Early Crowd Forecasting away from Stations by Geographically Complemented Regression using Transit Search and Mobility Logs

Soto Anno[1*], Kota Tsubouchi[2] and Masamichi Shimosaka[1]

[1]Department of Computer Science, Institute of Science Tokyo, 2-12-1 Ookayama, Meguro-ku, 152-8550, Tokyo, Japan.
[2]LY Corporation, Kioi Tower, 1-3 Kioicho, Chiyoda-ku, 102-8282, Tokyo, Japan.

*Corresponding author(s). E-mail(s): anno@miubiq.cs.titech.ac.jp;
Contributing authors: ktsubouc@lycorp.co.jp;
simosaka@miubiq.cs.titech.ac.jp;

Abstract

Forecasting crowd gatherings in advance, such as one week before they happen, plays a vital role in ensuring smooth mobility and public safety. Although early crowd forecasting has become possible by leveraging visitors' mobility schedules extracted from transit search logs, the forecasting area is limited to regions near railroad stations because the logs do not explicitly reflect, but only implicitly, the locations away from stations where people go after arriving. To address this issue, this paper presents an early crowd forecasting method capable of predicting crowding a week in advance in both station vicinities and areas away from stations by introducing an innovative crowd forecasting model called *geographically complemented multi-task Poisson regression (GCPR)*. Our method infers the flows of people after they arrive at railroad stations based on GPS-based mobility logs and transit search logs by leveraging the heterogeneous characteristics of nearby stations. Specifically, the model forecasts the number of visitors to an event one week in advance by using transit search logs recorded more than one week prior to the event, along with contextual features (such as day of the week) and time information. Furthermore, the model performs multi-task learning for station arrival schedules and mobility patterns, addressing the challenge of accurately predicting people flow to congestion points based on geographical and mobility proximity between stations and crowded areas. We conduct an empirical evaluation using a real-world dataset that includes 12 large-scale events held in Japan from 2019 to 2020, such as the Jingu Gaien Fireworks Festival, the

Comik Market 96, and the Rugby World Cup 2019. Results demonstrate that the GCPR can forecast crowd gatherings one week before their occurrence in areas previously challenging to predict, achieving up to 42% performance improvement over CityOutlook+, a state-of-the-art approach for early crowd forecasting.

Keywords: Crowd Forecasting, Crowd Dynamics, Active Population, Urban Computing, Mobility Logs, Transit Search Logs, GPS

1 Introduction

Early forecasting of crowding in a city is crucial for the realization of comfortable mobility and traffic management Lam et al (1999); Combinido and Lim (2012); Kim et al (2015); Hörcher et al (2017); Fedujwar and Agarwal (2024) as well as for safety purposes, such as crowd control Zheng et al (2014b); Hoang et al (2016); Lin et al (2019); Luca et al (2021a); Feliciani et al (2022); Jiang et al (2023); Minegishi (2023). Overcrowding in urban areas can lead to delays in public transportation Lam et al (1999); Kim et al (2015); Hörcher et al (2017); Fedujwar and Agarwal (2024), severe traffic jams Combinido and Lim (2012); Ali et al (2021), and sometimes threats to public safety, such as unexpected accidents due to stampedes (e.g., the fireworks festival at Akashi in 2001[1], the musical parade at Duisburg in 2010[2], the Halloween event at Itaewon in 2022[3]). Forecasting future crowd gatherings benefits various city stakeholders. For example, such forecasting enables event organizers and public security officials to allocate guards at forecasted peak congestion areas and times, helping to prevent accidents caused by stampedes. Furthermore, such forecasting allows event attendees to decide how to arrive at and leave the venues while avoiding congested areas and times.

With ubiquitous devices equipped with GPS sensors, the data-driven analysis of changes in the active population, that is, *urban dynamics*, has been conducted using GPS-based mobility logs Shimosaka et al (2015); Xia et al (2019); Xia and Li (2019), and forecasting crowdedness is now a central topic in this analysis. The simulation-based prediction of mobility flows Fan et al (2015); Zhang et al (2017); Jiang et al (2019); Lin et al (2019); Jiang et al (2023) is an example of such an approach. In these methods, the spatial distribution of mobility flows is learned using a deep neural network as a feature extractor to detect the signs of crowding that appear in the current mobility flows, which have been used for forecasting crowdedness several hours in advance. However, these methods cannot achieve early forecasting of crowdedness because the signs of crowding in the future (e.g., one week ahead) cannot be observed from the current flow of people.

In contrast, the transit search logs of railroads have been used as a promising leading indicator for the early forecasting of future crowding, such as one week ahead Konishi et al (2016); Anno et al (2020, 2023). However, such data does not

[1] https://en.wikipedia.org/wiki/Akashi_pedestrian_bridge_accident
[2] https://en.wikipedia.org/wiki/Love_Parade_disaster
[3] https://en.wikipedia.org/wiki/Seoul_Halloween_crowd_crush

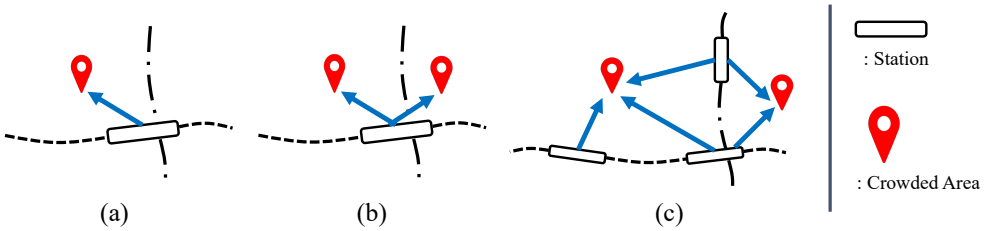


Fig. 1 Cases of the location of crowded areas and stations. (a) One crowded area near one station, (b) multiple crowded areas near one station, and (c) multiple crowded areas near multiple stations.

record the user's destination after arriving at the railroad station, and the forecast areas are limited to facilities adjacent to stations. In previous studies using transit search logs, the one-to-one rebate ratio between the active population in a crowded area and the intensity of the scheduled visits to a railroad station was studied Anno et al (2020, 2023). Thus, these methods can forecast crowding when only one nearest station is available in a crowded area (as shown in Fig. 1(a)) even if schedule information outside the station is unobtainable.

However, for cases of multiple crowded areas near a single station (Fig. 1(b)) or multiple nearest stations (Fig. 1(c)), early crowdedness forecasting cannot be performed for the aforementioned reason. This phenomenon is a critical concern because dense rail networks are typically constructed, and multiple crowded venues share multiple nearest stations in urban regions.

In this paper, we extend forecastable areas spatially by complementing population flows from the station by focusing on multiple stations nearby. More specifically, this study aims to forecast the active population based on GPS mobility logs for crowded events one week in advance with wider spatial coverage by leveraging transit search logs recorded more than one week prior to the event, along with contextual features (such as day of the week and holiday status) and time information.

However, there was a previously unencountered difficulty, in which the model tended to overlook population flows from large stations with consistently high passenger volumes and overfitted to schedule information extracted from stations experiencing a sudden surge of searches on event days, despite typically low user scales. Such minor stations show a significant increase in transit search logs on event days compared to non-event days, whereas major stations exhibit little difference. Consequently, the model excessively emphasized minor stations with relatively small absolute search log numbers, leading to underestimation and instability in crowd forecasting[4].

To overcome this problem, we propose *geographically complemented multi-task Poisson Regression (GCPR)* to forecast crowdedness early with wider spatial coverage, including both station vicinities and areas away from stations. The basic idea of GCPR is to capture how crowd congestion occurs in areas around stations and spreads geographically by introducing two key ideas: (1) *complementary learning*

[4]We provide a detailed case study on this issue using specific stations and log data in Section 3.3.

through multi-task regression and (2) *geographical and mobility proximity*. First, GCPR performed *complementary learning through multi-task regression*. Specifically, it performed simultaneous regression tasks with the active population as the primary task and the number of visitors to the station as the auxiliary task, so as not to overlook the schedule information of frequently used stations. Second, we introduced weighted optimization based on *geographical and mobility proximity*, that is, the heterogeneous characteristics between a target area and the stations, to solve the distribution problem of the proportion of people diffusing from multiple stations into a crowded area.

In terms of evaluation of the proposed method, this paper employs large-scale evaluation to confirm the effectiveness of the GCPR framework. Specifically, we empirically evaluate the forecasting performance of GCPR on 12 events that induced crowd gatherings in Japan using a large-scale real-world dataset.

The main contributions of this work can be summarized as follows:

- We propose a novel crowding forecast model called *GCPR* to enable early crowding forecasts with wider spatial coverage, including both station vicinities and areas away from stations.
- We design a complementary learning strategy based on multi-task regression with parameter sharing for capturing population flows from stations by focusing on multiple stations nearby.
- We formulate geography and mobility-weighted optimization in multi-task regression to represent a distribution of visitors to a crowded area from multiple stations.
- We evaluate the proposed methods in empirical experiments with datasets in the wild, that is, mobility logs and transit search logs on 12 events held in Japan, which highlights the predictive performance of the proposed method.

2 Related Work

First, we review the literature on **urban dynamics analytics**, which primarily deals with daily patterns of urban dynamics. We then highlight the differences between existing methods for **crowd forecasting** and our work. Finally, we review the literature on **time series forecasting**, also related to our work in that crowd forecasting can be viewed as the modeling of the time series of people flows.

2.1 Urban Dynamics Analytics

The increasing popularity of mobile device-based location history (e.g., GPS-based mobility logs) has enabled the analysis of urban dynamics. Because of the spatiotemporal properties of such data, numerous methods of tensor factorization or matrix factorization have been proposed Takeuchi et al (2013); Fan et al (2014); Zheng et al (2014a); Zhang et al (2015a). Researchers have also explored extracting urban dynamics patterns using mixture modeling Nishi et al (2014); Shimosaka et al (2016). However, these methods are not suitable for forecasting urban dynamics because of

[4] https://openstreetmap.org

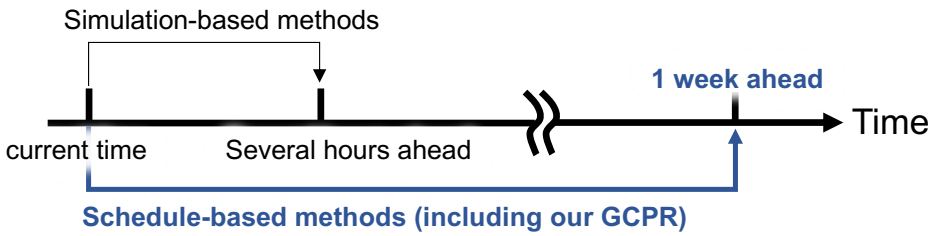


Fig. 2 Predictable time of the existing methods and GCPR.

the cold-start problem which is well-known in the recommendation system Lika et al (2014).

In contrast, studies have addressed the forecasting problem of urban dynamics using hidden Markov models Xia et al (2019); Xia and Li (2019) or generalized regression with external features Shimosaka et al (2015, 2019); Hayakawa et al (2021). Despite successful prediction results based on these methods, these studies have focused on modeling urban dynamics, such as active population patterns during commuting or shopping on a daily basis; therefore, the estimation of crowding during some events may be unstable.

2.2 Crowd Forecasting

Apart from analyzing daily urban dynamics, studies have explored the forecasting of crowded urban dynamics. Typical studies involve **simulation-based methods** assuming spatiotemporal autocorrelation in crowd densities/flows, where flows in a region and at a certain time are correlated with those in neighboring regions and subsequent times Hoang et al (2016); Zhang et al (2016); Luca et al (2021b). Thus, existing approaches use current crowd flows as covariates to predict crowd flows several hours in advance Zhang et al (2016, 2017); Lin et al (2019); Jiang et al (2023), as shown in Fig. 2. Although these methods show promising forecasting performance, the autocorrelation assumption is easily violated when the time steps between the input and output are large. For example, signs of crowded events one week from now do not appear in the current flow patterns. In Section 5, we compare our method with the DeepSTN+ Lin et al (2019) and MemeSTN Jiang et al (2023), which are state-of-the-art simulation-based crowd forecasting methods, and empirically reveal that our method can provide superior performance for early crowding forecasts.

To address the drawbacks of population flows in terms of the earliness of forecasting, the use of **schedule-based** features, such as route search history based on maps, or transit search applications, has been investigated Konishi et al (2016); Zhou et al (2018); Anno et al (2020, 2023). Such schedule information enabled us to capture future crowding at an early stage, and they could provide early crowding forecasts one week in advance, as shown in Fig. 2. However, this scheduling information primarily provides users' arrival plans at stations but is inadequate for tracking their visit plans to venues distant from stations. In such cases, estimating the number of visitors to

areas distant from stations using transit search logs is challenging, making early crowd forecasting with extensive spatial coverage difficult.

2.3 Time Series Forecasting

Time series forecasting (TSF), where the typical goal is to predict the data at future time steps based on data up to the current time step Lim and Zohren (2021), is roughly divided into two groups: short-term forecasting and long-term forecasting.

The short-term TSF technique has been applied in many fields, such as climate modeling Mudelsee (2019), biological sciences Stoffer and Ombao (2012), and finance Andersen et al (2005). With the emergence of deep learning technology in recent years, recurrent neural networks (RNN), such as LSTM Hochreiter and Schmidhuber (1997) and Gated Recurrent Unit (GRU) Chung et al (2014), are frequently used to model time-series data. Furthermore, convolutional neural networks (CNN) Krizhevsky et al (2012); Li et al (2021) and graph neural networks Defferrard et al (2016); Kipf and Welling (2016); Wu et al (2020a) have gained attention as TSF techniques for spatial data Bai et al (2018); Tran et al (2018); Wu et al (2019); Liu et al (2022), particularly for modeling spatiotemporal traffic flows Wu et al (2020b); Jiang et al (2021), air quality Du et al (2019), and abnormal events Wilson et al (2022). However, these techniques are designed for short-term forecasting, such as several minutes or hours ahead, which is not suitable for our purpose of early crowd forecasting.

In contrast to short-term forecasting, researchers have recently focused on long-term (or long-sequence) TSF, which aims to predict data in distant future time steps Chen et al (2023). Existing methods for long-term TSF have primarily focused on extracting long-term dependencies between input and output sequences, learning both short-term and long-term repeated patterns in time series Wu et al (2021). In particular, methods based on Transformers Vaswani et al (2017) have demonstrated their high model capacity for capturing long-term time series patterns Li and Moura (2020); Zhou et al (2021); Cirstea et al (2022). However, most existing methods for long-term TSF may fail to capture irregular patterns, such as crowding in the distant future, because these patterns deviate from regular ones (e.g., daily people flows during commuting hours), indicating that there is no longer a temporal dependency in the time series. Although the long-term TSF for irregular patterns has recently started to be explored by Li et al. Li et al (2024), this pioneering work is specifically designed for streamflow prediction in the hydrology field. Applying this method to crowd flow modeling is non-trivial, as this method is not designed to account for external factors such as time or visitors' schedules, which can significantly influence crowd flow changes during events.

3 Problem Setting and Baseline

In this section, we describe the problem setting of urban dynamics modeling for crowd forecasting, and its baseline. For clarity, Table 1 summarizes the notations used in this section and Section 4.

Table 1 Description of the important notations used in Section 3 and 4.

Symbol	Description
l, L	Single mesh and total number of meshes
t, T	Time segment and total number of time segments in a day
d, d'	Index representing a date
$y_{d,t}^{(l)}, \hat{y}_{d,t}^{(l)}$	Ground truth and predicted active population in mesh l on the date d and time segment t
s, S	Single station and total number of stations around the target mesh
$x_{d,t\|d'}^{(s)}$	Number of transit search logs in station s at date d and time t searched on date d'
$\boldsymbol{x}_d^{(s)}$	Schedule feature at station s on date d
$\boldsymbol{c}_d$	Context feature based on the condition of the date d
$\boldsymbol{t}$	Time feature on time segment t
$\lambda_{\boldsymbol{c}_d,t}^{(l)}$, $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{t}}^{(l)}$	Mean parameter of the Poisson distribution to model $y_{d,t}^{(l)}$ with $\boldsymbol{c}_d$ and $\boldsymbol{t}$, and the corresponding learning parameter
$\lambda_{\boldsymbol{c}_d,\boldsymbol{x},t}^{(l)}$, $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l)}$	Mean parameter of the Poisson distribution to model $y_{d,t}^{(l)}$ with $\boldsymbol{c}_d$, $\boldsymbol{x}_d^{(s)} (s = 1, ..., S)$ and $\boldsymbol{t}$, and the corresponding learning parameter
$\pi_{\boldsymbol{x},t}^{(s)}$, $\boldsymbol{W}_{\boldsymbol{x},\boldsymbol{t}}^{(s)}$	Mean parameter of the Poisson distribution to model $x_{d,t\|d}^{(s)}$ with $\boldsymbol{x}_d^{(s)}$ and $\boldsymbol{t}$, and the corresponding learning parameter
$\boldsymbol{U}^{(l)}, \boldsymbol{V}$	Two matrices decomposed from $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l)}$, where $\boldsymbol{V}$ is the shared parameter
$\boldsymbol{U}^{(s)}, \boldsymbol{V}$	Two matrices decomposed from $\boldsymbol{W}_{\boldsymbol{x},\boldsymbol{t}}^{(s)}$, where $\boldsymbol{V}$ is the shared parameter
$w_{l,s}^{\text{dist}}$, $w_{l,s}^{\text{sim}}$	Optimization weight corresponding to the geographical proximity and mobility proximity between mesh l and station s, respectively
$\alpha_{l,s}$	Weight for the optimization in multi-task learning
$\mathcal{L}^{(l)}$	Loss function to be minimized

3.1 Variable Definition and Problem Setting

We modeled urban dynamics, that is, the temporal change of an active population in a certain target area. To address urban dynamics forecasting for the detection of crowded areas, we considered the active population observed inside each square area (e.g., 500 m × 500 m), which we call the *mesh*. We count the number of mobile phone logs in each mesh at a certain duration and treat it as urban dynamics. Let l be a single mesh inside the target urban region, and each target region should be divided into L meshed areas (i.e., $l = 1, ..., L$). Furthermore, we divide one day into T time segments, denoted as t (i.e., $t = 1, 2, ..., T$). Specifically, T is set to 24, so the duration of t is one hour. The active population observed in mesh l on the date d and time segment t is defined as $y_{d,t}^{(l)}$.

We used transit search logs as a leading indicator of future crowding, following previous research Konishi et al (2016); Anno et al (2020, 2023). A transit search log is a tuple of the scheduled date d and time t, searching date d', and destination station s. Let $x_{d,t|d'}^{(s)}$ be the number of transit search logs of such searching histories. The number of scheduled visits in the station s at date d and time t searched on date d' can be represented by the number of transit search logs $x_{d,t|d'}^{(s)}$. Generally, d' denotes days prior to d (i.e., $x_{d,t|d-i}^{(s)}$ denotes the number of logs searched i days before the scheduled date d).

Fig. 3 shows an example of transit search log counts for the Rugby World Cup 2019 Final at Nissan Stadium on November 2, 2019, highlighting searches for the Kozukue station, a primary arrival point near the venue. Notably, an increase in transit search logs is observed at 16:00, one week before the event. We observe a similar trend in the transit search logs for other events used in our experiments presented in Section 5. In Section B, we provide visualizations of transit search log counts for the other events.

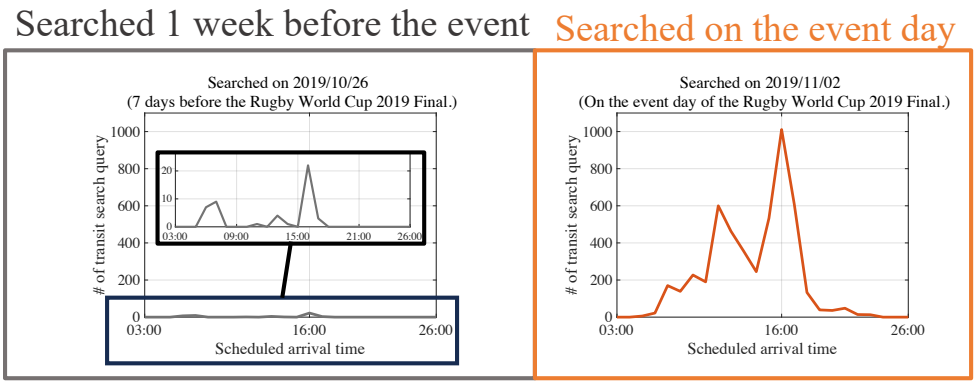


Fig. 3 Number of transit search logs searched for Kozukue station as a destination and Nov. 2, 2019, as an arrival date.

In this setting, we focused on the predictive problem of urban dynamics one week in advance (i.e., seven days before d). Specifically, given contextual information, such as the days-of-the-week or holiday-or-not, time information and scheduled visits $x^{(s)}_{d,t|d-i}(i = 7, 8, ...)$ for stations $s = 1, ..., S$ near mesh l[5], we forecast an active population $y^{(l)}_{d,t}$ on date d and time t in each mesh $l = 1, ..., L$.

3.2 Baseline Approach

For the regression problem of active population $y^{(l)}_{d,t}$, we used the bilinear Poisson regression model Shimosaka et al (2015), which is a promising approach for modeling active populations from external factors such as time and contextual information.

In this model, $y^{(l)}_{d,t}$ is assumed to follow the Poisson distribution as $y^{(l)}_{d,t} \sim \mathrm{Pois}(\cdot|\lambda^{(l)}_{\boldsymbol{c}_d,t})$, where $\boldsymbol{c}_d \in \mathbb{R}^C$ denotes the context feature represented by one-hot encoding based on the condition of the date d (such as days-of-the-week, or holiday-or-not). We define the time feature as $\boldsymbol{t} \in \mathbb{R}^T$, which is a smoothed variant of one-hot encoding on the time segment t as $\boldsymbol{t} = \{t_j | t_j = \mathcal{N}(j|\tau, \sigma^2), j = 1, ..., T\}$, where $\mathcal{N}(\cdot|\tau, \sigma^2)$

[5]In our experiment in Section 5, we extracted scheduled visits from the transit search logs of all train stations within a 2500-meter radius from the center of each mesh l and used them for our model construction. A detailed description of the experimental setup is provided in Section 5.

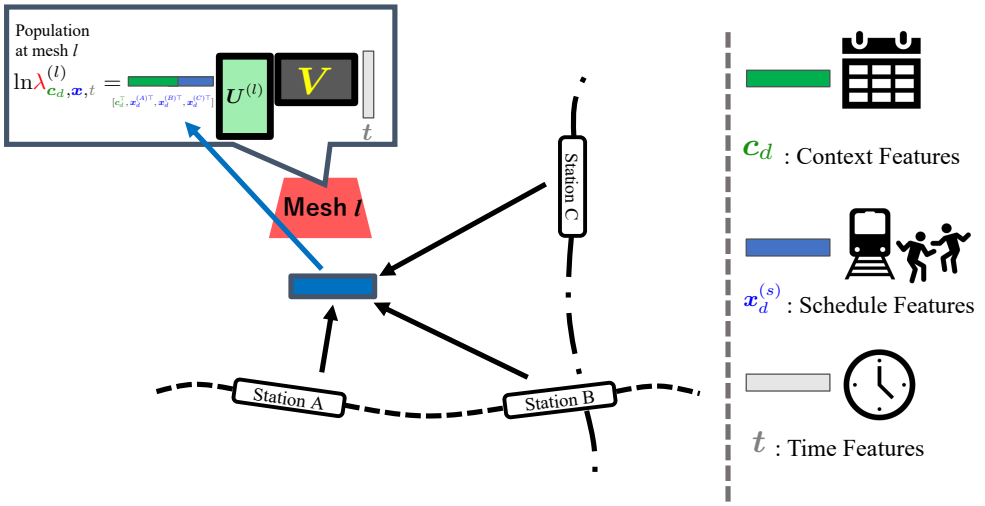


Fig. 4 Conceptual illustration of the baseline approach. This method aggregates schedule features (specified as a blue rectangle) from the nearest stations and models the population at mesh l.

indicates a Gaussian distribution with mean τ and variance σ^2. To handle the combination of the context feature and time feature, the bilinear Poisson regression uses a bilinear representation, as follows:

$$\ln \lambda^{(l)}_{\boldsymbol{c}_d,t} = \boldsymbol{c}_d^\top \boldsymbol{W}^{(l)}_{\boldsymbol{c},\boldsymbol{t}} \boldsymbol{t}, \tag{1}$$

where $\boldsymbol{W}^{(l)}_{\boldsymbol{c},\boldsymbol{t}} \in \mathbb{R}^{C \times T}$ is the learning parameter.

Although this model is suitable for extracting daily urban dynamics patterns, the model cannot predict the increase in active population during crowding because the model considers only static features and does not consider any clues of future crowd gatherings. Furthermore, the number of scheduled visits is not directly available on meshes without stations. Therefore, we used one of the simplest extensions of this model to incorporate future crowding features into its inputs.

As displayed in Fig. 4, we modeled the increase in the active population in the mesh outside the stations by using the transit search logs of multiple stations $s = 1, ..., S$ around mesh l. The schedule feature $\boldsymbol{x}_d^{(s)} \in \mathbb{R}^{p_w T}$ is at station s as $\boldsymbol{x}_d^{(s)} = \{x^{(s)}_{d,j|d-i} | i = p_d, p_d + 1, ..., p_d + p_w, j = 1, ..., T\}$, where p_d is the earliest day before the scheduled date d, and p_w is the range of days considered. The bilinear representation of the model is reformulated as follows:

$$\ln \lambda^{(l)}_{\boldsymbol{c}_d,\boldsymbol{x},t} = [\boldsymbol{c}_d^\top, \boldsymbol{x}_d^{(1)\top}, \boldsymbol{x}_d^{(2)\top}, ..., \boldsymbol{x}_d^{(S)\top}] \boldsymbol{W}^{(l)}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}} \boldsymbol{t}. \tag{2}$$

where $\boldsymbol{W}^{(l)}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}} \in \mathbb{R}^{(Sp_wT+C) \times T}$ is the learning parameter. This extension indicates that the model selectively incorporates the increase of scheduled visits to multiple surrounding stations, and infers crowd gathering at each mesh that occurs in conjunction

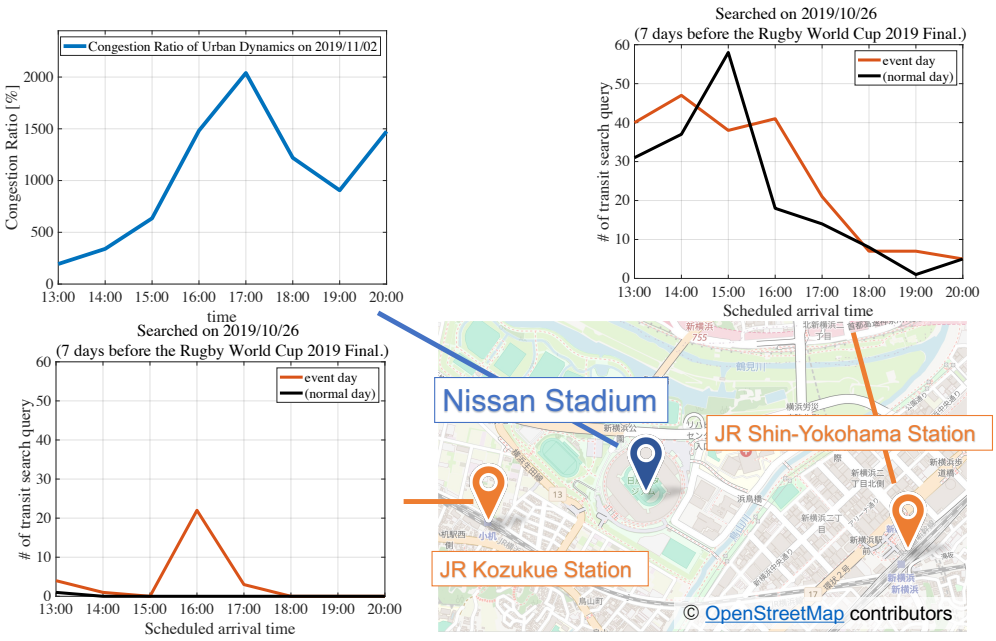


Fig. 5 Degree of crowding around Nissan Stadium and the number of transit search logs searched one week before the Rugby World Cup 2019 Final held on November 2, 2019, for the Kozukue and Shin-Yokohama stations, whose scheduled arrival date is November 2, 2019. Note that the number of transit search logs at the Kozukue station on a normal day is small because this station is ordinarily used by local commuters who rarely search transit routes for commuting. Map data: ©OpenStreetMap[7] contributors.

with congestion at neighborhood stations as complementary information. This phenomenon indicates that the learning parameter of the model represents the intensity of population flows from each surrounding station, and the distributional proportions of the inflows of the people into the mesh among multiple stations.

3.3 Challenges of Forecasting Outside Stations with the Baseline Approach

Although the baseline model presented in Section 3.2 uses transit search logs from multiple stations around the target mesh, this approach may result in an oversight of population spread from frequently used stations, as discussed in Section 1. To concretely describe this issue, as shown in Fig. 5, we provide a case study of the degree of crowding based on GPS mobility logs (blue), and the number of transit search logs obtained one week before the event day (red) and one week before a non-event day (black) in two types of stations[6], i.e., frequently used stations with a large number of users and infrequently used stations.

[6] In our experiment presented in Section 5, we extracted scheduled visits based on transit search logs for all train stations, including subway and suburban stations. This point is described in Section 5 in detail.

The figure shows that Nissan Stadium was 20 times more crowded on the event day than usual, indicating that the area is normally not so crowded, yet the number of people increases significantly during an event. In contrast, as shown in the same figure, many people are scheduled to visit the frequently used station (i.e., the Shin-Yokohama station) daily, whereas the number of search logs for the Kozukue station is usually small, because the Kozukue station serves fewer lines than the Shin-Yokohama station.

In such cases, the values for frequently used stations in the learning parameter $\boldsymbol{W}^{(l)}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}$ may become relatively smaller than those for infrequently used stations, because the model regresses the number of visitors from stations with many daily search logs to less crowded areas. Since the baseline method focuses on the increase in scheduled visits to stations on the event day, it overlooks frequently used stations with small increases, such as the Shin-Yokohama station.

Moreover, as shown in Fig. 5, the increase in transit search logs for the Kozukue station on the event day, compared to the non-event day, is more pronounced than that for the Shin-Yokohama station. As a result, the model tends to overfit to infrequently used stations with a small absolute number of transit search logs, such as the Kozukue station.

This phenomenon may cause an underestimation of the active population in the crowded mesh, because the model tends to overlook frequently used stations with many scheduled visits and to overfit infrequently used stations with a relatively small number of transit search logs. Therefore, this overlooking issue renders complementary learning challenging. Moreover, many stations with such distinct characteristics exist in urban areas with dense rail networks, highlighting the urgent necessity of addressing this issue for crowd forecasting in urban areas.

4 Proposed Method: GCPR

4.1 Basic Concept of GCPR

Motivated by the issue discussed in Section 3, we propose the geographically complemented multi-task Poisson regression (GCPR) model, whose schematic diagram for the early crowd forecasting framework is shown in Fig. 6. The GCPR model regresses the active population of the target mesh using context, time, and schedule features, as described in Section 3, while employing multi-task learning based on geographic completion to model the spread of congestion from the station to the target mesh. The GCPR model comprises the following three major stages:

- **Complementary Learning Through Multi-Task Regression** (Section 4.2)
- **Representing Distribution by Geographical and Mobility Proximity** (Section 4.3)
- **Parameter Learning Through Weighted Optimization** (Section 4.4)

As introduced in Section 4.2, GCPR solves the multi-task regression of the active population and station visits simultaneously so that it can extract not only the schedule features of infrequently used stations (i.e., stations with an absolute small number of users in the crowding) but also the schedule features of large-scale stations with

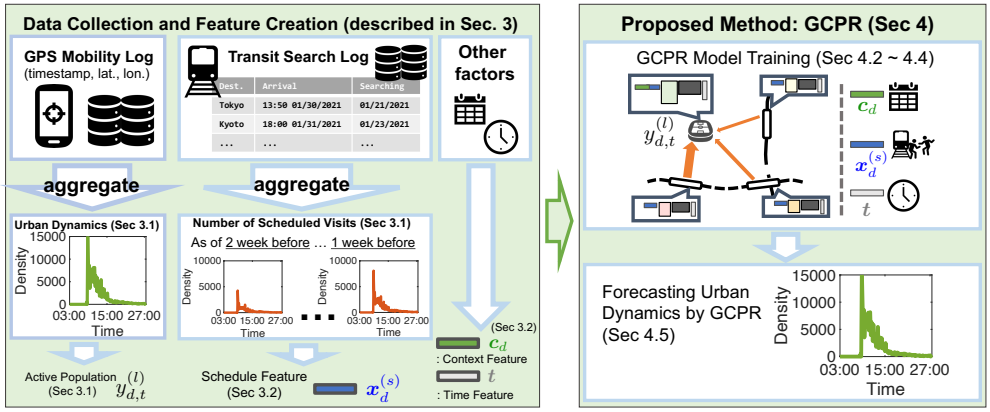


Fig. 6 Overview of the early crowd forecasting framework based on the proposed GCPR model. As described in Section 3, we collect the GPS-based mobility logs, count them for each mesh and time segment, and treat the number of logs as the active population. We also create time, context, and schedule features, each based on time, calendar information, and transit search logs, respectively. We then train the proposed GCPR model using these features and the active population. The model is subsequently used to forecast urban dynamics for each mesh.

numerous users (i.e., a large population flow from them) and achieve complementary population flow learning.

As discussed in Section 4.3, GCPR leverages these heterogeneous characteristics as weights for simultaneous optimization in multi-task learning. Intuitively, the proportion of the population flow to the mesh from each station can be described by the distance between the mesh and station, and whether congestion co-occurrence is observed during the same time period. To elucidate such characteristics, we introduce geographic and mobility proximity and solved the distribution problem of population inflows so that the model can capture various population inflows in different meshes sharing same surrounding stations and achieve spatially extensive and precise crowding forecasts.

4.2 Complementary Learning Through Multi-Task Regression

To address the aforementioned issue of the baseline method, we extend the method using multi-task learning that models the active population and station visits simultaneously. Multi-task learning involves learning multiple tasks in a single model so that knowledge obtained from auxiliary tasks can be used by a main task Caruana (1997); Sener and Koltun (2018).

In our setting, as shown in the top figure of Fig. 7, the primary task is to learn the active population at each mesh, while modeling the intensity of visits to surrounding stations serves as the auxiliary task. The primary task is formulated in Eq. (2), where the active population $y_{d,t}^{(l)}$ is regressed using the time feature, context feature, and schedule features of the surrounding stations.

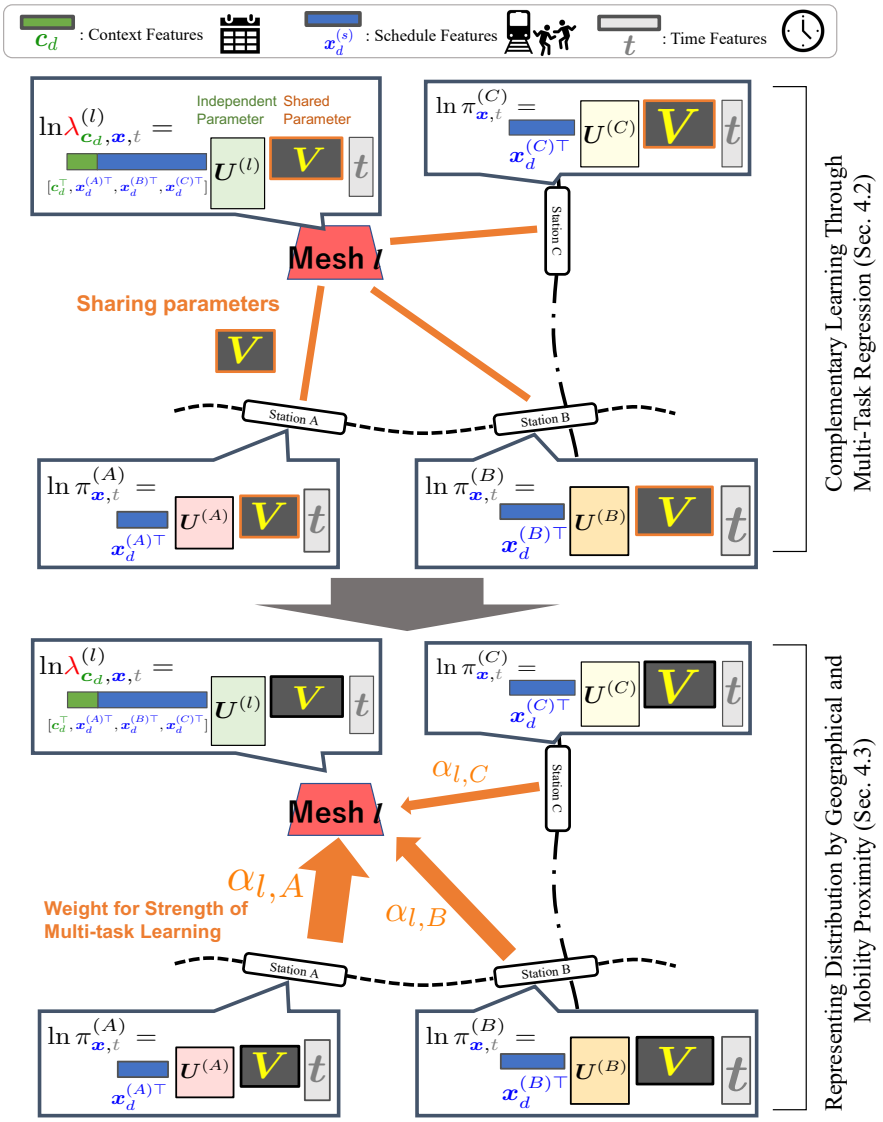
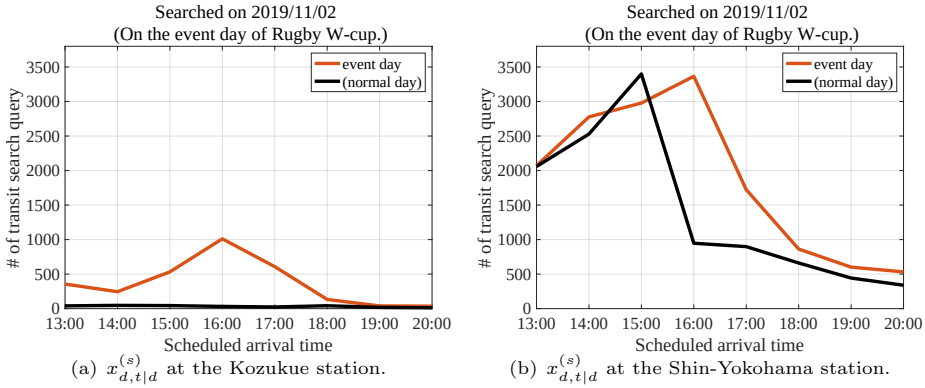


Fig. 7 Conceptual illustration of the key components of the GCPR model: complementary learning (Section 4.2), and weight optimization by representing population distribution by geographical- and mobility-based proximity (Section 4.3).

In the auxiliary task, we consider the number of people scheduled to visit the surrounding stations on the same date d and time segment t as the active population $y_{d,t}^{(l)}$. We defined the intensity of visits to station s as the number of transit search logs recorded on the same date as the scheduled visit, i.e., $x_{d,t|d}^{(s)}$, because more people search for transit on event days as shown in Fig. 3; hence, the number of logs on the event day is considered to accurately reflect station visitation patterns. Fig. 8 shows

examples of the intensity of visits to two stations, the Kozukue station and the Shin-Yokohama station, both on the non-event day and on the event day of the Rugby World Cup 2019 Final.

Although the total number of search logs up to the day of the event may be used as an indicator of visit intensity, there is a risk of overestimating this intensity because some individuals perform multiple searches, as reported by Anno et al (2020). Similarly, the number of GPS logs at a station may also be used; however, the GPS logs recorded around the station might, for example, represent visitors to a shopping mall adjacent to the station, and thus the number of such GPS logs does not necessarily indicate station visitors.



(a) $x^{(s)}_{d,t|d}$ at the Kozukue station. (b) $x^{(s)}_{d,t|d}$ at the Shin-Yokohama station.

Fig. 8 Visualization of the intensity of visits $x^{(s)}_{d,t|d}$ for the Kozukue station and the Shin-Yokohama station, where d is on November 2, 2019, when the Rugby World Cup 2019 Final was held.

The auxiliary task involves regressing $x^{(s)}_{d,t|d}$ based on the time feature, context feature, and schedule feature of station s. Similar to the baseline approach, $x^{(s)}_{d,t|d}$ is modeled using a bilinear Poisson regression, as follows:

$$x^{(s)}_{d,t|d} \sim \mathrm{Pois}(\cdot|\pi^{(s)}_{\boldsymbol{x},t}), \tag{3}$$

$$\ln \pi^{(s)}_{\boldsymbol{x},t} = [1, \boldsymbol{x}^{(s)\top}_{d}]\boldsymbol{W}^{(s)}_{\boldsymbol{x},\boldsymbol{t}}\boldsymbol{t}, \tag{4}$$

where $\boldsymbol{W}^{(s)}_{\boldsymbol{x},\boldsymbol{t}} \in \mathbb{R}^{(1+p_w T)\times T}$ is the learning parameter, and $\boldsymbol{x}^{(s)\top}_{d}$ is the schedule feature at station s.

As discussed in Section 4.1, our goal in multi-task learning is to estimate the number of visitors to the target mesh l from surrounding stations, including both large and small ones. We achieve this by sharing the learning parameters of the regression models between the primary and auxiliary tasks for all surrounding stations. Specifically, the learning parameters $\boldsymbol{W}^{(l)}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}$ and $\boldsymbol{W}^{(s)}_{\boldsymbol{x},\boldsymbol{t}}$ $(s = 1, ..., S)$ are decomposed into the multiplication of two matrices as follows:

$$\boldsymbol{W}^{(l)}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}} = \boldsymbol{U}^{(l)}\boldsymbol{V}^{\top}, \tag{5}$$

$$\boldsymbol{W}^{(s)}_{\boldsymbol{x},\boldsymbol{t}} = \boldsymbol{U}^{(s)}\boldsymbol{V}^{\top}, \tag{6}$$

where $\boldsymbol{U}^{(l)} \in \mathbb{R}^{(Sp_wT+C)\times K}$, $\boldsymbol{U}^{(s)} \in \mathbb{R}^{(1+p_wT)\times K}$ are the unique parameters for mesh l and station s, respectively, and $\boldsymbol{V} \in \mathbb{R}^{T\times K}$ is the shared parameter across all models. The mean parameters of the Poisson distribution, defined in Eq. (2) and Eq. (4), can be expressed as follows:

$$\ln \lambda^{(l)}_{\boldsymbol{c}_d,\boldsymbol{x},t} = [\boldsymbol{c}_d^{\top}, \boldsymbol{x}_d^{(1)\top}, \boldsymbol{x}_d^{(2)\top}, ..., \boldsymbol{x}_d^{(S)\top}]\boldsymbol{U}^{(l)}\boldsymbol{V}^{\top}\boldsymbol{t}, \tag{7}$$

$$\ln \pi^{(s)}_{\boldsymbol{x},t} = [1, \boldsymbol{x}_d^{(s)\top}]\boldsymbol{U}^{(s)}\boldsymbol{V}^{\top}\boldsymbol{t}. \tag{8}$$

Note that while the proposed model can forecast crowding one week in advance, it uses transit search logs recorded on the day of the event solely for training. Specifically, the number of transit searches on the day of the event serves as the target variable in a multi-task regression during model training. However, as described in Section 3, during forecasting, the model makes predictions based exclusively on scheduled visits derived from transit searches recorded strictly more than one week prior to the event. This process is formally detailed in Section 4.5.

The advantage of this approach is that it likely fits the higher intensity of station visits in the auxiliary tasks so that overlooking large and frequently used stations can be avoided. Imagine, for example, a station with $s = 1$, which is frequently used. In this case, the value corresponding to station $s = 1$ in the learning parameter $\boldsymbol{U}^{(l)}$ may be lower than those for other stations, as discussed in Section 3.3.

In contrast, the parameter $\boldsymbol{V}$, shared across models, captures the intensity of visits at the station $s = 1$, which is high even on a non-event day as shown in Fig. 8(b), as well as the active population in the mesh, which is less crowded on a non-event day as shown in Fig. 5. Therefore, the value of the learning parameter $\boldsymbol{W}^{(l)}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}$ (defined as the product of $\boldsymbol{U}^{(l)}$ and $\boldsymbol{V}$) for a frequently used station can be prevented from becoming unnecessarily small, thereby helping to avert the underestimation of crowding in the mesh. Thus, the model can learn people dispersing from the stations in a complementary manner, even if the models exhibit distinct characteristics in the scale of the number of users.

4.3 Representing Distribution by Geographical and Mobility Proximity

The multi-task learning through matrix factorization and parameter sharing provides complementary learning of population flows; however, the model learns the same auxiliary tasks across all meshes, and does not consider various characteristics of different meshes, that is, the difference between stations that are likely to be used or not. This phenomenon would cause the critical issue of the false detection of crowded areas when multiple crowded venues that share the same nearest stations exist.

To overcome this problem, we extend the multi-task learning framework with weighted optimization, which incorporates weights representing the heterogeneous characteristics between each mesh and station, as shown in the bottom figure of Fig. 7. The weights defined herein are employed to optimize the learning parameter via weighted optimization, as indicated in Eq. (11). Specifically, the weights determine the degree of fit of the shared parameter $\boldsymbol{V}$ to the visit intensity at each station. This issue is described in detail in Section 4.4.

To determine the heterogeneous characteristics between a mesh and stations, we introduce the concept of *geographical and mobility proximity*. The geographical proximity indicates that the smaller the distance between a station and a mesh, the larger the proportion of visitors in the mesh using the nearest station. Thus, the optimization weight $w_{l,s}^{\mathrm{dist}}$ corresponding to the geographical proximity is defined as follows:

$$w_{l,s}^{\mathrm{dist}} = \exp(-\eta_{\mathrm{dist}}\mathrm{dist}(l,s)), \tag{9}$$

where $\eta_{\mathrm{dist}} > 0$ and $\mathrm{dist}(\cdot,\cdot)$ represents the Manhattan distance between mesh l and the mesh containing station s. The geographical proximity weight takes a larger value as the distance between mesh l and station s decreases, as shown in Fig. 9, leading to stronger weighted optimization during parameter inference, as described in Section 4.4.

Mobility proximity represents that if the active population of mesh l increases simultaneously with the number of visitors to station s, that station is more likely to be used for visits to the mesh (even if the station is somewhat distant). Such cases are common due to differences in the size and convenience of stations. To quantify mobility proximity, we use the correlation coefficient between the urban dynamics at mesh l and the daily fluctuations in scheduled visits to station s. Thus, the optimization weight $w_{l,s}^{\mathrm{sim}}$ corresponding to mobility proximity is defined as follows:

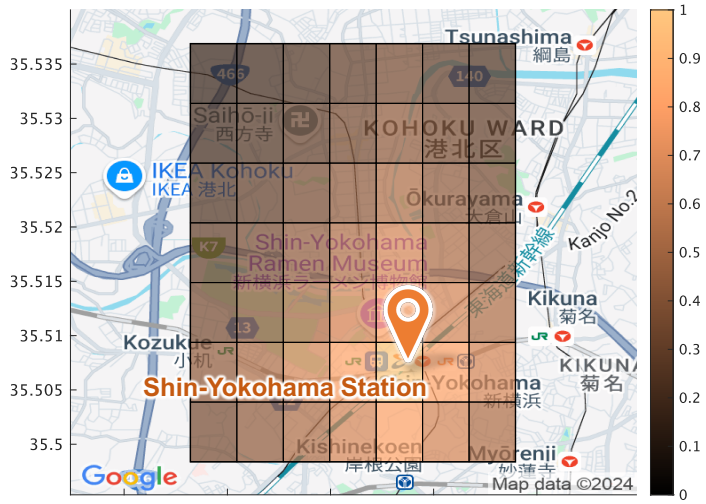
$$w_{l,s}^{\mathrm{sim}} = (1 + \mathrm{corr}(l,s))/2, \tag{10}$$

where $\mathrm{corr}(l,s)$ represents the correlation coefficient between $y_{d,t}^{l}$ and $x_{d,t|d}^{(s)}$ in the training dataset. The mobility proximity weight takes on a larger value when $y_{d,t}^{l}$ and $x_{d,t|d}^{(s)}$ are highly correlated. By definition, $0 < w_{l,s}^{\mathrm{dist}} \leq 1$ and $0 \leq w_{l,s}^{\mathrm{sim}} \leq 1$.

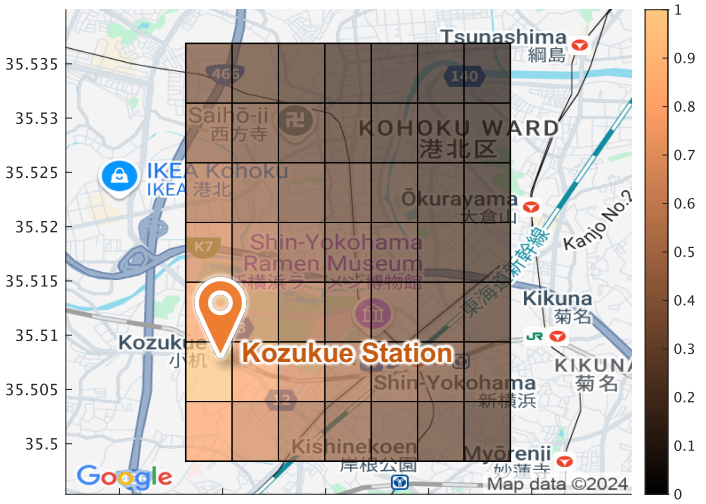
Note that the two proximity weights mentioned above are adaptive for each mesh-station combination but fixed across date and time. While these proximity weights are used in our experiment, custom metrics can be assigned based on additional geographical information (e.g., railways, stations, and residential areas) or event-related data to modify the weights.

4.4 Parameter Learning through Weighted Optimization

Parameter learning can be realized through the minimization of the loss function $\mathcal{L}^{(l)}(\boldsymbol{U}^{(l)}, \boldsymbol{U}^{(1:S)}, \boldsymbol{V})$ formulated by the summed weighted negative log-likelihood of the Poisson distribution. We first define the weights $\alpha_{l,s}(s = 1, ..., S)$ for the optimization in multi-task learning by the sum of the aforementioned two weights, that is, geographical and mobility proximity as $\alpha_{l,s} = s_{\mathrm{dist}} w_{l,s}^{\mathrm{dist}} + s_{\mathrm{sim}} w_{l,s}^{\mathrm{sim}}$, where s_{dist}



(a) Shin-Yokohama Station.



(b) Kozukue Station.

Fig. 9 Visualization of the weights based on geographical proximity. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight each station and overlay additional visualization data.

and s_{sim} are hyperparameters that determine the scale of each weight. Next, the loss function for the multi-task learning can be formulated as follows:

$$\begin{aligned}&\ln \mathcal{L}^{(l)}(\boldsymbol{U}^{(l)}, \boldsymbol{U}^{(1:S)}, \boldsymbol{V}) \\ &\quad = -\sum_{d}\sum_{t} \ln \mathrm{Pois}(y_{d,t}^{(l)} | \ln \lambda_{\boldsymbol{c}_d, \boldsymbol{x}, t}^{(l)}; \boldsymbol{U}^{(l)}, \boldsymbol{V})\end{aligned}$$

$$-\sum_{s=1}^{S}\sum_{d}\sum_{t}\alpha_{l,s}\ln \mathrm{Pois}(x^{(s)}_{d,t|d}|\pi^{(s)}_{\boldsymbol{x},t};\boldsymbol{U}^{(s)},\boldsymbol{V}), \tag{11}$$

where the first term corresponds to fitting the mean parameter of the Poisson distribution $\lambda^{(l)}_{\boldsymbol{c}_d,\boldsymbol{x},t}$ to the ground-truth active population $y^{(l)}_{d,t}$, while the second term corresponds to learning the scheduled station visits to the surrounding stations. In the second term, the larger the weight $\alpha_{l,s}$, the more strongly the visit intensity of the corresponding station is incorporated into the shared parameter $\boldsymbol{V}$.

Formally, we defined the parameter learning process with L2 regularization as follows:

$$\begin{aligned}\arg\min_{\boldsymbol{U}^{(l)},\boldsymbol{U}^{(1:S)},\boldsymbol{V}}\ &\mathcal{L}^{(l)}(\boldsymbol{U}^{(l)},\boldsymbol{U}^{(1:S)},\boldsymbol{V})\\ &+\Omega(\boldsymbol{U}^{(l)},\boldsymbol{U}^{(1:S)},\boldsymbol{V}),\end{aligned} \tag{12}$$

where $\Omega(\boldsymbol{U}^{(l)},\boldsymbol{U}^{(1:S)},\boldsymbol{V})=\gamma(||\boldsymbol{U}^{(l)}||^2_{\mathrm{Fro}}+||\boldsymbol{V}||^2_{\mathrm{Fro}}+\sum_s||\boldsymbol{U}^{(s)}||^2_{\mathrm{Fro}}),\gamma>0$ is the regularization term. $||\cdot||^2_{\mathrm{Fro}}$ is a Frobenius norm on the matrix. This optimization is not bi-convex over $\boldsymbol{U}^{(l)},\boldsymbol{U}^{(1:N)}$ and $\boldsymbol{V}$, but is convex over $\boldsymbol{U}^{(l)},\boldsymbol{U}^{(1:N)}$ with the given $\boldsymbol{V}$, and over $\boldsymbol{V}$ with given $\boldsymbol{U}^{(l)},\boldsymbol{U}^{(1:N)}$. Therefore, the model can be optimized in a similar manner to the alternating least squares algorithm Berry et al (2007), which is typically used in non-negative matrix factorization. In each convex programming, we used the L-BFGS method Liu and Nocedal (1989), which is a quasi-Newton's method Wright and Nocedal (1999).

4.5 Forecasting Urban Dynamics by GCPR

After optimizing the learning parameters, the predicted value of the urban dynamics is given by the mean parameter of the Poisson distribution. Let $\hat{\boldsymbol{U}}^{(l)}$ and $\hat{\boldsymbol{V}}$ denote the optimized learning parameters for mesh l. Then, the predicted value of the urban dynamics $\hat{y}^{l}_{d,t}$ can be formulated by exponentiating both sides of Eq. (7) as follows:

$$\hat{y}^{(l)}_{d,t}=\exp([\boldsymbol{c}_d^{\top},\boldsymbol{x}_d^{(1)\top},\boldsymbol{x}_d^{(2)\top},...,\boldsymbol{x}_d^{(S)\top}]\hat{\boldsymbol{U}}^{(l)}\hat{\boldsymbol{V}}^{\top}\boldsymbol{t}). \tag{13}$$

Although the weights $\alpha_{l,s}$, which combines the geographical proximity $w^{\mathrm{dist}}_{l,s}$ and mobility proximity $w^{\mathrm{sim}}_{l,s}$, are not directly used in Eq. (13), the equation includes the shared parameter $\hat{\boldsymbol{V}}$, whose degree of fit to each station is determined by the weights $\alpha_{l,s}$. Therefore, the predicted value $\hat{y}^{(l)}_{d,t}$ reflects the heterogeneity within each mesh in terms of patterns of station utilization, as described in Section 4.3.

5 Experiments

We conducted evaluation experiments to address the following research questions.

- RQ1: Can the GCPR accurately forecast crowding in areas, including those that are far from stations?

- RQ2: Is the GCPR more effective for early crowd forecasting compared to simulation-based methods?
- RQ3: Is the GCPR effective in scenarios with both a single crowded area or multiple crowded areas?
- RQ4: Can the proposed method suppress unnecessary reduction in the learning parameter values?
- RQ5: Do the proposed complementary learning methods via multi-task regression and weighted optimization with geographical and mobility proximity improve the performance of GCPR?

5.1 Experimental Settings

5.1.1 Dataset

Table 2 Overview of our two real-world datasets: GPS-based mobility logs and transit search logs.

Datasets	**GPS-based mobility logs**	**Transit search logs**
Sources	Mobile application by LY Corporation	Transit search engine by LY Corporation
Date Range	January 1, 2019 $\sim$ February 28, 2020 (423 days),	
Time Segments	24 time segments in a day (a single time segment of one hour).	
Area	Around event venues of 12 events (listed in Section A)	
Preprocessing	Counted by $500 \times 500m$ square mesh for each hour	Counted by each station in a target area for each hour
Other Details	Used as ground-truth urban dynamics to be predicted (as defined in Section 3.1)	Used for external factors, as is time and contextual information (as discussed in Section 3.2)

We evaluated our model on 12 events around the Greater Tokyo Area that resulted in large-scale crowd gatherings between 2019 and 2020 in Japan. The events are listed in Section A. We defined the target areas covering each region of interest corresponding to each event, and these areas were divided into $500 \times 500m$ square meshes. Furthermore, we set T to 24 to handle hourly changes in crowd congestion, as mentioned in Section 3.

To evaluate the performance of crowd congestion forecasting, we used two real-world datasets, including GPS-based mobility logs and transit search logs. The overview of these datasets are summarized in Table 2.

The GPS-based mobility logs were collected using a mobile application by LY Corporation from January 1, 2019 to February 28, 2020. Each record was completely anonymized by masking user IDs with dummies, and characterized by timestamp, latitude, and longitude. For each target event, we counted the number of mobility logs observed in each $500 \times 500m$ square mesh in the target area for each hour. This number of mobility logs per mesh and time segment was treated as active population, following the prior work Shimosaka et al (2015). We did not use any special preprocessing techniques, such as filtering or normalization. Furthermore, we did not use any dataset containing personally identifiable information for analyzing the data and building the model.

For users' schedule information, we used transit search logs, which were searched by train passengers. These logs were gathered from the transit search engine[8], also released by LY Corporation. The application was downloaded over 400,000 times, and the number of monthly users reached 18 million[9]. We collected the transit search logs for the same period as the mobility logs. Each record contains an anonymized user ID, a searching timestamp, a scheduled timestamp, and the destination station. For each event, we counted the number of search logs for each station in the target area, for each searching date, and for each time. This number of search logs was treated as scheduled visits. Analogous to the mobility logs, we did not use any special preprocessing steps for the data collection. Furthermore, we did not use any personal information to train the model.

For each mesh in the target area of the event, the model was trained using 90 days of training data up to one week before the target event, resulting in 2160 samples[10]. This setting is consistent with forecasts made one week in advance. For example, if the target event is the Tamagawa Fireworks Festival on October 5, 2019, the training data from June 29, 2019 to September 29, 2019 was used for model training. The model was then tested using one day of testing data from the event day, resulting in 24 samples[11].

5.1.2 Experimental Setups

As mentioned in Section 3, we considered one day as a 24-h period, and the number of time segments T was set to 24 (i.e., one time segment denotes a 1-h period). Following previous studies Shimosaka et al (2015), the start of a day was 3:00 AM, which had the least active population, and the end was 3:00 AM the next day (i.e., 27:00 in 24-h notation). To conduct urban dynamics forecasting one week in advance, we used the scheduled crowd dynamics observed one week in advance; thus, $p_d = 7$. We set $p_w = 7$ to consider the people's schedule patterns specified two weeks before the event day.

For each mesh l in the target area, we treated the train stations located within a 2500-meter radius of its center as the surrounding stations of mesh l. Therefore, the total number of stations S may vary among meshes. Furthermore, we extracted scheduled visits based on transit search logs for all types of train stations, including subway and suburban stations.

5.1.3 Model Setting

For the context denoted by $\boldsymbol{c}_d$, we used days-of-the-week, holiday-or-not, weekday-or-weekend features. Based on one-hot encoding, days-of-the-week feature is 7-dimensional vector and holiday-or-not and weekday-or-weekend features are two-dimensional vectors. We used the tensor product to compose these features into one vector so that $\boldsymbol{c}_d = \boldsymbol{c}_d^{(1)} \otimes \boldsymbol{c}_d^{(2)} \otimes \boldsymbol{c}_d^{(3)} \in \mathbb{R}^{28}$. For the regularization term, we set $\gamma = 0.01$. For the hyperparameter settings of the evaluation, we set $s_{\text{dist}} = 1.0$, $s_{\text{sim}} = 1.0$, $\eta_{\text{dist}} = 0.1$.

[8] https://transit.yahoo.co.jp/
[9] https://www.lycbiz.com/jp/column/display-guarantee/service-information/2020070330131360/
[10] 90 days × 24 samples per day.
[11] a day × 24 samples per day.

5.1.4 Evaluation Criteria

The evaluation criteria were based on two widely used metrics, namely, mean absolute error (MAE) and mean absolute percentage error (MAPE). The evaluation criteria are defined as follows:

$$\mathrm{MAE}_{(\mathrm{ev.})} = \mathbb{E}\left[\left|y_{d,t}^{(l)} - \hat{y}_{d,t}^{(l)}\right|\right],\ l \in \mathcal{L}_{\mathrm{event}}, t \in \mathcal{T}_{\mathrm{event}}, \tag{14}$$

$$\mathrm{MAPE}_{(\mathrm{all})} = \mathbb{E}\left[\left|y_{d,t}^{(l)} - \hat{y}_{d,t}^{(l)}\right| / y_{d,t}^{(l)}\right],\ t \in \mathcal{T}_{\mathrm{event}}, \tag{15}$$

where $y_{d,t}^{(l)}$ is the ground truth of active population, $\hat{y}_{d,t}^{(l)}$ is the prediction value on date d, $\mathcal{L}_{\mathrm{event}}$ is a set of meshes, which were the venues for each event, and $\mathcal{T}_{\mathrm{event}}$ is a set of time segments of duration on each event. $\mathrm{MAE}_{(\mathrm{ev.})}$ reveals the estimation error for the active population in event venues located away from stations, and $\mathrm{MAPE}_{(\mathrm{all})}$ represents the relative estimation error for the active population in the target meshes for each event. MAPE was used because the scale of the active population differs between the meshes around the event venues and other meshes.

Furthermore, to evaluate the forecasting performance in areas outside the event venues where crowding did not occur, and in the vicinity of stations where early crowding forecast had been realized in existing methods, the following two metrics are also used for evaluation:

$$\mathrm{MAE}_{(\mathrm{no.})} = \mathbb{E}\left[\left|y_{d,t}^{(l)} - \hat{y}_{d,t}^{(l)}\right|\right],\ l \notin \mathcal{L}_{\mathrm{event}}, t \notin \mathcal{T}_{\mathrm{event}}, \tag{16}$$

$$\mathrm{MAPE}_{(\mathrm{st.})} = \mathbb{E}\left[\left|y_{d,t}^{(l)} - \hat{y}_{d,t}^{(l)}\right| / y_{d,t}^{(l)}\right],\ l \in \mathcal{L}_{\mathrm{st.}}, t \in \mathcal{T}_{\mathrm{event}}, \tag{17}$$

where $\mathcal{L}_{\mathrm{st.}}$ is a set of meshes which includes stations. $\mathrm{MAE}_{(\mathrm{no.})}$ reveals estimation errors on the active population in the nonevent areas, and $\mathrm{MAPE}_{(\mathrm{st.})}$ represent the relative estimation error of active population only near the station, where existing methods where the early crowding forecast had already been achieved using existing methods.

When $y_{d,t}^{(l)}$ is small, it gives a very large penalty to MAPE, and it does not easily reveal effectiveness. Thus, when calculating MAPE-based criteria, the same strategy as Pan et al (2019) was adopted, filtering out samples with $y_{d,t}^{(l)} < 10$.

5.1.5 Comparison Methods

Firstly, we compare GCPR with the following baselines including the state-of-the-art early crowding forecast methods:

- **HA**: A historical average. This calculates the average values corresponding to the same time segments on the same day of the week within the training data.
- **BPReg** Shimosaka et al (2015): A baseline method of GCPR introduced in Section 3.2, which is an extension of bilinear Poisson regression with schedule feature.

- **CityOutlook+** Anno et al (2023): A state-of-the-art method for early crowding forecast with schedule features, which shows promising results on early crowding forecast **one week ahead** around the stations.

Additionally, we compare GCPR with the following simulation-based crowd forecasting methods, which achieve state-of-the-art performance in population flow prediction:

- **DeepSTN+** Lin et al (2019): A crowding forecast method **one hour ahead**. This method models spatiotemporal correlations of crowding patterns by leveraging an elaborate CNN-based architecture.
- **DeepSTN+***: A naive extension of DeepSTN+ Lin et al (2019) with the schedule feature for the crowding forecast **one week ahead**.
- **MemeSTN** Jiang et al (2023): A state-of-the-art crowd forecasting method **one hour ahead**. This method models spatiotemporal correlations of crowding patterns by leveraging GCN and GRU. Furthermore, this method employs attention-based memory module to capture historical anomaly crowding patterns.
- **MemeSTN*** Jiang et al (2023): A naive extension of MemeSTN Jiang et al (2023) for the crowding forecast **one week ahead**.

Note that the extension of DeepSTN+ and MemeSTN is denoted as DeepSTN+***** and MemeSTN*****, respectively.

5.2 Experimental Results

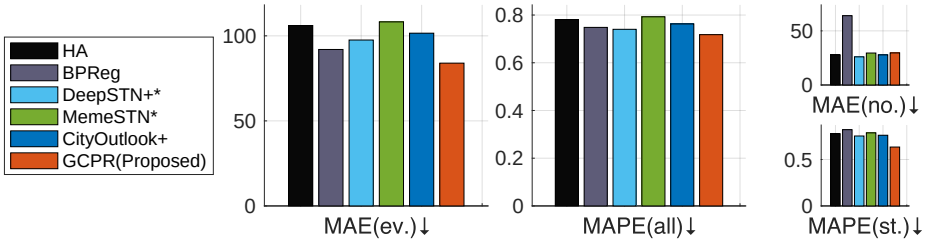
5.2.1 RQ1: Can the GCPR accurately forecast crowding in areas, including those that are far from stations?

We proposed overall performance evaluation results for one week ahead crowd forecasting over 12 events in Fig. 10, 11, and 12. Overall, the GCPR revealed promising performances in one week ahead of crowding forecast under events, even if the crowded areas were away from stations.

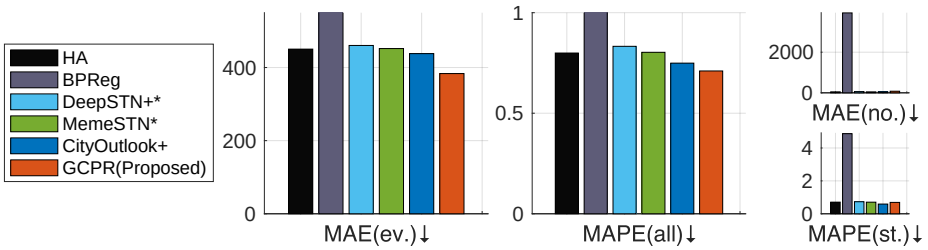
First, the proposed GCPR achieved the best performance on 10 over 12 events in $\mathrm{MAE}_{(\mathrm{ev.})}$, which is the highest percentage above the majority. Specifically, in terms of $\mathrm{MAE}_{(\mathrm{ev.})}$, the GCPR outperformed DeepSTN+* by 31% and MemeSTN* by 28% on average, indicating that GCPR could forecast crowding one week ahead more accurately than naive extensions of simulation-based methods originally designed for one-hour-ahead predictions. Furthermore, the GCPR outperformed CityOutlook+ by 21% on average and by 42% for the Jingu Gaien Fireworks Festival, the case with the highest improvement. This result indicates that GCPR could forecast crowding more accurately than the state-of-the-art one-week-ahead crowd forecasting method, even when the crowded event venues were distant from their nearest stations.

Second, the GCPR achieved the best performance on 6 over 12 events in $\mathrm{MAPE}_{(\mathrm{all.})}$, which is the highest percentage. (DeepSTN+ then achieved the best performance in four events.) Third, the GCPR achieved a forecast performance in $\mathrm{MAE}_{(\mathrm{no.})}$ and $\mathrm{MAPE}_{(\mathrm{st.})}$ that is close to that of the comparison methods.

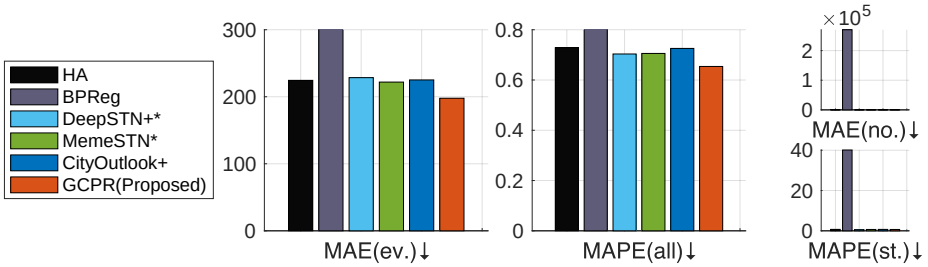
These results demonstrated the robustness of the forecasts by GCPR with both high forecasting performance in areas in which crowding occurred under various events



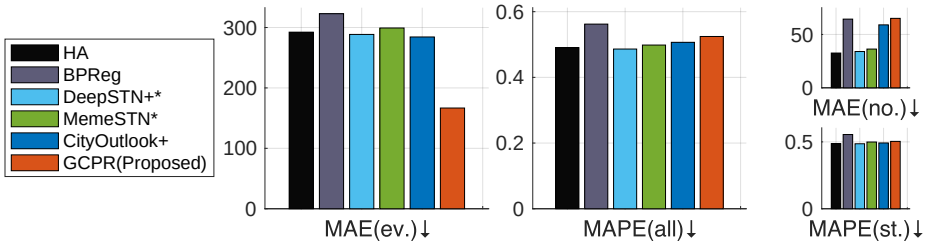
(a) Atsugi Ayu Matsuri Fireworks



(b) Adachi Fireworks Festival



(c) Tamagawa Fireworks Festival

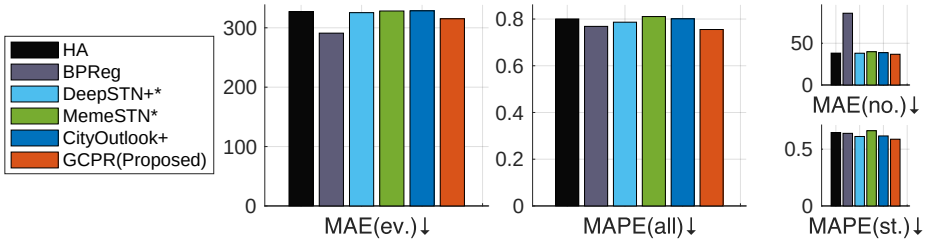


(d) Jingu Gaien Fireworks Festival

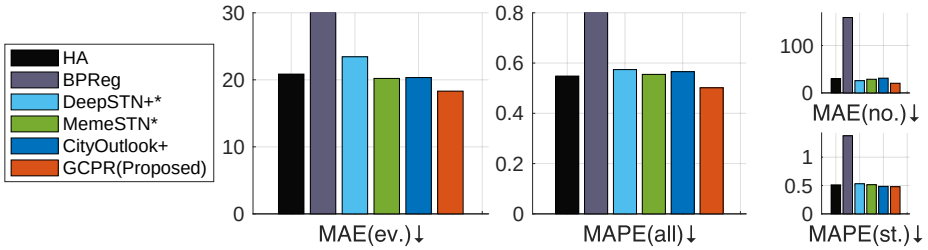
Fig. 10 Performance comparison of early crowding forecast in multiple events (1/3). The bars show the forecasting performance for each evaluation criterion, such as $\mathrm{MAE}_{(\mathrm{ev.})}$. Note that the smaller the value of the evaluation criterion, the better the method's forecasting performance.

and stable forecasting performance in other areas. Furthermore, GCPR achieved superior performance in forecasting congestion only in the vicinity of stations and in forecasting congestion in the entire region under various events.

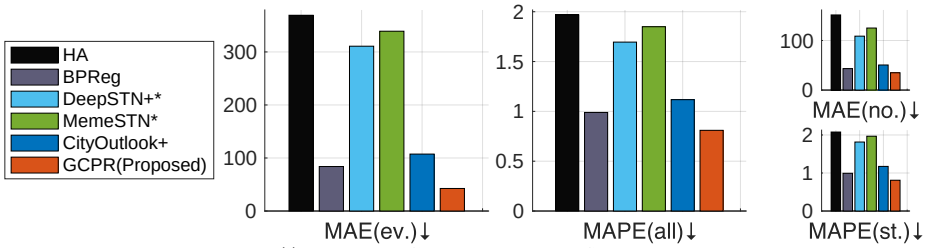
The GCPR did not achieve the best performance in all cases. Such results are discussed in Section 6.



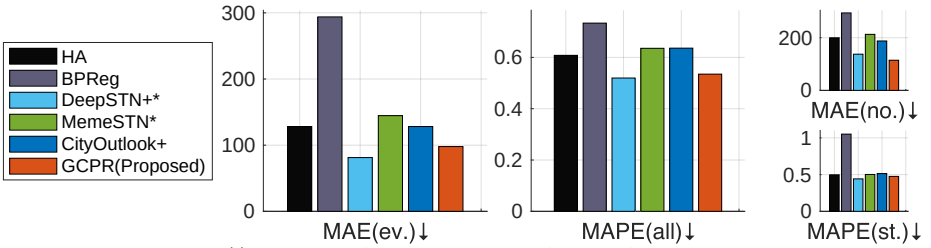
(g) Edogawa Fireworks Festival



(h) Comic Market 96 at the Tokyo Big Sight.



(i) New Year Celebration at Hie Shrine.

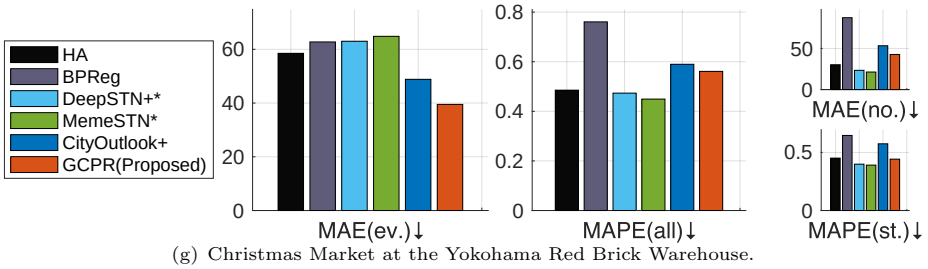


(j) Halloween Celebration around Shibuya Station.

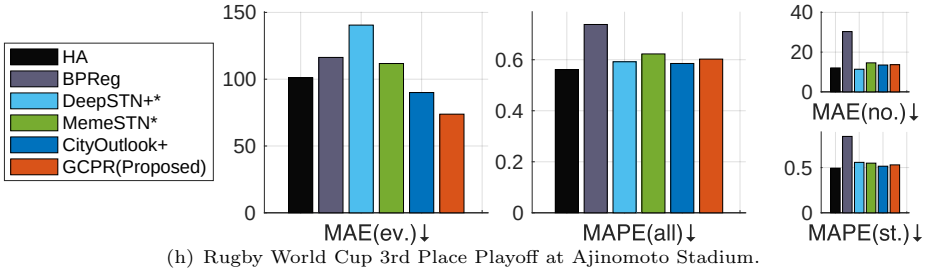
Fig. 11 Performance comparison of early crowding forecast in multiple events (2/3). The bars show the forecasting performance for each evaluation criterion, such as $\mathrm{MAE}_{(\mathrm{ev.})}$. Note that the smaller the value of the evaluation criterion, the better the method's forecasting performance.

5.2.2 RQ2: Is the GCPR more effective for early crowd forecasting compared to simulation-based methods?

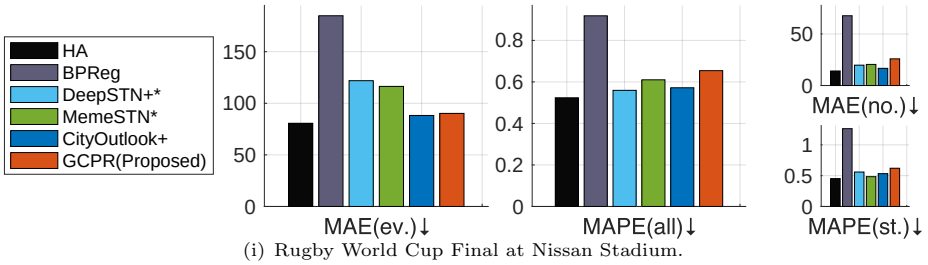
In this section, we demonstrate the efficacy of the GCPR for early crowd forecasting in comparison to DeepSTN+ Lin et al (2019) and MemeSTN Jiang et al (2023), which are the state-of-the-art simulation-based crowd forecasting methods. As reviewed in Section 2.2, these methods were designed to capture the autocorrelation patterns of the crowd flows. Therefore, we argue that the forecasting performance of these methods



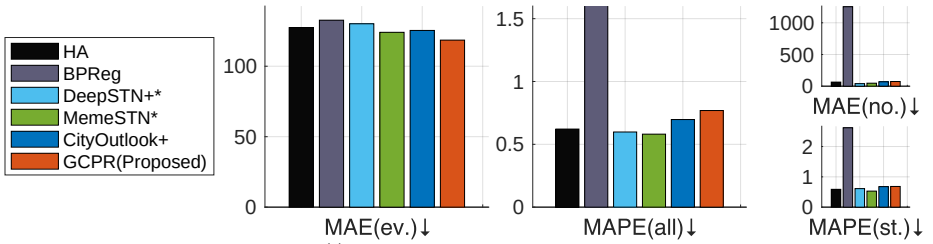
(g) Christmas Market at the Yokohama Red Brick Warehouse.



(h) Rugby World Cup 3rd Place Playoff at Ajinomoto Stadium.



(i) Rugby World Cup Final at Nissan Stadium.



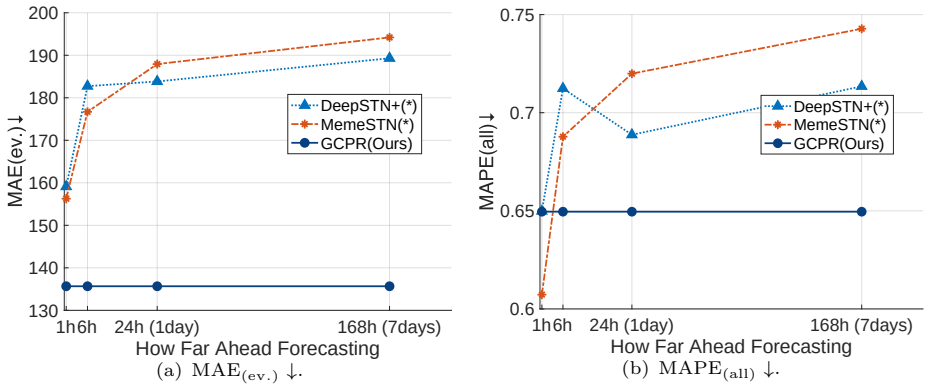
(j) 2019 Yokohama Marathon.

Fig. 12 Performance comparison of early crowding forecast in multiple events (3/3). The bars show the forecasting performance for each evaluation criterion, such as $\mathrm{MAE}_{(\mathrm{ev.})}$. Note that the smaller the value of the evaluation criterion, the better the method's forecasting performance.

degrades in one-week-ahead forecasting because the signs of crowded events occurring one week later do not appear in the current crowd flow patterns.

To empirically verify our claim and demonstrate the efficacy of the GCPR in early crowd forecasting, we compare the GCPR with DeepSTN+ Lin et al (2019) and MemeSTN Jiang et al (2023), and their naive extensions for one-week-ahead forecasting, which are denoted as DeepSTN+* and MemeSTN*, respectively.

Fig. 13 shows the comparison of the performance between GCPR and the state-of-the-art simulation-based methods. The performance of the GCPR was consistent



(a) $\mathrm{MAE}_{(\mathrm{ev.})}\downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})}\downarrow$.

Fig. 13 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods with respect to how far ahead in time each model forecasts the crowding. The performance measures visualized in the figure were calculated by averaging over 12 events. The event-wise evaluation results are provided in Section C.

over time because we used the same input features that were available one week in advance. Note that the results are averaged over 12 events; the event-wise evaluation results are provided in Appendix C.

These results confirm that the GCPR is effective for early crowd forecasting, while the state-of-the-art simulation-based methods experience reduced performance, especially as the forecasting horizon increases. Regarding the performance in $\mathrm{MAPE}_{(\mathrm{all})}$, MemeSTN achieved the best performance for the one-hour-ahead prediction but experienced a drastic reduction in performance for the one-week-ahead forecasting. The results suggest that while the autocorrelation-based approach is useful for short-term forecasting, external features, such as schedule features, are more suitable for long-term prediction.

Surprisingly, even for one-hour-ahead forecasting, the GCPR outperformed the simulation-based methods in terms of $\mathrm{MAE}_{(\mathrm{ev.})}$, which was calculated around event venues where the degree of crowding was relatively high compared to the surrounding areas. This result suggests the following two implications. First, forecasting irregular crowd flow patterns around crowded event venues is challenging, even for simulation-based methods. There were few irregular patterns in the training data from the last three months in our experimental setting, because most target events occur less than once a year. In such cases, the model's predictive performance tends to be unstable.

Second, the GCPR succeeded in suppressing such instability. This is probably because the irregular crowding patterns around event venues, which could not be identified simply by looking at past patterns, could be revealed by focusing on the complementary relationship between event venues and stations. Prior work has shown that the user's schedule patterns indicated by transit searches can reveal future irregular crowding patterns around stations Anno et al (2023). Therefore, GCPR could capture irregular crowding patterns around event venues by learning the complementary relationship between stations and event venues.

5.2.3 RQ3: Is the GCPR effective in scenarios with both a single crowded area or multiple crowded areas?

The GCPR learns the complementary relationship between event venues and the nearest stations, thereby enabling the forecasting of crowding in scenarios with both a single crowded area and multiple crowded areas. To empirically demonstrate this advantage, we perform case studies of early crowd forecasting for the following events:

- **The Rugby World Cup 2019 Final**, with three nearest stations and a single event venue, i.e., Nissan Stadium.
- **The Jingu Gaien Fireworks Festival**, with many stations around a single event venue.
- **The Tamagawa Fireworks Festival**, with five nearest stations around **two event venues**, where fireworks were launched simultaneously.

To emphasize crowd gatherings caused by the event, the ratio of the active population forecasted by each method $\hat{y}_{d,t}^{(l)}$ 1 week in advance and the active population observed on days without any events $\bar{y}_t^{(l)}$ in the area was used in the visualization with a color map. In "Ground Truth" in visualizations, the observed active population $y_{d,t}^{(l)}$ in the event was used instead of $\hat{y}_{d,t}^{(l)}$, which means that a real-time crowding detection is conducted.

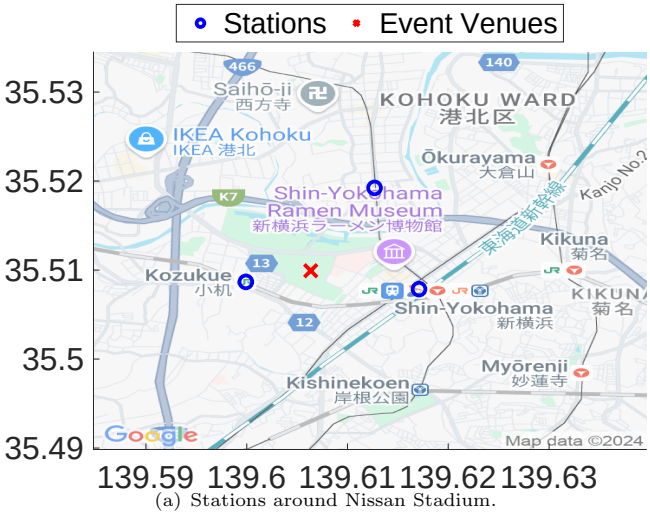
The Rugby World Cup 2019 Final

This event was held on November 2, 2019, at Nissan Stadium. The attendance reached 70,103, marking the highest of the tournament, which resulted in extreme crowding in the vicinity of the stadium [12]. Three nearby stations provided access to the venue: the Shin-Yokohama station, Kozukue station, and Kita-Shin-Yokohama station, as depicted in Fig. 14(a). The stadium opened at 15:00, and the game started at 18:00.

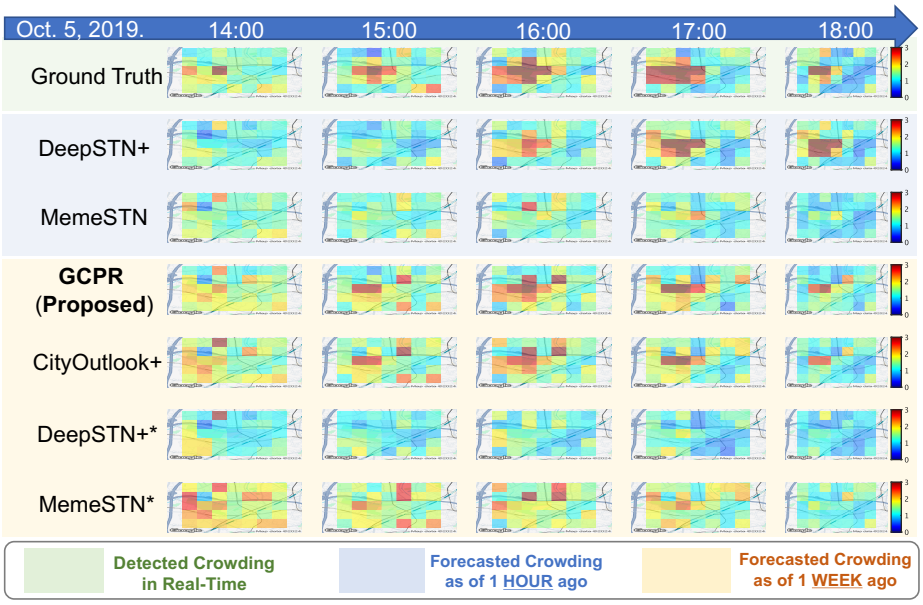
Fig. 14(b) shows a visualization of the crowding forecast for the GCPR and the comparison methods for the day. The figure presents, from top to bottom, the Ground Truth and the one-hour forecasts by DeepSTN+ and MemeSTN, followed by the one-week forecasts by GCPR, CityOutlook+, DeepSTN+*, and MemeSTN*. DeepSTN+'s one-hour forecasts (with a blue-colored background) are generally good. It predicts crowding from around 15:00, although it is a little later than the Ground Truth. In contrast, both GCPR and CityOutlook+ accurately predict the onset of crowding from 14:00, as well as its subsequent spread around and convergence at the stadium. However, DeepSTN+* (extended to one-week forecasts) does not forecast any crowding. Moreover, MemeSTN* (shown at the bottom of the figure) incorrectly detected the timing and locations of crowding.

This result indicates that CityOutlook+ and GCPR achieve the same forecasting accuracy as the one-hour forecast by DeepSTN+ for this event. There is no significant difference in the forecasting performance between CityOutlook+ and GCPR. Given that the event venue is a stadium, which frequently hosted high-attendance events, CityOutlook+ might have properly captured the historical crowding patterns even if

[12] https://www.world.rugby/news/538379/game-changing-rugby-world-cup-2019-is-a-record-breaker



(a) Stations around Nissan Stadium.



(b) Result of crowding forecast as of 1 hour ago and 1 week ago.

Fig. 14 Crowding forecast around Nissan Stadium on Nov. 2, 2019, when the Rugby World Cup 2019 Final was held. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

the event venues were distant from stations. This result will be discussed again in Section 6.

In contrast, the one-week forecast by DeepSTN+* and MemeSTN* performs poorly. As we discussed in Section 5.2.2, this result is because no signs of crowding around the venue appeared in the crowd flow patterns as of one week before the event.

The Jingu Gaien Fireworks Festival

This event was held on August 10, 2019, around Meiji Jingu Shrine in Tokyo, Japan, and attracted a total of 1 million visitors. This led to severe crowding around the venue on the event day, and many security guards were mobilized to prevent accidents. As illustrated in Fig. 15(a), several railroad stations are located around the venue, a detail officially confirmed by the event organizers[13]. Note that while the fireworks display occurs annually, the analysis herein is based on training data spanning only the three months preceding the event, excluding patterns from the previous year's display.

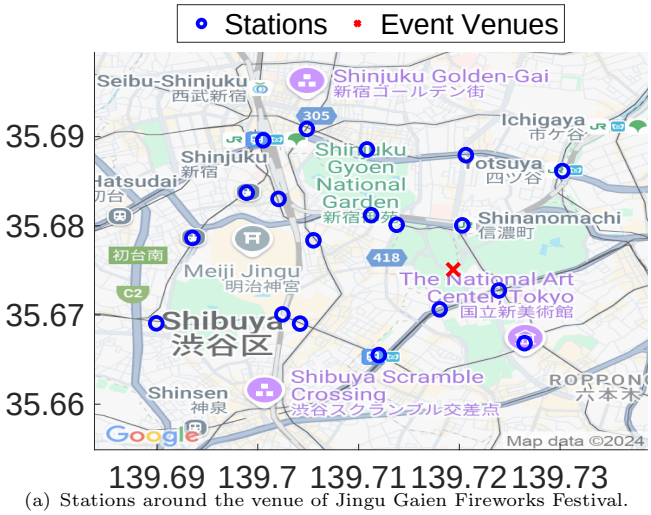
Fig. 15(b) displays the visualization of forecasting results in this event. As displayed in the result of Ground Truth, the number of people was 2 to 2.5 times larger than usual around the venue. None of the comparison methods predicted crowding at all. This may be due to the lack of crowding patterns around the venue in the training data (three months prior to the event). Nevertheless, the proposed method is able to forecast crowding from 18:00, albeit with an underestimation of the area due to the same reason. However, the proposed method also shows an early decay of crowding around 20:00. The reasons for this are discussed in Section 6.

The Tamagawa Fireworks Festival

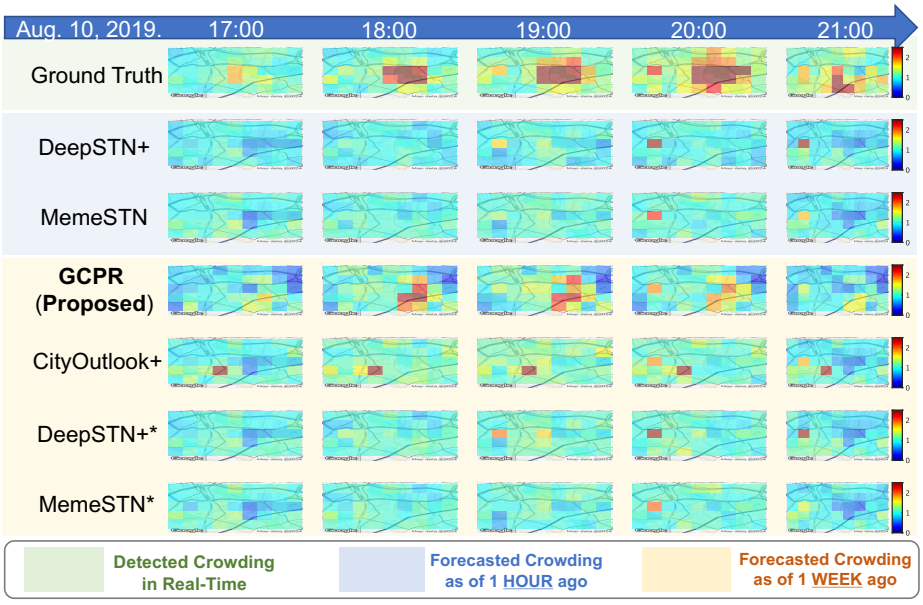
This event took place on October 5, 2019, on the banks of the Tama River. As illustrated in Fig. 16(a), fireworks were launched simultaneously at **two locations** during the event. Visitors disembarked at the railway station located between the two launch sites and proceeded to one of them. The total number of visitors was approximately 290,000. The pre-opening event started at 15:00, and the fireworks display began at 18:00[14]. It is important to note that while the fireworks display occurs annually, the analysis herein is based on training data spanning only the three months preceding the event, excluding patterns from the previous year's display.

Fig. 16(b) illustrates the forecast results: the Ground Truth data indicate that crowding begins at 14:00, with widespread crowding around the two sites evident after 15:00. DeepSTN+ and MemeSTN can partially predict crowding on the northwest side, although its forecast has limited spatial extent. CityOutlook+ fails to forecast any crowding, with the exception of some at the northwest site around 14:00. DeepSTN+*, on the other hand, does not predict crowding at the two sites but forecasts it from southwest to northeast around the train station. This may be attributed to DeepSTN+* using schedule features as input, with input flow patterns being one week old, making it challenging to pinpoint crowding locations outside the stations. Moreover, MemeSTN* only forecasts crowding around the stations. In contrast, GCPR can forecast crowding that occurs simultaneously at two different venues as well as around the stations. This indicates that GCPR can correctly model how people would disperse once they leave the stations, even with two different event venues near the stations.

[13] http://www.jinguhanabi.com/
[14] https://www.tamagawa-hanabi.com/program/



(a) Stations around the venue of Jingu Gaien Fireworks Festival.

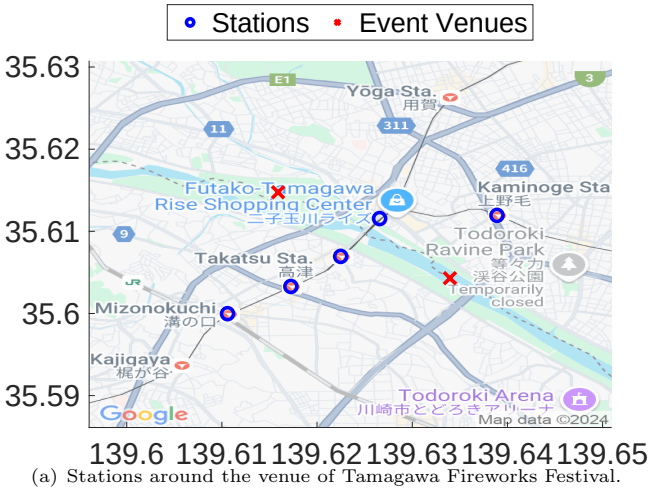


(b) Result of crowding forecast as of 1 hour ago and 1 week ago.

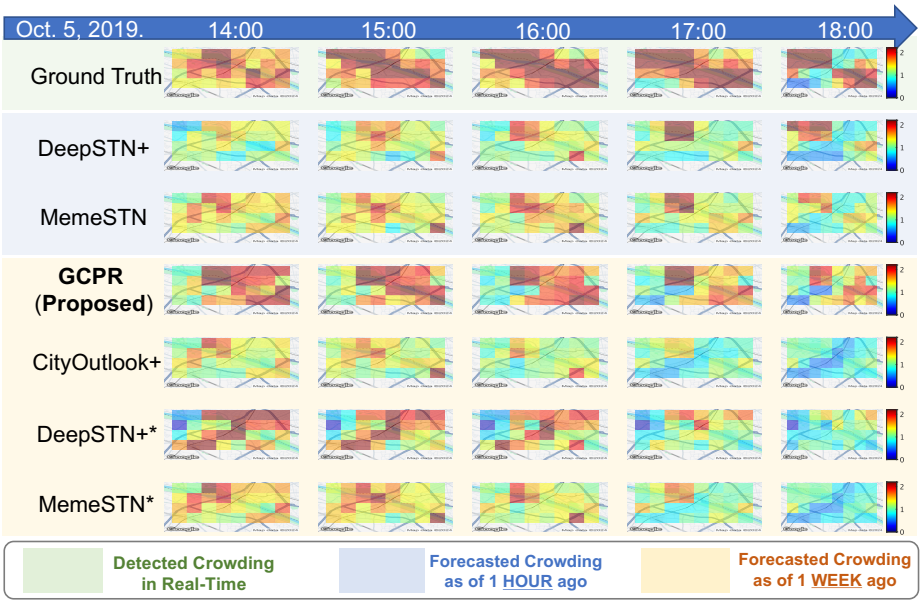
Fig. 15 Crowding forecast around the Jingu Gaien on Aug. 10, 2019, when the Jingu Gaien Fireworks Festival was held. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

5.2.4 RQ4: Can the proposed method suppress unnecessary reduction in the learning parameter values?

As introduced in Section 4, the proposed method employs parameter sharing to simultaneously regress the active population of the mesh and the intensity of station visits. Thus, we argued that the proposed method could suppress unnecessary reductions in the learning parameter values for frequently used stations.



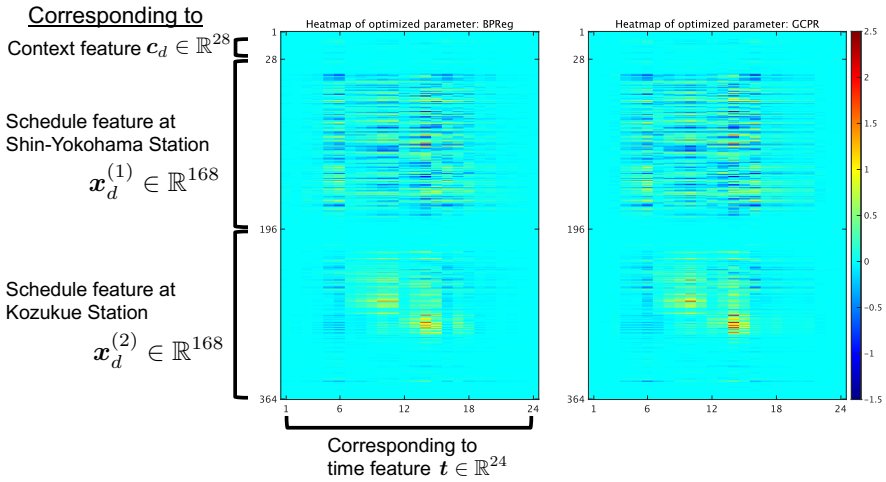
(a) Stations around the venue of Tamagawa Fireworks Festival.



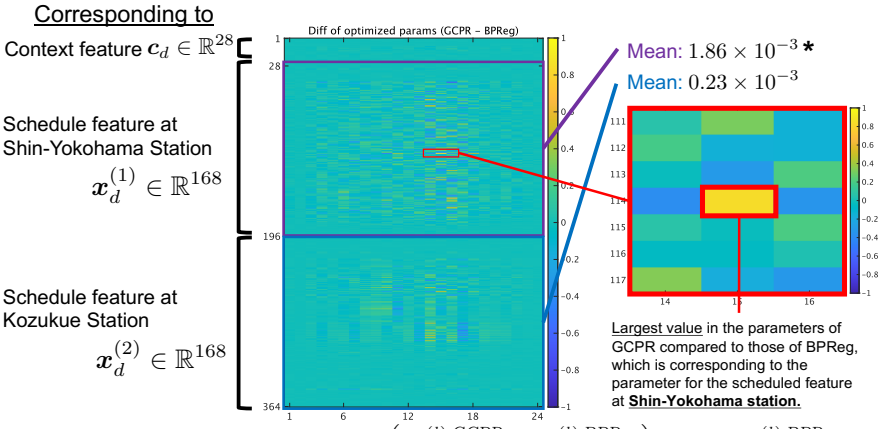
(b) Result of crowding forecast as of 1 hour ago and 1 week ago.

Fig. 16 Crowding forecast around the Tama River on Oct. 5, 2019, when Tamagawa Fireworks Festival and its pre-opening festival was held. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

To empirically validate this assumption and analyze which input features are most influential in forecasting crowding, we present the visualization of the learning parameters $\boldsymbol{W}^{(l)}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}$ (introduced in Eq. (2)) for BPReg, which is the baseline method presented in Section 3.2, and for the proposed GCPR presented in Section 4. We selected the Rugby World Cup Final held at Nissan Stadium as a case study for our analysis. For simplicity, in this experiment, we trained the two models by using the schedule features for two stations, that is, the Shin-Yokohama station (denoted as $\boldsymbol{x}^{(1)}_d$) and the



(a) Learning parameters $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l)} \in \mathbb{R}^{364\times 24}$ for BPReg (left) and GCPR (right), respectively.



(b) Difference of the learning parameters $\left(\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l),\mathrm{GCPR}} - \boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l),\mathrm{BPReg}}\right)$, where $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l),\mathrm{BPReg}}$ and $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l),\mathrm{GCPR}}$ are the learning parameters for BPReg and GCPR, respectively.

Fig. 17 Visualization of the learning parameters $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l)} \in \mathbb{R}^{364\times 24}$ for BPReg and GCPR, and their difference, in the case of the Rugby World Cup 2019 Final at Nissan Stadium. We trained the models by using the schedule features for the Shin-Yokohama station (denoted as $\boldsymbol{x}_d^{(1)}$) and the Kozukue station (denoted as $\boldsymbol{x}_d^{(2)}$), along with the context feature $\boldsymbol{c}_d$ and time feature $\boldsymbol{t}$. A single asterisk (*) in the bottom figure (b) indicates that the difference between the mean value and zero is statistically significant, as determined by a t-test at $p < 0.05$.

Kozukue station (denoted as $\boldsymbol{x}_d^{(2)}$), along with the context feature $\boldsymbol{c}_d$ and time feature $\boldsymbol{t}$. Thus, we set $S = 2$, $p_w = 7$, $T = 24$, and $C = 28$, resulting in $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l)} \in \mathbb{R}^{364\times 24}$.

Fig. 17 presents the visualization of the learning parameters for BPReg and GCPR, and their difference. Overall, we can confirm that the proposed method assigned higher

parameter values to the schedule features for the frequently used station (i.e., the Shin-Yokohama station), thereby confirming the effectiveness of the proposed multi-task learning.

First, Fig. 17(a) presents the visualization of the learning parameters for BPReg and GCPR. For both methods, larger parameter values are assigned to the schedule information than to the context feature, indicating that the schedule feature is relatively important for forecasting. Moreover, the parameter values for the Shin-Yokohama station schedule feature were generally smaller than those for the Kozukue station schedule feature. This result indicates that the model places more importance on the schedule features of the infrequently used station (i.e., the Kozukue station) than on those of the frequently used station (i.e., the Shin-Yokohama station).

In contrast to Fig. 17(a), Fig. 17(b) presents the difference in the learning parameters for the two methods, i.e., $\left(\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l),\mathrm{GCPR}} - \boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l),\mathrm{BPReg}}\right)$, where $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l),\mathrm{BPReg}}$ and $\boldsymbol{W}_{\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}}^{(l),\mathrm{GCPR}}$ denote the learning parameters for BPReg and GCPR, respectively. We have confirmed the statistical significance of the difference in the learning parameters for the Shin-Yokohama station using a t-test at the $p < 0.05$ level. Specifically, we have tested whether the values of the differences in the learning parameters come from independent random samples from normal distributions with a mean of zero. As a result, $p = 0.027 < 0.05$ for the Shin-Yokohama station, whereas $p = 0.871 > 0.05$ for the Kozukue station. We also found the largest differences in the learning parameters for the Shin-Yokohama station as shown in the red rectangle in the figure. These results indicate that the proposed method avoids omitting schedule information for frequently used stations.

5.2.5 RQ5: Do the proposed complementary learning methods via multi-task regression and weighted optimization with geographical and mobility proximity improve the performance of GCPR?

The proposed GCPR framework is characterized by two components: (1) complementary learning (**CL**) through multi-task regression, which implements the learning parameter sharing as introduced in Section 4.2, and (2) a weighted optimization (**WO**) incorporating geographical and mobility proximity as presented in Section 4.3. To analyze the contribution of these components, we conducted an ablation study by comparing the GCPR framework with its variants as follows:

- **BPReg**: This model did not employ either complementary learning through multi-task regression or weighted optimization with geographical and mobility proximity. Therefore, this model was exactly the same as the baseline method of GCPR as introduced in Section 3.2.
- **BPReg+CL**: This model employed complementary learning through multi-task regression with parameter sharing but did not employ weighted optimization with geographical and mobility proximity.

Table 3 compares the performance of GCPR with that of its variants. We can confirm the substantial contribution of complementary learning through multi-task

regression by comparing BPReg+CL with BPReg. This result suggests that the parameter sharing across the models enabled the model to capture the dispersion of people from the stations appropriately.

Furthermore, we can confirm that the weighted optimization with geographical and mobility proximity contributes to the performance gain by comparing GCPR with BPReg+CL. This result indicates that the consideration of the heterogeneity of each mesh enabled the model to properly capture crowding patterns in different meshes. From these results, we conclude that all components of the GCPR framework contribute to performance improvements.

Table 3 Ablation study of the GCPR framework.

Name	(1) CL (Section 4.2)	(2) WO (Section 4.3)	$MAE_{(ev.)} \downarrow$	$MAPE_{(all)} \downarrow$
BPReg	-	-	1061.0	2.676
BPReg+CL	✓	-	185.2	0.794
GCPR	✓	✓	**135.7**	**0.650**

6 Discussion

6.1 Findings from the experimental results

6.1.1 Forecasting Performance Evaluation

The experimental results revealed that the proposed complementary learning through multi-task regression can considerably improve predictive performance by comparing GCPR with the baseline method BPReg*. Such performance can be achieved by a sequence of simple convex optimizations by the proposed method without using more complex models with nonlinear transformations such as DeepSTN+ and MemeSTN.

In contrast, the comparison method achieved the best performance in $MAE_{(ev.)}$ and/or $MAPE_{(all.)}$ in some events. The events for which the proposed method outperformed the comparison method in $MAE_{(ev.)}$ and failed to outperform in $MAPE_{(all.)}$ include the Jingu Gaien Fireworks Festival, the Christmas Market at the Yokohama Red Brick Warehouse, the Rugby World Cup 3rd Place Playoff at Ajinomoto Stadium, and the 2019 Yokohama Marathon. This is attributed to the proposed method's poor forecasting performance for the normal active population, especially outside the event area. In fact, these events share poor forecasting performance of $MAE_{(no.)}$ in the proposed method, probably because the schedule feature becomes noise for forecasting normal patterns not associated with the increase in transit searches. It should be emphasized that forecasting non-congestion patterns falls outside the primary scope of this study. In practical scenarios, it is advisable to combine the GCPR method during events with approaches like those in Shimosaka et al (2015), which accurately predict non-congestion patterns in their absence.

During the Edogawa Fireworks Festival, the proposed method outperformed the comparative method in terms of $MAPE_{(all.)}$, but not in $MAE_{(ev.)}$. The results indicate that the performance of the proposed method is inferior to that of the baseline

method, BPReg. This inefficacy could stem from geographical complementation failing in certain meshes, such as the farthest northwestern and southeastern areas from stations. In fact, this event encompassed mesh regions that were the farthest from any station among all events. Hence, for deploying GCPR as a forecasting system in both urban and rural areas with diverse station densities, future studies should explore: (1) the correlation between station distribution around event sites and GCPR's forecasting accuracy; and (2) additional data sources for schedule information, like ticket sales or social media posts by users. The limitation and future work are discussed in Section 6.4 in detail.

Finally, the Rugby World Cup Final at Nissan Stadium was the event where the proposed method did not outperform the comparison method in terms of both $\mathrm{MAPE}_{(\mathrm{all.})}$ and $\mathrm{MAE}_{(\mathrm{ev.})}$. For this event, the historical average yielded better performance. Given the event venue's nature as a stadium, it is conceivable that events with high attendance frequently occurred in the training data. However, as the visualization results in Fig. 14(b) are satisfactory, this decline in performance is not deemed a significant issue.

6.1.2 Case Studies

For the Tamagawa Fireworks Festival depicted in Fig. 16, the degree of crowding was underestimated in regions distant from the station, notably in the northwest. This indicates a lack of effective geographical complementation for areas significantly distant from the station, echoing the observations from the Edogawa Fireworks Festival discussed in Section 6.1.1. Consequently, resolving this issue necessitates further investigation into additional data sources that could serve as indicators for detecting congestion in areas remote from stations, as mentioned in Section 6.1.1. The issue is comprehensively discussed in Section 6.4.

Fig. 15 depicts examples of inaccurate forecasts related to the early end of crowds at the Jingu Gaien Fireworks Festival. This issue could stem from the methodology that only considered the number of attendees expected to arrive at the venue based on transit search logs. To address this issue, it may be beneficial to incorporate visitor's departure schedules from the nearest stations as input features. This issue is also discussed in Section 6.4 in detail.

6.2 Reproducibility of GCPR in terms of Dataset

6.2.1 GPS-based Mobility Logs

The use of GPS location history to capture urban dynamics poses no issue regarding reproducibility. In recent years, several applications and services have been logging users' GPS-based locations, and numerous studies have been conducted using the mobility logs (e.g., "Konzatsu-Tokei (R)" from ZENRIN DataCom Co., Ltd. Jiang et al (2018, 2019), a mobile application from LY Corporation Shimosaka et al (2015); Anno et al (2020), dataset from Tencent Xia et al (2019)).

6.2.2 Transit Search Logs

In this study, we obtained people's schedule patterns from the transit search logs from LY Corporation; however, other records, such as route searching history on map applications (e.g., Google Maps, Yahoo! Map, ZENRIN Map, Japan Transit Planner, NAVITIME) or logs from travel reservation applications (e.g., Booking.com, Travelko), can also be used as scheduled patterns.

6.3 Applications of GCPR Framework

This study aims to extend the congestion forecasting system[15] that has already been used in LY Corporation's transit search application[16]. This system plays a crucial role in enabling users to avoid congestion, and organizations can implement measures against crowd gatherings. Although the existing congestion forecasting system can provide only forecasts inside the station, providing such forecasts not only for stations but also for various other places is essential for facilitating where users want to go. Moreover, this application was downloaded 40 million times in Japan, with 16 million daily active users. Therefore, the proposed congestion forecasting system can positively influence real-world traffic.

Currently, the forecasting method proposed by Konishi et al. Konishi et al (2016) is used internally in the system for providing forecasts inside stations. In addition to the current system, we plan to deploy the GCPR framework as a congestion forecasting system outside the stations. We also plan to analyze how the GCPR-based system effectively enhances the comfort of user mobility. This is part of our future work.

In addition to deploying the GCPR framework as the mobile application mentioned above, we expect urban planners or public safety officials to implement the GCPR-based crowd forecasting system by collaborating with industries that have real-world data, such as GPS mobility logs. Although the GCPR used GPS logs and transit search logs collected by LY Corporation, similar types of data collected by other companies can be used to implement this model, as discussed in Section 6.2.

6.4 Limitations and Future Work

Although the GCPR framework can provide early crowd forecasting even away from stations, we are aware that it may not work in the following three scenarios: (1) events held in areas with sparse stations or event venues distant from stations (as shown in the cases of the Edogawa Fireworks Festival and the Tamagawa Fireworks Festival, discussed in Section 6.1.1); (2) forecasting the end of crowded events (as shown in the case of the Jingu Gaien Fireworks Festival, discussed in Section 6.1.2); and (3) events/regions where people do not typically plan their visits in advance.

First, the GCPR may not accurately forecast crowding in areas with sparse stations or event venues distant from stations. Given that the GCPR assumes stations are located near the event venue, as is typical in urban areas, it may be more difficult to learn complementary relationships between stations and event venues that are distant from each other. Addressing this issue may require investigating the correlation

[15] https://blog-transit.yahoo.co.jp/congestion/
[16] https://transit.yahoo.co.jp/smartphone/app/

between station distribution around event venues and GCPR's forecasting accuracy. Furthermore, incorporating additional data beyond station-based schedules, such as ticket sales at the venue or social media posts, may be beneficial. In future studies, we plan to utilize such additional schedule information for early crowding forecasting.

Second, the GCPR may not accurately forecast the end of the crowded events because it leverages the people's arrival schedules at the stations, but does not consider how long visitors stay at the event venues. Intuitively, the number of people scheduled to arrive at the venue peaks at the beginning of the event. Conversely, focusing on the number scheduled to depart could help determine the congestion's duration and end. Therefore, future research seems to benefit from incorporating the departure schedules of attendees as covariates.

Third, the GCPR may be unable to forecast crowding in scenarios with events/regions where people do not typically plan their activities in advance. For example, spontaneous or informal gatherings, such as flash mobs and sudden celebrations following sports victories, have an inherently impulsive nature. Given that the GCPR implicitly relies on attendees searching for train routes in advance, it may be ineffective in predicting such sudden crowd gatherings. Additionally, in certain cultural or regional contexts, planning daily activities is less common, further limiting the model's effectiveness.

To address this issue, leveraging people's activity logs on social media would be beneficial. In fact, previous studies have revealed that such casual online behavior is indicative of actual event participation Du et al (2014); Zhang et al (2015b); Zhao et al (2015); Wu et al (2018); Mehmood et al (2020); Zhang and Hara (2023). Future work could explore integrating such social media data to better capture casual behaviors that visitors may perform in advance of actual events, enhancing the capabilities of early forecasting.

7 Conclusion and Future Work

In this study, we proposed GCPR for early crowding forecast one week in advance. Compared with the forecasting systems that had difficulty in achieving both early prediction and its wider spatial coverage, the proposed method extended the forecasting areas with the accurate prediction by leveraging multi-task learning with the domain knowledge derived from spatial and mobility proximity. The experimental results on real datasets revealed the model's superiority over existing methods, achieving the best performance on 10 over 12 events and outperforming the baseline approaches by up to 42% in the active population prediction in the event venues. In terms of practical implications, we plan to deploy the GCPR-based crowd-forecasting system as a mobile application and analyze how it enhances user mobility comfort. Furthermore, we expect urban planners or public safety officials to implement it in cooperation with companies that hold real-world data, such as route search logs, travel reservation logs, and GPS-based mobility logs.

Future work will attempt to integrate additional data sources (e.g., social media posts) to forecast crowding in areas with sparse stations or event venues distant from stations and to forecast unexpected crowd gatherings that are not publicly announced

in advance. We will also investigate users' departure schedules from stations to forecast the end of crowded events.

The data that support this study's findings will be available from LY Corporation, but they are subject to restrictions. These data were used under license for the current study and are not publicly available. However, the data will be available from the authors upon reasonable request and with the permission of LY Corporation.

Acknowledgements. This work was partially supported by JSPS KAKENHI Grant Number 22J22725. We would like to thank Editage (www.editage.jp) for English language editing.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data that support this study's findings will be available from LY Corporation, but they are subject to restrictions. These data were used under license for the current study and are not publicly available. However, the data will be available from the authors upon reasonable request and with the permission of LY Corporation.

It is against our corporate confidentiality to open the source code. However, we believe that the description of the proposed framework and the procedures for data preprocessing give enough information to build an equivalent system for the readers and other researchers and to achieve reproducibility.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was partly supported by JSPS KAKENHI Grant Number 22J22725.

Authors' contributions

S.A. was responsible for data processing, model formulation and implementation, experiments, and writing the first draft of the manuscript; K.T. was responsible for data processing, interpretation of experimental results, and writing the manuscript; M.S. was responsible for overall supervision of the research project, model formulation, interpretation of experimental results, and writing the manuscript.

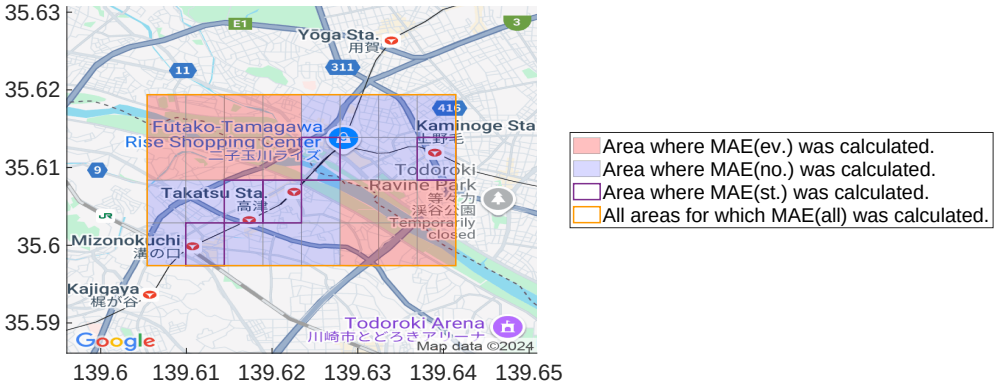


Fig. A1 Target meshed area for the Tamagawa Fireworks Festival and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

Appendix A Evaluated Events and Evaluation Strategy

In this study, we evaluated 12 congested events held in Japan. To ensure the reproducibility of the experiments, we describe the details of each event and the strategies we used to evaluate performance. The 12 events include five of Japan's most famous fireworks displays, the Rugby World Cup 2019, as well as an annual marathon, Comic Market, New Year celebration, and the Halloween Festival. Before conducting the experiment, the location and time of each event were investigated and used in the evaluation described in Section 5.1.4. The detailed information regarding the location and time of each event are displayed as follows.

Tamagawa Fireworks Festival [17]

This event occurred on September 5, 2019, and the fireworks were set off between 6:00 and 7:00 PM. Therefore, the time of the event was set as 4:00 to 8:00 PM. The event venue was the Tama Riverbed area around the fireworks launch and viewing sites. Fig. A1 shows the target meshed area for this event.

Jingu Gaien Fireworks Festival [18]

The event occurred on August 10, 2019, and the fireworks were set off between 7:30 and 8:30 PM. Therefore, the time of the event was set as 4:00 to 8:00 PM. The event venue was the area around the stadium in which fireworks were launched and spectators could watch them. Fig. A2 shows the target meshed area for this event.

[17] http://www.tamagawa-hanabi.com/
[18] http://www.jinguhanabi.com/

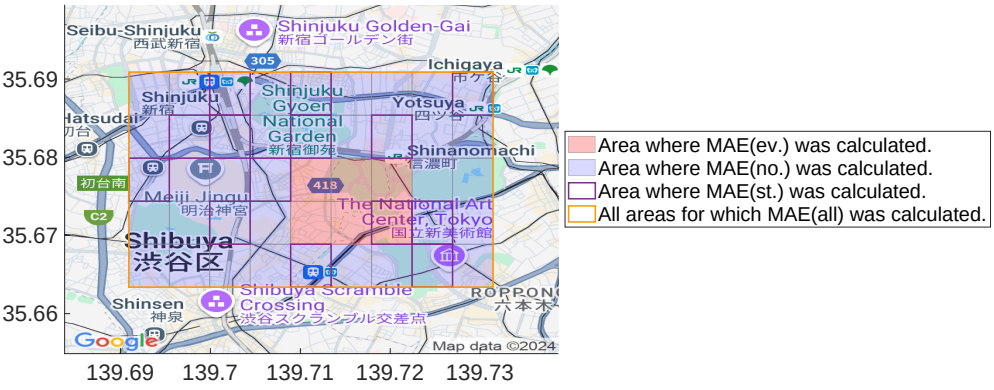


Fig. A2 Target meshed area for the Jingu Gaien Fireworks Festival and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

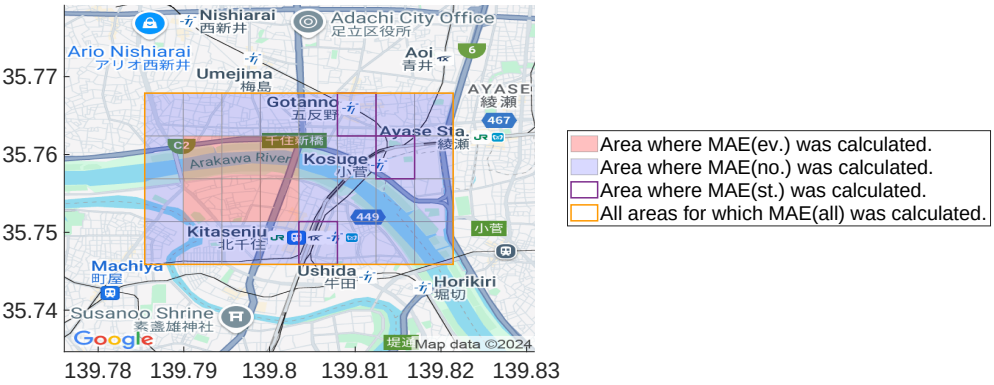


Fig. A3 Target meshed area for the Adachi Fireworks Festival and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

Adachi Fireworks Festival [19]

This event occurred on July 20, 2019, and the fireworks were set off between 7:30 and 8:30 PM. Therefore, the time of the event was set as 6:00 to 9:00 PM. The event venue was the area around the Arakawa riverbed where the fireworks were launched and the viewing area was located. Fig. A3 shows the target meshed area for this event.

Edogawa Fireworks Festival [20]

The event occurred on August 3, 2019, and the fireworks were set off between 7:15 and 8:30 PM. Therefore, the time of the event was set to be 6:00 to 9:00 PM. The

[19] https://www.adachikanko.net/event-hanabi41
[20] https://www.city.edogawa.tokyo.jp/e004/kuseijoho/kohokocho/press/2019/04/0410-1.html

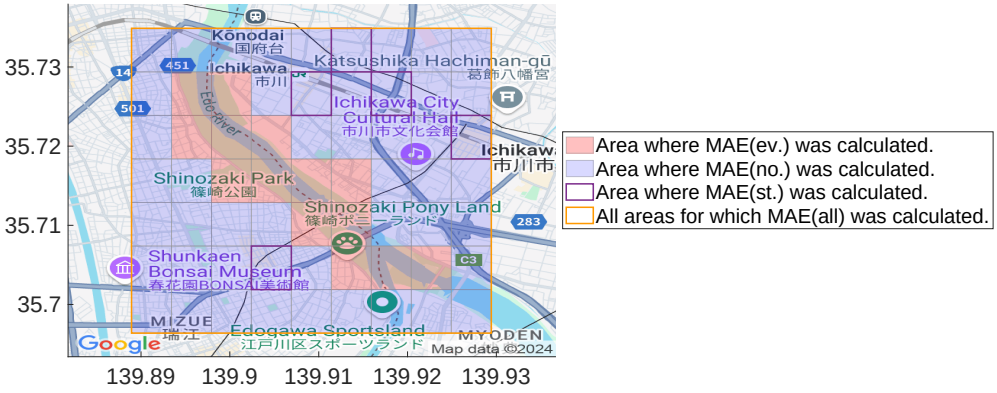


Fig. A4 Target meshed area for the Edogawa Fireworks Festival and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

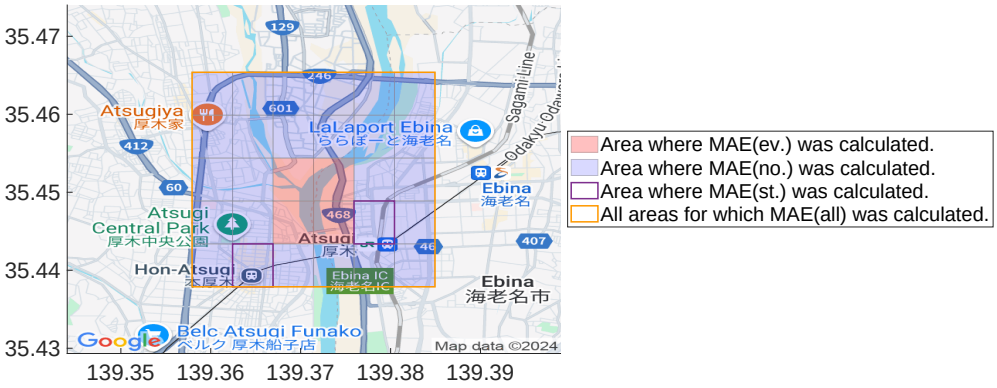


Fig. A5 Target meshed area for the Atsugi Ayu Matsuri Fireworks and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

event venue was the Edogawa riverbed area around the fireworks launch and viewing sites. Fig. A4 shows the target meshed area for this event.

Atsugi Ayu Matsuri Fireworks [21]

This event occurred on August 3, 2019, and the fireworks were set off between 7:00 and 8:30 PM. Therefore, the time of the event was set to be 6:00 to 9:00 PM. The event venue was the area around the Sagami Riverbed, which is the launch site and viewing site of the fireworks. Fig. A5 shows the target meshed area for this event.

[21] https://www.yeg-atsugi.jp/festival/

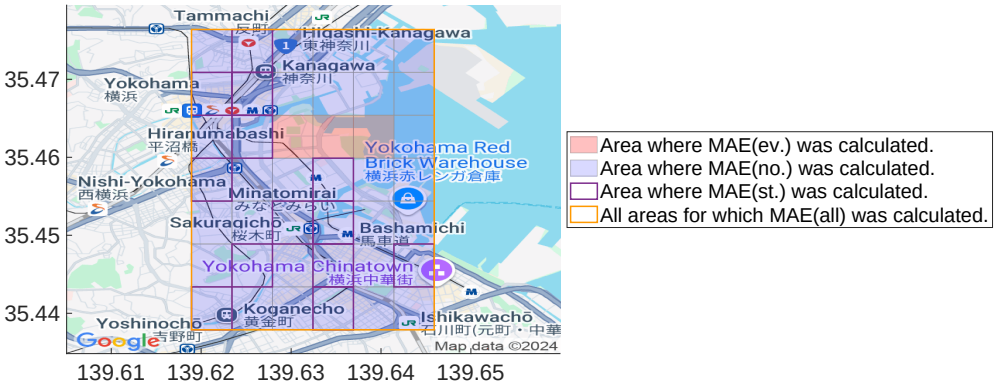


Fig. A6 Target meshed area for the 2019 Yokohama Marathon and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

2019 Yokohama Marathon [22]

The event occurred on November 10, 2019, with a marathon start time of 8:30 AM. The start time of the marathon was 8:30 AM. Therefore, the time of the event was set from 6 to 10 AM. The event venue was the Minato Mirai Bridge area in Yokohama City, which was the starting point of the marathon. Fig. A6 shows the target meshed area for this event.

Rugby W-Cup 3rd Place Playoff / Final [23]

Rugby World Cup was held in Japan in 2019. The third-place match was held on November 1, 2019, at Ajinomoto Stadium in Tokyo, Japan, with the stadium opening at 3:00 and kickoff at 6:00 PM. The stadium opened at 3:00 PM, and the kickoff time was 6:00 PM. Therefore, the event was held from 4:00 to 10:00 PM, and the event venue was around the stadium. The final match was held on November 2, 2019, at Nissan Stadium in Yokohama, Kanagawa Prefecture, Japan, with a kickoff time of 6:00 PM. The event was held from 3:00 to 10:00 PM, and the venue was around the stadium. Fig. A7 and Fig. A8 show the target meshed areas for the 3rd place playoff and the final, respectively.

Comiket 96 [24]

This event was held on Aug 11, 2019, at Tokyo Big Sight in Tokyo, Japan. Because the exhibition was held all day, the event time in this experiment was all day. The event was held around Tokyo Big Sight. Fig. A9 shows the target meshed area for this event.

[22] https://yokohamamarathon.jp/2019/
[23] https://www.rugbyworldcup.com/2019/teams/japan
[24] https://www.comiket.co.jp/info-a/C96/C96info.html

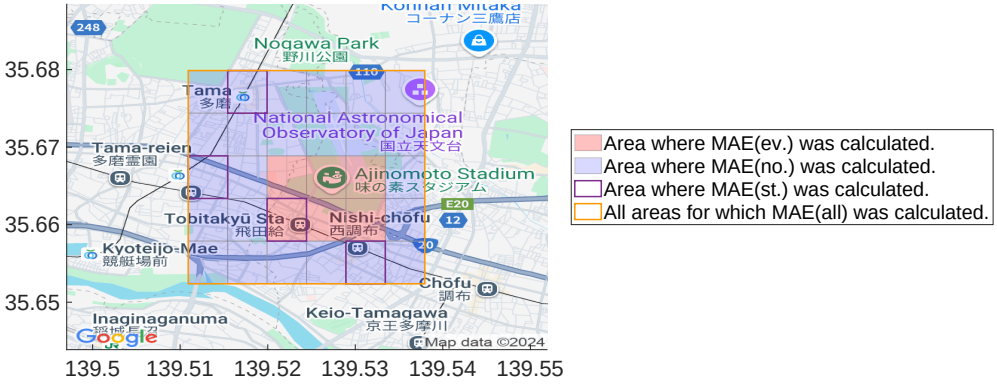


Fig. A7 Target meshed area for the Rugby W-Cup 3rd Place Playoff and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

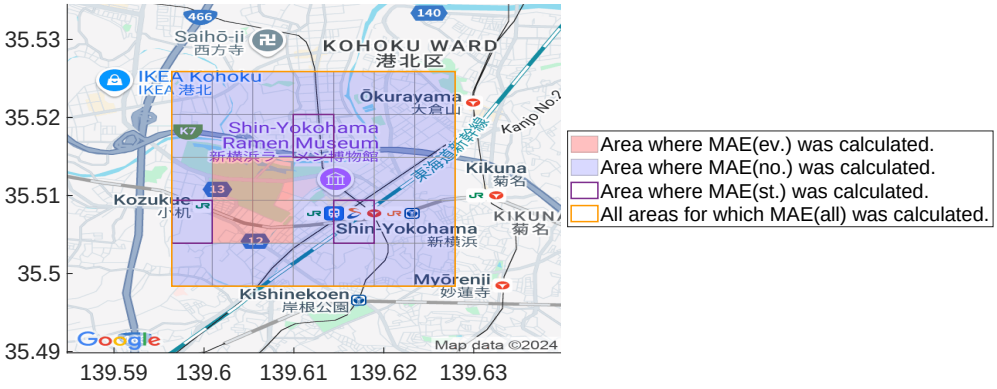


Fig. A8 Target meshed area for the Rugby W-Cup Final and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

New Year Celebration [25]

This event is an annual event in Japan, Hatsumode. In this experiment, we focused on the New Year Celebration at Hie Shrine in Akasaka, Tokyo, which attracts many worshippers in the Tokyo metropolitan area. In this experiment, we focused on the New Year Celebration at Hie Jinja Shrine in Akasaka, Tokyo, which attracts numerous visitors in the Tokyo metropolitan area. The event date was January 1, 2020, the event time was all day, and the event site was the area around Hie Jinja Shrine. Fig. A10 shows the target meshed area for this event.

[25] https://www.hiejinja.net/event/hatsumoude.html

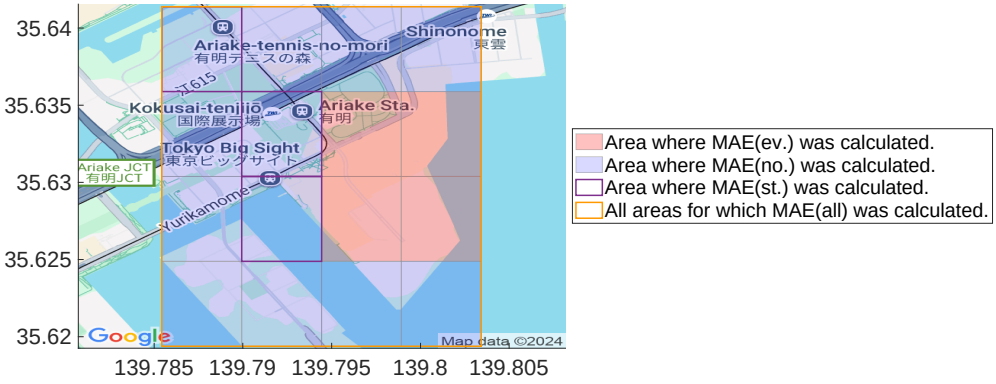


Fig. A9 Target meshed area for the Comiket 96 and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

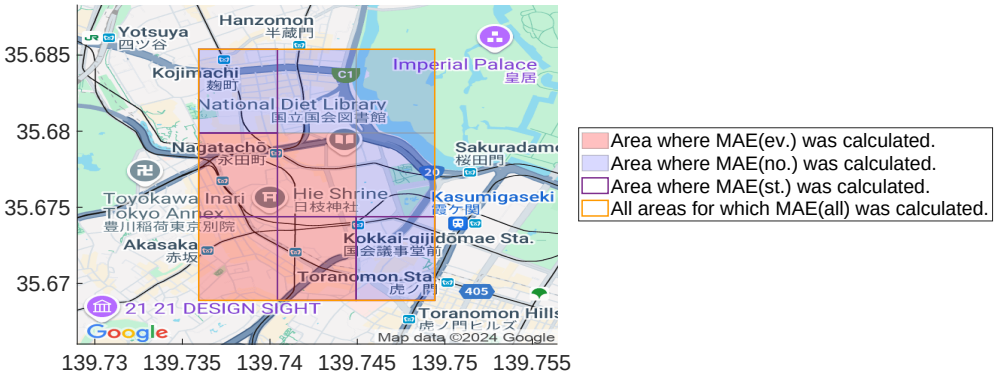


Fig. A10 Target meshed area for the New Year Celebration at Hie Shrine and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

Christmas Market [26]

This event was held during the Christmas season at the Red Brick Warehouse in Yokohama City, Kanagawa Prefecture. The date and time of the event is December 24, 2019, which is the most popular day of the event, the time of the event is from 5:00 to 10:00 PM, and the event site is around the Red Brick Warehouse. Fig. A11 shows the target meshed area for this event.

Halloween Celebration

In this event, many people gather around the Shibuya station in Tokyo to celebrate Halloween. The event date is October 31, 2019, and the time of the event to be

[26] https://www.yokohama-akarenga.jp/brickjournal/detail/61

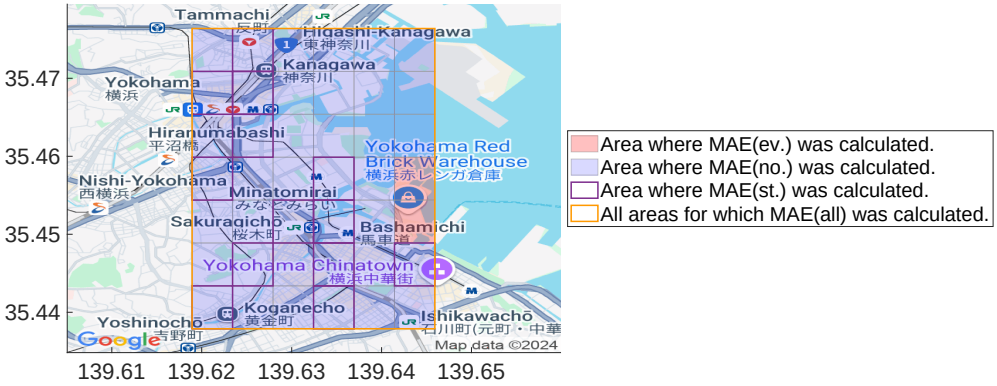


Fig. A11 Target meshed area for the Christmas Market and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

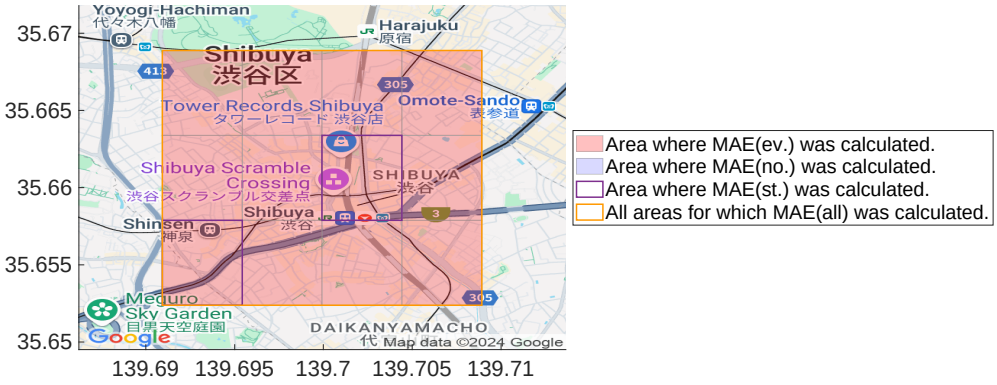
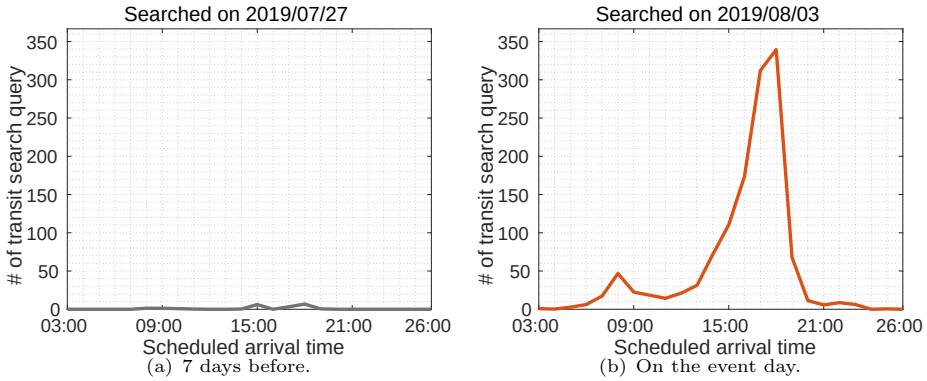


Fig. A12 Target meshed area for the Halloween Celebration around the Shibuya station and the specific meshes used for calculating the evaluation criteria. Background map: ©Google, Map data ©2024 Google. The map has been modified by the authors to highlight overlay additional visualization data.

evaluated is after 5:00 PM, when many people start to gather around the station. Fig. A12 shows the target meshed area for this event.

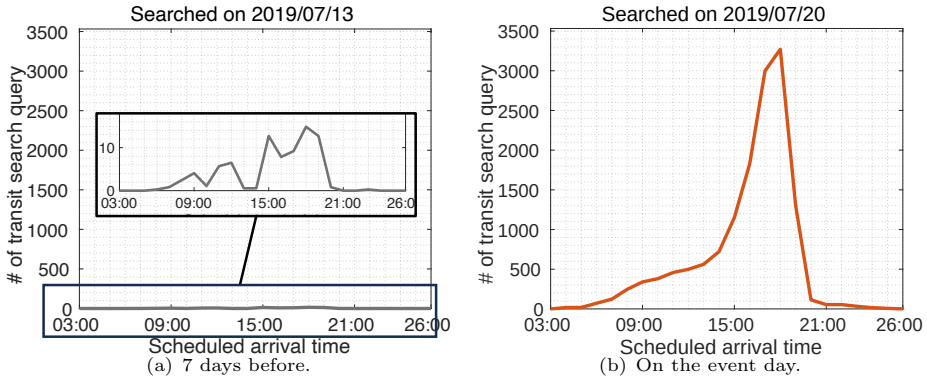
Appendix B Visualization of the Number of Transit Search Logs for Each Event

Here, we present the visualization of the number of transit search logs for each event except the Rugby World Cup Final, as visualized in Fig. 3. Figures from Fig. B13 to Fig. B23 confirm that there was an increase in transit search logs from one week before the event for all the target events.



(a) 7 days before. (b) On the event day.

Fig. B13 Number of transit search logs for attending the Atsugi Ayu Matsuri Fireworks. The logs were searched for Atsugi station as a destination and on August 3, 2019, as an arrival date.



(a) 7 days before. (b) On the event day.

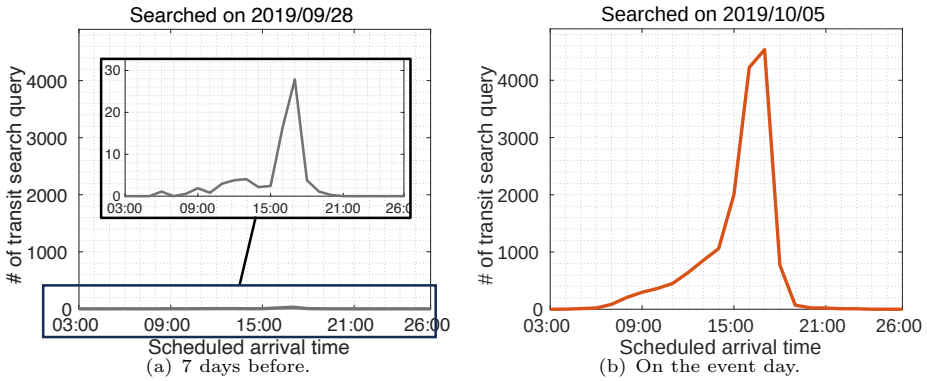
Fig. B14 Number of transit search logs for attending the Adachi Fireworks Festival. The logs were searched for Kita-Senju station as a destination and on July 20, 2019, as an arrival date.

Appendix C Performance Comparison between GCPR and Simulation-based Methods for Each Event

Figures from Fig. C25 to C35 present the event-wise performance comparison between the GCPR and the state-of-the-art simulation-based methods. The results demonstrate that the GCPR outperformed the simulation-based methods in most cases. For exceptional cases, such as the Jingu Gaien Fireworks Festival, refer to the discussion of experimental results in Section 6.1.1.

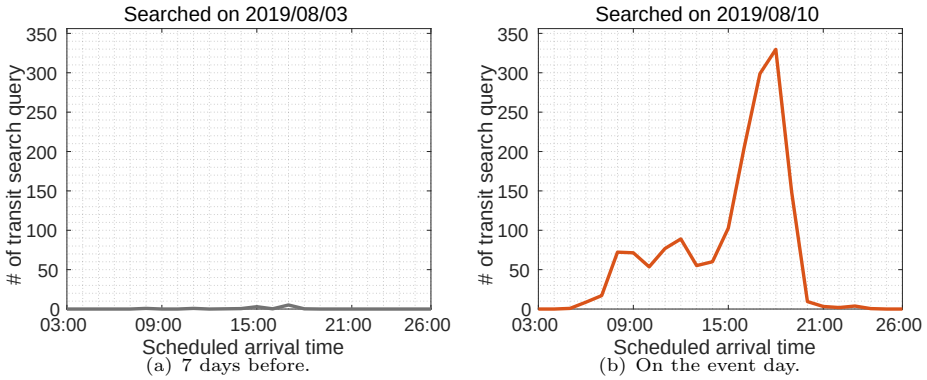
References

Ali A, et al (2021) Mobile crowd sensing based dynamic traffic efficiency framework for urban traffic congestion control. Sustainable Computing: Informatics and Systems



(a) 7 days before. (b) On the event day.

Fig. B15 Number of transit search logs for attending the Tamagawa Fireworks Festival. The logs were searched for Futako-Tamagawa station as a destination and on August 3, 2019, as an arrival date.



(a) 7 days before. (b) On the event day.

Fig. B16 Number of transit search logs for attending the Jingu Gaien Fireworks Festival. The logs were searched for Aoyama-Itchome station as a destination and on August 10, 2019, as an arrival date.

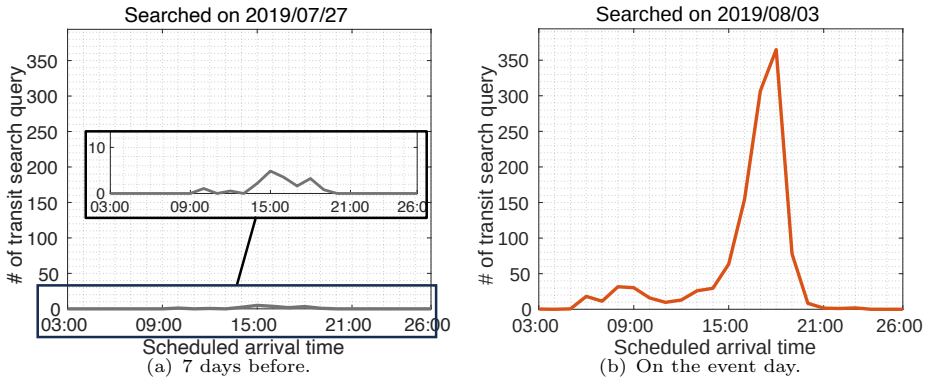
32:100608

Andersen TG, Bollerslev T, Christoffersen P, et al (2005) Volatility forecasting

Anno S, et al (2020) Supervised-CityProphet: Towards accurate anomalous crowd prediction. In: Proc. of SIGSPATIAL

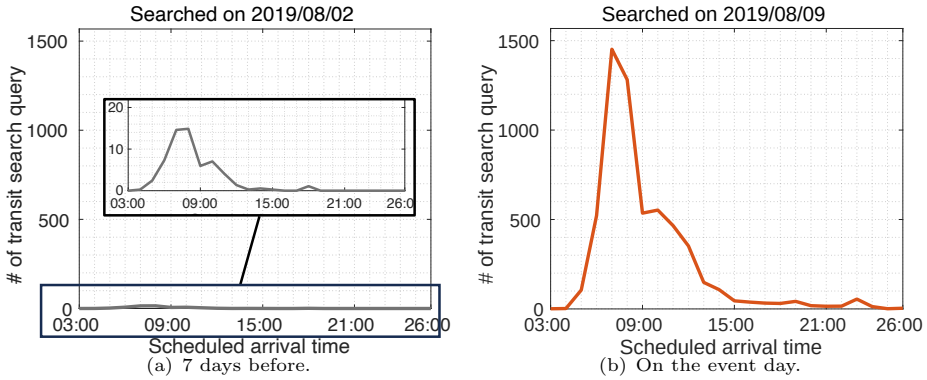
Anno S, et al (2023) CityOutlook+: Early crowd dynamics forecast through unbiased regression with importance-based synthetic oversampling. IEEE Pervasive Computing 22(4):26–34

Bai S, et al (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:180301271



(a) 7 days before. (b) On the event day.

Fig. B17 Number of transit search logs for attending the Edogawa Fireworks Festival. The logs were searched for Edogawa station as a destination and on August 3, 2019, as an arrival date.



(a) 7 days before. (b) On the event day.

Fig. B18 Number of transit search logs for attending the Comic Market 96 at the Tokyo Big Sight. The logs were searched for Kokusai-Tenjijo station as a destination and on August 9, 2019, as an arrival date.

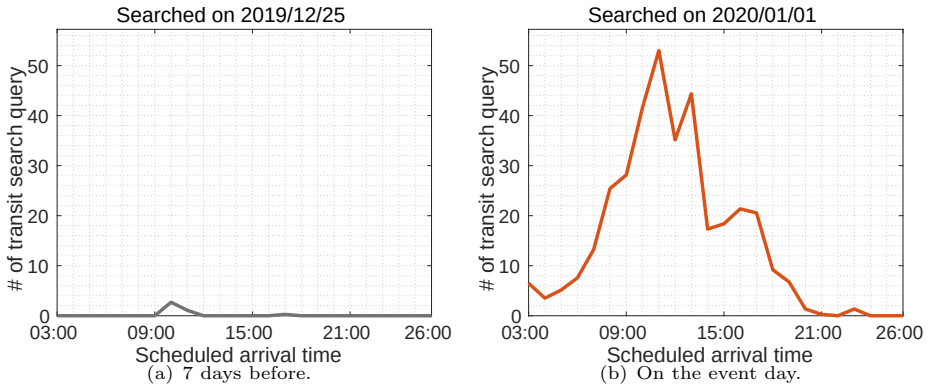
Berry MW, Browne M, Langville AN, et al (2007) Algorithms and applications for approximate nonnegative matrix factorization. Computational statistics & data analysis

Caruana R (1997) Multitask learning. Machine learning

Chen Z, et al (2023) Long sequence time-series forecasting with deep learning: A survey. Information Fusion 97:101819

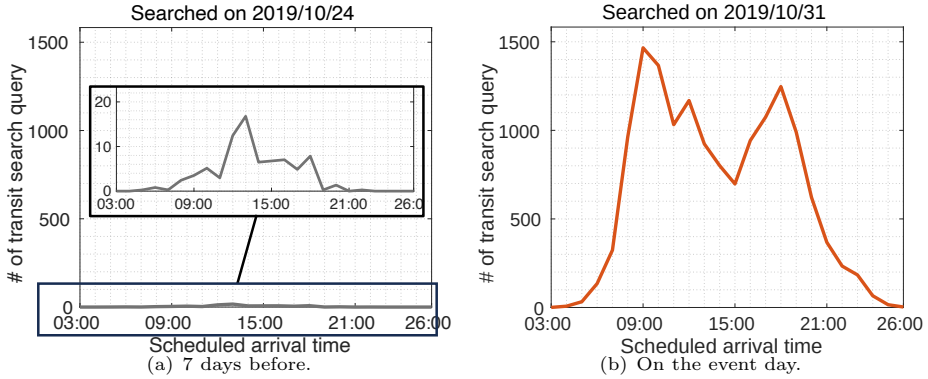
Chung J, et al (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:14123555

Cirstea RG, et al (2022) Triformer: Triangular, variable-specific attentions for long sequence multivariate time series forecasting. In: Proc. of IJCAI



(a) 7 days before. (b) On the event day.

Fig. B19 Number of transit search logs for attending the New Year Celebration at Hie Jinja. The logs were searched for Akasaka-Mitsuke station as a destination and on January 1, 2020, as an arrival date.



(a) 7 days before. (b) On the event day.

Fig. B20 Number of transit search logs for attending the Halloween Celebration around the Shibuya station. The logs were searched for Shibuya station as a destination and on October 31, 2019, as an arrival date.

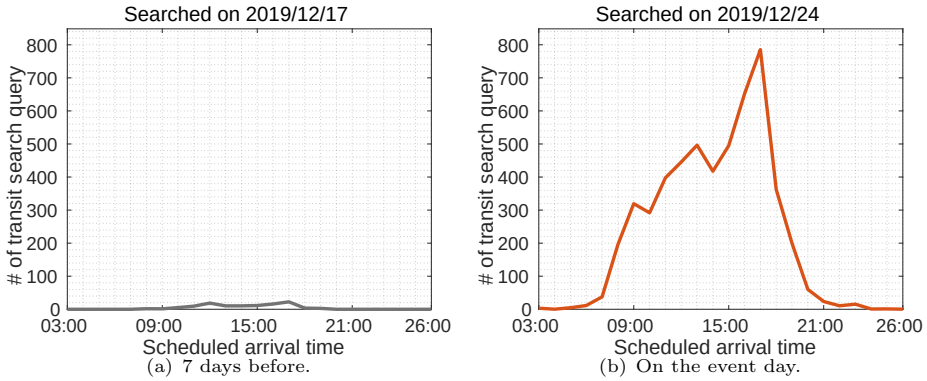
Combinido JSL, Lim MT (2012) Crowding effects in vehicular traffic. PLoS one 7(11):e48151

Defferrard M, et al (2016) Convolutional neural networks on graphs with fast localized spectral filtering. Advances in neural information processing systems 29

Du R, et al (2014) Predicting activity attendance in event-based social networks: Content, context and social influence. In: Proc. of UbiComp

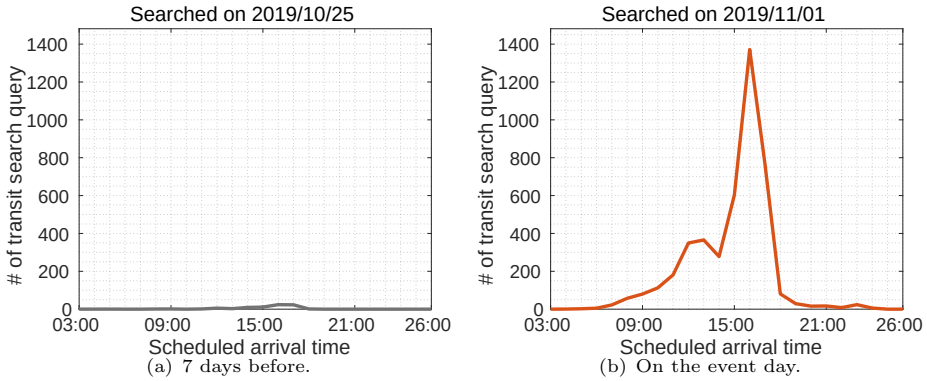
Du S, et al (2019) Deep air quality forecasting using hybrid deep learning framework. IEEE Transactions on Knowledge and Data Engineering 33(6):2412–2424

Fan Z, Song X, Shibasaki R (2014) Cityspectrum: A non-negative tensor factorization approach. In: Proc. of UbiComp



(a) 7 days before. (b) On the event day.

Fig. B21 Number of transit search logs for attending the Christmas Market at the Yokohama Red Brick Warehouse. The logs were searched for Minatomirai station as a destination and on December 24, 2019, as an arrival date.



(a) 7 days before. (b) On the event day.

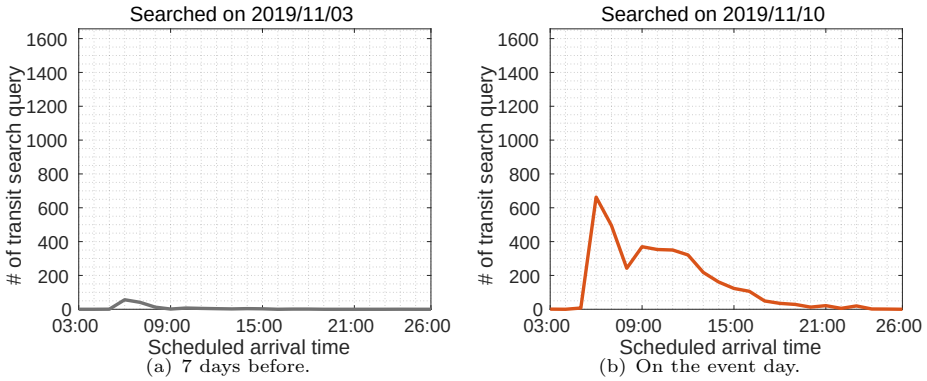
Fig. B22 Number of transit search logs for attending the Rugby World Cup 3rd Place Playoff at Ajinomoto Stadium. The logs were searched for Tobitakyu station as a destination and on November 1, 2019, as an arrival date.

Fan Z, et al (2015) Citymomentum: An online approach for crowd behavior prediction at a citywide level. In: Proc. of UbiComp

Fedujwar R, Agarwal A (2024) A systematic review on crowding valuation in public transport. Public Transport pp 1–31

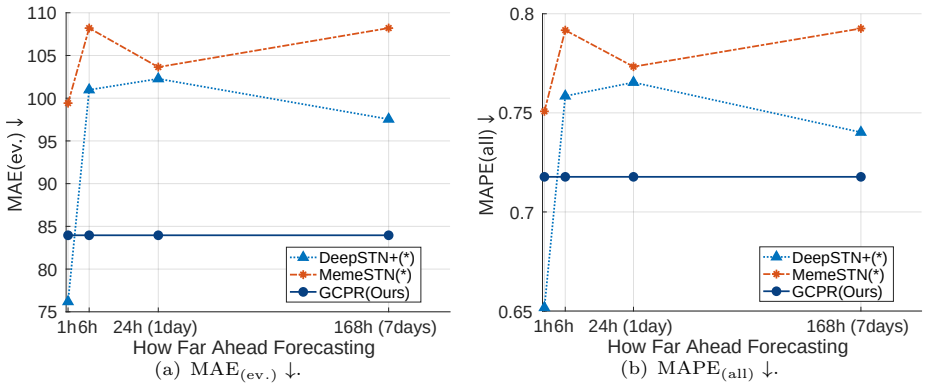
Feliciani C, et al (2022) Crowd control methods: Established and future practices. In: Introduction to Crowd Management: Managing Crowds in the Digital Era: Theory and Practice. Springer, p 167–216

Hayakawa Y, et al (2021) Simultaneous multiple poi population pattern analysis system with hdp mixture regression. In: Proc. of PAKDD



(a) 7 days before. (b) On the event day.

Fig. B23 Number of transit search logs for attending the 2019 Yokohama Marathon. The logs were searched for Minatomirai station as a destination and on November 10, 2019, as an arrival date.



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

Fig. C24 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Atsugi Ayu Matsuri Fireworks.

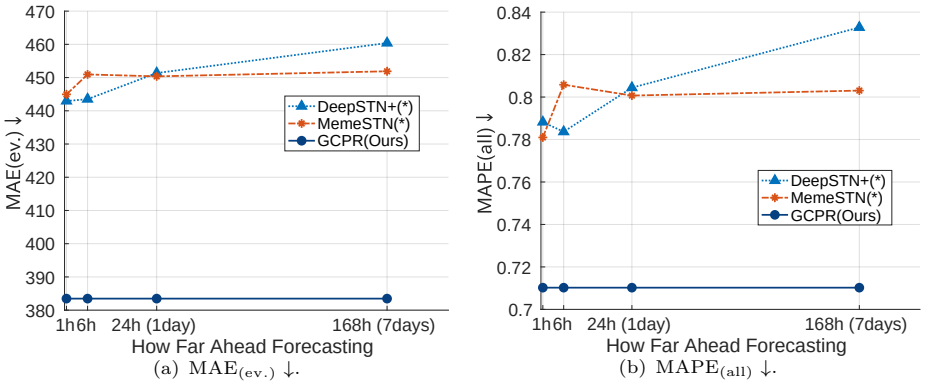
Hoang MX, et al (2016) FCCF: forecasting citywide crowd flows based on big data. In: Proc. of SIGSPATIAL

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural computation

Hörcher D, et al (2017) Crowding cost estimation with large scale smart card and vehicle location data. Transportation Research Part B: Methodological 95:105–125

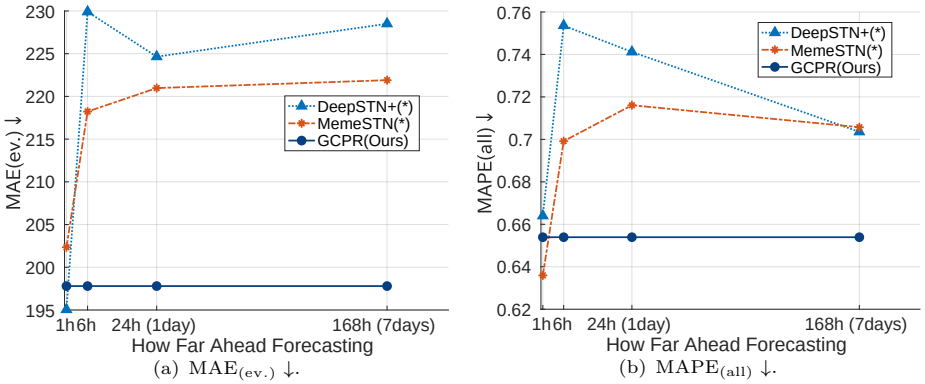
Jiang R, et al (2018) DeepUrbanMomentum: An online deep-learning system for short-term urban mobility prediction. In: Proc. of AAAI

Jiang R, et al (2019) DeepUrbanEvent: A system for predicting citywide crowd dynamics at big events. In: Proc. of SIGKDD



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

Fig. C25 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Adachi Fireworks Festival.



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

Fig. C26 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Tamagawa Fireworks Festival.

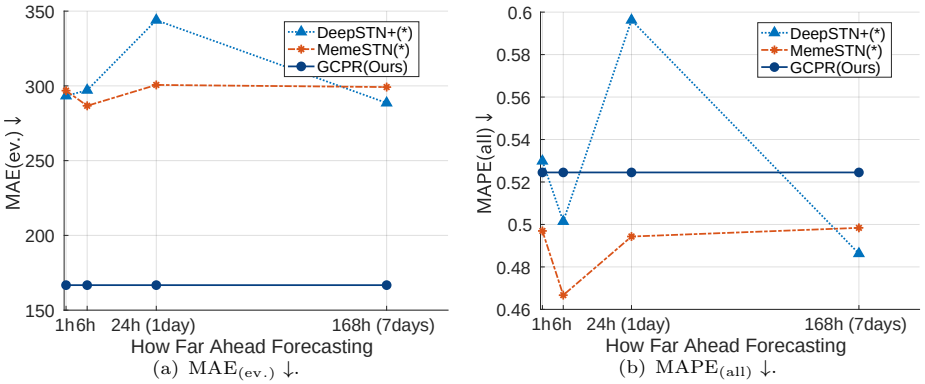
Jiang R, et al (2021) DL-Traff: Survey and benchmark of deep learning models for urban traffic prediction. In: Proc. of CIKM

Jiang R, et al (2023) Learning social meta-knowledge for nowcasting human mobility in disaster. In: Proc. of TheWebConf

Kim KM, Hong SP, Ko SJ, et al (2015) Does crowding affect the path choice of metro passengers? Transportation Research Part A: Policy and Practice 77:292–304

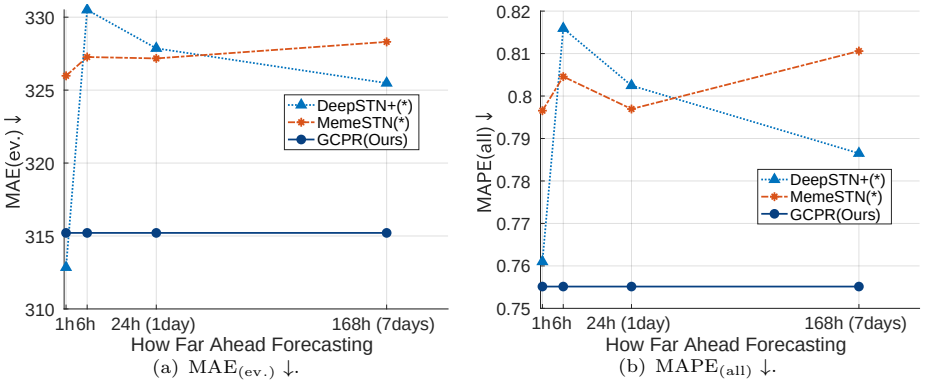
Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:160902907

Konishi T, et al (2016) CityProphet: City-scale irregularity prediction using transit app logs. In: Proc. of UbiComp



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

Fig. C27 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Jingu Gaien Fireworks Festival.



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

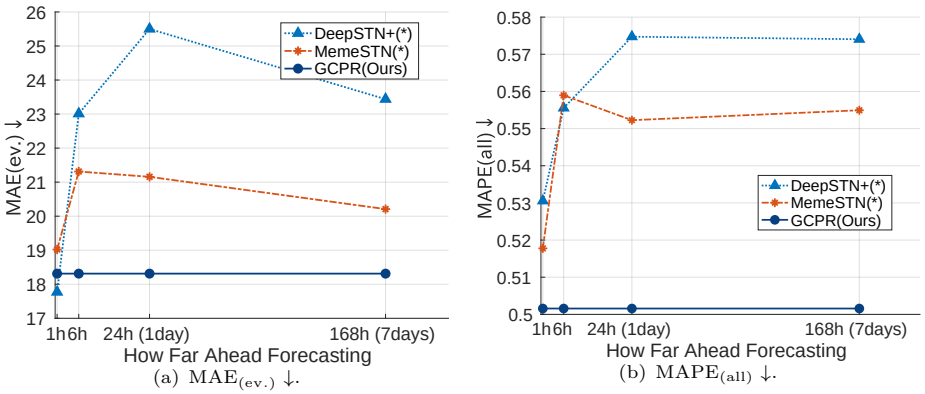
Fig. C28 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Edogawa Fireworks Festival.

Krizhevsky A, et al (2012) ImageNet classification with deep convolutional neural networks. Proc in NIPS 25

Lam WH, Cheung CY, Lam C (1999) A study of crowding effects at the hong kong light rail transit stations. Transportation Research Part A: Policy and Practice 33(5):401–415

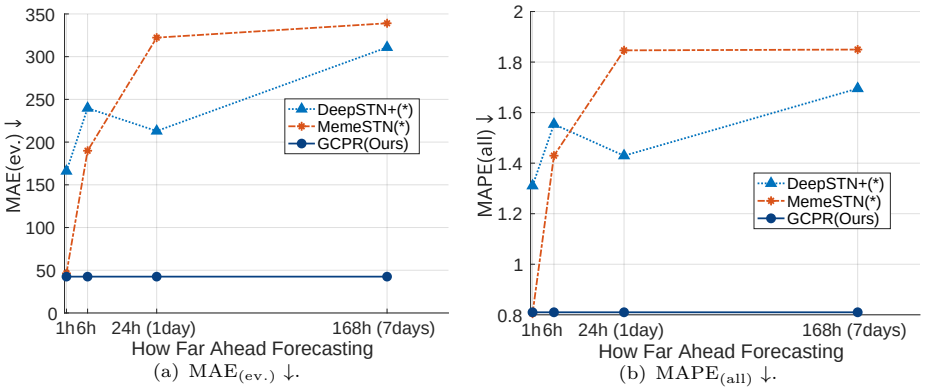
Li Y, Moura JM (2020) Forecaster: A graph transformer for forecasting spatial and time-dependent data. In: ECAI 2020. IOS Press, p 1293–1300

Li Y, et al (2024) Learning from polar representation: An extreme-adaptive model for long-term time series forecasting. In: Proc. of AAAI



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

Fig. C29 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Comic Market 96 at the Tokyo Big Sight.



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

Fig. C30 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the New Year Celebration at Hie Jinja.

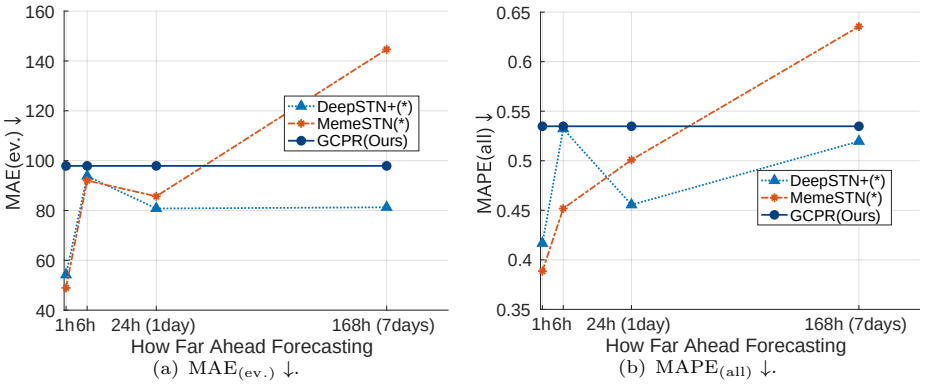
Li Z, et al (2021) A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems 33(12):6999–7019

Lika B, Kolomvatsos K, Hadjiefthymiades S (2014) Facing the cold start problem in recommender systems. Expert Systems with Applications

Lim B, Zohren S (2021) Time-series forecasting with deep learning: a survey. Philosophical Transactions of the Royal Society A 379(2194):20200209

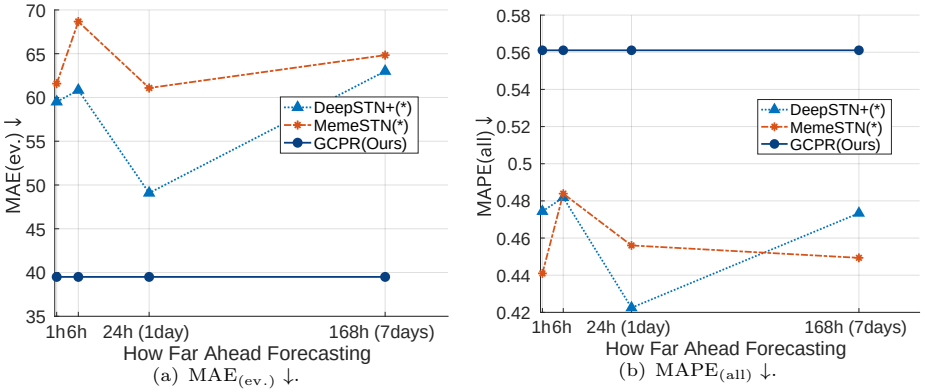
Lin Z, et al (2019) DeepSTN+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis. In: Proc. of AAAI

Liu DC, Nocedal J (1989) On the limited memory bfgs method for large scale optimization. Mathematical programming



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

Fig. C31 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Halloween Celebration around Shibuya station.



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

Fig. C32 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Christmas Market at the Yokohama Red Brick Warehouse.

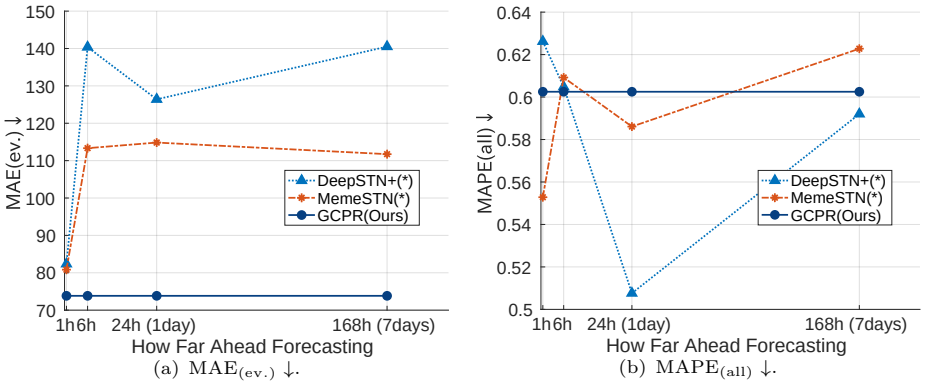
Liu Y, et al (2022) Multivariate time-series forecasting with temporal polynomial graph neural networks. Advances in neural information processing systems 35:19414–19426

Luca M, et al (2021a) A survey on deep learning for human mobility. ACM Computing Surveys (CSUR) 55(1):1–44

Luca M, et al (2021b) A survey on deep learning for human mobility. ACM CSUR 55(1)

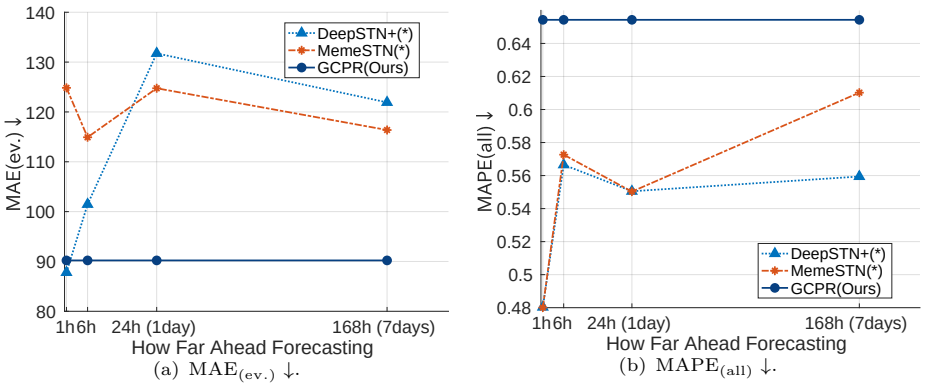
Mehmood U, et al (2020) Event attendance prediction using social media. In: Proc. of ACSW

Minegishi Y (2023) Tracer observation of egress and way-home crowd behavior at stadiums: From the perspective of crowd control in emergency evacuations. Japan



(a) $\text{MAE}_{(\text{ev.})} \downarrow$. (b) $\text{MAPE}_{(\text{all})} \downarrow$.

Fig. C33 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Rugby World Cup 3rd Place Playoff at Ajinomoto Stadium.



(a) $\text{MAE}_{(\text{ev.})} \downarrow$. (b) $\text{MAPE}_{(\text{all})} \downarrow$.

Fig. C34 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the Rugby World Cup Final at Nissan Stadium.

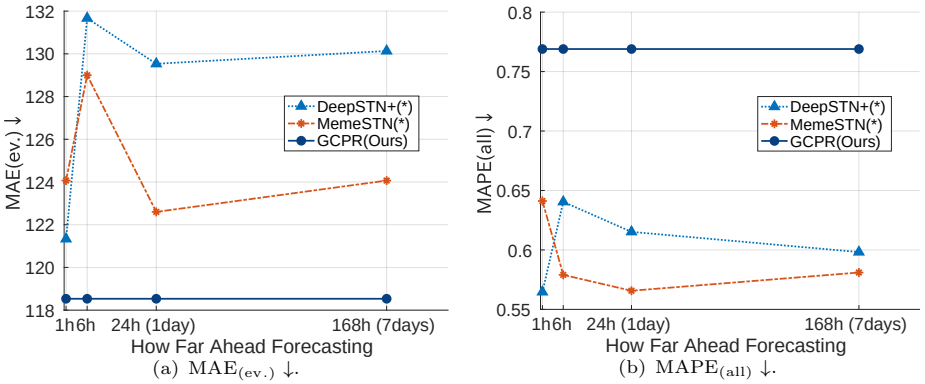
Architectural Review 6(1):e12358

Mudelsee M (2019) Trend analysis of climate time series: A review of methods. Earth-science reviews 190:310–322

Nishi K, et al (2014) Extracting land-use patterns using location data from smartphones. In: Proc. of the First International Conference on IoT in Urban Space

Pan Z, Wang Z, Wang W, et al (2019) Matrix factorization for spatio-temporal neural networks with applications to urban flow prediction. In: Proc. of CIKM

Sener O, Koltun V (2018) Multi-task learning as multi-objective optimization. Advances in neural information processing systems



(a) $\mathrm{MAE}_{(\mathrm{ev.})} \downarrow$. (b) $\mathrm{MAPE}_{(\mathrm{all})} \downarrow$.

Fig. C35 Performance comparison between GCPR and state-of-the-art simulation-based crowding forecast methods in the 2019 Yokohama Marathon.

Shimosaka M, et al (2015) Forecasting urban dynamics with mobility logs by bilinear poisson regression. In: Proc. of UbiComp

Shimosaka M, et al (2016) Coupled hierarchical dirichlet process mixtures for simultaneous clustering and topic modeling. In: Proc. of ECML-PKDD

Shimosaka M, et al (2019) Spatiality preservable factored poisson regression for large-scale fine-grained gps-based population analysis. In: Proc. of AAAI

Stoffer DS, Ombao H (2012) Special issue on time series analysis in the biological sciences

Takeuchi K, et al (2013) Non-negative multiple tensor factorization. In: Proc. of ICDM

Tran D, et al (2018) A closer look at spatiotemporal convolutions for action recognition. In: Proc. of CVPR

Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008

Wilson T, et al (2022) DeepGPD: A deep learning approach for modeling geospatio-temporal extreme events. In: Proc. of AAAI

Wright S, Nocedal J (1999) Numerical optimization. Springer Science

Wu H, et al (2021) Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in neural information processing systems 34:22419–22430

Wu X, et al (2018) Who will attend this event together? event attendance prediction via deep lstm networks. In: Proc. of SIAM SDM

Wu Z, et al (2019) Graph wavenet for deep spatial-temporal graph modeling. In: Proc. of IJCAI

Wu Z, et al (2020a) A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems 32(1):4–24

Wu Z, et al (2020b) Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proc. of SIGKDD

Xia T, Li Y (2019) Revealing urban dynamics by learning online and offline behaviours together. Proc of IMWUT

Xia T, et al (2019) Understanding urban dynamics via state-sharing hidden markov model. In: Proc. of WWW

Zhang F, Yuan NJ, Wilkie D, et al (2015a) Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. ACM Trans Intell Syst Technol

Zhang J, Zheng Y, Qi D (2017) Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proc. of AAAI

Zhang J, et al (2016) DNN-based prediction model for spatio-temporal data. In: Proc. of SIGSPATIAL

Zhang X, et al (2015b) Who will attend?–predicting event attendance in event-based social network. In: Proc. of MDM, IEEE

Zhang Y, Hara T (2023) Organized event participant prediction enhanced by social media retweeting data. In: Proc. of WI-IAT

Zhao L, et al (2015) Spatiotemporal event forecasting in social media. In: Proc. of SIAM SDM

Zheng Y, et al (2014a) Diagnosing new york city's noises with ubiquitous data. In: Proc. of UbiComp

Zheng Y, et al (2014b) Urban computing: Concepts, methodologies, and applications. ACM Trans on Intelligent Systems and Technology

Zhou H, et al (2021) Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proc. of AAAI

Zhou J, et al (2018) Early warning of human crowds based on query data from baidu maps: Analysis based on shanghai stampede. In: Big data support of urban planning and management. Springer