Title
Navigating Urban Dynamics: Exploring Urban Form, Motorways, and Transportation Equity

Author
Rezaei, Nazanin

Publication Date
2023

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

NAVIGATING URBAN DYNAMICS: EXPLORING URBAN FORM, MOTORWAYS, AND TRANSPORTATION EQUITY

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ENVIRONMENTAL STUDIES
with an emphasis in DATA SCEINCE

by

Nazanin Rezaei

December 2023

The Dissertation of Nazanin Rezaei is approved:

Professor Adam Millard-Ball, chair

Professor Chris Benner

Professor Elliott Campbell

Peter Biehl
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

NAVIGATING URBAN DYNAMICS: EXPLORING URBAN FORM, MOTORWAYS, AND TRANSPORTATION EQUITY

Nazanin Rezaei

This dissertation explores urban form metrics and their environmental outcomes, the impact of motorways on urban sprawl, and transportation accessibility across three interrelated chapters. Chapter 1 begins by examining the relationship between urban form metrics in 462 cities worldwide. By employing the K-Means clustering algorithm and statistical analysis, we uncover distinct typologies of urban form, highlighting the complex relationship between measures such as weighted density, street connectivity, compactness and two environmental outcomes—air pollution and green space access. Findings demonstrate that the emphasis on higher density comes with trade-offs in green space and $PM_{2.5}$ exposure, while street connectivity is crucial for reducing $PM_{2.5}$ emissions. Context-specific correlations caution against generalizations, providing a foundation for understanding urban processes and identifying effective policy responses.

Chapter 2 examines the impact of motorways on urban sprawl by investigating the correlation between motorway construction period and two key urban form metrics—street connectivity and density—across diverse understudied global contexts. By temporally separating the effects and utilizing a panel fixed effects model, we mitigated concerns related to endogeneity, allowing for a better

understanding of the causal impact of motorways. Results highlight the context-dependent nature of the impact of motorway construction and proximity on urban sprawl across various areas.

In Chapter 3, we introduce the Individual Experienced Utility-Based Synthesis (INEXUS) accessibility metric, designed to improve upon existing accessibility metrics by accounting for individual preferences and constraints within an agent-based regional transportation model. By differentiating between Potential INEXUS and Realized INEXUS, this metric offers a comprehensive view of transportation accessibility, considering the diverse needs of various subpopulations. We apply this approach to a case study involving alternative ridehail pricing scenarios, demonstrating its effectiveness in evaluating differences in accessibility within and between population groups.

These three chapters collectively contribute to a more comprehensive understanding of the complex dynamics at play in urban areas, offering insights for policymakers, urban planners, and researchers. The dissertation underscores the importance of context-specific urban planning and transportation policies, emphasizing the need for sustainability, equity, and adaptability in the face of urban challenges, providing insights for shaping more resilient and inclusive cities.

DEDICATION

To my parents

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all those who have played an important role in the successful completion of this dissertation.

Foremost, my deepest appreciation goes to my Ph.D. advisor, Adam Millard-Ball, whose guidance and supervision was pivotal in shaping the research presented in this dissertation. His unwavering support and exceptional mentorship have been instrumental throughout my Ph.D. studies.

I extend heartfelt thanks to my dissertation committee members, Chris Benner and Elliott Campbell, for their constructive feedback and dedicated support through the various milestones of my Ph.D. I am grateful to Carlos Dobkin, my data science advisor, whose guidance has been invaluable. I also thank Chris Barrington-Leigh for his contributions to street network sprawl and for providing server infrastructure. Moreover, I would like to thank Yang Liu for the insightful machine learning class in the computer science department, which enriched my research experience.

A special acknowledgment is reserved for the sustainable transportation initiative at Lawrence Berkeley National Lab, particularly Anna Spurlock, for providing the opportunity to work on transportation models and develop the INEXUS metric. The collaborative efforts within the Berkeley Lab's community, including Anna Spurlock, Annika Todd-Blick, Sydny Fujita, Cristian Poliziani, Zach Needell, Tom Wenzel, Juan David Caicedo, and Natalie Popovich, have been invaluable in this journey.

I express my gratitude to the professors and fellow students in the Environmental Studies department at the University of California, Santa Cruz for fostering a supportive and friendly academic environment. Special thanks to my cohort and incredible friends who consistently uplifted my spirits throughout this journey. Your friendship has been a crucial source of strength, particularly during challenging times when I was away from my family.

Most of all, I extend my heartfelt gratitude to my parents, from whom I have drawn everything, and to my dear brother and sister. Their support has been the bedrock of this academic endeavor. Special thanks to my beloved husband, Mehran, who has been a constant source of support throughout the process and always been there for me. This dissertation stands as a testament to the love and encouragement I have received from each of them.

Chapter specific acknowledgments follow the end of each chapter. The text of this dissertation includes reprints of the following previously published material:

Rezaei, N., & Millard-Ball, A. (2023). Urban form and its impacts on air pollution and access to green space: A global analysis of 462 cities. *PLoS ONE*, *18*(1 January), 1–26. https://doi.org/10.1371/journal.pone.0278265

The co-author listed in this publication directed and supervised the research which forms the basis for the dissertation. The PLOS journal applies the Creative Commons Attribution 4.0 International (CC BY) license that allow free and unrestricted use, to articles and other works they publish.

All co-authors of the third chapter of this dissertation, listed below, approve the inclusion of this work in the dissertation.

Rezaei, N., Todd-Blick, A., Fujita, K.S., Popovich, N., Needell, Z., Poliziani, C., Caicedo, J.D., Guirado, C., Spurlock, C.A. (2023). At the Nexus of Equity and Transportation Modeling: Assessing Accessibility through the Individual Experienced Utility-Based Synthesis (INEXUS) Metric (In review)

INTRODUCTION

Urban areas are at the forefront of a multifaceted interplay of factors with far-reaching impacts on society, the environment, and economies. This dissertation navigates through several key factors shaping urban environments, addressing the complexities of urban form metrics, the impact of motorways on urban sprawl, and the equity dimensions of transportation accessibility. The interconnectedness of these themes is explored through three distinct yet interrelated chapters, each contributing to a more detailed understanding of the challenges and opportunities inherent in urban environments.

Urban form, the physical layout of cities, plays a pivotal role in shaping environmental and societal dynamics. The adverse manifestation of urban form, often characterized by low density and scattered leapfrog development, is commonly referred to as urban sprawl. While urban sprawl has been extensively studied in the context of the United States, nowadays, it has become a more worldwide concern. Research outside the U.S. has begun to emerge, often relying on North American sprawl concepts. However, it remains unclear whether the extensive land conversion and varied sprawl measures observed in the U.S. are an appropriate model for other regions of the world.

Despite the growing recognition of cities' significant role in economic, political, and environmental systems worldwide (Schneider & Woodcock, 2008), there is a lack of comprehensive and comparative global-scale research on cities. The understanding of urban change at global scales has primarily been based on United Nations population figures for a long time, but these statistics do not usually provide information on the distribution, patterns, and scale of urban land-use change. While impressive strides have been made in studying global urban expansion over the last decade, a continued and intensified focus on this research area is essential. Studies, such as the one conducted by Angel et al. (2016), have highlighted that informality is becoming more frequent over time, and cities are expanding their territories faster than their populations (Angel et al., 2016). The phenomenon of unplanned development is mostly related to the cities in developing countries that have been unprepared for absorbing the high number of impoverished rural populations that are still increasing in informal settlements. Developing nations have witnessed significant urban sprawl in the past two decades, and projections indicate that an additional 900 million urban dwellers will be added to these areas by 2050 (L. Liu & Meng, 2020).

Using multiple sources and precise data can reduce the uncertainties connected with the transferability of current studies in this area. Some recent trends have played major roles in increasing interest in studying urban form and considerable progress in developing measures of it. Several software packages and geographic information systems (GIS) technology are now available to analyze spatial development patterns and predict the consequences. Finally, the availability of high-quality spatial data and access to aerial photographs and satellite images that reveal natural features such as topography, forest cover, and many details such as the color of the rooftops, have made significant progress in this field of study (Clifton et al., 2008; Lowry & Lowry, 2014).

Urban sprawl is associated with multiple environmental and social impacts such as increase in greenhouse gas emissions and disparity in wealth, which has led many researchers and policymakers to identify and attempt to implement alternative patterns of development. In turn, policy interventions such as urban growth boundaries, infill development, upzoning, and mixed-use land development are policies applied to mitigate adverse impacts that depend on what factors cause sprawl.

There are four major categories of concerns explicitly related to urban development, and a fifth related to public health, about sprawl: environmental impacts such as increased greenhouse gas emissions as well as more energy consumption; economic effects like inefficient movement of goods and people, wasted time, unproductive use of land; social costs such as increased racial and income discrimination as well as social isolation; political impacts like fragmented decision-making processes within urban areas, greater political conflict, and inequality; and finally health issues such as obesity due to the auto-dependent society, asthma and other respiratory diseases associated with proximity to freeways, and depression. Wilson and Chakraborty have identified four primary categories of environmental impacts attributed to urban sprawl—air, energy, land, and water (Wilson and Chakraborty, 2013). Sprawl is not only a cause of externalities but, in some instances, an environmental justice issue where those with the means to move out of the city experience lower health hazards exposures (Wilson and Chakraborty, 2013; Ewing, 2008).

From the health perspective, some studies discuss that contact with nature may offer benefits for both mental and physical health. Moreover, the experience of seeking refuge from the turmoil of city centers by relocating to suburban areas, where a sense of tranquil sanctuary prevails, can provide a soothing and restorative effect. In these regards, suburban lifestyles might offer health benefits. However, sprawl is commonly criticized for its adverse effects. The increase in car dependency resulting from urban sprawl is associated with health hazards, including air pollution, and motor vehicle crashes. Air pollution causes severe breathing problems, skin diseases, and other health problems. From the perspective of social health, low-density development is blamed for reducing social interaction and threatening the ways that people live together. Some articles have also demonstrated that sprawl can lead to obesity and depression (Bhatta, 2010; Ewing et al., 2014; Frumkin et al., 2004).

This dissertation seeks to address the gaps mentioned by undertaking a comparative analysis of urban form metrics, aiming to understand the transferability of existing studies to diverse regions worldwide. The opening chapter is dedicated to unraveling the correlations among various urban form metrics across 462 metropolitan areas worldwide with diverse economic characteristics, establishing a comprehensive global typology of urban forms. The relationship between various urban form metrics, including weighted density, density gradient slope, density gradient intercept, compactness, and street connectivity, are examined and the resulting typology of urban forms derived from this analysis challenges broad generalizations, emphasizing the need for context-specific urban form studies.

Notably, the study underscores the need for a multifaceted approach, incorporating indicators like street connectivity, to study urban form and its consequences. This chapter builds on the advances in data quality, using the Global Human Settlement Layer (GHSL) dataset to calculate urban form metrics.

Moreover, in the first chapter, we assess the impact of different urban form metrics on two key environmental factors: air pollution and green space access. The scrutinized environmental metrics include $PM_{2.5}$ emissions per capita, $PM_{2.5}$ exposure, green space per capita, and the percentage of individuals residing in high green areas. The discussion highlights critical environmental issues, particularly air pollution and insufficient access to green spaces. Environmental justice concerns are evident across countries with varying income levels, as lower-income countries exhibit higher exposure to $PM_{2.5}$ emissions, despite lower per capita emissions from the transportation sector.

Challenges arise when we want to measure urban form metrics, especially at the global scale, as there is a lack of data on CO_2 emissions and green spaces, with varying definitions in different geographic regions, hindering global-level city studies. Additionally, the absence of data on employment and urban land-use restricts the examination of factors such as polycentricity and mixed-use measures at the global scale. There is room for further improvement of the quantity and quality of data which can enhance the future studies regarding the urban form measures.

The findings of the first chapter lay a solid groundwork for subsequent research and policy responses geared toward fostering sustainable, context-aware urban environments. The subsequent chapters in this dissertation build upon these insights, offering a detailed understanding of the complexities inherent in urban form and its far-reaching implications on the well-being of urban communities.

Having explored the urban form metrics and their environmental impacts in the preceding chapter, we shift our focus to a pivotal driver of urban dynamics—the influence of motorways. This transition allows us to delve into the evolving landscape shaped by the interaction of urban structures and major transportation arteries.

There are multiple interrelated causal relationships identified between urban sprawl and different factors ranging from economic causes, policy-related reasons, social drivers, and historical-geographical causes. Population growth and other demographic drivers such as household size and proportion of children are also mentioned in some studies, but they are mainly recognized as the natural causes of urban growth, not necessarily urban sprawl (Colsaet et al., 2018; Wassmer, 2006). The degree to which sprawl in a particular place is caused by any of these drivers is important mainly in terms of policy intervention.

In the second chapter, we examine one of the drivers of urban sprawl which is the influence of motorways on urban sprawl. As urban areas grapple with growing populations and expanding economic activities, understanding how motorways impact urban form becomes essential. Contradictory findings in existing research necessitate a comprehensive exploration of the relationship between motorway construction and urban sprawl. By analyzing density and street connectivity metrics across diverse global contexts, this chapter contributes insights beyond localized contexts, enhancing our understanding of the dynamics between motorways and urban sprawl.

When discussing urban sprawl, it is essential to establish a consistent definition. Additionally, categorizing roads and highways based on their access levels, distinguishing between intracity and intercity highways, is crucial for assessing their impact on sprawl levels. Furthermore, expanding the scope of future research to include study areas in traditionally understudied regions, such as cities in Asia, Africa, or South America, holds significant promise. This is particularly important since a majority of existing studies in this field have predominantly focused on North America, Europe, and China.

As we conclude our investigation into motorways' impact on urban sprawl, we transition to the area of transportation accessibility. This transition forms a critical bridge between the macro-level influence of motorways and the micro-level experiences of urban residents. Existing accessibility metrics are often location-based, driven by observational data, and describe location-based accessibility (Hou et al., 2019) or transportation affordability (Center for Neighborhood Technology, 2017). While these metrics offer valuable insights into the transportation system, they fall short of representing the full distribution of policy impacts across individuals, and often do not fully account for the behavioral responses that may more realistically underpin actual accessibility.

To complement these metrics, we propose two individual-level metrics, collectively called the Individual Experienced Utility-Based Synthesis (INEXUS).

These metrics capture the Potential—the value of the complete choice set of available transportation modes for each person for every trip they take during a typical day—as well as the Realized—the achieved “happiness” from the selected mode of a given trip. The INEXUS specifications provide an agent-trip-level utility-based set of accessibility metrics that can complement other types of location-based accessibility metrics to generate a holistic understanding of variation in potential and realized experiences, and how they are valued by the individual travelers, across population groups and transportation scenarios.

By differentiating between Potential INEXUS and Realized INEXUS, this metric offers a more comprehensive view of transportation accessibility, considering the diverse needs of various population groups. The application of INEXUS to a case study involving alternative ridehail pricing scenarios demonstrates its effectiveness in evaluating differences in accessibility within and between population groups. This chapter highlights the importance of considering individual preferences and constraints in transportation modeling to ensure equitable access to goods and services. The insights gained from this research contribute to the development of more inclusive and responsive transportation policies.

In this dissertation, a range of analytical techniques has been employed, encompassing machine learning, econometrics, and statistical analysis methods such as Lowess regression, K-Means clustering, random forest regression, identification strategy, panel fixed effects model, and OLS regression. The primary sources of data in the initial two chapters include GHSL and OpenStreetMap. For the third chapter, open-source agent-based and activity-based models, namely BEAM and ActivitySim, were utilized to generate the output data.

In conclusion, this dissertation provides a multidimensional exploration of urban dynamics, emphasizing the need for context-specific approaches in urban planning and transportation policies. The implications extend beyond academia, providing insights for policymakers, urban planners, and researchers. This research contributes to a vision of smarter, more sustainable urban growth by delving into the intricate dynamics of urban environments.

Future research can be focused on adjustments to the description or measurement of urban sprawl. The results of the urban form typology can be further examined across cities of various sizes worldwide to gain insights into the specific characteristics that contribute to variations within and among the clusters. Research gaps in this area valuable to study deeper are related to economic causes of sprawl, evaluation of the mitigation policy implementation, comparative analysis of various growth management strategies, as well as international experience in tackling sprawling development. It is also worthwhile to study the historical segregation behind urban sprawl and other social causes in the U.S. and compare it to the drivers of sprawl in other countries.

CHAPTER 1

Urban Form and its Impacts on Air Pollution and Access to Green Space: A Global Analysis of 462 Cities

Abstract

A better understanding of urban form metrics and their environmental outcomes can help urban policymakers determine which policies will lead to more sustainable growth. In this study, we have examined five urban form metrics – weighted density, density gradient slope, density gradient intercept, compactness, and street connectivity – for 462 metropolitan areas worldwide. We compared urban form metrics and examined their correlations with each other across geographic regions and socioeconomic characteristics such as income. Using the K-Means clustering algorithm, we then developed a typology of urban forms worldwide. Furthermore, we assessed the associations between urban form metrics and two important environmental outcomes: green space access and air pollution.

Our results demonstrate that while higher density is often emphasized as the way to reduce driving and thus $PM_{2.5}$ emissions, it comes with a downside – less green space access and more exposure to $PM_{2.5}$. Moreover, street connectivity has a stronger association with reduced $PM_{2.5}$ emissions from the transportation sector. We further show that it is not appropriate to generalize urban form characteristics and impacts from one income group or geographical region to another, since the correlations between urban form metrics are context specific. Our conclusions indicate that density is not the only proxy for different aspects of urban form and multiple indicators such as street connectivity are needed. Our findings provide the foundation for future work to understand urban processes and identify effective policy responses.

Keywords: Urban form, air pollution, green space access, street connectivity, density, applied machine learning

1. Introduction

Urban form is one of the most central topics in the study of urbanism. Once the form of a city has been defined, it is challenging and expensive to alter it significantly (Liang et al., 2020). The physical shape and structure of cities may lead to adverse environmental impacts such as greenhouse gas emissions, economic effects such as high public service costs and related taxes, and social and political consequences like inequality and social segregation. Therefore, urban form has become a topic of widespread public policy and academic interest (Bhatta, 2010; Ewing, 2008; Wilson & Chakraborty, 2013). Moreover, progress towards smarter and more sustainable growth can only be made with a complete understanding of urban form and its relationship with air pollution (Bechle et al., 2011; Clark et al., 2011) and other environmental outcomes, as well as the policy measures required to respond to it.

Recent trends such as the availability of high-quality spatial data, access to aerial photographs and satellite images, the use of spatial analysis software packages, and geographic information system (GIS) technology have significantly contributed to the interest in studying urban form and progress in developing its measures (Clifton et al., 2008; Lowry & Lowry, 2014). A variety of spatial metrics have been developed to characterize and quantify urban form (Angel et al., 2016; Barrington-Leigh & Millard-Ball, 2020; Burchfield et al., 2006; Galster et al., 2001; Hamidi & Ewing, 2014; Sevtsuk & Amindarbari, 2013; Song & Knaap, 2004). The selection of the proper metrics requires an understanding of the available data, the scope and discipline of the research, and the characteristics of the geographical region.

Initially, density was the primary metric used to measure and characterize urban form, and it still is since it can be measured relatively easily and data is readily available. Many studies point to its association with vehicle ownership and travel (Hamidi & Ewing, 2014; Sohn et al., 2012). In addition, density is assumed to correlate with other aspects of urban form such as size, coverage, and compactness and is easy to understand and interpret. However, it is not evident whether density is a reasonable proxy for these different aspects of urban form in various contexts, or whether multiple indicators are needed. There are also problems with simple density measures since they may not accurately reflect the actual density faced by the individuals or firms involved. As an example, the metropolitan area of Flagstaff, Arizona, consists of the second-largest county in the U.S., but is spread across multiple forests, national parks, and monuments, meaning that the density experienced by the average person is much higher than average density as typically measured (people per square kilometer) (Duranton & Puga, 2020).

Where data are available, researchers have therefore gone beyond density and considered alternative and/or complementary measures. For example, Burchfield et al. (2006) use the percentage of undeveloped land in a square kilometer surrounding an average residential area as an indicator of urban form. Another group of researchers has used principal component analysis to combine many variables into higher-dimensional factors such as degree of mixed-use, degree of centrality, and street accessibility to operationalize urban form (Ewing et al., 2003). Urban form can also be measured by per capita land consumption, i.e., land area divided by population, which demonstrates the efficiency of growth in urban areas (Banai & DePriest, 2014). Recently, Barrington-Leigh and Millard-Ball (2020) have introduced a measure of urban form called the street-network disconnectedness index (SNDi). Their study demonstrates that connected street networks, such as grids, increase accessibility by walking, bicycling, and public transportation, as more destinations can be reached in a specific time. In contrast, dendritic or cul-de-sac-dominated networks favor traveling by private cars (Barrington-Leigh & Millard-Ball, 2020). Furthermore, Sevtsuk and Amindarbari (2013) defined nine metrics of urban form in their study: size, density, coverage, discontiguity, compactness, polycentricity, expandability, and land-use mix.

Patterns of urban form are extensively studied in the West, particularly the U.S. (Wheeler, 2008). In recent years, the shape of urban growth has also become a worldwide concern; however, it is unclear whether the results from studies in the U.S. generalize globally. It is useful to focus more on this area of research because cities are expanding their boundaries faster than their populations, and unplanned development is mainly related to the cities in developing countries that have been unprepared for absorbing the high number of often-impoverished rural people (Angel et al., 2016). Research outside the U.S. has begun to emerge, often relying on North American sprawl concepts to describe trends in Europe, China, and even diverse locations such as India (Goswami & Khire, 2016), Iran (Bagheri & Tousi, 2018; Ebrahimpour-Masoumi, 2012; Gouda et al., 2016; Zanganeh Shahraki et al., 2011), Japan (Pernice, 2007), Turkey (Beyhan et al., 2012; Terzi & Bolen, 2009), and Egypt (Gouda et al., 2016). The results of one recent review paper in this field show a clear unevenness of the geographic coverage since 85% of their reviewed articles focus on Europe, North America, or China (Colsaet et al., 2018). Moreover, there is limited comprehensive, comparative, and global-scale research on cities despite their significant role in economic, political, and environmental systems around the world (Schneider & Woodcock, 2008).

Among the previous studies in this area with global coverage are Atlas of Urban Expansion (Angel et al., 2016), Global Comparative Analysis of Urban Form (Huang et al., 2007), and the Global Homogenization of Urban Form (Lemoine-Rodríguez et al., 2020). Huang et al. (2007) have used 77 satellite images of metropolitan areas measuring spatial metrics such as complexity, centrality, compactness, and density. Atlas of Urban Expansion uses a sample of 200 cities from the UN Global Sample of Cities to measure some important attributes of cities such as density, compactness, and fragmentation. Similarly, Lemoine-Rodriguez et al. (2020) have applied landscape metrics such as ratio of open space, patch density, edge density, and landscape shape index to a global sample of 194 cities obtained from the Atlas of Urban Expansion database.

This paper builds on this literature by using a global dataset to first develop a typology of urban forms worldwide. As it is unclear how various measures of urban form relate to each other in different parts of the world, we compared different urban form metrics and examined how they co-vary with each other and across geographic regions and socioeconomic characteristics such as income. Second, we examine the relationship between our typology and two important policy outcomes: access to green space and air pollution, specifically $PM_{2.5}$.

Fine particulate matter ($PM_{2.5}$) refers to suspended particulates smaller than 2.5 microns in diameter capable of penetrating very deep into the respiratory tract (OECD, 2015). Among all air pollutants, $PM_{2.5}$ poses the greatest health risk globally, affecting more people than any other pollutant and increasing the risk of respiratory and cardiovascular diseases in chronic exposure. $PM_{2.5}$ can be of natural (i.e., sand and dust) or of anthropogenic source (i.e., combustion residuals), and its concentration is of high concern especially in urban agglomerations that are dense and growing rapidly (Florczyk et al., 2019). Two measures are important to consider: population exposure to air pollution by fine particulates and emissions of fine particulates from transportation. While the two measures are related, the first emphasizes the health *impacts*, while the second emphasizes the *contributions* of the transportation sector, given that urban form is most likely to affect emissions through its impacts on vehicle travel.

Access to green spaces in urban areas has been recognized as an essential component of the urban environment (Lee et al., 2015). Green spaces in cities mostly include semi-natural vegetation cover, such as street trees, lawns, parks, gardens, forests, green roofs (Florczyk et al., 2019). Therefore, this study aims to find the relationship between urban form and access to green spaces in large urban areas using two measures: green space per capita and the share of population living in areas of high green.

The following section reviews the literature on measuring urban form, mainly focusing on weighted density, density gradient, compactness, and street connectivity metrics. We then discuss our methods, including data characteristics, computational methods, and analysis methods such as Lowess regression, K-Means clustering, and random forest regression. Afterward, we present the results and discussion section examining correlations between urban form metrics, typology of the cities, the geographic distribution of the clusters, and environmental outcomes of urban form metrics (Figure 1.1). Finally, we conclude by suggesting that while higher density is often emphasized as the way to reduce driving (Brownstone & Golob, 2009; Ewing & Cervero, 2017) (due to higher access to active and public transit) (Gately et al., 2015; Newman & Kenworthy, 2021) and thus $PM_{2.5}$ transportation emissions, it comes with a downside: less green space access, and more $PM_{2.5}$ exposure. Moreover, street connectivity has a stronger association with reduced $PM_{2.5}$ emissions.

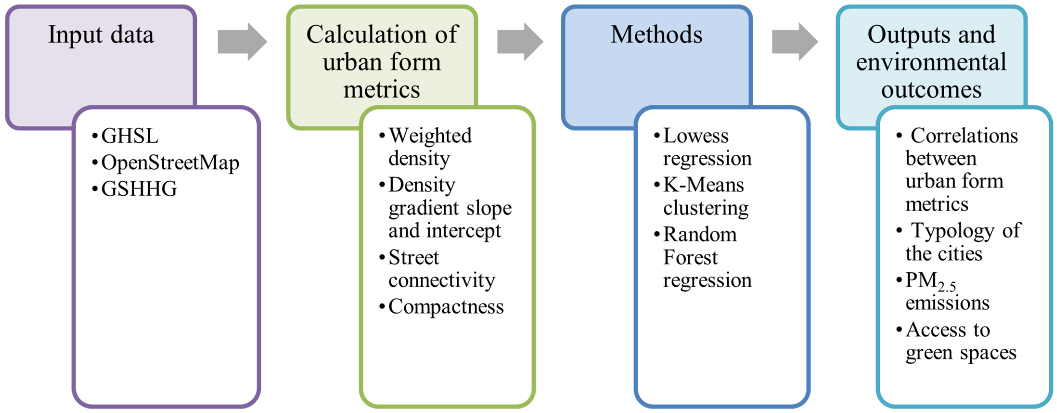


Figure 1.1. Flowchart of the input, process, and output of the study

2. Measuring urban form

Urban form refers to the physical characteristics of built-up areas, such as the shape, size, density, and configuration of settlements (K. Williams, 2014; Živković, 2019). Due to its negative environmental, social, and economic impacts, urban sprawl is considered as the unfavorable display of urban form. While there are many definitions and dimensions of sprawl, sprawl is used to describe a process with the result of low-density, scattered, leapfrog or isolated development, unplanned and haphazard land-use, endless commercial strip malls, and the lack of adequate transportation and housing options with high levels of automobile reliance at the periphery of urban areas (Angel et al., 2016; Baldassare, 2005; Ewing et al., 2003; Handy et al., 2002; Hayden, 2004; Wilson & Chakraborty, 2013). It describes cities where people need to drive long distances to conduct their daily lives or where employment is decentralized (Glaeser & Kahn, 2004). According to Galster et al. (2001), sprawl is a pattern of land-use in an urban area that exhibits insufficiencies within some or all of eight distinct dimensions, including continuity, density, centrality, concentration, clustering, nuclearity, mixed uses, and proximity. In this paper, we prefer the term "urban form" as it is more general and less value-laden than "urban sprawl", but our analysis could be interpreted as a study of urban sprawl.

Urban geographers, planners, economists, and other social scientists have developed numerous measures to quantify urban form characteristics, typically using remote sensing and/or census-based population datasets. In this section, we review the literature and discuss the motivation for several of these measures, focusing on the ones we employ in our analysis and treating others more briefly. The metrics of urban form that are measured in this research include street connectivity, weighted density, density gradient (intercept and slope), and compactness. Street connectivity is selected due to the major long-lasting impact of street networks on urban form. Density is the metric used in most similar global studies, and we use a variant, weighted density, in this research. Density gradient is another descriptive and analytical metric that demonstrates the change in population density across the city. Compactness is selected as it reflects the ratio of built-up area to the buildable area within the convex hull and is an important metric used in urban form studies. A brief definition of each measure is explained in Table 1.1.

Table 1.1. Urban form metrics used in this research (Angel et al., 2007; Barrington-Leigh & Millard-Ball, 2020; Duranton & Puga, 2020; Sevtsuk & Amindarbari, 2013)

Name of the Metric	**Definition**
Street Connectivity	We use the Street Network Disconnectedness Index (SNDi), which is a summary scalar measure for street-network sprawl
Weighted Density	The density in each one kilometer square grid cell, weighted by the population of that grid cell
Density Gradient	The change in population density in an urban area from the center to the periphery
Compactness	The built-up area of the city divided by the area of its convex hull (excluding unbuildable areas)

2.1. Street connectivity

Previous research in Europe and North America shows that disconnected urban street networks are strongly associated with increased vehicle travel, energy use, and CO_2 emissions. In one of their recent articles, Barrington-Leigh and Millard-Ball (2019) introduced a summary measure of street-network sprawl called the street-network disconnectedness index (SNDi). They used a large dataset of intersections and edges to calculate nodal degree for each intersection, dendricity, circuity, and sinuosity. Dendricity shows whether “a network is tree-like in that there is only one possible route to reach certain nodes”. Circuity means “the ratio of path length to straight-line distance between each node and other nodes in its vicinity”. Sinuosity is the ratio of length to end-to-end distance for every edge. They used a principal components analysis to assign empirical weights to their aggregate measures. They focus on the first principal component as an overall measure of street-network sprawl and refer to this measure as the Street-Network Disconnectedness Index (SNDi) (Barrington-Leigh & Millard-Ball, 2019).

2.2. Weighted density

Density is typically calculated as the number of people in a given geography (city, country, grid cell, etc.) divided by the land area. However, such a "simple density" measure may not accurately reflect the actual density faced by the individuals or firms involved. For example, the density of Canada does not reflect the experience of someone in central Toronto or Montreal. Moreover, density depends on the arbitrary shape and size of the units of analysis, such as whether parks, deserts, or other open spaces are included. Duranton and Puga (2020) propose that we address these challenges by measuring "experienced density" instead, i.e., by counting the population within a given radius around each individual. In addition to addressing the arbitrary nature of boundary lines, such experienced density also captures how close the typical individual is to others in an unevenly distributed population. In practice, experienced density is not feasible to calculate as the precise location of each individual is not known, but a weighted density measure is similar in practice – for example, weighting grid-cell level population densities by the population of that grid cell. It is essential to weight by population since otherwise, we would be calculating population within ten kilometers of the average place instead of within ten kilometers of the average person (Duranton & Puga, 2020).

2.3. Density gradient

Population density gradient has been extensively used in urban studies, both as a descriptive device and as a hypothesis-testing instrument. This measure represents a useful summary statistic of the distribution of population concentration in cities (Edmonston et al., 1985; Lewis et al., 2013). In other words, it indicates the unevenness of population distribution. As density gradients incorporate crucial aspects of urban land-use and overcome traditional constraints in measuring urban densities, they are potentially helpful for measuring urban sprawl. Moreover, the calculation of density gradients is generally independent of the dynamics of urban structures, such as shifts in the urban periphery and changes in its boundary over time, since they use the distance from activity centers as their basis (Torrens & Alberti, 2000).

Population density gradient depends on the spatial distribution function. According to Parr, the negative exponential function is a better fit for describing density in urban areas, while the inverse power function is better suited to the urban fringe and hinterland (Batty & Longley, 1994). Urban scholars often associate population density gradients from a negative exponential function with Alonso, Muth, and Mills' pioneering work, but it was actually Colin Clark who popularized this concept first in 1951 (Bertaud & Malpezzi, 2003). Angel et al. (2007) calculated density gradient by measuring the population density of small administrative areas. As a next step, they found the average population density in rings of increasing distance from the city center. Finally, they fitted a negative exponential curve to the average ring density points as a function of distance from the city center. The gradient is steep when a city's population density falls rapidly as the distance from the city center increases. When population density is pretty much the same in the center and everywhere else, the gradient is shallow (Angel et al., 2007; Torrens & Alberti, 2000).

2.4. Compactness

Compactness is considered as one of the most important aspects of geographic shapes and can be quantified in several ways (D'Acci, 2019; Maceachren, 1985). A common approach to calculate compactness is to measure the ratio of area to a circle (Angel et al., 2010) or the polygon's convex hull (Ogle et al., 2017). According to Angel et al. (2007), compactness is defined as the ratio of the observed built-up area to the observed buildable area within "the circle of minimum radius encompassing the consolidated built-up area of the city".

Sevtsuk and Amindarbari (2013) have proposed two new size metrics to consider when calculating the size of the urban centers. One is the area of the convex hull around all developed polygons. The convex hull is defined as "the smallest flat polygon that has no concavity in its perimeter and that fully contains all individual polygons of the urban extent of a city". The second metric is the unbuildable area within the convex hull. The authors suggest that the definition of "unbuildable" should include areas that are not buildable due to natural barriers, such as water bodies, steep slopes, or other natural limitations, but exclude those that result from policy decisions (e.g., Central Park in Manhattan). This definition will allow for a more reliable and consistent assessment of density (Sevtsuk & Amindarbari, 2013).

Built-up coverage shows how much of the total urban extent or a particular sub-area of the city is occupied by built-up areas. Estimating coverage within the convex hull minus unbuildable area is a consistent way to measure how much land is urbanized within a city. The exact shape of the hull and the extent of the unbuildable area varies from city to city. However, the estimation remains consistent and comparable since remaining vacant land within the hull can be said to be vacant voluntarily (Sevtsuk & Amindarbari, 2013). We consider the built-up coverage metric the same as compactness in our study. There are other definitions of compactness, but we use the convex hull method as it is flexible in cities with different shapes and topographical and water constraints. This metric is commonly estimated for building footprints - the percentage of the urban area covered by buildings.

Compactness is scale-dependent and image resolution can affect the absolute value of our compactness measure (Hamidi et al., 2015). However, we expect scale dependency of this metric to have limited impact on our results since we use a consistent image resolution across the study, and we examine the difference in compactness between the cities rather than the absolute value.

2.5. Other metrics

A number of other metrics have been used and presented in various studies, such as polycentricity, land-use mix, built-up density, and fragmentation. We discuss some of them briefly here, but do not use them in the empirical portion of this paper due to a lack of appropriate data or similarity to the other metrics selected.

2.5.1. Polycentricity

Polycentricity describes the degree of employment concentration in a city's sub-centers. This metric can also show how many significant job centers are within those sub-centers, what their size distribution is, and what employment percentage is located there. Polycentricity is hard to consistently quantify since the definition of a center is not obvious and the number of centers cannot necessarily be the sole indicator of polycentricity (Sevtsuk & Amindarbari, 2013). Moreover, calculating polycentricity requires access to high-resolution data on employment, which is not available at the global level. While we experimented with a calculation of residential polycentricity using population data, the results did not appear meaningful or relate to employment-based polycentricity measures.

2.5.2. Land-use mix

The planning literature has extensively discussed the effects of land-use mix on traffic congestion, transportation energy consumption, real estate values, and crime rates. Two popular types of metrics are used to capture land-use mixing: one is based on the number of different uses found in a given area, making it possible to rank areas according to the number of different land-uses they support, and the second evaluates a land-use pattern's heterogeneity or homogeneity by examining how equally each use occupies an area (Sevtsuk & Amindarbari, 2013).

Hamidi et al. (2015) proposed three variables for calculating land-use mix using principal component analysis: a measure of job-population balance, a measure of job mix, and an average weighted walk score for metropolitan areas. According to Sevtuk and Amindarbari (2013), an optimal land-use mix metric function should be estimated at the level of small intra-urban subdivisions rather than the entire city. The most accurate estimation is possible when the land-use data is given at pixel or raster cells level. We did not include the land-use mix metric in our study since we did not have access to any global datasets of employment numbers.

2.5.3. Urban extent density and built-up density

Urban extent density is the ratio of the total population of the city to its urban extent, measured in persons per hectare. Built-up area density is the ratio of the total population of the city to its built-up area, measured in persons per hectare. Urban extent density is always lower than built-up area density. Also, because the urban extent of the city contains its urbanized open space, urban extent density is not independent of the city's level of fragmentation, while built-up area density is. Two cities with the same population and the same built-up area will have the same built-up area density. If one city is more fragmented—its built-up area occupies only 40% of its urban extent—and the other city's built-up area occupies 80% of its urban extent, then the urban extent density in the former will be half that of the latter (Angel et al., 2016).

2.5.4. Fragmentation

Fragmentation measures to what extent a city's built-up area covers its urban extent or, how much built-up area within an urban extent is fragmented by urbanized open space. A more fragmented built-up area implies a lower urban extent density, greater distances between places within the city, and more open space disturbed. Angel et al. (2016) provide two measures of urban fragmentation that highly correlate with each other. One is the ratio of the built-up area within the urban extent of the city and its urban extent, called saturation. The other is openness, which is "the average share of open space pixels within the Walking Distance Circle (a circle with an area of 1km^2 and a radius of 564 meters) of every built-up pixel within the urban extent" (Angel et al., 2016). We decided not to use metrics such as built-up density and fragmentation as our measure of compactness addresses a similar aspect of urban form.

2.5.5. Different geometric measures

There are various geometric metrics that can be used to measure the shape of cities, such as circularity ratio, form ratio, elongation ratio, ellipticity index, and radial shape index. Circularity ratio, proposed by Miller in 1953, is calculated by multiplying the urban area by four and dividing it by the square of the perimeter (the length of urban boundary). Form ratio, suggested by Horton in 1932, is the ratio of area to the square of the length of the longest axis of a region. Elongation ratio, proposed by Werrity in 1969, is defined as the ratio of the length of the longest axis of a region to the length of its secondary axis. This ratio measures the extended degree of a region; the more extended the urban shape, the higher the ratio (Chen, 2011, 2022; Haggett et al., 1977; Y. Liu et al., 2012).

The ellipticity index, proposed by Stoddart in 1965, is the inverse of the compactness ratio. It is measured by dividing the area of the smallest circle to enclose the figure by the urban area (Chen, 2022; Haggett et al., 1977). The Boyce-Clark radial shape index measures shapes by comparing the relative lengths of regularly spaced radials extending outward from a central node to the values a circle would have (Cerny, 1975). Since these measures are similar in spirit to compactness and it is unclear whether they would have additional explanatory power with respect to outcomes such as emissions and green space, we have not used them in our study. However, using a wider range of urban form measures could be a topic of future work.

3. Methods

3.1. Data

Our primary data are drawn from the European Commission Joint Research Center's Global Human Settlement Layer (GHSL), which contains fine-scale global and multitemporal geospatial data on populations and built-up areas worldwide. By using GHSL, cities can be studied globally in a consistent and comparative way. By looking at population dynamics and the expansion of built-up areas, GHSL analyzes the growth of human settlements worldwide (Melchiorri & Siragusa, 2018).

GHSL data are available for 1975, 1990, 2000, and 2015 at three different scales: 30m, 250m, and 1km. The one-kilometer grid cells were used to calculate urban form metrics in this study. The GHSL project produces global population density grids (GHS-POP) by combining the built-up area grid (GHS-BUILT) with census data through spatial modeling techniques. The population data is collected by national censuses with heterogeneous criteria and heterogeneous update time (Florczyk et al., 2019). One novelty of our study is using the GHSL data to calculate different measures of urban form and compare them across the urban centers worldwide.

The GHSL dataset defines 13135 Urban Centers (UCs) worldwide. A UC is roughly equivalent to a city, but the boundaries are based on population and built-up intensity rather than administrative boundaries. UCs are spatially delineated by a "degree of urbanization" model, which considers population size, density, and contiguity of populated grid cells. Each UC is a cluster of grid cells such that each $1km^2$ grid cell has a density of at least 1,500 inhabitants, or 50% of the surface is built-up, and the cluster of grid cells is contiguous with a minimum population of 50,000 inhabitants. In this study, urban centers with more than 1 million population in 2015 were used, which includes 462 cities.

About 56 percent of the cities above 1 million studied in this research are located in Asia. 13.5 percent in Africa, 12 percent in Latin America and the Caribbean, 10 percent in Europe, 7.5 percent in Northern America, and less than one percent in Oceania. Around 23 percent of the cities above 1 million population are in high income countries, 34 percent are in upper-middle-income countries, 38 percent are in lower-middle-income countries, and 6 percent are in low-income countries.

The GHSL dataset also provides our two outcome measures: $PM_{2.5}$ and access to green space. In the GHSL data, $PM_{2.5}$ emissions are derived from the European Commission's in-house Emissions Database for Global Atmospheric Research that estimates anthropogenic greenhouse gas and particulate air pollutant emissions from 1970 to 2012 (Florczyk et al., 2019). The calculation of the emissions considers all human activities, except large-scale biomass burning and land use, land-use change, and forestry. Due to the bottom-up compilation methodology of sector-specific emissions applied consistently across all world countries, the results are comparable. Total emissions of $PM_{2.5}$ from the transport sector in 2012 is used in this study, which is calculated based on 2015 urban center boundaries, and expressed in tonnes per year. This measure is used to calculate the transportation emissions per capita. In addition, $PM_{2.5}$ concentration, based on the Global Burden of Disease (GBD) 2017 data on ambient air pollution from 1990 to 2015, is used to determine emissions exposure. This metric is expressed in $\mu g/m^3$. As the CO_2 emissions data is available only at the country level, and not at the city level, we did not include it in this study.

The data on green spaces has been produced by analyzing Landsat annual Top-of-Atmosphere (TOA) reflectance composites created by considering the highest value of the Normalized Difference Vegetation Index (NDVI) as the composite value (i.e., greenest pixel). Average greenness is estimated for 2014 and calculated based on the 2015 urban center boundaries. The access to green spaces is measured through a proxy metric, "generalized potential access to green areas", which measures the number of people living in high green areas at the generalization scale of the spatial data used in the assessment, regardless of their usability. The metric builds on the greenness metric derived from remote sensing Landsat imagery. The area of high green in 2015 is delineated per each urban center, and then intersected with the population grid of epoch 2015. Finally, the share of the urban center population living in this area is estimated from the total population of the urban center (Florczyk et al., 2019).

In addition to GHSL, we use three other data sources. First, the country classification by income level is based on 2016 GNI per capita from the World Bank. When it comes to income, the World Bank divides the world's economies into four income groups: high, upper-middle, lower-middle, and low (Florczyk et al., 2019). Second, shoreline data are from the Global Self-consistent, Hierarchical, High-resolution Geography (GSHHG) database, developed by Wessel and Smith and updated in June 2017 (https://www.soest.hawaii.edu/pwessel/gshhg/). We use the shoreline dataset to calculate the unbuildable areas within a convex hull to measure compactness. Third, street connectivity data are from Barrington-Leigh and Millard-Ball (2020), updated with the February 2022 vintage of the OpenStreetMap road network.

3.2. Computational approach

Our selected urban form metrics were calculated using SQL code from data stored in PostgreSQL/PostGIS database. Here, we provide a precise definition of each metric.

Street connectivity is measured by the Street Network Disconnectedness Index (SNDi), and we do not make further calculations here beyond aggregating to urban center boundaries. SNDi is an index that includes measures of circuity, the proportion of deadends, and other connectivity measures, as explained in Barrington-Leigh & Millard-Ball (2019). We reverse the sign for consistency with our other measures. Higher SNDi values mean more disconnected streets, but here we use "negative SNDi" so that a higher value indicates greater connectivity.

Weighted density is the density in each one kilometer square grid cell, weighted by the population of that grid cell (Eq. 1.1). Because the cells are one kilometer square in this study, density is the same as population.

$$weighted\ density = \frac{\sum_i (density_i \times population_i)}{\sum_i (population_i)} = \frac{\sum_i (population_i^{\ 2})}{\sum_i (population_i)}\ for\ each\ grid\ cell\ i \qquad (1.1)$$

To calculate density gradient, we use the negative exponential function, suggested by Mills, in Eq. 1.2, to estimate the intercept and slope parameters.

$$d(r) = d_0 e^{-\alpha r} \qquad (1.2)$$

where d(r) is the population density at distance r from the city center, d_0 is the population density at the city center, and α is the density gradient (Angel et al., 2007; Batty & Longley, 1994; Edmonston et al., 1985; Meier & McCoy, 2010). As the center of each urban center is not known with the available data in this research, and the highest-density pixel is not necessarily the center, we have used an approach that calculates a separate density gradient from every pixel in the urban center. We then assume that the center is the pixel with the largest density gradient. The density gradient is the regression line between density and distance from the center. As with any regression line, it has an intercept (the theoretical density at the center), as well as a slope. Therefore, we calculate two different urban form metrics: density gradient slope and density gradient intercept.

The convex hull of all urban centers are calculated, excluding the unbuildable areas around them due to bodies of water as given by the GSHHG shoreline dataset.

Compactness is then the ratio of the built area to the convex hull (Eq. 1.3). Figure 1.2 shows an example.

$$C_{BLT} = \frac{A_{BLT}}{(A_{CVX} - A_{UNB})} \tag{1.3}$$

A_{CVX}: the area of the convex hull around the developed polygons

A_{UNB}: the unbuildable area of the convex hull

A_{BLT}: the total area of all built-up polygons

C_{BLT}: the built-up coverage (compactness) (Sevtsuk & Amindarbari, 2013).

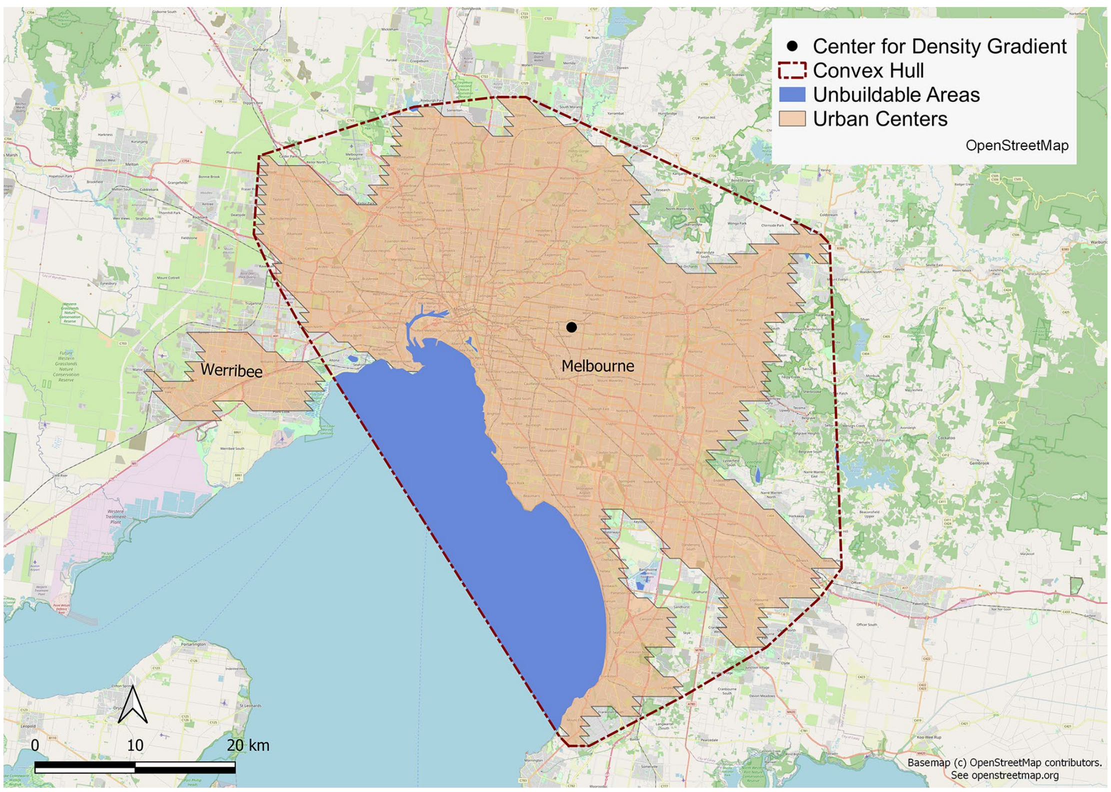


Figure 1.2. The convex hull of Melbourne, the unbuildable areas within the convex hull, and the center for density gradient

Base map and data from OpenStreetMap and OpenStreetMap Foundation

3.3 Analysis methods

After calculating urban form metrics in PostgreSQL, we assessed the correlations, and the results were plotted as a pairplot with the Seaborn package in Python. As a simple line does not produce a line of good fit in this case, non-parametric or Lowess smoothing was used to model the relationship between the variables that did not fit a predefined distribution and had a non-linear relationship. After standardizing the data by subtracting the mean and dividing by the standard deviation, we used an unsupervised machine learning algorithm, K-Means clustering, to identify a typology of cities (Jain, 2010). Cluster analysis is a descriptive tool with no objectively "right" number of clusters *k*. We chose *k* based on our substantive interpretation, increasing *k* until clusters were no longer qualitatively distinct. The other machine learning algorithm used in this research is random forest regression, which enables us to separate out the influence of urban form measures and show their associations with outcome variables.

4. Results and discussion

We begin with a descriptive analysis of how our different measures of urban form are distributed across various urban centers, how the measures correlate with each other, and how these correlations vary by income group and geographic region.

4.1. Correlation between urban form metrics

As might be expected, there is a strong positive correlation between the three density-related measures – weighted density, density gradient intercept, and density gradient slope. This correlation exists across all four income groups. For example, cities with high overall density also tend to have an even denser center and a steep gradient down to the urban periphery. Cities in high-income countries have the lowest weighted density, especially those in Northern America and Oceania, with cities in the other income groups having similar distributions. Cities in Northern America and Oceania also have the lowest density gradient slope and density gradient intercept compared to other areas, but European cities are more aligned with other income groups regarding the density gradient measures. This means that not all high-income countries follow similar trends in their urban form metrics, maybe because of different historical backgrounds and types of planning policies.

The other two measures, compactness and street connectivity, do not correlate as strongly with other metrics, and the correlation is different across income groups. Compactness is correlated with both density gradient metrics in upper-middle and lower-middle-income groups (Figure 1.3). Cities in Asia, mainly in Eastern and South-Eastern Asia, have the lowest levels of compactness compared to other regions, which means the ratio of built-up areas to the buildable areas in their convex hull is lower. Considering compactness along with density metrics can give us a sense of how sprawled or compact an urban area is.

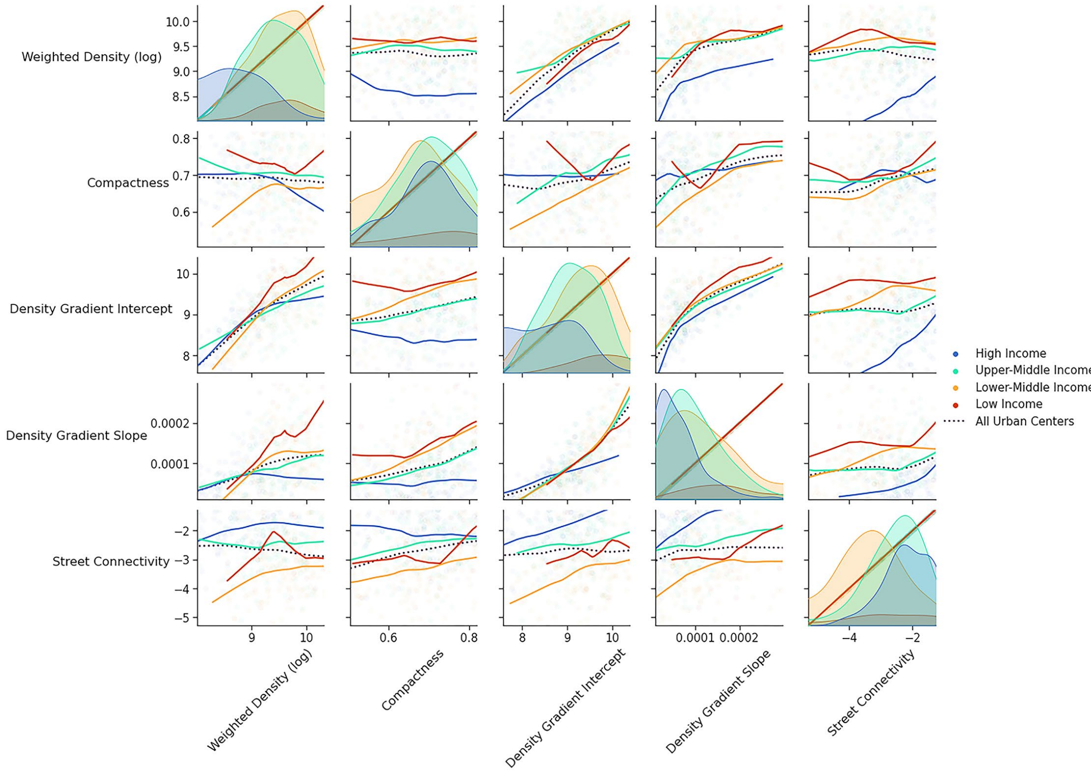


Figure 1.3. Correlation of urban form metrics across income groups

Street connectivity is correlated with density gradient metrics only in high-income countries, meaning that in these countries, having denser centers is associated with having more connected streets. This measure is lowest in cities located in lower-middle-income countries, especially in South-Eastern Asia and Southern Africa, and highest in high-income countries, particularly in Europe. Compared with cities in the other three income groups, cities in high-income countries display different characteristics. Having a correlation between different metrics can suggest that these metrics can be used interchangeably in those areas and for example if one of them is high, there is a high chance that the other correlated metric is high, too.

4.2. Typology of the cities

Cluster analysis captures some aspects of urban form that are not evident on the individual measures and distinguishes the cities with similar characteristics. The K-Means cluster analysis reveals five distinct types of city, as depicted in the radar plot of the cluster centroid (Figure 1.4). In the first cluster, which we call "High Density", all metrics are high, and except for street connectivity, are the highest of any other group.

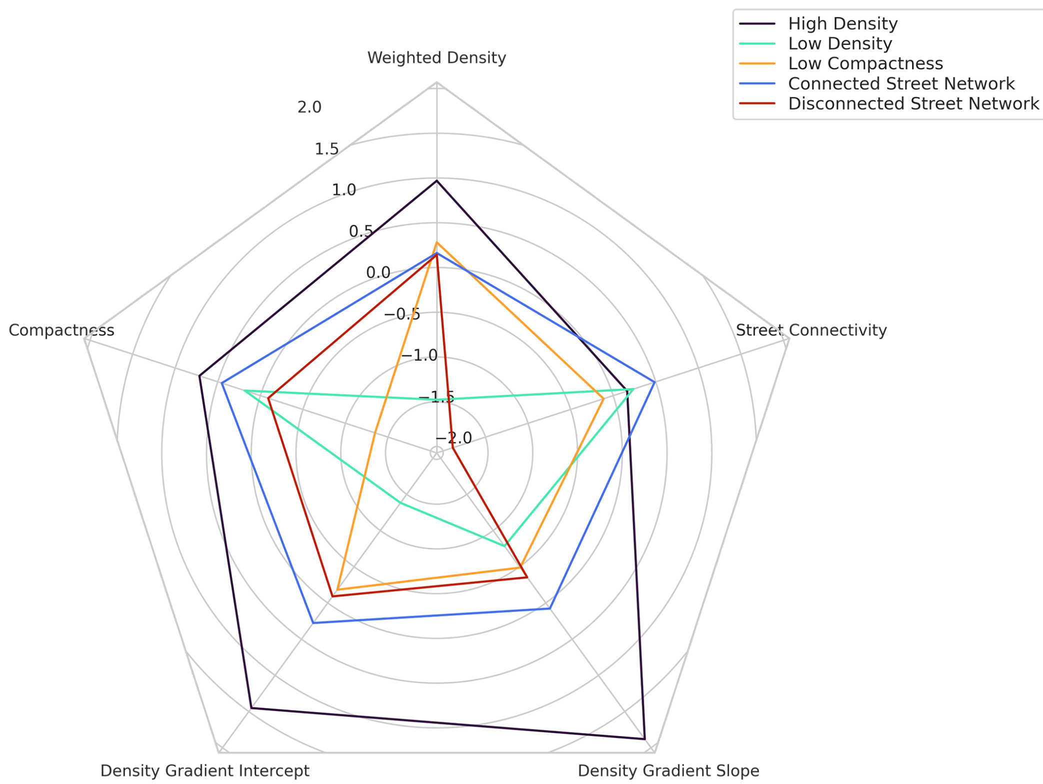


Figure 1.4. Radar plot of the cluster centroids

The second cluster, "Low Density", accounts for 18 percent of the cities and has the lowest weighted density, density gradient slope, and density gradient intercept. Their compactness and street connectivity are moderate, indicating that these low-density cities can also be compact and have connected streets.

The third cluster, "Low Compactness", has the lowest compactness among the clusters. Weighted density of the cities in this group is ranked second highest, and their density gradient and street connectivity is placed in the second lowest rank; showing that high weighted density does not necessarily lead to more compact cities. Agricultural lands, forests, and parks are calculated as potentially buildable areas and in cities like Rio de Janeiro, Brazil, Hong Kong, and Singapore, these protected areas may be one of the reasons for the low compactness metric. In some cities, such as Busan, South Korea, the irregular boundary shape might cause low compactness.

The fourth cluster, "Connected Street Network", includes 36 percent of the cities examined in this study, the largest number among all clusters. It is characterized by highly connected streets. The other metrics are at a moderate to high level compared to other clusters; density gradient and compactness are in the second rank.

The fifth cluster, "Disconnected Street Network"', has very low street connectivity and has the second lowest weighted density and compactness among other groups. The lowest number of cities among all clusters is in this group, with about 11.5 percent of the total large cities.

4.3. Geographic distribution of the clusters

While there is an example of each cluster in almost every world region, there are clear geographic patterns (Figures 1.5 and 1.6). The majority of the cities in the "High Density" cluster are located in Asia (71 percent), including 35 percent in India. There is only one European city in this cluster – Brussels, Belgium – which still has the lowest weighted density in this group but one of the highest connected streets among all urban centers in this study. The presence of this European city in the "High Density" group might be related to the postwar large-scale demolition of townhouses for development of the high-rise business district which symbolizes the transformation of Brussels from national capital to Europe's political capital (Baeten, 2001). Brussels' high level of density gradient slope is also related to this transformation which makes it different from most European cities.

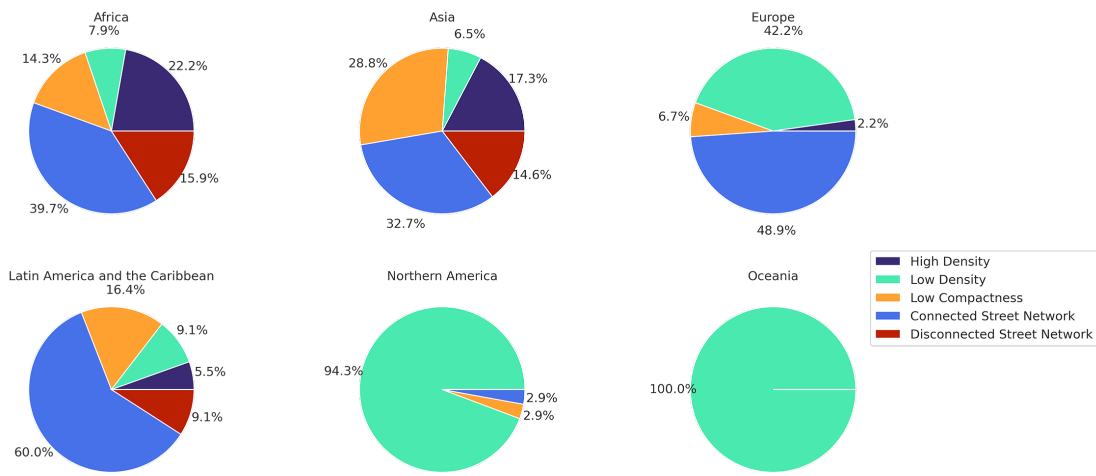


Figure 1.5. Distribution of clusters in different geographic regions

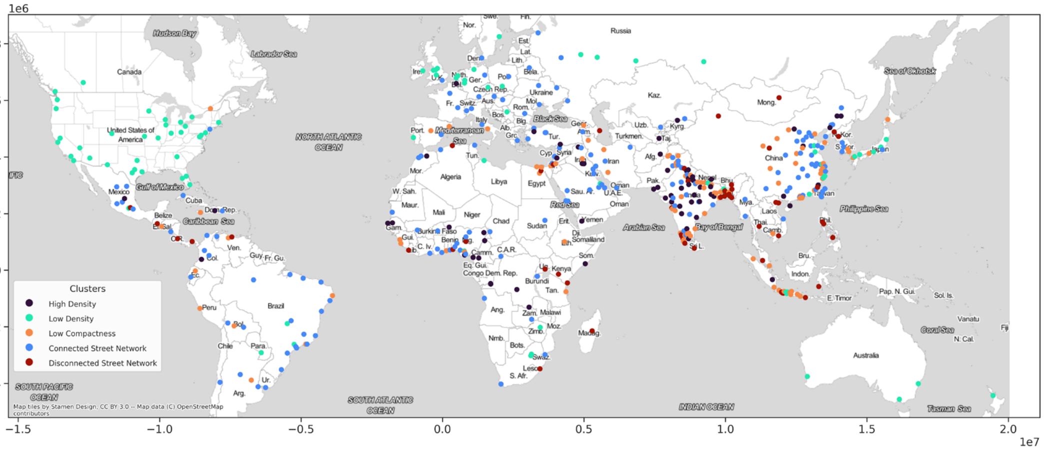


Figure 1.6. Typology of cities based on urban form metrics

Base map and data from OpenStreetMap and OpenStreetMap Foundation

75 percent of the cities in the “High Density” cluster are in low-income and lower-middle income countries. Kathmandu, one of the representatives of this cluster (i.e., a city close to the cluster centroid), is located in Nepal, one of the ten fastest urbanizing countries in the world. The traditional core area of Kathmandu consists of a densely built area including narrow streets which gradually shifts towards the periphery surrounded by a vast expanse of agriculture-dominated rural area. The population density is increasing rapidly, particularly in the Kathmandu Valley, along the main highways, and close to the border with India (Bakrania, 2015; Timsina et al., 2020). Surat, India, Najaf, Iraq, and Kinshasa, Democratic Republic of the Congo are other representatives of this cluster.

North America, especially the U.S., has the largest portion of cities in the “Low Density” group. Policies such as single-family zoning and automobile-oriented developments are two of the drivers of this low level of density. About a quarter of the cities in this cluster are in Europe, and 21 percent are in Asia (mainly in Eastern Asia and South-Eastern Asia including China, Japan, and Indonesia). 92 percent of the cities in this cluster are in high and upper-middle-income countries, suggesting that the income level of the countries affects the density of the cities. Two cities in this group, Indianapolis and Kansas City, both in the U.S., have the lowest levels of weighted density among all the cities studied. Los Angeles, the U.S., Sydney, Australia, Manchester, the U.K., and Toronto, Canada are among the representatives of this cluster.

77 percent of the cities in the "Low Compactness" group are located in Asia (mainly in India, China, and Indonesia). Just one city from Northern America—Montreal, Canada—is present in this group, mainly because of the presence of large green spaces within the convex hull, and the irregular shape of its boundary. Compared to other clusters, "Low Compactness" cities are more evenly distributed across income groups; 56 percent of the cities are in high and upper-middle income and 44 percent in low and lower-middle-income countries. Barcelona and Madrid, Spain, Tehran, Iran, Jakarta, Indonesia, and Lima, Peru are some representatives of this cluster.

More than half of the cities in the "Connected Street Network" cluster are located in Asia, and just one city is in Northern America (New York, U.S.). New York has the lowest density gradient slope and intercept among the cities of this cluster, but has the highest weighted density compared to other Northern American cities overall. While many other U.S. cities such as Boston and Philadelphia have gridiron street patterns in their central areas, they do not also have the relatively high density and compactness that defines cities in the “connected street network” cluster.

Beyond the U.S., about 70 percent of the cities in this cluster are in high- and upper-middle-income countries. Some metropolitan areas such as Istanbul, Turkey, Cape Town, South Africa, São Paulo, Brazil, London, the U.K, Paris, France, Tokyo, Japan, Isfahan, Iran, Rome, Italy, Moscow, Russia, Doha, Qatar, Seoul, South Korea, and Shanghai and Beijing, China are among the representatives of this cluster.

70 percent of the cities in the “Disconnected Street Network” cluster are located in South-Central Asia, South-Eastern Asia, and Eastern Asia, with India, Bangladesh, Indonesia, China, the Philippines, and Vietnam respectively at the highest levels. This cluster has no representatives in Europe, North America, or Oceania. 75 percent of the cities in this cluster are in low- and lower-middle-income countries, indicating that the income level of the countries can influence their street connectivity. One of the main reasons for the presence of South-Eastern Asian cities such as Manila, the Philippines in this cluster are gated communities, which are developer-driven settlements with access centered around the private automobile formed with the rise of a new middle-class in response to fears of crime, inadequate public services, and weak land-use regulation. Poor street connectivity can be seen as a way to generate social and physical exclusivity (Barrington-Leigh & Millard-Ball, 2020).

In Latin America and the Caribbean cities such as Guatemala City, Panama City, and San Salvador, the combination of weak planning institutions and high rates of crime has resulted in an increase in gated communities. For example, Guatemala City was originally established on a geometric grid under Spanish colonial rule, but has been transformed into a city characterized by informal settlements and by gated communities for the wealthy that are isolated from the rest of the city (Barrington-Leigh & Millard-Ball, 2020). The presence of African cities such as Lagos, Nigeria and Kampala, Uganda in this group is mainly due to the fast urbanization rate and large number of informal settlements as well as gated communities in these metropolitan areas (Ilesanmi, 2012; Vermeiren et al., 2012).

Africa, Latin America and the Caribbean, and China have representatives in all five clusters, but the majority of them are among the "Connected Street Network" group. Most Indian cities are in the "High Density" cluster. Most of the cities in Western Asia are among the cities in the "Connected Street Network" cluster. In South-Central Asia, the connectivity of street networks decreases from west to east. "Disconnected Street Network" and "Low Compactness" clusters represent most of the cities in South-Eastern Asia. No European cities are in the "Disconnected Street Network" group. The majority of European cities are in the "Connected Street Network" and "Low Density" clusters.

4.4. Environmental outcomes of urban form metrics

We now consider how the measures of urban form correlate with two environmental outcomes – access to green space and local air pollution. In addition to the scatter plots, the box plots are used to show the differences between our five clusters and highlight the outliers that are not evident from the scatter plots.

4.4.1. Urban form and access to green space

Not surprisingly, green space per capita is negatively correlated with all density metrics (Figure 1.7). After all, parks, housing with private gardens, and other green spaces will tend to reduce densities. Indianapolis, Kansas City, and Tampa have both the lowest levels of weighted density and the most green space per capita. However, there is no correlation between green space per capita and street connectivity, except in high-income countries. Nor is there a correlation with compactness.

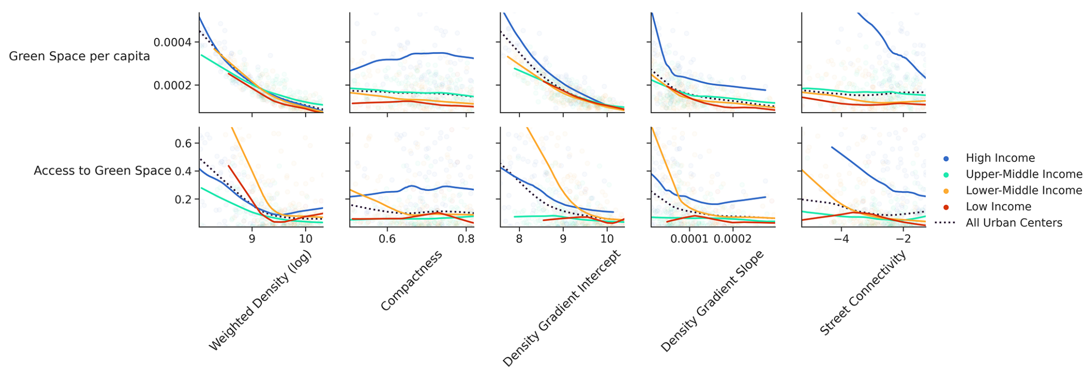


Figure 1.7. Correlation of urban form metrics with green space measures across income groups

In high-income countries, which have the highest levels of green space per capita worldwide, cities with lower density and less connected street networks tend to have higher green space per capita. In this income group, North American cities have much higher green space per capita compared to European and Asian cities. As shown in Figure 1.8, green space per capita is higher in the “Low Density” cluster, possibly because single-family zoning in American cities and other sprawling countries provides residents with larger areas of green space in the form of private gardens, but at the expense of reducing densities. The lowest level of green space per capita is found in cities in the “High Density” cluster and also in the countries with lower income, but the other three clusters have similar trends regarding green space per capita.

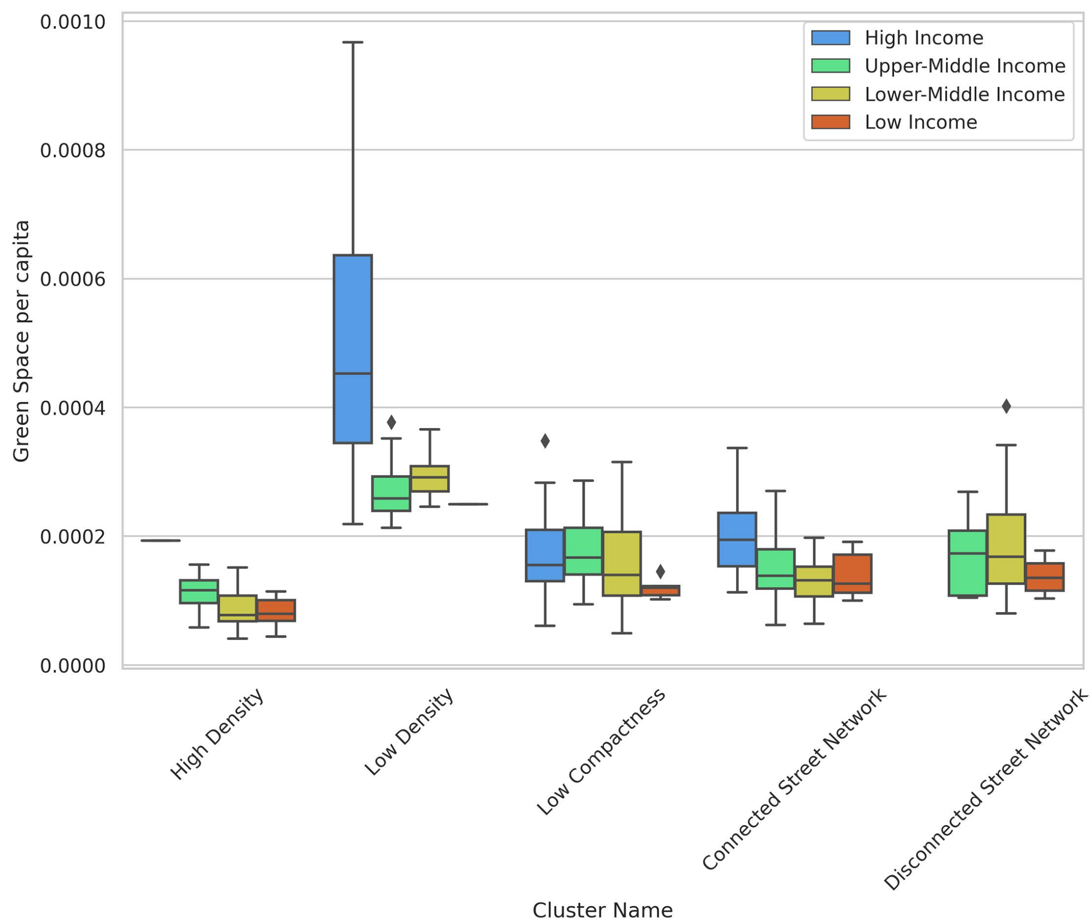


Figure 1.8. Box plot - green space per capita vs clusters across income levels

While in general, there is a negative correlation between density and green space per capita, some outliers suggest that it is possible to have both. Brussels, Belgium is an outlier in the “High Density” group which has a much higher green space per capita compared to the rest of the cities in this group. Based on our results, Brussels is one of the cities with the lowest degrees of sprawl.

A person’s access to green spaces is not necessarily determined by the amount of green space per capita, but rather by the percentage of the population that lives in areas with a high green cover. Therefore, we refer to the share of population that resides in high green areas as the “access to green space” metric. This measure is less affected by urban form metrics compared to green space per capita and the difference between various income groups is more significant. Lower-middle-income countries have the highest share of people living in high green areas. Cities in Bangladesh and India such as Noakhali and Thalassery are among the ones with the highest access to green space.

“High Density” and “Connected Street Network” clusters have lower levels of access to green space compared to other clusters (Figure 1.9). However, there are a large number of outliers such as Benin City, Nigeria; Yangon, Myanmar; Havana, Cuba; and Russian and Ukrainian cities from the “Connected Street Network” cluster, as well as Mbuji-Mayi, Democratic Republic of the Congo and Brussels from the “High Density” cluster. These cities have high levels of access to green space, but low green space per capita.

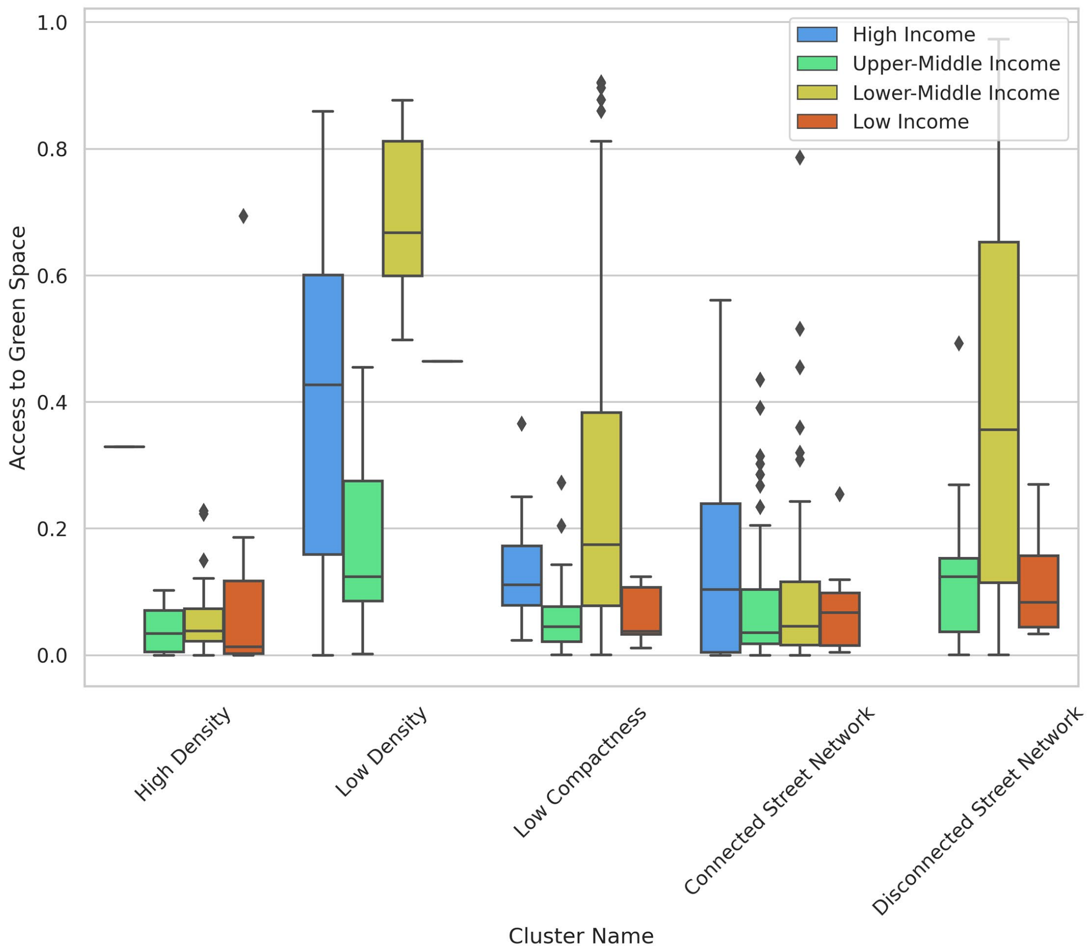


Figure 1.9. Box plot - share of population living in high green areas vs clusters across income groups

4.4.2. Urban form and air pollution

Two air pollution measures are examined in this study: $PM_{2.5}$ transportation emissions per capita and $PM_{2.5}$ concentration (an indicator of exposure). $PM_{2.5}$ transportation emissions per capita is higher in cities of high-income and upper-middle-income countries compared to the rest of the world. According to Figure 1.10, there is no noticeable relationship between urban form metrics and transportation emissions per capita overall. There is more variation in the emissions exposure between various income groups. $PM_{2.5}$ concentration is slightly correlated with weighted density in high-income and upper-middle-income groups and has no significant correlations with other urban form measures.

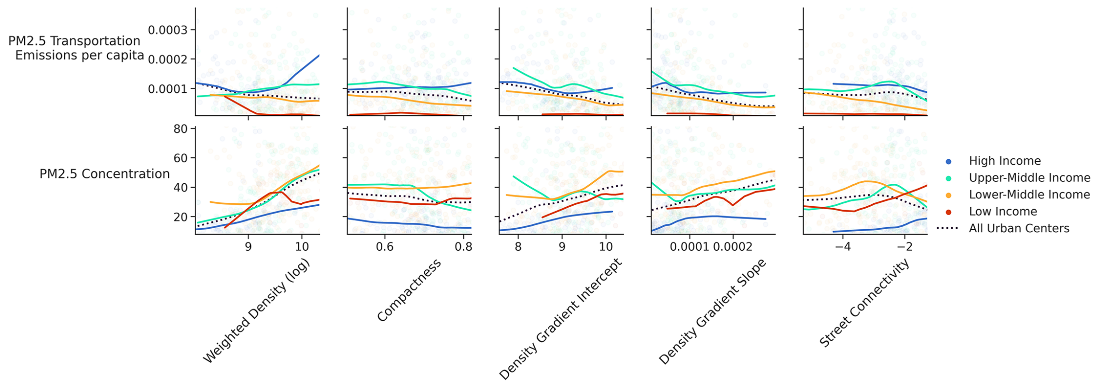


Figure 1.10. Correlation of urban form metrics with $PM_{2.5}$ measures across income groups

The boxplots illustrate some other aspects of how urban form metrics affect $PM_{2.5}$ measures. Cities with higher density have lower per capita transportation emissions (less driving), but higher $PM_{2.5}$ exposure because more people are living closer to emissions sources. Cities in the “Connected Street Network” cluster have lower transportation emissions per capita, but higher exposure to emissions compared to the cities in the “Disconnected Street Network” (Figure 1.11 and 1.12). Taichung, Taiwan is an outlier in the “Connected Street Network” cluster with high per capita transportation emissions. The main outlier in the “Low Density” cluster is Ontisha, Nigeria which has one of the lowest levels of transportation emissions per capita, but highest levels of emissions exposure. Giannadaki et al. (2016) reported that most of the emissions in Nigeria could be linked to natural sources such as Saharan desert dust or wildfires. Wood burning, vehicular emissions, particles generated from cooking using fossil fuels, firewood, industrial emissions, and unsustainable open waste dumpsites are other causes of high levels of emissions exposure in Nigerian cities (Opara et al., 2021).

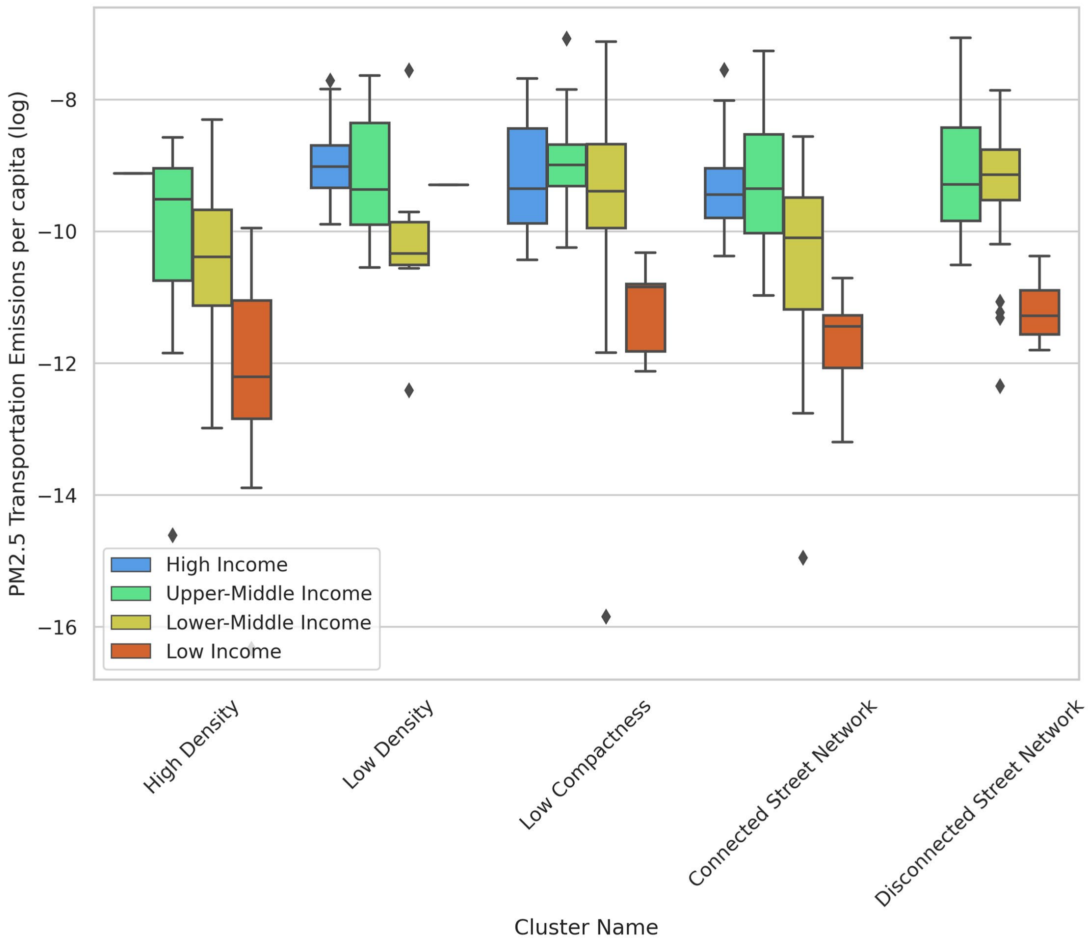


Figure 1.11. Box plot - $PM_{2.5}$ transportation emissions per capita vs clusters across income groups

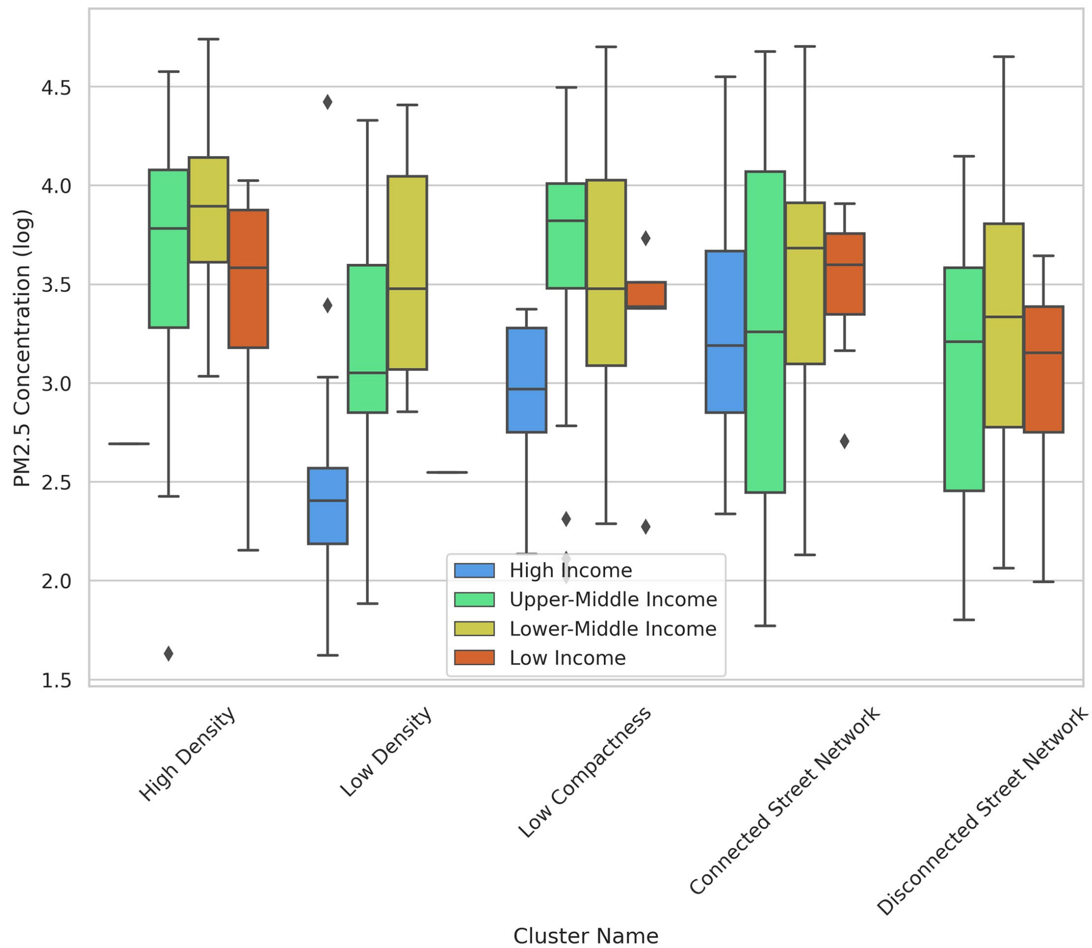


Figure 1.12. Box plot - $PM_{2.5}$ concentration vs clusters across income groups

The high-income group of the "Low Density" cluster has two outliers with high levels of emissions exposure: Katowice, Poland and Dammam, Saudi Arabia. In Katowice, vehicular emissions account for a considerable part of the total air pollution (Widziewicz & Loska, 2016). Alwadei et al. (2022) demonstrate that dust storms account for 42% of $PM_{2.5}$ mass concentration in Dammam. In addition to dust storms, other sources of atmospheric aerosols in Saudi Arabia include heavy oil combustion, resuspended soil, industrial emissions, traffic emissions and marine sources (Munir et al., 2016). Cities such as Doha, Qatar, Dubai, United Arab

Emirates, and Manama, Bahrain have high levels of $PM_{2.5}$ exposure due to similar drivers.

Among Northern American cities, Vancouver, Calgary, Miami, Portland, and Seattle have the lowest levels of $PM_{2.5}$ concentration. Stockholm, Sweden, Dublin, Ireland, and Copenhagen, Denmark, have the lowest exposure among European cities. These cities follow smart growth policies trying to reduce urban sprawl impacts.

4.5. Random Forest regression

As a final step, we have used random forest regression that helps with separating out the influence of each measure and showing its association with the outcome variables. When the data has a non-linear trend and extrapolation outside the range of the training data is not important, we can use random forest regression. We used the SHAP value as a united approach to explain the output of our random forest regression model. SHAP assigns each feature an important value for a particular prediction (Lundberg & Lee, 2017). On the beeswarm plots (Figure 1.13), the features are ordered by their predictive power, but we can also see how higher and lower values of the feature will affect the result. Each small dot on the plot represents a single observation. The horizontal axis represents the SHAP value and shows negative or positive effects of the variables, while the color of the point shows us if that observation has a higher or a lower value on the predictor variable, when compared to other observations.

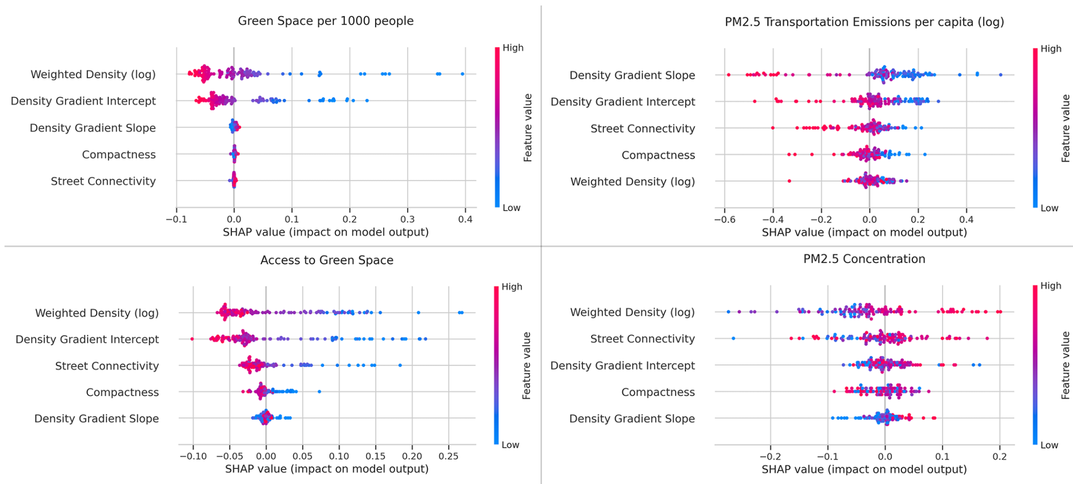


Figure 1.13. Random Forest regression - SHAP values for all environmental outcome measures

Density (weighted density and density gradient intercept) have the strongest association with green space. High values of weighted density and density gradient intercept have a high negative contribution to the level of green space per capita, while low values of these metrics have a high positive contribution. Similarly, access to green space has a negative correlation with weighted density, density gradient intercept, and street connectivity; however, the impact is less than the impact on green space per capita.

Density gradient slope and intercept followed by street connectivity are the best predictors of $PM_{2.5}$ transportation emissions per capita. Higher values of all three reduce $PM_{2.5}$ transportation emissions per capita. More compact cities also lead to lower transportation emissions per capita; however, the impact of compactness is not as high as street connectivity. Moreover, areas with higher weighted density and density gradient slope are likely to be more exposed to $PM_{2.5}$ emissions. The results of the SHAP values are consistent with the earlier analysis.

5. Conclusion

Improving our understanding of urban form is essential for progress towards smarter and more sustainable growth. An understanding of urban patterns, dynamic processes, and their relationships is a key objective of urban research, and urban management requires available and detailed information about the processes and patterns of urbanization. In addition, it allows urban planners and policymakers to decide what policy measures are needed to address problems caused by urban sprawl or to prevent future issues caused by their decisions about urban form.

In this study, we compared various urban form metrics, including weighted density, density gradient (slope and intercept), compactness, and street connectivity, using the GHSL dataset, and examined how they co-vary with each other and across geographic regions and socioeconomic characteristics. The differences in the metrics and their correlations among different geographic regions and income groups show that urban form metrics should be chosen in accordance with the characteristics of cities, and it is inappropriate to generalize results from one income group or geographic region to others.

We then developed a typology of urban forms, which resulted in five different clusters named “High Density”, “Low Density”, “Low Compactness”, “Connected Street Network”, and “Disconnected Street Network”. While there is an example of each cluster in almost every world region, there are clear geographic patterns. Perhaps unsurprisingly, "High Density" cities are largely found in Asia and "Low Density" cities in North America, but our results indicate that the majority of “Low

Compactness" and "Disconnected Street Network" cities are located in South-Central, South-Eastern and Eastern Asia.

Our study indicates that urban form is correlated with income level and its metrics can vary depending on the country's income level. Further, our findings demonstrate that density does not provide a complete picture of urban form, and multiple indicators are required. Density is often used as a proxy for multiple dimensions of urban form, and our results suggest that this is a reasonable approach in many high-income countries such as the U.S., but not appropriate in other parts of the world where cities can have (for example) high density but low compactness and poorly connected streets.

Density and street connectivity are more helpful for understanding urban form compared to compactness. Although compactness has been widely used as an urban form metric, it is not as informative as density metrics or street connectivity with the prevalent methods used to calculate it. Compactness is also very sensitive to the definition of urban boundary, which makes it more difficult to interpret and generalize. In cases such as Hong Kong, assuming agricultural lands, forests, and parks as potentially buildable areas can lower the compactness. However, in cases like Busan, South Korea, irregular boundary shape might lead to lower levels of compactness.

In the next step, we examined the relationship between our typology and two important policy outcomes: access to green space and air pollution, specifically $PM_{2.5}$. National income and geography have the largest impacts on these two measures, but our focus here is on the impacts of urban form after controlling for these factors. The results show that there is usually a trade-off between density and access to green space. Density (all three variables) have the strongest association with green space. Higher weighted density also increases $PM_{2.5}$ concentration and thus exposure. Street connectivity is associated with lower $PM_{2.5}$ emissions. It is therefore important that future urban planners and policy makers increase green space access in high-density, compact urban areas, which will help alleviate health problems resulting from higher emissions exposure.

Thus, while density is often emphasized as the way to reduce driving and thus $PM_{2.5}$, it comes with a downside: less green space access, and more exposure to $PM_{2.5}$. Moreover, street connectivity actually has a stronger association with reduced $PM_{2.5}$ emissions from the transportation sector, and does not have the same tradeoff with exposure to air pollution or access to green space. A combination of moderate densities and connected streets, therefore, may enable cities to reduce car dependence while maintaining the health and quality of life for their residents.

Acknowledgments

We are grateful to Christopher Barrington-Leigh for his contributions in providing the street network connectivity dataset and for his excellent comments during the research. We also thank Chris Benner and Carlos Dobkin for their helpful comments on earlier drafts of this paper. This article relies on the Global Human Settlements Layer (GHSL) from the European Commission Joint Research Center.

CHAPTER 2

Do Motorways Cause Urban Sprawl? Evidence from Four Global Contexts

Abstract

This study examines the impact of motorways on urban sprawl, focusing on two key urban form metrics—density and street connectivity—across diverse global contexts. The research addresses the existing gap in understanding the relationship between urban sprawl and motorway development. We contextualized the study within the broader urban development literature, considering traditional models, induced travel concepts, and existing empirical research. Case studies from Turkey, Ireland, Thailand, and Taiwan—selected based on diverse urban sprawl characteristics—provide insights into regional variations in the impact of motorways on urban form. Methodologically, the study employs Street-Network Disconnectedness Index (SNDi) for street connectivity and population density for density measurements. By temporally separating the effects and using a panel fixed effects model, we mitigated concerns related to endogeneity and better discern the causal impact of motorways on street connectivity and density. The findings indicate a strong impact of the construction of motorways in reducing density and street connectivity in Thailand. In Turkey, we find similar results, but the effects are smaller and generally not statistically significant. In Taiwan, our results are inconclusive, while in Ireland and Northern Island, the results are opposite of what we might expect. The conclusion highlights the context-specific effects of motorways on urban development and the need for further research, suggesting avenues for expanding case studies and

conducting comparative analyses between the impact of intercity and intracity highways on urban sprawl. Overall, the study broadens our understanding of the relationship between motorways and urban form on a global scale.

Keywords: Motorways, Urban sprawl, Density, Street connectivity, Panel Fixed Effects

1. Introduction

Transport infrastructure development plays a pivotal role in shaping urban form—the physical and spatial characteristics that define the layout of a city. As cities grapple with the challenges of accommodating growing populations and expanding economic activities, it becomes paramount to understand how specific elements of transportation infrastructure, such as roads, influence urban form. This understanding is particularly crucial for effective policy intervention as recent research underscores the significant influence of highways in altering the dynamics of cities. Highways have been found to impact various aspects of development and land use, extending beyond physical connectivity to influence patterns of population distribution, employment trends, and even local zoning policies (Baum-Snow, 2007; Duranton & Turner, 2012; M. À. Garcia-López, 2019; M. Á. Garcia-López et al., 2015). Added highway capacity can also lead to induced travel, facilitate faster car travel, and ultimately result in lower-density, more car-oriented development (Milam et al., 2017).

This study analyzes the impacts of motorways—controlled-access highways designed for high-speed vehicular movement—on urban form characteristics, particularly the phenomenon of urban sprawl. Due to its adverse environmental, social, and economic consequences, urban sprawl is regarded as the undesirable display of urban form. While various definitions and facets characterize sprawl, it commonly refers to a process resulting in low-density, dispersed, leapfrog, or isolated development, unplanned and haphazard land use, an abundance of commercial strip malls, and insufficient transportation and housing alternatives, coupled with a high reliance on automobiles at the periphery of urban areas (Angel et al., 2016; Hamidi et al., 2015; Rezaei & Millard-Ball, 2023; Wilson & Chakraborty, 2013).

Previous research has yielded conflicting conclusions in this area. While some studies posit a positive causal link between highway expansion and sprawling urban development over time, others find no significant impact on urban sprawl (Narayani & Nagalakshmi, 2023; Pratama et al., 2022; Zhang, 2001). Our study aims to address this discrepancy by ascertaining whether a positive causal relationship exists between motorways and urban sprawl, particularly in parts of the world where limited research has been conducted in this field.

Despite the importance of understanding the role of highways in inducing urban sprawl and their correlation with land development characteristics, our knowledge in this area remains limited, especially in countries outside Europe, the United States, and China. To bridge this research gap, our study explores the complex interaction between motorways and urban sprawl across diverse contexts. Through a comprehensive analysis of two urban form metrics—density and street connectivity—and time-series data of motorways, we shed light on the nuanced relationship between motorway infrastructure and the expansion of their surrounding urban areas. Our central research question explores how the construction of new motorways influences urban sprawl across different regions of the world, offering insights into the varying impacts of highway development.

In the following sections, we first provide a comprehensive review of relevant literature. Subsequently, we outline the methods and datasets used in this research, detailing the analytical framework. Finally, we present and discuss the results, drawing conclusions that contribute to our understanding of the dynamics between motorways and urban form.

2. Literature review

Traditional urban models, such as the monocentric model and bid-rent theory, offer foundational perspectives on the relationship between transportation infrastructure, land use patterns, and urban development processes. The monocentric model, developed by economist William Alonso, envisions a city with a central business district (CBD) and proposes a radial pattern of decreasing land prices and rents as one moves away from the CBD, reflecting the increasing cost of commuting (Arnott & McMillen, 2006). An extension of this model, the bid-rent theory, explores competition among land users and suggests that proximity to the CBD influences the maximum rent or price users are willing to pay (Arnott & McMillen, 2006).

These urban models provide a theoretical foundation for understanding how the development of highways may impact urban form. Highways facilitate efficient and rapid transportation, lowering the overall cost of commuting for individuals residing farther from the CBD, especially by alleviating time-related expenses linked with travel. As a result, the traditional radial pattern of decreasing land prices in monocentric models might experience modifications due to the accessibility facilitated by highways. Therefore, highways are likely to contribute to lower population densities in central areas and potentially expand the size of urban areas, as people find it more feasible to reside at greater distances from the CBD (Cervero, 2003; Levkovich et al., 2020; Noland & Lem, 2002).

Beyond these traditional models, the concept of induced travel enters the discussion. Rooted in transportation planning and economics, induced travel suggests that the construction or expansion of transportation infrastructure can stimulate increased travel demand, subsequently influencing various aspects of urban development, including congestion, land use patterns, and environmental considerations (Cervero, 2003; Cervero & Hansen, 2002; Volker et al., 2020).

The majority of empirical research supports these theoretical predictions and indicates that highways do lead to suburbanization and lower-density development. Government investment in infrastructure, particularly highway construction, is frequently cited as a key driver of sprawl. However, exceptions and regional variations challenge this consensus.

Studies exploring the impact of highway development on urban sprawl operate on diverse scales, ranging from national analyses to examinations of specific metropolitan areas. These investigations consider both intercity and intracity highways and employ a variety of metrics to measure the extent and nature of sprawl.

The nuanced nature of findings across studies may be attributed to varying definitions of urban form, distinctions in defining and categorizing roads, and the diverse scales employed in analysis.

Starting with the U.S. context, Baum-Snow's (2007) analysis of post-World War II suburbanization reveals a significant relationship between new highway construction and decreased central-city population growth. The interstate highway system, outlined in a 1947 plan, played a substantial role in shaping suburbanization patterns. Similarly, Milam et al. (2017) emphasize the long-term effects of highway growth in the United States, highlighting changes in land use development patterns characterized by dispersed, low-density arrangements that are heavily automobile-dependent. Moreover, Volker et al. (2020) discuss the flawed approach of expanding roadways solely to reduce congestion and emissions, emphasizing the phenomenon of induced travel. They reveal that the expansion of highway capacity can lead to increased vehicle travel due to decreased travel time costs, contributing to additional vehicle miles traveled (VMT) and influencing long-term land use patterns.

Contrastingly, Zhang's (2001) study on transportation accessibility and new housing in Chicago challenges the conventional perspective. His findings show limited impact and even a negative correlation between increased transportation accessibility, particularly from publicly-funded highway construction, and new developments in suburban areas. This indicates that the relationship between highways and suburbanization may not be universal and may vary across regions.

Burchfield et al. (2006) provide a nuanced examination of transportation's influence on urban sprawl. Analyzing major road density in the urban fringe across 275 U.S. metropolitan areas, they find no significant impact on sprawl patterns despite increased road density. Contrary to common beliefs, the determining factor for sprawl is not road density but whether the city center predates the dominance of automobiles. Essentially, Burchfield et al. (2006) challenge the prevailing narrative, suggesting that major road density may not be a key factor in urban sprawl dynamics across U.S. metropolitan areas.

Shifting focus to the European context, Garcia-López (2019) examines the impact of highways on urban sprawl across 579 European cities, revealing a positive correlation between highway stock and residential land area growth. The study notably utilizes the length of the highway network at the metropolitan level to explain patterns of sprawl. Oueslati et al. (2015), examining 282 European cities, also identify a positive correlation between highway density and developed land size, although no significant effect on land fragmentation is observed.

In the Asian context, Deng et al. (2008) find that highway density, along with various economic and demographic factors, significantly influences urban expansion patterns in China. Expanding this discourse, Pratama et al. (2022) delve into the causal relationship between highway expansion and urban sprawl in Indonesia's Jakarta Metropolitan Area (JMA). Their findings indicate that improved highway access contributes to increased urban sprawl, characterized by scattered developments, particularly within a 40-50 km radius from the city center. Moreover, intracity highways are identified as catalysts for urban expansion, defined as built-up land growth, in certain districts of the JMA over the past three decades.

The synthesis of these studies reveals points of consensus, indicating a general trend of highways contributing to suburbanization and lower-density development, particularly in the U.S. and Europe. However, exceptions and variations exist, influenced by factors such as geographic location, economic conditions, and methodological differences in defining and measuring urban sprawl. The literature review also underscores the importance of considering induced travel, different definitions of urban form, and analysis scales in understanding the nuanced relationship between highways and urban development.

In our study, we contribute to this body of research by adopting a global perspective, examining the impact of intercity highways in various regions on population density and street connectivity in grid cells near urban centers. This broad geographical scope aims to provide a more comprehensive understanding of how motorways influence urban development, especially in less studied regions.

3. Methods

3.1. Measures of urban form

Urban form encompasses the physical attributes of built-up areas such as their shape, size, density, and configuration. Urban sprawl—an undesirable manifestation of urban form—has been studied through various metrics, each capturing different aspects of the sprawling process. The choice of these metrics is often influenced by factors such as the scale of analysis, data availability, and the specific aspects of urban form impacting the study. In our study, we focus on two key measures, street connectivity and density. The selection of these metrics is driven by their relevance to the intricate dynamics of urban sprawl and the aspects of urban development that are particularly susceptible to the influence of motorways. Moreover, they are both widely employed in the literature and amenable to the scale of our analysis considering the availability of data.

3.1.1. Street connectivity

Street connectivity is a fundamental aspect of urban form, reflecting the layout and accessibility of road networks within a city. Prior research underscores a notable correlation between disconnected urban street networks and heightened vehicular travel, energy consumption, and CO_2 emissions (Barrington-Leigh & Millard-Ball, 2017, 2019; Marshall et al., 2014).

In our study, we employ the Street-Network Disconnectedness Index (SNDi), a metric designed to capture the degree of street network sprawl. This index, introduced by Barrington-Leigh and Millard-Ball (2019), was derived from a large dataset encompassing intersections and edges, enabling the computation of nodal degree for each intersection, along with dendricity, circuity, and sinuosity. Dendricity serves to ascertain whether a network exhibits tree-like attributes, featuring a solitary feasible route to access specific nodes. Circuity pertains to the ratio of path length to the straight-line distance between each node and neighboring nodes. Sinuosity represents the ratio of edge length to the end-to-end distance. By using principal components analysis, empirical weights were assigned to the aggregate metrics, with primary focus placed on the first principal component, which serves as an overall measure of street-network sprawl and is denoted as the Street-Network Disconnectedness Index (SNDi).

The choice of SNDi is motivated by its ability to encapsulate diverse dimensions of street connectivity and its compatibility with our global analysis. It requires no further computations within the scope of this study. We invert the sign of SNDi for consistency with our other metric, where higher values now indicate increased street connectivity, providing a clearer understanding of the influence of motorways on urban sprawl dynamics.

3.1.2. Density

Density, conventionally calculated as the population within a specific geographical area divided by the corresponding land area, serves as a key metric in comprehending urban form. It is often considered one of the primary indicators of sprawl due to its ability to capture an essential dimension of sprawl and is commonly used because of its ease of measurement (Hamidi & Ewing, 2014). Churchman (1999) suggests density is the most intuitive characteristic of urban form and a widely recognized indicator of urban sprawl (Galster et al., 2001; Lowry & Lowry, 2014).

Despite its widespread use, simple density may not fully capture the density experienced by individuals or businesses within a given area and can be influenced by arbitrary boundary lines and include non-residential spaces, prompting the proposal of alternatives like weighted density which provide a more robust measure that is less sensitive to the scale and the inclusion of non-urban land (Duranton & Puga, 2020).

However, for our study conducted at the grid cell level, where the scale is relatively small (one kilometer), simple density provides a meaningful representation of urban density patterns.

3.2. Identification strategy: panel fixed effects model

The potential endogeneity concerns arise from the relationship between motorways and urban sprawl. Motorways are likely to affect urban form. Moreover, in cities characterized by lower population density, there is a greater tendency for policymakers to prioritize the construction of motorways to accommodate the dominant car-centric infrastructure. Therefore, there is a dual causation dilemma: on one hand, motorways can be seen as a response to existing urban sprawl and low density, and on the other, they may contribute to further sprawl by facilitating easier suburban accessibility. This two-sided influence makes it challenging to disentangle the causal relationship between motorways and urban form metrics. Therefore, when investigating whether motorways cause urban sprawl, it becomes imperative to address potential endogeneity concerns by employing research designs to isolate the actual causal impact of motorway development on urban form.

We adopt a panel fixed effects model as our identification strategy to navigate potential endogeneity concerns, particularly in the context of street connectivity and density. Our rationale is grounded in the assumption that future land use does not affect plans for motorways. However, we do allow for past land use decisions to affect motorway building. The establishment of motorways is typically driven by economic, regional planning, and transportation considerations that vary across countries. The decision to construct motorways is less contingent on the specific urban development patterns of cities and more influenced by the existence of significant urban centers serving as crucial hubs for economic activities, population concentration, and transportation convergence. Therefore, concerns surrounding reverse causation and endogeneity hold limited relevance within the scope of our study. It is worth noting that these considerations might carry more weight in studies focused on assessing the effects of intracity highways.

Additionally, the temporal separation inherent in our analysis minimizes the likelihood of encountering endogeneity issues. The chronological sequence of motorway construction and the subsequent examination of grid cells before and after the road's establishment contribute to a robust identification strategy. By temporally separating the effects, we can mitigate concerns related to endogeneity and better discern the causal impact of motorways on street connectivity and density.

In our identification strategy, by examining the presence of roads temporally, with grid cells before the construction of the road acting as a control group, we aim to isolate the impact closest to the highways and observe its changes as the distance from the highway increases.

3.3. Case selection

Our four case study countries are drawn from different geographic regions and have varying urban sprawl characteristics, as measured by the Street-Network Disconnectedness Index (SNDi). We selected two regions characterized by lower sprawl levels—Turkey and Taiwan—and two regions with higher SNDi values—

Thailand and Ireland. The classifications provided by Angel et al. (2016) further supported our choice, categorizing these regions into the most and least sprawling areas.

Our choice of these four regions is substantiated by additional studies. The findings from Badhrudeen et al. (2022) reveal distinct clusters and groups within the road networks of our selected areas. Moreover, Rezaei and Millard-Ball's (2023) research highlights the more connected street networks in Turkey and Taiwan's metropolitan areas compared to other global regions. Istanbul, Turkey, emerges as a high-density urban center, Dublin, Ireland, falls into the low-density cluster, and Bangkok, Thailand, is classified within the disconnected street network group. Therefore, these four areas represent various clusters identified in the study.

3.4. Data sources

Our primary dataset for grid cells is the European Commission Joint Research Center's Global Human Settlement Layer (GHSL) 2023A data package (European Commission, 2023), providing extensive and temporally diverse geospatial information on global populations and built-up areas. The GHSL dataset identifies 13,135 Urban Centers (UCs) globally. A UC is conceptually akin to a city; however, its boundaries rely on population and built-up intensity rather than administrative boundaries. This dataset encompasses the years 1975, 1990, 2005, and 2020, presented at varying resolutions: 10m, 100m, and 1km. We obtain population density directly from the GHS-POP raster (part of the GHSL data package) at the $1km^2$ scale. We aggregated street connectivity to the same $1km^2$ grid cells. The scale of our analysis is at the grid cell level and we confined the grid cells to the boundaries of UCs.

The spatial analysis involves measuring the distance of each grid cell to major roads tagged as motorways in OpenStreetMap within each urban center. The primary dataset for our motorways is OpenStreetMap, which characterizes a motorway as a highway with restricted access, typically equipped with two or more lanes and an emergency hard shoulder. These controlled-access highways, synonymous with freeways, autobahns, expressways, interstates, and throughways (in the United States), are designed for high-speed vehicular traffic, and access is restricted to ramps.

Furthermore, the process of data collection on the construction dates of motorways in our selected regions was not straightforward. No centralized databases existed for this purpose, so we had to search across various online sources to compile the necessary information. One notable challenge was the scarcity of English-language sources, especially for regions such as Thailand, leading us to translate texts for relevant details. Moreover, in many instances, different sections of a single motorway were constructed in different years, further underscoring the complexities involved in assembling a comprehensive and accurate dataset for our study.

4. Results

In this section, we present the findings of our analysis for each selected region: Turkey, Ireland, Thailand, and Taiwan. We begin by providing descriptive statistics and graphically illustrating the relationship over time between development density and connectivity and motorways. Subsequently, we employ regression analysis to confirm the graphical trends, and we complement our results with pairplots, examining the correlation between urban form metrics and proximity to the nearest motorway in both pre- and post-motorway construction grid cells.

4.1. Turkey

With a strategic location bridging Europe and Asia, Turkey's motorways, known as the Otoyol network, serve as crucial connectors that can potentially influence the spatial dynamics of urban development. Spanning over 3,600 kilometers, the Otoyol network was first opened in 1973, traversing 29 out of 81 provinces in Turkey and collectively encompassing two-thirds of the nation's total population. Some sections of the motorways were constructed before 1973, later merging with additional motorways to form a comprehensive network. However, the majority of the motorways in Turkey were built after 1975.

Figure 2.1 presents the temporal development patterns of grid cells and motorways in Turkey. This map highlights the construction periods of urban grid cells and associated motorways, categorized into four time intervals: built before 1975, between 1975 and 1990, 1990 to 2005, and 2005-2020. The varying shades of blue indicate the motorway construction timing, with darker shades representing later construction periods. Urban grid cells are distinguished by color, with orange denoting those built before the nearest motorway and dark red indicating grid cells built after the nearest motorway. As can be seen in the figure, there is considerable variation in both the period of motorway construction and whether the motorway came before or after adjacent urban development. Moreover, motorways in Turkey are located in several urban centers, notably Istanbul, Ankara, and Izmir, but also several smaller cities.

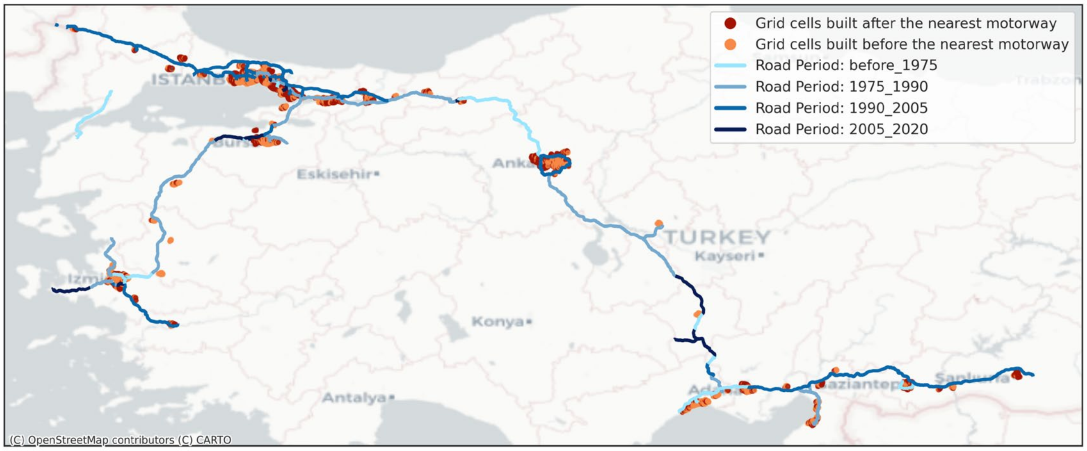


Figure 2.1. Grid cells and motorways development periods, Turkey

In Figure 2.2, we explore the correlation between urban form metrics—specifically street connectivity and density—and the proximity to the nearest motorway in grid cells developed before and after motorway construction. Grid cells developed before the motorway (blue lines) have higher density and more connected streets compared to those built after the motorway. As the distance from the motorway increases, the gap declines. This supports our hypothesis about how motorways affect urban development, although the difference is small. In addition, no apparent positive or negative trends indicate an absence of significant correlation and a clear pattern between the distance to the nearest motorway and density or street connectivity. If motorways affected adjacent urban development, we would expect to see a large gap between the red and blue lines in grid cells closest to the motorway, with the difference diminishing further away, but only a small effect is visible in the plots.

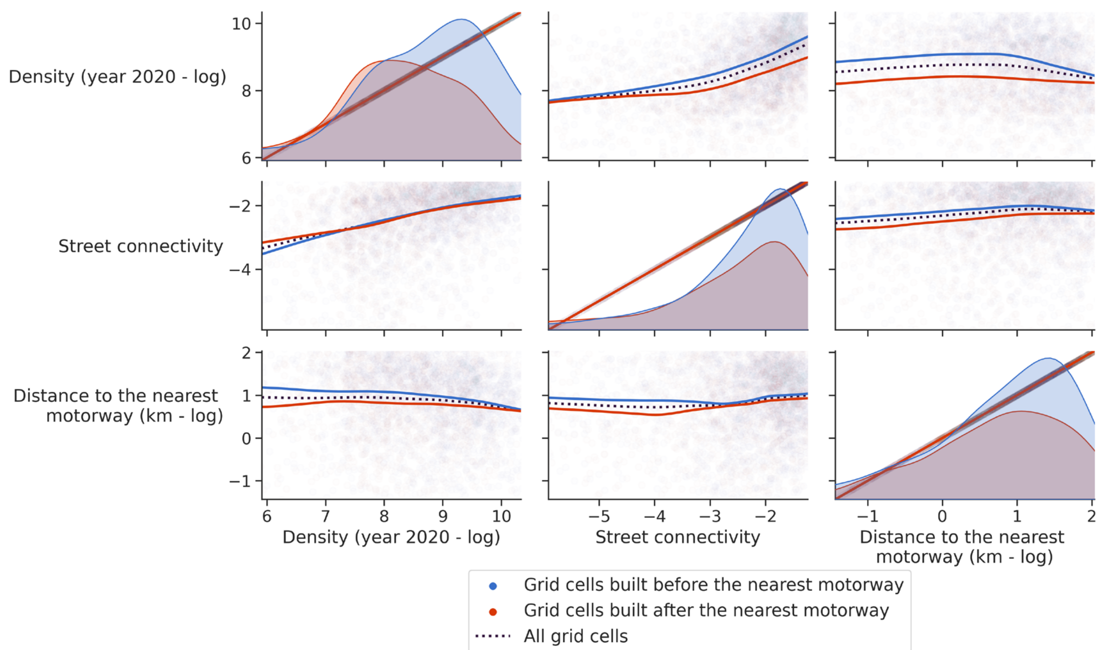


Figure 2.2. Correlation between urban form metrics and proximity to nearest motorway in pre- and post-motorway construction grid cells, Turkey

4.2. Ireland and Northern Ireland

Ireland, an island in northwestern Europe, shares a border with Northern Ireland, fostering a unique context for our study due to their proximity and historical ties. Examining their urban development patterns enables us to explore potential cross-border influences and shared trends.

In Ireland, the highest category of road is designated as a motorway, identifiable by the prefix M. While certain sections of the current motorways date back to the 1960s and 1970s, the inaugural segments of the Irish motorway system were opened in 1983.

Figure 2.3 provides insights into the temporal evolution of grid cells and motorways, offering a visual representation of their development periods in both Ireland and Northern Ireland. There is more potential impact of motorways in Northern Ireland compared to Ireland as the grid cells that were built after the motorways are more widespread near Belfast. The figure shows that there are fewer urban centers along the motorways in Ireland compared to Turkey.

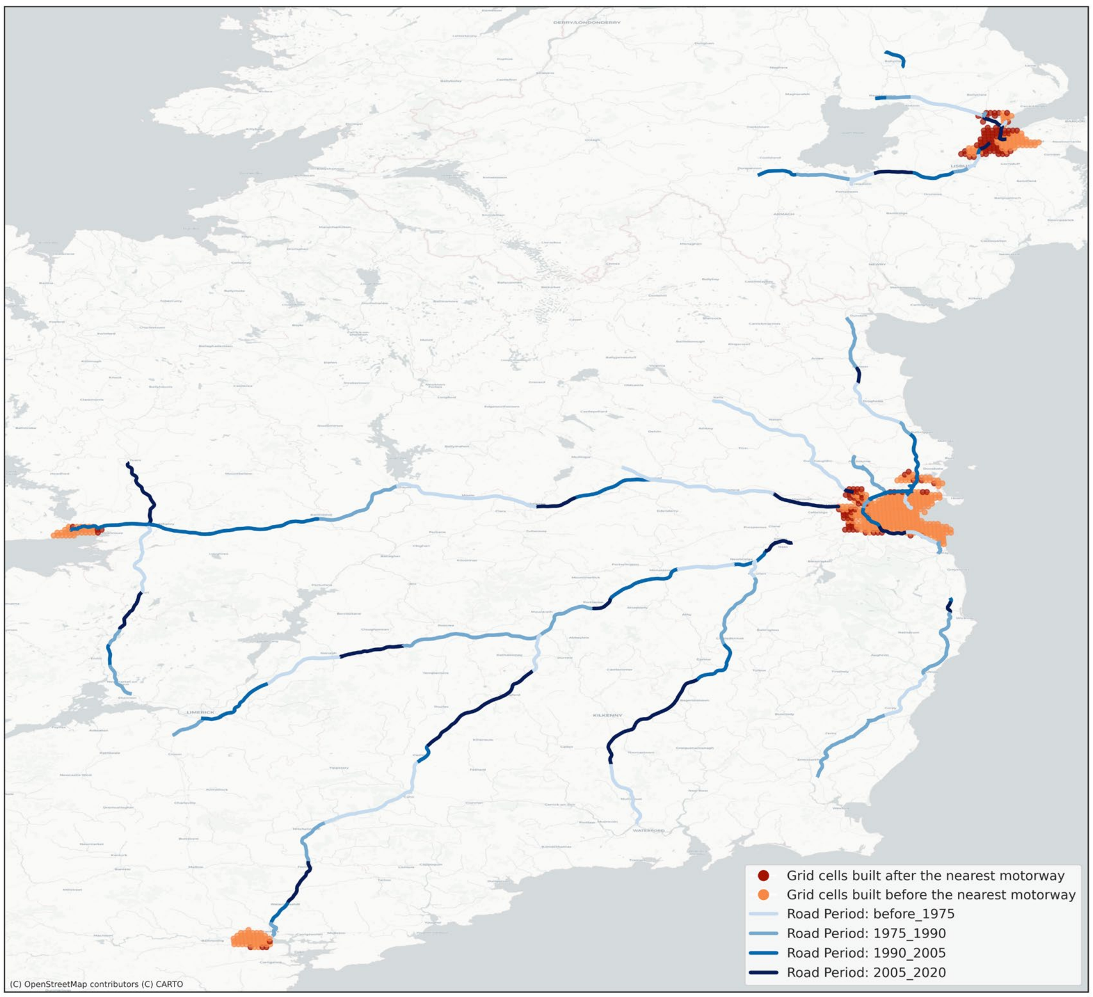


Figure 2.3. Grid cells and motorways development periods, Ireland and Northern Ireland

The pairplot in Figure 2.4 demonstrates the correlation between urban form metrics and the proximity to the nearest motorway in grid cells constructed before and after motorway development. Our findings show minimal noticeable differences in the density of the grid cells developed before and after the nearest motorway. However, the figure demonstrates that grid cells constructed after the motorway (depicted in red) have higher street connectivity, and the gap between the street connectivity of the grid cells before and after the motorway widens. These findings contradict our initial hypothesis, where we expected a decrease in street connectivity as the distance to the nearest motorway decreased for the grid cells built after the motorway.

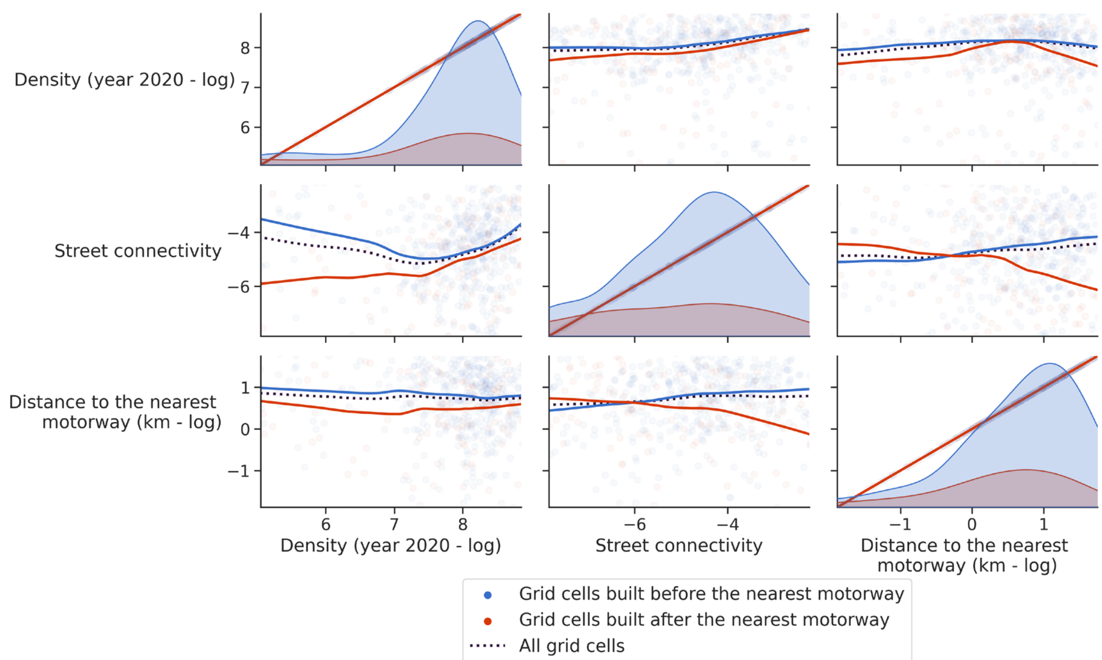


Figure 2.4. Correlation between urban form metrics and proximity to nearest motorway in pre- and post-motorway construction grid cells, Ireland and Northern Ireland

4.3. Thailand

In Thailand, situated in Southeast Asia, the major motorways are near Bangkok which makes this region different from our other case studies where motorways connect multiple cities. Coupled with the increase in the number of vehicles and the demand for limited-access motorways, the Thai Government issued a cabinet resolution in 1997 detailing the motorway construction master plan. Some upgraded sections of highways are being turned into motorways, while other motorways are being purpose-built. Therefore, the motorways are built more recently in this region compared to the other case studies and all have been built since 1975 (Figure 2.5).

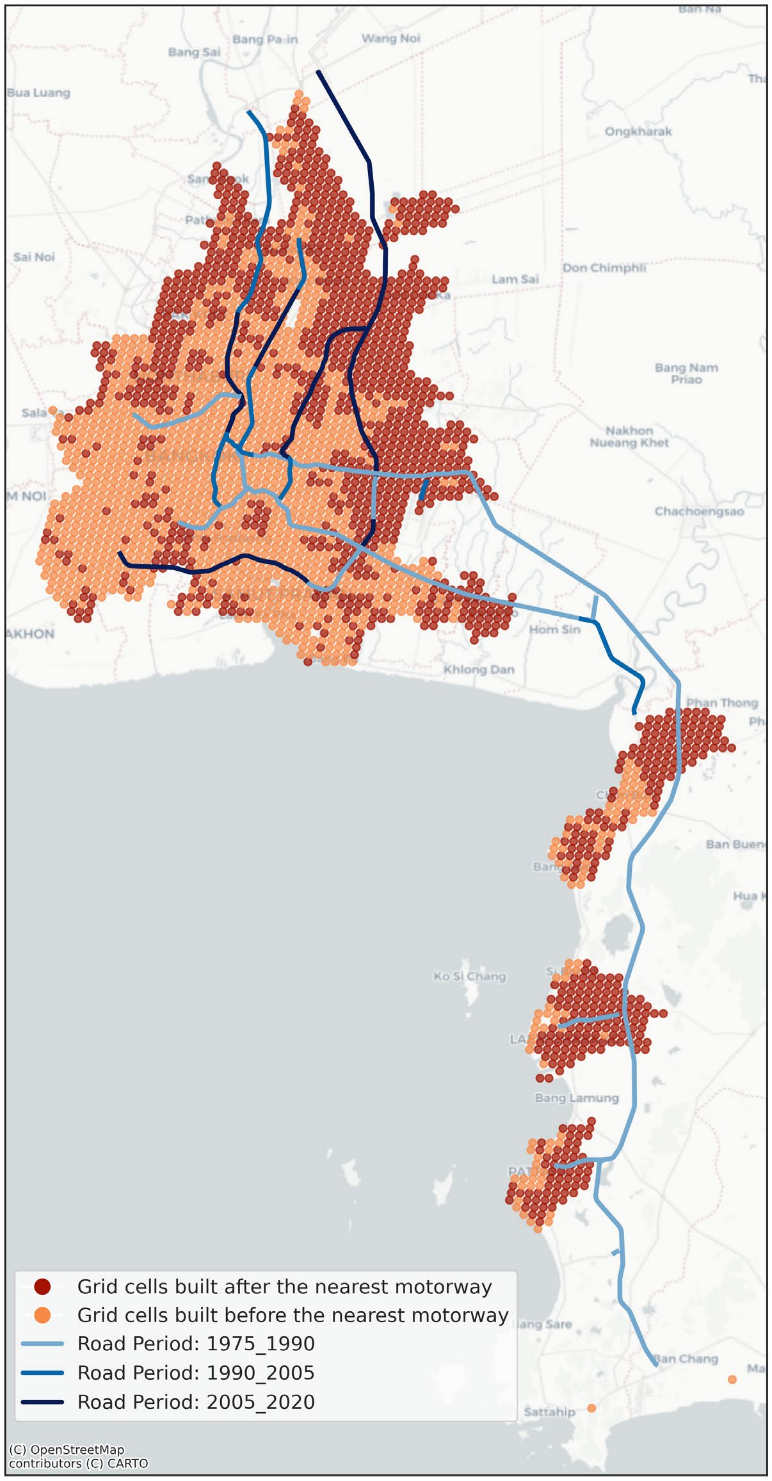


Figure 2.5. Grid cells and motorways development periods, Thailand

Our analysis reveals a pronounced gap between the red (after motorway) and blue (before motorway) lines in Figure 2.6. Close to a motorway, densities are substantially lower for grid cells developed after the motorway was built. This gap in density between grid cells built after and before the motorway narrows as we move away from the motorway, suggesting that the motorways reduce the density of urban development. A similar trend is observed regarding the gap in the street connectivity of two grid cell groups, although the initial gap is smaller than the gap observed in density. These findings suggest that the construction of new motorways tends to contribute to urban sprawl in Thailand.

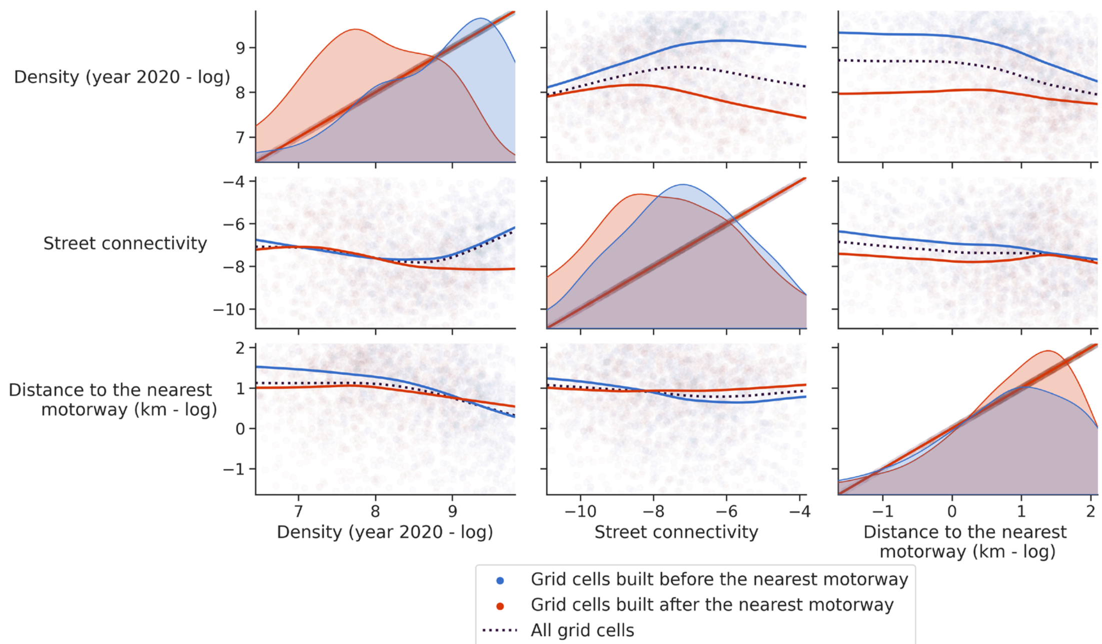


Figure 2.6.Correlation between urban form metrics and proximity to nearest motorway in pre- and post-motorway construction grid cells, Thailand

4.4. Taiwan

Taiwan, located in East Asia and characterized by mountainous terrain and vibrant urban centers, presents a distinctive context for understanding the relationship between motorways and urban patterns. The first motorway in Taiwan, and a predecessor to the national highways, was the MacArthur Thruway, built in 1964 between Keelung and Taipei. Construction on the first modern national highway, National Highway 1 began in 1971 and its first section was completed in 1974. Figure 2.7 demonstrates the grid cells and motorways development periods in Taiwan, showing that the majority of the motorways were built after 1975.

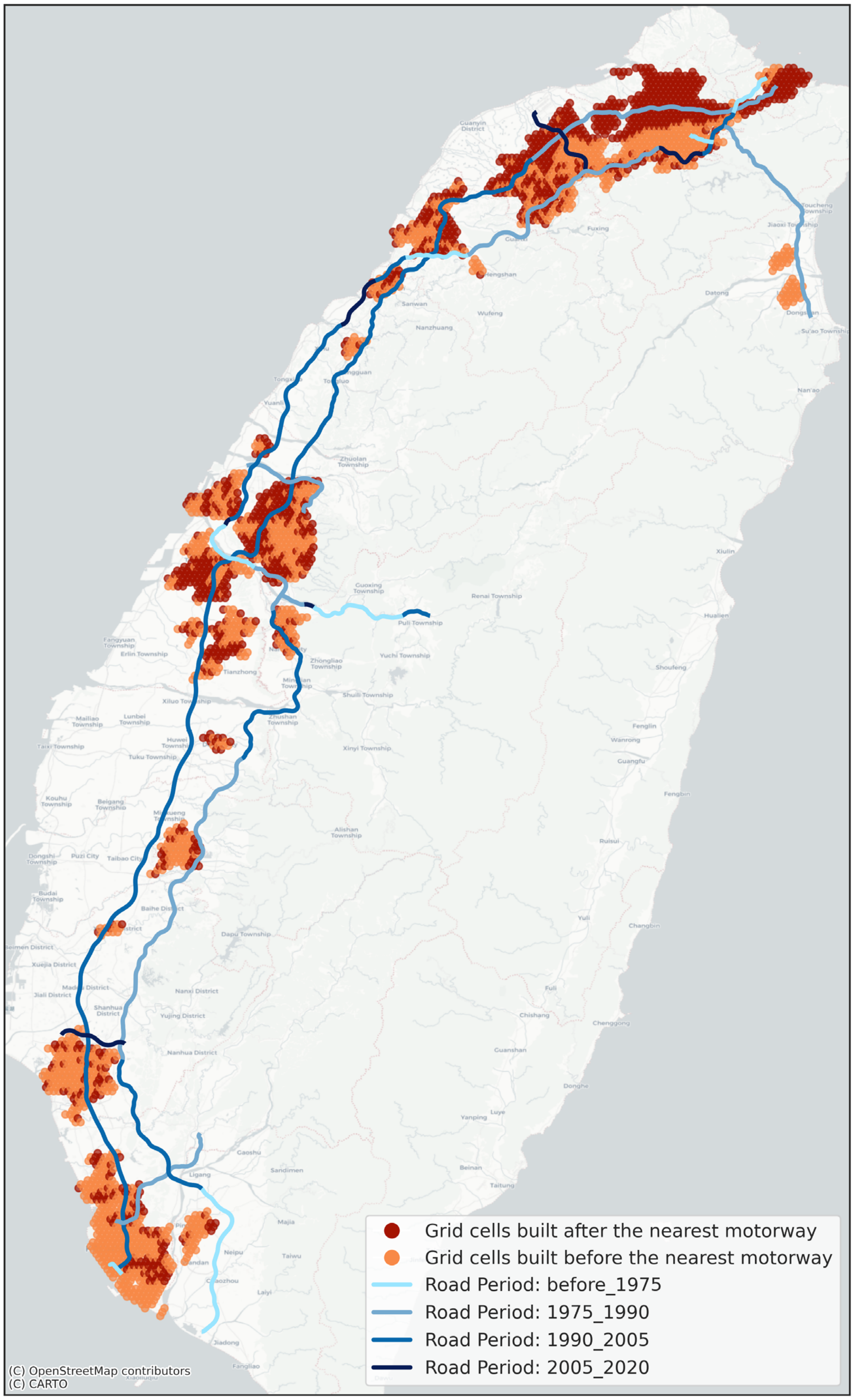


Figure 2.7. Grid cells and motorways development periods, Taiwan

Our examination presents patterns similar to Turkey, with a small gap close to the motorway, although in this case the gap does not narrow further away. Figure 2.8 illustrates the pairplots, confirming the absence of specific correlations between street connectivity, density, and the proximity to the nearest motorway in both pre- and post-motorway construction grid cells. The lack of distinct trends, as indicated by the absence of gaps or notable shifts in correlation lines, suggests a complex dynamic of factors shaping the development patterns around motorways in Taiwan.

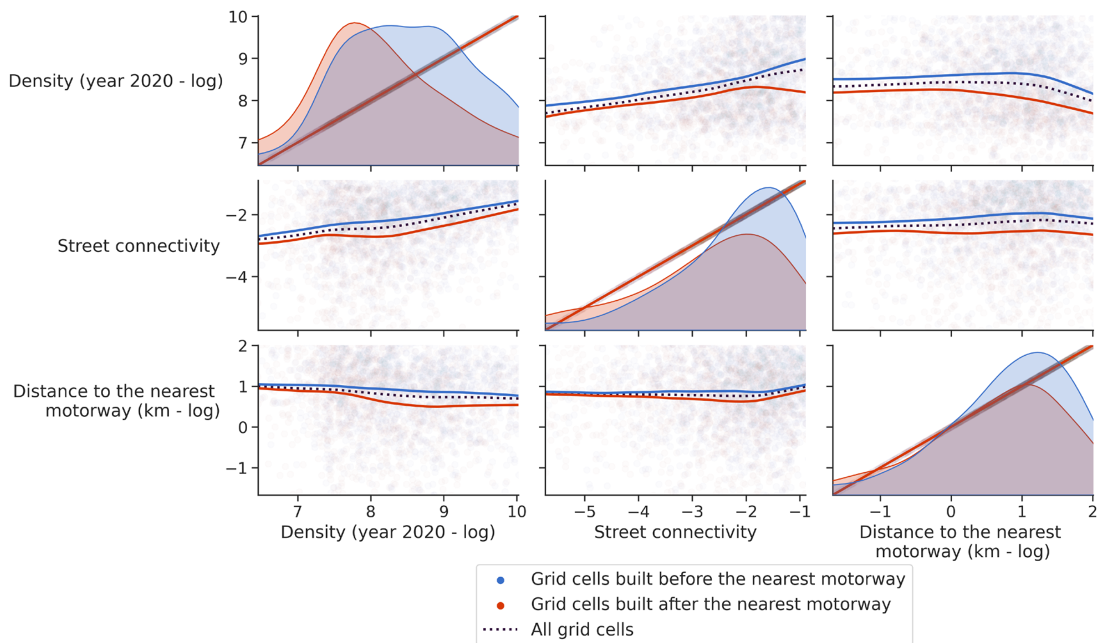


Figure 2.8. Correlation between urban form metrics and proximity to nearest motorway in pre- and post-motorway construction grid cells, Taiwan

4.5. Regression Analysis

The OLS regression analysis, summarized in Tables 2.1 and 2.2, explores the relationship between the built environment—specifically grid cell development after the construction of motorways—and the log of the distance to the nearest motorway.

The dependent variables are density and street connectivity. Our primary independent variables of interest are a dummy variable that indicates whether the grid cell was developed after the motorway and the interaction between this dummy variable and the log of distance to the nearest motorway. If the motorway reduces street connectivity or density, the coefficient on this interaction should be negative. We also control for the log of distance to the nearest city center, the log of distance to the nearest motorway, and the dummy variables for grid cell development over different periods.

Across the studied contexts, distinctive patterns emerge in the coefficients, offering insights into the association between the density of the grid cells built after the motorways and the distance to the nearest motorway. In the left-hand columns, all coefficients are negative, indicating that grid cells built after the motorway are less dense. However, this effect is only statistically significant in Thailand and Taiwan. In the right-hand columns, the coefficients in Thailand and Turkey are positive, indicating that (as our hypothesis suggests) the effect of the motorway declines with distance. However, in all case studies except Thailand, the effect is not statistically significant.

Table 2.1. Regression results on the impact of motorway proximity on urban density

	Dummy variable showing whether built after the nearest motorway			Interaction of grid cell built after the motorway and log of distance to the nearest motorway		
	Coefficients	Standard Error	P-value	Coefficients	Standard Error	P-value
Thailand	-0.7332	0.042	0.000	0.0987	0.029	0.001
Turkey	-0.5925	0.119	0.000	0.0902	0.070	0.200
Taiwan	-0.2339	0.195	0.230	-0.0236	0.089	0.790
Ireland and N Ireland	-0.2526	0.124	0.042	-0.0233	0.083	0.780

Notes: The dependent variable is density.
Other predictors: years of development, distance to the nearest large urban center, and log of distance to the nearest motorway

Table 2.2 reports coefficients for the same regression but with street connectivity as the dependent variable. As with density, the results for Thailand support our hypothesis – street connectivity is lower for grid cells built after the motorway, and the effect declines with distance. In Turkey, the signs of the coefficients are also in line with our hypothesis, but the effect is not statistically significant. In Taiwan, there is a statistically significant reduction in street connectivity for grid cells built after the motorway, but the effect does not decline with distance. In Ireland and Northern Ireland, the signs of the coefficients are the opposite to what our hypothesis would suggest.

Table 2.2. Regression results on the influence of motorway proximity on street connectivity

	Dummy variable showing whether built after the nearest motorway			Interaction of grid cell built after the motorway and log of distance to the nearest motorway		
	Coefficients	Standard Error	P-value	Coefficients	Standard Error	P-value
Thailand	-1.3792	0.089	0.000	0.3779	0.060	0.000
Turkey	-0.2088	0.113	0.064	0.0277	0.112	0.804
Taiwan	-0.4633	0.117	0.000	-0.0118	0.108	0.913
Ireland and N Ireland	0.0664	0.189	0.725	-0.5495	0.134	0.000

Notes: The dependent variable is street connectivity.
Other predictors: years of development, distance to the nearest large urban center, and log of distance to the nearest motorway

In Thailand, there is strong support that the construction of motorways has led to a decrease in both density and street connectivity in the areas built afterward, contributing to more sprawling areas. In Turkey, we find similar results, but the effects are smaller and generally not statistically significant. In Taiwan, our results are inconclusive, while in Ireland and Northern Island, they are the opposite of what we might expect. In summary, our regression analysis underscores the context-dependent nature of the impact of motorway proximity on urban sprawl across various regions.

5. Conclusion

As cities continue to face the challenges of accommodating growing populations and expanding economic activities, understanding the dynamics between motorways and urban form remains an essential element for sustainable urban development. Our findings contribute to the evolving body of knowledge that informs the future of urban development.

In summary, our analysis indicates that the influence of motorways on density and street connectivity is generally context-specific. While Turkey and Taiwan demonstrate non-significant associations, Thailand reveals a significant positive correlation for both sprawl metrics. Moreover, Ireland and Northern Ireland exhibit a significant negative correlation for street connectivity. This raises the possibility that motorways might affect urban development at the scale of the city as a whole, rather than having localized effects close to the motorway, suggesting a compelling area for future research. The counterintuitive discovery prompts an exploration into the broader effects of motorways on urban form, highlighting a global trend wherein motorways play a role in shaping the characteristics of cities on a holistic scale, extending beyond the immediate vicinity of the infrastructure.

This study examines specific contexts, aiming to explore the impact of the transportation infrastructure on urban sprawl on a global scale. The selection of case study regions—Turkey, Ireland, Thailand, and Taiwan—represents a deliberate effort to include regions that are often understudied in this area, contributing insights that extend beyond localized contexts.

The insights provided by our research pave the way for future investigations. Expanding the list of case studies and conducting a comparative analysis between intercity and intracity highways can be an intriguing avenue for further exploration. Moreover, examining the temporal aspects of motorway impact, considering longer-term trends, and incorporating a broader set of urban form metrics can enhance our understanding of the causal effect.

Acknowledgements

We extend our gratitude to Christopher Barrington-Leigh for his contributions in providing the street network connectivity dataset. We thank Carlos Dobkin and Chris Benner for their insightful comments on earlier drafts of this paper. This article relies on the Global Human Settlements Layer (GHSL) from the European Commission Joint Research Center and OpenStreetMap. We are also grateful to Kahaan Gandhi, Ciara Tang, Aneesh Raghavan, Laurie Jin, and Ruben Chung, the Science Internship Program (SIP) interns, for their efforts in collecting the motorway data of different regions.

CHAPTER 3

At the Nexus of Equity and Transportation Modeling: Assessing Accessibility through the Individual Experienced Utility-Based Synthesis (INEXUS) Metric

Abstract

We propose the Individual Experienced Utility-Based Synthesis (INEXUS) accessibility metric, which is developed to leverage an open-source agent-based regional transportation model. We include two specifications: the Potential INEXUS, which relates to an individual's potential set of mode alternatives and the Realized INEXUS, which reflects the optimal mode chosen by the agent. One advantage of using an agent-based approach is that it enables us to estimate individual agent-level behavior and travel needs. This addresses a commonly identified limitation of many existing accessibility metrics, which exhibit insensitivity to the heterogeneity of transportation preferences, opportunities, and constraints across subpopulations. While many system-level outcomes of interest may inform transportation planning, arguably an equally important consideration is that the system provides adequate and equitable access to goods and services for the broad spectrum of those traveling along its network. In many cases, average results do not reflect the experience of a majority – or even a significant – portion of the population. We apply our methods in a case study of alternative ridehail price scenarios to demonstrate the value of INEXUS distributions in evaluating differences in accessibility within and between population groups.

Keywords: Transportation accessibility, Equity, Utility-based metrics, Agent-based models

1. Introduction

When examining the effects of a new transportation policy, system-level outcomes summarizing congestion, capacity, aggregate or average vehicle miles traveled, person miles traveled, or mode splits can usefully inform planning. However, arguably an equally important consideration is that the system provides adequate and equitable access to jobs, goods, services, and education, for the broad spectrum of people traveling its network. Thus, we are concerned with more than average population-level impacts, as they may not reflect the experience of a significant portion of the population. Additionally, a policy, technology, or system design change may be intended to increase transportation system equity, e.g., by focusing on increasing accessibility of low-income groups to high-quality jobs, childcare, or grocery stores. In the transportation context, reporting only system-level impacts of new technologies such as autonomous vehicles (AVs) or ridehail service penetration masks the heterogeneity in experiences at the individual level, and precludes evaluating equity considerations.

Existing metrics are often location-based, driven by observational data, and describe location-based accessibility (Hou et al., 2019) or transportation affordability (Center for Neighborhood Technology, 2017). Moreover, many of these metrics or applications employ more normative definitions of accessibility, relying on assumptions about what should be considered a reasonable benchmark for travel distance, travel time, or cost (Páez et al., 2012). While these metrics offer valuable insights into the transportation system, they fall short of representing the full distribution of policy impacts across individuals, and often do not fully account for the behavioral responses that may more realistically underpin actual accessibility. Observational mode use data and surveys may provide information on variation across outcomes for different individuals, e.g., the chosen mode, travel times, or costs, for different population groups (Karlström & Franklin, 2009). However, these metrics do not assess an individual's happiness or "satisfaction" with their choices, including day-to-day level of satisfaction with their transportation experience and which set of transportation choices allow access to various destinations. A primary limitation of revealed measures of access or well-being is the lack of a baseline understanding against which to compare the outcomes. Though commonly used as accessibility measures, an individual's observed outcomes (e.g., mode chosen) only reveal part of the story. External factors constrain mode choice, including family wealth and car ownership, transportation system options, cost, congestion and travel time, hassle, etc. Research on voluntary versus involuntary car-free households indicates that these two groups make different residential location choices (Delbosc & Currie, 2012) and that, particularly in California, 79% of no-car households are involuntarily so (Brown, 2017), with car ownership strongly influenced by parking ratios and transportation accessibility (Millard-Ball et al., 2022).

We propose two individual-level metrics, collectively called the Individual Experienced Utility-Based Synthesis (INEXUS). These metrics capture the

Potential—the value of the complete choice set of available transportation modes for each person for every trip they take during a typical day—as well as the Realized—the achieved “happiness” from the selected mode of a given trip. The INEXUS specifications provide an agent-trip-level utility-based set of accessibility metrics that can complement other types of location-based accessibility metrics to generate a holistic understanding of variation in potential and realized experiences, and how they are valued by the individual travelers, across population groups and transportation scenarios.

To estimate these proposed metrics, we use an integrated set of agent-based models called BEAM CORE, where CORE stands for Comprehensive Regional Evaluator. This integrated modeling workflow consists of an agent-based regional transportation model, Behavior, Energy, Autonomy, and Mobility (BEAM), an open-source model developed by the Lawrence Berkeley National Laboratory (LBNL). BEAM is integrated with: ActivitySim, an open-source activity-based travel demand model; UrbanSim, an agent-based land-use model; and ATLAS, an agent-based household fleet and vehicle choice model. Unlike other techniques, agent-based models facilitate examination of both system-level results as well as the outcomes for all individual agents that utilize the system.

Advantages of this approach include 1) the ability to model the impact of a multitude of hypothetical and future transportation scenarios; 2) estimation of heterogeneous outcomes for each person or group and resulting (in)equity. Leveraging this approach to generate the INEXUS set of metrics, each converted into dollar units to enhance interpretability, enable transportation planners to understand not only how potential policy scenarios could change the overall system outcomes, but also individual or subpopulation-level behavior, wellbeing, and accessibility. This application of utility-based accessibility metrics implemented in a fully disaggregated agent-based regional simulation modeling framework enables a positivistic analysis of accessibility, thereby addressing a dearth of research taking this approach to accessibility (Páez et al., 2012). In addition, unlike most of the research applying positivistic accessibility analyses, which tend to focus on empirical or survey data as summarized in Páez et al. (2012), the application of our metric in a simulation framework enables a more flexible exploration of a range of potential what-if scenarios, policies, or system design changes.

The contributions of our development of the INEXUS metric include: 1) creating a framework to gain more value from the detailed, disaggregated, agent-based model outcomes and thereby contribute to accessibility research that is less well-represented, namely positivistic accessibility, accounting for a more complete representation of empirically-derived travel preferences and endogenous behavior change, 2) creating a set of building block metrics at the agent-trip level enabling reporting of distributions of individual outcomes allowing for identification of differential impacts on subgroups of interest, which can be examined ex-post by agent and trip attributes, 3) recognition of the potential for substantial differences between planned transportation choice expected utility and realized utility under system constraints, 4) the ability to compute differences in monetized values of consumer surplus and welfare across scenarios, 5) a consistent framework to compare accessibility, consumer surplus, and welfare across a broad range of policy, technology, and system design scenarios using the BEAM CORE integrated modeling system (BEAM and ActivitySim in particular); and 6) providing a multi-dimensional utility-based set of metrics that synthesize numerous trip characteristics, all of which contribute to the outcome.

Our paper proceeds as follows: section 2 provides background information defining types of metrics used to characterize transportation accessibility and orients our proposed INEXUS metric within the literature discussing similar metrics. In section 3, we describe our methods, including the transportation system and choice models used to define the agents, dataset format, construction of the INEXUS specifications, and extension of INEXUS to examine shifts in consumer and social welfare. In section 4, we present the rationale for several case studies of accessibility and welfare change under potential policy interventions and transportation utilization shifts; we discuss findings from these case studies, noting insights provided by the INEXUS metric. Section 5 concludes with discussion of potential applications of INEXUS in future research and policy analysis.

2. Background on accessibility metrics

This section presents a brief survey of existing metrics of accessibility that motivated the development of the INEXUS. Figure 3.1 provides a conceptual overview of the types of accessibility metrics established in the literature. Broadly, access can be measured from a geographic location (isochrone- or gravity-based) or for an individual traveler (utility- or activity-based). Each type provides insight into transportation system performance, albeit at different levels of analysis.

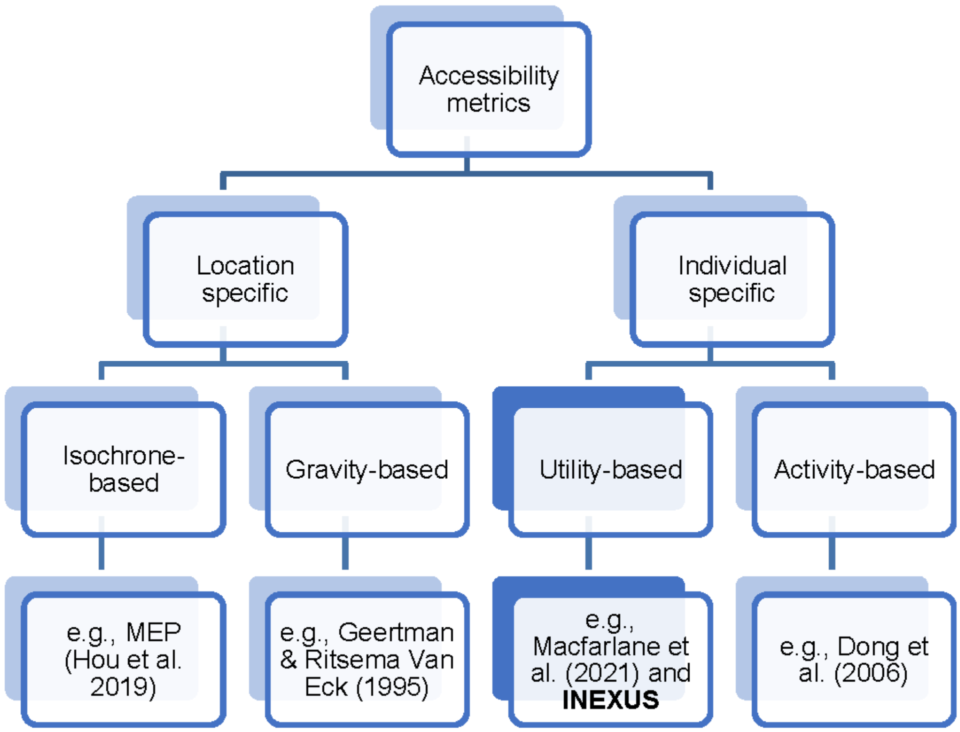


Figure 3.1. Accessibility metric overview

2.1. Location-based metrics

Location-based metrics estimate potential achievable outcomes, such as access to opportunities, for specific geographic locations based on geographic and transportation system attributes. Common types of location-based metrics include gravity-based measures, which calculate a location's accessibility by considering proximal opportunities weighted by a measure of effort required for access (Geertman & Ritsema Van Eck, 1995), isochrone measures, which estimate cumulative opportunities available within a given travel time, distance, or generalized cost

(Vickerman, 1974; Wachs & Kumagai, 1973), and composite indicators of energy and travel times (Hou et al., 2019). Location-based indicators are useful for identifying the maximum accessibility for travelers with unconstrained mobility. They can be either normative or positivistic, though are more likely to have normative definitions in which key thresholds of travel distance, time, or cost are assumed to be relevant benchmarks of access (Páez et al., 2012). They are typically measured with publicly available data and enable comparisons of land use, transit level of service, and potential access across different locations within a region. The more broadly applied location-based accessibility measures tend to be static and atemporal in their nature. However, recent studies such as Wang et al. (2018) and Järv et al. (2018) have introduced location-based space-time accessibility measures to capture temporal variation. Although location-based measures are widely used, they generally overlook the nuances of behavioral differences (Bills & Carrel, 2021). These measures often assume uniform accessibility for all travelers originating from the same starting point. Bills and Carrel (2021) propose an alternative perspective on measuring transit accessibility, focusing on the behavioral adaptations of individual travelers to travel time unreliability. Nevertheless, they acknowledge that more research is needed to refine the approach and explore its equity implications.

2.2. Individual-based metrics

Individual-based metrics provide an opportunity to understand not only how access varies across demographics, but also how transportation needs are distributed within the same or across different locations. These metrics are largely understudied because they require detailed, individual-level data on transportation needs, and tend to reflect accessibility that is positivistic, accounting for traveler preferences and behavioral responses—an area that has been less well-represented in the literature historically (Páez et al., 2012). Individual-based metrics are typically utility-based (Ben-Akiva & Lerman, 1979; Demitiry et al., 2022; Niemeier, 1997), but may be estimated using stated or revealed preference surveys (Farber et al., 2016). Individual-based metrics can be activity-based or path-dependent.

Utility-based measures are based on random utility theory (Domencich & McFadden, 1975), in which a portion of the traveler's utility (e.g., modal attributes) is observable and a portion is unobserved by the researcher. Logsums of the denominator of utility models are frequently used as a proxy for accessibility (Ben-Akiva & Lerman, 1979; Niemeier, 1997). This formulation generally implies that larger choice sets with higher utility alternatives result in higher levels of accessibility. The multinomial logit (MNL) specification of accessibility benefits from transferability to consumer welfare, as it can be viewed as the demand curve for a particular mode-destination alternative. Consumer surplus differences can be calculated for two scenarios based on the difference in expected maximum utility between a base condition and policy change scenario (H. C. W. . Williams, 1976). Utility-based measures differ in complexity, ranging from selecting among available destinations for a single trip to accounting for an individual's entire daily activity pattern choice set, inclusive of trip purposes, departure times, destinations, and modes.

Activity-based accessibility measures, introduced by Ben-Akiva and Bowman (1998), are derived from random utility theory and generated from the traveler's daily activity schedule. In contrast to other measures of accessibility, which traditionally capture only a single trip purpose at a time, Dong et al. (2006) introduced a utility and activity-based accessibility measure that takes into account the expected maximum utility from all activity patterns for the day, including destinations, modes, departure times, and subtours. Such rich detail provides a more holistic understanding of the individual's actual access to opportunities, which requires granular data on both the individual's activities and transportation system attributes.

Path-dependent measures of accessibility augment individual-based metrics by utilizing both the trip-tour origin as well as points along the route to define individual-level access (Kwan, 1998; Páez et al., 2012). They can also provide individual-path-specific accessibility measures, which require detailed data collection reflecting current and desired travel behavior.

We turn now to the INEXUS metrics, explaining their value in the context of existing transportation accessibility measures. It is important to note that these metrics only measure the utility of the trip, not the intrinsic benefits of the activity (Mokhtarian et al., 2015), and they are proposed to complement more common location-based metrics. Bills and Walker (2017) discuss the value of an individual utility metric ("distributional comparison measures… incorporating equity standards") and provide a demonstrative calculation using 2000 survey data points. They advocate for calculating equity indicators at the individual actor level, noting that mean or median indicators obscure key decision-making information. We agree that aggregate accessibility and equity measures mask important individual-level outcomes, potentially reducing their usefulness in identifying and addressing real-world disparities. We build off of (Bills & Walker, 2017) and propose two individual utility metric formulations, addressing potential and realized outcomes.

As mentioned above, the logsum measure is a well-known metric in the accessibility literature. It has been used in a variety of applications, including estimating travel demand (Harb et al., 2022), evaluating transportation projects (Guzman et al., 2023), and developing transportation policy (Gulhan et al., 2013). Our metrics and approach build from the utility-based accessibility literature, but advance the state of research on measuring transport accessibility in several ways:

1) By using data output from BEAM and ActivitySim, we enhance the analytical capabilities of these state-of-the-art transportation simulation and planning tools.

2) Because these models utilize recent data (2018), our process provides an accurate baseline for near-term calculations.

3) Scenarios can allow for multi-year analyses in which land use and travel demand can change over time. This flexibility enables INEXUS to be used to analyze cases where a range of behavioral responses—from residence and work location choices, household vehicle holdings, and trip generation, in addition to mode choice—can all be responsive to a given system intervention. It can also be applied in cases where land use, vehicle ownership, and travel plans are fixed, and only mode choice

behavior is subject to change (as we demonstrate in the case study presented in this paper), or anything in between.

4) When evaluating potential impacts of untried policies on a multi-year scenario with an evolving population, agent-based modeling benefits from the ability to consider large deviations from baseline on the fly, rather than relying on an anticipatory survey. Individual preferences in agent-based models can be updated to incorporate insights from future surveys.

5) The two INEXUS specifications allow for comparison between the distributions of idealized outcomes (Potential) and outcomes under individual, household, and system constraints (Realized).

6) INEXUS metrics are intentionally calculated at the individual level. When INEXUS metrics are calculated for a dataset, the result is a distribution of individual outcomes, which may be later aggregated by attributes of interest, rather than a population-level value. Treatment at the individual level is key to "revealing the winners and losers that result from transportation improvements, in comparison with average measures" (Bills & Walker, 2017). INEXUS metrics are well positioned to evaluate the potential for policy impacts to disproportionately affect historically disadvantaged communities.

While the concept used for the INEXUS draws from existing literature and we are not introducing a completely new methodology, we have chosen to refer to it with the name "INEXUS" due to the novel application case we present, with this metric incorporated into the BEAM CORE integrated agent-based modeling workflow. This choice aids in communication around this concept with the many stakeholders leveraging the BEAM CORE capabilities.

3. Methods

In this section, we outline the methods used to develop INEXUS distributions. We explain our use of BEAM and ActivitySim models, both of which interact with and draw from several other models, to derive the data necessary to estimate distributions of individual accessibility metrics (Figure 3.2). We define the Potential and the Realized INEXUS specifications, and explain how to derive welfare and equity computations and scenario comparisons using INEXUS.

Figure 3.2 illustrates how different models are integrated and leveraged to calculate the INEXUS metrics. These models interact to result in a transportation equilibrium defining mode choices and resulting system impacts of every individual and vehicle, at every time, within a typical weekday travel day. This equilibrium results from modeling the interaction, for every person and every trip, between residence location choice, household vehicle fleet composition choice, activity selection (including timing and location), and mode choice for accessing those activities (including both mandatory and non-mandatory trip purposes). Travel plans of individuals, including destination and mode selections, collectively exert pressure on both road and public transit networks, resulting in congestion and crowding. Key modules in the integrated suite then iterate until the system reaches equilibrium.

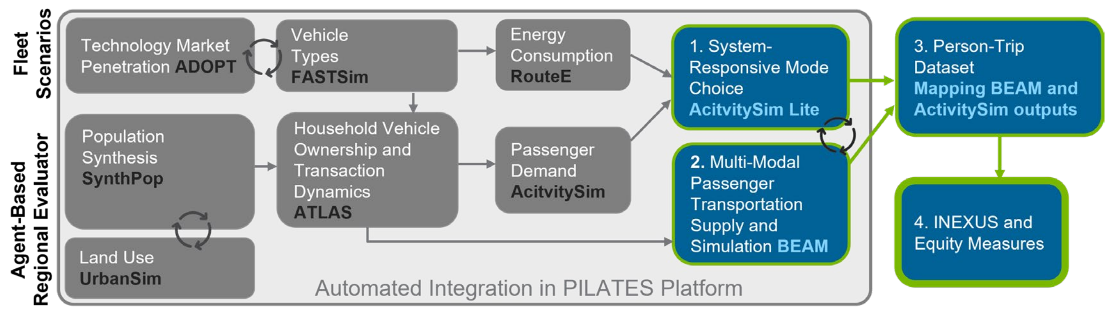


Figure 3.2. Developing INEXUS distributions from BEAM CORE outputs

As discussed in the previous section, BEAM CORE can be used to conduct multi-year scenario analyses in which all of these modules iterate and update each year, with behavioral responses leading to changes in residential and work location choice, vehicle ownership, travel demand and trip plans, and mode choice in response to a given system intervention. In such a case, all of the modules depicted in Figure 3.2 are run each simulation year. Alternatively, a short-run analysis can be conducted where only a subset of these models iterate as different scenarios are examined. For instance, in the case study presented in Section 4, in order to isolate only behavioral shifts in mode choice in response to ridehail price variations explored in the examined sensitivities, only the blue boxes in Figure 3.2 are re-run with each scenario, whereas the modules indicated as gray boxes are not re-executed. Outputs from the baseline run of the full set of models are used as static inputs to the sequence represented by the blue boxes for each scenario. The flexibility to apply the INEXUS in a wide variety of long or short-run analyses, depending on the needs of a given policy-maker, planner, or research question, is one of the strengths of this accessibility application.

3.1. Dataset: deriving an agent-level transportation choice

The INEXUS input dataset is composed primarily of outputs from BEAM and ActivitySim. The models are run for 10% of the 2018 population of the San Francisco Bay Area, defined as the nine California counties of Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma. The modeled output describes one typical weekday, and includes, by agent, every trip that the agent would take on a representative day. Our baseline scenario dataset consists of 2.5 million trips and 639,000 agents, with an average of about 4 trips per day per person. The modeled outputs include many variables, but we focus on those in Table 3.1.

Table 3.1. Key variables by process stage

Process:	Stage:	Key Variables	Variable Description
1. ActivitySim	Mediating & Moderating Variables	Income	Ranges from loss of 14,600 dollars to earning 1,397,000 dollars. mean = 121,427 standard deviation (sd) = 109,008
		Race	black, white, Asian, etc. (refer to PUMS data dictionary)
		Gender	male, female
		Age	0 to 94 years-old mean = 40 sd = 21
		Car ownership	none, one, two or more
		Household composition	household size, household income, number of workers, etc.
		Purpose of trip	work, shopping, school, etc.
2. BEAM	Trip Level Data	Departure and arrival time	Trips happen over one day starting at 5:00am and ending at 11:59pm.
		Fuel type	gasoline, electricity, diesel, etc.
		Vehicle type	conventional vehicles, hybrid electric vehicles, etc.
		Number of passengers	0 to 121
		Distance	mean = 11.1 km
		Mode choice (realized)	car, ride-hail, walk-transit, bike, etc.
3. Mapping BEAM & ActivitySim	Person-Trip Level Data	Total trip duration	mean = 16.3 minutes
		Wait time	For transit: mean = 15.7 minutes sd = 9.3 minutes For ridehail: mean = 7 minutes sd = 3.2 minutes
		In vehicle time	mean = 15.1 minutes
4. Distributions	Accessibility Metrics	Potential INEXUS	
		Realized INEXUS	
		Consumer surplus	

3.1.1. Activity-Based Model: ActivitySim

ActivitySim is an open-source activity-based travel demand model used by many Metropolitan Planning Organizations (MPOs) in the United States (Macfarlane &

Lant, 2021; Waddell et al., 2018). In activity-based models, each member of a population is simulated, with individuals completely disaggregated. Consequently, the model provides outputs relevant for evaluating the impact of policy changes on populations of interest, including historically marginalized groups (e.g., low-income communities) (Freedman & Hensle, 2021).

ActivitySim in its original form uses PopulationSim, an open-source population synthesizer, to create the synthetic population to forecast travel patterns (Freedman & Hensle, 2021). In the case of our implementation of ActivitySim, because it has been integrated into BEAM CORE with UrbanSim (Waddell, 2008), we use SynthPop (Ye et al., 2009) for population synthesis. SynthPop enables the efficient matching of both household-level and person-level characteristics when generating synthetic populations of interest.

ActivitySim outputs numerous moderating and mediating variables that describe each modeled agent, including demographics, travel patterns, and household characteristics. We use income to define groups of interest in the case study to follow, but any agent attribute output from ActivitySim or other upstream models in BEAM CORE could potentially be used to define a subgroup. Figure 3.3 depicts relationships between ActivitySim processes and inputs.

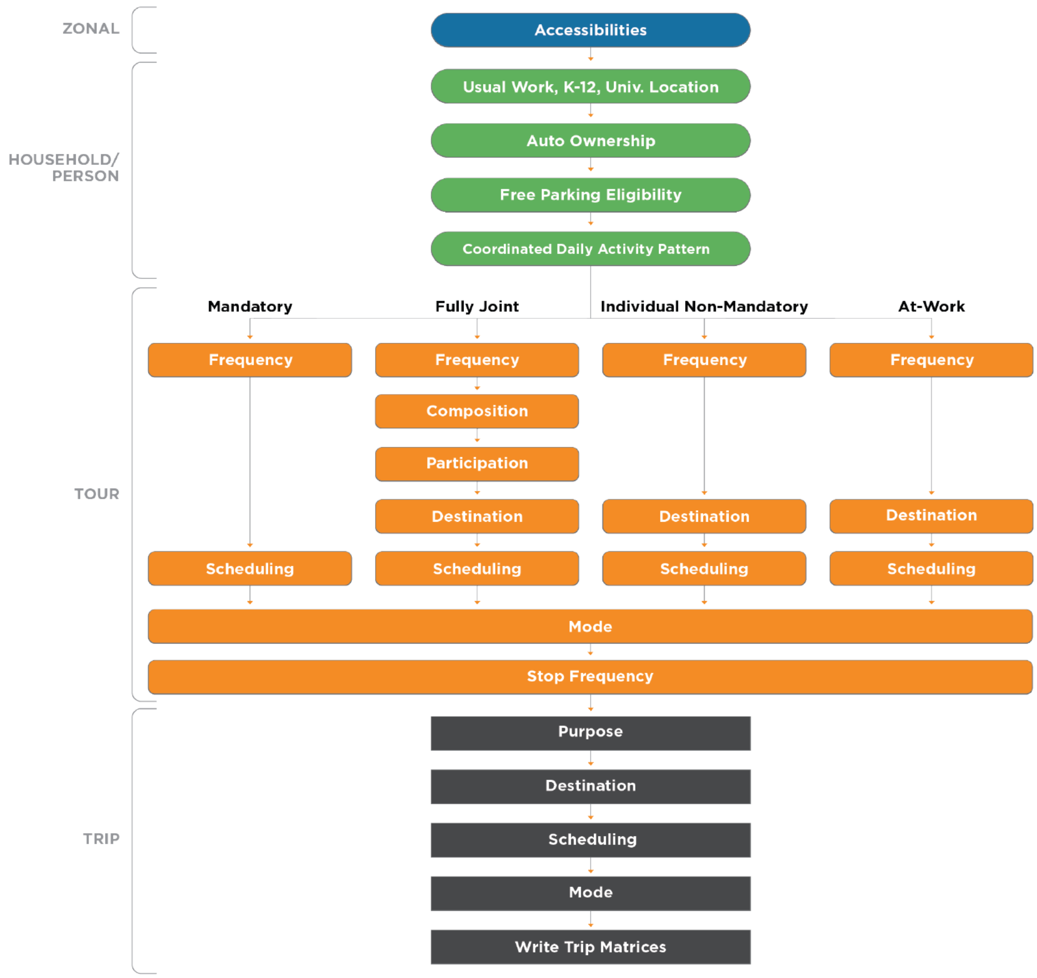


Figure 3.3. Components of ActivitySim modeling

3.1.2. Transportation System Model: BEAM

We also leverage BEAM (Laarabi et al., 2023), an open-source agent-based model ideally suited to produce results capturing detailed person-specific experiences, preferences, and constraints (Hsueh et al., 2021; Szinai et al., 2020). BEAM enables disaggregation of system-level impacts to agents, specific groups, and specific geographies. BEAM improves upon other open-source options, including its foundation model MATSim (Horni et al., 2016), with increased computational performance and by enabling integration of multiple new transportation innovations and paradigms (e.g., electric vehicles, ride-hailing) into a single scenario analysis.

3.1.3. Extracting output, creating a person-trip based dataset, and analysis from integration of BEAM with ActivitySim

Output from the equilibrium of all the combined agent-based models are extracted and processed to create a dataset in which each observation is one person-trip for the entire sample population and daily activities. We used "ActivitySim Lite", consisting of the ActivitySim mode choice model tightly integrated into the BEAM simulation such that BEAM and ActivitySim mode choice equilibrate through iteration. Using the BEAM output, we derived a new person-trip dataset, mapping trips onto people, grouping all trips for one person, and then mapping this to the ActivitySim data output to produce a rich dataset for each person-trip.

3.2. Proposed Metrics: Definitions of INEXUS

We assessed the ways to adapt best practices in estimating individual-level accessibility, as opposed to a location-based metric, and identified the utility-based specification as the most feasible for capabilities with BEAM CORE. To be suitable for policy decision-making and aid in interpretation, we defined INEXUS in dollar units. This allows policy-makers to determine, for example, if a transportation project's projected accessibility improvement benefits outweigh its costs. Following Train (2003), we use a conversion parameter α to translate utils into dollars. The parameter α_i is the marginal utility gained by consumer i from a one-dollar income increase (or cost decrease). Each person's utility is multiplied by $1/\alpha_i$ to convert it to dollar units. Below we introduce two metrics that together constitute the INEXUS suite of metrics: the Potential INEXUS and the Realized INEXUS. The Potential INEXUS is most closely related to the majority of other utility-based accessibility metrics derived in a similar way. It is based on the mode choice logsum, normalized by the parameter α, and therefore captures the monetized expected utility across the full set of modal alternatives. In contrast, the Realized INEXUS consists of the monetized utility only from the mode actually chosen by the agent. Many valuable analyses can consist of a comparison between outcomes across these two metrics. Examples of this are provided in Section 4.4 in the case study presented. In sum, the distinction between the Realized and the Potential INEXUS allows us to disentangle the mechanisms behind accessibility improvements, providing a more comprehensive view of the impact of a given intervention.

3.2.1. Potential INEXUS

The Potential INEXUS represents the monetized expected utility of an agent having all the mode choices available at the trip's beginning, when planning (i.e., based on the mode choice logsum) (Equation 3.1). Potential INEXUS is not limited to an agent's preferred or actually used mode of travel. The Potential INEXUS for agent i for trip n with transportation mode choice set $\mathrm{TC_{in}}$ captures the full modal option utility available to the agent for the given trip, specified as:

$$\text{Potential INEXUS}_{\mathrm{in}} = (1/\alpha_{\mathrm{i}})[\ln\left(\sum_{\forall k \in \mathrm{TC_{in}}} \exp\left(\mathrm{V_{ink}}\right)\right) + \mathrm{C}] \tag{3.1}$$

where TC_{in} is the transportation mode choice set available to agent i for trip n and V_{ink} is the deterministic portion of the utility of mode k for agent i for the set of modes available for trip n, $k \in TC_{in}$, and α_i is the marginal utility of an additional dollar. C is an unknown constant that represents the fact that the absolute level of utility cannot be measured. The random part of the utility is modeled through an error term ε_{ink} which is assumed to be independently and identically extreme value distributed (i.i.d.) and which, in expectation, reduces to the form above. The unit of measurement for the Potential INEXUS is dollars.

3.2.2. Realized INEXUS

The Realized INEXUS indicates the monetized utility of the trip taken, given the mode an agent actually used. Agents might take their planned mode, or may have to switch because, for example, a bus is full or no ridehail is available (Equation 3.2). The Realized INEXUS measures the utility experienced by agent i during trip n with mode choice set TC_{in} for the mode used k^*, R_{ni}, specified as:

$$\text{Realized INEXUS}_{in} = (1/\alpha_i)[\ln\left(\sum_{\forall k=k^*} \exp(V_{ink^*})\right) + C] = (1/\alpha_i)[(V_{ink^*}) + C] \qquad (3.2)$$

where V_{ink^*} is the observable portion of the utility function for agent i for trip n in which mode $k^*$is selected, and α_i is an additional dollar's marginal utility. Unknown constant C represents the immeasurability of the absolute level of utility, with the random part of the utility assumed to be i.i.d. extreme value distributed. This formulation follows the rationale explained above. The Realized INEXUS is also expressed in dollars.

3.2.3. Consumer Surplus and other welfare measures using INEXUS

A number of approaches have been used to measure social welfare, e.g., the degree to which people are better or worse off in a transportation scenario relative to the baseline. INEXUS lends itself well to direct comparisons of scenario outcomes and explicitly calculates the impact of different transportation scenarios on society, using various metrics of welfare. One common metric of welfare is Consumer Surplus (CS), which is defined as the amount of money a person would be willing to pay to attain a certain state of the world; it translates consumer utility into a monetary amount. Because the INEXUS is already monetized to be in "dollar" units, INEXUS is equivalent to CS. That is, the INEXUS represents each agent's individual CS for each trip they take.

Usually, CS is discussed as an average or overall welfare metric (i.e., INEXUS sum) for a population or a subpopulation, and policy decisions often focus on the change in CS — whether a particular policy makes individuals better or worse off and to what extent. To estimate how a population or subpopulation is affected by a transportation scenario relative to the baseline scenario, we subtract the total baseline CS from the total CS in the scenario of interest (Equation 3.3). The change in CS resulting from scenario V^1 changing to scenario V^2 , denoted as $\Delta \mathrm{P}_{\mathrm{in}}^{12}$, is specified as:

$$\Delta \mathrm{CS} \;=\; \Delta \mathrm{P}_{\mathrm{in}}^{12} = (1/\alpha_{\mathrm{i}}) \left[\ln \left(\sum_{\forall \mathrm{k} \in \mathrm{TC}_{\mathrm{in}}^{2}} \exp\,(\mathrm{V}_{\mathrm{ink}}^{2}) \right) - \ln \left(\sum_{\forall \mathrm{k} \in \mathrm{TC}_{\mathrm{in}}^{1}} \exp\,(\mathrm{V}_{\mathrm{ink}}^{1}) \right) \right] \quad (3.3)$$

When making policy decisions, we care about both the aggregate change in CS and equity. Equity assessment involves identifying which groups experience a net welfare increase and which encounter a net welfare decrease when moving from the baseline to the new scenario. A scenario in which high-income individuals benefit while low-income people are slightly worse off may result in an overall increase in welfare (i.e., aggregate CS increase), but this scenario may be undesirable from an equity perspective. This highlights the importance of assessing the change in CS for different subpopulations of interest. INEXUS is a metric that can be used to evaluate equity using a variety of different definitions, including equality, proportionality, and Rawlsian justice (Bills, 2022). The specific definition of equity used depends on the priorities specific to each application of the INEXUS in different policy or research settings.

4. INEXUS Case Study Application

To serve as a proof-of-concept of the value of INEXUS metric specifications in policy scenario simulation, we present a case study, and resulting INEXUS distributions for several subgroups of interest. We discuss insights regarding equity revealed through the power of the INEXUS metric.

4.1. Case study scenarios: ridehail price

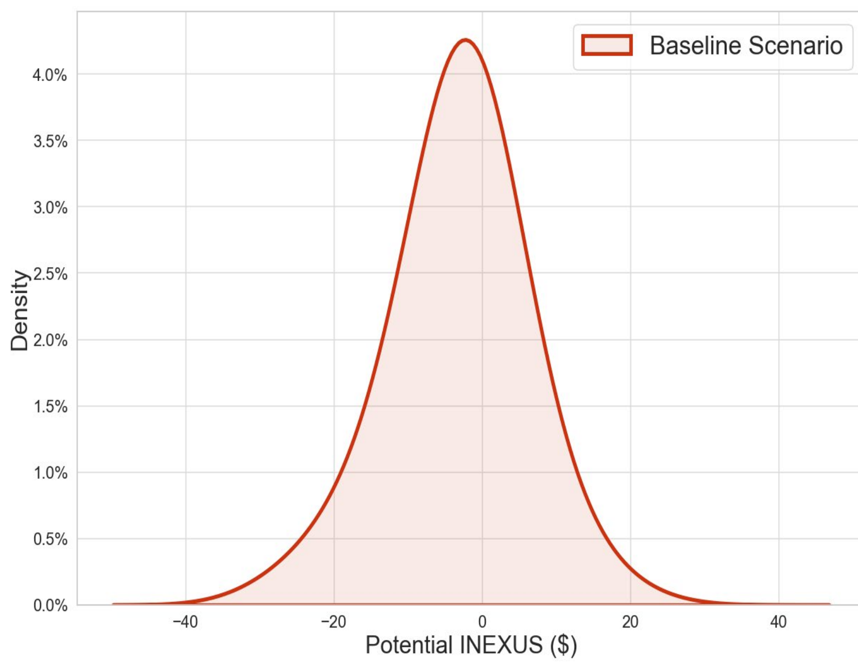
We evaluated the comparative impacts of hypothetical changes to ridehail price compared to baseline by using a sensitivity analysis in BEAM CORE in which the price of ridehail varied from 800% to 0% of baseline. Ridehail pricing makes for a compelling case study because of its flexibility and on-demand nature. We made a deliberate choice to use a single lever (ridehail price) and introduce an exaggerated range of variation for this case study to showcase the versatility and applicability of the INEXUS metric in this experimental setting. Exploration of the results from these scenarios may be relevant for transportation authorities and other stakeholders to assess future endogenous price changes, or to inform decisions with direct price effects. Plausible explanations for such alterations to ridehail price, though likely not to the extremes simulated for this example, include: policy-based (e.g., additional fees per ride to support public transit), market-based (e.g., business model change responding to market entrants), or technology-based (e.g., potential future ridehail automation).

Our case study takes place in the nine counties of the San Francisco Bay Area, as listed in Section 3.1. In our simulation, ridehail services are available everywhere in the region. However, vehicles tend to reposition themselves in such a way as to locate themselves in the highest demand areas, resulting in variation in the concentration and availability of ridehail vehicles designed to mimic real-world ridehail services. To calibrate our ridehail simulator, we utilized data recently released by the California Public Utilities Commission (CPUC) which includes detailed ride-level data on one year of all rides Uber provided throughout California (CPUC, 2020). We analyzed the 33 million weekday rides provided in the San Francisco Bay Area from September 2019 through March 2020, and used the analysis to calibrate the ridehail module in BEAM. Using the CPUC data, we established the baseline number of ridehail drivers/vehicles operating per day (17,000), and the baseline scenario base and per mile and per minute prices of solo and pooled ridehail rides. Furthermore, we calibrated various parameters for ridehail positioning and operation in BEAM to better match six target values observed in the CPUC data: total number of ridehail rides per day; average ride distance and wait time; fraction of ridehail VMT that was deadhead miles; fraction of rides that were matched pool requests; and number of daily rides provided per driver. There are about 35,000 ridehail trips and 415,000 kilometers of ridehail vehicle kilometers traveled (VKT) in the simulation with the 10% subsample of the population. These values therefore scale to 350,000 trips and 4,150,000 VKT representation of the baseline with a full population.

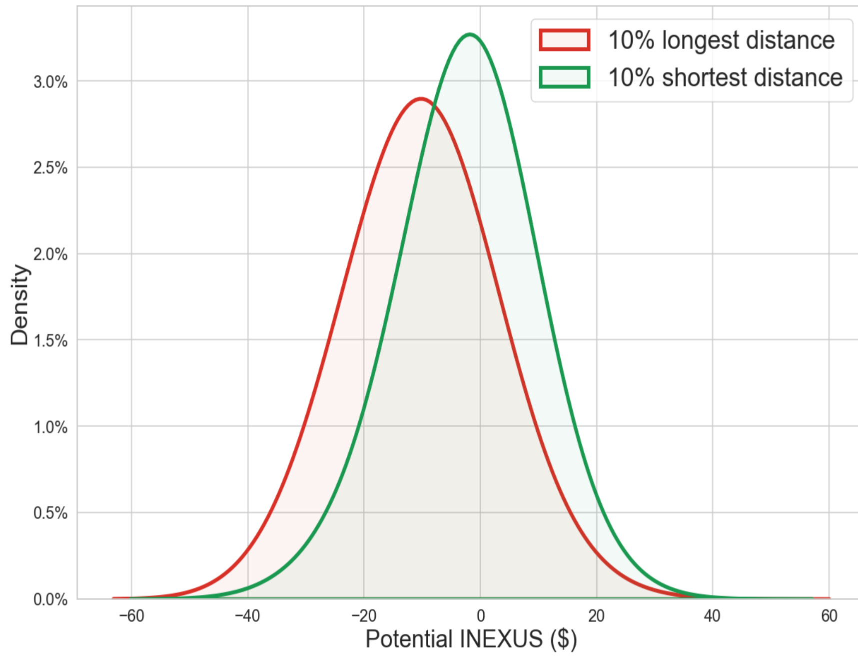
We examine changes to both the Potential and the Realized INEXUS, across different subgroups within the modeled population. We evaluate changes to the aggregated full utility of the modal options available to the individual for their trip as the price of ridehail changes (Potential), as well as the utility associated with the mode used (Realized).

4.2. Potential INEXUS: Baseline scenario distribution in aggregate and by subgroup

We begin by examining the Potential INEXUS distribution under the baseline scenario (Figure 3.4). Panel A presents the baseline Potential INEXUS distribution across the entire population. To provide further insights, we stratify the baseline scenario by trip distance (Panel B), trip type (Panel C), and transport mode (Panel D). Notably, substantial differences emerge between various groups based on trips' characteristics. For instance, non-mandatory trips have 78% higher Potential INEXUS values compared to mandatory trips, as observed from the distributions for those trips having more bulk to the positive side of the distribution. Outcomes are highly variable within distance, type, and mode choice groupings. The Potential INEXUS can highlight differences in transportation-related utility for subpopulations of individuals and trips, which can be defined across household, traveler, and trip characteristics (e.g., trip length, trip type, household income, vehicle ownership, transportation mode).

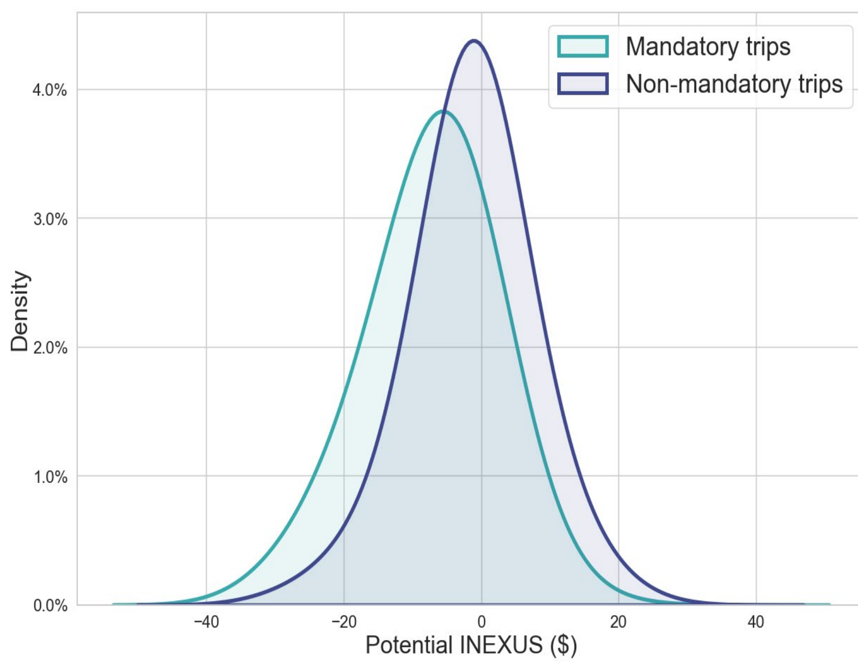


(A)



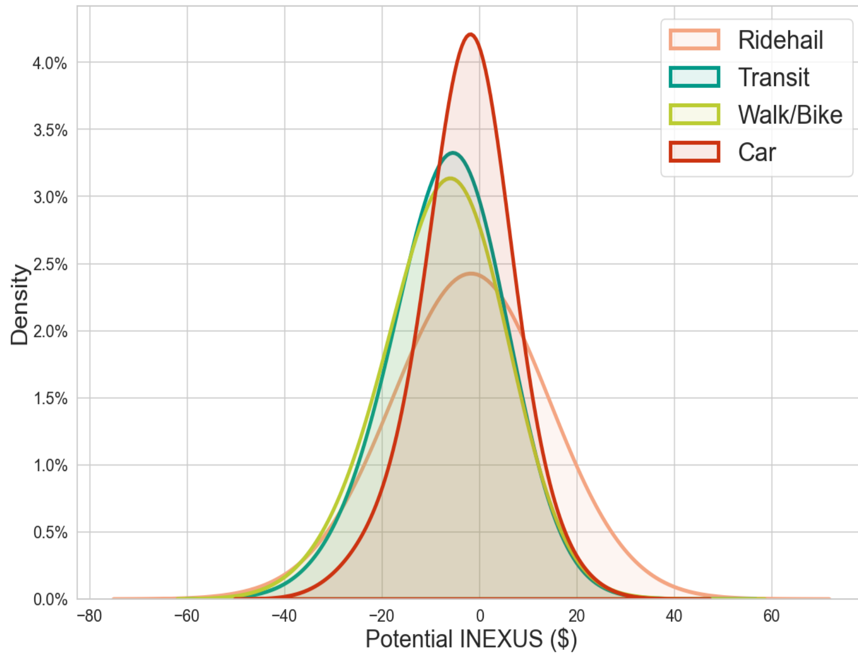
10% longest distance
10% shortest distance
Density
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
−60
−40
−20
0
20
40
60
Potential INEXUS ($)

(B)



Mandatory trips
Non-mandatory trips
Density
4.0%
3.0%
2.0%
1.0%
0.0%
−40
−20
0
20
40
Potential INEXUS ($)

(C)



(D)

Figure 3.4. Potential INEXUS distributions of the baseline scenario (A: aggregate, B: by trip distance, C: by trip type; D: by travel mode chosen)

4.3. Change in Potential INEXUS across scenarios

A multitude of factors such as residence location, mode availability, budget constraints, vehicle ownership, etc. contribute to inequities in the current transportation system. The ability to identify variation in accessibility distributions for different groups of interest can allow for more thorough examination of the degree of inequity in the baseline transportation system. Moreover, it facilitates the strategic design of policies to align the distribution of potential impacts more equitably.

Here, we examine how different types of individuals fare under modeled ridehail price increases and decreases. Figure 3.5 highlights how increasing the affordability of the backup mode (i.e., ridehail) disproportionately benefits low-income individuals. The top panel of Figure 3.5 shows how the distribution of the Potential INEXUS changes for the highest and lowest 10% income decile groups. Specifically, moving from baseline price to no-cost ridehail results in a 33% improvement in the median Potential INEXUS for the lowest income decile compared to 11.5% for the highest income decile (consistent with the fact that the distribution for the lowest income decile in the top panel of Figure 3.5 is shifted more towards the right than for the highest income decile). On the other hand, the change in the Potential INEXUS after increasing ridehail prices to 800% relative to the baseline is more similar in terms of distribution between these two income groups (bottom panel of Figure 3.5). The intuition behind these patterns is discussed further in Section 4.5.

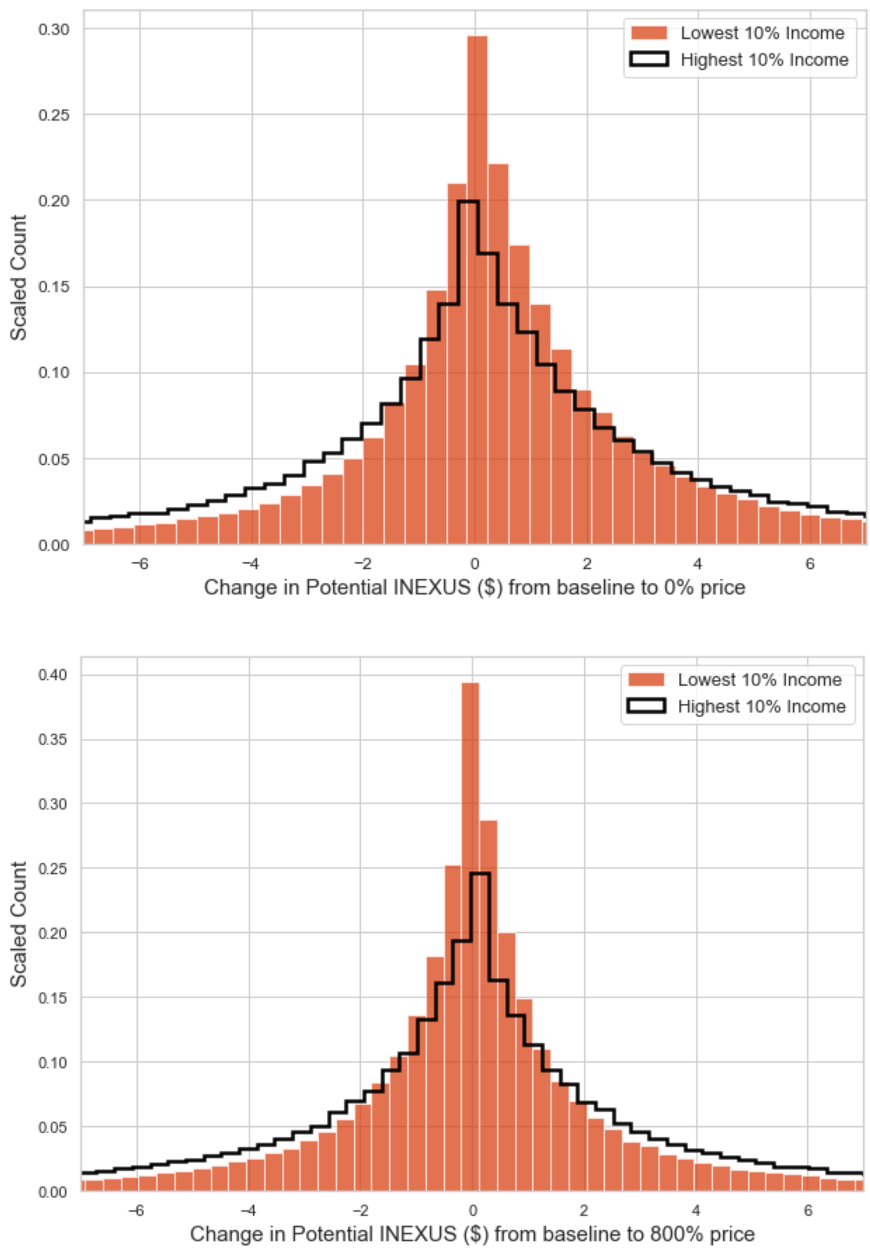


Figure 3.5. Distribution of Potential INEXUS relative to baseline across ridehail price scenarios by the income of travelers

4.4. Insights from comparing Potential and Realized INEXUS across scenarios

INEXUS metrics can be used to identify different types of benefits for various scenarios. We identify two types of benefits associated with changes in ridehail pricing: 1) freeride direct benefit, whereby travelers who utilized ridehail in both the baseline and pricing scenario received benefits without modifying their behavior, 2) indirect benefit, whereby some travelers who did not reoptimize their mode choice to take ridehail still benefit from the availability of a more appealing backup option. Figures 3.6 and 3.7, respectively, depict the Realized and the Potential INEXUS for individuals who did not modify their behavior in response to changing ridehail prices. The top cluster of individuals in Figure 3.6 used ridehail in all scenarios, and they experienced a direct freerider benefit from lower prices, as they received the benefit of lower prices without changing anything about their behavior. For the Potential INEXUS, individuals who utilized non-ridehail modes under all scenarios experienced increased Potential INEXUS values as ridehail became more affordable, demonstrating the value of a more affordable backup mode. This illustrates an indirect benefit associated with greater accessibility, even if a particular option is not actually utilized, which is similar to the option value concept discussed in Van Wee (2016). It can be viewed as a measure of household resilience to mode disruptions.

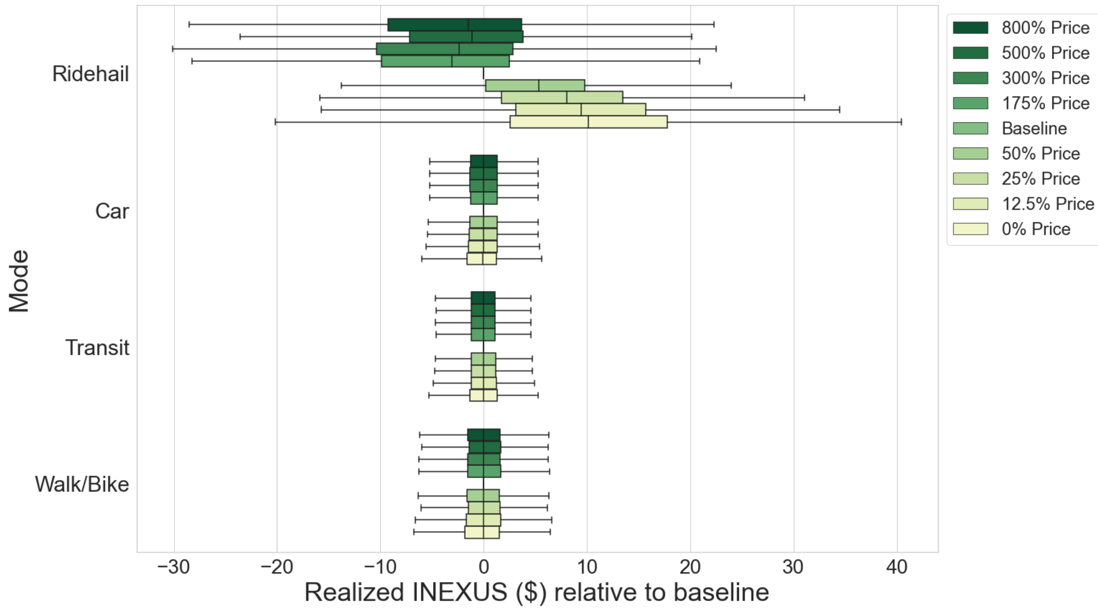


Figure 3.6. Realized INEXUS for travelers who do not change their mode from the baseline

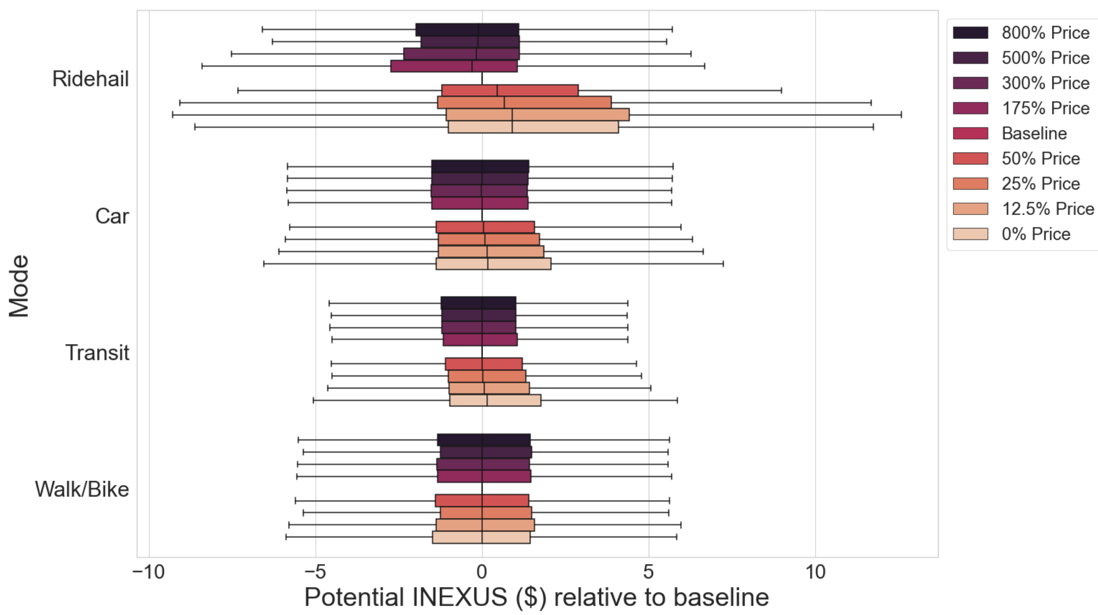


Figure 3.7. Potential INEXUS for travelers who do not change their mode from the baseline

4.5. Consumer Surplus as a welfare summary metric

In this section, we provide an example of using the INEXUS to explicitly calculate and quantify the CS impact of a transportation scenario (as defined in Section 3.2.3), for the entire population and for subpopulations of interest. Table 3.2 and Figure 3.8 illustrate the pricing scenario impacts on CS relative to the baseline. Figure 3.8 displays the change in CS by income decile, with income increasing from left to right; Table 3.2 provides numerical values. Both demonstrate that lower ridehail prices are associated with increased CS, while higher ridehail prices are associated with decreased CS. Equity impacts can also be assessed: while all benefit from lower ridehail prices, lower income deciles gain disproportionately more than high-income deciles.

Figure 3.8. Change in Potential INEXUS relative to baseline (i.e., change in consumer surplus) by income ranks reveals accessibility inequity

Table 3.2. Changes to consumer surplus under ride-hail scenarios

Scenario	Change in Consumer Surplus relative to the baseline scenario ($)					
	Total for the total population (A)	Per person for the total population in one day (B)	Total for the lowest 10% income (C)	Per person for the lowest 10% income in one day (D)	Total for the highest 10% income (E)	Per person for the highest 10% income in one day (F)
Ridehail Price 0%	$874,109.2	$1.35	$135,334.3	$1.93	$51,138.3	$0.82
Ridehail Price 12.5%	$410,963.9	$0.64	$70,561.5	$1.00	$40,531.8	$0.64
Ridehail Price 25%	$380,346.9	$0.59	$52,800.5	$0.75	$24,386.4	$0.39
Ridehail Price 50%	$120,572.9	$0.19	$19,966.8	$0.28	$13,421.0	$0.21
Baseline	$0	$0	$0	$0	$0	$0
Ridehail Price 175%	-$365,187.8	-$0.57	-$22,378.7	-$0.32	-$45,804.3	-$0.73
Ridehail Price 300%	-$403,472.2	-$0.62	-$20,974.2	-$0.30	-$46,588.9	-$0.74
Ridehail Price 500%	-$427,651.8	-$0.66	-$27,945.1	-$0.40	-$55,213.8	-$0.88
Ridehail Price 800%	-$479,743.5	-$0.74	-$15,574.4	-$0.22	-$80,364.9	-$1.28

Moreover, the CS sums the benefit (or detriment) of the change and quantifies it in concrete dollar amounts. This allows a social planner to estimate, in dollars, the benefit or cost of a proposed policy-induced change to the population (or subpopulations), aiding in decision-making either between similar scenarios (price levels) or quite different ones, so long as all have been converted to dollar values (e.g., ridehail price cap vs. school lunch program).

Our findings suggest that reducing or eliminating ridehail costs could potentially reduce transportation accessibility inequalities, particularly for financially constrained travelers. As demonstrated in Figure 3.8, lower ridehail prices disproportionately benefit low-income travelers (a low-income traveler on average would experience a benefit from a change from the baseline to free ridehail that is 2.35 times higher than the benefit to an average high-income traveler), in large part because ridehail becomes a substantially more feasible option compared to the baseline. In contrast, higher ridehail prices disproportionately reduce CS of higher income travelers (a high-income traveler on average would experience a cost from a change from the baseline to the highest ridehail price scenario that is 5.82 times the cost experienced by an average low-income traveler). Few low-income travelers used ridehail in the baseline, so this income group is not as directly affected by the price increase; a larger share of high-income travelers used ridehail in the baseline and are directly affected by a price increase.

While our scenario results hint at policy implications, our primary goal here is to demonstrate that INEXUS metrics are compatible with computing commonly used measures of equity, in addition to many other metrics. Due to the level of underlying detail, INEXUS values can be directly compared between scenarios for each person for each trip they take. Identification of exactly which people are made better or worse off allows for even deeper insights into affected groups and equity of scenario impacts.

5. Conclusions and Future Applications

5.1. The INEXUS metric in brief

In this study, we introduce a novel way of using models and econometric methods that expand the scope of what can be included in estimated impacts of potential policies or system design changes, using approaches to representing accessibility (utility-based, individual agent-based, and positivistic) that is understudied in the literature. In particular, our work extends the state of the art in application of accessibility metrics that meaningfully capture the distributional effects of transportation interventions on key subpopulations. We achieve this through our proposed Potential and Realized INEXUS metrics, which are person-trip specific, and therefore allow distributional estimates of accessibility changes. Examining INEXUS distributions, rather than average impacts, is particularly important when addressing equity for disadvantaged communities. In addition, valuable analyses can consist of a comparison between outcomes across these two metrics and the distinction between the Realized and the Potential INEXUS allows us to disentangle the mechanisms behind accessibility improvements, providing a comprehensive view of a given intervention's impact.

5.2. INEXUS case study insights

Case study insights fall into three broad areas. First, we can evaluate person-trip specific outcomes including the Potential INEXUS and the Realized INEXUS. We demonstrate the value of our rich data in examining outcomes and welfare impacts across groups. We also demonstrate the usefulness of monetizing INEXUS utility values to create metrics readily applied by policy-makers in planning and cost-benefit analysis.

Second, we showed that the INEXUS metrics can be used to assess distributional impacts, beyond simple average effects. We calculate detailed person-trip specific utility, allowing for rich, person-specific estimation of metrics that can be looked at with a fine lens in terms of equity across income, mode choice, etc. This specificity, in data and metrics, is critical to identifying impacted subpopulations and effectively targeting policies.

Third, we identify interesting equity-related results that may have policy relevance. We find that an average traveler in the lowest income decile received approximately 1.9 times the benefit from quartering ridehail prices as compared to the average traveler in the highest income decile (Table 3.2, columns C and E). In particular, while all income groups benefited, the free ridehail scenario resulted in an increase in per-person consumer surplus for the lowest income decile that is over 2.3 times that of the highest income decile.

5.3. The future of INEXUS research

Further examination of welfare and distributional impacts: While we have here discussed distributional impacts across income deciles, there are additional ways to consider welfare, distributional impacts, and equity using INEXUS. As mentioned above, INEXUS is a building block metric, but its value to grapple with questions pertaining to equity, justice, policy evaluation, planning, or any other application depends critically on the way the building block INEXUS values are formulated into specific metrics and insights. The formulation of various types of metrics, specifically with respect to equity (such as equality, proportionality, or Rawlsian justice (Bills, 2022)), have to be driven by the priorities of the policy-maker or researcher, and can have important implications for takeaways from a study. Careful consideration of INEXUS applications with an equity focus must take into account the underlying definitions of equity being employed. For example, making everyone's experience the same versus disproportionately benefiting those that have been historically disenfranchised are two different definitions of "equitable" outcomes with very different implications for different subpopulations. This is something that needs to be taken into account in any future work with INEXUS.

Other heterogeneity aspects: It will be informative to explore equity in accessibility across other characteristics, e.g., race, gender, vehicle ownership, presence of children in household, and disability status. In addition to income, race, and other demographics, future work could include examining differences in INEXUS across different geographies (e.g., census tracts), and how these values change under various policy scenarios.

Additional transportation policy and market changes: In addition to ridehail price changes, there are numerous potential policy, technology, and system design scenarios to investigate, spanning various transportation modes: 1) mass transit (e.g., expanding or improving bus service), 2) ridehail (e.g., fleet size, wheelchair accessibility), and 3) private vehicles (e.g., ownership rates, vehicle technology changes, regulations and mandates, connectivity and automation, micromobility use). Across modes, system-wide changes such as increased reliance on electrification and automation also have implications for traveler outcomes. Explorations of the impacts, including changes to INEXUS distributions for various population groups, are forthcoming for the specific topics of: ridehail fleet size and number of competing ridehail fleets in a region, wheelchair accessibility of ridehail vehicles, availability and use of telecommuting as a work mode, completed and planned Bay Area public transit improvement projects, and entry pricing for heavily traveled traffic cordons.

Exploration of different regions: While we have used the San Francisco Bay Area in this study, BEAM CORE can be used to model many different U.S. geographies; thus, INEXUS can be adapted to examine accessibility impacts of local and regional transportation changes across the country. Other parts of the world could also potentially be examined using INEXUS as well. The models we use are open-source, but feasibility will be conditional on the availability of appropriately detailed data. The necessary data to calibrate mode choices within the models, e.g., population demographics, household characteristics, and travel behaviors, are more likely to be available in open-source format for major cities. A comparison between model performance and outcomes between our SF Bay Area work and additional geographies both inside and outside the U.S. is of great interest for future exploration.

Declaration of Competing Interest

None.

Author Contributions

Author contributions are as follows: study conception and design: NR, ATB, AS, SF, NP; data generation: ZN, CP, JC; data collection, wrangling, analysis, interpretation, results visualization: NR, ATB; manuscript preparation: NR, ATB, AS, SF, CG.

Acknowledgements

This paper and the work described were sponsored by the U.S. Department of Energy (DOE) Vehicle Technologies Office (VTO) under the Systems and Modeling for Accelerated Research in Transportation (SMART) Mobility Laboratory Consortium, an initiative of the Energy Efficient Mobility Systems (EEMS) Program, under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes. The following DOE Office of Energy Efficiency and Renewable Energy (EERE) managers played important roles in establishing the project concept, advancing implementation, and providing ongoing guidance: David Anderson, Prasad Gupte, Heather Croteau, and Jacob Ward.

CONCLUSION

This dissertation has investigated the complexities of urban dynamics, probing the relationships between urban form metrics, the influence of motorways on urban sprawl, and the nuances of transportation accessibility. In three distinct chapters, the research has unearthed insights crucial for policymakers, urban and transportation planners, and researchers, emphasizing the necessity for context-specific strategies that prioritize sustainability, equity, and adaptability in urban development.

As we conclude this dissertation, it is essential to consider the future avenues of research in urban dynamics. The following areas present promising opportunities for further exploration:

- Fine-Grained Motorway Impact Studies: Investigate high level effects of motorways on the sprawl level of the large cities near them to understand variations in their impact on urban sprawl.
- Longitudinal Urban Form Analysis: Conduct longitudinal studies to track changes in urban form metrics over time, providing insights into evolving patterns and the dynamic nature of urban development.
- INEXUS in Different Contexts: Extend the application of the INEXUS metric to diverse urban scenarios and transportation modes and various geographic regions, exploring its effectiveness in capturing accessibility nuances across varied environments.

- Integration of Technological Advancements: Explore how emerging technologies, such as artificial intelligence and data analytics, can enhance urban planning and transportation strategies, optimizing decision-making processes.
- Climate Change Resilience: Assess the resilience of urban areas with various sprawl levels and from different urban form typology clusters to climate change and explore adaptive strategies that integrate environmental sustainability into urban planning.
- Impact of Urban Sprawl on Forest Loss and Habitat Fragmentation: Examine the ecological consequences of urban sprawl, focusing on its impacts on deforestation and habitat fragmentation and understand the broader environmental implications of urban form.
- Expand the Urban Form Metrics study: Considering the availability of data in the future, we can explore the calculation of additional urban form metrics at a global scale, adding various intriguing aspects to the study. Consider widening the definition of green space to encompass other forms of open space, making the metric more comprehensive and inclusive of diverse climate and geographic situations. This is also dependent on the availability of data at the interested scale of analysis.
- Policies for Improving Green Access in Higher Density Areas: Propose policies aimed at enhancing green access in higher density areas. Given the higher density areas' lower emissions, addressing green access challenges can amplify their positive impact on the environment and residents' health.

In summary, this dissertation concludes with a central theme that ties together the various research endeavors, namely a commitment to advancing sustainable, equitable, and adaptable urban futures from a global perspective. Beyond providing specific answers, the dissertation's significance lies in the thought-provoking questions it poses for future researchers, policymakers, urban and transportation planners, and environmental scientists. These questions prompt them to navigate the continually changing landscape of urban dynamics, emphasizing international studies and data-centric approaches. Moreover, highlighting the increasing integration of data science in urban and transportation planning, the overarching theme extends to leveraging data-driven insights to tackle global challenges and shape inclusive cities that meet the diverse needs of inhabitants worldwide.

REFERENCES

Alwadei, M., Srivastava, D., Alam, M. S., Shi, Z., & Bloss, W. J. (2022). Chemical characteristics and source apportionment of particulate matter (PM2.5) in Dammam, Saudi Arabia: Impact of dust storms. *Atmospheric Environment: X*, *14*(February), 100164. https://doi.org/10.1016/j.aeaoa.2022.100164

Angel, S., Blei, A. M., Parent, J., Lamson-Hall, P., & Galarza Sánchez, N. (2016). *Atlas of Urban Expansion—2016 Edition, Volume 1: Areas and Densities*. New York: New York University, UN Habitat, and Cambridge, MA: Lincoln Institute of Land Policy. http://atlasofurbanexpansion.org/data

Angel, S., Parent, J., & Civco, D. (2007). Urban sprawl metrics: An analysis of global urban expansion using GIS. *American Society for Photogrammetry and Remote Sensing - ASPRS Annual Conference 2007: Identifying Geospatial Solutions*, *1*, 22–33.

Angel, S., Parent, J., & Civco, D. L. (2010). Ten compactness properties of circles: Measuring shape in geography. *Canadian Geographer*, *54*(4), 441–461. https://doi.org/10.1111/j.1541-0064.2009.00304.x

Arnott, R. J., & McMillen, D. P. (Eds.). (2006). *A Companion to Urban Economics*. Blackwell Publishing Ltd. https://doi.org/10.1002/9780470996225

Badhrudeen, M., Derrible, S., Verma, T., Kermanshah, A., & Furno, A. (2022). A Geometric Classification of World Urban Road Networks. *Urban Science*, *6*(1), 11. https://doi.org/10.3390/urbansci6010011

Baeten, G. (2001). The Europeanization of brussels and the urbanization of "Europe" hybridizing the city empowerment and disempowerment in the EU district. *European Urban and Regional Studies*, *8*(2), 117–130. https://doi.org/10.1177/096977640100800203

Bagheri, B., & Tousi, S. N. (2018). An explanation of urban sprawl phenomenon in Shiraz Metropolitan Area (SMA). *Cities*, *73*, 71–90. https://doi.org/10.1016/j.cities.2017.10.011

Bakrania, S. (2015). Urbanisation and urban growth in Nepal. *GSDRC Applied Knowledge Services*, 1–24.

Baldassare, M. (2005). Reviewed Work(s): *Up against the Sprawl: Public Policy and the Making of Southern California* by Jennifer Wolch, Manuel Pastor and Peter Drier (Vol. 95). Taylor & Francis, Ltd. on behalf of the Association of American Geographers.

Banai, R., & DePriest, T. (2014). Urban Sprawl: Definitions, Data, Methods of Measurement, and Environmental Consequences. *Journal of Sustainability Education*, *7*(December), 1–15.

Barrington-Leigh, C., & Millard-Ball, A. (2017). More connected urban roads reduce US GHG emissions. *Environmental Research Letters*, *12*(4). https://doi.org/10.1088/1748-9326/aa59ba

Barrington-Leigh, C., & Millard-Ball, A. (2019). A global assessment of street-network sprawl. *PLoS ONE*, *14*(11), 1–16. https://doi.org/10.1371/journal.pone.0223078

Barrington-Leigh, C., & Millard-Ball, A. (2020). Global trends toward urban street-network sprawl. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(4), 1941–1950. https://doi.org/10.1073/pnas.1905232116

Batty, M., & Longley, P. (1994). *Fractal cities: A geometry of form and function*. London: Academic Press. https://doi.org/10.2307/3060277

Baum-Snow, N. (2007). Did Highways Cause Suburbanization? *The Quarterly Journal of Economics*, *122*(2), 775–805.

Bechle, M. J., Millet, D. B., & Marshall, J. D. (2011). Effects of Income and Urban Form on Urban NO2: Global Evidence from Satellites. *Environmental Science & Technology*, *45*(11), 4914–4919.

Ben-Akiva, M., & Bowman, J. L. (1998). Integration of an activity-based model system and a residential location model. *Urban Studies*, *35*(7), 1131–1153. https://doi.org/10.1080/0042098984529

Ben-Akiva, M., & Lerman, S. R. (1979). Disaggregate travel and mobility-choice modelsand measures of accessibility. In *Behavioural Travel Modelling* (pp. 654–679). Routledge. https://doi.org/10.4324/9781003156055-39

Bertaud, A., & Malpezzi, S. (2003). *The Spatial Distribution of Population in 48 World Cities: Implications for Economies in Transition*, 102.

Beyhan, B., Taubenböck, H., Suffa, S., Ullmann, T., Rauh, J., & Dech, S. (2012). Urban Growth and Sprawl of Mersin City , Turkey : Change Analysis Based on Earth Observation and Socio-Economic Data. *Megaron*, *7*(1), 3–25.

Bhatta, B. (2010). Causes and Consequences of Urban Growth and Sprawl. In *Advances in Geographic Information Science* (pp. 17–36). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-05299-6_2

Bills, T. S. (2022). Advancing the practice of regional transportation equity analysis: a San Francisco bay area case study. *Transportation*, *0123456789*. https://doi.org/10.1007/s11116-022-10355-z

Bills, T. S., & Carrel, A. L. (2021). Transit accessibility measurement considering behavioral adaptations to reliability. *Transportation Research Record*, *2675*(5), 265–278. https://doi.org/10.1177/0361198120986567

Bills, T. S., & Walker, J. L. (2017). Looking beyond the mean for equity analysis: Examining distributional impacts of transportation improvements. *Transport Policy*, *54*(June 2015), 61–69. https://doi.org/10.1016/j.tranpol.2016.08.003

Brown, A. E. (2017). Car-less or car-free? Socioeconomic and mobility differences among zero-car households. *Transport Policy*, *60*(September), 152–159. https://doi.org/10.1016/j.tranpol.2017.09.016

Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, *65*(1), 91–98. https://doi.org/10.1016/j.jue.2008.09.002

Burchfield, M., Overman, H. G., Puga, D., & Turner, M. A. (2006). Causes of sprawl: A portrait from space. *Quarterly Journal of Economics*, *121*(2), 587–633. https://doi.org/10.1162/qjec.2006.121.2.587

Center for Neighborhood Technology. (2017). *H + T Index Methods*.

Cerny, J. W. (1975). Sensitivity Analysis of the Boyce-Clark Shape Index. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *12*(1), 21–27. https://doi.org/10.3138/m5r3-22k4-27n8-227t

Cervero, R. (2003). Road expansion, urban growth, and induced travel: A path analysis. *Journal of the American Planning Association*, *69*(2), 145–163. https://doi.org/10.1080/01944360308976303

Cervero, R., & Hansen, M. (2002). Induced travel demand and induced road investment: A simultaneous equation analysis. *Journal of Transport Economics and Policy*, *36*(3), 469–490.

Chen, Y. (2011). Derivation of the functional relations between fractal dimension of and shape indices of urban form. *Computers, Environment and Urban Systems*, *35*(6), 442–451. https://doi.org/10.1016/j.compenvurbsys.2011.05.008

Chen, Y. (2022). Normalizing and classifying shape indexes of cities by ideas from fractals. *Chaos, Solitons and Fractals*, *154*, 111653. https://doi.org/10.1016/j.chaos.2021.111653

Churchman, E. (1999). Disentangling the psyche. *Journal of Planning Literature*, *13*(4), 53–86. https://doi.org/10.4324/9780429952692-3

Clark, L. P., Millet, D. B., & Marshall, J. D. (2011). Air quality and urban form in U.S. urban areas: Evidence from regulatory monitors. *Environmental Science and Technology*, *45*(16), 7028–7035.

Clifton, K., Ewing, R., Knaap, G. J., & Song, Y. (2008). Quantitative analysis of urban form: A multidisciplinary review. *Journal of Urbanism*, *1*(1), 17–45. https://doi.org/10.1080/17549170801903496

Colsaet, A., Laurans, Y., & Levrel, H. (2018). What drives land take and urban land expansion? A systematic review. *Land Use Policy*, *79*, 339–349. https://doi.org/10.1016/j.landusepol.2018.08.017

D'Acci, L. (Ed.). (2019). *The Mathematics of Urban Morphology*. Springer International Publishing. https://doi.org/10.1007/978-3-030-12381-9

Delbosc, A., & Currie, G. (2012). Choice and disadvantage in low-car ownership households. *Transport Policy*, *23*, 8–14. https://doi.org/10.1016/j.tranpol.2012.06.006

Demitiry, M., Higgins, C. D., Páez, A., & Miller, E. J. (2022). Accessibility to primary care physicians: Comparing floating catchments with a utility-based approach. *Journal of Transport Geography*, *101*(March). https://doi.org/10.1016/j.jtrangeo.2022.103356

Deng, X., Huang, J., Rozelle, S., & Uchida, E. (2008). Growth, population and industrialization, and urban land expansion of China. *Journal of Urban Economics*, *63*(1), 96–115. https://doi.org/10.1016/j.jue.2006.12.006

Domencich, T. A., & McFadden, D. (1975). *Urban travel demand - A behavioral analysis*. North-Holland.

Dong, X., Ben-Akiva, M. E., Bowman, J. L., & Walker, J. L. (2006). Moving from trip-based to activity-based measures of accessibility. *Transportation Research Part A: Policy and Practice*, *40*(2), 163–180. https://doi.org/10.1016/j.tra.2005.05.002

Duranton, G., & Puga, D. (2020). The economics of urban density. *Journal of Economic Perspectives*, *34*(3), 3–26. https://doi.org/10.1257/jep.34.3.3

Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, *79*(4), 1407–1440. https://doi.org/10.1093/restud/rds010

Ebrahimpour-Masoumi, H. (2012). Urban sprawl in Iranian cities and its differences with the western sprawl. *Spatium*, *27*, 12–18. https://doi.org/10.2298/SPAT1227012E

Edmonston, B., Goldberg, M. A., & Mercer, J. (1985). Urban Form in Canada and the United States: An Examination of Urban Density Gradients. *Urban Studies*, *22*(3), 209–217. https://doi.org/10.1080/00420988520080351

European Commission. (2023). *GHSL Data Package 2023*. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/098587 (JRC133256)

Ewing, R. (2008). Characteristics, causes, and effects of sprawl: A literature review. *Urban Ecology: An International Perspective on the Interaction Between Humans and Nature*, *21*(2), 519–535. https://doi.org/10.1007/978-0-387-73412-5_34

Ewing, R., & Cervero, R. (2017). "Does Compact Development Make People Drive Less?" The Answer Is Yes. *Journal of the American Planning Association*, *83*(1), 19–25. https://doi.org/10.1080/01944363.2016.1245112

Ewing, R., Meakins, G., Hamidi, S., & Nelson, A. C. (2014). Relationship between urban sprawl and physical activity, obesity, and morbidity - Update and refinement. *Health and Place*, *26*, 118–126. https://doi.org/10.1016/j.healthplace.2013.12.008

Ewing, R., Pendall, R., & Chen, D. (2003). Measuring Sprawl and Its Transportation Impacts. *Transportation Research Record*, *1831*(1831), 175–183. https://doi.org/10.3141/1831-20

Farber, S., Ritter, B., & Fu, L. (2016). Space-time mismatch between transit service and observed travel patterns in the Wasatch Front, Utah: A social equity perspective. *Travel Behaviour and Society*, *4*, 40–48. https://doi.org/10.1016/j.tbs.2016.01.001

Florczyk, A. J., Melchiorri, M., Corbane, C., Schiavina, M., Maffenini, M., Pesaresi, M., Politis, P., Sabo, S., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., & Zanchetta, L. (2019). *Description of the GHS Urban Centre Database 2015: Vol. Version 1.0.* Publications Office of the European Union. https://data.europa.eu/doi/10.2760/037310

Freedman, J., & Hensle, D. (2021). *Activity-Based Travel Demand Modeling Built by and for Users*.

Frumkin, H., Frank, L., & Jackson, R. J. (2004). *Urban Sprawl and Public Health: Designing, Planning, and Building for Healthy Communities*. Island Press.

Galster, G., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S., & Freihage, J. (2001). Wrestling sprawl to the ground: Defining and measuring an elusive concept. *Housing Policy Debate*, *12*(4), 681–717. https://doi.org/10.1080/10511482.2001.9521426

Garcia-López, M. À. (2019). All roads lead to Rome … and to sprawl? Evidence from European cities. *Regional Science and Urban Economics*, *79*(November 2018), 103467. https://doi.org/10.1016/j.regsciurbeco.2019.103467

Garcia-López, M. Á., Holl, A., & Viladecans-Marsal, E. (2015). Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics*, *85*, 52–67. https://doi.org/10.1016/j.jue.2014.11.002

Gately, C. K., Hutyra, L. R., & Wing, I. S. (2015). Cities, traffic, and CO2: A multidecadal assessment of trends, drivers, and scaling relationships. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(16), 4999–5004. https://doi.org/10.1073/pnas.1421723112

Geertman, S. C. M., & Ritsema Van Eck, J. R. (1995). Research article: GIS and models of accessibility potential: An application in planning. *International Journal of Geographical Information Systems*, *9*(1), 67–80. https://doi.org/10.1080/02693799508902025

Giannadaki, D., Lelieveld, J., & Pozzer, A. (2016). Implementing the US air quality standard for PM2.5 worldwide can prevent millions of premature deaths per year. *Environmental Health: A Global Access Science Source*, *15*(1), 1–11. https://doi.org/10.1186/s12940-016-0170-8

Glaeser, E. L., & Kahn, M. E. (2004). Chapter 56 Sprawl and urban growth. In *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2481–2527). Elsevier. https://doi.org/10.1016/S1574-0080(04)80013-0

Goswami, M., & Khire, M. V. (2016). Land Use and Land Cover Change Detection for Urban Sprawl Analysis of Ahmedabad City using Multitemporal Landsat Data. *International Journal of Advanced Remote Sensing and GIS*, *5*(1), 1670–1677. https://doi.org/10.23953/cloud.ijarsg.51

Gouda, A. A., Hosseini, M., & Masoumi, H. E. (2016). The Status of Urban and Suburban Sprawl in Egypt and Iran. *GeoScape*, *10*(1), 1–15. https://doi.org/10.1515/geosc-2016-0001

Gulhan, G., Ceylan, H., Özuysal, M., & Ceylan, H. (2013). Impact of utility-based accessibility measures on urban public transportation planning: A case study of Denizli, Turkey. *Cities*, *32*, 102–112. https://doi.org/10.1016/j.cities.2013.04.001

Guzman, L. A., Cantillo-Garcia, V. A., Oviedo, D., & Arellana, J. (2023). How much is accessibility worth? Utility-based accessibility to evaluate transport policies. *Journal of Transport Geography*, *112*(August), 103683. https://doi.org/10.1016/j.jtrangeo.2023.103683

Haggett, P., Cliff, A. D., & Allan, F. (1977). *Locational Analysis in Human Geography* (2nd edition). Edward Arnold.

Hamidi, S., & Ewing, R. (2014). A longitudinal study of changes in urban sprawl between 2000 and 2010 in the United States. *Landscape and Urban Planning*, *128*, 72–82. https://doi.org/10.1016/j.landurbplan.2014.04.021

Hamidi, S., Ewing, R., Preuss, I., & Dodds, A. (2015). Measuring Sprawl and Its Impacts: An Update. *Journal of Planning Education and Research*, *35*(1), 35–50. https://doi.org/10.1177/0739456X14565247

Handy, S., Paterson, R., Song, J., Rajamani, J., Jung, J., Bhat, C., & Kockelman, K. (2002). *Techniques for Mitigating Urban Sprawl: Goals, Characteristics, and Suitability Factors.* Center for Transportation Research, The University of Texas at Austin.

Harb, M., Walker, J. L., Malik, J., & Circella, G. (2022). Estimating short-term travel demand models that incorporate personally owned autonomous vehicles. *Travel Behaviour and Society*, *26*(October 2021), 279–289. https://doi.org/10.1016/j.tbs.2021.10.008

Hayden, D. (2004). *A Field Guide to Sprawl*. W. W. Norton and Company.

Horni, A., Nagel, K., & Axhausen, K. W. (2016). *The Multi-Agent Transport Simulation Title of Book : The Multi-Agent Transport Simulation MATSim*.

Hou, Y., Garikapati, V., Nag, A., Young, S. E., & Grushka, T. (2019). Novel and Practical Method to Quantify the Quality of Mobility: Mobility Energy Productivity Metric. *Transportation Research Record*, *2673*(10), 141–152. https://doi.org/10.1177/0361198119848705

Hsueh, G., Czerwinski, D., Poliziani, C., Becker, T., Hughes, A., Chen, P., & Benn, M. (2021). *Using BEAM Software to Simulate the Introduction of On-Demand, Automated, and Electric Shuttles for Last Mile Connectivity in Santa Clara County* (Issue January).

Huang, J., Lu, X. X., & Sellers, J. M. (2007). A global comparative analysis of urban form: Applying spatial metrics and remote sensing. *Landscape and Urban Planning*, *82*(4), 184–197. https://doi.org/10.1016/j.landurbplan.2007.02.010

Ilesanmi, A. (2012). The roots and fruits of gated communities in Lagos, Nigeria: social sustainability or segregation? *Sustainable Futures: Architecture and Urbanism in the Global South*, *June*, 105–112.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Järv, O., Tenkanen, H., Salonen, M., Ahas, R., & Toivonen, T. (2018). Dynamic cities: Location-based accessibility modelling as a function of time. *Applied Geography*, *95*(November 2017), 101–110. https://doi.org/10.1016/j.apgeog.2018.04.009

Karlström, A., & Franklin, J. P. (2009). Behavioral adjustments and equity effects of congestion pricing: Analysis of morning commutes during the Stockholm Trial. *Transportation Research Part A: Policy and Practice*, *43*(3), 283–296. https://doi.org/10.1016/j.tra.2008.09.008

Kwan, M. (1998). Accessibility : A Comparative Analysis Using a Point-based Framework. *GeoGraphical Analysis*, *30*(9512451), 191–216.

Laarabi, H., Needell, Z., Waraich, R., Poliziani, C., & Wenzel, T. (2023). *BEAM : The Modeling Framework for Behavior , Energy , Autonomy & Mobility* (Issue July).

Lee, A. C. K., Jordan, H. C., & Horsley, J. (2015). Value of urban green spaces in promoting healthy living and wellbeing: Prospects for planning. *Risk Management and Healthcare Policy*, *8*, 131–137. https://doi.org/10.2147/RMHP.S61654

Lemoine-Rodríguez, R., Inostroza, L., & Zepp, H. (2020). The global homogenization of urban form. An assessment of 194 cities across time. *Landscape and Urban Planning*, *204*, 103949. https://doi.org/10.1016/j.landurbplan.2020.103949

Levkovich, O., Rouwendal, J., & van Ommeren, J. (2020). The impact of highways on population redistribution: The role of land development restrictions. *Journal of Economic Geography*, *20*(3), 783–808. https://doi.org/10.1093/JEG/LBZ003

Lewis, R., Knaap, G., & Schindewolf, J. (2013). The Spatial Structure of Cities in the United States. *Lincoln Institute of Land Policy; Working Paper*, *WP13RL1*.

Liang, Z., Wu, S., Wang, Y., Wei, F., Huang, J., Shen, J., & Li, S. (2020). The relationship between urban form and heat island intensity along the urban development gradients. *Science of the Total Environment*, *708*, 135011. https://doi.org/10.1016/j.scitotenv.2019.135011

Liu, L., & Meng, L. (2020). Patterns of Urban Sprawl from a Global Perspective. *Journal of Urban Planning and Development*, *146*(2). https://doi.org/10.1061/(asce)up.1943-5444.0000558

Liu, Y., Song, Y., & Arp, H. P. (2012). Examination of the relationship between urban form and urban eco-efficiency in china. *Habitat International*, *36*(1), 171–177. https://doi.org/10.1016/j.habitatint.2011.08.001

Lowry, J. H., & Lowry, M. B. (2014). Comparing spatial metrics that quantify urban form. *Computers, Environment and Urban Systems*, *44*, 59–67. https://doi.org/10.1016/j.compenvurbsys.2013.11.005

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *2017-Decem*, 4766–4775.

Maceachren, A. M. (1985). Compactness of geographic shape: comparison and evaluation of measures. *Geografiska Annaler, Series B*, *67*(1), 53–67. https://doi.org/10.1080/04353684.1985.11879515

Macfarlane, G. S., & Lant, N. J. (2021). *Estimation and Simulation of Daily Activity Patterns for Individuals Using Wheelchairs* (Issue June).

Marshall, W. E., Piatkowski, D. P., & Garrick, N. W. (2014). Community design, street networks, and public health. *Journal of Transport and Health*, *1*(4), 326–340. https://doi.org/10.1016/j.jth.2014.06.002

Meier, P. M., & McCoy, T. H. (2010). Some Extensions of Mills's Method for Urban Population Density Gradient Estimation. *Geographical Analysis, 10*(2), 163–173. https://doi.org/10.1111/j.1538-4632.1978.tb00005.x

Melchiorri, M., & Siragusa, A. (2018). Analyzing Cities with the Global Human Settlement Layer: A Methodology to Compare Urban Growth Using Remote Sensing Data. A. Bisello, D. Vettorato, P. Laconte, & S. Costa (Eds.), *Smart and Sustainable Planning for Cities and Regions* (pp. 151–165). Springer International Publishing. https://doi.org/10.1007/978-3-319-75774-2_11

Milam, R. T., Birnbaum, M., Ganson, C., Handy, S., & Walters, J. (2017). Closing the induced vehicle travel gap between research and practice. *Transportation Research Record*, *2653*(January), 10–16. https://doi.org/10.3141/2653-02

Millard-Ball, A., West, J., Rezaei, N., & Desai, G. (2022). What do residential lotteries show us about transportation choices? *Urban Studies*, *59*(2), 434–452. https://doi.org/10.1177/0042098021995139

Mokhtarian, P. L., Salomon, I., & Singer, M. E. (2015). What Moves Us? An Interdisciplinary Exploration of Reasons for Traveling. *Transport Reviews*, *35*(3), 250–274. https://doi.org/10.1080/01441647.2015.1013076

Munir, S., Gabr, S., Habeebullah, T. M., & Janajrah, M. A. (2016). Spatiotemporal analysis of fine particulate matter (PM2.5) in Saudi Arabia using remote sensing data. *Egyptian Journal of Remote Sensing and Space Science*, *19*(2), 195–205. https://doi.org/10.1016/j.ejrs.2016.06.001

Narayani, A. R., & Nagalakshmi, R. (2023). Understanding urban sprawl trends in peri urban regions across global cities - survey of case studies. *Cities and Health*, *7*(3), 492–504. https://doi.org/10.1080/23748834.2022.2106771

Newman, P., & Kenworthy, J. (2021). Gasoline Consumption and Cities Revisited: What Have We Learnt? *Current Urban Studies*, *09*(03), 532–553. https://doi.org/10.4236/cus.2021.93032

Niemeier, D. A. (1997). Accessibility: An evaluation using consumer welfare. *Transportation*, *24*(4), 377–396. https://doi.org/10.1023/A:1004914803019

Noland, R. B., & Lem, L. L. (2002). A review of the evidence for induced travel and changes in transportation and environmental policy in the US and the UK. *Transportation Research Part D: Transport and Environment*, *7*(1), 1–26. https://doi.org/10.1016/S1361-9209(01)00009-8

OECD. (2015). *Environment at a Glance 2015: OECD Indicators.* OECD. https://doi.org/10.1787/9789264235199-en

Ogle, J., Delparte, D., & Sanger, H. (2017). Quantifying the sustainability of urban growth and form through time: An algorithmic analysis of a city's development. *Applied Geography*, *88*, 1–14. https://doi.org/10.1016/j.apgeog.2017.08.016

Opara, A. I., Akaolisa, C. Z., Akakuru, C. O., Nkwoada, A. U., Ibe, F. C., Verla, A. W., & Chukwuemeka, I. C. (2021). Particulate matter exposure and non-cancerous inhalation health risk assessment of major dumpsites of Owerri metropolis, Nigeria. *Environmental Analysis Health and Toxicology*, *36*(4), 12. https://doi.org/10.5620/eaht.2021025

Oueslati, W., Alvanides, S., & Garrod, G. (2015). Determinants of urban sprawl in European cities. *Urban Studies*, *52*(9), 1594–1614. https://doi.org/10.1177/0042098015577773

Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: Positive and normative implementations of various accessibility indicators. *Journal of Transport Geography*, *25*, 141–153.

https://doi.org/10.1016/j.jtrangeo.2012.03.016

Pernice, R. (2007). Urban Sprawl in Postwar Japan and the Vision of the City based on the Urban Theories of the Metabolists' Projects. *Journal of Asian Architecture and Building Engineering*, *6*(2), 237–244. https://doi.org/10.3130/jaabe.6.237

Pratama, A. P., Yudhistira, M. H., & Koomen, E. (2022). Highway expansion and urban sprawl in the Jakarta Metropolitan Area. *Land Use Policy*, *112*(November 2021), 105856. https://doi.org/10.1016/j.landusepol.2021.105856

Rezaei, N., & Millard-Ball, A. (2023). Urban form and its impacts on air pollution and access to green space: A global analysis of 462 cities. *PLoS ONE*, *18*(1 January), 1–26. https://doi.org/10.1371/journal.pone.0278265

Schneider, A., & Woodcock, C. E. (2008). Compact, dispersed, fragmented, extensive? A comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information. *Urban Studies*, *45*(3), 659–692. https://doi.org/10.1177/0042098007087340

Sevtsuk, A., & Amindarbari, R. (2013). *Measuring Growth and Change in Metropolitan Form.* MIT City Form Lab. http://web.mit.edu/11.521/papers/sevtsuk_measuring_growth_and_change_in_metropolitan_form.pdf

Sohn, J., Choi, S., Lewis, R., & Knaap, G. (2012). Characterising urban sprawl on a local scale with accessibility measures. *Geographical Journal*, *178*(3), 230–241. https://doi.org/10.1111/j.1475-4959.2012.00468.x

Song, Y., & Knaap, G. J. (2004). Measuring urban form: Is portland winning the war on sprawl? *Journal of the American Planning Association*, *70*(2), 210–225. https://doi.org/10.1080/01944360408976371

Szinai, J. K., Sheppard, C. J. R., Abhyankar, N., & Gopal, A. R. (2020). Reduced grid operating costs and renewable energy curtailment with electric vehicle charge management. *Energy Policy*, *136*, 111051. https://doi.org/10.1016/j.enpol.2019.111051

Terzi, F., & Bolen, F. (2009). Urban Sprawl Measurement of Istanbul. *European Planning Studies*, *17*(10), 1559–1570. https://doi.org/10.1080/09654310903141797

Timsina, N. P., Shrestha, A., Poudel, D. P., & Upadhyaya, R. (2020). Trend of urban growth in Nepal with a focus in Kathmandu Valley: A review of processes and drivers of change. *The UKRI GCRF Urban Disaster Risk Hub*.

www.tomorrowscities.org

Torrens, P. M., & Alberti, M. (2000). Measuring Sprawl. *Centre for Advanced Spatial Analysis, UCL*, 34.

Train, K. E. (2003). Discrete choice methods with simulation. In *Discrete Choice Methods with Simulation* (Vol. 9780521816). Cambridge University Press. https://doi.org/10.1017/CBO9780511753930

van Wee, B. (2016). Accessible accessibility research challenges. *Journal of Transport Geography*, *51*, 9–16. https://doi.org/10.1016/j.jtrangeo.2015.10.018

Vermeiren, K., Van Rompaey, A., Loopmans, M., Serwajja, E., & Mukwaya, P. (2012). Urban growth of Kampala, Uganda: Pattern analysis and scenario development. *Landscape and Urban Planning*, *106*(2), 199–206. https://doi.org/10.1016/j.landurbplan.2012.03.006

Vickerman, R. W. (1974). Accessibility, Attraction, and Potential: A Review of Some Concepts and Their Use in Determining Mobility. *Environment and Planning A: Economy and Space*, *6*(6), 675–691. https://doi.org/10.1068/a060675

Volker, J. M. B., Lee, A. E., & Handy, S. (2020). Induced Vehicle Travel in the Environmental Review Process. *Transportation Research Record*, *2674*(7), 468–479. https://doi.org/10.1177/0361198120923365

Wachs, M., & Kumagai, T. G. (1973). Physical accessibility as a social indicator. *Socio-Economic Planning Sciences*, *7*(5), 437–456. https://doi.org/10.1016/0038-0121(73)90041-4

Waddell, P. (2008). UrbanSim : An Evolving Planning Support System for Evolving Communities. *Policy Analysis*, 103–138. http://www.urbansim.org/papers/

Waddell, P., Boeing, G., Gardner, M., & Porter, E. (2018). *An Integrated Pipeline Architecture for Modeling Urban Land Use, Travel Demand, and Traffic Assignment*. *January*. http://arxiv.org/abs/1802.09335

Wang, Y., Chen, B. Y., Yuan, H., Wang, D., Lam, W. H. K., & Li, Q. (2018). Measuring temporal variation of location-based accessibility using space-time utility perspective. *Journal of Transport Geography*, *73*(August), 13–24. https://doi.org/10.1016/j.jtrangeo.2018.10.002

Wassmer, R. W. (2006). The influence of local urban containment policies and statewide growth management on the size of United States urban areas. In *Journal of Regional Science* (Vol. 46, Issue 1). https://doi.org/10.1111/j.0022-4146.2006.00432.x

Wheeler, S. M. (2008). The evolution of built landscapes in Metropolitan Regions. *Journal of Planning Education and Research*, *27*(4), 400–416. https://doi.org/10.1177/0739456X08315889

Widziewicz, K., & Loska, K. (2016). Metal induced inhalation exposure in urban population: A probabilistic approach. *Atmospheric Environment*, *128*, 198–207. https://doi.org/10.1016/j.atmosenv.2015.12.061

Williams, H. C. W. . (1976). Travel demand models, duality relations and User Benefit Analysis. *Journal of Regional Science*, *16*(8), 147–166.

Williams, K. (2014). Urban Form and Infrastructure : a Morphological Review. *Foresight, the UK Government Office for Science*, 58. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/324161/14-808-urban-form-and-infrastructure-1.pdf

Wilson, B., & Chakraborty, A. (2013). The environmental impacts of sprawl: Emergent themes from the past decade of planning research. *Sustainability (Switzerland)*, *5*(8), 3302–3327. https://doi.org/10.3390/su5083302

Ye, X., Konduri, K. C., Pendyala, R. M., Sana, B., & Waddell, P. (2009). Methodology to Match Distributions of Both Household and Person Attributes in Generation of Synthetic Populations. *Transportation Research Board Annual Meeting 2009*, *9601*(206), 1–24.

Zanganeh Shahraki, S., Sauri, D., Serra, P., Modugno, S., Seifolddini, F., & Pourahmad, A. (2011). Urban sprawl pattern and land-use change detection in Yazd, Iran. *Habitat International*, *35*(4), 521–528. https://doi.org/10.1016/j.habitatint.2011.02.004

Zhang, T. (2001). Community features and urban sprawl: The case of the Chicagometropolitan region. *Land Use Policy*, *18*(3), 221–232. https://doi.org/10.1016/S0264-8377(01)00018-7

Živković, J. (2019). Urban Form and Function. In W. Leal Filho, T. Wall, U. Azeiteiro, A. M. Azul, L. Brandli, & P. G. Özuyar (Eds.), *Good Health and Well-Being* (pp. 1–10). Springer International Publishing. https://doi.org/10.1007/978-3-319-71063-1_78-1