THE CURIOSITY CUP 2024

A Global SAS® Student Competition

Predicting EV Charging Demand: A Novel Approach Utilizing Traffic Flows

Team ORSOL

ABSTRACT

Despite the rapid growth in electric vehicle (EV) adoption and the corresponding need for extensive charging infrastructure, there remains a significant gap in the availability of granular, public data regarding the demand at individual EV charging stations. The scarcity of this data impedes reliable demand predictions for EV charging stations, leading to the risk of overcrowding that diminishes user experiences, and places heightened stress on power grids that amplify inefficiencies in energy distribution. Traditional forecasting approaches, while promising, struggle to address this issue effectively due to their high dependence on detailed, station-specific demand histories that are often unavailable. Hence, there is a pressing need for a novel approach that can accurately forecast EV charging station demand by leveraging alternative data sources and advanced analytical techniques. In response to this challenge, we propose an unprecedented approach that broadens the scope of analysis by leveraging predictions of traffic flow and trends along the I210 Corridor. Instead of relying on inaccessible direct EV charging station demand data, our method infers EV charging demand by estimating the number of EVs exiting the highway in need of charging and assessing demand at nearby charging stations. This approach uniquely leverages insights into seasonal variations and a deeper understanding of human driver behavior, moving beyond traditional deterministic models.

INTRODUCTION

The expedited adoption of EVs signals a significant shift towards sustainable transportation, necessitating a robust EV charging infrastructure. While the pace of EV adoption accelerates, the expansion and development of charging infrastructure struggle to keep up, highlighting a critical bottleneck in the transition to electric mobility. Simply adding more charging stations does not solve the problem in isolation; efficient energy distribution from power grids and the strategic placement are paramount to avoid bottlenecks and ensure sustainability. In addition, optimizing for reduced wait times and preventing congestion at charging stations has become a focal point of research, employing techniques like Real-Time Pricing (RTP) to manage demand efficiently and elicit optimal user experience. However, the development of such infrastructure faces a major challenge: the lack of detailed, public data on individual charging station demand.

The gap in this data not only hampers the strategic planning and optimization of charging infrastructure but also poses a risk of overcrowding at charging stations or exacerbated inefficiencies in energy distribution from power grids. Existing works on EV charging station demand forecasting [1], although have shown great promise, face a significant limitation in which these methods are heavily dependent upon detailed, station-specific usage histories. This reliance restricts their applicability in environments where such data are sparse or completely unavailable, undermining their effectiveness in predicting future demands accurately.

Recognizing the limitations of past forecasting approaches, this paper introduces a novel methodology that seeks to fill the void left by the lack of direct charging station demand data. By leveraging an alternative data source - the traffic flow and trends along the I210 corridor - we aim to propose a framework that incorporates the intricacies of traffic dynamics and EV driver behavior into EV charging station demand predictions. This framework utilizes advanced analytics and forecasting models - tested on the SAS Viya® for Learners (VFL) - to interpret the relationship between traffic patterns and charging station usage, allowing for the estimation of EV charging demands based on the volume and behavior of vehicles exiting the highway.

Data

DATASETS AND VALIDATION

The scope of our project focuses on evaluating the demand for EV chargers in California, the U.S. leader in EV adoption. Due to challenges in obtaining EV charging station usage data, influenced by privacy laws like California Consumer Privacy Act (CCPA) and the California Privacy Rights Act (CPRA), we shifted our analysis towards highway traffic data. Utilizing Caltrans PeMS data on busy highways near Los Angeles, we aimed to gauge EV charging demand. We zeroed in on a post mile range (PM range) of R32.50 to R36.33 for our study, believing traffic patterns here would reflect charging needs effectively.

DATA CLEANING AND PREPROCESSING

A comprehensive dataset was obtained from the California Department of Transportation Performance Measurement System (Caltrans PeMS). This dataset comprised 730 Excel files, each representing the traffic flow data for a single day. The structure of these files included 243 rows, delineating various PMranges, and 30 columns, featuring details such as sensor numbers, ramp types, and hourly traffic flow measurements. The initial phase of data preprocessing involved isolating Mainline data, focusing specifically on the range from PM R32.50 to R36.33. This procedure resulted in the identification of 8 distinct sensor data points. Subsequent to this filtration process, each preprocessed file was merged into a singular table formatted for time series analysis.

The final data table comprised 14,060 rows, calculated on the basis of 730 days multiplied by 24 hours and further multiplied by 8 sensors, and was organized into 4 columns: sensor identification, date/time of data collection, traffic flow, and holiday indicator. To ensure compatibility and synchronization with SAS Programming Studio, the date/time column was formatted to the specific notation of DDMMYY:H:MM:SS. Our final cleaned and processed data is shown in Table 1, where we would use this time series prepared data for our following forecasting tests

Sensor	Date_Time	Traffic_Volume	Holiday
773193	24APR18:22:00:00	3351	0
773193	24APR18:23:00:00	2182	0
773193	25APR18:00:00:00	1247	0
773193	25APR18:01:00:00	937	0
773193	25APR18:02:00:00	753	0
773193	25APR18:03:00:00	793	0
773193	25APR18:04:00:00	1314	0

Table 1: Hourly Traffic Volume Data by Vehicle Detection Sensor Along I210 Corridor (PM R32.50 - R36.33)

Methodology

SAS FORECAST STUDIO PIPELINE (TESTED 6 MODELS)

For this study, the SAS Forecast Studio was employed to develop and test various forecasting models. We selected the following models to test as shown in Figure 1. Each model was evaluated based on its forecasting accuracy, computational efficiency, and interpretability to ensure practical applicability in real-world traffic management systems.

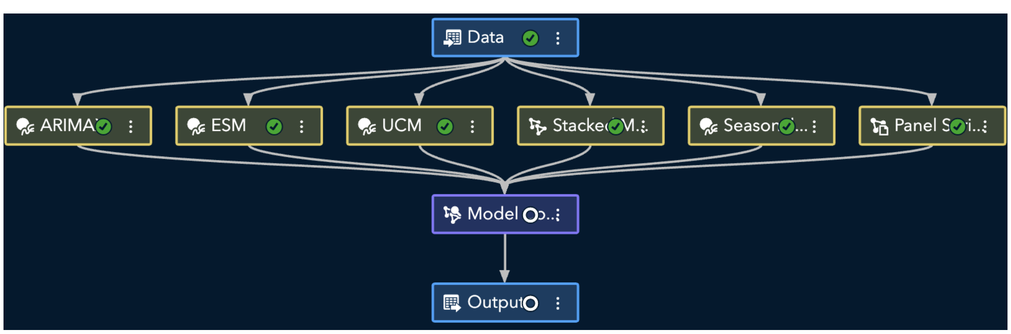


Figure 1: Forecast Pipeline (SAS VFL - Model Studio)

MODEL SELECTION JUSTIFICATION

A) Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX)
It was chosen due to its capacity to incorporate the impact of holidays, weather conditions, and other external events on traffic flow.

B) Unobserved Components Model (UCM)
UCM decomposes the time series into components such as trend, seasonality, and cycles. It was chosen to explicitly model and estimate these components, which are often present in traffic data.

C) Seasonal Model
Given the known daily and weekly patterns in traffic flow, seasonal models were tested to capture and utilize this recurring variability in the data.

D) Exponential Smoothing Model (ESM)
Known for its flexibility in modeling data with trends and seasonality, ESM was selected to adaptively forecast traffic flow based on recent patterns.

E) Stacked Model (Neural Network + CS)
This approach combines neural networks with conventional statistical (CS) models to leverage the strengths of both: the pattern recognition capabilities of neural networks and the explanatory power of statistical models.

F) Panel Series Neural Network
This type of model is well-suited for datasets with cross-sectional time-series data, which is typical in traffic data collected across various sensors or locations. Neural networks are capable of capturing complex nonlinear relationships in such data.

ESTIMATING NUMBER OF EV EXITING HIGHWAY THAT NEEDS TO CHARGE

This section of our study is crafted through an elaborate three-step methodology, meticulously designed to ascertain the demand for electric vehicle (EV) charging infrastructure in California. Our strategy leverages the principles of analytical methods prevalent in data science.

A) Proportion of Registered Electric Vehicles (EVs) on California Highways
To initiate, we compute the Probabilistic Ratio of EVs to the total vehicle population in California, utilizing data from the Department of Energy. This ratio is crucial for estimating the volume of electric vehicles traversing our specified PM range within the I210 Corridor.

$$R_{EV\ on\ Highway} = \frac{Total\ EV\ Registrations\ in\ California}{Total\ Vehicle\ Registrations\ in\ California} \quad (1)$$

B) Proportion of Vehicle leaving California Highways
Our analysis refines the vehicle exit rate from the highway by calculating the traffic flow differential between the initial and final Vehicle Detection Systems, adjusted for exit numbers and a correction factor. This provides a precise vehicular outflow estimate, vital for understanding the I210 Corridor's traffic and infrastructure requirements.

$$R_{Vehicle\ exiting\ highway} = \frac{VDS\ 1\ Traffic\ Flow - VDS\ 8\ Traffic\ Flow}{Number\ of\ Exit} * \frac{1}{VDS\ 1\ Traffic\ Flow} \quad (2)$$

C) Proportion of Electric Vehicles (EVs) needs a charge when leaving California Highways
Equation (3) refines our methodology to forecast EV charging station demand at key highway exits, incorporating assumptions due to privacy and data access constraints. By analyzing Traffic Count from Traffic Flow and Duration, and adjusting for EV proportions and exit rates, we account for the subset of EVs requiring immediate charging. This approach, tailored to meet privacy and legal requirements, delivers a precise and compliant assessment of charging infrastructure needs at critical locations.

$$Traffic\ Count = Traffic\ Flow\ (veh/hr) * Duration\ (hr)$$

$$R_{final} = Traffic\ Count * R_{EV\ on\ Highway} * R_{Vehicle\ exiting\ highway} * R_{EV\ that\ needs\ charge\ at\ exit} \quad (3)$$

$* R$: $Ratio\ in\ decimal$

Results

Results Data Visualization Process - 04/24/2018

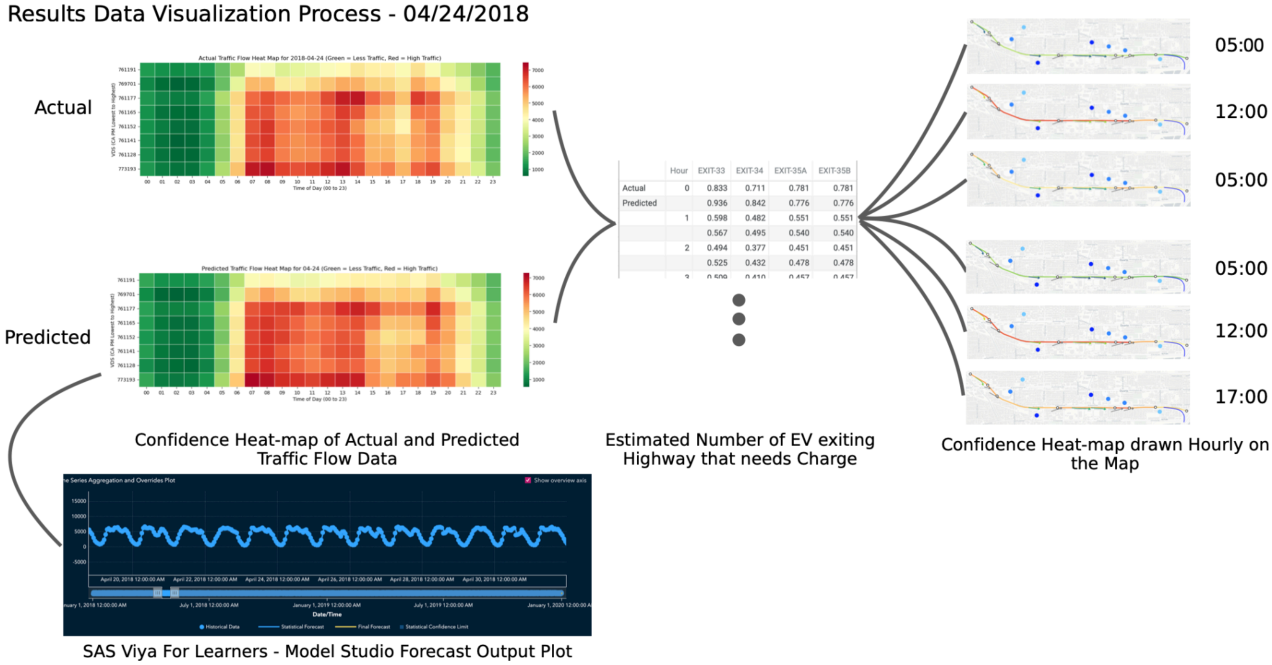


Figure 2: Data Visualization Process and Framework (APPENDIX contains enlarged map data and table)

Figure 2 serves as a comprehensive summary of our research findings, effectively illustrating the outcomes of our forecasting model through a series of heat maps that we created based on the forecasted values obtained from our SAS Model Studio Forecasting Pipeline, a detailed table and geographical maps to easily visualize the traffic flow. At the left of our framework are the two distinct confidence heat maps, one representing the actual traffic flow dataset and the other depicting the predicted traffic flow. These visual representations allow us to understand the discrepancies and alignments between anticipated and real-world traffic patterns. Central to the image is a table that encapsulates the extended analysis derived from the initial traffic flow datasets. This table calculates the appropriate "traffic count" using Equations 3, and extends this analysis to estimate the number of EVs exiting the highway in need of charging. At the right of our image includes columns of a confidence heat map - which represents an hour along the PM range we investigated - on to the geographical map along the I210 corridor, providing vivid examples of traffic flow at different times of the day.

Data Analyses

Champion	Model Name	Status	WMAE	WMAPE
★	Stacked Model (NN + TS)	Successful	308.7141	10.1561
	Panel Series Neural Network	Successful	379.8618	13.8365
	ESM	Successful	321.7428	11.2041
	Seasonal Model	Successful	334.5578	11.7876

Reported WMAPE for Champion Model = 10.1561%

Figure 3: Champion Model - Stacked (NN + TS) with reported WMAE and WMAPE (SAS VFL - Model Studio)

The champion model we selected for our forecasting tasks is the Stacked Model (NN + TS), which synergizes neural networks and time series analysis, as shown in Figure 3. The Stacked Model (NN + TS) leverages the neural networks' capacity to decipher complex nonlinear relationships and the time series analysis's adeptness at navigating temporal dynamics. The accuracy of this model is quantitatively gauged using the Weighted Mean Absolute Error (WMAE) and Weighted Mean Absolute Percentage Error (WMAPE) metrics. Our forecasting model demonstrates a high level of accuracy, as evidenced by a WMAPE of approximately 10%.

This indicates that, on average, the model's predictions deviate from actual values by only 10%, a strong indicator of its reliability in forecasting demand.

Figure 4 illustrates a segment of our final forecasting data over the Vehicle Detection Sensor (VDS) 761128, enabling detailed analysis of the model's performance across a specific timeframe. The examination of this segment validates the model's accuracy by aligning predicted values closely with the actual traffic data, with errors quantified on the graph. Notably, the model captures the inherent seasonality of EV driver behavior, with discernible peaks corresponding to key daily events. These peaks likely represent times when drivers are commuting to work, taking lunch breaks, and leaving work, reflecting predictable daily routines as shown in the reference lines in Figure 4. The graph's visual cues—such as the overlapping of the actual and predicted lines—reinforce the model's capability to anticipate fluctuations in the traffic flow along the investigated I210 corridor.

Our model captures and forecasts the dynamic patterns of traffic flow, which is crucial for estimating the number of electric vehicles (EVs) exiting the highway in need of charging. This capability is integral for deducing the demand at individual EV charging stations. The robustness of our model is further demonstrated by the map-integrated traffic flow figure presented in Appendix Figure 3. This figure showcases the geographical distribution of traffic and charging demand, providing a comprehensive overview that can inform infrastructure development and resource allocation for EV charging networks. The model's effectiveness in reflecting the temporal and spatial variations of traffic underlines its significance as a predictive tool for urban planning and energy management.

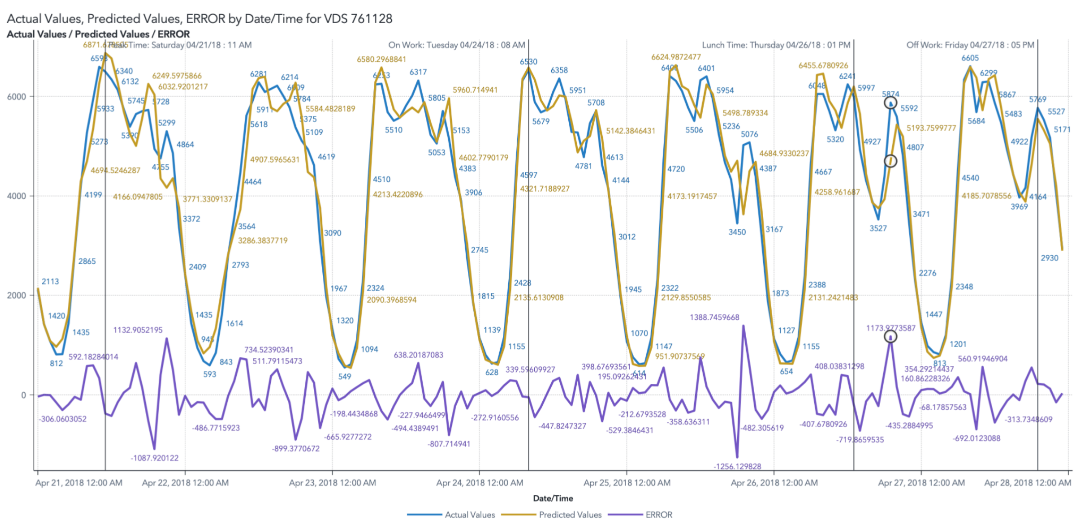


Figure 3: Visualization of VDS 761128 Data - 21 April- 27 April, 2018 - Actual/Predicted/Error Value (SAS VFL - Visual Analytics Explore and Visualize)

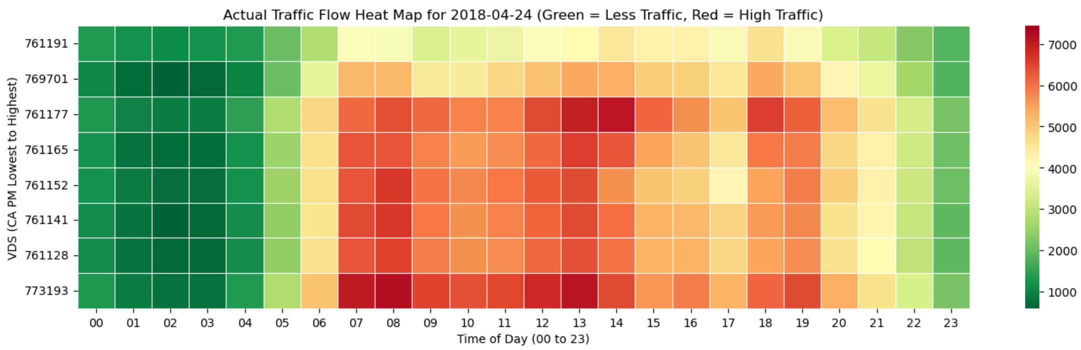
Conclusion

In light of the challenges presented by the rapid adoption of EVs, our study has successfully developed an innovative forecasting model that circumvents the data scarcity problem by utilizing traffic flow as a surrogate dataset of EV charging demand. The model's ability to predict traffic patterns with a WMAPE of 10.145% underscores its effectiveness and the constructed visual representation framework allows a deeper understanding of our findings. By intelligently estimating the demand for EV charging stations based on the traffic flow of the highway and exits, our approach opens avenues of new research directions towards fellow researchers who are tackling this same issue concurrently facing difficulty finding access to publicly accessible and detail-oriented charging station demand data. Currently, we estimate the number of EVs requiring charging based on assumptions and therefore in our future work, we aim to incorporate real-time data to improve the accuracy and reliability of our model, better predicting the individual charging station demands.

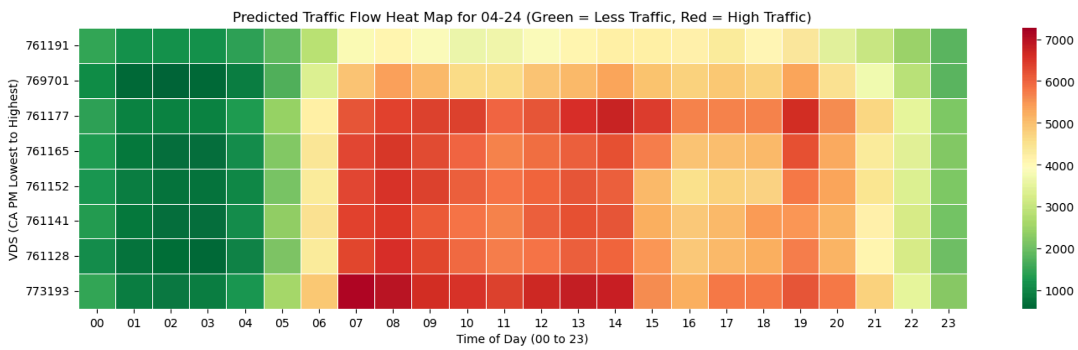
References

[1] Almaghrebi, Ahmad, et al. "Data-driven charging demand prediction at public charging stations using supervised machine learning regression methods." *Energies* 13.16 (2020): 4231.

Appendix



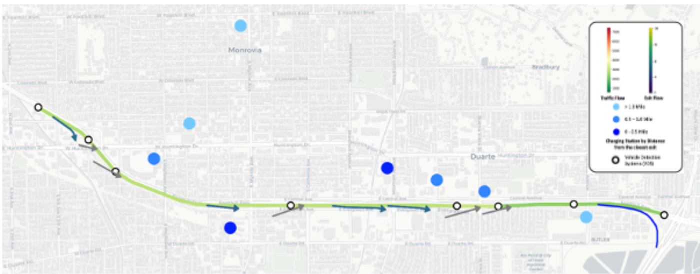
Appendix Figure 1: Actual Traffic Flow Heat Map for 04-24-2018

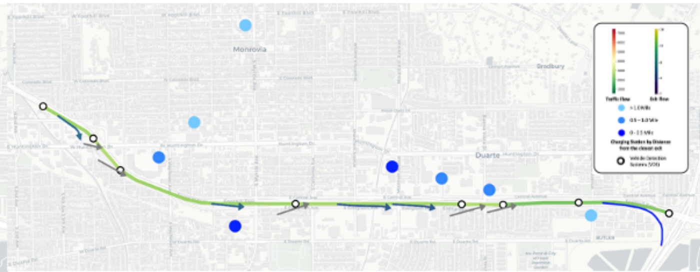


Appendix Figure 2: Predicted Traffic Flow Heat Map for 04-24-2018

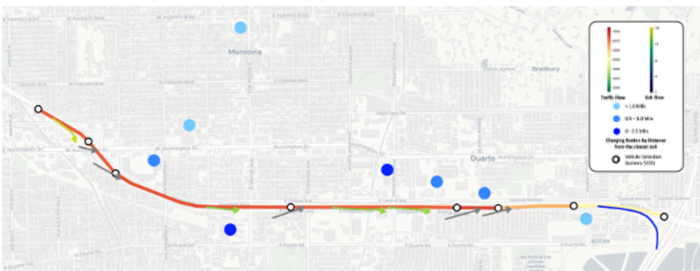
	Hour	EXIT-33	EXIT-34	EXIT-35A	EXIT-35B
Actual	0	0.833	0.711	0.781	0.781
Predicted		0.936	0.842	0.776	0.776
	1	0.598	0.482	0.551	0.551
		0.567	0.495	0.540	0.540
	2	0.494	0.377	0.451	0.451
		0.525	0.432	0.478	0.478
	3	0.509	0.410	0.457	0.457
		0.531	0.433	0.455	0.455
	4	0.849	0.727	0.765	0.765
		0.791	0.694	0.652	0.652
	5	1.744	1.521	1.622	1.622
		1.625	1.489	1.326	1.326
	6	3.219	2.871	2.931	2.931
		3.078	2.834	2.723	2.723
	7	4.425	4.035	3.981	3.981
		4.482	4.010	3.967	3.967
	8	4.528	4.184	4.187	4.187
		4.368	4.077	4.101	4.101
	9	4.111	3.723	3.764	3.764
		4.155	3.860	4.020	4.020
	10	4.007	3.587	3.638	3.638
		4.128	3.663	3.812	3.812
	11	4.092	3.663	3.738	3.738
		4.033	3.578	3.658	3.658
	12	4.301	3.868	3.930	3.930
		4.207	3.799	3.773	3.773
	13	4.476	4.045	4.051	4.051
		4.296	3.922	3.869	3.869
	14	4.060	3.780	3.588	3.588
		4.281	3.889	3.790	3.790
	15	3.550	3.336	3.212	3.212
		3.520	3.291	3.212	3.212
	16	3.702	3.285	3.067	3.067
		3.303	3.058	2.871	2.871
	17	3.321	3.043	2.626	2.626
		3.628	3.175	2.982	2.982
	18	3.855	3.500	3.441	3.441
		3.627	3.405	2.971	2.971
	19	4.042	3.632	3.687	3.687
		3.873	3.468	3.634	3.634
	20	3.350	2.948	3.104	3.104
		3.647	3.254	3.340	3.340
	21	2.930	2.649	2.712	2.712
		2.981	2.635	2.802	2.802
	22	2.111	1.913	1.993	1.993
		2.210	2.016	2.125	2.125
	23	1.375	1.225	1.300	1.300
		1.431	1.299	1.361	1.361

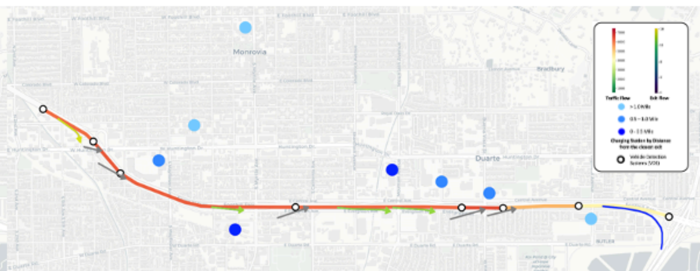
Appendix Table 1: Predicted Traffic Flow Heat Map for 04-24-2018



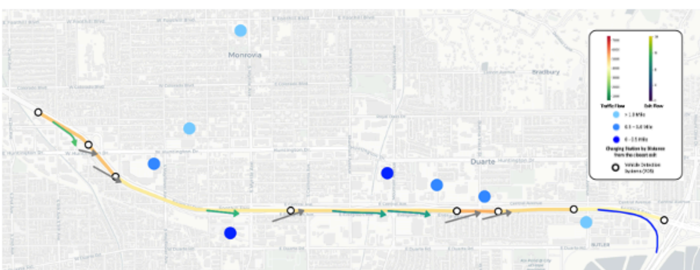


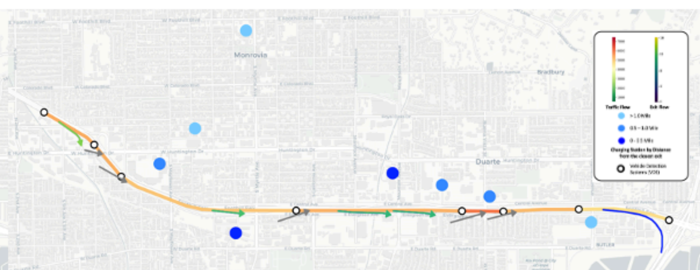
Actual Traffic Flow 05:00 **Predicted Traffic Flow 05:00**





Actual Traffic Flow 12:00 **Actual Traffic Flow 12:00**





Actual Traffic Flow 17:00 **Actual Traffic Flow 17:00**

Appendix Figure 3: Actual and Predicted Traffic Flow Comparison

https://youtu.be/HIgMC8Emtwk

Appendix Video 1: Actual & Predicted Heat Map Comparison

Description: This video shows the heat map comparison with actual and predicted data from 04/21/18 to 04/27/18

https://youtu.be/EYl2CKICv74

Appendix Video 2: Actual Traffic Flow Map

Description: This video shows the hourly I210 highway traffic flow on the map with actual data from 04/21/18 to 04/27/18

https://youtu.be/m5VfBFK6TCo

Appendix Video 3: Predicted Traffic Flow Map

Description: This video shows the hourly I210 highway traffic flow on the map with predicted data from 04/21/18 to 04/27/18