VTRC
Virginia Transportation Research Council

We Bring Innovation to Transportation

Development of Demand Estimation Models for the Virginia Department of Transportation's Park and Ride Facilities

http://www.virginiadot.org/vtrc/main/online_reports/pdf/21-r19.pdf

YIQING XU
Graduate Research Assistant

LANCE E. DOUGALD
Senior Research Scientist

JOHN S. MILLER
Principal Research Scientist

Final Report VTRC 21-R19

VIRGINIA TRANSPORTATION RESEARCH COUNCIL
530 Edgemont Road, Charlottesville, VA 22903-2454
vtrc.virginiadot.org

Standard Title Page - Report on Federally Funded Project

1. Report No.: FHWA/VTRC 21-R19	2. Government Accession No.:	3. Recipient's Catalog No.:
4. Title and Subtitle: Development of Demand Estimation Models for the Virginia Department of Transportation's Park and Ride Facilities		5. Report Date: May 2021
		6. Performing Organization Code:
7. Author(s): Yiqing Xu, Lance E. Dougald, and John S. Miller		8. Performing Organization Report No.: VTRC 21-R19
9. Performing Organization and Address: Virginia Transportation Research Council, 530 Edgemont Road, Charlottesville, VA 22903		10. Work Unit No. (TRAIS):
		11. Contract or Grant No.: 117664
12. Sponsoring Agencies' Name and Address: Virginia Department of Transportation, 1401 E. Broad Street, Richmond, VA 23219	Federal Highway Administration, 400 North 8th Street, Room 750, Richmond, VA 23219-4825	13. Type of Report and Period Covered: Final
		14. Sponsoring Agency Code:

15. Supplementary Notes:
This is an SPR-B report.

16. Abstract:

The construction and maintenance of park and ride lots represents a substantial public investment that if used judiciously can reduce congestion and emissions through the use of transit or the sharing of vehicle trips. With 297 lots scattered throughout Virginia, the Virginia Department of Transportation (VDOT) needs an approach for forecasting demand for these lots so that investments can be made wisely. Unfortunately, direct application of an existing approach yielded absolute differences (between forecast occupancy and observed occupancy) that depending on the VDOT district were 14 to 141 times larger than the observed occupancy. Calibrating an existing approach to Virginia-specific traffic volumes for the roadway serving the lot and the highest volume roadway within 2.5 miles of the lot reduced the scale of this error but still yielded forecasts where the mean testing error exceeded the mean occupancy for a majority of models.

Accordingly, 19 Virginia-specific models were developed that reflected distinct regions in Virginia. These models followed VDOT district boundaries for three of VDOT's nine districts (Lynchburg, Richmond, and Northern Virginia); planning district commission (PDC) or metropolitan planning organization (MPO) boundaries for four VDOT districts (Bristol, Culpeper, Salem, and Staunton); and urban/rural categorizations for two VDOT districts (Fredericksburg and Hampton Roads). A key finding was that determinants of occupancy varied by location. Statistically significant determinants included residents with a commute of 50+ miles (used in four models affecting 10% of Virginia's lots, such as those in the Lenowisco PDC in the Bristol District); the availability of transit service or the number of commuters who choose transit (used as a positive factor in seven models affecting more than one-half [151] of Virginia's lots, such as those in the urbanized portion of the Culpeper District); amenities such as lighting (a variable in two models reflecting 15% of Virginia's lots such as those in the low population density areas of the Fredericksburg District); traffic volume (a factor in five models representing 46% of Virginia's lots, such as those in the Lynchburg District); and the provision of bicycle spaces (a factor in the model for 78 of the Northern Virginia District lots, or about 26% of the statewide total). Thus, the models can help forecast how key changes (such as an increase in traffic, the introduction of transit service, or the addition of lighting) may influence demand at an existing lot.

The median-adjusted R-squared value (coefficient of determination) for the 19 models was 87%. The Richmond District was representative: a model based on the average 24-hour traffic volume of all facilities within 2.5 miles of the lot and the nearest peak hour expansion factor explained 86.7% of the variation in occupancy for the 11 lots in that district. When the models were tested on a dataset not used to build the models, the median ratio of mean testing error to mean occupancy was 56%. A typical model in this regard was for the lots in the Roanoke Valley-Alleghany Regional Commission (in the Salem District) where occupancy was based on the presence of transit service and the proportion of nearby residents with commutes of 25 to 50 miles: the mean testing error was 14 compared to a mean lot occupancy of 25, for a ratio of 56%. The models thus explained a portion of the variation in demand and can inform forecasts for new lots, although these results demonstrated that additional site-specific factors not included in each model also influenced demand.

17 Key Words: Park and ride, multimodal transportation, fringe parking, parking demand	18. Distribution Statement: No restrictions. This document is available to the public through NTIS, Springfield, VA 22161.

19. Security Classif. (of this report): Unclassified	20. Security Classif. (of this page): Unclassified	21. No. of Pages: 87	22. Price:

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

FINAL REPORT

DEVELOPMENT OF DEMAND ESTIMATION MODELS FOR THE VIRGINIA DEPARTMENT OF TRANSPORTATION'S PARK AND RIDE FACILITIES

Yiqing Xu
Graduate Research Assistant

Lance E. Dougald
Senior Research Scientist

John S. Miller
Principal Research Scientist

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Virginia Transportation Research Council
(A partnership of the Virginia Department of Transportation
and the University of Virginia since 1948)

Charlottesville, Virginia

June 2021
VTRC 21-R19

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation. Any inclusion of manufacturer names, trade names, or trademarks is for identification purposes only and is not to be considered an endorsement.

ABSTRACT

The construction and maintenance of park and ride lots represents a substantial public investment that if used judiciously can reduce congestion and emissions through the use of transit or the sharing of vehicle trips. With 297 lots scattered throughout Virginia, the Virginia Department of Transportation (VDOT) needs an approach for forecasting demand for these lots so that investments can be made wisely. Unfortunately, direct application of an existing approach yielded absolute differences (between forecast occupancy and observed occupancy) that depending on the VDOT district were 14 to 141 times larger than the observed occupancy. Calibrating an existing approach to Virginia-specific traffic volumes for the roadway serving the lot and the highest volume roadway within 2.5 miles of the lot reduced the scale of this error but still yielded forecasts where the mean testing error exceeded the mean occupancy for a majority of models.

Accordingly, 19 Virginia-specific models were developed that reflected distinct regions in Virginia. These models followed VDOT district boundaries for three of VDOT's nine districts (Lynchburg, Richmond, and Northern Virginia); planning district commission (PDC) or metropolitan planning organization (MPO) boundaries for four VDOT districts (Bristol, Culpeper, Salem, and Staunton); and urban/rural categorizations for two VDOT districts (Fredericksburg and Hampton Roads). A key finding was that determinants of occupancy varied by location. Statistically significant determinants included residents with a commute of 50+ miles (used in four models affecting 10% of Virginia's lots, such as those in the Lenowisco PDC in the Bristol District); the availability of transit service or the number of commuters who choose transit (used as a positive factor in seven models affecting more than one-half [151] of Virginia's lots, such as those in the urbanized portion of the Culpeper District); amenities such as lighting (a variable in two models reflecting 15% of Virginia's lots such as those in the low population density areas of the Fredericksburg District); traffic volume (a factor in five models representing 46% of Virginia's lots, such as those in the Lynchburg District); and the provision of bicycle spaces (a factor in the model for 78 of the Northern Virginia District lots, or about 26% of the statewide total). Thus, the models can help forecast how key changes (such as an increase in traffic, the introduction of transit service, or the addition of lighting) may influence demand at an existing lot.

The median-adjusted R-squared value (coefficient of determination) for the 19 models was 87%. The Richmond District was representative: a model based on the average 24-hour traffic volume of all facilities within 2.5 miles of the lot and the nearest peak hour expansion factor explained 86.7% of the variation in occupancy for the 11 lots in that district. When the models were tested on a dataset not used to build the models, the median ratio of mean testing error to mean occupancy was 56%. A typical model in this regard was for the lots in the Roanoke Valley-Alleghany Regional Commission (in the Salem District) where occupancy was based on the presence of transit service and the proportion of nearby residents with commutes of 25 to 50 miles: the mean testing error was 14 compared to a mean lot occupancy of 25, for a ratio of 56%. The models thus explained a portion of the variation in demand and can inform forecasts for new lots, although these results demonstrated that additional site-specific factors not included in each model also influenced demand.

TABLE OF CONTENTS

FINAL REPORT

DEVELOPMENT OF DEMAND ESTIMATION MODELS FOR THE VIRGINIA DEPARTMENT OF TRANSPORTATION'S PARK AND RIDE FACILITIES

Yiqing Xu
Graduate Research Assistant

Lance E. Dougald
Senior Research Scientist

John S. Miller
Principal Research Scientist

INTRODUCTION

Park and ride lots provide common locations for individuals to transfer from a low to a high occupancy travel mode. Such lots are thus a critical part of a multimodal transportation system, providing benefits to users (in the form of reduced travel costs); the roadway network (by increasing vehicle occupancy rather than the number of vehicles); and the environment by reducing emissions (Virginia Department of Transportation [VDOT], 2019a). Investments in park and ride lots are a part of Virginia's multimodal program, and demand at such lots affects their relative value when programming decisions are made, such as selection of candidate projects in SMART SCALE. For that reason, VDOT is interested in being able to forecast demand at these lots, especially as a function of potential influences such as lot amenities, traffic volume, or population.

The research team is aware of 297 park and ride lots in Virginia including state-owned, privately owned, and unofficial lots. Most lots are concentrated in urbanized areas (e.g., VDOT's Hampton Roads, Richmond, and Northern Virginia districts, as shown in Figure 1) and the I-81 corridor. VDOT performed lot audits in 1996, 2002, 2011, 2016, and 2018. These audits typically provided detailed information on each lot such as capacity (parking spaces); occupancy (vehicle counts); and other attributes such as lot surface type, lighting, transit service and shelters, bicycle accommodations, handicapped spaces, and parking fees (if applicable). Data obtained from such audits also enable VDOT to develop plans for future lots or consider expansion of existing lots based on demand forecasts.

Recent Virginia Forecasting Initiatives

Two initiatives prior to the current study to forecast demand at park and ride lots in Virginia did not yield useful information, at least on a small scale. First, the technical review panel (TRP) overseeing the current study reported that an approach employed by the Florida Department of Transportation (FDOT) (FDOT, 2012) yielded absolute errors for individual lots (difference between forecast and observed occupancy) that tended to be 3 to 4 times larger than the observed occupancy. In the Fredericksburg, Virginia, area, these errors were underestimates—forecast occupancy was between one-fifth and one-third observed occupancy; in VDOT's Staunton District, forecast occupancy was an overestimate of between 3 and 4 times observed occupancy (Mobayed, 2019). Second, and prior to the current study, VDOT's Transportation Mobility and Planning Division's (TMPD) application of a corridor-specific strategy (BMI et al., 2003) underestimated 2019 demand for Stafford County lots by a factor of 17 (where the 2003 model gave a forecast for 2020 demand and the research team presumed that the 2020 demand would be the same as the observed demand in 2019). The error associated with these demand estimation methods prompted interest from the TRP to develop new forecasting methods.

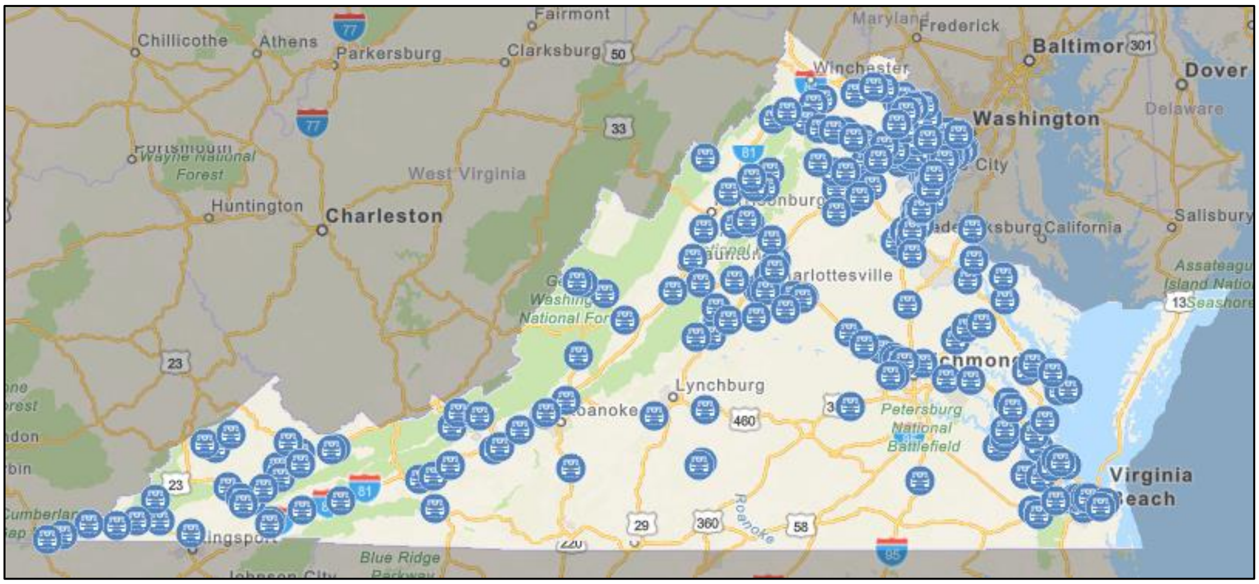

Figure 1. The 297 Park and Ride Lots in Virginia

Literature Review of Previous Forecasts

The literature included examples of models that were developed to forecast park and ride lot demand, and the research team sought to apply these types of models to Virginia. There were also several sources that although not explicitly developing a mathematical model for estimating occupancy based on observed behavior nonetheless informed the development of models that were later used in this particular study. This literature was categorized across five overlapping areas:

1. diversion models
2. regression-based models
3. other analytical approaches
4. lessons learned from forecasting
5. implications of the literature review.

Diversion Models

One approach is to estimate demand as some portion of the total traffic on the facility, or facilities, near the park and ride lot. This approach presumes that the only lot users are those who pass by the lot as part of their normal travel path, such that a road with no volume will yield a park and ride lot with zero occupancy. Five publications (FDOT, 2012; Nungesser and Ledbetter, 1987; Southwest Region Planning Commission et al., 2015; Turnbull, 1995; Vincent, 2007) referred to this approach as having been developed by the Institute of Transportation Engineers.

This approach has been defined in at least three slightly different ways in terms of the peak hour volume and the percent of traffic that is forecast to use the park and ride lot.

1. FDOT (2012) considered both a "primary" facility, defined as the "main commuting roadway" near the park and ride lot (with a diversion factor of 0.03), and "secondary" facilities that are commuting routes of "lesser importance" (with a diversion factor of 0.01). FDOT did not specify an explicit distance but rather stated: "The primary roadway is considered to be the main commuting roadway in the vicinity of the Park-and-Ride lot."

2. Nungesser and Ledbetter (1987) referred to a seemingly similar approach but with one key difference: the "secondary" facilities cited by FDOT (2012) were instead defined as "total peak period traffic on adjacent facilities (including the prime facility)." Vincent (2007) appeared to have used the same definition as Nungesser and Ledbetter (1987) but described the model only in a literature review. Nungesser and Ledbetter (1987) reported that for lots in Houston (Texas), calibration was essential, as the authors found that the diversion factors should be 0.083 (rather than 0.03) for the primary facility on which the lot is situated and -0.036 (rather than 0.01) for the other peak facilities, as the authors removed the primary facility from these other peak facilities for the purpose of calibration.

3. For New Hampshire lots, the Southwest Region Planning Commission et al. (2015) modified the Institute of Traffic Engineers method to use a single parameter—the percent of peak traffic passing by the facility or near the facility was 1% if two or fewer demographic thresholds were exceeded by communities near the lot or 3% if all five demographic thresholds were exceeded. If three or four thresholds were exceeded, the percentage was 2%. These thresholds are on a square mile basis: (1) population age 15-39 (47.3); (2) persons working outside the home (53.47); (3) number of households where housing costs exceed 30% of household income (22.68); (4) non-family households (19.28); and (5) households with 0 or 1 vehicles (20.8). One may suppose, for example, a community has 60 persons age 15-39 and 70 persons working outside the home per square mile, which exceeds Thresholds 1 and 2. For such a community, if Thresholds 3, 4, and 5 were not exceeded the percentage would be 1%. The authors indicated that this approach may be applied to more than one commuter facility near the lot.

Regression-Based Models

In addition to the diversion model, linear regression models have been tailored to specific park and ride lots. Nordstrom and Christiansen (1981) used a linear regression model to forecast park and ride lot demand in six cities in Texas as a function of population within the catchment area; a custom congestion index (which in turn was based on delay and average daily traffic [ADT]); and a third variable that was based on the number of spaces at the lot and the number of bus seats during peak periods. The authors explained that park and ride lots sometimes exceeded capacity. When all lots were aggregated for each region, the authors obtained good model fits for these variables, explaining 93% of the variance. For individual lots, errors with the training data (e.g., one compares the forecast to observed value for a lot used to build the model) ranged from a 198% underprediction to a 325% overprediction; when framed as absolute values, the average error (computed by the research team) was 43%. The authors concluded that if located properly in congested corridors, park and ride lots could capture "perhaps as much as 2.5% to 3.0" of the catchment area population.

Peng and Mohamad (2005) developed a model for forecasting demand at 12 light rail transit lots in Kuala Lumpur as a function of four variables: daily passengers using the light rail service, mean parking time in hours, whether there is a fee for parking, and whether there is feeder bus service. The authors noted that although all four variables could be forced into the model, stepwise regression showed that only two of the variables were significant (p = 0.05 or below): the number of passengers using light rail and whether there is a fee; this model explained 99% of the variation in the 12 lots. The stepwise model reduced the standard error of the chosen variables. For example, when the initial regression model was developed with all four variables, the binary variable for parking fee had a standard error of 55.8 and was not significant (p = 0.22). The stepwise method reduced this standard error to 33.0 and rendered it significant (p = 0.01). Thus, Peng and Mohamad (2005) articulated the value of judicious variable selection when building models.

Spillar (1997) developed five models that forecast demand for King County lots (Seattle area); the models yielded adjusted R-squared (R^2) values between 0.40 and 0.68 and were based on up to 31 lots; some of the models excluded the 9 lots for which demand exceeded capacity. Then, the authors performed an additional step not undertaken in any other studies reviewed by the research team: the authors assessed not the training error but rather the testing error—that is, the difference between a forecast value and observed value for a lot that was not used to build the initial model. The authors initially found that the models were not directly transferrable: the models (based on Seattle) tended to overestimate demand when applied to lots in Denver by a factor from 1.7 to 2.9 depending on the model. Based on this discrepancy the authors then determined a correction factor for each model to make it transferable to Denver and then reapplied the corrected Seattle models. For the best Seattle model (which had an adjusted R^2 of 0.68), after the correction factor was applied, the average testing error (174) was 1.07 times the average occupancy (162) for the 11 Denver lots where the model was tested. For a Seattle model with a lower adjusted R^2 of 0.45, testing on 12 Denver lots (the original 11 plus another) yielded an average testing error of 93, which was 0.47 times the average occupancy for the 12 lots (195).

Vincent (2007) also developed forecasts for park and ride lots served by rail in New Zealand, where the author employed binary variables to represent variation in demand by line. It should be noted that the combination of the intercept and the binary variable divided by the average occupancy was substantial; for example, for the "Western" line, the intercept (81.427) was about one-third of the average occupancy of 225.6 as tabulated by the research team; the ratio was smaller for one line (0.07) and larger for the other line (0.80). The model explained 96% of the variance, with key variables being the presence of express service, whether the station is patrolled, advertising, population, presence of a nearby lot (which would draw users away from an existing lot), and the fare; generally, the ratio of error based on the training data to the average occupancy was between 0.13 and 0.17 for the three lines. Of interest was that certain variables did not influence demand, notably, proximity to state highways and lighting, and that other variables should be removed because they had the wrong sign (e.g., a paved lot should increase occupancy relative to an unpaved lot, but this variable was removed as it had the wrong sign).

Other Analytical Approaches

Forecasting approaches may also be based on regional travel demand models; for example, FDOT (2012) showed how the share of home-based work auto trips can be used to estimate park and ride lot demand in urban centers. Virginia has also used regional models to forecast park and ride lot demand: in the I-395 corridor, the I-95/I-395 Transit/TDM Technical Advisory Committee (2008) used the Northern Virginia regional model, which had provided mode choice estimates for trips using the "drive access to transit" mode (as well as the "drive access to HOV" mode); these were post-processed to estimate demand. That same source reported the need to make some site adjustments; for example, the Horner Road Park and Ride Lot's forecast demand was increased by one-fifth to account for unique factors at that location. Vincent (2007) also used a regional model that was suitable for determining aggregate demand for park and ride lots by "sector" (e.g., groups of lots by geographic location), as opposed to individual lots; in that endeavor, cost and time data by mode were used.

Forecasts may also use sketch planning methods; for example, Jacobs (2010) forecast demand for park and ride lots that serve carpools based on existing use plus expected growth in interchange traffic volume and a further increase of 14% if high occupancy vehicle (HOV) lanes were in proximity of the lot. VHB Engineering, NC (2013) reported the use of a constant growth factor of 1.5% for existing park and ride lots in the Chapel Hill, North Carolina, area for the period 2011-2035. (Then, anticipated changes in demand from anticipated land development or better transit service may be used to adjust the forecast.)

Off-the-shelf models also exist for fixed guideway facilities. Cherrington et al. (2017) described such models, which may be appropriate for the Virginia Railway Express (VRE) and Washington Metropolitan Area Transit Authority (WMATA) lots in the Northern Virginia District; for example, one model (which the authors noted is detailed in Transit Cooperative Research Program Report 153) can be used to determine how the price of parking, bus service, bicycle access, pedestrian access, and the presence of transit-oriented development affect parking demand. For the purposes of this report, WMATA is referred to by its common name, i.e., "Metro."

Lessons Learned From Forecasting

The literature also suggests five key themes that should be considered in forecasting demand.

1. There is not necessarily a uniform catchment area.
2. There may be multiple market segments.
3. Socioeconomic variables should be included in the model.
4. The impact of amenities on demand is unknown.
5. Models are not usually directly transferrable.

There Is Not Necessarily a Uniform Catchment Area

Cherrington et al. (2017) reported that the sizes of catchment areas vary—an observation borne out by other literature. Beattie (2014) demonstrated the importance of considering variable catchment areas: surveys of bicyclists and pedestrians who parked at facilities serving shared use paths located in New York's Hudson Valley and "other parts of the country" showed catchment areas from 2 to 25 miles. In more urban locations, catchment areas can be smaller. Based on surveys of users of 35 lots in the San Francisco Bay Area (in California), Shirgoakar and Deakin (2005) found that more than two-thirds of lot users (70%) resided within 10 minutes of the lot, noting that this percentage was lower (48%) for users of the 3 lots associated with the heavy rail system (Bay Area Rapid Transit). Nelson\Nygaard Consulting Associates Inc. (2012) reported that previous work indicated that one-half of demand lived within 2.5 miles of the lot but also, in a footnote describing this value, stated that this value was "conservative" because surveys indicated that the average user would have a one-way driving distance of 6.2 miles. Nordstrom and Christiansen (1981) defined "catchment area" as a parabola, with sizes varying by location—e.g., 5 x 6 miles in Austin but 7 x 8 miles in Houston, with the larger number indicating the width of the parabola perpendicular to the direction of the central business district (CBD). Vincent (2007) used customized catchment areas when modeling demand for each park and ride lot: in New Zealand, the author used origin-destination data to determine the total lot patronage and then drew the catchment area for each lot at one-half of this patronage such that there was not a uniform radius for each lot.

There May Be Multiple Market Segments

Mouskos et al. (2007) suggested that when considering park and ride facilities that offer transit service, one could consider different market segments, such as individual drivers who take transit and individual drivers who may then take a common vehicle to a destination; by extension, one would expect attributes such as the presence of transit and the use of a carpool to have different influences on occupancy. The authors also pointed out that initial results from their model (based on travel time to the lot, overall travel time, cost, and fare) yielded "higher estimates" of demand at each facility owing to challenges in representing the network. The authors did not calibrate a new model but rather borrowed parameters from the literature to forecast park and ride lot demand via a nested logit model where users were placed into one of three modes: drive alone, use the lot then take transit, or walk to the lot then take transit.

Socioeconomic Variables Should Be Included in the Model

Cheu et al. (2012) developed a binary logit model from survey responses that sought to determine existing auto drivers' willingness to use a future park and ride facility where four variables increased stated willingness to use such a lot. Two variables were significant, with *p*-values below 0.05: (1) being both 24 years or younger and having an annual income of less than $25,000 ($p = 0.04$), and (2) having fewer cars in the household in increments of 0 to 5 ($p = 0.02$). In addition to the intercept, two other variables were included in the model: (1) not being in a two-person household ($p = 0.18$), and (2) having a longer commute time in increments of 0 to 9 minutes, 10 to 19 minutes, 20 to 34 minutes, and 35+ minutes ($p = 0.23$). This work was also reported in a shorter report (Cornejo et al., 2014). Elsewhere, demand has been shown to be relatively inelastic to cost: Desman Associates (2012) suggested that a 1% increase in lot fees for lots serving rail facilities would reduce parking demand—but by a much lower figure of 0.08%.

Southwest Region Planning Commission et al. (2015) developed a 15-step approach for forecasting demand at future park and ride lots based on a 2.5-mile parabolic area (with the vertex closest to the CBD); the percent of nearby 2-, 3-, 4-, and 5-person carpools; origin-CBD trips obtained from the "OnTheMap" application (U.S. Census Bureau, 2020b); and an assumption that one-fourth of all carpoolers will use the park and ride lot. Trip length appears to be a factor in urban areas: Shirgoakar and Deakin (2005) found that "most" commuters made long trips, with a travel time of 53 minutes, which the authors noted suggests a trip length of at least 30 miles; users of the heavy rail system (Bay Area Rapid Transit) reported a higher door-to-door travel time of 59 minutes. The Metropolitan Council (2010) developed forecasts for park and ride lots in the Minneapolis area based on population and employment forecasts, commuting data, and usage at existing lots; the authors reported that for the two lots for which they performed model validation, forecast and observed use were similar (within 1%).

The Impact of Amenities on Demand Is Unknown

Amenities, such as lighting, real-time information at the lot, cleanliness, and other factors that indicate the quality of the lot, may affect demand substantially or relatively little. Literature can be found to support both viewpoints:

- *Some sources reported that amenities are important*. Bos et al. (2004) used simulation results with an existing mode choice model for Nijmegen (in the Netherlands) to find that the provision of traffic information for a moderately congested corridor could increase park and ride lot occupancy by 1.4% to 6.4% if the lot offered a public transportation alternative, although the authors noted that the "quality" of the facility (e.g., maintenance and presence of staff) could affect this figure. The authors also noted that effective signage, availability of food and reading material, travel time information, and proximity to laundromats and shopping have improved lot usage (Bos et al., 2004). Shirgoakar and Deakin (2005) reported that after some discussion where fees were initially negatively perceived, focus groups composed of California users indicated their willingness to pay fees for certain amenities (e.g., "$1 to $2 for security, lighting, shelters, toilets, and so on"). The authors noted that other concerns included capacity, cleanliness, and transit service quality, including adherence to the schedule.

- *Other sources reported that amenities have at best a modest impact.* Rathbone (2006) stated that "amenities" did not materially affect the success of a park and ride lot: responses from 30 public transportation agencies who were asked for characteristics that defined lot success indicated proximity to roadways, frequent (and fast) transit service, and proximity to residential areas. The same survey results noted characteristics that improved some individual lots, such as capacity, presence of amenities, and whether parking was charged, but the author emphasized these were lesser than the role of "total travel time and convenience." Similar results were noted in the Northern Virginia District except that one amenity (lighting) was found to matter. BMI et al. (2003) assessed the effect of nine factors on the ratio of demand to capacity at park and ride lots in VDOT's Northern Virginia District: ownership; lighting; a phone being available; existence of bus service; bus shelters; connection to a bicycle route; the availability of "bike racks, bike lockers, or other amenities"; sidewalks; and being within one-fourth mile of an HOV lane. Based on the Wilcoxon rank test and a significance threshold of $p = 0.05$, only three of the nine attributes affected the ratio of demand to capacity: bus service, HOV lanes within one-fourth mile, and lighting.

Models Are Not Usually Directly Transferrable

The findings of Nungesser and Ledbetter (1987) and Spillar (1997) explained the difficulty of transferring models directly. The research team was able to test the latter approach because of its similarity to data available in VDOT's Northern Virginia District. Spillar (1997) had considered eight variables in the development of models for King County: the seven variables shown in Table 1 and an eighth variable that accounted for the presence of midday transit service at the lot. The research team then compared the correlation matrix between the independent variables that Spillar (1997) reported to the correlation matrix between similar independent variables used in the Northern Virginia District. In most cases, the results differed substantially; for instance, a fairly strong correlation between buses and population was noted for the Seattle data (0.614) compared to a fairly weak correlation in the Northern Virginia District (-0.083); by contrast, Virginia had a strong correlation between the number of adjacent lots within a 50% market area of the lot and the population (0.846), whereas Seattle had a weak correlation between these two variables (0.107).

Implications of the Literature Review

The literature helped identify candidate variables; for example, the aforementioned work by Rathbone (2006) suggested that one could consider several variables: access to highways, express lanes, or transit; presence of severe congestion; distance to homes; and distance to the CBD, the cost of parking, and safety. The fact that some variables were cited by multiple sources also led to their initial inclusion in the models; proximity to roadways was noted both in a comprehensive literature review by CTC and Associates (2010) and Rathbone (2006).

Table 1. Comparison of Correlation Coefficients for Seattle (top line of each row) and for Park and Ride Lots in the Northern Virginia District (bottom line of each row)

	Buses	Freeway	Adjacent Lots	Population	CarCost / TranCost	DistToCBD	Transit Speed
Buses[a]	1.000 1.000						
Freeway[b]	0.238 -0.206	1.000 1.000					
Adjacent Lots[c]	0.156 -0.016	0.254 -0.442	1.000 1.000				
Population[d]	0.614 -0.083	0.013 -0.404	0.107 0.846	1.000 1.000			
CarCost / TranCost[e]	0.115 -0.498	-0.096 0.474	-0.211 -0.372	-0.034 -0.235	1.000 1.000		
DistToCBD[f]	-0.142 -0.566	-0.011 0.460	-0.309 -0.238	-0.331 -0.085	0.822 0.916	1.000 1.000	
Transit Speed[g]	0.083 -0.047	0.080 0.287	-0.160 0.073	-0.058 0.042	0.796 -0.046	0.649 -0.042	1.000 1.000

[a] Number of AM peak buses whose destination is the central business district (CBD).
[b] Boolean variable to incorporate proximity to freeway (straight-line distance in miles from the lot to the nearest interstate access point).
[c] Number of adjacent lots observed in the 50th-percentile market area of the lot (the number of independent park and ride lots that are within the 2.5-mile radius of the designated park and ride lot).
[d] Total population in the 50th-percentile market area of the lot (the sum of the population for the portions of the block group that fell within a 2.5-mile buffer of the designated park and ride lot.
[e] Ratio of auto operating costs to transit costs (the time of auto operating divided by the time of transit).
[f] Straight-line distance between the lot and the CBD.
[g] Fastest transit schedule time between the lot and the CBD divided by the straight-line distance between the lot and the CBD where transit times were estimated using Google Maps.

RSG (2015) suggested that commuting patterns could be used; the authors examined origins and destinations based on longitudinal employment household dynamics data. A similar suggestion from the TRP led the research team to use the Census-based "OnTheMap" data (U.S. Census Bureau, 2020b). Thus, a large number of variables (i.e., 78, as shown in the Appendix) were considered in the development of Virginia-specific models.

The aforementioned studies provided a basis for understanding the types of variables that could influence demand; they also suggested two enhancements that further work might add to the state of the practice. One enhancement is informative: To what extent, if any, do additional amenities beyond time, cost, and convenience influence demand? The work by Rathbone (2006) and the earlier VDOT work (BMI et al., 2003) suggested that the answer is relatively little, although Bos et al. (2004) and Shirgoakar and Deakin (2005) suggested that amenities could matter in specific situations. Vincent (2007) suggested that safety was substantial, where patrolling a lot could increase occupancy by 64 (compared to an average occupancy of 185) for stations at a rail line in New Zealand—but the same study found that lighting did not have an impact, in contrast to the study by BMI et al. (2003), which found that lighting was the sole lot amenity that affected occupancy at park and ride lots in the Northern Virginia District (the other variables were bus service and HOV proximity). Answering this question might require the use of different models for different lots, just as Cherrington et al. (2017) and Beattie (2014) had reported when considering catchment areas.

The other enhancement is empirical: When one develops a model based on a particular dataset, what is the expected accuracy when this model is applied to a different dataset? The research team is aware of only one study that performed such an analysis: the King County work by Spillar (1997).

PURPOSE AND SCOPE

The purpose of this study was to develop models for forecasting occupancy (the number of used spaces) at Virginia park and ride lots. The study had three objectives: (1) to identify the variables that explain variation in occupancy at Virginia lots, (2) to quantify the accuracy of these models when applied to a dataset different from the one on which the models were developed; and (3) to demonstrate how to implement these models to forecast demand at both proposed lots and existing lots where some key attribute is expected to change in the future.

The scope of this study was limited in two ways. First, the models were based on 2018 lot occupancy and capacities provided by the VDOT Transportation Mobility Planning Division. These are collected every 1 to 2 years and represent a single site visit on one particular day such that weekly or daily variation is not captured. Second, candidate variables were constrained to publicly available datasets describing four sets of attributes: traffic, the facility, land use, and the social and economic composition of the population near the lot. As shown in the Appendix, traffic data represent volumes, indicators of congestion, and peaking (e.g., how the peak hour volume compares to the 24-hour volume). Land use data represent the location of the lot relative to employment sites, interstates, and other lots and characteristics of commuters: How far do they live from the lot, how far do they travel to work, and what mode of transportation do they use? Socioeconomic information represents characteristics of people near the lot (population, income, relative share of income spent on rent) and number of jobs near the lot. Facility attributes describe the lot itself, such as signing, lighting, the cost to park at the lot, and transit service. The 78 variables in the Appendix do not include disaggregate customer information, such as the income of user x, whether user y thinks amenities such as lighting are critical to the lot's attractiveness, or the factors that would cause nonuser z to use the lot.

METHODS

Four major tasks were undertaken to accomplish the study objectives:

1. Determine candidate variables.
2. Develop successively complex models.
3. Evaluate models.
4. Repeat analysis based on guidance from the TRP.

Determine Candidate Variables

Through several discussions with the TRP, 79 candidate variables related to each park and ride lot were identified, and supporting data were obtained. These variables were categorized as being related to the facility, traffic, land use, and demographics.

Facility Variables

The TRP provided a database of facility attributes for each park and ride lot such as the lot name, latitude and longitude, total number of parking spaces, and number of occupied parking spaces. Seven key facility-related variables were directly extracted from this database; the variable name is shown in brackets:

1. Whether transit service is provided to the lot, including cases where passengers might need to cross a street from the lot to reach the transit stop [TransitServiceAvailable where a value of 1 means yes].

2. Whether overnight parking is allowed [OvernightParkingAllowed, 1 means yes].

3. Number of transit lines serving the lot [NuofTranServicePP]. This includes bus, light rail, and heavy rail but does not incorporate frequency of service or ridership.

4. Whether the parking lot is lit [Lighting, 1 means yes].

5. Condition of the sign [SignCondition = 0 if no sign; 1 if poor; 2 if fair; 3 if good; 4 if very good].

6. Whether there is a fee to park in the lot [CostToPark, 1 means yes].

7. Whether the parking lot has covered bicycle parking facilities [BikeParkingisCovered where 1 means yes and 0 means no].

As noted in Table A6 in the Appendix, an additional 10 variables were considered but generally not used in the final modeling because they did not show promise—with one exception: the number of bicycle parking spaces was found to be useful in the Northern Virginia District model.

For 296 of the 297 lots, the VDOT database provided occupancy information. For 1 of the 297 lots (i.e., the Wiehle-Reston East Metro Station Garage), that database showed a value of 0, which the TRP noted was incorrect, especially given the lot's capacity of 2,300. For that lot only, another source was used: a web application developed by WMATA (2020) gave the "average daily parking transactions" for 2019 where Saturdays, Sundays, and holidays are excluded.

Traffic Variables

Traffic variables were in some cases provided by the TRP and in other cases extracted from VDOT's Pathways for Planning application (VDOT, 2019b, 2019c, 2019d, 2020a) and usually processed using ArcGIS Pro software (Version 2.4). In discussions with the TRP, a 2.5-mile radius emanating from the park and ride lot was chosen for obtaining 11 of these 12 traffic variables (Figure 2) given that previous work, such as that by Spillar (1997) and Nelson\Nygaard Consulting Associates Inc. (2012), had reported that that this ratio represented a substantial portion of the lot's users, and by extension, that one would expect traffic characteristics within this radius to be affected by park and ride lot demand. The TRP suggested that this radius might not be appropriate, however, for non-traffic variables, and thus other radii, as discussed later, were used for those other variables.

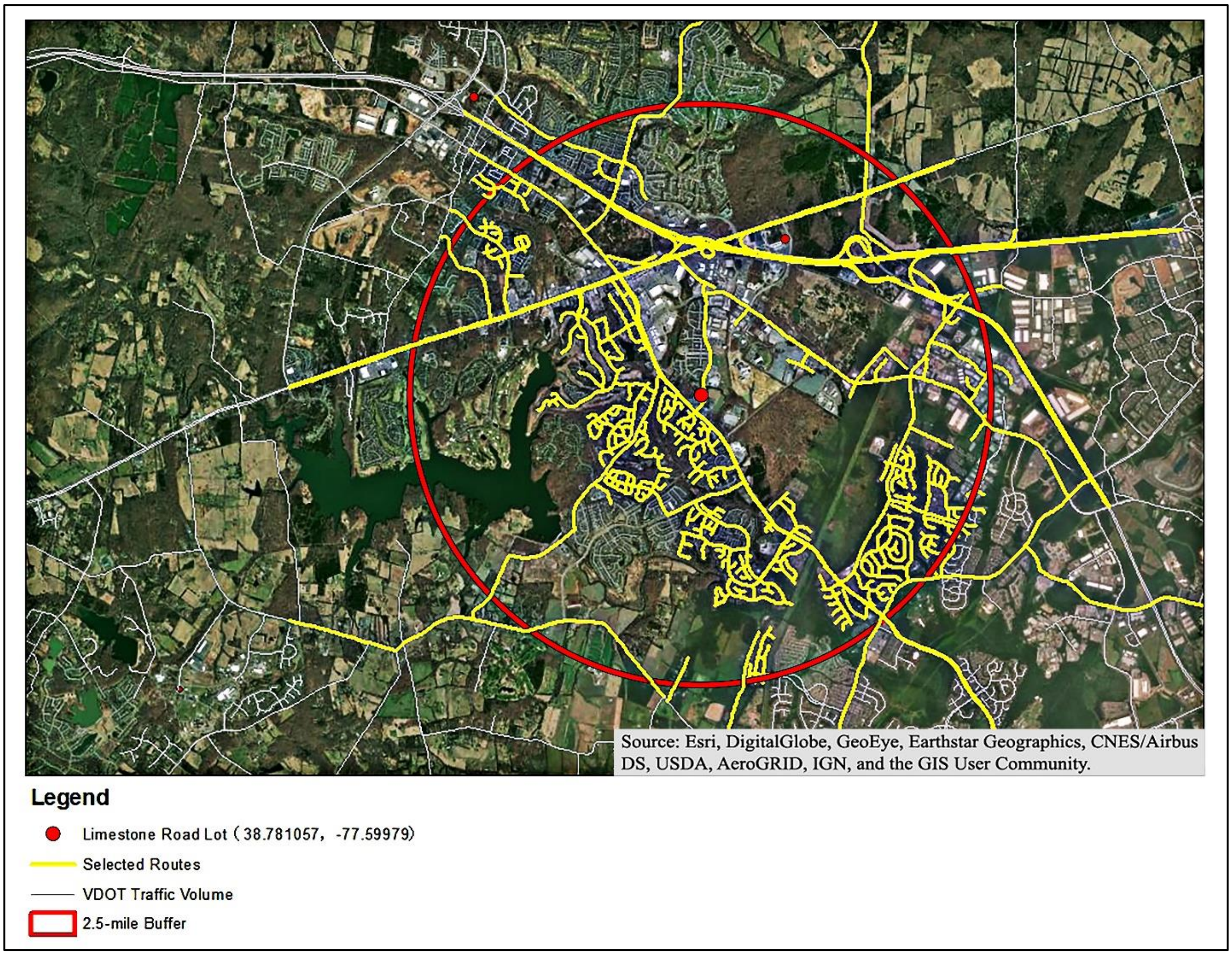

Figure 2. Example of Capturing Traffic Data With a 2.5-mile Radius Catchment Area

Twelve traffic-related variables were collected (variables are shown in brackets):

1. Mean value of 2018 ADT for roads within 2.5 miles of the lot [Average ADT].

2. Highest 2018 ADT for all roads within 2.5 miles of the lot [MAX ADT].

3. Sum of all 2018 ADT for roads within 2.5 miles of the lot [Sum ADT].

~~4.~~ 2018 ADT on the road providing a direct entrance to the lot. If there are two or more entrances, then this variable is the average of the ADTs [Closest ADT].

5. Mean ratio of 2018 volume (v) to capacity (c) for roads within 2.5 miles of the lot [V/C].

6. Mean 2018 peak hour factor (PHF) of roads within 2.5 miles of the lot [$PHF_{average}$].

7. PHF for the closest road and for the road within 2.5 miles with the highest ADT [$PHF_{closest}$, PHF_{max}].

8. Ratio of the peak hour volume to the ADT, commonly known as the K-factor, for the closest road, for the road within 2.5 miles with the highest ADT, and the mean value for all roads within 2.5 miles [$K_{closest}$, K_{max}, $K_{average}$].

9. Peak hour expansion factor (PHEF), provided by Mobayed (2020a), a surrogate for regional congestion indicating the length of time that the "peak period" (VDOT, 2020) lasts for interstates and arterial facilities. PHEF is determined by capturing the closest link with a PHEF value (regardless of the distance from the lot), given that PHEF is available for only some links in contrast to other variables [PHEF]. Thus, unlike the previous 11 traffic variables, PHEF was not related to the 2.5-mile radius.

Garber and Hoel (2009) distinguished between annual average daily traffic (AADT) and ADT in that that former is based on a continuous count station where volumes are sampled 24 hours per day 365 days per year, and the latter is based on a shorter-term count, such as a 48-hour count. In practice, however, as VDOT has very few continuous count stations, most annual traffic volumes reported by VDOT as AADT are ADTs that were then converted to an annual estimate based on the appropriate seasonal adjustment factors. This report uses the term *ADT* to denote a typical 24-hour volume. A map of 24-hour traffic count data (ADT) is available in the subsection of VDOT's Pathways for Planning application titled Transportation Planning Data (VDOT, 2019c).

Land Use Variables

Land use data were extracted from the American Community Survey (ACS) (Mobayed, 2020b; U.S. Census Bureau, 2020a); ancillary Census mapping applications (U.S. Census Bureau, 2020b); and employment centers used for VTrans, Virginia's Transportation Plan (CDM Smith, 2020). Most land use variables were collected 3 times for each lot, reflecting characteristics within a 2.5-mile radius, a 5-mile radius, and a 10-mile radius of the lot (see Figure 3). One reason for considering these different radii is that catchment areas may not be uniform—the 2.5 mile value is a good starting point, but others have defined different sizes (Nordstrom and Christiansen, 1981; Vincent, 2007).

Some data, such as Census blocks, block groups, and tracts, are in the form of polygons where a single block group might be partially, but not entirely, within the catchment area (see Figure 4). Additional GIS processing was performed to ensure that the estimated population outside the catchment area was not used. For example, if two tracts have 30 carpoolers and 100 carpoolers and the catchment area includes one-half of the first tract and one-third of the second tract, the number of carpoolers is estimated as 48 based on these two tracts (see Eq. 1).

$$30 \text{ Carpoolers } (1/2) + 100 \text{ Carpoolers } (1/3) = 48 \qquad \text{[Eq. 1]}$$

Thirty-four land use variables were collected (variables are shown in brackets):

1. Distance in miles to the next lot [DTNearestP].

2. Number of independent lots within the 2.5-mile buffer of the lot [NuofAdjLot].

3. The straight-line distance in miles from the lot to the nearest interstate access point [ProximityToIAP] (Mobayed, 2020c).

4. The straight-line distance in miles from the lot to the nearest express lanes [ProxToEL] (Mobayed, 2020d; 2020e). The straight-line distance is a relative indicator. For example, for the three park and ride lots shown in Figure 5, the straight-line distances to the Express Lanes access point (at the right of the figure) are all shorter than the travel path motorists will follow. In this case, although the straight-line distance from the Dunn Loring-Merrifield Station Metro Park and Ride lot to the access point for the express lanes is 2.3 miles, a motorist leaving that lot and seeking to access the express lanes will travel a longer distance, heading east on I-66 and then heading south on I-495. In practice, therefore, the variable "straight-line distance in miles from the lot to the nearest Express Lanes" is a surrogate for proximity, conveying the critical information that Dunn-Loring is considerably closer to the Express Lanes access point than Kutner Park.

5. The average of the area-weighted median commute times for all Census tracts within the catchment area, where for radii of 2.5, 5.0, and 10.0 miles the variables are [CommuteTime2], [CommuteTime5], and [CommuteTime10].

6. The average number of work trip carpoolers from all Census tracts within the catchment area for sizes of 2.5 miles ([Carpoolers2]), 5 miles ([Carpoolers5]), and 10 miles ([Carpoolers10]).

7. The average number of transit riders from all Census tracts within the catchment area [TransitRiders2], [TransitRiders5], [TransitRiders10].

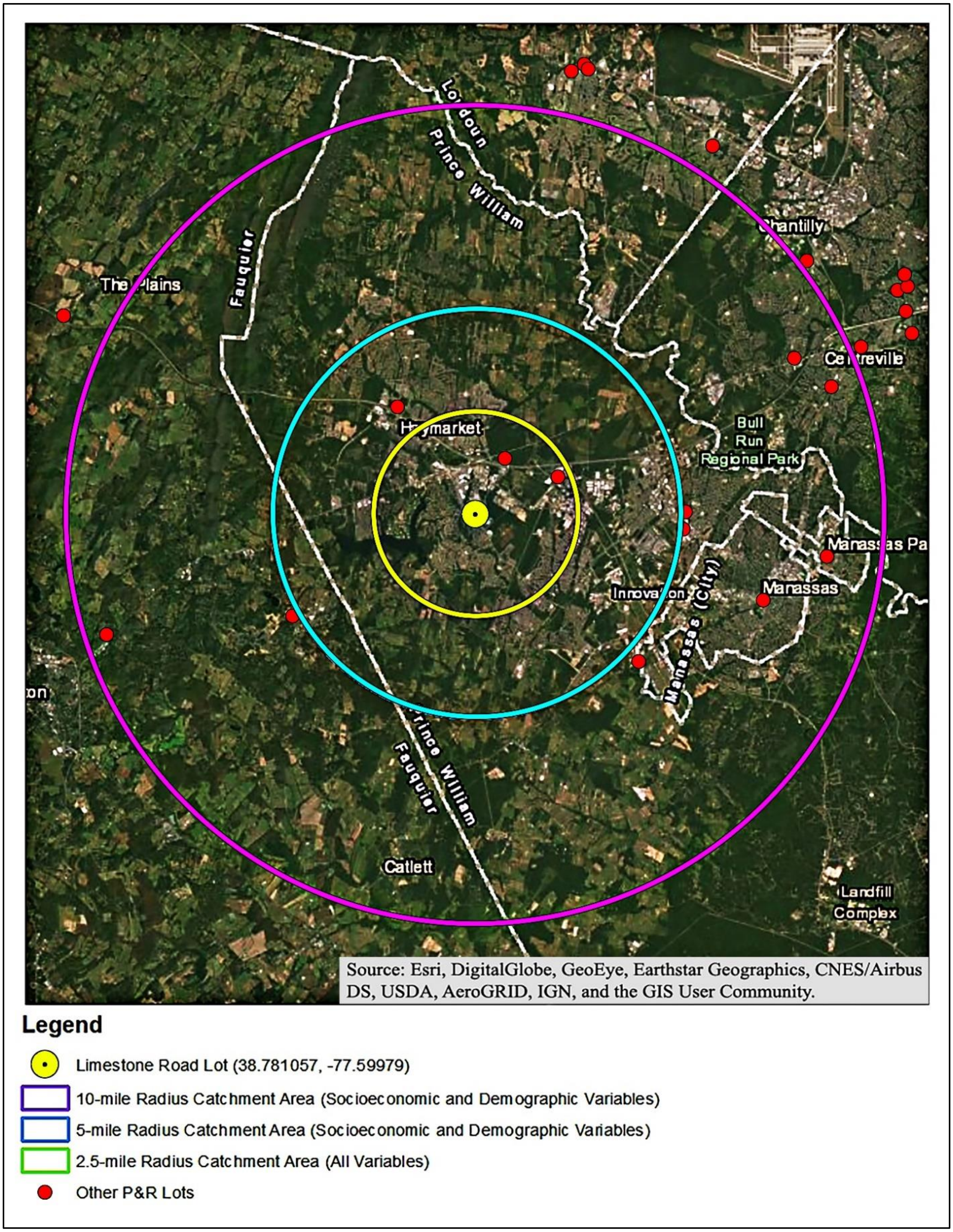

Figure 3. Example of Three Catchment Areas for the Limestone Road Park and Ride Lot

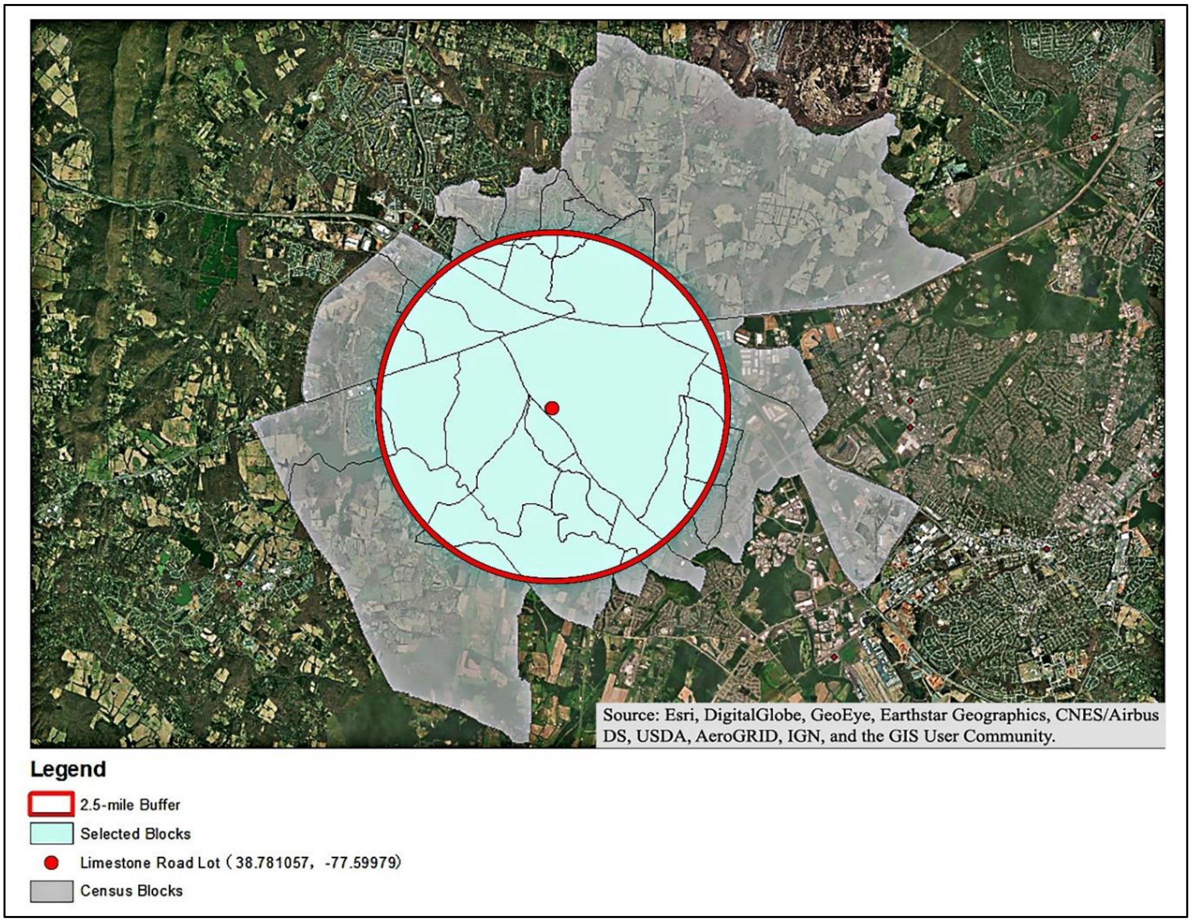

Figure 4. Example of Capturing Population With a 2.5-mile Radius Catchment Area

8. Distance to the nearest block group with at least 5,000 employees [DIST_BG_BigEmp].

9. Distances to the four nearest major employment centers, defined as being 1 of 379 VTrans employment centers (CDM Smith, 2020) and being within a block group of at least 10,000 employees [Dist_M1], [Dist_M2], [Dist_M3], [Dist_M4].

10. The sum of the square root of the distances to the four closest employment centers [Dist_Weight].

11. Number of commuters within the catchment area having jobs, jobs less than 10 miles away, jobs between 10 and 24 miles away, jobs between 25 and 50 miles away, and jobs more than 50 miles away. Such data can be derived from the Census mapping application OnTheMap (U.S. Census Bureau, 2020b) with some additional GIS processing. For example, Table 2 shows that there are 69,902 commuters (based on 2017 data) within 2.5 miles of the Dumbarton Oaks Park and Ride Lot. About one-half (33,925) have a commute of less than 10 miles, and slightly less than one-fifth (14,468) have a commute of more than 50 miles.

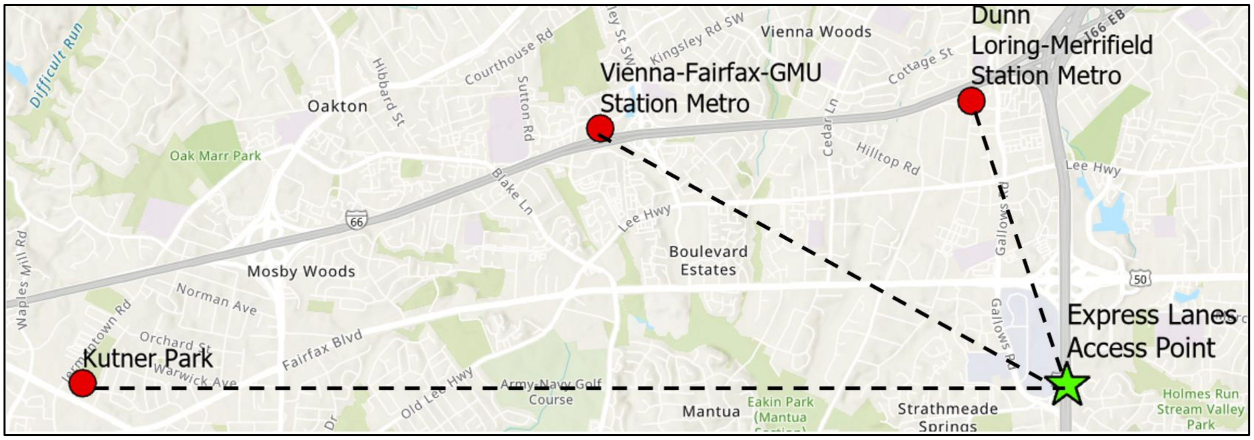

Figure 5. For the Variable "ProxToEL" (Straight-Line Distance in Miles From the Lot to the Nearest Express Lanes), Dunn Loring Has a Lower Value Than Kutner Park

Table 2. Number of Commuters With Jobs Within Various Radii of the Dumbarton Oaks Park and Ride Lot[a]

Radius (miles)	Less Than 10 Miles	10-24 Miles	25-50 Miles	More Than 50 Miles	Total
2.5	33,925	16,845	4,664	14,468	69,902
5	75,681	36,903	10,061	30,118	152,763
10	195,570	112,061	26,173	75,862	409,666

[a] [Rad2JobsTot], [Rad2JobsLT10Mi], [Rad2Jobs10to24Mi], [Rad2Jobs25to50Mi], [Rad2JobsGT50Mi], [Rad5JobsTot], [Rad5JobsLT10Mi], [Rad5Jobs10to24Mi], [Rad5Jobs25to50Mi], [Rad5JobsGT50Mi], [Rad10JobsTot], [Rad10JobsLT10Mi], [Rad10Jobs10to24Mi], [Rad10Jobs25to50Mi], [Rad10JobsGT50Mi].

Demographic Variables

As was the case with land use variables, demographic variables were developed for catchment radii of 2.5, 5.0, and 10.0 miles. Extensive GIS processing was often used to capture better each variable with respect to the park and ride lot. For example, if the area within 2.5 miles of a park and ride lot had just two tracts of 50,000 square feet and 100,000 square feet and the commute time in the former tract had an average value of 30 minutes and the latter had an average value of 40 minutes, the weighted commute time is about 36.7 minutes (Eq. 2).

$$(50{,}000 \text{ ft}^2 * 30 \text{ min} + 100{,}000 \text{ ft}^2 * 40 \text{ min})/(50{,}000 \text{ ft}^2 + 100{,}000 \text{ ft}^2) = 36.7 \qquad \text{[Eq. 2]}$$

Twenty-five demographic variables were collected (variables are shown in brackets):

- Population provided by Ling (2018) [POP2.5], [POP5], [POP10].
- Population density computed as people per square mile [PopDen] for the block group in which the lot is located.
- Employment (e.g., jobs) in the catchment area (Ling, 2018) [EMP2.5], [EMP5], [EMP10].

- The median monthly rent multiplied by 12 and divided by median household income (Mobayed, 2020b, U.S. Census Bureau, 2020a): [RentOverAllIncome2], [RentOverAllIncome5], [RentOverAllIncome10].

- Average percent of renters' household income spent on rent: [AvgPctOfRentIncomeOnRent2], [AvgPctOfRentIncomeOnRent5], [AvgPctOfRentIncomeOnRent10].

- Sum of minority, poverty, limited English population (LEP), and disabled populations from all Census tracts within the buffered area (Ling, 2018). With radii of 2.5, 5.0, and 10.0 miles, the 12 variables are [MinorityPop2], [PovertyPOP2], [LEPPop2], [EligDisadvPop2], [MinorityPop5], [PovertyPop5], [LEPPop5], [EligDisadvPop5], [MinorityPop10], [PovertyPop10], [LEPPop10], and [EligDisadvPop10]. VDOT asked that these variables not be used in the models, however, if at all possible, explaining that their inclusion might exacerbate inequity. For example, if a model showed that a higher value of LEP was associated with reduced demand, such a model could lead to VDOT not building lots in locations where such populations were high.

Develop Successively Complex Models

Five categories of models were developed:

1. existing diversion model
2. recalibrated diversion models
3. generalized ADT models
4. linear regression models
5. nonlinear regression models.

Existing Diversion Model

The initial approach for estimating occupancy was based almost entirely on the approach used by FDOT (2012) with one exception: FDOT (2012) noted that a diversion factor of 1% should be used for "secondary" facilities. To make this application consistent at the 297 lots in Virginia, the research team initially defined such "secondary" facilities as being represented by the variable Vprime_K, which is the peak period traffic on the facility having the highest ADT within 2.5 miles of the lot. Generally, the model assumes no change in travel patterns: the only users are those who pass the lot during the normal travel path, such that a road with no volume will yield a park and ride lot that has no occupancy. The model is defined in Equation 3:

$$\text{Occupancy} = a\,(\text{Vpeak}_K) + b\,(\text{Vprime}_K) \qquad \text{[Eq. 3]}$$

where

Vpeak_K = total peak-period traffic on the adjacent facility or facilities, where adjacent means there is an entrance to the lot

$Vprime_K$ = peak period traffic on the prime facility, which is the facility within the 2.5-mile radius that has the highest ADT

a, *b* = diversion factors of 0.03 and 0.01 for peak and prime traffic, respectively

$Vpeak_K$ = ADTpeak * K * D * Design period

$Vprime_K$ = ADTprime * K * D * Design period

ADTpeak = two-way ADT for the adjacent roadway facility

K = peak hour percentage

D = peak hour directional distribution of traffic

Design period = design period, the pronounced peak traffic period

ADTprime = two-way ADT for the prime roadway facility.

FDOT (2012) uses typical K-values (see Table 3), whereas the research team used facility-specific K-values (VDOT, 2019d). Although the aforementioned source is named for V/C ratios, it is a geographic layer that contains K-factors as a separate attribute. For the design period, Table 4 provides lookup values, where a higher traffic volume is correlated with a longer peak period. FDOT uses values of a = 0.03 and b = 0.01, representing a capture of 3% on adjacent facilities and a capture of 1% for the largest facility within 2.5 miles of the lot, where a and b were based on large cities in Texas (FDOT, 2012). FDOT also suggested that default values for the design period, K, and D may be used in lieu of site-specific data (see Tables 3 and 4) if such data are not available. Although the PHFs shown in Table 3 were not needed for the existing diversion model, they are provided here as they were useful for modifying that model later.

Table 3. Typical K-Factors, D-Factors, and PHFs

Roadway Class	K	D	PHF
Urban Freeway/Expressway	0.092	0.52	0.95
Urban Major and Minor Arterials	0.097	0.52	0.95
Urban Multi-Lane Highways	0.094	0.52	0.92
Transitioning Freeway/Expressway	0.094	0.52	0.92
Transitioning Major and Minor Arterials	0.097	0.52	0.88
Transitioning Multi-Lane Highways	0.097	0.52	0.88
Rural Freeway/Expressway	0.103	0.52	0.92
Rural Major and Minor Arterials	0.097	0.52	0.88
Rural Multi-Lane Highways	0.097	0.52	0.88

PHF = peak hour factor.

Table 4. Suggested Design Periods

Average Daily Traffic	Design Period
Above 50,000	60 min
35,000-49,999	45 min
Below 35,000	30 min

Recalibrated Diversion Model

The recalibrated diversion model presumes that parking demand is a direct function of the amount of traffic on roadways adjacent to the park and ride lot as was the case with the existing diversion model (FDOT, 2012). However, there are two differences when this is calibrated for Virginia. First, Virginia-specific values of diversion factors *a* and *b* are obtained where the values of *a* and *b* minimize the root mean square error (RMSE) given in Equation 4, where *n* is the sample size. Although the SPSS regression software tool was used to determine *a* and *b*, the research team did not use a criterion of statistical significance; rather, the independent variables V_{peakK} and V_{primeK} were always used and were never rejected, regardless of statistical significance).

$$\mathrm{RMSE} = \sum_{i}^{n} \sqrt{\frac{1}{n}(\mathrm{Y_i\ observed} - \mathrm{Y_i\ forecast})^2} \qquad \text{[Eq. 4]}$$

Second, two different approaches for estimating the peak hour volume were tested: one used the ADT multiplied by the K-factor, and one used the ADT multiplied by the PHF. Since the latter is the hourly volume during the maximum-volume hour of the day divided by the peak 15-minute flow rate within the peak hour, the product of (ADT)(PHF) yields a value that is typically between 88% and 95% of the ADT. Thus, even when there is a large variation in traffic volume during the peak hour, the product of (ADT)(PHF) yields a value that is similar to the ADT. By contrast, the product of (ADT)(K-factor) yields a value that is much smaller than the ADT.

The recalibrated diversion model requires the ADT on the adjacent roadways of the facility (where adjacent roadways each provide an entrance to the lot), the maximum ADT of all roadways within a 2.5-mile buffer, and the PHF and peak hour K-factor for the selected roads. Figure 6 illustrates such data for a hypothetical new lot where the ADT for the roads in red (shown in the large image) will yield the peak variable (assuming the lot will have three entrances) and those roads in white (shown in the small image) will yield the prime variable.

Similar to the existing diversion model, the recalibrated diversion model takes the form of either Equation 5 or Equation 6 where

$$\text{Occupancy} = a\,(\text{Vpeak}_K) + b\,(\text{Vprime}_K) \qquad \text{[Eq. 5]}$$

$$\text{Occupancy} = a\,(\text{Vpeak}_{PHF}) + b\,(\text{Vprime}_{PHF}) \qquad \text{[Eq. 6]}$$

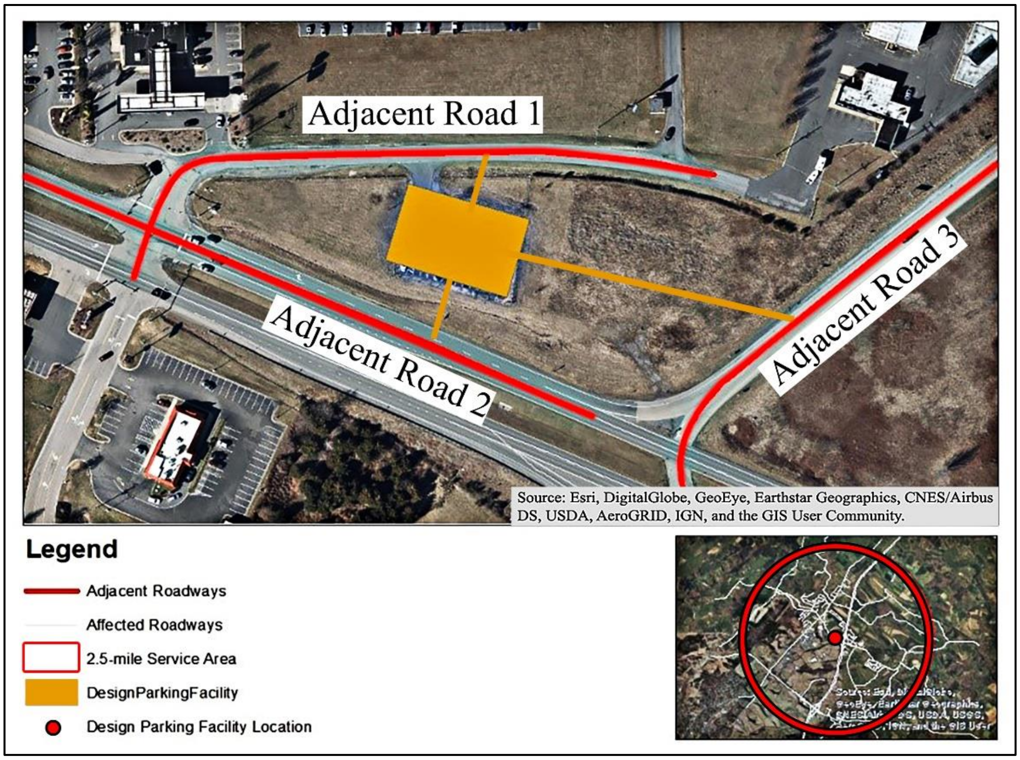

Figure 6. Three Adjacent Roads That Comprise VPeak$_K$ (red) and Candidate Roads From Which the Highest Volume (Vprime$_K$) Will Be Selected (Red or White)

This demand function has no intercept, just as the existing diversion model does not have an intercept. Because demand is a function of ADT, the model in Equations 7 through 10 presumes that an ADT of 0 means the parking lot occupancy should be 0.

$$Vpeak_K = ADT_{adjacent} \times K \times DP \div 60 \quad \text{[Eq. 7]}$$

$$Vpeak_{PHF} = ADT_{adjacent} \times PHF \times DP \div 60 \quad \text{[Eq. 8]}$$

$$Vprime_K = Max\ ADT \times K \times DP \div 60 \quad \text{[Eq. 9]}$$

$$Vprime_{PHF} = Max\ ADT \times PHF \times DP \div 60 \quad \text{[Eq. 10]}$$

where

$Vpeak_K$ = peak period traffic on the adjacent roadway (based on a K-factor)

$Vpeak_{PHF}$ = peak period traffic on the adjacent roadway (based on a PHF)

ADTadjacent = average daily traffic for the adjacent roadway

K = K-factor for the adjacent roadway

PHF = peak hour factor for the adjacent roadway

DP = design period (Table 4)

$Vprime_K$ = peak period traffic on the prime roadway (based on a K-factor)

$Vprime_{PHF}$ = peak period traffic on the prime roadway (based on a PHF)

Max ADT = ADT for the roadway that has the maximum ADT in the 2.5-mile catchment area.

For locations with complete data, the design period is based on values from Table 4, and roadway-specific K-factors and PHFs are determined from VDOT's Pathways for Planning (VDOT, 2019c). For locations with incomplete data, the design period, K-factor, and PHF may be chosen based solely on Tables 3 and 4 (FDOT, 2012).

Table 4 shows a suggested design period where the values were initially obtained from FDOT (2012); however, these values do not necessarily represent a particular peak period. Rather, this design period takes place "during the peak period when a facility experiences the highest traffic flows" (FDOT, 2012); the same source noted that although the values shown in Table 4 may be used, it is also possible to determine the length of this design period from observations of traffic flow. Presumably, therefore, such a design period could be inferred from other data sources; for example, if a roadway's segment planning time index (the 95th percentile travel time divided by the free flow travel time) is extremely high for 45 minutes, then 45 minutes would be the value of the design period. FDOT (2012) clarified that the design period, therefore, is an indication of the length of time during which a roadway is congested:

> The design period concept supports the theory that Park-and-Ride use is related to congestion levels, and is supported by observations showing arrivals at Park-and-Ride facilities during a well-defined time period. This postulates that motorists traveling during times of greatest congestion will have a greater propensity to utilize Park-and-Ride facilities.

Generalized ADT Models

Generalized ADT models are similar to the recalibrated diversion models in that some commuters who use a lot during their normal routes will arrive at these lots from adjacent streets. There are two ways to represent such adjacent traffic: the facility with the highest ADT within a 2.5-mile radius of the lot, and the ADT for the facility or facilities providing direct access to the lot.

However, four key changes were made for these models compared to those in the previous section. First, two other ways of representing traffic were also considered:—the average ADT for all facilities within 2.5 miles of the lot and the sum of these ADTs. Second, assumptions pertaining to the peak hour, such as the K-factor, the PHF, and the design period, were removed in case these extra values were not providing additional information. Third, in recognition that park and ride lot demand might be explained by factors other than traffic demand, an intercept was added. Fourth, a criterion of variables being statistically significant was added: only independent variables with $p \leq 0.05$ were included.

Four types of ADT models were considered (Eqs. 11-14):

$$\text{Occupancy} = \alpha \times \text{Average ADT} + \beta \qquad \text{[Eq. 11]}$$

$$\text{Occupancy} = \alpha \times \text{Sum ADT} + \beta \qquad \text{[Eq. 12]}$$

$$\text{Occupancy} = \alpha \times \text{Max ADT} + \beta \qquad \text{[Eq. 13]}$$

$$\text{Occupancy} = \alpha \times \text{Closest ADT} + \beta \qquad \text{[Eq. 14]}$$

The first three models were based on all roads within 2.5 miles of the facility, and the fourth model was based on the road or roads where the lot has a direct entrance. For example, a lot with two entrances may be considered. For this lot, one entrance is to a facility with an ADT of 1,000 and one entrance is to a facility with an ADT of 1,500. Further, a road with an ADT of 1,300 is located 2 miles away from the lot and there are no other roads within 2.5 miles of the lot. For this lot, therefore, the average ADT is 1,267; the sum ADT is 3,800; and the max ADT is 1,500. Because there are two roads that are closest to the facility (with ADTs of 1,000 and 1,500), the mean of these two values, i.e., 1,250, is chosen as the closest ADT.

Linear Regression Models

Regression methods are generally used to develop models from "unplanned experiments"—i.e., situations where the analyst does not have an ability to alter the independent variables. For instance, one cannot choose to have the number of lots divided equally into urban and rural areas; rather, one must work with data from the uneven number of rural and urban lots in Virginia. It was expected that grouping the lots into similar geographical markets would be productive (e.g., lots in one part of Virginia might be attractive because they enable carpooling for long distance commutes, whereas lots in another part of Virginia might be attractive because they provide access to transit service). However, it was not immediately obvious how to define a market. Thus, three different geographic boundaries were used for developing regression models: VDOT district, MPO, and PDC. All 78 input variables were considered, and stepwise linear regression via the SPSS software package was used to develop initial models where after identifying the variable that yields the highest adjusted R^2 is identified, the stepwise process adds more variables if those variables are statistically significant and increase the adjusted R^2 (Geert van den Berg, 2020).

There were some cases where additional experimentation was used to improve the model. For example, in the high population density portion of the Fredericksburg District, a model was initially developed from stepwise regression that showed a large constant along with two independent variables (the number of transit riders within 2.5 miles of the lot and the total population within 5 miles of the lot). Dropping these two variables enabled development of another model that included a much lower constant and two different independent variables (the number of persons below the poverty line living within 2.5 miles of the lot and the LEP within 5 miles of the lot). As another example, because some of the initial models contained only facility-specific variables, the research team developed a different model by dropping some of these facility-specific variables. For instance, because the number of transit lines serving the lot was already included, the research team sought to drop a related binary variable: presence of transit service.

Nonlinear Regression Models

If they provide a good fit to the data, linear models are generally preferable to nonlinear models for several reasons: the additive impacts of additional variables are more easily understood, the calibration process is more straightforward, the impact of the variables is easier to interpret, and incorrect assumptions about future conditions (e.g., an error in the forecast of ADT) are less problematic. However, for one district where all other modeling attempts did not yield acceptable performance, the Northern Virginia District with its 108 park and ride lots, the research team sought to improve the fit with a nonlinear model where the dependent variable was the square root of the occupancy, as suggested elsewhere (Montgomery, 2001). The research team further evaluated how removing 21 large parking lots mostly serving the VRE or the Metro system, which might have unique characteristics beyond the scope of the study, would affect model performance.

Evaluate Models

Each model was built twice. First, the final form of the model was determined by using the entire dataset. This gave the coefficient of determination, the standard error of the estimate, and the residual plots. An example of such a model is Equation 15, developed for the park and ride lots in the Richmond Regional PDC. Then, at random, 30% of the observations were removed, the model was recalibrated based on the remaining 70% of observations (an example is Eq. 16), and the accuracy of the model was tested on the remaining 30% of the observations.

$$\text{Occupancy} = 6.361 + 0.006 * \text{Average ADT} + 152.161 * \text{PHEF} \qquad \text{[Eq. 15]}$$

$$\text{Occupancy} = -1.516 + 0.007 * \text{Average ADT} + 125.554 * \text{PHEF} \qquad \text{[Eq. 16]}$$

Although Equations 15 and 16 are similar, there were other cases where the 70% calibrated equation and the full model were very different, as shown in Equations 17 and 18 for the Lynchburg District.

$$\text{Occupancy} = -1.472 + 0.475 * \text{PHEF} + 0.00192*\text{Average ADT} + 0.000049 * \text{POP5} \qquad \text{[Eq. 17]}$$

$$\text{Occupancy} = -9.191 + 1.108 * \text{PHEF} - 0.000296*\text{Average ADT} + 0.000459 * \text{POP5} \qquad \text{[Eq. 18]}$$

Thus, the use of the final form of the model (e.g., Eq. 15 or 17) and the training model (e.g., Eq. 16 or 18) yielded six criteria for evaluating model performance.

Criterion 1. Relatively Large Coefficient of Determination

The coefficient of determination, also known as an adjusted R^2, indicates the percentage of variation in occupancy explained by the independent variables; for example, a value of 0.496 means that 49.6% of the variation in occupancy is explained by the independent variables such that the highest value possible is 100%. The word "adjusted" signifies that a penalty is applied as the number of variables (q) in the model increases (see Eq. 19). This metric is based on the entire dataset.

$$1 - \frac{\sum_i^n (Y_i \text{ observed} - Y_i \text{ forecast})^2 \ \frac{1}{(n-q-1)}}{\sum_i^n (Y_i \text{ observed} - \overline{Y_i \text{ observed}})^2 \ \frac{1}{(n-1)}} \qquad \text{[Eq. 19]}$$

Generally if all other indicators of model performance were equal, one would prefer that the chosen model have a higher adjusted R^2 than other candidate models; Martin (2017) warned against using this as a sole criterion. There is no established threshold at which an R^2 becomes unacceptable, although other park and ride studies suggested that an R^2 of 0.7 was "relatively high" (Nungesser and Ledbetter, 1987), with Vincent (2007) generally reporting models that had an R^2 of 0.8. For a set of fairly homogenous lots, Peng and Mohamad (2005) reported values of 0.9.

The research team generally sought to have a coefficient of determination of at least 0.5, with higher values being preferable.

Criterion 2. Minimize the Standard Error of the Estimate

The standard error of the estimate, henceforth simply "standard error," is computed by squaring the difference between predicted and actual values, dividing by the number of samples, taking the square root (Lane, n.d.), and then dividing by the degrees of freedom (the sample size minus the number of independent variables) such that the standard error reflects the "precision" (Frost, n.d.) of the model in the units of occupancy. This standard error is thus roughly equivalent to the RMSE except the denominator is the number of samples minus the number of independent variables. Standard error is based on the entire dataset.

There is not a particular threshold at which the standard error of a candidate model becomes unacceptable. However, the research team generally sought to develop models where standard error was less than one-half the mean occupancy.

Criterion 3. Find the Acceptable Size of the Intercept

The size of the intercept compared to the mean occupancy indicates the amount of variability explained by the independent variables. For example, for a model where the mean occupancy is 10 and the intercept is 9, the independent variables are simply not offering much explanatory power.

There is not a particular threshold at which an intercept is too large to be useful. However, the research team generally sought to develop models where the intercept was less than one-half the mean occupancy.

Criterion 4. Keep the Mean Testing Error Small

The mean testing error is the average of the absolute value of the difference between the forecast and the observed value. Unlike the preceding three metrics, the mean testing error is not based on the full dataset. Rather, the mean error indicates how the model performs on a separate set of data not used to build the model—that is, 70% of the observations calibrate the model and that model is tested on the remaining 30% of observations.

There is not a particular threshold at which the mean testing error becomes intolerable. For larger lots, the research team generally sought to develop models where testing error was less than one-half the mean occupancy. For smaller lots, however, larger errors on a percentage basis were generally tolerated.

Criterion 5. Logical and Equitable Independent Variables

The utility of the independent variables depends on two factors: whether the independent variables show logical signs, and whether those independent variables can be used for decision making without adversely affecting equity? An example of a model that fails for the first factor might be a model based solely on an intercept and some negative coefficient multiplied by the ADT of the adjacent street, as such a model suggests (contrary to previous experience) that higher ADT is correlated with lower occupancy. An example of a model that fails for the second factor is a model showing a negative coefficient for the LEP near the lot, since such a model could lead eventually to a decision not to build or improve lots in areas with larger LEPs.

Generally, this criterion was applied more strictly than the previous four in the sense that models that exhibited a potential equity issue were avoided where possible. In some cases, modest deviations from this criterion were tolerated when the other candidate models failed in all other criteria.

Criterion 6. Lack of Bias

Plots of the residuals (the difference between observed and forecast occupancy) may be inspected in order to assess if there is sign bias or nonconstant variance (Martin, 2017). To detect sign bias, one can first divide the residual plots (see Figure 7) into rectangles ordered from left to right; blue was added in the figure to show these rectangles. For example, in Figure 7 (*left*), with regard to the horizontal axis only, the leftmost rectangle shows no points where the standardized predicted value is more than 2 standard deviations below the mean; the next rectangle shows five points where the standardized predicted value is between 1 and 2 standard deviations below the mean; and the rightmost rectangle shows one point where the standardized predicted value is more than 2 standard deviations above the mean. If for each rectangle the mean of the points based on the vertical axis is close to zero, then the model is generally unbiased. Figure 7 (*left*) suggests that the Bristol District model is unbiased as the residuals have a mean value that is relatively close to zero for each of the six rectangles. By contrast, Figure 7 (*right*) suggests that the Culpeper District model is biased as the residuals have an average value that is clearly above zero for the two rectangles where standardized predicted values are between 0 and -2.

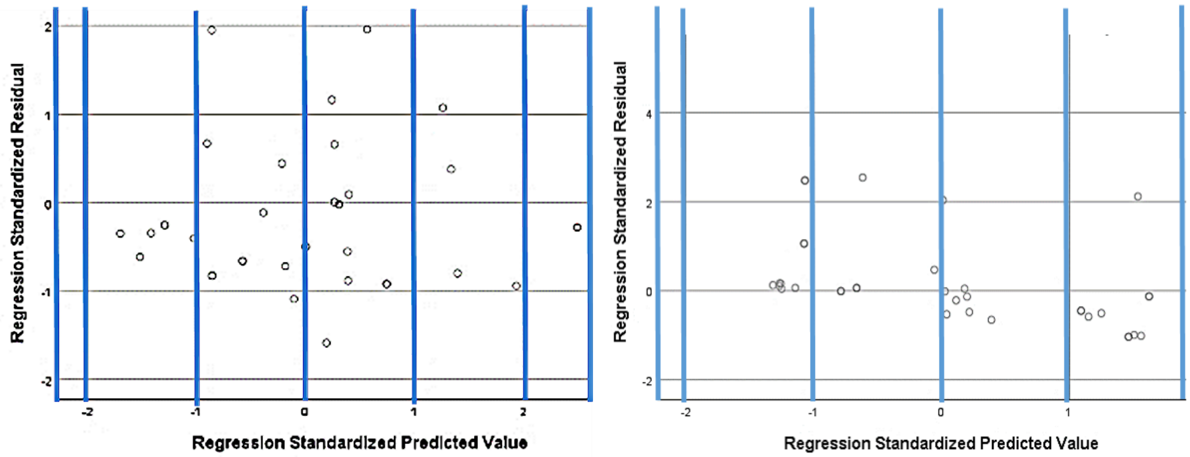

Figure 7. Homoscedastic Residual Plots That Unbiased (left) and Biased (right)

To detect constant variance, Martin (2017) also suggested that the spread of the residuals should be the same in the same six rectangles such that their relative variation is similar. In contrast to Figure 7 (*left*), the Salem District model in Figure 8 (*left*) shows that the relative variance of the residuals increases dramatically from the second leftmost rectangle (-2 to -1 standardized predicted values) to the fourth rectangle (0 to 1 standardized predicted values). Similarly, Figure 8 (*right*) for the Fredericksburg District model also shows heteroscedasticity: except for the standardized predicted values between 1 and 2, the spread of the residuals increases for each rectangle as one moves from left to right.

Although the detection of bias was not problematic, for datasets with a small sample size, visual inspection made determination of heteroscedasticity versus homoscedasticity difficult. For example, Figure 9 shows the Hampton Roads District model where no trend is evident based on these five residuals. In those cases where there was was no clear trend (indicating the model is heteroscedastic), the model was characterized as homoscedastic.

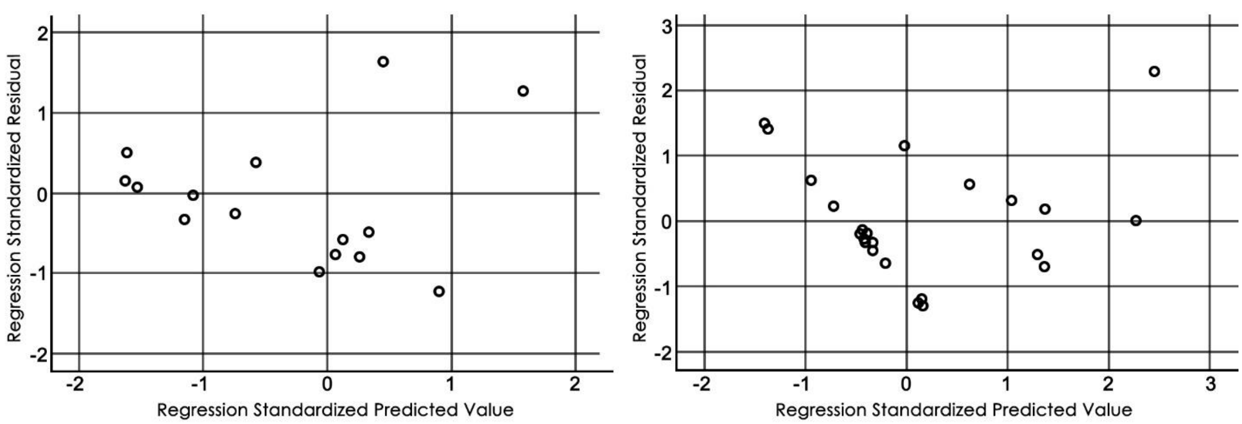

Figure 8. Heteroscedastic Residual Plots That Are Unbiased (left) and Biased (right)

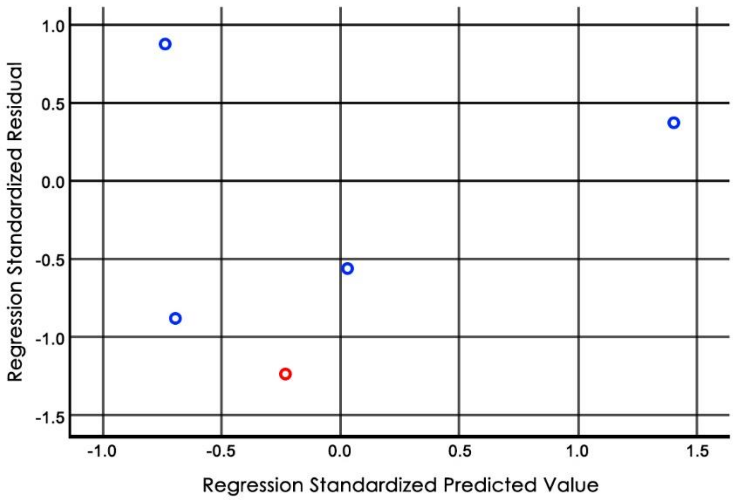

Figure 9. Hampton Roads District Model's Classification Is Unclear Because of the Small Number of Samples

The research team sought to avoid bias, generally preferring an unbiased model (with lower coefficient of determination) than a biased model with a higher coefficient of determination.

Repeat Analysis Based on Guidance From the TRP

The identification of variables, development of models, and evaluation of models in the first three tasks were performed iteratively. The TRP first asked that the research team experiment with applying the existing diversion model as used by FDOT (2012) beyond the initial two districts of Staunton and Richmond. It was hoped that this model could be applicable to Virginia, as it had been used successfully in Florida, either in its original form or as a recalibrated diversion model. Then, especially as the generalized ADT models, the linear regression models, and nonlinear regression models were pursued, the TRP helped identify potential data sources and alternative ways of computing these variables.

For instance, the decision to use a polygon proportion method for summarizing Census data, the use of three different radii for demographic variables, and the consideration of major employment centers through six different variables that entailed GIS processing all arose through applying initial models, seeing poor performance, and working with the TRP to identify possible solutions. The TRP also expressed a strong interest in the use of the models tailored to particular groups of lots, noting that in some regions, for instance, commute length by distance might be a key determinant but that this might not hold in other locations. In particular, one approach used in the Northern Virginia District was to exclude lots that served rail systems and then subdivide the remaining lots into those that offered transit service and those that did not.

One particular decision was how park and ride lots should be aggregated for the development of models, as aggregating by VDOT district is not necessarily the best way to analyze these lots given diverse commuting characteristics. Ultimately, four methods of aggregation were applied in an iterative fashion:

1. VDOT district
2. MPO boundary
3. PDCs
4. population density.

VDOT District Aggregation

The top one-third of Table 5 lists the 297 park and ride lots as distributed through the nine VDOT districts (see Figure 10), where capacity is the total number of parking spaces and occupancy is the number of occupied parking spaces on the day of the year the survey was conducted. For instance, the Bristol District has 29 lots, the largest of which has 80 spaces, although the mean lot size in the Bristol District is 26 spaces. The highest occupancy at any lot in that district was 23 spaces, although the mean was just 8. Table 5 shows variation by district: the mean lot capacity in the Lynchburg District was about one-tenth the mean lot capacity in the Fredericksburg District, which in turn was a bit more than one-half the capacity of the Northern Virginia District. The mean occupancy in the Salem District was about one-third the mean occupancy in the Richmond District, which in turn was about one-third of the mean occupancy in the Northern Virginia District. In short, the 297 lots had very different capacity and demand characteristics.

MPO Boundary Aggregation

The middle of Table 5 shows another way of aggregating some of the park and ride lots: by MPOs. These MPO boundaries (see Figure 11) encompass about two-thirds of the park and ride lots (194 of 297), although only six of the MPOs have enough lots (at least four) to make them suitable for development of a model with multiple variables.

PDC Boundary Aggregation

Virginia's 297 lots may also be categorized by PDC (shown in the bottom one-third of Table 5 and in Figure 12), although 2 of Virginia's 21 PDCs (West Piedmont and Accomack-Northampton, which are not shown) do not have any lots; 3 PDCs each have fewer than three lots such that modeling is not feasible (Central Virginia, Southside, and Commonwealth Regional Council); and 2 PDCs (Northern Neck and Crater) have three lots such that they are probably not reliable for testing purposes (since ideally one should set aside some lots for testing that are not used to develop the initial model). As is the case with the districts, there is substantial variation by PDC: the largest lots in the Northern Virginia District (5,144) and George Washington (1863) have three to ten more times as many spaces as the largest lots in Hampton Roads (504) and Richmond Regional (534), which in turn are roughly twice the size of the largest lots in New River Valley, Roanoke Valley-Alleghany, Northern Shenandoah, Rappahannock-Rapidan PDC, and Middle Peninsula PDC (200-300), which are larger than the capacities of lots in the remaining PDCs (less than 120). However, even in the largest lots, the occupancy has high variance (e.g., 0 to 3,795 in Northern Virginia).

Table 5. Park and Ride Lots by Capacity and Occupancy in Virginia

Grouping	Name	Lots	Capacity				Occupancy			
			Max.	Mean	Med.	Min.	Max.	Mean	Med.	Min.
VDOT District	Bristol	29	80	26	20	6	23	8	7	0
	Culpeper	29	211	38	23	7	120	16	6	0
	Fredericksburg	42	1848	275	43	13	1069	155	16	0
	Hampton Roads	30	504	96	63	10	181	41	22	0
	Lynchburg	8	70	22	20	6	8	3	2	0
	Northern Virginia	108	5144	519	234	12	3795	345	148	0
	Richmond	11	430	137	72	12	280	72	40	1
	Salem	15	270	44	23	12	84	20	12	0
	Staunton	25	258	58	31	5	131	26	11	0
Metropolitan Planning Organization (MPO) or Transportation Planning Organization (TPO)	Bristol MPO[a]	2	50	34	42	42	50	34	42	42
	Charlottesville-Albemarle MPO	8	104	11	32	23	56	0	11	5
	Fredericksburg Area MPO	25	1863	0	439	107	1069	0	254	49
	Hampton Roads TPO	29	504	0	118	66	181	0	44	23
	Harrisonburg-Rockingham MPO[a]	1	50	50	50	50	50	50	50	50
	Kingsport TPO[a]	1	80	80	80	80	16	16	16	16
	Central Virginia MPO[a]	1	78	78	78	78	4	4	4	4
	Northern Virginia District portion of the National Capital Region MPO	109	5144	0	516	232	3795	0	322	143
	New River Valley MPO[a]	2	267	23	145	145	84	17	51	51
	Richmond TPO	10	534	12	173	81	280	1	78	41
	Roanoke Valley TPO	4	239	20	80	30	72	12	32	22
	Staunton-Augusta-Waynesboro MPO[a]	2	120	35	78	78	55	24	40	40
Planning District Commission (PDC)	Lenowisco	9	80	8	31	20	18	1	8	6
	Cumberland Plateau	14	46	0	22	20	23	0	8	7
	Mount Rogers	7	50	6	30	28	20	2	10	9
	New River Valley	8	267	12	51	22	84	0	18	9
	Roanoke Valley-Alleghany	6	239	23	64	30	72	2	25	20
	Central Shenandoah	13	120	5	30	20	55	0	15	7
	Northern Shenandoah	12	257	10	97	66	131	3	39	20
	Northern Virginia	108	5144	12	519	234	3795	0	345	148
	Rappahannock-Rapidan	14	211	7	43	21	120	1	18	6
	Thomas Jefferson	20	104	6	30	23	56	0	12	4
	Central Virginia[a]	2	78	20	49	49	78	20	49	49
	Southside[a]	1	76	76	76	76	0	0	0	0
	Commonwealth Regional Council[a]	1	50	50	50	50	11	11	11	11
	Richmond Regional	10	534	12	173	81	280	1	78	41
	George Washington	28	1863	0	396	73	1069	0	228	25
	Northern Neck[a]	3	75	22	49	50	22	4	12	9
	Middle Peninsula	11	215	16	58	40	25	0	11	7
	Crater[a]	3	116	20	62	50	4	0	1	0
	Hampton Roads	27	504	0	115	66	181	0	46	23

Max. = Maximum; Med. = Median; Min. = Minimum.
[a] The small number of lots precludes developing a model exclusively for this group.

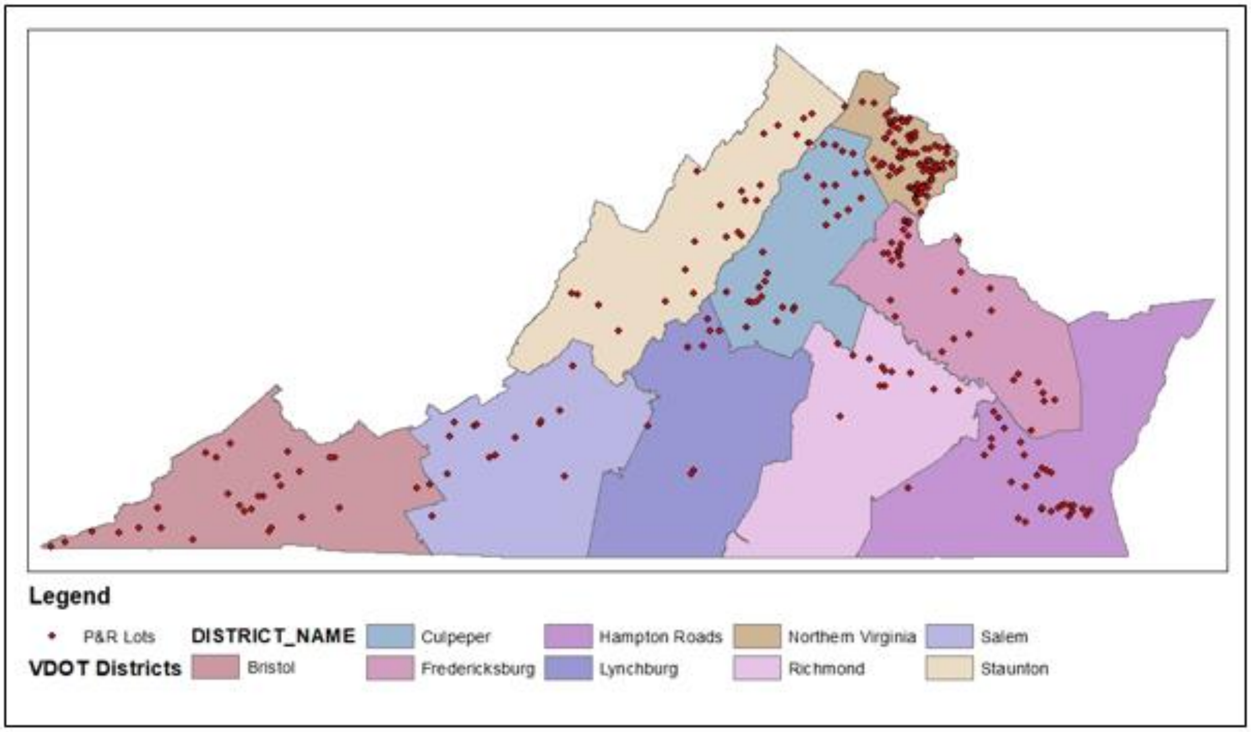

Figure 10. Park and Ride Lots by VDOT District

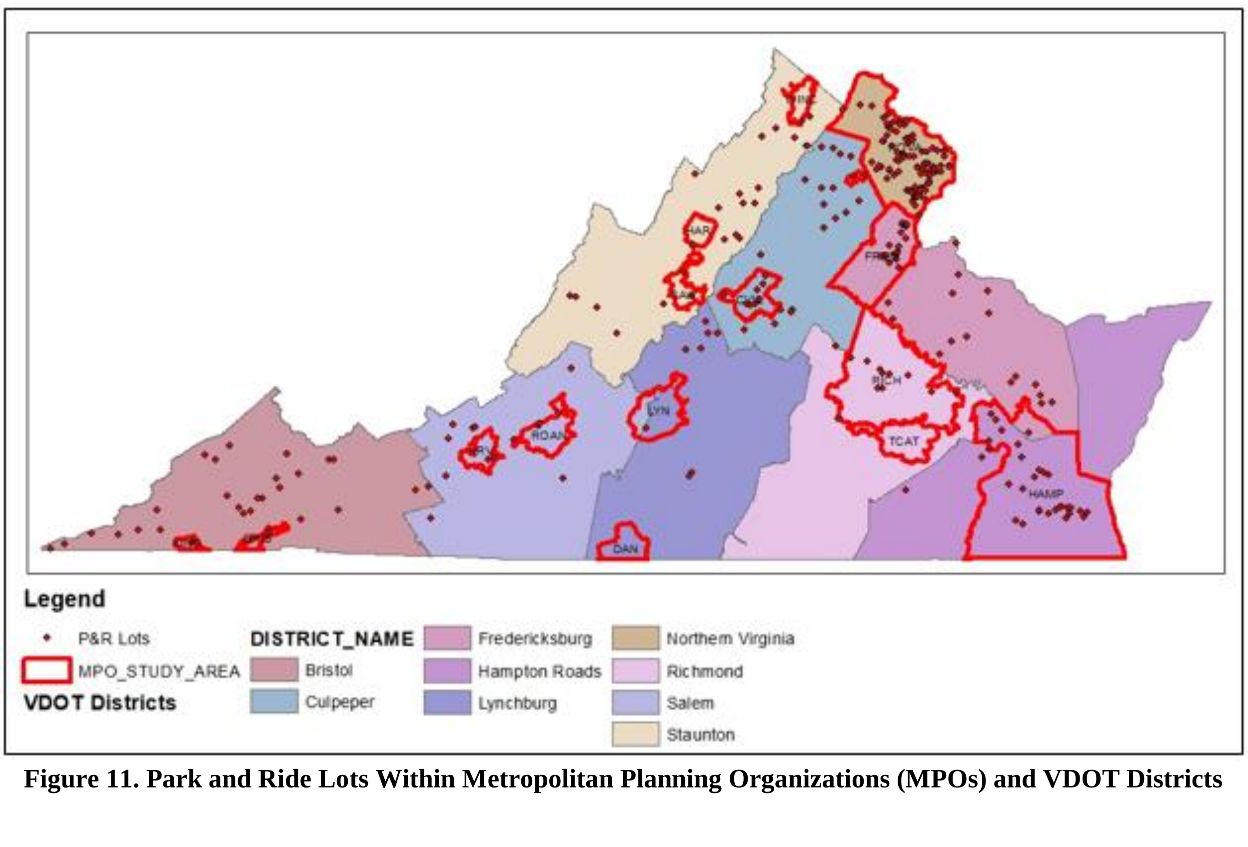

Figure 11. Park and Ride Lots Within Metropolitan Planning Organizations (MPOs) and VDOT Districts

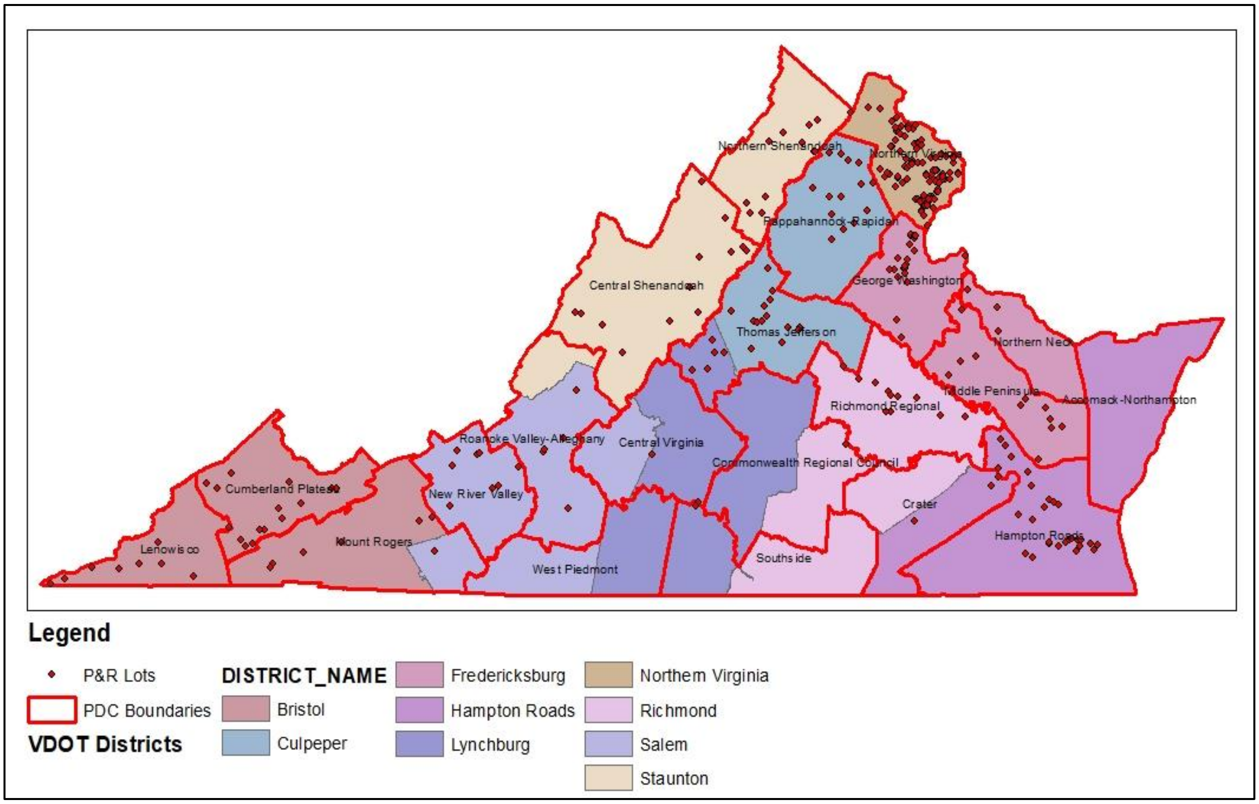

Figure 12. Park and Ride Lots Within Planning District Commissions (PDCs) and VDOT Districts

Population Density Aggregation

In the Fredericksburg and Hampton Roads districts, where initial model development seemed especially poor and where the districts appeared to have a substantial range of urban and rural areas, separate models for higher and lower density areas were developed where the population for each Census block group was divided by the geodesic area of the block group; the block groups are all in the coordinate system of Virginia State Plane South. Then, for each district, the Jenks method was used to classify the densities into low and high categories, as shown in Table 6. A similar approach was tried in the Richmond District; however, all lots were in the low density area. Maps for these districts are provided in the Appendix.

Table 6. Categorization of Park and Ride Lots Based on Block Group Density (People per Square Mile)

District	Low Density Range[a] (No. of lots)	High Density Range[a] (No. of lots)
Fredericksburg	0-2395 (30 lots)	2396-14669 (12 lots)
Hampton Roads	0-5374 (25 lots)	5375-67230 (5 lots)
Richmond	0-5883 (11 lots)	5884-28062 (0 lots)

[a] Ranges are rounded to the nearest integer.

The variables in Table 6 raise a potential question if a park and ride lot is located in an area that was formerly low density but is forecast to move to the high density category. As discussed later, one would want generally to use the density classification that will be applicable at the time the forecast is generated. To be clear, the regression models are based on correlation: one can never definitively prove that a given independent variable value leads to a given dependent variable value; one can only note that they are correlated.

RESULTS

The cumulative results from the four methods are presented in two sections. The first section gives the results of applying the existing diversion method directly to each VDOT district. The second section shows the models tailored to Virginia for each VDOT district (recalibrated diversion models, generalized ADT models, and models based on regression).

Application of the Existing Diversion Model

Table 7 shows the results of applying the existing diversion model to each VDOT district where each column refers to an absolute value. For example, in the Lynchburg District, when one computes the absolute value of the difference between the forecast occupancy and the mean occupancy for each of the eight lots, the mean difference is 423 spaces and the median difference is 385 spaces. With a mean occupancy of 3 and a median occupancy of 2, the district's mean absolute difference is about 141 times higher than the mean occupancy and its median absolute difference is about 192.5 times higher than the median occupancy. The two left columns suggest that the existing diversion model yielded the lowest errors for the Bristol and Lynchburg districts when only the number of occupied spaces was considered. However, when the fact that the average occupancy varies by district is considered, the next column suggests that the lowest errors are in the Fredericksburg and Northern Virginia districts. Overall, Table 7 shows that generally the error is many times larger than the occupancy for all districts when the existing diversion model is used.

Table 7. Error of Application of Existing Diversion Model in Virginia Districts[a]

District	|Mean Difference|	|Median Difference|	|Mean Difference| / Mean Occupancy	|Median Difference| / Median Occupancy
Bristol	542	296	67.750	42.286
Culpeper	1243	1133	77.688	188.833
Fredericksburg	2193	1837	14.148	306.167
Hampton Roads	2280	3801	55.610	172.773
Lynchburg	423	385	141.000	192.500
Northern Virginia	5266	4971	16.303	34.521
Richmond	2837	2805	39.403	70.125
Salem	1313	1203	65.650	100.250
Staunton	897	932	34.500	84.727

[a] All numbers presented are absolute values.

Development of Models for Each VDOT District

Tables 8 through 13 and 15 through 17 show the results of the models by VDOT district. All variables shown are statistically significant (p = 0.05 or lower) except for the models that are based exclusively on the V_{peak} and V_{prime} terms, such as Models 1 and 2 in Table 8. Each table gives the full model, coefficient of determination (adjusted R^2), mean occupancy, and the results of inspecting the plot of residuals, all of which are based on the full dataset. The tables also indicate the mean testing error, where the model was recalibrated based on 70% of the observations and then applied to the remaining 30% of observations (which were not used to build the model).

Each table also shows the recommended model for future application based on consideration of six criteria: (1) better adjusted R^2 than other models, (2) a standard error of less than one-half of the mean occupancy; (3) a mean testing error that is less than one-half of the mean occupancy; (4) residuals that are unbiased and homoscedastic; (5) less than one-half of the occupancy information contained in the intercept if the intercept is positive; and (6) variables with the proper sign and whose incorporation in a forecast will not cause VDOT to be inequitable when using occupancy forecasts to make investments in park and ride lots.

Bristol District

Table 8. Candidate Models for Park and Ride Lots in the Bristol District (29 Lots)

No.	Model (if not district-wide, number of sites and applicability)[a]	Adj. R^{2a}	Std. Error[a]	Mean |Error|[b]	Mean Occ.[a]	Residuals[a]
1	0.002 * $Vpeak_{PHF}$ + 0.000217 * $Vprime_{PHF}$	0.612[a]	6.564	4	8	Biased, Heteroscedastic
2	0.02 * $Vpeak_K$ + 0.002 * $Vprime_K$	0.623[a]	6.467	4	8	Biased, Heteroscedastic
3	5.823 + 0.001 * Closest ADT	0.207	5.709	4	8	Unbiased, Homoscedastic
4	20.963 + 46.327 * V/C – 1.037 * RentOverAllIncome5	0.385	5.029	2	8	Unbiased, Homoscedastic
5[c]	1.359 + 0.082 * Rad2_JobsGT50Mi (Lenowisco PDC, 9 lots)	0.839	2.613	5	8	Unbiased, Homoscedastic
6[c]	2.099 + 0.018 * Closest ADT - 0.362 * $Vpeak_K$ + 0.004 * Rad5_JobsLT10Mi (Cumberland Plateau PDC, 14 lots)	0.919	1.866	5	8	Unbiased, Homoscedastic
7[c]	-7.006 + 0.247 * TransitRiders2 + 0.004 * $Vprime_K$ (Mount Rogers PDC, 7 lots; includes 1 VDOT Salem lot)	0.668	3.544	5	10	Unbiased, Homoscedastic

Occ. = occupancy.
[a] Model, adjusted R^2, standard error, mean occupancy, and residuals are based on the entire dataset.
[b] |Mean error| is based on applying an earlier model, slightly different from that shown, developed from just 70% of the data to the remaining 30% testing dataset.
[c] Model is recommended for use.

Table 8 suggests that the best models for the 29 lots in the Bristol District are the three separate models for the PDCs that largely comprise the district. All three models met at least four of the six criteria established for a good model: higher adjusted R^2 compared to the other models (in this case, 0.67 to 0.92); less than one-half of the occupancy information contained in the intercept (e.g., with a mean occupancy of 8 to 10, the intercepts are all less than 4 to 5); residuals that are unbiased and homoscedastic; and a standard error that is less than one-half the mean occupancy. The Lenowisco PDC and Mount Rogers PDC models each met the fifth criterion: signs of independent variables are logical and will not cause policy challenges for forecasters. The Cumberland Plateau PDC model may only partially meet this criterion: the closest ADT variable is positive as expected, but the negative sign associated with the peak hour volume for this same facility is initially counterintuitive. Certainly there may be a good explanation for the sign as indicated: for example, it may be the case that use occurs outside the peak hour. None of the three chosen models met the sixth criterion of having a testing error less than one-half the mean occupancy. In this particular case, however, the fairly low mean occupancy values may make failing to meet this last criterion acceptable.

By comparison, each of the four models that were not selected had at least one substantial flaw. The first two models had residuals that were biased and heteroscedastic, and the third model had a large intercept (5.8) that contained more than one-half of the explanatory power of the model, given the mean occupancy of 8. The fourth model—which applies district-wide—failed two criteria: it had a lower percent of variance explained (about 39%) and a problematic independent variable in the form of a negative coefficient for rent divided by income. The former criterion is not a fatal weakness: despite the lower adjusted R^2, the fourth model has the lowest mean error of all seven models—about one-fourth of the mean occupancy. However, the latter criterion appears to present a potential equity challenge by VDOT: if in a given location rents relative to income rise, the forecast is that the occupancy will drop such that implementation could yield a reduction in services (the provision of a lot). Thus, the equity challenge with the fourth model appeared greater than the contradictory signs associated with the Cumberland Plateau PDC model.

Culpeper District

The four district-wide models in Table 9 for the 29 Culpeper District lots were all problematic: all models exhibited bias, all showed a testing error larger than the mean occupancy, and one did not yield a positive adjusted R^2. Dividing the Culpeper District into subareas—the Rappahannock-Rapidan PDC and the Thomas Jefferson PDC—was helpful in this regard, with one of these PDCs—Thomas Jefferson—being split further into the urbanized portion (the Charlottesville-Albemarle MPO) and a rural portion (which excludes the MPO). This subdivision yielded three models, although as was the case with the Bristol District, none met all six criteria.

Table 9. Candidate Models for Park and Ride Lots in the Culpeper District (29 Lots)

No.	Model (Limitations if not applicable to the entire district) [a]	Adj. R^{2a}	Std. Error[a]	Mean |Error|[b]	Mean Occ.[a]	Residuals[a]
1	-0.000266* $Vpeak_{PHF}$ + 0.001 * $Vprime_{PHF}$	0.227	26.214	24	16	Biased, Homoscedastic
2	-0.002 * $Vpeak_K$ + 0.004 * $Vprime_K$	0.200	26.664	24	16	Biased, Homoscedastic
3	0.000411 * Max ADT	0.267	26.061	29	16	Biased, Homoscedastic
4	12.744 + 0.001 * Average ADT	-0.031	26.061	29	16	Biased, Homoscedastic
5	0.004 * Average ADT (Thomas Jefferson PDC, 20 lots, includes 5 VDOT Staunton lots)	0.451	16.966	11	12	Biased, Homoscedastic
6	0.000066 * $Vpeak_{PHF}$ + 0.000293 * $Vprime_{PHF}$ (Charlottesville-Albemarle MPO, 8 lots)	0.149	19.447	6	11	Unbiased, Homoscedastic
7	-0.0002 * $Vpeak_K$ + 0.002 * $Vprime_K$ (Charlottesville-Albemarle MPO, 8 lots)	0.110	19.884	6	11	Unbiased, Homoscedastic
8[c]	-4.107 + 27.798 * OvernightParkingAllowed + 0.009 * TransitRiders2 (Charlottesville-Albemarle MPO, 8 lots)	0.583	12.307	4	11	Unbiased, Homoscedastic
9[c]	2.725 + 0.042 * Rad2_JobsLT10Mi (non-MPO portion of Thomas Jefferson PDC, 12 lots, includes 5 VDOT Lynchburg lots)	0.324	13.496	14	12	Unbiased, Homoscedastic
10[c]	6.141 – 0.065 * Carpoolers2 + 0.059 * TransitRiders2 + 0.005 * Rad5_Jobs10_to_24Mi + 8.229 * Lighting (Rappahannock-Rapidan PDC, 14 lots)	0.942	7.721	27	18	Unbiased, Homoscedastic

Occ. = occupancy.

[a] Model, adjusted R^2, standard error, mean occupancy, and residuals are based on the entire dataset.

[b] |Mean error| is based on applying an earlier model, slightly different from that shown, developed from just 70% of the data to the remaining 30% testing dataset.

[c] Model is recommended for use.

1. Model 10 for the lots in the Rappahannock-Rapidan PDC met four criteria (percent of variance explained, a standard error that is less than one-half the mean occupancy, and unbiased/homoscedastic residuals). Although the mean testing error of 27 was large, removal of two outliers in the testing dataset (whose errors are more than 3 standardized residuals from the mean) would reduce the mean error from 27 to 3. Without these two sites the coefficient of the full model would be similar to those shown in Table 9, of 5.651 – 0.083 * Carpoolers2 + 0.081 * TransitRiders2 + 0.007 * Rad5_Jobs10_to_24Mi + 9.886 * Lighting. Note this changed model is similar to model 10 in Table 9: for example the coefficient of Carpoolers 2 is -0.083 versus -0.065. The coefficients for the independent variables were plausible; although the negative sign associated with the number of carpoolers within 2.5 miles of the lot was counterintuitive in isolation, it is possible this was offset by the role of transit riders; both variables indicate areas within 2.5 miles of the lot. That said, the number of carpoolers was always higher than the number of transit riders within 2.5 miles of the lot in this region.

2. Model 8 for the Charlottesville-Albemarle MPO (which has eight lots) met five criteria. Notably, the lots in this area are sensitive to transit demand, which can be expected given that they are in the urbanized area of Charlottesville plus Albemarle County. The one weakness of Model 8 was the relatively large standard error compared to the mean occupancy, although the testing error was considerably better. The model's indication of overnight parking allowance increasing lot occupancy by 28 might not necessarily reflect a causal relationship; it is possible at this location that overnight parking restrictions indeed represented some other factor not available to the research team.

3. Model 9, which separates the 12 rural lots in the Thomas Jefferson PDC area from the 8 urban lots in the Charlottesville-Albemarle MPO, showed a testing error (14) larger than the mean occupancy (12), but it appeared better than alternative Model 5 in terms of the residual plots with a comparable mean testing error.

Fredericksburg District

There simply are no strong models for the 42 Fredericksburg District lots. Model 1 is an option, and certainly the poor coefficient of determination could be explained by the fact that other variables may influence this diverse set of lots—but the bias exhibited by the residuals makes such a model problematic. Splitting the 42 lots into high population density areas (12 lots) and low population density areas (30 lots) helped modestly but still yielded models with errors larger than the mean occupancy. A contributing factor for the high density portion may have been that most of the lots (10 out of 12) are VRE lots.

In high density areas, Models 2 and 3 showed higher standard errors and higher testing errors than the mean occupancy. Although it shows a low testing error, Model 4's negative coefficient for the number of transit riders within 2.5 miles of the lot appeared hard to explain in isolation. Because the model was already restricted to block groups of high population density, the fact that the transit ridership reduced forecast occupancy suggested that this mode choice might be a surrogate for some other unobserved variable. For example, because these high density areas include transit service, it may be the case that an increase in transit ridership signifies the substitution of transit trips for carpooling trips such that occupancy drops. However, the positive coefficient of population within 5 miles of the lot makes sense. Except for the fact that the testing error for Model 5 was higher than for Model 4, mathematically, Model 5 might be the best of the models in that part of the district. However, the negative coefficient associated with the population classified being in poverty within 2.5 miles of the lot could present an equity issue. Of these four models, Models 4 and 5 are the "least bad" options for the dozen lots in high density areas of the Fredericksburg District, and because of the equity concerns, Model 4 is recommended.

Table 10. Candidate Models for Park and Ride Lots in the Fredericksburg District (42 Lots)

No.	Model (Limitations if not applicable to the entire district)[a]	Adj. R^{2a}	Std. Error[a]	Mean |Error|[b]	Mean Occ.[a]	Residuals[a]
1	27.467 + 0.002 * Max ADT	0.132	244.598	125	155	Biased, Heteroscedastic
2	0.123 * $Vpeak_{PHF}$– 0.002 * $Vprime_{PHF}$ (high population density, 12 lots)	0.618 [a]	211.661	248	180	Unbiased, Homoscedastic
3	1.077 * $Vpeak_K$ – 0.013 * $Vprime_K$ (high population density, 12 lots)	0.638 [a]	205.985	232	180	Unbiased, Homoscedastic
4 [c]	1931.706 - 1.949 * TransitRiders2 + 0.000256 * POP5 (high population density, 12 lots)	0.979	68.041	35	180	Unbiased, Homoscedastic
5	-175.485 - 0.021 * PovertyPOP2 + 0.048 * LEPPop5 (high population density, 12 lots)	0.945	71.283	135	180	Unbiased, Homoscedastic
6	0.003 * $Vpeak_{PHF}$ + 0.002 * $Vprime_{PHF}$ (low population density, 30 lots)	0.618 [a]	247.256	106	145	Biased, Heteroscedastic
7	0.013 * $Vpeak_K$ – 0.022 * $Vprime_K$ (low population density, 30 lots)	0.294 [a]	248.815	109	145	Biased, Heteroscedastic
8	37.489 + 0.002 * Max ADT (low population density, 30 lots)	0.132	244.598	125	145	Biased, Heteroscedastic
9 [c]	-217.053 + 241.839 * Lighting + 329.448 * NuofTranServicePP +0.018 * POPDEN + 67.016 * PHEF (low population density, 30 lots)	0.790	120.285	36	145	Biased, Heteroscedastic
10	0.006 * $Vpeak_{PHF}$ + 0.002 * $Vprime_{PHF}$ (FAMPO, 25 lots)	0.325	329.988	285	254	Unbiased, Homoscedastic
11	0.048 * $Vpeak_K$ + 0.020 * $Vprime_K$ (FAMPO, 25 lots)	0.327	329.584	300	254	Unbiased, Homoscedastic
12	296.462 + 6.159 * NuofAdjLot + 375.304 * Lighting – 0.047 * Rad2_JobsLT10Mi (FAMPO, 25 lots)	0.600	200.791	203	254	Unbiased, Homoscedastic
13	-102.649 – 0.001 * EMP2.5 + 0.038 * Rad10_Jobs10_to_24Mi (George Washington PDC, 28 lots)	0.575	173.987	170	228	Unbiased, Heteroscedastic
14 [c]	1.872 * RentOverAllIncome2 – 0.999 * AvgPctOfRentIncomeOnRent5 (Middle Peninsula PDC, 11 lots)	0.667	7.468	11	11	Unbiased, Homoscedastic

Occ. = occupancy.
[a] Model, adjusted R^2, standard error, mean occupancy, and residuals are based on the entire dataset.
[b] |Mean error| is based on applying an earlier model, slightly different from that shown, developed from just 70% of the data to the remaining 30% testing dataset.
[c] Model is recommended for use.

For the low density area of the Fredericksburg District, all models were biased and heteroscedastic. However, Model 9 may be tolerable for forecasting purposes. This model, which applies to most sites (30 of 42) in the Fredericksburg District, met the four remaining criteria: coefficients that appear logical (e.g., transit service, lighting, and congestion increase occupancy), testing error below one-half the mean occupancy, a higher coefficient of determination than other models, and less than one-half of the occupancy information held in the intercept. Although the standard error for Model 9 is about four-fifths of the mean occupancy, it is nonetheless lower than for alternative models.

Model 14 is an option for the Middle Peninsula PDC portion of the district: the model has a relatively high mean error compared to the occupancy, but the low value overall may make it worth considering. That is, if one applies Model 9 (the recommended low density model) to the Middle Peninsula PDC sites (all of which are located in low density areas), the testing error exceeds 30. However, the last term is an equity concern: Model 14 suggests that fewer park and ride lots are needed in locations where the percent of income spent on rent rises (for locations within 5 miles of the park and ride lot). If Model 14 is not used, then Model 9 would be used for those Middle Peninsula PDC sites. For that reason, both Model 9 and Model 14 are listed as options for the Middle Peninsula PDC portion of the Fredericksburg District.

Hampton Roads District

Table 11. Candidate Models for Park and Ride Lots in the Hampton Roads District (30 Lots)

No.	Model (Limitations if not applicable to the entire district) [a]	Adj. R^{2a}	Std. Error[a]	Mean |Error|[b]	Mean Occ.[a]	Residuals[a]
1	1.576 + 0.000390 * Max ADT	0.226	42.239	14	41	Unbiased/ Homoscedastic
2	-0.000037 * $Vpeak_{PHF}$ + 0.001 * $Vprime_{PHF}$ (high population density, 5 lots)	0.678	58.684	78	84	Could not determine
3	-0.001 * $Vpeak_K$ – 0.008 * $Vprime_K$ (high population density, 5 lots)	0.689	57.667	68	84	Could not determine
4[c]	-541.193 + 0.052 * Carpoolers2 + 0.252 * Rad2_Jobs25_to_50Mi (high population density, 5 lots)	0.986	7.929	3	84	Unbiased/ Homoscedastic
5	-0.001 * $Vpeak_{PHF}$ + 0.000419 * $Vprime_{PHF}$ (low population density, 25 lots)	0.538	34.431	43	33	Unbiased/ Heteroscedastic
6	-0.014 * $Vpeak_K$ – 0.004 * $Vprime_K$ (low population density, 25 lots)	0.540	34.360	21	33	Unbiased/ Heteroscedastic
7	7.734 + 0.000142 * EligDisadvPop5 - 0.005 * Rad5_JobsLT10Mi + 0.016 * Rad5_JobsGT50Mi (low population density, 25 lots)	0.580	25.629	16	33	Unbiased/ Heteroscedastic
8[c]	8.341 + 0.000262 * Max ADT (low population density, 25 lots)	0.170	35.310	14	33	Unbiased/ Homoscedastic
9	23.183 + 0.001 * LEPPop5 + 103.384 * BikeParkingisCovered (HRTPO, 28 lots)	0.407	37.218	46	44	Biased/ Homoscedastic
10	0.000352 * $Vpeak_{PHF}$ + 0.000384 * $Vprime_{PHF}$ (HRTPO, 28 lots)	0.554	43.292	44	44	Unbiased/ Homoscedastic
11	0.003 * $Vpeak_K$+ 0.004 * $Vprime_K$ (HRTPO, 28 lots)	0.557	43.146	43	44	Unbiased/ Homoscedastic
12	25.671 + 100.831 * BikeParkingisCovered + 0.000113 * POP5 – 0.002 * Rad5_JobsLT10Mi (Hampton Roads PDC, 27 lots)	0.499	34.465	29	46	Unbiased/ Homoscedastic

Occ. = occupancy.
[a] Model, adjusted R^2, standard error, mean occupancy, and residuals are based on the entire dataset.
[b] |Mean error| is based on applying an earlier model, slightly different from that shown, developed from just 70% of the data to the remaining 30% testing dataset.
[c] Model is recommended for use.

As was the case with the Fredericksburg District, segmentation into high and low population density areas helped modestly. Most lots in Hampton Roads (25 of 30) are in low density locations, meaning that a generalized ADT model for the entire district (Model 1) and the generalized ADT model for the low density locations (Model 8) are similar in terms of model coefficients, standard error, testing error, and adjusted R^2. Only two models are recommended for future use:

1. For the 5 lots that are high density, Model 4 is recommended. However, the small sample size (4 lots for training and 1 lot for testing) means that factors other than the usual criteria of low testing error, higher coefficient of determination, and lack of bias shown by the residuals should be considered. In this particular case, the fact that the number of carpoolers increases the occupancy is logical.

2. For the 25 lots that are low density, Model 8 is recommended, primarily because of one criterion: the model is unbiased and homoscedastic. Although that model explains only 17% of the variation in occupancy (compared to 58% for Model 7), it is telling that both models performed similarly with the testing dataset. That said, the standard error is larger than the mean occupancy.

The existence of Models 9 and 12 comprises a cautionary tale regarding inferences that can be drawn from models and a recognition that there is not always a clearly best model. The 30 lots in the Hampton Roads District are largely similar to the 27 lots in the Hampton Roads PDC or the (25, 28, or 29) lots in the Hampton Roads Transportation Planning Organization (TPO). Although it is comforting that the same variable (covered bicycle parking) is present in both models with a similar coefficient in each, it is interesting that an additional variable in each model (LEP population in Model 9 and total population in Model 12, each within 5 miles of the lot) is included. It may be the case that segmentation by population density (Models 4 and 8) thus allows for different characteristics to be included in explaining occupancy in these two different areas (e.g., the number of carpoolers within 2.5 miles of the lot in high density areas as per Model 4, and ADT in low density areas as per Model 8).

That said, one could also argue that the small number of lots (5) in the high density locations renders segmentation by population less useful such that an alternative approach could be simply to adopt Model 12 (based on the PDC boundaries) and then not have an approach for estimating demand for the 3 remaining lots.

Lynchburg District

All four models in the Lynchburg District were similar with regard to four of the six criteria. They all met three criteria: (1) most of the information in the model is in the variables rather than the intercept; (2) all independent variables are logical (e.g., occupancy rises as a function of ADT); and (3) residuals suggest an unbiased model with constant variance. The models all failed one criterion: the mean testing error was larger than one-half the occupancy; it was either equal to or exceeded the occupancy.

Table 12. Candidate Models for Park and Ride Lots in the Lynchburg District (8 Lots)

No.	Model (Limitations if not applicable to the entire district)[a]	Adj. R^{2}[a]	Std. Error[a]	Mean |Error|[b]	Mean Occ.[a]	Residuals[a]
1	0.001 * $Vpeak_{PHF}$ + 0.000274 * $Vprime_{PHF}$	0.848	1.426	4	3	Unbiased/ Homoscedastic
2	0.006 * $Vpeak_{K}$ + 0.002 * $Vprime_{K}$	0.862	1.358	4	3	Unbiased/ Homoscedastic
3	0.879 + 0.0003 * Closest ADT	0.469	1.983	3	3	Unbiased/ Homoscedastic
4[c]	-1.472 + 0.475 * PHEF + 0.002 * Average ADT + 0.000049 * POP5	0.949	0.614	9	3	Unbiased/ Homoscedastic

Occ. = occupancy,
[a] Model, adjusted R^2, standard error, mean occupancy, and residuals are based on the entire dataset.
[b] |Mean error| is based on applying an earlier model, slightly different from that shown, developed from just 70% of the data to the remaining 30% testing dataset.
[c] Model is recommended for use. An option for 5 of these lots, however, is to use the rural portion of the Thomas Jefferson PDC model associated with the Culpeper District.

There were differences in two criteria: (1) the percent of variation in occupancy explained (a bit less than 50% for Model 3 compared to almost 95% for Model 4), and (2) size of the standard error (about 66% of the mean occupancy for Model 3 compared to 20% of the mean occupancy for Model 4).

For a district-wide approach, Model 4 is recommended as the better model based on the fact that it did well with regard to five of the six criteria and the fact that although its mean testing error was 3 times the occupancy, the error was nonetheless relatively small in absolute terms. A salient reason for preferring Model 4 to Model 3 is the reduced standard error of Model 4. However, the Lynchburg District includes 5 lots that are in the Thomas Jefferson PDC, and thus for those 5 lots, the rural portion of the Thomas Jefferson PDC model is also applicable.

As discussed previously, the testing error is developed based on a 70% model that is applied to the remaining 30% dataset. For example, after the form of Model 3 was determined from 100% of the data, a recalibration based on just 70% of the data gave -0.485 + 0.000419 * Closest ADT. That recalibrated model gave a forecast of 0 for one of two testing sites (compared to an observed value of 3) and a forecast of 1 at the other testing site (compared to an observed value of 4), and hence a mean testing error of 3 is shown for Model 3. This value of 3 is much lower than the mean testing error of Model 4, which was 9 based on the same two testing sites. However, in the full model shown in Table 12, the average difference between forecast and observed values based on all eight sites gave much lower values of 1.5 (Model 3) and 0.375 (Model 4). This result was consistent with the fact that Model 3 had about 3 times the standard error of Model 4. That said, one could also defend Model 3 if data requirements necessitated its use.

Northern Virginia District

Table 13. Candidate Models for Park and Ride Lots in the Northern Virginia District (108 Lots)

No.	Model (Limitations if not applicable to the entire district)[a]	Adj. R^{2a}	Std. Error[a]	Mean |Error|[b]	Mean Occ.[a]	Residual[a]
1	0.002 * $Vpeak_{PHF}$ + 0.002 * $Vprime_{PHF}$	0.239[a]	618.105	380	345	Unbiased/ Heteroscedastic
2	0.001 * $Vpeak_{K}$ + 0.000320 * $Vprime_{K}$	0.246	615.266	316	345	Unbiased/ Heteroscedastic
3	122.474 + 0.000049 * Sum ADT	0.039	609.839	396	345	Unbiased/ Heteroscedastic
4	104.685 + 928.503 * Cost to Park + 46.499 * NuofTranServicePP – 17.549 * ProxToIAP	0.572	406.927	318	345	Unbiased/ Heteroscedastic
5	29.919 + 0.19 * Average ADT + 48.053 * DTNearestP (*remove 11 lots based on residuals*)	0.065	225.980	168	191	Unbiased/ Homoscedastic
6	$(4.503 + 9.349$ * Bike Parking is Covered + 1.132 * NuofTranservicePP – 0.514 * ProximityToIAP+ 0.164 * RentOverAllIncome2$)^2$	0.513	(8.65065)	292	345	Unbiased/ Homoscedastic
7	(0.786 + 4.902 * BikeParkingIsCovered + 0.587 * NuofTranServicePP + 6.249 * TransitServiceAvailable$)^2$ (*remove 21 lots based on residuals and VRE = 87 lots*)	0.427	(5.56)	164	147	Unbiased/ Homoscedastic
8	-508.705 + 241.725 * BikeParkingIsCovered + 13.672 * NuofTranServicePP + 20.678 * AvgPctOfRentIncomeOnRent2 (*remove 21 lots based on residuals and VRE = 87 lots*)	0.447	140.933	166	147	Unbiased/ Heteroscedastic
9	(-7.614 + 0.330 * Bicycle Spaces + 0.616 * NuofTranServicePP + 0.586 * RentOverAllIncome2)2 (*remove 21 lots based on residuals and VRE = 87 lots*)	0.515	5.11787	85	147	Unbiased/ Homoscedastic
10[c]	(2.488 + 0.298 * Bicycle Spaces + 0.396 * NuofTranServicePP + 0.001 * Average ADT$)^2$ (*78 of the 87 lots with transit service*)	0.483	5.18047	139	163	Unbiased/ Homoscedastic
11[c]	(2.844 - 0.000071 * Dist_M2 + 1.128 * DTNearestP$)^2$ (*9 of the 87 lots without transit service*)	0.657	0.733375	4	4	Unbiased/ Homoscedastic

Occ. = occupancy.
[a] Model, adjusted R^2, standard error, mean occupancy, and residuals are based on the entire dataset.
[b] |Mean error| is based on applying an earlier model, slightly different from that shown, developed from just 70% of the data to the remaining 30% testing dataset.
[c] Model is recommended for use.

The method of developing models for the Northern Virginia District proceeded differently from that for the other districts because of the large number of lots, the widely varying characteristics of those lots, the urban nature of that district, and some initial challenges when district-wide models were developed. Thus, for this district only, five additional steps were followed in model development:

1. Identify lots with homogenous responses.
2. Develop a nonlinear model.
3. Apply a nonlinear model to lots with homogenous characteristics.
4. Segment the nonlinear model for lots with homogenous characteristics.
5. Consider a model for the excluded lots.

Identify Lots With Homogenous Responses

The first four models in Table 13 have large standard errors (relative to the mean occupancy), large testing errors, and heteroscedastic plots of residuals. Unlike with the other districts, challenges were not attributable to a small number of samples: the Northern Virginia District has 108 park and ride lots. Thus, clearly the first four linear models suffered from at least one of three possible drawbacks: missing an explanatory variable, incorrectly assuming only a linear relationship between occupancy and independent variables, or a dataset that requires segmentation in order to reduce scatter.

In Model 5, the research team considered this last possibility that some lots in the Northern Virginia District are fundamentally different from others—as mentioned previously, this district has a wide variance in annual estimates of occupancy from a value of 0 at the Leesburg II lot at the intersection of Crosstail Blvd. & Claudia Drive to a value of 3795 at the Vienna-Fairfax-GMU Station Metro lot (which was more than 80% of the lot's capacity of 4,467). To detect lots that may have fundamental differences from other lots, the standardized residuals from the regression equation in Model 4 (on the vertical axis) were plotted against the standardized predicted values (on the horizontal axis), as shown in Figure 13. Although most (97) lots are clustered to the left of Figure 13 around a standardized predicted value of 0, there are a few (10) lots that are clustered to the right of Figure 13 and one lot that is at 5 standardized residuals; these 11 lots are numbered in the figure.

These 11 lots were removed from the analysis (9 were Metro lots, 1 was the Ballston Parking Garage, and 1 was the Horner Road Commuter lot). A new model—Model 5—was built using just the remaining 7 lots. This removal of the 11 lots was successful in two ways: rendering constant variance (as exhibited by the change in the plot of the residuals), and lowering the mean testing error by almost one-half. However, the model was problematic in that an adjusted R^2 of just about zero means that for forecasting purposes, a better answer would be simply to take the mean occupancy of the dataset and use that as a forecast of occupancy.

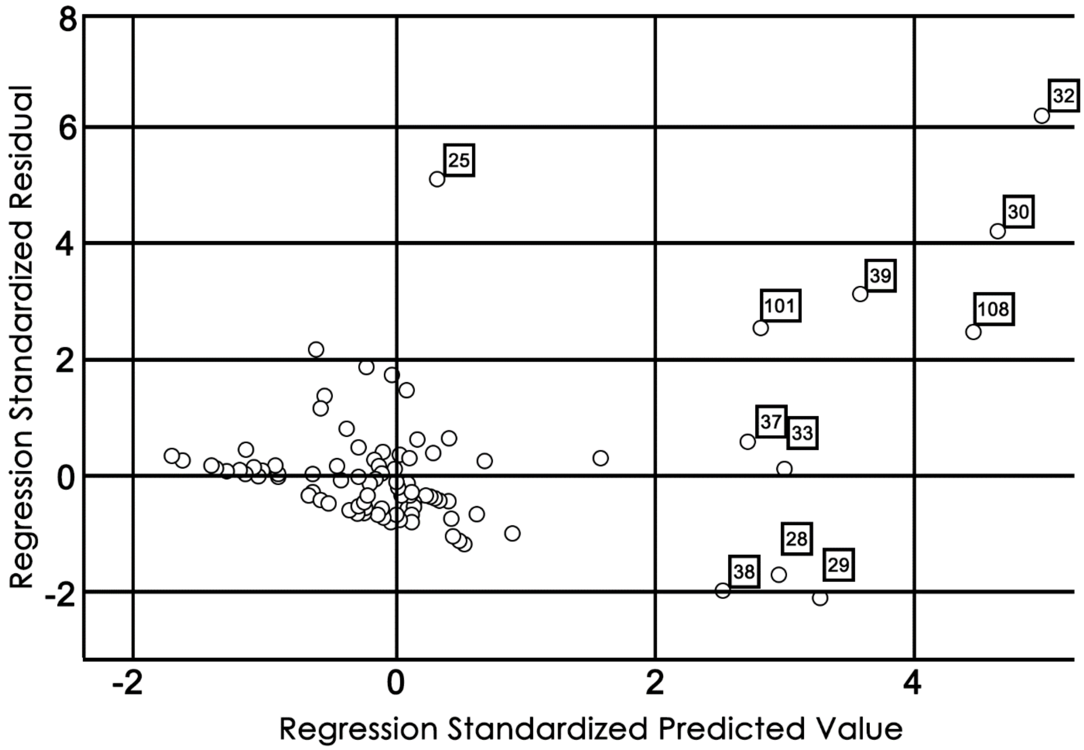

Figure 13. Eleven Lots That Appear to Be Outliers in the Northern Virginia District

Develop a Nonlinear Model

Accordingly, returning to the hypothesis that a nonlinear relationship might be present, Model 6 uses a second order transformation for all 108 park and ride lots such that the dependent variable was altered from occupancy to the square root of occupancy. Generally, this approach was successful in the sense that Model 6 had constant variance and the testing error, although large, was still less than the mean occupancy. The increase in the adjusted R^2 (0.45) showed that the model offered some explanatory power. The standard error is in the units of the dependent variable, so the smaller standard error associated with Model 6 compared to Model 5 reflects the fact that Model 6 gives error in terms of the square root of occupancy whereas Model 5 gives error in terms of occupancy. For instance, a Model 6 error of 8 might correspond to a Model 5 error of 64.

Apply a Nonlinear Model to Lots with Homogenous Characteristics

The research team then looked closely at the 11 lots that had been removed from the dataset in Model 5 and sought to identify manually any VRE lots, any Metro lots, and any lots with a capacity of more than 2,000 spaces, resulting in an additional 10 lots being removed, all of which were VRE lots, as shown in the latter portion (Last 10 lots) of Table 14. It is not surprising that models for the 87 lots in the Northern Virginia District might not adequately model demand at the Metro and VRE lots: others (e.g., Shirgoakar and Deakin, 2005) found that characteristics of users differ between general park and ride lots and those that exclusively serve rail facilities.

Table 14. Park and Ride Lots Considered for Removal in the Northern Virginia District

Basis for Removal	Park and Ride Lot	Capacity	Occupancy
Removed from Model 5 based on observations of residuals for Model 4	Ballston Public Parking Garage	2800	1914
	Dunn Loring-Merrifield Station Metro	2083	1076
	East Falls Church Metro	439	404
	Eisenhower Avenue Metro	625	196
	Franconia/Springfield WMATA/VRE	5144	2979
	Horner Road Commuter Lot	2293	2249
	Huntington Avenue Metro	3616	2315
	Van Dorn Street Metro	383	318
	Vienna-Fairfax-GMU Station Metro	4667	3795
	West Falls Church Metro	2058	1178
	Wiehle-Reston East Metro Station Garage	2300	2,295
Removed from Model 5 based on observations of residuals for Model 4 and also removed from Models 7 and 8 because of the presence of Virginia Railway Express (VRE)	Backlick Road VRE	220	215
	Broad Run/Airport VRE	1065	947
	Burke Center VRE	1516	929
	City of Manassas Park VRE	600	578
	City of Manassas VRE, 4 Lots	878	673
	Lorton Commuter Rail (VRE)	683	643
	Quantico VRE	195	244
	Rippon VRE	676	497
	Rolling Road VRE	368	401
	Woodbridge VRE	738	570

Then, a nonlinear model for the remaining 87 lots was developed, shown as Model 7. The net impact of both eliminating the 21 lots and using the nonlinear model is certainly preferable to doing neither: when Models 4 and 7 are compared, although the percent of variance explained is similar, the Model 7 shows a better fit in terms of the variance of the residuals being constant. The necessity of the nonlinear model is also confirmed by comparing Model 7 to Model 8, which similarly removes the 21 lots but uses a linear model: again, non-constant variance is observed in Model 8.

None of the six criteria, however, can justify choosing Model 7 over Model 6. Both nonlinear models have a similar coefficient of determination (around 51% or 43%), good residual plots, some independent variables that are reasonable (e.g., transit service and closeness to an interstate access point all increase occupancy), and relatively small intercepts (most of the occupancy comes from the variables rather than the intercept). Although the standard error and mean testing error associated with Model 7 were smaller than with Model 6, the mean lot occupancy was also smaller for Model 7. Thus, the best reason for continuing with Model 7 is that it excludes lots with fundamentally different service characteristics—mostly serving Metro and VRE.

Segment the Nonlinear Model for Lots With Homogenous Characteristics

The review of Model 7 also raised three concerns of practicality. The first was the provision of covered bicycle parking—the large coefficient caused the TRP to wonder if this might be a surrogate for some other phenomenon. The second was the use of two transit-related variables. To address the first concern, the binary variable BikeParkingIsCovered was replaced with the number of bicycle spaces (thereby making it a continuous variable). The second concern was the use of two transit-related variables. To address this concern, the variable TransitServiceAvailable was removed such that the only transit variable that remained was the number of transit providers. The third was the high occupancy rates relative to capacity; for some lots, for example, the Dumfries Road Commuter Lot at U.S. 1 and Route 234, occupancy (917) was almost equal to capacity (921), which suggests there may be latent demand such that occupancy would be higher. Although it is not possible to know this latent demand exactly, a presence of latent demand variable was added where the variable had a value of 1 if demand was 95% of capacity or higher; 11 of the 87 lots met this criterion. The result was Model 9, applicable to 87 lots. Both the number of bicycle spaces and the number of transit service providers were significant, although the presence of latent demand was not significant and thus was not included in the model. The mean testing error was reduced by about one-half (relative to Model 7), and the standard error also decreased.

Model 9 was then segmented into two separate submodels—one for the 78 lots with transit service (Model 10), and one for the 9 lots without transit service (Model 11). Although the mean testing error was larger for Model 10, this segmentation appeared appropriate given the very different occupancies for the lots in the two groups. When transit service is available (Model 10), key variables include number of bicycle spaces, number of transit service providers, and ADT. When transit service is not available (Model 11), none of these variables affected demand—rather, the explanatory variables were purely geographic (distance to the second largest employment center and distance to the nearest lot).

Consider a Model for the Excluded Lots

Additional experiments with developing a model for just the 21 excluded lots, however, were unsuccessful. No model for the 21 lots (Table 14, mostly VRE and Metro) could be developed that was statistically significant—that is, no independent variables remained in the model even if the intercept was removed and the model was forced through the origin. Accordingly, it appears that any preference for Model 6 over Models 10 and 11 does not reflect an improved ability to forecast demand for the 21 excluded lots. Thus, Models 10 and 11 are recommended for the 87 remaining lots in the Northern Virginia District, with an acknowledgment that at this point in time no good model exists for the 21 lots shown in Table 14. A map of these lots is provided in Figure A5 in the Appendix.

Richmond District

For the Richmond District, Model 4 is the best model, meeting five of the six criteria: less than one-half (actually about 10%) of the information contained in the intercept; independent variables that are logical (e.g., an increase in volume or congestion is associated with an increase in occupancy); a relatively high percent of variation explained (86.7%); a standard error that is less than one-half the mean occupancy; and a model that is unbiased and homoscedastic. In this district, however, the mean occupancy is much higher than the median occupancy of 40: neither this model nor many of the other models except Model 7 have a standard error that is less than one-half the median occupancy. The mean testing error of Model 4, however, is almost two-thirds the mean occupancy, meaning it does not meet the last criterion. As a practical matter, Model 4 simplifies implementation as it is applicable to the entire district. However, the last row of Table 15 shows the limits of a small dataset.

Table 15. Candidate Models for Park and Ride Lots in the Richmond District (11 Lots)

No.	Model (Limitations if not applicable to the entire district)[a]	Adj. R^{2a}	Std. Error[a]	Mean |Error|[b]	Mean Occ.[a]	Residuals[a]
1	-0.005 * $Vpeak_{PHF}$ + 0.001 * $Vprime_{PHF}$	0.549	73.411	77	72	Unbiased/ Heteroscedastic
2	-0.039 * $Vpeak_{K}$ + 0.011 * $Vprime_{K}$	0.527	75.150	80	72	Unbiased/ Heteroscedastic
3	-5.997 + 0.008 * Average ADT	0.637	51.881	45	72	Unbiased/ Homoscedastic
4[c]	4.083 + 0.006 * Average ADT + 151.906 * PHEF	0.867	31.467	45	72	Unbiased/ Homoscedastic
5	-0.005 * $Vpeak_{PHF}$ + 0.001 * $Vprime_{PHF}$ (Richmond Regional PDC or TPO, 10 lots)	0.538	77.861	47	78	Unbiased/ Homoscedastic
6	-0.039 * $Vpeak_{K}$ + 0.011 * $Vprime_{K}$ (Richmond Regional PDC or TPO, 10 lots)	0.516	79.698	55	78	Unbiased/ Homoscedastic
7	6.361 + 0.006 * Average ADT + 152.161 * PHEF (Richmond Regional PDC or TPO, 10 lots)	0.857	33.324	38 or 113	78	Biased/ Heteroscedastic

Occ. = occupancy.
[a] Model, adjusted R^2, standard error, mean occupancy, and residuals are based on the entire dataset.
[b] |Mean error| is based on applying an earlier model, slightly different from that shown, developed from just 70% of the data to the remaining 30% testing dataset.
[c] Model is recommended for use.

Models 4 and 7 are similar in terms of coefficients, independent variables, adjusted R^2, and standard error. The existence of Model 7 shows the limitations of assessing model performance with a small dataset. The area represented by the Richmond TPO and the area represented by the Richmond PDC each encompasses 10 of the 11 Richmond District lots; they exclude the Appomattox River lot. The research team tested the mean error for Model 7 twice: after determining the model shown in Table 8, 3 lots were removed randomly; a new model was recalibrated based on the remaining 7 lots; and then the mean error of that model, as applied to the 3 lots not used in model development, was determined. The first time, the mean error was 38; the second time, it was 113. To be clear, this difference resulted simply from the random selection process. The small number of samples can thus cause this variation, as well as inhibit a determination of the residual plot.

Salem District

Table 16 shows that Model 4 is potentially useful if one considers only statistical indicators: high adjusted R^2; standard error and mean testing error that are one-fifth and one-fourth of mean occupancy, respectively; and a good plot of residuals.

Table 16. Candidate Models for Park and Ride Lots in the Salem District (15 Lots)

No.	Model (Limitations if not applicable to the entire district)[a]	Adj. R^{2a}	Std. Error[a]	Mean |Error|[b]	Mean Occ.[a]	Residuals[a]
1	0.003 * $Vpeak_{PHF}$ + 0.000434 * $Vprime_{PHF}$	0.598 [a]	19.827	9	20	Unbiased/ Heteroscedastic
2	0.03 * $Vpeak_K$ + 0.004 * $Vprime_K$	0.583	20.205	10	20	Unbiased/ Heteroscedastic
3	-9.339 + 0.007 * Average ADT	0.450	18.367	8	20	Unbiased/ Heteroscedastic
4	-221.679 + 59.902 * TransitServiceAvailable – 0.014 * TransitRiders5 + 260.055 * PHF	0.960	4.924	4	20	Unbiased/ Homoscedastic
5	-0.018 * $Vpeak_{PHF}$ + 0.001 * $Vprime_{PHF}$ (Roanoke Valley TPO, 4 lots)	0.713	21.233	4	32	Biased/ Homoscedastic
6	-0.150 * $Vpeak_K$ + 0.009 * $Vprime_K$ (Roanoke Valley TPO, 4 lots)	0.778	18.705	4	32	Biased/ Homoscedastic
7	-2214.482 + 90.034 * AvgPctOFRentIncomeOnRent5 (Roanoke Valley TPO, 4 lots)	0.994	2.081	1	32	Unbiased/ Homoscedastic
8[c]	2.340 + 67.177 * TransitServiceAvailable + 0.000161 * POP5 (New River Valley PDC, 8 lots)	0.981	3.809	3	18	Unbiased/ Homoscedastic
9[c]	5.849 + 0.015 * Rad2_Jobs25_to_50Mi + 48.731 * TransitServiceAvailable (Roanoke Valley Alleghany Regional Commission, 6 lots)	0.881	8.4	14	25	Unbiased/ Homoscedastic

Occ. = occupancy.
[a] Model, adjusted R^2, standard error, mean occupancy, and residuals are based on the entire dataset.
[b] |Mean error| is based on applying an earlier model, slightly different from that shown, developed from just 70% of the data to the remaining 30% testing dataset.
[c] Model is recommended for use. The Salem District lot that is not covered by these models is reflected in the Mount Rogers PDC model associated with the Bristol District.

However, the contradictory impacts of transit in the coefficients (a higher occupancy for transit service being available yet a lower occupancy for the number of riders) is disconcerting. Further, given that the PHF tends to be within a tight range of 0.88 to 1.0 with a common value of 0.95, there exists the possibility that the intercept (-221.679) plus the term 260.055*PHF is itself responsible for most of the mean occupancy. Although these concerns are not fatal flaws, they suggest that either more explanation is needed or a different model, if available, is preferable.

Two PDCs (New River Valley with 8 lots and Roanoke Valley-Alleghany Regional Commission with 6 lots) account for 14 of the 15 Roanoke District lots (the 15th is in the Mount Rogers PDC). Models for the former two PDCs are shown in the last two rows of Table 16. Both models met five of the six criteria: relatively good adjusted R^2 (88% or higher); standard error below one-half the mean occupancy; good residual plots; low intercepts; and independent variables that are reasonable (e.g., transit service, population within 5 miles of the lot, and persons living nearby who have jobs that are 25 to 50 miles away are all associated with an increase in occupancy). Model 8 met a sixth criterion (mean testing error is low), and the testing error for Model 9 was slightly larger than one-half the mean occupancy. Accordingly, Models 8 and 9 are recommended in lieu of Model 4; the 1 lot not included therein is covered by the Mount Rogers PDC model, which also was recommended as part of the Bristol District lots. The small number of lots in Models 7, 8, and 9 requires some judgment about the evaluation of performance. Regarding Model 9, a case can be made that there are some residuals—that is, difference between forecast and predicted values for the testing dataset—that are not within 3 standard residuals, meaning they are outliers and are contributing to the larger testing error. However, the small sample size of four lots for training and two lots for testing indicates that as a practical matter, any further reduction in the dataset means that one is simply developing a site-specific model. Although Model 7 may also be considered, the fact that it is applicable for just four lots means that it is functioning also as almost a site-specific model where the sole determinant of occupancy is the percent of income spent on rent.

Staunton District

Models 4, 5, and 6 in Table 17 are all potential candidates for forecasting demand. All three models met three criteria: the models explain a large amount of scatter in the dataset (adjusted R^2 is above 80%), the plots of the residuals are unbiased and homoscedastic, and the mean testing error is less than one-half the mean occupancy. In terms of the fourth criterion, Models 5 and 6 have lower standard errors than Model 4, and this difference is more pronounced considering that the mean occupancy for the lots that are the subject of Model 6 is 50% higher than for the lots that are the subject of Model 4. The intercept is smaller for Model 5 (a desired trait) than for Models 4 and 6 (a larger negative value), but nonetheless more information is contained in the variables than in the intercept in terms of forecasting occupancy.

Table 17. Candidate Models for Park and Ride Lots in the Staunton District (25 Lots)

No.	Model (Limitations if not applicable to the entire district) [a]	Adj. R^{2a}	Std. Error[a]	Mean |Error|[b]	Mean Occ.[a]	Residuals[a]
1	0.003 * $Vpeak_{PHF}$ + 0.364 * $Vprime_{PHF}$	0.454	31.510	41	26	Unbiased/ Heteroscedastic
2	0.035 * $Vpeak_K$ + 0.006 * $Vprime_K$	0.425	32.336	41	26	Unbiased/ Heteroscedastic
3	-6.841 + 0.009 * Average ADT	0.485	23.727	21	26	Unbiased/ Heteroscedastic
4	-68.401 + 56.617 * TransitServiceAvailable + 2.215 * CommuteTime2 + 0.011 * EMP 2.5	0.807	15.020	9	26	Unbiased/ Homoscedastic
5[c]	0.516 + 0.027 * Rad2_JobsGT50Mi + 0.004 * Carpoolers10 (Central Shenandoah PDC, 13 lots)	0.893	6.178	6	15	Unbiased/ Homoscedastic
6[c]	-136.688 + 0.024 * Rad5_JobsLT10Mi + 3.516 * CommuteTime5 (Northern Shenandoah PDC, 12 lots)	0.898	13.609	13	39	Unbiased/ Homoscedastic

Occ. = occupancy.

[a] Model, adjusted R^2, standard error, mean occupancy, and residuals are based on the entire dataset.

[b] |Mean error| is based on applying an earlier model, slightly different from that shown, developed from just 70% of the data to the remaining 30% testing dataset.

[c] Model is recommended for use.

The decision as to whether to use a district model (Model 4) or the two PDC models that comprised the district (Models 5 and 6) hinges therefore on how one interprets the variables in the model. For the Central Shenandoah PDC (Model 5), these variables suggest a lot's occupancy is a function of its attractiveness to carpoolers, given the two variables of the number of commuters traveling a long distance (more than 50 miles) and the number of carpoolers. For the Northern Shenandoah PDC, one seemingly related variable (the length of the commuting time) and one very different variable (the number of jobs nearby) influenced demand. Thus. the two models appeared to identify characteristics that were applicable to each PDC. The benefit of this segmentation by PDC was also supported by the fact that two of the three variables in Model 4 (one relates to commute time and one relates to employment) are similar to the two variables in Model 6 and yet Model 6 has a lower standard error than Model 4, even with a higher mean occupancy. Both PDC models met all six criteria.

It was initially surprising that Model 4 showed that the availability of transit service was a significant determinant of lot availability whereas this variable was missing in Models 5 and 6. However, of the 25 lots in the Staunton District, only 1 (Crooked Run) had transit service, and the occupancy of that lot was 131—higher than any other lot in the district and roughly 30% higher than the next highest lot, with an occupancy of 103. After segmentation by PDC, Model 6 suggested that other variables for the PDC in which this lot is located, such as commute time (where the lot has the fourth highest commute time in the PDC), suffice for explaining occupancy. Thus, Models 5 and 6 are recommended for this PDC.

DISCUSSION

The results may be considered across six dimensions: (1) the failure of existing models when applied to Virginia sites; (2) the need for site-specific factors to forecast occupancy successfully; (3) the infeasibility of time series modeling; (4) the use of the models for forecasting demand; (5) the impacts of uncertainty on reported results; and (6) shorter-term future research needs.

The Failure of Existing Models for Virginia Sites

The application of the existing diversion model generally yielded results that could not be used in Virginia. Some clues about why this model failed are evident from both the literature and the earlier experience of TRP members. FDOT (2012) stated that the approach is best applied in areas where there is a limited number of commuting roadways because as the number of commuting roads increases, the accuracy of the forecast demand will decrease. TRP members had found that a modified version of the diversion model used by FDOT (2012) overestimated demand (in the Fredericksburg District) and underestimated demand (in the Staunton District). That modified version adds or subtracts an average error to the forecast based on experiments with several lots in the same district. Thus, it is not surprising that direct application of the existing diversion model could not work well in the remaining seven districts. This observation was not limited to Florida: the research team also considered adapting a model developed for

Washington State's King County (Spillar, 1997); however, the variables used in the King County model showed very different correlations than those found for the comparable Northern Virginia District variables.

The second set of models that calibrated parameters *a* and *b* (where *a* was the coefficient for the peak period traffic volume for the road(s) that provided a direct entrance to the lot and *b* was the coefficient for the peak hour traffic volume for the road within 2.5 miles of the lot that had the highest ADT) yielded better results. However, the testing error for such models remained large: in only four districts (Bristol, part of Fredericksburg, Hampton Roads, and Salem) was the mean testing error less than the mean occupancy. Of the 22 recalibrated models (two for each district except four for Fredericksburg and Hampton Roads districts, which were subdivided into high and low population density areas), the *p*-values were not significant except for 3 of the models, which were in the Bristol, Lynchburg, and Staunton districts. Further, for those 22 models that relied exclusively on $Vpeak_K$, $Vprime_K$, $Vpeak_{PHF}$, or $Vprime_{PHF}$, the median ratio of mean testing error to mean occupancy was 1.1 (an example is the Richmond District with 1 such model that incorporated the K-factor, giving a mean testing error of 80 compared to a mean occupancy of 72).

Interestingly, the two variations of the recalibrated diversion model used by the research team did not have a material difference. As discussed earlier, the original diversion model employed by FDOT (2012) used the K-factor to calculate $Vpeak_K$ and $Vprime_K$. Because it was initially easier for the research team to obtain the PHF, the team replaced the K-factor with the PHF in the recalibrated method, obtaining $Vpeak_{PHF}$ and $Vprime_{PHF}$. The research team had postulated that possibly the PHF could be a surrogate for congestion. Although the correlation between the K-factor and the PHF was strikingly low (-0.03), the correlation between $Vpeak_K$ and $Vpeak_{PHF}$ was 0.992 ($p < 0.01$) based on the 297 park and ride lots in Virginia, suggesting that the key discriminator in the second set of models was ultimately the traffic volumes—both the ADT adjacent to the lot (e.g., Vpeak) and the ADT for the heaviest traveled facility within 2.5 miles of the lot (e.g., Vprime).

The Need for Site-Specific Factors

The third, fourth, and fifth sets of models revealed that some variables have limited explanatory power. That is, occupancy is linearly correlated with certain variables (or certain combinations of variables), but this occupancy is influenced by site-specific factors not included in the model. One clue that such factors are missing is the pattern displayed by the residuals: if the sign of the error is unevenly distributed (e.g., biased such that residuals tend to be positive or negative) or if size of the error is uneven (e.g., heteroscedastic such that the variance changes as the size of the prediction changes), then the reason might be that a key variable is missing. In part through judicious selection of independent variables, the research team was able to develop models that were unbiased and homoscedastic with one exception: the model for the 30 lots in the Fredericksburg District that were in low population density areas (about 10% of Virginia's lots) as shown in Table 10 (see Model 9).

The value of replacing the VDOT districts in some cases with other appropriate commuting boundaries, such as the PDC or the MPO, was demonstrated in five of the nine VDOT districts—Bristol, Culpeper, part of Fredericksburg, Salem, and Staunton—where the TRP had suggested that perhaps in some locations, but not all, park and ride lot demand was influenced by long commutes. This was indeed the case for the park and ride lots in the Bristol District, where demand for lots in the Lenowisco PDC was associated with the number of commuters having a job more than 50 miles away, yet demand was not associated with far-flung commutes for lots in the Cumberland Plateau and Mount Rogers PDCs. In several of the PDC submodels, different factors explained occupancy than were thought to be the case with the district model; for instance, for the Staunton District, the district-wide model suggested transit service was a differentiator. However, when this district model was disaggregated into two submodels that aligned with the PDCs, it was found that commute time and commute distance were better indicators of occupancy.

If the models explained all of the variations in the observed occupancy, one would expect certain variables always to have a positive coefficient, such as the traffic variables, and the intercept always to be zero. However, in some cases, the best performing model nonetheless had variable signs that appeared counterintuitive. This means that additional factors not captured by the model explained some portion of the variation in occupancy rates. Although one can formulate possible reasons for these counterintuitive variables, one cannot prove them with the model alone. For example, the best performing model for the Rappahannock-Rapidan PDC showed a negative coefficient for the number of carpoolers within 2.5 miles of the lot (unexpected) along with a positive coefficient for the number of transit riders (expected). The model by itself did not indicate why this negative coefficient resulted. (One possible but unproven reason is that, at this particular site, an increase in carpools corresponds with a reduction in transit riders. Based on the 2001 National Household Travel Survey, Plotz et al. [2010] found that almost half (48%) of work-related HOV2 trips were made by members of the same household. If such an observation were applicable at this particular site, then one would expect the substitution of carpoolers for transit riders to reduce demand for spaces at the lot. However, the model cannot show causation and thus cannot prove this reason is valid.) That said, often the signs were plausible. For example, one may consider the best performing model for the high population density portion of the Hampton Roads District where the coefficient was positive for persons living within 2.5 miles of the lot where their jobs are 25 to 50 miles away. Logically, one would expect an increase in commuters with relatively far-off jobs to increase park and ride lot demand.

The models also revealed some surprises regarding the determinants of occupancy. Although several relationships were found as expected (e.g., commute length increases occupancy), it was surprising that variables representing distance to employment centers did not have a significant impact on occupancy as reflected in the models with the exception of a small number of lots that did not offer transit service and that were located in the Northern Virginia District. An example of those six variables is given in Figure 14 and Table 18 for one particular park and ride lot: Glenside Drive-Dumbarton.

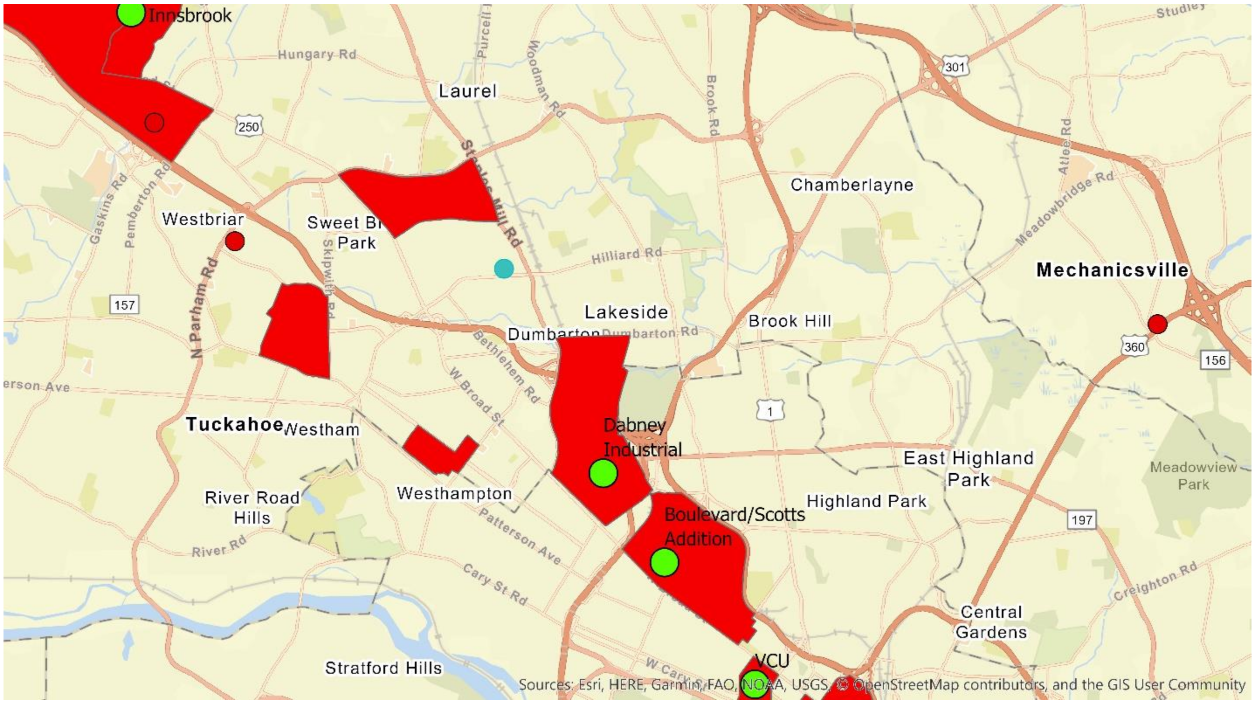

Figure 14. Example of Distances Available for the Glenside Drive-Dumbarton Park and Ride Lot (shown in blue). The lot is 2,769 ft from a block group of 5,000 jobs (shown in red) and is very close to the Dabney Industrial Major Activity Center and the Boulevard/Scott's Addition Major Activity Center.

Table 18. Examples of Distance Variables for the Glenside Drive-Dumbarton Park and Ride Lot[a]

Variable Name	Explanation	Value (ft)
DIST_BG_BigEmp	Distance to the nearest block group with at least 5,000 employees	2,769
Dist_M1	Distance to the VTrans Employment Center for the "Dabney_Industrial"	13,345
Dist_M2	Distance to the VTrans Employment Center for the "Boulevard_Scotts_Addition"	19,644
Dist_M3	Distance to the VTrans Employment Center for "Innsbrook"	26,561
Dist_M4	Distance to the VTrans Employment Center for the "VCU"	28,499
Dist_Weight	Sum of the square roots of the above four distances (Dist_M1, Dist_M2, Dist_M3 and Dist_M4)	587

[a] Names in quotation marks are regions as specified by CDM Smith (2020). Employment centers were calculated by the research team based on an overlay of data from CDM Smith (2020) with block groups of employment of 10,000 or more.

This finding does not mean that employment centers do not affect park and ride lot demand generally; however, that variable, as captured by the research team in this study, was not usually shown to be a predictor of the number of occupied spaces as found by VDOT's audit. For instance, a review comment regarding this material was that the Glenside Drive-Dumbarton lot is heavily used for vanpools that head toward Northern Virginia and Washington, D.C. That particular characteristic, which, makes this lot attractive for those vanpoolers was not explicitly captured by the variables in Figure 14.

Infeasibility of Time Series Modeling

The large number of lots in the Northern Virginia District, coupled with additional data at those locations, offered a unique opportunity to consider a different modeling approach: building a time series model. The errors shown for Model 7 in Table 13, with an average value of 164 spaces, suggested that one might have success with a model that simply used historical occupancies to forecast a future year occupancy. Examination of the historical data, however, suggested that a time series model was not appropriate with this particular dataset.

Seven of the 108 lots have annual data for each year from 1997-2018 inclusive, as shown in Figure 15. However, there does not appear to be a clear trend for the lots as a whole. For example, the Stone Road lot at U.S. 29 in Centreville saw occupancy almost double from 1997-2007, plateau from 2007-2010, and then decline steadily by almost 20% from 2010-2018. By contrast, occupancy of the Fairfax County Government Center lot declined for most of the observation period, with the 2016 occupancy being about one-third of the 1997 value, but then increased dramatically from 2016-2018. Although there may be some underlying patterns (e.g., declines in most lots are evident from 2011-2016), there was no obvious trend that appeared suitable for forecasting without additional investigation.

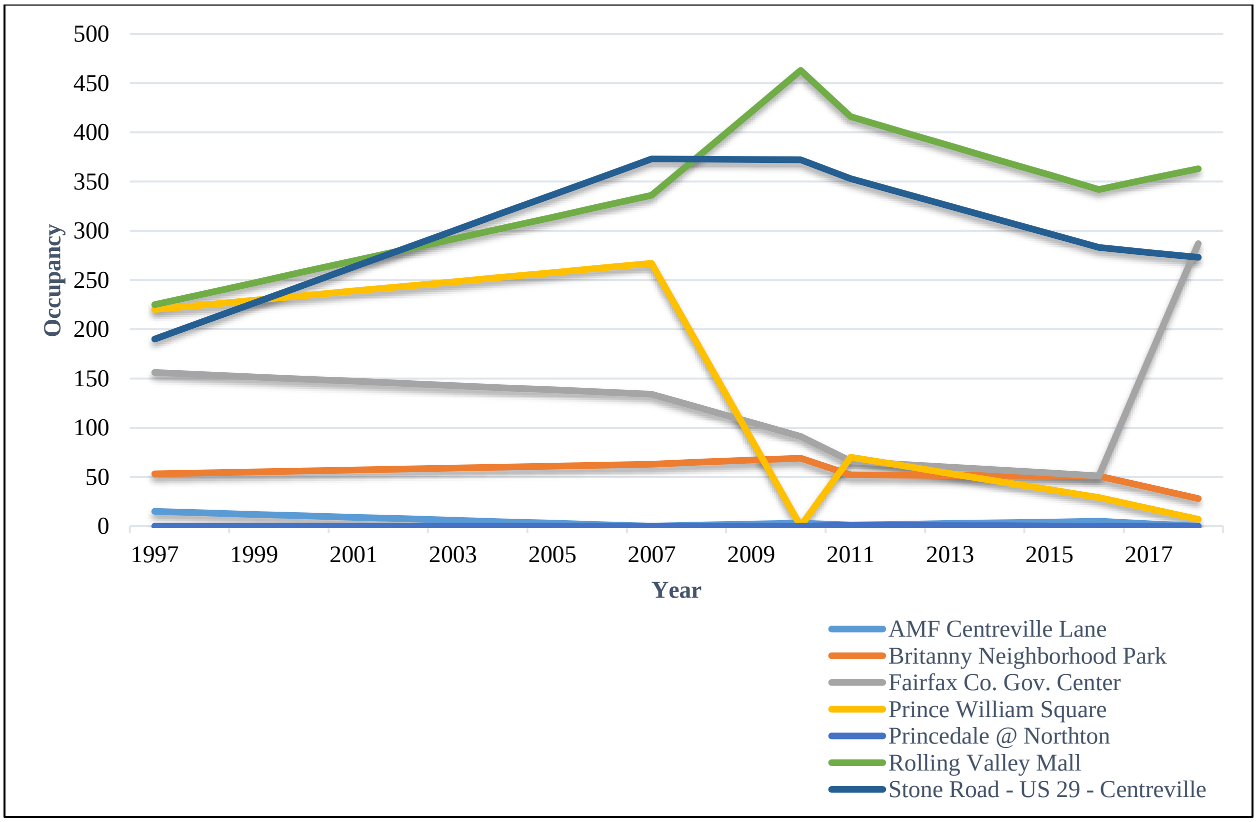

Figure 15. Yearly Occupancy Change for 7 Park and Ride Lots in the Northern Virginia District (1997-2018)

Figure 16 displays the estimated marginal means (where in this case there were no covariates such that the estimated marginal means were simply the occupancies) for 108 Northern Virginia District park and ride lots for the years 2010, 2011, 2016, and 2018. No trend was evident. This study did not explain the fluctuations shown in Figures 15 and 16: one possible reason is that the collection of demand values once per year contributed to these fluctuations, but it is also possible that there were other factors, such as construction, that affected demand at each lot. Figures 15 and 16 suggest that a time series model is not appropriate for this particular dataset.

Presumably, a variety of factors can influence occupancy: increases in transit ridership, improvements in the economy, and expansion of lots (as was the case with the 2018 expansion of the Fairfax County Government Center lot) may increase occupancy. For these reasons, graphs such as Figures 15 and 16 that show changes in occupancy will not by themselves indicate model feasibility. However, in part because the graphs also show decreases in occupancy, some of which are severe, the problem identified by the TRP and also noted in Mouskos et al. (2007) of collecting data only once per year (or less frequently) appears to be a partial reason for why a time series model may not be appropriate based on this dataset.

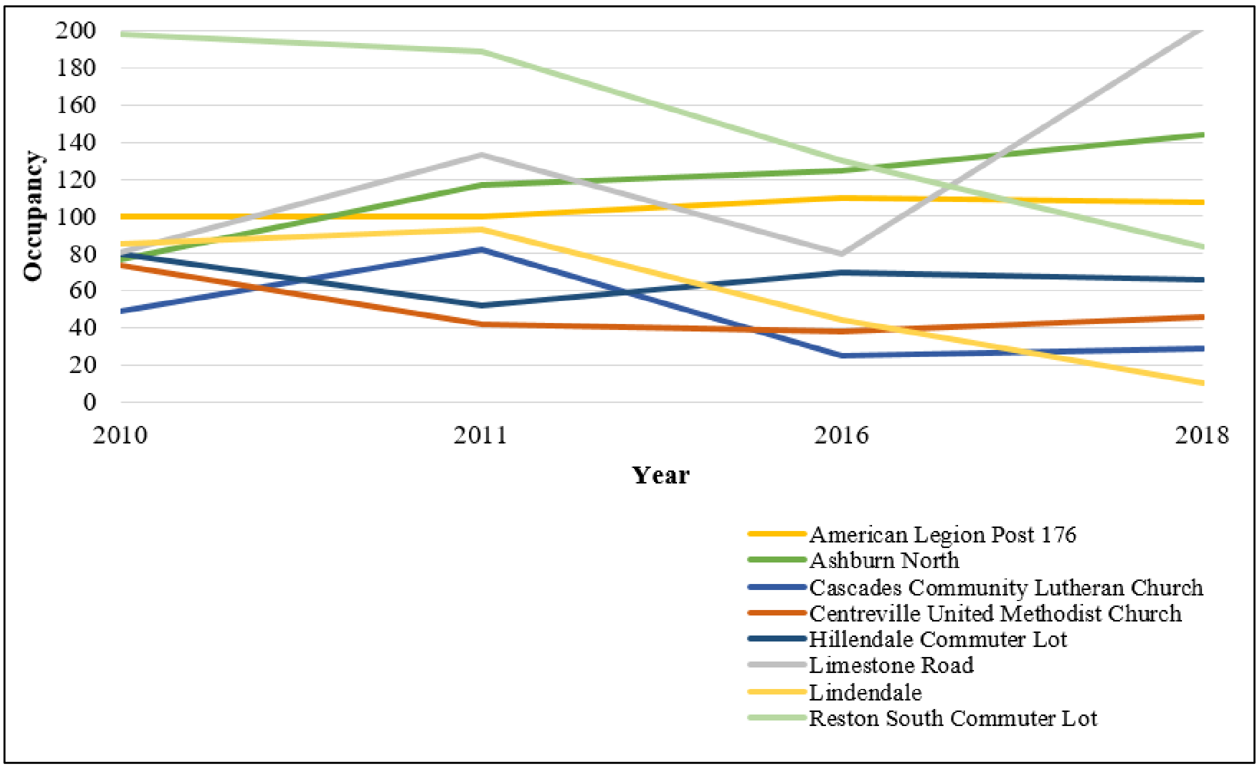

Figure 16. Occupancy Values in 2010, 2011, 2016, and 2018 for Select Northern Virginia District Lots. The Limestone Road Lot occupancy in 2018 was 202.

Using the Models for Forecasting Demand

There are two different ways to apply the forecast occupancy for each of the district models. One way is suitable for forecasting demand for a new lot that will be built, and the other way is suitable for forecasting demand for an existing lot as a function of some change that will occur in the future. Both ways require judgment concerning the appropriateness of the independent variables.

Demand Forecast for a New Lot

To use the model to forecast demand at a new lot, one can use the appropriate model shown in Tables 8 through17. For example, one could suppose that a new park and ride lot will be built in a low density location in the Fredericksburg District. The park and ride lot will have lighting but no transit service, will be in a Census block group with 1,000 people per square mile, and will have the nearest facility for which a peak hour expansion factor (PHEF) is available showing a value of 1.5. The forecast occupancy for such a lot comes from Table 10 (see Model 9 therein) and implemented as Equation 20.

$$
\begin{aligned}
&= -217.053 + 241.839 * \text{Lighting} + 329.448 * \text{NuofTranServicePP} + 0.018 * \text{POPDEN} \\
&\quad + 67.016 * \text{PHEF} \\
&= -217.053 + 241.839 * (1) + 329.448 * (0) + 0.018 * (1{,}000) + 67.016 * (1.5) \\
&= 143 \qquad \text{[Eq. 20]}
\end{aligned}
$$

Demand Forecast for a Change That Will Affect an Existing Lot

It is also possible to use the models to examine how changes in certain variables might affect occupancy at existing locations. For example, the Occoquan Commuter Lot (Old Hechinger's) can be considered. If the number of buses serving the lot doubles and the average ADT is forecast to increase 10%, what is the expected impact on occupancy? As this lot is in the Northern Virginia District, the suggested model is in Equation 21.

$$
(2.488 + 0.298*\text{Bicycle spaces} + 0.396*\text{NuofTranServicePP} + 0.001 * \text{Average ADT})^2 \qquad \text{[Eq. 21]}
$$

Presently, there are four transit lines and 12 bicycle spaces; the mean ADT for roads within 2.5 miles of the lot is 8,733. Thus, the forecast occupancy is 268, as shown in Equation 22.

$$
\begin{aligned}
&\text{Forecast demand (existing conditions)} \\
&= (2.488 + 0.298*12 + 0.396*4 + 0.001 * 8733)^2 \\
&= 268 \qquad \text{[Eq. 22]}
\end{aligned}
$$

With a 10% increase in ADT (from 8,733 to 9,606) and a doubling of transit service (from four to eight buses), the forecast occupancy is 355, as shown in Equation 23.

$$\text{Forecast demand (new conditions)} = (2.488 + 0.298*12 + 0.396*8 + 0.001 * 9606)^2$$
$$= 355 \qquad \text{[Eq. 23]}$$

However, at this lot, demand is presently 437. Thus, the new demand is forecast to be roughly 578 (Equation 24).

$$\text{New demand} = \text{Current demand} * \frac{\text{Forecast demand (new conditions)}}{\text{Forecast demand (existing conditions)}}$$
$$\text{New demand} = 437 * \frac{355}{269} = 578 \qquad \text{[Eq. 24]}$$

Thus, the increase in transit and ADT may lead to an increase in occupancy from 437 to 578. In this case, the change in demand for the existing facility is proportionate to the square of the change in transit service and ADT, owing to the nonlinear model recommended for the Northern Virginia District.

Necessary Judgment When Using These Models for Forecasting

One weakness of these models is that they cannot prove causality. The appeal of some of the simpler models, such as the diversion models used by FDOT (2012), is that the presumed relationship is at least plausible, especially for the simplest model where occupancy is simply some percentage of passby traffic. For some of the multivariate models herein, the relationship is at least plausible, such as an increase in the number of carpoolers living within 10 miles of the lot being associated with an increase in occupancy, as was the case in the Central Shenandoah PDC.

In other cases, however, especially with small datasets, some of the observed variables may represent other phenomena. For instance, in the Charlottesville-Albemarle MPO model, an overnight parking allowance increases occupancy by 28. Members of the TRP (Olivia Mobayed, personal communication, December 21, 2020) noted that this variable might be a surrogate for other factors such as (1) more signage (which might instill greater confidence from patrons); (2) the ability to use the lot for other purposes (e.g., some commuter lots are located at parks or other attractions); and (3) latent demand. As another example, it is possible that the number of bicycle spaces (in the Northern Virginia District) or the presence of covered bicycle parking (in the Hampton Roads District) represent these intangible elements.

For these reasons, care should be exercised when using these models. For example, although the number of transit lines serving the lot was shown to influence demand, one would not expect the addition of lesser used service to have the same impact as a fully used service. Thus, if new transit service was proposed and one wanted to update the forecast, one would consider whether the new service was likely to be used to a similar extent as existing services.

Impacts of Uncertainty in Reported Results

The development of multiple models provides an opportunity to conduct a rough uncertainty analysis at each lot. For example, one can also consider another model developed for the Northern Virginia District park and ride lots that was viewed as credible but not as useful as the previous model (Equation 21) because it is not limited to lots with transit service. That model is shown as Equation 25:

$$\text{Occupancy} = (-7.614 + 0.330 * \text{Bicycle spaces} + 0.616 * \text{NuofTranServicePP} + 0.586 * \text{RentOverAllIncome})^2 \qquad \text{[Eq. 25]}$$

For the same park and ride lot (Occoquan Commuter Lot), the median monthly rent multiplied by 12 and divided by median household income in the 2.5-mile radius catchment area is 21.79. If transit service were forecast to double and there will be a modest increase in rent relative to income such that the value of 21.79 becomes 24, Equation 25 yields a forecast of 134 under existing conditions and 235 for the new conditions. Further, the ratio of the latter to the former multiplied by the existing demand of 437 would suggest a new occupancy of 767, as shown in Equation 26:

$$\text{New demand} = 437 * \frac{235}{134} = 767 \qquad \text{[Eq. 26]}$$

The range of forecasts—that is, that an increase in transit service, ADT, and rent relative to income suggests an occupancy changing from a current value of 437 to either 578 or 767—reflects appropriately the uncertainty with these forecasts. On the one hand, transit service was a significant determinant of occupancy. On the other hand, even the best model for the Northern Virginia District that covered this lot explained only 48% of the variance in occupancy, suggesting that the factors here, although significant, are not the only determinants of the observed values.

For the models developed herein, the uncertainty tends to increase as the sample size decreases—that is, the very decision to develop submodels that are tailored to VDOT districts (as opposed to a statewide model), or to develop submodels tailored to PDCs (as opposed to district-wide models)—increases the uncertainty of the information presented herein. For instance, the mean testing error as reported from two VDOT districts—one case where there were 10 sites and one case where there were more than 10 times that number.

- As discussed previously, Model 7 in Table 15 for the Richmond Regional PDC or Richmond TPO was examined twice. The full model was developed that showed the model should contain variables: an intercept, the average ADT, and the PHEF. Then, the model was recalibrated (e.g., the same variables were retained but the coefficients changed) based on only seven sites—and the new model was tested on the remaining three sites. Doing this once yielded an error of 113, but doing this a second time (e.g., choosing three sites at random again) yielded an error of 38, such that there was a 66% change in the mean testing error.

- An earlier version of one of the models for the 108 lots in the Northern Virginia District had included the Wiehle Road lot with an occupancy of 0 rather than 2,295, as well as an independent variable for “bicycle parking is covered” (which was later excluded from the dataset). That earlier version had also been developed for the Northern Virginia District portion of the National Capital Region MPO, which

> reflected 109 lots. The models were developed independently but had almost identical coefficients. There was, however, one interesting observation in that the mean testing error changed from 301 to 267 depending on which model was used—not because of differences in the model but because of differences in which sites were selected for training versus testing. In short, this particular exercise showed that randomly selecting a different set of sites yielded a change in the testing error of about 34/301 or about 11%—a smaller change attributable to the larger number of testing sites.

In some cases, the small datasets also hampered an exact determination of whether the residuals showed constant variance; for instance, as noted for the Richmond TPO model (Model 7 in Table 15 for the Richmond District) with only 10 sites, the research team could not determine if the model was homoscedastic or heteroscedastic.

Future Research Needs

There are at least two future research directions that could be pursued.

1. *As discussed in this report, there were 21 lots for the Northern Virginia District that mostly served Metro and VRE facilities. A follow-up effort that examines these 21 lots could consider ridership on the lines (used as a predictive variable by Peng and Mohamad [2005] who considered park and ride lots near light rail facilities) and mode choice extracted from the regional model (used by the I-95/I-395 Transit/TDM Technical Advisory Committee [2008]). Further, coefficients for each line could at least be tested for their predictive power.* Webb et al. (2021) sought to determine a model for forecasting which park and ride lot would be used by transit users and found that two key predictor variables were (1) lots where the transit time relative to the time to drive from home to the lot was as high as possible, and (2) the sum of the home to lot distance plus lot to destination distance was as close to the home to destination distance as possible. Although neither variable is surprising in theory, Webb et al. (2021) suggested that users might select the lot that best met these two goals even at the expense of choosing the lot that minimized total travel time. Thus, such psychological factors could be investigated in a transit-focused study.

2. *There may be a need to understand better how occupancy changes on a daily or weekly basis.* This need is acute for locations where occupancy may be approaching capacity, such as the nine Northern Virginia District lots where occupancy is 95% of capacity or higher, as such locations may be experiencing latent demand. With data collected more than once per year, one could also begin to assess whether year-on-year changes reflected structural changes in demand or simply random variation. In some cases these seemed minor (e.g., the database showed occupancy changed from 58 to 55 at a Henrico County lot (intersection of Parham Road and Fordson Road) between the 2016 survey and the 2018-2019 survey), but in other cases these changes were dramatic, such as the occupancy shift from 196 to 280 during the same period at another Henrico County lot (Gaskins Road and Mayland Drive).

CONCLUSIONS

- *It is infeasible to apply models from the literature (FDOT, 2012) directly to Virginia, as such models yielded errors from 14 to 141 times the mean occupancy.* Error is the difference between the forecast occupancy and the observed occupancy for a separate testing dataset that was not used to build the model. Table 7 showed that direct application of such models yielded errors that were typically dozens of times larger than the observed occupancy for each VDOT district. For example, for the Culpeper District, direct application of the existing diversion model gave a mean error that was 78 times higher than the mean occupancy. This finding that calibration was essential was also noted by Nungesser and Ledbetter (1987), who found that recalibration of the diversion method adapted by FDOT (2012) materially changed the diversion parameters.

- *Generally, the use of site-specific traffic-related variables can yield forecast errors that range from roughly 0.3 to 1.8 times the mean occupancy.* For example, in the Hampton Roads District, where the mean occupancy was 41 (Table 11), inclusion of the maximum ADT of any facility within 2.5 miles of the park and ride lot yielded a mean testing error of 14; hence, error/occupancy = 0.3. In the Culpeper District, however, the use of the same variable led to a mean error of 29, which was 1.8 times the district's mean occupancy of 16 (Table 9).

- *The inclusion of socioeconomic variables further reduced forecast errors in a minority of cases.* In the Bristol, Salem, and Staunton districts, the development of models with independent variables that considered characteristics such as rent relative to income, transit service availability, commute time, number of transit riders, or jobs (with the variables referring to either a 2.5-mile or 5.0-mile radius of the park and ride lot) helped reduce the ratio of the error/occupancy to values of 0.20 (Salem), 0.25 (Bristol), and 0.35 (Staunton). In the Fredericksburg District, including socioeconomic variables (e.g., transit ridership, population density, and number of transit lines serving the lot) and developing two separate models for the rural and urban areas based on population density helped reduce the error/occupancy ratio from 0.8 (when only traffic variables were used) to 0.2 (see Table 10).

- *The development of models tailored to some type of anticipated commuting boundary rather than an administrative boundary improved models in some cases.* For 114 (more than one-third) of the 297 park and ride lots, some element of model performance was improved by explicitly replacing the VDOT district boundary with the PDC boundary or the MPO boundary. Although such boundaries are not a perfect surrogate for commuting patterns, they can be more appropriate than the VDOT district boundary for envisioning the travel market in some locations. In the Culpeper District, the aforementioned error/occupancy ratio was reduced from 1.8 to 1.5 and 0.9 by splitting the district into models for the Rappahannock-Rapidan PDC and the Thomas Jefferson PDC. In other cases, the improvement was not as dramatic but was helpful: for example, using separate models for the PDCs of Lenowisco, Cumberland Plateau, and Mount Rogers yielded a substantially better adjusted R^2 and a lower standard error than the district model.

- *The best models usually met most, but not all, of the ideal criteria for a good model.* Six idealized criteria were used to judge models once statistically significant ($p \leq 0.05$) variables were built: (1) less than one-half of the explanatory power is in the intercept (as opposed to the independent variables); (2) coefficients are logical and do not pose an equity challenge for VDOT should they use the models to make investment choices; (3) the standard error is less than one-half the mean occupancy; (4) the testing error is less than one-half the mean occupancy; (5) the coefficient of determination (adjusted R^2) is higher than for other models; and (6) residuals demonstrate lack of bias and constant variance. Typically each district's best models met five of these criteria; for example, in the Salem District, the model for the lots in the Roanoke Valley Allegany Regional Commission met all criteria except the fourth, where the mean testing error (14) was more than one-half the mean occupancy (25). The most challenging district for developing a model was for the portions of Fredericksburg that were not within the Middle Peninsula PDC (where four criteria were met); the best district was probably Staunton, where both the Northern Shenandoah and Central Shenandoah PDCs met all six criteria.

- *There was no single best way to develop models for forecasting occupancy.* Although the techniques of site-specific calibration, selection of different socioeconomic variables, use of commuting rather than administrative boundaries, and consideration of a nonlinear model each had a role in improving model performance, no single technique worked for all cases. For example, a nonlinear model was suitable in only one case—the Northern Virginia District—where other approaches had failed. Further, the use of MPO boundaries as a commute shed and the inclusion of the distance to the nearest employment center also did not improve model fit. In sum, the best models for the 297 park and ride lots were developed as follows (the "14-19" reflects the fact that for five lots in the Lynchburg District, one could use either those the district model or the Thomas Jefferson PDC model):

  — Develop a simple linear model applicable to the entire VDOT district: 14-19 lots.
  — Develop separate linear models for high and low population densities: 61 lots.
  — Develop a linear model applicable to the PDC or MPO: 114 lots.
  — Develop a nonlinear model (later segmented into two submodels): 87 lots.
  — No suitable model was found: 21 lots.

- *The best models confirmed some, but by no means all, a priori expectations of key influences of occupancy.* When the best models were considered, there were several cases where variables expected by the research team to affect demand were indeed found to be statistically significant; examples include some form of traffic volume (affecting demand for 58 lots); congestion as indicated by the PHEF (49 lots); the availability of amenities such as lighting (44 lots) or number of bicycle spaces (78 lots); and transit-related variables, such as availability of service (122 lots) or persons using transit (29 lots). However, the best models also showed a few surprises; for example, for 14 lots, the number of carpoolers was forecast to reduce occupancy (perhaps because of an interaction effect with transit service that increased occupancy); for 5 lots the number of transit riders reduced occupancy; and notably, the distance to major employment centers (Figure 14) was found to be significant for only 9 lots in the Northern Virginia District that did not offer transit service. In particular, it was noteworthy that there was no single variable that was useful for forecasting demand at all 297 lots.

- *The best models based on the methods used in this study may be more suitable for identifying the expected impact of a future trend than for making an exact forecast for a brand new lot.* For example, two recommended models can be considered: that of the Richmond District (average error of 45 spaces in Table 15), and that of the Lenowisco PDC (average error of 5 spaces in Table 8). Given that the former has an average occupancy of 72 and the latter has an average occupancy of 8, each of these models has an error/occupancy ratio of 0.63. Thus, using these models to forecast the occupancy of a new lot built in either location will yield a forecast error greater than one-half the occupancy. However, when the site-specific adjustment method discussed previously is used, each model can incorporate the impact of how changes in future key variables may affect occupancy. For instance, in the Richmond District, the finding that average ADT (within 2.5 miles of the park and ride lot) is significant suggests that an increase in an average ADT of 1,000 will in that area yield an increase in occupancy of about 6 spaces. The Lenowisco PDC model suggests another trend: an increase of 100 commuters who live within 2.5 miles of the lot who have jobs that are more than 50 miles away will increase forecast occupancy by about 8. Thus, these models may also be useful for forecasting the expected changes of key variables at existing lots.

- *Multiyear trends in occupancy could not be detected with the data available.* Figures 15 and 16, which are based on data collected once per year or less, do not show a consistent multiyear trend. One possible explanation is that there truly is no trend in occupancy. Another possible explanation is that a more frequent data collection program might allow one to isolate different sources of variation (seasonal, random, and yearly) in order to detect longer-term trends.

RECOMMENDATIONS

1. *The VDOT TMPD's Multimodal Section should consider using one of the following two approaches to forecast park and ride lot demand provided the lot does not solely serve passengers using the VRE or the Metro: (1) one of the suggested models shown in Table 19 provided in the form of a spreadsheet, or (2) a custom model developed as part of a site-specific study.* The reason for the first is that all models shown in Table 19, provided to TMPD in the form of a spreadsheet, incorporate variables that are shown to significantly affect park and ride lot demand ($p \leq 0.05$). This recommendation does not suggest that TMPD develop new models; rather, it suggests that TMPD use models developed in this study unless there is a desire to create new models. The reason for the second is that this study also showed that factors in addition to the variables identified in this study also may affect park and ride lot demand, and thus, if desired, new models could be developed. The models in Table 19 are not suitable for the 21 lots that mostly serve VRE and Metro shown in Table 14, along with the Horner Road Commuter Lot and the Ballston Public Parking Garage.

Table 19. Recommended Models for Application

District	Location	Model
Bristol	Lenowisco PDC	1.359 + 0.082 * Rad2_JobsGT50Mi
	Cumberland Plateau PDC	2.099 + 0.018 * Closest ADT - 0.362 * $Vpeak_K$+ 0.004 * Rad5_JobsLT10Mi
	Mount Rogers PDC[a]	-7.006 + 0.247 * TransitRiders2 + 0.004 * $Vprime_K$
Culpeper	Charlottesville-Albemarle MPO	-4.107 + 27.798 * OvernightParkingAllowed + 0.009 * TransitRiders2
	Non-MPO portion of the Thomas Jefferson PDC[b]	2.725 + 0.042 * Rad2_JobsLT10Mi
	Rappahannock-Rapidan PDC	6.141 – 0.065 * Carpoolers2 + 0.059 * TransitRiders2 + 0.005 * Rad5_Jobs10_to_24Mi + 8.229 * Lighting
Fredericksburg[c]	Middle Peninsula PDC [c]	1.872 * RentOverAllIncome2 – 0.999 * AvgPctOfRentIncomeOnRent5
	Population density $\geq$2396 people/mile^2	1931.706 - 1.949 * TransitRiders2 + 0.000256 * POP
	Population density $\leq$ 2395 people/mile^2	-217.053 + 241.839 * Lighting + 329.448 * NuofTranServicePP +0.018 * POPDEN + 67.016 * PHEF
Hampton Roads	Population density $\geq$ 5375 people/mile^2	-541.193 + 0.052 * Carpoolers2 + 0.252 * Rad2_Jobs25_to_50Mi
	Population density $\leq$ 5374 people/mile^2	8.341 + 0.000262 * Max ADT (low population density, 25 lots)
Lynchburg[b]	any	-1.472 + 0.475 * PHEF + 0.002 * Average ADT + 0.000049 * POP5
Northern Virginia[d]	Lots with transit service	$(2.488 + 0.298 * \text{Bicycle Spaces} + 0.396 * \text{NuofTranServicePP} + 0.001 * \text{Average ADT})^2$
	Lots without bus service	$(2.844 - 0.000071 * \text{Dist\_M2} + 1.128 * \text{DTNearestP})^2$
Richmond	Any	4.083 + 0.006 * Average ADT + 151.906 * PHEF
Salem[a]	New River Valley PDC	2.340 + 67.177 * TransitServiceAvailable + 0.000161 * POP5
	Roanoke Valley-Alleghany PDC	5.849 + 0.015 * Rad2_Jobs25_to_50Mi + 48.731 * TransitServiceAvailable
Staunton	Central Shenandoah PDC	0.516 + 0.027 * Rad2_Jobsto50Mi + 0.004 * Carpoolers10
	Northern Shenandoah PDC	-136.688 + 0.024 * Rad5_JobsLT10Mi + 3.516 * CommuteTime5

[a] One lot in the Mount Rogers PDC is in the Salem District. For that one lot, the Mount Rogers mode should be used.

[b] Five lots in the non-MPO portion of the Thomas Jefferson PDC are in the Lynchburg District. Either model is suitable for these five lots,

[c] In the portion of the Fredericksburg District covered by the Middle Peninsula PDC, the latter model is preferable in terms of showing a lower testing error. However, if the equity concern of that model's last term is problematic, then low population density Fredericksburg model could be used instead.

[d] The Northern Virginia District models are not applicable for the 21 lots (see Table 14) that exclusively serve Metro and the Virginia Railway Express as well as the Ballston Public Parking Garage and the Horner Road Commuter Lot.

Figure 17 shows the recommended models to use based on VDOT district, PDC, and MPO boundaries. When applying the models in Table 19 to an existing lot, users should consider incorporating existing occupancy information, as demonstrated in Equation 23. If the reason for generating the forecast is that some key independent variable is expected to change, such as an amenity at the lot, the introduction of transit service, or nearby commuter characteristics, the approach shown herein can help account for site-specific factors at the park and ride lot. For the two districts where different models are chosen based on population density, the question arises as to what approach one should use if the area is low density at present but will be high density by the time the lot is built. Logically, one would use the higher density model, and the coefficients appear to support this approach. For instance, in Hampton Roads, the highest ADT within 2.5 miles of the lot is the key variable for lower density lots, but for higher density lots, the key variables become carpool use and the number of commuters with jobs that were farther away (25 to 50 miles).

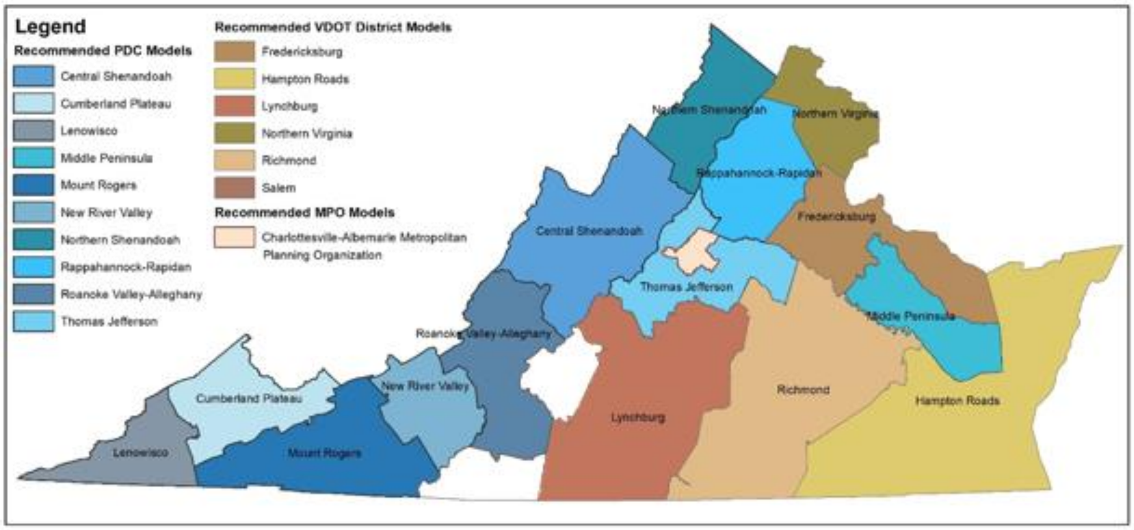

Figure 17. Recommended Models for Application. There are 5 lots in the rural portion of the Thomas Jefferson PDC that are also in the Lynchburg District. For those lots, both models are appropriate. No model has been developed that fits the 21 Northern Virginia District lots in Table 14 that exclusively serve Metro or the Virginia Railway Express. For the two areas marked "Salem," there are no park and ride lots; if new lots were to be constructed in those locations, one should consider adapting the adjacent models for Mount Rogers, New River Valley, or Roanoke Valley Alleghany. PDC = planning district commission.

2. *The VDOT TMPD's Multimodal Section should collect occupancy samples more than once per year*. Mouskos et al. (2007) noted that New Jersey's use of "only one visual survey per year is inadequate" because it does not allow one to capture variance by time of day or day of week. This same limitation also applies to the database available for this study. Cheu et al. (2012) supported multiple data collection instances per day, noting that a binary logit model that forecasts demand may not account for the fact that patrons will not use the facility at the same time (and if the lot is above capacity, latent demand will be missed). Forecasts generally do not account for such perturbations: for the Northern Virginia District, BMI et al. (2003) had forecast that demand would increase by about 50% from 2001-2020; however, data from 1997-2018 generally did not suggest a smooth linear increase (see Figures 15 and 16). Thus, more sampling can at least indicate an expected amount of variation at a given lot. Although the collection of additional. more frequent samples does not guarantee greater predictive power, the facts that others have observed such a need and the large variation shown in occupancies when data are collected only once per year suggest that this recommendation may yield better models for forecasting.

IMPLEMENTATION AND BENEFITS

Implementation

The chief effort involved in implementing Recommendation 1 is the development of spreadsheets that not only allow model application but also show the range of values. In this regard, spreadsheets have been developed to allow forecasters to apply these 19 models in different regions of Virginia. Forecasters can use the spreadsheets to forecast demand for a new lot (as done in Eq. 20) or they can use the spreadsheets to forecast demand as a function of a key change in some variable (as done in Eqs. 21-24). Figure 18 shows an excerpt of the spreadsheet that applies the example shown in Equations 21 through 24, where the impact of increased transit service and ADT for an existing lot is estimated. In future years, the values of the independent variables would need to be updated (e.g., in 2025, ADTAverage may have a different value than in 2020), but the spreadsheets can still be used provided the input data are updated.

The chief effort in implementing Recommendation 2 is the cost of data collection. Although more data are always desirable, a starting point could be to collect the data at least quarterly; this would help one capture some of the random variation. Then, with additional resources, one could begin to focus more intense efforts on higher occupancy lots, especially those where demand may be approaching capacity.

In the past, the cost for a site visit that included occupancy and other lot characteristics was on average \$500 per lot. In theory, therefore, implementing Recommendation 2 would increase data collection costs for VDOT. However, there may be opportunities to offset these additional expenses if certain data elements do not have to be collected with each visit or if technologies, besides manual counts, can be implemented; Dey et al. (2017) in a Washington, D.C., study, evaluated a variety of detection types such as closed-circuit television and cameras with global positioning systems and indicated that several of these technologies "show promise" (albeit for the case of on-street parking).

Model 10 Example 2	Suppose that the number of buses serving the Occoquan Commuter Lot (Old Hechinger's) will double and ADT will increase 10%. What is the expected impact on occupancy?	
	12	Existing data: Bicycle Spaces = 12
	4	Existing data: NuofTranServicePP = 4
	8733	Existing data: ADTAverage = 8733
	16	Square root of forecast occupancy with existing bus service
Eq. 22	*268*	Forecast occupancy with existing bus service
	12	New data: Bicycle Spaces = 21
	8	New data: NuofTranServicePP = 4
	9606	New data: ADTAverage increased 10%
Eq. 23	*355*	Forecast occupancy with new bus service and ADT
	437	Current occupancy
Eq. 24	578	New occupancy = Current Occupancy * Forecast occupancy with new bus service/Forecast occupancy with existing bus service

Figure 18. Example of Spreadsheet Implementing the Model Shown in Equations 21-24

Benefits

Park and ride lots have a wide range of agency costs depending on the price of land, whether the lot is owned outright or leased from other entities, whether the agency contributes to the cost of the lot, and the type of amenities provided at the lot. The Texas Transportation Institute (2020) reported that the new construction of park and ride lots can range from $30,000 to $50,000 per space. Park and ride lots can also yield notable mobility benefits that have a public impact. For example, if a lot enables a two-person carpool for a 20-mile commute, such a lot eliminates a bit more than 16 kg of carbon dioxide emissions per commuter per day assuming typical CO_2 emissions of 404 grams/vehicle (U.S. Environmental Protection Agency, 2018). Thus, being able to forecast where demand is likely to increase can help VDOT make these costly, but potentially highly beneficial, investments more wisely. Although the implementation of Recommendation 1 is designed to help forecast demand immediately, the implementation of Recommendation 2 could potentially improve forecasting accuracy in the longer term. The best models developed for this study have a median accuracy of about 56% (testing error divided by mean occupancy); it is possible that better data could yield more accurate models, especially in cases where true demand is above the number of spaces at the lot.

ACKNOWLEDGMENTS

This work was guided through the diligent efforts of the TRP consisting of Olivia Mobayed (champion), Terrell Hughes, and Shane Sawyer, all of VDOT's TMPD.

REFERENCES

Beattie, M. Memorandum to Peter Sanderson, AKRF Planning Consultants, August 29, 2014.

BMI, Burgess & Niple, Inc., and Fitzgerald & Halliday, Inc. *Northern Virginia Park & Ride Lot Feasibility Study*. Fairfax, VA, 2003.

Bos, I., Ettema, D., and Molin, E. Modeling Effect of Travel Time Uncertainty and Traffic Information on Use of Park-and-Ride Facilities. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1898, 2004, pp. 37-44.

Bowler, C.E., Noel, E.C., Peterson, R., and Christiansen, D. *Park-And-Ride Facilities—Guidelines for Planning, Design and Operation.* Federal Highway Administration, McLean, VA, 1986.

CDM Smith. VTrans Mid-Term Needs (Nodes), February 4, 2020. https://www.arcgis.com/home/item.html?id=cb2f8ae1f615453b9983e939666479bb. Accessed June 23, 2020.

Cherrington, L.K., Brooks, J., Cardenas, J., Elgart, Z., Galicia, L.D., Hansen, T., Miller, K., Walk, M.J., Ryus, P., Semler, C., and Coffel, K. *Decision-Making Toolbox to Plan and Manage Park-and-Ride Facilities for Public Transportation: Guidebook on Planning and Managing Park-and-Ride.* TCRP Report 192. Transportation Research Board, Washington, DC, 2017.

Cheu, K., Hernandez, S., Cornejo, L., and Perez, S. *Systematic Approach to Evaluate Potential Park and Ride Facilities.* University of Texas at El Paso, Department of Civil Engineering, 2012.

Cornejo, L., Perez, S., Cheu, R.L., and Hernandez, S. An Approach to Comprehensively Evaluate Potential Park and Ride Facilities. *International Journal of Transportation Science and Technology,* Vol. 3, No. 1, 2014, pp. 1-18.

CTC & Associates LLC. *Assessing Park-and-Ride Impacts.* Sacramento, CA, 2010.

Desman Associates. *Metrorail Parking Demand Study—Route 606 and 772 Stations.* Loudoun County, VA, 2012.

Dey, S.S., Darst, M., Pochowski, A., Pérez, B.O., and Sanchez, E.C. Hunt for Perfect Detection of Parking Occupancy Evaluation of On-Street Technology and Its Ability to Address Urban Challenges. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2645, 2017, pp. 12-23.

Florida Department of Transportation. *State Park-and-Ride Guide.* Tallahassee, 2012.

Frost, J. Standard Error of the Regression vs. R-squared, n.d. https://statisticsbyjim.com/regression/standard-error-regression-vs-r-squared/. Accessed October 25, 2020.

Garber, N.J., and Hoel, L.A. *Traffic and Highway Engineering,* 4th edition. Cengage, Toronto, Ontario, Canada, 2009.

Geert van den Berg, R. SPSS Stepwise Regression—Simple Tutorial, 2020. https://www.spss-tutorials.com/stepwise-regression-in-spss-example/. Accessed November 16, 2020.

I-95/I-395 Transit/TDM Technical Advisory Committee. *I-95/I-395 Transit/TDM Study Final Report.* Department of Rail and Public Transportation, Richmond, VA, 2008.

Jacobs. *FEIS Parking Demand.* Colorado Department of Transportation, Denver, 2010.

Lane, D. Online Statistics Education: A Multimedia Course of Study, n.d. http://onlinestatbook.com/2/regression/accuracy.html. Accessed October 25, 2020.

Ling, Z. Email to J.S. Miller, September 11, 2018.

Martin, J., De Adana, D.D.R., and Asuero, A.G. (2017). Fitting Models to Data: Residual Analysis, A Primer. Uncertainty Quantification and Model Calibration, 2017, 133.

Metropolitan Council. Metropolitan Council 2030 Park-and-Ride Plan. St. Paul, MN, 2010.

Mobayed, O. Email to L.E. Dougald, August 16, 2019.

Mobayed, O. Email to L.E. Dougald, June 26, 2020a.

Mobayed, O. Email to L.E. Dougald, May 6, 2020b.

Mobayed, O. Email to L.E. Dougald, June 17, 2020c.

Mobayed, O. Email to L.E. Dougald, June 12, 2020d.

Mobayed O. Email to L.E. Dougald, June 24, 2020e.

Montgomery, D.C. *Introduction to Statistical Quality Control*, 4th ed. John Wiley & Sons, New York, 2001.

Mouskos, K.C., Boile, M., and Parker, N. *Technical Solutions to Overcrowded Park and Ride Facilities.* FHWA-NJ-2007-011. New Jersey Department of Transportation, Trenton, 2007.

Nelson\Nygaard Consulting Associates Inc. *Central Minnesota Area Commuter Study: Final Report.* Minnesota Department of Transportation, St. Paul, 2012. http://www.dot.state.mn.us/d3/commuter/. Accessed December 15, 2020.

Nordstrom, J., and Christiansen, D.L. *Guidelines for Estimating Park-and-Ride Demand.* UMTATX81-1064 lF. Texas Transportation Institute, College Station, 1981.

Nungesser, L.G., and Ledbetter, N.P. *Procedures for Estimating Park-and-Ride Demand in Large Texas Cities.* Research Report 339-7. Texas Transportation Institute, College Station, 1987.

Peng, C., and Mohamad, D. The Development of Model Estimation to Determine Parking Needs at LRT Stations in Suburban Area. In *Proceedings of the Eastern Asia Society for Transportation Studies,* Vol. 5, 2005, pp. 877-890.

Plotz, J., Konduri, K.C. and Pendyala, R.M. To What Extent Can High-Occupancy Vehicle Lanes Reduce Vehicle Trips and Congestion? Exploratory Analysis Using National Statistics. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2178, 2010, pp. 170-176.

Rathbone, D.B. Characteristics of the Most Successful Park-and-Ride Lots in North America: This Week's Survey Results. *The Urban Transportation Monitor,* Vol. 20, No. 12, 2006, pp. 9-12.

RSG. Statewide Park-and-Ride Facilities Plan: Appendix, Vermont Agency of Transportation, Montpelier, 2015. https://vtrans.vermont.gov/sites/aot/files/planning/documents/planning/Appendix_2015-12.pdf. Accessed August 21, 2019.

Shirgoakar, M., and Deakin, E. Study of Park-and-Ride Facilities and Their Use in the San Francisco Bay Area of California. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1927, 2005, pp. 46-54.

Southwest Region Planning Commission, Strafford Regional Planning Commission, and Central New Hampshire Regional Planning Commission. *New Hampshire Park and Ride Toolkit: A Guide for Municipalities and Park and Ride Supporters,* 2015. https://cnhrpc.org/nh-park-and-ride-toolkit/. Accessed March 23, 2021.

Spillar, R.J. *Park-and-Ride Planning and Design Guidelines.* Parsons Brinckerhoff, New York, 1997.

Texas Transportation Institute. Mobility Investment Priorities: Park and Ride Lots, College Station, 2020. https://mobility.tamu.edu/mip/strategies-pdfs/system-modification/technical-summary/Park-And-Ride-Lots-4-Pg.pdf. Accessed November 5, 2020.

Turnbull, K. *NCHRP Synthesis 213: Effective Use of Park and Ride Facilities.* Transportation Research Board, Washington, DC, 1995. http://onlinepubs.trb.org/Onlinepubs/nchrp/nchrp_syn_213.pdf. Accessed December 18, 2020.

U.S. Census Bureau. Geography Program, 2020a. https://www.census.gov/programs-surveys/geography.html. Accessed May 6, 2020.

U.S. Census Bureau. OnTheMap, 2020b. https://onthemap.ces.census.gov/. Accessed July 2, 2020.

U.S. Environmental Protection Agency. Greenhouse Gas Emissions From a Typical Passenger Vehicle, 2018. https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle. Accessed November 5, 2020.

VHB Engineering, NC. *Transit and Parking Forecast.* Chapel Hill, NC, 2013.

Vincent, M. *Park and Ride: Characteristics and Demand Forecasting.* Research Report 328. Land Transport New Zealand, Waterloo Quay, Wellington, 2007.

Virginia Department of Transportation. *Park and Ride in Virginia,* 2019a. http://www.virginiadot.org/travel/parkride/about.asp. Accessed September 6, 2019.

Virginia Department of Transportation. AADT. Pathways for Planning, November 1, 2019b. https://vdotp4p.com/view. Accessed April 26, 2020.

Virginia Department of Transportation. Peak Hour Factor. Pathways for Planning, November 1, 2019c. https://vdotp4p.com/view. Accessed April 26, 2020.

Virginia Department of Transportation. Volume to Capacity Ratio. Pathways for Planning, November 1, 2019d. https://vdotp4p.com/view. Accessed April 26, 2020.

Virginia Department of Transportation. Peak Hour Expansion Factor Weekday 2019. Pathways for Planning, April 30, 2020. https://vdotp4p.com/view. Accessed July 3, 2020.

Washington Metropolitan Area Transit Authority. Average Daily Parking Transactions per Year, 2020. https://www.wmata.com/initiatives/ridership-portal/Parking-Data-Portal.cfm. Accessed December 21, 2020.

Webb, A., Tao, T., Khani, A., Cao, J., and Wu, X. *Impact of Transitways on Travel on Parallel and Adjacent Roads and Park-and-ride Facilities.* Minnesota Department of Transportation, St. Paul, 2021.

APPENDIX

SUMMARY OF DATA PROCESSING

This appendix illustrates the types of computations performed for this study and the resultant data. Key processing details are presented for only five variables, but the methods developed herein were generally applied to other sources. The manner in which these variables are processed is summarized in six steps:

1. Acquire data.
2. Compute values for variables.
3. Import variables into a GIS environment.
4. Implement appropriate scripts.
5. Export variables for use with statistical software.
6. Develop district-specific variables.

Acquire Data

Data were provided from several sources. For example, four ACS files were used to obtain, for each Census tract, five variables: average commute time, total number of carpoolers, total number of transit users, median rent divided by median household income, and average percent of renters' household income spent on rent. Each file was based on the 5-year period ending in 2018 and thus provided coverage for all Census tracts in Virginia. Three of the files were provided by Mobayed (2020b), and one file was obtained from the U.S. Census Bureau (2020) directly. The variables extracted were as follows for each of the 1,907 Virginia Census tracts:

Table DP04, Selected Housing Characteristics [midpoints in brackets]

- Median Rent (DP04_0134E)
- Percent spending less than 15% of income on rent (DP04_0137PE) [7.45]
- Percent spending 15.0-19.9% of income on rent (DP04_0138PE) [17.45]
- Percent spending 20.0-24.9% of income on rent (DP04_0139PE) [22.45]
- Percent spending 25.0-29.9% of income on rent (DP04_0140PE) [27.45]
- Percent spending 30.0-34.9% of income on rent (DP04_0141PE) [32.45]
- Percent spending 35% or more of income on rent (DP04_0142PE) [37.5]

Table S1903, Median Income in the Past 12 Months (in 2018 Inflation-Adjusted Dollars)

- Median household income (variable S1903_C03_001E)

Table S0801 Commuting Characteristics by Sex

- Total number of workers age 16+ (variable S0801_C01_001E)
- Percent of workers who carpool (variable S0801_C01_004E)
- Percent of workers who use transit (variable S0801_C01_009E)

Table B08303, Travel Time to Work

- Number of workers for whom a commute time is available (B08303_001E)

- Number of workers in each of the following travel time bins [midpoints in brackets]

— Less than 5 minutes	B08303_002E	[2]
— 5 to 9 minutes	B08303_003E	[7]
— 10 to 14 minutes	B08303_004E	[12]
— 15 to 19 minutes	B08303_005E	[17]
— 20 to 24 minutes	B08303_006E	[22]
— 25 to 29 minutes	B08303_007E	[27]
— 30 to 34 minutes	B08303_008E	[32]
— 35 to 39 minutes	B08303_009E	[37]
— 40 to 44 minutes	B08303_010E	[42]
— 45 to 59 minutes	B08303_011E	[47]
— 60 to 89 minutes	B08303_012E	[74.5]
— 90+ minutes	B08303_013E	[100]

Although there are 1,907 Census tracts, data are missing from 2% to 5% of these tracts as follows:

- For *average commute time*, *total number of carpoolers*, and *total number of transit users*, 32 tracts had no population, leaving 1,875 tracts (of 1,907 total tracts) with data for these variables.

- For *median rent divided by median household income*, there were 1,835 tracts with data. The reason was that for income, 1,868 tracts had no income data such that income would be reported as a "Null" value (an example was 51179010201). For 4 of the 1,868 remaining tracts, income was shown as "250,000+" and "2,500-," which the research team changed to "250,000" and "2,500," respectively. For rent, only 1,835 tracts had data. For 9 of those tracts, rent was shown as "3,500+," which the research team changed to "3,500." Then, when monthly rent was multiplied by 12 and then divided by income such that rent as a portion of income was reported, 1,835 values resulted (since cells with no rent data had to be excluded).

- For *average percent of renters' household income spent on rent*, 1,869 tracts had data.

Compute Values for Variables

For each Census tract, the five variables were then determined as follows:

1. Average commute time was determined by multiplying the number of commuters (variables B08303_002E through B08303_013E) by the travel time in brackets and then dividing the entire sum by the total number of commuters (B08303_001E).

2. Total number of carpoolers was determined by multiplying the total workers (S0801_C01_001E) by the percent who carpool (S0801_C01_004E).

3. Total number of transit users was determined by multiplying the total workers (S0801_C01_001E) by the percentage who use transit (S0801_C01_009E).

4. Median rent divided by median household income was determined by multiplying the median rent (DP04_0134E) by 12 and then dividing by the median household income (S1903_C03_001E). The median rent is for renters only, and the median household income includes renters and non-renters.

5. Average percent of renters' household income spent on rent was determined by multiplying the percent of people in each tract in each bin for portion of income spent on rent (DP04_0137PE through DP04_0142PE) by the portion in brackets.

For example, Tables A1, A2, A3, and A4 illustrate how these variables are computed using Census Tract 51540000700. As shown in Table A1, the average commuting time is 15.9 minutes. Table A2 shows the number of carpoolers (129.8) and transit users (66.8) in that tract. Table A3 shows that the median rent divided by median household income was 16.36%. Table A4 shows that the average percent of renters' household income spent on rent was 27.6%. These latter two variables reflect different information: Table A3 is median rent relative to all household incomes (some of which are renters and some of which are owners), whereas Table A4 is rent as a portion of that householder's income. These variables are also shown in Figure A1.

Table A1. Computing Average Commuting Time for Census Tract 51540000700

Time Bin	Midpoint	Frequency	Product
0-4 minutes	2	42	84
5-9 minutes	7	388	2,716
10-14 minutes	12	561	6,732
15-19 minutes	17	294	4,998
20-24 minutes	22	258	5,676
25-29 minutes	27	35	945
30-34 minutes	32	51	1,632
35-39 minutes	37	10	370
40-44 minutes	42	17	714
45-59 minutes	52	26	1,352
60-89 minutes	75	26	1,937
90+ minutes	100	0	0
Total		1,708	27,156
Average Commute Time			15.9

Table A2. Computing Carpoolers and Transit Users for Census Tract 51540000700

Variable	S0801_C01_001E	S0801_C01_004E	S0801_C01_009E
Description	Number of Workers 16 years and over	Percent carpooled	Percent public transportation (excluding taxicab)
Value	1966	6.6%	3.4%
Number		129.8	66.8

Table A3. Computing Median Rent Divided by Median Household Income for Census Tract 51540000700

Variable	Median Annual Household Income	Median Monthly Rent
Value	$78,565	1,071
Annual Rent/Annual Income		16.36

Table A4. Computing Percent of Renters' Household Income Spent on Rent for Census Tract 51540000700

Income Bin	Midpoint	Percent of People in Each Bin	Product
Less than 15%	7.45%	19.1	1.4
15.0%-19.9%	17.45%	5.7	1.0
20.0%-24.9%	22.45%	9.9	2.2
25.0%-29.9%	27.45%	9.4	2.6
30.0%-34.9%	32.45%	12	3.9
35% or more	37.50%	43.9	16.5
Total		100	27.6

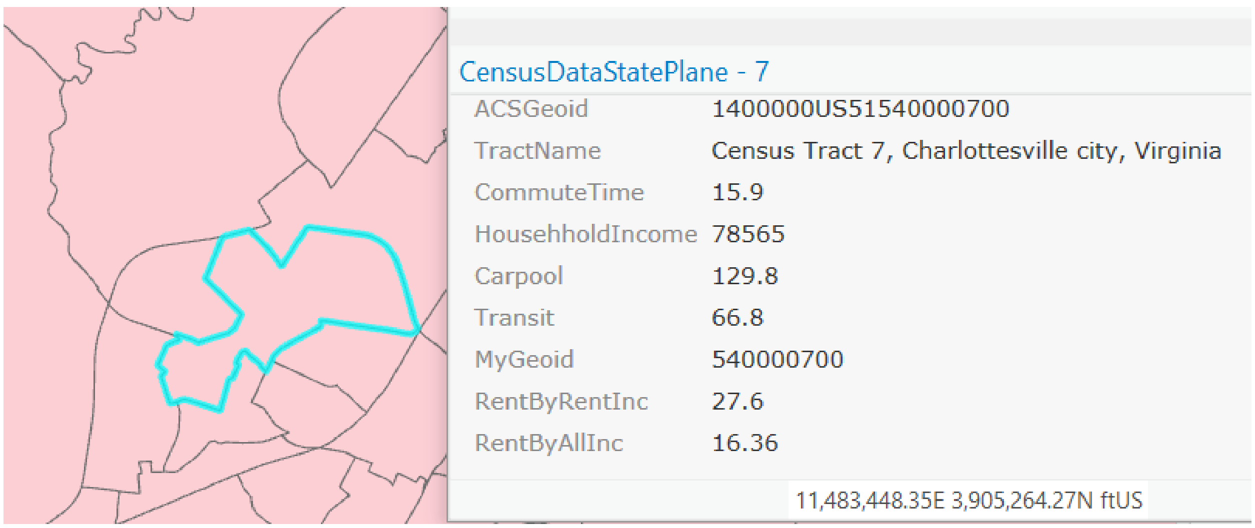

Figure A1. Example of Variables at the Census Tract Level

Import Variables Into a GIS Environment

As the data files used a geoid of the form "1400000US51001090100," a string function was performed to remove the "14000000US" from the geoid in order to make it comparable to the geoid used by the ACS. Then, a geographic file of Virginia Census tracts was obtained from the ACS geography program website (U.S. Census Bureau, 2020). That file also offers a geoid that enables one to link the geographic data (from the ACS) to the tabular data provided by VDOT after one converted one geoid from a text attribute to a numeric attribute.

The lots were then imported into GIS using their latitude/longitude coordinates. Both the lot layer and the Census tract layer were then projected from decimal degrees into Virginia State Plane coordinate system and the geodesic areas of the tracts was calculated. Then, buffers of 2.5 miles, 5.0 miles, and 10.0 miles were generated around each of the park and ride lots, a spatial join between the buffers and the Census layer was performed, and the geodesic areas of the resultant polygons was calculated.

Implement Appropriate Scripts

Similar types of geoprocessing were performed for the other datasets. For example, using the geodatabase (*SmartScale2018_PopEmp_Updated.gdb*) provided by Ling (2018), the LEP population can be found for each of the 10,073 Census block groups. The lots were then imported into GIS using their latitude/longitude coordinates. Both the lot layer and the Smart Scale Census block group layer were then projected from decimal degrees into the Virginia state plane projection. Then, a 2.5-mile buffer was generated around each of the park and ride lots, a spatial join between the buffers and the Census layer was performed, and the LEP population was determined by summing the number of LEP persons in this buffer. These data were exported to a spreadsheet for use in the statistical analysis with appropriate scripts being executed.

One difference from the socioeconomic variable processing was that the scripting performed for the six variables involving distance to activity centers did not require manipulations of polygons but did require joins and the use of the update cursor method given the need to compute the four closest activity centers for each lot. Another difference involved the commute distance obtained from the OnTheMap application, where commuting data appear to be represented as anonymized points where each point represents multiple commuters. In that particular case, the geoprocessing was considerably simpler; one simply tabulated the 52,090 points by performing appropriate spatial joins between them and the 2.5-, 5.0-, and 10.0-mile radii for each lot followed by cleaning the variables and rejoining the feature class to the lots feature class.

Scripts needed to be written for the GIS processing as these calculations were repeated for each of Virginia's 297 lots. The research team found that these operations could generally be completed with about 250 lines of scripting, depending on the geoprocessing tools available. Scripts are available from the research team.

Export Variables for Use With Statistical Software

The resultant variables were then exported such that for each park and ride lot, there were 78 independent variables. Table A5 provides the name, category, minimum, mean, and maximum values for each of these variables in the dataset. When submodels are being developed, the range of possible values is less than that shown in Table A5. For example, for the number of carpoolers within 2.5 miles of the park and ride lot, the values shown in Table A5 range from a low of 7 to a high of 10,787. For the high density sites in the Hampton Roads District, however, this variable ranged from 2,484 to 4,893. The spreadsheet that accompanies the regional models shown in Table 19 shows the low and high value for the variables used in those models, and those ranges will generally be tighter than those shown in Table A5.

Table A5. Domains of Independent Variables

Name	Category	Average	Min.	Max.
Average ADT	Traffic	7424	245	32836
Maximum ADT	Traffic	82311	1800	264000
Sum ADT	Traffic	2070648	12472	7535830
Closest ADT	Traffic	13759	40	120000
Average V/C Ratio	Traffic	0.337	0.043	0.545
K-factor Average	Traffic	0.096	0.074	0.156
PHF Average	Traffic	0.9252397	0.88	0.95
PHEF	Traffic	0.602	0	14
Kclosest	Traffic	0.0961849	0.065	0.173
Kmax	Traffic	0.0988179	0.088388	0.11225
PHFclosest	Traffic	0.91	0.88	0.95
PHFmax	Traffic	0.928896552	0.88	0.95
DIST_BG_BigEmp	Land Use	46894	0	342519
Dist_M1	Land Use	151754	318	1089929
Dist_M2	Land Use	173191	6673	1090649
Dist_M3	Land Use	183701	14005	1092017
Dist_M4	Land Use	193041	17008	1093702
Dist_Weight	Land Use	1445	410	4179
DTNearestP	Land Use	2.97	0.000054	18.57
NofAdjLot	Land Use	3	0	14
ProximityToIAP	Land Use	7.78	0.055	75.54
ProxToEL	Land Use	60.69	0.11	363.04
CommuteTime2	Land Use	34.49	18.23	54.25
CommuteTime5	Land Use	34.65	20.28	48.67
CommuteTime10	Land Use	35.05	23.19	44.66
Carpoolers2	Land Use	1878	7	10787
Carpoolers5	Land Use	6039	25	30061
Carpoolers10	Land Use	17356	51	64153
TransitRiders2	Land Use	1460	0	31095
TransitRiders5	Land Use	4607	0	69600
TransitRider10	Land Use	15589	0	109599
Rad2JobsTot	Land Use	16101.76	3	82901
Rad2JobsLT10M	Land Use	7859.285	2	47212
Rad2Jobs10to24Mi	Land Use	5020.124	0	31690
Rad2Jobs25to50Mi	Land Use	1404.111	0	5928
Rad2JobsGT50Mi	Land Use	1818.242	0	14468
Rad5JobsTot	Land Use	46451.89	39	252481
Rad5JobsLT10Mi	Land Use	22763.88	20	125195
Rad5Jobs10to24Mi	Land Use	14318.42	10	85378
Rad5Jobs25to50Mi	Land Use	4098.795	3	17035
Rad5JobsGT50Mi	Land Use	5270.799	2	30464
Rad10JobsTot	Land Use	141430.6	163	563000
Rad10JobsLT10Mi	Land Use	68619.42	57	273573
Rad10Jobs10to24Mi	Land Use	45417.23	63	201031
Rad10Jobs25to50Mi	Land Use	11765.96	22	39611
Rad10JobsGT50Mi	Land Use	15628.02	14	75862
POP2.5	Demographic	237330	144	1877898
EMP2.5	Demographic	128226	26	836493
POP5	Demographic	1764119	3036	9015716
EMP5	Demographic	856183	614	3871617
POP10	Demographic	451521	3273	1510888
EMP10	Demographic	281893	2887	1790271
AvgPctOfRentIncomeOnRent2	Demographic	26.211	15.187	31.343
AvgPctOfRentIncomeOnRent5	Demographic	26.045	15.2	30.091
AvgPctOfRentIncomeOnRent10	Demographic	25.965	15.2	29.698
RentOverAllIncome2	Demographic	19.779	9.46	40.388
RentOverAllIncome5	Demographic	19.408	11.999	30.678

RentOverAllIncome10	Demographic	18.929	12.61	25.528
MinorityPop2	Demographic	110634.39	9.424837	1032318.5
PorvertyPOP2	Demographic	17618.5	50.487422	153859.98
LEPPop2	Demographic	17437.779	0	186009.05
EligDisadvPop2	Demographic	278576.16	974.87869	2167457.8
MinorityPop5	Demographic	713575.98	43.772636	4169689.5
PovertyPop5	Demographic	99435.468	320.30825	533004.14
LEPPop5	Demographic	107214.6	0	570880.47
EligDisadvPop5	Demographic	1780127.9	2843.7905	9011394.1
MinorityPop10	Demographic	161964.07	105.05716	545528.21
PovertyPop10	Demographic	25947.025	433.89838	97892.395
LEPPop10	Demographic	21509.555	0	95303.296
EligDisadvPop10	Demographic	448401.92	3126	1487049
POPDEN	Demographic	3010.8097	8.347044	72456.204
OvernightParkingAllowed	Facility	0.2226027	0	1
NuofTranServicePP	Facility	2.1541096	0	27
BikeParkingisCovered	Facility	0.1130137	0	1
Lighting	Facility	0.7157534	0	1
SignCondition	Facility	2.5513699	0	4
CostToPark	Facility	0.0342466	0	1
TransitServiceAvailable	Facility	0.5171233	0	1

In addition to the variables shown in Table A5, there were 10 additional facility-related variables (see Table A6) that were not included in the final modeling described in the report. The reason is that early experiments suggested that these variables were not likely to influence occupancy. In these early experiments, the additional facility-related variables in Table A6 were excluded by the stepwise linear regression when building the Virginia district-specific models for nine VDOT districts plus one aggregation of Northern Virginia plus the City of Fredericksburg. However, the research team did later include the number of bicycle spaces in the lot for the Northern Virginia District.

Table A6. Additional Facility-Related Variables That Were Not Used in Model Development

Buses Parked	The number of buses parked in the lot
Trucks Parked	The number of trucks parked in the lot
Motorcycles Parked	The number of motorcycles parked in the lot
Illegally Parked Vehicles	The number of illegally parked vehicles in the lot
Bicycle Spaces	The number of bicycle spaces in the lot (This was used exclusively in the Northern Virginia District nonlinear model after TRP review)
Bikes Parked	The number of bikes parked in the lot
Bike Lane/Shared Use Path Leads to Lot	Whether there is bike lane/shared use path leads to the lot Yes = 1; No = 0
Shelter in Lot	Whether there is shelter in the lot Yes = 1; No = 0
Number of Auto Entrances	The number of auto entrances to the lot
Number of Auto Exits	The number of auto exits in the lot
Electric Vehicle Charging Stations Provided	Whether the electric vehicle charging is provided in the lot

Develop District-Specific Variables

As noted in Table 6, in three districts, park and ride lots were categorized based on the number of residents per square mile in the block group containing the lot. Figures A2, A3, and A4 show these lots for the Fredericksburg, Hampton Roads, and Richmond districts. As noted in Table 14, lots in the Northern Virginia District that are used for VRE or Metro were identified, and these are shown in Figure A5.

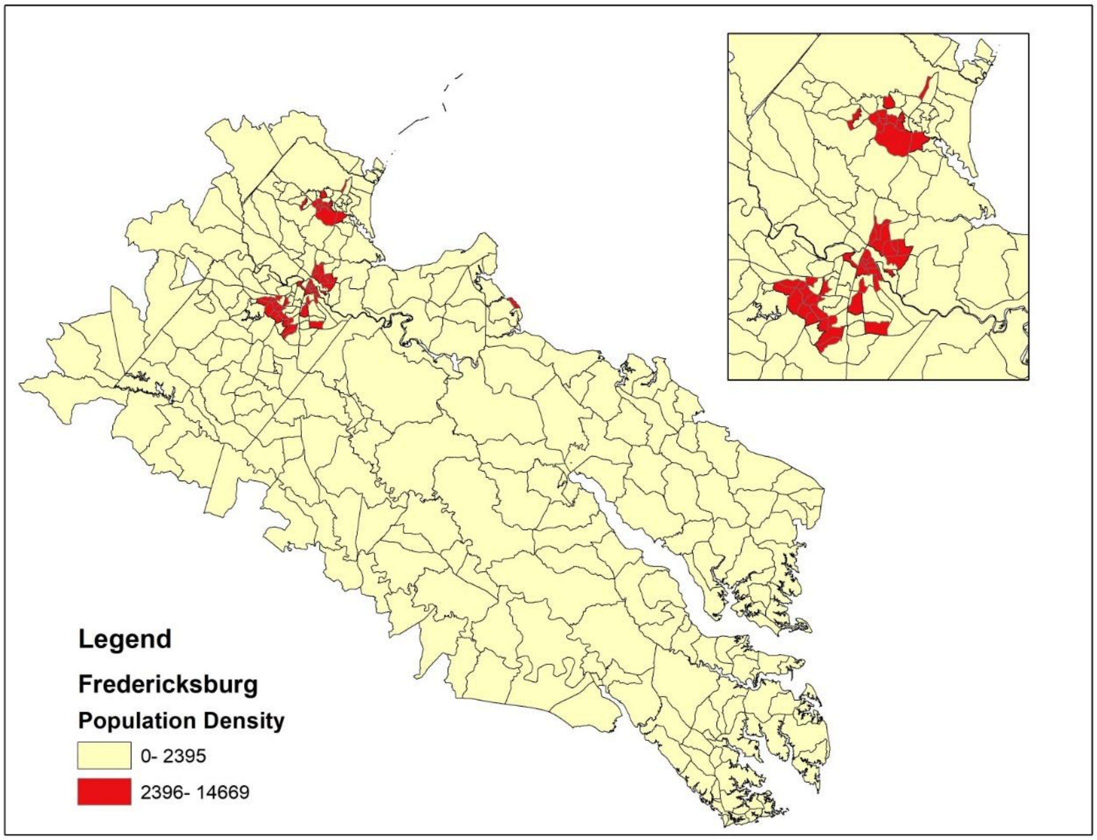

Figure A2. High and Low Population Density Block Groups in the Fredericksburg District

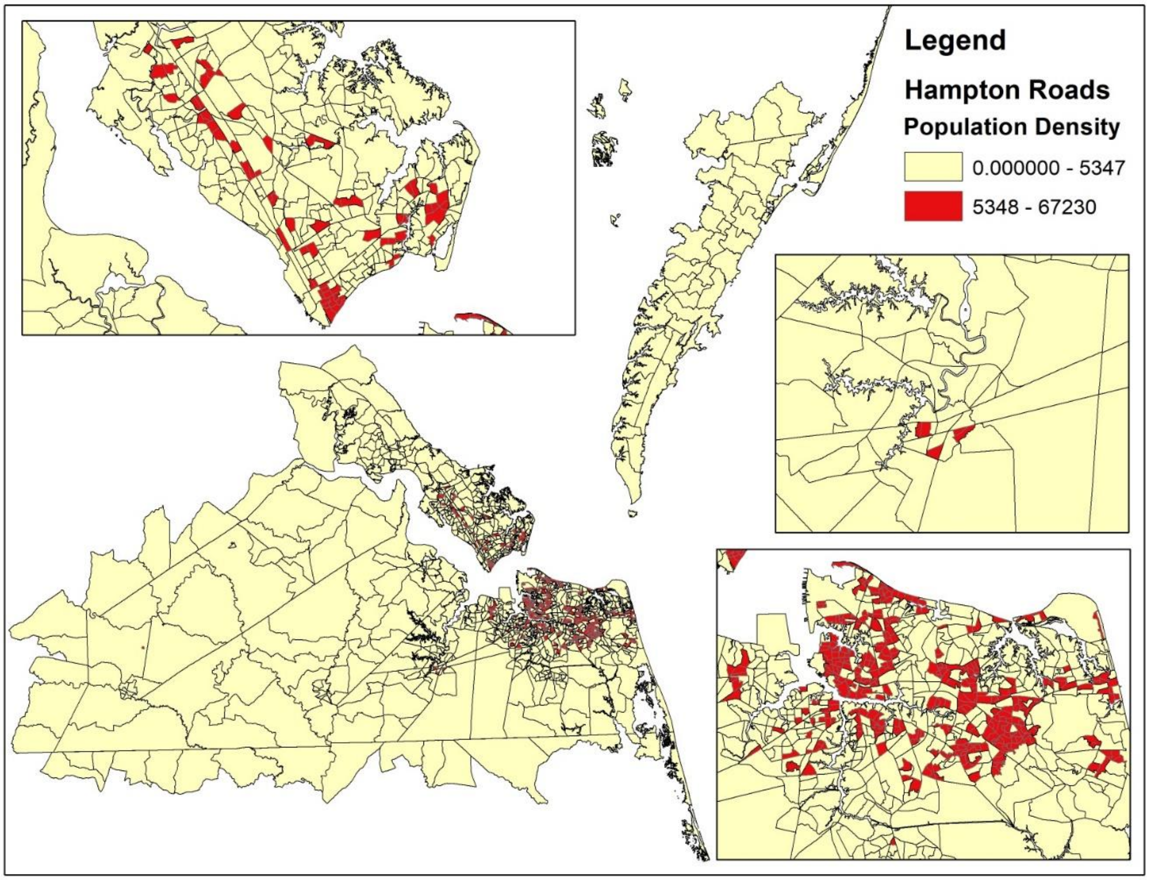

Figure A3. High and Low Population Density Block Groups in the Hampton Roads District

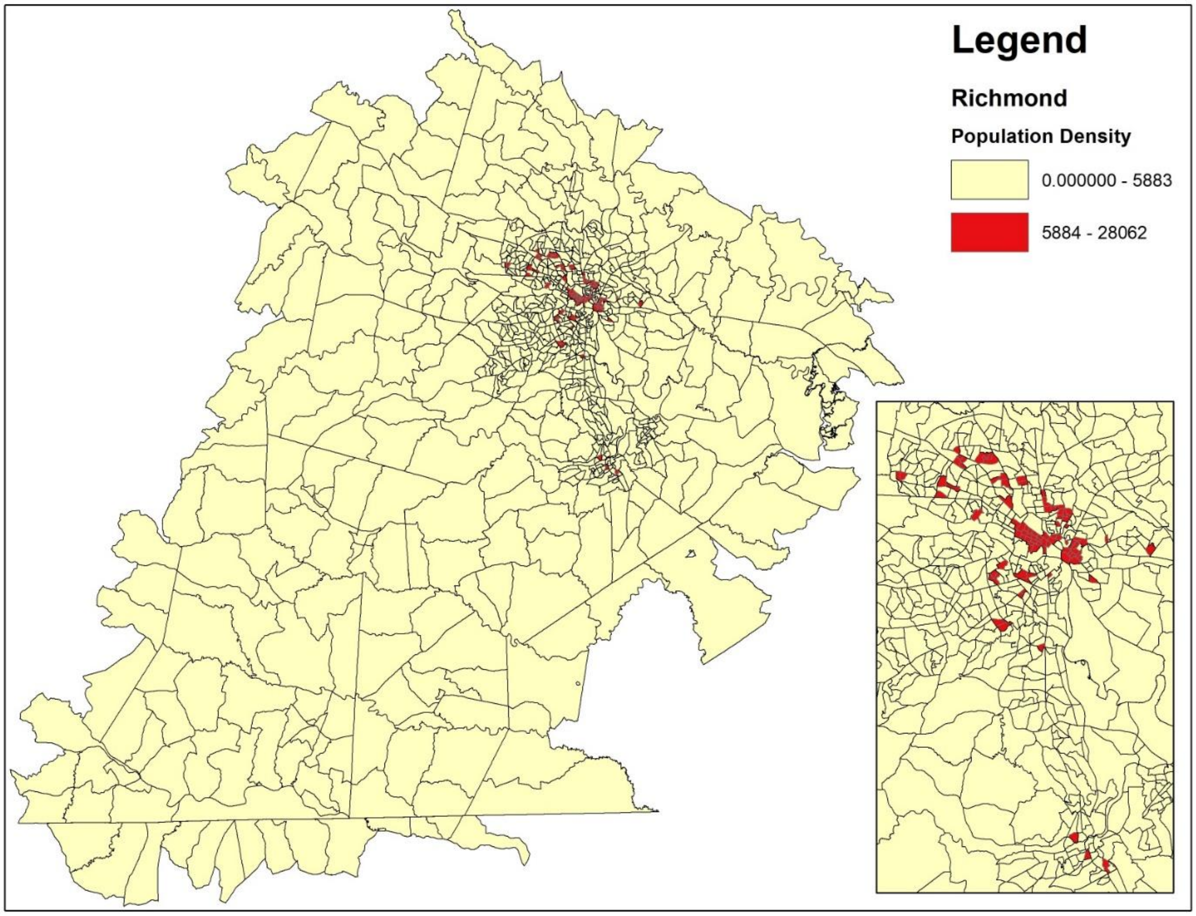

Figure A4. High and Low Population Density Block Groups in the Richmond District

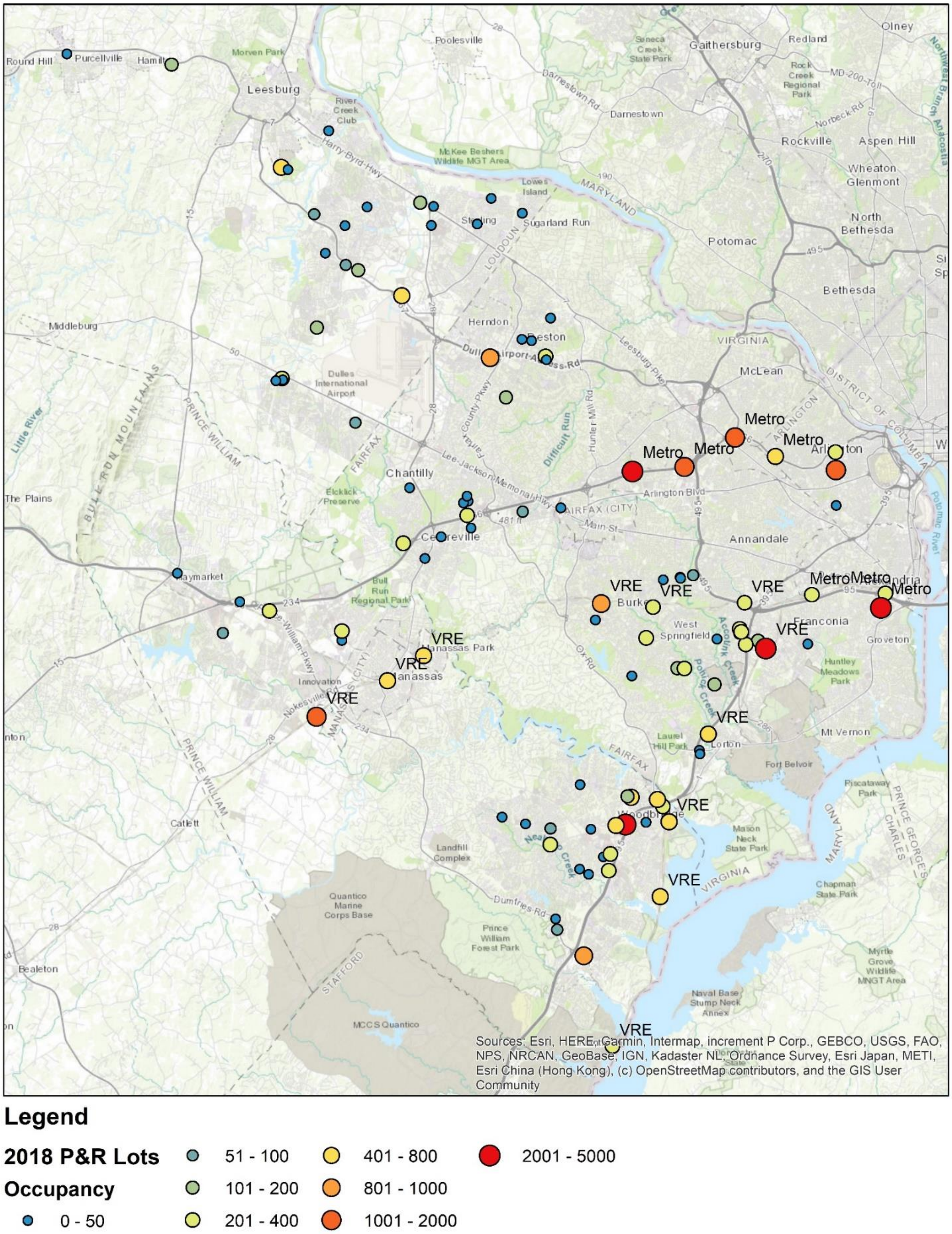

Figure A5. Lots in the Northern Virginia District Specific to Metro and VRE