Diffusion-aware Censored Gaussian Processes for Demand Modelling

Filipe Rodrigues
Technical University of Denmark
rodr@dtu.dk

Abstract

Inferring the true demand for a product or a service from aggregate data is often challenging due to the limited available supply, thus resulting in observations that are censored and correspond to the realized demand, thereby not accounting for the unsatisfied demand. Censored regression models are able to account for the effect of censoring due to the limited supply, but they don't consider the effect of substitutions, which may cause the demand for similar alternative products or services to increase. This paper proposes Diffusion-aware Censored Demand Models, which combine a Tobit likelihood with a graph-based diffusion process in order to model the latent process of transfer of unsatisfied demand between similar products or services. We instantiate this new class of models under the framework of GPs and, based on both simulated and real-world data for modeling sales, bike-sharing demand, and EV charging demand, demonstrate its ability to better recover the true demand and produce more accurate out-of-sample predictions.

1 Introduction

Learning well-specified probabilistic models capable of dealing with censored data is a long-standing challenge of the statistical sciences and machine learning. Censoring occurs when the value of a given observation or measurement is only partially known, with the true value being *latent*. Censoring is a process that occurs naturally in many research fields, such as economics, natural sciences, and medicine [Breen, 1996]. In this paper, we are particularly interested in censoring in the context of demand modeling. Demand modeling is crucial across various application domains, such as retail and e-commerce, energy, transportation and logistics, healthcare, telecommunications, manufacturing, tourism, etc., as it enables accurate forecasting, resource optimization, and strategic decision-making by understanding and predicting consumer needs, market trends, and system requirements. When modeling the aggregate demand for a product or service, censoring occurs whenever it runs out of supply, thus resulting in observations that are upper-bounded by the available supply (*satisfied demand*) and, therefore, don't correspond to the actual (*true*) demand [Heien and Wesseils, 1990], which is the quantity of interest to be modeled. From a learning perspective, if the dependent variable is censored for a non-neglectable fraction of the observations, then the parameter estimates obtained (e.g., by standard regression approaches such as OLS) will be inherently biased.

In the statistics literature, Tobit regression models [Amemiya, 1984; McDonald and Moffitt, 1980] constitute the main workhorse for handling censored observations. They provide a framework capable of accounting for left- and right-censored data through a carefully designed likelihood function. However, they rely on the assumption that individual censoring processes occur in isolation, which is unrealistic for many real-world applications. Consider the problem of modeling the charging demand of electric vehicles (EVs) in an area. If, at certain periods of the day, all chargers are occupied, then the observed charging demand may not correspond to the true demand, which will likely be higher than the value observed. Critically though, that *unsatisfied demand* above the available supply is not lost and, at least in a fraction of the cases, it will be transferred to the nearest available charger [Hipolito *et al.*, 2022], thus resulting in an unusually high demand being observed at that charging location. Besides leading to underestimation of the true demand in the original area, this may erroneously lead one to believe that the need for charging in nearby areas is higher than it actually is. More generally, we can consider the problem of modeling aggregate demand data for a given product or service. If the supply for that product reaches zero, then there is a high likelihood that that will result in a higher demand being observed for other similar/substitute products [Wan *et al.*, 2018] - e.g., supermarket customers buying more rucola because it has run out of lettuce to sell, or simply buying from a different brand.

This paper aims to model the ubiquitous process of demand transfer between related products or services due to censoring from an aggregate (market-level) perspective. It proposes a new class of models, which we refer to as Diffusion-aware Censored Demand Models, that integrates ideas from Tobit models [Tobin, 1958; Amemiya, 1984] with a graph diffusion process to model the latent process of transfer of unsatisfied demand (graph-based demand propagation). In doing so, we provide a statistically sound methodology to more accurately infer the latent true demand for the product subject to censoring by taking into account the unusually high demand for similar products, as well as to obtain adjusted estimates for the demand of the latter by accounting for the "spillover effect". By modeling the demand transfer process that occurs between similar products or services when their observed demand is subject to censoring due to limited supply, we are able to obtain more unbiased regression models of the true aggregate

demand. We instantiate this new class of models under the framework of Gaussian processes (GPs) and propose Diffusion-aware Censored GPs, thus allowing us to jointly model multiple correlated time-series of aggregate demand data corresponding to different products or services. We begin by empirically validating the proposed approach using artificial data under carefully constructed scenarios, whereby we analyze its strengths and limitations. We then demonstrate the proposed approach using three real-world datasets for modeling sales data, bike-sharing demand, and EV charging demand. Through these experiments, we pinpoint the advantages of the proposed approach over standard censored regression, thus demonstrating its ability to better recover the true demand and produce more accurate out-of-sample predictions.

2 Related Work

Censoring arises naturally in many research fields. Besides the cases in economics discussed above, it occurs, for example, in the natural sciences whenever the true value of interest is outside the measurable range of a measuring instrument. Similarly, in medicine, censoring arises, for example, when measuring the survival time in clinical trials, since one may know that the survival time surpasses the present moment, but one cannot know the total survival time. Due to the pervasiveness and, oftentimes, critical importance of censored variables, it is unsurprising that a significant portion of the literature is dedicated to them. Historically, learning well-specified models of censored data has always been an important focus of statistical research, with Tobit models constituting a unifying probabilistic approach for censored data. With the advances in machine learning, the limitations of Tobit models associated with its linearity assumptions have been relaxed through the use of non-linear models such as Gaussian processes [Basson *et al.*, 2023; Gammelli *et al.*, 2020] and Neural Networks [Hüttel *et al.*, 2023]. However, when using a Tobit likelihood, inference is no longer tractable. Therefore, a considerable portion of the literature is dedicated to developing approximate inference approaches (e.g., [Chib, 1992; Groot and Lucas, 2012]).

One domain in which censored data is prevalent is urban mobility, from which we draw several case studies. People want to travel from A to B at a given time using a given transportation mode, but they often face supply constraints that force them to, for example, opt for an alternative transportation mode, change their departure time, or not travel at all. Therefore, modeling and inferring true demand from censored demand observations in space and time is of vital importance for optimizing transportation systems. For example, [Gammelli *et al.*, 2020] proposed an approach based on independent Gaussian processes and a Tobit likelihood to model the demand for shared mobility services at individual locations, thus resulting in more accurate forecasts of the latent true aggregate demand. Similarly, [Hüttel *et al.*, 2022] proposed using neural networks and censored quantile regression models to model EV car-sharing usage. [Xie *et al.*, 2023] further combine the predictions of a censored GP model with an optimization procedure to efficiently allocate resources in bike-sharing systems.

Although these models are able to account for the censoring induced by the limited supply and adjust their estimates accordingly, they ignore the latent process of transfer of unsatisfied demand that the censoring process generates. For example, in bike-sharing systems, there is a high likelihood that users go to the nearest hub if there are no bikes available at the current hub, thus resulting in an abnormally high demand for the other hub. Researchers in econometrics are well aware of the importance of accounting for *substitution effects* and have proposed several approaches to model them [Wan *et al.*, 2018; Schaafsma and Brouwer, 2020; Domarchi and Cherchi, 2023]. However, their focus is on micro-econometric approaches that model individual choice behavior and, therefore, require individual choice data for calibration, which can be difficult and costly to obtain. On the other hand, market-level data of aggregate demand is typically readily available. Contrasting with the existing literature, this paper proposes a novel framework to model this demand transfer process from an aggregate (market-level) perspective, thus resulting in more unbiased regression models of the true (latent) aggregate demand.

3 Background: Censored Gaussian Processes

3.1 Censored Gaussian Processes

Given a dataset $\mathcal{D} = \{(\mathbf{x}_n, y_n, c_n)\}_{n=1}^{N}$, which, besides the covariates $\mathbf{x}_n$ and respective targets y_n, also contains a binary indicator variable c_n of whether the n^{th} observation is subject to censoring, the goal of censored regression [DeMaris, 2004; Greene, 2003] is to learn a function of the *latent* true (uncensored) value $f_n = f(\mathbf{x}_n)$. When modeling demand data, we are particularly interested in the case of right-censoring, where y_n is upper-bounded by a given threshold u_n corresponding to the available supply, such that

$$y_n = \begin{cases} f_n, & \text{if } f_n < u_n \\ u_n, & \text{if } f_n \geq u_n \end{cases} . \tag{1}$$

The censoring indicator variable c_n can then be formally defined as $c_n = \mathbb{1}(f_n \geq u_n)$. Traditionally, when modeling scalar-valued censored data, a reasonable choice for an observation model is the well-known Tobit likelihood [Tobin, 1958; Greene, 2003], or a type-I Tobit model according to the taxonomy of [Amemiya, 1984]. For right-censored data, the likelihood of y_n under a censored Gaussian distribution is given by

$$p(y_n \,|\, f_n) = \mathcal{N}(y_n \,|\, f_n, \sigma^2)^{(1-c_n)} (1 - \Phi(y_n \,|\, f_n, \sigma^2))^{c_n}, \tag{2}$$

where Φ is the Gaussian cumulative density function (CDF). Note that this likelihood can be generalized to other distributions such as Poisson or Negative Binomial as discussed, for example, in [Gammelli *et al.*, 2022]. Censored GP regression [Basson *et al.*, 2023; Groot and Lucas, 2012] then proceeds by placing a GP prior on $f(\mathbf{x}) \sim \mathcal{GP}(0, \kappa(\mathbf{x}, \mathbf{x}'))$, and performing Bayesian inference to compute the posterior distribution over the true (uncensored) values $\{f_n\}$ or the predictive distribution for a new test point $\mathbf{x}_*$.

4 Problem Formulation

We consider multivariate time-series data comprising input-output pairs, $\{\mathbf{X}^{(tp)} \in \mathbb{R}^{N_t \times N_p \times D}, \mathbf{Y}^{(tp)} \in \mathbb{R}^{N_t \times N_p}\}$, where N_t denotes the number of temporal points, N_p the number of products or services, and $D = 1 + D_p$ the input dimensions, with D_p being the number of product features. As it is common in the GP literature, we let $\mathbf{X} = \text{vec}(\mathbf{X}^{(tp)}) \in \mathbb{R}^{N \times D}$, $\mathbf{Y} = \text{vec}(\mathbf{Y}^{(tp)}) \in \mathbb{R}^{N \times 1}$, where $N = N_t N_p$ and the operator $\text{vec}(\cdot)$ simply converts the data into vector form, while preserving the ordering by time first and then

products. For convenience, we use $\mathbf{X}_{n,k} = \mathbf{X}^{(tp)}_{n,k}$ and $\mathbf{Y}_{n,k} = \mathbf{Y}^{(tp)}_{n,k}$ to index the aggregate demand data for product k at time index n. Our goal is then to learn a random function $f : \mathbb{R}^D \to \mathbb{R}$, on which we place a zero-mean GP prior with a fully-factorized likelihood, thus leading to the following generative process:

$$f(\mathbf{x}) \sim \mathcal{GP}(0, \kappa(\mathbf{x}, \mathbf{x}')), \quad p(\mathbf{Y} \mid \mathbf{f}) = \prod_{n=1}^{N_t} \prod_{k=1}^{N_p} p(\mathbf{Y}_{n,k} \mid \mathbf{f}_{n,k}), \quad (3)$$

where $\mathbf{f}_{n,k} = f(\mathbf{X}_{n,k})$, and we slightly abuse notation by writing $f(\mathbf{x}) = f(t, \mathbf{p})$ and $\kappa(\mathbf{x}, \mathbf{x}') = \kappa(t, \mathbf{p}, t', \mathbf{p}')$, with $\mathbf{p}$ denoting the product (or service) features describing their characteristics including their spatial location when demand is distributed in both space and time, as in the cases of demand for EV charging or Mobility-on-Demand services. Similarly to standard censored regression, our goal is two-fold: i) compute the posterior distribution over the (latent) true demand $\mathbf{f}$ given the observed demand $\mathbf{Y}$ (censored and subject to substitutions), and ii) predict the true demand $\mathbf{f}_*$ for a new test point $\mathbf{x}_*$ (e.g., for forecasting).

5 Diffusion-aware Censored Gaussian Processes

By leveraging the Tobit likelihood in Eq. 2, Censored GP regression (Section 3.1) allows us to account for the bias induced by censoring by essentially giving the latent function f added flexibility to take higher values for censored observations ($c_n = 1$). However, when modeling aggregate demand data, this approach doesn't account for the exchange of demand that may occur between similar products or services when some of them run out of supply, thus resulting in abnormal demand being observed for "substitute products" [Wan *et al.*, 2018; Schaafsma and Brouwer, 2020]. Let $\mathbf{f}_n = (f_{n,1}, ..., f_{n,N_p})$ denote a vector containing the true latent aggregate demand for the N_p products or services at time index n. We can formally define the concept of unsatisfied demand for a product k as: $\max(f_{n,k} - u_{n,k}, 0)$. We propose to model the process of transfer of unsatisfied demand across products as a non-linear diffusion process on a graph.

5.1 Transition Dynamics

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ be a weighted undirected graph, where $\mathcal{V}$ is a set of nodes $|\mathcal{V}| = N_p$ representing products or services, $\mathcal{E}$ is a set of edges, and $\mathbf{W} \in \mathbb{R}^{N_p \times N_p}$ is weighted adjacency matrix representing the similarity between products. We define the edge weights with the aid of kernel as

$$\mathbf{W}_{i,j} = \begin{cases} \kappa_{\text{diff}}(\mathbf{p}_i, \mathbf{p}_j), & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}, \quad (4)$$

with $\kappa_{\text{diff}}(\mathbf{p}_i, \mathbf{p}_j)$ being determined by an RBF kernel:

$$\kappa_{\text{diff}}(\mathbf{p}_i, \mathbf{p}_j) = \exp\{-\|\mathbf{p}_i - \mathbf{p}_j\|^2 / \ell^2_{\text{diff}}\}. \quad (5)$$

The underlying assumption is that products with similar characteristics are likely to be used as replacements when a given product is out of supply, with the lengthscale parameter ℓ_{diff} controlling how much customers are willing to tolerate differences in product characteristics $\mathbf{p}$. Diffusion of the unsatisfied demand to the same node (self-loops) is obviously not allowed. We then define the state transition matrix as

$$\mathbf{T} = \text{diag}(\mathbf{W}\mathbf{1}_{N_p})^{-1}\mathbf{W}, \quad (6)$$

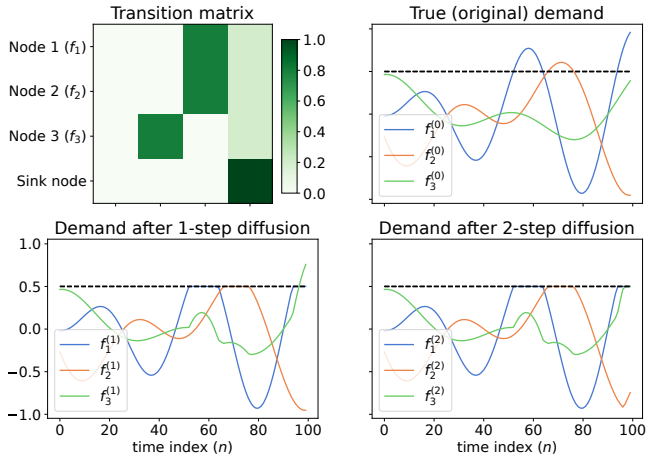


Figure 1: Diffusion process demo for 3 time-series using the transition matrix on the top-left (node 1 sends 80% of its unsatisfied demand to node 3 and 20% to the sink node, and so on). The true demand (top-right) that is above the supply (dashed line) gets transferred to the other nodes (incl. sink node) over the 2 diffusion steps (bottom row).

where $\mathbf{1}_{N_p}$ denotes an $N_p \times N_p$ matrix of ones. Eq. 6 essentially produces a row-normalized version of $\mathbf{W}$, such that the diffusion of unsatisfied demand from service i to service $j \neq i$ becomes: $\mathbf{T}_{i,j} = \kappa_{\text{diff}}(\mathbf{p}_i, \mathbf{p}_j) / \sum_{l=1}^{N_p} \kappa_{\text{diff}}(\mathbf{p}_i, \mathbf{p}_l)$. However, this assumes that all unsatisfied demand at a given node i is transferred to the neighboring nodes $j \in \mathcal{N}_i$, regardless of how different the service characteristics $\mathbf{p}_i$ are from $\mathbf{p}_j$. We can relax this assumption by introducing a "sink" node in the graph that connects to all the other nodes and extending $\mathbf{W}$ to be $(N_p+1) \times (N_p+1)$, with $\mathbf{W}_{i,\text{sink}} = \pi_{\text{sink}}, \forall i \neq \text{sink}$, while $\mathbf{W}_{\text{sink},i} = 0$ and $\mathbf{W}_{\text{sink,sink}} = 1$, in order to ensure that once demand enters the sink node it never leaves. Therefore, the parameter $\pi_{\text{sink}} \in \mathbb{R}$ controls the likelihood of customers simply giving up and deciding not to consume any alternative product when faced with the lack of supply for their intended product.

Based on $\mathbf{T}$, we propose to model the diffusion dynamics as

$$\begin{aligned} \mathbf{f}_n^{(i+1)} &= \mathbf{f}_n^{(i)} + \underbrace{\max(\mathbf{f}_n^{(i)} - \mathbf{u}_n, 0)\mathbf{T}}_{\text{incoming demand from others}} - \underbrace{\max(\mathbf{f}_n^{(i)} - \mathbf{u}_n, 0)}_{\text{outgoing demand}}, \\ &= \mathbf{f}_n^{(i)} + \max(\mathbf{f}_n^{(i)} - \mathbf{u}_n, 0)(\mathbf{T} - \mathbf{I}) \end{aligned} \quad (7)$$

where $\mathbf{u}_n$ is a vector of the corresponding available supply at time index n, thus making $\max(\mathbf{f}_n^{(i)} - \mathbf{u}_n, 0)$ the unsatisfied demand above the available supply at diffusion step i. If we assume that customers have no access to information about the available supply, we can apply the transition operator in Eq. 7 multiple times in order to simulate the multi-step process of trying to use a service, finding out that there is no available supply, and then searching for an alternative, as can happen, for example, for EV charging and Mobility-on-Demand [Unterluggauer *et al.*, 2023]. In practice, we limit the maximum number of diffusion steps to N_{diff}. Figure 1 illustrates two steps of this diffusion process for 3 time-series.

However, if we assume that customers have access to information about the available supply of all products (incl. alternatives), then it is unrealistic to consider a multi-step diffusion process. Instead, we propose introducing a time-dependent graph $\mathcal{G}_n = \{\mathcal{V}, \mathcal{E}_n\}$, such that $\mathcal{E}_n = \{(i,j) \mid u_{n,j} > y_{n,j}\}$, thereby not

allowing connections to products that are also out of supply at time index n, and applying a single diffusion step (Eq. 7) in $\mathcal{G}_n$. This modeling approach encodes the intuitive assumption that customers, when faced with a product out of supply, will look at all the available alternatives once and choose one or none (sink node).

5.2 Likelihood

After applying the graph diffusion process N_{diff} times, the additional demand observed at time index n due to the demand transfer process caused by the limited supply of other products or services can be computed as

$$\mathbf{d}_n = \max(\mathbf{f}_n^{(N_{\text{diff}})} - \mathbf{f}_n, 0). \tag{8}$$

Based on this "diffused demand" $\mathbf{d}_n$, we propose to model the likelihood of the observed demand values, accounting for censoring and for the diffusion of unsatisfied demand, as

$$p(\mathbf{Y}_n \mid \mathbf{f}_n) = \prod_{k=1}^{N_p} \mathcal{N}(y_{n,k} \mid f_{n,k} + d_{n,k}, \sigma^2)^{(1-c_{n,k})} \times (1 - \Phi(y_{n,k} \mid f_{n,k} + d_{n,k}, \sigma^2))^{c_{n,k}}. \tag{9}$$

This likelihood can be intuitively understood as a combination of: i) a Tobit formulation to account for the observed demand for an out-of-supply product k ($c_{n,k} = 1$) potentially being lower than the latent process expected it to be, and ii) an added term $d_{n,k}$ to the expected demand value of product k to account for the potentially extra (unexpected) demand observed due to similar products being out of supply. Kindly note that although Eq. 9 factorizes across services k, the likelihoods for different products are no longer independent due to the diffusion process, which ties them together. Also, note that using the likelihood in Eq. 9 implies access to information about the available supply $\mathbf{u}_n$, which is required for the diffusion process (Eq. 7). This contrasts with regular Tobit regression (Eq. 2), where only access to a censoring indicator variable $\mathbf{c}_n$ is required. Nevertheless, we argue that this is a reasonable assumption since, for most real-world demand modeling applications, supply information is typically readily available along with the demand data. Lastly, it should be noted that although we use Gaussian processes as the backbone for our proposed approach, the methodology described above can be generalized to other modeling approaches such as neural networks and linear models.

5.3 Inference

Scalability is often a concern when modeling large datasets with GPs. In our case, we are interested in jointly modeling potentially-long time-series of observed demand across different products or services, thus further aggravating this concern, since computing the posterior distribution $p(\mathbf{f} \mid \mathbf{X})$ typically has a cubic cost of $\mathcal{O}(N_t^3 N_p^3)$. Therefore, we leverage the work of [Hamelijnck *et al.*, 2021], which combines spatio-temporal filtering with natural gradient variational inference to achieve linear scalability with respect to time.

We begin by assuming that our kernel is both Markovian and separable between time and products, i.e. $\kappa(t, \mathbf{p}, t', \mathbf{p}') = \kappa_t(t, t')\kappa_{\mathbf{p}}(\mathbf{p}, \mathbf{p}')$. Under this assumption, the model in Eq. 3 can be reformulated as a state space model, thus reducing the computational scaling to linear in N_t. The GP prior can be written as a stochastic differential equation (SDE) [Chang *et al.*, 2020], which, in this case, requires marginalizing to a finite set of products, $\mathbf{P} \in \mathbb{R}^{N_p \times D_p}$, giving, $\mathrm{d}\bar{\mathbf{f}}(t) = \mathbf{F}\bar{\mathbf{f}}(t)\mathrm{d}t + \mathbf{L}\mathrm{d}\boldsymbol{\beta}(t)$, where $\bar{\mathbf{f}}(t)$ is the Gaussian distributed state over products $\mathbf{P}$ at time t, and $\mathbf{F}$ and $\mathbf{L}$ are the feedback and noise effect matrices. The function value $\mathbf{f}_n$ can be extracted from the state by a matrix $\mathbf{H}$ as $\mathbf{f}_n = \mathbf{H}\bar{\mathbf{f}}(t_n)$. For a fixed step size $\Delta_n = t_{n+1} - t_n$, the resulting discrete model is [Chang *et al.*, 2020]:

$$\begin{aligned} \bar{\mathbf{f}}(t_{n+1}) &= \mathbf{A}_n \bar{\mathbf{f}}(t_n) + \mathbf{q}_n, \\ \mathbf{Y}_n \mid \bar{\mathbf{f}}(t_n) &\sim p(\mathbf{Y}_n \mid \mathbf{H}\,\bar{\mathbf{f}}(t_n)), \end{aligned} \tag{10}$$

where $\mathbf{A}_n = \exp(\mathbf{F}\Delta_n)$ is the linear state transition matrix, and $\mathbf{q}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_n)$, with $\mathbf{Q}_n$ denoting the process noise covariance. If $p(\mathbf{Y}_n \mid \mathbf{H}\,\bar{\mathbf{f}}(t_n))$ is Gaussian, then Kalman smoothing algorithms to be employed to perform inference in linear time in N_t. Since the likelihood of our proposed model (Eq. 9) is non-Gaussian, we apply conjugate-computation variational inference (CVI) [Khan and Lin, 2017], whereby an *approximate likelihood*, that is conjugate to the prior (in this particular case, Gaussian), with free parameters $\tilde{\boldsymbol{\lambda}}$, is considered to enable efficient computation of the natural gradients of the evidence lower bound (ELBO).

As shown by [Khan and Lin, 2017], if the chosen approximate likelihood is conjugate to the prior, then performing natural gradient VI is equivalent to a two-step Bayesian update of the form:

$$\tilde{\boldsymbol{\lambda}} \leftarrow (1-\beta)\tilde{\boldsymbol{\lambda}} + \beta \frac{\partial \mathbb{E}_{q(\mathbf{f})}[\log p(\mathbf{Y} \mid \mathbf{f})]}{\partial \boldsymbol{\mu}}, \qquad \boldsymbol{\lambda} \leftarrow \boldsymbol{\eta} + \tilde{\boldsymbol{\lambda}}, \tag{11}$$

where $\boldsymbol{\lambda}$ are the natural parameters of the approximate posterior distribution (with corresponding mean parameters $\boldsymbol{\mu}$), $\boldsymbol{\eta}$ are the natural parameters of the prior, and $\tilde{\boldsymbol{\lambda}}$ are the natural parameters of the (approximate) likelihood contributions.

Let $\mathcal{N}(\tilde{\mathbf{Y}} \mid \mathbf{f}, \tilde{\mathbf{V}})$ be an approximate likelihood parameterised by covariance $\tilde{\mathbf{V}} = (-2\tilde{\boldsymbol{\lambda}}^{(2)})^{-1}$ and mean $\tilde{\mathbf{Y}} = \tilde{\mathbf{V}}\tilde{\boldsymbol{\lambda}}^{(1)}$, with $\tilde{\boldsymbol{\lambda}} = \{\tilde{\boldsymbol{\lambda}}^{(1)}, \tilde{\boldsymbol{\lambda}}^{(2)}\}$. The approximate posterior then has the form $q(\mathbf{f}) \propto \mathcal{N}(\tilde{\mathbf{Y}} \mid \mathbf{f}, \tilde{\mathbf{V}})\, p(\mathbf{f})$. Since the GP prior is Markov and the approximate likelihood factorizes across time [Hamelijnck *et al.*, 2021], the approximate GP posterior is also Markov [Tebbutt *et al.*, 2021]. Therefore, and since the approximate likelihood is now Gaussian, the marginals $q(\mathbf{f}_n)$ can be computed through linear filtering and smoothing applied to Eq. 10, but with $\tilde{\mathbf{Y}}_n \mid \bar{\mathbf{f}}(t_n) \sim \mathcal{N}(\tilde{\mathbf{Y}}_n \mid \mathbf{H}\,\bar{\mathbf{f}}(t_n), \tilde{\mathbf{V}})$ as the measurement model. Given the marginals, Eq. 11 can be used to give the new likelihood parameters $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{V}}$. The ELBO can be written as a sum of three terms:

$$\begin{aligned} \mathcal{L} = \mathbb{E}_{q(\mathbf{f})}\left[\log \frac{p(\mathbf{Y} \mid \mathbf{f})p(\mathbf{f})}{q(\mathbf{f})}\right] &= \sum_{n=1}^{N_t} \mathbb{E}_{q(\mathbf{f}_n)}[\log p(\mathbf{Y}_n \mid \mathbf{f}_n)] \\ &- \mathbb{E}_{q(\mathbf{f})}[\log \mathcal{N}(\tilde{\mathbf{Y}} \mid \mathbf{f}, \tilde{\mathbf{V}})] + \log \mathbb{E}_{p(\mathbf{f})}[\mathcal{N}[\tilde{\mathbf{Y}} \mid \mathbf{f}, \tilde{\mathbf{V}}]], \end{aligned} \tag{12}$$

where the first term corresponds to the expected log likelihood, the second term is the expected log *approximate likelihood*, and the third is the log marginal likelihood of the approximation posterior. As shown in [Hamelijnck *et al.*, 2021], these terms can also be computed efficiently based on the outputs of the filtering and smoothing algorithms. Computing the updates to the variational parameters in Eq. 11, as well as evaluating the ELBO (Eq. 12), requires computing $\mathbb{E}_{q(\mathbf{f}_n)}[\log p(\mathbf{Y}_n \mid \mathbf{f}_n)]$ and its gradients, for which we rely on quadrature. The gradients are back-propagated through the graph diffusion process using automatic differentiation.

5.4 Diffusion Parameters Learning

Along with the hyper-parameters from the kernels κ_t and $\kappa_{\mathbf{p}}$ of the GP prior, the proposed modeling approach introduces additional hyper-parameters to the likelihood due to the embedded diffusion process. The hyper-parameters of the likelihood in Eq. 9 are then: the observation variance σ^2, the diffusion lengthscale ℓ_{diff}, the sink node parameter π_{diff}, and the number of diffusion steps N_{diff}. While we assume N_{diff} to be fixed, the former three can also be learned from data. Although, for most domains, there probably is domain knowledge to guide the choice of the diffusion lengthscale ℓ_{diff} and the sink node parameter π_{diff} due to their natural interpretations from a behavioral point of view, in some scenarios, a data-driven approach may be preferred. In practice, this can be achieved by, for example, setting these hyper-parameters to reasonable initial values, and then performing type-II maximum likelihood estimation via maximizing the ELBO in Eq. 12 with gradient descent updates alternated with the natural gradient updates in Eq. 11.

6 Experiments

In this section, we empirically demonstrate the capabilities of the proposed Diffusion-aware Censored GP ("DCGP"), focusing on its ability to infer the true demand based on demand observations that are censored and subject to substitutions, as well as its ability to produce more accurate predictions of true demand. We consider two main variants of the proposed approach - one with fixed ℓ_{diff} and another where ℓ_{diff} is learned (referred to as "DCGP-f" and "DCGP-l", respectively), which we compare with Censored GPs ("CGP") and standard (Non-Censored) GPs ("NCGP") fitted to the observed demand. For reference, we also provide results for an "Oracle" GP fitted to the true (uncensored) demand. We compare the different modeling approaches across three metrics: RMSE, R-squared (R^2), and negative log predictive density (NLPD). Source code for all the experiments is provided at: https://bit.ly/3E2kdSJ

6.1 Artificial Data

Independent time-series. We begin by demonstrating the proposed approach using independently generated time-series data based on sinusoidal functions. We construct three datasets. In Dataset A, we draw samples from two sinusoidal functions (simulating the true demand) with some observation (Gaussian) noise added, and then simulate the censoring process by setting a constant censoring threshold (supply). The demand above the threshold is sent to the other time-series, and if the resulting demand for the other time-series now becomes higher than the threshold, then we throw away the unsatisfied demand (see Figure 2a). Dataset B is similar to Dataset A, but the threshold is sampled randomly from a uniform distribution for each time step, thus resulting in a stochastic supply that causes very sporadic supply shortages rather than shortages lasting for longer time intervals as in Dataset A. Dataset C is similar to Dataset A but considers three sinusoidal time-series (instead of two) mimicking the demand of three products. Whenever Product 1 or Product 3 are out of supply, their unsatisfied demand goes to Product 2, while the unsatisfied demand from the latter is distributed equally among Products 1 and 3. All models use the same underlying GP framework with independent GPs for each product (i.e. $\kappa_{\mathbf{p}}(\mathbf{p},\mathbf{p}') = \mathbb{1}(\mathbf{p}=\mathbf{p}')$) and $\kappa_t(t,t')$ being a Matern kernel. Details are in Appendix A.1.[1]

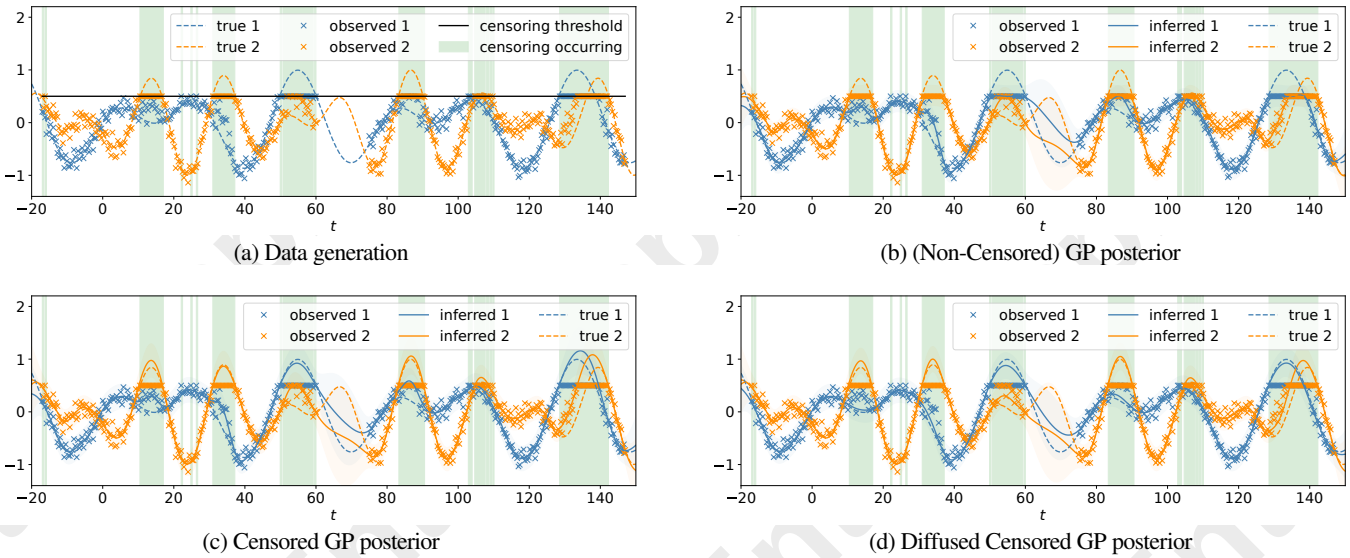
	Model	NLPD	RMSE	R^2	RMSE funct.
Dataset A	True GP (Oracle)	-1.771	0.099	0.960	0.029
	NCGP	0.853	0.188	0.858	0.163
	CGP	-0.674	0.162	0.895	0.130
	DCGP-f	-0.685	**0.108**	**0.953**	**0.051**
	DCGP-l	**-0.751**	0.118	0.944	0.071
Dataset B	True GP (Oracle)	-1.701	0.103	0.957	0.029
	NCGP	-0.505	0.188	0.857	0.158
	CGP	**-1.004**	0.148	0.912	0.111
	DCGP-f	0.911	0.107	**0.954**	0.038
	DCGP-l	0.590	**0.106**	**0.954**	**0.037**
Dataset C	True GP (Oracle)	-2.527	0.104	0.954	0.031
	NCGP	0.805	0.188	0.856	0.162
	CGP	-1.224	0.157	0.900	0.122
	DCGP-f	-1.311	**0.107**	**0.952**	**0.042**
	DCGP-l	**-1.380**	0.130	0.930	0.086

Table 1: Experimental results obtained for the three artificial datasets of independently-generated (uncorrelated) time-series. "RMSE funct." further measures the RMSE to the true underlying (noiseless) function.

Table 1 shows the obtained test set results (a more complete depiction of the results is provided in Appendix B.1). As expected, the Non-Censored GP performs the worst across the three datasets by simply fitting the observed data (see, e.g., Figure 2b). The Censored GP is able to infer that the demand should be higher than observed at censored locations (see Figure 2c), thus achieving better results when compared to the Non-Censored GP in Table 1. However, as Figure 2c shows, the Censored GP is unable to understand that the abnormally high demand observed in some time-series can be explained by the transfer of unsatisfied demand from other time-series that are subject to censoring. On the other hand, the proposed approach is able to account for this transfer and, to a great extent, recover the true underlying demand (Figure 2d), which can be also observed in the results in Table 1. Interestingly, for Datasets A and C, the best results were obtained with a fixed $\ell_{\text{diff}} = 1$, while for Dataset B learning ℓ_{diff} resulted in slightly better results. Across all datasets, the proposed approach is able to achieve results close to the "True GP (Oracle)" reference, thus demonstrating its ability to infer the true underlying functions from the observations.

Spatio-temporal data. We now consider artificial spatio-temporal demand data, whereby demand varies across space and time as, for example, in the cases of EV charging demand or Mobility-on-Demand services. We generate artificial true demand data by first sampling $N_p = 10$ 2-D spatial locations uniformly in the range $[-2,2]$, and then sampling 10 time-series of length $N_t = 400$ from a spatio-temporal GP with a separable kernel with a Periodic+Matern component over time and a Matern component over space, to which we then add Gaussian noise. We simulate the censoring process using a 2-state Markov model, where the states represent "with supply" and "out of supply" states, with the latter employing a fixed demand threshold. The unsatisfied demand is transferred to the nearby locations by simulating the diffusion process described in Section 5.1 with fixed diffusion parameters

[1] All appendices are available at: https://arxiv.org/abs/2501.12354.



(a) Data generation

(b) (Non-Censored) GP posterior

(c) Censored GP posterior

(d) Diffused Censored GP posterior

Figure 2: GP posteriors obtained by the different models for the artificial dataset (Dataset A) depicted in (a). The green-shaded areas indicate time indexes where at least one of the time-series is subject to censoring. As shown in (d), the proposed approach achieves the best approximation to the true demand.

	Model	NLPD	RMSE	R^2	RMSE funct.
Dataset ST-1	True GP (Oracle)	-0.031	0.273	0.924	0.178
	NCGP	0.327	0.368	0.869	0.313
	CGP	0.111	0.313	0.900	0.239
	DCGP-f	**-0.019**	**0.268**	**0.926**	**0.172**
	DCGP-l	0.047	0.281	0.920	0.191
Dataset ST-2	True GP (Oracle)	-0.031	0.273	0.924	0.178
	NCGP	0.285	0.358	0.877	0.300
	CGP	0.140	0.318	0.899	0.246
	DCGP-f (1-step)	0.020	0.282	0.920	0.193
	DCGP-l (1-step)	0.088	0.302	0.909	0.223
	DCGP-f (2-step)	**0.018**	**0.269**	**0.926**	**0.172**
	DCGP-l (2-step)	0.060	0.286	0.918	0.199

Table 2: Experimental results obtained for the two artificially-generated spatio-temporal datasets. "RMSE funct." further measures the RMSE to the true underlying (noiseless) function.

and with a sink node probability of 20%. In this scenario, the product features $\mathbf{p}$ correspond to the spatial locations. We generate two datasets: "Dataset ST-1" and "Dataset ST-2" using $N_{\text{diff}}=1$ and $N_{\text{diff}}=2$ diffusion steps, respectively. We use 130 evenly-sampled observations for training and the remaining for testing. We provide additional details, including figures, in Appendix A.2.

The results obtained are summarized in Table 2 (see Appendix B.1 for additional details). Despite the higher dimensionality of the problem and the correlations between time-series, the results remain similar to the smaller independent time-series experiment above. The Censored GP outperforms its Non-Censored counterpart, while the proposed approach further outperforms both (we plot examples of the inferred true demand by the different approaches in Appendix B.1). Importantly, even in the more challenging setting where ℓ_{diff} has to be learned, the proposed approach is still able to clearly outperform the other baselines. The non-learnable variant (fixed at $\ell_{\text{diff}} = 0.1$) is able to further outperform it, which is unsurprising given that the data was also generated with $\ell_{\text{diff}} = 0.1$. Therefore, this version corresponds to the scenario where domain knowledge can accurately drive the choice of ℓ_{diff}. We provide an in-depth study of the sensitivity to model misspecification in Appendix B.2. In Appendix B.3, we demonstrate the ability to also learn the hyper-parameter π_{diff}. Furthermore, the results in Table 2 also demonstrate that the proposed approach is able to effectively take into account multiple steps of diffusion, thus being able to further improve the true demand estimation in Dataset ST-2.

6.2 Supermaket Sales

We now turn to experiments on real-world supermarket vegetable sales data obtained from [Sup]. Unfortunately, it is impossible to have access to the true demand, as that would imply having access to the customers' intentions. Therefore, as it is standard practice in the censored demand modeling [Gammelli *et al.*, 2020; Hüttel *et al.*, 2022] and, more broadly, in the censored regression literature [Schmee and Hahn, 1979; Li and Wang, 2003], we resort to assuming the observed demand to correspond to the true demand, and simulate the censoring process based on it in a realistic manner. Concretely, we consider the demand time-series for two products and simulate supply shortages using a 2-state Markov model similar to the one used for the artificial spatio-temporal data. The unsatisfied demand for one product is then transferred to the other product. If the other product is also out of supply, then that demand is lost. We randomly sample 90% observations for training and leave 10% for testing. Additional experimental setup details are provided in Appendix A.3. We also experimented with different splits to study the impact of train set size - see Appendix B.4.

Table 3 shows the obtained results for the train and test sets. Note that, in censored regression, train set performance is particularly important, as it is an indicator of the ability to infer the true demand from the observed data. As the results in Table 3 show, the proposed approach is able to infer the true demand more accurately than the other baselines, thus resulting in substantially better

Model	Trainset NLPD	Trainset RMSE	Trainset R^2	Testset NLPD	Testset RMSE	Testset R^2
True GP	1.541	0.527	0.726	1.692	0.551	0.627
NCGP	2.285	0.735	0.469	2.066	0.669	0.446
CGP	2.166	0.714	0.499	2.010	0.653	0.472
DCGP-f	**1.740**	0.591	0.657	**1.968**	0.602	0.553
DCGP-l	1.896	**0.612**	**0.632**	2.103	**0.601**	**0.554**

Table 3: Experimental results obtained for supermarket sales dataset.

Model	Trainset NLPD	Trainset RMSE	Trainset R^2	Testset NLPD	Testset RMSE	Testset R^2
True GP	2.091	1.220	0.984	3.208	6.048	0.436
NCGP	2.674	4.016	0.805	3.297	6.603	0.332
CGP	2.552	4.197	0.769	3.252	6.436	0.365
DCGP-f	2.328	**2.983**	**0.885**	**3.171**	**5.977**	0.448
DCGP-l	**2.311**	2.992	0.882	3.175	5.982	**0.449**

Table 4: Experimental results obtained for bike-sharing dataset.

train and test set performance. The additional results with different train/test splits in Appendix B.4 further show that the superior performance of the proposed approach is observable even for smaller train sets, where some signs of overfitting start to emerge.

6.3 Bike Sharing Demand

We now consider a real-world problem where demand varies across space and time. Concretely, we consider the problem of building a model of the true demand for a hub-based bike-sharing system, where users rent bicycles on an on-demand basis from one of the many hubs available across the city through the use of an app. Bike-sharing systems are known to suffer from strong demand-supply imbalances (e.g., due to normal commuting patterns), thus leading to hubs frequently running out of bicycles. Naturally, if a hub is out of bicycles, the user either walks to the nearest hub with available bicycles or chooses an alternative transportation mode. Therefore, the problem of inferring and predicting the true demand is pivotal for daily operations (e.g., rebalancing [Liu *et al.*, 2016]) and for long-term strategic decisions (e.g., capacity planning and service expansion [Liu *et al.*, 2017]).

Since no ground truth data is available for this problem, we follow the same approach as in [Gammelli *et al.*, 2020] to simulate the censoring process based on the supply data and the observed demand data for 4 bicycle hubs in NYC - which is assumed to correspond to the true demand. The unsatisfied demand above the available supply is transferred to other hubs using the diffusion process from Section 5.1. Details are provided in Appendix A.4. The obtained results in Table 4 again clearly show that the proposed approach is able to more accurately infer the true demand (as indicated by the train set performance), even in the more challenging case where ℓ_{diff} is unknown. The improved estimate of the true demand for the train set then translates into more accurate predictions of the true demand in the test set. Interestingly, in this experiment, the Censored GP performs worse than its Non-Censored counterpart in the train set, which could be explained by its inability to account for the spillover effect to nearby hubs whenever a hub is out of supply. We also consider a simpler version of this experiment with just 2 hubs, which is easier to analyze, in Appendix B.5.

Model	Trainset NLPD	Trainset RMSE	Trainset R^2	Testset NLPD	Testset RMSE	Testset R^2
True GP	1.059	0.339	0.965	1.111	0.443	0.942
NCGP	1.806	1.408	0.407	1.817	1.424	0.397
CGP	1.721	1.357	0.447	1.740	1.375	0.431
DCGP-f	**1.599**	**1.215**	**0.562**	**1.628**	**1.244**	**0.540**
DCGP-l	**1.611**	**1.222**	**0.558**	**1.639**	**1.251**	**0.536**

Table 5: Experimental results obtained for EV charging dataset.

6.4 EV Charging Demand

Lastly, we consider the problem of inferring the true spatio-temporal demand for EV charging from observations consisting of (realized) charging events. We use the agent-based simulation tool, GAIA [Unterluggauer *et al.*, 2023], to generate EV charging data for the whole commune of Frederiksberg in Copenhagen. GAIA is based on the notion of a steady-state SoC distribution [Hipolito *et al.*, 2022] and a probabilistic decision-to-charge model to be able to simulate and analyze different charging strategies, thereby addressing the uncertainties resulting from the additional demand for EV charging by accounting for home charging availability, charging location, and the decision to charge in space and time. Since GAIA was carefully calibrated with real-world data sources, such as data from the energy distribution network and from the Danish National Travel Survey, and because it models individual agent behavior, it provides a unique environment for validating our proposed approach by allowing access to the true demand along with the realized (observable) demand. We use GAIA to simulate 4 weeks of EV charging event data. We partition the area of Frederiksberg into 9 clusters using k-means and time in 5-minute bins. We hold out 30% of data for testing. Appendix A.5 provides additional details on the experimental setup.

Table 5 shows the obtained results. Although the Censored GP already shows some improvement over the Non-Censored GP both in- and out-of-sample by accounting for the effect of censoring, the proposed approach is able to further improve the R^2 by more than 10% on top of the Censored GP, thus underscoring the importance of accounting for the process of transfer/substitutions when building models of the aggregated true demand for products/services for which reasonable alternatives exist. The learned value of ℓ_{diff} converged to a value of approx. 50.6m, which, albeit a bit conservative, is still a reasonable value of how much people are willing to drive to find the nearest charging station with available chargers. Fixing $\ell_{\text{diff}} = 100$m further led to a slight improvement.

7 Conclusion

This paper proposed Diffusion-aware Censored Demand Models, which combine a Tobit likelihood with a graph diffusion process in order to model the latent process of transfer of unsatisfied demand between similar products or services. Leveraging a GP framework, we showed that our proposed Diffusion-aware Censored GP is able to better recover the (latent) true demand based on the observations of the satisfied demand and produce more accurate out-of-sample predictions. Our experimental results, based on real-world datasets for supermarket sales, bike-sharing demand, and EV charging demand, underscore the broad applicability and potential of our proposed framework. In future work, we will explore even larger-scale applications by leveraging sparse GPs.

Acknowledgements

We thank Fabio Hipólito for providing assistance with setting up the agent-based simulation tool, GAIA, used in the experiments of the proposed approach.

References

[Amemiya, 1984] Takeshi Amemiya. Tobit models: A survey. *Journal of econometrics*, 24(1-2):3–61, 1984.

[Basson *et al.*, 2023] Marno Basson, Tobias M Louw, and Theresa R Smith. Variational tobit gaussian process regression. *Statistics and Computing*, 33(3):64, 2023.

[Breen, 1996] Richard Breen. *Regression models: Censored, sample selected, or truncated data*. Sage, 1996.

[Chang *et al.*, 2020] Paul E Chang, William J Wilkinson, Mohammad Emtiyaz Khan, and Arno Solin. Fast variational learning in state-space gaussian process models. In *2020 ieee 30th international workshop on machine learning for signal processing (mlsp)*, pages 1–6. IEEE, 2020.

[Chib, 1992] Siddhartha Chib. Bayes inference in the tobit censored regression model. *Journal of Econometrics*, 51(1-2):79–99, 1992.

[DeMaris, 2004] Alfred DeMaris. *Regression with social data: Modeling continuous and limited response variables*. John Wiley & Sons, 2004.

[Domarchi and Cherchi, 2023] Cristian Domarchi and Elisabetta Cherchi. Electric vehicle forecasts: a review of models and methods including diffusion and substitution effects. *Transport reviews*, 43(6):1118–1143, 2023.

[Gammelli *et al.*, 2020] Daniele Gammelli, Inon Peled, Filipe Rodrigues, Dario Pacino, Haci A Kurtaran, and Francisco C Pereira. Estimating latent demand of shared mobility through censored gaussian processes. *Transportation Research Part C: Emerging Technologies*, 120:102775, 2020.

[Gammelli *et al.*, 2022] Daniele Gammelli, Kasper Pryds Rolsted, Dario Pacino, and Filipe Rodrigues. Generalized multi-output gaussian process censored regression. *Pattern Recognition*, 129:108751, 2022.

[Greene, 2003] William H Greene. Econometric analysis. *Pretence Hall*, 2003.

[Groot and Lucas, 2012] Perry Groot and Peter JF Lucas. Gaussian process regression with censored data using expectation propagation. In *Proceedings of the Sixth European Workshop on Probabilistic Graphical Models*, pages 159–164, 2012.

[Hamelijnck *et al.*, 2021] Oliver Hamelijnck, William Wilkinson, Niki Loppi, Arno Solin, and Theodoros Damoulas. Spatio-temporal variational gaussian processes. *Advances in Neural Information Processing Systems*, 34:23621–23633, 2021.

[Heien and Wesseils, 1990] Dale Heien and Cathy Roheim Wesseils. Demand systems estimation with microdata: a censored regression approach. *Journal of Business & Economic Statistics*, 8(3):365–371, 1990.

[Hipolito *et al.*, 2022] F Hipolito, CA Vandet, and J Rich. Charging, steady-state soc and energy storage distributions for ev fleets. *Applied Energy*, 317:119065, 2022.

[Hüttel *et al.*, 2022] Frederik Boe Hüttel, Inon Peled, Filipe Rodrigues, and Francisco C Pereira. Modeling censored mobility demand through censored quantile regression neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 23(11):21753–21765, 2022.

[Hüttel *et al.*, 2023] Frederik Boe Hüttel, Filipe Rodrigues, and Francisco Câmara Pereira. Mind the gap: Modelling difference between censored and uncensored electric vehicle charging demand. *Transportation Research Part C: Emerging Technologies*, 153:104189, 2023.

[Khan and Lin, 2017] Mohammad Khan and Wu Lin. Conjugate-computation variational inference: Converting variational inference in non-conjugate models to inferences in conjugate models. In *Artificial Intelligence and Statistics*, pages 878–887. PMLR, 2017.

[Li and Wang, 2003] Gang Li and Qi-Hua Wang. Empirical likelihood regression analysis for right censored data. *Statistica Sinica*, pages 51–68, 2003.

[Liu *et al.*, 2016] Junming Liu, Leilei Sun, Weiwei Chen, and Hui Xiong. Rebalancing bike sharing systems: A multi-source data smart optimization. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1005–1014, 2016.

[Liu *et al.*, 2017] Junming Liu, Leilei Sun, Qiao Li, Jingci Ming, Yanchi Liu, and Hui Xiong. Functional zone based hierarchical demand prediction for bike system expansion. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 957–966, 2017.

[McDonald and Moffitt, 1980] John F McDonald and Robert A Moffitt. The uses of tobit analysis. *The review of economics and statistics*, pages 318–321, 1980.

[Schaafsma and Brouwer, 2020] Marije Schaafsma and Roy Brouwer. Substitution effects in spatial discrete choice experiments. *Environmental and Resource Economics*, 75(2):323–349, 2020.

[Schmee and Hahn, 1979] Josef Schmee and Gerald J Hahn. A simple method for regression analysis with censored data. *Technometrics*, 21(4):417–432, 1979.

[Sup, 2024] Sup. Supermarket sales data: Sales data of vegetables in supermarket. https://www.kaggle.com/datasets/yapwh1208/supermarket-sales-data, 2024. Accessed: 2024-04-23.

[Tebbutt *et al.*, 2021] Will Tebbutt, Arno Solin, and Richard E Turner. Combining pseudo-point and state space approximations for sum-separable gaussian processes. In *Uncertainty in artificial intelligence*, pages 1607–1617. PMLR, 2021.

[Tobin, 1958] James Tobin. Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, pages 24–36, 1958.

[Unterluggauer *et al.*, 2023] Tim Unterluggauer, F Hipolito, Jeppe Rich, Mattia Marinelli, and Peter Bach Andersen. Impact of cost-based smart electric vehicle charging on urban low voltage power distribution networks. *Sustainable Energy, Grids and Networks*, 35:101085, 2023.

[Wan *et al.*, 2018] Mingchao Wan, Yihui Huang, Lei Zhao, Tianhu Deng, and Jan C Fransoo. Demand estimation under multi-store multi-product substitution in high density traditional retail. *European Journal of Operational Research*, 266(1):99–111, 2018.

[Xie *et al.*, 2023] Na Xie, Zhiheng Li, Zhidong Liu, and Shiqi Tan. A censored semi-bandit model for resource allocation in bike sharing systems. *Expert Systems with Applications*, 216:119447, 2023.