Auckland Tauranga Wellington **Christchurch**

PATTLE DELAMORE PARTNERS LTD

Assessing the Impact of Gross Emitting Vehicles

Waka Kotahi NZ Transport Agency

solutions *for your environment*

Assessing the Impact of Gross Emitting Vehicles

Prepared for

Waka Kotahi NZ Transport Agency

July 2020

PATTLE DELAMORE PARTNERS LTD
Level 2, 134 Oxford Terrace
Christchurch Central, Christchurch 8011
PO Box 389, Christchurch 8140, New Zealand

Tel +64 3 **345 7100** Fax +64 3 **345 7101**
Website **http://www.pdp.co.nz**
Auckland Tauranga Wellington **Christchurch**

solutions for your environment

Quality Control Sheet

TITLE	Assessing the Impact of Gross Emitting Vehicles
CLIENT	Waka Kotahi NZ Transport Agency
VERSION	Final
ISSUE DATE	3 July 2020
JOB REFERENCE	C03954800
SOURCE FILE(S)	C03954800_R001_Final

DOCUMENT CONTRIBUTORS

Prepared by

SIGNATURE

Sima Bagheri   Jeff Bluett

Reviewed and Approved by

SIGNATURE

Jeff Bluett

Limitations:

This report has been prepared by Pattle Delamore Partners Limited (PDP) on the basis of information provided by Waka Kotahi NZ Transport Agency. PDP has not independently verified the provided information and has relied upon it being accurate and sufficient for use by PDP in preparing the report. PDP accepts no responsibility for errors or omissions in, or the currency or sufficiency of, the provided information.

This report has been prepared by PDP on the specific instructions of Waka Kotahi NZ Transport Agency for the limited purposes described in the report. PDP accepts no liability if the report is used for a different purpose or if it is used or relied on by any other person. Any such use or reliance will be solely at their own risk.

Acknowledgements

Pattle Delamore Partners Limited (PDP) wish to acknowledge the following people who made positive and valuable contributions to this project:

- Greg Haldane and Sharon Atkins (Waka Kotahi NZ Transport Agency) for their collegial project management, helpful ideas, and review comments;
- Paul Boulter (EEM Consulting) for his insightful external peer review of the report;
- James Eden and Iona Collins (Waka Kotahi NZ Transport Agency) for assistance with the provision of data from the Motor Vehicle Register; and
- Tim Green (PDP) R-coding and data analyst.

Executive Summary

Between 2003 and 2015, five road-side vehicle emission monitoring campaigns using remote sensing devices (RSD) were completed in Auckland, New Zealand. This assessment used data from the most recent campaign. The data set was augmented using additional data for the recorded plate details from the 2018 Motor Vehicle Register (MVR).

The objective of this project was to determine the impact that gross emitters have on emissions of pollutants harmful to human health from light duty vehicles within the vehicle fleet. The project consisted of:

- Stage 1: assessment and characterisation of the monitored fleet results, including repeatability of emissions measurements for gross emitting vehicles (GEVs), identification of GEV characteristics to create a predictive model for GEVs, and its application to the national fleet; and
- Stage 2: determination of the relative impact of emissions measurements for GEVs, including estimation of potential emission reduction benefits of GEV replacement.

Emissions of pollutants considered harmful to human health were assessed. These were carbon monoxide (CO), hydrocarbons (HC), nitrogen monoxide (NO), and particulate matter (PM). Vehicle emissions were ordered and the top 3% of emissions were categorised as GEVs. The overlap of GEVs for the pollutants considered was investigated and most GEVs were found to be GEVs for a single pollutant of the four studied. The highest crossover between pollutants was that between CO GEVs and HC GEVs.

Assessment of GEVs with more than one associated emissions record showed that second measurements for the same vehicle were higher than the majority of TEV measurements in the fleet indicating that GEVs tend to produce consistently higher emissions measurements.

Analyses of the data set were undertaken to find categories of vehicle characteristics for which the proportion of GEVs was higher than the overall proportion in the monitored vehicles data set. Assessment of the 2015 RSD data set showed different fuel types associated with higher emissions of different pollutants and vehicles older than 14 – 16 years, with odometer readings above 150,000 – 200,000 km, or having been tested to older emission standards to have higher proportions of GEVs than the overall monitored fleet.

Comparison between vehicles classified as GEVs and TEVs showed that GEVs had a higher rate of retirement from the fleet than TEVs, with removal or replacement of vehicles between 2015 and 2018 measured as 30-38% for GEVs and 15% for TEVs in the monitored fleet. GEVs for CO, HC and NO had slightly

lower average annual distance travelled (AADT) than their respective TEVs, whereas PM GEVs travelled further than TEVs as an annual average.

Regression tree analysis (RTA) was also undertaken for the data to identify clusters (regression tree nodes) of common criteria with the highest emissions. This analysis was used to construct a predictive GEV model. Data from the December 2019 MVR was sourced and assessed using the RTA-derived criteria to identify the regional distribution and AADT of vehicles modelled as potential GEVs for each pollutant.

Major cities were generally found to have lower proportions of modelled GEVs whereas Northland and West Coast regions had generally higher proportions of vehicles modelled as GEVs. For all pollutants, modelled GEVs were shown to have higher AADT values than corresponding TEVs

Finally, the pollutant emissions per vehicle per year was calculated and the difference between median GEV and TEV values was used to estimate the annual emission reduction benefit of replacing a GEV in the fleet. This was extended to quantify the potential national annual pollutant reductions available if it were possible to replace all GEVs in the fleet.

The details presented in this report demonstrate that the objective of this project has been met and the investigation of targeted issues successfully completed.

During this project, PDP and the external peer reviewer identified questions which, if investigated, would provide additional insight into the impacts and management of GEVs and therefore add value to the outcomes of this study. Using this experience, PDP have recommended that future work programmes consider a number of GEV related tasks.

Table of Contents

Table of Figures

Table of Tables

Appendices

1.0 Background and Project Objectives

1.1 Background

Roadside vehicle emission monitoring using remote sensing device (RSD) technology is recognised internationally as a useful and cost-effective method of collecting large amounts of real-world vehicle emission data. This is evidenced by its use by large international organisations such as the International Council on Clean Transportation[1] and the United States Environmental Protection Agency (USEPA) (2004).

Between 2003 and 2015, five RSD monitoring campaigns were completed in Auckland, New Zealand. The New Zealand RSD database has been widely used to characterise the types of vehicle that constitute Auckland's on-road vehicle fleet and quantify their emissions. While the five RSD monitoring campaigns have been undertaken in Auckland, the findings and conclusions of these projects have nationwide implications. Gaining an understanding of how real-world vehicle fleet emissions are changing with time can flag where additional vehicle emission-reduction strategies and policies may be required.

The New Zealand Motor Vehicle Register (MVR) contains over 5.3 million records of NZ-registered vehicles[2]. This incorporates a range of vehicle types including mopeds and motorbikes, caravans, busses, agricultural vehicles, and trucks as well as passenger cars, vans, and utility vehicles.

There are just under 45,000 valid vehicle emission measurements in the 2015 RSD database. The RSD measures the emissions of four pollutants: carbon monoxide (CO), hydrocarbons (HC), nitrogen monoxide (NO) and uvSmoke (a proxy for particulate matter (PM)). On-road emission measurements can be linked to individual vehicles using the vehicle's licence plate and data contained in the MVR. The linking of emission measurements with vehicle characteristics allows investigation into the impact of vehicle age, odometer reading, engine size, and emission-control technology on vehicle emissions.

Previous remote sensing campaigns have showed the increasing age of New Zealand's on-road vehicle fleet and the disproportionate impact of a small number of high or "gross" emitting vehicles. Understanding the characteristics of gross emitting vehicles (GEVs) and quantifying their emissions enables the identification of possible and effective interventions for these vehicles.

Integrating the information extracted from the RSD database on GEVs with the data contained in the MVR on a national level provides an even more powerful tool. Aligning these two sources of data will potentially allow us to advance our

[1] The Real Urban Emissions Initiative (TRUE)

[2] As at 6 June 2020.

understanding of GEVs by identifying their likely numbers, locations, and annual travel distances to allow better assessment of their impact on total emissions.

1.2 Project Objectives

The objective of this project was to determine the impact of GEVs on emissions of pollutants harmful to human health from the vehicle fleet across New Zealand. The following issues were identified for investigation:

- Prevalence of GEVs:
  - What proportion of the fleet are potentially GEVs?
  - How long do GEVs remain in the fleet?
  - What distance do GEVs typically travel in a year?
  - What is the regional distribution (geographical spread) of GEVs?
- Determination of replicability of GEV results;
- Determination of the potential impact of GEVs compared to emissions from the wider light duty vehicle (LDV) fleet; and
- Estimation of the emission reduction benefit of removing GEVs from the fleet.

The project was undertaken in two stages.

In Stage 1, the number, characteristics, regional distribution, and activity of GEVs was assessed, including:

- Defining the proportion of the fleet to be considered as GEVs as the highest 1, 2, or 3% of emissions measurements. Defining a set of GEVs for each of the four pollutants monitored by the RSD;
- Determining whether GEVs had been measured more than once in the 2015 RSD campaign and whether monitored emissions had changed;
- Defining characteristics of the GEV set for each RSD-monitored pollutant including vehicle age, fuel type, odometer reading, and emission control;
- Assessing the rate of retirement of GEVs within the monitored fleet by comparing the MVR's 'active registration' status for GEVs monitored in 2015 and for the same vehicles in late-2018;
- Calculation of the distance travelled by GEVs within the monitored fleet by comparing the MVR's 'odometer reading' data for GEVs monitored in 2015 and for the same vehicles in late-2018; and
- Establish a geographical (regional) distribution of vehicles listed in the MVR with GEV characteristics.

In Stage 2, the information developed in Stage 1 was used to:

- Determine the potential impact of harmful pollutant GEVs compared to the emissions from the wider LDV fleet; and
- Estimate the emission reduction benefit of removing GEVs from the fleet.

The results presented in this report have detailed the findings of the assessment for CO accompanied by a summary of the results and patterns observed for HC, NO and PM. Where figures for CO have been included in the body of the report, equivalent figures for HC, NO and PM are contained in Appendix H: Figures for Other Pollutants of Interest. Summary tables showing absolute values and percentages for all pollutants are contained in Appendix I: Summary Tables of GEV Characteristics.

The CO results were presented as a case study for the report because this pollutant clearly demonstrated the effects of fuel type, vehicle age, odometer reading and vehicle emission standards.

2.0 Stage 1

2.1 Data Preparation

The method used for the 2015 RSD monitoring was as previously detailed in Section 4 and Appendix B of NZ Transport Agency research report 596 (Bluett et al., 2016).

During the 2015 RSD monitoring campaign, a total of 54,539 emissions records were collected during 10 monitoring sessions over seven baseline sites. The speed and acceleration measurements were used to derive vehicle specific power (VSP). VSP is a performance measure for determining whether a vehicle is operating within an acceptable power range when measured by remote sensing. The emissions data from a vehicle was only considered valid if its VSP value fell between zero and 40 kW/tonne. This resulted in 44,826 valid 2015 RSD emissions records. Of these, 38,601 had recorded plate values associated with the record. The number of emissions records remaining following each data processing step is summarised in Table 1.

Table 1: 2015 RSD Monitoring Campaign Records	
Emissions Records Collected	54,539
Valid Emissions Records	44,826
Valid Records with Plate Details	38,601
Valid Records with Matched Plate Details (2015 & 2018)[1]	26,756

Table 1: 2015 RSD Monitoring Campaign Records	
Number of Vehicles with Matched Records (2015 & 2018)[2]	22,900

Notes:

1. *Reduction in number of records from 'Valid records with plate details' reflects various data cleaning steps as detailed.*
2. *Some vehicles went through the remote sensor more than once and therefore the number of individual vehicles captured is lower than the number of valid matched test results.*

Plate details for the 2015 RSD data had been saved from the MVR at the time of testing, and a data set combining full records of all vehicle plates for which emissions data had been collected was used for the analysis.

Several data cleaning steps and adjustments were undertaken to remove data outside the scope of the assessment and to ensure that incorrect emissions values would not have an influence on overall fleet results. These included:

- Updated HC values were used to replace the original HC readings[3];
- Vehicles of types other than 'Passenger Car/Van' and 'Goods Van/Truck/Utility' were not included in the analysis;
- Vehicles with a gross vehicle mass (GVM) outside the range of 500 kg – 3,500 kg were not included in the analysis to limit the study to light-duty vehicles;
- Only cars powered by petrol, diesel, and LPG were considered for the analysis; and
- Where more than one vehicle record existed for a single plate (i.e. the plate had previously been associated with a different vehicle), only the record with valid registration during the testing period in 2015 was retained.

A second data set was accessed from the MVR to provide 2018 data on the vehicles measured in 2015. These records were matched to the 2015 data set by plate, and odometer details and registration status in 2018 were added to the existing vehicle records. This was completed to allow investigation of the average annual distance travelled (AADT) and the rate of replacement of the vehicles assessed.

The vehicle data was matched by plate number to the emissions data. Where more than one emissions record existed for a single plate, the vehicle details corresponding to that plate were uploaded to each of the emissions records.

[3] The HC values initially recorded were investigated in depth and adjusted as detailed in Bluett et al. (2016) (Appendix C: Quality assurance of data, C4 2015 Hydrocarbon data).

This analysis resulted in a data set of 22,900 unique vehicles and 26,756 valid emissions records with complete corresponding vehicle data.

2.2 Definition of GEVs

This section outlines the initial assessment of the emissions data from the monitored fleet and the reasoning for selecting the top 3% of emissions to be categorised as GEVs.

The cleaned and matched data set was ordered from highest to lowest emission values and a cumulative distribution created. Each emissions measurement (in %CO) was divided by the sum of all of the emissions measurements to give the proportional contribution of the record to the total emissions. This was used to create a cumulative distribution plot to show the proportion of total emissions against the proportion of the vehicle fleet responsible. Estimation of the point on the charts at which the initial high rate of change began to reduce was used to define a cut-off point for characterisation of vehicles as GEVs or TEVs.

Figure 1 shows the cumulative distribution plot produced for CO emissions records from the 2015 RSD data set. Table 2 shows the associated rate of change in cumulative CO emissions for each percentage increase in valid readings up to 5%. The rate of change in cumulative CO emissions reduces significantly with each percentage increase of valid readings. Following visual assessment of the graphs and tables for each of the pollutants of interest, it was recommended to classify the highest 3% of valid emissions readings as GEVs, with the remainder of the vehicles classified as TEVs[4]. It is more typical for the highest 10% of emissions records to be used to define GEVs but the 3% definition was recommended as it captures a significant proportion of the total emissions while including a relatively small number of vehicles which could potentially be targeted through policy interventions. GEV and TEV labels were added to each of the emissions records for each of the four pollutants of interest. These were CO, HC, NO and PM.

In the example of CO, the 3% of valid vehicle emissions records classified as GEVs account for 44% of the total cumulative CO emissions.

[4] Email from Jeff Bluett (PDP) to Greg Haldane and Sharon Atkins (Waka Kotahi NZ Transport Agency), dated 17 June 2019. Relevant excerpts contained in Appendix A: Email Excerpts Outlining Reasoning for 3% GEV Cut-off Criteria.

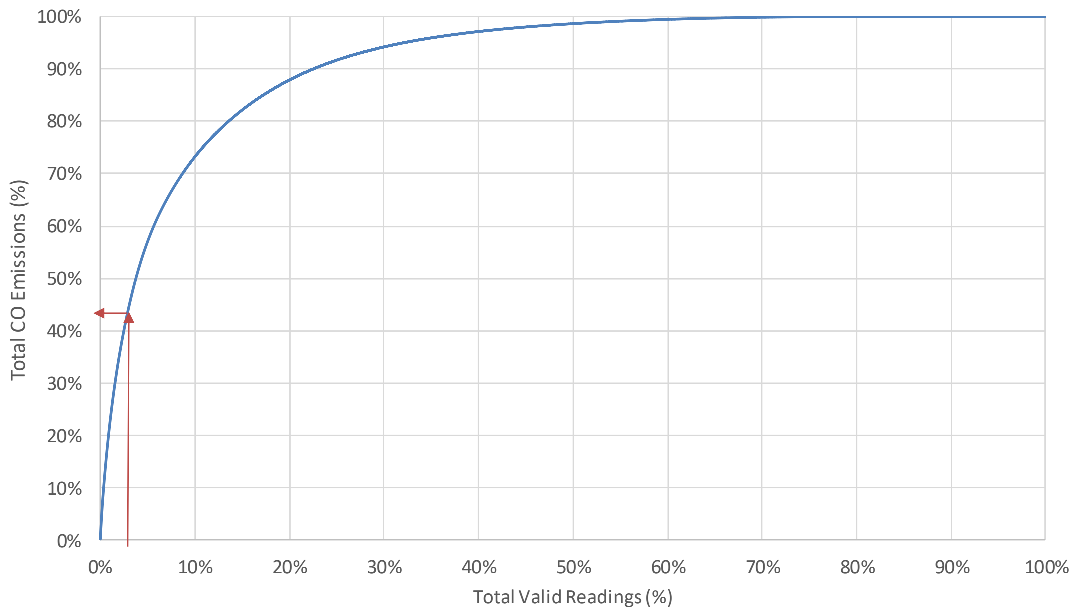


Figure 1: CO Cumulative Distribution Plot

Table 2: % total emissions at n% total valid readings (CO, 2015 RSD Data Set)		
Valid Readings	**Cumulative Emissions (%CO)**	**Rate of Change[1]**
1%	22%	22%
2%	35%	13%
3%	44%	9%
4%	52%	7%
5%	57%	5%

Notes:

1. *Rate of change is calculated stepwise by*

% Emissions(n) – % Emissions(n-1)

where n = % of valid readings.

For the initial cleaned data set of 26,756 records, this gave 803 records, resulting in a GEV count of 762 vehicles for CO and 741 – 779 vehicles for the other pollutants with variation due to some vehicles recording multiple emissions records. The top 3% of valid emissions records captured 25% of cumulative HC emissions, 26% of cumulative NO emissions, and 31% of cumulative PM emissions. Table 3 shows the absolute RSD measurement values associated with the 3% GEV cut-off.

Table 3: Measurement Values for GEV Classification (Top 3% of emissions, 2015 RSD Data Set)		
Pollutant	**GEV Values (≥)**	**Units**
CO	1.935	% CO
HC	469.81	ppm HC
NO	2367	ppm NO
PM	0.279	g PM/kg fuel

Categorisation of the emissions records as GEVs and TEVs for each pollutant allowed a series of analyses to be completed to investigate the characteristics of the vehicle subsets and look for similar characteristics in GEV vehicles for each pollutant.

2.2.2 Do GEVs Discharge High Amounts of All Pollutants?

This section investigates the crossover of vehicles classified as GEVs for the four pollutants of interest to identify whether there appears to be a strong relationship between GEVs of different pollutants.

Emissions labels were used to categorise each vehicle classified as a GEV for any pollutant dependent upon which pollutants they had recorded GEV emissions for. The information was then presented to show the number of vehicles recording high pollutant readings for the four pollutants of interest and the crossover of vehicles recording high readings for more than one of the pollutants.

Figure 2 shows the crossover of vehicles categorised as GEVs by the four pollutants of interest: CO, HC, NO and PM.

With the exception of HC, each pollutant has more vehicles classified as GEVs solely for that pollutant than in the combined crossover with other pollutants. The highest crossover between two pollutants is CO and HC with 163 vehicles (21% of both CO and HC GEVs). The crossover between HC and PM is over 10% of the total GEVs for each pollutant (12% of HC GEVs and 13% of PM GEVs) with 95 vehicles.

Crossover between three pollutants returned 6% of the total GEVs for CO, HC, and PM and for HC, NO and PM whereas the other two groups returned less than 1%. There were 9 vehicles recorded as GEVs for all four pollutants of interest (around 1%). The implications of this are that it is unlikely that it will be possible to develop a single policy to target GEVs of all pollutants and it is likely that pollutant-specific targeting of potential GEVs will be the most effective approach.

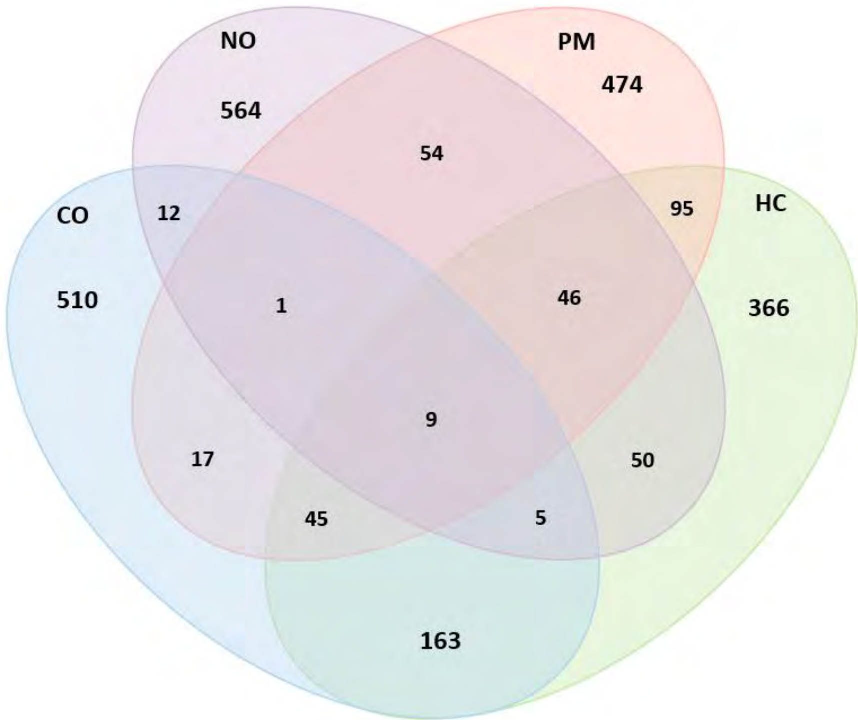


Figure 2: Venn diagram showing crossover of GEVs for the four pollutants investigated

2.3 Repeatability of GEV Measurements

This section evaluates the repeatability of GEV measurements with the intention of understanding whether vehicles that record an RSD emissions measurement that categorises them as a GEV will consistently produce a similar level of emissions that would be high enough to result in consistent categorisation as a GEV. This gives a useful indication of test to test variability of emissions from GEVs.

Figure 3 shows a box and whisker plot for CO GEVs in the 2015 RSD data set that recorded more than one emissions measurement. Equivalent figures for HC, NO and PM are contained in Appendix H: Figures for Other Pollutants of Interest.

There was a total of 166 CO GEVs in the data set with more than one associated emissions record. A vehicle was classified as a GEV when it had a single emissions measurement within the top 3% of valid readings so additional measurements may have been above or below the GEV cut-off value. The boxes show (from left to right), the highest emissions value recorded by the vehicle, the second highest value recorded by the vehicle, and the measurements of vehicles classified as TEVs in the 2015 RSD data set. This shows that for CO GEVs, the second highest measurements are significantly higher than those recorded in the TEV population. This can be inferred from the fact that there is no overlap of the interquartile range between the two sample groups. The median of the 'GEV 2nd Highest Record' series is higher than the top whisker of the TEV series which is indicative of the fact that 64% of the second measurements of these GEVs would be considered outliers if they were included in the TEV population. This indicates a strong relationship between vehicles categorised as CO GEVs, and repeated results of higher CO emissions than the majority of the population.

The HC GEVs in the 2015 RSD data set showed more of an overlap between second highest GEV measurements and TEV measurements. This indicates that the HC GEVs may produce less repeatable results than those shown by CO GEVs in the fleet.

NO and PM GEVs in the 2015 RSD data set showed less overlap between second highest GEV measurements and TEV measurements than both HC and CO GEVs. This shows a strong indicator for repeatability of emissions results for NO and PM GEVs. The number of GEVs in the data set with more than one associated emissions record varied from 147 vehicles for NO GEVs to 181 vehicles for PM GEVs.

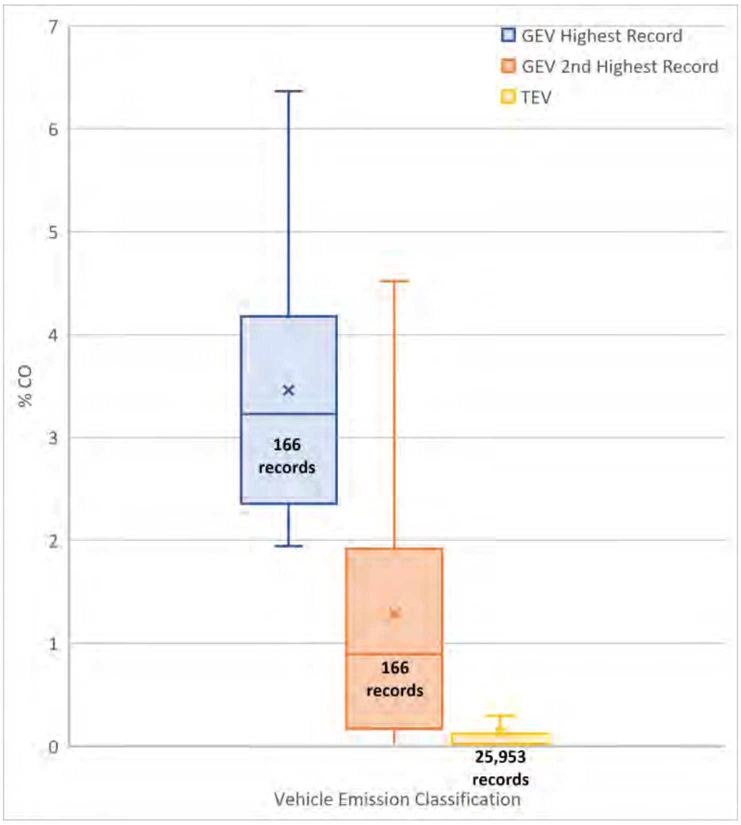


Figure 3: CO GEVs with duplicate measurements compared with TEV fleet emissions

To further check the repeatability of the measurements, the median and mean of the second highest records datasets were calculated and compared to the absolute GEV cut-off value as stated in Table 3 of Section 2.2. The percentages of second highest measurements exceeding the GEV cut-off value and that would be considered outliers in the TEV data set were also calculated. These results are shown in Table 4 below.

Table 4: Comparison of second highest reading from GEVs with multiple measurements and GEV cut-off values

Pollutant	CO	HC	NO	PM
Number of GEVs with multiple measurements	166	160	147	181
GEV cut-off value	1.935% CO	469.81 ppm HC	2367 ppm NO	0.279 g PM/kg fuel
Median of second highest record	0.894% CO	187 ppm HC	2127 ppm NO	0.212 g PM/kg fuel
Mean of second highest record	1.284% CO	237 ppm HC	2019 ppm NO	0.256 g PM/kg fuel
Percentage of second highest records exceeding GEV cut-off value	25%	14%	40%	32%
Percentage of second highest records which would be considered outliers in TEV data set	64%	35%	81%	81%

The percentage of second highest records exceeding the GEV cut-off value provides a quantitative indicator of the repeatability seen in the GEV results within the 2015 RSD data set. The best repeatability is observed in NO GEVs, with 40% of GEVs recording two emission measurements above the GEV cut-off value. The least repeatability is observed in HC GEVs with only 14% of GEVs recording a second emission measurement higher than the GEV cut-off value.

Assessment of GEVs with more than one associated emissions record showed that second measurements for the same vehicle were higher than the majority of TEV measurements in the fleet indicating that GEVs tend to produce consistently higher emissions measurements. While the second highest measurement from GEVs are frequently below the GEV cut-off value, they are still higher as a group than TEV emissions as shown by the lack of overlap between the interquartile ranges of the two sample groups. Given this result, PDP consider the use of RSD measurements to identify GEVs is a robust and pragmatic method.

2.4 Defining the Characteristics of GEVs

This section assesses patterns in the types of vehicles that were classified as GEVs within the 2015 RSD data set with regard to the characteristics of fuel type, vehicle age, odometer reading, and vehicle emission standards.

2.4.1 Vehicle Age

The age of the vehicles in the data set as at 2015 was calculated using the year of vehicle manufacture. The vehicle ages were separated into bins for every two-year increment in vehicle age. The number of GEVs in each two-year age category was then divided by the number of vehicles in the entire monitored fleet in the same category to find the proportion of the fleet that were GEVs for each age category. The proportion of GEVs in the total monitored fleet was also calculated. Proportions of GEVs in each age category were then plotted and compared to the incidence of GEVs in the overall monitored sample.

A similar process was repeated for fuel type, odometer reading and emission standard categories.

Figure 4 shows the proportion of CO GEVs vehicles within each age category. It is unsurprising that the proportion of GEVs increases with vehicle age as older emission standards allowed higher levels of emissions, and vehicles engines generally deteriorate as they age due to wear and tear. The graph shows that vehicles over 14 years old in the monitored fleet have a higher proportion of CO GEVs than the overall monitored fleet, with the proportion of CO GEVs increasing with age. Conversely, vehicles under 14 years old have a lower proportion of CO GEVs than that seen in the monitored fleet.

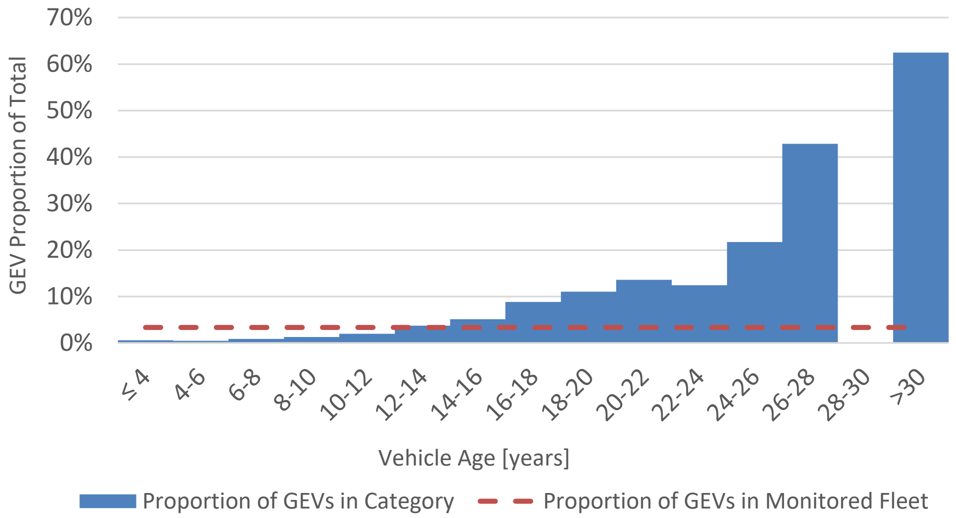


Figure 4: Vehicle Age of CO GEVs

The relationship with vehicle age is similar for HC, NO and PM GEVs, with increased proportion of GEVs correlating to increased vehicle age categories, and a vehicle age of around 14 years indicating the point at which the proportion of GEVs in the age category is higher than the proportion of GEVs in the monitored fleet as a whole. This age is slightly older for PM GEVs at about 16 years old, which may reflect the higher incidence of diesel vehicles in this group.

2.4.2 Fuel Type

The breakdown of fuel types of CO GEVs is shown in Figure 5 below. The number of vehicles of each fuel type classified as GEVs has been divided by the total number of vehicles of that fuel type in the monitored fleet. This allows comparison with the entire fleet proportion of GEVs at around 3% to view which fuel types have higher or lower proportions of GEVs when compared to the monitored fleet. The actual fleet proportion of GEVs is slightly higher than 3%. This is because 3% represents the proportion of emissions records categorised as GEVs from 26,756 records and these are associated with 22,900 vehicles. The frequency of multiple readings in GEV and TEVs causes the adjustment of the actual GEV proportion in the monitored fleet.

The results show that there are a slightly higher proportion (4%) of CO GEVs that are petrol-fuelled than in the monitored fleet and a significantly lower proportion of CO GEVs that are diesel-fuelled (0.1%) than in the monitored fleet.

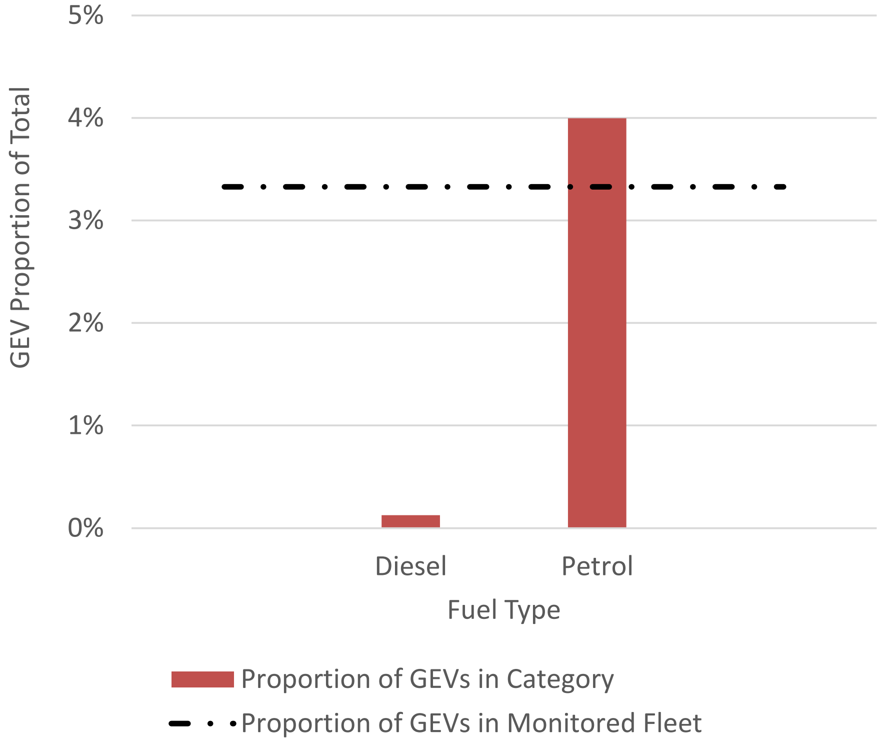


Figure 5: CO GEVs by Fuel Type

There is a similar proportion of HC GEVs that are petrol-fuelled and diesel-fuelled in the monitored fleet. Both categories appear to be slightly over 3%, this is due to the monitored fleet proportion of GEVs being 3.4% of the total vehicle count. This does not indicate that the number of petrol-fuelled and diesel-fuelled HC GEVs is similar, the split in the monitored fleet is roughly 83% petrol and 17% diesel and the split in the numbers of HC GEVs is similar.

When compared to the entire monitored fleet, there is a higher proportion of NO GEVs in the petrol-fuelled vehicles subset at 4%, and a lower proportion of NO GEVs in the diesel-fuelled vehicles subset at 2%. The most distinct difference from the monitored fleet proportion of GEVs occurs in diesel-fuelled PM GEVs. These have a proportion of 13% with petrol-fuelled vehicles having a lower proportion than that of the monitored fleet at 1%.

Absolute values for vehicle numbers of each fuel type in the GEV subset and the monitored fleet are shown in the summary tables in Appendix I: Summary Tables of GEV Characteristics. For vehicle fuel type, the percentage make-up of the monitored fleet in comparison to each of the GEV subsets is shown in Figure 6.

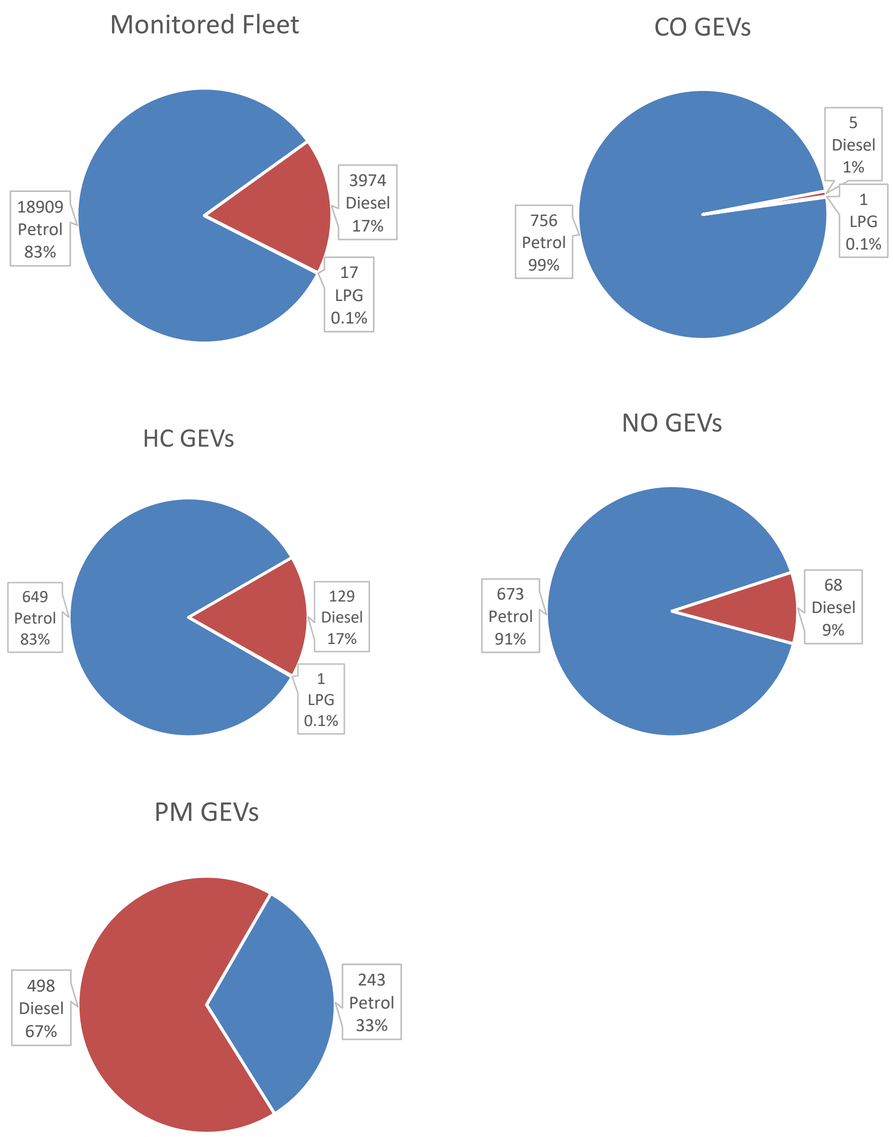


Figure 6: Proportion of vehicles by fuel type in (a) monitored fleet, (b) CO GEVs subset, (C) HC GEVs subset, (d) NO GEVs subset, (e) PM GEVs subset. Labels show number of vehicles, fuel type and proportion.

2.4.3 Odometer Reading

Figure 7 shows the proportion of CO GEVs within categories of odometer reading values. The proportion of GEVs in each category increases with increased odometer readings to a point, but then decreases when odometer readings rise above 350,000 km. One possible anecdotal reason for this may be that vehicles whose engines are still functional after this amount of travel have been either replaced or better maintained than vehicles which are retired from the fleet with lower odometer readings due to engine failure or other mechanical issues.

The data labels on the chart show the absolute number of GO GEVs in each category. The low numbers of vehicles in the categories with odometer readings above 350,000 km may result in skewed proportional values due to the increased effective weight of each additional vehicle on the smaller category. This may also help to explain the variation in proportions of CO GEVs reported for categories above 350,000 km.

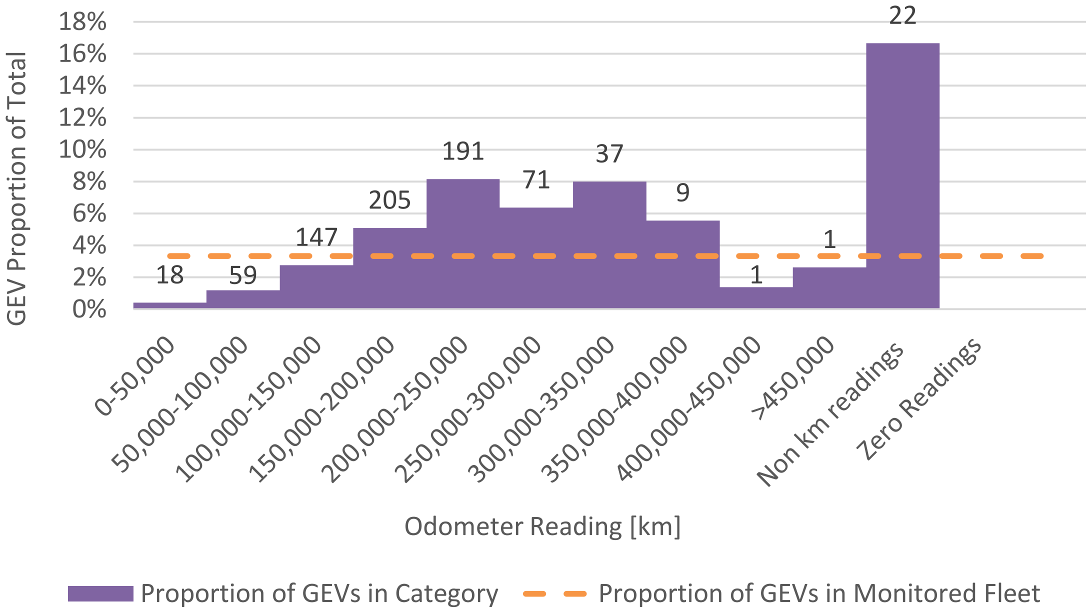


Figure 7: Vehicle odometer reading of CO GEVs (data labels show number of GEVs in each category)

The relationship between HC GEVs and odometer reading is similar to that of CO GEVs whereas NO and PM GEVs show a steady increase in proportion of GEVs with increased odometer reading categories. Vehicles with an odometer reading in the 150,000 – 200,000 km category and above show a higher proportion of GEVs for CO, HC and NO than that of the monitored fleet, whilst vehicles with an odometer reading of 200,000 – 250,000 km and above show a higher proportion of PM GEVs than the monitored fleet. Notably, the absolute PM GEV numbers in the higher odometer categories (> 350,000 km) are higher than those of CO, HC and NO GEVs. Again, this is likely due to the higher proportion of diesel vehicles included in the PM GEV group.

2.4.4 Vehicle Emission Standards

The vehicle emission standards listed for vehicles in the MVR for the monitored fleet were numerous and varied with time, fuel type and geographical origin of the vehicles. To prevent the fleet being divided into multiple small groups depending upon fuel type and country of origin, an assessment of the vehicle emission standards was completed, and standards from different countries were grouped into categories of similar standards as shown in Table 5. Additional categories were created for vehicles from countries with no standard specified on the MVR vehicle record. Petrol- and diesel-fuelled vehicles were considered separately, and the categories were grouped as early, intermediate, and recent emission standards with more stringent emissions restrictions being implemented over time. These three groups were colour-coded when plotted. This categorisation of vehicle emission standards was a pragmatic decision so there is some overlap in years between standards imposed in different countries.

Table 5: Grouping of Emission Standards into Categories

Category Name	Category 'Age'	Emission Standard Labels (from RSD Database[1]) Included
Pre-2000 Standards & Euro 2	Early	Pre-1993, 1993-94, 1997-99, pre-1998, 1998, Euro 2
2000-2005 Standards & Euro 3	Intermediate	2000-2002, 2002-04, 2005, Euro 3
2005 Standards Onwards, Euro 4 & Euro 5	Recent	2005-07, 2009, Euro 4, Euro 5

Notes:

1. Labels in RSD Database from 'trLabel' category.

Figure 8 and Figure 9 show the incidence of CO GEVs in the emission standard categories for petrol- and diesel-fuelled vehicles respectively. There is a clear reduction in the proportion of petrol-fuelled CO GEVs with more recent and more stringent standards. The four categories with a higher proportion of CO GEVs than the overall monitored fleet are pre- and post-1998 Japanese vehicles with unknown standards, pre-2000 standards and Euro 2, and pre-2003 European vehicles with unknown standards.

All diesel vehicle categories have a proportion of CO GEVs of less than 0.5%. However, there is still a downwards trend in the proportion of CO GEVs with more recent and more stringent standards.

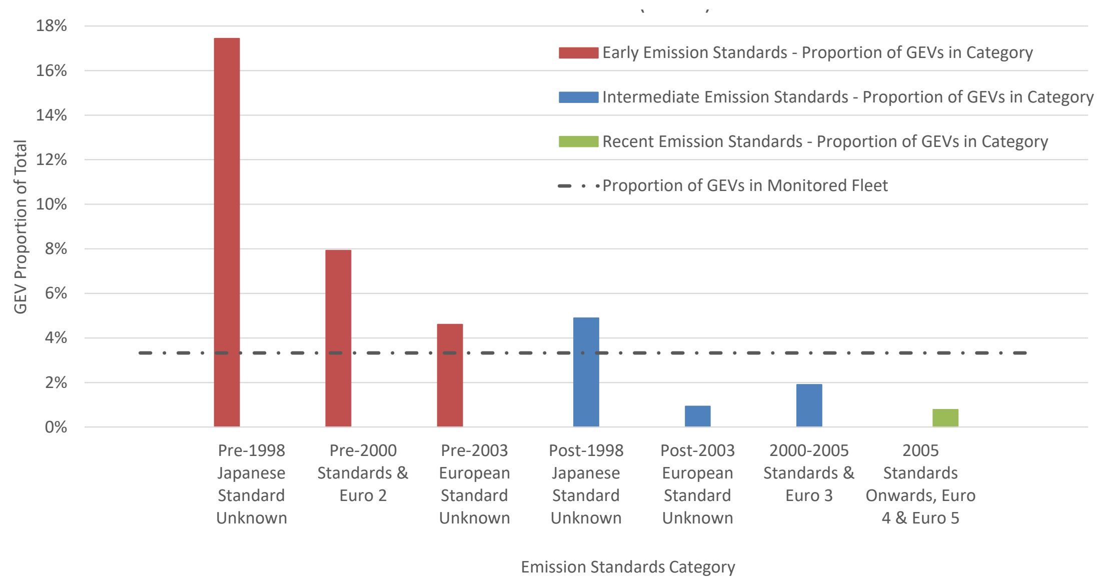


Figure 8: Emissions test regime categories of petrol-fuelled CO GEVs

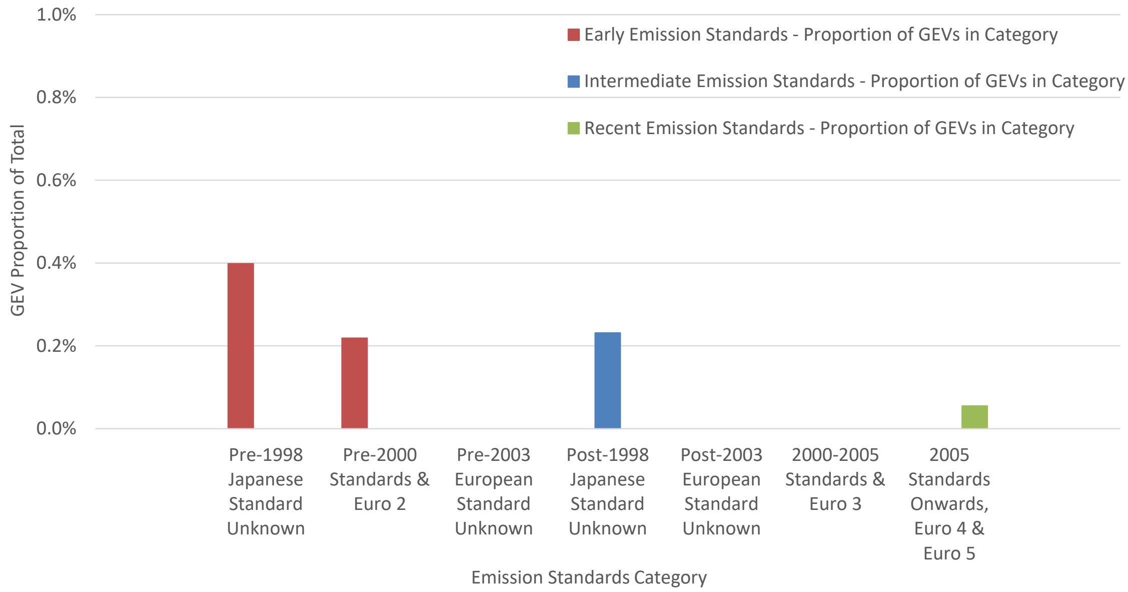


Figure 9: Emissions test regime categories of diesel-fuelled CO GEVs

The HC, NO and PM GEVs all showed similar trends, with relative proportion of GEVs reducing with inclusion of more recent and more stringent emission standards in the categories.

2.5 Retirement Rate of GEVs

This section assesses and compares the relative rate of retirement of GEVs and TEVs from the monitored fleet.

The vehicle registration status reported for 2018 was used to deduce whether vehicles with valid emissions recorded in the 2015 RSD data set were still part of the active vehicle fleet in 2018. This allowed the proportion of active and inactive vehicles to be plotted for GEVs and TEVs to provide a comparison of the rate that each was retired from the fleet during the 2015 – 2018 period.

Figure 10 shows the proportion of vehicles grouped as CO GEVs and TEVs that had active and inactive registration in 2018.

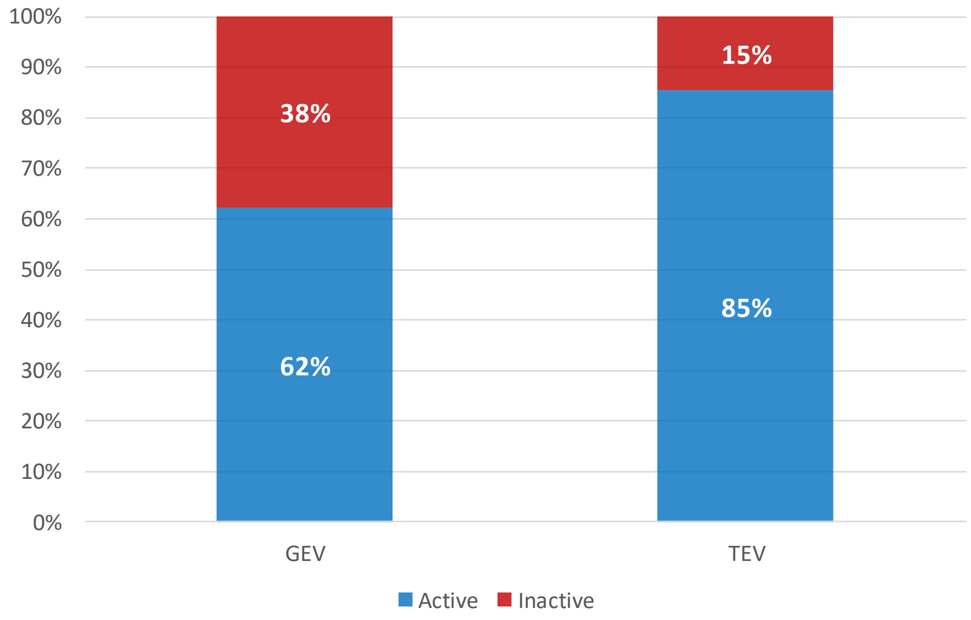


Figure 10: Proportion of CO GEVs and TEVs removed from the fleet during the three-year period between 2015 and 2018

Of the vehicles categorised as CO GEVs in 2015, 62% remained active in 2018 indicating a 38% retirement of GEVs from the fleet over a three-year period. In comparison, 85% of vehicles categorised as CO TEVs remained active in 2018.

For HC, NO and PM, 85% of all TEVs remained active in 2018. The proportion of GEVs remaining active was 68% for HC, 64% for NO and 70% for PM.

This analysis shows that GEVs are retired at a rate of about 2.5 times more quickly than TEVs over the period 2015-2018.

2.6 Distance Travelled by GEVs

This section compares the AADT for GEVs and TEVs in the monitored fleet. The AADT allows the calculation of total annual emissions (presented in Section 3.2).

The vehicle odometer data from 2015 and 2018 allowed for calculation of the annual distance travelled by each of the monitored vehicles. Null data[5] were omitted, along with data where the odometer had been flagged as unreliable, if the odometer units did not match between years, and if the 2018 odometer reading was lower than the 2015 reading. The difference between the two odometer readings was calculated and values recorded in miles were converted to kilometres. The time between the test dates on which the odometer readings were taken was calculated in years. The distance and number of years were then used to calculate the AADT in kilometres per year.

The difference between AADT for CO GEVs and CO TEVs is shown in Figure 11. This indicates that CO GEVs travelled approximately 2,000 km/yr or 14% less than CO TEVs during the 2015-2018 time period.

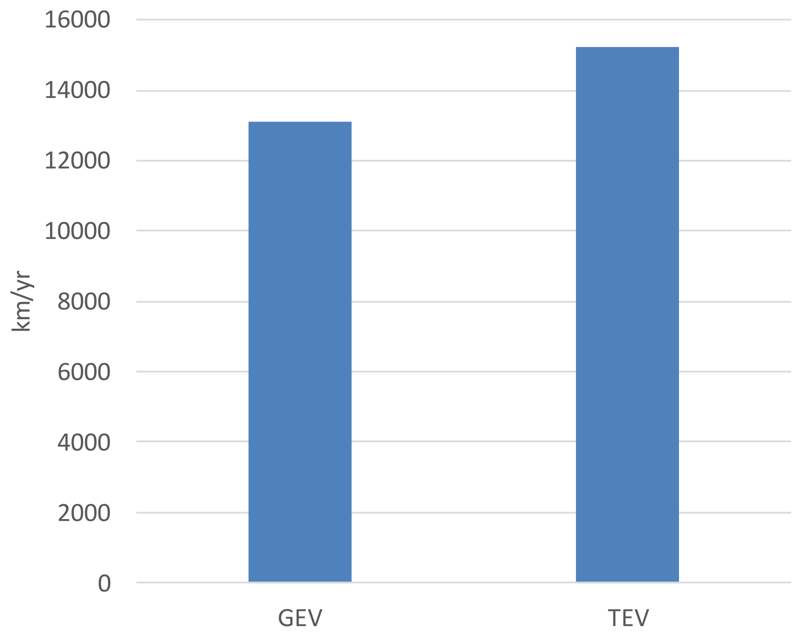


Figure 11: Annual average distance travelled (AADT) for CO GEVs vs TEVs

Assessment of HC GEVs showed that they travelled approximately 1,500 km/yr or 10% less than HC TEVs and NO GEVs travelled 730 km/yr or 5% less than NO TEVs. PM GEVs travelled further than PM TEVs by 750 km/yr (5%).

5 Including zero values, and values entered as the maximum reading of 999,999 km.

2.7 Identifying the Characteristics of GEVs

Analysis of the data was undertaken to identify which vehicle characteristics showed strongest correlation with vehicles categorised as GEVs. The analysis led to construction of a predictive GEV model, with definition of model parameters from measured vehicle characteristics. The predictive GEV model was used to estimate the number and location of GEVs in New Zealand's light duty vehicle fleet (Section 2.8).

Two methods were used to identify the characteristics of GEVs as detailed in sections 2.7.1 and 2.7.2.

2.7.1 Spreadsheet model using four factors for characterisation

Vehicle age, odometer reading, engine capacity and fuel type were chosen as parameters for characterisation of GEVs. GEVs from the monitored fleet were sorted in descending order for each of the selection parameters (except fuel type), and the cut-off value for the top 60% of values for each parameter was found for each pollutant. The 60% was chosen through iteration as it generally provided the best selectivity whilst minimising the number of TEVs incorrectly modelled as GEVs. The 60% cut-off was not used for the fuel type parameter due to the binary rather than continuous nature of the data. The value of each parameter at the 60% cut-off was used as the selection criteria for GEV characterisation for that pollutant.

A spreadsheet was set up to count vehicles from the entire monitored fleet and vehicles from the GEV fleet that had the GEV characteristics specified. Several performance indicators were calculated to allow interpretative strength of the modelling method. These included correctly modelled GEVs, correctly modelled TEVs, errors of commission, and errors of omission as defined in Section 2.7.4.

2.7.2 Regression Tree Analysis (RTA)

Regression tree analysis (RTA) was used to objectively identify the vehicle characteristics that were most important in affecting monitored emissions.

RTA can be used to look for relationships in complex data with very few assumptions about data distribution and interactions between data categories. The resulting tree models identify which predictor variables (vehicle characteristics) explain the most variation in the response variables (monitored vehicle emissions of CO, HC, NO and PM).

The models cluster the vehicles into groups or 'nodes' with similar emission values using rules based on the predictor variables. A more detailed description of the RTA approach used is contained in Appendix B: Description of Regression Tree Analysis GEV Modelling. The method used is based upon the full technical method documented by De'ath and Fabricius (2000).

GEV nodes were selected from RTA based on the highest median emission value of the nodes and the number of vehicles within the node. Whilst some nodes returned a low number of vehicles with very high emissions, it was considered that larger groups capturing a good proportion of GEVs should be included as well so that the numbers were more likely to be representative for a larger population size. Up to 5 nodes were selected for each pollutant, with an aim for the total modelled vehicles for each pollutant to be between 2% and 4% of the monitored fleet. The groups were characterised to compare their performance using various parameters as detailed in Section 2.7.4 below. The full selection criteria for each node identified as appropriate for GEV modelling using RTA are presented in Table 11 and Table 12 of Appendix C: Criteria Used for GEV Modelling.

2.7.3 Backfilling of Missing Power Values

During preliminary development of the RTA models, vehicle power was identified as having a strong correlation with GEVs. However, a large proportion of the vehicle records had null values associated with vehicle power in the MVR. Due to a requirement for removal of records with associated null values for the RTA modelling method, a method was developed to estimate vehicle power to allow back-filling of null values in the data set. This method is detailed in Appendix D: Backfilling of Missing Power Values. The retrospective addition of estimated power values increased the number of records in the final complete clean data set to 18,393.

There were multiple iterations of RTA for the 2015 RSD data set. The final method used the described method for backfilling of power values and entered power as a selection criterion. Power did not appear in any of the selection parameters for modelled groups. However, the addition of this as a possible selection criterion successfully improved the performance of the RTA modelling method for HC, NO and PM methods.

2.7.4 Comparison of GEV Modelling Methods

A number of performance indicators were calculated to allow interpretation of the strength of the two modelling methods. Comparison of these performance indicators for the modelling methods developed assisted in identifying the most effective modelling method for each pollutant.

Figure 12 provides a visual representation of the vehicle groups used for assessing model performance. The entire rectangle represents the monitored fleet, with the subsections representing:

A. Vehicles measured to be TEVs and modelled to be TEVs (i.e. correctly modelled TEVs);

B. Vehicles measured to be GEVs and modelled to be GEVs (i.e. correctly modelled GEVs);

C. Vehicles measured to be TEVs and modelled to be GEVs (i.e. 'errors of commission'); and

D. Vehicles measured to be GEVs and modelled to be TEVs (i.e. 'errors of omission').

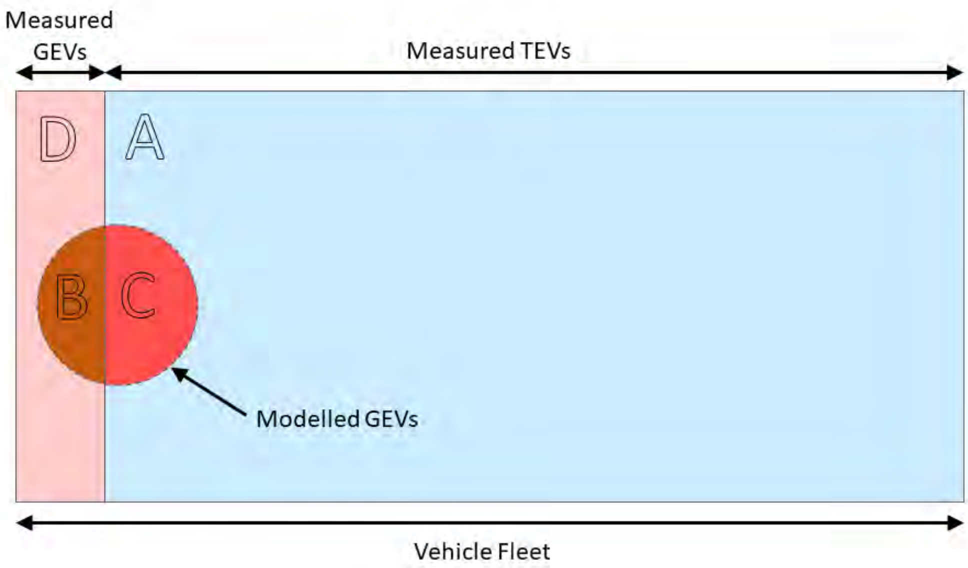


Figure 12: Visual representation of modelled GEVs and TEVs

The performance indicators used were proportional relationships of these groups to one another as follows:

- monitored vehicle fleet modelled as GEVs ($\frac{\text{Modelled GEVs}}{\text{Vehicle Fleet}}$);
- modelled GEVs correctly modelled ($\frac{b}{\text{Modelled GEVs}}$);
- measured GEVs correctly modelled ($\frac{b}{\text{Measured GEVs}}$); and
- measured TEVs incorrectly modelled as GEVs ($\frac{c}{\text{Measured TEVs}}$).

These parameters are shown for each of the modelling methods In Table 6 (spreadsheet model) and Table 7 (RTA) below.

It is important to note for this assessment that the top 3% of measured emissions values were categorised as GEVs, i.e. $\frac{\text{Measured GEVs}}{\text{Vehicle Fleet}}$ was always equal to 3%.

Table 6: Results of GEV modelling using spreadsheet model (by pollutant)

Number of vehicles	Pollutant			
	CO	HC	NO	PM
Measured GEVs	762	779	741	741
Measured TEVs	22138	22121	22159	22159
Modelled GEVs (b + c)	170	849	735	90
Correctly modelled GEVs (b)	35	108	87	36
Errors of Commission (c)	135	741	648	54
Proportion of fleet modelled as GEVs ($\frac{\text{Modelled GEVs}}{\text{Vehicle Fleet}}$)	0.7 %	3.7 %	3.2 %	0.4 %
Modelled GEVs correctly modelled ($\frac{b}{\text{Modelled GEVs}}$)	21 %	13 %	12 %	40 %
Measured GEVs correctly modelled ($\frac{b}{\text{Measured GEVs}}$)	5 %	14 %	12 %	5 %
Measured TEVs incorrectly modelled as GEVs ($\frac{c}{\text{Measured TEVs}}$)	0.6 %	3.3 %	2.9 %	0.2 %

Table 7: Results of GEV modelling using RTA (by pollutant)

Number of Vehicles	Pollutant			
	CO (5 nodes)	HC (4 nodes)	NO (4 nodes)	PM (2 nodes)
Measured GEVs	336	330	261	316
Measured TEVs	18057	18063	18132	18077
Modelled GEVs (b + c)	726	607	578	420
Correctly modelled GEVs (b)	104	51	79	112
Errors of Commission (c)	622	556	499	308
Proportion of fleet modelled as GEVs ($\frac{\text{Modelled GEVs}}{\text{Vehicle Fleet}}$)	3.9 %	3.3 %	3.1 %	2.3 %
Modelled GEVs correctly modelled ($\frac{b}{\text{Modelled GEVs}}$)	14 %	8 %	14 %	27 %
Measured GEVs correctly modelled ($\frac{b}{\text{Measured GEVs}}$)	31 %	15 %	30 %	35 %
Measured TEVs incorrectly modelled as GEVs ($\frac{c}{\text{Measured TEVs}}$)	3.4 %	3.1 %	2.8 %	1.7 %

The proportion of vehicles measured as GEVs that were correctly modelled as GEVs was between 5% and 14% for the spreadsheet model and between 15% and 35% for the RTA model. The associated proportion of vehicles measured as TEVs that were errors of commission (incorrectly modelled as GEVs) was between 0.2% and 3.3% for the spreadsheet model and between 1.7% and 3.4% for the RTA model. For comparison, previous assessment of RSD-measured GEVs using multivariate regression trees completed by Bluett et al. (2010) correctly modelled 40-80% of GEVs but had much higher errors of commission of 3-20%.

The proportion of measured GEVs correctly modelled is consistently higher using the RTA modelling. For CO and PM spreadsheet modelling, the proportion of the population in the modelled group was significantly lower than that of the HC and NO spreadsheet models and all of the RTA models. Although proportions of modelled GEVs correctly modelled for the CO and PM spreadsheet models seem high compared to those achieved by RTA modelling, this is due to the small size of the spreadsheet-modelled GEV group and the associated proportion of measured GEVs correctly modelled is low.

There is a less prominent difference between modelling methods for HC and NO GEVs but the proportions of measured GEVs correctly modelled are higher for RTA modelling while the proportions of the fleet modelled as GEVs and thus the proportions of measured TEVs incorrectly modelled as GEVs are lower.

This indicates that the RTA modelling method is stronger for all four of the pollutants.

2.8 Analysis of the National Fleet

A data set for the entire national fleet, as recorded for December 2019, was provided by the Corporate Support (Data Services) department of Waka Kotahi. This consisted of the data available for the MVR from the Waka Kotahi Open Data Portal with the addition of most recent odometer reading, the associated date, units and reliability and emissions test regime (emission standard) details. The data provided was restricted to the 'Passenger Car/Van' and 'Goods Van/Truck/Utility' vehicle type categories. Following a further request, additional values were added for the earliest odometer reading recorded within the previous 5 years and the date it was recorded.

The data was analysed in R and records were filtered using the following criteria:

- GVM above 500 kg and less than 3,500 kg; and
- Vehicle type recorded as 'Passenger Car/Van' or 'Goods Van/Truck/Utility' (as a check).

Data included information on the region and district (Territorial Land Authority, or TLA) in which each vehicle was registered. These were tidied to correct entries with multiple spelling options and to include records from previous TLAs that had since merged into the single 'Auckland Council' unitary authority.

Where an odometer was flagged as unreliable, or had null data[6], both odometer readings were set as null values. Where the odometer distance unit was recorded as miles, the values of the odometer readings were converted to kilometres and the units were updated accordingly. Unknown units were assumed to be kilometres. The difference in time between the first and last odometer reading provided was calculated in years. If the time calculated was less than 0.01 years or the difference in distance over the time period was less

[6] Including zero values, and values entered as the maximum reading of 999,999 km.

than 1 km, both odometer readings were set as null values. Remaining values were used to calculate the annual vehicle kilometres travelled (VKT) of vehicles in the fleet.

Total vehicle counts for regions and TLAs were completed, as well as counts of null records by region and TLA. The data set was complete with the exception of null odometer readings and associated annual VKT values.

The national fleet data was filtered to create a group with the specified GEV criteria and a second group of the remaining fleet (assumed to be TEV). The number of vehicles in each group was reported by region and by TLA. The AADT was found by calculating the average (mean) annual VKT excluding null values for both groups as a whole and by regions, and the null values in each of the groups was recorded as a total and by regions.

Outputs from the R analysis consisted of counts of national fleet vehicles by region and TLA, AADT of modelled GEVs and TEVs as a whole and by region, and null record counts by modelled GEVs and TEVs for each group. Average annual VKT values were combined with regional counts and null records to allow calculation of a combined AADT value for each pollutant as shown in Appendix E: Calculation of AADT (National Fleet).

Vehicle counts were grouped so that multiple nodes modelled for a single pollutant were combined and the GEV counts modelled by RTA were linked to GIS shapefiles to produce heat maps showing the variation in proportion of vehicles modelled for each pollutant by region and by TLA.

2.8.1 National Fleet

After cleaning of the December 2019 MVR national fleet data set to remove vehicles outside the weight range of 500 – 3,500 kg and vehicle types other than ‘Passenger Car/Van’ or ‘Goods Van/Truck/Utility’, 3.61 million records remained. If GEVs were classified as the top 3% of high-emitting vehicles, this would equate to over 108,000 GEVs on a national level.

The selection criteria identified through RTA (provided in full in Table 11 and Table 12 of Appendix C: Criteria Used for GEV Modelling) were used to identify the number of vehicles modelled as GEVs for each of the four pollutants. The number and proportion of the national fleet modelled as GEVs for each pollutant is shown in Table 8 alongside the proportion of the 2015 RSD data set that was modelled using the same selection criteria.

Table 8: Proportion of National Fleet Modelled as GEVs

Pollutant	Modelled GEVs		
	National Count	% of National Fleet[1]	% of 2015 RSD Data Set
CO	84,416	2.3 %	3.9 %
HC	61,730	1.7 %	3.3 %
NO	76,538	2.1 %	3.1 %
PM	42,144	1.2 %	2.3 %

Notes:
1. *Total population size of 3,610,504.*

The proportion of the national fleet modelled using the RTA criteria is consistently lower than the proportion modelled within the 2015 RSD data set. Despite these differences, it is considered that modelling of GEVs within the national fleet using the RTA selection criteria provides an objective and robust baseline for identifying potential GEVs.

2.8.2 Regional Distribution of GEVs

RTA model criteria were also used to identify the number of vehicles modelled as GEVs for each of the four pollutants on a regional and district level. Regional and district data has been presented on heat maps as shown in Figure 13 and Figure 14 for vehicles modelled as CO GEVs. Full regional and district heat maps for modelled GEVs for each pollutant are contained in Appendix F: Heat Maps Showing Proportion of Modelled GEVs at Regional and District Level

Figure 13 shows the proportion of vehicles modelled as CO GEVs using RTA-derived criteria by region for New Zealand. The higher proportions of CO GEVs in Gisborne, Northland and the West Coast followed by middle proportions in the central North Island generally align with lower gross domestic product (GDP) per capita values, indicating a less affluent population (Stats NZ, 2018). Modelling criteria for CO GEVs include older vehicles and high odometer readings which are vehicle properties likely to be more prevalent in communities with lower income levels.

For all pollutants assessed, Auckland and Wellington regions have consistently lower proportions of modelled GEVs compared to the rest of the country. Northland Regional Council consistently has one of the highest proportions of modelled GEVs and West Coast Regional Council has a high proportion of modelled GEVs for all pollutants except NO.

Although the proportions of GEVs in urban centres are often lower than those seen in the rest of the country, the corresponding number of vehicles is higher due to the density of vehicles in the urban centres. For example, although Auckland Region has one of the lowest regional proportions of modelled CO GEVs in the country at 2.1%, it also has the highest number of modelled GEVs at 24,729. Conversely, the highest proportion of modelled CO GEVs is seen in Gisborne Region at 3%, but this is representative of under 1,000 vehicles.

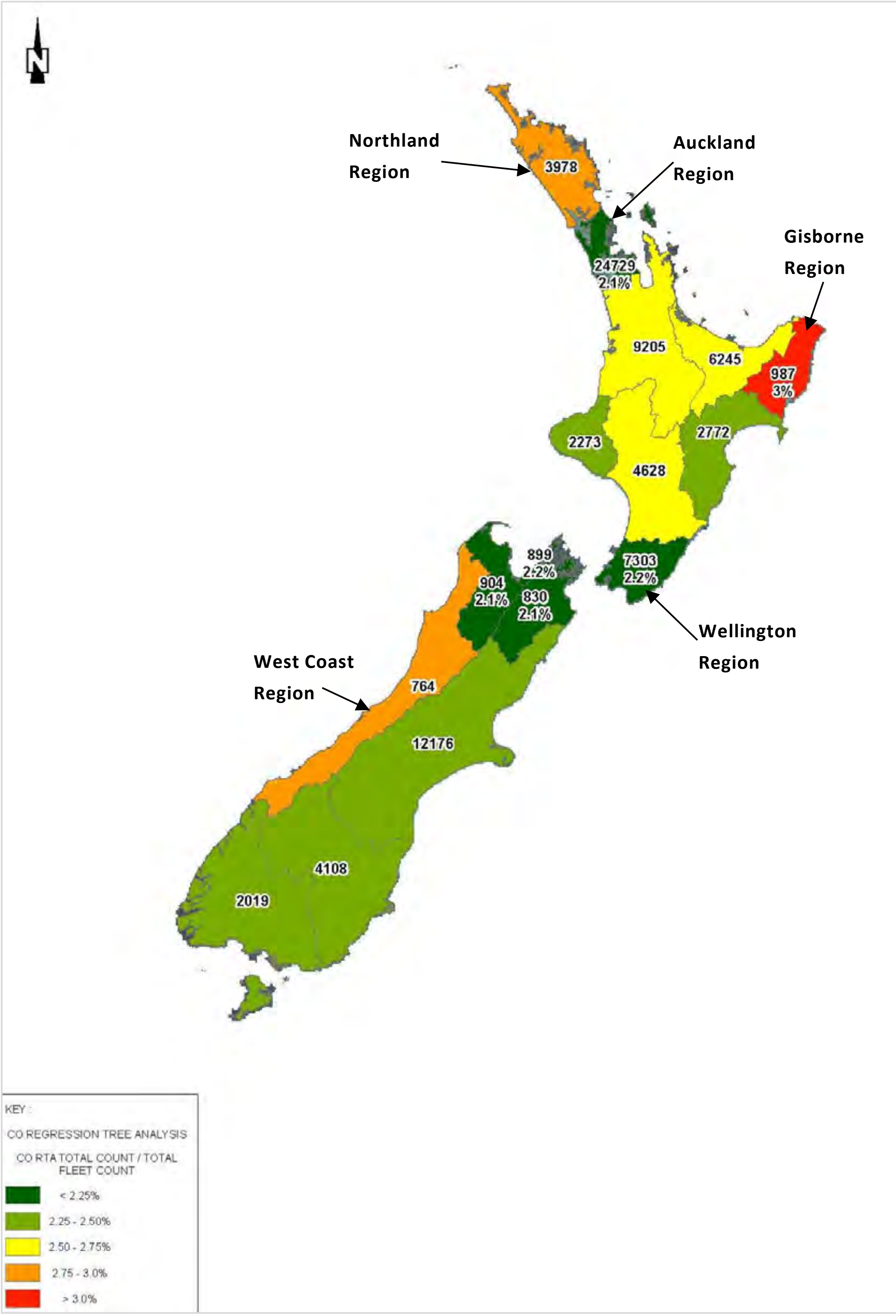


Figure 13: Proportion of national fleet modelled as CO GEVs by region

Figure 14 shows the RTA-derived proportion of vehicles modelled as CO GEVs by TLA for New Zealand. This shows that there is significant variation of modelled GEV proportions between districts.

The breakdown into TLAs shows that there is variation in the spread of modelled GEVs between districts within all regions for all pollutants. There is also a distinctive trend with vehicles modelled as PM GEVs in that main cities (particularly Auckland, Hamilton City, Wellington City, Christchurch City, Dunedin City, Palmerston North City and Nelson City) have distinctly lower proportions of modelled GEVs than the surrounding TLAs. This may be due to diesel-fuelled vehicles being a selection criterion for modelling PM GEVs and a higher prevalence of diesel-fuelled work and utility vehicles in farming and rural areas.

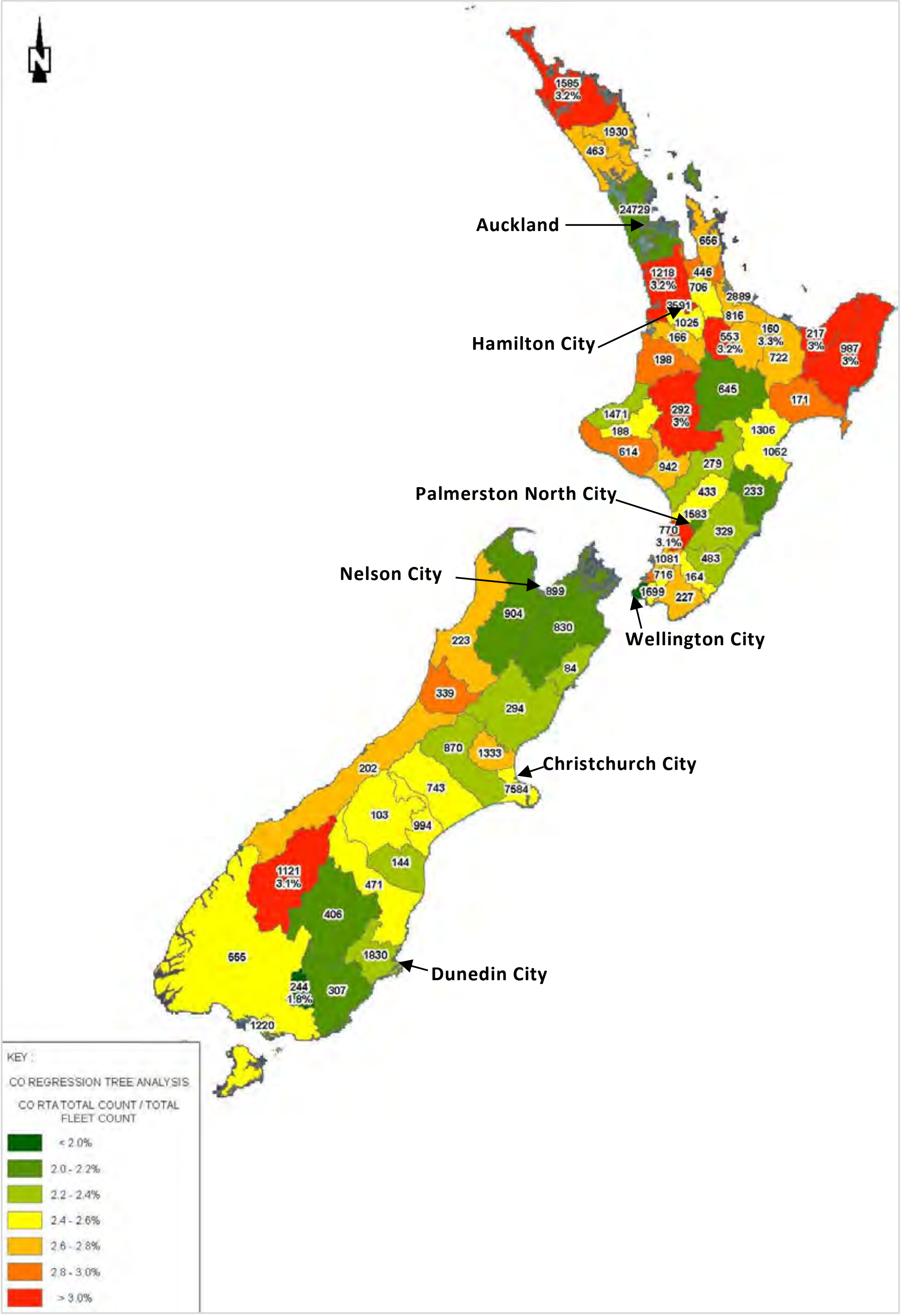


Figure 14: Proportion of national fleet modelled as CO GEVs by territorial land authority (TLA)

2.8.3 Distance Travelled by Modelled GEVs

The AADT was calculated for vehicles modelled as GEVs from the national fleet and compared to that of vehicles modelled as TEVs. This allowed comparison of the trends in AADT observed in the monitored fleet GEVs and TEVs (presented in Section 2.6) against the modelled GEVs and TEVs in the national fleet.

Figure 15 shows a regional comparison for AADT of vehicles modelled as CO GEVs against those modelled as TEVs. With the exceptions of the Chatham Islands, vehicles modelled as CO GEVs have a consistently higher AADT than those modelled as TEVs. This finding is inconsistent with the findings from the 2015 – 2018 RSD data set in which AADT of CO GEVs was lower than that of TEVs.

This trend of GEVs recording higher AADT values than TEVs was consistent with the other pollutants. When compared to results from the 2015 – 2018 RSD data set, this aligns with the findings for PM GEVs but is inconsistent with the findings for HC and NO GEVs.

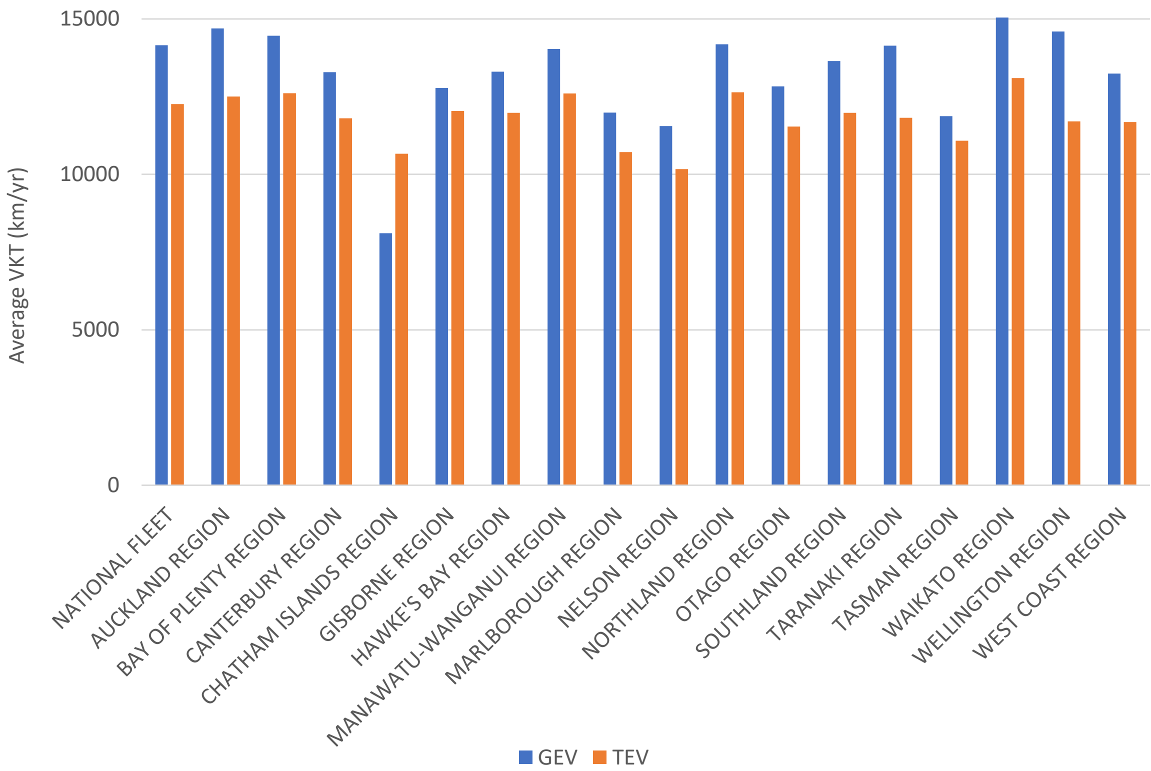


Figure 15: Annual average distance travelled (AADT) of national fleet vehicles by region and modelled as CO GEVs and TEVs

3.0 Stage 2

3.1 RSD Measurements of GEVs Compared to the Wider TEV Fleet

This section assesses how much more CO is emitted from GEVs than TEVs. Figure 16 shows a box and whisker plot comparing the emissions values of CO (reported as % CO) from CO GEVs and CO TEVs in the 2015 RSD data set. The ends of the box mark the 25th and 75th percentile values, the line in the middle of the box marks the median value, and the cross marks the mean value. The limits of the whiskers mark the most extreme values in the data set not classified as outliers, with outliers defined as values outside the range between the upper quartile plus one and a half times the interquartile range and the lower quartile minus one and a half times the interquartile range.

Figure 16 indicates that there is a statistically significant difference between emissions of CO from vehicles classified as GEVs and TEVs. The mean values are both higher than the median values. This indicates there are more vehicles with low emissions records than very high emissions records and that the frequency distributions for each of the data subsets are skewed to the right. Equivalent figures for HC, NO and PM are contained in Appendix H: Figures for Other Pollutants of Interest.

The HC, NO and PM emissions readings all show a statistically significant difference between GEV and TEV readings and all of the groups also display the same right-skewed behaviour with mean values higher than the respective median and a higher number of low emissions records than very high emissions records.

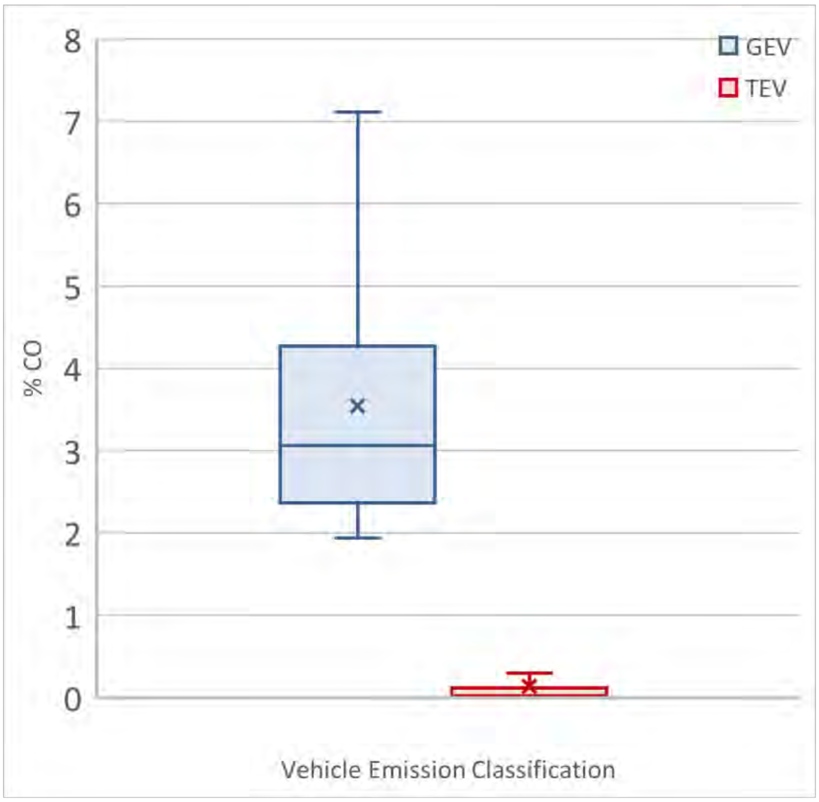


Figure 16: CO emissions from GEVs and TEVs (2015 RSD Data Set)

3.2 Potential Emission Reduction Benefits of GEV Replacement

This section compares the estimated annual emissions from GEVs and TEVs in the monitored fleet and provides an estimate for the potential emission reduction benefit of replacing a single GEV from the fleet with a TEV.

The emissions measurement was used to calculate an emission factor, the grams of pollutant released per kilogram of fuel burned for each vehicle, as described in Appendix G: Conversion of Emissions Data to Emission Factors. This was then converted into grams of pollutant released per litre of fuel by using a density of 750 g/L for petrol and 830 g/L for diesel.

Fuel efficiency data was taken from the Ministry of Transport fuel consumption data contained in the March 2019 version of the Vehicle Fleet Emissions Model (VFEM) and this was integrated into the existing data set. The fuel efficiency value assigned to each record was determined by a combination of vehicle type, year of manufacture, engine capacity, and fuel type. Fuel efficiency was given in litres per 100 kilometres so allowed conversion of the RSD-measured vehicle emissions figures into grams of pollutant per kilometre. The method used was based upon the Pokharel et al. (2002) method. At this point an estimated NOx figure was also introduced by scaling the NO emission in g/km by 1.11 for petrol-powered vehicles and by 1.25 for diesel vehicles. This was based upon assumed NO/NOx ratios of 90% for petrol and 80% for diesel (and that components of NOx that are not NO or NO_2 are negligible). These ratios are derived from ranges of 0.9 - 0.95 for the NO:NOx ratio of petrol vehicles and 0.75 – 0.85 for diesel vehicles (DEFRA, 2003, as cited in Bluett et al. (2010)).

Combining the AADT (km/yr) with the emissions rates in g/km allowed estimation of the annual emissions in grams of pollutant per year (g/yr). The annual emissions figures were used to deduce the potential emission reduction benefit of removing GEVs from the fleet and replacing them with TEVs.

Figure 17 shows the estimated annual CO emissions from CO GEVs and TEVs as recorded in the 2015 RSD data set. Equivalent figures for HC, NO and PM are contained in Appendix H: Figures for Other Pollutants of Interest.

The difference between the median annual emissions for GEVs and TEVs is 347 kilograms of CO per year (kg CO/yr) per vehicle. This provides an estimated annual emissions reduction benefit for replacement of a single CO GEV in the fleet.

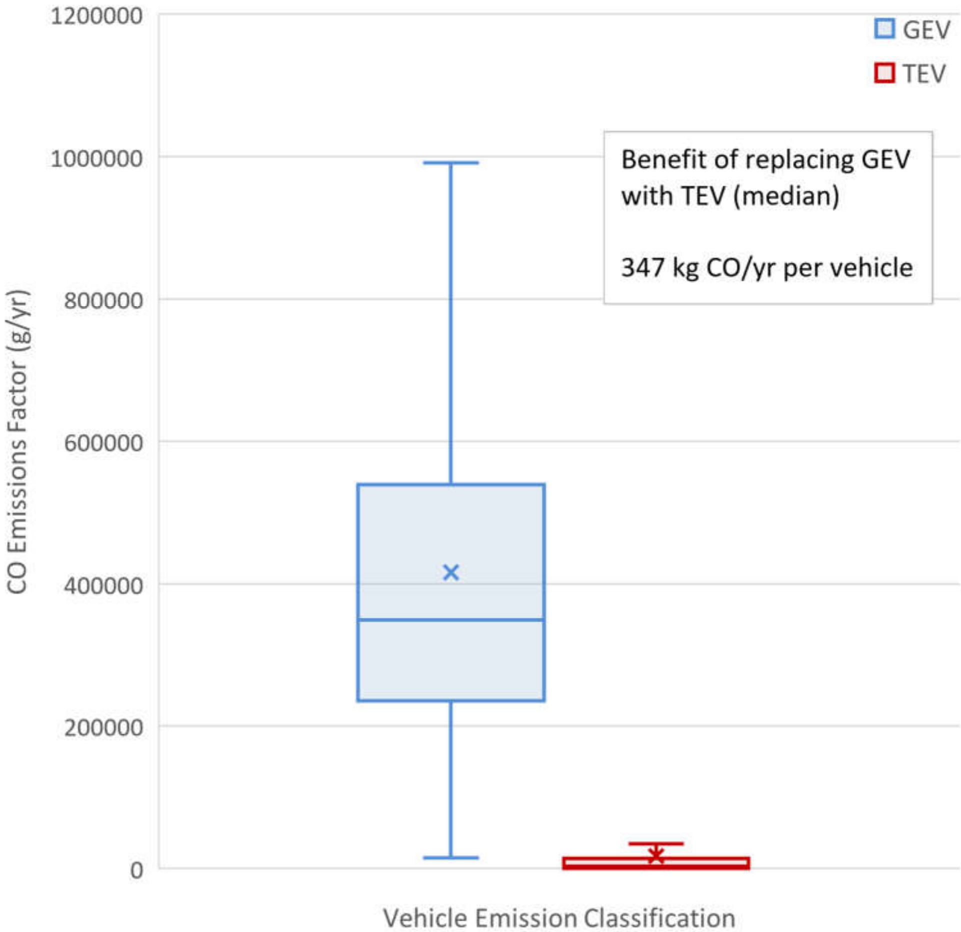


Figure 17: Estimated annual emissions of CO GEVs and TEVs

If a national fleet of 3.61 million LDVs is assumed (as described in Section 2.8.1), 3% of the population equates to over 108,000 GEVs nationally. Although identification and immediate replacement of all of these vehicles is very unlikely due to the difficulty in accurate identification of all of the vehicles and the high associated replacement cost, based upon the estimated annual emissions reduction benefit above they represent a potential annual reduction benefit of over 37,500 tonnes of CO.

The potential annual emissions reduction benefits for replacing a single vehicle and all GEVs nationally for each of the pollutants considered including CO (CO, HC, NO, NO_2, and PM) are shown in Table 9 below.

Table 9: Potential Annual Emissions Reduction Benefits

Pollutant	Annual Emissions Reduction Benefit	
	Per Vehicle	All GEVs in National Fleet[1]
CO	347 kg CO/yr	37,500 tonnes CO/yr
HC	22.7 kg HC/yr	2,500 tonnes HC/yr
NO	42.6 kg NO/yr	4,600 tonnes NO/yr
NO_2	4.8 kg NO_2/yr	520 tonnes NO_2/yr
PM	485 g PM/yr	53 tonnes PM/yr

Notes:
1. *Based upon national fleet of 3.61 million and 3% GEVs totalling 108,300 vehicles*

In summary, this analysis shows that placing a single GEV from the fleet with a TEV provides a significant emission reduction benefit and that this benefit would be amplified if all GEVs were removed from NZ's vehicle fleet. The analysis of the regional distribution of GEVs suggests that there may be an argument to support a geographically targeted approach to regulating GEVs.

4.0 Achievement of Project Objective and Completion of Tasks

The objective of this project was to determine the impact of GEVs on emissions of pollutants harmful to human health from the vehicle fleet across New Zealand. The following issues were identified for investigation:

- Prevalence of GEVs:
  - What proportion of the fleet are potentially GEVs?
  - How long do GEVs remain in the fleet?
  - What distance do GEVs typically travel in a year?
  - What is the regional distribution (geographical spread) of GEVs?
- Determination of replicability of GEV results;
- Determination of the potential impact of GEVs compared to emissions from the wider light duty vehicle (LDV) fleet; and
- Estimation of the emission reduction benefit of removing GEVs from the fleet.

The project was undertaken in two stages. A summary of findings from these is provided in the following two sections. The details presented in this report and the summary of findings from the study demonstrates that the objective of this project has been met and the investigation of targeted issues successfully completed.

4.1 Stage 1 – Assessment of Number, Characteristics, Regional Distribution, and Activity of GEVs

GEVs were defined as vehicles with the highest 3% of emissions readings from the 2015 RSD database for each pollutant. Defining GEVs as the highest 3% of emitters captures a significant proportion (>25%) of the total emissions while including a relatively small number of vehicles which could potentially be targeted through policy interventions.

The crossover of GEVs for the four pollutants considered was investigated and most GEVs were found to be GEVs for a single pollutant of the four studied. The highest crossover between pollutants was that between CO GEVs and HC GEVs.

Assessment of GEVs with more than one associated emissions record showed that second measurements for the same vehicle were higher than the majority of TEV measurements in the fleet indicating that GEVs tend to produce consistently higher emissions measurements. While the second highest measurement from GEVs are frequently below the GEV cut-off value, they are still higher as a group than TEV emissions as shown by the lack of overlap between interquartile ranges of the two sample groups. Given this result, PDP consider the use of RSD measurements to identify GEVs is a robust and pragmatic method.

Assessment of the 2015 RSD data set showed a higher proportion of petrol-fuelled vehicles in CO GEV and NO GEV subsets than in the overall monitored fleet. The PM GEV subset had a higher proportion of diesel-fuelled vehicles than the monitored fleet, and the HC GEV subset had a similar proportion of petrol- and diesel-fuelled vehicles to that of the monitored fleet. Vehicles categorised as older than 14 – 16 years, with odometer readings above 150,000 – 200,000 km, or having been tested to older emission standards were found to have higher proportions of GEVs than the overall monitored fleet.

Comparison between vehicles classified as GEVs and TEVs showed that GEVs had a higher retirement rate than TEVs, with removal or replacement of vehicles between 2015 and 2018 measured as 30-38% for GEVs and 15% for TEVs in the monitored fleet. GEVs for CO, HC and NO had slightly lower AADT value than their respective TEVs, whereas PM GEVs travelled slightly further than TEVs as an annual average.

Correlation of vehicle characteristics and GEVs was used to develop predictive GEV models using spreadsheet-based and RTA approaches. The predictive GEV model was used to estimate the number and location of GEVs in New Zealand's light duty vehicle fleet. Comparison of model performance on the 2015 monitored data set showed that RTA provided a stronger model than the spreadsheet-based approach.

Data from the December 2019 MVR was sourced and assessed using the RTA-derived criteria to predict the regional distribution and to calculate the AADT of vehicles modelled as potential GEVs for each pollutant.

Major cities were generally found to have lower proportions of modelled GEVs whereas Northland and West Coast regions had generally higher proportions of vehicles modelled as GEVs. For all pollutants, modelled GEVs were shown to have higher AADT than corresponding TEVs.

Although the proportions of GEVs in urban centres were generally found to be lower than those seen in the rest of the country, the corresponding number of vehicles was higher due to the density of vehicles in the urban centres.

4.2 Stage 2 –Relative Impact of GEVs, Estimation of Potential Emission Reduction Benefits, and Repeatability of Emissions Measurements

A comparison between emissions values for GEVs and TEVs was completed and found that vehicles classified as GEVs had significantly higher emissions than vehicles classified as TEVs for all pollutants. Each GEV and TEV subset had a higher mean value than the median. This indicated that there were more vehicles with low emissions records than very high emissions records and that the frequency distribution for each subset was skewed to the right.

Finally, the emissions of each pollutant per vehicle per year was calculated and the difference between median GEV and TEV values was used to estimate the annual emission reduction benefit of replacing a GEV in the fleet with a TEV. Estimated annual emission reduction benefit of replacing a single GEV in the fleet were 347 kg CO/yr for CO, 22.7 kg HC/yr for HC, 42.6 kg NO/yr and 4.8 kg NO_2/yr for NO and 485 g PM/yr for PM. This was extended to quantify the potential national annual pollutant reductions available if it were possible to replace all GEVs in the fleet with TEVs.

This analysis shows that placing a single GEV from the fleet with a TEV provides a significant emission reduction benefit and that this benefit would be amplified if all GEVs were removed from NZs vehicle fleet. The analysis of the regional distribution of GEVs suggests that there may be an argument to support geographically targeted approach to regulating GEVs.

5.0 Recommendations for Future Work

During this project, PDP and the external peer reviewer identified a number of questions which, if investigated, would provide additional insight into the impacts and management of GEVs and therefore add value to the outcomes of this study. Using this experience, PDP recommend that future work programmes consider the following tasks:

- Analyse the impact on emissions of vehicle age, odometer reading and emission standards separately for petrol and diesel vehicles. This could confirm and refine some of the findings of the current project;
- Develop a method for characterisation of GEVs/vehicle emissions that addresses the potential for non-independence of predictor variables. For example, if using vehicle age and emission standards as separate predictor variables, allowance should be made for the relationship between older emission standards and higher vehicle ages. A similar relationship is likely between high vehicle ages and high odometer readings;
- Conduct a more detailed investigation to confirm the finding that GEVs travel further each year than TEVs;
- Consider the impact of vehicle specific power (VSP) as measured by the RSD on GEVs;
- Assess the relative health benefits of GEV removal programmes which would target GEVs of specific pollutants, i.e. CO vs HC vs NO vs PM;
- Undertake a cost benefit analysis of replacing GEVs with TEVs;
- Review the Vehicle Emission Prediction Model to check that the model appropriately considers the impacts of GEVs on fleet average emissions;
- Investigate the potential benefits of a geographically targeted approach to regulating GEVs which considers, for example, the impact of districts with a relatively high proportion but a low number of GEVs;
- Engage policy analysts to assess if managing 3% of the LDV fleet as GEVs is practical and enforceable;
- Review vehicle inspection and maintenance (I/M) programmes (e.g. Warrant of Fitness) in NZ to identify if these could be potentially used to confirm and enhance the findings of this project;
- Consider the potential implications of the findings of this project on the heavy-duty vehicle fleet;
- Investigate any trends in GEVs by repeating and comparing the GEVs analyses undertaken in this report on an earlier RSD data set, e.g. 2009 vs 2015;
- Compare the findings of this study with international investigations on GEVs; and
- Consider if the RSD database may potentially be useful in assessing the impacts of GEVs on emissions of greenhouse gases from LDVs.

6.0 References

Bluett, J., Dey, K., & Kuschel, G. (2010). *Auckland vehicle fleet emission measurements 2005: The big clean up "Tune Your Car" campaign*. Prepared for Auckland Regional Council, NIWA Client Report CHC2010-092.

Bluett, J., Smit, R., & Aguiar, M. (2016). *Understanding trends in roadside air quality.* Golder Associates (New Zealand) Limited & The University of Queensland, NZ Transport Agency research report 596.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification algorithms and regression trees.* The Wadsworth statistics/probability series, 358.

De'ath, G., and Fabricius, K. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology 81*: 3178 – 3192.

Golder Associates (New Zealand) Limited. (2018). *Assessing the Impact of Gross Emitting Vehicles.*

Pokharel, S.S., Bishop G.A. and Stedman D.H. (2002). An on-road motor vehicle emissions inventory for Denver: an efficient alternative to modelling. *Atmospheric Environment 36,* 5177-5184.

Stats NZ. (2018) *Regional gross domestic product: Year ended March 2018.* Retrieved from https://www.stats.govt.nz/assets/Uploads/Regional-gross-domestic-product/Regional-gross-domestic-product-Year-ended-March-2018/Download-data/regional-gross-domestic-product-year-ended-march-2018.xlsx

United States Environmental Protection Agency (USEPA). (2004). *Guidance on Use of Remote Sensing for Evaluation of I/M Program Performance.* Office of Transportation and Air Quality, EPA420-B-04-010.

7.0 Glossary

Term	Definition
AADT	Annual average distance travelled
CO	Carbon monoxide
GEV	Gross emitting vehicle
GVM	Gross vehicle mass (kg)
HC	Hydrocarbons
LDV	Light duty vehicle (vehicle with a GVM under 3,500 kg)
MVR	Motor vehicle register
NO	Nitrogen monoxide
NO_2	Nitrogen dioxide
NO_x	Oxides of nitrogen
PM	Particulate matter
RSD	Remote sensing device
RTA	Regression tree analysis
TEV	Typical emitting vehicle
TLA	Territorial land authority
USEPA	United States Environmental Protection Agency
VFEM	Vehicle Fleet Emissions Model
VKT	Vehicle kilometres travelled
VSP	Vehicle specific power

Appendix A

Email Excerpts Outlining Reasoning for 3% GEV Cut-off Criteria

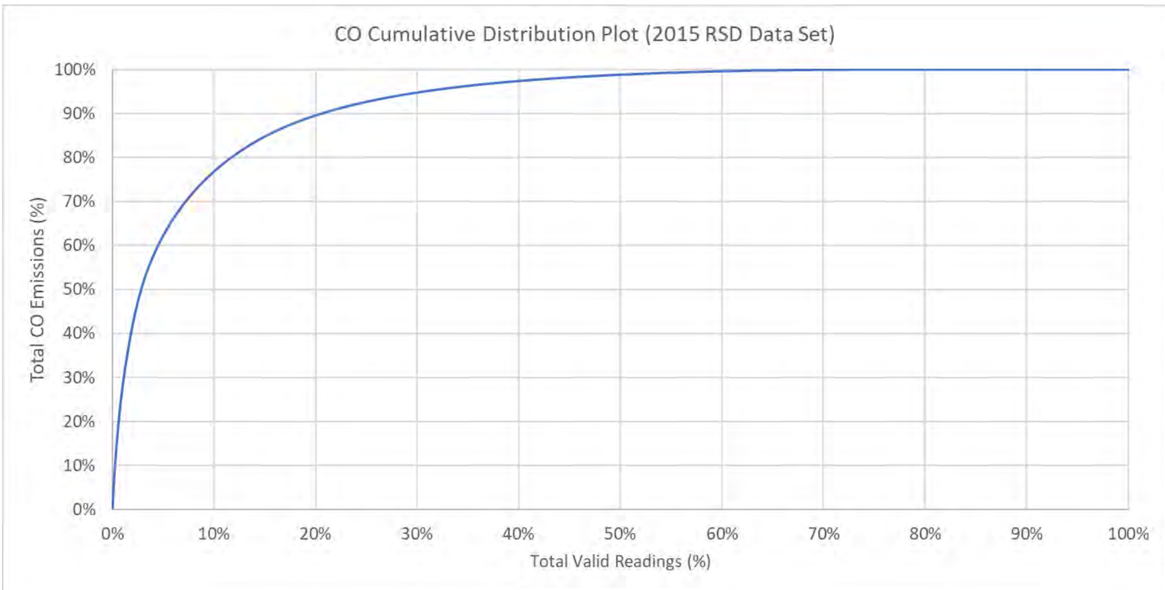
Gross Emitters - Cumulative Distribution Curves - Defining Gross emitters.

Please find attached a document which summarises the outcomes of our work on the cumulative distribution curves which we will be using to define the pollutant emission criteria for GEVs.

Having given consideration to these results we recommend the 97th percentile value (top 3 % of valid readings) be selected as the cut off point for defining GEVs. The tables included in the results summary show that below the 97th percentile value, the rate at which the % total emissions accumulate begins to slow. We believe the 97th percentile captures a very healthy proportion of the total fleet emissions while providing a pragmatic % of the fleet which policy could be targeted at.

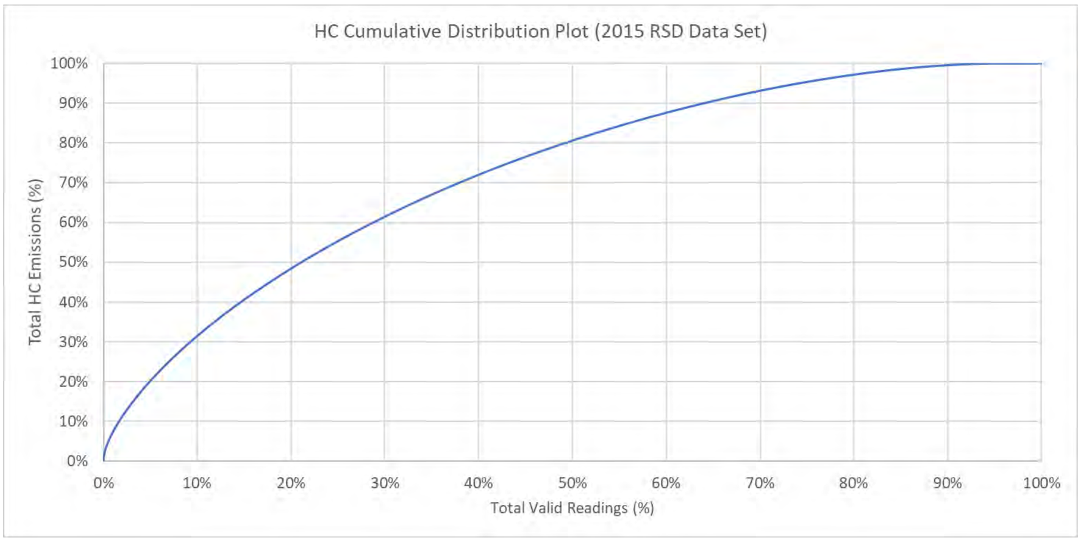
Please have a look at the attached result and let me know what you think. Once we have agreed on the GEV cut-off point we will begin to profile the GEVs.

CO Emissions



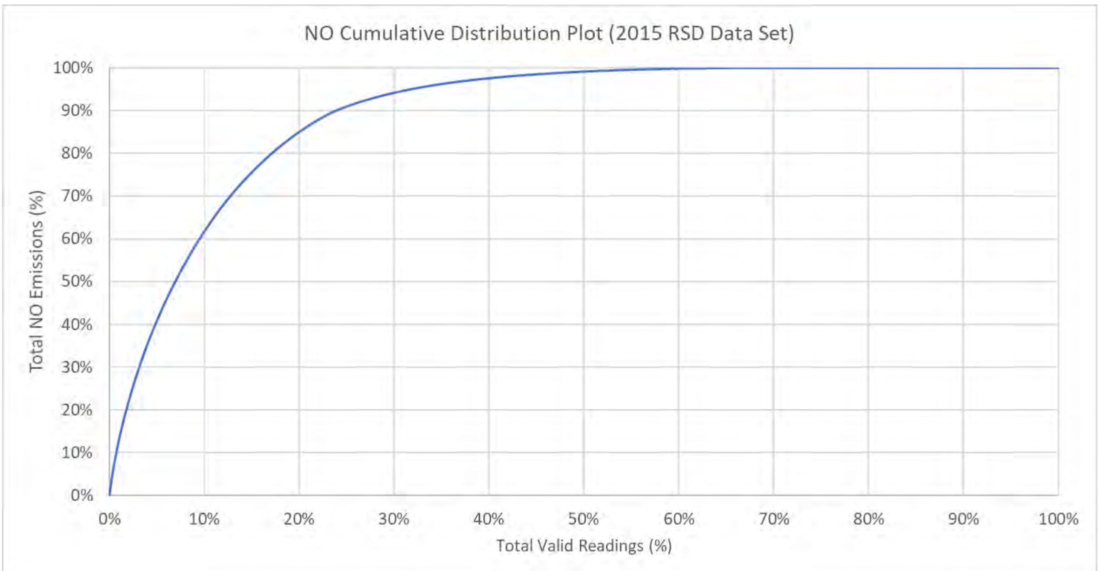
% emissions at n% total valid readings		
Valid Readings	**Total Emissions**	**Rate of change**
1%	28%	28%
2%	42%	14%
3%	51%	9%
4%	58%	6%
5%	62%	5%

HC Emissions



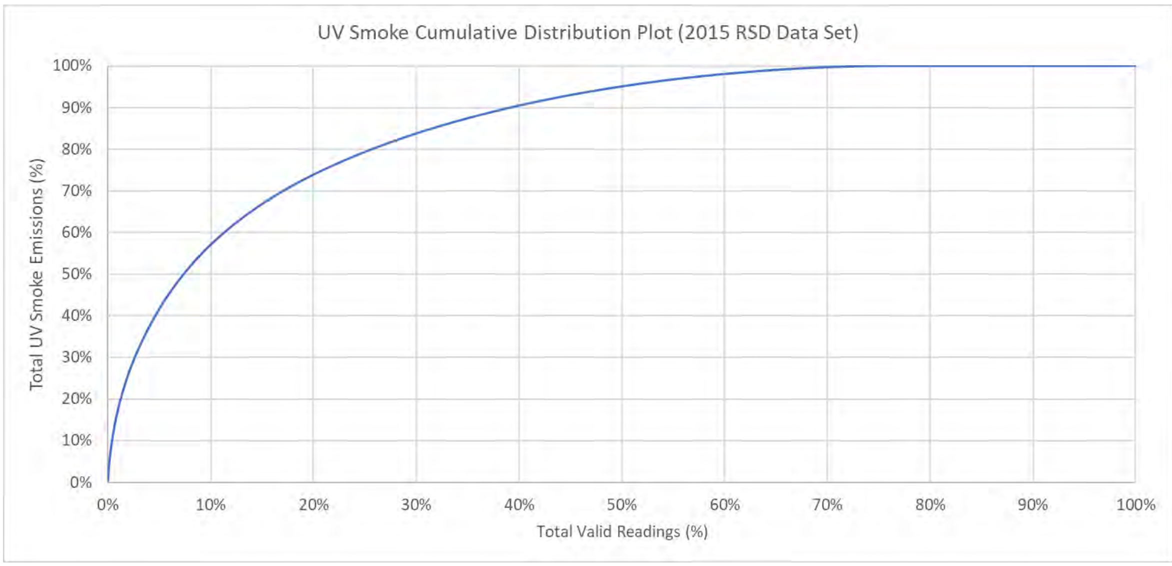
% emissions at n% total valid readings		
Valid Readings	Total Emissions	Rate of change
1%	7%	7%
2%	11%	4%
3%	15%	3%
4%	18%	3%
5%	20%	3%

NO Emissions



% emissions at n% total valid readings		
Valid Readings	Total Emissions	Rate of change
1%	13%	13%
2%	22%	9%
3%	29%	7%
4%	35%	6%
5%	41%	6%

uvSmoke Emissions



% emissions at n% total valid readings		
Valid Readings	Emissions	Rate of change
1%	18%	18%
2%	26%	8%
3%	32%	6%
4%	37%	5%
5%	42%	4%

Appendix B

Description of Regression Tree Analysis
GEV Modelling

Regressions tree analysis can be used to look for relationships in complex data with very few assumptions about data distribution and interactions. Tree models can work with more than one response variable at a time which makes it possible to identify a group of vehicle characteristics that explain particular emission profiles.

In this project, the predictor variables were vehicle characteristics which are detailed below. The response variables were the monitored vehicle emissions. The models cluster the vehicles into groups or 'nodes' with similar emission values using rules based on the predictor variables.

R code that was provided with NZ Transport Agency research report 596 (Bluett et al., 2016) was amended and the 2015 RSD data set was then used to replicate the multivariate regression tree approach used in Bluett et al (2010).

Ten parameters were selected for RTA including:

- fuel type;
- year of manufacture;
- vehicle make;
- engine capacity (cc);
- odometer reading (km);
- emissions test regime;
- emissions test regime label (a clustering of different test regimes and vehicle criteria);
- origin country;
- previously registered country; and
- power.

Preparatory data cleaning was undertaken to remove records in which there were null or zero values for any of the predictor variables used. Of the 2015 RSD records, 16,173 records remained in the final 'clean' data set.

RTA was used to identify nodes associated with groups of predictor variables (vehicle characteristics) that resulted in the highest response variables (vehicle emissions).

Every additional level of predictor variables resulted in an increase in the size of the RTA model. The size (number of levels) of each model was determined by finding the associated error with different numbers of levels and adjusting the size to allow one standard error more than the minimum possible error in the model. This approach was used to minimise the number of levels (and thus selection criteria per node) whilst maintaining an acceptable amount of error. This is the method previously described in Breiman et al (1984).

Appendix C

Criteria Used for GEV Modelling

Table 10: Modelled GEV Group Criteria - Access/Excel Method (by pollutant)

		Pollutant			
		CO	**HC**	**NO**	**UV smoke**
Selection Criteria	Vehicle Age (years)	$\geq$ 15	$\geq$ 15	$\geq$ 15	$\geq$ 14
	Vehicle Odometer (km)	$\geq$ 171,000	$\geq$ 171,500	$\geq$ 181,000	$\geq$ 182,500
	Fuel Type	Petrol	Petrol	Petrol	Diesel
	Engine Capacity	$\leq$ 1450	$\leq$ 1600	$\leq$ 1600	$\leq$ 2450
Monitored GEVs (a)		762	779	741	741
Monitored TEVs (b)		22138	22121	22159	22159
Modelled GEVs (c + d)		170	849	735	90
Monitored GEVs that fit model (c)		35	108	87	36
Monitored TEVs that fit model (d)		135	741	648	54
Proportion of population in modelled group $(\frac{c+d}{a+b})$		0.7 %	3.7 %	3.2 %	0.4 %
Modelled GEVs with high emissions $(\frac{c}{c+d})$		21 %	13 %	12 %	40 %
Monitored GEVs that fit model $(\frac{c}{a})$		5 %	14 %	12 %	5 %
Bycatch (TEVs modelled as GEVs) $(\frac{d}{b})$		0.6 %	3.3 %	2.9 %	0.2 %

Table 11: Modelled GEV Group Criteria – RTA Method (by pollutant, CO and HC)

Category	Pollutant								
	CO					HC			
	Node 15	Node 27	Node 51	Node 59	Node 75	Node 23	Node 29	Node 63	Node 121
Fuel Type					PETROL	DIESEL			
Year of Manufacture		≤ 2003		≤ 1994					
Engine Capacity				< 2274				1330.5 – 1395.5	≥ 1395.5
Odometer (km)			≥ 292611.5		≥ 205583		< 164580.5		≥ 203787
Make	CHRYSLER, DAIHATSU, JAGUAR, MITSUBISHI, NISSAN, SUBARU	MITSUBISHI, SAAB, VOLVO	AUDI, DAIHATSU, HONDA, ISUZU, LANDROVER, LEXUS, MAZDA, MERCEDES-BENZ, NISSAN, SUBARU, SUZUKI, TOYOTA, VOLKSWAGEN	ALFA ROMEO, AUDI, BMW, CHEVROLET, FORD, HONDA, ISUZU, JEEP, LANDROVER, MAZDA, MERCEDES-BENZ, OPEL, RENAULT, ROVER, SAAB, SUZUKI, TOYOTA, VOLKSWAGEN, VOLVO	ASTON MARTIN, AUDI, CADILLAC, CHERY, CHEVROLET, CHRYSLER, CITROEN, DODGE, FIAT, HOLDEN, HONDA, HYUNDAI, JAGUAR, JEEP, KIA, LAND ROVER, LEXUS, MAZDA, MERCEDES-BENZ, NISSAN, PORSCHE, ROVER, SKODA, SSANGYONG, SUZUKI, TOYOTA, VOLKSWAGEN, VOLVO	CHRYSLER, CITROEN, DODGE, FIAT, FORD, FOTON, HOLDEN, KIA, LANDROVER, LDV, MAZDA, MERCEDES-BENZ, MITSUBISHI, NISSAN, RENAULT	AUDI, CHRYSLER, CITROEN, DAIHATSU, FORD, HYUNDAI, ISUZU, JAGUAR, JEEP, MAZDA, MERCEDES-BENZ, MITSUBISHI, NISSAN, RENAULT, SKODA, SUBARU, SUZUKI, TOYOTA, VOLKSWAGEN, VOLVO		CHRYSLER, FORD, ISUZU, NISSAN, SUBARU, TOYOTA
Origin Country					AUSTRALIA, JAPAN				
Test Regime Label	JA, JBA, JC, JDB, JE, JGA, JGC, JGD, JGE, JGF, JGK, JT, JZ, JZL	EXXXXX, JGH, JHK, JLC, JS, JTA, JTC	EXXXXX, JGH, JHK, JLC, JS, JTA, JTC	JA, JBA, JC, JDB, JE, JGA, JGC, JGD, JGE, JGF, JGK, JT, JZ, JZL	A30/01, A37/01, A79/01, A79/02, A79/03, A80/01, A80/02, AZZZZZ, E02080, E03076, E72306, E91441, E98069, E98077, ECE83, EUR3A, EUR4, EUR4A, EUR5, EUR5A, EUR6, EXEMPT, EZZZZZ, J00/02, J333, JABF, JADF, JC08, JCBF, JDAA, JDAB, JDBF, JHG, JK, JKB, JKD, JKE, JKF, JKG, JKH, JKN, JQ, JTB, JU, JY, US2004, US98P, UZZZZZ	AZZZZZ, JZZZZZ	E98077, JC, JKB, JKH, JS	A79/04, E02080, E72306, E96069, E98069, E98077, EUR1, EUR3A, J02/04, J333, J555, JA, JABF, JADF, JC, JCBF, JDAA, JE, JGA, JGC, JGD, JGF, JK, JKB, JKC, JKD, JKE, JKF, JKG, JKH, JLC, JP, JS, JT, JY, JZ, JZA	

Table 12: Modelled GEV Group Criteria – RTA Method (by pollutant, NO and UV Smoke)

Category	Pollutant					
	NO				UV Smoke	
	Node 7	Node 47	Node 55	Node 159	Node 31	Node 61
Fuel Type	DIESEL, LPG	PETROL	DIESEL	PETROL	DIESEL	
Year of Manufacture				≤ 2002		
Engine Capacity			≥ 2998.5			
Odometer (km)		≥ 144137.5		≥ 138615		
Make	HYUNDAI, JEEP, KIA, RANGE ROVER, SUZUKI	CHRYSLER, DAIHATSU, FORD, MAZDA, SUZUKI	AUDI, BMW, CHRYSLER, CITROEN, DODGE, FIAT, FORD, FOTON, GREAT WALL, HOLDEN, ISUZU, JAGUAR, LAND ROVER, LANDROVER, LDV, MAHINDRA, MAZDA, MERCEDES-BENZ, MITSUBISHI, NISSAN, OPEL, PEUGEOT, RENAULT, SAAB, SKODA, SSANGYONG, TOYOTA, VOLKSWAGEN, VOLVO	ALFA ROMEO, CADILLAC, CUSTOMBUILT, SUZUKI, VOLKSWAGEN		FORD, MAZDA, MERCEDES-BENZ, MITSUBISHI, NISSAN, TOYOTA
Test Regime Label		E96069, EUR1, J02/04, J555, JA, JC, JDAA, JE, JGA, JGC, JGD, JGE, JGF, JH, JP, JT, JTC, JZL	E03076, E72306, EUR2, JE, JK, JKH	E91441, E98069, EUR2, EZZZZZ, J777, JABF, JGH, JGK, JHK, JLA, JLC, JTA, JTB, JZ, JZA, US98P	E02080, E96069, EXXXXX, JE, JKC, JS	AZZZZZ, EUR3A, EUR4A, J00/02, J333, J777, JKB, JKD, JKE, JQ, JY

Appendix D

Backfilling of Missing Power Values

Based upon initial RTA, vehicle power was found to be a strong predictor of GEVs of CO and HC.

Using records with valid power values from the data set previously used for GEV Fleet Characterisation, vehicles were separated dependent upon fuel type (petrol or diesel) and scatter plots were made to observe the relationship between vehicle power and GVM, and vehicle power and engine capacity. The petrol vehicles showed a weak correlation between power and GVM and a moderate-strong positive correlation (R^2 = 0.782) between power and engine size. The diesel vehicles showed a weak correlation between power and both GVM and engine size.

To further investigate the relationship between power and engine size, the data set was split depending upon vehicle year. The vehicle year was split into 5-year categories and each category was plotted as a separate series. Trendlines were added for each series and these showed an increasing gradient with increased year, i.e. for vehicles manufactured more recently, power increases more with increased engine size. This behaviour aligns with the expected improvements in technology resulting in development of more efficient engines over time. With the exception of the oldest vehicles in the data set (Pre-1996), the categories grouped by vehicle year also showed a better correlation than that between power and engine size alone (R^2 = 0.794 to 0.881). The resulting plot is shown in Figure 18 below, and the relationships are listed in Table 13. It was determined that these relationships should be used to back-fill the null values for power for petrol vehicles in the data set.

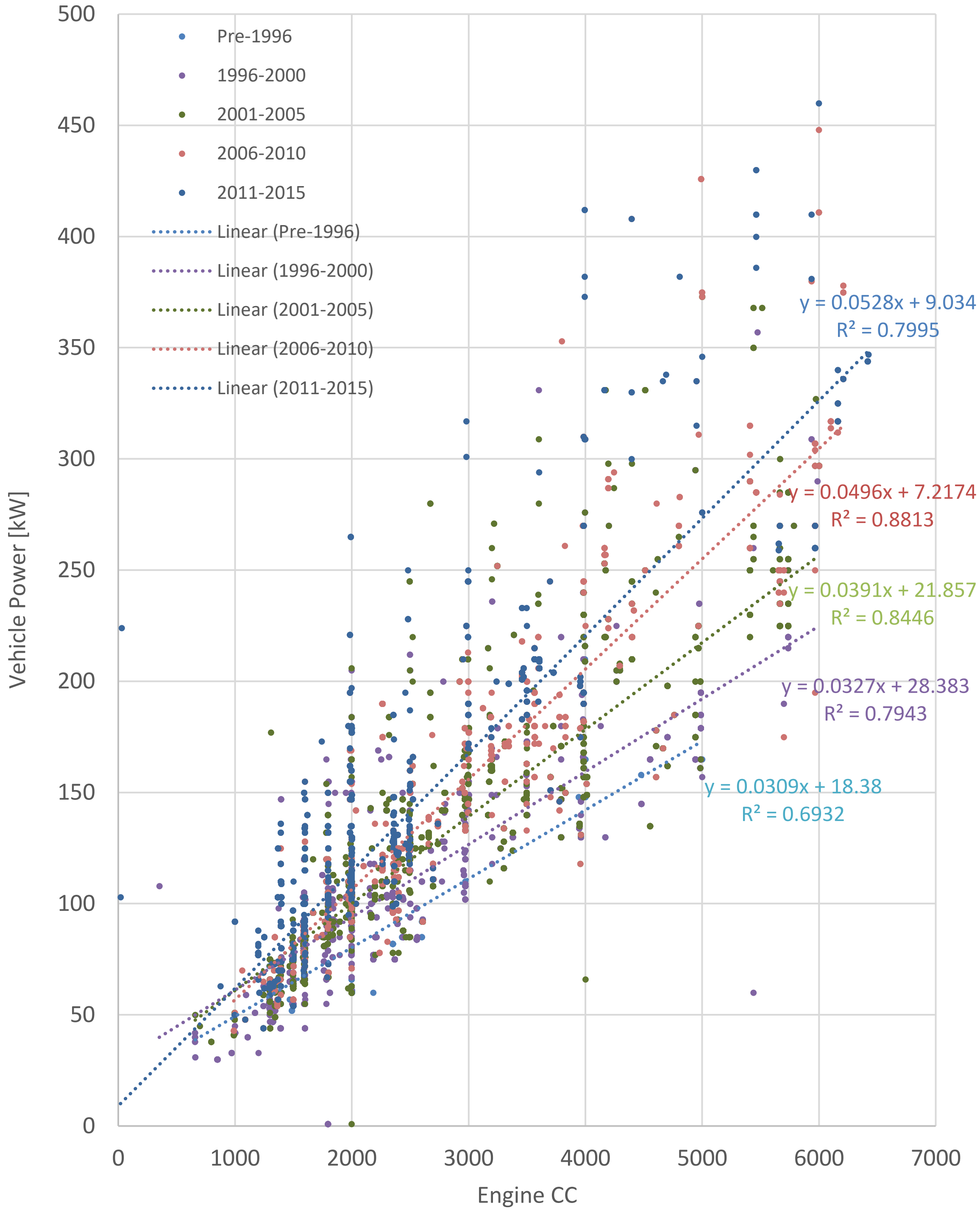


Figure 18: Scatter plot for petrol vehicles showing relationship between power (kW) and engine size, vehicles categorised by year (5-yearly groups)

Table 13: Relationships for back-filling power values from vehicle year and engine size (petrol vehicles)

Vehicle Year	R^2 value	Relationship
Pre-1996	0.693	$Power = 0.0309\ (engine\ CC) + 18.38$
1996 – 2000	0.794	$Power = 0.0327\ (engine\ CC) + 28.383$
2001 – 2005	0.845	$Power = 0.0391\ (engine\ CC) + 21.857$
2006 – 2010	0.881	$Power = 0.0496\ (engine\ CC) + 7.2174$
2011 – 2015	0.800	$Power = 0.0528\ (engine\ CC) + 9.034$

The weaker correlation between diesel vehicle power values and either GVM or engine size is likely due to the nature of diesel engines, as larger engines are often associated with higher torque (rotational power) whereas the vehicle power value (kW) is an indicator of linear power. From the initial scatter plots, there was an obvious abrupt cut-off at a GVM of 3,500 kg. This was due to the data set being filtered to contain only light passenger and light commercial vehicles (defined as GVM ≤ 3500 kg). To remove the bias introduced by using only a section of the fleet, a data set was extracted from the national fleet data containing all diesel vehicles with valid power values classified as 'Passenger Car/Van' or 'Goods Van/Truck/Utility'. This larger data set (>450,000 records vs ~11,500 records) gave a moderate positive correlation ($R^2 = 0.594$) between power and engine size. However, when split into categories by vehicle year, the increasing gradient trend with time observed for the petrol vehicles was not distinguishable. The data was further separated into two groups by classification. The 'Passenger Car/Van' set was plotted and showed a similar behaviour to that in the petrol vehicle data, albeit with weaker correlation (post-1996 R^2 values between 0.511 and 0.652). The resulting plot is shown in Figure 19 below, and the relationships are listed in Table 14. Values for vehicles newer than 2015 have been omitted as the data set that requires back-filling was created in 2015 so will not include newer vehicles. It was determined that these relationships should be used to back-fill the null values for power for diesel vehicles classified as 'Passenger Car/Van' in the data set.

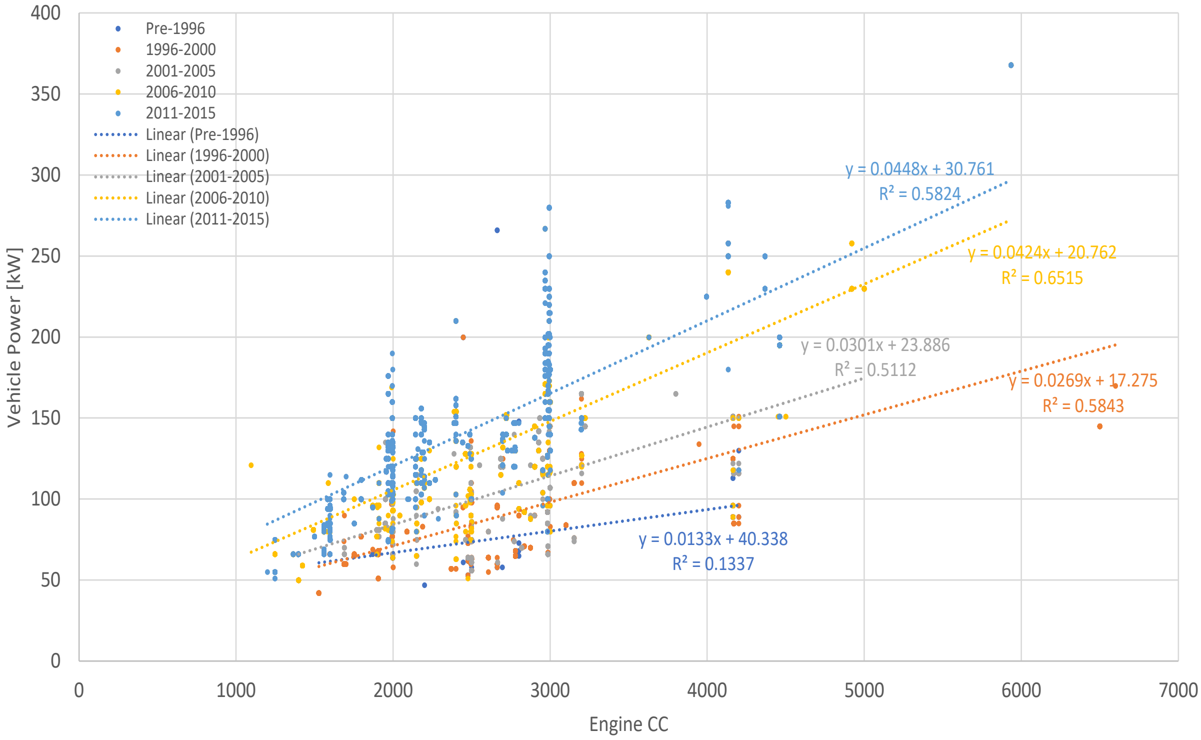


Figure 19: Scatter plot for diesel vehicles classified as 'Passenger Car/Van' showing relationship between power and engine size, vehicles categorised by year (5-yearly groups)

Table 14: Relationships for back-filling power values from vehicle year and engine size (diesel vehicles classified as 'Passenger Car/Van')

Vehicle Year	R^2 value	Relationship
Pre-1996	0.134	$Power = 0.0133\,(engine\ CC) + 40.338$
1996 – 2000	0.584	$Power = 0.0269\,(engine\ CC) + 17.275$
2001 – 2005	0.511	$Power = 0.0301\,(engine\ CC) + 23.886$
2006 – 2010	0.652	$Power = 0.0424\,(engine\ CC) + 20.762$
2011 – 2015	0.582	$Power = 0.0448\,(engine\ CC) + 30.761$

The data set for diesel vehicles classified as 'Goods Van/Truck/Utility' showed a similar relationship although separation between the trends of earlier vehicles was less distinct and the trend lines were closer to parallel than the more radial change previously observed. The correlation was stronger for this category of diesel vehicles than for the ‘Passenger Car/Van’ category. The resulting plot is shown in Figure 20 below, and the relationships are listed in Table 15. It was determined that these relationships should be used to back-fill the null values for power for diesel vehicles classified as ‘Goods Van/Truck/Utility’ in the data set.

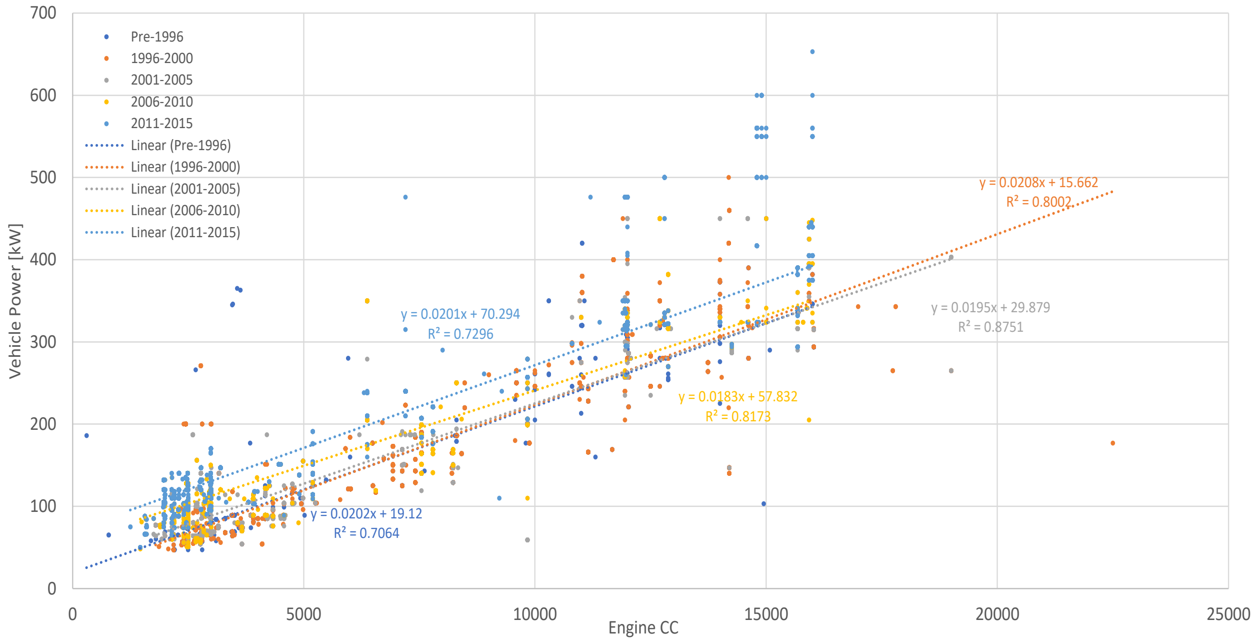


Figure 20: Scatter plot for diesel vehicles classified as ‘Goods Van/Truck/Utility’ showing relationship between power and engine size, vehicles categorised by year (5-yearly groups)

Table 15: Relationships for back-filling power values from vehicle year and engine size (diesel vehicles classified as 'Goods Van/Truck/Utility')

Vehicle Year	R^2 value	Relationship
Pre-1996	0.706	$Power = 0.0202\ (engine\ CC) + 19.12$
1996 – 2000	0.800	$Power = 0.0208\ (engine\ CC) + 15.662$
2001 – 2005	0.875	$Power = 0.0195\ (engine\ CC) + 29.879$
2006 – 2010	0.817	$Power = 0.0183\ (engine\ CC) + 57.832$
2011 – 2015	0.730	$Power = 0.0201\ (engine\ CC) + 70.294$

It should be noted that this assessment is not intended to describe the nature of any causative relationship between the parameters investigated. The purpose is limited to observation of correlation between parameters and increase of the correlation strength through categorisation of data. The trends described will be used to estimate a value for vehicle records with null power values in the 2015 emissions data set. These values will then be used as an optional parameter for RTA of this data set to investigate whether they are a strong predictor of whether a vehicle will be a GEV.

There are several limitations implicit in the method used for determining the trends above. These include but are not limited to:

- The existing power values in the data may be biased. The presence of a power value appears to be linked to vehicle make, model and year. It is unknown if the vehicles for which power data has been entered into the MVR records have implicit bias and what the effect of this may be on the resulting trends observed, i.e. if all Toyota vehicles have associated power data but Suzuki vehicles do not, the power data will trend towards vehicles that are similar to Toyotas.
- During this assessment, several outliers were noticed and investigated due to unusually high or low values. It was found that most of these appeared to be errors in data entry. A common issue seemed to be addition of an extra digit which effectively increased the value by a factor of 10. As all vehicles of the same type (make, model, year) are populated with the same power value, the impact of an error of this type on a common vehicle would be a significant skew in the data trend. Errors identified during the assessment were corrected as appropriate but there may be more of this type of error within the data set that are not as immediately obvious from observation of the scatter plots.

Appendix E

Calculation of AADT (National Fleet)

Calculation of AADT for the national fleet was undertaken as it reflects vehicle use which directly impacts the rate of vehicle emissions through its relationship with emission factors and fuel efficiency.

Average annual distance travelled (AADT) values were combined with regional vehicle counts and null records to allow calculation of a combined AADT for groups of modelled GEVs and TEVs for each pollutant.

The AADT value calculated for vehicles in each node within the R code omitted null values and calculated an average based upon the number of non-null values in the regional set. As multiple nodes were combined for each pollutant, this required calculation of an average AADT for more than one node. This was achieved by multiplying the node AADT by the number of non-null records for each node, adding this number for all nodes associated with the pollutant, and then dividing the total by the sum of non-null records for the combined nodes for that pollutant.

The calculation method below assumes combining of three nodes but was adjusted as necessary for more or fewer nodes.

$$A_{combined\ group} = \frac{[A_1 \times (N_1 - n_1)] + [A_2 \times (N_2 - n_2)] + [A_3 \times (N_3 - n_3)]}{(N_1 + N_2 + N_3) - (n_1 + n_2 + n_3)}$$

Where the subscript denotes the group;

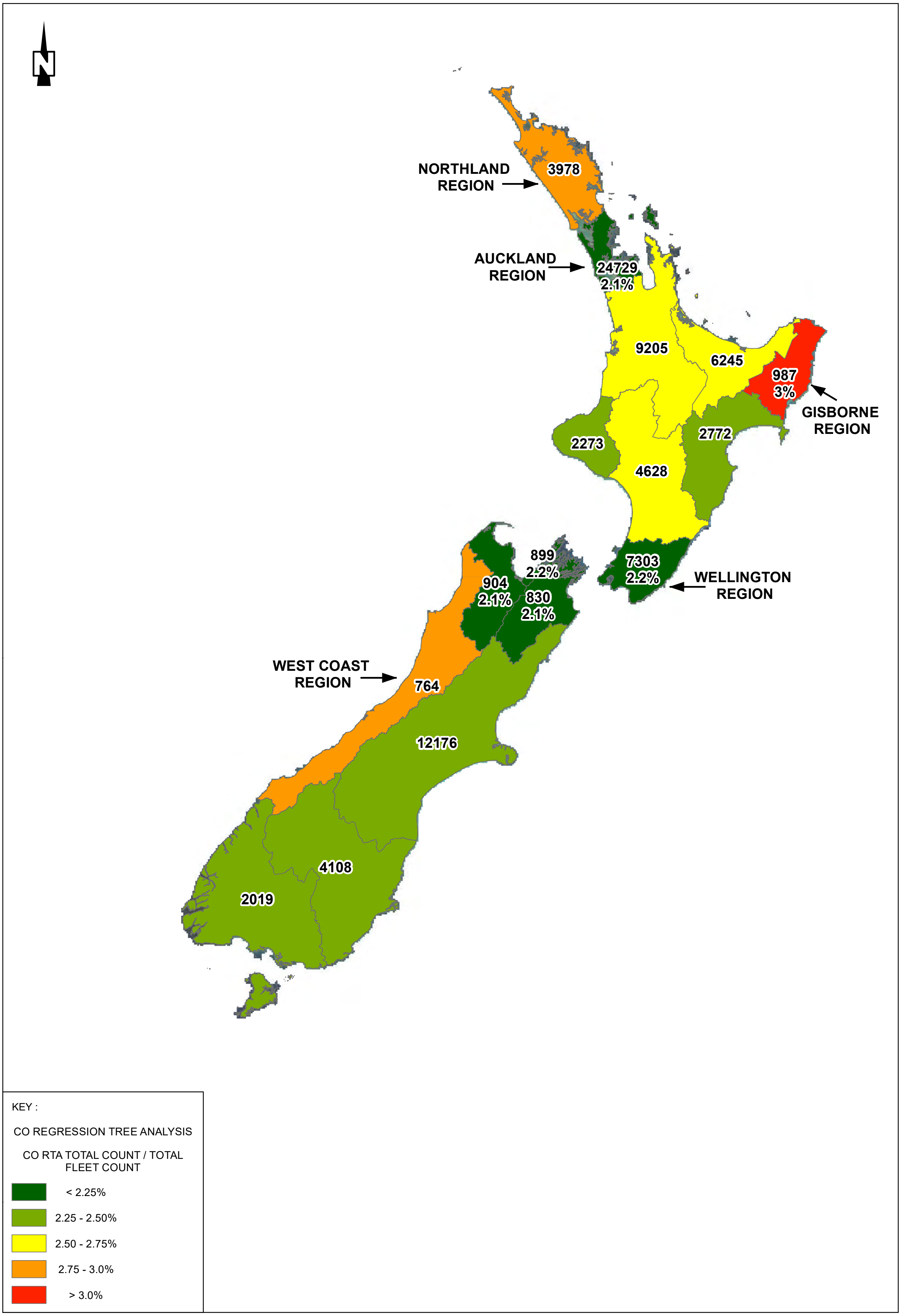
A is the AADT (of all non-null records);

N is the total number of records; and

n is the total number of null records

Appendix F

Heat Maps Showing Proportion of Modelled GEVs at Regional and District Level



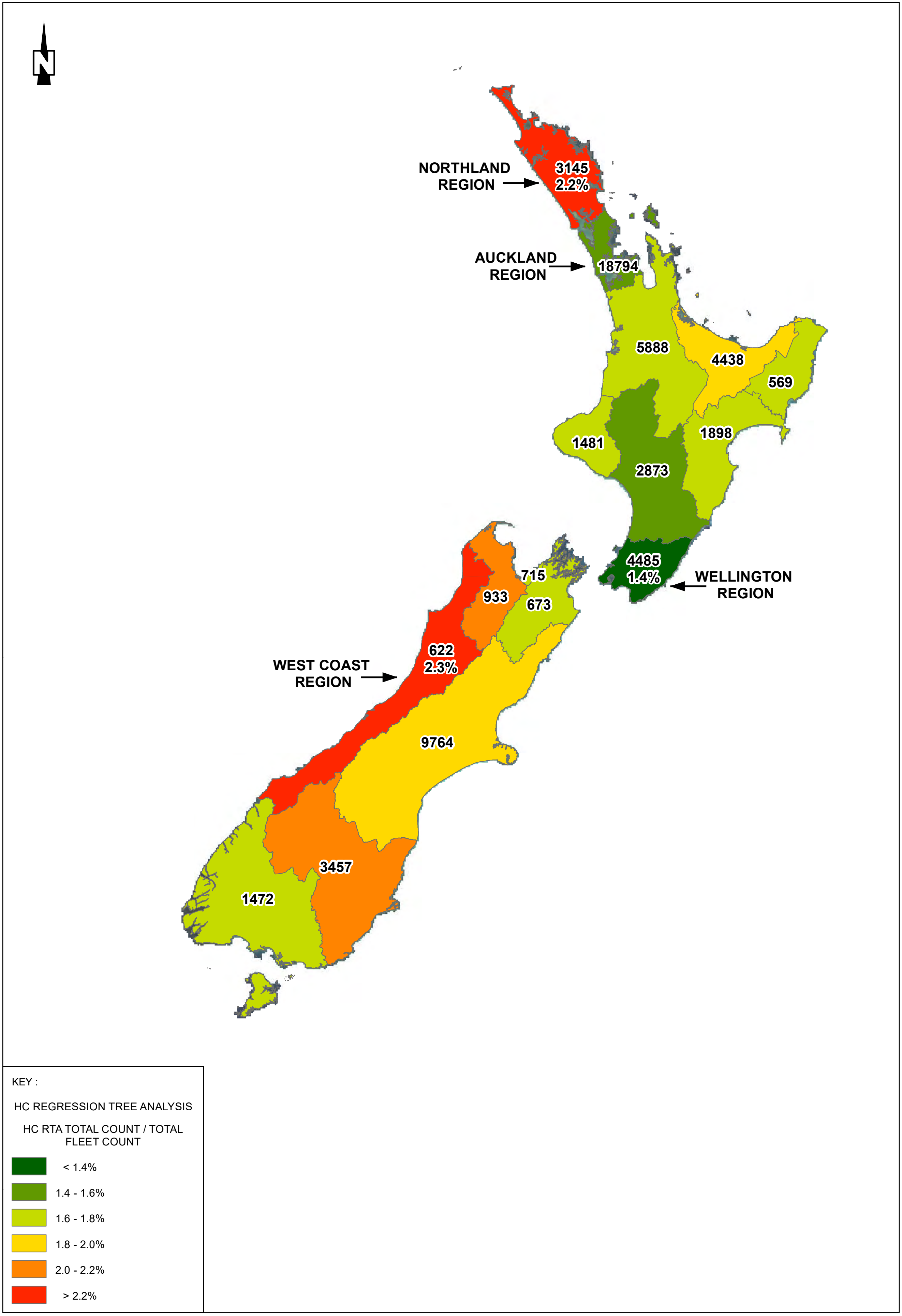
SOURCE:
1. AERIAL IMAGERY SOURCED FROM THE LINZ DATA SERVICE AND LICENCED BY [LISCENSOR] FOR RE-USE UNDER THE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENCE.
2. REGIONAL BOUNDARIES INFORMATION SOURCED FROM STATISTICS NZ.

FIGURE 1 : CO REGRESSION TREE ANALYSIS (REGIONAL)

SCALE : 1:8,000,000 (A4)

0 50 100 200 300

Kilometers



SOURCE:
1. AERIAL IMAGERY SOURCED FROM THE LINZ DATA SERVICE AND LICENCED BY [LISCENSOR] FOR RE-USE UNDER THE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENCE.
2. REGIONAL BOUNDARIES INFORMATION SOURCED FROM STATISTICS NZ.

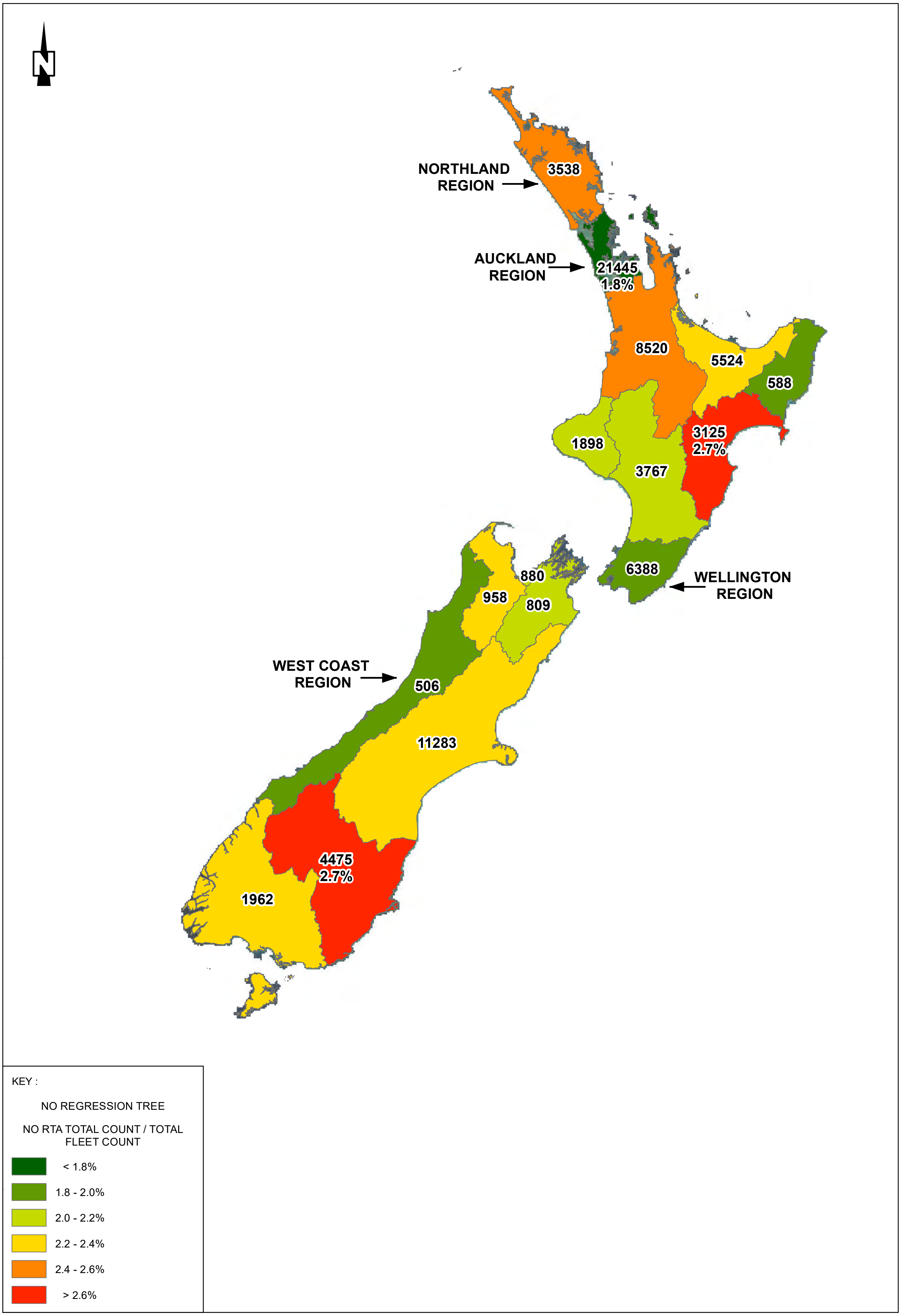
FIGURE 2 : HC REGRESSION TREE ANALYSIS (REGIONAL)

SCALE : 1:8,000,000 (A4)

0 50 100 200 300

Kilometers

C03954800_Z002A_RegionalHeatMaps_HC.mxd 25/06/2020 ISSUE 1

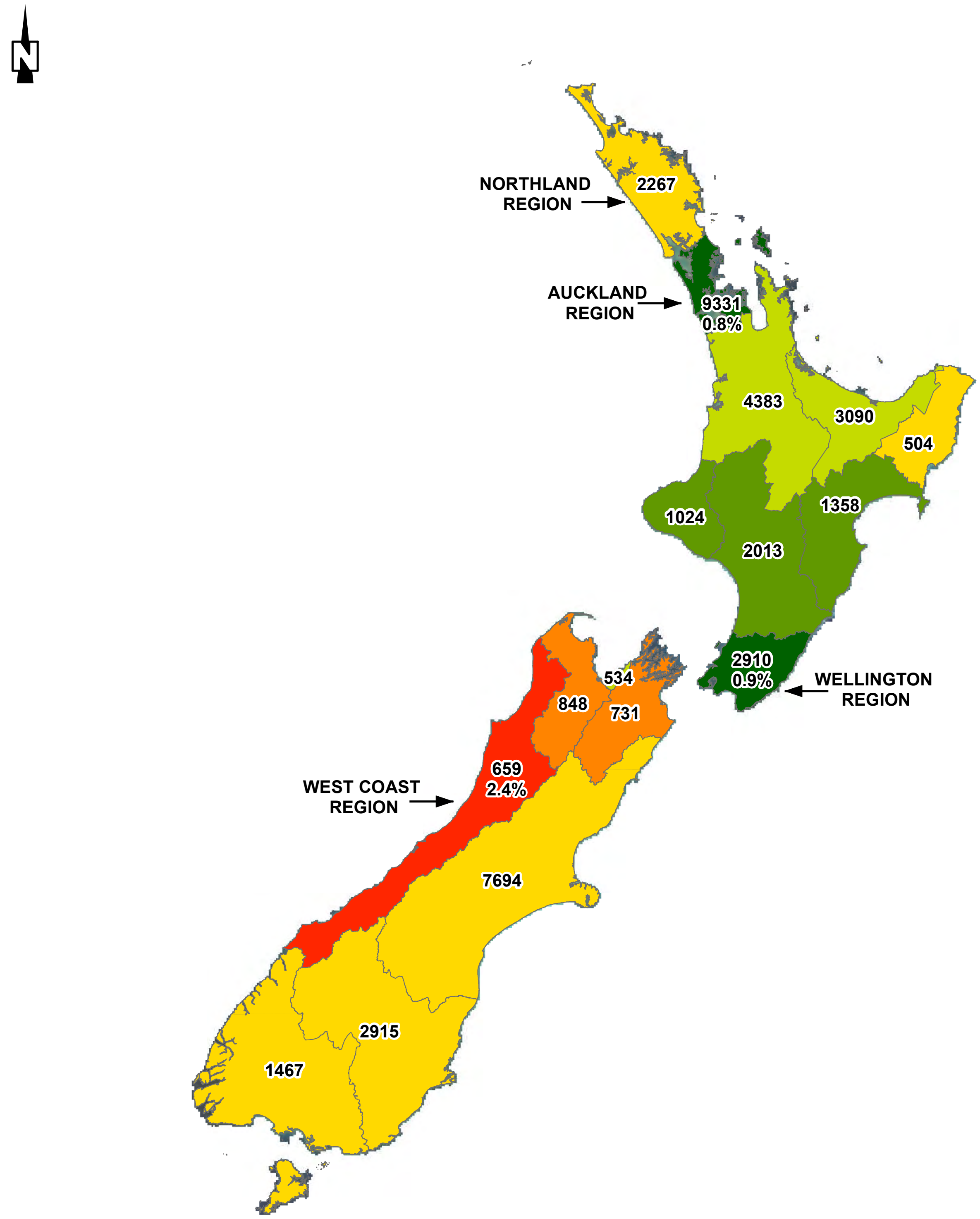


SOURCE:
1. AERIAL IMAGERY SOURCED FROM THE LINZ DATA SERVICE AND LICENCED BY [LISCENSOR] FOR RE-USE UNDER THE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENCE.
2. REGIONAL BOUNDARIES INFORMATION SOURCED FROM STATISTICS NZ.

FIGURE 3 : NO REGRESSION TREE ANALYSIS (REGIONAL)

SCALE : 1:8,000,000 (A4)

0 50 100 200 300
Kilometers



KEY :

UV SMOKE REGRESSION TREE ANALYSIS

UV RTA TOTAL COUNT / TOTAL FLEET COUNT

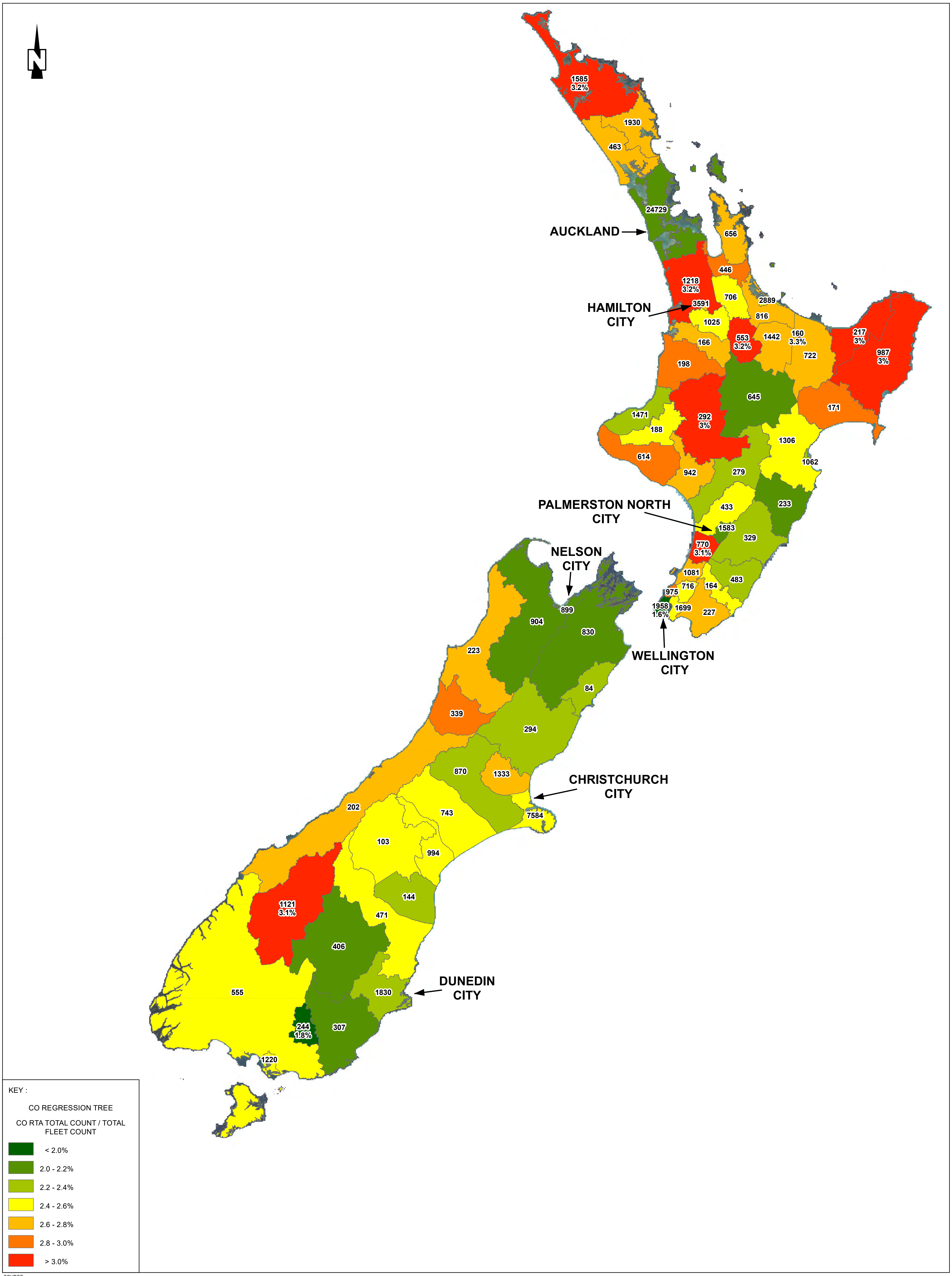
- < 1.0%
- 1.0 - 1.25%
- 1.25 - 1.5%
- 1.5 - 1.75%
- 1.75 - 2.0%
- > 2.0%

SOURCE:
1. AERIAL IMAGERY SOURCED FROM THE LINZ DATA SERVICE AND LICENCED BY [LISCENSOR] FOR RE-USE UNDER THE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENCE.
2. REGIONAL BOUNDARIES INFORMATION SOURCED FROM STATISTICS NZ.

FIGURE 4 : PM REGRESSION TREE ANALYSIS (REGIONAL)

SCALE : 1:8,000,000 (A4)

0 50 100 200 300
Kilometers

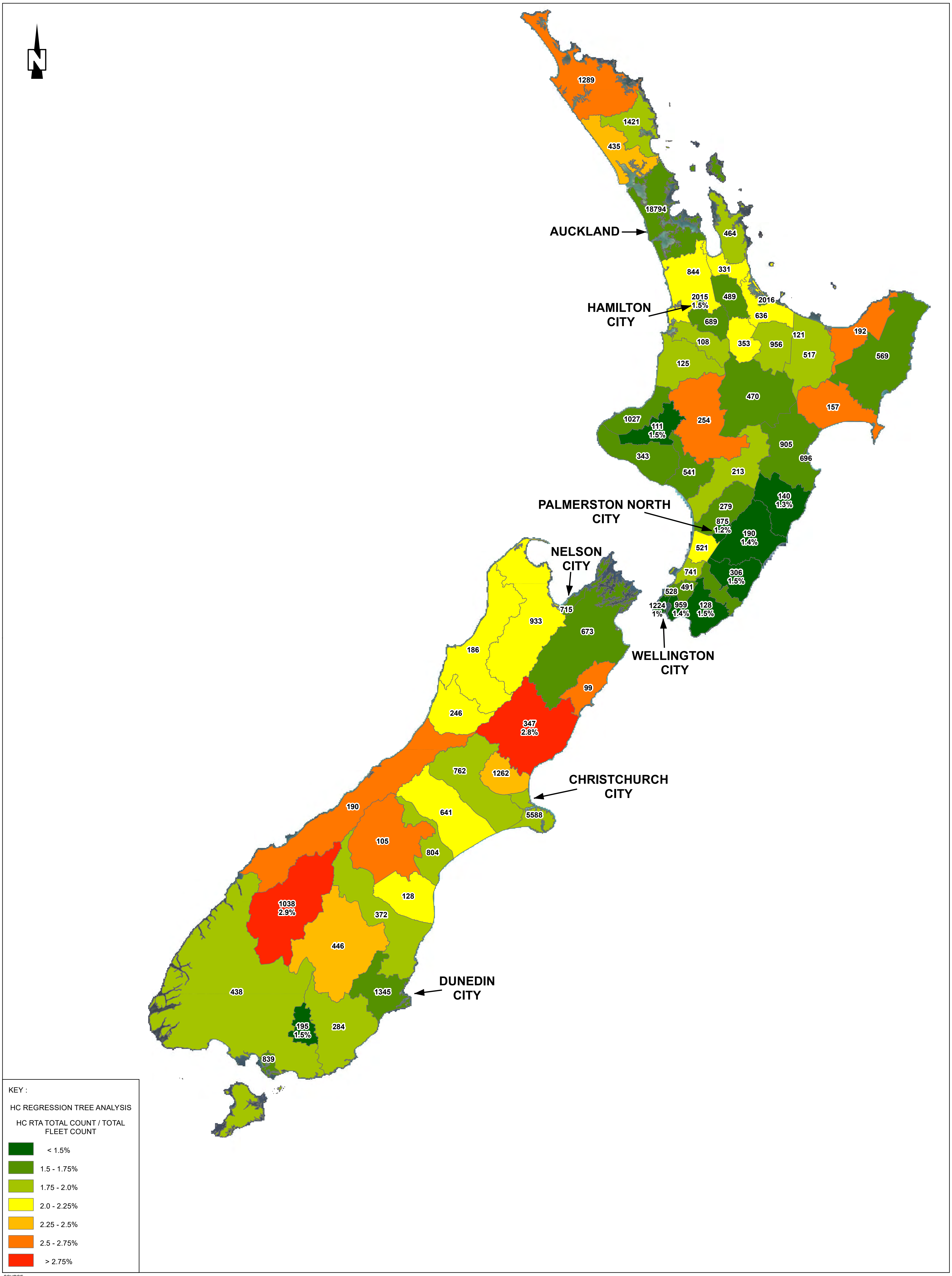


SOURCE:
1. AERIAL IMAGERY SOURCED FROM THE LINZ DATA SERVICE AND LICENCED BY [LISCENSOR] FOR RE-USE UNDER THE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENCE.
2. REGIONAL BOUNDARIES INFORMATION SOURCED FROM STATISTICS NZ.

FIGURE 5 : CO REGRESSION TREE ANALYSIS (TLA)

SCALE : 1:4,500,000 (A3)

0 50 100 200 300
Kilometers

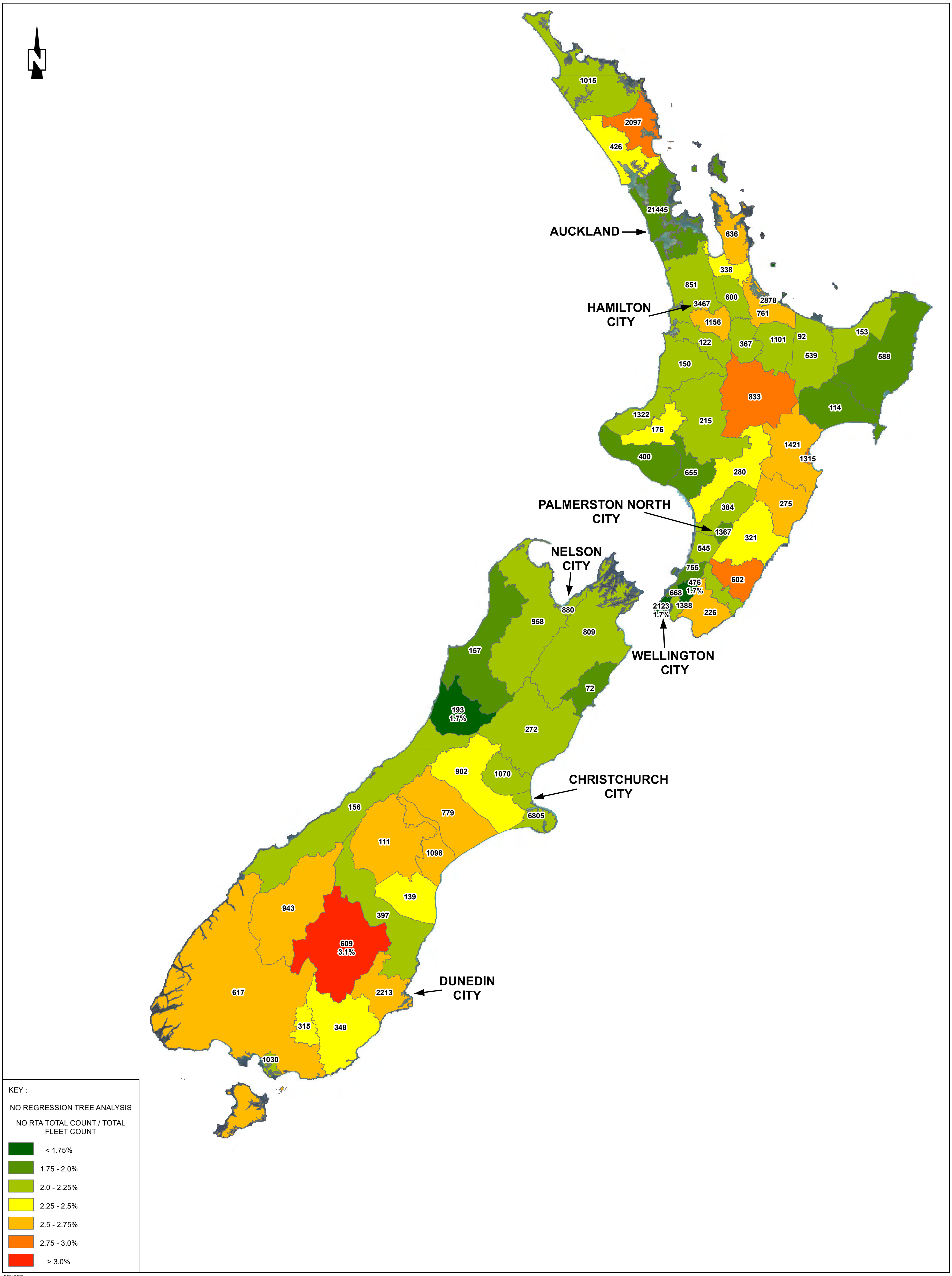


SOURCE:
1. AERIAL IMAGERY SOURCED FROM THE LINZ DATA SERVICE AND LICENCED BY [LISCENSOR] FOR RE-USE UNDER THE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENCE.
2. REGIONAL BOUNDARIES INFORMATION SOURCED FROM STATISTICS NZ.

FIGURE 6 : HC REGRESSION TREE ANALYSIS (TLA)

SCALE : 1:4,500,000 (A3)

0 50 100 200 300
Kilometers

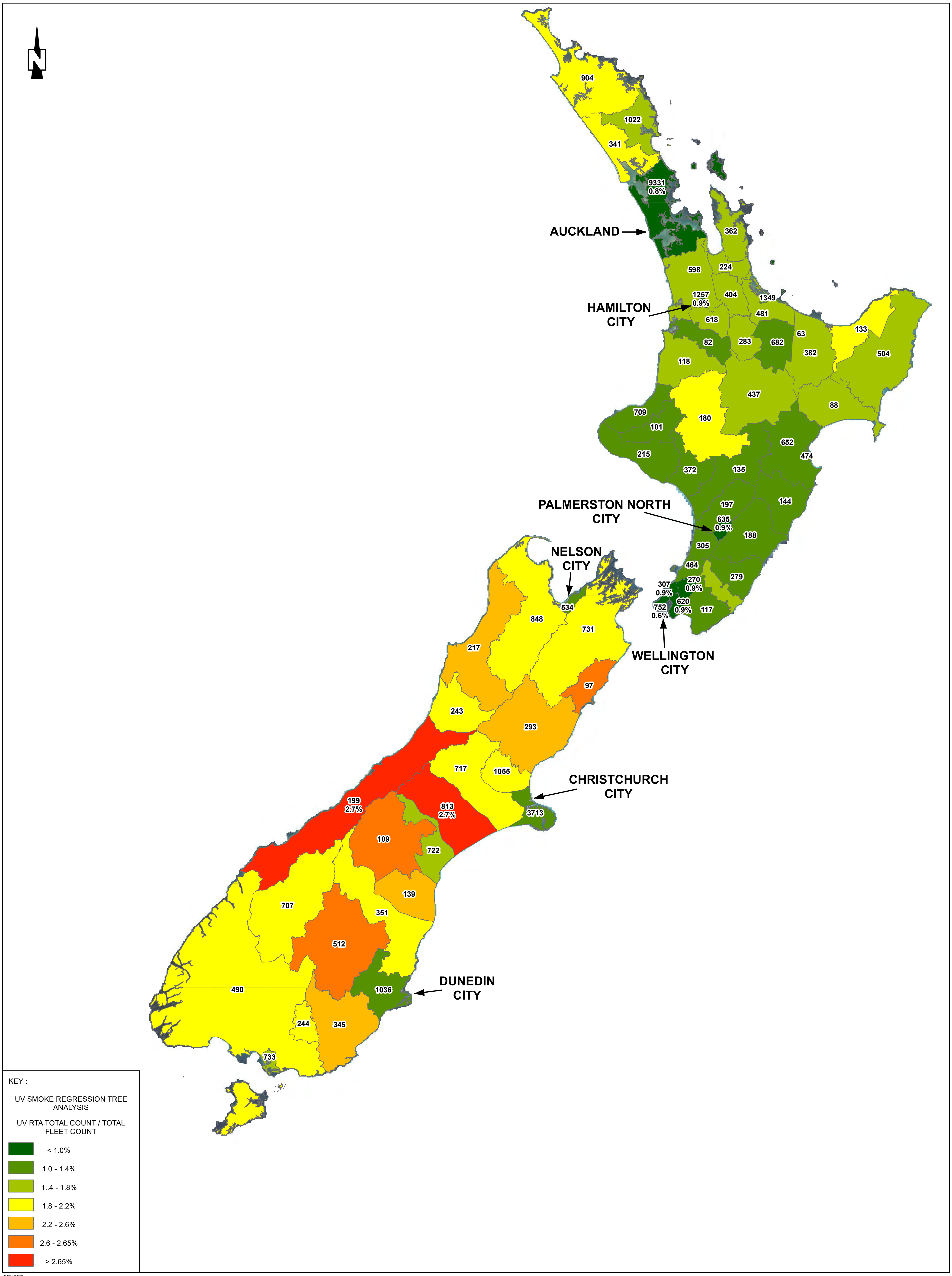


SOURCE:
1. AERIAL IMAGERY SOURCED FROM THE LINZ DATA SERVICE AND LICENCED BY [LISCENSOR] FOR RE-USE UNDER THE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENCE.
2. REGIONAL BOUNDARIES INFORMATION SOURCED FROM STATISTICS NZ.

FIGURE 7 : NO REGRESSION TREE ANALYSIS (TLA)

SCALE : 1:4,500,000 (A3)

0 50 100 200 300
Kilometers



SOURCE:
1. AERIAL IMAGERY SOURCED FROM THE LINZ DATA SERVICE AND LICENCED BY [LISCENSOR] FOR RE-USE UNDER THE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENCE.
2. REGIONAL BOUNDARIES INFORMATION SOURCED FROM STATISTICS NZ.

FIGURE 8 : PM REGRESSION TREE ANALYSIS (TLA)

SCALE : 1:4,500,000 (A3)

0 50 100 200 300
Kilometers

Appendix G

Conversion of Emissions Data to Emission Factors

In order to estimate the annual emissions from the vehicles in the data set, the emissions data was converted into emission factors indicating grams of pollutant per kilogram of fuel burned. This was done using the Pokharel et al. (2002) method as cited by Golder Associates (NZ) Limited (2018).

The emissions log reported CO as a percentage and HC and NO in parts per million. The parts per million values were divided by 10,000 to convert them to percentages. The RSD gas measurements (CO, HC, ad NO) were then converted into exhaust gas ratios of CO/CO_2, HC/CO_2, and NO/CO_2:

$$Q = \frac{CO\%}{CO_2\%}$$

$$Q' = \frac{HC\%}{CO_2\%}$$

$$Q'' = \frac{NO\%}{CO_2\%}$$

These ratios were then used to produce estimates of grams of pollutant per kilogram of fuel burned using the following equations as described by Pokharel et al. (2005).

$$\frac{g\ CO}{kg\ fuel} = \frac{28 \times Q \times 860}{\left(1 + Q + (2 \times 3Q')\right) \times 12}$$

$$\frac{g\ HC}{kg\ fuel} = \frac{2 \times 44 \times Q' \times 860}{\left(1 + Q + (2 \times 3Q')\right) \times 12}$$

$$\frac{g\ NO}{kg\ fuel} = \frac{30 \times Q'' \times 860}{\left(1 + Q + (2 \times 3Q')\right) \times 12}$$

Appendix H

Figures for Other Pollutants of Interest (HC, NO and PM)

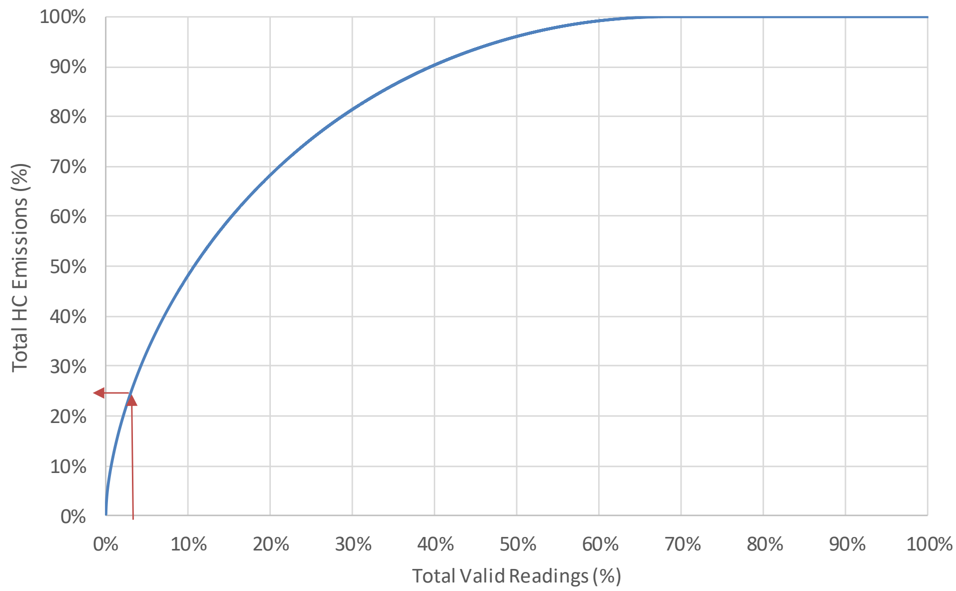


Figure 21: HC Cumulative Distribution Plot

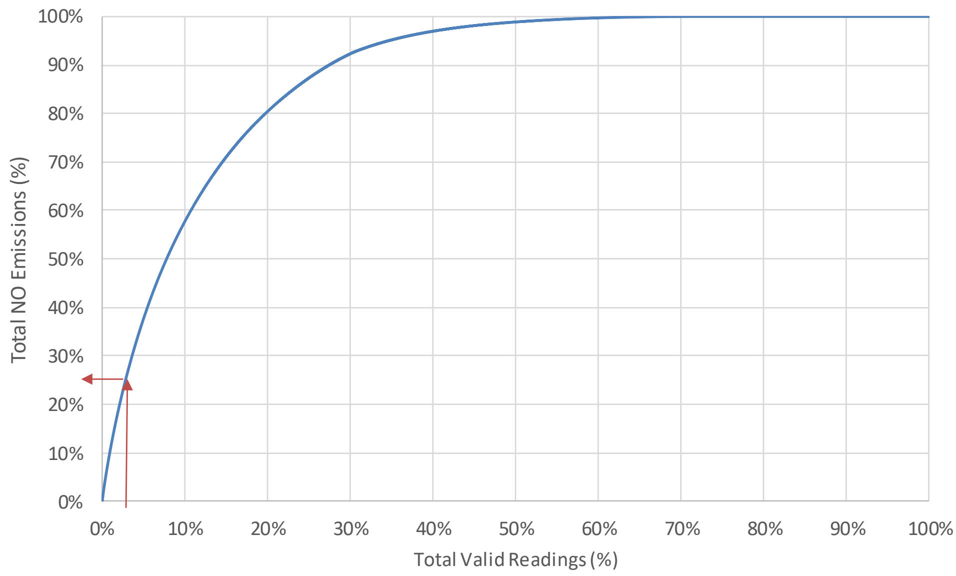


Figure 22: NO Cumulative Distribution Plot

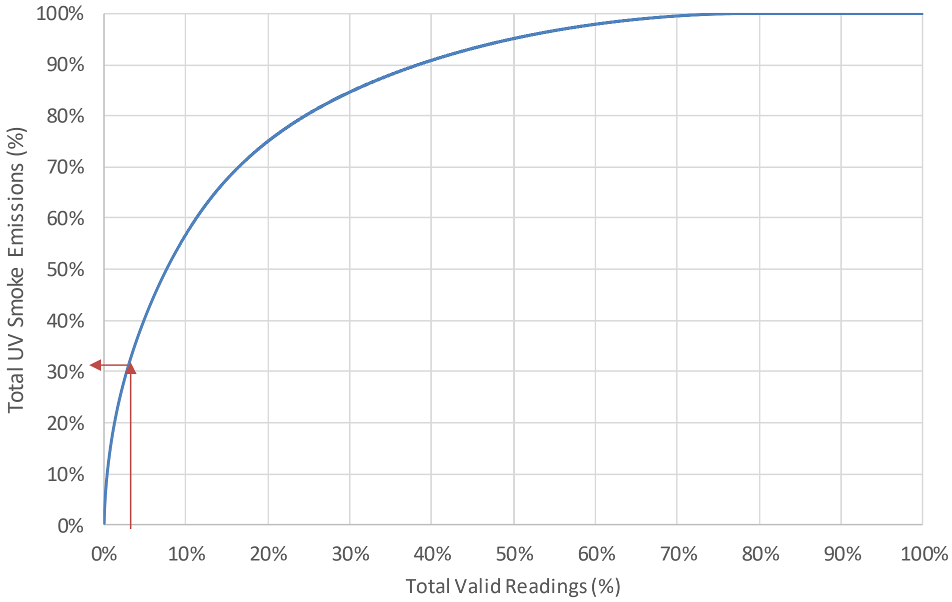


Figure 23: PM Cumulative Distribution Plot

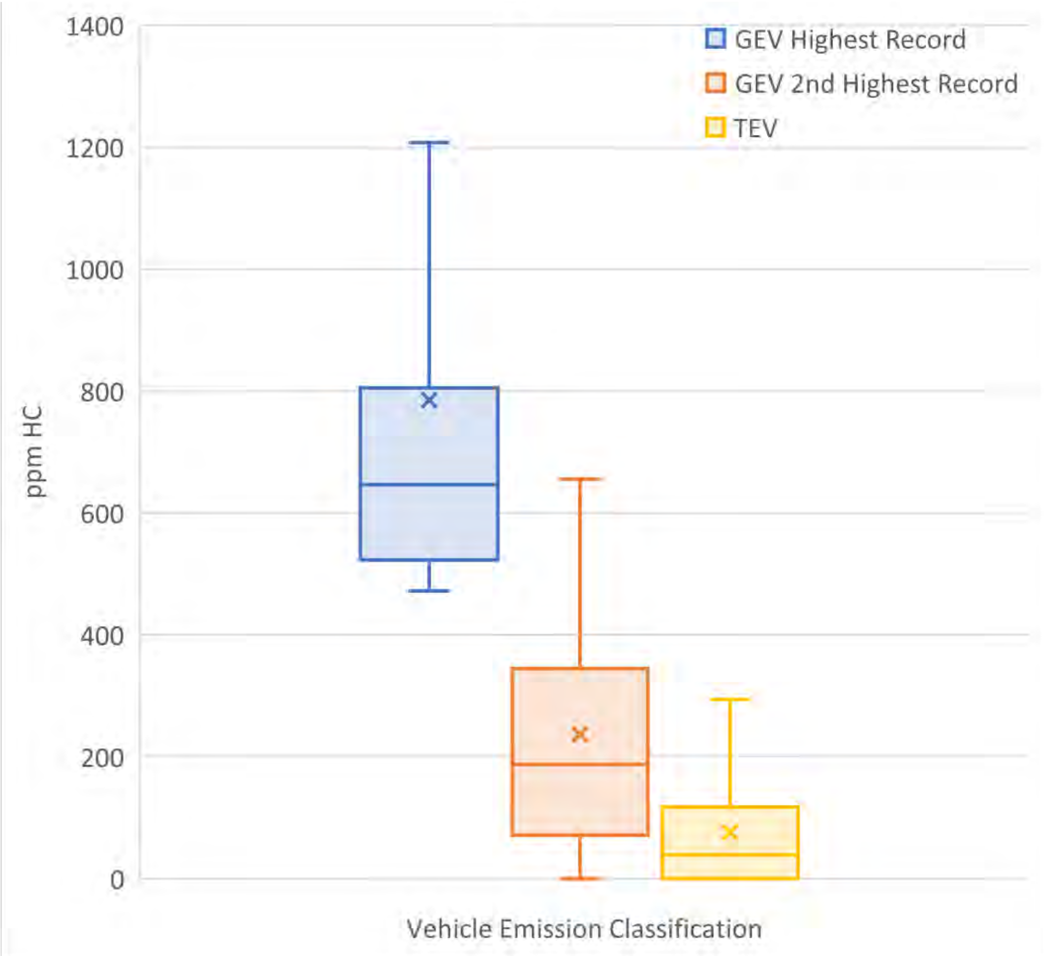


Figure 24: HC GEVs with duplicate measurements showing first highest and second highest emissions for vehicles classified as GEVs compared with HC TEV fleet emissions

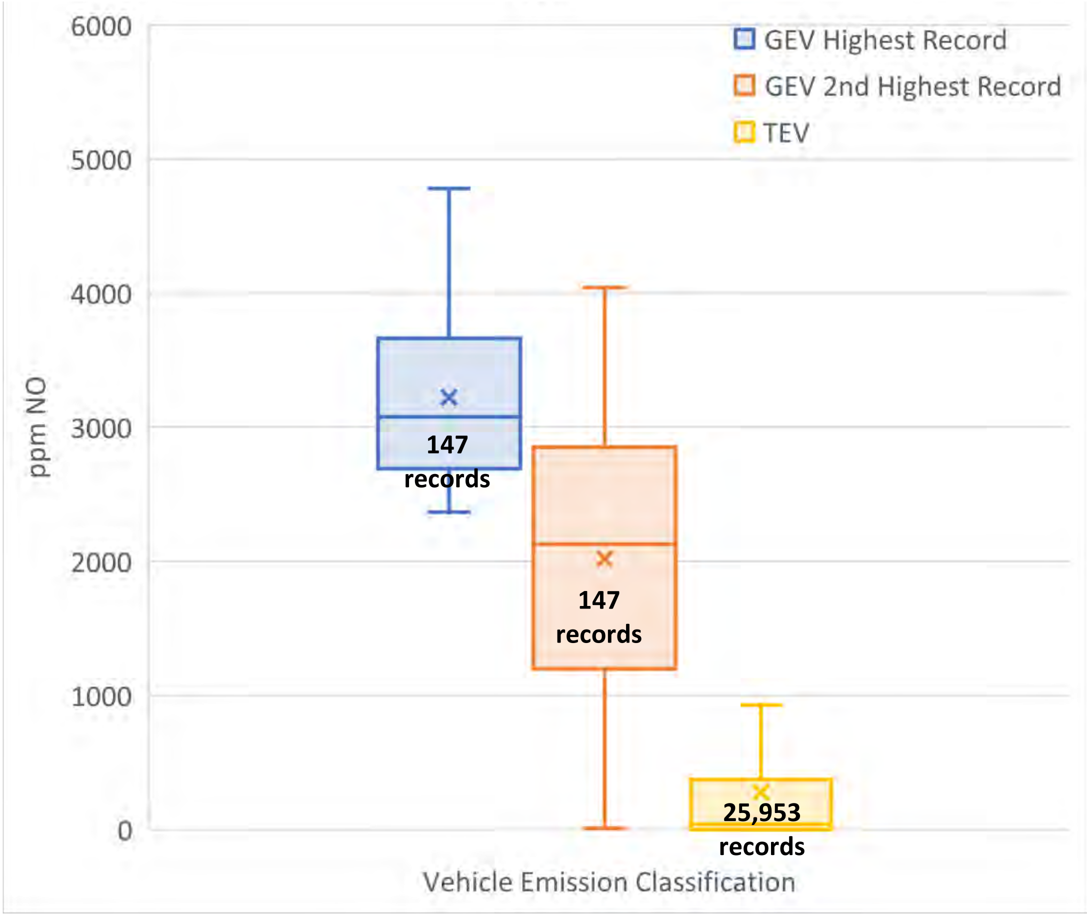


Figure 25: NO GEVs with duplicate measurements showing first highest and second highest emissions for vehicles classified as GEVs compared with NO TEV fleet emissions

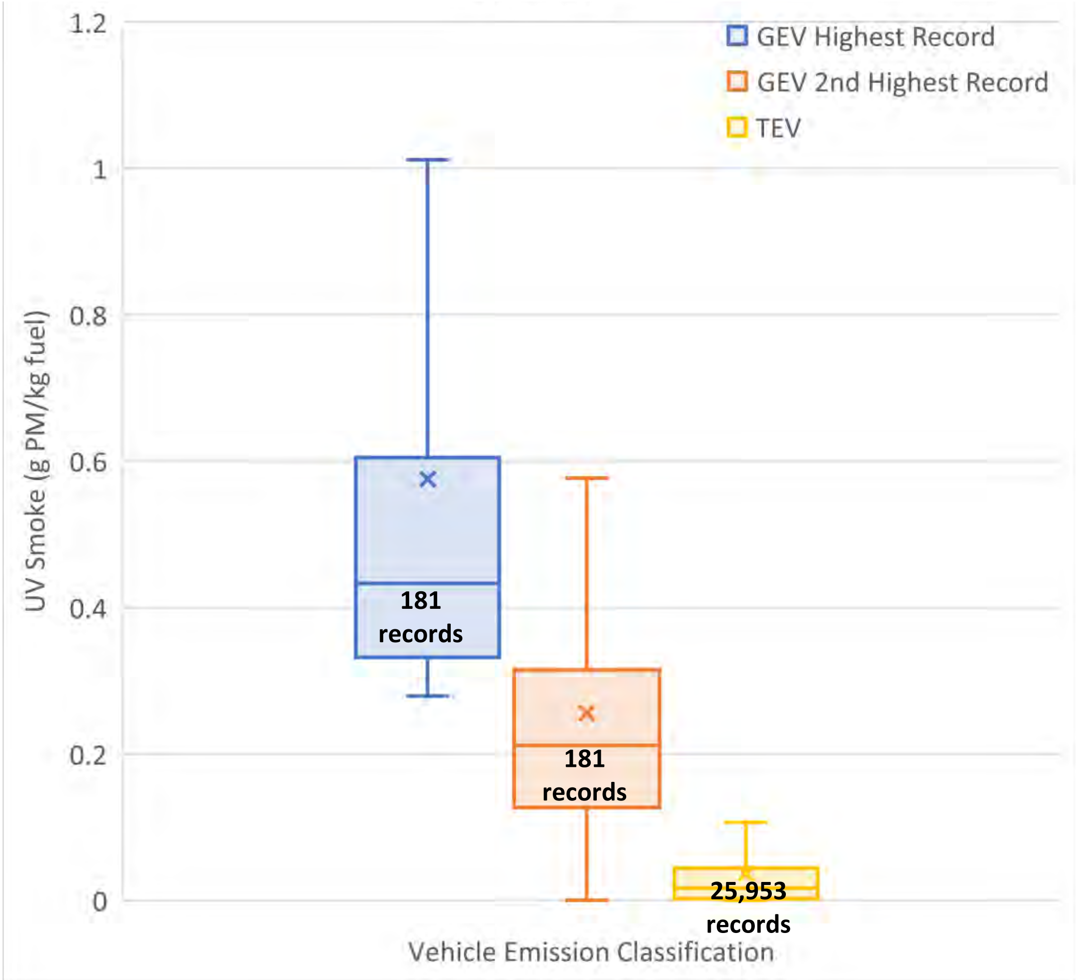


Figure 26: PM GEVs with duplicate measurements showing first highest and second highest emissions for vehicles classified as GEVs compared with PM TEV fleet emissions

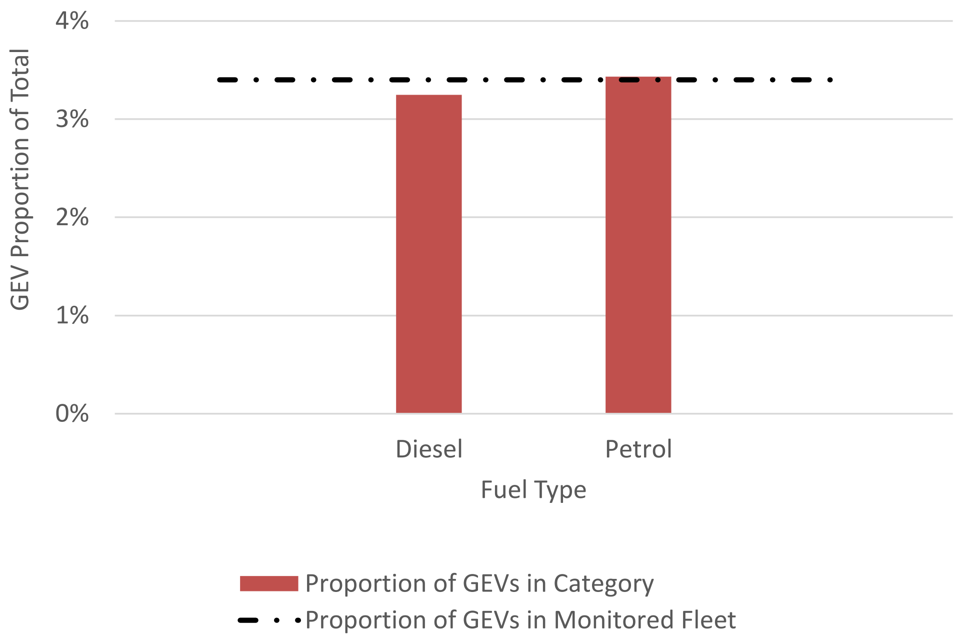


Figure 27: Fuel Types of HC GEVs

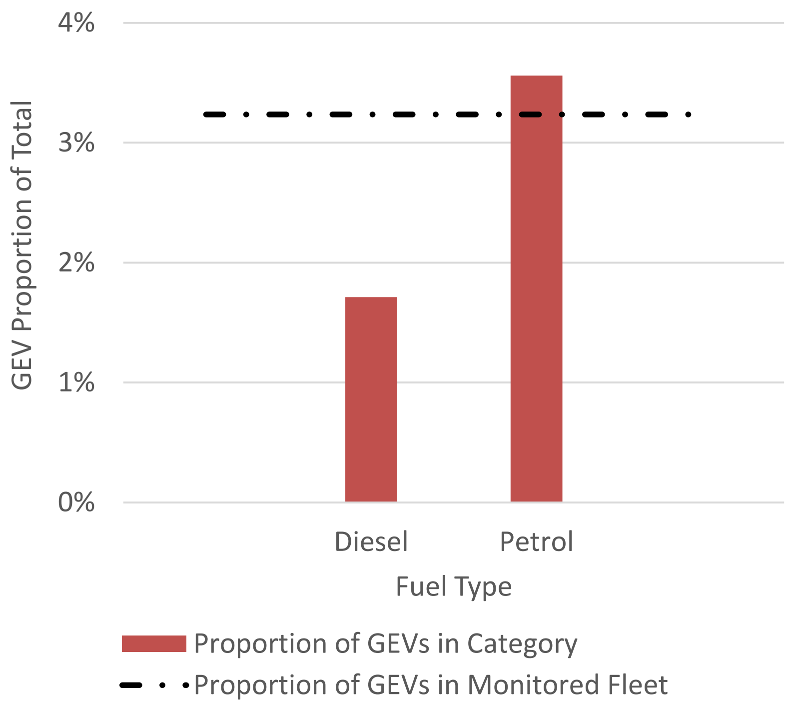


Figure 28: Fuel Types of NO GEVs

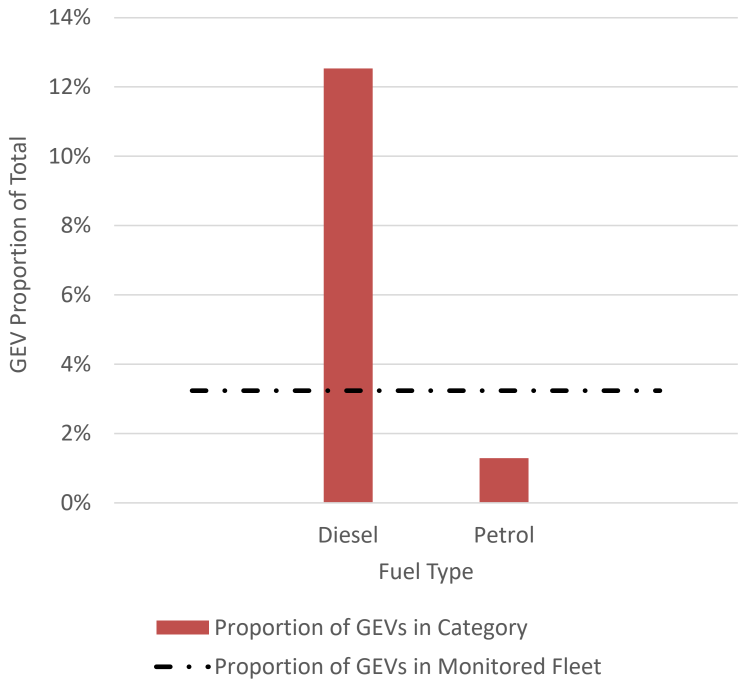


Figure 29: Fuel Types of PM GEVs

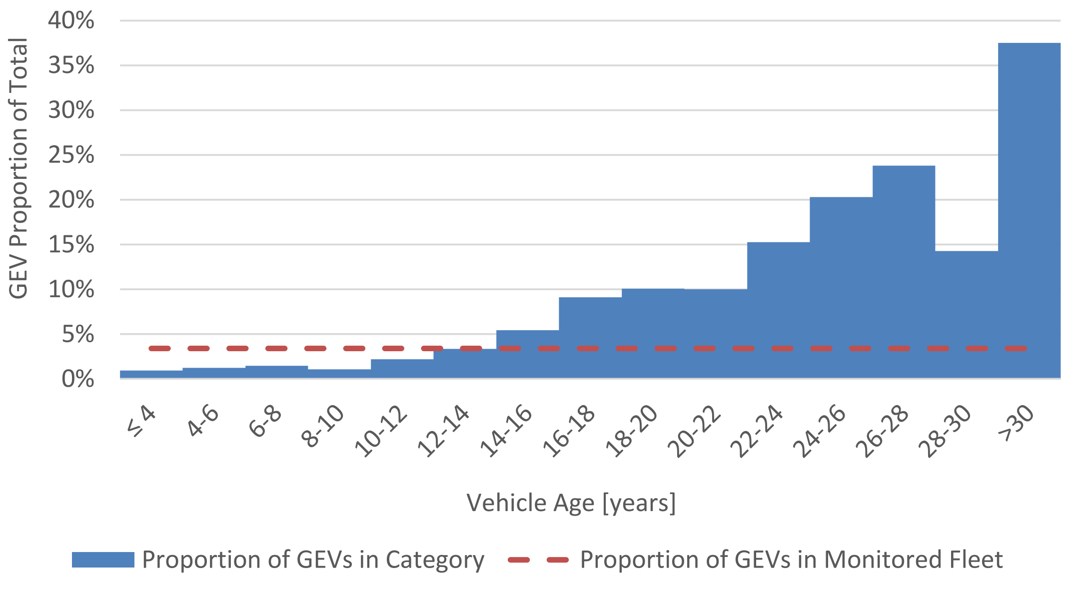


Figure 30: Vehicle Age of HC GEVs

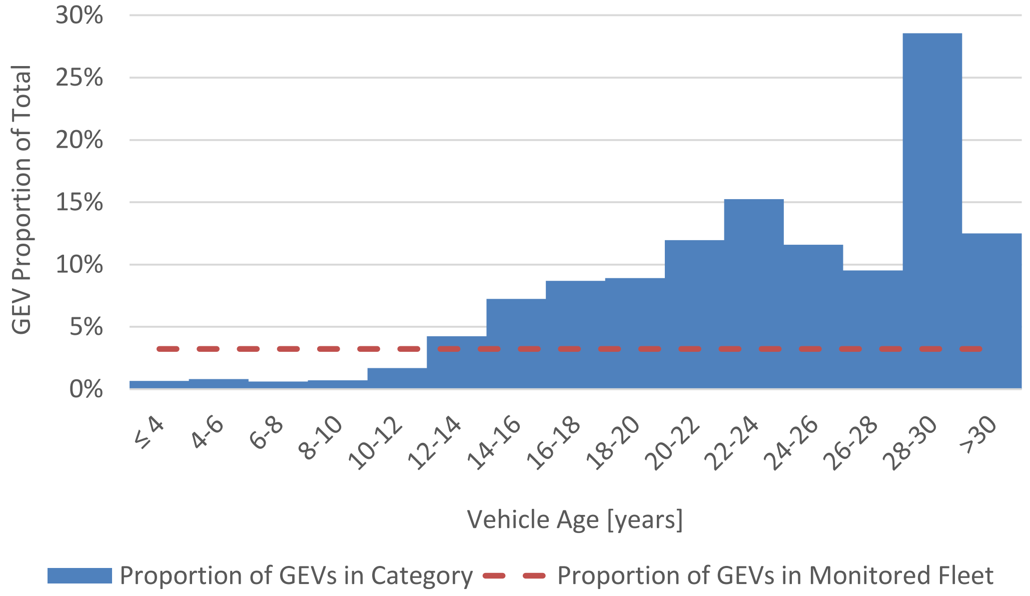


Figure 31: Vehicle Age of NO GEVs

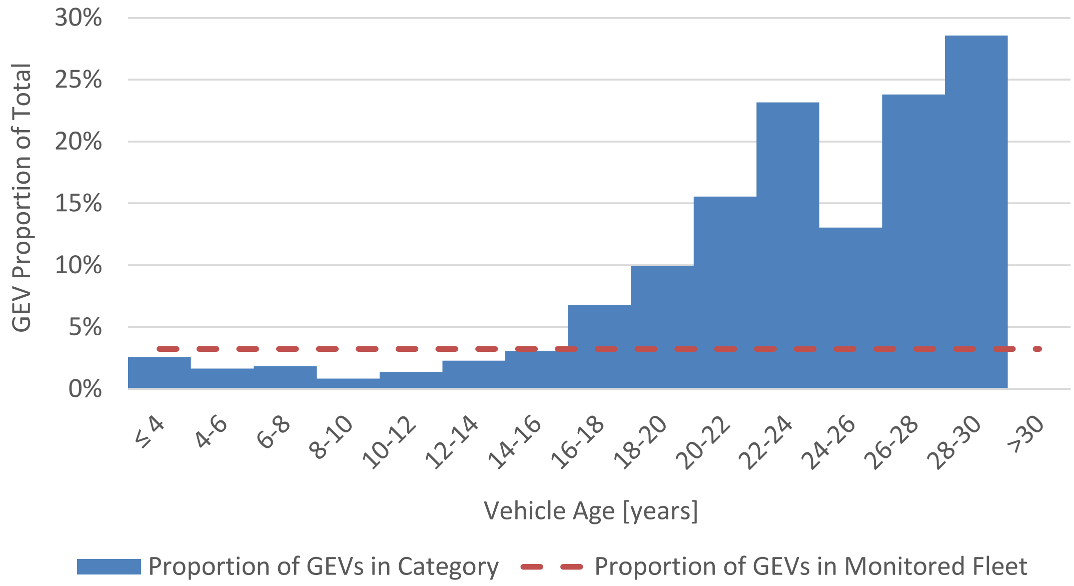


Figure 32: Vehicle Age of PM GEVs

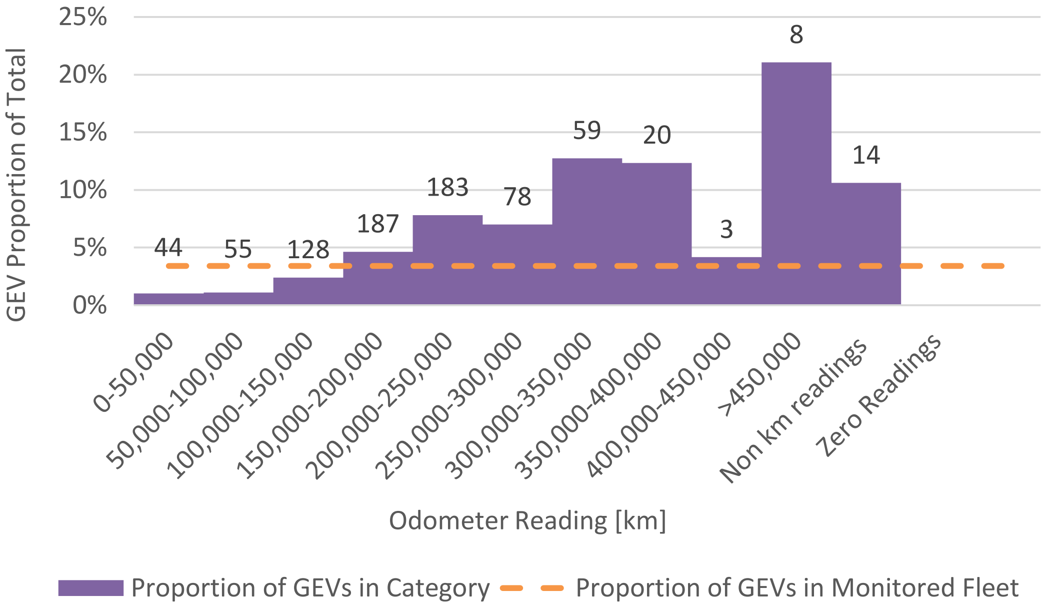


Figure 33: Vehicle odometer reading of HC GEVs
(data labels show number of GEVs in each category)

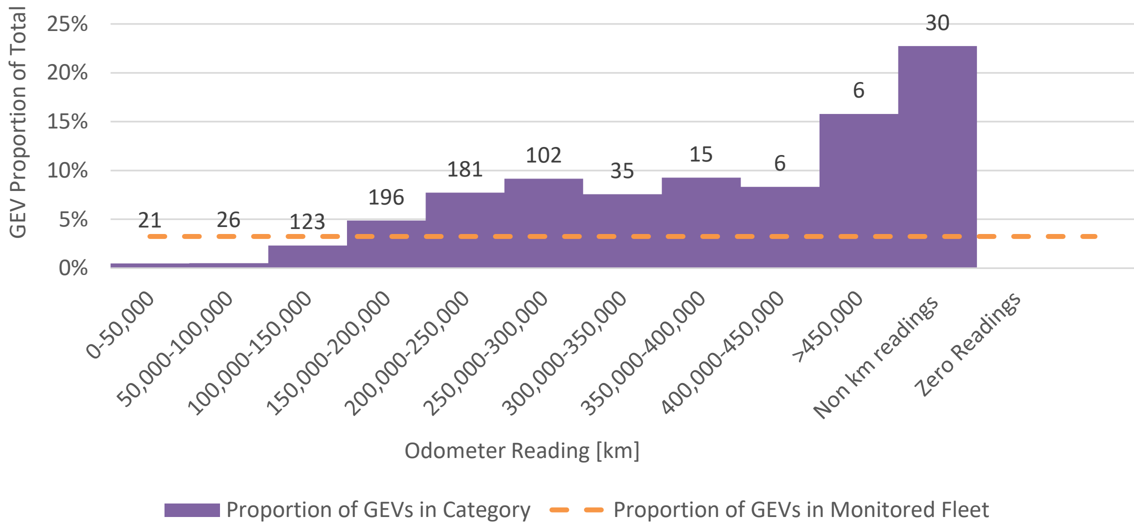


Figure 34: Vehicle odometer reading of NO GEVs
(data labels show number of GEVs in each category)

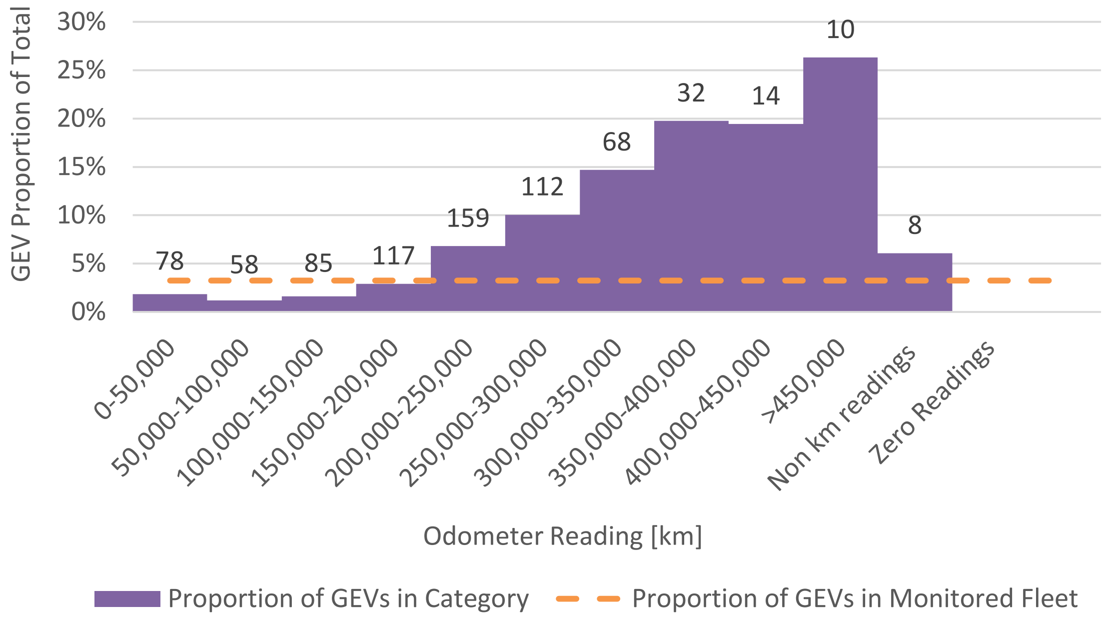


Figure 35: Vehicle odometer reading of PM GEVs
(data labels show number of GEVs in each category)

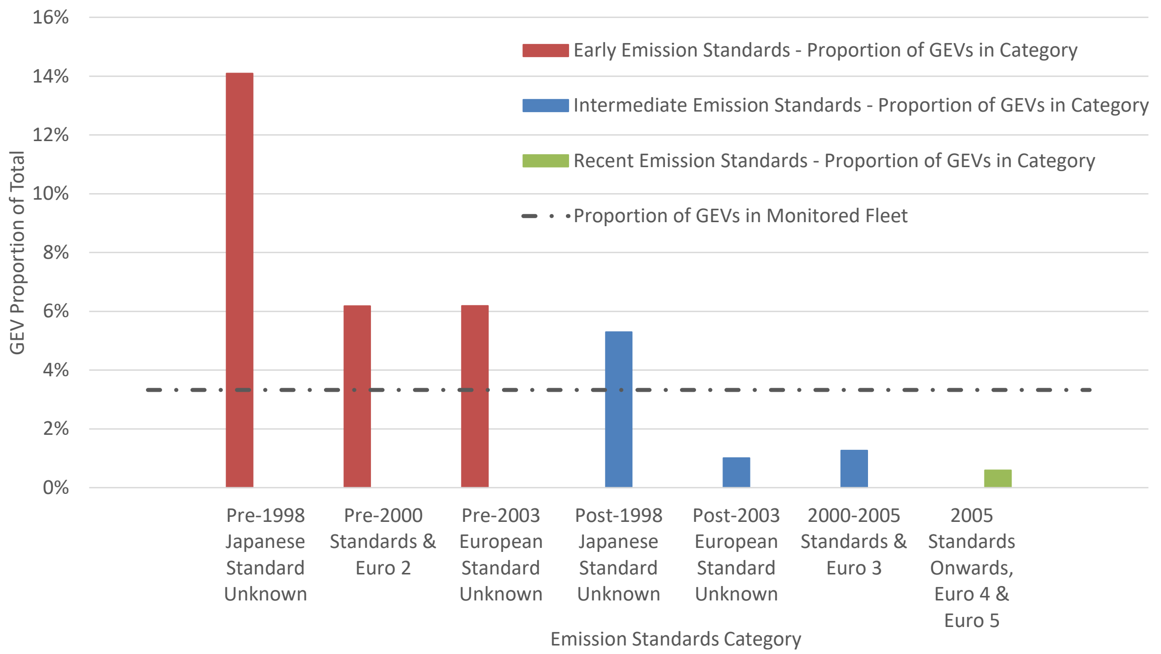


Figure 36: Emission standards categories of petrol-fuelled HC GEVs

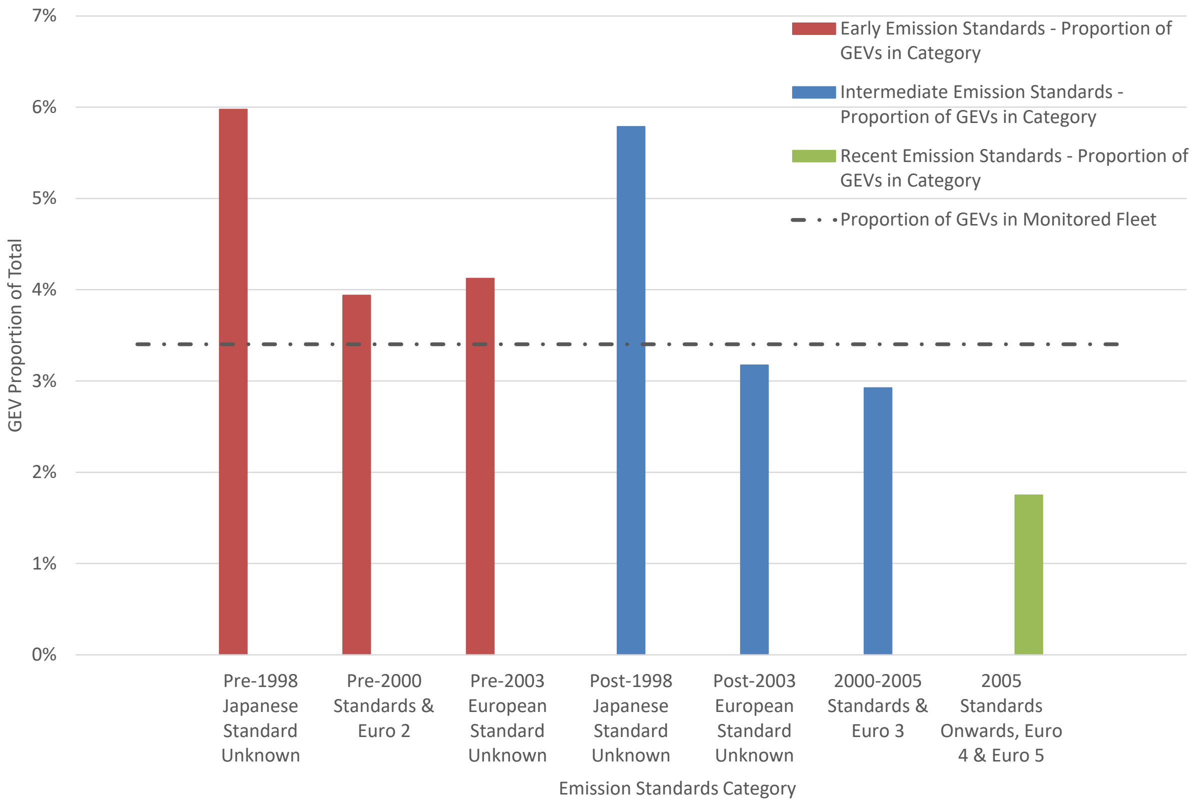


Figure 37: Emission standards categories of diesel-fuelled HC GEVs

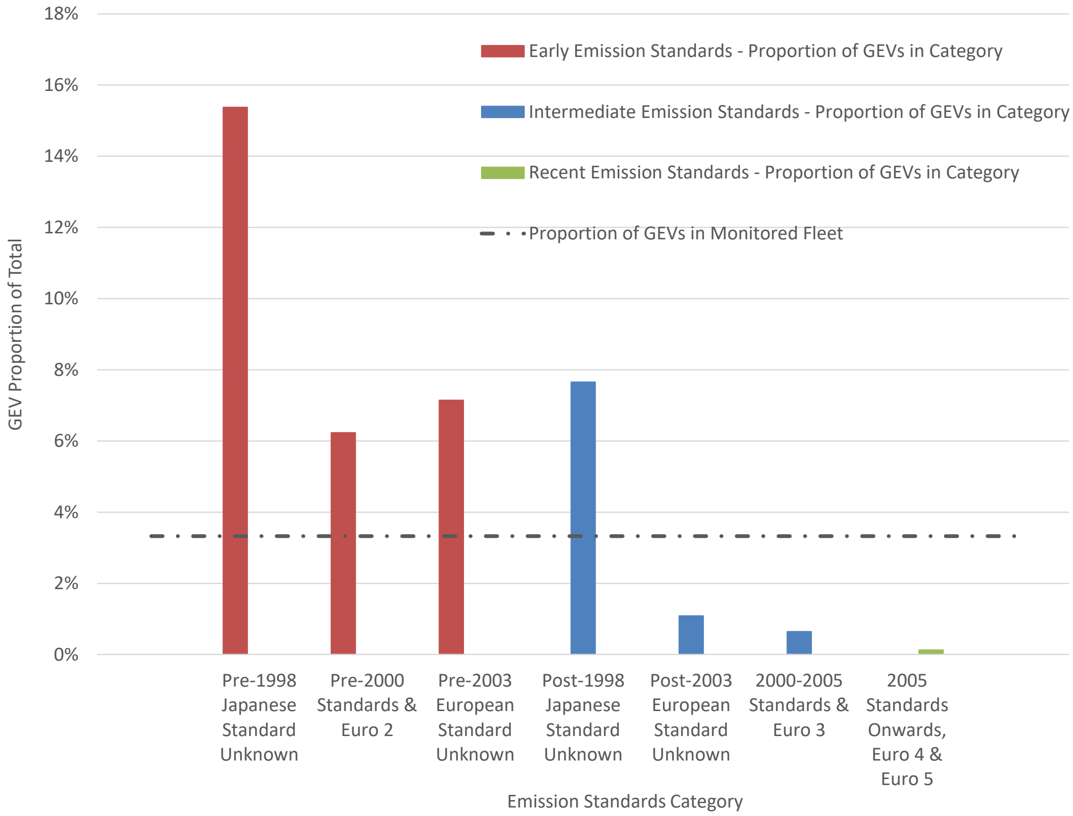


Figure 38: Emission standards categories of petrol-fuelled NO GEVs

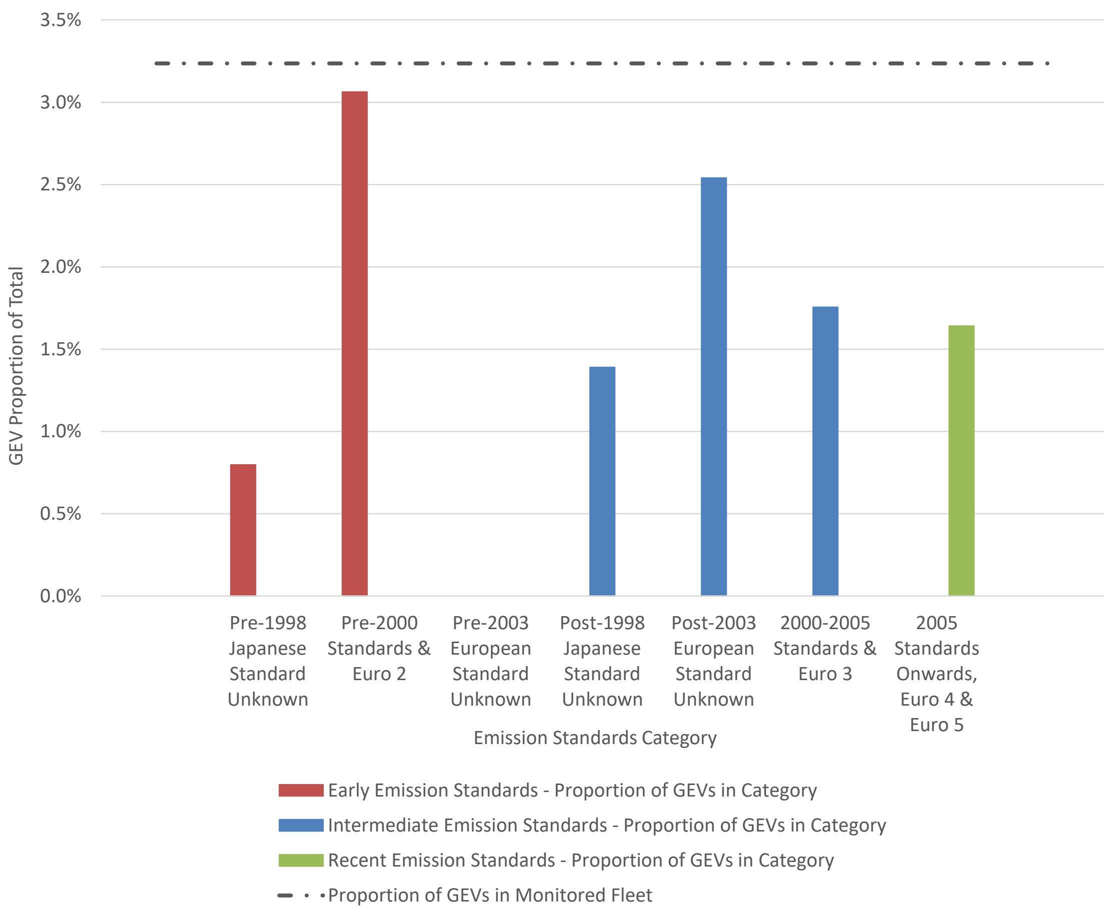


Figure 39: Emission standards categories of diesel-fuelled NO GEVs

Figure 40: Emission standards categories of petrol-fuelled PM GEVs

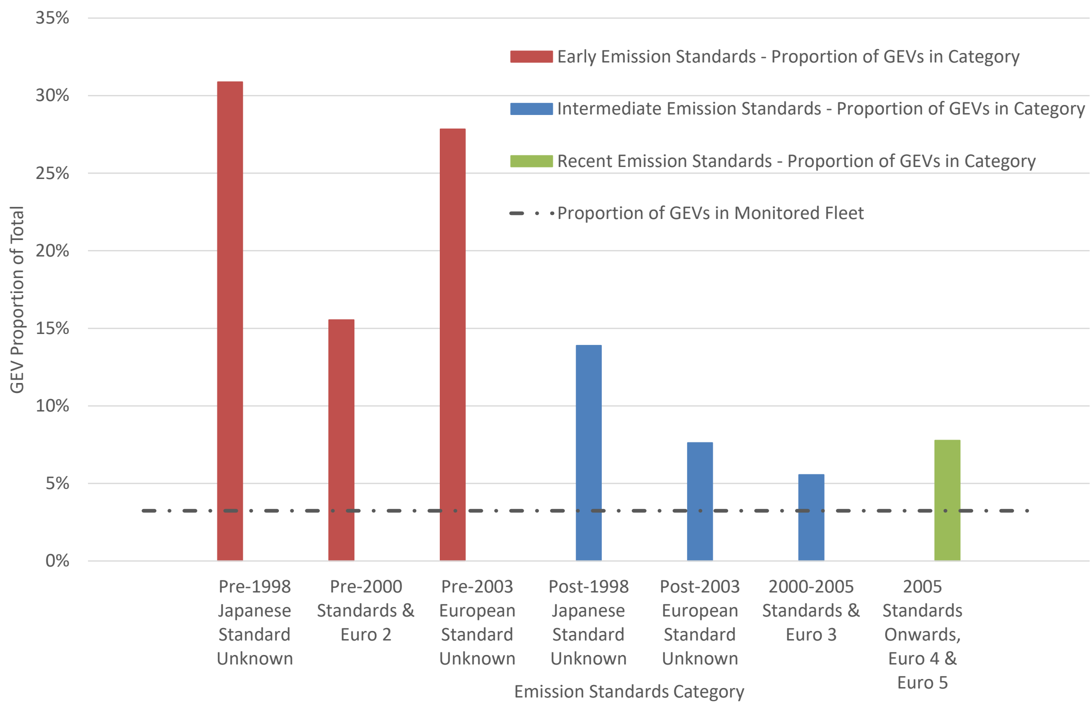


Figure 41: Emission standards categories of diesel-fuelled PM GEVs

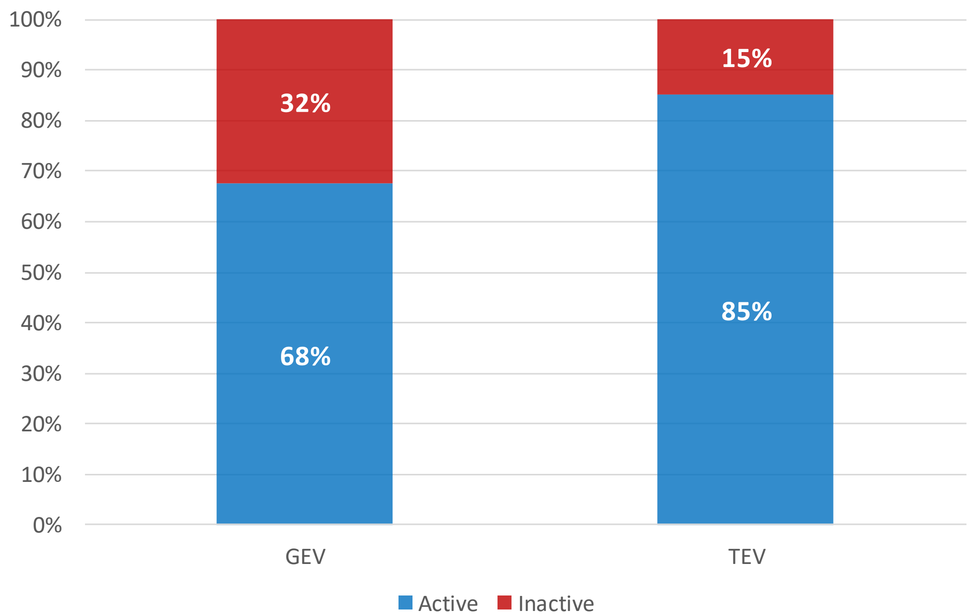


Figure 42: Proportion of HC GEVs and TEVs removed from the fleet during the three-year period between 2015 and 2018

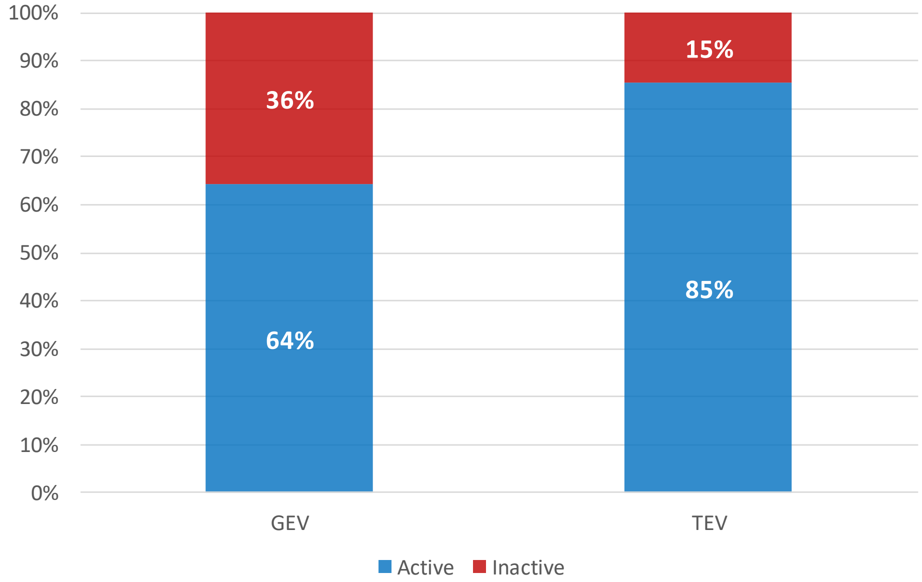


Figure 43: Proportion of NO GEVs and TEVs removed from the fleet during the three-year period between 2015 and 2018

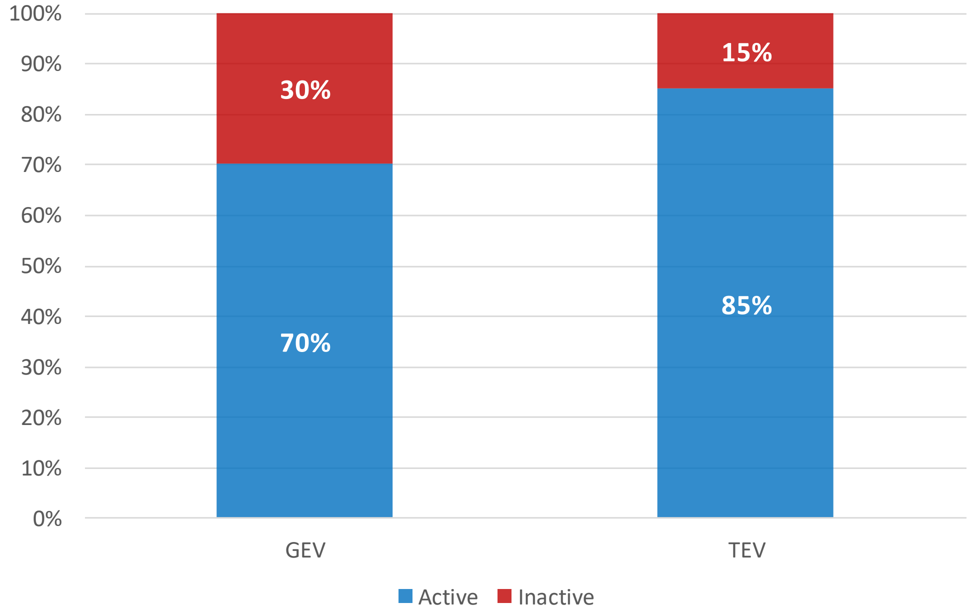


Figure 44: Proportion of PM GEVs and TEVs removed from the fleet during the three-year period between 2015 and 2018

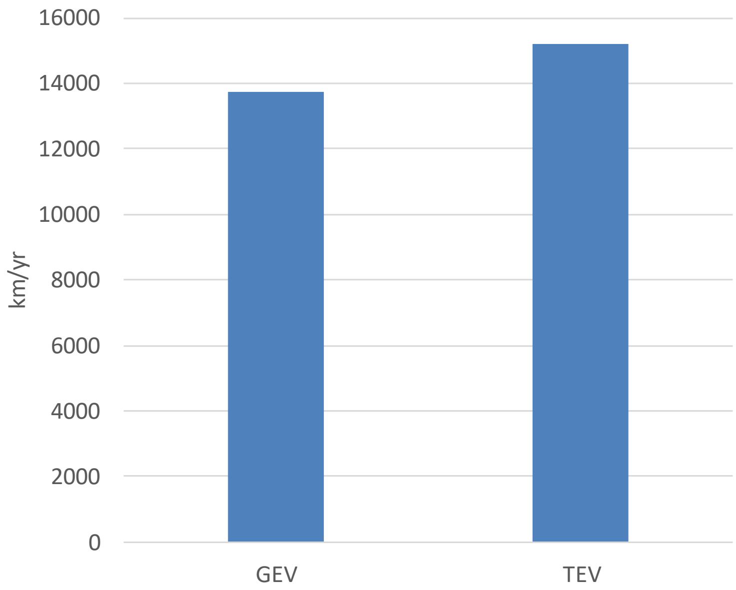


Figure 45: Annual average distance travelled (AADT) for HC GEVs vs HC TEVs

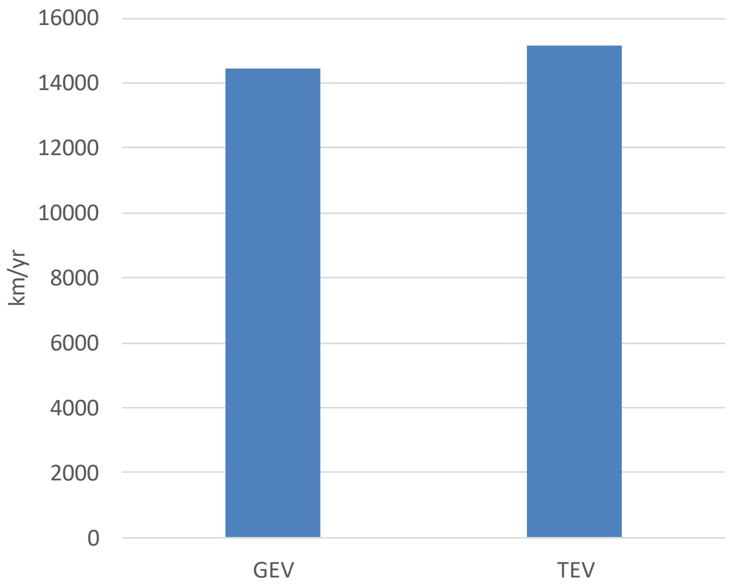


Figure 46: Annual average distance travelled (AADT) for NO GEVs vs NO TEVs

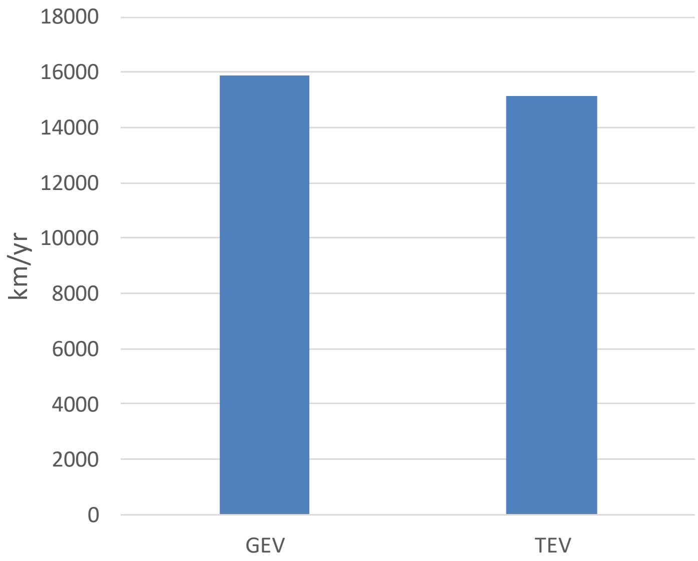


Figure 47: Annual average distance travelled (AADT) for PM GEVs vs PM TEVs

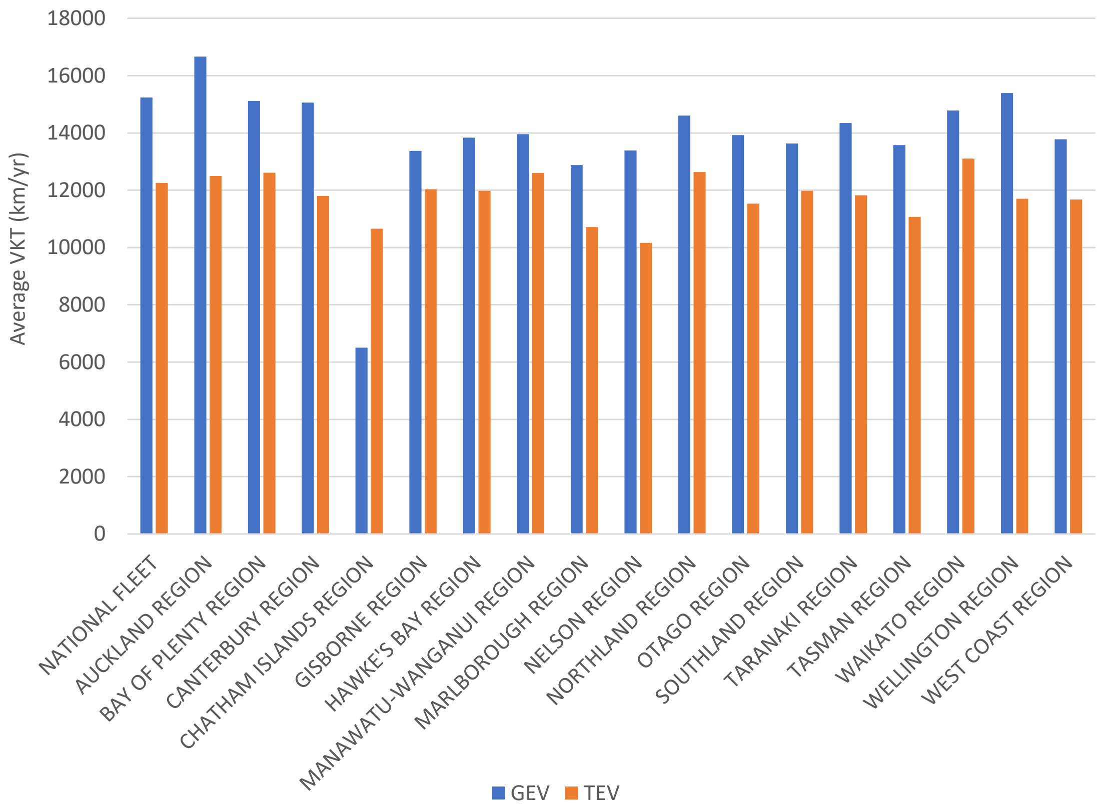


Figure 48: Annual average distance travelled (AADT) of national fleet vehicles by region and modelled as HC GEVs and HC TEVs using criteria derived from RTA

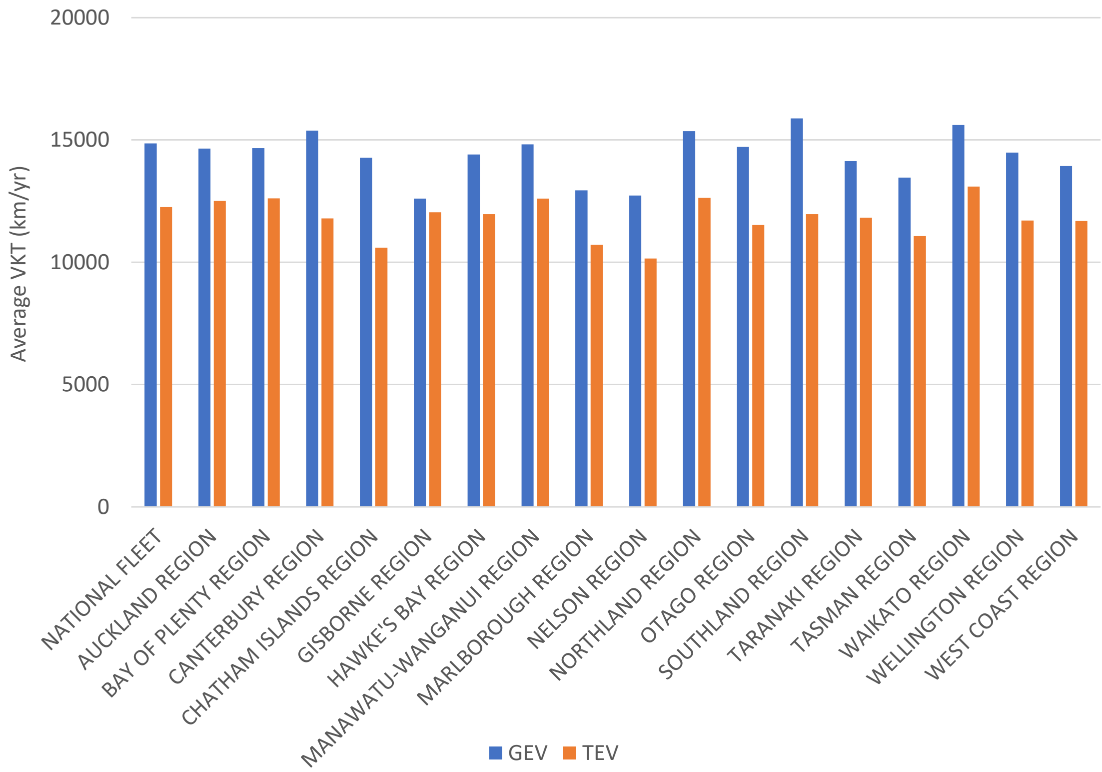


Figure 49: Annual average distance travelled (AADT) of national fleet vehicles by region and modelled as NO GEVs and NO TEVs using criteria derived from RTA

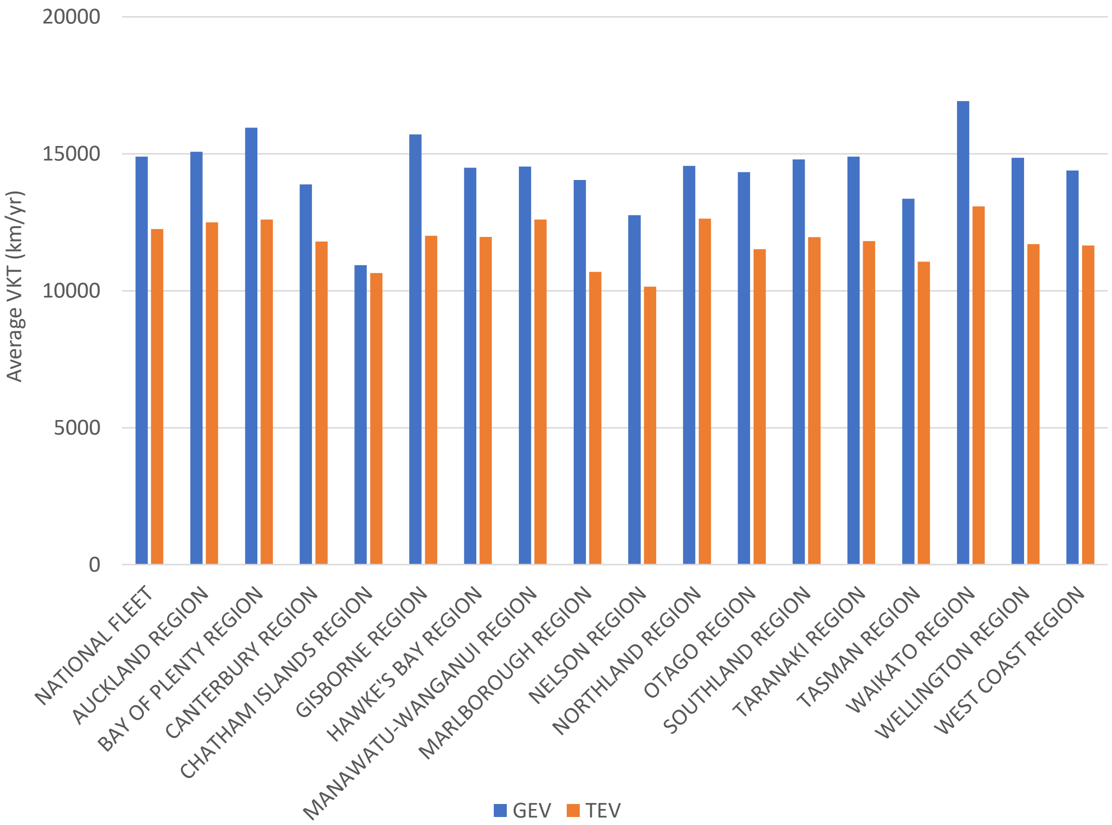


Figure 50: Annual average distance travelled (AADT) of national fleet vehicles by region and modelled as PM GEVs and PM TEVs using criteria derived from RTA

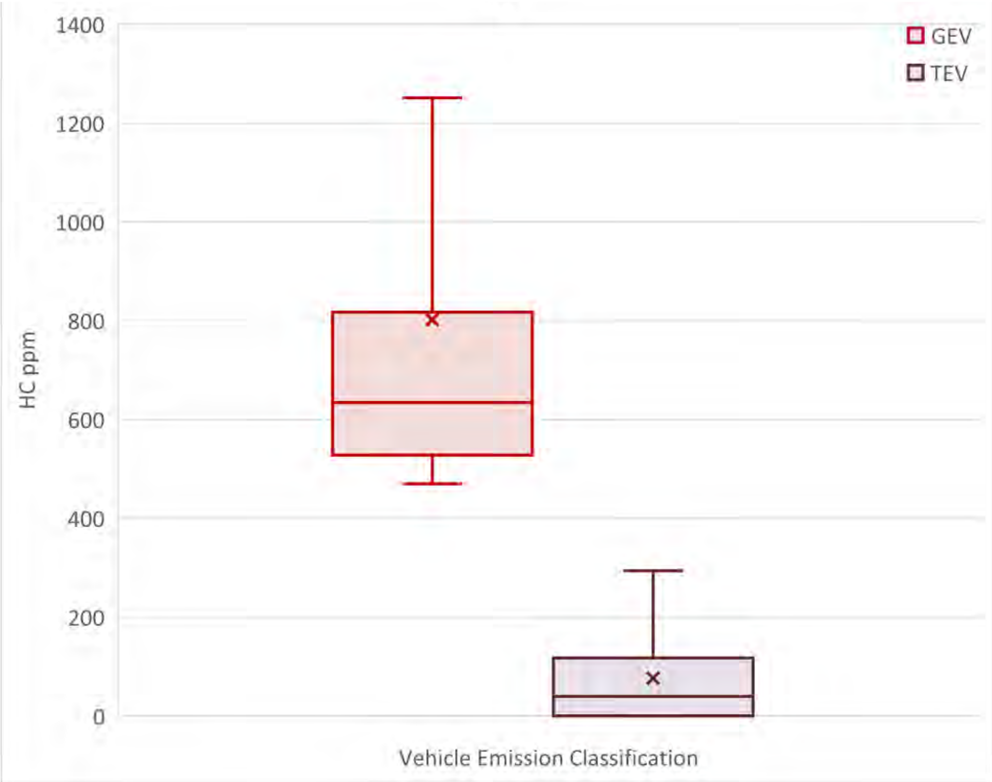


Figure 51: HC emissions from GEVs and TEVs

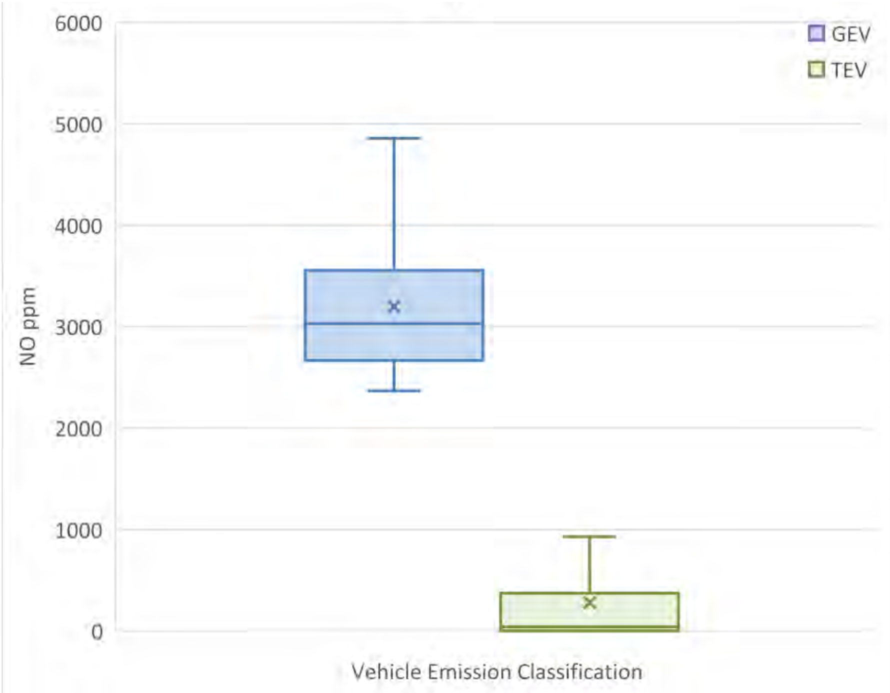


Figure 52: NO emissions from GEVs and TEVs

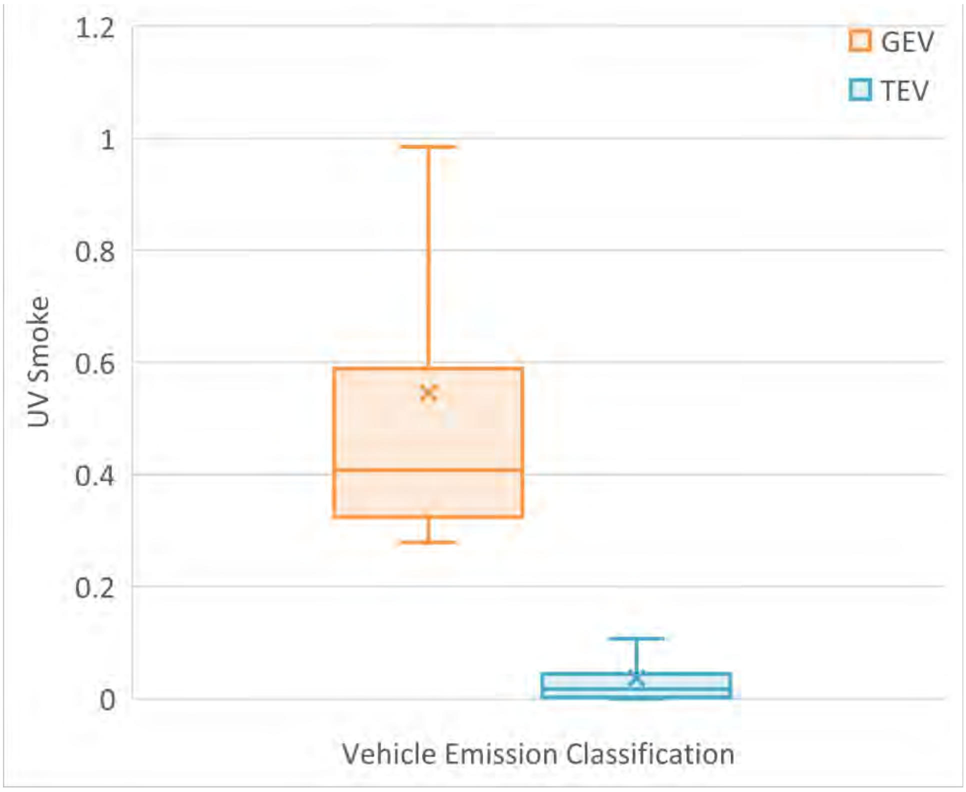


Figure 53: PM emissions from GEVs and TEVs

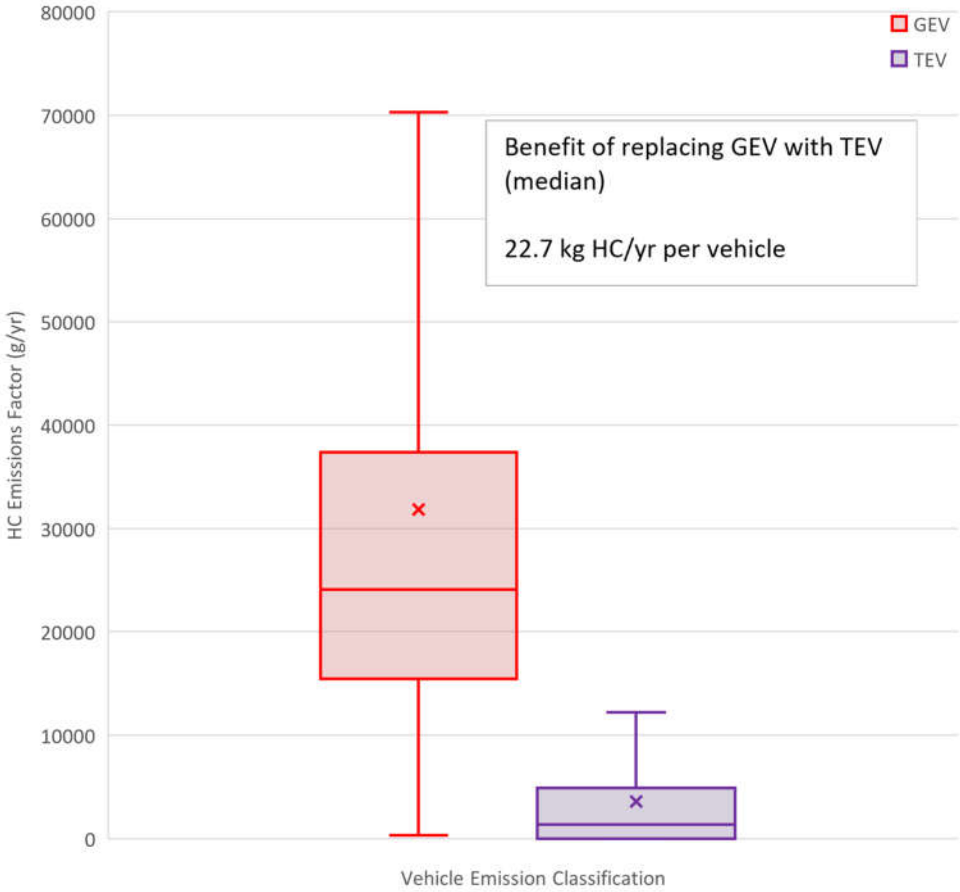


Figure 54: Estimated annual emissions of HC GEVs and HC TEVs

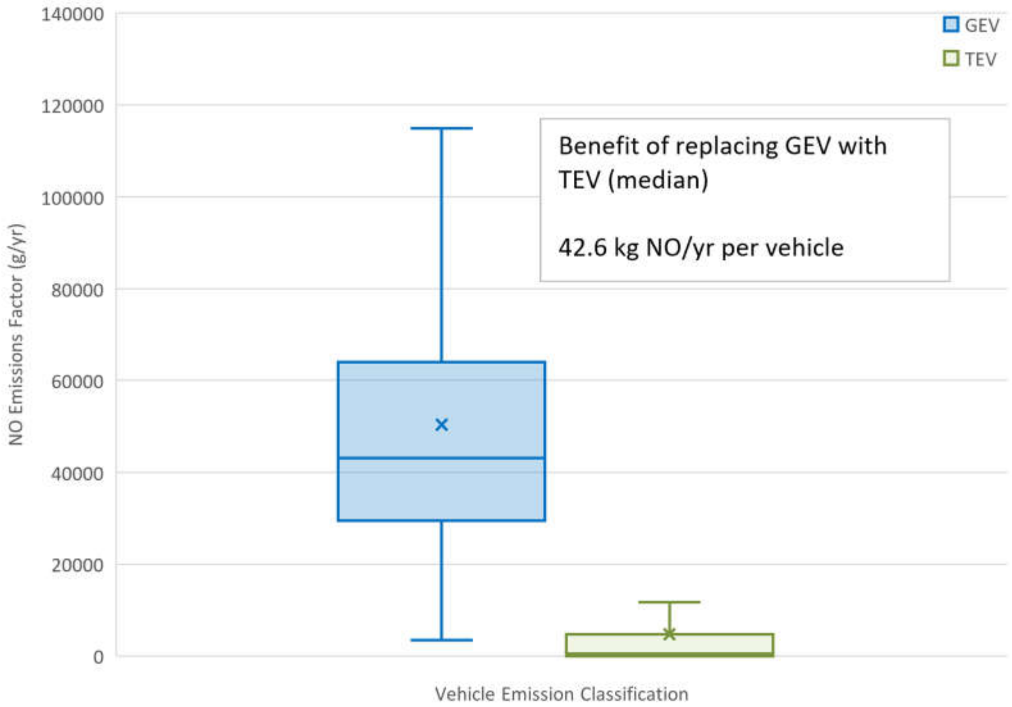


Figure 55: Estimated annual emissions of NO GEVs and NO TEVs

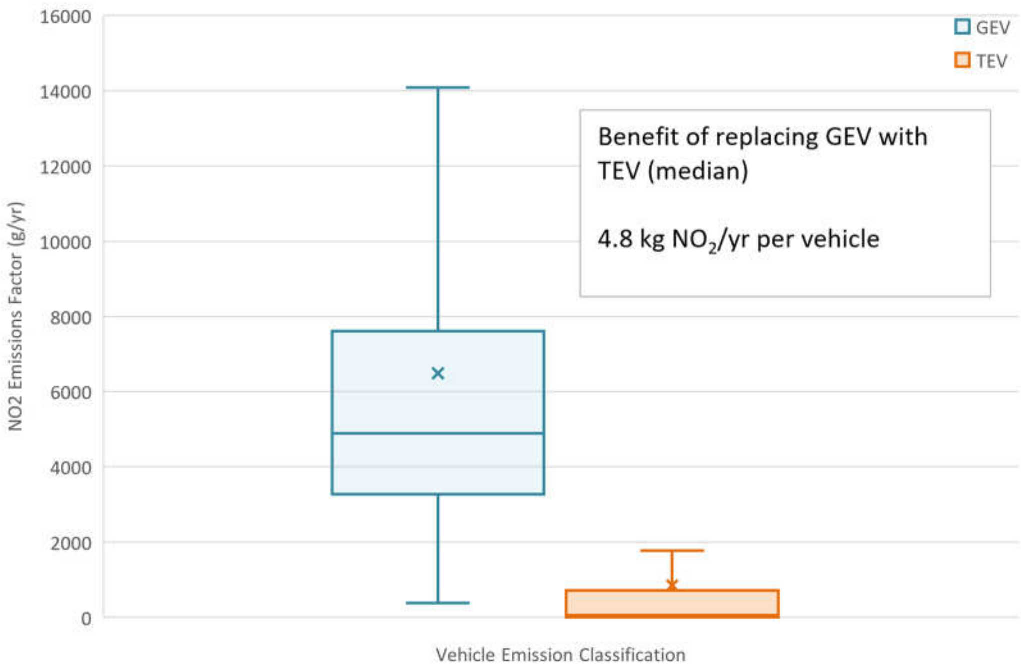


Figure 56: Estimated annual emissions of NO_2 GEVs and NO_2 TEVs

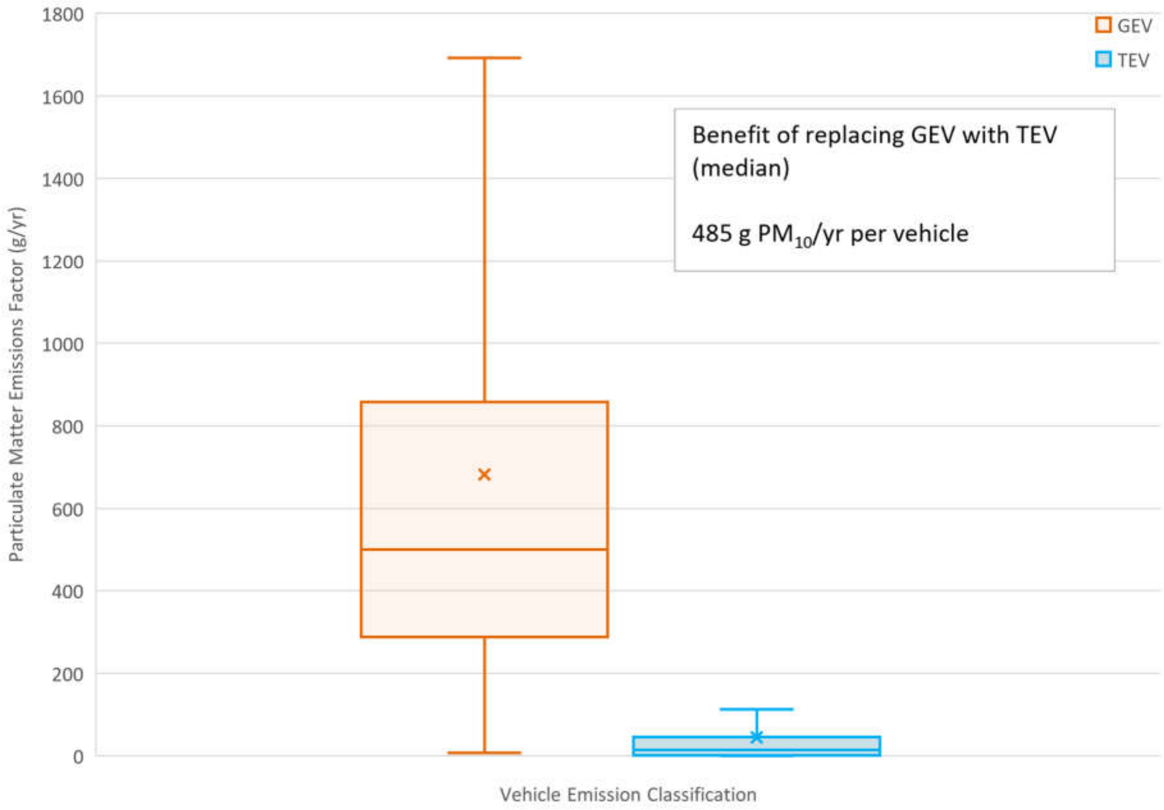


Figure 57: Estimated annual emissions of PM GEVs and PM TEVs

Appendix I

Summary Tables of GEV Characteristics

Table 16: Summary Table of GEV Characteristics

Pollutant	Fleet Count	CO GEVs Count	% CO GEVs in subset	HC GEVs Count	% HC GEVs in subset	NO GEVs Count	% NO GEVs in subset	PM GEVs Count	% PM GEVs in subset
Total Fleet	22,900	762	3.3%	779	3.4%	741	3.2%	741	3.2%
Fuel Type									
Petrol	18909	756	4.0%	649	3.4%	673	3.6%	243	1.3%
Diesel	3974	5	0.1%	129	3.2%	68	1.7%	498	12.5%
LPG	17	1	5.9%	1	5.9%	-	-	-	-
Vehicle Age									
≤ 4 years	4295	25	1%	40	1%	28	1%	111	3%
4-6 years	1513	7	0%	19	1%	12	1%	25	2%
6-8 years	2440	22	1%	36	1%	15	1%	45	2%
8-10 years	3504	44	1%	38	1%	25	1%	29	1%
10-12 years	2893	58	2%	64	2%	49	2%	40	1%
12-14 years	2401	89	4%	81	3%	102	4%	55	2%
14-16 years	1988	102	5%	108	5%	144	7%	61	3%
16-18 years	1844	163	9%	168	9%	160	9%	125	7%
18-20 years	1380	152	11%	139	10%	123	9%	137	10%
20-22 years	360	49	14%	36	10%	43	12%	56	16%
22-24 years	177	22	12%	27	15%	27	15%	41	23%
24-26 years	69	15	22%	14	20%	8	12%	9	13%
26-28 years	21	9	43%	5	24%	2	10%	5	24%
28-30 years	7	0	0%	1	14%	2	29%	2	29%
>30 years	8	5	63%	3	38%	1	13%	0	0%
Odometer Reading									
≤ 50,000 km	4286	18	0%	44	1%	21	0%	78	2%
50-000 – 100,000 km	4946	59	1%	55	1%	26	1%	58	1%
100,000 - 150,000 km	5312	147	3%	128	2%	123	2%	85	2%
150,000 – 200,000 km	4023	205	5%	187	5%	196	5%	117	3%
200,000 – 250,000 km	2341	191	8%	183	8%	181	8%	159	7%
250,000 – 300,000 km	1115	71	6%	78	7%	102	9%	112	10%
300,000 – 350,000 km	463	37	8%	59	13%	35	8%	68	15%
350,000 – 400,000 km	162	9	6%	20	12%	15	9%	32	20%
400,000 – 450,000 km	72	1	1%	3	4%	6	8%	14	19%

Table 16: Summary Table of GEV Characteristics

Pollutant	Fleet Count	CO GEVs Count	% CO GEVs in subset	HC GEVs Count	% HC GEVs in subset	NO GEVs Count	% NO GEVs in subset	PM GEVs Count	% PM GEVs in subset
> 450,000 km	38	1	3%	8	21%	6	16%	10	26%
Non-km readings	132	22	17%	14	11%	30	23%	8	6%
Zero readings	9	0	0%	0	0%	0	0%	0	0%
Emission Standards (Petrol)									
Pre-1998 Japanese Standard Unknown	1405	245	17%	198	14%	216	15%	94	7%
Pre-2000 Standards & Euro 2	2070	164	8%	128	6%	129	6%	50	2%
Pre-2003 European Standards Unknown	1260	58	5%	78	6%	90	7%	31	2%
Post-1998 Japanese Standard Unknown	2248	110	5%	119	5%	172	8%	51	2%
Post-2003 European Standard Unknown	1289	12	1%	13	1%	14	1%	0	0%
2000 – 2005 Standards & Euro 3	7446	142	2%	94	1%	48	1%	17	0.2%
2005 Standards Onwards, Euro 4 & Euro 5	3191	25	1%	19	1%	4	0.1%	0	0%
Emission Standards (Diesel)									
Pre-1998 Japanese Standard Unknown	502	2	0.4%	30	6%	4	1%	155	31%
Pre-2000 Standards & Euro 2	457	1	0.2%	18	4%	14	3%	71	16%
Pre-2003 European Standard Unknown	97	0	0%	4	4%	0	0%	27	28%
Post-1998 Japanese Standard Unknown	432	1	0.2%	25	6%	6	1%	60	14%
Post-2003 European Standard Unknown	315	0	0%	10	3%	8	3%	24	8%
2000 – 2005 Standards & Euro 3	342	0	0%	10	3%	6	2%	19	6%
2005 Standards Onwards, Euro 4 & Euro 5	1829	1	0.1%	32	2%	30	2%	142	8%

Notes:

1. *GEV proportions above the GEV proportion of the total fleet are shaded.*