Australasian Transport Research Forum 2017 Proceedings
27-29 November 2017, Auckland, New Zealand
Publication Website: http://www.atrf.info

# Reinforcement Learning based Driving Speed Control for Two Vehicle Scenario

Hasitha Dilshani Gamage[1] , Dr. Jinwoo (Brian) Lee[2]
Civil Engineering and Build Environment School, Queensland University of Technology, Brisbane, Australia
[1] hasithadilshani.dabaregamage@hdr.qut.edu.au, [2] brian.lee@qut.edu.au

Email for correspondence: hasithadilshani.dabaregamage@hdr.qut.edu.au

## Abstract

Dynamic speed guidance at signalised intersections to minimise the fuel consumption has been explored for many years. The vehicle manoeuvring in the urban setting will be mainly influenced by the traffic signal control and neighbouring vehicles, but its combined effect on the driving speed control has not been fully explored in the literature. In this paper, a Reinforcement Learning based speed control algorithm is proposed to provide the real-time fuel-optimal velocity. The control objective is to minimise the fuel consumption of a control vehicle at an isolated signalised intersection. Although the literature presents some early studies of eco-speed control, their application is limited to simple driving environments where there is no neighbouring traffic other than one control vehicle. The proposed algorithm is able to optimise the driving speed of a control vehicle taking into account its leading vehicle as well as the traffic signal timing condition. The analysis results show that the algorithm is able to effectively reduce the fuel consumption and complete stopping at a red signal. The saving of fuel ranged from 6% and 58%. The variation in the fuel saving is analysed by the control vehicle's approaching speed, traffic signal timing and leading vehicle status.

# 1. Introduction

The transport sector is responsible for 60% of global fuel consumption (Jollands et al., 2010) Fossil fuel is one of the major greenhouse gas contributors, and 88% of transport-related $CO_2$ emission is caused by the surface transport sector (Stanley et al., 2011). Therefore, it is essential to address the negative externalities caused by automobiles. Eco-driving was introduced to reduce fuel consumption and emission exhaust by enhancing driving habits. Early eco-driving practice focused on providing driving guidelines and training programs for drivers (Andrieu and Pierre, 2012, Strömberg and Karlsson, 2013, Beusen et al., 2009). However, drivers tend to fall back to their old driving behaviour quickly, as predicting such optimal manoeuvring under realistic driving environments is simply implausible (Kanok et al., 2010, Thew, 2007, Beusen et al., 2009).

In the urban setting, the traffic signal is the main source of periodic disruption of traffic flow. Responding to the changes in traffic flow imposed by traffic signals and neighbouring traffic, vehicles deviate from their optimal trajectory. Acceleration, deceleration and stop-and-go as a result of the traffic signal operation are the main contributors to excessive fuel consumption and emission exhaust (Barth et al., 2011, Mandava et al., 2009). The stop-and-go behaviour may incur additional emission exhaust of up to 14%, compared to vehicles travelling at a constant speed (Xia et al., 2012). Minimising unnecessary idling, acceleration and deceleration can effectively reduce fuel consumption and emission. However, current eco-driving practice requires considerable improvements to allow drivers to better respond to the traffic signal operation and neighbouring traffic to extend its application to wider and more realistic driving environments.

The emerging connected vehicle technology offers technological support to significantly advance the current practice of eco-driving by allowing to exchange real-time information among cars and between cars and infrastructures (Olia et al., 2016, Katsaros et al., 2011, Guler et al., 2014, Feng et al., 2015). More recent studies have proposed adaptive speed guidance algorithms, which provide driving speed recommendations on a real-time basis to cope with changing traffic control conditions, as an improvement to the earlier eco-driving practice (Barth et al., 2011, Rakha and Kamalanathsharma, 2011). Existing speed control algorithms provide individual control vehicles with target speeds, acceleration or deceleration limits and speed alerts to reduce vehicle fuel consumption levels. Various assumptions have been made regarding the driving environments; however, most of the speed control algorithms are limited to simple driving environments where there is no neighbouring traffic other than one control vehicle (Barth et al., 2011, Kamalanathsharma and Rakha, 2013, Hooker, 1988, Saerens, 2012, Zhang and Yao, 2015, Ozatay et al., 2012)

A number of optimisation methods have been applied to achieve various control objectives to provide vehicles with speed recommendations. Mandava et al. (2009) proposed an algorithm using a constrained optimisation technique. Asadi and Vahidi (2011) developed a cruise control system using model predictive control. Other optimisation techniques including; a Lagrange multiplier method (Wu et al., 2011), Pontryagin's minimum principle (Schwarzkopf and Leipnik, 1977, Wan et al., 2016, Ozatay et al., 2012), Dynamic Programming (Mahler and Vahidi, 2014, Mensing et al., 2011, Ozatay et al., 2013, Hellström et al., 2010, Kamalanathsharma and Rakha, 2013) and genetic algorithm (Chen et al., 2014), have been used for driving speed control in the literature. The aforementioned optimisation methods have drawbacks, including mathematical complexities, the need for an explicit model of the environment, or low computational efficiency (Mensing et al., 2011, Wan et al., 2016, Ozatay et al., 2012, Baskar et al., 2011).

Due to the complex and multi-factorial interactions among vehicles and with traffic signal control, machine learning techniques such as reinforcement learning (RL) have great potential for driving speed optimisation. RL is widely used, due to its generality and close relation to human thinking and learning behaviour through experience. RL has been employed to solve optimisation problems across a broad variety of engineering areas, but it has not been fully explored in the context of traffic engineering applications (Abdulhai and Kattan, 2003). A model-free RL technique called Q-learning offers significant advantages over traditional optimal control techniques including computational efficiency, easy implementation, and a continuous learning process (Sutton and Barto, 1998, Watkins and Dayan, 1992, Abdulhai and Kattan, 2003).

This paper presents a Q-learning based speed control algorithm. The control objective is to minimise the fuel consumption of a control vehicle at an isolated signalised intersection. Although the literature presents some early studies of eco-speed control, their application is limited to simple driving environments where there is no neighbouring traffic other than one control vehicle. The proposed algorithm is able to optimise the driving speed of a control vehicle taking into account its leading vehicle as well as the traffic signal timing conditions.

This proposed eco-speed control algorithm is significant in that: 1) it is able to account for the leading vehicle and the traffic signal operation in the driving speed optimisation; 2) it takes a model-free, self-learning approach for improved practicality and adaptability; and 3) it utilises an explicit fuel consumption model rather than an oversimplified surrogate measure. This study builds on the Q-learning agent, developed and calibrated through a comprehensive sensitivity test (Gamage and Lee, 2016). The enhanced algorithm is trained and tested using Aimsun microsimulation model. The algorithm is validated through a series of simulation tests under varying conditions. The performance is quantified in terms of fuel consumption, travel time, and complete stopping, which is then compared to a base scenario without the speed control for performance evaluation.

# 2. Methodology

## 2.1 Reinforcement Learning

RL is a closed-loop autonomous sequential decision-making algorithm that is inspired by human learning behaviour and decision-making process (Sutton and Barto, 1998). The decision-making of RL is refined through direct trial-and-error interactions with its environments without direct supervision. During the learning period, a learning agent attempts to perceive the current state $s_t$ of the environment and choose an action $a_t$ at each episode. The action results in changes in the state of the environment, which makes the agent to encounter a new state. A scalar reward value is given to assess the desirability of executing an action, while being on the given state space (Sutton and Barto, 1998). Figure 1 shows the agent-environment interaction.

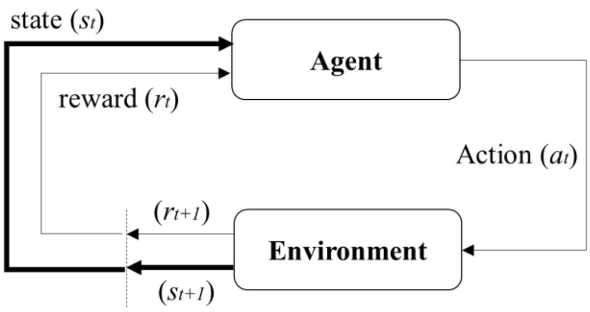



**Figure 1: Agent-environment interaction**

The state represents some characteristics of the environment, which are relevant to the control problem. The controller, learner or the decision maker, is called an agent. The objective of the action is to change the environment by moving from one state to another. To evaluate the impact of the performed action on the agent's final desirability, a reward is given. Normally, a positive reward is given on positive actions, and a negative reward is given on negative actions. Q-learning is a model-free Reinforcement learning technique. The agent's mapping from state to action is called as policy, $\pi$. The policy is improved iteratively through the agent's experience. The $Q(s,a)$ represents the expected sum of the discounted rewards of taking an action $a$ in the state $s$ by a certain policy ($\pi$). Many possible trials are executed during the training phase to confirm that the agent has learnt from enough experiences and convergence of each state-action pair. The control problems` ultimate goal is to find the optimal policy $\pi^*$ which determines the best control action when the agent is in a particular state,

$$\pi*(s) \in arg\ max\ a \in AQ(s,a) \qquad (1)$$

The value associated with a state-action pair is updated in a look-up table with its current value $Q(s_t, a_t)$, instance reward that receives for the executed action $r$ and with the expected return starting from that state$Q\ (s_{t+1}, a_{t+1})$. The Q-learning process may be expressed as below.

Algorithm 1 Pseudo code of Q-learning

---

Input: *set of states* $S$*, set of actions* $A$*, reward* $R$

Input: *Discount rate (*$\gamma$*), Learning rate* $(\alpha)$ *and action selection policy parameter*$(\varepsilon)$

Initialise $Q(s,a)$ *arbitrarily for every state* $s$ *and every action* $a$

**For** *each episode* **do**

*Initialise* $s$

*Repeat {for each step of the episode}*

*Choose* $a$ *from* $s$ *based on the policy derived*

*(e.g.* $\varepsilon - greedy$)

*Take action a, observe r, s.'*

$Q(s,a) \leftarrow Q(s,a) + \ \alpha[r + \gamma max_\alpha Q(s',a') - Q(s,a)]$

$\pi(s) \leftarrow arg\, max_{a \in A} Q(s,a)$

$s \leftarrow s'$;

**Until** $s$ *is terminal*

**End for**

---

Where $\gamma$ is called discounted rate $(0 < \ \gamma \leq 1)$ and $\alpha_t$ is called the learning rate $(0 < \ \gamma \leq 1)$

## 2.2. Algorithm Development

The control objective of the algorithm is to minimise the fuel consumption of the following controlled vehicle from its arrival on the intersection approach until crossing the stop line. It was assumed that the leading vehicle's trajectory was not impacted by the other traffic in the road. Hence, the leading vehicle travels at free-flow speed. The manoeuvring of the leading vehicle was determined by Aimsun without further modifications of the relevant parameters, and free-flow speed varies as a result of responding to the traffic signal timings.

Figure 2, illustrates the layout of the experimental intersection. The intersection approach was a single-lane, one-way street. The traffic signal operated a pre-timed timing plan with the 60-seconds of cycle time with two phases. For each episode, two vehicles – one leading non-control vehicle and one following controlled vehicle – were randomly generated 750 metres upstream of the intersection to maintain a random gap. The algorithm computes the optimal driving speed for the control vehicle at two control points at 300-metre (A) and 150-metre (B) upstream of the intersection. The target speeds were provided at two control points only to avoid frequent speed changes and potential fluctuations. It was also chosen to reduce the size of the state-space to minimise the computational load. To allow wide variations in the driving environment, the arrival speed of the control vehicle at the 300-metre point was chosen between 30 km/h and 54 km/h with an increment of 4 km/h.

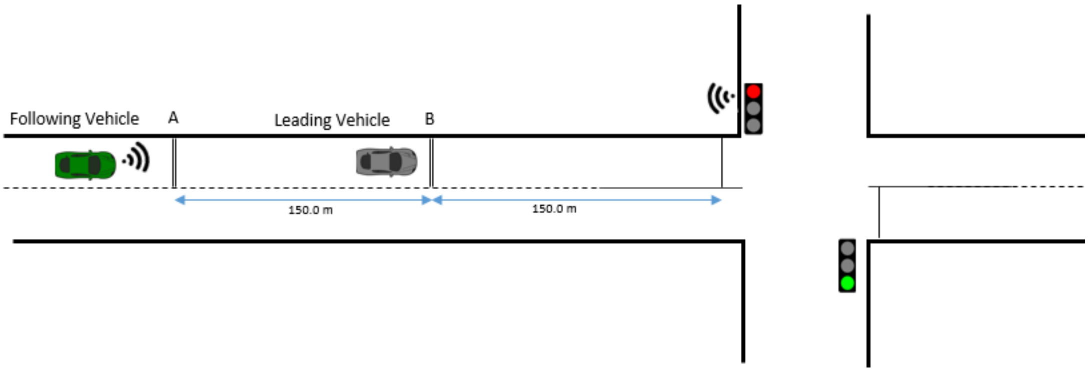



**Figure 2: Experimental Intersection Layout**

### *2.2.1 Problem Formulation*

To formulate the environment, each state is defined by a 4-quatiple, $s = (v_i, d_i, t_i, h_i)$; where, $v_i$ is the current speed, $d_i$ is the current position of the following vehicle, $t_i$ is the traffic signal timing status, and $h_i$ is the headway between the leader and the follower. The proposed speed control was measured states at discrete moments in the distance (measured using the sample distance $\Delta d$). Hence, at every $\Delta d$ the algorithm measured the state information. The speed was discretised into 2 kmh intervals, and traffic signal consists of 60 seconds of cycle timing and it was discretised into 1-second intervals. The time headway $(h_i)$ between the control vehicle and the leading vehicle was discretised into 1 second intervals between 1s and 30s.

The possible actions in this paper include changing the current driving speed to a target speed; restricted to +8 km/h and -10 km/h from the current speed between the maximum and minimum driving speeds, defined at 60 km/h and 20km/h, respectively. The agent used normal acceleration and deceleration speeds to reach the target speed and then continued cruising at that speed until the next decision point. The following controlled vehicle also adjusted the driving speed to maintain the safe distance from its leading car and also to respond to the traffic signal operation. The maximum acceleration and maximum deceleration were used as +2.2 m/s$^2$ and 2.7m/s$^2$ respectively.

As the main control objective is to minimise the fuel consumption along the vehicle trajectory, the reward function is inversely proportional to the cumulative fuel consumption experienced by the vehicle between two successive decision points. A penalty of 2 was given if the headway between leading and following vehicle is less than 2 seconds during the learning period. It should be noted that by reducing this extra component does not impact on the behaviour of optimal policy, as it is necessary to keep the minimum safe headway.

### *2.2.2 Vehicle Fuel Consumption*

Vehicle fuel consumption can be determined by a variety of factors including torque, air drag coefficient, temperature, driving speed, and so on. The microscopic fuel consumption model embedded in Aimsun was used in this paper. This model accounts for the time spent by each vehicle in the network during each simulation step in each of four operating modes; acceleration, deceleration, cruising and idling. The fuel consumption during idling, deceleration and acceleration is derived from Ferreira (1982), and cruising fuel consumption is derived from Akçelik (1983). During a given simulation time step the fuel consumed by a given vehicle $n$ is given by,

$$Fuel_n = f^n_{curise} + f^n_{acceleration} + f^n_{deceleration} + f^n_{idle} \qquad (2)$$

$$f^n_{curise} = k_1(1 + (\frac{v}{2v_m})^3) + k_2 v$$

$$f^n_{acceleration} = c_1 + c_2 a v$$

$c_1$+$c_2$ – The two constants in the equation for the fuel consumption rate for the accelerating vehicles (ml/s), $k_1$ and $k_2$ need to be determined empirically for each vehicle type. $v_m$ – the speed at which the fuel consumption rate, in ml/s, is at a minimum for a vehicle cruising at constant speed. It is assumed that the fuel consumption rate is constant for the idling and decelerating. This assumption can be easily relaxed (Osorio and Nanduri, 2015).

## 3. Simulation Test and Discussion

In this section, the simulation results of the proposed algorithm are discussed. The driving speed of the control vehicle was adjusted to achieve a recommended driving speed by RL. The initial control point is located 300 metres (point A) upstream of the intersection. Once a control vehicle enters into the control area, it follows the speed recommended by the algorithm until it reaches the next control point or discharges the intersection. The following controlled vehicle will attempt to achieve the recommended driving speed otherwise interrupted by the leading vehicle or the traffic signal control.

## 3.1 Reinforcement Learning Agent Training

There are two basic phases in developing the RL based eco-speed control algorithm. The first phase is training process, and the next is implementation phase. In this study, a lookup table was used to store Q-values during the training phase. The Q-values ($Q(s,a)$) for all state-action pairs were stored in a 2-dimentional matrix. Each row is represented by a single state and each column is represented by a single action. Each cell value represents a measure of how good or bad the executed action under the particular state condition. As the learning continued, the Q-values stored in the look up table progressed such that the most desired action will be the maximised the Q-value.

Initially, all the $Q(s,a)$ values were set to zero and the different rewards for various actions in states were observed and updated accordingly. The Q-value per state-action pair was then updated using the one-step equation as in below,

$$(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma max_{\alpha} Q(s',a') - Q(s,a)] \quad (3)$$

In other words, Q-values for a given state (s) under a particular action (a) is represented by immediate reward plus the maximum discounted future reward from the best future action taken in the following state. The training process was continued until all the states met the termination criteria. The termination criteria were defined by the following condition. If the absolute value of the Q-value function in iteration N and iteration N-1($_{\text{error}}$) was oscillating with a small variation (0.001) for few consecutive iterations, then the learning was terminated.

$$\Delta Q_t^N = \ | Q_t^N(s',a) - Q_t^{N-1}(s',a) \ | \qquad \forall\ t, s', a \quad (4)$$

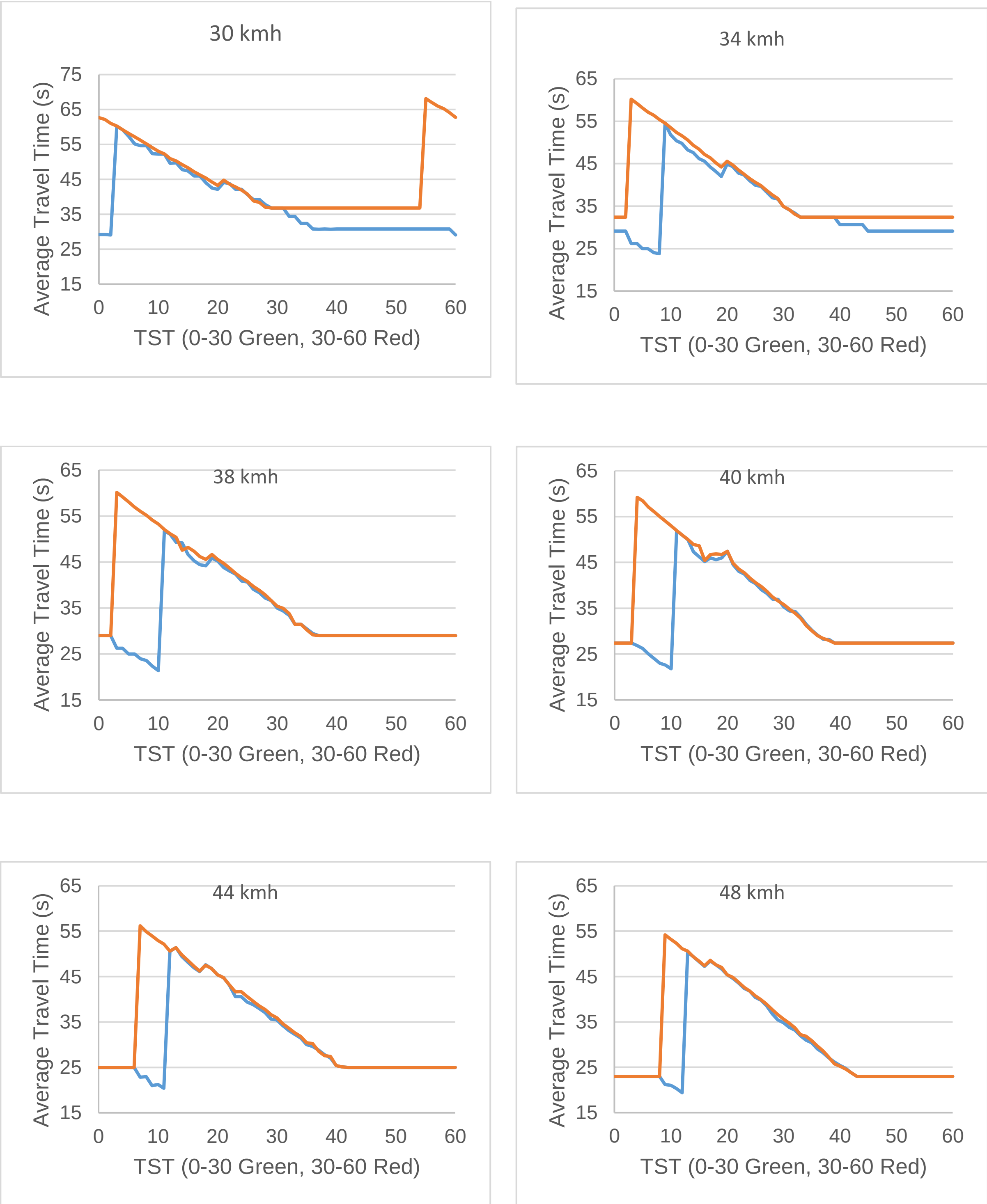
Once the $\Delta Q_t^N$ for all the state-action pairs were satisfied with the defined threshold value, the training process was terminated.  The agent was trained under different starting conditions which available due to the different traffic signal timing states and arrival speeds. This guarantees training with the all different time dependent variations of driving conditions. After all the training was completed the algorithm performance was evaluated during the implementation phase. The converged look-up table ($Q^*$) for each states$s \in S$, action $a \in A$ pair was used during the implementation phase.

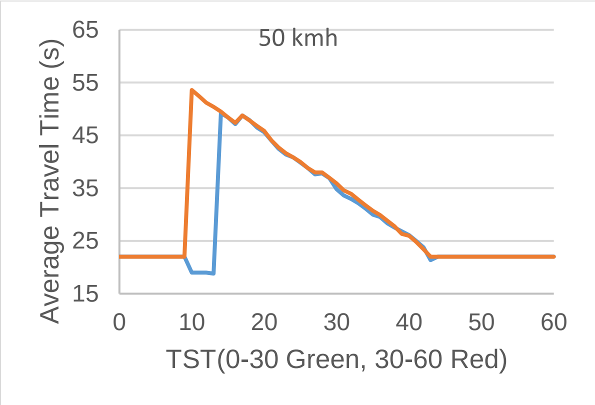
The fuel consumption of a total 2,400 control vehicles was compared with and without the speed control under the exactly identical test conditions (i.e., traffic signal status, control vehicle arrival speed, and leading vehicle status). The average fuel consumption, the average travel time, and the percentage of complete stopping during the red signal were collected for the test. The fuel consumption was estimated using the velocity and acceleration as presented in equation (2). The same fuel consumption parameters as in (Kamal et al., 2010)  were used in this study. The fuel consumption parameters used for this study were as follows. $c_1$ = 0.42,$c_2$ = 0.26, $F_i$= 0.333, $F_d$= 0.537, $F_1$ = 4.7, $F_2$=6.5 and $v_m$ = 60 km/h. The parameters related to the Q-learning algorithm were selected as $\omega$ = 0.8 and ε = 0.8.

## 3.2 Results and Analysis

In the following, overall algorithm performances were presented considering vehicle fuel efficiency, travel time and complete stopping at the intersection (this could be either due to the red signal or leading vehicle). The comparisons were made between the following controlled vehicle (vehicles with speed control algorithm) and the following uncontrolled vehicle (vehicles without speed control algorithm).

### *3.2.1 Impact on Travel Time*





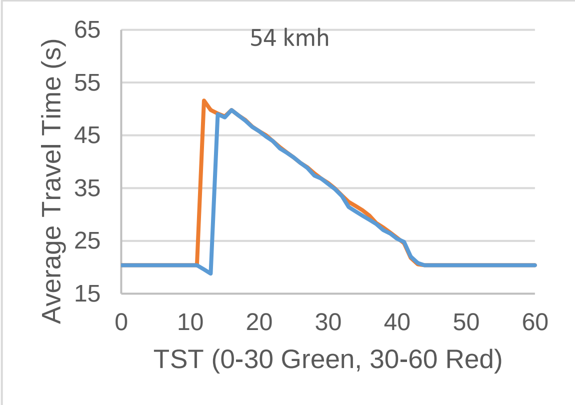
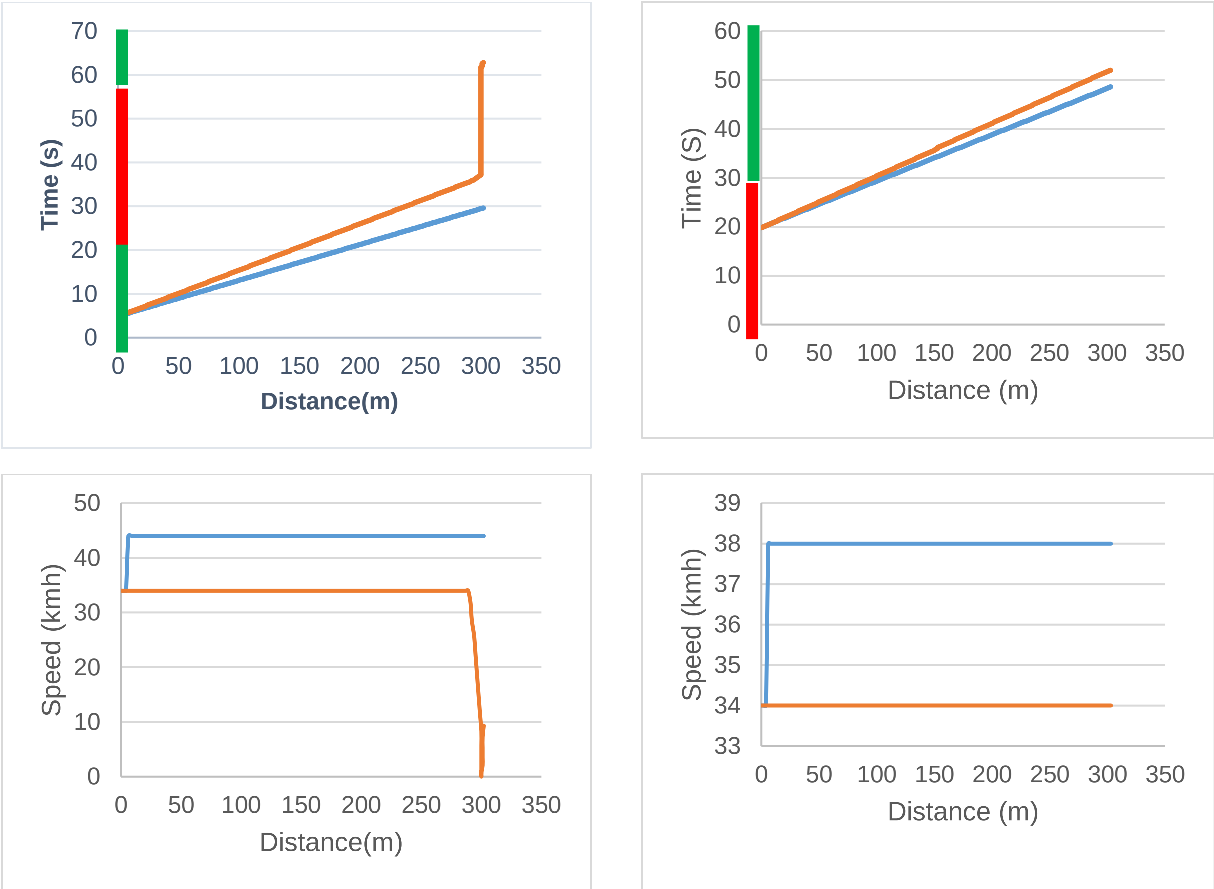




**Figure 2: Impact of Average Travel Time**
**Manoeuvring of a control vehicle with the speed control algorithm (Blue colour) and without the control algorithm (Orange colour)**

The algorithm's impact on the travel time of following controlled vehicles was analysed in this section. The impact was analysed considering two variables; traffic signal timing and arrival speeds of the following vehicles. The highest travel timing saving achieved was around 60% when the following controlled vehicles arrived at the beginning of the green signal. For demonstration purposes Figure 3 (a) and (b), show the manoeuvring of a following controlled vehicle with the arrival speed at 34km/h. In figure 3 (a), the following controlled vehicle avoided the complete stopping at the red signal by accelerating the driving speed. The following vehicle without the speed control had to wait for the next green time as a result of maintaining its driving speed.

The following controlled vehicles with 30km/h and 34 km/h arrival speed achieved significant travel time improvement at the end of the red phase as a result of discharging the intersection during the earliest possible green time by accelerating. Figure 3 (b) shows the vehicle manoeuvring when the red signal is presented. However, the travel improvement at the end of the green and at the beginning of the red was marginal. During above two signal timing periods, following controlled vehicle also decelerated or idled for a short period to avoid stopping at the intersection. The defined constraints (maximum possible deceleration and number of control points) made it unavoidable. Hence, there was no significant travel time improvement compared to the following vehicles without the speed control algorithm.




(a) (b)

**Figure 3: Manoeuvring of a control vehicle with the speed control algorithm (Blue colour) and without the control algorithm (Orange colour) (a) A vehicle approach the intersection during at the beginning of the green signal (b) A vehicle approach the intersection at the end of the red signal**

### *3.2.2 Impact on Fuel Consumption*

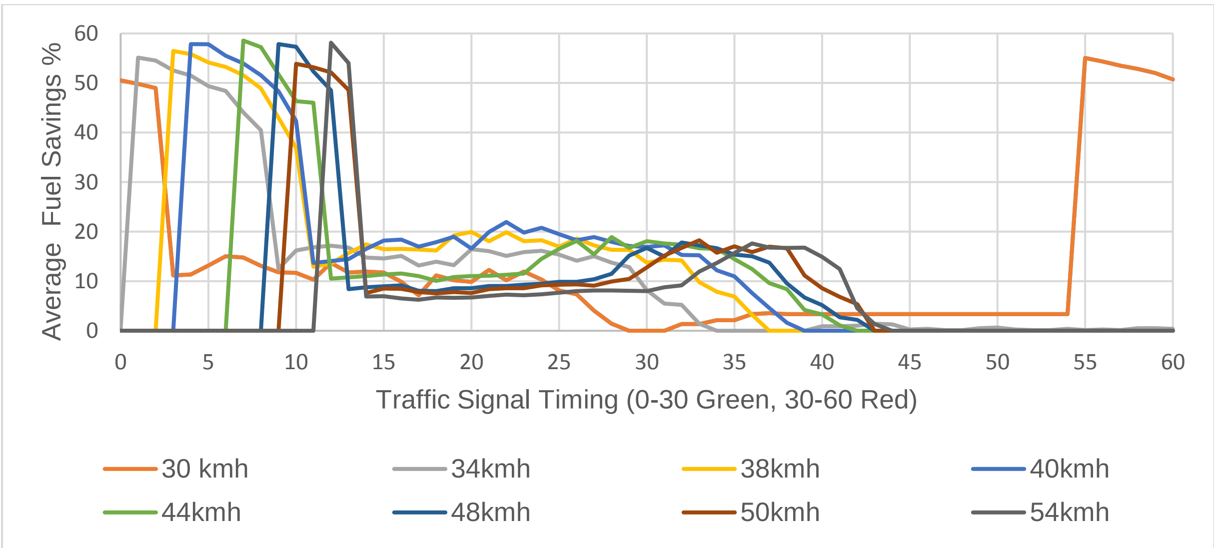



**Figure 4: Average Fuel Savings %**

The average fuel saving % is calculated using following equation,

AFS% = $\sum_{i=1}^{n}(FC_{controlled} - FC_{uncontrolled})/(\sum_{i=1}^{n} FC_{uncontrolled})$

Where i is the number of vehicles, and FC is the fuel consumption

The algorithm's impact on the fuel savings of control vehicles was analysed in this section. The impact was analysed considering two variables; traffic signal timings and arrival speeds of the following vehicle. The highest fuel saving achieved was around 58% when the following controlled vehicles arrived at the beginning of the green signal. Here, the following controlled vehicle used traffic signal information and leading vehicle information and discharged the intersection at green signal by accelerating the driving speed. The following vehicles without speed control react aggressively to reduce its speed and idle until next green signal. The average fuel saving was around 10 - 20% which was achieved at the end of green and at the beginning of red. It was found that the vehicles with speed control algorithm reduced its speed early to avoid an unnecessary stopping during the red phase.

As depicted in figure 4, fuel savings of around 52 % was achieved at the beginning of the green phase by accelerating to discharge the intersection within the current green phase. The reason for this behaviour is that, when the delay time required due to the red signal timing become smaller, the average vehicle speed becomes higher. When the vehicles without speed control algorithm maintain their current speed of 30 km/h, the vehicles with speed control algorithm accelerated from their current speeds and maintained higher average speeds. As a result of this, the following uncontrolled vehicle achieved significant fuel savings compared to the following uncontrolled vehicle. In contrast, there were no significant fuel savings at the end of the red phase, as the majority of the following vehicles maintained current speeds without further adjustments.

### *3.2.3 Impact on Vehicle Stopping at Red Signal*


**Figure 5: Relationship between vehicle arrival speeds and stopping at red**

Figure 5, shows the complete stopping at the red signal as a percentage of a total number of simulated vehicles under each arrival speed. As mentioned above, both the red phase and the leading vehicle dynamics were responsible for the complete stopping at the intersection. The compete stopping of following controlled vehicles showed an inclined trend as the speed was increased. The reason for the behaviour is due to the restrictions imposed by the number of speed control points and minimum deceleration level. In this study, the minimum deceleration limit was assumed as 2.7m/$s^{2,}$ and the number of speed control points was assumed as two. When the vehicle arrival speed was high, the lowest speed suggested by the algorithm was greater than the speed that was required to discharge the intersection. However, even at high arrival speeds, stopping % of following vehicles with speed control algorithm was significantly low compared to the following vehicles without speed control.

### *3.2.4 Leading vehicle Impact*

It is important to analyse further the impact of leading vehicle's behaviour on the following controlled vehicle as well as on the fuel economy. The time headway between leading and the following controlled vehicles and fuel saving % was plotted against 200 vehicles at two control points (point A and B). The headway variation of leading and the following vehicle was caused by the following vehicle's arrival speed, approaching time and leading vehicle's free-flow speed. The 200 vehicles consisted with 8 different arrival speeds at the end of the green phase.

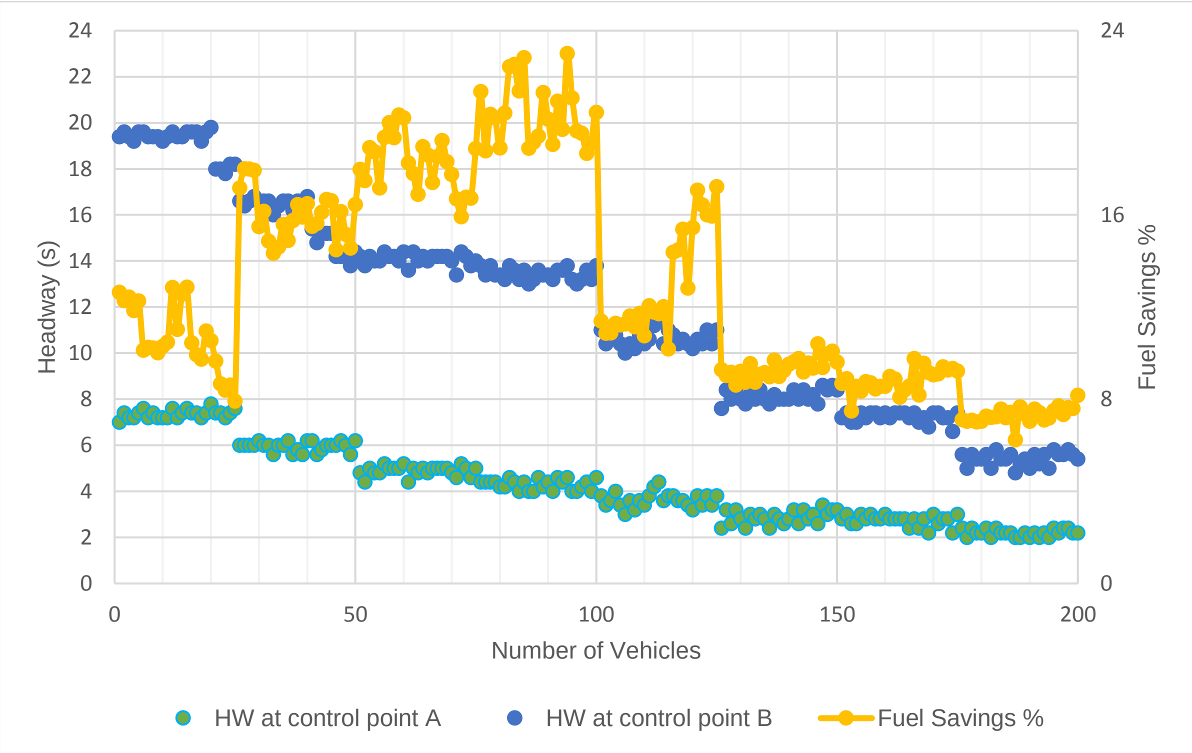



**Figure 6: Relationship between headway and fuel savings %**

As in Figure 6, the headway between leading and following vehicles were significantly large for first the 100 vehicles. The fuel savings were low when the headway was relatively large, as the behaviour of the following vehicle was not impacted by the leading vehicle's trajectory. When the headway between leading and following vehicle was around 13s-14s, significant fuel savings was recorded. In contrast, the fuel savings percentage was proportional to the headway variation of the next 100 vehicles. This behaviour is due to the fact that when the headway became smaller, the influence of the following controlled vehicle became more significant. Also, the low headway between leading and following vehicle restricted the possible target speeds that can be achieved by the following controlled vehicle.

# 4. Conclusion and Future Work

In this paper, a driving speed control algorithm was proposed to minimise the fuel consumption of vehicle at an isolated signalised intersection. The algorithm utilised the traffic signal timing and the leading vehicle information to compute the optimum driving speeds. The performance was analysed in terms of the average fuel savings percentage, the average travel time-saving % and the % of vehicles stopping at the red signal. The most significant fuel saving was approximately 58%. The algorithm also effectively reduced the travel time of the control vehicles up to 60% compared to the scenario without the speed control. The number of vehicles stopping at the red signal was reduced approximately 30% with the speed control algorithm. Also, the temporal and spatial constraints imposed by the leading vehicles dynamics significantly impacted on the following vehicle's fuel economy. The algorithm presented in the paper only considered a single leading vehicle. However, further analysis is required considering varying neighbouring traffic conditions to analyse the performance of the algorithm at more realistic driving environments.

# References

ABDULHAI, B. & KATTAN, L. 2003. Reinforcement learning: Introduction to theory and potential for transport applications. *Canadian Journal of Civil Engineering,* 30**,** 981-991.

AKÇELIK, R. Progress in fuel consumption modelling for urban traffic management. 1983. Australian Road Research Board.

ANDRIEU, C. & PIERRE, G. S. 2012. Comparing effects of eco-driving training and simple advice on driving behaviour. *Procedia-Social and Behavioral Sciences,* 54**,** 211-220.

ASADI, B. & VAHIDI, A. 2011. Predictive Cruise Control Utilizing Upcoming Traffic Signal Information for Improving Fuel Economy and Reducing Trip Time. *IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY,* vol.19.

BARTH, M., BORIBOONSOMSIN, K., MANDAVA, S. & HAITAO, X. 2011. Dynamic ECO-Driving for Arterial Corridors. *2011 IEEE Forum on Integrated and Sustainable Transportation Systems.* Vienna, Austria.

BASKAR, L. D., DE SCHUTTER, B., HELLENDOORN, J. & PAPP, Z. 2011. Traffic control and intelligent vehicle highway systems: a survey. *IET Intelligent Transport Systems,* 5**,** 38-52.

BEUSEN, B., BROEKX, S., DENYS, T., BECKX, C., DEGRAEUWE, B., GIJSBERS, M., SCHEEPERS, K., GOVAERTS, L., TORFS, R. & PANIS, L. I. 2009. Using on-board logging devices to study the longer-term impact of an eco-driving course. *Transportation research part D: transport and environment,* 14**,** 514-520.

CHEN, Z., ZHANG, Y., LV, J. & ZOU, Y. 2014. Model for Optimization of Ecodriving at Signalized Intersections. *Transportation Research Record: Journal of the Transportation Research Board***,** 54-62.

FENG, Y., HEAD, K. L., KHOSHMAGHAM, S. & ZAMANIPOUR, M. 2015. A real-time adaptive signal control in a connected vehicle environment. *Transportation Research Part C: Emerging Technologies,* 55**,** 460-473.

FERREIRA, L. 1982. Car Fuel Consumption in Urban Traffic. The Results of a Survey in Leeds using Instrumented Vehicles.

GAMAGE, H. D. & LEE, J. B. 2016. Machine learning approach for self-learning eco-speed control. *38th Australasian Transport Research Forum (ATRF).* Melbourne, Victoria, Australia.

GULER, S. I., MENENDEZ, M. & MEIER, L. 2014. Using connected vehicle technology to improve the efficiency of intersections. *Transportation Research Part C: Emerging Technologies,* 46**,** 121-131.

HELLSTRÖM, E., ÅSLUND, J. & NIELSEN, L. 2010. Design of an efficient algorithm for fuel-optimal look-ahead control. *Control Engineering Practice,* 18**,** 1318-1327.

HOOKER, J. N. 1988. Optimal driving for single-vehicle fuel economy. *Transportation Research Part A: General,* 22**,** 183-201.

JOLLANDS, N., WAIDE, P., ELLIS, M., ONODA, T., LAUSTSEN, J., TANAKA, K., DE T'SERCLAES, P., BARNSLEY, I., BRADLEY, R. & MEIER, A. 2010. The 25 IEA energy efficiency policy recommendations to the G8 Gleneagles Plan of Action. *Energy Policy,* 38**,** 6409-6418.

KAMAL, M. A. S., MUKAI, M., MURATA, J. & KAWABE, T. 2010. Ecological driver assistance system using model-based anticipation of vehicle–road–traffic information. *IET Intelligent Transport Systems,* 4**,** 244.

KAMALANATHSHARMA, R. K. & RAKHA, H. A. Multi-stage Dymaic Programming Algorithm for Eco-Speed Control at Traffic signalized Intersections. Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on, 2013. IEEE, 2094-2099.

KANOK, B., ALEXANDER, V. & MATHTHEW, B. 2010. Eco-Driving - Pilot Evaluation of Driving Behavior Changes Among U.S Drivers.

KATSAROS, K., KERNCHEN, R., DIANATI, M. & RIECK, D. Performance study of a Green Light Optimization Speed Advisory(GLOSA) application using an integrated copperative ITS simulaton platform. 7th International Wireless Communications and Mobile Computing Conference, 2011 Turkey.

MAHLER, G. & VAHIDI, A. 2014. An optimal velocity-planning scheme for vehicle energy efficiency through probabilistic prediction of traffic-signal timing. *Intelligent Transportation Systems, IEEE Transactions on,* 15**,** 2516-2523.

MANDAVA, S., BORIBOONSOMSIN, K. & BARTH, M. Arterial velocity planning based on traffic signal infomation under light traffic conditions. 12th International IEEE Conference on Intelligent Transportation Systems, 2009 USA.

MENSING, F., TRIGUI, R. & BIDEAUX, E. Vehicle trajectory optimization for application in ECO-driving. Vehicle Power and Propulsion Conference (VPPC), 2011 IEEE, 2011. IEEE, 1-6.

OLIA, A., ABDELGAWAD, H., ABDULHAI, B. & RAZAVI, S. N. 2016. Assessing the potential impacts of connected vehicles: mobility, environmental, and safety perspectives. *Journal of Intelligent Transportation Systems,* 20**,** 229-243.

OSORIO, C. & NANDURI, K. 2015. Energy-efficient urban traffic management: a microscopic simulation-based approach. *Transportation Science,* 49**,** 637-651.

OZATAY, E., ONORI, S., WOLLAEGER, J., OZGUNER, U., RIZZONI, G., FILEV, D., MICHELINI, J. & DI CAIRANO, S. 2013. Cloud-Based Velocity Profile Optimization for Everyday Driving: A Dynamic-Programming-Based Solution.

OZATAY, E., OZGUNER, U., ONORI, S. & RIZZONI, G. Analytical solution to the minimum fuel consumption optimization problem with the existence of a traffic light. ASME 2012 5th Annual Dynamic Systems and Control Conference joint with the JSME 2012 11th Motion and Vibration Conference, 2012. American Society of Mechanical Engineers, 837-846.

RAKHA, H. & KAMALANATHSHARMA, R. K. Eco-driving at signalized intersections using V2I communication. Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on, 2011. IEEE, 341-346.

SAERENS, B. 2012. *Optimal control based eco-driving.* Ph. D. thesis, Katholieke Universiteit Leuven.
SCHWARZKOPF, A. B. & LEIPNIK, R. B. 1977. Control of highway vehicles for minimum fuel consumption over varying terrain. *Transportation Research,* 11**,** 279-286.
STANLEY, J., LOADER, C. & BUSVIC. 2011. Road transport and climate change: Stepping off the greenhouse gas. *Transportation Research Part A: Policy and Practice,* 45**,** 1020-1030.
STRÖMBERG, H. K. & KARLSSON, I. 2013. Comparative effects of eco-driving initiatives aimed at urban bus drivers–Results from a field trial. *Transportation Research Part D: Transport and Environment,* 22**,** 28-33.
SUTTON, R. S. & BARTO, A. G. 1998. *Introduction to reinforcement learning*, MIT Press.
THEW, R. 2007. United evidence and research strategy:driving standards agency,CIECA
WAN, N., VAHIDI, A. & LUCKOW, A. 2016. Optimal speed advisory for connected vehicles in arterial roads and the impact on mixed traffic. *Transportation Research Part C: Emerging Technologies*.
WATKINS, C. J. C. H. & DAYAN, P. 1992. Q-learning. *Machine learning,* 8**,** 279-292.
WU, C., ZHAO, G. & OU, B. 2011. A fuel economy optimization system with applications in vehicles with human drivers and autonomous vehicles. *Transportation Research Part D: Transport and Environment,* 16**,** 515-524.
XIA, H., BORIBOONSOMSIN, K., SCHWEIZER, F., WINCKLER, A., ZHOU, K., ZHANG, W. & BARTH, M. Field operational testing of eco-approach technology at a fixed-time signalized intersection. Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on, 2012. IEEE, 188-193.
ZHANG, R. & YAO, E. 2015. Eco-driving at signalised intersections for electric vehicles. *IET Intelligent Transport Systems,* 9**,** 488-497.