Standardized Analysis of Urban Form

David J Quinn* and John E Fernández

***Abstract*—Despite the availability of detailed data for describing cities, there is a shortage of tools that enable standardized analyses of urban areas, and the associated resource consumption. The objective of this research is to contribute to the development of a standardized method for analysis, so that material and energy consumption can be quantified. Analyzing the resource consumption of the built environment is particularly relevant in cities that are rapidly growing, as short-term planning decisions will have long-term consequences for both the quality of life of the inhabitants, and the future energy and material use of the urban area.**

***Index Terms*—urbanization, resource usage, material, energy, geographic information systems**

I. Introduction

WHILE there has been a substantial amount of work published that examines the form of cities at a high-level [1]–[3], less work has been done examining the internal structure of cities in a way that enables comparison. Prior research has not focused on the functioning of the city with regard to resource consumption. This is partially due to inconsistencies in city-level datasets, and a shortage of available tools to easily perform this type of work. There is a need for simple analytical tools that can examine the physical structure of an urban area in a repeatable way, and provide some quantitative descriptions about the performance of that area.

Objective

In this paper a method of spatial analysis is developed and applied to a case study of Atlanta, Georgia. The objective of developing this approach is to facilitate easily repeatable analyses of cities. By using the same initial assumptions in each case, a comparison of the performance can be made. Spatial data which describes the location of people, roads and services are used to identify general measurements of the urban form. The analysis is implemented using a plugin developed for ArcGIS written in the Python programming language, enabling the methodology to be applied to any comparable dataset. This analytical approach tries to overcome the systematic problem that exists in many current urban analyses, where the results are not replicable, and the underlying assumptions are unclear.

Previous Work

Previous work analyzing cities has typically been done from a high-level perspective. The identification of population density gradients has been observed since 1951 [4], and more recent work has examined scaling patterns and internal mechanisms that result in particular forms of urban growth [5], [6]. While many analyses of groupings of cities consider the city as a homogenous entity [2], [3], some research has examined parameters that describe the physical form within the city [7]–[9]. Although there is limited policy relevance at the level of abstraction where cities are considered as part of a group, parameter ranges within cities can be used for purposes of comparison.

Much of the research to-date has focused on the *form* of the city (density, spatial configuration, physical form measurements), rather than measures of how the city *functions*, which considers where people go and how much energy they use due to their transportation mode. There is a need for research that focuses on quantifying the performance of cities, with regard to resource efficiency, as this has not been explored thoroughly in the published literature. Newman and Kenworthy [10] were amongst the first to observe a relationship between transportation energy and population density; their work identified that cities with higher densities used less energy for transportation per capita than cities with lower densities. However, city-boundary issues and city homogeneity are not adequately explained in this study, as the results are based on average values for each city. More recent work has examined relationships between urban form and transportation energy [11], [12], while the effects of New Urbanism design strategies and travel behaviors are linked to resource efficiency [13].

II. Geospatial and Statistical Analysis

In this section a brief description of the methods used to calculate the population density gradient, the road density gradient and distance to service gradient are described. Here, a gradient is a measure of the rate of change of the parameter of interest, with respect to a radial distance from the city center. These gradient measures can be related to the material required to provide the infrastructure, and the energy required for the daily functioning of the city. 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) when compared to the surrounding area. A 3D representation of the population density of Atlanta is shown in Fig. 1, where the population density gradient is apparent. The identification of this central area was the first step in this work, and the area with the highest business density was considered to signify the center of the city.

David J Quinn and Prof. John E. Fernández are with the Building Technology Group of the MIT Department of Architecture, Cambridge, MA 02139, USA.

*Corresponding author: djq@mit.edu

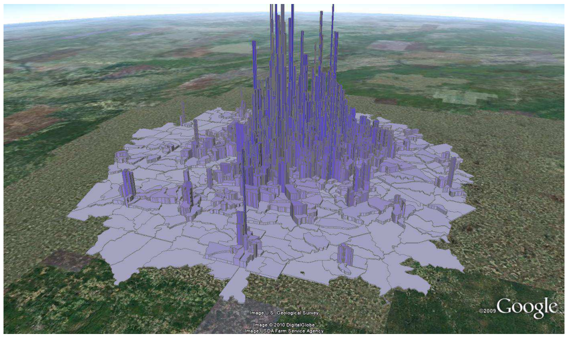

Fig. 1: 3D representation of Atlanta's population density

This study was conducted using *ArcGIS* for spatial analysis [14] and the open-source program *R*, for statistical analysis [15]. A plugin was developed for *ArcGIS*, written in the Python programming language and this plugin was used to analyze the geospatial data. A future objective of this work is to revise these scripts for *QGIS* [16], an open-source geographic information system, so that no proprietary software is necessary for the analysis. This paper explains in detail how these calculations were performed and provides the plugin that was developed for this work [17]. To illustrate this process, the city of Atlanta is analyzed using this plugin and the results are presented in Section III. The raw and processed data used in this analysis are also available for download [17].[1]

TABLE I: Data sources used in this analysis

Data	Source
Population Data	US Census (2000)
Road Data	US Census (2009)
Service Location	ESRI Business Analyst (2009)

Statistical analysis was used to examine the results from the geospatial analysis to identify patterns. *R* was used to read the output *.txt* files from the analysis which were explored graphically and numerically. Appropriate curves were fitted to the data using regression and these were plotted using the *R* package *ggplot2*.

A. 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 II. This includes a list of the inputs and outputs with a description of the filetype.[2]

[1]Business location data is not available due to licensing restrictions. A randomly generated example of business data is provided so that users can test the provided script to calculate the Service Distance Gradient, but the results will differ from Section III-C.

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

TABLE II: Structure of population density gradient calculation

	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	-

The structure of this analysis is described using pseudocode in Fig. 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 $1km$ and $50km$ are used. The user interface is shown in Fig. 3.

```
For DIST from 0 to MAXIMUM RADIUS:
    1 Make Buffer of size DIST around Center Polygon
    2 Clip Population count to Buffer
    3 Calculate Population Count and Area
    4 Write results to file
```

Fig. 2: Pseudocode for population density gradient calculation

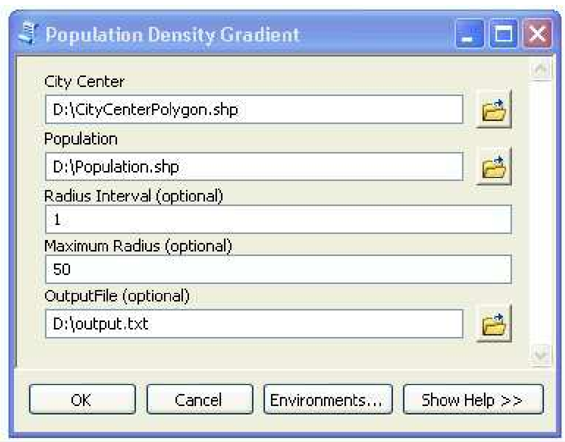

Fig. 3: Screenshot of population density gradient plugin

The population count is measured by the number of people in each blockgroup. To calculate the population density, this value is normalized by the area. This calculation assumes that people are uniformly distributed in each polygon.

B. Road Density Gradient

The road density gradient is calculated in a similar way to the population density gradient (Section II-A). Concentric circles are made from the city center, and the length of road in each ring is recorded (Fig. 4). The data required to calculate the road density gradient is listed in Table III.

The data requirements are also similar to Section II-A, 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

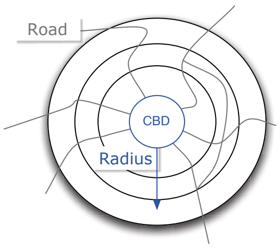

Fig. 4: Road density calculation

TABLE III: Structure of road density gradient calculation

	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	-

Fig. 5, and the user interface is structured in a similar way to Fig. 3.

```
For DIST from 0 to MAXIMUM RADIUS:
  Make Buffer of size DIST around center polygon
  Clip road network to Buffer
  Calculate Road Length and Area of buffer
  Write results to file
```

Fig. 5: Pseudocode for road density gradient calculation

C. Service Distance Gradient

The distance to services from households was calculated using raster[3] data. Fig. 6 illustrates this method of calculation. The vector based input data (polygons and points) were rasterized to square grid cells of 200m, and the distance to the nearest service was calculated for each cell. This distance is calculated by measuring the euclidean distance from each gridcell, to the nearest service or services of interest. Services were identified using the NAICS[4] classification system. To estimate the number of times a household visited a particular service in one week, data from the American Time Use Survey [19] was used.

Similar to the methods described in Section II-A and II-B, a gradient was calculated using a distance to the service or services. In addition, the population raster and the distance raster were multiplied to calculate the total distance travelled by households to services. The sum of all values in this raster was then divided by the total number of households in the city, to calculate an average value for a household. The data required for this calculation is listed in Table IV.

[3]A raster is a grid based method of representing data similar to a matrix, where each cell is georeferenced.

[4]The North American Industry Classification System (NAICS) is the standard used by Federal statistical agencies for 'classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy.' [18]

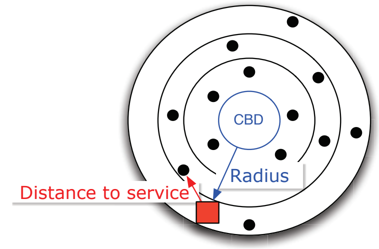

Fig. 6: Service distance calculation

TABLE IV: Structure of service density gradient calculation

	Data Type	Description	Required
Inputs	shapefile	polygon of city center	Y
	shapefile, points	service locations	Y
	numeric	radius interval	N
	numeric	maximum radius	N
	.txt	frequency of service visit (per week)	Y
Outputs	.txt	avg. dist. to service, at radius intervals	-
	.txt	avg. dist. to service for entire city	-

The structure of this analysis is described using pseudocode in Fig. 7, where lines 1-3 are calculated first, and lines 4 - 7 are repeated depending on the size of the radius interval, and the maximum radius chosen by the user. Three sources of inaccuracy in this calculation are the assumption that households visit the closest supermarket, the use of a euclidean distance measure rather than a road-network distance and the assumption that all trips are for a single purpose (no trip chaining).

```
1 Population count to raster (householdRaster)
2 Euclidean distance to supermarkets (distanceRaster)
3 HouseholdRaster * distanceRaster (householdDistanceRaster)

For DIST from 0 to MAXIMUM RADIUS:
   4 Make Buffer of size DIST around center polygon
   5 Clip householdDistanceRaster to Buffer
   6 Calculate total distance and Area
   7 Write results to file
```

Fig. 7: Pseudocode for distance to service gradient calculation

III. Example of Results

An example of the type of results that this analysis produces is illustrated using the city of Atlanta. The method can be applied to any city where the data is available and formatted appropriately. In this paper, the city center of Atlanta is used as the center point, but the analysis does not consider formal political boundaries. A maximum radius of $50\ km$ was chosen, with a radius interval of $1\ km$. The output of the analysis is fitted to specific curves and these curves all satisfy the following parameters: an *R-Squared* value > 0.90, a p-value < 0.0001 and a t-statistic > 2.

A. Population Density

Using the approach described in Section II-A, the population density gradient for Atlanta is shown in Fig. 8.

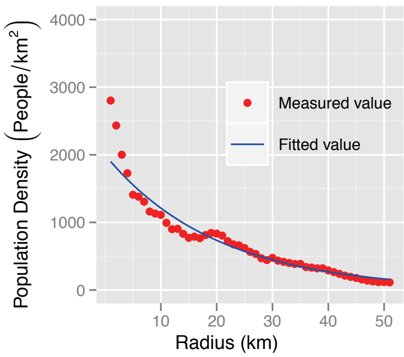

Fig. 8: Population density gradient for Atlanta.

The structure of the relationship identified was:

$$y = Ae^{rb} \tag{1}$$

where y is the population density, r is the distance from the urban center (a radial measure), and A and b are parameters to describe the curve. In the case of Atlanta, $A = 2199.60$ and $b = -0.059$.

B. Road Density

Applying the method described in Section II-B, a road density measure for Atlanta was calculated (Fig. 9). This calculation only considers local and secondary roads, so that the resource overhead of infrastructure for an area can be attributed to the people living there.

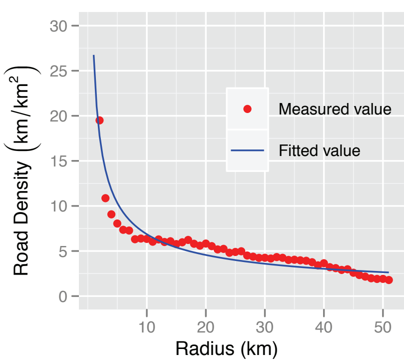

Fig. 9: Road density gradient for Atlanta.

The road density decreases as the radius increases, and this relationship was observed to follow a power law with the following structure:

$$y = Ar^b \tag{2}$$

where y is the road density, r is the radius from the urban center, and the parameters for Atlanta are $A = 26.76$ and $b = -0.591$. More detailed analysis of local road scaling parameters in US cities has been published by the authors [9]. This measurement can be used to estimate the amount of infrastructure required for a neighborhood when only the distance to a central business district is known.

As the population decreases more quickly than road density, (as functions of distance from the center) the local road length per person increases. This relationship is shown in Fig. 10 and can be described using the same structure as Eq. 1, with the parameters for Atlanta, $A = 4.63$ and $b = 0.024$. Fig. 10 illustrates how the road length per person changes in Atlanta as a function of distance from the center. This shows that three times as much local infrastructure is needed for an area of low population density, when compared to an area of high population density.

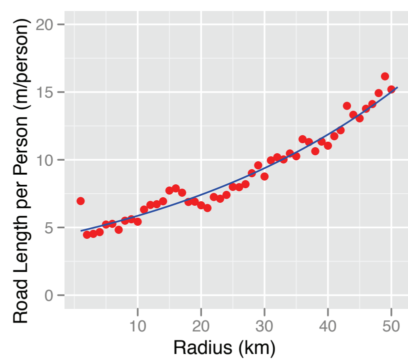

Fig. 10: Road density normalized by population density.

The quantity of material required for infrastructure can be calculated by multiplying the road-density by the area considered, and by the cross-sectional area of a typical road (Eq. 3).

$$Material = Ar^b \times Area \times Cross\text{-}Section \tag{3}$$

To estimate the amount of material required, an average width of local roads of 7.62 m is used.[5] A road of this type typically requires 125 mm of gravel and 60 mm asphalt [20], so the cross-sectional area of gravel is $7.62m \times 0.125m$ and $7.62m \times 0.06m$ for asphalt. A calculation of material required for local roads in Atlanta is shown in Table V. This illustrates the amount of construction material as a function of distance from the center. The specific road construction material used for local and secondary roads depends on state construction standards and local ground conditions. This calculation only accounts for the initial construction and does not include additional material inputs over the service life of the road bed.

C. Service Distance

Using the method described in Section II-C, a distance to service measure is calculated for Atlanta (Fig. 11). In this case, the distance to the nearest supermarkets (NAICS Code: 445110) was calculated. The average household visits a supermarket 1.173 times per week, based on data from the

[5]The road width value is based on a summary of guidelines for local road construction for the state of Massachusetts.

TABLE V: Construction material required for 1 km^2 of local road in Atlanta

Radius $[km]$	Pop. Density $[people/km^2]$	Road Per Person $[m/pers.]$	Asphalt Per Person $[m^3]$	Gravel Per Person $[m^3]$
5	1630	6.34	2.90	6
50	110	24.21	11.07	23

TABLE VI: Energy required by one household for weekly travel to a supermarket in Atlanta

Radius $[km]$	Pop. Density $[people/km^2]$	Distance to Supermarket $[km]$	Energy Per Trip $[MJ]$
5	1630	1.10	2.52
50	110	6.06	13.96

American Time Use Survey [19] so the weekly distance was multiplied by this factor. This relationship can be described using the same structure as Eq. 2, where the parameters for Atlanta are $A = 0.77$ and $b = 0.038$.

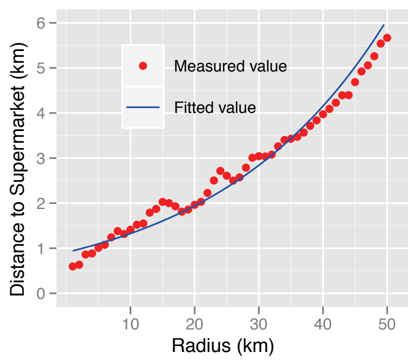

Fig. 11: Average distance travelled by one household per week to a supermarket in Atlanta

This measure can be used to estimate the average distance required to travel to a supermarket, when only the distance to a central business district is known. This method can be expanded to include other services, so that a more comprehensive measure of how far a household travels to reach typical services during the week can be estimated. In the case of Atlanta, the average distance traveled to supermarkets is 1.40 $km/week$. This low value is likely due to the NAICS classification of supermarkets, which considers 'Supermarkets and other grocery (except convenience) stores' without considering how large they are. In addition, this average value is skewed by the large number of people living close to the center of the city.

This distance measure can be related to the amount of energy used for transportation, based on the mode of transportation, and the typical MJ/km. Due to the mode choice in Atlanta being primarily automobiles ($> 95\%$), the average value of one household travelling 1.40 $km/week$ results in 3.2 MJ of energy being used per household per week on transportation to supermarkets. The value used for this calculation is 2.302 MJ/km, based on an average fuel efficiency of 22 mpg [21]. The difference between two areas of different population density can be observed in Table VI.

IV. Observations

It has been shown in this paper that population density (Eq. 1), road density (Eq. 2) and distance to services (Eq. 2), can be described in a city using a gradient measure. The existence of a population density gradient has been identified prevously [4], but typically this calculation is done on a case by case basis. In this paper we explicitly describe each step of the methodology and the assumptions involved in this calculation to formalize the approach.

A radial distance from the city center is used as input to estimate road density or distance to a service, once a central point within the city is identified. This relationship could be used to estimate the road length for a new development, or used to estimate the distance a household needs to travel to supermarket. The examples shown here are straightforward, but it is intended to develop the method so that more comprehensive parameters can be identified, which could be linked to resource efficiency targets. Ideally such an approach could be used to quickly estimate the resource efficiency of a city.

It is hoped that this type of analysis can help inform policies and zoning in cities. For example, a city may desire a minimum population density, so that the infrastructure investment per individual does not exceed a certain range. A zoning regulation that identifies a minimum population density goal could then be linked to a maximum infrastructure investment per person in a city. By highlighting the link between materials required for infrastructure, transportation energy and population density, the tradeoffs associated with urban policies can be considered.

V. Conclusions

This work develops a method to examine spatial data describing the physical form of cities, and explains how the form can be related to resource consumption measures. This is done by using road length to estimate construction material for infrastructure, and distances to supermarkets to estimate energy for transportation. The overall objective of this work is to identify gradients, and parameters of these gradients, so that they can be used to explore the functional performance of cities around the world. This methodology was illustrated using one case study; it is intended to apply this method of analysis to other cities, both in the US and other countries.

This paper describes both the process of examining parameters associated with a city, and a method of performing this calculation. In addition, it supplies the raw data and the analysis method for further exploration, verification and improvement. By identifying ranges of parameters in cities, it will provide local policy makers with a means of understanding how their city is performing, as well as providing a method of comparing the performance of cities to each other, both nationally and globally.

VI. ACKNOWLEDGMENT

The authors would like to thank Michael Quinn, John Quinn, Daniel Wiesmann, Chris Zegras, Mike Flaxman, Karen Noiva Welling, Noel Davis, Jonathan Krones and Tamas Abou-Abdo for their comments on this work.

REFERENCES

[1] L. Bettencourt and G. West, "A unified theory of urban living," *Nature*, vol. 467, no. 7318, pp. 912–913, Oct. 2010. [Online]. Available: http://dx.doi.org/10.1038/467912a

[2] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, and G. B. West, "Growth, innovation, scaling, and the pace of life in cities," *Proceedings of the National Academy of Sciences of the United States of America.*, vol. 104, no. 17, pp. 7301–7306, 2007.

[3] E. Decker, A. Kerkhoff, and M. Moses, "Global patterns of city size distributions and their fundamental drivers," *PLoS ONE*, vol. Issue 9, 2007.

[4] C. Clark, "Urban population densities." *Journal of the Royal Statistical Society*, vol. 114, pp. 490–496, 1951.

[5] M. Batty, "The Size, Scale, and Shape of Cities," *Science*, vol. 319, no. 5864, p. 769, 2008.

[6] J. D. Marshall, "Urban land area and population growth: A new scaling relationship for metropolitan expansion," *Urban Studies*, vol. 44, No. 10, pp. 1889–1904, 2007.

[7] M. Batty, R. Carvalho, A. Hudson-Smith, R. Milton, D. Smith, and P. Steadman, "Scaling and allometry in the building geometries of greater london," *The European Physical Journal B*, vol. 63, no. 3, p. 12, 2008.

[8] H. Samaniego and M. Moses, "Cities as organisms: Allometric scaling of urban road networks," *Journal of Transport and Land Use*, vol. 1, pp. 21–39, 2008.

[9] D. Quinn and J. Fernandez, "Estimating material usage of road infrastructure in US cities," in *IPBSA SimBuild Conference*. SimBuild, August 2010.

[10] P. Newman and J. Kenworthy, *Cities and automobile dependence: a sourcebook*. Gower, 1991.

[11] J. Holtzclaw, "Using residential patterns and transit to decrease auto dependence and costs," *Natural Resources Defense Council, San Francisco.*, 1994.

[12] R. Crane, "On form versus function: will the new urbanism reduce traffic, or increase it?" *Journal of Planning Education and Research*, vol. 15, no. 2, p. 117, 1996.

[13] R. Cervero and K. Kockelman, "Travel demand and the 3Ds: Density, diversity, and design," *Transportation Research Part D: Transport and Environment*, vol. 2, no. 3, pp. 199–219, Sep. 1997.

[14] E. S. R. I. (ESRI), *ArcGIS*, 10th ed., Environmental Systems Resource Institute, ESRI, Redlands, California., 2009.

[15] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011. [Online]. Available: http://www.R-project.org

[16] Quantum GIS Development Team, *Quantum GIS Geographic Information System*, 1st ed., Open Source Geospatial Foundation, 2011.

[17] [Online]. Available: http://urbmet.org/analysis

[18] (2011, April). [Online]. Available: http://www.census.gov/eos/www/naics/

[19] (2011, January). [Online]. Available: http://www.bls.gov/tus/

[20] R. Chudley and R. Greeno, *Building construction handbook*, 7th ed. Oxford: Butterworth-Heinemann, 2008.

[21] (2009). [Online]. Available: http://www.eia.doe.gov/emeu/aer/pdf/perspectives_2009.pdf