Authors
Zhang, Michael, PhD

Gao, Hang, PhD

Chen, Di

et al.

Publication Date
2024-06-01

DOI
10.7922/G2154FC5

RESEARCH REPORT

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

A Data-Driven Approach to Manage High-Occupancy Toll Lanes in California

Hang Gao, Ph.D., Transportation Technology and Policy, University of California, Davis
Di Chen, Ph.D. Candidate, Department of Civil Engineering, University of California, Davis
Yanlin Qi, Ph.D. Candidate, Transportation Technology and Policy, University of California, Davis
Michael Zhang, Ph.D., Professor, Department of Civil Engineering, University of California, Davis

June 2024

ITS UCDAVIS Institute of Transportation Studies

Report No.: UC-ITS-2023-06 | DOI: 10.7922/G2154FC5

Technical Report Documentation Page

<table>
<tr><td>1. Report No.
UC-ITS-2023-06</td><td>2. Government Accession No.
N/A</td><td colspan="2">3. Recipient's Catalog No.
N/A</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
A Data-Driven Approach to Manage High-Occupancy Toll Lanes in California</td><td colspan="2">5. Report Date
June 2024</td></tr>
<tr><td colspan="2">6. Performing Organization Code
ITS-Davis</td></tr>
<tr><td colspan="2">7. Author(s)
Michael Zhang, Ph.D. https://orcid.org/0000-0002-4647-3888
Hang Gao, Ph.D. https://orcid.org/0000-0002-8549-0170
Di Chen, Ph.D. Candidate
Yanlin Qi, Ph.D. Candidate</td><td colspan="2">8. Performing Organization Report No.
UCD-ITS-RR-24-33</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
Institute of Transportation Studies, Davis
1605 Tilia Street
Davis, CA 95616</td><td colspan="2">10. Work Unit No.
N/A</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
UC-ITS-2023-06</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
The University of California Institute of Transportation Studies
www.ucits.org</td><td colspan="2">13. Type of Report and Period Covered
Final Report (October 2022 – December 2023)</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code
UC ITS</td></tr>
<tr><td colspan="4">15. Supplementary Notes
DOI:10.7922/G2154FC5</td></tr>
<tr><td colspan="4">16. Abstract
Managing traffic flow in high-occupancy toll (HOT) lanes is a tough balancing act and current tolling schemes often lead to either under- or over-utilization of HOT lane capacity. The inherent linear/nonlinear relationship between flow and tolls in HOT lanes suggest that recent advances in machine learning and the use of a data-driven model may help set toll rates for optimal flow and lane use. In this research project, a data-driven model was developed, using long short-term memory (LSTM) neural networks to capture the underlying flow-toll pattern on both HOT and general-purpose lanes. Then, a dynamic control strategy, using linear quadratic regulator (LQR) feedback controller was implemented to fully utilize the HOT lane capacity while maintaining congestion-free conditions. A case study of the I-580 freeway in Alameda County, California was carried out. The control system was evaluated in terms of vehicle hours traveled and person hours traveled for solo drivers and carpoolers. Results show that the tolling strategy helps to mitigate congestion in HOT and general-purpose lanes, benefiting every traveler on I-580.</td></tr>
<tr><td colspan="2">17. Key Words
High occupancy toll lanes, traffic flow, traffic models, highway traffic control systems</td><td colspan="2">18. Distribution Statement
No restrictions.</td></tr>
<tr><td>19. Security Classification (of this report)
Unclassified</td><td>20. Security Classification (of this page)
Unclassified</td><td>21. No. of Pages
41</td><td>22. Price
N/A</td></tr>
</table>

Form Dot F 1700.7 (8-72) Reproduction of completed page authorized

About the UC Institute of Transportation Studies

The University of California Institute of Transportation Studies (UC ITS) is a network of faculty, research and administrative staff, and students dedicated to advancing the state of the art in transportation engineering, planning, and policy for the people of California. Established by the Legislature in 1947, ITS has branches at UC Berkeley, UC Davis, UC Irvine, and UCLA.

Acknowledgments

This study was made possible with funding received by the University of California Institute of Transportation Studies from the State of California through the Road Repair and Accountability Act of 2017 (Senate Bill 1). The authors would like to thank the State of California for its support of university-based research, and especially for the funding received for this project. The authors would also like to thank Elizabeth Rutman and Ashly Tam, Express Lanes Implementation and Operations of the Alameda County Transportation Commission for providing the research team the 680 and 580 HOT lane data, and Seth Karten for his careful editing of the report.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the State of California in the interest of information exchange. The State of California assumes no liability for the contents or use thereof. Nor does the content necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

A Data-Driven Approach to Manage High-Occupancy Toll Lanes in California

Hang Gao, Ph.D., Transportation Technology and Policy, University of California, Davis
Di Chen, Ph.D. Candidate, Department of Civil Engineering, University of California, Davis
Yanlin Qi, Ph.D. Candidate, Transportation Technology and Policy, University of California, Davis
Michael Zhang, Ph.D., Professor, Department of Civil Engineering, University of California, Davis

June 2024

UCDAVIS Institute of Transportation Studies

Report No.: UC-ITS-2023-06 | DOI: 10.7922/G2154FC5

Table of Contents

Table of Contents

List of Figures

List of Tables

Acronyms

GP	general-purpose [refers to a type of lane on a road]
HOT	high-occupancy toll [refers to a type of lane on a road]
HOV	high-occupancy vehicle
LQR	linear quadratic regulator
LSTM	long short-term memory
PHT	person hours traveled
RNN	recurrent neural network
SOV	single-occupancy vehicle
VHT	vehicle hours traveled
VMT	vehicle miles traveled

Executive Summary

Executive Summary

The study investigates traffic management and tolling strategies on the I-580 high-occupancy toll (HOT) ("express") lanes in Alameda County, California (Figure ES-1). Two key methods are employed: a long short-term memory (LSTM)-based traffic-speed prediction model and a linear quadratic regulator (LQR)-based tolling control system.

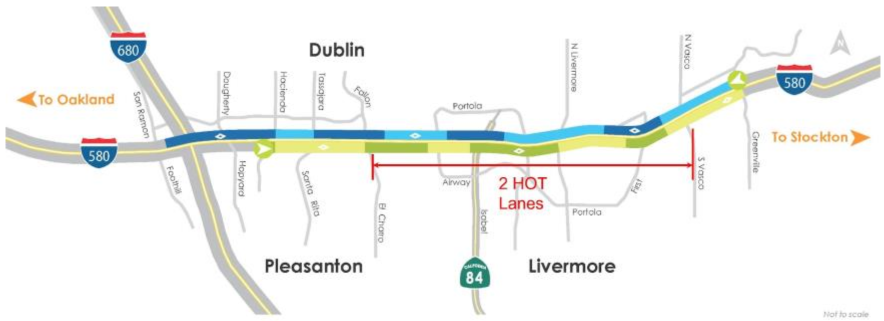

Figure ES - 1. Map of express lanes along I-580 in Alameda County, California. (Source: [Alameda County Transportation Commission](#))

The LSTM model accurately predicts traffic speeds in both HOT and general-purpose (GP) lanes, capturing temporal patterns and variations in traffic flow. This model serves as a valuable tool for understanding traffic dynamics and informing tolling strategies.

The LQR-based tolling control system integrates a lane choice model and traffic flow dynamics to formulate a multi-input, multi-output feedback control system. This system ensures convergence to optimal traffic states, reducing congestion and improving traffic flow in the HOT lane.

Analysis of the I-580 express lane dataset reveals several key findings:

- Tolling strategies significantly impact traffic flow, with the LQR tolling strategy reducing congestion levels in the HOT lane.
- Applying tolling only to the HOT lane mitigates congestion in both the HOT and GP lanes, improving traffic flow for all travelers.
- System performance analysis demonstrates improvements in vehicle hours traveled (VHT) and person hours traveled (PHT) after implementing tolling strategies. Solo drivers experience travel time savings regardless of lane choice, indicating benefits for all travelers.

Overall, the study highlights the importance of data-driven modeling and control strategies in managing traffic congestion and improving travel efficiency on express lanes. The combination of advanced modeling techniques and tolling strategies offers promising solutions for addressing traffic challenges and enhancing the transportation experience for people traveling by car.

Contents

1 Introduction

High occupancy vehicle (HOV) lanes allow carpool vehicles to bypass congestion and save travel time. However, they often face the dilemma of either under- or over-utilization (Li, 2001), leading to either the waste of precious road capacity or the loss of a travel-time advantage over general-purpose (GP) lanes. To address this issue, high-occupancy toll (HOT) lanes have gained popularity in recent years. HOT lanes, when properly priced, preserve the advantage of HOV lanes, which encourage carpooling and thereby reduce congestion and vehicle miles. Additionally, they allow single-occupancy vehicle (SOV) drivers to pay to use HOT lanes to make use of the spare capacity in HOT lanes. According to the Federal Highway Administration (FHWA), more than 100 miles of HOT lanes are operating in the US. California plays an important role in potentially optimizing HOT lane use, as it has the largest HOT lane system in the nation, including the I-15 FasTrak section in San Diego (Supernak et al., 1998) and I-580/I-680 sections in Alameda County (Alameda CTC, 2015). In addition potential to increase HOT lane miles in California is great: currently, more than 1,550 lane miles of HOV facilities are operational or under construction, and 560 lane-miles are being proposed by the California Department of Transportation. HOT lanes can be better than HOV lanes at allowing vehicles to maximize the use of road capacity and at offering SOV drivers more choices. Real-time pricing plays a key role in balancing the travel needs of carpoolers and SOV drivers and maintaining the travel advantage of HOT lanes. This balance can be difficult to achieve. To do this well one needs to design pricing schemes that (a) take into account the price sensitivity of SOVs, (b) predict the spare capacity of the HOT lanes, and (c) consider the operating conditions of GP lanes. Current pricing schemes tend to be one of two types: (1) simple schemes based on the volume or speed in HOT lanes, without consideration of the traffic conditions on GP lanes; or (2) black-box schemes that work by a protocol that vendors do not disclose to users. To the best of our knowledge, efforts are lacking to develop pricing schemes that meet the required needs (a-c) and are transparent to users. We plan to adopt a data driven approach to develop a new HOT lane pricing scheme by mining the data to reveal the underlying pattern between the traffic flow (on both HOT and GP lanes) and the toll rates, which we subsequently call the “flow-tolling pattern.”

2 Literature Review

2.1 Traffic Dynamics Prediction

Predicting traffic dynamics is an important problem in modern transportation systems. It requires a multifaceted exploration of the complex and changing nature of vehicular movement on freeways and urban roads. Substantial efforts have been made to provide accurate and timely predictions of traffic dynamics, with a primary focus on key targets such as traffic flow (number of vehicles per hour) and travel speed (Guo et al., 2020; Zheng et al., 2020; Zhao et al., 2019). Capturing time-series relationships has been recognized as an effective strategy for accurately predicting flow and speed, as target indicators of traffic dynamics (Shen et al., 2020; Lu et al., 2020). Studies have also delved into specifically predicting lane-level speed or flow (Li et al., 2023; Lu et al., 2020; Gu et al., 2019). However, the prediction of traffic dynamics on freeways, particularly with the existence of HOT lanes, remains relatively rare. In a study by Zahedian et al. (2021), traffic flow, travel time, and toll rates are integrated as input variables to forecast future toll rates. Despite this incorporation of tolling information, the intricate flow-toll pattern has yet to be comprehensively captured.

Extracting the dynamic relationship between traffic speed and tolling rates can be defined as a short- term sequential prediction problem. In practice, drivers respond not only to current tolls to make a lane change but also learn from their experiences of using HOT lanes, and traffic entering from upstream points takes time to affect downstream sections, and vice versa. These driving behaviors and vehicle entry dynamics result in a time delay and long-term dependency between the toll and traffic dynamics, which can be captured by sequence learning (Chen and Grant-Muller, 2001; Clegg et al., 1998). One such powerful sequence learning method is the long short-term memory (LSTM) approach (Zhao et al., 2017). The LSTM method is uniquely equipped to compute optimal time lags in traffic flow and capture the extensive temporal dependencies inherent in traffic flow and toll rates. As such, it may be able to characterize the dynamic interplay between tolling strategies and traffic speed in HOT lanes and guide transportation policy makers.

2.2 HOT Lane Tolling Strategy

Improving tolling strategies for HOT lanes requires a thorough comprehension of travelers' willingness to use the HOT lane. Carpoolers choose the HOT lane often because of no tolling fees with travel time benefits. Less is known about how solo drivers respond and adapt to the presence of HOT lanes, tolling, and traffic conditions. Liu et al. (2011) that tolls are elastic to solo drivers' lane choice behaviors. Also, solo drivers weigh time saving against the cost of tolls (Janson and Levinson, 2014). When the GP lane is congested due to heavy traffic, they are more willing to pay to access the HOT lane, seeking to save travel time. Or when the tolls are too high, they would rather suffer from congestion. In turn, it is essential to develop a tolling strategy that helps to benefit all travelers and also reduce delays on both HOT and GP lanes.

Numerous studies have concentrated their efforts on tolling strategies. Michalaka et al. (2011) formulated a robust pricing optimization problem to maximize the total throughput while controlling the congestion level in HOT lanes. The traffic dynamics were described by the cell transmission model, and the flow-rates in the GP and HOT lanes were estimated by a logit model. Beyond the optimization method, more robust approaches using feedback control have been applied. Zhang et al. (2008) applied a piecewise feedback control model to calculate the probability of choosing HOT lanes based on different speeds in the HOT and GP lanes. The price was estimated from the logit model. Yin and Lou (2009) proposed a feedback method and a self-learning method to determine the dynamic tolls to provide a free-flow traffic condition in the HOT lanes while maximizing the throughput on the freeway. For both methods, they used a logit model to capture the lane choice of solo drivers in a freeway segment with HOT and GP lanes, and a point queue model to capture traffic dynamics. Recently, Pandey et al. (2020) developed a deep reinforcement learning framework for dynamic tolling in HOT lanes with multiple points of access and origin-destinations. They formulated a partially observable Markov decision process and policy gradient methods to determine tolls as a function of real-time observations. However, existing literature, to our knowledge, has the following limitations: (1) The optimization and feedback strategies considered only a single toll controller, which is not able to smooth traffic along the corridor, as most HOT lanes can be accessible in multiple locations; (2) The reinforcement learning method does not capture the impacts of lane changes. We aim to handle these issues in our report.

3 Methodology

3.1 Traffic Speed Prediction for HOT and GP Lanes

Our approach is centered around the construction of a comprehensive traffic speed forecasting model. The aim is to create a holistic framework that enhances the accuracy of sequential traffic speed predictions. The main strategy is to propose a generic traffic speed forecasting model based on historical observations, including historical traffic speeds, dynamic toll rates, and auxiliary features such as time factors and neighboring spatial location information to make sequential traffic speed predictions.

Time series prediction analysis uses the characteristics of an event in a past period to predict the characteristics of the event in the future. Data on traffic dynamics are typically represented as a time series. Through the analysis and mining of these series data, the corresponding traffic dynamic forecasting model can be established to predict the trend of traffic speed in advance, which would enable its users to obtain forward-looking traffic information for travel. However, due to the sequential characteristics of traffic conditions, there is generally an internal strong correlation between traffic observations. Hence, the local time-series correlation needs to be well considered before building the forecasting model. Considering their temporal dependency, time series forecasting models are relatively complex for conventional prediction models, due to their nonlinear variation features. Also, unlike the prediction of a regression analysis model, the prediction of a time series model depends on the sequence of events being modeled. Moreover, the large quantity of traffic speed observation data generated by high-frequency sampling also makes traditional models unfeasible.

Neural networks, especially deep learning models such as LSTM, have been widely considered more powerful than conventional models for time-series forecasting (Fu et al., 2016; Cui et al., 2018; Qi et al., 2019). Hence, in the project, LSTM is used as the backbone of the traffic dynamic forecasting model.

3.1.1 LSTM Model

LSTM is a variant version of a recurrent neural network (RNN). For RNNs, the results of each hidden layer are related to the current input, and the last hidden layer is compared to the common neural network. With this method, RNN calculation results have the characteristics of remembering previous results while training and predicting. An LSTM model is a variant of RNN, which can deal with temporal dependency in the current input and last several hidden inputs. The typical structure of an LSTM memory block is shown in Figure 1. There are at least one memory cell and three nonlinear controlling gates within each memory block. The first two gates, the forget gate and the input gate, control the content state of the cell; the output gate, on the other hand, controls the amount of the cell state to be transferred to the current output value of the whole LSTM block. This process can be mathematically represented as Equation 5.

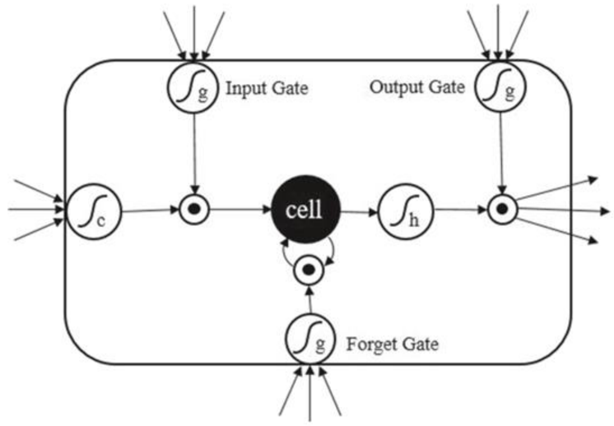

Figure 1. Structure of an LSTM memory block (with only 1 cell)

$$\boldsymbol{f}_t = sg\left(\boldsymbol{W}_f x_t + U_f h_{t-1} + b_f\right) \quad \text{(Eq. 1)}$$

$$i_t = sg(\boldsymbol{W}_i x_t + U_i h_{t-1} + \boldsymbol{b}_i) \quad \text{(Eq. 2)}$$

$$o_t = sg(\boldsymbol{W}_o x_t + U_o h_{t-1} + \boldsymbol{b}_o) \quad \text{(Eq. 3)}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot sc(\boldsymbol{W}_c \boldsymbol{x}_t + \boldsymbol{U}_c \boldsymbol{h}_{t-1} + \boldsymbol{b}_c) \quad \text{(Eq. 4)}$$

$$h_t = o_t \odot tanh\,(\boldsymbol{c}_t)\,, \quad \text{(Eq. 5)}$$

where $\mathbf{W}_f$, $\mathbf{W}_i$, $\mathbf{W}_o$, and $\mathbf{W}_c$ are the weight matrices for the input vector $\mathbf{x}_t$ at time step t. $\mathbf{U}_f$, $\mathbf{U}_i$, $\mathbf{U}_o$, and $\mathbf{U}_c$ are the weight matrices assigned to hidden state values from the previous block h_{t-1}. $\mathbf{b}_f$, $\mathbf{b}_i$, $\mathbf{b}_o$ and $\mathbf{b}_c$ are the bias vectors. To bring nonlinearity to the model, different types of activation functions are involved and represented as sg, sc and $tanh$. The symbol $\odot$ represents the multiplication by element of the matrix.

3.1.2 Model Architecture

The main objective of the traffic dynamic model is to propose a generic traffic speed forecasting model based on historical traffic dynamic observations and other auxiliary features to make traffic speed predictions for the HOT lane and the GP lane. The diagram in Figure 2 demonstrates the sequential steps of model input and output. For the sequential forecasting model, historical traffic-speed-related observations are inputted over a time step of 3, with the target traffic-speed output set at a length of 1. The time interval is configured to 15 minutes. The extended architecture, as shown in Figure 3, builds upon the basic LSTM structure.

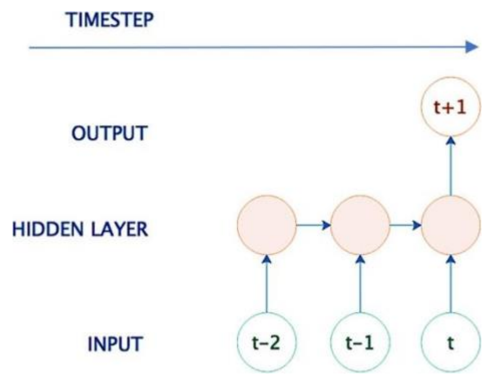

Figure 2. Diagram of sequential input and output steps for the traffic speed model

To enable traffic speed prediction, the sequential forecasting framework for traffic speed prediction is extended from the basic LSTM backbone (Figure 3).

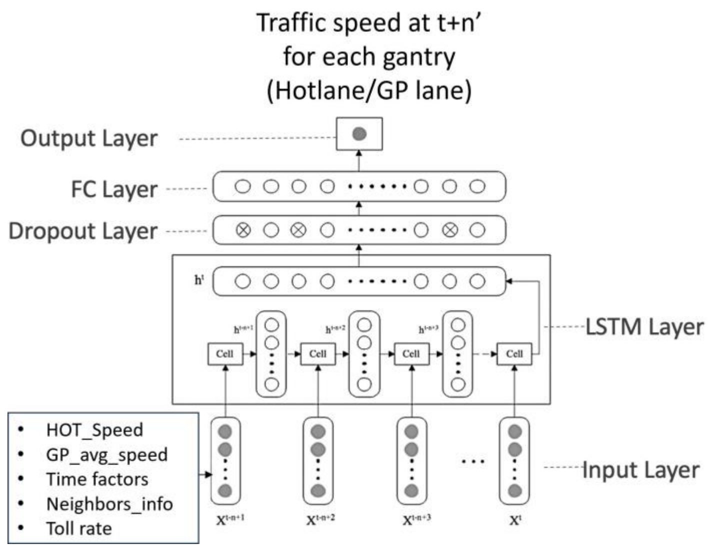

Figure 3. Architecture of the proposed traffic speed forecasting framework

3.2 Pricing Strategy

We sought to provide efficient online feedback tolling strategies for solo drivers in the HOT lane to achieve two primary goals: 1) optimizing the operational efficiency of the HOT lane by reducing congestion, and 2) incentivizing more solo drivers to choose the HOT lane, thereby alleviating congestion in the GP lane. To investigate how tolls affect solo drivers' decision-making, we introduce a logit model to estimate the relationship between toll rates, travel time differences, and lane choices of solo drivers. Then, building upon the Lighthill-Whitham-Richards (LWR) traffic flow model, we propose a multi-input, multi-output feedback control system based on a fundamental diagram for I-580 and I-680 to derive the spatial relationship between flow, speed, and density of the HOT lane. Integrating such models allows us to formulate a dynamic tolling price, guiding traffic toward an optimal speed state. Such convergence enhances the system efficiency and ensures a real-time tolling strategy to address congestion.

3.2.1 Lane Choice Model

We consider the lane choice model for solo drivers only. Following the model by Wang et al. (2020), the proportion of solo drivers choosing the HOT lane at each gantry (segment) i, $P_i(t)$is presented as:

$$P_i(t) \; = \; \frac{1}{1+e^{\pi w_i(t)-u_i(t)}}\,, \qquad \text{(Eq. 6)}$$

where π denotes the value of time, which is assumed as a constant representing the opportunity cost of the time a traveler spends on trips. This constant plays an important role in congestion pricing analysis, showing traveler's trade-off between cost and time. $u_i(t)$ represents the tolling rate we aim to control at gantry i. $w_i(t)$ is the travel time difference between the HOT and the neighboring GP lanes at gantry i, which is equivalent to:

$$w_i(t) = w_i^{GP}(t) - w_i^{H}(t) \qquad \text{(Eq. 7)}$$

Note that the travel time can be also derived using speed at gantry i as v_i:

$$w_i^{GP}(t) = \frac{L_i}{v_i^{GP}(t)} \qquad \text{(Eq. 8)}$$

$$w_i^{H}(t) = \frac{L_i}{v_i^{H}(t)} \qquad \text{(Eq. 9)}$$

Plugging these equations into (6), we get:

$$P_i(t) = \frac{1}{1+e^{\left(\left[\frac{\pi L_i}{v_i^{GP}(t)} - \frac{\pi L_i}{v_i^{H}(t)}\right] - u_i(t)\right)}} \qquad \text{(Eq. 10)}$$

The proportion of solo drivers in the HOT lane can be represented as the flow of solo drivers $q_{i,SOV}^{H}(t)$ in the HOT lane over their total flow $q_{i,SOV}^{T}(t)$ at gantry i:

$$P_i(t) = \frac{q_{i,SOV}^{H}(t)}{q_{i,SOV}^{T}(t)}\,, \qquad \text{(Eq. 11)}$$

where $q_{SOV}^{T}(t) = q_{SOV}^{H}(t) + q_{SOV}^{GP}(t)$. From these equations, we derive a nonlinear function, $u_{i(t)} = f\left(v_i^{GP}(t), v_i^{H}(t), q_{i,SOV}^{H}(t), q_{i,SOV}^{GP}(t)\right)$ that describes the relations between tolls, speeds and flows on both GP and HOT lanes for gantry i as:

$$u_i(t) = \pi L_i\left(\frac{1}{v_i^{GP}(t)} - \frac{1}{v_i^{H}(t)}\right) + ln\ ln\ \frac{q_{i,SOV}^{H}(t)}{q_{i,SOV}^{GP}(t)} \qquad \text{(Eq. 12)}$$

This equation describes how speed differences and lane choices of solo drivers on GP and HOT lanes affect the tolling rates.

3.2.2 Traffic Dynamic System

After deriving the lane choice models, we used the idea of traffic dynamic models proposed by Zhang and Ioannou (2016). By conservation law, the dynamics of density $\rho(t)$ for HOT lane gantry i is described by the differential equation:

$$\dot{\rho}_i(t) = \frac{q_i^{in}(t) - q_i^{out}(t)}{L_i} \qquad \text{(Eq. 13)}$$

For simplicity, in equation (13), we omit the label H, as the HOT lane. Therefore, Equation (13) shows that the change in density is proportional to the difference between inflow $q_i^{in}(t)$ and outflow $q_i^{out}(t)$. Since solo drivers decide which lane to choose at the starting point of each gantry based on tolls and congestion levels, we denote $q_i^{in}(t) = q_i(t)$ and $q_i^{out}(t) = q_{i+1}(t)$. Each gantry has its tolling strategy, both $q_i(t)$ and $q_{i+1}(t)$ can be regulated by tolls $u_i(t)$ and $u_{i+1}(t)$. By the fundamental diagram, we have the traffic flow equation:

$$q(t) = \rho(t)v(t) \qquad \text{(Eq. 14)}$$

With Equation (14), we derive the conservation law as:

$$\dot{\rho}_i(t) = \frac{v_i(t)\rho_i(t) - v_{i+1}(t)\rho_{i+1}(t)}{L_i} \qquad \text{(Eq. 15)}$$

In this equation, the speeds of gantry i and $i+1$ function as pivotal control variables to regulate the traffic flow. The fundamental diagram identifies a critical density point at which maximum throughput and optimum speed are achieved. These critical values are denoted as ρ^c, q_c, and v_c respectively. This control system is used to determine v_i, which regulates traffic flow to achieve ρ^c and q^c, which in turn can be regarded as an optimal state. To systematically achieve these objectives, we introduce an error system as: $e_i = \rho_i - \rho_i^c$, $z_i = v_i - v_i^c$. Substituting them into (15), we have:

$$\dot{e}_i(t) = \frac{z_i(t)\rho_i(t) + v_i^c e_i(t) - v_{i+1}^c e_{i+1}(t) - z_{i+1}(t)\rho_{i+1}(t)}{L_i} \qquad \text{(Eq. 16)}$$

Through linearization at a fixed point ρ^c, the above equation becomes:

$$\dot{e}_i(t) = \frac{z_i(t)\rho_i^c(t) + v_i^c e_i(t) - v_{i+1}^c e_{i+1}(t) - z_{i+1}(t)\rho_{i+1}^c(t)}{L_i} \qquad \text{(Eq. 17)}$$

If we define $e = [e_1, e_2, \ldots, e_N]^T$ and $z = [z_1, z_2, \ldots, z_N]^T$, then the error system can be written as:

$$\dot{e} = \begin{bmatrix} \frac{v_1^c}{L_1} & -\frac{v_2^c}{L_1} & & & \\ & \frac{v_2^c}{L_2} & -\frac{v_3^c}{L_2} & & \\ & & \ddots & \ddots & \\ & & & \frac{v_{N-1}^c}{L_{N-1}} & -\frac{v_N^c}{L_{N-1}} \\ & & & & \frac{v_N^c}{L_N} \end{bmatrix} e + \begin{bmatrix} \frac{\rho_1^c}{L_1} & -\frac{\rho_2^c}{L_1} & & & \\ & \frac{\rho_2^c}{L_2} & -\frac{\rho_3^c}{L_2} & & \\ & & \ddots & \ddots & \\ & & & \frac{\rho_{N-1}^c}{L_{N-1}} & -\frac{\rho_N^c}{L_{N-1}} \\ & & & & \frac{\rho_N^c}{L_N} \end{bmatrix} z \qquad \text{(Eq. 18)}$$

In a compact form, we set A as the state space of e, and B as the state space of z. Alternatively, $\dot{e} = Ae + Bz$.

3.2.3 LQR Control Design

The error system dynamics (18) are a set of linear differential equations to be controlled so that e converges to 0 which indicates states of all gantries will converge to their corresponding optimal ones (ρ^c, v^c). To achieve this, we implement the linear-quadratic regulator: LQR (Bemporad et al., 2002), a feedback controller given the following formulations:

$$J = \int_0^\infty (e^T Q e + z^T R z + 2e^T N z) dt \quad \text{(Eq. 19)}$$

$$z = -Ke \quad \text{(Eq. 20)}$$

$$K = R^{-1}(BP + N^T) \quad \text{(Eq. 21)}$$

$$0 = A^T P + PA - (PB + N)R^{-1}(B^T P + N^T) + Q \quad \text{(Eq. 22)}$$

Here, Q and R stand as two pre-determined cost function parameters, while N, K, and P are derived through Equations 19–22. In each control cycle, the LQR controller calculates z based on the current state e, determining optimal $[v_1, v_2, \dots, v_N]^T$. Subsequently, with the state variables v^{GP}, v^H, q^H_{SOV}, and q^{GP}_{SOV}, we obtain the feedback tolling price u using Equation (12).

4 Experiments and Results Discussion

4.1 I-580 HOT Lane Data General Information

In 2016, the Alameda County Transportation Commission (Alameda CTC) opened the I-580 Express Lanes—HOT lanes in the eastbound and westbound directions of I-580 in the Dublin-Pleasanton-Livermore area of eastern Alameda County (Figure 4).

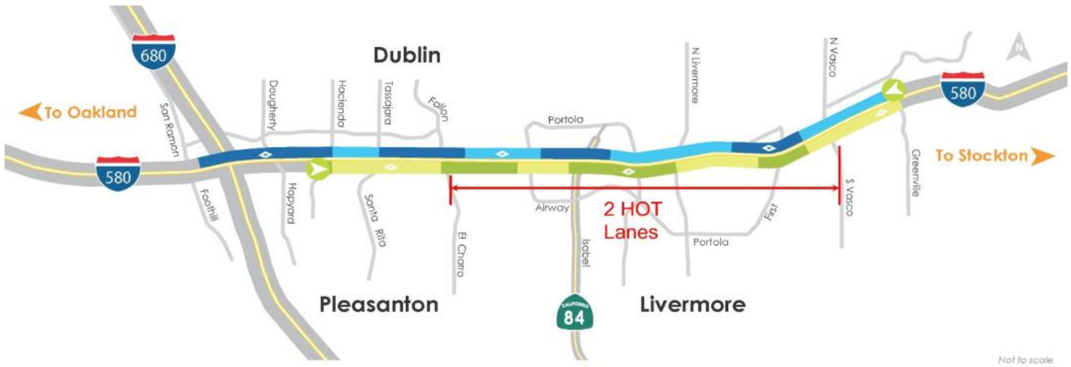

Figure 4. Map showing the areas of the HOT ("Express") lanes along I-580 (Source: Alameda County Transportation Commission)

The eastbound lanes run 11 miles from Hacienda Drive in Pleasanton to Greenville Road in Livermore; and the westbound lanes, 14 miles from Greenville Road in Livermore to San Ramon Road/Foothill Road in Dublin. There are two eastbound HOT lanes and one westbound HOT lane. The freeway segment is divided into several zones and gantries. An electronic toll system is installed on the gantries to collect trip data and show the price sign. All travelers must have a valid FasTrak account to use the HOT lanes to pay a toll. The I-580 Express datasets cover one month (23 weekdays) in October 2018. Details are described below.

- *Trip transaction data* consists of vehicle IDs, time, speed, and payment for the HOT lanes.
- *Traffic data* consists of speed and traffic volume every 15 minutes for all lanes.
- *Toll rate data* captures the dynamic toll amount every 3 minutes by zone. The tolling operation starts from 5 AM to 8 PM on weekdays and the toll amount ranges from $0 to $12.25 with 64 different values.

In our experiments, we chose the westbound HOT lane as an example to evaluate our proposed tolling strategy. In particular, the westbound HOT lane has 8 zones and 17 gantries. The toll amount can be changed at each gantry.

4.2 Exploratory Data Analysis

This section describes the studied freeway section and its underlying flow-toll patterns.

4.2.1 Time-Space Heatmap

Figure 5 shows heatmaps of the average traffic speed at each gantry at different times of day for the HOT lane and its adjacent GP lane. The speed is averaged over the 1-month study period (October 2018). In morning and evening peak hours, at most gantries, the speed in the HOT lane was faster than in the adjacent GP lane, indicating less congestion in the former. This finding suggests that the tolling strategy contributed to reduced congestion in the HOT lane. However, congestion still exists, which suggests the necessity of a tolling strategy to further improve the HOT lane efficiency.

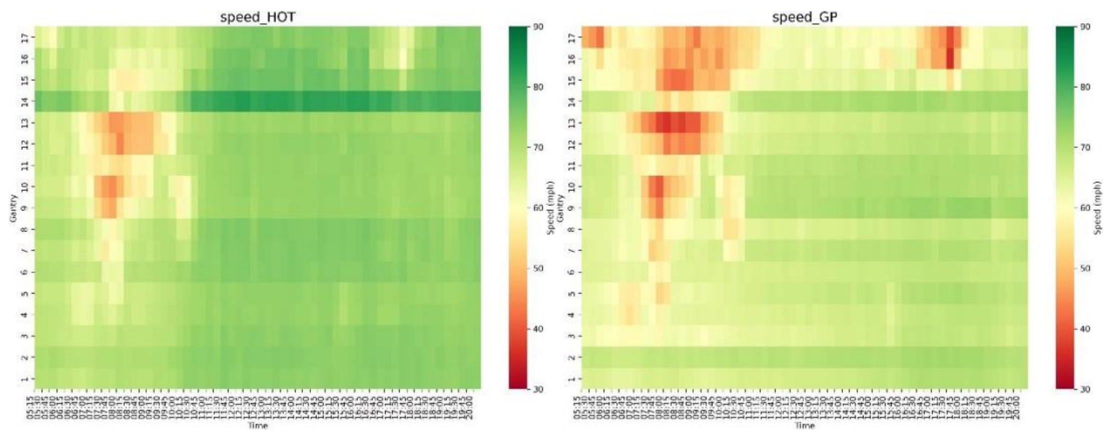

Figure 5. Space-time heatmap for the HOT lane (left) and adjacent GP lane (right).

4.2.2 Toll Rate and SOV Origin Trips Over Day

We plotted the fluctuations in SOV flow at origin and toll rates across various hours and days. The aim of this analysis is to investigate potential variations in recorded SOV flow and toll rates across different hours and days, ensuring that they do not remain constant. Understanding these variations is essential for enhancing existing tolling strategies, aiming not only to alleviate congestion but also to guarantee the effectiveness and adaptability of the tolling system to the specific demands observed on different days and at various times of the day. The figures presented below focus on gantry W1 as an illustrative example. The other gantries exhibit comparable patterns to W1.

In Figure 6, the green line represents actual data, and the dotted horizontal line represents the average value. Notably, within the same hour, there is observable variation in SOV origin flows over days. This suggests the necessity of a dynamic tolling strategy in response to these varying demands. In addition, the highest SOV origin flows occur during morning peak hours, indicating a concentration of single-occupancy commuters during this period.

According to Figure 7, within one hour, toll rates vary across days. This observation suggests that the tolling system is responsive to the dynamic changes in traffic flow. Moreover, toll rates reach their peak during the morning peak hours, suggesting an alignment with the peak hour traffic demands.

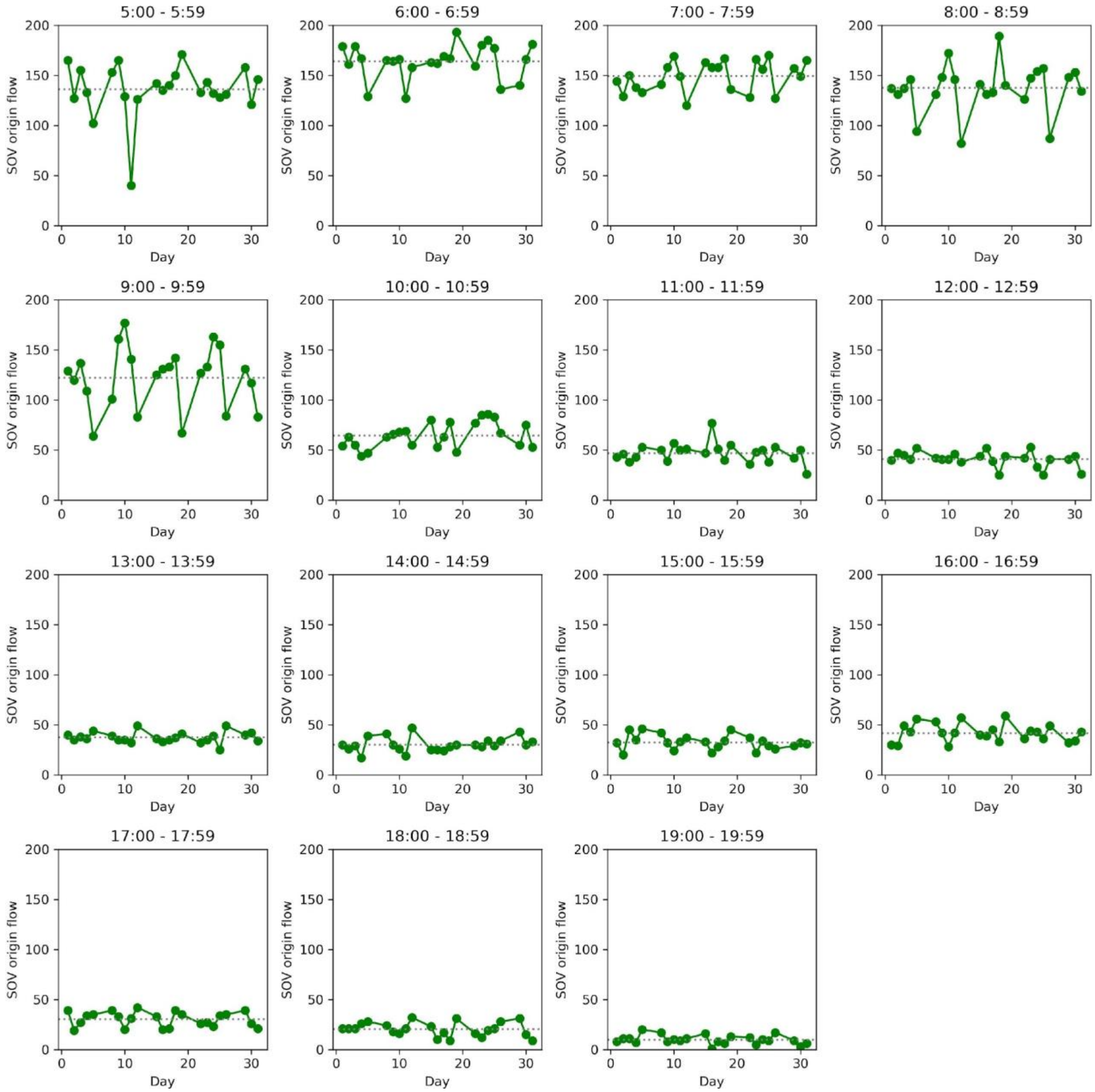

Figure 6. SOV origin flow over days at different hour of day

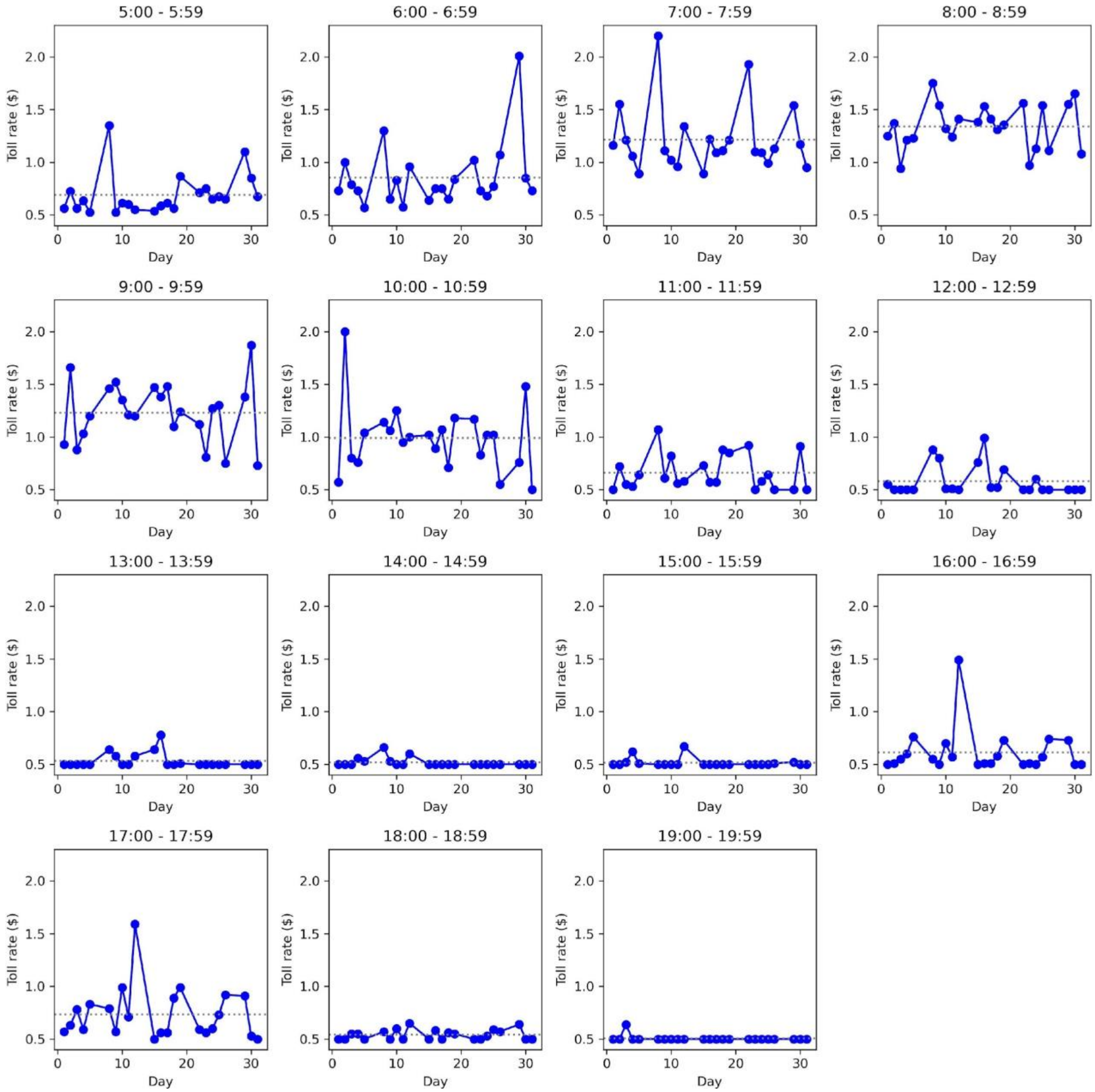

Figure 7. Toll rate over days at different hour of day

4.3 Correlation Analysis

In this section, we compute the correlation between the toll rate and SOV origin flow at different times of the day. We conduct a comparison between two correlation metrics, the Pearson's correlation coefficient and the Spearman's correlation coefficient, to assess the inherent relation (i.e., linear or non-linear) between toll rate and SOV flow.

The Pearson's coefficient is used to quantify a linear relationship between two continuous variables. It ranges from -1 to 1, indicating the strength and direction of the linear correlation. A value of -1 indicates a perfect negative linear correlation, 1 indicates a perfect positive linear correlation, and 0 denotes no linear correlation.

On the other hand, the Spearman's coefficient is suitable for dealing with relations that may not follow a linear pattern. Similar to the Pearson's coefficient, it ranges from -1 to 1, with -1 indicating a perfect negative monotonic correlation, 1 indicating a perfect positive monotonic correlation, and 0 indicating no monotonic correlation.

The correlation between toll rate and SOV origin flow is presented in Figure 8. From 6 to 7 AM, the Spearman method yields higher correlation values than the Pearson method. This observation suggests the existence of a nonlinear correlation between these variables during this time period. As the day progresses, the correlation values computed by both methods exhibit similar variation patterns. Although there are instances where Spearman's method produces slightly lower correlation values than Pearson's, generally the Spearman coefficients are higher. This pattern highlights the presence of a non-linear relation between toll rates and SOV origin flows. This validates our choice of employing non-linear modeling techniques, such as neural networks, to effectively capture this relationship. In addition, positive correlations are observed over the day, with higher values particularly evident during morning and evening peak hours.

4.4 Traffic Speed Prediction Result

4.4.1 Experiment Settings

In our experimental setup, the training set encompasses 80% of the available data, providing the LSTM model with ample exposure to the temporal dynamics within historical traffic speed patterns. The remaining 20% constitutes the validation set for fine-tuning the model and preventing overfitting. Additionally, a distinct test set, spanning an entire day of speed observations in HOT and GP lanes, is reserved for evaluating the model's performance on entirely unseen data. This chosen test dataset corresponds to the last day of the month, ensuring an evaluation on previously unencountered data, simulating real-world scenarios and gauging the model's generalizability.

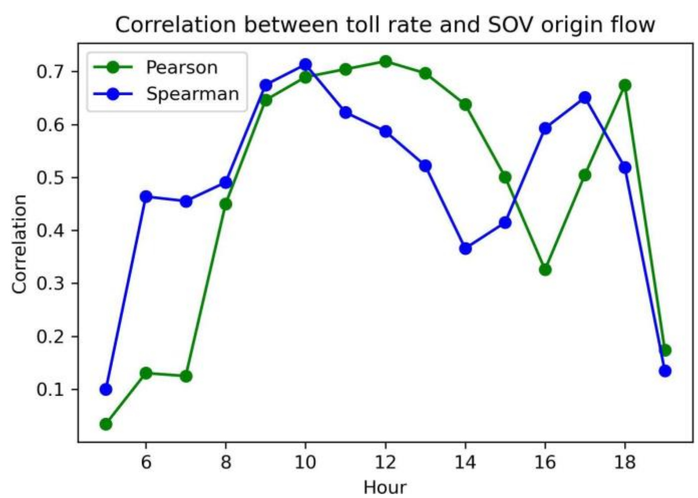

Figure 8. Correlation between toll rate and origin volume over time

The training and evaluation of the traffic speed prediction model is visualized in Figure 9, showing the model's learning trajectory.

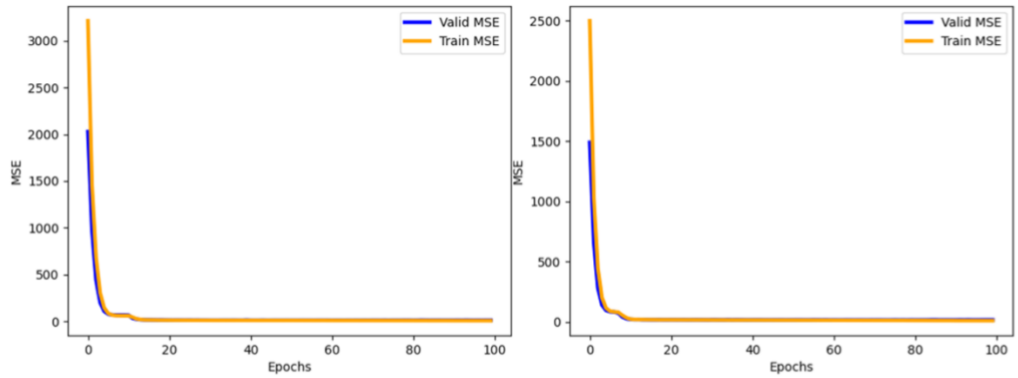

Figure 9. Training and evaluation loss for HOT-lane (left) and GP-lane (right) speed prediction

The model performance is summarized in table 1, including statistics on mean squared error (MSE) for both the training and validation sets. The traffic speed prediction model demonstrates effective learning from historical patterns, minimizing disparities in the training dataset, as indicated by the converging training loss. Simultaneous convergence of the validation loss highlights the model's adeptness at generalizing to new, unseen data—essential for real-world predictions in HOT lane speed and GP lane average speed scenarios. The convergence around the 30th epoch suggests a good balance between model performance and efficiency, showing stability and consistent predictive behavior.

Table 1. Comparison of model performance (mean squared error) on training and validation sets for HOT and GP lanes

	Train Loss	Validation Loss
HOT	7.82	10.62
GP	15.63	17.69

4.4.2 Traffic Speed Prediction Results

The traffic speed prediction results, illustrated in Figure 10, demonstrate the model's performance on the entire unseen testing dataset. Each prediction is generated by providing the historical observations within a 15 × 3-minute time window, showcasing the model's proficiency in mapping the intricate patterns within the data. The temporal prediction results for the HOT lane and the GP lane demonstrate good accuracy in capturing the evolving dynamics of traffic speed. This accuracy shows the model's ability to capture the nuances of congestion, especially prominent during the morning peak hours. From temporal patterns throughout the testing day from 6:00 to 20:00, a noticeable speed drop emerges during the critical hours of 7:00 to 9:00. The developed LSTM-based model captures this nuanced variation trend, showcasing its ability to reflect real-world congestion dynamics, especially pronounced during morning peak hours. While variations exist across different gantries, the model consistently captures the observable trend of the HOT lane speed generally exceeding the

GP lane speed. This alignment with real-world scenarios validates the model's capacity to forecast and accurately represent speed fluctuations.

In summary, the testing results highlight the model's effectiveness in real-world traffic scenarios and establish it as an effective tool for predicting traffic speed for the HOT lane and GP lane. The model demonstrates its capacity to identify temporal patterns, navigate variations during peak hours, and account for lane-specific speed differences. These strengths make it useful in diverse traffic conditions.

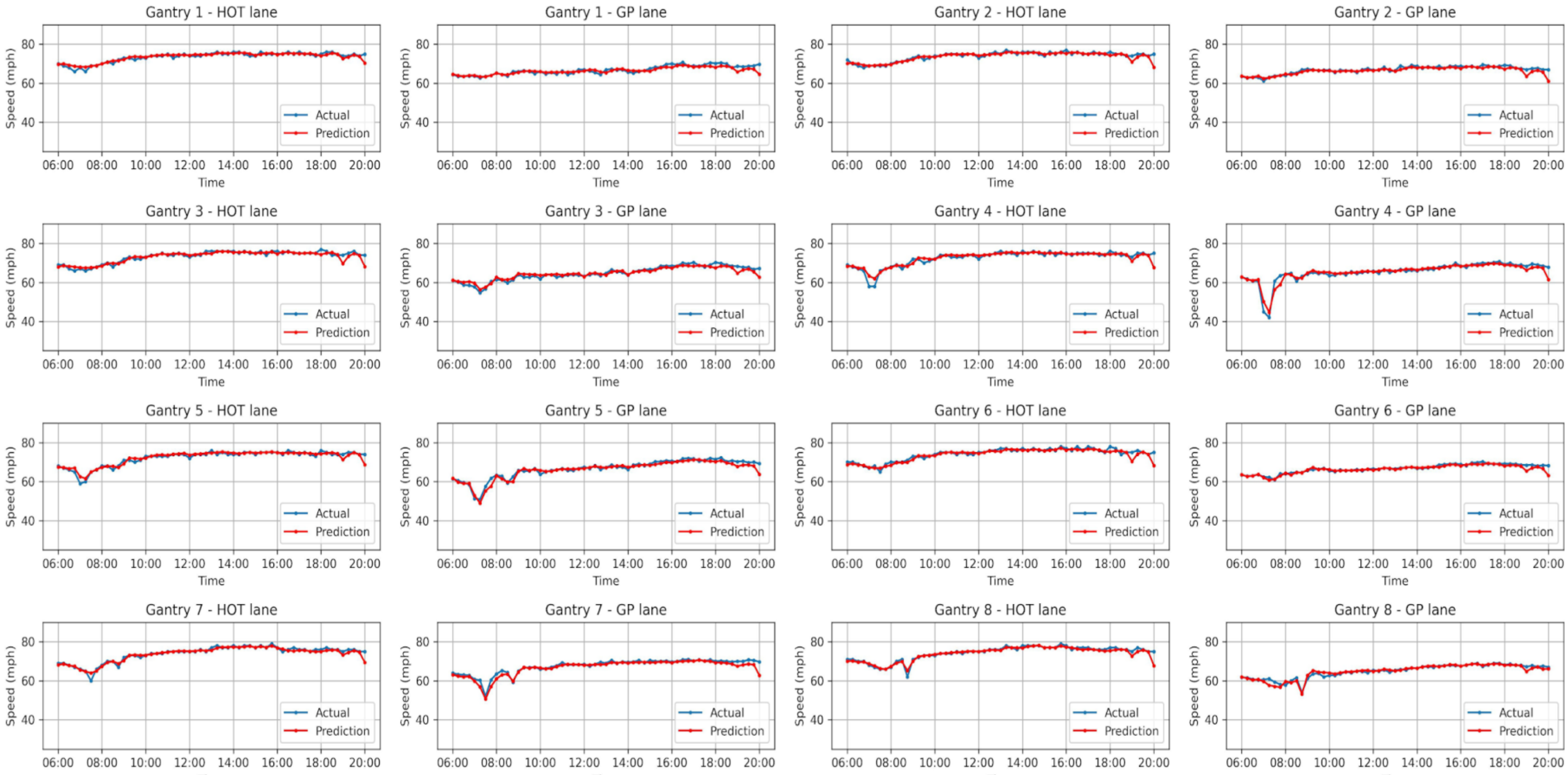

Figure 10A. Testing predictions of traffic speed in HOT lanes and GP lanes (gantries 1–8)

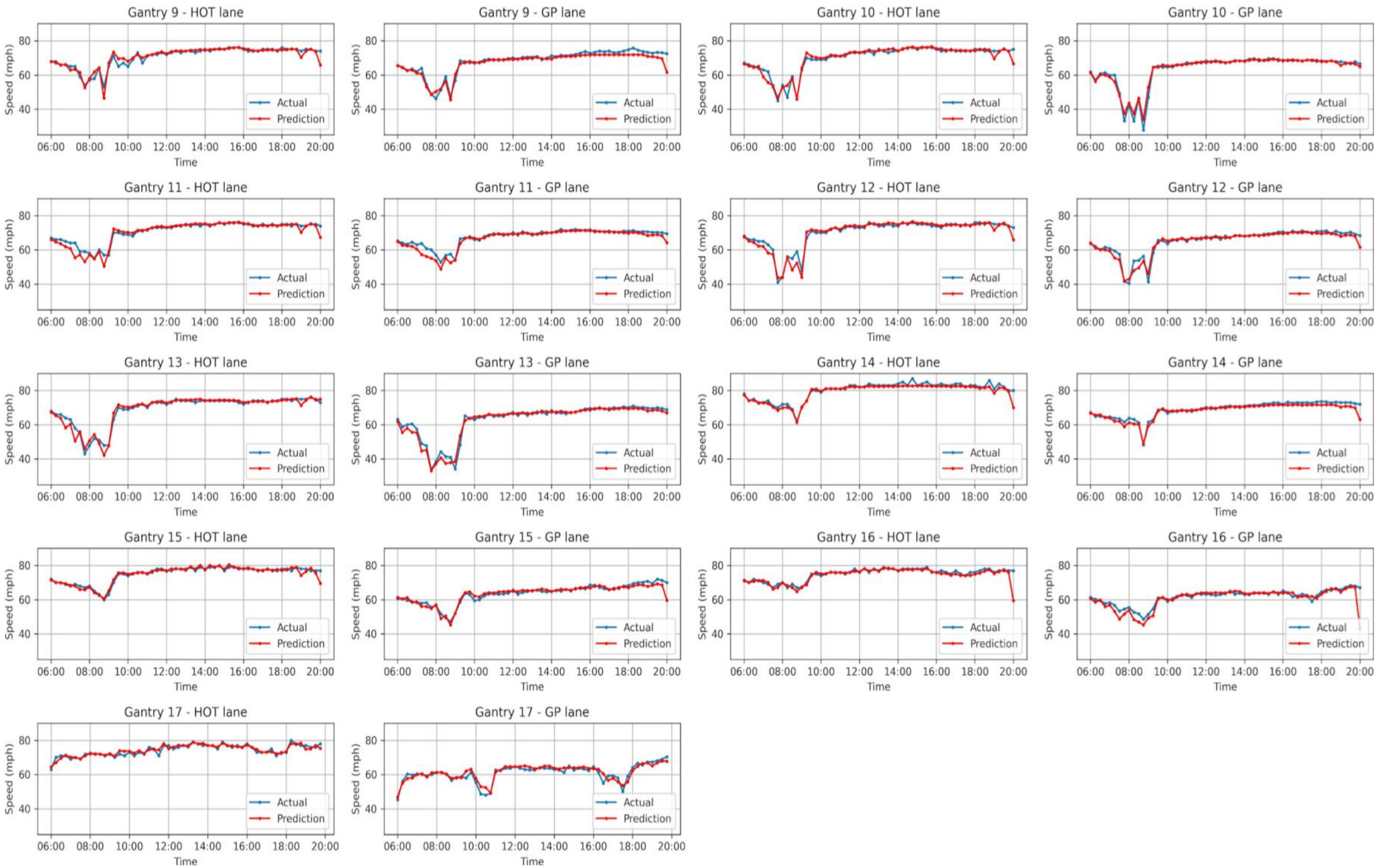

Figure 10B. Testing predictions of traffic speed in HOT lanes and GP lanes (gantries 9–17)

4.5 LQR Based Tolling Control Results

In this section, we evaluate the system's performance using speed visualization and two indices: vehicle hours traveled (VHT) and person hours traveled (PHT). We use PHT to quantify HOT lane benefits.

4.5.1 Speed Heatmap

Figures 11 and 12 display speed heatmaps for the HOT lane and its neighboring GP lane on Day 1, as an example. For the HOT lane, a notable morning peak hour spans from 7:00 AM to 9:00 AM. Under the pricing strategy implemented by Fastrak, a heavy congestion zone (speed around 30 mph) appears, extending from gantry 7 along gantry 14. This congested area covers Livermore, Isabel, Airway, and Fallon, posing significant challenges for morning commuters residing nearby or needing to pass through these zones. However, after implementing the proposed dynamic pricing, the severity of the congestion zone is diminished remarkably. Speed improves up to 50 mph and mostly to 60 mph in the HOT lane. This improvement signifies a more acceptable traffic flow during the morning commute and is more friendly to travelers who are willing to pay to use the HOT lane. These two plots demonstrate the effectiveness of the proposed dynamic pricing strategy in the HOT lane, which makes commuters more evenly distributed, therefore mitigating the congestion level. Regarding the GP lane, both morning peak (6:30 AM - 9:30 AM) and evening peak (17:00 PM - 18:00 PM) hours are observed. Interestingly, the congestion levels of morning and peak hours also exhibit a significant reduction. The impact of the proposed dynamic pricing strategy on speeds (and congestion levels) in the HOT and GP lanes indicates that this strategy would affect travelers' lane choices. Compared to the existing rate system, this pricing strategy leads more travelers to choose the HOT lane, reducing congestion levels in the GP lane.

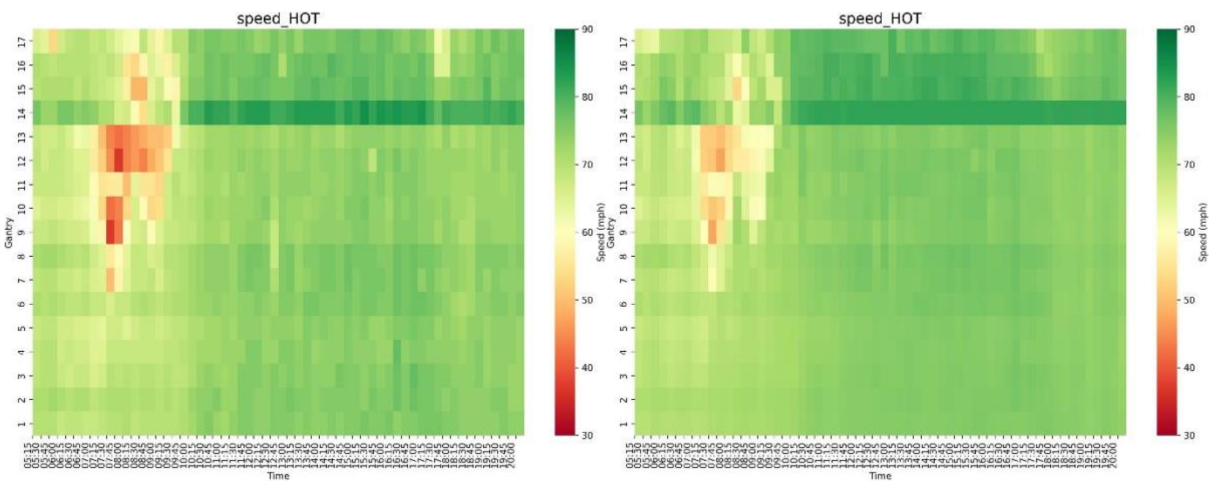

Figure 11. Average speed on Day 1 in the HOT lane before (left) and after (right) implementing the proposed tolling strategy

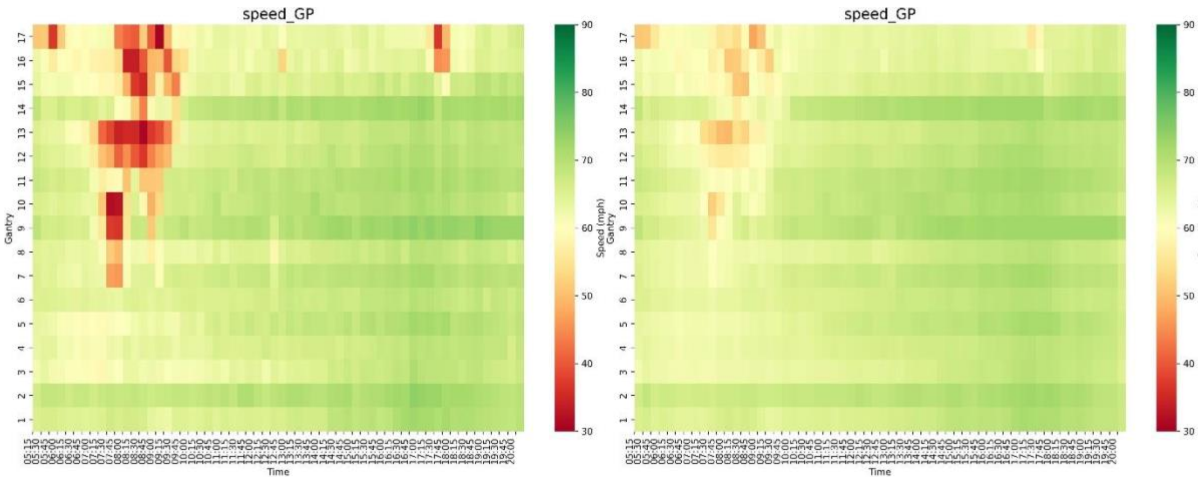

Figure 12. Average speed on Day 1 in the GP lane before (left) and after (right) implementing proposed tolling strategy

Daily traffic can be highly volatile due to random fluctuations and short-term factors like accidents. To avoid these variations from using a single day's result, we applied the dynamic pricing strategy using averaged data over one month to generate corresponding monthly average speeds as shown in Figures 13 and 14. The performance is consistent with Day 1.

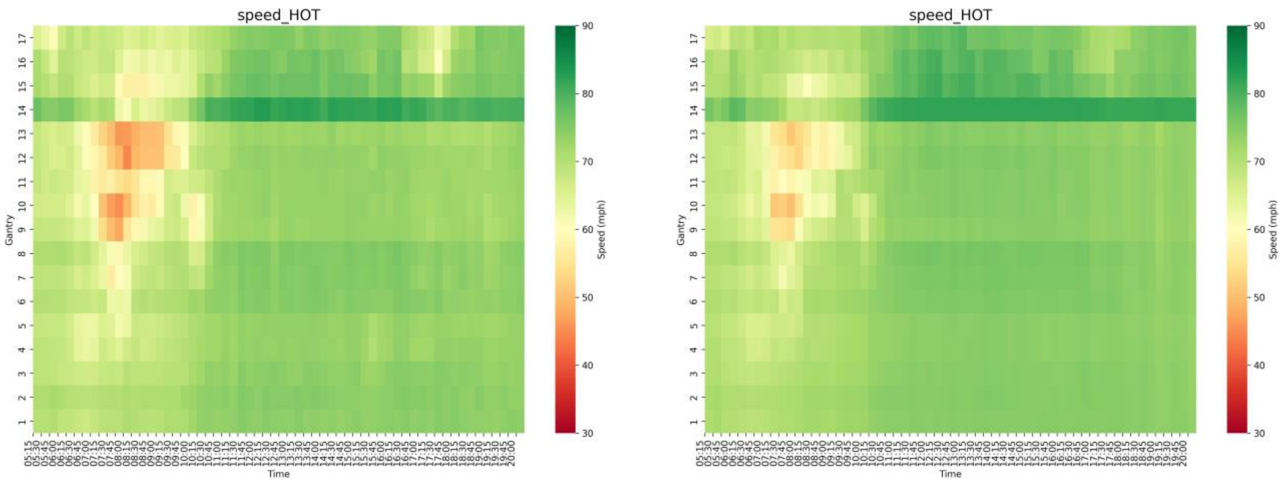

Figure 13. Average speed over one month in the HOT lane before (left) and after (right) implementing the proposed tolling strategy

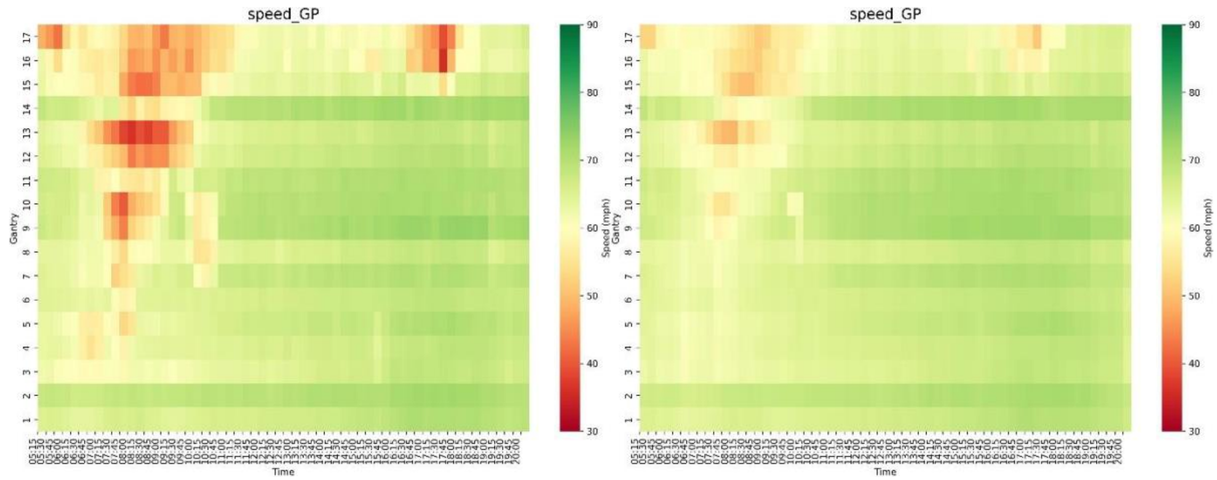

Figure 14. Average speed over one month in the GP lane before (left) and after (right) implementing the proposed tolling strategy

Figure 15 displays the average speed difference between the HOT and GP lanes before/after the LQR tolling strategy over 31 days for all gantries. The two plots demonstrate similar trends in which speeds in both lanes increase during morning and evening peak hours. The speed differences are greater and statistically significant for the morning peak and smaller and less significant for the evening peak hours. This might be due to the fact that most commuters to the San Francisco Bay Area community benefit districts use the westbound lanes (the lanes in this analysis) in the morning and eastbound lanes in the afternoon.

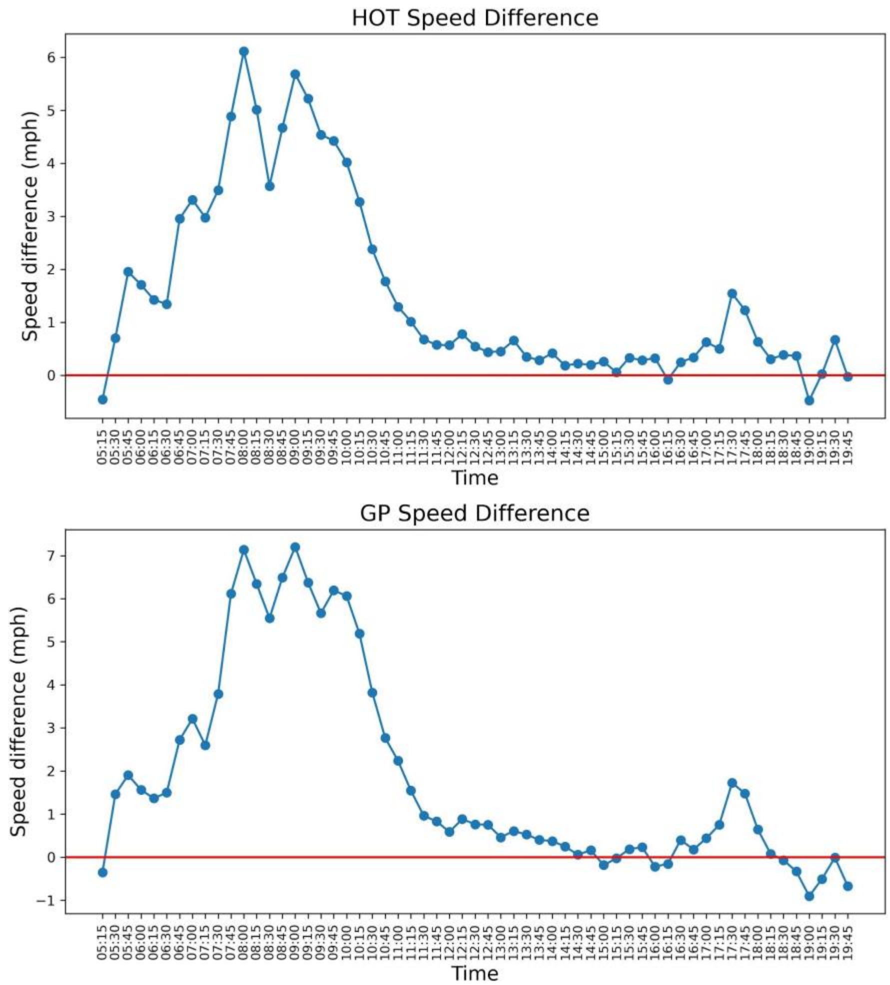

Figure 15. Speed differences in the HOT and GP lanes before (top) and after (bottom) implementing the proposed tolling strategy

4.5.2 Performance Analysis for Different Types of Travelers

Figure 16 shows the VHT in the HOT and GP lanes for all travelers. Figure 17 displays PHT in the HOT lane for all travelers, PHT in the HOT lane for solo drivers, and PHT in the HOT lane for carpoolers. Both figures provide VHT/PHT for each day in October, excluding weekends. The results in Figure 16 illustrate a fair reduction in VHT in both HOT and GP lanes, which indicates travelers in all lanes require less time to reach their destinations.

The decrease in travel time aligns with daily performance consistently, demonstrating the effectiveness and reliability of LQR based tolling strategy. Similarly, Figure 17 illustrates travel time advantages of tolling on different traveler types. Although the proposed tolling strategy is not explicitly designed for either solo drivers or carpoolers, the PHT reduction reveals a shared benefit for all travelers, despite their mode choices.

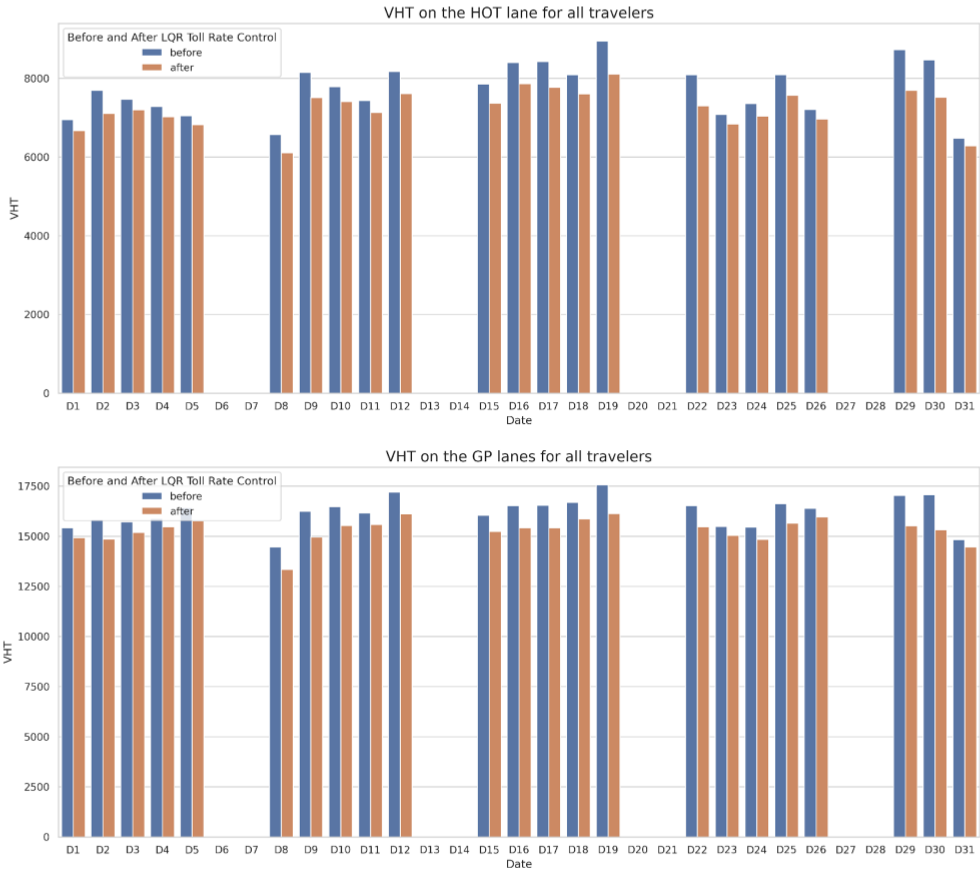

Figure 16. System performance on VHT

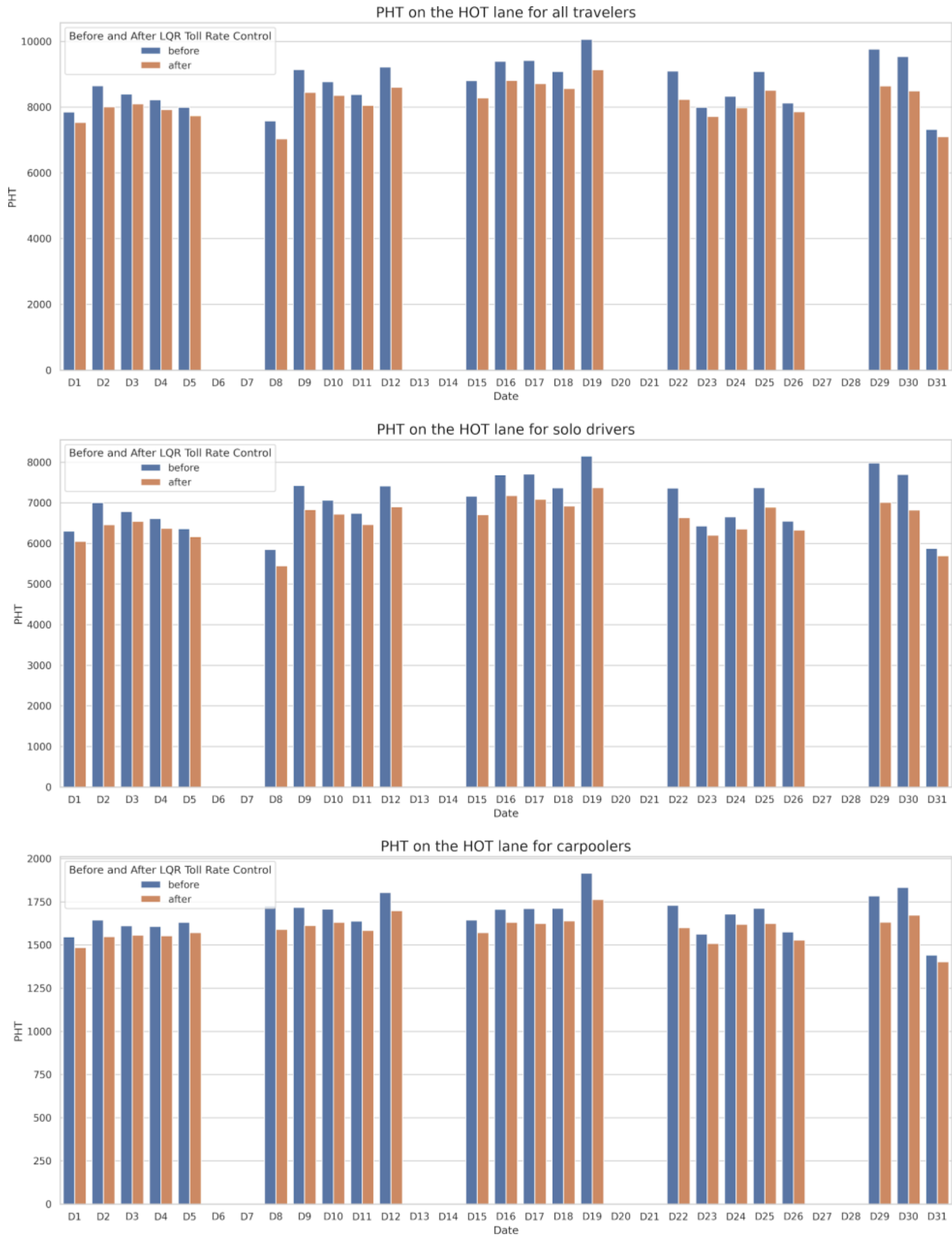

Figure 17. System performance on PHT

Table 2. System performance regarding VHT and PHT before and after implementation of the proposed tolling strategy

	VHT in HOT	VHT in GP	PHT in HOT	PHT for SOVs in HOT	PHT for HOVs in HOT
Tolling Before	177,794.82	372,720.79	200,251.82	161,598.69	38,653.13
Tolling After	166,550.65	352,196.73	187,855.69	151,194.15	36,661.53
Improvement	6.32%	5.51%	6.19%	6.44%	5.15%

Table 2 quantifies the system performance of VHT and PHT in GP and HOT lanes for solo drivers and carpoolers. Overall, the VHT and PHT reduction in the HOT lane is slightly higher than in the GP lane (6% versus 5%). Solo drivers traveling on the HOT lane also benefit more compared to carpoolers.

5 Conclusions

In this study, we developed an LSTM-based model to comprehensively understand and capture the intricate speed-toll patterns in HOT and GP lanes. This LSTM model serves as an effective evaluation tool, enabling us to assess the efficiency and performance of the proposed tolling strategies. By leveraging the temporal dependencies inherent in traffic data, the LSTM model provides a more nuanced understanding of the dynamic relationships between tolling policies and traffic speed. We propose a novel LQR based feedback control theoretical system to manage toll amounts in HOT lanes. We combined a lane choice model and traffic flow dynamics to formulate a multi-input, multi-output feedback control system. The system provides real-time control of an HOT lane corridor with multiple gantries and achieves the convergence to the optimal state, namely, critical density and speed.

To assess the efficacy of our proposed model, we use a dataset from the I-580 express lane as a case study, focusing specifically on the westbound direction featuring an HOT lane with 17 gantries. The exploratory data analysis reveals a non-linear correlation between SOV traffic flow and toll rates. This implies that the control of toll rates can significantly impact the traffic flow.

Based on the exploratory data analysis of the I-580 data, we reveal several findings by implementing our HOT lane management approaches:

- The LSTM-based traffic-speed prediction model is effective in capturing temporal patterns, navigating peak-hour variations, and accounting for lane-specific speed differences in real-world traffic scenarios. This establishes it as a robust tool for accurately predicting traffic speed in HOT and GP lanes.
- The tolling strategy directly affects the congestion level in the HOT lane. LQR tolling can change what would be light congestion in the HOT lane under current tolling methods to nearly congestion free. LQR tolling can also mitigate congestion in the GP lanes from a heavy/modest level to a light level.
- The LQR tolling strategy performs well on all weekdays in terms of decreasing VHT/PHT for carpoolers and solo drivers who choose the HOT or GP lane.

6 References

Bemporad, A., Morari, M., Dua, V., and Pistikopoulos, E. N. (2002). The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1):3–20.

Chen, H. and Grant-Muller, S. (2001). Use of sequential learning for short-term traffic flow forecasting.Transportation Research Part C: Emerging Technologies, 9(5):319–336.

Clegg, B. A., DiGirolamo, G. J., and Keele, S. W. (1998). Sequence learning. *Trends in cognitive sciences*, 2(8):275–281.

Cui, Z., Ke, R., Pu, Z., and Wang, Y. (2018). Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction. *arXiv preprint arXiv:1801.02143*.

Fu, R., Zhang, Z., and Li, L. (2016). Using lstm and gru neural network methods for traffic flow prediction. In *2016 31st Youth academic annual conference of Chinese association of automation (YAC)*, pages 324–328. IEEE.

Gu, Y., Lu, W., Qin, L., Li, M., and Shao, Z. (2019). Short-term prediction of lane-level traffic speeds: A fusion deep learning model. *Transportation research part C: emerging technologies*, 106:1–16.

Guo, K., Hu, Y., Qian, Z., Liu, H., Zhang, K., Sun, Y., Gao, J., and Yin, B. (2020). Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 22(2):1138–1149.

Janson, M. and Levinson, D. (2014). Hot or not: Driver elasticity to price on the mnpass hot lanes.Research in Transportation Economics, 44:21–32.

Li, B., Yang, Q., Chen, J., Yu, D., Wang, D., Wan, F., et al. (2023). A dynamic spatio-temporal deep learning model for lane-level traffic prediction. *Journal of Advanced Transportation*, 2023.

Li, J. (2001). Explaining high-occupancy-toll lane use. *Transportation Research Part D: Transport and Environment*, 6(1):61–74.

Liu, X., Zhang, G., Lao, Y., and Wang, Y. (2011). Quantifying the attractiveness of high-occupancy toll lanes with traffic sensor data under various traffic conditions. *Transportation research record*, 2229(1):102–109.

Lu, W., Rui, Y., and Ran, B. (2020). Lane-level traffic speed forecasting: a novel mixed deep learning model. *IEEE Transactions on Intelligent Transportation Systems*, 23(4):3601–3612.

Michalaka, D., Lou, Y., and Yin, Y. (2011). Proactive and robust dynamic pricing strategies for high-occupancy-toll (hot) lanes. Technical report.

Pandey, V., Wang, E., and Boyles, S. D. (2020). Deep reinforcement learning algorithm for dynamic pricing of express lanes with multiple access locations. *Transportation Research Part C: Emerging Technologies*, 119:102715.

Qi, Y., Li, Q., Karimian, H., and Liu, D. (2019). A hybrid model for spatiotemporal forecasting of pm2. 5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, 664:1–10.

Shen, L., Lu, J., Geng, D., and Deng, L. (2020). Peak traffic flow predictions: exploiting toll data from large expressway networks. *Sustainability*, 13(1):260.

Supernak, J., Golob, J. M., Kawada, K., and Golob, T. F. (1998). San diego's i-15 congestion pricing project: Preliminary findings.

Wang, X., Jin, W.-L., and Yin, Y. (2020). A control theoretic approach to simultaneously estimate average value of time and determine dynamic price for high-occupancy toll lanes. *IEEE Transactions on Intelligent Transportation Systems*, 22(11):7293–7305.

Yin, Y. and Lou, Y. (2009). Dynamic tolling strategies for managed lanes. *Journal of Transportation Engineering*, 135(2):45–52.

Zahedian, S., Nohekhan, A., and Sadabadi, K. F. (2021). Dynamic toll prediction using historical data on toll roads: Case study of the i-66 inner beltway. *Transportation Engineering*, 5:100084.

Zhang, G., Wang, Y., Wei, H., and Yi, P. (2008). A feedback-based dynamic tolling algorithm for high-occupancy toll lane operations. *Transportation Research Record*, 2065(1):54–63.

Zhang, Y. and Ioannou, P. A. (2016). Combined variable speed limit and lane change control for highway traffic. *IEEE Transactions on Intelligent Transportation Systems*, 18(7):1812–1823.

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., and Li, H. (2019). T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE transactions on intelligent transportation systems*, 21(9):3848–3858.

Zhao, Z., Chen, W., Wu, X., Chen, P. C., and Liu, J. (2017). Lstm network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2):68–75.

Zheng, C., Fan, X., Wang, C., and Qi, J. (2020). Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 1234–1241.