University of Cincinnati

Date: 3/28/2016

I, Zhuo Yao, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Civil Engineering.

It is entitled:

Smart Data Driven and Adaptive Modeling Framework for Quantifying Dynamic TAZ-based Household Travel Carbon Emissions

Student's name: **Zhuo Yao**

This work and its defense approved by:

Committee chair: Heng Wei, Ph.D.

Committee member: Andrew S Rohne, M.ENG.

Committee member: Steven Buchberger, Ph.D.

Committee member: Xinhao Wang, Ph.D.

19645

Smart Data Driven and Adaptive Modeling Framework for Quantifying Dynamic TAZ-based Household Travel Carbon Emissions

A dissertation submitted to the

Division of Research and Advanced Studies of the University of Cincinnati

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Civil and Architectural Engineering and Construction Management
College of Engineering and Applied Science

March, 2016

By

Zhuo Yao

M.S., Beijing University of Technology, China, 2008

B.S., Polytechnic University of the Philippines, the Philippines, 2004

Committee Members

Dr. Heng Wei, Chair

Dr. Steven Buchberger

Mr. Andrew Rohne

Dr. Xinhao Wang

ABSTRACT

The conventional carbon emission-modeling framework focuses on the link-based emissions and then aggregate to regional inventory. This approach is incapable of tracing emissions back to its geographical origin and providing information on areas where adaptive planning policies and strategies are needed. Recent studies also indicate potential deficiencies in converting four-step travel demand outputs into the inputs of emission models. Emission models often rely on four-step models for vehicle activity inputs. However, these models are mostly calibrated and validated using aggregated daily traffic data. No data sources are available to validate the models at the hourly or the most desired second-by-second level for emission estimates. The recent advancement of mobile device sensors and data transmitting technologies provide travel trajectories (e.g., latitude, longitude, speed, acceleration, altitude) collected from the users of smart phones or other GPS-enabled devices. The availability of such data sources will actually provide new opportunities of enhancing our understanding and modeling of the dynamics between land use pattern, travel behavior, and the associated environmental impacts. These trends call for the emergence of quick-response modeling framework that could be supported by the smart data source.

In this research, a research question is proposed to well direct the proposed research: is it positively possible to use the Smart-Data structured data sets to unveil the sophisticated dynamics between land use changes and its associated carbon emission impacts, if a smart data adaptive modeling framework for this attempt is well developed? The answer to the research question will benefit the integration of the actual and scenario-based land use visioning and planning, demographic changes, transportation emission analysis, and computer forecasting and evaluation of future scenarios. This research makes it possible to assess the household travel

carbon footprint and provides supportive models, and data sets for possible carbon emission mitigation through land use policies and adaptation. The quick response modeling framework using GPS survey data simulated Smart Data provides connections among land use, household socioeconomic and their travel carbon emissions. It is a practical tool for Metropolitan Planning Organizations (MPO) and other planning agencies to compare alternative planning scenarios with spatial details. The responsiveness, or sensitivity, of the model to changes in key inputs indicates whether the model can reasonably estimate the expected change in carbon emissions resulting from the changes in the socioeconomic characteristics.

This page intentionally left blank.

ACKNOWLEDGEMENTS

When minutes become hours, when days become years, I own the greatest gratitude to my advisor, Dr. Heng Wei for guiding and supporting me from the very beginning of the journey towards my Ph.D.. I am incredibly fortunate to have Dr. Wei as my advisor because he has been tirelessly giving insightful mentoring, ability of critical thinking, in-depth research training, and opportunities to engage in many research and teaching activities. He is the most caring person concerning my development in research, scholarship, and teaching. I own him a debt because this dissertation would not have been possible without his countless guidance and help. He is definitely the one individual that influences my life the deepest both professionally and personally. I will never forget his help in times of hardship; his motivation in times of lost and his rigorous requirements at all times. My deepest and most sincere gratitude go to Dr. Wei.

I am appreciative of Dr. Xinhao Wang for bringing me to the U.S. EPA sponsored research project. My entire dissertation is inspired from the preliminary results of the project. I would like to thank his great support and meaningful comments and discussions. My gratitude goes to Dr. Steven Buchberger for offering me so much guidance, help and continuous support of my research. I am very appreciative of Mr. Andrew Rohne for providing the OKI GPS survey data, which made this dissertation possible. Also, thank you for all the supports, meaningful discussions and continuous support on the dissertation.

I am also in debt to great help from my fellow colleagues and graduate researchers: Harikishan Perugu, Zhixia Li, and Ting Zuo. Special thanks to Mr. Hao Liu for his assistance, support and meaningful discussions.

Last, but not least, I would like to thank my wife, Huaiqing, for always being there with love and understanding. To my lovely daughter Claire, I really enjoyed the happiness you brought to me.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Background

A commonly accepted carbon footprint definition is: “A measure of the total amount of carbon dioxide (CO_2) and methane (CH_4) emissions of a defined population, system or activity, considering all relevant sources, sinks and storage within the spatial and temporal boundary of the population, system or activity of interest. Calculated as carbon dioxide equivalent (CO_2E) using the relevant 100-year global warming potential (GWP100)(Wright, et al., 2011)”. The United States Environmental Protection Agency (U.S. EPA) reported that the historical increase of CO_2 emissions from transportation end user sector is largely contributable to increased and imbalanced demand for land use and travel activities(USEPA, 2012). The current state of the practice for estimating carbon emission relies on the integration of two isolated modeling process: travel demand forecasting and emission estimating. The procedure employs an ad-hoc approach using average link-based speed and traffic volume from travel demand model as transportation activities related inputs to the MOVES model(USEPA, 2012; USEPA, 2010; FHWA, 2010). Climate change, land use and socioeconomic developments are principal variables that define the need and scope of adaptive engineering and management to sustain infrastructure development. It is of Federal (e.g. U.S. EPA) and state governments (e.g. California Air Resources Board) great interests to investigate research questions such as are the changes tangible? What are the actionable sciences for decision-making? What adaptation changes can be made in the planning horizon? Urban adaptation planning for climate change impacts requires data-driven, location-based analysis capability to estimate the spatial distribution of travel carbon emission contributing sources due to transportation activities. Therefore, household travel carbon emission modeling is viewed a pressing need to provide data and location-driven decision support to addressing the aforementioned research questions and analysis capabilities. The challenge remains in the theoretical representation of sensitive interactions between spatial-dependent land use and traffic activities and providing location-based carbon emission information for decision makers.

On the land use planning perspective, adaptive planning for scenarios that are likely considered in the urban planning process for visioning possible future developments in providing tools and data for decision support. Over the past two decades, a growing body of research has aimed at improving the understanding of the influence of socioeconomic changes and its corresponding changes in transportation pattern and its related emissions. Usually, such analyses investigate the interaction between specific attributes of land use, household travel demand and associated vehicle emissions(Cervero & Kockelman, 1997), while accounting for demographic factors cited as significant in the literature, including income, household size, and vehicle ownership using the available household travel survey data (Cervero, 1996; Frank, et al., 2000). However, in practice, it is very difficult for decision makers to make adaptive changes on land use policies (incentives or regulations) if the supporting traffic and emission data are link-based that is in an aggregated spatial and temporal level. Simply put, the link-based traffic is actually generated from other Traffic Analysis Zones (TAZ) which cannot be traced back by current aggregated traffic and emission modeling approach. The difficulty lies in the highly aggregated nature of link-based traffic and emission data. Although the activity-based travel demand model is capable of tracing the traffic back to its source (i.e. TAZ), the challenge remains in the absence of the activity-based model itself since it is new and expensive to implement. Therefore, there is a gap in between the need for household origin based travel demand and carbon emission analysis and the need for knowing where the link-based traffic came from.

On the travel demand modeling perspective, the growing complexity of travel demand patterns, as well as residential and firm location choice processes, adds to the challenge of better estimating carbon emissions and achieving more sustainable development patterns thus creating additional pressure on the evaluation of policies for decision makers. Indeed, there are few examples where these models have been extended with capabilities for simultaneous evaluation of emissions, air quality and exposure as a result of land-use and transport policy scenarios(Hatzopoulou, 2010). Despite significant efforts in data collection and developing trip-based, tour-based, and activity-based models, they still lack the capability of fully synthesizing emission impacts considering the spatial locations of household as a generation source. A trip-based model (the most commonly used approach) ignores some natural spatial and temporal constraints by modeling trip decisions separately and excluding the modeling of duration and time of day(Bowman & Ben-Akiva, 2000). The highly aggregated results undermined the capability of producing high resolution traffic data such as individual car operating conditions and the capability of tracing the vehicles on road to its origin source.

On the emission modeling perspective, traffic operation activity inputs to MOVES model are crucial in maximizing its capability to accurately reflect the greenhouse gas emission associated. Previous research(Song, et al., 2008; Frey, et al., 2008; Fulper, et al., 2011; Yao, et al., 2014) has proven that on-road traffic related emission varies with traffic operating conditions (i.e., speed, acceleration or deceleration). Recent studies indicate potential deficiencies in converting travel demand outputs into the emission model inputs. Emission models often rely on traditional travel demand models for vehicle activity input, but traditional travel demand models are mostly calibrated and validated by using aggregated total traffic data (Bachman, et al., 2000). Therefore, the hourly emission estimates may not be accurate because hourly VMT and speed variations are underrepresented as well as for the reason that aggregated inputs are used in the emission models (Bachman, et al., 2000; Bhat, et al., 2003). In addition, the real-world traffic data, especially location-based trip generations are in panel data format. Therefore, it contains unknown effects due to its spatial and temporal correlation (Nittal, et al., 2004; Wang, et al., 2010).

Transportation planning models use aggregated attributes such as total household and total employment based on TAZ. The attributes that are associated with the TAZs are spatially dependent as shown from previous studies(Bhat & Huimin, 2002; Miller, 1999; Lopes, et al., 2014). However, the effectiveness of incorporating spatial information into integrated land use, transportation and environmental impact analysis is not clear. In practice, it is very difficult for decision makers to make adaptive changes on land use policies (incentives or regulations) when the supporting traffic and emission data are link-based that is in an aggregated spatial and temporal level. To make plans that are adaptive to changes in socioeconomic of household, climate changes, etc. it is critical to know which TAZs are producing higher carbon emissions. Unfortunately, current modeling practices only quantify emission at a roadway link level and are incapable of tracing emissions back to its origin (Nittal, et al., 2004; Wang, et al., 2010). Therefore, there is a research gap between the need for knowing where emissions were generated from TAZ level and the current link-based outputs from travel demand models.

A large body of literature(Kockelman & Frazier, 2006; Wang, et al., 2012; Wang, et al., 2013) has proven that incorporating spatial factors into integrated land use and transportation applications are applicable and yields reliable results (Zhou & Kockelman, 2008; Zhou & Kockelman, 2009; Parent & LeSage, 2010). The spatial and temporal correlation characteristics which are originally introduced to the transportation field from Econometrics consider traffic activities, similar to its source generation, are spatially correlated. Several recent studies at the University of Cincinnati(Perugu, et al., 2012; Perugu, et al., 2013) indicate that using a spatial panel model is capable of achieving improved accuracy in both truck volume and Particulate Matter (PM $_{2.5}$) emission predictions.

Therefore, limited by the aggregated modeling assumptions and insufficient data support, it is difficult to make theoretical contributions connecting land use (i.e., socioeconomic) and household travel associated carbon emissions and makes it traceable to its origin. Besides, since the household travel survey data analyses are cross-sectional studies and are spatially and temporally dependent, the effectiveness of incorporating spatial information into the research is not clear. A method of modeling household travel associated carbon emissions, which accounts for spatial and temporal effects is needed.

It is an innovative effort to develop the conceptual framework of using Smart Data to unveil the dynamic interactions between land use (i.e., socioeconomic) changes and associated carbon emission impacts. The Smart Data-adaptive characteristic framework is capable of mining Big Data, which may be structured or unstructured, with great potential to unveil sophisticated urban phenomenon that could not be well understood by insufficient data. The results from this study offer insights of which types of land use planning policy practices are most highly associated with higher amount of VMT and associated carbon emissions, there are also potentials to reveal policy impacts that can be applied to integrated land use and transportation sustainability practices.

The results of this research are expected to add to the existing body of knowledge to foster faster and easier approaches of examining the how adaptive planning strategies could alleviate the effects of climate change from household travel carbon emissions. A spatial cross-sectional regression model is proposed to capture the spatial autocorrelation effects. The hypothesis, if tested to be true, will benefit the integration of the actual and scenario based land use visioning and planning, demographic changes, transportation emission analysis, and computer forecasting and evaluation of future scenarios. This research makes it possible to assess the household travel carbon footprint and provides supportive models, data for possible carbon emission mitigation through land use policies and adjustments. The quick response modeling framework using GPS survey data simulated Smart Data provides connections among land use, household socioeconomic and their travel carbon emissions. It is a practical tool for MPOs and other planning agencies' to quickly compare alternative planning scenarios with spatial details.

1.2 Research Motivation and Significance

This research is motivated to address the following research problems in the current understandings of land use (i.e., socioeconomic) changes and its dynamic interactions with household travel carbon emissions:

1) The conventional carbon emission-modeling framework focuses on the link-based emissions and then aggregate to regional quantities. This approach is incapable of tracing emissions back to its geographical origin and providing information on areas where adaptive planning policies are needed. However, planners need this information to assess the possible impact of changes in land use, socioeconomic of household, and their geographical relations to be adaptive to sustainability requirements, including their potential contribution to long-term climate changes. From the transportation planning point of view, it is critical to have the analytical capability of estimating the geographical distributions of the travel emission contributions at the Traffic Analysis Zone (TAZ) level.

2) Recent studies indicate potential deficiencies in converting travel demand outputs into the inputs for emission models. Emission models often rely on traditional travel demand models for vehicle activity inputs. However, traditional travel demand models are mostly calibrated and validated using aggregated daily traffic data. Although there are methods(Yao, et al., 2013; Wei, et al., 2009) to collect vehicle operating data, but they are mostly limited at either point or link sources and incapable of collectively capture a region's emission. No data are available to validate the models at the hourly level for emission estimates, because hourly VMT and speed variations are under-represented. In other words, it has actually resulted in daily aggregated inputs to be used in the emission models, which would inaccurately reflect the hourly emission features, if the presumed hourly traffic adjustment factors are used. Another drawback in the vehicle activity inputs is that when detailed vehicle trajectory data is not available, a default driving cycle will be applied. Since the default driving cycle is incapable to represent the local true driving operations, an inaccurate emission estimate would be inevitably resulted in the modeled traffic.

3) The recent advancement of mobile device sensors (e.g., GPS and accelerometer in smart phone) and data transmitting technologies (e.g., internet of things) could provide travel trajectories to be collected from the users of smart phones or other GPS devices. The availability of such data sources will actually provide new opportunities of enhancing our understanding and modeling of the dynamics between land use pattern, travel behavior, and the associated environmental impacts. This data is referred to as Smart Data, which is often massive on the real-time base. However, lots of research efforts are still needed to clarify how such data could be used to reveal the quantitative connection between the

"sustainability" and "development", and the way to extract useful information from the smart data source that decision makers have been craving for. Another challenge is how to make the data adaptive across multiple modeling platforms. This requires the support of rapid integration of either unstructured or semi-structured data, enabling quick response analytics and derives composite value from the data. Therefore, a Smart Data-driven and Smart Data-adaptive modeling framework is imperative to attempt to interconnect urban development and environmental impact.

4) The historical increase of CO_2 emissions from transportation sector is largely contributed to increased and imbalanced demand for land use and travel activities. Climate change, land use and socioeconomic developments are the principal variables that define the sustainable development. It is of Federal (e.g. U.S. EPA) and state governments' (e.g. California Air Resources Board) great interests to investigate research questions such as are the changes tangible? What are the actionable sciences for decision making? What adaptation changes can be made in the planning horizon? It is expected to have an integrated modeling framework such that a what-if scenario analysis is facilitated to answer those questions.

1.3 Goal and Objectives

The research question is whether the Smart-Data can unveil the sophisticated dynamics between land use changes and its associated carbon emission impacts, given that a smart data driven and adaptive modeling framework for this attempt is well developed. The following requirements for successfully investigating the research question have been met: (1) sample Smart Data (since the structure of the GPS-based Household Travel Survey (HTS) data is very close to the defined smart data, 2009-2010 Cincinnati HTS data is used as a substitute data source to simulate the

Smart Data), (2) data mining and/or relevant statistical analytic skills and theoretical background to quantify the spatiotemporal relationship between land use, household travel, and emission, and (3) computing programming skills to manipulate data and build analysis models in the multiple modeling environment such as Geographical Information System.

1) To achieve the goal of exploring the capability of unveiling sophisticated dynamics between land use changes and associated carbon emission impacts through testing the proposed research question, the following objectives are accordingly designed to fulfill:
2) to develop a generalized method to retrieve travel characteristics from the simulated Smart Data source by using the GPS-based travel trajectory data obtained from the household travel survey (HTS);
3) to use the resulting data for developing a generalized quantitative model structure to facilitate connecting the attributes of land use and household socioeconomic, and align household travel characteristics with the associated carbon emissions;
4) to provide a proof-of-concept study by using the Cincinnati HTS data, and demonstrate the capability of the proposed framework with respect to unveiling the connections between land use (i.e., socioeconomic) and carbon emissions through scenario-based analysis. The result of the proof-of-concept study will be used to conclude the proposed research question.

It is an innovative effort to develop the conceptual framework of using Smart Data to unveil the dynamic interactions between land use changes and associated carbon emission impacts. The Smart Data-adaptive characteristic framework is capable of mining Big Data, which may be structured or unstructured, with great potential to unveil sophisticated urban phenomenon that cannot be well interpreted with insufficient data. The results from this study offer insights on what types of land use planning policy practices are highly associated with larger VMT and carbon emissions. There are also potentials to help better understand the policy impact that can be possibly applied to integrated land use (i.e., socioeconomic) and transportation sustainability practices.

The results of this research are expected to add to the existing body of knowledge to foster faster and easier approaches of examining the how adaptive planning strategies could alleviate the effects of climate change from household travel carbon emissions. A spatial cross-sectional regression model is developed to capture the spatial autocorrelation effects. The research question, if tested to be true, will benefit the integration of the actual and scenario-based land use visioning and planning, demographic changes, transportation emission analysis, and computer forecasting and evaluation of future scenarios. This research makes it possible to assess the household travel carbon footprint and provides supportive models, and data sets for possible carbon emission mitigation through land use policies and adaptation. The quick response modeling framework using GPS survey data simulated Smart Data provides connections among land use, household socioeconomic and their travel carbon emissions. It is a practical tool for Metropolitan Planning Organizations (MPO) and other planning agencies to quickly compare alternative planning scenarios with spatial details.

1.4 Organization of the Dissertation

The dissertation is organized in the following sequence: Chapter 2 presents a review of literature on current practices of integrated land use, transportation and emission analysis, smart data characters and sources, household travel carbon emission related research and theoretical foundation of the spatial models. Chapter 3 presents the research question and detailed research methodology. Chapter 4 presents the results from Cincinnati GPS household travel survey data simulated smart data application to modeling the household cross-classification and zonal level carbon emissions. A stepwise variable selection procedure is used to identify the critical variables. The spatial autocorrelation of the variables is also examined to justify the need for including spatial information in the modeling process. Chapter 5 presents the investigation of different spatial model formulation and their performance evaluations. A set of goodness of fit measures is used to determine the best fit model. Chapter 6 presents the results from a scenario testing, sensitivity tests and calculated elasticity for regression variables. Chapter 7 wraps the dissertation by drawing conclusions of the entire research.

CHAPTER 2 LITERATURE REVIEW

2.1 The Call for Carbon Emission Reduction

The last decade has witnessed increased international, scientific and cooperative efforts (e.g., the Kyoto Protocol) in the world to address global climate change and reduce GHG emissions among countries since the late 1980s. Many cities and regions have introduced a wide range of regulations and incentives (Kamal-Chaoui & Robert, 2009) to address the challenge of climate change and carbon emission reduction. Countries across the globe committed to creating a new international climate agreement by the conclusion of the United Nation Framework Convention on Climate Change (UNFCCC) Conference of the Parties (COP21) in Paris in December 2015. The member countries have agreed to publicly outline a post-2020 climate actions they intend to take under a new international agreement, known as their Intended Nationally Determined Contributions (INDCs). The INDCs serves a clear path towards a low-carbon, climate-resilient future.

There are two major laws in California to reduce Greenhouse Gas (GHG) emissions on top of the Governor's executive order of achieving 40% GHG reductions of the 1990 levels by 2030. The first bill (Assembly Bill (AB) 32: Global Warming Solutions Act), signed by the Governor Arnold Schwarzenegger, in September 2006. AB32 is intended to reduce GHG emissions to 1990 levels by the year 2020 and to 80 percent below 1990 levels by 2050. Senate Bill (SB) 375 was passed by the state legislature and signed by Governor Schwarzenegger in September 2008.

SB 375 (California Air Resources Board, 2016) requires MPOs in California to develop a Sustainable Communities Strategy (SCS) as a major element of the Regional Transportation Plan (RTP) to reduce GHG emissions. SB375 acknowledges that the transportation sector contributes to the generation of GHG emissions and sets GHG reduction targets for each MPOs in California. It recommends that MPOs develop a set of SCSs to reduce GHG emissions from cars and light trucks through the integration of planning processes for multi-modal transportation, compacted land use, and high density housing. SB 375 offers local governments, regulatory relief and other incentives to encourage alternative land use development patterns and transportation alternatives.

A SCS, generally featured by diverse land uses, higher residential densities and building intensities, is a land use element in the Regional Transportation Plan (RTP). The land use plan element and its relevant sustainable transportation and land use strategies in the SCS would incentivize sustainable development, such as transit oriented development (TOD), mixed use development, provision of housing opportunities near job centers, job housing balance, the prioritized growth along transit corridors and hubs to utilize available capacity. As a result, transit use or walking becomes more popular, more vehicular VMT is shifted to transit and biking, and the planned reductions of GHG emissions can be achieved. However, planners still interested in where to locate the limited resources that would result in a higher amount of GHG reductions. This calls for a location traceable method to quantify the household travel carbon emissions.

2.2 Integrated Land Use, Transportation and Emissions Analysis

The integration of land use (i.e., socioeconomic), transportation and emissions has been studied for decades. It has been well accepted that transportation plays a role in land development, even in the mature and extensive networks of major urban regions within developed countries (Sadek, et al., 2011; Transportation Research Board, 2009).Although the impact of land use patterns on travel demand reasonably well understood (through trip generation and attraction), long-term changes in land use patterns and the effects of transport policy and system investment on land development and use remains elusive(Zhou, et al., 2010).

Changes in urban land form can lead to alterations on traffic pattern and transportation efficiency, population density and distribution(Redfearn, 2007; Roth, et al., 2011; Neuman & Smith, 2010). A variety of land use, employment and household allocation models that integrate with travel demand models have been developed for dedicated modeling purposes and methodologies. Several popular models include: PECAS (Production, Exchange and Consumption Allocation System) (Lee-Gosselin & Doherty, 2005; Hunt & Abraham, 2003), Kockelman et al.'s RUBMRIO (Random-Utility-Based Multi-Regional Input-Output) (Zhao & Kockelman, 2004), Gregor's LUSDR (Land Use Scenario DevelopeR) (Gregor, 2007), and Waddell's UrbanSim (Urban Simulation) (Waddell, et al., 2003). NCHRP Report 423A (Parsons Brinckerhoff Quade and Douglas, 1999) described eight leading land use models and focused on assessing the impacts of transportation projects. Duthie et al (Duthie, et al., 2007) compared gravity-based Integrated Transportation, Land Use Planning (ITLUP) type models with the UrbanSim models and tested the two systems in Austin, Texas. A conclusion drawn from the study is that UrbanSim is recommended for MPOs (Metropolitan Planning Organizations) for the integration of land use (i.e., socioeconomic) and transportation. However, the drawback is that sufficient resources and time need to be devoted to data collection and manipulation.

Previous studies (Wei, et al., 2012; Wei, et al., 2011; Yao, et al., 2014) at the University of Cincinnati have examined the effects of certain land use and transport policies on travel related emissions. Wei et al. (Wei, et al., 2012) developed an integrated framework to investigate advanced traffic management, operation, control technologies and systems' potential to reduce highway congestion and environmental impacts associated with vehicle travel. The framework is implemented through the development of a Scenario Based – Planning Support System (SB-PSS). Preliminary results show that the SB-PSS is a flexible methodology for the assessment of various regional policies' impacts on land use, transportation and emissions. The Hamilton County case study provides remarkable insights regarding potential development patterns, changes in land use and traffic, population and employment growth. The scenario evaluation system was built upon the guidelines of the EPA regarding sustainabile transportation measures is another implication of carbon emission quantification.

The results from this work offer specific recommendations as to which types of land use planning policy practices are most highly associated with higher amount of VMT, carbon emissions, sustainability scores. Further investigation into future improvement on the quantification of the sustainability with taking account into transportation infrastructure and its related adaptation could help to understand how changes in land use policy and travel behavior might corroborate to shape the overall sustainability under a specific designed scenario.

2.3 Household Travel Survey

Household travel surveys are designed to assist transportation planners and policy makers who need comprehensive data on travel and transportation patterns at national level or statewide. Basic information it gathers includes: purpose of the trip (work, shopping, social, etc.), means of transportation (car, walk, bus, subway, etc.), travel time of the trip, time of day/day of week, etc. The surveys are designed to fulfill the purpose of quantifying travel behavior, analyzing changes in travel characteristics, relating demographics to the travel behavior, etc. (Federal Highway Administration, 2009).

Lots of studies also used the Household Travel Survey data and attempted to connect land use, household demographics, and travel behaviors (Cervero & Landis, 1995; Newman &

Kenworthy, 1988; Bhat, et al., 2009; Bhat & Sen, 2006). A most common approach is to statistically link household demographics and socioeconomic characteristics with the number of trips associated with. Using the 2001 NHTS (National Household Travel Survey) data, Liu et al (Liu & Shen, 2011) investigated how urban land use characteristics influence on the household travel and the energy consumption associated with that. Their results demonstrate that accessibility explains more than the use of 3D (Density, Design, and Diversity measures) approach (Cervero & Kockelman, 1997). They also reported that there is a strong correlation between household characteristics, vehicle ownership, Vehicle Miles Traveled (VMT) and energy consumption. Lindsey et al (Lindsey, et al., 2011) investigated the relationship between household location on household patterns of vehicle miles traveled, and by extension, energy consumption and carbon emissions. They reported that VTM, energy use and carbon emissions increase as the residential distance from city center. Various scenarios show that with increases in privately vehicle fuel efficiency, the overall reduction in fuel use creates a more uniform spatial profile of energy/greenhouse gas emissions across the region.

Giaimo et al (Giaimo, et al., 2010) introduced the preliminary findings of the first largest GPS-based household travel survey – Greater Cincinnati Household Travel Survey (HTS). It is a proof of concept study for replacing travel diaries with a large-scale multi-day Global Positioning System (GPS) survey. They conclude that a representative sample of households can be recruited for a GPS-based survey, based on a comparison of pilot sample household characteristics with available Public Use Microdata Samples data, and response rates for difficult-to-reach households, such as cell phone-only, lower income, and zero-vehicle households can be improved with a cash incentive ($25). The preliminary results have proven that using the GPS-base travel survey is a viable approach for household travel survey, which is much reliable and informative comparing to traditional survey methods such as a computer-assisted telephone interview (CATI). As the project report (Stopher, et al., 2012) indicates, the survey identified 2,059 GPS complete and 549 GPS incomplete household travel surveys. The database includes 3,853 drivers, 60,900 trips with a daily motor vehicle trip rate of 7.60. The report also concludes that it is feasible to undertake a GPS-only household travel survey, achieving a high standard of representativeness of the population, while imputing trip mode and purpose at a sufficiently accurate level. The high level of accuracy attained in this survey for imputing mode and purpose with 96 percent on mode and around 90 percent on activity is far superior to other forms of surveys such as the self-report survey. The richness of the "ground-truth" of time, location (latitude and longitude), distance, speed, and routes data collected from this survey provided state-of-the-practice data to support research and studies such as development of activity-based travel demand model and even emission analysis.

2.4 Smart Data

The recent advancement of mobile device sensors (e.g., GPS and accelerometer in smart phone) and data transmitting technologies (e.g., Internet of things) have provided new opportunities for enhancing our understanding and modeling of interactions between land use patterns, the associated travel behavior, and environmental sustainability. It is anticipated that the data volume, velocity, veracity and value will lead to drastic changes in transportation and even our physical world. The collected Big Data with veracity and value that contains critical information is called Smart Data. A good example of Smart Data in transportation field includes location, velocity, and trajectory from GPS-based mobile devices such as smart phone, handheld computers, and wearable technologies(Corporate Partnership Board, 2015). Cell phone data, so far, is the most common Smart Data and with lots of applications in travel mode detection(Reddy, et al., 2010; Shin, et al., 2015; Liu, et al., 2008) and understanding travel pattern(Widhalm, et al., 2015). These data are often massive and real-time. Such datasets could be exploited and amalgamated to provide policy-relevant insights and operational improvements for informed data-driven decision making in sustainable land use and transportation development.

Smart Data, is not about data per se, but rather refers to the ways to analyze and make sense of it. From the transportation modeling perspective, Big Data is what we know about traveler behavior, while Smart Data is how we discover the underlying rationale and predict repetition of such behaviors. Smart Data can significantly reduce the most of the strong data dependent, and sophisticated transportation modeling works, and decrease time to insight and time to value for Big Data applications. Another challenge is how to choose desired itemized datasets from the Smart Data source so that the global reconstruction error could be minimized, or simply say, how to make the data adaptive across multiple data platform. This requires the support of rapid integration of either unstructured or semi-structured data (as most Big Data is), enabling quick response analytics and derive composite value from the data. Therefore, a Smart Data-driven and Smart Data-adaptive framework for connecting environmental impact and urban development is imperative and will be beneficial to finding the keys of trade-off between environmental sustainability and urban development demand.

A good example of the smart data is cell phone data, available from commercial data vendors (e.g., AirSage) at a high cost, or not yet open to public (e.g., ODOT Smartphone Survey). The Cincinnati GPS HTS datasets provided by OKI actually provided a great opportunity to make the research idea to be implemented. Smart data will become overwhelmingly available in the near future and enable not only household travel emissions analysis but also opportunities of population analytics. As more smart data emerges, the proposed modeling framework will be further reviewed, and improved. The following table (Table 2.1) compares the features of the Cincinnati GPS HTS data, ODOT smart phone survey data, and AirSage (a commercial cell phone data vendor) data.

Table 2.1 Comparison of Cincinnati GPS Survey Data and Smartphone Data

	Cincinnati GPS Survey(Stopher & Wargelin, 2012)	ODOT Smartphone Survey(Anderson & Giaimo, 2016)	AirSage (Helwagen, 2015)
Time Resolution	Second	Second	Second, minutes, hour, day, month
Trajectory	x,y (high accuracy GPS)	x,y, (Assisted GPS)	x, y (Cellphone Tower)
Origin	x,y (high accuracy GPS)	x,y, (Assisted GPS)	x, y (Cellphone Tower)
Destination	x,y (high accuracy GPS)	x,y, (Assisted GPS)	x, y (Cellphone Tower)
Sample size	5,502 Households	937 Households in Pilot Study	Covers, Verizon and Sprint cell phone users, over 100 million mobile devices
Sampling Duration	3 days per household in 2009/2010	7 days per household in 2014/2015	Continuous/Billions of records
Advantage	High accuracy trajectory	All trips validated	Largest sample available
Disadvantage	Small sample size, needs validation	Sample plan relevantly slow	Less accurate in trajectory

2.5 Spatial Regression Model and Transportation Applications

Spatial models typically refer to data containing time series observations over a type of spatial units such as TAZs, zip codes, regions, countries, and states. It is generally recognized that panel data are more informative since they contain more variations and less co-linearity among the variables. The use of panel data results in a greater availability of degrees of freedom, and hence increases efficiency in the estimation (Elhorst, 2010).

Hall et al (Kockelman & Frazier, 2006)identified that current land use, land cover (LULC) models fail to incorporate and integrate spatial and temporal correlations lies in the urban systems. To fill in the gap, they introduced the spatial linear and logistic regression model for panel data. They used downtown Austin's population data over years to predict the year 2020 population. A conclusion has drawn that spatial and temporal effects were shown to be highly statistically significant, suggesting that their recognition and formal inclusion in the models is likely to be of great value. Parent and LeSage (Parent & LeSage, 2010) applied spatial panel model with random effects to predict commuting times. They collected travel time to work, travel expenditures, traffic volume, lane miles and gas taxes to forecast the mean travel time to work for each state. The results find evidence of substantial of spatial spillovers and relatively weaker time dependence leading to much smaller time impacts accruing over future periods.

A very recent article by Chakir and Le Gallo (Chakir & Le Gallo, 2013) investigates how the introduction of spatial effects and individual heterogeneity in an aggregated land-use share model affects the predictive accuracy of land use models. They considered agricultural, forest, urban and other land uses in their investigation. And one of the conclusions drawn is that controlling for both unobserved individual heterogeneity and spatial autocorrelation outperforms any other specification in which spatial autocorrelation and/or individual heterogeneity are ignored.

Perugu et al (Perugu, et al., 2012) applied spatial panel model for modeling truck factors and for improved $PM_{2.5}$ estimations in a regional roadway network. The proposed methodology enables plotting the spatiotemporal distribution of $PM_{2.5}$ emissions in a subarea. They also reported that the methodology presented is scalable and transferable and holds technical promise in its application to different regions and different pollutants.

There are also studies that analysis the spatial transferability of a micro-behavioral model from the residential mobility component of the integrated land use, transportation, and environment (ILUTE) modeling system (Rahman & Habib, 2015) and explore the adaptive transportation supply and demand management system for integrated land use and transportation model(Ma & Lo, 2015). Handy et al has also investigated on the correlation of the built environment and travel behavior(Handy, et al., 2005) and the results suggest there is a causal relationship between the two.

The Spatial econometric deals with interaction effects among geographical units. The most simplistic general linear regression model without spatial interaction effects is:

$$y_{it} = X_{it}\beta + \mu + \varepsilon_{it} \tag{2.1}$$

where:

i is an index for the cross-sectional dimension (spatial units), with $i = 1, ..., N$,

t is an index for the time dimension (time periods), with $t = 1, ..., T$,

y_{it} is an observation on the dependent variable at i and t,

$\boldsymbol{X}_{it}$ an 1-by-K row vector of observations on the independent variables,

β is matching K-by-1 vector of fixed but unknown parameters.

Σ_{it} is an independently and identically distributed error term for i and t with zero mean and variance σ_2,

μ_i denotes a spatial specific effect.

When specifying interaction between spatial units, the model may contain a spatially lagged dependent variable or a spatial autoregressive process in the error term, known as the spatial lag and the spatial error model, respectively. The spatial lag model posits that the dependent variable depends on the dependent variable observed in neighboring units and on a set of observed local characteristics:

$$y_{it} = \delta\sum_{j=1}^{N} W_{ij}y_{it} + X_{it}\beta + \mu + \varepsilon_{it} \tag{2.2}$$

where:

δ is called the spatial autoregressive coefficient and,

W_{ij} is an element of a spatial weights matrix $\boldsymbol{W}$ describing the spatial arrangement of the units in the sample.

The spatial error model, on the other hand, posits that the dependent variable depends on a set of observed local characteristics and that the error terms are correlated across space:

$$y_{it} = X_{it}\beta + \mu + \phi_{it} \tag{2.3}$$

$$\phi_{it} = \rho \sum_{j=1}^{N} W_{ij}\phi_{it} + \varepsilon_{it} \tag{2.4}$$

where:

φ_{it} reflects the spatially autocorrelated error term and,

ρ is called the spatial autocorrelation coefficient.

Critical Inputs for Carbon Emission Inventory Modeling

Current carbon emission inventory modeling approach, utilizing an ad-hoc data structure, takes in data for land use (i.e., socioeconomic) forecasting, travel demand forecasting and MOVES emission estimation in turn. A list of critical inputs for carbon emission inventory modeling is shown in Table 2.2. Each data input is also included its level (High, Medium and Low) of potential of impact on emission inventory as well as its level of uncertainty in the current methodology.

Table 2.2 Critical Inputs for Carbon Emission Inventory Modeling

Category	Input	Impact on Carbon Emission Inventory Modeling	Uncertainty in Present Methodologies
Land Use (Socioeconomic)	Household per TAZ	High	Medium/High
	Employment per TAZ	High	Medium/High
	Low Trip Rate Employment Fraction per TAZ	Low	Medium

Category	Input	Impact on Carbon Emission Inventory Modeling	Uncertainty in Present Methodologies
	Medium Trip Rate Employment Fraction per TAZ	Low	Medium
	High Trip Rate Employment Fraction per TAZ	Low	Medium
	High School Enrollment per TAZ	Low	Medium
	University Enrollment per TAZ	Low	Medium
Travel Demand Forecasting	Trip Generation Rates	High	Medium
	Trip Attraction Rates	High	Medium
	Roadway Network Free-flow Speed	Low	Low
	Roadway Network Capacity	Low	Low
	Mode Choice Utility Function Parameters	High	Medium
MOVES Emission Model	Vehicle Miles Traveled	High	Medium/High
	Vehicle Hours Traveled	Medium/High	Medium/High
	VMT Fraction by Vehicle Type	High	Medium/High
	VMT Fraction by Time of Day	Medium/High	Medium/High
	VMT Fraction by Road Type	Medium/High	Medium/High
	VHT-based Speed Distribution	High	Medium/High
	Age Distribution	High	Low
	Fuel Type and Formulation	Medium	Low
	Average Speed	High	Low
	Driving Cycles	High	Medium
	Operating Mode Distribution	High	High
	Road Grade	Low	High

The most straightforward modeling approach in empirical work is to start with a non-spatial linear regression. Since the linear regression model is commonly estimated by ordinary least squares (OLS), it is often labelled the OLS model. The non-spatial linear regression model takes the form

$$Y = \alpha l_N + X\beta + \varepsilon \tag{2.5}$$

where Y denotes an N×1 vector consisting of one observation on the dependent variable for every unit in the sample (i=1, . . ., N), l_N is an N×1 vector of ones associated with the constant term parameter a, X denotes an N×K matrix of exogenous explanatory variables, with the associated parameters β contained in a K×1 vector, and ε= (ε_1, …, ε_N)^T is a vector of disturbance terms, where ε_i are independently and identically distributed error terms for all I with zero mean and variance σ^2.

Another approach is to begin with a more general model contains a nested special case with a series of simpler models that ideally should represent all the alternative economic hypotheses requiring consideration. Manski points out that three different types of interaction effects may explain why an observation associated with a specific location may be dependent on observations at other locations: (i) endogenous interaction effects, where the decision of a spatial unit (or its economic decision makers) to behave in some way depends on the decision taken by other spatial units; (ii) exogenous interaction effects, where the decision of a spatial unit to behave in some way depends on independent explanatory variables of the decision taken by other spatial units if the number of independent explanatory variables in a linear regression model is K, then the number of exogenous interaction effects is also K, provided that the intercept is considered as a separate variable; and (iii) correlated effects, where similar unobserved environ-mental characteristics result in similar behavior(Manski, 1993).

The Manski's model takes its form below:

$$y = \rho wy + al_N + X\beta + wX\theta + u \tag{2.6}$$

$$u = \lambda W_u + \varepsilon \tag{2.7}$$

or:

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \rho \times \begin{bmatrix} 0 & w_{12} & \dots & w_{1n} \\ w_{21} & 0 & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & 0 \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} + \alpha \times \begin{bmatrix} l_1 \\ l_2 \\ \dots \\ l_3 \end{bmatrix} + \begin{bmatrix} X_{11} & 0 & \dots & 0 \\ 0 & X_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & X_{nk} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} 0 & w_{12} & \dots & w_{1n} \\ w_{21} & 0 & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & 0 \end{bmatrix} \times \begin{bmatrix} X_{11} & 0 & \dots & 0 \\ 0 & X_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & X_{nk} \end{bmatrix} \times \begin{bmatrix} \theta_1 \\ \theta_2 \\ \dots \\ \theta_k \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{bmatrix} \tag{2.8}$$

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{bmatrix} = \lambda \begin{bmatrix} 0 & w_{12} & \dots & w_{1n} \\ w_{21} & 0 & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix} \tag{2.9}$$

where y denotes an N×1 vector consisting of one observation on the dependent variable for every unit in the sample (i=1, . . ., N),

ρ is called the spatial autoregressive coefficient

l_N is an N×1 vector of ones associated with the constant term parameter α,

w is a nonnegative N×N matrix of known constants describing the arrangement of the units in the sample,

wy denotes the endogenous interaction effects among the dependent variables,

wX is the exogenous interaction effects among the independent variables,

Wu is the interaction effects among the disturbance terms of the different spatial units.

λ is the spatial autocorrelation coefficient, while

θ represents a K×1 vector of fixed but unknown parameters, where K is number of independent explanatory variables in a linear regression model.

Figure 2.1 shows the variations of spatial cross-sectional models with respect to assumptions in the error distribution in the above parameters. Since no predeterminations on the error term distribution can be made, this study tested all the below spatial cross-section models and best model fits the data will be selected.

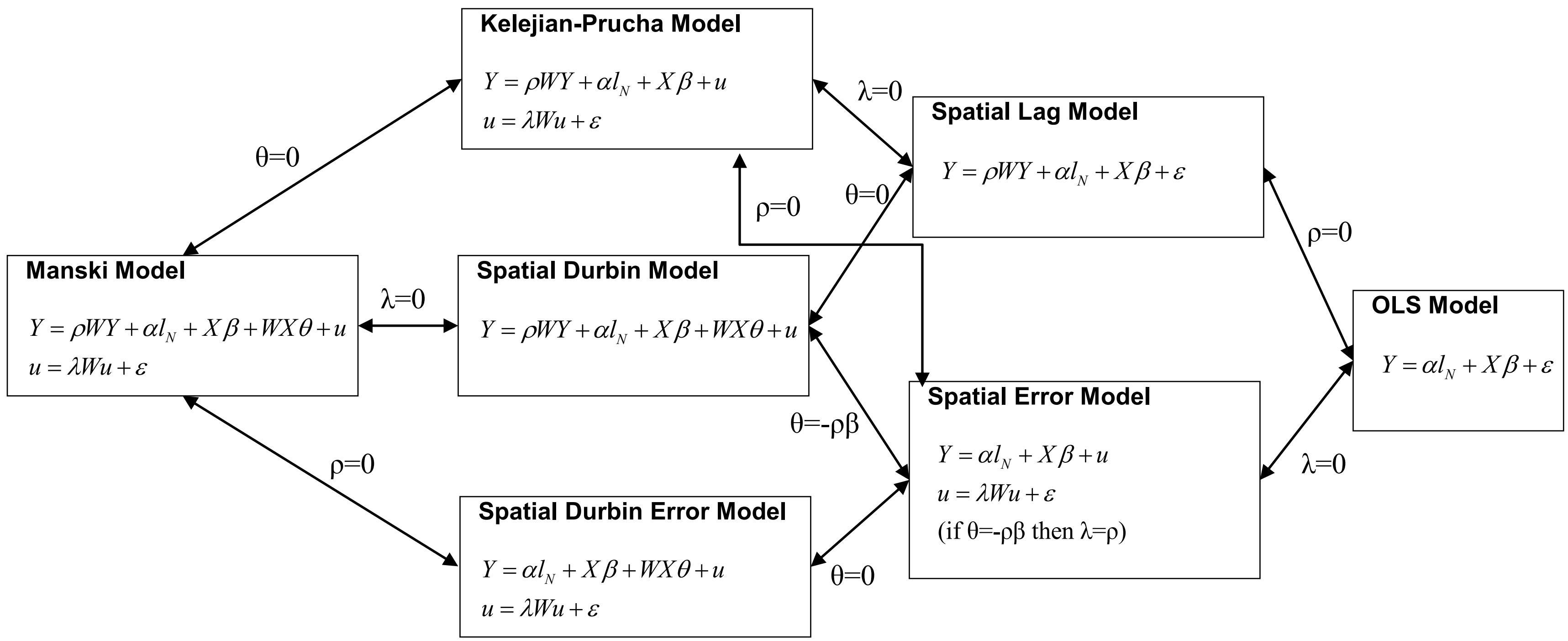


Figure 2.1 The Relationships between Different Spatial Dependence Models for Cross-section Data

2.6 Carbon Footprint Assessment Tools

There are several on-line carbon footprint assessment tools. Green Footstep (Rocky Mountain Institute, 2013) focuses on building carbon footprint. It uses predetermined CO_2 standards to estimate building construction, maintenance and operation, and site carbon emissions. The Empower (Carbon Trust Empower, 2013) developed by Carbon Trust is designed to engaging employees in cutting energy use, paper waste and travel. A more comprehensive online tool is the Carbon Footprint Calculator (Carbon Footprint Ltd, 2013). It covers the housing, flights, car, motorbike, bus and rail, and secondary carbon emissions. Its calculations for primary emissions are based on conversion factors sourced from U.S. EPA. The advantages of those online carbon footprint assessment tools are: convenient, user friendly etc. One common drawback of those online tools is that they have to use predetermined values of standards for the estimation.

Resource Systems Group (RSG) developed the smart growth area planning (SmartGAP) tool (Smith, 2013)to evaluate smart growth policies on travel demand. One of the objectives is to investigate the dynamics and inter - relationships of smart growth strategies with the performance of a transportation investment. The tool enables the comparison of performance of assumed scenarios to estimate fuel consumption, employment, vehicle miles traveled, vehicle hours of delay, transit trips, transit operating costs, and accidents by severity. Parsons and Brinckerhoff developed Carbon Footprint and Impact Evaluation Toolset which is a package of computing tools developed by PB to help clients make the connection between land use choices and carbon emissions. It includes modules to calculate emissions from building operations, traffic generated by land use patterns, and building construction. The tool is designed to analyze alternative land use/transportation scenarios, estimate amounts of greenhouse gas emissions and energy consumption, and provide real-time analysis and data visualization.

CHAPTER 3 METHODOLOGY

3.1 Smart Data Collection and Analysis

Figure 3.1 shows a Smart Data collection and analysis lifecycle. Data acquisition and recording is made possible because people increasingly leave a digital trace wherever they go (both voluntarily and involuntarily). The technology utilized in each phone call, text message, email, social media post, online search, and credit card purchase and many other electronic transaction reports on the user's location at a given point in time. The next step is to extract, clean, annotate and store the data. A range of techniques and tools (e.g., data fusion, data mining, optimization) have been developed, or adapted, to aggregate, manipulate and visualize those data. Within the context of transportation planning, spatial analytics typically extract the topological, geometric, or geographic properties encoded in a data set. New insights can emerge from the analysis of multiple data sources after the integration, aggregation and representation process. The analysis, modeling and visualization step uses knowledge-discovery approaches, including data mining (and the contribution of data mining to machine learning, network analysis and pattern recognition) and visualization techniques (e.g., geographical information system). Developing and testing models help test hypotheses regarding the impact and importance of different variables in real-world systems. By simplifying simulating real-world phenomena using Smart Data, models therefore, help to characterize, understand, quantify and visualize relationships that are difficult to correlate in complex systems such as the built environment.

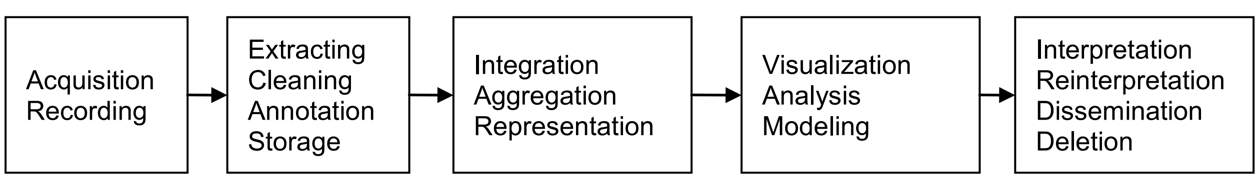


Figure 3.1 Framework for Smart Data Collection and Analysis.

3.2 The Research Question

First, the research explores the dynamics of land use, transportation, and emissions. The smart data modeling framework is developed with a non-spatial linear regression modeling approach and then to test whether the model needs to be extended with spatial interaction effects or not. This approach has been known as the specific-to general approach. The TAZ level basic model deals with determining variables that are of significance to the dependent variable (i.e.TAZ emissions). However, if the observations are spatially clustered in a certain degree, the estimates obtained from the correlation coefficient or Ordinary Linear Square (OLS) estimator will be biased and overly precise. The bias comes from the areas where there are higher concentrations of observations will have a greater impact on the model estimation, and will overestimate precision since the observations tend to be concentrated. Therefore, there are actually fewer number of independent observations than that being assumed.

Second, the existence of spatial effects among the variables is recognized from the location-rich information over time of the smart data source. To verify the applicability of the GPS simulated smart data and the utilization of spatial models, it is required to measure the spatial autocorrelation of the modeling variables using Moran's I (an indicator for spatial autocorrelation). In addition, the residuals from the OLS regression is also tested for spatial autocorrelation. If the spatial autocorrelation exists, it is recommended that the spatial model be applicable. The spatial information is then added to the basic OLS model so that the spatial autocorrelation effect is captured. Lastly, the effectiveness of the model is measured by using information-based goodness of fit.

Until now, the research question is tested. If the spatial model derived fits the data well, the test of research question of utilizing the Smart-Data structured data sets to unveil sophisticated dynamics between land use changes and its associated carbon emission impacts is true. However, if the model does not fit well, the research question shall be reinvestigated and more smart data is needed for further modeling exercise.

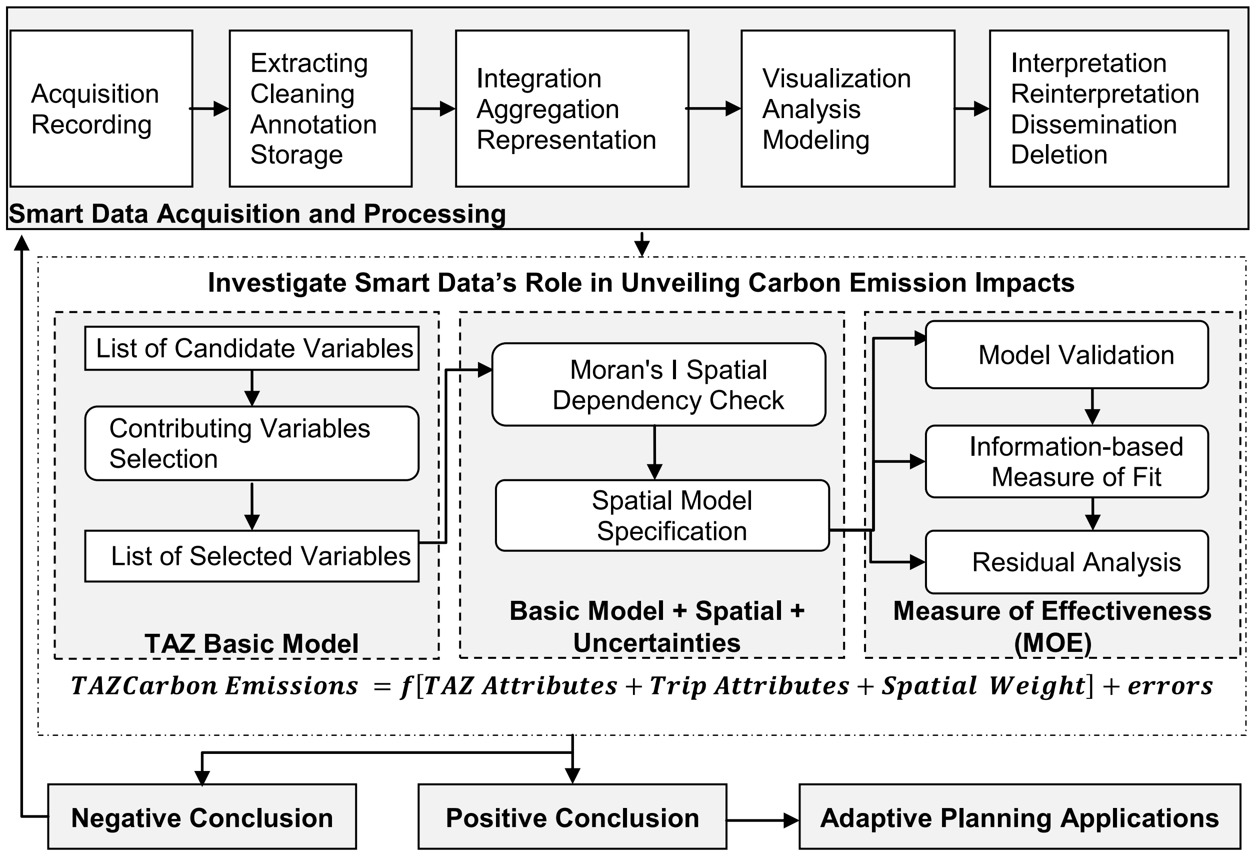


Figure 3.2 Framework for Household Travel Carbon Emission Modeling using Smart Data

The modeling framework starts with data acquisition and recording, which made possible because people increasingly leave a digital trace wherever they go (both voluntarily and involuntarily). The next step is to extract, clean, annotate and store the data. A range of techniques and tools have been developed, or adapted, to aggregate, manipulate and visualize those data. Within the context of transportation planning, spatial analytics typically extract the topological, geometric, or geographic properties encoded in a data set. New insights can emerge from the analysis of multiple data sources after the integration, aggregation and representation process. The analysis, modeling and visualization step uses knowledge-discovery approaches, including data mining and visualization techniques. Developing and testing models help answer the research question regarding the impact and importance of different variables in real-world systems. By simplifying simulating real-world phenomena using Smart Data, models therefore, help to characterize, understand, quantify and visualize relationships that are difficult to correlate in complex systems such as the built environment.

The mechanic for the smart data model framework development starts with a non-spatial linear regression model and then to test whether or not the model needs to be extended with spatial interaction effects. This approach is known as the specific-to general approach. The TAZ level basic model deals with determining variables that are of significance to the dependent variable (zonal CO_2 emissions). However, if the observations are spatially clustered in a certain degree, the estimates obtained from the correlation coefficient or OLS estimator will be biased and overly precise. The bias came from the areas with higher concentrations of events will have a greater impact on the model estimation and will overestimate precision since events tends to be concentrated, and therefore, there are actually fewer number of independent observations than that being assumed. Since the smart data provides location-rich information, it is imperative to recognize the existence of spatial effects among the variables. Spatial regression based modeling approach recognizes spatial dependency of the variables and are often more robust comparing to other modelling approach. Thus, the spatial information is added to the basic model and the spatial model is specified. The last step then is to measure the effectiveness of the model.

The prospective structure of smart data requires efficient access via large writing and reading. Transportation related (e.g. GPS data) smart data usually contains latitude and longitude of the traveler in a fixed time interval. This type of data along with supplementary data such as roadway grade, vehicle age will be sufficient to quantify the emissions. Thus, it is a very practical to use GPS-based HTS data to simulate the proposed smart data modeling framework.

A demonstrative case study testing the proposed smart data enabled modeling framework is completed. Figure 2 illustrates the heuristic framework of this GPS data simulated testing. The purpose of the methodology is to build up a linkage from household travel related carbon emissions with land use, socioeconomic, demographic, and spatial factors.

Since the smart data provides location-rich information, it is imperative to recognize the existence of spatial effects among the variables. Spatial regression based modeling approach recognizes spatial dependency of the variables and are often more robust comparing to other modelling approach. Thus, the spatial information is added to the basic model and the spatial model is specified. The last step then is to measure the effectiveness of the model. To verify the applicability of the GPS simulated smart data and the utilization of spatial models, it is required to measure the spatial autocorrelation of the modeling variables using Moran's I (an indicator for spatial autocorrelation). Then, the residuals from the OLS regression were also tested for spatial autocorrelation. If the spatial autocorrelation exists, it is recommended to use spatial model.

Thus, the research question is tested. If the spatial model derived fit the data, the conclusion of utilize the Smart-Data structured datasets to unveil sophisticated dynamics between land use changes and its associated carbon emission impacts is supported. However, if the model does not fit well, the research question is not supported and more smart data is needed for further modeling exercise.

3.3 Outline of Research Methodology

To fulfill the research gap identified in Chapter 1.2, an integrated approach is proposed by using the Greater Cincinnati Household Travel Survey Data. The purpose of the methodology is to build up a linkage from household travel related carbon emissions with land use, socioeconomic, demographic, and spatial and temporal factors. And to rapidly quantify the carbon emissions through simulation of scenario-based land use and socioeconomic changes.

The framework starts with GPS data which is the best available second-by-second vehicle trajectory and combine with the socioeconomics, demographics and spatial information. First, carbon emissions are calculated for the location specific households by using traditionally unavailable vehicle specific power (VSP) approach and the EPA MOVES model. Second, the TAZ level carbon emissions are aggregated by using the cross-classified household (by area type, number of workers, life cycle, and income level) travel emission rates and the zonal fractions of each household type. Third, the contributing variables are identified for spatial cross-sectional modeling including TAZ level attributes, trip level attributes and spatial weights. The spatial cross-sectional model is then estimated and evaluated. Finally, more policy and planning scenarios could be tested which provides justified land use patterns and associated household spatial distribution based on the adaptive planning.

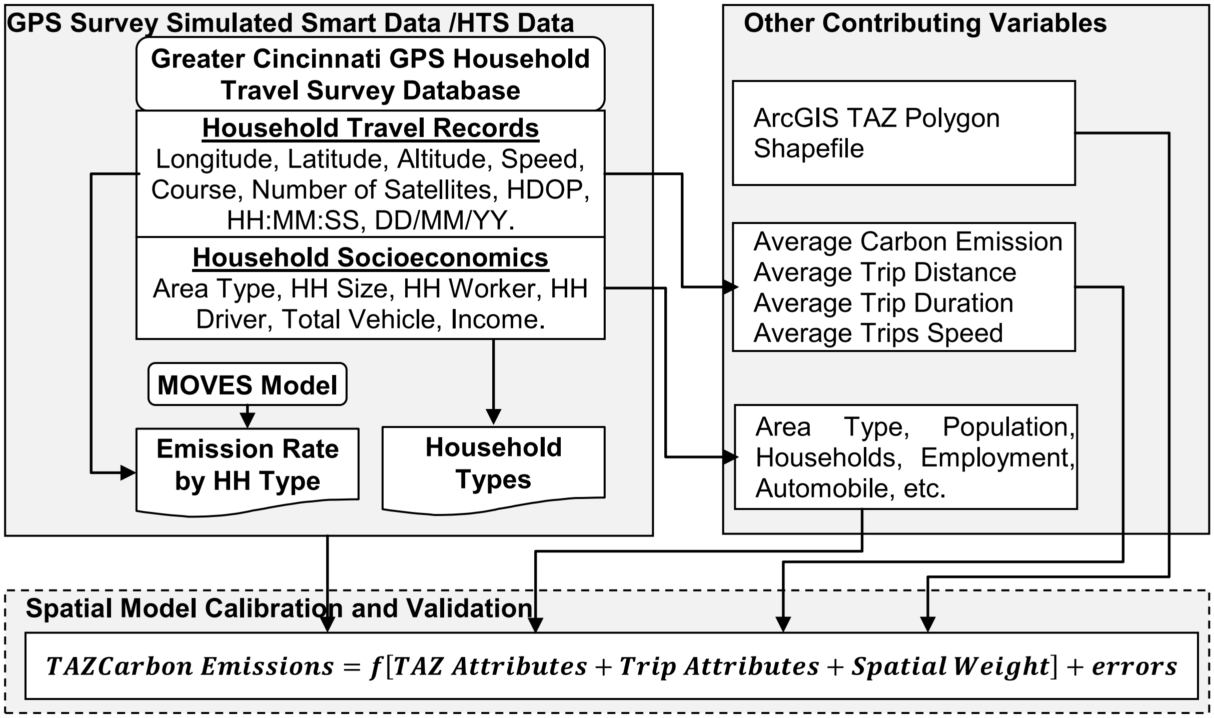


Figure 3.3 The Proposed Framework of Spatial Model-based Household Travel Carbon Emission Model.

Figure 3.3 illustrates the heuristic framework of this research. The household travel data processing module extracts household travel characteristics base on the survey database. The purposes are twofold. First, to calculate the carbon emissions from the location specific household by using traditionally unavailable vehicle specific power (VSP) approach and the EPA approved MOVES model. Second, the extracted trip features based on household socioeconomic data will be used to update the trip rates table for the customized travel demand model. The contributing variables, which produce contributing variables for spatial cross-sectional modeling, including TAZ level, trip level attributes and spatial weights. The spatial cross-sectional model will then be estimated. Thirdly, the policy and planning scenarios module will provide justified land use patterns and associated household spatial distribution based on the adaptive planning and carbon footprint assessment. The last part of this research will be using the scenario data generated in the previous step to forecast the carbon emissions of the scenario.

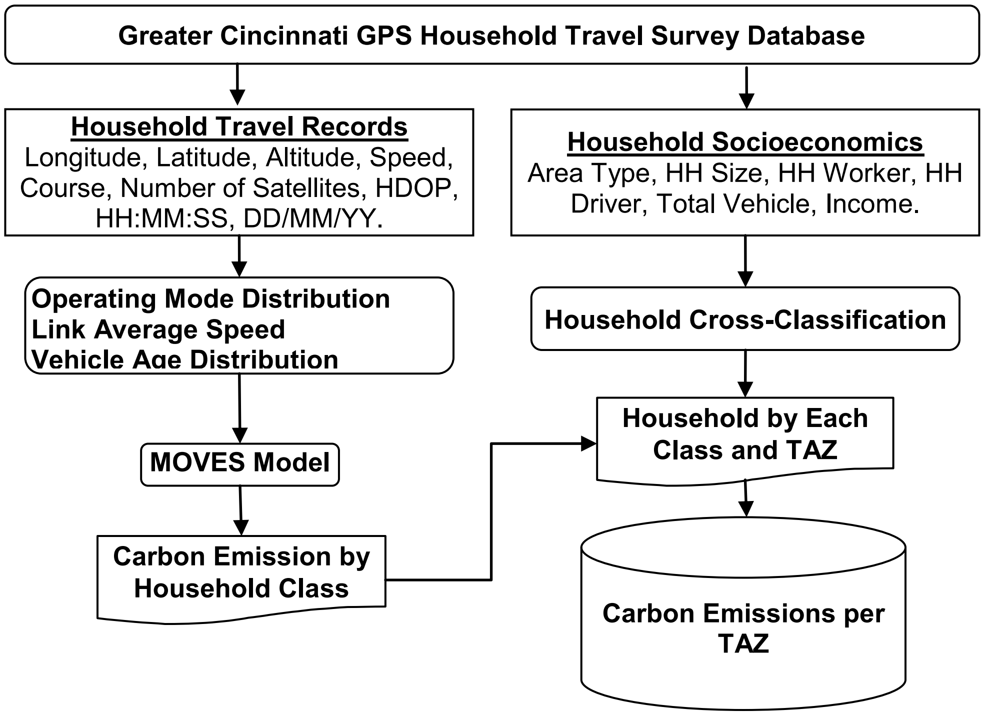


Figure 3.4 The Process of Household Travel Survey Data to Carbon Emissions per TAZ

The household travel carbon emissions are calculated using the MOVES model. Figure 3.4 illustrates the process for carbon emission calculation by household types. Two parts of the model inputs are derived from the greater Cincinnati GPS household travel survey database: household travel records, and the household socioeconomic. The household travel records are used to calculate the operating mode distribution, link average speed and the vehicle age distribution for MOVES model. The household socioeconomic are the basis for cross-classification based on household area type, workers, life cycle, and income level. Zonal aggregated carbon emissions associated with household travel are then calculated by using carbon emission by household class multiply by the number of that specific type of household in the TAZ.

The household travel carbon emissions will be extracted by using the key concept of the MOVES emission model – Vehicle Specific Power. The mathematical expression of VSP was first developed by J. L. Jiménez (Jiménez-Palacios, 1999) at the Massachusetts Institute of Technology. It is a mathematical representation of engine load against aerodynamic drag, acceleration, rolling resistance, plus the kinetic and potential energies of the vehicle, all divided by the mass of the vehicle. In practice, a generic set of coefficient values estimating VSP for a typical light duty fleet is applied as a useful basis for characterization (Frey, et al., 2006; Yao, et al., 2013). The VSP values are calculated by the following equation:

$$VSP = v \times [1.1a + 9.81 \times grade(\%) + 0.132] + 0.000302 \times v^3 \tag{3.1}$$

where:

v = vehicle speed (m/s)

a = vehicle acceleration/deceleration rate (m/s^2)

grade = vehicle vertical rise divided by the horizontal run (%)

As described above, the one of the key facts that made this study different from the previous studies is the availability of second-by-second GPS speed data. Table 3.1 shows a sample of the GPS household travel records.

By the combinations of speed and VSP representing real-world vehicle operating mode, MOVES adopted the 23 operating mode bins, plus additional operating modes for starts and evaporative emissions. Table 3.2 is the summary of the VSP bins which MOVES model implies. The VSP values are then binned into the below table and the operating mode distribution has therefore generated.

Table 3.1 Sample GPS Household Travel Records

Longitude	Latitude	Speed (km)	Course (degrees)	Number of Satellites	HDOP	Altitude (meters)	DD/MM/YY	HH:MM:SS	Distance (meters)
-84.4966	39.1716	85	266	9	0.93	167	31/8/2009	11:37:04	23
-84.4969	39.1716	85	266	9	0.93	167	31/8/2009	11:37:05	24
-84.4971	39.1716	85	266	8	1.2	168	31/8/2009	11:37:06	23
-84.4974	39.1716	83	264	8	1.2	168	31/8/2009	11:37:07	23
-84.4977	39.1715	83	262	9	0.93	169	31/8/2009	11:37:08	23
-84.4977	39.1715	81	260	9	0.93	170	31/8/2009	11:37:09	0
-84.4982	39.1715	81	258	8	1.1	170	31/8/2009	11:37:10	45
-84.4985	39.1714	80	256	8	1.2	170	31/8/2009	11:37:11	22
-84.4987	39.1714	80	254	8	1.2	170	31/8/2009	11:37:12	22
-84.499	39.1713	78	252	9	0.93	169	31/8/2009	11:37:13	22
-84.4992	39.1713	78	252	8	1.1	169	31/8/2009	11:37:14	21
-84.4994	39.1712	78	250	8	1.1	169	31/8/2009	11:37:15	21
-84.4997	39.1712	80	248	8	0.95	169	31/8/2009	11:37:16	21
-84.4999	39.1711	78	248	8	0.95	169	31/8/2009	11:37:17	22

Note, HDOP = Horizontal Dilution of Precision.

Table 3.2 Operating Mode Bins for MOVES Running Emissions

<table>
<tr><th rowspan="15">Vehicle Specific Power (VSP kw/ton)</th><th colspan="5">Instantaneous Speed (mi/h)</th></tr>
<tr><th></th><th>0</th><th>0-25</th><th>25-50</th><th>>50</th></tr>
<tr><td>>30</td><td rowspan="11"></td><td rowspan="8">Bin 16</td><td rowspan="2">Bin 30</td><td rowspan="2">Bin 40</td></tr>
<tr><td>30</td></tr>
<tr><td>27</td><td rowspan="2">Bin 29</td><td rowspan="2">Bin 39</td></tr>
<tr><td>24</td></tr>
<tr><td>21</td><td rowspan="2">Bin 28</td><td rowspan="2">Bin 38</td></tr>
<tr><td>18</td></tr>
<tr><td>15</td><td rowspan="2">Bin 27</td><td rowspan="2">Bin 37</td></tr>
<tr><td>12</td></tr>
<tr><td>9</td><td>Bin 15</td><td>Bin 25</td><td rowspan="2">Bin 35</td></tr>
<tr><td>6</td><td>Bin 14</td><td>Bin 24</td></tr>
<tr><td>3</td><td>Bin 13</td><td>Bin 23</td><td rowspan="3">Bin 33</td></tr>
<tr><td>0</td><td>Bin 1 (Idling)
Bin 0 (Braking)</td><td>Bin 12</td><td>Bin 22</td></tr>
<tr><td><0</td><td></td><td>Bin 11</td><td>Bin 21</td></tr>
</table>

Figure 3.5 shows an example of an operating mode distribution by VSP bins. Each household trip corresponding to an operating mode distribution and used as traffic activity inputs for the MOVES model.

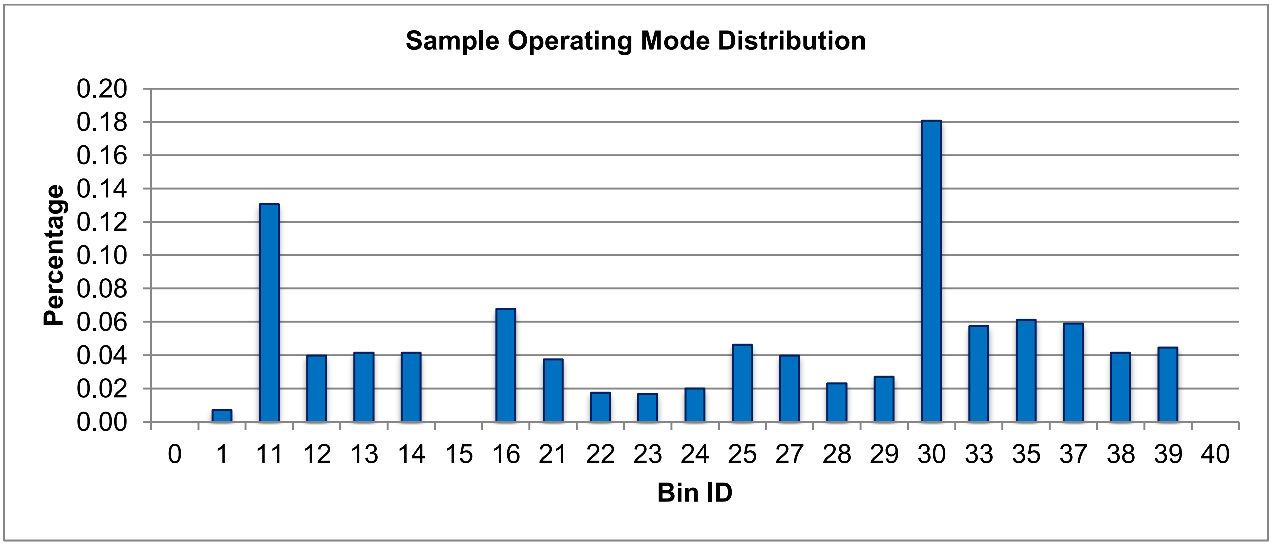


Figure 3.5 Sample Operating Mode Distribution

CHAPTER 4 MODELING HOUSEHOLD TRAVEL CARBON EMISSIONS

4.1 GPS Household Travel Survey Data

4.1.1. Geographical Boundary of the Survey

The preliminary study area for the integrated land use and household travel carbon emission analysis is Hamilton County, Ohio. The study area boundary is illustrated in Figure 4.1 below. The area contains 693 traffic analysis zones (TAZ) covering a total of 264,065.2 acres of land. Number of households is 355,777 and total employment of 525,862 in the year 2008.

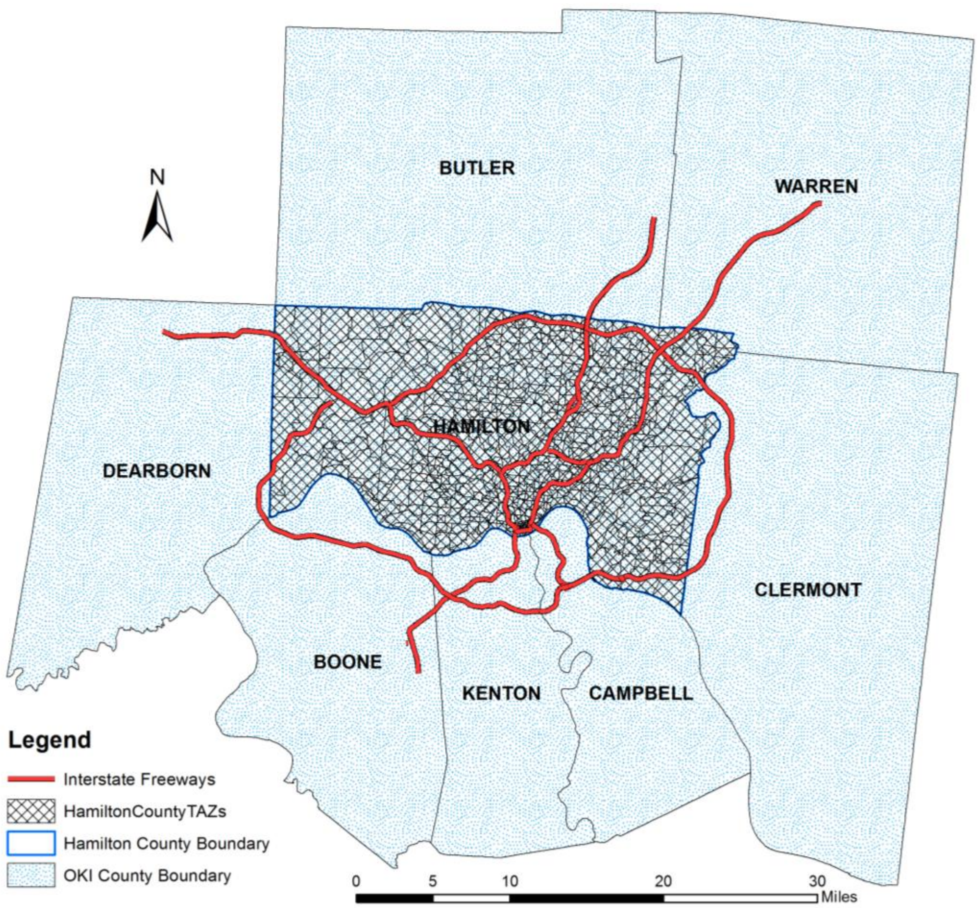


Figure 4.1 The Predefined Study Area of this Research

Figure 4.2 shows the locations household travel survey samples over the Hamilton County land use map. There are 2,697 sampled households within the county. Each household's land use information will also be included in the household stratification.

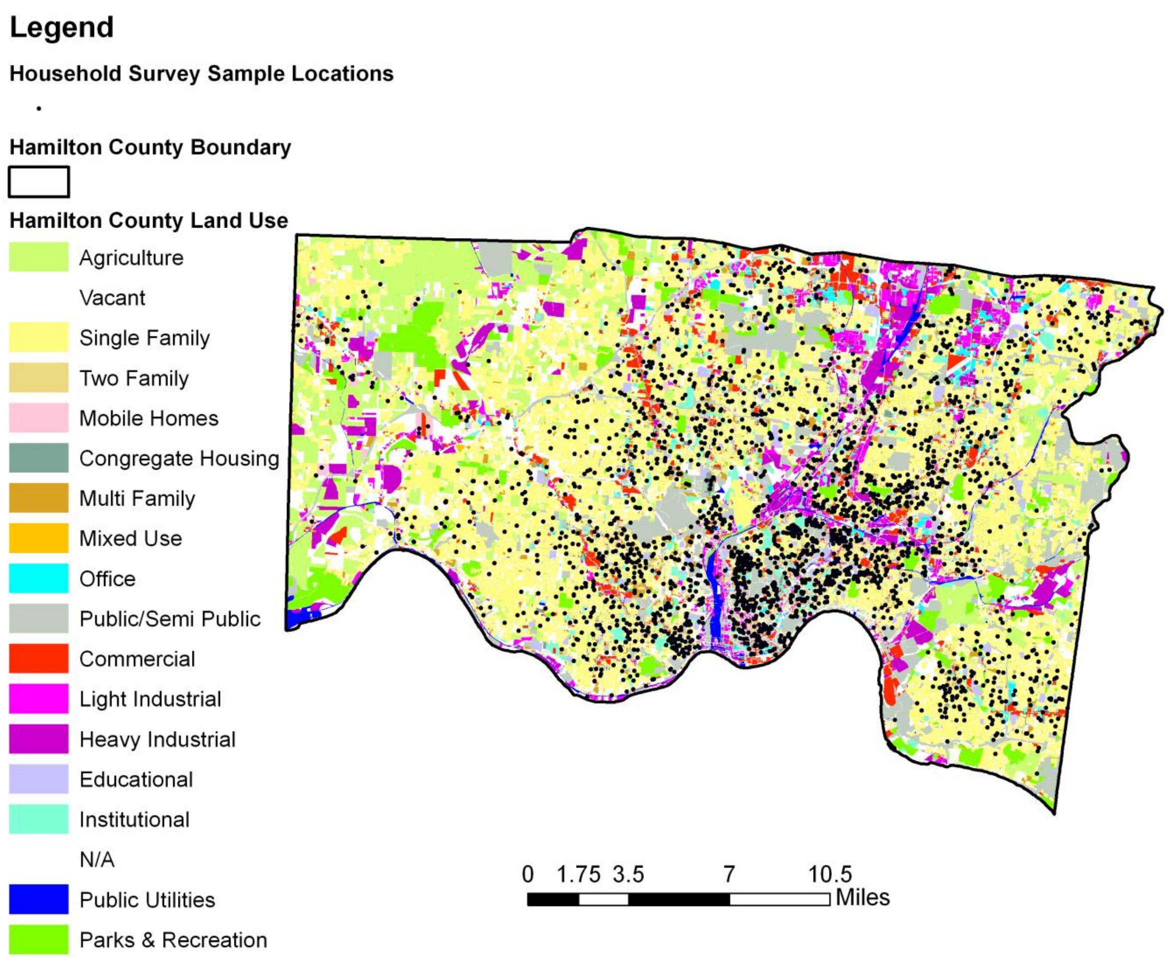


Figure 4.2 Household Travel Survey Sample Locations and Its Land Use Type for Hamilton County

4.1.2. Household Travel Survey Data Analysis and Modeling

The Greater Cincinnati Household Travel Survey covers households from all eight counties in the OKI region. A total of 5,502 households are included in the eight-county area. Figure 4.3 shows the location distribution of the sampled households. However, according to the final report of the Cincinnati household interview survey, the final number of households with GPS complete data collected is 2,059 and 549 for GPS incomplete (Stopher & Wargelin, 2012). According to the report, the completed sample region-wide was very representative by county, household size, number of vehicles, and the number of workers, with the exception that zero-vehicle households were underrepresented in completes by 5.3%, despite $25 incentives for completion.

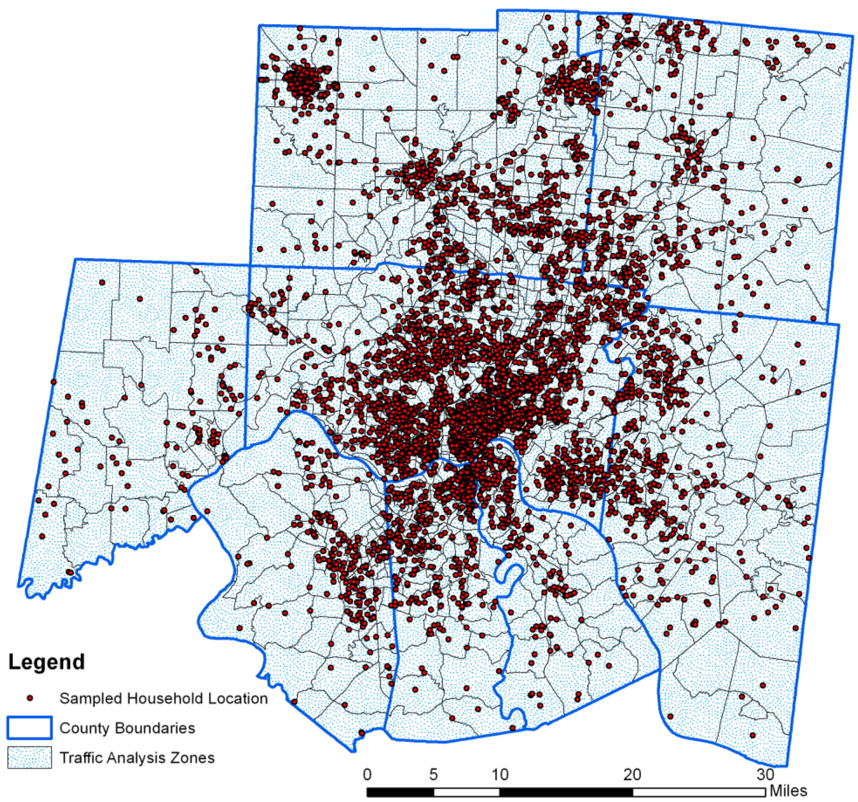


Figure 4.3 Spatial Distribution of Sampled Household Locations from Survey

4.2 Household Type Distribution by Area Type

The household type distribution is assumed to identical by area type. Each TAZ located in the same area type is assumed to have the same distribution of household types. Table 4.1 shows the household type distributions by area type.

Table 4.1 Household Type Distributions by Area Type

HHID	Area Type	Workers	Life Cycle	Income Group 1	Income Group 2	Income Group 3	Income Group 4
1	1	1	1	0.044850	0.013800	0.005914	0.003450
2	1	2	1	0.041400	0.101528	0.063578	0.066535
3	1	3	1	0.006407	0.034007	0.040907	0.103499
4	1	4	1	0.000493	0.001479	0.006900	0.017250
5	1	5	1	0.000000	0.000986	0.002464	0.002464
6	1	1	2	0.005914	0.001971	0.000000	0.000000
7	1	2	2	0.004436	0.006900	0.000986	0.000000
8	1	3	2	0.000000	0.000000	0.000000	0.000000
9	1	4	2	0.000493	0.000000	0.000000	0.000000
10	1	5	2	0.000000	0.000000	0.000000	0.000000
11	1	1	3	0.019714	0.019714	0.012321	0.005914
12	1	2	3	0.000000	0.000000	0.000000	0.000000
13	1	3	3	0.000000	0.000000	0.000000	0.000000
14	1	4	3	0.000000	0.000000	0.000000	0.000000
15	1	5	3	0.000000	0.000000	0.000000	0.000000
16	1	1	4	0.024643	0.005421	0.000493	0.002957
17	1	2	4	0.020207	0.021686	0.023164	0.042878
18	1	3	4	0.005421	0.018728	0.036964	0.120256
19	1	4	4	0.000493	0.001971	0.009857	0.023164
20	1	5	4	0.000000	0.000000	0.000986	0.004436
21	2	1	1	0.015407	0.008804	0.002201	0.002935
22	2	2	1	0.019076	0.073368	0.052091	0.067498
23	2	3	1	0.005869	0.027146	0.041820	0.102715
24	2	4	1	0.000734	0.003668	0.010271	0.027146
25	2	5	1	0.000000	0.000000	0.002935	0.000734
26	2	1	2	0.005869	0.002935	0.000000	0.000000
27	2	2	2	0.008070	0.004402	0.001467	0.002201
28	2	3	2	0.002935	0.000000	0.000000	0.001467
29	2	4	2	0.000000	0.000000	0.000734	0.001467
30	2	5	2	0.000000	0.000000	0.000734	0.000734
31	2	1	3	0.041820	0.057227	0.033749	0.024211

HHID	Area Type	Workers	Life Cycle	Income Group 1	Income Group 2	Income Group 3	Income Group 4
32	2	2	3	0.000000	0.000000	0.000000	0.000000
33	2	3	3	0.000000	0.000000	0.000000	0.000000
34	2	4	3	0.000000	0.000000	0.000000	0.000000
35	2	5	3	0.000000	0.000000	0.000000	0.000000
36	2	1	4	0.008804	0.002201	0.000734	0.003668
37	2	2	4	0.008804	0.021277	0.021277	0.044021
38	2	3	4	0.003668	0.015407	0.038151	0.133529
39	2	4	4	0.000000	0.000000	0.008070	0.027146
40	2	5	4	0.000000	0.000000	0.002201	0.006603
41	3	1	1	0.019417	0.012945	0.000000	0.000000
42	3	2	1	0.012945	0.061489	0.038835	0.064725
43	3	3	1	0.003236	0.022654	0.064725	0.126214
44	3	4	1	0.003236	0.000000	0.006472	0.029126
45	3	5	1	0.003236	0.000000	0.000000	0.000000
46	3	1	2	0.000000	0.000000	0.000000	0.000000
47	3	2	2	0.003236	0.003236	0.000000	0.000000
48	3	3	2	0.003236	0.003236	0.000000	0.000000
49	3	4	2	0.000000	0.000000	0.000000	0.003236
50	3	5	2	0.000000	0.000000	0.000000	0.000000
51	3	1	3	0.045307	0.048544	0.048544	0.042071
52	3	2	3	0.000000	0.000000	0.000000	0.000000
53	3	3	3	0.000000	0.000000	0.000000	0.000000
54	3	4	3	0.000000	0.000000	0.000000	0.000000
55	3	5	3	0.000000	0.000000	0.000000	0.000000
56	3	1	4	0.012945	0.003236	0.000000	0.003236
57	3	2	4	0.000000	0.022654	0.025890	0.032362
58	3	3	4	0.003236	0.019417	0.038835	0.126214
59	3	4	4	0.003236	0.000000	0.019417	0.016181
60	3	5	4	0.000000	0.000000	0.003236	0.000000

4.3 Household Cross-Classification

This modeling approach adopted OKI trip generation rates derived is to be derived from the survey data. The households in each zone are classified into household groups. The household groups are defined using four variables: Area Type, Number of Worker, Life Cycle and income group. “AT”, “W”, “LC” and “IC” are the index for value ranges for these variables. The dimensions for each variable are follows (Table 4.2):

Table 4.2 Household Cross-classification

Variable	Classification	Code
Area type	CBD & Urban, Suburban, Rural	corresponding to code (1, 2, 3)
Workers	0, 1, 2, 3, 4+	corresponding to code (1, 2, 3, 4, 5)
Life Cycle	Adult Household, Adult Student Household, Retiree, Household with Children	corresponding to code (1, 2, 3, 4)
Income Group	Less than \$25,000, \$25,000 to \$49,999, \$50,000 to \$74,999, \$75,000 or above.	corresponding to code (1, 2, 3, 4)

Figure 4.4 shows the household distributions by area type. The sample is dominated with household from area type of suburban area. Figure 4.5 shows the household distributions by number of workers. One and two workers households dominate this distribution. Figure 4.6 shows the household distribution by life cycle. The data show that the adult household occupies a large proportion of data while adult student households take a relevant small part. Figure 4.7 shows the distribution of the sampled households by income level. It appears that the higher household income levels are having a higher percentage.

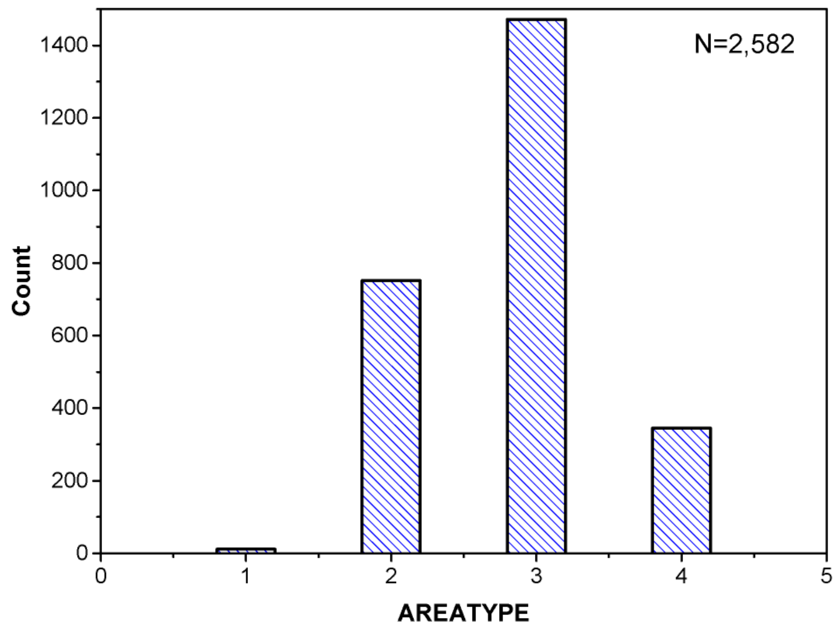


Figure 4.4 Household Distribution by Area Type
(Area Type 1= CBD, 2= Urban, 3=Suburban, 4= Rural)

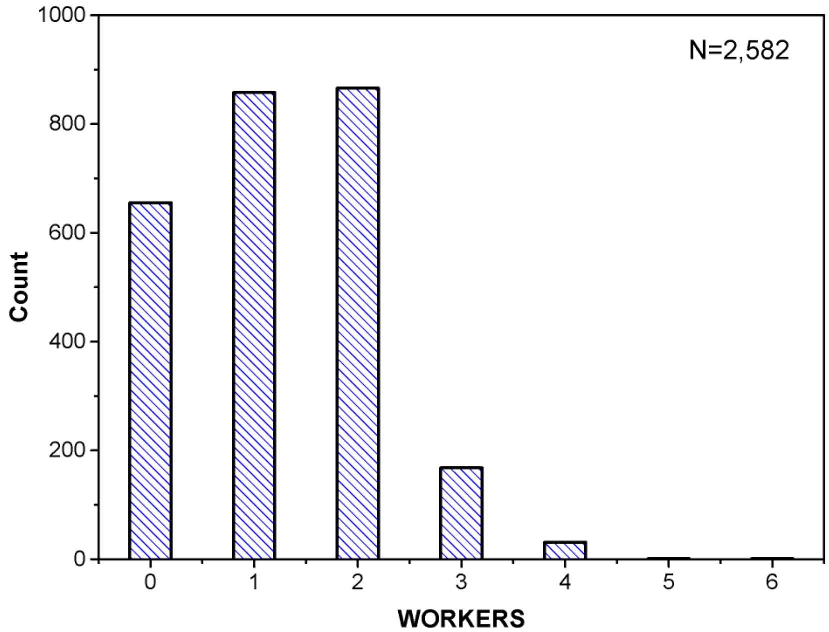


Figure 4.5 Household Distribution by Number of Workers

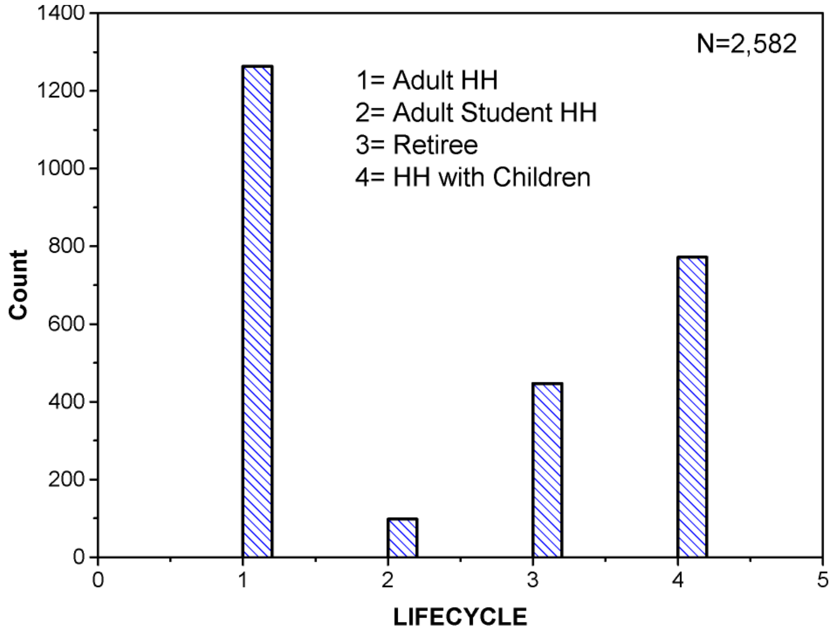


Figure 4.6 Household Distribution by Life Cycle

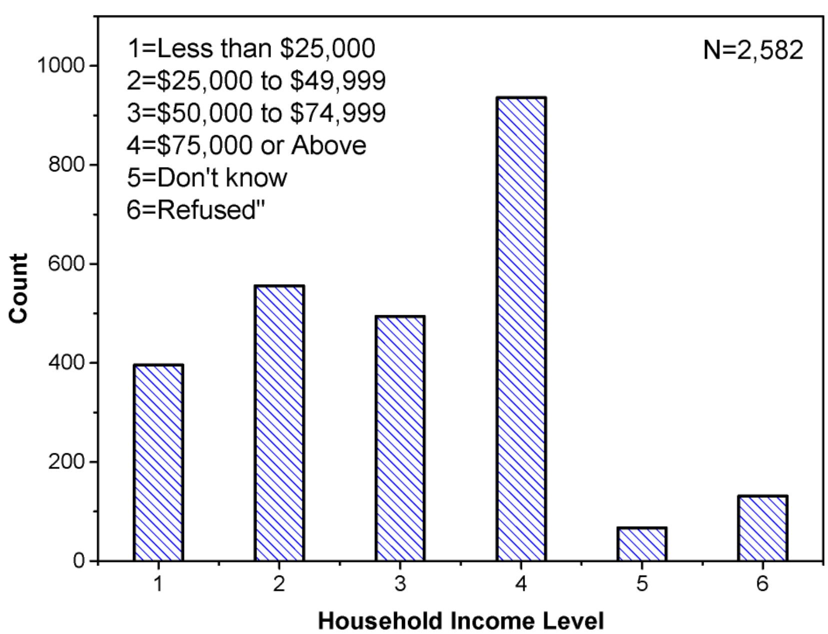


Figure 4.7 Household Distribution by Income Level

The household are then being stratified base on the above four variables.

4.4 Household Travel Carbon Emission Reasonableness Check

It is crucial to check the emission results from the survey data falls in the correct range by conforming to existing literature. A commonly recognized carbon emission from a gallon of gasoline is 8,887 grams (U.S. EPA, 2010) and the average fuel efficiency of U.S. light duty vehicles is 23.5 mpg(Bureau of Transportation Statistics, 2012). Therefore, the average carbon emission per mile is calculated using the equation below.

$$\mathrm{CO_2}\ \textit{emissions per mile} = \frac{CO_2\,\textit{per gallon}}{MPG} = \frac{8,887}{23.5^{*}} = 378.17\ \textit{grams} \tag{4.1}$$

Table 4.3 shows the empirical household travel carbon emission results from the survey data. The grams per mile emissions matches the EPA published carbon emission rates and the emission results from this calculation are valid.

Table 4.3 Household Travel Carbon Emissions Reasonableness Check

Workers	Carbon Emissions (lbs)	Trip Distance (miles)	Pounds per mile	Grams per mile
1	4.95	5.51	0.90	407.51
2	5.00	5.68	0.88	399.70
3	6.60	7.80	0.85	383.85
4	7.67	9.52	0.81	365.76
5	6.83	8.25	0.83	375.30

4.5 Carbon Emission Generation Rates by Household Type

Figure 4.8 shows the histogram of the daily aggregated household travel emissions from MOVES model. The mean is 0.00291, with median of 0.00255, minimum of 8.03E-7, maximum 0.01853 tons per day.

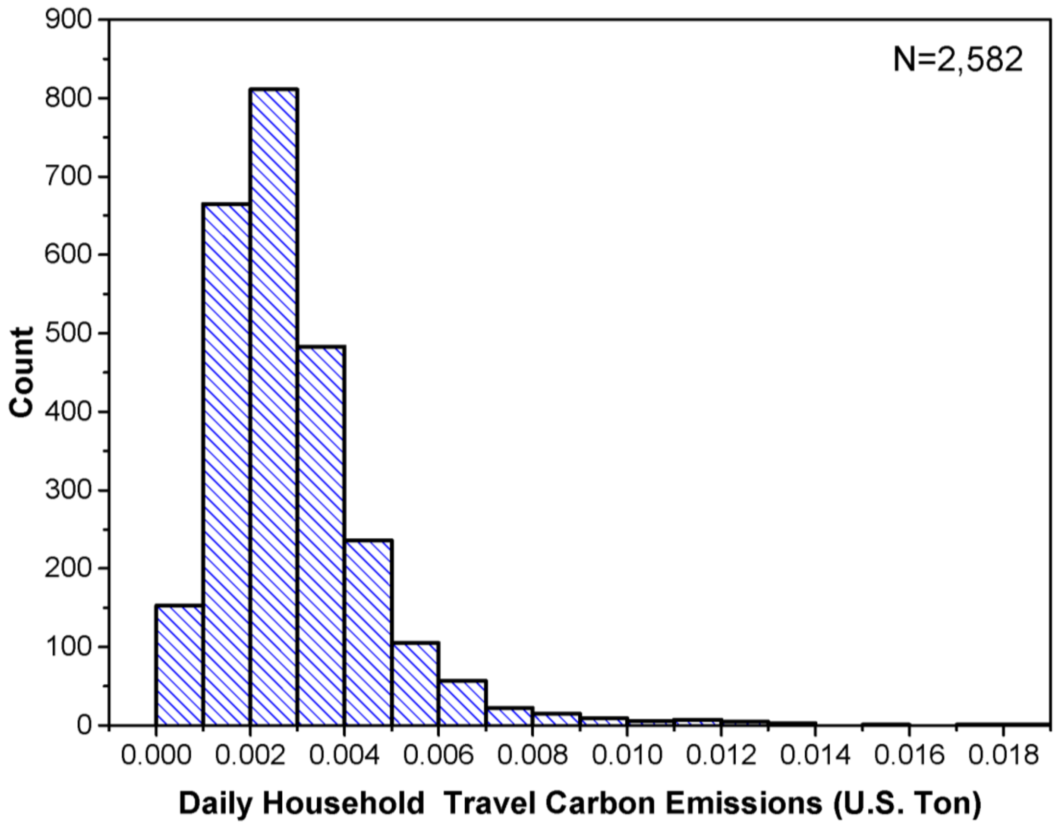


Figure 4.8 Histogram of Daily Household Travel Carbon Emissions

Figure 4.9 shows a summary of household travel carbon emissions by area type, household income, household life cycle and average travel speed.

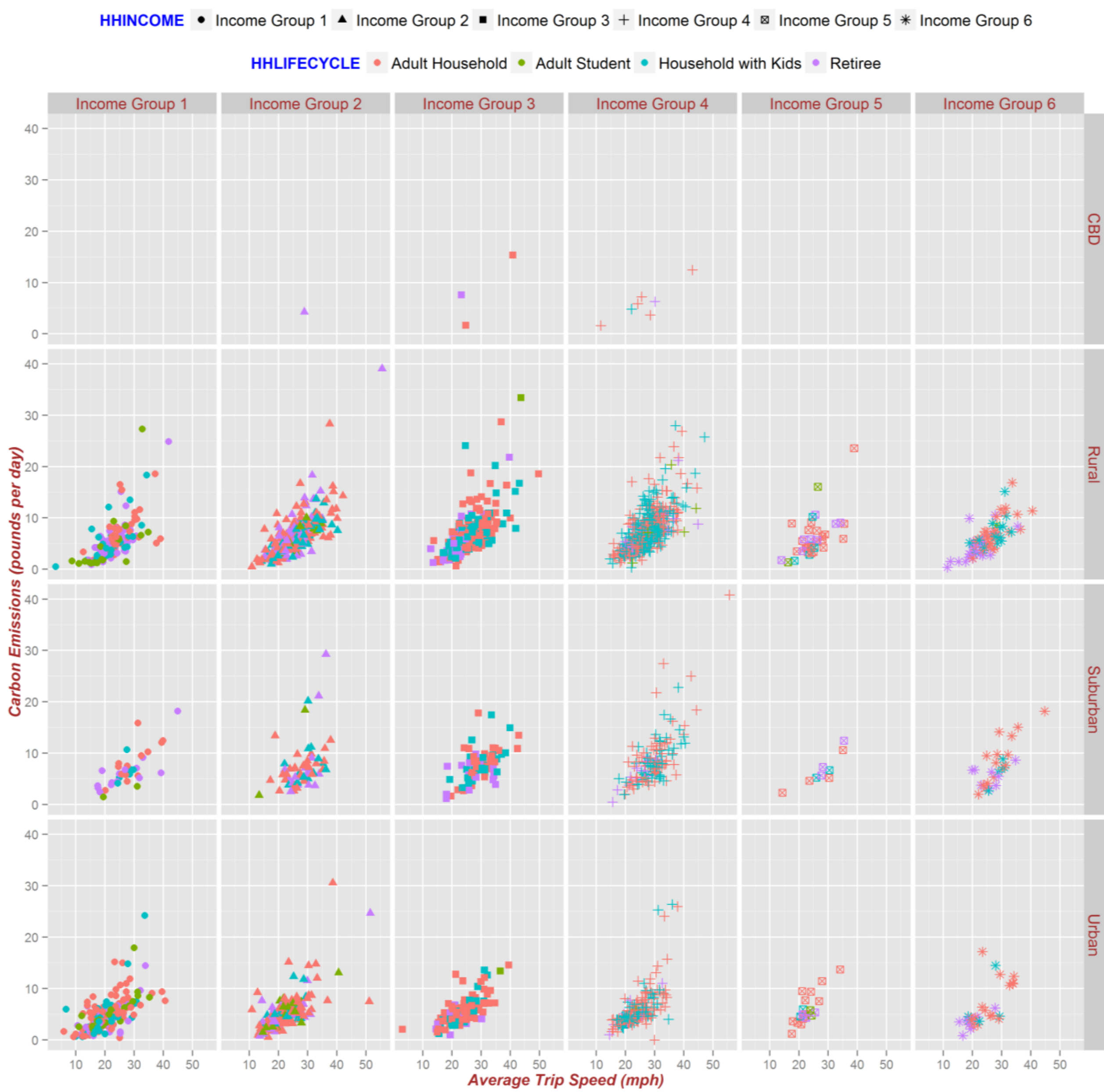


Figure 4.9 Household Travel Carbon Emission Distributions

Table 4.4 shows the results of the household travel carbon emission generation rates in pounds per day.

Table 4.4 Household Carbon Emission Generation Rates (Pounds/Day)

HHID	Area Type	Workers	Life Cycle	Income Group 1	Income Group 2	Income Group 3	Income Group 4
1	1	1	1	4.64131	4.91955	4.55828	5.79650
2	1	2	1	4.81988	5.07747	5.77688	6.31216

HHID	Area Type	Workers	Life Cycle	Income Group 1	Income Group 2	Income Group 3	Income Group 4
3	1	3	1	6.20307	5.42799	6.68672	6.36269
4	1	4	1	2.83918	6.79139	5.82965	6.89034
5	1	5	1	0.00000	3.37790	14.22109	3.63236
6	1	1	2	4.22531	3.89395	0.00000	1.35433
7	1	2	2	5.74784	5.18164	7.81896	0.00000
8	1	3	2	0.00000	0.00000	0.00000	0.00000
9	1	4	2	1.89193	1.25000	0.00000	0.00000
10	1	5	2	2.96627	0.00000	0.00000	5.41732
11	1	1	3	4.29514	5.44669	4.66340	5.44904
12	1	2	3	0.00000	0.00000	0.00000	0.00000
13	1	3	3	0.00000	0.00000	0.00000	0.00000
14	1	4	3	0.00000	0.00000	0.00000	0.00000
15	1	5	3	0.00000	0.00000	0.00000	0.00000
16	1	1	4	4.33797	5.06331	2.92980	4.99933
17	1	2	4	5.07307	5.39772	6.22435	6.61065
18	1	3	4	4.02470	5.44160	5.66269	6.42666
19	1	4	4	9.67274	6.60560	6.71683	5.88278
20	1	5	4	0.00000	0.00000	5.79194	14.07351
21	2	1	1	6.30902	4.09567	4.34251	6.93872
22	2	2	1	4.93542	6.02093	5.82540	6.50729
23	2	3	1	6.66593	5.66594	7.41647	6.95826
24	2	4	1	5.81686	5.46357	5.64396	5.99105
25	2	5	1	0.00000	0.00000	5.76829	6.37386
26	2	1	2	3.72302	5.40614	0.00000	0.00000
27	2	2	2	4.85571	6.22565	18.14048	10.81219
28	2	3	2	3.67699	0.00000	0.00000	6.53797
29	2	4	2	0.00000	0.00000	4.66028	5.48925
30	2	5	2	0.00000	0.00000	2.13660	1.08949
31	2	1	3	4.45467	5.67472	4.84195	5.85428
32	2	2	3	0.00000	0.00000	0.00000	0.00000
33	2	3	3	0.00000	0.00000	0.00000	0.00000
34	2	4	3	0.00000	0.00000	0.00000	0.00000
35	2	5	3	0.00000	0.00000	0.00000	0.00000
36	2	1	4	5.76837	5.21523	2.92980	5.11889
37	2	2	4	4.22806	5.45767	5.88147	6.88901
38	2	3	4	4.72798	5.44898	5.99503	6.35977
39	2	4	4	0.00000	0.00000	6.76653	5.55743
40	2	5	4	0.00000	0.00000	5.42137	7.87749
41	3	1	1	5.17021	6.91601	0.00000	0.00000
42	3	2	1	9.01389	6.40798	6.96355	9.88652

HHID	Area Type	Workers	Life Cycle	Income Group 1	Income Group 2	Income Group 3	Income Group 4
43	3	3	1	9.28870	5.36092	7.64528	7.42979
44	3	4	1	14.36542	0.00000	7.65829	10.13513
45	3	5	1	6.82608	0.00000	0.00000	0.00000
46	3	1	2	0.00000	0.00000	0.00000	0.00000
47	3	2	2	3.21112	16.69436	0.00000	0.00000
48	3	3	2	1.30651	1.63195	0.00000	0.00000
49	3	4	2	0.00000	0.00000	0.00000	5.24420
50	3	5	2	0.00000	0.00000	0.00000	0.00000
51	3	1	3	5.62989	7.26165	5.07731	5.21164
52	3	2	3	0.00000	0.00000	0.00000	0.00000
53	3	3	3	0.00000	0.00000	0.00000	0.00000
54	3	4	3	0.00000	0.00000	0.00000	0.00000
55	3	5	3	0.00000	0.00000	0.00000	0.00000
56	3	1	4	4.99809	9.88334	0.00000	4.40152
57	3	2	4	0.00000	5.58300	6.51314	6.39684
58	3	3	4	6.12788	8.25541	6.87408	7.88507
59	3	4	4	9.67274	0.00000	7.72350	10.97381
60	3	5	4	0.00000	0.00000	6.90362	5.77294

Figure 4.10 to Figure 4.21 shows the carbon emission generation rates of households by area type, number of workers, life cycle, and income level.

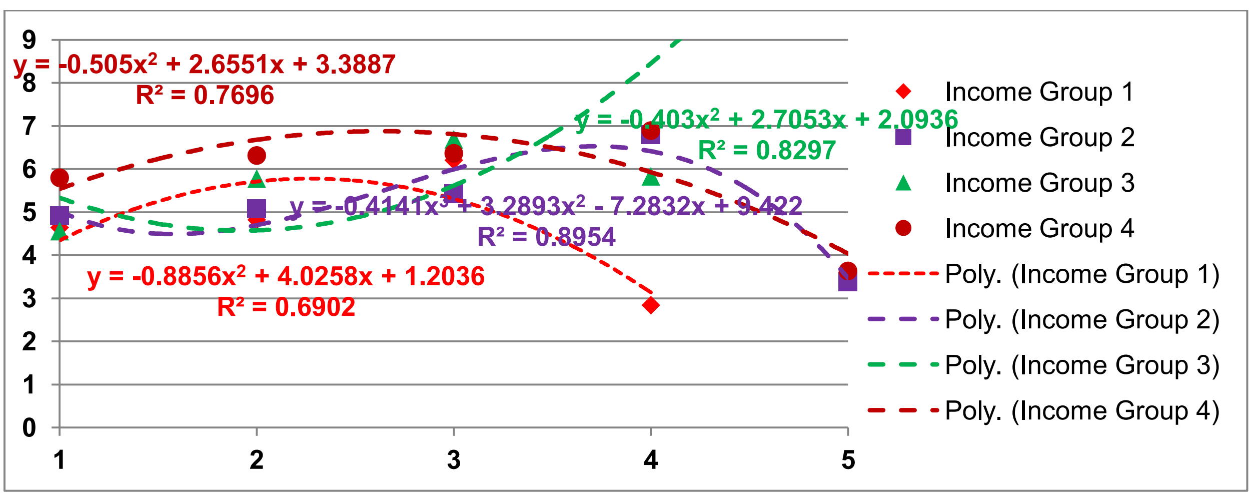


Figure 4.10 Carbon Emission Generation Rates for AT1 (CBD & Urban) Life Cycle 1 (Adult Household)

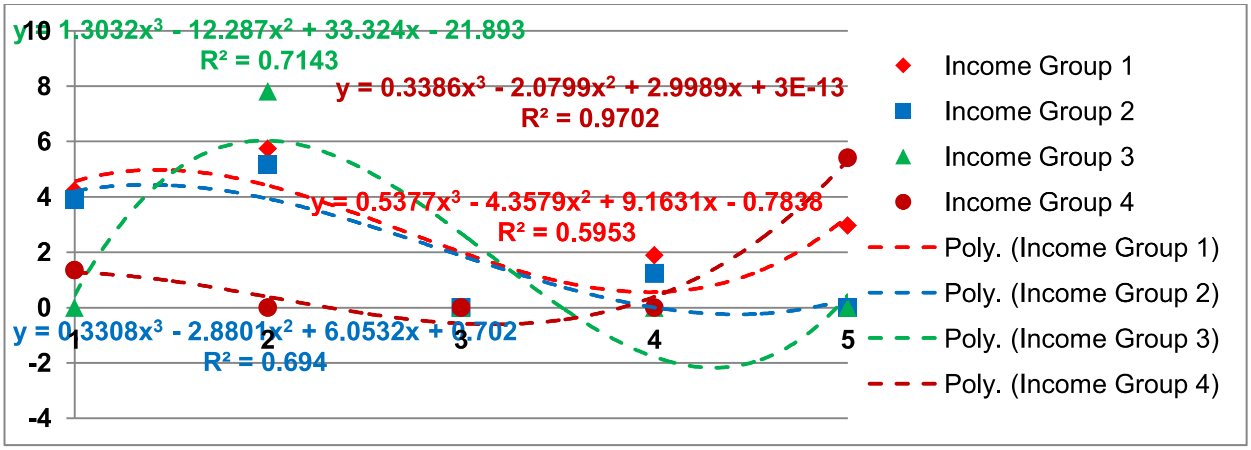


Figure 4.11 Carbon Emission Generation Rates for AT1 (CBD & Urban) Life Cycle 2 (Adult Student Household)

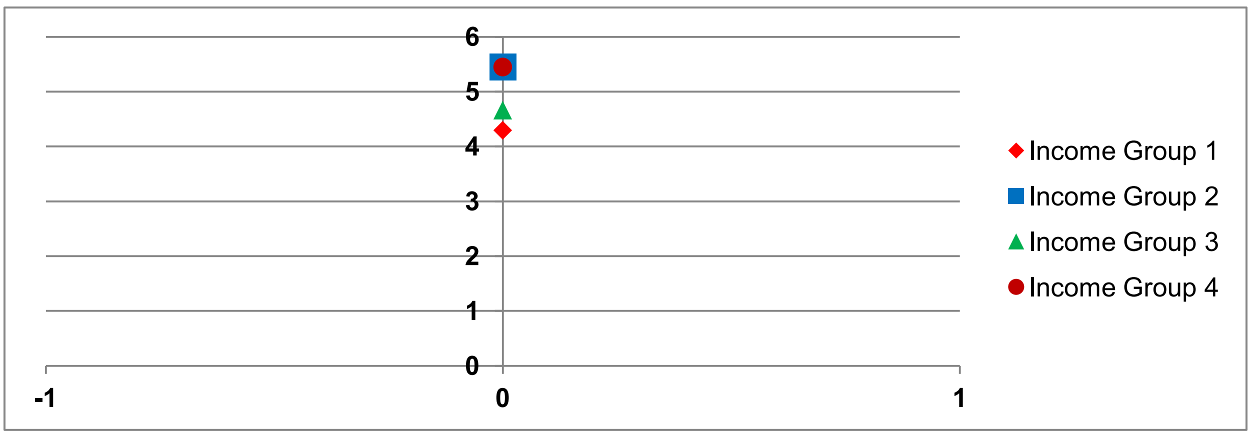


Figure 4.12 Carbon Emission Generation Rates for AT1 (CBD & Urban) Life Cycle 3 (Retiree)

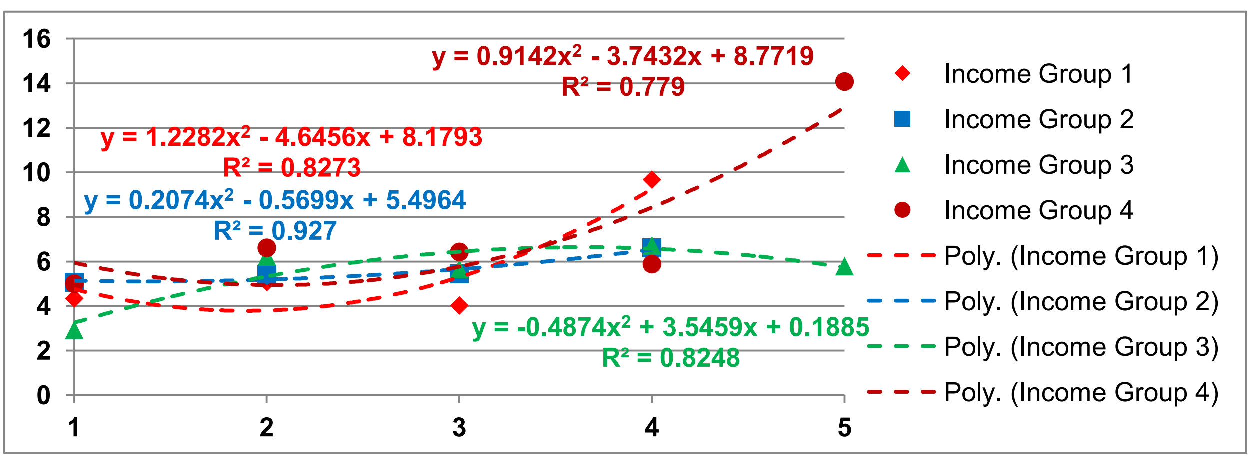


Figure 4.13 Carbon Emission Generation Rates for AT1 (CBD & Urban) Life Cycle 4 (Household with Children)

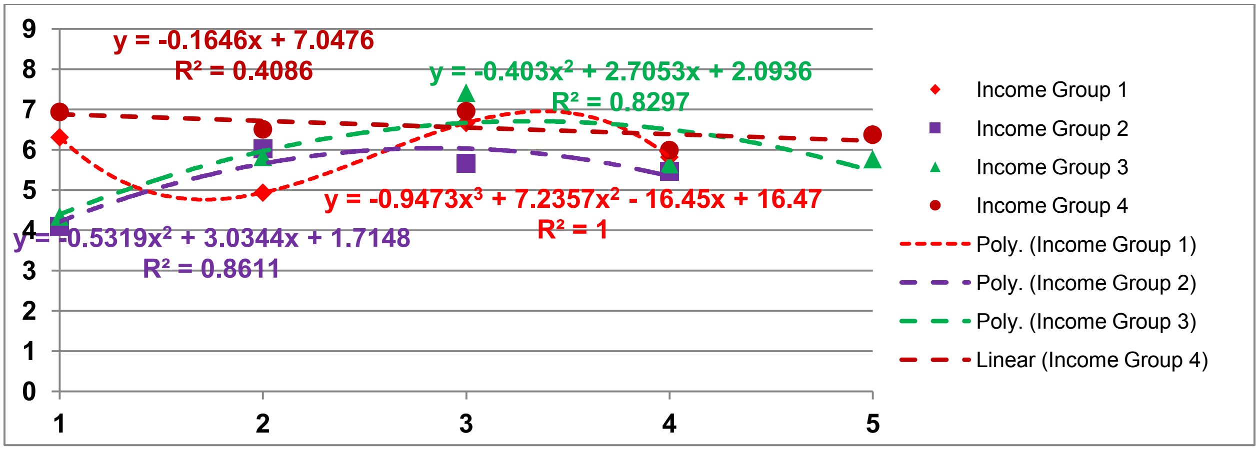


Figure 4.14 Carbon Emission Generation Rates for AT2 (Suburban), Life Cycle 1 (Adult Household)

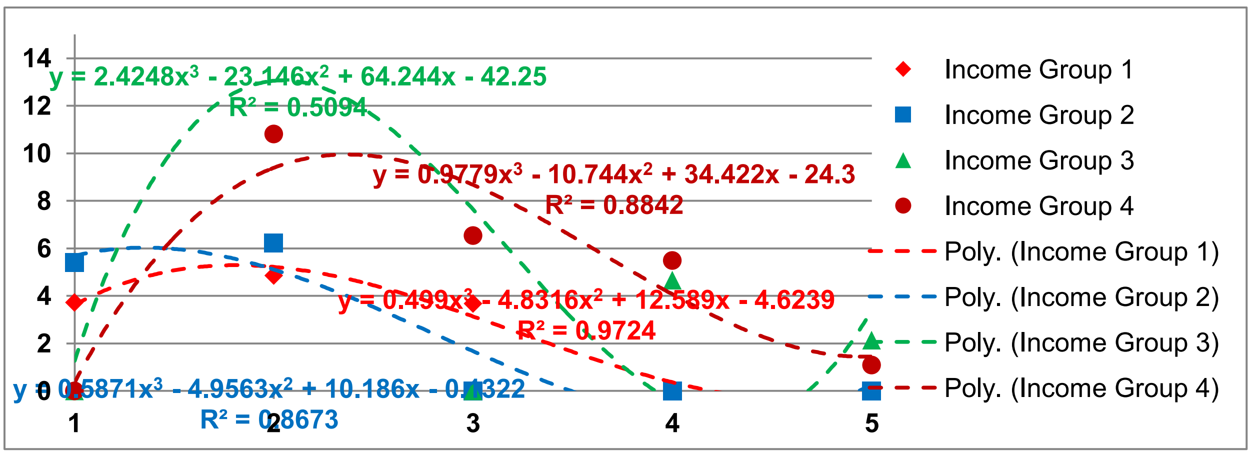


Figure 4.15 Carbon Emission Generation Rates for AT2 (Suburban), Life Cycle 2 (Adult Student Household)

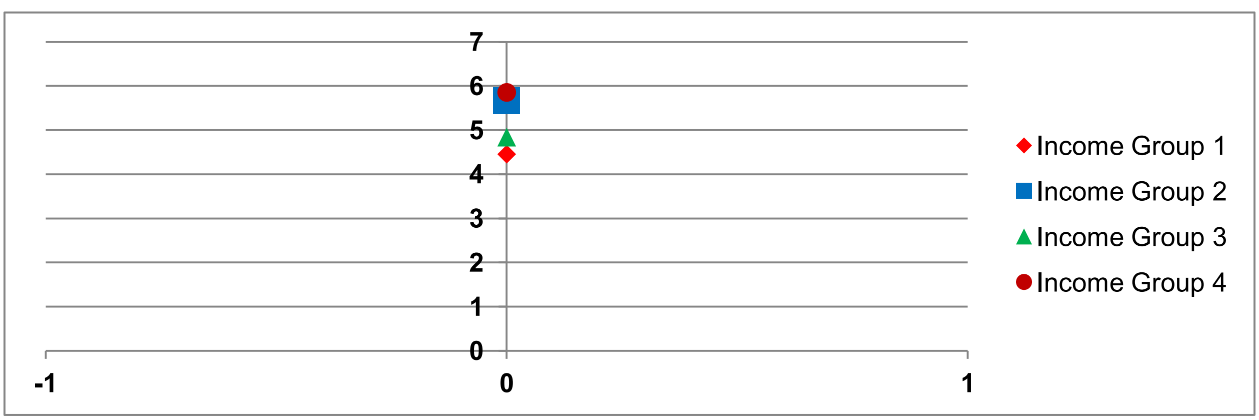


Figure 4.16 Carbon Emission Generation Rates for AT2 (Suburban), Life Cycle 3 (Retiree)

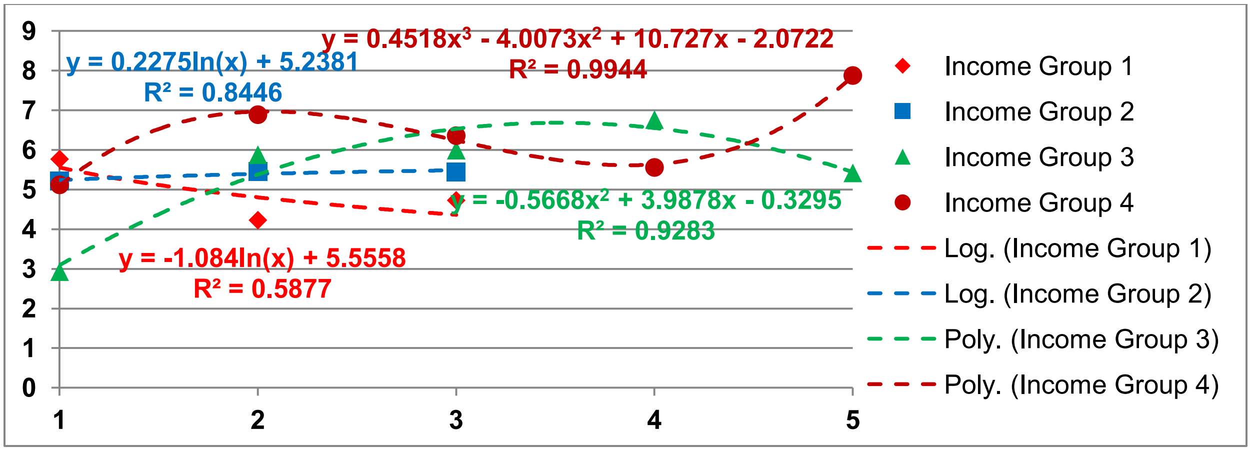


Figure 4.17 Carbon Emission Generation Rates for AT2 (Suburban), Life Cycle 4 (Household with Children)

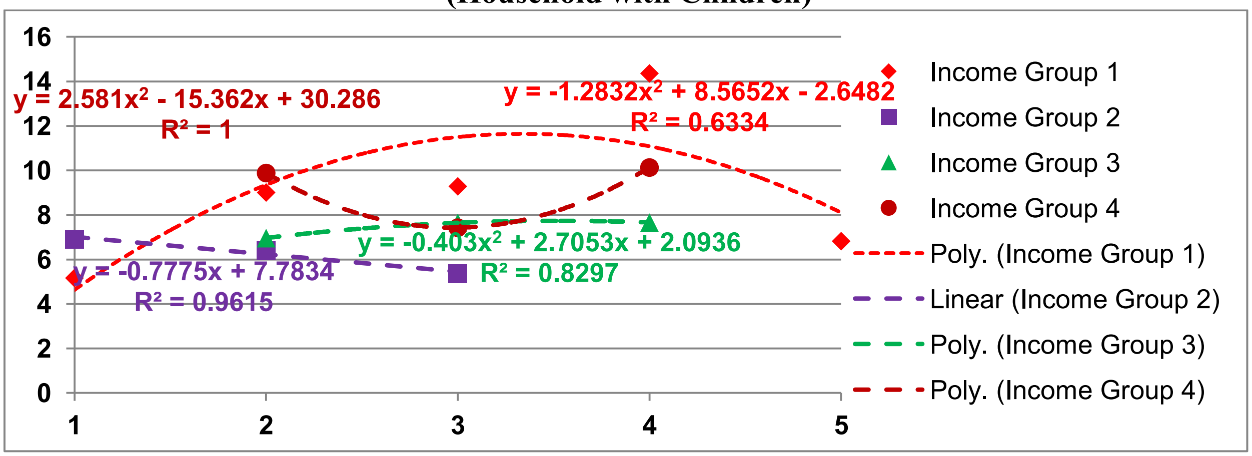


Figure 4.18 Carbon Emission Generation Rates for AT3 (Rural), Life Cycle 1 (Adult Household)

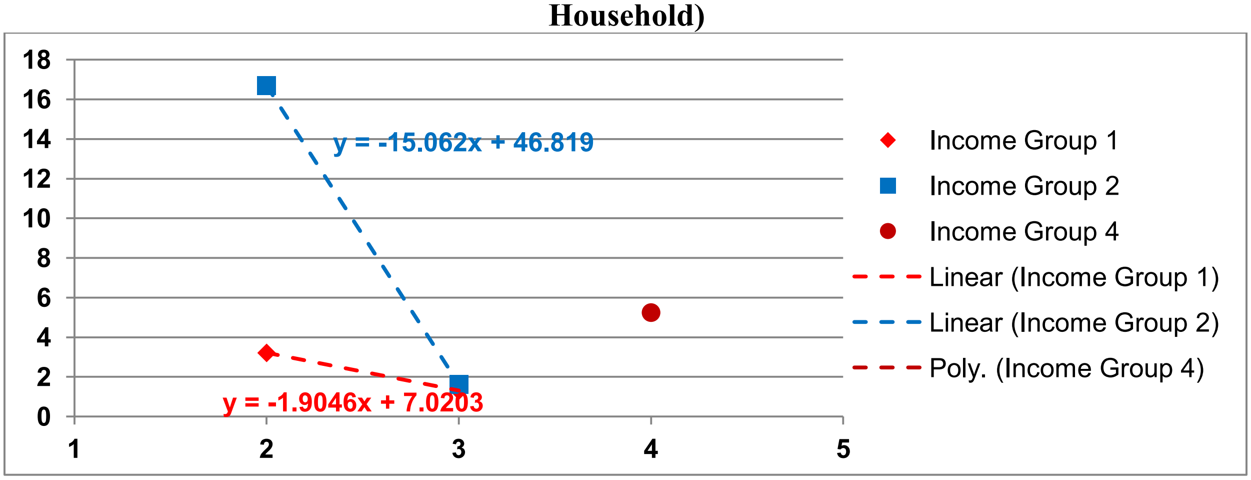


Figure 4.19 Carbon Emission Generation Rates for AT3 (Rural), Life Cycle 2 (Adult Student Household)

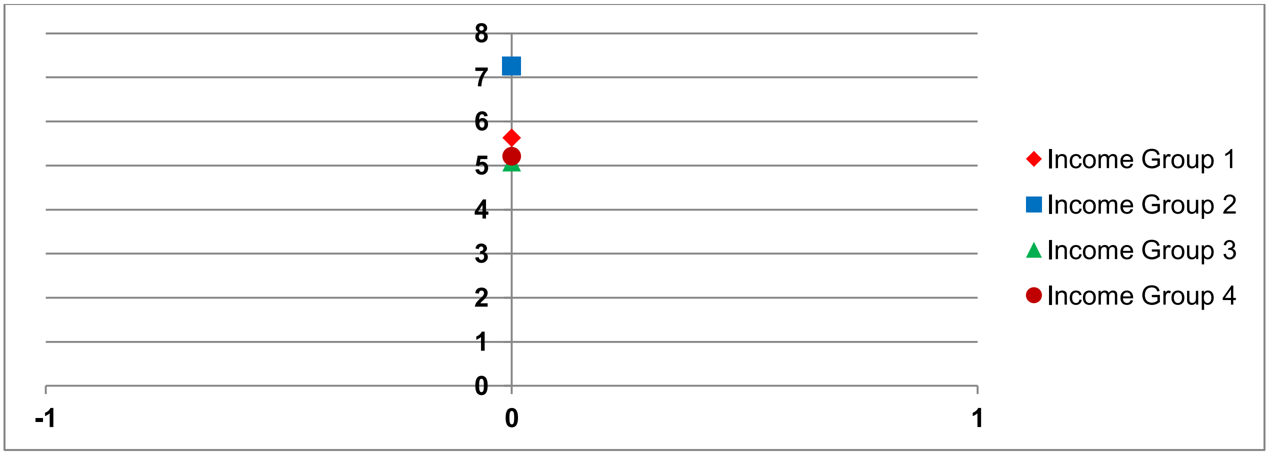


Figure 4.20 Carbon Emission Generation Rates for AT3 (Rural), Life Cycle 3 (Retiree)

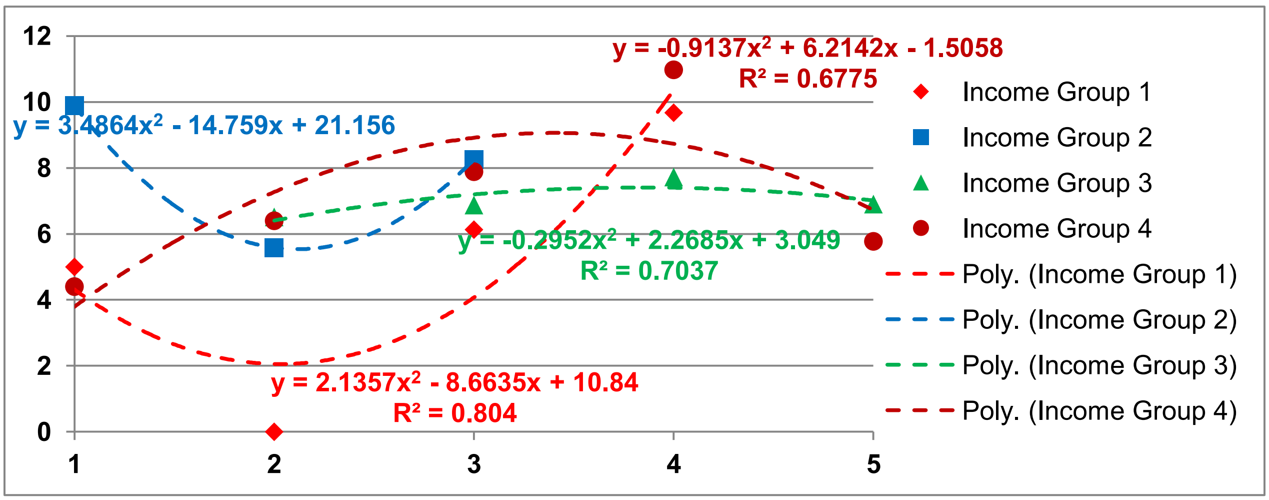


Figure 4.21 Carbon Emission Generation Rates for AT3 (Rural), Life Cycle 4 (Household with Children)

4.6 Zonal Carbon Emissions

The zonal carbon emissions are then calculated as below:

$$\begin{aligned}&Zonal\ Carbon\ Emission\ = \\ &Sum\ (Household\ No.\ \times Percentage\ of\ \ HH\ \ ID(i)\times Emission\ Rate\ of\ \ HH\ \ ID(i))\end{aligned} \tag{4.2}$$

The results of TAZ level carbon emissions are then presented in Figure 4.22 by showing its empirical distribution.

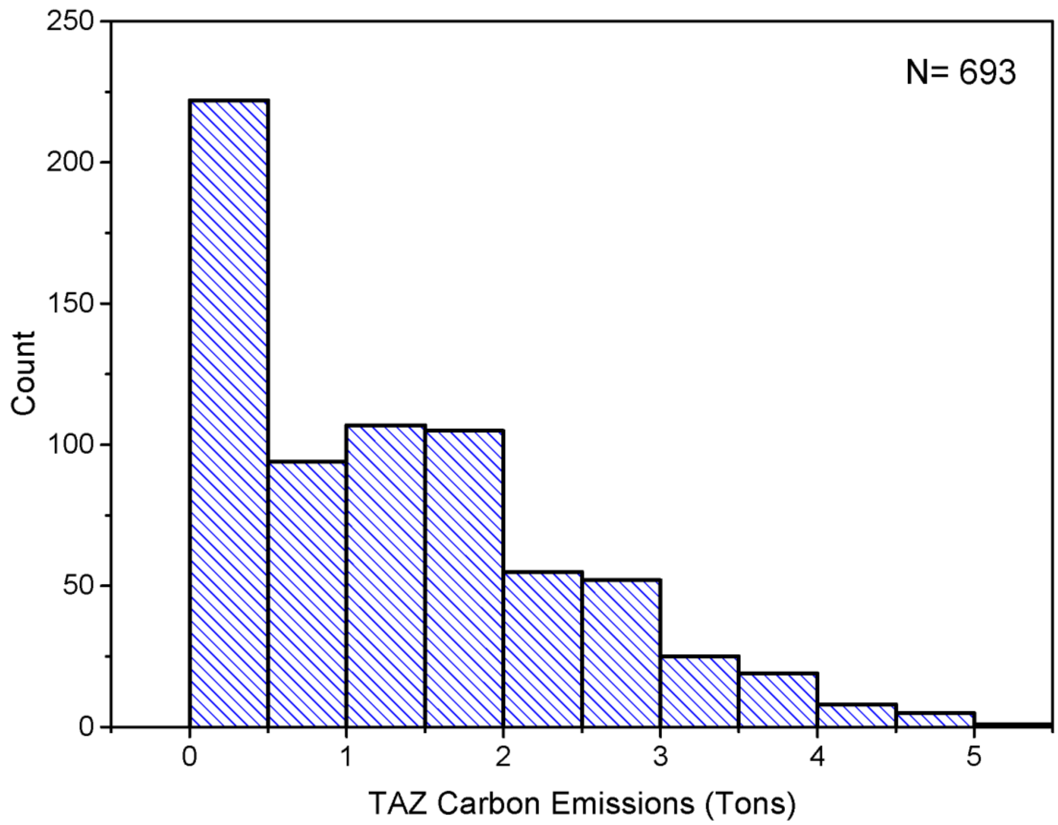


Figure 4.22 Histogram of TAZ Level Carbon Emissions

4.7 Contributing Variables for Household Travel Carbon Emissions

The contributing variables for household travel carbon emissions are categorized into two levels: TAZ (Macroscopic) and Trips (Microscopic). Those variables are included from the GPS household travel survey. Table 4.5 provides a list of descriptions for the contributing variables included in the preliminary study.

Table 4.5 List of Contributing Variables for Household Travel Carbon Emissions

Category	Code	Description	Units
TAZ	ACRES	TAZ Area in Acres	Acres
	AT	CBD, Urban, Suburban, Rural	NA
	POP	Population in Zone i	NA
	TOTAL_HH	Total Households in Zone i	Household
	TOTAL_EMPL	Total Employment in Zone i	Employment
	POP_DENSIT	Population Density	Person per acre

Category	Code	Description	Units
	EMP_DENSIT	Employment Density	Employment per acre
	TOTAL_AUTO	Automobiles in Zone i	Vehicle
	EMP_L	The low trip rate employment (Agriculture, Construction, Manufacturing, Mining, Transportation, Communications, Utility) in zone i	Employment
	EMP_M	The medium trip rate employment (Finance, Insurance, Real Estate, Public, Service, Wholesale Trade) in zone i	Employment
	EMP_H	The high trip rate employment (Retail) in zone i	Employment
	AVGWK	Average Worker Per Household	Worker
	AVGPER	Average Person Per Household	Person
	AVGAUTO	Average Auto Owned Per Household	Automobile
Trips	Avg_CarbEM	Trip Carbon Emission Average from Survey Data	U.S. Tons
	Avg_TRIPDI	Trip Distance Average from Survey Data	Miles
	Avg_TRIPSP	Trip Duration Average from Survey Data	Minutes
	Avg_TRIPDU	Trip Speed Average from Survey Data	Miles Per Hour

To identify the key contributing variables from the candidate variables listed in Table 4.5, the stepwise variable selection is used. The backward elimination, which involves starting with all candidate variables, testing the deletion of each variable using a chosen model comparison criterion, deleting the variable (if any) that improves the model the most by being deleted, and repeating this process until no further improvement is possible. Table 4.6 summarized the results from backward stepwise variable selection using the Akaike information criterion (AIC). AIC deals with the trade-off between the goodness of fit of the model and the complexity of the model. It is founded on information theory: it offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. For any statistical model, the AIC value is:

$$AIC = 2k - 2\ln(L) \tag{4.3}$$

where k is the number of parameters in the model, and L is the maximized value of the likelihood function for the model.

Table 4.6 Stepwise Variable Selection

Step	AIC	Dropped Variable
1	Start AIC=-937.43	EMP_L
2	Step AIC=-939.43	Avg_TRIPDI
3	Step AIC=-941.42	Avg_TRIPDU
4	Step AIC=-943.4	EMP_H
5	Step AIC=-945.25	AVGPER
6	Step AIC=-946.57	EMP_DENSIT

4.8 K-fold Cross-Validation of the OLS Model

K-fold cross validation is one way to improve over the holdout method. The data set is divided into k subsets, and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set and the other k-1 subsets are put together to form a training set. Then the average error across all k trials is computed. The advantage of this method is that it matters less how the data gets divided. Every data point gets to be in a test set exactly once, and gets to be in a training set k-1 times. The variance of the resulting estimate is reduced as k is increased. The disadvantage of this method is that the training algorithm has to be rerun from scratch k times, which means it takes k times as much computation to make an evaluation. A variant of this method is to randomly divide the data into a test and training set k different times. The advantage of doing this is that you can independently choose how large each test set is and how many trials you average over. A common k number for model cross validation is 10. However, since there are 693 TAZ in our dataset, a k = 9 is used to ensure each "fold" is equal.

Since the data are randomly assigned to a number of 'folds' (K=9 in this case). Each fold is removed, in turn, while the remaining data is used to refit the regression model and the deleted observations are predicted. Figure 4.23 is the validation plot showing the removed (folded) vs. fitted data. The validation plot shows a good validation since each removed vs. fitted data flows similar 45-degree line.

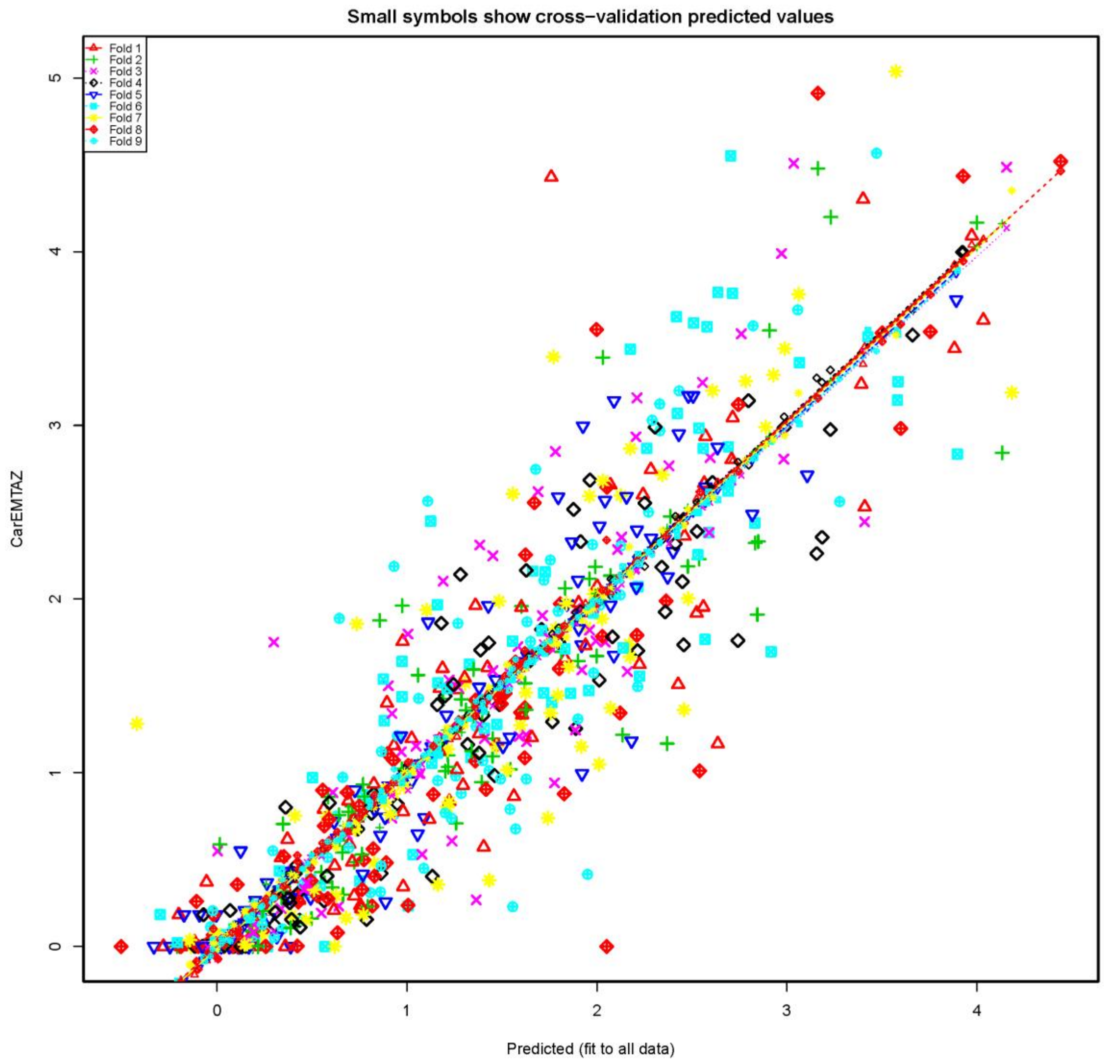


Figure 4.23 9-Fold Cross-Validation Results

CHAPTER 5 SPATIAL REGRESSION MODEL

5.1 OLS Regression Analysis Results

Table 5.1 shows the variables with its coefficient estimates. The R^2 (coefficient of determination) gives information about the goodness of fit of a model. In regression, the R^2 is a statistical measure of how well the regression line approximates the real data points. An R^2 of 1 indicates that the regression line perfectly fits the data. The linear model has a R^2 of 0.8002, which means the model is a good fit.

Table 5.1 OLS Regression Model and Coefficients

Variables	Estimate	Std. Error	t	value	Pr(>|t|)
(Intercept)	-1.61E-01	1.08E-01	-1.482	0.138675	
ACRES	7.61E-05	4.48E-05	1.697	0.09016	.
AT	1.87E-01	4.71E-02	3.965	8.10E-05	***
POP	4.57E-04	1.01E-04	4.535	6.80E-06	***
TOTAL_HH	4.20E-04	2.17E-04	1.932	0.053826	.
TOTAL_EMPL	-3.78E-05	2.48E-05	-1.52	0.128934	
POP_DENSIT	-2.13E-02	4.87E-03	-4.369	1.44E-05	***
EMP_DENSIT	4.21E-04	2.84E-04	1.485	0.13811	
TOTAL_AUTO	4.96E-04	1.04E-04	4.772	2.24E-06	***
EMP_M	2.40E-01	7.07E-02	3.39	0.00074	***
AVGWK	1.58E-01	8.50E-02	1.862	0.062971	.
AVGAUTO	-1.94E-01	7.48E-02	-2.594	0.009688	**
Avg_CarbEM	-7.54E+01	2.11E+01	-3.582	0.000365	***
Avg_TRIPSP	1.99E-02	3.56E-03	5.589	3.32E-08	***

(Residual standard error: 0.5001 on 679 degrees of freedom Multiple R-squared: 0.8002, Adjusted R-squared: 0.7964 F-statistic: 209.2 on 13 and 679 DF, p-value: < 2.2e-16)

Figure 5.1 is a diagnose plot of the fitted linear model. The first two plots (Residual and Normal Q-Q plots) describe the distribution of the residuals. Ideally, those two plots should be roughly normal. The Outliers (TAZ No. 669, 28, 198 and 231) are shown on the two plots. The scale-location plot, similar to the residuals versus fitted values, but it uses the square root of the standardized residuals. A good fit linear model should show roughly randomness in this plot. The last plot, residuals versus leverage, uses Cook's distance to identify points which have more influence than other points.

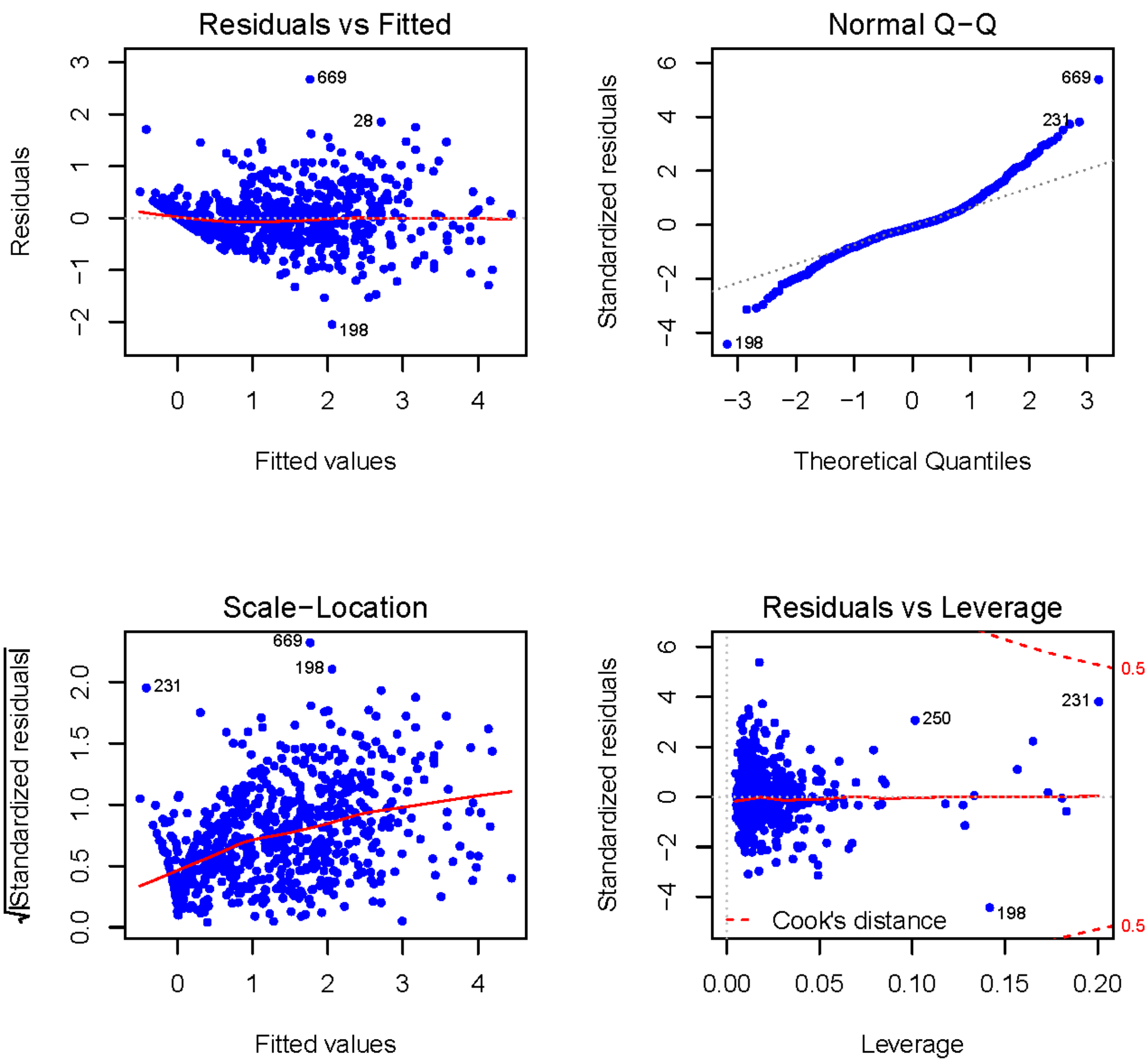


Figure 5.1 Diagnose Plot for OLS Regression Model

Generally, these are points that are distant from other points in the data, either for the dependent variable or one or more independent variables. Each observation is represented as a line whose height is indicative of the value of Cook's distance for that observation. There are no hard and fast rules for interpreting Cook's distance, but large values (which will be labeled with their observation numbers) represent points which might require further investigation.

5.2 Spatial Autocorrelation of the Variables

The first law of geography according to Waldo Tobler is “Everything is related to everything else, but near things are more related than distant things.”(Tobler, 1970) .This observation is embedded in the gravity model of trip distribution. It is also related to the law of demand, in that interactions between places are inversely proportional to the cost of travel, which is much like the probability of purchasing a good is inversely proportional to the cost. Spatial autocorrelation refers to the correlation of a variable with itself through space. If there is any systematic pattern in the spatial distribution of a variable, it is said to be spatially auto-correlated. Ordinary least squares (OLS) regressions assume that observations have been selected randomly. However, if the observations are spatially clustered in a certain degree, the estimates obtained from the correlation coefficient or OLS estimator will be biased and overly precise. The bias came from the areas with higher concentrations of events will have a greater impact on the model estimation and will overestimate precision since events tends to be concentrated, and therefore, there are actually fewer number of independent observations than that being assumed.

The most common measurement of spatial autocorrelation is the Moran’s autocorrelation coefficient (often denoted as I). It is an extension of Pearson-moment correlation coefficient to a univariate series(Cliff & Ord, 1973; Moran, 1950). Recall that Pearson’s correlation (denoted as ρ) between two variables x and y both of length n is:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i-1}^{n}(x_i - \bar{x})^2 \sum_{i-1}^{n}(y_i - \bar{y})^2\right]^{1/2}} \tag{5.1}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of both variables. ρ measures whether, on average, x_i and y_i are associated.

In the study of spatial patterns and processes, it is logically expected that close observations is more likely to be similar than those far apart. It is usual to associate a weight to each pair (x_i, x_j) which quantifies the above expectation (Cliff & Ord, 1981). In its simplest form, these weights will take values 1 for close neighbors, and 0 otherwise. The weights are sometimes referred to as a neighboring function with w_{ii} set to be 0. Moran's I can be interpreted as the correlation between variable, x, and the "spatial lag" of x formed by averaging all the values of x for the neighboring areal units (i.e. Polygons)

The Moran's autocorrelation coefficient I's measured by:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{5.2}$$

where w_{ij} is the weight between observation i and j , and S_0 is the sum of all w_{ij} :

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \tag{5.3}$$

The Moran's I varies on a scale between [-1, 1]. When the value is close to -1, it means high negative spatial autocorrelation; when the value is close to 0, it means no or minimal autocorrelation; when the value is close to 1, it suggests high positive spatial autocorrelation.

The null hypothesis is that the Spatial Autocorrelation (Moran's I) is that the data is completely spatial random. If the p-value is not statistically significant, the null hypothesis cannot be rejected. If the p-value is statistically significant, and the z-score is positive, the null hypothesis is rejected. Table 5.2 shows the Moran's I and its statistical testing results. Almost all the zonal attributes are determined as spatially dependent.

Table 5.2 Moran's I and its Spatial Dependency Check

Variables	Moran's I	P-Value	Z-Score	Null hypothesis	Spatially Dependent
AT	0.8705	0.0000	37.2544	Reject	Yes
AVGAUTO	0.7469	0.0000	29.0066	Reject	Yes
ACRES	0.5974	0.0000	23.6038	Reject	Yes
AVGWK	0.5387	0.0000	22.9845	Reject	Yes
EMP_DENSIT	0.5041	0.0000	21.4985	Reject	Yes
POP_DENSIT	0.4413	0.0000	18.5071	Reject	Yes
TOTAL_AUTO	0.2795	0.0000	12.4357	Reject	Yes
POP	0.2693	0.0018	11.1823	Reject	Yes
TOTAL_HH	0.2688	0.0002	11.4844	Reject	Yes
TOTAL_EMPL	0.2159	0.0000	10.0942	Reject	Yes
EMP_M	0.1874	0.0613	8.5768	Accept	No
Avg_TRIPSP	0.1803	0.0050	7.4908	Accept	No
Avg_CarbEM	0.1040	0.1040	5.2841	Accept	No

5.3 Spatial Regression Analysis Results

Table 5.3 Model Coefficients Comparison for OLS, SAR, SEM, SDM, SDEM, KPM and MAM Models shows the variable coefficients using the OLS, SAR, SEM and SDM.

Table 5.3 Model Coefficients Comparison for OLS, SAR, SEM, SDM, SDEM, KPM and MAM Models

Coefficients	OLS	SAR	SEM	SDM	SDEM	KPM	MAM
(Intercept)	-1.61E-01	-1.38E-01	-1.61E-01	-6.37E-02	-1.61E-01	-1.42E-01	-3.17E-02
ACRES	7.61E-05	5.66E-05	7.68E-05	2.17E-04	7.68E-05	7.37E-05	2.41E-04
AT	1.87E-01	1.93E-01	1.87E-01	3.00E-01	1.87E-01	1.88E-01	3.11E-01
POP	4.57E-04	4.59E-04	4.57E-04	5.50E-04	4.57E-04	4.75E-04	5.77E-04
TOTAL_HH	4.20E-04	4.68E-04	4.20E-04	6.05E-04	4.20E-04	4.65E-04	6.33E-04
TOTAL_EMPL	-3.78E-05	-4.22E-05	-3.78E-05	-5.08E-05	-3.78E-05	-4.19E-05	-4.96E-05

Coefficients	OLS	SAR	SEM	SDM	SDEM	KPM	MAM
POP_DENSIT	-2.13E-02	-2.18E-02	-2.13E-02	-2.55E-02	-2.13E-02	-2.21E-02	-2.68E-02
EMP_DENSIT	4.21E-04	3.78E-04	4.21E-04	3.75E-04	4.21E-04	3.72E-04	3.22E-04
TOTAL_AUTO	4.96E-04	4.78E-04	4.95E-04	2.56E-04	4.95E-04	4.56E-04	2.08E-04
EMP_M	2.40E-01	2.36E-01	2.40E-01	1.87E-01	2.40E-01	2.35E-01	1.83E-01
AVGWK	1.58E-01	1.70E-01	1.59E-01	1.91E-01	1.59E-01	1.74E-01	1.84E-01
AVGAUTO	-1.94E-01	-1.62E-01	-1.94E-01	-8.03E-02	-1.94E-01	-1.65E-01	-6.22E-02
Avg_CarbEM	-7.54E+01	-7.54E+01	-7.54E+01	-7.58E+01	-7.54E+01	-7.47E+01	-7.46E+01
Avg_TRIPSP	1.99E-02	2.02E-02	1.99E-02	1.98E-02	1.99E-02	1.99E-02	1.96E-0

5.4 Measurement of Effectiveness (MOE)

The goodness of fit measures in spatial regression models is slightly more complex due to the lack of standard measures such as the R^2. However, a common goodness of fit measures is the information-based measures. The information-based goodness of fit measures utilizes several model performance measures and ranks based on the values. The model with the lowest rank is considered a better fit than others. Table 5.4 shows the information based measures and their ranks for OLS, SAR, SEM, SDM, SDEM, KPM and MAM models. This ranking utilized AIC, Log Likelihood and Moran's I on Residuals as measures. For all three criteria, smaller values are better. Therefore, the SDEM model has the lowest summation of ranks and it fits the data better.

Table 5.4 Information-based Measure of Fit for Spatial Models

Model Type	AIC	Rank	Log Likelihood	Rank	Moran's I on Residuals	Rank	Total Rank
Ordinary Least Square ***(OLS)***	1092.5	4	-533.2	5	+0.026397503	5	14
Spatial Autoregressive Model ***(SAR)***	1088.4	3	-530.2	3	+0.056595454	6	12
Spatial Error Model ***(SEM)***	1093.1	5	-532.6	4	-0.000634720	1	10

Model Type	AIC	Rank	Log Likelihood	Rank	Moran's I on Residuals	Rank	Total Rank
Spatial Durbin Model ***(SDM)***	1065.3	2	-507.7	2	-0.008763889	4	8
Spatial Durbin Error Model ***(SDEM)***	1064.1	**1**	-532.6	**4**	-0.000634720	**1**	**6**
Kelejian-Prucha Model ***(KPM)***	N/A	3.5	N/A	3.5	-0.000340069	2	9
Manski Model ***(MAM)***	1093.8	5	-506.4	1	-0.0052871289	3	9

Figure 5.2 to Figure 5.8 illustrates how the residual from each model are quantitatively distributed. The residuals are divided into the following five categories and shown on the maps: [-2.5, -1.5), [-1.5, -0.5), [0.5, 0.5), [0.5, 1.5), [1.5, 2.5) in the unit of tons per day. It is observed that the SDEM model has smaller residuals comparing to the OLS, SAR, SEM KPM and MAM models.

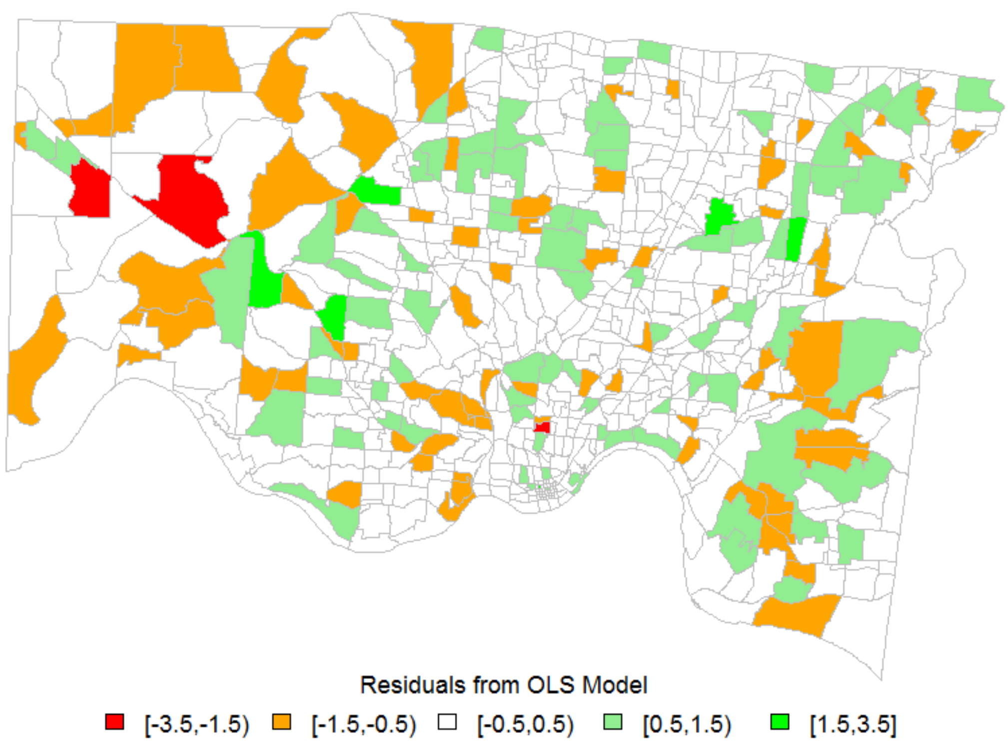


Figure 5.2 Residual Map for OLS Model

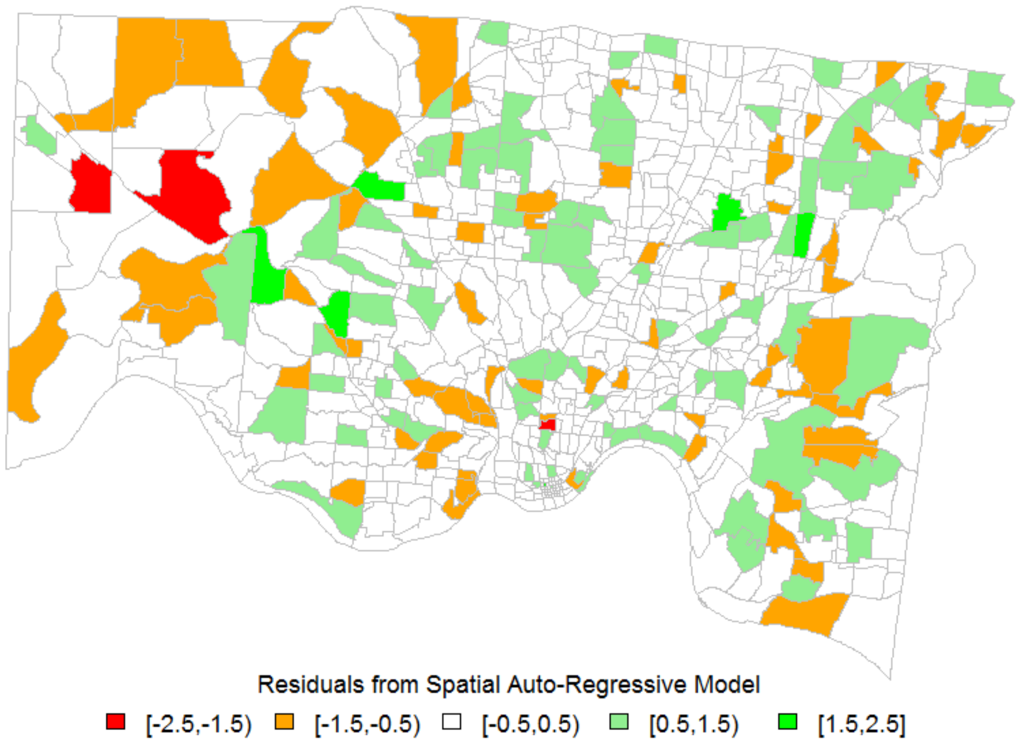


Figure 5.3 Residual Map for SAR Model

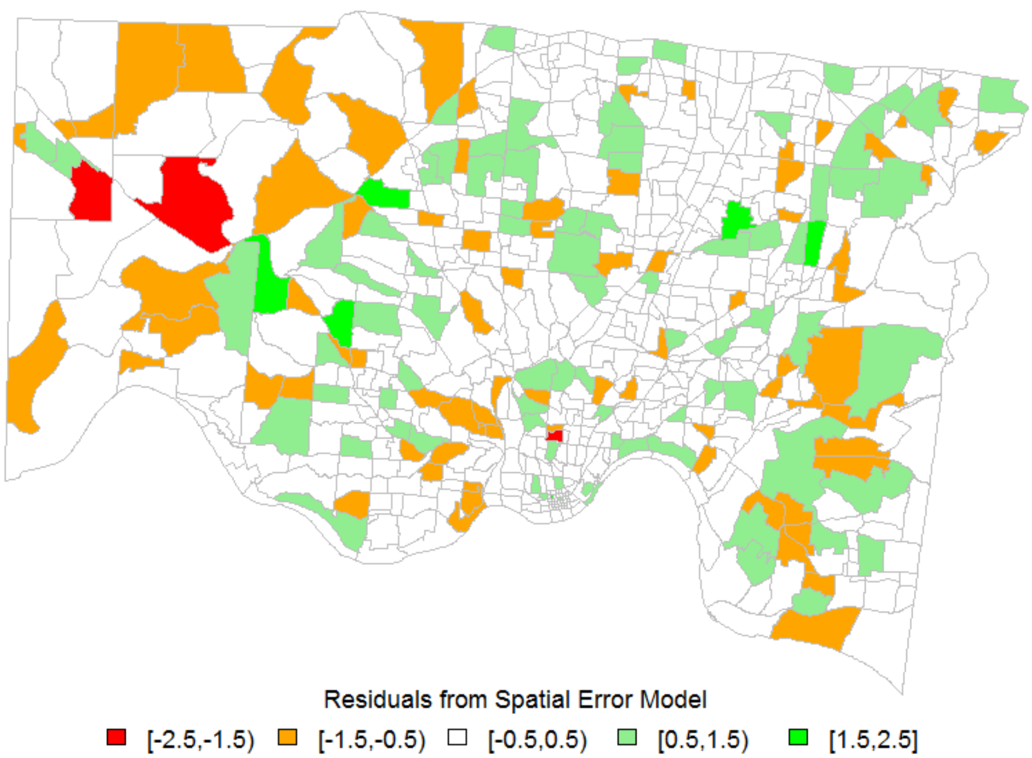


Figure 5.4 Residual Map for SEM Model

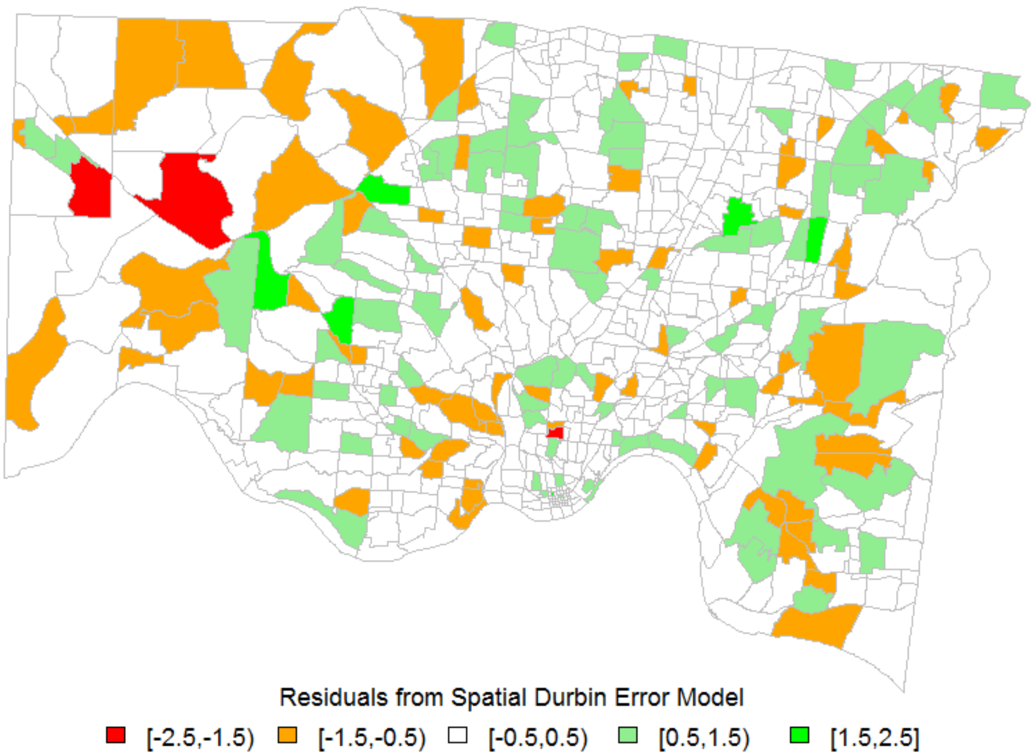


Figure 5.5 Residual Map for SDM Model

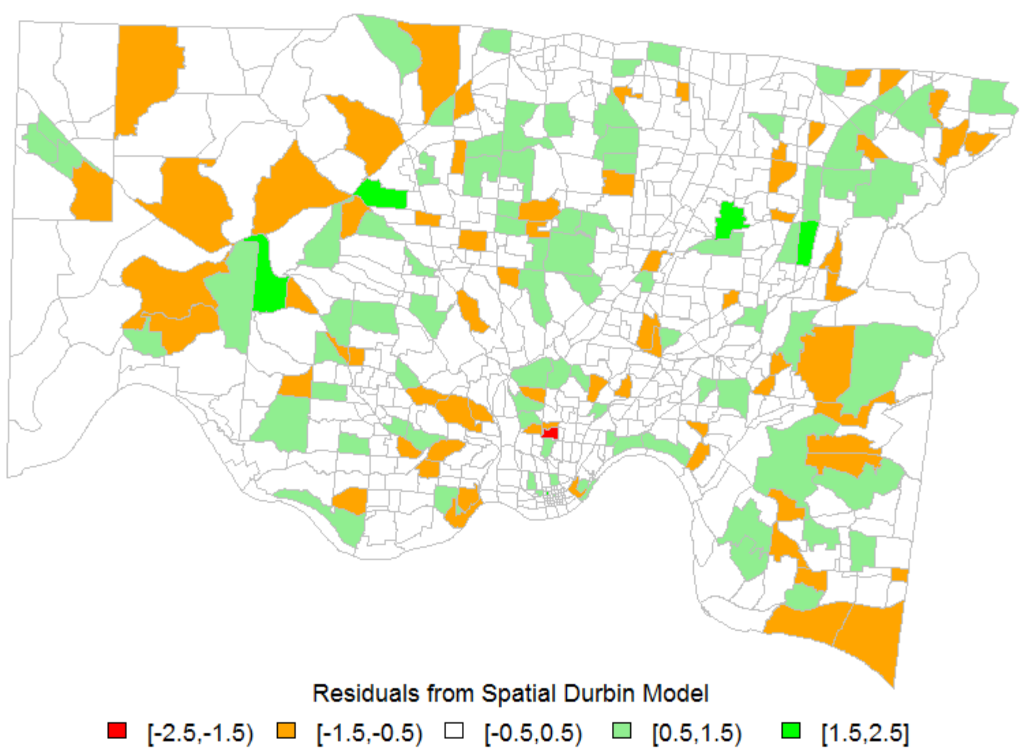


Figure 5.6 Residual Map for SDEM

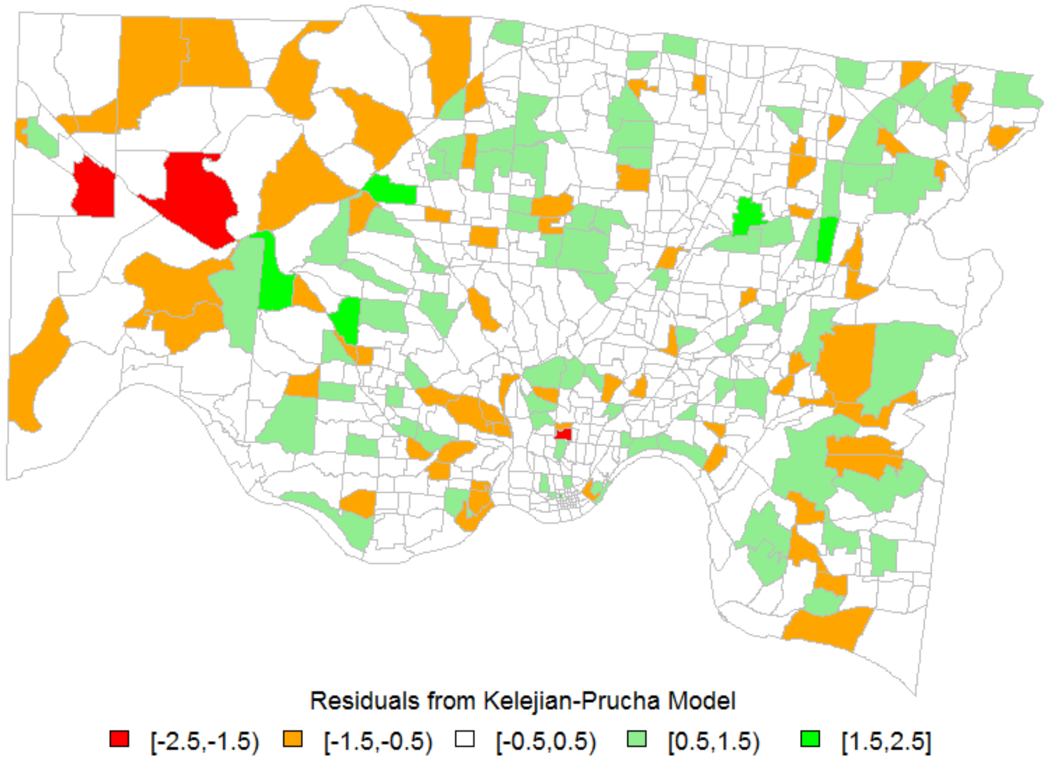


Figure 5.7 Residual Map for Kelejian-Prucha Model

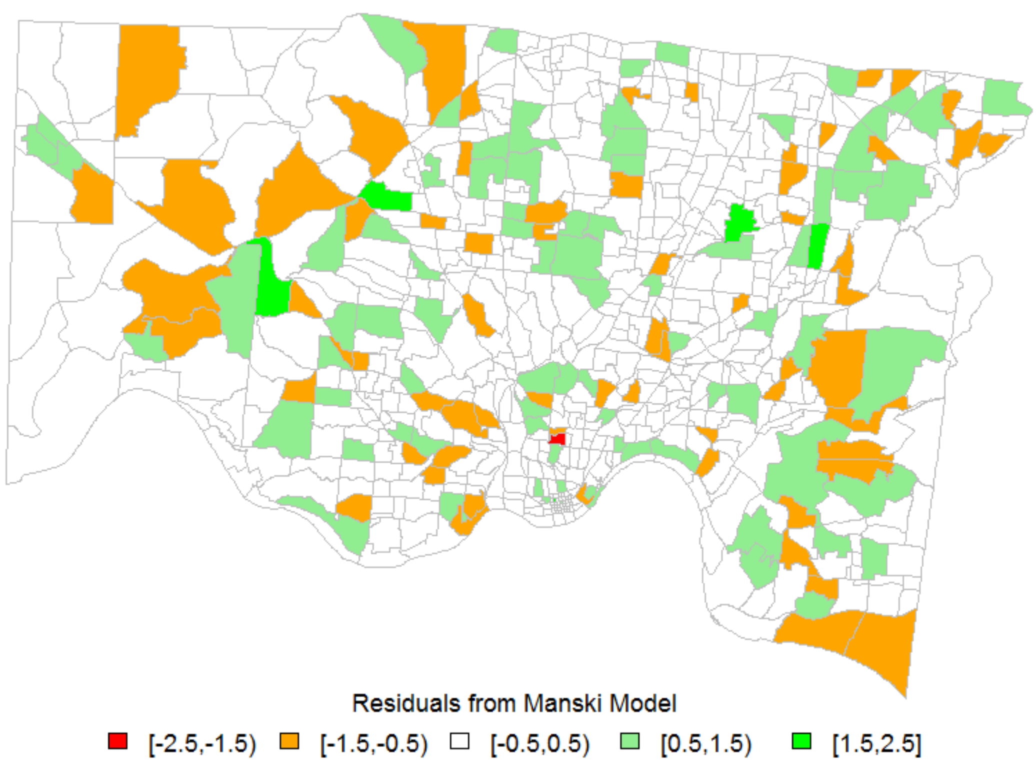


Figure 5.8 Residual Map for Manski Model

CHAPTER 6 SCENARIO TESTING AND SENSITIVITY ANALYSIS

Sensitivity analysis is helpful in determining how “sensitive” a model is to changes in the value of the parameters. Sensitivity analysis is used to build confidence in the model by studying the uncertainties that are often associated with parameters. It is wildly used after many models, from relatively simple linear regression models to sophisticated activity-based travel demand models, were established. The purpose of sensitivity analysis tests on the responsiveness of the spatial regression model to changes in selected input variables is to determine the level of impact. It is interesting to understand which is the most contributing variable of regional carbon emission and to what magnitude the impacts can be. The responsiveness, or sensitivity, of the model to changes in key inputs indicates whether the model can reasonably estimate the anticipated change in carbon emissions resulting from the changes in the land use characteristics. This analysis usually assumes one input variable change at a time and examines the range of output change.

6.1 Scenario Testing

The level of specificity, such as the land use change, and carbon emission analysis presented in this study enables more data and indicators to be developed for a given land use scenario. Such data and indicators can be incorporated into decision makers’ plans, policies, and ultimately, regulations, and its possible integration with project level environmental review processes. To further test the applicability of the developed model, a set of scenarios was developed to test if the model generates reasonable results. Two scenarios were designed with an assumed percentage of increase and decrease of average household density to mimic the dense and sparse development pattern. The scenarios preserved the total amount of household of 333,984 in Hamilton County for 2010 but spatially redistributed the households among the TAZs in order to achieve the changes. The scenario TAZ data were then plugged into the developed SDEM model and the total regional CO_2 emission is calculated. The scenarios and results are summarized in below table (Table 6.1):

Table 6.1 Summary of Scenario Testing Results

Scenarios	Household Density (HH per Acre)	Percentage Change from Base	CO_2 Emissions (Tons/day)	Percentage Change from Base
Dense Development	2.90	22.8%	872.47	-3.48%
Base	2.37	0.00%	904	0.00%
Sparse Development	1.86	-21.24%	938.42	3.80%

6.2 One-at-a-time Sensitivity Analysis

The Spatial Durbin Error model developed in chapter 5 is used to determine the model sensitivity. The numerical values of each dependent variable of the model were reduced or increased by 25 and 50 percent. The TAZ size in acres and area type have not been included in the sensitivity tests since change the values is almost equals to change the analytical scope of the study. Table 6.2 summarizes the results of the sensitivity analysis. The baseline regional carbon emission totals 904 tons per day. The minimum-modeled regional carbon emission is 721 tons, which is a 20.24% reduction when population reduced to 50 percent. The maximum-modeled regional carbon emission is 1,088 tons, which is a 20.35% increase when population increases 50 percent.

Table 6.2 Modeled Regional Household Travel Carbon Emissions (Tons/day)

Variables	-50%	-25%	Baseline	+25%	+50%
POP	721	813	904	996	1,088
TOTAL_HH	830	867	904	942	979
TOTAL_EMPL	914	909	904	899	894
POP_DENSIT	945	924	904	884	864
EMP_DENSIT	902	903	904	906	907
TOTAL_AUTO	764	834	904	975	1,045
EMP_M	855	880	904	929	953
AVGWK	845	875	904	934	964
AVGAUTO	1,009	957	904	852	800

Variables	-50%	-25%	Baseline	+25%	+50%
Avg_CarbEM	950	927	904	882	859
Avg_TRIPSP	811	857	904	951	998

Based on the summary of sensitivity test results, it is determined that population, total automobiles, average trip speed, total household number, average household worker, employment with medium trip rates and employment density (ordered from the largest impact to the smallest impact) are all positively related with regional household travel carbon emissions. Zonal level average auto ownership, average zonal carbon emission, population density, total employment and employment density (ordered from the largest impact to the smallest impact) are posing negative impacts on regional household travel carbon emissions.

6.3 Elasticities on Dependent Variables

Elasticity is widely used in many occasions where an estimation of the impact of changes in policy oriented model input changes. The concept of elasticity is usually applied in VMT reduction and other co-benefits while presenting the impact of changes in variables. In this dissertation, the elasticity is defined as the percentage change in regional household travel carbon emission divided by the percentage change in its corresponding variable. To be more specific, it is a measure impact of one percent change in input variable.

Table 6.3 shows the calculated elasticities of the developed spatial regression model. The elasticities range from -0.23 to 0.41, which is a reasonable range, since most literatures reporting the elasticities should be in between -1 and +1.

Table 6.3 Elasticities of Household Travel Carbon Emission Generation Model Variables

Variable	Elasticity
POP	0.41
TOTAL_AUTO	0.31
Avg_TRIPSP	0.21

Variable	Elasticity
TOTAL_HH	0.17
AVGWK	0.13
EMP_M	0.11
EMP_DENSIT	0.01
TOTAL_EMPL	-0.02
POP_DENSIT	-0.09
Avg_CarbEM	-0.10
AVGAUTO	-0.23

Figure 6.1 shows a ranking of the elasticities or impacts of the variables used in developing the spatial regression model. Population in a TAZ has the largest positive impact (0.41) on the regional household travel carbon emission. The variables having fewer impacts are total auto (0.31), average trip speed (0.21), total household (0.17), average worker (0.13), employment with medium trip rates (0.11), and employment density has the lowest impact (0.01) on the regional level. On the negative side, average household auto has the largest impact (-0.23), then the average zonal carbon emission (-0.10), population density (-0.09), and total employment (-0.02). This result provides a reference of which variable is more critical to regional household carbon emission reduction.

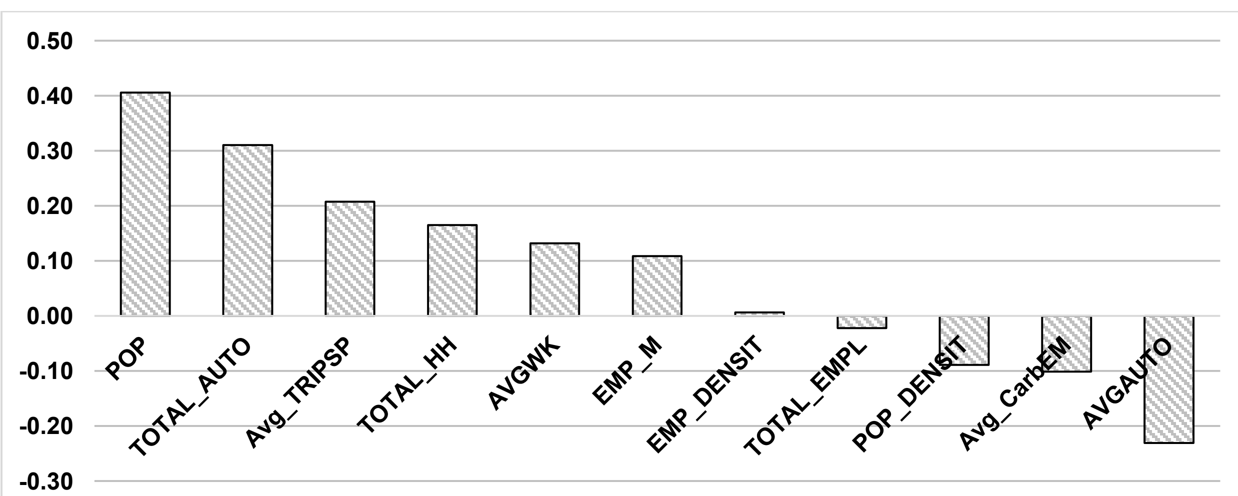


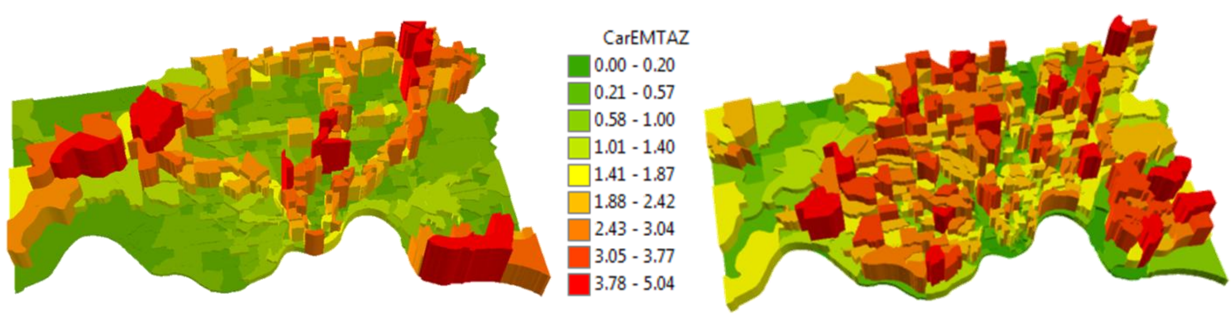
Figure 6.1 Elasticities Ranking of the Household Travel Carbon Emission Model

CHAPTER 7 CONCLUSIONS

Most cities and metropolitan areas these days confronting significant transportation related challenges, due to the increased population and travel demand. Those challenges include severe traffic-related air pollutions, unsustainable energy and resources consumption, excess recurrent and non-recurrent congestion, and increased accident risks. These challenges require a countermeasure to address the sustainability, urbanization, competitiveness, and mobility issues. This study has provided a proof-of-concept methodology, which is capable of quantifying land use change impacts on carbon emission. The proposed methodology has the capability to reveal the dynamic linkage between land use, transportation and emissions.

7.1 Discussion of the Modeling Framework

This proof-of-concept study testing the Cincinnati GPS HTS simulated smart data shows positive in support of the research question. The study demonstrates that the proposed modeling framework is capable of revealing the dynamic linkage between land use, transportation, and emissions thru scenario-based analysis.



Results from link-based traditional approach where the large emissions fall into the TAZs along freeways.

Results from the proposed TAZ-based approach shows which TAZ actually generates higher carbon emissions.

Figure 7.1 A Visual Comparison of Link-based and TAZ-based Approach.

The findings from this research provide a practical tool for the analysis of land uses visioning and planning alternatives. Especially, it enables a quick response analysis built on adaptive policies, enforced development regulations, associated travel patterns and their sequential carbon emissions. The proposed modeling framework can provide a TAZ level emission quantification, which is not possible to get from conventional modeling practice, as illustrated by Figure 7.1. The figure on the left shows the link-based traditional approach using the travel demand model to estimate the average speed and plug in the results of emission models. Once the link-based emissions being aggregated into TAZs, it is obvious that the TAZs among the I-71, I-75 and I-275 are showing high CO_2 emissions. However, this information is not helpful to planners since it is impossible to tell the source of high amount emissions. The figure on the right shows the result from the proposed TAZ-based approach. It shows the TAZs with higher amount of CO_2 emissions clearly.

The findings from this research provide insights on how land-uses planning alternatives built on adopted policies and enforced development regulations correlate with travel patterns and their sequential carbon emissions. Specifically, this method is capable of providing a TAZ level emission quantification approach, which is not possible from conventional modeling practice. The level of specificity, such as the land use change, and carbon emission analysis presented in this study enables more data and indicators to be developed. Such data and indicators can be incorporated into decision makers' plans, policies, and ultimately regulations and its possible integration with project level review processes. The findings from previous research provide insights on how land-uses planning alternatives built on adopted policies and enforced development regulations correlate with travel patterns and their sequential carbon emissions. The level of specificity, such as the land use change, and carbon emission analysis presented in this study enables more data and indicators to be developed. Such data and indicators can be incorporated into decision makers' plans, policies, and ultimately regulations and its possible integration with project level review processes.

While the results from this work offer specific recommendations as to which types of land use planning policy practices are most highly associated with higher amount of VMT, carbon emissions, there are also some potential to reveal policy impacts that can be applied to integrated land use and transportation sustainability practices. The proposed modeling framework is capable of unveil the dynamics between land use changes and related carbon emissions. Additional policy and planning scenarios could be tested and provides well informed smart data driven adaptive planning for sustainable development.

Although the results maybe pertaining to the specific dataset but it helps transportation decision makers to better connects the land use development and its related household socioeconomics with their carbon emission characteristics. Particular, the household travel carbon emission footprint quantification results made its contribution to current body of knowledge on the following: (1) provides accurate carbon emission results by using the best available traffic activity data inputs (VSP distributions) for emission modeling; (2) provides connections between household socioeconomics and their travel carbon footprint. The results showed using the cross-classification method is likely to use as carbon emission generation rates for the purpose of rapidly estimate household travel carbon footprint. Furthermore, the results from inter-life cycle differences further characterized the carbon footprints of the adult, adult student, retiree and households with children. The research suggests an important potential to provide solid grounds for analyzing, modeling of sustainable community strategies, adaptive planning policies etc.

7.2 Conclusions on Smart Data Applications

It is a proven success of using the Smart Data to modeling the sophisticated urban phenomenon, which is the land use, household travel and associated carbon emissions. The proposed framework utilizes the location-rich information to capture the vehicle fleet's trajectories at a very fine scale.

- It is critical to obtain the vehicle trajectory data at a very high time resolution. Preferably, this time resolution can be second-by-second. Thus, the modeling capacity of VSP-based MOVES model could be maximized.
- Smart Data is often massive and real-time. It requires advanced computing skills and database properties such as ACID (Atomicity, Consistency, Isolation, Durability) to assure the database transactions.
- Smart Data, once combined with other data such as socioeconomic and demographics, is capable of developing smart mobility for traffic operation efficiency and safety enhancement, data-driven public transportation, smart demand management for passengers and freight, etc. The potential applications are just endless.

7.3 Conclusions on Household Travel Carbon Emissions

This carbon emission rates calculated from the high resolution (second-by-second) GPS data set out empirical results from the best available household travel carbon emission and bridged the household socioeconomics. Carbon emission generation rates of households by area type, number of workers, life cycle, and income level provided grounds for estimating regional household travel carbon emission.

- The average household travel carbon emission for the Cincinnati region is 0.00291 ton per day, with median of 0.00255, minimum of 8.03E-7, maximum 0.01853 tons per day.
- Carbon emission generation rates by area type, number of workers, life cycle, and income level is similar to traditional trip generation rates. The carbon emission rates can be fit well into regression lines of polynomial functions. The R^2 ranges from 0.69 to 1.

7.4 Conclusions on Spatial Regression Model

The spatial regression-based model was developed based on finding the minimal model residuals and multiple information-based measure of fit. The goodness of fit measures in spatial regression models is slightly more complex due to the lack of standard measures such as the R^2. However, a common goodness of fit measures is the information-based measures. The information-based goodness of fit measures utilizes several model performance measures and ranks based on the values. The model with the lowest rank is considered a better fit than others. The information based measures and their ranks for OLS, SAR, SEM, SDM, SDEM, KPM and MAM models are summarized and presented. Some findings are summarized below:

- OLS model has a R^2 (coefficient of determination) of 0.8, which is a good fit. However, when examining the residuals on diagnosis plots, it was find that the residuals are still spatially correlated. This suggest that spatial models can fit the data better and reduce the residual spatial correlation.
- After performing spatial regressions, the information-based measure of fit based on AIC, log likelihood and Moran's I on residuals are compared and the best model fitting the given dataset is the Spatial Durbin Error Model. The SDEM has the lowest AIC and Moran's I on residuals compared to other candidate models. Residual maps (Figure 5.2-Figure 5.8) from the candidate models also confirm that SDEM has the minimal residuals.

7.5 Conclusions on Sensitivity/Elasticity of the Model

The responsiveness, or sensitivity, of the model to changes in key inputs indicates whether the model can reasonably estimate the anticipated change in carbon emissions resulting from the changes in the land use characteristics. And the dynamics drawn from the sensitivity test of the household travel carbon emission generation model can be concluded as follows:

- Population is the most influential variable in regional household travel carbon emissions. The impact of population is positive on the household travel carbon emissions. The calculated elasticity is 0.41, which means one percent change in population will cause 0.41 percent change in regional level carbon emissions.
- The variables having fewer impacts are total auto (0.31), average trip speed (0.21), total household (0.17), average worker (0.13), employment with medium trip rates (0.11), and employment density has the lowest impact (0.01) on the regional level.
- On the negative side, average household auto has the largest impact (-0.23), then the average zonal carbon emission (-0.10), population density (-0.09), and total employment (-0.02).

This research could benefit from completing the proposed modeling framework with testing case study is to apply it in a new context, location and/or culture. There is an urgent need for a quick-response tool for quantifying household travel carbon emissions in California for both state agencies and MPOs. Given the proposed modeling framework is data dependent and data adaptive, it will be a continuous work to tracking the new emergence of smart data and incorporate the data structure changes. The modeling framework shall be reviewed and new constructs should be added after each testing and modeling to assure its capability of adapting the changes in data. Another future direction is to incorporate latest theoretical development in spatial econometric in the modeling framework.

GLOSSARY

Air quality planning

The process by which state, and in some cases, regional, air quality planning agencies assess current and future air quality conditions and determine the "control strategies needed to reduce emissions and improve air quality. These agencies prepare State Implementation Plans (SIPs) and submit them to EPA for approval.

Transportation planning

The process by which state and local transportation agencies, along with metropolitan planning organizations, assess needs for future transportation infrastructure such as roads and transit systems. Federal regulations require states to demonstrate that planned transportation activities are consistent with or "conform" to the air goals outlined in the SIP.

Land use planning

The process by which local governments plan for future growth in communities and decide where and how development should occur within local boundaries. In some cases, regional planning agencies work with local governments to coordinate the planning efforts of neighboring municipalities.

Land use activities

Land use activities include all of the various actions that state and local governments or other entities take which affect the development of land use in a community or region. These land use activities result in patterns of land use that influence the transportation choices people make. In this guidance, land use activities that reduce reliance on motor vehicles (e.g., through shortening trip lengths or increasing accessibility of alternative modes of transportation) and that can also be shown to have air quality benefits may be accounted for in the air quality and transportation planning processes.

Land use activities include **land use policies**, defined as specific policies, programs, or regulations adopted or operated by government agencies and **land use projects** defined as specific developments.

Accounting for land use in the air quality and transportation processes

Where planning agencies can demonstrate, through modeling, that land use activities can be reasonably expected to have a positive impact on air quality, they can account for those benefits in the air quality and/or the transportation planning process.

Trip Generation

The trip generation step uses the land use assumptions to estimate the number of trip ends (productions and attractions) for each zone. The trips are generated by trip type, such as "home-based work," "home-based other" or "non-home based."

Trip Distribution

The trip distribution step links the productions with the attractions. Demand for travel between two zones is related to the number of trips in and out of the zone, and the amount of impedance (i.e., the effect of time, distance, and/or cost on travel activity).

Modal Choice

In some areas, the travel demand model also produces estimates of trips by mode (e.g., highway, transit, or other modes). Mode choice models may take into consideration factors such as demographic group, cost, trip purpose, and relative travel times.

Trip Assignment

Trip assignment involves assigning vehicle trips to specific links of the travel network. Travel demand models also estimate the speeds that vehicles travel, based on how congested the road network is.

REFERENCES

Anderson, R. & Giaimo, G., 2016. *First Large-Scale Smartphone-Based Household Travel Survey in the United States.* Washington D.C., Transportation Research Board.

Bachman, W., Sarasua, W., Hallmark, S. & Guensler, R., 2000. Modeling regional mobile source emissions in a geographic information system framework. *Transportation Research Part C: Emerging Technologies,* December, 8(1-6), pp. 205-229.

Bhat, C., Conoor, S. & Nerella, S., 2003. *GIS-Integrated Traffic Input Models for Mobile-Source Emissions Analysis.,* Austin: Transportation Research Board.

Bhat, C. & Huimin, Z., 2002. The spatial analysis of activity stop generation. *Transportation Research Part B: Methodological ,* 36(2), pp. 557-575.

Bhat, C. R., Sen, S. & Eluru, N., 2009. The Impact of Demographics, Built Environment Attributes, Vehicle Characteristics, and Gasoline Prices on Household Vehicle Holdings and Use. *Transportation Research Part B: Methodological,* 43(1), pp. 1-18.

Bhat, C. & Sen, S., 2006. Household Vehicle Type Holdings and Usage: An Application of the Multiple Discrete-Continuous Extreme Value (MDCEV) Model. *Transportation Research Part B: Methodological,* 40(1), pp. 35-53.

Bowman, J. & Ben-Akiva, M., 2000. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A,* jANUARY, 35(1), p. 1–28.

Bureau of Transportation Statistics, 2012. *National Average Fuel Efficiency of U.S. Light Duty Vehicles,* Washington D.C.: Research and Innovative Technology Administration.

California Air Resources Board, 2016. *Sustainable Communities.* [Online] Available at: http://www.arb.ca.gov/cc/sb375/sb375.htm [Accessed 31 12 2014].

Carbon Footprint Ltd, 2013. *A Breath of Fresh Air for Carbon Management.* [Online] Available at: http://www.carbonfootprint.com/ [Accessed 10 January 2013].

Carbon Trust Empower, 2013. *Carbon Trust Empower.* [Online] Available at: http://www.carbontrust.com/resources/tools/empower [Accessed 10 January 2013].

Cervero, R., 1996. Mixed Land-uses and Commuting: Evidence from the American Housing Survey. *Transportation Research Part A: Policy and Practice,* September, 30(5), pp. 361-377.

Cervero, R. & Kockelman, K., 1997. Travel Demand and the 3Ds: Density, Densign, Diversity. *Transportation Research Part D: Transport and Environment,* 2(3), pp. 199-219.

Cervero, R. & Kockelman, K., 1997. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment,* September, 2(3), pp. 199-219.

Cervero, R. & Landis, J., 1995. The Transportation-Land Use Connection Still Matters. *Access,* Volume 7, pp. 2-10.

Chakir, R. & Le Gallo, J., 2013. Predicting land use allocation in France: A spatial panel data analysis. *Ecological Economics,* August, Volume 92, pp. 114-125.

Cliff, A. D. & Ord, J., 1973. *Spatial autocorrelation.* 1st ed. London: Pion.

Cliff, A. & Ord, J., 1981. Spatial and temporal analysis: autocorrelation in space and time. In: E. Wrigley & R. Bennett, eds. *Quantitative Geography: A British View.* London: Routledge & Kegan Paul, pp. 104-110.

Corporate Partnership Board, 2015. *Big Data and Transport: Understanding and assessing options,* s.l.: International Transport Forum.

Duthie, J., Kockelman, K., Valsaraj, V. & Zhou, B., 2007. *Applications of Integrated Models of Land Use and Transport: A Comparison of ITLUP and UrbanSim Land Use Models.* Savannah, Georgia., Transportation Research Board.

Elhorst, J., 2010. Spatial Panel Model. In: M. Fischer & A. Getis, eds. *Handbook of Applied Spatial Analysis.* Berlin: Springer Berlin Heidelberg, pp. 377-407.

Federal Highway Administration, 2009. *Summary of Travel Trends: 2009 National Household Travel Survey,* Washington D.C.: FHWA.

FHWA, 2010. *Transportation Conformity: A Basic Guide for State and Local Officials.,* Washington,D.C.: Department of Transportation .

Frank, L., Stone, B. & Bachman, W., 2000. Linking Land Use with Household Vehicle Emissions in the Central Puget Sound: Methodological Framework and Findings. *Transportation Research Part D: Transport and Environment,* May, 5(3), pp. 173-196.

Frey, C., Rouphail, N. & Zhai, H., 2008. Link-Based Emission Factors for Heavy-Duty Diesel Trucks Based on Real-World Data. *Transportation Research Record: Journal of the Transportation Research Board,* November, 2058(4), pp. 23-32.

Frey, H., Rouphail, N. & Zhai, H., 2006. Speed- and Facility-Specific Emission Estimates for On-Road Light-Duty Vehicles on the Basis of Real-World Speed Profiles. *Transportation Research Record: Journal of the Transportation Research Board,* Volume 1987, pp. 128-137.

Fulper, C. et al., 2011. *Development of real-world data for MOVES Development: The Houston Drayage Activity Characterization Study.* Washington D.C., Transportation Research Annual Meeting.

Giaimo, G., Anderson, R., Wargelin, L. & Stopher, P., 2010. Will It Work? Pilot Results from First Large-Scale Global Positioning System-Based Household Travel Survey in the United States. *Transportation Research Record: Journal of the Transportation Research Board,* 2176(1), pp. 26-34.

Gregor, B., 2007. Land Use Scenario DevelopeR: Practical Land Use Model Using a Stochastic Microsimulation Framework. *Transportation Research Record,* 2003(1), pp. 93-102.

Handy, S., Cao, X. & Patricia, M., 2005. Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment,* 10(6), pp. 427-444.

Hatzopoulou, M. M. E. J., 2010. Linking an activity-based travel demand model with traffic emission and dispersion models: Transport's contribution to air pollution in Toronto.. *Transportation Research Part D: Transport and Environment,,* pp. 15(6), 315.

Helwagen, J., 2015. *Population Analytics: Innovations in Data Collection.* [Online] Available at: http://www.airsage.com/Population-Analytics/ [Accessed 12 December 2015].

Hunt, J. & Abraham, J., 2003. *Design and Application of the PECAS Land Use Modelling System.* Sendai, Japan., Transportation Research Board.

Jiménez-Palacios, J., 1999. *Understanding and Quantifying Motor Vehicle Emissions with Vehicle Specific Power and TILDAS Remote Sensing,* Cambridge, Mass: MIT.

Kamal-Chaoui, L. & Robert, A., 2009. *Competitive Cities and Climate Change,* s.l.: OECD Regional Development Working Papers No 2.

Kockelman, K. & Frazier, C., 2006. Spatial Econometric Models for Panel Data: Incorporating Spatial and Temporal Data. *Transportation Research Record: Journal of the Transportation Research Board,* Volume 1902, pp. 80-90.

Kockelman, K. & Frazier, C., 2006. Spatial Econometric Models for Panel Data : Incorporating Spatial and Temporal Data. *Transportation Research Record: Journal of the Transportation Research Board,* December, 1902(2005), pp. 80-90.

Lee-Gosselin, M. & Doherty, S. eds., 2005. *Integrated Land Use and Transportation Models: Behavioral Foundations.* Boston: Elsevier.

Lindsey, M., Schofer, J., Durango-Cohen, P. & Gray, K., 2011. The effect of residential location on vehicle miles of travel, energy consumption and greenhouse gas emissions: Chicago case study. *Transportation Research Part D: Transport and Environment,* 16(1), pp. 1-9.

Liu, C. & Shen, Q., 2011. An empirical analysis of the influence of urban form on household travel and energy consumption. *Computers, Environment and Urban Systems,* 35(5), pp. 347-357.

Liu, H., Danczyk, A., Brewer, R. & Starr, R., 2008. Evaluation of Cell Phone Traffic Data in Minnesota. *Transportation Research Record: Journal of the Transportation Research Board ,* Volume 2086, pp. 1-7.

Lopes, S., Brondino, N. & Da Silva, A., 2014. GIS-Based Analytical Tools for Transport Planning: Spatial Regression Models for Transportation Demand Forecast. *International Journal of Geo-Information,* Issue 3, pp. 555-583.

Manski, C., 1993. Identification of endogenous social effects: the reflection problem. *Review of Economic Studies,* Volume 60, pp. 531-542.

Ma, X. & Lo, H. K., 2015. Adaptive Transport Supply and Demand Management Strategies in an Integrated Land Use and Transport Model. *Transportation Research Record: Journal of the Transportation Research Board,* Volume 2, pp. 11-20.

Miller, H. J., 1999. Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T).. *Geographical Analysis,* 31(4), pp. 373-399.

Moran, P., 1950. Notes on continuous stochastic phenomena. *Biometrika,* Volume 37, pp. 17-23.

Neuman, M. & Smith, S., 2010. City Planning and Infrastructure: Once and Future Partners. *Journal of Planning History,* February, 9(1), pp. 21-42.

Newman, P. & Kenworthy, J., 1988. The Transport Energy Trade-Off: Fuel-Efficient Traffic versus Fuel Efficient Cities. *Transportation Research A: Policy and Practice,* 22-A(3), pp. 163-174.

Nittal, N. et al., 2004. *Estimating Truck Volumes on State Highways – A Statistical Approach.* Washington D.C., Transportation Research Board Annual Meeting.

Parent, O. & LeSage, J., 2010. A spatial dynamic panel model with random effects applied to commuting times. *Transportation Research Part B: Methodological,* 44(5), pp. 633-645.

Parsons Brinckerhoff Quade and Douglas, 1999. *Land Use Impacts of Transportation - A Guidebook,* Washington, D.C.: Transportation Research Board.

Perugu, H., Wei, H. & Rohne, A., 2012. Modeling Roadway Link PM2.5 Emissions with Accurate Truck Activity Estimate for Regional Transportation Conformity Analysis. *Transportation Research Record: Journal of the Transportation Research Board,* October, Volume 2270, pp. 87-95.

Perugu, H., Wei, H. & Rohne, A., 2012. Modeling Roadway Link PM2.5 Emissions with Accurate Truck Activity Estimate for Regional Transportation Conformity Analysis. *Transportation Research Record: Journal of the Transportation Research Board,* October, Volume 2270, pp. 87-95.

Perugu, H., Wei, H. & Yao, Z., 2013. *An Improved Methodology for Modeling Truck Contribution to Regional Air Quality.* Columbus, Transportation Research Board Annual Meeting.

Perugu, H., Wei, H. & Yao, Z., 2014. *Modeling Truck Activity using Short-term Traffic Counts for Reliable Estimation of Heavy-Duty Truck Emissions in Urban Areas.* Washington D.C., Transportation Research Board Annual Meeting, pp. 14-4350.

Rahman , M. & Habib, M., 2015. Spatial Transferability of a Micro Residential Mobility Model for the Integrated Land Use , Transport and Environment Modeling System. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 902, pp. 1-17.

Reddy, S. et al., 2010. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks,* 6(2), pp. 1-27.

Redfearn, C., 2007. The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area. *Journal of Urban Economics,* May, 61(3), pp. 519-541.

Rocky Mountain Institute, 2013. *Green Footstep:An assessment tool for reducing carbon emissions from building construction projects.* [Online] Available at: http://www.greenfootstep.org/ [Accessed 10 January 2013].

Roth, C., Kang, S. M., Batty, M. & Barthelemy, M., 2011. Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *PLoS ONE 6(1): e15923,* January, 6(1), pp. 1-8.

Sadek, A., Wang, Q., Su, P. & Tracy, A., 2011. *Reducing Vehicle Miles Traveled Through Smart Land-use Design,* Albany: New York State Department of Transportaton.

Shin, D., Aliaga, D. & Turner, B., 2015. Urban sensing: Using smartphones for transportation mode classification. *Computers, Environment and Urban Systems,* Volume 53, pp. 76-86.

Smith, C., 2013. *Smart Growth Area Planning (SmartGAP) Tool: Pilot Test for Thurston County, Washington.* Kansas City, Missouri., Resource Systems Group.

Song, G., Yu, L. & Wang, Z., 2008. *A Practical Modeling Approach for Evaluation of Fuel Efficiency for Road Traffic.* Washington D.C., Transportation Research Board Annual Meeting.

Stopher, P. & Wargelin, L., 2012. *GPS-Based Household Interview Survey for the Cincinnati, Ohio Region,* Columbus: Ohio Department of Transportation.

Stopher, P. & Wargelin, L., 2012. *GPS-Based Household Interview Survey for the Cincinnati, Ohio Region,* Columbus: FHWA.

Stopher, P. et al., 2012. *GPS-Based Household Interview Survey for the Cincinnati, Ohio Region,* Columbus, Ohio.: RSG.

Tobler, W., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography,* 46(2), pp. 234-240.

Transportation Research Board, 2009. *Driving and the Built Environment: The Effects of Compact Development on Motorized Travel,* Washington D.C.: National Academy of Science.

U.S. EPA, 2010. *Greenhouse Gas Emissions from a Typical Passenger Vehicle,* Washington D.C.: United States Environmental Protection Agency.

USEPA, 2010. *Transportation Conformity Guidance for Quantitative Hot-spot Analyses in PM2.5 and PM10 Nonattainment and Maintenance Areas,* Washington, D.C.: United States Environmental Protection Agency.

USEPA, 2012. *Fast Facts: U.S. Transportation Sector Greenhouse Gas Emissions 1900-2010,* Washington, D.C.: United States Environmental Protection Agency.

USEPA, 2012. *Using MOVES to Prepare Emission Inventories in State Implementation Plans and Transportation Conformity,* Washington, D.C.: United States Environmental Protection Agency.

Waddell, P. et al., 2003. Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics,* 3(1), pp. 43-67.

Wang, H., Jang , K. & Chan, C., 2010. *Analysis and Multi-Level Modeling of Truck Freight Demand.* Washington D.C., Strategic Highway Research Program 2 (SHRP2) Freight Modeling and Data Symposium.

Wang, Y., Kockelman, K. & Damien, P., 2013. A spatial autoregressive multinomial probit model for anticipating land-use change in Austin, Texas. *The Annals of Regional Science,* December.

Wang, Y., Kockelman, K. & Wang, X., 2012. Anticipation of Land use Change Through Use of Geographically Weighted Regression Models for Discrete Response. *Transportation Research Record: Journal of the Transportation Research Board,* pp. 111-123.

Wei, H., Li, Z. & Ai, Q., 2009. Observation-based study of intersection dilemma zone natures. *Journal of Transportation Safety and Security,* 1(4), pp. 282-295.

Wei, H. et al., 2011. *Integrated Approach to Explore Roadway Air impact of Travel Demand and Traffic Operation for Transportation Sustainability Analysis.* Minneapolis, MN., s.n.

Wei, H. et al., 2012. *Framework for Integrating Traffic-Source Emission Estimates into Sustainability Analysis.* Beijing, China, s.n.

Widhalm, P. et al., 2015. Discovering urban activity patterns in cell phone data. *Transportation,* 42(4), pp. 597-623.

Wright, L., Kemp, S. & Williams, I., 2011. Carbon footprinting: towards a universally accepted definition. *Carbon Management,* February, 2(1), pp. 61-72.

Yao, Z., Wei, H., Liu, H. & Li, Z., 2013. Statistical vehicle specific power profiling for urban freeways. *Procedia-Social and Behavioral Sciences,* 96(6), p. 2927–2938.

Yao, Z. et al., 2013. Developing operating mode distribution inputs for moves with a computer vision-based vehicle data collector. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 2340, pp. 49-58.

Yao, Z. et al., 2014. Sensitivity analysis of project level MOVES running emission rates for light and heavy duty vehicles. *Journal of Traffic and Transportation Engineering,* 1(2), pp. 62-71.

Yao, Z. et al., 2014. *Scenario-based Carbon Footprint Inventory Tool for Urban Sustainable Development Decision Support: The Cincinnati Case Study.* Washinton D.C. , Transportation Research Board.

Zhao, Y. & Kockelman, K., 2004. The random-utility-based multiregional input–output model: solution existence and uniqueness. *Transportation Research Part B: Methodological,* 38(9), pp. 789-807.

Zhou, B. & Kockelman, K., 2008. Neighborhood impacts on land use change: a multinomial logit model of spatial relationships. *The Annals of Regional Science ,* pp. 321-340.

Zhou, B. & Kockelman, K., 2009. Predicting the distribution of households and employment: a seemingly unrelated regression model with two spatial processes. *Journal of Transport Geography,* September, 17(5), p. 369–376.

Zhou, B., Kockelman, K. & Lemp, J., 2010. Applications of Integrated Transport and Gravity-Based Land Use Models for Policy Analysis. *Transportation Research Record: Journal of the Transportation Research Board,* January, 2133(2), pp. 123-132.

APPENDIX 1: List of Abbreviations

AADT	Average Annual Daily Traffic
ABS	Absolute Value
ADT	Average Daily Traffic
AWI	Aging Water Infrastructure
CA	Cellular Automata
CARB	California Air Resource Board
CBD	Central Business District
CO_2	Carbon Dioxide
ESRI	Environmental Systems Research Institute
FHWA	Federal Highway Administration
GIS	Geographical Information System
HBO	Home-based Others
HBSC	Home-based SChool
HBU	Home-based University
HBW	Home-based Work
HCM	Highway Capacity Manual
LOS	Level of Service
LUM	Land Use Model
MAPE	Mean Absolute Percentage Error
MCMC	Markov Chain Monte Carlo
MCSA	Multi-Criteria Suitability Analysis
MOVES	Motor Vehicle Emission Simulator
MPO	Metropolitan Planning Organization

NAAQS	National Ambient Air Quality Standards
OKI	Ohio Kentucky Indiana Regional Council of Governments
OTAQ	USEPA Office of Transportation and Air Quality
PM	Particular Matter
RAIA	Roadway Air Impact Analysis
SB-PSS	Scenario-Based Planning Support System
SIP	State Implementation Plans
TAZ	Traffic Analysis Zone
TDM	Travel Demand Model
USEPA	United States Environmental Protection Agency
USGS	United States Geological Survey
VMT	Vehicle Miles Traveled
WRAP	Water Resources Adaptation Program

APPENDIX 2: OKI GPS Survey Data Release Agreement

Ohio-Kentucky-Indiana Regional Council of Governments

Data Release Form

CONDITIONS FOR RELEASE

Confidential Data

The data that this release form pertains to includes confidential information. Such confidential information may include names, phone numbers, addresses, GPS coordinates, and/or email addresses. Because of the nature of the data and the ongoing work that OKI does with the public, it is in the best interest to keep this data secure. However, OKI understands that this data has a value to OKI projects and academic research. As such, OKI will, at its discretion, make this data available to interested parties. However, this data must be kept confidential.

Terms and Conditions

1. The recipient is responsible for maintaining the security of the data consistent with normal IT security practices. Any and all fees (legal or otherwise), settlements, etc. associated with failure to maintain security of the data will be the responsibility of the recipient.
2. Should a data breach occur, the recipient will contact OKI immediately.
3. The recipient will use the data provided solely for the project it was provided for. If the recipient wishes to use the data for another project, the recipient **must** get permission from OKI in writing.
4. The recipient will not redistribute the data without written permission from OKI.
5. To the extent permitted by Ohio law, recipient shall indemnify and hold harmless OKI, its officers, employees, and agents, against any and all claims, damages, liabilities, and court awards, including all costs, expenses, and attorney's fees, incurred as a result of the recipients use of the data.
6. OKI retains the right to change, update, or revoke permission to use the data at any time. Should OKI revoke permission to use the data, or OKI changes or updates these terms and the recipient finds the new terms unacceptable, the recipient must delete their copy of the data.
7. OKI makes no warranty or representation of the accuracy or suitability for a particular purpose of the data provided. Any opinions expressed by the recipient based on the data are the sole responsibility of the recipient.

The undersigned represents that he/she is authorized to enter into this agreement on behalf of the recipient and bind the recipient to the above Terms and Conditions

Type of data: Cincinnati GPS Household Travel Survey Data.

Recipient Name: Zhuo Yao

Recipient Title: Graduate Student.

Recipient Email: Yaozo@mail.uc.edu.

Recipient Firm: University of Cincinnati

Recipient Signature: [signature]

Date: April,01.2013 Phone Number: 513-382-6110

Project for which data is requested:

Recipient's Ph.D. dissertation.

Accepted by OKI

Name: Andrew Rollins

Signature: [signature]

Date: 4/1/13