

Development of the Foundation for a Supplementary Ramp Metering Algorithm to Address Negative Impacts of Ramp Metering

Nima Amini¹, Lauren Gardner and Travis Waller

¹ School of Environmental and Civil Engineering

The University of New South Wales, Sydney, Australia

Email for correspondence: n.amini@unsw.edu.au

Abstract

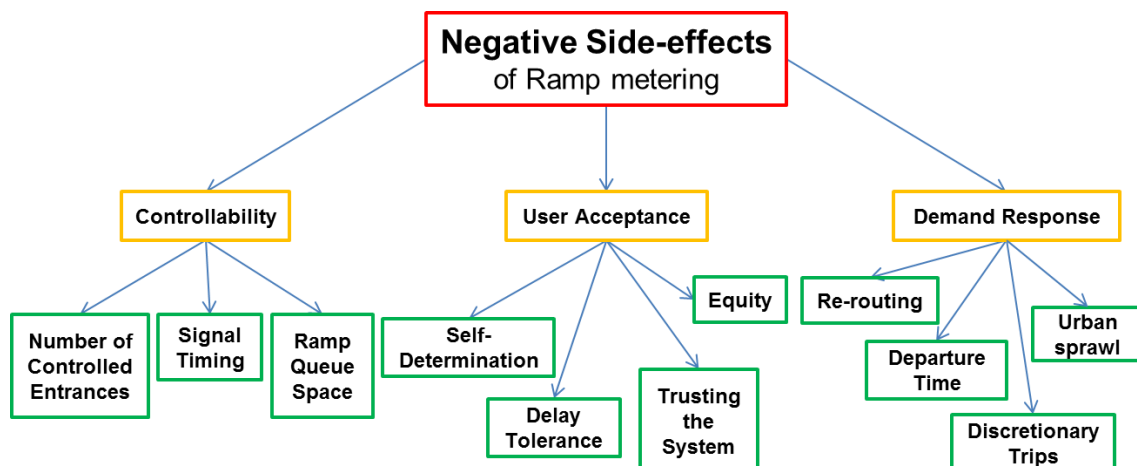
Ramp metering has been shown to be a useful tool for increasing the overall efficiency of the freeway system. However, they can have negative side-effects such as user rejection, encouragement of congestion inducing trip choices, and even urban sprawl. To reduce the burden of these side-effects, this paper develops the building blocks of a supplementary algorithm which is able to address some of these issues. An analytical framework is developed to better justify the choices made in developing the algorithm and is used to compare the proposed solution to the system optimal and the typical equity strategy adopted in the literature. The results show that some of the side-effects examined are addressed whilst still operating close the system optimal efficiency.

1. Introduction

A ramp meter is a set of traffic lights placed along freeway entrance on-ramps. The main difference between standard traffic lights and ramp meters is that short cycle lengths (typically between 4 to 18 seconds) are used to regulate the release of vehicle/s (one vehicle or a small cluster of vehicles) into the freeway (Burley and Gaffney 2013). Ramp meters fall within the broader field of Intelligent Transport Systems (ITS) (Piotrowicz and Robinson 1995).

Whilst the benefits of ramp metering has been shown through both simulation (Papageorgiou and Kotsialos 2002, Kotsialos and Papageorgiou 2004, Amini, Aydos et al. 2015, Amini, Grzybowska et al. 2015) and field data (Haj-Salem and Papageorgiou 1995, Levinson and Zhang 2006, Faulkner, Dekker et al. 2013), many of the side-effects associated with ramp metering continue to be a concern. Figure 1 classifies and summarizes these side effects.

Figure 1: Negative Side-effects of Ramp Metering



Majority of the negative impacts of ramp meters are due to the uneven distribution of ramp metering delays between the onramps. Examples of negative side effects include:

- An increase in long discretionary trips during the peak period (ZhangI and Levinson 2002)
- Large ramp metering delays can cause motorists to re-route through local streets (i.e. rat-run) to use an alternative onramp with lower delay (Piotrowicz and Robinson 1995).
- Whilst trips at onramps that are metered appear to experience peak spreading, the unmetered entry points experience peak narrowing (Papageorgiou and Papamichail 2007).
- The inequity resulting from the current methods of distributing ramp metering delays among onramps have also been sighted as a major concern for motorists and the road authorities (Arnold Jr 1998, Jacobson, Stribiak et al. 2006, Mizuta, Roberts et al. 2014).

(Amini, Gardner et al. 2016) suggested that if the ramp metering delays were distributed according to the level of congestion caused by each onramp many of the negative side-effects may be alleviated. This concept was based on horizontal equity or cost-responsibility (as referred to in congestion tolling). A methodology was also developed for calculating the “*equitable*” ramp metering delays (in this paper referred to as the *ideal RM delays*) based on the same concept. However, the previous paper only focused on evaluating existing ramp metering systems and did not explore the methodology of implementing the proposed concept. The aim of this research is to fill this gap by *developing a supplementary ramp metering algorithm to address the negative side effects of ramp metering*.

In the process of developing the supplementary algorithm an analytical framework is developed to isolate the effects of the proposed ramp metering strategy. The framework is a direct response to the concerns raised regarding how ramp metering algorithms have been evaluated in the literature (Zhang and Levinson 2004). Typically the only validation step taken is simulating the final product (i.e. ramp metering strategy). Whilst simulations allow for the evaluation of complex and interacting variables, they do not indicate why an algorithm performs the way it does.

The proposed analytical framework determines how the proposed RM strategy performs in comparison to existing strategies. To isolate the implications of the PS strategy, an idealized scenario is assumed where ramp meters are able to perfectly maximize throughput on the freeway. These assumptions result in deterministic mathematical equations and closed form solutions for different strategies given any demand and network configuration. The framework is also used to compare the Proposed Solution (PS) against two other strategies, namely System Optimal (SO) and Typical Equity (TE). The assumptions are then removed sequentially until a generalized optimization program and its associated solution method is developed.

Three *ramp metering strategies* are examined. These strategies differ only in how they distribute the *RM delays* when maximizing the bottleneck/s utilization. Their objectives are defined as:

1. *System Optimal (SO)* - minimize Total System Travel Time (TSTT).
2. *Typical Equity (TE)* – the *RM delays* are the same across all onramps.
3. *Proposed Solution (PS)* – distribute the *RM delays* between the onramps (and in time) according to the level contribution to the bottlenecks. Detailed reasons for selecting this objective is summarized in (Amini, Gardner et al. 2016).

2. Assumptions in the Analytical Framework

In this section the primary assumptions of the analytical framework and the associated *terminologies* are defined.

A metering *update-period* (also referred to as *control period* or *traffic cycle*) is a short period of time during which the same ramp metering rate is applied. During each update-period traffic conditions are measured by detectors and used to calculate the metering rate to be

applied in the following update-period. Update-periods used in ramp metering range between 15 to 60 seconds (Papageorgiou and Papamichail 2007). In the analytical framework an update-period of one hour is defined to simplify the formulations presented and to exadurate the effects.

The analytical framework is assumed to operated under the *steady-state condition*, such that the travel time between onramps and any part of the freeway (i.e. bottleneck/s) is instantaneous as long as the freeway is congestion free.

The *bottleneck patrons* of onramp i , refers to the number of vehicles that depart the onramp and travel through bottleneck b (see Figure 3). In addition, the *proportion of bottleneck patrons* ($\lambda_{i,b}$) refers to the *bottleneck patrons* divided by the total number of vehicles departing onramp i . The *non-bottleneck patrons* exit the freeway via offramps upstream of the bottleneck.

Dead time occurs because each motorist on an onramp has to wait their turn to enter the freeway regardless of their destination. Dead time refers specifically to the ramp metering delay experienced by motorists that are *non-bottleneck patrons*. As metering *non-bottleneck patrons* does not improve the bottleneck condition, *dead time* only serves to increase the inefficiency of the ramp metering system.

Unlimited controlability is assumed. Thus, no limit is imposed on the queue storage space or maximum metering rates. The minimum metering rate is equal to zero (i.e. ramp closure).

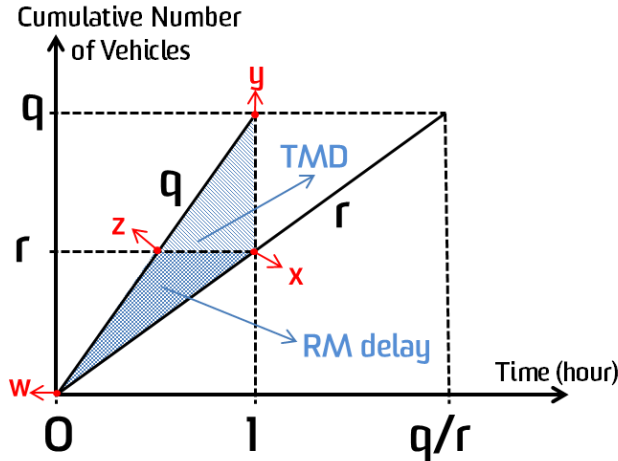
Ramp Metering (RM) delay is defined to be the average ramp metering delay experienced by vehicles exiting an onramp during one update-period. A vertical queue at the stopline of the onramps is assumed. The Input-Output diagram (shown in Figure 2) is used to calculate the RM delay of each onramp (Lawson, Lovell et al. 1997). The area between the input and output lines is the experienced delay (Lovell and Windover 1999). As shown in Figure 2, the RM delay (δ) is the area of the triangle $w-x-z$ divided by the metering rate (r). The left-over delay (triangle $x-y-z$) is the delay incurred by vehicles that are unable to depart the onramp during the update-period. The Total Metering Delay (TMD), is defined to be the sum of the area $w-x-y$ across all onramps. The TMD is used to compare the efficiency of each ramp metering strategy. These variables can be derived by calculating the area of the respective triangles.

$$\delta = \frac{\text{area}(w-x-z)}{r} = \frac{\text{base} \times \text{height}}{2r} = \frac{\left(1 - \frac{r}{q}\right) \times r}{2r} = 0.5 \left(1 - \frac{r}{q}\right) \quad 1$$

$$\Delta = \text{area}(w-x-y) = \frac{\text{base} \times \text{height}}{2} = \frac{(q-r) \times 1}{2} = 0.5(q-r) \quad 2$$

$$TMD = \sum_i \Delta_i \quad \forall i \in I \quad 3$$

Figure 2: Input-Output Diagram of an Onramp with a Ramp Meter



It is assumed that the controllers are able to realise *perfect metering*, meaning that maximum bottleneck throughput is achieved at all times. In other words, if the traffic demand at a bottleneck is larger than its capacity, the controllers meter the demand so that the total amount of traffic at the bottleneck is exactly at capacity (i.e. just before capacity drop). Thus, maximizing the bottleneck throughput is a pre-condition of the ramp metering strategies examined. The perfect metering assumption can be summarized mathematically:

$$C = \sum_i \lambda_i r_i \quad i \in I$$

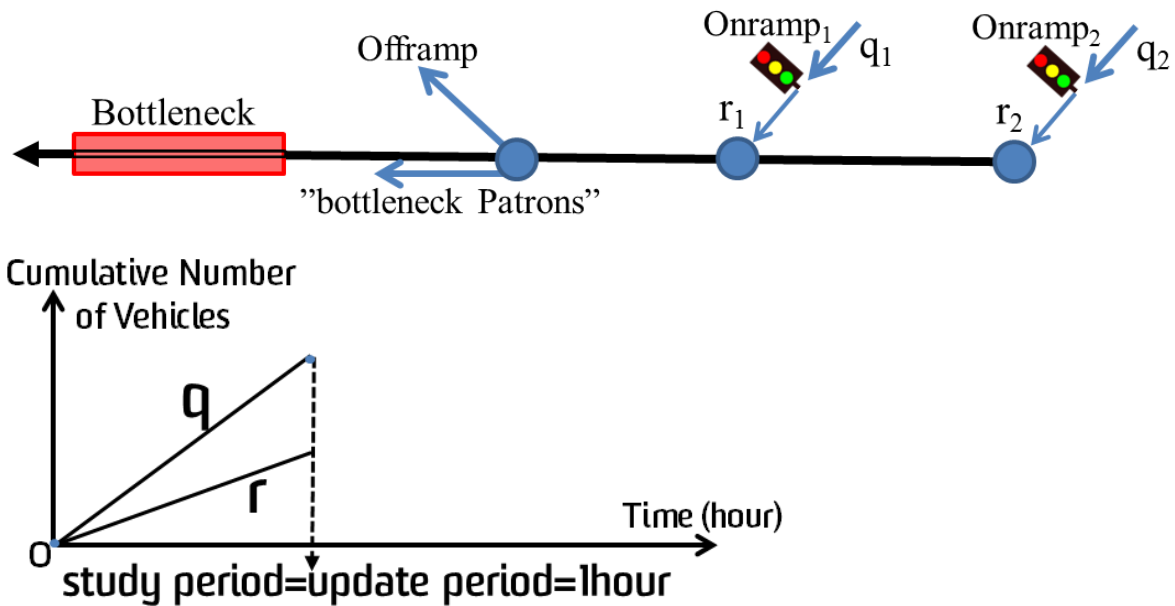
4

3. Development of the PS Strategy

3.1 Simple Network

The network shown in Figure 3 has the advantage of allowing for exact solutions to be solved numerically.

Figure 3: Simple Network (one bottleneck)



The PS strategy has three interacting variables all of which are dependant on the decision variable (i.e. metering rates). Firstly the bottleneck utilization is to be maximized using

Equation (4). Second, the RM delays (Equation 1). Third, the *onramp contribution* to the bottleneck and is comprised the following two ratios:

1. *P1* is the direct contribution from an onramp to the bottleneck (i.e. punishment for contributing to congestion). It is defined as the *bottleneck patrons* from onramp *i* divided by the total number of motorists in the bottleneck:

$$P1 = \frac{\lambda_i r_i}{\lambda_1 r_1 + \lambda_2 r_2}$$

2. *P2* is the proportion of *bottleneck patrons* (i.e. compensation for not contributing to congestion). It is defined as the *bottleneck patrons* from onramp *i* divided by the total vehicles departing the onramp:

$$P2 = \frac{\lambda_i r_i}{r_i} = \lambda_i$$

The final contribution of onramp *i* is defined in Equation 5. Thus, the contribution increases as more vehicles from the onramp enter the bottleneck. The contribution is reduced as its ratio of bottleneck patrons is reduced (i.e. to reduce the burden non-bottleneck patrons):

$$contribution_i = P1 \times P2 = \frac{\lambda_i r_i}{\lambda_1 r_1 + \lambda_2 r_2} \times \lambda_i \quad 5$$

The objective of the PS strategy is to distribute the RM delay (required to maintain the bottleneck at capacity) among the onramps according to their contribution. Thus, the according to the PS strategy, the RM delay to be applied at onramp *i* is:

$$\delta_i = Contribution_i \times \frac{[\delta_1 + \delta_2]}{Contribution_1 + Contribution_2} \quad 6$$

As an alternative, the total RM delay (i.e.) was considered instead of the average RM delay (i.e.). The advantage of using total RM delay is that the TMD remains the same after redistributing it using Equation 6. However, the total RM delay can be reduced by releasing less vehicles (i.e. more restrictive metering), which would increase the average RM delay. As the objective of the PS strategy is to take into account the delay experienced by individual drivers, the average RM delay is utilized.

Substituting Equation 5 and 1 into Equation 6:

$$\begin{aligned} \delta_i &= 0.5 \times \left(1 - \frac{r_i}{q_i}\right) = \frac{\lambda_i^2 r_i}{\lambda_1 r_1 + \lambda_2 r_2} \times \frac{0.5 \times \left(1 - \frac{r_1}{q_1}\right) + 0.5 \times \left(1 - \frac{r_2}{q_2}\right)}{\frac{\lambda_1^2 r_1}{\lambda_1 r_1 + \lambda_2 r_2} + \frac{\lambda_2^2 r_2}{\lambda_1 r_1 + \lambda_2 r_2}} \\ &= \lambda_i^2 r_i \times \frac{0.5 \times \left[2 - \frac{r_1}{q_1} - \frac{r_2}{q_2}\right]}{\lambda_1^2 r_1 + \lambda_2^2 r_2} \quad 7 \end{aligned}$$

Substituting Equation (4) in Equation (7) and solving for r_1 :

$$0.5 \times \left(1 - \frac{r_1}{q_1}\right) = \lambda_1^2 r_1 \times \frac{0.5 \times \left[2 - \frac{r_1}{q_1} - \frac{r_2}{q_2}\right]}{\lambda_1^2 r_1 + \lambda_2^2 r_2}$$

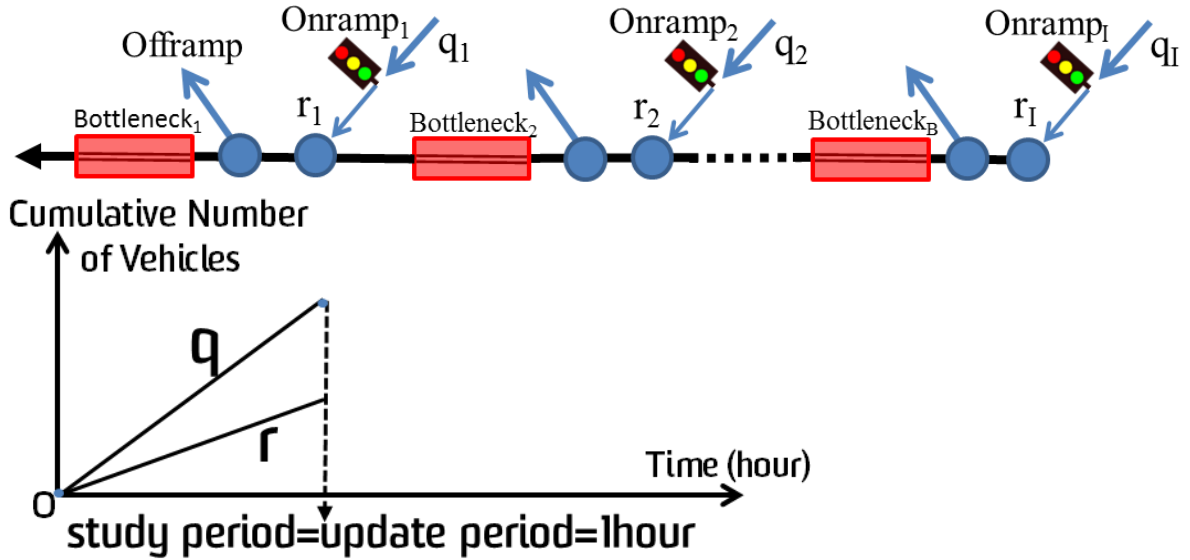
$$\begin{aligned} \Rightarrow 0 &= r_2 \times \left(\lambda_2^2 - \frac{\lambda_2^2 r_1}{q_1} + \frac{\lambda_1^2 r_1}{q_2}\right) - \lambda_1^2 r_1 = \frac{C - r_1 \lambda_1}{\lambda_2} \times \left(\lambda_2^2 - \frac{\lambda_2^2 r_1}{q_1} + \frac{\lambda_1^2 r_1}{q_2}\right) - \lambda_1^2 r_1 \\ &= r_1^2 \left[\frac{\lambda_1 \lambda_2}{q_1} - \frac{\lambda_1^3}{\lambda_2 q_2}\right] + r_1 \left[\frac{\lambda_1^2 C}{\lambda_2 q_2} - \frac{\lambda_2 C}{q_1} - \lambda_1 \lambda_2 - \lambda_1^2\right] + [\lambda_2 C] \end{aligned} \quad 8$$

r_1 in Equation 8 can be solved using the standard quadratic formula. r_2 can be solved using:

$$r_2 = \frac{C - r_1 \lambda_1}{\lambda_2} \quad 9$$

3.2 Generic Network

Figure 4: Generic Network



This section explores the solution method for the generic network shown in Figure 4. The generic network (especially the possible existence of multiple bottlenecks) can result in a number of challenges in meeting the objectives of the PS strategy. These issues are systematically addressed and the properties of the solution methods examined.

It is easy to see that the algebraic solution method adopted for the simple network is too complicated to adopt for the generic network. Thus, its optimization program is derived and solved to global optimum using the Couenne solver in AMPL (Belotti 2009).

The optimization program for the PS strategy is bi-objective, because it requires maximum bottleneck utilization and to distribute the RM delays according to the strategy's objective.

Derivation of the Contributions for a Generalized Network:

The PS strategy requires that the RM delays are distributed between the onramps according to its contribution to the freeway congestion. The contribution of onramp i to bottleneck b can be calculated by generalizing Equation 5 to a multi-bottleneck network. However, it may result in unintentional bias between bottlenecks, because the summation of the contributions to bottleneck b may not equal unity (unless $\lambda_{i,b}=\lambda_{j,b}=1$). i.e.:

$$\sum_{i \in I} \frac{\lambda_{i,b}^2 r_i}{\sum_{j \in I} (\lambda_{j,b} r_j)} \neq 1$$

To remove the unintentional bias between bottlenecks, it is necessary that the summation of all onramp contributions to bottleneck b is equal to one. Otherwise each bottleneck could have a different weight according to the values of $\lambda_{i,b}$'s. Thus, the bottleneck weights are normalized as follows (N.B. normalization of Equation 5 for a simple network is not required as there is only one bottleneck):

$$contribution_{i,b} = \frac{\left(\frac{\lambda_{i,b}^2 r_i}{\sum_{j \in I} \lambda_{j,b} r_j} \right)}{\sum_{k \in I} \left(\frac{\lambda_{k,b}^2 r_k}{\sum_{j \in I} \lambda_{j,b} r_j} \right)} = \frac{\lambda_{i,b}^2 r_i}{\sum_{k \in I} (\lambda_{k,b}^2 r_k)} \quad 10$$

The total contribution of onramp i is equal to the summation of the normalized contribution to each bottleneck:

$$contribution_i = \sum_{b \in B} \left[\frac{\lambda_{i,b}^2 r_i}{\sum_{k \in I} (\lambda_{k,b}^2 r_k)} \right] \quad 11$$

The Objectives of the PS Strategy Cannot be Perfectly Satisfied:

The “ideal” RM delay ($\hat{\delta}_i$) according to the PS strategy is calculated as follows:

$$\begin{aligned} \hat{\delta}_i &= Contribution_i \times \frac{[\delta_1 + \delta_2]}{Contribution_1 + Contribution_2} \\ &= \sum_{b \in B} \left[\frac{\lambda_{i,b}^2 r_i}{\sum_{k \in I} (\lambda_{k,b}^2 r_k)} \right] \times \frac{\sum_{j \in I} (\delta_j)}{\sum_{j \in I} \sum_{b \in B} \left[\frac{\lambda_{j,b}^2 r_j}{\sum_{k \in I} (\lambda_{k,b}^2 r_k)} \right]} \quad 12 \end{aligned}$$

For any solution the *ideal* RM delay can be calculated. The term *ideal* RM delay ($\hat{\delta}_i$) is used because it is typically impossible to realize the *ideal* RM delay across all onramps. The *ideal* RM delay ($\hat{\delta}_i$) is distinguished from the *Optimized* RM delay (δ_i). The *Optimized* RM delay

is associated with the solution that will result in the minimum gap with *ideal* RM delay. Thus, the objective of the PS strategy is to find the *Optimized* RM delay.

There are a number of methods available for measuring the gap ($\hat{\delta}_i - \delta_i$) (Creedy 1998, Ramjerdi 2005, Raffinetti, Siletti et al. 2014, Litman 2015). Whilst some consider the total gap in the system (e.g. Least absolute deviation), others focus purely on the distribution of the gap (e.g. Normalized Gini coefficient). The Root Mean Squared Error (RMSE) is a well-known method used for measuring the differences between values (e.g. a set of predicted and measured values), and is highly sensitive to outliers due to its squared term. The latter is highly desirable in this context, as the minimization of RMSE tends towards solutions that distributes the gap relatively equally among the onramps. Whilst the MinMax approach ensures the minimum gap for the onramp with the largest gap, RMSE also takes into account the total gap created in the system. RMSE is selected as the objective function of the PS strategy due to its desired properties, relative simplicity, and having the same unit as the measured variable.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i \in I} (\hat{\delta}_i - \delta_i)^2}$$

The objective function can be simplified by removing the constant term $\frac{1}{n}$, and the square root (minimising inside the square root is equivalent to minimizing the square root itself).

$$Min(RMSE) = Min\left(\sum_{i \in I} (\hat{\delta}_i - \delta_i)^2\right) \quad 13$$

Subject to:

$$\sum_{i \in I} (r_i \lambda_{j,b}) \leq C_b \quad \forall b \in B \quad 14$$

$$r_i \geq 200 \quad \forall i \in I \quad 15$$

$$r_i \leq q_i \quad \forall i \in I \quad 16$$

Constraint 14 ensures that the bottleneck capacities are never exceeded. Constraints 15 and 16 ensure minimum metering flows and flow conservation respectively.

Maximising Bottleneck Utilization:

The above program minimizes RMSE without maximizing the utilization of bottleneck/s. In keeping with the objective of developing a supplementary algorithm, it is assumed that the nearest upstream onramp to each bottleneck (i.e. *responding onramp*) is responsible for regulating the bottleneck inflow at capacity. Thus, the freeway utilization is maximized by restricting the metering rates at *responding onramps* as constraints. Whilst the metering rates at the remaining *non-responding onramps* are the decision variables in the RMSE minimization problem as discussed above. In other words, the onramp closest to each bottleneck is responsible for keeping it at capacity and its impact on RMSE is controlled indirectly through the combined metering rate of onramps further upstream.

The metering rates of the *responding onramps* are calculated according to Equation 17 (similar to Equation 11). This way the efficiency of every bottleneck is ensured, whilst distributing the metering task in order to minimise RMSE.

$$r_i = \frac{C_b - \sum_{j \in U^i} (\lambda_{j,b} r_j)}{\lambda_{i,b}}$$

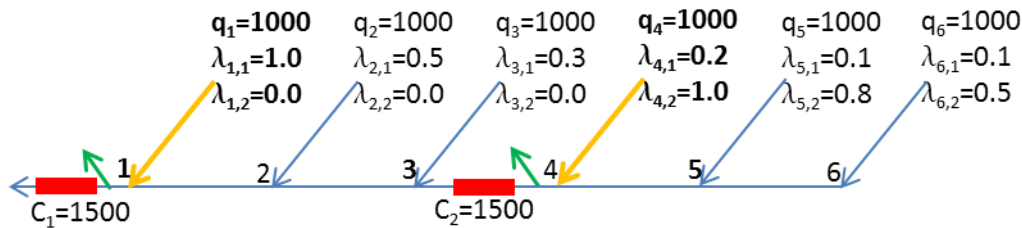
17

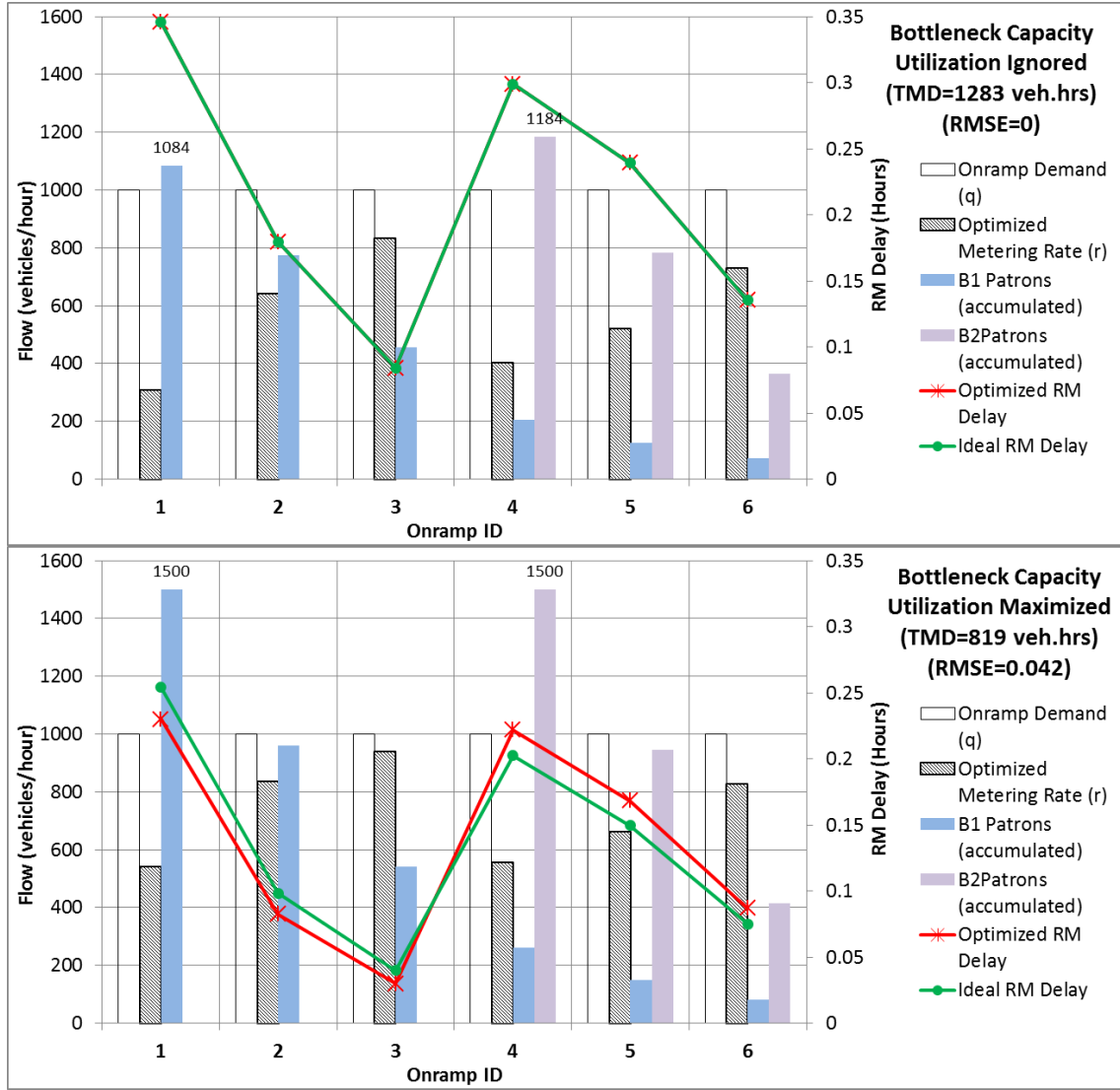
It is theoretically possible that the *responding onramp* is unable to maintain the bottleneck at capacity (e.g. due to limited demand) and the bottleneck would experience capacity drop. However, such a scenario is unlikely as the increased bottleneck accumulation would result in the contributing onramps being assigned significantly more RM delay until the bottleneck is at capacity again. Regardless, the ideal *responding onramp* assigned to optimize a bottleneck would have the highest level of controllability over the bottleneck (which is typically the nearest upstream onramp).

To demonstrate the proposed method of maximizing the freeway utilization the experiment shown in Figure 5 is devised. The *responding onramps* are highlighted as orange onramps. The green off-ramps (placed immediately before the bottlenecks) represent all upstream off-ramps up to the next upstream bottleneck. The experiment is first solved using program 13 to 16, which does not consider the utilization of the two bottlenecks. The results show that RMSE of zero can be achieved. However, the throughput of both bottlenecks is less than 1200 vehicles/hour, much lower than the available capacity of 1500 vehicles/hour. Thus, the system is said to be “starved”, a highly undesirable outcome as the supply of the road network is wasted.

The same experiment was solved by adding Constraint 17 to the program, thus maximising the utilization of the bottlenecks. The results show a small gap (RMSE=0.042 hours) between *ideal* and *optimized* RM delays. This is because the metering rates are forced to completely utilise the bottleneck capacities (i.e. the throughput of both bottlenecks is at capacity). In other words the constraint of maximizing the freeway bottleneck (Constraint 17) limits the solution space which may result in some degree of discrepancy between the ideal and optimized delays. As a result the Total Metering Delay is reduced significantly (more than 30% in this scenario) (Figure 5).

Figure 5: Impact of Maximizing Bottleneck Utilization in the PS Strategy





Permanent Versus Removable Bottlenecks:

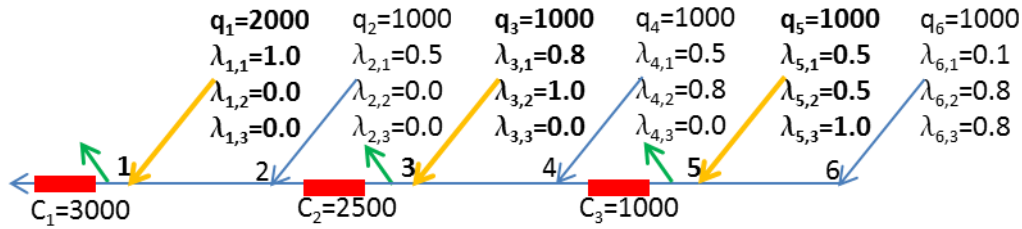
In a multi-bottleneck freeway, optimising the outflow of one bottleneck (according to the PS strategy) may affect the demand at another bottleneck. It is possible that the demand at the affected bottleneck is reduced such that it is unnecessary for its *responding onramp* to continue metering it. More specifically, if the metering rate of a *responding onramp* is calculated to be larger than or equal to its demand (i.e. $r_i \geq q_i$ according to Equation 17), the bottleneck is assumed to be operating in the non-congested regime and is no longer considered a bottleneck. This condition is more likely to occur at bottlenecks with a demand close to capacity. It can be caused by upstream bottleneck/s: the combination of optimized metering rates responding to an upstream bottleneck result in a large portion of traffic exiting the freeway between the two bottlenecks. It can also be caused by downstream bottleneck/s: the onramp contributions to downstream bottleneck/s are large enough that the optimized RM delays result in the demand of an upstream bottleneck to be lower than its capacity.

To demonstrate the above condition the scenario shown in Figure 6 is examined. The maximum traffic that can exit bottleneck 3 is $C_3 = 1000$ vehicles/hour. Thus, the maximum demand of bottleneck 2 is ($\lambda_{6,2}q_6 + \lambda_{4,2}q_4 + \lambda_{3,2}q_3 = 2600$) vehicles/hour (due to $\lambda_{5,2} < \lambda_{6,2}$ and $q_6 = C_3$). The maximum demand of bottleneck 2 is close to its capacity $C_2 = 2500$

vehicles/hour. The following describes how the influences of upstream and downstream bottlenecks reduce the demand of bottleneck 2 such that $r_2 \geq q_2$ (according to Equation 17):

- Effect on bottleneck 2 due to bottleneck 3 (upstream): $\lambda_{5,1} > \lambda_{6,1}$ and $\lambda_{5,3} > \lambda_{6,3}$, thus the contribution of onramp 5 is higher than onramp 6. It is expected that $r_5 < r_6$ after the metering rates are optimized. Thus, the demand for bottleneck 2 is reduced (since $\lambda_{5,2} < \lambda_{6,2}$).
- Effect on Bottleneck 2 due to Bottleneck 1 (downstream): onramp 4 contributes to Bottleneck 1 and 2, resulting in a larger *optimized* RM delay for this onramp, further reducing the demand for Bottleneck 2.

Figure 6: Scenario for Highlighting the effects of Permanent and Removable Bottlenecks



The discussion above highlights that Bottleneck 2 is close to its capacity and that the metering required at the other two bottlenecks would reduce the demand at Bottleneck 2 below its capacity. The top chart of Figure 7 shows the solution for the Program 13 to 17. Constraint 17 imposes an increase of the metering rate at onramp 4, otherwise Constraint 16 would not be satisfied for onramp 3 (i.e. $r_3 \geq q_3$). These constraint restrictions have resulted in an increase of RMSE. Thus, enforcing Constraint 17 on a bottleneck with low demand can significantly restrict the realisation of the PS strategy objectives. A low demand bottleneck should be treated like any other (free-flowing) segment of the freeway, otherwise it artificially increases RMSE simply because it was pre-selected as a bottleneck location.

To increase the flexibility of the *responding onramps* when their bottleneck demand is low, Constraint 17 is replaced with Constraint 18 and 19. The Dummy Variables (DV) are binary variables and allow the solution to bound r_i when Constraints 15 and 16 are active.

Constraint 19 ensures that only one of the DV's is equal to one and the other two equal zero.

$$r_{i_b} = DV1_{i_b,b} \times 200 + DV2_{i_b,b} \times q_{i_b} + DV3_{i_b,b} \times \frac{C_b - \sum_{j \in U^i} (\lambda_{j,b} r_j)}{\lambda_{i_b,b}} \quad 18 \quad \left. \begin{array}{l} \forall b \in B \\ \text{and } i_b \text{ is the} \\ \text{onramp} \\ \text{immediately} \\ \text{upstream of } b \\ \text{(i.e. responding} \\ \text{onramps)} \end{array} \right\} \quad 19$$

$$DV1_{i_b,b} + DV2_{i_b,b} + DV3_{i_b,b} = 1$$

The middle chart of Figure 7 shows the optimized solution given the program 13 to 16 including Constraints 18 and 19. Bottleneck 2 is now able to operating below capacity (i.e. a total inflow of 2382 vehicles/hour and a capacity of 2500 vehicles/hour). Constraint 18 allows the meter at onramp 3 to switch off (i.e. $r_3 = q_3$), without effecting upstream onramp. However, the resulting RMSE is even larger than applying Constraint 17, because onramp 3 contributes to Bottleneck 1 but does not experience any RM delay.

To deal with the above issue, it is proposed that when $r_i \geq q_i$ (according to Equation 17), the bottleneck is removed from the system. More specifically contributions to it are set to zero and its *responding onramp* is released from its duty (i.e. constraint to maximise its bottlenecks utilization). Thus, this ramp meter is free to be optimized directly as a decision variable (in the same way as other *non-responding onramps*). To achieve this, Constraints 18 and 19 are modified as follows:

$$\begin{aligned}
 r_{i_b} &\leq DV1_{i_b,b} \times 200 + DV2_{i_b,b} \times q_i + DV3_{i_b,b} \times \frac{C_b - \sum_{j \in U^i} (\lambda_{j,b} r_j)}{\lambda_{i_b,b}} & 20 \\
 r_i &\geq DV3_{i,b} \times \frac{C_b - \sum_{j \in U^i} (\lambda_{j,b} r_j)}{\lambda_{i,b}} & 21 \\
 \frac{C_b - \sum_{j \in U^i} (\lambda_{j,b} r_j)}{\lambda_{i_b,b}} &> DV2_{i_b,b} \times q_{i_b} & 22 \\
 DV1_{i_b,b} + DV2_{i_b,b} + DV3_{i_b,b} &= 1 & 23 \\
 BC_b &= DV1_{i_b,b} + DV3_{i_b,b} & 24 \\
 \left[C_b - \sum_{j \in I} (\lambda_{j,b} r_j) \right] \times (DV1_{i_b,b} + DV3_{i_b,b}) &= 0 & 25 \\
 \sum_{c \in B} (DV1_{i_c,c} + DV3_{i_c,c}) &\geq 1 & 26
 \end{aligned}$$

$\forall b \in B$
 and i_b is the
 onramp
 immediately
 upstream of b
 (i.e. *responding*
onramps)

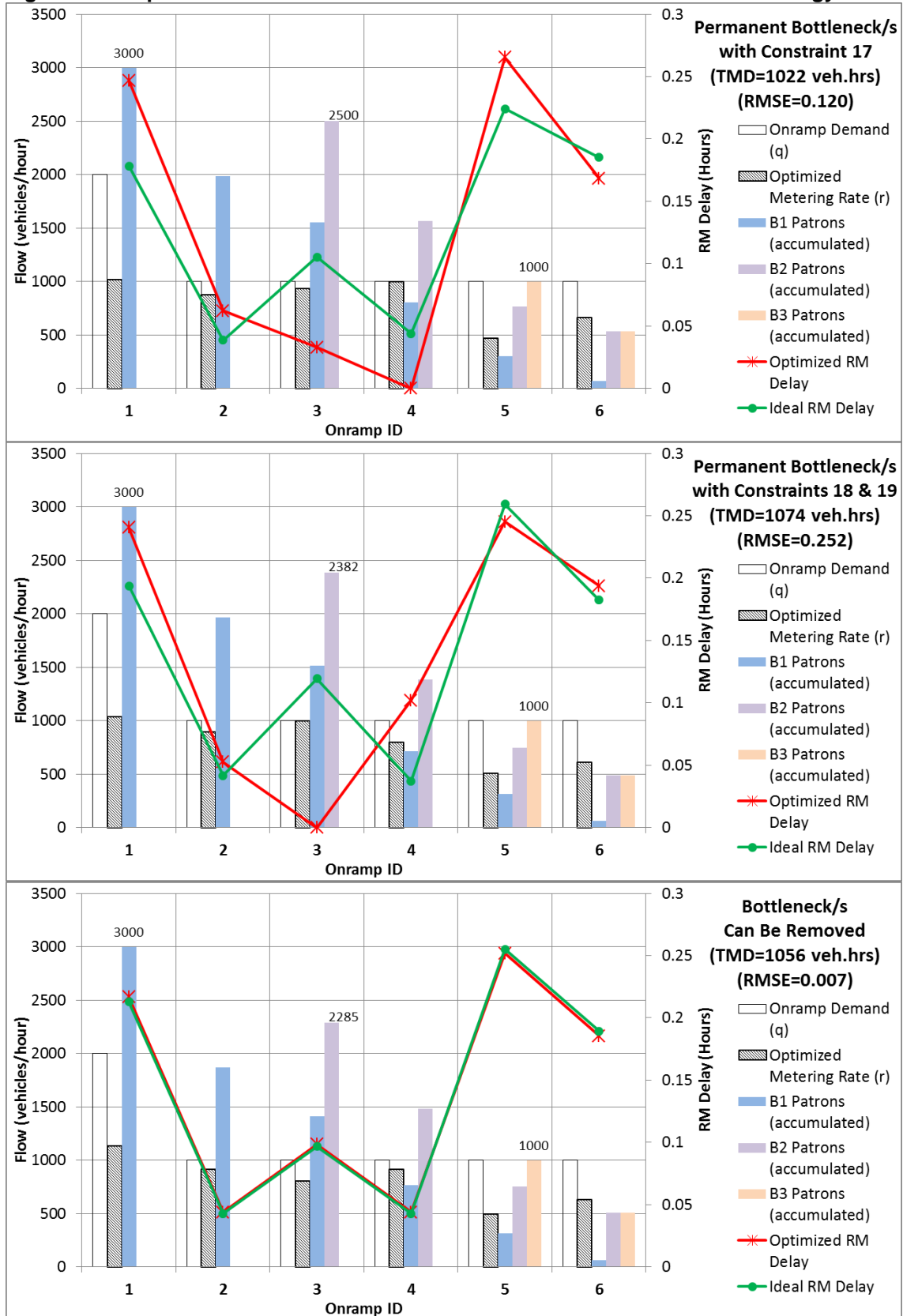
If $DV1=1$ then r_i is bounded to Constraint 15 (i.e. $r_i = 200$). The inequality in Constraint 20 and 21 allows r_i to be restricted to Constraint 17 when $DV3=1$ (i.e. when its bottleneck is active). If $DV2=1$ then it is bounded by $200 \leq r_i \leq q_i$ (i.e. its bottleneck is removed), thus r_i is a decision variable with the same level of freedom as the *non-responding onramps*. Constraint 22 ensures that the only time $DV2=1$ is when $r_i \geq q_i$ (i.e. r_i can only be “free” variable when the condition to remove its bottleneck is met). Constraint 23 forces only one of the DV’s equal one. Constraint 24 ensures that if a bottleneck is removed then any contributions to it are removed from the objective function by setting the variable $BC_b = 0$ (N.B. $BC_b = 1$ if $r_i < q_i$ for its *responding onramp*). Constraints 25 and 26 ensure that at least one bottleneck is forced to be at capacity. If all bottlenecks are allowed to operate below capacity, the objective of maximizing bottleneck utilisation would not been achieved (i.e. it would be possible for all bottlenecks to be removed by metering all bottlenecks below capacity).

In order to allow the BC_b variable to remove or include the contributions to a bottleneck, it is multiplied by the normalized bottleneck contribution (i.e. multiplied by Equation 10). Thus, the *ideal* RM delay ($\hat{\delta}_i$) of the Objective Function 13 is calculated using Equation 27 instead of Equation 12.

$$\hat{\delta}_i = \sum_{b \in B} \left[\frac{BC_b \lambda_{i,b}^2 r_i}{\sum_{k \in I} (\lambda_{k,b}^2 r_k)} \right] \times \frac{\sum_{j \in I} (\delta_j)}{\sum_{j \in I} \sum_{b \in B} \left[\frac{BC_b \lambda_{j,b}^2 r_j}{\sum_{k \in I} (\lambda_{k,b}^2 r_k)} \right]} \quad 27$$

To highlight the effects of removing a bottleneck, the solution to Program 13 to 16 ($\hat{\delta}_i$ calculated using Equation 27) including Constraints 20 to 26 is shown in the bottom chart of Figure 7. The demand for Bottleneck 2 (i.e. 2285 vehicles/hour) is still below capacity. However, r_3 is no longer restricted to the traffic conditions at Bottleneck 2 and is effectively a “free” variable in the same manner as the other *non-responding onramps*. The result shows that the flexibility of removing in-active bottlenecks significantly reduces RMSE. The resulting TMD (i.e. 1056 vehicle.hours) is slightly larger than the TMD obtained when assuming a permanent bottleneck and applying Constraint 17 (i.e. 1022 vehicle.hours). The latter method is in effect increasing the freeway throughput (i.e. disregarding the PS strategy until the low demand bottleneck is activated). A distinction is made between increasing the freeway throughput simply to activate a bottleneck and maximizing the bottleneck utilization when it is already active (which is one of the objectives of the PS strategy).

Figure 7: Comparison of Permanent and Removable Bottlenecks in the PS Strategy



All Bottlenecks are Not Equal:

Equation 10 normalizes the contribution to each bottleneck, therefore all bottlenecks are considered equal. However, as shown in Figure 7, some bottlenecks carry a larger number of vehicles. In addition, once the perfect metering assumption is removed, each bottleneck may contain a different level of congestion. Thus, when calculating the contribution to congestion across multiple bottlenecks, it is necessary to distinguish between bottlenecks by weighing them according to their throughput and level of congestion.

Two methods for intentionally differentiating between bottlenecks were explored, including the average *bottleneck delay* and *bottleneck count*. Both of which can be easily calculated using count detectors at the entrance and exit of the bottleneck segment. The advantage of using the average *bottleneck delay* (of the vehicles exiting the bottleneck segment in an update-period) is that it is sensitive to the congestion level and uses the same unit as the decision variable (i.e. delay). However, *bottleneck count* is able to capture both the congestion level and the throughput difference between bottlenecks. Thus, *Bottleneck Count* (BC_b) is selected to weigh bottlenecks. Due to the *perfect metering* assumption (i.e. bottleneck queue build-up does not occur), BC_b is equal to the bottleneck outflow during an update-period. The complete program 28 to 38 of PS strategy for a generic network is presented below. Constraint 36 includes the total bottleneck outflow as a weighting factor.

Figure 8 highlights the difference between equal bottleneck weights and bottlenecks weighted using bottleneck counts. The demand scenario is the same as Figure 6, and the top chart in Figure 8 is the same solution as the bottom chart in Figure 7. The bars in Figure 8 highlight the contribution to each bottleneck. There are no contributions to Bottleneck 2 as it has been removed as described previously. The middle chart in Figure 8 highlights the solution with weighted bottlenecks. The contributions to Bottleneck 3 are reduced by a third when weighted. The resulting change in contribution pushes a larger *RM delay* at onramps contributing to Bottleneck 1. However, the capacity Constraint 29 restricts this desire. The bottom chart of Figure 8 shows the changes to on the *optimum* RM delay due to the bottleneck weights. Onramp 1 is assigned more delay as it is the main contributor to Bottleneck 1. The total contribution of onramp 5 is reduced more than onramp 6, because onramp 5 is the main contributor to Bottleneck 3 (the weight of which is reduced). Thus, onramp 6 is assigned relatively more RM delay when bottleneck weights are included.

$$Min(RMSE) = Min \sum_{i \in I} \left(\frac{\sum_{b \in B} \left[\frac{BC_b \lambda_{i,b}^2 r_i}{\sum_{k \in I} (\lambda_{k,b}^2 r_k)} \right] \times \sum_{j \in I} (\delta_j)}{\sum_{j \in I} \sum_{b \in B} \left[\frac{BC_b \lambda_{j,b}^2 r_j}{\sum_{k \in I} (\lambda_{k,b}^2 r_k)} \right]} - \delta_i \right)^2 \quad 28$$

Subject to:

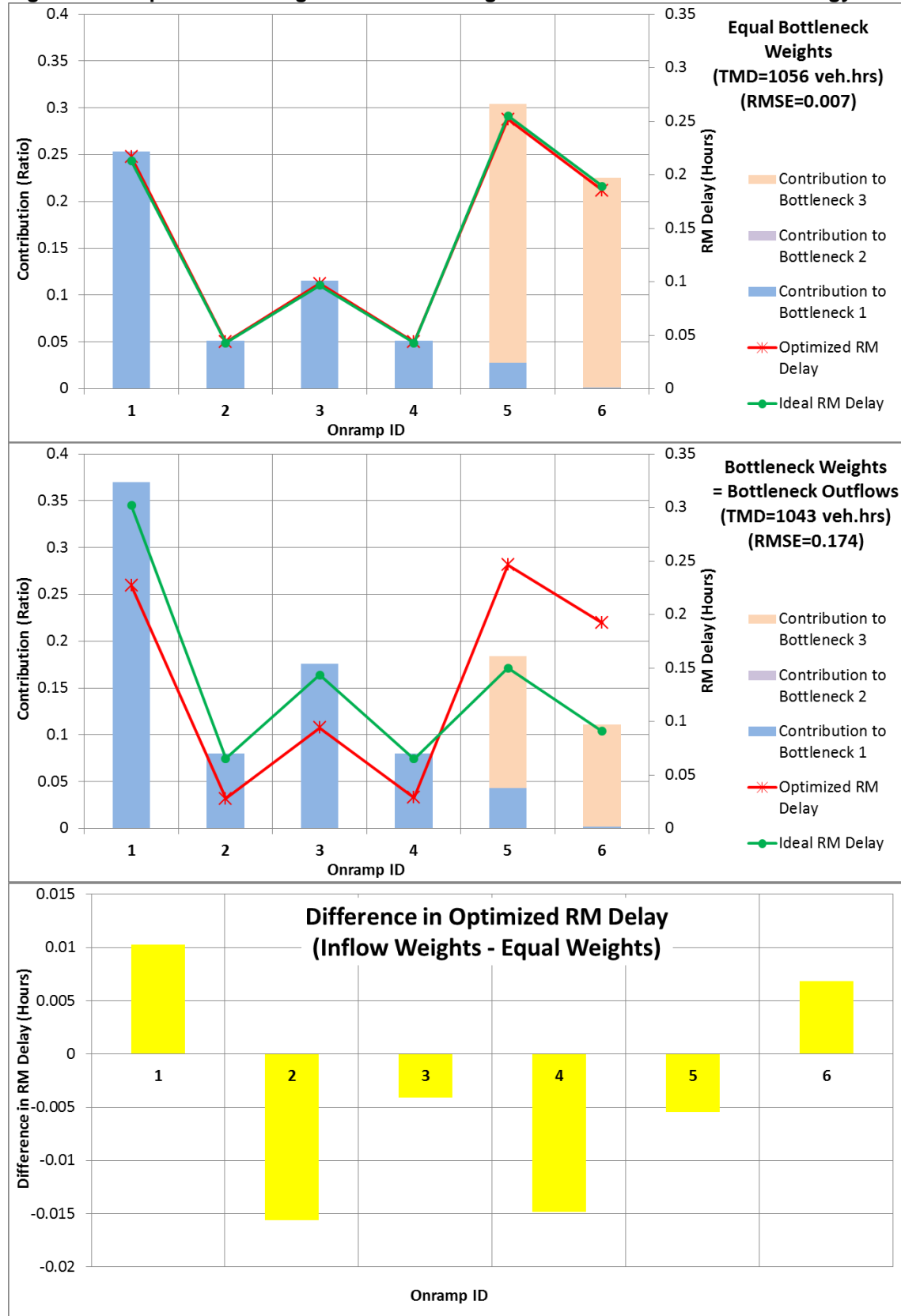
$$\sum_{i \in I} (r_i \lambda_{j,b}) \leq C_b \quad \forall b \in B \quad 29$$

$$r_i \geq 200 \quad \forall i \in I \quad 30$$

$$r_i \leq q_i \quad \forall i \in I \quad 31$$

$$\left. \begin{aligned} r_{i_b} &\leq DV1_{i_b,b} \times 200 + DV2_{i_b,b} \times q_i + DV3_{i_b,b} \times \frac{C_b - \sum_{j \in U^i} (\lambda_{j,b} r_j)}{\lambda_{i_b,b}} & 32 \\ r_i &\geq DV3_{i,b} \times \frac{C_b - \sum_{j \in U^i} (\lambda_{j,b} r_j)}{\lambda_{i,b}} & 33 \\ \frac{C_b - \sum_{j \in U^i} (\lambda_{j,b} r_j)}{\lambda_{i_b,b}} &> DV2_{i_b,b} \times q_{i_b} & 34 \\ DV1_{i_b,b} + DV2_{i_b,b} + DV3_{i_b,b} &= 1 & 35 \\ BC_b &= \sum_{j \in U^i} (\lambda_{j,b} r_j) \times (DV1_{i_b,b} + DV3_{i_b,b}) & 36 \\ \left[C_b - \sum_{j \in I} (\lambda_{j,b} r_j) \right] &\times (DV1_{i_b,b} + DV3_{i_b,b}) = 0 & 37 \\ \sum_{c \in B} (DV1_{i_c,c} + DV3_{i_c,c}) &\geq 1 & 38 \end{aligned} \right\} \begin{array}{l} \forall b \in B \\ \text{and } i_b \text{ is the} \\ \text{onramp} \\ \text{immediately} \\ \text{upstream of } b \\ \text{(i.e. responding} \\ \text{onramps)} \end{array}$$

Figure 8: Comparison of Weighted and Non-weighted Bottlenecks in the PS Strategy



4. Comparison with SO and TE Strategies

4.1 Derivation of the System Optimal (SO) Program

In the SO strategy the objective is to minimize TSTT, which is equivalent to maximising throughput (Papageorgiou 1980). The the SO strategy can be achieved by solving the optimization program 39 to 42.

$$\text{Min} \left(- \sum_{i \in I} r_i \right) \quad 39$$

Subject to:

$$\sum_{i \in I} (r_i \lambda_{j,b}) \leq C_b \quad \forall b \in B \quad 40$$

$$r_i \geq 200 \quad \forall i \in I \quad 41$$

$$r_i \leq q_i \quad \forall i \in I \quad 42$$

To maximize throughput, the objective function maximises the metering rates across all onramps, whilst ensuring the freeway demand is at or below the capacity of all bottlenecks (i.e. Constraint 40). Constraints 41 and 42 ensure minimum metering flows and flow conservation respectively.

4.2 Derivation of the Typical Equity (TE) Program

The TE strategy can be solved with the optimization program for SO if the following constraint is added to the program. Constraint 43 requires that all onramps have equal RM delays (derived from Equation 1).

$$\frac{r_1}{q_1} = \frac{r_i}{q_i} \quad \forall i \in I \quad 43$$

Objective Function 39 ensures optimum utilization of the freeway capacity, and Constraint 43 ensures the RM delays are distributed according to the TE strategy.

4.3 Comparative Results

As an example the scenario shown in Figure 9 is optimized (globally) for the SO, TE, and the final PS strategy programs using AMPL. The results of the three strategies are highlighted in Figure 10. The *ideal* RM delay is calculated using Equation 27 and is utilised only in the PS strategy. However, it is displayed with the optimized results of the SO and TE strategies for comparison purposes, and to highlight the interaction between contributions and the *optimized* RM delays.

Figure 9: Example Scenario for Comparison of the Three RM Strategies

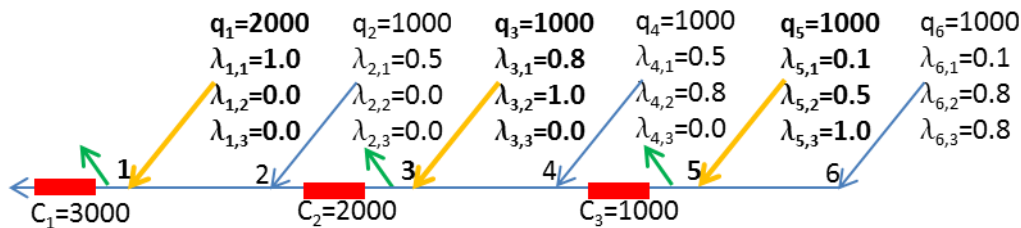
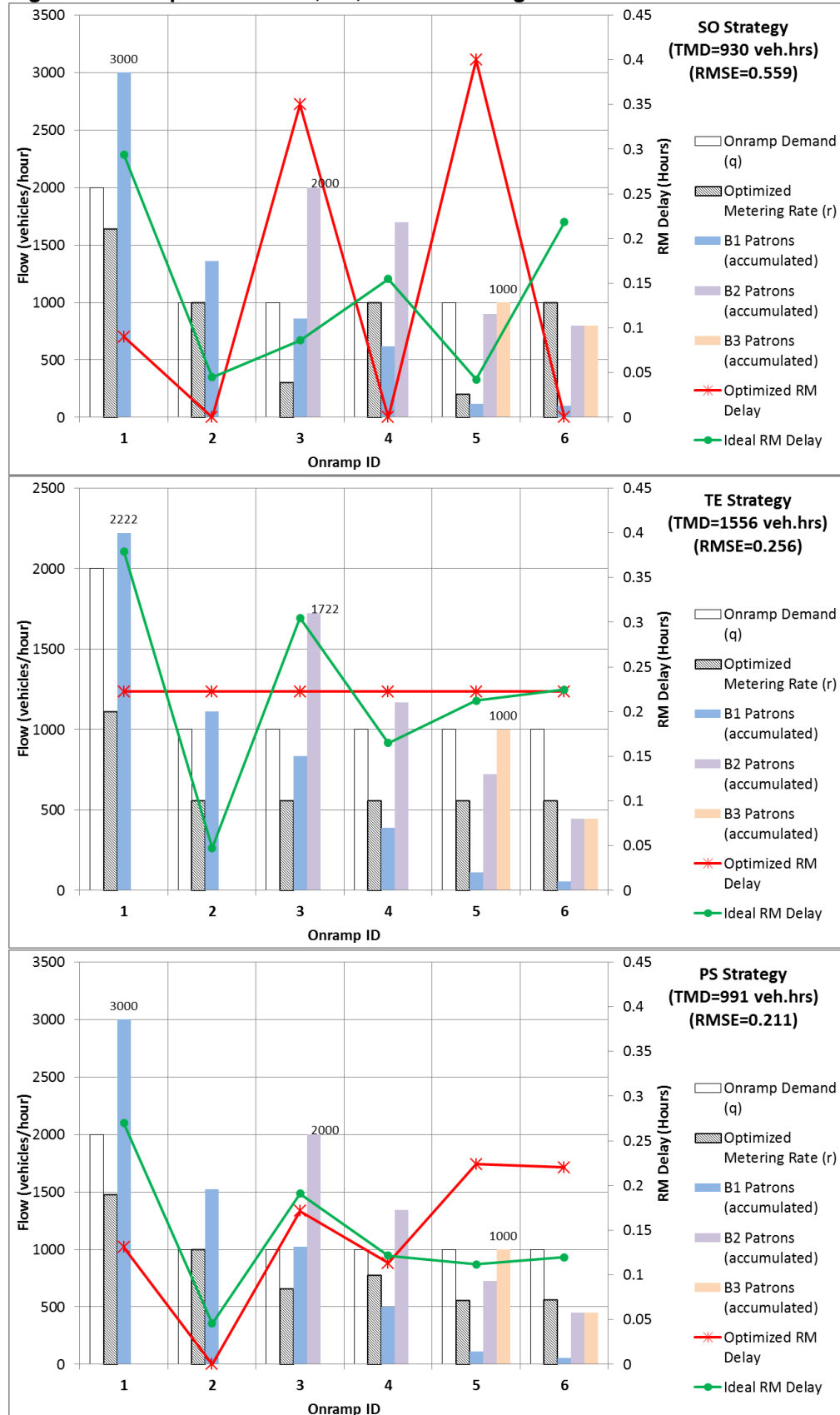


Figure 10: Comparison of SO, TE, and PS Strategies



All three bottlenecks operate at capacity in solution of the SO strategy. The strategy assigns the entire metering task for Bottleneck b to the onramp with the largest proportion of *bottleneck patrons* ($\lambda_{i,b}$). As described in Framework 1 this strategy minimizes *dead time*. However, the *ideal* RM delay shows that the onramps that are switched off have a large contribution in comparison to the metered onramps. As a result a large RMSE=0.559 hours is observed.

All bottlenecks are at capacity in the PS strategy. The constraints that maximise the bottleneck utilisation, limit the degree to which the *ideal* RM delay can be realized. However, in comparison to the SO strategy most ramp meters are switched on and *RM delays* are distributed according to the onramp contributions as best as possible. The resulting TMD=991 vehicle.hours is 7% more than the SO strategy (as compared to 67% in the TE strategy), whilst the RMSE=0.211 hours which is less than half the RMSE of the SO strategy (and smaller than the TE strategy).

The PS strategy reduces other side-effects as well. For example, the maximum RM delay experienced by an individual vehicle is 0.45, 0.44, 0.8 hours for the PS, TE, and SO strategies respectively. In addition, the maximum onramp queue at the end of the one hour period is 524, 899, 800 vehicles for the PS, TE, and SO strategies respectively.

5. Conclusions

The foundation of a supplementary algorithm is developed with the aim of alleviating some of the negative side effects of ramp metering. The generalized optimization program and its associated solution algorithm is fit for incorporation into a Model Predictive Control (MPC) framework to realise the PS strategy in the real world (i.e. subject of future research). In addition an analytical framework has been developed and used to examine the properties of the proposed solution. The proposed solution (PS) was compared to System Optimal (SO) and Typical Equity (TE) ramp metering strategies. The comparison reveals that the proposed strategy can operate close to system optimal efficiency, whilst significantly improving equity, reducing maximum RM delay and the maximum onramp queue length. In future research its impact on other side effects such as diversions and urban sprawl will be examined.

References

- Amini, N., C. Aydos, K. Wijayaratna, H. Grzybowska and S. T. Waller (2015) *A Network Wide Evaluation of the SCATS Ramp Metering System Using Microsimulation*. Australian Transport Research Forum (ATRF) Sydney, Australia.
- Amini, N., L. Gardner and S. T. Waller (2016) "A New Horizontal Equity Measure for Ramp Meters." *Transportation Research Record: Journal of the Transportation Research Board*(In Press).
- Amini, N., H. Grzybowska, K. Wijayaratna and S. T. Waller (2015) *Systemic Evaluation of the HERO-Based Ramp Metering Algorithm Using Microsimulation*. IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas de Gran Canaria, Spain.
- Arnold Jr, E. (1998) "Ramp metering: a review of the literature." *Contract* 21535: 6940.
- Belotti, P. (2009) *Couenne: a user's manual*, Technical report, Lehigh University.
- Burley, M. and J. Gaffney (2013) *Managed Freeways: Freeway Ramp Signals Handbook*. VicRoads.
- Creedy, J. (1998) *The dynamics of inequality and poverty : comparing income distributions*. Cheltenham
- Northampton, Mass., Cheltenham
- Northampton, Mass. : E. Elgar.
- Faulkner, L., F. Dekker, D. Gyles, I. Papamichail and M. Papageorgiou (2013) "Evaluation of HERO Ramp Metering Installation at the M1/M3 Freeway in Queensland,

- Australia." *Transportation Research Record: Journal of the Transportation Research Board*.
- Haj-Salem, H. and M. Papageorgiou (1995) "Ramp metering impact on urban corridor traffic: Field results." *Transportation Research Part A: Policy and Practice* 29(4): 303-319.
- Jacobson, L., J. Stribiak, L. Nelson and D. Sallman (2006) Ramp management and control handbook.
- Kotsialos, A. and M. Papageorgiou (2004) "Efficiency and equity properties of freeway network-wide ramp metering with AMOC." *Transportation Research Part C: Emerging Technologies* 12(6): 401-420.
- Lawson, T., D. Lovell and C. Daganzo (1997) "Using input-output diagram to determine spatial and temporal extents of a queue upstream of a bottleneck." *Transportation Research Record: Journal of the Transportation Research Board*(1572): 140-147.
- Levinson, D. and L. Zhang (2006) "Ramp meters on trial: Evidence from the Twin Cities metering holiday." *Transportation Research Part A: Policy and Practice* 40(10): 810-828.
- Litman, T. (2015) Evaluating transportation equity, Victoria Transport Policy Institute.
- Lovell, D. J. and J. R. Windover (1999) "Analyzing Freeway Traffic under Congestion: Traffic Dynamics Approach." *Journal of Transportation Engineering* 125(4): 373-375.
- Mizuta, A., K. Roberts, L. Jacobsen and N. Thompson (2014) Ramp Metering: A Proven, Cost-Effective Operational Strategy-A Primer, FHWA.
- Papageorgiou, M. (1980) "A new approach to time-of-day control based on a dynamic freeway traffic model." *Transportation Research Part B: Methodological* 14(4): 349-360.
- Papageorgiou, M. and A. Kotsialos (2002) "Freeway ramp metering: an overview." *Intelligent Transportation Systems, IEEE Transactions on* 3(4): 271-281.
- Papageorgiou, M. and I. Papamichail (2007) "Handbook of Ramp Metering. Document for EUROpean RAMP Metering Project (EURAMP)." *Rapport Technique*.
- Piotrowicz, G. and J. Robinson (1995) Ramp Metering Status in North America. 1995 Update.
- Raffinetti, E., E. Siletti and A. Vernizzi (2014) "On the Gini coefficient normalization when attributes with negative values are considered." *Statistical Methods & Applications*: 1-15.
- Ramjerdi, F. (2005) *An evaluation of the performances of equity measures. Proceedings of the 45th ERSA Conference*.
- Zhang, L. and D. Levinson (2004) "Optimal freeway ramp control without origin-destination information." *Transportation Research Part B: Methodological* 38(10): 869-887.
- Zhangl, L. and D. Levinson (2002) "Estimation of the Demand Responses to Ramp Metering." *American Society of Civil Engineers*.

Notations

Sets

- I Set of all onramps
- T Set of all discretised update-periods during the study
- B Set of all freeway bottleneck
- U^i Set of onramps upstream of onramp $i \in I$

Parameters

$C_{b,t}$	Capacity of bottleneck $b \in B$ during update-period $t \in T$ (vehicles/hour)
q_i	Inflow demand for onramp $i \in I$ (vehicles/hour)
$\lambda_{i,b}$	Proportion of bottleneck patrons from onramp $i \in I$ and bottleneck $b \in B$
n	The total number of onramps included in the study
τ	The total number of update periods included in the study
β	The total number of bottlenecks included in the study

Variables

i	the index of an individual onramp $i \in I$
t	the index of an individual discretised update-period $t \in T$
b	the index of an individual bottleneck $b \in B$
$r_i(t)$	Metering rate at onramp $i \in I$ during update-period $t \in T$ (vehicles/hour)
$\delta_{i,t}$	Optimized (applied) RM delay of onramp $i \in I$ during update-period $t \in T$ (seconds per vehicle)
$\hat{\delta}_{i,t}$	Ideal RM delay of onramp $i \in I$ during update-period $t \in T$ (seconds per vehicle)
$DV_{i,b,t}$	Dummy variable associated with onramp $i \in I$ and bottleneck $b \in B$ during update-period $t \in T$
$BC_{b,t}$	Bottleneck Count (weighing factor) associated with bottleneck $b \in B$ during update-period $t \in T$. Comprised of bottleneck outflow plus accumulation of vehicles in the bottleneck segment during an update period.