

Uncovering Censorship: Poisson Regression for Estimating Latent Demand in EV chargers

Carl Johan Astrup {s203331}, Julian Róin Skovhus {s174045}

Abstract—The task of estimating electric vehicle (EV) charging demand presents substantial challenges as a consequence of data censoring in charging records. The censoring occurs due to the threshold imposed by the number of charging plugs at a charging station as observed demand cannot be higher than the available plug supply. Conventional predictive models are biased toward the supply of charging plugs and prove insufficient when attempting to extrapolate beyond the observed data. Consequently, there is a gap between the predicted demand and the true latent demand that can lead to ill-advised conclusions for infrastructure expansion strategies and charging supply management. To counter this problem, we propose to use a censorship-aware demand model based on censored Poisson likelihood functions adequate for discrete and censored data. We apply the likelihood function with spatiotemporal deep learning models and investigate their performance under varying conditions. Our methodology is tested on real-world data where the censorship-aware models exhibit superior performance to their censorship-unaware counterparts. We showcase how accounting for data censorship can significantly improve demand model applications for estimating utilization rates and queue lengths at charging stations.

Index Terms—censored regression, Poisson, EV charging, spatio-temporal, neural networks, queueing systems.

I. INTRODUCTION

THE climate crisis has become a global and collective problem that demands action on regional and international levels. In 2012, the contribution of the transportation sector to greenhouse gas emissions was estimated to be almost 20%, making it the largest stand-alone contributing sector [1]. Electrification of the vehicle fleet is cited as the most important factor in reducing emissions from transportation, while also improving public health and reducing operation costs [2]. In 2022, electric vehicles (EVs) accounted for 14% of worldwide automobile sales [3], and to fulfill the Paris Agreement's net-zero emission objective by 2050, it is estimated that EVs need to comprise 60% of global vehicle sales by 2030 [4], [5]. This growth indicates a projected surge in the EV market and, correspondingly, a tenfold increase in the demand for public charging infrastructure is expected by 2023 [6]. However, many potential customers cite range anxiety as a mitigating factor for acquiring an EV, and insufficient charging infrastructure is thought to be one of the major obstacles to widespread EV adoption [7]. Therefore, it is essential to exploit the vast amount of charging data gathered during the last decade to drive the forthcoming transformation of the transportation sector.

EV charging is usually categorized into four types: home, work, long-distance, or public charging [7]. Regardless of which type we consider, the deployment and operation of the stations are affected by various structural and technological factors such as electricity production, grid capacity, plug type, parking availability, and demand. In this work, we focus on public charging due to its important role in the charging infrastructure within urban areas [7]. A challenging problem for industry practitioners is to estimate the charging demand at existing and new charging station locations which is vital for decision-making. This is due to the phenomenon of data censorship which splits data into two sets of values: the observed and censored data. It occurs when data is clipped by lower, intervals, or upper thresholds [8]. It is not restricted to charging data but occurs in many fields. In the context of charging data, censorship occurs when no plugs are available, limiting the observed demand from charging records to data below the charging capacity (see Fig. 1). Traditional models, biased towards supply rather than demand, cannot extrapolate beyond the charging plug threshold and thus might give over-conservative estimates of the demand, making them suboptimal for stakeholder decision-making [9]. An accurate estimate of short- and long-term demand can assist EV charging operators in expansion strategies and implementation of dynamic pricing. Another emerging technology, that can benefit from accurate demand forecasts, is grid-free mobile EV charging stations that can be deployed to handle surges in demand and reduce pressure on the grid [10]. They can also be deployed temporarily at new locations to estimate demand at potential expansion sites.

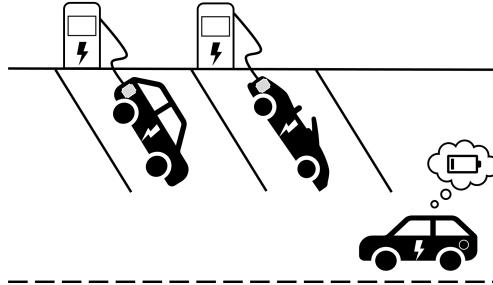


Fig. 1. An illustration of a customer arriving when every plug at a station is in use. The true demand is censored in the charging records since two charging events are noted, but the actual demand is three

To provide accurate demand forecasts and deal with the bias in the data induced by censorship, we employ censorship-aware models that can extrapolate the information from supply data to learn the latent demand. Rather than working directly with the consumption data, we propose to discretize

the data into charging events, i.e., counting the number of plugs in use at a point in time. A clear advantage of using discrete data for modeling charging demand is the ability to define a distinct censoring threshold from the number of available plugs and relate it to the observed number of charging events. Additionally, we can take advantage of the statistical properties of discrete count distributions to estimate utility rates and queue lengths for charging stations. We propose to use a censored Poisson regression model in a time-series context using past observations to predict current and future charging demand. We implement the neural network (NN) based models gated recurrent unit (GRU), long short-term memory (LSTM), and temporal graph convolutional network (T-GCN). For each model architecture, we develop a censorship-aware model, which distinguishes between observed and censored data, and a censorship-unaware version, treating observed and censored observations identically. We test various censoring schemes and forecast horizons to investigate the model performance for aware and unaware models. We explore several use cases of the aware models, including forecasting capabilities, estimation of utility, and queue length using queuing theory.

To summarize, our contributions, in order of importance, are:

- We propose to use a censorship-aware Poisson likelihood function to estimate charging demand by extrapolating information from the supply data of charging stations.
- Methodologies to assess the key parameters for stakeholders, i.e., utilization rates and queue lengths at charging stations, directly from charging records.
- A framework to preprocess censored data for modeling and training aware models.

The paper is structured as follows: in Section II, we provide an overview of the existing research on charging infrastructure and censored modeling. We describe the concepts of censored regression and graph neural networks for time series prediction in Section III. In Section IV, we present the construction of the Poisson likelihood function for NNs and graph neural networks (GNNs) and outline basic concepts of queuing systems for estimating queues at charging stations. We describe the data, censoring schemes, and model development and applications in Section V. The results and related discussion are covered in Section III, before moving on to a conclusion in Section VII and future work in Section VIII.

II. LITERATURE REVIEW

In this section, we cover the relevant literature and research on the main topics of this paper; EV charging infrastructure, model censorship, and graph representation learning.

A. Analyzing charging infrastructure

As a result of the rapidly expanding demand for EV charging, the topic of optimal placement of new charging stations has received substantial attention as a research topic in recent years [11]–[14]. The problem can roughly be separated into urban charging and long-trip charging although similar

methods can be applied. For long-trip charging, i.e., onset and destination differ, many studies employ traffic analysis to estimate traffic patterns and, in turn, the expected utility rates of potential stations [15]–[17]. Naturally, a higher traffic density will most likely lead to higher demand. In 2019, Fredriksson et al. [18] proposed a method for long-trip charging expansion. They formulated the problem for long-distance, non-urban charging, as a route network coverage problem with the objective of covering the most probable routes in a transportation network.

When urban charging distributors expand the existing charging capacity at current or new locations, various factors are considered, such as demand, convenience, cost, and clientele. One of the main methods mentioned in the literature for expansion is maximum expected cover models, which are widely used to analyze public facility locations [19]. This approach is similar to the route network coverage problem used for long-trip charging planning, but rather than identifying the most probable routes, other factors such as residents, offices, amenities, parking spaces, etc., are used to identify optimal locations. Although various formulations of this optimization problem exist, the overall objective is to minimize the distance to charging stations from any point in the city. A study in Lisbon applied the methodology to propose sites for the expansion of urban EV infrastructure by the maximum area covered within a given distance to the nearest station [20]. Similarly, the methodology has been adjusted and applied to Amsterdam and Brussels, where positional data from existing charging stations were used to compare with their modeled locations [21]. Certain station operators are known to also use energy modeling, proximity to amenities, and traffic analysis to evaluate locations and expected utility rates [22].

More recent studies have begun to leverage the vast amounts of data collected from public stations over the past decade to inform and guide decisions regarding the expansion of charging infrastructure [11]–[14]. Given charging records, utilization rates of the stations are a commonly used metric by charging distributors and governments to drive expansion strategies. From a customer perspective, the utilization should be as low as possible to ensure plug availability. From a business perspective, a higher utilization rate will lead to increased return on investment (ROI) for charging distributors. An optimal rate is dependent on location, charger type, electricity price, and expected ROI. According to a study by PwC, the optimal utilization rate balancing supply and demand for slow (level 2) chargers is around 25%–35% for a 10% ROI with a fixed electricity price [23]. Utility rates for public charging stations vary greatly from 10%–50% depending on location and nearby amenities [12]. Various European studies have found low mean utilization rates of public chargers between 15–20%, with the highest utilization in residential areas [13], [14]. The tendency towards expanding the charging network while the utilization rate is low is driven by the expected increase in EVs in the coming years. By acquiring market share now, the charging distributors ensure the ROI in the coming years when utilization rates presumably increase [24].

B. Censorship-aware models

Several of the mentioned articles cite difficulty in quantitatively estimating the expected demand in new and existing charging locations, both for urban and long-trip charging. This is caused by censorship of charging supply data, which occurs when no plugs are available at a charging station making the demand above the capacity unobservable through charging records. As a consequence, data censorship induces bias in the data with varying effects on predictive models [25]. Considering supply data for estimating demand censored above a threshold, ignoring censorship would lead to lower estimates than the true demand. In the context of time series predictions, naive approaches to deal with censorship involve preprocessing the data to exclude observations where censoring possibly occurred to deal with censorship bias [26]. However, excluding too many observations will result in performance degradation, thus inspiring research on more sophisticated methods. A second approach, commonly used for left-censored data in biology, uses non-parametric maximum likelihood estimation (MLE) to estimate censored values [27].

A third approach is censorship-aware models which attempt to distinguish between censored and uncensored observations assuming latent and observed data belong to the same distribution. They have been extensively studied for many years, with the Tobit model from 1958 being one of the most widely used today [28]. The model assumes that the underlying or latent variable is related to the explanatory features in a linear manner, incorporating Gaussian noise [28], and it has, in different variations, been used especially in survival analysis [29], [30]. Research has also been conducted in the area of credit risk management, specifically focusing on the unpredictability of “good” credit card users who may default in the future [31]. This presents a scenario of censored outcomes. Similarly, in retail, demand censoring occurs when customers’ potential to purchase more products is hampered due to stock-outs [32]. This is analogous to censoring induced by unavailable plugs.

Arrivals at EV charging stations have been modeled as a Poisson process in various studies [33], [34]. Specifically, arrivals are described as a non-homogenous process i.e. the rate parameter, λ , varies with time. It is essential to allow λ to change as charging demand is highly variational throughout the day and week (see Fig. 13 in Appendix C) [13], [14]. Censored modeling has been used to estimate arrivals at shared mobility hubs, which exhibit similar behavior to EV charging clusters which inspires us to attempt censored modeling for estimating latent charging demand [8], [9]. As the charging events are discrete and non-zero, Poisson regression appears to be a good choice of model. To the best of our knowledge, censorship-aware Poisson-Tobit-type models have *not* been employed to estimate latent EV charging demand [35].

C. Deep learning for EV charging data

Deep learning (DL) models have proven useful for predicting complex time series data with the introduction of

LSTM and GRU and provide an alternative to traditional autoregressive integrated moving average (ARIMA) models [36], [37]. Recently, Hüttel et al. [9] applied censored quantile regression on bike-sharing time series data using various DL architectures showing better performance than an autoregressive (AR) alternative. Hüttel et al. [11] also used GNN architectures to predict EV charging supply, outperforming non-spatial models, supporting the idea that spatial models are suitable for estimating EV charging at the station level. Both works inspire the idea to use spatiotemporal DL models for censored charging data but our approach differs in two significant ways. Primarily, rather than modeling the demand in electricity consumption, we focus on the demand as discrete charging events, thus increasing the interpretability related to the key parameters of the stakeholders, such as utilization rates and queue lengths. Secondly, we adopt the censorship-aware Poisson likelihood function to be used with GNNs to augment the temporal information with spatial dependencies, leading to better long-term forecasts.

III. BACKGROUND

In this section, we cover the key theoretical concepts related to censorship-aware models and deep learning (DL) in the context of time series.

A. Tobit-type models

Censoring occurs when we do not have access to the true measurement, denoted as y_i^* , but observe y_i instead. Data can either be left, right, or interval-censored, i.e., y_i^* is below a threshold τ_l , y_i^* is above a threshold τ_r , or y_i^* is between values τ_l and τ_r . For right-censored data, y_i is defined as:

$$y_i = \begin{cases} y_i^*, & \text{if } y_i^* < \tau_i \\ \tau_i, & \text{if } y_i^* \geq \tau_i \end{cases} \quad (1)$$

In the Tobit model for right-censored data, we are given a matrix of independent variables, denoted as $\mathbf{X} \in \mathbb{R}^{n \times k}$, where n represents the number of observations and k corresponds to the number of features. Each observation comprises a feature vector, \mathbf{x}_i . The latent, or true, variable, y_i^* , in which we are interested, is associated with each feature vector \mathbf{x}_i . The Tobit model uses a relationship between the latent variable y_i^* and the corresponding feature vector \mathbf{x}_i , which is conceptualized through some model $f_\theta(\mathbf{x}_i)$ and a normally distributed noise term ε_i with mean zero and variance σ^2 [8], [38],

$$y_i^* = f_\theta(\mathbf{x}_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \quad (2)$$

where θ denotes the model parameters.

An indicator variable d_i of whether a sample is censored or not is given as:

$$d_i = \begin{cases} 1, & \text{if } y_i \geq \tau_i \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Given n censored targets $\mathbf{y} \in \mathbb{R}^n$, the model parameters θ and the independent variables \mathbf{X} , the likelihood function for the Tobit Type-1 can be formulated.

All observations are assumed to be independent, and thus,

the probability density function (PDF) ϕ and cumulative distribution function (CDF) Φ of the normal distribution can be used to express the likelihood [28]:

$$\mathcal{L}(\theta|\mathbf{X}, \mathbf{y}) = \prod_{i=1}^n \left[\phi(y_i|f_\theta(\mathbf{x}_i))^{1-d_i} + (1 - \Phi(y_i|f_\theta(\mathbf{x}_i)))^{d_i} \right] \quad (4)$$

The logic behind this formulation is to maximize the simultaneous likelihood of a predicted value being equal to the observed non-censored value, and the likelihood of a predicted value exceeding or matching the censoring threshold. When an observation y_i is not censored, i.e., $d_i = 0$, only the densities in the first term of the likelihood function are activated. For maximum likelihood estimation, it can conveniently be expressed as a log-likelihood leading to the formulation [39]:

$$\log \mathcal{L}(\theta|\mathbf{X}, \mathbf{y}) = \sum_{i=1}^n [(1 - d_i) \log \phi(y_i|f_\theta(\mathbf{x}_i)) + d_i \log (1 - \Phi(y_i|f_\theta(\mathbf{x}_i)))] \quad (5)$$

B. Time series prediction

The charging events form a time series and the number of charging events y_T^* at time T are dependent on previous charging events. The dependency can be summarized as the conditional probability,

$$p(y_T^*|y_{T-1}^*, \dots, y_{T-t}^*) \quad (6)$$

where the index t denotes how many previous time steps still carry information about the number of events at time T . Since the emergence of RNNs, several alternative DL models have been developed to model complex time series data. A well-documented problem of the RNNs is the exploding and vanishing gradients, which often occur when the time series under study are complex. However, the more recent models, GRU and LSTM, mitigate this problem by using information gates, thus making DL suitable for modeling complex time series. We refer to Appendix A for a detailed description of their working principles.

C. Temporal graph neural networks

In this work, we use the temporal graph convolutional network (T-GCN) proposed by Zhao et al. for traffic prediction [40]. The hybrid temporal and spatial model is used to improve the forecasting of traffic speed on roads by taking into account the spatial dependence of the road network. The topology of the network is represented by an undirected graph $G = (V, E)$ with vertices V connected by edges E . T-GCN has an architecture with two graph convolutional network (GCN) layers followed by a GRU cell. We define a single layer of a Graph Convolutional Network as:

$$\mathbf{H}^{(l+1)} = act\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) \quad (7)$$

where $\mathbf{H}^{(0)} = \mathbf{X}$. \mathbf{X} is the feature matrix of the network with dimensions $\mathbf{X} \in \mathbb{R}^{p \times n \times k}$ where p is the number of nodes and n is the length of the time series. act denotes any activation

function, $\mathbf{W}^{(l)}$ is the weight-matrix for the l 'th layer, $\tilde{\mathbf{D}} = \sum_j \tilde{\mathbf{A}}_{ij}$ is the degree matrix of the graph with self-connections and $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_P$ is the adjacency matrix of the graph G , with added self-connections in form of the identity matrix \mathbf{I}_P [41]. The adjacency matrix denotes the weight between nodes and neighboring nodes. In the T-GCN architecture, it is defined before training using distance, correlation, or other metrics [11]. Other GNN architectures, such as graph attention networks, do not use adjacency matrices to denote the spatial dependencies but rather use attention coefficients through a feedforward NN [42].

The GCN layer is used to create an embedding of the topological layout of the graph, where the GRU is utilized to track the dynamic variation in time series data within the clusters, thereby modeling temporal dependence. A depiction of the model is shown in Fig. 2. The overall structure of

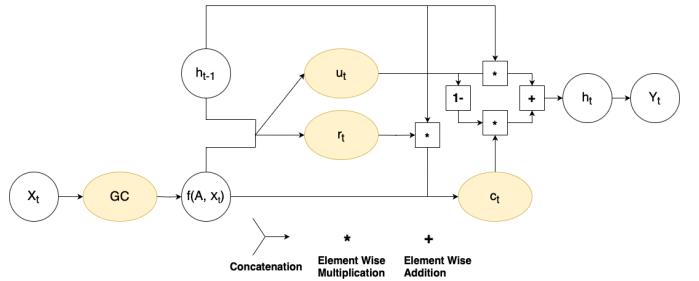


Fig. 2. T-GCN architecture visualized with an input X_t being the node attributes at time t . GC notates a graph convolution layer and $f(A, X_t)$ notates the embedding created by the GCN layer

the model is recurrent with hidden states being sequentially computed for each sample in the sequence. The states are propagated to the following cell for the next sample in the sequence. Inside the T-GCN cell, the nodes' features are fed to the GC layer where the graph embeddings are created and passed on to the GRU cell. This is done for each sequence in the time series, which in the end, yields a final hidden state. The hidden state is passed through a linear layer to produce the prediction. We define $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}$ as a preprocessing step. In the case of GCNs, two layers are typically used,

$$f(\mathbf{A}, \mathbf{X}) = act\left(\hat{\mathbf{A}} \text{ ReLu}\left(\hat{\mathbf{A}}\mathbf{X}\mathbf{W}^{(0)}\right)\mathbf{W}^{(1)}\right) \quad (8)$$

where ReLu is the rectified linear unit. We can then define the GRU cells,

$$\begin{aligned} \mathbf{u}_t &= \sigma_s(\mathbf{W}_u[f(\mathbf{A}, \mathbf{X}_t), \mathbf{h}_{t-1}] + \mathbf{b}_u) \\ \mathbf{r}_t &= \sigma_s(\mathbf{W}_r[f(\mathbf{A}, \mathbf{X}_t), \mathbf{h}_{t-1}] + \mathbf{b}_r) \\ \mathbf{c}_t &= \tanh(\mathbf{W}_c[f(\mathbf{A}, \mathbf{X}_t), (\mathbf{r}_t * \mathbf{h}_{t-1})] + \mathbf{b}_c) \\ \mathbf{h}_t &= \mathbf{u}_t * \mathbf{h}_{t-1} + (1 - \mathbf{u}_t) * \mathbf{c}_t \end{aligned}$$

where σ_s denotes the sigmoid activation function, \tanh denotes the hyperbolic tangent activation function, and $*$ denotes the element-wise multiplication.

IV. METHODOLOGY

In this section, we cover the construction of the likelihood functions and a brief introduction to queuing systems.

A. Poisson regression

The arguments for using discrete charging sessions rather than electricity consumption for estimating demand are the following. Charging consumption data is highly dependent on the type of car, the type of charging station, and the load on the electrical grid. However, a metric that is independent of the aforementioned characteristics, is the usage of available plugs, and can readily be related to demand. Additionally, a censoring threshold can easily be defined from the charging capacities. Charging events are discrete and non-negative and thus Poisson regression is a suitable model [43].

We describe the number of charging events in time interval i as $y_i \sim \text{Poisson}(\lambda_i)$, where the events are assumed to be independent and have the following probability mass function (PMF) f [43]:

$$f(y_i = k; \lambda_i) = P(y_i = k | \lambda_i) = \frac{\lambda_i^k e^{-\lambda_i}}{k!}, \quad \text{for } k = 0, 1, \dots \quad (9)$$

where $\lambda_i > 0$ is the mean or rate parameter for the Poisson distribution. The CDF is given as:

$$F(y_i \leq k; \lambda_i) = \sum_{j=0}^{\lfloor k \rfloor} P(y_i = j | \lambda_i) = e^{-\lambda_i} \sum_{j=0}^{\lfloor k \rfloor} \frac{\lambda_i^j}{j!} \quad (10)$$

To model a Poisson distribution target y_i^* , we use Poisson regression, which is a Generalized Linear Model using the log link function. The OLS is substituted with some model $f_\theta(x_i)$, so we have

$$\begin{aligned} \log y_i &= f_\theta(x_i) \\ y_i &= \exp\{f_\theta(x_i)\} \end{aligned} \quad (11)$$

where θ is estimated using MLE to maximize the Poisson log-likelihood function,

$$\begin{aligned} L(\theta | \mathbf{x}, \mathbf{y}) &= \prod_{i=1}^n f(y_i | \exp\{f_\theta(x_i)\}) \\ \log L(\theta | \mathbf{X}, \mathbf{y}) &= \sum_{i=1}^n \log f(y_i | \exp\{f_\theta(x_i)\}) \end{aligned} \quad (12)$$

B. Censored likelihood function

For censored Poisson regression, a modified likelihood is needed to reflect the probability density of the unobserved values $y_i^* \geq \tau_i$. We use the Tobit likelihood function from eq. (4) modified to Poisson regression [35]:

$$\begin{aligned} \mathcal{L}^*(\theta | \mathbf{X}, \mathbf{y}) &= \prod_{i=1}^n \left[f(y_i | f_\theta(x_i))^{1-d_i} \right. \\ &\quad \left. + (1 - F(y_i - 1 | f_\theta(x_i)))^{d_i} \right] \end{aligned} \quad (13)$$

Similarly to eq. (5), the censored Poisson log-likelihood is given as:

$$\begin{aligned} \log \mathcal{L}^*(\theta | \mathbf{X}, \mathbf{y}) &= \sum_{i=1}^n [(1 - d_i) \log f(y_i | f_\theta(x_i)) \\ &\quad + d_i \log (1 - F(y_i - 1 | f_\theta(x_i)))] \end{aligned} \quad (14)$$

where $f(\cdot)$ is the Poisson PMF, $F(\cdot)$ is the Poisson CDF and d_i is the indicator defined in eq. (3). Notice that we are using $1 - F(y_i - 1 | \hat{y}_i)$ as we are dealing with a discrete distribution. We use MLE to train the neural network $\text{NN}(\mathbf{x}_i | \theta) = f_\theta(\mathbf{x}_i)$ by minimizing the negative log-likelihoods using backpropagation.

C. Constructing the graph

We construct the graph letting each node V represent a charging cluster and the edges E represent the connectivity between the clusters. The edge weight, stored in the adjacency matrix \mathbf{A} between nodes V_i and V_j , is determined by the Haversine distance $h(\cdot)$ [11]:

$$\mathbf{A}_{i,j} = \begin{cases} \exp(-h(V_i, V_j)) & \text{if } h(V_i, V_j) < d_{min} \\ 0 & \text{otherwise} \end{cases}$$

where d_{min} is the minimum distance required for two nodes to be connected. The adjacency matrix \mathbf{A} controls the weight applied to information from neighboring nodes during the graph convolutions in eq. (7).

We extend the architecture of the T-GCN model to integrate the censored likelihood functions to make the T-GCN censorship-aware. Specifically, a rate parameter, λ of a Poisson distribution is estimated for each time step, for each node in the graph. Consequently, the log-likelihood for graphs becomes:

$$\begin{aligned} \log \mathcal{L}(\theta | \mathbf{X}, \mathbf{y}, G) &= \sum_{i=1}^n \sum_{v=1}^{|V|} [(1 - d_{i,v}) \log f(y_{i,v} | f_\theta(\mathbf{x}_{i,v})) \\ &\quad + d_{i,v} \log \{1 - F(y_{i,v} - 1 | f_\theta(\mathbf{x}_{i,v}))\}] \end{aligned} \quad (15)$$

where $|V|$ denotes the number of nodes in the graph G .

D. Simulating customer arrivals

An application of an EV charging demand model is to estimate queue lengths at charging stations. Queueing occurs if no plugs are available and new customers arrive that cannot be attended to. We estimate queue length as the difference between the number of available plugs, τ , and the model prediction \hat{y} :

$$QL = \begin{cases} \hat{y} - \tau & \text{if } \hat{y} - \tau > 0 \\ 0 & \text{otherwise} \end{cases}$$

We apply queuing systems to estimate queue length by simulating the charging network during a day. Specifically, we use an M/M/c queue to simulate a queuing system where customers arrive at the charging clusters according to a Poisson process and are attended to with exponential service times at c plugs operating in parallel [44]. The arrival rates are estimated from the model predictions as the difference in the number of charging sessions between time steps. The service times are estimated from the plug-in time, which is given in the data. Arrival rates and service times exhibit strong daily patterns and therefore we sample the values depending on the time

of day. We do not assume customers await the availability of slow chargers and thus, if unattended to upon arrival due to unavailability, customers are rejected rather than included in a queue and demand will consequently be lost. A depiction of the M/M/c system with four charging stations with rejections is illustrated in Fig. 3. However, queuing behavior is prevalent at fast-charging stations [45]. Therefore, a distinct M/M/c system, capable of handling priority queues and estimating waiting times, is necessary for fast-charging stations. We do not explore M/M/c queuing systems for fast-charging stations further in this work.

An overview of the pseudo-code of the simulation is provided in Appendix A-C.

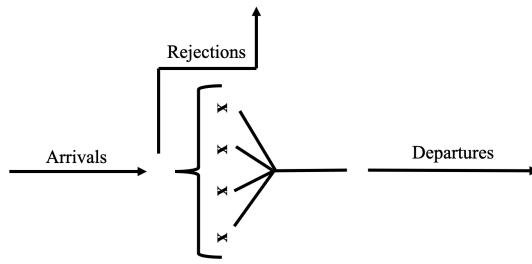


Fig. 3. M/M/c system with 4 plugs. If a customer arrives at a time with all four plugs busy, then the customer is rejected

V. EXPERIMENTAL DESIGN

In this section, we outline the design of our experiments to develop, test, and validate applications of censorship-aware models.

A. Data and preprocessing

We conducted our experiments using an open-source dataset containing charging records obtained from 37 EV-charging plugs, distributed at eight addresses in Palo Alto, CA [46]. The data contain static information related to the stations and dynamic information like plug-in start time, plug-in end time, charging time, electricity consumption, etc. All charging stations offer 208-240V electric services and charging times from empty to 80% in 4-10 hours for battery electric vehicles [47]. They are usually referred to as type level 2 charging stations.

The discretization of the data is done by counting unique charging events within 30-minute time intervals. An in-depth description of this can be found in Appendix C. We spatially aggregate the charging plugs into charging clusters based on their address to achieve a total number of eight clusters. The stations in a cluster are within a radius of a few meters from each other and we assume they exhibit similar usage patterns. By aggregating the clusters, the capacity increases from one or two to the sum of the plugs of each station in a cluster. Increasing elements of the set of the dependent variable, we provide the model with more information on the overall usage patterns than if we strictly performed binary or tertiary regression. Inspecting the distance between clusters they can further be separated into two “superclusters” containing

Webster, Bryant, Hamilton, and High in the first supercluster and Ted and Cambridge in the second. Rinconada and MPL remain “isolated”. Based on the distance, the charging clusters, as nodes, are connected by vertices if they are within a radius of 1.5 km (see Fig. 4). A discussion of the construction of the graph is presented in Section VIII.

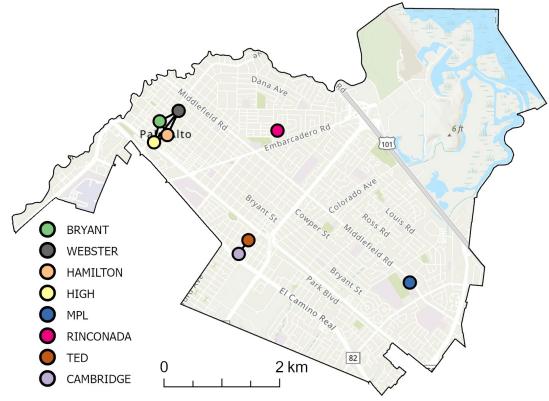


Fig. 4. Charging clusters from the Palo Alto dataset with graph edges between the clusters shown

B. Censoring for model development

Ideally, we would want access to the genuine demand for EV charging, unimpeded by any real-world censoring. However, such a scenario is almost unattainable since historical data inevitably contains some level of inherent censorship. Therefore, for the purposes of our experiment, we operate under the assumption that the existing historical data accurately represents the true demand y_i^* . This approach allows us to manually introduce censorship into the data and assess our trained models to the true, unobserved demand y_i^* . Given a chosen censoring threshold characterized by the number of charging plugs τ_i , we observe the true observation y_i^* if $y_i^* < \tau_i$, otherwise τ_i . We discuss the implications of our assumptions in Section VI-A.

We apply four different censoring methods, which we can characterize into Static and Dynamic methods. In the former, the plug capacity is kept at a censor threshold τ . In the latter, the plug capacity is reduced with an integer δ below the plug capacity. In Table I, we summarize the censoring configurations that we test for each cluster. For each censor strategy, the percentages of observations that are equal to or above the capacity imposed by the censoring threshold are shown.

TABLE I
CENSORING PERCENTAGES FOR EACH CLUSTER IN THE TEST DATA, FOR EACH CENSORING STRATEGY, TOGETHER WITH MEAN CENSORSHIP FOR EACH STRATEGY. THE NUMBER AFTER THE CENSOR STRATEGY INDICATES τ OR δ DEPENDING ON THE STRATEGY

	Bryant	Cambridge	Hamilton	High	MPL	Rinc.	Ted	Webster
Capacity	9	10	3	8	4	3	8	6
Dyn. 1	0.054	0.002	0.502	0.046	0.115	0.216	0.034	0.167
Dyn. 2	0.094	0.049	*	0.105	0.304	*	0.062	0.265
Static 3	0.382	0.388	*	0.368	0.115	*	0.267	0.366
Static 2	0.485	0.514	*	0.471	0.304	*	0.400	0.477

For example, the censor scheme Static 2 reduces the number of plugs available at every cluster to two. The dynamic scheme Dynamic 2 reduces the number of plugs available at a cluster with six plugs to four and at a cluster with four plugs to two. Since there are rarely any charging events during the night, censoring percentages around 30% actually correspond to around 50% of observations during the day being censored assuming that nighttime low periods last 8-10 hours. The thresholds are ranging from little to no censoring to moderate censoring around 30%. Dynamic 1 and 2 exhibit similar censor percentages except for clusters MPL and Webster, which see a large increase in censorship.

C. Model development

When defining the T-GCN model for our charging network, we chose to only use one GCN layer, as the resulting graph is small. An increase in the number of layers could result in an aggregation of information from every node within each individual node. This extensive aggregation poses the risk of diluting relevant node information, even for very distinct nodes, thus potentially compromising the efficacy of the model predictions [48].

The censorship-aware and unaware models are trained by minimizing the negative log-likelihood defined in equation (14) and (12) respectively. L2-regularization $\|\mathbf{w}\|_2^2$ is added to the loss function to avoid overfitting. For the development process, the data is split into three distinct subsets: training data, validation data, and test data in a split of 70/20/10 percent. The training set is used to train the weights of the models and build predictive capabilities. The validation set is used to tune hyperparameters and to ensure that the model is not over- or underfitting during the training phase. The test set is for evaluating the final performance of our model. It is an unbiased assessment of the model's generalization capabilities and accurately represents how well the model is expected to perform on unseen data. To optimize the models, Bayesian hyperparameter tuning was performed for every censorship-aware model architecture [49]. For a brief introduction to Bayesian Optimization, see Appendix B. The configurations of the models are summarized in Table V in Appendix B.

Every aware and unaware model architecture was evaluated on the true “test” loss, referring to the uncensored observations y_i^* . We also measure the metrics mean squared error (MSE) and mean absolute error (MAE) between the predictions \hat{y}_i and the true, uncensored observations y_i^* :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i^* - \hat{y}_i)^2 \quad (16)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i^* - \hat{y}_i| \quad (17)$$

The DL models T-GCN, GRU, and LSTM are compared to an AR model trained with a NN. All models are trained on an NVIDIA Tesla V100 32GB GPU. Our implementation of

the models can be found at <https://github.com/Fiskehandleren/Thesis>.

D. Model applications

After assessing model performance across architectures, censorship levels, and forecast leads, we conduct several experiments to explore applications of censorship-aware models in the domain of EV charging. The main experiments are summarized below:

- **Day-ahead forecasts:** Ability of censorship-aware models to forecast demand 30 minutes and 24 hours ahead
- **Utilization rate:** Calculate utilization rates of charging clusters to assess the model's ability to provide insight into the expansion of charging infrastructure at current and new locations
- **Lost demand and “queue lengths”:** Using properties of discrete data to estimate lost demand using model predictions and queuing systems

The first experiment is used to assess the capabilities of aware models for increasing forecast leads. Being able to predict several hours and days ahead can enable charging operators to use dynamic pricing to even out utility across clusters. Mobile charging trucks can also be used to deal with excess demand given that the peak hours and excess plug quantity are known in advance [50]. The second experiment compares the capabilities of the aware and unaware models in estimating the utilization rate, which is a key value used by charging distributors to manage plug capacities. The utilization rate at a cluster k is calculated using the estimated charging events \hat{y} and the plug capacity p for every time step t in a day T :

$$U_{k,T} = \frac{\sum_{i=1}^t \hat{y}_{k,i}}{\sum_{i=1}^t p_{k,i}} \quad (18)$$

To produce a utilization plot, the average across all days in the test set, for different hypothetical cluster capacities, is calculated. The final experiment focuses on estimating the lost demand as covered in Section IV-D.

VI. RESULTS

In this section, we initially evaluate the performance of the censorship-aware models against each other and their unaware counterparts, on the different censoring strategies defined. We then present the model applications mentioned in Section V-D. In the following, we refer to the censorship-aware T-GCN as T-GCN-A, while T-GCN will denote the censorship-unaware T-GCN.

A. Model performance and robustness

We compare the model performance for forecast leads of 30 minutes and 24 hours for the four censoring schemes. We also compare the aware and unaware models. The mean and standard deviation of losses and metrics from four model runs are presented in Table VII. The key takeaway is the identical or lower losses of the censorship-aware models compared to their unaware counterparts. Dynamic 1 and

TABLE II

LOSS TABLE FOR 30-MINUTE AND 24-HOUR FORECAST LEADS. WE DENOTE THE MEAN AND STANDARD DEVIATION FROM FOUR DIFFERENT WEIGHT INITIALIZATIONS AND TRAINING SET SHUFFLES. BOLD METRICS COMPARE AWARE AND UNAWARE MODEL VERSIONS, WHILE UNDERLINED METRICS HIGHLIGHT THE TOP MODEL FOR EACH CENSOR SCHEME

FL	Censoring	Model	CPNLL		PNLL		MAE		MSE	
			Aware	Unaware	Aware	Unaware	Aware	Unaware	Aware	Unaware
1	Dynamic 1	T-GCN	0.97 ± 0.00	—	−0.09 ± 0.00	−0.08 ± 0.01	0.68 ± 0.02	0.68 ± 0.01	1.10 ± 0.05	1.08 ± 0.03
		LSTM	0.93 ± 0.00	—	−0.10 ± 0.01	−0.14 ± 0.00	0.68 ± 0.03	0.56 ± 0.00	1.04 ± 0.1	0.63 ± 0.01
		GRU	0.88 ± 0.00	—	−0.16 ± 0.01	−0.19 ± 0.12	0.58 ± 0.01	0.51 ± 0.02	0.82 ± 0.04	0.58 ± 0.04
		AR	1.12 ± 0.04	—	0.17 ± 0.03	−0.02 ± 0.01	1.21 ± 0.05	0.79 ± 0.03	6.49 ± 1.98	1.87 ± 0.21
	Dynamic 2	T-GCN	0.90 ± 0.02	—	−0.07 ± 0.01	−0.07 ± 0.00	0.73 ± 0.03	0.70 ± 0.00	1.30 ± 0.09	1.14 ± 0.01
		LSTM	0.85 ± 0.00	—	−0.09 ± 0.00	−0.13 ± 0.00	0.71 ± 0.01	0.58 ± 0.00	1.11 ± 0.06	0.68 ± 0.00
		GRU	0.82 ± 0.01	—	−0.15 ± 0.01	−0.21 ± 0.06	0.61 ± 0.02	0.54 ± 0.01	0.91 ± 0.02	0.67 ± 0.01
		AR	0.94 ± 0.02	—	0.09 ± 0.07	−0.05 ± 0.00	1.07 ± 0.12	0.70 ± 0.01	3.23 ± 0.95	0.98 ± 0.07
	Static 3	T-GCN	0.67 ± 0.01	—	−0.07 ± 0.00	0.06 ± 0.01	0.73 ± 0.01	0.84 ± 0.00	1.30 ± 0.03	2.02 ± 0.03
		LSTM	0.63 ± 0.00	—	−0.06 ± 0.01	0.00 ± 0.00	0.80 ± 0.02	0.80 ± 0.00	1.39 ± 0.07	1.67 ± 0.01
		GRU	0.59 ± 0.00	—	−0.13 ± 0.01	−0.01 ± 0.01	0.66 ± 0.02	0.78 ± 0.01	1.12 ± 0.08	1.84 ± 0.04
		AR	0.66 ± 0.01	—	0.10 ± 0.03	0.01 ± 0.00	1.16 ± 0.07	0.80 ± 0.01	3.52 ± 0.50	1.53 ± 0.01
	Static 2	TGCN	0.51 ± 0.00	—	−0.02 ± 0.01	0.24 ± 0.01	0.83 ± 0.02	0.98 ± 0.00	1.59 ± 0.05	2.79 ± 0.02
		LSTM	0.50 ± 0.00	—	−0.01 ± 0.01	0.21 ± 0.01	0.89 ± 0.02	0.98 ± 0.01	1.73 ± 0.09	2.66 ± 0.03
		GRU	0.46 ± 0.01	—	−0.09 ± 0.02	0.20 ± 0.01	0.75 ± 0.02	0.93 ± 0.03	1.41 ± 0.10	2.75 ± 0.08
		AR	0.51 ± 0.01	—	0.14 ± 0.04	0.16 ± 0.01	1.21 ± 0.08	0.95 ± 0.01	3.97 ± 0.80	2.36 ± 0.06
48	Dynamic 1	T-GCN	1.20 ± 0.01	—	0.11 ± 0.00	0.24 ± 0.02	0.91 ± 0.01	0.98 ± 0.01	1.66 ± 0.03	2.16 ± 0.05
		LSTM	1.24 ± 0.01	—	0.16 ± 0.01	0.14 ± 0.57	0.96 ± 0.02	0.93 ± 0.20	1.79 ± 0.07	1.65 ± 0.82
		GRU	1.25 ± 0.06	—	0.18 ± 0.06	0.11 ± 0.01	0.95 ± 0.04	0.88 ± 0.01	1.89 ± 0.21	1.53 ± 0.05
		AR	1.29 ± 0.03	—	0.24 ± 0.05	0.23 ± 0.01	1.16 ± 0.09	0.99 ± 0.03	2.93 ± 0.69	2.15 ± 0.21
	Dynamic 2	T-GCN	1.14 ± 0.02	—	0.12 ± 0.01	0.25 ± 0.02	0.91 ± 0.00	0.99 ± 0.01	1.71 ± 0.06	2.20 ± 0.04
		LSTM	1.19 ± 0.03	—	0.18 ± 0.03	0.15 ± 0.00	0.96 ± 0.03	0.94 ± 0.01	1.84 ± 0.14	1.67 ± 0.02
		GRU	1.20 ± 0.04	—	0.20 ± 0.04	0.11 ± 0.01	0.97 ± 0.03	0.88 ± 0.01	1.97 ± 0.19	1.53 ± 0.05
		AR	1.19 ± 0.02	—	0.21 ± 0.03	0.17 ± 0.02	1.10 ± 0.05	1.00 ± 0.01	2.47 ± 0.32	1.85 ± 0.07
	Static 3	T-GCN	0.88 ± 0.01	—	0.13 ± 0.01	0.38 ± 0.02	0.92 ± 0.01	1.11 ± 0.01	1.73 ± 0.01	2.88 ± 0.05
		LSTM	0.91 ± 0.01	—	0.17 ± 0.02	0.29 ± 0.02	0.99 ± 0.02	1.08 ± 0.02	1.89 ± 0.09	2.45 ± 0.06
		GRU	0.88 ± 0.02	—	0.13 ± 0.03	0.27 ± 0.05	0.92 ± 0.04	1.01 ± 0.02	1.71 ± 0.17	2.30 ± 0.12
		AR	0.89 ± 0.00	—	0.16 ± 0.01	0.21 ± 0.02	1.03 ± 0.02	1.03 ± 0.01	2.03 ± 0.10	2.12 ± 0.09
	Static 2	T-GCN	0.72 ± 0.01	—	0.15 ± 0.02	0.59 ± 0.03	0.98 ± 0.02	1.22 ± 0.01	1.90 ± 0.08	3.55 ± 0.09
		LSTM	0.76 ± 0.01	—	0.21 ± 0.03	0.44 ± 0.01	1.02 ± 0.02	1.18 ± 0.00	2.07 ± 0.10	3.10 ± 0.02
		GRU	0.73 ± 0.01	—	0.17 ± 0.02	0.47 ± 0.02	0.96 ± 0.02	1.16 ± 0.01	1.91 ± 0.12	3.23 ± 0.08
		AR	0.73 ± 0.01	—	0.19 ± 0.02	0.36 ± 0.03	1.09 ± 0.03	1.14 ± 0.01	2.27 ± 0.14	2.86 ± 0.13

2 apply almost the same amount of censoring hence the similar performance. For schemes Static 2 and 3, there is a clear improvement in the censorship-aware models for both forecast leads. We now proceed with details on the performance of censorship-aware models. The AR model is outperformed by the DL models in any scenario. The GRU model outperforms the other models for 30-minute predictions across all censoring schemes, while T-GCN performs better for 24-hour forecasts. There is no clear model outperforming the other alternatives across forecast leads and censor levels. Surprisingly, T-GCN does not outperform the other models on short-term predictions. We intuitively expect the models to benefit from information from the graph, since charging data exhibit spatial correlation [11]. We attribute this to two factors. Primarily, the adjacency matrix controlling the contribution in predictions from the other nodes was based on the distance between clusters. It could be beneficial to incorporate the significance of the features of cluster i for cluster j with careful consideration of the characteristic charging patterns of the clusters. Furthermore, adopting graph attention network models, which have previously been identified as being capable of learning importance coefficients through the attention mechanism, could be advantageous to learn the spatial charging patterns [42]. Secondly, one-step predictions are a relatively easy task for charging data which is highly seasonal. Therefore, the complex model structure of T-GCN might overfit even though we applied regularization and saw no signs of overfitting in the loss curves. Additionally,

in the original paper implementing the T-GCN architecture, the model demonstrates that the performance of the T-GCN surpasses that of the GRU model, except when measuring for MAE, across all forecast horizons. This factor significantly contributes to the observed performance disparity [40].

Since the T-GCN models perform better at 24-hour forecasts, which are more suitable for many applications, we use T-GCN and T-GCN-A for the remainder of the analysis. The Hamilton cluster exhibits a high degree of censoring when just a single plug is removed with the scheme Dynamic 1. This weakens the aforementioned assumption of the observed demand not being censored by inherent real-world censoring (see Section V-B). As a result, the aware models overpredict relative to y^* as seen in Fig. 5, and thus the validity of the loss metric is questionable. When inspecting the loss metrics for the 30-minute, censorship-aware forecasts for Dynamic 1 and Dynamic 2, we see that Hamilton predictions contribute 20.41% and 22.29% of the sum of MAE respectively. Compared to the unaware predictions, these contributions are at 14.73% and 14.24% respectively. This discrepancy suggests that the censorship-aware model struggles with accurately forecasting this particular cluster when compared to the remaining clusters. If Hamilton is excluded from the dataset, a noteworthy reduction in MAE is observed. For the T-GCN-A, the MAE decreases from 0.73 to 0.67. Conversely, for the censorship-unaware T-GCN, the decrease is more marginal, reducing from 0.70 to 0.68.

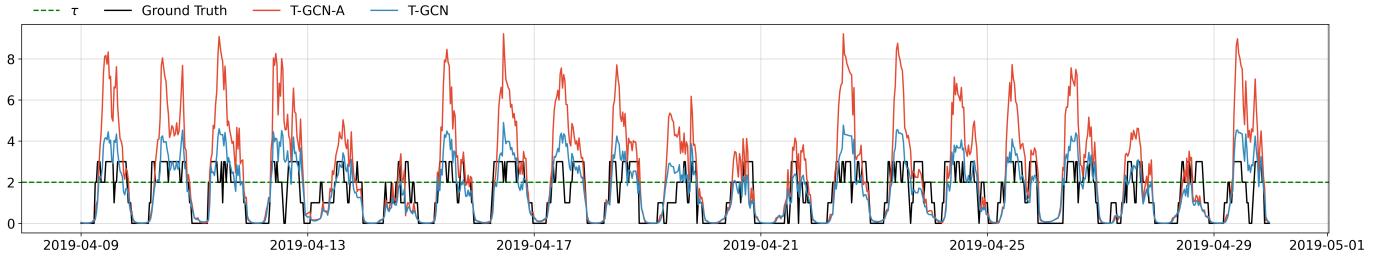


Fig. 5. 30-minute forecasts made by T-GCN and T-GCN-A on the Hamilton cluster, using the Dynamic 2 censoring strategy

As such, the exclusion of the Hamilton cluster results in the T-GCN-A demonstrating similar or marginally better performance. Thus, the exclusion of clusters with inherent real-world censoring in the data would lead to greater validity of the evaluation metrics.

Inspections of the T-GCN-A short-term predictions highlight that an “overreaction” happens when the model encounters multiple consecutively censored samples. This leads the next prediction to be much higher than the censoring threshold, leading to large fluctuations in the predictions. This is especially the case for clusters subject to high degrees of censorship, such as Hamilton, as can be seen in Fig. 5. In the figure, it is also evident that the unaware T-GCN model is able to forecast above τ . This can be attributed to the significantly higher demand prevalent among the nodes surrounding Hamilton, which remain uncapped within the dynamic censorship scheme, as opposed to their constrained status under the static censorship scheme.

B. Model applications: utilization and queue lengths

In this subsection, we will highlight the results of the experiments introduced in Section V-D. We are especially interested in testing the capabilities at moderate to high censorship levels thus all results are produced from predictions using the Static 2 censor scheme. This infers that, in a practical context, a charging cluster only houses two plugs.

1) *Day-ahead forecasts:* When inspecting Table VII it is apparent that the increase in MAE from forecast lead 30 minutes to 24 hours, is in the region of 0-0.3 which is relatively low. In Fig. 6 the predicted charging events for the Bryant cluster with nine plugs, visualized alongside the predictions of T-GCN and T-GCN-A for forecast leads 30 minutes, 12 hours, and 24 hours, are shown. It is apparent how the T-GCN-A model can accurately predict above the censoring threshold, while the unaware model is unable to extrapolate past the threshold. Although T-GCN-A only observes the true data until $\tau > 2$, the model predicts daily peak demand will be around six simultaneous charging events, while T-GCN predicts demand to be around 2. As the forecast lead increases, the T-GCN-A is less reactive to information from new time steps and the predictions are smoother, but the forecasts remain close to the observed demand when inspected visually. Day-ahead forecasts are useful to implement dynamic pricing across the charging network to encourage customers to use clusters with low demand through advantageous prices

[51]. This could increase the overall utility rate of the charging network. They can also be used to deploy mobile charging stations to increase plug capacity at clusters with demand exceeding the capacity. By doing so, the lost demand due to plug unavailability is decreased.

2) *Utilization rate:* A plot illustrating the utilization rate can serve as an effective tool for industry professionals to identify the frequency of usage for various clusters across their charging network. City planners also use it to ensure that adequate charging is available for EV owners. As mentioned in Section II, utilization rates are highly dependent on various factors. To balance business incentives and availability for citizens, we compare the utility rates to a 20% benchmark. As illustrated in Fig. 7, the model delineates the utilization for the high-capacity cluster Bryant, for T-GCN and T-GCN-A. The average and standard deviation of the utilization is reported. In addition to the rate estimated from model predictions, the actual utilization is also presented for comparison. The training of the models has been conducted using the Static 2 censoring scheme.

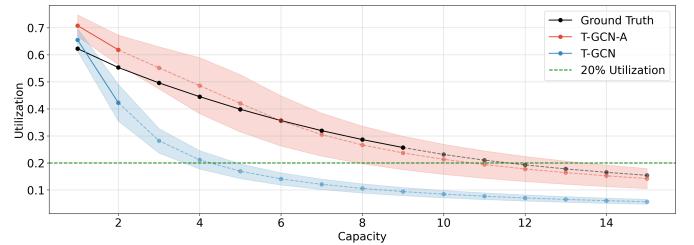


Fig. 7. Utilization rates of Bryant as estimated by the aware and unaware T-GCN model compared to the true utilization rate. Variations in these rates are indicated by the shaded regions representing standard deviations. Points, where the model-estimated theoretical capacity exceeds the observed capacity, are denoted by dotted lines. Additionally, these dotted lines also represent the cluster capacity for the genuine demand curve in the analysis.

T-GCN-A lies much closer to the true utilization. Compared to the 20% utilization rate benchmark, both aware and true utilization estimate an increase in capacity to ten to eleven plugs, while T-GCN estimates four plugs are sufficient. However, four plugs would not be sufficient from a business point of view since the true demand exceeds the estimated demand and additional plugs could easily be constructed while still staying above the 20% utility threshold. In Table III the optimal number of plugs to stay above the 20% utilization are summarized for each cluster, according to the observed data, T-GCN-A and T-GCN.

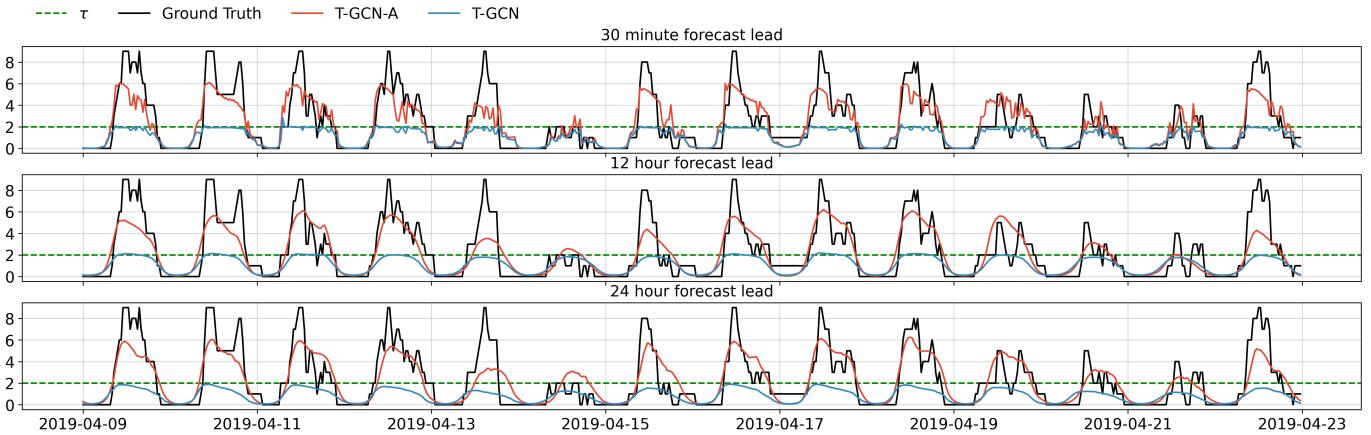


Fig. 6. Time series of cluster Bryant for several forecast leads for the test set censored using the Static 2 strategy. The censoring threshold is indicated with τ

A way to utilize these results for choosing areas to expand EV chargers would be to pick a low-capacity cluster where the predicted optimal capacity far exceeds the current capacity by most.

TABLE III

EXPERIMENT SHOWING THE OPTIMAL NUMBER OF CLUSTERS BASED ON A TARGET OF 20% UTILIZATION ACCORDING TO THE ACTUAL DATA, AWARE T-GCN PREDICTIONS, AND UNAWARE T-GCN PREDICTIONS. THE CURRENT NUMBER OF PLUGS IN A CLUSTER IS ALSO DISPLAYED

	Bryant	Cambridge	Hamilton	High	MPL	Rinc.	Ted	Webster
Capacity	9	10	3	8	4	3	8	6
True	12	13	8	11	6	4	9	10
TGCN-A	11	11	11	11	8	5	10	11
TGCN	4	5	5	5	4	4	5	5

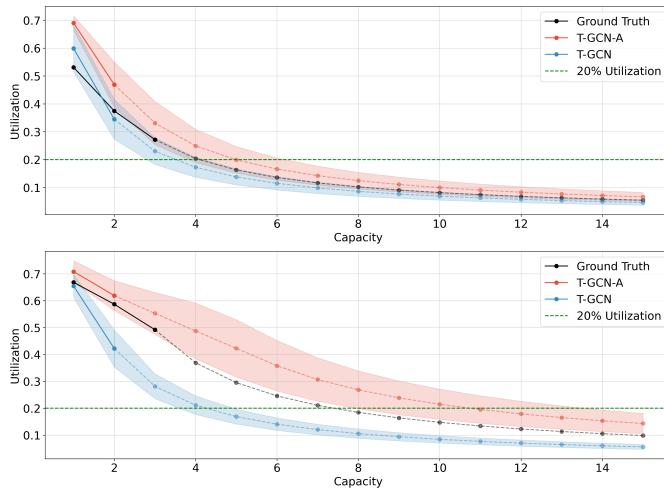


Fig. 8. Utilization rates of Rinconada (upper) and Hamilton (lower) as estimated by the aware and unaware T-GCN model compared to the true utilization rate. The data is censored using the Static 2 censoring strategy and the standard deviation of the utilization estimation is shown in shaded colors

Another use case for estimating utilization is the deployment of mobile charging stations to explore new potential charging locations. The station can be placed at new sites to collect charging data for a few months, to then be included as a cluster in the charging network. Assuming there are no existing locations at Rinconada and Hamilton, both having true

capacity three, charging stations with plug capacity two are placed at each site, respectively. Inspecting their utilization rate in Fig. 8 it is apparent that both locations appear to be good locations to install new charging stations, as the utilization rate is above the desired 20% threshold at capacity two. According to the unaware model, a capacity of four is sufficient at both locations. The aware model estimates that five plugs are sufficient at Rinconada, but up to 8-11 are needed at Hamilton according to T-GCN-A. Therefore, the aware models provide valuable insight into the optimal capacity of clusters at new locations. When the censorship level is low, as for the Rinconada cluster with 6.5% censorship, the performance of T-GCN and T-GCN-A is similar.

3) *Queue lengths:* Directly estimating the queue length from historical data can help EV charging distributors understand temporal patterns of lost demand and ultimately employ dynamic pricing and manage charging trucks in a similar fashion to the forecasts discussed in Section VI-B1. Fig. 9 shows the mean queue lengths and 95% confidence interval during a day for another high-capacity cluster, Webster, where the capacity is reduced to two plugs using the censor scheme Static 2. The true queue length is estimated from the true observations. It is apparent that queuing only occurs during the daytime when the demand for public charging is high. The estimated queue lengths lie close to the actual queue lengths, especially during the peak hours in the morning. The demand is highest around 10:00 AM with a mean queue length of around 1. Unsurprisingly, T-GCN does not estimate queue length at any time during the day. As mentioned in Section V-D, the queue length is an estimate for lost customers for level 2 charging stations. Depending on the location of a cluster, different daily demand patterns are expected. Webster is located in a commercial area with several businesses, where employees charge their vehicles upon arrival in the morning. Demand for public charging in residential areas is usually highest during the afternoon as people return from work.

The interpretation of the queue lengths in Fig. 9 is not that approximately one customer is rejected every day during

peak hours since a charging event usually lasts several hours. Thus, an alternative approach to estimating rejected customers is using queuing systems.

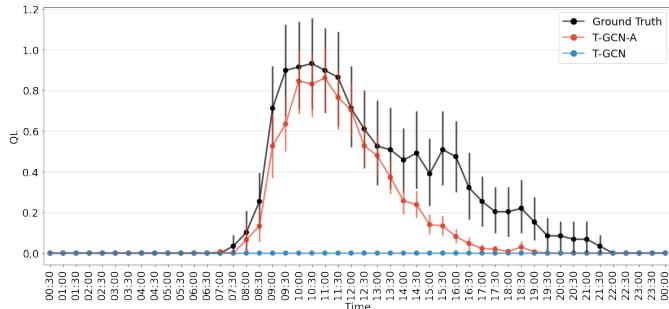


Fig. 9. Estimated queue lengths for cluster Webster, as predicted by T-GCN-A and T-GCN, together with the ground truth

In Fig. 10 we have simulated Webster with the arrival rates estimated from the predictions of T-GCN and T-GCN-A assuming that Webster has four plugs available. For this particular simulated day, two plugs were in use for the majority of the day according to the true observations. It is evident that the low arrival rates estimated by T-GCN lead to zero plugs being in use for the majority of the day.

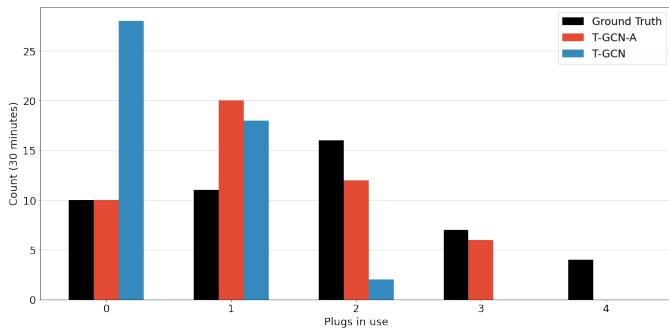


Fig. 10. Simulated count of time intervals spent with 0 to 4 plugs in use at cluster Webster

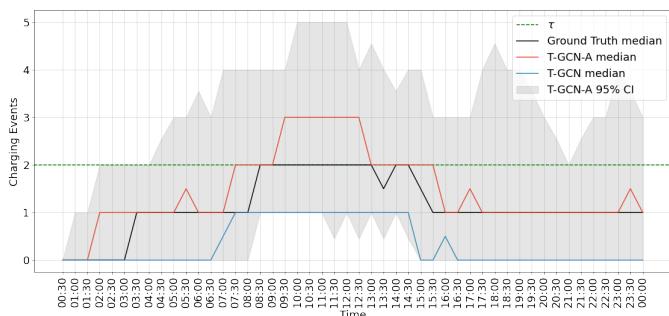


Fig. 11. Median of T-GCN and T-GCN-A shown alongside the true median for the Webster cluster. The 95% CI of T-GCN-A and plug capacity in the simulation is also shown

We run the simulation 50 times (corresponding to 50 simulated days), allowing any arriving customer to charge regardless of the current number of charging events. The realizations of the T-GCN-A can be seen in Appendix Fig.

15. Fig. 11 shows the median with a 95% confidence interval alongside the median of the true observations. The horizontal line τ indicates the plug capacity given by the Static 2 censor scheme.

We assume any customers arriving while two cars are already charging are rejected. The estimated average number of rejections by the models is 0.58 by true observations, 0.38 by T-GCN-A, and 0 by T-GCN. Table IV shows the median number of rejected customers across clusters when simulating 50 days. The number of available plugs is kept constant at two rejecting any customers arriving above the threshold.

TABLE IV
SIMULATED REJECTED CUSTOMERS WITH ARRIVAL RATES ESTIMATED FROM THE TRUE OBSERVATIONS AND PREDICTIONS FROM T-GCN-A AND T-GCN RESPECTIVELY

	Bryant	Cambridge	Hamilton	High	MPL	Rinc.	Ted	Webster
True	3 ± 1.6	3.5 ± 2.06	2 ± 1.77	4 ± 1.60	1 ± 1.47	1 ± 1.73	3 ± 1.52	3 ± 1.70
T-GCN-A	1 ± 1.38	1 ± 1.58	2 ± 1.46	2 ± 1.50	1 ± 1.43	0 ± 0.82	1 ± 1.47	2 ± 1.12
T-GCN	0 ± 0.50	0 ± 0.46	0 ± 0.38	0 ± 0.50	0 ± 0.58	0 ± 0.46	0 ± 0.48	0 ± 0.35

The median number of rejected customers across clusters is in general higher using the arrival rate from the true observations. According to the T-GCN-A model, customers are likely to be rejected at every cluster. The median is zero at Rinconada, which is also the cluster with the lowest censorship level. The median number of rejected customers estimated from T-GCN is zero across all clusters. This is unsurprising since the unaware model rarely predicts above the censoring threshold.

VII. CONCLUSION

The effect of censoring in public EV charging data can have a significant effect on the reliability of predictive models and ultimately lead to ill-informed decision-making for industry practitioners and city planners. In this paper, we addressed the problem of estimating latent demand from charging supply data by proposing censorship-aware models that are able to extrapolate to unobserved, censored demand. We used several model architectures (AR, GRU, LSTM and T-GCN) at different censor levels and forecast leads. The proposed censored Poisson likelihood function was tested on charging data from eight charging clusters in Palo Alto. The aware models demonstrated noticeable improvement in performance over the unaware models at low censoring levels and above (above 5%). With little to no censoring (below 5%) the aware and unaware models perform similarly. When real-world censoring is present in the data, the aware models are prone to overpredict when compared to the “true” demand which must be accounted for when assessing the performance of censorship-aware models.

In the context of censorship-aware models, the GRU model performed better for one-step predictions (30 min), while T-GCN was the best performer when forecasting 48 steps ahead (24 hr). We provide several ways to potentially harness the full potential of spatial correlation between charging clusters when using graph learning for censored regression. The likelihood function is trivially adaptable to

other datasets and even other domains, where right censoring occurs.

We showcase various model applications that can aid industry practitioners in informed expansion strategies, implementation of dynamic pricing, and managing mobile charging trucks. The applications highlight how censorship-aware models can elevate the insight gained from the increasing amount of charging data recorded every day, eventually leading to better EV infrastructure.

VIII. FUTURE WORK

In future research, we plan to study alternative ways to construct the graph used in the T-GCN model. We propose to use the cross-correlation of clusters time series to capture temporal dependencies, rather than solely relying on the distance to express similarity between charging clusters. This methodology could improve predictive performance, by unveiling intricate correlations between charging events, although it necessitates careful handling due to potential spurious correlations and computational complexity. Alternatively, it would be interesting to compare the performance of the T-GCN architecture to the performance of an Attention-based model, where the spatial dependencies are learned in training.

We saw good initial results with the Poisson likelihood function and deemed it useful for predicting the demand of charging events. However, alternative likelihood functions could have been applied to evaluate the performance of the Poisson likelihood function. For example, a non-parametric approach like quantile regression or a normal Tobit model could also have been used for discretized charging data [8], [9].

Ideally, to make a better evaluation of these models, a dataset with no censorship should be used. When using a censoring scheme to validate the model forecasts, the assumption of the real-world demand not being censored has to be met. It is virtually impossible to eliminate any inherent censorship within the data, but schemes to preprocess data or “inform” the model to ignore certain censored observations could be explored.

REFERENCES

- [1] S. Abraham, K. Ganesh, A. S. Kumar, and Y. Ducqd, “Impact on climate change due to transportation sector – research prospective,” *Procedia Engineering*, vol. 38, pp. 3869–3879, 2012, iINTERNATIONAL CONFERENCE ON MODELLING OPTIMIZATION AND COMPUTING. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877705812023582>
- [2] M. Yuan, J. T. Zinck, H. Lund, and Y. Liang, “The electrification of transportation in energy transition,” *Energy*, vol. 236, p. 121564, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360544221018120>
- [3] I. E. Agency, “Global ev outlook 2023,” <https://www.iea.org/reports/global-ev-outlook-2023>, Paris, 2023, license: CC BY 4.0.
- [4] International Energy Agency. (2022, 9) Electric vehicles. International Energy Agency. [Online]. Available: \url{https://www.iea.org/reports/electric-vehicles}
- [5] L. Chen, G. Msigwa, M. Yang, A. I. Osman, S. Fawzy, D. W. Rooney, and P.-S. Yap, “Strategies to achieve a carbon neutral society: a review,” *Environmental Chemistry Letters*, vol. 20, no. 4, pp. 2277–2310, 8 2022. [Online]. Available: <https://doi.org/10.1007/s10311-022-01435-8>
- [6] P. Kampshoff, A. Kumar, S. Peloquin, and S. Sahdev. (2021) Building the electric vehicle charging infrastructure america needs. [Online]. Available: <https://www.mckinsey.com/industries/public-sector/our-insights/building-the-electric-vehicle-charging-infrastructure-america-needs>
- [7] H. Engel, R. Hensley, S. Knupfer, and S. Sahdev. (2018) Charging ahead: Electric vehicle infrastructure demand. [Online]. Available: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/charging-ahead-electric-vehicle-infrastructure-demand>
- [8] D. Gammelli, I. Peled, F. Rodrigues, D. Pacino, H. A. Kurtaran, and F. C. Pereira, “Estimating latent demand of shared mobility through censored gaussian processes,” *Transportation Research Part C: Emerging Technologies*, vol. 120, p. 102775, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0968090X20306859>
- [9] F. B. Hüttel, I. Peled, F. Rodrigues, and F. C. Pereira, “Modeling censored mobility demand through quantile regression neural networks,” *CoRR*, vol. abs/2104.01214, 2021. [Online]. Available: <https://arxiv.org/abs/2104.01214>
- [10] V. Chauhan and A. Gupta, “Scheduling mobile charging stations for electric vehicle charging,” in *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2018, pp. 131–136.
- [11] F. B. Hüttel, I. Peled, F. Rodrigues, and F. C. Pereira, “Deep spatio-temporal forecasting of electrical vehicle charging demand,” *CoRR*, vol. abs/2106.10940, 2021. [Online]. Available: <https://arxiv.org/abs/2106.10940>
- [12] B. Borlaug, F. Yang, E. Pritchard, E. Wood, and J. Gonder, “Public electric vehicle charging station utilization in the united states,” *Transp. Res. D Transp. Environ.*, vol. 114, no. 103564, p. 103564, Jan. 2023.
- [13] M. Gellrich, A. Block, and N. Leikert-Böhm, “Spatial and temporal patterns of electric vehicle charging station utilization: a nationwide case study of switzerland,” *Environmental Research: Infrastructure and Sustainability*, vol. 2, no. 2, p. 021003, jun 2022. [Online]. Available: <https://dx.doi.org/10.1088/2634-4505/ac6a09>
- [14] C. Hecht, S. Das, C. Bussar, and D. U. Sauer, “Representative, empirical, real-world charging station usage characteristics and data in germany,” *eTransportation*, vol. 6, p. 100079, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2590116820300369>
- [15] P. Sadeghi-Barzani, A. Rajabi-Ghahnavieh, and H. Kazemi-Karegar, “Optimal fast charging station placing and sizing,” *Applied Energy*, vol. 125, pp. 289–299, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261914003171>
- [16] C. Vandet and J. Rich, “Optimal placement and sizing of charging infrastructure for evs under information-sharing,” *Technological Forecasting and Social Change*, vol. 187, 2023.
- [17] F. Ahmad, A. Iqbal, I. Ashraf, M. Marzband, and I. Khan, “Optimal location of electric vehicle charging station and its impact on distribution network: A review,” *Energy Reports*, vol. 8, pp. 2314–2333, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352484722001809>
- [18] H. Fredriksson, M. Dahl, and J. Holmgren, “Optimal placement of charging stations for electric vehicles in large-scale transportation networks,” *Procedia Computer Science*, vol. 160, pp. 77–84, 2019, the 10th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2019) / The 9th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2019) / Affiliated Workshops. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S187705919316618>
- [19] M. Daskin, “A maximum expected covering location model: Formulation, properties and heuristic solution,” *Transportation Science*, vol. 17, pp. 48–70, 02 1983.
- [20] I. Frade, A. Ribeiro, G. Gonçalves, and A. Pais Antunes, “Optimal location of charging stations for electric vehicles in a neighborhood in lisbon, portugal,” *TRR*, vol. 2252, 12 2011.
- [21] S. Wagner, M. Götzinger, and D. Neumann, “Optimal location of charging stations in smart cities: A point of interest based approach,” *International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design*, vol. 3, pp. 2838–2855, 01 2013.
- [22] “Supercharging — tesla,” accessed: June 10, 2023. [Online]. Available: \url{https://www.tesla.com/en/_NZ/support/supercharging}

- [23] R. Telang, A. Singh, H. Le, and A. Higashi. (2021) Electric vehicles and the charging infrastructure: a new mindset? [Online]. Available: <https://www.pwc.com/us/en/industrial-products/publications/assets/pwc-electric-vehicles-charging-infrastructure-mindset.pdf>
- [24] M. Békés, M. Fruk, E. Hildebrandt, K. Kalstad, K. Mantel, F. Nägele, and S. Ramanathan, "What norway's experience reveals about the ev charging market," 2023, accessed: June 25, 2023. [Online]. Available: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/what-norways-experience-reveals-about-the-ev-charging-market>
- [25] K. Suresh, C. Severn, and D. Ghosh, "Survival prediction models: an introduction to discrete-time modeling," *BMC Medical Research Methodology*, vol. 22, no. 1, p. 207, Jul 2022. [Online]. Available: <https://doi.org/10.1186/s12874-022-01679-6>
- [26] A. Lewbel and O. Linton, "Nonparametric censored and truncated regression," *Econometrica*, vol. 70, no. 2, pp. 765–779, 2002. [Online]. Available: <http://www.jstor.org/stable/2692291>
- [27] R. A. Canales, A. M. Wilson, J. I. Pearce-Walker, M. P. Verhougstraete, and K. A. Reynolds, "Methods for handling left-censored data in quantitative microbial risk assessment," *Appl. Environ. Microbiol.*, vol. 84, no. 20, Oct. 2018.
- [28] J. Tobin, "Estimation of relationships for limited dependent variables," *Econometrica*, vol. 26, no. 1, pp. 24–36, 1958. [Online]. Available: <http://www.jstor.org/stable/1907382>
- [29] E. T. Lee and O. T. Go, "Survival analysis in public health research," *Annu. Rev. Public Health*, vol. 18, no. 1, pp. 105–134, 1997.
- [30] G. Salinas-Escudero, M. F. Carrillo-Vega, V. Granados-García, S. Martínez-Valverde, F. Toledano-Toledano, and J. Garduño-Espinoza, "A survival analysis of covid-19 in the mexican population," *BMC Public Health*, vol. 20, no. 1, p. 1616, 10 2020. [Online]. Available: <https://doi.org/10.1186/s12889-020-09721-2>
- [31] H. J. Noh, T. H. Roh, and I. Han, "Prognostic personal credit risk model considering censored information," *Expert Systems with Applications*, vol. 28, no. 4, pp. 753–762, 2005. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417404001794>
- [32] L. Chen and A. J. Mersereau, "Analytics for operational visibility in the retail store: The cases of censored demand and inventory record inaccuracy," *Retail Supply Chain Management: Quantitative Models and Empirical Studies*, pp. 79–112, 2015. [Online]. Available: https://doi.org/10.1007/978-1-4899-7562-1_5
- [33] T. Alghamdi, D. Said, and H. Mouftah, "Profit maximization for evses-based renewable energy sources in smart cities with different arrival rate scenarios," *IEEE Access*, vol. PP, 03 2021.
- [34] F. Aboshady, I. Pisica, and C. Axon, "A stochastic model for estimating electric vehicle arrival at multi-charger forecourts," *Energy Reports*, vol. 8, pp. 11569–11578, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352484722017292>
- [35] J. V. Terza, "A tobit-type estimator for the censored poisson regression model," *Economics Letters*, vol. 18, no. 4, pp. 361–365, 1985. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0165176585900539>
- [36] K. Cho, B. van Merriënboer, Ç. Gülcöhre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *CoRR*, vol. abs/1406.1078, 2014. [Online]. Available: <http://arxiv.org/abs/1406.1078>
- [37] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 11 1997. [Online]. Available: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [38] F. B. Hüttel, F. Rodrigues, and F. C. Pereira, "Mind the gap – modelling difference between censored and uncensored electric vehicle charging demand," 2023. [Online]. Available: <https://arxiv.org/abs/2301.06418>
- [39] R. Kass and P. Vos, *Geometrical Foundations of Asymptotic Inference*, ser. Wiley Series in Probability and Statistics. Wiley, 1997. [Online]. Available: <https://books.google.dk/books?id=7RzvAAAAMAAJ>
- [40] 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*, vol. 21, no. 9, pp. 3848–3858, 9 2020. [Online]. Available: <https://doi.org/10.1109/TITS.2019.2935152>
- [41] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, 2017. [Online]. Available: <https://openreview.net/forum?id=SJU4ayYgl>
- [42] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," *6th International Conference on Learning Representations*, 2017.
- [43] M. A. Pinsky and S. Karlin, "1 - introduction," in *An Introduction to Stochastic Modeling (Fourth Edition)*, 4th ed., M. A. Pinsky and S. Karlin, Eds. Boston: Academic Press, 2011, pp. 1–45. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780123814166000010>
- [44] ———, "9 - queueing systems," in *An Introduction to Stochastic Modeling (Fourth Edition)*, 4th ed., M. A. Pinsky and S. Karlin, Eds. Boston: Academic Press, 2011, pp. 447–494. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780123814166000010>
- [45] F. H. Malik and M. Lehtonen, "Minimization of queuing time of electric vehicles at a fast charging station," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 2017, pp. 1–6.
- [46] City of Palo Alto, "Electric Vehicle Charging Station Usage (July 2011 - Dec 2020)," <https://data.cityofpaloalto.org/dataviews/257812/electric-vehicle-charging-station-usage-july-2011-dec-2020/>, 2 2021, [Online; accessed 2023-04-12].
- [47] U.S. Department of Transportation, "Charger types and speeds," 2023. [Online]. Available: <https://www.transportation.gov/rural/ev/toolkit/ev-basics/charging-speeds>
- [48] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, 2019.
- [49] B. Shahriari, K. Swersky, Z. Wang, R. Adams, and N. De Freitas, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the Institute of Radio Engineers*, vol. 104, no. 1, pp. 148–175, Jan. 2016, publisher Copyright © 1963-2012 IEEE.
- [50] SparkCharge, "About Us," <https://www.sparkcharge.io/pages/about-us>, Year of Access.
- [51] Spirii, "Dynamic pricing from spirii: The future of ev charging costs explained," Spirii, Bragesgade 8b, 2200 Copenhagen, 2023, accessed: 2023-06-26. [Online]. Available: <https://spirii.com/en/resources/blog/dynamic-pricing-from-spirii-the-future-of-ev-charging-costs-explained>
- [52] F. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural computation*, vol. 12, pp. 2451–71, 10 2000.
- [53] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014.
- [54] S. Paul, "Bayesian hyperparameter optimization - a primer," 2020, accessed: 2023-06-30. [Online]. Available: <https://wandb.ai/site/articles/bayesian-hyperparameter-optimization-a-primer>
- [55] C. D. Sa. (2019) Gaussian processes and bayesian optimization. Cornell University, CS 4787. [Online]. Available: <https://www.cs.cornell.edu/courses/cs4787/2019sp/notes/lecture16.pdf>

APPENDIX A BACKGROUND AND ALGORITHMS

A. LSTM

The Long Short-Term Memory (LSTM) model introduced the usage of information gates in RNNs, making the model capable of learning long-term dependencies [37]. The model uses three gates; *forget*, *update* and *output*. The LSTM incorporates a cell state c_t which can be thought of as the memory of the RNN model, with the gates controlling what information that should be stored in it or discarded from it. An extension of the LSTM added an additional gate, f_t , referred to as the forget gate [52]. This gate controls what information is to be discarded from the cell state by deciding the importance of the input x_t and the previous hidden state h_t defined as:

$$f_t = \sigma_s(\mathbf{W}_f[x_t, h_{t-1}] + \mathbf{b}_f) \quad (19)$$

where σ_s is the sigmoid activation function, x_t is the input at time t , and \mathbf{W}_f are the weights learned by the network using back-propagation through time. The output out the gate is $f_t \in (0, 1)$, where 0 implies forgetting the part in the cell state and 1 indicates it should be kept. The LSTM improves on the issue of vanishing gradients compared to the plain RNN model architectures, as the forget-gate does additive updates to the cell state instead of multiplication of hidden state with some weights.

We then have the input layer, which decides what new information should be stored in the cell state, again weighed by using the sigmoid function.

$$i_t = \sigma_s(\mathbf{W}_i[x_t, h_{t-1}] + \mathbf{b}_i) \quad (20)$$

The information which to update the cell state with, candidate values, are created by a tanh layer,

$$\tilde{c}_t = \tanh(\mathbf{W}_{\tilde{c}}[x_t, h_{t-1}] + \mathbf{b}_{\tilde{c}}) \quad (21)$$

$\tilde{c}_t \in (-1, 1)$. We can then update the cell state, by forgetting part of the old cell state, and updating it with the candidate state, weighed by worthiness by the input layer

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (22)$$

To decide which parts of the cell state to use in the new hidden state, the output gate o_t is another sigmoid layer that weighs this:

$$o_t = \sigma_s(\mathbf{W}_o[x_t, h_{t-1}] + \mathbf{b}_o) \quad (23)$$

Finally, a new hidden state can be defined by using the relevant part of the cell state, as judged by the output layer, and running it through a tanh activation function, to maintain the values in the range (-1, 1):

$$h_t = \tanh(c_t) * o_t \quad (24)$$

B. GRU

A gated recurrent unit (GRU) was first introduced by Cho et al. and is similar to the LSTM in the sense that it uses gates to filter information from previous time steps, but has a reduced number of gates to improve computational efficiency [36]. It makes use of two gates; the reset and update gate. For a single GRU-layer, the hidden state at time t is given as the weighted sum between existing and new states [53]:

$$h_t = (1 - u_t)c_t + u_t h_{t-1} \quad (25)$$

using the hidden state at $t-1$ and a candidate hidden state at state t , c_t , and the update gate, u_t , given as:

$$u_t = \sigma_s(\mathbf{W}_u[x_t, h_{t-1}] + \mathbf{b}_u) \quad (26)$$

which similarly to the input gate in LSTM, decides how much of past information should be kept in the state, thus helping the GRU learn long-term dependencies. When the entries of the update gate are close to 1, the previous hidden state is kept.

The candidate hidden state c_t is given as:

$$c_t = \tanh(\mathbf{W}[x_t, r_t * h_{t-1}] + \mathbf{b}_c) \quad (27)$$

where \tanh denotes the hyperbolic tangent. The reset gate determines how much of the previous state should be remembered, and is given by:

$$r_t = \sigma_s(\mathbf{W}_r[x_t, h_{t-1}] + \mathbf{b}_r) \quad (28)$$

If the entries in r_t are close to 1, it forgets the previously computed state, while being close to 0 implies completely retaining the previous state. The hidden state h_t is thus propagated through time carrying information about previous time steps. It is passed through an activation function to produce a prediction, \hat{y} .

C. Queueing system

The outline of the algorithm used to simulate the clusters as servers in an M/M/c queue

Algorithm 1 M/M/c Queue Simulation

```

1: random.seed(seed)                                ▷ Set seed for reproducibility
2: for t in 0 to day do                            ▷ Sample rates
3:   Sample arrival rate λ from predictions
4:   Sample service times μ from data
5: end for
6: for i in 0 to time_horizon do          ▷ Process each time step
7:   Update queue length
8:   Remove departed customers from the queue
9:   if Arrival at i then
10:     if Charging station available then
11:       Calculate departure time for the customer
12:       Add departure time to the queue
13:       Sort queue based on departure times
14:     else
15:       Increment the count of rejected customers
16:     end if
17:   Add the completed customer to the list of completed
18:   customers
19: end for
20: Return the list of completed customers, queue length, and counts
    of rejected customers

```

APPENDIX B MODEL DEVELOPMENT

A. Bayesian Optimization

For tuning the hyperparameters of the model, we used the out-of-the-box method *Sweeps* of Weights and Biases using Bayesian optimization as a method for searching the hyperparameter space [54]. The problem of finding optimal parameters θ^* for an objective function $f(\theta)$ can be boiled down to a global maximization problem [49],

$$\theta^* = \arg \max_{\theta \in \Theta} f(\theta) \quad (29)$$

where Θ is the design space spanned by the hyperparameters that we wish to optimize. Assume we have observed

$$\mathcal{D}_n = \{(\theta_i, y_i)\}_{i=1}^n \quad (30)$$

with y_i being the loss of the model given parameters θ . Here, we assume $f(\theta_1), f(\theta_2), \dots, f(\theta_n)$ as observed.

Bayesian optimization uses a cheap surrogate model to estimate the computational heavy $f(\theta) \notin \mathcal{D}_n$, which in our case is the validation loss. The surrogate model is assumed to be a Gaussian process,

$$y_i \sim \mathcal{N}(0, \Sigma), \quad y_i \notin \mathcal{D}_n \quad (31)$$

with the covariance matrix being

$$\Sigma = \begin{bmatrix} K(\theta_1, \theta_1) & K(\theta_1, \theta_2) & \cdots & K(\theta_1, \theta_D) \\ \vdots & \vdots & \ddots & \vdots \\ K(\theta_D, \theta_1) & K(\theta_D, \theta_2) & \cdots & K(\theta_D, \theta_D) \end{bmatrix} \quad (32)$$

where $K(\theta, \theta')$ is the kernel function of the hyperparameter set θ and θ' . K is our uncertainty estimate of the surrogate model which is used to pick the next set of hyperparameters. The Matérn kernel is used in the Weights and Biases implementation of Bayesian Optimization,

$$K(\theta_i, \theta_j) = \frac{1}{2^{\zeta-1}\Gamma(\zeta)}(2\sqrt{\zeta}||\theta_i - \theta_j||)^{\zeta}H_{\zeta}(2\sqrt{\zeta}||\theta_i - \theta_j||) \quad (33)$$

where Γ and H_{ζ} are the Gamma and Bessel functions of order ζ . The Gaussian process is modeling the probability of the performance of model f at a new hyperparameter θ^* , given our observed hyperparameters and performances, thus giving rise to the marginal distribution:

$$f(\theta^*)|\mathcal{D}_n \sim \mathcal{N}(\mu(\theta^*), \sigma(\theta^*))$$

where

$$\mu(\theta^*) = \mathbf{k}_*^T \Sigma^{-1} y, \quad \sigma(\theta^*) = K(\theta^*, \theta^*) - \mathbf{k}_*^T \Sigma^{-1} \mathbf{k}_*$$

where is the mean and standard deviation of the hidden variable $f(\theta^*)$, and [55]

$$\mathbf{k}^* = [K(\theta_1, \theta_*) \ K(\theta_2, \theta_*) \ \dots \ K(\theta_D, \theta_*)]^T \quad (34)$$

The acquisition function is another key part of Bayesian optimization. This function defines the next set of hyperparameters to evaluate, influenced by the observations made thus far. We can calculate the probability of a proposed hyperparameter θ^*

leading to a lower loss than our best-performing hyperparameter $\hat{\theta}$,

$$\begin{aligned} \mathbb{P}(f(\theta^*) < f(\hat{\theta})) &= \mathbb{P}\left(\frac{f(\theta^*) - \mu(\theta^*)}{\sigma(\theta^*)} < \frac{f(\hat{\theta}) - \mu(\theta^*)}{\sigma(\theta^*)}\right) \\ &= \Phi\left(\frac{f(\hat{\theta}) - \mu(\theta^*)}{\sigma(\theta^*)}\right) \end{aligned}$$

The selection process is guided by the principle of *maximized expected improvement*, which is given analytically as follows:

$$EI(\theta^*) = -\left(\phi\left(\frac{f(\hat{\theta}) - \mu(\theta^*)}{\sigma(\theta^*)}\right) + \frac{f(\hat{\theta}) - \mu(\theta^*)}{\sigma(\theta^*)} \cdot \Phi\left(\frac{f(\hat{\theta}) - \mu(\theta^*)}{\sigma(\theta^*)}\right)\right) \cdot \sigma$$

The next hyperparameter θ_{t+1} to evaluate in the real model, is then chosen based on,

$$\theta_{t+1} = \arg \max_{\theta_t} EI(\theta_t \in \Theta) \quad (35)$$

The algorithm evaluates this candidate, adds the hyperparameters and the losses to the observations, and finds the next hyperparameter to be assessed. This is done for a certain number of iterations.

B. Model Configurations

The resulting configurations for hyperparameter tuning are summarized in Table V.

TABLE V
HYPERPARAMETERS FOR MODELS TRAINED WITH FORECAST LEAD 30 MINUTES. AR, GRU, AND LSTM WERE TRAINED FOR 30 EPOCHS AND THE TGCN WAS TRAINED FOR 20 EPOCHS

Model	Seq. length	LR	Weight decay	Hyperparameters		
				Covariates	Hid. dim.	Batch size
AR	336	0.0006	0.0647	-	-	32
GRU	336	0.0007	0.0027	Yes	238	51
LSTM	336	0.0006	0.0507	Yes	163	60
TGCN	336	0.0001	0.0001	Yes	512	128

APPENDIX C DATA AND VISUALIZATIONS

A. Data processing

We conducted our experiments using an open-source dataset obtained from electric vehicle charging stations located in Palo Alto, CA [46]. The data consists of almost 260.000 entries corresponding to individual charging sessions with 32 columns amounting to around a memory size of 65MB.

In the original data, there was a ninth cluster, Sherman, which was constructed in December 2020 and only contained data for a single month. As a result, the cluster was excluded from the analysis meaning that we consider eight clusters for the model. Although the charging infrastructure we consider is expanded over time, it is important to note that the dataset does not include data from every charging station in Palo Alto.

We used the plug number for each station to extrapolate the

number of plugs available within a cluster. Some charging plugs were used in very few sessions, and we concluded that these plugs may have been used for testing or were under maintenance leading to low usage. As a result, we correct the number of available plugs accordingly. After correction of the plug capacity, some clusters experienced periods with coinciding charging events above the plug capacity for a few time periods. This mishap occurs due to the time resolution of one minute, i.e. counting the maximum number of charging events per minute, which we used to create the 30-minute intervals. Given that a customer leaves and arrives in the same minute using the same plug, a total of two simultaneous charging events would be reported, even though they used the same plug. We correct these observations to be equal to the plug capacity amounting to a correction of 2% of the observations. We assume that for data owned by an industry practitioner, there would naturally be data on how many functioning plugs there would be at any charging station.

B. Temporal charging patterns

We explore the temporal and spatial variations for some periods post-discretization. Initially, we explored the time series of clusters Bryant, MPL, and Webster during the first two weeks in May 2019 in Fig. 12. A similar weekly trend across clusters is the decrease in charging events on weekends. Additionally, all clusters experience little to no usage during the night, with demand peaking in the morning and afternoon.

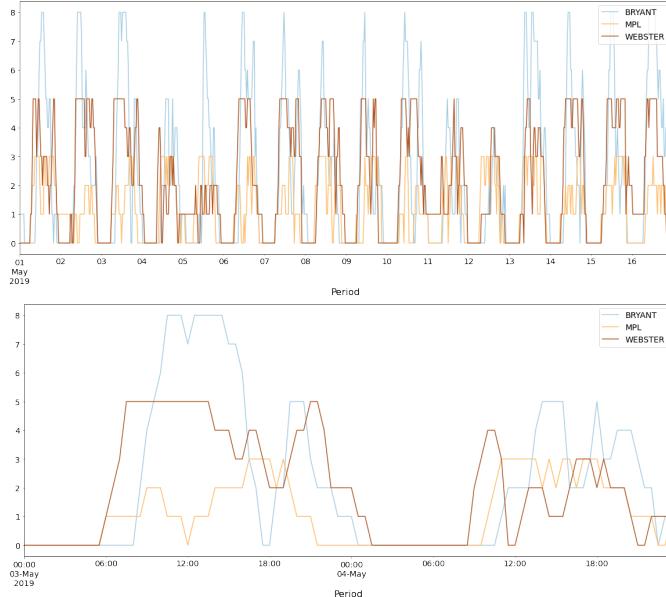


Fig. 12. The time series of clusters Bryant, MPL, and Webster during two weeks (upper) and two days (lower) of May 2019. The y-axis denotes the number of ongoing charging events and the x-axis denotes time on a 30-minute resolution.

It is also interesting to notice that the clusters exhibit different daily patterns. From the randomly chosen two weeks, it is evident that the clusters are fully utilized during For example, the usage of MPL seems to have similar demand for weekdays and weekends while Bryant and Webster have

clear dips.

Aggregating the charging events into daily patterns we can see the clear dips for days 5 and 6 (Saturday and Sunday) in Fig. 13. It is interesting to note that clusters MPL and Rinconada do not dip as much. They are also located in more residential areas and farther from the commercial area of the other clusters. Therefore, weekend usage is more expected. This is also visible in the daily pattern, where MPL and Rinconada (and to some extent, Ted) exhibit steady mean charging events throughout the day.

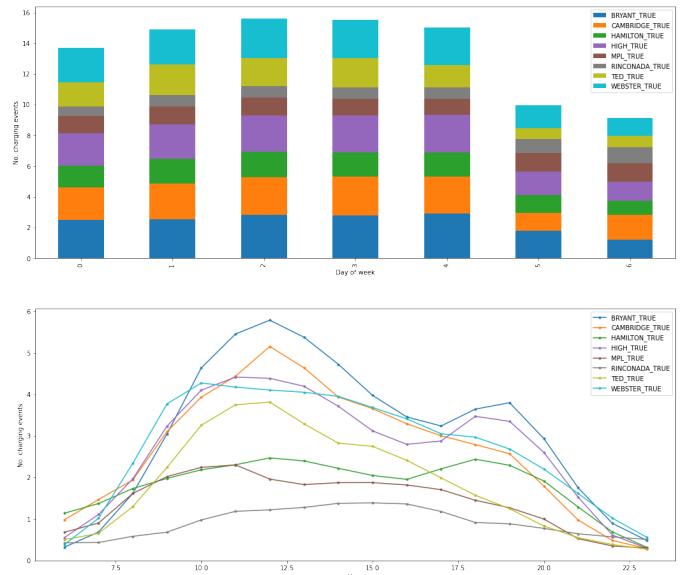


Fig. 13. Weekly and hourly patterns across clusters

The other clusters have a strong peak in the morning and a small peak in the evening. The morning peak is attributed to employees arriving at work while the evening peak is due to people going shopping or eating out at restaurants.

C. Censoring strategy

In Table VI, we have summarized the censoring configurations that we tested for each cluster. For each censor strategy, the percentages of observations that are equal to or above the capacity imposed by the censoring threshold in the training data are shown.

TABLE VI
CENSORING PERCENTAGES FOR EACH CLUSTER, FOR EACH CENSORING STRATEGY, TOGETHER WITH MEAN CENSORSHIP FOR EACH STRATEGY

Capacity	Bryant	Cambridge	Hamilton	High	MPL	Rinc.	Ted	Webster
Dyn. 1	0.034	0.112	0.203	0.019	0.053	0.167	0.005	0.190
Dyn. 2	0.124	0.193	0.203	0.084	0.138	0.167	0.015	0.302
Static 3	0.129	0.234	0.203	0.117	0.140	0.167	0.077	0.374
Static 2	0.270	0.350	0.203	0.199	0.323	0.167	0.114	0.484

In Fig. 14, we have visualized the Dynamic 1 strategy throughout the time period across the eight clusters.

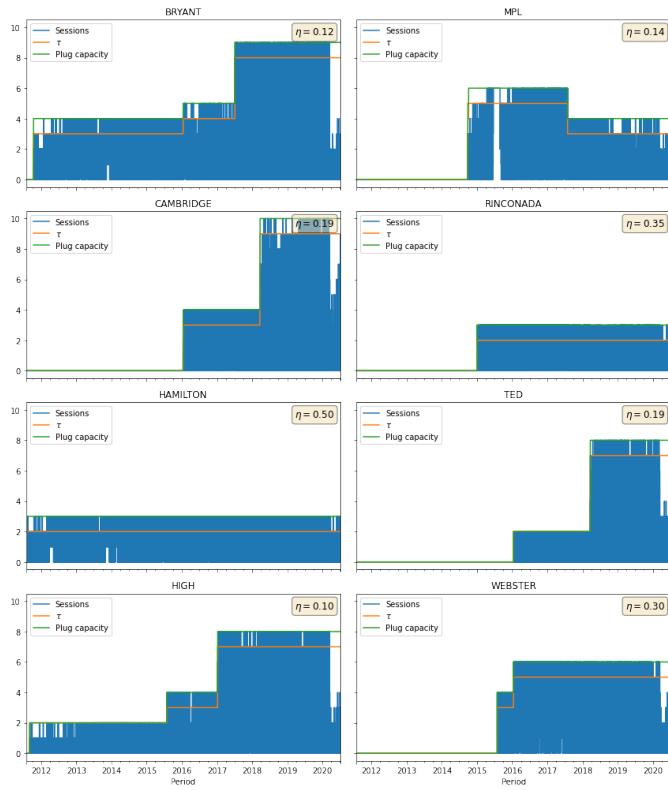


Fig. 14. Time series of charging events for all clusters throughout the period. The plug capacity and the censoring threshold τ are also shown

D. Results

In this section, we present a selection of visualizations that were not incorporated in the main report, in an effort to maintain its conciseness.

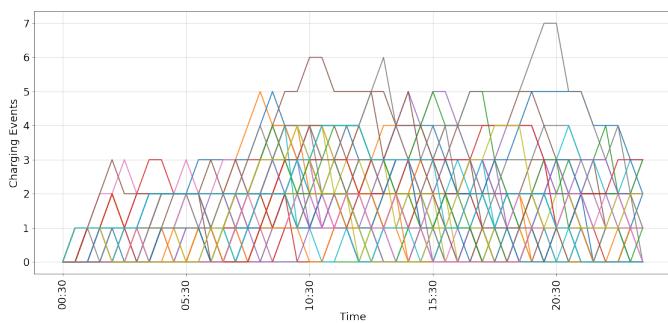


Fig. 15. Realizations of T-GCN-A coming from 50 simulated days (the median and CI of the realizations are shown in Fig. 11. Each line is a simulation.

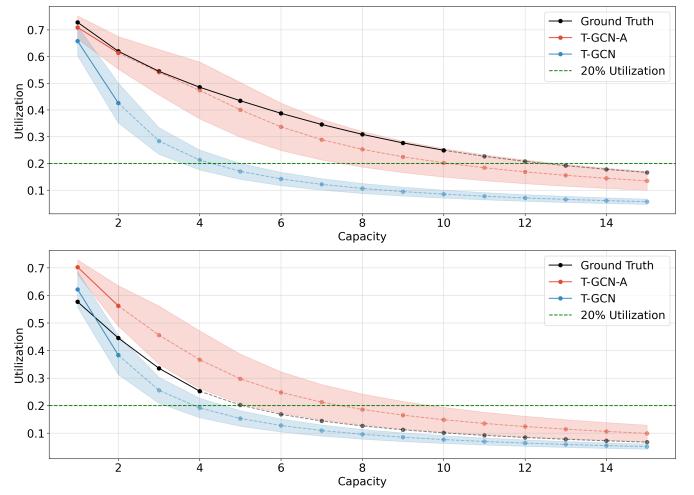


Fig. 16. Utilization rates of Cambridge (upper) and MPL (lower) as estimated by the aware and unaware T-GCN model compared to the true utilization rate. The standard deviation is shown in shaded colors

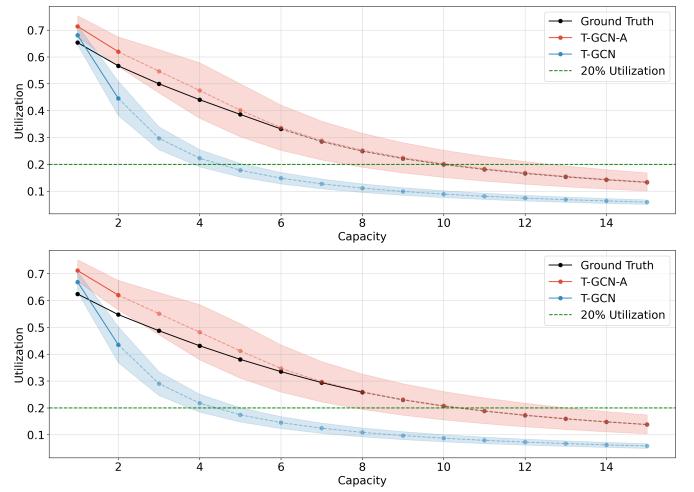


Fig. 17. Utilization rates of Webster (upper) and High (lower) as estimated by the aware and unaware T-GCN model compared to the true utilization rate. The standard deviation is shown in shaded colors

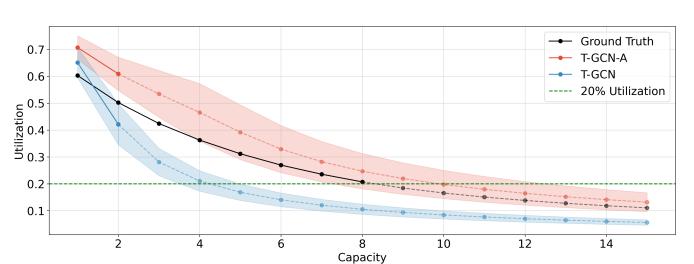


Fig. 18. Utilization rates of Ted as estimated by the aware and unaware T-GCN model compared to the true utilization rate. The standard deviation is shown in shaded colors

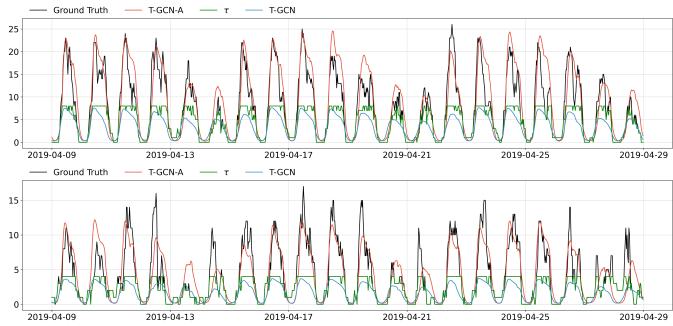


Fig. 19. Forecasts aggregated on superclusters Webster, Bryant, Hamilton and High (upper) and Ted and Cambridge (lower)

TABLE VII

LOSS TABLE FOR 30-MINUTE AND 24-HOUR FORECAST LEADS. WE DENOTE THE MEAN AND STANDARD DEVIATION FROM FOUR DIFFERENT WEIGHT INITIALIZATIONS AND TRAINING SET SHUFFLES. BOLD METRICS COMPARE AWARE AND UNAWARE MODEL VERSIONS, WHILE UNDERLINED METRICS HIGHLIGHT THE TOP MODEL FOR EACH CENSOR SCHEME

FL	Censoring	Model	PNLL		MAE		MSE	
			Aware	Unaware	Aware	Unaware	Aware	Unaware
1	Static 3	T-GCN	-0.07 ± 0.00	0.06 ± 0.01	0.73 ± 0.01	0.84 ± 0.00	1.30 ± 0.03	2.02 ± 0.03
		LSTM	-0.06 ± 0.01	0.00 ± 0.00	0.80 ± 0.02	0.80 ± 0.00	1.39 ± 0.07	1.67 ± 0.01
		GRU	-0.13 ± 0.01	-0.01 ± 0.01	0.66 ± 0.02	0.78 ± 0.01	1.12 ± 0.08	1.84 ± 0.04
		AR	0.10 ± 0.03	0.01 ± 0.00	1.16 ± 0.07	0.80 ± 0.01	3.52 ± 0.50	1.53 ± 0.01
	Static 2	TGCN	-0.02 ± 0.01	0.24 ± 0.01	0.83 ± 0.02	0.98 ± 0.00	1.59 ± 0.05	2.79 ± 0.02
		LSTM	-0.01 ± 0.01	0.21 ± 0.01	0.89 ± 0.02	0.98 ± 0.01	1.73 ± 0.09	2.66 ± 0.03
		GRU	-0.09 ± 0.02	0.20 ± 0.01	0.75 ± 0.02	0.93 ± 0.03	1.41 ± 0.10	2.75 ± 0.08
		AR	0.14 ± 0.04	0.16 ± 0.01	1.21 ± 0.08	0.95 ± 0.01	3.97 ± 0.80	2.36 ± 0.06
48	Static 3	T-GCN	0.13 ± 0.01	0.38 ± 0.02	0.92 ± 0.01	1.11 ± 0.01	1.73 ± 0.01	2.88 ± 0.05
		LSTM	0.17 ± 0.02	0.29 ± 0.02	0.99 ± 0.02	1.08 ± 0.02	1.89 ± 0.09	2.45 ± 0.06
		GRU	0.13 ± 0.03	0.27 ± 0.05	0.92 ± 0.04	1.01 ± 0.02	1.71 ± 0.17	2.30 ± 0.12
		AR	0.16 ± 0.01	0.21 ± 0.02	1.03 ± 0.02	1.03 ± 0.01	2.03 ± 0.10	2.12 ± 0.09
	Static 2	T-GCN	0.15 ± 0.02	0.59 ± 0.03	0.98 ± 0.02	1.22 ± 0.01	1.90 ± 0.08	3.55 ± 0.09
		LSTM	0.21 ± 0.03	0.44 ± 0.01	1.02 ± 0.02	1.18 ± 0.00	2.07 ± 0.10	3.10 ± 0.02
		GRU	0.17 ± 0.02	0.47 ± 0.02	0.96 ± 0.02	1.16 ± 0.01	1.91 ± 0.12	3.23 ± 0.08
		AR	0.19 ± 0.02	0.36 ± 0.03	1.09 ± 0.03	1.14 ± 0.01	2.27 ± 0.14	2.86 ± 0.13

TABLE VIII
CENSORSHIP FOR TEST SET

	BRYANT	CAMBRIDGE	HAMILTON	HIGH	MPL	RINCONADA	TED	WEBSTER
Dynamic 1	0.054	0.020	0.502	0.046	0.115	0.216	0.034	0.167
Dynamic 2	0.094	0.049	0.502	0.105	0.304	0.216	0.062	0.265
Static 3	0.382	0.388	0.502	0.368	0.115	0.216	0.267	0.366
Static 2	0.485	0.514	0.502	0.471	0.304	0.216	0.400	0.477

TABLE IX
CENSORSHIP FOR TRAIN SET

	BRYANT	CAMBRIDGE	HAMILTON	HIGH	MPL	RINCONADA	TED	WEBSTER
Dynamic 1	0.067	0.252	0.535	0.105	0.178	0.355	0.262	0.291
Dynamic 2	0.171	0.366	0.535	0.166	0.377	0.355	0.269	0.391
Static 3	0.374	0.341	0.535	0.423	0.348	0.355	0.316	0.503
Static 2	0.518	0.492	0.535	0.542	0.544	0.355	0.349	0.632