TN18 - WELLINGTON TRANSPORT ANALYTICAL TOOLS 2019-22 UPDATE – VEHICLE AVAILABILITY MODEL

PREPARED FOR GREATER WELLINGTON REGIONAL COUNCIL

August 2022

Stantec | JACOBS

QUALITY STATEMENT

PROJECT MANAGER	PROJECT TECHNICAL LEAD
Ali Siddiqui	Julie Ballantyne

PREPARED BY		
Dave Hunter	....	25/8/2022
CHECKED BY		
Julie Ballantyne	....	25/8/2022
REVIEWED BY		
Julie Ballantyne	....	25/8/2022
APPROVED FOR ISSUE BY		
Julie Ballantyne	....	25/8/2022

CHRISTCHURCH
Hazeldean Business Park, 2 Hazeldean Road, Addington, Christchurch 8024
PO Box 13-052, Armagh, Christchurch 8141
TEL +64 3 366 7449, FAX +64 3 366 7780

REVISION SCHEDULE

Rev No.	Date	Description	Signature or Typed Name (documentation on file)			
			Prepared by	Checked by	Reviewed by	Approved by
1	1 Aug 2022	Draft	DH	JEB	JEB	JEB
2	25 Aug 2022	Final Cars/Household Updated to Cars/Person	DH	JEB	JEB	JEB

Greater Wellington Regional Council

TN18 - Wellington Transport Analytical Tools 2019-22 update – Vehicle availability model

CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. Introduction

This technical note is part of a series documenting the 2019-2022 update of components of the Wellington Regional Transportation Planning Analytical Tools. The higher-level Analytical Tools are maintained and operated by Greater Wellington Regional Council (GWRC), who is the client for this project. This project is being delivered by Stantec and Jacobs, supported by GWRC transport planners.

The approach adopted to estimate production trip ends is a category model. A category model has two dimensions which divide households/people into different groups where their behaviour is expected to differ.

Vehicle availability is the first dimension of the two-dimensional category model.

Regression models will be fitted to determine if the average vehicle availability per household can be calculated (rather than input). Predictors could include household size, geographic location/accessibility, average household income, number of adults, number of children, etc. While income may be a strong predictor for vehicle availability, it will likely be difficult to forecast. If a predictive model cannot be robustly estimated, then average vehicle availability per household or person for each zone will be an input to the model. The base year data will be sourced from Census. Forecast years will require judgement but can be sensitivity tested.

Once the average vehicle availability per household or person is produced, vehicle allocation curves will be applied. These will convert an average zonal vehicle availability to numbers of persons in households with zero, one, two, three plus vehicles. The source data to fit these curves is Census.

2. Zonal Average Vehicle Availability

2.1 Available Vehicles per Household

Census data for vehicles available to the households within each model area was obtained. The number of houses within each zone that had zero, one, two, three, four and five+ vehicles available were obtained. An estimate of the average number of vehicles that were available to each zones' households were calculated by multiplying through the number of households in each availability group by the relevant number of surveyed households. The average number of vehicles in the 5+ households was obtained from the previous model's data.

To create a model that estimates the zonal average vehicle availability for each zone, a regression model linking zonal vehicle availability to various zonal land use and accessibility characteristics was produced. Census information for each model zone (820 zones) was regressed against its comparable average vehicle availability. The zone characteristics initially assessed against were:

- Average zonal persons per household
- Percent of total persons in each age group (0-19, 20-29, 30-64, 65-69, 70+)
- Average zonal percentage in workforce (Employed, Not employed)
- Average zonal percentage in study (In study, Not in study)
- Average zonal median personal income
- Zonal public transport accessibility

PT accessibility is a measure calculated by using PT generalised cost and employment. Starting from the closest destination zone, the PT generalised cost to that zone is multiplied by employment in that zone, which is then summed with the second closest zone, third closest zone, etc until a pre-determined number of jobs is reached. "Closest" is defined as the smallest generalised cost. This is a measure which should differentiate central city/town locations with good PT access from more suburban locations.

The regression analysis was completed using the latest version of MiniTab (a specialist statistical software). Trip ends for each purpose were regressed against all of the land use variables initially. The regression was a forward stepwise regression with a constant of zero forced in each instance. Variables were only permitted to have non-negative coefficients in the analysis.

After initial efforts to regress vehicles per household to all factors, it was found that no combination of explanatory variables resulted in a relationship that produced a trend that went through the origin when comparing the calculated average vehicle availability against the actual Census average vehicle availability.

The closest of the unsatisfactory relationships for the whole model area is shown in Figure 2-1. This relationship compared zonal average vehicles per household with a combination of median zonal household income, the log_{10} of PT accessibility, zonal percent of population in the 0-19yr, 30-64yr, 65-69yr and 70+yr age groups.

This relationship was considered inadequate for use in projecting average zonal vehicle availability due to the relatively poor R^2 of 0.6651 (calculated when not forced through the origin) and, more importantly, the relationship tended to overpredict availability in zones with low vehicle availability and underpredict zones with high availability by up to +/- 0.62 vehicles per household.

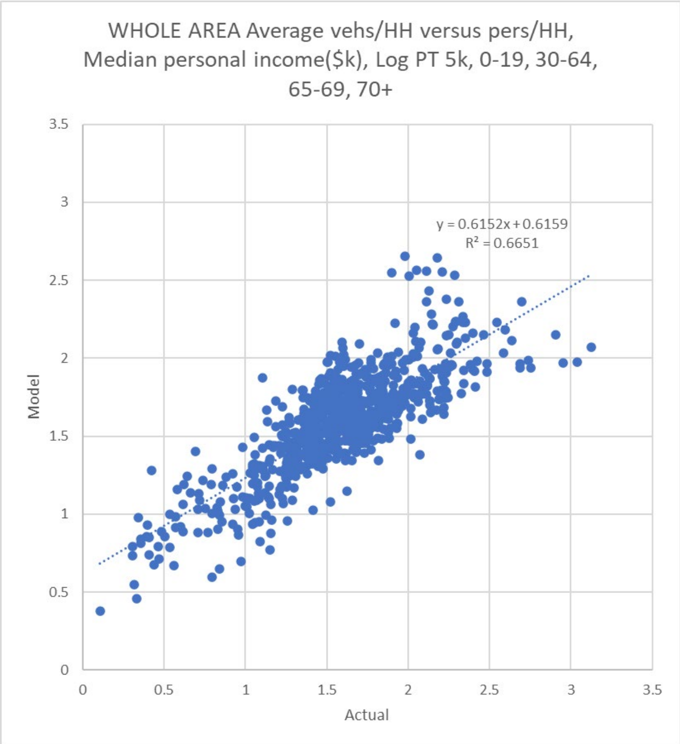
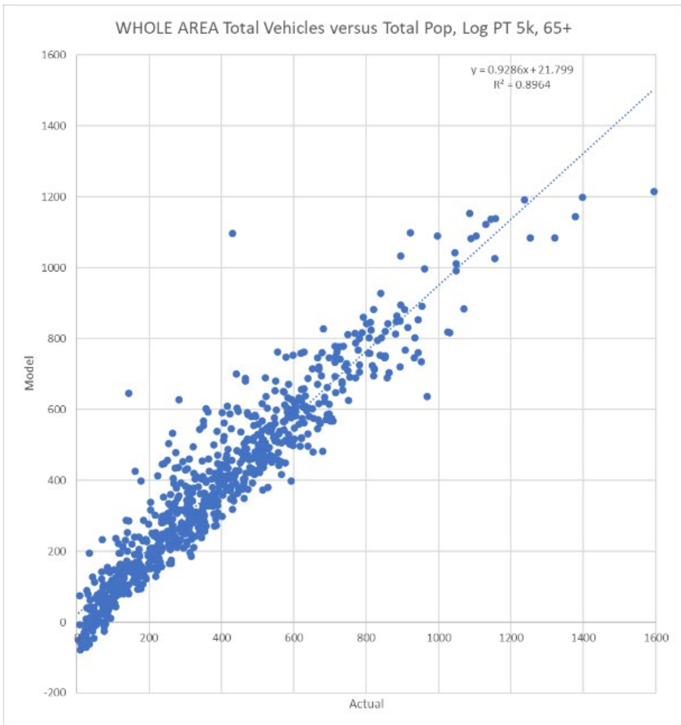


Figure 2-1: Initial Average Vehicle per Household Regression Analysis Comparison

2.2 Available Vehicles

To find an alternative relationship, it was decided to investigate a possible relationship with total zonal vehicle availability (rather than vehicles per household or per person). Census-based zonal total available vehicles were regressed against the same zonal characteristics as used in the previous assessment with the addition of zonal total population. This assessment, again using Minitab, found that a strong relationship was found for the entire model area using a combination of total zonal population, log_{10} of PT accessibility to reach 5000 jobs, and zonal percent of population in the 65+ age group. Figure 2-2 shows the plot of this relationship along with its relevant coefficients.

This relationship has an acceptable R^2 of 0.8964 (calculated when not forced through the origin) and shows that whilst its slope is 0.93 instead of ideally 1.0, it very nearly passes through the origin.



Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Total Pop	0.59691	0.00632	94.49	0.000	2.69
Log PT 5k	174.9	11.5	15.19	0.000	2.66
65+	347.6	27.3	12.73	0.000	2.34

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
84.9636	96.72%	96.71%	96.67%

Figure 2-2: Whole Model Area Zonal Total Vehicle Regression Analysis Comparison

To check that this relationship correctly predicted the number of vehicles available on a sectoral basis, the results were amalgamated into 11 sectors. The total Census sectoral vehicles available were compared to the equivalent model predicted vehicles available.

Table 2-1 shows that whilst reasonably good overall, it was found that this relationship significantly overpredicted vehicle availability in the CBD and Southern suburbs whilst significantly underpredicting vehicle availability in Petone and Upper Hutt.

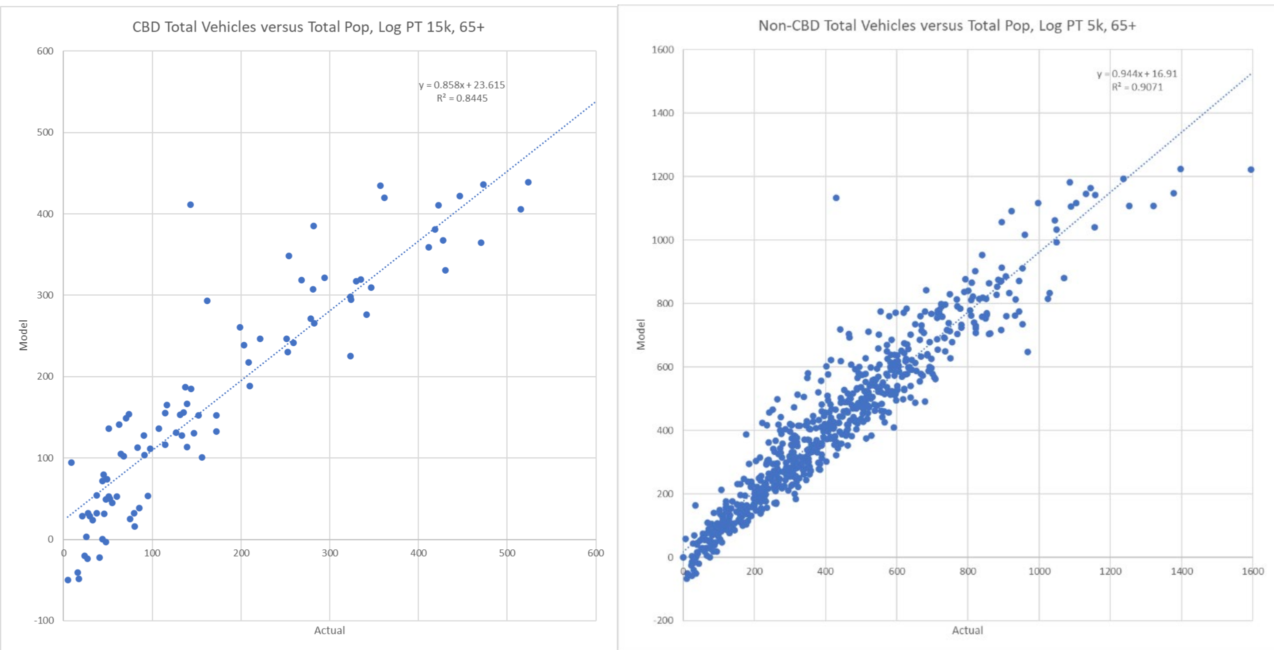
Table 2-1: Sectoral Predicted Total Vehicles Fit Comparison

Sector	Census Vehicles	Predicted Vehicles	% Error
CBD	16812	20538	22%
Eastern suburbs	19300	19703	2%
Southern suburbs	17110	19078	12%
North and Western suburbs	22810	22414	-2%
Tawa and Johnsonville	31688	29584	-7%
Porirua	33078	33736	2%
Kapiti	37196	36182	-3%
Petone	15944	13432	-16%
Lower Hutt	47228	46088	-2%
Upper Hutt	28587	26124	-9%
Wairarapa	33115	31356	-5%
Total	302868	298235	-2%

Given the relative success of this process, it was decided to separate the CBD from the rest of the model area and attempt to find separate relationships for both the CBD and Non-CBD areas.

Data from each area was again regressed against zonal characteristics and the following relationships were found:

- CBD - total zonal population, log10 of PT accessibility of 15000 jobs, and zonal percent of population in the 65+ age group
- Non-CBD - total zonal population, log10 of PT accessibility of 5000 jobs, and zonal percent of population in the 65+ age group



Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Total Pop	0.3759	0.0136	27.71	0.000	1.64
Log PT 15k	428.8	64.7	6.63	0.000	1.98
65+	485.8	70.1	6.94	0.000	1.88

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
60.8640	93.48%	93.27%	92.69%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Total Pop	0.60892	0.00643	94.68	0.000	2.89
Log PT 5k	155.0	12.9	11.97	0.000	2.88
65+	281.6	27.2	10.34	0.000	2.42

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
80.6978	97.32%	97.31%	97.27%

Figure 2-3: Split Model Area Zonal Total Vehicle Regression Analysis Comparison

Again, these relationships showed an acceptable R^2 of 0.8445 and 0.9071 (calculated when not forced through the origin) for the CBD and Non-CBD respectively but showed slopes of 0.86 and 0.94 instead of ideally 1.0. Both did however very nearly pass through the origin.

To check that these relationships correctly predicted the number of vehicles available on a sectoral basis, the results were again amalgamated into the same 11 sectors used previously. The total Census sectoral vehicles available were compared to the equivalent model predicted vehicles available. Table 2-2 shows that whilst the CBD prediction was improved, the non-CBD relationship again significantly overpredicted vehicle availability in the Southern suburbs whilst again significantly underpredicting vehicle availability in Petone and, in a lesser degree, Upper Hutt.

Table 2-2: Sectoral Predicted Total Vehicles Fit Comparison for Separate CBD and Non-CBD Regressions

Sector	Census Vehicles	Predicted Vehicles	% Error
CBD	16812	16645	-1%
Eastern suburbs	19300	20202	5%
Southern suburbs	17110	19608	15%
North and Western suburbs	22810	22937	1%
Tawa and Johnsonville	31688	30276	-4%
Porirua	33078	34296	4%
Kapiti	37196	35932	-3%
Petone	15944	13799	-13%
Lower Hutt	47228	46977	-1%
Upper Hutt	28587	26394	-8%
Wairarapa	33115	31248	-6%
Total	302868	298313	-2%

Having been unable to find any sufficiently accurate predictive regression models for either zonal available vehicles per household or total zonal vehicles, it was concluded that no predictive model was good enough to estimate either model input on a geographically robust basis. In discussion with the client, it is therefore agreed that available vehicles per household will be a direct input into the model and not calculated.

In multiplying the proportions of vehicles per household with the proportions of population by age to form the population matrix for the trip production category model, it was found that vehicles per household did not result in a good fit. So the vehicle allocation model was refitted to be vehicles per person. Average zonal vehicles per person will remain an input to the model. The next section reports the updated vehicles per person allocation model.

3. Vehicle Allocation Model

Having obtained the average vehicle availability of a model zone (through input rather than prediction), this average is used to estimate the proportion of persons in the zone that have households with zero, one, two or three or more vehicles available. This process relies on the development of a set of probability curves that represent the proportion of households at any zonal average level of vehicle availability that fall into each of the different availability levels.

As detailed earlier, census data for vehicles available to the households within each model area was obtained. The number of persons within households in each zone that had zero, one, two and three+ vehicles available were obtained. The zonal total number of vehicles and persons were also used to obtain a zonal average vehicles per person rate.

The percent of persons within each zone that fell into each of the availability levels was calculated along with the relevant overall zonal average vehicle availability. The zonal percentages for each availability level were each plotted against the zonal average to determine any trends and formulate a relationship equation. Figure 3-1 shows the raw zonal level trends for each availability level along with preliminary curves of best fit.

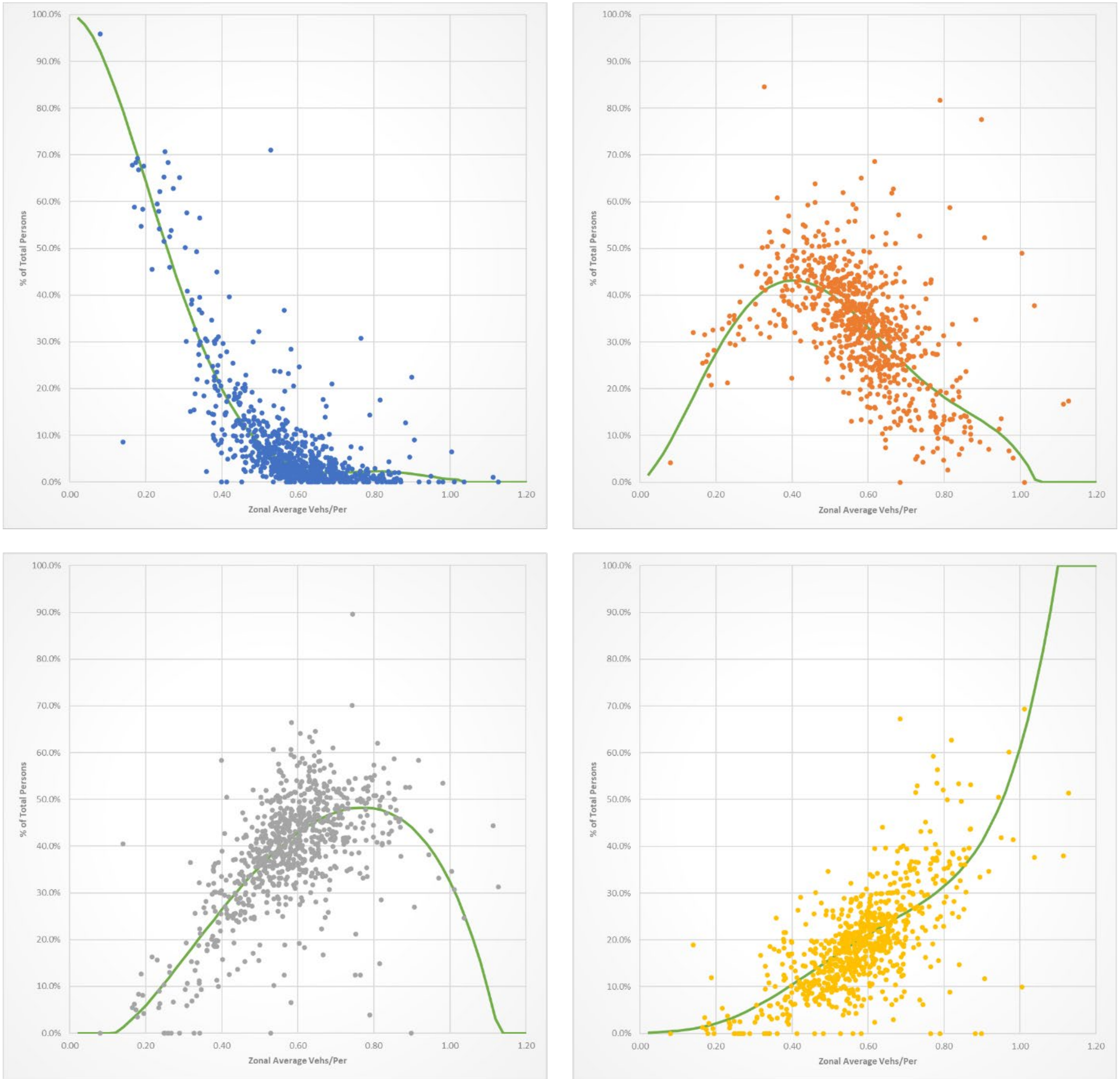


Figure 3-1: RAW Zonal Level Vehicle Availability Trends

Clearly the data shows strong trends within each availability level and produces correlation equations with high levels of confidence. However, to remove some of the "noise" from the data and further improve the confidence in each of the relationship equations, the raw zonal data was aggregated into zonal average bands, or "bins" representing binned increases of 0.02 vehicles/person from 0 up to 1.20.

Figure 3-2 shows the binned data trends for each availability level. The plotted binned data reveals very strong trends for each vehicle availability level. The level of correlation for each of the levels is significant with R^2 values all in excess of 0.93. The third power smoothed curves are shown along with their equations and fit.

For these curves to be used predictively in the model, there are two constraints that need to be satisfied:

- The sum of household/population proportions for each zonal average value, must equal one
- When each proportion, for a zonal average, is multiplied by the number of vehicles they represent, the total must equal the zonal average

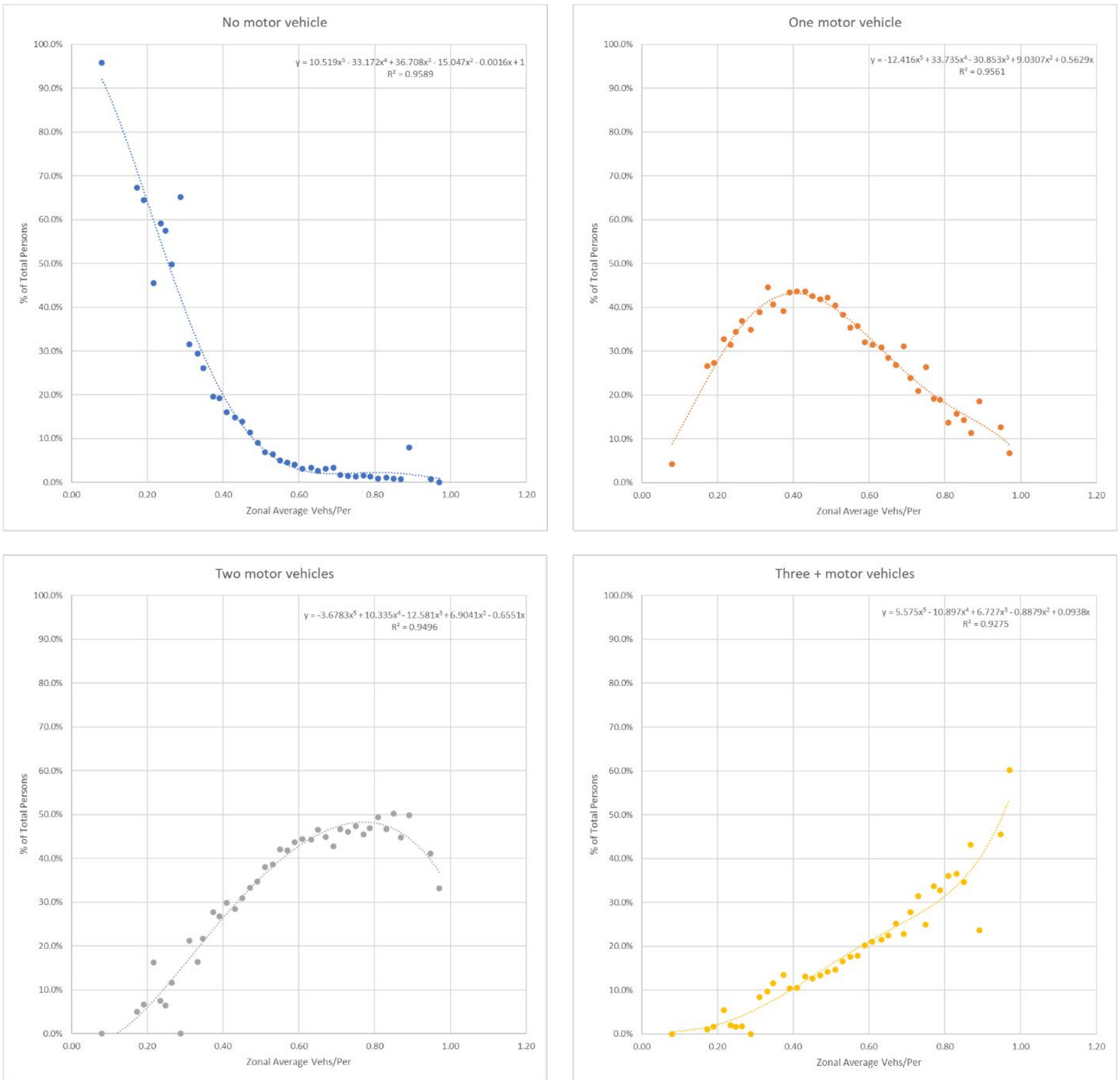


Figure 3-2: Binned Vehicle Availability Trends

The resulting curves are shown below overlaid on the raw data in Figure 3-3.

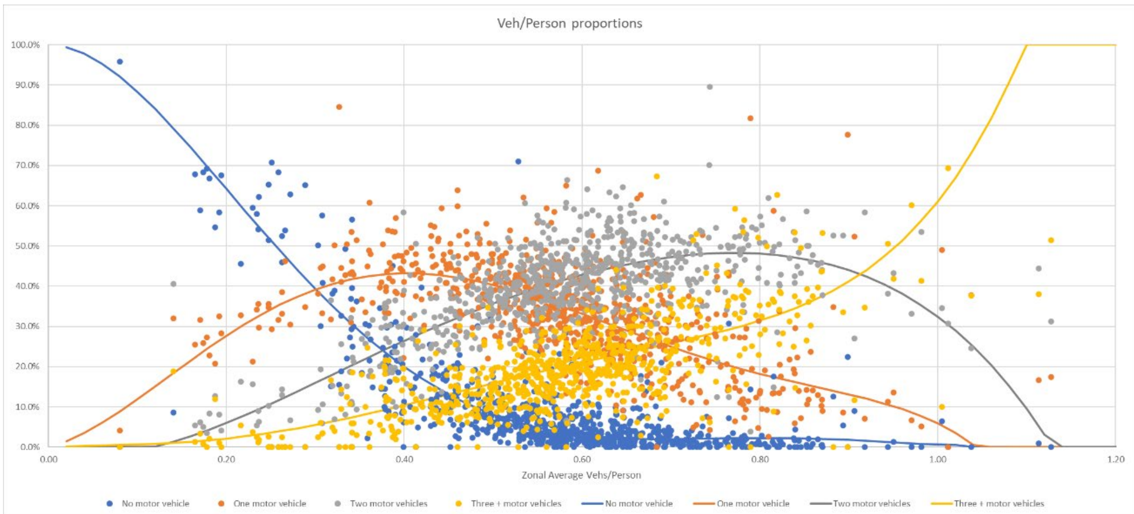


Figure 3-3: Smoothed Vehicle Availability Curves Against RAW Data

For use in the model, a set of fifth-power polynomial equations have been calculated representing the fitted curves for each vehicle availability group. These take the form of:

$$Y = ax^5 + bx4 + cx3 + dx^2 + ex + constant$$

Table 3-1 details the coefficients used in the above equation form for each of the curves. The high number of decimal places contributes to representing the curve more accurately.

As these equations can produce extreme results at high levels of average vehicles per person which can break the need for all percentages to sum to 100%, two limiting conditions have been included:

1. If the zonal average vehicles per person >1.035, force the percent of persons in 0 vehicles household to 0%.
2. Restrict the maximum percent of persons in any household category to 100%.

Table 3-1: Goodness of Fit Statistics for Calibrated Volume Delay Curves

Vehicle Category	a	b	c	d	e	Constant
0 Vehicles	10.51919035	-33.17237996	36.70785814	-15.04688527	-0.001580745	1
1 Vehicles	-12.41589423	33.73454516	-30.85349363	9.030724395	0.562898944	-
2 Vehicles	-3.678292129	10.33503203	-12.58137938	6.904053097	-0.655143356	-
3+ Vehicles	5.574996012	-10.89719723	6.727014862	-0.887892218	0.093825157	-

4. Summary

Regression techniques were applied to estimate average vehicles per household or vehicles per zone. Region-wide, a satisfactory model was produced using total zonal population, $\log_{10}$ of PT accessibility to reach 5000 jobs, and zonal percent of population in the 65+ age group. This model, however, overestimated vehicles in the CBD and Southern suburbs – both areas for future infrastructure and transport system improvement. In discussion with the client, it was agreed that a predictive model would not be used and average vehicles per household per zone would be an input to the model.

Curves were produced to take the average vehicles per households and convert this into the number of households with zero, one, two, or three plus vehicles available. In combining this data with population by age in the second stage of the model, it was found that vehicles per person was a better fit than vehicles per household. So the allocation curves were rebuilt.

Curves were developed that take the average vehicles per person by zone and translate this into the proportion of persons within the zone with zero, one, two, or three plus vehicles available. Applying these proportions to the number of persons in each zone creates the first dimension of the category model to which trip production rates by purpose are applied.

Appendices

Appendix A Client Comment and Consultant Response

No.	Comment By	Comment	Response
1	Andrew Ford	Generally, all good as per previous version.	Noted
2		Over-prediction of vehicle availability in CBD / southern suburbs and under-prediction in Petone / Upper Hutt is interesting, maybe a research topic to understand drivers behind these differences (not for model update, more for general WTAU work).	Noted
3	Ian Clark	I am interested in exploring any potential implications of moving from household to persons for this component of the model. Does it introduce any complications, relative to other components of the model that still use households?	Vehicles per person aligns better with the family structure and trip generation models. The only potential implication I can think of, is that we tend to associate vehicles as a household commodity. But it's just a matter of multiplying households by the people per household to convert to vehicles per person. So we are comfortable there are no negative implications.
4		And like Andy, I was interested in the geographic differences in the vehicle availability. There appears to be an income/affluence issue here, which in itself sounds like a bit of a minefield I'm not sure we want to get into.	Agreed

Christchurch

Hazeldean Business Park, 3 Hazeldean Road
Addington, Christchurch 8024
PO Box 13-052, Armagh
Christchurch 8141
Tel +64 3 366 7449
Fax +64 3 366 7780

Please visit **www.stantec.com** to learn more about how Stantec design with community in mind.