DEVELOPMENT OF A NEW VISIONEVAL LAND USE MODEL AND PILOT APPLICATIONS IN OREGON

Final Report

SPR 868

Oregon Department of Transportation

DEVELOPMENT OF A NEW VISIONEVAL LAND USE MODEL AND PILOT APPLICATIONS IN OREGON

Research Report
SPR 868

by

Liming Wang
Portland State University

and

Brian Gregor
Oregon System Analytics, LLC

for

Oregon Department of Transportation
Research Section
555 13th Street NE, Suite 1
Salem OR 97301

and

Federal Highway Administration
1200 New Jersey Avenue SE
Washington, DC 20590

May 2026

Technical Report Documentation Page

1. Report No.

FHWA-OR-RD-26-12

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle

Development of a New VisionEval Land Use Model and Applications to Evaluate Climate Strategies in Oregon

5. Report Date

-May 2026

6. Performing Organization Code: PSU

7. Author(s)

Liming Wang, Portland State University, ORCID

Brian Gregor, Oregon System Analytics, LLC (OSA)

8. Performing Organization Report No.

9. Performing Organization Name and Address

Oregon Department of Transportation
Research Section
555 13th Street NE
Salem, OR 97301

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

12. Sponsoring Agency Name and Address

Oregon Dept. of Transportation
Research Section
555 13th Street NE, Suite 1
Salem, OR 97301

Federal Highway Admin.
1200 New Jersey Avenue SE
Washington, DC 20590

13. Type of Report and Period Covered

Final Report

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This project upgraded the land use module for the VisionEval framework to enhance the evaluation of integrated transportation and land use policies and their impacts. The project developed a new strategic land use model utilizing NHTS and Smart Location Database data from 2011 and 2019, applying advanced statistical and machine learning methods to predict land use changes. Key innovations included land use transition model and allocation models for dwelling units and employment, along with D3 (Urban Design) and D4 (Level of Transit Service) classification methods. Project deliverables comprised a new VELandUse R package, a comprehensive data package with enriched variables, and pilot applications demonstrating the model's effectiveness. These enhancements significantly improved ODOT's capabilities for detailed land use modeling and regional strategy evaluation, while providing user-friendly tools for scenario creation and policy evaluation.

17. Key Words

Land use model, strategic tools, performance-based transportation planning

18. Distribution Statement

Copies available from NTIS, and online at www.oregon.gov/ODOT/TD/TP_RES/

19. Security Classification (of this report)

Unclassified

20. Security Classification (of this page)

Unclassified

21. No. of Pages

176

22. Price

SI* (Modern Metric) Conversion Factors
Approximate Conversions to SI Units

Physical Quantity	Symbol	When You Know	Multiply By	To Find	Symbol
Length	n	inches	25.4	millimeters	mm
Length	ft	feet	0.305	meters	m
Length	yd	yards	0.914	meters	m
Length	mi	miles	1.61	kilometers	km
Area	in2	square inches	645.2	square millimeters	mm2
Area	ft2	square feet	0.093	square meters	m2
Area	yd2	square yard	0.836	square meters	m2
Area	ac	acres	0.405	hectares	ha
Area	mi2	square miles	2.59	square kilometers	km2
Volume	fl oz	fluid ounces	29.57	milliliters	mL
Volume	gal	gallons	3.785	liters **	L
Volume	ft3	cubic feet	0.028	cubic meters	m3
Volume	yd3	cubic yards	0.765	cubic meters	m3
Mass	oz	ounces	28.35	grams	g
Mass	lb	pounds	0.454	kilograms	kg
Mass	T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
Temperature (exact degrees)	oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	oC
Illumination	fc	foot-candles	10.76	lux	lx
Illumination	fl	foot-Lamberts	3.426	candela/m2	cd/m2
Force and Pressure or Stress	lbf	poundforce	4.45	newtons	N
Force and Pressure or Stress	lbf/in2	poundforce per square inch	6.89	kilopascals	kPa

*SI is the symbol for the International System of Measurement
** Volumes greater than 1000 L shall be shown in m^3

SI* (Modern Metric) Conversion Factors
Approximate Conversions from SI Units

Physical Quantity	Symbol	When You Know	Multiply By	To Find	Symbol
Length	mm	millimeters	0.039	inches	in
Length	m	meters	3.28	feet	ft
Length	m	meters	1.09	yards	yd
Length	km	kilometers	0.621	miles	mi
Area	mm2	square millimeters	0.0016	square inches	in2
Area	m2	square meters	10.764	square feet	ft2
Area	m2	square meters	1.195	square yards	yd2
Area	ha	hectares	2.47	acres	ac
Area	km2	square kilometers	0.386	square miles	mi2
Volume	mL	milliliters	0.034	fluid ounces	fl oz
Volume	L	liters	0.264	gallons	gal
Volume	m3	cubic meters	35.314	cubic feet	ft3
Volume	m3	cubic meters	1.307	cubic yards	yd3
Mass	g	grams	0.035	ounces	oz
Mass	kg	kilograms	2.202	pounds	lb
Mass	Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
Temperature (exact degrees)	oC	Celsius	1.8C+32	Fahrenheit	oF
Illumination	lx	lux	0.0929	foot-candles	fc
Illumination	cd/m2	candela/m2	0.2919	foot-Lamberts	fl
Force and Pressure or Stress	N	newtons	0.225	poundforce	lbf
Force and Pressure or Stress	kPa	kilopascals	0.145	poundforce per square inch	lbf/in2

For More Information see: https://www.fhwa.dot.gov/publications/convtabl.cfm

ACKNOWLEDGEMENTS

The author gratefully acknowledges the members and friends of our Technical Advisory Committee (TAC), whose expertise and thoughtful feedback shaped key decisions throughout this project:

- Alexander Bettinardi, ODOT
- Jeffrey Hood, Oregon Metro
- Zachary Horowitz, ODOT
- Cody Meyer, DLCD
- Jeremy Raw, FHWA
- Jonathan Slason, RSG
- Colin Smith, RSG
- Tara Weidner, ODOT

Eun Jun Choi, graduate student at Portland State University, made essential contributions to this project. We are grateful for the data and technical support provided by staff at ODOT, FHWA, National Renewable Energy Laboratory, RSG, and SKATS. Finally, we extend special thanks to Josh Roll, ODOT Research Section Coordinator, for his excellent management and continuous support throughout the project.

DISCLAIMER

This document is disseminated under the sponsorship of the Oregon Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Oregon and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the view of the authors who are solely responsible for the facts and accuracy of the material presented. The contents do not necessarily reflect the official views of the Oregon Department of Transportation or the United States Department of Transportation.

The State of Oregon and the United States Government do not endorse products of manufacturers. Trademarks or manufacturers' names appear herein only because they are considered essential to the object of this document.

This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The main goal of this SPR project was to design, implement, and test a new strategic land use model in VisionEval so ODOT can evaluate transportation and land use policies in a more integrated way. The resulting upgrades improve ODOT's ability to analyze how land use and transportation strategies affect each other and influence GHG emissions. They also expand VisionEval's ability to represent land use at a more detailed level, evaluate regional land use strategies, and assess the social equity implications of land use and transportation strategies. In particular, Use Case #3 is a major contribution because it extends VisionEval to predict plausible land use change over time using real Bzones (real geographic analysis zones rather than simulated zones; see Chapter 2) and user constraints, which is especially important for larger-scale VE-State-style applications.

The key objectives of this project included:

- Designing, documenting, and implementing a new strategic land use model for the VisionEval framework
- Utilizing the latest and historical National Household Travel Survey (NHTS) and Smart Location Database (SLD) (2011 & 2019) for model estimation and validation
- Identifying and selecting optimal statistical and machine learning methods for predicting land use changes
- Applying rigorous testing and validation methods to ensure practical implementation
- Enabling user-friendly creation and evaluation of land use scenarios alongside transportation policies
- Piloting applications of the new land use model with existing VisionEval modules to assess various climate strategies

The report then moves from model design to implementation and application. It explains the design choices behind the land use type, D3, D4, and allocation models, documents how the new functionality is implemented in VisionEval, and uses pilot applications in Oregon to show both the added capabilities of the new framework and the current limits that remain visible in the results.

This report begins with a literature review of existing land use models and documents the extensive data sources utilized in the project. It reports the model design and selection of each modeling component, including the land use types, D3 (Urban Design) and D4 (Level of Transit Service) classification, and allocation models of dwelling units and employment. It then describes implementation of the new land use model in the VisionEval framework and concludes with the pilot applications of the new land use modules in Oregon. Besides this report, the project developed a new VELandUse module as an R package for the VisionEval framework, as well as a data package including a rich set of variables.

1.0 LITERATURE REVIEW

1.1 PLACE TYPE

Place types classify urban areas based on their defining characteristics, aiding in the prediction of future urban developments by tracking changes across different place types and identifying relationships between them. However, the definitions of place types vary across literature and are sometimes used interchangeably with neighborhood types. In this literature review, we examined the definition and usage of place types through two dimensions, such as the place type defined by the Smart Growth Area Planning (SmartGAP) tool.

In this report, the distinction is used pragmatically rather than as a strict methodological boundary. We use *place type* when referring to planning-oriented typologies such as the SmartGAP and ODOT frameworks that organize land use into named categories, often using area-type and development-type dimensions. We use *neighborhood type* when referring to the broader empirical literature that clusters areas based on built environment, land use, and sometimes socioeconomic characteristics. The two concepts overlap in practice, but this convention helps distinguish the policy-oriented typologies most relevant to VisionEval from the wider classification literature reviewed below.

1.1.1 History of "Place Type"

Place types defined in the SmartGAP tool have two dimensions: area type and development type. This classification method was first used in the Caltrans Smart Mobility Handbook (2010), and this report also classified place types based on two dimensions: community design and regional accessibility. Community design refers to the characteristics of development use, design, and location that encourage the multimodal transportation system at neighborhood and area scales, while regional accessibility extends the concept of community design to regional, interstate, and international scales (Caltrans, 2010). Seven place types were derived based on the strength and weaknesses of the elements constituting community design and regional accessibility. In addition, there are a total of 14 place types in terms of sub-classification.

SmartGAP is a scenario-planning tool that summarizes household and business characteristics within an area and uses them to estimate travel outcomes (Outwater, 2014). The tool defines the built environment as a set of 13 place types. The SmartGAP tool is similar to that suggested by Caltrans (2010) in that the two dimensions are defined by spatial scale, but there are differences in how to define each dimension, how to measure, and how to classify. Two dimensions consist of area types (urban core, clos-in community, suburban, and rural) and development patterns (residential, commercial, mixed-use, transit-oriented development, and greenfield). Area types are the types of areas where people or firms reside, and development patterns are classified according to population and employment density and urban form characteristics.

The Oregon Department of Transportation (ODOT) adopted the concept of place types measured in two-dimensional form in land use-built form, which is the input variable of scenario modeling such as the Regional Strategic Planning Model (RSPM) and Rapid Policy Analysis Tool (RPAT). Area types defined in these tools indicate how locations are situated with respect to the urban core and based on the three elements of destination accessibility, design, and density, areas are classified into four types: Regional center, Close-in-Community, Suburban/Town, and Low Density/Rural. Development type classifies the general nature of land use utilizing space. Specifically, it has six classifications: residential, employment, mixed-use, mixed-use-high, transit-oriented development, and low density/ Rural, using four indicators: density, design, diversity, and transit. A total of 13 place types were created by integrating these two dimensions.

1.1.2 Practical application of Place Type

Similar concepts to the place types defined above have been applied as primary space development inputs in various scenario-planning tools, such as UrbanFootprint and Envision Tomorrow. Within these tools, however, place types have flexible definitions compared to theoretical classifications. The UrbanFootprint model basically provides 35 Place Types using block resolution and 50 Building Types using parcel resolution. The building type can be newly defined according to the user's preference, and a new place type can be created by mixing several building types. Similarly, the Envision Tomorrow model consists of building prototypes and development types as primary inputs to spatial development planning. Users can create a specific building by entering building-related information (physical properties, financial properties, etc.) into an excel file of the building prototype. Users can also create development types consisting of different building prototypes to implement scenarios.

1.2 NEIGHBORHOOD TYPE

As more data becomes available, efforts are being made to classify neighborhood types by considering various variables beyond just urban design elements. Many studies define and develop their own neighborhood types and use them in their research. Each study derives different types by adopting different spatial units (such as census tracts, census block groups, and parcels) and variable configurations. By classifying neighborhood types, these studies aim to enhance our understanding of urban change and people's behavior, and to facilitate the formulation of policies and strategies for urban development.

1.2.1 History of Neighborhood types

According to Song and Knaap (2007), neighborhood typologies have been developed by urban designers to make better urban policies. Neighborhood types are classified by components that make up an area, such as built environment, land use, and street design. Based on these types of areas, we can identify the relationship between urban environments and people's behavior to help formulate appropriate policies. However, there is no official definition of neighborhood type yet, and there is no consensus on how to measure it (Song and Knaap,2007; Lin and Long, 2008; Frost et al., 2018). In particular, efforts to classify neighborhood types have been actively made in the field of transportation research. These papers measure the relationship between traffic behavior and built environment characteristics by neighborhood types to better understand people's behavior and improve the transportation system and urban environment (Lin and Long, 2008; Salon, 2015; Voulgaris et al. 2016). Song and Knaap's (2007) study began an attempt to quantitatively classify neighborhood types. Since then, with the advent of various data, a different number of neighborhood types using different spatial units have evolved.

Song and Knaap (2007) aimed to improve the classification of neighborhood types using quantitative methods. This study was the first to classify neighborhood types from rural to urban. In particular, this study differs from other studies in that it sets individual parcels as spatial units rather than using fixed spatial units such as census tracts, traffic analysis zones (TAZs), and zip codes. This reduces the bias caused by the unit of analysis. In addition, 21 measures of urban form were calculated, and neighborhood types were classified using statistical methodologies such as factor analysis and cluster analysis. Here, factor analysis means reducing many related variables into a smaller number of underlying dimensions, while cluster analysis means grouping places that look similar across those dimensions. Using these classified neighborhood types, this study identified spatial patterns of newly built single-family home development in the Portland Oregon metropolitan area.

Shay and Khattack (2007) created a neighborhood typology to identify how car ownership and travel behavior differ between neighborhood types. This study states that it is necessary to explain people's choice of neighborhood type in order to understand the long-term trend of travel behavior, and emphasizes the need for neighborhood type classification. Neighbor typologies were classified through factor and cluster analysis using 34 measures at the block group level. This study employed a more extensive range of variables than previous research to classify neighborhood types, incorporating additional factors such as socio-economic scale and employment density.

Lin and Long (2008) redefined neighborhood types and explored the relationship between household travel and neighborhood environments across the United States. The author considered the neighborhood to be synonymous with the census tract, concluding that the definition of census tracts ("Census tracts … are designed to be homogeneous with respect to population characteristics, economic status, and living conditions.") provides the most accurate representation of neighborhood characteristics. Neighborhood types are defined as clusters of census tracts that share similar socioeconomic, demographic, and land use characteristics. In contrast to previous studies, this research incorporated more comprehensive and detailed socioeconomic characteristics. Moreover, this study utilized log-likelihood clustering, a clustering method that can handle both continuous and categorical variables, to classify neighborhood types without first reducing the variables through factor analysis.

Salon et al. (2014) conducted a detailed investigation of changes in Vehicle Miles Traveled (VMT) in relation to changes in California's land use and transportation system characteristics, utilizing neighborhood types. Their findings confirmed that travel choices are strongly influenced by the built environment in specific types of neighborhoods. This study utilized the same methodology as Song and Knapp (2007) for classifying neighborhood types, which involved factor analysis and K-means cluster analysis. However, the variables used in this study were different, and the spatial unit of the neighborhood was based on the census tract. The content of Salon (2015) is similar to that of the report by Salon et al. (2014), but there are differences in how variables are composed to classify neighborhood types. Furthermore, Salon et al. (2014) utilized census data from 2000, whereas Salon (2015) used data from 2010. Due to the dissimilar variable configurations and data years used, the two studies arrived at different numbers of neighborhood types. Specifically, in Salon (2015), Urban and Suburb classifications have more detailed sub-classifications than in Salon et al. (2014). For instance, Urban is further divided into Urban Low Transit Use and Urban High Transit Use, whereas Suburb is categorized into multifamily and single-family housing types in Salon et al. (2014).

Ralph et al. (2016) developed a neighborhood typology to examine the relationship between Millennials' travel behavior and the built environment. The study excluded socioeconomic variables from the neighborhood type classification to focus solely on the relationship between travel behavior and the built environment. The methodology used in the study involved factor analysis and cluster analysis, combined with a latent class model, a statistical method used to identify unobserved subgroups in the data, that incorporated categorical variables such as traveler type. In a similar vein, Voulgaris et al. (2017) aimed to identify the impact of the built environment on travel behavior and emphasized the importance of utilizing neighborhood types including various built environment variables. This study focused on physical characteristics while excluding socioeconomic factors, and employed the commonly used factor analysis and cluster analysis methods.

Frost et al. (2018) enhanced the classification methodology for typology and broadened the evaluation criteria to comprehensively analyze the impact of diverse urban forms on economic, social, and environmental results across all census tracts in California. This paper utilized several techniques to classify neighborhood types, including Principal Component Analysis, Cluster Analysis, Threshold analysis, and Overlay analysis. Principal Component Analysis is another dimension-reduction method similar in purpose to factor analysis. Threshold analysis was used to identify cutoff values that help distinguish one type from another, and overlay analysis was used as a visual check on whether the resulting types made geographic sense. Compared to previous studies that relied solely on factor analysis and cluster analysis, this methodology is more sophisticated. The authors clarified the differences between each type through threshold value analysis, and visually evaluated the validity of the classification method by overlaying the types.

1.2.2 Variables/ Data Resources

The variables used for classifying neighborhood types can be broadly classified into seven types (see table 00). The first is population and employment density factors. This group of variables includes housing density, population density, and employment density, and Lin and Long (2008) consider detailed variables such as income, age, occupation, and race. Land-use mix factors are the second group of variables. These variables determine the mix of land uses by calculating the number of jobs by industry sector (e.g., retail, office, industrial, manufacturing, service, entertainment, etc.) or area occupancy by usage. The third is building and regional factors, which include variables such as the current state of use of the building, the age of the building, the value of the building, and the proportion occupied by single-family or multi-family dwellings. The fourth through sixth groups of variables are related to transportation. Street and Road design factors include variables such as road density, intersection density, and pedestrian density to indicate the degree of road connectivity in the area. Travel mode factors include the number or percentage of users by transport mode (e.g., car, transit, walk, and bike) or the distance to public transport stops, the number of stops, and the density of public transport services. Regional accessibility factors measure the number of jobs or other facilities (such as restaurants, supermarkets, etc.) that can be reached by public transportation or car within a specific time, or measure the distance to the facilities. Lastly, the natural environment factors consist of the area of open areas and the area of the tree canopy.

The variables mentioned above are derived from various sources, including Census data, the Census Transportation Planning Package, the EPA Smart Location Database, the MapQuest API, Metro's Regional Land Information System, the American Community Survey, the Charlotte-region travel survey, and data provided by ESRI Corporation. These variables were created by either directly using raw data or applying secondary processing.

Taken together, this review informs the proposed VisionEval classification method by combining the strengths of the two traditions. In this project, we call the new classification **land use type** to distinguish it from existing place type frameworks and from the broader neighborhood type literature. The proposed method largely retains the policy-facing categories used in place type systems, but defines them using a more data-driven approach that is more common in neighborhood type research. As discussed in Section 5.1, this allows the new land use type definition to stay recognizable for planning applications while grounding the category definitions in observed data and travel-related outcomes.

1.3 METHODS

Papers that classify neighborhood types typically use factor analysis and cluster analysis. Factor analysis and principal component analysis involve identifying common patterns across many related variables and reducing them to a smaller number of summary dimensions. This helps address collinearity while retaining much of the key information in the original variables (Salon et al., 2014). Cluster analysis is then used to group places with similar characteristics. Many studies use K-means cluster analysis, which requires the researcher to specify the number of clusters in advance and then compares candidate solutions. The log-likelihood clustering method is also used because it can handle both continuous and categorical variables, unlike K-means, which only works with continuous variables (Lin and Long, 2008). Frost et al. (2018) developed a more sophisticated method that uses threshold analysis and overlay analysis to clarify differences between neighborhood types and evaluate their validity.

2.0 MODEL DESIGN AND USE CASES

2.1 PURPOSES OF VELANDUSE PACKAGE AND THE GOAL AND OBJECTIVES OF THE RESEARCH PROJECT

The VE model framework addresses Bzone-level land development and related matters using two different packages: the VELandUse package used for VE-RSPM models and the VESimLandUse package used for VE-State models. VE-State uses a different package than VE-RSPM because Bzone-level input requirements were deemed too onerous for large regional and statewide applications. Both packages serve the same functions and produce the same outputs, but the VELandUse package does so by modeling real Bzones while the VESimLandUse package models simulated Bzones. In summary, both packages do the following:

1. **Assign Housing Units and Jobs to Bzones:** Identify the Bzone locations of housing units by type (single-family, multi-family) and jobs by sector (retail, service, other);

2. **Assign Household Residences**: Assign the synthesized households to residence Bzones;

3. **Assign Worker Job Sites:** Assign household workers to jobsite Bzones;

4. **Calculate Bzone-Level Density, Diversity, and Destination Proximity Measures**: Calculate land use density, diversity, and proximity measures from the Bzone allocations of households and jobs that are important inputs to modules that simulate household vehicle ownership and multimodal travel.

5. **Calculate Bzone-Level Transportation Measures:** Process user inputs for transportation facility, service, and regulatory characteristics that are applied at a Bzone level including:

    1. Bike and pedestrian network characteristics,

    2. Parking availability and pricing;

    3. Travel demand management programs; and,

    4. Car service levels

    5. transit service

NOTE: In the implemented package design described in Chapter 6, the new VELandUse package replaces the old VELandUse and VESimLandUse packages for the land use functions covered by this project, while the functions in VETransportSupply remain separate and unchanged.

The goal of this project is to replace the existing VELandUse and VESimLandUse modules with a new VELandUse module that can be used for all applications. This will be achieved by meeting the following objectives.

2.1.1 Objective #1: Replace Synthetic Bzones in VE-State Models with Real Geography

Currently the VELandUse package is not used in VE-State models because it would be too onerous for users to provide inputs at the Bzone level for large regions or entire states. To eliminate this burden, the VESimLandUse package synthesizes Bzones and enables users to input values according to general area type categories that module functions then assign to the synthesized Bzones. The simulation of Bzones by modules in the VESimLandUse package expanded upon methods developed in the GreenSTEP model for assigning land use characteristics to households from likely distributions. Bzone simulation enabled large scale regional and state models to be implemented in the VisionEval model system with a minimal amount of additional code. However, since synthetic Bzones have no real locational attributes – other than that they are located within real Azones – modeling capabilities are limited in some important respects. For example:

1. Models can't account for Bzone spatial relationships such as the relationships between worker residence and jobsite.

2. Simulated relationships such as the simulation of proximity to jobs and households are less accurate than what would be the case with real Bzones.

3. Models are not able to use geographic information that might be informative such as topography and freeway proximity.

4. Simulated Bzones present challenges for developing scenario assumptions such as parking pricing and car service availability because the geographical extent of the assumptions can't be visualized or controlled.

The research project will remove these limitations by eliminating simulated Bzones. All models will have real Bzones. The problem of users having to compile inputs at the Bzone level for large regional or state models will be eliminated by allowing them to specify inputs by land use types (see below) rather than at the Bzone level. In addition, using real Bzones for all models will increase the functionality of large-scale models. For example, it will enable inter-region commuting to be modeled.

This change also improves how VE-State-style policy inputs can be represented. Instead of applying parking, demand-management, car-service, and similar assumptions only to synthesized Bzones within broad categories, users can continue to specify those policies by LocType or by the broader land use type dimensions and then have them applied to actual mapped Bzones. That makes the spatial extent of a policy scenario visible and easier to control, while still avoiding the need to hand-code every Bzone in a large model. The four-county Use Case #3 pilot follows this pattern by keeping the larger-scale VE-State controls but attaching land use and policy inputs to real Census Block Group Bzones, which in turn makes it possible to represent worker residence-job relationships that cross metropolitan or county boundaries.

2.1.2 Objective #2: Make the Bzone Allocations of Housing Units and Employment Sensitive to Transportation Facilities and Services

Currently VE models do not address the effects of transportation facilities and services on the distribution of housing units and jobs. In the case of VE-RSPM models, the user exogenously specifies the numbers of housing units by type and jobs by sector in each Bzone, and VE populates those housing units with households. These exogenous Bzone housing unit forecasts may or may not realistically reflect what is likely to occur given scenario assumptions about the transportation system and other conditions. For example, a housing forecast which assumes increasing development density over time may be inconsistent with concurrent assumptions of extensive freeway expansion.

Although the new VELandUse module will continue to allow users the option of specifying jobs and housing at a Bzone level (see use case #1 below), it will also allow users to model how development is likely to occur over time given scenario assumptions about population growth and the future provision of transportation facilities and services as well as other significant factors (see use cases #2 & #3 below).

The use case #2 & #3 plan-based land use models that will be developed in this research will include transportation facility and service metrics, to the extent that they are found to have discernible effects on land development patterns. Multimodal facility metrics now included in VE models (e.g. transit service levels, bike/pedestrian network extent, freeway extent) will be considered as well as other available metrics that are likely to have a discernible relationship to land development (e.g. proximity to freeway interchange).

This design requirement has implications for package organization as well as for model specification. Earlier design discussion considered creating a combined VETransportAndLandUse package, but the implemented approach described in Chapter 6 is different: the new VELandUse package replaces the older VELandUse and VESimLandUse packages, while VETransportSupply remains a separate package with unchanged functions. In that implemented structure, the new land use procedures use transport-supply outputs and newer explanatory variables such as distance to freeway interchanges and fixed-guideway transit, but the transport modules themselves are not relocated. The implementation chapter also shows that several legacy VELandUse functions remain in place, while the older place-type functions that are not part of this project scope are deprecated or superseded. That package structure keeps the report aligned with the current codebase rather than the earlier alternative package concept.

2.1.3 Objective #3: Support Use Cases at Different Scales and Levels of Specificity

The package will be designed to support three use cases that allow users to model at different geographic scales and levels of specificity (Table 2.1).

The interaction between user-specified and predicted land use information differs across the three use cases. In Use Case #1, the user specifies housing and employment directly by Bzone, so land use types are not the primary control on the simulation. In Use Case #2, the user specifies the land use type dimensions by Bzone, including location type, area type, and diversity type, and the model then allocates housing and employment conditional on those fixed land use inputs. In Use Case #3, the model starts from base-year conditions and predicts future land use types over time through the PredictLUTypes and PredictLocTypes steps; those predicted land use types are then used in the same downstream allocation steps that Use Case #2 applies to user-specified land use types. The pilot applications in Chapter 7 follow this same structure: the SKATS Use Case #2 scenarios replace direct Bzone housing and employment inputs with land-use-type-based inputs, while the four-county Use Case #3 scenario predicts future AreaType, DivType, and LocType before allocating housing and employment.

The transportation variables also differ by use case. In Use Case #1, transportation-related inputs are applied directly by Bzone and mainly include pedestrian-oriented network design (D3bpo4), parking supply and price, paid-parking and cash-out shares, commute-options participation, car-service level, and the transit service and road-mile assignments produced in the transport supply package. In the newer implementation used for Use Cases #2 and #3, transportation conditions can also be represented through discrete D3Lvl and D4Lvl classifications derived from the underlying D3 and D4 measures. In Use Case #2, those transportation and policy variables are specified by land use type rather than by Bzone, and the allocation models can also use transportation accessibility variables such as distance to a freeway interchange or distance to a fixed-guideway station, as reflected in the current model specification and pilot scenarios. In Use Case #3, the model uses the same land-use-type-based transportation inputs as Use Case #2, but it also allows transportation facility and service changes to influence predicted land use type transitions over time before housing and employment are allocated.

2.1.3.1 Use Case #1 - User Specifies Bzone Employment and Housing by Type

This use case is intended for situations where detailed Bzone-level land use inputs are already available and should be preserved directly, such as historical validation work or scenarios based on an adopted official forecast.

In this use case, users specify, at the Bzone level, numbers of housing units by type (SF & MF) and number of jobs by type (retail, service, other) for every Bzone. Models then assign households and worker job sites to Bzones as a function of those user inputs. Other user inputs assign related Bzone characteristics at the Bzone level. The user specifies these inputs for every forecast year. This is the current VE-RSPM modeling process that is not suitable for large regions and states because of the level of effort required to compile inputs at the Bzone level. It should be noted that although land use inputs may be provided for a sequence of years, the sequence of those inputs will have no effect on the modeling procedures. The results for a year will not be affected by the results for previous years in the sequence. It should also be noted that the distributions of housing and employment are unaffected by transportation facilities and services metrics. Despite these limitations, use case #1 is to be preferred in instances where VE is being used to model historical conditions or an ‘official’ land use forecast.

2.1.3.2 Use Case #2 - User Specifies Bzone Land Use Types

This use case is intended for scenario planning where stakeholders want to work with generalized land use patterns rather than full Bzone-level housing and employment inputs, while still allowing the model to respond to transportation conditions and other explanatory variables.

In this use case, users specify land use types at the Bzone level (see below for description of land use types). Users specify all other required inputs by land use type rather than by Bzone. Those inputs then get assigned to Bzones based on the user specified Bzone land use types. Use case #2 model procedures assign numbers of housing units by type and jobs by type to Bzones as a function of the user assigned land use types and other relevant characteristics such as transportation facilities and services. Then, as with use case #1, model procedures assign households and worker job sites to Bzones. This use case, by making model inputs less specific, will enable a model with real Bzones to be applied to large regions and states (see Objective #1 above). It will also make the allocations of housing and jobs sensitive to transportation facilities and services inputs (see Objective #2 above). Users will be able to override these modeled allocations where more detailed information on planned developments or regulatory development constraints need to be reflected in the model. It should be noted that although land use inputs may be provided for a sequence of years, the sequence of those inputs will have no effect on the modeling procedures. The results for a year will not be affected by the results for previous years in the sequence.

2.1.3.3 Use Case #3 - Models are Used to Predict Future Land Uses Subject to Constraints

This use case is intended for exploratory planning and policy analysis when future land use scenarios have not yet been mapped in detail and the goal is to understand what patterns of land use change are plausible under different growth, transportation, and constraint assumptions.

In this use case, models predict how land uses could plausibly change over time given starting conditions and user input assumptions regarding Azone population and employment changes, transportation facilities and services changes, and other user specified constraints (for example, limiting the transition of rural land to urban use). The nature of the constraints and how they are specified as inputs are yet to be determined. As with use case #2, other Bzone inputs are specified by land use type. They are assigned to Bzones based on the land use types assigned by use case #3 model procedures. Given the specified assumptions and constraints, models will predict how Bzone land use types might plausibly transition over time to a future end state that is consistent with those given input assumptions. The transitions that occur in one time period will be a function of Bzone attributes in the previous time period and input assumptions about household and employment growth occurring between the periods. This is an important difference between this use case and the other use cases. Whereas use cases #1 and #2 will allow land use patterns to be specified for any future year (e.g. 2050) without consideration of how those land use patterns would come to be, use case #3 will model future land use as a series of predicted land use transitions occurring over time starting from the base year.

For Use Case #2, user specified land use type constraints in the input steer the model rather than guarantees that a specified density or land use outcome will always be achieved at Bzone level. For Use Case #3, the models predict land use transitions based on current land use type, household and employment growth, and transportation conditions before allocating housing and employment to Bzones.

Unlike use case #1 where the user specified Bzone employment and dwelling units in the input, use case #2 and #3 allow the model to determine the number of housing units and jobs in each Bzone that respond to transportation accessibility, existing land use conditions, and growth assumptions, so a user-specified land use pattern may conflict with what those other factors support. Where that happens, we suggest an iterative resolution process: compare the simulated outputs with the intended land use pattern, then revise land use inputs, transportation inputs, growth assumptions, or policy constraints and rerun the scenario. For some applications, running a scenario through both Use Case #2 and Use Case #3 may be useful. The user can then compare the land use pattern from Use Case #2 with the land use pattern from Use Case #3 and decide which one is more appropriate for their application and revise the inputs accordingly.

Table 2.1: Comparison of the three VELandUse use cases by purpose, inputs, predicted outputs, and transportation role

Use Case	Primary purpose	User-specified items	Predicted items	Required inputs	Optional or additional inputs	Transportation variables used	Main outputs
#1 User specifies Bzone employment and housing by type	Historical validation and official forecast applications where detailed Bzone inputs should be preserved directly	Bzone housing units, Bzone jobs, and related Bzone characteristics for each forecast year	Household residence assignment and worker job-site assignment	Detailed Bzone housing, employment, household-income, parking, car-service, demand-management, urban-mix, and D3bpo4 inputs	Future enhancement noted in text: some Bzone inputs could potentially be specified by land use category instead	Transportation-related inputs can be supplied by Bzone, but they do not drive the distribution of housing and employment in this use case	Standard VELandUse outputs for households, workers, Bzone summaries, and downstream land use measures
#2 User specifies Bzone land use types	Scenario planning where users work with generalized land use patterns rather than full Bzone housing and employment inputs	Bzone land use types plus other inputs specified by land use type	Bzone housing and employment allocations, then household residence and worker job-site assignment	Bzone location, area, and diversity type; household income quartiles; and land-use-type inputs for parking, car service, demand management, urban mix, D3bpo4, and related variables	Users may override modeled allocations to reflect planned developments or regulatory constraints	Transportation facilities and services can influence housing and employment allocation; additional significant variables may also be included, such as distance to a freeway interchange	Standard VELandUse outputs plus simulated land use inputs needed by downstream VE modules
#3 Models predict future land	Exploratory planning and policy analysis	Base-year Bzone land use inputs,	Future land use type transitions	Base-year Bzone housing and employment,	Additional land use constraints or policy inputs	Transportation facility and service changes	All land use inputs required by downstream

Use Case	Primary purpose	User-specified items	Predicted items	Required inputs	Optional or additional inputs	Transportation variables used	Main outputs
uses subject to constraints	where future land use patterns are not yet mapped in detail	shared land-use-type inputs, growth assumptions, and user constraints or policy assumptions	over time, followed by housing and employment allocation and household and worker assignment	household income quartiles, transition-model variables, and the land-use-type inputs shared with Use Case #2	are anticipated but not yet fully specified	are part of the transition assumptions, along with other significant explanatory variables	VE modules, including future location, area, and diversity types plus Bzone housing and employment outputs

2.2 APPLICATION OF USE CASES

In the case of modeling historical conditions, use case #1 is most appropriate because housing and jobs data will be available and the results will most accurately reflect past conditions. Use case #1 will also be most appropriate for modeling the impacts of an 'official' forecast that represents an adopted land use plan or planning stakeholder consensus.

Use case #2 would be applied for modeling studies where use case #1 details are not available but where stakeholders have mapped out one or more general planning scenarios through a planning charrette process or other land use envisioning process. Those scenarios would be represented by the assignment of land use types to Bzones in the study area. The use case #2 models would allocate housing units by type and employment by type to Bzones, and then assign households to residence Bzones and workers to job site Bzones. The completion of a use case #2 study will simulate all the land use inputs required by other VE modules.

VisionEval users may carry out use case #2 studies iteratively because there is no guarantee that the distribution by land use types of resulting household and job assignments will match the land use type inputs. This could be the case for several reasons such as the following:

- The changes required for a land use scenario to be realized might entail more regional household and employment growth than is assumed to occur.
- The distributions of households and jobs envisaged by a scenario might not occur because changes to factors that significantly affect those distributions might not be substantial enough.

In an iterated process, the land use type outputs would be compared to the land use type inputs and adjustments would be made to the inputs to bring the outputs into better correspondence. In some cases the adjustments might take the form of scaling back scenario expectations, as when the amount of growth assumed to occur may be insufficient to realize substantial changes in existing land use patterns. In other cases, transportation and land inputs might be revised to create the pressures for changes from status quo patterns.

Use case #3 would be applied in cases where general or detailed land use scenarios have not been developed. It would most often start with running use case #3 model procedures with inputs that represent a 'business as usual' or 'planned trends' case. That would provide information about the land use patterns that could plausibly develop as a consequence of continuing expected trends. This would be followed by a series of model runs having different combinations of inputs for variables found to have significant effects on land development (e.g. transportation facilities and services) to provide information about the sensitivities of outcomes to inputs.

The information provided by a use case #3 study like this would be very useful for a planning charrette or other land use envisioning process. It would provide a set of Bzone land use type scenarios that stakeholders can modify in the envisioning process. It would also provide useful information about the types and degrees of changes to influential factors that might be necessary to achieve those scenarios. In this instance, the use case #3 study would be followed by a use case #2 study.

Alternately, since use case #3 procedures provide all the land use inputs required by other VE modules, the user may skip the use case #2 procedures and proceed directly to running other VE modules with those outputs.

2.3 GENERAL LAND USE TYPES

A consistent general land use type system is necessary in order to make full use of the national land use (Smart Location Database) and travel (National Household Travel Survey) datasets, and to create a geographically transferrable set of models for accomplishing the goal of replacing the existing VELandUse package with a new package that can be used for all VE models and that achieves the objectives outlined above. The proposed system classifies land use in 3 dimensions: location type, area type, and diversity type as follows:

- **Location Type** specifies the larger-scale location of the Bzone:
  - Urban (i.e., Census urban definition)
  - Town
  - Rural
- **Area Type** identifies the context of the Bzone within its location, indicating its development density and proximity to other Bzones in the location:
  - Regional Center
  - Center
  - Inner
  - Outer
  - Fringe
- **Diversity Type** identifies the composition of land uses in the Bzone and its immediate vicinity:
  - Balance of housing and jobs
  - Balance of job types

2.4 VELANDUSE PACKAGE OUTPUTS

2.4.1 Base VELandUse Package Outputs

The modules in the new VELandUse package will produce all the outputs produced by the current VELandUse modules. This is a necessity to avoid having to reestimate and reprogram other VE modules. Following is a listing of outputs organized by the modules that produce them.

- **PredictHousing** module assigns households spatially
    - Assigns households to a housing type (SF or MF) and Bzone.
    - Calculates Bzone totals of households in single family units, households in multifamily units, total households, total population, group quarters population, and workers.
- **LocateEmployment** module locates worker job sites
    - Creates a worker table that includes the following data for each worker: household ID; worker ID; job location Bzone, Azone, and Marea; distance between residence and work.
    - Scales service jobs, retail jobs, and total jobs by Bzone to match the number of workers.
- **AssignLocTypes** module identifies household location types
    - Identifies the location type of each household (urban, town, rural).
    - Identifies the name of the metropolitan area where the household resides.
    - Sums the population by location type for each Bzone
    - Sums the population by location type for each metropolitan area
    - Sums the household income by location type for each metropolitan area
- **AssignDevTypes** module assigns a development type (urban, rural) to each household based on whether the household is located in an urbanized area boundary. *Note that this development type does not have the same meaning or purpose as 'development type' in the place types typology. It predates the use of place types, coming from the early GreenSTEP model and is still used by some other VE modules. This module function might be eliminated to reduce possible confusion.*
- **Calculate4DMeasures** module calculates several land use metrics used by the household multimodal travel module and other VE modules
    - Calculates land use density measures including population density (D1B), employment density (D1C), and activity density (D1D).
    - Calculates land use diversity measures including the ratio of jobs to households (D2A_JPHH), the ratio of workers living in the zone to jobs located in the zone (D2A_WRKEMP), and the entropy of activities in the zone: households, retail employment, service employment, other employment (D2A_EPHHM).

- Calculates a destination proximity measure (D5) which is the harmonic mean of jobs within 2 miles and population within 5 miles of the Bzone centroid.
- Assigns the pedestrian-oriented network design values (D3bpo4) to Bzones from user input values.

- **CalculateUrbanMixMeasure** module assigns an "urban-mix" neighborhood attribute corresponding to the urban category of the urban/rural indicator measure in the 2001 National Household Travel Survey. This measure is used by the original (not multimodal) household travel module and by several other modules.
    - Estimates an urban-mix neighborhood probability for each Bzone.
    - Assigns households to urban mixed neighborhood type as a function of the probability of the Bzone in which they are located.
    - Allows users to set an urban mix probability (proportion) target for each Bzone.
- **AssignParkingRestrictions** module identifies parking restrictions and prices affecting households at their residences, workplaces, and other places they are likely to visit in the urban area. Assignments are made based on user inputs by Bzone.
    - Number of free parking spaces per single family dwelling unit
    - Number of free parking spaces per multi-family dwelling unit
    - Number of free parking spaces per group quarters resident
    - Proportion of non-work trips to the Bzone which pay for parking
    - Proportion of workers working at jobs in the Bzone who pay for parking
    - Proportion of worker paid parking in cash-out_buy-back program
    - Average daily parking cost
- **AssignDemandManagement** module assigns demand management program participation to households and to workers based on user inputs.
    - Assign households to individualized marketing program participation based on input targets for their residence Bzones.
    - Assign workers to strong employee commute options program participation based on input targets for their jobsite Bzones.
- **AssignCarSvcAvailability** module assigns a level of car service available to each household (high or low).
    - Assign a car service availability to each household based on Bzone input values.

2.4.2 New VELandUse Package Outputs

The following new outputs will be produced by modules in the new VELandUse package:

1. **Bzone Land Use Types:** In addition to producing the outputs listed above, the new VELandUse procedures will save to the model datastore the location type, area type, and diversity type of each Bzone. That information is required in order to implement use cases #2 and #3. It will also allow VE outputs to be stratified by land use as well as transportation measures. This can highlight the different travel behavior outcomes related to the combination of land use and transportation factors.

2. **Numbers of single-family dwelling units, multi-family dwelling units, and group quarters units by Bzone**: While this is an input for use case #1 modules, it is an output for use case #2 & #3 modules. Outputting these data from use case #2 and #3 modules will enable the existing PredictHousing and Locate Employment modules to be used with few or no changes.

3. **Total employment, retail employment, and service employment by Bzone:** Same note as for #2 above.

4. **Bzone dwelling unit proportions by location type (urban, town, rural is remainder) for each housing type (single family, multi-family, group quarters):** Same note as for #2 above.

2.5 VELANDUSE PACKAGE INPUTS

2.5.1 Inputs Common to All Use Cases

In a significant departure from existing VE procedures, all future use cases will use real Bzone geography. Currently only the VE-RSPM model uses real Bzone geography. The VE-State model simulates Bzones within Azones. Changing to real Bzone geography will enable a consistent set of model procedures to be shared across all model scales. For example, the procedures of the LocateEmployment module that allocates workers to job sites in VE-RSPM models can also be used to model intercity commuting in larger scale regional models. In addition, real geography is necessary to model the effects of transportation services and facilities on land use. Therefore, all use cases for all model scales will have the following geographic inputs:

- Geographic coordinates (latitude, longitude) of Bzone centroids; and,

- Developable areas that are urban, town, and rural location types by Bzone. In the case of larger scale models like large metropolitan area models and state models, it is likely that this input will be simplified by assigning only one location type to each Bzone.

2.5.2 Use Case #1 Inputs

The use case #1 inputs are as follows:

1. Numbers of single-family dwelling units, multi-family dwelling units, and group quarters units by Bzone.

2. Total employment, retail employment, and service employment by Bzone.

3. Bzone dwelling unit proportions by location type (urban, town, rural is remainder) for each housing type (single family, multi-family, group quarters)

4. Bzone household (non-group quarters) proportions by corresponding Azone income quartile.

5. Pedestrian-oriented network design values (D3bpo4) by Bzone.

6. Targets for the proportion of households in 'urban-mixed' neighborhoods by Bzone.

7. Average number of free parking spaces for each single-family dwelling unit, for each multi-family dwelling unit, and for each group quarters resident by Bzone.

8. Daily parking price by Bzone.

9. Proportion of workers paying for parking and proportion of workers participating in a cash-out-buy-back program by Bzone.

10. The proportion of non-work trips to Bzone paying for parking.

11. The proportion of workers working in Bzone who participate in a strong employee commute options program.

12. The proportion of households residing in Bzone who participate in a strong individualized marketing program.

13. The level of car service (low or high) in each Bzone.

It might be useful to users to allow specification of inputs 3-13 by general land use category instead of by Bzone. This could greatly simplify the process of developing user inputs. Since the use case #2 and #3 process will enable this, it might not take much additional work to enable it for use case #1.

2.5.3 Use Case #2 Inputs

The use case #2 inputs are as follows:

1. Location type, area type, and diversity type by Bzone.

2. Bzone household (non-group quarters) proportions by corresponding Azone income quartile.

3. Pedestrian-oriented network design values (D3bpo4) by land use type.

4. Targets for the proportion of households in ‘urban-mixed’ neighborhoods by land use type.

5. Average number of free parking spaces for each single-family dwelling unit, for each multi-family dwelling unit, and for each group quarters resident by land use type.

6. Daily parking price by land use type.

7. Proportion of workers paying for parking and proportion of workers participating in a cash-out-buy-back program by land use type.

8. The proportion of non-work trips paying for parking by land use type.

9. The proportion of workers who participate in a strong employee commute options program by jobsite land use type.

10. The proportion of households who participate in a strong individualized marketing program by residence land use type.

11. The level of car service (low or high) by land use type.

12. Inputs for other variables found to be statistically significant in modeling the location of housing and employment (e.g. distance to freeway interchange)

2.5.4 Use Case #3 Inputs

Use case #3 will have the following inputs for the model base year only. Those inputs will not be necessary for other model years because values for the previous model year will be used.

1. Numbers of single-family dwelling units, multi-family dwelling units, and group quarters units by Bzone.

2. Total employment, retail employment, and service employment by Bzone.

3. Bzone dwelling unit proportions by location type (urban, town, rural is remainder) for each housing type (single family, multi-family, group quarters)

4. Bzone household (non-group quarters) proportions by corresponding Azone income quartile.

5. Inputs for other variables found to be statistically significant in modeling the transition of land use types.

The following use case #3 inputs will be the same as the corresponding using case #2 inputs.

6. Targets for the proportion of households in ‘urban-mixed’ neighborhoods by land use type.

7. Average number of free parking spaces for each single-family dwelling unit, for each multi-family dwelling unit, and for each group quarters resident by land use type.

8. Daily parking price by land use type.

9. Proportion of workers paying for parking and proportion of workers participating in a cash-out-buy-back program by land use type.

10. The proportion of non-work trips paying for parking by land use type.

11. The proportion of workers who participate in a strong employee commute options program by job site land use type.

12. The proportion of households who participate in a strong individualized marketing program by residence land use type.

13. The level of car service (low or high) by land use type.

14. Inputs for other variables found to be statistically significant in modeling the transition of land use types.

In addition, use case #3 will have yet to be specified inputs which represent land use constraints/policies that influence the assignment of dwelling units by type and employment by type. Examples of such constraints/policy inputs include constraint on urban expansion, preferences for greater densities, and preferences for greater expansion.

3.0 DATA PROCESSING AND EXPLORATION

This chapter documents the data sources and processing steps used in our analysis. Our data pipeline combines multiple national datasets to create a comprehensive dataset at the block group level, for circa 2010 and 2017 respectively. The 2009 and 2017 NHTS are processed and then joined with block group-level data.

3.1 DATA PROCESSING

Table 3.1summarizes the main source datasets used in the pipeline, Table 3.2 summarizes the main groups of processed variables carried forward into the exploratory analysis and later model-estimation steps, and Table 3.4 provides a compact bridge from those processed variables to the major VELandUse model components described later in Chapters Chapter 5 and Chapter 6. These tables are intended as reader guides; later chapters provide more detailed discussion of exact specifications and estimation results.

Table 3.1: Main source datasets used in the data-processing pipeline

Source	Geographic level	Time period used	Main variables or products used in this project
Census geography and data	Census block, block group, urban area	2000, 2010, 2020 geography; circa 2010 and 2020 attributes	Bzone geography, population, housing, household characteristics, urban area definitions, crosswalks across vintages
Smart Location Database (SLD) v2 and v3	Block group	circa 2010 and circa 2020	Built environment measures including D1 density, D2 diversity, D3 design, D4 transit, D5 accessibility, and related buffer-based metrics
LEHD LODES	Census block, aggregated to block group	2010 and 2017	Workplace and residence employment, sector-group employment, and block-to-block employment patterns used to build Bzone employment measures
National Transit Database (NTD)	Urbanized area	2009 and 2017	Transit supply measures including annual vehicle revenue miles and related system-level service indicators
Highway Performance Monitoring System (HPMS)	Urbanized area	2008 and 2017	Freeway lane miles and related roadway supply measures
National Household Travel Survey (NHTS)	Household, linked to residential block group	2009 and 2017	Household travel outcomes including daily VMT and other household, person, vehicle, and trip variables used in exploratory analysis
Fixed-guideway station inventory and transit-stop data	Point locations	contemporary network files matched to analysis periods	Distance to transit stops and fixed-guideway stations
OpenStreetMap and related network processing	Network and point locations	contemporary network files matched to analysis periods	Distance to freeway ramps and related roadway-access variables
USGS elevation data	Raster, aggregated to block group	contemporary elevation surface	Steep-slope share and other terrain constraints
HHS poverty guidelines	Household-size thresholds	2009 and 2017 applicable thresholds	Poverty classification used for NHTS households

Table 3.2: Main groups of processed variables used in the data chapter and later model-development steps

Variable group	Representative variables	Main role in this report
Geography and crosswalks	GEOID, block-to-block-group links, 2010-2020 crosswalks, urban area relationships	Define analysis Bzones, harmonize geography across years, and support joins across source datasets
Household and demographic measures	population, households, household size, income quartile, poverty status	Describe Bzone population and support travel-behavior stratification
Housing and land development measures	dwelling units, housing type, developable area, urban/town/rural location shares	Support land use classification and later allocation of housing to Bzones
Employment and activity measures	total employment, retail employment, service employment, workplace and residence employment	Support employment allocation, diversity measures, and job-housing relationships
Built environment intensity and mix measures	D1B, D1C, D1D, D2A_JPHH, D2A_EPHHM, D2C_WREMIX	Describe density and diversity conditions used in classification and model estimation
Transportation supply and accessibility measures	UZAAVRM, freeway lane miles, D4A, D4C, DistToStop, DistToFgwSta, distance to ramps	Represent transit, roadway, and access conditions used in exploration and later model inputs
Design and walkability measures	D3BPO4, intersection-based design measures, urban mix indicators	Represent pedestrian-oriented design and related neighborhood form
Terrain and physical constraints	steep slope share, developable land indicators	Capture physical constraints on development patterns
Travel outcome measures	household DVMT, day-of-travel summaries, mode-related trip totals	Provide dependent variables for exploratory travel analysis and comparison across place types

Table 3.3: Summary of variables used in each major VELandUse model component

Model component	Main outputs	Main variables used	Main source datasets
LoadLUType	Base or user-specified AreaType and DivType by Bzone	Bzone identifiers, user-supplied land use type assignments, geography crosswalks where needed	Census geography; scenario input files
PredictLocType	Future LocType (Urban, Town, Rural)	Current LocType, Bzone geography, development context, and transition-model predictors derived from SLD and related geographic variables	Census geography, SLD, processed geographic attributes
PredictLUType	Future AreaType and DivType	Current AreaType, DivType, population, households, employment, density and accessibility measures such as D1D_hmcbuf, D5, PctSteepSlope, DistToRamp, DistToCBD, and DistToFgwSta	SLD, Census, LEHD, OSM/network processing, fixed-guideway stations, elevation data
AssignD3D4Levels	Discrete D3Lvl and D4Lvl classes	Design and transit predictors including D1D_hmbuf, D2A_JPHH_hmbuf, D3BPO4, D4A, D4C, DistToStop, DistToFgwSta, and HasFgwTransit	SLD, transit-stop data, fixed-guideway station data, transit-service processing
AllocateDU	SFDU, MFDU, GQDU by Bzone	LUType plus Bzone accessibility and development-pattern variables, with LUType-level dwelling-unit control totals	SLD, Census housing measures, scenario input files
AllocateEmployment	TotEmp, RetEmp, SvcEmp, and other employment allocations by Bzone	LUType plus Bzone accessibility and development-pattern variables, with sector-group employment control totals and prior/base-year employment distributions	LEHD, SLD, Census, scenario input files
PredictHousing and LocateEmployment	Household residence Bzones, worker job-site Bzones, work-distance measures	Allocated dwelling units and jobs by Bzone, household synthesis outputs, worker and household attributes, Bzone and Marea geography	Census, household synthesis outputs, allocated land use outputs

Model component	Main outputs	Main variables used	Main source datasets
CalculateUrbanMixMeasure	Urban-mix indicators by Bzone	Bzone household and employment totals, proximity measures, and assigned land use inputs	SLD-style built environment measures, allocated Bzone totals, processed geography
Existing policy-assignment modules (AssignParkingRestrictions, AssignDemandManagement, AssignCarSvcAvailability)	Parking, TDM, and car-service attributes by household, worker, or Bzone	User-specified or land-use-type-based policy inputs linked to Bzones and land use types	Scenario input files, Bzone geography, land use type assignments

\One point that needs to be explicit is the difference between the diversity variables carried in the processed datasets and the diversity variables actually emphasized in the current model specification. In this report, the main diversity variable used for land use type definition and several related model steps is a job-balance measure, especially D2A_JPHH and its buffered variants such as D2A_JPHH_2mbuf and D2A_JPHH_hmbuf. Those variables describe the balance between jobs and households and are used directly in the diversity-type definition and in the AssignD3D4Levels module. By contrast, broader employment-mix or entropy-style measures such as D2A_EPHHM and worker-employment mix measures such as D2C_WREMIX are retained in the processed data and remain useful as descriptive diversity measures, but they are not the primary diversity variables driving the current land use type definition documented in Chapters Section 5.2 and Chapter 6.

Because the D3 and D4 variables appear repeatedly across later chapters, Table 3.4summarizes the main measures used in the current report, their units, and how they should be interpreted. Appendix 1 (Chapter 9) provides the glossary-style definitions for the individual variables, while the model and implementation chapters explain how these continuous measures are converted into the discrete D3Lvl and D4Lvl classes.

Table 3.4: Main D3 and D4 variables used in the report, with units and interpretation

Variable	What it represents	Units or scale	Interpretation in this report
D3BPO4	Pedestrian-oriented four-leg intersection density	Intersections per square mile	Higher values indicate a finer-grained, more walkable street network and stronger urban-design support for walking
D1D_hmbuf	Buffered activity density used with D3 classification	Households plus employment intensity in a half-mile buffer	Higher values indicate denser surrounding activity and strengthen higher D3 classifications when combined with design measures
D2A_JPHH_hmbuf	Buffered jobs-per-household balance used with D3 classification	Ratio	Values near balance help distinguish mixed, walk-supportive settings from predominantly residential or employment-only settings
D4A	Distance from population-weighted centroid to the nearest transit stop	Meters in the SLD source data	Lower values indicate better stop proximity and generally support higher transit-service classification
D4C	Aggregate transit service frequency near the block group	Vehicles per hour within 0.25 miles during the evening peak	Higher values indicate more transit service and generally support higher D4 classifications
DistToStop	Distance to the nearest transit stop	Miles in the current model implementation	Lower values indicate better local transit access
DistToFgwSta	Distance to the nearest fixed-guideway station	Miles in the current model implementation	Lower values indicate better access to higher-capacity transit service
HasFgwTransit	Indicator for fixed-guideway transit availability in the metropolitan area	Binary indicator	Distinguishes places with and without access to a fixed-guideway transit system context
D3Lvl	Discrete urban-design class predicted by AssignD3D4Levels	Ordinal levels 1 to 7	Higher levels indicate stronger urban-design conditions supportive of walking
D4Lvl	Discrete transit-service class predicted by AssignD3D4Levels	Ordinal levels 1 to 7	Higher levels indicate stronger transit-service conditions

The thresholds for the discrete D3Lvl and D4Lvl classes are not fixed by hand in this data chapter. Instead, they are estimated in the decision-tree procedures documented in Chapter Chapter 5 and implemented in AssignD3D4Levels in Chapter Chapter 6. In other words, Chapter 3 documents the continuous input measures and their units, while the later chapters document the empirical cutoffs used to translate those measures into ordinal design and transit classes.

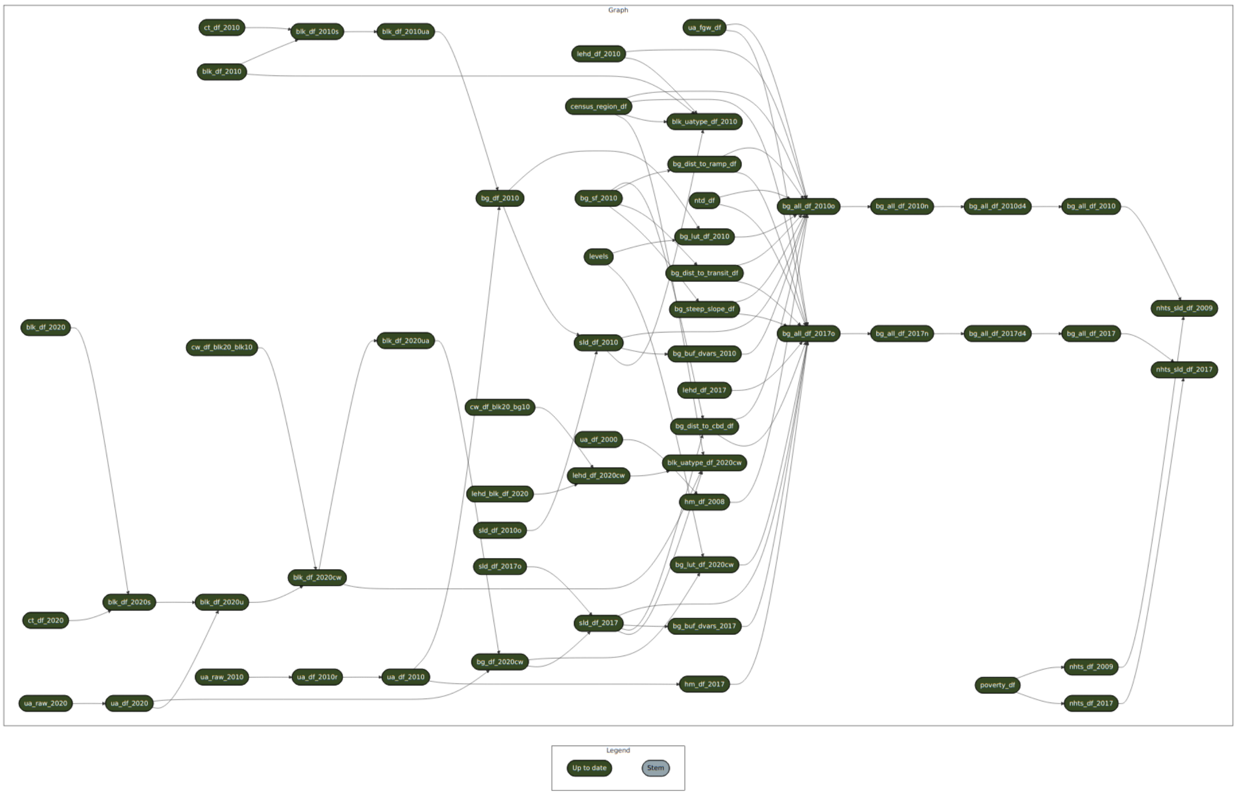

Figure 3.1: Data processing pipeline for combining Census, SLD, NHTS, NTD, and HPMS data sources

Figure 3.1 shows the complete data processing pipeline. The pipeline consists of three main stages:

1. **Data Collection** (top): Individual data sources are loaded and processed separately, including:
    - Urban Areas (UA) data from 2000, 2010, and 2020
    - National Transit Database (NTD)
    - Highway Performance Monitoring System (HPMS)
    - Census block and block group geography and data
    - Smart Location Database (SLD) versions 2 and 3
    - LEHD data (2010 and 2020)

- Fixed Guideway Transit Stations by Center for TOD

2. **Geographic Processing** (middle): Various spatial calculations and joins are performed:
    - Distance calculations (distance to ramps, CBD, transit stops, fixed guideway stations)
    - Block-to-UA relationships
    - Census geography crosswalks
    - Census Block Urban Area type prediction and aggregation
    - Steep slope computation
3. **Integration** (bottom): All data sources are combined into two comprehensive datasets:
    - 2009 period: 2010 CBG dataset, as well as joined NHTS 2009 with circa 2010 block group data
    - 2017 period: 2020 CBG dataset, as well as joined NHTS 2017 with circa 2020 block group data

The arrows in the visualization show data dependencies, with each node representing a discrete processing step. Interactive elements allow exploration of specific pipeline components and their relationships.

The data pipeline is implemented using the targets R package (Landau, 2021), which provides a Make-like pipeline toolkit for reproducible data processing. Key features of our implementation include:

- **Modularity**: Each processing step is implemented as a separate R function
- **Dependency Management**: Automatic tracking of data dependencies
- **Caching**: Intelligent caching of intermediate results
- **Parallel Processing**: Capability to run independent steps in parallel
- **Reproducibility**: Version-controlled pipeline definition and processing functions

The pipeline configuration is stored in _targets.R, with individual processing functions organized in the code/data-pipeline/ directory. This setup ensures reproducibility and makes it easy to update specific components when new data becomes available.

3.1.1 Urban Areas Data

3.1.1.1 Processing Steps

1. Load raw UA boundary files for each decade
2. Reclassify 2010 UAs using 2020 criteria
3. Categorize UAs into Urbanized Areas (UZA) and Urban Cluster (Town) based on UA population.

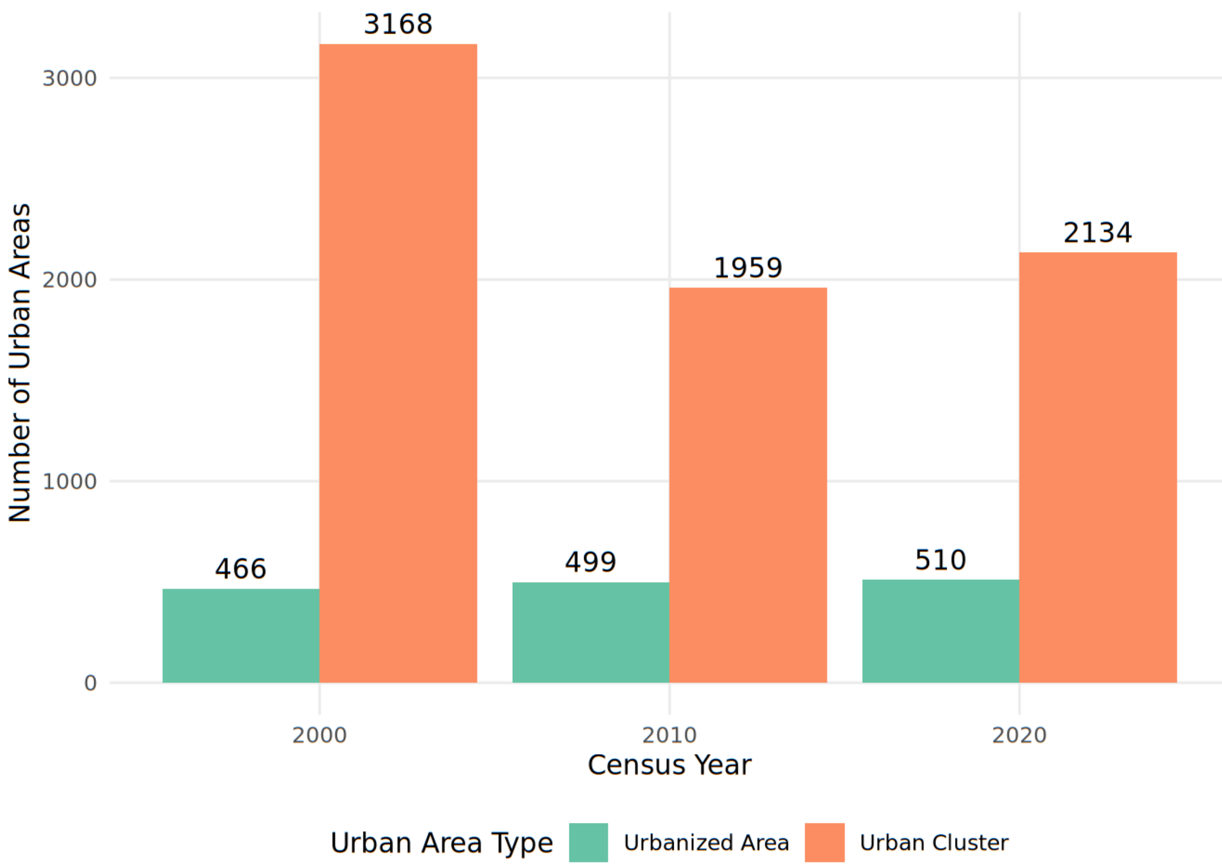

Figure 3.2: Number of Census-designated Urbanized Areas and Urban Clusters by year (2010 and 2020)

Figure 3.2 shows the distribution of Urban Areas by type across three Census years. The number of Urbanized Areas (population > 50,000) has steadily increased from 2000 to 2020, while Urban Clusters (population ≤ 50,000) show a notable decrease. Note that the Census Bureau revised urban area criteria in each of these Census years.

3.1.2 Transportation Infrastructure Data

3.1.2.1 National Transit Database (NTD)

The National Transit Database provides transit system data for urban areas including Operating Expenses (VehOp), Vehicles Operated in Maximum Service (VehAv), Annual Vehicle Revenue Miles (AVRM), Annual Vehicle Revenue Hours (AVRH), Unlinked Passenger Trips (UPT), and Passenger Miles (PM).

Figure 3.3 shows the distribution of Annual Vehicle Revenue Miles (AVRM) by year.

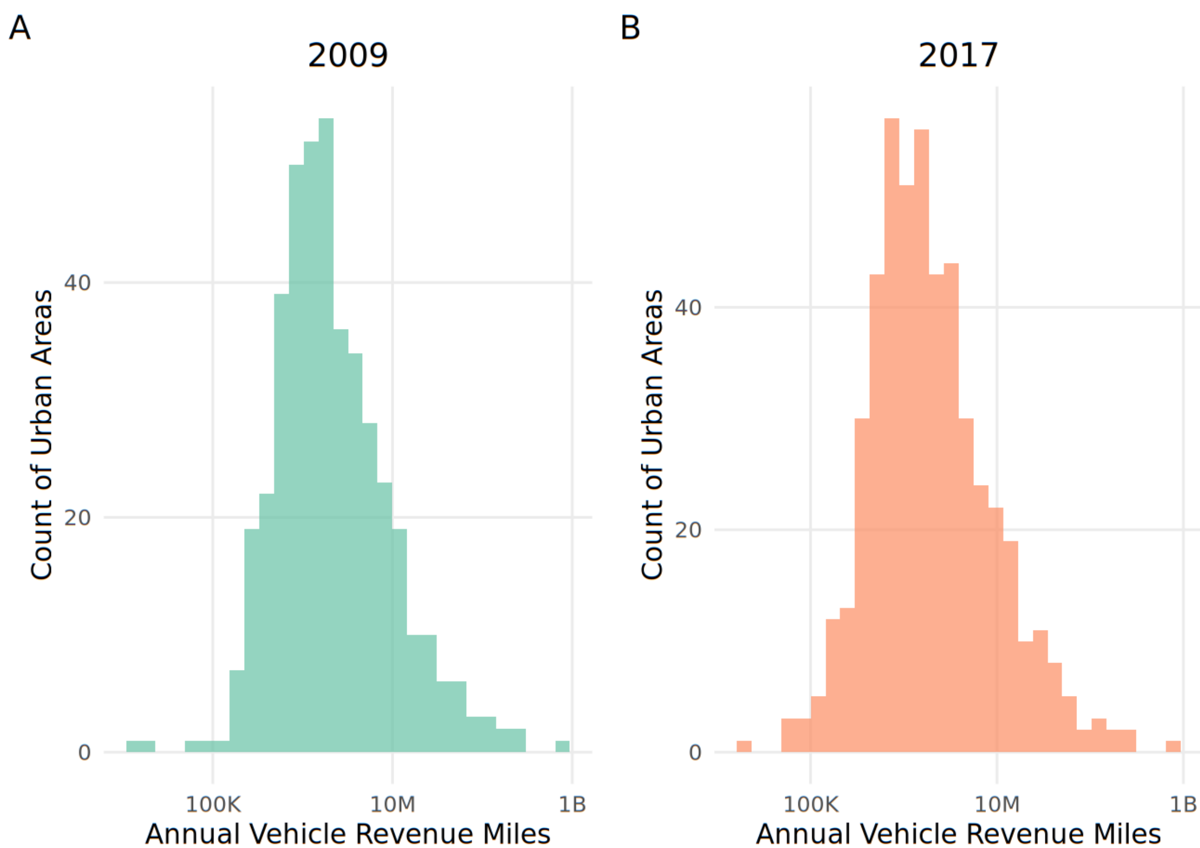

Figure 3.3: Distribution of annual transit vehicle revenue miles per capita by year (2009 and 2017)

3.1.2.2 Highway Performance Monitoring System (HPMS)

HPMS provides a number of UZA level data on roadway (e.g. Freeway Lane Miles) and travel (e.g. AADT, DVMT). We focus on freeway lane miles for 2008 (since 2009 table is not available, 2008 is the year closest to 2009 NHTS) and 2017 from HPMS table HM-72.

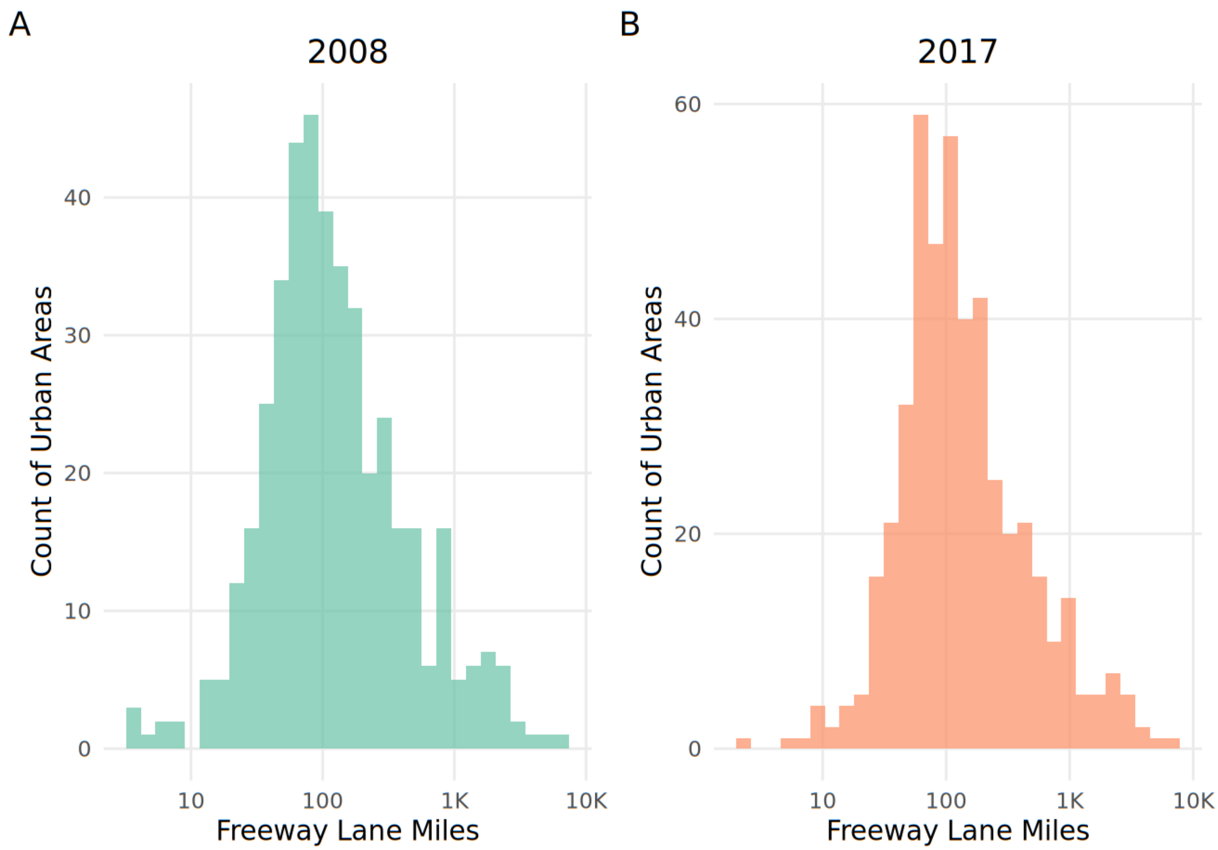

Figure 3.4: Distribution of freeway lane miles per capita by year (2008 and 2017)

Figure 3.4 shows the distribution of freeway lane miles across urban areas in 2008 and 2017. The distributions are right-skewed, with most urban areas having relatively few freeway lane miles and a small number of large urban areas having substantially more. Comparing with the transit service distribution in Figure 3.3, we can see that both transportation infrastructure measures follow similar patterns, reflecting the size distribution of urban areas in the United States. The distributions remain relatively stable between the two time periods, suggesting that the relative differences in transportation infrastructure provision across urban areas have persisted over time.

3.1.3 Geographic Data Processing

3.1.3.1 Census Geographies

We use multiple Census geographic levels:

1. Block level data (2010, 2020)

2. Block group level data (2010)

3. Crosswalks between 2010-2020 geographies

4. Cenus regions and divisions

3.1.3.2 Smart Location Database (SLD)

The EPA's Smart Location Database provides US nationwide built environment measures. We use both version 2 (circa 2010) and version 3 (circa 2017).

3.1.3.3 Distance Calculations

Several distance variables are computed for each block group:

1. Distance to highway ramps (using latest OSM data)
2. Distance to Central Business District (CBD), using city hall locations and SLD employment centers
3. Distance to transit stops and fixed guideway transit stations

3.1.3.4 Terrain/Steep Slope

We use USGS elevation data to compute steep slope for each block group and aggregate to get percentage of block group area with steep slope.

3.1.4 LEHD Origin-Destination Employment Statistics (LODES)

Workplace and residential employment data from the Census LEHD program:

- Job counts by census block
- Employment by sector group, particularly retail, service, and entertainment sectors
- Data from 2010 and 2017

Because LEHD workplace locations are confidentiality-protected, the published block-level employment locations are not exact point representations of true job sites. In practice, this means the LEHD-based employment patterns used in this report should be interpreted as approximate spatial distributions rather than parcel- or address-precise employment maps. That caveat matters most when later chapters show mapped employment patterns at fine geography, including the pilot Bzone maps, even though aggregation to block groups reduces some of the visual noise.

3.1.5 Travel Behavior Data

3.1.5.1 National Household Travel Survey (NHTS)

The 2009 and 2017 NHTS data is loaded and joined with confidential residential block group GEOID. Select person, vehicle, and trip level variables are aggregated to the household level.

3.1.5.2 Poverty Guidelines

Federal poverty guidelines by household size from HHS is used to classify NHTS households whether their income is below the federal poverty line.

3.2 DATA EXPLORATION

This section explores key patterns in our processed datasets, focusing on household travel behavior and its relationship with built environment characteristics. We examine both temporal changes between 2009 and 2017, and spatial variations across different urban contexts.

3.2.1 HH VMT in 2009 and 2017 NHTS

Figure 3.5 shows the distribution of daily household vehicle miles traveled (VMT) in both survey years. The distributions are heavily right-skewed, with most households traveling relatively short distances (under 50 miles per day) and a long tail of high-mileage households. The overall pattern is similar between 2009 and 2017, though there are some subtle differences in the distributions.

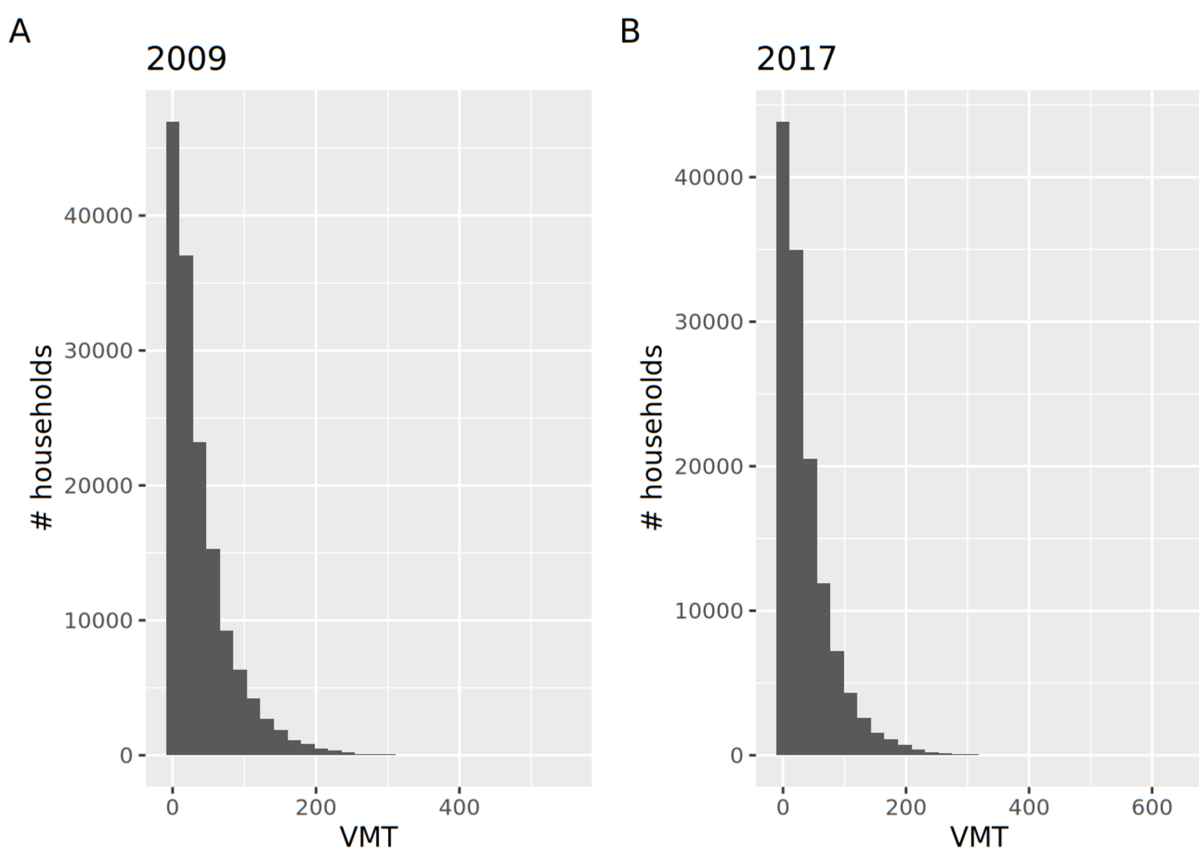

Figure 3.5: Distribution of average daily household vehicle miles traveled (2009 and 2017)

3.2.2 Household VMT by Day of Travel in 2009 and 2017 NHTS

The day-of-week patterns in household VMT (Figure 3.6) reveal consistent weekly travel rhythms. Both survey years show similar patterns with: - Higher travel on weekdays, particularly mid-week - Lower travel on weekends, especially Sundays - Relatively stable confidence intervals suggesting consistent patterns across households The persistence of these patterns between 2009 and 2017 suggests stable weekly travel routines despite broader societal changes.

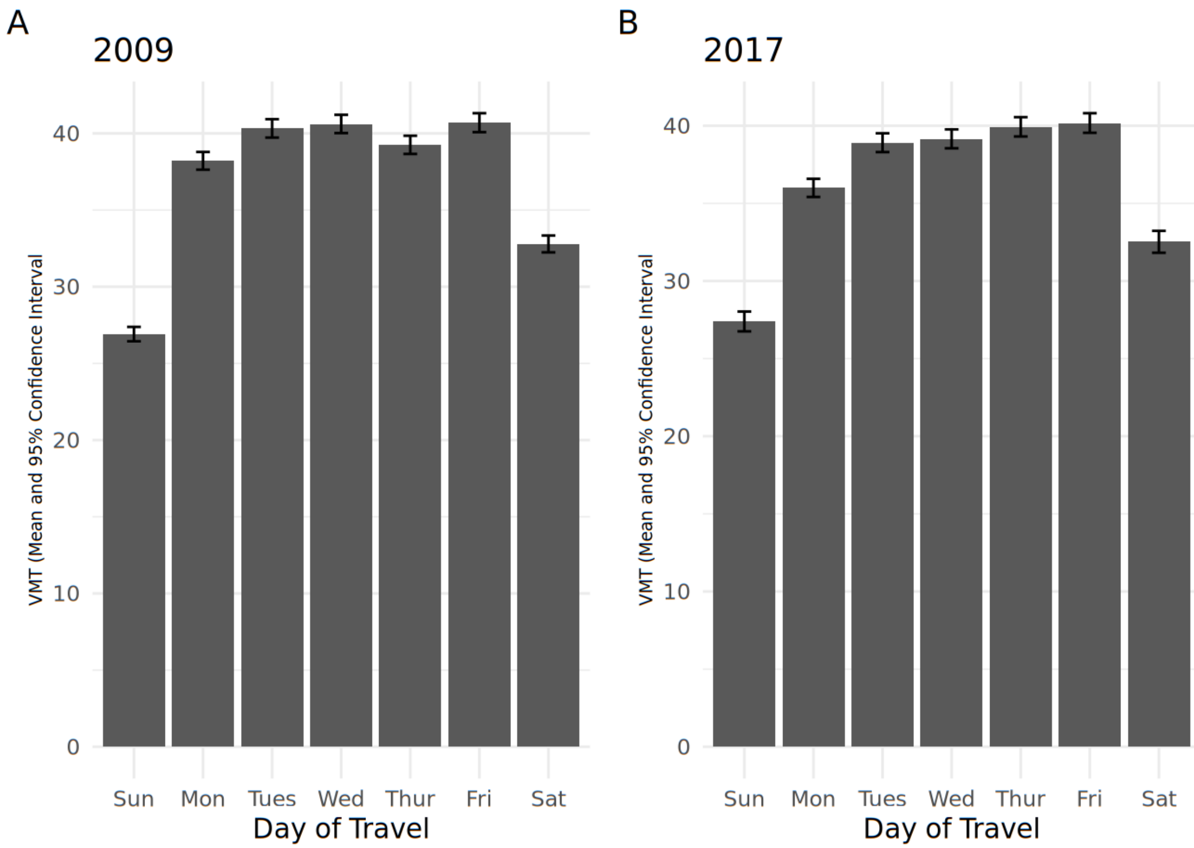

Figure 3.6: Household vehicle miles traveled by day of week (2009 and 2017)

3.2.3 Correlation among SLD Variables

The correlation matrices for Smart Location Database variables (Figure 3.7 and Figure 3.8) reveal strong correlation between built environment measures - especially the measures in the same D categories. Key patterns include: - High correlation between density measures (population, employment, housing) - Strong relationships between transit accessibility and development intensity - Consistent correlation structures between 2009 and 2017.

3.2.3.1 SLD v2 (2009)

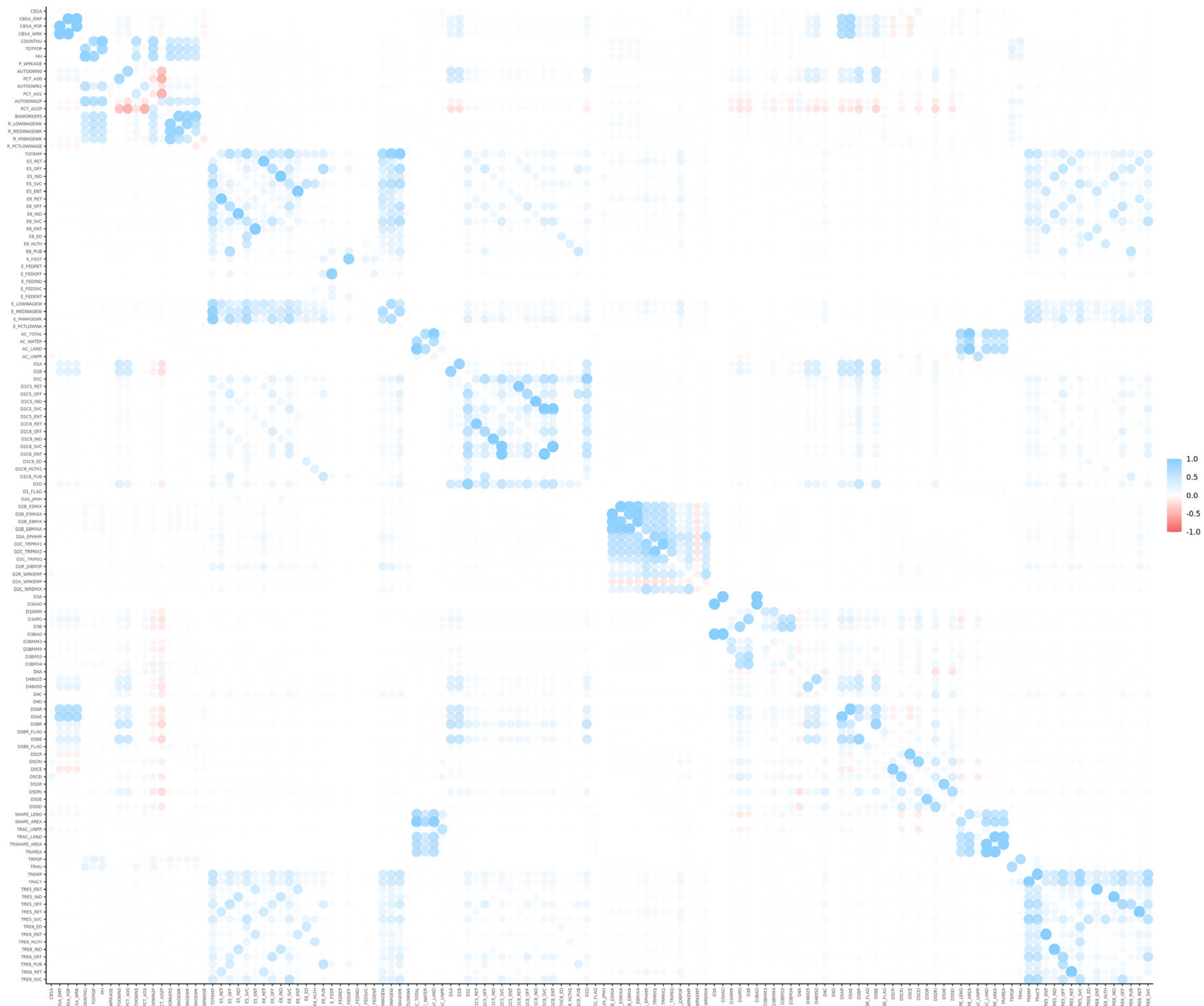

Figure 3.7: Pairwise correlation matrix for Smart Location Database variables (2009)

3.2.3.2 SLD v3 (2017)

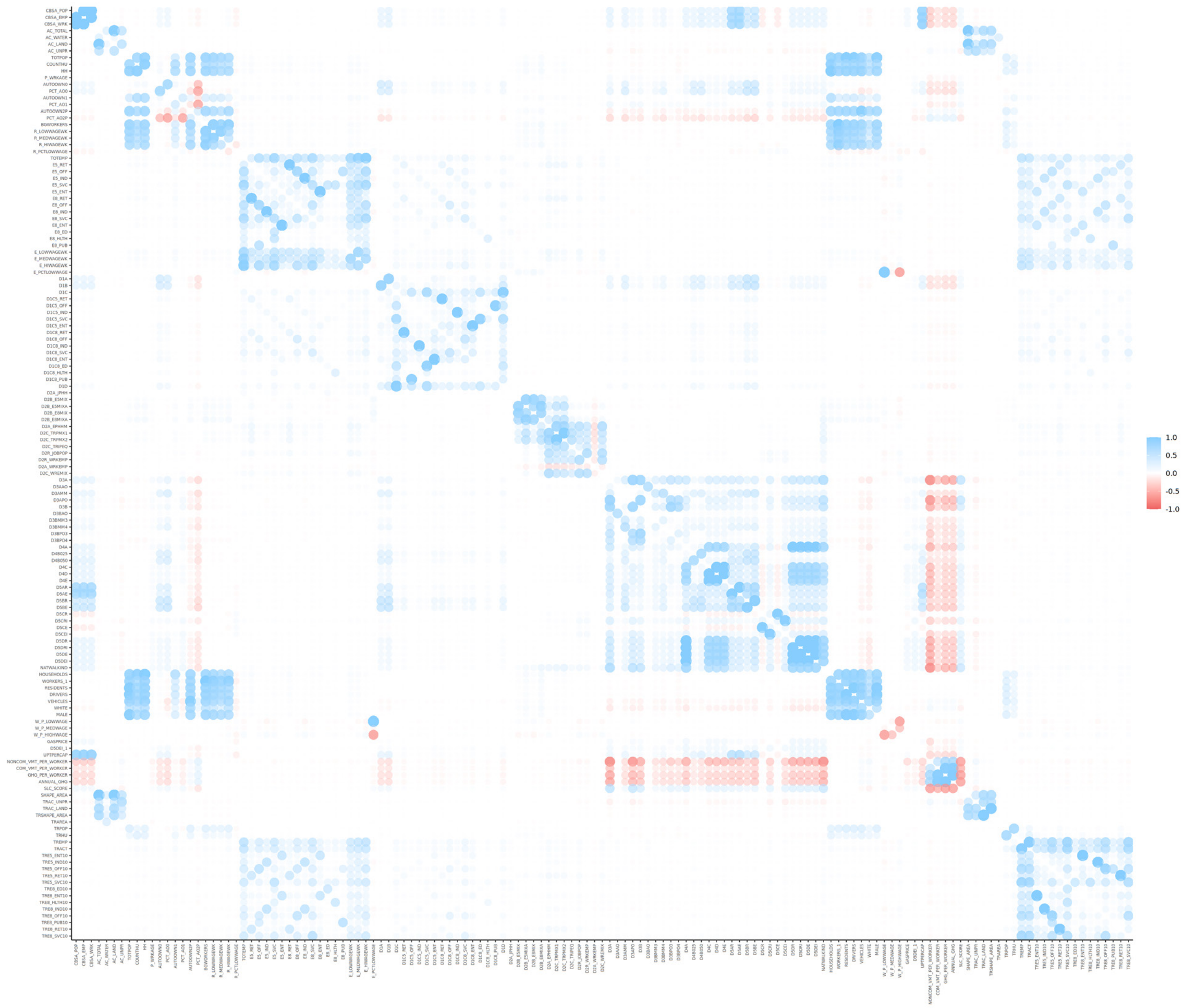

Figure 3.8: Pairwise correlation matrix for Smart Location Database variables (2017)

3.2.4 Household VMT by VEState Place Type

Figure 3.9 shows how household VMT varies across different place types, as defined by the VEState definition. The patterns reveal: - Clear gradients in travel behavior across the urban-rural spectrum - Generally higher VMT in less urbanized areas - Lower VMT in dense, mixed-use environments - Consistent patterns between 2009 and 2017, though with some variation in specific place types

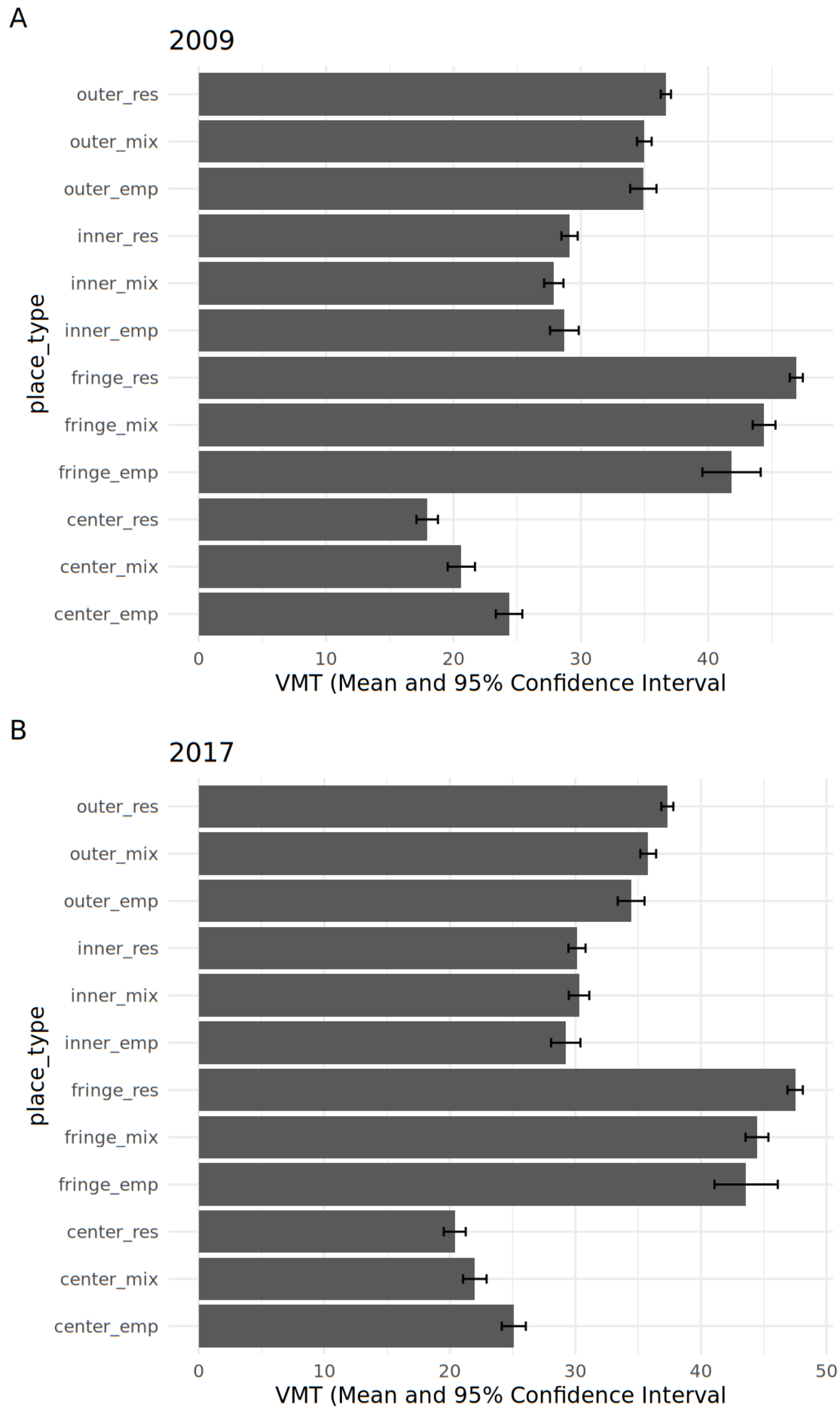

Figure 3.9: Household vehicle miles traveled by VEState place type (2009 and 2017)

4.0 MODEL ESTIMATION AND SELECTION

This chapter explores the land use type definition and the empirical model components that support the main VELandUse functions described in Chapter Chapter 6. In particular, it covers the classification and transition logic used by PredictLUType, PredictLocType, and AssignD3D4Levels, along with the allocation models used by AllocateDU and AllocateEmployment. LoadLUType is the companion function used when land use types are supplied directly by the user, and UpdateLUType is the bookkeeping step that updates land use type fields between simulation years rather than a separately estimated statistical model.

4.1 LAND USE TYPE DEFINITION AND MODELING

This section explores Land Use Type definitions using the 2010 (V2) and 2018 (V3) Smart Location Database. Note that after internal discussion, we decided to use the term land use type instead of place type to avoid confusion of similar concepts in different applications.

4.1.1 Land Use Type Classification Methods

As discussed in the literature review section (Chapter 2), there have been a few land use (place) type classification methods (definitions), including the place type definition in SmartGAP/RPAT, VEState, and the method used by Oregon DLCD/ODOT in the Climate Friendly Area work.

4.1.1.1 VEState Place Type

The Place Type definition used in VEState (https://github.com/VisionEval/VisionEval/blob/master/sources/modules/VESimLandUse/inst/module_docs/CreateSimBzoneModels.md):

There are three dimensions to the place type system. - Area types identify the relative urban nature of the SimBzone: center, inner, outer, fringe. - Development types identify the character of development in the SimBzone: residential (res), employment (emp), mix. - Location type identifies whether a Bzone is located in an urbanized area (Urban), a smaller urban-type area (Town), or a non-urban area (Rural).

Area types are designated based on a combination of activity density and destination accessibility levels. Each is split into 4 levels. Area type is determined by 16 combinations of those levels. Following are the activity density level definitions:

- Very Low (VL): 0 to 0.5 households and jobs per acre
- Low (L): Greater than 0.5 to 5 households and jobs per acre
- Moderate (M): Greater than 5 to 10 households and jobs per acre
- High (H): Greater than 10 households and jobs per acre

Following are the destination accessibility level definitions:

- Very Low (VL): 0 to 2,000 units
- Low (L): Greater than 2,000 units to 10,000 units
- Moderate (M): Greater than 10,000 units to 50,000 units
- High (H): Greater than 50,000 units

The following table classifies area types by activity density levels and destination accessibility levels. Rows in the table represent activity density levels and columns represent destination accessibility levels.

Table 4.1: Area type classification by activity density and destination accessibility levels in VEState

	Very Low	**Low**	**Moderate**	**High**
Very Low	fringe	fringe	outer	outer
Low	fringe	outer	outer	inner
Moderate	outer	outer	inner	inner
High	outer	inner	center	center

As with activity density, these profiles are simplified by discretizing the D2A_JPHH variable into the following 5 activity mix levels:

- primarily-hh: greater than 4 households per job
- largely-hh: less than 4 households to 2 households per job
- mixed: less than 2 households per job to 2 jobs per household
- largely-job: greater than 2 jobs per household to 4 jobs per household
- primarily-job: greater than 4 jobs per household

Development type is determined by collapsing the mix levels from 5 to 3 as follow:

Table 4.2: Development type classification by activity mix levels in VEState

Development Type	**Mix Levels**
mix	mixed
res	primarily-hh & largely-hh
emp	primarily-job & largely-job

4.1.1.2 DLCD/ODOT Place Type

Oregon DLCD/ODOT uses a place type definition (https://www.oregon.gov/ODOT/Planning/Documents/Oregon-Place-Types-Classification.pdf) suitable for communities in Oregon, especially in their Climate Friendly Area work. Compared to VEState definition, DLCD/ODOT definition has a different classification of area types: Regional Center, Close In Community, Suburban/Town, and Low Density/Rural.

And there are also more development types: in addition to Residential, Employment, and Mixed, it includes Mixed High and TOD.

One limitation of both place type definitions is that they focus on identifying different types of communities (land use), but less on the difference on travel outcomes of residents.

In VELandUse project, we refine the land use type definition based on the VEState topology, and utilize machine learning method to select variables and determine cutoffs such that the land use type is most useful in predicting travel outcomes (Vehicle Miles Travelled).

4.1.2 Proposed Land Use Type Definition

We adopted the VEState terms for land use types, but used a more data-driven approach for definition. The land use types in the new VELandUse module comprises of:

- Area Type: Regional Center, Center, Inner, Outer, and Fringe
- Diversity Type: Residential, Employment, and Mixed
- Location Type: Urban, Town, and Rural.

The following sections discuss the method to classify area type, diversity type, and location type and to model their transition in turn.

Land use type transition refers to how the land use type of a BZone changes over time in VELandUse simulation. Later in this section, we will discuss how to model land use type transition. We examine how land use types changed between 2010 (SLD v2) and 2018 (SLD v3) to understand patterns and factors associated with land use type transition. The estimated models will be used to predict future land use types for Bzones in use case #3 (Section 3.5.4).

4.2 AREA TYPE AND DIVERSITY TYPE CLASSIFICATION

We chose the decision tree method to both select variables and to determine cutoffs for area type and diversity type definition. In the current approach, we use a customized decision tree to classify block groups into area type with the best prediction accuracy for daily household VMT (NHTS). We experimented with D variables in SLD with various buffer configurations (no buffer; 1/4 mile centroid buffer, 1/2 mile centroid buffer, 1 mile centroid buffer, 2 mile centroid buffer; 1/4 mile polygon buffer, 1/2 mile polygon buffer, 1 mile polygon buffer, 2 mile polygon buffer), allowing at most 1 variable from each of the D categories (density, diversity, and destination accessibility). In the same process, the decision trees also choose cutoffs for area types.

In earlier iterations, we experimented with a two-step process:

1. We fit a decision tree of daily household VMT (NHTS) on the built environment variables (D variables in Smart Location Database - SLD) and pick D variables on the top of variable importance ranking (at most one from each D category). The variables selected include:

    - *D1D_hmbuf* - activity density (employment + households) within half-mile buffer;

    - *D5 Destination Accessibility* - harmonic means of employment within 2 miles and population within 5 miles, as a proxy for accessibility measure within the need for network information;

    - *D2R_WRKEMP_1mbuf* - Workers-to-employment ratio within 1 mile buffer (centroid buffer);

    - *D2A_JPHH_2mbuf* - Jobs-to-households ratio within 2 mile buffer (centroid buffer).

2. For each selected variable, we run the decision tree method again to select cutoffs for each variable.

We also tested area type definition with cutoffs segmented by UZA size group (large, small and non-UZA). Neither of the methods provided significantly better prediction accuracy, but introduced more complexity. Therefore, we favor the current approach because it simplifies the process in selecting variables and determining cutoffs simultaneously and provides similar prediction accuracy.

4.2.1 Area Type Definition

Our method first defines regional centers for large UZAs (Population >= 1,000,000) using a decision tree of daily household VMT (NHTS) on the built environment variables (D variables). The regional center decision tree selects both the variables and cutoffs to minimize the residual sum of squares of DVMT. Figure 4.1 shows the regional center decision tree, and Table 4.3 summarizes the resulting rule in a compact table.

The decision-tree figures in this section should be read from top to bottom. Each internal node shows the variable used for the next split and the threshold that divides the sample into two groups. Following one branch means the block group satisfies that condition; following the other means it does not. The upper splits therefore identify the variables with the largest role in separating different travel-behavior contexts, while the terminal leaves represent the final groups that are later translated into the reported area type or diversity type categories. In practical terms, a split near the top of the tree indicates a stronger organizing variable for the classification than a split that appears only near the bottom.

For consistency with the later maps and tables, the leaves of the area-type trees should be interpreted as the empirical source of the final regional center, center, inner, outer, and fringe labels, while the leaves of the diversity tree are later translated into the reported res, mix, and emp categories. The trees therefore show the decision rules, and the later maps and tables show those same leaves after they have been converted into the named categories used throughout the report.

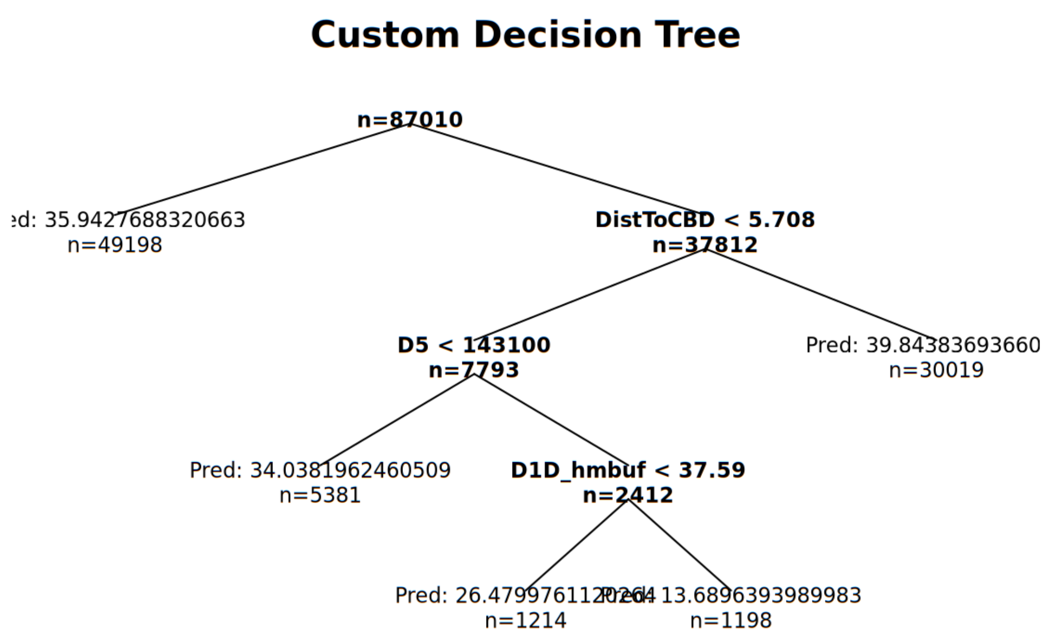

Figure 4.1: Decision tree for identifying regional centers based on UZA size, distance to CBD, destination accessibility, and activity density

Table 4.3: Regional center classification rule derived from the fitted decision tree

Rule	Result
UZASize != "small" and DistToCBD < 5.71 and D5 >= 143140 and D1D_hmbuf >= 37.59	**regional center**

Rule	Result
All other combinations	**not regional center**

Only one terminal branch maps to the regional center label, so the table highlights the positive rule directly while the tree preserves the full fitted structure.

For non-regional center non-rural block groups (Urban or Town block groups), we then use another decision tree of daily household VMT (NHTS) on the built environment variables (D variables) to classify block groups into four area types: center, inner, outer, fringe. Similar to the regional center decision tree, the area type decision tree selects both the variables and cutoffs to minimize the residual sum of squares of DVMT. Figure 4.2 shows the area type decision tree, Table 4.4 summarizes the resulting thresholds, and Figure 4.3 shows a map of block group area types in Oregon.

For interpretation, the important point is not only which variable appears in the tree, but also where it appears. When D5 or D1D_hmbuf appears near the top of the tree, it means accessibility or activity density is doing most of the work in separating lower-VMT from higher-VMT contexts. The leaves should therefore be read as empirically defined combinations of accessibility and intensity that are later labeled as center, inner, outer, or fringe, rather than as arbitrary bins chosen in advance.

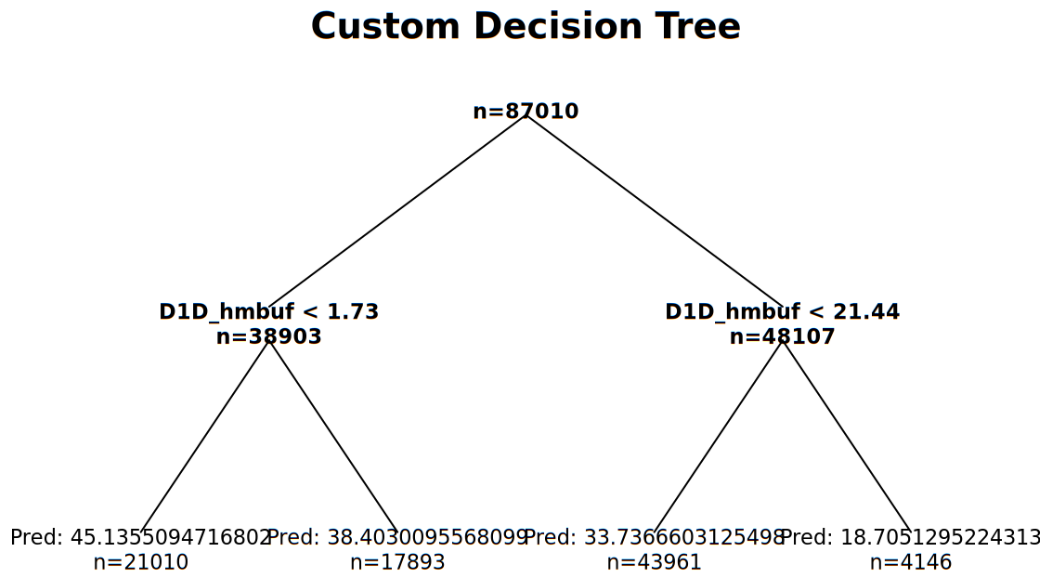

Figure 4.2: Decision tree for classifying block groups into area types (center, inner, outer, fringe)

Table 4.4: Area type classification rules derived from the fitted decision tree

D5	D1D_hmbuf	Area Type
>= 16404	>= 21.44	**center**
>= 16404	< 21.44	**inner**
< 16404	>= 1.73	**outer**
< 16404	< 1.73	**fringe**

As with the Diversity Type section, the chart and table are shown together on purpose: the chart documents the estimated tree structure, while the table makes the final threshold logic easier to scan and describe accessibly.

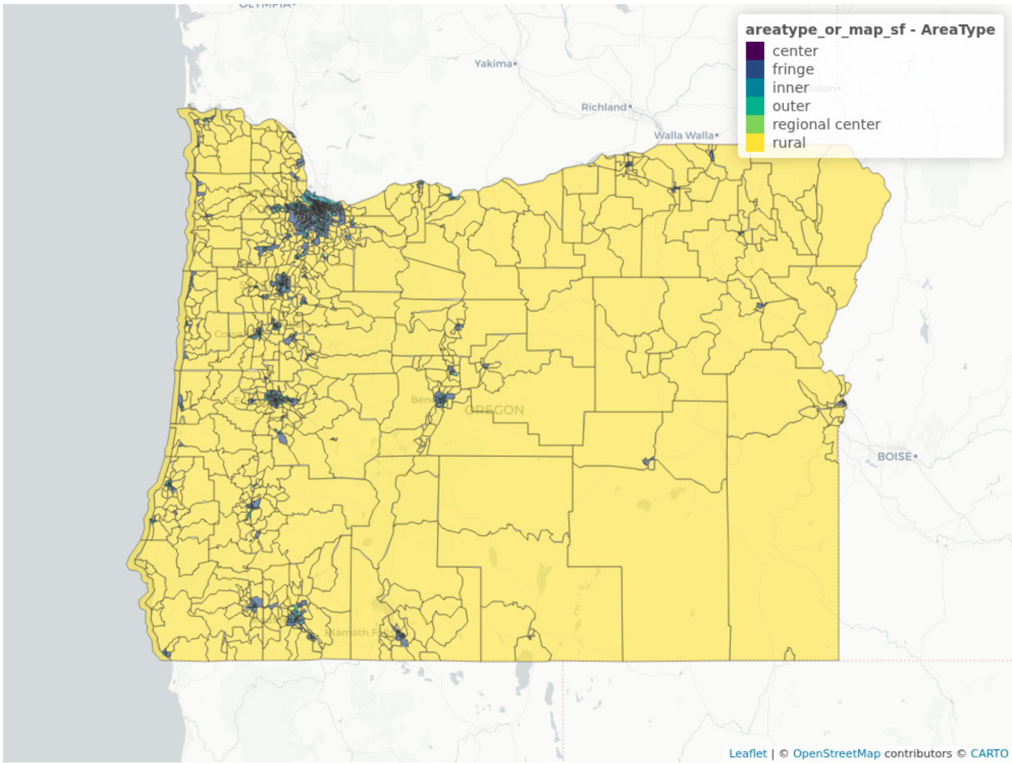

Figure 4.3 Interactive map of Oregon Census block groups colored by proposed area type classification

4.2.2 Diversity Type

Similar to the Development Type in VEState, Diversity Type in the Land Use Type definition is based on the mix of employment and households. We used the same decision-tree approach as for Area Type to identify the most informative variable and cutoff values. In the current specification, the fitted tree reduces to a single variable, D2A_JPHH_2mbuf (jobs per household within a 2-mile buffer of a Census Block Group). Figure 4.4 shows the fitted tree, and Table 4.5 summarizes the same result in a more compact ADA-friendly form.

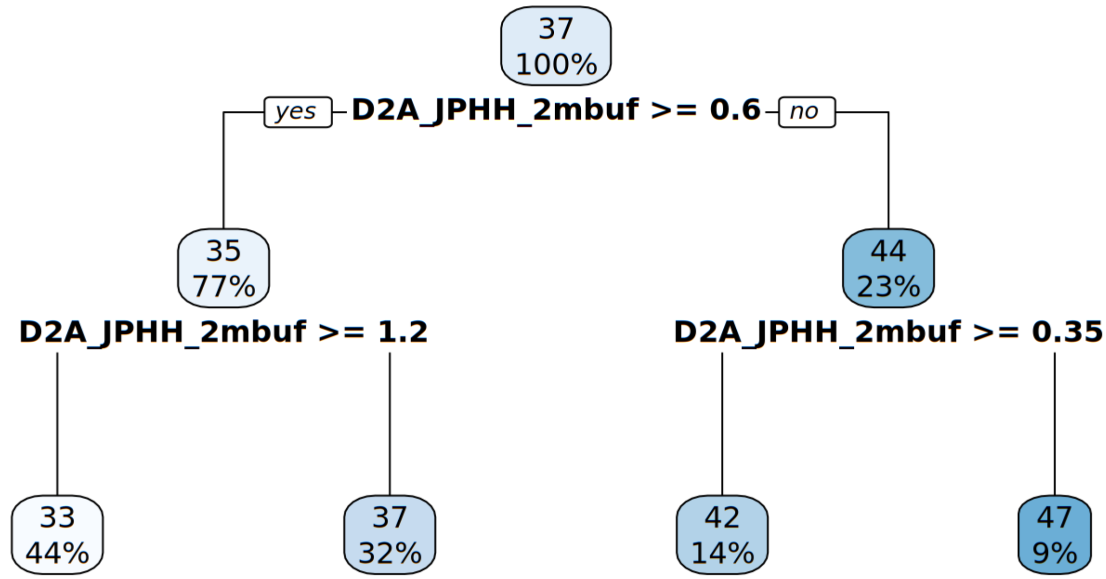

Figure 4.4: Decision tree for classifying block groups into diversity types (employment, mixed-use, residential)

Table 4.5: Diversity Type classification rule based on D2A_JPHH_2mbuf

D2A_JPHH_2mbuf range	Diversity Type
< 0.5	res
0.5 to 2.0	mix
> 2.0	emp

The chart and table are intentionally shown together here: the chart documents the fitted tree structure, while the table makes the final threshold rule easier to scan and describe accessibly.

4.2.3 Land Use Type (Combining Area Type and Diversity Type)

We combine the area type and diversity type to create the Land Use Type classification. Figure 4.5 shows a map of block group Land Use Types in Oregon.

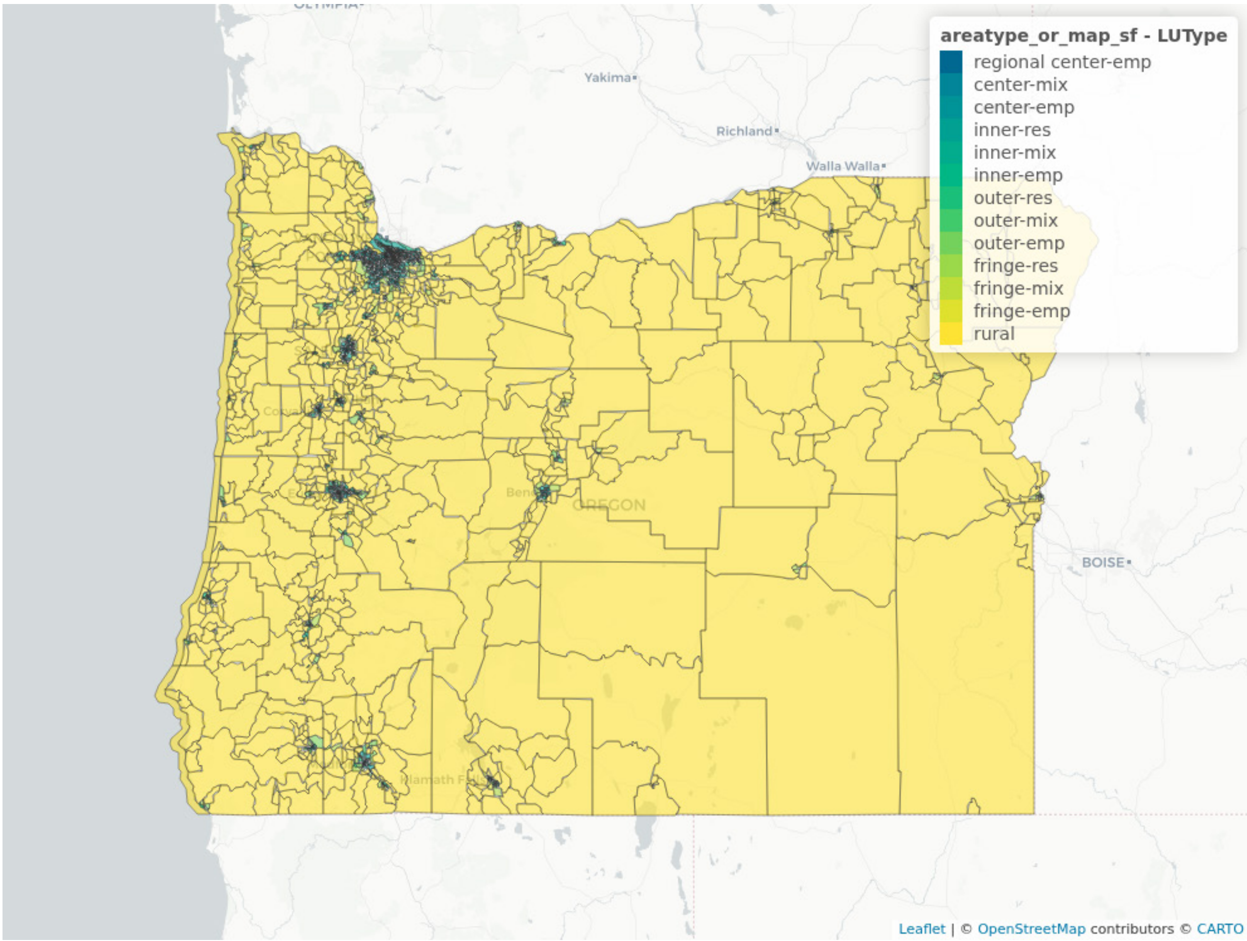

Figure 4.5 Interactive map of Oregon Census block groups colored by proposed land use type classification

4.2.4 Compare Land Use Type Classification Methods

In this section, we compare the proposed Land Use Type with VEState Place Type and relate both to the Oregon Place Type frameworks used in ODOT tools such as RSPM and RPAT. A crude comparison of the two quantitative classification methods is to compare the household VMT prediction accuracy based on RMSE and (R^2). Figure 4.6 shows average household VMT by Place Type and Land Use Type, respectively. Table 4.7 shows the household VMT prediction accuracy of the two methods. The proposed definitions have similar prediction accuracy as the VEState definition.

Table 4.6: Comparison of VEState, Oregon Place Type, and proposed land use type frameworks

Framework	Main structure	How categories are defined	Relation to the proposed method
VEState place type	Two dimensions: area type and development type	Fixed threshold bins on density, accessibility, and jobs-housing balance variables	Closest direct predecessor. The proposed method keeps the same two-dimensional logic, but estimates the category rules empirically rather than fixing them a priori.
Oregon Place Types (RSPM / RPAT)	Policy-facing area and development/place type categories	Hand-crafted planning categories intended for scenario inputs and communication	Similar in planning purpose and broad ordering, especially in the area-type dimension. The proposed method is narrower and more data-driven, with simpler res / mix / emp diversity classes instead of the fuller Oregon development-type menu.
Proposed Land Use Type	Two dimensions: AreaType and DivType	Data-driven thresholds estimated from observed built-environment and travel relationships	Designed to remain recognizable to VEState and Oregon users while grounding the classification in observed block-group data and travel outcomes.

The main alignment is therefore structural rather than one-to-one. The proposed AreaType categories preserve the same broad progression from more central and accessible places to more peripheral ones, which makes them interpretable to users familiar with VEState and Oregon Place Types. The main departure is in how the categories are defined: instead of fixed bins or a richer planning taxonomy such as TOD or mixed-use-high, the new method uses empirical thresholds and a simpler diversity dimension so it can be estimated consistently from observed data and then carried forward in simulation.

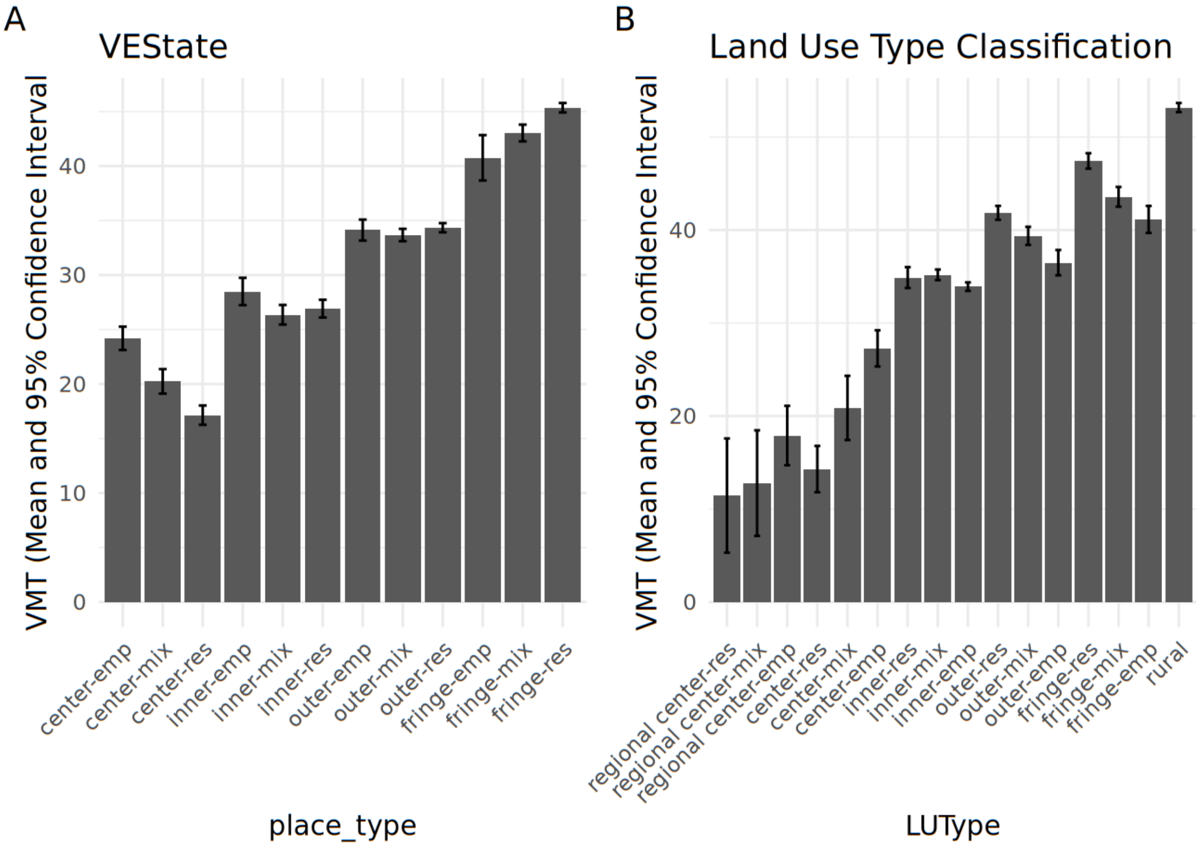

Figure 4.6: Comparison of mean household VMT by land use type classification: (A) VEState versus (B) proposed definition

Table 4.7: R^2 and RMSE comparison of household VMT prediction by VEState and proposed land use type definitions

Method	\(R^2\)	RMSE
VEState	0.0319	42.47
Proposed Land Use Type	0.0367	42.73

The VMT comparison suggests that the proposed method performs at least as well as the inherited VEState place type for this application while remaining closer to the Oregon planning frameworks in naming and interpretation than a purely statistical clustering method would be. In that sense, the new classification is intended as a bridge: it preserves the policy-facing logic of Oregon and VEState place types, but replaces ad hoc threshold choices with rules estimated from observed built environment and travel data.

4.2.4.1 Mapping Place Type and Land Use Type

In this section, we map the different definitions for three regions in Oregon - Portland (Figure 4.7), Salem (Figure 4.8), and Rogue Valley (Figure 4.9) - for visual comparison (A live version of the maps is available at http://sapporo.usp.pdx.edu/files/a0fb89).

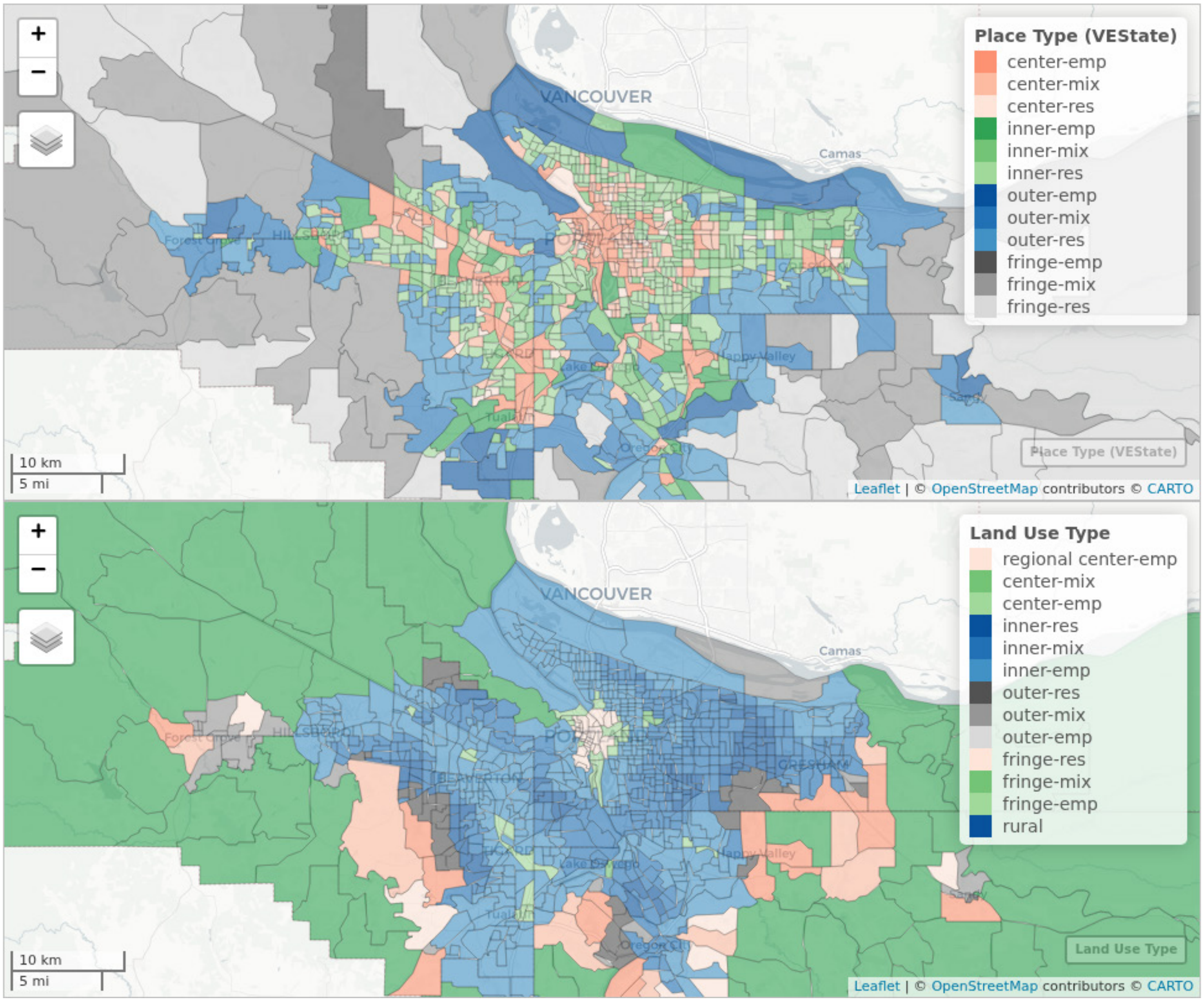

Figure 4.7: Choropleth map of Portland metro area Census block groups colored by predicted land use type classification

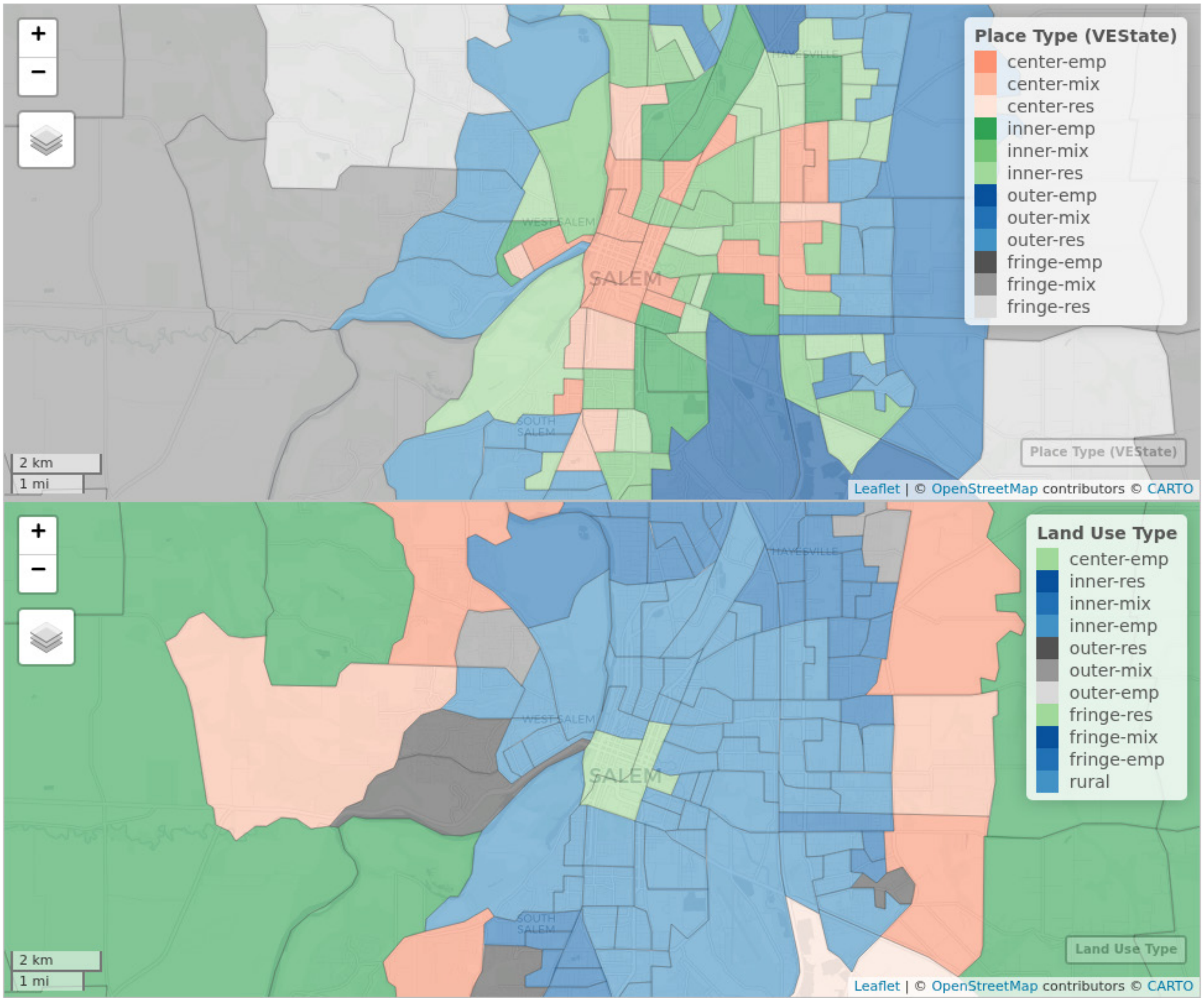

Figure 4.8: Choropleth map of Salem area Census block groups colored by predicted land use type classification

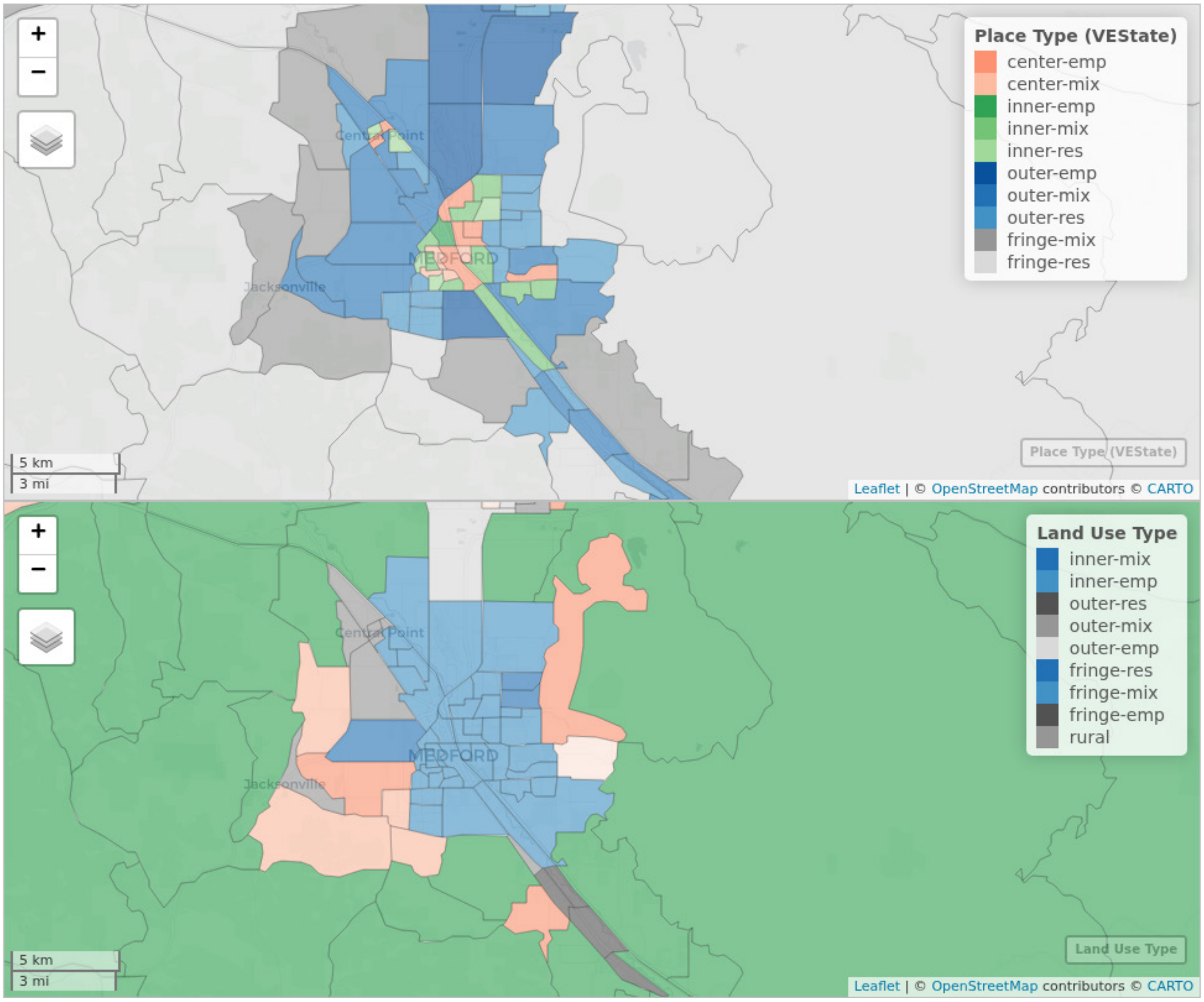

Figure 4.9: Choropleth map of Rogue Valley area Census block groups colored by predicted land use type classification

These side-by-side maps also illustrate an important limitation of the current classification in small urban and rural contexts. The regional center rule is only estimated for large urbanized areas, and the four-way center / inner / outer / fringe area-type tree is estimated only on non-rural Urban and Town block groups. As a result, the proposed classification is most strongly supported for larger metropolitan settings with enough variation in accessibility and density to identify clear thresholds. In smaller metros and at rural edges, the same buffered measures can smooth local variation and produce patterns that look more diluted or less intuitive than the inherited VEState place type maps. The later pilot results in this report show why that matters in practice: some small-area and rural-edge classifications and transitions should be interpreted cautiously as provisional model behavior rather than as fully validated planning categories.

4.3 PREDICTLUTYPE: LAND USE TYPE TRANSITION MODEL

PredictLUType is the VELandUse function that predicts how AreaType and DivType change over time in Use Case #3. In this section, we examine how land use types changed between 2010

(SLD v2) and 2018 (SLD v3) to understand the patterns and factors that influence the transition. We then estimate models using observed SLD v2 and v3 data to predict future land use types for Bzones.

4.3.1 Explore land use type transition

Figure 4.10 shows the transition of Land Use Types between the SLD v2 (2010) and SLD v3 (2018) with complete SLD data. The Land Use Types for v2 and v3 are classified using the method described in Section 4.1.

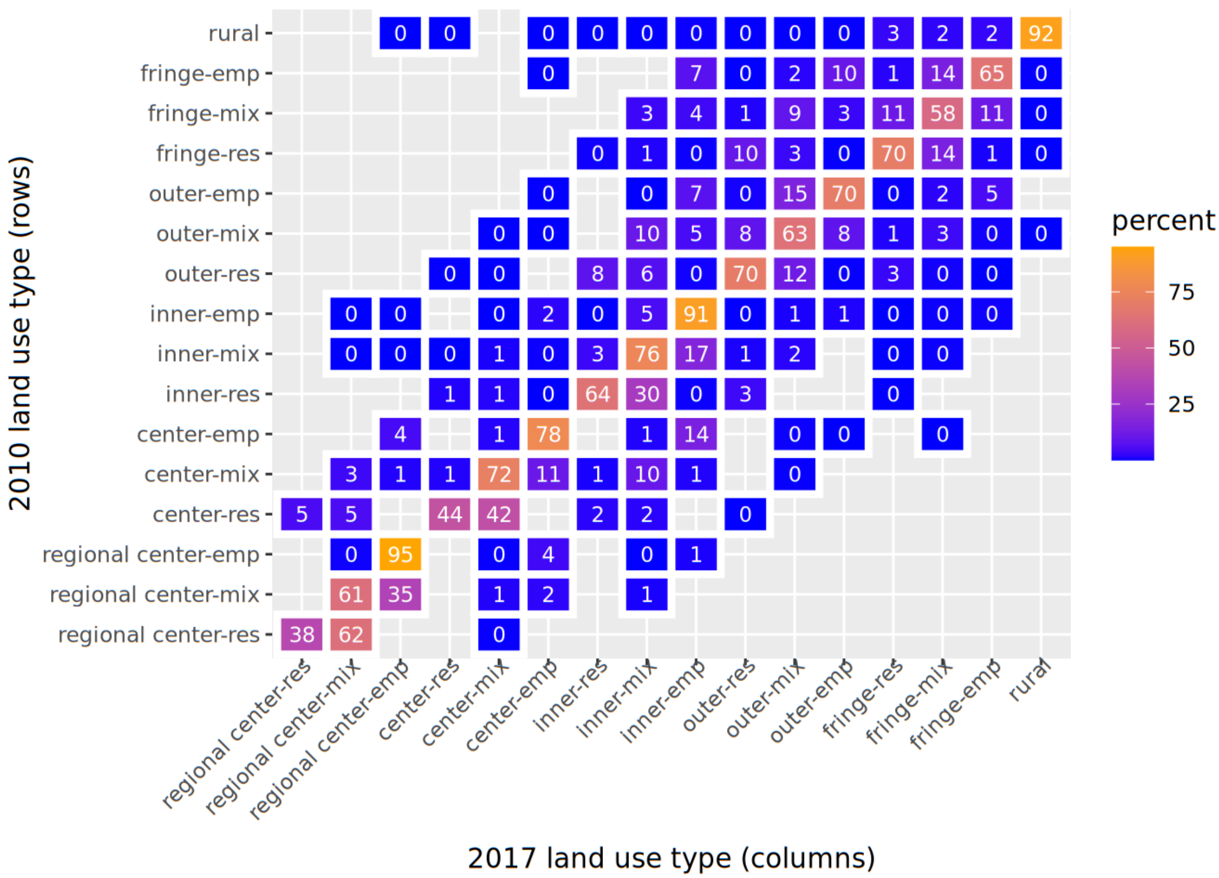

Figure 4.10: Land Use Type transitions between 2010 (SLD v2) and 2017 (SLD v3)

Throughout the transition displays in this chapter, the earlier or observed category is shown in the rows and the later or predicted category is shown in the columns. That keeps the reading direction consistent across transition heat maps and confusion matrices.

4.3.2 Land Use Type Transition Models

For use case #3 (Section 3.5.4), the VELandUse model predicts future Land Use Types from base year or previous simulation year inputs.

Here we evaluate different model structures that balance prediction accuracy and interpretability - the selected model should not be overly complex or hard to interpret. For this reason, we focus on the decision tree-based method and discrete choice models (Multinomial Logit Models).

4.3.2.1 Decision Tree Method

We use decision trees to model the transition of Land Use Types between SLD v2 (2010) and SLD v3 (2018). Decision trees are chosen for their interpretability and ability to handle non-linear relationships. The model predicts the Land Use Type in 2018 based on these potential variables:

1. Land Use Type characteristics in 2010:
    - Area type (density and accessibility levels)
    - Diversity type (jobs-housing balance)
2. Socioeconomic characteristics:
    - Number of households (HH)
    - Total population (TOTPOP10)
    - Total employment, and employment by sector group (service and retail) (EMPTOT)
3. Built environment measures:
    - Activity density (employment + housing units) (D1D_hmbuf)
    - Destination accessibility (D5)
    - Percent steep slope (pct_steep_slope)
    - Distance to highway ramp (dist_to_ramp)
    - Distance to CBD (dist_to_cbd)
    - Distance to freeway station (dist_to_fgw_sta)

The model is trained on 75% of the data and tested on the remaining 25%. The variable importance plot (Figure 4.12) and decision tree visualization (Figure 4.11) below show the key factors driving land use type transitions. The accuracy metrics (Table 4.10) indicate how well the model predicts land use types in the test dataset. The tree diagram and variable importance plot are generated using the rpart and vip packages, along with prediction accuracy metrics.

The transition tree should be read in the same top-to-bottom way as the classification trees elsewhere in this chapter. A split near the root indicates a variable that does the most work distinguishing likely future land use outcomes; lower splits provide more local refinements after the major separation has already occurred. Each terminal leaf represents a subset of Bzones with similar observed transition behavior, so the tree is useful not only for prediction but also for showing which combinations of current land use, accessibility, slope, and distance variables are most associated with different future land use types.

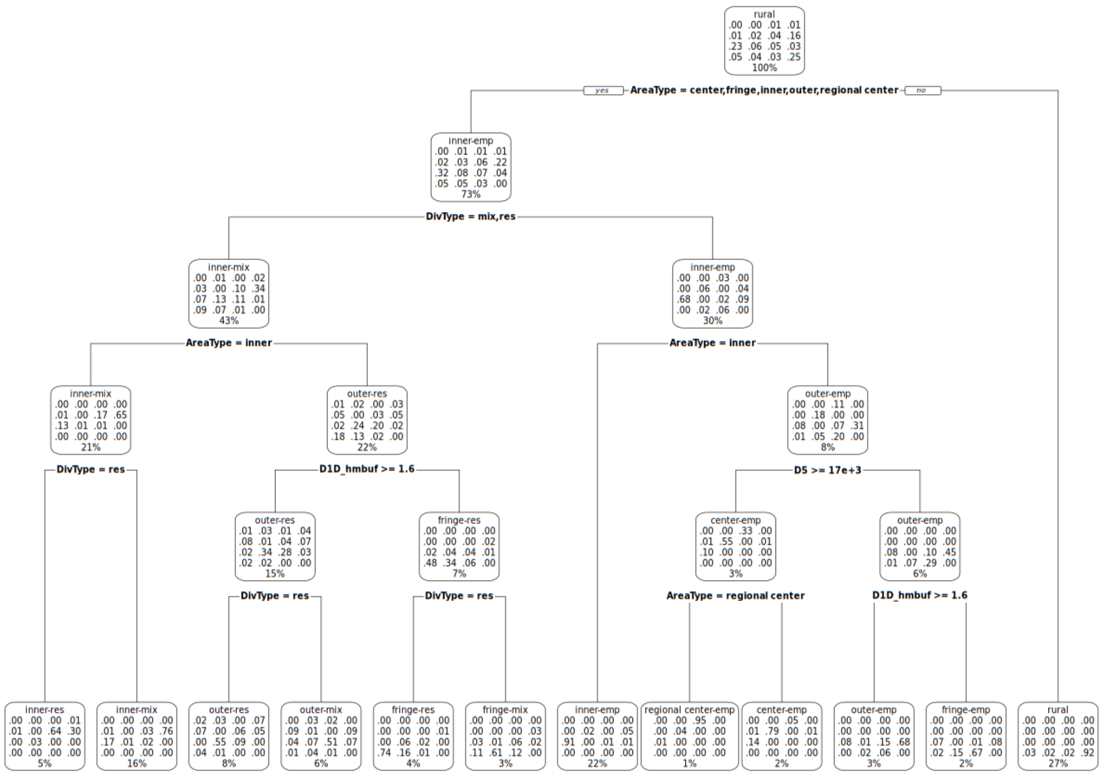

Figure 4.11: Decision tree for predicting land use type transitions from 2010 to 2017

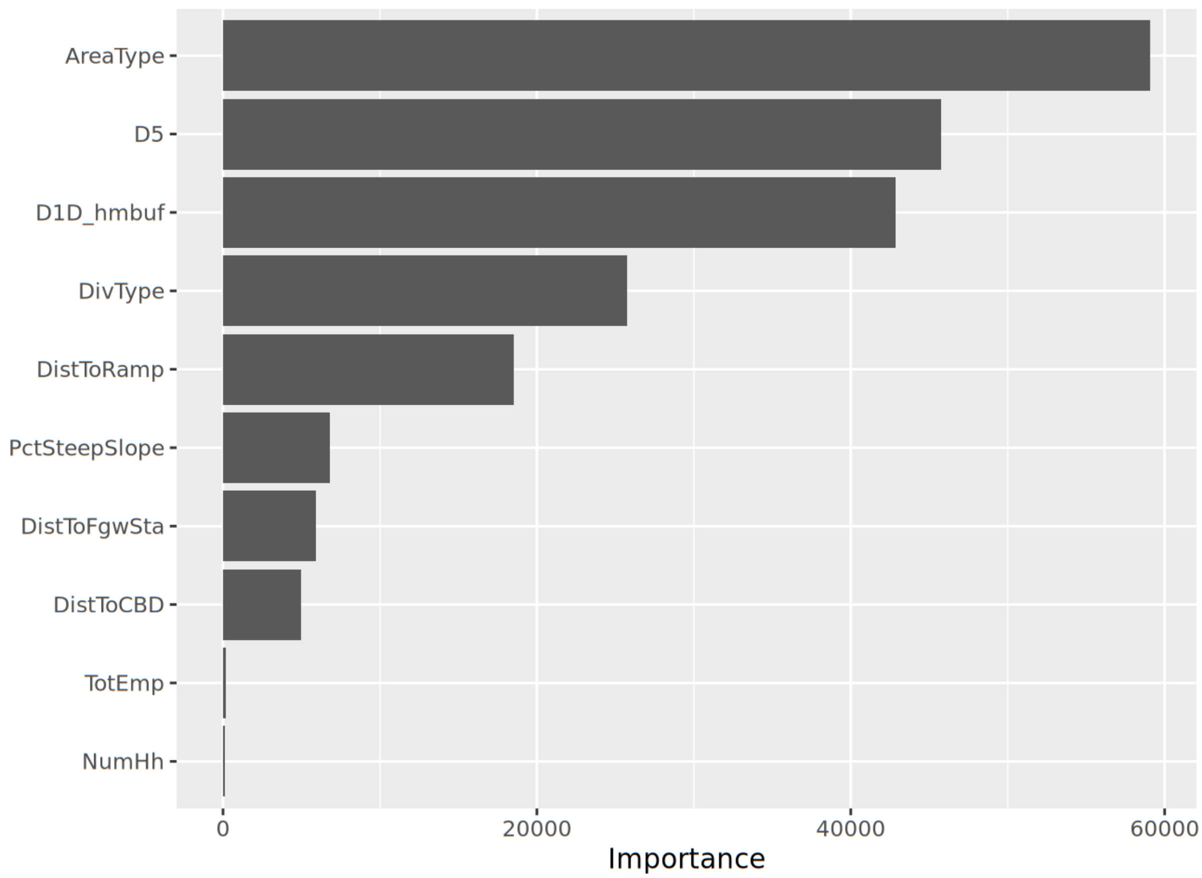

Figure 4.12: Variable importance for the land use type transition decision tree model

4.3.2.2 LASSO-Regularized Multinomial Logit Models of Area Type and Diversity Type

As an alternative to the decision tree method, we use LASSO-regularized multinomial logit models to analyze and predict land use type transitions between 2010 (SLD v2) and 2017 (SLD v3). The models estimate the probability of a Bzone transitioning from one land use type to another, with LASSO regularization (L1 penalty) and automatic variable selection. The cv.glmnet() function from the glmnetUtils package is used with 10-fold cross-validation to select the optimal regularization parameter (lambda). Similar to the decision tree method, the model is based on the same set of potential variables as the decision tree method. The models are trained on 80% of the data and validated on the remaining 20%.

For multinomial logit models, we estimate separate models for area type transitions and diversity type transitions instead of a single model for land use type transitions, as the two types of land use are conceptually distinct and may be influenced by different factors. Two separate models also address the issue of violating the assumption of independence of irrelevant alternatives (IIA) with a single multinomial logit model directly predicting the probability of Land Use Type transitions.

In the training set, we filter to include only feasible transitions: area types can only stay the same or increase in density (e.g., a fringe area can transition to outer, but an inner area cannot transition to fringe) as the reverse transitions are rare in the real world and in the observed data.

The current transition mechanics are therefore asymmetric. For AreaType, the estimation sample removes reverse transitions and several large jumps, so the fitted model is intentionally monotonic: denser types can be maintained or intensified, but not relaxed. By contrast, the current DivType estimation does not impose an analogous monotonic filter, so reverse changes among res, mix, and emp remain possible in principle if the estimated probabilities support them.

The implementation described in Chapter Chapter 6 also clarifies how these estimated transitions are applied in simulation. PredictLUType uses the transition models to determine the most probable future AreaType and DivType for each Bzone from its current or base-year attributes, then stores those class assignments for the target year. In other words, the current land use type transition process is probability-based in estimation but deterministic in assignment: unlike PredictLocType, the implementation chapter does not describe an optional stochastic draw from the predicted class probabilities.

This point matters for repeated future simulation years. The transition models here are estimated from observed change between the historical SLD vintages, then reapplied to whatever current-year attributes are available at the next simulation step. The report does not currently estimate separate short-step versus long-step transition models or any explicit time-scaling parameter. In practice, that means multi-period forecasts approximate long-run change by iterating the same one-step transition relationship on updated Bzone conditions rather than by scaling a single transition probability over a longer horizon.

For the Area Type Transition Model, model coefficients are reported below. Prediction accuracy metrics are summarized in Table 4.10.

In early iterations, we also used the mlogit R package to estimate simple multinomial logit models without regularization. mlogit was prone to convergence issues with large datasets, so we switched to use glmnetUtils instead.

For the Diversity Type Transition Model, model coefficients (Table 4.9) are reported below. Prediction accuracy metrics are summarized in Table 4.10.

Table 4.9: Diversity Type Transition LASSO Multinomial Model Coefficients

	res	mix	emp
(Intercept)	-0.5206	1.2784	-0.7578
NumHh	0.0000	0.0000	0.0002
TotEmp	-0.0005	0.0000	0.0002
AreaTypecenter	0.0000	0.3578	-0.2001
AreaTypefringe	0.0000	-0.0780	0.0000
AreaTypeinner	-0.5139	0.0000	0.1335
AreaTypeouter	0.0472	0.0000	-0.2300
AreaTyperegional center	-0.2273	0.2881	0.0000
DivTyperes	3.5491	0.0000	-2.1131
DivTypeemp	-0.3208	0.0000	3.9261
D1D_hmbuf	0.0000	0.0000	0.0000
D5	0.0000	0.0000	0.0000
PctSteepSlope	0.0026	-0.0020	0.0000
DistToRamp	0.0000	0.0006	-0.0085
DistToCBD	-0.0135	0.0000	0.0096
DistToFgwSta	-0.0012	0.0000	0.0007

Table 4.10: Prediction accuracy of land use type transition models

Model	.metric	.estimate
Decision Tree (LUType)	accuracy	0.7876267
Decision Tree (LUType)	kap	0.7473872
MNL (AreaType)	accuracy	0.9320305
MNL (AreaType)	kap	0.9044396
MNL (DivType)	accuracy	0.2789281
MNL (DivType)	kap	-0.0771525
MNL (LUType)	accuracy	0.2630515
MNL (LUType)	kap	0.1650520

For this table, the most useful metric is overall accuracy, since these models predict discrete categories rather than continuous values. In this context, the practical reading is comparative rather than absolute: a model is more useful if it improves materially on the alternatives while remaining interpretable and behaviorally plausible. Because this is a multi-class transition problem with many possible outcomes, middling accuracy values are not automatically poor; the question is whether the selected model performs credibly relative to the alternatives and is good enough for the downstream simulation task.

Both decision tree and LASSO-regularized multinomial logit models have their strengths and limitations for land use type classification. The decision tree model offers better interpretability through its clear hierarchical structure and decision rules, making it easier to understand and explain the transition process. It also naturally handles non-linear relationships and interactions between variables. However, decision trees can be sensitive to small changes in the training data and may not capture smooth transitions between land use types.

The LASSO-regularized multinomial logit model provides automatic variable selection through L1 regularization, retaining only the most predictive variables. The model uses cross-validation to select the optimal regularization parameter, balancing model complexity and prediction accuracy. This approach avoids the computational singularity issues that can arise with standard MNL estimation when predictors are highly correlated.

We choose the regularized multinomial logit model for Land Use Type transition, as it is a model type that's been used in VisionEval (for example, the housing type choice model) and avoids the need to introduce a new model type. The LASSO regularization provides the additional benefit of automatic variable selection and improved numerical stability.

4.4 PREDICTLOCTYPE: LOCATION TYPE CLASSIFICATION AND TRANSITION

As discussed in Section 4.2, we use the same three location types in VERSPM: urban, town and rural. The current VERSPM uses Census Tract population density thresholds (< 161 people/sqm, <1241 people/sqm, and > 1241 people/sqm, respectively) to define Rural, Town and Urban.

For PredictLocType in the new VELandUse module, we take advantage of the fact that the urban-rural classification corresponds to Census Bureau's urban-rural classification and switch to use this official classification and data source.

However, we cannot directly use the Census Bureau's urban area classification, as the criteria for urban area changed substantially between 2010 and 2020 (https://www2.census.gov/geo/pdfs/reference/ua/Census_UA_CritDiff_2010_2020.pdf), besides the Census Bureau got rid of the Urban Cluster class altogether in 2020 (https://www2.census.gov/geo/pdfs/reference/ua/Census_UA_CritDiff_2010_2020.pdf). To address these issues, we train a random forest model to predict the urban area classification using the 2020 data and then apply the classifier to predict the urban area classification for 2010. We re-classify urban areas with population less than 50,000 as urban clusters. To be consitent with VisionEval geography, we use the Census block group as the primary unit of analysis and aggregate Census Bureau's block-level urban area classification to block groups.

4.4.1 Data Summary

We prepare data for the random forest location type classification model by joining block-level data with block group and census tract aggregation, including demographics, built environment (housing and network density), and employment data from LEHD. We also add Census region and division information.

A summary of the block-level variables is shown in Table 4.11:

Table 4.11: Summary statistics of block-level variables by urban/rural classification in 2020

Characteristic	Overall N = 10,791,057[1]	Urban N = 3,705,420[1]	Town N = 911,217[1]	Rural N = 6,174,420[1]
HOUSING	2 (0, 12)	11 (0, 26)	6 (0, 14)	0 (0, 4)
POP	4 (0, 29)	29 (0, 68)	13 (0, 32)	0 (0, 9)
ALAND	25,102 (8,184, 186,767)	17,553 (7,878, 37,129)	13,738 (6,809, 30,979)	68,845 (9,104, 714,537)
POPDEN	0.0 (0.0, 4.2)	5.3 (0.0, 11.6)	3.0 (0.0, 7.1)	0.0 (0.0, 0.1)
HUDEN	0.02 (0.00, 1.81)	2.11 (0.00, 4.37)	1.43 (0.00, 3.17)	0.00 (0.00, 0.07)
BG_POPDEN	0.3 (0.0, 3.2)	4.8 (2.2, 8.9)	1.5 (0.6, 3.7)	0.0 (0.0, 0.2)
BG_HUDEN	0.12 (0.02, 1.39)	2.04 (0.90, 3.72)	0.73 (0.26, 1.75)	0.02 (0.01, 0.07)
BG_ALAND_AC	5,698 (495, 34,770)	323 (149, 898)	872 (300, 2,812)	26,453 (9,360, 85,978)
BG_POP	1,283 (912, 1,866)	1,531 (1,061, 2,266)	1,230 (892, 1,789)	1,173 (854, 1,655)
BG_HOUSING	591 (427, 848)	643 (447, 954)	579 (422, 824)	569 (417, 794)
CT_POPDEN	0.2 (0.0, 2.7)	4.5 (2.1, 8.2)	0.9 (0.3, 2.1)	0.0 (0.0, 0.1)
CT_HUDEN	0.10 (0.02, 1.18)	1.93 (0.87, 3.45)	0.39 (0.15, 0.95)	0.02 (0.01, 0.06)
CT_ALAND_AC	22,146 (1,782, 101,758)	1,026 (506, 2,629)	5,459 (2,011, 15,040)	83,312 (34,081, 215,094)
CT_POP	4,243 (2,966, 5,837)	4,858 (3,495, 6,514)	4,534 (3,380, 6,026)	3,855 (2,677, 5,334)
CT_HOUSING	1,918 (1,392, 2,582)	2,035 (1,479, 2,731)	2,080 (1,601, 2,703)	1,827 (1,320, 2,477)
totEmp	232 (94, 594)	345 (111, 1,069)	430 (174, 942)	182 (82, 400)
agrResEmp	0 (0, 12)	0 (0, 0)	0 (0, 3)	5 (0, 25)
nonArgResEmp	210 (80, 557)	342 (110, 1,060)	418 (168, 923)	156 (67, 355)
svcEmp	59 (16, 181)	118 (35, 359)	120 (41, 309)	36 (10, 104)
retEmp	13 (2, 49)	22 (3, 103)	34 (8, 105)	9 (1, 29)
entEmp	10 (0, 50)	24 (1, 104)	31 (4, 98)	6 (0, 26)
CT_totEmp	919 (437, 1,910)	1,273 (523, 2,987)	1,733 (961, 2,853)	724 (378, 1,354)
CT_agrResEmp	8 (0, 43)	0 (0, 2)	4 (0, 26)	26 (6, 72)
CT_nonArgResEmp	851 (396, 1,819)	1,261 (517, 2,963)	1,687 (925, 2,795)	644 (330, 1,250)
CT_svcEmp	273 (114, 620)	445 (182, 1,077)	570 (276, 1,018)	195 (85, 394)
CT_retEmp	73 (25, 189)	115 (31, 330)	174 (74, 366)	53 (21, 118)
CT_entEmp	62 (17, 175)	113 (31, 288)	155 (63, 303)	39 (12, 103)
DIVISION				
East North Central	1,666,950 (15%)	625,990 (17%)	172,426 (19%)	868,534 (14%)

Characteristic	Overall N = 10,791,057[1]	Urban N = 3,705,420[1]	Town N = 911,217[1]	Rural N = 6,174,420[1]
East South Central	825,832 (7.7%)	201,565 (5.4%)	86,958 (9.5%)	537,309 (8.7%)
Middle Atlantic	941,302 (8.7%)	484,639 (13%)	73,707 (8.1%)	382,956 (6.2%)
Mountain	1,179,615 (11%)	287,659 (7.8%)	91,864 (10%)	800,092 (13%)
New England	401,202 (3.7%)	205,609 (5.5%)	23,435 (2.6%)	172,158 (2.8%)
Pacific	1,102,340 (10%)	533,204 (14%)	80,167 (8.8%)	488,969 (7.9%)
South Atlantic	1,717,956 (16%)	712,477 (19%)	115,970 (13%)	889,509 (14%)
West North Central	1,473,430 (14%)	234,339 (6.3%)	139,388 (15%)	1,099,703 (18%)
West South Central	1,482,430 (14%)	419,938 (11%)	127,302 (14%)	935,190 (15%)
REGION				
Midwest	3,140,380 (29%)	860,329 (23%)	311,814 (34%)	1,968,237 (32%)
Northeast	1,342,504 (12%)	690,248 (19%)	97,142 (11%)	555,114 (9.0%)
South	4,026,218 (37%)	1,333,980 (36%)	330,230 (36%)	2,362,008 (38%)
West	2,281,955 (21%)	820,863 (22%)	172,031 (19%)	1,289,061 (21%)

[1] Median (Q1, Q3); n (%)

4.4.2 Model Training

We create a stratified split of the data by state to ensure geographic representation in both training and test sets. The training set is 80% of the data, while the test set is 20% of the data. We use the group_initial_split function from the rsample package to create the split.

For classification, we use the random forest model because it is a powerful and flexible method, and it is less sensitive to overfitting than other methods. Compared to the decision tree model, random forest is less interpretable, but it can capture more complex relationships between variables. In the case of location type classification, since the intrepretability is not a priority, we prefer the random forest model. We use the ranger package to train the random forest model. After tuning, we use 2000 trees and maximum depth of 30 in the random forest model, and we use the impurity parameter to calculate variable importance.

4.4.2.1 Random Forest

Table 4.12 shows the variable importance for the random forest model.

Table 4.12: Variable importance for the random forest location type classification model

Variable	Importance
CT_POPDEN	660760.35
CT_HUDEN	596738.14
CT_ALAND_AC	459396.76
BG_POPDEN	452772.03
BG_HUDEN	335738.57
D3A	325424.03
BG_ALAND_AC	262081.95
POPDEN	99316.45
HUDEN	92487.54
CT_agrResEmp	90618.82
CT_nonArgResEmp	59463.86
BG_POP	56422.82
CT_totEmp	55735.12
CT_POP	53648.88
nonArgResEmp	51493.55
CT_retEmp	50401.75
CT_svcEmp	50209.38
CT_HOUSING	47609.03
CT_entEmp	46647.98
totEmp	46516.12
BG_HOUSING	40536.45
ALAND	38994.18
POP	35285.82
STATEFP	33690.47
svcEmp	32983.35

Variable	Importance
agrResEmp	32000.92
HOUSING	31014.43
retEmp	28468.04
entEmp	24959.49
DIVISION	23202.09
REGION	18339.59

Table 4.13: Prediction accuracy of the random forest location type classification model

	Urban	Town	Rural
Urban	656649	559	19526
Town	6527	138879	23429
Rural	17841	10077	1150515

Table 4.13 shows the performance metrics for the random forest model. The random forest model has a high accuracy of 96.15%.

4.4.2.2 Spatial Validation

We then examine and compare how the random forest model performs spatially by analyzing its predictions in Salem, OR (Figure 4.13) and Corvallis, OR (Figure 4.14).

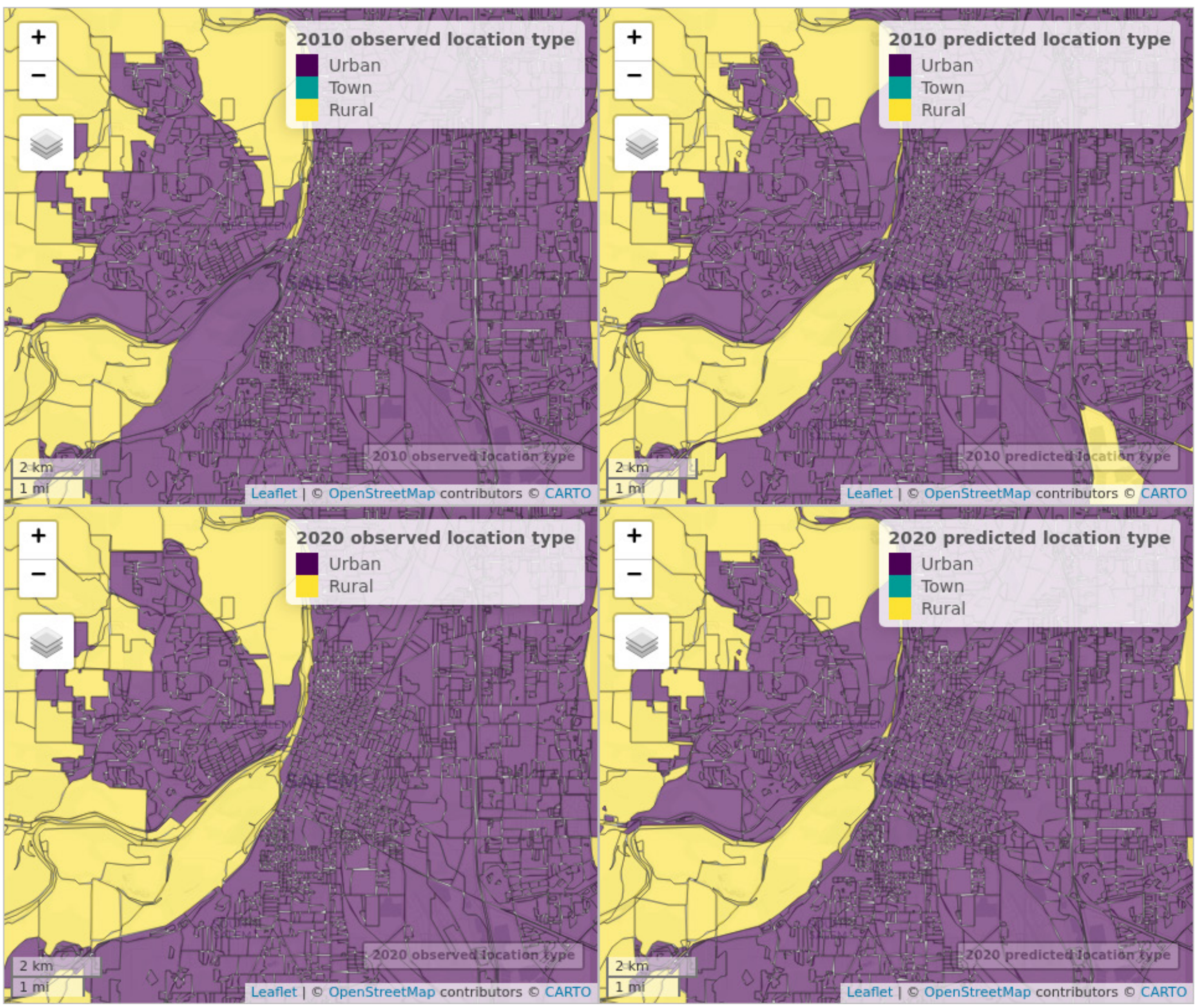

Figure 4.13: Map of Salem location type classification (2010 Census Block Geography)

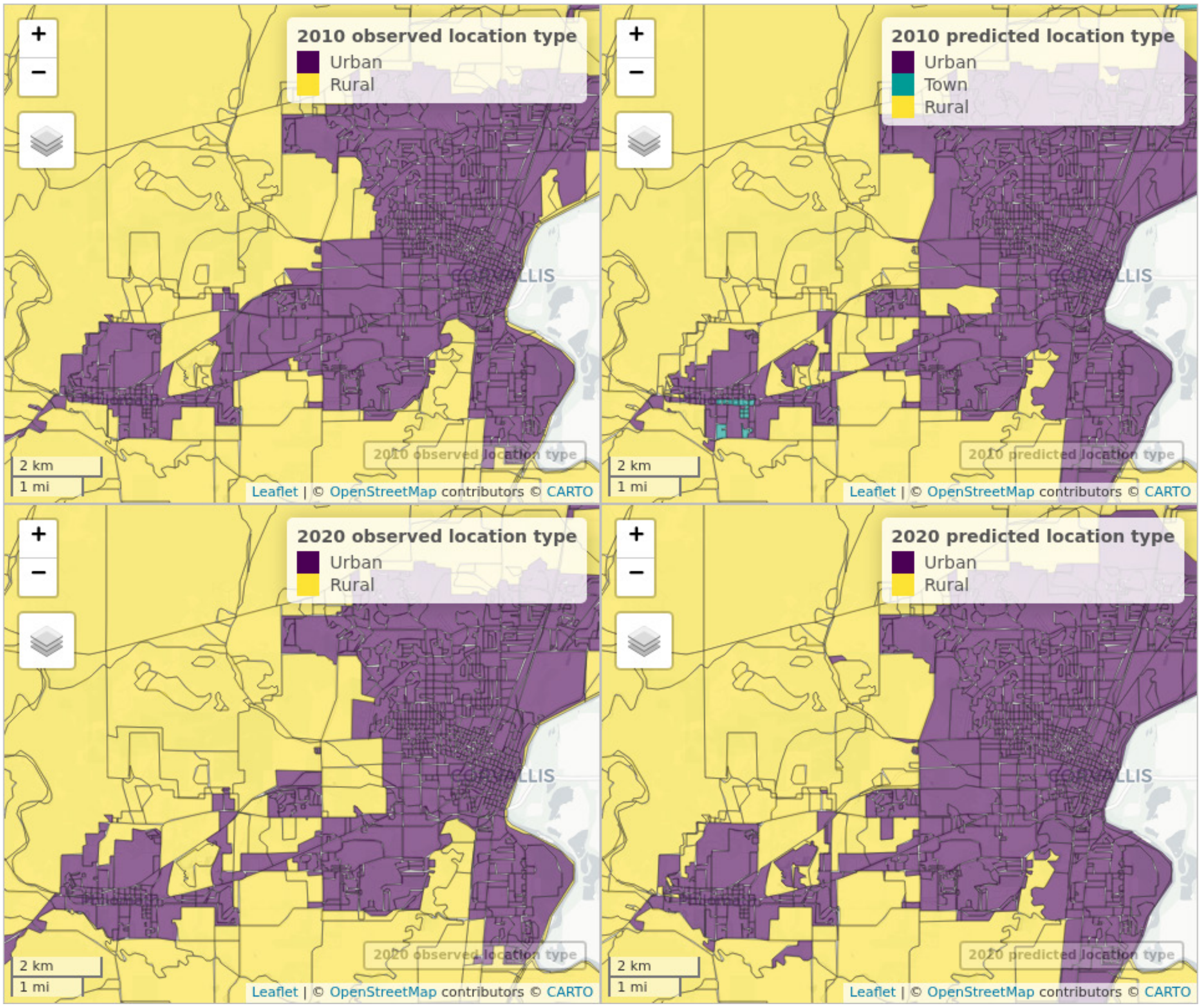

Figure 4.14: Map of Corvallis location type classification (2010 Census Block Geography)

The synchronized maps above show: 1. 2010 observed location type (top left) 2. 2010 predicted location type (top right) 3. 2020 observed location type (bottom left) 4. 2020 predicted location type (bottom right)

In the HTML version of the report (https://sapporo.usp.pdx.edu/files/a0fb89/), the maps are live and synchronized so you can zoom and pan all views simultaneously to compare how each model captures the urban-cluster-rural classification at the block level. In general, the random forest model performs well for prediction urban area types, as shown by the higher accuracy and reasonable spatial patterns.

4.4.3 Location Type

After predicting urban area types (U/C/R) for blocks, we aggregated the block-level urban area type predictions to the block group level (UATYPE).

Figure 4.15 and Figure 4.16 shows the maps of block group location type for 2010 and 2020 for Salem and Corvallis.

NULL

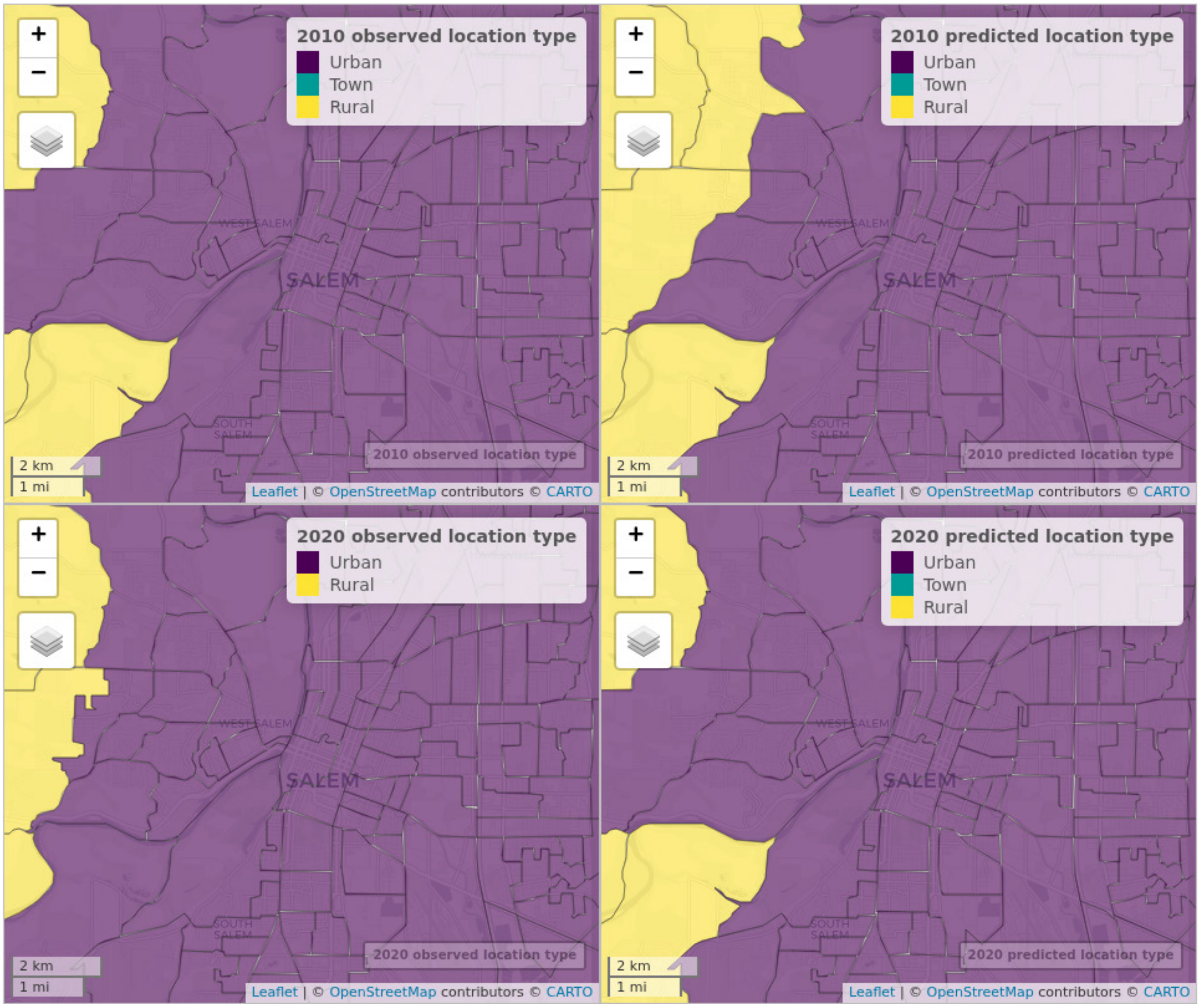

Figure 4.15: Map of Salem land use type classification

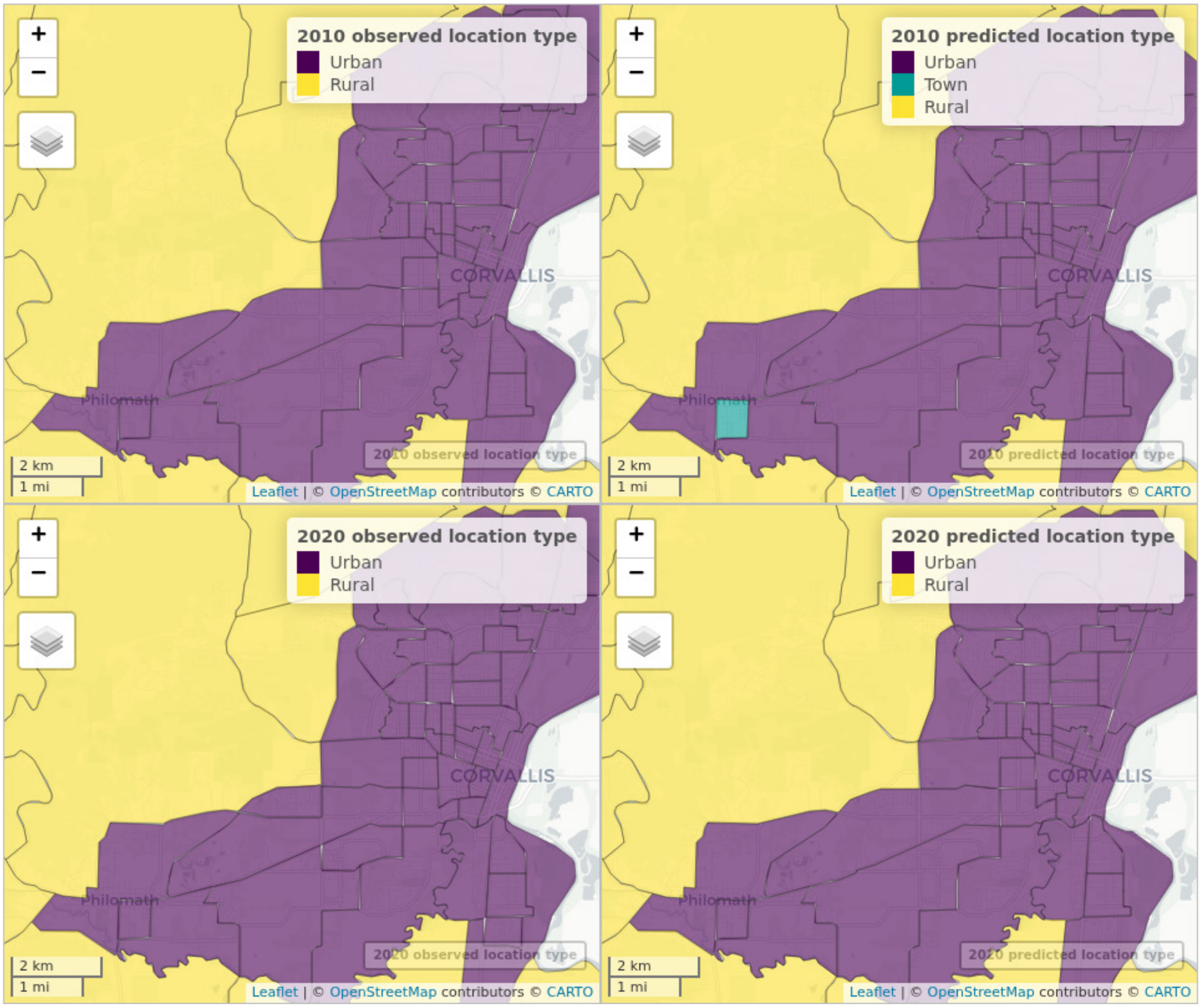

Figure 4.16: Choropleth map showing observed and predicted location types for Corvallis block groups

4.4.4 Location Type Transition

Adopted the same method in AreaType and Diversity Type transition models, we model the transition of location type at the block group (Bzone) level. Figure 4.17 shows the transition matrix of block group location type between 2010 and 2020.

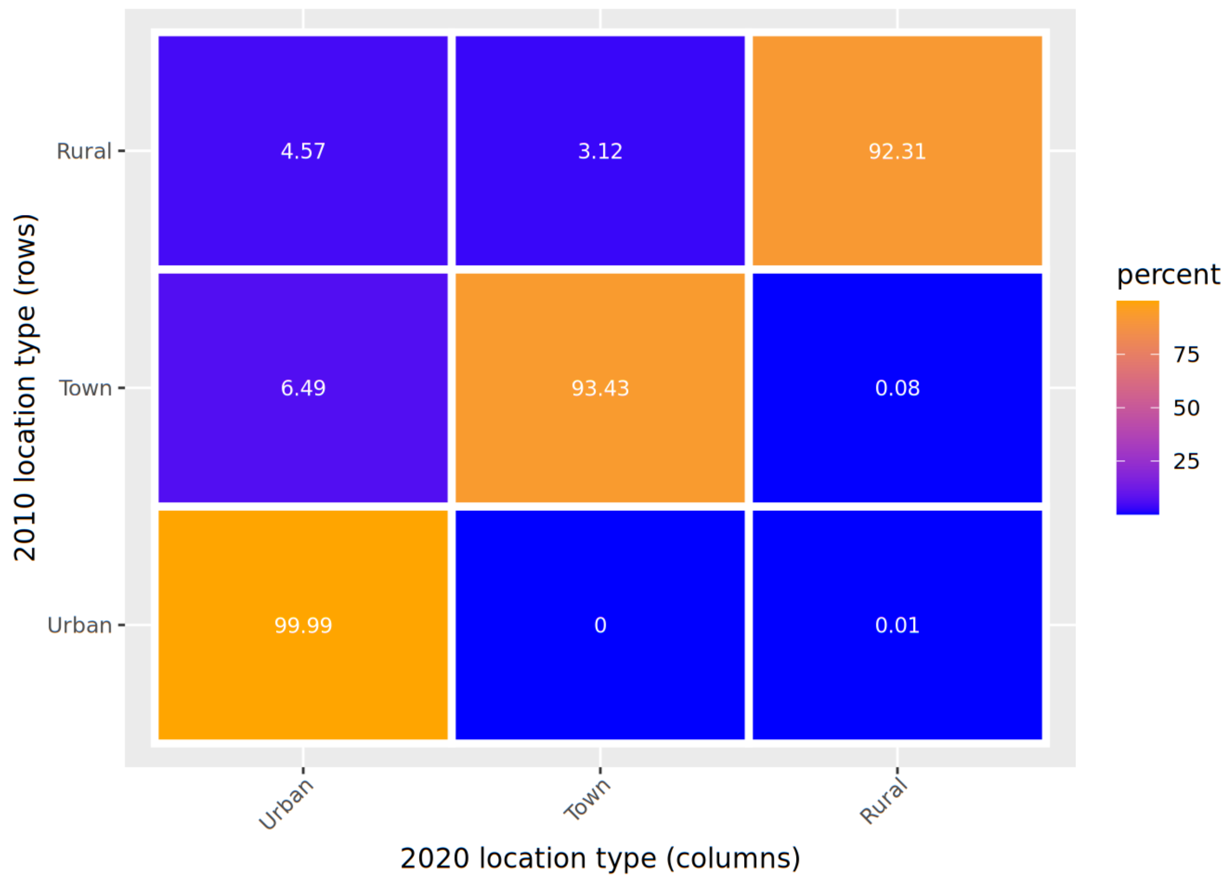

Figure 4.17: Block group location type transition matrix between 2010 and 2020

Table 4.14 shows the LASSO multinomial logit regression results of Location Type transition model, including the following variables:

- Activity (Housing + Employment) Density (D1D)
- Destination accessibility (D5)
- Distance to CBD (DistToCBD)
- Distance to Ramp (DistToRamp)
- Percentage Block group Area of Steep Slope (> 20%) (PctSteepSlope)
- Distance to Transit Stop (DistToStop)
- Distance to Fixed Guideway Transit Station (DistToFgwSta)

Table 4.14: LASSO multinomial logit model coefficients for block group location type transitions

	Urban	Town	Rural
(Intercept)	2.282	-1.177	-1.105
LocTypeUrban	8.778	-0.116	0.000
LocTypeTown	0.000	5.842	-0.860
LocTypeRural	-2.733	0.000	3.226
D1D	0.000	0.006	-0.098
DistToCBD	0.040	-0.141	0.000
DistToRamp	-0.105	0.000	0.000
PctSteepSlope	-0.020	0.000	0.002
DistToStop	-0.045	0.000	0.001
DistToFgwSta	0.002	0.000	0.000
NumHh	0.000	0.000	-0.001
TotEmp	0.000	0.000	0.000

Table 4.15: Prediction accuracy of multinomial logit model for location type transitions

.metric	.estimator	.estimate
accuracy	multiclass	0.973
kap	multiclass	0.946

The LASSO-regularized multinomial logit model avoids the convergence issues we saw with the unregularized specification and achieves a prediction accuracy of 97.34%.

4.5 ASSIGND3D4LEVELS: D3 AND D4 CLASSIFICATION

This section documents the classification logic used by AssignD3D4Levels, the VELandUse function that converts continuous urban-design and transit-service measures into discrete D3Lvl and D4Lvl categories. The purpose of these classes is to improve data quality, simplify scenario inputs, and better align the design and transit measures with observed travel behavior.

Currently, VisionEval directly simulates two measures of urban design and quality of transit service: - D3bpo4 (Pedestrian-Oriented Network Design) through VESimLandUse (https://github.com/VisionEval/VisionEval/blob/main/sources/modules/VESimLandUse/inst/module_docs/Simulate4DMeasures.md) - D4c (aggregate peak period transit service) through VETransportSupply (https://github.com/VisionEval/VisionEval/blob/main/sources/modules/VETransportSupply/inst/module_docs/AssignTransitService.md)

While they represents the best individual measure available for each built environment dimension, they do not necessarily represent the measures associated with actual travel behavior.

Furthermore, the implementation makes it hard for users to create scenarios as neither measure is intuitive nor regularly used.

The new implementation aims to overcome both limitations by first classifying urban design and quality of transit service for each Bzone into discrete categories and then either modeling the categories for future simulation years or reading them in from user-provided scenario inputs.

See Table 9.2 in the Appendix for a lookup table between variable names and their definitions.

We explore methods for classifying Census block groups (CBGs) based on their D3 (urban design) and D4 (transit service) characteristics. The goal is to develop a classification method that is both interpretable and meaningfully related to travel behavior outcomes.

One method is to cluster CBGs into groups that are close to other CBGs across the dimensions of D3 and D4 measures. The limitations of clustering methods include:

- Interpretation challenges - clusters often lack clear meaning or interpretation
- Arbitrary cluster counts - the number of clusters must be predetermined
- Unclear behavioral relationships - clusters may not meaningfully relate to travel outcomes
- Missing data sensitivity - clustering methods typically cannot handle missing values

Another method is to use a decision tree-based approach, similar to the approach we used for classifying land use types. The advantages of a tree-based method are:

1. The classification rules are transparent and easy to interpret.
2. The resulting categories have clear relationships with travel behavior outcomes (e.g., transit use, walking trips).
3. The classification can be easily applied to new data or scenarios.
4. Tree-based method can handle missing value.

In this analysis, we use decision trees to classify block groups based on their D3 and D4 characteristics. Separate decision trees are trained on transit and walking outcomes. For each outcome, we explore a few different outcome variables:

1. For quality of transit service:
    - Share of transit commuters from ACS;
    - Presence of NHTS households making any transit trips;
    - Presence of NHTS households used transit in the last 30 days.

2. For urban design:
   - o Share of walking commuters from ACS;
   - o Presence of NHTS households making any walking trips.

Note that even though we use transit and walking outcomes in training decision trees, the purpose is not to predicting transit or walking outcomes, but to use them as “ground truth” in classifying urban design and quality of transit service based on the rationale that the best built environment supportive of transit and walking are more likely to see the highest transit and walking use, and vice versa.

4.5.1 D4 (Quality of Transit Service)

4.5.1.1 D4 Decision tree trained on CBG transit commuting share from ACS

Transit share ~ D4A + D4C + DistToStop + DistToFgwSta + UZAAVRM

The transit commuting share variable from ACS has the best spatial coverage - theoretically cover every block group, even though the margin of error for many block groups is high. The resulted decision tree is shown below (Figure 4.18) and the variable importance scores are shown inTable 4.17.

To read the tree inFigure 4.18, start at the root node and follow the branch that matches the block group’s value for the splitting variable shown at that node. Each successive split refines the classification using another transit-related variable, and the terminal leaves represent the final transit-service groupings. Splits near the top of the tree indicate the most influential thresholds for separating stronger from weaker transit-service contexts; in this case, readers should pay particular attention to whether the first splits are being driven by service frequency (D4C) or distance-based access measures such as D4A, DistToStop, and DistToFgwSta.

The numerical leaf values in the plotted tree are not the final labels used elsewhere in this section. After prediction, those leaves are converted into ordered transit-service classes and carried forward into the box plots, maps, and saved outputs. In other words, the tree figure shows the threshold structure, while the later class labels show the same leaves after they have been converted into a consistent ordinal coding.

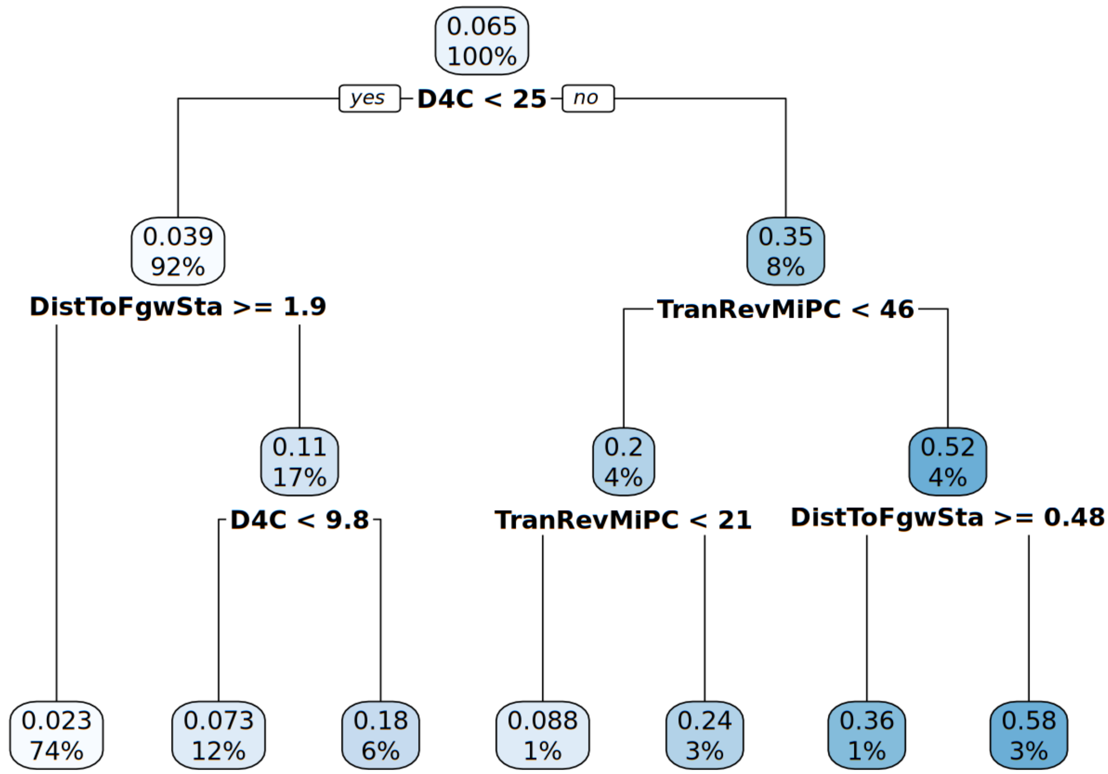

Figure 4.18: Decision tree for classifying transit service quality based on CBG transit commuting share

We can use \(R^2\) (rsq) and root mean squared error (rmse) to measure the prediction accuracy of the decision tree on an out of sample test set (Table 4.16). The variable importance scores (Table 4.17) shown below indicate how much each variable contributes to making accurate predictions - higher scores mean that variable plays a bigger role in the decision tree's prediction process. Note that while the variable importance scores show all variables that were considered by the algorithm during training, only the most predictive variables end up being used in the final decision tree splits. Variables with lower importance scores may be excluded from the tree if they don't provide enough additional predictive power beyond what's already captured by the primary splitting variables, or if including them would violate the minimum bucket size or maximum depth constraints we set in the rpart.control() parameters.

Table 4.16: Prediction accuracy of decision tree for transit service quality classification based on ACS transit commuting share

.metric	.estimate
rsq	0.630
rmse	0.082

For the ACS-based regression trees, higher R^2 and lower RMSE indicate better fit. In practical terms here, an R^2 around 0.6 means the tree is capturing a substantial share of the cross-sectional variation in transit share, while a much lower value would indicate that the available built-environment variables are only weakly tracking the outcome. RMSE should be read in the units of the outcome itself, so it is most useful for comparing competing models trained on the same dependent variable rather than for comparison across different outcomes.

Table 4.17: Variable importance for transit service quality decision tree based on ACS transit commuting share

Variable	Importance
D4C	958.6
DistToFgwSta	494.8
TranRevMiPC	284.1
DistToStop	56.1
D4A	23.8

Below we summarize key variables by class in a series of box plots (Figure 4.19).

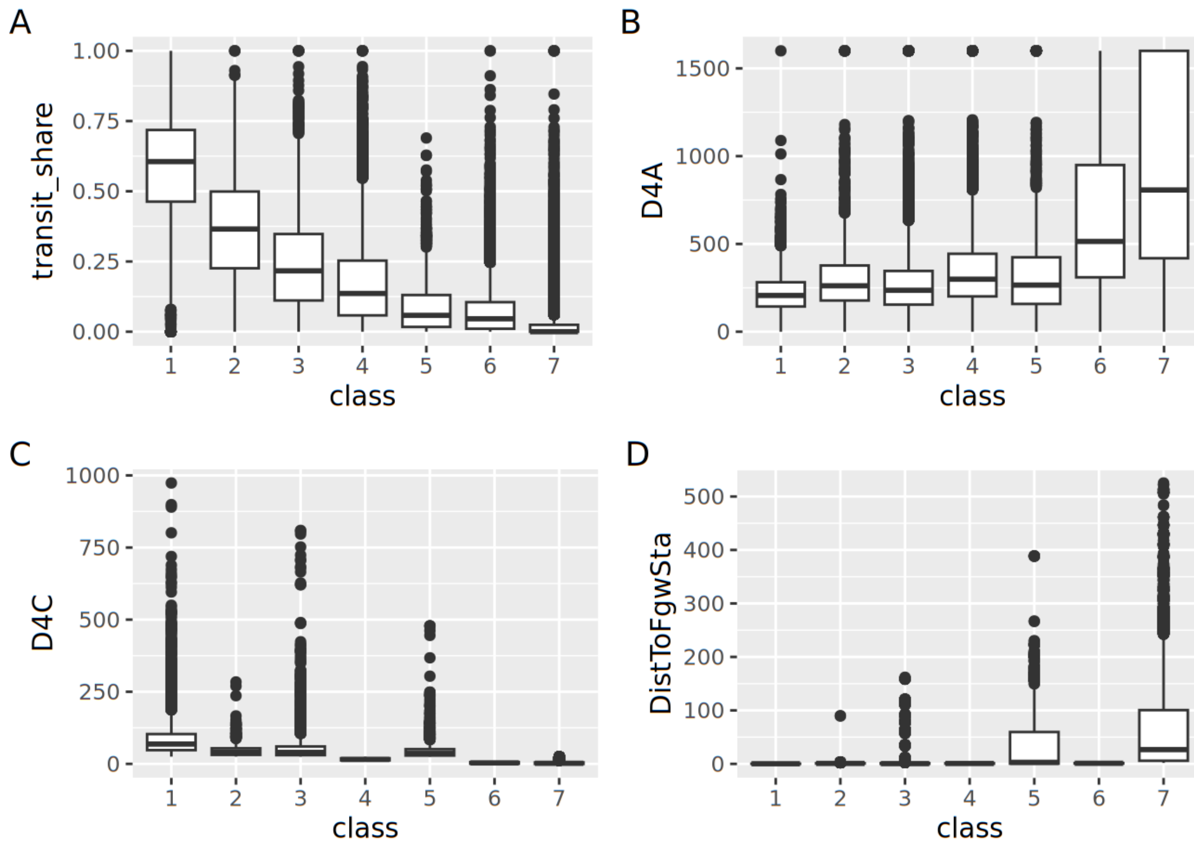

Figure 4.19: Distribution of key variables by transit service quality class classified on CBG transit commuting share

We use interactive map (Figure 4.20) to examine the spatial pattern of classification results (Interactive map only vailable in html format).

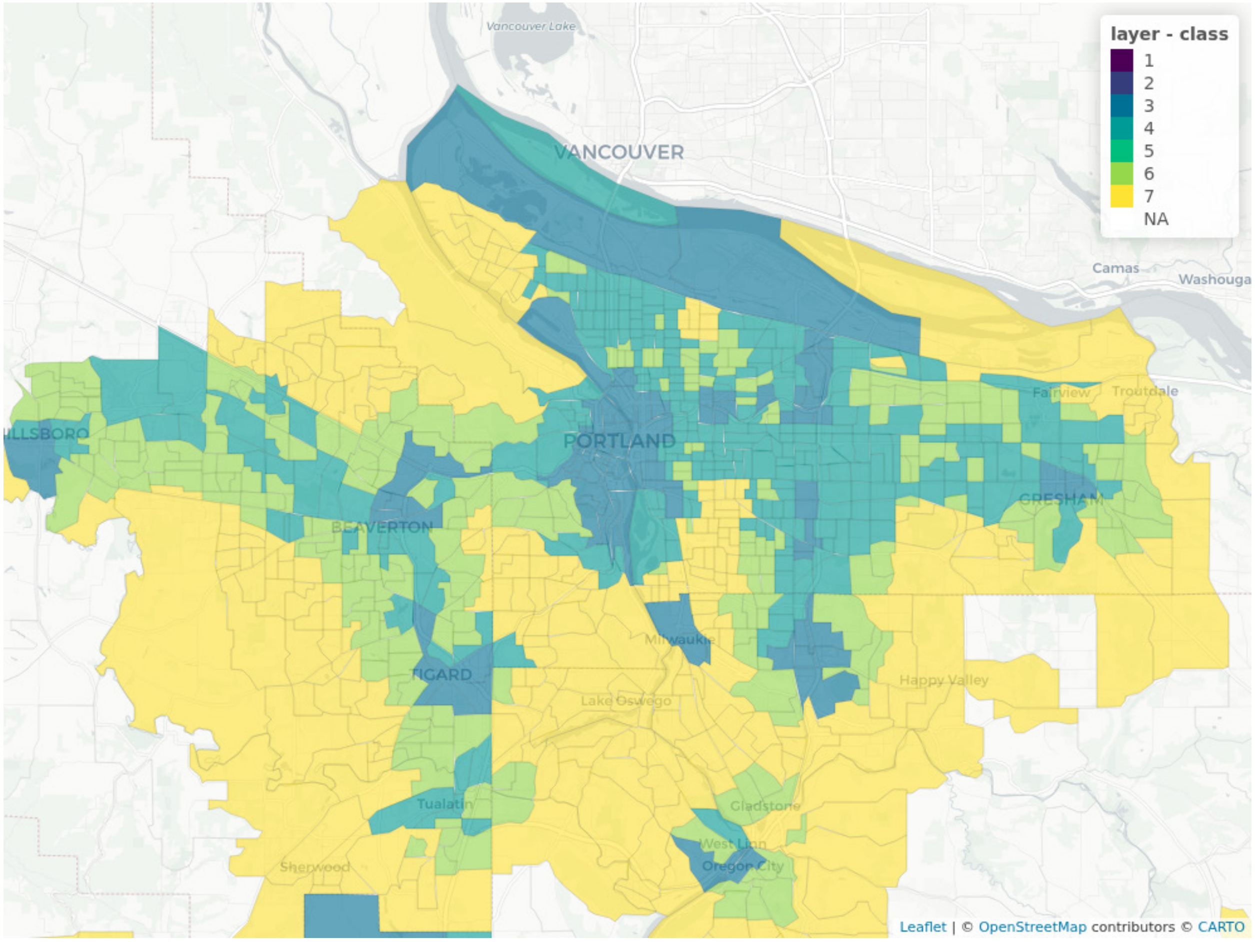

Figure 4.20: Choropleth map of Oregon and Washington block groups colored by transit service quality class based on CBG transit commuting share

4.5.1.2 D4 Decision Tree trained on NHTS households making at least one transit trip

To consider all transit trips, we use the joined NHTS-SLD dataset and train a decision tree on whether a CBG containing NHTS households making at least one transit trip.

Any NHTS Transit trip ~ D4A + D4C + DistToStop + DistToFgwSta + HasFgwTransit + TranMiPC

One problem here is that the number of CBGs including at least one household making a transit trip in the survey day is small. Another problem is that we attribute transit trips to the households' residence CBG, which may cause distortion, as a household member taking a transit trip anywhere will be attributed to their residence CBG.

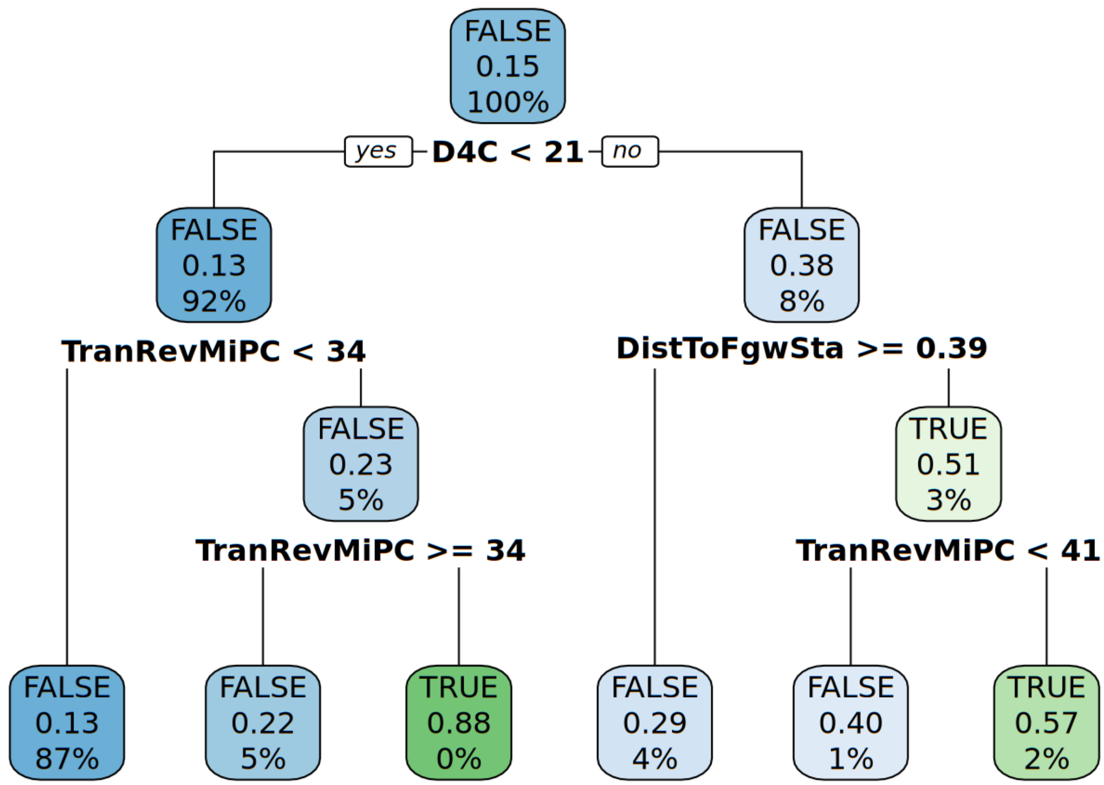

Figure 4.21: Decision tree for classifying transit service quality based on NHTS transit trip occurrence

Since the outcome here is binary (either a household makes transit trips or not), we can evaluate how well our model predicts transit use with three metrics: precision measures how often the model is correct when it predicts transit use, recall measures what percentage of actual transit users the model correctly identifies, and the f1-score provides a balanced combination of both metrics. We calculate these metrics using a separate test dataset that wasn't used to train the model (Table 4.18).

Table 4.18: Prediction accuracy of decision tree for transit service quality classification based on NHTS transit trip occurrence

.metric	.estimator	.estimate
precision	binary	0.498
recall	binary	0.061
f_meas	binary	0.108

For these binary transit-use models, precision tells us whether positive predictions are trustworthy, while recall tells us whether the model is finding most of the actual transit users. In sparse-outcome settings like this one, low recall is often the more serious weakness because a model can look acceptable on precision while still missing most of the observed positives. The f_meas statistic is therefore the best compact summary here: values close to zero indicate weak classification performance, while higher values indicate a better balance between finding transit users and avoiding false positives.

Table 4.19: Variable importance for transit service quality decision tree based on NHTS transit trip occurrence

Variable	Importance
D4C	319.58
DistToFgwSta	128.84
TranRevMiPC	85.36
DistToStop	8.44
D4A	2.69

Below we summarize key variables by class in a series of box plots (Figure 4.22).

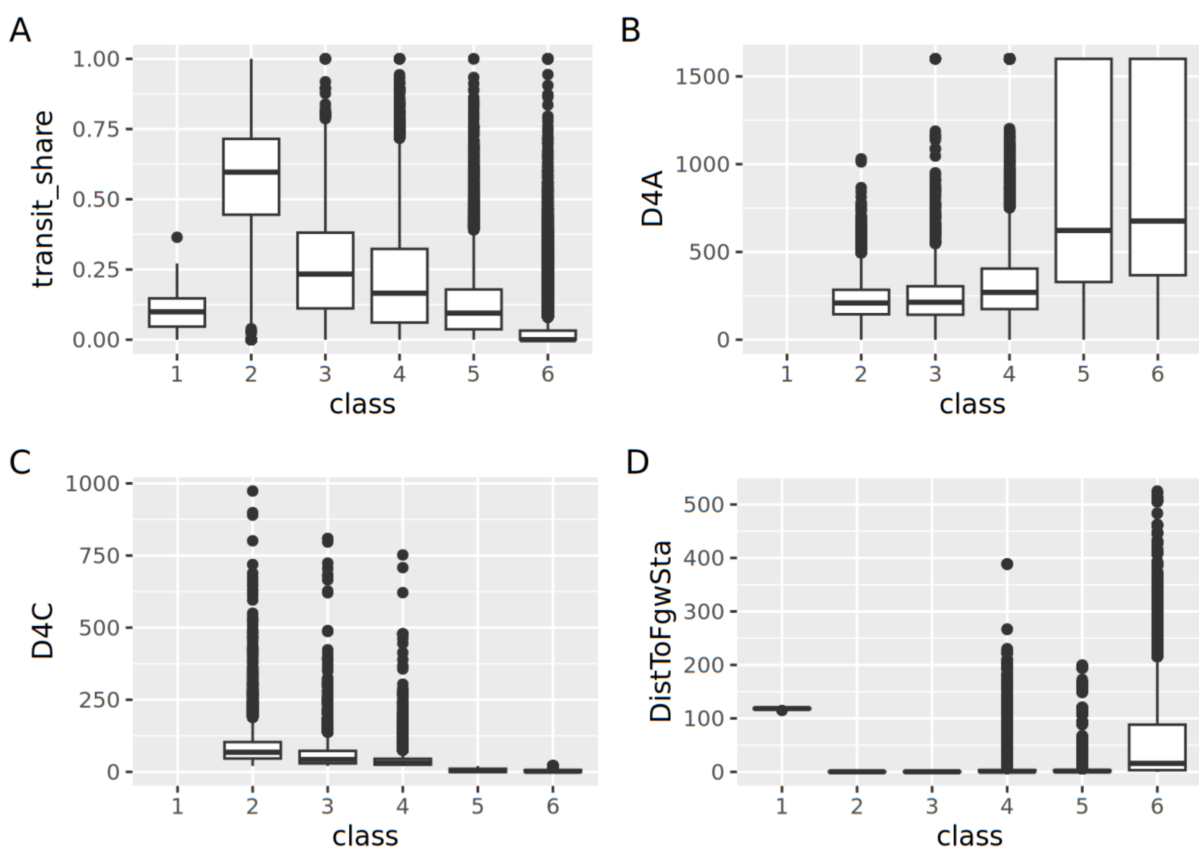

Figure 4.22: Distribution of key variables by class based on NHTS households making any transit trips

Again an interactive map for examining the spatial pattern of classification results (Available only in html format. Not showing in the report in Word format):

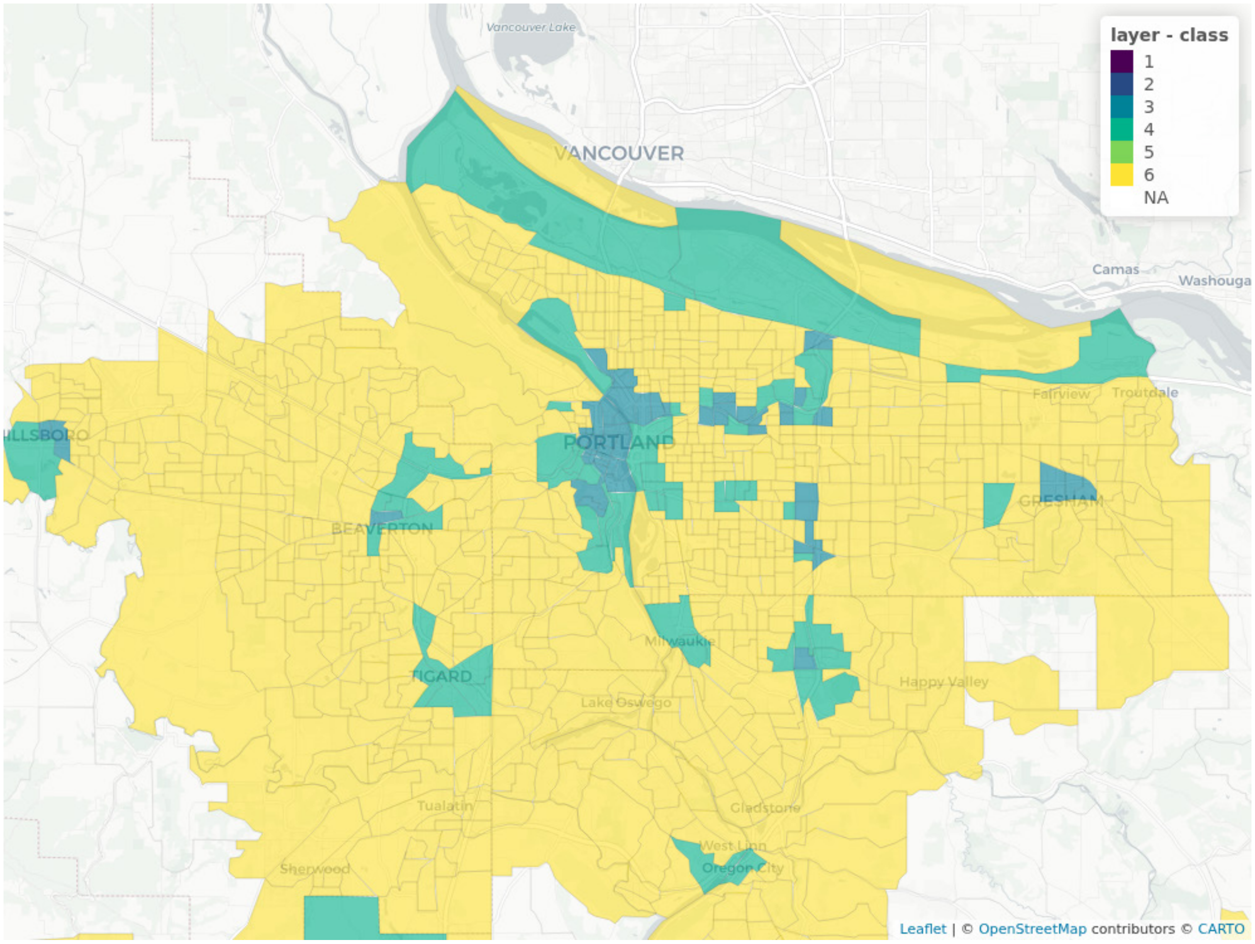

Figure 4.23: Choropleth map of block groups colored by transit service quality class based on NHTS transit trip occurrence

4.5.1.3 D4 Decision Tree trained on NHTS households making transit trips in the last 30 days

To overcome the problem of small size of NHTS households making at least one transit trip in the survey day, we explore using whether CBG containing NHTS households making at least one transit trips in the last 30 days (NHTS 2017 includes a question on the frequency of transit usage in the last 30 days for each household member).

Using transit in the last 30 days ~ D4A + D4C + DistToStop + DistToFgwSta + UZAAVRM

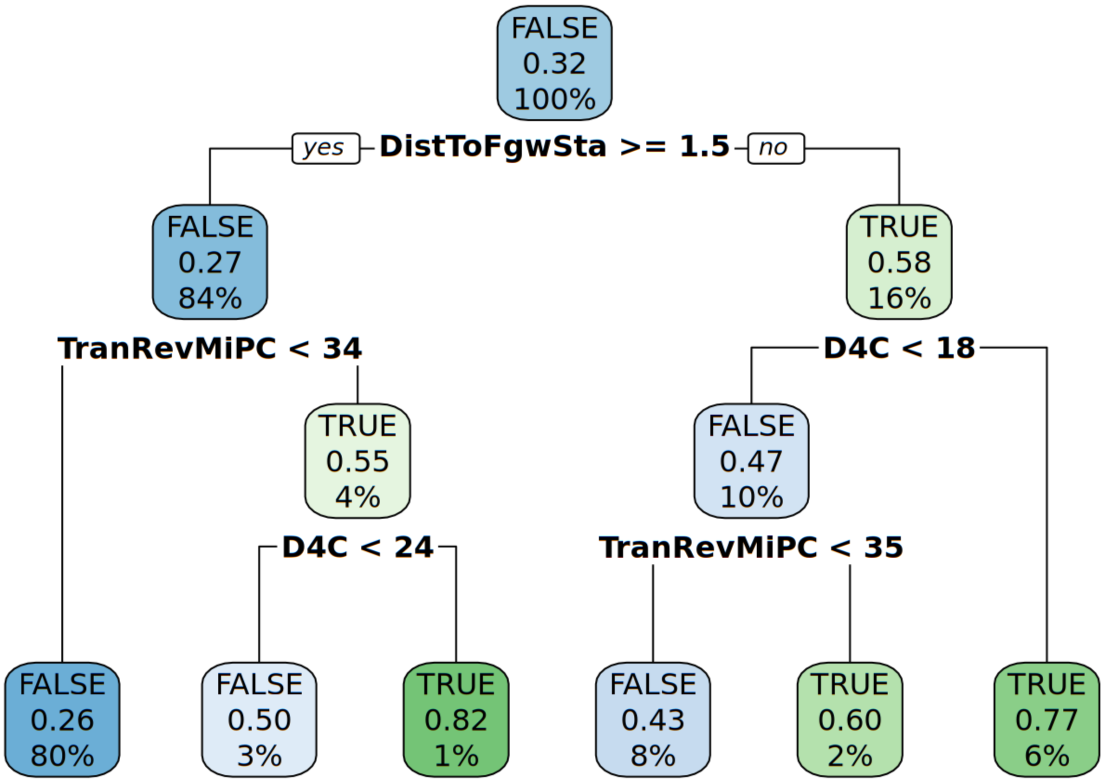

Figure 4.24: Decision tree for classifying transit service quality based on 30-day NHTS transit use

Similar to the decision tree trained on any transit trip, we can use precision, recall, and f1_score to measure the prediction accuracy of the decision tree on a test set (Table 4.20).

Table 4.20: Prediction accuracy of decision tree for transit service quality classification based on 30-day NHTS transit use

.metric	.estimator	.estimate
precision	binary	0.703
recall	binary	0.187
f_meas	binary	0.295

The same interpretation applies to the 30-day model. The main question is whether the gain in recall is large enough to make the classifier more useful in practice, since a model with somewhat lower precision can still be preferable if it identifies substantially more of the true positive cases.

Table 4.21: Variable importance for transit service quality decision tree based on 30-day NHTS transit use

Variable	Importance
DistToFgwSta	974.7
TranRevMiPC	275.4
D4C	264.6
DistToStop	57.5
D4A	42.0

Below we summarize key variables by class with a series of box plots (Figure 4.25).

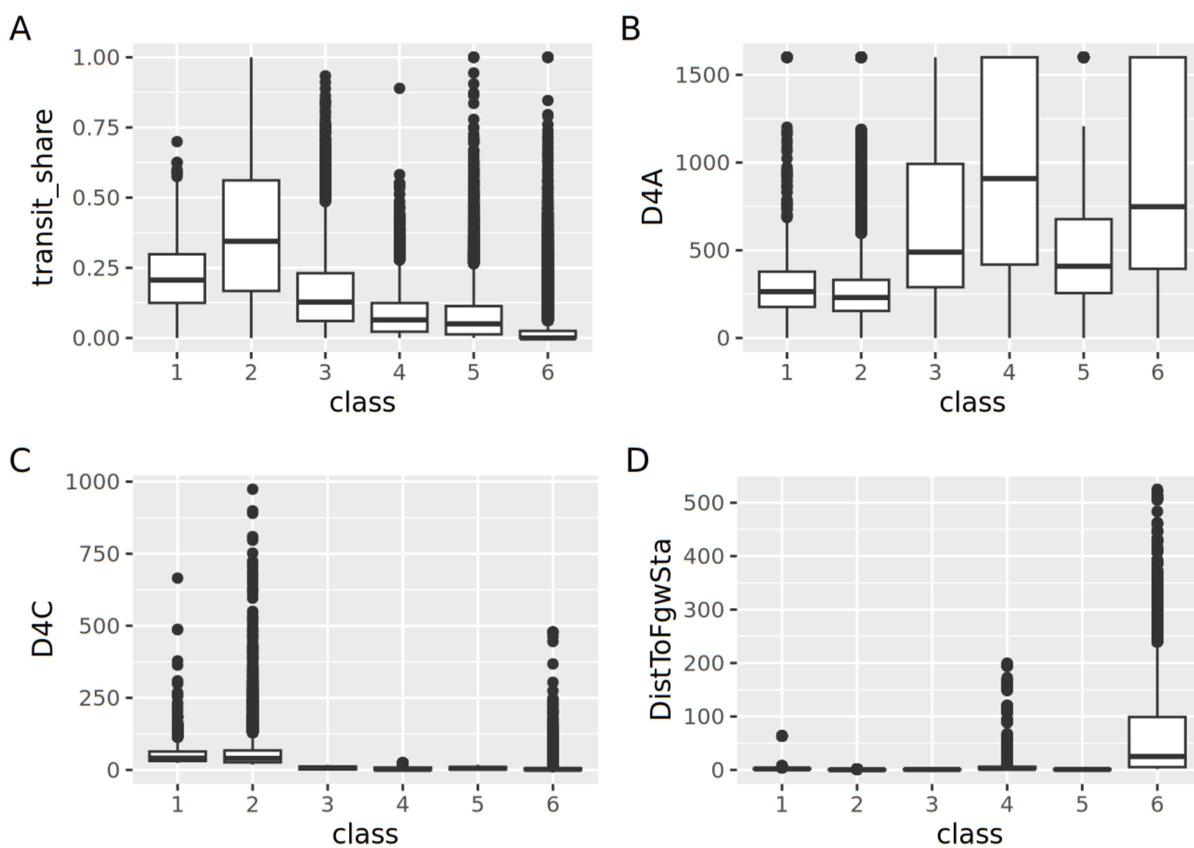

Figure 4.25: Distribution of key variables by transit service quality class classified on NHTS households making any transit trips in the last 30 days

An interactive map showing the spatial pattern of classifying CBGs based on NHTS households making any transit trips in the last 30 days (Available only in html format).

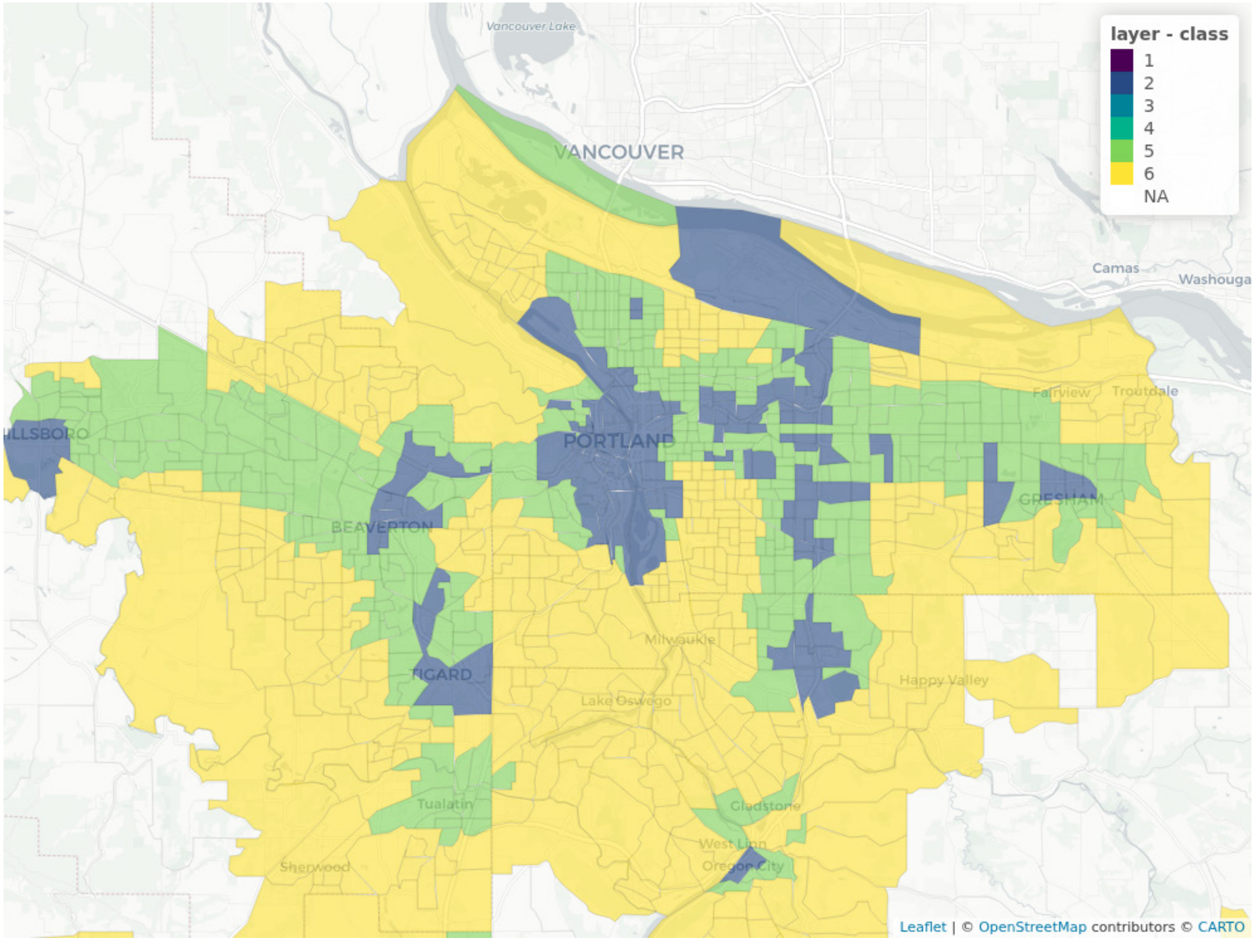

Figure 4.26: Choropleth map of block groups colored by transit service quality class based on 30-day NHTS transit use

4.5.1.4 Summary of D4 Classification

Table 4.22: Comparison of prediction accuracy across D4 transit service quality classification models

Outcome	rsq	rmse	precision	recall	f_meas
Transit Share (ACS)	0.63	0.082			
Any NHTS Transit Trip			0.498	0.061	0.108
Any NHTS Transit Trip in the Last 30 days			0.703	0.187	0.295

Putting metrics together (Table 4.22), we can see that the model performance metrics reveal distinct patterns across different transit outcome measures. The ACS transit share model, using a regression approach, achieves a moderate R^2 of 0.63 with an RMSE of 0.082, indicating it explains about 63% of the variance in transit commute mode share. In contrast, the classification models for NHTS transit trips show varying levels of success. The model predicting any NHTS transit trip achieves relatively low performance with a precision of 0.52 and particularly low recall of 0.06 (F1 score = 0.10), suggesting it struggles to identify actual transit users. However, the model predicting transit use in the last 30 days performs notably better, with improved precision (0.66) and recall (0.23), resulting in a higher F1 score of 0.34. This improvement likely reflects the larger sample of transit users when considering a 30-day window rather than a single day, though the still-low recall suggests challenges in identifying all transit users. The higher performance of the ACS-based model may be attributed to its focus on regular commute patterns rather than all-purpose transit trips captured in NHTS.

In short, the ACS-based D4 model is reasonably strong for this application, because it captures a substantial share of cross-sectional variation and is also the easiest outcome to interpret spatially. The NHTS-based D4 models are better treated as diagnostic checks than as preferred production models, because their recall remains too low to support a strong claim that they are reliably identifying the full set of transit-supportive contexts.

Figure 4.27 shows the variable importance comparison for the three models.

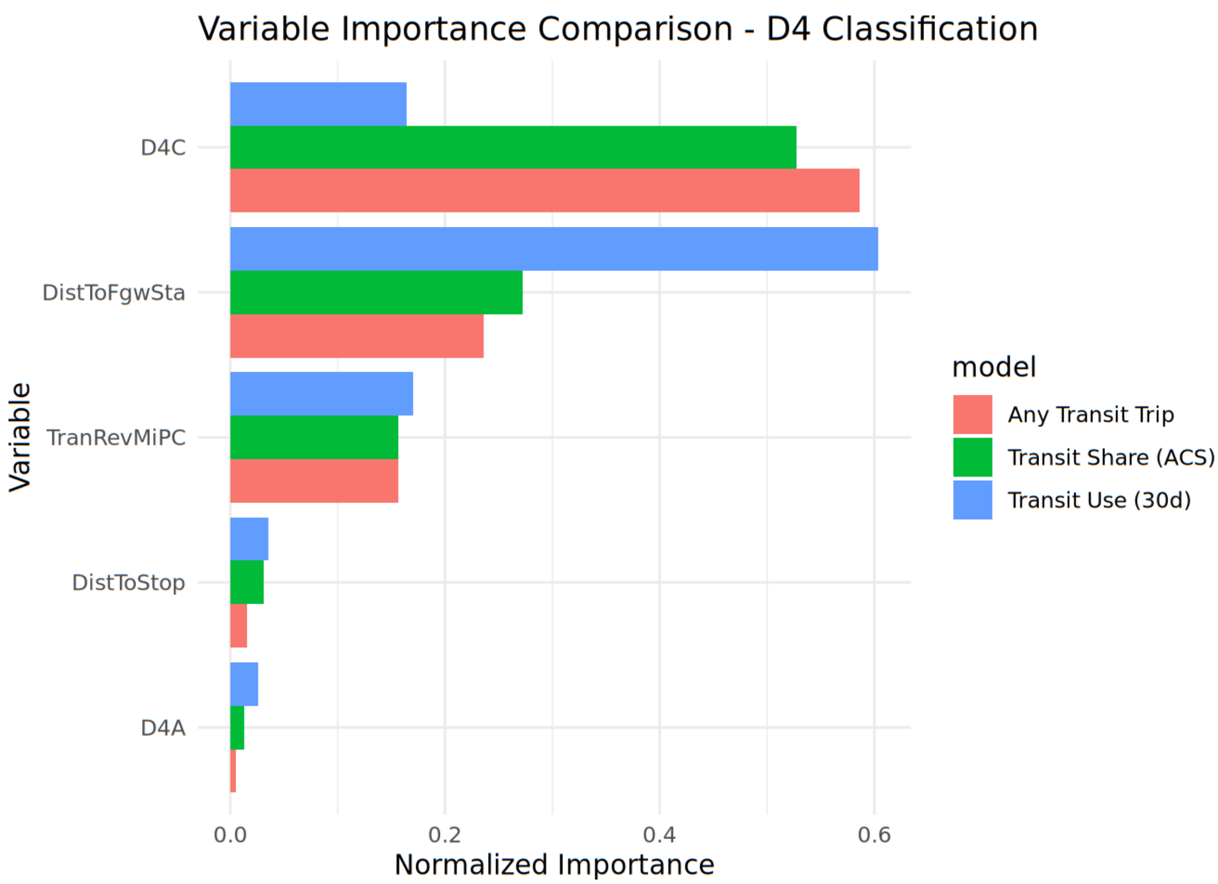

Figure 4.27: Variable importance comparison across D4 transit service quality classification models

Given the result, we prefer classifying quality of transit service using CBG transit commuting share variable as the ground truth, acknowledging its limitations discussed above.

4.5.2 D3 (Urban Design)

We use a similar method to classify urban design using walking as the “ground truth” outcome. Because we only have very limited number of urban design features available in SLD, we decide to include a D1 - density variable and D2 - land use mix in the input features besides the D3 - design variables. In addition, we experiment with CBG variables as well as a buffered version of each variable (half mile buffer from centroid of each CBG). The decision tree can pick either the buffered or non-buffered version depending on which is better at predicting the ground truth variable.

4.5.2.1 D3 Decision Tree trained on CBG walking commuting share from ACS

Walking share ~ D1D_hmbuf + D2A_JPHH_hmbuf + D3BPO4

Similar to the transit commuting share variable, walking commuting share from ACS has good spatial coverage, with a similar high margin of error problem for many block groups. The resulted dicsion tree is shown below (Figure 4.28).

The D3 tree is interpreted the same way: start at the top, read each split as a threshold rule, and treat the final leaves as empirically defined urban-design classes. Here the most informative splits tell the reader whether the classification is being driven primarily by street-network design (D3BPO4), surrounding activity density (D1D_hmbuf), or the local jobs-housing balance (D2A_JPHH_hmbuf). When one of those variables appears higher in the tree, it means that variable is more important to the final class structure than variables that appear only in lower branches.

As with the D4 tree, the plotted leaf values are then converted into ordered class labels before they appear in the later box plots, maps, and saved outputs. The important point for interpretation is therefore the relative ordering of the leaves and the thresholds that separate them, not the raw numeric prediction shown in the regression-tree node itself.

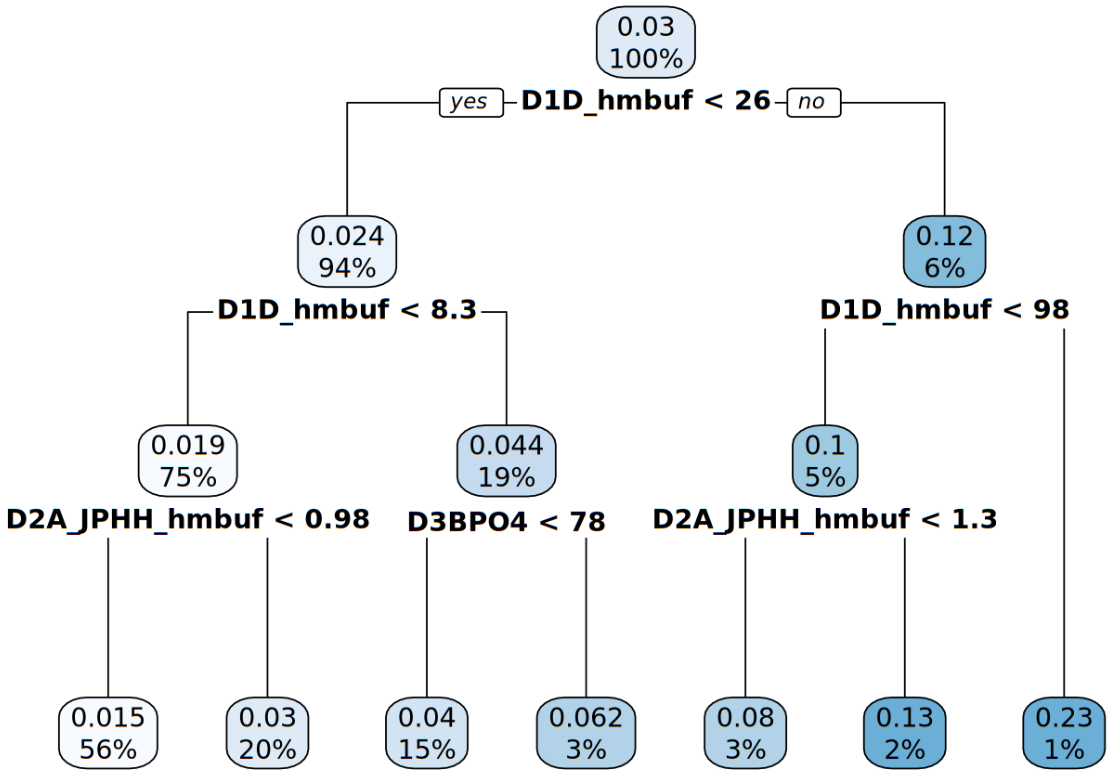

Figure 4.28: Decision tree for D3 urban design quality classification based on walking commuting share

\(R^2\) (rsq) and root mean squared error (rmse) for the prediction accuracy of the decision tree on an out-of-sample test set (Table 4.23):

Table 4.23: Prediction accuracy of decision tree for urban design quality classification based on ACS walking commuting share

.metric	.estimator	.estimate
rsq	standard	0.192
rmse	standard	0.059

The interpretation is the same as for the ACS-based D4 model: higher R^2 and lower RMSE indicate better fit, but the practical question is whether the model captures enough variation to be useful for classification. In this section, an R^2 well below the D4 transit-share model should be read as evidence that walking-supportive design is harder to recover from the available variables alone.

Summarize key variables by class in a series of box plots (Figure 4.29):

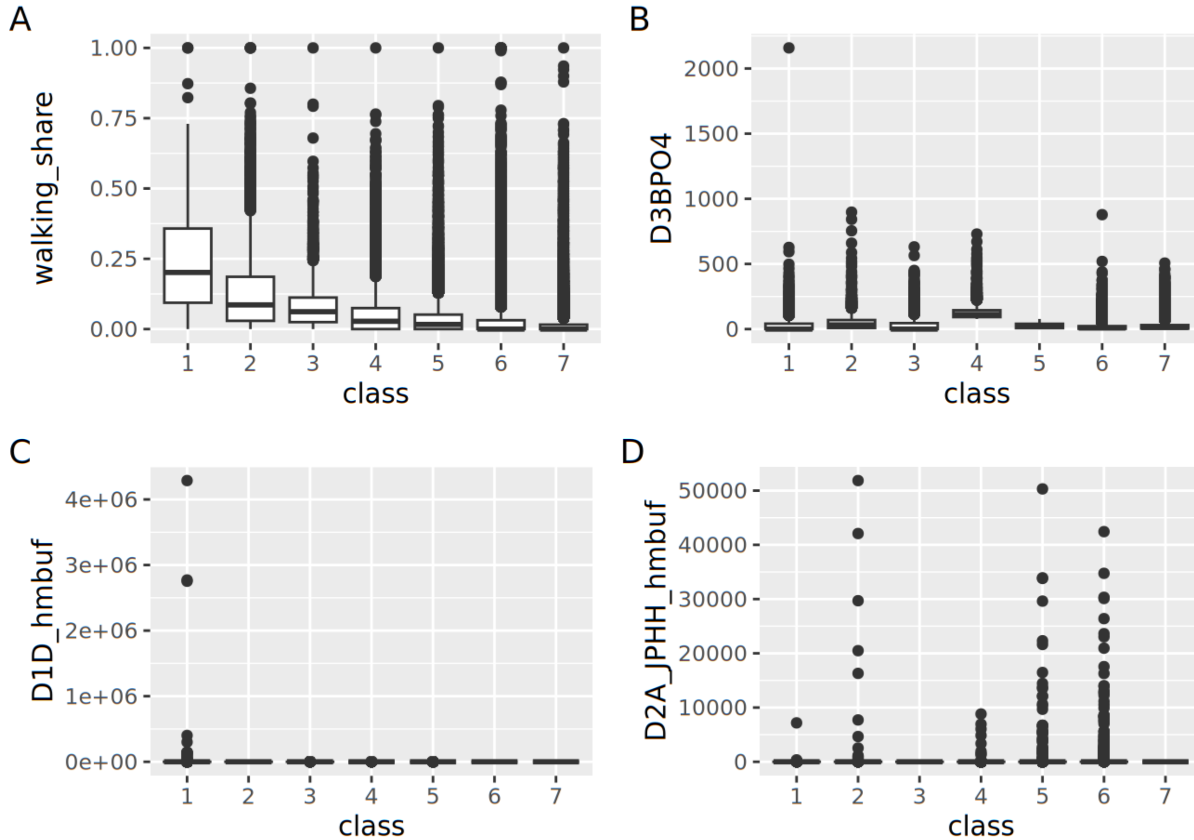

Figure 4.29: Distribution of key variables by urban design quality class classified on CBG walking commuting share

An interactive map (Figure 4.30) showing the spatial pattern of classifying CBGs based on walking commuting share from ACS (Available only in html format):

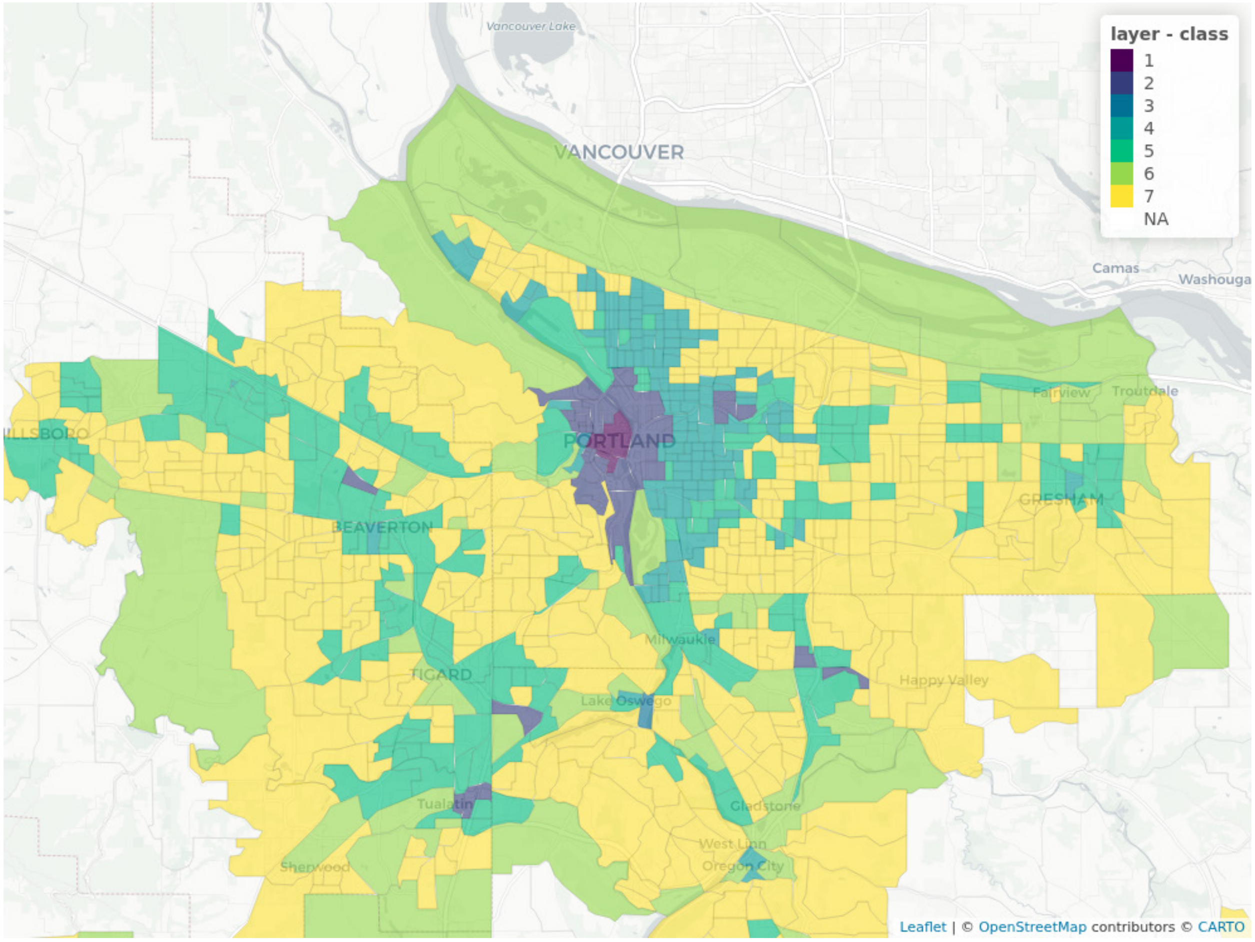

Figure 4.30: Choropleth map of block groups colored by urban design quality class based on CBG walking commuting share

4.5.2.2 D3 Decision Tree trained on any NHTS walking trip

Similar to the decision tree trained on NHTS households making at least one transit trip, here we train a decision tree on NHTS households making at least one walking trip. The resulted dicsion tree is shown below (Figure 4.31).

any NHTS walking trip ~ D1 + D2 + D3* + UZA*

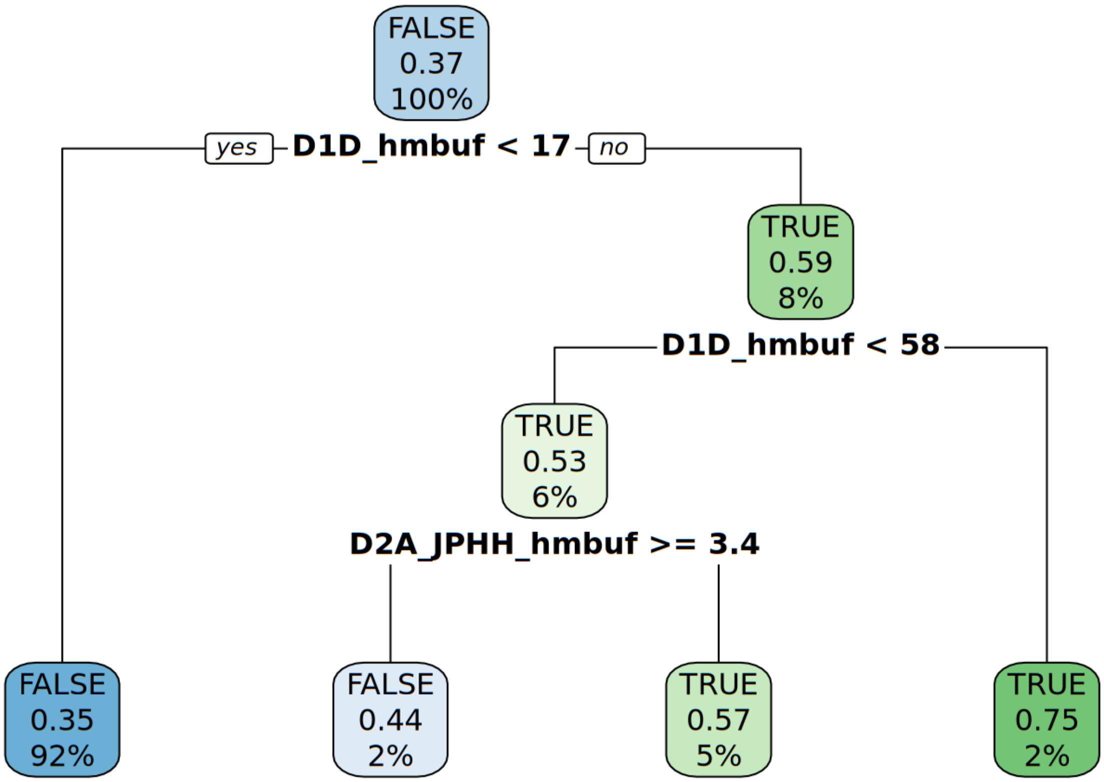

Figure 4.31: Decision tree for classifying block groups based on NHTS household walking trip occurrence

Similarly, we use precision, recall, and f1_score to measure the prediction accuracy of the decision tree on an out-of-sample test set (Table 4.24).

Table 4.24: Prediction accuracy of decision tree for urban design quality classification based on NHTS walking trip occurrence

.metric	.estimator	.estimate
precision	binary	0.612
recall	binary	0.115
f_meas	binary	0.194

As with the transit models, low recall is the most important warning sign here because it means the classifier is missing most of the block groups where walking activity is actually observed. Precision without recall would not be enough for a strong production model.

Summarize key variables by class in a series of box plots (Figure 4.32):

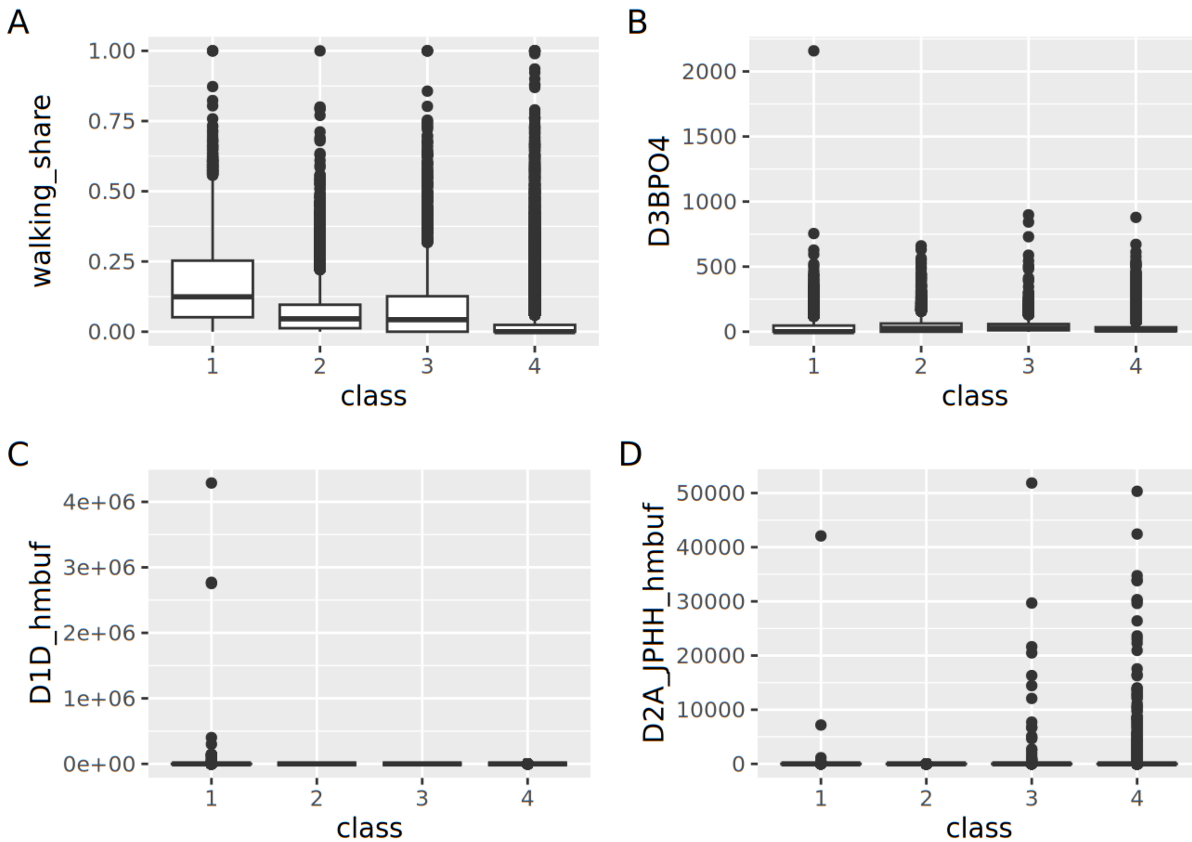

Figure 4.32: Distribution of key variables by class based on NHTS households making any walking trips

An interactive map (Figure 4.33) showing the spatial pattern of classifying CBGs based on NHTS households making any walking trips (Available only in html format):

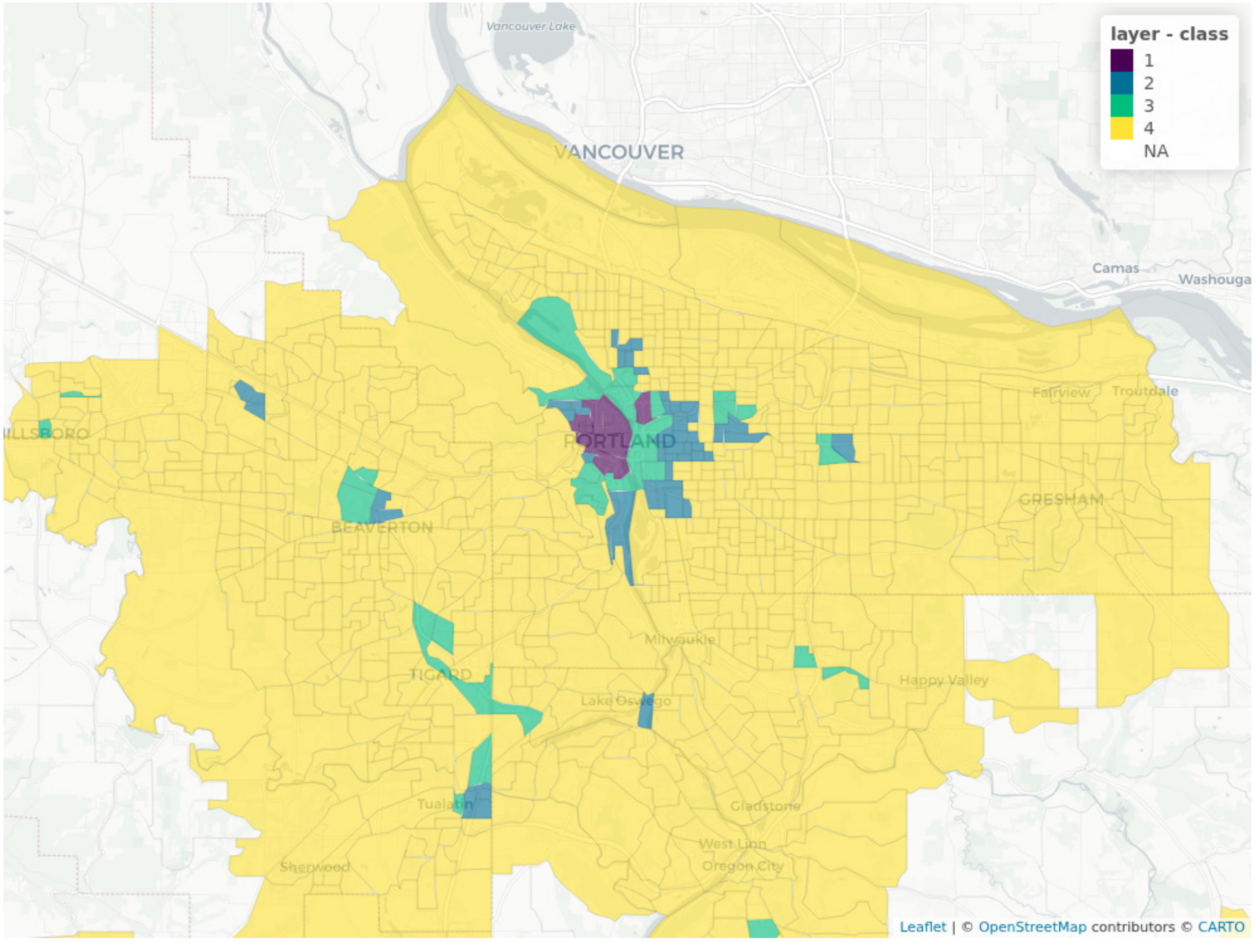

Figure 4.33: Choropleth map of block groups colored by urban design quality class based on NHTS walking trip occurrence

4.5.2.3 Summary of D3 Classification

Table 4.25: Comparison of prediction accuracy across D3 urban design quality classification models

Outcome	rsq	rmse	precision	recall	f_meas
Walking Share (ACS)	0.192	0.059			
Any NHTS Walking Trip			0.612	0.115	0.194

The model performance metrics, gathered together inTable 4.25, show different patterns between the ACS walking share model and the NHTS walking trip model. The ACS walking share model achieves a relatively low R^2 of 0.19 with an RMSE of 0.06, indicating it explains only about 19% of the variance in walking commute mode share. The classification model for any NHTS walking trip shows moderate precision (0.61) but very low recall (0.12), resulting in a poor F1 score of 0.19. This suggests that while the model has some ability to correctly identify areas where walking trips occur (precision), it misses the vast majority of areas with walking activity (recall). The low performance metrics across both models indicate that available variables may not be sufficient to fully capture the factors that influence walking behavior, or that there might be significant noise in both the ACS walking share and NHTS walking trip data.

For interpretation, that means the D3 models are weak to moderate rather than strong. They still provide usable structure for a simplified ordinal D3 classification, but they should not be read as highly accurate behavioral models of walking. This matters for the rest of the report: the D3 classes are useful abstractions for scenario inputs, but the fit statistics suggest caution in over-interpreting small differences between adjacent classes.

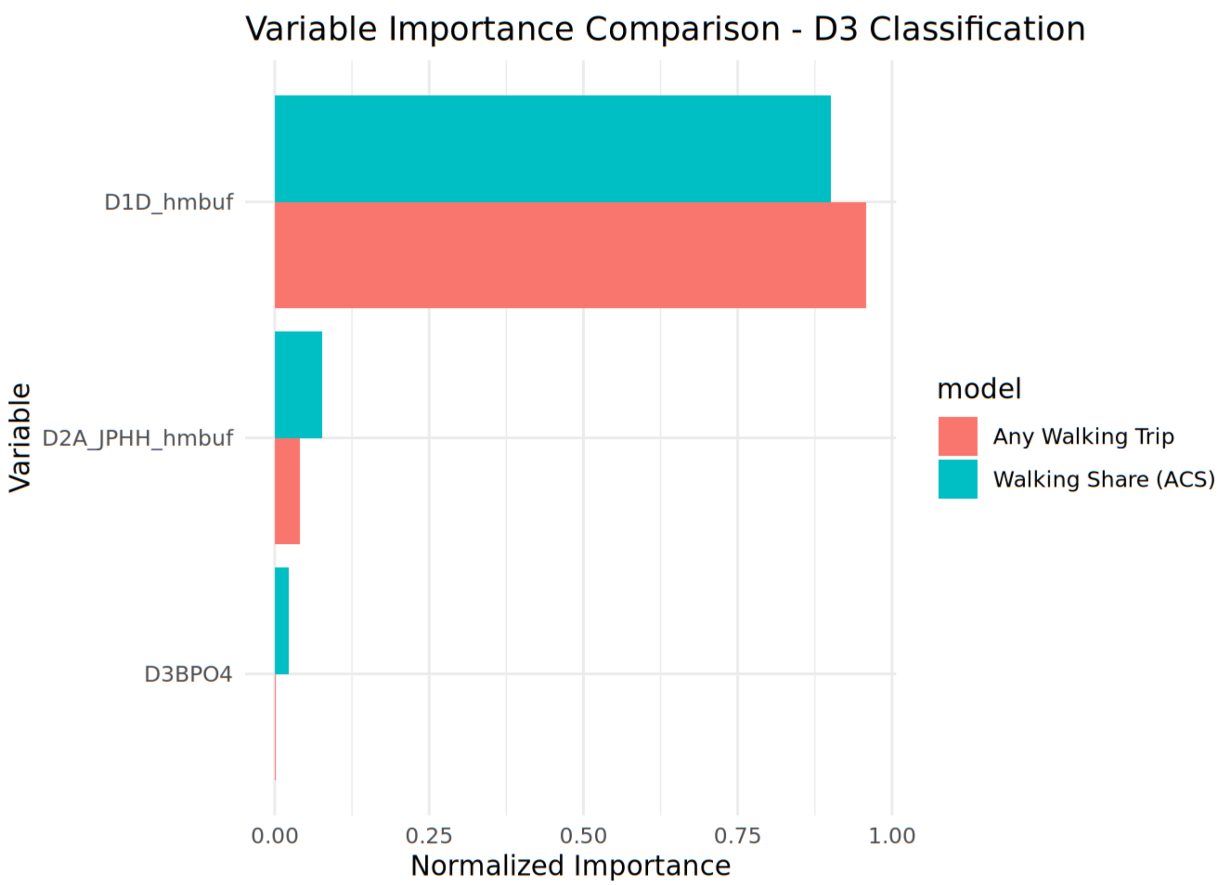

Figure 4.34: Variable importance comparison across D3 urban design quality classification models

The variable importance comparison (Figure 4.34) reveals that both models rely heavily on intersection density measures (D3BPO4 and related variables) and activity density (D1D_hmbuf). The buffered versions of these variables (with _hmbuf suffix) tend to be more important than their non-buffered counterparts, suggesting that the walking environment of the surrounding area matters more than just the characteristics of the immediate block group.

Given these results, we prefer classifying urban design quality using the ACS walking share as the ground truth, as it provides better spatial coverage and more stable estimates compared to the NHTS walking trips, despite its limitation of only capturing commute trips.

4.6 ALLOCATEDU AND ALLOCATEEMPLOYMENT: HOUSING AND EMPLOYMENT ALLOCATION

This section covers the allocation models implemented in AllocateDU and AllocateEmployment. For Use Cases #2 and #3, these functions predict Bzone housing and employment given the Bzone land use type, which is either provided directly by the user in Use Case #2 or predicted by the transition modules in Use Case #3.

In this section, we explore a few different options to model household and employment allocation at the block group level:

1. Direct prediction models: Predict the number of households and jobs in year \(t\) based on year \(t-1\) values and other inputs;

2. Change models: Predict the absolute or percentage changes in households and jobs between year \(t\) and \(t-1\);

We evaluate these approaches based on their prediction accuracy and ability to capture key patterns in the observed data. The models use block group land use type, socioeconomic characteristics, and built environment measures as predictors.

4.6.1 Household Allocation

Figure 4.35 shows the relationship of households in SLD v3 (2017) and v2 (2010) in a few different ways. It is clear that there is a high correlation between the two at 0.94. However, a direct model may not be good at the predicting the changes in households.

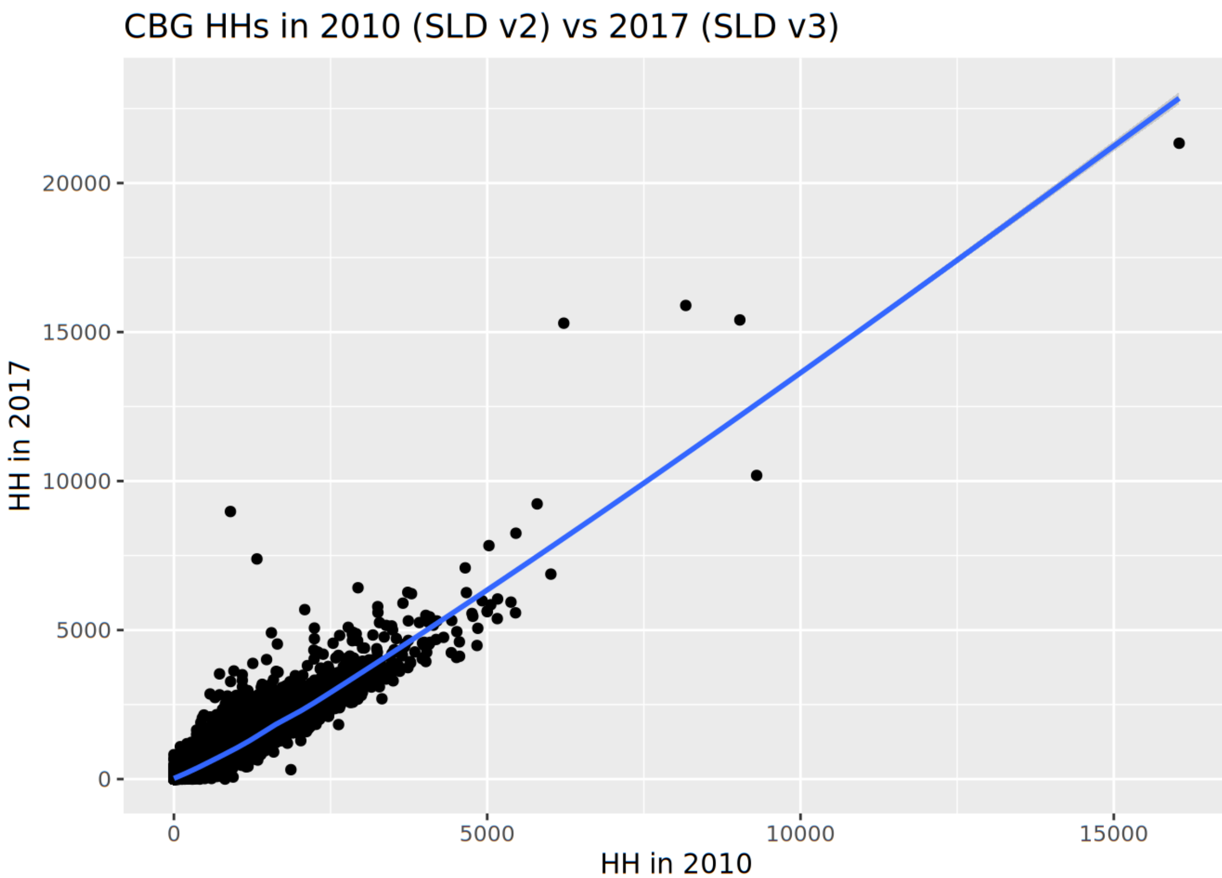

Figure 4.35: Census block group household counts in 2010 (SLD v2) versus 2017 (SLD v3)

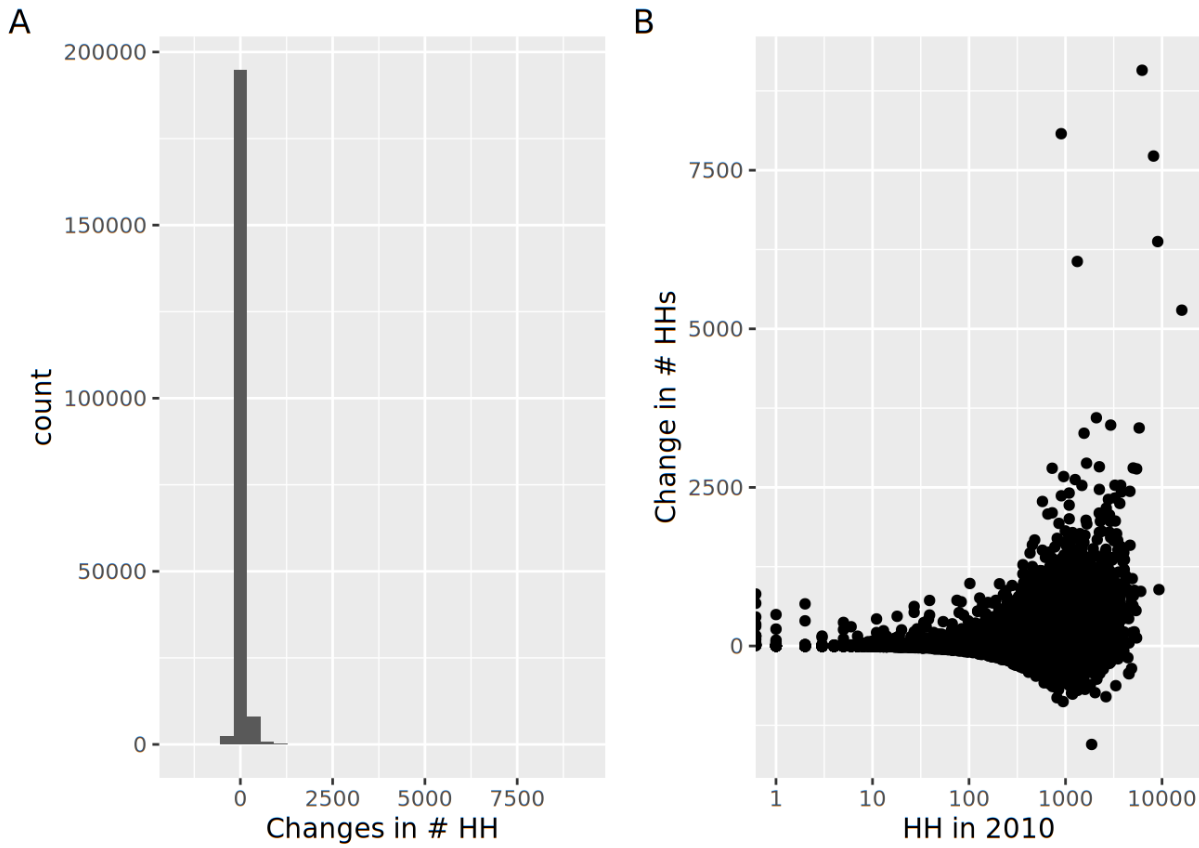

Figure 4.36: Distribution of household count changes by block group from 2010 to 2017: (A) histogram of absolute changes and (B) changes versus 2010 household count

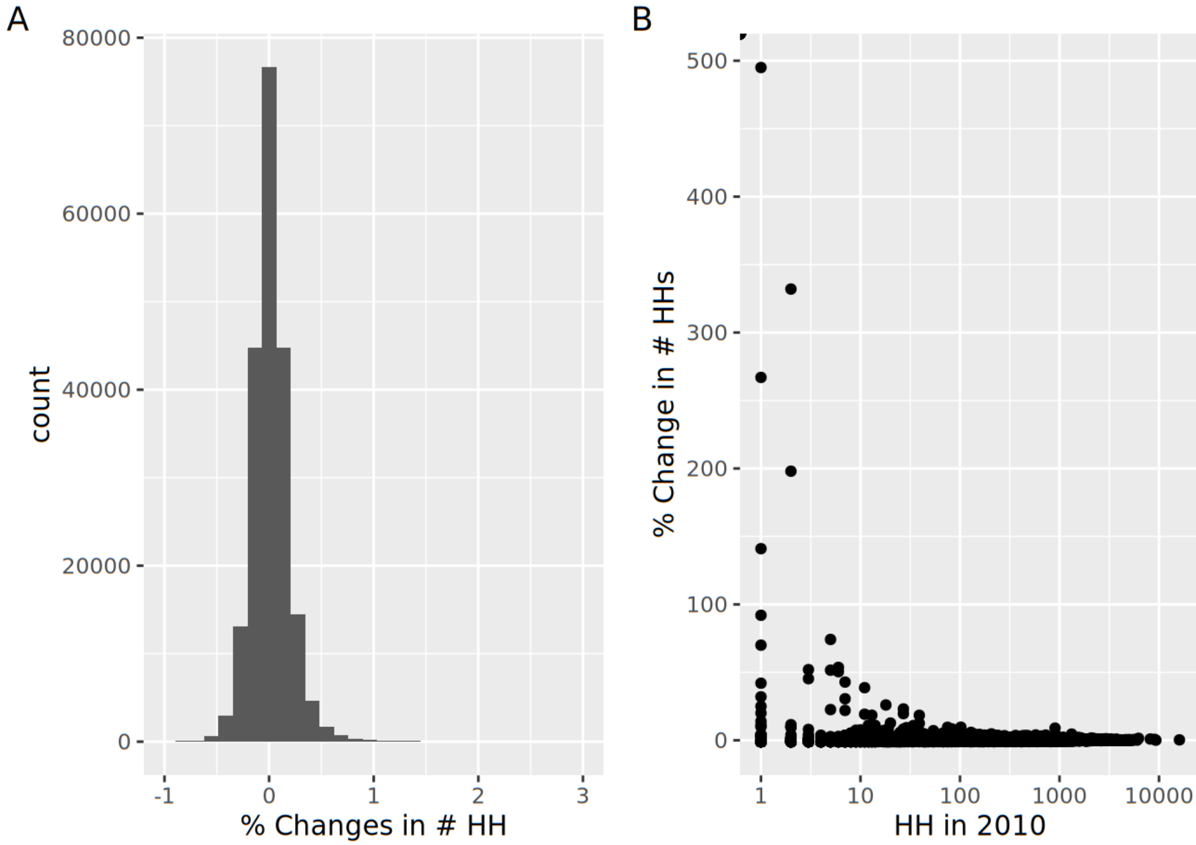

Figure 4.37: Distribution of percentage household changes by block group from 2010 to 2017: (A) histogram of percentage changes and (B) percentage changes versus 2010 household count

4.6.1.1 Models of households

Direct Model of Number of Households

Table 4.26 is the results of an OLS regression model of HH_v3. We include state dummies (variables starting with "STATEFP") to capture the likely different economic and politic environment in each state. Besides land use type, we include percent steep slope (pct_steep_slope), distance to highway ramp (dist_to_ramp) and distance to CBD (dist_to_cbd), and distance to fixed guideway station (dist_to_fgw_sta) variables.

The model performs well on the surface, with a \(R^2\) of 0.89 in out-of-sample validation, but it is not doing well in replicating the difference between HH_v3 and HH_v2, with a \(R^2\) of 0.1. We also try a Poisson regression model, but it perform worse than the OLS model.

The distinction between those two statistics is important. A high out-of-sample R^2 for household levels means the model reproduces the broad cross-sectional ranking of places fairly well. A much lower R^2 for changes means the same model is weak at explaining which specific Bzones gain or lose households over time. In this section, readers should therefore treat the level models as comparatively strong and the change models as comparatively weak.

Table 4.26: Ordinary least squares regression model for predicting 2017 household counts

Characteristic[1]	HH_v3 OLS			HH_v3 GLM		
	Beta[1]	95% CI[1]	p-value[1]	log(IRR)[1]	95% CI[1]	p-value[1]
TotEmp	0.01	0.01, 0.01	<0.001	0.00	0.00, 0.00	<0.001
NumHh	1.1	1.1, 1.1	<0.001	0.00	0.00, 0.00	<0.001
RetEmp	0.00	0.00, 0.01	0.3	0.00	0.00, 0.00	<0.001
SvcEmp	0.00	-0.01, 0.00	<0.001	0.00	0.00, 0.00	<0.001
PctSteepSlope	-0.44	-0.51, -0.36	<0.001	0.00	0.00, 0.00	<0.001
DistToRamp	-0.38	-0.44, -0.32	<0.001	0.00	0.00, 0.00	<0.001
DistToCBD	0.06	-0.04, 0.15	0.2	0.00	0.00, 0.00	<0.001
DistToFgwSta	-0.07	-0.08, -0.06	<0.001	0.00	0.00, 0.00	<0.001
D5	0.00	0.00, 0.00	<0.001	0.00	0.00, 0.00	<0.001
rural * DivType	-17	-29, -5.3	0.005	-0.02	-0.03, -0.02	<0.001
center * emp	-28	-41, -16	<0.001	-0.06	-0.06, -0.05	<0.001
fringe * emp	-11	-24, 0.95	0.071	0.04	0.03, 0.04	<0.001
inner * emp	-28	-39, -16	<0.001	-0.10	-0.11, -0.10	<0.001
outer * emp	-27	-39, -15	<0.001	-0.12	-0.12, -0.11	<0.001
regional center * emp	-16	-30, -1.0	0.036	0.19	0.18, 0.19	<0.001
rural * emp						
center * mix	-30	-43, -17	<0.001	0.04	0.03, 0.04	<0.001
fringe * mix	-3.6	-16, 8.4	0.6	0.10	0.10, 0.10	<0.001
inner * mix	-30	-42, -18	<0.001	-0.12	-0.12, -0.11	<0.001
outer * mix	-25	-37, -14	<0.001	-0.02	-0.02, -0.02	<0.001
regional center * mix	-22	-38, -6.6	0.006	0.01	0.00, 0.02	<0.001
rural * mix						

Characteristic[1]	HH_v3 OLS			HH_v3 GLM		
	Beta[1]	95% CI[1]	p-value[1]	log(IRR)[1]	95% CI[1]	p-value[1]
center * res	-9.8	-22, 2.5	0.12	-0.03	-0.04, -0.03	<0.001
fringe * res	14	2.3, 26	0.020	0.19	0.19, 0.20	<0.001
inner * res	-26	-37, -14	<0.001	-0.16	-0.17, -0.16	<0.001
outer * res	-29	-41, -17	<0.001	-0.07	-0.08, -0.07	<0.001
regional center * res						
rural * res						

[1] Out of Sample Validation: HH_v3 OLS rsq=0.89, HH_delta rsq=0.09; GLM rsq=0.61, HH_delta rsq=0.07

Abbreviations: CI = Confidence Interval, IRR = Incidence Rate Ratio

4.6.2 Housing Units (SFDU, MFDU) Allocation

Figure 4.38 shows the relationship of housing units in SLD v3 (2017) and v2 (2010). It is clear that there is a high correlation between the two.

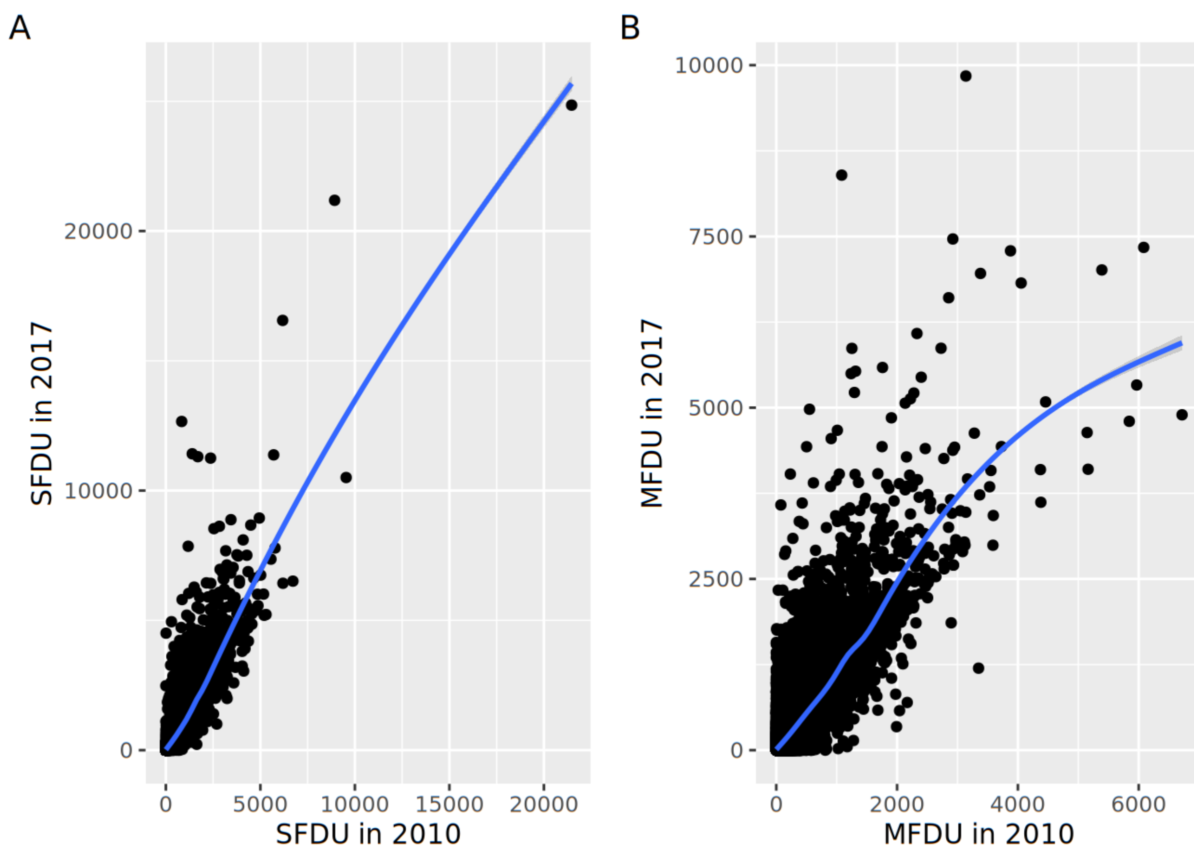

Figure 4.38: Census block group dwelling units in 2010 versus 2017: (A) single-family and (B) multi-family

4.6.2.1 Models of Housing Units

Direct Model of Number of Households

Table 4.27 shows the results of OLS and GLM regression models of SFDU_v3 and MFDU_v3. Similar to the Household Allocation Model, we include state dummies (variables starting with "STATEFP") to capture the likely different economic and politic environment in each state. Besides land use type, we include percent steep slope (PctSteepSlope), distance to highway ramp (DistToRamp) and distance to CBD (DistToCBD), and distance to fixed guideway station (DistToFgwSta) variables.

Table 4.27: Out of sample OLS and GLM model fit R2 for single family and multiple family dwelling units

	OLS	GLM
SFDU counts	0.84	0.23
SFDU changes	0.1	0.06
MFDU counts	0.82	0.38
MFDU changes	0.12	0.04

The OLS models have good \(R^2\)s in out-of-sample validation, but it is not doing well in replicating the changes between dwelling units, with a \(R^2\) around 0.1. The GLM models with Poisson link perform worse than OLS in out-of-sample validation. The OLS models are selected due to its accuracy and simplicity.

The same reading rule applies here: high R^2 for final dwelling-unit totals indicates that the models are useful for recovering broad spatial allocation patterns, while low delta R^2 indicates they are much less reliable for explaining the exact amount of change from one period to the next. That is good enough for the report's allocation purpose, but it is not evidence of a strong short-run change model.

4.6.3 Employment

Figure 4.39 shows the relationship of employment in SLD v3 (2017) and v2 (2010) in a few different ways. Like households, the two have a high correlation at 0.91.

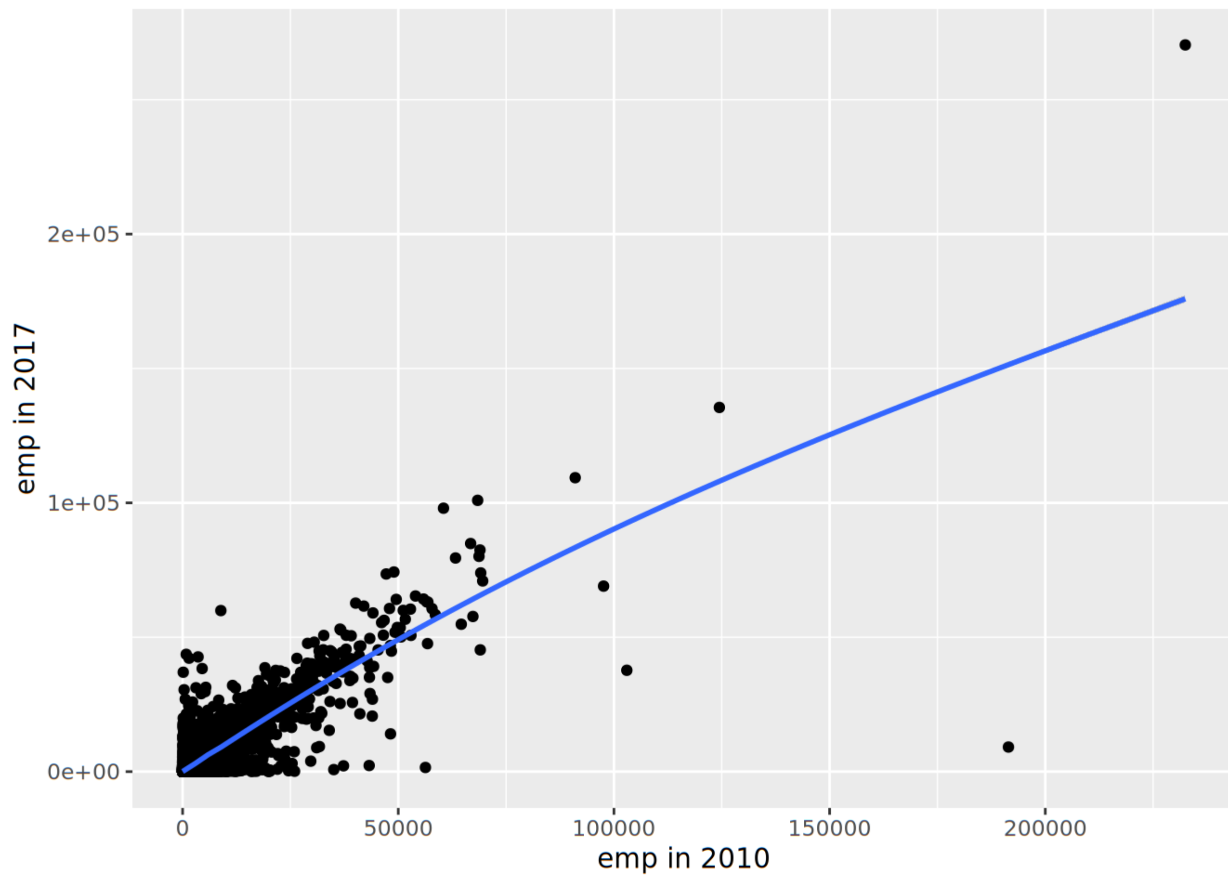

Figure 4.39: Census block group total employment in 2010 (SLD v2) versus 2017 (SLD v3)

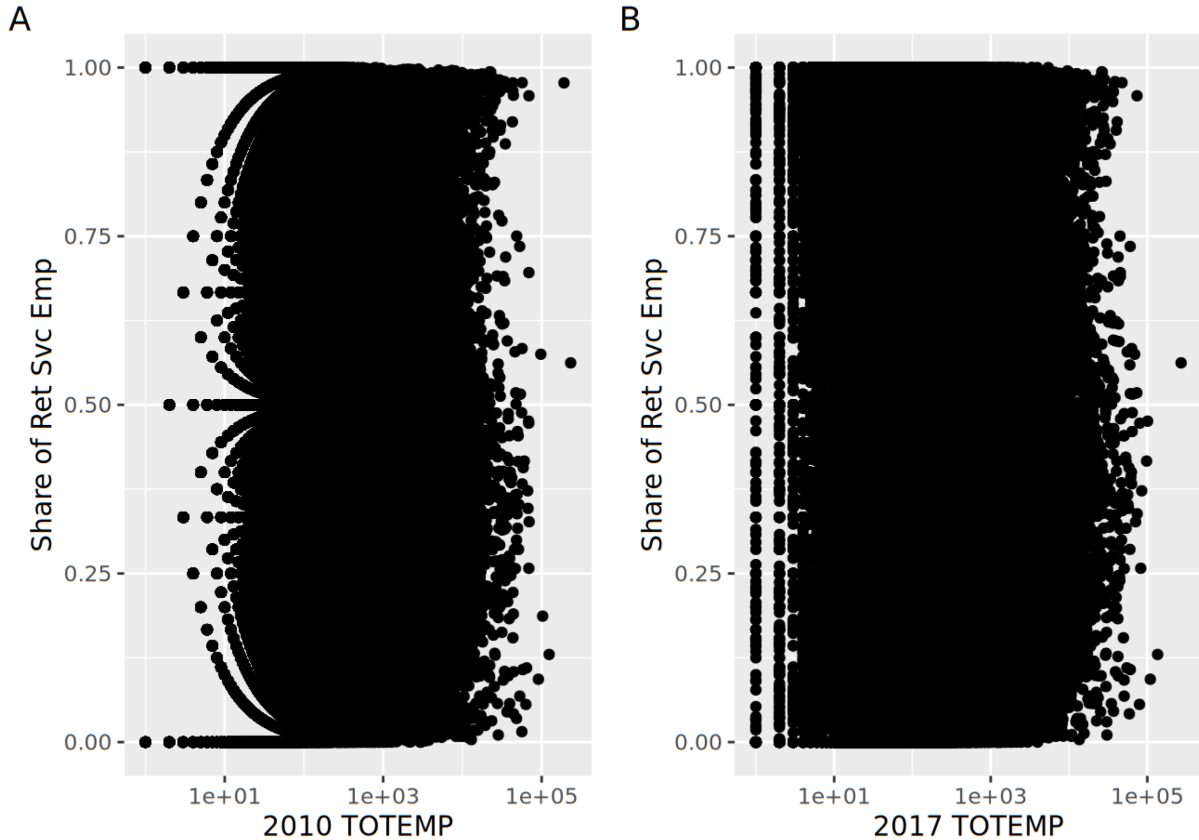

Figure 4.40: Share of retail and service employment by total employment: (A) 2010 and (B) 2017

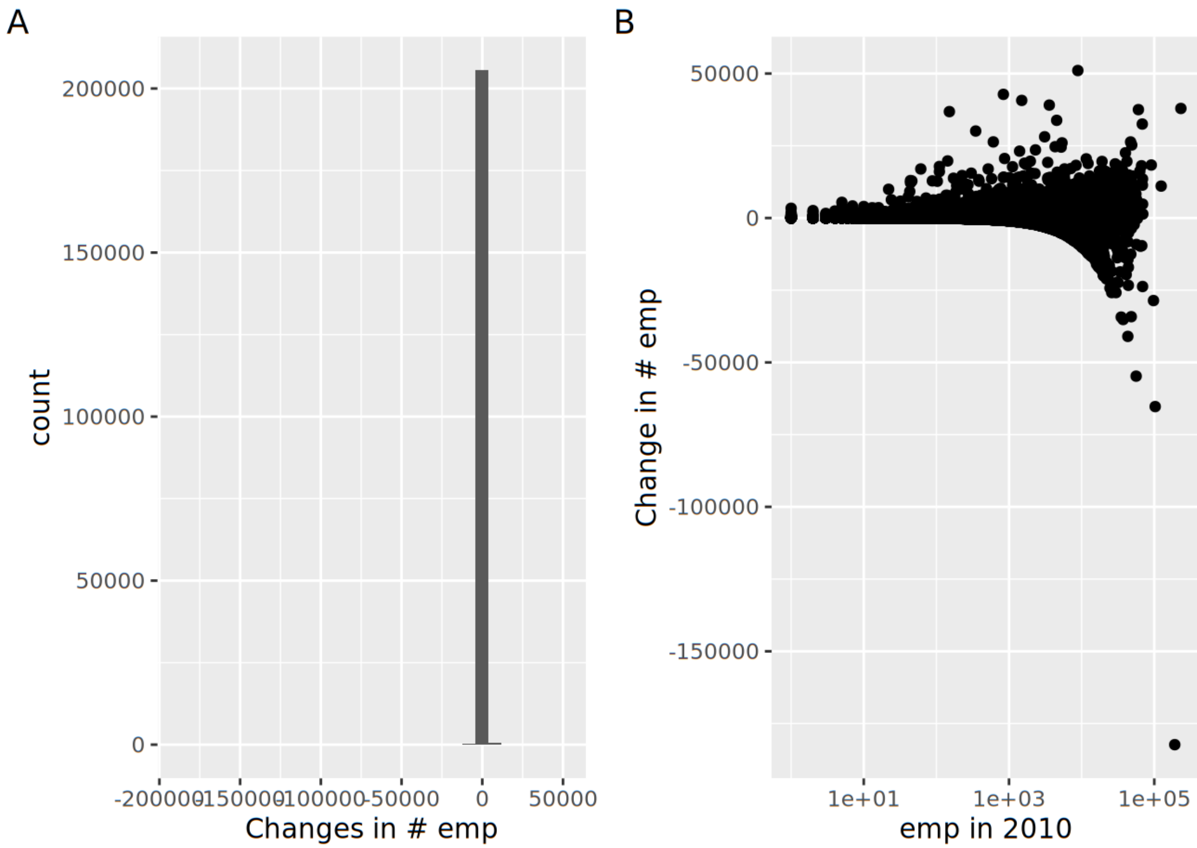

Figure 4.41: Distribution of employment changes by block group from 2010 to 2017: (A) histogram of absolute changes and (B) changes versus 2010 total employment

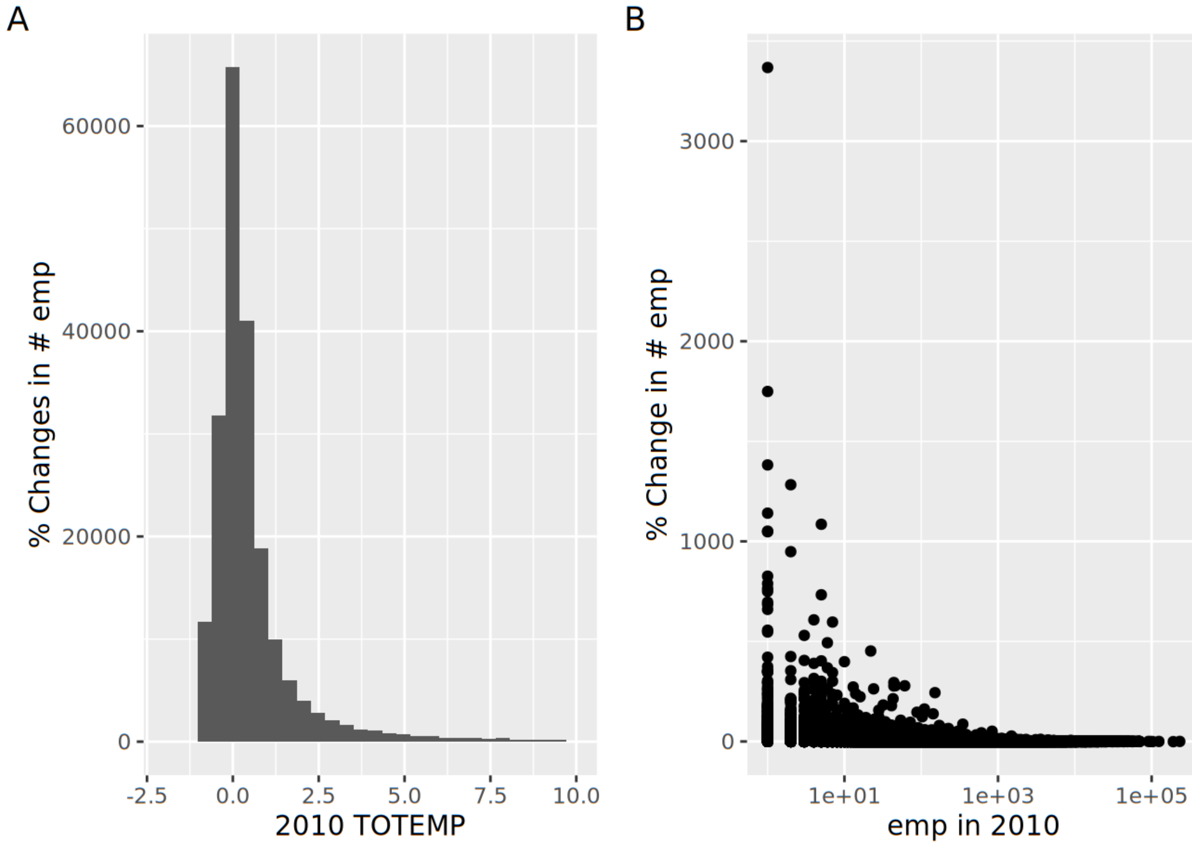

Figure 4.42: Distribution of percentage employment changes by block group from 2010 to 2017: (A) histogram of percentage changes and (B) percentage changes versus 2010 total employment

4.6.3.1 Models of Employment

Below is the results of a Generalized Linear Model (GLM) regression model of employment by sector group. We similarly include state dummies (variables starting with "STATEFP") to capture the likely different economic and politic environment in each state. Besides land use type, we include percent steep slope (PctSteepSlope), distance to highway ramp (DistToRamp) and distance to CBD (DistToCBD) variable.

The OLS models have good \(R^2\)s in out-of-sample validation, but it is not doing well in replicating the changes between employment, with a \(R^2\) lower than 0.1. The GLM with Poisson link perform worse than OLS. The OLS models are selected due to its accuracy and simplicity.

Again, the practical interpretation is that the employment allocation models are better at reproducing broad end-state employment levels than they are at reproducing the exact magnitude of employment change. For this application, that makes them serviceable allocation models, but the low change-fit statistics should be read as a clear limitation rather than as a minor weakness.

Table 4.28: Out of sample OLS and GLM model fit R^2 for retail, service, and other employment

	OLS	GLM
RetEmp	0.78	0.24
SvcEmp	0.42	0.1
NRSEmp	0.78	0.1

5.0 VELANDUSE MODULE IMPLEMENTATION

This chapter documents the VELandUse package implemented for the VisionEval framework (https://github.com/VisionEval-Dev/VisionEval). The new package is currently hosted at https://github.com/cities-lab/VisionEval-Dev and will be submitted to the offical VisionEval repo for inclusion.

5.1 OVERVIEW

The VELandUse module is an R package that enhances land use modeling within the VisionEval framework. It supports land use scenarios with different levels of specificity and simulates their impacts on travel demand. The module aims to provide enhanced capability for VisionEval to simulate how land use policies and developments influence travel behavior and outcomes.

5.1.1 Key Features

- **Integration with VisionEval**: Seamlessly integrates with the VisionEval framework to provide enhanced land use modeling capabilities alongside other existing VisionEval modules.
- **Policy Sensitivity**: Designed to assess the impacts of various land use policies.
- **Multiple Use Case Support**: Supports 3 use cases with different levels of input specificity, as discussed in Chapter 3.
- **Utilizing best data available**: The package uses the best data available for estimating model parameters and providing initial base year inputs, as discussed in Chapter 4.

5.1.2 SubModels

As a VisionEval module, VELandUse provides the following functions:

- AllocateDU: Allocate dwelling units to Bzone based on Land Use Type (predicted or user-specified).
- AllocateEmployment: Allocate employment to Bzone based on Land Use Type (predicted or user-specified).
- AssignD3D4Levels: Predict D3 (Urban Design) and D4 (Transit) levels
- LoadLUType: Loads land use type (AreaType and DivType) from csv file.
- PredictLUType: Simulates changes in land use type (AreaType and DivType).
- PredictLocType: Simulates changes in location type (LocType).

- UpdateLUType: Updates land use type (AreaType and DivType) based on current simulation year attributes.

These functions are in the current VELandUse module and remain unchanged: - AssignCarSvcAvailability - AssignDemandManagement - AssignParkingRestrictions - Calculate4DMeasures - CalculateUrbanMixMeasure - LocateEmployment - PredictHousing

These functions are in the current VELandUse module and will be deprecated or superseded: - AssignDevTypes - CalculateBasePlaceTypes - CalculateFuturePlaceTypes

5.2 USAGE

5.2.1 Installation

At this point, the package can be installed from GitHub using the remotes package:

When the package is submitted to the VisionEval repo, it is expected to be included in the VisionEval installer.

Like any other visioneval module, to use the new VELandUse module, users will invoke the submodels/functions through a visioneval::runModule call:

5.2.2 Practical Overview by Use Case

The detailed module documentation later in this chapter is organized by submodel, but users typically implement VELandUse by use case rather than by individual function. Table 5.1 therefore summarizes the practical workflow for the three use cases introduced in Chapter Chapter 3.

Table 5.1: Practical implementation overview by use case

Use case	Typical situation	User provides directly	Typical VELandUse sequence	Main result
Use Case #1	Detailed base-year or forecast application with explicit Bzone land use inputs	Bzone housing, employment, and land use inputs	Legacy or external land use inputs are loaded, then downstream land use and travel-support modules use those Bzone values	Bzone-level land use inputs are preserved and carried into downstream VE modules
Use Case #2	Scenario planning with land use assumptions specified by type rather than by Bzone	AreaType, DivType, LocType, and LUType-level housing and employment controls	LoadLUType -> PredictLocType or user-supplied LocType -> AllocateDU -> AllocateEmployment -> AssignD3D4Levels and downstream support modules	Bzone housing and employment are allocated from user-specified land use patterns
Use Case #3	Exploratory future scenario where detailed future Bzone land use is not available	Base/current-year Bzone attributes plus scenario-sensitive transport and control inputs	PredictLUType -> PredictLocType -> AllocateDU -> AllocateEmployment -> AssignD3D4Levels and downstream support modules	Future Bzone land use types, location types, housing, employment, and support variables are simulated from current conditions

In practical terms, the main branching point is whether future land use types are supplied by the user or predicted by the model. Use Case #1 keeps explicit Bzone inputs. Use Case #2 supplies land use types but lets VELandUse allocate households and jobs within those constraints. Use Case #3 asks VELandUse to predict the future AreaType, DivType, and LocType first, then uses those predicted classes in the same downstream allocation and support steps.

5.2.3 Required Setup Files

In addition to the normal VisionEval scenario structure, VELandUse requires different input files depending on whether land use is provided directly by Bzone, provided by land use type, or predicted from base/current-year Bzone data. The packaged example models under VELandUse/inst/models/inputs-usecase1, inputs-usecase2, and inputs-usecase3.* provide representative file sets for each workflow.

Table 5.2: Main setup files and new land use inputs by use case

File group	Representative filenames	Use cases	Purpose
Existing direct Bzone controls	bzone_dwelling_units.csv, bzone_employment.csv	Use Case #1; base/current year for Use Case #3	Supply dwelling units and employment directly by Bzone when those values are already known
User-supplied land use types by Bzone	bzone_lutypes.csv or bzone_land_use_types.csv	Use Case #2; often base/current year for Use Case #3	Provide AreaType, DivType, and LocType when land use types are loaded rather than predicted
LUType control totals	lutype_dwelling_units.csv / lut_dwelling_units.csv, lutype_employment.csv / lut_employment.csv	Use Cases #2 and #3	Provide housing and employment totals by land use type for AllocateDU and AllocateEmployment
Base/current-year Bzone attributes for prediction	bzone_distances.csv, bzone_steep_slope.csv, bzone_transit_service.csv	Especially Use Case #3	Supply the explanatory variables needed by PredictLUType, PredictLocType, and AssignD3D4Levels when future land use is predicted
Downstream VE support inputs retained from standard models	bzone_parking.csv, bzone_travel_demand_mgt.csv, bzone_carsvc_availability.csv, bzone_network_design.csv, plus normal Azone/Marea/Region files	All use cases	Continue to support parking, demand management, car service, network, and transport-supply steps after land use has been loaded, allocated, or predicted

Use Case #2 replaces direct future bzone_employment.csv and bzone_dwelling_units.csv controls with LUType-level control files lutype_employment.csv and lutype_dwelling_units.csv along with bzone_lutypes.csv, while Use Case #3 removes the need for bzone_lutypes.csv beyond the base year, and instead relies on the land use transition and location-type models to predict future land use types. In the pilots, those base/current-year Bzone attributes were loaded from the SKATS VisionEval inputs and augmented with data from Census, LEHD, SLD, and transport-support data.

5.2.4 Outputs and Datastore Workflow

Like other VisionEval modules, VELandUse writes its outputs into the datastore. Across all three use cases, VELandUse updates the data table in the VisionEval datastore for each simulation year, typically the Bzone table in results/Datastore/<year>/Bzone/. That means the VELandUse outputs are available to later VELandUse steps and to downstream VE modules. Table 5.3 lists the main land use and support fields referenced below.

Table 5.3: Main VELandUse outputs as stored in the datastore

Output Group	Bzone Fields	Functions	Downstream Use
Land Use Classifications	AreaType, DivType, LocType	LoadLUType, PredictLUType, PredictLocType	Bzone land use and location types used by allocation, and performance metrics
Classification Probabilities	.AreaTypeProbCenter, .AreaTypeProbInner, .AreaTypeProbOuter, .DivTypeProbEmp, .DivTypeProbMix, .LocTypeProbTown, .LocTypeProbUrban	PredictLUType, PredictLocType	Diagnostics and selection of near-threshold Bzones for scenario applications
Housing Allocations	SFDU, MFDU, GQDU	AllocateDU	Used by PredictHousing and AssignParkingRestrictions modules, and performance metrics
Employment Allocations	TotEmp, RetEmp, SvcEmp	AllocateEmployment	Used by LocateEmployment and AssignParkingRestrictions modules, and performance metrics
Design and Transit Levels	D3Lvl, D4Lvl	AssignD3D4Levels	Used by housing and employment allocations, land use type transition

5.2.5 Using VELandUse Outputs as Explicit Bzone Inputs

Besides standard VisionEval workflow, one practical use of VELandUse module for VE-State is to treat a VELandUse run as a scenario-building step and then freeze the resulting target-year

Bzone pattern as explicit input for later model runs. In other words, Use Case #2 or Use Case #3 can be used to convert land-use-type controls, policy assumptions, and transport-sensitive land use responses into the same kind of explicit Bzone inputs that Use Case #1 VE-State workflow expects. The main conversion needed is data exchange: extract the target-year Bzone fields from the datastore, write them back out in the standard VisionEval csv input format, and then reuse those files in new scenario runs.

In practice, the workflow is:

1. Run VELandUse through the target year using Use Case #2 or Use Case #3, depending on whether future land use types are user-specified or predicted.

2. Export the target-year Bzone fields from the datastore: SFDU, MFDU, and GQDU into bzone_dwelling_units.csv; and TotEmp, RetEmp, and SvcEmp into bzone_employment.csv, and AreaType, DivType, and LocType into bzone_lutypes.csv if needed.

3. Import the extracted files along with other VE inputs into the inputs directory for a new VE scenario.

4. Use those exported files as the fixed land use inputs in the new VE scenario. This lets VE-State applications preserve a land use pattern generated by VELandUse, while still using the usual downstream VisionEval workflow and the unchanged VETransportSupply functions.

This export-and-reuse pattern is especially useful when a planner wants to move from an exploratory scenario to a more controlled scenario. For example, a Use Case #3 run can be used to generate a plausible future Bzone land use types under new policy conditions, and that simulated land use types, housing and employment allocations can then become the explicit inputs for other scenarios. If desired, users can also modify the simulated inputs (Land Use Types, housing and employment allocations) for specific Bzones in alternative scenarios.

5.2.6 Model Estimation

All scripts used to estimate model parameters are available in the data-raw folder. Unlike many visioneval modules, the model estimation scripts are not run when the package is built to speed up installation, as some of the estimation scripts take a long time to run. Users can replicate the model parameters and estimate process by running the scripts in the data-raw folder. They can also modify the scripts to customize the model parameters, specification, and/or data sources used for model estimation.

To customize the model parameters or structures, users can follow these steps:

1. **Prepare Data**: Load and process data relevant to land use modeling. The current data processing is managed by the targets data pipeline (Chapter 4).

2. **Customize Model Formula**: Edit the model estimation scripts to adjust the formulae in corresponding model.

3. **Re-estimate Models**: Run the estimation scripts to update model parameters based on new data or formula.

4. **Modify Prediction Specifications**: Ensure that prediction scripts align with the updated model structures and parameters.

5. **Rebuild and Reinstall Package**: Compile the package with the new model specifications and reload it for use in VisionEval.

5.3 CODE REPOSITORY

The source code for the VELandUse package is currently available on GitHub: https://github.com/cities-lab/VisionEval-Dev and will be submitted as a pull request to the offical VisionEval repo for inclusion.

5.4 MODULE DOCUMENTATION

Following visioneval module development guidelines, documentation of the module is embedded in the package and rendered into markdown documents (available in the inst/module_docs folder) when the package is built. The module documentation is provided here for completeness.

5.4.1 AllocateDU Module

This module assigns dwelling unit targets specified by Housing Type (SF, MF, GQ) and Land Use Type to Bzones.

The module carries out the following series of calculations to assign dwelling units by type (SFDU, MFDU, GQDU) to Bzones:

1. Predict Initial Bzone Dwelling Units by Type: For each dwelling unit type (SFDU - Single Family Dwelling Units, MFDU - Multi-Family Dwelling Units, GQDU - Group Quarters Dwelling Units), an Ordinary Least Squares (OLS) regression model is applied. These models use Bzone characteristics (such as accessibility, existing development patterns) and the Bzone’s assigned Land Use Type (LUType) as predictor variables to generate an initial forecast of the number of dwelling units of that type within each Bzone. This initial forecast serves as the seed for the subsequent adjustment step.

2. Allocate Dwelling Units using Iterative Proportional Fitting (IPF): An simplified IPF process is used to adjust the initial Bzone dwelling unit predictions (from Step 1) to match specified control totals. The **seed matrix** for the IPF is the predicted Bzone dwelling units by type (SFDU, MFDU, GQDU) from Step 1. The **margin control totals** are the target dwelling unit numbers by type (SFDU, MFDU, GQDU) for each LUType, read from the lutype_dwelling_units.csv input file. The IPF algorithm iteratively samples from the seed matrix until the sum of dwelling units in Bzones, grouped by their LUType, matches the LUType-level targets for each dwelling unit

type. This ensures the final allocation respects both the spatial patterns suggested by the OLS models and the aggregate targets specified by LUType.

5.4.1.1 Model Parameter Estimation

An OLS model is used to predict Bzone dwelling units for each housing type (SFDU, MFDU, GQDU) based on LUType and other Bzone varialbes. The model is estimated using data from Smart Location Database.

5.4.1.2 How the Module Works

The module carries out the following series of calculations to assign dwelling units by type (SFDU, MFDU, GQDU) to Bzones:

1. Predict Initial Bzone Dwelling Units by Type: For each dwelling unit type (SFDU - Single Family Dwelling Units, MFDU - Multi-Family Dwelling Units, GQDU - Group Quarters Dwelling Units), an Ordinary Least Squares (OLS) regression model is applied. These models use Bzone characteristics (such as accessibility, existing development patterns) and the Bzone's assigned Land Use Type (LUType) as predictor variables to generate an initial forecast of the number of dwelling units of that type within each Bzone. This initial forecast serves as the seed for the subsequent adjustment step.

2. Allocate Dwelling Units using Iterative Proportional Fitting (IPF): An IPF process is used to adjust the initial Bzone dwelling unit predictions (from Step 1) to match specified control totals. The **seed matrix** for the IPF is the predicted Bzone dwelling units by type (SFDU, MFDU, GQDU) from Step 1. The **margin control totals** are the target dwelling unit numbers by type (SFDU, MFDU, GQDU) for each LUType, read from the lutype_dwelling_units.csv input file. The IPF algorithm iteratively scales the seed matrix until the sum of dwelling units in Bzones, grouped by their LUType, matches the LUType-level targets for each dwelling unit type. This ensures the final allocation respects both the spatial patterns suggested by the OLS models and the aggregate targets specified by LUType.

3. Finalize Bzone Dwelling Units: The output of the IPF process is the final allocated number of dwelling units (SFDU, MFDU, GQDU) for each Bzone. These values are then stored in the datastore.

5.4.2 AllocateEmployment Module

This module assigns employment targets (TotEmp, SvcEmp, and RetEmp), specified by Land Use Type, to Bzones.

There are two steps for the allocation: 1. Allocation models predict Bzone employment by sector group based on LUType and other Bzone varialbes 2. An simplified IPF process based on the distribution of predict Bzone employment to meet the specified employment targets (by Azone or LUType);

The IPF process scales employment to Bzone based on the base/previous year employment by land use type and sector group (Svc, Ret, and Non-Svc-Ret) as well as the employment by type and sector group in current simulation year (an input). The former is used as the seed matrix for the IPF, while the latter matches the margin control totals.

5.4.2.1 Model Parameter Estimation

A OLS model is used to predict Bzone employment based on LUType and other Bzone varialbes. The model is estimated using data from Smart Location Database.

5.4.2.2 How the Module Works

The module carries out the following series of calculations to assign employment by sector group (e.g., Retail, Service, Other) to Bzones:

1. Predict Initial Bzone Employment by Sector: For each employment sector group (e.g., Retail - RetEmp, Service - SvcEmp, Non-Retail-Service/Other - NrsEmp), an Ordinary Least Squares (OLS) regression model is applied. These models use Bzone characteristics (such as accessibility, existing development patterns) and the Bzone's assigned Land Use Type (LUType) as predictor variables to generate an initial forecast of the number of jobs within each Bzone for that sector. This initial forecast serves as the seed for the subsequent adjustment step.

2. Allocate Employment using Iterative Proportional Fitting (IPF): An IPF process is used to adjust the initial Bzone employment predictions (from Step 1) to match specified control totals. The **seed matrix** for the IPF is the predicted Bzone employment by sector group (from Step 1). The **margin control totals** are the employment targets by sector group (TotEmp, RetEmp, SvcEmp) by LUType reading from lutype_employment.csv.

   The IPF algorithm iteratively scales the seed matrix until it converges with both the LUType-level and employment sector group targets. This ensures the final allocation respects both the spatial patterns from the OLS models and the aggregate targets.

3. Finalize Bzone Employment: The output of the IPF process is the final allocated number of jobs for each Bzone, broken down by each employment sector group (e.g., TotEmp, RetEmp, SvcEmp). These values are then stored in the datastore.

5.4.3 AssignD3D4Levels Module

This module assigns D3 and D4 levels to Bzones. Section Section 5.5 explains the classification logic in detail, while Appendix Chapter 9 gives the glossary definitions for the underlying variables.

5.4.3.1 Model Parameter Estimation

The module uses two decision tree models to predict D3 (urban design) and D4 (transit) levels:

1. D3 Level Decision Tree (Urban Design):
    - Predicts urban design level based on:
        - D1D_hmbuf (activity density within half mile buffer of Bzone centroid)
        - D2A_JPHH_hmbuf (Job per HH within half mile buffer of Bzone centroid)
        - D3BPO4 (intersection density in terms of pedestrian-oriented intersections having three legs per square mile)
    - Model parameters:
        - minbucket = 1000 (minimum observations in terminal nodes)
        - maxdepth = 3 (maximum tree depth)
        - cp = 0.001 (complexity parameter)
2. D4 Level Decision Tree (Transit):
    - Predicts transit level based on:
        - D4A (distance from the population-weighted centroid to nearest transit stop)
        - D4C (aggregate frequency of transit service within 0.25 miles of CBG boundary per hour during evening peak period)
        - DistToStop (distance to nearest transit stop, miles)
        - DistToFgwSta (distance to nearest fixed guideway station, miles)
        - HasFgwTransit (indicator for whether the Marea has fixed guideway transit service)
    - Model parameters:
        - minbucket = 500 (minimum observations in terminal nodes)
        - maxdepth = 3 (maximum tree depth)
        - cp = 0.01 (complexity parameter)

5.4.3.2 How the Module Works

The module carries out the following steps to assign D3 (urban design) and D4 (transit) levels to Bzones:

1. Predict D3 Level (Urban Design): A decision tree model is used to predict the urban design level (D3) for each Bzone based on:
    - D1D_hmbuf (activity density within half mile buffer of Bzone centroid)
    - D2A_JPHH_hmbuf (Job per HH within half mile buffer of Bzone centroid)
    - D3BPO4 (intersection density in terms of pedestrian-oriented intersections having three legs per square mile) The model predicts one of four levels (1-4), where higher levels indicate better urban design supporting walkability.
2. Predict D4 Level (Transit): A separate decision tree model is used to predict the transit service level (D4) for each Bzone based on:
    - D4A (distance from the population-weighted centroid to nearest transit stop)
    - D4C (aggregate frequency of transit service within 0.25 miles of CBG boundary per hour during evening peak period)
    - DistToStop (distance to nearest transit stop, miles)
    - DistToFgwSta (distance to nearest fixed guideway station, miles)
    - HasFgwTransit (indicator for whether the Marea has fixed guideway transit service) The model predicts one of seven levels (1-7), where higher levels indicate better transit service quality.
3. Finalize Bzone Levels: The predicted D3 and D4 levels are stored in the datastore for each Bzone. These levels are used by other modules to model travel behavior and accessibility.

The decision tree models were trained using data from the Smart Location Database (SLD) and transit and walking commuting shares from the American Community Survey (ACS) data.

For implementation, AssignD3D4Levels can be used to prepare the base year D3Lvl and D4Lvl inputs for Bzones. For simulation years, the module can assign D3Lvl and D4Lvl if users provides the underlying inputs for them. Table 5.4 below summarizes where the module expects each input to come from. Alternatively, users can create scenarios by modifying base year D3Lvl and D4Lvl for selected Bzones, for example, upgrading D4Lvl for Bzones near stations as part of a transit-oriented development scenario.

Table 5.4: Additional transport-related and prerequisite inputs used by AssignD3D4Levels

Input field(s)	Source	Why needed
D3bpo4	bzone_network_design.csv	Direct network-design input used by the D3 decision tree
D4A, D4C	bzone_transit_service.csv	Direct transit access and service inputs used by the D4 decision tree
DistToStop, DistToFgwSta	bzone_distances.csv	Distance-based access measures used by the D4 decision tree; DistToFgwSta also appears in other land use transition logic
HasFgwTransit	marea_fgw_transit.csv	Metro-area fixed-guideway indicator used by the D4 decision tree
D1D_hmbuf, D2A_JPHH_hmbuf	Current-year Bzone datastore fields, typically refreshed through Calculate4DMeasures after housing and employment are updated	Buffered density and jobs-housing measures used by the D3 decision tree

5.4.4 PredictLUTypes Module

This module predicts future year land use types (AreaType and DivType) for Bzones. It is designed for scenarios where future Bzone-level land use types are not explicitly provided by the user (use case #3).

The land use type for a Bzone is composed of two main components predicted by this module: 1. **AreaType**: Categorizes Bzones: urban center, inner, outer, or fringe based on their development intensity and D5 characteristics. 2. **DivType**: Categorizes Bzones by their mix of activities: residential, mixed-use, or employment-focused.

The module uses predictive models to determine the most probable AreaType and DivType for each Bzone in the target year, based on various Bzone attributes from the current/base year.

5.4.4.1 Model Parameter Estimation

Two separate models are used for predicting land use type components:

1. **AreaType Transition Model**: This model predicts the future AreaType of a Bzone. Details of this model's estimation can be found in the data-raw/AreaTypeTransitionModel.R script. <txt:AreaTypeTransitionModel$Summary>

2. **DivType Transition Model**: This model predicts the future DivType of a Bzone. Details of this model's estimation can be found in the data-raw/DivTypeTransitionModel.R script. <txt:DivTypeTransitionModel$Summary>

These models are typically estimated using historical data and Bzone characteristics to learn patterns of land use change.

5.4.4.2 How the Module Works

The module performs the following steps to predict future land use types for each Bzone:

1. **Load Bzone Data**: The module retrieves current/base year data for all Bzones. This includes existing AreaType, DivType, population, employment, household numbers, density measures (D1D_hmcbuf, D5), and geographic attributes (PctSteepSlope, DistToRamp, DistToCBD, DistToFgwSta).
2. **Load Prediction Models**: The pre-estimated AreaTypeTransitionModel and DivTypeTransitionModel are loaded from the package datasets.
3. **Predict Future AreaType**: For each Bzone, the AreaTypeTransitionModel is applied using the Bzone's characteristics as input variables. This yields a predicted AreaType for the future year.
4. **Predict Future DivType**: Similarly, for each Bzone, the DivTypeTransitionModel is applied using the Bzone's characteristics as input variables. This yields a predicted DivType for the future year.
5. **Store Predicted LUTypes**: The predicted AreaTypes and DivTypes for each Bzone are then stored in the datastore for the target year.

5.4.5 PredictLocTypes Module

This module predicts land use types for bzones for a future year. Land use types comprises of a combination of: - four area types: urban center, inner, outer, and fringe - three diversity types: residential, mixed-use, and employment - three location types: urban, town (urban cluster), and rural.

This module supports land use scenario use case #3, in which users doesn't specify future year land use type for each bzone and relies on this module to predict land use types for each bzone.

5.4.5.1 Model Parameter Estimation

The location type transition model is used to predict the probability of a Bzone transitioning to different location types (Urban, Town, Rural) based on Bzone characteristics.

The model is estimated using data from Smart Location Database and other sources.

5.4.5.2 How the Module Works

The module carries out the following series of calculations to predict location types for Bzones:

1. Predict Location Type Probabilities: For each Bzone, the module uses a location type transition model to predict the probability of the Bzone being classified as Urban, Town, or Rural.

2. Convert Probabilities to Location Types: The probabilities are converted to location type assignments using either:

- Maximum probability selection (if MaxProb = TRUE)

- Random sampling based on probabilities (if MaxProb = FALSE)

3. Enforce Transition Logic: The module enforces a hierarchical transition logic where:

- Transitions can only occur to the same level or a higher level

- The hierarchy is: Rural -> Town -> Urban

- If a predicted transition violates this logic, the original location type is retained

4. Store Results: The final location type assignments and their associated probabilities are stored in the datastore.

6.0 PILOT APPLICATIONS

The project team conducted a pilot study and tested land use scenarios in the Salem-Keizer Area Transportation Study (SKATS) region. The team created several hypothetical scenarios and ran them with the new VELandUse module along with existing VE modules. The outcomes were studied for sensitivity and reasonableness. This chapter documents the process and results of the pilot study.

6.1 PILOT APPLICATION SELECTION PROCESS

The project team conducted a thorough evaluation to select appropriate pilot applications for the VELandUse module. After considering factors including data availability, model tractability, and alignment with project objectives, the Salem-Keizer Area Transportation Study (SKATS) region was identified as the area for the pilot study.

6.1.1 Selected Pilot Region

The SKATS metropolitan area was chosen for testing Use Cases #1 and #2, while the surrounding four-county area was selected for the larger-area Use Case #3 pilot. This selection offers several advantages:

- Comprehensive data availability with a 2021 base year and 2050 end year;
- Existing documentation and model infrastructure;
- Opportunity to isolate and evaluate the VELandUse module's performance.

6.1.2 Testing Approach

The pilot application integrates the VELandUse module with the VETransportSupply module for transit service assignment, deliberately excluding other VE modules to better isolate and assess the new module's behavior.

To test model sensitivities, the following hypothetical scenarios are created and tested:

1. **Transit Infrastructure Scenarios**: Testing fixed guideway transit along key corridors including Commercial Street through downtown Salem and River Road into Keizer.

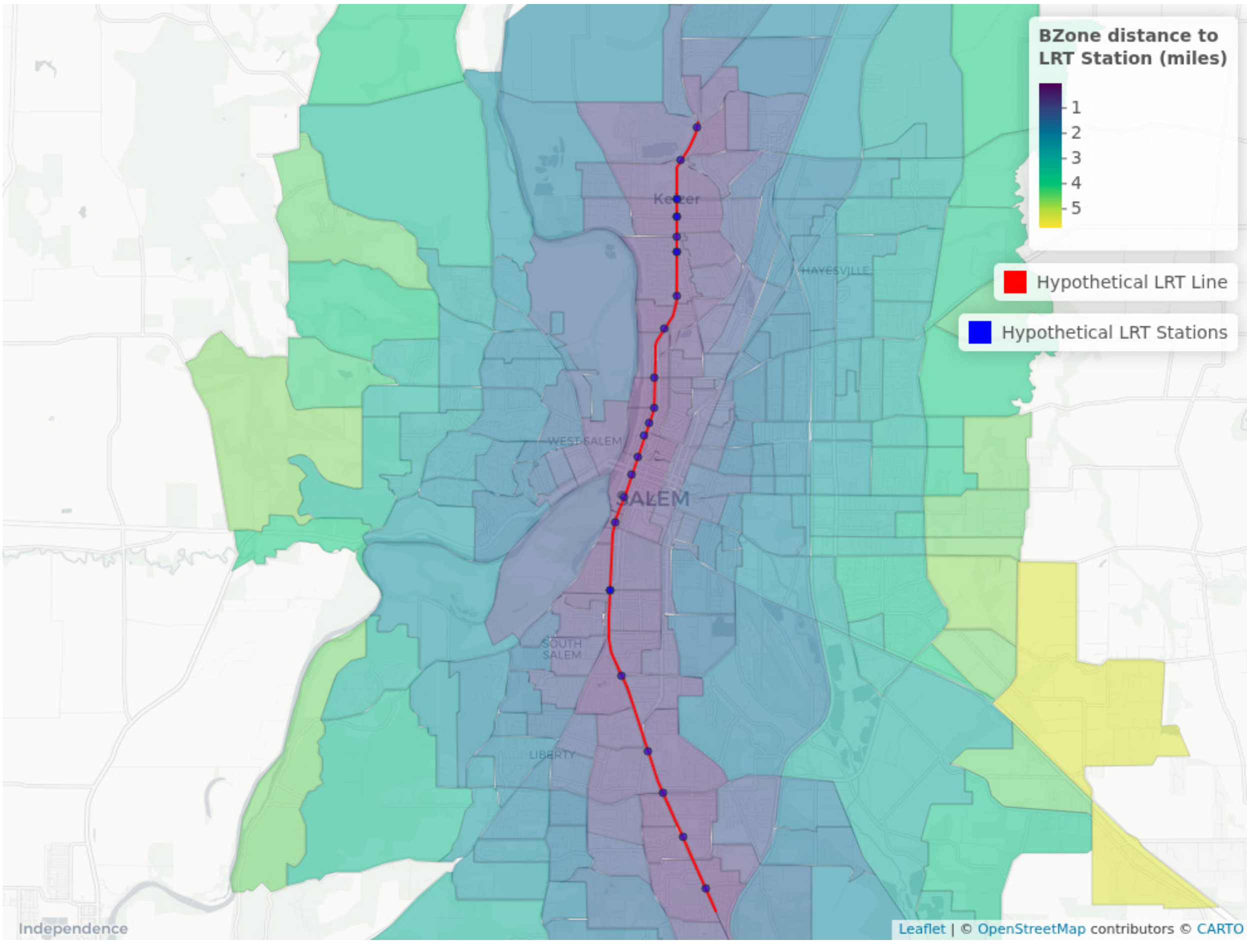

Figure 6.1: Interactive map of SKATS Bzone boundaries colored by distance to hypothetical light rail transit stations, with the line running north-south through downtown Salem and into Keizer

The rendered corridor map makes the intended geography explicit: the hypothetical LRT line follows the Commercial Street and River Road spine through central Salem, crosses the downtown area, and continues north toward Keizer. This is the main location to watch in the later employment and dwelling-unit difference maps.

2. **Highway Infrastructure Scenarios**: Comparing outcomes with and without the Kuebler interchange

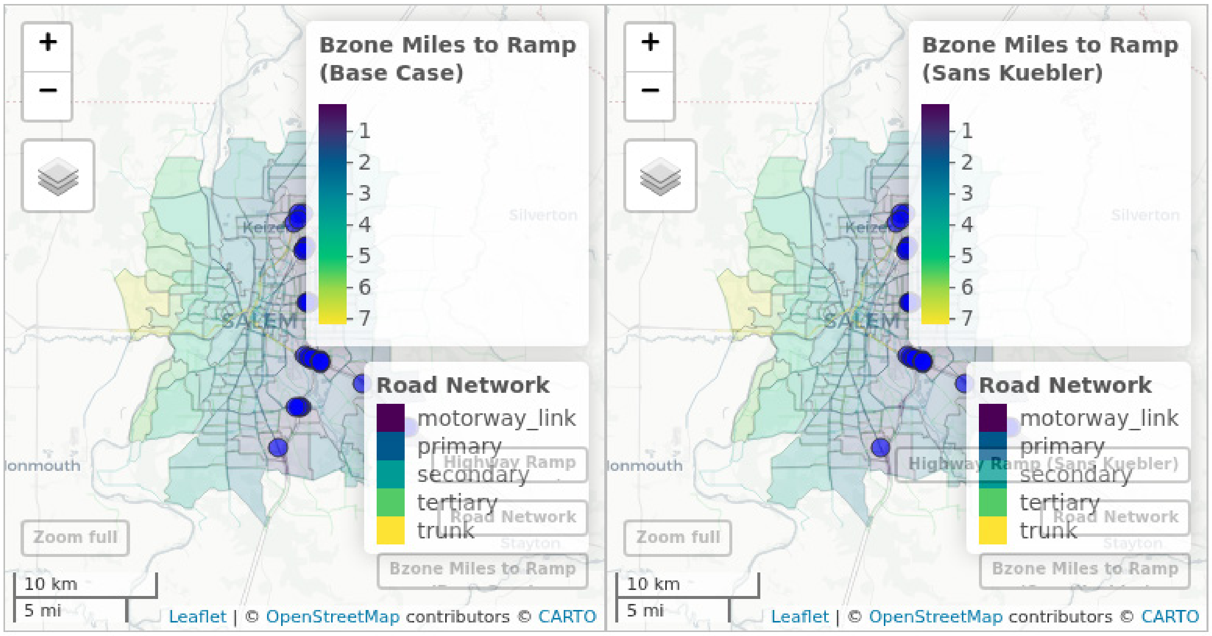

Figure 6.2: Side-by-side interactive maps comparing Bzone distances to highway ramps with and without the Kuebler interchange, centered on the southeast Salem I-5 access change

The Kuebler scenario shows a localized accessibility change near the southeast Salem Kuebler Boulevard and I-5 area. That localized geography is important when interpreting the later Kuebler scenario maps, because the most relevant shifts should appear near that corridor rather than across the entire metro area.

3. **Four-County Regional Analysis**: Implementing a VEState-like application with real BZone for broader regional assessment

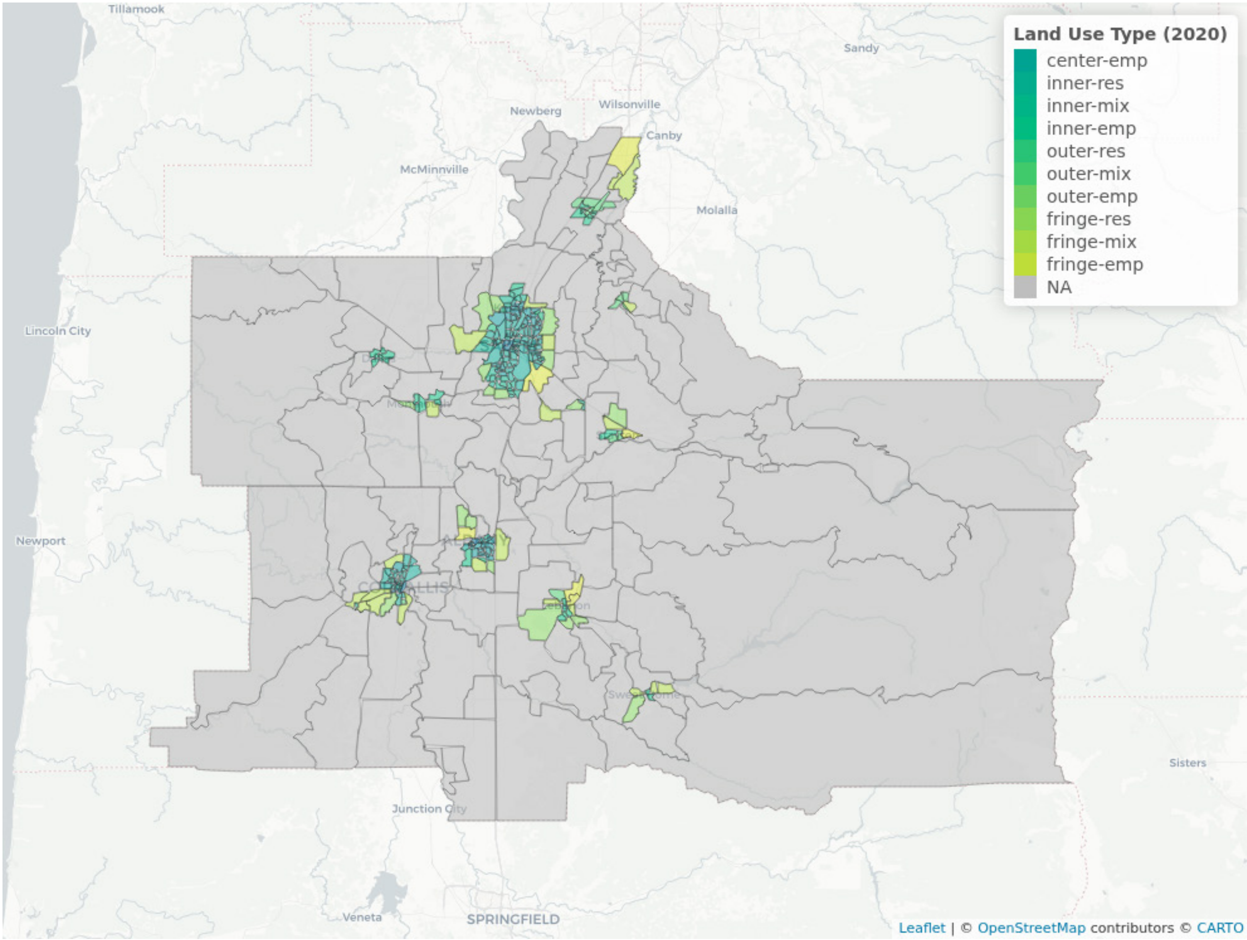

Figure 6.3: Interactive map of Census block groups in the four-county area (Marion, Polk, Benton, Linn) colored by 2020 land use type classification

The four-county base map shows that the real-Bzone pilot is dominated by the Salem-Keizer cluster, with smaller urban clusters around Corvallis, Albany, and Lebanon. Those locations provide the main geographic reference points for interpreting the later regional transition maps.

6.2 PILOT APPLICATION RESULTS

We ran the two sets of pilot applications (SKATS and four-county area) separately and compared the results with the respective existing VE modules as a test of the new VELandUse module. The results are shown below.

Table 6.1 summarizes the pilot scenarios before the detailed maps and module lists. This is useful because Chapter 6 combines both within-region scenario tests in SKATS and a larger-scale VE-State-style test in the four-county area.

Table 6.1: Pilot scenarios and corresponding use cases

Pilot scenario	Geography and purpose	Corresponding use case	Main distinguishing feature
SKATS base case	Salem-Keizer RSPM comparison case used as the local benchmark	Use Case #1	Existing workflow with direct Bzone housing and employment inputs
SKATS hypothetical LRT scenario	Salem-Keizer sensitivity test for new fixed-guideway transit along the Commercial Street and River Road corridor	Use Case #2	User-specified land use types and LUType control totals replace direct future Bzone housing and employment inputs
SKATS Kuebler interchange scenario	Salem-Keizer sensitivity test for a highway-access change near the Kuebler interchange	Use Case #2	Same land-use-type-based workflow as the LRT case, but with a different transport intervention
Four-county VE-State base case	Legacy large-area comparison across Marion, Polk, Benton, and Linn counties	Existing VE-State comparison case	Existing synthesized-Bzone workflow used as the reference point for the new real-Bzone pilot
Four-county real-Bzone scenario	Large-area pilot showing the new package on real Census Block Group Bzones	Use Case #3	Future AreaType, DivType, and LocType are predicted on real Bzones before housing and employment are allocated

6.2.1 Salem-Keizer Area Transportation Study (SKATS)

The SKATS pilot is based on the VESKATSRSPM model, which was developed by RSG for the Salem-Keizer Area Transportation Study (SKATS) area. The base case (Use Case #1) uses the existing VELandUse modules, along with VESimHouseholds and VETransportSupply modules, while the two alternative scenarios (Use Case #2) use the new VELandUse module, again along with VESimHouseholds and VETransportSupply modules.

In other words, the SKATS pilot is a local three-scenario comparison: one benchmark run using direct Bzone land use inputs (Use Case #1), one transit-oriented sensitivity test (Use Case #2), and one highway-oriented sensitivity test (Use Case #2). The purpose is not to forecast a single preferred future, but to check whether the new land-use-type-based workflow produces plausible spatial responses under clearly defined infrastructure changes.

6.2.1.1 Model Steps of Use Case #1: (Existing Modules)

1. VESimHouseholdsSKATS
    - CreateHouseholds
    - PredictWorkers
    - AssignLifeCycle
    - PredictIncome
2. VELandUse (current)
    - PredictHousing (VELandUseDLSKATS)
    - LocateEmployment
    - AssignLocTypes
    - Calculate4DMeasures
    - CalculateUrbanMixMeasure
    - AssignCarSvcAvailability (VELandUseDLSKATS)
    - AssignParkingRestrictions
    - AssignDemandManagement
3. VETransportSupply
    - AssignTransitService → AssignRoadMiles

6.2.1.2 Model Steps of Use Case #2: (New Modules)

1. VESimHouseholdsSKATS
    - CreateHouseholds
    - PredictWorkers
    - AssignLifeCycle
    - PredictIncome
2. VELandUse (new)

- LoadLUTypes
- AllocateDU
- AllocateEmployment
- PredictHousing (VELandUseDLSKATS)
- LocateEmployment
- AssignLocTypes
- Calculate4DMeasures
- CalculateUrbanMixMeasure
- AssignCarSvcAvailability (VELandUseDLSKATS)
- AssignParkingRestrictions
- AssignDemandManagement

3. VETransportSupply
    - AssignTransitService
    - AssignRoadMiles

6.2.1.3 Base Case (Use Case #1)

Figure 6.4 shows the 2050 employment by Bzone for the base case. Note that the Bzone employment (TotEmp, RetEmp and SvcEmp) for each simulation year are specified in the bzone_employment.csv file as part of the inputs.

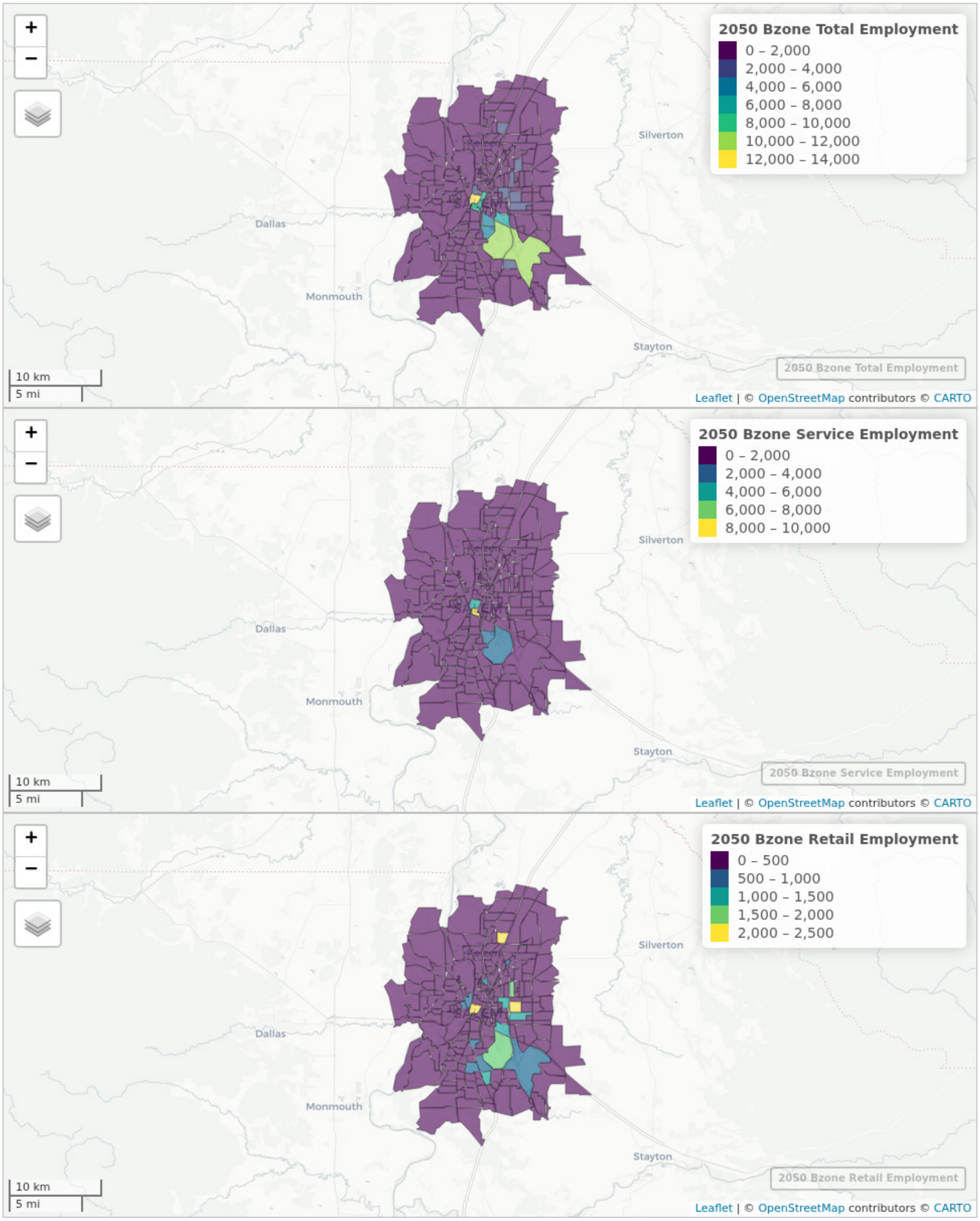

Figure 6.4: Synchronized interactive maps showing 2050 total, service, and retail employment by Bzone for the SKATS base case scenario

Similarly, Figure 6.5 shows the 2050 dwelling unit (SFDU, MFDU and GQDU) by Bzone for the base case. Note that, like employment, the dwelling unit (SFDU, MFDU and GQDU) by Bzone are specified in the bzone_dwelling_units.csv file as part of the inputs. GQDU is not shown in the map because it is not specified (all zeros) in the inputs for the SKATS VERSPM project.

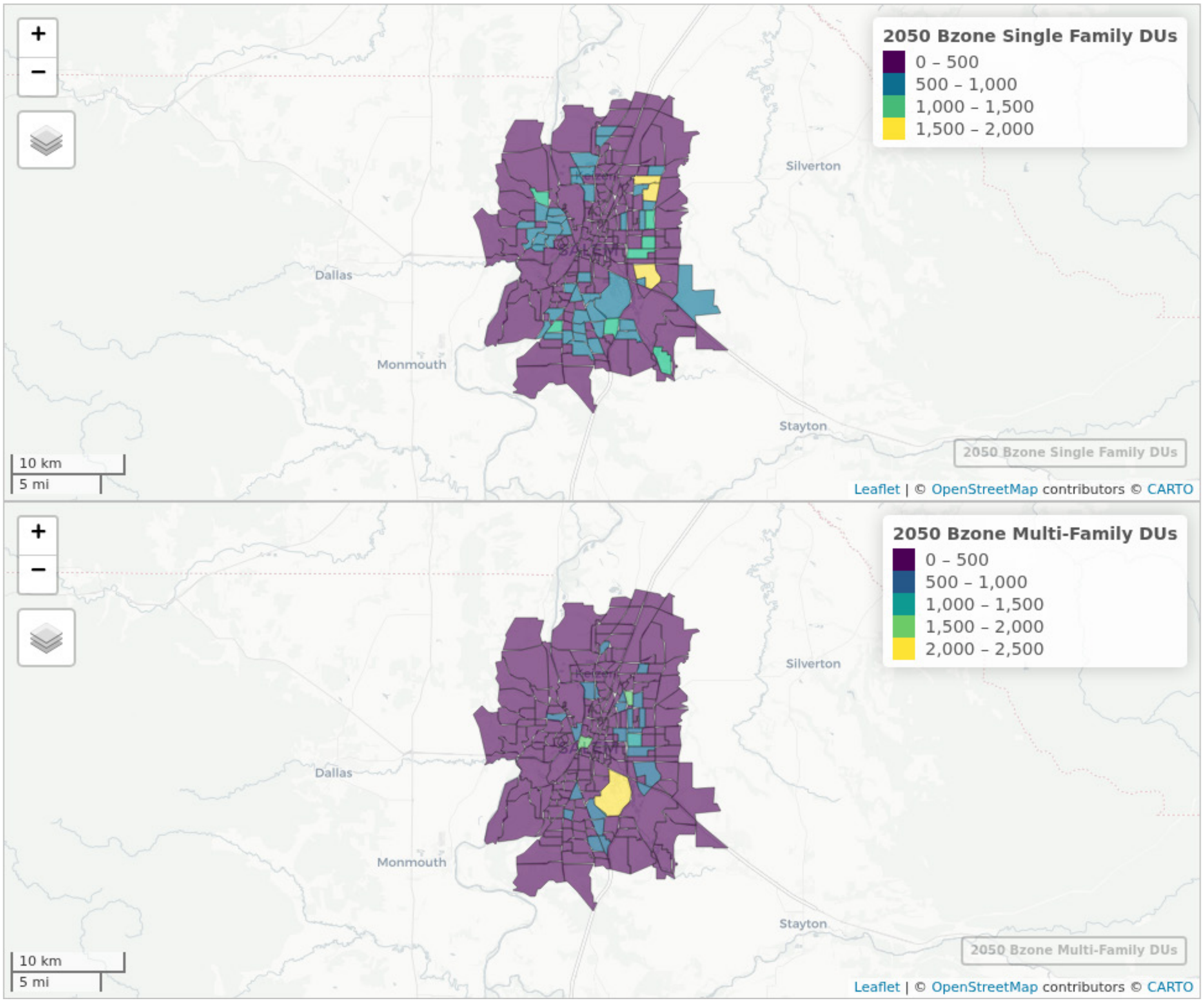

Figure 6.5: Synchronized interactive maps showing 2050 single-family and multi-family dwelling units by Bzone for the SKATS base case scenario

6.2.1.4 Hypothetical LRT Scenarios

Instead of specifying the employment and DUs by Bzone in bzone_employment.csv and bzone_dwelling_units.csv as in use case #1, we specify employment and dwelling units by land use type (Area Type and Diversity Type) in lutype_employment.csv and lutype_dwelling_units.csv, which are then allocated to bzones in AllocateEmployment and AllocateDU submodules.

Figure 6.6 shows the difference in 2050 employment by Bzone between the LRT scenario and the base case, while Figure 6.7 shows the corresponding difference in 2050 DUs by Bzone.

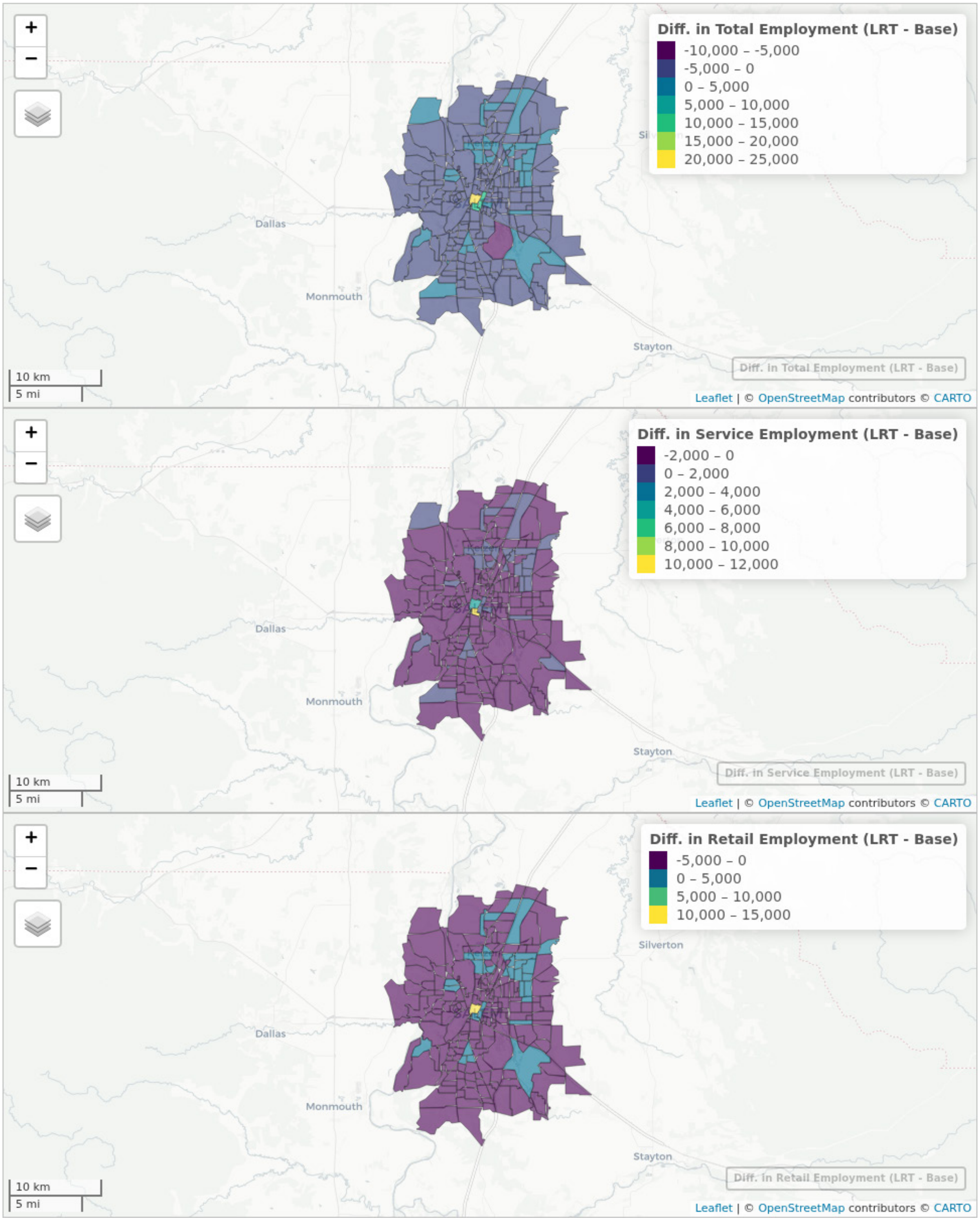

Figure 6.6: Synchronized interactive maps showing the difference in 2050 total, service, and retail employment by Bzone between the LRT scenario and the base case

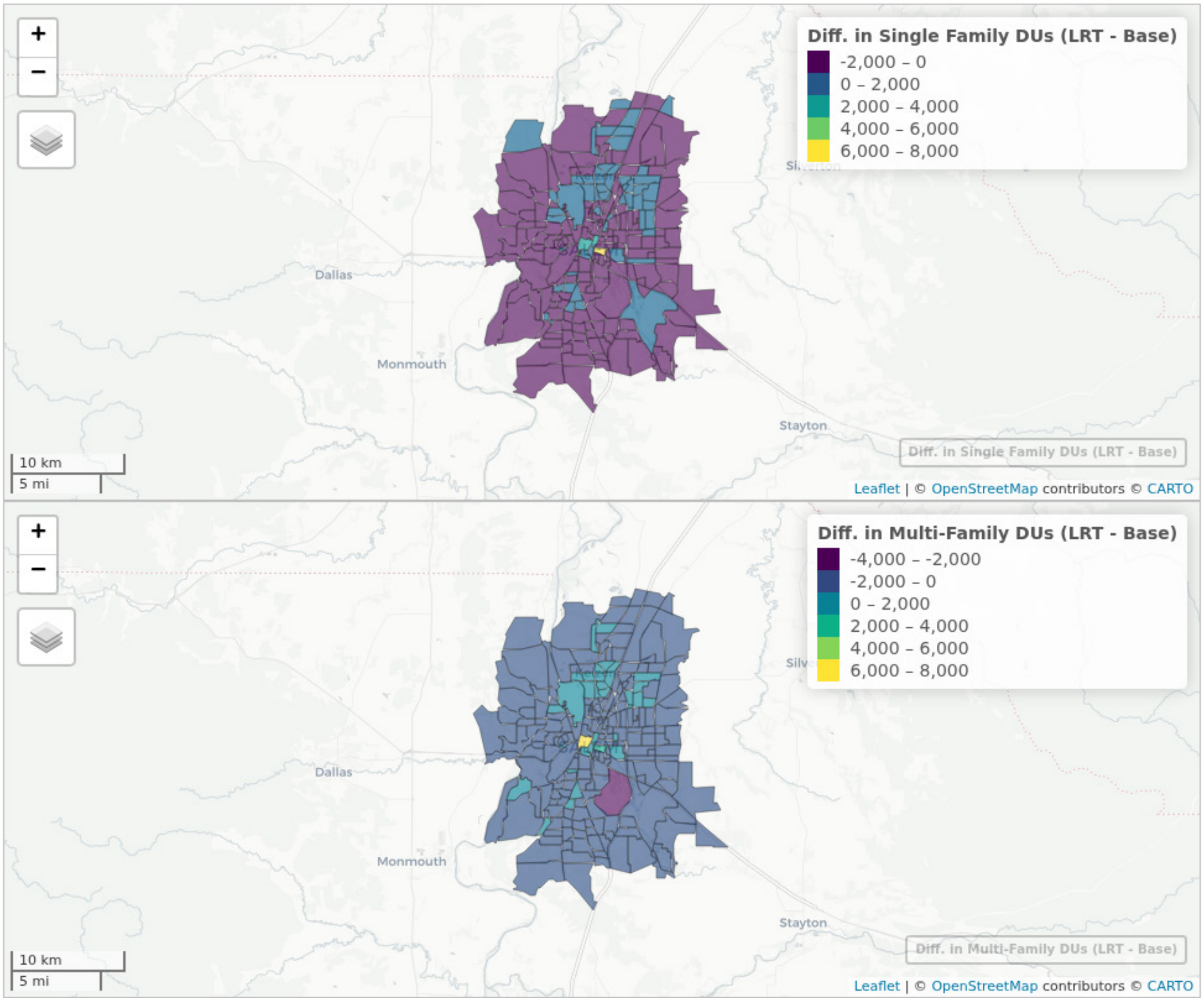

Figure 6.7: Synchronized interactive maps showing the difference in 2050 single-family and multi-family dwelling units by Bzone between the LRT scenario and the base case

Figure 6.8 summarizes the same LRT-to-base comparison as a distribution of Bzone-level changes, which makes it easier to see whether the scenario effect is widespread or concentrated in a smaller subset of zones.

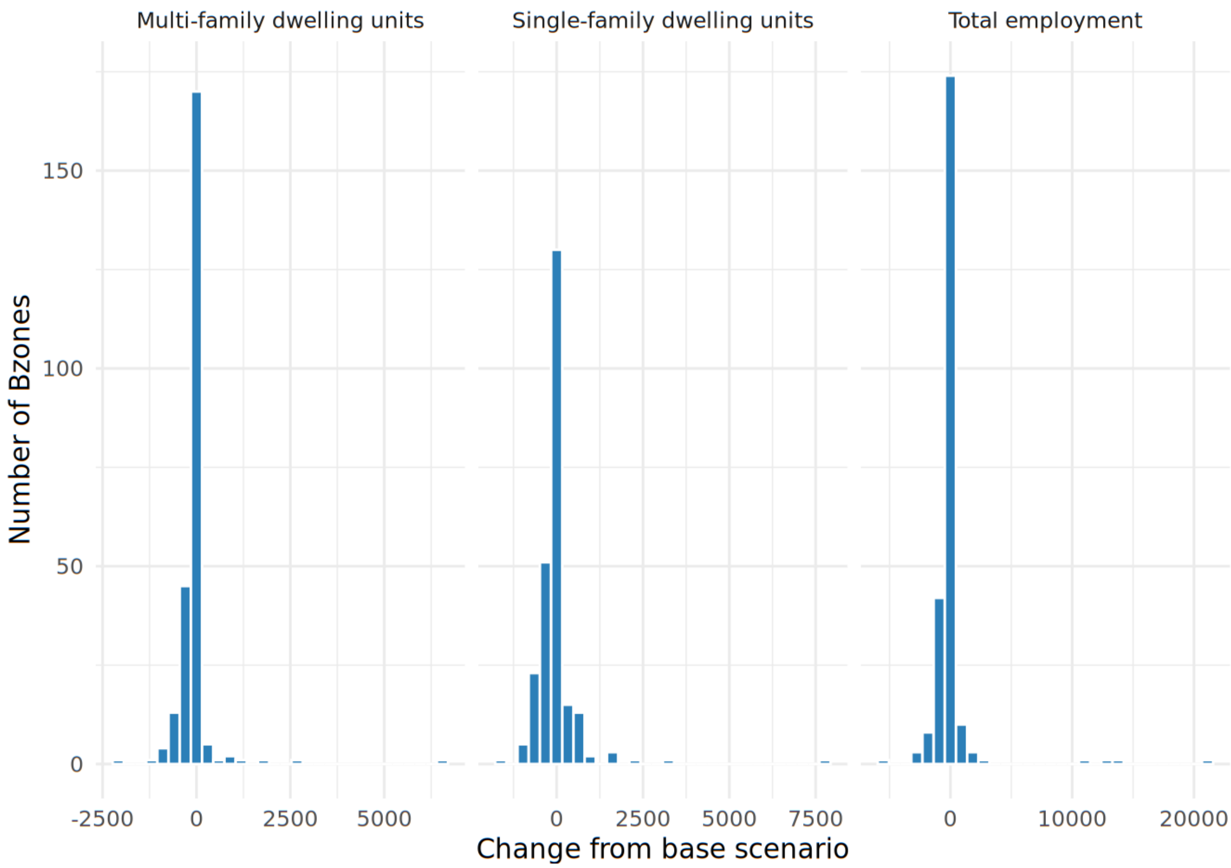

Figure 6.8: Distribution of Bzone-level changes in the SKATS hypothetical LRT scenario relative to the base case

Taken together, the LRT maps and histogram indicate a visible but still limited redistribution effect. Most SKATS Bzones remain close to the base case, while a smaller subset shows positive or negative shifts, which is consistent with a scenario that reallocates growth within fixed regional control totals rather than creating a regionwide increase. The effect is therefore meaningful in location but muted in aggregate scale: the scenario changes where housing and employment are placed more than it changes the overall magnitude of growth. The rendered maps also suggest that the visible shifts are organized along the same north-south Salem-to-Keizer corridor shown earlier in Figure 6.1, rather than appearing uniformly across the SKATS region.

6.2.1.5 Kuebler Interchange Scenarios

Like the LRT scenario, we also replaced bzone employment and dwelling units inputs with employment and DUs by land use type in the Kuebler Interchange scenario.

Figure 6.9 shows the difference in 2050 employment by Bzone between the Kuebler interchange scenario and the base case, while Figure 6.10 shows the corresponding difference in 2050 DUs by Bzone.

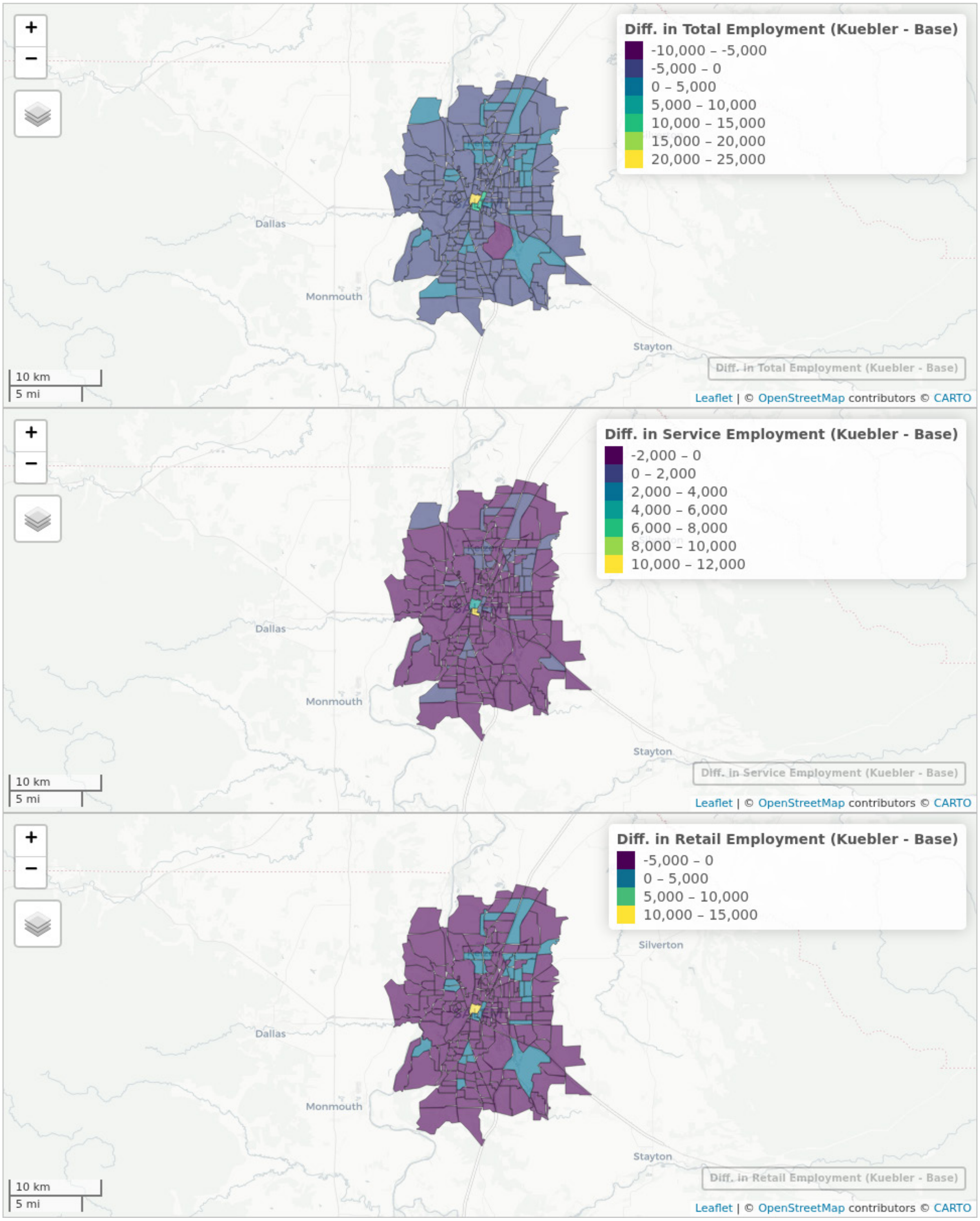

Figure 6.9: Synchronized interactive maps showing the difference in 2050 total, service, and retail employment by Bzone between the Kuebler interchange scenario and the base case

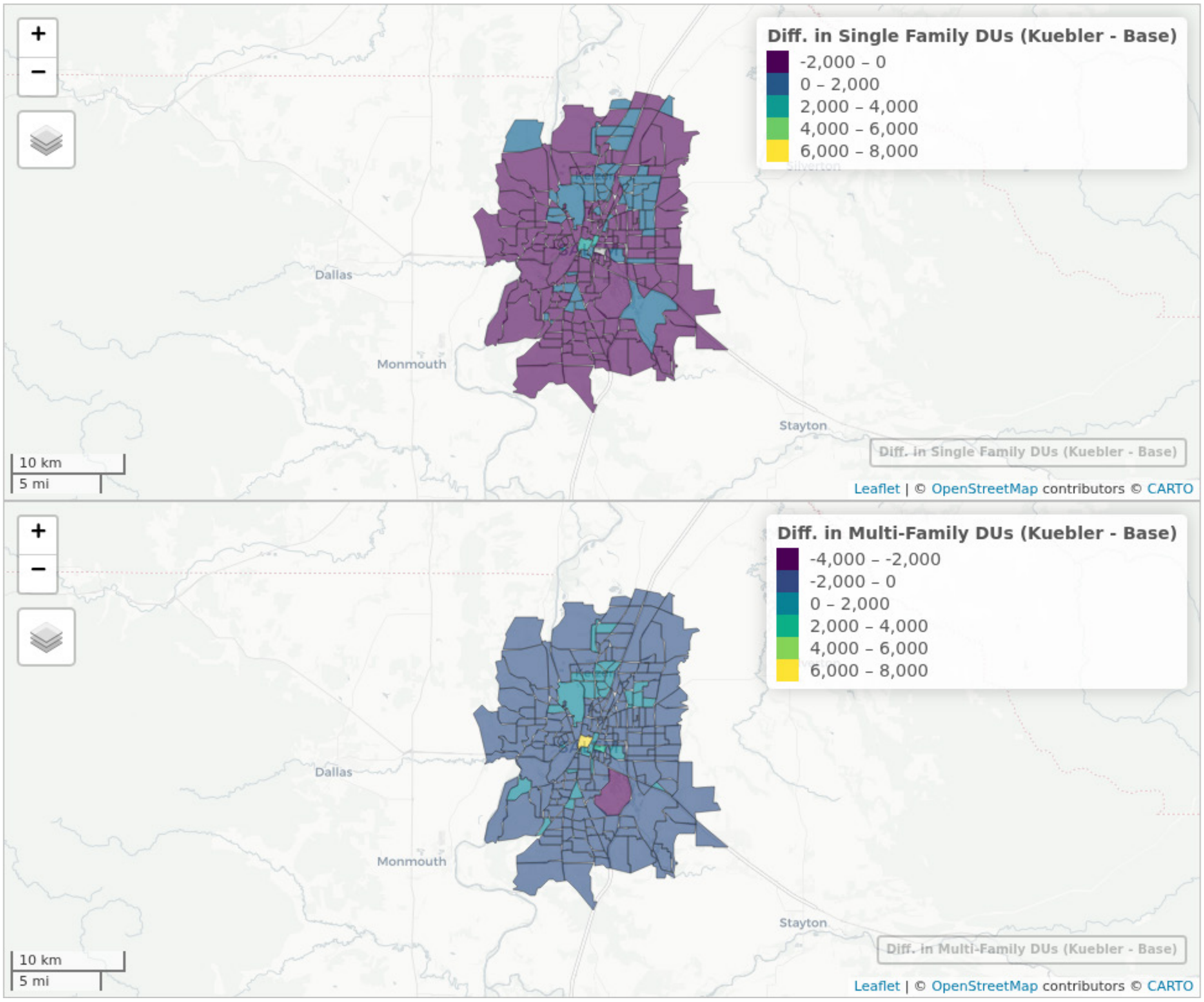

Figure 6.10: Synchronized interactive maps showing the difference in 2050 single-family and multi-family dwelling units by Bzone between the Kuebler interchange scenario and the base case

Figure 6.11 provides the same Bzone-level change distribution for the Kuebler scenario, making it easier to compare how diffuse or localized its effects are relative to the LRT case.

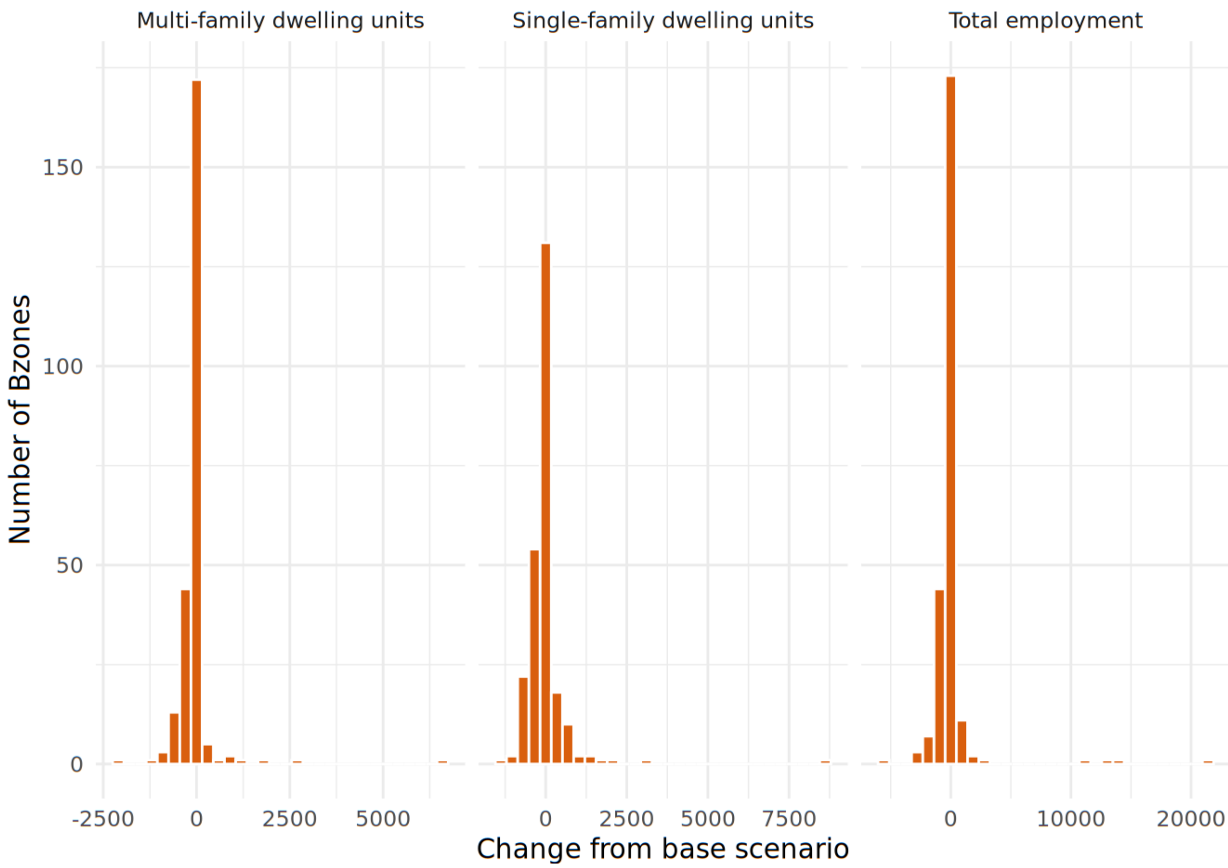

Figure 6.11: Distribution of Bzone-level changes in the SKATS Kuebler interchange scenario relative to the base case

The Kuebler scenario shows a similar pattern: the effect is visible, but it is concentrated in a relatively small share of Bzones rather than spread evenly across the region. This is the kind of result we would expect from a localized accessibility change. The maps suggest that the scenario is sensitive enough to redirect some housing and employment, but not so aggressive that it overwhelms the broader land use pattern inherited from the base case. As expected fromFigure 6.2, the most relevant changes should be read as a southeast Salem interchange effect rather than a metro-wide restructuring.

6.2.2 Four-County Area

The four-county pilot application uses data from the VEState model based on the VEState variants developed by RSG for Oregon, but trimmed down to Marion, Polk, Benton, and Linn counties. The base case uses the existing VESimLandUse module, along with VESimHouseholds and VESimTransportSupply modules, while the alternative scenario (Use Case #3) uses real Bzones (Census Block Groups) with the new VELandUse module, along with VESimHouseholds and VETransportSupply modules.

This larger pilot is therefore a comparison between the original VE-State structure and the new real-Bzone workflow. The original VE-State case remains the synthesized-Bzone reference, while the new scenario is the report's Use Case #3 demonstration of predicting future land use types and related allocations directly on real Bzones.

6.2.2.1 Model Steps of VEState Base Case

1. VESimHouseholds
    - CreateHouseholds
    - PredictWorkers
    - AssignLifeCycle
    - PredictIncome
2. VESimLandUse
    - CreateSimBzones
    - SimulateHousing
    - SimulateEmployment
    - Simulate4DMeasures
    - SimulateUrbanMixMeasure
    - AssignParkingRestrictions
    - AssignCarSvcAvailability
    - AssignDemandManagement
3. VESimTransportSupply
    - SimulateTransitService
    - SimulateRoadMiles

6.2.2.2 Model Steps of Use Case #3

1. VESimHouseholds
    - CreateHouseholds
    - PredictWorkers
    - AssignLifeCycle
    - PredictIncome

2. VELandUse (new)
   - PredictLUTypes
   - PredictLocTypes
   - AllocateDU
   - AllocateEmployment
   - PredictHousing
   - LocateEmployment
   - Calculate4DMeasures
   - CalculateUrbanMixMeasure
   - AssignCarSvcAvailability
   - AssignParkingRestrictions
   - AssignDemandManagement
3. VETransportSupply
   - AssignTransitService
   - AssignRoadMiles

6.2.2.3 Current VEState Modules

The current VEState model uses synthesized Bzones by dividing Azones with a specified median Bzone size by location type (Urban, Town, and Rural). Thus it is not possible to map the resulting employment and DUs by Bzone in the current VEState model.

6.2.2.4 New VELandUse with Real Bzone (Use Case #3)

For the Use Case #3 pilot, we use the 2010 Census Block Groups (CBG) as the real Bzones and populate the base year (2010) data from Census, LEHD, and Smart Location Database (SLD) data. Future year inputs are derived from those in the original VEState inputs.

Figure 6.12 shows the base year (2020) employment by Bzone, while Figure 6.13 shows the base year dwelling units.

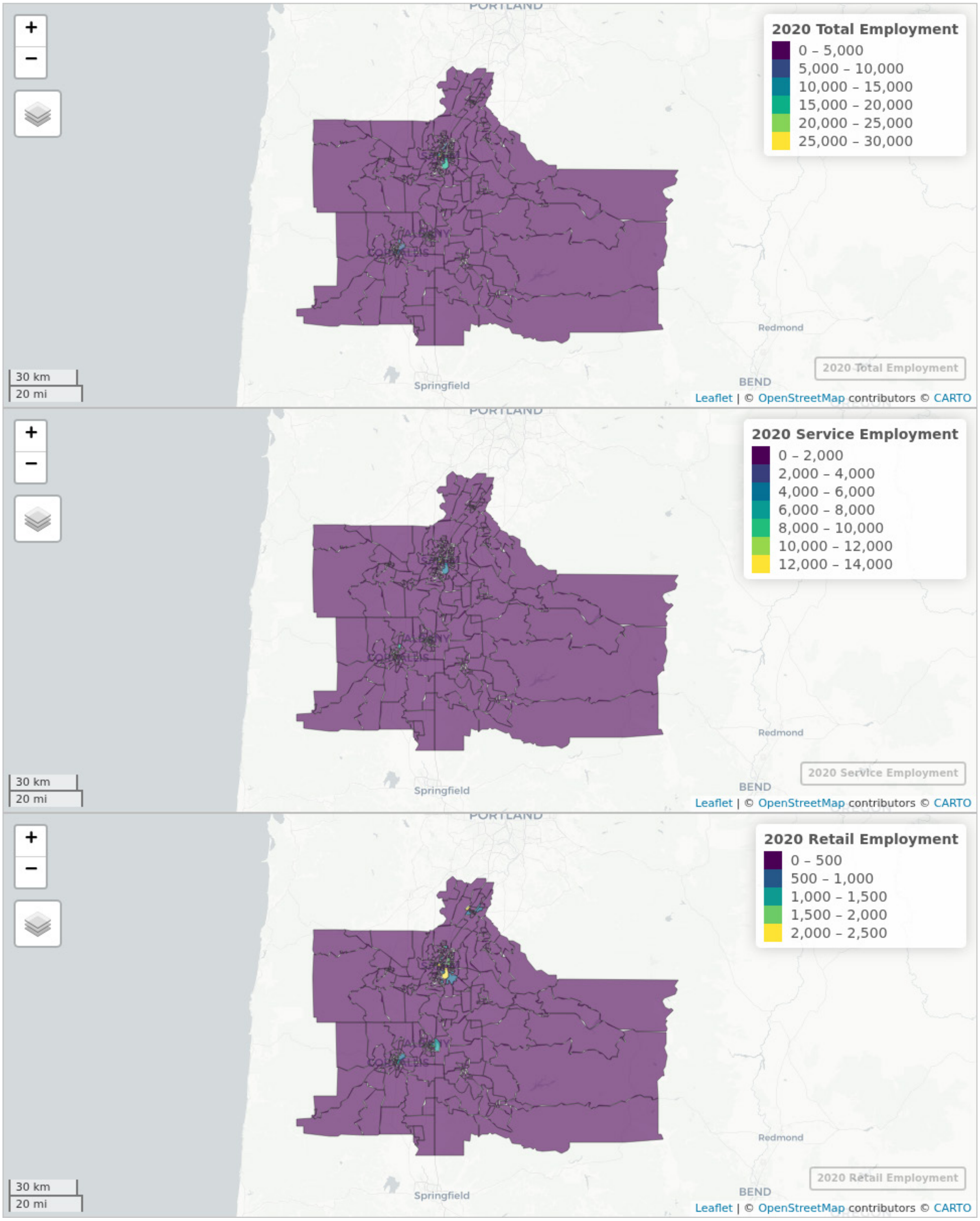

Figure 6.12: Synchronized interactive maps showing 2020 base year total, service, and retail employment by Bzone for the four-county Use Case 3 scenario

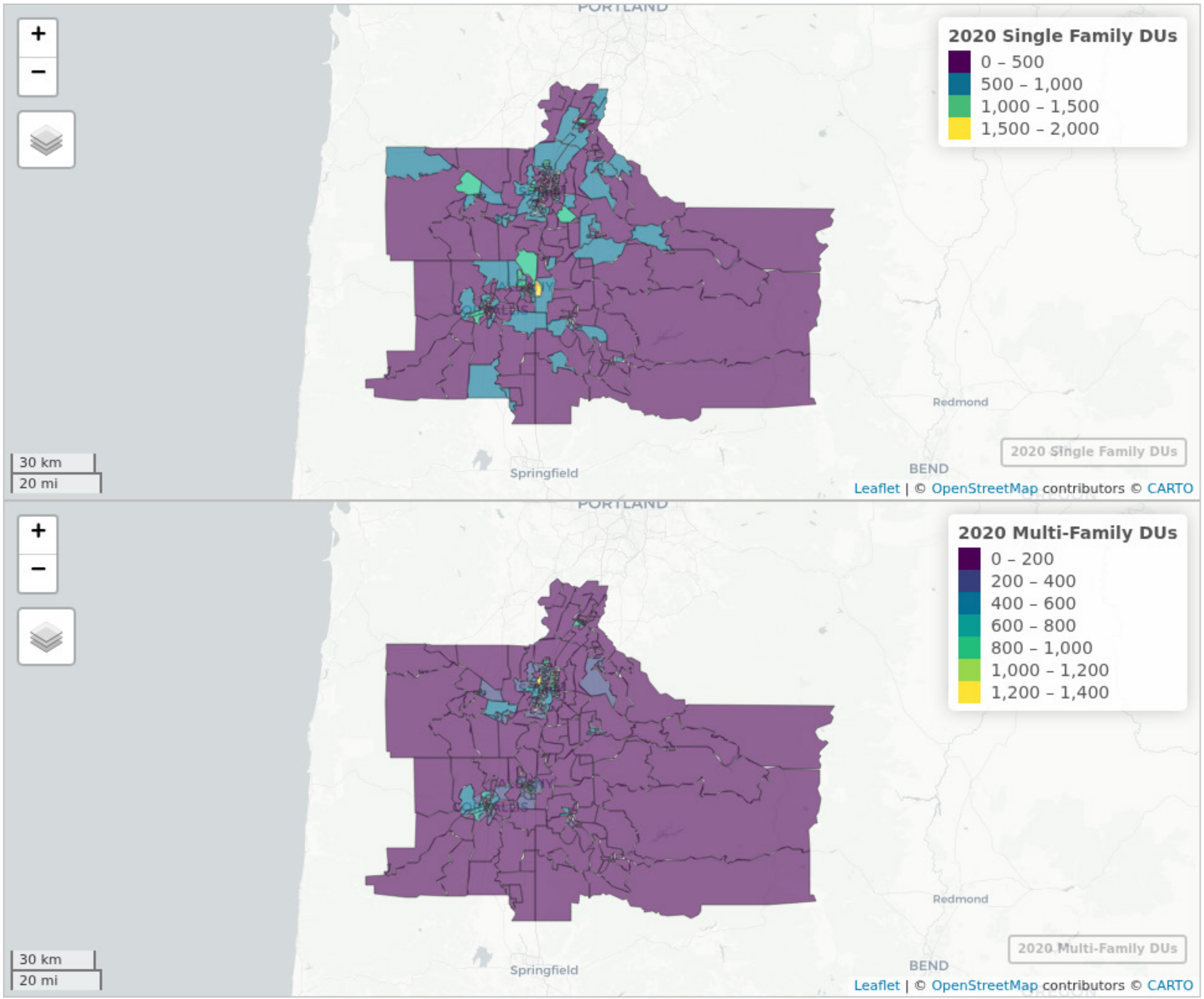

Figure 6.13: Synchronized interactive maps showing 2020 base year single-family and multi-family dwelling units by Bzone for the four-county Use Case 3 scenario

Figure 6.14 shows the difference in 2050 employment relative to the 2020 base year, while Figure 6.15 shows the corresponding difference in dwelling units.

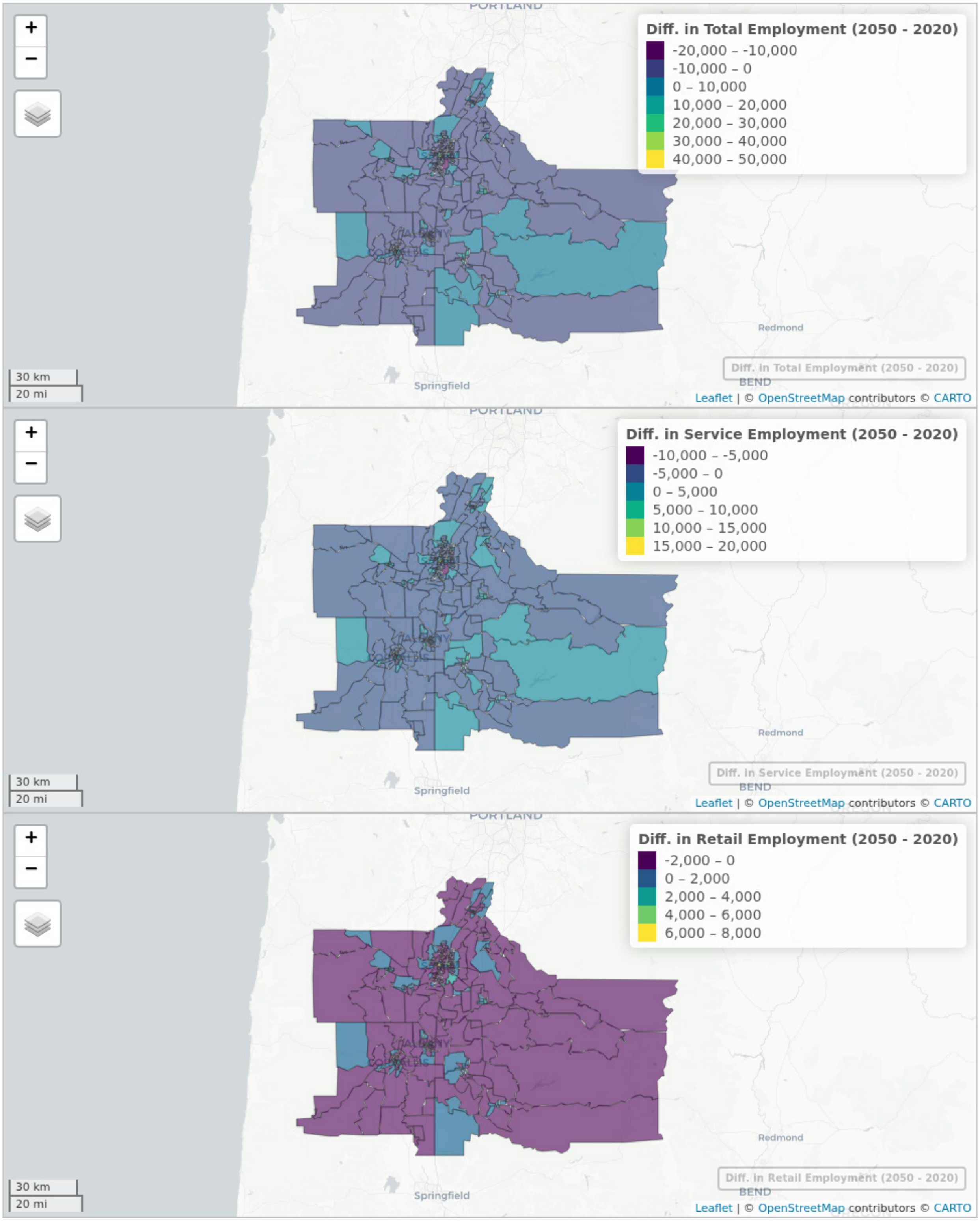

Figure 6.14: Synchronized interactive maps showing the difference in total, service, and retail employment by Bzone between 2050 and the 2020 base year for the four-county Use Case 3 scenario

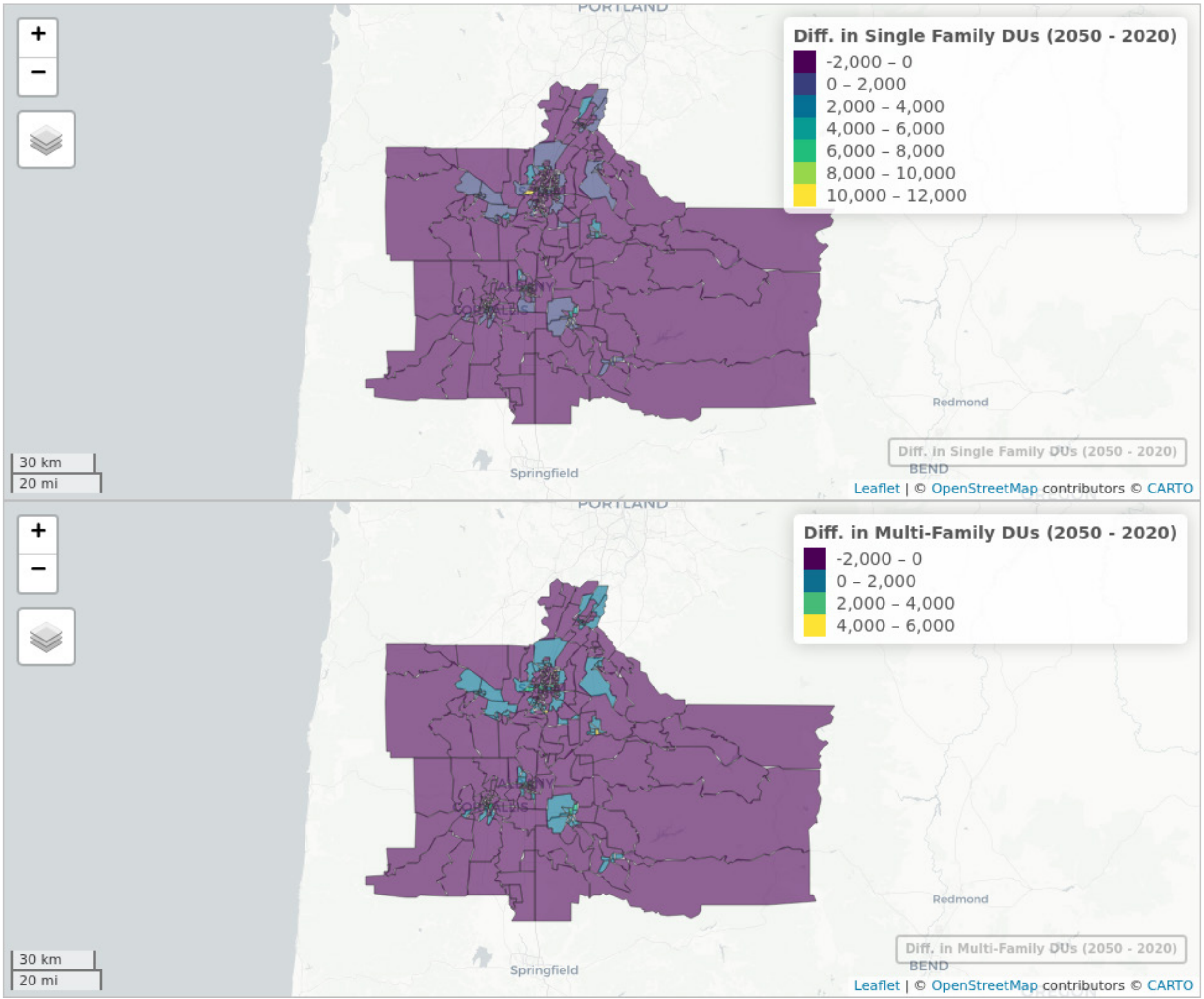

Figure 6.15: Synchronized interactive maps showing the difference in single-family and multi-family dwelling units by Bzone between 2050 and the 2020 base year for the four-county Use Case 3 scenario

With Use Case #3, we can show Area Type and Diversity Type by Bzone and their transitions. Figure 6.16 shows the Area Type and Diversity Type by Bzone for the Use Case #3.

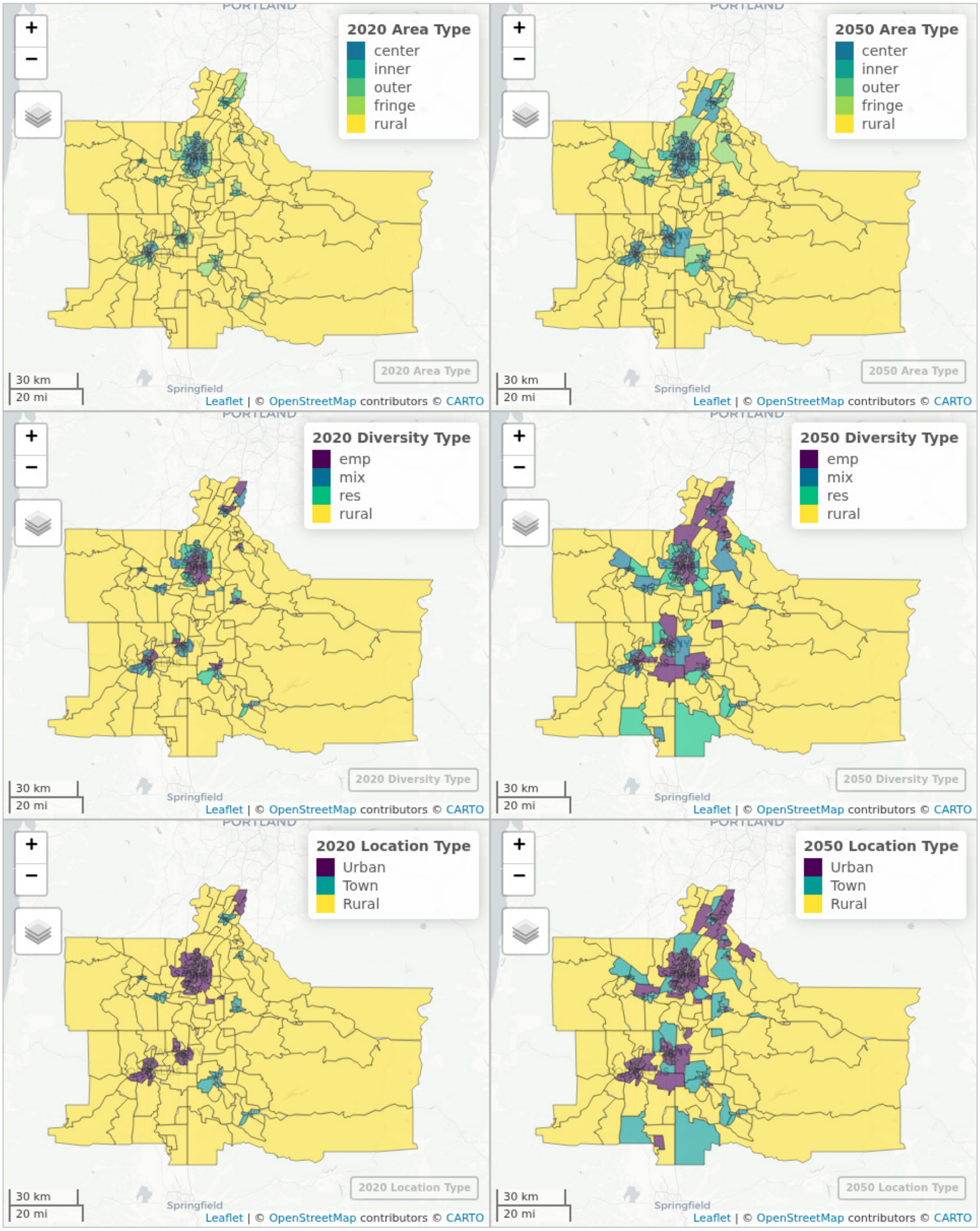

Figure 6.16: Synchronized interactive maps showing 2020 and 2050 Area Type, Diversity Type, and Location Type by Bzone for the four-county Use Case 3 scenario

The four-county Use Case #3 pilot shows a different pattern from the SKATS sensitivity tests. Here the main story is not a small redistribution around a single intervention, but a broader pattern of predicted land use evolution on real Bzones. The rendered maps suggest considerable transition of rural Location Type to town and urban surround existing urban areas in the region, consequently transition of rural Diversity Type to mix and res in these areas. In other words, the model appears to be more aggressive in transition rural land use types to urban land use types. Possible explanations for this pattern are that the transition models were estimated using nation-wide data and do not necessarily reflect Oregon-specific development pattern under tighter state-wide land use planning, and the Census Bureau's changes in urban area definitions between 2010 and 2020 may also contribute to this pattern.

This also means that the specific anomaly in an earlier version of the report, namely obvious rural-to-center or rural-to-inner jumps in the Salem area, is not prominent in the current output. InFigure 6.16, the downtown Salem center area remains in place, and most visible changes around Salem appear to be more incremental shifts among adjacent urban categories such as fringe, outer, and inner, rather than implausible direct jumps from clearly rural blocks into the urban core. Figure 6.18 likewise shows only scattered high-probability transition locations rather than a broad anomalous ring of Salem-area rural Bzones moving toward intense mixed urban development. Based on the current results, any remaining unusual transitions appear to be isolated cases, not a dominant pattern in the pilot output.

The same rendered results also help explain why some effects look muted. In the SKATS pilots, the visible changes are limited partly because the scenarios redistribute growth within fixed regional housing and employment totals and partly because the underlying interventions (hypotheticall LRT and remval of Kuebler Interchange) are geographically narrow, so a broad metro-wide shift would actually be less plausible than the concentrated changes shown in the maps. That said, if a larger shift in employment and housing to the LRT corridor is expected, a user can create an alternative scenario with a more aggressive shift in employment and housing to expected land use types in the lutype_employment.csv and lutype_dwelling_units.csv input.

In the four-county Use Case #3 pilot, the relative stability of LocType and the tendency for most visible change to occur among adjacent AreaType and DivType categories suggest a fairly conservative transition process. That pattern is consistent with a model that is being trained on nation-wide land use change data.

The PredictLUType submodule also outputs prediction probabilities for each dimension of the land use type, (Area Type, Diversity Type, and Location Type). It is possible to use the information to identify Bzones at the borderline of transitioning to different land use types. The information is stored in Bzone columns with a .Prob prefix. For example, probability for Bzone transitioning to mix Diversity Type is in column .ProbDivTypeMix. Figure 6.18 below is a map demonstrating the probability of Bzone transitioning to mix Diversity Type in 2050 (excluding Bzones already in mix Diversity Type).

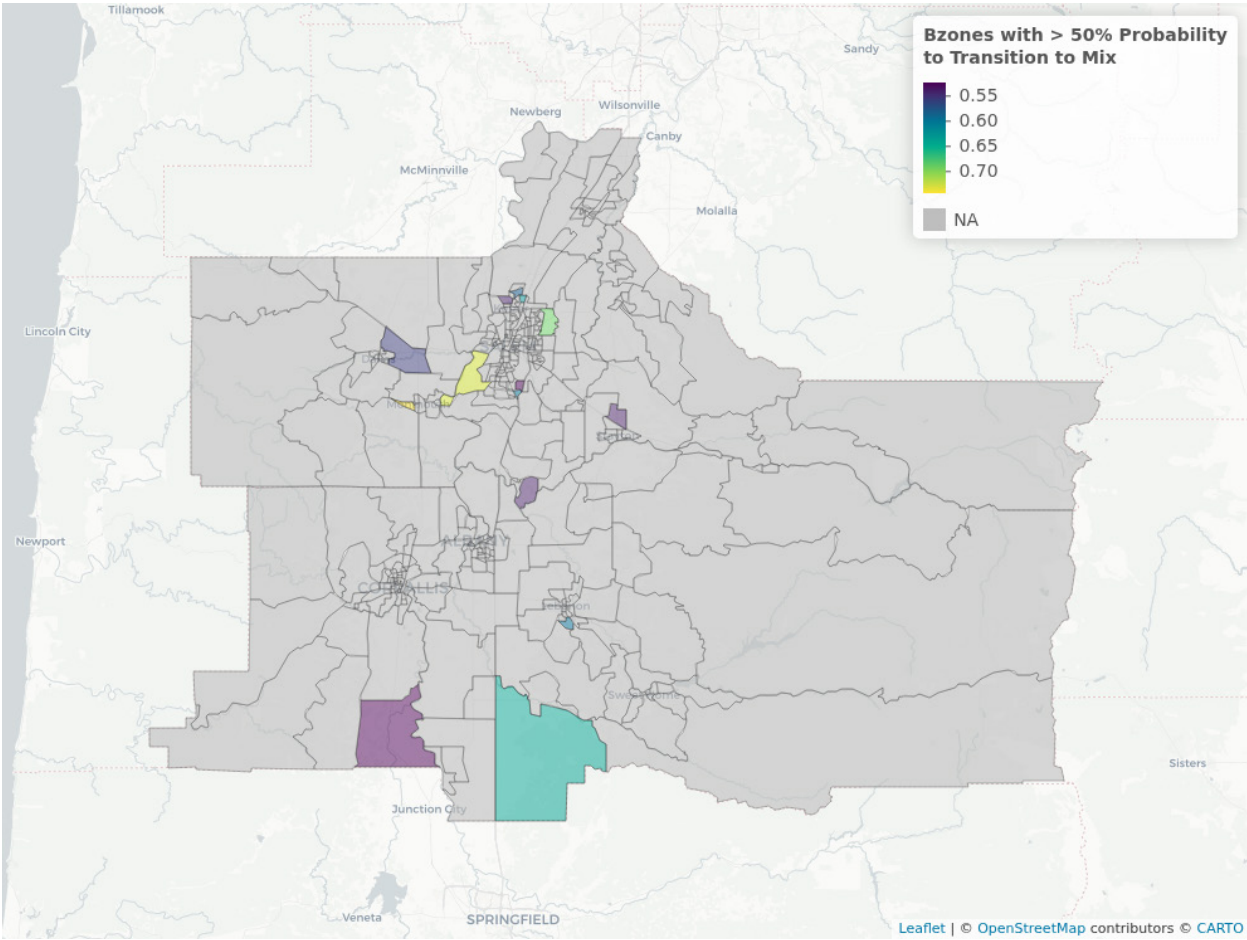

Figure 6.18: Interactive map highlighting Bzones with greater than 50 percent probability of transitioning to mixed-use diversity type by 2050

The pilot applications demonstrate that the new VELandUse module produces spatially plausible responses to both transit and highway infrastructure changes at the SKATS metropolitan scale, as well as coherent land use type predictions and employment and housing allocations across the larger four-county area using real Census Block Group Bzones. In the SKATS tests, the difference maps confirm that employment and dwelling unit shifts are concentrated near the infrastructure interventions — along the hypothetical LRT corridor and around the Kuebler interchange area — rather than distributed uniformly, which is a key indicator of model sensitivity. The four-county Use Case #3 pilot further shows that the module can predict future Area Type, Diversity Type, and Location Type transitions on real geography using base-year data, and that the resulting 2050-versus-2020 growth patterns align with expected urbanization trends. Taken together, these results provide confidence that the land-use-type-based workflow is a viable replacement for, and improvement upon, the direct Bzone input approach and the synthesized-Bzone approach used in the existing VE modules.

7.0 REFERENCES

Appleyard, B., Stanton, J., & Allen, C. (2020). *Toward a guide for smart mobility corridors: Frameworks and tools for measuring, understanding, and realizing transportation land use coordination*. Mineta Transportation Institute. https://doi.org/10.31979/mti.2020.1805

California Department of Transportation. (2010). *Smart mobility: A Caltrans handbook*.

Frost, A. R., Appleyard, B., Gibbons, J., & Ryan, S. (2018). Quantifying the sustainability, livability, and equity performance of urban and suburban places in California. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(3), 130–144. https://doi.org/10.1177/0361198118791382

Landau, W. M. (2021). The *targets* R package: A dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *6*(57), 2959. https://doi.org/10.21105/joss.02959

Lin, J., & Long, L. (2008). What neighborhood are you in? Empirical findings of relationships between household travel and neighborhood characteristics. *Transportation*, *35*(6), 739–758. https://doi.org/10.1007/s11116-008-9167-7

Outwater, M. (2014). *Effect of smart growth policies on travel demand*. Transportation Research Board.

Ralph, K., Voulgaris, C. T., & Brown, A. (2017). Travel and the built environment: Insights using activity densities, the sprawl index, and neighborhood types. *Transportation Research Record: Journal of the Transportation Research Board*, *2653*(1), 1–9. https://doi.org/10.3141/2653-01

Ralph, K., Voulgaris, C. T., Taylor, B. D., Blumenberg, E., & Brown, A. E. (2016). Millennials, built form, and travel insights from a nationwide typology of U.S. neighborhoods. *Journal of Transport Geography*, *57*, 218–226. https://doi.org/10.1016/j.jtrangeo.2016.10.007

Salon, D. (2015). Heterogeneity in the relationship between the built environment and driving: Focus on neighborhood type and travel purpose. *Research in Transportation Economics*, *52*, 34–45. https://doi.org/10.1016/j.retrec.2015.10.008

Salon, D., Boarnet, M. G., & Mokhtarian, P. (2014). Quantifying the effect of local government actions on VMT.

Shay, E., & Khattak, A. J. (2007). Automobiles, trips, and neighborhood type: Comparing environmental measures. *Transportation Research Record: Journal of the Transportation Research Board*, *2010*(1), 73–82. https://doi.org/10.3141/2010-09

Song, Y., & Knaap, G.-J. (2007). Quantitative classification of neighbourhoods: The neighbourhoods of new single-family homes in the Portland metropolitan area. *Journal of Urban Design*, *12*(1), 1–24. https://doi.org/10.1080/13574800601072640

Southern California Association of Governments. (2020). *Scenario planning model*.

Voulgaris, C. T., Taylor, B. D., Blumenberg, E., Brown, A., & Ralph, K. (2016). Synergistic neighborhood relationships with travel behavior: An analysis of travel in 30,000 US neighborhoods. *Journal of Transport and Land Use*, *10*(1). https://doi.org/10.5198/jtlu.2016.840

APPENDIX A: TERM AND VARIABLE DEFINITIONS

This appendix collects terms, abbreviations, and variables used across the active report chapters. It supplements in-text definitions by providing a single reference point for items that may otherwise appear without an immediate explanation.

Table A.1: General terms and abbreviations used in the report

Term	Definition
VisionEval (VE)	Open-source strategic modeling framework used to evaluate transportation, land use, and related policy scenarios.
Bzone	Basic land-use analysis zone used by VisionEval for land use inputs, outputs, and spatial measures.
Azone	Aggregate zone above the Bzone level used for broader demographic, economic, and scenario inputs.
Marea	Metropolitan area identifier used in VisionEval to group related Azones and Bzones.
VE-RSPM	VisionEval implementation used for Regional Strategic Planning Model applications that work with real Bzones.
VE-State	VisionEval implementation used for large regional and statewide applications.
VELandUse	VisionEval package that allocates land use, households, and employment using real Bzones.
VESimLandUse	Legacy VisionEval package that works with simulated Bzones, mainly for VE-State models.
RSPM	Regional Strategic Planning Model.
RPAT	Rapid Policy Analysis Tool.
CBG	Census Block Group.
SLD	Smart Location Database published by the U.S. Environmental Protection Agency.
NHTS	National Household Travel Survey.
ACS	American Community Survey.
LEHD	Longitudinal Employer-Household Dynamics employment data.

Term	Definition
Location Type (LocType)	Broad settlement context of a Bzone, such as Urban, Town, or Rural.
Area Type (AreaType)	Relative urban intensity of a Bzone within its location type, such as center, inner, outer, fringe, or regional center.
Diversity Type (DivType)	Land use composition category of a Bzone, such as residential, employment, or mixed.
LUType	Combined land use type label created from AreaType and DivType.
Urban mix	Neighborhood attribute used in older VisionEval travel modules to represent an urban mixed-use setting.
Fixed guideway transit	Transit service operating on rail or other dedicated guideway facilities, represented in this report through distance and availability variables.
TDM	Travel demand management policies such as commute programs, parking management, and related demand-shaping measures.

Table A.2: Report-specific variables and datastore fields used in the report

Name	Definition
D1	Density measures of the built environment.
D2	Diversity measures of the built environment.
D3	Design measures of the built environment.
D4	Transit measures of the built environment.
D5	Accessibility to destinations.
SFDU	Single-family dwelling units by Bzone.
MFDU	Multi-family dwelling units by Bzone.
GQDU	Group quarters dwelling units by Bzone.
TotEmp	Total employment by Bzone.

Name	Definition
RetEmp	Retail employment by Bzone.
SvcEmp	Service employment by Bzone.
D1B	Population density.
D1C	Employment density.
D1D	Gross activity density, usually defined as employment plus housing units on unprotected land.
D2A_JPHH	Jobs per household.
D2A_WRKEMP	Ratio of workers living in the zone to jobs located in the zone.
D2A_EPHHM	Entropy-based land use mix measure using households and employment categories.
D3BPO4	Pedestrian-oriented four-leg intersection density measure.
D3bpo4	Same D3BPO4 measure when referenced with the mixed-case naming used in some VisionEval inputs and text.
D3Lvl	Ordinal urban-design class predicted by AssignD3D4Levels, with higher values indicating stronger walk-supportive design conditions.
D4C	Aggregate frequency of transit service near a Census Block Group boundary during the evening peak period.
D4Lvl	Ordinal transit-service class predicted by AssignD3D4Levels, with higher values indicating stronger transit-service conditions.
D5	Destination proximity measure based on nearby jobs and population.
DistToStop	Distance to the nearest transit stop.
DistToFgwSta	Distance to the nearest fixed guideway transit station.
HasFgwTransit	Indicator for whether the metropolitan area has fixed guideway transit service.
UZAAVRMPA	Urbanized area annual transit vehicle revenue miles per acre.
UZAAVRMPC	Urbanized area annual transit vehicle revenue miles per capita.

Name	Definition
UZAPOP	Urbanized area population.
UZAAREA	Urbanized area land area.
UZAPOPDEN	Urbanized area population density.
DistToRamp	Distance to the nearest freeway ramp or interchange.
DistToCBD	Distance to the central business district.
PctSteepSlope	Share of land area with steep slope constraints.
D1D_hmbuf	Activity density within a half-mile buffer around the Bzone centroid.
D1D_hmcbuf	Buffered activity density measure used in area-type classification and related model estimation.
D2A_JPHH_hmbuf	Jobs-per-household measure calculated within a half-mile buffer around the Bzone centroid.
.Prob prefix	Prefix used for probability outputs written by PredictLUType for possible future land use type outcomes.
.DivTypeProbMix	Predicted probability that a Bzone transitions to the mix diversity type.

Table A.3: Selected Smart Location Database variables referenced in the report

Name	Definition
D1D	Gross activity density (employment + HUs) on unprotected land
D2A_JPHH	Jobs per household
D3A	Total road network density
D3AAO	Network density in terms of facility miles of auto-oriented links per square mile
D3AMM	Network density in terms of facility miles of multi-modal links per square mile

Name	Definition
D3APO	Network density in terms of facility miles of pedestrian-oriented links per square mile
D3B	Street intersection density (weighted, auto-oriented intersections eliminated)
D3BAO	Intersection density in terms of auto-oriented intersections per square mile
D3BMM3	Intersection density in terms of multi-modal intersections having three legs per square mile
D3BMM4	Intersection density in terms of multi-modal intersections having four or more legs per square mile
D3BPO3	Intersection density in terms of pedestrian-oriented intersections having three legs per square mile
D3BPO4	Intersection density in terms of pedestrian-oriented intersections having four or more legs per square mile
D4A	Distance from population-weighted centroid to nearest transit stop (meters)
D4C	Aggregate frequency of transit service within 0.25 miles of CBG boundary per hour during evening peak period
DistToStop	Distance to nearest transit stop (meter)
DistToFgwSta	Distance to nearest fixed guideway transit stop (meter)
HasFgwTransit	Binary indicator for presence of fixed guideway transit in an Urban Area
UZAAVRMPA	Urban Area annual transit vehicle revenue miles per acre
UZAAVRMPC	Urban Area annual transit vehicle revenue miles per capita
UZAPOP	Total population of Urban Area
UZAAREA	Total land area of Urban Area (sq miles)
UZAPOPDEN	Population density of urban area (persons per sq mile)