

Arterial Traffic Flow Prediction: A Deep Learning Approach with Embedded Signal Phasing Information

*Victor Chan
Qijian Gan
Alexandre Bayen*

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2020-68

<http://www2.eecs.berkeley.edu/Pubs/TechRpts/2020/EECS-2020-68.html>

May 27, 2020



Copyright © 2020, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**Arterial Traffic Flow Prediction: A Deep Learning Approach with
Embedded Signal Phasing Information**

by Victor Chan

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences,
University of California at Berkeley, in partial satisfaction of the requirements for the
degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

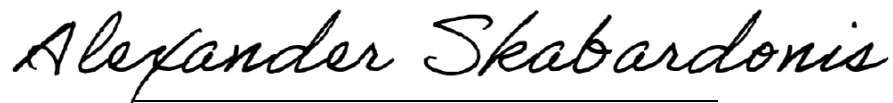


Professor Alexandre Bayen
Research Advisor

5/22/2020

(Date)

* * * * *



Professor Alexander Skabardonis
Second Reader

5/22/2020

(Date)

Abstract

Arterial Traffic Flow Prediction: A Deep Learning Approach with Embedded Signal Phasing

Information

by

Victor Chan

Master of Science in Electrical Engineering and Computer Science

University of California, Berkeley

Professor Alexandre Bayen

Accurate and reliable prediction of traffic measurements plays a crucial role in the development of modern intelligent transportation systems. Due to more complex road geometries and the presence of signal control, arterial traffic prediction is a level above freeway traffic prediction. Many existing studies on arterial traffic prediction only consider temporal measurements of flow and occupancy from loop sensors and neglect the rich spatial relationships between upstream and downstream detectors. As a result, they often suffer large prediction errors, especially for long horizons. We fill this gap by enhancing a deep learning approach, Diffusion Convolutional Recurrent Neural Network, with spatial information generated from signal timing plans at targeted intersections. Traffic at signalized intersections is modeled as a diffusion process with a transition matrix constructed from the green times of the signal phase timing plan. We apply this novel method to predict traffic flow from loop sensor measurements and signal timing plans at an arterial intersection in Arcadia, CA. We demonstrate that our proposed method yields superior forecasts; for a prediction horizon of 30 minutes, we cut the MAPE down to 16% for morning peaks, 10% for off peaks, and even 8% for afternoon peaks. In addition, we exemplify the robustness of our model through a number of experiments with various settings in detector coverage, detector type, and data quality.

ARTERIAL TRAFFIC FLOW PREDICTION: A DEEP LEARNING APPROACH WITH EMBEDDED SIGNAL PHASING INFORMATION

Victor Chan, Qijian Gan, Alexandre Bayen
Department of Electrical Engineering and Computer Science
University of California, Berkeley
{victorchan, qgan, bayen}@berkeley.edu

May 27, 2020

ABSTRACT

Accurate and reliable prediction of traffic measurements plays a crucial role in the development of modern intelligent transportation systems. Due to more complex road geometries and the presence of signal control, arterial traffic prediction is a level above freeway traffic prediction. Many existing studies on arterial traffic prediction only consider temporal measurements of flow and occupancy from loop sensors and neglect the rich spatial relationships between upstream and downstream detectors. As a result, they often suffer large prediction errors, especially for long horizons. We fill this gap by enhancing a deep learning approach, Diffusion Convolutional Recurrent Neural Network, with spatial information generated from signal timing plans at targeted intersections. Traffic at signalized intersections is modeled as a diffusion process with a transition matrix constructed from the green times of the signal phase timing plan. We apply this novel method to predict traffic flow from loop sensor measurements and signal timing plans at an arterial intersection in Arcadia, CA. We demonstrate that our proposed method yields superior forecasts; for a prediction horizon of 30 minutes, we cut the MAPE down to 16% for morning peaks, 10% for off peaks, and even 8% for afternoon peaks. In addition, we exemplify the robustness of our model through a number of experiments with various settings in detector coverage, detector type, and data quality.

Keywords Traffic prediction · Arterial intersections · Signal phase timing data · DCRNN

1 Introduction

The problem of efficient transportation has typically been a hardware and civil engineering problem, as companies have developed faster and cleaner cars, built carefully-designed freeways, and architected roads in cities. With the rise of intelligent transportation systems (ITS), the problem has shifted focus to the fields of mathematics, statistics, and computer science. As governments install more sensors in road networks and collect ever-increasing amounts of data, research has begun to concentrate on designing improved prediction and control techniques. Traffic flow and speed prediction has numerous applications, such as freeway metering, travel time prediction, intelligent intersection signal control, and traffic simulation software. With accurate traffic flow forecasts, cities can better plan logistics and allocation of resources for construction, road development, and safety. Predictions can also be leveraged to optimize signal control at intersections, saving commuters valuable time and reducing consumption of gas and electricity.

Historically, there have been a wide variety of models in use for traffic flow prediction. Although they are often grouped into parametric and nonparametric categories, or classified as statistics or machine learning, most models are closely-related and have overlapping properties [24]. Statistical methods have been applied heavily to traffic flow prediction. Ahmed and Cook used the Box-Jenkins method to fit an ARIMA model to freeway occupancy in 1979 [1]. Since then, many other techniques in the ARMA family have been applied to predict both freeway and arterial data [22, 45, 46, 47]. Kalman filters [36] and exponential smoothing [44] have been actively explored for decades. Typically, they rely on strong priors and assumptions about the data distributions; as a result, traffic experts usually must carefully

select and structure the models. Because of this, parametric methods present a trade-off between easy interpretation and practicality [40].

Although they used to have poor performance in traffic flow prediction, nonparametric methods have enjoyed a surge in popularity as hardware has been upgraded and more powerful algorithms have been developed. They are typically more flexible and accurate than comparable parametric models. Nearest neighbors regression lies in the realm of nonparametric learning and begins to cross over into the machine learning category [6, 54, 12]. Other nonparametric methods are also prevalent in the literature: principal component analysis [10], decision trees [2], support vector machines [16, 4], linear dynamical systems [3], and fuzzy rule-based systems [13, 25].

Even these machine learning methods have been overshadowed by the rise of neural networks, of which there are countless architectures to choose from. Deep learning has gained much traction in recent years as data has become more readily available and computer power has exponentiated. Simple feed-forward neural networks have evolved into convolutional neural networks, long short-term memory, and graph convolutions. State-of-the-art algorithms utilize strategies such as meta-learning and distillation [49, 37], residual connections [56], attention [50], and adversarial training [57, 58]. Deep architectures open the door for a new generation of nonparametric models that are constantly improving prediction accuracy.

Overall, most prediction methods are very proficient at forecasting freeway data. Freeways are a mostly-closed system, with leakages only from on-ramps and off-ramps. Traffic often flows smoothly from one sensor to the next with few interruptions, and even congestion is predictable. Thus, freeway traffic data is typically smooth and clean. In contrast, arterial traffic is much noisier and more difficult to predict. At intersections, traffic signals and stop signs introduce exogenous factors that affect the speed and movement of cars. Moreover, elements such as pedestrians, bikes, parking, and driveways further complicate traffic patterns. While there is a lot of existing literature that focuses on freeway traffic prediction, less work explores the same topic for arterial traffic.

One strategy that has proved useful in overall traffic flow prediction is the graph convolution, which applies to the setting of predicting a label for a graph, given a set of graphs with their associated labels [5]. In the most general case, the graphs are directed and weighted, and the labels can be associated with any part of the graph, including the nodes, edges, and the graphs themselves. The spatial information from graph convolutions is significant in arterial traffic flow prediction because the detector graph is much more complicated than that of freeways. Graph convolutions have spatial structure built into the architecture, so they naturally account for the spatial relationships between detectors when predicting traffic flow.

Another consideration is the inclusion of different types of data as input for prediction. Most models treat the data as a time series, thus relying only on historical values of the data to forecast future values. Sometimes, extra features such as date, time, day of week, and exogenous events are included [3]. We employ signal phase timing data from the traffic signals at our study site. Previously, signal phase timing data has been combined with detailed traffic knowledge to develop a system of equations to predict traffic flow [41]. However, the model only applied to very short-term predictions, as it was intended for real-time signal control.

In this study, we focus on the Diffusion Convolutional Recurrent Neural Network (DCRNN) [28], which implements a graph convolution as a diffusion process on a network of traffic detectors with a diffusion convolution operation. The diffusion convolution is integrated with a seq2seq architecture of Gated Recurrent Units to predict future traffic flow. We apply DCRNN to predict arterial traffic flow for detectors with full coverage. In order to adapt DCRNN to arterial traffic, we use novel signal phase timing data to construct the weighted transition matrix of the graph. Instead of modeling transition probabilities with road distances, which are not suitable for intersections, we calculate the phase green time fraction from the planned green time and cycle length. We demonstrate that using signal phase timing information reduces prediction error, especially for long horizon predictions. Moreover, we find through many ablation studies that the model does indeed learn the relationships between the detectors in the network.¹

The rest of this thesis is organized as follows. In section 2, we summarize current literature on traffic flow prediction, especially in deep learning. In section 3, we present the model we use and our strategy to append signal phase timing data. In section 4, we introduce the study site and dataset used in this report. In section 5, we analyze our arterial traffic flow forecasts and evaluate their effectiveness through many ablation studies. In section 6, we draw conclusions based on our analysis.

¹Code available at https://github.com/victorchan314/arterial_traffic_flow_predictor

2 Related Work

2.1 ARMA Models

As mentioned in the introduction, numerous statistical methods for traffic flow prediction have been investigated. However, we focus on models in the ARMA family, which have seen much success in general time series prediction. Although not the first, Ahmed and Cook applied an ARIMA(0, 1, 3) model to forecast freeway occupancy and flow in 1979 [1]. Hamed, Al-Masaeid, and Said extended the model to predict arterial flow [22]. Williams and Hoel showed that a weekly seasonal difference could make freeway flow stationary, thus cementing the theoretical justification for fitting ARMA models to traffic data [46]. The field has been further expanded by the application of exogenous data to standard ARIMA models [45, 47].

2.2 Deep Learning

Recurrent Neural Networks Because traffic data is a time series, it makes sense to apply recurrent neural networks (RNN) to the prediction problem to learn temporal patterns. Due to the vanishing gradient problem, Ma *et al.* applied the Long Short-Term Memory (LSTM) architecture to demonstrate that larger window sizes are important [33]. Tian and Pan take advantage of the recurrent connections of LSTM to allow variable-length inputs [42]. Fu, Zhang, and Li achieved higher prediction accuracy using GRU and LSTM compared to ARIMA models [17]. Zhao *et al.* use an origin-destination correlation matrix, which captures the correlation between different observation points in the detector network, as input to LSTM [60]. The above methods all deal with freeway data; in contrast, Mackenzie, Roddick, and Zito use a sparse distributed representation of data with binary vectors and show that it is comparable to LSTM for arterial flow prediction [34].

Recurrent Convolutional Neural Networks In order to use the spatial information encoded within the data, RNN methods simply include data from multiple sensors and rely on RNN to extract that information, which is not very effective. To directly access spatial information, models evolved to synthesize RNNs with convolutional neural networks (CNN). Yu *et al.* splits an image of a road network into 10-meter pixels with the value set to the average speed of the links in that pixel; the output of a 2D convolution is fed into an LSTM to learn temporal relationships [53]. Du *et al.* extends this further by merging the CNN and RNN features and feeding them into a dense layer for a final prediction [14]. Yao *et al.* take a different approach by using start and end flow values for two channels of the grid [50]. The output is modulated with a learned Flow Gate Mechanism before being fed into an LSTM. In addition, an attention mechanism is used to derive the periodic trends in the data; these features are merged with the LSTM outputs in a dense layer to produce the final forecasts.

Graph Convolutions While CNNs consider spatial relationships in a more proper way, they still bear an inevitable mismatch with traffic data. Road networks are inherently graphs and not grids—they are not accurately represented as images. To this end, we turn to the graph convolution, which is perfect for learning traffic data. The graph structure is explicitly baked into the architecture of Graph Convolutional Networks (GCN) instead of being implicitly included with the data or imprecisely approximated with images.

Atwood and Towsley defined the Diffusion Convolutional Neural Network, which uses the power series of the degree-normalized transition matrix to model a diffusion process; DCNN outputs high-quality results for citation graph datasets [5]. Ruiz, Gama, and Ribeiro designed the Gated Graph Convolutional Recurrent Neural Network (GGCRNN), which stores banks of linear transforms modulated by filter taps in order to keep the number of parameters independent of the size of the graph [39]. In addition, gating reminiscent of the forget gate in LSTMs updates the hidden state from the previous hidden state and the input. Chen *et al.* use attention twice: once to aggregate all hops of the graph convolution into a hierarchical representation for each node, and another time to pool the node representations into an embedding matrix to be used for classification [8].

Graph convolutions have been applied to freeway data as well. Li *et al.* adapted the DCNN with a seq2seq GRU architecture to create the DCRNN [28]. Yu, Yin, and Zhu use a first order Chebyshev polynomial approximation of the filters for efficiency [51]. This allows them to sandwich a spatial graph convolution bottleneck between two blocks that apply a convolution and a Gated Linear Unit for temporal features. Zhao *et al.* use a graph convolution to extract features of road networks that are fed into a GRU for highway speed predictions [59]. Wu *et al.* add an adaptive layer to the forward and backward processes in the DCRNN [48]. They then apply the WaveNet dilated convolution with gating to learn temporal relationships. Yu, Yin, and Zhu pool traffic features with the maximum weight matching algorithm and compress the temporal dimension with dilated skip connections [52]. The embeddings are expanded back out in a U-Net fashion and fed into downstream architectures. Chen *et al.* combine graph convolutions, GRUs, skip connections, gates, and hop-links to forecast freeway flow [7]. Two recent works incorporate attention into GCNs. Guo *et al.* stack

a temporal attention mechanism, a spatial graph convolution, and a temporal standard convolution into a block [21]. They then accept recent, daily-periodic, and weekly-periodic data points as inputs to forecast freeway flow. In contrast, Fang *et al.* concatenate traffic data along the time axis before applying multiple tensor convolutions, a localized graph convolution, and a global spatial correlation [15]. Fewer works take advantage of GCNs to forecast arterial data. Cui *et al.* use the graph convolution with an adjacency matrix of nodes in a k -hop neighborhood to extract features of the graph before feeding them into an LSTM [11]. The forget gate of the LSTM is also a graph convolution so as to retain the important structure of the hidden state. Guo *et al.* optimize the Laplace matrix in Graph Convolution Gated Recurrent Unit cells and show that the learned matrices have high correlation with physical proximity [20].

Other Deep Learning Methods It is worth noting that there are also many other deep learning methods that have shown promise in traffic flow prediction, even though they do not always leverage the graph convolution. In 2005, Vlahogianni, Karlaftis, and Golias trained a multilayer perceptron (MLP) with a genetic algorithm for step size, momentum, and hidden units to predict arterial flow in Athens [43]. Kumar, Parida, and Katiyar show decent results with just a vanilla MLP and vehicle classification, speed, day of week, and time of day features in addition to flow [26]. Fusco *et al.* compare implicit models, such as a SARIMA model with points conditioned by a Bayes Network (SARIMA-BN), to explicit traffic models constructed by simulators [18]. Lv *et al.* train a reconstruction loss with a KL divergence term in a stacked autoencoder to predict freeway flow [30]. Polson and Sokolov encourage sparseness with a LASSO penalty on a vector autoregression model, with ridge regression and dropout regularization [38]. Zhang, Zheng, and Qi feed the output of three ResNets, each accepting an input of different seasonality to learn periodic trends, into a linear layer to measure crowd inflow and outflow for pedestrians and taxicabs [56]. Ma *et al.* encode average speed of cars into images where the width and height represent time and space [32]. They then use a standard CNN to predict speed with high accuracy. CapsNet extends this work by utilizing capsules in the CNN to overcome complex link geometries and low resolutions [31]. The output of the CapsNet is fed into a Nested LSTM to capture long-term spatio-temporal patterns. In order to reap the benefits from multiple models, Zhan *et al.* use a consensus ensemble system with ARMAX, Partial Least Squares, support vector regression, kernel ridge regression, and Gaussian process regression to prune outlier predictions [55]. They illustrate that the different models all bring unique advantages to the table when predicting arterial flow.

The most recent works have incorporated elements of deep reinforcement learning and unsupervised learning into arterial traffic flow prediction. Yao *et al.* train MetaST on datasets from many different cities in order to learn the overall distribution of arterial volume [49]. Each city is divided into grid squares that are associated with a global cluster through attention. An ST-Mem block maps the cluster to a feature vector that is concatenated to the output of an ST-Net. In order to predict flow for a new city, the parameters for the ST-Net and ST-Mem are initialized from the global models before being finetuned with city data. Pan *et al.* take a different approach with meta-learning; they use a seq2seq architecture, where the encoder and decoders are composed of the same components: an RNN to extract features of flow and speed data, node and edge meta-knowledge learners (NMK and EMK learners) to convert graph characteristics to feature vectors, a Meta Graph Attention Network (Meta-GAT) to learn relationships between different elements of the graph, and a Meta-GRU that generates weights from the NMK learner to discover temporal correlations [37]. They show that nodes that are closer in the network also have high correlation in the embeddings in the NMK learner. He, Chow, and Zhang turn to a seq2seq architecture with LSTMs that feed into attention layers [23]. The output of these layers is included in the LSTM gates to learn important long-term features. In order to deal with blurry features that causes graph convolutional networks to learn only the general trend and not peaks and valleys, Zhang *et al.* devise two models to take advantage of the "mode collapse" problem of the Generative Adversarial Network (GAN) [57, 58]. GCGAN uses a seq2seq architecture for both the generator and discriminator, combining LSTMs with bi-directional random walk with restart GCNs to reproduce and classify traffic network matrices. TrafficGAN constructs a generator with a CNN to learn spatial features, LSTM for temporal features, and another CNN to generate new traffic network matrices. They use a deformable convolution kernel to learn the shape of the filters, determined by relationships between nodes of the graph.

Newer deep learning methods can combine GCNs with deep unsupervised learning to learn feature representations of traffic. These methods provide flexible and expressive models that, if designed and trained properly, can easily outperform parametric and statistical methods. The additional parameters of these models also provide a way to incorporate extra data, such as signal phase timing information. There is still much to be explored and much room for improvement, especially with arterial traffic prediction.

3 Method

Most conventional traffic prediction methods only exploit temporal information to predict future information. When time series data is involved, the layout of sensors is typically ignored in order to generate predictions. Oftentimes,

methods rely on the underlying architecture to detect any spatial relations that could be useful for prediction. However, spatial information is not directly built into the architecture.

Recently, a new method, DCRNN [28], has been proposed to directly integrate spatial information, such as sensor layouts, into the architecture. In particular, it models freeway traffic between sensors as a diffusion process, with upstream sensors sending cars to downstream sensors. The parameters for the transition matrix rely on physical properties of the road network, such as road distances between sensors, unlike other graph convolutions that use binary weights or learn the weights; thus, we are able to insert our signal phase timing data into the model. This spatio-temporal property of DCRNN has also been utilized in other applications and fields: travel time estimation [27], ride-hailing demand [19], air quality forecasting [29], and distributed fleet control in reinforcement learning [35].

Different from the aforementioned studies, we apply DCRNN to arterial traffic prediction. We are one of the first studies to do so; moreover, we are the first to use signal phase timing data with deep learning for arterial traffic prediction. We establish that it is possible to model traffic at adjacent arterial intersections as diffusion processes if the architecture is correctly constructed with the right parameter settings. A more detailed description of our model architecture is provided in the following subsections.

3.1 DCRNN

DCRNN relies on the title diffusion process to incorporate spatial information into the architecture. This is represented with a transition matrix between the network of sensors which, when multiplied with the state vector at time t , outputs the data point for time $t + 1$. The transition matrix defines a *diffusion convolution* operation that replaces matrix multiplications in a seq2seq RNN to comprise the DCRNN.

Let us define D as the number of detectors in our network and F as the number of features from each detector (flow, occupancy, etc.). Let us also define H as the *prediction horizon*, the number of time steps that we predict into the future, and W as the *window size*, the number of time steps we use to predict. Then each data point is an $X \in \mathbb{R}^{D \times F}$, and our goal is to learn a model that uses input $(X^{(t-W+1)}, \dots, X^{(t)})$ to predict $(X^{(t+1)}, \dots, X^{(t+H)})$.

We represent our system as a weighted directed graph $G = \{V, E, \mathbf{W}\}$, where the detectors are the vertices and the arterial roads are the edges. In our graph, $|V| = D$, and the transition matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ is a weighted adjacency matrix, with entry $\mathbf{W}_{i,j}$ representing the likelihood of transitioning from node i to node j . These weights do not have to be probabilities and do not need to be normalized; they must simply be some function that is larger for nodes j that are more likely destinations of cars in node i . We define $\mathbf{D}_O = \text{diag}(\mathbf{W}\mathbf{1})$ and $\mathbf{D}_I = \text{diag}(\mathbf{W}^\top\mathbf{1})$, where $\mathbf{1} \in \mathbb{R}^N$ is the all-ones vector and the *diag* function takes in a vector and constructs a square matrix with the entries of the vector along its main diagonal. Thus, $\mathbf{D}_O, \mathbf{D}_I \in \mathbb{R}^{N \times N}$ are the normalization matrices for the forward and reverse diffusion processes, since traffic flow is affected by both upstream and downstream detectors.

These diffusion processes are represented as random walks on G with a restart probability $\alpha \in [0, 1]$. Then the stationary distribution \mathcal{P} of the forward diffusion process is

$$\mathcal{P} = \sum_{k=0}^{\infty} \alpha(1 - \alpha)^k (\mathbf{D}_O^{-1}\mathbf{W})^k \quad (1)$$

The DCRNN model uses a truncated K -step diffusion process with learned weights for each step. The diffusion process, which we denote by \mathfrak{F}_θ , is parameterized by $\theta \in \mathbb{R}^{K \times 2}$ and acts on an input $X \in \mathbb{R}^{D \times F}$ to produce an output $Y \in \mathbb{R}^D$.

$$\mathfrak{F}_\theta(X; G, f) = \sum_{k=0}^{K-1} \left(\theta_{k,0} (\mathbf{D}_O^{-1}\mathbf{W})^k + \theta_{k,1} (\mathbf{D}_I^{-1}\mathbf{W}^\top)^k \right) X_{:,f} \quad \text{for } f \in \{1, \dots, F\} \quad (2)$$

To incorporate diffusion convolutions into a model of the network, we use a Gated Recurrent Unit (GRU) [9], but with matrix multiplications replaced by the diffusion convolution. This constitutes the *Diffusion Convolutional Gated Recurrent Unit* (DCGRU). Multiple DCGRUs are then stacked together in a seq2seq architecture, which finalizes the structure of DCRNN (Figure 1). In our paper, we use two cells in the encoder and two cells in the decoder. We feed in a sequence of W inputs $X \in \mathbb{R}^{D \times F}$, and the next H outputs (with earlier outputs recursively fed into the DCRNN to generate later outputs) are the predictions. The network is trained with backpropagation from loss incurred by our labeled data points. The authors also use scheduled sampling during training to switch between using ground truth labeled outputs and predictions from the DCRNN to generate later predictions.

Through the diffusion convolution and transition matrix, spatial information is baked into the architecture and not learned, allowing the model to only learn parameters for the relationships between the spatial and temporal information. Because of this embedded architecture, we need to train the model with a new transition matrix if detectors are added to

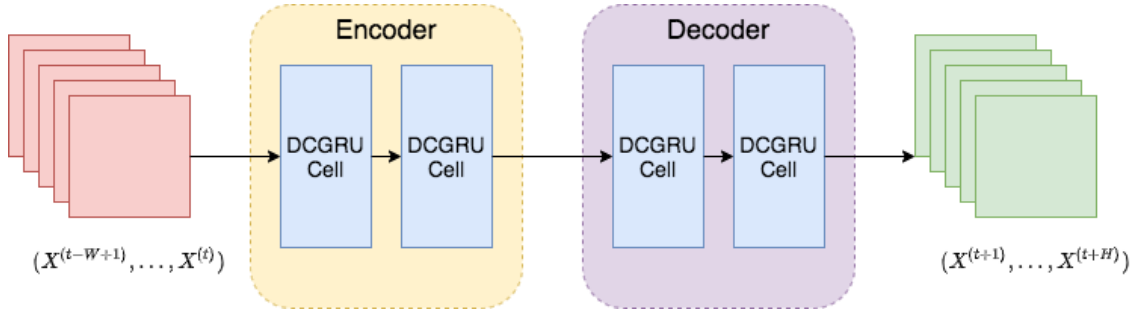


Figure 1: The seq2seq architecture of DCRNN. The encoder and decoder are both composed of multiple DCGRU cells.

or removed from the network. In order to adapt DCRNN for use with arteries, we modify several components, including the transition matrix (detailed below).

3.2 Intersections

Layout and Properties Arterial *intersections* are composed of a series of *inbound approaches* and *outbound approaches* (typically four for both). Each inbound direction has a predefined set of allowed *movements* such as Through, Left Turn, and Right Turn. These movements connect inbound approaches to outbound approaches and dictate the flow of traffic in the intersection. A phase, defined for a set of movements, is a period of time during which the traffic lights are green. At every point in time, a combination of phases (usually two) are active; they are carefully chosen so that no two protected directions intersect. These phases are active for a number of seconds before they are made inactive, and another set of phases is made active. In between, the traffic lights involved in the first phase turn yellow and red during a window of time that is included in the phase length. In order to maintain an organized and easily understandable road network, the pattern of phases repeats. See Figure 3 for examples of signal phase cycles. The total time it takes for one cycle of phases is known as the *cycle length*.

Signal Phase Timing Each set of phases in the signal phase cycle at the intersection is active for a controllable amount of time. One set of phases and phase lengths is known as a *plan*. Because plans may be set to vary between intersections, time, and date, they are recorded for each intersection and time period. Signal phase timing data explains the different plans at an intersection; it includes intersection ID, plan, start time, end time, cycle length, and the length of each individual phase. Note that a group of intersections in the same region typically cycle through the same preset list of phases, with the only difference being the amount of time spent in each phase.

Across intersections, these differences between phases may be major, with some smaller intersections omitting phases that might be utilized at larger intersections. For example, a less busy intersection might not allow protected left turns and instead would allow unprotected turns during the Through phase. Within a single intersection, phase times vary across plans. Morning and afternoon rush hours, known as *peak* hours, may call for plans that allow more time for Through directions compared to off peak and evening hours. Weekend plans usually simulate off peak plans.

In order to take advantage of the full state of the intersection, we incorporate signal phase timing data into our model. This data represents the amount of time allotted to each inbound approach to send cars to each outbound approach. With signal phase timing data, we are able to model the state of an intersection at any given point in time. When combined with traffic flow data, we have a full view into intersection state, including but not limited to everyday traffic patterns, seasonal traffic patterns, and even one-time exogenous events such as accidents.

3.3 Transition Matrix

The weights for the transition matrix in [28] are based off of the road network distances between sensors. These distances are appropriate parameterizations for the diffusion model, as longer distances are correlated with slower diffusion rates. However, for an arterial intersection, road distance becomes inappropriate. Intersection roads are closely clustered and never more than several hundred feet across, rendering any variation in distance insignificant.

Instead, we use *phase green time fraction*, defined as the fraction of a cycle during which cars are allowed to travel from the inbound sensor to the outbound sensor. The phase green time fraction is calculated for each unique combination of intersection and plan. We use the fraction of time and not the actual number of seconds in order to normalize between busy intersections with longer cycles and smaller intersections with shorter cycles. In addition, because DCRNN

Method	Window Size											
	15 min			30 min			1 hr			2 hr		
Unnormalized, Flooded	3.93%	7.17%	9.61%	3.4%	6.49%	8.9%	3.23%	6.29%	9.07%	3.11%	6.13%	9.02%
	3.93%	7.14%	9.56%	3.39%	6.42%	8.87%	3.26%	6.39%	9.26%	3.07%	6.05%	8.92%
Normalized, Flooded	3.93%	7.13%	9.57%	3.39%	6.43%	8.86%	3.24%	6.36%	9.27%	3.11%	6.15%	9.1%
	3.91%	7.13%	9.55%	3.39%	6.45%	8.94%	3.25%	6.34%	9.2%	3.1%	6.14%	9.07%
Exponential, Flooded	3.93%	7.1%	9.51%	3.39%	6.43%	8.85%	3.24%	6.32%	9.13%	3.09%	6.09%	8.97%
	3.92%	7.09%	9.45%	3.4%	6.44%	8.91%	3.24%	6.3%	9.11%	3.13%	6.12%	8.98%
Unnormalized, Unflooded	3.92%	7.14%	9.54%	3.38%	6.44%	8.92%	3.23%	6.33%	9.07%	3.1%	6.19%	9.11%
	3.93%	7.17%	9.62%	3.41%	6.49%	8.97%	3.22%	6.3%	9.12%	3.12%	6.18%	9.16%
	5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m
	Horizon											

Table 1: MAPE for predictions on afternoon peak data with different transition matrix designs for DCRNN. Each experiment was run twice.

assumes a static transition matrix, we use the planned green time for the signal phase timing plan and not the actual green time.

Let P represent the set of phases active at a set of intersections of interest. For a phase $p \in P$, we let p_{in} denote the inbound direction and p_{out} denote the set of outbound directions of the phase. Let $d^{(i)}$ denote the i th detector in our dataset of D detectors, $d_{dir}^{(i)}$ denote the direction of detector i , $I_{d^{(i)}}$ denote the intersection of detector i , and $\text{adj}(d^{(i)}, d^{(j)})$ be a boolean denoting whether detector j is directly downstream from detector i , i.e. there is a direct path from detector i to detector j . Let $L_g(I, p)$ denote the green time of phase p of intersection I , and $L_{yr}(I, p)$ denote the yellow/all-red time of the same context. We assume that the intersection is not one-way, so that at any given moment, two synchronized phases are active; thus, the total cycle length $L(I)$ for intersection I is equal to half of the sum of the lengths of all phases for dual-ring signal control.

$$L(I) = \frac{1}{2} \sum_{p \in P} L_g(I, p) + L_{yr}(I, p) \quad (3)$$

We compute the weights of the transition matrix as follows:

$$W_{i,j} = \begin{cases} 1 & \text{if } I_{d^{(i)}} = I_{d^{(j)}} \\ \frac{\sum_{p \in P} \mathbb{1}_{\text{adj}(d^{(i)}, d^{(j)})} \mathbb{1}_{d_{dir}^{(i)} = p_{in}} \mathbb{1}_{d_{dir}^{(j)} \in p_{out}} (L_g(I_{d^{(i)}}, p) + L_{yr}(I_{d^{(i)}}, p))}{\frac{1}{2} \sum_{p \in P} L_g(I_{d^{(i)}}, p) + L_{yr}(I_{d^{(i)}}, p)} & \text{o.w.} \end{cases} \quad (4)$$

Because we use phase green time fraction instead of road distances, we do not transform the weights with the Gaussian kernel as in [28]. Instead, we leave the probabilities as the weights for the graph. We experimented with several different variations and transformations for the transition matrices. The first alteration we tested was row-normalizing the transition matrix so that the weights would represent probabilities (Eqn 5). Although this wouldn't make a difference for the forward diffusion process, it would change the backward process. In addition, we experimented with applying a Gaussian kernel to an exponential transform of the weights to simulate road distances (Eqn 6); we used $\alpha = 6$ and $\beta = 100$. In tandem with both of these, we zeroed out values in the matrix less than a threshold (Eqn 7), as was done in [28]. The prediction errors are available in Table I. Of the three variants evaluated, none of them caused significant improvements in prediction error, although the thresholded matrix overall had relatively consistently low error. Thus, we use thresholding in our model as well with $\varepsilon = 0.1$ as the limit.

$$W' = \text{diag}(W(\mathbf{1}))^{-1} W \quad (5)$$

$$W'_{ij} = \exp\left(-\frac{(\beta \exp(\alpha(1 - W_{ij})))^2}{\text{Var}(W)}\right) \quad (6)$$

$$W'_{ij} = \begin{cases} W_{ij} & \text{if } W_{ij} \geq \varepsilon \\ 0 & \text{o.w.} \end{cases} \quad (7)$$

In our transition matrices, we incorporated signal phase timing information for Through, Left Turn, and Right Turn directions. However, for both the upstream and downstream directions, we did not include U-Turns in our model. Overall, U-Turns contribute little flow to the data, especially during congested peak hours. In order to avoid this noise and not have to incorporate additional sensors in the opposite direction in our network, we ignored U-Turns.

3.4 Flow Prediction

Detector Types In traffic analysis, there are several measures of interest: flow, occupancy, and speed. *Flow* indicates the number of vehicles that pass a detector over a specified period of time. *Occupancy* is defined as the fraction of time in that interval during which a car is detected on the sensor. *Speed* is defined as the average speed of vehicles in that interval.

In our study, we use two types of detectors: *advance detectors*, placed in lanes about 100-200 feet before the intersection, and stopbar detectors, located just before the intersections. Both types of detectors measure all three metrics, but speed is less reliable than flow and occupancy. Advance detectors provide very reliable measurements for both flow and occupancy, while stopbar detectors are reliable for occupancy but noisier for flow. These differences stem from the design and location of the detectors; advance detectors are well-suited to detect cars moving towards the intersection, while stopbar detectors are perfect for sensing cars that have stopped at the intersection.

Flow Prediction We use flow data in our model because most of the detectors in our network are advance detectors. In some cases, we include occupancy measurements during training in order to determine whether occupancy provides any benefit for flow prediction, but we disregard occupancy predictions, as the results are not as accurate as those of flow. Predicting flow instead of speed does not introduce any major changes to the methodology.

3.5 Baselines

In this study, we use four different baselines: Constant Mean, Seasonal Naive, ARIMAX, and Gated Recurrent Unit (GRU). The naive baselines are included to provide an upper bound on prediction error. With the remaining baselines, we cover both statistical methods and machine learning methods.

3.5.1 Naive Baselines

The naive baselines handle each time series independently between detectors. With Constant Mean, the average value of the training data is taken for each detector, aggregated across time, and used as the constant prediction for that detector for all offsets and time stamps. With Seasonal Naive (sometimes referred to as Historical Average), we take the mean for each combination of detector and time stamp. Predictions are set to the mean value corresponding to the particular detector and time stamp.

3.5.2 ARIMAX

ARIMAX falls into the camp of statistical methods. In general, autoregressive moving average (ARMA) models predict future values of time series from previous values. The autoregressive part $AR(p)$ uses an affine combination of the last p data points to predict. The moving-average part $MA(q)$ uses an affine combination of the last q error terms to predict. The standard $ARMA(p, q)$ model combines both the AR and MA terms.

$$X_t = c + \epsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (8)$$

We included an integrated $I(d)$ term, which applies the ARMA model to the first order difference of our data to enforce stationarity. Finally, we incorporated exogenous $X(b)$ data corresponding to the Fourier terms for a specific seasonality. To tune the parameters (p, d, q, b) , we analyzed the correlation and autocorrelation of the data to determine a set of potential parameter values. We then fit ARIMAX models to each set of parameters and compared the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and prediction error on a validation set. All three metrics resulted in a final order of $(2, 1, 0, 1)$. We experimented with a Seasonal $S(P, D, Q, s)$ term, but did not see any improvements over ARIMAX. In addition, we ran all of our experiments with an additional online version of ARIMAX, but there were no significant differences between the two methods; we omit online ARIMAX for clarity.

3.5.3 GRU

Our GRU architecture follows the architecture used in [28]. In fact, the architecture is the DCRNN model, but using standard matrix multiplications instead of diffusion convolutions. We assume weights of the correct dimensions: W_r , U_r , and b_r for the reset gate r , W_u , U_u , and b_u for the update gate u , and W_c , U_c , and b_c for the candidate activation. At time t , we denote the input by $X^{(t)}$, the candidate activation by $C^{(t)}$, and the hidden state by $H^{(t)}$. Finally, we

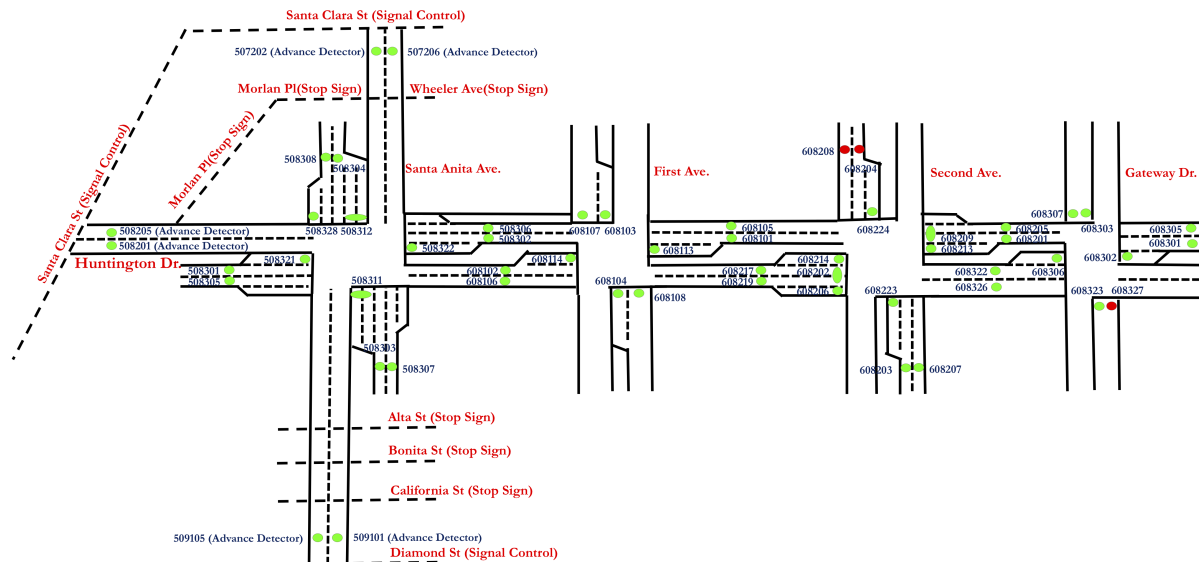


Figure 2: Detector layout at the study site in Arcadia. The intersections we examine in this study are 5083 and 6081.

denote the sigmoid function by σ . Thus, we have the following equations for the GRU cells.

$$\begin{aligned}
 r^{(t)} &= \sigma(W_r X^{(t)} + U_r H^{(t-1)} + b_r) \\
 u^{(t)} &= \sigma(W_u X^{(t)} + U_u H^{(t-1)} + b_u) \\
 C^{(t)} &= \tanh(W_c X^{(t)} + U_c(r^{(t)} \odot H^{(t-1)}) + b_c) \\
 H^{(t)} &= u^{(t)} \odot H^{(t-1)} + (1 - u^{(t)}) \odot C^{(t)}
 \end{aligned} \tag{9}$$

These GRU cells are combined in a seq2seq architecture to form the final GRU model. Like with DCRNN, we use two cells in the encoder and two cells in the decoder. Thus, the only difference between DCRNN and GRU is the diffusion convolution with the incorporation of signal phase timing data.

4 Study Site and Dataset

4.1 Study Site

The data used in this report is part of a larger dataset collected for the I-210 Connected Corridors Project. The project dataset includes traffic flow data from stopbar and advance detectors, maps of the cities and sensor layouts, and the corresponding signal timing sheets. We surveyed detectors along Huntington Dr. between Santa Clara St. and Gateway Dr. in the city of Arcadia (Figure 2).

In particular, we focus on detectors 508302 and 508306. These detectors were selected because they are heavily covered by both advance and stopbar detectors in both the upstream and downstream directions. The upstream lanes are covered in the Through direction by detectors 608101 and 608105, in the Right Turn direction by detector 608107, and in the Left Turn direction by detector 608104. The downstream lanes are covered in the Through direction by detectors 508201 and 508205, in the Right Turn direction by detectors 507202 and 507206, and in the Left Turn direction by detectors 509101 and 509105. These last two sets of advance detectors for the downstream turn directions are several blocks down; while there are some leakages that prevent the system from being fully closed, they are only at minor intersections with stop signs. We call this ideal situation the Full Information scenario. See Figure 3 for the signal phase cycle and Table 2 for the signal timing plans.

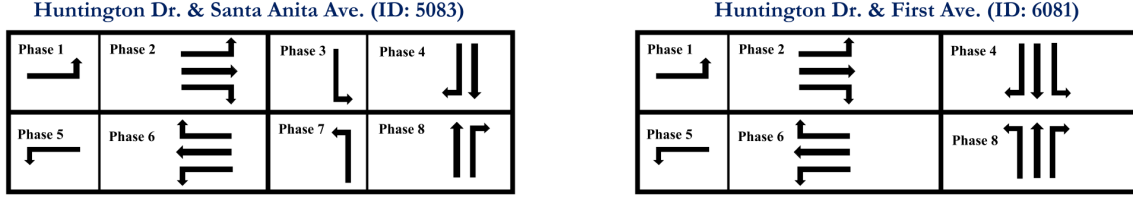


Figure 3: The signal phases for the upstream (6081) and downstream (5083) intersections at our study site.

Huntington Dr & Santa Anita Ave (ID: 5083)										
Plan Name	Activation Time	Cycle (sec)	Phase 1&5		Phase 2&6		Phase 3&7		Phase 4&8	
			G (sec)	Y+R (sec)	G (sec)	Y+R (sec)	G (sec)	Y+R (sec)	G (sec)	Y+R (sec)
E	0:00-6:00 & 21:00-24:00	110	20	3	27	5	20	3	27	5
P_1	9:00-15:30 & 19:00-21:00	120	15	3	39	5	14	3	36	5
P_2	6:00-9:00	120	11	3	46	5	11	3	36	5
P_3	15:30-19:00	120	15	3	41	5	12	3	36	5

Huntington Dr & First Ave (ID: 6081)										
Plan Name	Activation Time	Cycle (sec)	Phase 1&5		Phase 2&6		Phase 3&7		Phase 4&8	
			G (sec)	Y+R (sec)	G (sec)	Y+R (sec)	G (sec)	Y+R (sec)	G (sec)	Y+R (sec)
E	0:00-6:00 & 21:00-24:00	90	20	3	28	4	NA	NA	31	4
P_1	9:00-15:30 & 19:00-21:00	120	10	3	74	4	NA	NA	25	4
P_2	6:00-9:00	120	10	3	74	4	NA	NA	25	4
P_3	15:30-19:00	120	10	3	74	4	NA	NA	25	4

Table 2: Signal timing plans at the two intersections in Arcadia. "G", "Y", and "R" stand for "Green Time", "Yellow Time", and "All-Red Time", respectively. The green times provided in the table are the maximum ones from the controller settings.

4.2 Dataset

The dataset includes data from 1/1/2017 to 12/31/2017 aggregated into 5 minute intervals. Both advance and stopbar detectors measure volume and occupancy, from which we can calculate flow and occupancy percentage (we use the term *occupancy* to refer to occupancy percentage). Visualizations of the data from an advance detector (Figure 4) confirm that the measurements are highly cyclical. Note that flow measurements are the cleanest, while occupancy measurements are slightly noisier. Furthermore, data for the morning and afternoon peaks are more consistent and have larger magnitude changes than off peak data. We also visualize flow and occupancy for a stopbar detector (Figure 5) and illustrate the relative noisiness of the flow measurements compared to those of the advance detector. Notably, the stopbar detector sees much higher occupancy values compared to the advance detector, located further upstream from the intersection. Using flow and occupancy, we plot the flow-occupancy graphs for the data (Figure 6) and note that they exhibit the trapezoidal shape that is typical of traffic fundamental diagrams. The morning peak reaches congestion and queue spillback far more often than the other two periods. Each period has its own set of signal phase timing plans, which explains the varying parameters.

Detector health and signal phase timings are collected at a granularity of one day. The signals at these intersections use four different plans: P_1 , P_2 , P_3 , and E , which correspond respectively to off peak, morning peak, afternoon peak, and nighttime (Table 2). As expected, data from P_2 and P_3 have larger magnitude than data from P_1 and E and exhibit very obvious cyclical patterns. The P_2 and P_3 plans are only active on weekdays, so we train and predict only on weekday data for the morning and afternoon peaks.

Preprocessing The sensor data was combined with the detector health data to filter out spurious data. Because the DCRNN requires data for all of the detectors at each timestamp, we kept data only from days where all 12 detectors in our system were healthy. However, detector 508201 provided data for only one day out of the entire year; as a result, we ignored the detector at the expense of introducing another impurity in our closed system. Otherwise, the detectors

Detector 508302 Data in the Month of August

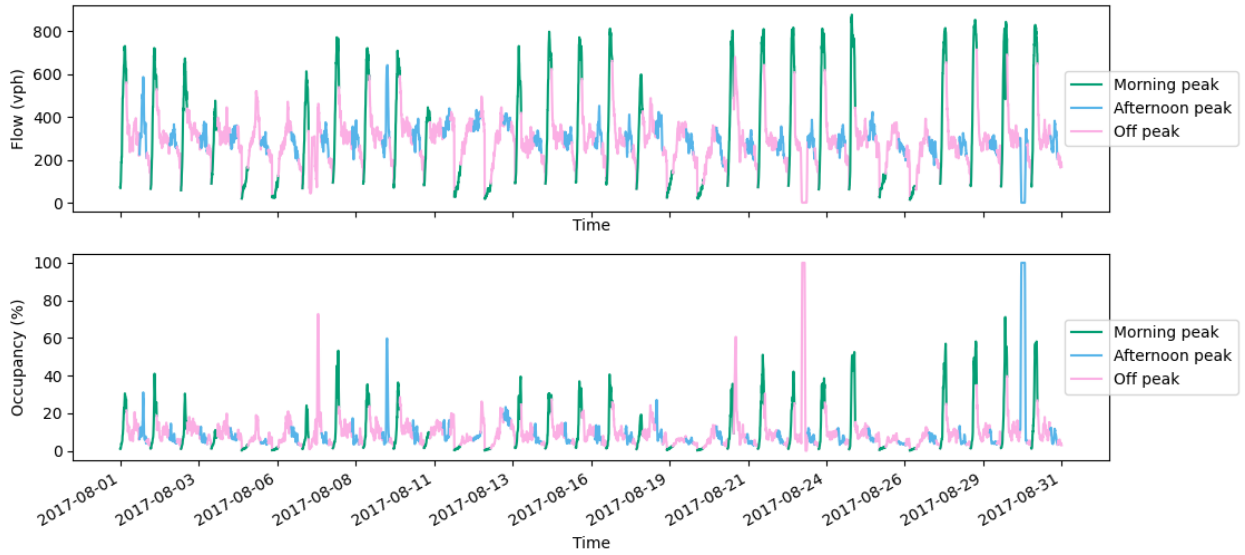


Figure 4: Flow and occupancy measured in the month of August by detector 508302.

Detector 608107 Data in the Month of August

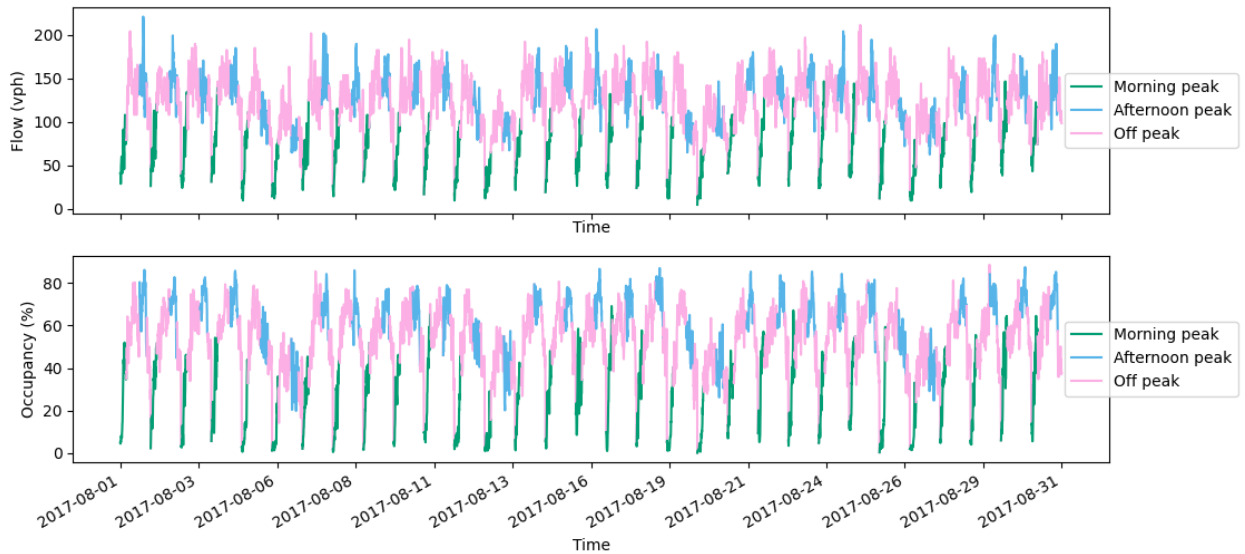


Figure 5: Flow and occupancy measured in the month of August by detector 608107. Compared to detector 508302, the flow measurements are much noisier, although the occupancy measurements are similar.

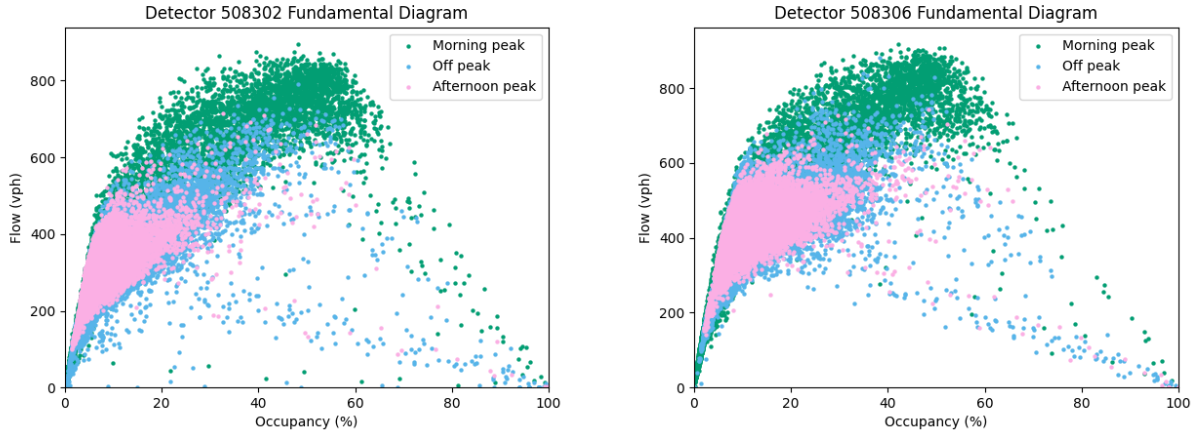


Figure 6: Fundamental diagrams for our two detectors of interest: 508302 and 508306.

were fairly healthy, with only a few outages throughout the year. Luckily, only a few days were filtered out in this way, so not much data was dropped.

5 Analysis

For our experiments, we reused most of the original DCRNN code², with some adaptations to test variations of the algorithm. The hyperparameters used in the original paper produced positive results for us. We train for 100 epochs with batch size 64, decay the learning rate by 0.1 from an initial 0.1 every 10 epochs, and use scheduled sampling with an inverse sigmoid decay function to slowly introduce predicted data as labels. We focus on the Mean Absolute Percentage Error (MAPE) in order to have a normalized metric of prediction performance, although we include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) for reference. The experiments were run in parallel, so we do not have hard numbers for training times, but for the shorter plans, training models ranged from about thirty minutes for window size 15m to an hour for window size 2hr.

Experiments were conducted separately on three of the traffic signal timing plans: P1, P2, and P3. The difference in traffic flow between the three periods of the day is significant, so we trained separate models for each plan instead of learning one model for the entire dataset. However, data from the nighttime plan E is far more sparse and noisy; moreover, because nighttime periods exhibit little congestion, it is less useful to predict for them than for the daytime, so we did not run experiments on plan E.

All experiments predict with a horizon of six, which is equivalent to half an hour. We test four different window sizes: fifteen minutes (15m), half an hour (30m), one hour (1hr), and two hours (2hr), corresponding to three, six, twelve, and 24 data points. Because the lengths of some periods during the plans were short, for example only three hours or 36 points for the morning peak, we included a *start buffer* of the previous plan’s data at the beginning of each period of each plan in order to have access to more data points. The length of the start buffer was equal to the window size that we were training. For example, for data from plan P2 with a 1hr window size, we included the hour of data from plan E from 5:00 to 6:00. This way, the number of data points for each of the different window sizes would remain the same. We set values less than the threshold of $\varepsilon = 0.1$ in the transition matrices to 0.

In order to test situations in which full data is not available, we ran experiments with augmented data. We explored two types of scenarios: missing detectors and unhealthy detector data. In the former case, we simulate a situation in which we do not have full information by creating a new transition matrix with a subset of detectors and predicting using data from only that subset of detectors. In the latter case, we use the same transition matrix as the full information case, but zero out part of the data to simulate a situation in which the number of detectors is fixed, but some of the detectors do not provide good data during training. These ablations illustrate a more realistic situation; cities may not have the resources to cover all streets with detectors, and even then, real-world detectors may occasionally fail, as exemplified by our data.

²Code available at <https://github.com/victorchan314/DCRNN>

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
DCRNN	MSE	521.04	2725	10481	485.99	2831	9832	379.28	2444	8958	347.63	2199	8348
	RMSE	22.83	52.2	102.38	22.05	53.22	99.16	19.48	49.44	94.65	18.64	46.89	91.37
	MAE	16.33	34.37	64.03	15.13	33.97	60.54	13.9	31.75	57.18	13.4	30.67	55.83
	MAPE	5.38%	11.33%	17.44%	5.14%	11.23%	16.81%	5.06%	10.52%	16.26%	4.58%	9.63%	15.67%
Constant Mean	MSE	68980	65071	59296	68980	65071	59296	68980	65071	59296	68980	65071	59296
	RMSE	262.64	255.09	243.51	262.64	255.09	243.51	262.64	255.09	243.51	262.64	255.09	243.51
	MAE	238.82	230.16	213.45	238.82	230.16	213.45	238.82	230.16	213.45	238.82	230.16	213.45
	MAPE	137.6%	123.2%	105.0%	137.6%	123.2%	105.0%	137.6%	123.2%	105.0%	137.6%	123.2%	105.0%
Seasonal Naive	MSE	69022	65112	59355	69022	65112	59355	69022	65112	59355	69022	65112	59355
	RMSE	262.72	255.17	243.63	262.72	255.17	243.63	262.72	255.17	243.63	262.72	255.17	243.63
	MAE	238.77	230.1	213.44	238.77	230.1	213.44	238.77	230.1	213.44	238.77	230.1	213.44
	MAPE	137.5%	123.2%	105.1%	137.5%	123.2%	105.1%	137.5%	123.2%	105.1%	137.5%	123.2%	105.1%
ARIMAX	MSE	1006	6915	25247	1006	6915	25247	1006	6915	25247	1006	6915	25247
	RMSE	31.72	83.16	158.89	31.72	83.16	158.89	31.72	83.16	158.89	31.72	83.16	158.89
	MAE	23.93	63.59	122.44	23.93	63.59	122.44	23.93	63.59	122.44	23.93	63.59	122.44
	MAPE	7.3%	17.81%	28.63%	7.3%	17.81%	28.63%	7.3%	17.81%	28.63%	7.3%	17.81%	28.63%
GRU	MSE	578.2	2846	9373	466.49	2809	9668	385.91	2412	8968	348.2	2355	9414
	RMSE	24.05	53.35	96.82	21.6	53.0	98.33	19.64	49.11	94.7	18.66	48.53	97.03
	MAE	17.34	36.63	63.92	15.42	36.07	64.96	14.34	33.43	61.49	13.41	33.05	64.42
	MAPE	5.39%	11.1%	17.46%	4.97%	10.94%	17.63%	4.56%	10.15%	16.96%	4.37%	9.98%	17.5%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 3: Test prediction errors for the morning peak with full information and flow only.

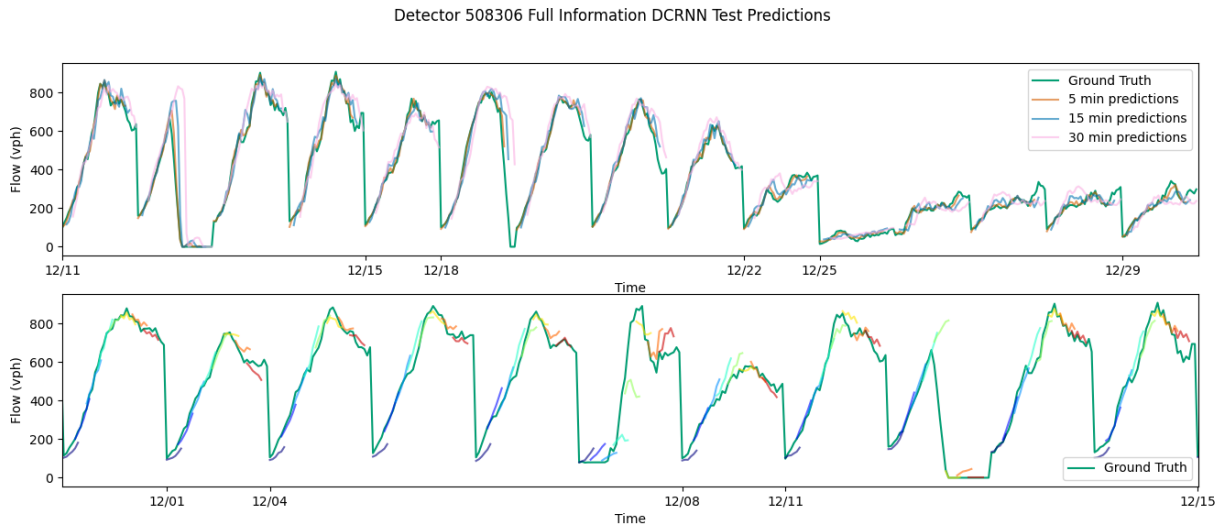


Figure 7: Test predictions of DCRNN on flow measured in the month of December by detector 508306. The model was trained with a window size of 1hr and a horizon of 6 data points, or half an hour. The top graph shows predictions across time for horizons of 5 minutes, 15 minutes, and 30 minutes. The bottom graph shows predictions for all 6 horizons at various points in the time series. For neatness sake, we limit the predictions depicted in the bottom graph to every 4 time steps.

5.1 DCRNN with Full Information

The DCRNN was our model of choice because we had access to full information with which we could populate a transition matrix. From the signal phase timing data, we knew the distances between all pairs of detectors in the network. We hypothesized that with a closed and fully-observed system, we would achieve higher accuracy compared to other flow prediction methods.

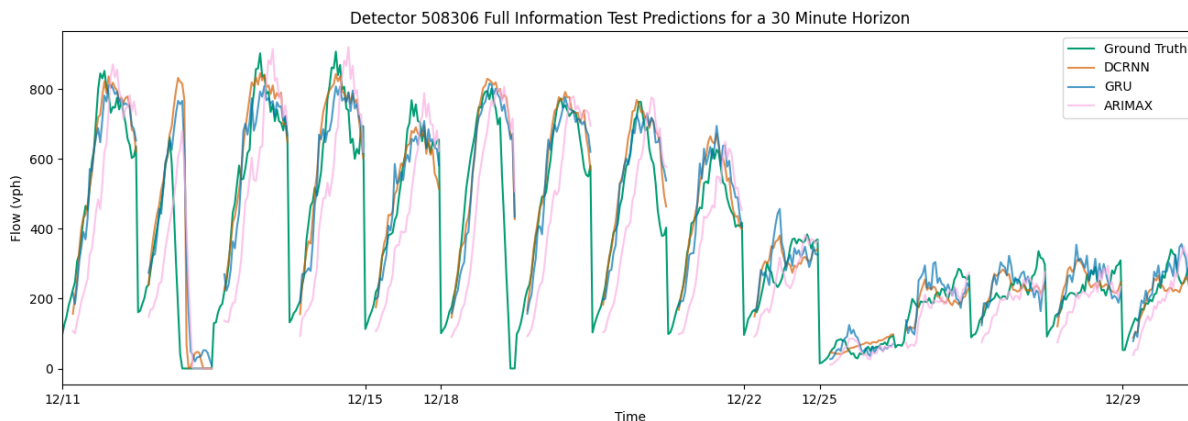


Figure 8: Test predictions of DCRNN, GRU, and ARIMAX on flow measured in the month of December by detector 508306. The model was trained with a window size of 1hr and a horizon of 6 data points, or half an hour. Here we show the predictions with a horizon of 6, where DCRNN improves the most compared to the other methods.

Morning Peak The results for the morning peak are available in Table 3. We note several trends in DCRNN results. Prediction error increases drastically as we predict flow at longer horizons. Unsurprisingly, data points further from input data points have higher entropy. Moreover, in general, error decreases as window size increases. However, in some situations, error increases when we use a 2hr window. This seems to be a common pattern across other experiments as well. There is enough variation in the training procedure that performance saturates once window size reaches 1hr. Because using longer windows requires longer time to train the model, a practical implementation of DCRNN could use a shorter window and still achieve near-optimal performance. For this location, 1hr is the plateau for window size.

We compare DCRNN results to that of ARIMAX and GRU. ARIMAX has decent performance on the data. However, for the longer-horizon 15m and 30m predictions, DCRNN performs much better than ARIMAX. The difference is less pronounced between DCRNN and GRU. In fact, GRU performance is very close to DCRNN performance, achieving lower error in several cases, even though the difference is within the bounds of random variation. Notably, the gap narrows for 15m predictions, and for 30m predictions, DCRNN achieves lower error.

The difference between DCRNN and the other models becomes clear when we predict for longer horizons. DCRNN is able to learn long-term temporal characteristics of the system for prediction more effectively than GRU, the next-best model, can. As a result, DCRNN consistently outperforms GRU for long horizon predictions, even when GRU achieves lower error for the 5m and 15m predictions. Moreover, while GRU error plateaus at the 1hr window size, DCRNN error continues to drop when we use a 2hr window size, indicating that with more training data, we would see even better predictions from DCRNN. Because traffic during the morning peak is very irregular and complicated, the model requires more examples to learn the rich patterns present in the data.

The comparison between DCRNN predictions and the ground truth is shown in Figure 7. All three horizons result in predictions that are very close to the ground truth, although the half-hour predictions are clearly less accurate. We note a consistent issue across both of the graphs, which we dub the *overshooting problem*. This is more clearly depicted in the bottom graph, in which we plot at regular time steps the sequence of predictions for the next six time steps.

When the ground truth data maintains its cyclical course, the predictions follow very closely. Even when the model exhibits sawtooth edges, the predictions match the ground truth. However, for irregular changes in shape, the model's predictions at each time step continue along the previous trajectory and diverge from the ground truth. For example, on 12/11, traffic grinds to a halt in the middle of the morning peak, likely due to an exogenous event such as an accident. However, the model's predictions overshoot this drop and continue upwards in the direction that traffic flow would usually travel. After seeing the atypically low measurements, the model tries to correct by predicting a very sudden sharp downward trajectory. At this point, the output of DCRNN is actually negative, overshooting the flattening out of flow, so we clip the value to zero when predicting. While DCRNN is able to learn the cyclical patterns, it struggles to adapt to outliers and rare exogenous events.

In Figure 8 we see the comparison between DCRNN and its two closest baselines for half-hour predictions. ARIMAX exhibits the cyclical shape of the ground truth data, but lags behind by several time steps, and thus is unable to predict the future accurately. GRU is similar to DCRNN, except that it tends to not follow the ground truth pattern as closely as

does DCRNN. Especially during the congested morning periods, DCRNN reaches the same amplitude that the ground truth reaches, but GRU peaks too soon and turns downwards before the ground truth flow subsides. Even during winter vacation, from Christmas (12/25) to New Year’s Eve (12/31), when traffic flow has quieted down, DCRNN adapts to the smaller peaks more quickly than GRU does.

Other Times In both the afternoon peak (Table 5) and off peak (Table 4) results, we recognize many of the same trends present in the morning peak experiments. Longer windows consistently produce lower prediction errors, but not past 1hr. For the off peak plan, we see the same pattern as with the morning peak. DCRNN outperforms all of the other models except for with a 2hr window, where GRU achieves lower error for 5m and 15m horizon predictions. However, for the afternoon peak plan, DCRNN outperforms every other model. Thus, the results substantiate the hypothesis that a major advantage of DCRNN is its ability to learn the pattern of traffic flow for long horizon predictions. The ARIMAX and GRU models are able to memorize the gist of the data trends, but fail to understand finer details of the data.

One notable difference between these two periods and the morning peak is the drastically lower prediction error, often dipping below four percent, even for the baselines. The magnitude of data from outside the morning peak is much lower, and therefore the peaks are not as pronounced. Because there is less up-and-down variation in the data, the trends are easier for the models to learn and predict. Moreover, there is less contrast between different days of the week. Traffic during the afternoon peak is simpler than during the morning peak, so even with a limited amount of data, DCRNN enjoys the full benefits of the signal phase timing plans.

For the afternoon peak, DCRNN still has a clear improvement over GRU, especially for the 30m predictions. However, for the off peak plan, although DCRNN consistently achieves lower error than GRU, the decrease is far smaller. We believe that this is due to the fact that signal timing plans have the most substantial effect on traffic flow when there is congestion but not queue spillback. During these peak hours, cars are fully subject to the signal phases. Thus, the transition matrix representing the diffusion process comes into play. During the off peak hours, there is little congestion, and therefore little benefit from modeling the intersection.

Flow and Occupancy In theory, if we provide additional sensor information to the model, it should perform at least as well as before. At the very least, the model could zero out all of the weights that are multiplied with the additional data, which would simply result in a model with the original data. Thus, if we append occupancy data to the flow data to predict future traffic flow, the error should not be greater.

This expectation is realized in the results for the off peak and afternoon peak periods (Tables 6 and 8), where the results are very similar when occupancy is included in the input data. There is no significant improvement in the error, but the results are also not worse. However, when we include occupancy for the morning peak (Table 7), the error increases by a not insignificant amount.

For the morning peak, the MAPE for 30m predictions increases by one percent, and GRU also regresses by two percent. Notably, we still see the same trend as before, where DCRNN always outperforms the other baselines for 30m predictions, but is occasionally outperformed by GRU for 5m and 15m predictions. However, DCRNN and GRU both suffer when occupancy is added as an extra feature. We do not fully understand why this phenomenon occurs; however, our hypothesis is that occupancy data from the detectors is noisy enough that it affects the long-term relationships that the models learn.

5.2 DCRNN with Incomplete Information

We chose a very specific group of detectors for our experiment because they represented a system that was very close to being fully-observed for both upstream and downstream lanes. However, it may not be practical for cities to install sensors for every lane on every road to define a closed system. Even in our network, detectors 508302 and 508306 are the only two detectors with this type of coverage, and the system is still not fully closed. Therefore, we reran all of the experiments while omitting either the upstream or downstream sensors. The results are available in Tables 9, 10, and 11

No Upstream When upstream detectors are omitted, the results are fairly similar to those of the full network. The differences are not significant and could be attributed to random variation. This occurs for all plans, both when we include occupancy and when we only use flow. There are two data points that have much worse performance: for the morning peak, window size 2hr, and prediction horizon 15m, for both flow only and flow and occupancy. This is likely just a random spurious data point. Otherwise, omitting upstream detectors does not make a major difference in prediction error. However, we note that performance for the no upstream case exhibits the saturating behavior that we saw with GRU. With more data, it is likely that the full information case would surpass the case of omission.

No Downstream Surprisingly, performance on the morning peak consistently improves by a slight amount when downstream detectors are omitted. This pattern does not appear for the off or afternoon peaks.

This phenomenon is likely caused by the fact that while the upstream detectors cover all of the lanes that cars can enter from, all three downstream directions have some sort of imperfection. The Through direction has only one out of two sensors healthy, while both the Left Turn and Right Turn sensors are several blocks down, with multiple intersections with stop signs in between. Because this violates the diffusion assumptions of the transition matrix, it is not as fitting of a model as we would like it to be. As a result, if we merely omit the data, DCRNN achieves lower error because the transition matrix holds for the remaining detectors.

5.3 Unhealthy Detectors

In real-life applications, detectors may not always be operational, and will fail with some probability. The invalid data from these detectors can usually be easily detected; numbers are often far too low or far too high to be legitimate data. For our study, we simulate unhealthy detectors by zeroing out some portion of the data. We experimented with two types of unhealthy detector data scenarios: unhealthy detectors, in which some subset of detectors are zeroed out for the entire training period, and unhealthy days, in which data for all detectors is zeroed out on a randomly-selected proportion of days. We divide the unhealthy detectors scenario into two further categories for detector directions and detector types.

For all scenarios, we tested the augmented data in two different ways. First, we used the trained models from the full information scenario to predict with the augmented data to determine their robustness to unreliable data. Second, we trained a new model with the unhealthy data to determine the impact of spurious data on prediction quality.

Unhealthy Detector Directions There are two directions for the sensors of interest: Upstream and Downstream. For each direction, we explore augmenting data for three groups of interest: all detectors, Through detectors, and Turn detectors. These results are available in Tables [12](#), [13](#), and [14](#).

In general, prediction accuracy is lower and variance is higher when unhealthy detectors are present. Although there are many cases in which the unhealthy scenario performs similarly to the full information scenario, overall, the full information case achieves low error most consistently. When using the full information models for the morning peak, error blows up when we zero out all of the detectors in either direction. Notably, error does not increase as significantly when only Upstream Turn (stopbar detectors have noisier flow measurements) or Downstream Through (there is only one detector in this group) detector measurements are augmented. The results from retraining DCRNN on augmented data are similar to the full information case, albeit slightly worse overall, especially with shorter window sizes and longer prediction horizons. There are several individual cases where MAPE jumps to over 30%. The consistent presence of multiple outliers suggests that unhealthy data negatively affects predictions. Unsurprisingly, retraining on the augmented data far outperforms predicting on augmented data with the full information models, although it is still unable to close the gap. Clearly, the full information models have learned to rely on data from both upstream and downstream detectors for the most accurate predictions.

We see the same patterns for the afternoon peak. Zeroing out more detectors results in less accuracy. However, zeroing out the single detector in the Downstream Through direction or zeroing out the stopbar detectors in the Upstream Turn direction causes much less increase in error compared to the other scenarios. The unhealthy directions scenario performs consistently worse than the full information scenario, but retraining on the augmented data reduces the impact. The off peak period also exhibits the same trends, but they are less pronounced because the data has less variance and is easier to predict.

Unhealthy Detector Types In our network, we have advance and stopbar detectors. Advance detectors produce reliable measurements for both flow and occupancy; on the other hand, while stopbar detectors produce reliable occupancy measurements, their flow measurements are noisier. We explored augmenting the data from different types of sensors. Results are available in Tables [15](#), [16](#), and [17](#).

Zeroing out stopbar detector data causes a noticeable increase in prediction error for longer horizon predictions, although the difference is not significant for shorter horizons. Retraining on the augmented data brings the errors closer together, but does not close the gap. Even though the flow information from stopbar detectors is relatively noisy, it is still important in incorporating data from the Upstream Turn directions. However, most of the detectors in the network are advance detectors. Understandably, zeroing out this data causes a very significant escalation in prediction error. Surprisingly, short horizon predictions are not affected when the models are retrained. This suggests that while flow information from detectors 508302 and 508306 (our detectors of interest) is sufficient for one-step predictions, flow information from the other detectors is crucial for long horizon predictions.

When we include occupancy data, the error from unhealthy detectors increases slightly, but not by an unreasonable amount. Because stopbar detectors produce reliable occupancy data, zeroing out stopbar detector data causes a larger increase in prediction error compared to when only flow is considered. Overall, retraining outperforms predicting on the augmented data with the full information models.

These trends are present for all three plans, but they are the most pronounced for the morning peak. The two trends present for all of the scenarios are that unhealthy data causes prediction accuracy and consistency to diminish, and that retraining models provides some mitigation of that decline.

Unhealthy Proportion of Days Our final scenario investigates setting the detector data to zero on a certain proportion of days. Our data on detector health only stores information at a granularity of one day, so we decided to zero out days of data instead of specific data points. We analyzed four percentages of data to augment: 5%, 10%, 25%, and 50%; the results for this experiment are available in Tables [18](#), [19](#), and [20](#). Because of the length of time required to run experiments, we only tested one random seed for each scenario; thus, small and inconsistent improvements or deterioration in prediction accuracy are noted as possibly due to random variation.

The outcomes of this scenario are not very surprising. Using the full information matrix to predict on the augmented data performs worse, with MAPE surging as a larger portion of the data is augmented. With all plans, having 5% unhealthy data raises MAPE by 5%, and having 10% unhealthy data increments MAPE by another 5%. Having 25% and 50% unhealthy data significantly curtails performance for all prediction horizons. After retraining, the errors are still higher than when there is no unhealthy data, although the accuracy is much closer. In some cases, even with 50% unhealthy data, DCRNN matches the original scenario. Still, the results exhibit very high variation, even if some isolated cases achieve good performance. The 10% retrain scenario for the morning peak results in very similar results to the original scenario; however, because the 5% retrain scenario has a significant decline in accuracy, we believe that the similarity stems from randomness. In addition, the 5% and 10% retrain scenarios for the afternoon peak do not manifest effects of the augmented data. We believe this is caused by the less pronounced variation in the daily peaks of afternoon traffic.

We can conclude from these results that data quality is of utmost importance to our model. While having more training data is necessary to boost performance, the data must be carefully preprocessed to filter out bad data and avoid detrimental effects. Unlike in the previous unhealthy detectors scenarios, where the model is robust even with some unhealthy detectors, here the quality of the data itself is degraded. This highlights the distinction between accounting for several misbehaving detectors and handling low-quality data. While the model is able to absorb some of the impact and produce decent results in some cases, it produces much more accurate and consistent results when no data is corrupt.

Of note are the retrained predictions for a window size of 2hr with both the morning and off peak periods. In some cases, the MAPE decreases significantly compared to the other window sizes. While we do not know why this occurs, we hypothesize that although 50% of the data has been zeroed out, having 24 data points in the window allows DCRNN to better learn relationships between data points. This way, it can distinguish between scenarios where data is corrupt for an entire day and where data is simply low for a small period of time.

5.3.1 Other Experiments

We also tested several DCRNN setup variations to investigate whether these alterations would generate more accurate predictions. First, we trained a different DCRNN for each day of the week, as [10](#) discovered significant changes in the traffic profile between each day of the week. Second, we explored single-horizon predictions. Instead of predicting all six horizons at once, we trained six different models, each of which predicted one of the six horizons.

Neither of these experiments improved upon the original model. Splitting by days of the week resulted in similar performance. Training models for single-horizon predictions actually performed slightly worse than the normal model, especially for long horizon predictions. Having all of the labels to guide the model during training allows it to learn the full relationship between each point in time instead of just jumping ahead to a specific horizon without the intermediate context. The original DCRNN model has the structure and expressiveness to represent traffic in our system without these extra modifications.

6 Conclusion

Arterial traffic prediction is far more challenging compared to freeway traffic prediction. Spatial information plays a much more salient role and must be effectively applied to optimize prediction accuracy. In this study, we explored using signal phase timing data to generate a weighted adjacency matrix based on traffic signal green times. Combined with our graph convolutional model of choice, the DCRNN, we show that the signal phase timing data enhances arterial flow

predictions, especially long horizon forecasts. We achieve MAPE as low as 16% for a prediction horizon of 30 minutes for morning peak congestion. For afternoon peak and off peak data, we achieve MAPE lower than 8% and 10% for the same horizon. Signal phase timing data defines the relationships between detectors in the network and allows the model to learn long-term temporal relationships for long horizon predictions.

In addition, we tested numerous variations of the measurements and the detector network to investigate the effects of detector coverage, detector type, and data quality on prediction performance. One surprising discovery is that detector coverage is overshadowed by detector proximity and precise measurements; as a result, we saw no significant decrease in performance even after omitting stopbar detectors and detectors several blocks away.

In our ablations with augmented data, prediction error skyrocketed when we simulated unhealthy data for even just one or two detectors. However, after we retrained on the augmented data, errors dipped back down to the levels in the full information scenarios. This reveals that when presented with more information, our model makes good use of it to generate excellent predictions, but it is also robust to faulty detectors when training. However, at least some of the detectors must be relatively reliable, or else accuracy will take a heavy hit. If entire days of data are zeroed out, errors soar, even when only 5% of days are augmented. Although the data can include some anomalies, it must at the very least be relatively consistent throughout the entire dataset.

As expected, omitting advance detectors rendered the model ineffective for long horizons. Simulated unhealthy stopbar detectors had less of an impact, although the difference was slightly larger when occupancy measurements were included. Short horizon predictions were not particularly affected by these changes, even with all advance detectors zeroed out, indicating that the graph structure of the network is most useful for long horizon predictions.

In the future, we can study extensions and variations of this work. We can train deeper and more expressive models to better learn complex patterns. The area of deep unsupervised learning is burgeoning, and because traffic network matrices are polynomial with respect to the number of detectors and the size of the graph, it would be very useful to find a compressed feature representation for the entire network state. This would be particularly beneficial for the signal phase timing data. DCRNN applies a static transition matrix, so we used planned green times; however, traffic plans are dynamic and reactive to traffic conditions, so the actual green times are different for each point in time. With a latent embedding, we could encode the signal phases for each data point instead of aggregating them into a single static matrix. Some newer graph convolutional architectures, such as Graph WaveNet [48], allow adaptive filters, so they can be applied to the problem as well.

Another prospective direction is to include even more varied types of information, such as pedestrian activity at intersections. In addition, DCRNN allows prediction of all detectors at once. We could examine flow forecasts for an entire network of sensors, even one that isn't a closed system. Flow predictions can also be applied to signal control applications to determine the effect of forecasts on travel time and queue length on urban roads. Arterial traffic predictions have many applications, so we must leverage all the data and technology in our toolbox to tackle the challenge.

References

- [1] Mohamed S. Ahmed and Allen R. Cook. Analysis of freeway traffic time-series data by using box-jenkins techniques. *Transportation Research Record*, 721:1–9, 1979.
- [2] Walaa Alajali, Wei Zhou, Sheng Wen, and Yu Wang. Intersection traffic prediction using decision tree models. *Symmetry*, 10(9):386, Sep 2018.
- [3] Pramod Anantharam, Krishnaprasad Thirunarayan, Surendra Marupudi, Amit Sheth, and Tanvi Banerjee. Understanding city traffic dynamics utilizing sensor and textual observations. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 3793–3799. AAAI Press, 2016.
- [4] M. T. Asif, J. Dauwels, C. Y. Goh, A. Oran, E. Fathi, M. Xu, M. M. Dhanya, N. Mitrovic, and P. Jaillet. Spatiotemporal patterns in large-scale traffic speed prediction. *IEEE Transactions on Intelligent Transportation Systems*, 15(2):794–804, 2014.
- [5] James Atwood and Don Towsley. Diffusion-convolutional neural networks. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 1993–2001. Curran Associates, Inc., 2016.
- [6] Pinlong Cai, Yunpeng Wang, Guangquan Lu, Peng Chen, Chuan Ding, and Jianping Sun. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies*, 62:21 – 34, 2016.

- [7] Cen Chen, Kenli Li, Sin Teo, Xiaofeng Zou, Kang Wang, Jie Wang, and Zeng Zeng. Gated residual recurrent graph neural networks for traffic prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:485–492, 07 2019.
- [8] F. Chen, S. Pan, J. Jiang, H. Huo, and G. Long. Dagen: Dual attention graph convolutional networks. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2019.
- [9] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014.
- [10] Samuel Coogan, Christopher Flores, and Pravin Varaiya. Traffic predictive control from low-rank structure. *Transportation Research Part B: Methodological*, 97:1 – 22, 2017.
- [11] Z. Cui, K. Henrickson, R. Ke, and Y. Wang. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12, 2019.
- [12] P. Dell’Acqua, F. Bellotti, R. Berta, and A. De Gloria. Time-aware multivariate nearest neighbor regression methods for traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):3393–3402, 2015.
- [13] Loukas Dimitriou, Theodore Tsekeris, and Antony Stathopoulos. Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. *Transportation Research Part C: Emerging Technologies*, 16(5):554 – 573, 2008.
- [14] S. Du, T. Li, X. Gong, Y. Yang, and S. J. Horng. Traffic flow forecasting based on hybrid deep learning framework. In *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pages 1–6, 2017.
- [15] Shen Fang, Qi Zhang, Gaofeng Meng, Shiming Xiang, and Chunhong Pan. Gstnet: Global spatial-temporal network for traffic flow prediction. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 2286–2293. International Joint Conferences on Artificial Intelligence Organization, 7 2019.
- [16] X. Feng, X. Ling, H. Zheng, Z. Chen, and Y. Xu. Adaptive multi-kernel svm with spatial–temporal correlation for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(6):2001–2013, 2019.
- [17] R. Fu, Z. Zhang, and L. Li. 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, 2016.
- [18] G. Fusco, C. Colombaroni, L. Comelli, and N. Isaenko. Short-term traffic predictions on large urban traffic networks: Applications of network-based machine learning models and dynamic traffic assignment models. In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pages 93–101, 2015.
- [19] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:3656–3663, 07 2019.
- [20] K. Guo, Y. Hu, Z. Qian, H. Liu, K. Zhang, Y. Sun, J. Gao, and B. Yin. Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12, 2020.
- [21] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:922–929, 07 2019.
- [22] Mohammad M. Hamed, Hashem R. Al-Masaeid, and Zahi M. Bani Said. Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering*, 121(3):249–254, 1995.
- [23] Z. He, C. Chow, and J. Zhang. Stann: A spatio–temporal attentive neural network for traffic prediction. *IEEE Access*, 7:4795–4806, 2019.
- [24] M.G. Karlaftis and E.I. Vlahogianni. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3):387 – 399, 2011.
- [25] Xiangjie Kong, Zhenzhen Xu, Guojiang Shen, Jinzhong Wang, Qiuyuan Yang, and Benshi Zhang. Urban traffic congestion estimation and prediction based on floating car trajectory data. *Future Generation Computer Systems*, 61:97 – 107, 2016.

- [26] Kranti Kumar, Manoranjan Parida, and Vinod Kumar Katiyar. Short term traffic flow prediction in heterogeneous condition using artificial neural network. *Transport*, 30(4):397–405, 2015.
- [27] Yaguang Li, Kun Fu, Zheng Wang, Cyrus Shahabi, Jieping Ye, and Yan Liu. Multi-task representation learning for travel time estimation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, page 1695–1704, New York, NY, USA, 2018. Association for Computing Machinery.
- [28] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations (ICLR '18)*, 2018.
- [29] Yijun Lin, Nikhit Mago, Yu Gao, Yaguang Li, Yao-Yi Chiang, Cyrus Shahabi, and José Luis Ambite. Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '18, page 359–368, New York, NY, USA, 2018. Association for Computing Machinery.
- [30] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang. Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):865–873, 2015.
- [31] X. Ma, H. Zhong, Y. Li, J. Ma, Z. Cui, and Y. Wang. Forecasting transportation network speed using deep capsule networks with nested lstm models. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12, 2020.
- [32] Xiaolei Ma, Zhuang Dai, Zhengbing He, Jihui Ma, Yong Wang, and Yunpeng Wang. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17(4):818, Apr 2017.
- [33] Xiaolei Ma, Zhimin Tao, Yin Hai Wang, Haiyang Yu, and Yunpeng Wang. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54:187 – 197, 2015.
- [34] J. Mackenzie, J. F. Roddick, and R. Zito. An evaluation of htm and lstm for short-term arterial traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(5):1847–1857, 2019.
- [35] Takuma Oda and Yulia Tachibana. Distributed fleet control with maximum entropy deep reinforcement learning. 2018.
- [36] Iwao Okutani and Yorgos J. Stephanedes. Dynamic prediction of traffic volume through kalman filtering theory. *Transportation Research Part B: Methodological*, 18(1):1 – 11, 1984.
- [37] Zheyi Pan, Yuxuan Liang, Weifeng Wang, Yong Yu, Yu Zheng, and Junbo Zhang. Urban traffic prediction from spatio-temporal data using deep meta learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 1720–1730, New York, NY, USA, 2019. Association for Computing Machinery.
- [38] Nicholas G. Polson and Vadim O. Sokolov. Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 79:1 – 17, 2017.
- [39] L. Ruiz, F. Gama, and A. Ribeiro. Gated graph convolutional recurrent neural networks. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5, 2019.
- [40] Brian L Smith, Billy M Williams, and R [Keith Oswald]. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, 10(4):303 – 321, 2002.
- [41] Daniel Jian Sun and Lun Zhang. Vehicle actuation based short-term traffic flow prediction model for signalized intersections. *Journal of Central South University*, 19, 01 2012.
- [42] Y. Tian and L. Pan. Predicting short-term traffic flow by long short-term memory recurrent neural network. In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, pages 153–158, 2015.
- [43] Eleni I. Vlahogianni, Matthew G. Karlaftis, and John C. Golias. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies*, 13(3):211 – 234, 2005.
- [44] Billy Williams, Priya Durvasula, and Donald Brown. Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record*, 1644:132–141, 01 1998.
- [45] Billy M. Williams. Multivariate vehicular traffic flow prediction: Evaluation of arimax modeling. *Transportation Research Record*, 1776(1):194–200, 2001.
- [46] Billy M. Williams and Lester A. Hoel. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, 129(6):664–672, 2003.

- [47] Cheng-Ju Wu, Thomas Schreiter, Roberto Horowitz, and Gabriel Gomes. Traffic flow prediction using optimal autoregressive moving average with exogenous input-based predictors. *Transportation Research Record*, 2421(1):125–132, 2014.
- [48] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. pages 1907–1913, 08 2019.
- [49] Huaxiu Yao, Yiding Liu, Ying Wei, Xianfeng Tang, and Zhenhui Li. Learning from multiple cities: A meta-learning approach for spatial-temporal prediction. In *The World Wide Web Conference, WWW '19*, page 2181–2191, New York, NY, USA, 2019. Association for Computing Machinery.
- [50] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, and Zhenhui Li. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:5668–5675, 07 2019.
- [51] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. pages 3634–3640, 07 2018.
- [52] Bing Yu, Haoteng Yin, and Zhanxing Zhu. St-unet: A spatio-temporal u-network for graph-structured time series modeling, 2019.
- [53] Haiyang Yu, Zhihai Wu, Shuqin Wang, Yunpeng Wang, and Xiaolei Ma. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors*, 27:1501, 06 2017.
- [54] Wu Yuankai, Huachun Tan, Jin Peter, Bin Shen, and Bin Ran. Short-term traffic flow prediction based on multilinear analysis and k-nearest neighbor regression. pages 556–569, 07 2015.
- [55] H. Zhan, G. Gomes, X. S. Li, K. Madduri, A. Sim, and K. Wu. Consensus ensemble system for traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 19(12):3903–3914, 2018.
- [56] Junbo Zhang, Yu Zheng, and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI'17*, page 1655–1661. AAAI Press, 2017.
- [57] Y. Zhang, S. Wang, B. Chen, and J. Cao. Gcgan: Generative adversarial nets with graph cnn for network-scale traffic prediction. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2019.
- [58] Y. Zhang, S. Wang, B. Chen, J. Cao, and Z. Huang. Trafficgan: Network-scale deep traffic prediction with generative adversarial nets. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12, 2019.
- [59] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–11, 2019.
- [60] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu. Lstm network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2):68–75, 2017.

A Additional Results

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
DCRNN	MSE	289.71	974.75	1842	235.67	920.15	1851	206.54	794.93	1774	246.19	802.61	1721
	RMSE	17.02	31.22	42.92	15.35	30.33	43.03	14.37	28.19	42.13	15.69	28.33	41.49
	MAE	13.33	23.52	31.32	11.9	22.51	31.21	11.22	21.48	30.79	12.22	21.68	30.6
	MAPE	4.18%	7.58%	10.53%	3.72%	7.26%	10.48%	3.54%	7.01%	10.44%	3.86%	6.94%	10.02%
Constant Mean	MSE	9605	8875	8437	9605	8875	8437	9605	8875	8437	9605	8875	8437
	RMSE	98.01	94.21	91.86	98.01	94.21	91.86	98.01	94.21	91.86	98.01	94.21	91.86
	MAE	73.73	72.44	71.93	73.73	72.44	71.93	73.73	72.44	71.93	73.73	72.44	71.93
	MAPE	26.93%	26.85%	27.23%	26.93%	26.85%	27.23%	26.93%	26.85%	27.23%	26.93%	26.85%	27.23%
Seasonal Naive	MSE	9633	8897	8459	9633	8897	8459	9633	8897	8459	9633	8897	8459
	RMSE	98.15	94.33	91.97	98.15	94.33	91.97	98.15	94.33	91.97	98.15	94.33	91.97
	MAE	73.88	72.53	72.06	73.88	72.53	72.06	73.88	72.53	72.06	73.88	72.53	72.06
	MAPE	26.97%	26.87%	27.27%	26.97%	26.87%	27.27%	26.97%	26.87%	27.27%	26.97%	26.87%	27.27%
ARIMAX	MSE	358.68	1483	3337	358.68	1483	3337	358.68	1483	3337	348.53	1443	3310
	RMSE	18.94	38.52	57.77	18.94	38.52	57.77	18.94	38.52	57.77	18.67	37.99	57.54
	MAE	14.48	28.37	41.78	14.48	28.37	41.78	14.48	28.37	41.78	14.33	28.07	41.69
	MAPE	4.51%	8.83%	13.31%	4.51%	8.83%	13.31%	4.51%	8.83%	13.31%	4.47%	8.73%	13.28%
GRU	MSE	310.7	1055	2107	247.61	948.38	2039	220.41	879.1	2018	197.81	825.01	2008
	RMSE	17.63	32.48	45.9	15.74	30.8	45.17	14.85	29.65	44.92	14.06	28.72	44.81
	MAE	13.66	24.69	34.07	12.16	23.29	33.39	11.52	22.47	33.2	10.92	21.71	32.82
	MAPE	4.29%	7.77%	10.84%	3.82%	7.33%	10.66%	3.62%	7.07%	10.59%	3.44%	6.89%	10.52%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 4: Test prediction errors for off peak with full information and flow only.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
DCRNN	MSE	330.6	942.6	1591	268.6	857.35	1448	227.86	757.32	1346	216.88	733.63	1401
	RMSE	18.18	30.7	39.89	16.39	29.28	38.06	15.1	27.52	36.69	14.73	27.09	37.44
	MAE	14.51	24.27	30.94	13.02	23.17	29.51	11.98	21.78	28.41	11.75	21.41	28.94
	MAPE	4.01%	6.69%	8.44%	3.62%	6.37%	8.09%	3.3%	6.0%	7.82%	3.26%	5.94%	8.0%
Constant Mean	MSE	4985	4955	4953	4985	4955	4953	4985	4955	4953	4985	4955	4953
	RMSE	70.61	70.39	70.38	70.61	70.39	70.38	70.61	70.39	70.38	70.61	70.39	70.38
	MAE	50.95	51.27	51.65	50.95	51.27	51.65	50.95	51.27	51.65	50.95	51.27	51.65
	MAPE	16.15%	16.32%	16.56%	16.15%	16.32%	16.56%	16.15%	16.32%	16.56%	16.15%	16.32%	16.56%
Seasonal Naive	MSE	5000	4973	4975	5000	4973	4975	5000	4973	4975	5000	4973	4975
	RMSE	70.72	70.53	70.53	70.72	70.53	70.53	70.72	70.53	70.53	70.72	70.53	70.53
	MAE	51.12	51.43	51.78	51.12	51.43	51.78	51.12	51.43	51.78	51.12	51.43	51.78
	MAPE	16.19%	16.36%	16.6%	16.19%	16.36%	16.6%	16.19%	16.36%	16.6%	16.19%	16.36%	16.6%
ARIMAX	MSE	375.2	1248	2587	375.2	1248	2587	375.2	1248	2587	375.2	1248	2587
	RMSE	19.37	35.34	50.86	19.37	35.34	50.86	19.37	35.34	50.86	19.37	35.34	50.86
	MAE	15.34	28.01	40.17	15.34	28.01	40.17	15.34	28.01	40.17	15.34	28.01	40.17
	MAPE	4.22%	7.76%	11.27%	4.22%	7.76%	11.27%	4.22%	7.76%	11.27%	4.22%	7.76%	11.27%
GRU	MSE	371.43	1205	2415	299.3	1163	2456	262.24	1090	2434	235.82	994.51	2338
	RMSE	19.27	34.72	49.15	17.3	34.11	49.57	16.19	33.02	49.34	15.36	31.54	48.36
	MAE	15.27	27.63	38.66	13.8	27.21	39.04	12.89	26.3	38.95	12.27	25.08	38.16
	MAPE	4.21%	7.7%	10.89%	3.8%	7.51%	10.98%	3.55%	7.25%	10.89%	3.37%	6.92%	10.7%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 5: Test prediction errors for the afternoon peak with full information and flow only.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
DCRNN	MSE	306.62	1045	1997	242.99	889.57	1897	229.61	833.3	1736	221.86	773.11	1651
	RMSE	17.51	32.33	44.7	15.59	29.83	43.56	15.15	28.87	41.67	14.89	27.8	40.64
	MAE	13.46	24.13	32.5	12.04	22.53	31.73	11.65	21.93	31.29	11.43	21.27	30.68
	MAPE	4.24%	7.69%	10.55%	3.79%	7.2%	10.34%	3.68%	7.05%	10.22%	3.62%	6.89%	10.01%
Constant Mean	MSE	9605	8875	8437	9605	8875	8437	9605	8875	8437	9605	8875	8437
	RMSE	98.01	94.21	91.86	98.01	94.21	91.86	98.01	94.21	91.86	98.01	94.21	91.86
	MAE	73.73	72.44	71.93	73.73	72.44	71.93	73.73	72.44	71.93	73.73	72.44	71.93
	MAPE	26.93%	26.85%	27.23%	26.93%	26.85%	27.23%	26.93%	26.85%	27.23%	26.93%	26.85%	27.23%
Seasonal Naive	MSE	9633	8897	8459	9633	8897	8459	9633	8897	8459	9633	8897	8459
	RMSE	98.15	94.33	91.97	98.15	94.33	91.97	98.15	94.33	91.97	98.15	94.33	91.97
	MAE	73.88	72.53	72.06	73.88	72.53	72.06	73.88	72.53	72.06	73.88	72.53	72.06
	MAPE	26.97%	26.87%	27.27%	26.97%	26.87%	27.27%	26.97%	26.87%	27.27%	26.97%	26.87%	27.27%
ARIMAX	MSE	358.68	1483	3337	358.68	1483	3337	358.68	1483	3337	348.53	1443	3310
	RMSE	18.94	38.52	57.77	18.94	38.52	57.77	18.94	38.52	57.77	18.67	37.99	57.54
	MAE	14.48	28.37	41.78	14.48	28.37	41.78	14.48	28.37	41.78	14.33	28.07	41.69
	MAPE	4.51%	8.83%	13.31%	4.51%	8.83%	13.31%	4.51%	8.83%	13.31%	4.47%	8.73%	13.28%
GRU	MSE	307.84	1060	2189	250.86	979.25	2120	222.67	905.88	1998	204.48	873.2	2019
	RMSE	17.55	32.56	46.79	15.84	31.29	46.05	14.92	30.1	44.71	14.3	29.55	44.94
	MAE	13.65	24.85	34.65	12.21	23.67	33.96	11.59	22.81	33.38	11.1	22.29	33.44
	MAPE	4.28%	7.85%	11.05%	3.83%	7.44%	10.83%	3.6%	7.22%	10.9%	3.45%	7.06%	10.89%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 6: Test prediction errors for off peak with full information and both flow and occupancy.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
DCRNN	MSE	588.97	3033	9908	501.47	3287	10631	496.57	2993	9865	691.94	3737	10755
	RMSE	24.27	55.08	99.54	22.39	57.33	103.11	22.28	54.71	99.33	26.3	61.14	103.71
	MAE	17.49	37.15	62.26	15.93	38.34	65.62	15.84	36.07	61.84	17.97	37.56	61.93
	MAPE	5.57%	11.4%	17.58%	5.24%	11.48%	17.5%	5.03%	10.75%	16.81%	5.88%	11.93%	16.92%
Constant Mean	MSE	68980	65071	59296	68980	65071	59296	68980	65071	59296	68980	65071	59296
	RMSE	262.64	255.09	243.51	262.64	255.09	243.51	262.64	255.09	243.51	262.64	255.09	243.51
	MAE	238.82	230.16	213.45	238.82	230.16	213.45	238.82	230.16	213.45	238.82	230.16	213.45
	MAPE	137.6%	123.2%	105.0%	137.6%	123.2%	105.0%	137.6%	123.2%	105.0%	137.6%	123.2%	105.0%
Seasonal Naive	MSE	69022	65112	59355	69022	65112	59355	69022	65112	59355	69022	65112	59355
	RMSE	262.72	255.17	243.63	262.72	255.17	243.63	262.72	255.17	243.63	262.72	255.17	243.63
	MAE	238.77	230.1	213.44	238.77	230.1	213.44	238.77	230.1	213.44	238.77	230.1	213.44
	MAPE	137.5%	123.2%	105.1%	137.5%	123.2%	105.1%	137.5%	123.2%	105.1%	137.5%	123.2%	105.1%
ARIMAX	MSE	1006	6915	25247	1006	6915	25247	1006	6915	25247	1006	6915	25247
	RMSE	31.72	83.16	158.89	31.72	83.16	158.89	31.72	83.16	158.89	31.72	83.16	158.89
	MAE	23.93	63.59	122.44	23.93	63.59	122.44	23.93	63.59	122.44	23.93	63.59	122.44
	MAPE	7.3%	17.81%	28.63%	7.3%	17.81%	28.63%	7.3%	17.81%	28.63%	7.3%	17.81%	28.63%
GRU	MSE	594.37	2915	9875	468.2	2845	9783	397.06	2519	9613	348.15	2374	9125
	RMSE	24.38	53.99	99.38	21.64	53.34	98.91	19.93	50.19	98.05	18.66	48.73	95.53
	MAE	17.5	36.91	66.41	15.58	36.77	65.51	14.47	34.42	65.34	13.41	33.41	63.22
	MAPE	5.61%	11.84%	19.5%	5.08%	11.39%	18.75%	4.78%	10.8%	18.49%	4.59%	10.93%	19.23%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 7: Test prediction errors for the morning peak with full information and both flow and occupancy.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
DCRNN	MSE	357.93	1062	1722	306.42	915.98	1599	283.82	891.11	1542	264.05	786.63	1412
	RMSE	18.92	32.6	41.51	17.5	30.27	39.99	16.85	29.85	39.27	16.25	28.05	37.58
	MAE	15.0	25.81	32.31	13.92	23.97	31.0	13.28	23.43	30.42	12.91	22.11	28.98
	MAPE	4.16%	7.09%	8.84%	3.83%	6.62%	8.52%	3.66%	6.41%	8.24%	3.57%	6.12%	7.99%
Constant Mean	MSE	4985	4955	4953	4985	4955	4953	4985	4955	4953	4985	4955	4953
	RMSE	70.61	70.39	70.38	70.61	70.39	70.38	70.61	70.39	70.38	70.61	70.39	70.38
	MAE	50.95	51.27	51.65	50.95	51.27	51.65	50.95	51.27	51.65	50.95	51.27	51.65
	MAPE	16.15%	16.32%	16.56%	16.15%	16.32%	16.56%	16.15%	16.32%	16.56%	16.15%	16.32%	16.56%
Seasonal Naive	MSE	5000	4973	4975	5000	4973	4975	5000	4973	4975	5000	4973	4975
	RMSE	70.72	70.53	70.53	70.72	70.53	70.53	70.72	70.53	70.53	70.72	70.53	70.53
	MAE	51.12	51.43	51.78	51.12	51.43	51.78	51.12	51.43	51.78	51.12	51.43	51.78
	MAPE	16.19%	16.36%	16.6%	16.19%	16.36%	16.6%	16.19%	16.36%	16.6%	16.19%	16.36%	16.6%
ARIMAX	MSE	375.2	1248	2587	375.2	1248	2587	375.2	1248	2587	375.2	1248	2587
	RMSE	19.37	35.34	50.86	19.37	35.34	50.86	19.37	35.34	50.86	19.37	35.34	50.86
	MAE	15.34	28.01	40.17	15.34	28.01	40.17	15.34	28.01	40.17	15.34	28.01	40.17
	MAPE	4.22%	7.76%	11.27%	4.22%	7.76%	11.27%	4.22%	7.76%	11.27%	4.22%	7.76%	11.27%
GRU	MSE	372.78	1197	2335	296.94	1144	2367	260.78	1033	2151	236.86	979.0	2223
	RMSE	19.31	34.61	48.33	17.23	33.82	48.66	16.15	32.15	46.39	15.39	31.29	47.16
	MAE	15.31	27.59	38.28	13.77	26.97	38.27	12.85	25.62	36.69	12.28	25.0	37.55
	MAPE	4.24%	7.72%	10.83%	3.8%	7.48%	10.76%	3.54%	7.09%	10.35%	3.38%	6.96%	10.64%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 8: Test prediction errors for the afternoon peak with full information and both flow and occupancy.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
Full Information, Flow only	MSE	289.71	974.75	1842	235.67	920.15	1851	206.54	794.93	1774	246.19	802.61	1721
	RMSE	17.02	31.22	42.92	15.35	30.33	43.03	14.37	28.19	42.13	15.69	28.33	41.49
	MAE	13.33	23.52	31.32	11.9	22.51	31.21	11.22	21.48	30.79	12.22	21.68	30.6
	MAPE	4.18%	7.58%	10.53%	3.72%	7.26%	10.48%	3.54%	7.01%	10.44%	3.86%	6.94%	10.02%
No Upstream, Flow only	MSE	309.56	1024	1926	241.06	884.79	1756	221.46	842.33	1777	214.58	854.02	1878
	RMSE	17.59	32.01	43.9	15.53	29.75	41.91	14.88	29.02	42.16	14.65	29.22	43.34
	MAE	13.72	24.19	31.83	12.06	22.41	30.6	11.57	22.03	31.12	11.34	22.07	31.41
	MAPE	4.29%	7.75%	10.59%	3.77%	7.14%	10.09%	3.63%	7.13%	10.42%	3.5%	6.99%	10.39%
No Downstream, Flow only	MSE	284.41	946.74	1797	227.36	864.49	1805	208.05	838.81	1852	191.78	736.98	1647
	RMSE	16.86	30.77	42.39	15.08	29.4	42.49	14.42	28.96	43.05	13.85	27.15	40.59
	MAE	13.26	23.55	31.31	11.77	22.24	31.06	11.24	21.66	31.05	10.84	20.75	30.13
	MAPE	4.2%	7.62%	10.51%	3.71%	7.17%	10.45%	3.53%	6.99%	10.48%	3.39%	6.63%	9.75%
Full Information, Flow and Occupancy	MSE	306.62	1045	1997	242.99	889.57	1897	229.61	833.3	1736	221.86	773.11	1651
	RMSE	17.51	32.33	44.7	15.59	29.83	43.56	15.15	28.87	41.67	14.89	27.8	40.64
	MAE	13.46	24.13	32.5	12.04	22.53	31.73	11.65	21.93	31.29	11.43	21.27	30.68
	MAPE	4.24%	7.69%	10.55%	3.79%	7.2%	10.34%	3.68%	7.05%	10.22%	3.62%	6.89%	10.01%
No Upstream, Flow and Occupancy	MSE	323.32	1040	1937	253.22	890.87	1783	227.95	855.36	1796	220.8	806.94	1720
	RMSE	17.98	32.26	44.02	15.91	29.85	42.23	15.1	29.25	42.38	14.86	28.41	41.48
	MAE	13.95	24.65	32.94	12.26	22.75	31.37	11.66	22.18	31.21	11.48	21.84	31.28
	MAPE	4.37%	7.85%	10.69%	3.84%	7.23%	10.14%	3.66%	7.05%	10.08%	3.58%	7.05%	10.13%
No Downstream, Flow and Occupancy	MSE	312.92	1044	1964	254.8	904.88	1779	253.5	889.25	1816	220.19	767.11	1613
	RMSE	17.69	32.33	44.33	15.96	30.08	42.19	15.92	29.82	42.63	14.84	27.7	40.17
	MAE	13.66	24.7	33.19	12.13	22.86	31.65	12.0	22.68	31.89	11.32	21.44	30.59
	MAPE	4.29%	7.97%	11.05%	3.83%	7.37%	10.45%	3.76%	7.28%	10.84%	3.57%	6.89%	10.14%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 9: Comparison of DCRNN test prediction errors between the full information, no upstream, and no downstream scenarios for off peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
Full Information, Flow only	MSE	521.04	2725	10481	485.99	2831	9832	379.28	2444	8958	347.63	2199	8348
	RMSE	22.83	52.2	102.38	22.05	53.22	99.16	19.48	49.44	94.65	18.64	46.89	91.37
	MAE	16.33	34.37	64.03	15.13	33.97	60.54	13.9	31.75	57.18	13.4	30.67	55.83
	MAPE	5.38%	11.33%	17.44%	5.14%	11.23%	16.81%	5.06%	10.52%	16.26%	4.58%	9.63%	15.67%
No Upstream, Flow only	MSE	528.3	2676	9098	448.39	2752	9166	404.72	2373	8701	3423	6142	11257
	RMSE	22.98	51.74	95.39	21.18	52.47	95.74	20.12	48.71	93.28	58.51	78.37	106.1
	MAE	16.51	34.06	58.25	14.97	33.84	57.78	14.56	31.89	56.86	30.44	45.9	64.4
	MAPE	5.36%	10.86%	15.91%	4.89%	10.67%	16.11%	4.82%	9.93%	15.75%	9.44%	14.06%	17.83%
No Downstream, Flow only	MSE	536.21	2741	9198	417.91	2586	9003	364.03	2300	8249	327.75	2240	8617
	RMSE	23.16	52.36	95.91	20.44	50.86	94.89	19.08	47.97	90.83	18.1	47.33	92.83
	MAE	16.78	35.33	60.44	14.59	33.47	59.57	13.78	31.79	56.23	13.03	30.91	56.78
	MAPE	5.45%	11.45%	16.77%	5.12%	10.83%	16.55%	4.65%	10.14%	15.59%	4.37%	9.7%	15.37%
Full Information, Flow and Occupancy	MSE	588.97	3033	9908	501.47	3287	10631	496.57	2993	9865	691.94	3737	10755
	RMSE	24.27	55.08	99.54	22.39	57.33	103.11	22.28	54.71	99.33	26.3	61.14	103.71
	MAE	17.49	37.15	62.26	15.93	38.34	65.62	15.84	36.07	61.84	17.97	37.56	61.93
	MAPE	5.57%	11.4%	17.58%	5.24%	11.48%	17.5%	5.03%	10.75%	16.81%	5.88%	11.93%	16.92%
No Upstream, Flow and Occupancy	MSE	552.72	3143	9989	490.13	2927	9910	564.89	3220	10459	7252	8144	13422
	RMSE	23.51	56.06	99.95	22.14	54.11	99.55	23.77	56.75	102.27	85.16	90.25	115.85
	MAE	16.86	36.95	62.11	15.67	36.18	62.31	16.72	36.78	62.7	42.71	50.98	70.04
	MAPE	5.36%	11.49%	17.94%	5.05%	10.95%	17.09%	5.3%	11.09%	16.76%	14.87%	17.81%	22.39%
No Downstream, Flow and Occupancy	MSE	599.94	3365	10526	546.82	3545	12051	488.61	3153	10473	498.32	2809	10180
	RMSE	24.49	58.01	102.6	23.38	59.55	109.78	22.1	56.15	102.34	22.32	53.01	100.9
	MAE	17.24	37.11	61.77	16.33	37.65	66.2	15.78	35.12	60.81	15.81	34.14	61.54
	MAPE	5.72%	11.46%	17.18%	5.05%	11.06%	17.2%	5.06%	10.7%	16.17%	5.32%	10.67%	16.55%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m
Horizon													

Table 10: Comparison of DCRNN test prediction errors between the full information, no upstream, and no downstream scenarios for the morning peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
Full Information, Flow only	MSE	330.6	942.6	1591	268.6	857.35	1448	227.86	757.32	1346	216.88	733.63	1401
	RMSE	18.18	30.7	39.89	16.39	29.28	38.06	15.1	27.52	36.69	14.73	27.09	37.44
	MAE	14.51	24.27	30.94	13.02	23.17	29.51	11.98	21.78	28.41	11.75	21.41	28.94
	MAPE	4.01%	6.69%	8.44%	3.62%	6.37%	8.09%	3.3%	6.0%	7.82%	3.26%	5.94%	8.0%
No Upstream, Flow only	MSE	342.33	954.42	1591	273.3	893.16	1529	258.29	850.59	1489	246.59	848.98	1563
	RMSE	18.5	30.89	39.9	16.53	29.89	39.11	16.07	29.16	38.6	15.7	29.14	39.55
	MAE	14.72	24.45	30.89	13.09	23.39	30.02	12.66	22.8	29.5	12.42	22.73	30.53
	MAPE	4.05%	6.72%	8.45%	3.6%	6.45%	8.28%	3.48%	6.31%	8.27%	3.42%	6.32%	8.6%
No Downstream, Flow only	MSE	331.03	944.69	1568	266.86	868.3	1511	225.89	757.48	1366	215.65	748.14	1401
	RMSE	18.19	30.74	39.61	16.34	29.47	38.88	15.03	27.52	36.97	14.68	27.35	37.44
	MAE	14.53	24.43	30.54	13.0	23.38	30.05	11.98	21.87	28.54	11.7	21.55	28.75
	MAPE	4.02%	6.76%	8.32%	3.6%	6.43%	8.25%	3.32%	6.06%	7.94%	3.24%	5.93%	7.9%
Full Information, Flow and Occupancy	MSE	357.93	1062	1722	306.42	915.98	1599	283.82	891.11	1542	264.05	786.63	1412
	RMSE	18.92	32.6	41.51	17.5	30.27	39.99	16.85	29.85	39.27	16.25	28.05	37.58
	MAE	15.0	25.81	32.31	13.92	23.97	31.0	13.28	23.43	30.42	12.91	22.11	28.98
	MAPE	4.16%	7.09%	8.84%	3.83%	6.62%	8.52%	3.66%	6.41%	8.24%	3.57%	6.12%	7.99%
No Upstream, Flow and Occupancy	MSE	357.97	1062	1743	287.58	900.53	1638	281.96	886.24	1531	262.49	799.08	1461
	RMSE	18.92	32.59	41.75	16.96	30.01	40.47	16.79	29.77	39.14	16.2	28.27	38.23
	MAE	15.04	25.88	32.39	13.42	23.6	31.11	13.28	23.02	29.82	12.86	21.94	29.09
	MAPE	4.13%	7.09%	8.85%	3.68%	6.45%	8.56%	3.68%	6.33%	8.13%	3.55%	6.02%	7.98%
No Downstream, Flow and Occupancy	MSE	352.88	1048	1727	288.99	913.23	1614	306.4	911.13	1598	322.51	852.29	1510
	RMSE	18.79	32.38	41.56	17.0	30.22	40.18	17.5	30.18	39.98	17.96	29.19	38.86
	MAE	14.92	25.68	32.13	13.53	23.81	31.01	13.96	23.95	31.12	14.17	23.09	29.94
	MAPE	4.13%	7.1%	8.8%	3.74%	6.56%	8.51%	3.87%	6.67%	8.6%	3.96%	6.41%	8.3%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m
Horizon													

Table 11: Comparison of DCRNN test prediction errors between the full information, no upstream, and no downstream scenarios for the afternoon peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
None	MSE	289.71	974.75	1842	235.67	920.15	1851	206.54	794.93	1774	246.19	802.61	1721
	RMSE	17.02	31.22	42.92	15.35	30.33	43.03	14.37	28.19	42.13	15.69	28.33	41.49
	MAE	13.33	23.52	31.32	11.9	22.51	31.21	11.22	21.48	30.79	12.22	21.68	30.6
	MAPE	4.18%	7.58%	10.53%	3.72%	7.26%	10.48%	3.54%	7.01%	10.44%	3.86%	6.94%	10.02%
Upstream	MSE	744.97	2837	5461	514.31	2501	6474	1388	5602	12387	1500	2995	4940
	RMSE	27.29	53.27	73.91	22.68	50.02	80.47	37.26	74.85	111.3	38.74	54.73	70.29
	MAE	21.23	40.96	57.75	17.59	39.26	62.96	32.31	61.91	82.03	33.07	44.46	55.65
	MAPE	6.59%	12.49%	17.58%	5.43%	11.73%	18.67%	9.49%	17.69%	23.16%	9.67%	13.02%	16.08%
Upstream Through	MSE	391.28	1233	2123	320.71	1331	2529	385.13	1152	2441	547.57	1574	2521
	RMSE	19.78	35.13	46.08	17.91	36.49	50.29	19.62	33.95	49.41	23.4	39.68	50.22
	MAE	15.75	27.57	35.46	14.02	28.06	38.72	15.4	25.73	37.92	18.74	31.1	38.22
	MAPE	5.03%	8.88%	11.56%	4.28%	8.43%	12.17%	4.62%	7.94%	12.52%	5.67%	9.3%	11.72%
Upstream Turn (Stopbar)	MSE	306.82	1061	1928	243.98	960.42	1925	215.7	869.86	1896	258.11	828.87	1742
	RMSE	17.52	32.58	43.91	15.62	30.99	43.88	14.69	29.49	43.55	16.07	28.79	41.74
	MAE	13.7	24.81	32.81	12.12	23.46	32.94	11.48	22.62	32.64	12.54	22.12	31.23
	MAPE	4.34%	8.12%	11.14%	3.79%	7.56%	11.01%	3.63%	7.39%	10.94%	3.95%	7.06%	10.29%
Downstream	MSE	474.99	3109	10529	423.49	3365	18696	431.45	3846	18541	470.39	3494	12130
	RMSE	21.79	55.77	102.61	20.58	58.02	136.73	20.77	62.02	136.17	21.69	59.11	110.14
	MAE	17.16	47.37	91.71	15.74	47.49	121.36	16.42	50.44	117.27	17.2	49.69	94.8
	MAPE	5.47%	15.63%	29.92%	4.88%	15.84%	38.11%	5.15%	15.77%	35.5%	5.43%	15.66%	29.17%
Downstream Through	MSE	303.38	1075	2015	260.22	1064	2095	222.02	1014	2536	276.62	937.26	2100
	RMSE	17.42	32.79	44.89	16.13	32.63	45.78	14.9	31.85	50.36	16.63	30.61	45.83
	MAE	13.66	24.95	33.71	12.42	23.96	33.17	11.64	23.99	37.31	12.87	23.26	33.5
	MAPE	4.28%	8.0%	11.11%	3.82%	7.6%	10.94%	3.62%	7.55%	11.82%	4.02%	7.31%	10.57%
Downstream Turn	MSE	387.71	2384	7221	323.91	2756	9994	427.2	2959	9748	436.37	2089	6397
	RMSE	19.69	48.83	84.98	18.0	52.5	99.97	20.67	54.4	98.73	20.89	45.72	79.99
	MAE	15.38	40.31	72.78	14.15	42.63	84.75	16.63	44.93	83.55	16.51	37.01	66.13
	MAPE	4.96%	13.57%	24.37%	4.51%	14.0%	26.88%	5.19%	14.0%	25.66%	5.12%	11.58%	20.59%
Upstream, Retrain	MSE	295.95	971.59	1881	232.96	848.21	1771	207.44	782.5	1719	194.89	794.05	1777
	MAE	13.42	23.92	32.18	11.84	22.19	31.04	11.21	21.32	30.48	10.91	21.34	31.06
	MAPE	4.23%	7.66%	10.52%	3.73%	7.04%	10.03%	3.54%	6.8%	9.91%	3.42%	6.96%	10.55%
Upstream Through, Retrain	MSE	292.14	988.8	1907	237.99	917.54	1861	208.69	789.73	1722	189.3	728.53	1621
	RMSE	17.09	31.45	43.68	15.43	30.29	43.14	14.45	28.1	41.5	13.76	26.99	40.27
	MAE	13.39	24.01	32.25	12.0	22.93	31.89	11.27	21.4	30.58	10.75	20.66	29.99
	MAPE	4.21%	7.75%	10.91%	3.79%	7.5%	10.81%	3.58%	6.87%	9.98%	3.38%	6.67%	9.81%
Upstream Turn (Stopbar), Retrain	MSE	286.79	980.46	1890	231.73	859.58	1789	207.48	806.8	1752	185.91	717.07	1578
	RMSE	16.93	31.31	43.48	15.22	29.32	42.3	14.4	28.4	41.87	13.63	26.78	39.74
	MAE	13.28	23.72	31.79	11.89	22.29	31.05	11.23	21.52	30.76	10.69	20.47	29.82
	MAPE	4.18%	7.72%	10.74%	3.77%	7.29%	10.52%	3.53%	6.98%	10.45%	3.37%	6.57%	9.67%
Downstream, Retrain	MSE	290.82	979.63	1912	232.63	898.33	1846	202.48	808.79	1834	188.08	757.68	1721
	RMSE	17.05	31.3	43.74	15.25	29.97	42.98	14.23	28.44	42.83	13.71	27.53	41.49
	MAE	13.35	23.66	31.77	11.84	22.36	31.15	11.13	21.41	30.82	10.65	20.72	30.38
	MAPE	4.22%	7.63%	10.6%	3.7%	7.16%	10.39%	3.48%	6.91%	10.39%	3.27%	6.62%	10.23%
Downstream Through, Retrain	MSE	280.82	917.02	1789	224.54	807.04	1681	206.19	840.5	1889	189.77	777.99	1782
	RMSE	16.76	30.28	42.3	14.98	28.41	41.01	14.36	28.99	43.46	13.78	27.89	42.22
	MAE	13.13	23.25	31.31	11.65	21.61	30.13	11.21	21.72	31.34	10.73	20.87	30.45
	MAPE	4.16%	7.46%	10.19%	3.68%	6.85%	9.71%	3.52%	7.03%	10.56%	3.31%	6.59%	10.21%
Downstream Turn, Retrain	MSE	280.86	896.74	1756	229.39	868.21	1823	200.29	751.79	1657	189.42	785.04	1831
	RMSE	16.76	29.95	41.91	15.15	29.47	42.7	14.15	27.42	40.71	13.76	28.02	42.8
	MAE	13.12	22.95	30.91	11.78	22.24	31.2	11.05	20.9	30.12	10.77	20.89	30.61
	MAPE	4.17%	7.34%	10.01%	3.7%	7.18%	10.5%	3.49%	6.68%	9.79%	3.38%	6.77%	10.34%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 12: Comparison of DCRNN test prediction errors for unhealthy detector direction scenarios for off peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
None	MSE	521.04	2725	10481	485.99	2831	9832	379.28	2444	8958	347.63	2199	8348
	RMSE	22.83	52.2	102.38	22.05	53.22	99.16	19.48	49.44	94.65	18.64	46.89	91.37
	MAE	16.33	34.37	64.03	15.13	33.97	60.54	13.9	31.75	57.18	13.4	30.67	55.83
	MAPE	5.38%	11.33%	17.44%	5.14%	11.23%	16.81%	5.06%	10.52%	16.26%	4.58%	9.63%	15.67%
Upstream	MSE	5615	16843	36355	4320	13891	26010	6061	21124	40582	2265	7528	19387
	RMSE	74.93	129.78	190.67	65.73	117.86	161.28	77.86	145.34	201.45	47.6	86.77	139.24
	MAE	56.56	102.45	158.61	50.59	93.03	132.27	59.89	116.34	163.18	33.48	66.34	109.61
	MAPE	12.2%	22.95%	36.95%	12.13%	24.01%	35.51%	12.46%	23.63%	32.98%	8.05%	15.58%	22.98%
Upstream Through	MSE	1362	5880	13378	1066	4980	11473	1085	5144	12193	649.05	3018	8988
	RMSE	36.91	76.68	115.67	32.66	70.57	107.11	32.95	71.73	110.42	25.48	54.94	94.81
	MAE	27.02	57.36	86.74	23.76	51.75	77.53	24.58	52.63	79.09	18.19	38.71	63.35
	MAPE	7.2%	14.67%	21.67%	6.57%	14.12%	21.15%	6.53%	13.47%	19.17%	5.59%	11.48%	17.02%
Upstream Turn (Stopbar)	MSE	529.34	2814	11399	496.13	3048	11347	373.55	2438	9234	358.46	2203	8561
	RMSE	23.01	53.05	106.77	22.27	55.21	106.53	19.33	49.38	96.1	18.93	46.94	92.53
	MAE	16.49	35.09	69.42	15.35	36.27	69.82	13.81	31.95	59.88	13.75	31.28	58.05
	MAPE	5.58%	11.92%	20.03%	5.29%	12.27%	20.19%	4.94%	10.61%	17.76%	4.91%	10.45%	17.18%
Downstream	MSE	3542	18432	46540	2929	13507	47369	1822	11337	33847	3782	10911	28956
	RMSE	59.52	135.77	215.73	54.13	116.22	217.64	42.69	106.48	183.98	61.5	104.46	170.17
	MAE	43.39	104.55	182.5	43.13	99.15	188.67	32.18	85.83	156.21	44.51	82.03	143.48
	MAPE	11.3%	29.78%	56.92%	14.48%	37.38%	61.55%	8.75%	23.49%	41.59%	9.18%	17.94%	29.75%
Downstream Through	MSE	620.73	3297	11300	611.02	3301	11084	447.69	2552	8726	647.56	3627	9940
	RMSE	24.91	57.42	106.3	24.72	57.46	105.28	21.16	50.53	93.41	25.45	60.23	99.7
	MAE	17.96	41.14	75.01	17.97	41.43	75.22	15.29	34.06	62.18	18.9	44.85	72.45
	MAPE	5.68%	12.74%	21.59%	5.77%	13.85%	22.4%	5.03%	10.77%	17.54%	5.39%	11.78%	17.91%
Downstream Turn	MSE	1359	7831	23366	844.31	6222	23255	516.96	4130	14221	1001	5211	13105
	RMSE	36.87	88.5	152.86	29.06	78.88	152.5	22.74	64.27	119.25	31.64	72.19	114.48
	MAE	27.77	69.25	124.37	23.18	63.4	122.54	17.09	50.06	96.06	23.41	54.95	88.33
	MAPE	8.54%	22.87%	44.31%	10.13%	28.2%	46.46%	6.12%	16.5%	30.01%	6.19%	13.62%	20.7%
Upstream, Retrain	MSE	2336	29131	96774	442.62	11028	82302	460.22	4658	32624	4256	13432	45491
	RMSE	48.34	170.68	311.09	21.04	105.02	286.88	21.45	68.25	180.62	65.24	115.9	213.29
	MAE	36.57	160.41	288.83	15.06	89.31	249.89	14.79	52.18	149.96	48.13	90.9	174.49
	MAPE	9.25%	69.92%	126.23%	5.13%	52.29%	128.9%	4.91%	28.45%	75.02%	10.71%	20.0%	32.11%
Upstream Through, Retrain	MSE	1588	6616	14883	440.6	2842	9724	375.97	2760	10308	837.77	5070	13597
	RMSE	39.86	81.34	122.0	20.99	53.31	98.61	19.39	52.54	101.53	28.94	71.21	116.61
	MAE	29.22	61.21	91.38	14.73	34.28	60.84	13.8	35.13	65.69	21.62	53.76	82.57
	MAPE	7.54%	15.42%	20.64%	4.86%	10.93%	17.09%	5.06%	11.62%	19.29%	6.74%	14.94%	19.91%
Upstream Turn (Stopbar), Retrain	MSE	542.48	2816	10108	481.32	2994	10771	387.45	2508	8988	322.38	2240	8544
	RMSE	23.29	53.07	100.54	21.94	54.72	103.79	19.68	50.08	94.81	17.95	47.34	92.44
	MAE	16.56	35.21	63.75	15.06	34.53	63.0	14.05	33.13	59.19	12.79	30.61	56.52
	MAPE	5.42%	11.54%	17.47%	5.19%	11.25%	16.85%	4.97%	10.97%	17.73%	4.37%	9.61%	16.14%
Downstream, Retrain	MSE	541.23	2721	10504	5167	21385	52159	922.92	10350	18804	356.83	2147	12208
	RMSE	23.26	52.17	102.49	71.88	146.24	228.38	30.38	101.74	137.13	18.89	46.34	110.49
	MAE	16.54	34.5	63.98	55.76	125.02	204.72	24.75	86.23	106.46	13.58	31.65	82.86
	MAPE	5.41%	11.24%	17.11%	12.71%	28.68%	43.97%	10.55%	29.16%	33.27%	4.74%	10.01%	19.85%
Downstream Through, Retrain	MSE	525.84	2829	11545	439.05	2736	9493	381.16	2489	9758	450.2	2264	8754
	RMSE	22.93	53.19	107.45	20.95	52.31	97.43	19.52	49.9	98.79	21.22	47.59	93.57
	MAE	16.41	35.01	65.2	14.74	33.96	61.16	13.84	32.13	60.41	15.21	32.11	59.75
	MAPE	5.37%	11.62%	17.65%	5.04%	10.77%	16.78%	4.69%	10.25%	16.32%	5.13%	10.11%	16.31%
Downstream Turn, Retrain	MSE	2709	8954	21648	726.99	3665	15353	389.59	2330	8475	707.99	6136	19559
	RMSE	52.06	94.63	147.14	26.96	60.54	123.91	19.74	48.28	92.06	26.61	78.33	139.86
	MAE	37.72	74.78	121.09	19.74	41.75	79.14	14.05	31.66	59.04	18.83	51.24	82.41
	MAPE	8.56%	18.06%	28.77%	6.65%	13.9%	22.34%	4.85%	10.1%	15.67%	6.1%	15.14%	21.87%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m
Horizon													

Table 13: Comparison of DCRNN test prediction errors for unhealthy detector direction scenarios for the morning peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
None	MSE	330.6	942.6	1591	268.6	857.35	1448	227.86	757.32	1346	216.88	733.63	1401
	RMSE	18.18	30.7	39.89	16.39	29.28	38.06	15.1	27.52	36.69	14.73	27.09	37.44
	MAE	14.51	24.27	30.94	13.02	23.17	29.51	11.98	21.78	28.41	11.75	21.41	28.94
	MAPE	4.01%	6.69%	8.44%	3.62%	6.37%	8.09%	3.3%	6.0%	7.82%	3.26%	5.94%	8.0%
Upstream	MSE	4539	20118	35750	2806	9880	16441	2013	8750	23077	2935	14699	26095
	RMSE	67.37	141.84	189.08	52.97	99.4	128.22	44.87	93.54	151.91	54.18	121.24	161.54
	MAE	54.91	125.18	169.81	43.44	84.04	108.55	37.2	80.13	133.45	46.06	108.69	141.42
	MAPE	13.23%	30.89%	42.47%	10.53%	20.55%	26.67%	9.29%	20.0%	33.63%	11.33%	27.34%	35.23%
Upstream Through	MSE	1061	4240	5805	530.28	2551	4917	520.16	2067	4724	505.63	2442	4110
	RMSE	32.58	65.12	76.2	23.03	50.51	70.12	22.81	45.47	68.73	22.49	49.42	64.11
	MAE	25.86	51.85	60.87	18.39	41.82	58.79	18.26	37.14	56.96	17.98	40.17	53.44
	MAPE	6.54%	12.84%	15.34%	4.82%	10.82%	15.47%	4.79%	9.86%	15.42%	4.68%	10.37%	14.55%
Upstream Turn (Stopbar)	MSE	354.71	1078	1774	285.45	1036	1954	244.56	883.73	1797	230.62	866.45	1789
	RMSE	18.83	32.85	42.12	16.9	32.19	44.21	15.64	29.73	42.39	15.19	29.44	42.3
	MAE	15.01	26.16	32.85	13.43	25.45	34.8	12.39	23.66	33.41	12.11	23.36	33.29
	MAPE	4.12%	7.18%	9.02%	3.7%	6.99%	9.62%	3.43%	6.55%	9.3%	3.38%	6.54%	9.42%
Downstream	MSE	430.26	1666	7320	407.72	5173	18889	681.01	3269	15910	560.6	1488	4863
	RMSE	20.74	40.82	85.56	20.19	71.93	137.44	26.1	57.18	126.14	23.68	38.58	69.74
	MAE	16.47	33.1	73.11	16.14	63.47	128.69	20.98	47.6	115.73	19.07	30.36	55.89
	MAPE	4.51%	9.11%	19.23%	4.43%	16.79%	34.36%	5.61%	12.54%	30.64%	5.47%	8.6%	15.65%
Downstream Through	MSE	344.4	1059	1868	291.01	973.69	1685	249.95	851.32	1582	238.98	815.3	1724
	RMSE	18.56	32.55	43.22	17.06	31.2	41.06	15.81	29.18	39.78	15.46	28.55	41.53
	MAE	14.71	25.77	33.69	13.62	24.6	31.8	12.57	23.02	31.01	12.26	22.49	32.28
	MAPE	4.07%	7.14%	9.28%	3.76%	6.76%	8.7%	3.49%	6.38%	8.58%	3.39%	6.26%	8.97%
Downstream Turn	MSE	383.19	1423	3561	353.98	3002	9993	409.67	1645	6181	319.99	1366	4331
	RMSE	19.58	37.73	59.68	18.81	54.8	99.97	20.24	40.56	78.63	17.89	36.97	65.81
	MAE	15.56	30.27	48.96	15.12	45.89	90.45	16.33	32.76	67.36	14.23	29.73	54.85
	MAPE	4.3%	8.36%	13.15%	4.19%	12.16%	24.09%	4.56%	9.0%	18.18%	3.95%	8.22%	15.0%
Upstream, Retrain	MSE	29136	95673	133861	287.15	1408	4739	537.61	18827	47875	225.93	2678	60463
	RMSE	170.69	309.31	365.87	16.95	37.53	68.84	23.19	137.21	218.81	15.03	51.75	245.89
	MAE	164.86	298.92	355.11	13.44	29.62	55.56	18.75	128.72	207.14	12.0	41.69	225.68
	MAPE	44.31%	79.45%	94.5%	3.7%	7.85%	14.17%	5.22%	35.05%	55.99%	3.31%	12.06%	65.04%
Upstream Through, Retrain	MSE	475.18	1940	5264	465.78	4406	22398	235.24	992.25	4051	257.88	1382	7997
	RMSE	21.8	44.05	72.56	21.58	66.38	149.66	15.34	31.5	63.65	16.06	37.18	89.43
	MAE	17.41	35.13	58.56	17.5	57.04	137.35	12.19	24.95	53.46	12.72	29.23	73.68
	MAPE	4.81%	9.79%	16.4%	5.09%	16.87%	39.92%	3.38%	7.01%	15.03%	3.5%	8.06%	19.86%
Upstream Turn (Stopbar), Retrain	MSE	328.65	926.02	1574	264.05	859.23	1477	231.54	749.69	1338	219.64	815.6	1791
	RMSE	18.13	30.43	39.68	16.25	29.31	38.44	15.22	27.38	36.58	14.82	28.56	42.32
	MAE	14.45	24.17	30.74	12.91	23.21	29.78	12.04	21.59	28.16	11.85	22.61	33.14
	MAPE	3.98%	6.68%	8.44%	3.56%	6.37%	8.11%	3.32%	5.93%	7.71%	3.3%	6.35%	9.58%
Downstream, Retrain	MSE	330.62	997.65	1673	3378	11581	80409	240.24	1031	2783	218.05	1119	5947
	RMSE	18.18	31.59	40.91	58.12	107.62	283.57	15.5	32.11	52.76	14.77	33.45	77.12
	MAE	14.49	25.04	31.7	50.96	92.7	273.45	12.33	25.02	41.23	11.82	26.44	63.86
	MAPE	3.99%	6.89%	8.77%	14.82%	24.17%	73.13%	3.42%	6.69%	11.66%	3.27%	7.05%	16.79%
Downstream Through, Retrain	MSE	325.8	915.72	1554	261.42	837.92	1443	231.78	792.92	1421	233.86	909.07	1739
	RMSE	18.05	30.26	39.43	16.17	28.95	38.0	15.22	28.16	37.7	15.29	30.15	41.7
	MAE	14.39	23.97	30.53	12.86	22.96	29.6	12.11	22.31	29.06	12.19	23.93	32.75
	MAPE	3.96%	6.59%	8.34%	3.55%	6.34%	8.12%	3.35%	6.1%	7.94%	3.37%	6.55%	8.87%
Downstream Turn, Retrain	MSE	326.89	921.2	1505	266.62	873.39	1488	1980	31277	95184	211.9	760.05	1482
	RMSE	18.08	30.35	38.8	16.33	29.55	38.58	44.5	176.85	308.52	14.56	27.57	38.51
	MAE	14.42	24.07	30.17	12.95	23.41	29.87	39.4	169.34	298.07	11.66	21.84	29.67
	MAPE	3.98%	6.66%	8.26%	3.57%	6.43%	8.15%	10.9%	46.48%	84.85%	3.22%	6.0%	8.2%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 14: Comparison of DCRNN test prediction errors for unhealthy detector direction scenarios for the afternoon peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
None, Flow only	MSE	289.71	974.75	1842	235.67	920.15	1851	206.54	794.93	1774	246.19	802.61	1721
	RMSE	17.02	31.22	42.92	15.35	30.33	43.03	14.37	28.19	42.13	15.69	28.33	41.49
	MAE	13.33	23.52	31.32	11.9	22.51	31.21	11.22	21.48	30.79	12.22	21.68	30.6
	MAPE	4.18%	7.58%	10.53%	3.72%	7.26%	10.48%	3.54%	7.01%	10.44%	3.86%	6.94%	10.02%
Stopbar, Flow only	MSE	306.82	1061	1928	243.98	960.42	1925	215.7	869.86	1896	258.11	828.87	1742
	RMSE	17.52	32.58	43.91	15.62	30.99	43.88	14.69	29.49	43.55	16.07	28.79	41.74
	MAE	13.7	24.81	32.81	12.12	23.46	32.94	11.48	22.62	32.64	12.54	22.12	31.23
	MAPE	4.34%	8.12%	11.14%	3.79%	7.56%	11.01%	3.63%	7.39%	10.94%	3.95%	7.06%	10.29%
Advance, Flow only	MSE	583.11	3987	10643	563.4	3036	9338	1278	5945	13572	2309	4200	7126
	RMSE	24.15	63.15	103.17	23.74	55.1	96.64	35.75	77.11	116.5	48.06	64.81	84.42
	MAE	19.09	52.66	88.32	18.98	45.23	79.26	28.66	56.24	81.83	34.22	49.7	67.88
	MAPE	6.1%	16.64%	27.34%	6.21%	15.14%	26.18%	8.3%	16.56%	25.38%	9.4%	15.14%	22.17%
Stopbar, Flow only, Retrain	MSE	286.79	980.46	1890	231.73	859.58	1789	207.48	806.8	1752	185.91	717.07	1578
	RMSE	16.93	31.31	43.48	15.22	29.32	42.3	14.4	28.4	41.87	13.63	26.78	39.74
	MAE	13.28	23.72	31.79	11.89	22.29	31.05	11.23	21.52	30.76	10.69	20.47	29.82
	MAPE	4.18%	7.72%	10.74%	3.77%	7.29%	10.52%	3.53%	6.98%	10.45%	3.37%	6.57%	9.67%
Advance, Flow only, Retrain	MSE	298.39	1001	1893	236.82	878.57	1795	209.07	789.14	1729	191.68	771.4	1784
	RMSE	17.27	31.64	43.51	15.39	29.64	42.37	14.46	28.09	41.58	13.84	27.77	42.24
	MAE	13.52	24.12	32.32	11.97	22.43	31.39	11.26	21.43	30.83	10.83	21.1	31.14
	MAPE	4.25%	7.7%	10.74%	3.75%	7.21%	10.44%	3.53%	6.79%	9.92%	3.39%	6.86%	10.5%
None, Flow and Occupancy	MSE	306.62	1045	1997	242.99	889.57	1897	229.61	833.3	1736	221.86	773.11	1651
	RMSE	17.51	32.33	44.7	15.59	29.83	43.56	15.15	28.87	41.67	14.89	27.8	40.64
	MAE	13.46	24.13	32.5	12.04	22.53	31.73	11.65	21.93	31.29	11.43	21.27	30.68
	MAPE	4.24%	7.69%	10.55%	3.79%	7.2%	10.34%	3.68%	7.05%	10.22%	3.62%	6.89%	10.01%
Stopbar, Flow and Occupancy	MSE	1710	3578	6011	987.81	2406	4176	744.44	2491	5389	1101	3919	10519
	RMSE	41.36	59.82	77.54	31.43	49.06	64.63	27.28	49.92	73.41	33.19	62.61	102.57
	MAE	27.95	43.21	58.69	22.21	35.65	48.3	19.85	38.9	58.74	24.66	50.98	85.61
	MAPE	7.29%	11.98%	16.67%	6.11%	10.18%	14.24%	5.65%	11.64%	17.62%	6.98%	15.18%	24.75%
Stopbar, Flow and Occupancy, Retrain	MSE	312.57	961.24	1822	277.76	949.02	1900	255.0	865.79	1767	305.39	870.26	1643
	RMSE	17.68	31.0	42.69	16.67	30.81	43.59	15.97	29.42	42.04	17.48	29.5	40.54
	MAE	13.56	23.53	31.54	12.3	22.82	31.59	11.82	21.96	31.02	12.75	22.32	30.51
	MAPE	4.26%	7.52%	10.28%	3.84%	7.28%	10.28%	3.68%	6.98%	10.07%	4.06%	7.11%	9.92%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m
Horizon													

Table 15: Comparison of DCRNN test prediction errors for unhealthy detector type scenarios for off peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
None, Flow only	MSE	521.04	2725	10481	485.99	2831	9832	379.28	2444	8958	347.63	2199	8348
	RMSE	22.83	52.2	102.38	22.05	53.22	99.16	19.48	49.44	94.65	18.64	46.89	91.37
	MAE	16.33	34.37	64.03	15.13	33.97	60.54	13.9	31.75	57.18	13.4	30.67	55.83
	MAPE	5.38%	11.33%	17.44%	5.14%	11.23%	16.81%	5.06%	10.52%	16.26%	4.58%	9.63%	15.67%
Stopbar, Flow only	MSE	529.34	2814	11399	496.13	3048	11347	373.55	2438	9234	358.46	2203	8561
	RMSE	23.01	53.05	106.77	22.27	55.21	106.53	19.33	49.38	96.1	18.93	46.94	92.53
	MAE	16.49	35.09	69.42	15.35	36.27	69.82	13.81	31.95	59.88	13.75	31.28	58.05
	MAPE	5.58%	11.92%	20.03%	5.29%	12.27%	20.19%	4.94%	10.61%	17.76%	4.91%	10.45%	17.18%
Advance, Flow only	MSE	33244	70757	107324	25770	67927	114705	28429	51313	75882	42920	76633	164024
	RMSE	182.33	266.0	327.6	160.53	260.63	338.68	168.61	226.53	275.47	207.17	276.83	405.0
	MAE	135.2	213.14	274.69	121.08	212.8	285.96	123.59	179.63	232.6	156.81	230.24	348.72
	MAPE	25.27%	44.25%	64.44%	25.54%	51.12%	70.16%	22.6%	36.85%	51.65%	28.74%	44.98%	64.92%
Stopbar, Flow only, Retrain	MSE	542.48	2816	10108	481.32	2994	10771	387.45	2508	8988	322.38	2240	8544
	RMSE	23.29	53.07	100.54	21.94	54.72	103.79	19.68	50.08	94.81	17.95	47.34	92.44
	MAE	16.56	35.21	63.75	15.06	34.53	63.0	14.05	33.13	59.19	12.79	30.61	56.52
	MAPE	5.42%	11.54%	17.47%	5.19%	11.25%	16.85%	4.97%	10.97%	17.73%	4.37%	9.61%	16.14%
Advance, Flow only, Retrain	MSE	555.91	2833	9805	465.54	3051	11714	369.06	2345	8579	351.59	2399	9283
	RMSE	23.58	53.23	99.02	21.58	55.24	108.23	19.21	48.44	92.62	18.75	48.98	96.35
	MAE	17.2	36.97	64.49	15.07	34.86	63.28	13.69	31.54	56.7	13.42	32.41	60.41
	MAPE	5.59%	12.06%	19.42%	5.12%	11.04%	16.54%	4.73%	10.47%	16.38%	4.69%	10.5%	17.25%
None, Flow and Occupancy	MSE	588.97	3033	9908	501.47	3287	10631	496.57	2993	9865	691.94	3737	10755
	RMSE	24.27	55.08	99.54	22.39	57.33	103.11	22.28	54.71	99.33	26.3	61.14	103.71
	MAE	17.49	37.15	62.26	15.93	38.34	65.62	15.84	36.07	61.84	17.97	37.56	61.93
	MAPE	5.57%	11.4%	17.58%	5.24%	11.48%	17.5%	5.03%	10.75%	16.81%	5.88%	11.93%	16.92%
Stopbar, Flow and Occupancy	MSE	3156	9057	15645	2589	7901	14746	3803	9275	16938	6267	11612	18796
	RMSE	56.18	95.17	125.08	50.89	88.89	121.43	61.67	96.31	130.15	79.17	107.76	137.1
	MAE	38.73	71.04	95.62	34.35	64.17	91.99	42.19	69.64	98.64	52.93	80.19	106.47
	MAPE	8.5%	15.73%	22.83%	7.69%	14.79%	22.0%	8.59%	15.19%	22.36%	11.14%	20.41%	27.7%
Stopbar, Flow and Occupancy, Retrain	MSE	566.74	3272	10459	487.57	2925	9831	470.38	2773	9878	448.67	2936	10106
	RMSE	23.81	57.2	102.27	22.08	54.09	99.15	21.69	52.67	99.39	21.18	54.19	100.53
	MAE	16.99	38.32	65.11	15.74	36.15	62.62	15.61	33.74	59.79	15.02	35.36	61.98
	MAPE	5.38%	11.63%	18.03%	5.07%	10.92%	17.32%	5.03%	10.19%	16.18%	4.92%	10.76%	16.74%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 16: Comparison of DCRNN test prediction errors for unhealthy detector type scenarios for the morning peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
None, Flow only	MSE	330.6	942.6	1591	268.6	857.35	1448	227.86	757.32	1346	216.88	733.63	1401
	RMSE	18.18	30.7	39.89	16.39	29.28	38.06	15.1	27.52	36.69	14.73	27.09	37.44
	MAE	14.51	24.27	30.94	13.02	23.17	29.51	11.98	21.78	28.41	11.75	21.41	28.94
	MAPE	4.01%	6.69%	8.44%	3.62%	6.37%	8.09%	3.3%	6.0%	7.82%	3.26%	5.94%	8.0%
Stopbar, Flow only	MSE	354.71	1078	1774	285.45	1036	1954	244.56	883.73	1797	230.62	866.45	1789
	RMSE	18.83	32.85	42.12	16.9	32.19	44.21	15.64	29.73	42.39	15.19	29.44	42.3
	MAE	15.01	26.16	32.85	13.43	25.45	34.8	12.39	23.66	33.41	12.11	23.36	33.29
	MAPE	4.12%	7.18%	9.02%	3.7%	6.99%	9.62%	3.43%	6.55%	9.3%	3.38%	6.54%	9.42%
Advance, Flow only	MSE	27754	61899	86778	15696	44891	75657	9661	37882	66832	14352	41499	85650
	RMSE	166.6	248.8	294.58	125.28	211.88	275.06	98.29	194.63	258.52	119.8	203.71	292.66
	MAE	155.85	239.8	283.23	115.1	201.39	263.3	89.82	184.17	244.72	111.15	196.36	283.35
	MAPE	39.67%	62.93%	74.7%	28.96%	52.08%	68.96%	22.59%	47.54%	63.71%	28.22%	51.72%	75.41%
Stopbar, Flow only, Retrain	MSE	328.65	926.02	1574	264.05	859.23	1477	231.54	749.69	1338	219.64	815.6	1791
	RMSE	18.13	30.43	39.68	16.25	29.31	38.44	15.22	27.38	36.58	14.82	28.56	42.32
	MAE	14.45	24.17	30.74	12.91	23.21	29.78	12.04	21.59	28.16	11.85	22.61	33.14
	MAPE	3.98%	6.68%	8.44%	3.56%	6.37%	8.11%	3.32%	5.93%	7.71%	3.3%	6.35%	9.58%
Advance, Flow only, Retrain	MSE	19069	26305	27588	267.88	856.87	1439	268.72	890.98	1727	213.97	719.15	1334
	RMSE	138.09	162.19	166.1	16.37	29.27	37.94	16.39	29.85	41.56	14.63	26.82	36.54
	MAE	126.69	150.22	149.08	13.0	23.2	29.39	12.97	23.53	31.73	11.69	21.13	28.14
	MAPE	31.69%	37.95%	37.26%	3.59%	6.35%	8.02%	3.61%	6.55%	8.88%	3.23%	5.8%	7.73%
None, Flow and Occupancy	MSE	357.93	1062	1722	306.42	915.98	1599	283.82	891.11	1542	264.05	786.63	1412
	RMSE	18.92	32.6	41.51	17.5	30.27	39.99	16.85	29.85	39.27	16.25	28.05	37.58
	MAE	15.0	25.81	32.31	13.92	23.97	31.0	13.28	23.43	30.42	12.91	22.11	28.98
	MAPE	4.16%	7.09%	8.84%	3.83%	6.62%	8.52%	3.66%	6.41%	8.24%	3.57%	6.12%	7.99%
Stopbar, Flow and Occupancy	MSE	3315	10066	14238	1438	2811	6093	3953	10725	16863	4918	8638	16517
	RMSE	57.58	100.33	119.33	37.92	53.03	78.06	62.88	103.56	129.86	70.13	92.95	128.52
	MAE	46.33	85.36	101.01	29.68	42.39	63.07	53.85	91.11	115.2	56.66	78.82	111.02
	MAPE	11.16%	20.8%	24.66%	7.37%	10.73%	15.66%	13.48%	23.06%	29.34%	13.61%	19.32%	27.48%
Stopbar, Flow and Occupancy, Retrain	MSE	500.69	1807	3113	440.16	1588	2471	282.31	864.69	1551	844.78	3094	4247
	RMSE	22.38	42.52	55.8	20.98	39.86	49.71	16.8	29.41	39.39	29.07	55.62	65.18
	MAE	17.78	34.45	45.02	16.68	31.32	38.66	13.36	23.2	30.37	24.23	47.0	52.5
	MAPE	4.71%	9.02%	11.82%	4.67%	8.57%	10.54%	3.71%	6.34%	8.22%	7.24%	13.39%	15.4%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m

Table 17: Comparison of DCRNN test prediction errors for unhealthy detector type scenarios for the afternoon peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
None	MSE	289.71	974.75	1842	235.67	920.15	1851	206.54	794.93	1774	246.19	802.61	1721
	RMSE	17.02	31.22	42.92	15.35	30.33	43.03	14.37	28.19	42.13	15.69	28.33	41.49
	MAE	13.33	23.52	31.32	11.9	22.51	31.21	11.22	21.48	30.79	12.22	21.68	30.6
	MAPE	4.18%	7.58%	10.53%	3.72%	7.26%	10.48%	3.54%	7.01%	10.44%	3.86%	6.94%	10.02%
5% days	MSE	6943	7127	7398	6461	7305	8787	6409	6377	6877	6047	6139	6698
	RMSE	83.33	84.43	86.02	80.39	85.47	93.74	80.06	79.86	82.93	77.76	78.36	81.85
	MAE	30.48	39.73	46.6	28.48	39.11	48.51	27.78	36.8	44.99	28.21	36.75	44.82
	MAPE	9.03%	12.22%	14.88%	8.38%	12.04%	15.54%	8.2%	11.36%	14.4%	8.38%	11.3%	14.16%
10% days	MSE	14959	15271	15631	14051	14972	17196	14184	13990	13412	13409	12606	12819
	RMSE	122.31	123.58	125.03	118.54	122.36	131.14	119.1	118.28	115.81	115.8	112.28	113.22
	MAE	49.47	58.87	65.91	47.0	57.7	68.15	46.61	55.54	62.45	46.4	53.65	61.11
	MAPE	13.98%	17.24%	19.91%	13.2%	16.89%	20.68%	13.12%	16.29%	18.96%	13.11%	15.66%	18.3%
25% days	MSE	31827	31006	30053	30044	28184	26486	31136	29014	25176	29817	26819	25062
	RMSE	178.4	176.09	173.36	173.33	167.88	162.75	176.45	170.34	158.67	172.68	163.77	158.31
	MAE	95.13	103.17	109.2	91.66	98.81	104.97	93.08	100.11	103.27	91.2	95.18	99.78
	MAPE	28.28%	31.35%	33.73%	27.15%	29.95%	32.55%	27.58%	30.36%	31.99%	26.93%	28.34%	30.19%
50% days	MSE	64566	61113	55947	63645	59554	53368	63985	57816	51973	64057	58219	52410
	RMSE	254.1	247.21	236.53	252.28	244.04	231.02	252.95	240.45	227.98	253.1	241.29	228.93
	MAE	178.75	181.78	183.18	177.58	181.45	181.87	178.16	179.3	180.6	177.6	178.31	178.88
	MAPE	52.0%	54.0%	55.56%	51.72%	54.02%	55.47%	51.74%	53.09%	54.83%	51.41%	52.24%	53.45%
5% days, Retrain	MSE	6242	6758	8194	5943	6117	6732	6445	6069	5717	4017	3040	3540
	RMSE	79.01	82.21	90.52	77.1	78.21	82.05	80.28	77.91	75.61	63.38	55.14	59.5
	MAE	29.31	39.07	47.59	27.55	36.95	44.98	27.75	36.0	42.45	23.5	29.79	37.72
	MAPE	8.72%	12.13%	15.32%	8.13%	11.37%	14.52%	8.14%	11.04%	13.58%	6.94%	9.25%	12.27%
10% days, Retrain	MSE	14947	15238	15802	10550	8836	9042	11719	10403	10596	10851	8917	9569
	RMSE	122.26	123.44	125.71	102.72	94.0	95.09	108.26	102.0	102.94	104.17	94.43	97.83
	MAE	49.01	57.42	64.23	41.5	47.15	53.88	42.87	49.37	56.96	41.18	46.2	54.7
	MAPE	13.86%	16.7%	19.08%	11.65%	13.78%	16.43%	12.02%	14.44%	17.4%	11.49%	13.41%	16.69%
25% days, Retrain	MSE	32063	31815	31840	32021	31762	31825	32004	31715	31809	11745	8862	9625
	RMSE	179.06	178.37	178.44	178.95	178.22	178.4	178.9	178.09	178.35	108.38	94.14	98.11
	MAE	94.65	101.45	106.69	93.6	100.64	106.58	93.17	100.1	106.54	57.55	56.29	62.69
	MAPE	28.07%	30.49%	32.47%	27.75%	30.25%	32.46%	27.61%	30.08%	32.44%	16.19%	16.43%	18.64%
50% days, Retrain	MSE	53277	51093	60301	20565	17142	17309	66103	65323	65099	18409	5730	5445
	RMSE	230.82	226.04	245.56	143.41	130.93	131.57	257.11	255.59	255.15	135.68	75.7	73.8
	MAE	159.03	160.71	179.29	93.69	89.83	94.32	179.01	182.91	186.89	89.18	51.94	54.15
	MAPE	44.97%	45.88%	52.46%	25.27%	24.8%	26.68%	51.94%	53.59%	55.28%	24.29%	17.7%	19.78%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m
Horizon													

Table 18: Comparison of DCRNN test prediction errors for unhealthy proportion of days scenarios for off peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
None	MSE	521.04	2725	10481	485.99	2831	9832	379.28	2444	8958	347.63	2199	8348
	RMSE	22.83	52.2	102.38	22.05	53.22	99.16	19.48	49.44	94.65	18.64	46.89	91.37
	MAE	16.33	34.37	64.03	15.13	33.97	60.54	13.9	31.75	57.18	13.4	30.67	55.83
	MAPE	5.38%	11.33%	17.44%	5.14%	11.23%	16.81%	5.06%	10.52%	16.26%	4.58%	9.63%	15.67%
5% days	MSE	14641	17375	25541	14372	16907	24847	13496	15684	22663	14128	16236	22534
	RMSE	121.0	131.82	159.82	119.89	130.03	157.63	116.17	125.24	150.55	118.86	127.42	150.12
	MAE	38.96	57.03	87.83	37.61	57.03	86.58	35.54	52.96	78.53	36.0	52.95	78.05
	MAPE	9.96%	15.81%	23.0%	9.68%	16.07%	22.85%	9.27%	14.53%	20.9%	9.31%	14.43%	20.55%
10% days	MSE	28657	32034	40657	28228	31381	40856	26997	29608	37234	27829	30265	36696
	RMSE	169.29	178.98	201.64	168.01	177.15	202.13	164.31	172.07	192.96	166.82	173.97	191.56
	MAE	60.37	79.62	113.86	59.32	81.96	116.87	56.78	75.1	103.22	57.12	74.12	99.0
	MAPE	14.26%	20.38%	29.59%	14.26%	21.9%	30.55%	13.31%	18.72%	25.82%	13.5%	18.37%	24.01%
25% days	MSE	67695	73032	84780	66843	72835	88252	65637	70341	80496	66181	70204	79101
	RMSE	260.18	270.24	291.17	258.54	269.88	297.07	256.2	265.22	283.72	257.26	264.96	281.25
	MAE	126.0	148.44	188.25	126.49	156.46	198.23	122.26	144.05	177.03	122.37	141.1	168.4
	MAPE	30.27%	37.71%	50.39%	31.57%	42.6%	53.68%	28.85%	35.02%	43.84%	28.98%	33.25%	38.9%
50% days	MSE	132730	136562	140309	131101	137287	148521	132997	140446	150422	132719	140942	155808
	RMSE	364.32	369.54	374.58	362.08	370.52	385.39	364.69	374.76	387.84	364.31	375.42	394.73
	MAE	231.25	248.27	271.39	231.07	255.17	281.95	231.61	255.4	283.8	232.41	257.8	291.35
	MAPE	53.01%	55.98%	65.77%	53.98%	61.98%	71.26%	52.89%	57.37%	63.7%	52.72%	56.02%	60.88%
5% days, Retrain	MSE	518.32	2688	9286	442.7	2847	10106	380.14	2494	8918	14056	16671	24280
	RMSE	22.77	51.85	96.37	21.04	53.37	100.53	19.5	49.94	94.44	118.56	129.12	155.82
	MAE	16.18	34.26	61.25	14.75	34.27	62.16	13.73	31.83	57.07	35.46	53.75	80.33
	MAPE	5.29%	11.08%	17.27%	4.94%	10.98%	16.8%	4.83%	10.48%	16.42%	8.96%	13.91%	19.88%
10% days, Retrain	MSE	555.11	2834	10625	500.77	2865	9764	376.7	2314	8591	349.52	2267	8622
	RMSE	23.56	53.24	103.08	22.38	53.53	98.82	19.41	48.11	92.69	18.7	47.61	92.86
	MAE	16.68	35.26	63.64	15.23	33.82	59.98	13.87	31.07	56.37	13.45	31.09	56.46
	MAPE	5.37%	11.34%	17.1%	5.1%	11.04%	16.48%	4.88%	10.26%	16.19%	4.75%	10.09%	16.19%
25% days, Retrain	MSE	65707	67705	78239	450.88	2831	9959	59363	55690	60913	69018	68433	75535
	RMSE	256.33	260.2	279.71	21.23	53.21	99.8	243.65	235.99	246.81	262.71	261.6	274.84
	MAE	122.79	138.18	167.1	14.83	34.18	61.47	116.9	126.53	147.01	125.27	137.73	161.22
	MAPE	28.38%	31.39%	35.65%	4.97%	11.17%	17.1%	26.83%	28.84%	31.71%	29.41%	31.2%	34.28%
50% days, Retrain	MSE	579.94	2855	10267	135355	146649	163479	135327	146556	163226	135344	146603	163388
	RMSE	24.08	53.44	101.33	367.91	382.95	404.33	367.87	382.83	404.01	367.89	382.89	404.21
	MAE	17.1	35.83	64.84	232.14	255.43	286.99	231.7	254.66	285.54	231.9	254.26	284.23
	MAPE	5.52%	11.54%	17.79%	53.26%	55.86%	59.15%	53.11%	55.62%	58.69%	53.31%	55.62%	58.08%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m
Horizon													

Table 19: Comparison of DCRNN test prediction errors for unhealthy proportion of days scenarios for the morning peak.

Method	Metric	Window Size											
		15 min			30 min			1 hr			2 hr		
None	MSE	330.6	942.6	1591	268.6	857.35	1448	227.86	757.32	1346	216.88	733.63	1401
	RMSE	18.18	30.7	39.89	16.39	29.28	38.06	15.1	27.52	36.69	14.73	27.09	37.44
	MAE	14.51	24.27	30.94	13.02	23.17	29.51	11.98	21.78	28.41	11.75	21.41	28.94
	MAPE	4.01%	6.69%	8.44%	3.62%	6.37%	8.09%	3.3%	6.0%	7.82%	3.26%	5.94%	8.0%
5% days	MSE	6895	6742	7341	5591	5620	7206	7322	7422	7409	6890	6172	6383
	RMSE	83.04	82.11	85.69	74.78	74.97	84.89	85.57	86.15	86.08	83.01	78.56	79.9
	MAE	30.91	39.84	46.61	27.66	36.89	44.64	29.25	38.21	43.96	28.5	36.12	43.03
	MAPE	7.95%	10.42%	12.18%	7.1%	9.64%	11.74%	7.49%	9.97%	11.59%	7.31%	9.48%	11.45%
10% days	MSE	13380	13312	14103	11182	10954	13688	14415	14914	14289	13615	12650	12621
	RMSE	115.67	115.38	118.76	105.75	104.67	117.0	120.07	122.12	119.54	116.69	112.48	112.34
	MAE	47.58	57.58	66.2	43.2	52.2	62.1	46.75	56.22	61.41	45.59	53.01	59.83
	MAPE	12.33%	15.1%	17.44%	11.18%	13.7%	16.42%	12.11%	14.77%	16.31%	11.82%	13.98%	16.0%
25% days	MSE	40387	43465	47163	38586	40193	46635	43582	46923	45932	42768	45927	48388
	RMSE	200.97	208.48	217.17	196.44	200.48	215.95	208.77	216.62	214.32	206.8	214.31	219.97
	MAE	111.11	128.08	145.92	107.61	118.88	134.22	112.7	124.44	128.64	111.81	123.61	133.31
	MAPE	28.48%	33.27%	38.47%	27.56%	30.88%	35.47%	28.87%	32.25%	33.86%	28.67%	32.12%	35.2%
50% days	MSE	77080	80395	86854	82157	89188	97112	83024	85272	83734	81922	92675	106903
	RMSE	277.63	283.54	294.71	286.63	298.65	311.63	288.14	292.01	289.37	286.22	304.43	326.96
	MAE	201.14	221.11	246.42	207.09	224.32	243.83	206.86	214.98	217.43	206.46	227.05	252.9
	MAPE	52.85%	58.37%	65.7%	54.59%	59.68%	65.65%	54.48%	56.93%	58.06%	54.35%	60.21%	67.63%
5% days, Retrain	MSE	329.1	938.63	1611	267.98	867.17	1511	227.87	774.6	1441	218.8	720.91	1390
	RMSE	18.14	30.64	40.14	16.37	29.45	38.87	15.1	27.83	37.97	14.79	26.85	37.29
	MAE	14.47	24.28	31.19	13.01	23.32	30.03	12.02	22.02	29.36	11.78	21.19	29.0
	MAPE	3.99%	6.71%	8.55%	3.58%	6.38%	8.15%	3.32%	6.05%	8.09%	3.26%	5.83%	8.02%
10% days, Retrain	MSE	330.14	926.45	1548	266.69	867.45	1489	229.32	759.86	1345	213.52	741.76	1426
	RMSE	18.17	30.44	39.35	16.33	29.45	38.6	15.14	27.57	36.68	14.61	27.24	37.77
	MAE	14.49	24.11	30.57	12.94	23.3	30.05	12.05	21.67	28.25	11.62	21.41	29.11
	MAPE	4.01%	6.63%	8.34%	3.57%	6.42%	8.27%	3.34%	5.94%	7.77%	3.21%	5.94%	8.26%
25% days, Retrain	MSE	340.46	1019	1887	5891	2499	6055	7822	4595	6916	41430	41781	42608
	RMSE	18.45	31.92	43.45	76.76	49.99	77.82	88.44	67.79	83.17	203.55	204.4	206.42
	MAE	14.76	25.25	33.06	43.47	36.47	52.56	49.38	43.67	55.82	110.13	117.97	125.95
	MAPE	4.14%	7.21%	9.83%	11.14%	10.7%	14.09%	12.62%	11.58%	14.97%	28.34%	31.03%	34.52%
50% days, Retrain	MSE	78290	78425	78220	58413	39359	35858	8549	6738	5141	29364	17168	26057
	RMSE	279.8	280.04	279.68	241.69	198.39	189.36	92.46	82.09	71.7	171.36	131.03	161.42
	MAE	201.85	206.45	209.11	173.07	145.33	140.59	61.46	56.75	53.87	120.34	93.63	118.82
	MAPE	53.2%	54.68%	55.9%	44.99%	37.38%	36.32%	16.12%	19.44%	15.83%	30.52%	23.9%	30.42%
		5m	15m	30m	5m	15m	30m	5m	15m	30m	5m	15m	30m
Horizon													

Table 20: Comparison of DCRNN test prediction errors for unhealthy proportion of days scenarios for the afternoon peak.