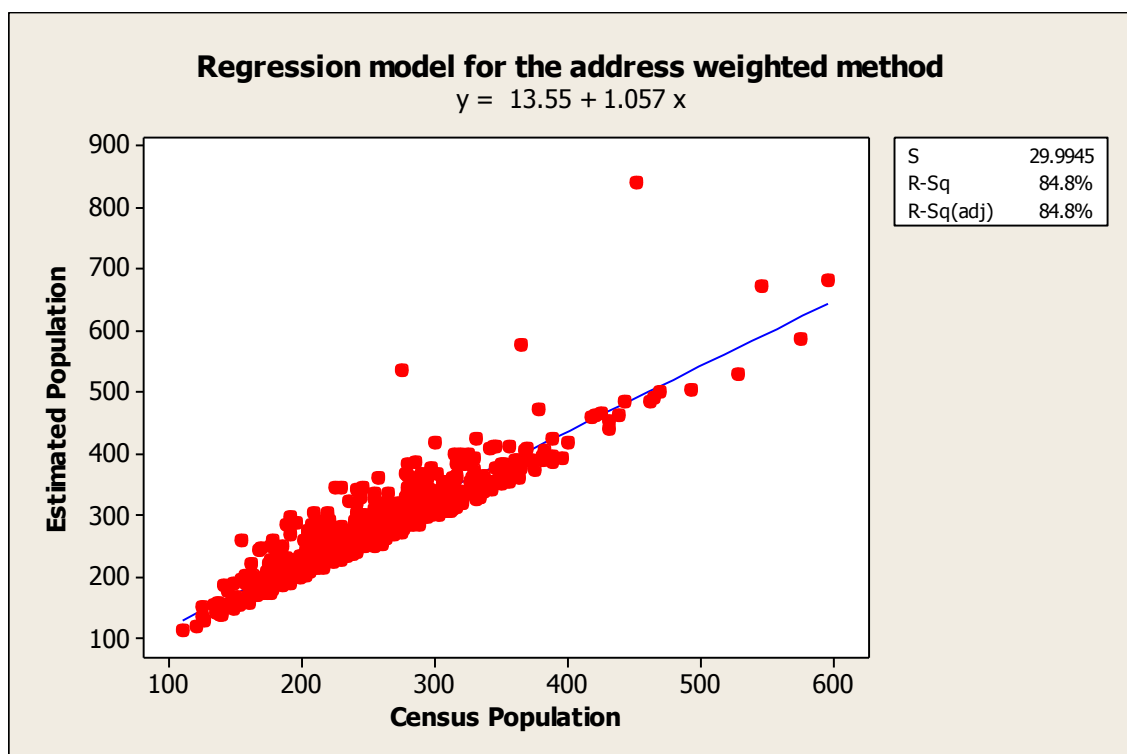


Figure 5.10: Regression model for the address-weighted dasymetric mapping interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.

Table 5.4: Accuracy measurements for the address-weighted dasymetric method obtained population estimates for the Borough of Kensington and Chelsea.

Accuracy Measurements	Accuracy Results
Regression Coefficients	1.056
R Squared	0.847
Standard Error	29.99
RMSE	0.796
Relative Error (RE)	0.1105
Percentage Error (PE)	11.05%

Overall, the address-weighted dasymetric methods obtained much more precise population estimates than the other interpolation methods used in the study region. It shows that an address based interpolation method can be a strong alternative dasymetric method when ancillary inputs are available for both source and target areas. The results of these studies attest that building point datasets can be used as control variables to disaggregate population totals from source zones to target areas.

5.2.5 The address-weighted method with occupancy information

This section combines the residential building address point data with building occupancy information in order to disaggregate population totals from source zones to target zones in the study region. The main objective was to evaluate the performance of address point data with building occupancy information as external inputs in the dasymetric interpolation methodology. The address layer data shows the location of residential housing units within the source zones. Additionally, average population count per housing unit and building occupancy information are combined to an interpolation framework to distribute population totals to individual housing units within the given target zones. The address-weighted dasymetric model was used for the Borough of Kensington and Chelsea in order to disaggregate population values from census wards to output areas.

The 2011 UK census values were located to output areas based on the control variables of building address and occupancy information. The estimated population values of output areas were mapped (see Figure 5.11). As can be seen from the legend in figure 5.11, predicted values were grouped into five classes. The yellow colour represents

lower estimated values and the dark brown colour represents a higher estimated population value of output areas. The population estimation map demonstrates that population totals were distributed based on the number of residential housing units and occupancy information. As a result, more people were located to output areas, which consist of a large number of occupied residential units, and less people were located to output areas which consist of less numbers of occupied residential units.

The estimated population values were compared with the true population counts to assess the accuracy of the occupancy based estimation model. The accuracy is calculated by using several measurements that were used in previous estimation methods. Table 5.5 shows the results of the accuracy measurements. The improved address-weighted dasymetric model significantly improved the accuracy of areal interpolation with the availability of residential building points and occupancy information. The best results of the population estimates were obtained for the study region by using building occupancy information as control variables. A correlation coefficient of 0.998 was provided and the lowest RMSE value was obtained for all the output areas. All accuracy analyses suggest that the relationship is very strong between predicted and actual population counts. The results show that a very low number of people are placed incorrectly. It determines that the estimated values are very close to the actual population counts. These results reveal that the areal interpolation methods with ancillary data produced similar estimation results as reported in previous studies (Zandbergen and Ignizio, 2010, Hawley and Moellering, 2005).

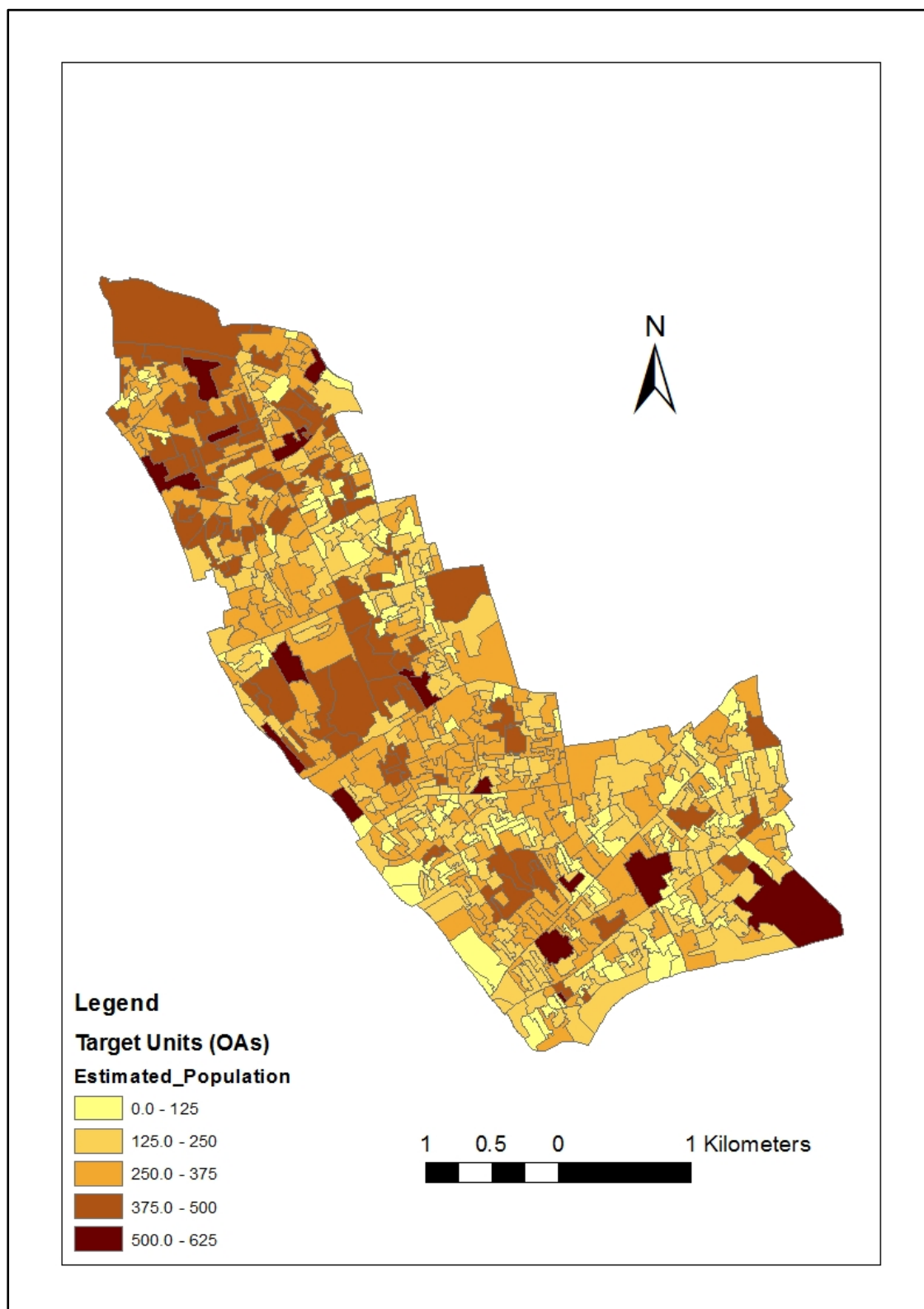


Figure 5.11: Population estimation results of the refined address-weighted dasymetric mapping method in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service.

In order to evaluate the relationship between actual population counts and the estimated population totals, a regression analysis was carried out for the building address point and the occupancy based interpolation approach. The predicted population values were plotted in the y axis and the true values were plotted in the x axis in the scatterplot (see Figure 5.12). The point patterns were moving from lower left to upper right along the regression line. There are not any important outliers in the point patterns. The figure suggests that the relationship is very strong between building occupancy based population estimates and the actual value. This analysis represents the performance of the building address based dasymetric process depends on the accurate number of residential buildings and the occupancy information for the study region. In addition to this, the usage of realistic occupancy information mostly decreased the overestimated population totals in target areas where there is a significant number of non-occupied housing units. This can be clearly seen comparing the scatterplots of the address-weighted (Section 5.2.4) and the refined address-weighted (Section 5.2.5) dasymetric frameworks.

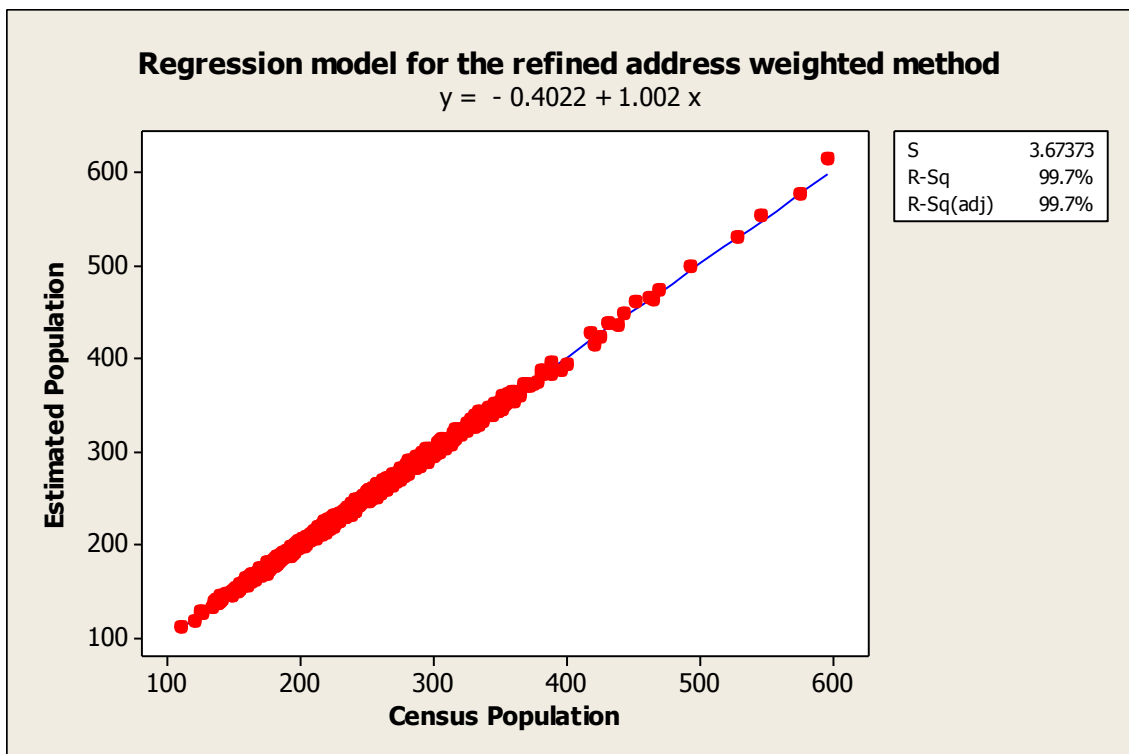


Figure 5.12: Regression model for the refined address-weighted dasymetric mapping interpolation method. The estimated population totals of the Borough of Kensington and Chelsea at output area level versus the actual population of output areas released by census.

Table 5.5: Accuracy measurements for the improved address-weighted dasymetric method obtained population estimates for the Borough of Kensington and Chelsea.

Accuracy Measurements	Accuracy Results
Regression Coefficients	1.002
R Squared	0.997
Standard Error	3.67
RMSE	0.238
Relative Error (RE)	0.0121
Percentage Error (PE)	1.21%

With the help of occupancy information, occupied housing units were used to estimate population totals, and there are not any significant overestimated and underestimated population totals in the occupancy information based population estimates. This is because non-occupied housing units are not considered and the population totals are disaggregated to the occupied housing units correctly. The address-weighted dasymetric model with residential address points and occupancy information provides consistently precise interpolation results in small-area population estimates. These outcomes show that the choice of ancillary input data is important to determine the location of residential housing units. The address points based dasymetric with occupancy data provide the best estimates among the tested areal interpolation. The tested dasymetric interpolation models show a marginal improvement over area-weighted interpolation. Similar results were obtained by Langford, 2013. The availability of address point data provides an opportunity to perform interpolation models in the studies of population estimates, at least in the UK context.

5.3 Comparison of Areal Interpolation Approaches

This present the comparison of the estimation results of areal interpolation approaches for the Borough of Kensington and Chelsea. In this section, the results of areal interpolation techniques were compared in order to observe the performance of each population disaggregation method. Firstly, the differences between the area-weighted method and the binary dasymetric mapping technique were provided. Secondly, the building volume based estimation method and the address-weighted dasymetric approach were compared for measuring the accuracy of population estimation models.

The estimation maps show that there are inevitable differences between all implemented interpolation techniques. This is because each method requires different ancillary input data and their assumption is different in the distribution of population totals. Comparison results demonstrate that the conducted interpolation approaches and the chosen external input data perform well in distributing population counts in the study area. In the following sub-sections, the results of comparisons are described in detail.

5.3.1 Comparison of areal weighting and the binary dasymetric methods

In this section, the results of the area-weighted and binary methods were compared. Estimation differences show the superiority and weakness of applied interpolation methods in the same environments. The areal weighting and the binary methods obtained estimation results were provided in previous sections (Section 5.2.1 and 5.2.2). Residual maps were created by subtracting actual population totals reported by census from the estimated population. Figure 5.13 demonstrates the residual map of the area-weighted method obtained population totals, and Figure 5.14 shows the residual map of the binary dasymetric method provided population estimates at output areal level. The residual maps demonstrate how well interpolation methods estimate population totals of target units. In the residual maps, the estimation errors were grouped into five classes with the labelled colour changing from blue to red as the population residuals increase. Blue tones represent negative estimation errors and red tones represent positive estimation errors. The middle class of the distribution is coloured yellow and represents output areas with less than 25 people incorrectly placed. The spatial error distribution helps to visualise higher negative, moderate and higher positive estimation errors. The results of mapping the differences between these interpolation approaches in the case of the Borough of Kensington and Chelsea show that population totals disaggregated according to the assumption of each interpolation.

The spatial distribution of population estimate errors obtained from the areal weighting interpolation model shows that the population was overestimated in areas where non-residential areas cover large parts of the study region. Looking at the areal imagery of the study region, it can be easily seen that the overestimated areas mostly overlap with these non-populated zones (see, Figure 3.9 in Chapter 3). These green areas are located in the northern and western part of the Borough of Kensington and Chelsea. The other apparent feature is the non-built-up areas close to the boundary of study regions (see

Figure 3.11 in page 69). The method indicates that large negative residuals were obtained in areas where actual population totals are high, and the area of output areas are smaller compare to the output areas close to the boundary of the study region. As mentioned in the preceding chapter, the area-weighted method assumes that populations are distributed homogeneously within the source zones. Because of this reason the areal weighting method assigns population totals everywhere in target zones and underestimate population totals in small-sized target areas. The only superiority of the areal weighting method is that it does not require any external input data. In order to obtain relatively precise population estimates, population totals must be distributed uniformly, but this is not in reality.

The binary dasymetric method assumes heterogonous population distribution within the source zones, unlike the area-weighted interpolation. This method places the people to only built-up areas. This study notes that the differences between areal weighting and binary population mapping are smaller in the target units that are predominantly inhabited areas. The differences were larger in the larger-sized target units that cover both inhabited and uninhabited zones. The spatial distribution of population estimation errors obtained from the area-weighted (Figure 5.13) and the binary dasymetric mapping method (Figure 5.14) reveal that populations were overestimated in the large-sized output areas and underestimated in the small-sized output areas. From the figures, orange coloured output areas appear much more in the residual map of binary method obtained estimates. This proves that the binary method generates better population estimation results with the relatively accurate imagery classification as control variables. The area-based interpolation method obtained the worst estimation results within the applied interpolation models. Similar results can be found in Chapter 4. All interpolation results suggest that the areal interpolation models show substantial difficulty in predicting population count in areas where population is distributed heterogeneously because of the variations in topography and regional development within source zones (Brinegar and Popick, 2010, Wu et al., 2008, Mennis and Hultgren, 2006b, Harvey, 2002).

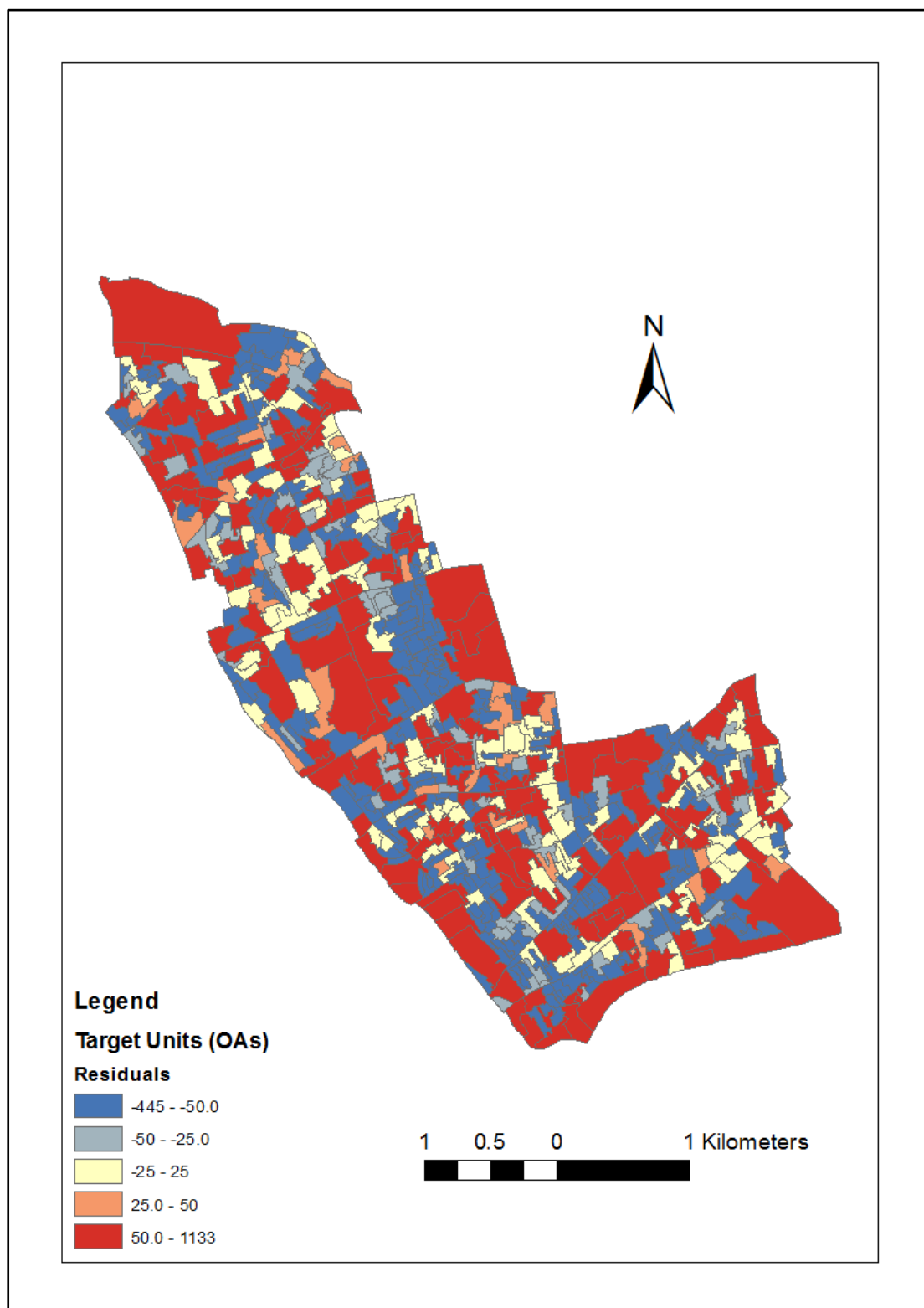


Figure 5.13: Residual map of areal weighting method obtained population estimates in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

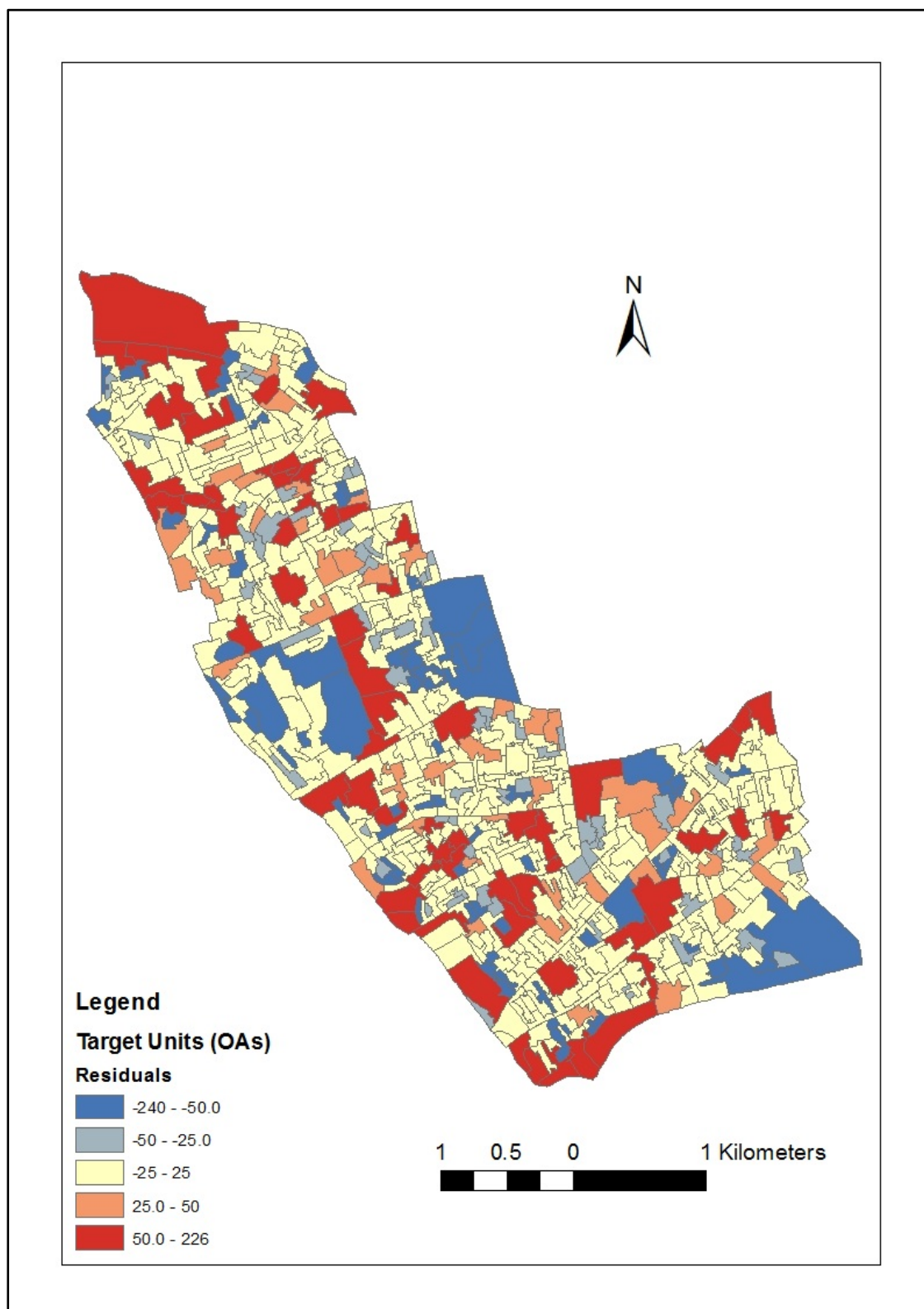


Figure 5.14: Residual map of the binary dasymetric mapping method obtained population estimates in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

In order to test the relationship between population estimation residuals and target units, spatial autocorrelation analysis of Moran's I was applied. Figure 5.15 and Figure 5.16 show the spatial autocorrelation results of the areal weighting and the binary dasymetric mapping methods. The residuals of the areal interpolation based population estimates has a Moran's I value of 0.020894 and the residuals of binary mapping method has a value of 0.033158. The analysis results show that the residual of areal weighted method based estimations are randomly distributed over target zones in the Borough of Kensington and Chelsea. Autocorrelation analysis also indicates a clustered distribution of the binary mapping yielded population residuals. It means that the correlation between population totals and the space depends on ancillary data of the binary masking image used as control variables to disaggregate population totals.

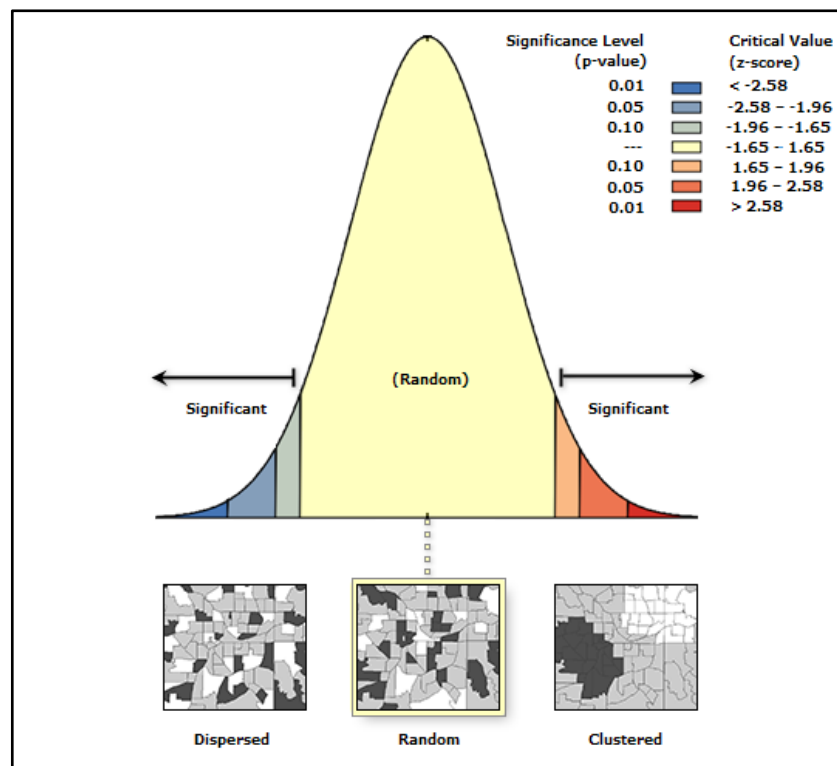


Figure 5.15: Spatial autocorrelation of the areal weighting method provided population residuals in the Borough of Kensington and Chelsea

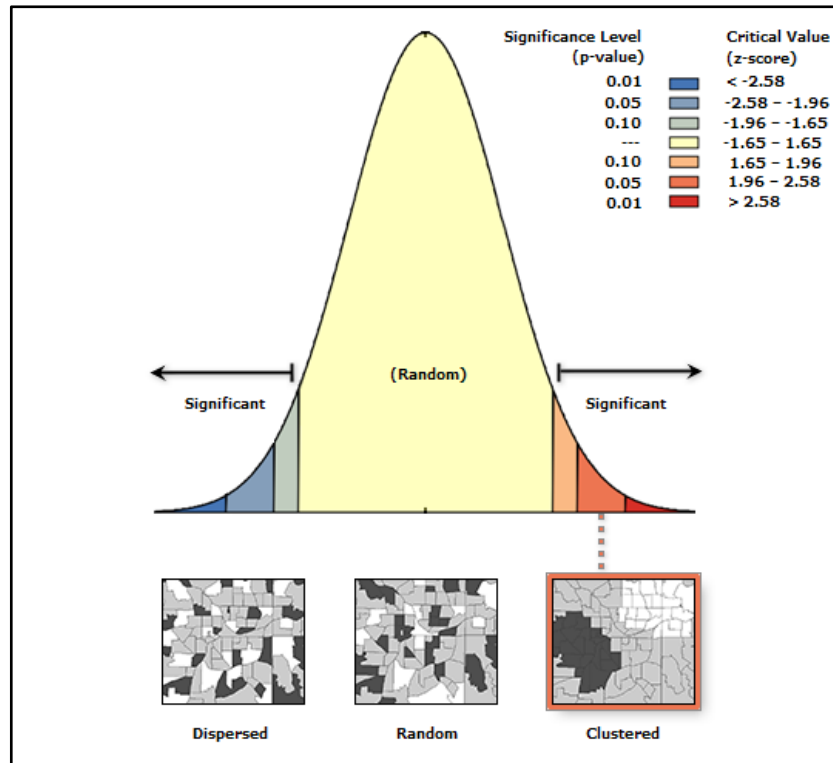


Figure 5.16: Spatial autocorrelation of the binary dasymetric mapping provided population residuals in the Borough of Kensington and Chelsea

5.3.2 Comparison of the volumetric and the address-weighted methods

This section compares volumetric estimation and address-weighted dasymetric techniques. First of all, the census reported true population was subtracted from the estimates population values to observe overestimated and underestimated target units. These residual maps help to demonstrate the difference between both dasymetric mapping methods. Additionally, they allow measuring of the performance of implemented interpolation methods in the study region. Figure 5.17 shows the spatial pattern of errors captured using volumetric estimation and Figure 5.18 shows the spatial pattern of errors produced using the address-weighted dasymetric method. The legend of the estimated error maps display residuals grouped into five classes to show overestimated and underestimated target units. Figure 5.19 shows the spatial pattern of errors captured using the occupancy based address weighted method. The legend of this residual map was re-structured into three classes to show negative and positive residuals. Negative residuals were shown by blue colours and positive residuals were shown by red colours. Yellow colours show the output areas where residuals of population predictions lie between -25 and +25. The spatial distribution of estimation

errors reveals that these approaches yielded different residuals in output areas. This is because assumptions of interpolation models already described in Chapter 3 were used to obtain different population estimates based on the weighing factor of ancillary input data.

The volumetric estimation method places people by using building volumes as control variables. Therefore, larger building volumes are subject to positive estimation residuals and smaller building volumes are subject to a negative estimation error. While spatial distribution of errors reveals that there are considerable negative estimation errors and positive estimation errors over the study region (see Figure 5.17), the address-weighted dasymetric mapping model using address point data produced the lowest error patterns compared to other tested models (see Figure 5.18). This was due to the use of address data as external input data that provides the location of building units by showing where people actually reside. The volumetric estimation method obtained less accurate estimation results than the two address-weighted models that use residential address points as ancillary variables. In Figure 5.17, the spatial distributions of error patterns reveal that a few of the output areas were coloured yellow and most of the output areas were labelled in an orange colour all over the study region. In the central and north easterly parts of the study region, the most output areas were labelled in a brown and dark brown colour. This residual map shows that estimated population totals are close to census reported actual counts, and there are not significant negative estimation errors. However, there are various positive estimation errors due to the assumption that all residential houses are occupied. This interpolation estimation result is understandable because residential building address points theoretically indicate where people actually live. Similar results were obtained when the address based method was applied for the City of Leicester.

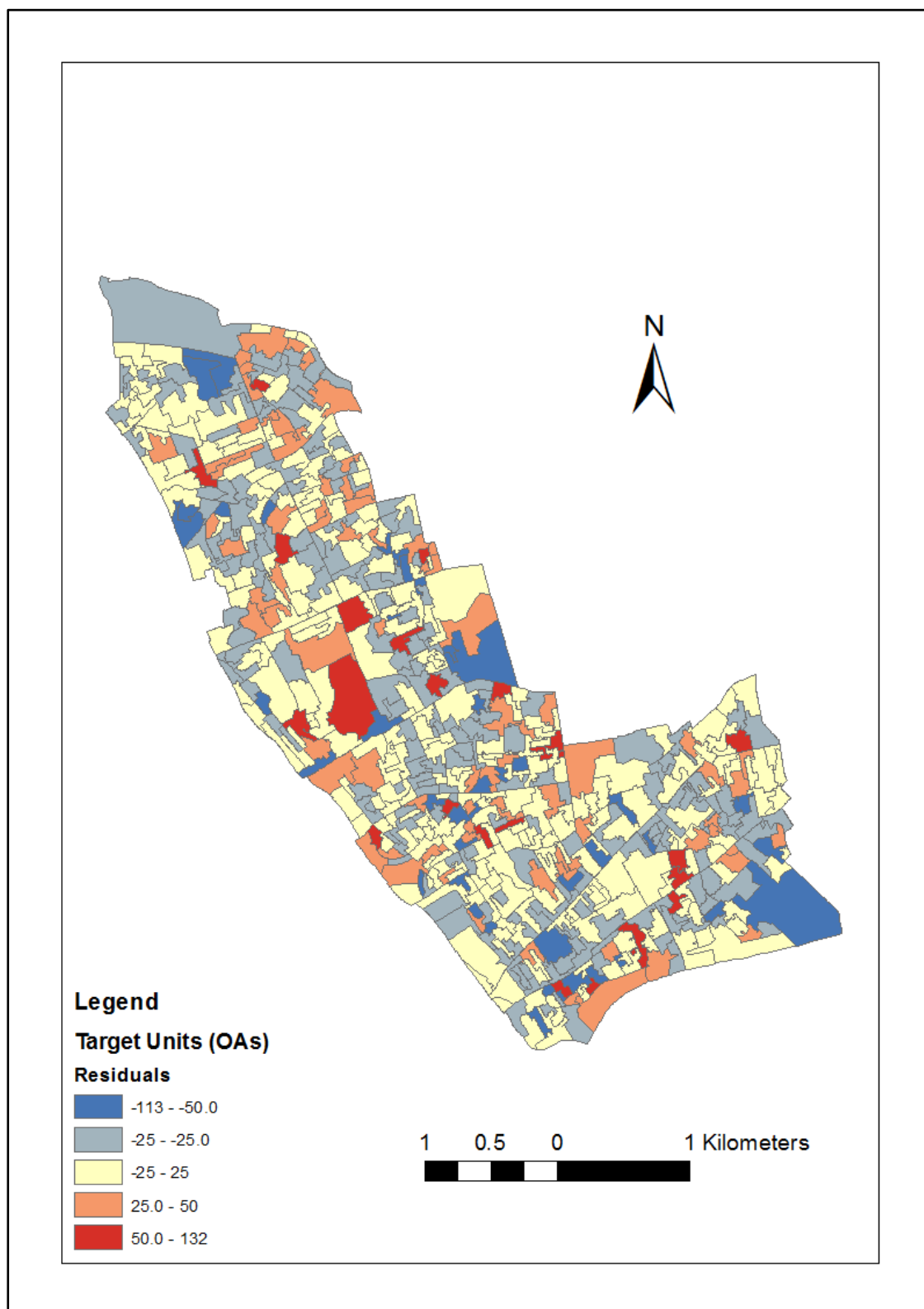


Figure 5.17: Residual map of volume based dasymetric mapping method obtained population estimates in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

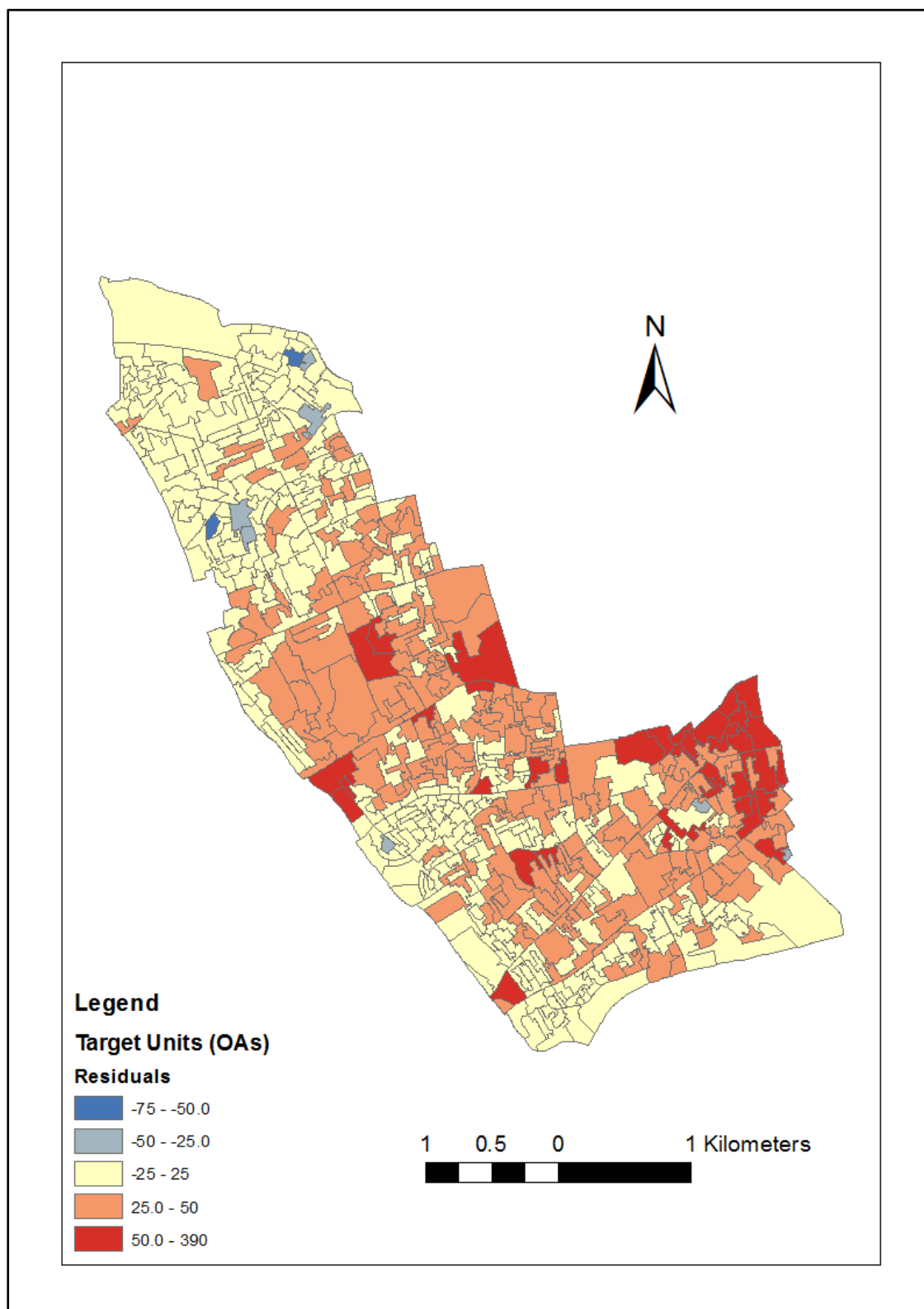


Figure 5.18: Residual map of the address-weighted dasymetric mapping method obtained population estimates in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

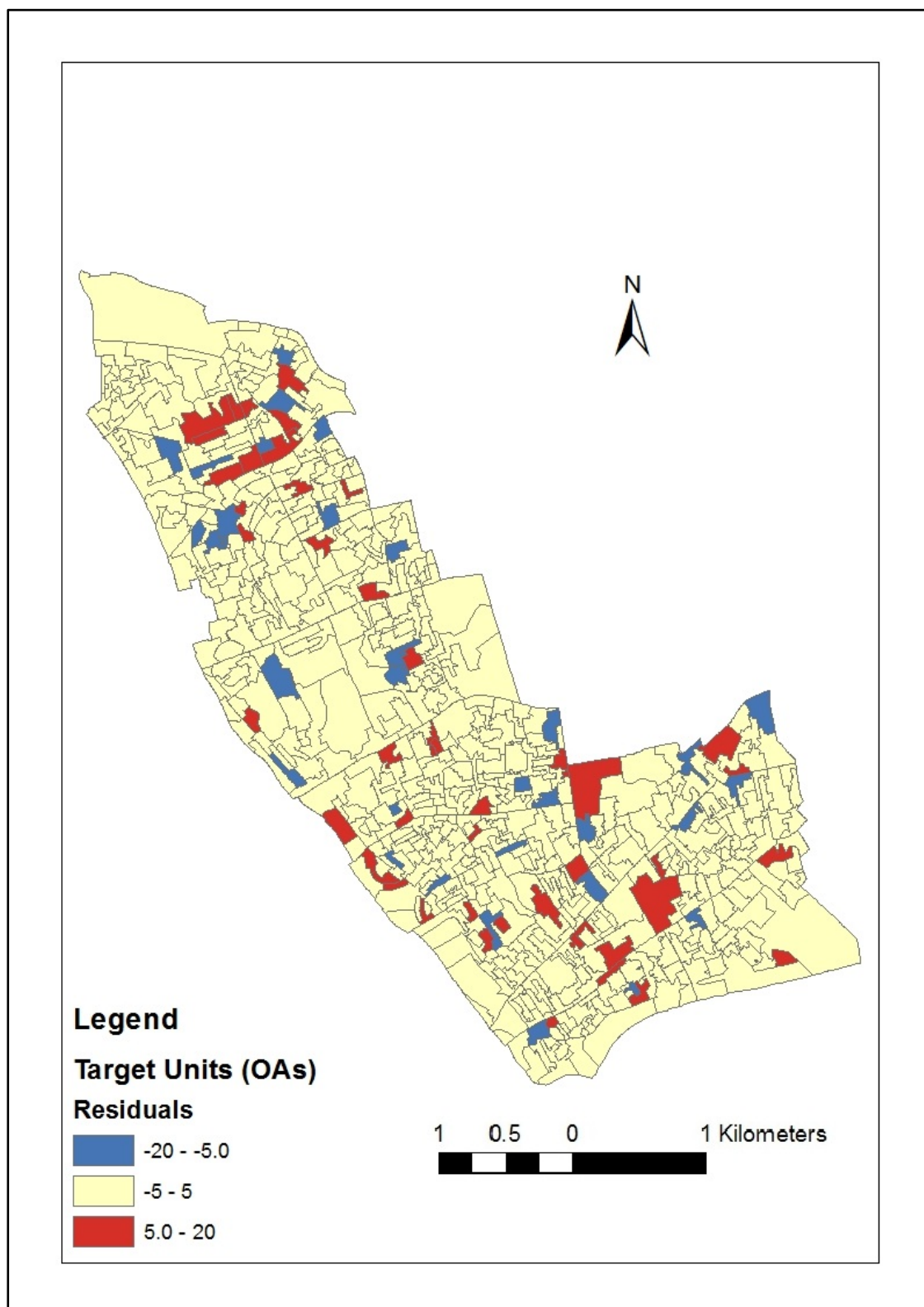


Figure 5.19: Residual map of the address-weighted dasymetric mapping method obtained population estimates using occupancy datasets in the Borough of Kensington and Chelsea. © Crown Copyright/ database right 2013, An Ordnance Survey/EDINA supplied service

In order to measure the relation between estimation residuals and local areas, spatial autocorrelation analysis of Moran's I was also applied for the results of the volumetric estimation and the address-weighted dasymetric models. The spatial autocorrelation of the results of volumetric estimation, and the address-weighted and the occupancy data based disaggregation model are shown in figures 5.20, 5.21 and 5.22. The residuals of building volume based population estimates has a Moran's I value of 0.000495, building addresses based estimations has a value of 0.298596 and occupancy datasets based estimations has a value of -0.025765. The results of autocorrelation analysis display that the residual of volumetric method based estimations were randomly distributed over target zones in the Borough of Kensington and Chelsea. Therefore, there is no negative and positive correlation between the observed value and neighbouring locations. However, the result of autocorrelation analysis shows a clustered distribution of the address-weighted method obtained population residuals. The application result indicates a random distribution of population residuals in the building volumes and building address points used to estimate population totals. Therefore, there is not any negative and positive correlation between the observed value and neighbouring locations. However, the residuals of the address-weighted model are spatially clustered when building addresses and occupancy information were combined and used as ancillary datasets to disaggregate population values.

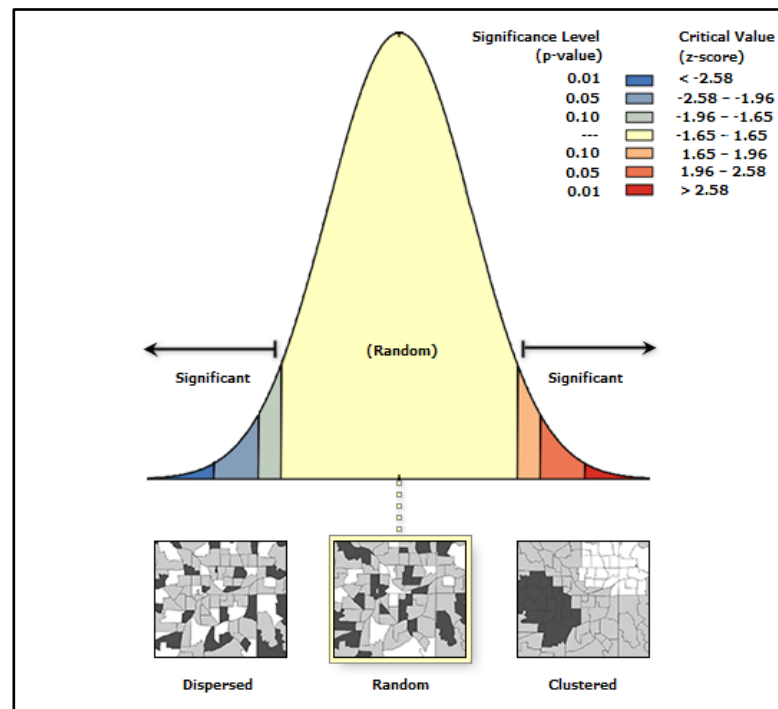


Figure 5.20: Spatial autocorrelation of the volumetric estimation model provided population residuals in the Borough of Kensington and Chelsea.

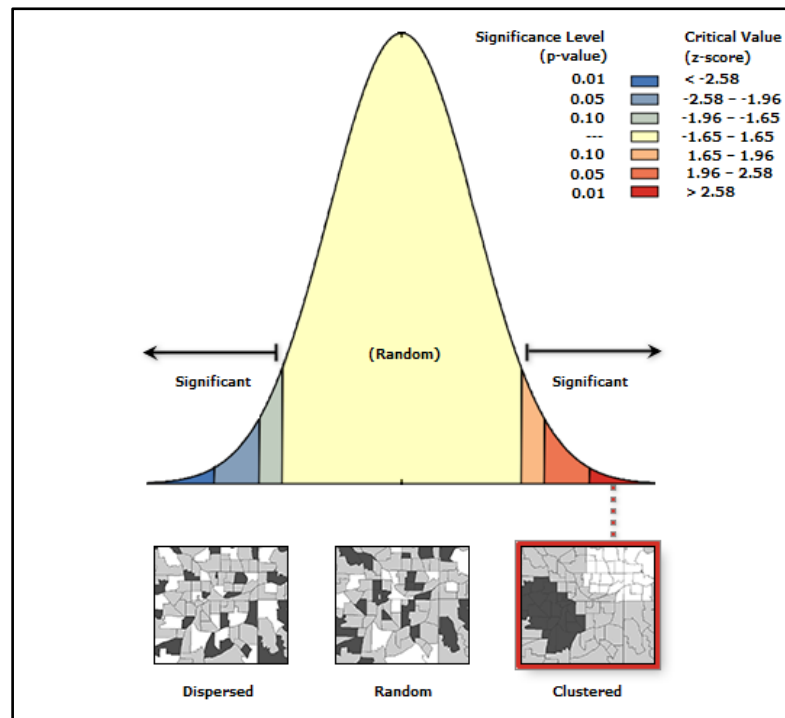


Figure 5.21: Spatial autocorrelation of the address-weighted dasymetric method provided population residuals in the Borough of Kensington and Chelsea.

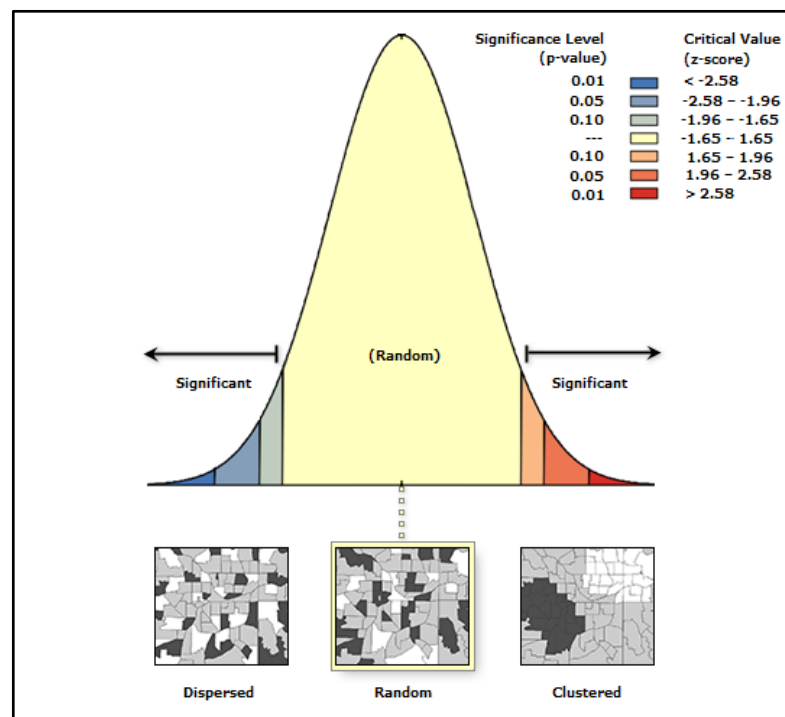


Figure 5.22: Spatial autocorrelation of the address-weighted method using occupancy data obtained population residuals in the Borough of Kensington and Chelsea.

5.4 Summary

The same areal interpolation techniques – the area-weighted, land cover based, building volume based, the address-weighted method, occupancy information, and address point data based interpolation methods – were conducted in the Borough of Kensington and Chelsea to estimate population totals at output areas level. This was done to find the most appropriate estimation method based on data availability. The estimation results show that the assumption of the population estimation model makes different population estimates in target areas. The estimations confirm that the external data resources help to estimate population totals when they contain information to determine the location of residential uses. Thus, the performance of the areal interpolation methods was tested by applying these interpolation methods in the Borough of Kensington and Chelsea.

The refined address-weighted and volumetric dasymetric interpolation methods perform better than the binary dasymetric and areal weighting interpolation approaches. Theoretically, national housing unit address points datasets supply accurate locations of individual housing units and every non-residential unit within an authority. It can be considered that the density of housing unit address points is closely related to population density (Zandbergen, 2011). Building address point datasets have potential to define residential housing units, and building volume information may determine different housing height and characteristics. However, the binary dasymetric mapping model using land use data as control variables results in less accurate dasymetric methods than the building volume based estimation and the address-weighted estimation techniques. When these techniques are employed to any other areas, different estimation results may likely emerge. However, this dasymetric mapping method can be applied to other intended areas where ancillary inputs are available. The address-weighted method was employed by using the ancillary data of residential address points, occupancy information and average housing unit in order to select the most convenient dasymetric method in small-area population estimates. Dasymetrically estimated population totals generated with different ancillary information were compared. The estimated population totals by the address-weighted were most close to the actual population figures than the population estimates obtained from the other areal interpolation model.

To perform a statistical relationship between estimated and actual population totals, regression analysis was applied to define the correlation coefficients. For a validation and accuracy assessment, the estimation results are required to be compared with known census population counts for the target areas (output areas). Due to data availability of the actual population of output areas it was possible to validate results. Higher accuracy resulted when building address points and occupancy information was integrated into the address-weighted dasymetric model based on the RMSE. Although dasymetric interpolation techniques produce better estimation results, the accuracy of population distribution varies with the used additional variables. In this case, the occupancy of building information produced the least error at output area level population estimates. However, the best estimated population values were obtained by making use of the address-weighted dasymetric approach, where R^2 values are greater than 0.97.

The different assumptions for the dasymetric mapping models generate different population estimate surfaces. Differences in estimation models show that the prediction results fully rely on the applied interpolation approach and the used control variables. This can be seen from the residual maps of estimated population. The address-weighted and volumetric dasymetric interpolation methods perform better than other used areal interpolation approaches. Theoretically, national housing unit address point datasets supply the accurate location of individual housing units and every non-residential unit within an authority. It can be recognised that the density of housing unit address points is closely related to population density (Zandbergen, 2011). The building address point datasets also have potential to define residential housing units and building volume information to determine different housing height and characteristics. The comparison of dasymetric areal interpolation techniques shows that volume and the number of housing units is a more stable population indicator. On the contrary, a simple areal weighting interpolation method gives the worst population interpolation results. This technique does not use any ancillary data besides source and target units. Therefore, population may distribute to non-residential land use area by applying this approach. When additional datasets do not exist, a simple areal interpolation method may be used to disaggregate population counts to smaller geographies.

CHAPTER 6

Discussion

6.1 Introduction

As discussed in the literature review, various forms of areal interpolation approaches have been attempted in recent times to disaggregate population totals from source zones to an intended spatial level of target zones (Alahmadi et al., 2013, Azar et al., 2013, Bentley et al., 2013, Langford, 2013, Upegui and Viel, 2012, Kim and Yao, 2010, Su et al., 2010, Briggs et al., 2007, Langford, 2007, Harvey, 2002, Fisher and Langford, 1996, Xie, 1995, Wright, 1936). Each interpolation technique makes use of different external input data in order to disaggregate population counts. To date, land cover map, aerial photograph and cadastral datasets have been used as ancillary inputs in interpolation methodology. Hitherto, researchers have very recently used residential building address points and residential building volume information for generating population estimates with the availability of building address points (Sridharan and Qiu, 2013, Zandbergen, 2011, Tapp, 2010). The study has focused on using novel ancillary datasets in spatial population distribution models to improve small-area population estimates. In addressing the aim of the study, a number of research questions that were raised by the authors of underpinning research are advanced in this study, they were:

1. Do high resolution aerial photography-derived land use/cover datasets and LiDAR-derived building volumes used as external datasets in binary dasymetric mapping and volumetric estimation methods increase the accuracy of population estimates?
2. How well do the datasets of residential housing units perform as control variables in the process of dasymetric mapping in order to obtain small area population estimates?
3. In the address-weighted dasymetric model, how much does the availability of occupied housing unit's datasets improve the estimate of population?

4. How well do the models of areal interpolation with innovative ancillary datasets perform in high density and low density areas of population and what are the limiting factors in the approaches developed?

The research questions were addressed by applying spatial disaggregation models and the results of population estimates obtained by using areal interpolation approaches were provided in Chapter 4. The generated population estimates were compared with census reported actual population values at output area level and finest geographical output in census hierarchy, in order to determine the closeness of estimated population totals to the known population values. These interpolation methodologies were applied to London in order to observe the performance of these used methods in different urban environments and the results of the estimation were presented in Chapter 5. This study also attempted to determine the effect of ancillary data inputs in order to obtain population estimates in areas where population density and the structure of settlements are different.

This chapter provides an evaluation of the population estimation results. These results are discussed in the following sections. Section 6.2 discusses the disaggregation results of existing areal interpolation techniques and compared the estimation results for both case areas. Section 6.3 provides the estimation results of the novel dasymetric mapping method and compared their results in the study regions. Finally, Section 6.4 provides a summary of the areal interpolation methods adopted and summarises the findings in this chapter.

6.2 Discussion of the Results of Existing Approaches

The results reported here make use of the 2011 United Kingdom Census demographic data to spatially disaggregate population information from census wards to output areas. This research found that the aggregated population totals of source zones can be distributed to smaller target zones by using population disaggregation models. The process of dasymetric mapping models demonstrates that the external input data of disaggregation frameworks aid in disaggregating population totals to smaller spatial units. The proposed dasymetric approaches perform mostly better than the areal weighting interpolation method. This is because dasymetric mapping processes use ancillary data, while the area-weighted method distributes population totals everywhere in source zones within the study region. In reality there are residential and non-

residential uses and the area-weighted method does not achieve accurate population distribution by disaggregating population counts to every location. The review of literature in Section 2.5 has indicated that previous researchers discussed the importance of dasymetric mapping methodology for obtaining population estimates in intended spatial units (Bajat et al., 2013, Petrov, 2012, Brinegar and Popick, 2010, Briggs et al., 2007, Bielecka, 2005, Chen, 2002). The interpolation methods applied in this thesis obtained population estimates similar to those previously used disaggregation models (Mennis, 2009, Langford, 2007, Eicher and Brewer, 2001). Furthermore, using high resolution remotely sensed products, this research achieved much better estimation results compared to the area-weighted predicted population totals.

The effect of external input data for the spatial disaggregation process was examined in both study regions. The impact of aerial photographs as ancillary data in the dasymetric mapping process was evaluated. Classified remotely sensed data were used as control variables to distribute population totals in the binary dasymetric mapping process used in past research (see chapter 2 and 3). The image classification results demonstrate that high resolution imagery data are relatively accurate in discriminating built-up areas from non-built-up land cover. The disadvantages of high resolution imagery data are that they require more computer storage and may be time consuming in the process of image classification compared to low resolution multispectral imagery produces based on the size of research area.

The accuracy measurement of the disaggregation models used are summarised in Table 6.1 for the City of Leicester and Table 6.2 for the Borough of Kensington and Chelsea by disaggregating from census wards to output areas. The accuracy is measured using root mean squared error (RMSE), relative error (RE) and percentage error (PE). These accuracy measurements allow the comparison of population estimation results. The area-weighted method yielded the least accurate population estimates with a RMSE value of 9.123 and with an RE score of 0.41 for the City of Leicester and with an RMSE of 4.510 and an RE of 0.40 for the Borough of Kensington and Chelsea. These measurements indicate that the correlation is relatively low between the predicted and the actual population counts in both study regions. The interpolation method showed substantial improvement in using the binary dasymetric method compared to the method using only zonal boundaries in both study regions. With the binary masks of land cover data as control variables, an RMSE score of 0.769 and an RE value of 0.08 are obtained

for Leicester City and an RMSE of 0.899 and an RE of 0.12 are achieved for the Borough of Kensington and Chelsea. Furthermore, the mean estimation error of output areas were quantified for each deployed interpolation technique (see Figure 6.1). This figure shows that the estimation error of population was largely reduced by using the control variables of aerial imagery in the binary mapping. In various succeeding studies, similar results on spatially distributed population estimates were obtained by showing a strong relationship between the binary mask of land cover data and the population distribution (Mennis, 2009, Eicher and Brewer, 2001, Fisher and Langford, 1996). The performance tests show that the estimated values are relatively close to actual population counts using the external data inputs of remote sensing imagery.

Table 6.1: The measurements of accuracy for population disaggregation results from census wards to output areas for the City of Leicester

Interpolation Techniques	Regression Coefficients	R squared	RMSE	RE	PE
Area-weighted using boundaries of source and target zones	0.851	0.116	9.123	0.4101	41.01%
Binary dasymetric using classified land cover form aerial photography	0.922	0.780	0.642	0.0825	8.25%
Volumetric estimation using residential building volumes obtained from LiDAR data	0.901	0.735	0.769	0.0841	8.41%
The address-weighted using building address point data from OS address layer 2 datasets only	0.981	0.937	0.160	0.0407	4.07%
The address-weighted using building address point data from OS address layer 2 datasets and occupancy information from ONS	0.999	0.998	0.064	0.0108	1.08%

Note: Mean population of target zones is 340 for Leicester.

Table 6.2: The measurements of accuracy for population disaggregation results from census wards to output areas for the Borough of Kensington and Chelsea

Interpolation Techniques	Regression Coefficients	R squared	RMSE	RE	PE
Area-weighted using boundaries of source and target zones	1.041	0.175	4.510	0.4054	40.54 %
Binary dasymetric using classified land cover from aerial photography	1.048	0.682	0.899	0.1260	12.60 %
Volumetric estimation using residential building volumes obtained from LiDAR data	0.928	0.818	0.746	0.0773	7.73%
The address-weighted using building address point data from OS address layer 2 datasets only	1.056	0.847	0.796	0.1105	11.05 %
The address-weighted using building address point data from OS address layer 2 datasets and occupancy information from ONS	1.002	0.997	0.238	0.0121	1.21%

Note: Mean population of target zones is 252 for Kensington and Chelsea.



Figure 6.1: The histogram of errors of estimated population for both study areas. a) Areal weighting, b) the binary mapping, c) the volumetric estimation, d) the address-weighted, and e) the address-weighted interpolation method with occupancy data inputs.

The date of the ancillary imagery data may be the main error source that has affected the results of the population estimates of the study. Aerial images were taken by OS a year before the conduct of the UK 2011 census. However, even as the year of the imagery data is in 2010, it may not affect the estimation results. The binary method also has some limitations. The first inevitable shortcoming of the land cover based mapping method is misclassified pixels of land cover classes based on the decision of analysts in the classification process. The second limitation is the difficulty of differentiation of high-rise and low-rise building units. The Borough of Kensington and Chelsea has relatively high-rise buildings, on the other hand Leicester has a few high-rise building blocks. This limitation has been reduced by using the LiDAR-derived building height information as ancillary data for the volumetric estimation model. The third challenging issue is that built-up areas may consist of institutional, commercial and industrial building structures. The land cover classification based on the population disaggregation method distributes population counts by considering built-up areas as residential units. For both case studies, built-up areas were used as control variables in the binary method to spatially distribute population totals at output areas. The results were validated comparing the predicted values with census reported population values. In conclusion, the different assumptions of population disaggregation models generated different population estimates in the study regions. With the ancillary input data of imagery products, estimation results have considerably improved compared to the interpolation method that uses zonal boundaries only. It has been clearly seen in the comparison of the estimation results that areal weighting distributes population totals to output areas based on the size of target units; while the results of binary mapping uses binary dasymetric zones to distribute population totals.

6.3 Discussion of the Results of Novel Approaches

This section discusses the results of novel dasymetric mapping approaches on spatially disaggregated population totals. The results of these novel approaches have been compared in terms of estimation differences. The main finding is that population totals can be disaggregated from larger source zones to smaller target units using high resolution external data inputs that constitute spatial location of population. The spatial disaggregation models were used in order to evaluate the performance of ancillary data inputs for distributing aggregated population totals. The benefit and superiority of using control variables was explained in the preceding successful studies (Bentley et al., 2013,

Langford, 2013, Leyk et al., 2013b, Sridharan and Qiu, 2013, Zandbergen, 2011, Lwin and Murayama, 2010). The dasymetric mapping techniques performed much better than the areal weighting interpolation by using additional information. The availability of actual population totals makes it easier to validate the accuracy of estimation results for each proposed areal interpolation method, and it helps in comparing the result of estimations with each other.

The literature has shown that the data of building address points have been used very recently in dasymetric methodology (Zandbergen, 2011, Tapp, 2010). Furthermore, LiDAR-derived building height information has been utilised several times in the methodology of population estimation (Lu et al., 2011b, Dong et al., 2010, Lu et al., 2010, Silván-Cárdenas et al., 2010). Similarly and more recently, Sridharan and Qiu (2013) and Lwin and Murayama (2010) used the building volume data as ancillary data in order to disaggregate population totals to smaller spatial units. This volumetric estimation and the address-weighted dasymetric process obtained highly good estimation results similar to the studies of Sridharan and Qiu (2013) and Zandbergen (2011). At this resolution of interpolation process, the address-weighted interpolation technique shows an improvement over the current areal interpolation approaches with sufficient accuracy. The dasymetric methodologies obtained very close estimation results to census reported actual population totals, by using high resolution building attributes and occupancy information.

The work reported in this thesis discussed the impact of the used ancillary input datasets to estimate population counts using different interpolation methods. First of all, the effect of building height information on spatial disaggregation models has been evaluated. The literature review of volumetric estimation demonstrates that LiDAR-derived height information allows the differentiation of high-rise and low-rise buildings to distribute population counts based on the weighting factor of each building. Compared to remote sensing imagery, the weakness of LiDAR data is that height data are expensive, require large computer storage and require more time to make the data useable as control variables in the dasymetric process. Secondly, the effect of building address point data on the estimated population totals has been tested. The results illustrate that address point performed well, giving detailed information of residential housing units when used as ancillary data in the dasymetric process. This accuracy has not been achieved using remote sensing imagery data. This is because it is difficult to

differentiate low-rise buildings from high-rise buildings. Although the use of building address point data as control variables for the dasymetric mapping process may require additional occupancy information for more accuracy on spatially distributed population estimates.

The performance of the areal interpolation processes was measured by comparing the spatially distributed population estimates with census reported actual value of output areas. Table 6.1 and Table 6.2 above have shown the results of the accuracy assessment of interpolation methods for the City of Leicester and the Borough of Kensington and Chelsea, respectively. These measurements aim to observe the performance of each method. Considering the disaggregation of population totals from census wards to output areas, estimation results show that the address-weighted process with occupancy information performs most well with an RMSE value of 0.064 and an RE score of 0.010 for the City of Leicester and an RMSE of 0.238 and an RE of 0.012 for the Borough of Kensington and Chelsea, by yielding the closest estimation values to actual values. With the availability of occupancy information, the used disaggregation method showed a considerable improvement over the other interpolation processes used in the current study. The results show the most accurate population estimates were obtained when the occupancy information was used in the areal interpolation process. Secondly, the address-weighted process with only building address points yielded good estimation results with an RMSE value of 0.160 and an RE score of 0.04 for the City of Leicester and an RMSE of 0.796 with an RE of 0.11 for the Borough of Kensington and Chelsea. These results are similar to the estimation results of Zandbergen (2011) who used high resolution address points in the dasymetric mapping method. The success of the address-weighted method can be attributed to the ability of the address layer data to differentiate residential building address points from non-residential building structures.

The second most important reason is that the residential address point's data have defined the location of residential houses where people reside. The residential building data are much more detailed than residential land use and road networks in identifying housing units where people actually reside (Qiu et al., 2010). Zandbergen (2011) used residential address points as ancillary data without considering occupied housing structures, however, the current study used residential housing units and the integration of occupancy information and address points considering non-occupied houses. Thirdly, with residential building volumes used as external data input in the dasymetric method,

an RMSE score of 0.769 with an RE value of 0.08 for the City of Leicester and an RMSE of 0.746 with an RE value of 0.07 for the Borough of Kensington and Chelsea are obtained, performing relatively well in population estimates. This result is similar to what Sridharan and Qiu (2013) obtained by using the volumetric estimation process in the case of the Borough of Kensington and Chelsea. When integrated with LiDAR derived building height information and building footprint information, the dasymetric model achieved a more accurate estimation of population totals. These results suggest that if the height data are available for the intended study area, it may have been used as external input on the spatially population disaggregation process. Additionally, the mean estimation error of output areas was detailed for each used interpolation technique in Figure 6.1. This table shows that the smallest estimation error was yielded by the address-weighted disaggregation model. Finally, as stated by Langford (2013), the selection of external inputs in dasymetric mapping provides a considerable difference in the results of population disaggregation process.

The main concern is that LiDAR derived digital elevation models for the City of Leicester were compiled in 2002 and the ones for the Borough of Kensington and Chelsea were archived in 2012. This means that the old dated external data inputs may account for potential error and consequently affect population estimation results. When using only the building height information, it is difficult to differentiate residential building units from non-residential manmade structures. This is another major error source in the dasymetric mapping process. OS VectorMap District data were used to identify residential buildings. The 2012 dated 'OS Address layer 2' data were obtained for both case areas. The address point data has building usage information and identifies the residential housing units. However, non-occupied housing units are the potential error sources. This has been reduced by using building occupancy information. The address-weighted method was used twice. At first, building address points were used alone as a control variable, followed by the integration of address points and occupancy information used in the disaggregation process. With these two processes, substantial improvement was brought to the estimation result.

Many published papers showed that different forms of dasymetric mapping methods have been used in population disaggregation and population estimations with the availability of external data inputs (see Section 2.5). The choice of an interpolation algorithm for population estimation largely depends on the availability of different types

of ancillary information of the intended study area. If the address point data are available, the address-weighted data can be preferable method because of the simplicity of the methodology. The other reason is that the dasymetric methods consider that people live in build-up areas, while the address-weighted method considers that people live in residential units within the build-up areas. This is the main superiority of the address-weighted method over dasymetric mapping techniques. In this study, the address-weighted dasymetric mapping model integrated with residential building address point, and occupancy information indicated significant development over existing interpolation techniques in both case sites. The superiority of this approach depends heavily on accurate average household sizes and the number of residential housing units. The complementary findings of the current study showed that the core contribution of using building attribute information as an ancillary data in dasymetric population mapping model improves dasymetric estimation accuracy with very reasonable estimation results, at least in the UK context, due to the availability of national address point data of each addressable unit. This is because, ‘the address layer 2’ of OS data has the ability to discriminate residential and non-residential building structures at UK output area level, which is the finest spatial unit. Based on these findings, this research suggests that precise population estimates may be obtained in different geographical scales, particularly in areas where building attributes and occupancy data are available.

6.4 Summary

Five interpolation approaches were used in each study region to disaggregate population totals from census wards to output areas based on the UK census hierarchy. Owing to the different measurement functions of interpolation methods, they obtained different estimation results for output areas. The 2011 census reported population counts for these target zones allows validating the accuracy of estimated population values for both study regions. Furthermore, the actual values permit comparison with the spatial distribution of estimated error patterns of each disaggregation model. By taking the advantage of the actual population, the differences of estimation results based on areal interpolation models were discussed in this thesis. Additionally, the effects of utilised ancillary input data as control variables in the disaggregation process were discussed and stressed. The discussion chapter clearly shows that the results of areal interpolation techniques may have advantages and disadvantages based on the differences in

population density and the spatial characteristics of environments. As stated by Langford (2013), their performance may alter with specific conditions.

Important outcomes have been obtained when creating dasymetric population surfaces and they were pointed out in the results chapters (Chapters 4 and 5) and discussed in this chapter. The most challenging issue of the dasymetric mapping method is to determine the residential uses in source areas. For the purpose of estimation population totals, the address-weighted dasymetric method has proved to be the most accurate of all interpolation approaches used. Additionally, neither the volumetric estimation method nor the existing binary methods have obtained more accurate results than the area-weighted interpolation techniques. The superior performance of the address-weighted model mostly results from using address points datasets to define residential housing units and the additional information of occupancy rate assists to spatially distribute population totals to occupied housing units. This may be understood by comparing the accuracy assessment of all spatially disaggregating models. The dasymetric model using the point data of residential buildings shows significant improvement over existing interpolation approaches in both case sites. The superiority of this process needs to be further examined by applying the same technique to the areas where address point data is available. The advantages and disadvantages of these interpolation models and recommendations for future work will be concluded in the next chapter (Chapter 7).

CHAPTER 7

Conclusion

This chapter brings together results about improving population estimation models. The aim of this research was to develop novel ancillary datasets for the disaggregation of population totals to better estimate population. The research has used volumetric estimation and the address-weighted dasymetric models to obtain spatially distributed population surfaces to generate small-area population totals. LiDAR-derived building volumes and OS obtained building address point datasets and occupancy information has been used as sources of control variables in the process of interpolations. These estimation models and the data of ancillary inputs have been explained in the research methodology. The research has presented and evaluated the viability of using building volumes and building address points products in the dasymetric population distribution model. The proposed interpolation methods were implemented in two local authority areas in the UK (the City of Leicester and the Royal Borough of Kensington and Chelsea). For validation of spatially distributed population surfaces, the areal interpolation provided estimation results were compared with census reported actual population counts of output areas. The address-weighted dasymetric mapping model using residential building information indicates significant improvement over existing interpolation techniques in the case sites. However, the superiority of this approach is completely based on accurate average household sizes and the number of occupied residential housing units.

As indicated in previous studies and supported by the current work, a dasymetric framework can be used in population estimation and in creating the distribution of population maps on the intended scale. The results of tested interpolation techniques show that disaggregation methodology provides the most accurate estimation results with the help of Ordnance Survey based data of building address points and occupancy information, as control variables. The two improved methods have shown a higher performance in dasymetric mapping. The address-weighted model performs consistently better than other applied interpolation methods in both case areas. The tests have also

demonstrated that the application of occupancy information rate value can be used as ancillary data to substantially improve the accuracy of population estimation results. This work has refined and presented the Ordnance Survey data based population estimation model within the dasymetric mapping techniques. The study has shown that ancillary data inputs and the assumptions of different areal interpolation methods have impacts on spatially distributed population surfaces. With the integration of LiDAR derived building height information and building footprint information, the dasymetric model achieved a more accurate estimation of population totals. As indicated in the results and discussion chapters (4, 5, and 6), intelligent dasymetric mapping models have produced promising results with better accuracy than those previously adopted interpolation techniques. These refined estimation models have been tested for several densely and sparsely populated settlements; thus, an address-weighted framework based on OS data and ONS product has great potential to estimate population. The results also show that an address-weighted dasymetric method performs best when occupancy information is utilised. The accuracy of population estimates appears to be mainly influenced by the type and quality of the ancillary datasets, and the interpolation method adopted. Ordnance Survey acquired building address point datasets have the potential to be used with the proposed method to provide larger scale population estimates.

7.1 Key Findings of the Disaggregation Process

This thesis uses selected areal interpolation methods in the City of Leicester and the Borough of Kensington and Chelsea in order to address the research questions raised in Chapter 2. This section provides important, key findings based on the research objectives and aims that have emerged in the process of interpolations. The main objective of the study is to find the most appropriate population disaggregation model for obtaining population estimates. The different assumption of each model and the distribution of population densities in different urban environments are significantly important for estimation population totals in small-area level. This is because the used control variables may not work as a good predictor at the same spatial scale. The residential housing units are the finest geographical units where people reside. The complementary findings of the current study demonstrates the core contribution of using building attribute information, as an ancillary data in the dasymetric population mapping model improves dasymetric estimation accuracy. Additionally, inaccuracies in

additional datasets should be investigated. The accuracy of datasets plays key roles for successful population estimates in intended spatial scale.

The actual census population totals for the City of Leicester and the Borough of Kensington and Chelsea released at census ward level were successfully distributed to output areas, which is the finest spatial unit using external data inputs. There is no benefit to using areal weighting interpolation when external inputs data are available to extract non-residential uses and redistribute population totals only to defined residential zones. In both densely populated and sparsely populated areas, the technique generates the least accurate results. This is because the process of area-weighted technique distributes population counts only based on the size of target areas. The binary mapping technique obtains better population estimates of target areas in the City of Leicester compared to the Borough of Kensington and Chelsea. Aerial image classification indicates that Leicester City has more segregated land use than the Borough of Kensington and Chelsea. Digital elevation models show low-rise buildings are dominant in the City of Leicester (see Chapter 3). For this reason, population is most likely distributed more homogeneously compared to London.

Population density may change in residential areas due to the different spatial characteristics of building units. Therefore, aerial photography derived built-up areas may work as a good predictor to generate spatial distributed population surfaces in areas where the patterns of population distribution is uniform. This uniformity is because two-dimensional products consider only the horizontal extent of building structures when disaggregating interested variables over the surface. However, the volumetric estimation model shows clear improvement over the binary mapping method in the Borough of Kensington and Chelsea. The LiDAR-derived building height data were used to calculate building volumes to distribute population counts. The datasets consider the horizontal and vertically stacked structures that have been used as indicator variables to estimate population totals. The study region of the Borough of Kensington and Chelsea, cover both high and low-rise buildings. With the availability of LiDAR data, building volumes can be used to estimate population estimates in areas, such as Kensington and Chelsea.

The other important outcome is that interpolation frameworks provided different predicted population values in different urban areas. In both study regions, the address-

weighted dasymetric method provided different estimation results when only building address points were used as control variables. This is most probably as a result of varying the numbers of non-occupied housing units. Accuracy measurements and scatterplots have shown that the method generates more accurate results in the City of Leicester. The address-weighted approach obtained very similar population estimates by integrating occupancy information into the dasymetric mapping process. Population counts can be located to occupied housing units within the target zones with the availability of accurate occupancy information by extracting non-occupied housing units.

Overall, demographic datasets of national censuses are mostly published based on the census dissemination geographies. Population estimation models are usually needed when the user defined areas are dissimilar to census output units. In the current study, the application of the different dasymetric mapping models are based on remotely sensed products and Ordnance Survey based datasets. The generated results were evaluated using statistical and visual approaches. With the availability of actual census population figures, the results were validated and compared with each single interpolation technique. The models achieved reasonable results; the study demonstrates the use of dasymetric models to create the representation of population totals. The two improved interpolation techniques presented have proven to show better performance in dasymetric mapping. Through integration with LiDAR derived building height information and building footprint information, a dasymetric model achieved a more accurate estimation of population totals. Also, by applying dasymetric population mapping techniques to the City of Leicester, this research demonstrates that address point datasets and building volume information are able to improve the quality of population estimation significantly. OS-derived address point datasets have the potential to identify the smallest residential units. Therefore, dasymetric methods can be used to assign population to each residential housing unit.

By using dasymetric methods (the address-weighted and volumetric estimation techniques), realistic population estimation can be calculated and distribution maps produced due to the strength of the dasymetric process in the differentiation of residential units and non-residential structures. It is understood that the quality of chosen datasets is important to show unpopulated areas. This study shows that the quality of chosen ancillary datasets and applied methodology influence the performance

of dasymetric areal interpolation to distribute population counts. Thus, this research contributes to the literature on dasymetric areal interpolation methods for population estimation. The estimate of population results indicates that the address-weighted dasymetric interpolation method is the most appropriate approach for spatially disaggregating the known population information in both study areas. The increasing availability of ancillary datasets provides significant potential enlargement to the adaptation of dasymetric population interpolation tools (Langford, 2013, Langford, 2005, Martin et al., 2000).

7.2 Research Question Revisited

The main aim of this research, highlighted in Chapter 1, is to use novel ancillary datasets as control variables in dasymetric modelling to derive population estimates within the UK. The key findings have been summarised in the preceding section. Now, the research questions are revisited to assess whether the issues raised have been successfully addressed.

1. Do high resolution aerial photography-derived land cover/use datasets and LiDAR-derived building volumes used as external dataset in binary dasymetric mapping and volumetric estimation methods increase the accuracy of population estimates?

The study has presented binary and volumetric models to spatially disaggregate population totals from census ward to output areas within the UK. Two study areas have been chosen for the purpose of examining these disaggregation approaches due to their differences in the density of building block and population. The classified aerial images derived binary dasymetric zones and the building volumes have been used to spatially disaggregate population totals in the given study areas. The classification of aerial images and total volumes of building blocks directly affected estimation results. Population totals have been spatially distributed, assuming a simple linear relationship between population counts, building volumes, and the total area of populated binary zones. On the one hand, the binary dasymetric mapping method generated slightly better estimation results in Leicester, and on the other hand, the volumetric estimation method generated better population estimations in the Borough of Kensington and Chelsea. However, population density may change over an administrative area based on

the density of residential building blocks, and mixed use of building blocks may affect estimation outcomes. In Kensington and Chelsea, this factor would most likely affect the results of the dasymetric mapping model. This question has been addressed by obtaining reasonable population estimated in both study areas.

2. How well do the datasets of residential housing units perform as control variables in the process of dasymetric mapping in order to obtain small area population estimates?

The research has presented an address based estimation model in order to estimate small-area population totals within the UK context. Residential housing units have been used as external variables in the process of dasymetric population mapping. The residential housing units are the finest geographical phenomena that indicate where people actually reside. The estimation models, using the ancillary datasets of accurate residential housing blocks, have generated more precise population estimates than the binary and volumetric approaches. However, the datasets of address points as a control variable are more sophisticated in population estimates, but unoccupied households are the main source of error in housing unit based population estimates. Again, this research question was successfully examined in the two study areas.

3. In the address-weighted dasymetric model, how much does the availability of occupied housing unit's datasets improve the estimate of population?

The study has investigated the potential of OS datasets and local city council derived occupancy data (the combination of building address points and occupancy information) as control variables in the address-weighted interpolation method. These datasets consist of attribute information to refine the occupied residential housing units within each output area. Hence, the total number of occupied households have been multiplied by the average household size for the purpose of estimation population totals. Binary and volumetric estimation models have produced price estimation results, but with the advent of occupancy datasets, the address weighted approach provided the best estimation results with great accuracy for both study areas.

4. How well do the models of areal interpolation with innovative ancillary datasets perform in high density and low density areas of population and what are the limiting factors of the approaches developed?

The developed dasymetric models were applied in both study areas for the purpose of comparison of interpolation models. As can be seen from the interpolation results in Chapter 4 and Chapter 5, the areal interpolation method generates similar results. On the one hand, the binary dasymetric method generates slightly better estimation results for the City of Leicester than Kensington and Chelsea. This could be due to the mixed use of building units which are more dominant in Kensington and Chelsea. On the other hand, the volumetric estimation model produced considerably more precise estimation results for Kensington and Chelsea than Leicester City due to the following reasons. First of all, the volumetric method establishes a linear relationship between population counts and building volumes: low-rise and high-rise housing units are mixed in Leicester City and high-rise building units are dominant in Kensington and Chelsea. Secondly, in bigger housing units, fewer populations can reside in the suburban areas of Leicester City. This is mostly related with the income of people, since higher income people live in bigger building complexes. Thirdly, white Britons are a majority in Kensington and Chelsea, while they are a minority in the City of Leicester (NeSS). The differences (variations) in ethnic origin probably account for population density changes in target areas in Leicester. All these differences directly affect the performance of volumetric estimation. When occupancy information is used as an additional dataset, the address-weighted method generated very close population counts to actual population figures in both study areas. Thus, this research has met the objectives of the study successfully.

7.3 Contributions of this Research

This research has added to the literature on improving the accuracy of population estimation models, shedding light on improvements made in dasymetric areal interpolation methods for population estimation. Two major contributions can be identified. The first main contribution of this study is that high resolution ancillary datasets assist in obtaining precise population estimates at small-area spatial resolution. The data of building address points and building volumes have advantages over other remote sensing datasets for use as a source of ancillary data in the dasymetric mapping process. With the help of these datasets, residential housing units can be defined within target zones. LiDAR-derived building volumes have been applied as control variables in the population disaggregation process. Secondly, the research evaluates the benefit of

building usage and building occupancy information as control variables in the dasymetric mapping process. In this implementation, a dasymetric mapping model provides significantly better estimation results by integrating occupancy information with building address points. Results obtained from this integration have proven that census estimation results can be very close to actual population figures with high accuracy.

7.4 Limitations of the Study

The implemented interpolation methodologies do have several limitations. The major shortcoming of the research is that areal interpolation methods with various ancillary datasets do not provide standard accuracy over the case areas. The spatial characteristics and different population density of study areas may have an impact on spatially distributed population surfaces. Secondly, there is the generalisation problem inherent in using limited case studies. That means, based on the conclusions drawn from two case areas, it may be hard to accept that the dasymetric mapping techniques performed are a good model to estimate population totals. The volumetric estimation and the address-weighted dasymetric models may yield less accurate estimation results in other urban environments. The third drawback of the interpolation method is that spatial disaggregation methods have been implemented in both case areas where validation data is available. The performances of these models need to be examined in different regions where validation data is not available. A fourth limitation of the dasymetric techniques reported here is that population estimates depend heavily on the number of residential housing units within the target areas which can be highly problematic where the national datasets of building address points do not exist. A fifth and final shortcoming is that the author has used census population counts. The censuses record population counts based on where people live at night. Census population counts are associated with only residential address locations (Martin, 2011, Martin et al., 2009), the methods estimate night time population and do not consider daytime population. These drawbacks are the starting point of future work.

7.5 Recommendations for Future Work

The following are suggested areas for extension of this research in the future. First, the method developed in this study used several datasets including aerial photography, the

boundaries of census output areas, building volumes and building address data in order to disaggregate the aggregated population figures. These ancillary variables may be inaccurate or errors can be introduced during the process of determining the residential building locations. These errors may affect the performance of the proposed dasymetric interpolation technique. For example, mislabelling of land cover and building usage types is an area that requires evaluation in future research. The performance of innovative dasymetric mapping methods will be further evaluated using a different resolution of ancillary datasets in different regions (more complex and larger areas) including the superiority of disaggregation methods.

Secondly, population density is spatio-temporal variable. Therefore, daytime and night-time population information may be a requirement for some applications, such as emergency management and market delineation and transportation construction and so on. With an expanded availability of high resolution residential building information, school point data, bus stop information and detailed land use information, areal interpolation models can be employed to estimate both night-time and daytime population distribution. Ordnance Survey acquired building address point datasets have the potential to be used with the proposed method to provide larger scale population estimates within the UK. The applicability of these building address points, as control variables in the process of population disaggregation, will be examined in areas where building usage information is available. Lastly, the feasibility of these implemented disaggregation models will be tested for defining optimal facility locations for public and government facilities where population information is not available at the intended spatial scale.

References

- ALAHMADI, M., ATKINSON, P. & MARTIN, D. 2013. Estimating the spatial distribution of the population of Riyadh, Saudi Arabia using remotely sensed built land cover and height data. *Computers, Environment and Urban Systems*, 41, 167-176.
- ALSHWESH, I. O. 2014. *GIS-based interaction of location allocation models with areal interpolation techniques*. PhD, University of Leicester.
- ANDERSON, S. J., TUTTLE, B. T., POWELL, R. L. & SUTTON, P. C. 2010. Characterizing relationships between population density and nighttime imagery for Denver, Colorado: issues of scale and representation. *International Journal of Remote Sensing*, 31, 5733-5746.
- ANSELIN, L., SYABRI, I. & KHO, Y. 2010. GeoDa: an introduction to spatial data analysis. In: FISCHER, M. M. & GETIS, A. (eds.) *Handbook of Applied Spatial Analysis*. Springer Berlin Heidelberg.
- ANTONARAKIS, A. S., RICHARDS, K. S. & BRASINGTON, J. 2008. Object-based land cover classification using airborne LiDAR. *Remote Sensing of Environment*, 112, 2988-2998.
- AZAR, D., ENGSTROM, R., GRAESSER, J. & COMENETZ, J. 2013. Generation of fine-scale population layers using multi-resolution satellite imagery and geospatial data. *Remote Sensing of Environment*, 130, 219-232.
- AZAR, D., GRAESSER, J., ENGSTROM, R., COMENETZ, J., LEDDY, R. M., SCHECHTMAN, N. G. & ANDREWS, T. 2010. Spatial refinement of census population distribution using remotely sensed estimates of impervious surfaces in Haiti. *International Journal of Remote Sensing*, 31, 5635-5655.
- BAJAT, B., KRUNIĆ, N., PETROVIĆ, M. S. & KILIBARDA, M. 2013. Dasymetric modelling of population dynamics in urban areas *Geodetski Vestnik* 57, 777-792.
- BENTLEY, G. C., CROMLEY, R. G. & ATKINSON-PALOMBO, C. 2013. The Network Interpolation of Population for Flow Modeling Using Dasymetric Mapping. *Geographical Analysis*, 45, 307-323.
- BIELECKA, E. A dasymetric population density map of Poland. In *International Cartographic Conference*, 2005.
- BLOOM, L. M., PEDLER, P. J. & WRAGG, G. E. 1996. Implementation of Enhanced areal interpolation using mapinfo. *Computers and Geosciences*, 22, 459-466.

- BRACKEN, I. & MARTIN, D. 1989. The generation of spatial population distributions from census centroid data. *Environment and Planning A*, 21, 537–543.
- BRIGGS, D. J., GULLIVER, J., FECHT, D. & VIENNEAU, D. M. 2007. Dasymetric modelling of small-area population distribution using land cover and light emissions data. *Remote Sensing of Environment*, 108, 451-466.
- BRINEGAR, S. J. & POPICK, S. J. 2010. A Comparative Analysis of Small Area Population Estimation Methods. *Cartography and Geographic Information Science*, 37, 273-284.
- CAI, Q., GERARD, R., BHADURI, B., BRIGHT, E. & COLEMAN, P. 2006. Estimating Small-Area Populations by Age and Sex Using Spatial Interpolation and Statistical Inference Methods. *Transactions in GIS*, 10, 577–598.
- CAMPBELL, J. B. 2002. *Introduction to Remote Sensing*, London, Taylor & Francis.
- CHEN, K. 2002. An approach to linking remotely sensed data and areal census data. *International Journal of Remote Sensing*, 23, 37-48.
- CHEN, Y., SU, W., LI, J. & SUN, Z. 2009. Hierarchical object oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research*, 43, 1101-1110.
- COCKINGS, S., FISHER, F. P. & LANGFORD, M. 1997. Parameterization and visualization of the errors in areal interpolation. *Geographical Analysis*, 29, 314-328.
- COCKINGS, S., HARFOOT, A. & HORNBY, D. 2011. Towards 2011 output geographies: Exploring the need for, and challenges involved in, maintenance of the 2001 output geographies. *Population Trends*, 138, 38 - 49.
- COCKINGS, S., HARFOOT, A., MARTIN, D. & HORNBY, D. 2013. Getting the foundations right: spatial building blocks for official population statistics. *Environment and Planning A*, 45, 1403-1420.
- COMBER, A., PROCTOR, C. & ANTHONY, S. 2008. The Creation of a National Agricultural Land Use Dataset: Combining Pycnophylactic Interpolation with Dasymetric Mapping Techniques. *Transactions in GIS*, 12, 775-791.
- COOK, L. 2004. The Quality and Qualities of Population Statistics, and the Place of the Census. *Area*, 36, 111-123.
- CROMLEY, R. G., EBENSTEIN, A. Y. & HANINK, D. M. 2009. Estimating Components of Population Change from Census Data for Incongruent Spatial/Temporal Units and Attributes. *Spatial Science* 54, 89-100.

- DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1-38.
- DENG, C., WU, C. & WANG, L. 2010. Improving the housing-unit method for small-area population estimation using remote-sensing and GIS information. *International Journal of Remote Sensing*, 31, 5673-5688.
- DONG, P., RAMESH, S. & NEPALI, A. 2010. Evaluation of small-area population estimation using LiDAR, Landsat TM and parcel data. *International Journal of Remote Sensing*, 31, 5571-5586.
- DU, G., ZHANG, S. & ZHANG, Y. 2006. Character and causes of population distribution in Shenyang city, China. *Chinese Geographical Science*, 16, 217-222.
- EICHER, C. L. & BREWER, C. A. 2001. Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science*, 28 125-138.
- ERENER, A. & DÜZGÜN, H. S. B. Prediction of Population in Urban Areas by Using High Resolution Images. International Conference on Recent Advances in Space Technologies, 2009 Istanbul.
- FISHER, P. F. & LANGFORD, M. 1996. Modeling Sensitivity to Accuracy in Classified Imagery: A Study of Areal Interpolation by Dasymetric Mapping*. *The Professional Geographer*, 48, 299-309.
- FISHER, P. F. & LANGFORD, M. 1995. Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation *Environment and Planning A* 27, 211 – 224
- FLOWERDEW, R. 1991. Spatial data integration. *Geographical information systems*, 1, 375-387.
- FLOWERDEW, R., FENG, Z. & MANLEY, D. 2007. Constructing data zones for Scottish Neighbourhood Statistics. *Computers, Environment and Urban Systems*, 31, 76-90.
- FLOWERDEW, R. & GREEN, M. 1989. Statistical Methods for inference between incompitable zonal systems *In: GOODCHILD, M. F. & GOPAL, S. (eds.) Handing geographical infromation: Methodology and Potential Applications.* New york: Longman.

- FLOWERDEW, R. & GREEN, M. 1991. Data integration: Statistical methods for transferring data between zonal systems. *In*: I. MASSER & M. B. BLAKEMORE (eds.) *In handling geographical information*. Logman Scientific and Technical.
- FLOWERDEW, R. & GREEN, M. 1992. Developments in areal interpolation methods and GIS. *The Annals of Regional Science*, 26, 67-78.
- GALLEGO, J. & PEEDELL, S. 2001. Using CORINE Land Cover to map population density. *Towards agri-environmental indicators*. EEA. Italy.
- GERLAND, P., RAFTERY, A. E., ŠEVČÍKOVÁ, H., LI, N., GU, D., SPOORENBERG, T., ALKEMA, L., FOSDICK, B. K., CHUNN, J., LALIC, N., BAY, G., BUETTNER, T., HEILIG, G. K. & WILMOTH, J. 2014. World population stabilization unlikely this century. *Scienceexpress*.
- GOODCHILD, M. F. 1986. *Spatial autocorrelation*, Geo Books.
- GOODCHILD, M. F., ANSELIN, L. & DEICHMANN, U. 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, 25, 383-397.
- GOODCHILD, M. F. & LAM, N. S.-N. 1980. Areal interpolation; A variant of the traditional spatial problem. *Geo-Processing*, 1, 297-312.
- GREGORY, I. N. 2002. The accuracy of areal interpolation techniques: standardising 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems*, 26 293-314.
- GREGORY, I. N. & ELL, P. S. 2005. Breaking the boundaries: geographical approaches to integrating 200 years of the census. *Royal Statistical Society*, 168, 419-437.
- GREGORY, I. N., MARTI-HENNEBERG, J. & TAPIADOR, F. J. 2010. Modelling long-term pan-European population change from 1870 to 2000 by using geographical information systems. *Journal of Royal Statistics Society*, 173, 31-50.
- GRIFFITH, D. A. 2013. Estimating missing data values for georeferenced Poisson counts. *Geographical Analysis*, 45, 259-284.
- HARDIN, P. J., JACKSON, M. W. & JENSEN, R. R. 2008. Modelling housing unit density from land cover metrics: a Midwestern US example. *Geocarto International*, 23, 393-411.

- HARVEY, J. T. 2002. Estimating census district populations from satellite imagery: Some approaches and limitations. *International Journal of Remote Sensing*, 23, 2071-2095.
- HAWLEY, K. & MOELLERING, H. 2005. A Comparative Analysis of Areal Interpolation Methods. *Cartography and Geographic Information Science*, 32, 411-423.
- HAY, S. I., NOOR, A. M., NELSON, A. & TATEM, A. J. 2005. The accuracy of human population maps for public health application. *Trop Med Int Health*, 10, 1073-86.
- HIGGS, G., FRY, R. & LANGFORD, M. 2012. Investigating the implications of using alternative GIS-based techniques to measure accessibility to open space. *Environment and Planning B: Planning and Design*, 39, 326-343.
- HOLLAND, D. Developing a national height database. Symposium on Geospatial Theory, Processing and Applications, 2002 Ottova.
- HOLT, J. B., LO, C. P. & HODLER, T. W. 2004. Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. *Cartography and Geographic Information Science*, 31, 103-121.
- HORNER, B. L. 2008. Comparison of Population Distribution Models using Areal Interpolation on Data with Incompatible Spatial Zones. *Papers in Resource Analysis*, 11, 1-15.
- KAR, B. & HODGSON, M. E. 2012. A Process Oriented Areal Interpolation Technique: A Coastal County Example. *Cartography and Geographic Information Science*, 39, 3-16.
- KIM, H. & CHOI, J. 2011. A Hybrid Dasymetric Mapping for Population Density Surface using Remote Sensing Data. *Journal of the Korean Geographical Society*, 46, 67-80.
- KIM, H. & YAO, X. 2010. Pycnophylactic interpolation revisited: integration with the dasymetric-mapping method. *International Journal of Remote Sensing*, 31, 5657-5671.
- KITCHIN, R. & TATE, N. J. 2000. *Conducting Research Into Human Geography: Theory, Methodology, and Practice*, London, Prentice Hall.
- LAM, N. S.-N. 1983. Spatial interpolation methods: A review. *The American Cartographer*, 10, 129-149.

- LANGFORD, M. 2005. Dasymetric mapping of industry and commerce across Wales, UK, using geospatial integration of national mapping agency products.
- 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, 161-180.
- LANGFORD, M. 2007. Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems*, 31, 19-32.
- LANGFORD, M. 2013. An Evaluation of Small Area Population Estimation Techniques Using Open Access Ancillary Data. *Geographical Analysis*, 45, 324-344.
- LANGFORD, M. & HARVEY, J. T. 2001. The use of Remotely Sensed Data for Spatial Disaggregation of Published Census Population Counts. *IEEEASPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, 260-264.
- LANGFORD, M. & HIGGS, G. 2006. Measuring Potential Access to Primary Healthcare Services: The Influence of Alternative Spatial Representations of Population. *The Professional Geographer*, 58, 294–306.
- LANGFORD, M., HIGGS, G., RADCLIFFE, J. & WHITE, S. 2008. Urban population distribution models and service accessibility estimation. *Computers, Environment and Urban Systems*, 32, 66-80.
- LANGFORD, M. & UNWIN, D. J. 1994. Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal*, 31, 21-26.
- LEYK, S., BUTTENFIELD, B. P. & NAGLE, N. N. Uncertainty in Demographic Small Area Estimates. Spatial Accuracy Assessment 2010 Leicester, UK.
- LEYK, S., BUTTENFIELD, B. P. & NAGLE, N. N. 2013a. Modeling Ambiguity in Census Microdata Allocations to Improve Demographic Small Area Estimates. *Transactions in GIS*, 17, 406-425.
- LEYK, S., BUTTENFIELD, B. P., NAGLE, N. N. & STUM, A. K. Landcover Data to Derive Related Ancillary Variables for Dasymetric Modeling. AutoCarto 2012 2012 Columbus, Ohio, USA.

- LEYK, S., BUTTENFIELD, B. P., NAGLE, N. N. & STUM, A. K. 2013b. Establishing relationships between parcel data and land cover for demographic small area estimation. *Cartography and Geographic Information Science*, 40, 305-315.
- LEYK, S., NAGLE, N. N. & BUTTENFIELD, B. P. 2013c. Maximum Entropy Dasymetric Modeling for Demographic Small Area Estimation. *Geographical Analysis*, 45, 285-306.
- LI, G. & WENG, Q. 2005. Using landsat ETM+ imagery to measure population density in Indianapolis, Indiana, USA. *Photogrammetric Engineering & Remote Sensing*, 71, 947-958.
- LIAO, Y., WANG, J., MENG, B. & LI, X. 2010. Integration of GP and GA for mapping population distribution. *International Journal of Geographical Information Science*, 24, 47-67.
- LIN, J., CROMLEY, R. & ZHANG, C. 2011. Using geographically weighted regression to solve the areal interpolation problem. *Annals of GIS*, 17, 1-14.
- LIU, X. 2011. Accuracy Assessment of Lidar Elevation Data Using Survey Marks. *Survey Review*, 43, 80-93.
- LIU, X., CLARKE, K. & HEROLD, M. 2006. Population Density and Image Texture: A Comparison Study. *Photogrammetric Engineering & Remote Sensing* 72, 187-196.
- LIU, X. & LIU, Y. 2008. The accuracy assessment in areal interpolation: An empirical investigation. *Science in China Series E: Technological Sciences*, 51, 62-71.
- LIU, X. H. 2003. *Estimation of the spatial distribution of urban population using high-resolution satellite imagery*. PhD, University of California.
- LIU, X. H., KYRIAKIDIS, P. C. & GOODCHILD, M. F. 2008. Population-density estimation using regression and area-to-point residual kriging. *International Journal of Geographical Information Science*, 22, 431-447.
- LO, C. P. 2008. Population estimation using Geographically weighted regression. *GIScience and Remote Sensing*, 45, 131-148.
- LONGLEY, P. 2005. *Geographic Information Systems and Science*, New York, Wiley.
- LU, D., HETRICK, S. & MORAN, E. 2011a. Impervious surface mapping with Quickbird imagery. *International Journal of Remote Sensing*, 32, 2519-2533.
- LU, D., WENG, Q. & LI, G. 2006. Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing*, 27, 3553-3570.

- LU, Z., IM, J. & QUACKENBUSH, L. 2011b. A volumetric approach to population estimation using lidar remote sensing. *Photogrammetric Engineering and Remote Sensing*, 77, 1145-1156.
- LU, Z., IM, J., QUACKENBUSH, L. & HALLIGAN, K. 2010. Population estimation based on multi-sensor data fusion. *International Journal of Remote Sensing*, 31, 5587-5604.
- LWIN, K. & MURAYAMA, Y. 2009. A GIS approach to estimation of building population for micro-spatial analysis. *Transactions in GIS*, 13, 401-414.
- LWIN, K. & MURAYAMA, Y. 2010. Development of GIS tool for dasymetric mapping. *International Journal of Geoinformatics*, 6, 11-18.
- MAANTAY, J. & MAROKO, A. 2009. Mapping urban risk: flood hazards, race, & environmental justice in New York". *Applied Geography*, 29, 111-124.
- MAANTAY, J. A., MAROKO, A. R. & HERRMANN, C. 2007. Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDs). *Cartography and Geographic Information Science*, 34, 77-102.
- MADHAVAN, B. B., WANG, C., TANAHASHI, H., HIRAYU, H., NIWA, Y., YAMAMOTO, K., TACHIBANA, K. & SASAGAWA, T. 2006. A computer vision based approach for 3D building modelling of airborne laser scanner DSM data. *Computers, Environment and Urban Systems*, 30, 54-77.
- MANSOUR, S., MARTIN, D. & WRIGHT, J. 2012. Problems of spatial linkage of a geo-referenced Demographic and Health Survey (DHS) dataset to a population census: A case study of Egypt. *Computers, Environment and Urban Systems*, 36, 350-358.
- MARTIN, D. 1996. *Geographic information systems: socioeconomic applications*, London, Routledge.
- MARTIN, D. 2004. Neighbourhoods and area statistics in the post 2001 census era. *Area* 36, 136-145.
- MARTIN, D. 2006. Last of the censuses? The future of small area population data. *Transactions of the Institute of British Geographers* 31, 6-18.
- MARTIN, D. 2010. Understanding the social geography of census undercount. *Environment and Planning A*, 42, 2753-2770.
- MARTIN, D. 2011. Directions in Population GIS. *Geography Compass*, 5, 655-665.

- MARTIN, D., COCKINGS, S. & HARFOOT, A. 2013. Development of a geographical framework for census workplace data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176, 585-602.
- MARTIN, D., COCKINGS, S. & LEUNG, S. Population 24/7: building time-specific population grid models. the European Forum for Geostatistics, 2009 Netherlands.
- MARTIN, D., LLOYD, C. D. & SHUTTLEWORTH, I. G. 2011. Evaluation of grided population models using 2001 Northern Ireland census data. *Environment and Planning A*, 43, 1965-1980.
- MARTIN, D., NOLAN, A. & TRANMER, M. 2001. The application of zone-design methodology in the 2001 UK Census. *Environment and Planning A*, 33, 1949-1962.
- MARTIN, D., TATE, N. J. & LANGFORD, M. 2000. Refining Population Surface Models: Experiments with Northern Ireland Census Data. *Transactions in GIS*, 4, 343-360.
- MCCLEARY, G. F. 1969. *The dasymetric method in thematic cartography*. PhD, University of Wisconsin.
- MENNIS, J. 2002. Using Geographic Information Systems to Create and Analyze Statistical Surfaces of Population and Risk for Environmental Justice Analysis. *Social Science Quarterly*, 83, 281-297.
- MENNIS, J. 2003. Generating Surface Models of Population Using Dasymetric Mapping. *The Professional Geographer*, 55, 31-42.
- MENNIS, J. 2009. Dasymetric mapping for estimating population in small areas. *Geography Compass*, 3, 727-745.
- MENNIS, J. 2010. Multidimensional Map Algebra: Design and Implementation of a Spatio-Temporal GIS Processing Language. *Transactions in GIS*, 14, 1-21.
- MENNIS, J. & HULTGREN, T. 2005. Dasymetric Mapping for Disaggregating Coarse Resolution Population Data.
- MENNIS, J. & HULTGREN, T. 2006a. Intelligent Dasymetric Mapping and Its Application to Areal Interpolation. *Cartography and Geographic Information Science*, 33, 179-194.
- MENNIS, J. & HULTGREN, T. 2006b. 'Intelligent' Dasymetric Mapping and Its Comparison to Other Areal Interpolation Techniques.

- MENNIS, J., MASON, M. J. & CAO, Y. 2013. Qualitative GIS and the visualization of narrative activity space data. *International Journal of Geographical Information Science*, 27, 267-291.
- MESEV, V. 2005. Identification and characterisation of urban building patterns using IKONOS imagery and point-based postal data. *Computers, Environment and Urban Systems*, 29, 541-557.
- MILIAREISIS, G. & KOKKAS, N. 2007. Segmentation and object-based classification for the extraction of the building class from LIDAR DEMs. *Computers and Geosciences*, 33, 1076-1087.
- MOORE, D. S. 2010. *The basic practice of statistics*, New York, W. H. Freeman and Company.
- MORTON, T. A. & YUAN, F. 2009. Analysis of population dynamics using satellite remote sensing and US census data. *Geocarto International*, 24, 143-163.
- MROZINSKI, R. D. & CROMLEY, R. G. 1999. Singly- and doubly- constrained methods of areal interpolation for vector-based GIS. *Transactions in GIS*, 3, 285-301.
- MUGGLIN, A. S. & CARLIN, B. P. 1998. Hierarchical Modeling in Geographic Information Systems: Population Interpolation over Incompatible Zones. *Journal of Agricultural, Biological, and Environmental Statistics*, 3, 111-130.
- MURAKAMI, D. & TSUTSUMI, M. 2011. A New Areal Interpolation Method Based on Spatial Statistics. *Procedia - Social and Behavioral Sciences*, 21, 230-239.
- MURAKAMI, D. & TSUTSUMI, M. 2012. Practical spatial statistics for areal interpolation. *Environment and Planning B: Planning and Design*, 39, 1016-1033.
- NAGLE, N. N., BUTTENFIELD, B. P., LEYK, S. & SPIELMAN, S. 2014. Dasymetric Modeling and Uncertainty. *Annals of the Association of American Geographers*, 104, 80-95.
- O'SULLIVAN, D. & UNWIN, D. J. 2003. *Geographic Information Analysis*, New Jersey, Wiley.
- OPENSHAW, S. 1984. The Modifiable Areal Unit Problem CATMOG 38 (Geo Books, Norwich).
- ORFORD, S. & RADCLIFFE, J. 2007. Modelling UK residential dwelling types using OS Mastermap data: A comparison to the 2001 census. *Computers, Environment and Urban Systems*, 31, 206-227.

- OWENS, P. M., TITUS-ERNSTOFF, L., GIBSON, L., BEACH, M. L., BEAUREGARD, S. & DALTON, M. A. 2010. Smart density: A more accurate method of measuring rural residential density for health-related research. *International Journal of Health Geography*, 9, 8.
- PETROV, A. 2012. One Hundred Years of Dasymetric Mapping: Back to the Origin. *The Cartographic Journal*, 49, 256-264.
- PETROV, A. N. 2008. Setting the Record Straight: On the Russian Origins of Dasymetric Mapping. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 43, 133-136.
- POULSEN, E. & KENNEDY, L. W. 2004. Using Dasymetric Mapping for Spatially Aggregated Crime Data. *Journal of Quantitative Criminology*, 20.
- QIU, F. & CROMLEY, R. 2013. Areal Interpolation and Dasymetric Modeling. *Geographical Analysis*, 45, 213-215.
- QIU, F., SRIDHARAN, H. & CHUN, Y. 2010. Spatial Autoregressive Model for Population Estimation at the Census Block Level Using LIDAR-derived Building Volume Information. *Cartography and Geographic Information Science*, 37, 239-257.
- REIBEL, M. 2007. Geographic Information Systems and Spatial Data Processing in Demography: a Review. *Population Research and Policy Review*, 26, 601-618.
- REIBEL, M. & AGRAWAL, A. 2007. Areal Interpolation of Population Counts Using Pre-classified Land Cover Data. *Population Research and Policy Review*, 26, 619-633.
- REIBEL, M. & BUFALINO, M. E. 2005. Street Weighted Interpolation Techniques for Demographic Count Estimation in Incompatible Zone Systems.
- ROBINSON, A. C., CHEN, J., LINGERICH, E. J., MEYER, H. G. & MACEACHREN, A. M. 2005. Combining Usability Techniques to Design Geovisualization Tools for Epidemiology. *Cartography and Geographic Information Science*, 32, 243-255.
- ROBINSON, A. H. 1955. The 1837 Maps of Henry Drury Harness. *The Geographical Journal* 121, 440-450.
- ROBINSON, S., C. 2004. *Modelling the distribution of population using Ordnance Survey vector data sets*. PhD, University of Leicester.
- SADAHIRO, Y. 1999. Accuracy of areal interpolation: A comparison of alternative methods *Journal of Geographical Systems* 1, 323-346.

- SAPORITO, S., CHAVERS, J. M., NIXON, L. C. & MCQUIDDY, M. R. 2007. From here to there: Methods of allocating data between census geography and socially meaningful areas. *Social Science Research*, 36, 897-920.
- SCHMID, T. & MÜNNICH, R. T. 2013. Spatial robust small area estimation. *Statistical Papers*.
- SCHROEDER, J. P. & VAN RIPER, D. C. 2013. Because Muncie's Densities Are Not Manhattan's: Using Geographical Weighting in the EM Algorithm for Areal Interpolation. *Geographical Analysis*, 45, 216-237.
- SHUTTLEWORTH, I. G., LLOYD, A. C. D. & MARTIN, D. J. 2011. Exploring the implications of changing census output geographies for the measurement of residential segregation: the example of Northern Ireland 1991–2001. *Royal Statistical Society*, 174, 1–16.
- SILVÁN-CÁRDENAS, J. L. & WANG, L. 2006. A multi-resolution approach for filtering LiDAR altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61, 11-22.
- SILVÁN-CÁRDENAS, J. L., WANG, L., ROGERSON, P., WU, C., FENG, T. & KAMPHAUS, B. D. 2010. Assessing fine-spatial-resolution remote sensing for small-area population estimation. *International Journal of Remote Sensing*, 31, 5605-5634.
- SLOCUM, T. A. 2009. *Thematic Cartography and Geovisualization*, Pearson Prentice Hall.
- SMITH, S. K. & CODY, S. 1994. Evaluating the Housing Unit Method: A Case Study of 1990 Population Estimates in Florida. *Journal of the American Planning Association*, 60, 209-221.
- SMITH, S. K. & LEWIS, B. B. 1980. Some new techniques for applying the housing unit method of local population estimation *Demography*, 17, 323-339.
- SMITH, S. K. & LEWIS, B. B. 1983. Some new techniques for applying the housing unit method of local population estimation: Further evidence. *Demography* 20, 407-413.
- SMITH, S. K. & MANDELL, M. 1984. A comparisons of population estimation methods: Housing unit versus Component II, Ration correlation and Administrative records. *Journal of the American Statistical Association*, 79, 282-289.

- SMITH, S. K., NOGLE, J. & CODY, S. 2002. A regression approach to estimating the average number of persons per household. *Demography*, 39, 697-712.
- SOHN, G. & DOWMAN, I. 2007. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 43-63.
- SRIDHARAN, H. & QIU, F. 2013. A Spatially Disaggregated Areal Interpolation Model Using Light Detection and Ranging-Derived Building Volumes. *Geographical Analysis*, 45, 238-258.
- SU, M. D., LIN, M. C., HSIEH, H. I., TSAI, B. W. & LIN, C. H. 2010. Multi-layer multi-class dasymetric mapping to estimate population distribution. *Science of the Total Environment*, 408, 4807-16.
- TAPP, A. F. 2010. Areal Interpolation and Dasymetric Mapping Methods Using Local Ancillary Data Sources. *Cartography and Geographic Information Science*, 37, 215-228.
- TIAN, Y., YUE, T., ZHU, L. & CLINTON, N. 2005. Modeling population density using land cover data. *Ecological Modelling*, 189, 72-88.
- TOBLER, W. 1979. Smooth Pynophylactic interpolation for geographical regions. *Journal of American Statistical Association*, 74, 519-536.
- UPEGUI, E. & VIEL, J. F. 2012. GeoEye imagery and Lidar Technology for Small-area population estimation: An epidemiological viewpoint. *Photogrammetric Engineering & Remote Sensing*, 78, 693-702.
- URAL, S., HUSSAIN, E. & SHAN, J. 2011. Building population mapping with aerial imagery and GIS data. *International Journal of Applied Earth Observation and Geoinformation*, 13, 841-852.
- VALENTE, P. 2010. Census taking in Europe: how are populations counted in 2010? *Population and Societies*, 467.
- VIEL, J. F. & TRAN, A. 2009. Estimating denominators: satellite-based population estimates at a fine spatial resolution in a European urban area. *Epidemiology*, 20, 214-22.
- WANG, L. & WU, C. 2010. Population estimation using remote sensing and GIS technologies. *International Journal of Remote Sensing*, 31, 5569-5570.
- WATMOUGH, G. R., ATKINSON, P. M. & HUTTON, C. W. 2013. Exploring the links between census and environment using remotely sensed satellite sensor imagery. *Journal of Land Use Science*, 8, 284-303.

- WENG, Q. 2012. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, 117, 34-49.
- WHITE, I. 2009. The 2011 Census taking shape: methodological and technological developments. *Population Trends*, 136, 64-72.
- WRIGHT, J. K. 1936. A method of Mapping Densities of Population: With Cape Cod as an Example. *Geographical Review*, 26, 103-110.
- WU, C. & MURRAY, A. T. 2003. Estimating impervious surface distribution by spectral mixture analysis. *Remote Sensing of Environment*, 84, 493-505.
- WU, C. & MURRAY, A. T. 2005. A cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems*, 29, 558-579.
- WU, C. & MURRAY, A. T. 2007. Population estimation using Landsat enhanced thematic Mapper Imagery. *Geographical Analysis*, 39, 26-43.
- WU, S.-S., QIU, X. & WANG, L. 2005. Population Estimation Methods in GIS and Remote Sensing: A Review. *GIScience and Remote Sensing*, 42, 80-96.
- WU, S.-S., QIU, X. & WANG, L. 2006. Using Semi-variance Image Texture Statistics to Model Population Densities. *Cartography and Geographic Information Science*, 33, 127-140.
- WU, S.-S., WANG, L. & QIU, X. 2008. Incorporating GIS Building Data and Census Housing Statistics for Sub-Block-Level Population Estimation. *The Professional Geographer*, 60, 121-135.
- XIAOYE, L. 2008. Airborne LiDAR for DEM generation: some critical issues. *Progress in Physical Geography*, 32, 31-49.
- XIE, Y. 1995. The overlaid network algorithms for areal interpolation problems *Computer, Environments and Urban Systems* 19, 287-306.
- XIE, Z. 2006. A Framework for Interpolating the Population Surface at the Residential-Housing-Unit Level. *GIScience and Remote Sensing*, 43, 1-19.
- YAGOUB, M. M. 2006. Application of remote sensing and Geographic information systems (GIS) to population studies in the gulf: A case of Al Ain City (UAE). *Journal of the Indian Society of Remote Sensing*, 34, 7-21.
- YOUNG, C., MARTIN, D. & SKINNER, C. 2009. Geographically intelligent disclosure control for flexible aggregation of census data. *International Journal of Geographical Information Science*, 23, 457-482.

- YUAN, Y., SMITH, R. M. & LIMP, W. F. 1997. Remodeling census population with spatial information from landsat TM imagery. *Computer, Environment and Urban Systems*, 21, 245-258.
- ZANDBERGEN, P. A. 2011. Dasymetric mapping using high resolution address point datasets. *Transactions in GIS*, 15, 5-27.
- ZANDBERGEN, P. A. & IGNIZIO, D. A. 2010. Comparison of dasymetric mapping techniques for small-area population estimates. *Cartography and Geographic Information Science*, 37 199-214.
- ZHAN, F. B., TAPIA SILVA, F. O. & SANTILLANA, M. 2010. Estimating small-area population growth using geographic-knowledge-guided cellular automata. *International Journal of Remote Sensing*, 31, 5689-5707.
- ZHANG, C. & QIU, F. 2011. A point-based intelligent approach to areal interpolation. *The Professional Geographer*, 63, 262-276.
- ZHOU, G., SONG, C., SIMMERS, J. & CHENG, P. 2004. Urban 3D GIS From LiDAR and digital aerial images. *Computers and Geosciences*, 30, 345-353.
- ZHOU, Y., ZHANG, C. & QIU, F. 2012. The development of an areal interpolation ArcGIS extension and a comparative study. *GIScience and Remote Sensing*, 49, 644-663.

Web references

- ENVI User's Guide, http://aviris.gl.fcen.uba.ar/Curso_SR/biblio_sr/ENVI_userguid.pdf
- Office for National Statistics (ONS), <http://infuse2011.mimas.ac.uk/>
- ONS, <http://www.neighbourhood.statistics.gov.uk/>
- Ordnance Survey (OS), <http://borders.edina.ac.uk/ukborders/>
- OS, <http://census.ukdataservice.ac.uk/get-data/boundary-data.aspx>
- OS MasterMap® <http://digimap.edina.ac.uk/digimap/home>

Appendices

Appendix 1: Classification accuracy assessment report for the Leicester City.

Confusion Matrix: Z:\PhD_Data\THESIS
CHAPTERS\Dasymetric_binary_method\leicester_classification.tiff

Overall Accuracy = (900679/1007593) 89.3892%
Kappa Coefficient = 0.7797

Ground Truth (Pixels)

Class	Green Space	Tree Cover	Urban Area	Water	Total
Unclassified	0	0	0	0	0
Green Space	644043	15333	3752	0	663128
Tree Cover	58867	53104	3548	501	116020
Urban Area	10981	594	194499	19	206093
Water	1412	4359	7548	9033	22352
Total	715303	73390	209347	9553	1007593

Ground Truth (Percent)

Class	Green Space	Tree Cover	Urban Area	Water	Total
Unclassified	0.00	0.00	0.00	0.00	0.00
Green Space	90.04	20.89	1.79	0.00	65.81
Tree Cover	8.23	72.36	1.69	5.24	11.51
Urban Area	1.54	0.81	92.91	0.20	20.45
Water	0.20	5.94	3.61	94.56	2.22
Total	100.00	100.00	100.00	100.00	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Green Space	2.88	9.96	19085/663128	71260/715303
Tree Cover	54.23	27.64	62916/116020	20286/73390
Urban Area	5.63	7.09	11594/206093	14848/209347
Water	59.59	5.44	13319/22352	520/9553

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Green Space	90.04	97.12	644043/715303	644043/663128
Tree Cover	72.36	45.77	53104/73390	53104/116020
Urban Area	92.91	94.37	194499/209347	194499/206093
Water	94.56	40.41	9033/9553	9033/22352

Appendix 2: Classification accuracy assessment report for the Borough of Kensington and Chelsea

Confusion Matrix: Z:\PhD_Data\THESIS
CHAPTERS\Dasymetric_binary_method\Kensington_classification.tiff

Overall Accuracy = (318443/355172) 89.6588%

Kappa Coefficient = 0.8548

Ground Truth (Pixels)

Class	Green Space	Tree Cover	Water	Urban Area	Total
Unclassified	0	0	0	0	0
Green Space	57020	24272	0	330	81622
Tree Cover	6125	58058	730	667	65580
Water Bodies	12	2180	146161	2301	150654
Urban Area	27	85	0	57204	57316
Total	63184	84595	146891	60502	355172

Ground Truth (Percent)

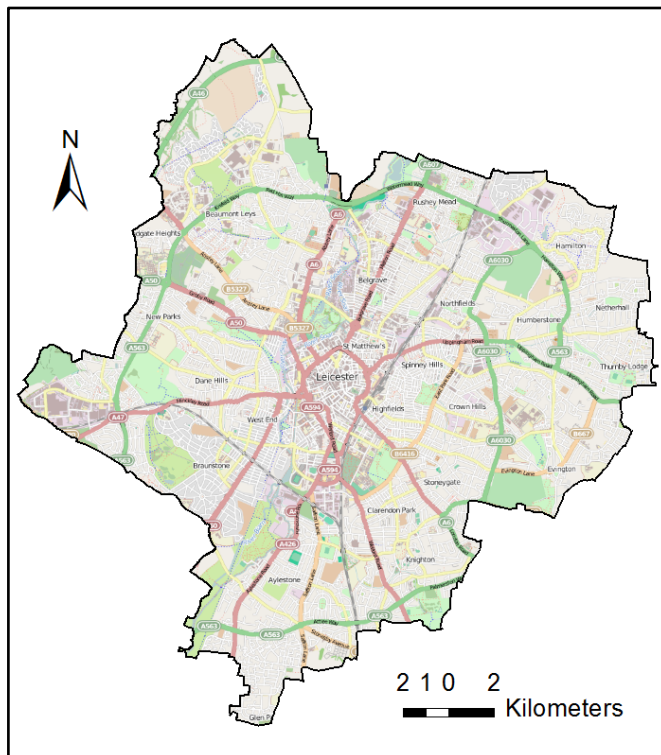
Class	Green Space	Tree Cover	Water	Urban Area	Total
Unclassified	0.00	0.00	0.00	0.00	0.00
Green Space	90.24	28.69	0.00	0.55	22.98
Tree Cover	9.69	68.63	0.50	1.10	18.46
Water	0.02	2.58	99.50	3.80	42.42
Urban Area	0.04	0.10	0.00	94.55	16.14
Total	100.00	100.00	100.00	100.00	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Green Space	30.14	9.76	24602/81622	6164/63184
Tree Cover	11.47	31.37	7522/65580	26537/84595
Water	2.98	0.50	4493/150654	730/146891
Urban Area	0.20	5.45	112/57316	3298/60502

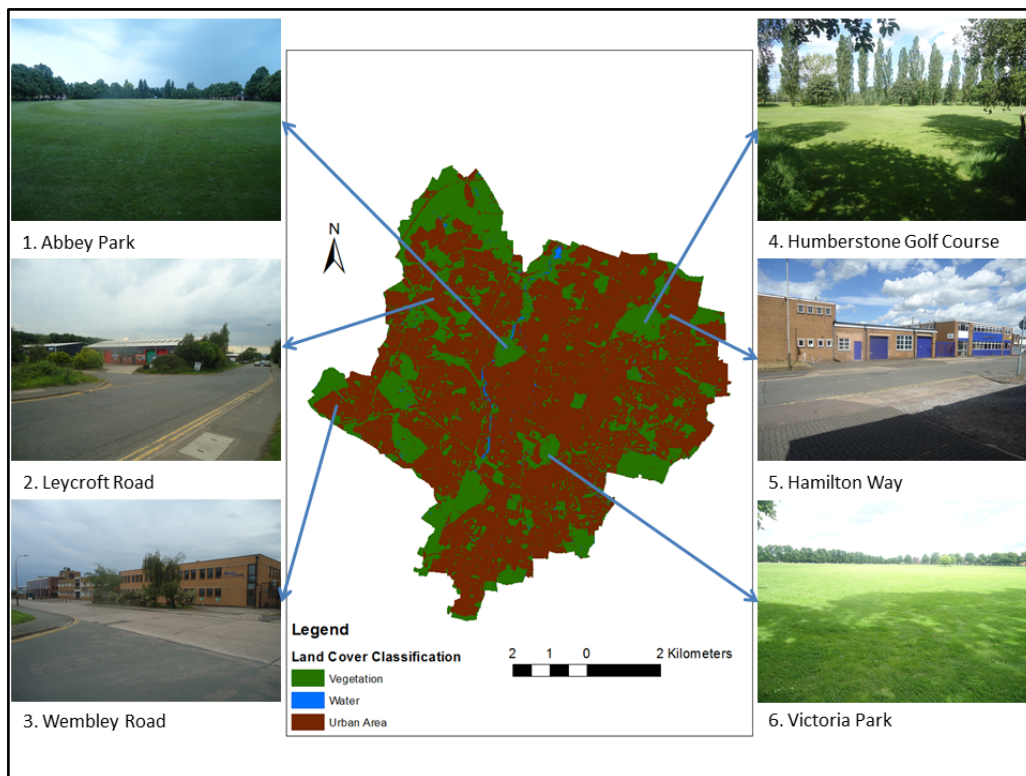
Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Green Space	90.24	69.86	57020/63184	57020/81622
Tree Cover	68.63	88.53	58058/84595	58058/65580
Water	99.50	97.02	146161/146891	146161/150654
Urban Area	94.55	99.80	57204/60502	57204/57316

Appendix 3: The validation of image classification for Leicester City

1. City map of Leicester.



2. The classified image and pictures from the selected areas.



Appendix 4: Codes for Accuracy Measurements

R programming codes were used to provide values of RMSE.

```
data<- read.csv ("the_City_of_Leicester.csv", as.is=T)
```

```
data<- read.csv ("the_Borough_of_Kensington_and_Chelsea.csv", as.is=T)
```

```
data
```

```
RMSE <- function(x,y){
```

```
  for(i in 1:length(x)){
```

```
    for(j in 1:length(y)){
```

```
      r<- sqrt(sum((x[i] - y[j]))^2/length(x))
```

```
    }
```

```
  }
```

```
  return(r)
```

```
}
```

```
x <- data$x
```

```
y <- data$y
```

```
RMSE(x,y)
```

Appendix 5: Spatial Autocorrelation Reports for the residuals of areal interpolation methods for the City of Leicester

1. The areal weighting method

Moran's Index: -0.005069

z-score: -0.347717

p-value: 0.728053

Given the z-score of -0.35, the pattern does not appear to be significantly different than random.

Global Moran's I Summary

Moran's Index: -0.005069

Expected Index: -0.001033

Variance: 0.000135

z-score: -0.347717

p-value: 0.728053

Dataset Information

Input Feature Class: Target Units (OAs)

Input Field: RESIDUALS THE AREAL WEIGHTING

Conceptualization: INVERSE_DISTANCE

Distance Method: EUCLIDEAN

Row Standardization: False

Distance Threshold: 614.767609 Meters

Weights Matrix File: None

Selection Set: False

2. The binary mapping method

Moran's Index: 0.176294

z-score: 15.291245

p-value: 0.000000

Given the z-score of 15.29, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

Moran's Index: 0.176294

Expected Index: -0.001033

Variance: 0.000134

z-score: 15.291245

p-value: 0.000000

Dataset Information

Input Feature Class: Target Units (OAs)

Input Field: RESIDUALS THE BINARY

Conceptualization: INVERSE_DISTANCE

Distance Method: EUCLIDEAN

Row Standardization: False

Distance Threshold: 614.767609 Meters

Weights Matrix File: None

Selection Set: False

3. The volumetric estimation method

Moran's Index: -0.003600

z-score: -0.221306

p-value: 0.824854

Given the z-score of -0.22, the pattern does not appear to be significantly different than random.

Global Moran's I Summary

Moran's Index: -0.003600

Expected Index: -0.001033

Variance: 0.000135

z-score: -0.221306

p-value: 0.824854

Dataset Information

Input Feature Class: Target Units (OAs)

Input Field: RESIDUALS THE VOLUMETRIC

Conceptualization: INVERSE_DISTANCE

Distance Method: EUCLIDEAN

Row Standardization: False

Distance Threshold: 614.767609 Meters

Weights Matrix File: None

Selection Set: False

4. The address-weighted method using building address points only

Moran's Index: 0.066480

z-score: 5.877464

p-value: 0.000000

Given the z-score of 5.88, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

Moran's Index: 0.066480

Expected Index: -0.001033

Variance: 0.000132

z-score: 5.877464

p-value: 0.000000

Dataset Information

Input Feature Class: Target Units (OAs)

Input Field: RESIDUALS_THE ADDRESS WEIGHTED

Conceptualization: INVERSE_DISTANCE

Distance Method: EUCLIDEAN

Row Standardization: False

Distance Threshold: 614.767609 Meters

Weights Matrix File: None

Selection Set: False

5. The address-weighted method using the integration of building address points and occupancy information

Moran's Index: -0.015153

z-score: -1.212512

p-value: 0.225316

Given the z-score of -1.21, the pattern does not appear to be significantly different than random.

Global Moran's I Summary

Moran's Index: -0.015153

Expected Index: -0.001033

Variance: 0.000136

z-score: -1.212512

p-value: 0.225316

Dataset Information

Input Feature Class: Target Units (OAs)
Input Field: RESIDUALS _THE ADDRESS WEIGHTED 2
Conceptualization: INVERSE_DISTANCE
Distance Method: EUCLIDEAN
Row Standardization: False
Distance Threshold: 614.767609 Meters
Weights Matrix File: None
Selection Set: False

Appendix 6: Spatial Autocorrelation Reports for the residuals of areal interpolation methods for the Borough of Kensington and Chelsea

1. The areal weighting method

Moran's Index: 0.020894

z-score: 1.346419

p-value: 0.178167

Given the z-score of 1.35, the pattern does not appear to be significantly different than random.

Global Moran's I Summary

Moran's Index: 0.020894

Expected Index: -0.001587

Variance: 0.000279

z-score: 1.346419

p-value: 0.178167

Dataset Information

Input Feature Class: Target Units (OAs)

Input Field: RESIDUALS THE AREAL WEIGHTING

Conceptualization: INVERSE_DISTANCE

Distance Method: EUCLIDEAN

Row Standardization: False

Distance Threshold: 281.692214 Meters

Weights Matrix File: None

Selection Set: False

2. The binary mapping method

Moran's Index: 0.033158

z-score: 2.071850

p-value: 0.038279

Given the z-score of 2.07, there is less than 5% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

Moran's Index: 0.033158

Expected Index: -0.001587

Variance: 0.000281

z-score: 2.071850

p-value: 0.038279

Dataset Information

Input Feature Class: Target Units (OAs)

Input Field: RESIDUALS THE BINARY

Conceptualization: INVERSE_DISTANCE

Distance Method: EUCLIDEAN

Row Standardization: False

Distance Threshold: 281.692214 Meters

Weights Matrix File: None

Selection Set: False

3. The volumetric estimation method

Moran's Index: 0.000495

z-score: 0.124224

p-value: 0.901138

Given the z-score of 0.12, the pattern does not appear to be significantly different than random.

Global Moran's I Summary

Moran's Index: 0.000495

Expected Index: -0.001587

Variance: 0.000281

z-score: 0.124224

p-value: 0.901138

Dataset Information

Input Feature Class: Target Units (OAs)

Input Field: RESIDUALS THE VOLUMETRIC

Conceptualization: INVERSE_DISTANCE

Distance Method: EUCLIDEAN

Row Standardization: False

Distance Threshold: 281.692214 Meters

Weights Matrix File: None

Selection Set: False

4. The address-weighted method using building address points only

Moran's Index: 0.298596

z-score: 18.437767

p-value: 0.000000

Given the z-score of 18.44, there is less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

Moran's 0.298596

Index:

Expected -0.001587

Index:

Variance: 0.000265

z-score: 18.437767

p-value: 0.000000

Dataset Information

Input Feature Class: Target Units (OAs)

Input Field: RESIDUALS_THE ADDRESS WEIGHTED

Conceptualization: INVERSE_DISTANCE

Distance Method: EUCLIDEAN

Row False

Standardization:

Distance Threshold: 281.692214 Meters

Weights Matrix File: None

Selection Set: False

5. The address weighted method using the integration of building address points and occupancy information

Moran's Index: -0.025765

z-score: -1.437615

p-value: 0.150543

Given the z-score of -1.44, the pattern does not appear to be significantly different than random.

Global Moran's I Summary

Moran's Index: -0.025765

Expected Index: -0.001587

Variance: 0.000283

z-score: -1.437615

p-value: 0.150543

Dataset Information

Input Feature Class: Target Units (OAs)
Input Field: RESIDUALS _THE ADDRESS WEIGHTED 2
Conceptualization: INVERSE_DISTANCE
Distance Method: EUCLIDEAN
Row Standardization: False
Distance Threshold: 281.692214 Meters
Weights Matrix File: None
Selection Set: False

