

# Capturing the Built Environment - Travel Interaction for Strategic Planning: Development of A Multi-Modal Travel Module for Regional Strategic Planning Model (RSPM)

*Liming Wang<sup>1</sup>, Brian Gregor<sup>2</sup>, Huajie Yang<sup>1</sup>, Tara Weinder<sup>3</sup>, and Tony Knudson<sup>3</sup>*

*<sup>1</sup> Toulon School of Urban Studies and Planning, Portland State University; <sup>2</sup> Oregon System*

*Analytics, LLC; <sup>3</sup> Oregon Department of Transportation*

## **Abstract**

Integrated land use and transportation models have evolved along a spectrum from simple sketch planning models to complex microsimulation models. While each has its niche, they are largely unable to balance the flexibility and realism of microsimulation and the speed and interactivity of simple models. The Regional Strategic Planning Model (RSPM) aims to fill this gap by taking a microsimulation approach while making other simplifications in order to model first-order effects quickly. It enables planners to consider the robustness of prospective policies in the face of future uncertainties by accepting a broad range of inputs and allowing rapid simulations of many scenarios.

This paper introduces the RSPM and shows how new land use and multi-modal transportation

sensitivities have been incorporated through the conversion to the new VisionEval open-source framework. Land use and transportation interactions in the RSPM are reviewed and the development of a new multi-modal travel demand module with improved land use sensitivities is highlighted. The use of a unique nationwide dataset combining the 2009 NHTS, EPA’s Smart Location Database, and metropolitan transit and roadway data is explained. The paper concludes with the results of validation and sensitivity tests, and a discussion of future work.

## 1 Introduction

Integrated land use and transportation models of urban areas have evolved along a spectrum with simple sketch planning models on one end, to complex dynamic microsimulation models on the other. Simple sketch planning tools, such as CommunityViz, Envision Tomorrow+, and UrbanFootprint, provide rapid order-of-magnitude estimates for impacts of transportation and land use scenarios with limited information (Avin, Cambridge Systematics, Inc, & Patnode, 2016), while dynamic disaggregate microsimulation models, such as UrbanSim and ILUTE, aim to predict more detailed effects with higher precision and theoretical consistency through sophisticated modeling of individual behavior and multi-modal network details (Hunt, Kriger, & Miller, 2005). In a more recent review, Acheampong & Silva (2015) summarize the theoretical foundations and the status of 28 operational land-use-transportation interaction (LUTI) models. They particularly discuss the challenges of microsimulation LUTI models including long execution time and uncertainties with respect to model outputs resulting “from model misspecification, imperfect input information, and innate randomness in events and behaviors that are being modeled”. After discussing the topology and evolution of various LUTI models, Kii et al (2016) conclude that directions for future LUTI include “a simplified and essential model for national, regional and global assessment of urban land use”. While the

landscape of LUTI models is crowded, these reviews agree that there is a gap in this landscape that balances the flexibility and realism of microsimulation models with the speed and interactivity of simpler models.

The Regional Strategic Planning Model (RSPM) aims to occupy the niche between more aggregated sketch planning tools and more complex LUTI models. Sketch planning tools are primarily driven by correlation or univariate elasticities, but unable to reflect the interconnectedness between different urban phenomena, while full-fledged disaggregate models integrated with travel demand models aim at capturing the interdependence in the urban system at the cost of model complexity, detail, and run time. Advanced Activity-Based models capture interacting policies by microsimulation of each individual's daily trips by purpose using a detailed multi-modal network. Traditional travel demand models incorporate the network but policies and their interactions are limited by an aggregate treatment of the average behavior of "groups" of individuals.

RSPM uses microsimulation and interacting model components to enable better accounting of policy interactions and the social and spatial heterogeneity of effects than do sketch planning tools. The advantage of a microsimulation approach is in the interpretability of model results and the flexibility in introducing additional disaggregate variables and utilizing alternative model structures (Donnelly, Erhardt, Moeckel, & Davidson, 2010; Waddell, 2011). At the same time, RSPM simplifies the disaggregate modeling approach from that of more complex LUTI models. RSPM forecasts only overall travel based on urbanized area travel conditions and household attributes, avoiding network detail that leads to long run times. Thereby striking a balance between rapid computation and accurate representation of how different types of households will change travel behavior in response to policies and investments. These simplifications enable more comprehensive analysis consistent with the uncertainties inherent in long-range planning, and by allowing a larger set of

factors and many more scenarios to be evaluated in a timely manner. This enables the planning decision space and potential policy tradeoffs to be broadly explored. It also enables the development of web-based interactive decision-support tools to give planners and decision-makers the ability to better understand prospects and tradeoffs and resilience to alternative futures(Gregor, 2016).

The RSPM is an offshoot of the GreenSTEP model, a microsimulation modeling package originally developed to assist Oregon Department of Transportation (ODOT) and other state agencies develop statewide transportation strategies and policies for reducing greenhouse gas (GHG) emissions to meet state goals (Oregon State Legislature, 2015). The RSPM was developed from GreenSTEP to model metropolitan areas at a finer level of geographic detail to assist with the development of scenario plans covering policy actions relevant at the regional level, and reporting for regional goals, including Oregon's adopted metropolitan GHG targets (Oregon State Legislature, 2010). The model was also rebranded to reflect the fact that it provides analytical support for much more than GHG mitigation planning since it models a large number of transportation factors and produces a rich set of performance metrics. The RSPM and GreenSTEP models are both operational and have been used for a number of studies to support the metropolitan area planning as well as state plans, rules, and legislation in Oregon (Oregon Department of Transportation, Planning Section, 2012; Pietz & Gregor, 2014) as well as for long-range visioning in the Atlanta region. Other operational models that have branched off the original GreenSTEP model code base include the Energy and Emissions Reduction Policy Analysis Tool (EERPAT)(FHWA Office of Planning, Policy, and Realty, 2016), and the Rapid Policy Analysis Tool (RPAT) which was developed as part of the SHRP2 C16 project for evaluating the effects of smart growth policies on travel demand (Outwater, Collin, J. Walters, B. Welch, & R. Cervero, 2014).

Current efforts are merging the GreenSTEP family of tools into an open source common software

platform, VisionEval, to support multi-agency collaboration in the development of strategic planning models (Gregor, Weidner, & Raw, 2016). In conjunction with migrating the RSPM to the VisionEval platform, ODOT commissioned research and development to expand and improve the land use model components of the RSPM, increase the sensitivity of travel forecasting to land use variables, and improve multi-modal travel modeling capabilities. This paper describes the results of that work. The remainder of the paper provides an overview of land use and transportation model components of the RSPM as implemented in the VisionEval platform, then documents the specification, validation, and sensitivity testing of the new multi-modal travel demand module, before concluding with a discussion of the new module's integration with the RSPM and future research.

## **2 Overview of the RSPM and Modeled Land Use and Transportation Interactions**

Estimation of the GreenSTEP and RSPM vehicle ownership, daily VMT, and several other submodels relied heavily on the national sample of the 2001 National Household Travel Survey (NHTS). The use of the 2001 NHTS in the development of these models is documented in GreenSTEP and RSPM technical documents (Gregor, 2015) and an earlier paper (Clifton & Gregor, 2012). A benefit of using the nationwide data was that it opened up possibilities for transferring the model to other states and metropolitan areas in the U.S. Unfortunately, few land use variables are included in the 2001 NHTS datasets (i.e. population density and general location with respect to the metropolitan core). Although land use attributes are important variables in a number of the GreenSTEP and RSPM submodels, the small number of variables limits the land use and policy sensitivity of those models.

It has been a goal of this paper to improve the land use interactivity of the RSPM and the ability of the model to predict consequences for non-auto mode travel (i.e. walk, bike, public transit) that are highly dependent on built form as well as automobile travel. To that end, we conduct research and development to accomplish this goal in concert with the conversion of the RSPM to the new VisionEval framework. This open-source framework, implemented in the R statistical programming language (R Core Team, 2017), is being built to facilitate collaboration in the development of strategic planning models. Submodels are modules that can be written to interact with other submodels in a plug-and-play manner. Details on the framework design are available in the project repository (Gregor, 2017). The objectives of the research described herein were to create new modules and revise existing modules in the VisionEval framework to increase the land use and transportation interactivity of the RSPM and enable multimodal transportation analysis.

The development of improved land use and transportation modules for the RSPM was made possible by a unique dataset that was created by joining three nationwide datasets: the 2009 NHTS, the Smart Location Database, and transportation supply information from National Transit Database and the Texas Transportation Institute (TTI). This dataset enabled travel models to be estimated which are sensitive to a number of land use attributes characterizing the density, diversity, design, destination accessibility, and distance to transit (5Ds) of places. Modules have been created to model accommodate land use attributes and enable them to influence the interacting transportation modules. Following are highlights of RSPM model components and the land use and transportation modules included within them as context for understanding the enhancements gained from the models described in the remainder of the paper. Figure 1 provides a summary overview of the sequence of model steps; all RSPM steps are provided here for completeness, but those not directly related to this paper are not discussed.

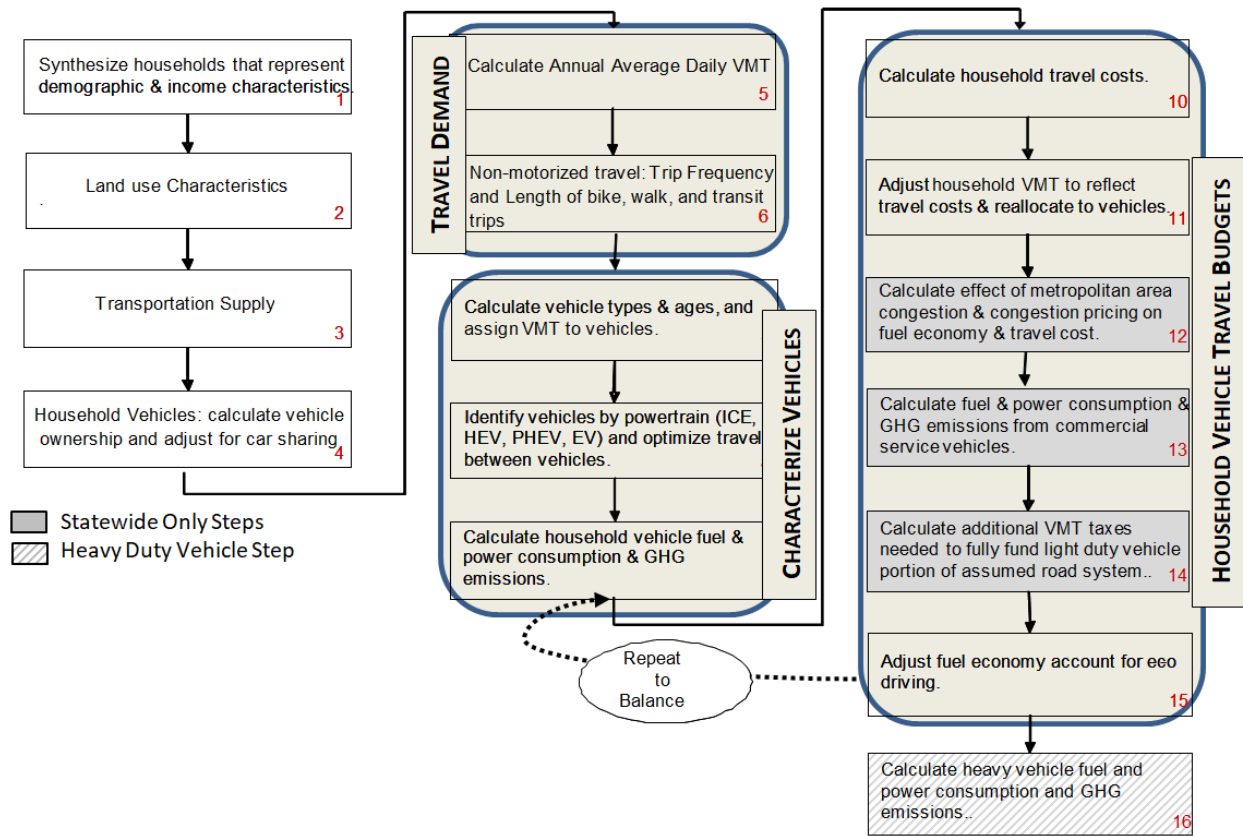


Figure 1: Overview of RSPM Steps

**Household Synthesis:** Households are synthesized having persons in each of 6 age brackets from a regional projection of population by age cohort using an iterative proportional fitting process. Workers by age bracket are similarly synthesized. The age and worker attributes are used to assign a household life cycle category. Household income is modeled as a function of the number of household workers and their ages and the average per capita income of households in the region.

**Land Use Characteristics:** Land use scenarios are created by specifying the numbers of single family and multi-family dwellings units and the numbers of jobs in 3 sectors by model zone (e.g. census block group). In addition, the income distribution of households in each zone is specified as the proportion of households by regional income quartile. A logit model assigns each household to a housing type and an iterative proportional fitting process assigns households to zones based on dwelling unit and income characteristics of each zone. Several density and diversity measures are calculated from the zonal allocations of population, employment, and zonal areas. Destination accessibility is also calculated based on those data as well as an interzonal distance matrix. Zonal network design and distance to transit attributes are inputs to the model.

**Transportation Supply:** The transportation supply (i.e. congestion) model is a networkless aggregate equilibrium model. It allocates daily vehicle miles traveled (VMT, 1 mile = 1.61 kilometers) among roadway functional classes (freeway, arterial, other) and congestion levels using relationships derived from urban mobility information collected by the Texas A&M Transportation Institute (Texas A&M Transportation Institute, 2015). Each congestion level for each functional class is associated with an average trip speed. Daily VMT is allocated between freeways and arterials as a function of the ratio of average trip speeds. The effect of pricing is modeled by converting prices into time equivalents. The effects of operations programs (e.g. ramp metering) on average speeds and the effect of speeds on fuel economy are accounted for based on research using EPA MOVES model



simulations by Clifton and Biggazzi (2011). Inputs to the model are the numbers of lane-miles of freeways and of arterials, the deployment of operations programs, prices (per vehicle mile) for traveling in severely and extremely congested conditions, and bus-equivalent transit revenue-miles. In addition, freeway capacities are adjusted in response to the deployment of driverless (autonomous) vehicles. Outputs from the model include VMT and VHT (vehicle-hours traveled) by roadway functional class, vehicle type, and congestion level.

**Travel Demand:** Modules calculate household travel by light-duty vehicles and by walking, bicycling, and public transit. The light-duty vehicle travel model calculates daily vehicle miles traveled by the household as a function of the characteristics of the household (e.g. size, age, income, vehicle availability, distance to transit), the land use characteristics of the zone where the household resides (e.g. density, diversity, design, destination accessibility), and urbanized area transportation supply levels (e.g. freeway lane-miles, transit-revenue miles, transit accessibility). Details about the estimation, validation, and sensitivity testing of this model are described below. The walking, bicycling and public transit models likewise are functions of household, land use, and transportation supply characteristics. The alternative mode travel models are also sensitive to the amount of household vehicle travel. This enables the effects of travel demand management programs and household travel budgets (explained below) to be translated into changes in travel by mode.

**Household Vehicles:** The RSPM includes several household vehicle models because of the strong relationships between vehicle characteristics, vehicle emissions, and the cost of vehicle travel. In addition to modeling the number of vehicles each household owns, the RSPM models the types (auto vs. light truck), ages, and powertrains (ICE, HEV, PHEV, EV) of vehicles, along with statewide fuel mix and associated fuel and electricity carbon intensities. The models also allocate travel between vehicles, and for plug-in hybrid electric vehicles, the proportions of travel powered by electricity

and gasoline. In addition, the RSPM models the potential for households to use carsharing services and the services of shared autonomous vehicles (SAV) to substitute for some or all of their vehicles. This is accomplished by modeling the marginal ownership cost per mile of travel and comparing that with the cost of using a carshare service, to determine a household's autos owned. Land use characteristics are included in several of these models and indirectly affect several others through effects on household vehicle travel.

**Household Vehicle Travel Budgets:** The RSPM models the effects of prices on household vehicle travel (fuel prices, parking prices, VMT taxes, etc.) using a household budgeting approach. This addresses the 'rebound effect' where improved vehicle fuel economy increases the amount of vehicle travel. The budget model is based on Consumer Expenditure Survey data which show that household spending on transportation has historically been fairly stable and that households shift expenses between transportation budget categories when gasoline prices fluctuate. The budget model establishes a maximum household budget for variable transportation costs and adjusts household vehicle travel to remain within the budget. The walk/bike/transit models are run after the budget model to account for the effect of vehicle travel costs on those travel modes. Finally, the model includes an optional feedback loop which calculates supplemental VMT taxes needed to cover deficits between roadway costs and fuel taxes - as vehicles consume less fuel per mile of travel - and adjusts travel in response.

The functionality described above has evolved with the development of the multi-modal model described in this paper. This includes the inclusion of built form variables as Land Use Characteristics and their use in predicting Travel Demand by mode and opportunities for their use in other modules to improve their sensitivity to land use.

Thus, the main aspect of full LUTI models that is missing from the RSPM is feedback from

transportation system performance to land use. Future land uses are specified through inputs regarding the numbers of housing units (by type), and jobs (by sector), and relative attractiveness (by household income quartile) for each zone. In other words, housing supplies are fixed inputs with respect to type and location. They are not derived from either household demand or accessibility. There are several reasons for this approach. First, the original development focus of GreenSTEP and the RSPM was on incorporating policy levers and other exogenous influences that are likely to have the greatest influences on future GHG emissions. Although the development of the freeway system in the U.S. has greatly affected land use and travel patterns in the past, the system is for all practical purposes built out and there is little appetite for considering more than incremental expansions (in most regions of the U.S. at least). It was determined at the time that adding feedback would substantially increase the model complexity while providing relatively little information gain to planners given the types of scenarios of interest to them. Second, the RSPM was developed to support a scenario planning approach, where users specify what-if scenarios in terms of land use and transportation inputs and the model estimates travel and environmental outcomes of those scenarios. This is not to say that the effects of transportation on land use are viewed as being insignificant, only that accounting for those effects in models that were initially created to support long-range planning for reducing GHG emissions remains a secondary consideration. As the RSPM and statewide GreenSTEP models expand, the value of and prospects for incorporating feedbacks from and consistency between transportation to land use will increase, which will be implemented in a more elaborated land use model in a future version of RSPM.

The travel demand module has been improved to better capture the built environment - travel interaction for regional strategic planning and is the main focus of this remainder of this paper. Details of VisionEval and other RSPM modules can be found on the Github project repository for VisionEval (Gregor, 2017) and technical documents for RSPM (Gregor, 2015).

### 3 Travel Demand Module

As being described above, the travel demand module captures the effects of household characteristics, built environment and urbanized area transportation supply levels on multi-modal household travel outcomes.

Multi-modal travel and its relationship to household characteristics, built environment, and transportation supply is a well-researched topic in the literature. Not only are there hundreds of original research papers on the topic, but there are now multiple reviews and syntheses of previous research (see, for example, Ewing & Cervero, 2001, 2010; Stevens, 2017). VMT, trip frequency, and trip length are the most common travel outcomes modeled (Ewing & Cervero, 2001). For our multi-modal travel module, we model VMT, trip frequencies and trip length for transit, bike, and walk travel. There are more than a dozen model types used in the literature; the most common model structures are linear regression model, logistic regression model and count model (Poisson regression and negative binomial regression) (Ewing & Cervero, 2010). Due to the space limit, we focus on the VMT model in the remainder of this paper. Results for the non-driving models are available in our project report (Wang, 2017).

In operational modeling systems, there are numerous approaches to model VMT ranging from full-fledged 4-step travel demand models and activity-based models to simplistic elasticity-based sketch planning models. Consistent with the design goal of RSPM, our travel demand module aims to balance flexibility, realism, speed, and interactivity. To achieve this end, we directly model long-term VMT as a function of household characteristics, built environment, and transportation supply for individual households in our data. For strategic planning tools like RSPM, annual average daily VMT (AADVMT) is more useful than modeling VMT on the day of the survey and

approximating average or annual VMT, which is commonly done in practice due to data availability or limitation. For example, GreenSTEP and the RSPM currently synthesize AADVMT for each household because the 2001 NHTS estimates of annual VMT are incomplete (available for less than half of the records) with questionable data quality (Clifton & Gregor, 2012).

### 3.1 Structure of AADVMT model

To estimate and validate our AADVMT model, we first calculate AADVMT for a household from an estimate of annual miles driven for each vehicle in the household (AVMT):

$$\text{AAADVMT}_h = \frac{\sum_{v_h=0}^{V_h} \text{AVMT}_{v_h}}{365}, \quad (1)$$

where

- $\text{AAADVMT}_h$  is the annual average daily VMT for household  $h$ ,
- $v_h \in \{0, \dots, V_h\}$  indexes vehicles in household  $h$ ,
- $V_h$  is number of vehicles in the household  $h$ , and
- $\text{AVMT}_{v_h}$  is the annual VMT driven for vehicle  $v_h$ .

In model estimation, household AADVMT  $\text{AAADVMT}_h$  computed with Equation (1) is then regressed on independent variables including household characteristics, built environment, and transportation supply:

$$\text{AAADVMT}_h = f(\text{SD}_h, \text{BE}_h, \text{TS}_{R_h}), \quad (2)$$

where

- $AADVMT_h$  is the annual average daily VMT for household  $h$ ,
- $SD_h$  represents the demographic and social-demographic characteristics of household  $h$ ,
- $BE_h$  is the built environment variables (of various geographical resolution) of household  $h$ , and
- $TS_{R_h}$  is the transportation supply of the region where household  $h$  resides.

In term of model structure options for household AADVMT model ( $f(\cdot)$  in Equation (2)), we consider three of the most commonly used structures in the literature (Ewing & Cervero, 2010): linear and transformed linear regression models, and a hurdle model, as well as the model structure used in the current version of the travel demand module of RSPM: 2-step models of binomial logit and linear/non-linear regression model.

### 3.1.1 Linear regression model

In a linear regression model of AADVMT, the dependent variable is modeled as a linear function of the independent variables:

$$AADVMT_h = X_h \beta^{lm} + \epsilon_h, \quad (3)$$

where

- $X_h$  is the independent variables that are composed of  $SD_h$ ,  $BE_h$ , and  $TS_{R_h}$ ,
- $\beta^{lm}$  is the linear regression coefficients to be estimated,
- $\epsilon_h$  is the error term of the model, assumed to be independent, identically and Normally

distributed (iid).

This is the simplest model form used for modeling VMT and is widely used in the literature (see, for example, Frank & Engelke, 2005; Kockelman, 1997; Sun, Wilmot, & Kasturi, 1998). However, the iid Normal assumption of the error term may be violated due to the nature of VMT measure.

### 3.1.2 Transformed linear regression models

We tested two forms transformed linear regression models: a semi-log model and power-transformed linear model, in which the dependent variable is log- and power-transformed, respectively. The right-hand side of the regression is otherwise similar to the linear regression model in Equation (3). These transformations are commonly used to address the violation of iid Normal assumption in a linear regression model.

### 3.1.3 2-step models

The current travel demand module of RSPM uses a 2-step approach to model survey day VMT: a binomial logistic regression model on whether a household has non-zero VMT and, for households with non-zero VMT, a power-transformed regression model of VMT (Clifton & Gregor, 2012). For the new travel demand module, we test the same 2-step model structure with AADVMT in the place of survey day VMT:

$$\Pr(\text{AAVMT}_h = 0) = \frac{\exp(X_h \beta^{2S1})}{1 + \exp(X_h \beta^{2S1})}, \text{ and} \tag{4}$$
$$\text{AAVMT}_h^\delta = X_h \beta^{2S2}, \text{ if } \text{AAVMT}_h > 0,$$

where

- $\Pr(\text{AADVMT}_h = 0)$  is the probability of *AADVMT* for household  $h$  equal 0, which is modeled by a logistic regression model;
- $X_h$  is the independent variables that are composed of  $\text{SD}_h$ ,  $\text{BE}_h$ , and  $\text{TS}_{R_h}$ ;
- $\beta^{2S1}$  and  $\beta^{2S2}$  are the coefficients to be estimated for the first and second step of the 2-Step models, respectively; and
- $\delta$  is the power parameter for the power-transformed regression, which is determined via the Box-Cox transformation (Box & Cox, 1964).

### 3.1.4 Hurdle model

Since VMT can only take values equal to or greater than 0 and households with zero VMT may be qualitatively different from those with positive VMT, there are applications of hurdle models to modeling of VMT (see, for example, Ewing et al., 2015). Similar to 2-step models described in Equation (4), a hurdle model has two regimes, one generating the zeros and one generating the positive values (W. H. Greene, 2011, pp. 821–826). The difference is that in a hurdle model, the estimation process maximizes the likelihood of these two regimes jointly, instead of doing it independently like in the 2-step models.

$$\Pr(\text{AADVMT}_h = 0) = \frac{\exp(X_h \beta^{h1})}{1 + \exp(X_h \beta^{h1})}, \text{ and}$$

$$\Pr(\text{AADVMT}_h = j) = (1 - \Pr(\text{AADVMT}_h = 0)) \frac{\exp(-\lambda_h) \lambda_h^j}{j! [1 - \exp(-\lambda_h)]}, j = 1, 2, \dots, \text{ with} \quad (5)$$

$$\lambda_h = \exp(X_h \beta^{h2}),$$



where

- $\Pr(\text{AADVMT}_h = 0)$  and  $\Pr(\text{AADVMT}_h = j)$  are the probability of *AADVMT* for household  $h$  equal 0 and  $j$  respectively;
- $X_h$  is the independent variables that are composed of  $\text{SD}_h$ ,  $\text{BE}_h$ , and  $\text{TS}_{R_h}$ ;
- $\beta^{h1}$  and  $\beta^{h2}$  are the coefficients to be estimated for the two regimes of the hurdle model, respectively; and
- $\lambda_h$  is the mean VMT for household  $h$  and is modeled as  $\exp(X_h\beta^{h2})$ .

### 3.2 Variable and Model Selection

We rely on previous research on travel behavior to ensure behavior validity of our models and aim to select at least one variables from each group of household characteristics, built environment, and transportation supply variables with moderate correlations. For each model structure, we loosely use a forward step-wise variable selection process to enter these variables into the specification to achieve highest prediction accuracy (measured by rmse - root mean square errors) while monitoring coefficients' statistical significance: we first control for household's socio-demographic characteristics, and gradually add built environment and transportation supply variables. Among the built environment variables, we aim to include at least one variable from each of the 5D categories of built environment measures. When no variable in a category is statistically significant at 5% significance level, we still include one that is marginally significant but provides the best improvement in prediction accuracy.

After finding the best model specification for each model structure, we use the 5-fold cross-validation to select the best model structure with highest prediction accuracy among 4 possible options. A

$k$ -fold cross-validation randomly partitions a sample evenly into  $k$  subsamples. In  $k$  iterations, each one of the  $k$  subsamples is reserved for cross-validation (testing) in turn, while the remaining  $k - 1$  subsamples are combined and used for estimation (training).  $k$ -fold validation is an efficient approach for cross-validation with low variance. 5- or 10-fold cross-validation is commonly used (Hastie, Tibshirani, & Friedman, 2016).

### 3.3 Data

We combine three nationwide datasets to create a unique dataset for model estimation and testing.

#### 3.3.1 NHTS

The 2009 NHTS (U.S. Department of Transportation, Federal Highway Administration, 2009) collected trips taken by all members of a surveyed household in a 24-hour period, as well as their socio-demographic characteristics. The 2009 NHTS included 150,145 households, 308,901 household members, and 1,079,763 trips.

The 2009 NHTS also includes odometer readings, as well as other attributes, of all vehicles in a household. Oak Ridge National Laboratory (ORNL) creates estimates of annual miles driven for each vehicle utilizing the odometer readings, self-reported annual miles driven, vehicle attributes, and household socio-demographic characteristics. ORNL validates the imputed annual VMT against highway statistics and 2001 NHTS and concludes the imputation “improves upon available data . . . [and provides] better estimates for a given vehicle” (Oak Ridge National Laboratory, 2011, p48).

Figure 2 shows histograms for AADVMT.

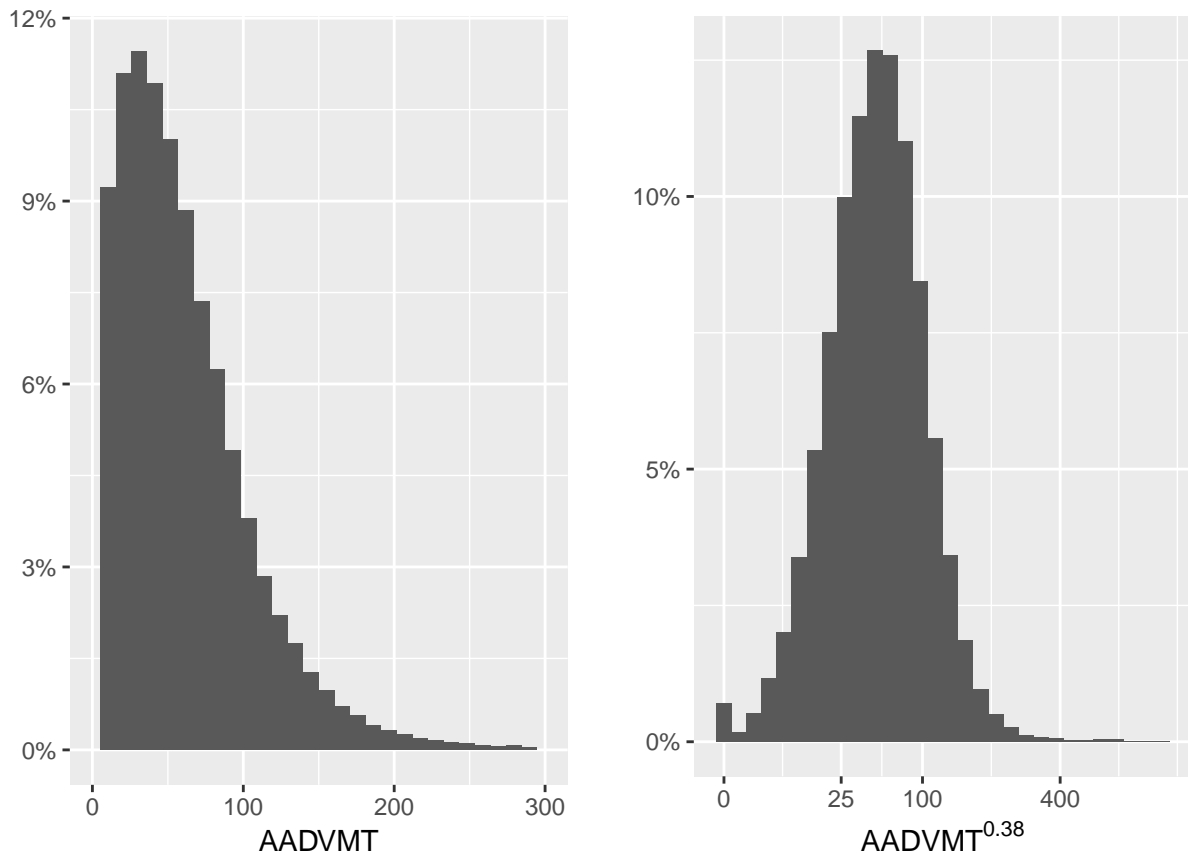


Figure 2: Histogram of household annual average daily VMT in the 2009 NHTS data

We are able to access the confidential residence Census Block Group (2010 geography) for all households in the 2009 NHTS. This information allows us to join household characteristics and travel outcomes in NHTS with the Smart Location Database to create a unique nationwide dataset with rich household characteristics, travel and built environment information. This unique dataset allows us to overcome a limitation that plagues similar research utilizing the NHTS data: it has a very limited set of built environment variables and previous studies resort to using urbanized areas as their unit of analysis (Cervero & Murakami, 2010; Glaeser & Kahn, 2008).

### **3.3.2 Smart Location Database**

The Smart Location Database (SLD) is a US nationwide database with extensive built environment variables organized around the 5D categorization - Density, Diversity, Design, Destination, Distance to transit. It includes more than 90 attributes summarizing characteristics such as housing density, diversity of land use, neighborhood design, destination accessibility, transit service, employment, and demographics. Most attributes are available for every census block group in the United States (Ramsey & Bell, 2014).

### **3.3.3 Transportation Supply Data**

Even though the SLD contains some measures of the transportation supply, such as transportation network density and access to transit stops and services, they are local measures at the Census block group level. There is likely network effect of transportation supply - the more complete a transportation network, the higher its utility to travelers - that cannot be captured by the local measures (Levinson & Krizek, 2008). To address this, we use the urbanized area level transportation

supply measures including freeway lane-kilometers (converted from original lane-miles), annual transit vehicle revenue kilometer (converted from original revenue miles), etc from Texas A&M Transportation Institute’s Urban Mobility Report (Texas A&M Transportation Institute, 2015).

Since these data are only available for urbanized areas (UZA), we segment the NHTS data similar to what Gregor did (2015): a UZA segment with complete information of household characteristics, built environment and transportation supply and a non-UZA segment for which regional transportation supply information is missing.

After joining these three datasets, we have a household-level dataset with about 200 variables. Table 1 presents a select subset of these variables with descriptions, source, and summary statistics. Note this is an incomplete list of variables that appear in at least one of the models we present in this paper.

Table 1: Variables, their source, description and summary statistics

Name	Source	Description	Mean	std dev
AADVMT	NHTS	Annual average daily VMT	60.01	48.87
ntrips.Transit	NHTS	Transit trips during the day of survey	0.17	0.76
ntrips.Bike	NHTS	Biking trips during the day of survey	0.07	0.49
ntrips.Walk	NHTS	Walking trips during the day of survey	0.70	1.60

Name	Source	Description	Mean	std dev
Age0to14	NHTS	Number of household members younger than 14	0.20	0.58
Age65Plus	NHTS	Number of household members older than 65	0.57	0.75
CENSUS_D	NHTS	Census division classification for home address: New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, or Pacific		
DRVRCNT	NHTS	Number of drivers in household	1.80	0.78
HHSIZE	NHTS	Count of household members	2.34	1.24
LIF_CYC	NHTS	Household life cycle classification: Single, Couple w/o children, Couple w/ children, or Empty Nester		
LogIncome	NHTS	log total household income	10.72	0.87
VehPerDriver	NHTS	Number of vehicles per licensed driver	1.12	0.57
WRKCOUNT	NHTS	Number of workers in household	0.93	0.89

Name	Source	Description	Mean	std dev
D1B	SLD	Gross population density (people/acre) on unprotected land	6.15	16.02
D2A_EPHHM	SLD	Employment and household entropy	0.47	0.23
D2A_WRKEMP	SLD	Household Workers per Job, as compared to the region	10.05	31.81
D3bpo4sqkm	SLD	Intersection density in terms of pedestrian-oriented intersections having four or more legs per square kilometer	33.95	59.53
D4c	SLD	Aggregate frequency of transit service within 400 meters of block group boundary per hour during evening peak period	25.70	65.44
D5ar	SLD	Jobs within 45 minutes auto travel time, time- decay (network travel time) weighted	85004.28	123761.90
D5cr	SLD	Employment accessibility expressed as a ratio of total MSA accessibility	0.00	0.01
TRPOPDEN	SLD	Census tract population density	5.53	15.04

Name	Source	Description	Mean	std dev
EMPTOT_5	SLD	Total employment within 8 kilometers buffer of block group	31773.32	95420.00
FwyLanekmP1k	TTI	Urbanized area freeway	1.12	0.62
	UMR	lane-kilometers per 1,000 person		
TranRevKmP1k	TTI	Urbanized area transit annual	26.95	19.13
	UMR	vehicle revenue kilometers per 1,000 person		

## 4 Results

To eliminate potential outliers, we exclude observations whose AADVMT value is above the 99% percentile. With this cutoff, we exclude 1403 observations (out of 150145 ) with AADVMT values ranging from 221.588 to 800.462. Observations with missing values in dependent or independent variables are also excluded. We use the variance inflation factor (VIF) to filter independent variables with serious multi-collinearity ( $VIF > 10$ )(Menard, 2001, p. 76).

Table 2 shows the model goodness-of-fit ( $R^2$  for linear regression models or pseudo- $R^2$  for non-linear models) and root mean squared error (rmse) for each of 5-fold cross-validation. As shown in Table 2, the power-transformed model and the 2-step models have the best accuracy (lowest rmse) in cross-validation. We choose the power-transformed model for its simplicity, easy of use, and computational performance. The power parameter is determined via the Box-Cox transformation (Box & Cox, 1964):  $\delta = 0.38$ .



Table 2: k-fold cross-validation for model structure selection

Model Type	Segment	$k_{th}$ -fold	rmse	$R^2$	pseudo- $R^2$
2 step models	UZA	1	29.27	0.46	0.59
hurdle model	UZA	1	30.54		0.33
power-transformed linear regression model	UZA	1	29.31	0.45	
semi-log regression model	UZA	1	32.51	0.41	
2 step models	UZA	2	28.91	0.46	0.61
hurdle model	UZA	2	29.69		0.33
power-transformed linear regression model	UZA	2	28.89	0.46	
semi-log regression model	UZA	2	31.43	0.42	
2 step models	UZA	3	29.24	0.46	0.60
hurdle model	UZA	3	29.91		0.33
power-transformed linear regression model	UZA	3	29.29	0.46	
semi-log regression model	UZA	3	31.63	0.42	
2 step models	UZA	4	29.48	0.46	0.58
hurdle model	UZA	4	31.35		0.34
power-transformed linear regression model	UZA	4	29.46	0.46	
semi-log regression model	UZA	4	33.79	0.42	
2 step models	UZA	5	28.98	0.46	0.57

Model Type	Segment	$k_{th}$ -fold	rmse	$R^2$	pseudo- $R^2$
hurdle model	UZA	5	29.84		0.34
power-transformed linear regression model	UZA	5	28.94	0.45	
semi-log regression model	UZA	5	31.67	0.41	
2 step models	non-UZA	1	32.78	0.47	0.41
hurdle model	non-UZA	1	34.04		0.12
power-transformed linear regression model	non-UZA	1	32.79	0.47	
semi-log regression model	non-UZA	1	36.00	0.43	
2 step models	non-UZA	2	32.77	0.47	0.41
hurdle model	non-UZA	2	33.67		0.12
power-transformed linear regression model	non-UZA	2	32.75	0.47	
semi-log regression model	non-UZA	2	35.46	0.43	
2 step models	non-UZA	3	32.77	0.47	0.39
hurdle model	non-UZA	3	33.74		0.12
power-transformed linear regression model	non-UZA	3	32.74	0.47	
semi-log regression model	non-UZA	3	35.50	0.43	
2 step models	non-UZA	4	32.97	0.47	0.39
hurdle model	non-UZA	4	34.70		0.12

Model Type	Segment	$k_{th}$ -fold	rmse	$R^2$	pseudo- $R^2$
power-transformed linear regression model	non-UZA	4	32.96	0.47	
semi-log regression model	non-UZA	4	37.30	0.43	
2 step models	non-UZA	5	33.38	0.47	0.41
hurdle model	non-UZA	5	34.95		0.12
power-transformed linear regression model	non-UZA	5	33.35	0.47	
semi-log regression model	non-UZA	5	37.37	0.43	

Table 3 presents the final estimation results for the power-transformed model. All the households characteristics covariates have expected sign. For the VehPerDriver variable (number of vehicles per driver in a household), we use a cubic spline on  $\log(\text{VehPerDriver}+1)$  to capture the non-linear effect of households' vehicle ownership on driving. The k-fold cross-validation helps ensure that specifications with such non-linear transformations do not overfit the sample.

Figure 3 shows the non-linear elasticity of AADVMT to VehPerDriver: the x-axis is the percentage change to current vehicle per driver level in NHTS households and the y-axis the average AADVMT over all UZA or non-UZA households. It is clear that the elasticity is non-linear. In the NHTS, UZA households have  $\text{AADVMT} = 48.291$  (where the UZA elasticity curve intersects with  $x = 0$ ). As VehPerDriver reduces by 100% (in this case, all households owns 0 vehicle), AADVMT drops to 5.555, a 88.498% drop. However, if VehPerDrive increases by 100%, AADVMT increases to 66.675, a 38.068% increase. The next 100% increase in VehPerDriver brings even less increase in AADVMT. It also shows that households in non-UZA have higher elasticities than UZA households.

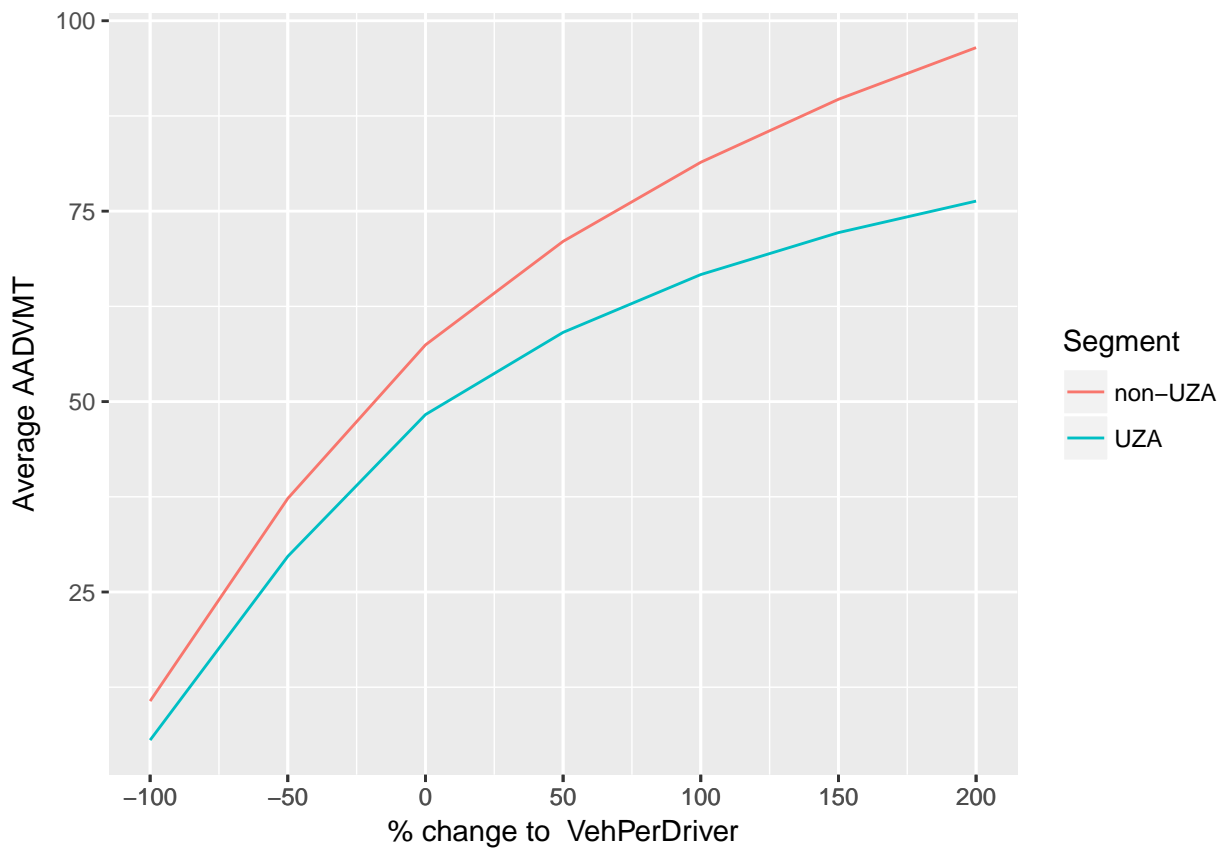


Figure 3: AADVMT elasticity to VehPerDriver. The elasticity is non-linear and non-UZA households have higher elasticities to VehPerDriver than UZA households

Most coefficients for the 5D variables have expected sign and relative small magnitude as documented in the literature (Ewing & Cervero, 2010; Stevens, 2017), except for D1B (population density) and D2A\_EPHHM (land use diversity) variable for the non-UZA segment. After controlling for TRPOPDEN (tract population density) and employment within 8-kilometer buffer (TOTEMP\_5), coefficients for D1B and D2A\_EPHHM are positive. In addition to each 5D variables by themselves, we also test interaction terms across different 5D categories. The significant and positive coefficient for the interaction term of UZA transit revenue kilometers per 1,000-person (TranRevKmP1k) with a localized measure of the frequency of transit service (D4c) indicates that good local access to transit services in a UZA with extensive transit network have an extra effect on reducing driving. Similarly, the interaction term of D1B and D2A\_EPHHM for the non-UZA model has an expected negative sign, which may indicate that density or diversity alone in non-UZA areas is not effective in reducing driving, but the two working together is. Excluding the interaction term of D1B:D2A\_EPHHM does not change the sign or significance of the coefficients for D1B or D2A\_EPHHM.

## 5 Conclusion and Discussion

In this paper, we introduce the RSPM suite of tools: its origin and niche of focusing on long-term effects and enabling rapid scenario simulations with a flexible microsimulation approach, its evolution, and current development. We then focus the recently improved multi-modal travel demand module for the RSPM that better captures the land use-travel interaction. We document the process and techniques we utilize to do model selection, validation and sensitivity testing with the household AADVMT model. We aim for the simplest model with the best predicting power and behavior validity. We end up choosing the power-transformed method among four model structure options with a k-fold cross-validation process. The cross-validation process also ensures we don't overfit the

sample data as we consider transformations to capture non-linear effects and interaction effects. We believe this paper is the first that benchmarks alternative model structures for modeling AADVMT using a rigorous cross-validation process.

We utilize a unique dataset by joining three US nationwide datasets: the 2009 NHTS, the Smart Location Database, and transportation supply information from National Transit Database and TTI. We believe the household AADVMT model presented in this paper is the first model utilizing such a nationwide dataset with high resolution built environment variables while controlling for households' socio-demographic characteristics. Previous nationwide research in the US is either limited to use only the coarse measures of built environment (for example, Gregor, 2015) or aggregate household travel to a larger geography like UZA (for example, Cervero & Murakami, 2010; Glaeser & Kahn, 2008) while discarding most other household characteristics.

However, using a large sample with many variables is not without challenges, including the curse of dimensionality and spurious correlation. With the number of variables we have, it is almost at the limit of what is possible with a manual model selection process, especially since we want to consider non-linear effects and interactions between variables. Model selection techniques such as stepwise regression and Bayesian model averaging are helpful only to an extent in this case. Machine learning techniques may be a potential solution and are a direction we will explore in our future research. With the large sample size of more than 150,000 in the 2009 NHTS, it is very easy to find a statistically significant correlation between an independent variable and the dependent variables. However, not all significant correlations are meaningful. In this paper, we rely on prediction accuracies in cross-validations in addition to model goodness-of-fit to guide the model selection and avoid overfitting, but the process is very onerous with a large number of independent variables.

Although not included in this paper, the models for non-auto travel, along with the AADVMT model, made up the multi-modal travel demand module that are implemented for the open source VisionEval software framework. We hope that its modular structure, openly available source code, and documentation will enable modifications and extensions of this work by the research community.

Finally, the new travel demand module is applied to the Rogue Valley MPO (the MPO area including Medford in southern Oregon with more than 70,000 households) for a performance check, sensitivity tests, and external validation. The module takes seconds on a modern PC of moderate configuration with the RVMPO data. The new travel demand module has better sensitivity to built environment than the current RSPM module. We also compare the predictions of VMT and non-motorized travel from the new module against the information in Oregon Household Activity Survey. The new module produces perfect predictions in aggregation (predicted VMT of 41.8 vs observed VMT of 36.7) and by market segments. Overall, the new module has demonstrated improvement in all aspects compared. Details of the module testing are available in a project report (Wang, 2017). A comparison with full-fledged travel demand models is left for future research, as they are not readily available for RVMPO.

## 6 Acknowledgements

The research project received financial support from Oregon Department of Transportation (SPR 788) and National Institute for Transportation and Communities (NITC-881).

## References

Acheampong, R. A., & Silva, E. (2015). Land use - transport interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land Use*. <https://doi.org/10.5198/jtlu.2015.806>

Avin, U., Cambridge Systematics, Inc, & Patnode, P. (2016). *Sketch Planning Tools for Regional Sustainability* (No. NCHRP 08-36/Task 117). Washington, D.C.: The AASHTO Standing Committee on Planning. Retrieved from [http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36\(117\)\\_FR.pdf](http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36(117)_FR.pdf)

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252.

Cervero, R., & Murakami, J. (2010). Effects of built environments on vehicle miles traveled: Evidence from 370 US urbanized areas. *Environment and Planning A*, *42*(2), 400–418. Retrieved from <http://epn.sagepub.com/content/42/2/400.short>

Clifton, K., & Bigazzi, A. (2011). *Refining GreenSTEP: Impacts of Vehicle Technologies and ITS/Operational Improvements on Travel Speed and Fuel Consumption Curves*. Portland, OR: Portland State University. Retrieved from [http://ppms.trec.pdx.edu/media/project\\_files/GreenSTEP\\_Congestion\\_Final\\_Report.pdf](http://ppms.trec.pdx.edu/media/project_files/GreenSTEP_Congestion_Final_Report.pdf)

Clifton, K., & Gregor, B. (2012). Development of Decision Tool for Strategies to Reduce Greenhouse Gas Emissions. *Transportation Research Record: Journal of the Transportation Research Board*,



2291, 124–134. <https://doi.org/10.3141/2291-14>

Donnelly, R., Erhardt, G. D., Moeckel, R., & Davidson, W. A. (2010). *NCHRP Synthesis 406: Advanced Practices in Travel Forecasting*. Washington, D.C.: The National Academies Press. Retrieved from <https://www.nap.edu/catalog/22950/advanced-practices-in-travel-forecasting>

Ewing, R., & Cervero, R. (2001). Travel and the Built Environment: A Synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, 1780, 87–114. <https://doi.org/10.3141/1780-10>

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment - A Meta-Analysis. *Journal of the American Planning Association*, 76(3), 265–294.

Ewing, R., Tian, G., Goates, J. P., Zhang, M., Greenwald, M. J., Joyce, A., . . . Greene, W. (2015). Varying influences of the built environment on household travel in 15 diverse regions of the United States. *Urban Studies*, 52(13), 2330–2348. <https://doi.org/10.1177/0042098014560991>

FHWA Office of Planning, Policy, and Realty. (2016). *EERPAT - Energy and Emissions Reduction Policy Analysis Tool Version 3.0* (No. FHWA-HEPN-10). Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration. Retrieved from [https://www.planning.dot.gov/FHWA\\_tool/default.aspx](https://www.planning.dot.gov/FHWA_tool/default.aspx)

Frank, L. D., & Engelke, P. (2005). Multiple Impacts of the Built Environment on Public Health: Walkable Places and the Exposure to Air Pollution. *International Regional Science Review*, 28(2), 193–216. <https://doi.org/10.1177/0160017604273853>

Glaeser, E. L., & Kahn, M. E. (2008). *The Greenness of Cities: Carbon Dioxide Emissions and Urban Development* (Working Paper No. 14238). National Bureau of Economic Research.

<https://doi.org/10.3386/w14238>

Greene, W. H. (2011). *Econometric Analysis* (7 edition). Boston: Pearson.

Gregor, B. (2015, December). GreenSTEP & RSPM Model Version 3.5 Technical Documentation. Retrieved March 26, 2017, from [https://github.com/gregorbj/RSPM/tree/master/Version\\_3.5/Documentation](https://github.com/gregorbj/RSPM/tree/master/Version_3.5/Documentation)

Gregor, B. (2016). RVMPO Metropolitan Planning Area Scenario Viewer. Retrieved March 26, 2017, from <https://regorbj.github.io/RSPM-Viewer/>

Gregor, B. (2017). *VisionEval Model System and Framework*. US Federal Highway Administration. Retrieved from <https://github.com/VisionEval/VisionEval>

Gregor, B., Weidner, T., & Raw, J. (2016). VisionEval: A Model System and Software Framework for the Collaborative Development of Regional Strategic Planning Models. In. Presented at the 6th Transportation Research Board Conference on Innovations in Travel Modeling, Denver, CO.

Hastie, T., Tibshirani, R., & Friedman, J. (2016). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* (2nd edition). New York, NY: Springer.

Hunt, J. D., Kriger, D. S., & Miller, E. J. (2005). Current Operational Urban Land-Use: A Review. *Transport Reviews*, 25(3), 329–376. <https://doi.org/10.1080/0144164052000336470>

Kii, M., Nakanishi, H., Nakamura, K., & Doi, K. (2016). Transportation and spatial development: An overview and a future direction. *Transport Policy*, 49(Supplement C), 148–158. <https://doi.org/10.1016/j.tranpol.2016.04.015>

Kockelman, K. (1997). Travel Behavior as Function of Accessibility, Land Use Mixing, and Land

Use Balance: Evidence from San Francisco Bay Area. *Transportation Research Record: Journal of the Transportation Research Board*, 1607, 116–125. <https://doi.org/10.3141/1607-16>

Levinson, D. M., & Krizek, K. J. (2008). *Planning for Place and Plexus: Metropolitan Land Use and Transport* (1 edition). New York; London: Routledge.

Menard, S. (2001). *Applied Logistic Regression Analysis* (2nd edition). Thousand Oaks, Calif: SAGE Publications, Inc.

Oak Ridge National Laboratory. (2011). *Developing a Best Estimate of Annual Vehicle Mileage for 2009 NHTS Vehicles*. Washington, D.C.: Federal Highway Administration. Retrieved from <http://nhts.ornl.gov/2009/pub/DerivedAddedVariables2009.pdf>

Oregon Department of Transportation, Planning Section. (2012). *Oregon Statewide Transportation Strategy: A 2050 Vision for Greenhouse Gas Emissions Reduction* (No. Volum 2 Technical Appendices). Salem, OR. Retrieved from [https://www.oregon.gov/ODOT/TD/OSTI/docs/STS/Oregon\\_Statewide\\_Transportation\\_Strategy.pdf](https://www.oregon.gov/ODOT/TD/OSTI/docs/STS/Oregon_Statewide_Transportation_Strategy.pdf)

Oregon State Legislature. Chapter 85 Oregon Laws SB 1059 (2010).

Oregon State Legislature. Oregon Revised Statutes 468A.205 (2015).

Outwater, M., Collin, S., J. Walters, B. Welch, & R. Cervero. (2014). *SHRP 2 Report S2-C16-RR-1: Effect of Smart Growth Policies on Travel Demand*. Washington, D.C.: The National Academies Press. Retrieved from <https://www.nap.edu/catalog/22616/effect-of-smart-growth-policies-on-travel-demand>

Pietz, A., & Gregor, B. (2014). Oregon Statewide Transportation Strategy. *Transportation Research*

*Record: Journal of the Transportation Research Board*, 2454, 45–52. <https://doi.org/10.3141/2454-06>

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Ramsey, K., & Bell, A. (2014). *Smart Location Database: Version 2.0 User Guide*. Washington, DC 20460: US Environment Protection Agency. Retrieved from <https://www.epa.gov/smartgrowth/smart-location-mapping#SLD>

Stevens, M. R. (2017). Does Compact Development Make People Drive Less? *Journal of the American Planning Association*, 83(1), 7–18. <https://doi.org/10.1080/01944363.2016.1240044>

Sun, X., Wilmot, C., & Kasturi, T. (1998). Household Travel, Household Characteristics, and Land Use: An Empirical Study from the 1994 Portland Activity-Based Travel Survey. *Transportation Research Record: Journal of the Transportation Research Board*, 1617, 10–17. <https://doi.org/10.3141/1617-02>

Texas A&M Transportation Institute. (2015). *Urban Mobility Report*. College Station, TX.

U.S. Department of Transportation, Federal Highway Administration. (2009). *2009 National Household Travel Survey*. Retrieved from <http://nhts.ornl.gov>

Waddell, P. (2011). Integrated Land Use and Transportation Planning and Modelling: Addressing Challenges in Research and Practice. *Transport Reviews*, 31(2), 209–229. <https://doi.org/10.1080/01441647.2010.525671>

Wang, L. (2017). *Performance-Based Planning and Decision Making* (No. SPR-788). Salem, OR: Oregon DOT.

Table 3: Estimation results for the power-transformed non-linear model

	<i>AADVMT</i> <sup>0.38</sup>	
	UZA	Non-UZA
DRVRCNT	0.719*** (0.011)	0.755*** (0.011)
HHSIZE	0.004 (0.008)	0.018** (0.009)
WRKCOUNT	0.178*** (0.008)	0.167*** (0.007)
LogIncome	0.250*** (0.007)	0.299*** (0.006)
Age0to14	0.095*** (0.011)	0.097*** (0.011)
Age65Plus	-0.066*** (0.008)	-0.073*** (0.007)
ns(log1p(VehPerDriver), 3)1	2.770*** (0.047)	2.730*** (0.043)
ns(log1p(VehPerDriver), 3)2	5.870*** (0.192)	5.600*** (0.171)
ns(log1p(VehPerDriver), 3)3	2.950*** (0.208)	3.530*** (0.173)
LIF_CYCEmpty Nester	-0.227*** (0.016)	-0.188*** (0.015)
LIF_CYCParents w/ children	0.034* (0.017)	0.019 (0.017)
LIF_CYCSingle	-0.186*** (0.020)	-0.176*** (0.020)
log1p(TRPOPDEN)	-0.028*** (0.010)	-0.039*** (0.013)
log1p(EMPTOT_5)	-0.057*** (0.005)	-0.037*** (0.003)
CENSUS_DEast South Central	0.079 (0.054)	0.087*** (0.030)
CENSUS_DMiddle Atlantic	-0.108*** (0.027)	-0.168*** (0.022)
CENSUS_DMountain	-0.084*** (0.026)	-0.111*** (0.029)
CENSUS_DNew England	-0.130*** (0.045)	-0.025 (0.030)
CENSUS_DPacific	-0.077*** (0.023)	-0.202*** (0.024)
CENSUS_DSouth Atlantic	0.018 (0.023)	0.030 (0.019)
CENSUS_DWest North Central	-0.030 (0.058)	-0.057** (0.025)
CENSUS_DWest South Central	0.073*** (0.024)	0.084*** (0.021)
FwyLanekmP1k	0.040*** (0.015)	
TranRevKmp1k	-0.0004 (0.0003)	
D1B	-0.001*** (0.0004)	0.010*** (0.004)
D2A_WRKEMP	-0.0003** (0.0001)	
D3bpo4sqkm	-0.0002** (0.0001)	
D5cr	-12.000*** (2.930)	
TranRevKmP1k:D4c	-0.00000*** (0.00000)	
D2A_EPHHM		0.044* (0.026)
I(D5ar/1000)		-0.00002 (0.0002)
D1B:D2A_EPHHM		-0.026*** (0.007)
Constant	-1.900*** (0.112)	-2.220*** (0.094)
Observations	47,288	55,103
R <sup>2</sup>	0.456	0.464
Adjusted R <sup>2</sup>	0.456	0.464
Residual Std. Error	0.979 (df = 47258)	1.000 (df = 55076)
F Statistic	1,366.000*** (df = 29; 47258)	1,834.000*** (df = 26; 55076)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01  
Std error in parentheses.