Estimating Material and Energy Intensities of Urban Areas

by

David James Quinn

Bachelor of Science in Civil Engineering
University College Dublin, 2005

Master of Science in Building Technology
Massachusetts Institute of Technology, 2008

Submitted to the Department of Architecture
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Architecture: Building Technology

at the

Massachusetts Institute of Technology

June 2012

Signature of Author: ______________________________

Department of Architecture
May 4, 2012

Certified by: ______________________________

John E. Fernández
Associate Professor of Architecture and Building Technology
and Engineering Systems
Thesis Supervisor

Accepted by: ______________________________

Takehiko Nagakura
Associate Professor of Design and Computation
Chair, Departmental Committee on Graduate Students

Thesis Committee:

John E. Fernández
Associate Professor of Architecture and Building Technology
and Engineering Systems
Massachusetts Institute of Technology
Thesis Supervisor

Leslie Keith Norford
Professor of Building Technology
Massachusetts Institute of Technology
Thesis Reader

P. Christopher Zegras
Associate Professor of Urban Planning and Transportation
Massachusetts Institute of Technology
Thesis Reader

Michael Flaxman
Assistant Professor of Urban Technologies and Information Systems
Massachusetts Institute of Technology
Thesis Reader

Estimating Material and Energy Intensities of Urban Areas

by

David James Quinn

Submitted to the Department of Architecture
on May 4, 2012 in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Architecture: Building Technology

Abstract

The objective of this thesis is to develop methods to estimate, analyze and visualize the resource intensity of urban areas. Understanding the resource consumption of the built environment is particularly relevant in cities that are rapidly growing, as the urban forms that emerge have long-term consequences for both the quality of life of the inhabitants, and their future material and energy demands.

This work was completed by assembling datasets of cities from around the world, identifying geometric patterns in the built environment, relating these geometric patterns to material and energy intensities, and illustrating these intensities in a visually intuitive way. This thesis describes a standardized analytical approach to assess the physical characteristics of the built environment, enabling comparisons to be made between cities. This approach provides a preliminary assessment of resource intensities that may be useful for decision-makers to compare differences among a variety of urban forms.

Finally, a new web-map visualization tool has been developed that enables users to gain an understanding of the resource intensity of 40 cities in the USA. This tool allows the user to explore the resource intensity of urban areas using a web-browser, and to dynamically generate reports that can compare areas within a city, or entire cities, to each other.

Thesis supervisor: John E Fernández
Title: Associate Professor of Architecture and Building Technology and Engineering Systems

Acknowledgements

I would first like to thank my advisor, Professor John E Fernandez. Over the course of my Masters and PhD, John patiently allowed me to explore many avenues of research, while remaining supportive and encouraging at all times.

I would also like to thank my committee - Professors Les Norford, Mike Flaxman and Chris Zegras for their advice and encouragement.

I have had the pleasure of working closely with two collaborators (and friends) - Daniel Wiesmann from the Institute Superio Technico and Juan Jose Sarralde from the University of Cambridge.

I have greatly enjoyed my time at MIT. The opportunity to take many classes, and to work with peers from a variety of backgrounds was very rewarding. I have greatly appreciated the collaborative and non-competitive culture within MIT where ideas and skills are shared freely. This culture of openness continues to inspire me.

I have enjoyed working with my colleagues in Urban Metabolism at MIT, (Karen, Noel, Jonathan and Artessa). My colleagues and friends in the Building Technology Lab have been great companions, in particular Teri, Caitlin, Samar, Timothy, Bruno, Kate, Jenn, Zara, Ale., Tea, Adam, Sian and Nick. The administrative support staff in the Building Technology (Kathleen Ross, Alex Golledge and Alexandra Mulcahy) and in Architecture (Renée Caso) have also been very helpful during my time here.

The technical staff in the School of Architecture and Planning have also been extremely helpful. I have learned much from the eternally patient Duncan and all of the CRON crew (Rickie, Phil, Eduardo, Tom and Jesse). Daniel and Lisa in the GIS-Lab were also very helpful during my time here, in particular when I started to explore the world of GIS.

I was fortunate to be supported financially over the course of my studies; initially as a teaching assistant through the Building Technology Program and then as a recipient of an MIT Presidential Fellowship. I was also supported by the MIT Portugal Program and most recently, as a recipient of a Martin Fellowship for Sustainability.

Finally, on a personal level, I would like to thank all the people close to me for their support and encouragement. The collective Quinn Clann have been extremely supportive during my time in the US (Joan & Joe, John & Katie, Michael & Pauline) as well as Eileen & Tony.

Kathryn has been encouraging and supportive as I reached the end of my PhD, as have my wonderful room-mates (Kate, Alex & Karen). I received encouragement from my close friends who are both near and far (Ed, Nathan, Dave, Chelsea, Adam, Olwen, Sara, Johnny, Meredith, Ben, Beaudry, Cormac, Nuno, Joana, Roisin & Deirdre). I have shared many enjoyable times with friends from MIT, which made the experience much more enjoyable and who I look forward to staying in contact with (Ted, Beth, Dimitris, Beaudry, Shawn, Nabil, Danial, Rory M, Rory C, Conor, Frank & Tony).

Contents

1 Introduction **17**
1.1 Resource Intensity of Cities . . . 19
1.2 Analysis . . . 20
1.3 Data . . . 21
1.4 Visualization . . . 23

2 Literature Review **25**
2.1 Global Trends . . . 26
2.2 Industrial Ecology . . . 28
2.3 Urban Metabolism . . . 30
2.4 Complexity Theory . . . 31
2.5 System Dynamics . . . 33
2.6 Land Use and Transportation . . . 36
2.7 Urban Information Systems . . . 41
2.8 Research Question . . . 43

3 Urban Characterization: City Level **45**
3.1 Methodology and Data . . . 47
3.2 Population Density Gradient . . . 50
3.3 Road Density Gradient . . . 53
3.4 Grid Level Measurements . . . 58
3.5 Characterizing the City using Building Measurements . . . 64
3.6 Conclusions . . . 68

4 Urban Parameters: Neighborhood Scale **69**
4.1 Urban Parameter Relationships . . . 69
4.2 Residential Building Prediction . . . 79
4.3 Validation of Parameter Relationships . . . 82
4.4 Conclusions . . . 85

5 Resource Intensity **87**
5.1 Material Intensity . . . 88
5.2 Energy Intensity . . . 94

5.3 Neighborhood Typologies . . . 102
5.4 Combined Material and Energy Patterns . . . 106
5.5 Conclusions . . . 109

6 Discussion **111**
6.1 Resource Intensity and Urban Efficiency . . . 111
6.2 Application to Planning . . . 119
6.3 Conclusions . . . 123

7 Web-based Spatial Analysis **125**
7.1 Tools to Assist with Pathways for Sustainability . . . 126
7.2 Neighborhood Visualizer . . . 131
7.3 Technical Details . . . 135
7.4 Dynamic Report Generation . . . 138
7.5 Dissemination of Analysis . . . 141
7.6 Survey and Analysis of User Behavior . . . 143
7.7 Methods of Analysis, Data and Future Technology . . . 154
7.8 Conclusions . . . 159

8 Conclusions **161**
8.1 General Conclusions . . . 161
8.2 Specific Conclusions . . . 162
8.3 Analysis Processes and Data . . . 165
8.4 Future Work . . . 166
8.5 Concluding Comments . . . 167

Appendices **171**

A Data **171**
A.1 Spatial Data . . . 171
A.2 Classification References . . . 172

B ArcGIS Plugin **175**
B.1 ArcGIS: Population and Road Density Gradient Analysis Tool . . . 175

C Spatial Analysis: Examples of Code **179**
C.1 Python Scripting for ArcGIS . . . 179
C.2 Spatial Database Analysis . . . 180

D Survey Data **183**

E Bibliography **189**

List of Figures

1.1 Low density housing (London). 17
1.2 High density housing (London). 17
1.3 Hypothesized influence of data complexity on predictive error, and the levels of abstraction that are used when analyzing urban areas. 18

2.1 Global historic and projected urban and rural population. Image from United Nations (2010). 26
2.2 Energy for transportation. 36

3.1 Population density and road network patterns in the USA. These visualizations use data from the US Census (2000). 45
3.2 Map of the USA showing the cities analyzed in this chapter. 46
3.3 Atlanta service density, with the peak density marked with red. 48
3.4 Schematic of the analytical technique used to measure the radial distribution of population density and road density. 49
3.5 3D illustrations of the population density of Atlanta. 50
3.6 Population density gradient measure for Atlanta. 51
3.7 Population density gradient for 40 US cities. 51
3.8 Local road density gradient. 53
3.9 Road density gradient for roads. 53
3.10 Total road length measure from the city center, for 40 US cities. 54
3.11 Road per person. 54
3.12 Local road patterns for cities Boston and Dallas at a distance of 5 km from the city center. 55
3.13 Raster population density representation of two cities. The maximum value of each horizontal and vertical band is shown on the plot along each axis. 58
3.14 Map of the USA showing the city gradients for 40 cities. 59
3.15 Road density for two cities. 60
3.16 Linear road density for 40 cities, plotted against population density. 61
3.17 Road length per person, illustrated using two cities. 62
3.18 Road length per person for 40 cities, plotted against population density. 63
3.19 General view of land-use and population density. 64

3.20 Relationship between building height and building distance for three cities. . . 65
3.21 *Spacemate* criteria. Image from spacemate.nl 65
3.22 Spacemate land use types. Image from Berghauser-Pont (2010) 65
3.23 Physical limitations of urban space. *Floors* shows the relationship between GSI and FSI, and *Open Space Ratio* illustrates the relationship between these parameters. 66
3.24 Using *The Spacemate* approach for three cities. FSI and GSI were calculated at a 300m grid-cell. All buildings used are shown in the boxes above each plot. . 67

4.1 Built environment measures for road width estimation. 70
4.2 Map of data for each city used in this analysis. 71
4.3 Clipping vector measurements to a 250m grid illustrated using a sample of data from New York. 71
4.4 Relationship between road length and total road area. 73
4.5 Histograms of road-width for the four cities examined. 73
4.6 Road network diagram. In the *Simplified Road Network*, $\sum i = 1$, $\sum i_c = 4$ and in the the *Typical Road Network*, $\sum i = 2$, $\sum i_c = 5$ as intersections with $i = 2$ are ignored. 74
4.7 Road segments (black) and intersections (red). A grid with 250m spacing is shown. 75
4.8 Relationship between intersections and connections. 76
4.9 Relationship between centerline road density and the intersection density. . . . 76
4.10 Population density and road length. 77
4.11 Road length per person, calculated using block group level averages. 77
4.12 Road length per person for all 40 US cities calculated per block group using 106305 observations. 78
4.13 Predicted and actual areas for three urban areas. 83
4.14 Relationship between people and residential building area. 84
4.15 Building Height and Volume prediction. 84

5.1 Histograms of road area per person for varying population density levels. . . 89
5.2 Road area fraction for varying population density levels per city. 90
5.3 Total material content in road infrastructure per grid-cell. 91
5.4 Built area per person measurements. 91
5.5 Building area per person histograms for varying population density levels. This is a cross-section of the data shown in Figure 5.4 (a). 92
5.6 Calculating the euclidean distance to services. The red square represents one grid cell, and the small circles represent business locations. The euclidean distance to each service was measured from each grid-cell. 94
5.7 This combined distance to services for each grid-cell is shown for Atlanta in 3D. 95
5.8 Parameters used in model. 96
5.9 Estimated distances for four cities in MA. The error between the predicted and empirical values had a mean value of 12.78% and a standard deviation of 10. . 97

5.10 Estimated VKT per capita for varying population density levels. 98
5.11 Empirical and estimated values. The empirical data is from the state of MA, the estimated values are all values for the 40 cities. 99
5.12 Comparison of predicted values with city level empirical data 99
5.13 Estimated yearly New York residential electricity consumption per zip-code, normalized by the building area. Data from nycopendata.socrata.com. 100
5.14 New York yearly electricity consumption per block-group, normalized by the building area. Data from (Howard et al., 2012) 101
5.15 Yearly total of electricity consumption per zip code for New York. Data from nycopendata.socrata.com 101
5.16 Objective of neighborhood typology analysis. In this section, the focus is on material usage. 102
5.17 Typologies and dimensions of measurement of the urban form 104
5.18 Map of London showing typologies. The range of each variable is shown in Figure 5.17. The black lines represent the motorway around London. 104
5.19 Examples of typologies from the London Metro area 104
5.20 Local road area and residential building surface area. 105
5.21 Infrastructure material normalized per household. 105
5.22 Explanation of plot structure. 106
5.23 Combined material and energy measures per household for New York and San Francisco. 107
5.24 Combined material and energy measures per household for US cities. 108

6.1 New York metropolitan area and boroughs. 112
6.2 Material and energy intensities per household for New York, considering the metro area and the borough. 113
6.3 Histogram of energy intensities for New York. 114
6.4 Histogram of material intensities for New York. 114
6.5 American Community Survey data illustrating the mode choice fraction of workers. The values shown here are the average per 1km group. 117
6.6 Distances within 30 *km* considering weekly frequency of service use. 118
6.7 Intersection density plotted against non-motorized mode-choice from the ACS. The values shown here are the average values per intersection group. 118

7.1 Accelerating and democratizing the learning cycle with regard to resource intensity. 125
7.2 Initial concept sketch (*November 2010*). 132
7.3 *Neighborhood Visualizer v0.1* interface. The input parameters on the left-hand side are for demonstration purposes (*January 2011*). A *MySQL* database was used to store data. 132
7.4 *Neighborhood Visualizer v0.2* interface. The input parameters on the left-hand side provide the user ways to adjust the mix of urban forms that are present (*June 2011*). 133

7.5 *Neighborhood Visualizer v0.3* interface. This version enables analysis of 40 cities, and the dynamic generation of reports. The database was changed to PostgreSQL with PostGIS (*December 2011*). The default analysis box is shown here with the results from a query displayed for New York City. 133
7.6 *Neighborhood Visualizer v0.9* (*April 2012*). The user can choose between three different *heatmap* options to overlay a population density measure, an energy measure or a material measure. The term *heatmap* was used as a colloquial term to describe the raster overlay. 134
7.7 *Neighborhood Analysis Visualizer v0.9* with the *help* options expanded (*April 2012*). Choosing the interactive demo enables the user to step through the process of analysis as well as dynamic explanations of the tool's features. 134
7.8 urbmet.org site structure 136
7.9 Vector grid used to convert the raw spatial information. The centroid of buildings is stored, and each measure is calculated for each part of the road segment per cell. 137
7.10 Raster grid of road data. 137
7.11 Report for Boston, with comparative measurements at the neighborhood scale. 139
7.12 Report for Boston with comparative measurements at the city scale. 140
7.13 Accessing data using a desktop GIS via a Web Map Service or Web Coverage Service. 141
7.14 Connecting to the WMS service using *QGIS*. This *url*s are listed in Table 7.4. . 142
7.15 Viewing the WMS service using *QGIS*. Here population density for Boston is loaded into the data window using the WMS *url*. 142
7.16 Site visitors and areas that were queried. Users examined areas that were not analyzed as part of this research. 145
7.17 Examples of queries that users performed. Queries outside the analysis area did not return any data, just a message that data was not available for the user's area of interest. 146
7.18 These calculations were only performed for IP addresses or queries that were within in the US. 147
7.19 Area of query, based on the order that queries were performed by each user. There were approximately four queries per user. 148
7.20 Average query area per interaction (error bars show one standard deviation). . 149
7.21 Illustration of query areas. A base-map of aerial imagery from *Bing* maps is used here. 149
7.22 Histogram of query area per profession. The mean values for each facet are: Architects: 22.4, Engineers: 19.3, Planners: 20.3 150
7.23 Areas examined by users. 150
7.24 Behavior of user #1. 151
7.25 Behavior of user #2. 151
7.26 Behavior of user #3. 152
7.27 Behavior of user #4. 152

8.1 The four steps in this analysis. (This figure is repeated from Chapter **??**). . . . 162

B.1 Screenshot of population density gradient plugin 175
B.2 Pseudocode for population density gradient calculation. 176
B.3 Road density calculation (Figure repeated from Chapter 3). 177
B.4 Pseudocode for road density gradient calculation. 177

C.1 Pseudocode for spatial query to identify all the local roads that are within a block-group of the census. 180
C.2 SQL code example illustrating spatial query. 180
C.3 SQL code example illustrating how the number of roads that made an intersection were recorded. 180
C.4 SQL code to remove duplicate lines. 181
C.5 Connection count of road network for a sample of *New York*. 181
C.6 Pseudocode for spatial query to identify all the local roads that are within a block-group of the census. 181
C.7 SQL code example illustrating spatial join. 182
C.8 SQL code illustrating multi-band raster query. *POLYGON* is assumed to be a valid polygon that intersects the raster for a non-null result to be returned. . . 182

1 *Introduction*

The objective of this work is to contribute to the measurement of the resource intensity of urban areas. I achieve this by providing spatially detailed measurements of material and energy use in urban areas, so that the use of resources can be compared using the same initial assumptions. I hope that this contribution will lead to an improved understanding of human-environmental interactions, resulting in a better understanding of global urbanization patterns. The results from these analyses may assist urban planners in making informed decisions about material and energy use, and assist with improving resource efficiency. In addition, this analysis can facilitate rapidly urbanizing cities to understand the resource consumption consequences of urban development patterns.

Figure 1.1: Low density housing (London).

Figure 1.2: High density housing (London).

Figure 1.1 and 1.2 illustrate the type of problem that I hope this work will help assess. Here we can see two different urban form patterns, with different population densities, building sizes and infrastructure requirements. If we consider the resources consumed within the building, there are several metrics that help us to assess the efficiency in a straightforward manner if we focus purely on the dimension of energy. If we start to consider the overall urban system, including the building and infrastructure, as well as the energy required to enable the households to function within this urban system, the problem becomes more complicated due to the multiple dimensions (material, energy, social, environmental) that need to be considered. When considering these multiple dimensions, it is difficult to argue for the most efficient urban form. However, if the ranges of resource intensity measures are made explicit and stakeholders combine these dimensionally different criteria using collective values, an argument can be made for an urban form that attempts to minimize the resources used, while satisfying the societal criteria.

General patterns of resource use due to the spatial configuration of cities have not frequently been explored in detail. Often, aggregated per capita values are used when resource consumption is considered within a city. This aggregation conceals spatial variations within the city, which is problematic when local policies are considered. In this analysis, a method is developed to assess urban performance use high-resolution spatial measurements. In addition, this work does not use political boundaries of cities, but identifies urban boundaries using population density thresholds. Hence, the focus is on assessing the overall behavior of the urban system and does not consider arbitrary political boundaries.

I consider this work to be relevant for the following reasons. The first is due to rapid global urbanization, resulting in increasing demand for material and energy. The second reason is due to the difficulty in assessing the resource demands of urban areas. This is due to several factors, but the main reason is that our data gathering processes reflect the current economic transaction system, rather than measurements of the flow of resources through our society. Typically our political systems are not structured in such a way that they enable resource consumption data to be measured at high-levels of spatial resolution. The challenge in predicting the resource consumption of urban areas is accentuated by the difficulty in assembling comprehensive spatial datasets at high levels of resolution. While some of the data-gathering problems are mitigated by bottom-up sensing and distributed sensors, this approach is not sufficiently widespread to be used across many cities. The approach taken in this thesis is to examine several different data sources from national and city governments and to explore how these data can be used to predict urban resource flows. I have chosen to categorize these data into three loose hierarchical levels of data complexity, which are shown in Figure 1.3.

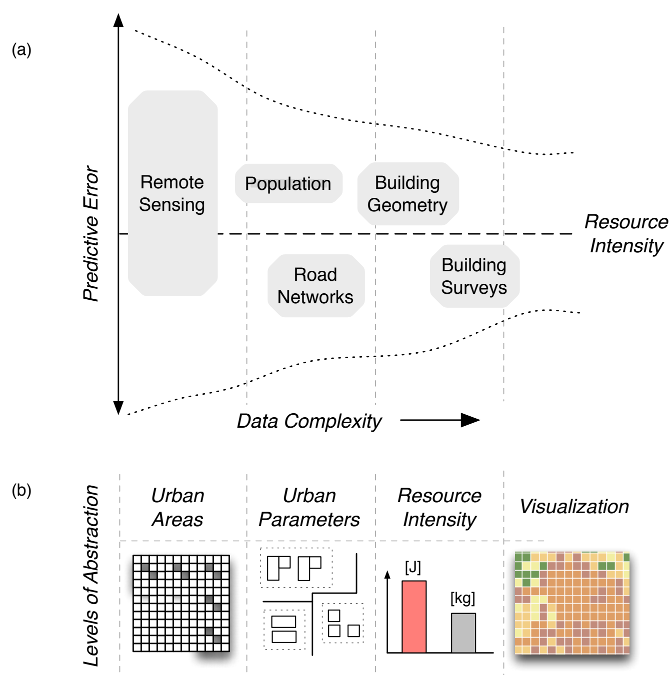


Figure 1.3: Hypothesized influence of data complexity on predictive error, and the levels of abstraction that are used when analyzing urban areas.

The term *Data complexity*, shown in Figure 1.3-a is considered to be a combination of both the level of detail of data and the ease of accessibility from a logistical perspective (both availability and access). Here, the influence of more detailed data is assumed to have a cumulative effect on reducing the error associated with a *Resource Intensity* prediction. One objective of this work is to examine the difference between predictions of energy and material use of the built environment at varying levels of detail. The term *Levels of Abstraction* considers the four different approaches used in this work. The first level, *Urban Areas* focused on general urban form characterization. The second level, *Urban Parameters* focused on identifying parametric descriptors of the urban form, and the third level, *Resource Use* explored how these patterns can be related to ranges of material intensity for buildings and roads, and ranges of energy use for building functioning and transportation. The final abstraction level is the visualization of this data in an intuitive (and iterative) way that can facilitate learning. This analysis draws on data from 42 cities in the US and UK, and is performed at a high-level of spatial resolution using building and road level spatial information.

Due to the inherent complexity of cities, and the challenges associated with gathering urban-level data, my work evolved to focus on the use of spatial measurements to examine material and energy flows in urban areas (Figure 1.3-b). As a result, I have worked on developing methods to analyze material and energy use at the neighborhood scale (Quinn and Fernández, 2010, 2011b; Quinn et al., 2011), and to visualize how these resources are used (Quinn and Fernández, 2011a). This analytical approach tries to overcome the systematic problem that exists in many urban assessments where the analysis is not replicable. By providing the source-code to the analysis, the assumptions and methodology can be evaluated and reviewed. This work has involved land-use classification (Wiesmann and Quinn, 2011) and the development of several other script-based analytical approaches using geospatial analysis techniques. These methods were applied to the data sources summarized in Figure 1.3 to calculate the resource intensity.

I identified a need for the results of this analysis, due to the lack of urban data currently available. To facilitate sharing the results of this work an interactive web-tool was developed, and the data was also published as a Web Map Service (Open Geospatial Consortium, 2012b) and a Web Coverage Service (Open Geospatial Consortium, 2012a). The results of this analysis can be viewed using a web-browser, or the data can be accessed directly using desktop GIS software.

1.1 Resource Intensity of Cities

The objective of this work is to develop a method of analysis that can assess the material and energy intensity of different types of urban areas in a generalizable way. The motivation of this work is focused on relating top-down work that examines cities at a global scale, and bottom-up localized studies that consider urban attributes at the building level. Performing a standardized analysis of each urban system means that the resource intensity of each area can be compared.

Three challenges to analyzing resource consumption in cities are the lack of useful available data, the spatial resolution in which information about urban resource consumption is gathered and the computational power needed to analyze spatial data. While remote sensing data and geographic information systems (GIS) enable analysis at a high levels of resolution, a fundamental challenge arises in relating calculations that describe the structure of the system, to the actual functioning of the system. These barriers to analysis have been significantly reduced with increased computing power and more easily available data sources. Although it is becoming easier to perform analyses of urban resource intensities at high spatial resolutions, there are still many technical challenges as standard desktop tools are not capable of handling such large datasets.

This work attempts to relate the structure of the urban form to the overall urban system performance by examining discrete units of the spatial structure of each city. A goal of this research is to assist with the identification of upper and lower boundaries of resource intensity measures that are likely to occur in a city. These boundaries can then provide the basis for policy makers to identify reasonable urban resource efficiency targets.

1.2 *Analysis*

The resources examined in this work are solely resource intensity measures that can be directly attributed to the residential built environment, summarized in Table 1.1. No socio-economic measures were considered. The exclusion of socio-economic measures enables the analysis to be more generalizable as it relies on fewer measurements.

Table 1.1: Resource intensity measures attributable to the residential built environment.

Material	Energy
Road Infrastructure	Transportation
Buildings	Gas
	Electricity

Over the course of this work, it was observed that identifying patterns to relate these heterogeneous data sources would in itself be a valuable contribution to the literature on urban resource flows due to the shortage of available data. To achieve this, several specific approaches were developed to perform this analysis. These are summarized in Table 1.2.

In this thesis, I am also clearly documenting the procedures that I used to perform this analysis and providing samples of the code used in the Appendices. This is in part due to the fact that a heterogeneous grouping of software tools was used, but also due to the fact that processing large volumes of spatial data has many challenges. Many of the software tools used in this work are open-source, and have not been widely adopted by non-

Category	*Level of Abstraction*	*Description*
Land Cover	urban level	Statistical method of predicting land-use using remote sensing data; R-package *rasclass* developed by Daniel Wiesmann and David Quinn. The output from this R-package was not used in this work.
Urban Categorization	urban level	Categorizing cities using population and infrastructure density gradients; identifying upper and lower boundaries using discrete measurements of infrastructure and population.
Road Width	neighborhood parameter	Prediction of average road width.
Dwelling Units	neighborhood parameter	Prediction of number of dwelling units per area.
Building Area/Heights	neighborhood parameter	Prediction of residential built area and average building height, per block group.
Neighborhood Typologies	neighborhood parameter	Identification of neighborhoods with similar physical characteristics.
Building Material	resource consumption	Estimate of material used to construct buildings.
Road Material	resource consumption	Estimate of material used to construct road infrastructure.
Transportation Energy	resource consumption	Estimate of energy use per household due to private auto travel.

Table 1.2: Summary of analytical contributions.

technical specialists for the purposes of urban resource analysis. It is hoped that clear descriptions of how these tools were used will contribute to their documentation and adoption.

1.3 *Data*

This work assumes that there are many detailed data sources that can be used to describe the urban form. As will be discussed further in Chapter 3 and 4, in many cases comprehensive datasets are not available. However, it is still important to iterate through the process of analysis to identify specifically what data is needed and how it should be measured to enable systematic assessment of neighborhoods. It can also be assumed that future datasets will be more detailed and accurate, as it is becoming easier to gather and analyze data. This project succeeded due to a well-designed interactive database, user-generated data from volunteers with an effective quality-control mechanism and public domain datasets. The second reason for this approach is that it is an important academic contribution to understand if data can be interpolated, or applied from one case to another, and where the limitations of this approach are. Due to global data constraints, this analysis used data from cities in the USA and the UK. A future goal of this general research is to explore whether the patterns observed in these industrialized countries also holds true for other cities.

It is important to consider how data is collected from a human organizational perspective. In this work (Figure 1.3), *Data Complexity* was considered to be a combination of the accessibility of the data (based on the experience of conducting this research) and the sophistication associated with gathering, storing and distributing the data. All data used in this work was from a top-down organization at the city or national level.

This work examines how these data sources can be related to each other in a cumulative way. In Figure 1.3, *Predictive Error* is considered to be the difference between the actual resource use and the predicted value using a combination of these data sources. The data types used in this work are summarized in Table 1.3.

Category	Type	Description	Social Characteristics of Data
Remote Sensing	raster	This data can be categorized into two groups; passive and active. Passive remote sensing detects natural radiation that is emitted and is typically used by satellites to identify vegetation and urban areas. Active remote sensing emits energy in order to scan an object and detects that object based on how the energy scattered. A commonly used active remote sensing method is LiDAR (Light Detection and Ranging) which is used for topological identification and geometric measurements of buildings.	Not influenced by administrative measurements; originally used to identify vegetation change, now used for land-use classification. Globally available, free at various levels of resolution though high resolution data is typically not free.
Population Data	survey, vector	Census data is the most common record of the population. While this describes where people are (depending on the country the frequency and fineness of data varies); it can be closely linked to urban form measurements.	Globally available, usually free but some countries charge for the administrative boundaries. As this is collected locally, it can be manipulated and depending on the political motivations may be biased.
Road Networks	vector	Center-lines, road-widths; lanes.	Ubiquitous global data source. After geographic boundaries, road centerline data is the most widely available spatial dataset and available without restriction from OpenStreetMap and VMap.
Building Geometries	vector	This can consist of 2-D building outlines; 3-D building shapes and survey data describing the characteristics of the buildings.	2D building data is becoming more commonly available, but the availability of data depends on the culture of sharing information. OpenStreetMap is a good barometer (and source) of the availability of this data.
Building Level Data	surveys, utility records, smart meters	Detailed information about the individual building. This type of data is difficult to acquire.	Privacy concerns are frequently cited as a cause for restricting access to these datasets, but it is likely that there are several factors, including the fact that the data is perceived to have commercial value.

Table 1.3: Summary of data types used.

1.4 Visualization

Effective visualization strategies are important so that we can gain an understanding about complex systems, and develop an intuition about the performance of urban systems with multiple dimensions. The goal of visualizing urban resource intensity data is to illustrate how material and energy is being used within a city, and to help citizens understand the resource intensities associated with the urban area that they inhabit. Using a web-based approach reduces the access barrier to accessing this tool, as there is no proprietary software required and the user is likely to have some familiarity with a web-based map. The interface provides a user-friendly, cross-platform way of exploring urban areas, so that a user can learn what parameters influence resource consumption at the neighborhood scale, using a location that they are familiar with to provide a frame of reference.

The motivation for developing this tool was due to the lack of suitable analysis tools available to the planning community that enable quantitative analysis of material and energy use in urban areas. I hope that the provision of this information can influence planning decisions positively as more participants consider the functioning of their neighborhoods from a resource intensity perspective. This visualization tool demonstrates some of the trade-offs associated with specific urban configurations and illustrates the variation of material and energy per capita or per household within a city. Currently, 40 US cities can be explored using this tool which is accessible at urbmet.org.

This tool is structured in two components, with two different target audiences. One component, consisting of the processed data, is intended to be used by an expert, such as a planner, engineer or architect who is familiar with GIS. The other component focuses on visualizing the results, in a user-friendly way. The visualization component is intended primarily as a pedagogical method to facilitate discussion about urbanization patterns (though preliminary analysis is possible, and the user can generate *pdf* reports). A web-based visualization tool democratizes access to resource intensity information, and has the potential to engage non-experts in the discussion about urban sustainability.

2 Literature Review

In this chapter, I start with an overview of global trends that are relevant to this work, considering global trends (Section 2.1) and then review several methods of analysis that have been used to analyze urban systems. The broader context of these methodological approaches is that cities are no longer regarded as disordered systems but rather a form of urban complexity that has its basis in the regular ordering of size and shape across many spatial scales. These methodologies are applicable to the study of urban resource usage, as they can be used to identify patterns of behavior that can assist with improved levels of urban sustainability. This summary describes a growing body of work that is contributing to the development of an integrated theory with regard to how cities develop function and evolve (Batty, 2008). I discuss how a variety of approaches have been applied to urban systems, and identify what the strengths and weakness of these approaches are.

I introduce the conceptual approaches used to analyze resource flows in urban areas, by discussing the fields of *Industrial Ecology* and *Urban Metabolism* in Sections 2.2 and 2.3. I then discuss more applied methods of analysis, namely *Complexity Theory* and *System Dynamics* in Sections 2.4 and 2.5. Both *Complexity Theory* and *System Dynamics* have been used to examine mechanisms within cities that are relevant to the flow of resources.

I then discuss approaches from the field of *Land Use and Transportation* in Section 2.6 and summarize analysis methods from relevant work. Finally, I consider *Urban Information Systems* and discuss aspects of spatial data analysis and planning-support systems in Section 2.7, describing the technical issues associated with this analysis.

Overall, I consider this work to be a multi-disciplinary endeavor that is not based in any specific methodology. I have applied and combined a variety of established methodologies with the goal of contributing to a deeper understanding of urban systems. I conclude this chapter with a summary of work that is directly relevant to this research (Section 2.8), and formally state the research objectives that I wish to address in this work.

2.1 Global Trends

When the overall global context of urbanization is considered, this work is particularly relevant. Due to rapid urbanization throughout the world (Weisz and Schandl, 2008; Krausmann et al., 2009), and the global shift from an agrarian socio-metabolism to an industrial socio-metabolism (Krausmann et al., 2008), it is projected that there will be a dramatic increase in consumption of material and energy. Demographic trends project that the majority of population growth will be in cities (United Nations, 2011), suggesting that by 2050 there will be an additional 3 billion urban dwellers, with the majority of this growth occurring in developing countries (Figure 2.1).

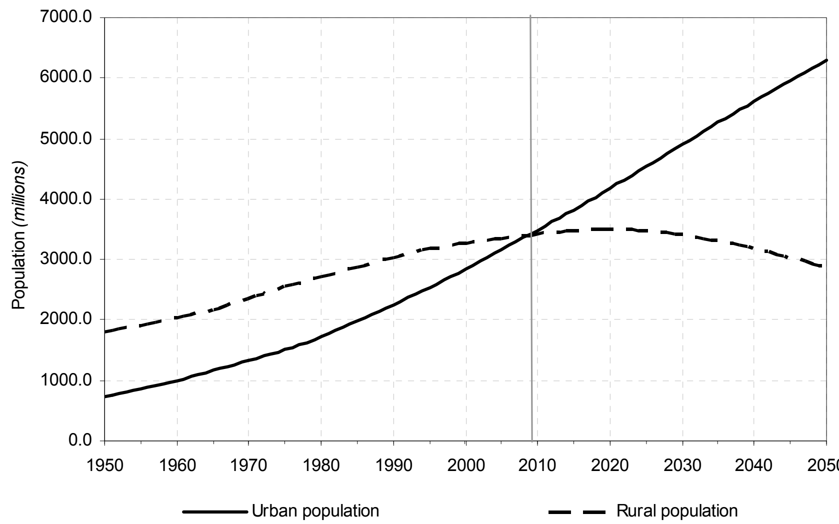


Figure 2.1: Global historic and projected urban and rural population. Image from United Nations (2010).

Despite the significance of resource use in urban areas, Decker et al. (2000) observe that there have been few cross-cutting comparisons applied to the development or growth of mega-cities. Decker et al. (2000) consider mega-cities from the perspective of biological metabolism and ecosystem succession, and emphasize that little work has been done which considers cities from a systems perspective. At the global scale, changes in resource usage are examined by Krausmann et al. (2008, 2009). Weisz and Schandl (2008) discuss the importance of reaching harmonized standards and approaches for countries throughout the world so that urban resource can be characterized more accurately.

Countries, which have changed their resource consumption patterns dramatically over the last several decades have been examined extensively (Kovanda and Hak, 2008; Niza and Ferrão, 2006; Schulz, 2007) with the general (unsurprising) observation that economic growth results in increased resource consumption. However, the effects of urbanization are not limited just to increases in the bulk flows of material and energy. Erb et al. (2008) and Pataki et al. (2006) consider urbanization within countries from a spatial perspective, and examine the consequence of land-use change and how it relates to national carbon flows, which also has implications at a global scale.

McGranahan and Satterthwaite (2003) consider global urbanization and the sometimes opposing forces of *development* and *environment*:

> Although rapid urbanization is seen as a problem, it is generally the nations that have urbanized most in the last 50 years that have the highest average life expectancies or the largest increase in their life expectancies.

Due to economic disparity, McGranahan and Satterthwaite (2003) argue that we need global regulation to prevent cities from exporting their problems to poorer countries. Brand (2009, 2006) shares McGranahan and Satterthwaite (2003)'s positive perspective on the consequences of development, arguing that slums have a positive overall effect as they are areas with intense economic activity while being extremely resource efficient. McGranahan and Satterthwaite (2003) argue that cities need to consider the impact of their resource use on the ecosystem around the city, which is typically outside the administrative boundary. This is also discussed by Newman and Jennings (2008) who explores how cities should be organized based on principles from nature.

2.2 *Industrial Ecology*

Industrial Ecology has been described by Coelho and Ruth (2006) as the

> systematic analysis and design of human activities and the environment with the implicit goal of optimizing the total industrial cycle: from raw material input through the creating of a finished product to waste output and back to the economy

This approach views human systems as being analogous to natural systems. As the biological analogy is typically applied to individual, discrete organisms, Fischer-Kowalski (1998) suggests that that the concept of metabolism, needs to be expanded beyond the material and energetic flows associated with living things. Fischer-Kowalski (1998) suggests that this description should encompass all of the interactions associated with human society and should include both the resources used by the individual and by society. This proposed concept is termed 'societal metabolism'.

The background of material flow analysis (MFA) is traced by Fischer-Kowalski (1998) and Fischer-Kowalski and Hüttler (1999). Fischer-Kowalski and Hüttler (1999) suggest that MFA may become one of the most important methodologies for describing and analyzing environmental problems. Approaches for quantifying the metabolism of physical economies are discussed by Daniels and Moore (2001) and Daniels (2002). Daniels and Moore (2001) review nine approaches that can be used to analyze resource and energy flows and discuss how these flows can be considered in relation to the environment. Daniels and Moore (2001) define MFA in the following way:

> the systematic physical tracking of material flows associated with the socioeconomic metabolism

where flows are considered to refer to both energy and materials. Daniels (2002) provides examples of these approaches to illustrate how such calculations can be made, schematically. While Daniels (2002) considers this methodological approach to be useful for aggregating data, the authors are not suggesting that it provides a comprehensive evaluation of sustainable development.

Extensions to the MFA approach have been undertaken by Bouman et al. (2000) who compares three different approaches that can be used to analyze material flows. These approaches are *Substance Flow Analysis* (SFA), *Life-Cycle Analysis* (LCA) and *partial-equilbrium models* (PEA). The goal of this comparison is to consider how these approaches can best be combined, by examining elements of each that are complementary. Bouman et al. (2000) conclude that using a combination of approaches is useful, but one significant difficult with the LCA approach is related to the assumptions around the life-span of a product. As a result the magnitude of an impact can be increased or decreased depending on what the product life-span is.

A more complex approach to the analysis of resource flows, is developed by Goßling-Reisemann (2008) who examines how thermodynamic measures (exergy and entropy) can be directly applied to material and energy flows, so that the degradation of resources can be measured. He argues that the standard measures of throughput and resource use (LCA, MFA, SFA) do not consider how the resource has been changed after it has passed through the boundary that is defined. Goßling-Reisemann (2008) proposes that entropy be used as measure of the potential loss of utility and argues that the more irreversible a process is, the more potential utility is lost. Goßling-Reisemann (2008) explores whether exergy is an appropriate measure for ecological costs (defined as the cumulative depletion of non-renewable exergy resources). Exergy and entropy measures can be used in two ways (1) to define the amount of waste being produced and (2) to identify thermodynamic inefficiencies in a process. Goßling-Reisemann (2008) concludes that an entropy measure is appropriate as it quantifies the transformation process.

Frequently, environmental regulations focus on the symptoms of environmental problems, rather than tackling mechanisms within the system that result in this behavior. Wernick and Irwin (2005) suggest that MFA is an approach that can allow regulatory structures to move from being reactive and only responding to problems when they are identified, to one that considers the overall system with the ability to identity the source of the problem.

While these *Industrial Ecology* approaches are useful to assess the static state of a system, they are less focused on modeling the driving mechanisms within the system. Hence, an overall weakness with MFA is that there remains a gap between the methodology, and relating it to social or economic processes. While MFA can be useful for identifying problems within in a system, there is not an explicit link between MFA and how it can be related to regulatory policies.

2.3 *Urban Metabolism*

The concept of urban metabolism was first described by Wolman (1969) as a means to estimate material and energy flows in cities. This concept has been further advanced by Fischer-Kowalski (1998) and Fischer-Kowalski (1998) who examined the concept of societal metabolism. This concept considers the material and energy requirements necessary for human society to function. Urban metabolism can assist with identifying material and energy flows, and these flows can then be used to characterize cities. Decker et al. (2000) describes the similarities between cities and living organisms, as cities 'transform raw materials, fuel, and water into the built environment, human biomass and waste'. Studying the urban metabolism provides an understanding of these transformations, as well as enabling measures of resource efficiency. An example of such a measure is to consider the 'degree of circularity of resource streams' which was suggested by Kennedy (2007). Urban metabolism analyses inform how well equipped the local, or global, capacity of the planet can fulfill our demands (Giradet, 1992). Fischer-Kowalski and Hüttler (1999) defined urban metabolism in the following way:

Sum of Material and Energetic Inputs = Sum of Outputs + Changes in Stock

Examining the metabolic flows of cities allow us to consider at what rate resources are being depleted and at what rate waste is being produced. Huang and Hsu (2003) use a more specific definition of the urban metabolism, considering it to be the process of transforming all the materials and commodities for sustaining the city's economic activity.

Niza et al. (2009) study the urban metabolism of the city of Lisbon and propose a methodological approach for urban material flows. The mix of materials that Niza et al. (2009) consider are biomass, fossil fuels, metals and nonmetallic minerals. One of the challenges faced by Niza et al. (2009) was to calculate the material flows of a city, where no formal borders exist. A general problem of urban metabolism studies is defining methodologically standards, even in terms of defining the spatial boundary. Without this standardization it is difficult to compare results. The issue of consistency and repeatability is discussed in detail by Weisz and Steinberger (2010), as well the issue of urban boundary definitions.

Kennedy et al. (2011) provide a comprehensive review of the field of *Urban Metabolism* and consider the future direction of research which is anticipated to integrate social, health and economic indicators into the urban metabolism framework, while considering resource flows. Characterizing the resource consumption of urban areas as a complex set of metabolic functions, requires a broad perspective and a diversity of methods. While I consider urban metabolism to be a useful framework for identifying resource flows in cities, the actual tools need to perform these types of analyses are still at an early stage of development.

2.4 Complexity Theory

One approach to examine cross-city patterns has been to apply methods from the field of complexity theory. Historically, complexity theory has been employed in the analysis of urban systems in several ways. Large scale integrated computational models, the development and application of which have ebbed and flowed since the seminal work by Lowry (1964), have entered into an era marked by an increasing focus on micro-simulation, aided by enhancements in computing power, behavioral theory, and econometrics (Wegener, 2004). While work has been done on developing an integrated theory describing how cities evolve and develop from a complexity perspective (Batty, 2005; Bettencourt et al., 2007), there has been less emphasis in the literature on relating complexity patterns to urban resource-efficiency. While global city size (and scaling) has been examined thoroughly, as well as population density gradient patterns; less work has focused on identifying patterns at the neighborhood scale that are common for all cities. Batty (2008) considers complexity theory applicable to city planning as it can be used to consider urban processes from a variety of scales, removing the need for top-down plans. West et al. (1997) examined scaling patterns in cities, and considered the parallel scaling relationships between cities and biological organisms.

The size distribution of cities throughout the world has been frequently studied (Decker et al., 2007) with the overall metropolitan size considered to be influenced by a combination of factors, such as transportation technology (Hanson and Giuliano, 2004), climate (Glaeser et al., 2001), innovation cycles (Bettencourt et al., 2007) and randomness (Gabaix, 1999). Glaeser (1998) refers to two opposing forces that influence the urban size; agglomerating forces which cause concentration in density (reduced transportation costs, information spillovers, learning), and congesting forces which disperse this density (such as living and pollution costs, and crime). In addition to identifying physical scaling relationships, Bettencourt et al. (2007) refer to the increasing returns to scale with regard to innovation in cities (measured by patents, inventors, R & D), while Gabaix (1999) explains the basis for city size distribution (often referred to as Zipf's law) due to random growth patterns. Within the city, a relationship between the distance from the central business district (CBD) and population density has been observed for some time; this is referred to as a population density gradient, and has been discussed in the literature (Clark, 1951; Ingram, 1998; Bertraud, 2004; Marshall, 2007). Ingram (1998) observed in cities throughout the world that the population density gradient is becoming flatter over time, and Marshall (2007) observed that 'newcomers to urban areas occupy about twice the land area per capita of existing residents'. This is important when considering future urban growth as the physical area of the city does not grow linearly as a function of the number of people.

More recent work which examines intra-city scaling has been done by Bettencourt et al. (2007) who observed that as cities grow in size, physical networks tend to grow more slowly, due to economies of scale. As a result, the physical infrastructure required does not increase as quickly as the population. Bettencourt et al. (2007) review some physical scaling parameters that have been observed within cities, and discuss how these can be interpreted. The key observations are with regard to what is scaling sub-linearly, and what is scaling super-linearly. Information (measured by patents, inventors and R &D) scale with values greater than 1, which means that the larger the city, the more information is being generated. Interestingly some physical parameters scale at 1 (housing, water consumption) while others are < 1 (Road Surface). Bettencourt et al. (2007) observed the Road Surface scaling parameter to be 0.83 (where Road Surface, R, and Population, P, are related using the relationship, $R \approx P^{0.83}$). This means that as a city becomes larger the amount of road surface is proportionally smaller. Samaniego and Moses (2008) also examined road scaling at a macro-level for a range of cities.

While there has been a substantial body of research that has focused on the behavior of cities from a complexity perspective (Batty, 2008; Samaniego and Moses, 2008; Bettencourt et al., 2007; Batty, 2005), there has been little work done which bridges the gap between the emergent behavior of cities and analyses at the neighborhood scale that is directly relevant to policy-makers or developers.[1] Although representing the formation and dynamics of cities at the micro-scale can be a rigorous way of describing the evolution of cities, (Wu, 2007) observes that the policy relevance is not immediately obvious from such an approach. As a result, it can be difficult to direct relate the results of complexity analysis to policy as there are many abstractions necessary to make the work applicable.

[1] Here, I consider policy-makers to include those involved in the development of regulatory guidelines at a local scale; while developers are considered to be those who operate within the regulatory framework.

2.5 *System Dynamics*

System Dynamics (SD) is a methodology for understanding the dynamic behavior of complex systems, by describing systems as stocks and flows, and developing chains of causality that identify positive or negative feedback loops. It is a useful approach for considering causality within a system and also for explaining non-linear patterns. SD is used in situations in which a particular phenomena of dynamic complexity has been identified and is in need of a deeper explanation to understand unintended consequences, time lags and other unanticipated results (Sterman, 2000).

SD was first used to analyze a city by Forrester, who developed a model called *Urban Dynamics* (Forrester, 1969). This model, which considered the city aggregated into several components, explored the influence of various policies, and was further developed by Alfeld and Graham (1976). Due to philosophical disagreements and technical limitations, *SD* was not widely accepted by the planning community as a valid approach (Alfeld, 1995). However, more recent applications of *SD* relating to planning can be seen in Emmi et al. (2005), the interaction between urban and ecological systems (Deal, 2001) and the prediction of urban growth patterns (Güneralp and Seto, 2008; Han et al., 2009).

Two examples with direct policy applications can be seen in Han et al. (2009) and Güneralp and Seto (2008). Han et al. (2009) create an SD model of urban growth in Shanghai, China, illustrating that it can closely predict some future patterns of growth, while Güneralp and Seto (2008) examine urban growth in Shenzen, China. Both of these cities are rapidly growing, both in terms of population and land-area and similar to the earlier urban dynamics model a sectoral approach is used.

One of the difficulties regarding urban analysis, stems from the limited interaction between fields of study that use specialized scientific analysis such as ecology and climatology, and how these results are communicated to policy-makers (Costanza and Ruth, 1998; Andrews, 2000). There is a lack of inter-disciplinary work within these formal sciences, as well as a lack of a consistent framework (Pataki et al., 2006). Coelho and Ruth (2006) propose a framework similar to an SD model structure, while Pickett et al. (2001) describe such an approach in greater detail. Ruth and Coelho (2007) identify the challenges that cities face with the increased risk due to climate change, and what is necessary to be considered for greater urban resilience.

There is a consensus from various fields of study (ecology, resilience and policy) that the feedbacks from ecological systems where cities are located, should be considered endogenous, rather than viewed as exogenous factors. This is addressed by Newman et al. (2009) and Clifton et al. (2008) who view it from a resource constraint issue; by Pickett et al. (2001) from an ecology perspective; and by Pickett et al. (2004) and Krasny and Tidball (2009) who view the importance of making city-dewellers more aware of the role in the ecosystem through education.

Urban resilience is one approach to considering how cities relate to the ecosystem within which they are located. Diamond (2005) provides several examples of societies that collapsed due to their failure to adapt. The collapse of the residents on Easter Island a much studied System Dynamics model illustrating an overshoot and collapse scenario (Sterman, 2000), due to society's inability to change their behavior despite ecological feedback with regard to the carrying capacity and to anticipate the future. These tales can be considered useful fables for our current society as we have the ability to learn from previous tragedies. Diamond (2005) does provide a positive example of how societies can adapt and has some positive examples of the forms this adaptation can take.

How cities respond to slow-moving phenomena that can undermine their survival is a critical challenge (Newman et al., 2009). Newman et al. (2009) considers a resilient city as one which can respond to natural or to human disasters. Considering this from the perspective of how much external energy is required for the city to function[2], Newman et al. (2009) suggests that cities that are more sustainable, also are more resilient, considered from the perspective of changes in oil prices and availability. Although this takes a long-term view of a diminishing resource (oil), Hurricane Katrina provides an interesting example of a lack of resilience with regard to transportation in the short-term. After Hurricane Katrina, there was an inadequate public transit system, and all roads were filled with cars which lead to chaos; all of which could have been diminished if there was less of a reliance on one mode of transportation, in this case automobiles (Newman et al., 2009). Although Folke et al. (2005) suggest that a 'resilient social-ecological system may make use of crisis as an opportunity to transform into a more desired state', a social-ecological system that is not resilient, may simply fail.

[2] Atlanta requires 782 gallons of gasoline per person while Barcelona requires 64 gallons of gasoline per person

A common theme amongst the literature that has reviewed societies that collapsed (Diamond, 2005; del Moral and Walker, 2007; Newman et al., 2009), is that systems of governance failed to respond to signals from the ecological world which demonstrated increased vulnerability. The main conclusion from the *Stern Report on Climate Change* (Stern, 2006), is that the benefits of strong, early action on climate change considerably outweigh the costs. Although there is an awareness of the future problems that humanity faces due to climate change (Stern, 2006) there has been little coordinated effort at a global scale. To quote del Moral and Walker (2007) again, 'The will of a society can only be derived from a political and economic calculation that demonstrates that action is more valuable than inaction.'

Meadows and AtKisson (1997) examine the delays in societal response to environmental problems (Table 2.1). They state that 'a system that experiences change faster than it can respond is a system out of control' and highlight the fact the challenge for humans to solve problems technologically, is due to the rate at which problems are generated.

	Clean Air Act [years]	Toxic Waste Dumping [years]	Ozone Depletion [years]
Time from clear signal of problem to meaningful law	22	38	58
Time from first law to improvement	20	17+	50 - 80

Table 2.1: Time delays in responding to environmental problems.

It is important to consider how long it has taken for national and international consensus to be reached on these issues, particularly when estimating the magnitude of the effects. Diamond (2005) explores why societies have been so slow to act in the face of environmental problems which have resulted in disaster, despite being aware of the need for corrective action. Diamond (2005) argues that it is a combination of three factors:

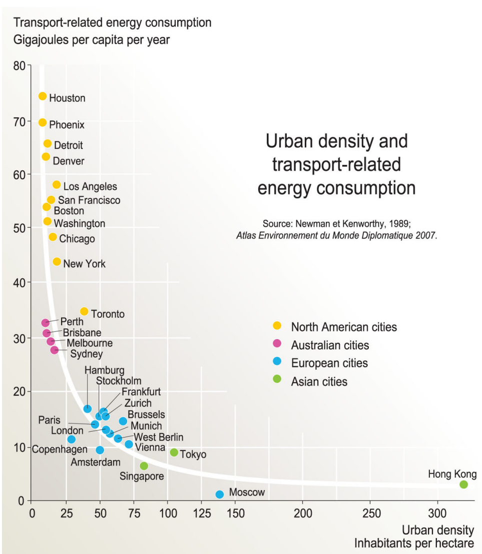
- *False analogies* - people base their view of the future on the past.
- *Landscape amnesia or creeping normalcy*: it takes a long time for environmental degradation as it frequently occurs slowly. If you live in such an ecosystem, you do not notice the small changes occurring.
- *Tragedy of the commons*: First observed by Hardin (1968), this problem occurs when the individual benefit is greater than the societal cost, and the individuals within the system try to prevent regulations that will affect their short term interests.

Similar to technological lock-in, cities need to consider how policies will influence future planning and market-decisions. One of the hallmarks of resilience is diversity and adaption to changing environments. Considering resilience from this perspective, is particularly relevant when the long-time scales that are required for infrastructure investment are considered. Path dependence is an inevitable consequence of the widespread adoption of a particular technology, and should be considered in a city context, as it is important to consider how easily alternatives can be adopted in a resource constrained urban environment.

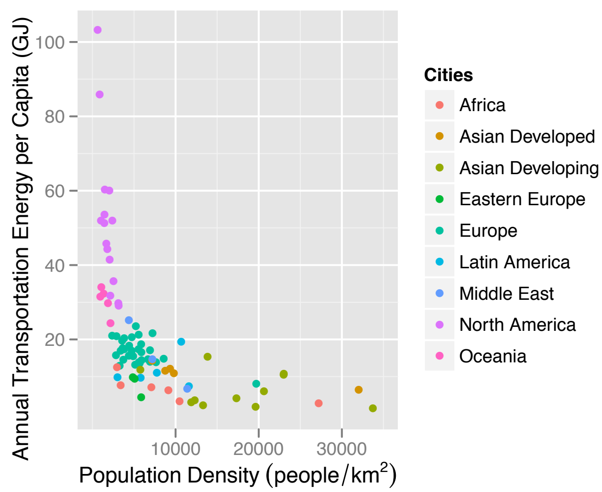
In contrast to *complexity theory*, *SD* is a top-down approach to modeling the system, and as a result it is heavily influenced by the decisions that the model-builder makes. Forrester failed to use a participatory approach in his initial efforts, which alienated the planning community. I believe that *SD* would be beneficial for participatory planning as it can be assist in the simplification of complexity, as well as illustrating unintended consequences and long-term effects that are not immediately obvious. In addition, it is a useful tool or the integration of other model approaches and it facilities the understanding of a system from a high-level prospective.

2.6 *Land Use and Transportation*

In the field of Land Use and Transportation, much published work has compared the functioning of urban areas to each other, using city-level average measures. This can be seen in the much-cited work by Newman and Kenworthy (1989) that explores the relationship between population density and energy; a similar pattern of behavior can be observed in the Millennium Cities Database (Figure 2.2). Bento et al. (2005) use measures of the urban form and public transportation supply in 114 urban areas, and conclude that population centrality, jobs-housing balance, road density, and city shape significantly influence automobile distances traveled. Lefèvre and Mainguy (2010) explore the relationship between population density and transportation modes, and consider how the general spatial structure of cities influences resource consumption. Hankey and Marshall (2010) examine how sprawl and infill development influence daily travel patterns and annual vehicle kilometers travelled (VKT).



(a) Figure from UNEP based on data from Newman and Kenworthy (1991)



(b) Millenium Cities Database (Vivier, 2000)

Figure 2.2: Energy for transportation.

More recent work has examined relationships between urban form and transportation energy Holtzclaw (1994); Crane (1996), while the effects of New Urbanism design strategies and travel behaviors are linked to resource efficiency Cervero and Kockelman (1997). Ingram (1998) identified a similar relationship to the work by Newman and Kenworthy (1991) using a 'residential floor area per person'. Although the work by Newman and Kenworthy (1991) has been criticized, few alternatives have yet been provided using an improved methodology (Zhang, 2004). Mindali et al. (2004) disagree with

The discussions and references in this section are based on a class taught by Professor Chris Zegras at MIT, Spring 2010.

the conclusions from Newman and Kenworthy (1989) and suggest that the relationship between population density and transportation energy is more complex than originally suggested. Boarnet and Crane (2001) describe many of the subtleties associated with understanding VKT and urban form, and the challenges associated with developing policies to reduce VKT.

While Vehicle Kilometers Traveled (VKT) is considered to be a primary indicator of transport system performance and allows us to account for trip-chaining by automobile, it misses the phenomena of possible local importance Zegras (2007), and it does not consider the type of vehicle (from heavy truck to hybrid), emissions or congestion. Nevertheless, in a data-scarce environment VKT can be converted into a measurement of energy, which is a way of relating transportation to resource consumption. Although Neuman (2005), citing Bouwman, argues in *The Compact City Fallacy* that there is not a significant difference between transportation energy consumption between urbanized and rural areas, the results from Cervero and Kockelman (1997), Cervero and Duncan (2006) and the comprehensive review by Ewing and Cervero (2001) suggest otherwise, as does the work in this thesis. The built environment is thought to influence travel demand along three principal dimensions, density, diversity, and design Cervero and Kockelman (1997). I review these three factors of the built environment.

Density: Density, in its most simple form is a measure of a unit per area. A simple measure for describing the structure of a city is the gross population density, with more refined measures considering people per area of built space, block density, street density Dill (2004). Sprawl, although difficult to define[3], can be considered to be an urbanized area with a low population density. Ewing (1997) discusses the challenges of measuring sprawl, which illustrates the difficulty of applying a qualitative judgment to a quantitative measure.

[3] For example, Ewing (1997) refers to 17 varying definitions of sprawl.

Diversity: Neuman (2005) considers diversity to be an indicator of health, whether for an ecosystem, urban community, or organization. Clifton et al. (2008) suggests that urban economies are more stable when diverse which is consistent with the informal argument made by Jacobs (1965). Schelling (1978) identified some properties of self-organization in a city, with the interesting conclusion that a diverse community has delicate balance (a small zone of stable equilibrium) for this diversity to be maintained. Levinson and Krizek (2008), emphasize the importance of achieving sufficient diversity in a neighborhood, and for the urban form to be sufficiently flexible so that it can adapt to different purposes over its lifespan making it easier to ensure that there is diversity. They consider a permeable neighborhood like a grid, easier to adapt than a tree layout.

Design: Levinson and Krizek (2008) suggest that Brand's approach[4], and his six principles *Site, Structure, Skin, Services, Space* and *Stuff* should be expanded to include where the building is located by considering a sev-

[4] Described in How Buildings Learn Brand (1995), Brand (2006) describes the rate at which cities are renewed as follows: in Europe, cities replace 2 to 3 per-cent per year of their material fabric (buildings, roads, and other construction) by demolishing and rebuilding it. This means, in effect, that a wholly new city takes shape every 50 years.

enth concept, the *Street*. Cervero and Kockelman (1997) consider aspects of streets (network type, four-way intersections), pedestrian and cycling provision (side- walks, cycle-lanes), and site-design to be important design elements when considering the influence of neighborhoods on transportation. Levinson and Krizek (2008) quoting Scheer, argue that urban designers should focus on slow-moving elements such as space and stuff, rather than building design. Dill (2004) reviews some other design measures such as block length, block size, grid pattern, pedestrian route directness, effective walking area as well as mentioning some quantitive approaches that are used from other disciplines such as geography and landscape-ecology for examining networks. Specific network design measures that are also considered are connected node ratio and the link-node ratio.

2.6.1 Measuring Urban Form

There is a growing importance associated with understanding quantitative measures of the urban form, to justify interventions or policies. Jacobs (1965) argued that the physical and social fabric of cities was destroyed by gentrification and highways, but had only qualitative and anecdotal data to support her position. I believe that access to technology (such as GIS, databases, and GPS) makes it easier to analyze and justify Jacob's arguments. Clifton et al. (2008) provide a comprehensive overview of urban form measures, and draw on other related disciplines such as landscape ecology and transportation planning in addition to considering some of the scaling patterns discussed earlier, such as the population density gradient.

Clifton et al. (2008) concludes that there has been much progress in developing and computing measures of the urban form, which can describe its behavior at a variety of scales. The scale of measurement is influenced by the discipline as landscape ecologists consider the area of an ecological region, while urban designers consider the urban form from the street or neighborhood. The issue of scale and unit definition[5] is a common problem in geography with regard to accurately representing and analyzing data by using an appropriate spatial scale Horner and Murray (2002). Horner and Murray (2002) explore how the number of traffic analysis zones influence the estimated value (in this case, excess commuting) with the conclusion that data should be as disaggregated as possible for this type of analysis.

[5] This is also referred to as the Modifiable Aerial Unit Problem (MAUP).

Aesthetics, safety and other socio-economic factors are harder to capture in abstract measures of the built environment, as well as measures to quantify the functioning of the physical form. Many of the computational approaches that try to explain why the urban form is structured in a particular way frequently are validated by comparing only the physical form or parameters that describe the form (Crucitti et al., 1991; Gastner and Newman, 2006) while economic models frequently fail to consider this aspect (Lucas and Rossi-Hansberg, 2002). A modeling approach that considers both economic factors and urban form is UrbanSim (Waddell, 2002).

At the city-scale, rather than using the population density gradient as means to approximate the urban form, Bento et al. (2005) develop a population centrality measure which weighted the distance that people are from the CBD. Then, they examine the distribution of employment by calculating a balance between jobs and housing (similar to the Gini coefficient calculation) to provide a measure of sprawl. They conclude that the individual effects are small (a 10% increase in population centrality results in a 1% reduction in the probability that a worker drives to work), but that the combine effects of population centrality and public transportation can have a significant combined effect on travel demand.

Zhang (2004) discusses how the land-use mix can influence travel behavior, although the influence on some relationships is still ambiguous. As travel is considered to be a derived demand (Banister, 2008), the urban form influences transportation patterns, but quantifying this influence and controlling for confounding factors is difficult. One of the difficulties, with regard to estimating whether the built environment (BE) is a causal factor in influencing mode selection is discussed by Zhang (2004). This is due to the difficulty in parametrizing the BE into separate components that can be examined, the lack of consistency in previous studies examining this Crane (1996), the issue of self-selection Handy et al. (2005, 2006); Mokhtarian and Cao (2008) and the problem of appropriately measuring human behavior. Mokhtarian and Cao (2008) explore this relationship in detail, concluding that more studies considering longitudinal data are needed to fully explore causal relationships.

Handy et al. (2006) explore the issue of self-selection further by examining eight neighborhoods, and find that increases in accessibility, particularly close proximity to potential destinations such as shops and services, as being most important factors which will increase walking. They also observe that nearly 32% of the respondents say they are walking more now than they were before they moved or one year ago, while fewer than 18% percent say they are walking less. Mokhtarian and Cao (2008) also consider self-selection, and examine whether this can explain the different patterns of behavior between suburban residents and residents in traditional neighborhoods. They examine the criteria that have been used to measure these parameters and explore whether observed patterns of travel behavior can be attributed to the built environment itself, or to residents choosing to live in these locations as they are attracted to these travel modes.

Newman and Kenworthy (1989, 1999) are oft-cited for the observation with regard to gasoline and population density, which suggested that gasoline consumption was inversely correlated with (gross) population density. Although this has been criticized as being overly simplistic as socio-economic factors are not considered, few alternatives have yet been provided using an improved methodology Zhang (2004). Ingram (1998) illustrated a similar relationship using a residential floor area per per person, which illustrated similar findings to Newman and Kenworthy (1989).

Angel (2008) suggests that an important planning strategy for governments in rapidly urbanizing cities is for them to assume that this growth will occur, and to buy land for infrastructure before the urban growth happens, so that the infrastructure can shape the urban growth, but also so that they can buy land while it is still affordable. This predicted urban growth[6] highlights the importance of understanding the interrelationships between urban form, transportation and resource-use.

[6] According to Angel (2008) between 2000 and 2030, cities in developing countries will double their population from 2 to 4 billion and at least triple their total built-up area.

Zhang (2004) argues that when market forces do not result in the optimal mix of density or land-uses, land-use planning is seen as the second best approach to attack the problem. Handy et al. (2006) observed that land-use made a difference in determining whether residents perceived walking as an option available to them or not which is important in the individuals mode choice decision, which is an important first step. However, there is not a clear consensus on how the built environment influences VKT. Crane (1996) reviews many papers that illustrate New Urbanism design principles that can either increase or decrease VKT.

Song and Knaap (2004) assess policies that influence neighborhood growth patterns, by examining several quantitative measures of urban form and compute these for neighborhoods of varying age in Portland, Oregon. They examine whether Portland's strategies are preventing sprawl, and conclude that many of the measures associated with improved mobility and reduced automobile use have improved since the 1990s based on the regions growth strategies. Ewing (1997) estimates that a doubling of population density results in a 25-30 percent lower level of VKT, while Holtzclaw (1994) observes that the difference between 50 dwellings/hectare (urban densities) and 12.5 dwellings/hectare (sub-urban densities) was a 40 percent increase in travel.

Newman et al. (2009) considers it important for cities not to draw resources from outside their bio-region, while planners recognize the need to prevent housing construction in floodplains. Newman et al. (2009) refers to a phenomenon called transit leverage described by Davoudi et al. (2009)[p.74] whereby one passenger km of transit use replaces between five to seven passenger km in a car due to more direct travel. This is illustrated in the Millennium Cities Database (Vivier, 2000) where a linear increase in public transport results in an exponential decline in VKT. To achieve this, Newman et al. (2009) argue that we need a combination of dis-incentives (such as rising fuel prices) and incentives (more public transit available). Zegras (2010) identifies that household ownership of one vehicle is based primarily on income, however additional vehicle ownership is influenced by land-use mixes, dwelling unit densities and proximity to CBD.

2.7 *Urban Information Systems*

Here I briefly discuss the technology related to planning, analyzing urban systems and public participation. Arnstein (1969) explores citizen participation in the paper '*A Ladder of Citizen Participation*' and structures the involvement that citizens can have into eight levels: *Manipulation, Therapy, Informing, Consultation, Placation, Partnership, Delegated Power,* and *Citizen-Control*. While Arnstein (1969) observes that while people are rarely openly against citizen participation, she provides many examples where the involvement was not meaningful, and argues that 'participation without redistribution of power is pointless'. One technological approach that has emerged from the planning field to assist with the analysis of urban systems at varying scales, such as a city or a region is a set of tools referred to as Planning Support Systems (PSS) (Bishop, 1998). The name, PSS, is suggestive of how technology can assist planning, rather than a technology that should be applied to a problem. PSS can be broadly defined to include 'technology-based solutions useful to planners or more narrowly as GIS-based models for examining urban futures' and can encompass analysis, design, participatory planning, communication and visualization (Brail, 2008).

Klosterman (Brail, 2008, 85 - 99) considers PSS to have emerged from three general approaches: (1) *Planning as Design*, (2) *Planning as Applied Science*, and (3) *Planning as Reasoning Together*. Klosterman considers (1) to have emerged from the traditional disciplines of engineering and construction, while (2) was a result of new quantitative techniques from the emerging fields of regional science, urban economics and operations research. Lee's critique of large models (Lee, 1973) would appear to be targeted primarily at the *Planning as Applied Science* approach. *Planning as Reasoning Together* requires planners to act in a fundamentally different role where their goal is to facilitate involvement and inform a citizen-lead approach, which is the approach that Klosterman favors. Large scale urban modeling has evolved considerably since Lee's critique as the technology associated with PSS is becoming more easily available to the public.

PSS can be classified generally into systems that consider single criteria or multiple criteria. Typically, single-criteria models facilitate sketch-planning can be manipulated in real-time, while those with multiple dimensions interacting are more complex and require substantial simulation time. For example, *What if?* (Brail, 2008) can be used to present the tradeoffs associated with how factors can be weighted, making this appropriate for public-meeting discussions, as the dimensions of measurement are independent (i.e. how many roads will there be, is separate from how many businesses are in the area). In contrast, *UrbanSim* (Waddell et al., 2003) is more of a black-box type approach where factors are combined using historic estimates, linear regression, and expert judgement to conceptualize the problem and requires substantial simulation time. *UrbanSim* is more similar to Klosterman's category, *Planning as*

Applied Science, as it uses sophisticated quantitative techniques.

2.7.1 Analysis Methods and Data Considerations

While spatial information systems have been used for many years to inform decision makers about possible causes of problems (Tufte, 1982), this domain has been closely related to the fields of planning, engineering and architecture. Despite the recent innovations in web-mapping software, it is still challenging to perform spatial analysis using a web-browser. While Batty (Brail, 2008)[p.29] observes that visualization is 'drifting toward web-based models', he also notes that there is a lot of software fragmentation associated with *Planning Support Systems*.

While hardware development opens up future computing potentials, it is software which eventually gives the planner or citizen access to the technology, information and democratizes access to information. Individuals are becoming more familiar with Google Maps and the web-based mapping paradigm, yet there are still gaps in full public engagement in the planning process. A future challenge, is to achieve a balance between public-participation, data-accuracy, and proportional representation amongst digitally literal citizens and rigorous quantitative analysis.

There are several technical and conceptual challenges with regard to representing the physical environment at different scales. The scales considered in this research are the city level, the neighborhood level and a range of urban form parameters at the building level. The methods of representing such cross-scale information have not been standardized, with most research using a mixture of vector and raster data. Although there have been efforts by Kolbe et al. (2005) to use a format referred to as *CityGML*, this has not been widely adopted. It is important to reach a consensus on the best approaches to represent this information, so that researchers can benefit from the semantic information associated with data. It is important that data standards with open, non-proprietary formats are used for urban analysis, so that the analysis is not limited to a particular software. For example, the dominant form of spatial data storage is the shape-file, which is proprietary.

Within the urban space, access to data is a frequent challenge. While user-generated data holds great opportunities, one challenge with these data is maintaining quality and accuracy. One example of a well developed data-management hierarchy is the Cornell Bird project, where data is vetted by editors with various levels of expertise before it is compiled and verified. (*Project FeederWatch*: birds.cornell.edu/). Another example of a dataset with high-quality user-generated data is OpenStreetMap (openstreetmap.com). OpenStreetMap (OSM) is a global vector map of the world with freely available data. OSM have developed and modified many software editors to enable users to contribute data to this project using multiple methods and to cross-check the results from a variety of sources.

2.8 Research Question

Based on this review of the current state of research and considering global trends, I reiterate the key points which led me to formulate my research question. First, I consider research in the domain of urban resource analysis to be relevant due to increasing per person material and energy use as well as projected population growth in urban areas. Cross-city spatial patterns are not frequently studied (Kennedy et al., 2009; Decker et al., 2000) and many cross-city studies use one number per city (Bettencourt et al., 2007; Newman and Kenworthy, 1991) which does not capture spatial variation. In addition, acquiring standardized city level data that is related to resource consumption is difficult and there is a lack of standardization in defining urban boundaries (Weisz and Steinberger, 2010). There is a need for research that focuses on quantifying the resource efficiency of cities at a high spatial resolution, as this has not been explored thoroughly in the published literature.

Within the planning community, there have been many discussions about the holistic assessment of cities, and several approaches have emerged to facilitate the identification of sustainable urban forms (PSS, GIS, etc). While the important elements have been identified, there has been less focus on the process of repeatable urban analysis. Examples of best-practice guidelines and strategies are usually guidelines without numeric criteria (USGBC, 2009). Numerous *Sustainability Metrics* are published both by governments, and organizations yet the vast majority of these reports do not consider spatial variation within cities. Instead they typically use a per capita measure for each city which does not take into account spatial heterogeneity.

Finally, from the policy perspective, there is a need to identify the minimum and maximum levels of material and energy needed for the functioning of neighborhoods. Establishing reasonable targets within these bounds can then be accomplished, which can enable cities to move towards greater resource efficiency.

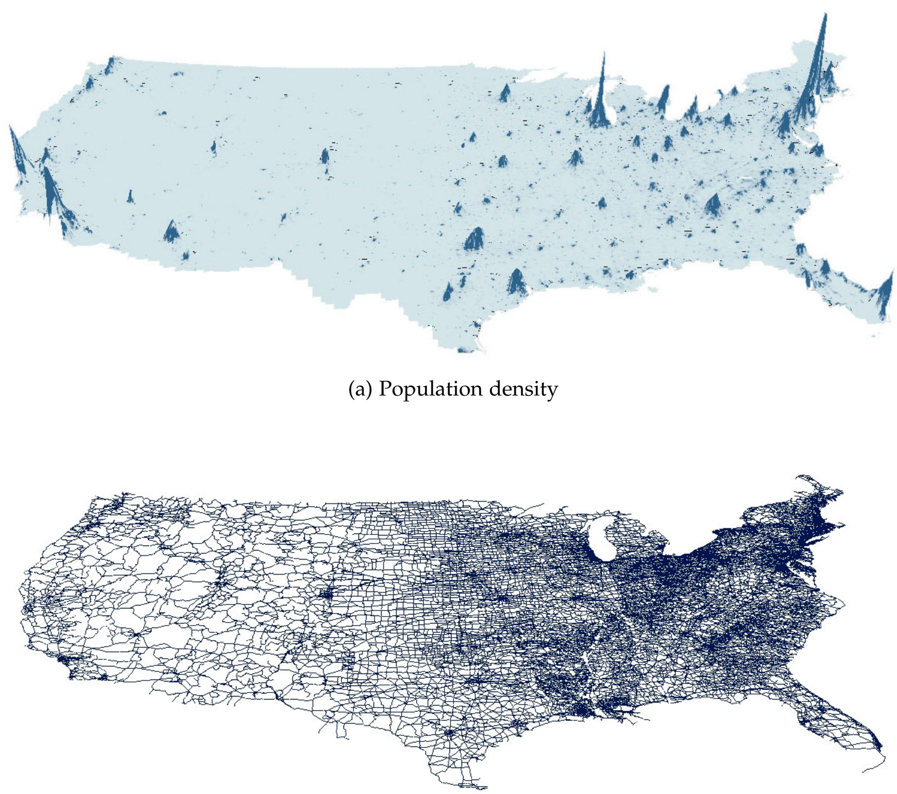
Considering this broader context, my research question is, can one establish credible estimates of urban resource intensity through the careful examination of physical attributes?

While answering this question, I also wanted to to (1) identify general upper and lower ranges of resource intensities that can be used to characterize the urban form, (2) identify material and energy intensities at a high spatial resolution, (3) make the results of this analysis available using web-mapping standards and (4) visualize the results of this analysis in an interactive and intuitive manner.

3 Urban Characterization: City Level

In this chapter, I describe several approaches that can be used to characterize urban areas. In Section 3.1, I describe the background, methodology and data used in this analysis. I then explain in detail how population density gradients (Section 3.2) and road density gradients (Section 3.3) were calculated for 40 US cities. In Section 3.4, I describe an alternative approach for urban characterization by calculating population and road measurements using measures from a grid. Finally in Section 3.5, I describe how building-level measurements can provide a more nuanced method to characterizing urban areas, and illustrate this approach using data from three cities.

The contents of this chapter have been presented at two conferences Quinn and Fernández (2010) and Quinn and Fernández (2011b).

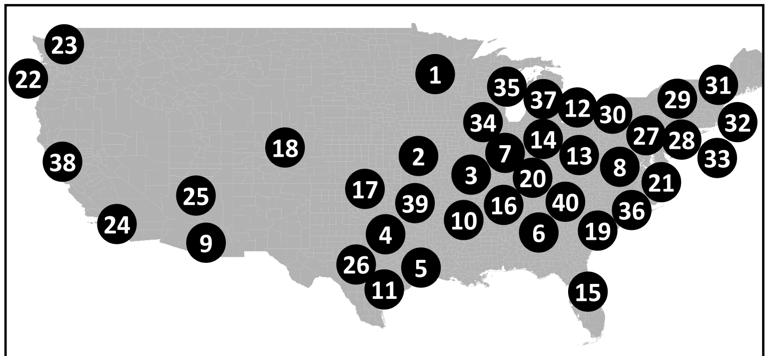


(a) Population density

(b) Road network

Figure 3.1: Population density and road network patterns in the USA. These visualizations use data from the US Census (2000).

Figure 3.2: Map of the USA showing the cities analyzed in this chapter.



#	City	#	City	#	City	#	City
1	Minn.-St.Paul	11	San Antonio	21	Richmond	31	Lawrence
2	Kansas City	12	Detroit	22	Portland	32	Boston
3	St Louis	13	Columbus	23	Seattle	33	Providence
4	Dallas	14	Cincinnati	24	Los Angeles	34	Chicago
5	Houston	15	Orlando	25	Flagstaff	35	Milwaukee
6	Atlanta	16	Nashville	26	Austin	36	Greensboro
7	Indianapolis	17	Oklahoma City	27	Philadelphia	37	Flint
8	Pittsburgh	18	Denver	28	New York	38	San Francisco
9	Phoenix Mesa	19	Charl.-Gastonia	29	Springfield	39	Tulsa
10	Memphis	20	Louisville	30	Cleveland	40	Knoxville

Figure 3.1 (a) and (b) show the population density and the road network for the USA. Here a visual correlation can be observed between the population density and the road network. I examine the relationship between these two measures in detail for 40 US cities which are shown in Figure 3.2.

Through the identification of cross-city spatial patterns, empirical ranges of population and road network measures can be identified. This approach can then be used to identify reasonable upper and lower boundaries of urban measurements in data scarce environments.

The methods of analysis in this chapter were combined into a plugin that I developed for ArcGIS to enable this analysis process to be repeated and shared. This plugin is described in detail in Appendix B.1.1.

In this chapter (and in this work overall), there has been a focus on the process of analysis so this work can be repeated and easily shared. To achieve this, all of the spatial and statistical analysis was performed using a script-based approach. Hence the analysis shown here is not a series of case studies; it is a description of the analysis process and a discussion of the results of the analysis, with the goal that these methods can be easily applied elsewhere.

The final aspect of this work was an effort to condense complex spatial measurements into easily interpretable graphics. These graphics can be used to quickly gain a preliminary insight into an urban area and also facilitate city level comparisons.

3.1 *Methodology and Data*

To perform this analysis, 40 US cities were analyzed (Figure 3.2). The data sources for this analysis are listed in Table 3.1 and the map projection used was Albert Equal Area. All calculations involving population data were performed on data at the census block-group level. The US Census Metropolitan Statistical Area Census (US Census, 2000) was used as an initial boundary for each city. Then, block-groups[1] with a population density greater than 300 people/km^2 were used to identify the general urban area. The final selection of contiguous block-groups was chosen using graphical inspection and included some areas with population densities less than 300 people/km^2.

[1] A block-group is a unit used by the US Census and consists of approximately 1500 people.

For this reason, the areas of the cities used in this study are not coincident with formal municipal boundaries. City names are used loosely, as the name does not correspond to political boundaries but it is used to identify the general area of the census-defined city boundary. The definition of urban boundaries has been discussed in the literature when considering urban resource flows (Kennedy et al., 2009) and is a frequently discussed topic in urban research (Weisz and Steinberger, 2010) due to the lack of a formal definition for a city. Using a numeric threshold to define the boundary of the urban system (such as population density), instead of a political boundary, makes it easier to compare the results of this analysis.

This work was conducted using *ArcGIS 9.3* for spatial analysis (ESRI, 2009) and the open-source program *R*, for statistical analysis (R Development Core Team, 2011). *ArcGIS* was controlled using computer code written in the *Python* language. The output from the analysis was written to a text file. Then *R* was used to read in the these text files and statistical analysis was then used to explore the results graphically and numerically and to identify patterns. Appropriate curves were fitted to the data using regression and these were plotted using the *R* package *ggplot2* (Wickham, 2009). A description of the plugin that was developed to perform this work is listed in Appendix B.1.

Data	Source	Year	Webpage
Population Data	US Census	2000	`http://www.census.gov`
Metro Areas	US Census	2000	`http://www.census.gov/geo/www/cob/bdy_files.html`
Road Data	US Census	2007	`ftp://ftp2.census.gov/geo/tiger`
Service Location	ESRI Business Analyst	2007	`http://www.esri.com/software/bao`
Business Classification	NAICS	-	`http://www.census.gov/eos/www/naics/`

Table 3.1: Data sources used in this analysis.

A gradient density measurement describes the rate of change of the parameter of interest, with respect to a radial distance from the city center. Population density gradients have been observed in many cities. Clark (1951) observed, and was instrumental in formalizing the concept of a population density gradient in cities. This approach measured the concentration of people as function of distance from the city center. The underlying assumption for calculating gradients in cities is based on the assumption that a city is mono-centric, with the existence of a central area with a high population or business density (frequently referred to as a central business district or CBD). One of the challenges of using this approach is the need to identify an urban center which may not be possible in polycentric cities. While it is possible to discern distinct governmental and business centers in many cities, the urban population and services may be distributed and concentrated around multiple urban structures and districts. In this work, the criteria used to define an urban center is straightforward, however there are several different approaches that can be used (Bento et al., 2005). The identification of this central area was the first step in this work.

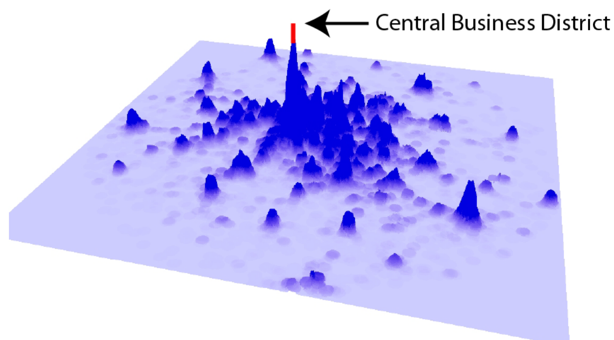


Figure 3.3: Atlanta service density, with the peak density marked with red.

The center of the city was defined as the area with the highest service density, using point data that identified the location of services.[2] These services were identified using the North American Industrial Classification System (NAICS) listed in Appendix A.1. The density of services was calculated for each city using grid-cells at a 200m resolution, by counting the number of services that were within a 1.5 km radius of each grid-cell. The US Census Metropolitan Statistical Area was used as an initial boundary for each city. Then, grid-cells with values that were within the highest 5% of all service densities were converted into a polygon which was considered to represent the CBD. These polygons were then used for all subsequent population density and road density gradient analyses in this chapter. The service density and CBD for Atlanta are illustrated in Figure 3.3.

[2] This collection of services are identified by *LEED for Neighborhood Development* (USGBC, 2009) as being important services for communities. These services are discussed in greater detail in Chapter 6.

To calculate the population density, a radial buffer was created at 1km intervals from the CBD, and all census blocks that were within, or intersected the buffer, were summed, up to a maximum radial distance of 80 km. For census blocks that were intersected by the buffer (Figure 3.4), the fraction of area that fell within the buffer was used to estimate the number of people living there using Equation 3.1.

$$pop_{new} = (area_{new}/area_{old}) \times pop_{old} \tag{3.1}$$

This equation assumed that the number of people living in each census block was uniformly distributed. The population density values appear significantly lower than typical population density values, as all areas of the city were included, such as water bodies, parkland and non-residential areas.

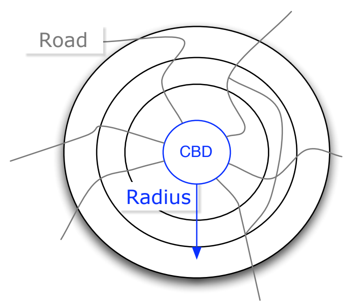


Figure 3.4: Schematic of the analytical technique used to measure the radial distribution of population density and road density.

For road calculations, the length of road that fell within each buffer and the road type was recorded. This work considered primary, secondary and local roads (Appendix A.2.2). A schematic of how this road density calculation was performed is illustrated in Figure 3.4.

As mentioned previously, the existence of population density gradients in cities has been widely observed and studied, but is reproduced here with an analytical method that can be examined and easily replicated. While this overall process of analysis was applied to 40 cities, a special emphasis is placed on Atlanta, Georgia to clearly illustrate the results of these measurements.

3.2 *Population Density Gradient*

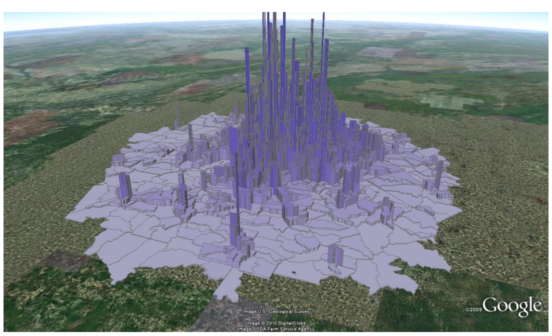
As previously mentioned, Clark (1951) formalized a description of the population density gradient. Assuming that the urban center can be identified, the population density can be related to the radial distance from the CBD using the relationship listed in Equation 3.2:

$$y = Ae^{br} \tag{3.2}$$

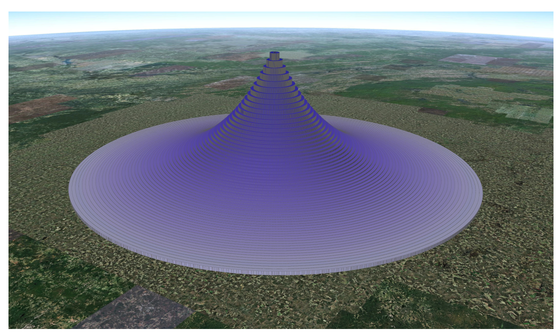
where y is the population density, r is the radius from the CBD, and A and b are specific parameters to describe the curve for each city.

Using this approach, the population density gradient was estimated for the 40 cities in this study (Figure 3.2) by fitting measured data to the curve described by Equation 3.2. For these 40 cities, the parameters for this curve are listed in Table 3.2, with the associated R-squared, p-value and t-statistic. The fitted curves for all cities are shown in Figure 3.7.

A population density gradient pattern can also be illustrated graphically using a geo-visualization technique, where the population density is used to extrude a polygon to a certain height. Using the original census data at the block level, the population density for Atlanta is visualized in Figure 3.5 (a), with the estimated population density gradient illustrated in Figure 3.5 (b). Figure 3.5 (b) uses a curve of the form described in Equation 3.2, with the parameters listed in Table 3.2. The 3D representation of the raw population density data in Figure 3.5 (a) illustrates the existence of a population density gradient.



(a) Population density using raw data.



(b) Population density using estimated parameters.

Figure 3.5: 3D illustrations of the population density of Atlanta.

The results from Atlanta are shown in Figure 3.6. Here, the parameters that describe the curve are $A = 2199.60$ and $b = -0.06$. Considering the results in Table 3.2, it can be observed that this population density gradient exists for all cities examined (Figure 3.7) with the exception of Providence, and Lawrence. The low R^2 for these two cities is likely due to the fact that a clear urban center was not identified, and the population of the adjacent larger urban area of Boston obscured any local gradient.

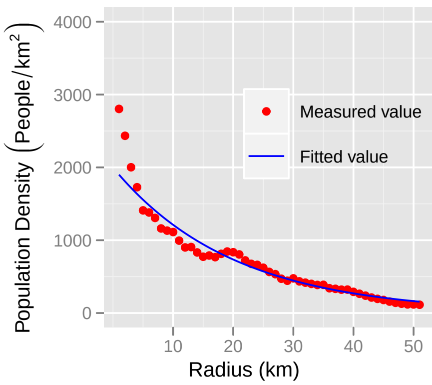


Figure 3.6: Population density gradient measure for Atlanta.

This analysis illustrates that there are clear cross city patterns of behavior when considering the population density variation as a function of distance from CBD. As this process of analysis was scripted, the method can be applied to any city where the data is available and formatted appropriately.

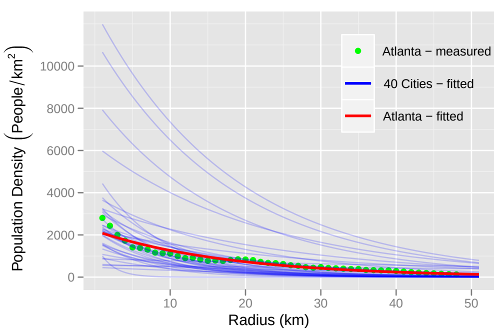


Figure 3.7: Population density gradient for 40 US cities.

To illustrate an example of how this analysis could be used, consider a group of planners who are working in a data scarce environment and are curious about the population density of an area 20 km from the city center. There is no other data available for this city, but several other cities in this country have recently had a population survey. Using data from this survey, the planners consider the likely population density in the city center to be 8000 $people/km^2$ and decide to estimate the population density of the area of interest using a population density gradient approach. There is uncertainty about the rate at which the population density decreases, so the planners decide to use a value of -0.09 for b as it lies midway between the most extreme cases in this work. The population density for this area can now be estimated using Equation 3.2 with the following values:

$$\text{Population Density} = 8000 * e^{-0.09*20} \tag{3.3}$$

In this case, Equation 3.3 predicts the population density of this area to be 1322 $people/km^2$.

Table 3.2: Population density parameters.

City Names	A [Person/km^2]	b [$1/km$]	Adj.-R-Sq.	p-value	t-test
Atlanta	2200	-0.06	0.9721	0.0000	-43.38
Austin	2237	-0.11	0.9670	0.0000	-29.69
Boston	2193	-0.03	0.7350	0.0000	-14.83
Charlotte-Gastonia	1148	-0.06	0.8840	0.0000	-17.70
Chicago	8387	-0.06	0.9558	0.0000	-39.47
Cincinatti	2516	-0.09	0.9690	0.0000	-33.55
Cleveland	2155	-0.05	0.8353	0.0000	-17.33
Columbus	3100	-0.12	0.9083	0.0000	-15.77
Dallas	2398	-0.05	0.9700	0.0000	-44.45
Denver	5005	-0.12	0.9100	0.0000	-18.02
Detroit	3845	-0.05	0.9460	0.0000	-33.75
Flagstaff	1648	-0.50	0.9991	0.0000	-81.95
Flint	1828	-0.16	0.9840	0.0000	-35.06
Greensboro	469	-0.04	0.5444	0.0000	-7.87
Houston	3162	-0.07	0.9406	0.0000	-27.29
Indianapolis	2424	-0.10	0.9308	0.0000	-19.43
Kansas City	2434	-0.09	0.9232	0.0000	-19.32
Knoxville	988	-0.09	0.9223	0.0000	-18.26
Lawrence	869	-0.01	0.1194	0.0010	-3.42
Los Angeles	6249	-0.04	0.9347	0.0000	-33.63
Louisville	2717	-0.13	0.9576	0.0000	-22.80
Memphis	1680	-0.09	0.9262	0.0000	-18.78
Milwaukee	3445	-0.09	0.9110	0.0000	-21.00
Minneapolis	4148	-0.10	0.9518	0.0000	-26.30
Nashville	1467	-0.09	0.9670	0.0000	-29.67
New York	11263	-0.06	0.9285	0.0000	-32.04
Oklahoma	1795	-0.11	0.8623	0.0000	-12.30
Orlando	2343	-0.10	0.9360	0.0000	-20.62
Philadelphia	12648	-0.05	0.9450	0.0000	-36.87
Phoenix	2347	-0.07	0.7315	0.0000	-10.36
Pittsburgh	2248	-0.08	0.9320	0.0000	-24.29
Portland	3449	-0.11	0.9107	0.0000	-18.09
Providence	558	-0.01	0.0310	0.0642	-1.88
Richmond-Petersburg	2233	-0.12	0.9786	0.0000	-33.87
Saint Louis	2670	-0.08	0.9575	0.0000	-29.29
San Antonio	3696	-0.13	0.8790	0.0000	-13.51
San Francisco	3355	-0.04	0.7800	0.0000	-16.77
Seattle	2613	-0.06	0.9481	0.0000	-32.01
Springfield	613	-0.02	0.5302	0.0000	-9.49
Tulsa	1699	-0.12	0.9755	0.0000	-30.30
Mean	3056.05	-0.09	-	-	-
Standard Deviation	2555.17	0.08	-	-	-

3.3 *Road Density Gradient*

To explore the relationship between the road density and urban structure a road density gradient measure was explored. As described in Figure 3.4, the total road length per buffer was calculated using *ArcGIS* and the data was then analyzed using *R*. After exploring a range of curve fitting estimates, these road density measurements were observed to follow a power law with the following structure:

$$y = Ar^b \tag{3.4}$$

where y is the road density, r is the radius from the CBD, and A and b are parameters for each city. The results of this analysis for Atlanta are

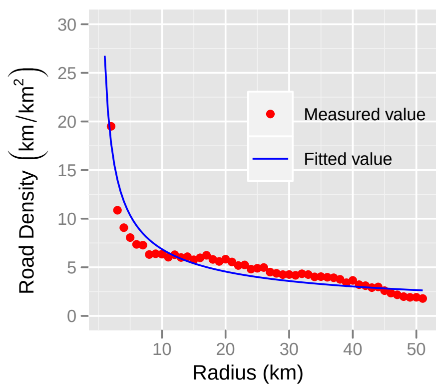
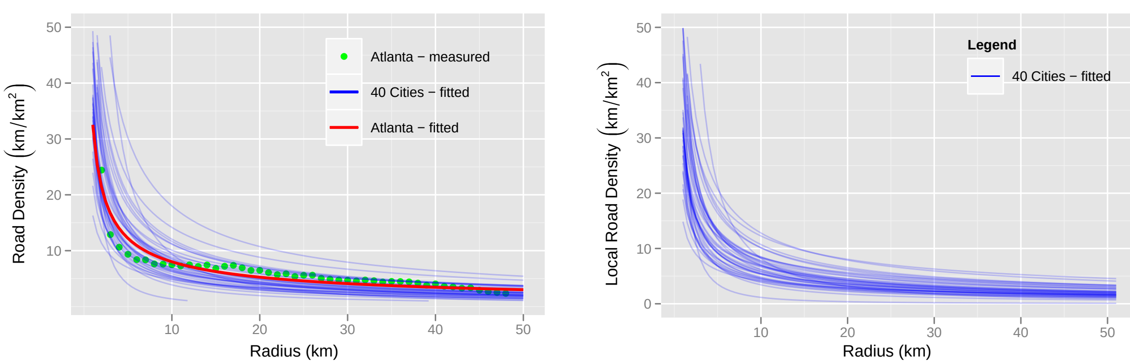


Figure 3.8: Local road density gradient.

illustrated in Figure 3.8. The parameters to describe this curve are $A = 25.13$ and $b = -0.54$. A road-density gradient pattern was observed for all cities, with the fitted curves shown in Figure 3.9 (a). The parameters for each city are listed in Table 3.3.



(a) All roads.

(b) Local roads.

Figure 3.9: Road density gradient for roads.

A similar pattern of behavior was observed for local roads which can also be described using Equation 3.4. The fitted curves for local roads are illustrated in Figure 3.9 (b), with the parameters describing these curves listed in Table 3.4.

One other interesting pattern that was observed when performing these calculations is illustrated in Figure 3.10. Figure 3.10 demonstrates that the total road length increases linearly as a function of distance from the CBD, when measured radially. The slope of this linear relationship varies from 134 (Flagstaff) to 1041 (Los Angles). The R^2 value is > 0.99 with the exception of Boston ($R^2 = 0.70$); all values are significant at the 99% confidence interval.

Figure 3.10: Total road length measure from the city center, for 40 US cities.

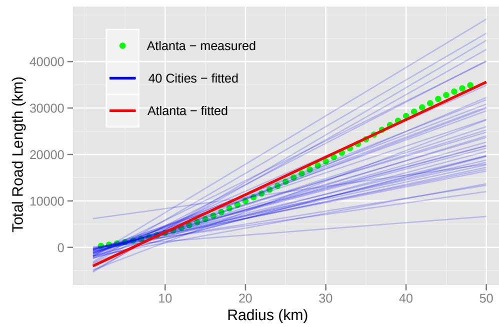
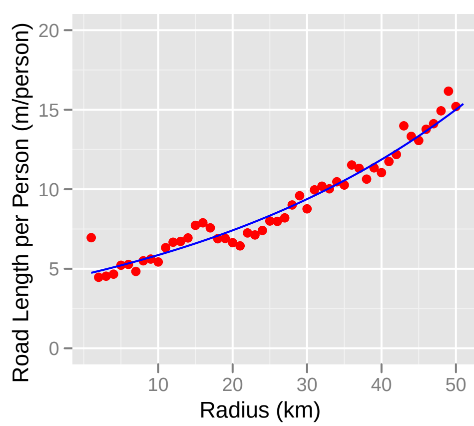


Figure 3.11 illustrates how the road length per person changes in Atlanta, as a function of distance from the center. As the population decreases more quickly than road density, the local road length per person increases. The measurements shown in Figure 3.11 can be described using the same curve as Eq. 3.2. In the case of Atlanta, these parameters are $A = 4.63$ and $b = 0.024$. Figure 3.11 demonstrates that three times as much local infrastructure is needed for an area with low population density, when compared to an area of high population density (the population density gradient for Atlanta is shown in Figure 3.6).

Figure 3.11: Road per person.



This measurement can be used to estimate the amount of road infrastructure in a neighborhood, when only the distance to the CBD is provided. To illustrate this using a specific case, road networks for Boston and Dallas are shown in Figure 3.12 at a distance of 5 km from the CBD. Visually we can see that there is a higher road density in Boston than Dallas. Using parameters from Table 3.4 and Equation 3.4 we can estimate that the local road-density for Boston is 21.52 km/km^2, while the local road density in Dallas is 12.93 km/km^2.

Figure 3.12: Local road patterns for cities Boston and Dallas at a distance of 5 km from the city center.

Newman and Kenworthy (1989) use the linear miles of road per capita to categorize cities and suggest that this measure can illustrate what the primary transportation mode is. This work starts to explore how the rate of change in road density from the CBD could be an alternative approach to the classify cities. As local roads have also been examined in this section, the resources required to provide this infrastructure can be directly attributed to the local population, as it is assumed that they are the primary users of these roads.

Similar to the caveats associated with the population density measurements, this approach is only applicable to cities that have an easily identifiable CBD.

Table 3.3: Parameters to describe the road density for all roads.

City Names	**A** $[km/km^2]$	**b** $[km/km]$	**Adj.-R-Sq.**	**p-value**	**t-test**
Atlanta	25.13	-0.54	0.8299	0.8299	-16.11
Austin	22.19	-0.68	0.9599	0.9599	-26.37
Boston	104.45	-1.14	0.4389	0.4389	-5.68
Charlotte-Gastonia	18.57	-0.51	0.9711	0.9711	-36.70
Chicago	73.13	-0.66	0.8077	0.8077	-17.30
Cincinatti	25.81	-0.65	0.9684	0.9684	-32.76
Cleveland	26.35	-0.57	0.9245	0.9245	-26.66
Columbus	26.33	-0.67	0.8821	0.8821	-13.44
Dallas	32.89	-0.56	0.8788	0.8788	-20.88
Denver	35.97	-0.68	0.7824	0.7824	-10.61
Detroit	47.12	-0.65	0.7517	0.7517	-13.96
Flagstaff	17.65	-1.08	0.9389	0.9389	-8.82
Flint	21.88	-0.80	0.9472	0.9472	-18.48
Greensboro	13.52	-0.48	0.8366	0.8366	-16.03
Houston	45.24	-0.64	0.8436	0.8436	-15.78
Indianapolis	25.10	-0.57	0.8724	0.8724	-13.63
Kansas City	30.02	-0.61	0.8509	0.8509	-13.12
Knoxville	16.18	-0.54	0.9688	0.9688	-28.98
Lawrence	19.63	-0.68	0.9486	0.9486	-17.74
Los Angeles	52.92	-0.62	0.9216	0.9216	-30.29
Louisville	23.06	-0.64	0.9039	0.9039	-14.42
Memphis	18.92	-0.59	0.8888	0.8888	-14.72
Milwaukee	36.92	-0.73	0.8845	0.8845	-17.97
Minneapolis	26.94	-0.52	0.8243	0.8243	-12.67
Nashville	25.63	-0.72	0.9850	0.9850	-43.64
New York	43.74	-0.64	0.7874	0.7874	-17.03
Oklahoma	22.54	-0.58	0.9124	0.9124	-15.51
Orlando	23.18	-0.57	0.8448	0.8448	-12.38
Philadelphia	24.42	-0.62	0.8519	0.8519	-21.20
Phoenix	25.85	-0.48	0.6798	0.6798	-9.04
Pittsburgh	32.93	-0.65	0.9635	0.9635	-33.32
Portland	39.93	-0.75	0.8806	0.8806	-15.15
Providence	89.72	-1.21	0.5680	0.5680	-8.49
Richmond-Petersburg	22.26	-0.54	0.9365	0.9365	-18.84
Saint Louis	32.92	-0.63	0.9254	0.9254	-21.45
San Antonio	34.46	-0.71	0.9316	0.9316	-18.10
San Francisco	41.30	-0.62	0.9086	0.9086	-27.87
Seattle	35.49	-0.62	0.9134	0.9134	-24.11
Springfield	20.97	-0.68	0.8785	0.8785	-23.77
Tulsa	22.65	-0.66	0.9707	0.9707	-27.00
Mean	33.10	-0.662	-	-	-
Standard Deviation	18.83	0.157	-	-	-

City Names	A $[km/km^2]$	b $[km/km]$	Adj.-R-Sq.	p-value	t-test
Atlanta	26.76	-0.59	0.8113	0.0000	-15.13
Austin	26.90	-0.78	0.9638	0.0000	-27.80
Boston	198.33	-1.38	0.4424	0.0000	-5.72
Charlotte-Gastonia	20.74	-0.58	0.9699	0.0000	-35.90
Chicago	49.83	-0.61	0.8155	0.0000	-17.74
Cincinatti	31.82	-0.77	0.9691	0.0000	-33.14
Cleveland	31.36	-0.66	0.9180	0.0000	-25.51
Columbus	31.94	-0.78	0.8856	0.0000	-13.67
Dallas	35.08	-0.62	0.8844	0.0000	-21.44
Denver	49.87	-0.82	0.8213	0.0000	-11.98
Detroit	55.41	-0.71	0.7619	0.0000	-14.34
Flagstaff	47.60	-1.62	0.9561	0.0000	-10.49
Flint	33.73	-0.99	0.9607	0.0000	-21.57
Greensboro	14.84	-0.55	0.8194	0.0000	-15.09
Houston	40.89	-0.69	0.9433	0.0000	-27.67
Indianapolis	30.90	-0.67	0.8828	0.0000	-14.29
Kansas City	40.54	-0.77	0.8973	0.0000	-16.22
Knoxville	18.86	-0.63	0.9609	0.0000	-25.78
Lawrence	25.44	-0.82	0.9055	0.0000	-12.80
Los Angeles	64.13	-0.70	0.9356	0.0000	-33.68
Louisville	27.76	-0.76	0.8966	0.0000	-13.85
Memphis	21.60	-0.67	0.8774	0.0000	-13.94
Milwaukee	54.01	-0.89	0.8996	0.0000	-19.42
Minneapolis	34.62	-0.65	0.8418	0.0000	-13.49
Nashville	24.05	-0.75	0.9687	0.0000	-29.97
New York	56.39	-0.75	0.8205	0.0000	-18.91
Oklahoma	31.32	-0.73	0.9192	0.0000	-16.21
Orlando	28.61	-0.67	0.8588	0.0000	-13.09
Philadelphia	29.63	-0.70	0.8238	0.0000	-19.12
Phoenix	30.21	-0.56	0.6450	0.0000	-8.37
Pittsburgh	45.18	-0.79	0.9733	0.0000	-39.16
Portland	51.08	-0.86	0.8995	0.0000	-16.69
Providence	30.94	-0.76	0.9432	0.0000	-23.07
Richmond-Petersburg	28.47	-0.71	0.9385	0.0000	-19.17
Saint Louis	44.72	-0.76	0.9486	0.0000	-26.15
San Antonio	50.92	-0.92	0.9430	0.0000	-19.95
San Francisco	49.91	-0.69	0.9195	0.0000	-29.86
Seattle	39.04	-0.66	0.9162	0.0000	-24.54
Springfield	23.77	-0.77	0.8907	0.0000	-25.23
Tulsa	31.63	-0.83	0.9666	0.0000	-25.25
Mean	40.22	-0.765	-	-	-
Standard Deviation	28.23	0.197	-	-	-

Table 3.4: Parameters to describe the road density for local roads.

3.4 *Grid Level Measurements*

A set of software tools similar to those described in Section 3.1 was used for the analysis in this section. Each city was split into grid-cells, using a grid with a cell-size of 250m. The average number of grid-cells per city was $200,000$. The results of this analysis were plotted in *R* using the *ggplot2* package (Wickham, 2009) and the *rasterVis* package (Lamigueiro and Hijmans, 2011).

As city-wide grid measurements are used, there is no need to identify a CBD in the city, which makes this approach easier to apply in all cases. In Figure 3.13, the population density for Boston and Los Angeles is shown for a 60 *km* square. The maximum population density values from horizontal row of cells and vertical columns of cells are plotted along the x-axis and y-axis.

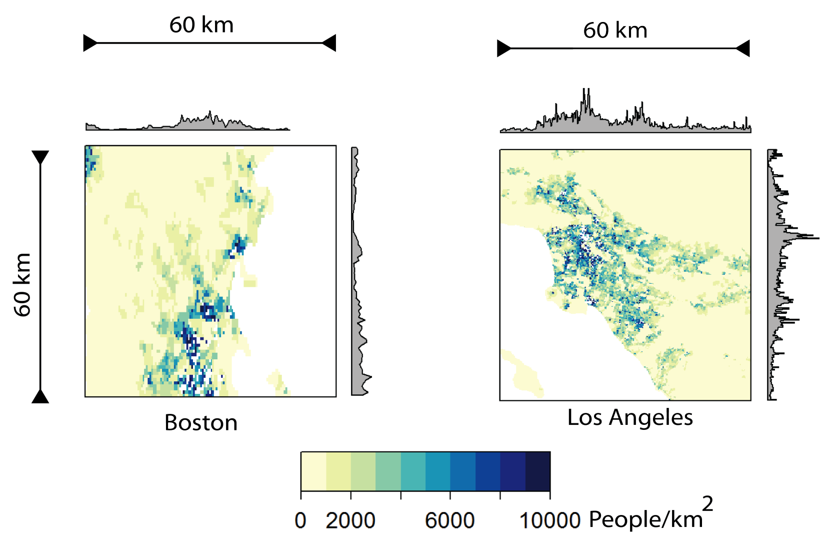


Figure 3.13: Raster population density representation of two cities. The maximum value of each horizontal and vertical band is shown on the plot along each axis.

All 40 cities are examined at the same scale using a 60 *km* square, shown in Figure 3.14. While, it is hard to see a population density gradient for some of the smaller, lower-density cities this is due to the fact that the scale needed for larger cities causes the measure to be almost indistinguishable for lower density cities. This method of data representation provides an interesting insight into the structure of the city and compresses a large number of spatial measures into an easily interpretable pattern.

This approach provides more insight than the population density gradient approach, as it is not dependent on the CBD and several dimensions of measurement are shown in an intuitive way.

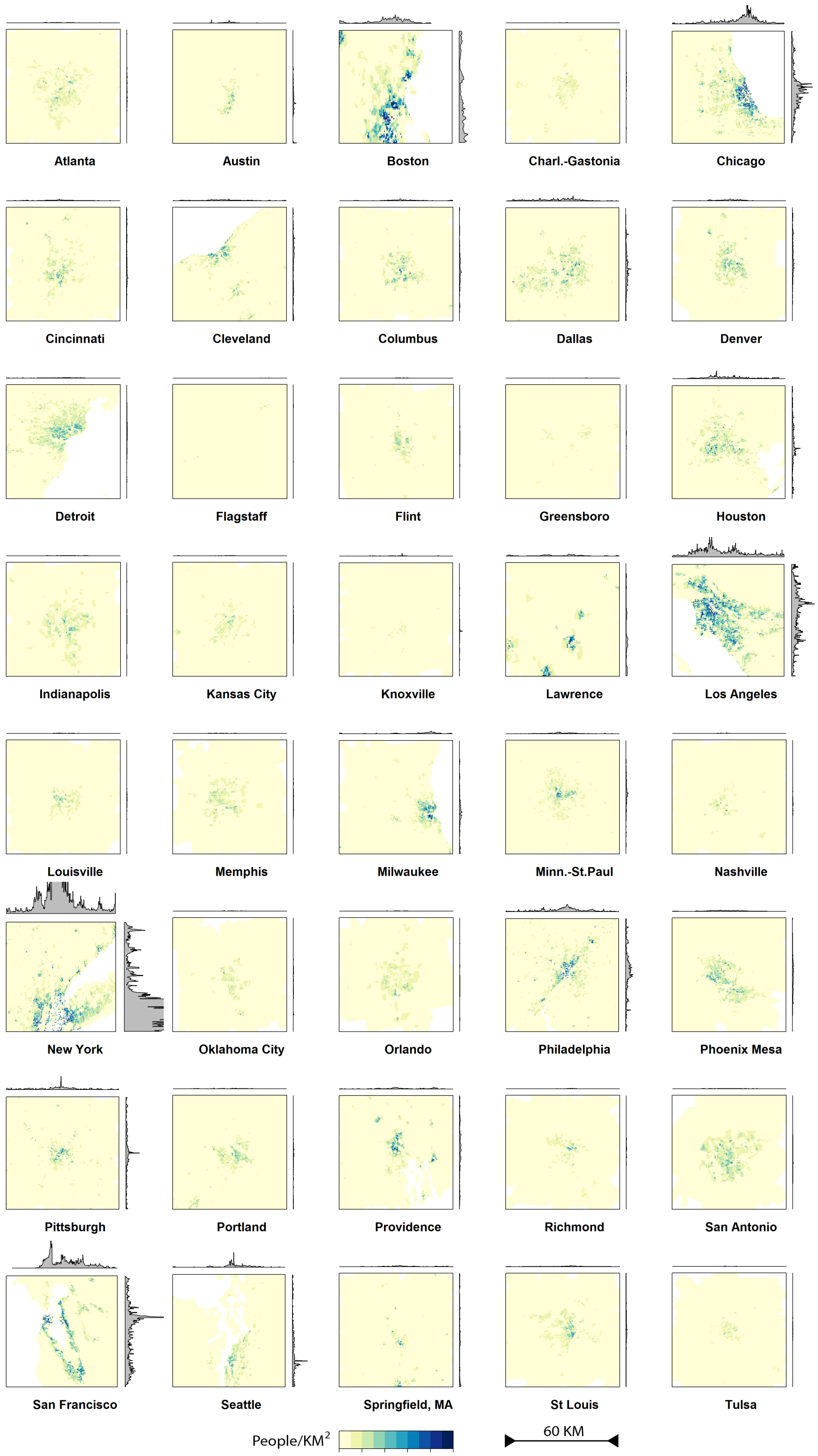


Figure 3.14: Map of the USA showing the city gradients for 40 cities.

3.4.1 *Road Density per City*

For each city, the total linear road length per grid cell was calculated for the entire urban area. This measure of infrastructure intensity includes all road types, from local roads to freeways. The resulting road patterns are shown in Figures 3.15 and 3.16. A two-dimensional histogram was used to display the data so that the frequency of each sample can be observed.[3] Each plot can be interpreted as an abstract representation of the spatial road-population structure of a city.

[3] A scatterplot with opacities was originally used to display this data, but due to the number of points over-plotting obscured the pattern.

This measurement is shown for Boston and Los Angeles in Figure 3.15. Here, the difference can be seen in terms of both the number of samples (Boston has a smaller total area, hence a lower number of observations) and also the variation. As the population density is used in this image, Figure 3.16 illustrates how much of the city lies within areas of low population density and how much is within compact areas of higher population densities.

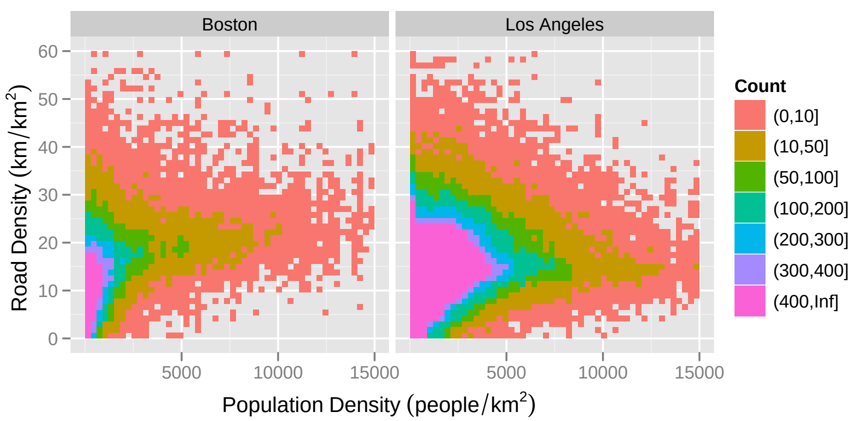


Figure 3.15: Road density for two cities.

One conclusion from these graphs is that the size of the city may also influence the achievable compactness. For example, both Boston and New York are two cities that are considered to be compact, yet each has a high level of infrastructure intensity in suburban areas, as well as large amounts of low-density areas.

Figure 3.16 illustrates the difference between large and small cities, as well as large cities with low population densities such as Atlanta, Houston and Dallas. Based on the sample count of the histogram, it can be observed that Atlanta, Houston and Dallas are large cities, but do not have areas within the city with high population densities.

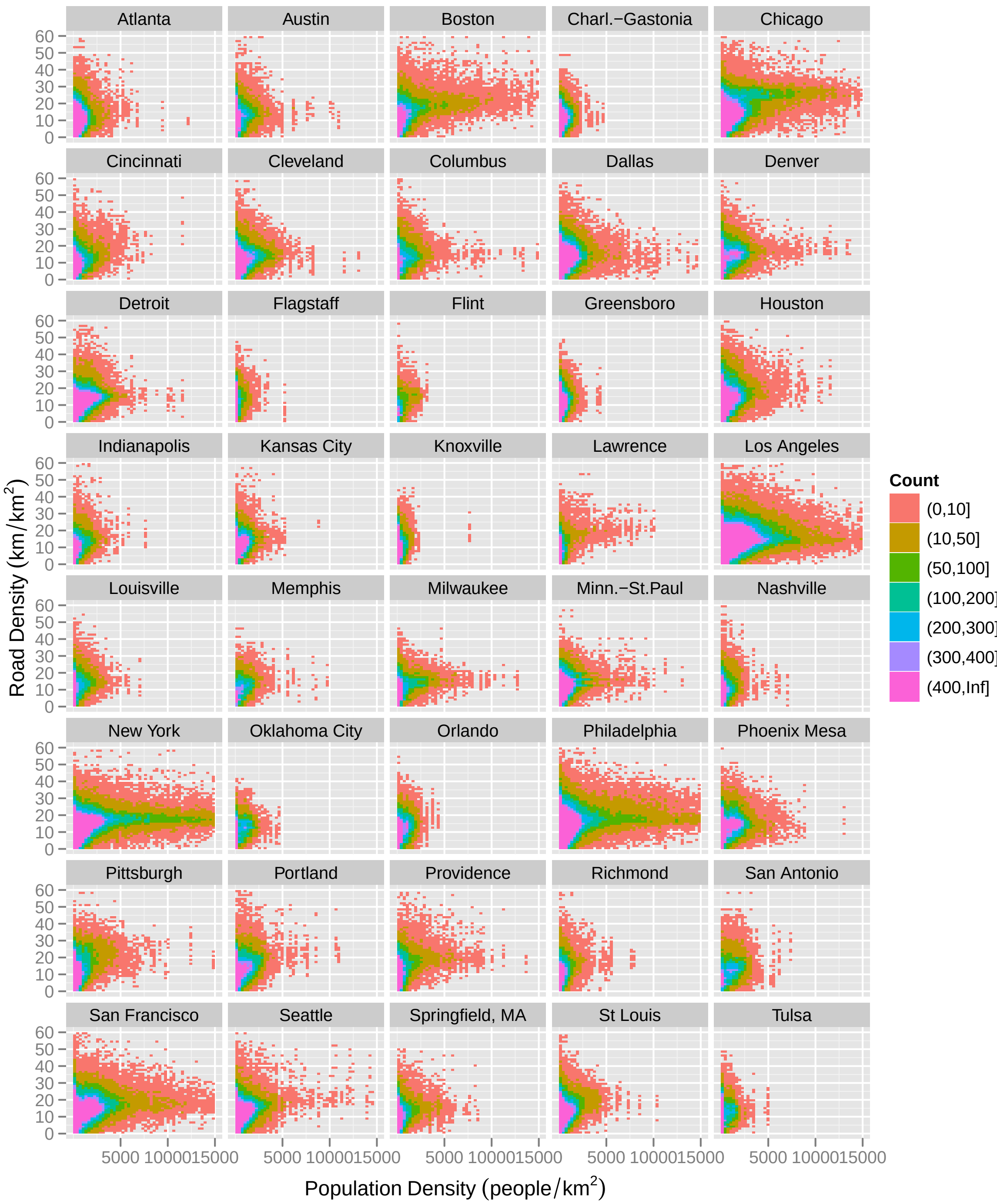


Figure 3.16: Linear road density for 40 cities, plotted against population density.

3.4.2 *Road Density per Person*

For each city, the total linear road length per person was calculated for each grid cell. The resulting road patterns are shown in Figures 3.17 and 3.18. A two-dimensional histogram was used to display the data so that the frequency of each sample could be observed.

The *Road per Person* measurement is shown for Boston and Los Angeles in Figure 3.17. Interestingly, the trend for both cities appears to be very similar (aside from a larger number of samples for Los Angles). This trend suggests that there is a clear upper boundary with regard to the road per person measure in cities. Similar to the analysis in Section 3.4.1, as the population density is used, Figure 3.17 illustrates how much of the city lies within areas of low population density and how much is within compact areas of higher population densities.

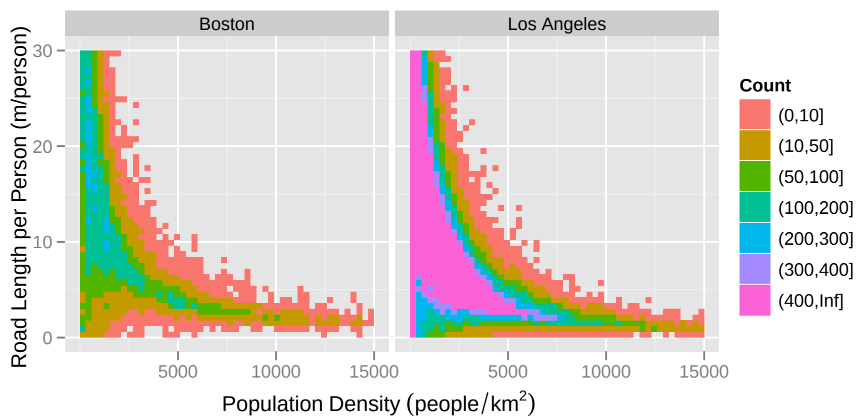


Figure 3.17: Road length per person, illustrated using two cities.

By examining the configuration of the plots for all 40 cities (Figure 3.18), one can discern several things. First, these plots illustrate a consistent pattern for all 40 cities, suggesting that there is a maximum amount of road per person, to enable the functioning of that urban area. Furthermore, this upper bound is similar for all cities.

As the population density increases, the demand for other physical infrastructure also increases. Hence, one explanation of this upper bound, could be due to competing uses of space in higher-density areas, due to buildings, green-space and other (non-road) infrastructure. At higher densities building height and floor/area ratios increase, while the available ground area remains the same which causes the road per person measure to decrease further.

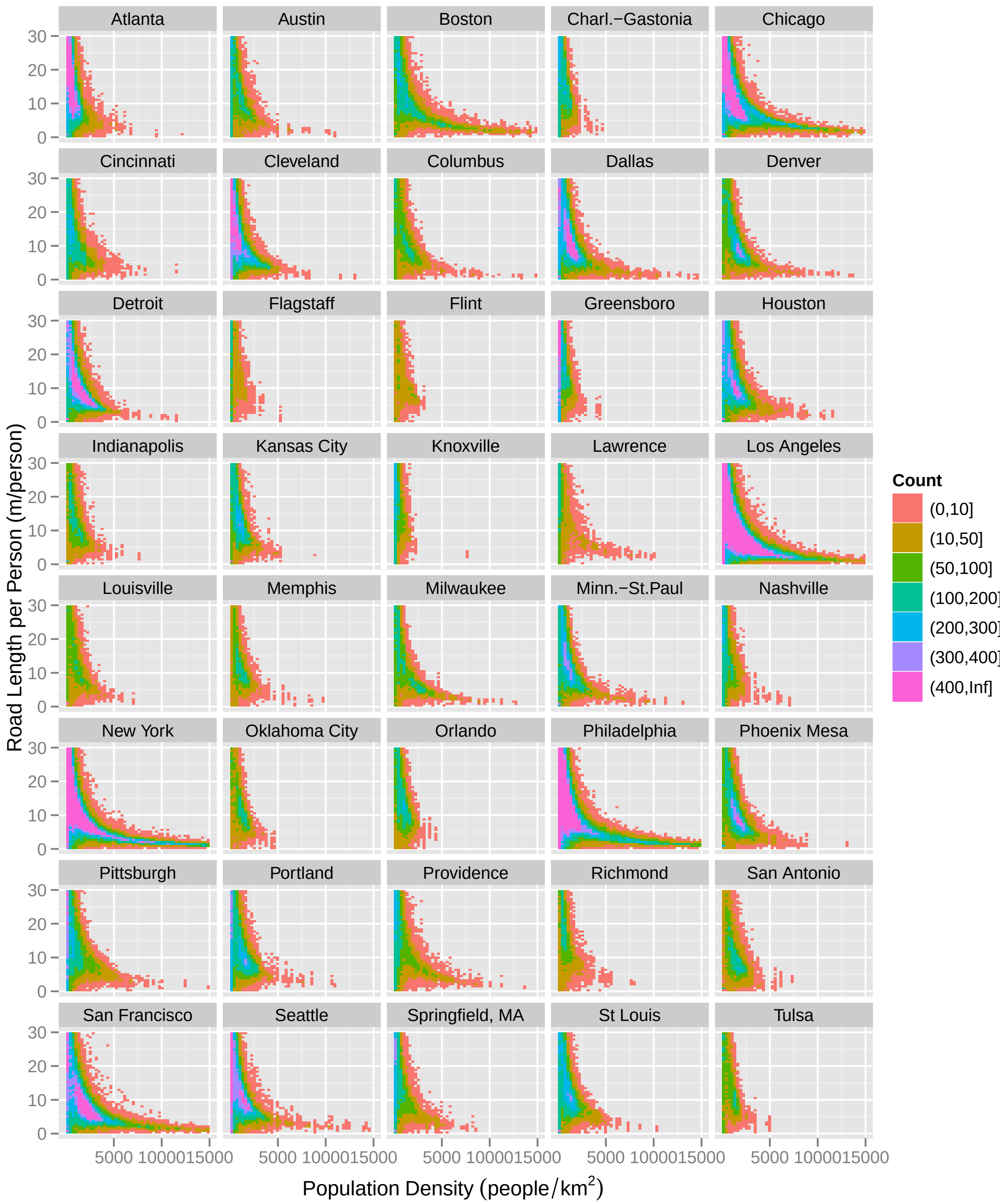


Figure 3.18: Road length per person for 40 cities, plotted against population density.

3.5 *Characterizing the City using Building Measurements*

While Sections 3.2, 3.3 and 3.4 considered discrete measurements of infrastructure and people, and the relationship between them, this section discusses an approach which combines several building-level measures to identify empirical patterns in the built environment. Again, this approach can be used to characterize a city using building-scale measures, and observing the distribution of these patterns.

Here the characterization of the urban form is being performed using only physical measures from buildings and the space around them. The objective of this approach is to identify characteristics of the urban area, using only physical measures without considering information that describes the functioning of these areas. Using this approach, some preliminary observations[4] from London data are shown in Figure 3.19. Figure 3.19 illustrates that even in a city with a high population density (London) there are large fractions of area which are not used by buildings or infrastructure. Areas that are likely to be centers of commerce can be seen with the spikes in the purple color.

[4] *Free* space is considered here to be undefined.

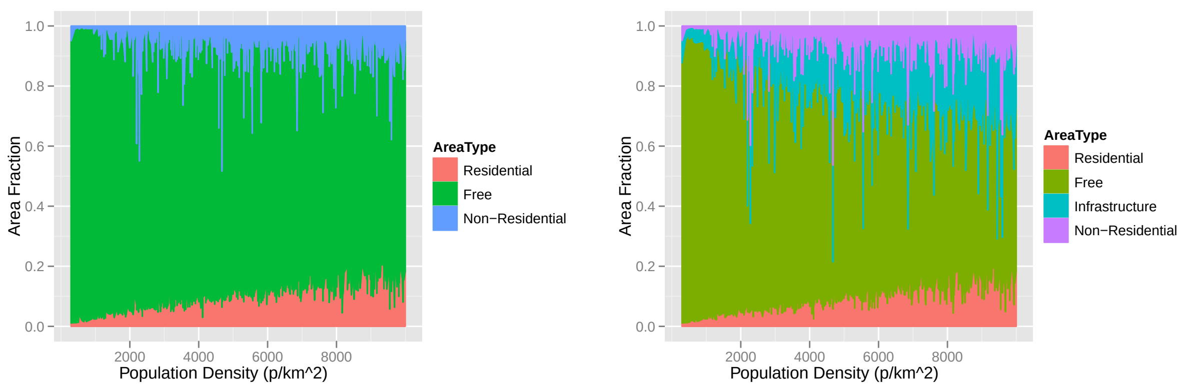


Figure 3.19: General view of land-use and population density.

Using building data for London, Manchester and Boston the relationship between building height and nearest neighbor distances was explored (Figure 3.20). This analysis was performed using *PostGIS*, which enables a more accurate nearest-neighbor calculation than *ArcGIS*, as *PostGIS* uses edge-distances instead of polygon centroids. This measurement considered the perimeter of any part of the building to calculate the nearest-neighbor distance, and calculated the minimum distance to the nearest perimeter. Figure 3.20 illustrates that a rough pattern exists between these two variables.

Using a method to describe the urban form developed by Berghauser-Pont (2010) called *The Spacemate*, the urban morphology of London, Manchester and Boston was explored using a specific set of measurements. This method of describing urban form is useful as it can intuitively show how building density, height and open space are interrelated. The *Spacemate* is calculated using the following three criteria:

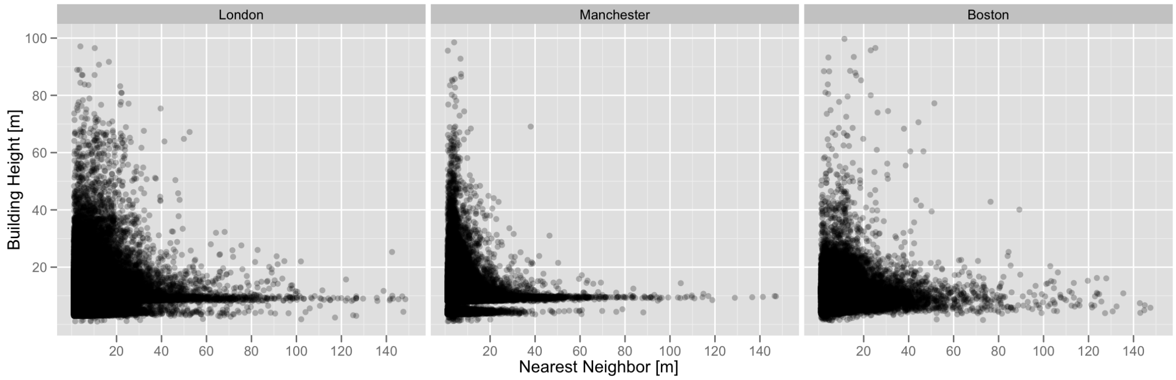


Figure 3.20: Relationship between building height and building distance for three cities.

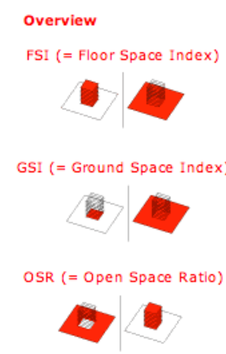


Figure 3.21: *Spacemate* criteria. Image from space-mate.nl

1. Floor Space Index (FSI) = Gross Floor Area / Area
2. Ground Space Index (GSI) = Footprint / Area
3. Open Space Ratio (OSR) = (Area - Footprint) / Gross Floor Area

These measures can be combined on the same plot (illustrated in Figure 3.22) as the measures are interrelated. To explain how this calculation is performed I explain these measures sequentially. Figure 3.23 shows a plot of *GSI* (x-axis) against *FSI* (y-axis). Dividing the *GSI* by the *FSI* results in the number of floors (shown in Figure 3.23 as *L*), which is a linear relationship. In addition, constant values of the *Open Space Ratio* (OSR) can also plotted on the graph.

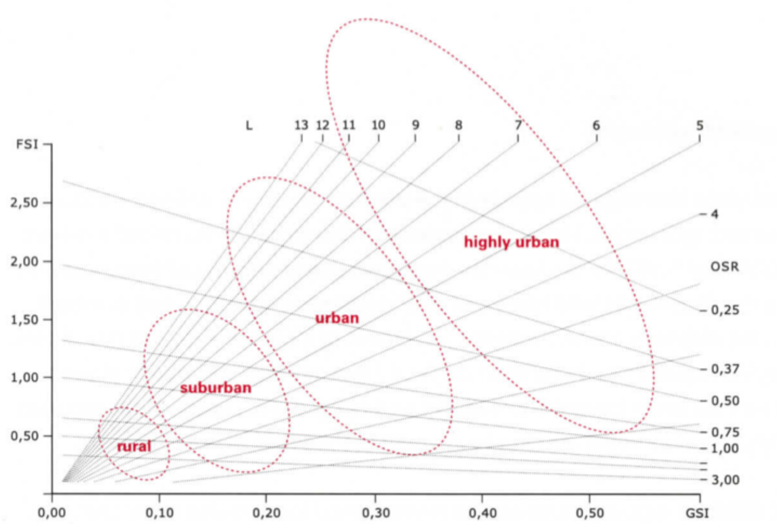


Figure 3.22: Spacemate land use types. Image from Berghauser-Pont (2010)

These measures were calculated for Boston, London and Manchester using these measurements, and are shown in Figure 3.24. The data for Boston was limited to the center of the metro area (consisting of 160,000 buildings), while data for London and Manchester was available for the entire city.

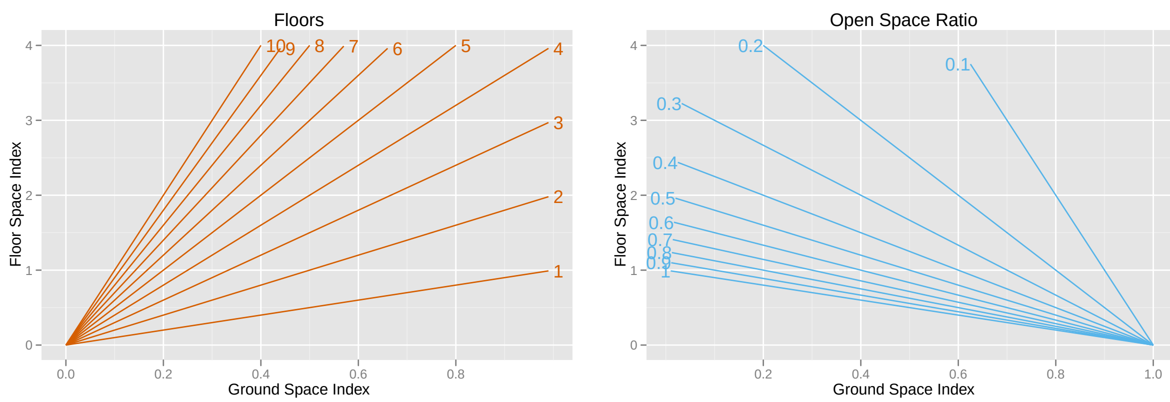


Figure 3.23: Physical limitations of urban space. *Floors* shows the relationship between GSI and FSI, and *Open Space Ratio* illustrates the relationship between these parameters.

From the data shown in Figure 3.24, we can start to develop an understanding of the structure of the city. For example, although less data points are available for Boston, it can be seen that the city is less dense when compared to London or Manchester, as the ranges of *FSI* and *GSI* are lower. In addition, it can be observed that a low *GSI* results in buildings with a lower number of floors (which is to be expected). When the *FSI* starts to increase above 0.2 we can see the the minimum number of floors increases to three stories in all cities.

Overall, this approach is useful as it describes the city using building level measurements. We can start to understand what the city looks like, based on the building heights, and how much space there is between buildings, when measures of *GSI* and *FSI* are considered. The number of tall buildings can gives us an idea about the size of the CBD.

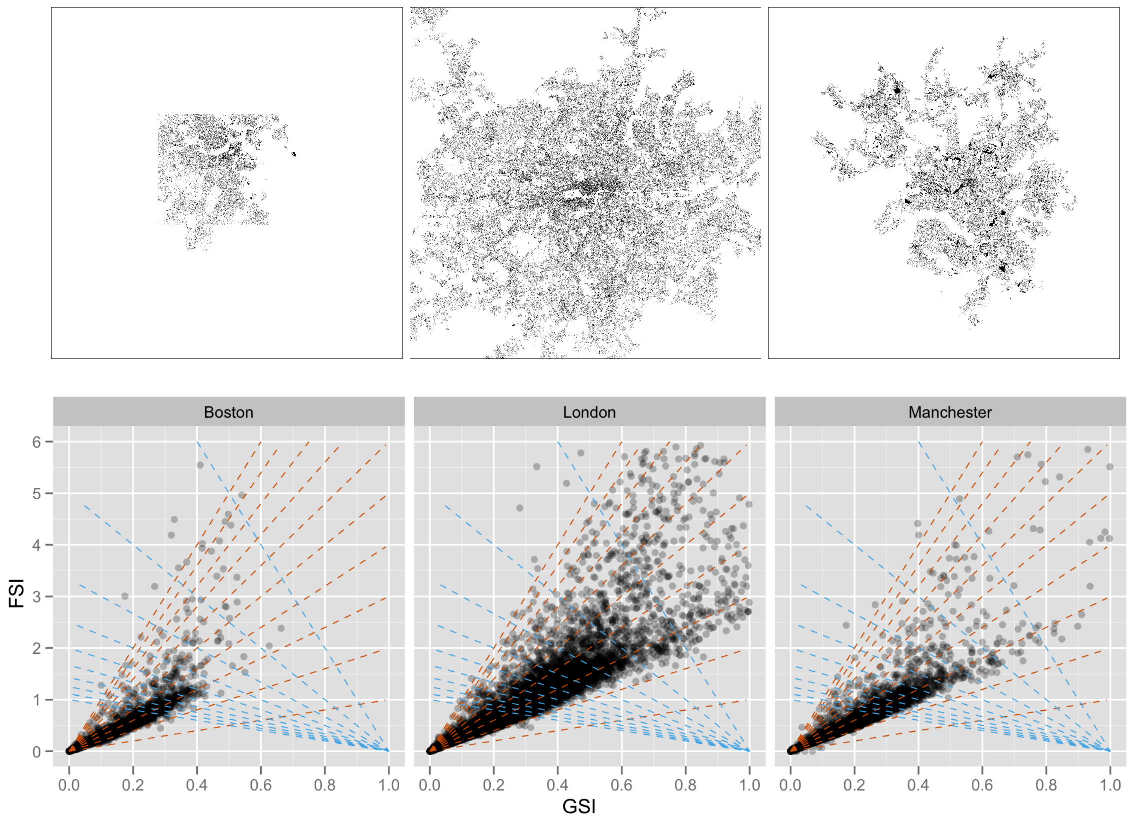


Figure 3.24: Using *The Spacemate* approach for three cities. FSI and GSI were calculated at a 300m grid-cell. All buildings used are shown in the boxes above each plot.

3.6 Conclusions

This chapter describes the process of characterizing urban areas, focusing initially on 40 US cities. I started by exploring density gradient measures within cities, considering general patterns of population and road networks. This gradient approach required the identification of a central business district. While this is a reasonable assumption for mono-centric cities, the assumption does not hold when poly-centric cities are considered. I then approached the characterization of cities using grid level measurements. Finally, building-level calculations were performed to explore empirical ranges of height, volume and open-space. This building level analysis used data from three cities (Boston, London and Manchester).

One objective of this analysis process was to provide a standardized method to characterize a city so that repeatable analyses of cities can be performed.[5] By using the same initial assumptions in each case, a comparison of cities can be made. This analytical approach tries to overcome the systematic problem that exists in many urban analyses, where the results are not replicable, and the underlying assumptions are unclear.

[5] A GIS plugin was developed for *ArcGIS*, written in the *Python* programming language and this plugin was used to analyze the geospatial data (Appendix B.1.1). A future objective of this work is to revise these scripts for *QGIS* (Quantum GIS Development Team, 2011), an open-source geographic information system, so that no proprietary software is necessary for the analysis.

A building level analysis for the entire city, enables the viewer to learn about the spatial structure of the city, as they can intuitively understand the distribution of building configurations. In this way the viewer can quickly observe how tall buildings are, and what the overall form that the city takes. When this is combined with an understanding of the population density gradient, and the infrastructure per person, the viewer can quickly develop a comprehensive understanding of the city.

This work enables the viewer to quickly assess how cities compare to each other. Upper and lower boundaries can also be observed for population and road density gradients. Similarly, an upper boundary for the linear length of road per person was observed, when grid level measurements were used, considering population density. The approach of examining the overall urban area enables the viewer to develop an understanding of the overall distribution of this city. It is clear that high-population density areas have a significantly lower infrastructure requirement per person, when compared to low-population density areas.

This chapter illustrated several different approaches that can be used to simplify the spatial complexity of urban areas. These simplified measures can then be represented in an intuitive way so that clear patterns of behavior are revealed.

Through the identification of upper and lower ranges of parameters, cities can be compared to each other, both nationally and internationally. In addition, the empirical relationships, and parameters that describe these relationships, can be used to estimate values for urban areas where little data is available.

4 Urban Parameters: Neighborhood Scale

The objective of this chapter was to identify relationships amongst geometric measures of the built environment, and to develop a repeatable methodology that could be used to predict measurements of road area, building area, and building height. In Section 4.1, I discuss built environment measurements that are often unavailable from urban datasets. I then discuss how several geometric relationships were identified in Sections 4.1. These relationships were then used to predict built environment characteristics, where data was not available in Section 4.2. The accuracy of these predictions was then assessed using empirical data (based on data availability).

Comprehensive datasets from four cities (Boston, Chicago, London and New York) were used to identify geometric relationships, and these relationships were then used to predict measures for the other 37 cities. This analysis was performed at the block-group[1] level. A block-group is a unit used by the US Census and consists of approximately 1500 people. The block group level was considered to be a useful unit of measurement as it can be related to other data sources such as the *US Census*, the *American Community Survey* and the *UK Census*.

[1] I refer to the UK Census unit the *Lower Layer Super Output Area* (LLSOA) as a block group in this chapter.

4.1 Urban Parameter Relationships

One of the challenges in performing this analysis was acquiring comprehensive spatial and demographic data. Spatial and demographic information was assembled from over 60 cities (Appendix A.1). After assessing the measurements that were available, I decided to focus on 40 cities from the USA (these are the same cities listed in Figure 3.2) and one city from the UK (London). This set of 41 cities was chosen as it provided a dataset with a variety of urban forms and the opportunity to test whether consistent relationships between built environment parameters could be identified, and the opportunity to test if these relationships could be generalized across cities. Based on the data gaps identified from the 60 cities, the focus of this section was to predict the following three measures at a high spatial resolution.

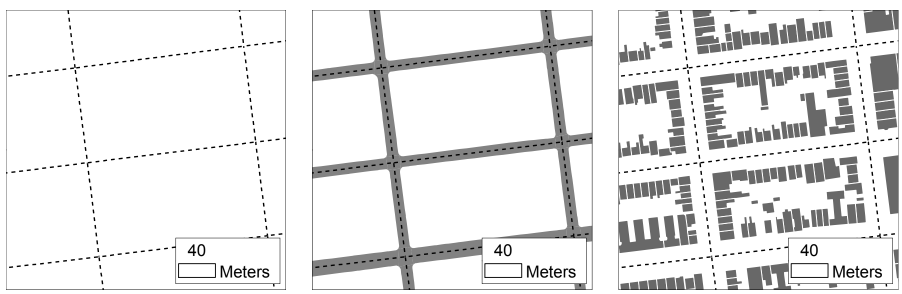
1. *Road-Bed Area*

   Road centerline measurements are a common measure that is often freely available, globally. While the road type is sometimes included in this information, the road width was not available for many of the cities examined. In Section 4.1.1, I examine if there is a geometric relationship that can be used to estimate the road-bed area.

2. *Residential Built Area*
   Building footprints were not easy to obtain for many cities. The ability to estimate the residential built area is useful where building information data is not available due to cost, or due to rapid changes in the physical environment that are not recorded.

3. *Average Building Height*
   Similarly, building heights are not easily to obtain. Six million building footprints gathered from non-commercial sources, and only 1.5 million building footprints had heights associated with them. Hence, developing a method to estimate the average building height was considered to be a useful academic contribution. In Section 4.2.2, I examine whether a relationship between the building height and measures of road area, residential built area and population density exists.

Figure 4.1: Built environment measures for road width estimation.

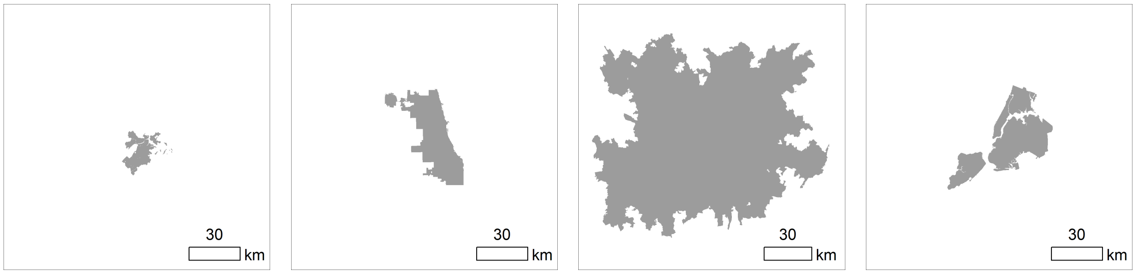


(a) Road centerlines (b) Road-bed (c) Buildings and road centerlines

This process of analysis used comprehensive datasets from Boston, Chicago, London and New York, where road centerlines, road-bed polygons, building area and building height was available, as well as land-use information. The total number of measurements for each city is listed in Table 4.1 with a map of the data available for each city shown in Figure 4.2. These comprehensive datasets enabled statistical models to be identified which were then applied to less complete datasets from other cities.

Table 4.1: Summary of city data. The number of blocks reflects the data availability rather than the actual city size.

City	Buildings	Intersections	Blocks	Avg. Block Area [km^2]
Boston	94 584	14 015	664	0.303
Chicago	427 353	61 233	2 178	0.312
London	964 881	452 459	4 726	0.782
New York	971 439	315 057	6 234	0.944



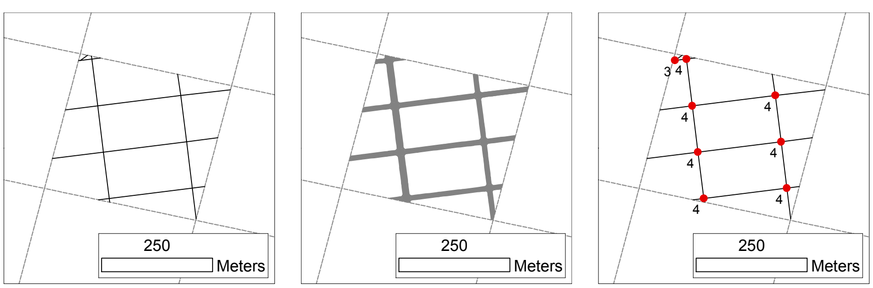
(a) Boston (b) Chicago (c) London (d) New York

Figure 4.2: Map of data for each city used in this analysis.

Spatial measurements were calculated for each block group, and also calculated using a 250m grid. These measurements involved calculating the total linear road length, the road bed area and other measurements of the built environment which are summarized in Table 4.2. An illustration of the process of clipping the spatial data to each grid cell is shown in Figure 4.3. A comprehensive list of all the data used in this chapter is provide in Appendix A.2.

Measure	Process	Unit
Road centerline	Linear length of road	m
Road area	Total road surface area	m^2
Total building area	Total residential building area	#
Average building height	Average height of all buildings	m
Intersections	Number of intersecting roads	#
Intersection count	Number of roads per intersection	#
Intersection nearest neighbor	Euclidean distance to the nearest n intersections (from $n = 1$ to $n = 8$)	m

Table 4.2: Summary of spatial measurements performed, per area of analysis. The method of performing these calculations is discussed in the following sections in this chapter. The area units used were a 250m grid-cell and a block-group.



(a) Centerlines (b) Road-bed (c) Intersections

Figure 4.3: Clipping vector measurements to a 250m grid illustrated using a sample of data from New York.

4.1.1 *Road Width Measurements*

To approximate the road width, urban areas with data describing both the road-centerlines and road-bed polygons[2] were examined. Three approaches were used to estimate the road-width. Eq. 4.3 and 4.4 use geometric measures of vector data, while Eq. 4.4 was based on empirical aggregated measurements. The first approach was as follows:

[2] Road-bed polygons contain the exact information about where the road ends and the curb begins. The accuracy of the road-bed data was not examined. An approach to validate this data could be to use the *RasClass* package (Wiesmann and Quinn, 2011) to identify road-bed areas using remote sensing data.

$$P = 2l + 2w \tag{4.1}$$

where P is the perimeter, l is the length, and w the width. The area A is approximately equal to $w \times l$, assuming that the centerline is straight for short segments. w and l can then be identified from the quadratic:

$$2w^2 - wP + 2A = 0 \tag{4.2}$$

which has the following solution:

$$w = (P \pm \sqrt{P^2 - 16A})/4. \tag{4.3}$$

The second approach assumed that the polygon was long and thin, so that $2l + 2w$ could be approximated to equal $2l$, resulting in the following value for w:

$$w = 2A/P \tag{4.4}$$

While Eq. 4.3 and Eq. 4.4 should produce similar results, the difficulty is that neither w and l are known, so the minimum value was chosen to equal w. The third approach (Eq. 4.5) was to calculate the total road-bed area per grid cell (Figure 4.1 (b)) or block-group and to divide the road-bed area by the sum of linear length of road, l, contained in that spatial unit (Figure 4.1 (a)). w can then be described using the following equation:

$$w = \frac{\sum A}{\sum l} \tag{4.5}$$

Analysis

Measurements from Figure 4.1 (a) and (b) are used for these calculations. All three approaches described in Eqs. 4.3, 4.4 and 4.5 were calculated. Using Eq. 4.5, Figure 4.4 illustrates that there is a clear linear relationship between road-centerlines and the total road-bed area per block-group, where the slope of the line can be used to relate the linear road-length to an average road-width value. Using the approaches described in Eq 4.3 and Eq 4.4, did not result in any consistent estimate of road width, so these approaches were disregarded as it was assumed that the initial assumptions were invalid.

A regression model of the relationship is given in Eq. 4.6. A linear relationship was first tested of the form

$$\sum A = \beta_0 + \beta_1 \sum l \tag{4.6}$$

However this model was mis-specified as it failed to satisfy the *pregibon* test

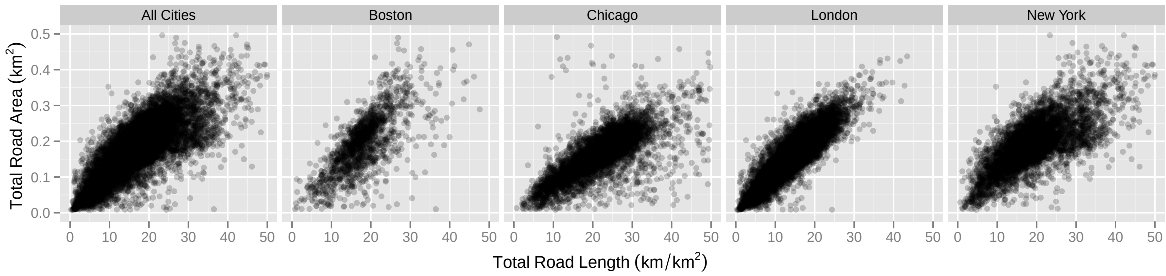


Figure 4.4: Relationship between road length and total road area.

for linearity, so a non-linear term was included (Eq. 4.7).

$$\sum A = \beta_0 + \beta_1 \sum l + \beta_2 \sum l^2 \tag{4.7}$$

The values for this regression are summarized in Table 4.3. The fact that such a strong relationship with a high R^2 is observed in four different cities, and that the coefficients (with the exception of the intercept, β_0, which does not affect the slope) are of similar magnitude and sign suggests that this relationship will hold true elsewhere.

City	β_0	β_1	β_2	Adj. R Squared
Boston	-151.2	12.84	-0.0016	0.7047
Chicago	352.6	8.70	-0.0011	0.6946
London	271.2	11.70	-0.0010	0.9687
New York	2383.0	9.07	-0.0009	0.8253

Table 4.3: Summary of regression analysis for Equation 4.7. The p-value and test were significant at the 99th percentile confidence interval.

Using the same regression model structure, the parameters for a model using the data from all cities are listed in Table 4.4. Using Eq. 4.5, the average road width measure is illustrated in Figure 4.5 as a histogram.

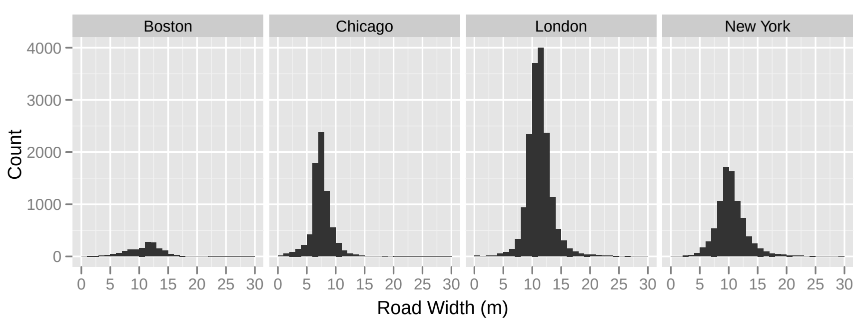


Figure 4.5: Histograms of road-width for the four cities examined.

Table 4.4: Relationship between linear road length and road area for all cities, calculated at the block group level. The p-value and test were significant at the 99th percentile confidence interval.

City	β_0	β_1	β_2	Adj. R Squared
All Cities	480.0	8.86	-0.046	0.6888

4.1.2 *Intersection Density*

Intersection density is an important measure for measuring the walkability of urban form (Dill, 2004). In a review paper by Ewing and Cervero (2010) the authors emphasize the importance of how intersection density influences the walkability of urban neighborhoods. I examined the intersection density for all cities in this work, using the road network to perform calculations. I explore some empirical relationships between the intersection density and other measurements related to the road network structure. This approach was used to explore if any relationships between built environment measurements and road network properties could be identified. The intersection measurements were calculated with the goal of categorizing the built environment, and with the hope of applying these measures to prediction models in this analysis. Road networks from the same four cities are used to examine built environment measurement relationships. In this case two of the cities have grid-systems (New York and Chicago), while two are non-gridded (Boston and London). These patterns are illustrated in Figure 4.7.

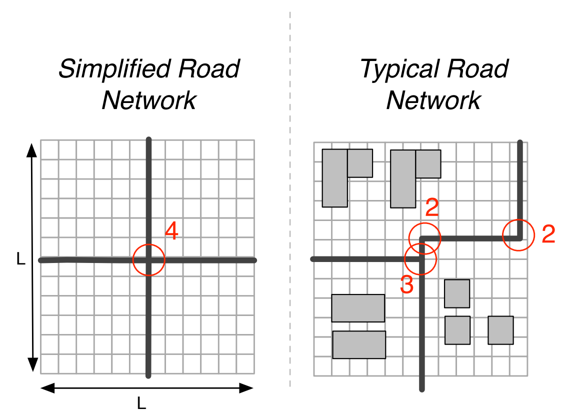


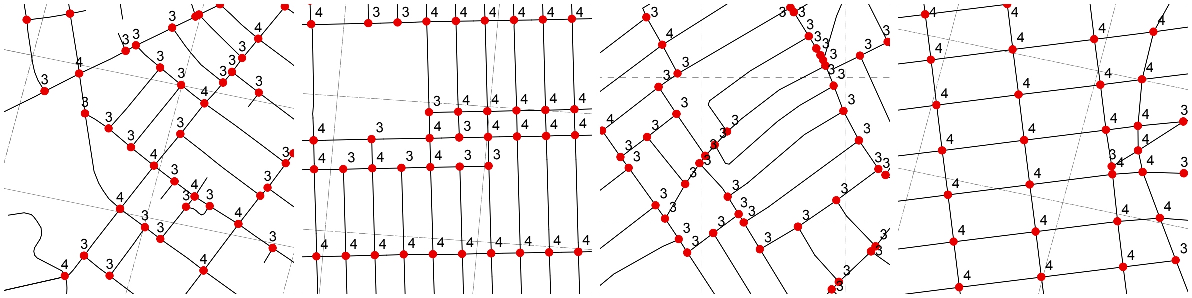
Figure 4.6: Road network diagram. In the *Simplified Road Network*, $\sum i = 1$, $\sum i_c = 4$ and in the the *Typical Road Network*, $\sum i = 2$, $\sum i_c = 5$ as intersections with $i = 2$ are ignored.

In the *Simplified Road Network* example (Figure 4.6), the road length of each segment is l, the number of intersections i and the intersection count i_c. As the road length can be used to predict the road width, I am interested in the relationship between the average road length and the number of intersections. I hypothesize that that the average road length, l_a can be related to the intersection density using the following equation:

$$\frac{\sum l}{\sum count} \propto \frac{\sum i}{Area} \tag{4.8}$$

or where p_i is an empirically observed constant for each city.

$$l_a = \frac{\sum i}{Area} p_i \tag{4.9}$$



(a) London (b) New York (c) Chicago (d) Boston

Figure 4.7: Road segments (black) and intersections (red). A grid with 250m spacing is shown.

Analysis

I explore the hypothesis described in Eq. 4.9 which attempts to relate the average road length to the intersection density. The motivation to estimate the number of intersections using a simpler approach is due to the technical challenges associated with performing road intersection calculations. Road network calculations are difficult to perform using GIS, due to software limitations. While*ArcGIS* provide a tool to perform this analysis (*Network Analyst*), it is a plugin for the most costly version of *ArcGIS*. In the open-source GIS world *PGRouting* is available as an extension for *PostGIS*, but this library is several versions behind the current *PostGIS* release.[3] The method described in this section is a useful quasi-network theory approach to identifying road network properties.

[3] Alternatively, there are many non-GIS packages available for this type of analysis; *NetworkX* (a python library) and *Gephi* (standalone network analysis tool) are two popular options.

There were several steps required to calculate the number of intersections. First the road network was split into road segments at every vertex so that each road segment was a straight line, rather than a multi-point road segment. Duplicate road segments were removed by checking if they had an identical geometry when compared to any other line segments (this is a common occurrence, and the problem is illustrated in Appendix C.4). The *SQL* code for the procedure to remove duplicate line segments is listed in Appendix C.4.

PostGIS was then used to perform a calculation that counted any line strings (or vertices) that touched each other. Intersections (i) were identified where vertices touched each other, and the number of connecting lines (i_c) was recorded (Appendix C.3). Any intersection that had a value $i_c = 2$ was disregarded as this is considered to be the a continuation of a road, but due to the storage method of line segments it is recorded as an intersection, with a road count of 2. The mean value of the the number of connections per intersection is summarized in Table 4.1.2. As expected, the mean number of roads per intersection in gridded cities is closer to 4.0, when compared to cities without a grid. In addition to the intersection identification and road count, a linear measurement of the euclidean distance to the nearest n

City	Avg. Node
Boston	3.29
Chicago	3.49
London	3.10
New York	3.58

Table 4.1.2

intersections was performed for every node on the road network, from $n = 2$ to $n = 8$.

The relationship between the number of intersections per area (referred to as *Intersection Density* and the total number of roads connected per intersection is illustrated in Figure 4.8. This illustrates that there is a similar pattern of behavior for all cities, with an obvious linear relationship. The interpretation of this image is that the number of roads per intersection, is constant for each city.

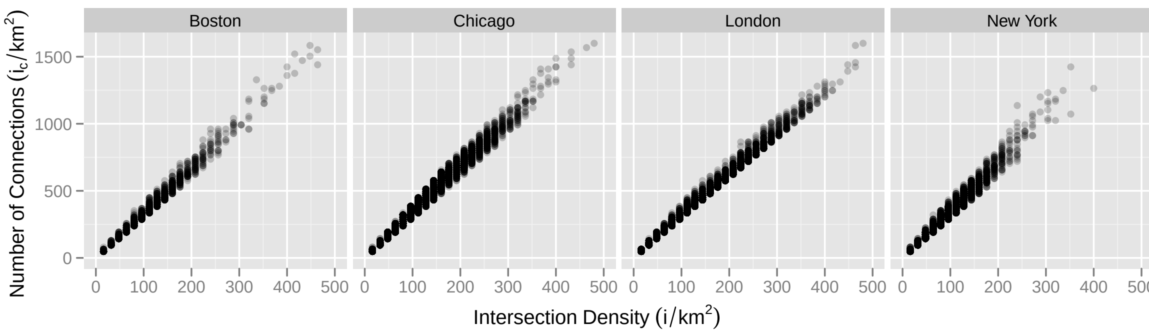


Figure 4.8: Relationship between intersections and connections.

The hypothesis that the average road length per area could be related to the intersection density did not hold as no clear pattern could be observed. However, a relationship between the *Linear Road per Area* (road density) and the intersection density was observed, shown in Figure 4.9.

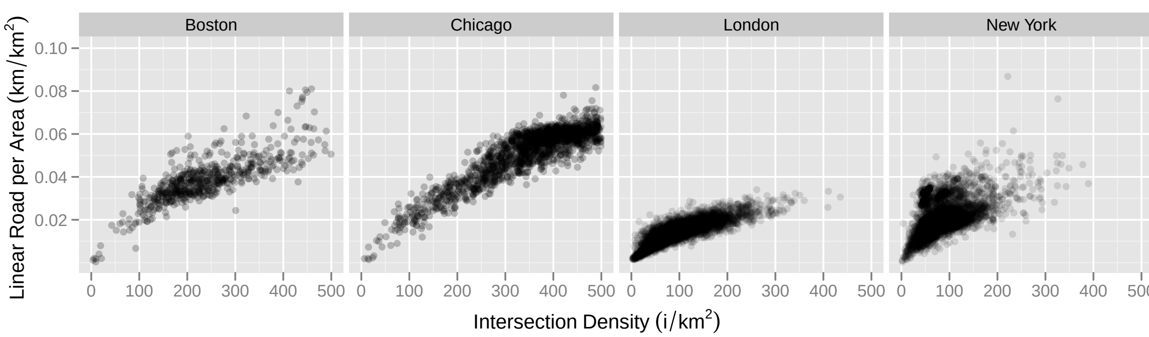


Figure 4.9: Relationship between centerline road density and the intersection density.

It is striking that the road density for Chicago is significantly higher than that of London, for high intersection densities. This is perhaps due to the differences between a gridded network and non-gridded network.

4.1.3 *Population Density*

Similar to the work described in Chapter 3, there are clear relationships between the population density and other urban form measures per block group. These non-linear relationships between *Road Density* and *Road per Person* are illustrated in Figures 4.10 and 4.11.

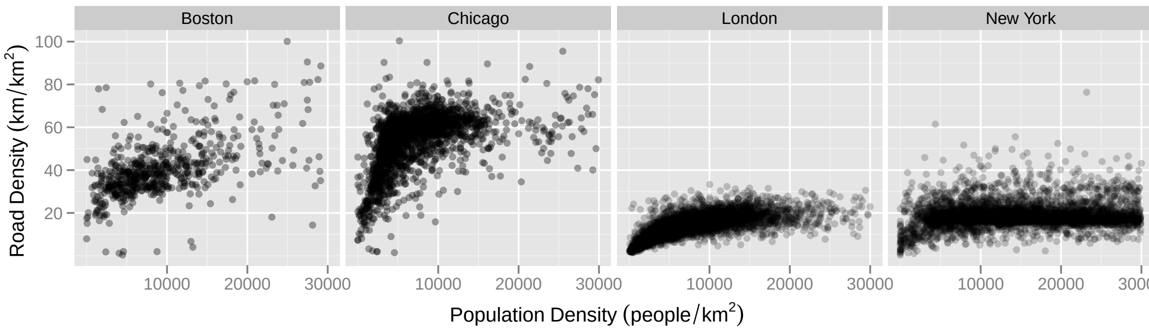


Figure 4.10: Population density and road length.

As there is an almost linear relationship between road-area and road length (described in Section 4.1.1), the plot of population density and road area is similar to Figure 4.10. When the *Road per Person* is considered, it can be observed that the amount of infrastructure per person decreases as the population density increases and that there is a consistent pattern for all cities (Figure 4.11). At high density levels, the road per person drops below 5 m per person.

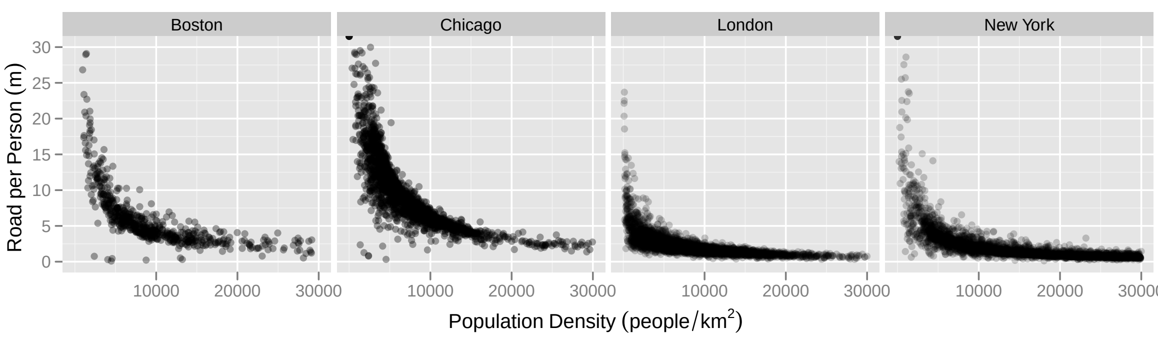
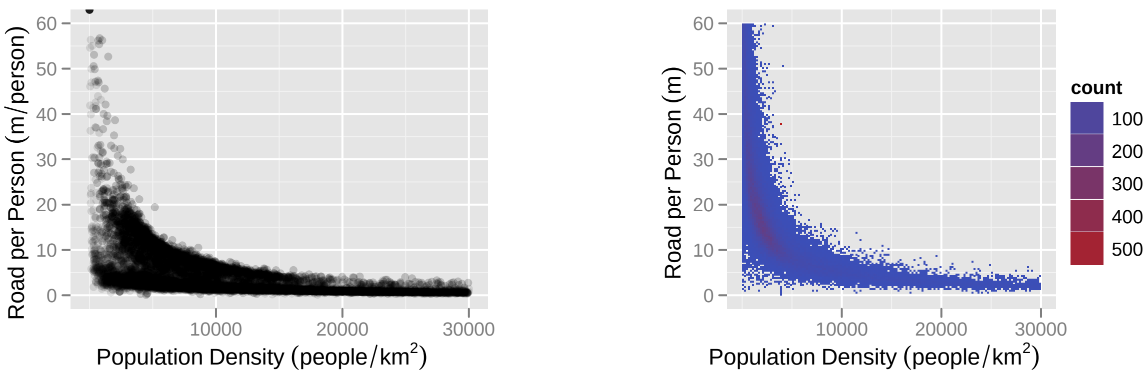


Figure 4.11: Road length per person, calculated using block group level averages.

This is calculated for 106305 block groups from the 40 US cities used in this study. A clear upper and lower boundary can be seen (Figure 4.12), as well as the frequency of observations.

To consider how efficient the infrastructure is, it is necessary to consider the unit of service that it provides. Here, the unit of service is considered to be linear road length per person. If we assume that the infrastructure is used equally by the residents of a city, we can normalize by the number of people in that area to approximate an infrastructure efficiency measure, as shown in Figure 4.12. For each city, a similarly shaped curve can be observed which



(a) Scatterplot

(b) Histogram

Figure 4.12: Road length per person for all 40 US cities calculated per block group using 106305 observations.

suggests there is an upper and lower limit to the amount of road per person, once a certain density level is achieved. While the maximum amount of road per person is most likely a result of the planning regulations and standards in each area, this visualization approach illustrates the value of characterizing the spatial relationship between population and road infrastructure, as an approach to considering urban resource efficiency.

4.2 Residential Building Prediction

In this section, models to predict the residential building area and average building height at the block-group level, are identified. These models use a combination of spatial measures from Boston, Chicago, London and New York. Several non-linear statistical approaches were explored (including multinomial logistic regression and classification trees), however the statistical approaches used were multi-variate linear regressions. This approach was chosen as it was possible to predict building parameters with a small number of spatial measurements, and an acceptable level of accuracy and is an easily transferable approach. For data parsimony reasons, the only census measurements used were the total number of people and the land area of each block group (the number of housing units per block group was not used).

4.2.1 Residential Building Area

Through a preliminary exploration of relationships between variables, it was observed that the number of buildings per block-group could be related to the total building area per block group. A regression model was identified to predict the number of buildings per block group (Eq. 4.10). All distances are calculated in m, and all density measures were calculated per m^2.

$$\begin{aligned} \text{Num. Buildings.} &= \beta_0 + \beta_1 \, log(\text{Pop.}) + \beta_2 \text{ Intersections} \\ &+ \beta_3 \, log(\text{Land Per Person}) + \beta_4 \, log(\text{Road Dens.}) \end{aligned} \tag{4.10}$$

β_0	β_1	β_2	β_3	β_4	Adj. R Squared
-1143	119.1	1.325	86.59	2670	0.4616

Table 4.5: Parameters to predict the number of buildings per area. The p-value and test were significant at the 99th percentile confidence interval and satisfied the Pregibon test.

It can be seen that as the *Population* increases, the *Number of Buildings* increases. Similarly, as the count of *Intersections* increases, the *Number of Buildings* increases. As the *Land per Person* increases, the *Number of Buildings* also increases. The sign of the β_3 coefficient is less obvious as it is a result of buildings in higher density areas, being shared by more people. For example, a suburban neighborhood has a higher number of buildings per person when compared to a high-density area with multi-family dwellings in the city center. Finally, as the *Road Density* increases the *Number of Buildings* increases. A more general interpretation of this relationship, is that the factors which describe the level of urbanization (intersections, land per person and road density), in combination with the total number of people, enable the number of buildings to be estimated.

Next, a model was identified to predict the total residential built area (Eq. 4.11). The residential built area was normalized by the total land area so that the regression predicted a building area fraction. This model is first identified using the empirical number of buildings as one of the parameters. The R^2 of the model when using the predicted number of buildings from Eq. 4.10 is also stated.

$$\begin{aligned} \frac{\text{Residential Built Area}}{\text{Area}} &= \beta_0 + \beta_1 \, log(\text{Pop. Dens.}) + \beta_2 \, log(\text{Road Dens.}) \\ &+ \beta_3 \text{ Intersection Dens.} + \beta_4 \, (\frac{\text{Num. Buildings}}{\text{Area}})^2 \end{aligned} \tag{4.11}$$

Table 4.6: Parameters to predict the number of buildings per area. The p-value and test were significant at the 99th percentile confidence interval, and both of these models satisfy the Pregibon test.

β_0	β_1	β_2	β_3	β_4	Adj. R Squared
0.05111	3.595e-06	2.315	199.8	10270	0.6476

The relationship in Eq. 4.11 shows that as all of the independent variables increase, the *Residential Built Area* increases. This is logical, as an increase in *Population Density* or the *Number of Buildings* will cause an increase in the square footage of residential buildings. Similarly (and less obviously) an increase in *Road Density* or *Intersection Density* results in an increase in *Residential Built Area*. When the predicted *Number of Buildings* was used to estimate the *Residential Built Area*, the R^2 decreased to 0.5520. To calculate the predicted residential built area per block group, the predicted area fraction was multiplied by the total block group area.

4.2.2 *Average Building Height*

The final urban form measurement that was considered was a method to estimate the average building height. With this measure, in combination with the residential building area, the total building volume can also be estimated. The building volume can be then be used to estimate the resource intensity of material and energy. A model was identified to estimate the average building height at the block group level. This approach used geometric measurements and some of the predicted values observed in the earlier models. An equation with the following structure was identified:

$$\begin{aligned} log(\text{Avg.Height}) &= \beta_0 + \beta_1 \text{ Pop. Dens.} + \beta_2 \, log(\text{Built Area per Person}) \\ &+ \beta_3 \text{ Road Dens.} + \beta_4 \text{ Intersection Dens.} \end{aligned} \tag{4.12}$$

Table 4.7: Parameters to predict the average height. The p-value and test were significant at the 99th percentile confidence interval, and this model satisfies the Pregibon test.

β_0	β_1	β_2	β_3	β_4	Adj. R Squared
3.176	1.183e-05	-0.3006	0.04744	0.04078	0.4371

From the parameters listed in Table 4.7 it can be observed that as *Population Density* increases, the *Average Height* increases. As the *Built Area per Person* increases the *Average Height* decreases. An increase in either *Road Density* or *Intersection Density* result in an increase in the *Average Height*. When the predicted area value is used from Eq. 4.11, the R^2 value for the *Average Height* decreases to *0.3349*. Models with logical parameters were identified with R^2 values close to 0.7 but these models did not satisfy linearity requirements. Using this relationship to predict the average height per block group (Eq. 4.12), and the previous relationship to predict the residential area (Eq. 4.11) per block-group, the average building volume could be predicted.

4.3 Validation of Parameter Relationships

The spatial measurements listed in Table 4.2 were then calculated for 37 cities in the USA, at the block-group level. There were a total of 106,305 block groups used in this analysis. The data used for these calculations came from the 2010 US Census, and the *2010 Road TIGER Line* data. The projection used was *Albers Equal Area Conic* (SRID:102008). These spatial and statistical measurements were then used to predict the amount of residential built area per block group, and the average height of buildings per block group.

To perform this analysis, the computational requirements of the analysis need to be considered. These datasets were loaded into a *PostGreSQL* database and spatial analyses were performed using the PostGIS extension. Using a spatial database enabled large amounts of data to be processed in a timely manner. It took approximately six hours to load the USA road network dataset into the database, and six to twelve hours for every major spatial operation. For example, to calculate the linear road length per block group took one day using a 6 core CPU of 3.2GHz with 18 GB of RAM. This process used PostGreSQL 9.1 (64 bit), with PostGIS 2.0 (alpha) running on Windows 7. This process was primarily CPU intensive. The statistical analysis was performed using *R*. The *RPostGreSQL* driver was used for interoperability between *PostGreSQL* and *R*; this enabled data to be read from the database, analyzed, and written back to the database.

Using this workflow, the regression models identified to predict the *Residential Building Area* (Eq. 4.11) and the *Average Building Height* (Eq. 4.12) were used to predict the building area and average building height for 106,305 block groups. To calculate the accuracy of these predictions, samples of data from three cities, Los Angeles, Seattle and Huston was used. Data from these cities was not used in the initial model specification. Data for Seattle and Houston was downloaded from the cities respective websites; the data for Los Angeles was provided upon request for a small sample of residential parts of the city.[4]

[4] Mark Greninger from the LA City GIS department kindly provided 10402 buildings from twenty block groups that I randomly chose. The LA building model will be in the public domain by the end of 2012.

4.3.1 Built Area Error

To estimate the accuracy of the residential area prediction model listed in Eq. 4.11, the total building footprint area was calculated per block group for several urban areas where data was available, and compared to the values predicted from the model. In these building datasets, there was no information about whether the buildings were residential or not, and the building outlines are from 3 - 7 years ago. The predicted and actual areas are shown for Los Angeles, Seattle and Houston in Figure 4.13. The red line illustrates what a perfect prediction value would be. In the cases of Los Angeles and Seattle the points appear to be centered around the line, while in Houston the values appear to be slightly overestimated.

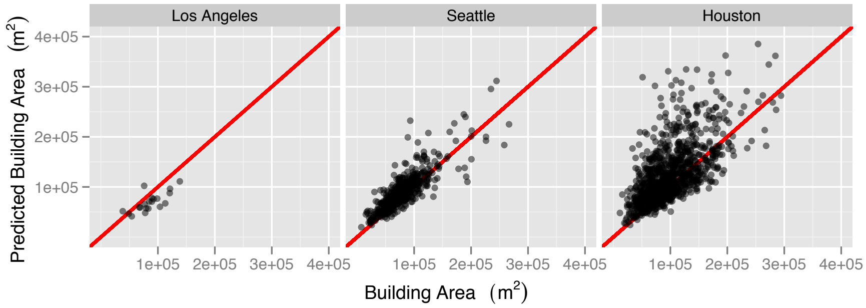


Figure 4.13: Predicted and actual areas for three urban areas.

The magnitude of the errors for each city, per block-group are summarized in Table 4.8. The error was calculated by measuring the absolute error using the *RMSE* and then normalized[5] by the average block group area, or by the total city area. When the *RMSE* is calculate at the city level, the magnitude of the errors can be observed in Table 4.9. While the error at the block-group level ranges from 0.2-0.56, when the relationship between the number of people, and the total residential building area is observed in Figure 4.14 this appears to be a satisfactory prediction.

[5] The error normalized by the average observation is sometimes referred to as *relative error* and also referred to as the *coefficient of variation*.

City	RMSE km^2	Avg. Block Group km^2	RMSE / Avg. Block Group -
Los Angeles	0.024	0.086	0.28
Seattle	0.034	0.085	0.40
Houston	0.055	0.099	0.56

Table 4.8: Summary of errors for predicted urban areas at the block-group level.

City	RMSE km^2	City Area km^2	RMSE / City Area -
Los Angeles	0.33	1.72	0.20
Seattle	4.27	48.40	0.08
Houston	27.06	99.85	0.27

Table 4.9: Summary of errors for predicted urban areas at the city-level.

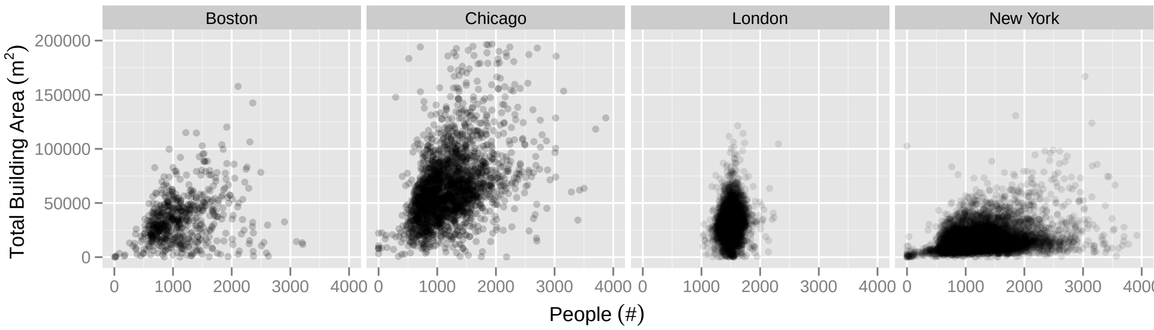


Figure 4.14: Relationship between people and residential building area.

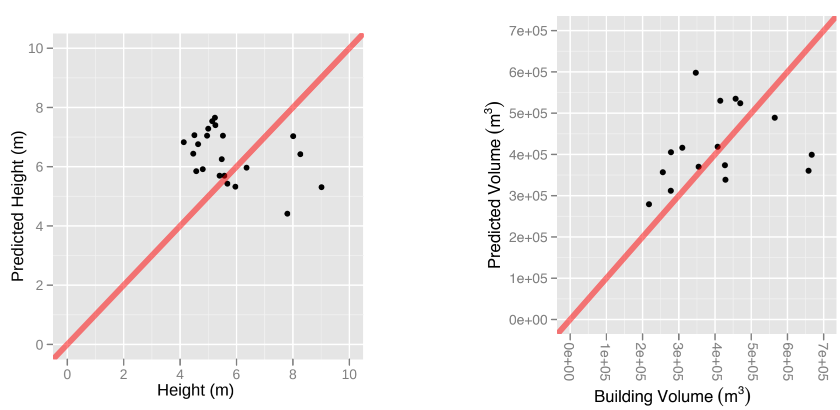
4.3.2 *Height and Volume Prediction Error*

The only data that was available to compare the average height estimates (not considering New York, Boston or Chicago) was the dataset from Los Angeles. The average building height per block group was calculated and compared to the predicted value using Eq. 4.12, illustrated in Figure 4.15. The estimated building volume per block group was compared to the predicted values (Figure 4.15) using this data. Similarly the Average Height errors are summarized in Table 4.10. Unfortunately this data sample is quite small so it is difficult to assess how this model predicts buildings with greater variability than the sample of low-density residential buildings from Los Angeles.

Table 4.10: Summary of Average Height errors using data from Los Angeles.

City	RMSE	Avg. Building Height	RMSE / Avg. Building Height
	m	m	-
Height	1.86	5.57	0.33

Figure 4.15: Building Height and Volume prediction.



(a) Predicted and actual height for twenty block-groups.

(b) Predicted and actual volume for twenty block-groups.

4.4 Conclusions

In this chapter, the objective was to identify relationships amongst geometric measures of the built environment, and to develop a repeatable methodology that could be used to predict measurements of road area, building area, and building height.

Using data from four cities with comprehensive datasets, multi-variate linear regression models were identified, and these models were used to predict the road area, the building area and the average building height for 37 cities at the block-group scale. The accuracy of these models was then assessed by comparing the predicted values to empirical data (although this was limited, due to data availability).

These statistical models used geometric measurements of the road network, and a measure of the population density. No other socio-economic measures were included. The use of a small number of variables, enables this approach to be easily applied to other datasets. This approach can be used by researchers and planners to estimate characteristics of the built environment. This procedure is of benefit in areas where data is not available, or when the cost of acquiring the data is too high.

5 Resource Intensity

In this chapter, I examine the relationships between urban form measurements and the material and energy intensities associated with these measurements. In Section 5.1 I discuss how geometric measurements of the urban form can be related to the quantity of construction material stock that exists in cities.[1] In Section 5.2 I discuss how the energy intensity of cities varies spatially.

In Sections 5.3 and 5.4, I illustrate how these measures can be used in combination to assess the resource intensity of several different urban forms, at the neighborhood and city scale. Section 5.3 focuses on spatial patterns of the material stock in London, while Section 5.4 considers material and energy measurements for 40 US cities at the block-group level.

[1] I am not considering any flows of construction material; the purpose of this work is to estimate the stock of material. Energy calculations consider the total energy use in one year.

This chapter examines material calculations for two different countries. In the analysis of residential buildings in the USA, the predicted and measured geometric measurements from Chapter 4 are used to estimate the quantity of material necessary to construct residential buildings and roads. In the UK analysis a similar approach is used, with the geometric measurements of road area and buildings already available for London. In both cases, these geometric measurements are converted into kilogram measures of construction material using data from construction standards, building surveys and tax-assessor information. It should be emphasized that these estimates are not validated or compared to the existing conditions. These material calculations are preliminary estimates of the material stock associated with a variety of geometric measurements.

This chapter also discusses residential energy use; both energy use within the building and the energy associated with private automobile travel. My primary contribution in this section, is the development of a method to predict the *Vehicle Kilometers Travelled* (VKT) and associated energy, using a variety of spatial measurements (Section 5.2.1). In Section 5.2.3, residential building energy consumption is discussed. While some observations are made using empirical data from New York, a published and validated model from Efficiency 2.0 (Min et al., 2010) is referred to in this thesis and used in combination with the estimates of energy from vehicular transportation.[2]

[2] Zeke Hausfather of *Efficiency 2.0* kindly shared the data from a study that estimated the average household energy consumption at the zip code level for the USA.

5.1 Material Intensity

Using geometric measurements of roads and buildings, I estimate the material content per block-group. This measure can then be normalized by the number of people or households to provide an intensity measure. A description of data used to convert geometric data into material units is described in the following sections.

5.1.1 Material for Road Infrastructure

To estimate the material content of roads, the road-bed area and typical road construction standards are used. This calculation only accounts for the initial construction and does not include additional material inputs over the service life of the road bed. The approach to calculating the road-area per block group has been discussed in Section 4.1.1 and this value is multiplied by the cross-sectional area of a typical road and the density of the material (Eq. 5.1).

$$\text{Material} = \text{Predicted Area} \times \text{Cross Section} \times \text{Material Density} \tag{5.1}$$

Several data sources were used to estimate the amount of material used to construct infrastructure and residential buildings. Data was gathered from a range of road-construction standards and guidelines in the US and UK. Chudley and Greeno (2008) specify 125 mm of gravel and 60 mm asphalt for typical road cross-sections. The specific road construction material used for local and secondary roads depends on state construction standards and local ground conditions. These data sources came from came from guidelines for road construction published by UK county councils (Aberdeenshire Council, 2008; Cambridgeshire City Council, 2007; London Borough of Croydon, 2009; Thurrock Borough Council, 2005; Worcestershire City Council, 2011), aggregated surveys of existing buildings for Communities and Government (2010), and typical construction methods (McMorrough, 2006). These values are summarized in Table 5.1. These values were used to convert geometric measurements of different road types into kg of materials.

Table 5.1: Road construction and material properties.

	Asphalt	Gravel	Unit
Thickness	0.06	0.125	$[m]$
Density	2200	1922	$[kg/m^3]$

A calculation of the material required per m^2 of road area is shown in Table 5.2. Considering a *road area per person* measure, the infrastructure requirements per person for a range of population densities are given in Table 5.3. The variation of road per person and population density can be seen in Figure 5.2 for all 40 cities. These measures illustrate how the amount of construction material varies when population density is considered. The upper and lower ranges of road area per person are shown in Figure 5.2, with histograms of slices at 10000, 20000 and 30000 $people/km^2$ shown in Figure 5.1. The average values from Figure 5.1 are used in Table 5.3 to summarize the material required for road infrastructure at varying population density levels.

Asphalt	Gravel	Total	Unit
132	240.25	372.25	$[kg/m^2]$

Table 5.2: Material requirement per m^2 of road area.

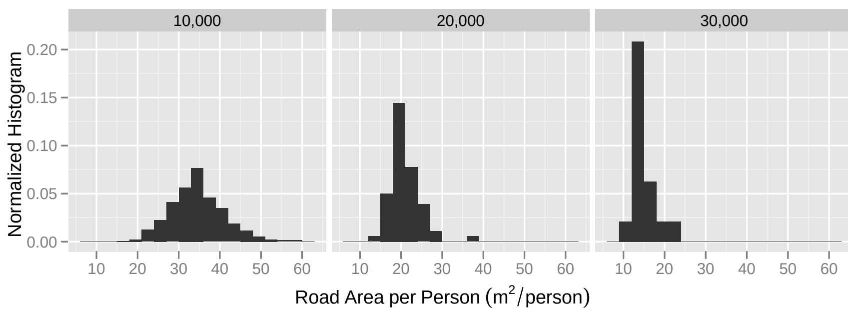


Figure 5.1: Histograms of road area per person for varying population density levels.

Pop. Density $[people/km^2]$	Average Road Area Per Person $[kg]$	Material Per Person $[kg]$
10 000	34.64	12894
20 000	21.02	7825
30 000	14.91	5550

Table 5.3: Construction material required for roads, considering population density levels.

These material requirements are now applied to the predicted road area values. An example of the material per grid cell is shown for New York and Boston in Figure 5.3. Here the total road area per grid-cell is multiplied by the kg/m^2 values listed in Table 5.2. To calculate the intensity per block-group, the total road area was predicted and this area was multiplied by the material per area. This value was then normalized by the number of people or households per block group. This infrastructure measurement is subsequently used in Section 5.4 where I discuss the material intensity of cities.

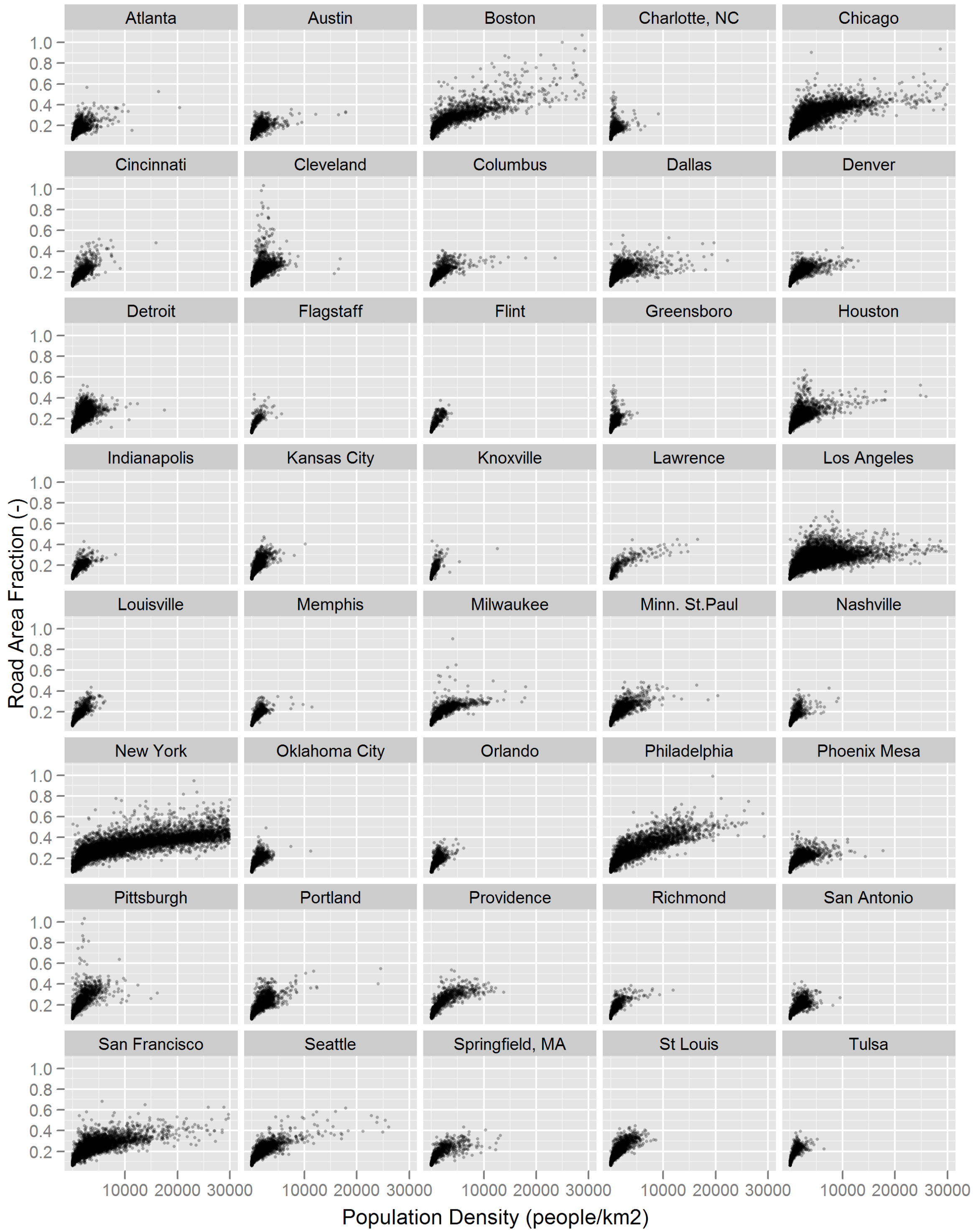
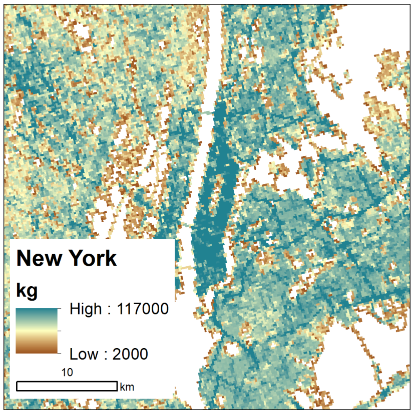
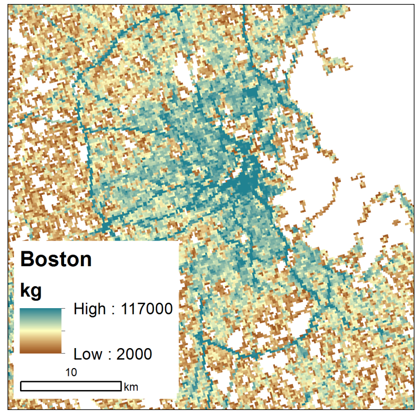


Figure 5.2: Road area fraction for varying population density levels per city.



(a) New York

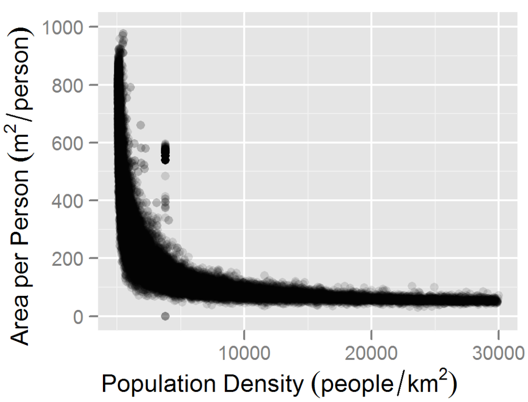


(b) Boston

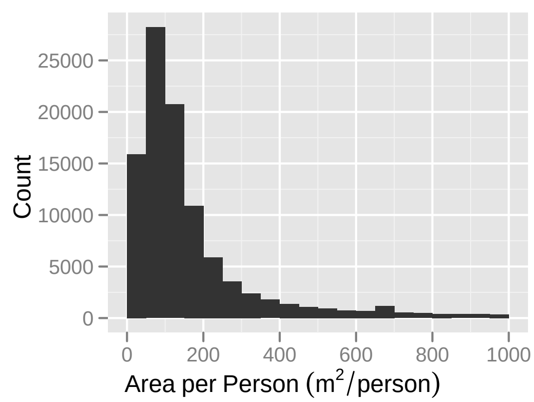
Figure 5.3: Total material content in road infrastructure per grid-cell.

5.1.2 *Building Material Intensity*

The models identified in Section 4.2.1, were used to estimate the geometric measurements of floor, roof and wall area using Equations 4.11 and 4.12. The variation of building area per person shown is illustrated in Figure 5.4.



(a) Building Area per Person



(b) Histogram of Building Area per Person

Figure 5.4: Built area per person measurements.

A set of slices is taken through Figure 5.4 (a) at population densities of 10000, 20000 and 30000 $people/km^2$. Histograms of these slices are shown in Figure 5.5 to illustrate the per person variation of built area. Using the average height, and the average building proportions (Table 5.5) the geometric measurements per building are listed in Table 5.4.[3] The measures shown in Table 5.4 are for three population density levels; this histogram illustrates that there are clear differences in the residential area per person, at varying population density levels.

[3] While population density is a factor in the area prediction model (Eq. 4.11) there are several other factors that are also involved in the building area estimation.

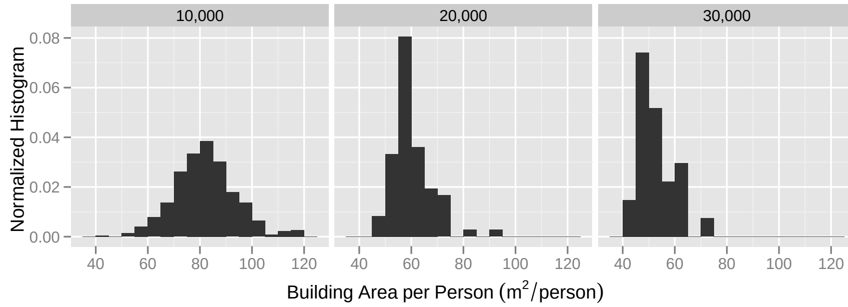


Figure 5.5: Building area per person histograms for varying population density levels. This is a cross-section of the data shown in Figure 5.4 (a).

Table 5.4: Geometric areas per person, considering population density levels.

Pop. Density [$people/km^2$]	Building Area [$m^2/person$]	Roof/Floor Area [$m^2/person$]	Perimeter [$m/person$]
10 000	81.89	33.95	30.96
20 000	59.94	20.30	22.66
30 000	52.94	15.64	20.02

In Table 5.4, the building area is the total floor space in a building; the roof/floor area is based on the building footprint. The average *area/perimeter* ratio for buildings in the four cities analyzed in detail was 2.645; this value was used to estimate the perimeter of buildings. This perimeter was then multiplied by the average height to provide an estimate of the wall area.

5.1.3 *Material for Building Construction*

Using several assumptions about the overall construction details, conversion factors for each residential building element are listed in Table 5.6. In the US, the mixture of the building stock was chosen using data from new construction mixes over the last 30 years (US Census Bureau, b). A more accurate approach would be to use regionally specific values. Only two construction types were considered, *wood stud* and *masonry cavity*. Based on a survey of the American Housing Survey (US Census Bureau, a), it was estimated that 70% of the residential building stock consisted of wood-stud with 30% masonry.

In the UK, residential buildings were identified from the UK building

Table 5.5: Average building proportions per city

City	Area/Perim.
Boston	2.51
Chicago	2.29
London	3.10
New York	2.68

House Type	Element	Masonry $[kg/m^2]$	Glazing $[kg/m^2]$	Timber $[kg/m^2]$
Masonry Cavity	Wall	480	10	-
	Roof	-	-	21
	Floor	-	-	32
Wood Stud	Wall	-	10	27
	Roof	-	-	21
	Floor	-	-	32

Table 5.6: Material requirement per m^2 of floor, roof and wall area. Data from (National Association of Home Builders (U.S.) and Bank of America Home Equity, 2007; Chudley and Greeno, 2008)

dataset UK Ordnance Survey (2010a). Using data from the English Condition Housing Survey (EHCS), information was obtained about the construction type of these residential houses. Then, based on typical construction methods for wall, floor, and roofs (Table 5.6) these conversion factors were used to convert the geometric measurements of buildings, into kilograms of materials. In the UK, as over 92% of houses in the EHCS were constructed of masonry (masonry cavity 64.7%, solid masonry 27.2%) a typical masonry house was used to estimate the material used by the average house in each cluster type. The average house in London had a glazed area of 37% of the wall area.

5.2 Energy Intensity

In this section, I examine variations in population density, transportation infrastructure and services for 40 US cities, and relate these variations to transportation infrastructure and vehicular travel patterns. The objective is to identify the effect of the spatial structure at the meso-scale on vehicle kilometers travelled (VKT). Geometric measurements of the road network are used in addition to the distance to a range of services. A model is identified and validated at the block group level for the state of Massachusetts, and the parameters used in this model are then applied to other urban areas in the US. Aggregate city level VKT predictions are then compared to empirical VKT data for US metro areas.

5.2.1 Transportation Energy

A *distance to services* measure was calculated for each grid-cell using the euclidean distance to nineteen mixed-use services.[4] The North American Industrial Classification System (NAICS, 2011) was used to classify the services listed in Table 5.7. The location of each service was identified from ESRI Business Analyst data (ESRI Business Analyst, 2008). The euclidean distance from each grid-cell to each of the nineteen services was calculated (Figure 5.6) and these values were then summed together. The *distance to services* measure is the total distance from each grid-cell to all nineteen services. This distance is used as an index to characterize the urban form and represents the availability of nearby services.

[4] This list of services is from the USGBC's Leadership in Energy and Environmental Design for Neighborhood Development guidelines, where they suggest developments should be within a certain distance of services on the 'Diverse Uses' list.

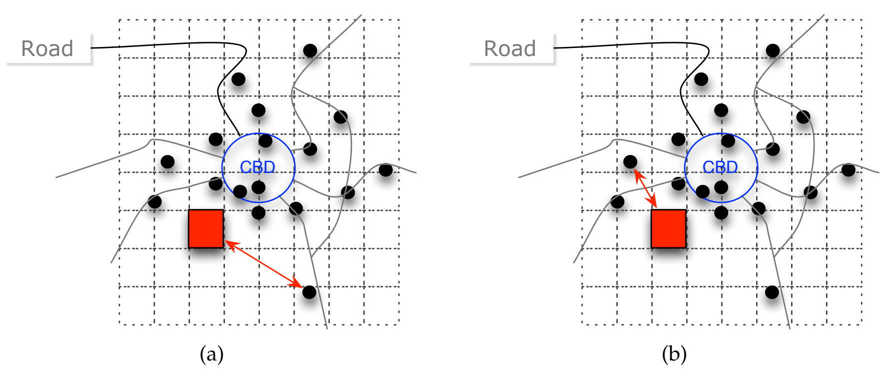


Figure 5.6: Calculating the euclidean distance to services. The red square represents one grid cell, and the small circles represent business locations. The euclidean distance to each service was measured from each grid-cell.

Figure 5.7 illustrates the *distance to services* in 3D for each grid cell in Atlanta. It can be observed that the distance to services is low in the city center (near the CBD) and high towards the city's perimeter. It should be noted that the surface is quite rough, with many local valleys and peaks, which are likely influenced by road network patterns.

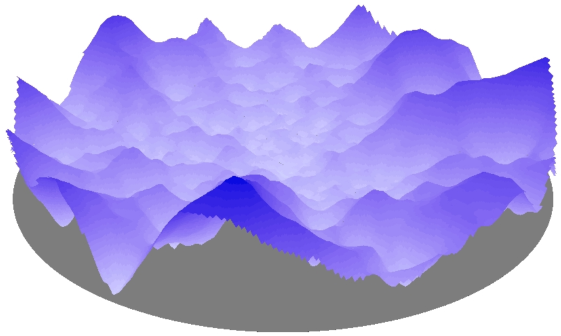


Figure 5.7: This combined distance to services for each grid-cell is shown for Atlanta in 3D.

Service	
Bank	Child care facility
Community center	Convenience store
Hair care	Hardware Store
Theater	Laundry/dry cleaner
Library	Medical/Dental office
Pharmacy	Place of worship
Police/Fire station	Post Office
Restaurant	School
Senior care facility	Supermarket
Health club/Rec. facility	

Table 5.7: Distance was measured at a 250m grid-cell to the following services. These values were then averaged per block group.

5.2.2 *Data*

Empirical data from the Metropolitan Authority Planning Council (MAPC) of Boston was used in this work. This dataset was composed of the total annual kilometers driven by passenger vehicles at a 250m grid-cell, using 2007 Registry of Motor Vehicles data. This mileage data was gathered during annual emissions testing which is done for each registered vehicle in the state of Massachusetts. These values were geocoded to a 250m grid-cell based on the owners' address, by the MAPC. The total kilometers travelled per grid-cell was aggregated to the block group for this analysis.

When the raw data is examined in Figure 5.11 (a) it can be observed that there is an upper bound for the distance that people travel based on population density using empirical data. I explore if this pattern of behavior can be explained using geometric measurements and population density values. It can be seen in Figure 5.11 (a) that although there is a general reduction in VKT as population density increases, there is also significant variation within the data, particularly at low population densities.

The objective of this model is to predict VKT primarily using spatial measurements. Figure 5.8 shows the relationship between the empirical VKT data per household aggregated to the block group level, and a range of geometric parameters calculated at the block-group level. Household density is also shown in Figure 5.8. These geometric measures consider several road network properties and housing density, but do not consider other demographic or socio-economic subtleties.

A model to predict VKT is identified using this collection of spatial measurements per block group, the *distance to services* measure, and a housing-unit measure from the census.[5] This relationship is described in Equation 5.2, and the parameters for the regression are shown in Table 5.8.

[5] *Housing Units* had a stronger correlation with VKT than *Population Density* and was used in this regression model

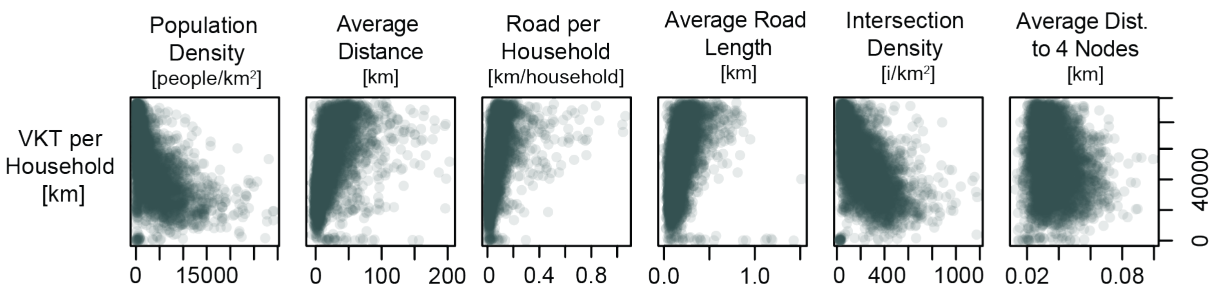


Figure 5.8: Parameters used in model.

$$
\begin{aligned}
\text{VKT per Household} &= \beta_0 + \beta_1 \text{ Pop. Dens} + \beta_2 \text{ Avg. Dist.} + \beta_3 \text{ Avg. Dist.}^2 \\
&+ \beta_4 \text{ Road per Household} + \beta_5 \text{ Avg. Road Length} \\
&+ \beta_4 \text{ Intersection Dens.} + \beta_5 \text{ Intersection Dens.}^2 \\
&+ \beta_4 \text{ Avg. Node}_4 \text{ Dist.} + \beta_5 \text{ Avg. Node}_4 \text{ Dist.}^2 \qquad (5.2)
\end{aligned}
$$

Table 5.8: Summary of regression relating service distance, population density and road length to VKT. $R^2 = 0.4306$, $N = 4507$. This model was calibrated using data outside the four urban areas shown in Figure 5.9.

	Estimate	Std. Error	t value	Pr(>t)
(Intercept)	3.542e+04	3.432e+03	10.321	0.000
Pop. Dens	-1.076	0.1119	-9.620	0.000
Avg. Dist.	527.5	38.84	13.581	0.000
Avg. Dist.^{2}	-3.445	0.238	-14.471	0.000
Road per Household	9551	1160	8.237	0.000
Avg. Road Length	18080	3186	5.675	0.000
Intersection Dens.	-62.31	4.958	-12.568	0.000
Intersection Dens.^{2}	0.052	0.0041	12.672	0.000
Avg. Node$_4$ Dist	6.44e+05	1.324e+05	4.862	0.000
Avg. Node$_4$ Dist.^{2}	-6.59e+06	1.438e+06	-4.583	0.000

The model was estimated using data from the state of Massachusetts, where empirical VKT data was available. Four urban areas (Amherst, Boston, Springfield and Worcester) were excluded so that they could be used to compare the accuracy of the model's prediction. Then the model was validated by comparing the predicted VKT values to empirical VKT values for these four urban areas in Massachusetts. The predicted values were estimated at the block group level and summed for each area. Figure 5.9 shows a comparison of the empirical and predicted values for cities in Massachusetts.

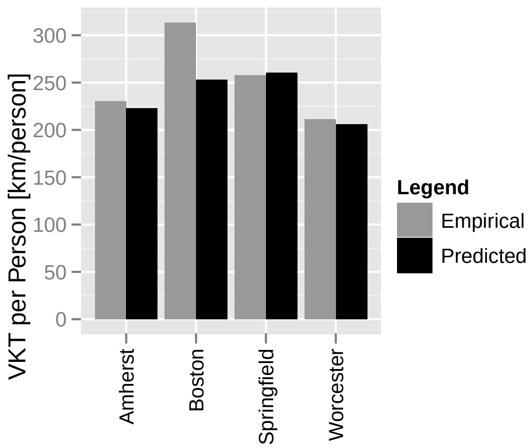


Figure 5.9: Estimated distances for four cities in MA. The error between the predicted and empirical values had a mean value of 12.78% and a standard deviation of 10.

Then, the regression model described in Eq. 5.2 was applied to all of the 40 US cities analyzed in this work. Again this calculation was performed at the block-group level. Figure 5.10 illustrates the predicted VKT values per block-group, for each city. Similar to the road per person relationship for each city, it can be seen that the overall VKT varies significantly depending on what part of the city is being considered. Figure 5.11 (a) illustrates the same general trend for empirical data from Massachusetts, alongside the predicted VKT values (Figure 5.11 (b)).

To compare the predicted VKT values for each city, the average VKT per person for each city is plotted with data from the Texas Transportation Congestion survey (Schrank and Lomax, 2011) in Figure 5.12. The data from Schrank and Lomax (2011) uses empirical data from Federal Highway Administration which records traffic on freeways and arterials, for many urban areas. This data is aggregated by Schrank and Lomax (2011) for the urban area. Figure 5.12 illustrates that the predicted values are of a similar order of magnitude to the empirical data; the *RMSE* error is 3942 km. When the *RMSE* value is normalized by the mean VKT per person for all cities (11012 km per person), this value is 0.36. Despite the earlier criticism of using mean values to compare urban areas, this is the only way that this empirical data can be compared to the predicted values, due to the unit of aggregation. Then using national values, the VKT per household was converted into kWh units of energy.

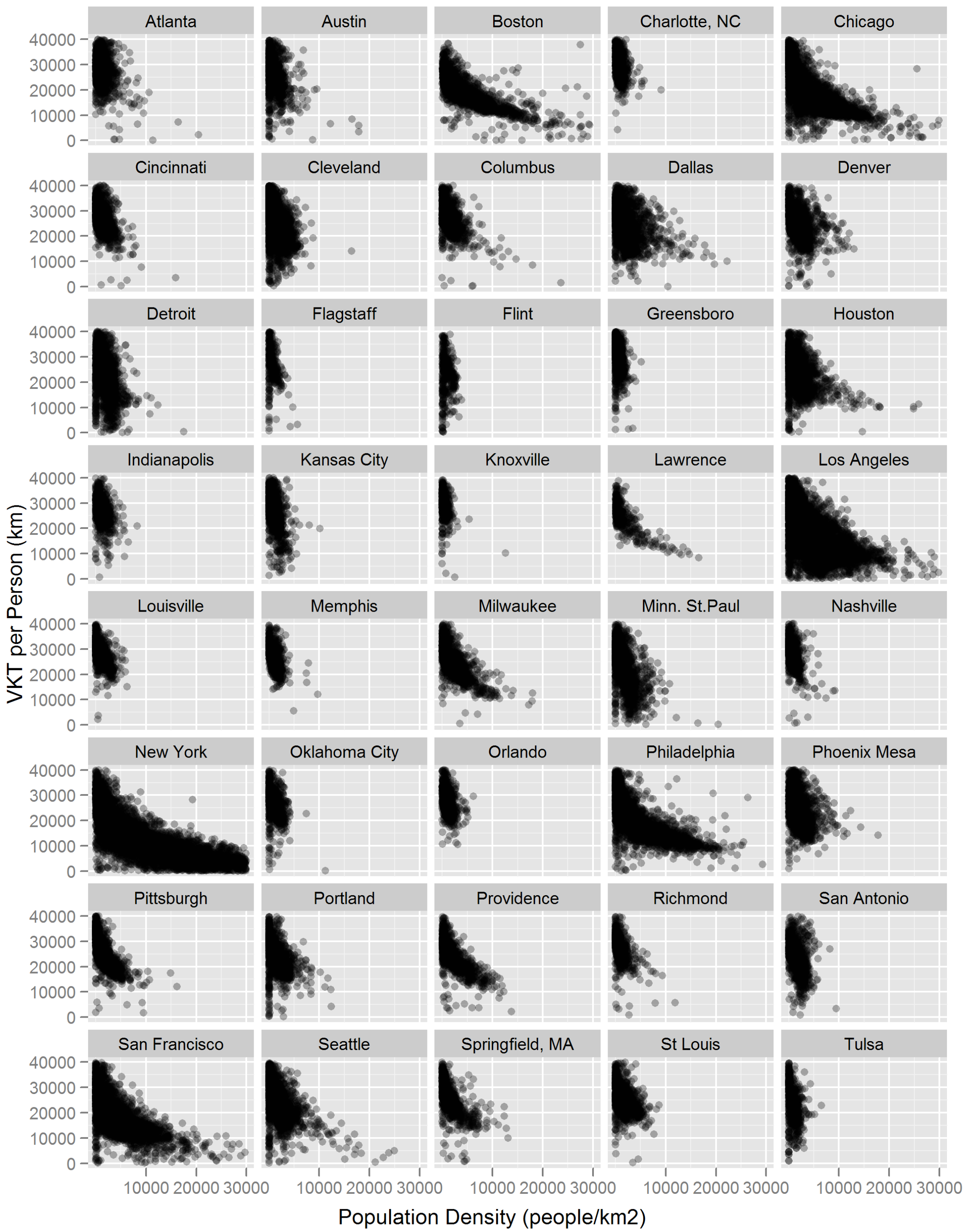
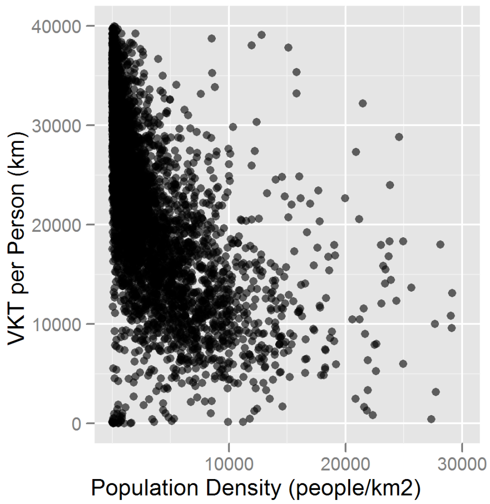
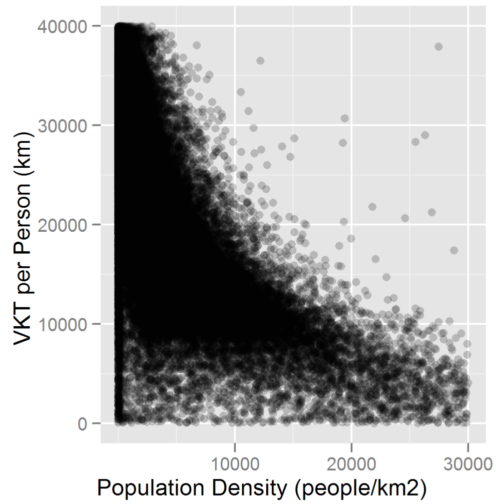


Figure 5.10: Estimated VKT per capita for varying population density levels.

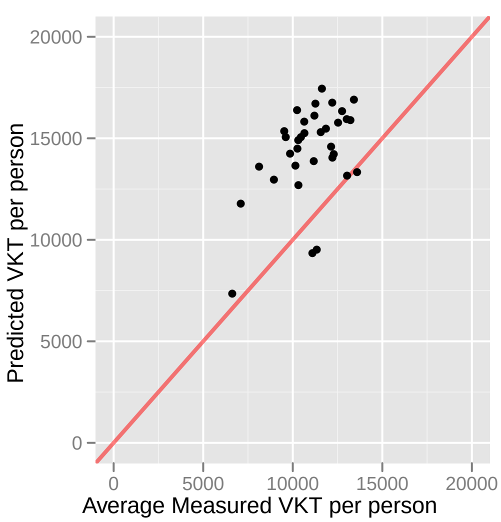


(a) MA empirical data

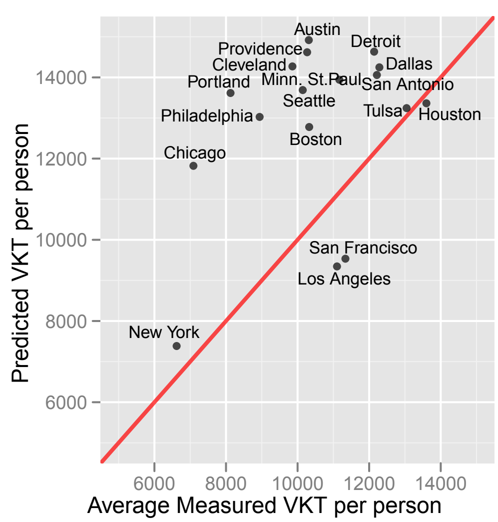


(b) Predicted data for 40 cities

Figure 5.11: Empirical and estimated values. The empirical data is from the state of MA, the estimated values are all values for the 40 cities.



(a) Predicted and Empirical VKT



(b) Predicted and Empirical VKT (labels)

Figure 5.12: Comparison of predicted values with city level empirical data

5.2.3 *Residential Building Energy Use*

The initial goal of this analysis was to use the geometric predictions of built area and volume to estimate energy consumption, when the average household size and climate were considered. However, due to the lack of available data this process was not undertaken. In addition, building stock variability and occupancy behavior would not have been captured.

Some exploratory analysis using zip-code level electricity data from New York was examined (nycopendata.socrata.com) and is illustrated in Figure 5.13 and 5.14, as well as data from (Howard et al., 2012).[6] Figure 5.20 illustrates the distinctive linear nature of electricity consumption, with respect to building area.

[6] Data from Howard et al. (2012) provided courtesy of Professor Stephen Hammer.

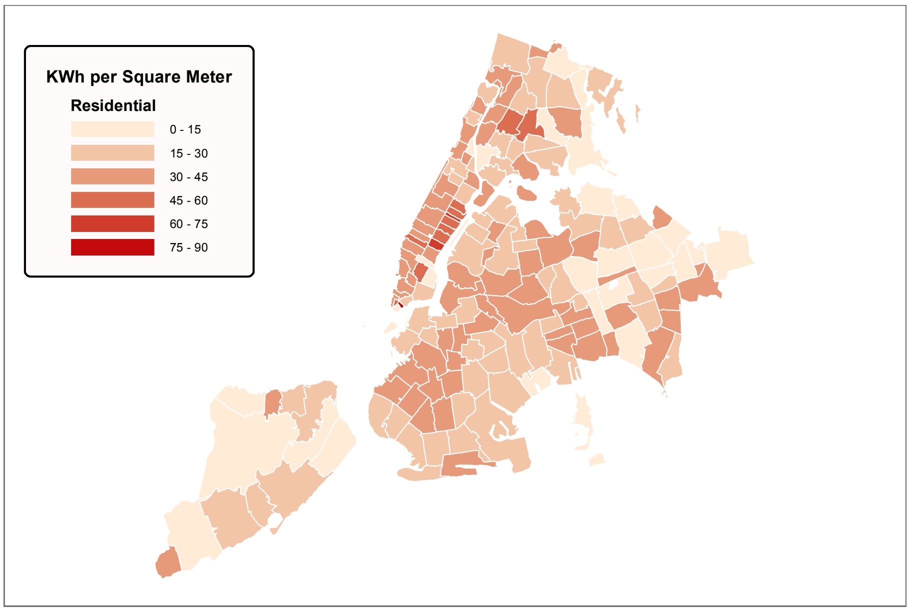


Figure 5.13: Estimated yearly New York residential electricity consumption per zip-code, normalized by the building area. Data from nycopendata.socrata.com.

Due to the lack of national energy data at a high resolution, data from a published and validated model from *Efficiency 2.0* (Min et al., 2010) is used in this research. This model uses data from the Residential Energy Consumption Survey (RECS) data to estimate average household electricity and gas consumption at the zip code, as well as socioeconomic measures, climate factors and fuel sources. This residential energy consumption data was spatially joined to the block-groups for all 40 cities and was then used in combination with the transportation energy calculations.

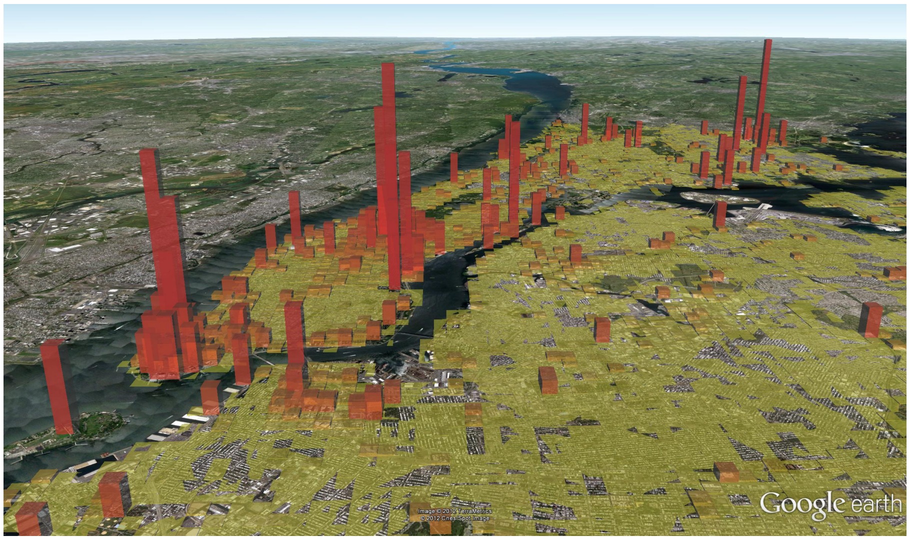


Figure 5.14: New York yearly electricity consumption per block-group, normalized by the building area. Data from (Howard et al., 2012)

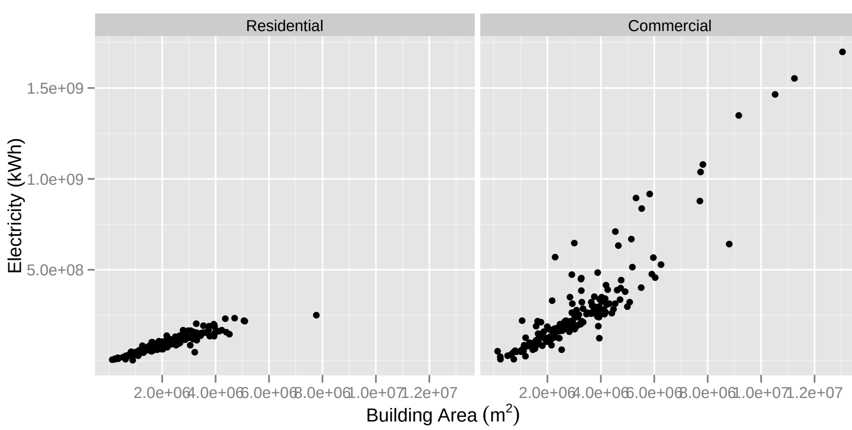


Figure 5.15: Yearly total of electricity consumption per zip code for New York. Data from nycopendata.socrata.com

5.3 *Neighborhood Typologies*

The work in this section comes directly from a conference paper titled 'Estimating Resource Consumption using Urban Typologies' presented at CISBAT 2011. The authors of this paper were David Quinn, Daniel Wiesmann and Juan José Sarralde. The work listed in this section is a collaboration between the three authors.

This section describes an approach developed to estimate the material intensity of neighborhoods, using representative urban typologies. Typologies are identified using geometric parameters that describe the physical environment. These typologies are then used to estimate the resource consumption of neighborhoods. The objective of this methodology is to identify relationships between urban form parameters and resource consumption, and to explore how varying these parameters influences the resource intensity and efficiency of neighborhoods. In this section, the focus is on measuring the material required to construct the infrastructure and buildings for each typology.

To estimate resource consumption, the material intensity of different urban forms was explored. In this analysis, measurements from the London building stock are analyzed. The analysis of material intensity considers construction materials used in buildings and urban infrastructure. By comparing the resource demands of the various urban typologies (shown in Figure 5.16), links between urban form and resource usage can be identified.

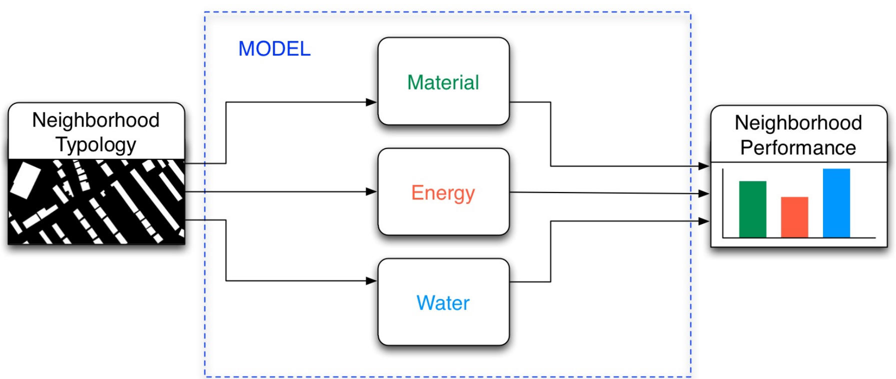


Figure 5.16: Objective of neighborhood typology analysis. In this section, the focus is on material usage.

In this analysis, vector data describing the 3D geometry of buildings and roads was available. Road vector road data (with information about the road type) was downloaded from UK Ordnance Survey (2010b). 3D building data from the The Geoinformation Group Ltd. (2010) was also used. When estimating the road area using a linear vector source, one frequent inaccuracy is due to the road width variation throughout the city. In this case, one of the available data sets (The Geoinformation Group Ltd., 2010) also had roadbed data available in a polygon format, which was gathered using remote sensing data.

5.3.1 *Methodology*

This analysis is structured in two parts. The first part describes the process of identifying typologies using the clustering of physical characteristics that describe urban form. The second part of this work describes how these physical characteristics are then converted into units of material. The spatial unit of analysis in this study was the *Lower Layer Super Output Area* (LLSOA), which is a spatial unit defined by the UK census bureau.[7] In this study approximately 5000 LLSOA units were used, which make up the greater London area. The average number of people, buildings and area per LLSOA used in this study, are summarized in Table 5.9.

[7] The LLSOA unit is an equivalent size to the block-group used in the US analysis.

Table 5.9: Average measures per LLSOA.

Measure	Unit	Count
Population	-	1500
Buildings	-	178
Area	$[km^2]$	0.49

5.3.2 *Typology Identification*

This section describes the methodology that was used to identify urban typologies. Physical parameters which describe the urban form are listed in Table 5.10. Then a statistical clustering technique was used to identify groups in the data. Clusters were identified from these physical parameters. These clusters were then used to estimate the amount of material required to construct each typology.

Table 5.10: Dimensions used for clustering calculations.

Category	Description
Plot Ratio	Total floor space / LLSOA area
Green Space Fraction	Total green space / LLSOA area
Built Area Fraction	Total built footprint / LLSOA area
Average Building Height	Average height of buildings in LLSOA

The k-means algorithm was used to identify clusters. This algorithm partitions data into k number of clusters (where k is chosen based on graphical observation) using n observations. In this case, three cluster groups were chosen and the statistical language R was used. The implementation of k-means was the default McQueen implementation which functions by iteratively partitioning the data until it reaches convergence.

5.3.3 *Typologies*

Using the physical characteristics described in Table 5.10 for clustering, resulted in the clusters shown in Figure 5.17. The results from the clustering

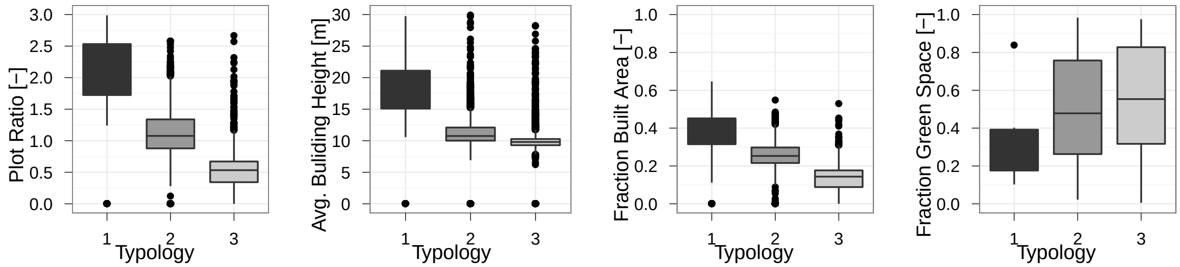


Figure 5.17: Typologies and dimensions of measurement of the urban form

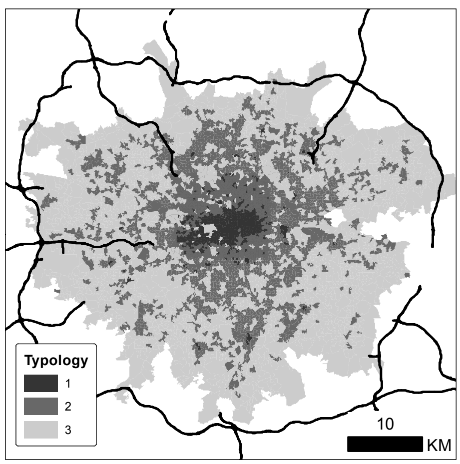


Figure 5.18: Map of London showing typologies. The range of each variable is shown in Figure 5.17. The black lines represent the motorway around London.

(a) Typology 1

(b) Typology 2

(c) Typology 3

Figure 5.19: Examples of typologies from the London Metro area

are also shown in a map of the greater London area (Figure 5.18). Representative examples of individual clusters are shown in 3D in Figure 5.19.

5.3.4 *Material Requirements per Cluster*

Using the geometric measurements of each cluster, and data from Tables 5.2 and 5.6, material was estimated for roads and buildings. Surface area is calculated as the total amount of wall, floor and roof area for each household. These material estimates are shown in Figures 5.20 and 5.21.

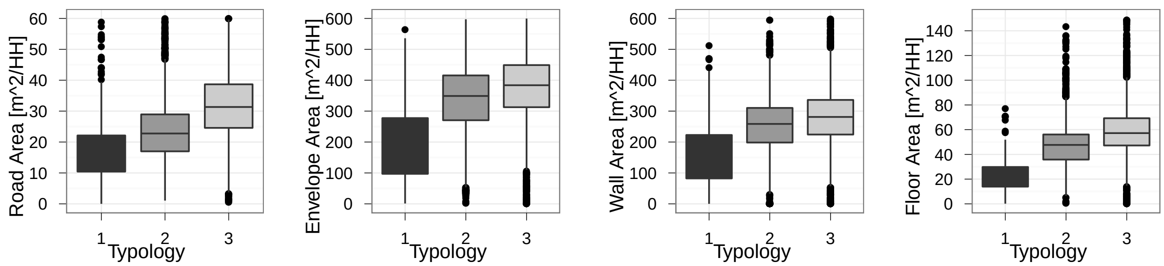


Figure 5.20: Local road area and residential building surface area.

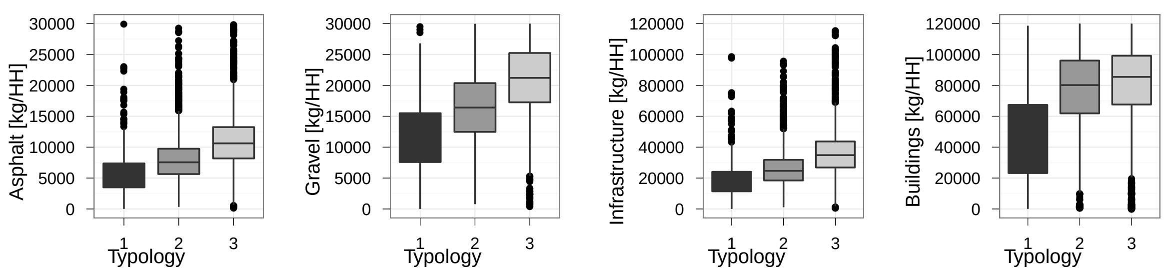


Figure 5.21: Infrastructure material normalized per household.

The scale of the spatial calculations here is quite aggregated as each LLSOA consists of several hundred buildings. Nevertheless it is still sufficient to identify clear patterns of behavior and to pick up on the density patters in Figure 5.18. Though the measurement of material for construction an objective was to develop an efficiency measure, where the resources required for a certain typology, could be attributed to the residents of that neighborhood. However, the fabric of the urban form is not typically homogeneous as there is usually a mixture of land-use types, where residential and commercial areas are mixed. Attributing the physical infrastructure overhead to residents requires measurements of just the built environment that contains people.

In the Section 5.4, both material and energy are considered at the same spatial resolution. A future work extension of this work could be an exploration of water patterns based on local typologies, however this work is not done in this thesis.

5.4 *Combined Material and Energy Patterns*

Using the average block-group measurements of material and energy (described in Section 5.1 and 5.2), we can explore material and energy intensities of cities in combination with population density. To illustrate how these dimensions of measurement can be combined, a schematic of the plot structure is shown in Figure 5.22. In Figure 5.22, each point represents the average household value per block group. *Material per Household* is the sum of the average material for road infrastructure and buildings per household; similarly *Energy per Household* is the sum of the average building energy and average transportation energy per household. The population density is illustrated using a color scheme of low (blue) to high (red). All of these measurements are per block-group.

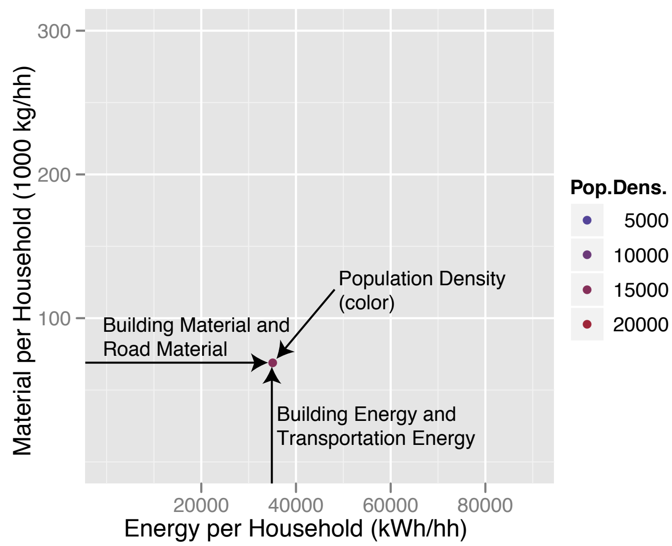


Figure 5.22: Explanation of plot structure.

This relationship is illustrated using data from two cities (New York and San Francisco) in Figure 5.23. It can be observed that both cities have similar trends, as the scatterplot shape is similar; however, the absolute values are different. For high population densities (greater than 20000 $people/km^2$, there are small number of block groups with an energy intensity less than 15000 kWh per household in San Franciso. In New York, the lowest energy consumption values start at 17000 kWh, these values then increase to a much higher level than San Francisco. This could be explained in part by the climate, but also by the fact that the high-density areas in San Francisco still have buildings with a low-height.

Similar to Figures 5.22 and 5.23, Figure 5.24 shows material and energy measures per household, colored by population density for all 40 cities. Considering the patterns for all 40 cities in Figure 5.24. As is expected, the higher density areas typically use less material and energy. This can be explained by

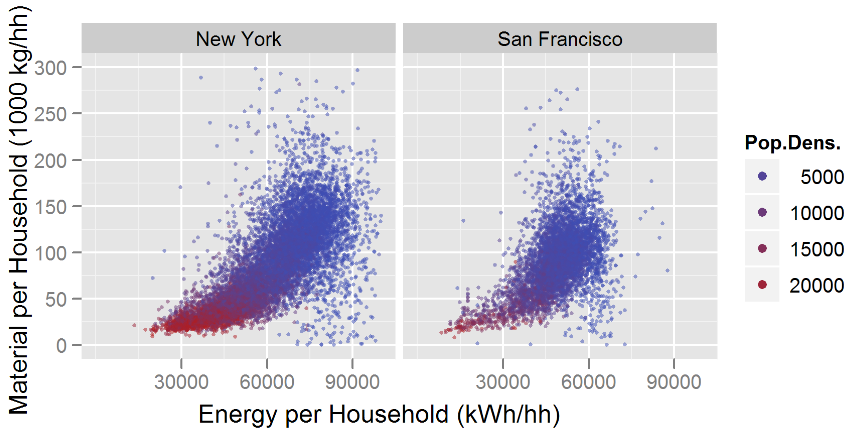


Figure 5.23: Combined material and energy measures per household for New York and San Francisco.

the fact that in high population density areas, the building area per person is smaller, and the infrastructure per person is lower. Similarly, less energy is (generally) used within smaller buildings, and less energy is used by private transportation. As the energy per household includes both building energy and vehicle transportation, as the population density drops, the energy share due to transportation becomes slightly more dominant.

Two general patterns can be observed for all 40 cities in Figure 5.24. Large cities, with greater population densities end up with a distinctive curved shape on the plot. On the other hand, the cluster of points for low-density cities is more nondescript; instead of a clear pattern, there is just a general grouping of points in an almost circular shape. This is particularly true for smaller cities with lower population densities. If the population density images from Section 3.4 are considered, we can appreciate that there is a negligible population gradient for some of these urban areas.

Interestingly, some large cities without higher-density levels (Atlanta, and Detroit, for example) do not appear to have any areas with compact development, and the overall average appears to be much higher when compared to other cities.

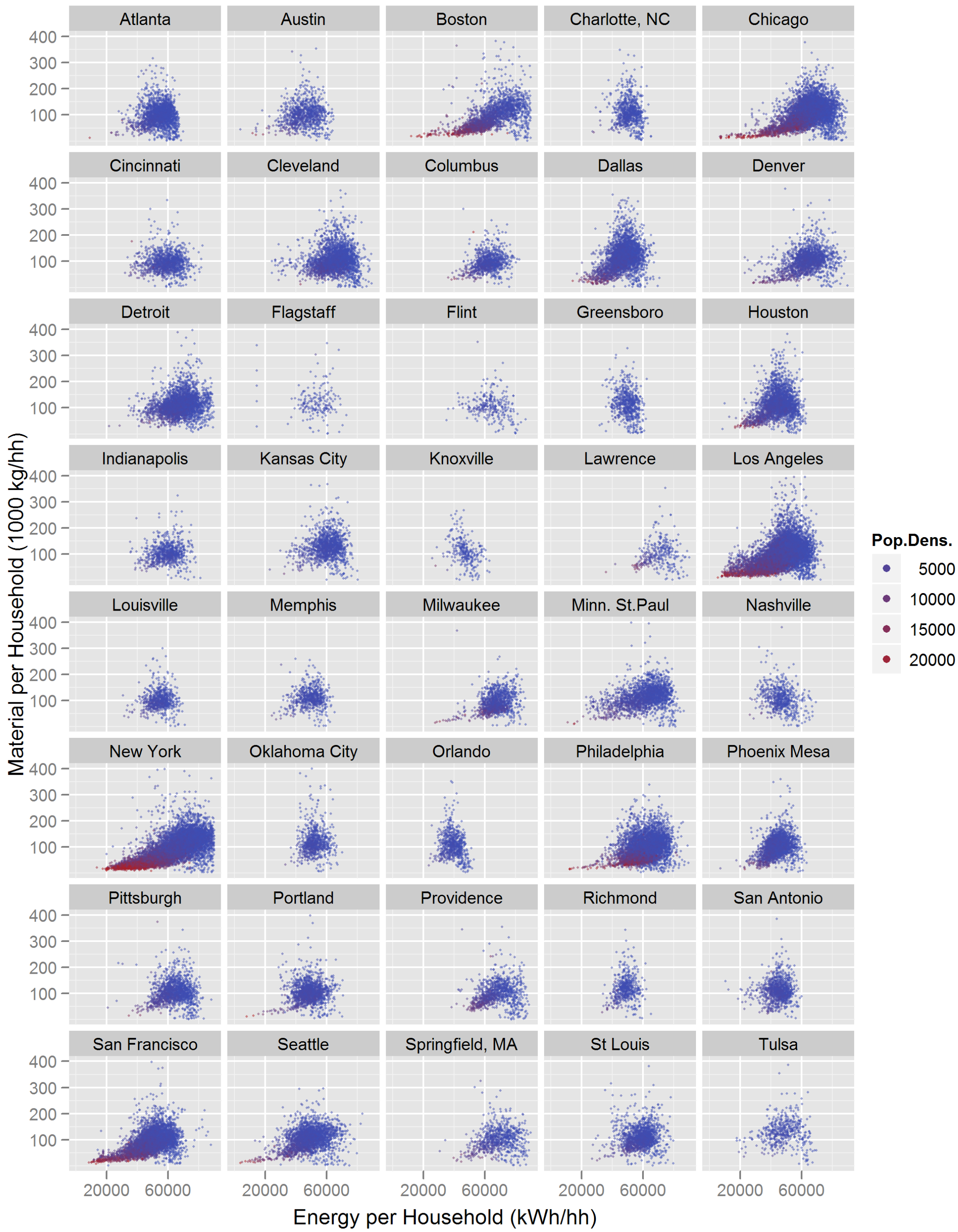


Figure 5.24: Combined material and energy measures per household for US cities.

5.5 Conclusions

In this chapter, I examined relationships between geometric measurements of the urban form, and material and energy intensities. Geometric measurements of the urban form were related to the material stock of roads and buildings, as well as the spatial variation of energy intensities. I then explored combinations of buildings and infrastructure measures for London to identify neighborhood typologies. Then block-group level measures of material and energy were combined to illustrate distinctive patterns for the 40 US cities examined.

The material stock due to roads and building is estimated in Section 5.1. These estimates use the predicted geometric values from Chapter 5, in combination with survey data and building construction standards. These material estimates focus only on the material required to construct roads and residential buildings. However, the yearly material flows due to new infrastructure construction, replacement, maintenance are likely to be substantial. Brand (2006) states that European cities replace 2 to 3% of their material fabric each year due to demolition and rebuilding of buildings, roads, and other construction. While this is likely to vary regionally due to many reasons (construction material, climate), considering the renewal of the building stock is one approach that would extend this initial assessment to predict material flows due to roads and buildings.

In Section 5.2, energy measurements per block-group are considered. This includes measures both within building energy use, and the energy used due to private auto-transportation.

In the literature review, I refer to the fact that few studies focus on holistic measurement of cities. In Figure 5.24, I illustrate the importance of considering both within-building measures (material for structure, energy use within structure) and the resources need to for that household to function depending on its location (material for infrastructure, energy for transportation). Here, I have tried to emphasize that it is important to consider both the material and energy requirements within the building, but to also consider the material and energy requirements associated with the functioning of the building in the form of the infrastructure needed to get there; and also the energy required to travel in that area.

Again, this work illustrates cross-city patterns, and identifies upper and lower boundaries of resource intensities that are specific to each city. The identification of these boundaries is important as it identifies baseline measurements of material and energy intensities, which can be used to define the maximum achievable efficiency in that particular city.

6 Discussion

In this chapter, I discuss the important themes and observations that have emerged from this work. In Section 6.1, I consider the issue of resource efficiency in cities, using the results from Chapters 4 and 5. I discuss the considerations associated with city level measurements, and the tradeoffs associated with compactness in urban form. In Section 6.2, I discuss how planners can incorporate this work and what policy strategies have been used previously.

This work focused on identifying relationships within the built environment, however the underlying causal mechanisms were not identified due to the complex nature of urban economics, geography and regulation. Nevertheless, while there is significant variation within different parts of each city, I have shown that there are common patterns across all cities. As these values are based on empirical (and estimated) measures, this work cannot be directly incorporated to the urban design process, as causal mechanisms are not identified; however, it is likely that new designs will fall within these upper and lower bounds of resource intensities.

In this work, I have used parsimonious statistical models that are not complex. The tradeoff of this approach is that the models identified are more inaccurate at prediction (although still statistically significant) when compared to more complex statistical approaches. However, due to the small number of variables required and the fact that multi-variate linear regression was used, this approach ensures that these models can be easily applied to other datasets. As I focused on the physical characteristics of the built environment, and disregarded any socio-economic measurements (other than counts of people and housing), these parsimonious models can be easily applied to datasets from other cities and countries. This approach is particularly relevant in data scarce environments.

It should be noted, that the data used to train these models came from cities in the US and UK which are both OECD countries. Further research is need to confirm that these models are appropriate when examining cities from other countries.

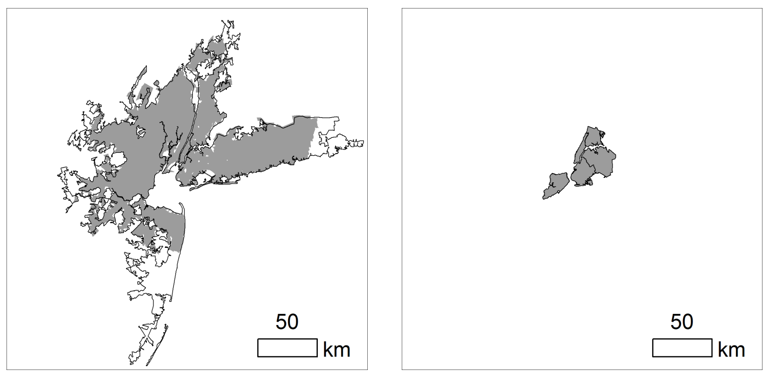
6.1 Resource Intensity and Urban Efficiency

I now discuss three aspects of resource intensity and urban efficiency. The first approach (Section 6.1.1) considers political boundaries and explore how the boundary definition influences the average values that represent the city. The second approach described in Section 6.1.2, considers some of the tradeoffs associated with compact urban form. Finally, in Section 6.1.3 I discuss transportation patterns related to the modeling approach used in this work.

6.1.1 Political Boundaries

When considering the resource intensity of a city, it is important to consider the overall urban system in a regional context, rather than just focusing on historical political boundaries. For example, New York is frequently cited as having a low energy use value per capita (EIA, 2012) when compared to other cities in the US. The importance of how this boundary is calculated, is illustrated when we consider the political boundary of New York (consisting of the five boroughs: Manhattan, Queens, Brooklyn, The Bronx and Staten Island) and the metropolitan area political boundary defined by the US Census (this boundary is defined using a contiguous population density threshold). These boundaries are shown in Figure 6.1.

Figure 6.1: New York metropolitan area and boroughs.



(a) Metropolitan Area (areas analyzed in this work are shaded gray, while the black outline is the metro area boundary.

(b) Boroughs

When the resource intensity measurements for the New York metropolitan area and the five NY boroughs are compared, we can observe a significant difference, illustrated in Figure 6.2. We can observe that a greater portion of the high resource intensity households are located in the lower density areas of the city (Figure 6.2). As the boroughs have a high population density, which results in smaller building volumes per person and less private transportation energy due to the ability to walk or take public transportation, this results in a lower mean value per household for energy. Similarly there is less material required for buildings, and the infrastructure is heavily used resulting in a low road infrastructure material measurement per capita. It is also worth considering that there is an empirical limit of 40% of the land area being covered by road infrastructure; this means that as the population reaches very high-levels, the road per person value constantly decreases. A possible explanation of this phenomenon is that one the population density is above a certain level, alternative modes of transportation become more prevalent.

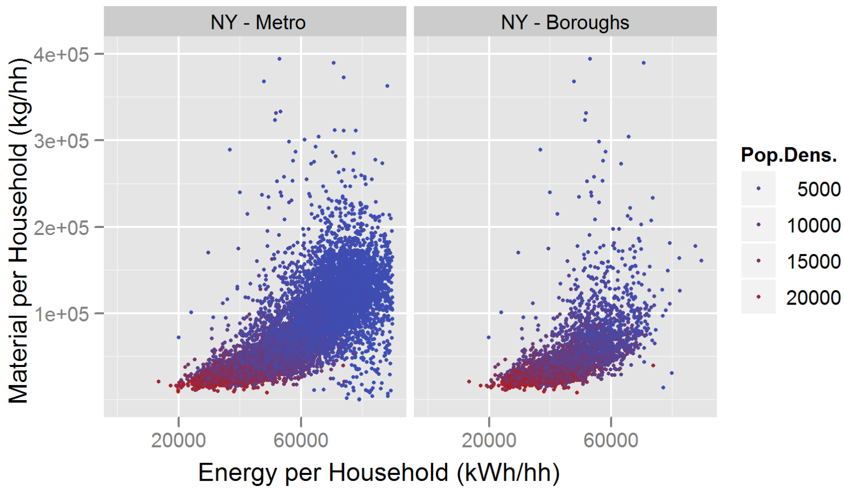


Figure 6.2: Material and energy intensities per household for New York, considering the metro area and the borough.

To illustrate these resource intensity measures in greater detail, histograms of each area for the average energy and material per household are shown in Figures 6.3 and 6.4. Figure 6.3 shows a histogram of the average energy consumption per household, for the metropolitan and borough areas. A similar pattern can be seen in Figure 6.4 where a histogram of the average building and road infrastructure considered. The average value per area is shown in Table 6.1.

Table 6.1: Average resource intensities per household in New York.

	Material kg [1000 kg]	Energy kWh [kWh]
Metro	82.55	58 187
Borough	46.67	43 280

To conclude this argument, including the overall urban region is important when considering average resource intensity measurements per city. This issue is frequently overlooked when researchers use city level analysis. To solve this problem, more rigorous criteria should be applied when defining the boundary of a city for analysis. For example, an alternative approach is to consider the overall distribution of an urban system and to define the boundary using some cutoff threshold. While the census metropolitan areas are defined using a population density threshold, this measurement does not result in established political districts changing their boundaries in response to this criteria which would be necessary to implement policies. To achieve effective sustainability policies, I believe that it is essential for cities to

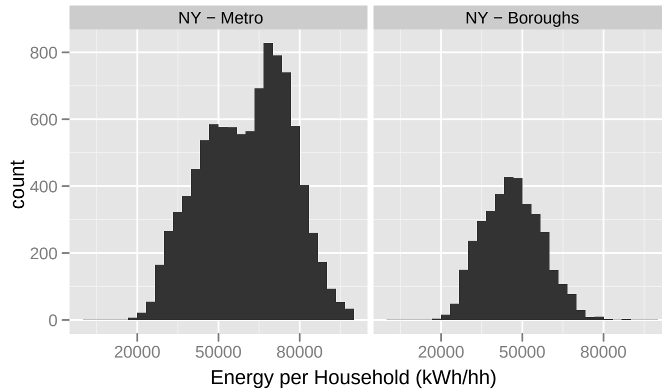


Figure 6.3: Histogram of energy intensities for New York.

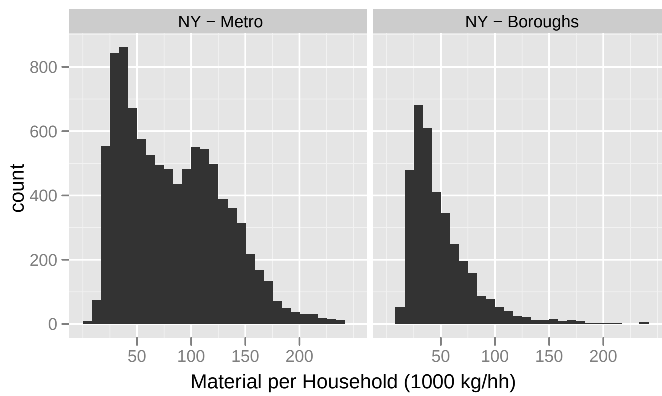


Figure 6.4: Histogram of material intensities for New York.

consider the functioning of the overall system, and to redefine their boundaries based on these types of criteria. Without considering the overall urban system, it is hard to assess whether policies are targeting areas where they would have the greatest impact.

6.1.2 *Compact Urban Form*

In this sub-section I discuss three main points. The first point considers how the resource requirements of infrastructure should be attributed, depending on the scale of measurement. The second point considers some of the tradeoffs associated with high-density urban populations and the third point considers the unit of service for transportation.

Attribution of Infrastructure: When considering infrastructure, the overall performance of the urban system is important, as discussed in the previous section. While this analysis suggests that higher population density ranges hold a greater potential for resource efficiency, there are some subtleties in this research that are not considered. In particular, this work focused on measuring discrete units of the urban system, and did not consider infrastructure from the city level perspective.

High population densities require infrastructure that may not be located directly where the population is located. This analysis does not consider external transportation infrastructure requirements, as it considers each area unit to be independent and discrete. For example, while a population density of 10,000 $people/km^2$ in Boston has an observed minimum requirement of 40 km of linear road per km^2, there are additional external costs to provide the infrastructure to serve this area that are not attributed to the individual area in these calculations. As some infrastructure facilities national transportation, with some for regional travel and some for local travel, a refinement of this calculation would be to justify the cost of shared infrastructure normalizing it according to the purpose that it serves. A suggested attribution scale is listed in Table 6.2.

Road Type	Attribution
Interstate	National
Expressway	City
Local Street	Neighborhood
Minor Road	Neighborhood

Table 6.2: Attribution of infrastructure cost.

However, as the dominant road-area per block group is due to local roads, this will not result in a significant change in the results, but it it will slightly increase the infrastructure intensity of areas with high population densities, and slightly decrease the intensity of lower density areas. A further extension of this work would be to consider the material requirements of non-road infrastructure that is more common in higher-density areas, such as rail, bridges and subway tunnels.

Tradeoffs of High-Density Development: Another perspective that Ewing and Rong (2008) highlight is the consequence of the interactions between higher-density developments when compared to urban sprawl. Ewing and Rong (2008) consider how higher density settlements require more air-conditioning due to the urban heat island effect, and examine whether this outweighs the savings resulting from high-density development. Ewing and Rong (2008) observe that compact urban development has a double benefit as people live in smaller dwellings typically reducing transportation energy use and emissions 'by 20 to 40 percent relative to sprawl' with a similar impact on residential energy use and emissions. Hence Ewing and Rong (2008) conclude that the tradeoffs associated with higher densities are worth it. Based on the results in this work (which does not consider any interactions between buildings), I agree with these conclusions.

Considering the material demand, many studies refer to cost savings for infrastructure at high densities (Ewing, 1997) and the associated tradeoffs. For areas of low-population density the use of septic tanks, open drainage and rural cross sections may cause the infrastructure per area to decrease while for areas of extremely high densities, the need for special high-rise structures and infrastructure causes it to increase Ewing (1997). The exact shape of the material demand function is difficult to estimate, as are the material requirements. The conclusions from this work show that the geometric area of roads and buildings per person are significantly lower at population densities greater than 5000 people/km^2. While more detailed calculations of how the material per component varied based on building type were not considered, the area per person reduction and the shared wall and floorspace are likely to compensate for the increased material demand to achieve a higher structural performance.

Infrastructure Efficiency: While roads and people compete for available space, identifying empirically where these limits are is an interesting result. In Figure 4.11, I illustrate that there is an upper limit for the fraction of the city taken up by road area. For material and cost to be considered per capita, a measure of the unit of service provided by the infrastructure needs to be defined so that an efficiency baseline can be calculated. Shin et al. (2009) examine this with regard to the functionality of the land area, and illustrate that when private auto is the dominant mode choice, buildings need to be taller so that they can provide more productive area, due to the space required for road area.

Considering the functioning of roads from the perspective of transportation efficiency is a useful extension of this work and would provide a more refined approach for examining the functioning of road-infrastructure. One approach could be to normalize the linear road length by VKT per person. In this way, a unit of service could be provided that takes into account the physical size of the road infrastructure and also how much it is used by local residents.

6.1.3 Transportation Patterns

Identifying upper and lower bounds of VKT, is an important factor when considering energy use in urban areas. This work shows that there are upper boundaries of VKT per household, when population density alone is considered (Figure 5.11). While the predictive capacity of such an approach is limited, due to the uncertainty of estimation at lower population densities, it can still be helpful as an approach to generally categorize the city.

In general, this approach does not consider the subtleties in behavior associated with demographic profiles or socio-economic patterns.[1] Including this data would enable the models used to be more precise for specific cases, at the cost of reducing the generalizability of the models, due to data scarcities. However, I argue that this work is still relevant, as there are physical limits to travel constraints. For example, it is unlikely that an individual will walk or cycle to visit a supermarket that is fifteen miles away; in this case, some mode of motorized transportation is probably used.

[1] Though some components of this are captured endogenously as households with higher income levels can afford larger homes.

To explore this hypothesis, data from the American Community Survey (ACS) that records the mode-choice of workers (U.S. Census Bureau, 2010) is plotted against the *Distance to Services* measure[2] in Figure 6.5. The x-axis shows the average cumulative distance to services per block group. On the y-axis is data from the ACS which gives the fraction of workers who walk or cycle to work. This illustrates the trend that a proximity to services, makes it more likely that commuters will walk or cycle and that there is a clear upper limit for walking or cycling. The ACS data only considers modes of transportation to and from work, but these mode choices are probably related to the mode choices for non-work trips.

[2] This calculation is described in Section 5.2.1. This distance assumes that an individual visits each of the services listed in Table A.2.1, once per week.

One interpretation, of this relationship[3] is that a reduction in the distance to a range of services would result in an increase in the fraction of people walking or cycling to work. This relationship is also non-linear with a very clear upper-boundary, which is an important consideration when considering the consequences of implementing a policy to reduce the distances to services, through land-use zoning or tax incentives. Decrease the *Distance to Services* from 20km to 15km has a much greater impact, than decreasing the distance from 40km to 20km. In fact, decreasing the distance to services from 40km to 20km appears to make no difference.

[3] Not considering the more nuanced causality arguments.

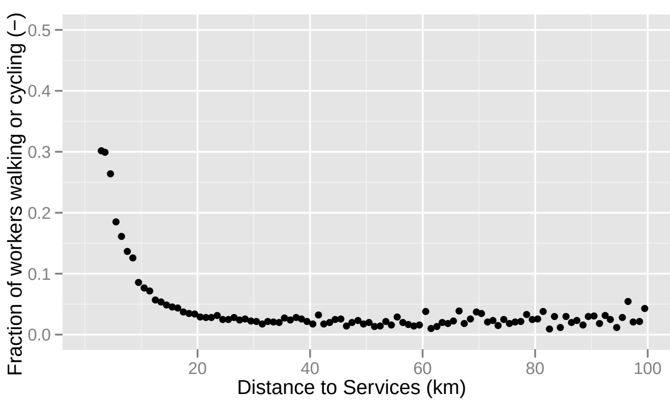
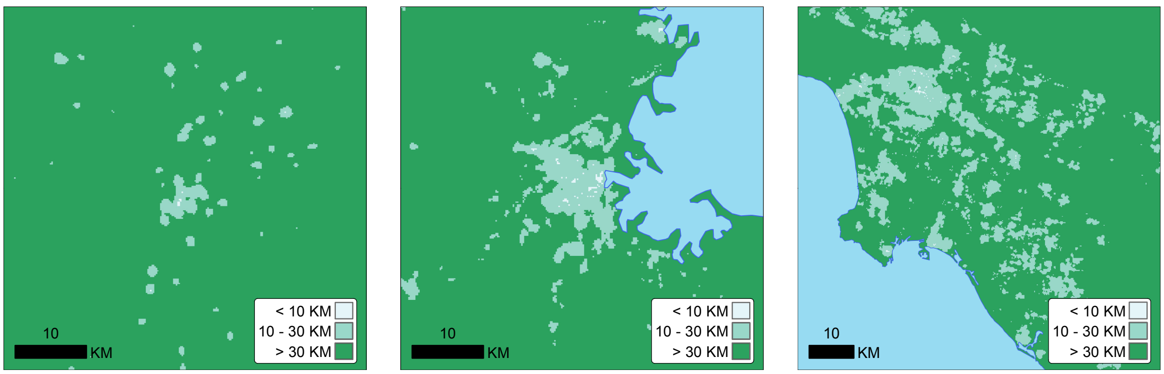


Figure 6.5: American Community Survey data illustrating the mode choice fraction of workers. The values shown here are the average per 1km group.

While there are many subtleties associated with mode choices, Figure 6.5 shows that there are upper limits which influence the feasibility of walking or cycling. The *Distance to Services* measure is shown in Figure 6.6 for three cities. This simplified estimate is one approach to assess what parts of the city it is possible to rely on walking or cycling to access these services. Traveling less than 10 km per week is considered a reasonable walking distance, $10 - 30$ km is considered reasonable for cycling, and more than 30 km per week is considered to require motorized transportation. These estimated values appear to fall within the empirical ranges observed in Figure 6.6. Where distances greater than 30 km per week need to be covered, providing public transportation would be challenging due to the low population density of these areas.

In Figure 6.7, the intersection density per block-group is plotted against the fraction of workers who walk or cycle, using data from the ACS. This is motivated by the review paper by Ewing and Cervero (2010), where the authors concluded that intersection density was important in influencing the mode-fraction of people who walked (where an increase in intersection density was correlated with a a higher walking or cycling mode-choice). Figure 6.7 agrees with this observation. Again a non-linear pattern can be observed, with a density of 300 $intersections/km^2$ appearing to be a minimum threshold value; above this intersection density the mode-fraction of workers who walk or cycle increases. While this transportation analysis uses many

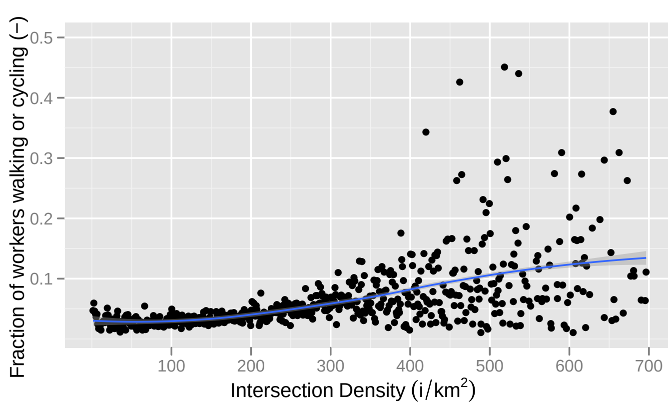


(a) Atlanta (b) Boston (c) Los Angeles

Figure 6.6: Distances within 30 *km* considering weekly frequency of service use.

generalizations, the overall goal is to relate geometric measurements to limits of feasible non-motorized mode choices. While more sophisticated models to predict travel behavior and mode choice exist, this work is primarily focused on exploring general trends that can be used to preliminary assess urban areas. I have illustrated that two geometric measures in this work can be related to empirical patterns of mode choice using a *Distance to Services* measure and a *Road Intersection Density* measure.

Figure 6.7: Intersection density plotted against non-motorized mode-choice from the ACS. The values shown here are the average values per intersection group.



6.2 Application to Planning

I consider this work to be relevant to three specific planning issues. Firstly, this work enables macro-scale patterns of cities to be compared to each other, and illustrates the variation within each city. The second application for this work, discusses policy strategies and resource intensities targets. Thirdly, by providing numeric values describing resource intensities, I have provided a mechanism for planners to consider digital precedents associated with a design from a quantitative perspective and to illustrate what the resource intensities of a design choice would be.

6.2.1 City Scale Comparisons

In this work, I have shown that while there is a significant variation of resource intensities within different parts of each city, there are common patterns across all cities. This has been achieved by analyzing a large number of urban areas using a large range of city samples. While this type of observation has been made about population density gradients for some time (Clark, 1951), I have refined this calculation by removing the need to identify a CBD. Other cross-city spatial patterns have not been explored thoroughly in the published literature. A useful contribution from this work is the observation of local geometric relationships, as well as the relationship between these geometric patterns, with material and energy measures, that can be used to estimate upper and lower bounds of resource intensities.

The actual reasons for resource intensity variation within cities are complex. The argument of population density resulting in increased efficiency is illustrated clearly in Figure 5.22, which supports the observations of Bettencourt et al. (2007) with regard to the economies of scale that can be observed in large cities. Figure 5.22 illustrates why some of this efficiency occurs, as smaller buildings, and less road infrastructure per person reduce the material demand, while smaller buildings require less energy, and people are closer together requiring less energy for transportation. However, these efficiencies also have their limits due to physical constraints. This is illustrated by the fact that the road-area does not go above 40% of the total area in the city center (Figure 4.11). This suggests that once the road reaches capacity, it is necessary for alternative modes of transportation to be provided. In larger, developed cities subways and trains provide additional transportation capacity, but in rapidly urbanizing cities with very high densities and few mass-transit options, extreme congestion can occur. Hence, it is important to balance increasing population densities with longer term infrastructure strategies that can result in the efficient movement of people.

Whether cities have a economies of scale by virtue of being large, or just by having achieved a high population densities due to regulation or emergent urban economic patterns, is an open question. However, what this work highlights is that the variation of these measurements is typically continuous, per city. For policy makers, this emphasizes that cities should not be considered as homogeneous entities. The spatial variation is an important aspect to be considered when the functioning of the overall system is being evaluated. In addition, it is also important to consider how the city influences the zone outside its immediate area. Newman and Kenworthy (1999) discuss how cities should fit within their ecological watershed; a sustainable city needs to consider issues of future growth (or decay) and whether the boundaries of the cities should be fixed or flexible to enable adaptation.

Several caveats should be noted about the methods used in this research. Chief among these is the issue of discounting and obscuring local variations due to the nature of the data employed for this analysis. This is problematic for addressing local variation within particular cities and among cities of contrasting climatic, topographic, governmental and socioeconomic attributes. Another variation not considered, are the local regulations and standards that may determine the contrasting densities and configurations of infrastructure between cities, and even within a particular city. Hence, it is important to consider this research as a static snapshot of the urban form that does not address the more complex question of urban growth formation.

6.2.2 Policy Strategies for Sustainable Urban Forms

There is a significant shortage of information available to researchers that describes resource use in the built environment, and there are few baseline metrics about the minimum energy or material required for neighborhoods to function. If we compare this to the state of building analysis twenty or thirty years ago we can see a similar trend, where the first phase of sustainable building design was component optimization (improved wall insulation material and window construction), the next phase was focused on the overall system and considered how the components interacted and the final phase focused on the functioning of the overall system. Eventually, performance metrics which consider the energy use per unit of area to achieve a certain comfort level were identified.[4] In this way, the performance of a building can be benchmarked against what the optimal measurement range is. To achieve the development of truly sustainable neighborhoods we need to consider a similar metric-driven approach (although a broader range of measurements needs to be considered than are described in this research).

[4] For example, the *passivehaus* standard is a well known building performance metric which requires that buildings do not use more than 15 kWh/m^2 per year.

The approach of using an absolute metric, similar to the *passivehaus* standard, is a challenging and optimistic goal, as it does not consider climatic variations or market forces. However, through the identification of the local resource intensity distribution, as well as upper and lower bounds of material and energy intensities, planners and policy makers can gain a better understanding of feasible targets for sustainability policies within their geographical region. When the range of resource intensities across a city is considered (Figure 6.2), it can be observed that that there are specific areas of a city that should be targeted through planning regulation.

In the domain of urban sustainability, I have observed two general strategies. The first is a high-level strategy with general numeric targets about reduction.[5] The second approach is at a more local scale and in the US is driven by third party assessment criteria. I will discuss these third party metric approaches and illustrate them using *LEED for Neighborhood Development* (LEED-ND) (USGBC, 2009) and *WalkScore* (WalkScore, 2009) as examples. LEED-ND is a rating system that quantifies certain urban form metrics at the neighborhood scale. LEED encourages compact, mixed-use developments and uses a checklist to assess new developments. *WalkScore* is a walkability ranking system that considers proximity to services, sidewalk quality, and a range of other criteria that are considered important for walking. Both LEED-ND and *WalkScore* are popular metrics as they provide a useful way of assessing neighborhoods that the individual can relate to. Both strategies encourage the location of new developments near existing services, and encourage strategies that reduce auto-travel.

[5] A common example is the following top-down policy statement: *City X plans to reduce CO_2 by 20% by 2050,*

As the influence of LEED-ND is at the neighborhood-level it is challenging to imagine how it would result in city-wide change. Nevertheless, it is an encouraging example of how the market has adopted several measures that have significant social and environmental benefits, and incorporated them into the pricing structure. However, LEED-ND is essentially just a softer version of a resource intensity metric driven approach. An optimistic view of urban sustainability is that the market would drive such behavior. A study by Cortright (2009) shows that WalkScore is positively correlated with property value illustrating that walkability is valued by the market. Another consideration, is that LEED-ND can be a catalyst for nodes of high density development. Realistically, due to the path-dependence associated with the road-network structure, it is unlikely that this market force can significantly change the existing infrastructure. Hence, this approach is likely to have greatest impact on new developments.

Effective sustainability strategies for cities are likely a combination of high-level strategy, strategic zoning and the adoption of more third-party sustainability metrics (as these metrics are useful reflections of aspects of the property market). Firstly, I believe that cities should analyze the performance of the overall urban area, and identify the worst performing areas. Then they should try to change these areas using the policy instruments of taxation and zoning. The goal of these policies would be to first achieve some minimum population density requirement and a certain mix of land-uses. The next approach would be to encourage the market to become involved, so that metrics like *LEED-ND* and *WalkScore* start becoming incorporated in the assessment of these areas. Although this suggested approach does focus on spatially discrete areas without considering the overall regional performance, I believe this to be a necessary first step for intervention.

6.2.3 Digital Precedents at the Neighborhood Scale

The ability to interactively explore urban areas enables planners to consider resource intensity measures when assessing an existing neighborhood, or a proposed urban plan. For example, a planner can explore an area that he or she is familiar, and examine material and energy intensities. The planner can then consider what factors result in these material and energy profiles.

The ability for planners to incorporate resource intensity measures dynamically as they consider design precedents, is an approach that I hope facilitates more powerful arguments for resource efficiency. For example, when a planner presents a proposed urban design, they can also present an estimate of the resource intensities based on comparable neighborhood types.[6] They can also explore and discuss the criteria that are important which influence this process. This ability to compare the functioning of neighborhoods and to look synoptically at other neighborhoods and cities is a significant contribution as it places energy and material intensities on an equal footing with other numeric measures, by providing numeric criteria that can be included in the discussion.

[6] They can also reference the source of this data, and interested parties can explore the data and analysis process further.

Once planners consider material and energy measurements to be relevant, they can access a GIS layer of energy or material measurements using a format that they are familiar with, from *urbmet.org*.[7] Even if a planner is unfamiliar with energy measurements but is familiar with the digital mapmaking process, they can incorporate this information into their workflow. Hence, if a planner decides that energy or material are important criteria and wishes to use data of this type in their work, the barrier to action is low as the data is freely available and easily accessible using standard tools that they are familiar with. The data is provided in such a format that it can be easily introduced to the technical planning process. In addition, this approach enables planners to illustrate the non-linear relationships between planning-type criteria that are controlled through zoning (such as population-density or floor-area ratios) and to related these measurements to the corresponding material and energy demands of the result patterns of urban form.

[7] Accesing the data using the *WMS* or *WCS* mapping feature.

In the future, I hope that the use and incorporation of data from this research into the planning process, will encourage greater data availability. More widely accessible data can then be used to improve the accuracy of resource intensity measurements and estimates. If data from *urbmet.org* is used in highly visible fora, more policy-makers and citizens can gain a better understanding of why these measurements are important and necessary, illustrating how this type of data can inform the planning process. Ideally this will lead to policy makers and citizens gaining a better understanding of resource intensity measurements.

6.3 Conclusions

Analyzing the resource consumption of the built environment is necessary in cities, both those that are mature and those that are rapidly growing, as short-term planning decisions have long-term consequences for both the quality of life of the inhabitants, and the future energy and material use of the urban area. When considering the assessment of an overall urban area, the political boundary is an important consideration. There are several tradeoffs associated with high-density urban development (urban heat island effect) for example, but this negative does not outweigh the positives. While the relationship between VKT and population density is complex, there are geometric constraints that influence the minimum distance that households need to travel to access services, and these distances influence the overall VKT of both the individual and the city. As there have been few cross-city high-resolution analyses of cities it is hoped that this work contributes to quantitative arguments about the benefit of compact urban form.

Second, this research has been completed from the perspective of a comparative analysis between these 40 cities. Arriving at conclusions regarding the nature of resource consumption is a primary goal here. Identifying the presence of floors of minimum and the edges of maximum consumption in a variety of population and service density conditions is an important result that can be used to understand the nature of urban resource consumption. One element of this understanding is progress toward a more nuanced understanding of the efficacy of compactness for urban resource efficiency. In addition, using the work here to expand beyond the US will be one of the next steps for this work. More cities around the world need to be examined to see if the parameters that describe US cities are similar for regions that are rapidly urbanizing. If urban areas behave globally in similar ways as found here, this would be a useful conclusion. Incorporating fuel and maintenance costs for suburban transportation and infrastructure provision are necessary considerations and will help to make these measures more significant.

I argue that cities are merely combinations of ranges of resource intensities from a continuous range. From this perspective, the concept of a sustainable city becomes more vague and the focus of policy measures should be targeted at the worst-performing areas of a city, assuming that the problem is local to that spatial area. However, this perspective considers the city purely from a static perspective, and does not consider the mechanisms that result in the formation of these urban forms. The rules and systems that result in this urban formation are of critical importance in new cities, but in existing cities where the structure is already defined and difficult to change it is useful to think about it from the more abstract perspective of resource intensity. It is hoped that this work will be of assistance for cities that create numerical targets for sustainability plans.

This work explores what empirical limits can be identified using some general assumptions, so that cities can be compared against each other. The benefit of this work, developing a standard procedure for analyzing a city, is also a limitation, due to the general nature of the analysis. This work does not consider socio-economic variation, or human behavior. It also does not consider more detailed urban form measures, or the type of businesses that may exist in various areas; for example big-box stores, compared to local corner stores would influence the proximity to services.

While *LEED ND* and similar approaches that emphasize higher density developments and mixed use indirectly address this, a more energy driven zoning policy would be beneficial to accelerate the process of resource intensity reduction. The concept of energy driven zoning to reduce peak loads and to encourage metric driven planning policies is somethings that sustainable neighborhoods should move towards.

The availability of planning precedents is important. Architects internalize and apply these to design, yet policy makers frequently have to deal with imperfect data, and few comparative examples. Enabling users to compare areas of a city, and to examine how these areas fit within an overall distribution is a useful means of comparison. I hope that this research can contribute to the discussion on resource efficient urban form, and influence what data is made available by policy makers and individuals.

Clearly, both the overall size of the city, and the population density are important aspects of resource efficiency. Banister (2008) states that 'empirical research has concluded that the key parameters of the sustainable city are that it should be over 25,000 population (preferably over 50,000), with medium densities (over 40 persons per hectare)'. This chapter illustrates the benefit of this minimum population density. In addition, ensuring that this population density is achieved will ensure that the distance to services will be within a reasonable range for a large fraction of the population to walk or cycle.

If a similarly strong correlation can be identified between neighborhoods and LEED-ND certification, it is possible that the market will move towards an urban form that results in a better quality of life.

In addition, this approach does not take into account urban economic factors that influences business location choice, but these economic factors can be shaped through planing regulation, in particular mixed-use zoning, and minimum population density criteria.

7 Web-based Spatial Analysis

In this chapter, I discuss why there is a need for an online visualization and analysis tool to enable the exploration of resource consumption measures. I argue that there is a shortage of tools available that can inform both the general public and specialists about the resource intensity of distinct urban forms. To achieve this, a visualization tool was developed that can display large amounts of spatial data using a web-browser.[1] At the time of writing, over 3800 unique users[2] have visited this web-tool with each user examining an average of four urban areas. In addition, several newspapers and cities have contacted us about this work.

This visualization work has been done in close collaboration with Daniel Wiesmann. The initial design and implementation was developed by David Quinn (Figures 7.2 and 7.3), all of the subsequent conceptual development and technological implementation was done collaboratively. This includes the interface design, the code to implement the design, as well as server-side code, server-maintenance and database administration.

[1] This tool is referred to as the *Neighborhood Visualizer*

[2] There have been over 8000 unique visitors to the site, but only 3800 unique users queried areas. Some of the other site-visitors are non-human, such as web-index spiders, etc.

First I discuss the reasons for developing a web-tool to facilitate the exploration of material and energy intensities for a range of urban areas in Section 7.1. In Section 7.1, I propose how a tool like this can be applied to real-world planning situations using three scenarios. The conceptual development of this tool is discussed in Section 7.2. In Section 7.3, I discuss the technical details associated with displaying spatial data using a web-browser. In Section 7.4 I describe the approach which enables users to generate dynamic reports of the urban area they examined. In Section 7.5, I discuss how this analysis can be accessed through desktop GIS software using a web-mapping service protocol. In Section 7.6, I describe a survey I developed to gather feedback on this tool to assess learning, identify functions which could be improved, as well as monitoring user behavior. Finally, in Section 7.7, I discuss the technological issues associated with this work, considering analysis methods, data availability and strategies for managing urban information in the long term.

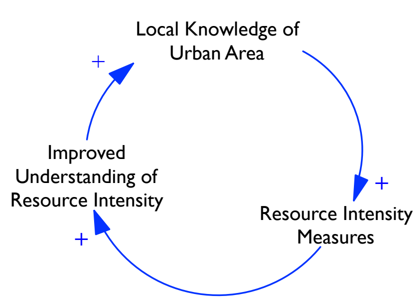


Figure 7.1: Accelerating and democratizing the learning cycle with regard to resource intensity.

The objective of this tool was to create a scaleable method of representing urban data in a way that enabled users to learn about empirical patterns of urban development. While this does not enable the user to envision how a particular area can be changed, it does illustrate what the typical resource intensity is, for a range of urban forms. By visualizing this data for 40 US cities, the user may be able to identify a neighborhood that they are familiar with,[3] so that a quantitative measurement of material and energy use can be related to the user's personal knowledge of that neighborhood. The intention of relating this local knowledge with quantitative resource intensity metrics is the ability to accelerate the user's learning about their neighborhood (Figure 7.1), without the need for a particular skill-set or software tools. This assessment tool can facilitate learning through comparative analysis of neighborhoods or cities, based on this sample of cities and urbanization patterns. In addition, the dynamic reporting feature facilitates the comparison of a particular neighborhood to other neighborhoods in that city by providing a *pdf* that users can save.

[3] Assuming that the user has lived or visited any of the cities used in this study.

7.1 Tools to Assist with Pathways for Sustainability

In this section, I propose how an interactive web-tool can be used in several fictional scenarios to assist with the planning process. I consider the use of this web-tool from the perspective of a variety of users: (1) Concerned Citizens, (2) a Planning Department that is considering zoning guidelines for a greenfield development and (3) a Planning Department that is considering how to change the functioning of a low-density residential area. In each scenario, I discuss how the ability to perform basic analysis and generate reports enables the users to gain a better understanding of resource intensity measures. As a result, criteria which consider energy and material are considered to be equivalent to other planning criteria. After describing these scenarios in detail, I discuss what features a web-tool of this type should have.

7.1.1 Concerned Citizens - Scenario 1

In Scenario 1, consider a local community meeting where a new development is being proposed by a developer. The developer wants to build a large low-density development in the suburbs of the city. The city is currently in favor of this development, as they will benefit in the short-term from an increased tax-base and subsidized infrastructure (the developer has agreed to pay for some of the costs associated with new road construction). Due to the scale of the development, the developer is required to hold public consultations as part of the planning process. In these public consultations the developer presents the plan for this development and answers questions. Typically, the developer has a deeper range of technical knowledge than the audience.

A local group of citizens disagree with the proposed development. Firstly, they argue that such a low-density development will lead to increased private automobile travel and will not result in any local amenities being developed due to the single-use nature of the plan. Secondly they argue that the overall cost to the city is not being considered properly as the infrastructure per person is significantly higher than elsewhere in the city. Finally, they argue that such a development promotes excessive energy consumption due to large single-family homes, and due to the energy required for transportation by residents in this area. The citizens argue that it is unlikely that individuals can walk or cycle to services with this urban form pattern, as services are unlikely to be located within a reasonable distance, due to the population density.

How citizens can use this tool

To substantiate their arguments, the citizens use a web-based GIS tool to analyze parts of the city that have both high-density mixed-use development, and low-density single-use development. The citizens generate several reports using this web-tool and print out several *pdf* reports in color to distribute at the meeting.

Through these reports, they illustrate the infrastructure required for this development and demonstrate that the infrastructure per person requirement is 3 to 4 times higher, when compared to other parts of the city. Over the course of this discussion, they also emphasize that when road-maintenance is considered, the yearly cost of maintaining this infrastructure is significantly higher than a more compact alternative. Due to the unequal distribution of infrastructure, the citizens argue that this new development is being subsided by residents in other high-density parts of the city.

The developer is surprised that the citizens are so well-equipped with numeric arguments about the resource intensity measurements presented, and argues that market demands this type of development, and that the economic activity due to this development will outweigh any future maintenance costs. The planning department considers the arguments from both sides and reflects on the data presented.

7.1.2 Zoning Requirements for a Greenfield Development - Scenario 2

In this scenario, the planning board of a city are considering what the zoning requirements of new greenfield developments should be. To frame the discussion, the planning board holds a meeting and use a web-based GIS tool and a projector to explore various parts of their city and discuss the ranges of population, material and energy consumption. They are interested in identifying numeric targets for new developments with regard to the amount of material and energy needed for a neighborhood to function.

While they recognize that there are many intangibles associated with the design process (and their zoning procedure will not disregard other non-numeric considerations), they are interested in an approach that will enable a quantitative assessment of proposed designs, so that they can better estimate material and energy intensities of future households using these guidelines.

How a Planning Board can use this tool

From this exploration, the Planning Board identifies some key metrics that they are considering for future developments. In addition to the current guidelines which specify population densities and floor/area ratios, they are considering the inclusion of the following criteria:

1. a maximum linear road length per person (10 $m/person$)

2. a maximum distance to a range of services (< 30 km)

3. a minimum intersection density (> 300 $intersections/km^2$)

Ensuring that new developments satisfy these built environment criteria is considered to a be a worthwhile objective by the Planning Board, as these requirements will reduce the resource intensity per household when compared to average city values.

7.1.3 Modifying an Existing Neighborhood - Scenario 3

In this scenario, a city agency is interested in creating a more sustainable neighborhood with a low energy and material intensity per household. The city is already interested in the success of a particular neighborhood, in part due to the low VKT and general satisfaction of residents in the neighborhood. External assessments using show that walking and cycling are common mode choices. This is correlated with *ACS* survey data and the area has a high ranking on the *Walkscore* index. The city is trying to decide the best strategy to improve a low-density neighborhood which has a dominant mode choice of autos. Despite the area having a low-population density it is the same distance from the central-business district as the more walkable neighborhood.

The city realizes that the road area covers a significant percentage of the overall area and that the street grid has a very low number of intersections per area. The consequences of path-dependencies as a result of the existing infrastructure makes it difficult to change the existing urban development patterns. However, through incrementally changing the infrastructure as part of an overall master-plan, the resource efficiency of a neighborhood can gradually increase.

How a Planing Board can use this tool

An example of how a metric driven approach is being directly applied, can be seen in Washington DC (Gelinne, 2011) where the city is specifically targeting the redevelopment of an area with the objective of achieving a *Walkscore* greater than 90/100. This demonstrates that a numerical benchmark can be a useful redevelopment criteria, even when the underlying mechanism is not understood. Hence, including more built environment criteria with numeric targets (similar to the values listed in Section 7.1.2) can be a useful process to measure the target and progress towards the target.

The members of this board, are interested in considering how a more comprehensive metric driven approach to measuring the built environment can be used to gauge the success of a long-term master-plan and consider using this approach.

7.1.4 *Using the* Neighborhood Visualizer *for Planning*

Here, I briefly summarize the proposed use-cases of this tool, based on these three hypothetical narratives. In *Scenario 1*, a web-based tool enables citizens without specialist skills, to generate reports that provide numeric arguments to justify their arguments.[4] I believe that this tool democratizes access to this information as it is no longer just a developer or planner who has control over the information required for a preliminary assessment of a development.

[4] While, this tool does require numeric literacy in assessing these arguments, it does assume a certain familiarity with a computer, web-browsing, access to the internet and a printer.

In *Scenario 2*, the members of a planning board can start to develop an intuitive understanding of the resource intensities of various parts of their city. Using this improved knowledge about the material and energy required for a range of urban forms, they can then specify criteria that they would like future developments to satisfy. Future plans can then be assess using these numeric criteria. While this non-dynamic approach does not consider the underlying mechanisms necessary to create a functioning, vibrant neighborhood, it is likely that these built environment measurements will result in material and energy intensities that lie within empirical upper and lower bounds that have been previously observed.

In *Scenario 3*, the planning board discusses how the numeric guidelines identified in *Scenario 2* are relevant as additional metrics that can be applied to a redevelopment plan. This simple use case example considers the limited number of levers that a policy-maker can influence, but does not consider the more sophisticated second order effects that may result. Similar to *Scenario 2*, it does not consider the essential components of what is necessary for a vibrant economy. However, it does provide a reasonable bounding of the problem. Nevertheless, there are many complications due to the multiple path dependencies associated with urban form, and development patterns can take a long time to change once established.

In general, one consequence of making these measurements explicit is that users (both planners and the general public) can have a clearer understanding of the resources associated with urban development patterns. One use of these resource intensity measurements could be to convert these measures into costs as this is relevant to both the city, and the individual. The individual directly pays for the energy they used for transportation while the city pays for the construction and maintenance of infrastructure. However the citizen is indirectly paying for the cost of this infrastructure through their taxes. As the actual cost of providing this service is not as transparent as it should be when considering overall urban form patterns making a clearer connection between the cost of providing various development patterns can motivate individuals to care more about the planning strategies that a city adopts.

From the individual's perspective, the cost of fuel (and time) is frequently not internalized until fuel rises. When it does increase, and is factored in to the market prices of neighborhoods, compact mixed use neighborhoods will undoubtedly be more attractive. These associated costs are considered in cities with budgetary crises as difficult decisions are made about where services such as roads, sewers and policing are no longer provided, to areas of very low density.

7.2 Neighborhood Visualizer

The objective of developing a web-based GIS tool was to demonstrate some of the tradeoffs associated with specific urban configurations through the exploration of a dynamic web-map. This web-tool provides a user-friendly way of exploring these tradeoffs, so that the user can learn what parameters influence resource consumption at the neighborhood scale. This tool uses the resource intensity measures described in Chapter 5. The user has the ability to normalize the measurements displayed by people or households (for example kWh per person or household). The objective of the tool is for the user to develop an understanding about the relationship between population density, material and energy intensities through the exploration of a range of different urban areas.

A working version of this website can be seen at urbmet.org

To use this tool, the user chooses a city that they wish to explore from any of the 40 US cities currently analyzed. The map zooms into the center of the chosen city and displays a box approximately 6 by 4 *km*. If the user presses the *Analyze* button, measures of population, energy and material are displayed on-screen. These values are summarized as both bar-charts, and as a *heatmap* overlay. The user can choose which measure (population, energy or material) to overlay on-screen. The user also has the option of drawing a polygon of any size or shape on-screen and analyzing this area. The interface shown in Figure 7.7, has several interactive help functions built-in to facilitate the user navigate the site. The user can also learn about the site features available using an interactive demo when the start page is loaded.

Examples of other spatially discrete urban sustainability tools can be seen at the following links: urbanecomap.org and walkscore.com

The tool has gone through several iterations, both in terms of the objectives, interface design and the implementation of the server-side spatial analysis. The first conceptual sketch is shown in Figure 7.2 where the initial idea was more focused on a *scenario analysis* type tool. After the first implementation (Figure 7.3) this was refined to focus more on resource intensity measures (though a dynamic design component was still included at this stage). Further iterations of the interface are shown in Figures 7.4, 7.5, 7.6 and 7.7 with the final objective of the tool focused on illustrating a range of measurements, rather than a tool where the user could change input parameters. On the server-side, the significant technical iterations were moving from using a *MySQL* database to a *PostgreSQL* database with *PostGIS*. *PostGIS* is a spatial extension for *PostgreSQL* that enables sophisticated spatial analysis to be performed on a database.

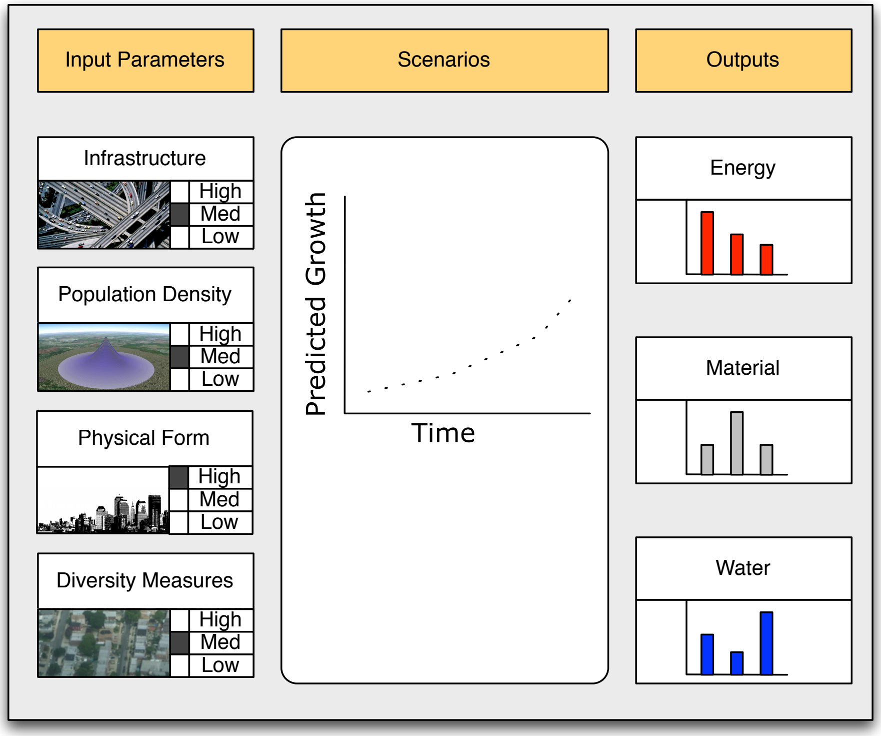


Figure 7.2: Initial concept sketch (*November 2010*).

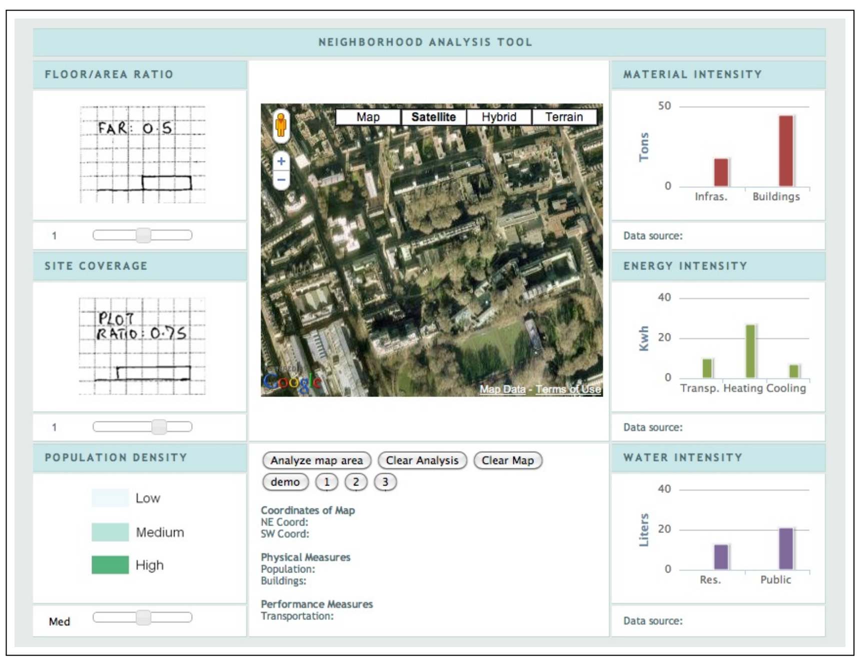


Figure 7.3: *Neighborhood Visualizer v0.1* interface. The input parameters on the left-hand side are for demonstration purposes (*January 2011*). A *MySQL* database was used to store data.

Figure 7.4: *Neighborhood Visualizer v0.2* interface. The input parameters on the left-hand side provide the user ways to adjust the mix of urban forms that are present (*June 2011*).

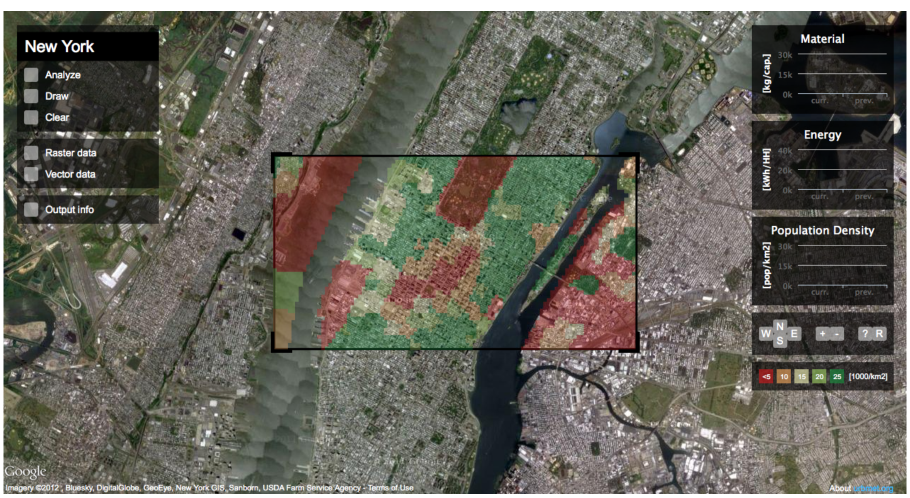


Figure 7.5: *Neighborhood Visualizer v0.3* interface. This version enables analysis of 40 cities, and the dynamic generation of reports. The database was changed to PostgreSQL with PostGIS (*December 2011*). The default analysis box is shown here with the results from a query displayed for New York City.

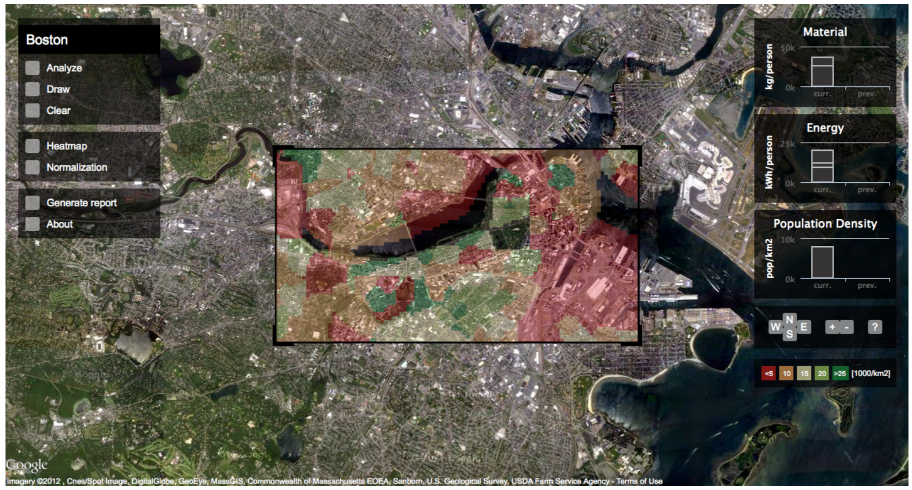


Figure 7.6: *Neighborhood Visualizer v0.9* (*April 2012*). The user can choose between three different *heatmap* options to overlay a population density measure, an energy measure or a material measure. The term *heatmap* was used as a colloquial term to describe the raster overlay.

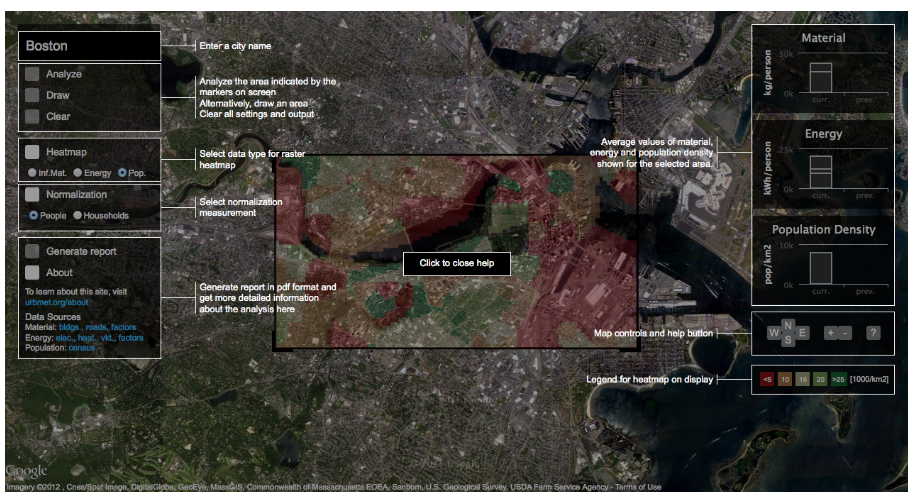


Figure 7.7: *Neighborhood Analysis Visualizer v0.9* with the *help* options expanded (*April 2012*). Choosing the interactive demo enables the user to step through the process of analysis as well as dynamic explanations of the tool's features.

7.3 Technical Details

The *Neighborhood Visualizer* uses an open-source stack of software to display data and to perform spatial analysis. This tool combines both raster and vector analysis, and is extremely fast, while sacrificing only a small amount of accuracy for the speed provided.[5] In addition, access is provided to the underlying data using desktop GIS tools using web-mapping standards. This web-visualizer is different from many other web-map tools for several reasons. Unlike many other complex web-maps, it uses *JavaScript* instead of *Adobe Flash*, enabling the site to function in most browsers[6] whether on a computer or tablet (assuming that *Javascript* is enabled). This tool was also checked for cross-compatibility on *Firefox*, *Chrome*, *Safari* and *Internet Explorer* on the *Windows* and *Mac* operating system.[7] The tool is also easily scaleable, as more data can be loaded into the database, and the site interface can interact with the spatial data in the same way without any performance reduction.

[5] The code for both the interface (web-browser) and back-end (server) was written specifically for this tool by David Quinn and Daniel Wiesmann.

[6] There have been a few minor problems with old versions of web-browsers (> 4 years). Many of the problems that we are aware of have been fixed.

[7] The four most popular web-browsers that have been used to access this site are *Chrome*: 38.5% *Firefox*: 26.4%, *Safari*: 15.3% and *Internet Explorer*: 9.4%.

Javascript was used to create an interactive interface in a web-browser, along with *HTML* and *CSS*. A summary of the libraries and software used in this website is listed in Table 7.1. The range of libraries and languages listed in Table 7.1 is necessary to achieve the desired site-functionality. The approximate number of lines of code for each language is listed in Table 7.2.

Javascript:	Openlayers, JQuery, Raphael, HighCharts
Database:	PostGreSQL 9.1 (64 bit), PostGIS 2.0
Raster display:	MapServer 6.0

Table 7.1: Software and libraries used in this website. All code is open-source, with the exception of *HighCharts* which is free for non-commercial use.

The *OpenLayers* library was used as it provides flexibility for interacting with a range of commercial web-mapping APIs. Currently, satellite imagery from *Google* is used as a base-map, however by replacing two lines of code, a base-map from *OpenStreetMap* or *Bing* could be substituted. Using the *Openlayers* library provides a layer of abstraction for interacting with a commercial web-mapping API. This abstraction layer reduces the rate at which the code used in this website will become obsolete due to changes in the commercial web-mapping API.

Languge	Lines of Code
Javascript:	1500
SQL	800
PHP	1200

Table 7.2: Approximate size of site (this does not consider the code written for data preparation and analysis), and does not include third party libraries.

The overall structure of the website is shown in Figure 7.8. When a user visits an urban area in one of the 40 cities and presses the *Analyze* button, a query is sent to the server with the latitude and longitude coordinates of the bounding-box on-screen.[8] Then, a spatial query is performed on the *PostGreSQL* database using *PostGIS*. The result is returned to the browser and the data is displayed on-screen in the form of a *heatmap* overlay, as well

[8] The default bounding box is a rectangle positioned in the center of the screen. The size of the box is exactly one third of the map, shown in the browser extents. The user also has the ability to draw any closed polygon on-screen.

as updating the three charts. The user can also examine a second area and compare the results as the two most recent results are displayed on-screen.

The raster overlay (previously referred to as a *heatmap*) is provided using a *WMS* service. One process used to speed up the display of data was to separate the *WMS* service from the *PostGIS* query. The raster overlay query is handled directly by *MapServer*, while the *PostGIS* query is handled separately. In this way, both queries are performed simultaneously using different computers and the results are returned to the user's screen almost simultaneously. Depending on the number of concurrent users, the *PostGIS* query sometimes can take longer, as this is running on a single spatial database without any load-balancing. The *MapServer* query is less computationally intensive and is handled by a web-host which uses load-balancing.

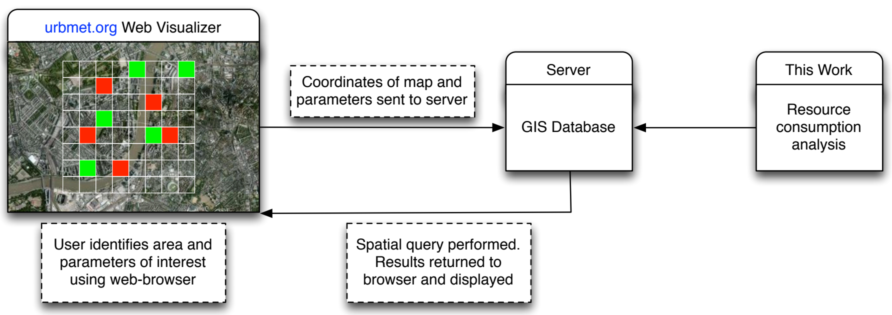


Figure 7.8: urbmet.org site structure

Good security practices were followed, in particular with regard to database queries which were performed using *SQL* parametrization rather than *PHP* string expansion which reduces the likelihood of *SQL* injection attacks.

To run server-side commands, the *PHP* language was used. The server-side code receives data from the web-browser and runs commands based on user requests; these included both spatial query requests and requests to generate reports. Several technical approaches were tested to identify the quickest method to display data in response to a spatial query in a timely manner. Initially, vector data was used, however the resulting query time took between 2-10 seconds for small areas; this response time increased as larger areas were included. To accelerate the response time, only raster data was used. The tradeoff of using raster data is that it is more inaccurate, however the rasterization approach used here tries to preserve as much information as possible in the vector-to-raster conversion process. The first step of converting this data, involved an abstraction of vector data to a raster grid, while retaining more information than is typically stored when using a raster grid. This approach removes some of the geometric details associated with the data, but retains the relevant information required for this calculation process, illustrated in Figure 7.9. The process described in Figure 7.9 is different than directly rasterizing the vector data, as more specific grid-level measurements are stored. Using an integer grid reduced the total space required for storing this data and increased the query speed.

Using a spatial query to access the data stored in a multi-band raster

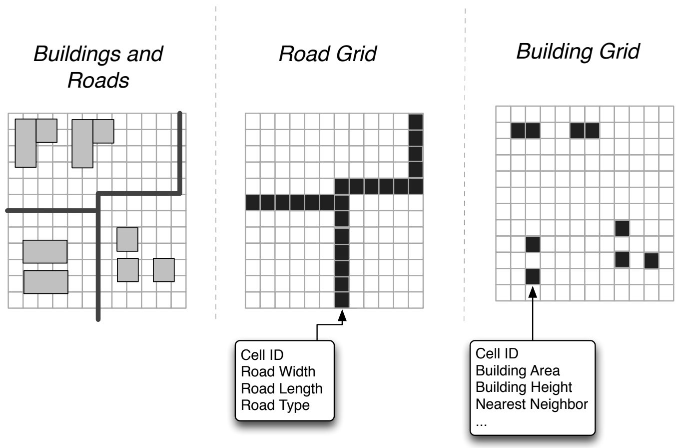


Figure 7.9: Vector grid used to convert the raw spatial information. The centroid of buildings is stored, and each measure is calculated for each part of the road segment per cell.

enabled many measurements to be returned using one spatial query. As spatial queries are the most time-consuming and processor-intensive step in this analysis, reducing the number of queries per user interaction resulted in significant speed increases. Instead of performing a spatial query on ten datasets, a query was performed on one dataset which returned ten values for the area of interest. This reduced the time that the user had to wait from seconds to milliseconds.[9]

[9] The requested data is displayed on screen in less than one second when using a two year old laptop with a fast internet connection.

The process of generating multi-band rasters is illustrated in Figure 7.10. These rasters were identical with the same resolution and extents, and were combined using *ArcGIS*. Three bands are shown in Figure 7.10 to illustrate how road information can be split into each layer, using details calculated from the raw vector information. These rasters are then stored in a multi-band raster and the information about each layer can be accessed through one spatial query. An example of the spatial query used to access the multi-band raster is listed in Appendix C.8.

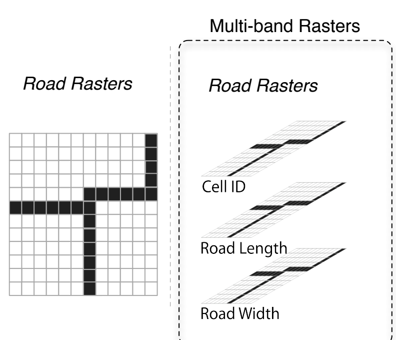


Figure 7.10: Raster grid of road data.

Finally, although large amounts of spatial vector data were used to perform this analysis, the final rasters for all cities are quite small (< 10 GB). This is an important consideration when considering the global scaleability of this analysis both in terms of computation and cost.

7.4 *Dynamic Report Generation*

All data and reports provided on this web-site are released under a *Creative Commons* non-commercial license. This allows users to freely use this data and to modify it, so long as they preserve the terms of the license agreement.

Using a similar approach to the on-screen display of data, the user also has the ability to generate *pdf* reports dynamically, of the area which they examine. The technical difference here is that a *pdf* file is generated on the server, and this file is subsequently returned to the user. The *pdf* is generated using the typesetting engine LaTeX, with graphs generated dynamically using the statistical language *R*. The overall process of generating this file takes approximately 10 seconds; this is due to the *PHP* script calling *R*, then LaTeX, and then returning the *pdf* file to the user's browser. The *R* process reads in a data-file for each city which is then used to generate five histograms that are customized for each area analyzed, and incorporated in each report.

Presenting data using an interactive format enables the tool to dynamically respond to the user's behavior. While this dynamic response is obvious for panning and zooming as there is visual feedback in the form of the map changing scale and location, resource intensity measurements can be dynamically changed depending on the zoom level and the size of the area examined. This is useful, as the accuracy and importance of certain measurements varies at different zoom levels and enables scale-dependent comparisons to be made.

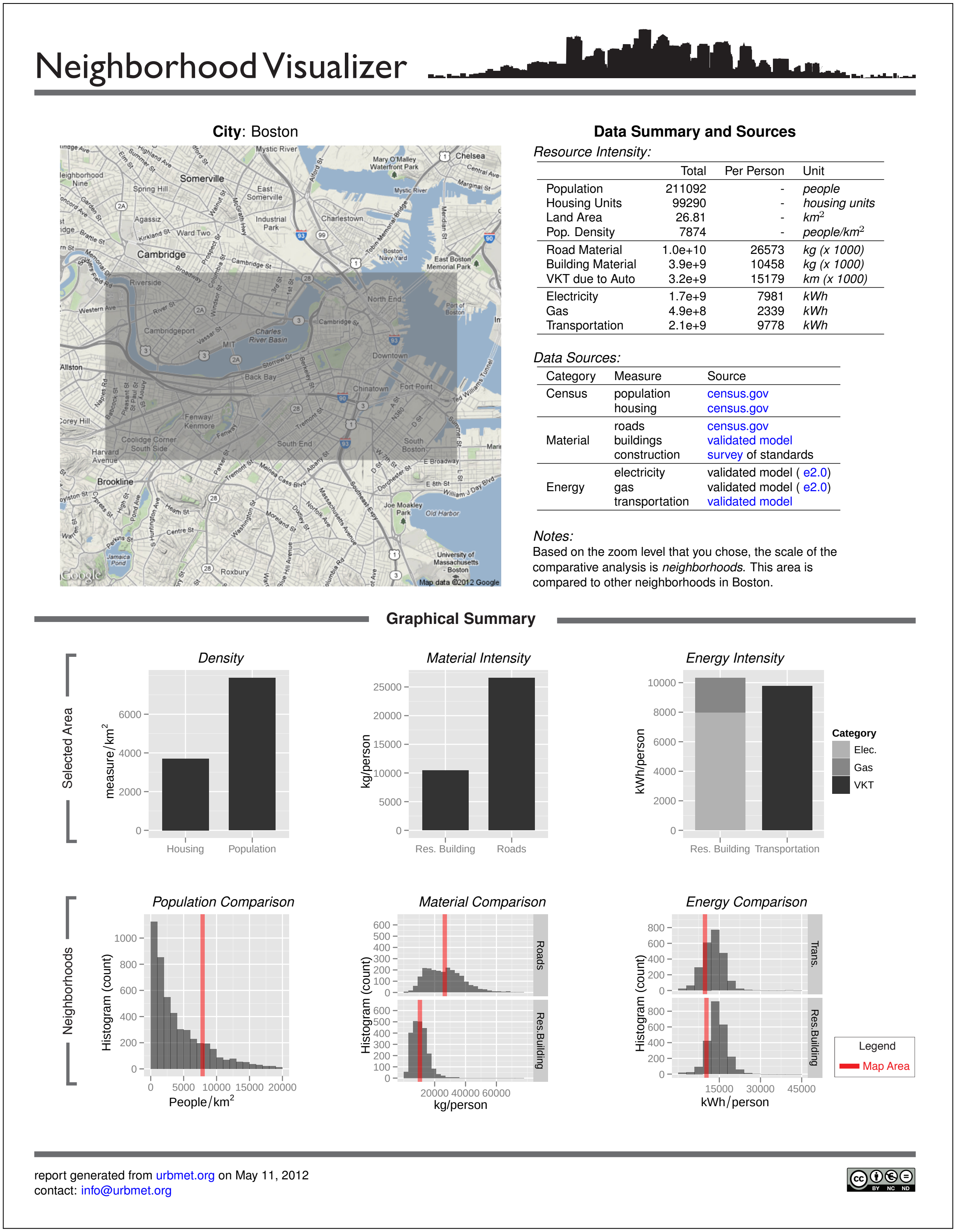
Table 7.3: Two zoom scales are considered. This is illustrated using images for New York, using Google aerial photographs.

Resolution	Illustration	Measurements
City		City level histograms. The chosen city is compared to other cities.
Neighborhood		Neighborhood level histograms. The area examined is compared to other block groups within the city.

This approach considers the spatial scale that the viewer perceives to influence the type of information presented about the resource intensity of the area examined. The scales for this dynamic reporting are described in Table 7.3. An example of a report for each resolution is shown in Figures 7.11 and 7.12. These reports are produced directly from urbmet.org and can be replicated by visiting the website.[10] These reports can be generated for any of the 40 cities analyzed.

[10] To refine this reporting process, multiple layouts were tested. Over 500 iterations were used to refine the report template and to develop the final version.

The benefit of this approach is that a sustainability logic can be encoded into the analysis. In this case, the logic is dependent on the physical scale, but it could also be tailored to the criteria that the user is most interested in, such as the cost of infrastructure. The risk of such an approach is that it reflects the website creators' perspectives and is not a reflection of the community's values. It is important that this approach is carefully applied so that it can be representative of the values of the community that it aims to serve.



Neighborhood Visualizer

City: Boston

Map data ©2012 Google

Data Summary and Sources

Resource Intensity:

	Total	Per Person	Unit
Population	211092	-	*people*
Housing Units	99290	-	*housing units*
Land Area	26.81	-	*km²*
Pop. Density	7874	-	*people/km²*
Road Material	1.0e+10	26573	*kg (x 1000)*
Building Material	3.9e+9	10458	*kg (x 1000)*
VKT due to Auto	3.2e+9	15179	*km (x 1000)*
Electricity	1.7e+9	7981	*kWh*
Gas	4.9e+8	2339	*kWh*
Transportation	2.1e+9	9778	*kWh*

Data Sources:

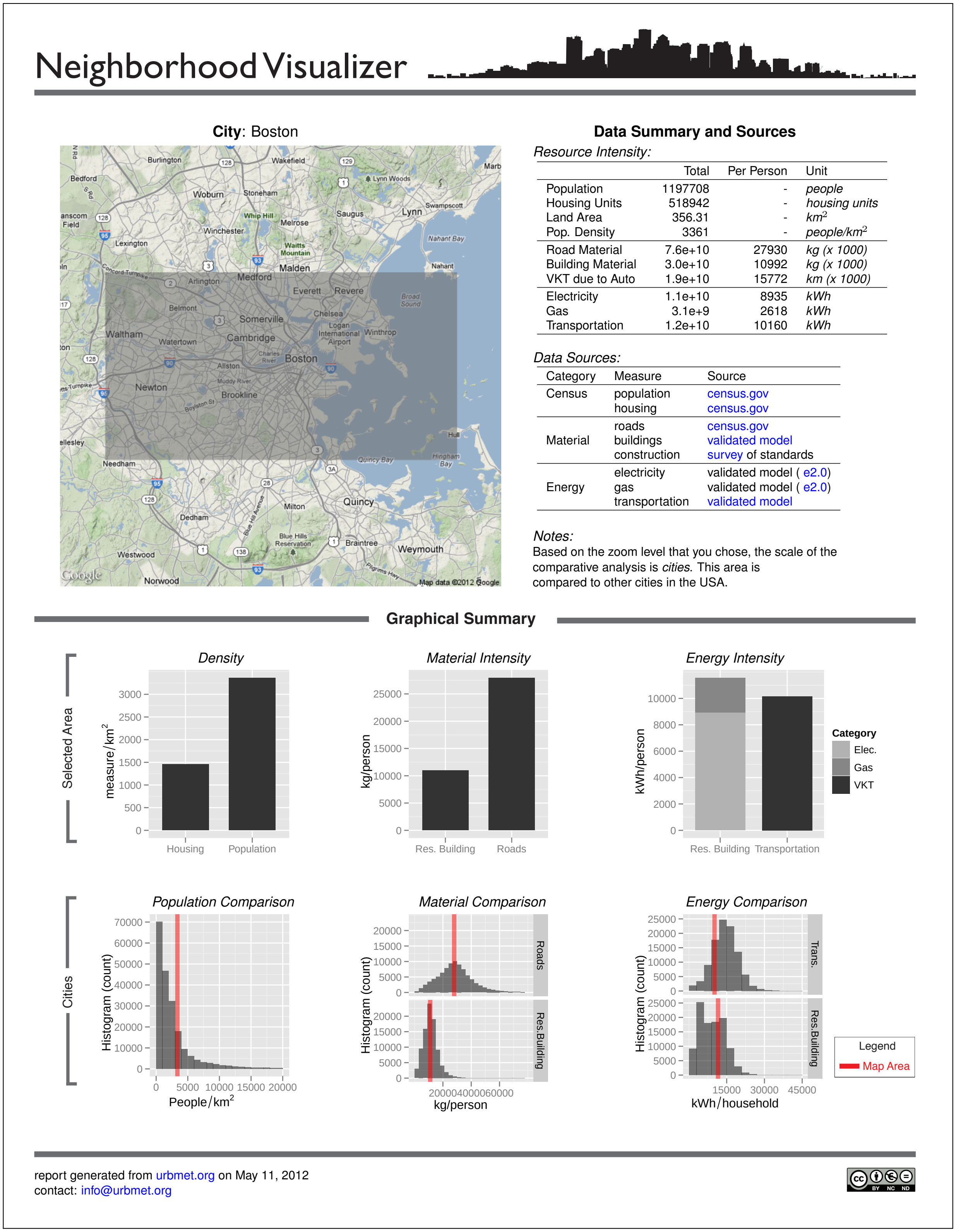
Category	Measure	Source
Census	population	census.gov
	housing	census.gov
Material	roads	census.gov
	buildings	validated model
	construction	survey of standards
Energy	electricity	validated model (e2.0)
	gas	validated model (e2.0)
	transportation	validated model

Notes:
Based on the zoom level that you chose, the scale of the comparative analysis is *neighborhoods*. This area is compared to other neighborhoods in Boston.

Graphical Summary

report generated from urbmet.org on May 11, 2012
contact: info@urbmet.org

Figure 7.11: Report for Boston, with comparative measurements at the neighborhood scale.



Neighborhood Visualizer

City: Boston

Data Summary and Sources

Resource Intensity:

	Total	Per Person	Unit
Population	1197708	-	*people*
Housing Units	518942	-	*housing units*
Land Area	356.31	-	*km*^{2}
Pop. Density	3361	-	*people/km*^{2}
Road Material	7.6e+10	27930	*kg (x 1000)*
Building Material	3.0e+10	10992	*kg (x 1000)*
VKT due to Auto	1.9e+10	15772	*km (x 1000)*
Electricity	1.1e+10	8935	*kWh*
Gas	3.1e+9	2618	*kWh*
Transportation	1.2e+10	10160	*kWh*

Data Sources:

Category	Measure	Source
Census	population	census.gov
	housing	census.gov
Material	roads	census.gov
	buildings	validated model
	construction	survey of standards
Energy	electricity	validated model (e2.0)
	gas	validated model (e2.0)
	transportation	validated model

Notes:
Based on the zoom level that you chose, the scale of the comparative analysis is *cities*. This area is compared to other cities in the USA.

Graphical Summary

report generated from urbmet.org on May 11, 2012
contact: info@urbmet.org

Figure 7.12: Report for Boston with comparative measurements at the city scale.

7.5 *Dissemination of Analysis*

An additional purpose of this website is to disseminate the results of this research, so that the underlying data can be accessed by researchers in an easily accessible way. The structure of this portion of the website is shown in Figure 7.13. I have started to share this analysis by publishing the results as a Web Map Service (WMS) (Open Geospatial Consortium, 2012b) and a Web Coverage Service (WCS) (Open Geospatial Consortium, 2012a).

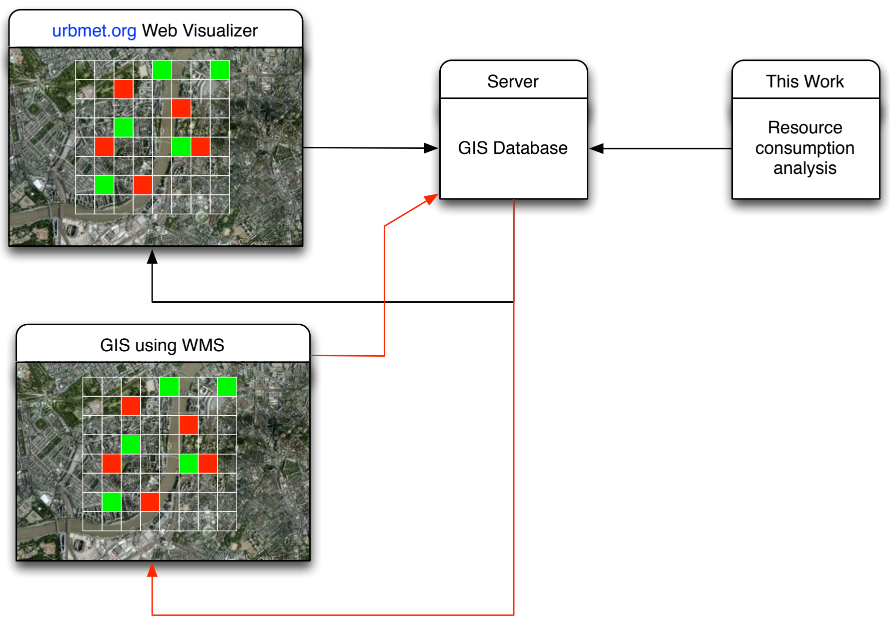


Figure 7.13: Accessing data using a desktop GIS via a Web Map Service or Web Coverage Service.

These links are not urls that a web-browser recognizes. Instead they can be connected to using GIS software (for example QGIS, OpenJump or ArcGIS) which enable a user to load in a data-layer from a web-service. The WMS addresses of this analysis are listed in Table 7.4. Using these standardized protocols enables the rapid dissemination of this analysis.

Measure	WMS Address
Material	`http://urbmet.org/mapserv.cgi?map=material.map`
Energy	`http://urbmet.org/mapserv.cgi?map=energy.map`
Population	`http://urbmet.org/mapserv.cgi?map=population.map`

Table 7.4: WMS addresses for accessing the results of this data as a base layer (the actual data is not returned with this query, just an image overlay).

Similarly the raw data will also be accessible using a *WCS*. While the WMS GetMap request returns a map image, a *WCS GetCoverage* request returns the raw raster data in *GeoTIFF* format. This enables the user access to the underlying rasters which can be loaded into a desktop GIS program and analyzed. The WCS addresses of this analysis are listed in Table 7.5.

Figure 7.14 and 7.15 illustrate how *GIS* software can be used to connect

Table 7.5: WCS addresses for accessing the results of this data as a *GeoTiff*.

Measure	WCS Address
Material	http://urbmet.org/mapserv.cgi?map=material_wcs.map
Energy	http://urbmet.org/mapserv.cgi?map=energy_wcs.map
Population	http://urbmet.org/mapserv.cgi?map=population_wcs.map

to these datasets. The web-address listed in Table 7.4 is entered into the map-browser (Figure 7.14) and an image of the data is returned to the user's desktop GIS (Figure 7.15). This data can be styled according to the user's preferences and incorporated with other spatial data into a map.

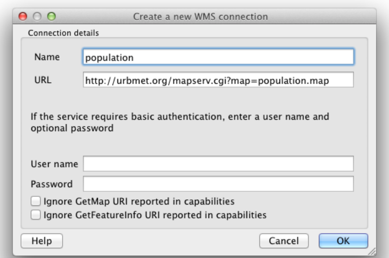


Figure 7.14: Connecting to the WMS service using *QGIS*. This *urls* are listed in Table 7.4.

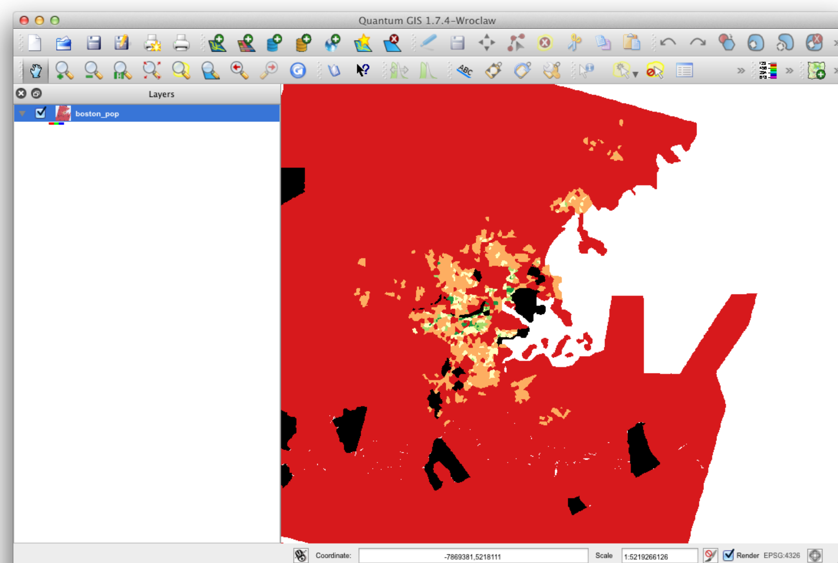


Figure 7.15: Viewing the WMS service using *QGIS*. Here population density for Boston is loaded into the data window using the WMS *url*.

7.6 *Survey and Analysis of User Behavior*

To explore how users interacted with this site, a survey was designed and implemented using a website that was custom made for the purpose (survey.urbmet.org). The purpose of the survey was twofold. The first purpose was to explore how users interacted with this site, and to test if it assisted with learning about sustainability measurements. The second purpose was to identify how the site could be improved.

The survey site was custom made using several *Javascript* libraries and data was recorded on a *PostgreSQL* database. Temporary cookies were used to track the user behavior. In addition, *Google Analytics* was used to monitor some aspects of the site usage. Users were asked a series of questions, introduced to the tool, and then asked the same set of questions. The survey structure is shown in Table 7.6.

Step	Process
1	Eight questions on urban measures (+ information about user)
2	User explores urban areas using *urbmet.org* (40 cities to choose from)
3	Eight questions on urban measures asked again

Table 7.6: Survey structure.

290 users took part in Step 1, and explored the tool in Step 2. The breakdown of users who visited the site is listed in Table 7.7. The total number of users that answered questions at Step 1 and Step 3, and had cookies enabled was 125.

Profession	Count	Experience using GIS	Average Experience Level Scale: Novice (1) - Expert (3)
Architect	128	52%	1.3
Engineer	122	45%	1.4
Planner	19	63%	1.5

Table 7.7: Measures used to track-user behavior. Users were not obliged to answer every question. There was an option for users to self-identify as *Policy Makers* and *Other*; no users chose these options so they are omitted from the survey analysis. Of the users who had experience using GIS, their average skill level was calculated using a scale from 1 to 3.

The results of the survey are shown in Table 7.8, with the percentage of users who answered the questions correctly shown. The results are inconclusive. For some questions, there is a significant improvement (Question 1[a] for example), but for other questions the values are zero or below. It is not clear whether the tool facilitates learning about some measures, or whether these patterns are not significantly clear to be observable through the exploration of a small sample of values. The answers for Question 2[b] seem to suggest that the tool results in users having a worse estimate of this specific pattern after using the web-tool. It was not possible to answer Question 2[a] directly from the data, but it could be deduced based on the normalized infrastructure material measurement. Possible reasons why users answered the question the second time with a lower accuracy might be due to unclear phrasing, or the user not exploring enough areas on the site to identify a trend.

	Survey Questions	Step 1 (%)	Step 3 (%)	Change (%)
1 [a]	What do you think the typical range of population density in US cities is?	44.83	70.43	25.6
1 [b]	What do you think the population density of suburban housing is?	61.21	65.79	4.58
2 [a]	In the city center, what fraction of land is covered by roads?	50.86	44.25	-6.61
2 [b]	In the city center, do you think the linear road length per person is higher or lower than the suburbs?	80.17	63.16	-17.01
3 [a]	In the USA, how many kilometers per year does the average household travel by automobile?	63.25	68.38	5.13
3 [b]	How does this travel distance change if you live in the center of a city?	97.41	91.15	-6.26
4 [a]	If you live in the city center, what fraction of your total energy use is due to transportation?	24.56	34.21	9.65
4 [b]	How much energy (electricity, gas & transportation) do you think a suburban household uses when compared to a household in the city center?	84.48	76.72	-7.76

Table 7.8: Responses to questions before and after taking the survey.

Users were also asked to comment on whether this tool would be useful for their work or research. The answers are summarized in Table 7.9. Overall, it seems that users found this tool useful.

Table 7.9: Survey question on usefulness of tool.

	Yes	Maybe	No
Would you find this tool useful for your work or research?	34%	48%	18%

7.6.1 *Comments and Improvements*

A comment box was also provided to allow users to suggest improved functionality and suggested changes. Several users requested the inclusion of water information. Several commenters also requested access to the underlying data. While access to the data is documented on urbmet.org/about of the site (where the users can connect to the underlying data using a GIS tool[11]) perhaps this could be made more obvious.

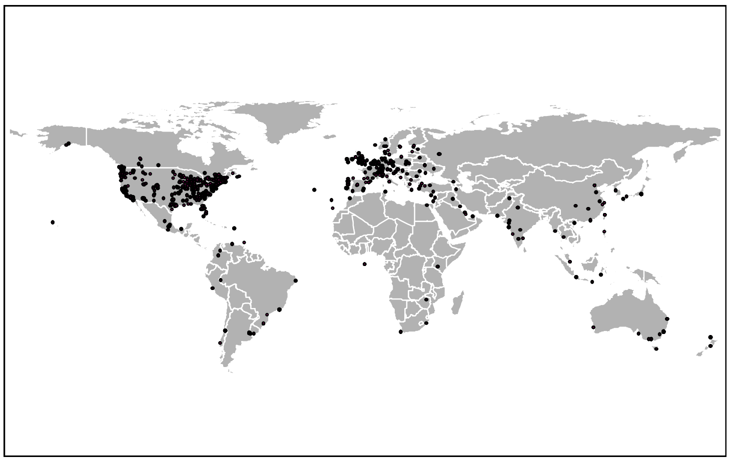
[11] This is discussed in greater detail in Section 7.5.

In general, there were few complaints about speed, and aside from some operating-system/browser version issues early on, the majority of users suggested functionality changes, rather than criticisms of the analysis. There were several color scheme suggestions, with some users disliking the color reversal of scale depending on the measurement type. In addition, one color-blind user pointed out that the colors used were not color-blind friendly. Another common request was for more baseline measures to be shown to the user as they explore areas within the browser. All users comments are included in Appendix D.

7.6.2 *User Behavior*

The location of users and the centroids of queries are shown in Figure 7.16. 3500+ users visited urbmet.org (at the time of writing) with each user performing approximately four queries.[12] The number of users per day is currently 75-100. The bounding boxes of all queries made are shown in Figure 7.17 (a). The queries for five large US cities are shown in Figure 7.17 (b) - (f). Most queries occur near the center of the city. The website automatically takes the user to the this central location, however it can be seen that users do explore a wide range of scales. A base-map from OpenStreetMap is used here to illustrate the city area.

[12] In general, users searched areas where data was available from this research, but some users queried areas in other countries, and extremely large global areas. If data was not available for the requested area, a message was returned to the user, stating that data was not available for the current location.



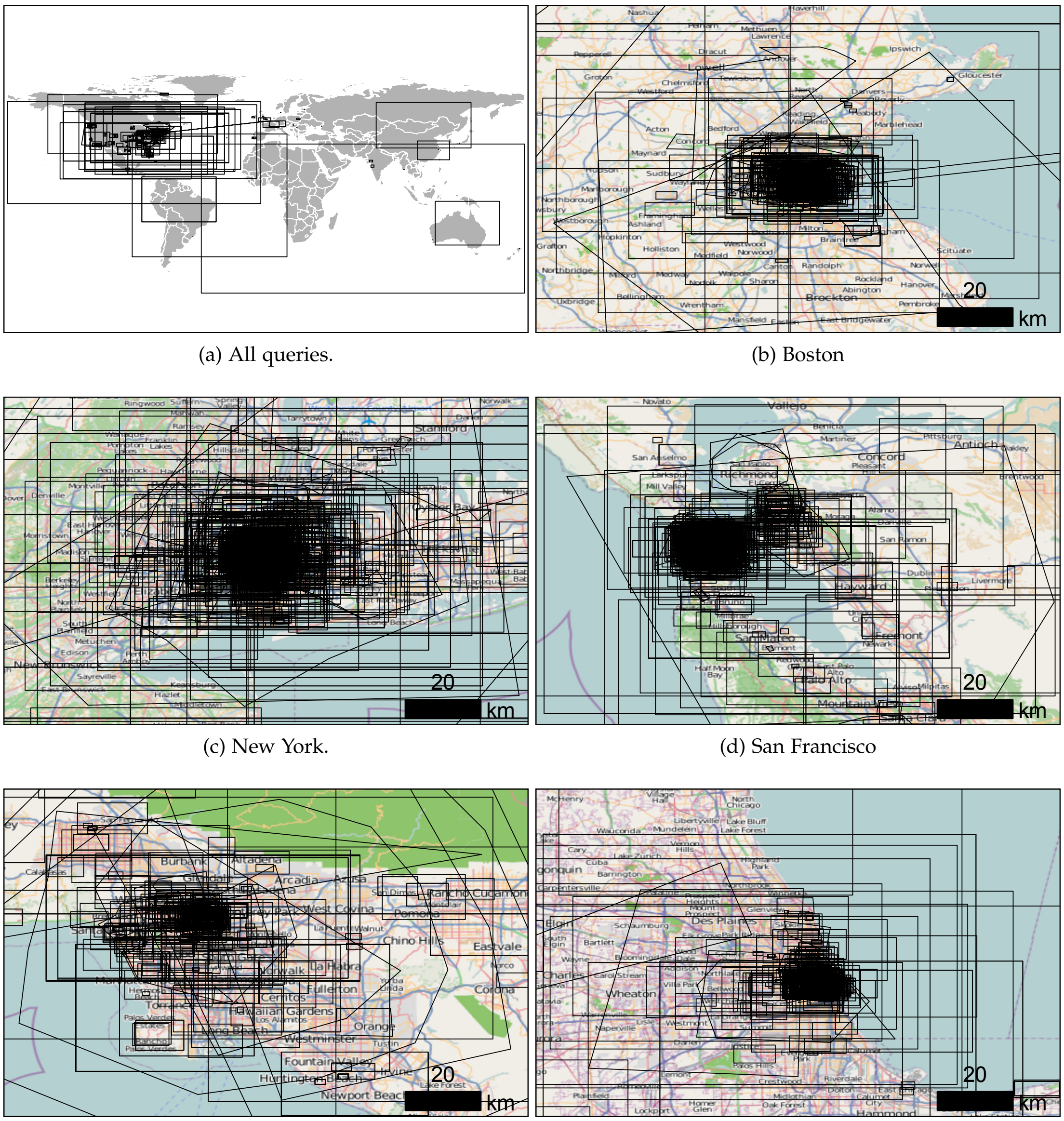
(a) Location of the first 3,000 users.

(b) Polygon centroids of 12,168 queries.

Figure 7.16: Site visitors and areas that were queried. Users examined areas that were not analyzed as part of this research.

Some simple monitoring tracked the behavior of users who visited the site by assigned a unique identifier to each user. Any queries that were made were recorded and associated to that user's behavior. Then, by geocoding[13] the IP address of the user an approximate latitude and longitude was identified.

[13] To perform this calculation the IP addresses were geocoded to *Class C* using a database provided by *IP2Location* which identifies the general metro area of IP addresses. A *Python* script was written to identify the latitude and longitude for each IP address using this database. This geocoding only used the first parts of the IP address to identify the user location. This was due to both privacy reasons, and the fact that the non-commercial version of *IP2Location* does not enable more accurate geocoding.



(a) All queries.

(b) Boston

(c) New York.

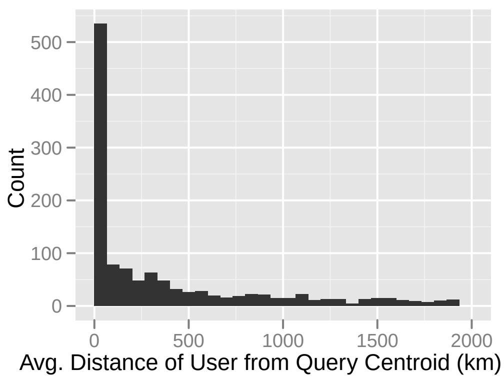
(d) San Francisco

(e) Los Angeles

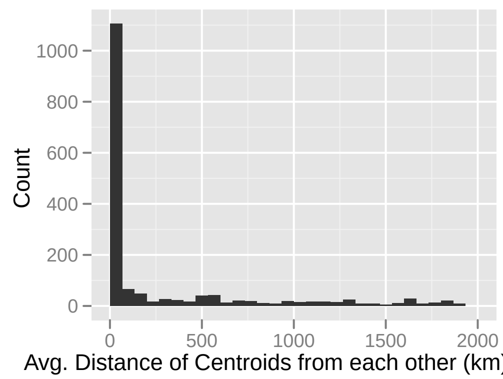
(f) Chicago

Figure 7.17: Examples of queries that users performed. Queries outside the analysis area did not return any data, just a message that data was not available for the user's area of interest.

Figure 7.18 (a) shows a histogram of the average distance from the geocoded IP address of the user, to the centroids of the polygons that were queried. Figure 7.18 (a) illustrates that a significant number of users visited areas that they were near their physical location (no non-US users were included in these distance calculations). This result confirms that users are interested in exploring resource intensity measures for their neighborhood. However, based on the survey analysis it is not conclusive whether their learning was improved or not through the exploration.



(a) Average distance to query area from geocoded user-location.



(b) Average distance from centroids to each other, grouped by user.

Figure 7.18: These calculations were only performed for IP addresses or queries that were within in the US.

Based on the general patterns of behavior that were observed from site-usage statistics, typically users examined four areas and zoomed in slightly, decreasing the size of the area that they examined. The area of the query is illustrated in Figure 7.19 (a). The sequential query size is shown in each facet of this plot, and the mean value per query for all users is shown in Figure 7.20. Figure 7.19 illustrates the sequential queries by 3000 users the size of the area viewed by the user. The average value of these queries is shown in Figure 7.20. The number of queries diminishes for each step. In general, users frequently searched areas in close proximity to the area they start exploring initially (Figure 7.19 (b)).

A unique ID was used to identify each user, and then each users's queries were ordered by timestamps. The distribution of Query 1 to Query 6 is shown in Figure 7.19. As the number of queries reduced at each interval, the sum of the histogram becomes smaller. We can see that the user starts to examine areas at a certain scale (this initial scale is obviously based on the site preset values), but the user then starts to zoom in slightly, to examine a smaller area than they started off with (Figure 7.20). A small number of queries with areas greater than 100 km^2 were not included in the average values shown in Figure 7.20 as they distorted the more refined patterns that exist at a smaller scale. Some users queried substantial fractions of the entire

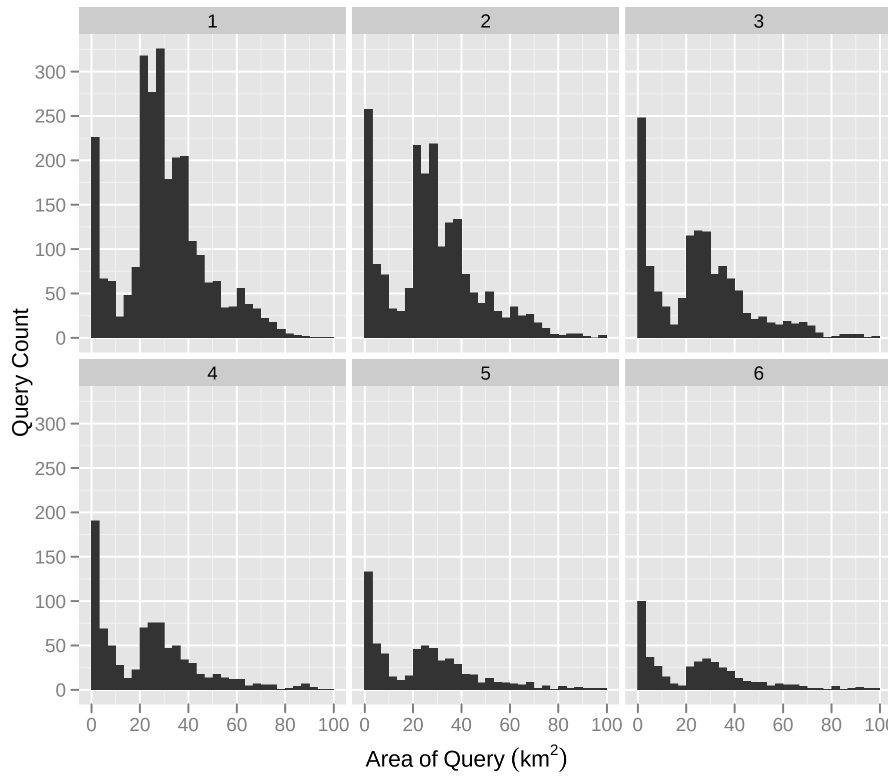


Figure 7.19: Area of query, based on the order that queries were performed by each user. There were approximately four queries per user.

world, which skews the mean value disproportionately.

From the histograms in Figure 7.19, there appear to be two types of users. One group of users is interested in zooming in to very small areas, while the other group is more interested in medium-sized areas. To illustrate the size of these areas, examples of queries from the Cambridge/Boston area are shown with an area less than 1 km^2 and areas in the range of 23 - 27 km^2 (Figure 7.21). These areas correspond to the two peaks in the histogram of Query 1 (Figure 7.19). While zoom level increments in the web-map may have some influence on the area size chosen, perhaps the chosen areas can also be explained using human cognition and the scale that humans understand urban patterns.

Using the professions identified in the survey, I explored if there was any noticeable difference in the query areas between architects, engineers and planners. A histogram of queries is shown in Figure 7.22 using data from 202 users. Aside from the fact that a smaller number of users are analyzed in this histogram, there is no noticeable difference between this pattern of behavior and the pattern in Figure 7.19; the bi-modal pattern exists in all three histograms.

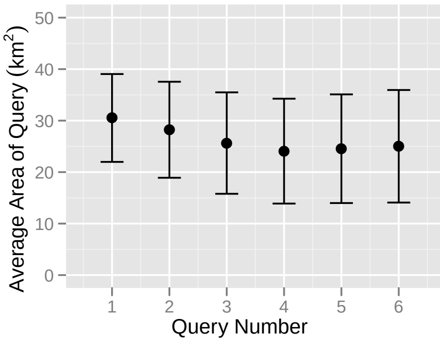
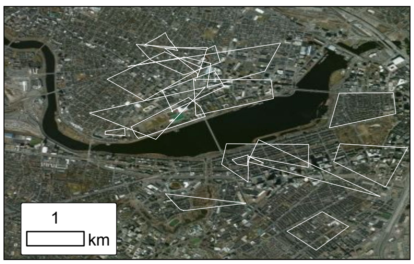
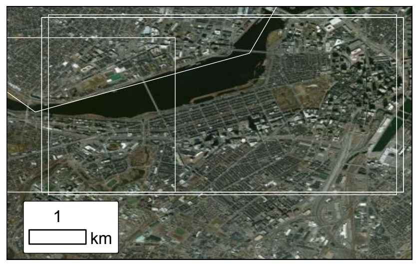


Figure 7.20: Average query area per interaction (error bars show one standard deviation).



(a) Queries with areas less than 1 km^2



(b) Queries with areas between 20-30 km^2

Figure 7.21: Illustration of query areas. A base-map of aerial imagery from *Bing* maps is used here.

To illustrate some specific user behaviors, four random users were chosen with the condition that they made at least four queries in the US. A map of the USA (Figure 7.23) shows the general location of each of these queries. These queries are shown chronologically in Figures 7.24 - 7.27. [14]

[14] Any duplicate queries were removed, and the queries are shown with an OpenStreetMap base-map.For example, some users pressed the *Analyze* button several times without changing the location of the analysis box.

User 1: In Figure 7.24 we can see several queries that this user made. They first started to explore New York city using the default bounding-box generated by the site, centered on Manahattan. They then examined a similarly sized box, for a different part of Manhattan. Then they moved to Portland and explored the city at two different spatial scales - a regular box, and then hand-drawing a small triangle to explore a small area of the city.

User 2: In Figure 7.25 we can see several queries that *User 2* made. Again, the user started to explore New York city using the default bounding-box generated by the site. They then zoomed out and explore more of New York city. Then they moved to Syracuse, NY and explored an urban area at two different spatial scales using the auto-generated boxes. The areas that they chose to explore are not currently in the database, so no results were returned for Queries 3 and 4.

User 3: In Figure 7.26 we can see several queries that *User 3* made around the city of Los Angeles. The user first examined an area in the city center (as defined by the website). They then explored three other areas in close proximity to the first location, varying the scale slightly.

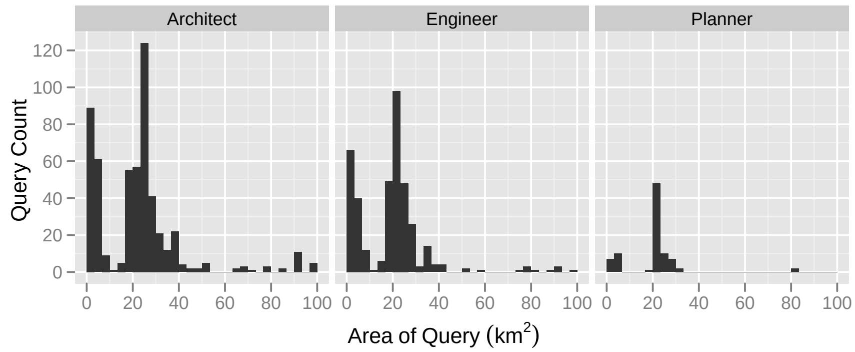


Figure 7.22: Histogram of query area per profession. The mean values for each facet are: Architects: 22.4, Engineers: 19.3, Planners: 20.3

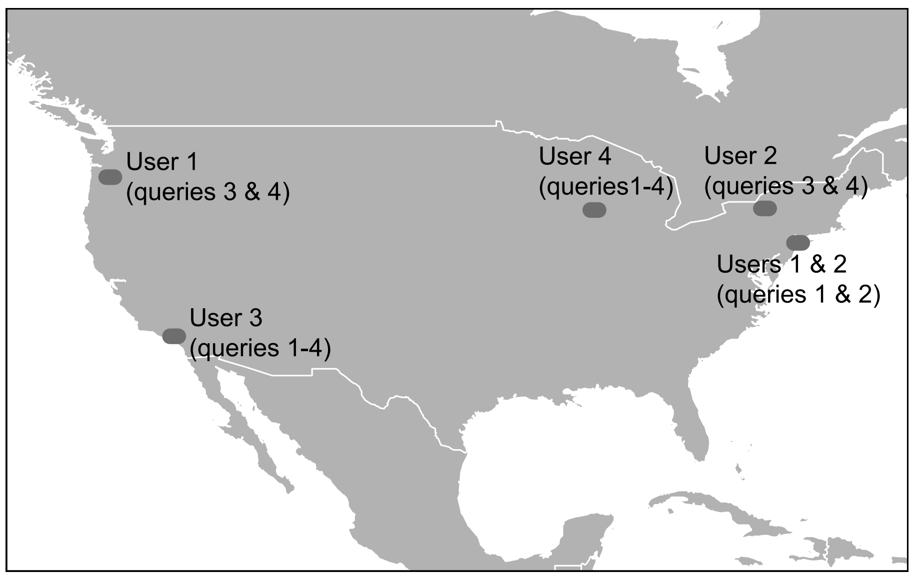


Figure 7.23: Areas examined by users.

User 4: In Figure 7.27 we can see several queries that *User 4* made. Here the user examined several areas around Milwaukee. The user started by using two auto-generated boxes in the city. Continuing to examine the city at the same scale, the user then drew an area to exclude the water boundaries. They then finally examined another part of the city using an auto-generated box at the same scale.

A general observation about the individual users (and the city queries shown in Figure 7.17) is that users are interested in higher-resolution measurements, rather than city level measures. This is also shown in the histograms shown in Figures 7.19 and 7.20. In addition, users are interested in custom areas through the drawing of polygons, and exploring the areas that they are interested in at a range of scales. All of these interactions would not be feasible using a paper map.

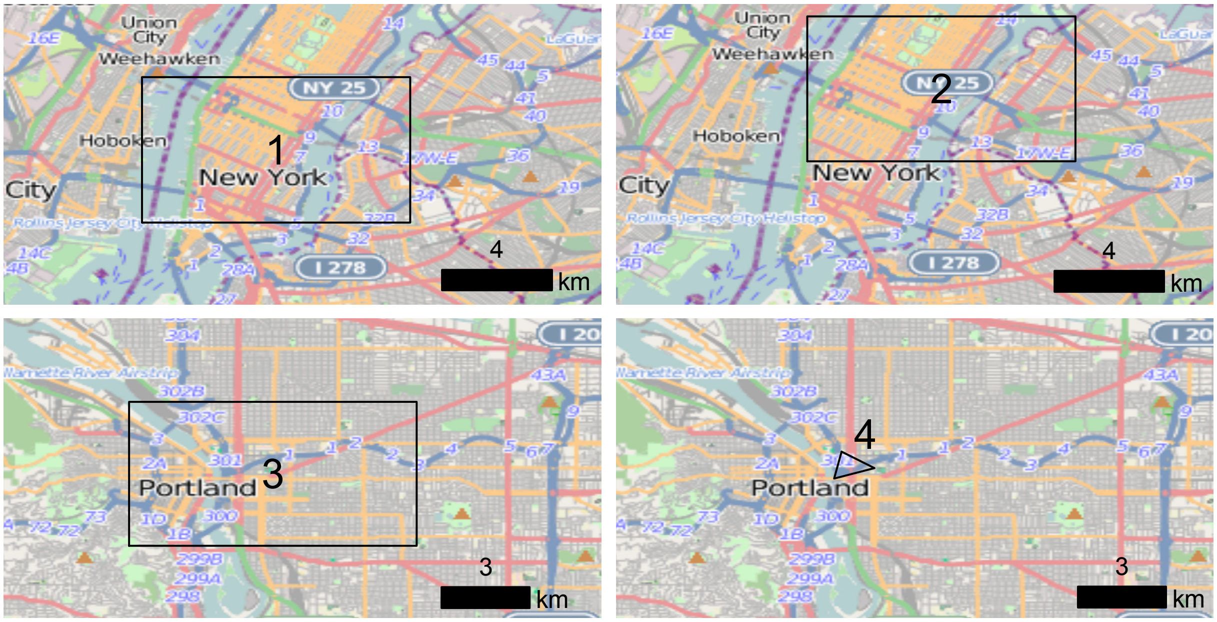


Figure 7.24: Behavior of user #1.

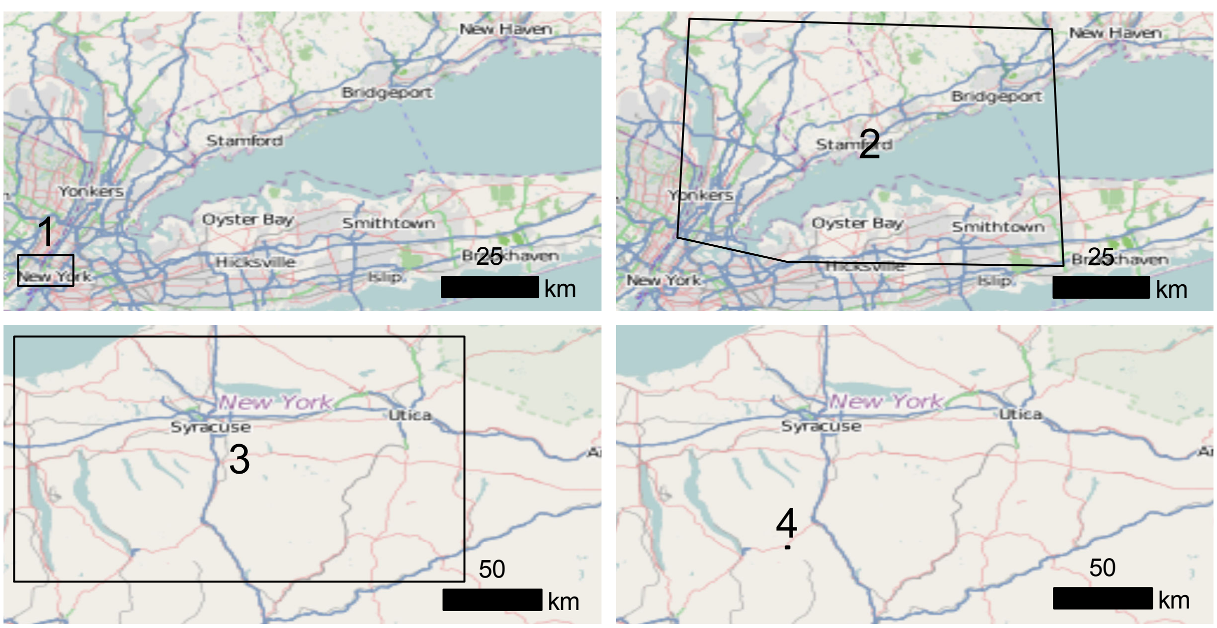


Figure 7.25: Behavior of user #2.

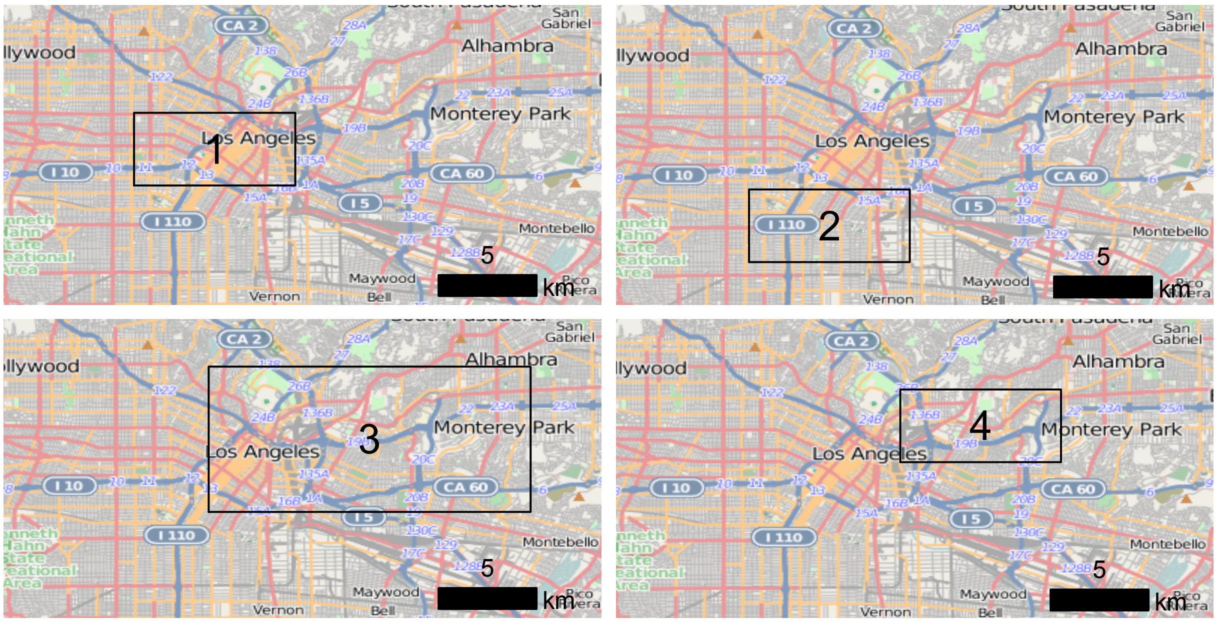


Figure 7.26: Behavior of user #3.

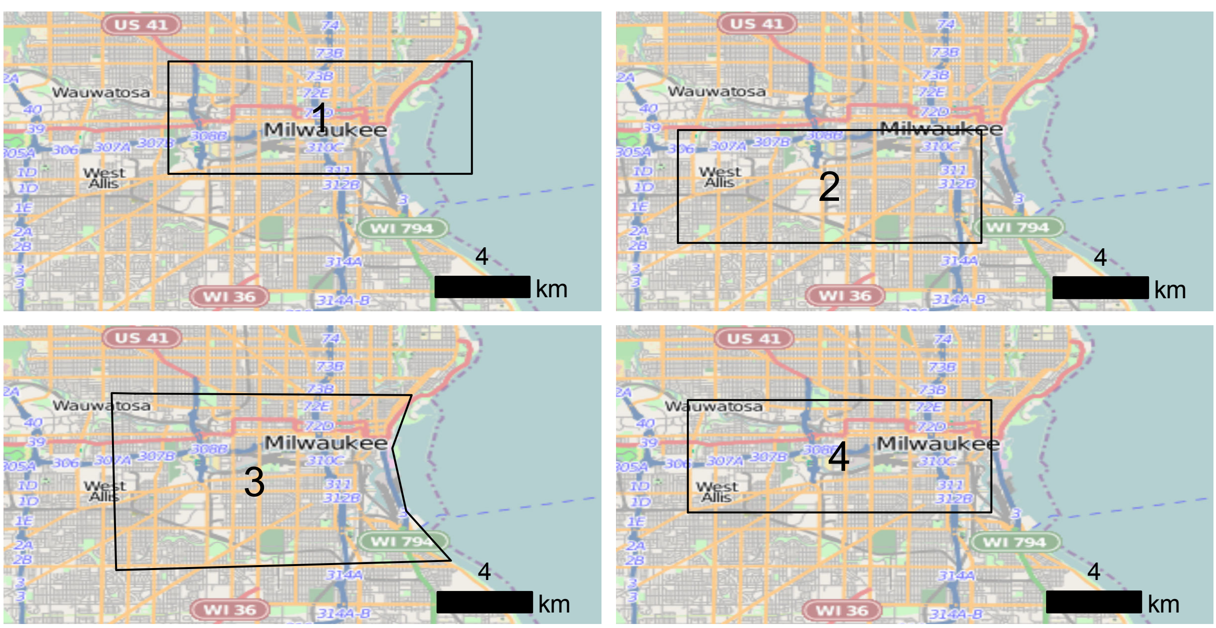


Figure 7.27: Behavior of user #4.

7.6.3 Neighborhood Visualizer *Improvements*

After considering this tool through the lens of the three scenarios, several improvements become apparent. First, more baseline data measurements per city could be included to assist users to understand whether they are above or below significant thresholds; this would encourage a broader consideration of absolute resource intensities required for neighborhood functioning and not just consider how the area analyzed compared to other neighborhoods. These baseline measurements should include error bars to illustrate the accuracy of the estimate.

Some of the measurements used to predict resource intensities (listed in Chapter 4) would be useful to include on a second page of the report. As the responses to the survey questions illustrated, there are some city trends that are difficult to observe using a small number of measurements. Hence larger macro patterns of behavior could be emphasized (such as road length per person, for example). This current lack of clearly identifying trends may result in an obfuscation of the overall argument, unless the user can understand the bigger picture. While this was attempted with the inclusion of histograms, I believe that more data representation of this type would help.

After exploring these hypothetical use cases, I realize that the reporting features should include more conventional planning metrics and criteria that could be used to assess urban plans. For example, while I state the importance of intersections per km^2 as a metric, this measurement is not included in the report though it would be a useful planning criteria to consider. The measurements here are very focused on the quantitative assessment of resource usage. Other measures that could be included are *Floor Area Ratios*, road intersection counts and other spatial criteria that can be easily understood and measured from a plan, which would enable a planner to match a proposed development to a specific case generated from the report.

Finally, the ability to share a link to a report would be a useful function as a user can argue a particular point and then include the *url* as a reference; whether the discussion is electronic or on printed media. In addition, social networking is an effective means of sharing a message quickly; ensuring that a reference can be tailored to illustrate a particular point would facilitate users to share interesting observations (and perhaps errors) in the data.

7.7 Methods of Analysis, Data and Future Technology

In this section, I discuss three themes related to the analysis process and data. I start by discussing the standardization of analysis methods and the tools that facilitate this standardization. I then discuss the issue of data availability and the associated technological and political issues. I conclude by discussing how these themes can be considered when urban analysis is considered in the long-term, and the influence that citizen engagement, democratization of information and user-generated data will have on this process.

7.7.1 Formalizing the Analysis Process

The academic community that focuses on resource usage patterns in cities has contributed much to the conceptualization and standardization of what should be measured using methodologies such as Life-Cycle Assessment (LCA) and Material Flow Analysis (MFA). Less work has been done considering how spatial variation within the city should be considered, or even how to apply these methodologies to smaller areas. It is important for similarly rigorous methods of analysis such as MFA or LCA to be developed that consider the spatial dimension. The reasons for this lack of standardization are partially due to the availability of data and partially due to researchers using traditional methods of analysis. There is a need for more work to emphasize the process of analyze using spatial data so that standards can be developed.

I believe that is important for researchers to move away from using a *graphical user interface* or GUI, and to move towards using a script-based approach for spatial analysis. Just as Microsoft Excel has caused users to mix data and results in a way that often makes it difficult to separate the raw data from the final results, pointing and clicking to perform GIS analysis is not a rigorous way of conducting research. There is no audit trail and after performing a task once, the user cannot easily repeat that task, or share the process. I believe that a script-based approach should be considered to be a fundamental step for GIS analysis that is used in research, as it facilities the sharing of the analysis process.

Next, I believe that it is important to encourage the use of open-source tools for researchers and centers of learning when performing GIS analysis. There are several reasons for this. The first is that researchers should understand how calculations are being performed and should be able to cite the algorithms being used. This is not possible with proprietary GIS software. The second reason is due to data formats. The most commonly used spatial data format, the *shapefile* (developed by ESRI) is not documented completely.[15] Although ESRI (2011) claim that they wish to promote a hybrid approach to GIS analysis which encourages open-source technology, in reality this is not true as the file formats which ESRI promotes, are not open.

[15] For example, the quadtree method of spatial indexing for the *shapefile* is still being reverse-engineered.

One of the problems with large companies becoming involved in the development of commonly used standards for data-sharing and the standardized communication protocols is technological lock-in. Ensuring that open-data formats are used, prevents this problem. Hence, it is extremely important that communication protocols are standardized using open-standards, particularly when the data is being transmitted in non-human readable formats, such as binary data. This is particularly relevant when dealing with urban-data, where the data is anticipated to have a long time span. Another argument in favor of open-source GIS tools is that they are technically better than the proprietary alternatives for large scale spatial analysis.[16]

Similar to the world of software, if a company that uses a proprietary file-format ceases to exist, the danger is that the data stored in this format is lost. For example, the benefits of storing data in a *.csv* file compared to a *.xls* file are clear. In ten years time, it will be possible to open the .csv file; however it is not clear if the *.xls* file will be usable. There is already a *.xlsx* format, and no guarantee for backwards compatibility in the software program *Microsoft Excel*.

[16] Speed tests comparing *PostGIS* and *Oracle* have shown *PostGIS* to be far quicker. ArcGIS file formats and analysis tools are not efficient for large-scale data analysis.

To conclude, not only do the underlying tools need to be transparent, but the process of analysis (the underlying code used for the analysis) should also be clearly documented and shared to further research in the academic field, as discussed recently by Hsu (2012). Through the use of open-source analysis tools, and clearly documented public file formats, researchers from other domains can easily access this information and data without significant difficulty. Equally importantly, money is not a barrier to accessing the analysis tools that professionals use. While it is not realistic to assume that the general public will be able to use GIS software with ease, solely by virtue of its availability, it is important to ensure that this remains an option. In addition, more creative and innovative uses of both data and analysis methods are possible when there are few restrictions on who can access the analysis methods and raw data.

7.7.2 Data Availability and Information Flows

Overall, the approach of assembling these datasets and making predictions of energy and material accessible through a web-browser democratizes access to this information. No specialist tools are needed to explore material and energy patterns, or to generate dynamic reports summarizing the analysis (though it is harder to tailor this general analysis to the specific needs of all users). It is hoped that this work can help researchers access data that exists commercially, but is expensive to access.

From my experience of assembling large amounts of urban spatial data there are three groups that can be used to categorize cities. The first group is a city with no web presence for data accessibility. This is more common amongst small cities, but there are some large cities where the data is simply not easily available online.[17] The second category of city, provides access to data via an interactive web map, typically using a flash-based interface that does not allow access to the underlying data. Frequently, websites of this type do not provide data in a format that is easily usable.[18]

The final data sharing structure that cities use is one where the raw data is easily available and searchable. New York, Chicago and Portland are cities that are using websites of this type,[19] and in my experience this data man-

[17] For example, while the City of Cambridge provides extensive and detailed GIS data, but the primary means of accessing it is through purchasing a low-cost CD of data.

[18] To clarify the distinction between the visualization tool that I have developed, and these interfaces - the *Neighborhood Visualizer* is built around explaining three specific measures in combination; material, energy and population, and the underlying data is available using web-mapping protocols.

[19] A commonly used software for this purpose is made by the company Socrata; this company develops websites that allow the download of data. The Socrata model of content management system for data enables searching, tags and feedback to be incorporated into the data.

agement system is the most useful. However, this functionality is unrelated to the quality of the data. Cities frequently publish data that is not useful, as the assumptions, or metadata is not adequately described. The structure of the *Socrata* website helps address this problem as the user can view the most popular datasets, and ask questions about some measurements. Nevertheless, the best approach would be to ensure that comprehensive metadata is always included.

Many of the datasets that describe the built environment at the necessary level of resolution for resource consumption analysis, are expensive and owned by private companies. It is hoped that this research can help accelerate the process of more detailed resource consumption data being provided at an aggregate level that is still useful to researchers for analysis. Publishing slightly aggregated information (at the block group for example) can help inform policy decisions, while the aggregated values can avoid containing personally identifying information. For example, the UK has an admirable strategy in this regard as it provides block-group level information for the entire country that contains a range of measures at a high resolution. This includes information about the area of residential and commercial buildings, in addition to publishing yearly electricity and gas consumption measures.

Many cities still charge commercial rates for researchers to access data. For example, the City of Chicago does not provide access[20] to parcel information at the city level. The intention is for the city to sell the information to advertisers or market-researchers. This demonstrates a flawed approach to how local governance should function. Firstly, I do not consider the city to be owners of the data; they are merely caretakers of this record of public information and they should not be allowed to sell it for profit. Secondly, I believe that providing information about land-use patterns and detailed information about buildings is of greater societal benefit to the local community than selling it to advertisers. I believe that federal transparency laws should ensure that data of this nature is publicly available for research and planning purposes.

[20] In response to a request for parcel-data I received the following terse response from the planning department in Chicago:*"The parcel data is proprietary"*.

Overall, one problem with urban level data is due to inconsistencies in metadata. Metadata is frequently incomplete and there is not a good mechanism for reporting this back to the city (aside from the *Socrata* software tool). In the survey of US cities I was unaware of any city providing an interactive map that conforms to a WMS or WFS standard. I believe that this could be another model that cities should consider providing; an example of such a web-map is shown in this work.

Graham (2012) discusses an interesting strategy that *Wikipedia* are considering adopting, referred to as Wikidata. The *Wikimedia Foundation* are considering using a central repository of referenced information that be can be drawn on by many articles; when an update is made to that data value, the change is passed on to all articles referring to that numerical fact. I propose that a similar approach would be of benefit for urban data. In the same way that review papers are published in academic journals, there is a need for peer-reviewed measures of urban form, from around the world. Through the identification of upper and lower bounds of these measurements, assumptions and estimated values can be improved, as well as the accuracy of urban models.

Finally, there is a shortage of comprehensive urban databases for analysis. In the fields of statistics, image recognition and machine learning there are canonical datasets that are used to test and benchmark new methods of analysis. Such a dataset does not exist for urban areas. Developing and providing a canonical database could enable more researchers from different fields to become involved, with the objective of predicting energy or material consumption.[21] A focus on identifying the most significant measurements of urban form (including socio-economic measures) and measures related to resource consumption, could have far reaching consequences for the urban sustainability community.

7.7.3 *Analyzing Future Cities*

The methods used to analyze cities are rapidly changing as the use of ubiquitous distributed sensors can gather data at a much faster rater than was ever previously possible. Although the datasets used in this research were substantial with 3.5 million building polygons and 100,000 block groups analyzed, the data streams available from real-time data are orders of magnitude larger. Nevertheless, it is important to appreciate the fundamental difference between real-time information, and developing an understanding of long-term measurements that are not dynamic. For example, while users can provide feedback about the functioning of a particular system (such as the performance of a bus route), there is a difference in providing feedback within a system, and the creation of a new system. While feedback can be useful for performance measurement and local-optimization, this approach does not currently have the ability to drive changes within the system to the point where it can be considered within the larger framework of infrastructure planning.[22]

The process of real-time gathering and data sharing is of great relevance due to the involvement of both large companies with centralized infrastructure and hierarchies, and individuals with autonomy (researchers, civic-leaders and hacktivists[23]). There is much discussion about the tension between large companies (such as Cisco, IBM and Siemens) forming partnerships with cities while gathering and controlling urban data streams. While it is likely that a compromise between the top-down (large organizations) and bottom-up (individuals) will result in the optimal solution, there are few agents in the middle ground. Technological commentator Anthony Townsend suggests that the field of urban planning would be well positioned to coordinate the two sides to reach a compromise (considering the

[21] The movie rental company *NetFlix* achieved fame in the statistical/machine learning community when they sponsored a competition to see who could improve on their prediction algorithm. A comparable approach could be for a a city to sponsor a similar competition, with the goal of predicting some urban sustainability measurements.

[22] The organization *MySociety.org* are a UK based charity with the objective of encouraging residents of an area to become involved in civic activities in their local area. They have developed several heavily-trafficked websites that encouraged citizens to become involved in local issues (both infrastructural and political), and through this involvement have moved towards driving community involvement for transportation changes, as well as encouraging transparency in parallel.

[23] Activists who use technology for social issues (either to radically change the system, or to improve the system) but using cutting edge technology to achieve their goals, sometimes repurposing the existing technology innovatively.

additional complications of privacy and transparency), but highlights that the planning community has historically been slow in embracing technological innovation. Johnson (2010) describes the process of reusing an existing innovation for a new purpose as *exaptation*; a process which he has identified in many human innovations. I believe that the process of exaptation will be beneficial for cities, but that this technological innovation is dependent upon the existing infrastructure to first exist, and secondly for this infrastructure be sufficiently open, to enable innovative repurposing.

There have been many innovative and interesting developments that rely on volunteers to generate data and to provide feedback. Two such examples are *SeeClickFix*[24] and the *311* service[25] adopted by many US cities. An example of a detailed global dataset that raw data that is publicly available and widely used, is OpenStreetMap (openstreetmap.org).

One final thought associated with the future analysis of cities and data, is the concept of the minimum *information density* needed to achieve a certain level of accuracy. Considering the diagram in Figure 1.3 describing the process of analysis when this work was introduced. I am proposing that there is a minimum threshold of information that is necessary depending on the achievable accuracy of analysis. This information could be considered to be a basic target that all cities should strive to achieve.

To summarize, the future of analysis in cities will likely involve a mix of tensions over data and privacy as active citizens develop and use new tools, while there is a dependance on large companies to provide the fundamental infrastructure.

[24] *SeeClickFix* is a website (seeclickfix.com) that provides a means for users to contact their local government about problems that need fixing in their community. Based on this crowd-sourcing approach individuals can provide quantitative feedback based on their geographic location. A smartphone app is also available.

[25] *311* is a means for cities to receive feedback, from a site such as *SeeClickFix*. There is a proposed protocol for this communication method open311.org which is a result of the data stream being open, and a desire for standardization.

7.8 Conclusions

The objective of this tool is to educate non-experts and to influence the planning process using a learning feedback loop, illustrated in Figure 7.1 through the display of information about resource intensities of urban areas. It provides a platform agnostic way to assess the resource intensity of urban areas as well as providing users with a means to explore urban areas that they are familiar with. This tool provides a user-friendly way of accessing multiple urban measurements and has the ability to disseminate results to researchers rapidly. I describe the iterations that this tool went through as it was developed, and discuss the technical details associated with this web-tool. A suite of open-source tools and libraries were used in this work, and a fast, spatial analysis tool was developed.

I introduce the reasons why a web-tool that facilitates basic GIS analysis of material and energy patterns in urban areas is useful, using a series of scenarios to illustrate how it can be used. I give some hypothetical examples of how a tool like this can be applied. The ability for non-specialists to generate reports that justify their arguments democratizes access to this information. The main reason is a democratization of a data, and the ability for non-experts to justify their arguments backed by facts, using recent data and sophisticated spatial analysis. This provides them with a similar level of technical expertise to experts, which can make their argument more credible.

Providing analysis tools that enable the user to interact with the data enables the user to derive their own conclusions. This is a more powerful learning approach than observing the same values in a static report. This is in part due to the medium; I believe that static maps greatly limit our understanding of spatial data. Google Earth and Google Maps have radically changed how the general public and planners perceive the world and how spatial data can be represented. Dynamic methods of representation and exploration facilitate users to build a new type of intuition about the urban environment, in addition to the recent emergence of spatially situated tools that enable users to directly interact with their environment using smartphones.

As was observed from the user-tracking, many users examined urban areas that they were near. In addition, if a user performed several queries, they were likely to search near the location of their first query. There is also a tendency for users to zoom in slightly with subsequent queries. In general, users were not interested in city level measurements. This may have been due to the user background as it seems likely that the general public are more interested in the details of a particular area that they can relate to, rather than the overall urban performance.

A survey was performed on users of this site. Initial 250 people responded to this survey, the effects of this tool on learning are inconclusive. For some measures, there is a significant improvement (in particular population density measurements). There are several possible reasons why these results are not more clear; (1) some of the units used are unfamiliar to the US audience, (2) the patterns can be hard to identify from a small sample number, (3) not enough baseline measurements were provided so that the user can gain an intuitive understanding of whether the area that they examine is above or below average. Overall, users are interested in such a tool for their work or research.

This website received a reasonable amount of traffic, with over 3,800 users during the three weeks of operation (prior to the completion of this thesis). The site was featured on the websites of *The Guardian* (The Guardian, 2012) and *The New York Times* (New York Times, 2012), as well as being featured on several other prominent websites such as *Infosthetics* (Infosthetics, 2012), *O'Reilly Media* (O'Reilly Media, 2012) and the innovation and business website *Fast Company* (Fast Company, 2012). After this initial publicity, the number of visitors to the site leveled off to approximately 100 people per day.

A form was also provided on the website where users could request the inclusion of their city. One senior planner requested the inclusion of two cities from California, with an offer of data for both of these cities. In addition, an energy efficiency organization asked us to consider issues associated with state-level energy analysis, considering dynamic report generation targeted at energy efficiency measures.

Overall, it appears that there is a substantial interest in this tool, based on the number of users thus far, contact from policy makers and some news coverage. I believe that the rapid dissemination of results from this analysis can benefit research in the domain of urban analysis, by making data more easily available and accessible to both specialists and non-specialists.

8 Conclusions

In this chapter, I summarize the arguments that I have made throughout this work and discuss what I consider to be the important emerging themes. The objective of this research was to develop repeatable methods of spatial analysis so that material and energy intensities at the neighborhood scale could be compared using the same initial assumptions. Over the course of this research, I have developed and refined standardized analysis methods that characterize neighborhoods using data from 41 cities.[1] These analyses identified spatial and geometric patterns within cities and related these physical attributes to material and energy use. I hope that this analysis can inform the planning process and facilitate the inclusion of quantitative measures of sustainability in policies that consider the overall urban system.

[1] This included 40 cities in the US and one from the UK.

8.1 General Conclusions

Considering this work from the perspective of data scarcity, the first general conclusion is that I have identified cross-city patterns, with clear upper and lower boundaries. This is discussed in Chapter 3, where I describe the process of identifying urban characteristics using geometric measurements of roads and population. These upper and lower boundaries can be used to identify appropriate ranges of road infrastructure, when analyzing data-scarce cities.

I have also demonstrated how road network measures and population density can be used to describe characteristics of the built environment. This spatial analysis process is described in detail in Chapter 4. Using these spatial measurements, I identify statistical models to estimate the road area, the residential building area and the average building height, using 3.5 million buildings.

In Chapter 5, I illustrate how these geometric measurements can be related to the resource intensity of cities, using the analysis of the built environment described in Chapter 4. To estimate the material content of roads and buildings, I use construction standards and survey data. To estimate the residential energy use I consider private automobile travel and the energy use within buildings. The energy used by private automobile travel is estimated using measurements of the spatial configuration of the city, in combination with the population density. The building energy data uses the results from another national US study.[2] I combine block-group level estimates of transportation energy with zip-code level estimates of building energy, to estimate the average energy per household, per block-group. Then, using these estimates of material and energy for all 40 US cities, I explore how these values vary throughout the city and discuss trends amongst the cities examined.

[2] This study was performed by *Efficiency 2.0*.

In Chapter 6, I show how the urban boundary definition influences the average values used to describe a city. I also discuss how this analysis can assist planners in identifying what policy targets are feasible, and how an understanding of material and energy use for urban typologies can contribute to an understanding of the consequences of urban patterns.

In Chapter 7, I discuss the need for new visualization and analysis tools, how they can be used to inform the planning process, and democratize access to information. Then I describe how an interactive web-based tool was developed and how it can be used to explore material and energy intensities. This web-tool is used to display the resource intensity measurements described in Chapters 4 and 5. I demonstrate that there is a high level of interest in this web-tool, and examine how users interacted with it. In addition, I make several observations about methods of analysis and data, considering short and long-term trends. I emphasize the challenges associated with data acquisition when performing urban-scale research, and discuss how cities, corporations and open-data standards will influence future access to data.

8.2 *Specific Conclusions*

In this section, I describe the specific conclusions for each phase of the analysis, shown in Figure 8.1. These steps consists of characterizing urban areas, identifying geometric parameters using spatial analysis, relating these geometric patterns to material and energy use and visualizing these measurements using an interactive web-tool.

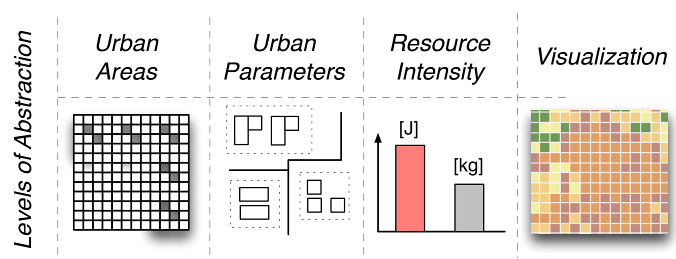


Figure 8.1: The four steps in this analysis. (This figure is repeated from Chapter **??**).

8.2.1 *Urban Area Characterization*

The characterization of urban areas was done using three steps. The first step involved identifying the boundary of urban areas using a population density threshold. Next, the population density and road density gradients were examined using concentric rings measured from the CBD.[3] An alternative approach to characterizing the urban form was developed using measurements from a 250 grid to identify a range of road density and road per person measurements.

[3] A method to identify the CBD was developed using the density of services.

This non-spatial representation of grid calculations, enables the viewer to understand the distribution of behavior within a city. Clear cross-city spatial patterns with upper and lower boundaries of infrastructure measurements can also be observed. In addition, by considering the overall urban area the viewer can understand the overall distribution of these spatial measures. These measures were combined into information-dense diagrams of a city that can be used to quickly characterize the overall structure of the urban form, and enable rapid comparisons to other cities.

8.2.2 Urban Parameter Identification

These attributes can be related to the resource intensity of cities. Three specific measurements that I have focused on identifying at a high-level of resolution are:

- road area (Section 4.1.1)
- residential building area (Section 4.2.1)
- average building height (Section 4.2.2)

These predicted values were all performed at the block-group level. The *road-area* is shown to be highly-correlated with linear road length, which is an expected result. The R^2 of model used to predict road-area is 0.6888. The *Residential Building Area* estimates can be predicted with an R^2 value of 0.6476. The R^2 for the model used to predict the *Average Building Height* is 0.4371.

8.2.3 Resource Intensity Measures

I have shown how the building volume varies per population density and correspondingly how the building material demand changes based on these demands, and also how infrastructure intensity varies. When energy is considered similar patterns hold true. This work provides strong quantitative arguments for minimum levels of density to be required in urban areas. I have also identified a range of geometric measurements that can be used to predict VKT per household and validated this model using a large sample from the state of Massachusetts.

The first pattern of behavior is a similar curve shape for large US cities, with high-population density areas. The second pattern, is that smaller low density cities do not have significant variation in terms of the overall resource intensity. I illustrate that the spatial variation of resource intensity within cities should be considered for large cities with high-population densities in the core, but that this variation is significantly lower for smaller, low-density cities.

While the general trends of resource intensities are not surprising, it is important to emphasize that cross-city patterns were observed in this analysis. This consistency across patterns is a useful result. Additionally, this work emphasizes the non-linear relationships that occur between resource intensities and other measures (such as population density) which is another important conclusion.

In Chapter 7, I discuss in greater detail the material and energy use patterns that can be observed. I illustrate how this pattern of behavior varies for 40 cities in the USA, and identify some specific ranges of material and energy use that are likely to occur. This is of particular relevance when we consider that the highest resource intensity areas are typically on the perimeter of the urban boundary. In addition, a substantial percentage of VKT can be explained due to the spatial configuration of the urban form.

8.2.4 Visualization

[4] This work was done in partnership with Daniel Wiesmann.

Over the course of this work, I have developed[4] a web-tool that satisfies the demand for the visualization and analysis or urban resource intensities. The first version of this tool contains data for 40 cities in the USA and can be viewed at urbmet.org. Within the first three weeks of launching this tool, we have had over 3500 users, and have been featured on the websites of two newspapers (The New York Times and The Guardian) in addition to being featured on several other high profile blogs, both in the domain of urban sustainability, but also in the field of data visualization and web-mapping. The number of visitors to this website illustrates the significant interest and demand for this type of work. Based on a survey of initial users who used the site, there were requests for the inclusion of more data, in particular water.

User behavior was also tracked on the site and it was observed that users frequently visit the city that they are located within. In addition the area size that the user examined gradually decreased as the user zoomed in to the initial area of interest. Also, users typically examined areas that were close to the initial area that they chose. In general, the vast majority of users were interested in high-resolution measurements and less interested in city-level analyses.

I believe that the field of web-mapping and global data sharing is one which can have a significant contribution to the understanding of human-physical systems which can lead to an improved understanding of sustainable resource use around the world.

8.3 Analysis Processes and Data

In this work there were three main technical considerations. The first was a focus on repeatable analysis methods, the second was on the tools that facilitated the sharing of the analysis process and visualization while the third considered issues of data availability.

Analysis Process: In this work I have placed much emphasis on describing the analysis process to perform this work in such a way that it was repeatable and easily shareable. All of the analysis was performed using scripts, where the set of commands were documented and written to a file. Using this approach enables the analysis to be applied to many cities, and to ensure that a record of the analysis steps is maintained. As the data used to identify the statistical models in this research came from industrialized cities, it may be necessary to refine the parameters used with samples of data from the cities of interest.

The underlying process of analysis is extremely important. Closed source proprietary tools with GUIs are not appropriate for researchers who wish to share their analytical process. While a certain level of technical sophistication is necessary for researchers to perform their analysis in such a way that it can be shared, documenting this process of analysis is of fundamental importance as more urban researchers who focus on the analysis of urban areas can learn from the process.

Analytic Tools: Based on my survey of previous research and currently available GIS tools, I identified that there was a need for better tools for planning that could be used to assess energy and material use. As a result, the *Neighborhood Visualizer* tool evolved to satisfy this demand and to provide an interactive way for users to explore a range of resource intensity measures.

This work identified that there was a need to provide tools that enable the synthesis of information. Based on feedback from users, the majority found this web-tool useful and requested the addition of more information (in particular, measurements about water intensity). This is one tool, to facilitate this type of urban exploration; I hope that many more tools of this type become available.

There should be better GIS tools available to encourage users to explore patterns from larger datasets.

Data Availability: I hope that future research on the built environment is less focused on the issue of data accessibility, and more focused on research and the analysis of patterns and resource consumption. In part, this work is a necessary iteration in the cycle of identifying what data is necessary for this research. My hope is that this work contributes to the identification of useful measurements that are considered to be important by cities, based on some of the calculations I have performed. In reality, there is likely to be a convergence of user-generated data and the release or more aggregated high-resolution city data due to an appreciation of the benefit associated with facilitating urban analysis. This does not make this work irrelevant, as heuristics can be identified from this theoretical approach and used to check the accuracy of user-generated data. One of the major challenges with user-generated data is ensuring that it satisfies some quality assurance standard. The more theoretical approach developed in this research can enable such user-generated data to be checked.

One of the biggest challenges in this work, was acquiring data to perform the analysis. Fortunately, several collaborators kindly shared datasets, datasets were downloaded from a multitude of websites, and other datasets were available through university licenses. Reducing the barrier to data accessibility will only improve the research that can be undertaken in the urban space.

To enable other researchers to use the results of my analysis, I am making the results of this work available using a digital mapping standard. This enables other researchers and planners to directly incorporate the results of my analysis into their work using a web-browser to visualize and intuitively learn about the analysis, or to connect directly to the data using GIS software.

While analyzing and providing urban resource intensity measures at a global scale is challenging, there are examples of projects that provide global data at high-levels of resolution (OpenStreetMap, 2012; Patterson and Kelso, 2012; Center for International Earth Science Information Network, 2012). These projects are feasible due to technological advances in analysis methods, data availability and data-gathering processes; the success of these projects motivates me to examine cities at a global level. I believe that there is a need for similar types of datasets that focus exclusively on urban resource usage patterns and my long-term goal is to facilitate the provision of international datasets that describe urban areas at a high level of resolution.

8.4 Future Work

Over the course of this research, I have developed and refined standardized analysis methods that characterize neighborhoods from many cities. Future work can explore the application of these processes to more urban areas, particularly in non-OECD countries, which are likely to have even less data available. It is of even greater relevance to apply these analytical methods to cities that are rapidly urbanizing.

My goal is to extend the analysis that I have developed during this research to create a freely accessible web-based tool that will provide quantitative analyses of material and energy use in neighborhoods around the world. This tool can help urban planners make informed decisions about development patterns as they consider material and energy requirements of urban patterns. This analysis will facilitate the preliminary assessment of resource consumption in urban area around the world. In the long-term this analysis can illustrate how access to resource intensity information can result in policy-changes.

As policy-makers gain a better understanding of how this data can be used, resource consumption measures can become a more accepted metric for assessing the sustainability of cities. These measurements could also be incorporated into scenario generation tools so that policy makers can understand the future energy and material demand at a high spatial resolution.

However, the parametrization and estimation of physical characteristics of the built environment is just one step in the overall analysis process of cities. It is important to consider more complex interactions in the urban space. Research that examines the urban heat island effect, considers renewable energy potential with regard to urban form and considers urban airflow patterns, all use aggregated urban form measures. The estimates of urban form parameters from this work could be directly incorporated into more complex urban models. In addition, building level energy models could incorporate the geometric estimates of the built form, to estimate energy consumption.

Finally, future work could benefit from remote-sensing data that can be used to identify the types of materials used for road construction, and validate the roadbed area predictions. This data is currently lacking in national US databases, so identifying the road type would be useful in calculations of albedo, and identifying impermeable surfaces. The predicted amount of material in roads could also be validated using remote sensing. The *RasClass* package (Wiesmann and Quinn, 2011) is one method that could be used to improve the assumptions about the materials used in building infrastructure.

8.5 Concluding Comments

I believe that there is an urgent need for simple analytic tools that can examine the physical structure of an urban area and provide some quantitative descriptions about the resource intensity of that area. This approach can help urban planners make informed decisions about development patterns as they consider the material and energy requirements of urban patterns while working in data scarce environments.

This work has sought to identify relationships that describe the emergent patterns that occur in cities. I have identified upper and lower ranges of behavior for some geometric measures, and discussed how these can be related to material and energy measurements. In addition, I have developed and validated models to estimate the road-bed area, the residential building area and the average building height per block group.

The potential impacts of this work are two-fold. The knowledge acquired can be used in urban design by helping to improve design guidelines for new sustainable neighborhoods; in addition it can contribute to policy development, by identifying areas relevant for regulation for improved urban resource efficiency.

One of the objectives of this work was to democratize the access to data, that is related to urban sustainability. The provision of a web-tool enables non-specialists to explore urban patterns. The web GIS capability enables specialists to access the underlying data.

As policy-makers gain a better understanding of how this data can be used, I hope that resource consumption measures will become a more accepted metric for assessing the sustainability of cities globally. I believe that this work demonstrates how researchers can use high-resolution spatial data and I hope that it will encourage the release of aggregated high-resolution data from cities.

Appendices

A Data

A.1 Spatial Data

A collection of data has been gathered for other cities listed in Table A.1. This collection of data contains six million building footprints, but one of the key measurements missing is building height data. This data is in addition to the national level data for the 40 US cities that are analyzed in this work.

Continent	Country	City	Footprints	Height	Roads	Census	Parcel Data	Boundaries
Europe	UK	London	1000000	Y	Y	Y	Y	Y
		Manchester	375000	Y	Y	Y	N	Y
	France	Paris	OSM	-	Y	?	-	Y
		Toulouse	-	-	Y	-	-	Y
		Lyon	-	-	Y	-	-	Y
	Germany	Berlin	-	-	Y	-	-	Y
		Leipzig	-	-	Y	-	-	Y
		Stuttgart	-	-	Y	-	-	Y
		Gelsenkirchen	-	-	Y	-	-	Y
	Spain	Madrid	-	-	Y	-	-	Y
		Barcelona	-	-	Y	-	-	Y
	Denmark	Copenhagen	?	?	Y	-	?	Y
	Switzerland	Zurich	?	?	Y	-	?	Y
		Bern	-	-	Y	-	-	Y
	Portugal	Lisbon	Y	Y	Y	-	-	Y
North America	USA	Boston	160000	Y	Y	Y	Y	Y
		New York	1000000	?	Y	Y	Y	Y
		Chicago	800000	N	Y	Y	Y	Y
		Los Angeles	10000	Y	Y	Y	Y	Y
		Seattle	280000	N	Y	Y	Y	Y
		San Francisco	-	-	Y	Y	-	Y
		Miami	-	-	Y	Y	-	Y
		Atlanta	160000	N	Y	Y	-	Y
		Bloomington	40000	N	Y	Y	-	Y
		Spokane	138000	N	Y	Y	-	Y
		Kitsap	120000	Y	Y	Y	-	Y
		Park City	5000	N	Y	Y	-	Y
		Philadelphia	400000	N	Y	Y	Y	Y
		PA - 6 Counties	2400000	N	Y	Y	-	Y
South America	Brazil	Sao Paulo	?	?	Y	-	?	Y
	Chile	Santiago	-	-	Y	-	-	Y
Asia	Singapore	Singapore	-	?	Y	?	?	Y
Africa	Kenya	Nairobi	?	N	Y	?	?	Y

Table A.1: Data for each city; 'Y' = downloaded; 'N' = not available; '-' = unknown; '?' = awaiting response; OSM = OpenStreetMap

Data	Source	Year	Webpage
Population Data	US Census	2000	http://www.census.gov
Road Data	US Census	2009	ftp://ftp2.census.gov/geo/tiger
Metro Areas	US Census	2000	http://www.census.gov/geo/www/cob/bdy_files.html
City Data	Boston City GIS	2009	http://www.cityofboston.gov/maps/
City Data	Cambridge City GIS	2009	http://www.cambridgema.gov/gis.aspx
City Data	New York City GIS	2009	https://nycopendata.socrata.com/
City Data	Chicago City GIS	2009	https://data.cityofchicago.org/
State Data	Massachusetts GIS	2009	http://www.mass.gov/mgis/download.htm

Table A.2: Data sources used in Chapter 4.

A.2 *Classification References*

A.2.1 *NAICS*

The North American Industry Classification System was used in Table A.3 to identify services from LEED-ND.

Table A.3: NAICS codes used to identify services that match LEED-ND criteria.

Service	**NAICS Code**
Bank	521110, 522110
Child care facility	624410
Community/Civic center	624110, 624120
Convenience store	445120,447110, 447110, 447190
Hair Care	812112
Hardware store	444130
Health club or outdoor recreation facility	713940
Laundry/Dry Cleaner	81231, 81232
Library	519120
Medical/Dental office	621111, 621112
Pharmacy (stand-alone)	446110
Place of worship	813110
Police/Fire station	922160, 922120
Post-Office	491110
Restaurant	722110
School	611110
Senior care facility	624120
Supermarket	445110
Theater	512131

A.2.2 *Road Nomenclature*

http://www.census.gov/geo/www/tiger/cfcc_to_mtfcc.txt

Primary roads - MTFCC = S1100: Primary roads are generally divided, limited-access highways within the interstate highway system or under state management, and are distinguished by the presence of interchanges. These highways are accessible by ramps and may include some toll highways.

Secondary roads - MTFCC = S1200: Secondary roads are main arteries, usually in the U.S. Highway, State Highway or County Highway system. These roads have one or more lanes of traffic in each direction, may or may not be divided, and usually have at-grade intersections with many other roads and driveways. They often have both a local name and a route number.

Local Neighborhood Road, Rural Road, City Street - MTFCC = S1400: Generally a paved non-arterial street, road, or byway that usually has a single lane of traffic in each direction. Roads in this feature class may be privately or publicly maintained. Scenic park roads would be included in this feature class, as would (depending on the region of the country) some unpaved roads.

B ArcGIS Plugin

B.1 ArcGIS: Population and Road Density Gradient Analysis Tool

A brief description of the plugin developed to perform population density and road density gradient calculations is provided in the following two sections.

B.1.1 Population Density Gradient

The population density gradient is calculated by creating concentric circles or rings, around the polygon representing the center of the city, and measuring the number of people in each ring. The data required to calculate the population density gradient is listed in Table B.1. This includes a list of the inputs and outputs with a description of the filetype.[1]

[1] A shapefile is a commonly used geospatial vector data format. Each vector in the shape-file can have numerical or categorical attributes associated with it.

	Data Type	Description	Required
Inputs	shapefile	polygon of city center	Y
	shapefile	polygon, with population measure	Y
	numeric	radius interval	N
	numeric	maximum radius	N
Outputs	.txt	population density, at radius intervals	-

Table B.1: Structure of population density gradient calculation

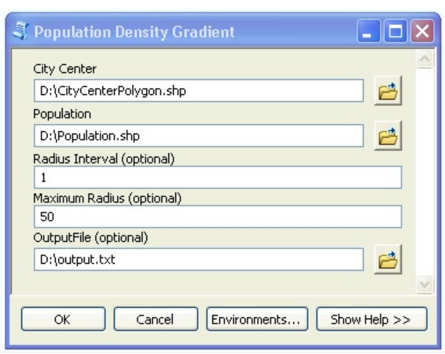


Figure B.1: Screenshot of population density gradient plugin

The structure of this analysis is described using pseudocode in Fig. B.2. The program loops through steps 1-4, until it has performed the calculation the required number of times, based on the user input value of the radius interval and maximum radius. If no value is provided by the user, for either the radius or maximum radius, default values of 1*km* and 50*km* are used. The user interface is shown in Fig. B.1. The population count is measured

Figure B.2: Pseudocode for population density gradient calculation.

```
Make Buffer of size DIST around Center Polygon
Clip Population count to Buffer
Calculate Population Count and Area
Write results to file
```

by the number of people in each block-group. To calculate the population density, this value is normalized by the area. This calculation assumes that people are uniformly distributed in each polygon.

B.1.2 Road Density Gradient

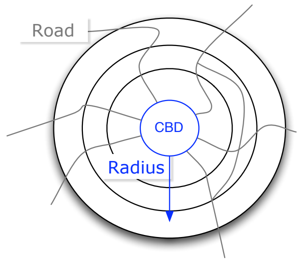


Figure B.3: Road density calculation (Figure repeated from Chapter 3).

The road density gradient is calculated in a similar way to the population density gradient (Section **??**). Concentric circles are made from the city center, and the length of road in each ring is recorded (Fig. B.3). The data required to calculate the road density gradient is listed in Table B.2.

	Data Type	Description	Required
Inputs	shapefile	polygon of city center	Y
	shapefile	vector, representing road network	Y
	numeric	radius interval	N
	numeric	maximum radius	N
Outputs	.txt	road density, at radius intervals	-

Table B.2: Structure of road density gradient calculation

```
Make Buffer of size DIST around center polygon
Clip road network to Buffer
Calculate Road Length and Area of buffer
Write results to file
```

Figure B.4: Pseudocode for road density gradient calculation.

The data requirements are also similar to Section **??**, with the exception of a vector file representing the road network, instead of polygons with population count attributes. The structure of the program is described using pseudocode in Fig. B.4, and the user interface is structured in a similar way to Fig. B.1.

C Spatial Analysis: Examples of Code

In this section, I list the details of the spatial analysis performed in this work. The majority of the spatial analysis undertaken in this research was performed using a *PostgreSQL* database with the spatial extension *PostGIS*. Both of these database technologies are open source. *ArcGIS* and *QGIS* were both used for various tasks, but the large-scale spatial was done using spatial databases, rather than desktop GIS programs.

C.1 Python Scripting for ArcGIS

ArcGIS has support for the *Python* programming language, enabling sets of commands to be written that make it easier to batch process analysis, or to perform tasks that the memory or file-format cannot process at once. While this is useful as it facilitates repeatable analyses processes to be developed, there are some difficulties when using ArcGIS for urban analysis.

Due to the proprietary nature of the underlying algorithms it is not possible to examine how the calculations are being performed, which can be problematic. For example, one problem that emerged during this work, is the fact that *ArcGIS* does not document how a nearest-neighbor calculation is being performed. As a result, through trial and error it was observed that the distance calculated was a centroid nearest-neighbor calculation. For more complex spatial analysis, it is unsatisfactory for the calculation process not to be clearly documented.

A second problem that emerged was due to the size of the datasets analyzed in this work. *ArcGIS* has many limitations on how data can be stored, and for large datasets involving table merging it is extremely slow. While an effort was made initially to write code loops that iterated through datasets that were split up into small pieces, this process was tedious and error prone. Instead, *PostGIS* was used for the majority of the spatial analysis that used vectors.

C.2 Spatial Database Analysis

Due to the size of the datasets analyzed conventional desktop GIS tools were unable to handle the data. Spatial databases were used, which facilitate the storing of large amounts of data, both spatial and non-spatial and performing operations on the data. *PostgreSQL* was using with the spatial extension *PostGIS 2.0*. This enables spatial queries to be written in the *SQL* syntax. Examples of some queries are shown in Figures C.2, C.7 with the underlying logic of the queries written in pseudocode in Figures C.1 and C.6. Two commonly used processes in this work were *spatial intersections* and *spatial joins*.

Figure C.1: Pseudocode for spatial query to identify all the local roads that are within a block-group of the census.

```
-- Make a new table to store the results
-- Identify roads that fall within a polygon border
-- Limit selection to roads that have the attribute 'MTFCC' = 1400
-- Calculate the total length of roads that fall into this polygon area
-- Store these road lengths in the new table
```

A *spatial intersection* is illustrated in Figure C.1 with the *SQL* code shown in Figure C.2. The variable referred to as *geom* is the default name of the *geometry* column stored in the database. Geometries are stored using a binary representation and can be accessed using various functions and commands. In general, these procedures were significantly quicker than the equivalent performed using shapefiles or geodatabases.

Figure C.2: SQL code example illustrating spatial query.

```
CREATE TABLE local_roads_per_blockgroup as
        SELECT
                SUM(ST_Length(ST_Intersection(b.geom, r.geom))),
                b.block_id
        FROM
                block_groups as b,
                usa_roads as r
        WHERE
                ST_Intersects(b.geom, r.geom)
        AND
                r.mtfcc = 'S1400'
        GROUP BY
                b.block_id;
```

Figure C.3: SQL code example illustrating how the number of roads that made an intersection were recorded.

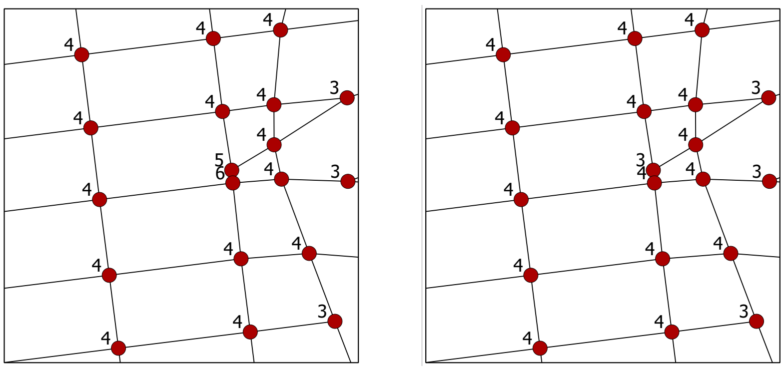
```
CREATE TABLE ny_roads_split_no_dups_i_count_touches as
SELECT
    ST_Intersection(a.geom, b.geom),
    Count(Distinct a.gid)
FROM
    dup_lines as a,
    dup_lines as b
WHERE
    ST_Touches(a.geom, b.geom)
AND
    a.gid != b.gid
GROUP BY
   ST_Intersection(a.geom, b.geom)
```

```sql
-- make index on table
CREATE INDEX idx_roads_dup ON dup_lines USING gist (geom);

-- delete duplicate lines
delete from dup_lines
        where gid IN (
             select mt1.gid
             from dup_lines mt1, dup_lines mt2
             where st_equals(mt1.geom, mt2.geom)
             and mt1.gid < mt2.gid
        )
```

Figure C.4: SQL code to remove duplicate lines.

When performing road network analysis, after splitting the road multi-lines into separate pieces, it was necessary to remove duplicates. The code to remove duplicate geometers is shown in Figure C.4. The intersection count before and after duplicates are removed is illustrated in Figure C.5.



(a) road duplicates (b) road duplicates removed

Figure C.5: Connection count of road network for a sample of *New York*.

```sql
-- Make a new table to store the results
-- Identify polygons that fall within a certain area
-- Store the polygons that are inside each zip-code polygon, with zip-code
```

Figure C.6: Pseudocode for spatial query to identify all the local roads that are within a block-group of the census.

In Figure C.2 the *mtfcc = 'S1400'* is included so only roads that are classified as *Local* are considered. Similarly the *spatial join* described in Figure C.6 is implemented using the code shown in Figure C.7. This query illustrates how a series of parcels from a tax assessment database can be related to each other.

See Appendix A.2.2 for a more detailed description of road types

An example of a spatial join is shown in Figure C.7. This illustrates how a set of tax assessment parcels can be assigned an attribute from the polygon that they fall within.

The *SQL* code for a multiband spatial query is shown in Figure C.8. This was used as part of the urbmet.org analysis process when a user examined a particular area.

Figure C.7: SQL code example illustrating spatial join.

```
CREATE TABLE
        summary_zip AS
SELECT
        b.*,                                        -- get all the values of 'b'
        z.zipcode                                   -- get zip codes
FROM
        tax_parcels as b
JOIN
        zip_codes as z
ON
        ST_Contains(z.the_geom, b.the_geom)         -- polygon intersection clause
```

Figure C.8: SQL code illustrating multi-band raster query. *POLYGON* is assumed to be a valid polygon that intersects the raster for a non-null result to be returned.

```
SELECT
        band,
        SUM((stats).sum) as sum,
        SUM((stats).count) as count,
        AVG((stats).mean) as avg
FROM (
        SELECT
          band,
          ST_SummaryStats(ST_Clip(r.rast, band, POLYGON, NULL, TRUE)) as stats
        FROM
          generate_series(1,10) As band,
          urbmet.raster_data AS r
        WHERE
          ST_Intersects(r.rast, POLYGON)
        ) AS foo
GROUP BY
  band
ORDER BY
  band
```

D Survey Data

Based on the survey conducted for Section **??**, users were given the opportunity to provide feedback on functionality changes and usability problems. All of the comments submitted are included in Table D.

Table D.1: Survey feedback and suggested functionality changes.

User	Functionality	Improvement
4	Can you change the size of the rectangle ? (area you're analyzing?) and can you rotate it?	I don't know how you can calculate the road length-/person with this. Also maybe make it more obvious that when you mouse over the bar graphs you get more information. It took me a while to discover it.. maybe different color parts of the bars?
6		It would be nice if you could export the base data for data analysis.
8		Didn't get the map to work.
11	Water Use	
13		Have the option of Regenerating the heat map automatically when the map is moved.
14	If I think of something I'll tell you ;)	What is the draw function for? When I tried to use it to analyze the polygonon I had drawn it went back to the rectangular shape and used that instead.
15	the ability of dragging and changing the frame of analysis zone by hand	
17	Help help help. This comes from a first time user with no experience in the field of course, but I didn't really understand many of the abbreviations. Also, I didn't really pay attention at the numbers (pop density, etc) because they are very small and in the bottom part. If this tool is purely for research purposes: - I think it is great as it is - I love how specific areas can be hand-drawn, and i think the metrics are fairly specific - I would like to be able to set my own limits for the color-coding. In particular, I would like to have a function that resets the color coding for the limits of the area I'm looking at. If you're trying to educate people about what matters, there's still a long way to go. Since urban metabolism is not my field, I didn't feel the tool was actively helping me to learn much. I didn't get a feeling for what numbers really matter. (on a side note, I *loved* the layout and cleanliness of this survey)	Maybe I should have take the demo, but it was hard for me to grasp what I was looking at exactly. I found it confusing that color coding was red for low values. I understand that red means "bad", but it took me a while to adapt to that (I couldn't understand why streets were red in population density...).

21	Yes, but too difficult to describe. Will write something up later for developers.	Yes, but too difficult to describe. Will write something up later for developers.
24	Include rural data.	Allow comparisons of areas of different sizes
25		water usage, water pollution
28	various modes of transportation? water consumption would be very interesting.	material is a vague term to me... but I chose to explore without a demo, perhaps some examples or definition is given there.
29		Had a bit of problem with selecting exactly the place I needed, but that is no problem if you get used to it. Overall, Fantastic!
31	I think that this has a lot of potential. However I think the the terms used, the graphs, colors denoting density, etc. need to really be clarified. Maybe I didn't spend enough time using the demo, but it just didn't seem very clear...maybe I'm just slow, but I'd like to think that with my background I would have been able to understand it more easily.	I minored in urban planning, did an independent research project on city planning, have worked as a landscape architect, and in my previous position used GIS software for the majority of my work. However the visualizer tool still was confusing to me. What does the "normalizing" option on the left hand side mean? And I know the rest of the world uses the metric system, but all of my professional work has dealt with imperial units. When you initially draw a box around a neighborhood, and the areas are graphed by color, I found it confusing. I believe there was a legend in the lower right corner - the units were 1000/km squared? 1000 what? It is not immediately clear. Also I believe that the higher end of this spectrum was colored green, but I could be mistaken. It would be more intuitive if the higher numbers i.e. higher density, higher energy usage, whatever, were red and descended to green. Also when I generated a PDF report, I found it difficult to understand units on the graphs and what I was comparing exactly. I was also confused by some of the terminology - the term "residential materials" or "building materials" or whichever was used, I can't remember exactly, was not clear or at least those aren't terms I've heard used in my field. Does that include hardscape - sidewalks, park lots?
32	Water consumption would be an interesting addition, provided you can find the data. Energy intensity per building floor area would be extremely useful, though collecting the data would be a significant undertaking (combining sources such as Zillow for for-sale housing, REIS for rental housing, CoStar for office, etc. would yield this, though it would be costly/time consuming).	

35	I wonder if it could expand into water and food availability, as well as waste - liquid, trash, etc... When will data be available for other countries?	Sorry to complain - about the tool and the survey but... it is not very user-friendly. It is not well-explained. Even as someone technically inclined, it was not straightforward. As for the survey, I'm an engineer and architect, student and professional - across those two and other fields. Shouldn't we be allowed to check all that apply? Furthermore, I cannot say that this too helped me to improve my knowledge of the things you are surveying... in part because it's not clear how to use it. For instance, when it takes the density of an area, is it the area that you drew or the box as the case may be... or is it the city in question? That said, I like the data that you are collecting and trying t present. I will use this in my work - as a student, professional and teacher. Thank you.
37	a. Batch processing across a number of analysis areas and auto plot generation across a trajectory of points? b. Historical trends?	a. Maybe allow for comparison of more than two locations. b. I understood that "heatmap" referred to the "filled contour plot" of population density or normalized energy consumption. This name might be misconstrued as heat output. I've not seen this name for filled contour plots, though I'd believe it is a standard term in some fields.
40	There's lots of potential in the tool and it's very well design and intuitive. Imagine more information about water consumption and maybe pollution.	Difficulty in analyzing the city patterns like the questions about if there's more roads in the city than in the suburbs are still unanswered for me.
41	A tool to establish comparison between areas.	
43		The coloring of the population/km is for the lower density red and the higher green. For the energy and material the coloring is the other way around, green is low and red is high. This is misleading.
46		- popovers when a cursor is resting on the map, providing metrics for that particular point
50		After exploring the tool, I still didn't feel like I knew how to answer the questions relating roads to population.
51	an option to use miles instead of km	
52		I think it would be nice to be able to pull up the search bar at any time, and I would have loved to play with the demos as well as the explore mode, but I couldn't get back to one after I'd clicked on the other.
55		It's not particularly intuitive. I didn't get much out of it. Several functions did not seem to work or took long enough to load that they seemed not to work.
58		There is a slight ambiguity in the city center as to whether we're dividing over the local resident population or the daytime population. I assume its the resident population, which really skews the resource intensity up, since in many cities the CBD has relatively few residents, and a lot of infrastructure.

59		make the graphs on the right more colorful and easier to read
60	Water use would be interesting to compare.	Make it navigable with arrow keys. Also, if you could save regions so you could compare multiple areas instead of just two at a time, that would be neat.
63	Renewable energy potential	no, it did crash on me once though
64	See above.	There was a lot of interesting information to process, so I may have to get back to you on this. One thought though: Is there a way to give the user the option to automatically highlight a specific region (the first thought was district/city/county limits)? It seems like this could make this a stronger analytical tool, rather than an educational piece.
65	The GIS portion didn't work too well on my Apple iOS device, other survey takers might also have trouble viewing the information on their mobile device.	I understand this is about GIS and it wants people to get a chance to utilize GIS as a resource, but at the very end it might help to inform the survey taker of the correct answers.
72	Include tendencies to have an idea how the population, energy consumption, etc, evolved in the area Include info on type of area: residential, industrial, green spaces...	It looks great! Great great job. Here are some thoughts on how to improve it: How to use the tool is still not straightforward. Clearly state that your are comparing areas: for instance use "compare" instead of analyze. Give an idea on the area of your selection while drawing. Give some reference on the map, like refer where the city center is: not all cities are easy to recognize from the top. Improve legends on the report. Include reference values on the report to have an idea of magnitude on the report: "kWh a coal plant produces", for example. This all depends on the audience you are aiming at.
74		If this is for an American audience, it would probably be more useful to use miles as units rather than kilometers. Also, at least in the demo, the scale doesn't show up in the window with the map, I have to scroll to see it. Some of the survey questions (like roads-miles/person) are probably not easily answerable based on the tool. Or at least I couldn't easily see how you might make such a calculation.
77		Not being in the field some terminology was not sufficiently clear. For instance what is "VKT due to auto". probably it is perfectly OK for the typical user.
78	do you show the % of impermeable surface/roads?	I was having trouble getting it to do what i wanted-
80	It would be nice to have a "show me cities like this" tool	Found a small typo in the tool. On the draw option, the instruction should read "Double click to finish" (not finnish).
		Add a benchmark indiction in the plots (e.g. US average values)
82		The report should show both areas that I compared, instead of only showing the last area selected. This will help me understand the differences between the two areas.

84	identification of open green space for potential productive use.	
85	Instead of having to draw outlines every time, it would be cool to have a drop-down list of neighborhoods. i.e. in NYC, I could choose "Brooklyn Heights", "Financial District", etc. While the organization and presentation quantitative data is just flat-out amazing, I reckon a lot of people will want to use this to make qualitative assessments. i.e. Is my neighborhood better than the national average? Than the city average? Is living in Boston more sustainable than than in SF? Perhaps you can figure out a way to explicitly tell them, and make this a snappy way to solve arguments or give policy-makers a quick synopsis of a detailed, messy issue. I'm pretty sure your neighborhood histograms do this, but I don't see what the input sample is. Is this showing me a dist. over neighborhoods in the country? the city?	I wasn't sure how to interpret your kWh/HH numbers. Is this annual usage, monthly usage..? Perhaps a popup with a layman's description when we hover over the units would help non-tech folks
87		Please understand while the following is critical, I'm very impressed with the data you're able to pull very easily. I think it's a great framework that needs work on the information presentation side. Metric comparisons are not very easy to interpret for Americans (sad but true) While the information is rich, the presentation could use a lot of work. e.g. It's unclear why the weight of the roads is more important than the length for analysis purposes... How many miles of road per person seems more relevant, even if the ratio is the same it's more meaningful. The "Neighborhoods" graph at the bottom of the report is very difficult to understand. Does the red line show my neighborhood while the histograms refer to a reference case? That should be stated. I would strongly recommend a second page that describes how to interpret the results and show some basis for comparison (e.g. your results vs. US avg reference vs. European reference). is 7175 kg per person a lot for building material consumption or not? I'm not sure and I work in facility design and construction!
90	always show a benchmark (avg. or max, etc). At every scale. People are just not familiar with metrics.. they dont have an order of magnitude of what is reasonable.	Include the distribution charts shown during the BT lab presentation, to allow users to identified were are they compared to the whole city or other cities for a given metric.
93		I didn't explore that much. When I have the chance to use it with a certain purpose I will be able to make some useful suggestions.
96	The report function is great. I'd like to see it include an analysis of both selected areas and a comparison of them. Also, in the reports, it'd be great to have commas at the '000s place.	

97	Further zoom and/or street labels.	I didn't understand the Material per person aspect. I also could have used some help finding the city center- even just being able to zoom in further, or having street labels would have helped with this.
98		yayy! nice work, gents! the survey format was so clear to follow. the one piece that'd be helpful is explaining what the different colors mean on each of the maps. granted I'm probably not your core audience so you can ignore that. but it'd help to see a few more hand holding on how to best utilize the tool.
100		No suggestions, this tool isn't applicable to my work.
110		A video showing practical example of problem (question asked) and solution (analysis of data) could be helpful.
111	More information on a per home basis- breakdown of where different information is coming from right as you see it being developed, the capability to switch the bar graphs you see right away without having to generate the report for more detailed information.	I would absolutely love to have more granular home info- but I know it isn't out there- I could maybe help add it in there! My ability to learn dynamically was primarily limited to the bar graphs I saw as I moved along- in the demo you might want to show someone a good "city" example and a good "suburb" example to get them started on the information. The questions you ask relate to some metric that can only be found when you generate the pdf- some might not get to this step.
115		I could not actually get the tool to work for my city.
116		Define "infrastructure" and "energy" in the legend.
119	It would be wonderful to be able to look at this data longitudinally. Could you recreate the model in ten year intervals looking back at data from the last 50 years? This would be valuable for researchers, such as myself, wanting insights on how changes in the built environment have changed energy use patterns. Data on energy use in commercial buildings is an essential next step. Energy use in industry would also be nice.	The ability to generate an Excel-based report with numerical values that could be compared across cities would be valuable for those wanting analyze the results from your tool.
120		The colour scale of the number of habitants should be inversed. Red lists as low density, but is usually associated with high values.

E Bibliography

Aberdeenshire Council (2008, May). Transportation and infrastructure: Standards for road construction consent and adoption.

Alfeld, L. (1995). Urban dynamics-the first fifty years. *System Dynamics Review 11*(3), 199–218.

Alfeld, L. and A. Graham (1976). *Introduction to Urban Dynamics*. Wright-Allen Press Cambridge, MA.

Andrews, C. J. (2000). Features - Restoring Legitimacy to the Systems Approach - Past misapplications of the systems approach in the policy arena prompt us to rethink the the role of systems concepts. *IEEE technology & society magazine. 19*(4), 38.

Angel, S. (2008, June). An arterial grid of dirt roads. *Cities 25*(3), 146–162.

Arnstein, S. R. (1969). A Ladder of Citizen Participation. *Journal of the American Planning Association 35*(4), 216–224.

Banister, D. (2008, March). The sustainable mobility paradigm. *Transport Policy 15*(2), 73–80.

Batty, M. (2005). *Cities and complexity : Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. Cambridge, Mass.: MIT Press.

Batty, M. (2008). The Size, Scale, and Shape of Cities. *Science 319*(5864), 769.

Bento, A., M. Cropper, A. Mobarak, and K. Vinha (2005). The effects of urban spatial structure on travel demand in the united states. *Review of Economics and Statistics 87*(3), 466–478.

Berghauser-Pont, M. (2010). *Spacematrix*. Rotterdam: NAi Publishers.

Bertraud, A. (2004). The spatial organization of cities: Deliberate outcome or unforeseen consequence?

Bettencourt, L. M. A., J. Lobo, D. Helbing, C. Kuhnert, and G. B. West (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America. 104*(17), 7301–7306.

Bishop, I. D. (1998). Planning support: hardware and software in search of a system. *Computers, Environment and Urban Systems 22*(3), 189 – 202.

Boarnet, M. and R. Crane (2001). *Travel by design: The influence of urban form on travel*. Oxford University Press, USA.

Bouman, M., R. Heijungs, E. van der Voet, J. C. J. M. van den Bergh, and G. Huppes (2000). Material Flows and Economic Models: An Analytical Comparison of SFA, LCA and Partial Equilibrium Models. *Ecological Economics : The Journal of the International Society for Ecological Economics. 32*(2), 195 – 216.

Brail, R. (2008). *Planning support systems for cities and regions*. Cambridge Mass.: Lincoln Institute of Land Policy.

Brand, F. (2009). Critical natural capital revisited: Ecological resilience and sustainable development. *Ecological Economics 68*(3), 605–612.

Brand, S. (1995). *How buildings learn: what happens after they're built*. Penguin Books.

Brand, S. (2006). City Planet. Technical report, www.strategy-business.com.

Cambridgeshire City Council (2007, October). Cambridgeshire design guide: For streets and public realm.

Center for International Earth Science Information Network (2012, January). Gridded Population of the World.

Cervero, R. and K. Kockelman (1997, September). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment 2*(3), 199–219.

Chudley, R. and R. Greeno (2008). *Building construction handbook* (7th ed. ed.). Oxford: Butterworth-Heinemann.

Clark, C. (1951). Urban population densities. *Journal of the Royal Statistical Society 114*, 490–496.

Clifton, K., R. Ewing, G. Knaap, and Y. Song (2008). Quantitative analysis of urban form: a multidisciplinary review. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability 1*(1), 17.

Coelho, D. and M. Ruth (2006). Seeking a unified urban systems theory. In *The Sustainable City IV: Urban Regeneration and Sustainability*, Tallin, Estonia, pp. 179–188.

Cortright, J. (2009, August). Walking the walk: How walkability raises home values in u.s. cities. electronic.

Costanza, R. and M. Ruth (1998). Using dynamic modeling to scope environmental problems and build consensus. *Environmental management 22*(2), 183–195.

Crane, R. (1996). On form versus function: will the new urbanism reduce traffic, or increase it? *Journal of Planning Education and Research 15*(2), 117.

Crucitti, P., V. Latora, and S. Porta (1991). Centrality in networks of urban streets. *American Institute of Physics 16, No.1*.

Daniels, P. (2002). Approaches for Quantifying the Metabolism of Physical Economies: A Comparative Survey, Part II: Review of Individual Approaches. *Journal of Industrial Ecology 6*(1), 65–88.

Daniels, P. and S. Moore (2001). Approaches for Quantifying the Metabolism of Physical Economies: Part I: Methodological Overview. *Journal of Industrial Ecology 5*(4), 69–93.

Davoudi, S., J. Crawford, and A. Mehmood (2009). *Planning for Climate Change: Strategies for Mitigation and Adaptation for Spatial Planners*. Earthscan/James & James.

Deal, B. (2001). Ecological urban dynamics: the convergence of spatial modelling and sustainability. *Building Research & Information 29*(5), 381.

Decker, E., A. Kerkhoff, and M. Moses (2007). Global patterns of city size distributions and their fundamental drivers. *PLoS ONE Issue 9*.

Decker, H., S. Elliott, F. A. Smith, D. R. Blake, and F. S. Rowland (2000). Energy and material flow through the urban ecosystem. *Annual Review of Energy and the Environment 25*, 685–740.

del Moral, R. and L. R. Walker (2007). *Environmental disasters, natural recovery and human responses*. Cambridge, UK; New York: Cambridge University Press.

Diamond, J. (2005). *Collapse*. Penguin Group USA.

Dill, J. (2004). Measuring network connectivity for bicycling and walking. In *Measuring Network Connectivity for Bicycling and Walking*, Measuring Network Connectivity for Bicycling and Walking. Transportation Research Board.

EIA (2012, March). New York - Quick Facts.

Emmi, P., C. Forster, and J. Mills (2005). Insights into the dynamics of a carbon-based metropolis. In *Proceedings of the 23rd International Conference of the Systems Dynamics Society, Boston*.

Erb, K., S. Gingrich, F. Krausmann, and H. Haberl (2008). Industrialization, Fossil Fuels, and the Transformation of Land Use: An Integrated Analysis of Carbon Flows in Austria 1830–2000. *Journal of Industrial Ecology, 12 5*(6), 686–703.

ESRI (2009). *ArcGIS* (10 ed.). ESRI, Redlands, California.: Environmental Systems Resource Institute.

ESRI (2011, February). ESRI News.

ESRI Business Analyst (2008). http://www.esri.com/software/bao.

Ewing, R. (1997). Is Los Angeles-style sprawl desirable? *Journal of the American Planning Association 63*(1), 107–126.

Ewing, R. and R. Cervero (2010). Travel and the Built Environment. *Journal of the American Planning Association 76*(3), 265–294.

Ewing, R. and F. Rong (2008). The impact of urban form on us residential energy use. *Housing Policy Debate 19*(1), 1–30.

Fast Company (2012). The Neighborhood Visualizer Maps The Resource Intensity Of Your City.

Fischer-Kowalski, M. (1998). Society's Metabolism. The Intellectual History of Materials Flow Analysis, Part I, 1860-1970. *Journal of Industrial Ecology 2*(1), 61–78.

Fischer-Kowalski, M. and W. Hüttler (1999). Society's Metabolism. The Intellectual History of Materials Flow Analysis, Part II, 1970–1998. *Journal of Industrial Ecology 2*(4), 107–136.

Folke, C., T. Hahn, P. Olsson, and J. Norberg (2005). Adaptive governance of social-ecological systems. *Annual Review of Environment and Resources 30*(1), 441–473.

for Communities, D. and L. Government (2010, October). English Housing Survey 2008 Housing Stock Report. This is the detailed report of findings relating to the housing stock from the new survey, and builds on results reported in the EHS Headline Report published in February 2010.

Forrester, J. (1969). *Urban Dynamics*. MIT Press.

Gabaix, X. (1999). Zipf's Law For Cities: An Explanation. *Quarterly Journal of Economics 114*(3), 739–767.

Gastner, M. and M. Newman (2006). Optimal design of spatial distribution networks. *Physical review.E, Statistical, nonlinear, and soft matter physics 74*(1).

Gelinne, D. (2011, November). Bicycle and Pedestrian Planning? There's an App for That!

Giradet, H. (1992). *The Gaia Atlas of Cities*. Gaia Books.

Glaeser, E. (1998). Are cities dying? *The Journal of Economic Perspectives 12*(2), 139–160.

Glaeser, E., J. Kolko, and A. Saiz (2001). Consumer city. *Journal of Economic Geography 1*(1), 27.

Goßling-Reisemann, S. (2008). What Is Resource Consumption and How Can It Be Measured? *Journal of Industrial Ecology 12*(1), 1088–1980.

Graham, M. (2012, April). The Problem With Wikidata. http://www.theatlantic.com/technology/archive/2012/04/the-problem-with-wikidata/255564/.

Güneralp, B. and K. C. Seto (2008, October). Environmental impacts of urban growth from an integrated dynamic perspective: A case study of shenzhen, south china. *Global Environmental Change 18*(4), 720–735.

Han, J., Y. Hayashi, X. Cao, and H. Imura (2009, 6). Application of an integrated system dynamics and cellular automata model for urban growth assessment: A case study of shanghai, china. *Landscape and Urban Planning 91*(3), 133–141.

Handy, S., X. Cao, and P. Mokhtarian (2005, November). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment 10*(6), 427–444.

Handy, S., X. Cao, and P. Mokhtarian (2006). Self-selection in the relationship between the built environment and walking: Empirical evidence from northern california. *Journal of the American Planning Association 72*(1), 55–74.

Hankey, S. and J. D. Marshall (2010, September). Impacts of urban form on future US passenger-vehicle greenhouse gas emissions. *Energy Policy 38*(9), 4880–4887.

Hanson, S. and G. Giuliano (2004). *The geography of urban transportation*. Guilford Press.

Hardin, G. (1968). The tragedy of the commons. *Science 162*, 1243–48.

Holtzclaw, J. (1994). Using residential patterns and transit to decrease auto dependence and costs. *Natural Resources Defense Council, San Francisco.*.

Horner, M. and A. Murray (2002). Excess commuting and the modifiable areal unit problem. *Urban Studies 39*(1), 131–139.

Howard, B., L. Parshall, J. Thompson, S. Hammer, J. Dickinson, and V. Modi (2012, February). Spatial distribution of urban building energy consumption by end use. *Energy and Buildings 45*(0), 141–151.

Hsu, J. (2012, April). Secret Computer Code Threatens Science: Scientific American.

Huang, S. and W. Hsu (2003, April). Materials flow analysis and emergy evaluation of Taipei's urban construction. *Landscape and Urban Planning 63*(2), 61–74.

Infosthetics (2012). Neighborhood Visualizer: Revealing Material and Energy Use in Cities.

Ingram, G. (1998). Patterns of metropolitan development: what have we learned? *Urban studies 35*(7), 1019–1035.

Jacobs, J. (1965). *The death and life of great American cities:[the failure of town planning]*. Penguin Books.

Johnson, S. (2010). *Where good ideas come from : the natural history of innovation*. New York: Riverhead Books.

Kennedy, C. (2007, April). Urban metabolism.

Kennedy, C., S. Pincetl, and P. Bunje (2011, August). The study of urban metabolism and its applications to urban planning and design. *Environmental Pollution 159*(8–9), 1965–1973.

Kennedy, C., J. Steinberger, B. Gasson, Y. Hansen, T. Hillman, M. Havránek, D. Pataki, A. Phdungsilp, A. Ramaswami, and G. Mendez (2009). Methodology for inventorying greenhouse gas emissions from global cities. *Energy Policy In Press, Corrected Proof*.

Kolbe, T., G. Gröger, and L. Plümer (2005). CityGML: Interoperable Access to 3D City Models. In *Geo-information for Disaster Management*, pp. 883–899.

Kovanda, J. and T. Hak (2008). Changes in Materials Use in Transition Economies. *Journal of Industrial Ecology 12*(5-6), 721–738.

Krasny, M. E. and K. G. Tidball (2009). Applying a resilience systems framework to urban environmental education. *Environmental Education Research 15*(4), 465.

Krausmann, F., M. Fischer-Kowalski, H. Schandl, and N. Eisenmenger (2008). The Global Sociometabolic Transition Past and Present Metabolic Profiles and Their Future Trajectories. *Journal of Industrial Ecology 12*(5-6), 637–656.

Krausmann, F., S. Gingrich, N. Eisenmenger, K. H. Erb, H. Haberl, and M. Fischer-Kowalski (2009). Growth in global materials use, GDP and population during the 20th century. *Ecological Economics 68*(10), 2696–2705.

Lamigueiro, O. P. and R. Hijmans (2011). rasterVis: 0.10.

Lee, D. (1973). Requiem for large-scale models. *Journal of the American Planning Association 39*(3), 163–178.

Lefèvre, B. and G. Mainguy (2010, April). Urban Transport Energy Consumption: Determinants and Strategies for its Reduction. *S.A.P.I.EN.S* (2.3).

Levinson, D. M. and K. J. Krizek (2008). *Planning for Place and Plexus : Metropolitan Land Use and Transport*. New York; London: Routledge.

London Borough of Croydon (2009). Design brief and specification for road and sewer works in new streets.

Lowry, I. S. (1964). *A model of metropolis*. Santa Monica, Calif.: Rand Corp.

Lucas, R. and E. Rossi-Hansberg (2002). On the internal structure of cities. *Econometrica 70*(4), 1445–1476.

Marshall, J. D. (2007). Urban land area and population growth: A new scaling relationship for metropolitan expansion. *Urban Studies 44, No. 10*, 1889–1904.

McGranahan, G. and D. Satterthwaite (2003). Urban Centers: An Assessment of Sustainability. *Annual Review of Environment and Resources 28*(1), 243–274.

McMorrough, J. (2006). *Materials, structures, and standards : all the details architects need to know but can never find*. Gloucester Mass.: Rockport Publishers.

Meadows, D. and A. AtKisson (1997). Delay times: How long does it take to respond?

Min, J., Z. Hausfather, and Q. F. Lin (2010, October). A HighâĂŘResolution Statistical Model of Residential Energy End Use Characteristics for the United States. *Journal of Industrial Ecology 14*(5), 791–807.

Mindali, O., A. Raveh, and I. Salomon (2004, February). Urban density and energy consumption: a new look at old statistics. *Transportation Research Part A: Policy and Practice 38*(2), 143–162.

Mokhtarian, P. and X. Cao (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B 42*(3), 204–228.

NAICS (2011, April). North American Industry Classification System.

National Association of Home Builders (U.S.) and Bank of America Home Equity (2007). *Study of life expectancy of home components*. [Washington, DC]; [S.l.]: National Association of Home Builders ; Bank of America.

Neuman, M. (2005, September). The Compact City Fallacy. *Journal of Planning Education and Research 25*(1), 11–26.

New York Times (2012, April). On Our Radar: Urban Energy Patterns.

Newman, P., T. Beatley, and H. Boyer (2009). *Resilient cities : responding to peak oil and climate change*. Washington, DC: Island Press.

Newman, P. and I. Jennings (2008). *Cities as Sustainable Ecosystems*. Island Press.

Newman, P. and J. Kenworthy (1989). *Cities and automobile dependence : a sourcebook*. Aldershot, Hants., England; Brookfield, Vt., USA: Gower Technical.

Newman, P. and J. Kenworthy (1991). *Cities and automobile dependence: a sourcebook*. Gower.

Newman, P. and J. R. Kenworthy (1999). *Sustainability and cities : overcoming automobile dependence*. Washington, D.C.: Island Press.

Niza, S. and P. Ferrão (2006). A transitional economy's metabolism: The case of Portugal. *Resources, Conservation and Recycling 46*(3), 265–280.

Niza, S., L. Rosado, and P. Ferrao (2009). Urban Metabolism: Methodological Advances in Urban Material Flow Accounting Based on the Lisbon Case Study. *Journal of Industrial Ecology 13*(3), 384–405.

Open Geospatial Consortium (2012a, January). OpenGIS Web Coverage Service (WCS) Implementation Specification.

Open Geospatial Consortium (2012b, January). OpenGIS Web Map Service (WMS) Implementation Specification.

OpenStreetMap (2012, January). Openstreetmap.org.

O'Reilly Media (2012). Visualizing cities' energy usage, population density, and material intensity.

Pataki, D. E., R. J. Alig, A. S. Fung, N. E. Golubiewski, C. A. Kennedy, E. G. McPherson, D. J. Nowak, R. V. Pouyat, and P. Romero Lankao (2006). Review: Urban Ecosystems and the North American Carbon Cycle. *Global Change Biology 12*(11), 2092–2102.

Patterson, T. and N. V. Kelso (2012, January). Natural Earth.

Pickett, S., M. Cadenasso, J. Grove, C. Nilon, R. Pouyat, W. Zipperer, and R. Costanza (2001). Urban Ecological Systems: Linking Terrestrial Ecological, Physical, and Socioeconomic Components of Metropolitan Areas. *Annual Review of Ecology and Systematics 32*(1), 127–157.

Pickett, S. T. A., M. L. Cadenasso, and J. M. Grove (2004, October). Resilient cities: meaning, models, and metaphor for integrating the ecological, socio-economic, and planning realms. *Landscape and Urban Planning 69*(4), 369–384.

Quantum GIS Development Team (2011). *Quantum GIS Geographic Information System* (1.6 ed.). Open Source Geospatial Foundation.

Quinn, D. J. and J. E. Fernández (2010, August). Estimating Material Usage of Road Infrastructure in US Cities. In *IPBSA SimBuild Proceedings*. SimBuild Conference.

Quinn, D. J. and J. E. Fernández (2011a, October). Relating Urban Form to Resource Consumption: Analysis and Visualization Methods. In *ACSP Conference*. ACSP.

Quinn, D. J. and J. E. Fernández (2011b, May). Standardized Analysis of Urban Form. In *Proceedings of the International Symposium on Sustainable Systems and Technology*. IEEE.

Quinn, D. J., D. Wiesmann, and J. J. Sarralde (2011, September). Estimating Resource Consumption Using Urban Typologies. In *CISBAT Proceedings*. EPFL.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ruth, M. and D. Coelho (2007). Understanding and managing the complexity of urban systems under climate change. *Climate Policy 7*, 317–336.

Samaniego, H. and M. Moses (2008). Cities as organisms: Allometric scaling of urban road networks. *Journal of Transport and Land Use 1*, 21–39.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York: Norton.

Schrank, D. and T. Lomax (2011). 2010 Annual Urban Mobility Report: Information on long-term congestion trends.

Schulz, N. (2007). The Direct Material Inputs into Singapore's Development. *Journal of Industrial Ecology 11*(2), 117–131.

Shin, Y., V. Vuchic, and E. Bruun (2009). Land consumption impacts of a transportation system on a city. *Transportation Research Record: Journal of the Transportation Research Board 2110*(-1), 69–77.

Song, Y. and G. Knaap (2004). Measuring urban form. *Journal of the American Planning Association 70*(2), 210–225.

Sterman, J. (2000). *Business Dynamics : Systems Thinking and Modeling for a Complex World*. Boston: Irwin/McGraw-Hill.

Stern, N. H. (2006). The Stern review on the economics of climate change. http://www.hm-treasury.gov.uk.

The Geoinformation Group Ltd. (2010, February). UK Map. Fulbourn, Cambridge, UK.

The Guardian (2012, April). Interactive: MIT researchers map energy use and building material intensity across US cities.

Thurrock Borough Council (2005, August). Housing estate road construction specification.

Tufte, E. (1982). *The Visual Display of Quantitative Information*. Cheshire CT: Graphics Press.

UK Ordnance Survey (2010a, March). OS MasterMap. http://edina.ac.uk/mastermap/.

UK Ordnance Survey (2010b, March). OS VectorMap District. https://www.ordnancesurvey.co.uk/.

United Nations (2010, March). World Urbanization Prospects The 2009 Revision. Booklet.

United Nations (2011). Growth projections 2010-2050.

US Census (2000). www.census.gov.

U.S. Census Bureau (2010). American Community Survey. http://www.census.gov/acs/www/.

US Census Bureau, D. I. S. American Housing Survey (AHS), AHS Main. http://www.census.gov/housing/ahs/. US Census Bureau information on the subject of American Housing Survey (AHS). Main section.

US Census Bureau, M. C. D. Characteristics of New Housing. http://www.census.gov/construction/chars/completed.html. US Census Bureau Characteristics of New Housing website.

USGBC (2009). Leadership in environmental and energy design - neighborhood development. electronic.

Vivier, J. (2000). Millenium cities database for sustainable transport : an innovative project in the service of UTIP members. *Public transport international. 49*(4).

Waddell, P. (2002). Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association 68*(3), 297–314.

Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke, and G. Ulfarsson (2003). Microsimulation of urban development and location choices: Design and implementation of urbansim. *Networks and Spatial Economics 3*(1), 43–67.

WalkScore (2009). Find a walkable place to live.

Wegener, M. (2004). Overview of land use transport models. *Handbook of transport geography and spatial 5*(3), 127–146.

Weisz, H. and H. Schandl (2008). Materials Use Across World Regions. *Journal of Industrial Ecology 12*(5/6), 629–636.

Weisz, H. and J. K. Steinberger (2010, August). Reducing energy and material flows in cities. *Current Opinion in Environmental Sustainability 2*(3), 185–192.

Wernick, I. and F. Irwin (2005). Material Flows Accounts: A Tool for Making Environmental Policy. *World Resources Institute*.

West, G. B., J. H. Brown, and B. J. Enquist (1997). A general model for the origin of allometric scaling laws in biology. *Science No. 5309-122*.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Number 978-0-387-98140-6. Springer New York.

Wiesmann, D. and D. J. Quinn (2011, September). *rasclass*: Supervised Raster Image Classification.

Wolman, A. (1969). The metabolism of cities. *Scientific American*.

Worcestershire City Council (2011, March). Highways specification.

Wu, F. (2007). Book Review: Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals. *Progress in Human Geography 31*(1), 113–115.

Zegras, C. (2010). The built environment and motor vehicle ownership and use: evidence from Santiago de Chile. *Urban Studies*.

Zegras, P. C. (2007, October). As if Kyoto mattered: The clean development mechanism and transportation. *Energy Policy 35*(10), 5136–5150.

Zhang, M. (2004). The role of land use in travel mode choice: Evidence from Boston and Hong Kong. *Journal of the American Planning Association 70*(3), 344–360.