On Simulation and Optimization of Freeway Network Operations

B. Wiwatanapataphee, Yong Hong Wu, C Gu
Curtin University of Technology

Progress from last PSG meeting

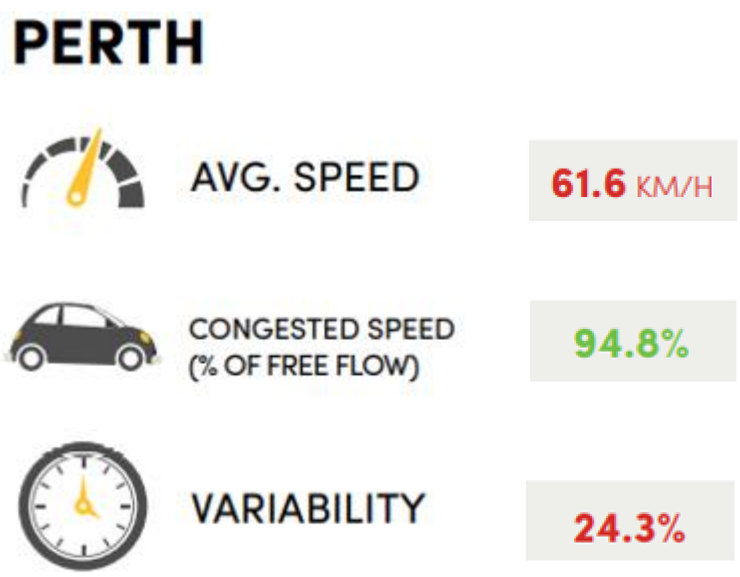
1 Freeway traffic control via Ramp Metering and Variable Speed Limit using a mesoscopic model

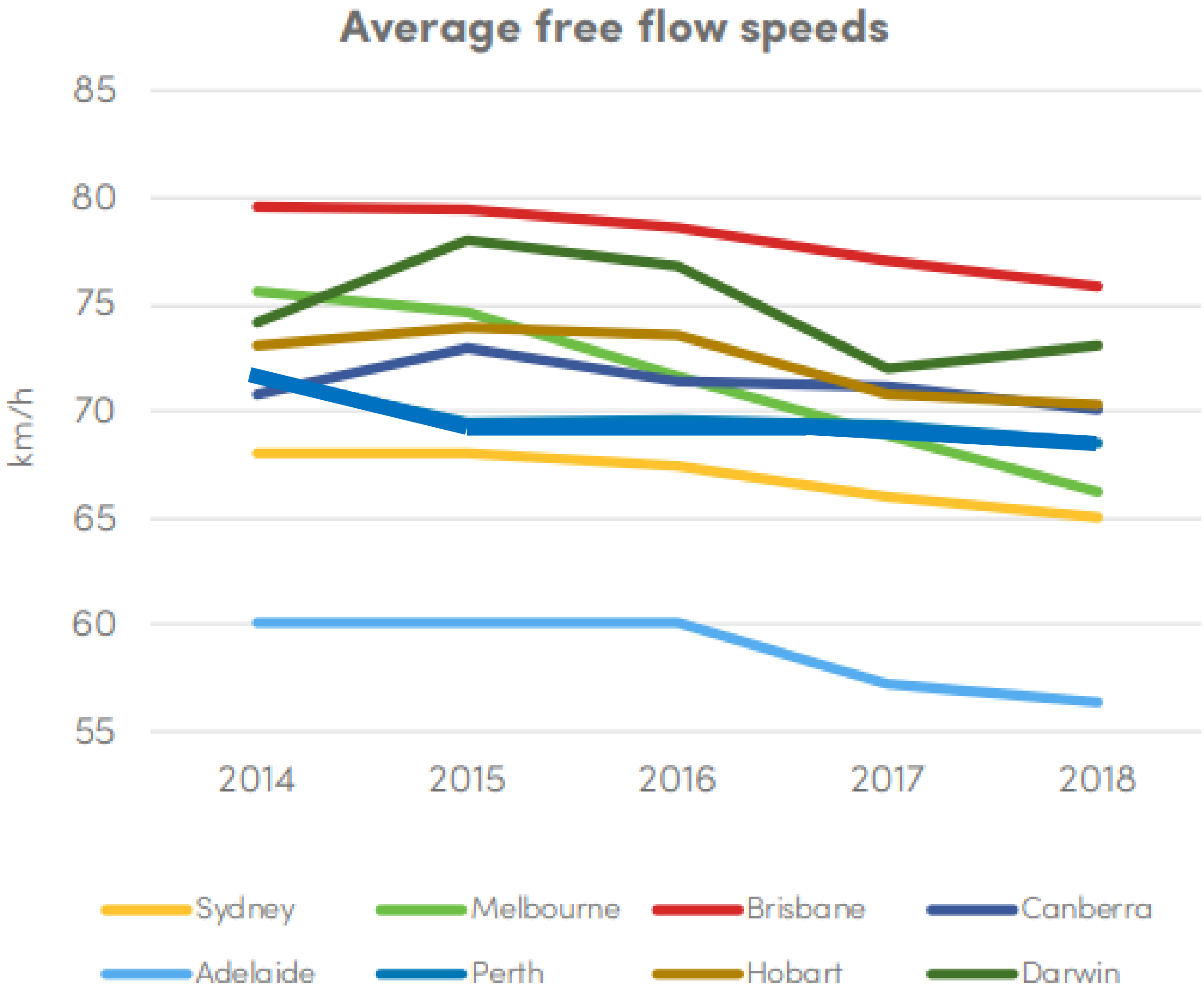
2. Control of ramp metering based on reinforcement learning

Freeway Traffic Control via Ramp Metering and Variable Speed Limit using a mesoscopic model

B. Wiwatanapataphee & YH Wu
Curtin University of Technology

Background





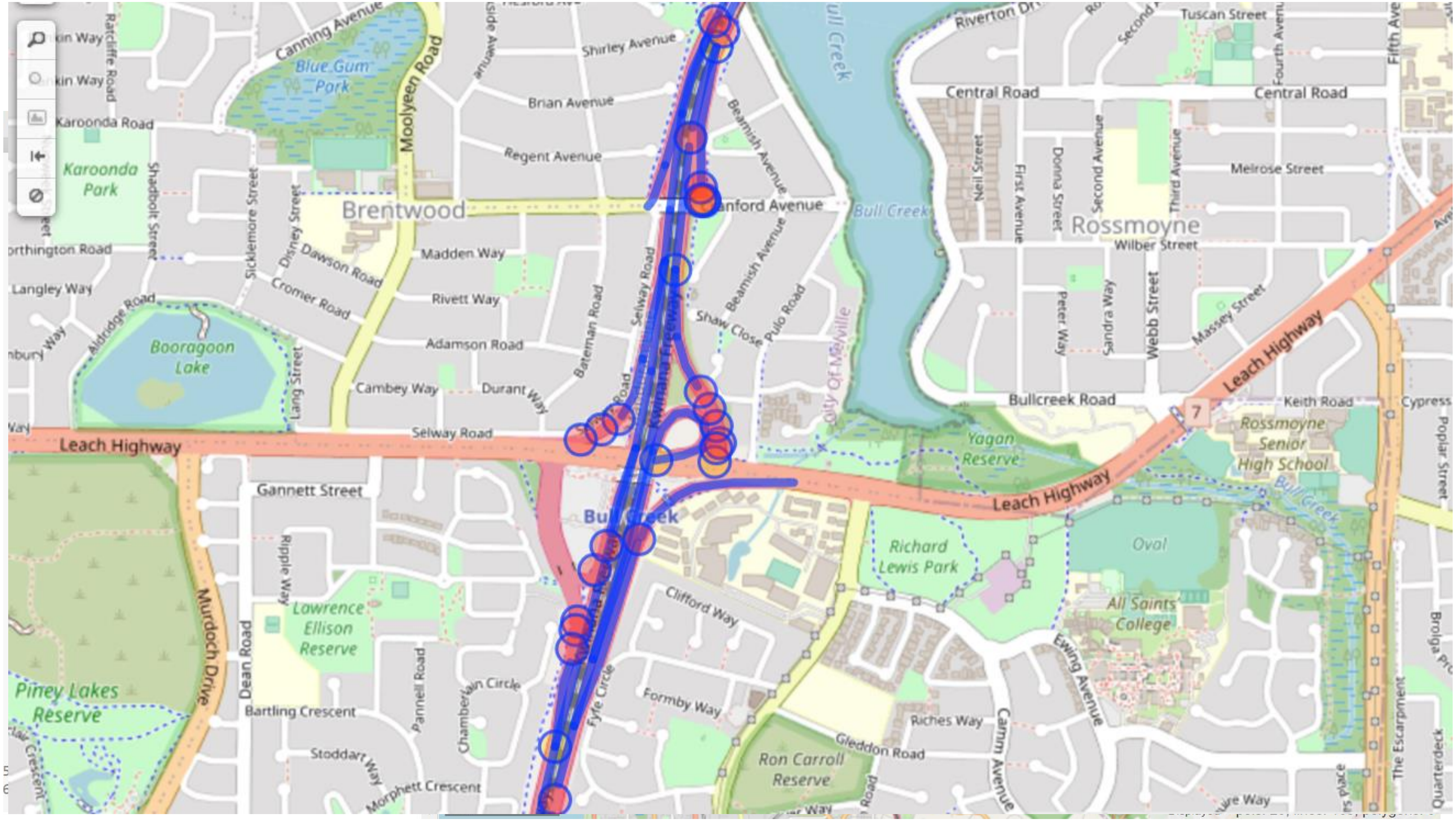
https://www.mainroads.wa.gov.au/about-main-roads/news-media/smart-freeway-technology-upgrades/

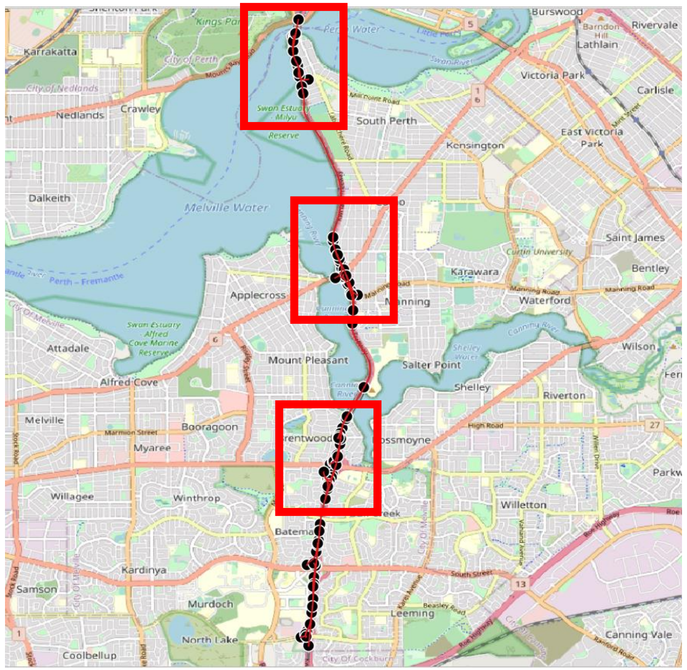
Traffic monitoring and Control systems

- Traffic congestion has a significant impact on economic activity throughout much of the world.
- An essential step towards active congestion control is the creation of accurate, reliable traffic monitoring and control systems.
- These systems usually run algorithms which rely on mathematical models of traffic used to power estimation and control schemes.

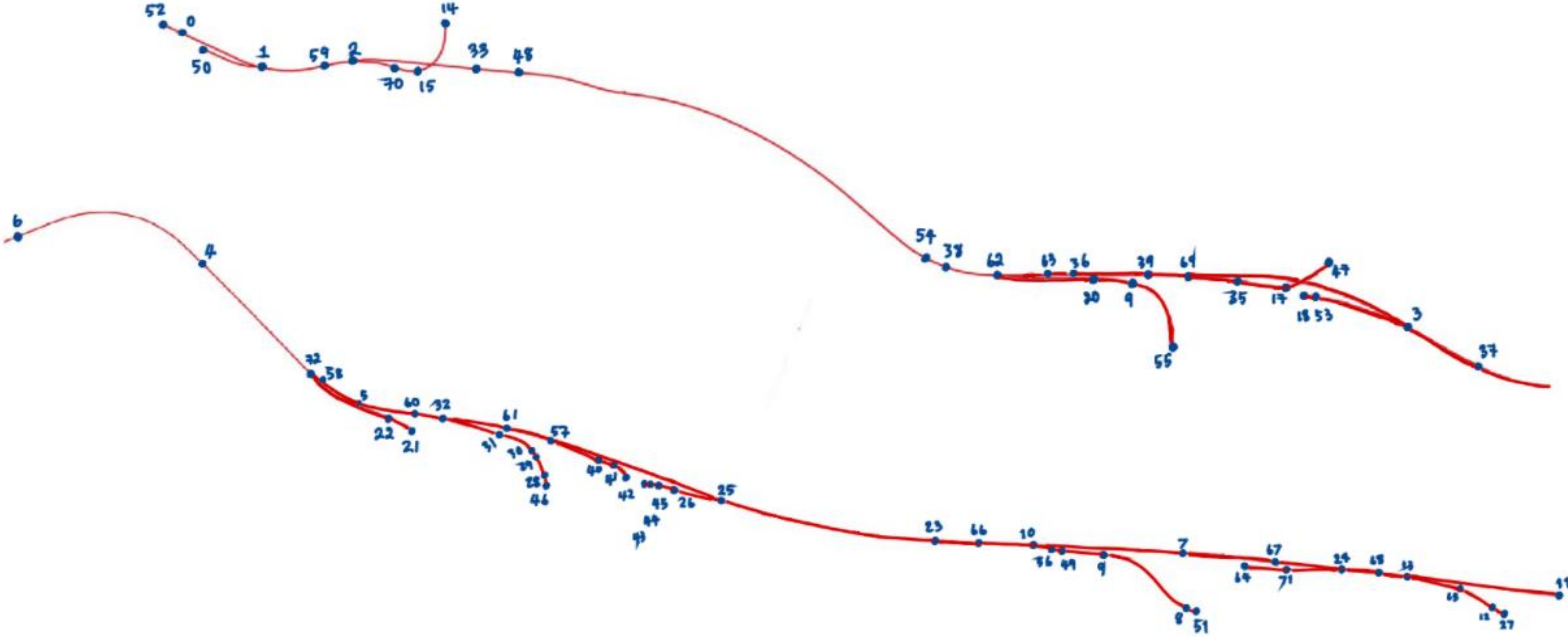
Road Network

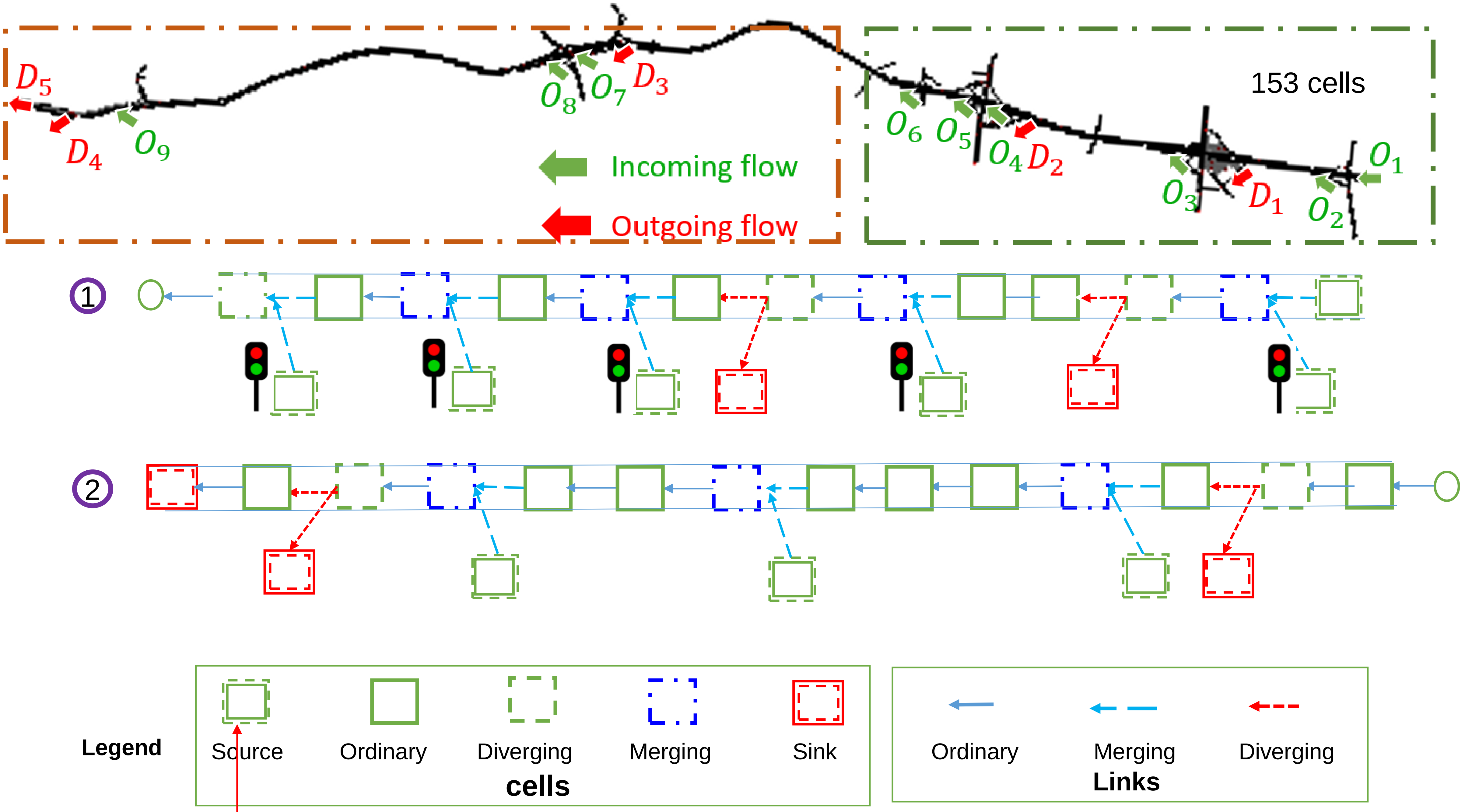

Map
Satellite
© 2021 Google
© 2021 Google
© 2021 Google
Google
Keyboard shortcuts
Imagery ©2021 TerraMetrics
Terms of Use
Report a map error


Canning Avenue
Blue Gum Park
Moolyeen Road
Shirley Avenue
Brian Avenue
Regent Avenue
Riverton Drive
Tuscan Street
Central Road
Central Road
Fourth Avenue
Fifth Ave
Ratcliffe Road
Karoonda Road
Karoonda Park
Shadbolt Street
Sicklemore Street
Disney Street
Brentwood
Beamish Avenue
Bull Creek
Neil Street
First Avenue
Donna Street
Second Avenue
Third Avenue
Melrose Street
Rossmoyne
Wilber Street
Madden Way
Dawson Road
Cromer Road
Rivett Way
Adamson Road
Cambey Way
Durant Way
Bateman Road
Selway Road
Kwinana Freeway
Shaw Close
Pulo Road
City Of Melville
Langley Way
Aldridge Road
Booragoon Lake
Lang Street
Peter Way
Sandra Way
Webb Street
Massey Street
Leach Highway
Bullcreek Road
Keith Road
Cypress
Leach Highway
Selway Road
Yagan Reserve
Rossmoyne Senior High School
Gannett Street
Bull Creek
Leach Highway
Richard Lewis Park
Oval
Ripple Way
Murdoch Drive
Lawrence Ellison Reserve
Clifford Way
All Saints' College
Dean Road
Pannell Road
Fyfe Circle
Formby Way
Ewing Avenue
Piney Lakes Reserve
Bartling Crescent
Riches Way
Camm Avenue
Stoddart Way
Gleddon Road
Ron Carroll Reserve
Morphett Crescent
Poplar Street
The Escarpment
Quarterdeck


Karrakatta
Nedlands
Crawley
Dalkeith
Melville Water
Applecross
Attadale
Alfred Cove
Mount Pleasant
Melville
Booragoon
Myaree
Willagee
Winthrop
Kardinya
Samson
Murdoch
North Lake
Coolbellup
Burswood
Rivervale
Lathlain
Victoria Park
Carlisle
South Perth
Kensington
East Victoria Park
Saint James
Bentley
Karawara
Waterford
Salter Point
Shelley
Riverton
Wilson
Willetton
Leeming
Canning Vale


Brian Avenue
Regent Avenue
Brentwood
Cranford Avenue
Madden Way
Ric Vosper Reserve
Rivett Way
Beddington Way
Adamson Road
Bateman Road
Selway Road
Cambey Way
Durant Way
Moolyeen Road
Mandurah PSP
Kwinana Freeway
Beamish Avenue
The Esplanade
Freeth Court
Shaw Close
Pulo Road
Brentwood Avenue
Spinaway Crescent
Bull Creek
Neil Street
Creekview Close
Bateman Park
Yagan Reserve
Leach Highway
Kwinana Freeway On Ramp
Amur Place
Electra Street
Venus Street
Noalimba Crescent
Mandalia Crescent
Clifford Way
Spitfire
Richard Lewis Park
Camm Avenue
Hartung Way
Salmond Way
Riches Way
Gleddon Road
Ron Carroll
Gemmell Park
Formby Way
Bull Creek Drive
Fyfe Circle
Saxon Place
Wray Close
Chamberlain Circle
Rummer Way
Pannell Road
Barling Crescent





Estimated demand using CNN deep learning every 15 minutes

The DE-CTM optimisation model:

$$TTS = \min \left[\sum_{t=\tau}^{T} \sum_{i \in C} T L_i k_i(\alpha, \beta, t) \right]$$

$\boldsymbol{\alpha} \in [\boldsymbol{\alpha}_{\boldsymbol{min}}, \boldsymbol{\alpha}_{\boldsymbol{max}}]$: probability from upstream normal cell;

$\mathbf{1} - \boldsymbol{\alpha}$: probability from upstream merge cell

$\boldsymbol{\beta} \in [\boldsymbol{\beta}_{\boldsymbol{min}}, \boldsymbol{\beta}_{\boldsymbol{max}}]$: VSL rate

PDE:

$$\frac{n_i \partial k_i}{\partial t} + \frac{\partial (n_i q_i)}{\partial x} = g(\alpha, \beta, t)$$

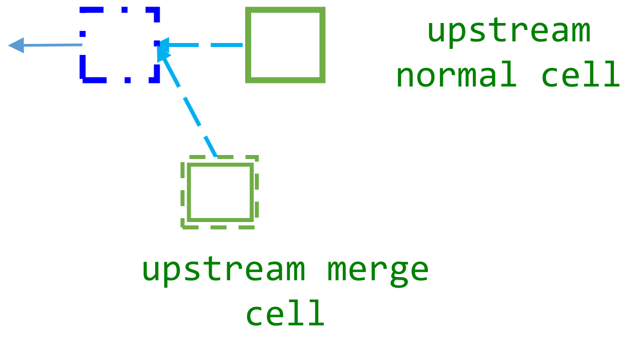
$$\boldsymbol{r}_{\boldsymbol{i},\boldsymbol{j}}(\boldsymbol{t}) = \mathbf{min} \left[(\mathbf{1} - \boldsymbol{\alpha}) \boldsymbol{q}_{\boldsymbol{i}-\mathbf{1}}(\boldsymbol{t}); \quad \boldsymbol{Q}_{\boldsymbol{r}}; \quad \boldsymbol{m}_{\boldsymbol{i}}(\boldsymbol{t} - \mathbf{1}) + \frac{\boldsymbol{T}}{\boldsymbol{L}_{\boldsymbol{j}}} \left(\boldsymbol{d}_{\boldsymbol{i},\boldsymbol{j}} + \boldsymbol{r}_{\boldsymbol{i},\boldsymbol{j}}(\boldsymbol{t} - \mathbf{1}) \right) \right], \ \boldsymbol{j} = \mathbf{1}, \ldots, \mathbf{8}$$

Constraints:

$$\sum_{t=\tau}^{\tau + H_c - 1} T \left\{ \sum_{i \in C_M} r_{i,j}(t) + n_i (q_i(t) - Q_i) \right\} \leq 0, \qquad j = 1, \ldots, 8$$

$$\sum_{t=\tau}^{\tau + H_c - 1} T \left\{ \sum_{i \in C_{VSL}} n_i k_i(t) \left(\boldsymbol{\beta}(\boldsymbol{i}, \boldsymbol{t}) \, v_f \right) - n_i Q_i \right\} \leq 0$$

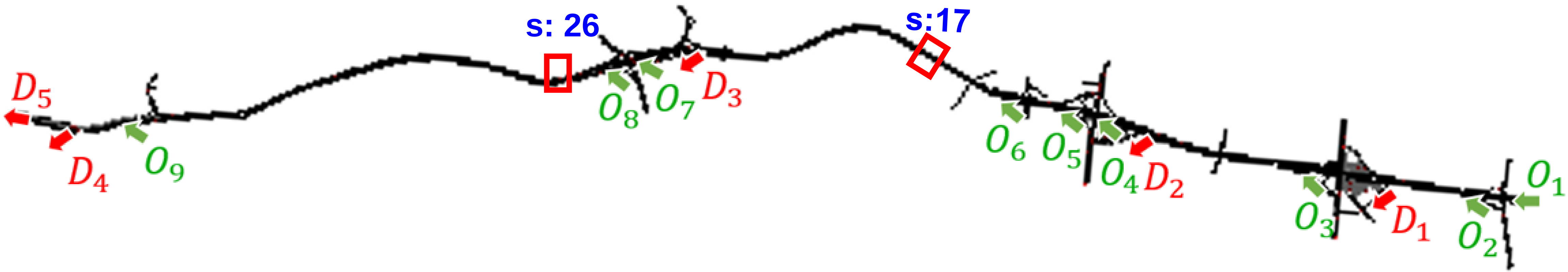
$$0 \leq k_i(t) \leq k_{jam}$$



Some results $\alpha_m \in [-0.25, 0.25], \beta_{VSL} \in [0.25, 1.0]$

Let q_{max} = 1800 vph on freeway q_{max}= 1300 vph on on/off-ramp

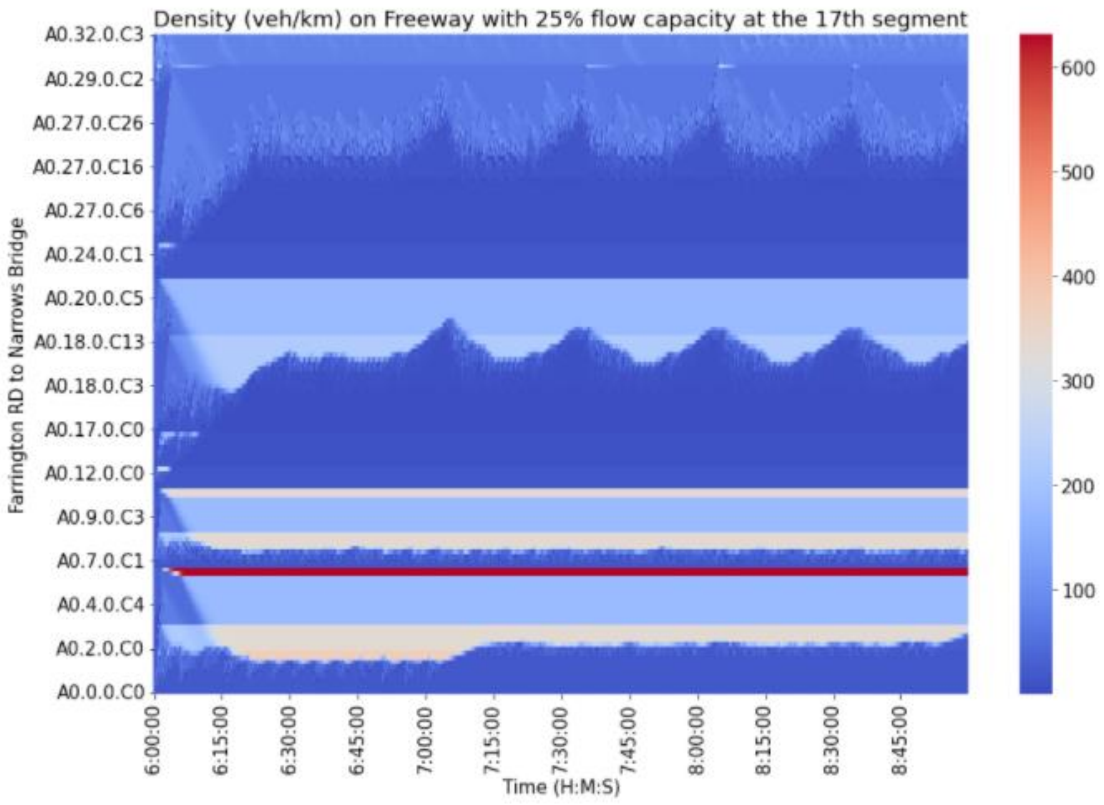
Case study		Segment	Free speed (km/hr)	Capacity(veh/hr)
I	a	All segments	100	1800
	b	All segments	70	1800
II	a	17	100	450
	b	26	100	450
III	a	17	70	450
	b	26	70	450



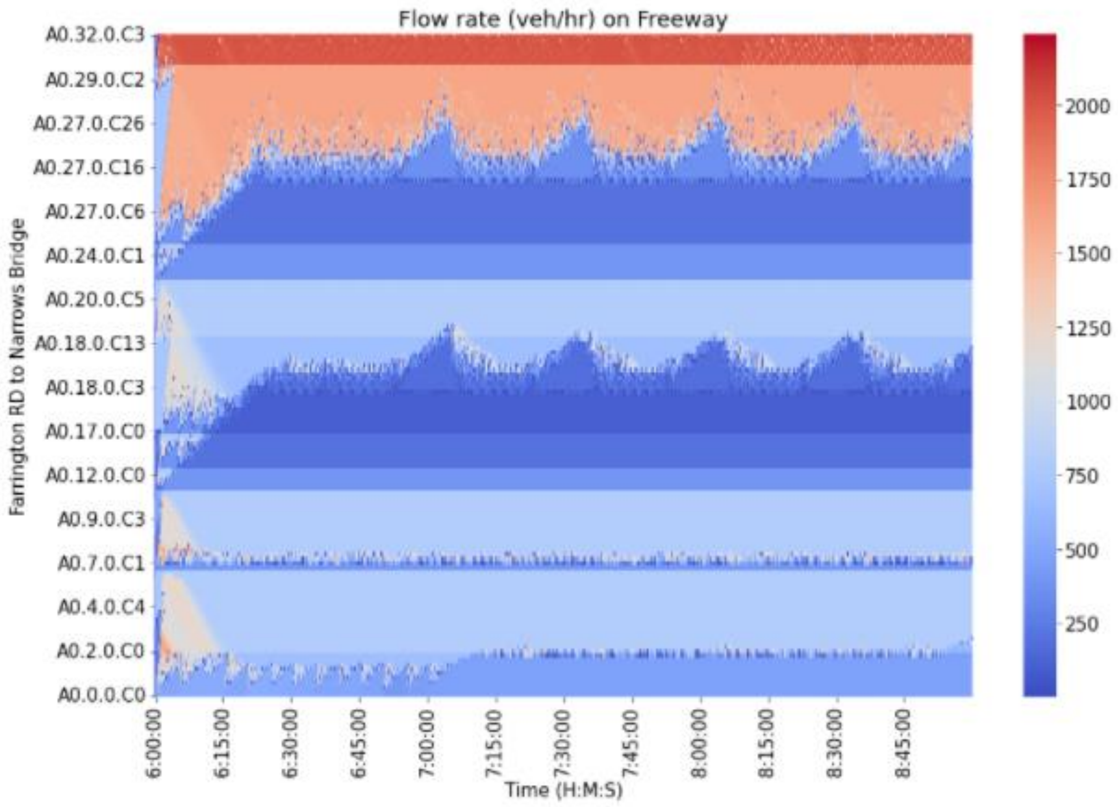
Case I(b): Heatmap plot of density and flow rate

Free speed 100

Traffic density (veh/km)

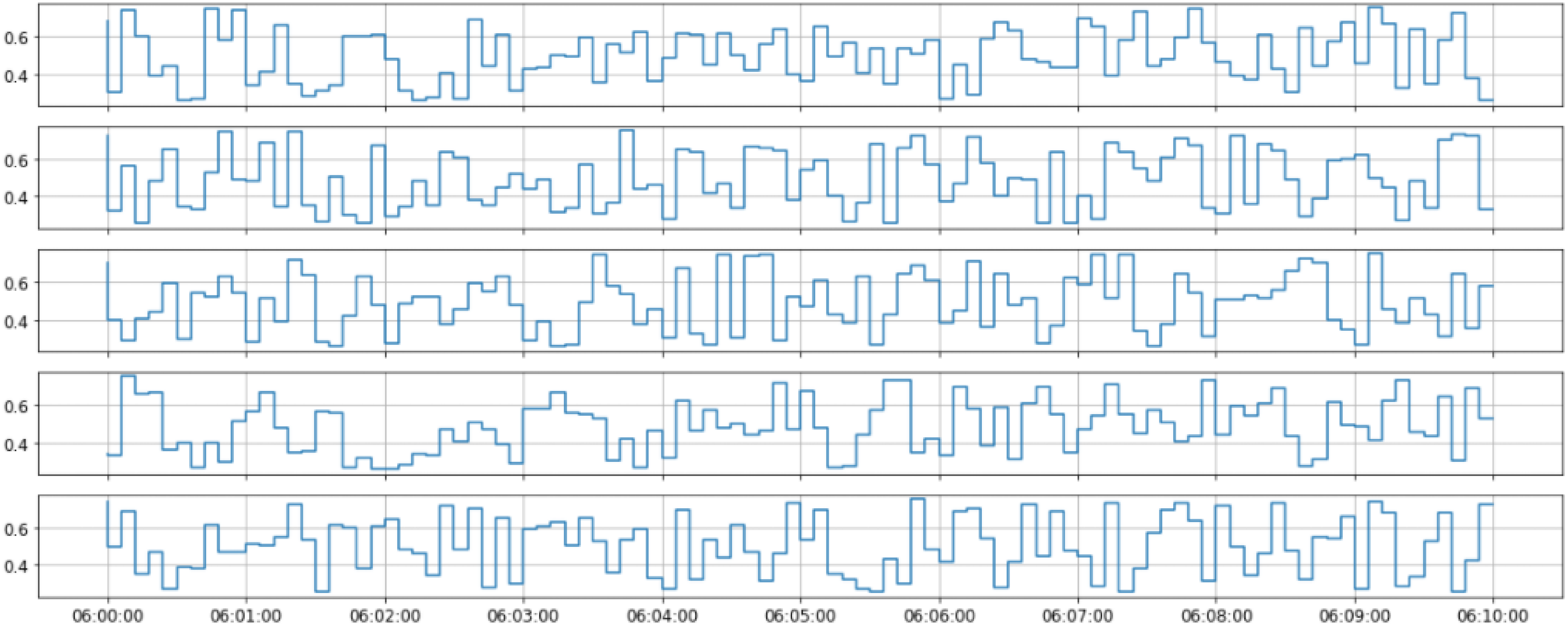


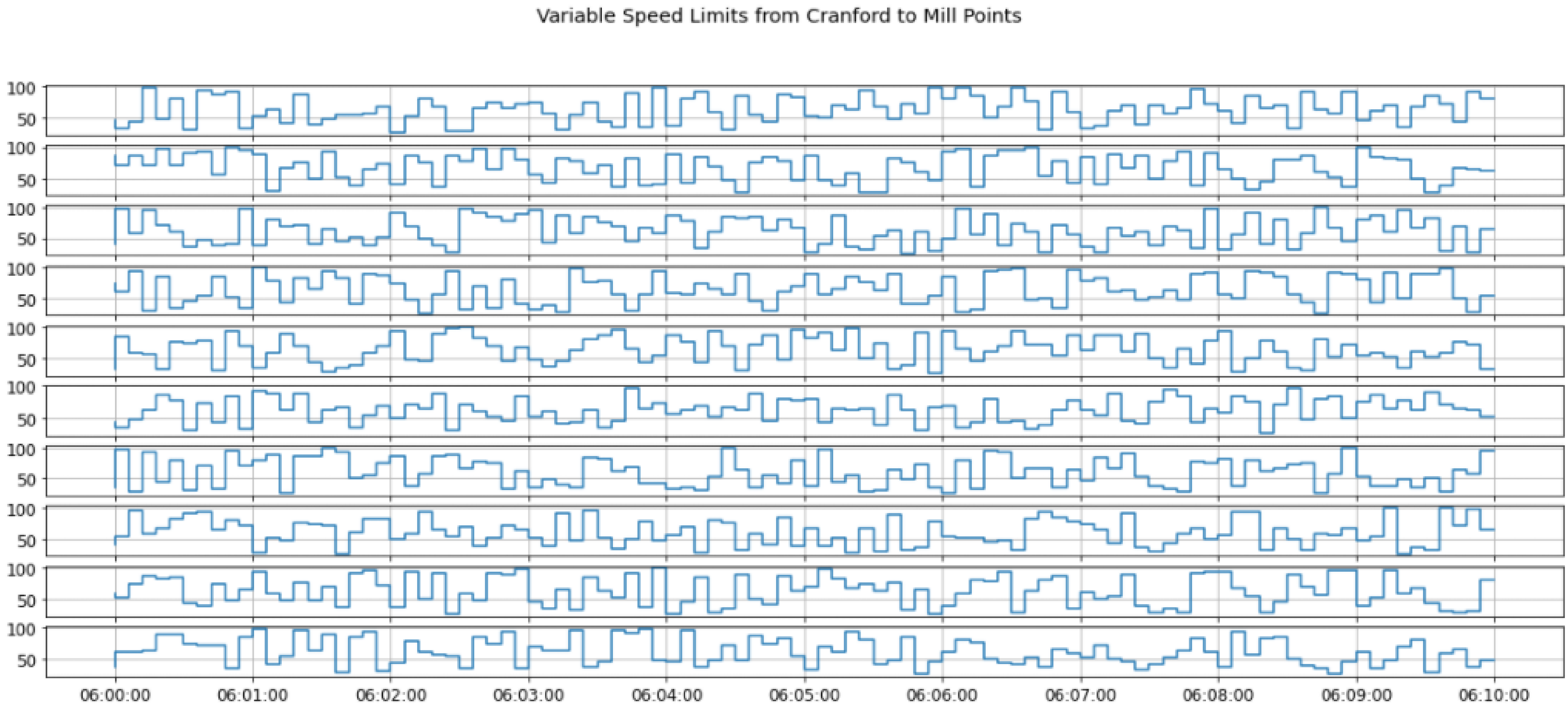
Traffic flow rate (veh/hr)



Case I(a): RM control and Variable speed limit

Ramp Metering at five on-ramps from Farrington to Cranford

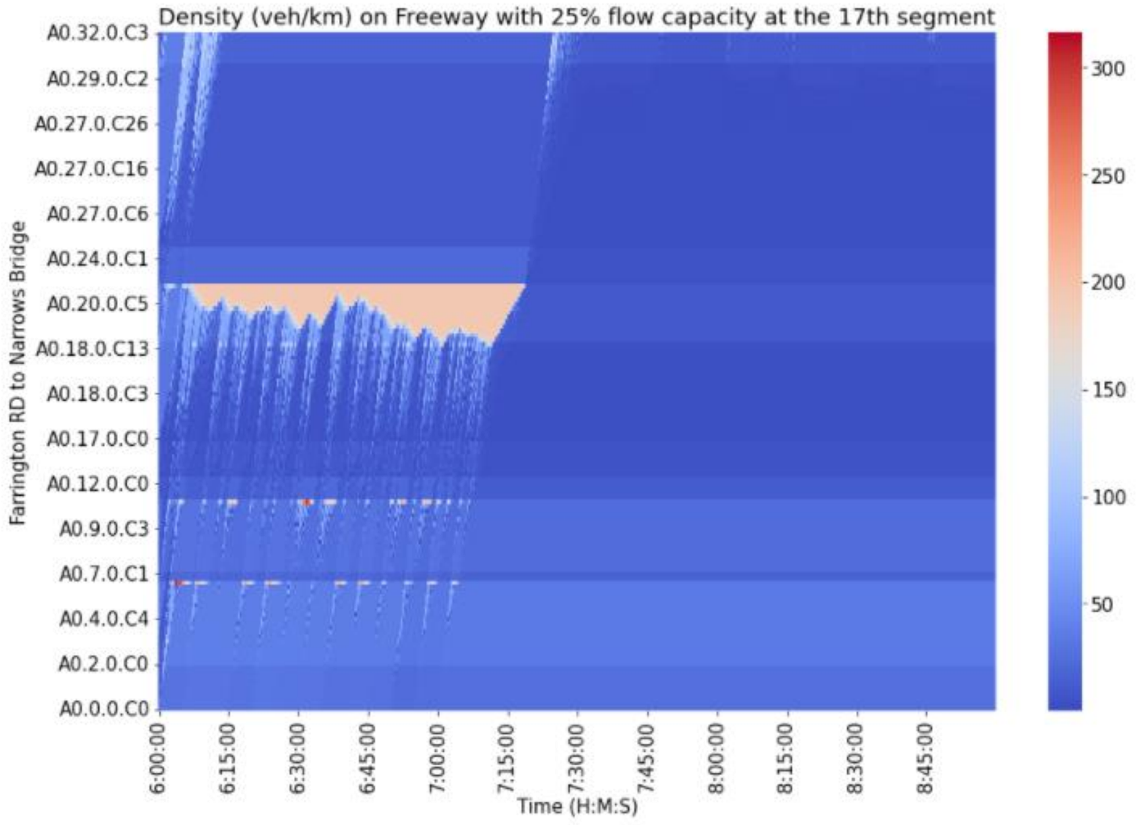



Variable Speed Limits from Cranford to Mill Points
100
50
06:00:00
06:01:00
06:02:00
06:03:00
06:04:00
06:05:00
06:06:00
06:07:00
06:08:00
06:09:00
06:10:00

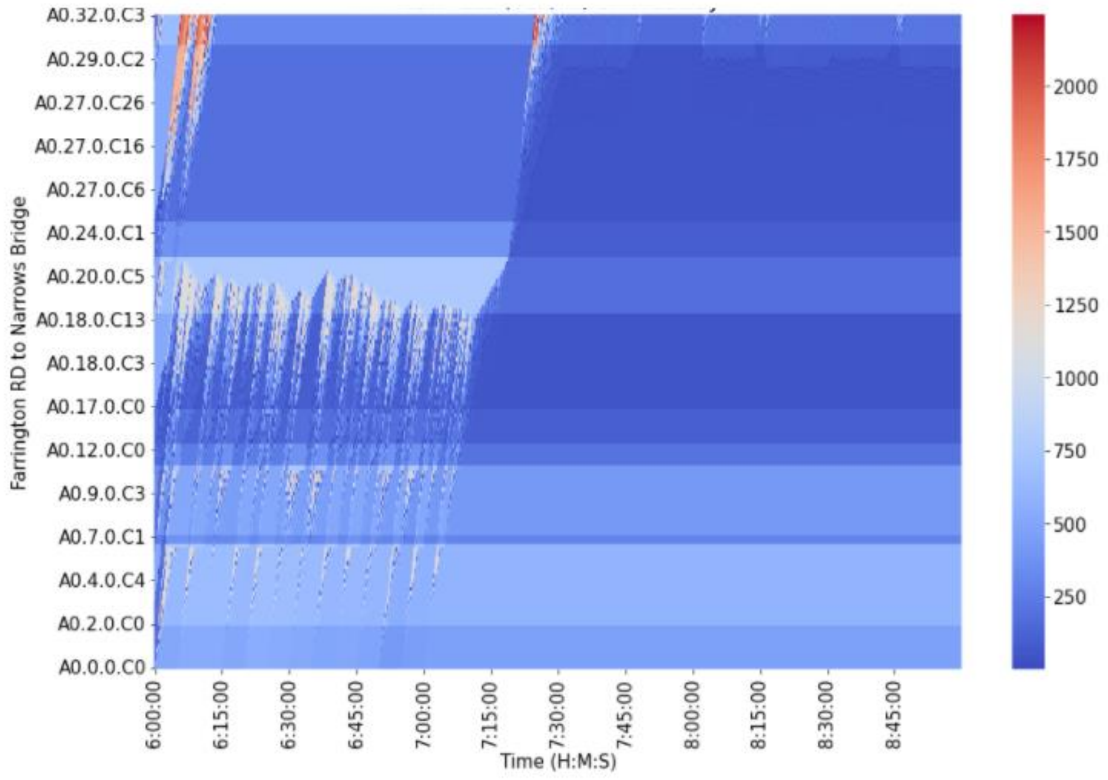
Case I (a): Heatmap plot of density and flow rate

Free speed 70

Traffic density (veh/km)

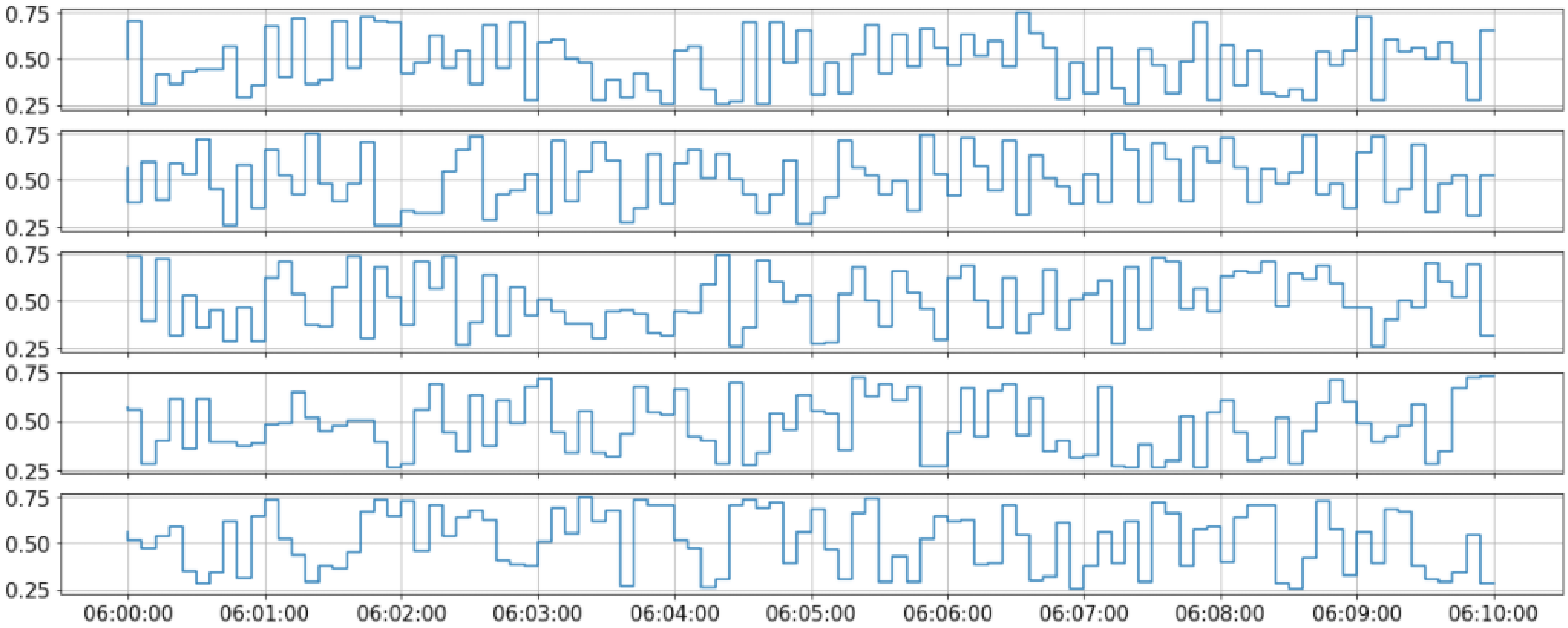


Traffic flow rate (veh/hr)

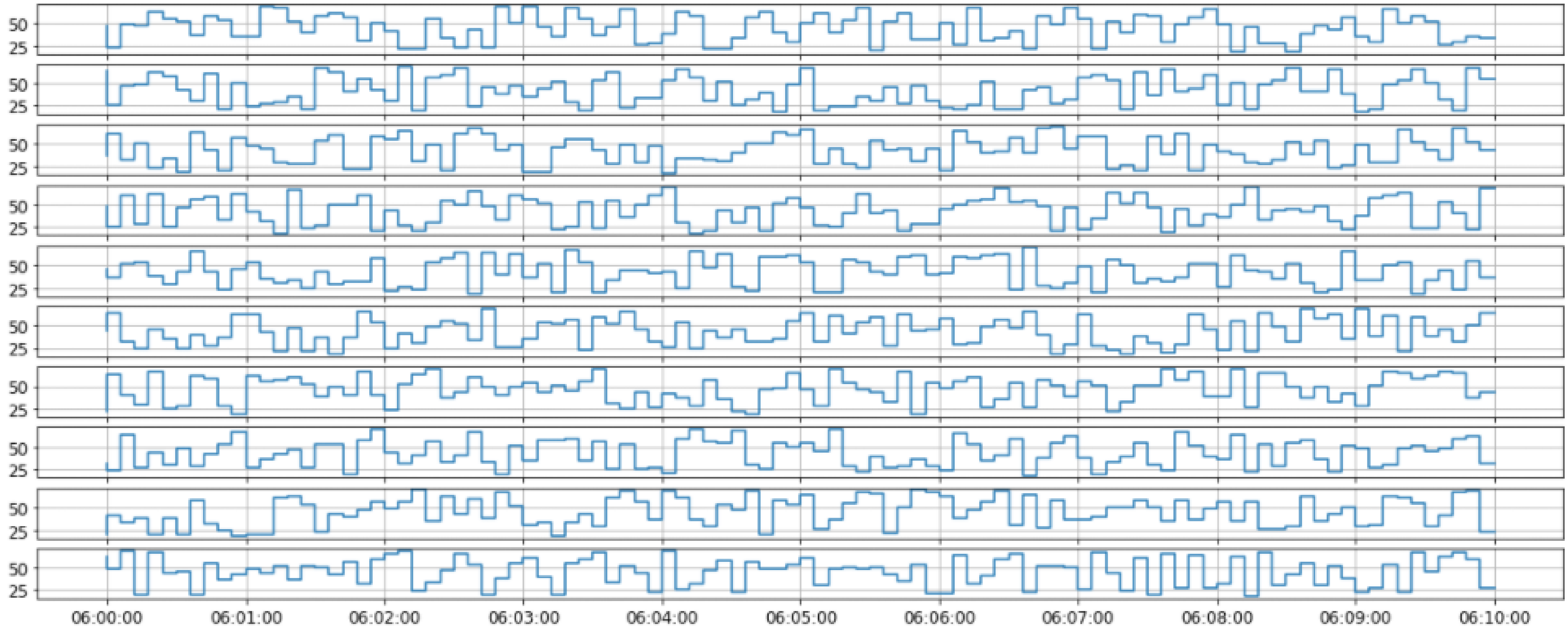


Case I (b): RM control and Variable speed limit

Ramp Metering at five on-ramps from Farrington to Cranford



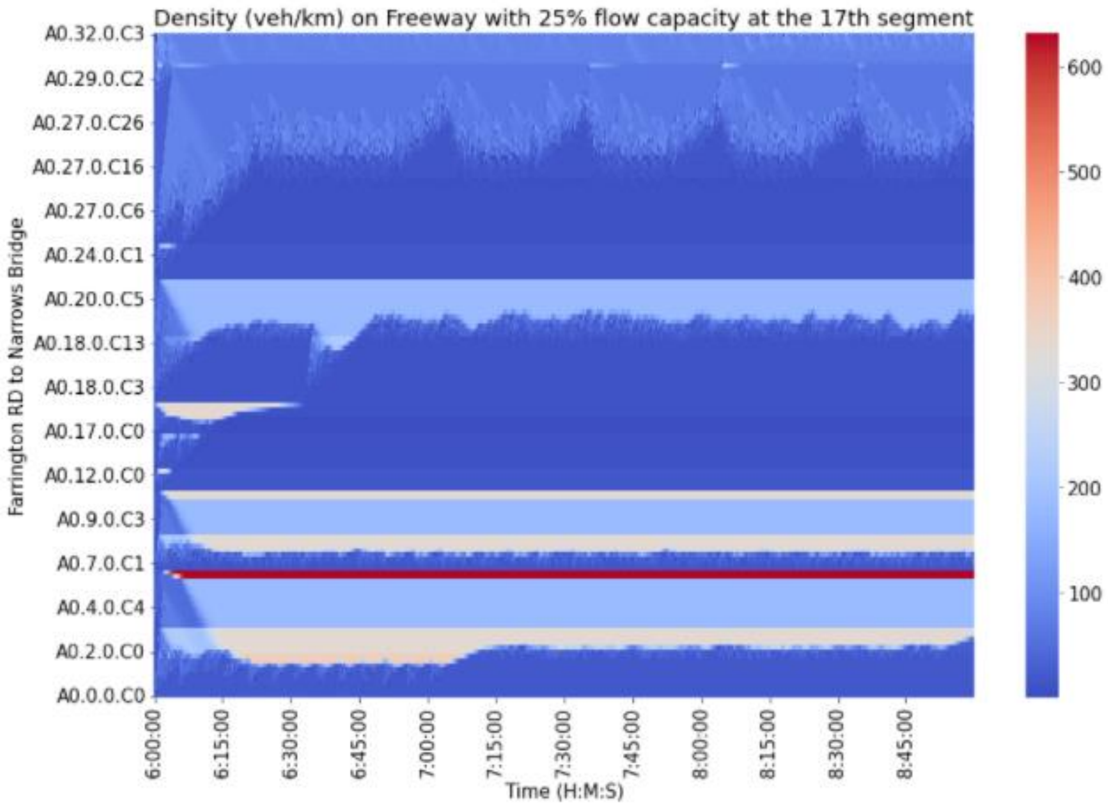
Variable Speed Limits from Cranford to Mill Points



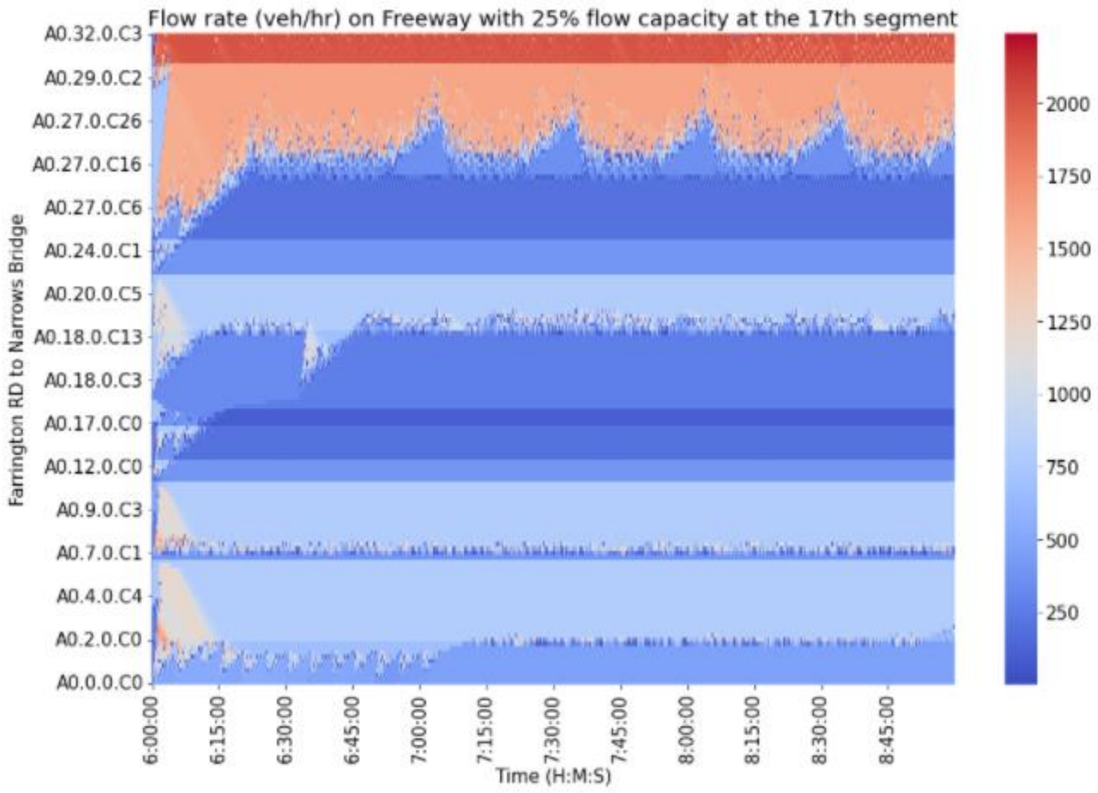
Case II(a): Heatmap plot of density and flow rate

Segment 17

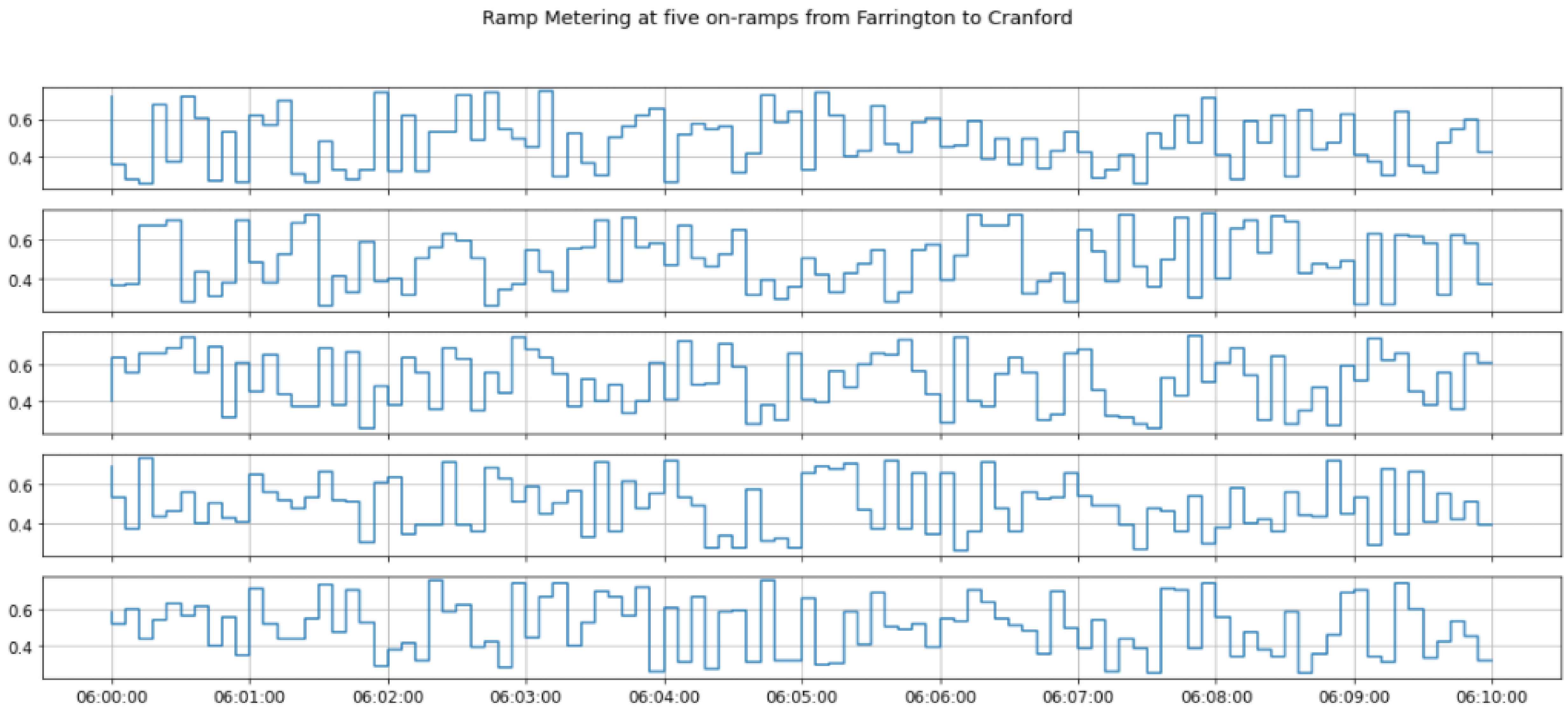
Traffic density (veh/km)



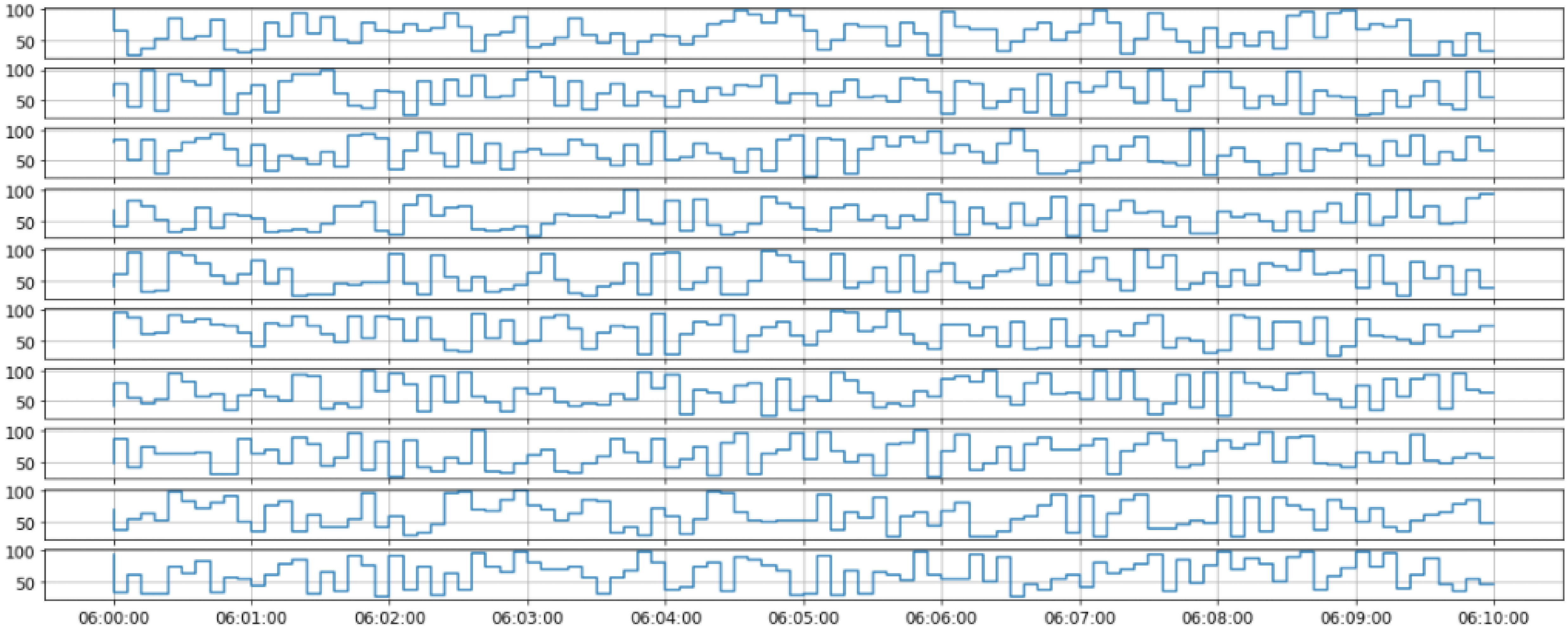
Traffic flow rate (veh/hr)



Case II(a): RM control and Variable speed limit


Ramp Metering at five on-ramps from Farrington to Cranford
0.6
0.4
06:00:00
06:01:00
06:02:00
06:03:00
06:04:00
06:05:00
06:06:00
06:07:00
06:08:00
06:09:00
06:10:00

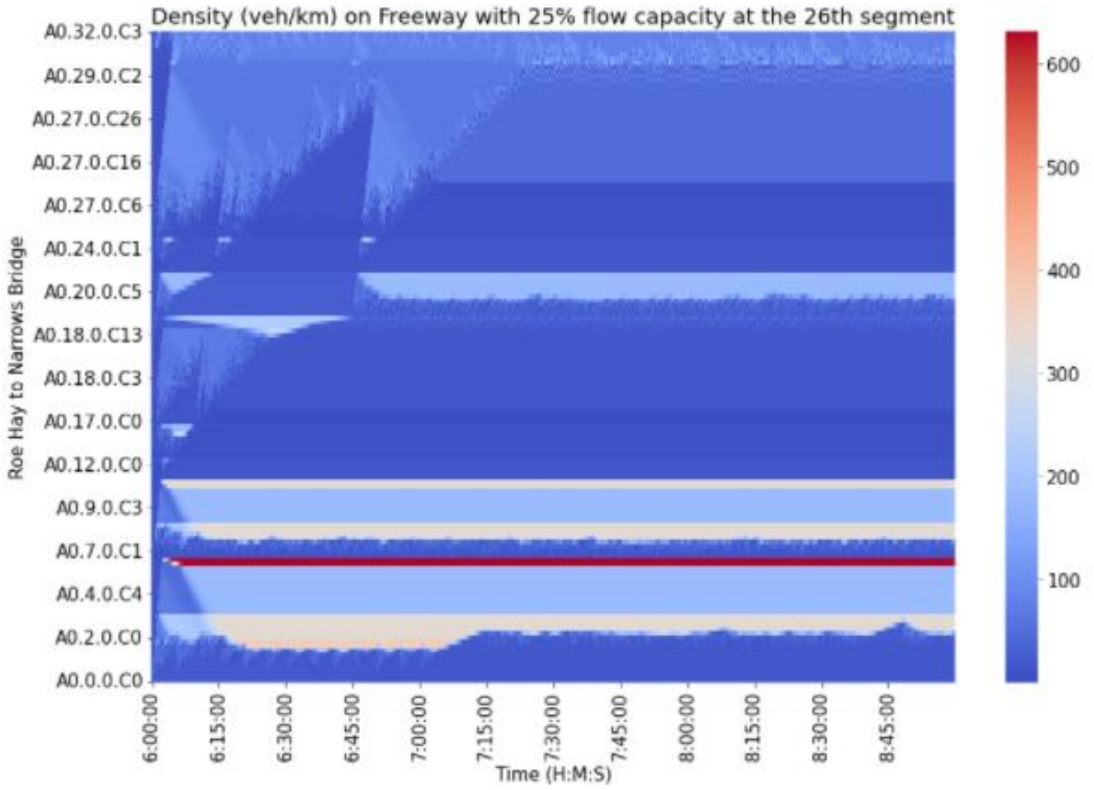
Variable Speed Limits from Cranford to Mill Points



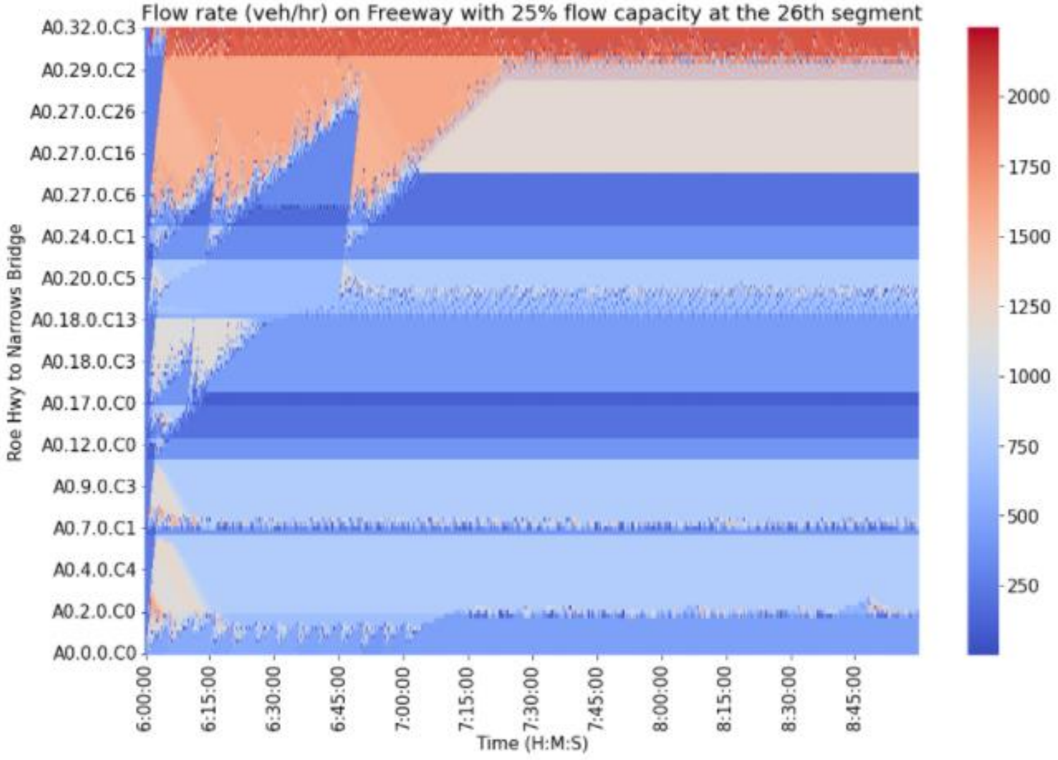
Case II(b): Heatmap plot of density and flow rate

Segment 26

Traffic density (veh/km)

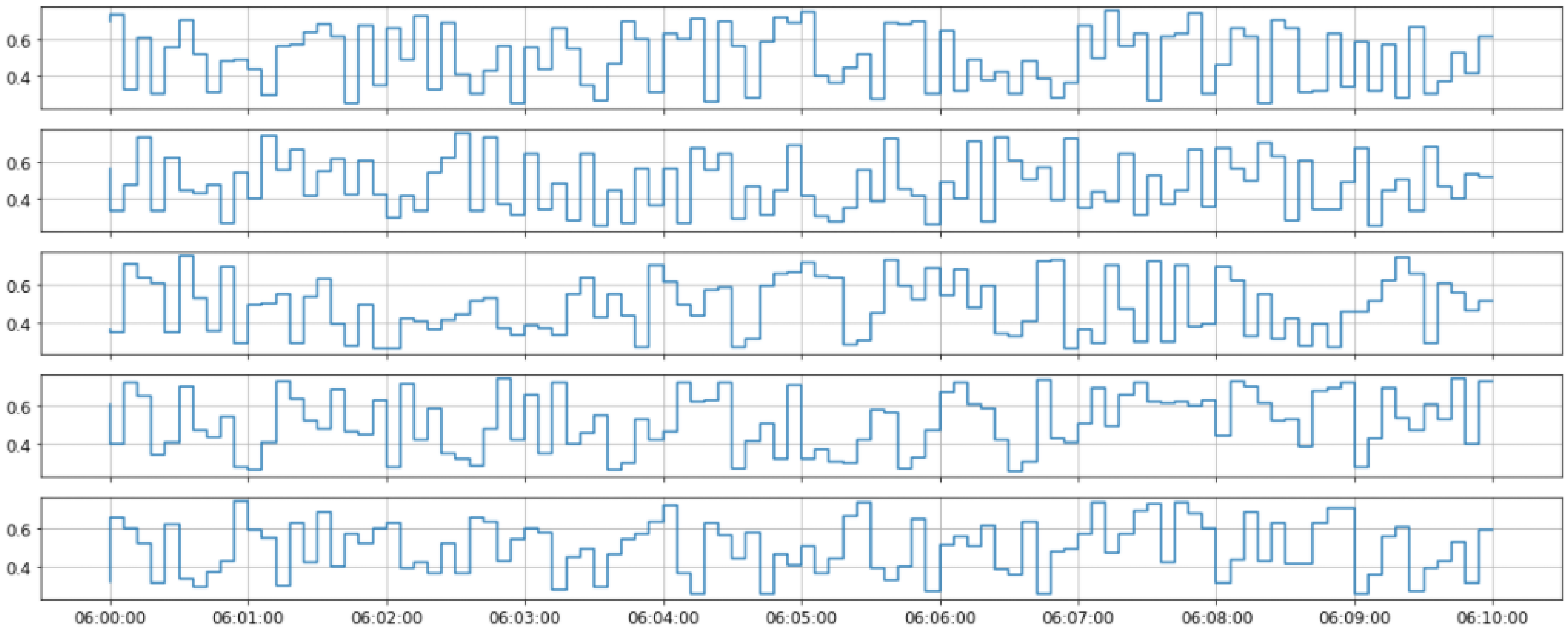


Traffic flow rate (veh/hr)

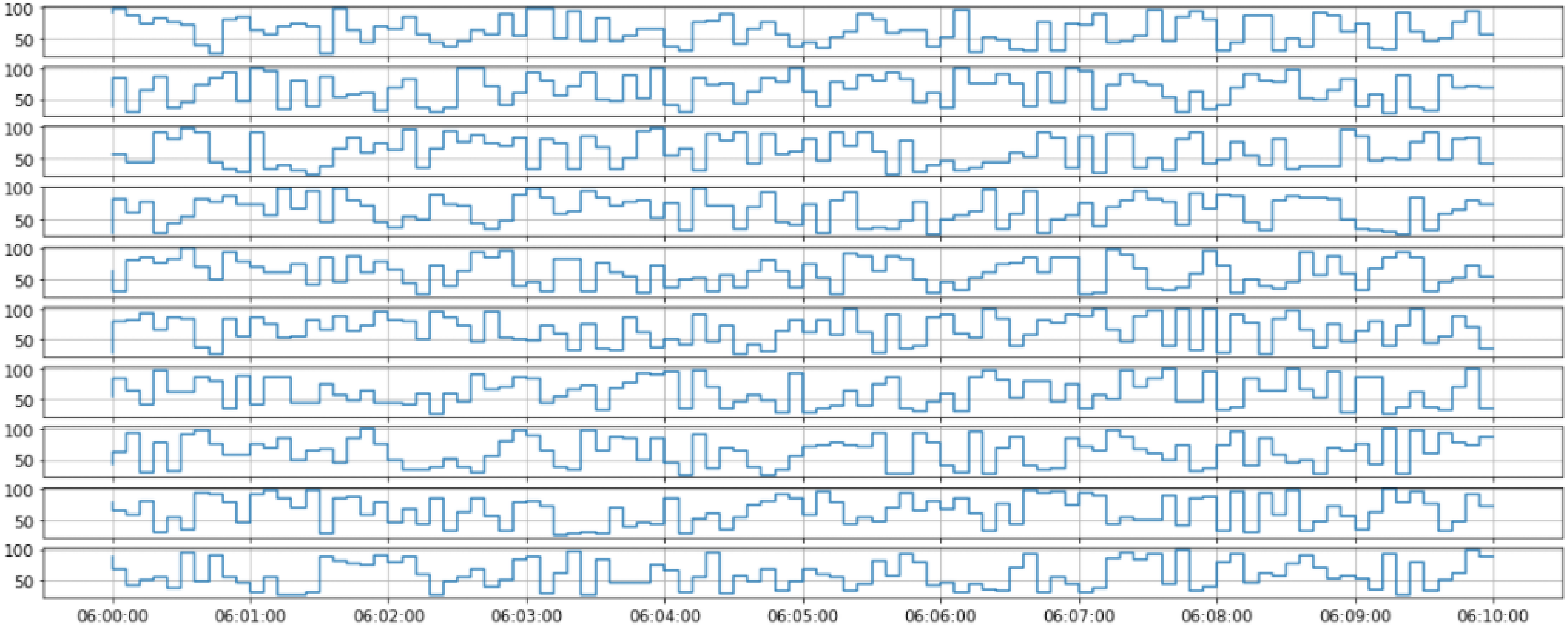


Case II(b): RM control and Variable speed limit

Ramp Metering at five on-ramps from Farrington to Cranford



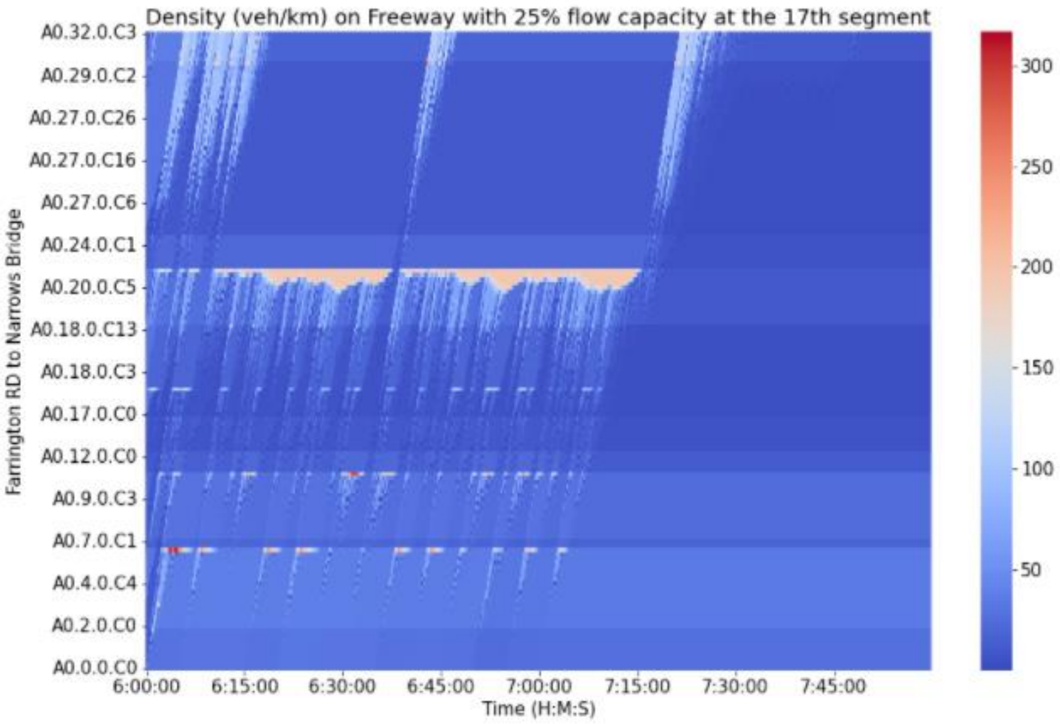
Variable Speed Limits from Cranford to Mill Points



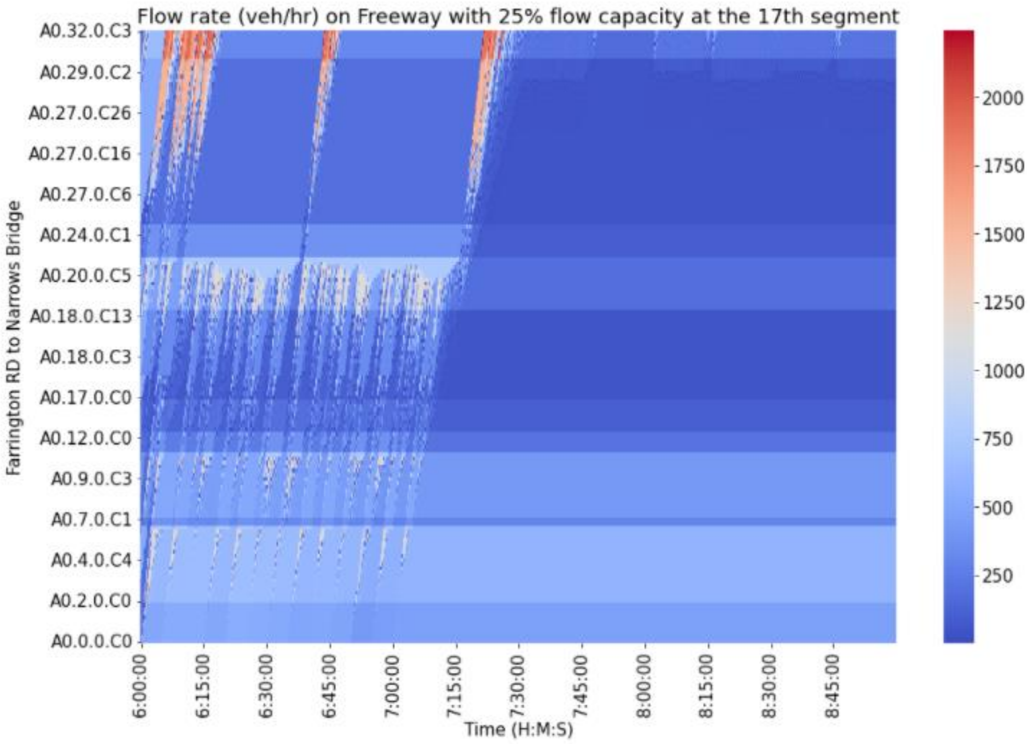
Case III(a): Heatmap plot of density and flow rate

Segment 17

Traffic density (veh/km)

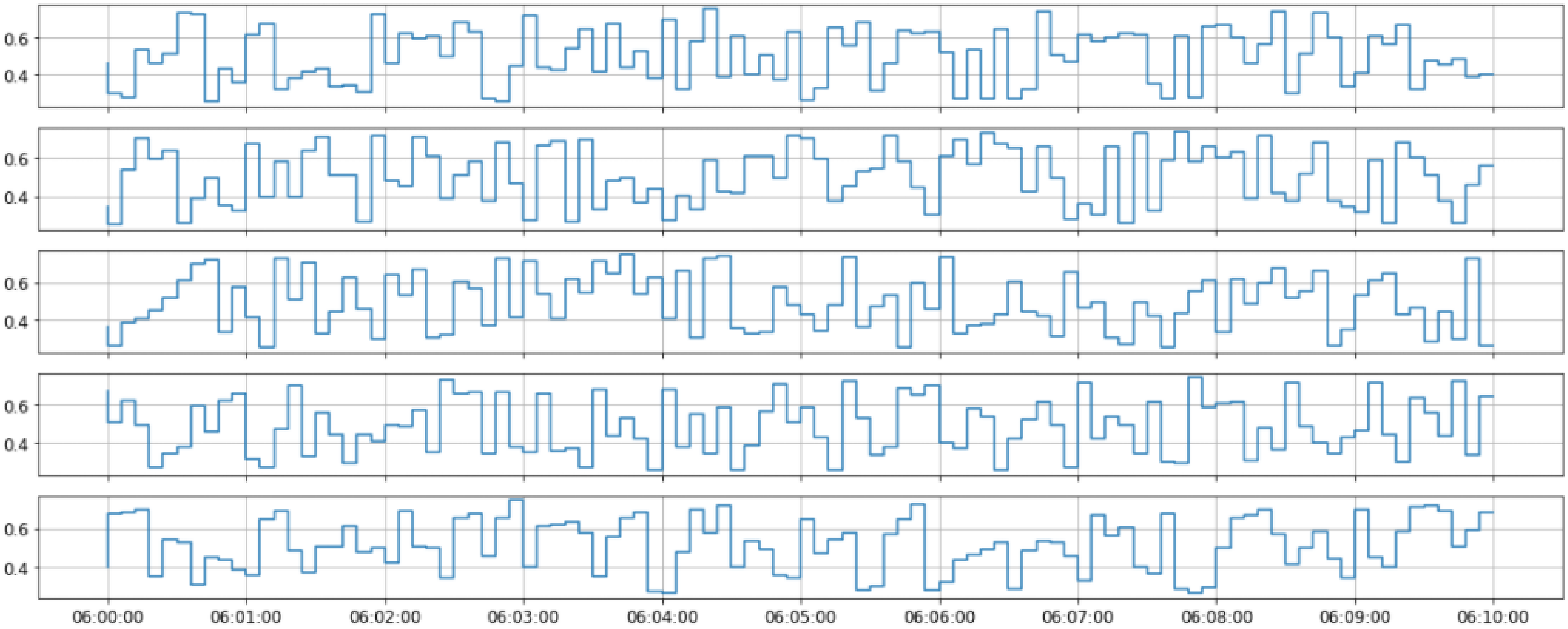


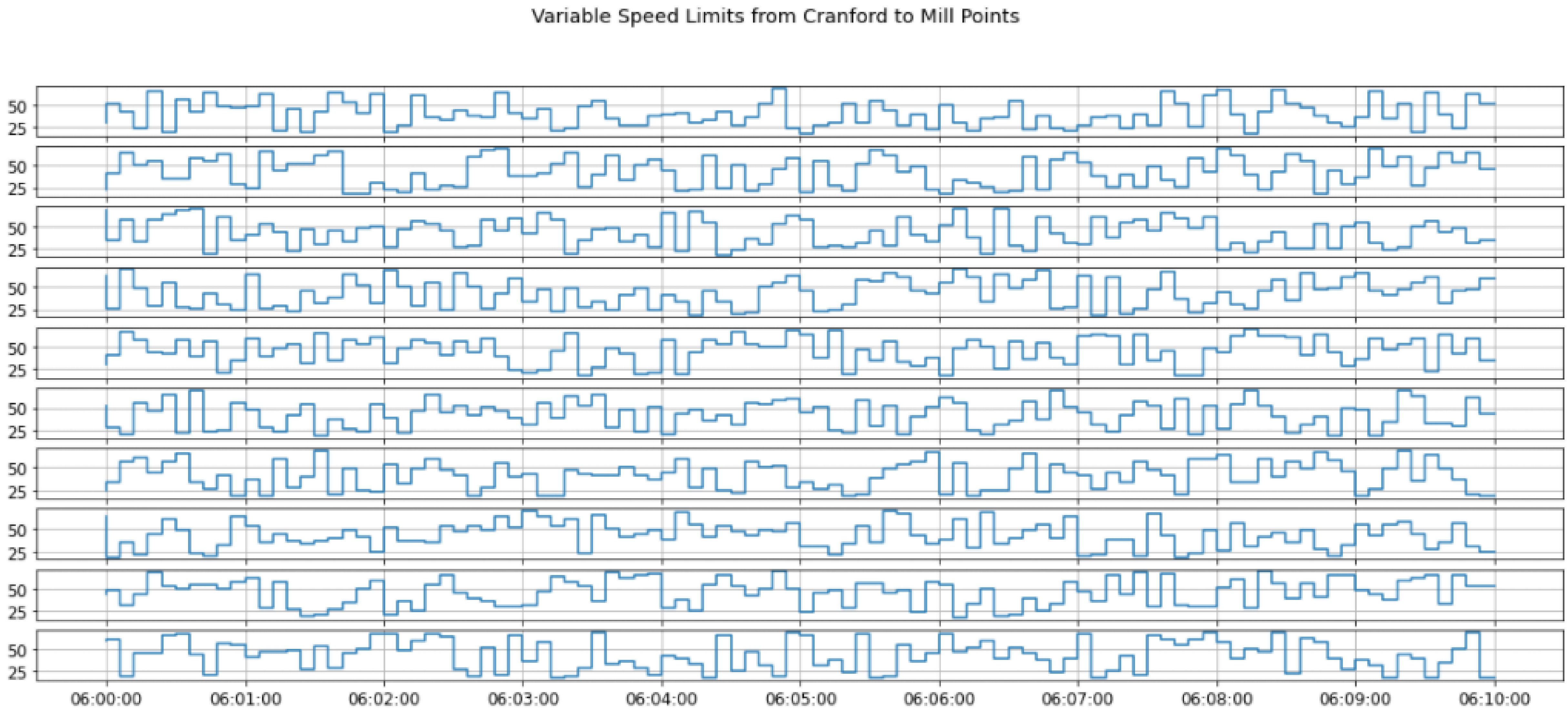
Traffic flow rate (veh/hr)



Case III(a): RM control and Variable speed limit

Ramp Metering at five on-ramps from Farrington to Cranford

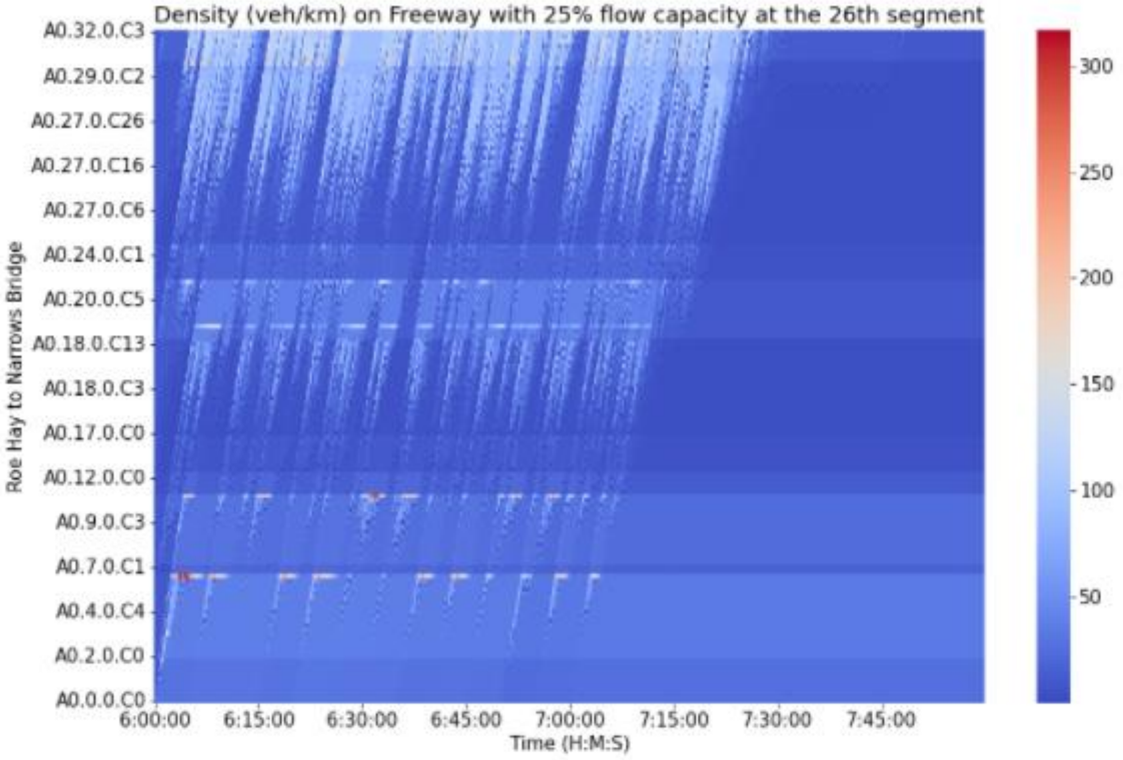



Variable Speed Limits from Cranford to Mill Points
50
25
06:00:00
06:01:00
06:02:00
06:03:00
06:04:00
06:05:00
06:06:00
06:07:00
06:08:00
06:09:00
06:10:00

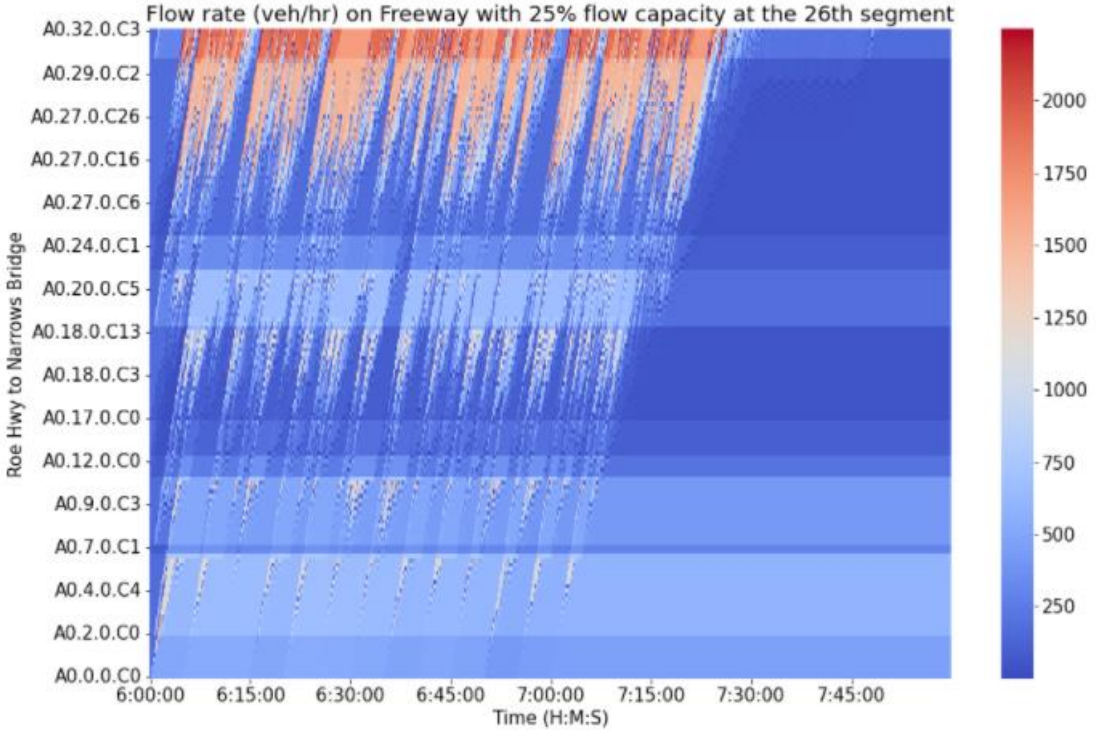
Case III(b): Heatmap plot of density and flow rate

Segment 26

Traffic density (veh/km)

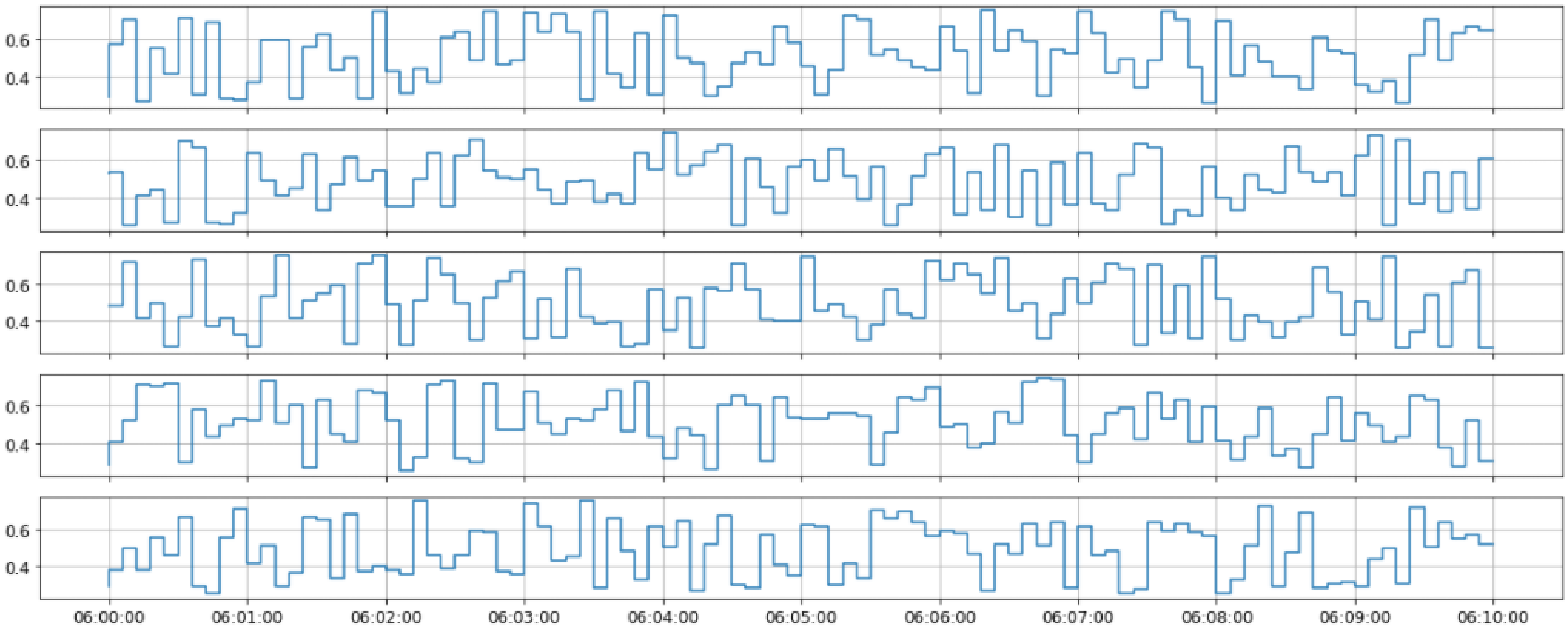


Traffic flow rate (veh/hr)

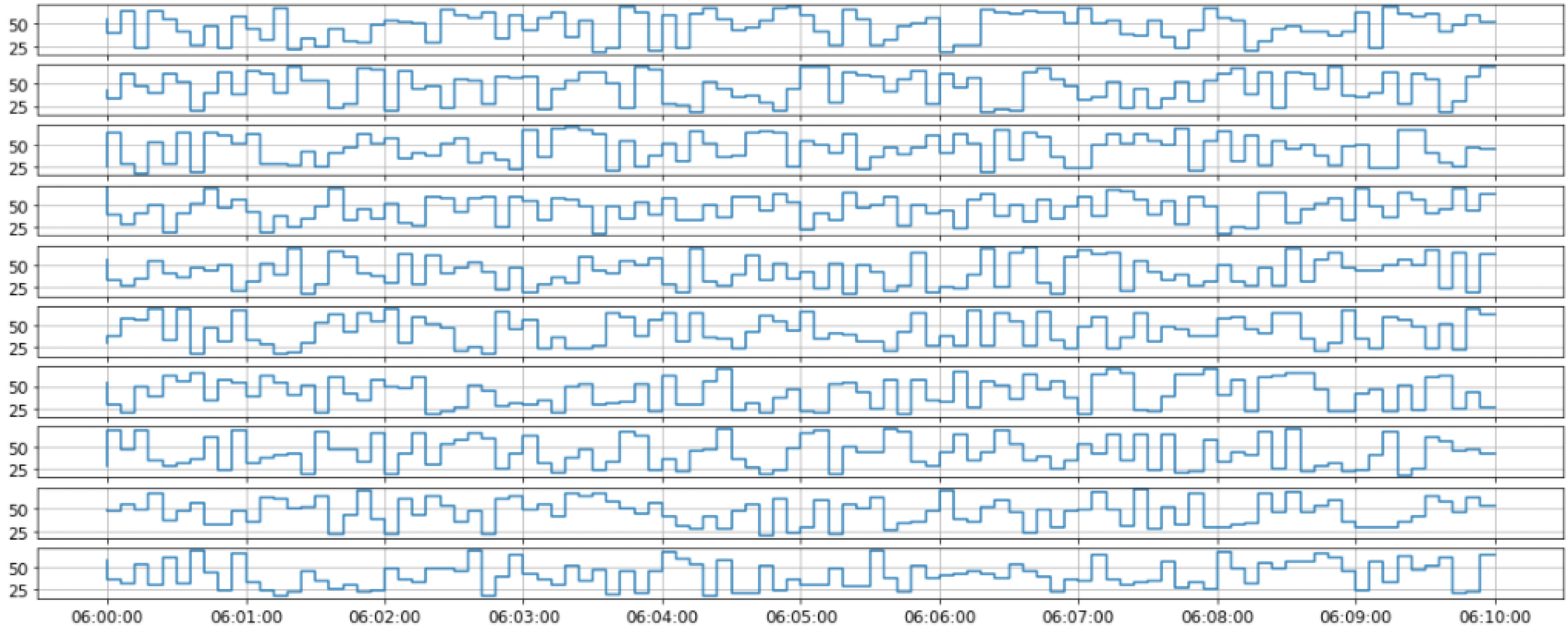


Case III(b): RM control and Variable speed limit

Ramp Metering at five on-ramps from Farrington to Cranford



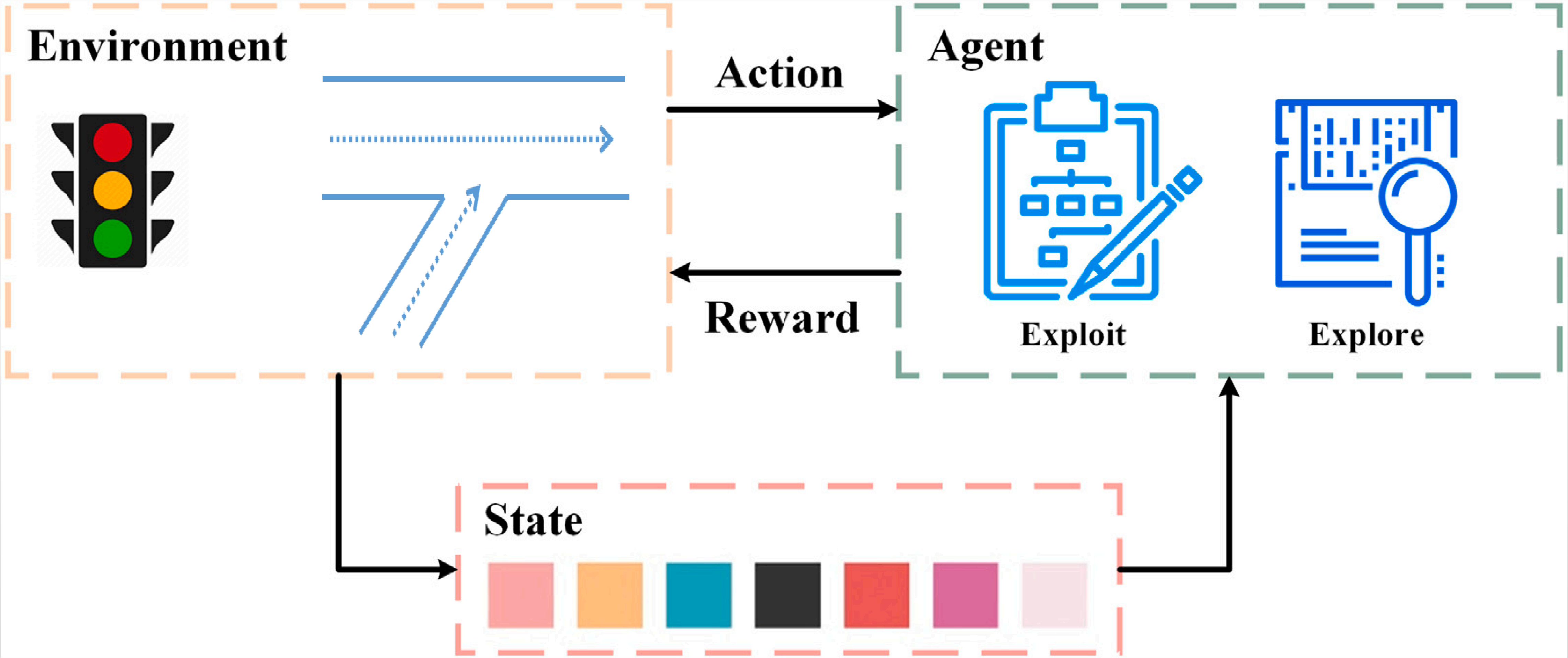
Variable Speed Limits from Cranford to Mill Points



2. Control of ramp metering based on reinforcement learning

PhD Student: C. Gu; Supervisors: YH Wu & B Wiwatanapataphee

Control of ramp metering based on reinforcement learning



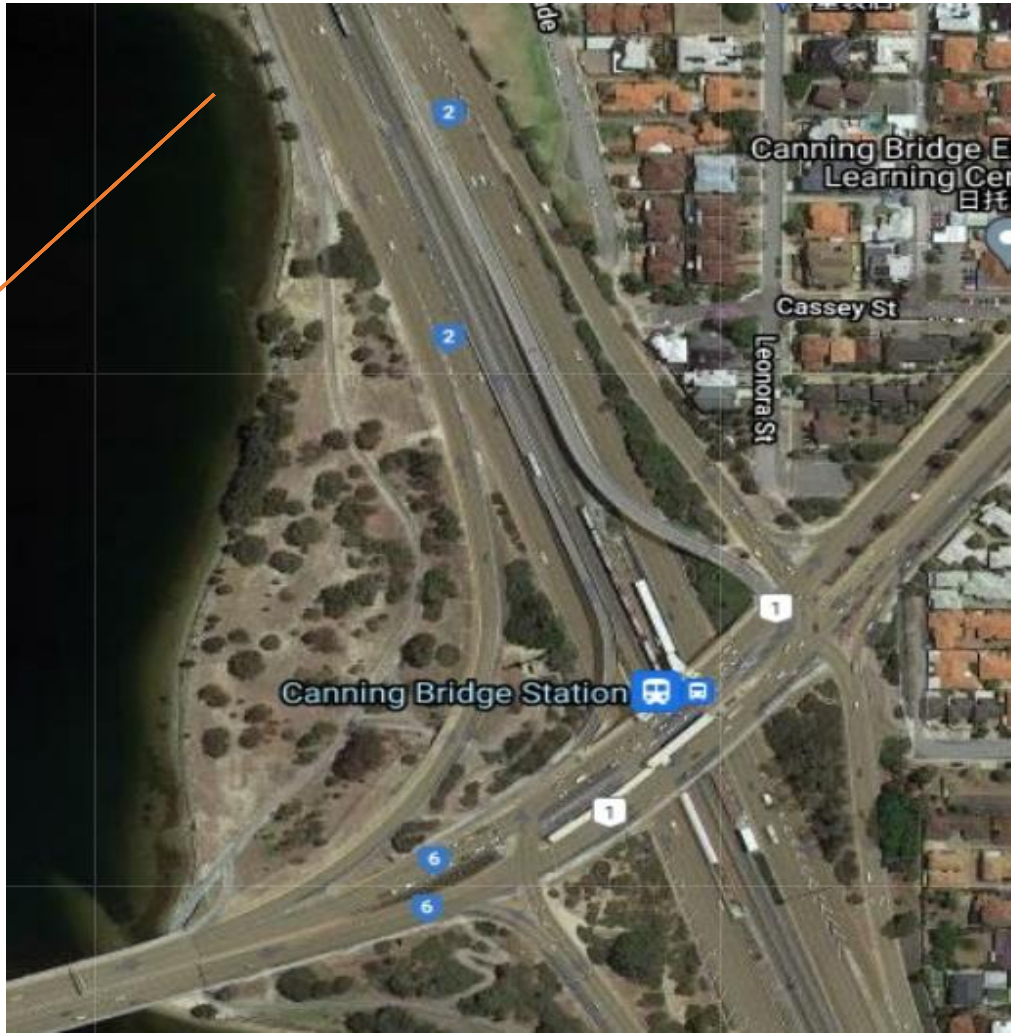
state: queue length, mean waiting time, mean speed....

action: switching signal phase, $a \in \{0,1\}$

policy: $\pi(a \mid s)$

$\pi(1 \mid s) = 0.8$

$\pi(0 \mid s) = 0.2$



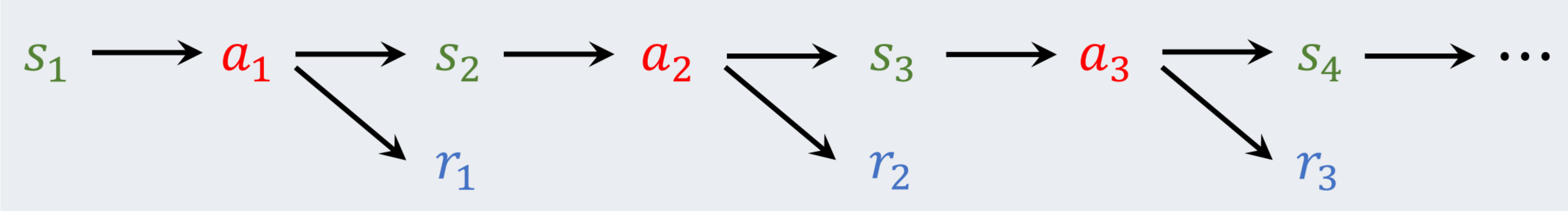
reward: mean waiting time, mean time, total time spent....

state transition: old state $\longrightarrow$ new state $S' \sim p(\cdot \mid s, a)$

- (state, action, reward) trajectory:

$$s_1, a_1, r_1, \quad s_2, a_2, r_2, \quad \cdots \quad , \quad s_n, a_n, r_n.$$

- One episode is from the the beginning to the end



Randomness in Returns

Definition: Discounted return (at time t).

- $U_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots + \gamma^{n-t} R_n$.

At time t, the rewards, $R_t, \cdots, R_n$, are random, so the return U_t is random.

- Reward R_i depends on S_i and A_i.
- States can be random: $S_i \sim p(\cdot \mid s_{i-1}, a_{i-1})$.
- Actions can be random: $A_i \sim \pi(\cdot \mid s_i)$.
- If either S_i or A_i is random, then R_i is random.

Action-Value Function $Q_\pi(s, a)$

Definition: Discounted return.

- $U_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots + \gamma^{n-t} R_n$.

Definition: Action-value function.

- $Q_\pi(s_t, a_t) = \mathbb{E}[\, U_t \mid S_t = s_t, A_t = a_t \,]$.

- $Q_\pi(s_t, a_t)$ depends on s_t, a_t, π, and p.
- $Q_\pi(s_t, a_t)$ is dependent of $S_{t+1}, \cdots, S_n$ and $A_{t+1}, \cdots, A_n$.

State-Value Function $V_\pi(s)$

Definition: Discounted return.

- $U_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots + \gamma^{n-t} R_n.$

Definition: Action-value function.

- $Q_\pi(s_t, a_t) = \mathbb{E}[U_t \mid S_t = s_t, A_t = a_t].$

Definition: State-value function.

- $V_\pi(s_t) = \mathbb{E}_A[Q_\pi(s_t, A)]$

Action-Value Functions $Q(s, a)$

Definition: Discounted return (aka cumulative discounted future reward).

- $U_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \gamma^3 R_{t+3} + \cdots$

Definition: Action-value function for policy π.

- $Q_\pi(s_t, a_t) = \mathbb{E}[U_t | S_t = s_t, A_t = a_t]$.

Definition: Optimal action-value function.

- $Q^\star(s_t, a_t) = \max_\pi Q_\pi(s_t, a_t)$.
- Whatever policy function π is used, the result of taking a_t at state s_t cannot be better than $Q^\star(s_t, a_t)$.

Approximate the Q Function

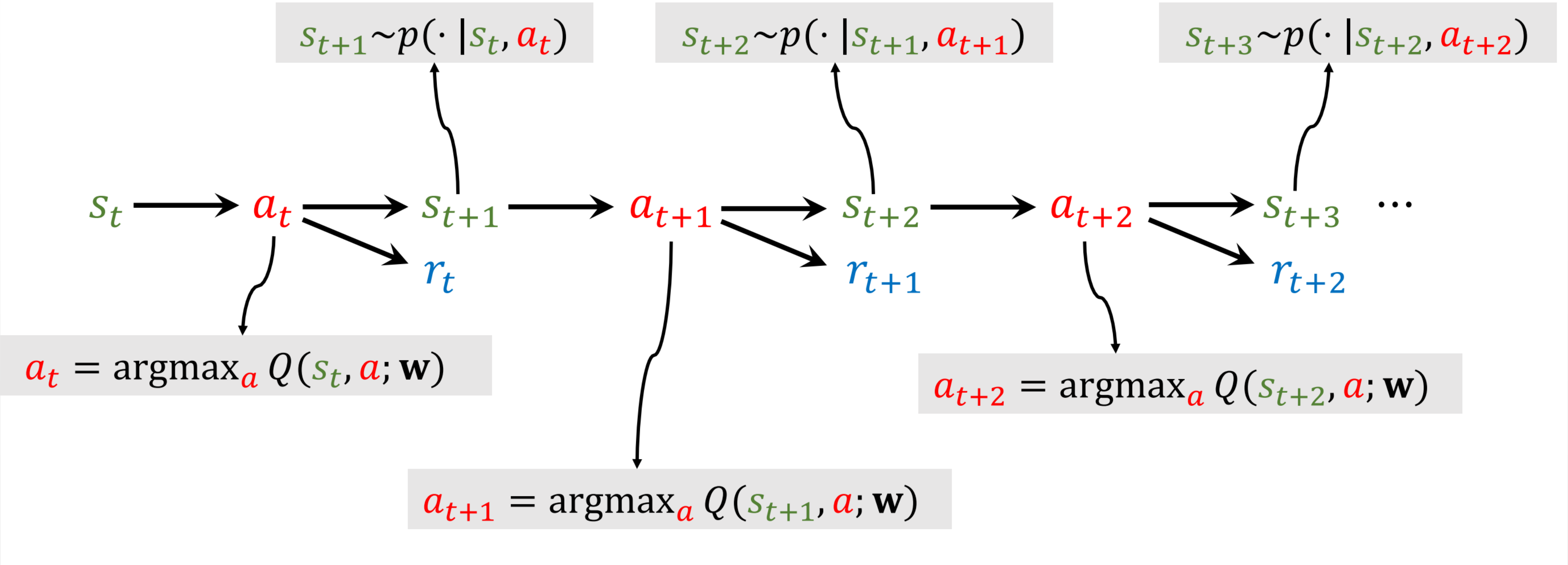
Goal: Win the game (≈ maximize the total reward.)

Question: If we know $Q^\star(s, a)$, what is the best action?

- Obviously, the best action is $a^\star = \underset{a}{\operatorname{argmax}}\, Q^\star(s, a)$.

Challenge: We do not know $Q^\star(s, a)$.

- Solution: Deep Q Network (DQN)
- Use neural network $Q(s, a; \mathbf{w})$ to approximate $Q^\star(s, a)$.



$s_{t+1} \sim p(\cdot \mid s_t, a_t)$
$s_{t+2} \sim p(\cdot \mid s_{t+1}, a_{t+1})$
$s_{t+3} \sim p(\cdot \mid s_{t+2}, a_{t+2})$
s_t
a_t
s_{t+1}
a_{t+1}
s_{t+2}
a_{t+2}
s_{t+3}
$\cdots$
r_t
r_{t+1}
r_{t+2}
$a_t = \text{argmax}_a \, Q(s_t, a; \mathbf{w})$
$a_{t+2} = \text{argmax}_a \, Q(s_{t+2}, a; \mathbf{w})$
$a_{t+1} = \text{argmax}_a \, Q(s_{t+1}, a; \mathbf{w})$

How to apply TD learning to DQN?

Identity: $U_t = R_t + \gamma \cdot U_{t+1}$.

TD learning for DQN:

- DQN's output, $Q(s_t, a_t; \mathbf{w})$, is an estimate of U_t.
- DQN's output, $Q(s_{t+1}, a_{t+1}; \mathbf{w})$, is an estimate of U_{t+1}.
- Thus, $\underbrace{Q(s_t, a_t; \mathbf{w})}_{\text{Prediction}} \approx \underbrace{r_t + \gamma \cdot Q(s_{t+1}, a_{t+1}; \mathbf{w})}_{\text{TD target}}$.

Train DQN using TD learning

- Prediction: $Q(s_t, a_t; \mathbf{w}_t)$.
- TD target:

$$y_t = r_t + \gamma \cdot Q(s_{t+1}, a_{t+1}; \mathbf{w}_t)$$

$$= r_t + \gamma \cdot \max_a Q(s_{t+1}, a; \mathbf{w}_t).$$

- Loss: $L_t = \frac{1}{2}[Q(s_t, a_t; \mathbf{w}) - y_t]^2$.
- Gradient descent: $\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha \cdot \frac{\partial L_t}{\partial \mathbf{w}} \Big|_{\mathbf{w}=\mathbf{w}_t}$.

Temporal Difference (TD) Learning

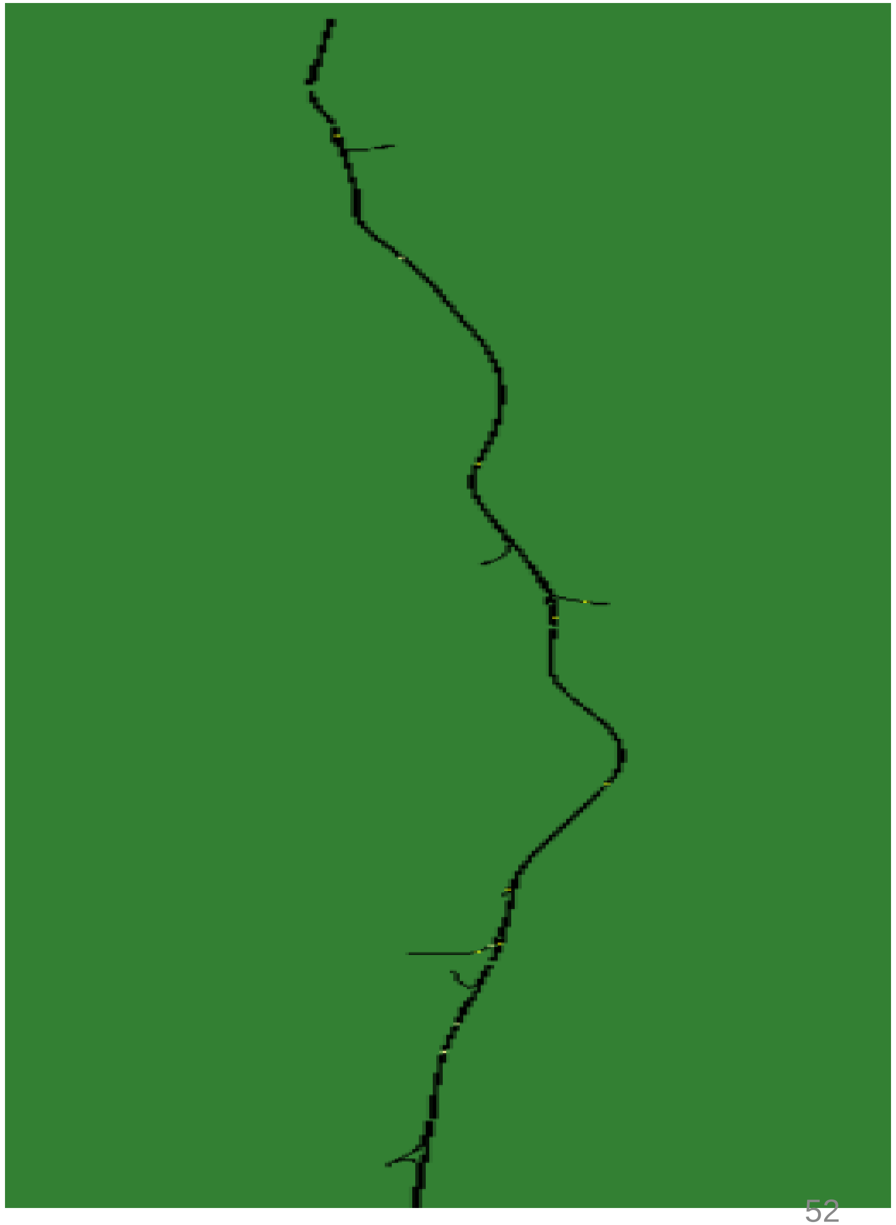
Algorithm: One iteration of TD learning.

1. Observe state $S_t = s_t$ and perform action $A_t = a_t$.
2. Predict the value: $q_t = Q(s_t, a_t; \mathbf{w}_t)$.
3. Differentiate the value network: $\mathbf{d}_t = \frac{\partial\, Q(s_t, a_t; \mathbf{w})}{\partial\, \mathbf{w}} \Big|_{\mathbf{w}=\mathbf{w}_t}$.
4. Environment provides new state s_{t+1} and reward r_t.
5. Compute TD target: $y_t = r_t + \gamma \cdot \max_a Q(s_{t+1}, a; \mathbf{w}_t)$.
6. Gradient descent: $\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha \cdot (q_t - y_t) \cdot \mathbf{d}_t$.

Test: $\min \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{i=1}^{I} w_{ijt}$

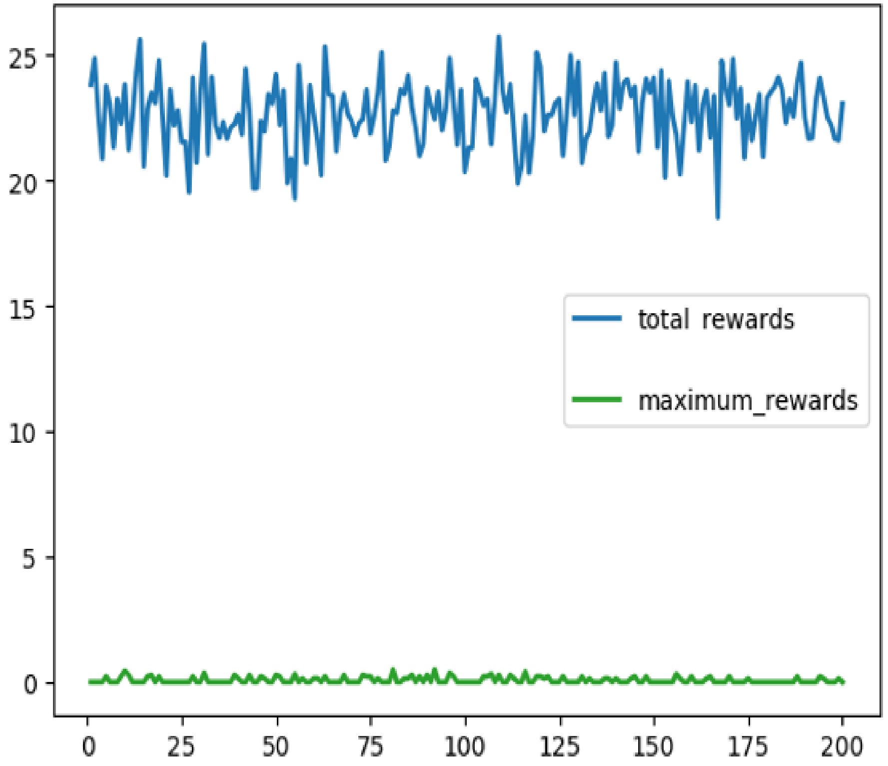
w_{ijt} : the waiting time of vehicle i on ramp j for time t

$T = 7200, \quad J = 7$



Result:


25
20
15
10
5
0
0
20
40
60
80
100
total rewards
maximum_rewards


25
20
15
10
5
0
0
25
50
75
100
125
150
175
200
total rewards
maximum_rewards

ACKNOWLEDGEMENTS

The project is funded by the Australian Research Council, Main Roads WA, and Roads and Maritime Services NSW.

Acknowledgement is made to the Australian Research Council, Main Roads WA, Roads and Maritime Services NSW, and the Australia's Sustainable Built Environment National Research Centre (SBEnrc) for their support.


THANK YOU
For Your Attention