Transportation Research Part A 190 (2024) 104242

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Situational-aware multi-graph convolutional recurrent network (SA-MGCRN) for travel demand forecasting during wildfires

Xiaojian Zhang [a], Xilei Zhao [a], Yiming Xu [a,b], Daniel Nilsson [c], Ruggiero Lovreglio [d,*]

[a] Department of Civil and Coastal Engineering, University of Florida, United States of America
[b] School of Architecture, The University of Texas at Austin, United States of America
[c] Department of Civil and Natural Resources Engineering, University of Canterbury, New Zealand
[d] School of Built Environment, Massey University, New Zealand

ARTICLE INFO

Keywords:
AI
Wildfire evacuation
GPS data
Travel demand forecasting
Real-time

ABSTRACT

Natural hazards, such as wildfires, pose a significant threat to communities worldwide. Real-time forecasting of travel demand during wildfire evacuations is crucial for emergency managers and transportation planners to make timely and better-informed decisions. However, few studies focus on accurate travel demand forecasting in large-scale emergency evacuations. To tackle this research gap, the study develops a new methodological framework for modeling highly granular spatiotemporal trip generation in wildfire evacuations by using (a) large-scale GPS data generated by mobile devices and (b) state-of-the-art AI technologies. Based on the travel demand inferred from the GPS data, we develop a new deep learning model, i.e., Situational-Aware Multi-Graph Convolutional Recurrent Network (SA-MGCRN), along with a model updating scheme to achieve real-time forecasting of travel demand during wildfire evacuations. The proposed methodological framework is tested using a real-world case study: the 2019 Kincade Fire in Sonoma County, CA. The results show that SA-MGCRN significantly outperforms all the selected state-of-the-art benchmarks in terms of prediction performance. Our finding suggests that the most important model components of SA-MGCRN are weekend indicator, population change, evacuation order/warning information, and proximity to fire, which are consistent with behavioral theories and empirical findings. SA-MGCRN can be directly used in future wildfire events to assist real-time decision-making and emergency management.

1. Introduction

Wildfires are posing a growing threat to communities across the U.S. and worldwide (Haghani et al., 2022). Studies have found that the consequences of wildfires have been aggravated in recent years, mainly due to accelerating climate change risks and rapid expansion of the wildland-urban interface (WUI) (Kuligowski et al., 2020; Zhao et al., 2022; Radeloff et al., 2018). For example, in recent years, Sonoma County in California experienced multiple large wildfires (e.g., the 2017 Tubbs Fire, the 2019 Kincade Fire, and the 2020 Glass Fire), which forced thousands of people to evacuate nearly every year. Wildfires are also becoming a threat in winter season as demonstrated by the 2021 Marshall Fire. In terms of structures lost, the Marshall Fire (December 30, 2021 – January 1, 2022) is the most destructive fire in Colorado history (Thy Vo and Prentzel, 2022). To reduce the risk of wildfires and enhance the resilience of WUI communities, we need to improve real-time decision support for emergency managers by accurately forecasting the *travel demand* (including evacuation trips and other types of trips) during wildfire evacuations.

* Correspondence to: School of Built Environment, Massey University, East Precinct Albany Expressway, SH17, Albany, Auckland 0632, New Zealand.
E-mail addresses: xiaojianzhang@ufl.edu (X. Zhang), xilei.zhao@essie.ufl.edu (X. Zhao), yiming.xu@utexas.edu (Y. Xu), daniel.nilsson@canterbury.ac.nz (D. Nilsson), r.lovreglio@massey.ac.nz (R. Lovreglio).

https://doi.org/10.1016/j.tra.2024.104242
Received 15 November 2023; Received in revised form 9 July 2024; Accepted 31 August 2024
Available online 10 September 2024

Significant research gaps exist in terms of travel demand forecasting in large-scale emergency evacuations. To date, evacuation demand (i.e., time-dependent evacuation trip generation) is generally estimated by using survey data and logistic regression (Murray-Tuite and Wolshon, 2013), both of which have some limitations. Specifically, despite generating valuable insights on household evacuation behavior, survey data have relatively small sample sizes (e.g., only hundreds of data points) and provide a low-resolution timeline (e.g. 2–6 h resolution) of household decisions during the evacuation (Zhao et al., 2022; Lovreglio et al., 2020; Fu et al., 2007). Based on the survey data, logistic regression is commonly used to model the binary evacuation decision (to evacuate or not) of households (Murray-Tuite and Wolshon, 2013; Kuligowski et al., 2020), but logistic regression cannot capture the potential nonlinearities and interactions between the evacuation decision and independent variables and thus often lead to inaccurate predictions (Zhao et al., 2020a). In addition to trips for evacuation purposes, other types of trips, such as background trips and intermediate trips, play essential roles in measuring transportation network performance and estimating travel time (Murray-Tuite and Wolshon, 2013; Zeigler et al., 1981). However, nearly no prior work has estimated the total amount of trips generated during evacuations (McGhee et al., 2006; Murray-Tuite and Wolshon, 2013).

To tackle these research gaps, we aim to develop a new real-time trip generation model including evacuation trips and other types of trips for wildfire evacuation based on GPS data and AI technologies. GPS data typically contain millions of location signals from mobile devices such as smartphones and smartwatches. As such, GPS data can provide large-scale highly-granular spatiotemporal trajectories of people's movements during an evacuation (Zhao et al., 2022). More importantly, recent technological advancements in GPS tracking and super-computing services have made real-time delivery of GPS data possible (Datarade, 2023). Generally, GPS data are generated in real-time on mobile devices, uploaded to a server, and subsequently passed through certain processing algorithms. While there might be some delay in this process, it is generally not substantial, especially for high-frequency data (Smart Traffic, 2021). This allows the end-users to get the (near) real-time feed from the GPS data providers and conduct timely research (e.g., daily updated human mobility patterns during Covid-19 pandemic (Xiong et al., 2020; Hu et al., 2021)). To date, GPS data has shown great potential for estimating and understanding evacuation behavior for different types of disasters (Horanont et al., 2013; Yabe and Ukkusuri, 2020; Zhao et al., 2022). We believe with the GPS data, hourly trip generation at the zonal level can be inferred, which can then be used to model travel demand.

With the inferred trips, we then develop a new deep learning model, i.e., **S**ituational-**A**ware **M**ultiple **G**raph **C**onvolutional **R**ecurrent **N**etwork (**SA-MGCRN**), along with a model updating scheme to achieve real-time forecasting of travel demand during wildfire evacuations. The proposed model uses a Graph Convolutional Network (GCN) based on environmental similarity graph, and demographic similarity graph to attach spatial dependency to temporal inputs (e.g., historical travel demand). The output of GCN is subsequently processed with the weather information, evacuation order/warning, fire progression, weekend indicator (i.e., a binary variable indicating if it is Sat./Sun.) and historical travel demand information by the Gated Recurrent Unit (GRU) to capture temporal dependency. Then, a fully-connected layer is used to generate the final prediction for the hourly trip generation at the zonal (i.e., census tract) level. To account for the delivery delay issue (which is common among GPS providers) of GPS data, we propose a new model updating scheme to ensure the situational awareness of the AI model. The proposed model and the model updating scheme are empirically evaluated by a real-world case study of Kincade Fire, Sonoma County, CA, where we employ a large-scale GPS dataset. An approach for inferring trip generations from GPS data and a corresponding validation method are also developed. The results of this study can be directly used by emergency managers to support real-time decision-making in wildfire evacuations.

The remaining paper is structured as follows: Section 2 reviews the related prior work. Section 3 introduces the methodology that includes three major parts, i.e., trip generation inference, travel demand forecasting, and model updating scheme. Section 4 describes the case study and Section 5 presents the results of the case study. Section 6 discusses the key findings of the research and concludes the paper.

2. Literature review

In this section we provide an overview of existing AI applications for evacuation modeling (Section 2.1) and the key actors affecting evacuation wildfire decision-making (Section 2.2).

2.1. AI applications in evacuation modeling

Machine learning techniques provide researchers powerful tools to model human behaviors in both normal conditions (Zhao et al., 2020b; Xu et al., 2021; Zhang et al., 2024a,c) and emergencies such as wildfires, hurricanes and earthquakes (Zhao et al., 2020a; Bergado et al., 2021; Zhao et al., 2021a,b; Wang et al., 2019a; Lo et al., 2009; Zhang et al., 2024b; Sun et al., 2024a). When the target variable is properly specified and the parameters are appropriately set, machine learning methods usually produce much more accurate predictions compared to the traditional statistical models (Sun et al., 2024b). For example, Zhang et al. (2024b) showed that machine learning model significantly outperforms traditional statistical models in terms of prediction accuracy in determining evacuation decisions during earthquakes. This advantage arises from their flexible model structures, which can effectively capture the complex underlying relationships behind the data (Zhao et al., 2020a,b; Xu et al., 2021). Also, certain explainable machine learning models such as tree-structured models can also offer more powerful interpretability than traditional statistical models, since they are able to identify nonlinear relationships (Zhang et al., 2024b; Sun et al., 2024a; Zhao et al., 2020a). For example, Zhao et al. (2020a) applied random forest (a tree-based machine learning model) to model individuals' pre-evacuation decisions in building fires. The results showed strong nonlinear relationships between key factors (such as the time elapsed after the

alarm has started and the decision maker's personal group size) and building occupants' decisions which were not identified with traditional statistics in previous studies (Lovreglio et al., 2015). More recently, Zhao et al. (2021a) has applied random forest to model and explain individual-level evacuation decisions for wildfires and the random forest model showed superior results than the traditional statistical model. However, we should note that this interpretation strength is usually dependent on the specific model architecture (such as tree structures) and not usually generalizable to all machine learning methods (Molnar, 2020; Sun et al., 2024b).

Recently, deep learning has gained wide popularity in emergency management studies such as traffic speed and volume forecasting during emergency events, as it can produce highly accurate real-time predictions (Rahman and Hasan, 2018; Roy et al., 2021; Huang et al., 2021). For example, Rahman and Hasan (2023) developed a dynamic graph convolutional long short-term memory neural network (DGCN-LSTM) model to predict traffic speeds in freeways during Hurricane Irma's evacuation. Roy et al. (2021) used an LSTM model to forecast the traffic volumes of interstate highways during Hurricanes Matthew and Irma evacuation. Nguyen et al. (2019) developed a deep learning model that consists of two LSTM models to forecasting people's needs during Hurricanes Sandy, Harvey, and Irma using social media and weather data. Li et al. (2022) adopted an LSTM model and a Bidirectional Encoder Representations from Transformers (BERT) model to identify people's evacuation intent during Hurricane Irma using social media data. However, nearly no prior work has utilized deep learning to forecast real-time travel demand during wildfire evacuation, probably due to the lack of appropriate data and/or suitable design of deep learning architectures.

2.2. Factors associated with evacuation decision-making

The decision-making process to evacuate or stay during a wildfire has been widely investigated in the literature using different qualitative and quantitative approaches, such as individual interviews, focus group interviews, questionnaires/surveys and GPS data inference (Kuligowski, 2021; Wu et al., 2022). These methodologies have been applied to study numerous wildfires around the globe, uncovering a range of factors that influence the decision-making process for homeowners' protective actions (Lovreglio et al., 2019, 2020; Kuligowski et al., 2022; Vaiciulyte et al., 2022). These contributing factors include both internal factors (e.g., demographics, perspectives, memories of previous experiences and attitudes) and external factors (e.g., disaster alerts and warnings as well as cues from the physical and social environment) (Lovreglio et al., 2019). Recent studies have synthesized these factors into conceptual models, particularly based on the Protective Action Decision Model framework (Lindell and Perry, 2012), to elucidate their impact on different phases of the decision-making process (Folk et al., 2019; Lovreglio et al., 2019, 2020; Kuligowski et al., 2022).

While focusing on external factors, several researchers have reported that the type and number of warnings and cues could prompt the householders' decision to evacuate. For instance, the presence of mandatory or voluntary evacuation orders has been associated with a higher likelihood of evacuation (Kuligowski et al., 2022; Lovreglio et al., 2019; McLennan et al., 2011; Mozumder et al., 2008; Strawderman et al., 2012; Whittaker et al., 2016; Wong et al., 2022), especially when issued by sources that the public trusts (Kuligowski et al., 2020; Templeton et al., 2023). Furthermore, visible smoke, flames, and information regarding the firefront location also triggered more householders to evacuate (McCaffrey et al., 2018; McLennan et al., 2012). However, Lovreglio et al. (2020) did not find significant evidence that the distance of the fire has a significant impact on the evacuation decision. This assertion may stem from data reliability issues, as the study relied on postal codes rather than exact addresses to estimate the distance of the fire from homes. Moreover, characteristics of the built environment, such as being located in a high fire risk zone and the size of property parcels, have also been identified as influential factors in evacuation decisions (Wu et al., 2022). Overall, these sets of empirical evidence inform us to incorporate external factors such as evacuation order/warning and fire proximity as key components in deep learning model design.

Internal factors also have shown strong correlations to the decisions to evacuate or stay. Demographic factors (such as gender, age, racial composition and income) have been identified as key determinants of evacuation decisions (Kuligowski et al., 2020; Paveglio et al., 2014; Whittaker et al., 2016; Wong et al., 2022; Wu et al., 2022). For example, Wu et al. (2022) found that census block groups with a higher presence of bachelor's degree (or above) holders are more likely to have a higher evacuation rate. Similarly, Paveglio et al. (2014) found a propensity for households with higher incomes to opt for evacuation. When focusing on the zone-level (e.g., census tract) analysis, these demographic factors could be readily assessable through official reports such as the American Community Survey (ACS). Therefore, they can be effectively utilized in the development of deep learning models (e.g., embedded in a graph) for demand prediction and planning purposes (Xu et al., 2023). Previous experience and attitudes with disasters can also play important roles in shaping evacuation decision-making behavior (Mozumder et al., 2008; Strawderman et al., 2012; Kuligowski et al., 2022). However, it can be impractical to collect these individual-level data from a large scale of households in real-time to assist in emergency response.

3. Methodology

This section introduces the process of trip generation inference using GPS data, the deep learning architecture for travel demand forecasting in evacuations, and the model updating scheme. The overall framework of this study is presented in Fig. 1. We first clean the GPS trace data by removing duplicated data points and inaccurate data points.[1] Then, we implement a trip generation

[1] As GPS records usually have spatial measurement errors (Zhang et al., 2016), some data providers such as Gravy Analytics would label the accuracy of the latitude and longitude of a GPS record. Modelers can choose an error threshold to filter out highly inaccurate data points. Note that this step can be skipped if the data provider does not provide this data field.

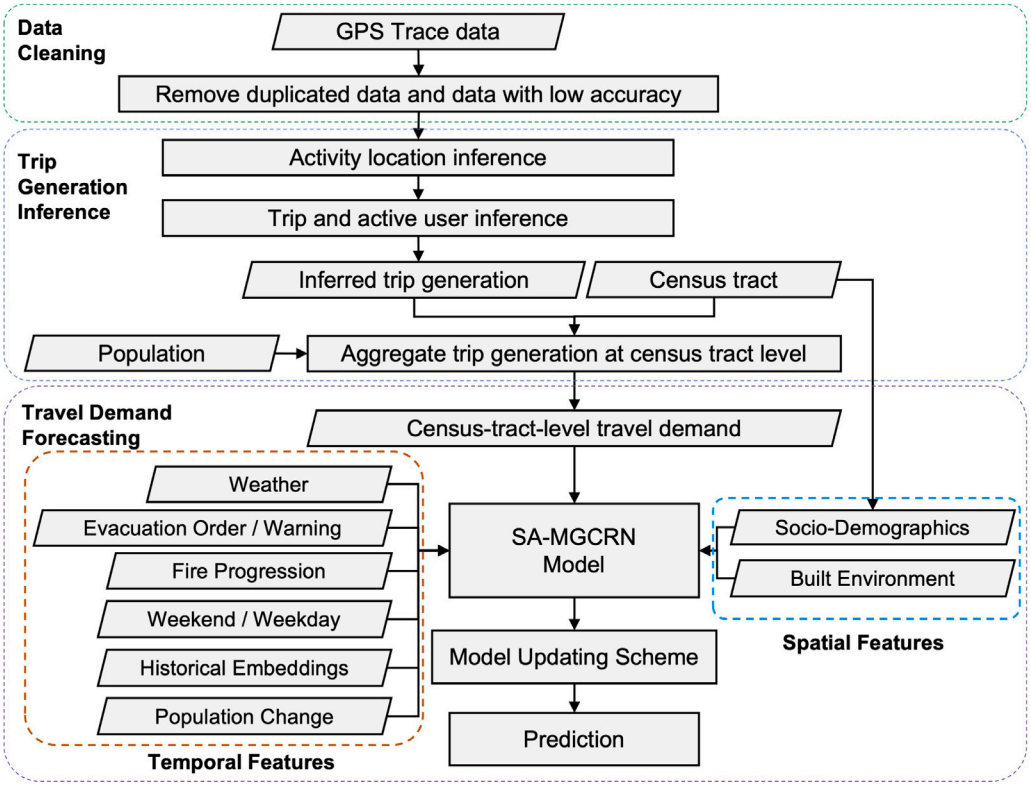

Fig. 1. Overall methodological framework.

inference algorithm to extract census-tract-level active users and trip generations based on the GPS trace data. After that, we train a SA-MGCRN model to forecast the travel demand at the census-tract level. The inputs of the SA-MGCRN model include historical census-tract-level travel demand, temporal features (i.e., weather, evacuation order/warning, fire progression, day of the week, historical demand information, and population change), and spatial features (i.e., socio-demographics and built environment). The proposed model will be updated regularly (e.g., daily) to quickly adapt to the fast-changing situations during wildfires. The details of the trip generation inference algorithm, the SA-MGCRN model, and the model updating scheme are explained in detail in the following subsections.

3.1. Trip generation inference

In this study, we assume that if a GPS device user stayed at a place for a period of time, the user carried out a certain activity at that place. Based on this assumption, we can extract activity locations by clustering using GPS trace data. This study applied an incremental clustering method (Wang and Chen, 2018; Alexander et al., 2015) to infer activity locations. Incremental clustering is a stream clustering approach that processes the data points in a sequence successively. Given a sample of GPS trace data $\boldsymbol{p} = [p_0, p_1, \ldots, p_n]$, the clustering process starts from the first data point p_0. We first create a new cluster C_0 centering at p_0 with radius of R. Then we check if the next data point p_1 is in cluster C_0 by calculating the distance d_1 between p_1 and center of C_0. If $d_1 < R$, we aggregate p_1 to cluster C_0, and update the center of C_0. Otherwise, we create another cluster C_1 centering at p_1 with radius of R. We perform the same procedure for the data points in the GPS trace data sample $\boldsymbol{p}$ successively. After that, we check the time duration t_i of each cluster C_i (i.e., time duration that the GPS device user spends in cluster C_i). If t_i is no less than a given threshold T_c, the cluster C_i is inferred as an activity location. Note that we need to set proper values for cluster radius R and the time threshold T_c to obtain reliable inferred results. In this study, we set the cluster radius R and the time threshold T_c to 100 m and 5 min, respectively (Chen et al., 2014; Wang et al., 2019a). The activity location inference process is illustrated in Fig. 2.

We assume that there is a trip between every two consecutive activity locations in a GPS trace trajectory. Based on this assumption and the inferred activity locations, we can further infer trips by linking time-adjacent activity locations, and thus obtain trip origins and destinations. We further aggregate the trip origins to calculate the hourly census-tract-level trip generations for each day. Since each activity corresponds to a unique ID, we also infer the number of active users on a census-tract-level basis for each timestamp. We denote the inferred (from GPS data) trip generations and the number of active users at hour i for day j in census tract k as M_{ijk}^{GPS} and U_{ijk}^{GPS}. It is important to note that GPS data is generated by only a small portion of the population and therefore cannot

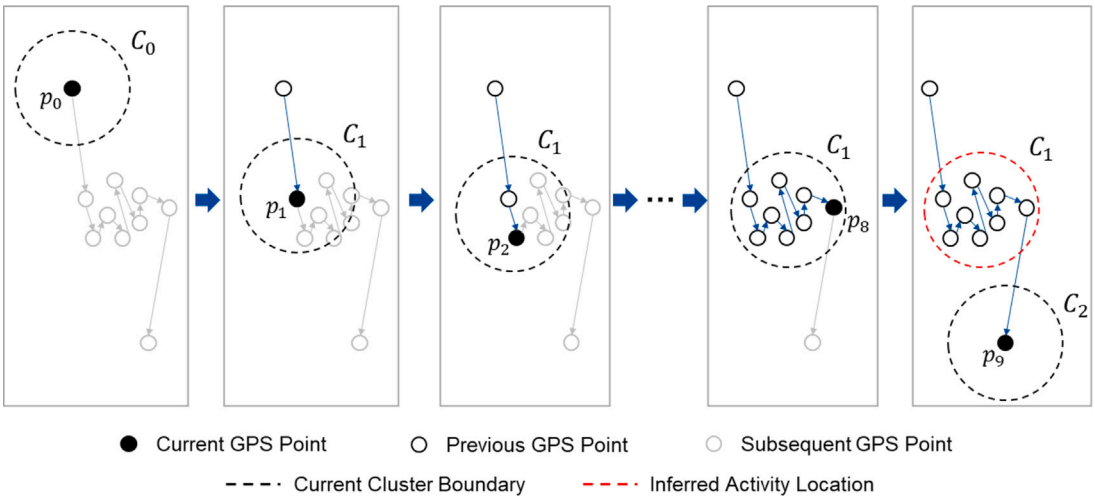

Fig. 2. An example to illustrate the incremental clustering method to infer activity locations.

accurately represent overall travel demand at the census-tract level for the whole population. To address this issue, we use the following procedure.

First, we calculate the average number of active users *per day* during all workdays prior to the wildfire for each census tract k, denoted by U_k^{GPS}. This value is used to represent the overall number of active users for each census tract. Mathematically, it can be written as:

$$U_k^{GPS} = \frac{1}{|S|} \sum_i \sum_j U_{ijk}^{GPS}, j \in S \tag{1}$$

where S refers to the set of workdays prior to the wildfire and correspondingly, $|S|$ refers to the number of workdays (i.e., the size of the set S).

Next, we divide the inferred trip generations (i.e., M_{ijk}^{GPS}) by the assumed overall number of active users per day (i.e., U_k^{GPS}) to obtain the inferred trip generation rate per person for each census tract at every timestamp. Specifically,

$$R_{ijk}^{GPS} = M_{ijk}^{GPS} / U_k^{GPS} \tag{2}$$

where the inferred trip generation rate per person at hour i for day j in census tract k is denoted by R_{ijk}^{GPS}. Finally, we multiply the trip generation rate per person by the number of population in each census tract (denoted by P_k) to estimate the travel demand, or trip generations, of all populations. More formally, trip generations for the whole population at hour i for day j in census tract k is calculated as:

$$M_{ijk}^{All} = R_{ijk}^{GPS} \times P_k \tag{3}$$

By using this procedure, we can obtain a more accurate estimation of travel demand at the census-tract level for the whole population while taking into account the limitations of GPS data.

3.2. Travel demand forecasting

Formally, the travel demand forecasting problem in this study can be formulated as follows.

Definition 1. Let the historical travel demand matrix $\boldsymbol{X}_t^{N\times T}$ represents historical travel demand for all the N census tracts from time $t-T+1$ to time t. $\boldsymbol{X}_t = [\boldsymbol{x}_{t-T+1}, \dots, \boldsymbol{x}_t]$ where $\boldsymbol{x}_t = [x_t^1, x_t^2, \dots, x_t^N]$ denote the travel demand of all census tracts at time t, and x_t^i denote the travel demand of census tract i at time t.

Definition 2. We denote the temporal feature matrix $\boldsymbol{C}_t^{K\times T}$ consists of several temporal features from time $t-T+1$ to time t, including weather condition, evacuation order/warning in each census tract, fire progression, weekend indicator and historical travel demand information, where K is the number of temporal features.

Definition 3. The spatial correlation graph $G = (V, E)$ describes the spatial and property correlation between areas. This graph is fused by two graphs to represent environmental and demographic similarities between the areas. In graph $G = (V, E)$, $V = \{v_1, v_2, \dots, v_N\}$ is a set of nodes (i.e., census tracts). E is a set of edges. Note that the graph G is an undirected graph. If two nodes in G are correlated (e.g., functionally similar), there is an edge between these two nodes. An adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ can be used to represent the graph G. The element $a_{ij} \in \boldsymbol{A}$ is 1 if there is an edge between node i and node j; otherwise, 0.

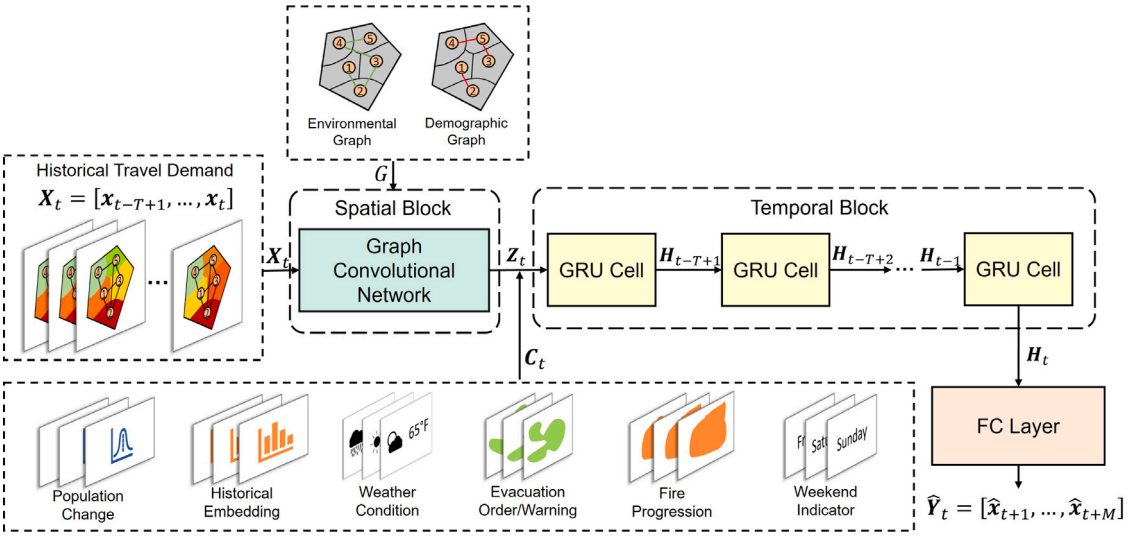

Fig. 3. Architecture of the SA-MGCRN model.

Problem. Given a historical travel demand (i.e., evacuation, background and intermediate trips) matrix $\boldsymbol{X}_t$, a temporal feature (such as weather conditions) matrix $\boldsymbol{C}_t$, and a spatial correlation (such as build environment similarity) graph G, learn a function $f : \mathbb{R}^{N\times T} \rightarrow \mathbb{R}^{N\times M}$ that maps historical travel demand of all census tracts to the travel demand in next M time intervals:

$$\boldsymbol{Y}_t = [\boldsymbol{x}_{t+1}, \dots, \boldsymbol{x}_{t+M}] = f(\boldsymbol{X}_t, \boldsymbol{C}_t, G) \tag{4}$$

where N is the number of census tracts, and T is time sequence length of input historical travel demand data.

3.2.1. Overview of model architecture

We propose a **S**ituational-**A**ware **M**ultiple **G**raph **C**onvolutional **R**ecurrent **N**etwork (**SA-MGCRN**) model to solve the travel demand forecasting problem. The model architecture is presented in Fig. 3. The proposed SA-MGCRN model is composed of a spatial block, a temporal block, and a fully connected (FC) layer. The spatial block is a graph convolutional network (GCN) based on the spatial correlation graph G. The graph G is fused by two graphs including environmental graph and demographic graph. The GCN takes historical travel demand $\boldsymbol{X}_t$ as input, and generates output $\boldsymbol{Z}_t$ by convolutional computation. The output $\boldsymbol{Z}_t$ is then concatenated with temporal feature matrix $\boldsymbol{C}_t$. After that, the temporal block uses a sequence of gated recurrent unit (GRU) cells to processes $[\boldsymbol{Z}_t, \boldsymbol{C}_t]$ successively. The output of temporal block $\boldsymbol{H}_t$ is subsequently processed by a FC layer to generate the prediction $\hat{\boldsymbol{Y}}_t = [\hat{\boldsymbol{x}}_{t+1}, \dots, \hat{\boldsymbol{x}}_{t+M}]$. The details of the GCN model and the GRU model are introduced in the following subsections.

3.2.2. Graph Convolutional Network (GCN)

Travelers in areas with similar properties are more likely to have similar travel behavior (Zhang and Zhao, 2022; Zhang et al., 2024c). Therefore, spatial dependencies are essential to travel behavior prediction (Zhao et al., 2019; Zhang and Zhao, 2022; Xu et al., 2023). Graph convolutional network (GCN) is a widely-used model to capture spatial dependencies. GCN performs convolutional operations based on a graph, thus, it can handle graph-structured data (Zhao et al., 2019). Since the census tracts do not have a regular spatial structure but can be represented by a graph, we use GCN to capture spatial dependencies in this study.

Given an adjacency matrix $\boldsymbol{A}$ and the historical travel demand $\boldsymbol{X}_t$, GCN performs convolutional operations using a filter in the Fourier domain. The filter is applied on each node of the graph, thus capturing spatial dependencies between the node and its adjacent nodes. The GCN model is constructed by multiple convolutional layers:

$$\boldsymbol{H}^{l+1} = \sigma(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}^{l}\boldsymbol{W}^{l}) \tag{5}$$

where $\boldsymbol{H}^l$ is the output of layer l and $\boldsymbol{H}^0 = \boldsymbol{X}_t$, $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}$ is the adjacency matrix of the graph G with self-connections, $\boldsymbol{I}$ is the identity matrix, $\tilde{\boldsymbol{D}}$ is the diagonal node degree matrix of $\tilde{\boldsymbol{A}}$, and $\boldsymbol{W}^l$ is a layer-specific trainable matrix. $\sigma(\cdot)$ denotes an activation function, such as the $ReLU(\cdot) = max(0, \cdot)$ (Nair and Hinton, 2010). A two-layer GCN model (Kipf and Welling, 2016) is used in this study to capture spatial dependencies. We first calculate $\hat{\boldsymbol{A}} = \tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}$ in a pre-processing step. The forward model then takes the form:

$$\boldsymbol{Z} = f(\boldsymbol{X}, \boldsymbol{A}) = softmax(\hat{\boldsymbol{A}}\ ReLU(\hat{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W}^0)\ \boldsymbol{W}^1) \tag{6}$$

where $\boldsymbol{W}^0 \in \mathbb{R}^{C\times H}$ is the input-to-hidden weight matrix, C is the number of input channels (i.e., a C-dimensional feature vector for each node), H is the number of hidden units, $\boldsymbol{W}^1 \in \mathbb{R}^{H\times F}$ is the hidden-to-output weight matrix, F is the number of filters, $\boldsymbol{Z} \in \mathbb{R}^{N\times F}$ is the output convolved matrix, and N is the number of nodes. The softmax activation function, defined as $softmax(x_i) = \frac{1}{z}e^{x_i}$ with $z = \sum_i e^{x_i}$, is applied row-wise.

3.2.3. Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) (Cho et al., 2014) is a developed RNN model to process sequence data such as time series and speech signal data. Using the gating mechanism to memorize long-term information, GRU can well handle the vanishing gradient problem that may happen in traditional RNN models. Compared with other RNN models such as LSTM (Hochreiter and Schmidhuber, 1997), GRU is faster to compute but still offers comparable performance in prediction (Chung et al., 2014). Therefore, we use GRU to perform temporal prediction in this study.

GRU uses two gates, reset gate and update gate, to determine what information should be kept and passed to the output. The reset gate controls how much of the previous state information to remember, and the update gate determines how much of the past information needs to be passed to the new state. The structure of a GRU cell is presented in Fig. 4.

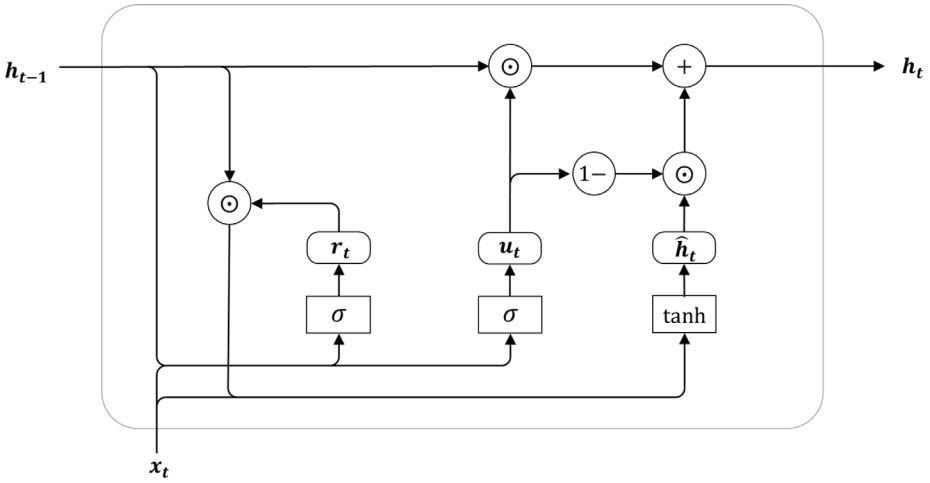

Fig. 4. Structure of a GRU cell. h_{t-1} is the previous hidden state. x_t is the current input. r_t is the reset gate and u_t is the update gate. $\hat{h}_t$ is the candidate hidden state and h_t is the current hidden state. $\odot$ is the element-wise product operator. σ is the sigmod activation and $tanh$ is the tangent hyperbolic activation.

The GRU cell takes current input $\boldsymbol{x}_t$ and the previous hidden state $\boldsymbol{h}_{t-1}$ as inputs, and outputs the new hidden state $\boldsymbol{h}_t$. Given $\boldsymbol{x}_t$ and $\boldsymbol{h}_{t-1}$, the reset gates $\boldsymbol{r}_t$ and the update gate $\boldsymbol{u}_t$ are calculated by fully connected layers with the sigmoid (Nwankpa et al., 2018) activation function σ. Mathematically, for a given time step t, the reset gates $\boldsymbol{r}_t$ and the update gate $\boldsymbol{u}_t$ are computed by:

$$\boldsymbol{r}_t = \sigma(\boldsymbol{x}_t \boldsymbol{W}_{xr} + \boldsymbol{h}_{t-1} \boldsymbol{W}_{hr} + \boldsymbol{b}_r) \tag{7}$$

$$\boldsymbol{u}_t = \sigma(\boldsymbol{x}_t \boldsymbol{W}_{xu} + \boldsymbol{h}_{t-1} \boldsymbol{W}_{hu} + \boldsymbol{b}_u) \tag{8}$$

where $\boldsymbol{W}_{xr}$, $\boldsymbol{W}_{hr}$, $\boldsymbol{W}_{xu}$, $\boldsymbol{W}_{hu}$ are weight parameters and $\boldsymbol{b}_r$, $\boldsymbol{b}_u$ are biases.

Then we integrate the reset gate $\boldsymbol{r}_t$ with the previous hidden state $\boldsymbol{h}_{t-1}$ and current input $\boldsymbol{x}_t$ to generate the candidate hidden state $\hat{\boldsymbol{h}}_t$:

$$\hat{\boldsymbol{h}}_t = tanh(\boldsymbol{x}_t \boldsymbol{W}_{xh} + (\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}) \boldsymbol{W}_{hh} + \boldsymbol{b}_h) \tag{9}$$

where $\boldsymbol{W}_{xh}$, $\boldsymbol{W}_{hh}$ are weight parameters, $\boldsymbol{b}_h$ is the bias, and $\odot$ is the element-wise product operator.

We use the update gate $\boldsymbol{u}_t$, the candidate hidden state $\hat{\boldsymbol{h}}_t$, and the previous hidden state $\boldsymbol{h}_{t-1}$ to compute the current hidden state $\boldsymbol{h}_t$. The hidden state $\boldsymbol{h}_t$ is the output of the GRU cell and will be passed forward. The current hidden state $\boldsymbol{h}_t$ is computed by:

$$\boldsymbol{h}_t = \boldsymbol{u}_t \odot \boldsymbol{h}_{t-1} + (1 - \boldsymbol{u}_t) \odot \hat{\boldsymbol{h}}_t \tag{10}$$

3.2.4. Integrating GCN and GRU

We integrate GCN and GRU to capture the spatial dependencies and temporal dependencies simultaneously. Let $\boldsymbol{A}$ denote the matrix of spatial correlation graph G; let $\boldsymbol{X}_t$ denote the historical travel demand. We first use GCN to process $\boldsymbol{A}$ and $\boldsymbol{X}_t$ using Eq. (6):

$$\boldsymbol{Z}_t = f(\boldsymbol{X}_t, \boldsymbol{A}) = softmax(\hat{\boldsymbol{A}}\ ReLU(\hat{\boldsymbol{A}} \boldsymbol{X}_t \boldsymbol{W}^0)\ \boldsymbol{W}^1) \tag{11}$$

Then we concatenate the output matrix $\boldsymbol{Z}_t$ with the temporal feature matrix $\boldsymbol{C}_t$ to generate the input $\boldsymbol{Z}'_t = [\boldsymbol{Z}_t, \boldsymbol{C}_t]$ of GRU. The calculation process in the GRU cell at time t can be expressed as:

$$\boldsymbol{R}_t = \sigma(\boldsymbol{Z}'_t \boldsymbol{W}_{ZR} + \boldsymbol{H}_{t-1} \boldsymbol{W}_{HR} + \boldsymbol{b}_R) \tag{12}$$

$$\boldsymbol{U}_t = \sigma(\boldsymbol{Z}'_t \boldsymbol{W}_{ZU} + \boldsymbol{H}_{t-1} \boldsymbol{W}_{HU} + \boldsymbol{b}_U) \tag{13}$$

$$\widehat{\boldsymbol{H}}_t = tanh(\boldsymbol{Z}_t'\boldsymbol{W}_{ZH} + (\boldsymbol{R}_t \odot \boldsymbol{H}_{t-1})\boldsymbol{W}_{HH} + \boldsymbol{b}_H) \tag{14}$$

$$\boldsymbol{H}_t = \boldsymbol{U}_t \odot \boldsymbol{H}_{t-1} + (1 - \boldsymbol{U}_t) \odot \widehat{\boldsymbol{H}}_t \tag{15}$$

where $\boldsymbol{R}_t$ is the reset gate, $\boldsymbol{U}_t$ is the update gate, $\boldsymbol{H}_{t-1}$ is the hidden state of the previous time step, $\widehat{\boldsymbol{H}}_t$ is the candidate hidden state, $\boldsymbol{H}_t$ is the current hidden state, $\boldsymbol{W}_{ZR}$, $\boldsymbol{W}_{HR}$, $\boldsymbol{W}_{ZU}$, $\boldsymbol{W}_{HU}$, $\boldsymbol{W}_{ZH}$, $\boldsymbol{W}_{HH}$ are weight parameters, and $\boldsymbol{b}_R$, $\boldsymbol{b}_U$, $\boldsymbol{b}_H$ are biases.

3.2.5. Constructing spatial correlation graph

We use two graphs to represent the spatial correlations between areas.

The first one is the environmental similarity graph $G_F = (V, E_F)$ which is constructed by linking two environmentally similar areas i and j by an edge $e_F^{i,j} \in E_F$. The environmental similarity is evaluated by the same metrics (i.e., *Pearson coefficient*) in Tang et al. (2021). In this study, we use the number of residential parcels, median parcel size (per race), high fire risk zone indicator and local responsibility area (LRA) indicator to estimate the environmental similarity. Let $\boldsymbol{A}_F$ denote the adjacency matrix of graph G_F, the element $a_F^{i,j} \in \boldsymbol{A}_F$ is given by:

$$a_F^{i,j} = \begin{cases} 1, & \text{if } sim(\boldsymbol{p}_i, \boldsymbol{p}_j) > d_F \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

where $\boldsymbol{p}_i \in \mathbb{R}^{1\times n}$ and $\boldsymbol{p}_j \in \mathbb{R}^{1\times n}$ are the environmental feature vectors of nodes i and j respectively, n is the number of features, $sim(\cdot)$ is the calculation function of the Pearson coefficient, and d_F is the threshold parameter, which is set to 0.9 to control the sparsity of the graph.

Similarly, the demographic similarity graph $G_D = (V, E_D)$ is constructed by connecting two nodes i and j with similar demographic characteristics by an edge $e_D^{i,j} \in E_D$. Let $\boldsymbol{A}_D$ denote the adjacency matrix of graph G_D, the element $a_D^{i,j} \in \boldsymbol{A}_D$ is given by:

$$a_D^{i,j} = \begin{cases} 1, & \text{if } sim(\boldsymbol{q}_i, \boldsymbol{q}_j) > d_D \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

where $\boldsymbol{q}_i \in \mathbb{R}^{1\times m}$ and $\boldsymbol{q}_j \in \mathbb{R}^{1\times m}$ are the demographic feature vectors of nodes i and j respectively, m is the number of demographic features, $sim(\cdot)$ is the calculation function of the Pearson coefficient, and d_D is the threshold parameter, which is also set to 0.9.

The environmental similarity and the demographic similarity are then fused to jointly represent the spatial correlations between areas. Specifically, the fused graph $G = (V, E)$ is established by connecting two areas i and j by an edge $e_{i,j} \in E$. Given $\boldsymbol{A}$ as the adjacency matrix of graph G, the element $a^{i,j} \in \boldsymbol{A}$ is calculated by:

$$a^{i,j} = \begin{cases} 1, & \text{if } sim(\boldsymbol{p}_i, \boldsymbol{p}_j) + sim(\boldsymbol{q}_i, \boldsymbol{q}_j) > d_F + d_D \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

In this way, the fused graph G can largely maintain the sparsity and identify the areas with similar environmental and demographic characteristics.

3.3. Model updating scheme

Ideally, the GPS data providers should deliver near real-time GPS data with minimal delay. In this way, we can update our model, in a timely manner, to ensure that it can quickly adapt to the fast-changing situations during wildfires and thus better facilitate real-time decision-making. However, in real-world applications, data delivery could also be delayed (i.e., from when data is generated to ready-to-use status) for certain hours (Gravy Analytics, 2023), due to the latency in data transfer, the time for data processing, and/or legal constraints (Smart Traffic, 2021). This study recognizes the pivotal importance of timeliness in addressing real-world challenges (Wand and Wang, 1996) and therefore takes into account both situations—one where data are delivered promptly and another where it is subject to delays.

Fig. 5 illustrates the model updating scheme for both with and without delay. Suppose the GPS data is delivered on an hourly basis. The model takes a sequence of data as input (shown as the blank box) and forecasts the travel demand for the following hour (shown as the red bar). By the end of each day, we re-train/update the forecasting model with the available (observed) data (as shown by the green bar). Then, we use the updated model to predict future travel demand on the following day. On the other hand, if the data delivery is delayed by multiple hours (as shown by the gray bar), we have to use less available data to train the model and the performance of the model is expected to be compromised. Note that there could be a trade-off between the model update frequency and the computational efficiency. Frequent recalibrations demand significant computational resources and time. Recalibrating with minor updates (such as hourly updates) could result in less reliable estimations due to the limited new information, especially if there are delivery delays. To achieve near real-time accurate travel demand predictions during wildfires, we believe updating the model promptly to learn the most up-to-date information and the computational efficiency should be equally important. Therefore, updating the model at an appropriate frequency can ensure that the data is efficiently processed while still maintaining timely updates that reflect the latest available information. For practical and demonstration purposes, we chose a daily recalibration schedule in this study. While critical features, such as detailed fire progression information or updated evacuation

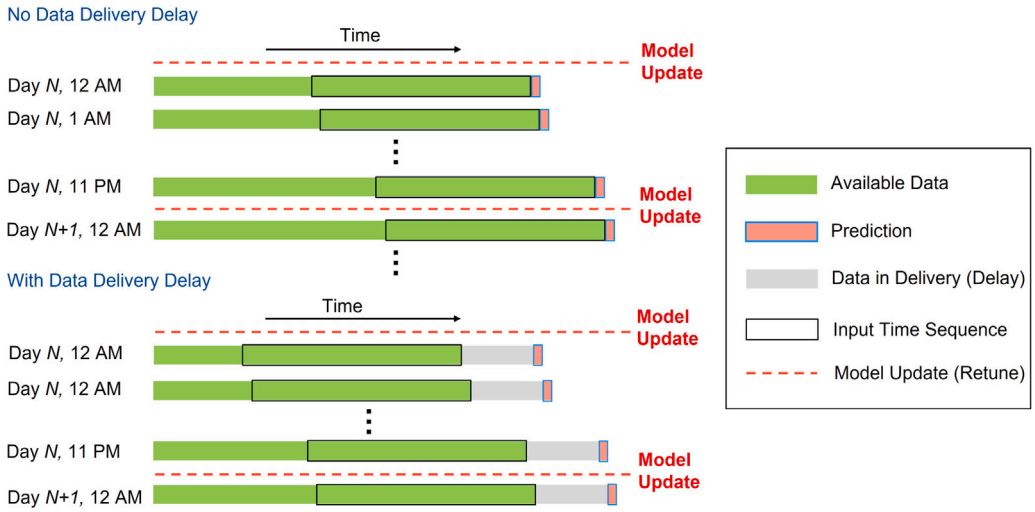

Fig. 5. Model updating scheme considering data delivery delay. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

warnings, can be sometimes updated more frequently, our current approach demonstrates the feasibility of this concept. As better data become available and computational capabilities are enhanced, we believe the model can be recalibrated more frequently to further improve accuracy.

For the case study (see Section 4), since the wildfire started on October 23 and was fully contained on November 6, we stopped updating the model after generating predictions for November 6 and updated the model 14 times (i.e., 14-day duration of wildfire) in total. We consider three scenarios: (1) no delay, (2) 24-h delay, and (3) 48-h delay, and then compare the model's performance (see Section 5.4).

The model updating scheme enables the model to take in timely information as new data become available (Shahraki et al., 2022). During emergencies like wildfires, the situation is ever-changing, e.g., fire spread, weather, evacuation order/warning, among others, and the changing situation will greatly impact a householder's evacuation decision-making and travel behavior (Lindell and Perry, 2012). Therefore, it is highly beneficial to recalibrate the forecasting model frequently in order to adapt to the rapidly changing situation and thus produce more accurate predictions to facilitate better-informed real-time decision-making.

4. Data

We selected the 2019 Kincade Fire, Sonoma County, CA, as the case study. This section introduces the study site and the wildfire in Section 4.1 and describes the dataset in Section 4.2.

4.1. Study site

Sonoma County is located in Northern California, U.S. The population estimate of Sonoma County was 494,336 in 2019. Its county seat and largest city is Santa Rosa. The highway system of Sonoma County consists of U.S. Highway 101, and State Highways 1, 12, 37, 116, 121, and 128. The Kincade Fire started northeast of Geyserville at 9:27 p.m. on October 23, 2019 and was fully contained at 7:00 p.m. on November 6, 2019. The fire burned 77,758 acres, destroyed 374 structures, damaged 60 structures, and caused 4 injuries (Sonoma Operational Area and the County of Sonoma, Department of Emergency Management, 2020). As the fire spread, the mandatory evacuation order was first issued in Geyserville at 10:00 a.m. on October 26, and then the evacuation warnings and orders grew to encompass nearly all of Sonoma County in the following days, making it the largest evacuation in Sonoma County's history. The study site and the fire perimeter are shown in Fig. 6.

4.2. Data description

The GPS data[2] was provided by Gravy Analytics and built on privacy-friendly mobile location data. Gravy's location data platform processes raw location signals from multiple data providers representing over 150 million U.S. mobile devices. The location information was aggregated into cells (the size is 4.77 m × 4.77 m) and represented by a 9-character geohash. The timestamp of

[2] The GPS data underwent Gravy's cleansing processes and was optimized with Gravy Location Data Forensics–filtering and categorizing inaccurate and even fraudulent location signals. This enabled us to identify and use only the cleansed location signals relevant to this project.

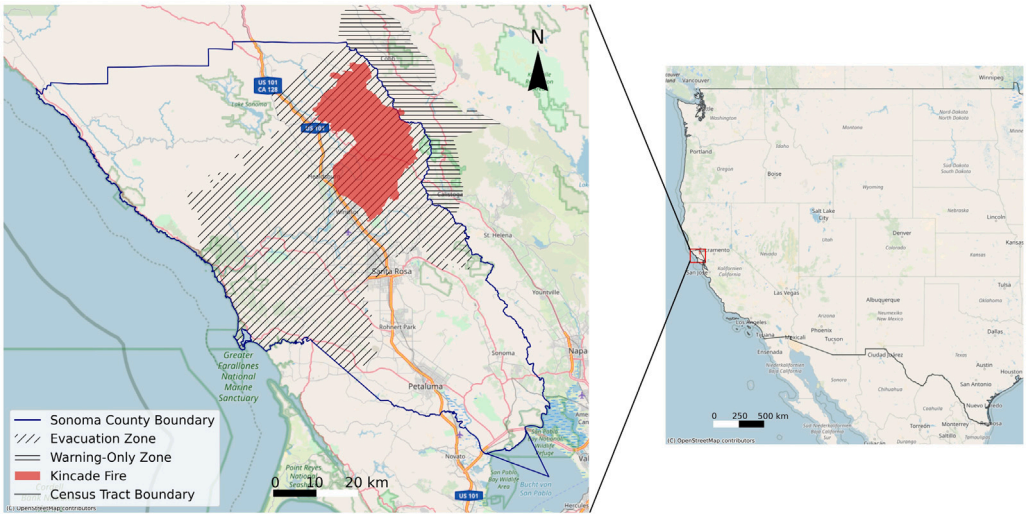

Fig. 6. Sonoma County and the Kincade Fire perimeter.

the signal was encoded as epoch time in milliseconds rounded down to the floor minute. The accuracy of the signal location was indicated by the Forensic Flag (to measure the GPS error). The GPS error was usually caused by spoofed locations (e.g., some mobile applications may generate fake location information), IP address derived signals, Cell Tower Derived (CTD) signals, abnormal signals density, and GPS signal floating. In this study, we adopt the same data cleaning process as in Zhao et al. (2022). Records with signal location error no more than 250 m were retained. After the data cleaning process, we included 100,913,550 GPS signal records in Sonoma County, CA from October 16, 2019 to November 13, 2019 for analysis.

The hourly weather condition data were collected using the *OpenWeather*[3] API. The parcel data (for functional similarity evaluation), the evacuation order and warning information, and the fire progression information were provided by the county authorities. The demographic data were collected from the *American Community Survey 2015–2019 5-year* estimates data. Table 1 presents the descriptive statistics of the input variables used in the case study. The functional variables, the demographic variables, the fire distance, and the evacuation order/warning are processed at the census-tract level, and the weather condition variables are at the county level. In this study, the fire distance is defined as the shortest Euclidean distance between the fire perimeter and the boundary of each census tract. The evacuation order/warning is a dummy variable that indicates if a census tract is under mandatory evacuation order or warning at a given timestamp. The weekend indicator is also a dummy variable indicating weekday or weekend. We also included historical travel demand information, i.e., historical embeddings, as two additional inputs. Specifically, historical embeddings 1 refers to the overall historical average travel demand for each census tract before given timestamp while historical embeddings 2 refers to the only last-four-hour historical average travel demand for each census tract. Finally, we calculate the population change (i.e., the number of active users on each day after wildfire divided by the average number of active users per weekday before wildfire) to capture the population movement during wildfire evacuation.

5. Results

In the following section, we will sequentially introduce the results of the inferred trip generations, the validation of the trip-generation results, the implementation settings of SA-MGCRN, the brief profile of all benchmark models, the prediction performance of SA-MGCRN and benchmark models across three delay scenarios and the predictive contributions of the major components embedded in SA-MGCRN via an ablation study.

5.1. Trip generation inference

We inferred 457,100 trips generated by 4894 unique users in Sonoma County from October 18, 2019 to November 13, 2019 using the GPS data. On average, each user generated 3.54 trips per day. This value is close to the results produced by National Household Travel Survey (NHTS), where the average number of trips per individual per day is estimated at around 4 (United States Department of Transportation, 2023). The average trips estimated from GPS data is slightly lower than survey data, which is consistent with existing findings (Wang et al., 2019b). Then, we aggregated these trip origins at the census-tract level on an hourly basis and consider them as the trip generations. Sequentially, we further inferred the trip generations (i.e., travel demand) for the whole population

[3] https://openweathermap.org/api.

Table 1
Descriptive statistics of input variables.

Variables	Mean	Std.	Min	Max	Category
Residential area count	1 653.091	713.102	227.000	4 333.000	Environmental similarity graph
Median parcel size (per arce)	0.518	1.308	0.033	9.355	
High fire risk zone	0.152	0.360	0.000	1.000	
If the zone is inside LRA	0.697	0.462	0.000	1.000	
Population density (per sq. mile)	3 339.788	3 192.117	7.155	12 474.630	Demographic similarity graph
Proportion of the white population	0.766	0.129	0.382	0.955	
Proportion of population with BA's degree and above	0.363	0.127	0.121	0.636	
Median household income (US dollar)	83 823.667	20 522.564	49 856.000	145 147.000	
Proportion of households own 0 car	0.075	0.068	0.000	0.320	
Median age	43.801	8.960	23.000	71.600	
Fire distance	181.731	102.993	0.000	250.000	Temporal variables
Evacuation order/warning	0.101	0.301	0.000	1.000	
Weekend indicator	0.308	0.462	0.000	1.000	
Temperature	56.762	12.803	33.400	90.600	
Feels like temperature	56.521	12.574	33.400	86.400	
Wind speed	4.017	4.314	0.000	29.600	
Sea level pressure	1 016.715	3.230	1 005.900	1 023.900	
Humidity	58.437	29.248	8.920	100.000	
Visibility	8.752	2.416	0.000	9.900	
Cloud cover	17.321	26.958	0.000	100.000	
UV index	1.535	2.201	0.000	8.000	
Historical embeddings 1	207.002	73.883	77.194	478.749	
Historical embeddings 2	187.741	69.326	70.367	430.327	
Population Change	0.819	0.225	0.000	1.706	

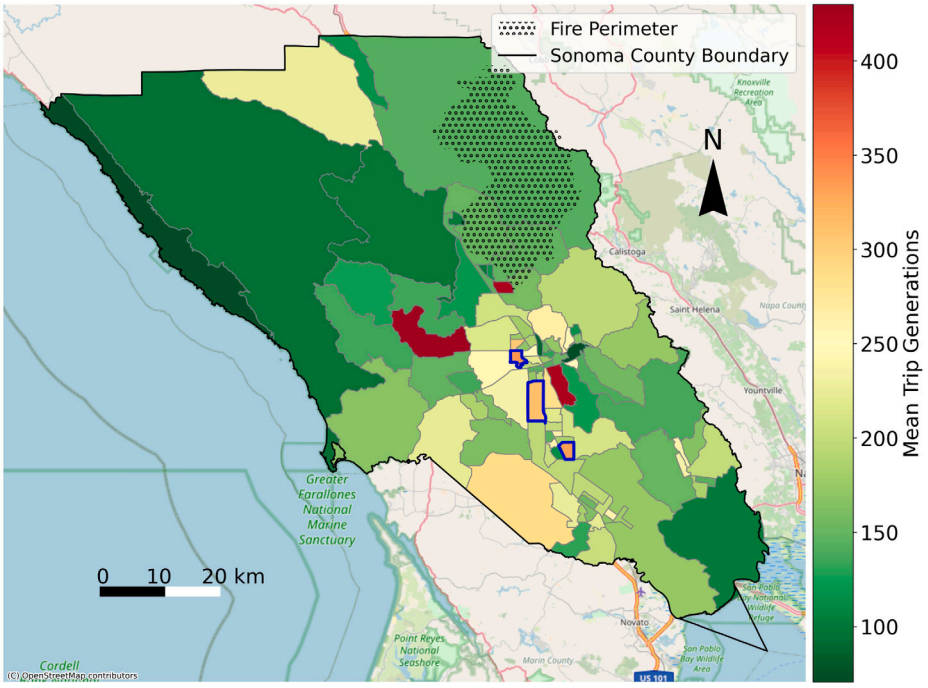

Fig. 7. Spatial distribution of mean trip generation in Sonoma County at census-tract level. Three census tracts with blue-highlighted boundary were selected to show the model's prediction performance (please refer to Section 5). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

following the procedure described in Section 3.1. The spatial distribution of trip generations is presented in Fig. 7. The mean of hourly inferred travel demand for each census tract is 187.644, the standard deviation is 69.335, the maximum value is 429.975, and the minimum value is 70.612.

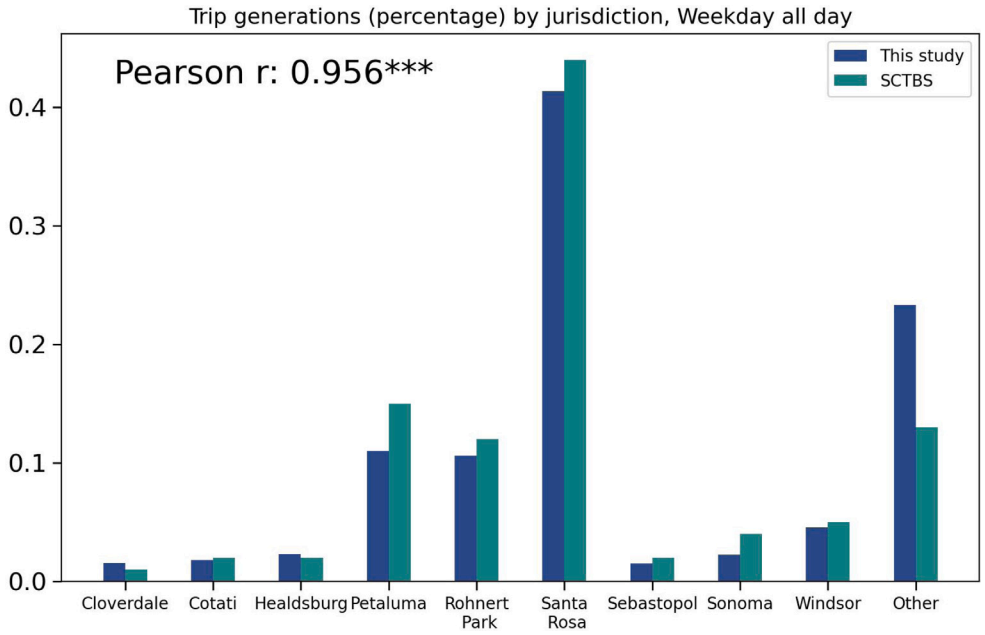

Fig. 8. Trip generations (percentage) by jurisdiction on weekday all day.

5.2. Validation of trip generation results

Before we deploy the trip generation results into modeling process, we need to ensure their validity and reliability. Therefore, we contrasted the results of the inferred trip generations with findings from Sonoma County Travel Behavior Study (SCTBS) (Sonoma County Transportation Authority, 2023). Sonoma County Transportation Authority (SCTA) conducted this comprehensive travel behavior study in 2019 with local jurisdictions. Thus, we believe that SCTBS can serve as a reliable reference for examining the validity of our trip generation results. SCTBS used large-scale GPS data derived from mobile devices to identify the patterns of trip generation and destinations, trip length and distance, major trip purposes and other critical travel information. The GPS data used for SCTBS were provided by StreetLight[4] and Cuebiq,[5] which were obtained from March, 2017 to May, 2017.

Specifically, we compared the percentage of trip generations by jurisdiction (nine major cities and other areas) across three scenarios: weekday all day, weekday AM peak (6 AM–10 AM) and weekday PM peak (3 PM–7 PM). ote that in SCTBS (Sonoma County Transportation Authority, 2023), the weekday only includes Tuesday, Wednesday and Thursday. To keep consistent with SCTBS, our validation also only considered weekday as Tuesday to Thursday. Fig. 8 presents the trip generation comparison for weekday all day. For the other two scenarios, please refer to Figs. A.1 and A.2 in Appendix A. Results showed that our inferred trip generations largely align with the results reported by SCTA. Fig. 8 showed that for nine major cities, the percentage of trip generations derived by our proposed algorithm tends to be slightly underestimated (but almost consistent) compared with results from SCTBS. While our proposed algorithm may overestimate the trip generations in other (unincorporated) areas. The Pearson correlation calculated by these two sources of trip generations is 0.962 and holds strong statistical significance. To further examine the differences of trip generations estimated by this study and SCTBS, we used the Kolmogorov–Smirnov test (K–S test) (Kolmogorov, 1933; Smirnov, 1948). The p-value calculated by K-S test can be used to indicate if we can or cannot reject the null hypothesis that two data samples follow the same distribution. Results showed that the corresponding p-value (0.99) is well above the commonly-used 0.05 threshold, implying that the distributions of trip generations estimated by our study and SCTBS are statistically similar. Results of other scenarios almost replicated this finding. Therefore, we believe our inference of trip generation within Sonoma County is reliable for follow-up analysis.[6]

[4] https://www.streetlightdata.com/.

[5] https://www.cuebiq.com/.

[6] Note that previous studies usually compared the trip-related characteristics derived from GPS data with traditional travel surveys (Wang et al., 2019b; Sinclair et al., 2023). However, due to the availability of reliable references, we only compared our results with the findings from another GPS dataset, which was provided by StreerLight and Ceubiq. These two data providers have been widely applied for large-scale spatiotemporal mobility analysis (Coleman et al., 2022; Monz et al., 2021; Wu et al., 2021; Pepe et al., 2020) and partnered with researchers, federal agencies, and Department of Transportation to assist traffic estimations (Turner et al., 2020; Adler et al., 2017).

Table 2
Prediction performance of SA-MGCRN and benchmark models across all census tracts in Sonoma County.

Models	Delay = 0 h		Delay = 24 h		Delay = 48 h	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
HA	94.498	125.599	97.134	128.808	99.313	131.303
ARIMA	108.474	149.929	110.499	151.676	112.921	155.256
SVM	84.700	114.182	88.329	118.906	90.653	122.423
RF	83.855	114.376	88.396	120.406	91.837	125.060
GBDT	88.635	122.338	94.499	129.915	98.166	134.970
MLP	86.763	117.022	92.944	125.969	100.307	135.496
LSTM	88.075	117.734	91.693	121.726	94.248	125.027
SA-MGCRN	**77.347**	**106.016**	**81.189**	**111.579**	**85.866**	**118.107**

5.3. Model setting

The case studies were conducted with Adam optimizer in a Pytorch environment (Paszke et al., 2019) using an Ampere A-100 GPU. We tuned the hyperparameters, including batch size, sequence length and learning rate, of SA-MGCRN model for each instance (i.e., each date and each data delivery data scenario). We adopted an early-stopping technique to prevent the model from overfitting and improve the training efficiency. We initially set the number of training epochs as 500. If the trained model does not have an accuracy improvement during the last 50 training epochs, we consider the model has already converged. In our case study, we updated (i.e., retune) SA-MGCRN 14 times (from October 23, 2019 to November 6, 2019) in total. Note that as discussed in Section 3.3, at each update, we re-trained the model across all hyperparameters to make sure it captured the timely information of evacuation order, fire distance, weather, etc. The testing set was set as the travel demand on the following day after each model updates. The validation set was set as the travel demand of the day before the testing set day and the training set was set as the travel demand data before validating set day.

5.4. Model comparison

In this section, we compared the proposed SA-MGCRN model with several benchmark models. The details of these models are described as follows. Note that all the models were fine-tuned and we also applied the proposed model updating scheme to these benchmark models.

- **HA**: Historical Average is one of the most fundamental statistical models for time series prediction. HA predicts the demand in a specific timestamp period by averaging historical observations (Hyndman and Athanasopoulos, 2018).
- **ARIMA**: Auto-Regressive Integrated Moving Average is a statistical time series prediction model. ARIMA fits a parametric model based on historical observations to predict future demand. The order of ARIMA was set to (3,1,1) in the case study.
- **SVR**: Support Vector Regression is a machine learning model that uses the same principle as Support Vector Regression (SVM) but for regression problems. We used the Radial Basis Function (RBF) kernel here. The cost was set to 10.
- **GBDT**: Gradient Boosting Decision Tree is a tree-based ensemble machine learning model. In this case study, the number of trees was set to 300, the maximum depth was set to 5, and the learning rate was set to 0.05.
- **RF**: Random Forest is another tree-based ensemble machine learning method. In this model, the number of trees was set to 100, and we consider all features when looking for the best split.
- **MLP**: Multiple Layer Perceptron is a classical feedforward artificial neural network. In the case study, we used an MLP model with 64 neurons in one hidden layer, followed by a dropout layer to prevent the model from overfitting. Both the learning rate, batch size, the sequence length are fine-tuned at each trial.
- **LSTM**: Long Short-Term Memory is another widely used neural network based on the gating mechanism. Both the learning rate, batch size, the sequence length are fine-tuned for each trail.

We compared and evaluated the prediction performance of all the benchmarks and SA-MGCRN using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics can be defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \tag{19}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{20}$$

where y are the observed trip generations and $\hat{y}$ are the predicted trip generations.

The model performance results across all areas are shown in Table 2. First, when there are no data delivery delay issues, the SA-MGCRN model significantly outperformed all the benchmark models. MLP showed better prediction performance than the statistical models (i.e., HA and ARIMA) and the classical machine learning models (i.e., SVR, GBDT, and RF). Among the classical machine learning models, the RF model had the best performance. LSTM delivered a relatively lower predictive performance,

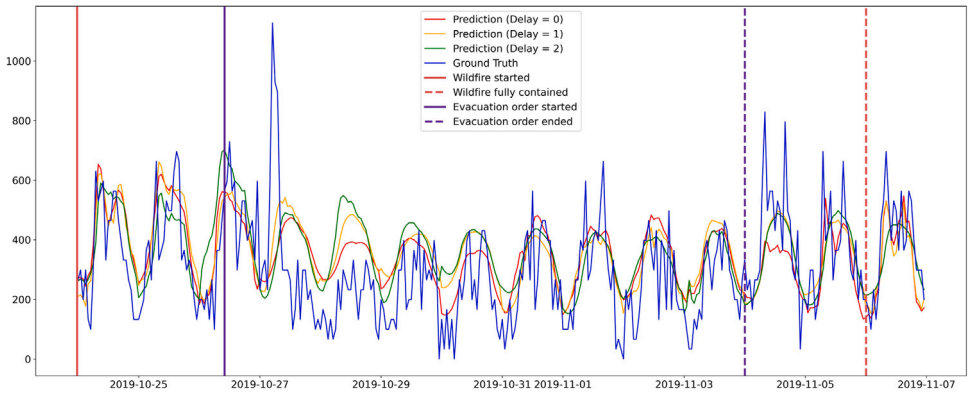

Fig. 9. Hourly prediction performance of SA-MGCRN considering data delivery delay for census tract ID: 6097153006. Please refer to the census tract with highlighted boundary in Fig. 7.

Table 3
Prediction performance of SA-MGCRN and benchmark models across areas under evacuation orders or warnings in Sonoma County.

Models	Delay = 0		Delay = 1		Delay = 2	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
HA	92.122	124.083	96.101	129.468	99.338	133.447
ARIMA	109.239	153.590	112.407	156.546	114.820	160.216
SVM	87.907	118.658	93.511	125.690	96.516	130.059
RF	88.092	120.698	94.557	129.195	99.497	135.492
GBDT	93.143	129.611	100.692	139.322	106.566	147.046
MLP	91.142	123.155	98.924	134.323	108.276	145.984
LSTM	89.402	120.080	93.848	124.937	96.582	128.333
SA-MGCRN	**77.967**	**107.690**	**83.428**	**115.603**	**89.648**	**123.776**

probably because its complex model structure is not appropriately applicable to the small-size training set ($n = 624$). The prediction accuracy across all models showcased a decreasing trend when the days of delay gradually increase. For example, MAE of SA-MGCRN increased from 77.347 to 81.189 and 85.866 when there are 24 and 48 h delay. This is intuitive since more delay usually leads to the absent awareness of the ongoing (wildfire) situation, which inevitably restricts the model's prediction performance. Notably, our proposed SA-MGCRN model was still the best-performing one among all models even if there exist data delivery delay issues. We also illustrated the hourly prediction performance of SA-MGCRN during the wildfire (i.e., October 24 to November 6, 2019) of three census tracts. We show one cnesus tract here as an example in Fig. 9, for the other two, please refer to Figs. C.1 and C.2 in Appendix C. The results showed that the prediction accuracy of the SA-MGCRN model can well pick up the temporal trend of the trip generation rate under different levels of delay. However, we observed that SA-MGCRN struggled to forecast some extreme values, especially the highest demand that occurred on 5 am, October 27 when the evacuation order involving the most population (186,651) was issued (see Fig. 9). This is probably because the wildfire-related data (e.g., evacuation order and fire distance) was unobserved for SA-MGCRN until the wildfire started (after October 26). The predictive error significantly dropped as the model gradually updated itself by accounting for wildfire-related information.

We also compared the prediction performance of SA-MGCRN and benchmark models across areas under evacuation orders or warnings in Sonoma County (please refer to Fig. 6 for the spatial distribution of these areas), as shown in Table 3. The results showed that all models have relatively lower prediction performance than the performance across all areas shown in Table 2. However, our proposed model, i.e., SA-MGCRN, was still the *best-performing* model in no delay, 24-h delay, and 48-h delay scenarios.

5.5. Ablation study

We also conducted an ablation study for SA-MGCRN with no data delivery delay. The ablation study examines the performance of the model by removing certain components to see the contribution of the removed components (Meyes et al., 2019). The component is considered to be more significant when its removal leads to an increase in model error. In such a scenario, the model relies relatively more on this component for making predictions. In this ablation study, we generated nine models by sequentially removing the environmental similarity graph, demographic similarity graph, weekend indicator, weather information, evacuation order/warning information, fire distance, two historical embeddings, and population change.

The performance of the nine ablated models is presented in Table 4. We calculated the importance of each component by two performance metrics, i.e., MAE and RMSE, to ensure the robustness of the results. The results suggested that all components contribute to the prediction accuracy, with the weekend indicator, popotation change, and evacuation order/warning information, fire distance being the top 4. Weekend information was found to be the most influential component in SA-MGCRN, which is

Table 4
Results of ablation study.

Methods	MAE	Importance % (by MAE)	RMSE	Importance % (by RMSE)
SA-MGCRN	**77.347**	–	**106.016**	–
W/O environment similarity	77.492	1.501	106.147	0.899
W/O demographic similarity	77.477	1.340	106.210	1.331
W/O weekend indicator	**79.706**	**24.318**	**109.700**	**25.300**
W/O weather information	78.067	7.425	107.535	10.431
W/O evacuation order/warning	**79.357**	**20.724**	**108.081**	**14.183**
W/O fire distance	**78.563**	**12.537**	**107.534**	**10.425**
W/O historical embeddings 1	78.113	7.905	107.541	10.473
W/O historical embeddings 2	78.024	6.985	107.658	11.273
W/O population change	**79.021**	**17.264**	**108.300**	**15.683**

Notes: Historical Embeddings 1 was calculated as the entire historical average at the given timestamp; Historical Embedding 2 was calculated as the last four hour average. The importance of the kth component was calculated as $e_k \times 100\% / \sum_{i=1}^{n} e_i$, where e_k refers to the increased predictive error after removing the kth component.

consistent with the existing literature (Xie et al., 2021; Tang et al., 2019). It accounted for 24.318% predictive importance for MAE and 25.300% for RMSE. According to the results, the prediction error considerably increased compared with the original model when removing the population change (MAE increases 17.264% and RMSE increases 15.683%). Removing wildfire-related information, i.e., evacuation order/warning, showcased considerable predictive power, as results demonstrated that it accounted for 20.724% and 14.183% of importance calculated by MAE and RMSE. Removing fire distance also notably increased the prediction error. Specifically, the predictive importance of fire distance calculated by MAE and RMSE was 12.537% and 10.425%, respectively. A relatively small increase in prediction error (less than 2% for MAE and RMSE) occurred when we removed the environmental similarity or demographic similarity graph. This finding indicated that although graphs can improve the model performance, their contributions are relatively limited. Weather information accounted for a relatively small strength in terms of forecasting travel demand (7.425% for MAE and 10.431% for RMSE). Two historical embeddings (i.e., historical travel demand information) displayed similar contributions in predicting travel demand under wildfire evacuations. Specifically, historical embeddings 1, i.e., the entire historical average of travel demand only accounted for around 7.905% of predictive importance regarding MAE and 10.473% regarding RMSE. Historical embeddings 2, i.e., the last-four-hour average of travel demand, contributed 6.985% and 11.273% of predictive importance calculated by MAE and RMSE.

6. Discussion and conclusion

In this study, we developed a Situational-Aware Multiple Graph Convolutional Recurrent Network (SA-MGCRN) model to forecast the real-time travel demand in wildfire evacuations while accounting for data delivery delay issues. The proposed model incorporates multiple dimensions of information including historical census-tract-level travel demand, temporal features (i.e., weather, evacuation order/warning, fire progression, weekend indicator and population change), and spatial features (i.e., socio-demographics similarity graph and built environment similarity graph). The proposed model, which is based on the framework in Fig. 3, uses these inputs to make accurate predictions on an hourly basis for the following day. Given the difficulties of having reliable traffic data during wildfire emergencies (Melendez et al., 2021), this study utilized GPS data collected from mobile devices to derive trips (for daily travel or evacuation purposes) for each census tract using the clustering approach illustrated in Fig. 2.

Using GPS data for real-time trip generation forecasting in wildfire scenarios is a novel and promising approach. This type of information provides emergency managers and planners with critical and reliable information on what is happening in the areas of interest and facilitates them to decide where and which type of emergency management strategies (e.g., traffic signal retiming, extending the area under evacuation order or activating contraflow) should be prioritized to keep mobility under emergent scenarios. Furthermore, the proposed SA-MGCRN model can provide them with accurate simulations (i.e., predictions) regarding the pattern of travel demand in the following day and therefore help pre-evaluate the effectiveness of the possible decisions. As such, this work takes a substantial step forward in the literature on the use of GPS data for wildfire evacuation management. To date, GPS historical data have been mostly used only to investigate wildfire evacuation after several months from the events (Zhao et al., 2022). Therefore, this work represents the first attempt to use GPS big data and AI to provide near real-time forecasting of human behavior to emergency managers and evacuation modelers during a wildfire disaster. In fact, the output of the SA-MGCRN can be used as an input for existing wildfire evacuation models for near real-time evacuation simulations of different if-then scenarios.

SA-MGCRN has shown strong capabilities in coping with spatial and temporal dependencies. Specifically, SA-MGCRN first employs GCN model to account for spatial correlations between census tracts using several graphs (i.e. environmental similarity graph and demographic similarity graph). Then, a GRU model is integrated to extract temporal dependencies behind multiple temporal variables including historical demand, weather, evacuation order/warning information, fire progression, weekend indicator and population change. The case study of 2019 Kincade Fire showed that SA-MGCRN outperformed all state-of-the-art benchmark models (see Tables 2, 3 and Fig. 9). In real-world applications, the GPS data usually suffer delivery delay issues, which may impede the real-time decision-making process. This study accounts for this issue by considering both delay and non-delay scenarios. Results showed that more data delivery delays will trigger an accuracy drop and our proposed model, i.e., SA-MGCRN, was still the best-performing model among both statistical and machine learning models.

Previous studies have highlighted the critical role of social cues and crisis communication in determining protective action decision-making (Lindell and Perry, 2012), or more specifically, evacuation behavior (Kuligowski et al., 2022; Lovreglio et al., 2019). Our study further reinforces this finding by demonstrating that the evacuation order and warning information is indeed a critical (the third most important) predictor for accuracy in forecasting travel demand under wildfire evacuation. In addition, this work showcased the strong relevance of the fire distance when forecasting traffic demand during wildfire evacuations, which aligns with a recent study (Wu et al., 2022). While another previous work by Lovreglio et al. (2020) did not show evidence that this variable was significant in the decision to evacuate. This can be possibly attributed to the difference in the adopted approaches. Lovreglio et al. (2020) used a statistical model to assess the impact of the distance between the households and the *final* firefront. However, this study used a deep learning approach and considered that the fire distance changes dynamically on a *daily* basis depending on the new fire front location. We believe that fire distance changes quickly throughout the fire progression and naturally influences the household's decision to evacuate or not. Our proposed model can well capture these spatiotemporal dynamics and therefore highly relies on this component. Besides fire distance, the results revealed that the other dynamic features such as population change and evacuation order/warning information, play more important roles than the static features like environment and demographic factors. This finding aligns with many previous studies, which have indicated that in (near) real-time travel demand prediction tasks, dynamic features typically contribute more significantly to predictive performance (Xu et al., 2023, 2024; He and Shin, 2020; Song et al., 2023). We believe that with the model updating scheme, the model can learn and adapt to the evolving wildfire patterns through the dynamic features, making them more critical for accurate predictions.

The predictive strength of SA-MGCRN, together with the proposed model updating scheme, hold implications for both practical and policy applications. For example, the SA-MGCRN model can significantly enhance traffic management during wildfire scenarios if combined with traffic models (Rohaert et al., 2023b). Previous studies have found that the traffic moved slower during evacuation scenarios than routine scenarios, due to the surge of evacuation flow and the reduction of the road capacity (Rohaert et al., 2023b; Dixit and Wolshon, 2014; Rahman et al., 2021). Emergency managers can use the real-time travel demand predictions of SA-MGCRN to enhance existing traffic prediction models (such as DGCN-LSTM proposed in Rahman and Hasan (2023)) or traffic simulators (such as Chen et al. (2020)). This allows for a more comprehensive understanding of the potential traffic dynamics (i.e., speed, flow, and capacity) during wildfire evacuations on major evacuation hotspots or corridors (Rohaert et al., 2023a). Accordingly, they may develop plans to identify traffic flow bottlenecks, optimize traffic signal timing and implement contraflow operations more effectively, which ensures reduced congestion and faster movement (Chen et al., 2020; Jha et al., 2004; Parr and Kaisar, 2011; Xie and Turnquist, 2011). Accurate travel demand predictions generated by SA-MGCRN can be used to facilitate future wildfire evacuation planning (Murray-Tuite and Wolshon, 2013; Pel et al., 2012). For example, the real-time trip generation predictions allow emergency mangers to dynamically delineate evacuation zones, assign evacuation routes, and issue evacuation orders or warnings based on current traffic conditions, travel demand patterns and projected fire spread. Also, the travel demand predictions can serve as essential inputs for wildfire evacuation planning tools, e.g., agent-based models (Wahlqvist et al., 2021; Barnes et al., 2021; Siam et al., 2022), which take the evacuation demand or travel demand as a key component. The powerful capability to forecast travel demand also allows authorities to communicate more effectively with the public. Department of Transportation (DOT) of multiple states tried to enhance the existing Intelligent Transportation System for use in hurricane evacuations to disseminate travel information to evacuees (Wolshon et al., 2005). By forecasting travel patterns and identifying traffic hotspots during wildfire evacuations, emergency planners can provide timely information about expected traffic conditions and advise alternative evacuation routes, which will collectively enhance public compliance and safety (Pel et al., 2010).

Overall, our findings showed that the proposed methodological framework can generate highly accurate predictions to facilitate real-time emergency management. However, there are several limitations that require follow-up work. For example, we observed that the trip generations estimated by GPS data are relatively less than travel surveys. This could somewhat introduce immeasurable biases in the follow-up modeling process. Future research may consider exploring the underlying reasons for this type of underestimation and develop solutions to account for the biases. In this study, fire cues were accounted for just by using the distance from the fire perimeter. However, wildfire generates smoke and embers, which can reach areas before the fire front reaches them, depending on the weather conditions (Kochanski et al., 2019; Ronchi et al., 2017). Also, this study did not include social influence as a component of SA-MGCRN, as capturing social interactions by solely using large-scale GPS data can be challenging. In fact, social influence plays a crucial role in determining evacuation behavior as people may respond when others are responding (Lovreglio et al., 2016; Reneke and Reneke, 2013; Zhang et al., 2024b). In future work, solutions need to be developed to include social influence characteristics into the current modeling framework to further improve prediction accuracy. Additionally, the fire distance is measured by the distance from the centroid of each census tract to the fire front instead of the actual distance from a householder to the fire front. This may create potential uncertainties. Fig. 9 showed that SA-MGCRN has relatively limited capabilities to deal with extreme values. One possible explanation is that the size of the training sample is relatively small ($n = 624$). We believe that with more observations, SA-MGNRN can further address the extreme values and produces even higher prediction accuracy. Notably, SA-MGCRN has the worst predictive performance at the start of the wildfire, due to limited previous wildfire-related information for training. To address this issue, future studies may consider integrating human intelligence into machine learning (i.e., human-in-the-loop AI Monarch, 2021) to further reduce the prediction error (Zhang and Zhao, 2022). Furthermore, fairness-enhancing techniques can be included in the model to make sure every population group's travel demand is accurately modeled (Zhang et al., 2024a). Another limitation is that we only used one case study to evaluate our proposed model's performance and the results may not be directly transferable to other contexts (Pel et al., 2012). Future studies may consider designing methods to account for the transferability issues and using transfer-learning techniques to transfer the useful information learned from our case study to the future studied ones (Rahman and Hasan, 2023).

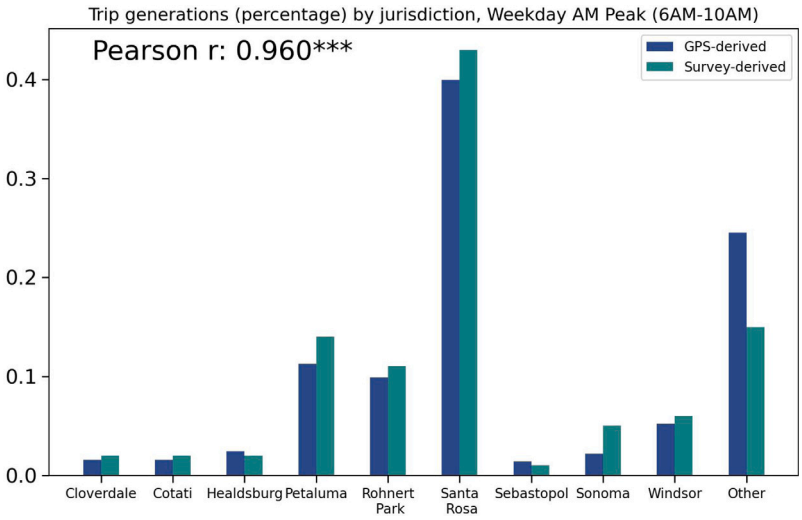

Fig. A.1. Trip generations (percentage) by jurisdiction on weekday AM peak.

CRediT authorship contribution statement

Xiaojian Zhang: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Xilei Zhao:** Writing – original draft, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Yiming Xu:** Visualization, Methodology, Formal analysis, Data curation. **Daniel Nilsson:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Ruggiero Lovreglio:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization.

Acknowledgments

This work was performed under the following financial assistance award No. 60NANB20D182, 60NANB21D180, 60NANB22D177, and 60NANB23D222 from U.S. Department of Commerce, National Institute of Standards and Technology (NIST). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NIST. We would like to thank Ruoyang Xiong for his contributions to the previous draft of the paper. During the preparation of this work the authors used ChatGPT in order to check grammar errors and improving the language. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Appendix A

Figs. A.1 and A.2 present the comparison of the percentage of trip generations by jurisdiction during the weekday morning peak and weekday evening peak hours. These comparisons are based on data obtained from both SCTBS and our own dataset.

Appendix B

We tested the CPU vs GPU inference time for predicting one-day travel demand and the training time per epoch, and considered with/without data delivery delay issues. The results are shown in Table B.1. As we may have expected, a slightly worse model compensates for its computational advantages. As shown in this table, the single-layer MLP exhibits the fastest computational speed for both training and inference, across both CPU and GPU settings. With the increasing complexity of the model architecture, the training and inference times also increase. This is evident in the performance of the LSTM and SA-MGCRN models, which require more computational resources due to their more sophisticated structures. We also investigated the effect of data delivery delay on the training and inference times. Our results indicate that the training and inference times are reduced when delays occur. This is because less data is available for training during these periods, which subsequently reduces the computational load.

Practitioners can choose between models based on their available computational resources and specific needs. For instance, a relatively simpler model, with a faster inference time, is ideal for environments with limited computational power or where rapid predictions are essential. Meanwhile, the SA-MGCRN model, although slightly slower, still enjoys a decent computational speed regarding both training (i.e., model updating) and inference (i.e., predictions), offers superior accuracy and is better suited for scenarios requiring detailed and accurate travel demand predictions.

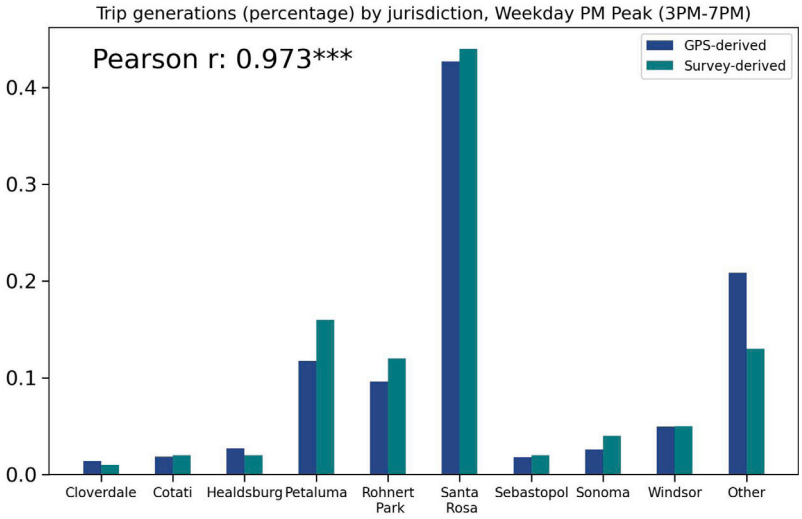

Fig. A.2. Trip generations (percentage) by jurisdiction on weekday PM peak.

Table B.1
CPU vs. GPU average per epoch training time and inference time (Unit: $\times 10^{-2}$ s).

	CPU		GPU	
	Training per epoch	Inference	Training per epoch	Inference
MLP	2.516	0.505	2.491	0.898
LSTM	19.462	2.696	18.792	1.878
SA-MGCRN (Delay = 0 h)	41.476	8.891	41.686	7.520
SA-MGCRN (Delay = 24 h)	32.261	4.734	30.152	4.274
SA-MGCRN (Delay = 48 h)	32.272	6.064	31.075	4.399

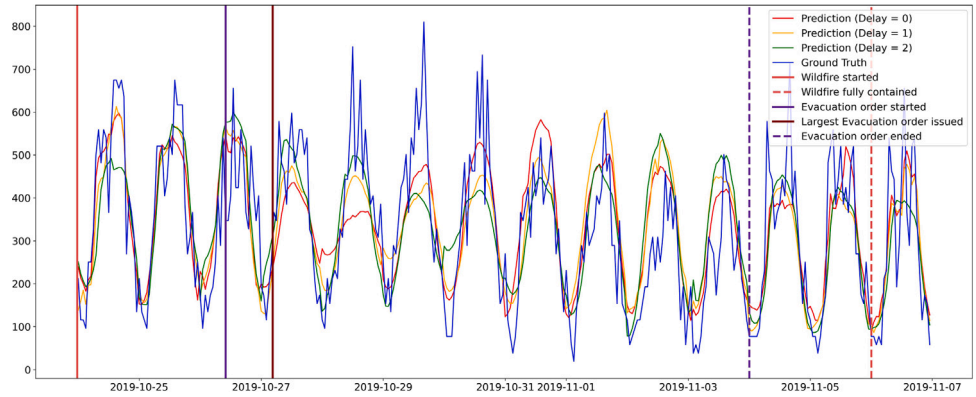

Fig. C.1. Hourly prediction performance of SA-MGCRN considering data delivery delay for census tract ID: 6097153200. Please refer to the census tract with highlighted boundary in Fig. 7.

Appendix C

The following two figures present the hourly prediction performance of SA-MGCRN considering data delivery delay for another two census tracts here for better illustrating the how model performance changes as updating goes on.

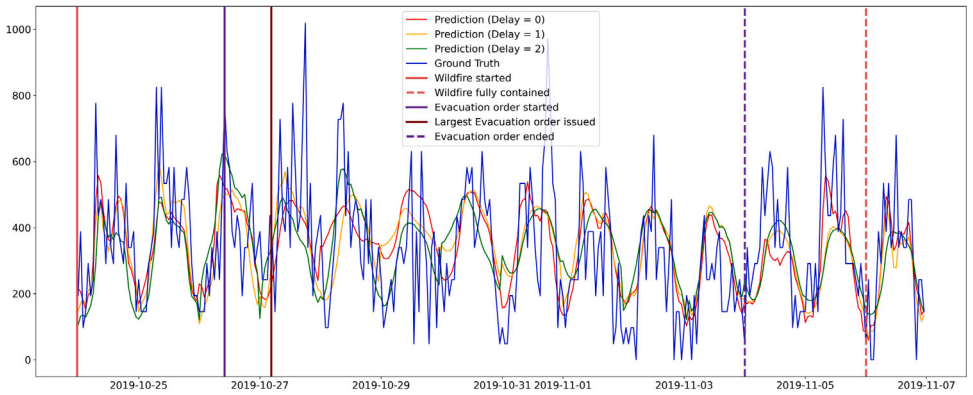

Fig. C.2. Hourly prediction performance of SA-MGCRN considering data delivery delay for census tract ID: 6097151311. Please refer to the census tract with highlighted boundary in Fig. 7.

References

Adler, T., Bernardin, V., Dumont, J., Flake, L., Sadrsadat, H., et al., 2017. The Promise and Limitations of Locational App Data for Origin-Destination Analysis: A Case Study. Technical Report, United States. Federal Highway Administration.

Alexander, L., Jiang, S., Murga, M., González, M.C., 2015. Origin–destination trips by purpose and time of day inferred from mobile phone data. Transp. Res. C 58, 240–250.

Analytics, G., 2023. Gravy analytics. URL: https://gravyanalytics.com/our-location-data/. (Accessed 25 September 2023).

Barnes, B., Dunn, S., Pearson, C., Wilkinson, S., 2021. Improving human behaviour in macroscale city evacuation agent-based simulation. Int. J. Disaster Risk Reduct. 60, 102289.

Bergado, J.R., Persello, C., Reinke, K., Stein, A., 2021. Predicting wildfire burns from big geodata using deep learning. Saf. Sci. 140, 105276.

Chen, C., Bian, L., Ma, J., 2014. From traces to trajectories: How well can we guess activity locations from mobile phone traces? Transp. Res. C 46, 326–337.

Chen, Y., Shafi, S.Y., Chen, Y.-f., 2020. Simulation pipeline for traffic evacuation in urban areas and emergency traffic management policy improvements through case studies. Transp. Res. Interdiscip. Perspect. 7, 100210.

Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y., 2014. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.

Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Coleman, N., Gao, X., DeLeon, J., Mostafavi, A., 2022. Human activity and mobility data reveal disparities in exposure risk reduction indicators among socially vulnerable populations during COVID-19 for five US metropolitan cities. Sci. Rep. 12 (1), 15814.

Datarade, 2023. Mobile location data - data category. https://datarade.ai/data-categories/mobile-location-data. (Accessed 06 November 2023).

Dixit, V., Wolshon, B., 2014. Evacuation traffic dynamics. Transp. Res. C 49, 114–125.

Folk, L.H., Kuligowski, E.D., Gwynne, S., Gales, J.A., 2019. A provisional conceptual model of human behavior in response to wildland-urban interface fires. Fire Technol. 55 (5), 1619–1647.

Fu, H., Wilmot, C.G., Zhang, H., Baker, E.J., 2007. Modeling the hurricane evacuation response curve. Transp. Res. Rec. 2022 (1), 94–102.

Haghani, M., Kuligowski, E., Rajabifard, A., Kolden, C.A., 2022. The state of wildfire and bushfire science: temporal trends, research divisions and knowledge gaps. Saf. Sci. 153, 105797.

He, S., Shin, K.G., 2020. Dynamic flow distribution prediction for urban dockless e-scooter sharing reconfiguration. In: Proceedings of the Web Conference 2020. pp. 133–143.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Horanont, T., Witayangkurn, A., Sekimoto, Y., Shibasaki, R., 2013. Large-scale auto-GPS analysis for discerning behavior change during crisis. IEEE Intell. Syst. 28 (4), 26–34.

Hu, S., Xiong, C., Yang, M., Younes, H., Luo, W., Zhang, L., 2021. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. Transp. Res. C 124, 102955.

Huang, D., Wang, S., Liu, Z., 2021. A systematic review of prediction methods for emergency management. Int. J. Disaster Risk Reduct. 62, 102412.

Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: Principles and Practice. OTexts.

Jha, M., Moore, K., Pashaie, B., 2004. Emergency evacuation planning with microscopic traffic simulation. Transp. Res. Rec. 1886 (1), 40–48.

Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Kochanski, A.K., Mallia, D.V., Fearon, M.G., Mandel, J., Souri, A.H., Brown, T., 2019. Modeling wildfire smoke feedback mechanisms using a coupled fire-atmosphere model with a radiatively active aerosol scheme. J. Geophys. Res.: Atmos. 124 (16), 9099–9116.

Kolmogorov, A., 1933. Sulla determinazione empirica di una lgge di distribuzione—ScienceOpen. Inst. Ital. Attuari, giorn.

Kuligowski, E., 2021. Evacuation decision-making and behavior in wildfires: Past research, current challenges and a future research agenda. Fire Saf. J. 120, 103129.

Kuligowski, E.D., Walpole, E.H., Lovreglio, R., McCaffrey, S., 2020. Modelling evacuation decision-making in the 2016 chimney tops 2 fire in Gatlinburg, TN. Int. J. Wildland Fire 29 (12), 1120–1132.

Kuligowski, E.D., Zhao, X., Lovreglio, R., Xu, N., Yang, K., Westbury, A., Nilsson, D., Brown, N., 2022. Modeling evacuation decisions in the 2019 kincade fire in California. Saf. Sci. 146, 105541.

Li, X., Hasan, S., Culotta, A., 2022. Identifying hurricane evacuation intent on Twitter. In: Proceedings of the International AAAI Conference on Web and Social Media. Vol. 16, pp. 618–627.

Lindell, M.K., Perry, R.W., 2012. The protective action decision model: Theoretical modifications and additional evidence. Risk Anal.: Int. J. 32 (4), 616–632.

Lo, S., Liu, M., Zhang, P., Yuen, R.K., 2009. An artificial neural-network based predictive model for pre-evacuation human response in domestic building fire. Fire Technol. 45 (4), 431–449.

Lovreglio, R., Kuligowski, E., Gwynne, S., Strahan, K., 2019. A modelling framework for householder decision-making for wildfire emergencies. Int. J. Disaster Risk Reduct. 41, 101274.
Lovreglio, R., Kuligowski, E., Walpole, E., Link, E., Gwynne, S., 2020. Calibrating the wildfire decision model using hybrid choice modelling. Int. J. Disaster Risk Reduct. 50, 101770.
Lovreglio, R., Ronchi, E., Nilsson, D., 2015. A model of the decision-making process during pre-evacuation. Fire Saf. J. 78, 168–179.
Lovreglio, R., Ronchi, E., Nilsson, D., 2016. An evacuation decision model based on perceived risk, social influence and behavioural uncertainty. Simul. Model. Pract. Theory 66, 226–242.
McCaffrey, S., Wilson, R., Konar, A., 2018. Should I stay or should I go now? Or should I wait and see? Influences on wildfire evacuation decisions. Risk Anal. 38 (7), 1390–1404.
McGhee, C.C., Grimes, M.C., et al., 2006. An Operational Analysis of the Hampton Roads Hurricane Evacuation Traffic Control Plan. Technical Report, Virginia Transportation Research Council.
McLennan, J., Elliott, G., Omodei, M., 2011. Issues in Community Bushfire Safety: Analyses of Interviews Conducted by the 2009 Victorian Bushfires Research Task Force. Bundoora, AU.
McLennan, J., Elliott, G., Omodei, M., 2012. Householder decision-making under imminent wildfire threat: stay and defend or leave? Int. J. Wildland Fire 21 (7), 915–925.
Melendez, B., Machiani, S.G., Nara, A., 2021. Modelling traffic during lilac wildfire evacuation using cellular data. Transp. Res. Interdiscip. Perspect. 9, 100335.
Meyes, R., Lu, M., de Puiseau, C.W., Meisen, T., 2019. Ablation studies in artificial neural networks. arXiv preprint arXiv:1901.08644.
Molnar, C., 2020. Interpretable Machine Learning. Lulu. com.
Monarch, R.M., 2021. Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI. Simon and Schuster.
Monz, C., Creany, N., Nesbitt, J., Mitrovich, M., 2021. Mobile device data analysis to determine the demographics of park visitors. J. Park Recreat. Adm. 39 (1).
Mozumder, P., Raheem, N., Talberth, J., Berrens, R.P., 2008. Investigating intended evacuation from wildfires in the wildland–urban interface: application of a bivariate probit model. For. Policy Econ. 10 (6), 415–423.
Murray-Tuite, P., Wolshon, B., 2013. Evacuation transportation modeling: An overview of research, development, and practice. Transp. Res. C 27, 25–45.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Icml.
Nguyen, L., Yang, Z., Li, J., Pan, Z., Cao, G., Jin, F., 2019. Forecasting people's needs in hurricane events from social network. IEEE Trans. Big Data 8 (1), 229–240.
Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S., 2018. Activation functions: Comparison of trends in practice and research for deep learning. arXiv preprint arXiv:1811.03378.
Parr, S.A., Kaisar, E., 2011. Critical intersection signal optimization during urban evacuation utilizing dynamic programming. J. Transp. Saf. Secur. 3 (1), 59–76.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.
Paveglio, T., Prato, T., Dalenberg, D., Venn, T., 2014. Understanding evacuation preferences and wildfire mitigations among northwest Montana residents. Int. J. Wildland Fire 23 (3), 435–444.
Pel, A.J., Bliemer, M.C., Hoogendoorn, S.P., 2012. A review on travel behaviour modelling in dynamic traffic simulation models for evacuations. Transportation 39, 97–123.
Pel, A.J., Hoogendoorn, S.P., Bliemer, M.C., 2010. Evacuation modeling including traveler information and compliance behavior. Procedia Eng. 3, 101–111.
Pepe, E., Bajardi, P., Gauvin, L., Privitera, F., Lake, B., Cattuto, C., Tizzoni, M., 2020. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. Sci. Data 7 (1), 230.
Radeloff, V.C., Helmers, D.P., Kramer, H.A., Mockrin, M.H., Alexandre, P.M., Bar-Massada, A., Butsic, V., Hawbaker, T.J., Martinuzzi, S., Syphard, A.D., et al., 2018. Rapid growth of the US wildland-urban interface raises wildfire risk. Proc. Natl. Acad. Sci. 115 (13), 3314–3319.
Rahman, R., Bhowmik, T., Eluru, N., Hasan, S., 2021. Assessing the crash risks of evacuation: A matched case-control approach applied over data collected during Hurricane Irma. Accid. Anal. Prev. 159, 106260.
Rahman, R., Hasan, S., 2018. Short-term traffic speed prediction for freeways during hurricane evacuation: a deep learning approach. In: 2018 21st International Conference on Intelligent Transportation Systems. ITSC, IEEE, pp. 1291–1296.
Rahman, R., Hasan, S., 2023. A deep learning approach for network-wide dynamic traffic prediction during hurricane evacuation. Transp. Res. C 152, 104126.
Reneke, P.A., Reneke, P.A., 2013. Evacuation Decision Model. US Department of Commerce, National Institute of Standards and Technology ....
Rohaert, A., Janfeshanaraghi, N., Kuligowski, E., Ronchi, E., 2023a. The analysis of traffic data of wildfire evacuation: the case study of the 2020 glass fire. Fire Saf. J. 141, 103909.
Rohaert, A., Kuligowski, E.D., Ardinge, A., Wahlqvist, J., Gwynne, S.M., Kimball, A., Bénichou, N., Ronchi, E., 2023b. Traffic dynamics during the 2019 kincade wildfire evacuation. Transp. Res. D 116, 103610.
Ronchi, E., Gwynne, S.M., Rein, G., Wadhwani, R., Intini, P., Bergstedt, A., 2017. e-Sanctuary: Open Multi-Physics Framework for Modelling Wildfire Urban Evacuation. Fire Protection Research Foundation Quincy.
Roy, K.C., Hasan, S., Culotta, A., Eluru, N., 2021. Predicting traffic demand during hurricane evacuation using real-time data from transportation systems and social media. Transp. Res. C 131, 103339.
Shahraki, A., Abbasi, M., Taherkordi, A., Jurcut, A.D., 2022. A comparative study on online machine learning techniques for network traffic streams analysis. Comput. Netw. 207, 108836.
Siam, M.R., Wang, H., Lindell, M.K., Chen, C., Vlahogianni, E.I., Axhausen, K., 2022. An interdisciplinary agent-based multimodal wildfire evacuation model: Critical decisions and life safety. Transp. Res. D 103, 103147.
Sinclair, M., Maadi, S., Zhao, Q., Hong, J., Ghermandi, A., Bailey, N., 2023. Assessing the socio-demographic representativeness of mobile phone application data. Appl. Geogr. 158, 102997.
Smart Traffic, 2021. Location-based services apps. URL: https://www.smatstraffic.com/2021/08/23/location-data/#:~:text=Location%2DBased%20Services%20Apps,-Location%20data%20coming&text=This%20data%20updates%20every%201,frequency%20location%2Dbased%20services%20apps. (Accessed 25 September 2023).
Smirnov, N., 1948. Table for estimating the goodness of fit of empirical distributions. Ann. Math. Stat. 19 (2), 279–281.
Song, J.-C., Hsieh, I.-Y.L., Chen, C.-S., 2023. Sparse trip demand prediction for shared e-scooter using spatio-temporal graph neural networks. Transp. Res. D 125, 103962.
Sonoma County Transportation Authority, 2023. Sonoma county travel behavior study. URL: https://scta.ca.gov/planning/sonoma-county-travel-behavior-study/. (Accessed 21 September 2023).
Sonoma Operational Area and the County of Sonoma, Department of Emergency Management, 2020. 2019 Kincade fire after action report. In: 2019 Kincade Fire After-Action Report and Improvement Plan.
Strawderman, L., Salehi, A., Babski-Reeves, K., Thornton-Neaves, T., Cosby, A., 2012. Reverse 911 as a complementary evacuation warning system. Nat. Hazards Rev. 13 (1), 65–73.
Sun, Y., Huang, S.-K., Zhao, X., 2024a. Predicting hurricane evacuation decisions with interpretable machine learning methods. Int. J. Disaster Risk Sci. 15 (1), 134–148.

Sun, Y., Zhao, X., Lovreglio, R., Kuligowski, E., 2024b. AI for large-scale evacuation modeling: promises and challenges. In: Interpretable Machine Learning for the Analysis Design Assessment and Informed Decision Making for Civil Infrastructure. Elsevier, pp. 185–204.
Tang, J., Gao, F., Liu, F., Zhang, W., Qi, Y., 2019. Understanding spatio-temporal characteristics of urban travel demand based on the combination of GWR and GLM. Sustainability 11 (19), 5525.
Tang, J., Liang, J., Liu, F., Hao, J., Wang, Y., 2021. Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network. Transp. Res. C 124, 102951.
Templeton, A., Nash, C., Spearpoint, M., Gwynne, S., Hui, X., Arnott, M., 2023. Who and what is trusted in fire incidents? The role of trust in guidance and guidance creators in resident response to fire incidents in high-rise residential buildings. Saf. Sci. 164, 106172.
Thy Vo, D.G., Prentzel, O., 2022. Marshall fire officially becomes Colorado's most destructive, with 991 homes and businesses burned, officials confirm. Colo. Sun URL: https://coloradosun.com/2022/01/01/marshall-fire-snow-displaced-residents-power-outage/.
Turner, S., Tsapakis, I., Koeneman, P., et al., 2020. Evaluation of StreetLight Data's Traffic Count Estimates from Mobile Device Data. Technical Report, Minnesota. Dept. of Transportation. Office of Policy Analysis, Research ....
United States Department of Transportation, 2023. National household travel survey daily travel quick facts. URL: https://www.bts.gov/statistical-products/surveys/national-household-travel-survey-daily-travel-quick-facts. (Accessed 21 September 2023).
Vaiciulyte, S., Hulse, L.M., Galea, E.R., Veeraswamy, A., 2022. Exploring 'wait and see'responses in french and Australian WUI wildfire emergencies. Saf. Sci. 155, 105866.
Wahlqvist, J., Ronchi, E., Gwynne, S.M., Kinateder, M., Rein, G., Mitchell, H., Bénichou, N., Ma, C., Kimball, A., Kuligowski, E., 2021. The simulation of wildland-urban interface fire evacuation: The WUI-NITY platform. Saf. Sci. 136, 105145.
Wand, Y., Wang, R.Y., 1996. Anchoring data quality dimensions in ontological foundations. Commun. ACM 39 (11), 86–95.
Wang, F., Chen, C., 2018. On data processing required to derive mobility patterns from passively-generated mobile phone data. Transp. Res. C 87, 58–74.
Wang, J., Wang, F., Ban, X.J., Chen, C., 2019b. Comparative Analysis of Big and Small (Survey) Data for Deriving Human Mobility Patterns. Technical Report.
Wang, F., Wang, J., Cao, J., Chen, C., Ban, X.J., 2019a. Extracting trips from multi-sourced data for mobility pattern analysis: An app-based data example. Transp. Res. C 105, 183–202.
Whittaker, J., Eriksen, C., Haynes, K., 2016. Gendered responses to the 2009 black saturday bushfires in Victoria, australia. Geogr. Res. 54 (2), 203–215.
Wolshon, B., Urbina Hamilton, E., Levitan, M., Wilmot, C., 2005. Review of policies and practices for hurricane evacuation. II: Traffic operations, management, and control. Nat. Hazards Rev. 6 (3), 143–161.
Wong, S.D., Broader, J.C., Walker, J.L., Shaheen, S.A., 2022. Understanding California wildfire evacuee behavior and joint choice making. Transportation 1–47.
Wu, Y., Mooring, T.A., Linz, M., 2021. Policy and weather influences on mobility during the early US COVID-19 pandemic. Proc. Natl. Acad. Sci. 118 (22), e2018185118.
Wu, A., Yan, X., Kuligowski, E., Lovreglio, R., Nilsson, D., Cova, T.J., Xu, Y., Zhao, X., 2022. Wildfire evacuation decision modeling using GPS data. Int. J. Disaster Risk Reduct. 83, 103373.
Xie, C., Turnquist, M.A., 2011. Lane-based evacuation network optimization: An integrated Lagrangian relaxation and tabu search approach. Transp. Res. C 19 (1), 40–63.
Xie, C., Yu, D., Zheng, X., Wang, Z., Jiang, Z., 2021. Revealing spatiotemporal travel demand and community structure characteristics with taxi trip data: A case study of New York city. PLoS ONE 16 (11), e0259694.
Xiong, C., Hu, S., Yang, M., Luo, W., Zhang, L., 2020. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. Proc. Natl. Acad. Sci. 117 (44), 27087–27089.
Xu, Y., Ke, Q., Zhang, X., Zhao, X., 2024. ICN: Interactive convolutional network for forecasting travel demand of shared micromobility. GeoInformatica 1–26.
Xu, Y., Yan, X., Liu, X., Zhao, X., 2021. Identifying key factors associated with ridesplitting adoption rate and modeling their nonlinear relationships. Transp. Res. A 144, 170–188.
Xu, Y., Zhao, X., Zhang, X., Paliwal, M., 2023. Real-time forecasting of dockless scooter-sharing demand: A spatio-temporal multi-graph transformer approach. IEEE Trans. Intell. Transp. Syst..
Yabe, T., Ukkusuri, S.V., 2020. Effects of income inequality on evacuation, reentry and segregation after disasters. Transp. Res. D 82, 102260.
Zeigler, D.J., Brunn, S.D., Johnson, Jr., J.H., 1981. Evacuation from a nuclear technological disaster. Geogr. Rev. 1–16.
Zhang, X., Ke, Q., Zhao, X., 2024a. Travel demand forecasting: A fair AI approach. IEEE Trans. Intell. Transp. Syst..
Zhang, Z., Manjourides, J., Cohen, T., Hu, Y., Jiang, Q., 2016. Spatial measurement errors in the field of spatial epidemiology. Int. J. Health Geogr. 15 (1), 1–12.
Zhang, X., Zhao, X., 2022. Machine learning approach for spatial modeling of ridesourcing demand. J. Transp. Geogr. 100, 103310.
Zhang, X., Zhao, X., Baldwin, D., McBride, S., Bellizzi, J., Cochran, E.S., Luco, N., Wood, M., Cova, T.J., 2024b. Modeling protective action decision-making in earthquakes by using explainable machine learning and video data. Sci. Rep. 14 (1), 5480.
Zhang, X., Zhou, Z., Xu, Y., Zhao, X., 2024c. Analyzing spatial heterogeneity of ridesourcing usage determinants using explainable machine learning. J. Transp. Geogr. 114, 103782.
Zhao, X., Lovreglio, R., Kuligowski, E., Nilsson, D., 2021a. Using artificial intelligence for safe and effective wildfire evacuations. Fire Technol. 57 (2), 483–485.
Zhao, X., Lovreglio, R., Nilsson, D., 2020a. Modelling and interpreting pre-evacuation decision-making using machine learning. Autom. Constr. 113, 103140.
Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., Li, H., 2019. T-gcn: A temporal graph convolutional network for traffic prediction. IEEE Trans. Intell. Transp. Syst. 21 (9), 3848–3858.
Zhao, X., Xu, Y., Lovreglio, R., Kuligowski, E., Nilsson, D., Cova, T.J., Wu, A., Yan, X., 2022. Estimating wildfire evacuation decision and departure timing using large-scale GPS data. Transp. Res. D 107, 103277.
Zhao, X., Xu, N., Yang, K., Kuligowski, E., Lovreglio, R., Nilsson, D., Brown, N.A., 2021b. Modeling evacuation behavior in the 2019 kincade fire, Sonoma County, California. In: Natural Hazards Center Quick Response Grant Report Series. Vol. 326.
Zhao, X., Yan, X., Yu, A., Van Hentenryck, P., 2020b. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. Travel Behav. Soc. 20, 22–35.