Development and Evaluation of a Framework for Linking Traffic Simulation and Emission Estimation Models

by
Matthew Muresan

A thesis
presented to the University Of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, ON, Canada, 2015

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.
I understand that my thesis may be made electronically available to the public.

ABSTRACT

The need to understand the effect of policy decisions on environmental indicators is strong. The emergence of new technologies brought about by connected vehicle technologies, which are difficult to evaluate in field settings, means that policies must often be evaluated with software models. In these cases, however, the transportation model and the emissions model are often separate, and multiple different ways to connect these models are possible. Although the estimations provided by each model will vary, each method also differs in terms of the computational time.

This research is motivated by the need to understand the consequences of choosing a particular method to link a traffic and emissions model. Within the literature, aggregated approaches that simply use average speeds and volumes are often selected for their convenience and lower data needs. A number of different scenarios were therefore constructed to compare the estimates of these aggregated approaches to other methods that use disaggregated data, such as the use of individual discrete trajectories, the use of a velocity binning scheme that characterises networks based on their velocity profile or the use of a clustering algorithm developed for this study. This research presents a clustering algorithm that can be used to reduce the computational loads of an emissions estimation process without loss of accuracy.

The results of the analysis highlight the consequences of choosing each approach. Aggregated approaches produce unreliable estimates as they are backed by assumptions that may not be valid in every case. Using individual trajectories creates high computational loads and may not be feasible in all cases. The wealth of data available from a traffic microsimulation mean that using an aggregated approach neglects to utilise the full potential of the model; however, the hybrid approaches presented in this research (clustering and velocity binning) were found to make excellent use of this data while still minimizing computational demands.

Two are better than one;
because they have a good reward for their labour.
For if they fall, the one will lift up his fellow.

— Ecclesiastes 4:9-10

ACKNOWLEDGEMENTS

I would like to express sincere thanks and appreciation to my supervisor, Dr. Liping Fu, who has provided guidance, motivation and substantial help in the development of my research and research ideas. Dr. Fu's guidance and advice has been invaluable in this endeavour, and this thesis would not have been possible without his support.

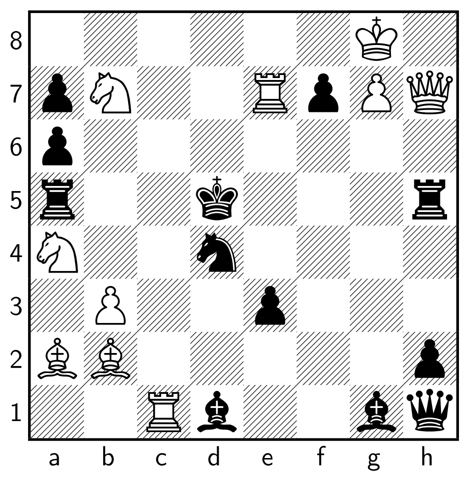
I would also like to thank the members of my family, including my wife Tiffany, parents Radu and Ngamta, siblings Joel, John-Luke, Heman Neginoth and Julia. Their support and understanding throughout my studies has been a strong motivational force.

I am also grateful for the friendship, advice and support of my colleagues at the iTSS lab, especially Kamal Hossain, who helped me become familiar with academic life and the requirements for successful research. I look forward to working together with my colleagues in the iTSS lab every day, and their help has been a significant help to the success of this thesis.

For me, the period surrounding my study has been a time of significant change, and I thank God for all the accomplishments He has helped me complete, both with my academic and non-academic life.

In every thing give thanks:
for this is the will of God in Christ Jesus concerning you.
— 1 Thesselonians 5:18

For my wife, Tiffany, who has always supported me throughout my studies. I love you.



White to move, mate in two

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

AS Average Speed Based

CMEM Comprehensive Modal Emissions Model

CALINE4 CAlifornia LINE Source Dispersion Model, Version 4

CH_4 Methane

CO Carbon Monoxide

CO_2 Carbon Dioxide

COM Component Object Model

CV Connected Vehicle

GUI Graphical User Interface

HC Hydrocarbons

HOV High Occupancy Vehicle

IPCC Intergovernmental Panel on Climate Change

ITS Intelligent Transportation System

MOVES MOtor Vehicle Emissions Simulator

NOx Nitrogen Oxides

PM Particulate Matter

SDS Surveillance Driving Sequence

SHO Source Hours Operating

SHP Source Hours Parked

SPaT Signal Phase and Timing

SQL Structured Query Language

UI User Interface

US EPA United States Environmental Protection Agency

V2I Vehicle-to-Infrastructure

V2V Vehicle-to-Vehicle

VB Velocity Binning

VMT Vehicle Miles Travelled

VSP Vehicle Specific Power

VT Vehicle Trajectory Based

Part I

BACKGROUND AND LITERATURE REVIEW

This part is divided into two chapters. The first chapter gives some background on the major tools used in this project. These tools include VISSIM, a traffic microsimulation model, and the United States Environmental Protection Agency (US EPA)'s MOVES, as well as a few other minor tools. Where relevant, this section highlights specific details about these tools. The literature review explores some of the current literature in the area, including situations where these models have been applied.

1 BACKGROUND

Greenhouse gas emissions have received considerable attention from decision makers in the past few decades. The transportation sector represents a major source of greenhouse gas emissions, which have grown substantially over the past few decades. The Intergovernmental Panel on Climate Change (IPCC) estimates that between 1970 and 2004 transportation emissions have grown by 120 per cent, and without changes in current energy use patterns emissions from transportation will continue to increase at a rate of 2 per cent per year [44]. To better understand these effects, transportation researchers continue to develop and refine the models used to estimate transportation-related emissions with substantial progress made over the past few decades. Consequently, today's transportation researchers have a variety of tools at their disposal that can be used to assess and evaluate policy proposals. Because of the complexity of a transportation system, different modelling platforms are normally used to model each aspect of the transportation system. In the context of emission estimation, this often involves the use of two (or more) models: a transportation model capable of simulating traffic operations on a network and an emissions model capable of estimating vehicle generated emissions. While this approach gives the transportation researcher flexibility, as the emissions model can be used for other purposes (such as field data), it also creates the need to connect the output from one model to the other. The need for this connection creates additional possibilities and considerations, as the manner in which the models are connected affects both the accuracy of the estimate and the computational loads.

1.1 TRANSPORTATION MODELS

Transportation models can be broadly categorised into two major groups: micro-simulation models capable of simulating the behaviour of individual vehicles and macro-simulation models that estimate the behaviour of a network at an aggregate level. The advantages and disadvantages of either type of model is typical of any similarly framed problem: micro-scale models are able to capture and provide more disaggregate features of a system, such as the behaviour of individual vehicles, as well as the larger-scale data that macro-scale models can provide, such as network operating characteristics. A variety of different models are available on today's market, and differ in the technical aspects governing their operation and computational demand. VISSUM and EMME are two examples of macro-scale transportation models while VISSIM, CORSIM and Paramics are examples of micro-scale transportation models. The selection of one alternative over the other is often determined by the needs of the project as well as considerations of cost (including considerations of currently owned software). This research project focusses primarily on developing a framework to connect traffic and emissions models by utilising vehicle trajectories, assuming a micro-scale simulation approach is necessary. For reasons of convenience and cost, VISSIM was selected for use in this study; however, the framework explored in this research could easily be adapted to fit other models as well.

1.2 POLLUTION MODELS

The need to accurately estimate emissions before making policy decisions has led to the development of a number of different models and approaches. As is the case for transportation models, emissions models can also be compared in terms of macro-scale and micro-scale models. One of the most prominent modelling platforms available for emissions today is the US EPA's MOVES model. MOVES has been developed by the US EPA through substantial research and also includes databases containing default values that can be used to quickly develop robust estimates. As an emissions model, MOVES combines elements of macro-scale models and micro-scale models, and can be used to estimate emissions at national and county levels (similar to traditional emissions factor models) and at smaller project scales. Prior to the development of MOVES, the US EPA's main model for the estimation of transportation emissions was the MOBILE model [59], which is an example of a macro-scale model that estimates emissions without giving consideration to the detailed aspects of a transportation system. These models are often criticized for their inaccurate modelling of driver behaviour, as they are based on average driving characteristics [5].

In contrast, micro-scale models such as MOVES or the Comprehensive Modal Emissions Model (CMEM) [57] are capable of estimating emissions on a per-second basis. Unlike macro-scale models, micro-scale models can be used to estimate the effect of subtle changes in the operation of transportation system, such as driver behaviour or the operation of individual intersections. Of course, this increased ability comes with a price: an increase in the data required for a successful and accurate estimation of emissions and a decrease in the size of network that can be successfully modelled with today's computer hardware. In the real world it is often difficult to obtain highly detailed information on a vehicle's operation, and therefore many models (such as MOVES) also include options to make an estimation with more aggregated data. For example, MOVES is able to use extremely disaggregate data, such as a record of a vehicle's speed on a per-second basis, or more aggregated data, such as a vehicle's average speed, to arrive at an emissions estimate. Aggregated approaches can be very useful when analysing and evaluating situations that are not suited to an highly macro-scaled approach (such as emissions factor models) or to highly micro-scaled approaches (such as using accurate vehicle position data).

1.3 MODEL INTEGRATION

While the data requirements for analyses of real-world situations can be very high, when coupled with a traffic-micro simulation model, a micro-scale pollution model can be used as a powerful tool. In the literature these models can then be used as part of a "bottom-up" approach to emissions modelling, where scenarios or policy alternatives are evaluated at a micro-scale level before being considered in a region-wide macro-scale model. This coupling can be achieved in a variety of different ways, and the approach selected varies depending on the needs of the situation. Micro-simulation models are able to provide substantial data, but the scale of data available can create substantial computational burdens, especially on large or complex networks. The result is a trade-off where analysing more aggregate data simplifies the analysis and reduces the computational burden while analysing more disaggregate data increases the accuracy of the estimations but also increases the complexity of the analysis. Although individual modelling platforms work differently, the basic principles of their operation are very similar, and methods developed to link particular combinations of models can often be extended to incorporate other models as well.

1.4 RESEARCH OBJECTIVES

Broadly speaking there are three main methods to integrate transportation and pollution models so that the results of a traffic simulation can be used in an emission model: a disaggregate approach that uses the trajectories of individual vehicles, an aggregate approach that uses average speeds and volumes on roadways, and hybrid approaches that perform some aggregation. Aggregate approaches reduce computational burdens and accuracy while disaggregate methods increase computational burdens and accuracy. This research therefore seeks to investigate and quantify these trade-offs. In addition to this, this research has also sought to develop hybrid approaches that reduce computational burdens while minimising the impact on the accuracy of emission estimates.

To accomplish these objectives a reliable integration method must first be developed and tested using common approaches demonstrated in the literature. After developing sufficient background, the proposed hybrid method must also be developed and incorporated into an integration tool that can be used to reliably link the two chosen models. This research uses VISSIM due to its availability and MOVES as it is one of the newest models available. After an integration tool has been successfully created, a number of scenarios must then be devised to quantitatively assess the performance differences of these methods.

A link between a transportation model and an emission model can be a very useful tool when evaluating real-world policy proposals. Within the field of transportation, connected vehicle technologies have received considerable attention. While many of the benefits of connected vehicle technologies are safety related, the increased information and connectivity available also has the potential to effect positive changes in the environmental impact of vehicles. This research therefore also seeks to apply linked transportation and emissions models to evaluate the environmental impact of an ECO-driving system that uses connected vehicle technologies in a simulated environment.

1.5 THESIS ORGANISATION

This thesis is divided into three parts consisting of eight chapters.

The first part contains introductory material and includes two chapters: this chapter (Chapter 1) introduces the research problem and gives some basic background on the concepts discussed in the remainder of this work while Chapter 2 gives some details on relevant operational characteristics of MOVES as well as previous studies that effectively link a micro-scale traffic and pollution model. This chapter also discusses current proposals of connected vehicle technologies and studies on ECO-driving systems. Finally Section 2.7 provides detail on data-clustering methods, including the general operating principle and previous applications of the algorithm chosen for application in this thesis.

The second part details the methodologies and research process used and one chapter. In this part, Chapter 3 gives an overview of the process developed to link VISSIM and MOVES and also includes detail on the tool developed to evaluate an ECO-driving system.

Finally, the third part gives the results and provides a discussion on their implications. In this part, Chapter 4 provides the results of the evaluation comparing the different integration methods, and includes some discussion on the appropriateness of each method. Chapter 5 provides the results of the ECO-driving evaluation and Chapter 6 provides a final discussion of the work done in this research and areas needing additional attention.

2 LITERATURE REVIEW

Air pollution, and specifically air emissions from transportation vehicles, is an area that has seen much research in the past few decades. Indeed, increased public awareness and concern for environmental issues means that controlling and understanding the effects of pollution has become a significant focus. In the transportation sector, understanding the scale and effects of pollution is often done through the use of various models. Before developing a framework to link these models, it is important to review the process by which the models being used operate on as well as previous applications of this concept in the literature.

2.1 EMISSIONS MODELLING

As mentioned briefly in Section 1.2, emissions models can be broadly classified into two groups depending on the scale of their analysis: macroscopic and microscopic. The selection of a model from either of these two categories depends on research needs, with macroscopic models often suitable for large-scale analyses such as national or province-wide emission estimates and microscopic models often suitable for smaller subsets of a larger network. Macro-scale models such as the US EPA's MOBILE model often use an emissions factor approach when estimating emissions. In this approach, the model estimates an *emission factor* in either mass of pollutant per unit time or mass of pollutant per distance travelled. These values can then be combined with estimates of network parameters (such as total Vehicle Miles Travelled (VMT)) to develop vehicle emissions inventories for a particular network [61]. As one of the first approaches developed for emission modelling, emission factor models such as MOBILE are the product of extensive research and data.

2.2 MICROSCOPIC EMISSION MODELS

In contrast to macroscopic emissions modelling, microscopic models were developed to address a number of limitations of the existing emissions factor approach. Emissions factor models are unable to accurately characterise driving behaviour (as they are based on average driving characteristics). These models operate on an aggregate level, and ignore the effects of individual driver behaviour, such as individual acceleration rate patterns. These models are often also based on predetermined driving cycles (e.g. the Federal Test Procedure) which form base emission rates that are then adjusted through correction factors for aspects such as speed, fuel type, temperature, etc [5]. While capable of providing regional level estimates, these models are not suitable for use in analysing the complex effects of micro-scale policy proposals, such as ramp metering, signal timing and coordination changes, and emerging Intelligent Transportation System (ITS) technologies, and as such other approaches have been developed in the past to address these challenges.

2.2.1 *Modal Emissions Modelling*

Unlike emission factor models, modal emissions modelling is a micro-scale approach to the emission estimate process that subdivides a vehicle's operating pattern into *modes*. In these models a vehicle's operating *mode* describes elements of its operational behaviour, such as cruising, idling, accelerating, decelerating. [5]. The number of modes considered varies from model to model, and often include further subdivision on the basis of speed or acceleration rate. Many of the micro-scale emission models developed and employed in the research community, such as MOVES [59] and CMEM [57], are modal emissions models. Research into the emission patterns of vehicles operating in each of these modal states has been conducted in the past, and these results have been used by various agencies to develop microscopic modal emissions models.

2.2.1.1 *The Autmobile Exhaust Emission Analysis Model*

The first modal emissions model developed was the "Automobile Exhaust Emission Modal Analysis Model". Developed in 1974 by the US EPA [8], this model allowed the estimation of vehicle emissions over an arbitrary driving sequence. The model also allowed for predictions of fuel economy through the use of a carbon-balance equation of the resulting emissions [33]. Emissions data from over one thousand light duty vehicles (manufactured between 1957 and 1971) were used to build the model [8]. The model inputs were based on the Surveillance Driving Sequence (SDS), which includes 37 discrete driving modes (derived from five steady-state speeds). These 37 modes were expanded into a continuous function through regression that allows estimation of emissions for any combination of speed and acceleration, which can then be integrated over a vehicle's drive cycle to arrive at an estimate of its emissions [33].

2.2.1.2 *CALINE4*

Subsequent models were built based on the framework established in the US EPA's and were more computationally efficient, simpler, and more straightforward. One of the earliest examples of this is the CAlifornia LINE Source Dispersion Model, Version 4 (CALINE4) model. CALINE4 was developed by the California Department of Transportation in the 1980's and can be used for dispersion modelling of Carbon Monoxide (CO) (which it is still used for today), Particulate Matter (PM) and Nitrogen Oxides (NOx) [11]. Although the CALINE4 final estimates of the model are derived from a Gaussian diffusion based dispersion model, the inputs for this final stage are provided by a modal emissions model that was patterned after an earlier model developed by the Colorado Department of Highways called CDOH [8]. The CALINE4 model defined four operating states for vehicles: deceleration, idling, acceleration, and cruising. The model uses the time-in-mode approach to develop the emissions profiles of all the vehicles in the modelled network. The modal emissions factors for these states were derived from SDS data restricted to the California region and included around 80 vehicles. Unlike previous approaches, which average the emissions of vehicles for particular operating conditions, the underlying model factors were developed using a disaggregated approach.

2.2.1.3 *CMEM*

In addition to these models, the modal approach has been used to develop other models, such as VEHSIME [39] and, more recently, CMEM. CMEM was originally developed by a team at the University of California, Riverside in the 1990's to meet the demand for a micro-scale emission model [57]. As was the case with the CALINE4 model, the CMEM project was conceived with the aim of developing a *modal* emissions model for light-duty vehicles such as passenger cars and small trucks that could interface with existing transportation models and datasets [6]. CMEM was also developed to address issues present in other modal emissions models [6]. Models such as the US EPA's described in Section 2.2.1.1 depend solely on a vehicle's speed and acceleration and cannot adequately consider other variables such as road grade or weather. CMEM was therefore designed to use a physical modelling approach that subdivides emission processes into components that correspond to the underlying physical and chemical processes for each operating state. Under this approach, separate models were developed for each of the engine/emissions technologies modelled, including different engine combinations (e.g. spark ignition, diesel), fuel delivery systems (e.g. fuel injection), emission control systems, and catalyst usage (e.g. no catalyst, oxidation catalyst). The physical models are then combined with vehicle operating parameters such as temperature, air density, road grade, air conditioning use, etc to arrive at estimates of emission rates [6].

The CMEM model gives the researcher the ability to specify a number of different network parameters. The vehicle characteristics described previously form the basics of the fleet distribution in the modelled network, and can be specified directly (or a default distribution can be used). At its highest temporal resolution, second-by-second information on vehicles can be provided (e.g. per-second speed, grade, etc) from which emissions estimates are generated [6]. CMEM was designed to interface with existing traffic micro-simulation models easily. Software plug-ins to interface the model with popular micro-simulation platforms such as PARAMICS have also been developed [13].

The CMEM model has been validated by comparing its emission estimates to both field measurements and emission measurements of other models established at the time (e.g. MOBILE). As part of the model's validation, Barth et al. compared the emissions predictions of the CMEM model to independently generated emissions from vehicles following pre-defined driving schedules. and showed that the the CMEM model made emission estimates that were generally within the range of the tested vehicles [7]. The model was found to under-estimate emissions for some high-emitting vehicles, however. When compared to other established models (such as the US EPA's MOBILE model) it was found that it provided similar emission estimates at low to medium speeds. However, the models deviated from each other at very low and very high speeds. At very low speeds CMEM generally provided estimates that were lower than the models compared. At very high speeds, CMEM provided estimates that were higher for some pollutants (e.g. Hydrocarbons (HC)) but lower for other pollutants (e.g. NOx) [7].

2.2.2 *Limitations of a Micro-scale Approach*

Micro-scale models in general suffer from a number of limitations. Although these models can be integrated with micro-scale traffic models (such as VISSIM, CORSIM, etc), there are limits in their ability to estimate larger, regional level emissions. In the case of emissions models, substantial inputs on the operating characteristics of individual vehicles is required, and this data may not be available or may be difficult to obtain from a traffic simulation model. The complexity of these models mean they do not scale linearly with network size and are therefore only viable for small-scale analyses [6].

2.3 THE MOTOR VEHICLE EMISSIONS SIMULATOR (MOVES)

MOVES is a software package produced and made available by the US EPA [60]. The software is available for download from the US EPA's website free of charge and can be used to model a variety of emissions from mobile sources. MOVES replaces the US EPA's former tool for estimating emissions factors from highway sources, namely, MOBILE. MOBILE was originally developed in 1978 and was one of the first models to generate highway vehicle emission factors. The MOBILE model was continuously updated, as technology progressed, forming the basis for MOVES, which superseded it in 2010 [60]. Fundamentally MOBILE is an emissions inventory model and operates at a macro level only. While retaining the macro level modelling capabilities of MOBILE, MOVES incorporates new features to simulate emissions at smaller scales as well. Modelling in MOVES is done at three possible scales: National, County and Project. At the project level, MOVES is able to use detailed information about vehicle trajectories and operating states to estimate the emissions of a micro-level network. This allows it to be used in tandem with a traffic model and forms the basis of the work done for this research project. The fundamental principles behind the software's operation, implementation and its uses is publicly available from the US EPA's website. The following sections will serve to highlight critical aspects of the model's operation, limitations and capabilities. The specific operating aspects of MOVES are important to understand, at least at a broad level, as the accuracy and validity of any method used to process the data before it is input into MOVES is affected by its operating behaviour. This is particularly important in the context of this project, which seeks to compare and contrast the different ways of connecting MOVES to a transportation microsimulation model.

MOVES is derived from the large body of work that the US EPA has done on previous models and the substantial data it has collected on emission rates across the country. This data is included with MOVES as a database that can provide default values for many aspects of the model on a county-by-county basis. Readers interested in learning more about the specific operating principles of MOVES are encouraged to read the US EPA's Design and Implementation Plan [60].

Combustion Products	Hydrocarbon Evaporation	Other
Running Exhaust Start Exhaust Crankcase	Diurnal Hot Soak Resting Loss Running Loss Refueling	Brake Wear Tire Wear

(a) The emissions processes used in MOVES.

Emission Process	Operating Mode Parameter(s)
Running Exhaust, Brake Wear, Tire Wear	Average Speed and Vehicle Specific Power (VSP)
Start Exhaust, Hot Soak	Soak Time
Diurnal	Tank Pressure
Running Loss	Time Since Start

(b) Total Activity by Emissions Process.

Figure 1: Emissions Processes and Their Operation (from [59])

2.3.1 *Model Design Philosophy*

The MOVES design is modular, general purpose, data driven, easy-to-use, and high performance [59]. The general design framework applied by the US EPA in MOVES progresses through the following four stages.

2.3.1.1 *Calculate the Total Activity for a Given Emission Process*

In MOVES, an emission process is defined as a unique emission pathway [59]. Each emission process is handled separately within the model and may not always produce output on the same pollutant as another process. In essence, each emission process can be likened to a sub-model, having its own set of inputs (which may be the same as another emissions process) and outputs. A list of the emissions processes used in MOVES is shown in Figure 1a.

The total activity is defined as the product of a population and a per-source activity. An activity's definition depends on the emission process under consideration. Most activities are characterised on the basis of source-time, which is further subdivided into Source Hours Operating (SHO) or Source Hours Parked (SHP). This selection stands in contrast to other possible schemes, such as the more common VMT, as it allows emissions that do not vary over distance to be aggregated together (e.g. idle time). Of course, given an average speed, it is possible to convert between a distance-based unit system or a time-based one; thus, the systems are interchangeable from the user's perspective. MOVES permits a user to specify a variety of different units in both the input and output stage, but is designed to characterize each process at the software level based on source-time Agency [59]. Figure 1 details the activity basis for each emissions process.

2.3.1.2 *Distribute the Total Activity into Source and Operating Mode Bins*

The concept of operating modes and source bins is fundamental to the MOVES model. From a software perspective, the concept of binning represents a way to

Emission Process	Total Activity Basis
Running Exhaust, Brake Wear, Tire Wear, Running Loss, Crankcase	Source Hours Operating (SHO)
Start Exhaust	Number of Starts
Diurnal, Hot Soak	Source Hours Parked (SHP)
Resting Loss	Source Hours (SH)
Refueling	Gallons of Fuel Used

(a) Operating Mode Parameters by Emission Process.

Use Type	Type of Bin	Bin Parameters	Example
Passenger Car	Source Bin	Fuel Type	Gasoline
		Mileage	High
		Technology	Fuel Injection/3-Way Catalyst
		Standard	Tier 1
		Emitter Category	Normal
	Operating Mode Bin	Power Bin	Vehicle Specific Power =13 to 16 kW/ton

(b) Example of Source and Operating Mode Bins.

Figure 2: Source operating mode bins (from [61]

categorize data and permit modelling at a very large scale. This method of organization is one of the distinctions between MOBILE and MOVES. The change permits the development of modal emission rates that do not depend on additional modelling analyses (such as regression) and removes the dependence that the MOBILE model had on correction factors [20, 59].

An operating mode is defined as the "breakdown of total activity necessary to reflect differences in emission rates" [61]. As shown in Figure 2a, the operating modes subdivided from the activity of an emissions process varies depending on the process. This organization is also reflected in the databases holding the emissions rates, which, in MOVES, are also categorized based on their associated operating mode. Each of the associated operating mode parameters is then generated, be it from default data (soak times, start times), user supplied data, or calculations based on inputs. Like operating bins, source bins allow for discrete categorization of the model's parameters, but subdivided by source type rather than activity. Source types themselves are specific classes of on-road or off-road vehicles, such as, for example, a passenger car, and are subdivided on parameters such as fuel type or mileage. The resulting combination of a source bin and operating mode bin is thus unique, such as the one shown in Figure 2, forming the basis for all of the parameters in the model. Additional information about these processes can be found in reports done by the US EPA and studies done by other researchers [61, 20].

2.3.1.3 *Calculate an Emission Rate*

After traffic input has been divided into their respective operating mode bins, MOVES uses this input to calculates an emission rate. It is at this stage that MOVES considers additional parameters that modeller can specify, such as weather. An emission rate is estimated for each element (e.g. link, depending on the scale) specified by the modeller in each mode of operation.

2.3.1.4 *Aggregate Emission Rates Across all Modes*

The MOVES model ultimately aggregates all the data using the distinct source and operating bins discussed in Section 2.3.1. Data is aggregated according to the distributions specified by the user (or assumed by the software), including volume, fuel type distribution, fleet age distribution, etc. The basic principle can be characterized with the following equation:

$$\xi_{Total,u} = T_u \times \sum_{n=1}^{N} R_{u,n} \times B_{u,n}$$

Where ξ is Emissions, T is total activity, u is the use type (or source type), N is the total number of bins, R is the emission rate, and B is the bin distribution.

2.3.2 *Data Interaction*

One of the most obvious changes between MOBILE and MOVES is the use of a relational rather than flat file database. Specifically, the MOVES model stores all its input and output data in an SQL compatible database. The popular open-source database system, MySQL, is included with the MOVES setup package [60], but theoretically any SQL database could be used. The choice of an SQL type database allows both the MOVES software and the user to interact with the data using SQL queries, opening the door for the development of custom tools that take advantage of SQL's features.

Input Data

The MOVES model permits users to supply their own input data, but also provides default databases with its software packages. Input databases contain the following [59]:

- Total Activity Information
- Operating Mode Distributions
- Source Bin Distributions
- Meteorology Data
- Fuel Data
- Emission Rate Information

Some of the information in these databases, such as weather, is stratified according to county and state boundaries. Any analysis conducted in MOVES must therefore also specify a geographic location before any of the provided data can be used. Since MOVES is developed for an American context, only data on US counties are provided, and any non-US data must be provided by the modeller. While a version of the MOBILE model adapted for a Canadian context exists(version 6.2c) [17], similar adaptations for MOVES have not yet been developed. As this research project does not involve location-constrained case studies, an arbitrary US county can been selected to for analysis.

Vehicle Data

While MOVES does estimate emissions on a per-second basis, it is still able to accept per-second data on vehicles for project level analyses. In addition to per-second data, MOVES also accepts more macro-scale data, such as average speeds, velocities and lengths for links in a network. This allows traffic models to be connected in a variety of different ways; the primary aim of this research project is to evaluate the accuracy and computational requirements for each of these methods.

2.3.3 *Broader Implications*

MOVES allows modelling at multiple scales, including the option to work at the national, state, county and project level. At the project level, MOVES can act as a sort of emissions micro-simulator and accepts input in three major formats: average speeds, drive schedules and operating mode distributions. As described in previous sections, MOVES fundamentally operates on these so-called "operating mode" schedules. As such, inputs other than operating mode distributions are automatically converted by the software into this format before any analysis is undertaken. For the case of average speeds, MOVES automatically converts an average speed into an operating mode distribution using a "typical" or "default" configuration. This is important, as the emissions and accuracy of the model is strongly influenced by the discrete parameters of a vehicle's operation, such as acceleration or deceleration, and so assuming a constant velocity would likely lead to an underestimation of emissions [54].

2.3.4 *Emission Estimate Differences Between Competing Models*

Although many of the models discussed have similar operating principles, often each of these models provide emission estimates that can be substantially different from each other. As discussed in Section 2.2.1.3, the CMEM was designed to address the limitations of emission factor models, such as MOBILE. The modal approach in the CMEM framework forms the basis of MOVES's own approach to micro-scale simulation, and was also designed to address limitations of previous models by extending them through a multi-scale approach. As a recently developed model, there have not been many studies that compare the output of MOVES to other established models, especially at the micro-scale. A recent study by Chamberlin et al. found that CMEM and MOVES provided comparable estimates for NOx emissions but widely different results for CO emissions [13]. The researchers identified a number of items considered in the MOVES model that are not considered in the CMEM model, such as weather, fuel type, additional pollutant processes, more robust emission rate source data and different approaches to modelling emissions.

As discussed in Section 2.2.1.3, CMEM uses an analytical approach derived from models of the physical processes of combustion to build the model; in contrast, MOVES uses a statistical approach that groups vehicles by specific power and speed [13].

Other studies have been done comparing the macro-scale estimates of MOVES. For example, a study by Bai et al. compared emission estimates from MOVES at the county level in terms of Carbon Dioxide (CO_2) and Methane (CH_4) to EMFAC, which is a mobile source emissions model used in California. Their study found that while MOVES produces similar estimates for CO_2, it provided estimates that were less than half those of EMFAC for CH_4. These results were estimated for conditions in 2002; projections to the future are possible in both modelling frameworks, and the researchers also generated emission estimates for 2030. In this instance, the researchers found MOVES provided CO_2 estimates that were 40% higher and CH_4 estimates that were nearly double those of EMFAC. However, these evaluations were conducted at a preliminary stage of MOVES's development [3], and the results of the study will likely be different if it were repeated with the current model. The researchers also identified several differences between the modelling approaches of each of the platforms, and noted that MOVES uses a combination of a Vehicle Specific Power (VSP) approach with speed bins and vehicle operating times rather than speed correction factors and VMT commonly used by other models like EMFAC and MOBILE.

Another recent study compared the output of MOVES to its predecessor model, MOBILE. The study, conducted by Sonntag and Gao [55], revealed that the estimates of each of the models differ most prominently at low speeds (less than 20 mph). At these speeds, the researchers found that MOVES predicts higher emissions for most of the pollutants modelled, particularly from heavy-duty vehicles. PM emissions were also found to vary more strongly with a vehicle's speed in MOVES, an aspect that was not similarly noticed in MOBILE. At the county level, the researchers also found that MOVES provided NO_x estimates that were higher and HC estimates that were lower than MOBILE. As was the case with the previous study, the researchers completed their evaluation on a preliminary version of MOVES.

A need still exists in transportation literature for additional evaluations of the MOVES model. Although the US EPA has developed the model using substantial data, emission estimates will differ between the various models, and understanding these differences will allow researchers and model developers to improve the current state of emission modelling.

2.4 TRANSPORTATION MODELS

Accurate models of transportation systems has been the focus of research, and a number of different systems have been developed by both private, public and research entities. As with emissions models, transportation models can be differentiated by scale, including micro and macro-scale models. Popular micro-scale systems include VISSIM [51], PARAMICS [50], and SUMO [16] while popular macro-scale systems include VISSUM [51], EMME [26], and MATSim [43]. Although the theory behind models may be similar, each of these models may vary in their exact implementation, and consequently the approximation of traffic behaviour produced by each model may be different. As the focus of this research work is on micro-scale simulation, the next few subsections will briefly discuss competing micro-simulation models, including studies comparing their output.

2.4.1 *PARAMICS*

PARAMICS was originally developed as a project of the Edinburgh Parallel Computing Centre, with the aim of developing a model that used supercomputing techniques, including parallel data processing and analysis [50, 12]. In the original proposal by Cameron et al. [12] PARAMICS was developed to address the limitations of macroscopic models, which could not properly simulate real traffic behaviour in congested situations. Macro-scale models were also recognised as unable to properly reproduce the dynamic and fluctuating nature of transportation systems, and as a result were not suited for certain types of analyses. PARAMICS was created to demonstrate that micro-scale analyses could be conducted on large geographic areas and could predict and model congestion accurately. The researchers in Cameron et al. used real-world data from the Scottish Trunk Network to evaluate their model, which contained data on around 150,000 vehicles. The model was developed for the CM-200 super-computer and pioneered parallelism in computing to solve transportation problems. The original PARAMICS model included separate models for each of the major areas of driver behaviour: speed modification/control, car-following, gap acceptance, overtaking, and lane changing. Each of these models included equations that governed a vehicles behaviour and could be evaluated during the model's evaluation time-steps; these equations considered elements such as driver reaction, impedance of vehicles ahead, etc. The initial model developed lacked sufficient data for proper calibration, but verification of the travel behaviour generated by the model, such as flow density relationships showed that it approximated real-world behaviour with accuracy. At the same time, the researchers recognised that there were limits to the accuracy of their models. For example, the car following behaviour used in the original model had a limited ability to model the "wave"-like behaviour of vehicle motion commonly seen in congested traffic flow. Today PARAMICS is developed and maintained by a private entity [50], but continues to see use in research settings.

2.4.2 *Simlation of Urban MObility (SUMO)*

SUMO is an open-source multi-modal traffic simulation package developed by researchers at the Institute of Transportation Research at the German Aerospace Centre [31]. SUMO has the ability to import networks from other popular transportation suites (such as VISSIM, MATSim) or from other GIS-based network formats, such as OpenStreetMap or GIS Shapefiles. As a microscopic traffic simulator, SUMO simulates each vehicle individually and specifies explicit routes for them, including origin, intermediate and destination nodes/roads. Unlike other micro-scale simulation models, SUMO is able to simulate large-scale networks, including networks of entire cities. The simulative behaviour of SUMO's traffic model is derived primarily from a model proposed by Stefan Krauß[31, 32], which is notable for its simplicity and high execution speed. The model is not without limitations, including conservative gap sizes, low gap acceptance during lane changing, and poor scaling if the simulation time step is changed.

2.4.3 *VISSIM*

VISSIM was developed in the early 1990's and is a discrete-stochastic traffic simulator. It is currently developed by the PTV Group [51]. At the time of development, VISSIM distinguished itself in its ability to model both fixed-time traffic signals and signals that interfaced with traffic through virtual detectors [18]. The original model included two separate programs, a traffic flow model and a signal control model. For the traffic flow component, VISSIM originally used the car-following model proposed by Wiedemann [63], which is a psycho-physical spacing model. Under this model, a faster vehicle approaching a slower vehicle decelerates until reaching a threshold value. This threshold value is a function of speed difference and spacing, but allows for a bunching "wave" effect since it simulates the inability of drivers to perceive small speed and distance differences [18]. In addition to this model, the complex rules governing lane-changing behaviour were also included in the program. Since its original development, the program has been modified to incorporate additional newer models and to take advantage of technological developments in computing. VISSIM models transportation networks through the use of one-way links, enabling nearly any network to be physically represented in the software. Today's version of VISSIM also includes the ability to simulate pedestrains and non-road transit vehicles like trams.

2.4.4 *Estimate Differences among Commonly Used Models*

The myriad of available simulation models mean that for a given problem, a variety of reasonable estimates can be obtained. As is the case with emissions models, researchers have therefore compared the outputs of each of these models to better understand their limitations and differences. For example, a recent study by Maciejewski [37] compared the output of SUMO, VISSIM, and TRANSIMS. In terms of modelled network capacity, the researchers found that the SUMO model had a capacity that oscillated between 85 to 100% of the measured traffic flow on the network they were modelling, depending on whether or not vehicle parameters were thoroughly calibrated. In contrast, VISSIM and TRANSIMS had modelled capacities that were between 130 to 140% of the measured traffic flow. SUMO's model also had a greater number of vehicles present in the network than either VISSIM or TRANSIMS for scenarios with identical volumes and network configurations. Despite the differences, the researchers generally found that traffic features identifiable in one model were similarly identifiable in the other. For example, the researchers noted that for all models, capacity issues began to appear first at the same intersection and movement for simulations run in each of the models.

2.5 INTEGRATING TRAFFIC MICRO-SIMULATION MODELS WITH MOVES

To better understand the complex relationship between traffic and pollution, it is often necessary to link a traffic model and pollution model, such as MOVES, together. Since models vary in their implementation, ad hoc methods must often be developed to link specific model combinations. However, despite their differences, many models share common elements and a method developed and tested on one specific combination of models could be then applied on another. As discussed in Section 2.3.2, there are a variety of different ways to connect the output of a traffic model to MOVES, and broadly speaking they can be grouped into three major categories based on the way vehicle data is input into MOVES: aggregated approaches, disaggregate approaches and hybrid approaches that are partially aggregated.

2.5.1 *Aggregated Integration Methods*

The approach selected can be important, as emissions estimates will vary depending on specific traffic characteristics, such as speed [9], and methods that obscure these aspects can lead to inaccurate estimates. In a recent study, Marsden et al. [42] develop a model that can be used to estimate emissions in real-time using loop detector data. Their model was designed to account for the different vehicle operating modes (acceleration, cruising, idling, deceleration) as well as many other aspects such as the state of repair of the vehicle's emission control system and the type of engine (gasoline or diesel powered). The results of their study compares the results of their model and highlights the inadequacy of an average-speed approach to emissions modelling; they note that macro-scale models are unable to model the detailed effects of new technology has on driving patterns. They also note that high-emitters can be a significant contributor to emissions at an intersection, as in one of their scenarios 10% of the vehicles produced 50% of the CO emissions [42].

In the context of aggregated approaches, a number of studies have also been done in MOVES that take advantage of the wealth of data available. Many of these studies employ aggregated approaches and often aggregate data from traffic models, generating lists of links with average speeds and volumes. For example, Xie et al. [66] developed an integrated tool to link the PARAMICS model to MOVES. To evaluate their tool, they modelled a section of a motorway in Greenville and used an origin-destination travel matrix to generate travel demands. Their analysis was conducted at the project level, and used the default settings in MOVES for attributes such as vehicle age distributions and fuel formulations. The output from the PARAMICS model was analysed using Microsoft Access, but input tables were simplified and only included link volumes, average speeds and vehicle type distributions. Although the study used an aggregated approach, they demonstrated the usefulness of an integration framework by evaluating the effects of a shift in fuel use to compressed natural gas for transit vehicles. In the scenarios they tested, they estimated that emissions would decrease by 34% if 40% of busses switched from diesel to compressed natural gas [66].

One of the major limitations of an average speed approach is the loss of time-based granularity in analysis data, especially if the analysis in question covers a multi-hour period. At the project level, MOVES limits the time-frame of an emissions analysis to a one hour time period. To estimate emissions for multi-hour periods, some studies have used MOVES to provide individual estimates of emissions for each time period and then aggregating (or analysing) these results for a daily period. One study to use this approach was conducted by Veeregowda et al. [62] The researchers attempted to analyse the effect a mega mixed-use development project in New York City would have on the surrounding area's air quality. The researchers focus on $PM_{2.5}$ emissions and use the CORSIM model to simulate the traffic effects. The authors analysed an entire 24 hour day's traffic, and used MOVES to estimate emissions factors for each of the day's hours. The subsequent emissions factors were then aggregated together to generate four emissions factors, one for the over-night period, one for each of the AM and PM peak periods, and one for the mid-day period. Although the use of an average speed approach results in the loss of some of CORSIM's simulated data, the authors indicate that this method is infeasible due to the scale of data that would need to be analysed and due to computational limits [62].

2.5.2 *Hybrid and Disaggregate Approaches*

Disaggregate approaches are not often employed in the literature due to the extensive data and computational requirements. Research at a micro-scale into CO_2 and other emissions is relatively new, and the methods and models applied continue to improve and evolve as technology changes.

Despite this, a few studies have been done that integrate MOVES with a micro-scale traffic model using various levels of disaggregate data. One of the most recent studies to do this was conducted by Abou-Senna et al. [2]. Their study compared and analysed various ways to link a VISSIM micro-simulation model with MOVES. The authors evaluated three different ways to link VISSIM and MOVES together, each using one of the three main ways of providing input to MOVES for project-level analyses. The first approach utilised average speeds and volumes, the second used second-by-second link drive schedules, and finally the third used an external analysis to provide operating mode distributions directly to MOVES. The authors compared the estimates of CO, NO_x, PM2.5, PM10 and CO_2 for of each those methods. The authors evaluated the estimates of these pollutants as vehicles traversed 11 links of a network designed to compare the methods. The first method is typical of the approaches discussed in the previous section. In the second method, the authors did not provide fully disaggregate link drive schedules to MOVES for analysis, but rather grouped similarly performing vehicles together. As discussed in Section 2.3, MOVES fundamentally operates at an operating mode level; the authors recognise this and designed their final method to capitalise on the ability of a modeller to specify this distribution directly. The authors developed an application to generate the input for this method which they called VISSIM/MOVES Integration Software (VIMIS) and use the concepts of VSP to generate operating mode distributions externally. One of the benefits of this approach is that it has the potential to substantially reduce computational burdens. The authors note that vehicle trajectory records for their test network could reach sizes of 10 gigabytes and are not accessible using conventional programs. Through the use of their ex-

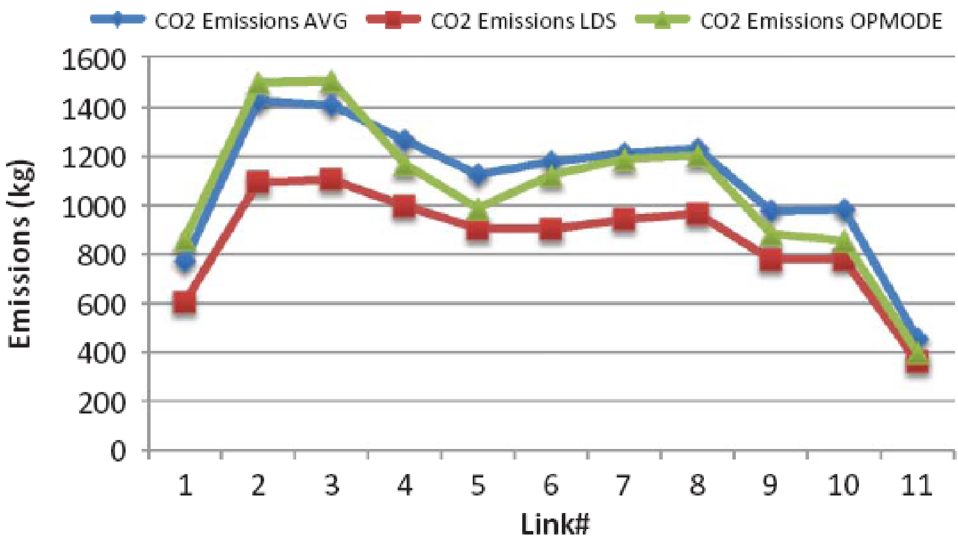


Figure 3: The effect of input method on emissions estimates, results from [2]

ternal analysis program, the subsequent tables generated for analysis are only 300 kilobytes. The subsequent analysis conducted by the authors revealed that on an overall basis the average speed approach produced the highest emissions estimates while the link drive schedules produced the lowest estimates. The results of the op-mode distribution generated by their VIMIS software was in the middle of the two methods. The results of their study for CO_2 emissions has been included here for reference in Figure 3, illustrating the aforementioned trends of their study [2]. Their analysis, shown in Figure 3, highlights the fact that an average speed approach can often still be used to great effect in understanding overall emissions trends, though as the authors note some of the differences in estimates can be substantial. The researchers concluded that, for the case of their scenario, providing MOVES with average speeds generally results in an over-estimation of emissions while the grouped link drive schedules resulted in lower emissions. The researchers assert that the operating mode approach is more accurate as it considered operating parameters provided by VISSIM that are ignored by MOVES when converting drive schedules, though they acknowledge that this would be difficult to validate in field settings. The model they developed was subsequently applied in an analysis at a test-bed prototype of Florida's I-4, which is a limited access highway corridor [1].

As one of the few detailed studies analysing the different ways to link traffic and emissions models, the method proposed by Abou-Senna et al. [2] is of particular interest, and indeed no other comparable studies were found in the extensive literature review conducted. Despite its detail, their study focusses strongly on the development of their VIMIS tool and as a result some of the details of their analysis's methodology were not clear. For example, although the authors mention that they group similarly performing vehicles together before inputting link drive schedules to MOVES, the authors did specify how vehicles were grouped together. Due to their network's scale, it also appears it was infeasible to consider each vehicle individually in MOVES, and thus the authors make no comparisons using link drive schedules that fully describe vehicles travelling on their network. As their analysis and evaluations focus on motorways, the effect of extensive periods of idling, stop-and-go situations caused by signals and congestion are not discussed in their study.

Although no other studies have been done that compare different integration methods, a recent study by Papson et al. [49] applied a novel approach to aggregate data from SYNCHRO before inputting it in MOVES. As was the case with the study done by Abou-Senna, the authors process the output of a model externally into four modes: accelerating, decelerating, cruising and idle. Unlike the approach in Abou-Senna, the authors did not use this process to specify an operating mode distribution, rather they use it to build a representative link drive schedule. The representative schedule was divided into four links, each representing one of the aforementioned activity modes. The advantage of their approach is a reduction in the data required to arrive at an estimate of emissions. A traffic analysis in SYNCHRO is not a micro-level analysis, however SYNCHRO is able to provide some performance metrics from urban networks, including the control delay of traffic movements and vehicle queues. The approach, called "Time in Mode Analysis", then uses these variables to generate estimates of the time vehicles spend in each of the four modes. While aggregate in nature, this approach is able to provide more accurate representations of vehicle travel patterns than the simple use of average speeds [49].

2.5.3 *Integration Applications for Policy Evaluations*

An effective link between a traffic micro-simulation model and a pollution model allows a greater depth of analysis than would be possible from each model individually; but, ultimately an integration framework is only useful if it can be applied to solve problems. In previous literature, these frameworks have been successfully used to evaluate the effect of policy decisions and proposals. For example, a recent study by Int Panis et al. [27] aimed to determine what effect an Intelligent Speed Adaptation system, which caps the maximum speed of a vehicle to the local limit, would have on emissions of CO, NO_x, etc. To evaluate the proposed system, they modelled a real-world transportation network in Ghent, Belgium using DRACULA, which is a network-wide traffic microsimulation package and estimated emissions using an empirical-based model. They tested different levels of market penetration for the proposed system but ultimately found that while the proposed system successfully reduced average speeds, the effect on emissions was not so clear, and no statistically significant reductions were observed [27]. Despite this, their approach successfully demonstrated the potential of an integrated approach in the analysis of an emerging technology, and the results of their study will be useful for policy and decision makers.

Another study by Boriboonsomsin et al. [10] evaluated the effect of High Occupancy Vehicle (HOV) lane configurations on vehicle emissions. The authors integrated the PARAMICS model with the CMEM model and evaluated a 12 mile section of State Route 91 E in California. The researchers developed an integrated framework and used it to evaluate the differences in emissions between HOV lanes that are access controlled with separate ingress/egress sections and configurations where access is continuous. Their evaluation found that continuous access to HOV lanes had lower emissions than access-controlled ones, primarily due to the effect of weaving concentrated in the egress/ingress sections. Unlike studies done using MOVES, the CMEM model is designed to be used with traffic-microsimulation models and in this study the authors use the trajectories of individual vehicles. While this feature of the CMEM model may sound attractive, the model has some limitations that stem from its development. The model was developed through data collected by 343 light duty vehicles tested in laboratory situations. Since data was collected in laboratory settings using a dynamo-meter, it does not accurately represent real-world driving conditions. CMEM is also unable to estimate emissions from heavy goods vehicles such as trucks and busses and does not estimate particulate emissions [48].

A study done by Rakha et al. [52] has also shown that the CMEM model can exhibit abnormal behaviour. The authors used data from databases provided by the Oak Ridge National Laboratory and US EPA to compare the MOBILE5a, MOBILE6, VT-Micro and CMEM models. In particular, the authors found that CMEM exhibited strange behaviours in its estimates of CO emissions at low speeds and high acceleration rates. The study also found that CMEM underestimated emissions for acceleration manoeuvres when compared to the databases' field data. In contrast, the authors found that the performance of the MOBILE6 and VT-Micro was in line with the US EPA field data.

2.6 CONNECTED VEHICLES

Connected vehicles are a much-discussed topic in current literature. Broadly speaking, there are two major forms of connected vehicle technologies, with each serving a distinct purpose: Vehicle-to-Vehicle (V2V) communications and Vehicle-to-Infrastructure (V2I) communications. V2V technologies allow vehicles to communicate with each other, exchanging information to support technologies such as collision warnings, emergency braking, lane change warnings and blind spot warnings [47, 28]. These technologies have been demonstrated to have significant safety benefits when used in a connected vehicle environment, with the potential to address the causes of about 81 per cent of vehicle crashes [47]. Existing proposals, however, require dedicated short-range communications operating at 5.9 GHz to be installed on each participating vehicle.

Similarly, V2I technologies allow vehicles to exchange information with surrounding infrastructure. Like V2V technologies, both infrastructure and vehicles must be equipped with a communications device to participate [47, 28]. V2I technologies have the potential to address the causes of about 26 per cent of all crashes [47]. V2I technologies proposed in the literature can broadcast Signal Phase and Timing (SPaT) information, transit information, pedestrian movements, and traffic information to participating vehicles [47, 28].

Beyond their safety benefits, these technologies could provide network operators, drivers and researchers with data on an unparalleled scale. Their low latency and dedicated nature means more data could be exchanged than would be possible over shared communication systems such as cellphone-based systems. Many different ways to take advantage of these systems have been proposed. For ex-

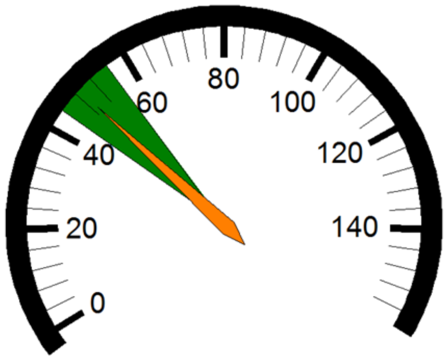


Figure 4: A Speed Advice Speedometer

ample, previous studies have demonstrated the possibility of connected vehicle technologies to optimise signals while allowing priority to be yielded to emergency vehicles [23], estimate queue length [35], and implement strategies such as ECO-Driving (see section Section 2.6.1).

2.6.1 *ECO-Driving*

As discussed in Section 2.6, V2I technologies can be used to broadcast and collect information to and from vehicles. Broadcasting SPaT information to vehicles has been demonstrated to have potential safety benefits (e.g. a vehicle is aware that a signal is red and takes action to prevent red-light running) [47], but implementations in literature have also demonstrated the potential of this information to reduce emissions. ECO-driving systems proposed in the literature use connected vehicle technologies to exploit the link between driving behaviour and reduce overall emissions. Driving behaviour is modified through an advisory system that selects optimal speeds for a driver to follow. A 2014 study by Hobeika et al. [24] highlights this potential by modelling the emissions of drivers categorised in three groups: aggressive, moderate and defensive. Their study found that if aggressive drivers change their habits and become defensive drivers, HC, CO and emissions of NOx would decrease by 15 to 21 per cent. Another study by De Vlieger et al. [15] found that aggressive driving can increase fuel consumption by 40% and emissions by a factor 8. They also found that aggressive driving can increase the incidence of traffic jams and that the effect of driving behaviour is most pronounced on gasoline fuelled vehicles. ECO-driving systems expand on this by not only attempting to reduce instances of aggressive driving, but also through speed advice on approach to intersections.

2.6.1.1 *Interaction with Drivers*

Speed advice can be provided in a variety of ways, but common proposals in the literature often present this information to drivers on the speedometer [64] in a manner similar to that in Figure 4. In this way, whenever the driver checks his or her speed, the recommended speed they should travel at is also presented with it.

The nature of ECO-driving systems is optional in nature, and ultimately it remains to be seen whether such systems would gain widespread acceptance. However, past research has suggested that drivers may be willing to adopt more eco-friendly approaches to driving. An extensive survey of over 5000 individuals done in 2012 by Tommer and Hotl on drivers in Europe found that drivers believed speed advice systems similar to the ones discussed previously were useful. Respondents generally agreed that these systems could help save fuel and contribute positively to the environment. Despite this, the respondents generally also affirmed their desire for such systems to be optional; many respondents also indicated that such systems are not worth paying extra for, though the results were mixed and varied depending on the country of the respondent. In addition to the issue of adoption, the authors also suggest that such systems may not be beneficial in all cases. The authors note that such systems could pose a distraction to the driver and could have other safety implications [56].

2.6.1.2 *Evaluations in the Literature*

ECO-driving seeks to reduce instances of acceleration and deceleration and can operate through a device that recommends speeds to a driver [64]. Systems such as these continue to get attention from various organisations and governmental agencies. For example, a current project under-way at the United States Department of Transportation called AERIS is applying a simulation-based approach to evaluate the feasibility of strategies such as ECO-driving on a real-world corridor. The ECO-driving analysis conducted as part of the AERIS program aims to evaluate the effectiveness of providing driving advice to drivers. While the full research has not been published yet, as part of the program a PARAMICS-based traffic simulation model based on a real-world network in Palo Alto, California is being developed. Preliminary work has demonstrated a 4% reduction in fuel consumption and a 6% reduction in travel time, depending on traffic conditions, on a hypothetical motorway segment. On an arterial network, preliminary modelling work has demonstrated fuel savings of 5%, but a 2% increase in travel time [58].

Other studies have also demonstrated the potential of this method through both simulations and field demonstrations. For example, a study by Xia et al. conducted in 2012 used both a simulation-based and field-based approach to evaluate an ECO-driving system [64]. Simulations were conducted in PARAMICS, and field tests were conducted at a test intersection. The results revealed reductions of around 14 per cent in both fuel consumption and CO_2 emissions. Both the simulation and field tests were conducted for single intersections with signals operating on a fixed time schedule. This limits the applicability of the method to real-world situations. Other studies have been conducted in the past as well. A study by Li et al. in 2009 demonstrated that an individual vehicle has the potential to reduce its emissions and fuel consumption by up to 7 per cent and 8 per cent respectively [36]. Still yet, another study by Barth et al. showed potential reductions in fuel consumption and CO_2 emissions by about 12 per cent [4]. Another study done by Xia [65] analysed the effect of a number of different parameters on an advanced algorithm that included V2I-based communications. The researchers conducted a sensitivity analysis and found that communication range had a strong effect on their algorithm's effectiveness. Fuel savings ranged from 30 per cent with infinite range to below 5 per cent with only 200 metres of communication.

2.6.1.3 *Algorithmic Theory*

The literature available on ECO-driving includes a number of different ways to calculate a vehicle's optimal speed. These methods vary in terms of their complexity and potential to reduce vehicle emissions. According to Mandava et al., the simplest ECO-driving model solves the following basic equation [40]:

$$\max\left(\nu = \frac{D}{t_p}\right) \qquad \text{Where:} \begin{cases} t_p \in [t_g, t_r) \text{ or } t_p = t'_g \text{ if } s = \text{red} \\ t_p \in [0, t_r) \text{ or } t_p = t'_g \text{ if } s = \text{green}) \\ \nu \leqslant \nu_{\text{limit}} \end{cases} \tag{1}$$

Where ν is the optimised speed of the vehicle, D is the distance to the stop bar, ν_{limit} is the speed limit, t_g and t_r are the green and red times respectively of the signal and t_p is the phase time being considered for optimisation. This equation represents the simplest approach to ECO-driving, and does not consider elements such as deceleration time, queue length, or other limitations. This algorithm also considers only the signal ahead of the vehicle, but other studies have shown that more optimal paths may be found if multiple signals are considered in the evaluation [14]. The downside of this approach is that the intended path a vehicle wishes to choose must be known, and that relevant SPaT (or speed) information also be provided to the vehicle. In an advisory context, drivers may also be more distrustful of the system if it appears to intentionally skip green lights that would otherwise be achievable.

2.7 DATA CLUSTERING

Often when dealing with large amounts of data, it becomes necessary to implement some form of clustering. Traffic micro-simulation models have the potential to produce large amounts of data. A vehicle traversing a network can have thousands of data points associated with its trajectory, and a large network can have thousands of vehicles entering it per hour. The computational requirements for calculating the emissions of each discrete vehicle modelled in the simulation therefore increases exponentially. Clustering reduces this computational load. This section explores potential algorithms available in the literature that could be adapted to group individual trajectories together. Particular attention is given to the algorithm selected for this work, k-means clustering, but some exploration of alternatives is also presented.

2.7.1 *K-Means Clustering*

As an algorithm, the k-means process is relatively simple to understand. As its name implies, the aim of a k-means process is to partition N observations in k sets such that the in-set variance of all the sets is a minimized. While many variations of the algorithm exist, initial proposals followed the same basic procedure: First, in initial seeding of k groups is chosen, each of which are assigned an initial "mean" value. Sample points are then successively added to the group whose mean it is closest to. After adding all points, the mean of the groups is recalculated. Each of the k-means therefore represent the mean of the k groups selected at the onset of the model (giving rise to the algorithm's name). An iterative process thus follows where after assigning all the points, the means are then updated and the points are re-assigned. This process could continue until there is no substantial change in the assignment process, or the means themselves do not change significantly [38, 22]. It is important to note, however, that while a k-means algorithm will converge, this convergence may not be to the optimal solution and could oscillate indefinitely [38].

To determine the nearest mean to an individual point, a sum-of-squares approach is usually employed. Sums-of-squares can be thought of as a representation of the Euclidean distance between two points, and thus minimizing its value is akin to selecting the "nearest" point. The mathematical aim of the algorithm is therefore to reduce in-group sum of squares across all groups [21]. This does not always need to be done through the iterative procedure discussed previously. Hartigan and Wong suggest an algorithm which switches points between clusters as an alternative the iterative procedure [21].

As all the k-means algorithms discussed still require an initial seeding, it should be no surprise therefore that the selection of an appropriate seed can improve both the algorithm's results and efficiency. Initial groups could be selected randomly [38], sequentially from the dataset (e.g. first x points), or evenly from the data [21]. In their algorithm, Hartigan and Wong recommend ordering the points based on their distance from the mean, and then selecting points at even intervals based on how many groups are desired [21]. This approach has the advantage of ensuring that no group will be empty after the initial assignment (since the point selected for the initial mean will definitely be added to the group).

2.7.1.1 *Selecting a Value for* k

There are many approaches to the selection of the value of k in the literature, and some of the simplest involve formulas based on the number of points in the data. For example, one possible rule of thumb suggested is to use the following equation [41]:

$$k = \sqrt{n/2} \tag{2}$$

While approaches such as these have merits, they ultimately have no relationship to the data and may not always result in an optimal choice. Another possible approach is the use of a goodness-of-fit value or some other indicator of the differences between individual members of clusters and the average trajectory. These methods ultimately form a "heuristical" approach to the selection of k, called the "elbow" method. In these approaches the appropriate indicator value chosen is graphed against the cluster count on an ordinary x-y plot. As the cluster count increases the value of this indicator value will change, but this change will begin to decrease and eventually the graph will visibly "flatten" [30]. This flattening occurs due to the diminishing returns of increasing the cluster count. Ultimately the selection of the cluster count with this method will not be "precise" as the chief limitation of this approach is that the determination is often made visually using ad hoc rules. As such, more advanced methods proposed in the literature exist and the problem of selecting optimum values of k is an area where many implementations will differ from each other. For example, a recent study in the literature used a machine-learning approach using hidden Markov models that automatically adjusts k and evaluates each tested value, discarding those that did not satisfy the conditions of the Markov model. In their study, the researchers programmed heuristic rules into their algorithm to guide the selection and evaluation of k [53].

2.7.1.2 *Applications in the literature*

An understanding of past applications of a k-means approach can provide insight into its ability to solve the problems of this research. Since its development, the k-means algorithm has seen various applications and improvements across multiple fields. Within the field of transportation, current research work is still being done that uses various forms of k-means algorithms. For example, a study in 2014 by Lentzakis et al. [34] compared a weighted k-means approach and a k-harmonic means approach to group links in a network into clusters based on their traffic parameters. Their study revealed that centre-based clustering algorithms (such as k-means) can be very effective in partitioning urban traffic networks. Of the two methods studied, the weighted approach was also found to perform better. Another study by Ferreira et al. [19] in 2013 further expand on the original k-means algorithm by taking a vector-based approach. While fundamentally similar to the k-means algorithm originally proposed, their algorithm applied techniques developed in computer graphics and visualisation and they demonstrate its effectiveness on a number of different datasets, including GPS traces of people and vehicles. Finally, as mentioned previously, Saunier and Sayed demonstrate a vehicle tracking algorithm for intersections and traffic-conflict detection that uses a k-means clustering approach [53]. In their study, the number of clusters is generated automatically using a heuristic approach. A clustered approach was necessary in their setting as the automatic vehicle software would often trace a path for a vehicle in the video frame that was different than a previous vehicle moving in the same manner. Clustering allowed these trajectories to be grouped together, and the heuristic approach for cluster seeding removes the need of the modeller to guess an appropriate number for the final cluster count. While these trajectories are different than those that would be generated by a micro-simulation software package, many of the concepts are similar. Their approach demonstrates the applicability, simplicity, and power of a k-means approach.

2.7.2 *Other Clustering Algorithms*

There are always a number of different approaches that can be taken to solving a problem, and data clustering is no exception. Indeed, as discussed in the previous section, there are many algorithms that build on the concepts of the k-means approach (such as Ferreira et al. [19]). Broadly speaking, clustering algorithms can be grouped into two groups: *hierarchical* or *partitional*. The k-means algorithm is a partitional algorithm. In contrast, hierarchical algorithms function either by, starting with each individual data point, progressively merging similar pairs of points to form a cluster hierarchy or, starting with all the data points in one cluster, by dividing clusters into smaller clusters successively. Each successive stage of the algorithm, be it divisive or agglomerative, represents a hierarchical level. The most popular hierarchical algorithms are the *single-link* and *complete-link* algorithms, whereas k-means is the most popular partitional algorithm [29].

2.7.3 *Conclusion*

The k-means is a popular algorithm due to its simplicity when compared to other alternatives, but as mentioned in previous sections, suffers from the issue of requiring the researcher to specify a value for the cluster count (k). K-means is also limited by the method employed to generate the clusters. For example, the use of Euclidean distance means that certain data relationships (e.g. spirals vs. lines) may not be clustered in the best manner.

Part II

RESEARCH METHODOLOGY

After developing a sound background in the literature surrounding the topic, an application was developed to link VISSIM and MOVES together. This section details the methodology used to link VISSIM and MOVES together, including descriptions of the scenarios used to evaluate the possible methods. The methodology developed through this method was subsequently applied to evaluate the potential of an ECO-driving system. This section also details the methodology and algorithms used to evaluate this proposal.

3 VISSIM-MOVES INTEGRATION

Using the concepts and ideas explored in the literature review, a technical framework to link the output of VISSIM to MOVES was developed. MOVES is able to analyse data at varying levels of aggregation and has multiple levels of analysis, including national, county and project levels. For this research, the analysis in MOVES was conducted using a project level analysis. As discussed in Chapter 1, MOVES always operates at an "activity" level, regardless of the modelling scope and aggregation level.

3.1 INTEGRATION FRAMEWORK OVERVIEW

A successful analysis in MOVES can be done using a variety of different methods, and each method differs in the manner in which the output from the traffic simulator is prepared for input. The approaches one can take to link traffic and emissions models can be broadly classified on a spectrum that ranges from fully disaggregate to aggregate. The most disaggregated of approaches is to simply use the raw output from the traffic model in the emissions model while the most aggregated approaches treat the traffic micro-simulation model similar to a macro-simulation model where only total volume and average speed are used. Hybrid approaches are also possible and fall somewhere in-between, and can include methods such as data clustering or velocity binning. The level of aggregation provided results in a trade-off where increasing the level of disaggregation simultaneously increases both the accuracy of the estimation and the computational time. To quantitatively assess this trade-off, this study compares and contrasts the results of different types of integration, including the following three that have been commonly used in the literature:

1. AS: These methods model a network of links and use link volumes and link average speeds to arrive at emissions estimates. These approaches are the simplest that can be employed and often have the lowest computation requirements. As discussed previously, when MOVES is provided with average speeds, it uses a default operating mode profile to divide the time a vehicles spend on the network into the various operating modes it simulates (e.g. accelerating, decelerating, idling, cruising). This is markedly better than estimating emissions solely as a "cruising speed", but still suffers from inaccuracies depending on the networks being modelled. This accuracy is further affected by how the network is partitioned in MOVES

2. VB: In this method, vehicles traversing a network are stratified according to their average speed. There are a variety of different methods to achieve this stratification, as stratification could be provided for individual links (or sub-links) or simply for the network-at-large. Since MOVES automatically partitions vehicles into operating mode distributions, this method may interfere

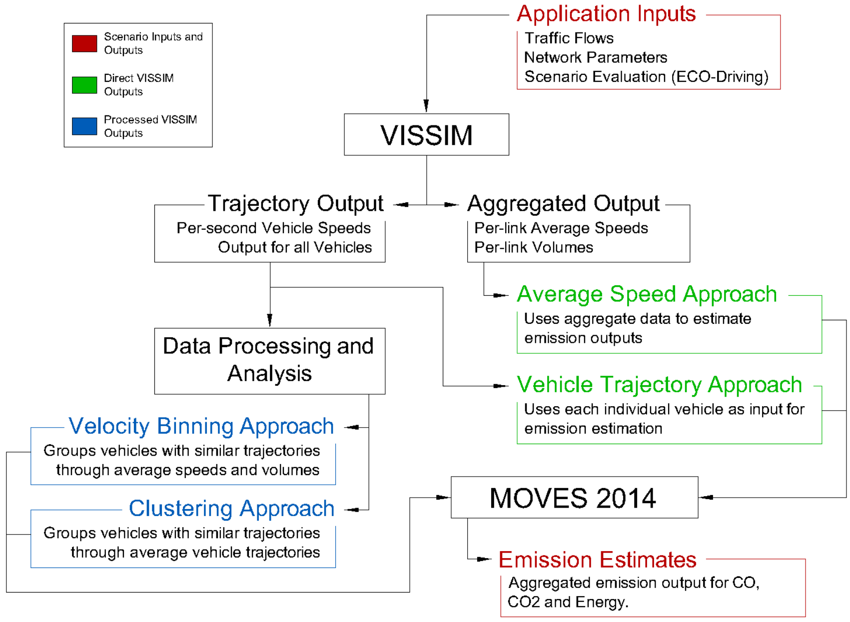


Figure 5: Overview of the Integration Framework

with this process as the speed a vehicle travels at directly affects its operating mode distribution. For example, lower average speeds may be caused by frequent stops at traffic signals or by lower speed limits; however, these two scenarios would still have different emissions profiles.

3. Vehicle Trajectory Based (VT): In this method, MOVES is provided with vehicle trajectories for each vehicle directly. This method is the most disaggregate possible, and trajectory outputs from a simulation model (e.g., VISSIM) are only converted to a format that MOVES can accept with no additional aggregation. In this method each vehicle is modelled as an individual link with a volume of 1 and a link drive schedule that fully describes its speed for each second. Estimating emissions using the trajectories of individual vehicles is very time consuming, making it infeasible to model networks of large scale.

In addition to these methods, a hybrid approach has also been proposed and tested. This approach uses the K-means algorithm discussed in the literature review to aggregate individual trajectories further. The use of trajectories also precludes the need to develop and apply an understanding of the complex relationship between a vehicle's trajectory and its operating mode, as trajectories are directly interpreted by MOVES. This approach will also simplify the analysis, as the data generated by the micro-simulation model (which can be substantial) is condensed.

The relationship of these approaches in the overall context of the analysis can be seen in Figure 5. As shown in this figure, a successful analysis first starts with the definition of the goals of the scenario and the specification of all relevant inputs. In the case of this study, scenarios take the form of networks designed to compare the accuracy of each of the methods (Section 3.7), scenarios designed to test the computational time of the clustering algorithm, and scenarios applying the method to evaluate an ECO-Driving system based on Connected Vehicle (CV) technologies (Section 3.7.3). These inputs are required for a microsimulation analysis to be conducted, which in this study is conducted using VISSIM.

VISSIM can provide two types of output, time-based data on all operating parameters of vehicles (such as speed, acceleration, etc), and aggregated statistics on the network. In the case of this study, the methods of linking the output of VISSIM to MOVES can thus be further categorised by whether or not subsequent data analysis is required. This is illustrated in Figure 5, as the AS and VT approaches only require that VISSIM's output be reformatted such that MOVES can understand it. In contrast, the VB and Clustering approaches require additional processing and analysis, either using spreadsheet programs such as Excel, or through the implementation of another algorithm.

At the end of every analysis, MOVES provides an output database table containing emission estimates for the requested parameters. In the case of this analysis, output was requested for CO, CO_2, and Energy Consumption. Ultimately, each of these methods have their own advantages and disadvantages, and each varies in the complexity of the approach and the amount of external data processing that is required. These methods are discussed in greater detail in the following sections, including descriptions of the process by which data is extracted, the process by which data is analysed, and the process by which data is encoded to be imported into MOVES.

3.2 APPLICATION INPUTS

The primary goal of this research project was the development of an integration framework to link VISSIM and MOVES, and to then test and evaluate that framework. This research project therefore includes three scenarios: The first scenario is designed to compare the emission estimates of all the integration approaches to each other in terms of perceived accuracy and computational time, with the VT approach forming the reference baseline. The second scenario is designed to evaluate the computational performance of the algorithm developed to cluster vehicle trajectories, with the aim of assessing its potential for use on larger-scale networks. The final scenario applies the framework to evaluate a technology based on connected vehicle systems, namely, ECO-driving.

3.3 VISSIM

VISSIM is a complex program and is able to simulate many aspects of a transportation system's operation. VISSIM includes a comprehensive user-interface that allows the modeller to interact with the software to configure the simulation (see Figure 6). For the purposes of this analysis the majority of VISSIM's customisable options were left at their default settings, including all settings for the car following model. The tools provided by the UI allow the modeller to specify the location, direction, width and orientation of links, speed limits, rights of way, traffic lights, vehicle routes and other features. Specific configurations are detailed in Section 3.7 for each of the evaluation scenarios developed for this research project.

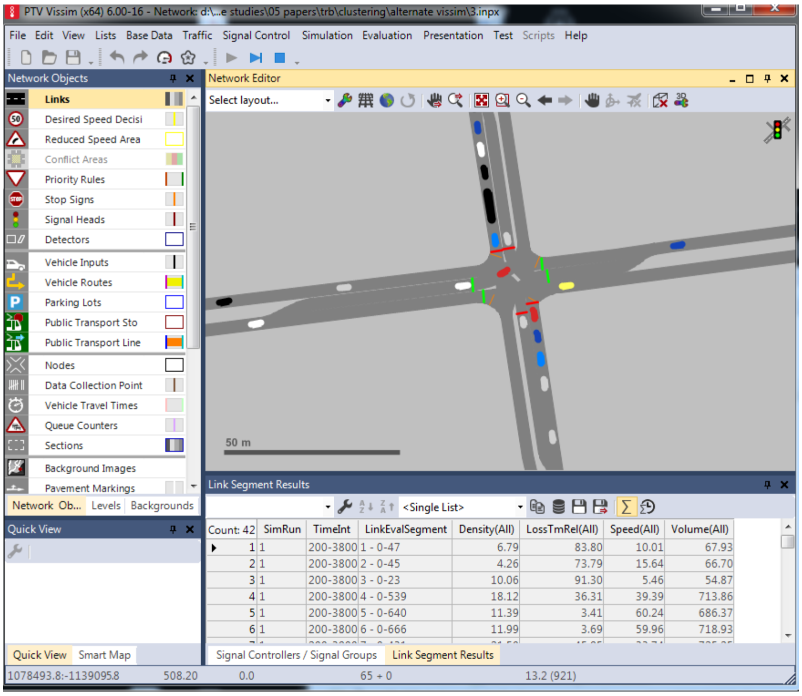


Figure 6: VISSIM's main UI

3.4 AGGREGATED OUTPUT

At the simplest level, vehicles and simulated components can be thought of in an aggregate manner. In this approach, the average speed and total volume of vehicles on each link can be extracted from the micro-simulation model. This approach is the simplest possible and provides the quickest computation times in MOVES.

3.4.1 *Obtaining Data from VISSIM*

VISSIM is readily able to provide this information on a link-by-link basis, and this information can be extracted directly from VISSIM in a format that can be easily converted to one understandable by MOVES. In the diagram shown in Figure 6 the tool that provides this information is shown (Link Segement Results, the list-box in the bottom right), and information can be exported from this tool into Excel for further analysis.

The scale of aggregation employed in this approach is up to the modeller, therefore, and a variety of different combinations are possible. The networks shown in Figure 11 can be aggregated in two major ways. The first is to consider each leg as an independent link (e.g. giving 8 links in total in Network 1) and the second is to consider each direction of travel as an independent link (e.g. giving a total of 4 links in Network 1). While seemingly more detailed, the first method may not be superior. This is because MOVES does not assume that vehicles always cruise at the provided average speeds, as discussed in Section 2.3.3, but rather that some default driving regime is followed. Therefore, a more disaggregated approach may not always provide the best results.

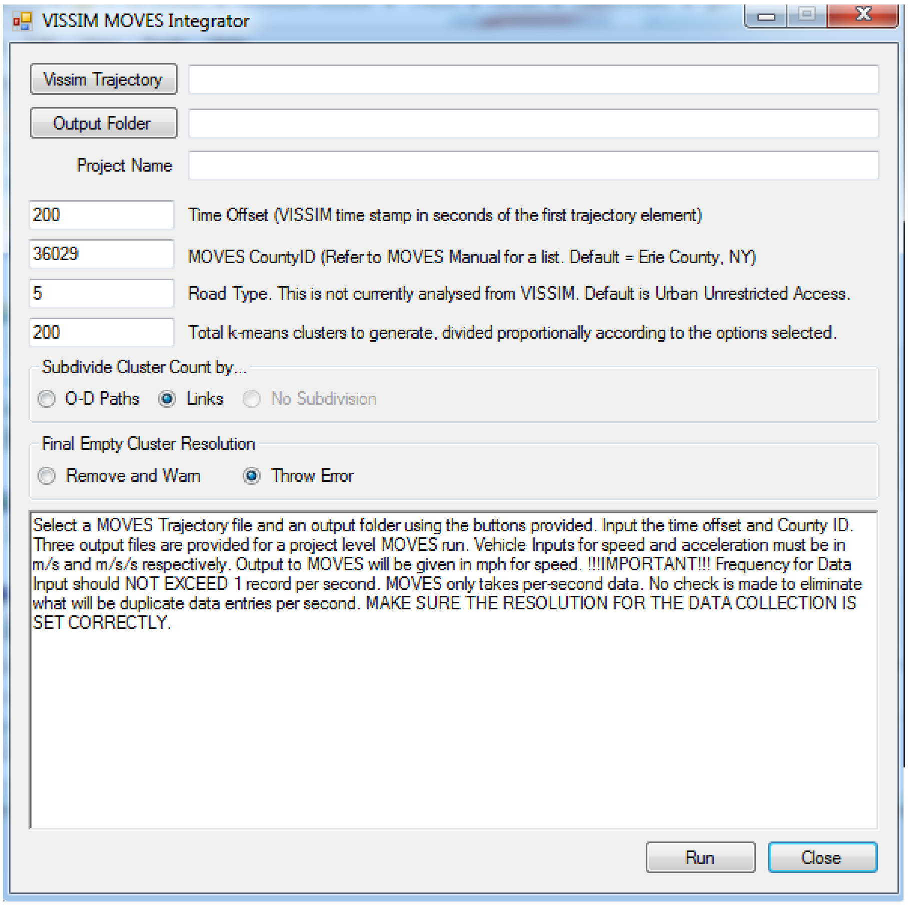


Figure 7: Integration Tool GUI

This study therefore evaluates both methods of organizing the network, though ultimately traffic analysts and modellers may select one arrangement over another due to reasons other than accuracy and increased data disaggregation (such as existing networks already coded as complete links).

3.5 Trajectory Output: An Integrated Tool

While simple methods like AS approaches can be readily prepared solely using VISSIM and Excel, alternative methods require additional tasks that can be tedious to do without automation. Therefore, an integrated tool was developed for this study to translate the outputs from VISSIM into formats usable by MOVES. As indicated in Section 3.6.1, MOVES requires three tables of input to be provided to it for each scenario. The integrated tool therefore analyses the output of VISSIM (a vehicle record file for all vehicles on the network, Figure 32 on page 89) and uses it to build the three tables required. This tool was developed using Visual C# and includes an interactive Graphical User Interface (GUI) for ease of use. A sample of the GUI is shown in Figure 7. The GUI allows a user to navigate to and select the VISSIM trajectory file to be used in the analysis and also to specify a folder to place the output files. A variety of options for the program are also configurable. The "Project Name" field allows the user to specify the name of the project that will be included in all output files (for example, if the name was "foo" then all tables would have names starting with "foo"). The "Time Offset" field is required as it allows a user to specify the time range that the data operates at. Since vehicles in a VISSIM simulation only appear at vehicle inputs located at the starts of links, it is customary to allow vehicles to flow through the network for a time before recording data. Timestamps in the output will then be shifted by a certain amount of time, and this factor allows that to be accounted for since MOVES only permits second stamps that range from 0 to 3600. The "County ID" field is included in many of the output tables, but is not a feature of VISSIM and so must be specified manually. While VISSIM has the ability to model different road types, this is not analysed by the tool and thus a default must be specified for all links. For most networks this will not be an issue as often only a small area with only one road type will be simulated. Finally, the last field specifies the number of clusters to be used when generating tables for a clustered analysis. This tool has been used in some way to complete all the remaining analyses, the specifics of which are discussed in the following sections.

3.5.1 *Using the Raw Output*

As is the case with the aggregated approaches, the use of raw trajectories requires little processing of the VISSIM output. For this scenario the tool simply uses the output of VISSIM to generate the three tables required for a sucessful MOVES run. Individual vehicles are modelled in MOVES as individual links with volumes of one. MOVES is not a fully micro-scale model, and the subsequent analysis that it conducts is computationally intensive. Indeed, while little processing is required by the tool a successful run in MOVES can take well over an hour, depending on the network's size. Despite this, this method can be useful when analysing small networks and is taken as a representation of the most accurate estimation of emissions possible in this study and the results of all other methods are compared to the output of this scenario. The integrated tool therefore always generates output for this scenario in the form of the three tables mentioned earlier: a link drive schedule table which contains discrete trajectory information for all vehicles (see Section B.1.2); a table of links which contains the distance individual vehicles travel and their average speed (see Section B.1.1) and a table of sources on each link which essentially identifies each vehicle's type (Section B.1.2).

3.5.2 *Velocity Binning*

The VB approach seeks to aggregate individual vehicles on the basis of the speed they travel, but fundamentally the approach is analytically similar to the AS approach as a network of links is provided without any trajectories to MOVES. These links contain the average velocity, length, and volume of vehicles in each respective bin. Despite its similarities to the AS approach the VB method's inputs were still derived through the use of individual vehicle trajectories. The VB approach takes advantage of one of the data tables created when generating input for each individual vehicle (as described in Section 3.5.1), which includes a table that contains the distance and average speed of each vehicle on the network. Vehicles listed in this data table were then grouped together with other vehicles of similar velocity using ordinary functions in Microsoft Excel. For this study, vehicles travelling on each network and in each scenario were grouped into 7 different bins. For the case of the first network, these bins were spaced at intervals of 5 kph starting at 15 kph and going to 45 kph. The larger size of the second network made this increment infeasible, however, and an increment of 2 kph was used instead, beginning at 22 kph and going until 34 kph. After grouping individual vehicles, their velocities and travel distances were averaged together to arrive at an input table of 7 links. A sample data table can be seen in Appendix B.

3.5.3 *A Clustered Approach*

Clustering the individual trajectories of vehicles from the raw data is a simple way of reducing the computational load. In the literature review a k-means approach was discussed extensively. A k-means approach is a potentially powerful way of reducing the computational load while retaining a higher level of disaggregation than using average speeds or volumes. The k-means approach groups data points in k clusters. In the context of this study, each trajectory can be thought of as a data point, and the k clusters generated by the algorithm can be thought of as k "links", with the number of trajectories in each cluster corresponding to its volume. Unlike other aggregated approaches, each "link" in this approach is a fully described trajectory representing the average of each vehicle in the cluster. A program to implement the k means algorithm was developed as an extension to the program described in Section 3.5. This extension was also developed in C# and runs immediately after the VISSIM input data has been parsed. The subsequent sections describe each step of the algorithm used to cluster the trajectories as coded in the extension. There are 4 major steps.

1. Set the parameters of the clustering algorithm.
2. Initialize the k-means algorithm and seed the clusters.
3. Run the k-means algorithm.
4. Export the results when the halting condition is met.

3.5.3.1 *Set the parameters of the clustering algorithm*

As mentioned previously, the results of the k-means algorithm depends on the number of clusters the modeller wishes to create. There are a variety of different ways to derive this number, but its selection is primarily a function of the level of detail required counter-balanced with the computational load. It is ultimately the modeler's decision as to what constitutes an optimal value for k, but generally the aim is to select a value that provides the highest disaggregation while also minimizing the computational loads that will be incurred. The F-ratio is the ratio between the explained and unexplained variance; this ratio will be higher when cluster counts are low and lower as cluster counts increase. While the F-test is often used to evaluate samples when conducting an ANOVA, in this instance the F-value is used as a heuristic to select a value for k. This method is a variant of what is commonly called the "elbow method", where an appropriate cluster count is selected visually from a point in the plot where flattening is observed to occur [30]. Arriving at an estimate using this approach requires multiple iterations of the algorithm with different values of k, so for this study multiple values of k ranging from 20 to 400 were tried. The results of this approach are discussed further in Section 4.1

Beyond the selection of k, the extension developed also contains two different ways to pre-process the data. To reduce computational loads and to allow individual assessment of link segments, it may be preferable to first partition a vehicle's trajectory according to the links those trajectory portions travel on. This is the recommended default for the program, and the result is that the total cluster count will be partitioned to each of the links in the network according to the number of vehicles that travel on them. Clustering will then be done on the link-level, with multiple clusters representing a single link. Alternatively, the program contains a simple O-D path analysis algorithm that can be applied to analyse smaller networks. If this option is selected, vehicles are grouped with other vehicles who travel on an identical set of network links (i.e. an identical path through the network) and the subsequent clustering continues at the O-D level. As was the case with the link-based method, total cluster counts are partitioned evenly to each of the O-D paths according to the volume on each path.

3.5.3.2 *Initialize the k-means algorithm and seed the clusters*

In the literature describing the k-means approach, particular attention is given to the methods by which the algorithm is initialized. Because of the algorithm's design, an initial condition must be provided and then an iterative approach is used to arrive at a final solution. The selection of a poor initial condition can create additional computational loads as the algorithm will take longer to converge. Extremely poor selections of the initial conditions can create further issues if successive iterations trap the results of the convergence test in a local minimum, resulting in a sub-optimal solution. Therefore, in this research project the k-means algorithm was seeded such that initial clusters were spaced evenly. Trajectories were ordered according to the raw sum square difference from a null trajectory. In this work, the sum of squares for any trajectory is calculated according to the following equation:

$$SS_i = \sum_{j=1}^{n} \left(v_{av,j} - v_{i,j}\right)^2 \tag{3}$$

In this equation SS_i is the sum of squares for trajectory i. For a given trajectory, there are n data points representing each individual speed measurement. Therefore, $v_{av,j}$ and $v_{i,j}$ represent the individual speed measurements at point j of the average trajectory and trajectory i respectively, where j increments from 1 to length of the trajectory, n. Of course, in the initial seeding no average trajectory is being considered, and the equation reduces to the following form:

$$SS_i = \sum_{j=1}^{n} \left(v_{i,j}\right)^2 \tag{4}$$

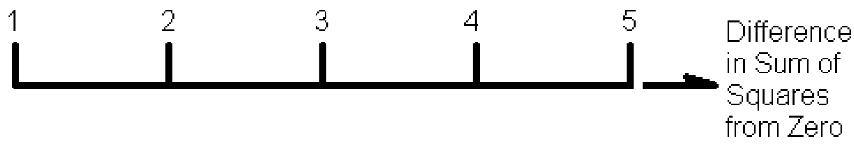


Figure 8: Sum Square Continuum

The manner in which these equations are written means that the "trajectories" are not trajectories in the sense of a representation of position with respect to time, but rather these trajectories trace the speed of a vehicle travels with respect to time. The choice of speed over position does not affect the result: the order of "trajectories" would be the same regardless for vehicles on the same path (as speed affects position). However, since speed-based trajectories on different paths may be similar to those on other paths, vehicles were first divided into groups based on their paths. The number of clusters for each path was then proportionally assigned based on the number of vehicles in the path (with a minimum of 1 cluster for each path). Then, as illustrated in Figure 8, If the trajectories are ordered within their path-groupings according to the results of equation 4, then a continuum of the trajectories can be imagined, from which initial trajectories can be selected at intervals of N/k, where N is the total number of trajectories and k is the desired number of clusters. It is important to note, however, that while vehicles were first divided according to the paths they took this division is not necessary in an emissions analysis. Emissions are not dependant on the location of the vehicle, rather they are affected mostly by the speed and acceleration a vehicle travels at.

3.5.3.3 *Run the k-means algorithm*

After initialization, the algorithm proceeds to assign trajectories to each of the clusters. Sums of squares are computed for all trajectories sequentially for each of the available clusters. The order of selected trajectories is arbitrary, and in this case is simply the order by which the vehicles appear in the simulation. Mathematically, the process to derive the sum of squares is the same as that listed in equation 3. Since vehicles travelling at different speeds will cover ground slower or faster than each other, this means that the "length" (i.e. the number of data points) of the trajectories may not be the same even for vehicles on the same path. Consequently, it is necessary to devise a method to accommodate trajectories of differing length. For this work, when either the average trajectory ($v_{av,j}$) or vehicle trajectory ($v_{i,j}$) for point j does not exist, that point's value is taken to be zero instead. After the sums of squares have been computed, the trajectory is assigned to the cluster with the lowest sum of square, which represents the group it is most similar with. The average trajectory of the cluster is then recomputed for each point j according to the following equation:

$$v_{av,j} = \sum_{i=1}^{n} \frac{v_{i,j}}{n} \tag{5}$$

As was the case in equation 3, $v_{av,j}$ is the average velocity at point j, $v_{i,j}$ is the velocity of trajectory i at point j and n is now the total number of trajectories in the cluster group represented by v_{av}. In this equation, values for v_j are computed such that v_{av} is as long as the longest member in its cluster. A value of zero is substituted in the above equation where no data point exists for trajectory i.

3.5.3.4 *Export the results when the halting condition is met*

The halting condition for a k-means algorithm can be a complicated topic. Depending on the parameters seeded to the algorithm, it is possible that a solution to the algorithm be unattainable in a reasonable amount of time. Original proposals of the algorithm indicate that the algorithm is to continue until no further changes to the cluster assignments is observed [22]. This is the logic applied for this project. More specifically, the stopping criterion is met when the difference between successive iterations for all average trajectories is less than 2.5 per cent. Furthermore, an additional constraint is imposed requiring all clusters to have a minimum volume of 1. In a k-means algorithm, it is very possible that, especially for large cluster sizes, successive iterations will create clusters that have volumes of zero. Therefore, after each iteration clusters with no volume are reseeded with means from the largest cluster and iterations continue until no zero-volume clusters remain. To prevent the algorithm from iterating forever, a cap of 999 iterations was imposed in the implementation of the algorithm, though convergence was often observed within 10 iterations for nearly all cluster sizes.

A successful run of the clustering algorithm will create the three tables required for input into MOVES, and the process for running MOVES after completing the clustering process then becomes nearly identical to the process described in Section 3.5.1.

3.5.4 *Running an Integrated Simulation*

The tools and procedures developed and described in the previous sections ultimately come together when a simulation is run to completion. Because the tools are separate from each other, there is no need to run them all at the same time, and indeed, additional runs in MOVES can be made by analysing the data from VISSIM in a different way. For each of the proposed scenarios discussed in Section 3.7.1 the following procedure was followed:

- First, two networks were coded in VISSIM according to the diagram in Figure 11. Volumes were then set according to each of the planned volume scenarios, beginning with the low and progressing on to the moderate and variable volume scenarios.

- After all scenarios ran successfully in VISSIM, the resulting outputs were imported into Excel spreadsheets and CSV files for analysis using the integrated tool discussed in Section 3.5.

- MOVES input tables were prepared directly from VISSIM's output data using Excel for the AS method.

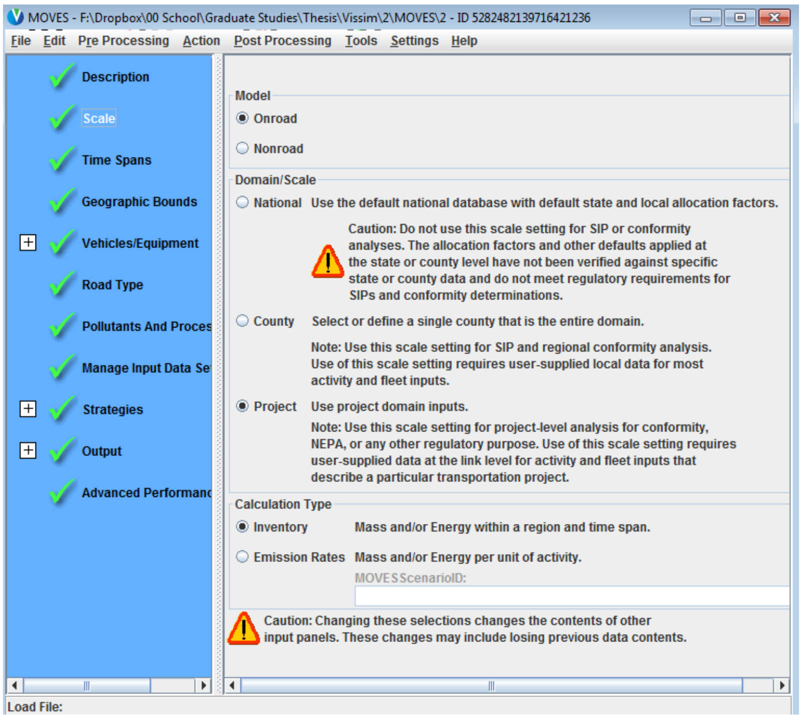


Figure 9: Main Menu of MOVES

- The integrated tool was used to generate MOVES input tables for the individual trajectory method and the k-means cluster-based approach. The "Elbow" method was used to select the value of k that would be used in MOVES for analysis.
- The data tables generated by the integrated tool for the individual trajectory method were used to generate the input for the VB method.
- A RunSpec was set-up in MOVES using the common configurations discussed in Section 3.6.1. Individual data tables were imported into MOVES and separate estimations were done for each of the data tables generated.

3.6 MOVES

In MOVES simulations are prepared as "RunSpecs" using a graphical editor installed with the software. This editor, shown in Figure 9, allows customisation of any detail or setting available in MOVES.

In Figure 9, the list-menu on the left details every major component of the RunSpec that must be configured. These configurations affect the results of the emission estimates, as the inclusion, exclusion or modification of certain parameters will change the model's behaviour. The following sub-section details the most important settings configured in MOVES that affect the model's behaviour, and are common to all model runs.

3.6.1 *Common MOVES RunSpec Configurations*

The output of all of the individual scenarios discussed Section 3.7 are eventually analysed in MOVES. With the exception of volume inputs, these scenarios all share common settings in MOVES that allow their output to be compared directly. Volume inputs are entered in a separate tool from the menu items shown in Figure 9, and as such these settings are generally the same across all scenarios.

For the case of the "Scale" tab, as discussed extensively in previous sections the "Project" scale was selected. The scale tab also allows setting of the calculation type, which was set to "Inventory"; selecting this option means MOVES will tabulate total emissions across the network for all vehicles.

Settings in the "Time Spans" tab are partially constrained by the selection a project level analysis, as MOVES restricts the analysis to a single hour (to simulate additional hours, additional runs must be executed). For this analysis, the year was set to 2014, the month was set to March, the day was set to a weekday, and the hours were set to 0900h to 0959h. At this scale, these settings are generally only used by MOVES to select appropriate weather conditions to include in the simulation.

The "Geographic Bounds" tab allows selection of a county from any US State. For the purposes of this analysis Erie County, NY was arbitrarily selected due to its proximity to Canada. There are no Canadian regions included in the MOVES databases by default, but as this analysis does not include any real-world data, this selection has been made solely to satisfy the requirements of a MOVES run.

The "Vehicles/Equipment" tab allows specification of the vehicles running on the network from a myriad of different options. These options include combinations of fuel and vehicle size, such as Diesel powered Long-haul Trucks, or Compressed Natural Gas powered Transit Vehicles. While the options are vast, any combination of vehicle selected must also be included in the VISSIM simulation as vehicle type distributions (or individual vehicle types) must also be specified with volume inputs. For simplicity only two types of vehicles have been included in this analysis: Gasoline Powered Passenger Cars and Gasoline Powered Light Commercial Trucks. These vehicles are a primary component of urban traffic; VISSIM's default vehicle distribution also includes only two vehicle types (Passenger Cars and Commercial Vehicles) and the addition of more vehicles would not increase the strength of the analysis.

The "Road Type" tab allows the type of road included in the network to be specified. For a project level analysis, these road types are also included later when volume data is imported. For this analysis, only Urban Unrestricted Access roadways were included.

The "Pollutants and Processes" tab allows the modeller to specify which pollutants should be included in the model's results. MOVES is able to estimate the quantities of a large number of pollutants from different states of vehicle operation. In addition to running exhausts, these states include start exhausts, extended idle exhausts (such as at parking), evaporative fuel losses, etc. However, the selection of additional operation states increases the data requirements. For example, if "extended idle exhausts" are requested for each of the pollutants modelled, MOVES must also be provided with an estimate of idle time to include in the calculations.

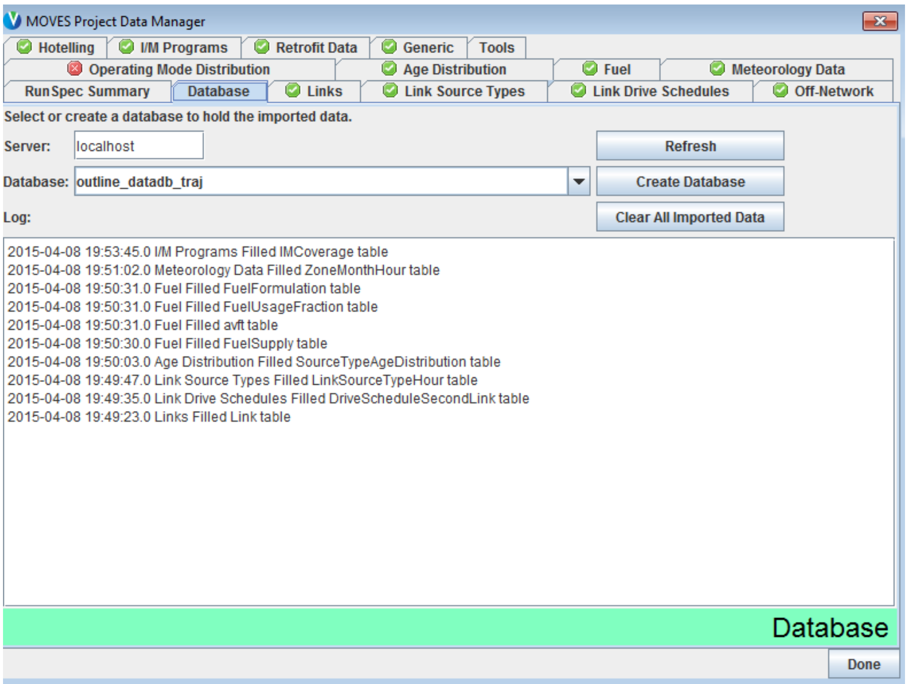


Figure 10: MOVES Data Importer

Furthermore, the selection of additional pollutants increases the model run time and can also increase the amount of data required. For simplicity, only "running exhaust" emissions were selected for Carbon Monoxide, Total Energy Consumption and Carbon Dioxide.

Finally, the "Strategies" and "Advanced Performance" tabs contain options not relevant to this project and no additional configurations were selected from these tabs for any runs. The "Manage Input Data Sets" and "Output" tabs allow configuration of the databases used to provide vehicle inputs and emissions estimates respectively. A database schema was created for each scenario's input and output.

Additional Data

MOVES also provides a specialized data importing tool that is able to import tabular files (such as CSV files or Excel Files) into a database format. The graphical interface of this data importer is shown in Figure 10. The data importer contains several tabs that can be used to export default data or import formatted data from VISSIM. Vehicle trajectory data must be formatted in a drive schedule table that include a column for the link the vehicle is travelling on, a column indicating the vehicle's speed, a column indicating the grade the vehicle is currently traversing over, and a column indicating the time those measurements were taken at. A Sample data table has been included in Section B.1.2 on page 86.

Additionally, all scenarios also require a table of links to be provided that includes the link number, county number, road type, link length, link volume, link average speed, and the link average grade (see Section B.1.1 on page 83. Although this table is required for inputs into MOVES it differs slightly in its function between the different methods. When a drive schedule is provided for a link, MOVES will use the drive schedule to estimate emissions rather than the average speed provided in this table (which must still be provided). Furthermore, if individual trajectories are provided a link is also provided in this table for each individual vehicle, and the volume is set to 1 (see Figure 27).

Finally, a table must also be provided that specifies the vehicle distributions on each link as a decimal (see Section B.1.3 on page 86). In the case where individual trajectories are used, as was the case with the link table, a single vehicle is travelling the link and so its type distribution includes only one vehicle (see Figure 30. This type distribution comes directly from VISSIM and the integrated tool translates VISSIM's built in vehicle types to a matching type in MOVES. On all other methods, the *sourceTypeHourFraction* is used to indicate what percentage of vehicles travelling the link are of each type. For a particular link, the sum of the values in *sourceTypeHourFraction* must be 1. When using AS methods, these values are simply set to 0.9 for passenger vehicles (MOVES vehicle type 21) and 0.1 for commercial vehicles (MOVES vehicle type 32). On all other methods, this value depends on the percentage of vehicles travelling the link, and in the case of the clustered method this is directly provided by the integrated tool.

Once created, all these tables are imported using the "Links", "Link Source Types", and "Link Drive Schedules" tabs on the project data manager.

In addition to these tables, the project data manager allows default tables for Fuel, Age Distribution and Weather to be exported for inspection and import. Even if default data is to be used, the modeller must export these tables first. For the case of this study, default data was not modified for any of these tabs.

3.7 EVALUATION SCENARIOS

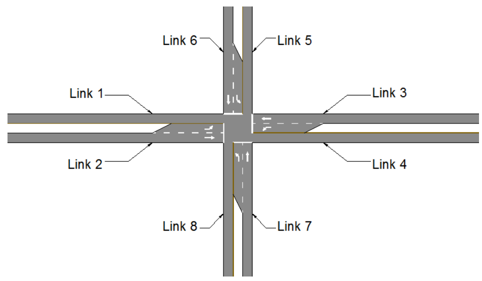
The methods discussed in the previous sections were then applied in three different scenarios, with the first scenario focussing on comparing the computational time and accuracy of each the methods, the second focussing on evaluating the computational performance of the clustering approach, and the third focussing on an application of this approach to asses an emerging CV technology.

3.7.1 *Scenario 1: Evaluating Accuracy*

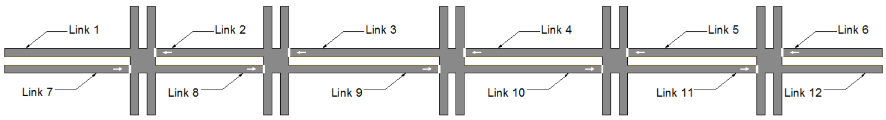
The accuracy of the methods used to connect MOVES to VISSIM will vary depending on the network being analysed. This is because, as discussed previously, when using aggregate methods MOVES makes assumptions about the behaviour of vehicles on the network that affects estimates. The geometry of the network has an influence on this, as the number of signals and their layout, including elements such as left turn lanes, right turn ramps, etc will affect how drivers operate their vehicle.

3.7.1.1 *Network Configurations*

To evaluate the differences in the estimates provided by each of the methods discussed in Section 3.1, two different networks were devised. These networks were devised with the aim of evaluating differences in terms of sensitivity, and to obtain



(a) Network 1



(b) Network 2

Figure 11: Vissim Test Networks

an idea of the effect of network size on the calculation time. These networks are shown in Figure 11.

The first network, shown in Figure 11a, is a single intersection with four two-lane approaches. The signal at the intersection operates on a 90 second cycle and provides equal time to all approaches. This signal has been configured with VISSIM's Ring Barrier Controller which accurately simulates the operating principles of modern North American traffic signals. Turning movements were permitted on this network, and left turn lanes were also included to accommodate turning vehicles (though no protected/permissive left turn phases were included). 10 per cent of vehicles were set to turn left and 10 per cent to turn right at the intersection. To facilitate the subsequent analysis, each link in VISSIM was modelled as a distinct link. This approach allows average speed and volume to be obtained natively from VISSIM for each link.

The second network, shown in Figure 11b, is more complex and includes five signal controlled intersections. To simplify the analysis, only volume flowing on the main corridor was analysed for all scenarios. Turning movements were therefore not considered in this method. In this network, links were also modelled distinctly between the various intersections (in contrast to defining a single link spanning all intersections).

These two network orientations were specifically selected to analyse the effect of increasing complexity on emissions estimates. On a complete corridor, such as the one modelled in Network 2, vehicles make multiple stops if they are required to stop at multiple traffic lights. In contrast, vehicles on Network 1 will only ever have to stop at the single intersection, and thus even if both scenarios have the same average speed, the emissions profile may be different between them. Network 2 was designed to simulate corridor with directional flow, and signals were timed according to the Highway Capacity Manual to give progression in the east-west direction. The signal progression was developed for the medium volume scenario and remained unchanged for all test scenarios. Signals for this network were timed on the assumption of low volume for non-corridor directions (500 vehicles per

SCENARIO	NETWORK 1 VOLUME	NETWORK 2 VOLUME
Low Volume	350 vph, all approaches	750 vph EB, 500 vph WB
Medium Volume	700 vph, all approaches	700 vph EB and WB
Variable Volume	500 vph for the first 20 min, 700 vph for the next 20 min, 800 vph for the last 20 min,	500 vph EB and 250 vph WB for the first 20 min, 700 vph EB and 350 vph WB for the next 20 min, 900 vph EB and 450 vph WB for the last 20 min.

Table 1: Network Volumes

hour) and using a cycle length of 90 seconds. On both networks, speed limits were also set to 60 kilometres per hour for all simulated links. Vehicle composition was set at 10 per cent light commercial vehicles and 90 per cent passenger cars. All links drawn in VISSIM were also set to have widths of 3.3 metres.

3.7.1.2 *Volumes*

In addition to differences in network geometry, multiple volume scenarios were also tested. Higher volumes can cause congestion and increase average speeds, which can also affect driver behaviour and emissions, while lower volumes provide less impedance to travelling vehicles and increase average speed. For this study, three different volume scenarios were tested on each of the two networks, including a high volume scenario, low volume scenario and variable volume scenario. These volumes are detailed in Table 1 and are different for each network.

Each volume and network combination was tested in VISSIM for one hour (as MOVES restricts input at the project level to one hour). In the case of the variable volume scenario, the volume was changed every 20 minutes and were designed to start at a low uncongested volume, and rise steadily to a higher congested volume. The high volume scenarios were designed to operate the intersection at or near capacity (i.e. vehicles queuing at a red light are generally able to clear the intersection when it is green). Conversely the low volume scenarios operate well below capacity and vehicles queuing at a red light are always able to clear the intersection on green.

3.7.2 *Scenario 2: Evaluating the Computational Performance of Clustering*

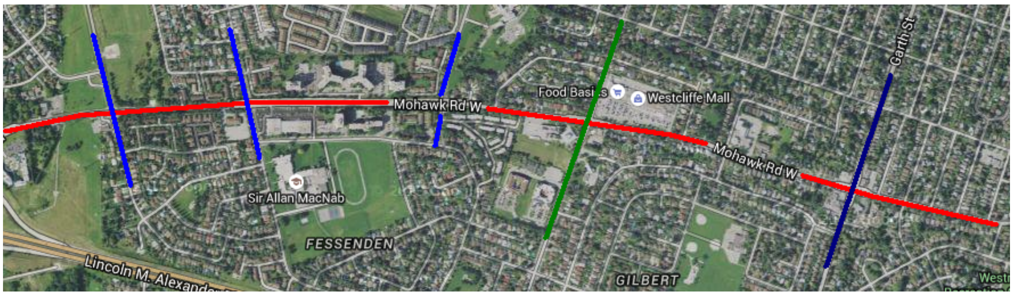
Similar to scenario 1, scenario 2 was designed to evaluate the computational performance of the integration framework. This scenario, however, focussed exclusively on the potential for the clustered approach to reduce computational burdens as networks increase. To evaluate this potential, a series of networks increasing in size were coded into VISSIM. The base network consisted of the same network shown in Figure 11a. Larger networks were built by placing additional intersections at the end of the base network (and subsequently added intersections) and adding additional feeder links. The result is a corridor-type that increases in size and has a shape similar to the network in Figure 11b. Unlike the network in Figure 11b, these test networks were designed with equal spacing and identical signal timing patterns for all intersections, and ranged in size from 2 intersections to 16 intersections. Output from the VISSIM simulations run on this network were then used to create clustered MOVES input files using the clustering algorithm, and the computation time required to estimate these clusters was evaluated.

3.7.3 *Scenario 3: ECO-Driving Analysis*

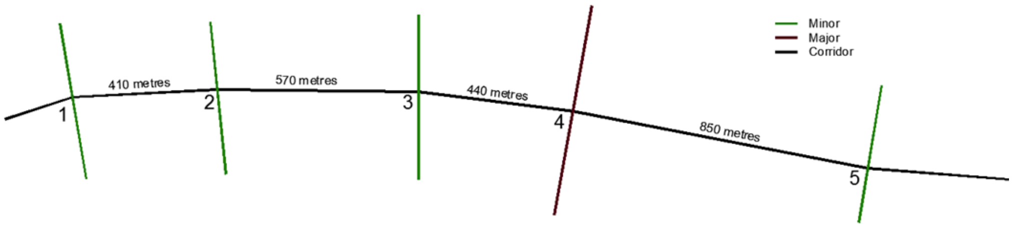
Although the focus of this scenario is an application of the integration framework developed, it is important to note that specific aspects of the scenario may merit more rigorous study. The benefits of an ECO-driving approach is based on the premise that a vehicle is receiving speed advice and then reacts to that advice using human driving characteristics. Additionally vehicles following this vehicle will also react and slow down. All these aspects are governed parameters of the micro-simulation model that govern driver behaviour. For this scenario, however, no focus was given to setting these parameters. Instead, default settings were used in VISSIM. Speed advice was simulated by dynamically altering a vehicle's "desired speed" parameter, which in VISSIM represents the speed a vehicle wishes to attain and hold at that particular instant. This speed does not need to be the speed limit and is overridden by other aspects of driver behaviour, such as the need to stop at a traffic light or yield the right of way to another vehicle.

This analysis ultimately sought to evaluate the benefits of an ECO-driving simulation in a realistic corridor. Many previous studies have used test scenarios such as one way corridors or single intersections to evaluate the merit, but such evaluations may over-state the benefits of an ECO-driving system as they often ignore elements of signal design such as coordination and progression. For this study a corridor was designed in VISSIM using a real-world roadway in Hamilton, Ontario as the model. Representations of this corridor is shown in Figure 12. The network contains 5 intersections, of which four were classified as "minor" and one was classified as "major". Past research has shown that high congestion can decrease the effectiveness of ECO-based systems [25, 36] and that low volumes can also decrease their effectiveness [65]. With the exception of the major road, all roadways in the network (including the corridor) are two-lane roads. All roadways also have left turn lanes to allow left turning vehicles to queue. The major road is a larger road and has four lanes and also contains separate left turn lanes. While the layout of the roadways follows the real-world roadways, specific geometrical features have been omitted or modified to simplify the analysis.

There are a number of different parameters that can be tested in such models, but limitations of time and resources often mean focusing on a specific area. Since volume-based effects are not the focus of this research, volumes were selected to reflect moderate use conditions. The volumes used in this study are shown in Table 2. The main corridor has directional volumes with more volume flowing in the eastbound direction. All minor streets have the same volume (200 vph), while the major street has the highest volume (as it was a four lane road). These volumes were synthesised specifically for this study and do not reflect actual use conditions. The distinction of "major" and "minor" intersections was set primarily for the purposes of signal coordination and timing. All signals were timed according



(a) Terrain Map of Network



(b) Network Properties and Links

Figure 12: Vissim Test Networks

DIRECTION	VOLUME
EB Corridor (1 to 5)	500 vph
WB Corridor (5 to 1)	350 vph
Minor Street (both ways)	200 vph
Major Street (both ways)	750 vph

Table 2: Network Volumes

to the procedure specified in the Highway Capacity Manual (hand calculations). Signal offsets were also set to give progression for vehicles travelling across the corridor, and signals at all intersections except the major one have identical cycle lengths (90 seconds). In contrast, the signal at the major intersection has a different cycle length and gives more time to the through movements of the major street's direction. These differences in signal timing are common in the real world, where signals along one corridor cannot always be timed to allow progression on every street in every direction.

Market penetration can have the strongest effect on this system, and its effectiveness on a large-scale often depends on this; therefore a number of different values for the penetration rate were tested.

3.7.3.1 *A Dynamic Testing Tool*

While the integrated tool described in Section 3.5 allows for VISSIM and MOVES to be linked together, a separate tool is required to implement the ECO-driving system. As the proposed ECO-driving system modifies the behaviour of vehicles based on parameters of the simulation (singal phase and timing, vehicle position, etc) it was necessary to create a tool that could interact with VISSIM while it is running. VISSIM natively provides an interface that can support this operation, called the Component Object Model (COM). COM allows a program written in a specific language (e.g. Python or C#) to communicate with the VISSIM application within the program using calls in the language.

A program was therefore written to communicate with VISSIM and implement an ECO-driving system. This program was written in C# and was designed as a separate program from the integrated tool described in Section 3.5. When programming an application using COM, an iterative loop structure is normally used and in each iteration of this loop the VISSIM simulation is advanced a time-step and the program runs any commands it may have to do. This is reflected in the program's general structure shown in Listing 1.

The code in this listing is taken directly from the application developed to implement the ECO-driving system. Using the COM system, a VISSIM file that contains all the initial settings for the network (such as volumes, signal timing and link layout) is loaded by the program and interfaced with. In Listing 1 this is visible by the declaration of the *VissimOperations* delegator. The "delegator" handles all the commands directed at VISSIM, and immediately on declaration it's constructor runs the commands shown in Listing 2. Interaction with VISSIM is ultimately facilitated by the *VissimClass* class, which becomes available to the programmer when VISSIM's COM interface is linked to the program. As soon as a call is made to *VissimClass.LoadNet* VISSIM immediately loads and begins running. The program does not continue until VISSIM loads completely; to the user this process is clearly visible, as an ordinary VISSIM window similar to the one shown in Figure 6 will be shown.

Despite granting the ability to interact with VISSIM, the uses of *VissimClass* are very restrictive and aspects of a simulation cannot be accessed in the same manner as ordinary variables in a program through features of the C# language. For example, vehicles on the network are contained in the *VissimClass.Net.Vehicles*, but

Listing 1: Excerpt of the Program's General Operation

```
static void Main(string[] args) {
  try {
    /* This passes the VISSIM file to a class that      */
    /* starts VISSIM with the loaded network. The class */
    /* is bound as a variable and can be called later   */
    VissimOperations Delegator =
      new VissimOperations(@"C:\VissimFiles\2.inpx");

    /* This starts a loop iterating until the desired   */
    /* runs are complete                                */
    for (int run = 0; run < total_runs; run++) {
      /* The delegator runs a single simulation step in */
      /* VISSIM. After that step is run it calls checks */
      /* the parameters of the simulation and throws an */
      /* error if something is wrong                    */
      if (!Delegator.VissimSingleStep()) {
         throw new Exception("Error running a step...");
      }
    }
  }
  catch (Exception e) {
    Console.WriteLine("Error");
  }
}
```

Listing 2: The VissimOperations Function

```
public VissimOperations(string location) {
  this.vissim_simulation = new VissimClass();
  this.rnd = new Random();
  this.current_seed = rnd.Next(1, 20000);

  /* Tells VISSIM to load the file passed to it        */
  this.vissim_simulation.LoadNet(location);

  /* After loading the file, specialised data strut-   */
  /* are used to hold information about the network.   */
  this.SignalTracker = new SignalPhaseTracker(vissim_simulation.Net);
  this.VehicleTracker = new VehicleTracker();

  /* As in an ordinary VISSIM run, we need to let the  */
  /* simulation play out a little before saving data   */
  this._runJunkSteps();
}
```

Listing 3: The VissimSingleStep Function

```
public bool VissimSingleStep() {
  /* Calls a function that advances the simulation and  */
  /* updates the program's record of signal states      */
  this._simulate_update_step();

  /* To reduce computational load, vehicles are only    */
  /* polled every 5 steps (1 simulated second). After   */
  /* polling vehicles, the program sets desired speeds  */
  if (this._current_step%5 == 0) {
    this._update_vehicles();
    this._update_ECO_desired_speed();
  }
  this._current_step++; //increment the step counter.
  return true;
}
```

this list cannot be interacted with through commands such *for* loops, rather the *Iterator* attribute of *VissimClass.Net.Vehicles* (*VissimClass.Net.Vehicles.Iterator*) is used to cycle through the vehicles in the list. Another example of these limitations can be found in the procedure to retrieve vehicle attributes. Vehicles retrieved through the *Iterator* are interacted with through the *IVehicle* class. To retrieve an attribute, a call is made to *IVehicle.get_AttValue("string")*, a sub-function of the *IVehicle* class, with "string" being substituted for the desired vehicle attribute. So, to retrieve a vehicle's position a call must be made to *IVehicle.get_AttValue("Pos")*. Because these values are returned from this "generic" function, they must also be converted to appropriate C# data type (such as integers, floats, doubles, etc). This "disconnect" of VISSIM objects from the language means that all attributes must first be imported into operable data types before actions and analyses can be done on them quickly, as accessing attributes repeatedly using the COM interface incurs a computational penalty. This disconnect also affects software debugging tools as variables and attributes held within VISSIM are not accessible while debugging unless imported first. Despite this disconnect, the *VissimClass* allows complete control of many attributes of the simulation, and documentation is provided in VISSIM's help file (though this documentation can be hard to follow).

After the simulation has been loaded, as is the case in scenarios where VISSIM is run without an external program, it is necessary to first allow vehicles entering the network to populate all links completely (as vehicles only enter at source points and links will start empty). The commands in Listing 2 thus end when this is complete and control returns to the next line after *VissimOperations* in Listing 1. The iterative procedure described at the start of this section commences, where each iteration of the *for* loop makes a call to *VissimOperations.VissimSingleStep* (shown in Listing 3. As discussed previously, limitations in the COM interface means that important variables and aspects of the simulation are tracked in separate data structures for use in the program. The computational intensity of polling VISSIM is fully evident when polling all vehicles in the network. The call to *_simu-*

late_update_step() therefore only updates the program's record of signal states (in addition to advancing the VISSIM simulation a single step). By default, a single step in VISSIM translates to 0.2 simulated seconds; that is to say, for each "step" VISSIM simulates what would happen in 0.2 seconds if the simulation was a reflection of the real world. As ECO-driving proposals in the real world often only receive information on a per-second basis, to reduce computational loads it was decided to update ECO-driving speeds and estimates every five time steps (reflected in the modulo operation shown at the end of Listing 3). Even with this reduction, the ECO-driving program substantially increases a VISSIM run's operating time.

After updating the position of vehicles, new vehicles detected in the system are assigned as "normal" or "ECO-driving" vehicles using a random number generator. The proportion of vehicles following the ECO-driving logic is set according to the scenario being tested. All Vehicles following an ECO-driving regime then have their position checked. For all scenarios tested a range of 400 metre was selected as the maximum distance a vehicle could be from a traffic light before it follows an ECO-driving regime. Vehicles within 400 metres of a traffic signal then have their speed set according to the following equation:

$$\max\left(\nu = \frac{D}{t_p + t_q}\right) \qquad \text{And:} \begin{cases} t_p \in [t_g, t_r) \text{ or } t_p = t_g' \text{ if } s = \text{red} \\ t_p \in [0, t_r) \text{ or } t_p = t_g' \text{ if } s = \text{green}) \\ \nu \leqslant \nu_{\text{limit}} \end{cases} \tag{6}$$

Similar to the equation given in Section 2.6.1.3, in this equation ν is the optimised speed of the vehicle, D is the distance to the stop bar, ν_{limit} is the speed limit, t_g and t_r are the green and red times respectively of the signal and t_p is the phase time being considered for optimisation. However, unlike the equation in Section 2.6.1.3, this equation includes an additional term, t_q, which is a buffer value included to account for the possibility of a queue at the traffic signal. In this study it is assumed that no knowledge of the queue length is available, and the value for t_q also accounts for the time it takes for the existing queue to dissipate. This value was set at five seconds; if this value is set too high it can negatively affect the performance of the algorithm (as vehicles will slow down more than is required), and if it is too low the vehicle will have to slow down too much when it encounters vehicles ahead.

After determining the appropriate ECO-driving speed, the desired speed of these vehicles in VISSIM is changed to reflect this target. If no successful speed can be found using an ECO-driving approach, then no modifications are made. When a vehicle's desired speed is set in VISSIM, VISSIM adjusts the speed of the vehicle using its own internal logic.

3.7.3.2 *Integration with MOVES*

The program created to implement the ECO-driving approach interacts and complements VISSIM, but the output provided by VISSIM is the same as that described in Chapter 3 and the relationship between all the components is visualised in Figure 13. This implementation used a fully disaggregate approach, and the input into MOVES consisted of the raw vehicle trajectories. This approach, while considered the most accurate possible, added to the motivation for the development of

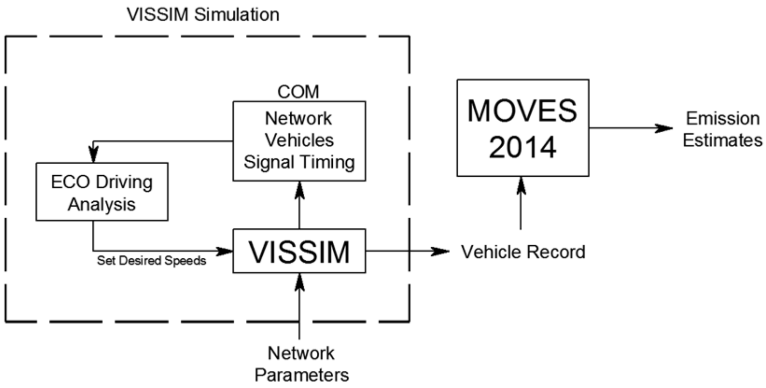


Figure 13: VISSIM's main UI

an alternative approach that would reduce computational burdens. Individual scenarios in this analysis took multiple hours to complete, with some runs in MOVES taking well over 6 hours on the computer used to estimate emissions.

Part III

RESULTS, DISCUSSION AND FUTURE WORK

The methods discussed in the previous part were then applied to arrive at emissions estimates. The results of the different integration methods were compared to each other on the basis of accuracy and computation time. For the case of the ECO-driving algorithm, the results of the ECO-driving process are compared to a base-line case to determine if the proposal has merit as an emissions-reduction strategy. Finally, this part closes with a discussion of the implications of these results and the potential for future work.

4 RESULTS: SCENARIO 1 AND 2

For scenario 1, the framework discussed in Chapter 3 was used to evaluate 6 sub-scenarios on two networks. These scenarios include a high, low and varying volume scenario for each of the two networks designed in VISSIM. The VISSIM outputs were then given as input into MOVES, generating output from five different methods on three pollutants (CO, CO_2 and Total Energy). For scenario 2, a number of networks progressively increasing in size were tested to evaluate the computational demands of the algorithm. The following sections discuss these results and their implications, starting with the results from scenario 1[1].

4.1 SELECTING AN APPROPRIATE VALUE FOR k

As mentioned in Section 3.5.3.1, a number of different values for k were explored in the analysis process. This values ranged from a minimum of 20 on the low end to a maximum of 400 on the high end. While the clustering algorithm generated input tables for MOVES for each of these values, only one of these should be selected for each scenario tested. Computational limitations mean that it is infeasible for each cluster value to be evaluated in MOVES, and doing so would defeat the purpose of the clustering. As such, it is necessary to devise a method that allows selection of the cluster count simply. In Section 2.7.1.1 a few possible methods for the selection of k are discussed, and ultimately for this project it was determined that a simpler method should be used in this evaluative phase of the integration tool, namely, the "elbow" method. As discussed in Section 2.7.1.1, this method is a heuristic that selects a value of k according to the point where a graph of the indicator coefficient and cluster count selected changes sharply. One possible indicator that can be used is the F statistic, which is the ratio of the between-group variance to the within-group variance. Ordinarily, the F statistic can be calculated using the following equation:

$$F = \frac{\sum\limits_{i} \frac{(\bar{Y}_i - \bar{Y})^2}{(K-1)}}{\sum\limits_{ij} \frac{(Y_{ij} - \bar{Y}_i)^2}{(N-K)}} \tag{7}$$

In this equation $\bar{Y}$ is the average trajectory for all vehicles in the sample and $\bar{Y}_i$ is the average trajectory for cluster i. K is the total number of clusters and N is the number of vehicles in the entire sample. Finally Y_{ij} is an individual trajectory j in cluster i. Because individual trajectories are not a single point, in actuality this

1 Most of these results will also be presented at the 95th Annual Meeting of the Transportation Research Board [46].

equation is expanded to accommodate the individual trajectory elements. Therefore, $(\bar{Y}_i - \bar{Y})^2$ is calculated using the following procedure:

$$(\bar{Y}_i - \bar{Y})^2 = \sum_{p=1}^{P} (\bar{Y}_{i,p} - \bar{Y}_p)^2 \tag{8}$$

Where p is the individual entry of the cluster average trajectory $\bar{Y}_i$ and the overall average trajectory $\bar{Y}$ and P is the total number of entries in trajectory $\bar{Y}$. Similarly, $(Y_{ij} - \bar{Y}_i)^2$ is expanded using the following procedure:

$$(Y_{ij} - \bar{Y}_i)^2 = \sum_{p=1}^{P} (Y_{ij,p} - \bar{Y}_{i,p})^2 \tag{9}$$

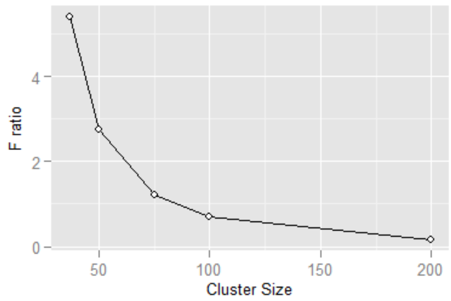
Where p now represents the individual entry of trajectories Y_{ij} and the cluster average trajectory $\bar{Y}_i$ with P being the total number of entries in trajectory $\bar{Y}_i$. As the cluster count changes, it is obvious that the result of Equation 7 will eventually reach a floor value that is close to zero. This occurs because the value of $N - K$ begins to approach zero when the cluster count approaches the value of the number of total data points, causing the denominator of the equation to tend towards infinity.

4.1.1 *Results for All Scenarios*

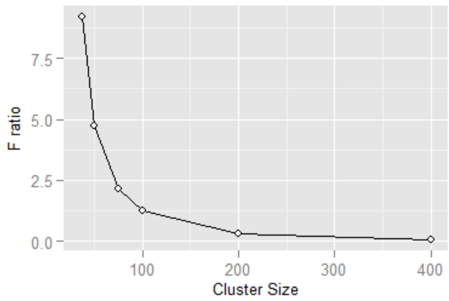
For all clustered results, the F ratios calculated according to Equation 7 was also done. The resulting graphs were then plotted using R, a statistical software package to estimate the optimal value for the cluster count.

4.1.1.1 *Results for Network 1*

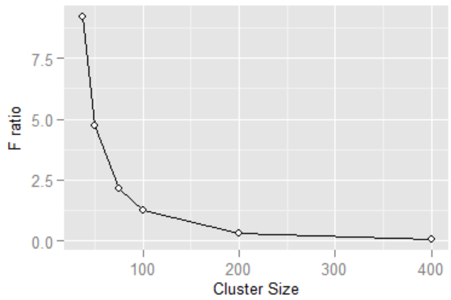
Because the volumes of Network 1 and Network 2 are different the optimal values for the cluster count may be different between the two networks. The complexity of the clustering analysis is influenced firstly by the number of individual vehicle trajectories in the network (as these are the elements clustered) and secondly by the length of those trajectories (as these increase the difficulty of finding a solution to the clustering problem). In the case of Network 1, the resulting F ratios for each of the different cluster counts and volume scenarios is shown in Figure 14. By applying the heuristic discussed in Section 4.1, values of 100 were selected for the optimum cluster count in the Medium and Variable volume scenarios (Figure 14b and Figure 14c respectively) while a value of 75 was selected for the Low Volume scenario (Figure 14a). These values all correspond to the location in the graphs marked by a sudden change in the rate of decrease of the F ratio. It is logical to expect a lower value for k in the low volume scenarios as the number of vehicles to cluster will be lower. Although the variable and medium volume scenarios are different, the total number of vehicles running through the scenarios is similar, so the resulting optimal cluster count is similar.



(a) Low Volume



(b) Medium Volume



(c) Variable Volume

Figure 14: F Values from Network 1

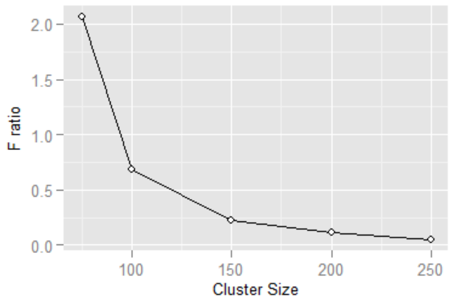
4.1.1.2 *Results for Network 2*

Finally, the same process can be applied to Network 2, with the results shown in Figure 15. Similar to Network 2, the medium volume scenario and variable volume scenarios have similar optimal cluster counts, and an application of the heuristic resulted in the selection of a value of 150. In contrast, a cluster count of 100 was selected for the low volume scenarios. The increased complexity of Network 2 when compared to Network 1 has likely resulted in the higher cluster counts, as both networks have similar volumes. Once the optimal value of k was selected using this heuristic, the input files associated with those values were then used as inputs in MOVES to generate the results discussed in the rest of this chapter.

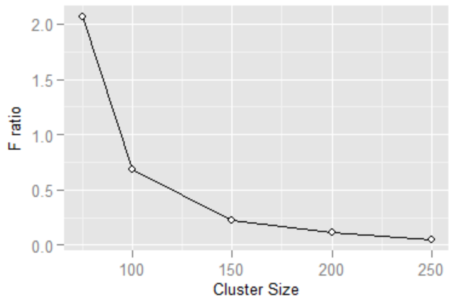
4.2 RESULTS FROM NETWORK 1

When the process described in Chapter 3 is applied to network 1 (shown in figure Figure 11a, the results shown in Figure 16 are obtained. These results include emissions estimates from MOVES on CO_2 and CO emissions as well as the total energy consumed by vehicles on the network. In these results the estimates derived from using VISSIM's un-aggregated vehicle record (see Section 3.5.1 is highlighted in red. Similarly, the results from medium and variable volume scenarios can be seen in Figure 17 and Figure 18 respectively. the values used to generate these graphs and the percentage difference between each method and the individual trajectory method have also been included in the Appendix. The data tables in

(a) Low Volume



(b) Medium Volume

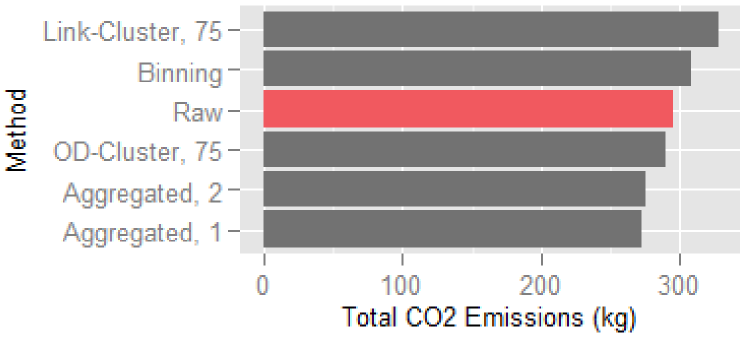


(c) Variable Volume

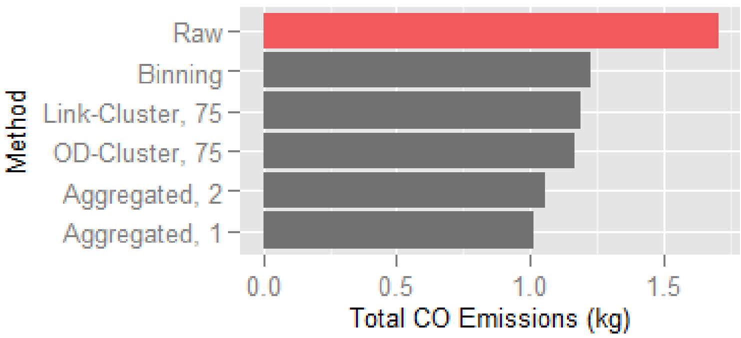
Figure 15: F Values from Network 2

that section, particularly those highlighting the difference between the estimates provided by each method are better at highlighting an interesting facet of the data: an underestimation of the fuel consumed provided an identical underestimation of CO_2 emissions. For example, if the VB approach estimates a fuel consumption that is 4.4% higher (see Table 3) than when individual trajectories are used, then it will also estimate CO_2 emissions that are 4.4% higher. This is not to say that these estimates are *identically* higher or lower. Indeed, for this particular example, but for all practical purposes the difference is very minute. This was always the case when comparing the different methods to each other (but not always the case in every analysis conducted using MOVES). The reasons behind this difference are not clear, especially since CO estimates are dramatically lower for all methods, however as the majority of a vehicle's emissions will be CO_2, it does make sense to some degree that an estimate of lower fuel consumption will correspond to a similarly lower estimate of CO_2 emissions.

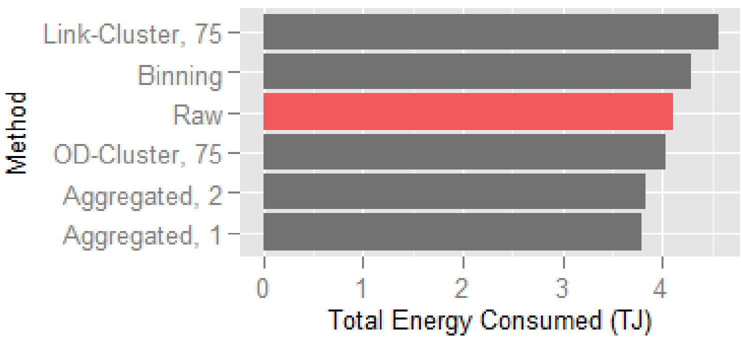
In general, the results of this network show that the clustering and VB approaches have similar performances in terms of accuracy, as for all volume scenarios their estimates for CO_2 emissions and total energy consumed were within 5% of the estimates provided when raw trajectories were used. When compared to each other, the VB approach always gave higher estimates than the clustered approach, and, with the exception of CO emissions, gave higher estimates than when individual trajectories were used. In all cases, pre-processing vehicle trajectories and subdividing them by links resulted in higher emissions estimates than



(a) CO_2

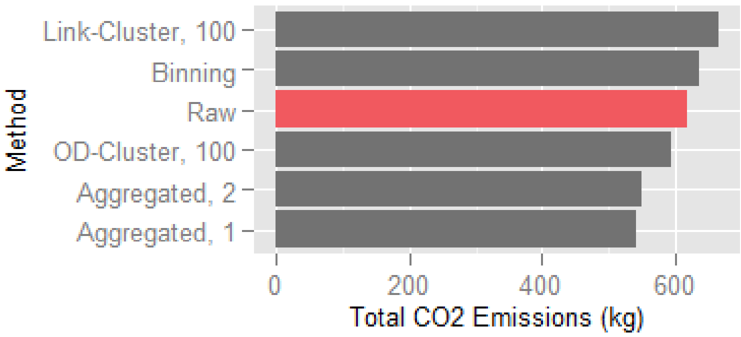


(b) CO

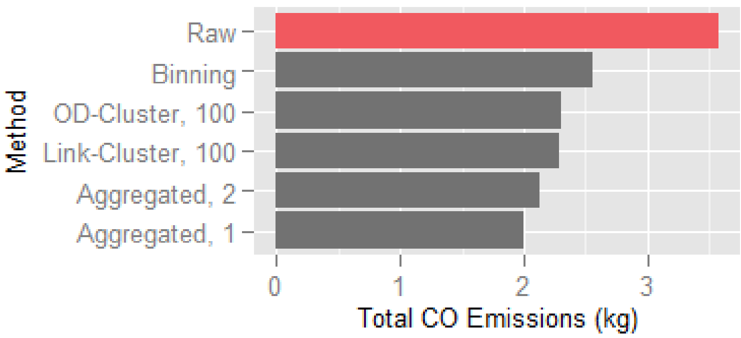


(c) Energy

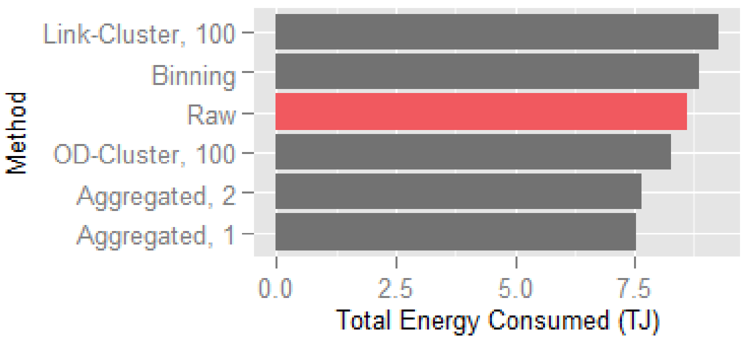
Figure 16: Results from Network 1, Low Volume Scenario



(a) CO_2

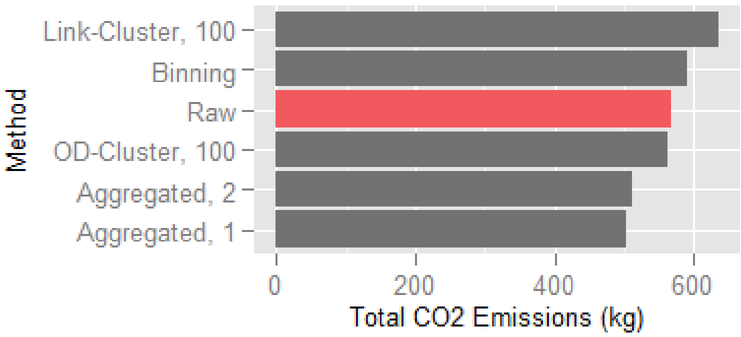


(b) CO

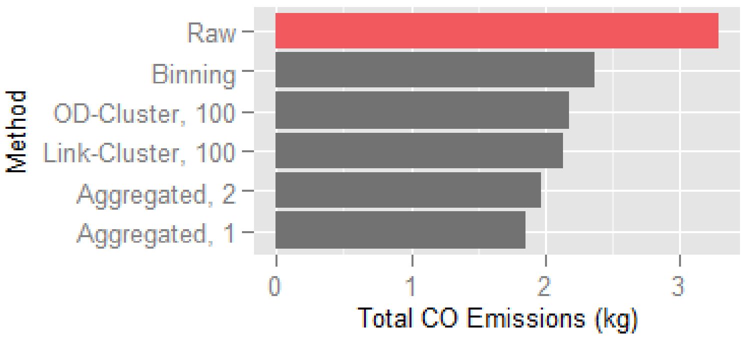


(c) Energy

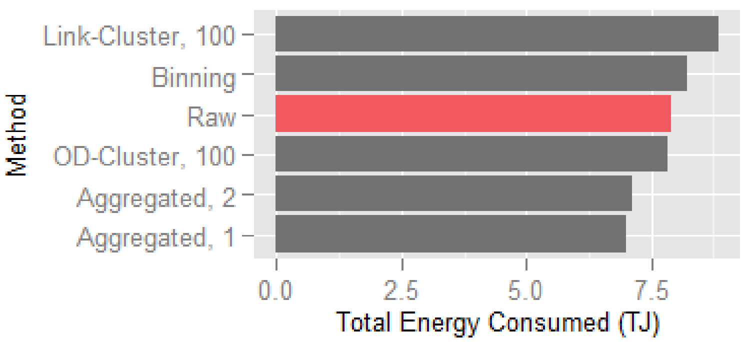
Figure 17: Results from Network 1, Medium Volume Scenario



(a) CO_2



(b) CO



(c) Energy

Figure 18: Results from Network 1, Variable Volume Scenario

when O-D pairs were used. For CO_2 and Energy estimates this makes a link-based approach less accurate.

The AS approaches performed the poorest of all the approaches tried, though to some extent this was expected. Estimates of CO_2 emissions and total energy consumed from these approaches was always 5% below those provided when individual trajectories were used, and for some volume scenarios were up to 12% below those estimates. In general the AS approaches performed similar to each other; however, the increased disaggregation provided by the second method appears to have increased the accuracy of the approach to a certain extent. In all cases in Network 1, the second method (modelling each leg of an approach separately) provided estimates that were closer to those when individual trajectories were used than the first method.

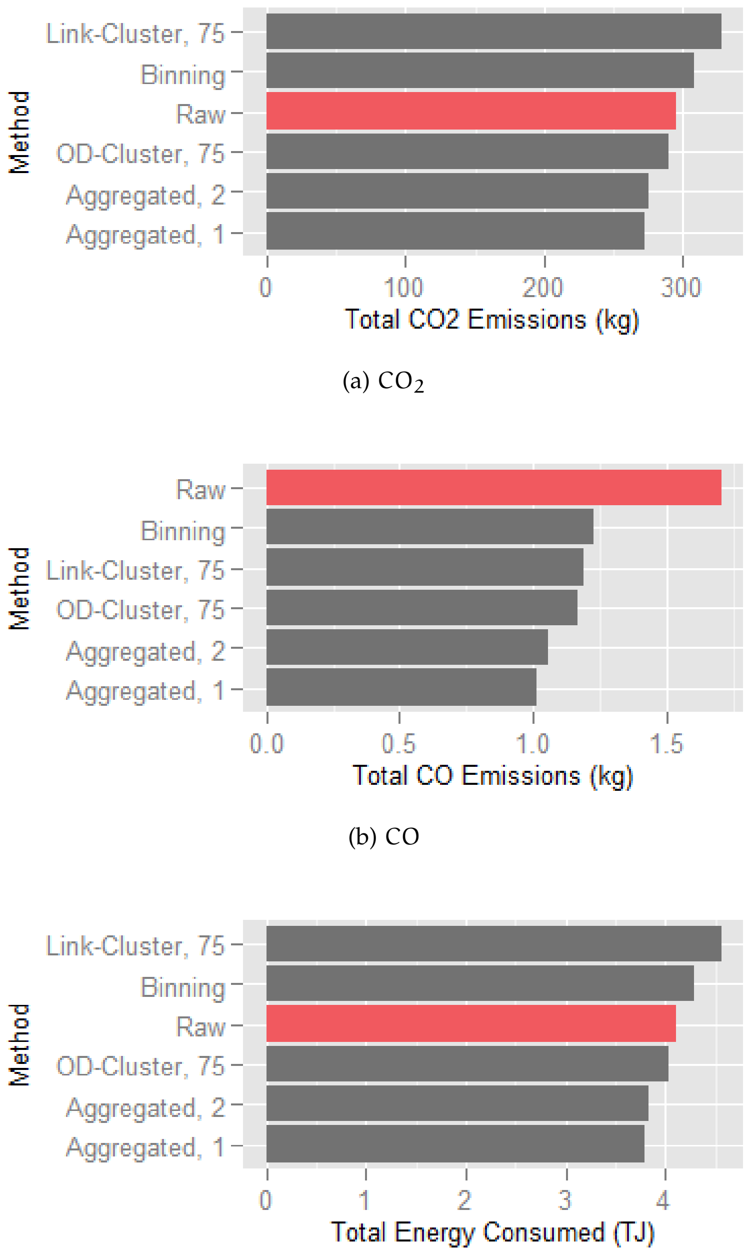
All methods generally substantially under-estimate CO emissions. The reasons for this under-estimation are not immediately apparent and more investigation is necessary to determine why this occurred. Previous studies [24] have shown that CO emissions are highly sensitive to the manner in which vehicles are operated in, and since these emissions are orders of magnitude below those of CO_2 emissions they appear to react more sensitively when any aggregation is performed. Surprisingly, despite providing estimates for CO_2 emissions and Energy consumption that were greater than those of the VB approach, a link-based approach to clustering provided similar estimates of CO emissions to the O-D approach.

The effect of volume on emissions estimates is generally not substantial on the accuracy of most methods, except for the cases of the AS approaches. As can be seen by the figures discussed previously, AS approaches under-estimated emissions more severely when the volume was higher (see Figure 17 and Figure 18) than when the volume was lower (see Figure 16). This suggests that the default operating mode distribution MOVES uses when estimating emissions based on average speeds may not consider the effects of traffic well.

4.3 RESULTS FROM NETWORK 2

Although somewhat similar, the results from Network 2 (shown in Figure 19, Figure 20, and Figure 21) possess a number of noticeable differences. In contrast to the results of Network 1, estimates for CO emissions for all methods are generally closer to those provided when individual trajectories are used. This is particularly true for the VB approach where for the medium and variable volume scenarios the estimates provided were around 6% higher when individual trajectories were used. This is markedly different than the estimates in Network 1, as all methods provided estimates that were more than 15% below those when individual trajectories were used.

Although the estimates for the low volume scenario were all similar to each, the case is dramatically different for the medium and variable volume scenarios on this network. For these volume scenarios, the cluster approach continued to provide an estimate for CO_2 and total energy consumption that was similar to the estimates provided by the individual trajectories, but all other approaches had mixed or poor performance. While the VB approach's estimates for CO emissions were the closest, the approach substantially over-estimated emissions of CO_2 and

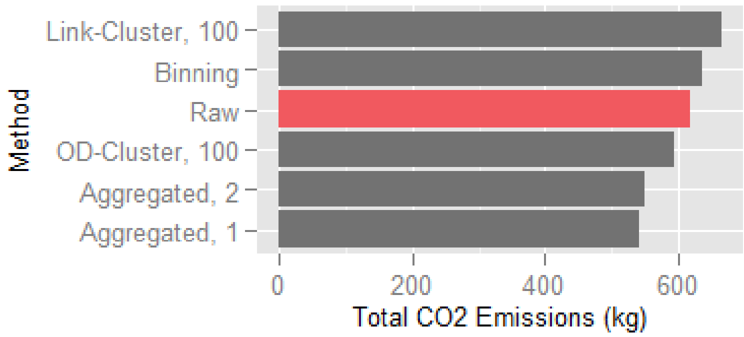


(a) CO_2

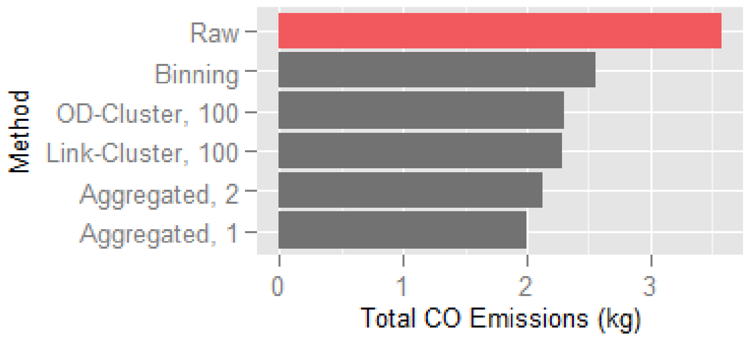
(b) CO

(c) Energy

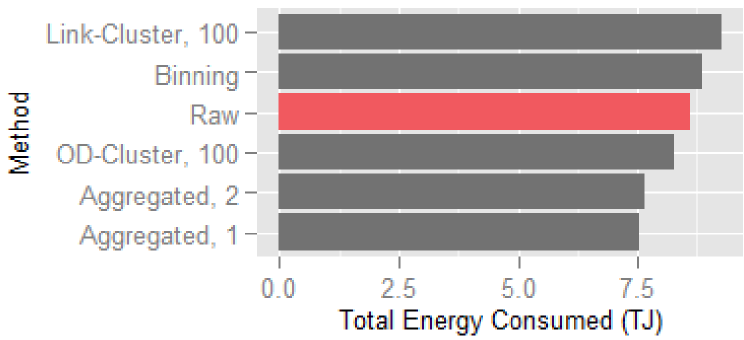
Figure 19: Results from Network 2, Low Volume Scenario



(a) CO_2

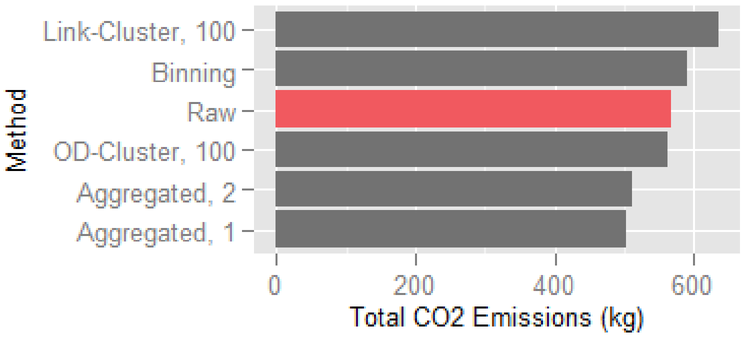


(b) CO

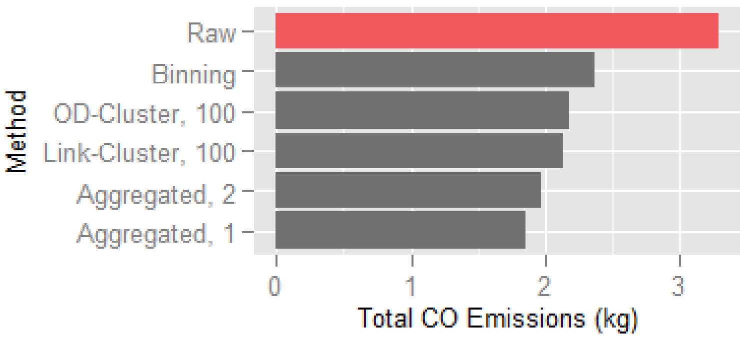


(c) Energy

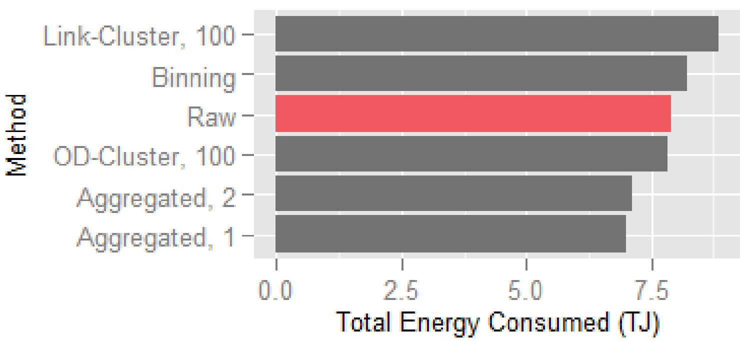
Figure 20: Results from Network 2, Medium Volume Scenario



(a) CO_2



(b) CO



(c) Energy

Figure 21: Results from Network 2, Variable Volume Scenario

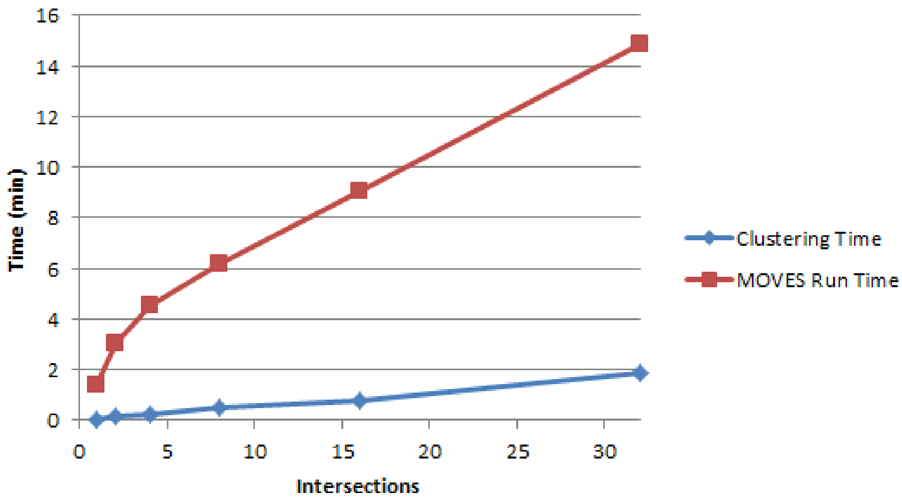
total energy consumption for these two volume scenarios. As was the case for network 1, the aggregated approaches performed similarly to each other, and the second method generally had estimates that were closer to those of the individual trajectory method. The high accuracy the AS methods demonstrated in the variable volume scenario is interesting, and it is apparent that the default operating mode distribution for this method lends itself well to this scenario, where a certain amount of congestion occurs on a larger network but is not severe or excessive. The variable performance of AS methods however mean they are unreliable in situations where an accurate estimate is required.

As was the case in Network 1, the estimates provided from the data-clustering algorithm were generally similar to those of the individual trajectory method (within 6%) for CO_2 and total energy consumption, but were not very accurate for CO estimations. Again, a link-based approach to trajectory clustering resulted in higher estimates of CO_2 emissions and Energy consumption, despite providing an estimate for CO emissions that was similar to the O-D approach.

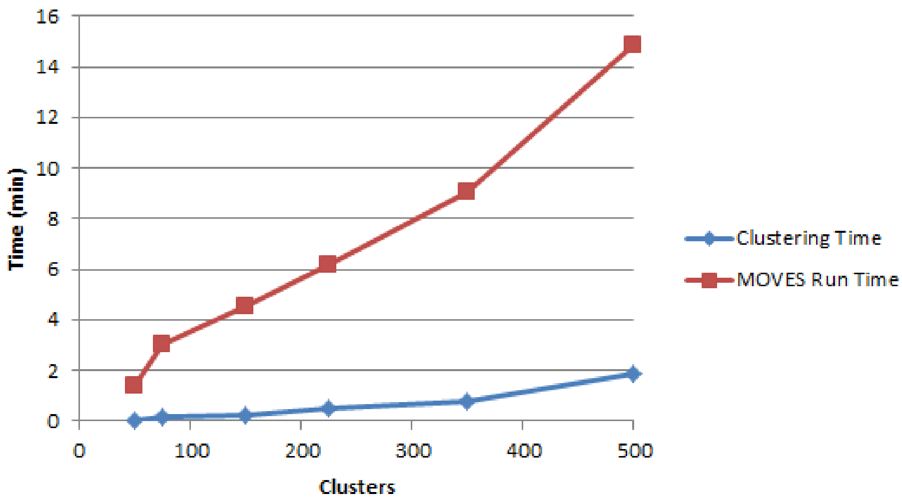
4.4 results of scenario 2

As described in Section 3.2, the second scenario was designed to evaluate the computational scaling of the algorithms. As the clustering algorithm is iterative in nature, the time it takes to generate output can be somewhat variable. This variability increases with both cluster count and volume. To evaluate this variability, a series of networks were tested solely with the clustering algorithm. The networks tested had sizes ranging from 1 intersection to 32 intersections. The results of the clustering analysis were also used to obtain emission estimates in MOVES, solely for the purpose of evaluating the time the subsequent estimation in MOVES would take. The results can be seen in Figure 22.

It is important to note that, in this figure, the cluster count (k) is also increasing with network size, ranging from 50 for the smaller network to 500 for the larger network. The time of the subsequent estimation in MOVES is most strongly influenced by this number, as each additional cluster is an additional link, including a drive schedule, that must then be modelled in MOVES. In general, the model estimation time scales linearly. The estimation time in MOVES forms the majority of the time required for a successful analysis, and the time required for this analysis increases faster than the time required to generate the clustered inputs. On the balance, this means that the time savings when compared to the use of individual vehicles will be more pronounced from a proportional perspective as the network size increases. In general, the time required to generate the clustered input is loosely tied to the cluster count; however, if the value of k is too high or too low, then the algorithm will have difficulty converging, and estimation time will be high. For optimal values of k, much of the program's time is spent parsing the vehicle trajectory data. As the program subdivides all vehicles according to their trajectory as vehicles are parsed, there is generally no difference in estimation time between the O-D based approach and the link based approach. Separate trials conducted on the 32 intersection network with different values of k did not differ substantially in estimation time, and from a user's perspective the majority



(a) Time scaling by Intersection Count



(b) Time scaling by Cluster Count

Figure 22: Execution Times of the Clustering Process

of the time required for a successful estimation will come from data preparation, including configuring MOVES and VISSIM.

In terms of MOVES execution times, for these networks the estimation time varies more strongly with volume, which in the case of cluster counts is directly related to the value of k. A higher value of k increases the execution time in MOVES in a roughly linear manner. Although the network increases in length (resulting in larger vehicle trajectories), in this test network the volume does not change substantially from one scenario to the next (only vehicles on the corridor are modelled). As such, when un-clustered individual vehicle trajectories are used in MOVES, the run times do not vary substantially from run to run. Using individual trajectories generally results in an execution time of 40 to 45 minutes for this network (about 1200 unique vehicles pass through the network over the analysis time). The volume-based trend in execution times shown in Figure 22 appears to be linear in nature, at least for the network tested here; and so, if the value of k (or the total volume) were doubled the execution time in MOVES would be similarly doubled. In contrast, doubling the trajectory length does not necessarily result in a doubling of the execution time, as the increased network size did not drastically increase computation time when compared to the effect of volume.

These results support the rationale that data clustering can be used to dramatically reduce computational time for a successful MOVES run. A data clustering approach to the problem affords the modeller more control over the computation time (e.g. by specifying a low value for k) and scales better with network size than MOVES does. Although these tests are limited, larger scale tests will be conducted in future research work. The constraints of micro-simulation can ultimately more of a concern than the data clustering process, as developing test networks in VISSIM can require substantial time and data.

4.5 DISCUSSION

Although the VB approach had estimates that were sometimes better than those of the clustering algorithm, their results (much like those of the AS approaches) were variable and influenced by the dynamics of the simulation. Since the VB approach still uses average speeds to arrive at estimates for emissions, the method still depends on a default operating mode distribution; as such, when simulated conditions such as volume change substantially the accuracy of the estimates also appears to change. This effect is most prominently visible in the difference between the estimates provided in Network 2's variable volume scenario (see Figure 21) and low volume scenarios (see Figure 19). In the low volume scenario each method produces estimates for CO_2 emissions that are lower than those provided when individual trajectories are used. However, in the variable volume scenario, the AS and VB approaches (which are based on the use of average speeds) instead provide emissions estimates that are now higher than those of the individual trajectory method. In contrast the clustered approach provides consistent performance regardless of the volume or network configuration, with CO_2 and total energy estimates always no more than 6% below those provided when individual trajectories are used.

Despite this, the AS methods can still be useful in situations where accuracy is not as important but where computational time is a concern. Excluding preparation time, a successful run of MOVES that uses individual trajectories ranged from 30 minutes for low volumes on the simple network to 2 hours for the variable volumes on the larger network. In contrast, the AS and VB approaches took less than 5 minutes to run after the data and input was prepared while the clustering approach took between 5 to 10 minutes to run in MOVES after data and input were prepared. All methods require some data preparation before input into MOVES, though the integrated tool developed for this project simplifies the task extensively for the individual trajectory and clustered approaches. Despite this, due to the need to decide a value for k, inputs for the clustered approach require additional time to prepare (between 5 to 15 minutes per scenario) though this could easily be reduced through improvements to the integrated tool.

In terms of pre-processing before data clustering, it is clear that the link-based approach provides estimates that are often less accurate than the O-D approach. This is likely due to the manner in which the network was designed in VISSIM. The use of connector links to join approach legs of an intersection mean that portions of a vehicles acceleration and deceleration profile are split across multiple links. This means that a full and complete acceleration profile from rest to cruise speed is never modelled in MOVES as one continuous trajectory, but rather as two separate pieces, with the small connector holding half of acceleration and deceleration profile as well as many vehicle's at rest portions. Future implementations of this algorithm may benefit from a trajectory analysis approach whereby vehicle trajectories are first separated into segments representing various operating modes, and then clustering can be run on these segments.

The results of this analysis demonstrated that the proposed clustering system has the potential reduce computational requirements while maintaining a consistent level of performance and accuracy. The clustered approach had the most consistent performance of all the methods tested (when compared to the use of individual trajectories) and has the potential to reduce computational times on larger more complicated networks. Additional improvements are still possible, with further investigation of the method's performance with other pollutants and by improving its ability to provide accurate estimates for minor pollutants such as CO.

5 RESULTS: SCENARIO 3

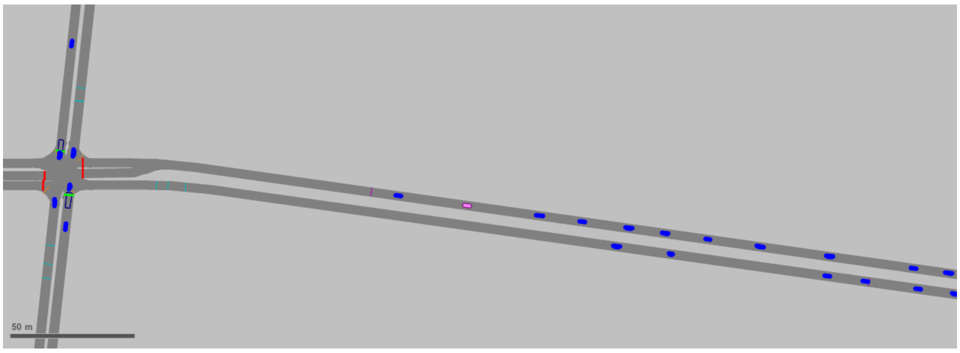
The final scenario evaluated was an application of the developed framework to analyse an ECO-driving system. It includes four sub-scenarios developed to evaluate different aspects of the system on the tested corridor. Each sub-scenario evaluated a different market penetration rate. The primary goal of the evaluation was to understand of the benefits of an ECO-driving system on a real-world network with the secondary goal of evaluating the effect of market penetration on emissions[1].

5.1 PRELIMINARY OBSERVATIONS OF ECO-DRIVING BEHAVIOUR

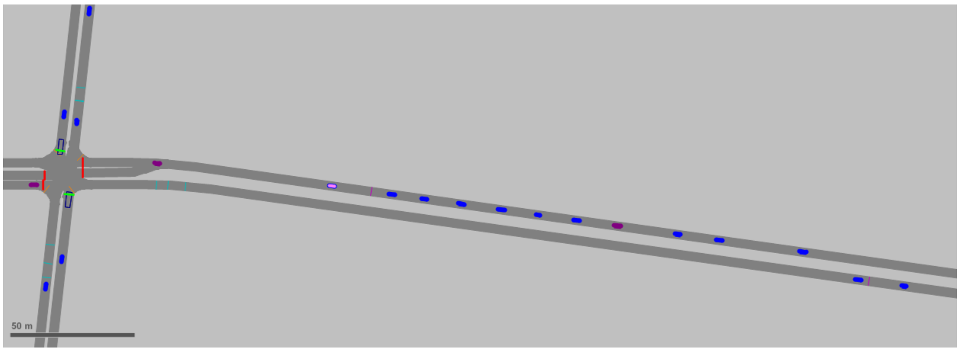
During the development of the ECO-driving algorithm, a number of cursory observations were conducted in VISSIM while the ECO-driving algorithm was running. These observations were conducted with the goal of verifying that the coded algorithm was working and could successfully affect the travel behaviour of vehicles. During the development of the algorithm, this step was important, especially as a complete run of the algorithm for a one hour simulated period could take multiple hours in the real-world and a subsequent analysis in MOVES would take even longer. The screen captures shown in Figure 23 were observed from the VISSIM simulation and highlight the operation of the algorithm. In Figure 23a the second vehicle on approach to the intersection in the westbound direction was randomly selected to operate as an ECO-driving vehicle. This vehicle is highlighted in the figure as a magenta vehicle. Upon entering the control area of the ECO-driving system, the vehicle immediately slows down as the current signal facing it is red. After a few seconds have passed, we see that in Figure 23b the vehicle ahead of the ECO-driving vehicle has almost arrived at the intersection's stop bar and is deceleration (indicated by the magenta colour in VISSIM) to stop at the red light. The ECO-driving vehicle has successfully reduced its speed and by the time the conflicting direction's signal has changed to amber we see that a substantial queue of vehicles also "following" an ECO-driving algorithm has formed behind the original ECO-driving vehicle. Finally, by the time the signal changes to green in Figure 23d the vehicle has successfully cleared the intersection and resumes normal operation.

While this image sequence highlights one of the better observations of the algorithm's operation there were also many cases where the operation was not so successful. For example, in some cases where vehicle queues waiting at the intersection were long, vehicles following ECO-driving speeds are blocked by the dissipating queue and forced to slow down or stop completely. In the absence of incorporating queue length into the algorithm, the frequency of this behaviour is controlled largely by the buffer term in Equation 3.7.3.1.

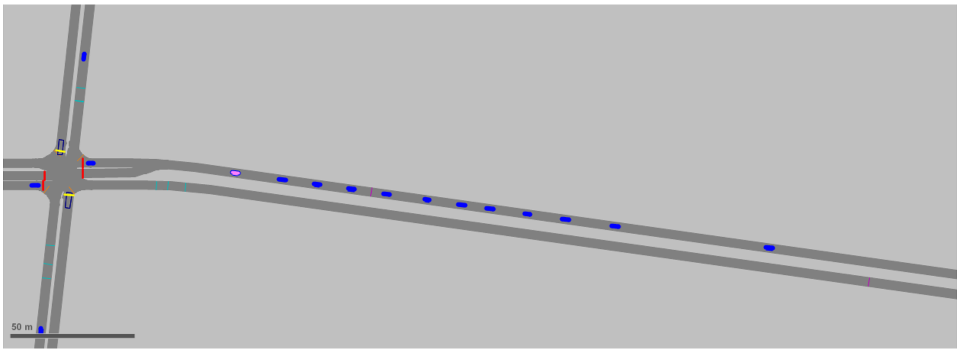
1 The results of this analysis were presented at the 2015 Annual CSCE Conference [45].



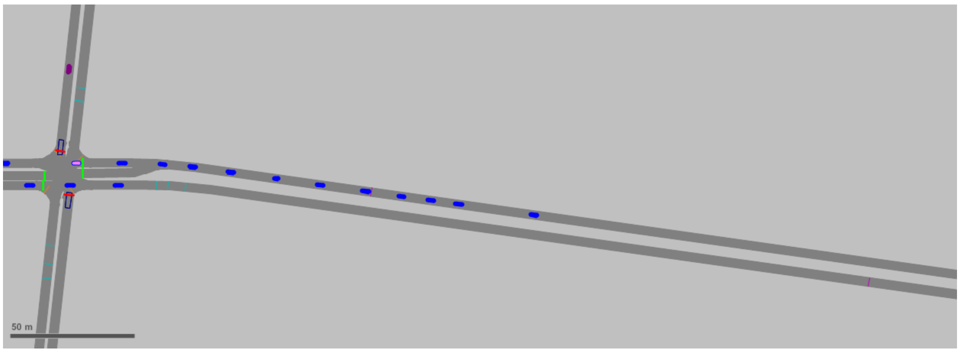
(a) On approach to intersection, just after the algorithm has started running



(b) Vehicle slows down, a queue begins to form behind it



(c) The opposing signal changes to yellow and green is imminent



(d) ECO-driving algorithm completes successfully and vehicle resumes normal operation

Figure 23: ECO-driving, Simulation Example

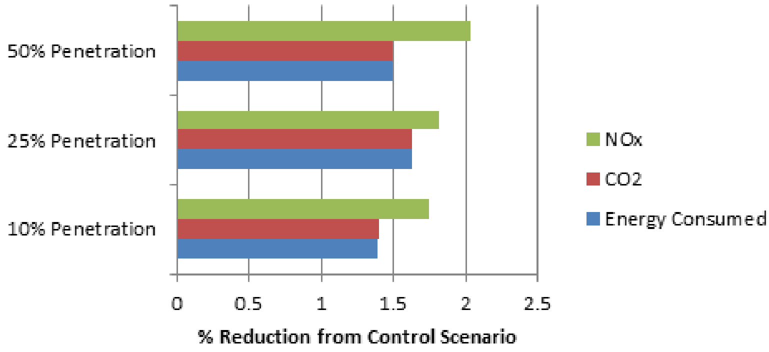


Figure 24: Emissions Reductions by Market Penetration

5.2 RESULTS OF THE ECO-DRIVING ANALYSIS

After successfully evaluating and developing the ECO-driving algorithm, the results of the VISSIM simulation were analysed and input into MOVES. For this analysis, the most disaggregate approach possible was selected and individual vehicles were modelled separately in MOVES. As the effect of ECO-driving is subtle, this approach was selected to ensure that the variations in driver behaviour could be fully captured. This analysis was also conducted before rigorous development and evaluation of the clustered analysis discussed Chapter 3 was completed, and the exceptionally high computational demands observed in this analysis were a primary motivating factor for the development of an alternative approach.

As per the methodology described in Section 3.7.3, estimates from MOVES were obtained for four scenarios, three with penetration rates of 10%, 25%, and 50% respectively. To minimize computational burdens, estimates were only made for two pollutants (CO_2 and NO_x) as well as total fuel consumption. The results of the analysis are summarized in Figure 24, and complete data tables have been included in Appendix A.

While many of the studies reviewed in the literature have cited emissions and fuel consumption rates reductions in excess of 4%, the results of this analysis are somewhat more muted, ranging from 1.3% to 2%. This reduced performance stems largely from the fact that vehicles are quickly able to form platoons and capture "green waves" that allow them to stop only at one or two traffic signals. Additionally, the effect of the buffer term in Equation 3.7.3.1 likely has some effect on the process as well. Many of the studies conducted in Section 2.6.1 do not deal with the issue of queues, either neglecting to mention them or ignoring their effect as it only becomes pronounced as volume increases. Indeed, the authors of one of the studies mentioned in Section 2.6.1 (Xie et al) indicate that as congestion becomes an issue on a network, the benefit of an ECO-driving system decreases as vehicles encounter unexpected acceleration/deceleration [65].

Despite these limitations, in general, a corresponding reduction in fuel consumption is also followed by a reduction in CO_2 emissions by approximately the same amount. NO_x emissions were observed to decrease more substantially, and these reductions are caused by the decreased incidences of sharp accelerations or decelerations which affects minor pollutants more strongly. The effect of varying the penetration rate was mixed, and conclusive statements cannot be made about its effect. As penetration rate increased from 10% to 25%, emissions and fuel consumption decreased, but only marginally. However, when the penetration rate was increased to 50%, CO_2 emissions and energy consumption increased instead (though NO_x emissions decreased further). While this result is somewhat counter-intuitive, it is important to note that ECO-driving vehicles that follow other ECO-driving vehicles will have a substantially reduced effect on system-wide emissions, as if they would already follow and ECO schedule if the leading vehicle modifies its behaviour. Additionally, the effect of the buffer term shown in Equation 3.7.3.1 could be detrimental in this case. This term will lead to increased separation between vehicles since ECO-driving vehicles following other ECO vehicles will follow their own speed advice, which could be lower due to the buffer term, rather than maintain the minimum following distance.

5.3 DISCUSSION

More research is still needed and additional scenarios should be run to further investigate some of the aspects of an ECO-driving system's operation revealed in this analysis. The buffer term was largely set through cursory examinations of the network's operation and an estimation of the average queue dissipation time at all intersections, and may not be set at an optimal value. In investigation into incorporating queue length estimates is also worthwhile and could be implemented in VISSIM to test theoretical operational characteristics. Other parameters of the simulation were also set using values gleaned from the literature, including the maximum range of the ECO-driving system. Finally the design of the network is restrictive as it includes mostly two lane roads. Although this may be the case in real-world networks, such networks also include multi-lane roadways where vehicles are able to pass an ECO-driving vehicle and thereby reduce the effectiveness of the system.

6 FINAL DISCUSSION AND FUTURE WORK

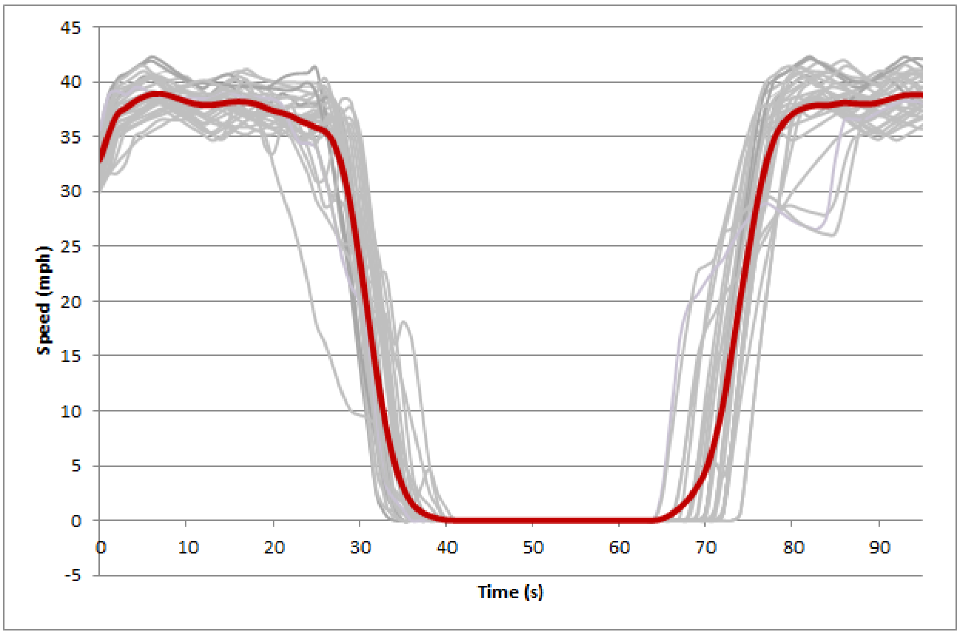
6.1 MAJOR FINDINGS

At the onset, this research's primary aim was to develop a system that can be used to integrate the VISSIM traffic microsimulation platform with the MOVES emissions estimator. The subsequent survey of the literature and initial attempts revealed that there are a variety of different ways to accomplish this linkage and each of them have disadvantages in terms of computational requirements, data requirements, and accuracy.

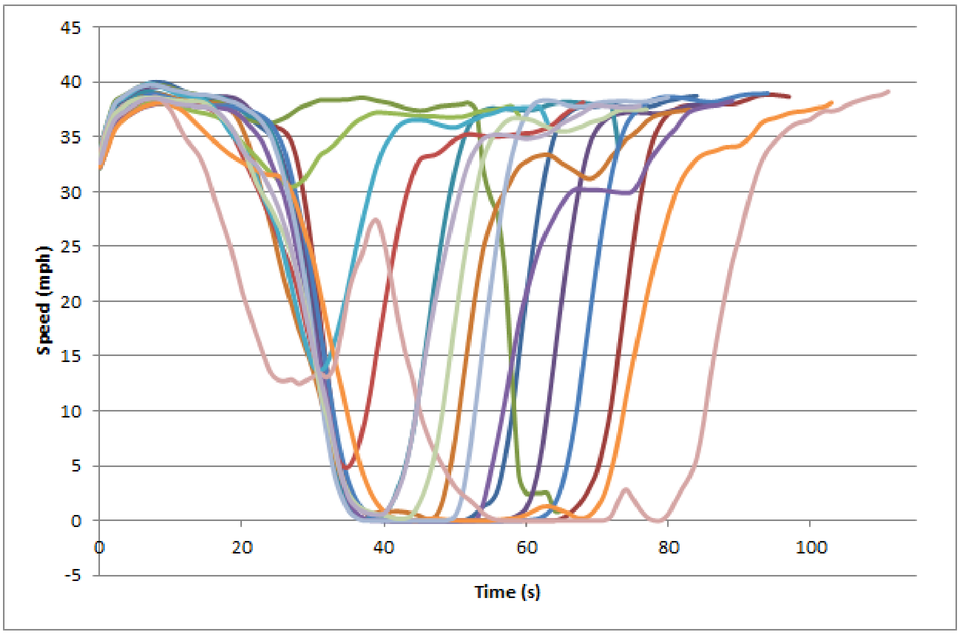
This research was successful at comparing the three major categories of approaches, namely, aggregated approaches, hybrid approaches, and disaggregate approaches. In the literature reviewed, this study is the first to use individual vehicle trajectories in a MOVES-based project-level analysis of a multiple signal intersection and is also one of the few that specifically compares the differences in the estimates that each method provides. Using individual vehicle trajectories in MOVES is very useful at a small scale, and, while mentioned in guides put out by the US EPA, it has seen few applications in the literature.

Although it provides MOVES with substantial data, the use of individual trajectories in MOVES results in excessive computational burdens, especially on large networks. The clustering algorithm proposed in this study succeeds in creating a compromise that captures some of the aspects of driver behaviour while also reducing computational burdens. This proposal is one of the first of its kind in the literature, and applies aspects of data clustering to a trajectory-based context. k-means algorithms such as the one proposed in this study have been successfully applied to solve vehicle trajectory-based problems in the past, but the proposals in this study represent a novel application that uses them to simplify an analysis and reduce computational burdens.

The algorithm's cluster estimates are also generally logical, especially when one considers the parameters of the problem: namely, grouping vehicles which may have drastically different travel patterns together. For example, Figure 25 shows the resulting cluster estimations for one of the scenarios tested on the single intersection network. In Figure 25a we see the individual vehicles included in the estimation of a particular cluster's average trajectory. Although some isolated vehicles deviate noticeably from the cluster's estimated trajectory, the trajectories in this group are very similar. Of particular note, however, is the grouping of vehicles travelling in a somewhat wide range of speeds (vehicles included often deviate from the average trajectory's speed by as much as 5 mph). This means that some of the detail of driver behaviour, particularly some of the traits of aggressive driving (e.g. consistently exceeding the speed limit), are lost when the clustering algorithm is run. Despite this, some variation is still captured, as the other clusters representing this O-D pair also include regions where the speed is higher or lower than this example (see Figure 25b). It is important to note that as time-speed trajecto-



(a) Vehicles present in a sample cluster



(b) Clusters present in an O-D pair

Figure 25: Cluster Compositions

ries, these trajectories are not all the same length, and trajectories where the signal does not impede the vehicle's movement will be noticeably shorter. The clustering algorithm captures the variation of arrival at different phases of a signal's red timing, and multiple different durations of "zero speed" explain the majority of the differences between individual clusters.

The overall analysis has also demonstrated that while AS approaches can be unreliable, the amount of data and computational time required to apply them is substantially lower than all other methods. Hybrid approaches such as the VB approach or the clustered approach require additional data, but are not as computationally intensive as direct use of individual vehicle trajectories. When an analysis is conducted using a traffic micro-simulator, complete information is often available and often the only obstacle is computational power and time. As such, hybrid approaches represent a compromise that still utilises the increased power of a micro-simulation model while also limiting the computational power required to run an evaluation.

Finally, the developed framework has many different potential applications, and has the potential to assist policy makers evaluating competing proposals and new technologies. CV technologies are an emerging field and generate substantial interest in both the Transportation and wider community. The results of this study show that an ECO-driving system built on CV technologies has the potential to reduce emissions, even if such systems are not equipped on all vehicles.

6.2 LIMITATIONS AND FUTURE WORK

While the integrated tool developed has been successfully tested in this research, some improvements could still be made. As a functional tool, additional attention is required in the development of its UI. As a consequence of some programming design decisions one major disadvantage of the UI is that users are unable to interact with it while it is running. Although it will provide updates to the user through a text-box, the window itself freezes and cannot be moved. This can lead users to conclude the application is not working correctly. This issue can be fixed by improving the way the programmed GUI interacts with the core components of the program. Although the integrated tool calculates the F-statistic value, a researcher must still run the tool multiple times and change the cluster count manually to obtain a representative sample. This creates duplicate tables and additional unneeded computational burdens, as tables will also be generated for the raw input. No options are currently provided to disable certain output components.

While on the subject of GUIs, it is also worth mentioning that no GUI was developed for the ECO-driving algorithm, and the creation of one would be required before such a tool could be used by other researchers. This tool was developed as a console-based application, and it could potentially be incorporated into the tool designed to integrate VISSIM and MOVES.

Beyond cosmetic issues, the analysis conducted also has a number of other limitations. The use of a heuristic in the k-means approach means that the value of k selected could vary between different modellers. Within the literature, applications of the k-means approach often apply some form of a heuristic, however the reliance of the algorithm on a user's discretion could be removed. Future implementations will therefore work to improve this aspect of the clustered approach. The results analysis also revealed potential issues with the link-based approach to pre-processing the data before clustering. As mentioned, full and complete acceleration profiles that include vehicles going from rest to cruise speed were never modelled as a continuous trajectory in MOVES when using a link-based approach because of the way the network was coded in VISSIM. Future implementations of this algorithm would benefit from a different approach, such as the implementation of a trajectory analysis algorithm that partitions trajectories according to their operating regime. This approach may also remove oddities introduced by MOVES's dependence on speed-time profiles, which creates trajectories of unequal length that results in some clusters being used solely to capture operational variations (such as vehicles stopping for shorter times due to arriving at a signal later) rather than vehicle operation characteristics.

In the ECO-driving analysis, more research is needed, including a full sensitivity analysis of all the major factors affecting the models. The analysis conducted in the study has only focussed on a few aspects: market penetration and network geometry in relation to previous studies. Other factors, some of which have been explored in other studies, include assessments of the effect of communication's range, volume, queue length, the number of lanes on a link and the speed limit.

Part IV

APPENDIX

A

DATA TABLES

The following data tables in this chapter are directly aggregated from MOVES for each scenario. Data tables present here include those from the evaluation of the integrated tool discussed in Chapter 4 and ECO-Driving discussed in Chapter 5

A.1 DATA TABLES FROM THE INTEGRATED TOOL

METHOD	CO (KG)	CO_2 (KG)	ENERGY (MJ)
Aggregated, 1	1.02	272.35	3790
Aggregated, 2	1.06	275.70	3836
Binning	1.23	308.60	4294
OD-Cluster, 75	1.17	289.09	4023
Link-Cluster, 75	1.19	327.57	4558
Raw	1.71	295.53	4112

Table 3: Network 1, Low Volume Results

METHOD	CO (KG)	CO_2 (KG)	ENERGY (MJ)
Aggregated, 1	2.00	541.11	7529
Aggregated, 2	2.13	550.79	7664
Binning	2.56	637.96	8877
OD-luster, 100	2.31	594.28	8269
Link-Cluster, 100	2.29	666.43	9273
Raw	3.58	617.95	8599

Table 4: Network 1, Medium Volume Results

METHOD	CO (KG)	CO_2 (KG)	ENERGY (MJ)
Aggregated, 1	1.86	503.65	7008
Aggregated, 2	1.97	512.36	7129
Binning	2.37	589.87	8208
OD-Cluster, 100	2.18	562.10	7821
Link-Cluster, 100	2.14	636.33	8854
Raw	3.30	569.57	7925

Table 5: Network 1, Variable Volume Results

METHOD	CO (%)	CO_2 (%)	ENERGY (%)
Aggregated, 1	−40.48	−7.84	−7.84
Aggregated, 2	−38.08	−6.71	−6.71
Binning	−28.03	4.42	4.42
OD-Cluster, 75	−31.53	−2.18	−2.18
Link-Cluster, 75	−30.28	10.83	10.83

Table 6: Network 1, Low Volume Per Cent Difference from Individual Trajectory Estimates

METHOD	CO (%)	CO_2 (%)	ENERGY (%)
Aggregated, 1	−44.04	−12.43	−12.43
Aggregated, 2	−40.56	−10.87	−10.87
Binning	−28.32	3.24	3.24
OD-Cluster, 100	−35.44	−3.83	−3.83
Link-Cluster, 100	−35.92	7.81	7.81

Table 7: Network 1, Medium Volume Per Cent Difference from Individual Trajectory Estimates

METHOD	CO (%)	CO_2 (%)	ENERGY (%)
Aggregated, 1	−43.46	−11.57	−11.57
Aggregated, 2	−40.11	−10.04	−10.04
Binning	−28.13	3.57	3.57
OD-Cluster, 100	−33.98	−1.31	−1.31
Link-Cluster, 100	−35.12	11.72	11.72

Table 8: Network 1, Variable Volume Per Cent Difference from Individual Trajectory Estimates

METHOD	CO (KG)	CO_2 (KG)	ENERGY (MJ)
Aggregated, 1	2.85	777.69	10821
Aggregated, 2	2.93	782.24	10885
Binning	3.42	858.89	11951
OD-Cluster, 100	2.86	843.34	11735
Link-Cluster, 100	2.99	938.02	13052
Raw	3.87	871.70	12129

Table 9: Network 1, Low Volume Results

METHOD	CO (KG)	CO_2 (KG)	ENERGY (MJ)
Aggregated, 1	2.85	777.69	10821
Aggregated, 2	2.93	782.24	10885
Binning	4.49	1131.24	15741
OD-Cluster, 150	3.22	941.38	13099
Link-Cluster, 150	3.39	1054.49	14673
Raw	4.32	969.37	13489

Table 10: Network 1, Medium Volume Results

METHOD	CO (KG)	CO_2 (KG)	ENERGY (MJ)
Aggregated, 1	2.86	779.27	10843
Aggregated, 2	2.93	782.24	10885
Binning	3.42	858.89	11951
OD-Cluster, 150	2.47	715.29	9953
Link-Cluster, 150	2.69	827.80	11518
Raw	3.24	734.91	10226

Table 11: Network 1, Variable Volume Results

method	CO (%)	CO_2 (%)	energy (%)
Aggregated, 1	−26.23	−10.78	−10.78
Aggregated, 2	−24.36	−10.26	−10.26
Binning	−11.51	−1.47	−1.47
OD-Cluster, 100	−26.17	−3.25	−3.25
Link-Cluster, 100	−22.71	7.62	7.62

Table 12: Network 1, Low Volume Per Cent Difference from Individual Trajectory Estimates

method	CO (%)	CO_2 (%)	energy (%)
Aggregated, 1	−33.96	−19.77	−19.77
Aggregated, 2	−32.29	−19.30	−19.30
Binning	3.79	16.70	16.70
OD-Cluster, 150	−25.46	−2.89	−2.89
Link-Cluster, 150	−21.6	8.78	8.78

Table 13: Network 1, Medium Volume Per Cent Difference from Individual Trajectory Estimates

method	CO (%)	CO_2 (%)	energy (%)
Aggregated, 1	−11.61	6.04	6.04
Aggregated, 2	−9.55	6.44	6.44
Binning	5.82	16.87	16.87
OD-Cluster, 150	−23.57	−2.67	−2.67
Link-Cluster, 150	−16.65	12.63	12.63

Table 14: Network 1, Variable Volume Per Cent Difference from Individual Trajectory Estimates

A.2 DATA TABLES FROM THE ECO-DRIVING ANALYSIS

METHOD	CO (KG)	CO_2 (KG)	ENERGY (MJ)
Control	175.26	19040	259332
10%	172.20	18775	255734
25%	172.10	18730	255121
50%	171.71	18756	255487

Table 15: Total Aggregated Emissions by ECO-Driving Market Share

METHOD	CO (%)	CO_2 (%)	ENERGY (%)
10%	1.39	1.40	1.75
25%	1.63	1.63	1.81
50%	1.49	1.50	2.03

Table 16: Per Cent Emissions Reductions by Market Share

B

FIGURES AND EXAMPLES

These figures include illustrations of the UIs and inputs into the various programs.

B.1 INPUT TABLES FOR MOVES

Much of the intermediate input prepared for MOVES is in comma separated files exported by the integration tool. These files can be extremely large, therefore only portions of them are included here to show their general format.

B.1.1 *Sample Table of Links*

The table of links is common to all methods used to input data to MOVES. Every method must provide this table, though it differed slightly between the different methods.

	A	B	C	D	E	F	G	H	I	J	K
1	linkID	countyID	zoneID	roadTypeID	linkLength	linkVolume	linkAvgSpeed	linkDescription	linkAvgGrade		
2	2	36029	360290	5	0.668091885	1	14.48637049	15	0		
3	3	36029	360290	5	0.67450547	3	18.52767141	20	0		
4	4	36029	360290	5	0.644957548	201	23.2134726	25	0		
5	5	36029	360290	5	0.720663869	342	27.3188706	30	0		
6	6	36029	360290	5	0.723302405	264	32.33089948	35	0		
7	7	36029	360290	5	0.712501079	454	37.52093048	40	0		
8	8	36029	360290	5	0.720561345	108	40.85820449	45	0		
9											
10											
11											

Figure 26: Sample Link Table for VB or AS Methods. This table specifically used for a VB input.

	A	B	C	D	E	F	G	H	I	J
1	linkID	countyID	zoneID	roadTypeID	linkLength	linkVolume	linkAvgSpeed	linkDescription	linkAvgGrade	
2	103	36029	360290	5	0.598287067	1	39.84518871	vehicle_gen_link	0	
3	109	36029	360290	5	0.595211281	1	41.98306199	vehicle_gen_link	0	
4	117	36029	360290	5	0.592933956	1	31.31618742	vehicle_gen_link	0	
5	119	36029	360290	5	0.601132947	1	34.80891682	vehicle_gen_link	0	
6	126	36029	360290	5	0.604581556	1	41.80883066	vehicle_gen_link	0	
7	132	36029	360290	5	0.603894941	1	21.14553571	vehicle_gen_link	0	
8	136	36029	360290	5	0.60223588	1	23.62617543	vehicle_gen_link	0	
9	137	36029	360290	5	0.607253451	1	24.87954829	vehicle_gen_link	0	
10	152	36029	360290	5	0.599523596	1	35.29422284	vehicle_gen_link	0	
11	156	36029	360290	5	0.591694321	1	37.98049408	vehicle_gen_link	0	
12	164	36029	360290	5	0.602164422	1	37.99828269	vehicle_gen_link	0	
13	165	36029	360290	5	0.592812789	1	38.08527706	vehicle_gen_link	0	
14	167	36029	360290	5	0.591560726	1	39.40512161	vehicle_gen_link	0	
15	173	36029	360290	5	0.600937215	1	24.05349626	vehicle_gen_link	0	
16	180	36029	360290	5	0.601017993	1	27.72418131	vehicle_gen_link	0	
17	188	36029	360290	5	0.606532661	1	39.64656587	vehicle_gen_link	0	
18	196	36029	360290	5	0.59629868	1	41.24410883	vehicle_gen_link	0	
19	202	36029	360290	5	0.603503477	1	22.21258825	vehicle_gen_link	0	
20	203	36029	360290	5	0.608872122	1	24.11858982	vehicle_gen_link	0	
21	204	36029	360290	5	0.601070809	1	23.77721333	vehicle_gen_link	0	
22	216	36029	360290	5	0.599362039	1	32.59421903	vehicle_gen_link	0	
23	224	36029	360290	5	0.608440269	1	40.49837519	vehicle_gen_link	0	
24	237	36029	360290	5	0.603655713	1	23.38506595	vehicle_gen_link	0	
25	238	36029	360290	5	0.606321394	1	23.98923633	vehicle_gen_link	0	
26	248	36029	360290	5	0.600514682	1	30.00686473	vehicle_gen_link	0	
27	250	36029	360290	5	0.595242349	1	30.60984587	vehicle_gen_link	0	
28	255	36029	360290	5	0.607514427	1	41.19780683	vehicle_gen_link	0	
29	258	36029	360290	5	0.604190092	1	39.49597188	vehicle_gen_link	0	

Figure 27: Sample Link Table Containing Individual Trajectories. This is structurally identical to the previous one, but each link is a unique vehicle and so has a volume of 1. Link numbers are derived from VISSIM and not every number is represented.

	A	B	C	D	E	F	G	H	I
1	linkID	countyID	zoneID	roadTypeID	linkLength	linkVolume	linkAvgSpeed	linkDescription	linkAvgGrade
2	4	36029	360290	5	0.609828205	60	35.40944931	vehicle_clust_link	0
3	65540	36029	360290	5	0.609318607	62	21.09183944	vehicle_clust_link	0
4	131076	36029	360290	5	0.610629981	30	31.40388936	vehicle_clust_link	0
5	196612	36029	360290	5	0.608461751	35	27.72742548	vehicle_clust_link	0
6	262148	36029	360290	5	0.6091413	50	24.09794499	vehicle_clust_link	0
7	5	36029	360290	5	0.611098477	207	19.99962585	vehicle_clust_link	0
8	6	36029	360290	5	0.612028731	71	23.4394443	vehicle_clust_link	0
9	7	36029	360290	5	0.60882062	52	24.90634709	vehicle_clust_link	0
10	8	36029	360290	5	0.68222473	26	24.07856724	vehicle_clust_link	0
11	9	36029	360290	5	0.825193892	40	28.841785	vehicle_clust_link	0
12	65545	36029	360290	5	0.824721015	44	31.58512226	vehicle_clust_link	0
13	131081	36029	360290	5	0.824636902	72	35.76745406	vehicle_clust_link	0
14	196617	36029	360290	5	0.825244795	41	23.39281186	vehicle_clust_link	0
15	262153	36029	360290	5	0.825129834	38	26.05678284	vehicle_clust_link	0
16	10	36029	360290	5	0.824748287	34	30.60927498	vehicle_clust_link	0
17	65546	36029	360290	5	0.827124294	29	23.63216916	vehicle_clust_link	0
18	131082	36029	360290	5	0.826661144	117	35.85525269	vehicle_clust_link	0
19	196618	36029	360290	5	0.826663136	26	32.70322135	vehicle_clust_link	0
20	262154	36029	360290	5	0.826773043	34	26.33972396	vehicle_clust_link	0
21	327690	36029	360290	5	0.826812822	33	28.08049071	vehicle_clust_link	0
22	11	36029	360290	5	0.825424011	44	27.51418783	vehicle_clust_link	0
23	12	36029	360290	5	0.690312112	24	18.97044651	vehicle_clust_link	0
24	13	36029	360290	5	0.691536213	1	39.51643277	vehicle_clust_link	0
25	14	36029	360290	5	0.744467984	22	22.52176484	vehicle_clust_link	0
26	15	36029	360290	5	0.671659685	33	23.02837735	vehicle_clust_link	0
27	16	36029	360290	5	0.740860643	8	27.21534247	vehicle_clust_link	0
28	17	36029	360290	5	0.67601362	25	14.40031846	vehicle_clust_link	0

Figure 28: Sample Link Table For Clustered Input. Although similar to the previous table, the links in this table are the averages of the individual clusters.

B.1.2 *Sample Table of Drive Schedules*

This table is only used for the clustered and individual trajectory input methods. It contains per-second speed information.

linkID	secondID	speed	grade
103	0	34.09097	0
103	1	37.17794	0
103	2	39.10171	0
103	3	39.57147	0
103	4	40.01886	0
103	5	40.46624	0
103	6	40.91363	0
103	7	41.09259	0
103	8	40.6452	0
103	9	40.19781	0
103	10	39.75042	0
103	11	39.30304	0
103	12	38.92276	0
103	13	39.37014	0
103	14	39.8399	0
103	15	40.28729	0
103	16	40.73468	0
103	17	40.91363	0
103	18	40.46624	0
103	19	40.01886	0
103	20	39.57147	0
103	21	39.10171	0
103	22	39.10171	0
103	23	39.57147	0
103	24	39.88464	0
103	25	40.1307	0
103	26	40.57809	0
103	27	41.04785	0

Figure 29: Sample Drive Schedule Table. Link IDs must correspond to an entry in the table of links.

B.1.3 *Sample Table of Link Source Distributions*

This table outlines the type distribution of vehicles on the network. It is required for all input methods and is derived from settings made in VISSIM.

linkID	sourceTypeID	sourceTypeHourFraction
103	21	1
109	21	1
117	21	1
119	21	1
126	21	1
132	21	1
136	21	1
137	21	1
152	21	1
156	21	1
164	21	1
165	21	1
167	32	1
173	21	1
180	21	1
188	21	1
196	21	1
202	21	1
203	21	1
204	21	1
216	21	1
224	21	1
237	21	1
238	21	1
248	21	1
250	21	1
255	21	1
258	21	1
264	21	1

Figure 30: Link Sources Sample Table, Individual Trajectories. In this case the value of *sourceTypeHourFraction* is always 1 as links represent individual vehicles.

	A	B	C
1	linkID	sourceTypeID	sourceTypeHourFraction
2	4	21	0.816666667
3	4	32	0.183333333
4	65540	21	0.935483871
5	65540	32	0.064516129
6	131076	21	0.9
7	131076	32	0.1
8	196612	21	0.942857143
9	196612	32	0.057142857
10	262148	21	0.9
11	262148	32	0.1
12	5	21	0.893719807
13	5	32	0.106280193
14	6	21	0.943661972
15	6	32	0.056338028
16	7	21	0.846153846
17	7	32	0.153846154
18	8	32	0.038461538
19	8	21	0.961538462
20	9	21	0.975
21	9	32	0.025
22	65545	21	0.909090909
23	65545	32	0.090909091
24	131081	21	0.916666667
25	131081	32	0.083333333
26	196617	21	0.902439024
27	196617	32	0.097560976

Figure 31: Link Sources Sample Table, Clustered Trajectories. In this case the value of *sourceTypeHourFraction* represents the fraction of vehicles of a particular source type; each link's entries must always sum to 1.

B.2 OTHER FIGURES

	A	B	C	D	E	F	G	H	I	J
1	linkID	sourceTypeID	sourceTypeHourFraction							
2	103	21	1							
3	109	21	1							
4	117	21	1							
5	119	21	1							
6	126	21	1							
7	132	21	1							
8	136	21	1							
9	137	21	1							
10	152	21	1							
11	156	21	1							
12	164	21	1							
13	165	21	1							
14	167	32	1							
15	173	21	1							
16	180	21	1							
17	188	21	1							
18	196	21	1							
19	202	21	1							
20	203	21	1							
21	204	21	1							
22	216	21	1							
23	224	21	1							
24	237	21	1							
25	238	21	1							
26	248	21	1							
27	250	21	1							
28	255	21	1							
29	258	21	1							
30	264	21	1							

Figure 32: Raw Output of VISSIM. This output is configurable, but if certain columns are not included then the integrated tool will not run.

BIBLIOGRAPHY

[1] H. Abou-Senna and E. Radwan. Vissim/moves integration to investigate the effect of major key parameters on co2 emissions. *Transportation Research Part D: Transport and Environment*, 21:39–46, 2013. ISSN 13619209. doi: 10.1016/j.trd.2013.02.003.

[2] H. Abou-Senna, E. Radwan, K. Westerlund, and C.D. Cooper. Using a traffic simulation model (vissim) with an emissions model (moves) to predict emissions from vehicles on a limited-access highway. *Journal of the Air and Waste Management Association*, 63(7):819–831, 2013. ISSN 10962247. doi: 10.1080/10962247.2013.795918.

[3] Song Bai, Douglas Eisinger, and Deb Niemeier. Moves vs. emfac: A comparison of greenhouse gas emissions using los angeles county. In *Transportation Research Board 88th Annual Meeting, Paper*, pages 09–0692, 2009.

[4] M.a b Barth, S.a Mandava, K.b Boriboonsomsin, and H.a Xia. Dynamic eco-driving for arterial corridors. pages 182–188, Vienna, 2011. ISBN 9781457709906. doi: 10.1109/FISTS.2011.5973594. cited By 10; Conference of 2011 IEEE Forum on Integrated and Sustainable Transportation Systems, FISTS 2011 ; Conference Date: 29 June 2011 Through 1 July 2011; Conference Code:86294.

[5] Matthew Barth, Feng An, Joseph Norbeck, and Marc Ross. Modal emissions modeling: A physical approach. *Transportation Research Record: Journal of the Transportation Research Board*, 1520:81–88, 1996. doi: 10.3141/1520-10.

[6] Matthew Barth, Feng An, Theodore Younglove, C Levine, G Scora, Marc Ross, and Thomas Wenzel. *Development of a comprehensive modal emissions model*. National Cooperative Highway Research Program, Transportation Research Board of the National Academies, 2000.

[7] Matthew Barth, Carrie Malcolm, Theodore Younglove, and Nicole Hill. Recent validation efforts for a comprehensive modal emissions model. *Transportation Research Record: Journal of the Transportation Research Board*, (1750): 13–23, 2001.

[8] Paul E Benson. Caline4-a dispersion model for predicting air pollutant concentrations near roadways. Technical report, 1984.

[9] H. Bogo, D.R. Gomez, S.L. Reich, R.M. Negri, and E. San Roman. Traffic pollution in a downtown site of buenos aires city. *Atmospheric Environment*, 35(10):1717 – 1727, 2001. ISSN 1352-2310. doi: http://dx.doi.org/10.1016/S1352-2310(00)00555-0.

[10] K. Boriboonsomsin and M. Barth. Impacts of freeway high-occupancy vehicle lane configuration on vehicle emissions. *Transportation Research Part*

D: Transport and Environment, 13(2):112–125, 2008. ISSN 13619209. doi: 10.1016/j.trd.2008.01.001.

[11] CALTRANS. Caline4 software, February 2014. URL http://www.dot.ca.gov/hq/env/air/software/caline4/calinesw.htm.

[12] G.D.B.a Cameron and G.I.D.b Duncan. Paramics - parallel microscopic simulation of road traffic. *Journal of Supercomputing*, 10(1):25–53, 1996. ISSN 09208542.

[13] Robert Chamberlin, Ben Swanson, Eric Talbot, Jeff Dumont, and Stephen Pesci. Analysis of moves and cmem for evaluating the emissions impact of an intersection control change. In *Transportation Research Board 90th Annual Meeting*, number 11-0673, 2011.

[14] G.a De Nunzio, C.C.b De Wit, P.c Moulin, and D.d Di Domenico. Eco-driving in urban traffic networks using traffic signal information. pages 892–898, Florence, 2013. Institute of Electrical and Electronics Engineers Inc. ISBN 9781467357173. doi: 10.1109/CDC.2013.6759995.

[15] I De Vlieger, D De Keukeleere, and J.G Kretzschmar. Environmental effects of driving behaviour and congestion related to passenger cars. *Atmospheric Environment*, 34(27):4649 – 4655, 2000. ISSN 1352-2310. doi: http://dx.doi.org/10.1016/S1352-2310(00)00217-X.

[16] DLR Institute of Transportation Systems. Sumo - simulation of urban mobility, 2015. URL http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/.

[17] Environment Canada. Tool information: Mobile 6.2c, 2 2012. URL http://data.tc.gc.ca/archive/eng/programs/environment-urban-menu-eng-1799.htm#4.1.11.

[18] Martin Fellendorf. Vissim: A microscopic simulation tool to evaluate actuated signal control including bus priority. In *64th Institute of Transportation Engineers Annual Meeting*, pages 1–9, 1994.

[19] N.a Ferreira, J.T.b Klosowski, C.E.b Scheidegger, and C.T.a Silva. Vector field k-means: Clustering trajectories by fitting multiple vector fields. *Computer Graphics Forum*, 32(3 PART2):201–210, 2013. ISSN 01677055. doi: 10.1111/cgf.12107.

[20] H. Christopher Frey, Alper Unal, Jianjun Chen, and Song Li. Modeling mobile source emissions based upon in-use and second-by-second data: Development of conceptual approaches for epa's new moves model. Number 69577. North Carolina State University, 2003.

[21] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28 (1):pp. 100–108, 1979. ISSN 00359254.

[22] John A Hartigan. *Clustering Algorithms*. Wiley & Sons, Inc., 1975.

[23] Q.a b He, K.L.c Head, and J.c Ding. Multi-modal traffic signal control with priority, signal actuation and coordination. *Transportation Research Part C: Emerging Technologies*, 46:65–82, 2014. ISSN 0968090X. doi: 10.1016/j.trc.2014.05.001.

[24] A.G.a Hobeika, H.a Jung, and S.b Bae. Contribution of aggressive drivers to automobile tailpipe emissions under acceleration and braking conditions. *Journal of Transportation Engineering*, 141(2), 2014. ISSN 0733947X. doi: 10.1061/(ASCE)TE.1943-5436.0000736.

[25] S. Ilgin Guler, M. Menendez, and L. Meier. Using connected vehicle technology to improve the efficiency of intersections. *Transportation Research Part C: Emerging Technologies*, 46:121–131, 2014. ISSN 0968090X. doi: 10.1016/j.trc.2014.05.008.

[26] INRO Software. Inro emme, 2015. URL https://www.inrosoftware.com/en/.

[27] L.a Int Panis, S.a Broekx, and R.b Liu. Modelling instantaneous traffic emission and the influence of traffic speed limits. *Science of the Total Environment*, 371(1-3):270–285, 2006. ISSN 00489697. doi: 10.1016/j.scitotenv.2006.08.017. cited By (since 1996)65.

[28] Intelligent Transportation Systems, Joint Program Office. Connected vehicle research in the united states, May 2015. URL https://web.archive.org/web/20141229174506/http://www.its.dot.gov/connected_vehicle/connected_vehicle_research.htm.

[29] Anil K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651 – 666, 2010. ISSN 0167-8655. doi: http://dx.doi.org/10.1016/j.patrec.2009.09.011.

[30] David J Ketchen and Christopher L Shook. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal*, 17(6):441–458, 1996.

[31] Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. Recent development and applications of sumo–simulation of urban mobility. *International Journal On Advances in Systems and Measurements*, 5(3&4), 2012.

[32] Stefan Krauß. *Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics*. PhD thesis, Universitat zu Koln., 1998.

[33] Paul Kunselman, HT McAdams, CJ Domke, and ME Williams. Automobile exhaust emission modal analysis model; calspan corporation; buffalo, ny; prepared for the environmental protection agency (document 460/3-74-005), office of mobile source air pollution control. *Ann Arbor, MI*, 1974.

[34] A.F. Lentzakis, R. Su, and C. Wen. Time-dependent partitioning of urban traffic network into homogeneous regions. pages 535–540. Institute of Electrical and Electronics Engineers Inc., 2014. ISBN 9781479951994. doi: 10.1109/ICARCV.2014.7064361.

[35] J.-Q. Li, K. Zhou, S.E. Shladover, and A. Skabardonis. Estimating queue length under connected vehicle technology. *Transportation Research Record*, (2356):17–22, 2013. ISSN 03611981. doi: 10.3141/2356-03.

[36] Meng Li, Kanok Boriboonsomsin, Guoyuan Wu, Wei-Bin Zhang, and Matthew Barth. Traffic energy and emission reductions at signalized intersections: a study of the benefits of advanced driver information. *International Journal of Intelligent Transportation Systems Research*, 7(1):49–58, 2009.

[37] Michal Maciejewski. A comparison of microscopic traffic flow simulation systems for an urban area. *Transport Problems*, 5(4):27–38, 2010.

[38] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif., 1967. University of California Press.

[39] H Maldonado, M Carlock, E Yotter, and M Ahuja. Air resources board's research projects to improve the mobile source emissions inventory. *A&WMA The Emission Inventory: Perception and Reality*, pages 51–70, 1993.

[40] S.a Mandava, K.b Boriboonsomsin, and M.a Barth. Arterial velocity planning based on traffic signal information under light traffic conditions. pages 160–165, St. Louis, MO, 2009. ISBN 9781424455218. doi: 10.1109/ITSC.2009.5309519.

[41] Kanti Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. 1980.

[42] Greg Marsden, Margaret Bell, and Shirley Reynolds. Towards a real-time microscopic emissions model. *Transportation Research Part D: Transport and Environment*, 6(1):37 – 60, 2001. ISSN 1361-9209. doi: http://dx.doi.org/10.1016/S1361-9209(00)00012-2.

[43] MATSim Development Group. Matsim, 2015. URL http://www.matsim.org/.

[44] B. Metz, O.R. Davidson, P.R. Bosch, R. Dave, and L.A. Meyer. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007, 2007. URL http://www.ipcc.ch/pdf/assessment-report/ar4/wg3/ar4_wg3_full_report.pdf.

[45] M Muresan, S. M. K. Hossain, and L Fu. A model-based evaluation of an eco-driving system using connected vehicle technologies. 2015.

[46] M Muresan, S. M. K. Hossain, L Fu, and Chaozhe Jiang. A trajectory-clustering based method for integrating microscopic traffic simulation and emission estimation models. in-press.

[47] Wassim G. Najm, Jonathan Koopmann, John D. Smith, and John Brewer. Frequency of target crashes for intellidrive safety systems, October 2010. URL http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2010/811381.pdf.

[48] R.B. Noland and M.A. Quddus. Flow improvements and vehicle emissions: Effects of trip generation and emission control technology. *Transportation Research Part D: Transport and Environment*, 11(1):1–14, 2006. ISSN 13619209. doi: 10.1016/j.trd.2005.06.003.

[49] A.a Papson, S.b Hartley, and K.-L.c Kuo. Analysis of emissions at congested and uncongested intersections with motor vehicle emission simulation 2010. *Transportation Research Record*, (2270):124–131, 2012. ISSN 03611981. doi: 10.3141/2270-15.

[50] Portrait Software International. Quadstone paramics, 2015. URL http://www.paramics-online.com/.

[51] PTV Group. Ptv group: Transportation planing, traffic engineering and traffic simulation, 2015. URL http://vision-traffic.ptvgroup.com/en-us/home/.

[52] Hesham Rakha, Kyoungho Ahn, and Antonio Trani. Comparison of mobile5a, mobile6, vt-micro, and cmem models for estimating hot-stabilized light-duty gasoline vehicle emissions. *Canadian Journal of Civil Engineering*, 30(6):1010–1021, 2003. doi: 10.1139/l03-017.

[53] N. Saunier and T. Sayed. Automated analysis of road safety with video data. *Transportation Research Record*, (2019):57–64, 2007. ISSN 03611981. doi: 10.3141/2019-08. cited By 20.

[54] Robin Smit, Muriel Poelman, and Jeroen Schrijver. Improved road traffic emission inventories by adding mean speed distributions. *Atmospheric Environment*, 42(5):916 – 926, 2008. ISSN 1352-2310. doi: http://dx.doi.org/10.1016/j.atmosenv.2007.10.026.

[55] Darrell Bruce Sonntag and H Oliver Gao. The moves from mobile: Preliminary comparisons of epa?s current and future mobile emissions models. In *Transportation Research Board 86th Annual Meeting*, number 07-3090, 2007.

[56] S. Trommer and A. Holtl. Perceived usefulness of eco-driving assistance systems in europe. *IET Intelligent Transport Systems*, 6(2):145–152, 2012. ISSN 1751956X. doi: 10.1049/iet-its.2011.0154.

[57] UC Riverside. Comprehensive modal emission model (cmem), 2008. URL http://www.cert.ucr.edu/cmem/index.html.

[58] United States Department of Transportation. Applications for the environment: Real-time information synthesis (AERIS), 2015. URL http://www.its.dot.gov/aeris/aeris_progress.htm.

[59] United States Environmental Protection Agency. Draft design and implementation plan for epa's multi-scale motor vehicle and equipment emission system (moves), October 2002. URL http://www.epa.gov/otaq/models/ngm/p02006.pdf.

[60] United States Environmental Protection Agency. Moves(motor vehicle emission simulator). [Online], October 2013. URL http://www.epa.gov/otaq/models/moves/.

[61] United States Environmental Protection Agency. Description and history of the mobile highway vehicle emission factor model. [Online], May 2013. URL http://www.epa.gov/otaq/models/mob-hist.htm.

[62] Babu Veeregowda, Teresa Lin, and Joshua Herman. An efficient approach to epa?s moves hot-spot emissions analysis using comprehensive traffic modeling. In *Transportation Research Board 94th Annual Meeting*, number 15-3281, 2015.

[63] R. Wiedemann. Simulation des verkehrsflusses schriftenreihe des instituts für verkehrswesen. 1974.

[64] H.a Xia, K.a Boriboonsomsin, F.b Schweizer, A.b Winckler, K.c Zhou, W.-B.c Zhang, and M.a Barth. Field operational testing of eco-approach technology at a fixed-time signalized intersection. pages 188–193, Anchorage, AK, 2012. ISBN 9781467330640. doi: 10.1109/ITSC.2012.6338888.

[65] H.a Xia, G.b Wu, K.b Boriboonsomsin, and M.J.b Barth. Development and evaluation of an enhanced eco-approach traffic signal application for connected vehicles. pages 296–301, The Hague, 2013. ISBN 9781479929146. doi: 10.1109/ITSC.2013.6728248.

[66] Yuanchang Xie, Mashrur A Chowdhury, Parth Bhavsar, and Yan Zhou. An integrated tool for modeling the impact of alternative fueled vehicles on traffic emissions: a case study of greenville, south carolina. In *Transportation Research Board 90th Annual Meeting*, number 11-3880, 2011.