CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering

BACHELOR THESIS

Filip Kubiš

Application of Spatiotemporal Modeling Used in Robotics for Demand Forecast

Department of Cybernetics

Thesis supervisor: **Ing. Tomáš Vintr** May, 2020

Author statement for undergraduate thesis:

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, date............................. ....

BACHELOR‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Kubiš Filip** Personal ID number: **478203**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Cybernetics**

Study program: **Open Informatics**

Branch of study: **Computer and Information Science**

II. Bachelor’s thesis details

Bachelor’s thesis title in English:

Application of Spatiotemporal Modeling Used in Robotics for Demand Forecast

Bachelor’s thesis title in Czech:

Použití metod modelování časoprostoru v robotice pro predikci poptávky

Guidelines:

1) Research methods for spatiotemporal modeling used in robotics and intelligent transportation domain and research relevant datasets.
2) Research methods for demand forecasting used in urban mobility and their potential use in „before-demand“ fleet allocation.
3) Research relevant datasets of transport demand.
4) Select a set of suitable methods for spatio-temporal prediction of the transporation demand.
5) Select a set of performance and quality indicators of the transportation service.
6) Design an evaluation tool, capable of automated calculation of the aforementioned indicators for the seleted forecasting methods and datasets.
7) Using the tool, perform a comparison of the predictive ability of the aforementioned methods.
8) Discuss the impact of the prediction accurracy on the quality of transportation service.

Bibliography / sources:

[1] Donovan, Brian; Work, Dan (2016): New York City Taxi Trip Data (2010-2013). University of Illinois at Urbana-Champaign.
[2] Tsao, Matthew, et al. "Model predictive control of ride-sharing autonomous mobility on demand systems." Proc. IEEE ICRA 2019.
[3] Tsao, Matthew, Ramon Iglesias, and Marco Pavone. "Stochastic model predictive control for autonomous mobility on demand." In ITSC, 2018.
[4] A. Wallar, et al.: "Vehicle Rebalancing for Mobility-on-Demand Systems with Ride-Sharing," 2018 IEEE/RSJ IROS.
[5] Krajnik, Vintr et al.: "Warped hypertime representations for long-term autonomy of mobile robots." IEEE Robotics and Automation Letters (2019).
[6] Krajnik et al.: „Chronorobotics: Representing the structure of time for service robots.“ In IJCRAI 2019.

Name and workplace of bachelor’s thesis supervisor:

Ing. Tomáš Vintr, Artificial Intelligence Center, FEE

Name and workplace of second bachelor’s thesis supervisor or consultant:

Date of bachelor’s thesis assignment: **10.01.2020** Deadline for bachelor thesis submission: **25.05.2020**

Assignment valid until: **30.09.2021**

Ing. Tomáš Vintr Supervisor’s signature	doc. Ing. Tomáš Svoboda, Ph.D. Head of department’s signature	prof. Mgr. Petr Páta, Ph.D. Dean’s signature

III. Assignment receipt

The student acknowledges that the bachelor's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the bachelor's thesis, the author must state the names of consultants and include a list of references.

.	
Date of assignment receipt	Student's signature

Acknowledgements

First and foremost, I would like to thank Tomáš Vintr for being an excellent supervisor. I thank him for spending so much time with me discussing an endless amount of ideas for the thesis. He pushed me repeatedly to make the thesis better and provided me with plenty of valuable feedback. I had a great time working with him.

I would like to thank Jan Blaha for his inventive ideas during my and Tomáš' brainstorming sessions.

I would like to thank Tomáš Krajník for giving me the chance to work on this topic.

And lastly, I would like to thank Jiří Ulrich without whom my cooperation with Tomáš Vintr would never have happened.

Abstrakt

Cílem této práce je prozkoumání metod časoprostorového modelování, jež jsou v současné době používány pro predikci poptávky v oblasti přepravy, a přezkoumání způsobů testování kvality těchto metod. Práce ukazuje, že metody, které jsou v současnosti používány pro predikci poptávky, ve své podstatě obsahují vážné chyby. Proto jsou navrženy dvě nové metody testování prediktivních modelů, které tyto chyby nemají. Na problém predikce poptávky jsou v práci, kromě běžně používaných metod, aplikovány metody z oblasti chronorobotiky. Dále je navržena nová prediktivní metoda. Experimenty v této práci demonstrují problémy současných vyhodnocovacích metod. Další experimenty porovnávají jednotlivé modely. Výsledky ukazují, že současně používané metody evaluace nemotivují tvorbu co nejpřesnějších modelů. Navržená prediktivní metoda ve výsledcích experimentů dopadla vždy mezi dvěma nejlepšími.

Abstract

The goal of this work is to explore spatiotemporal modeling methods currently used in transportation demand forecasting and to examine the means of quality testing of these models. Furthermore, the thesis unveils fundamental issues in the way demand predicting methods are currently tested. The thesis proposes two new methods of demand prediction quality evaluation. These methods address the issues mentioned above. Chronorobotic modeling methods are applied to the domain, and a new predictive method is proposed. The experiments in this work demonstrate the shortcomings of the currently used evaluation methods. Moreover, the experiments evaluate the quality of various predictive models. The experiments reveal that the currently used evaluation methods do not incentivize the creation of the best possible models. Furthermore, the proposed method ranked as one of the two best methods in the experiments.

Contents

List of Figures

1 Introduction

With the recent rise of mobility-on-demand (MoD) services and the looming take-off of widespread autonomous mobility, much research has been directed towards the design of efficient algorithms to coordinate MoD fleets, minimizing service costs and passenger discomfort along the way.

MoD encompasses services that enable users to request a journey through an interface such as an app. The customer's request is processed and a solution is offered (such as a taxi trip). The payment for the service happens through the same unified interface. As a general concept, MoD is a departure from privately owned transportation (such as personal cars). Transportation is then approached as a service.

A common problem of these services (and of many others such as any vehicle-sharing) is a fleet imbalance. Fleet imbalance is a disadvantageous placement of the vehicles in places with little demand. This problem has been addressed by multiple fleet rebalancing approaches[1, 2] that seek to distribute an automobile or a bike[3] fleet based on a criterion. In many recent works, the criterion used for rebalancing is future demand prediction[1, 4–6]. In other words, these works focus mainly on finding ways in which to utilize predictions for fleet routing, not necessarily on demand prediction itself. This work does not aim to design rebalancing algorithms, but rather to focus on modeling demand.

The problem of demand prediction was covered by a multitude of papers in and of itself[7–18]. It was also briefly addressed in some of the works on rebalancing[1, 4–6]. Many different approaches to demand forecasting have been described in recent papers, including frequentist approaches[4], deep learning[7], various regression methods[18], and time-series predictors. Nevertheless, all of the works have a commonality: they all discretize both the time and area into a spatiotemporal grid. The predictive methods are then applied to the time-series corresponding to demand on a segment of the map, losing spatial context. While some address this issue through the use of convolutions or clustering[5, 17], there is still information loss incurred by the discretization. A method able to utilize observed phenomena in space and time without discretization would be better fitting for spatiotemporal modeling.

The works focusing solely on demand prediction almost exclusively create short-horizon predictions models[7–9, 11–13, 15–18]. That is, they only make predictions for one time segment ahead of the training data. While this approach surely provides a tactical advantage to a ride-hailing service provider, it does not provide them with a long-term strategic edge, such as being able to anticipate a spike in demand in a few hours or even days.

Recently introduced methods of spatiotemporal modeling in the field of robotics (specifically chronorobotics) can model periodic phenomena[19]. These methods are not yet applied to the domain of transportation. However, these methods have a few properties, which may prove to be advantageous, namely: continuity and inherent periodicity detection[19]. The continuity characteristic enables the creation of models that do not reduce the space-time into multiple discrete one-dimensional time-series as is currently commonly practiced.

Periodicity detection should, in turn, help create models that are firstly able to make accurate long-term predictions and, secondly, do not require constant recalibration with new data for each prediction. The issue at hand is that these methods were developed for sparse data and that they may not be able to utilize all data in a large dataset. The other problem is the novelty of these methods, meaning that some are not yet perfectly optimized and refined. While these methods may not replace the state-of-the-art methods for short-term predictions, they might prove to be very useful in creating long-term models to be used hand in hand with the current methods.

An important question that this thesis tries to address is: How do we assess the quality of the models? The works mentioned above on rebalancing[1, 4–6] usually compared the in-simulation performance of methods that used predictions for rebalancing to those that did not. The apparent problem of simulations is that a simulation is a black box in many ways (especially to an outsider). Many factors may influence the results, especially the in-between step of using the prediction to manage a fleet (which in each simulation may be different). These extra steps create a distance between the predictive model and the evaluation criterion (such as the average time to serve a passenger). Other works use error metrics such as mean absolute error (MAE) and root-mean-square error (RMSE) to evaluate predictions in a discretized space-time. This adjustment turns the demand prediction problem into a time-series prediction problem. While the results of such evaluation are rigorous and relatively easily interpreted, they do not reflect spatial relations. Furthermore, they were used to asses models built on compressed training data using compressed test data.

While very handy and sensible when comparing time-series prediction quality, these metrics are unable to compare a discrete model to a continuous one thoroughly. Such used metrics are thus incapable of rewarding a nuanced continuous model (demonstrated in Sections 3.1 and 4.7.1). An ideal criterion of quality would preserve spatial relations, use testing data without compression, not penalize models for finer resolution (if discretization is used), and would be easily interpreted. The thesis briefly looks into the theory of continuous forecast scoring[20] and shows the difficulties of applying conventional scoring rules to the evaluation of demand forecasting. The thesis suggests a way of using error metrics to satisfy the properties as mentioned above. Furthermore, a simple test measuring the ability of predictive models to correctly position vehicle fleet is proposed. The thesis additionally proposes a new method of spatiotemporal modeling based on gaussian mixture models.

Experiments that test the prediction quality of various models using different evaluation methods are conducted. The initial hypotheses are then examined against the results. The Future work section (6) considers possible future research in the area in light of the theses findings.

2 Related work

This section is an introduction to the current state of research in the following areas:

1.) Demand modeling in transportation.

2.) Chronorobotic spatiotemporal modeling.

3.) Prediction quality evaluation

The first part is a brief survey of the recently used methods of modeling transportation demand, especially in the area of MoD. It starts with a high-level overview of approaches that seem to be common to all the recent contributions and concludes with all the details in which they differ. The second part introduces the specifics of the particular chronorobotic modeling methods and their current uses. It also demonstrates the link between its current domain and demand modeling. In the third part, the thesis focuses on the evaluation metrics recently used to asses predictive models in transportation. Lastly, it introduces the theory behind scoring rules (functions that return a numerical evaluation of a forecast based on observation) and the desirable qualities of proper scoring rules.

2.1 Demand modeling in transportation

2.1.1 Approach to data

The approach overwhelmingly used in recent works has been to discretize the data both space-wise and time-wise into spatiotemporal segments and make predictions for these segments[1, 4, 7–13, 15–18, 21]. The discretization of time into equally long stretches remained consistent across the works, the length of these stretches varied between 15 minutes and one hour [4, 8]. The way of splitting the area was much more diverse. The simplest methods were different variations of division into a rectangular grid, the most interesting one being geohashing[8, 22], which assigns a hash to a pair of latitude, longitude. Geohash, in effect, splits the area into a rectangular grid with resolution affected by parameter choice. Another approach was the selection of n points in the road network and assigning each demand event to one of those points. A more sophisticated way of choosing the stations[1] attempted to cover the area with as little stations as possible sufficiently. A sufficient cover meant that each point of the road network is reachable from one of the stations in less than t seconds. An interesting approach[15] was choosing only certain venues (such as sports arenas or music clubs) for which predictions were made using not only past demand data but also textual information about events at these venues. Rodrigues et al. were not the only ones to use contextual data in their forecasts. A common approach in the most recent work is the utilization of information about current weather [16–18, 23].

2.1.2 Modeling methods

The nature of the models is, to a large, extent determined by the approach to the data. In the case of demand modeling in transportation, the models used in recent work are mostly predictors of future values of time-series (of multiple time-series, where each corresponds to an area segment). The meaning of the values being the volume of demand in the area during the specified time-window. The recent works almost universally cover predictions for one time-window following training data [7–9, 11–13, 15–18]. Nevertheless, there were some fundamental differences in these models:

1.) Utilization of outside parameters such as current weather to fine-tune predictions. This use of parameters was mostly the case of neural-network (NN) models.

2.) Use of spatial relationships. Some time-series[21] and convolutional NN[7, 13, 16, 17] models used past data from neighboring segments for their next prediction.

3.) Destination prediction. Meaning that the model made predictions for $Origin \times Destination$ pairs[1, 4, 16].

4.) Model updating on new observations. Burke[24] distinguishes two types of learning. Online, which uses each new observation to update itself, and offline, a fixed model that does not update itself. The main focus of recent work is online models, with the exception being historical offline models [1, 4].

2.1.3 Use of models

The last significant difference is the way that the models were used. Works focusing solely on demand prediction[7–18] used the models to predict demand volumes. The works which used the predictive models in simulations [1, 4, 6] aimed to create probabilistic models that could be used to sample future prediction from for the simulation.

2.1.4 List of models

Following is an outline of the recently used methods.

1.) Neural Networks(NN). In the works from the last two years, this has been an overwhelming approach. NN usually take into account spatial context as well as external events such as weather.

2.) Autoregressive time-series predictors such as ARIMA (often used as a baseline [9, 12]) and STARIMA, which has been used for traffic prediction [21].

3.) Regressive models, such as support vector regression (SVR), are often used as baseline methods. Interestingly, in the experiments of Jiang et al.[18] SVR has achieved better prediction accuracy than baseline NN.

4.) Historical models use the average of historical demand to make predictions about the future demand volume or to create a probabilistic model. These methods model a fixed period (such as a day or a week). The predictions are then made e.g., for the third hour of Thursday.

5.) Stochastic models. These models create a distribution for each pair $(station, time)$, which is used for demand sampling, Tsao et al. [6] use Poisson distribution to determine demand volume at each station at a given time.

2.2 Spatiotemporal modeling in chronorobotics

Chronorobotics seeks to improve robots' functionality by creating robust models of time-dependent phenomena in the robot's environment[19]. They model the changing environment as functions of time. A successful approach enabling long-term autonomy[19, 25, 26] of the robots was modeling the environmental changes by periodic functions, as a lot of these changes are periodical. These methods can detect significant periodicities in the modelled phenomena, enabling them to make long term predictions. They have been successful in anomaly detection eighteen weeks after training on two weeks of data[25]. These methods have already been applied to the domain of spatiotemporal modeling by predicting human presence[26] to avoid crowds or to find a person as fast as possible[27]. So the extension of chronorobotic modeling approaches to the domain of taxi demand prediction is quite natural. Frequency Map Enhancement (FreMEn)[28] and Warped Hypertime (WHyTe)[29] are the modeling tools that will be used in this work.

2.2.1 Properties of the above methods

Following is a list of properties of these two methods, which may prove to be desirable for spatiotemporal demand modeling and especially for long term predictions.

1.) Continuity These methods can be used to create nuanced continuous demand density models (as opposed to the currently used discrete time-series predicting models).

2.) Periodicity modeling FreMEn and WHyTe both detect and model periodicity, enabling realistic long term models.

3.) Parsimony Vintr et al. [26] show that a WHyTe model (occupying 2.4 KiB) had a better prediction accuracy of human presence than any grid-based model (each of those was at least an order of magnitude larger than the WHyTe model).

2.3 Prediction quality evaluation

2.3.1 Evaluation in recent works on the topic

Let us first discuss the methods used in recent works on the topic. The first commonly used method was an evaluation in simulation[1, 4, 6]. However, the works above did not use simulation to compare different predictive models. They used the simulation to demonstrate the effect of having a predictive model. The second approach was the use of widely used error metrics such as RMSE and MAE. The use of this approach was quite natural due to the reduction of the problem to the measuring of the quality of time-series predictions. This evaluation approach is recommended for the evaluation of time-series prediction [30]. However, there are many questions at hand: Do we truly evaluate the quality of the spatiotemporal model when we evaluate the quality of some arbitrary time-series (the fineness of the grid or the number of stations can be manipulated, and it does not seem to be chosen based on a rigorous method, as evidenced by substantial differences in recent works)? How valid is it as a test of spatiotemporal prediction quality?

Wang et al. [31] show the shortcomings of mean square error (MSE) in the domain of image processing. They show various striking examples of how an original image can be manipulated in a way that results in relatively small MSE but leaves the image nearly unrecognizable. Moreover, they demonstrate that two images that are nearly indistinguishable to the naked eye can have relatively high MSE. They suggest that MSE should not be discarded, but used instead alongside more sophisticated methods, such as structural similarity in the domain of image processing. Could it be that it is also the case in the domain of spatiotemporal demand modeling? Let us explore some other methods of prediction evaluation.

2.3.2 Probabilistic predictions evaluation

There is an extensive theory behind the testing of probabilistic prediction quality laid out by Gneiting and Raftery[20]. They describe scoring rules as functions that measure the quality of fit of a Probabilistic model to a sample. Their work also shows how to compute the expected value of a scoring rule for pair of probability measures P (the probabilistic forecast) and Q (the probabilistic measure from which a sample would be drawn). A strictly proper scoring rule is then a function that for constant Q maximizes its value if and only if $P = Q$. In other words, the expected value of the score is the highest for a correct prediction.

Applicability of strictly proper scoring rules to demand prediction If we can reduce the problem of demand prediction into the problems of demand volume and demand distribution prediction (which can be represented as a continuous probability density), then we could apply the available tools to evaluate the demand distribution prediction. However, there are many problems when trying to use these as metrics to evaluate different demand predicting models. Some tools are only applicable to discrete distributions, some only to univariate distributions[20, 32]. Others require complicated arithmetics to be applied effectively to any type of distribution[32], which might not even be possible for some of the models. The exception to those is the logarithmic score, which is the logarithmic likelihood of a model. The logarithmic score is tough to interpret when applied across different distributions. Moreover, it has the undesirable property of predicting minus infinity every time an event to which the model assigned zero probability happens. A model of demand prediction can represent demand quite well even if it unable to predict some events. This property is also the reason it is highly inconvenient to use any likelihood-based criterion such as Akaike information criterion[33] as a quality metric in this case.

2.3.3 Final remarks

Due to the limitations of modeling demand predictors as a set of discrete time-series predictors and the ensuing evaluation using error metrics, this work tries to look for other testing methods. Also, the possibility to apply scoring rules used in the evaluation of probabilistic forecasts to the testing of various demand predicting models is minimal. These were the main reasons why one of the main goals of this theses became to suggest new ways (see Section 3.2) of testing demand prediction quality to be used alongside established methods.

3 Proposed methods

As mentioned in previous section, error metrics such as MAE and RMSE are the overwhelmingly used tests of prediction quality. Section 2.3.1 described how this is the recommended approach when the spatiotemporal prediction problem is reduced to time-series forecasting. This section evaluates the popular approach towards demand prediction evaluation critically. Firstly, it is demonstrated how discretization causes loss of valuable information. Afterward, two new methods of model evaluation are proposed. These methods address the problems from which evaluation-on-grid suffers. See Section 4.3.1 for a detailed description of evaluation-on-grid (the overwhelmingly used evaluation method for demand prediction in recent works). The proposed methods should be suitable for a fair evaluation of any spatiotemporal demand predicting model regardless of it being continuous or discrete (and irrespective of resolution in case of discrete models) Lastly, a new spatiotemporal modeling method based on gaussian mixture models (GMM) is proposed.

3.1 Discretization issues

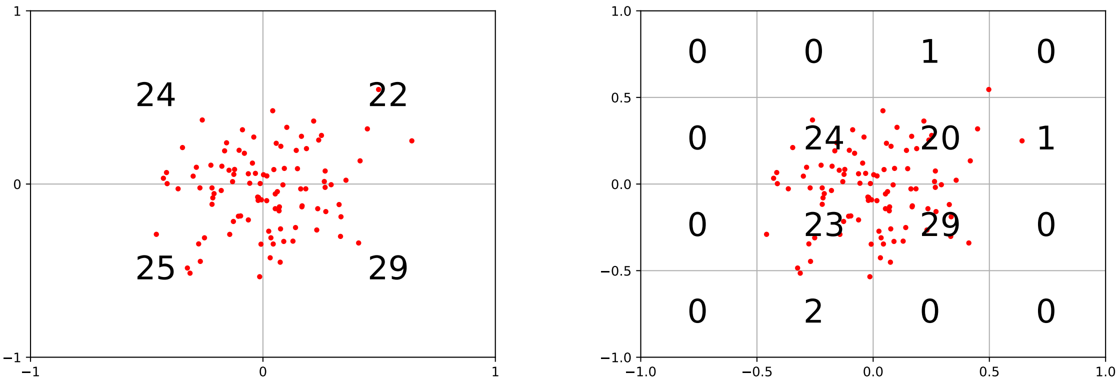

Figure 1: Visualization of the issues of discretization. Cluster of observations on the border of cells.

Figures 1 and 2 show the information loss caused by discretization on simple examples. The figures also show how this may, to a certain degree, be mitigated by using a fine enough resolution. However, with finer resolutions, the data of each time-series will be more volatile. The volatility is caused by the irregularities of demand at different parts of the cell balancing each other less with decreasing cell size. In other words, to create a discrete model as accurate as possible, it would be advantageous just to predict the total volume of demand (1x1 spatial grid). On the other hand, if one wanted a discrete model that provides as much information about the demand, this model would have as fine a resolution as possible. To illustrate this, let us compare two recent works. In the first one,

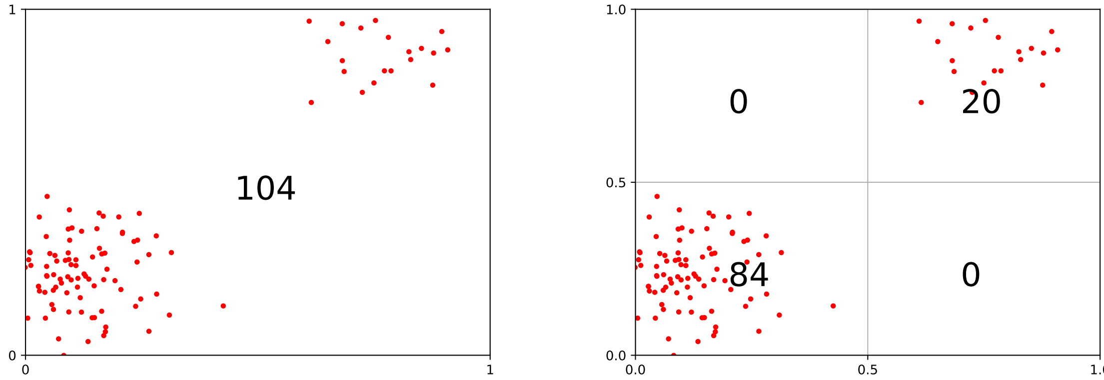

Figure 2: Another discretization issue. Multiple clusters of observations in one cell.

Yao et al.[7] focus solely on creating accurate demand models, evaluated using error metrics with no intention shown in their work of using their model for fleet positioning. They use a rather coarse discretization into regions of $0.7\,\text{km} \times 0.7\,\text{km}$. While very sophisticated and accurate, their model has no chance of fully capturing small but dense areas of high demand due to the nature of data preprocessing. Alonso-Mora et al.[4] contrarily create demand models to use in simulation; their motivation is to create models as informative as possible to improve their routing algorithm in simulation. Their model, albeit a simple historical model, is built on a relatively fine discretization where the distance between two cell centers is at most 150 meters. This twenty-fold difference in region size is a significant example of this ambiguity of intentions brought about by evaluating discrete models on a grid. Should a criterion that does not punish a model for inaccuracies showed in figures 1 and 2 be used as a standalone performance indicator? This incongruity is a serious reason why it is essential to look into new ways of evaluating demand models.

3.2 Proposed evaluation methods

In this section, two alternative methods of evaluation are presented. The first one is Random area (RA) evaluation, which evaluates models' ability to predict demand in different areas. The second one, called the Fleet placement test (FPT), is based on fleet sampling from the predictive model and evaluating how well it serves demand in a short time-window. The section about FPT (3.2.2) describes how it differs from simulation and its simple results interpretation.

3.2.1 Random area evaluation

Random area evaluation method compares the ability of different models to predict demand on randomly generated areas. For discrete models, the prediction over each area is

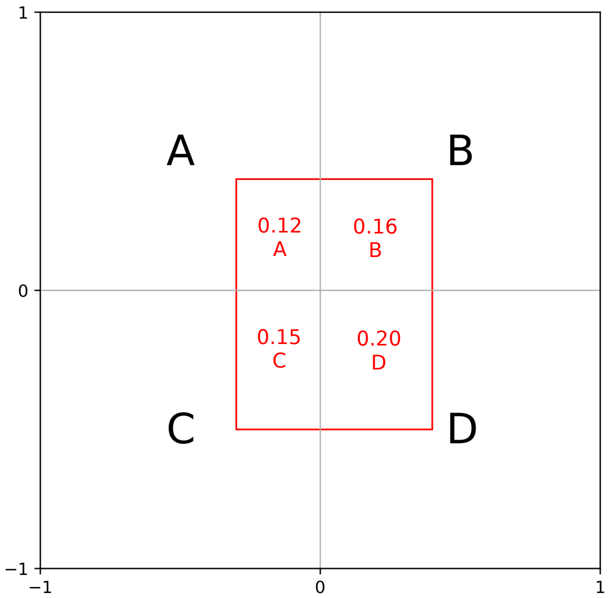

Figure 3: Visualization of RA computation for discrete models.

a linear combination of model predictions p_M for the whole grid, where prediction for each cell is multiplied by its proportion in the random area (see Figure 3). Each discrete model (M) comprises of time-series predictors for m areas $a_i, i = 1 \ldots m$, that together cover the whole area A for which predictions are made. So demand prediction of the discrete model for a random area (A_r) at time t is calculated as follows:

$$p_M\left(A_r, t\right) = \sum_{i=1}^{m} \frac{V\left(a_i \cap A_r\right)}{V\left(a_i\right)} \cdot p_M\left(a_i, t\right), \tag{1}$$

where V is the 2-dimensional volume of the area.

The prediction for a random area is computed via integration of demand density D_t at time t over the random area in case of a spatially continuos model:

$$V_t(A) = \int_{x \in A_r} D_t(x) dx dt \tag{2}$$

The predictions are then compared to the true values using RMSE to receive a score (see Section 4.3.2 for RMSE definition).

This method expands evaluation-on-grid by asking models to predict volumes in areas that do not correspond to the grid cells. It addresses many of the problems of evaluation-on-grid, such as disregard for spatial relations and discouraging finer model creation, while remaining easy to interpret. The compression of test data is also much less severe, as any observation from test data may belong to any area that contains it in the case of this experiment.

On the other hand, it is necessary to implement a random area predictor for each model (or at least for each kind of space discretization) to use this method. Also, its computation could be relatively complex, especially for discrete models with irregular cells.

3.2.2 Fleet placement test

This method measures the efficiency of the assignment of the fleet to demand in a short time-window. The steps are the following:

- Demand in a specific short time-window is considered, and a fleet is sampled from model predictions (for the same time-window).
- The cars in the fleet are assigned to serve one demand each.
- The result is obtained by computing distance traveled per customer served.

This method is not dynamic in time and only performs one optimal assignment. This temporal stasis is opposed to a simulation. So the method is rigorous, and the results are not influenced by other factors such as assignment strategy the way they are in a simulation. The strengths of this method are:

1.) Straightforward result interpretation. It is the average distance a taxi has to travel to serve one customer.

2.) Link to the use of the predictions. As the predictions should be used for before-demand fleet positioning and rebalancing, it is natural that the quality testing method should evaluate the ability to place a fleet well.

3.) It does not reward dishonesty (or laziness) in model creation as opposed to error metrics, which may incentivize coarse discretization and thus less informative models (see 3.1 and 4.8.1).

FPT with specified demand volume To perform this test for a time-window Δt, we denote $D = \{\mathbf{d_i}\}_{i=1}^{n}$ as the demand in t, $S = \{\mathbf{s_i}\}_{i=1}^{n}$ the sample of the model at Δt. Both S and D are sets of vectors from $\mathbb{R}^2$. Ro draw samples from any model, simple rejection sampling, as described by Neal[34], was used.

Rejection sampler randomly generates points in $\mathbb{R}^{n+1}$ for a distribution in $\mathbb{R}^n$; if the generated point is below the graph of the distribution, it is accepted into the sample (illustrated in Figure 4). The points are generated until enough samples are produced.

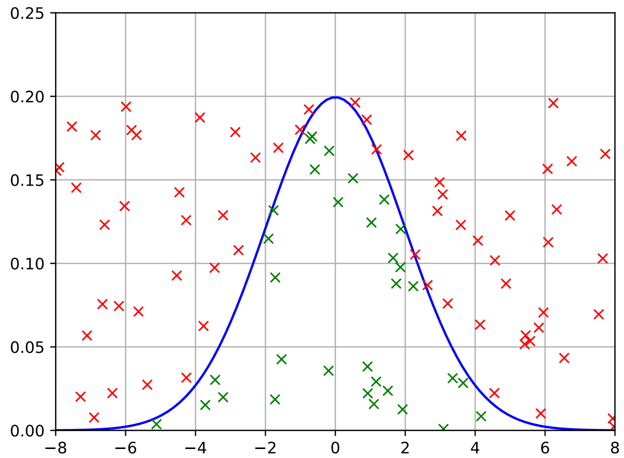

Figure 4: Visualization of rejection sampling using a univariate normal distribution. The green points are accepted and the red ones rejected.

An assignment is any bijective function $A_f : S \rightarrow D$. Let us define the cost of the assignment as:

$$c(A_f) = \sum_{i=1}^{n} dist(\mathbf{s_i}, A_f(\mathbf{s_i})), \tag{3}$$

where *dist* is a Manhattan distance function. This distance serves as a good estimate of the length of the route, especially in the chosen scenario (see Section 4.5). In future work, the distance function might be improved through using a more sophisticated model of a city. An assignment is optimal if its cost is minimal out of all possible assignments. To use the optimal assignment linear_sum_assignment function from scipy[35] optimize library was utilized. This implementation uses the Hungarian algorithm[36] to find the optimum.

It is evident that this test only measures the quality of demand density prediction. This method can be supplemented by other methods measuring the quality of total volume prediction, such as any error metric mentioned thus far. Another way would be to contain it in the test.

FPT-v - FPT with demand volume predictions Each model has to also predict volume in this testing scenario. If the predicted volume is larger than the demand, the superfluous taxis have to return to a station. Contrarily, if the predicted volume is smaller than the demand, the extra taxis have to be deployed from the station. This means that the size of the drawn sample would then correspond to the predicted demand volume. Any difference in the sizes of S and D sets would be equalized by supplementing the smaller set with multiple instances of vector $\mathbf{s_0}$. This vector represents the station from which extra taxis must be deployed (or to which they return). (Note: this turns the set into a multiset).

$$D = \{\mathbf{d_i}\}_{i=1}^{n}, \qquad S = \{\mathbf{s_i}\}_{i=1}^{m} \tag{4}$$

If $n < m$:

$$D' = D \cup \{\mathbf{x_i}\}_{i=1}^{m-n}, \quad \forall i: \ \mathbf{x_i} = \mathbf{s_0}, \tag{5}$$

and if $n > m$:

$$S' = S \cup \{\mathbf{x_i}\}_{i=1}^{n-m}, \quad \forall i: \ \mathbf{x_i} = \mathbf{s_0} \tag{6}$$

The cost of the optimal assignment is then computed the same way as was described in the previous paragraph. The output value of both FPT and FPT-v is the distance traveled per customer served, expressed in meters.

3.3 Proposed spatiotemporal modeling method

This work suggests a new approach towards spatiotemporal demand modeling using gaussian mixture models (GMM) called: Time-window GMM (TW GMM). This approach, similar to historical models (see 4.4.1), creates a spatial model for each time-window in a certain period. This spatial model (M_t) for time-window t is a pair of total volume prediction and a gaussian mixture probabilistic model (v_t, G_t).

3.3.1 Model fitting

Time-window GMM model with a period T, time-window length, and n components fits a GMM for each time-window. The total volumes are fitted analogically to the historical models (see 4.4.1). The GMM for a time-window is trained on the training dataset data assigned to that time-window. It is fitted using the expectation-maximization (EM) algorithm. This work uses scikit-learn [37] implementation of EM GMM.

3.3.2 Using the model for predictions

Let us define the demand density ($D_t(\mathbf{x})$) at a point $\mathbf{x}$ in time-window t as the value of probability density function $G_t(\mathbf{x})$ multiplied by the volume of demand in time-window t (v_t):

$$D_t(\mathbf{x}) = v_t \cdot G_t(\mathbf{x}) \tag{7}$$

The total demand $V_t(A)$ at an area A in time-window t is computed via integration of demand density over A:

$$V_t(A) = \int_{\mathbf{x} \in A} D_t(\mathbf{x}) d\mathbf{x} = v_t \cdot \int_{\mathbf{x} \in A} G_t(\mathbf{x}) d\mathbf{x} \tag{8}$$

4 Experiments

This section offers a detailed description of the used datasets, evaluation tool, used error metrics, models, and the setup of different experiments.

4.1 Datasets

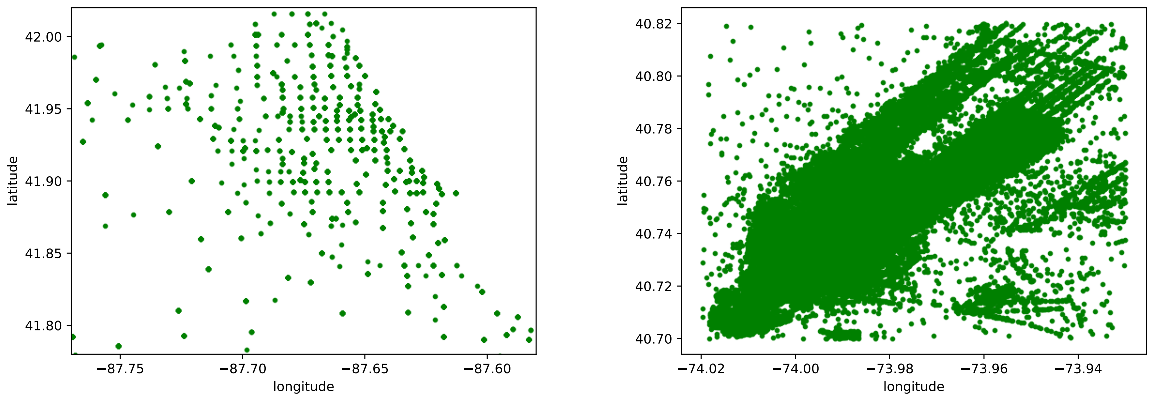

Figure 5: Visualization of the raw datasets from the cities of Chicago (left) and New York (right) shows that the Chicago data has already been somehow discretized.

The data used in the experiments throughout this work comes from New York taxi trip data[38] collected between 2010 and 2013. This dataset is very suitable for the needs of this work as it offers very dense data (over 10 million measurements per month). This density of data enables experimentation with different discretization resolutions. The raw New York data do not seem to be crudely masked or compressed, unlike the Chicago dataset[39]. This phenomenon is best seen in Figure 5, which is a visualization of one month of data from both cities. Other considered datasets such as uber trips from New York[40] were relatively small. Some datasets used in recent work were not openly available such as the commonly used[7, 17, 18] data from Chinese company Didi Chuxing.

4.2 Automated Evaluation tool

A unified evaluation tool (Evaluator) capable of an automatic model evaluation was implemented for the thesis. The tool is implemented in Python 3 programming language (as are all the predictive models). An interface defines how the tool communicates with the predictive models. All the predictive models inherit from the Predictor class, which ensures that the models follow the defined interface. The Evaluator class is initiated with a dictionary of predictive models. Evaluator contains the following methods:

- **train**: takes training data and times and trains all the predictive models.
- **different test methods**: take testing data and times and evaluate all the models.
- **visualization methods**: show the results of the models in the performed tests.
- **save and load**: save or load the models.

4.3 Evaluation

4.3.1 Evaluation-on-grid

The evaluation method used throughout the recently published work on demand prediction is the evaluation-on-grid. Evaluation-on-grid means that the models predict future values of demand in each cell of the grid (for any amount of time-windows). These predictions are then compared to the actual values from the test dataset using error metrics. Following is a short description of error metrics used in this work.

4.3.2 Root mean square error (RMSE)

RMSE[30] is defined for a sequence of predictions $P = (p_i)_{i=1}^n$ and a sequence of true values $R = (r_i)_{i=1}^n$ as:

$$RMSE(P, R) = \sqrt{\frac{\sum_{i=1}^n (p_i - r_i)^2}{n}} \tag{9}$$

4.3.3 Mean absolute error (MAE)

MAE[30] is defined for a sequence of predictions $P = (p_i)_{i=1}^n$ and a sequence of true values $R = (r_i)_{i=1}^n$ as:

$$MAE(P, R) = \frac{\sum_{i=1}^n |p_i - r_i|}{n} \tag{10}$$

4.4 Models

A wide array of model types were included (time-series tool, historical models, regressor, spatio-temporal continuous model). Following is a table of the used models.

Model	Description
Historical	Weekly and Daily and Mean models
Prophet [41]	Time-series forecasting tool by Facebook
Zeros	Predicts zeros only
SVR	Support vector regressor
FreMEn	see Section 4.4.4
WHyTe and WHyTeS	see Section 4.4.5
Time-window GMM	see Section 3.3

4.4.1 Historical models

Historical models base their prediction on the mean value of the predicted phenomena at a similar time within a specified period in the past. Each historical model is defined by its period (T) and time-window length (l_t). The period of each model is split into time-windows of length l_t. Each measurement from the training data can be assigned to one of these time-windows. Because each time-window may appear multiple times in the training data, the predicted value for a time-window is the amount of data assigned to it divided by the number of appearances of that time-window in the training data.

4.4.2 Prophet

Prophet is a time-series predicting tool by Facebook [41], which is fitted by providing an array of equally spaced times and corresponding values for these times. For each cell in the spatial grid, a Prophet model is fitted.

4.4.3 Support vector regression

Same as when using Prophet, an SVR model was fitted for each cell of the grid (the scikit-learn [37] implementation of SVR was used).

4.4.4 FreMEn

The FreMEn model implemented in this work is based on the algorithm described by Krajnik et al. [28]. A FreMEn model was fitted for each cell in the grid. Each FreMEn model is defined by its mean value, around which it oscillates (α_0) and by n periodicities, which are chosen from a set of candidate periodicities.

4.4.5 WHyTe and WHyTeS

The implementation of WHyTe models in this work is based on the algorithm description by Vintr et al. [25]. The WHyTe model is a gaussian mixture model with n components fitted on the points from the training dataset, whose time coordinate is projected onto multidimensional vector space. A WHyTe model was fitted for each cell.

WHyTeS [26] is a version of the WHyTe algorithm over spatiotemporal space, where the position and of each training datum is projected onto multidimensional vector space. WHyTeS is continuous in space and time. Therefore, one WHyTeS model was trained to create the model.

4.5 General experiment setup

The training dataset consisted of data from June 2010, and the testing dataset is made up of data from July 2010. The length of each time segment was set to 10 minutes in all the experiments for evaluation-on-grid or random areas. 1x1 grid (1 cell), 10x10 grid (100 cells), and 200x200 grid (40000 cells) were used in the experiments. The sizes of the random areas used in the random area test ranged from 1/40000 to 1/25 of the total area. The maximum amount of demands in one iteration of the FPT is 200 due to the limitations caused by computing optimal assignments. The resulting value is the total distance covered during all the iterations divided by the total customers served. The station in the fleet placement test with volumes was placed into the centroid of the training dataset.

4.6 Experiment 1

It was hypothesized in Section 3.1 that finer discretization yields more informative models. Another hypothesis was that it becomes more challenging to evaluate these models conventionally on grids with finer resolutions. This experiment explores how different testing methods evaluate discrete models at different resolutions. The methods used are RMSE and MAE evaluation-on-grid and random area test. These methods are used to compare three historical models (with the Zeros model as baseline) at three different resolutions. A graph showing how the models rank at each resolution is included for all the evaluation methods.

4.6.1 Experiment 1 - Results

Figures 6 and 7 both show that using evaluation-on-grid becomes unreliable for very fine resolutions, as the models' order changes at fine resolutions for each method. What is worse, this change in order is different for each of the two metrics. While the Zeros model at 40 thousand cells is the best model, according to MAE, it is the worst, according to RMSE.

So this unreliability naturally turns researchers off from using very fine models when evaluating on a grid. However, Figure 8 shows that models get increasingly accurate in demand predictions for random areas with finer resolution. That could mean that the

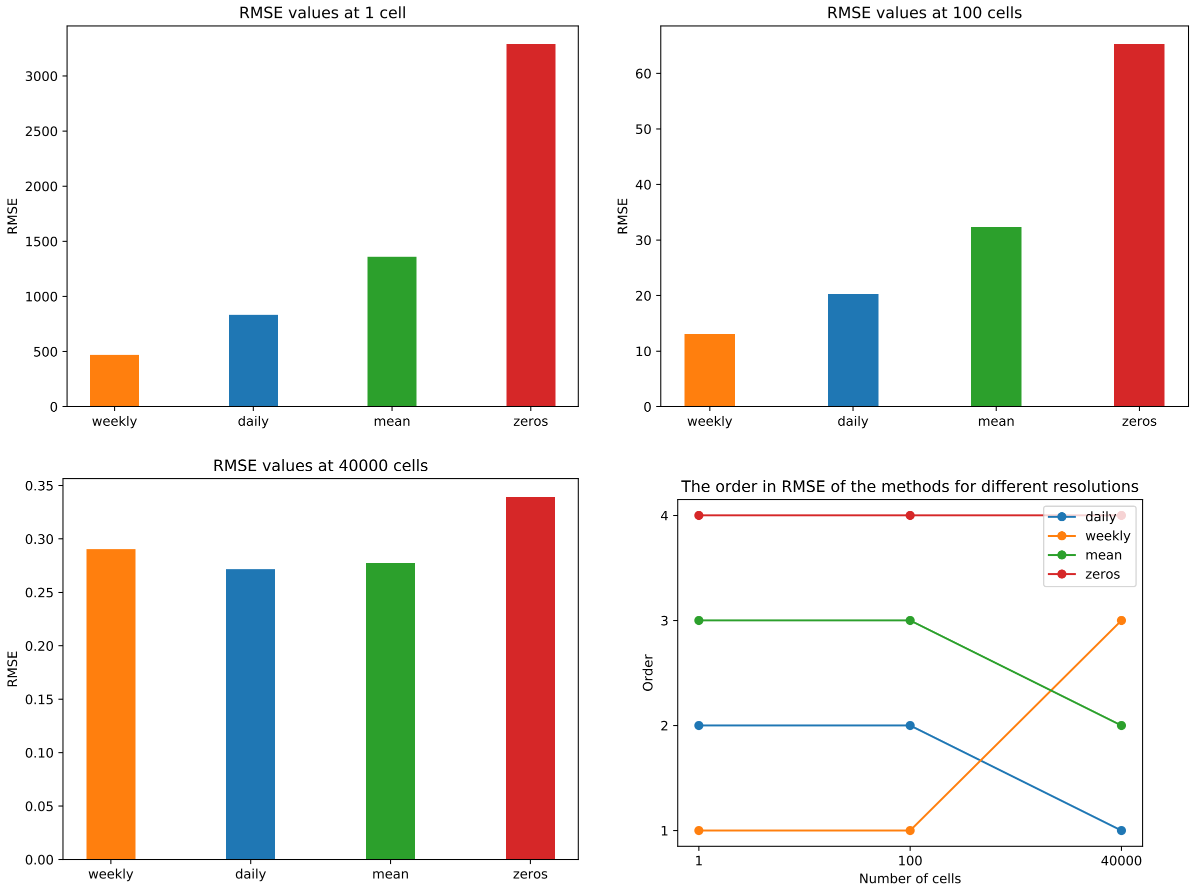

Figure 6: RMSE values of different models at different resolutions and the order (lower is better) of the models at different resolutions.

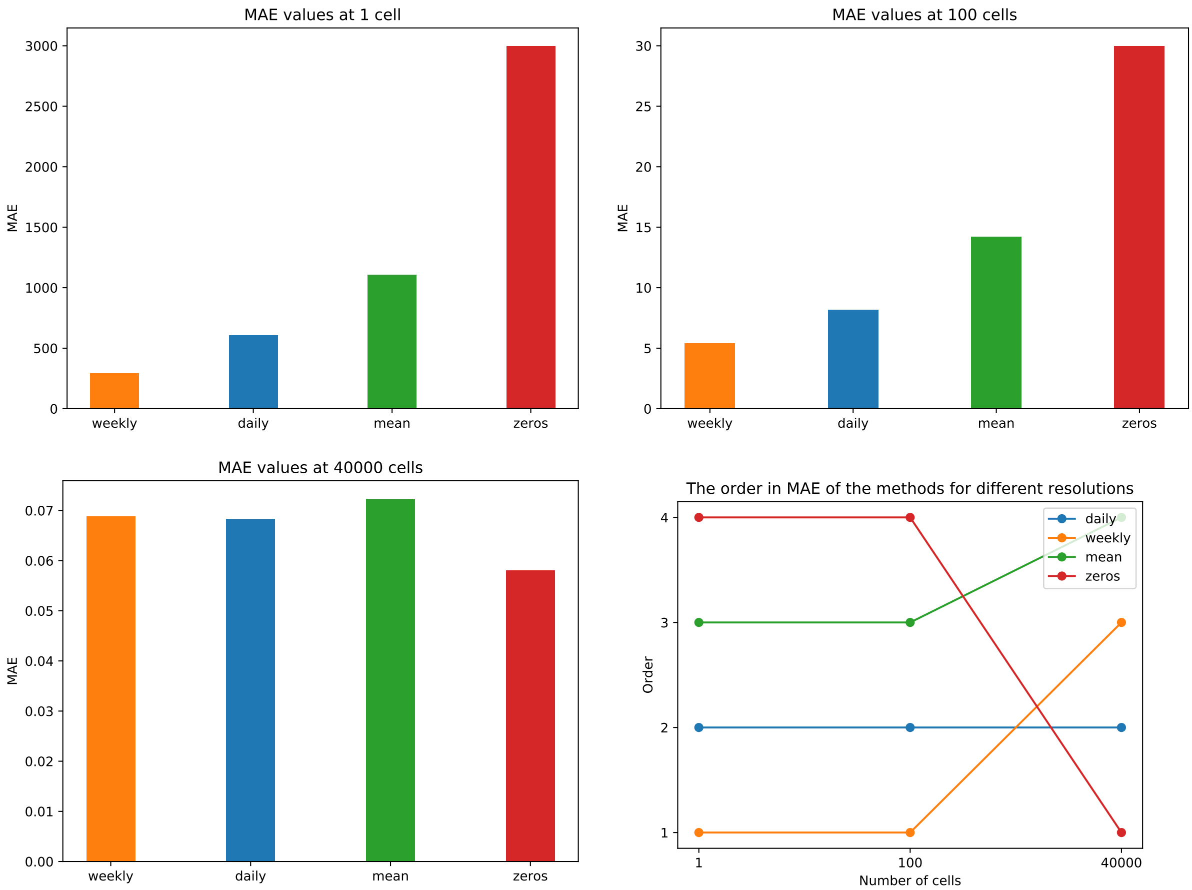

Figure 7: MAE values of different models at different resolutions and the order (lower is better) of the models at different resolutions.

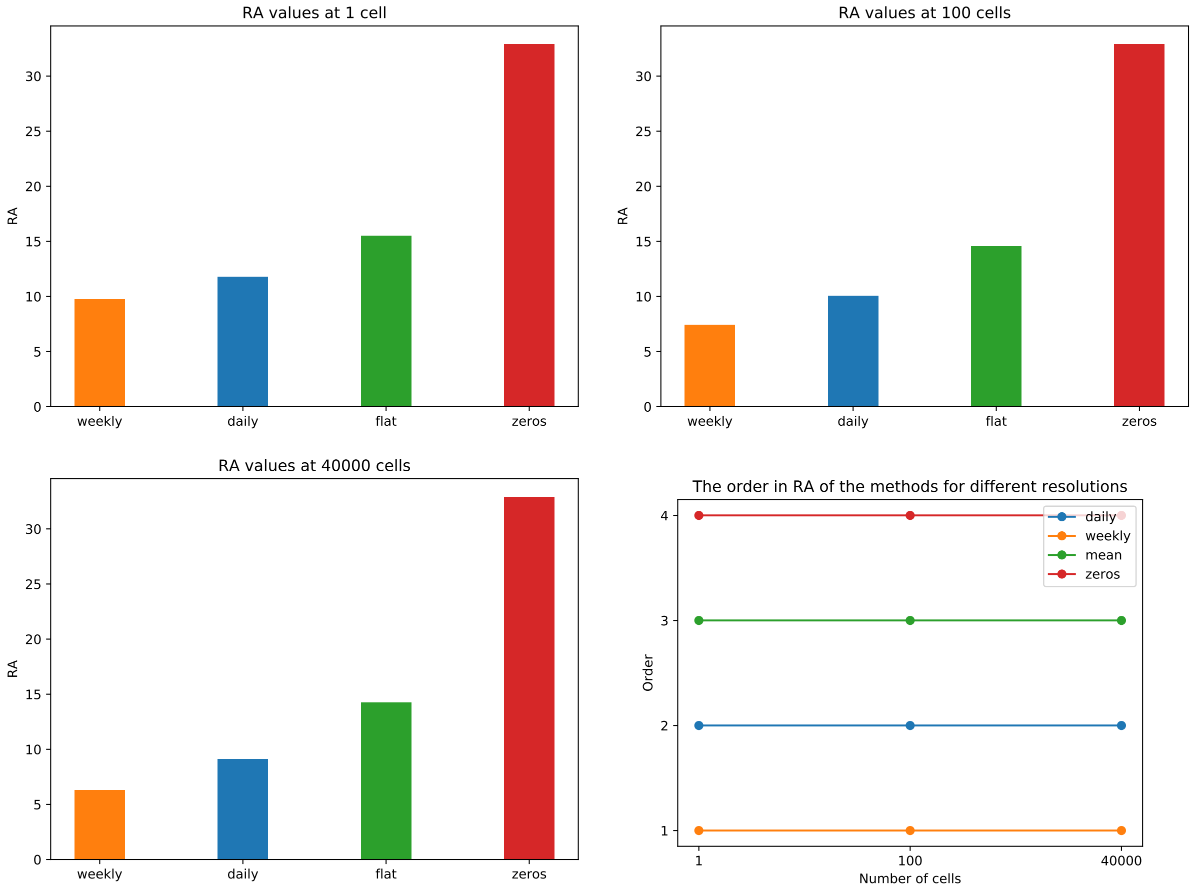

Figure 8: RA values of different models at different resolutions and the order (lower is better) of the models at different resolutions.

current way of testing does not encourage the creation of models as informative as possible. Furthermore, the RA is shown to be robust to the inaccuracy at fine resolutions in contrast to evaluation-on-grid (see Figure 8).

4.7 Experiment 2

This experiment compares discrete models trained on a 10x10 grid and the two spatially continuous models. The evaluation is performed using the RA method and the evaluation-on-grid by MAE and RMSE. The experiment provides the analysis of prediction accuracy of different models at the same resolution (in case of discrete models) and a comparison of how the evaluation methods evaluate conceptually different models (discrete and continuous). Prophet model was not used evaluated using RA due to its computational demands.

4.7.1 Experiment 2 - Results

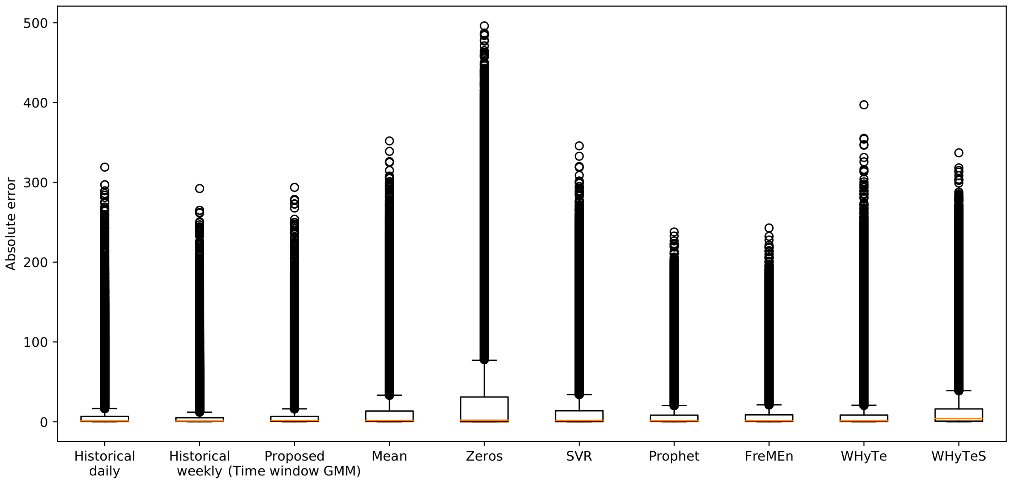

Figure 9: Box plot of absolute errors of different models.

Figure 9 shows that none of the models can fully capture the intricacies of demand fluctuation. This inability is demonstrated by the large number of absolute error outliers that each model has.

Table 1 shows how different models compare to each other. The Weekly model is the most accurate, with TW GMM being a close second when evaluated on a grid. However, TW GMM is the most accurate, with Weekly being a close second when evaluated using the RA method. This result seems to imply (perhaps not surprisingly) that evaluation-on-grid favors models trained on the same grid. The theory can be further supported by comparing

Model	MAE	RMSE	RA
Zeros	29.97	65.28	32.91
Mean	14.20	32.25	14.56
Daily	8.18	20.18	10.05
Weekly	**5.40**	**12.98**	7.42
SVR	14.87	34.85	16.71
Prophet	8.79	19.75	-
FreMEn	8.63	19.10	9.51
WHyTe	9.88	23.70	10.23
WHyTeS	14.11	27.76	9.90
TW GMM	6.61	14.97	**7.16**

Table 1: Table of different models tested on a 10x10 grid. Error metrics and RA used for evaluation. All discrete models trained on a 10x10 grid

the results of WHyTeS to the Daily and WHyTe models. WHyTeS is worse than both the discrete models in the case of evaluation-on-grid. However, WHyTeS beats both in random area predictions. When two discrete models are compared using RA and MAE, the results are very similar. The only two discrete models whose order changes are FreMEn and Daily. But the models were very close in both tests. Moreover, the order of the two models is also different when using RMSE than when using MAE. This fact that the difference in results is significant only when comparing a continuous model to a discrete one suggests that evaluation-on-grid benefits the discrete models.

4.8 Experiment 3

This experiment compares continuous models and discrete models of different resolutions. The models are compared using the fleet placement test, both with and without a given volume (see Section 3.2.2 for description). Furthermore, random area test and evaluation-on-grid at different resolutions are added. MAE-n means MAE when evaluating on a grid with n cells. Each discrete model in this experiment is trained on a 10x10 (100 cells) grid unless further specified. Prophet model was not tested in this experiment due to its computational demands.

4.8.1 Experiment 3 - Results

Table 2 shows that all the models which capture demand density reasonably well all rank similarly on fleet placement test without volume prediction. The exception is the models that are trained on a 1x1 grid and thus place the taxis uniformly. The finding suggests that there are diminishing returns in improving demand density predictions. It might also mean

Model	FPT	FPT-v	RA	MAE-1	MAE-100	MAE-40k	Size
Daily-1	352.4	524.3	11.80	606.6	37.98	0.068	1.2kB
Daily-100	124	195.5	10.05	606.6	8.18	0.104	112.6kB
Daily-40K	123.8	188.8	9.12	606.6	8.18	0.068	43.9MB
Weekly-1	351.9	443.4	9.76	**292.2**	37.82	**0.056**	8kB
Weekly-100	127.1	152.4	7.42	**292.2**	**5.40**	0.103	787.6kB
Weekly-40K	124.8	**148.6**	**6.29**	**292.2**	**5.40**	0.068	307.6MB
WHyTe	124.9	198.9	10.23	626.4	9.88	0.104	75.4
WHyTeS	129.7	225.7	9.90	644.9	13.03	0.105	3.8kB
FreMEn	128.9	203	9.51	609.16	8.63	0.105	70.8kB
SVR	128.8	331.6	16.70	1182.3	14.87	0.114	322.2kB
TW GMM	**123.7**	149.1	7.16	**292.2**	6.61	0.097	1.6MB

Table 2: Discrete models of different resolution and continuous models compared using different testing methods. Values of FPT and FPT-v are in metres.

that correct volume prediction might play a more significant role in running an efficient taxi service than having a perfect demand density model. This hypothesis is, to a certain extent, tested in the fleet placement test with volume predictions. The test shows that models that correctly predict volume (shown by MAE-1) are able to provide their service more efficiently.

The experiment also shows that it is challenging to choose the best model, according to evaluation-on-grid (or to choose a resolution in case of a discrete model). Evaluation on a coarse grid results in the models trained on this grid having similar results as finer models (or same in case of historical models). Evaluation on a fine grid is not very valid, as demonstrated in the first experiment. Also, evaluation-on-grid favors models trained on the same grid in general, as demonstrated in the second experiment. To choose the best model random area test or even fleet placement test with volumes seem to be more appropriate.

4.9 Discussion of model quality

The experiments reveal a few intriguing discoveries about the different models and their quality. The ability to model significant periodicities seems to be very important. The two best models have a hard-coded weekly periodicity. Other methods might also detect this periodicity. However, they may not be able to capture the demand fluctuations during the period entirely. The importance of period detection is best captured in the Daily models. There is a glass ceiling in the performance that the Daily models are not able to break, no matter how many cells they contain. The other finding is that the continuous methods are disadvantaged when in evaluation-on-grid (when the discrete models are trained on the same grid). Additionally, continuous models are relatively memory efficient. The WHyTeS model was the second-smallest, but ranked much better in all the test in secon experiment. The TW GMM method performed better in the proposed tests than Weekly-100. If the TW GMM object were stripped of parameters unnecessary for predictions before being saved, it would take up less than half of its current space. This decrease would mean that it would perform better than the weekly model and would be more memory efficient. The proposed test showed that it is preferable to use finer discrete models if possible. Finally, in most of the rankings, the proposed method ranked second or first, while never generating the largest models (memory-wise).

5 Conclusion

This thesis began with three main goals. Firstly, to examine methods currently used for demand modeling and evaluation of these demand predictions. Secondly, to apply chronorobotic modeling methods to the domain of demand modeling. Thirdly, to create an automatic evaluation tool able to compare them to a set of methods used in state of the art.

The evaluation method across the recent works from the field was the same (using error metrics to evaluate prediction accuracy on a grid). However, intuitively this method did not seem to encourage the creation of as informative models as possible. The thesis demonstrated the issues of the method. Furthermore, two novel methods that do not suffer the same maladies were proposed. The methods enable a fair comparison of different classes of methods (e.g., continuous and discrete). The thesis showed that chronorobotic methods could indeed be applied to the demand modeling domain. Moreover, the thesis proposed a new predictive method derived from the chronorobotic paradigm. The proposed method combines relative memory efficiency with good predictive qualities. These qualities are demonstrated in Section 4.8.1, where it ranks as one of the two best methods.

The experiments presented in the thesis were performed using the automatic evaluation tool.

The future work section (6) suggests many possible routes to be taken that could improve both demand prediction evaluation and especially demand modeling.

The main takeaways are the following. We need to choose testing methods that encourage honesty in model creation (a model that receives a good score is suitable for practical use). The method used across recent works does not necessarily do so. Therefore, we need to look for better ways to evaluate demand prediction.

Coupling of the domains of transportation demand predicting in MoD and chronorobotics benefits both areas of research. The MoD domain gains a different view of the problem and new methods of long-term prediction. Chronorobotics gains a new field of application and the opportunity to design more generalized predictive methods.

This thesis is an effort to make a step in the right direction so that the way that demand modeling methods are created and evaluated makes sense in the light of how they will be used.

6 Future work

This section considers possible future improvements in three key areas: Prediction evaluation, Time-window GMM model, and demand modeling.

6.1 Prediction evaluation

This thesis offered two novel approaches to demand prediction evaluation, which encourage the creation of models that are as informative as possible. However, each of these has its problems. The random area test requires that the models are extended by predictors of demand on an area in a specified time-window. The fleet placement test with given volumes is not very good at distinguishing models once they reach a certain degree of quality. When used with volumes, the test contains an arbitrary factor (the return of unused taxis to the station), which introduces a distance between the prediction and the test result. They are the first attempt to introduce testing, which encourages honesty in model creation, into the field. This work has shown that they are capable methods of model evaluation, probably more so than the currently used ones. However, there might be methods which might be even better and simpler tests.

It might be the case that simulations ultimately prove to be the best way to test the quality of these models. Nevertheless, a comparison of how different simulations utilize the same models will be necessary.

6.2 Time-window GMM

The model is currently discrete in the time dimension. The natural extension is finding a way to make the model spatiotemporally continuous, which might prove to be useful in other domains. The next step would be an efficient implementation and release as an open-source tool for spatiotemporal modeling.

6.3 Demand modelling

One possible direction is to split demand volume and density predictors. This approach would enable more focused modular methods that could be combined.

Since data discretization incurs a penalty in prediction accuracy, as repeatedly demonstrated in this thesis, the natural way of using past demand data would be to sample them directly. When deployed in practice, the models are used for sampling. That means the approach would be natural. In this case, the model would, for example, choose time-window from training data most similar to the one for which it predicts, then a volume predicting model would determine how many samples should be drawn from the training data. This approach has many advantages. Firstly, it places the taxis in places where there was a demand for them in the past. Secondly, these models could use neural networks in a way that makes sense for informative model creation. That means they would be used to predict volume and could utilize contextual data (such as weather). Furthermore, they could be used to choose similar times from the past from which the demand prediction would be drawn.

7 Extra chapter: Creating spatio-temporal models for social distancing

The writing of this thesis coincided with the unfortunate breakout of the COVID-19 pandemic. The recent developments have greatly influenced our lifestyles due to the comprehensive efforts to stop the spreading of the disease. One of the most effective preventive measures is social distancing [42, 43], the avoidance of close contact with other people. Koo et al.[43] show that the vast majority of cases in Singapore are linked to clusters that formed due to lack of social distancing. To follow preventive measures is crucial, especially for at-risk people (such as the elderly and the people with diabetes)[44] and their household members. Social distance is best maintained while staying at home, but for some people, leaving home for tasks such as shopping for food and prescription drugs or visiting a doctor is unavoidable. Other people might not be willing to stay at home despite being part of the at-risk group. In any case it is necessary to help people keep social distances when they leave their homes.

One such way of helping is providing the people with information about the crowdedness at different times of the places they are about to visit. This information could help them plan their errands in such a way that they visit these places when they are the least busy. That requires the ability to forecast the crowdedness of different places, which in turn requires the data about crowdedness of places upon which the forecast models could be based. The FreMEn contra COVID initiative of the Laboratory of chronorobotics at the Faculty of Electrical Engineering facilitated this effort to create a system of anonymous data collection about crowded places and then using the data to help people through Nebojsa mobile app, which provides predictions about crowdedness.

The application of chronorobotic methods to this specific problem is included as a part of this thesis.

7.1 Application of chronorobotic methods to crowdedness modeling

Let us first introduce this problem by showing its main differences and similarities to the problem of transportation demand modeling.

The main difference stems from the different usage and, thus, different requirements. The displacement of a taxi by a few tens of meters when rebalancing a fleet does not severely damage the quality of the service. In the case of modeling crowds, spatial precision is of utmost importance. A supermarket might have completely different visiting patterns than the park right next to it. Firstly, correct labeling of the data (each datum has a specific location type) is important, and secondly, data is relevant only within a particular location. These spatial limitations mean that making spatially continuous models would not make much sense in this case.

However, once the data for a specific location is extracted, it is possible to model its crowdedness the same way as the taxi demand in a specific cell of the spatial grid.

How then is the data for a specific location extracted? Different ways of achieving that were suggested. One of them was using a combination of data labels and geohash (briefly described in Section 2.1.1). However, geohash is not well suited to this problem. Firstly, it has a very limited set of resolutions. Secondly, a location for which prediction is made might lie on the boundary of two different geohash areas. The approach chosen in the end was filtering nearby and correctly labeled data for each request and creating a model from the data.

FreMEn method was used to model the crowdedness, as the amount of data collected was relatively small (compared to the taxi data) and was not evenly placed. FreMEn is robust against sparse and uneven data (unless the data is extremely uneven).

7.2 Summary

This section described the application of FreMEn, a chronorobotic modeling method, to the task of modeling crowdedness in order to simplify social distancing for people. The crowd modeling was achieved by creating temporal models for each request by filtering out relevant data and fitting model on to them.

References

[1] Alex Wallar, Menno Van Der Zee, Javier Alonso-Mora, and Daniela Rus. Vehicle rebalancing for mobility-on-demand systems with ride-sharing. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4539–4546. IEEE, 2018.

[2] Michal Cáp and Javier Alonso Mora. Multi-objective analysis of ridesharing in automated mobility-on-demand. *Proceedings of RSS 2018: Robotics-Science and Systems XIV*, 2018.

[3] Jasper Schuijbroek, Robert C Hampshire, and W-J Van Hoeve. Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research*, 257(3):992–1004, 2017.

[4] Javier Alonso-Mora, Alex Wallar, and Daniela Rus. Predictive routing for autonomous mobility-on-demand systems with ride-sharing. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3583–3590. IEEE, 2017.

[5] Matthew Tsao, Dejan Milojevic, Claudio Ruch, Mauro Salazar, Emilio Frazzoli, and Marco Pavone. Model predictive control of ride-sharing autonomous mobility-on-demand systems. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 6665–6671. IEEE, 2019.

[6] Matthew Tsao, Ramon Iglesias, and Marco Pavone. Stochastic model predictive control for autonomous mobility on demand. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3941–3948. IEEE, 2018.

[7] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. Deep multi-view spatial-temporal network for taxi demand prediction. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[8] Neema Davis, Gaurav Raina, and Krishna Jagannathan. A multi-level clustering approach for forecasting taxi travel demand. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 223–228. IEEE, 2016.

[9] Luis Moreira-Matias, Joao Gama, Michel Ferreira, Joao Mendes-Moreira, and Luis Damas. On predicting the taxi-passenger demand: A real-time approach. In *Portuguese Conference on Artificial Intelligence*, pages 54–65. Springer, 2013.

[10] Eric J Gonzales, Ci Yang, Ender Faruk Morgul, Kaan Ozbay, Mineta National Transit Research Consortium, et al. Modeling taxi demand with gps data from taxis and transit. Technical report, Mineta National Transit Research Consortium, 2014.

[11] Kai Zhao, Denis Khryashchev, Juliana Freire, Claudio Silva, and Huy Vo. Predicting taxi demand at high spatial resolution: Approaching the limit of predictability. In

2016 IEEE International Conference on Big Data (Big Data), pages 833–842. IEEE, 2016.

[12] Yongxin Tong, Yuqiang Chen, Zimu Zhou, Lei Chen, Jie Wang, Qiang Yang, Jieping Ye, and Weifeng Lv. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1653–1662, 2017.

[13] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3656–3663, 2019.

[14] Bei Pan, Ugur Demiryurek, and Cyrus Shahabi. Utilizing real-world transportation data for accurate traffic prediction. In *2012 IEEE 12th International Conference on Data Mining*, pages 595–604. IEEE, 2012.

[15] Filipe Rodrigues, Ioulia Markou, and Francisco C Pereira. Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach. *Information Fusion*, 49:120–129, 2019.

[16] Lingbo Liu, Zhilin Qiu, Guanbin Li, Qing Wang, Wanli Ouyang, and Liang Lin. Contextualized spatial–temporal network for taxi origin-destination demand prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(10):3875–3887, 2019.

[17] Xiaoyuan Liang, Guiling Wang, Martin Renqiang Min, Yi Qi, and Zhu Han. A deep spatio-temporal fuzzy neural network for passenger demand prediction. In *Proceedings of the 2019 SIAM International Conference on Data Mining*, pages 100–108. SIAM, 2019.

[18] Shan Jiang, Wentian Chen, Zhiheng Li, and Haiyang Yu. Short-term demand prediction method for online car-hailing services based on a least squares support vector machine. *IEEE Access*, 7:11882–11891, 2019.

[19] Tomáš Krajník, Tomáš Vintr, George Broughton, Filip Majer, Tomáš Rouček, Jiří Ulrich, Jan Blaha, Veronika Pěčonková, and Martin Rektoris. Chronorobotics: Representing the structure of time for service robots. 2019.

[20] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

[21] Xinyu Min, Jianming Hu, Qi Chen, Tongshuai Zhang, and Yi Zhang. Short-term traffic flow forecasting of urban network based on dynamic starima model. In *2009 12th International IEEE conference on intelligent transportation systems*, pages 1–6. IEEE, 2009.

[22] Gustavo Niemeyer. Geohash. *Retrieved June*, 6:2018, 2008.

[23] Fei Wu, Hongjian Wang, and Zhenhui Li. Interpreting traffic dynamics using ubiquitous urban data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–4, 2016.

[24] Edmund K Burke, Matthew Hyde, Graham Kendall, Gabriela Ochoa, Ender Özcan, and John R Woodward. A classification of hyper-heuristic approaches. In *Handbook of metaheuristics*, pages 449–468. Springer, 2010.

[25] Tomáš Vintr, Kerem Eyisoy, Vanda Vintrová, Zhi Yan, Yassine Ruichek, and Tomáš Krajník. Spatiotemporal models of human activity for robotic patrolling. In *International Conference on Modelling and Simulation for Autonomous Systesm*, pages 54–64. Springer, 2018.

[26] Tomáš Vintr, Zhi Yan, Tom Duckett, and Tomáš Krajník. Spatio-temporal representation for long-term anticipation of human presence in service robotics. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 2620–2626. IEEE, 2019.

[27] Tomáš Krajník, Miroslav Kulich, Lenka Mudrová, Rares Ambrus, and Tom Duckett. Where's waldo at time t? using spatio-temporal models for mobile robot search. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2140–2146. IEEE, 2015.

[28] Tomáš Krajník, Jaime P Fentanes, Joao M Santos, and Tom Duckett. Fremen: Frequency map enhancement for long-term mobile robot autonomy in changing environments. *IEEE Transactions on Robotics*, 33(4):964–977, 2017.

[29] Tomáš Krajník, Tomáš Vintr, Sergi Molina, Jaime Pulido Fentanes, Grzegorz Cielniak, Oscar Martinez Mozos, George Broughton, and Tom Duckett. Warped hypertime representations for long-term autonomy of mobile robots. *IEEE Robotics and Automation Letters*, 4(4):3310–3317, 2019.

[30] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice.* OTexts, 2018.

[31] Zhou Wang and Alan C Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117, 2009.

[32] Florian Ziel and Kevin Berk. Multivariate forecasting evaluation: On sensitive and strictly proper scoring rules. *arXiv preprint arXiv:1910.07325*, 2019.

[33] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

[34] Radford M Neal. Slice sampling. *Annals of statistics*, pages 716–717, 2003.

[35] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake Vand erPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[36] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[38] B Donovan and DB Work. New york city taxi trip data (2010-2013), 2014.

[39] City of Chicago. Taxi trips: City of chicago: Data portal, May 2020.

[40] FiveThirtyEight. Uber pickups in new york city, Nov 2019.

[41] Sean J Taylor and Benjamin Letham. Forecasting at scale. *The American Statistician*, 72(1):37–45, 2018.

[42] Michael Greenstone and Vishan Nigam. Does social distancing matter? *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-26), 2020.

[43] Joel R Koo, Alex R Cook, Minah Park, Yinxiaohe Sun, Haoyang Sun, Jue Tao Lim, Clarence Tam, and Borame L Dickens. Interventions to mitigate early spread of sars-cov-2 in singapore: a modelling study. *The Lancet Infectious Diseases*, 2020.

[44] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al. Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The lancet*, 2020.