Are the data good enough? Spatial and temporal modeling of evacuee behavior using GPS data in a small rural community

Raei, Bahareh; Kinateder, Max; Bénichou, Noureddine; Gomaa, Islam; Wang, Xin

Publisher's version / Version de l'éditeur:

https://doi.org/10.1016/j.ijdrr.2024.105054

International Journal of Disaster Risk Reduction, 116, C, 2024-12-20

International Journal of Disaster Risk Reduction 116 (2025) 105054

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

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

Are the data good enough? Spatial and temporal modeling of evacuee behavior using GPS data in a small rural community

Bahareh Raei [a], Max Kinateder [b], Noureddine Bénichou [b], Islam Gomaa [b], Xin Wang [a,*]

[a] University of Calgary, Calgary, Alberta, Canada
[b] National Research Council Canada, Ottawa, Ontario, Canada

ARTICLE INFO

Keywords:
Wildfire
GPS trajectories
Map visualization
Movement analysis
WUI
Evacuation

ABSTRACT

The growing frequency and intensity of wildfires pose significant challenges for Canadian communities in the Wildland-Urban Interface (WUI). This case study explores the evacuation dynamics (traffic movement) during the 2021 wildfire in Lytton, BC. Drawing upon Global Positioning System (GPS) data, the study proposes novel methodologies for departure time analysis and traffic prediction tailored to the local wildfire context. The methodology offers insights into the temporal and spatial movement patterns of residents within the community before and during the wildfire event. By employing stay point detection and temporal analysis techniques, the study quantifies evacuation behavior, shedding light on departure times and evacuation trends. In addition, the study presents a comprehensive methodology for predicting traffic dynamics on highways during wildfire evacuations beyond the case study. Leveraging GPS data and machine learning techniques, the proposed approach integrates spatial and temporal analyses with predictive modeling to forecast traffic conditions accurately. Our analysis revealed that the southbound exit roads to the highway experienced significant traffic congestion during the wildfire. Overall, this work contributes to our understanding of WUI community preparedness and evacuation by providing insights into evacuation behaviors, preferred routes, and potential traffic challenges. However, the results also clearly exposed the limitations of GPS data from smaller communities in sparsely populated areas. Further research is needed to enhance our comprehension of wildfire responses, especially within underserved communities in rural areas. This will allow for the improvement of predictive models and the creation of more efficient evacuation planning strategies. Such advancements are crucial in mitigating risks and ensuring the safety of WUI residants.

1. Introduction

The increasing threat of wildfires to both people and the environment is a serious concern due to global warming and climate change [1–4]. In recent years, Canada has witnessed a notable rise in the frequency, scale, and intensity of wildfires. These trends are projected to worsen, exacerbating their harmful impacts [5]. A prime example occurred in 2021, when Canada experienced a record high temperature of 49.5 °C (121.1 degrees Fahrenheit) in Lytton, BC. The average high temperature in this area for June is 24 degrees Celsius. This unprecedented heat was followed by a devastating wildfire that destroyed over 83,000 hectares of land and 151 homes [6].

* Corresponding author.
E-mail addresses: bahareh.mohammadraei@ucalgary.ca (B. Raei), max.kinateder@nrc-cnrc.gc.ca (M. Kinateder), noureddine.benichou@nrc-cnrc.gc.ca (N. Bénichou), Islam.Gomaa@nrc-cnrc.gc.ca (I. Gomaa), xcwang@ucalgary.ca (X. Wang).

https://doi.org/10.1016/j.ijdrr.2024.105054
Received 28 May 2024; Received in revised form 7 October 2024; Accepted 4 December 2024

Available online 20 December 2024

With the rising threat of wildfires, it is crucial for public agencies and communities to better understand evacuation behavior [7]. Effective wildfire evacuation management relies on capturing behavioral data, especially on roads, to estimate traffic conditions and prevent bottlenecks [8,9]. Accurate traffic management can enhance evacuation efficiency and optimize evacuation orders, thus improving community preparedness.

Traditionally, research in this field has relied on survey or interview data [3,10–12], which, while valuable, are often limited by memory bias and access constraints following large-scale evacuations [7]. Such limitations underscore the need for alternative methods of data collection to provide real-time, scalable insights into evacuation patterns. To address these challenges, this study employs GPS data to examine the evacuation behavior of residents in Lytton, BC. GPS data provides timestamped location information from mobile devices, offering a more comprehensive understanding of evacuation timing, routes, and destinations. This method is particularly beneficial in sparsely populated areas like Lytton, where survey-based approaches may not provide the necessary granularity [13].

Given these challenges, alternative means of collecting empirical data are needed. Global Positioning System (GPS) data has been proposed to better understand Wildland-Urban Interface (WUI) wildfire evacuations [9,13]. This method harnesses location data shared by mobile devices such as mobile phones with third parties. Timestamped location information of devices can be used as a proxy for people's movements, which enables us to gain insights into evacuation timing, destination, and routes. These data overcome some of the limitations of survey data and typically offer access to larger and more representative samples, allowing for a more comprehensive understanding of evacuation patterns [9,13]. However, GPS data has thus far been primarily employed in densely populated areas, e.g. [13–15]. In contrast, large parts of the WUI are relatively sparsely populated [16].

This paper presents several contributions to the field of disaster management and transportation planning, focusing on the analysis of evacuation behavior and the prediction of traffic dynamics during wildfire evacuations.

1. The study introduces a new departure time analysis methodology aimed at understanding evacuation behavior during wildfire events. The analysis combines temporal and spatial methodologies to investigate movement patterns of residents before and during a wildfire event. By employing stay point detection and temporal analysis techniques, departure times of residents are quantified. This departure time analysis offers insights into evacuation patterns, rates, and trends, providing a better understanding of the dynamics of evacuations in the face of natural disasters.
2. The paper presents a comprehensive methodology for predicting traffic dynamics during wildfire evacuations. Leveraging GPS data and machine learning techniques, the proposed methodology integrates spatial and temporal analyses of people movement with predictive modeling for traffic prediction. Specifically, the methodology encompasses preprocessing of GPS data, calculation of hourly flow on highways, synchronization of flow and speed data, training a linear regression model, and subsequent speed prediction. This approach enables the accurate forecasting of traffic conditions during wildfire events.
3. The developed methodologies were applied to a case study, namely the evacuation from the community of Lytton in British Columbia, Canada during the 2021 wildfire season.

The remainder of the paper is organized as follows: Section 2 provides an overview of related works. Section 3 describes the case study and data used in this research, detailing the context of the Lytton wildfire and the sources of GPS data utilized for departure time analysis and traffic prediction. Section 4 outlines the methodology employed for both the departure time analysis and traffic prediction. Section 5 presents the results of the analyses conducted, followed by a discussion of the findings and their implications and limitations and potential avenues for future works in Section 6. Finally, this paper concludes in Section 7 with a summary of the key findings.

2. Related works

Numerous studies have employed diverse datasets and modeling methodologies to forecast the decision-making process concerning evacuations during wildfire events. Many of these investigations rely on survey data [17–19]. For instance, Kuligowski et al. [18] examined the Kincade fire in Sonoma County, California, aiming to predict household perceptions and evacuation determinations. Employing regression analysis, they identified various factors influencing risk perception during evacuation decisions, encompassing pre-fire safety perceptions, household composition, income, education, and threat assessment. Furthermore, logistic regression analysis unveiled that risk perception, tenure, household composition, income, education, evacuation directives, fire signals, pre-fire safety perceptions, and home-ownership significantly influenced evacuation choices. Additionally, Xu et al. [19] conducted a study on the same case, employing various machine learning models with survey data. The findings indicated that all machine learning models outperformed logistic regression in predicting evacuation decisions. Zheng et al. [20] used the dynamic fan-grid partitioning method to simulate safe evacuation through building bottlenecks, and they found the fan-grid partitioning method effectively characterizes evacuation parameters and improves simulation accuracy for safe evacuations during emergencies. Li et al. [21] worked on the Tahoe Donner neighborhood in Truckee, California using various data sources, traffic simulation models, and geographic information systems to develop a data-driven wildfire evacuation model, integrating household vehicle ownership and second home occupancy rates to estimate evacuation times. They found evacuation time estimates vary significantly depending on the mean number of vehicles per home and the occupancy rate of second homes in resort areas.

2.1. GPS data in evacuation modeling

GPS data are ubiquitous in our everyday life and are used by millions of people, for example in mobile phone-based navigation assistants. In many cases, users of such apps are sharing their location data with third parties who make these data available to researchers and others. Typically, these data include timestamped, latitude, and longitude information about the devices being used, making them an interesting data source for understanding large-scale movement dynamics, such as those occurring during wildfire evacuation. There are many studies that discuss different analysis techniques that can be used to estimate evacuees' decisions on evacuation using GPS data, such as Wu et al. [13] who studied the 2019 Kincade wildfire, by analyzing GPS datasets collected from mobile devices comparing the results with survey data. They proposed a methodology to dedicate home locations of mobile device users and developed a regression model to examine how built environment variables affect evacuation rates. The results show that the use of GPS data is a valuable complement to existing methods. Zhao et al. [9] worked on the same case study, they used a new methodology to systematically identify different groups of wildfire evacuees. They found that self-evacuees and shadow evacuees, residents who begin evacuation before an official evacuation order is issued, accounted for more than half of the evacuees in that wildfire. Yabe et al. [22] analyzed GPS trajectories of mobile phones after major earthquakes in Japan. They found that an individual's likelihood of evacuating was strongly influenced by the seismic intensity they experienced, with consistent evacuation probabilities observed across different earthquake profiles and urban characteristics, primarily determined by seismic intensity. Moreover, the distance individuals evacuated was not dependent on the seismic intensities they experienced. Muhammad et al. [23] studied the tsunami event in Kochi, Japan using stochastic tsunami hazard modeling and agent-based evacuation simulations and found higher ground evacuation points effectively reduce the number of affected people.

2.2. Trajectories and movement analysis

GPS data has also been used to support management or routine and disaster evacuation traffic. Melendez et al. [24] worked on the 2017 Lilac wildfire in California, they explored cellular data to predict vehicular densities on evacuation routes in wildfire-prone regions. Rohaert et al. [25] collected traffic flow data during the Kincade Fire evacuation and found that during evacuations, vehicle speed decreased by approximately 3.5 km/h compared to routine traffic days, regardless of traffic density.

Kong et al. [26] introduced a robust approach using floating car trajectory data and compared two different optimization algorithms. This method showcased superiority in accuracy, immediacy, and stability compared to previous approaches [19], making it a promising tool to support traffic management. Additionally, Wang et al. [27] presented the visual analysis, and the model introduced by Ferreira et al. [28] provided interactive systems that leverage GPS trajectories for understanding and managing urban traffic congestion. These contributions collectively advance the understanding of evacuation strategies, traffic management, and urban dynamics during emergency scenarios. Zhang et al. [29] analyzed a post-concert egress event, introduced a destination choice model that incorporates spatiotemporal traffic resilience dynamics, guiding evacuees towards less congested routes and suitable destinations.

GPS data, when recorded at a sufficient level of detail, affords detailed trajectory and movement analysis. Likewise, Ahmad et al. [15] used connected vehicle data to evaluate traffic performance during evacuation events. This shift in focus underscores a more comprehensive approach to understanding and managing evacuations. In another innovative approach, Polakis and Tsouchlaraki [30] utilized GIS and GPS measurements to identify optimal pedestrian assembly points for residential settlement evacuations by integrating the Dijkstra algorithm for route optimization and developing a method for evacuating large crowds in groups.

These collaborative efforts demonstrate a shift towards incorporating more extensive and diverse data sources, as well as modeling trajectories and movements during environmental disasters, reflecting an advancement in the comprehension and management of evacuation dynamics during emergencies.

2.3. Wildfire evacuations vs. Other disaster evacuations

One critical distinction between wildfire evacuations and other types of disaster evacuations, such as hurricanes or floods, is the abrupt and unpredictable nature of wildfire departures. In hurricanes and floods, residents often receive advanced warnings, and evacuation orders are phased over several days, allowing for more gradual, organized departures, often modeled using S-curves to describe demand. In contrast, wildfires tend to force sudden, mass evacuations, where departure patterns are more erratic, and residents may have only minutes to evacuate. This makes wildfire evacuations more prone to congestion and bottlenecks, as seen in events like the 2017 Lilac and Kincade fires.

Yang et al. [31] applied machine learning models, including deep learning techniques, to predict traffic flow during emergency evacuations, which is highly relevant to wildfire scenarios. Their research showed that deep learning models performed better at predicting congestion compared to traditional statistical models, demonstrating the potential of advanced predictive techniques in disaster contexts. Anyidoho et al. [32] compared multiple models for predicting population behavior during hurricane evacuations using datasets from four hurricanes. Their study found that the Dynamic Discrete Choice (DDC) model performed best for predicting spatial and temporal evacuation patterns, while the Logistic Regression with S-curves (LR-S) model was most effective for estimating total evacuation rates. By incorporating Scurve methodologies, they provided a framework for forecasting evacuation timing in response to official orders, a useful approach that could potentially inform wildfire evacuation studies.

Similarly, Song and Yan [33] developed a disaster evacuation demand curve model based on the Susceptible-Infective (SI) model, originally used to simulate epidemic spread, to capture the influence of social contagion during evacuations. Their work highlights the importance of early and strong warning systems, as the model showed that issuing a second, more urgent evacuation order resulted in a substantial increase in evacuation rates, aligning with real-world observations from the Tianjin disaster. These studies demonstrate how social contagion and the timing of evacuation orders can dramatically influence evacuation patterns,

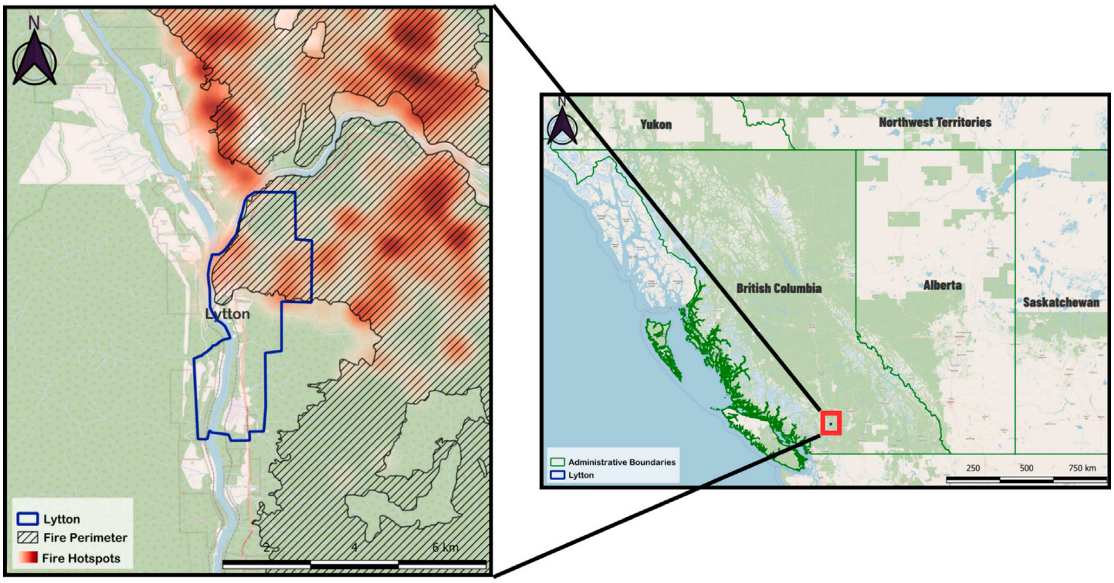

Fig. 1. Case study: Lytton Fire perimeter and hotspots heatmap.

which could be vital for refining models used in wildfire scenarios. Traffic management strategies for wildfires must account for this unpredictability and rely on real-time updates to adjust traffic flow dynamically. Wildfire evacuations typically require rapid decisions and often lead to "shadow evacuations", where residents evacuate before official orders are issued. This can further complicate traffic predictions, as real-time traffic models are essential in managing congestion and ensuring efficient routes during these evacuations.

3. Case study

Most prior work focuses on large and relatively densely populated areas in the Western United States [9,13,18,19,25,34]. However, an open question remains whether these findings are transferable to other, less densely populated regions. To contribute to closing this gap, we analyzed a GPS dataset covering the 2021 wildfire that affected the village of Lytton in British Columbia, Canada. This section provides details about the present case study by introducing the study site in Section 3.1 and describing the GPS data and the data cleaning in Section 3.2.

3.1. Study site exploration

Lytton is a village in British Columbia, approximately 260 km northeast of Vancouver. It is located on the east side of the Fraser River Canyon, on an elevated area overlooking the confluence of the Thompson and Fraser Rivers. The study site, the fire perimeter, and the fire hotspot heatmap are shown in Fig. 1. Lytton is part of a transportation hub in Western Canada with access to major railway lines and highways. At the time of the wildfire (June 2021), the village's population was around 210 residents, with a mix of residential homes, businesses, and public buildings. The residential areas mainly consist of wooden houses with garages, sheds, and workshops, maintaining a distance of 2–5 meters between structures within and between lots [6].

In the days leading up to the devastating fire in Lytton on June 30, 2021, the temperatures soared consistently above 40 °C, hitting a record 49.5 °C on June 29, much higher than average temperatures during this time of year. The extreme heat extended into the nearby inland region which also experienced similarly high temperatures. The challenging terrain, especially the steep slopes of the Fraser River canyon, combined with intense sunlight, created strong slope-canyon winds. These winds, strongest in the late afternoon, became even more powerful due to low humidity, which dropped below 15 percent. On June 30, Lytton faced afternoon winds reaching around 35 km/hour, with gusts exceeding 50 km/h. The Lytton Creek Fire ignited in the afternoon of the same day, persisting until early August and burning an area exceeding 83,000 hectares. The fire's initial hours brought significant destruction to the Lytton community, where 151 homes and businesses were completely destroyed. The fire spread rapidly and reached the village's boundaries in less than an hour after ignition [6].

3.2. Data description

The raw dataset was obtained from a location data service[1] and contained GPS data covering the area −122.2598 to −118.9610 longitude and from 51.0607 to 49.3998 latitude, collected over a span of two months, between June 1st and July 31st, 2021. For

[1] https://www.quadrant.io/mobile-location-data

Table 1
GPS data attributes.

Data attribute	Definitions
Device ID	Unique device advertising identifier
Latitude	Provides the north–south positioning of the vehicle
Longitude	Provides the east–west positioning of the vehicle
Timestamp	Records the time and date of each data point
Horizontal Accuracy	Margin of error for the GPS point in meters

the specific area of Lytton, the analysis focused on coordinates from 50.33139 to 50.12481 latitude and −121.67108 to −121.34816 longitude, ensuring that GPS points both within and around Lytton were covered. This range was selected to include points along roads leading to and from Lytton that may have slight positional inaccuracies.

The raw dataset encompassed various attributes including device ID, latitude, longitude, timestamp, horizontal accuracy, device operating system, OS version, and country code. The final columns that we need for this research and we kept them after cleaning the data can be seen in Table 1. This raw dataset consisted of approximately 49 million records from 234,000 devices, provided in a CSV file format. Note that the data presented refer to the location and movement of devices, not people. However, it is assumed that mobile devices can be used as a proxy for people's locations.

4. Methodology

Fig. 2 gives an overview of the data processing steps to estimate departure times and predict the movement of evacuees using GPS data. In Section 4.1 we explain data cleaning and then discuss the evacuation rate during the wildfire in Section 4.2. Finally, the traffic prediction model will be explored in Section 4.3.

4.1. Data cleaning

The first step in data cleaning involved identifying and removing any missing or incorrect GPS records. To ensure the integrity of the dataset, we defined "unreliable or unusable data points" based on specific criteria. This included:

- **Duplicated Records:** GPS points where the same latitude and longitude were recorded multiple times for a single device within a short time frame were considered duplicates. These duplicate entries were removed to prevent redundancy and ensure that each device's movement was accurately represented.
- **Single or Two-Point Entries:** Device IDs associated with only one or two GPS records were excluded from the dataset. These records were considered unreliable as they did not provide enough data to track movement or determine evacuation behavior. Approximately 24,000 devices had a single GPS record, and 15,000 devices had two records, and these entries were removed to maintain the accuracy of the dataset.

To further improve data quality, a threshold of 1000 m was established for horizontal accuracy. This threshold was selected because GPS devices can experience positional errors due to terrain and other environmental factors, especially in areas like Lytton. If the horizontal accuracy of a GPS record exceeded 1000 m, the location was considered too imprecise for tracking movement during an evacuation. This threshold ensured that we were working with data points accurate enough for analysis while retaining enough records to conduct a meaningful study. The 1000-meter cutoff was chosen based on typical GPS accuracy observed in similar studies [35,36], ensuring a balance between data reliability and dataset size.

Following these cleaning steps, the dataset was refined to include 41,578,631 records, providing a robust and reliable foundation for the study's objectives (see top of Fig. 2).

4.2. The departure analysis

Effective disaster management necessitates a comprehension of individuals' evacuation patterns during emergencies. Human trajectories exhibit significant regularity in both time and space, with individuals frequently returning to a few key locations [37]. Various algorithms have been developed for this purpose, including clustering-based [38,39], differential-based [40–42], and probabilistic-based [8,43] strategies. In this study, we opted for a differential-based approach due to its favorable characteristics, such as low computational complexity and suitability for real-time detection on smartphones [44]. By leveraging differential-based algorithms, stay points can be automatically identified from an individual's GPS trajectory by identifying spatial regions where the individual spends a duration exceeding a specified threshold. We established a predefined distance threshold to determine whether an individual had likely evacuated the area. Specifically, individuals were considered to have evacuated if their location was detected beyond a certain distance from the village. This threshold was set to ensure that the movement detected represented a significant displacement, rather than daily routine travel, and indicated an actual evacuation event.

The distance threshold was chosen based on a combination of factors, including the geographical layout of Lytton and its surrounding areas, as well as typical travel patterns observed in the dataset. This approach ensured that only meaningful evacuations were captured in the analysis. We applied this distance consistently to analyze the behavior of evacuees during the wildfire.

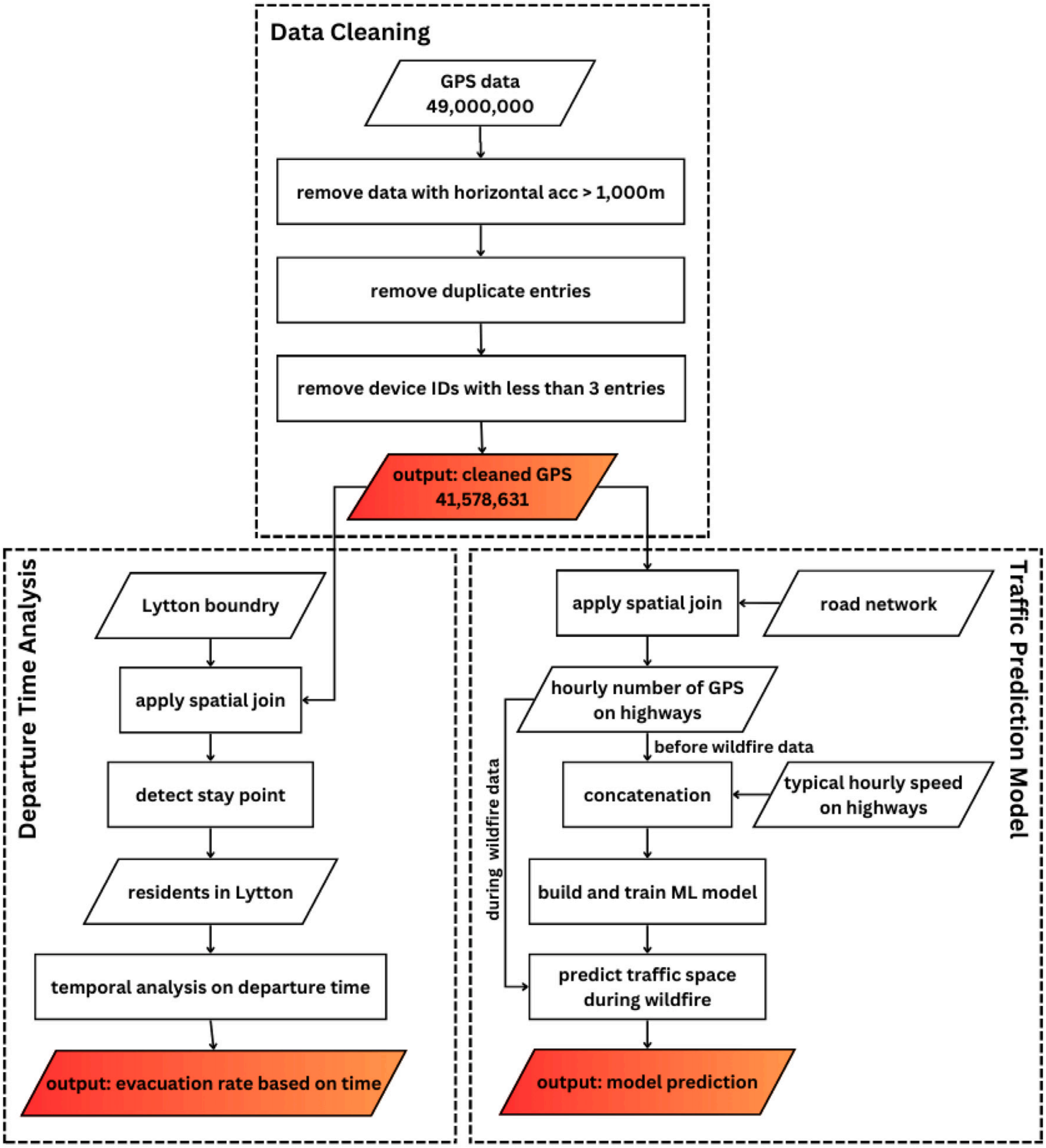

Fig. 2. Overview of the data processing procedure.

The differential-based strategy for detecting stay point locations involves calculating the spatial difference between consecutive GPS points using the Haversine formula [45,46], where d_i represents the distance between consecutive points. The Haversine formula for calculating distance between two points $(\text{lat}_1, \text{lon}_1)$ and $(\text{lat}_2, \text{lon}_2)$ on the Earth's surface is given by:

$$d_i = 2 \cdot R \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\text{lat}}{2}\right) + \cos(\text{lat}_1) \cdot \cos(\text{lat}_2) \cdot \sin^2\left(\frac{\Delta\text{lon}}{2}\right)}\right) \tag{1}$$

Here, R is the Earth's radius (mean radius = 6,371 kilometers), $\Delta\text{lat} = \text{lat}_2 - \text{lat}_1$, and $\Delta\text{lon} = \text{lon}_2 - \text{lon}_1$.

Incorporating time intervals between consecutive GPS points, we define the adjusted distance d_{adjusted} to include time as a factor in the analysis. Let Δt_i represent the time interval between consecutive GPS points. The adjusted distance d_{adjusted} is calculated as follows:

$$d_{\text{adjusted}} = d_i + s \times \Delta t_i \tag{2}$$

Where s is the average speed of movement between consecutive GPS points.

Thresholds for distance (τ_d) and time (τ_t) are then set to define stay points, with a potential stay point detected if $d_{\text{adjusted}} > \tau_d$ and the time interval (Δt_i) exceeds τ_t. Individuals are considered to have evacuated if they are estimated to be staying more than a certain distance (r) away from their origin location.

The evacuation rate offers a robust framework for analyzing evacuation behavior and provides valuable insights for disaster management.

The calculation of evacuation rates involved determining the proportion of individuals who were identified as evacuated compared to the total number of users estimated to be residing in the community under consideration. This metric provides insights into the temporal dynamics of evacuation, highlighting periods of heightened evacuation activity relative to the population size. We can further calculate the rate at which residents complied with an evacuation order in a given location. The evacuation compliance rate α_t on a given time period t can be calculated by:

$$\alpha_t = \frac{M_t}{N_t} \tag{3}$$

where M_t is the number of evacuees who left during time period t in a given location, and N_t is the total number of residents living in a given location during time period t.

4.3. Traffic prediction model

Understanding and predicting traffic speed during wildfire evacuations is crucial for efficient emergency response planning and ensuring the safety of affected populations. By leveraging GPS data and external sources of traffic information, such as average speed data from platforms like HERE Maps,[2] we developed a methodology to predict traffic speeds during wildfire evacuations. Initially, hourly average typical speed data is collected from HERE Maps for a period of one week, providing typical speed values for each hour of the day. This dataset serves as a reference for normal traffic conditions on highways. Subsequently, the average speed data is synchronized with our flow data obtained during a normal week before the wildfire event. By aligning the flow and speed data for the same time intervals, we can analyze the relationship between normal flow and average speed on highways. The method comprises GPS data map matching, flow calculation, synchronization of flow and speed data, training a regression model, and speed prediction.

Due to inherent limitations of raw GPS data such as data inaccuracy, spareness, and missing alignment with road segments, the first step in analyzing GPS records on the road segments is map matching [47]. The map matching process is essential for accurately associating raw GPS records with specific road segments, enabling the estimation of traffic flow and speed on highways [48]. Here, we use a stepwise corresponding map-matching algorithm to match the GPS records on the corresponding road segments section for further analysis:

Step 1 - Discretization and definition of study regions. The study area is partitioned into a grid with square cells measuring 50 meters on each side [49]; each cell is assigned a numerical identifier. Then the number of road sections in each grid cell is counted.

Step 2 - Record localization. We start with the first GPS record. Firstly, we determine the grid cell where the GPS record is located. Then we check if there are roadway segments within the designated grid cell or its adjacent eight cells, Step 3 is executed; otherwise, it is assumed that the record is not located on any road, and the assessment proceeds to the next GPS record.

Step 3 - On-road verification. In this step, we determine whether the GPS record aligns with the chosen road segments. If there are multiple road sections in the same grid cell as the GPS record or its neighboring cells, we use a buffer around the GPS record to determine the closest road segments to the GPS record. If the closest road segments is among our desired road segment, we proceed to Step 4, otherwise, we go back to Step 2.

Step 4 - Direction of travel inference. We check the direction between the current GPS record with the corresponding GPS record to determine the direction of the movement to use it as the probable direction of travel on the road (e.g., south- or northbound). Subsequently, the algorithm continues with the next GPS record (Step 2) until all entries have been analyzed.

Following map matching, the hourly flow of vehicles on highways is calculated. This process involves aggregating individual trajectories to estimate the number of vehicles passing through each highway segment per hour. The hourly flow (F_t) on a specific highway segment (H) at time t is computed as the sum of binary indicators denoting whether a GPS point falls within the bounds of the highway segment:

$$F_t(H) = \sum_{i=1}^{N_t} \delta(x_i, H) \tag{4}$$

[2] https://www.here.com/

Where N_t is the total number of vehicles observed at time t, and $\delta(x_i, H)$ is a binary function indicating whether GPS point x_i falls within the bounds of highway segment H.

Next, to establish a relationship between flow and speed, hourly *average* traffic speed data obtained from the external source (HERE Map dataset, which provides typical traffic data of non-emergency days and contains an average speed for each segments of the road per hour), is synchronized and concatenated with the hourly flow data. This synchronization ensures that each hourly observation of flow corresponds to the corresponding average speed measurement. The synchronized dataset is represented as:

$$D = \{(F_t, S_t)\}_{t=1}^{T} \tag{5}$$

Where D is the synchronized dataset, F_t is the hourly flow on highways at time t, and S_t is the corresponding average speed at time t. This synchronized dataset serves as the basis for training a linear regression model to predict traffic speed based on observed flow. The training dataset consists of hourly data collected during a normal week, comprising 168 data points representing each hour of the week. To train the model, we utilize the scikit-learn[3] library in Python, which provides efficient tools for machine learning tasks. We used 10-fold cross-validation to evaluate the performance of our machine learning model. In 10-fold cross-validation, the dataset is randomly divided into 10 equally sized subsets, or "folds". The model is trained and validated 10 times, each time using a different fold as the validation set and the remaining 9 folds as the training set. This process ensures that each data point is used for validation once and for training nine times. The performance metrics are then averaged over the iterations to provide a more robust estimate of the model's performance. This approach allows us to assess the performance of the model on unseen data and mitigate overfitting.

Once the synchronized dataset is prepared, a linear regression machine learning model [45,50] is trained to establish the relationship between flow and speed on highways. The relationship between traffic flow and speed was modeled using linear regression. This approach was chosen because, given the size and characteristics of our dataset, linear regression provided the best results. The dataset, while sufficient for analysis, was not extensive enough to necessitate more complex models, and linear regression offered a balance between simplicity and performance. The model was able to capture the key relationships between traffic flow and speed effectively and provided reliable results for predicting traffic dynamics during the evacuation. The model learns the underlying patterns and dependencies between the two variables, enabling the prediction of speed based on observed flow. The linear regression model can be represented as:

$$S_t = \beta_0 + \beta_1 \cdot F_t + \epsilon_t \tag{6}$$

Here, S_t represents the predicted speed at time t, F_t denotes the observed flow at time t, β_0 and β_1 are the intercept and slope coefficients, respectively, and ϵ_t represents the error term.

In this study, we minimized the Root Mean Squared Error (RMSE) during model parameter estimation. RMSE is calculated as the square root of the average of the squared differences between predicted and actual values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

Where n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value. Minimizing RMSE helps in reducing the overall error in predictions and ensures that the model accurately captures the relationship between flow and speed.

MAE represents the average absolute deviation between predicted and actual values, offering insights into the precision of the model predictions. It is calculated as the average of the absolute differences between predicted and actual values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

Where n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

R-squared measures the proportion of the variance in the target variable that is explained by the independent variables. It indicates the goodness of fit of the model to the data, with higher values suggesting a better fit. R-squared is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{9}$$

Where: n is the number of observations, y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the actual values.

Once the linear regression model is trained, it can be utilized to predict the speed on highways given the hourly flow data for a specific date and time during wildfire evacuation. The model does not assume that traffic during evacuations will flow the same as on normal traffic days. Instead, it takes into account the observed flow during the evacuation and generates an estimate of the corresponding speed, accounting for potential traffic delays or congestion that may occur due to the emergency situation. The goal is to provide accurate predictions of traffic speeds under evacuation conditions based on real-time or past data.

By following this methodology, traffic speed during wildfire evacuations can be predicted with known accuracy and precision using GPS data.

[3] https://scikit-learn.org/

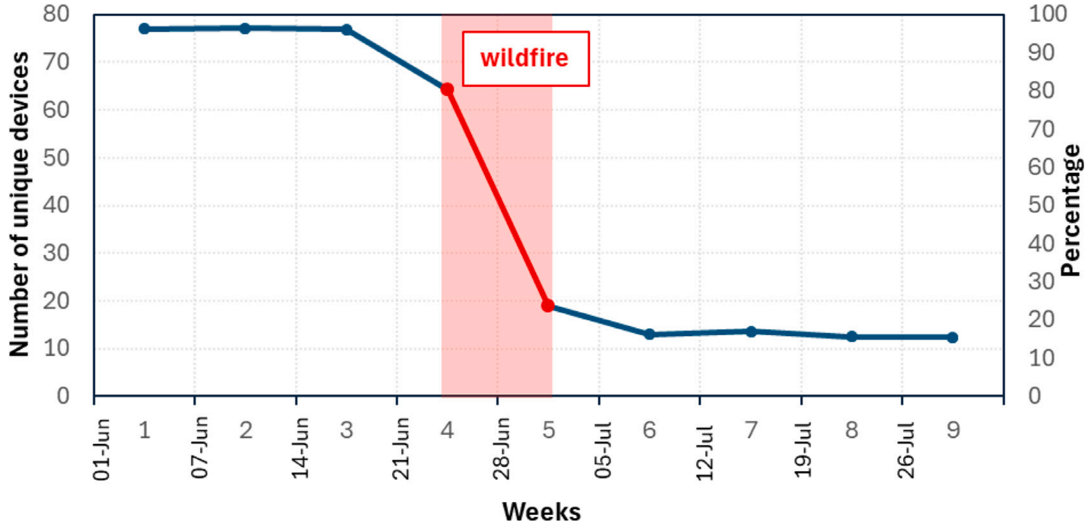

Fig. 3. The average rate of number of unique device IDs per week during June and July.

5. Results

5.1. Departure time results

This section presents the average GPS records over the 2 months study period impacted by the Lytton wildfire 2021, from June 1st to July 31th, as illustrated in Fig. 3.

In this subsection, we focus on analyzing departure time results using the differential-based strategy for detecting stay point locations as outlined in Section 4.2. Firstly, we identify stay points within Lytton, revealing a total of 80 unique device IDs.We assumed that each unique device ID belongs to one resident. Comparing this finding with the official population statistics[4] states a population of 210 for Lytton before the fire, our dataset covers approximately 38% of the population. To compute the departure rate, we calculate the weekly average of unique individuals and divide it by the total number of unique individuals identified (80), as illustrated in Fig. 3.

In analyzing the percentage rates of average stay points in Lytton across different Weeks, notable fluctuations are observed. Fig. 3 represents the average number of unique devices per week, with each point in the line chart reflecting the middle of the corresponding week. As can be seen in Fig. 3, preceding the wildfire event, during Weeks 1, 2, and 3, the percentage rates remain relatively consistent, ranging from approximately 96% to 96.45%. This consistency suggests stable residency patterns among individuals within the community during this period. However, a significant deviation is observed during Week 4 (June 21–27) and Week 5 (June 28-July 4), highlighted as red, where the most significant changes occurred and they are placed in a red rectangle to emphasize that these two weeks saw the most dramatic shifts in evacuation behavior. In the weeks of the wildfire, week 4 and week 5, the percentage rate drops drastically to approximately 80% and 23.5%, respectively. This decline likely indicates a disruption in normal residency patterns, possibly due to evacuation measures or other wildfire-related factors, such as the record-breaking increase in temperature as the region approached June 30. The extreme heat, which set new records daily, may have prompted some residents to evacuate early.

Additionally, smoke and poor air quality caused by nearby fires from the north of Lytton could also have been factors prompting early self-evacuations. The presence of these fires contributed to worsening air quality, which may have driven residents to leave the area before the official evacuation order.

Subsequently, in post-wildfire weeks, the percentage rates continue to decrease, indicating a prolonged impact on residency patterns following the wildfire event. Notably, during the weeks immediately following the wildfire (Weeks 6 to 7), the percentage rates reached as low as approximately 16%. This substantial decrease suggests a continued disruption in residency patterns, possibly due to ongoing evacuation efforts, damage assessment, or community rebuilding processes. Overall, the observed departure rate trends provide valuable insights into the dynamic nature of evacuation behavior and community response to wildfire emergencies.

With a focus on Weeks 4 (June 21–27) and 5 (June 28-July 4), the most significant reduction in departure rates is observed. Delving deeper into these Weeks, it is essential to analyze the daily variations to better understand the temporal dynamics of evacuation behavior. By examining the departure rates on a daily basis during Weeks 4 and 5, we can gain a more granular understanding of how evacuation patterns evolved in response to the escalating wildfire threat.

[4] https://www.tnrd.ca/

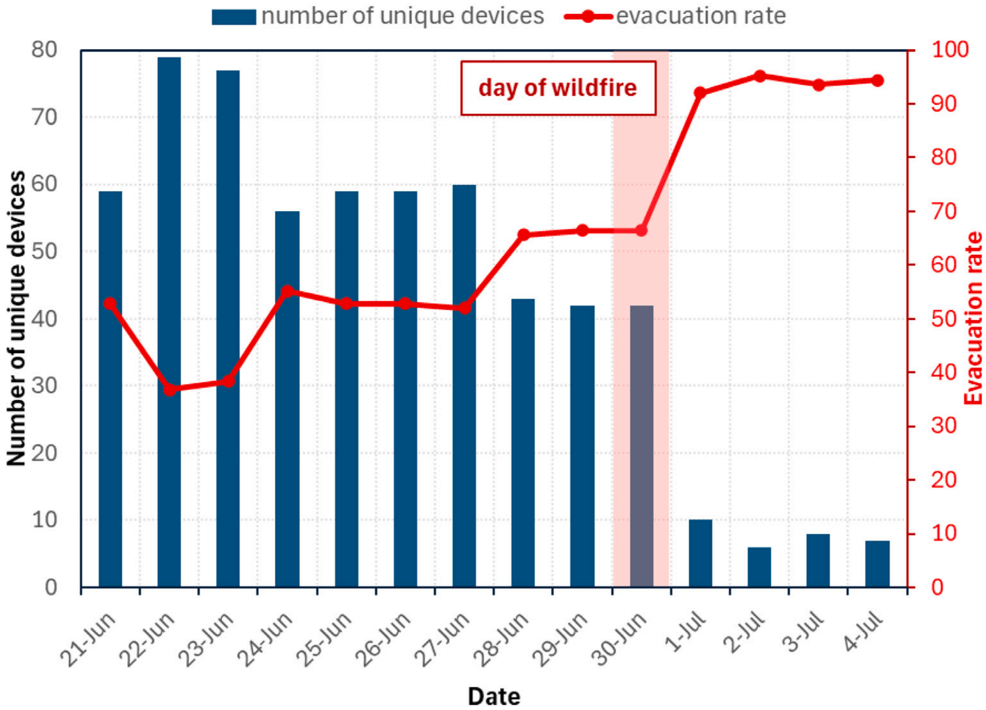

Fig. 4. The number of unique device IDs and the rate of evacuees in Lytton over Week 4 and Week 5.

Analyzing the evacuation behavior during the critical period spanning from June 27 to July 4 in Lytton provides valuable insights into the temporal dynamics of community response to the wildfire threat. As can be seen in Fig. 4, commencing on June 27, we observe a relatively stable number of unique devices, suggesting that evacuation preparations were not yet underway. However, a slight decrease in the number of devices is noticeable on June 28 and 29, indicating a potential early stage of evacuation readiness among some individuals. The evacuation rate increased on June 29, indicating a substantial proportion of the population considering or initiating evacuation measures.

On the day of the wildfire, June 30, a notable surge in evacuation activity is evident. Despite the relatively stable number of unique devices leading up to this date, the evacuation rate sharply rises to 66.4%, suggesting a sudden and significant response to the immediate threat posed by the advancing wildfire. Interestingly, the evacuation rate on July 1 jumps to 92%, indicating a substantial evacuation effort occurring concurrently with or immediately following.

Following the wildfire event, from July 2 to July 4, the number of unique devices decreases markedly, indicating a continued evacuation trend. The evacuation rates remain high, hovering above 90%, suggesting a swift and comprehensive evacuation response in the aftermath of the wildfire. This pattern indicates a proactive approach to evacuation among the majority of residents, with many opting to evacuate early in response to the immediate threat posed by the wildfire. Overall, the analysis reveals a remarkable cohesion and efficiency in evacuation behavior within the community, with a significant portion of the population taking prompt action to ensure their safety in the face of the wildfire emergency.

5.2. Traffic prediction model

In this section, we present the development and evaluation of a traffic speed prediction model aimed at forecasting traffic speed on three main exit roads from the village during wildfire evacuations. Fig. 5 shows the three main roads connecting Lytton to the surrounding highways, the TransCanada Highway North (red), the TransCanada Highway South (green), and Highway 12 (yellow). These roads served as the primary exit routes during the wildfire evacuation. The TransCanada Highway South was the most heavily used, as it leads to cities in the south and provided safer passage, with fewer immediate threats from the wildfire. The TransCanada Highway North usage was limited due to the proximity of the Sparks Lake Wildfire to the north. Similarly, Highway 12 saw lower usage due to the McKay Creek Fire in that direction.

Our model aims to capture the relationships between the number of evacuees using each road (traffic flow) and traffic speed, providing actionable insights for emergency management agencies. By predicting how traffic flow affects road capacity, the model helps identify when road closures or bottlenecks might occur based on the number of evacuees using the same road simultaneously. This information can assist both emergency managers in planning responses and evacuees in choosing safer, less congested routes.

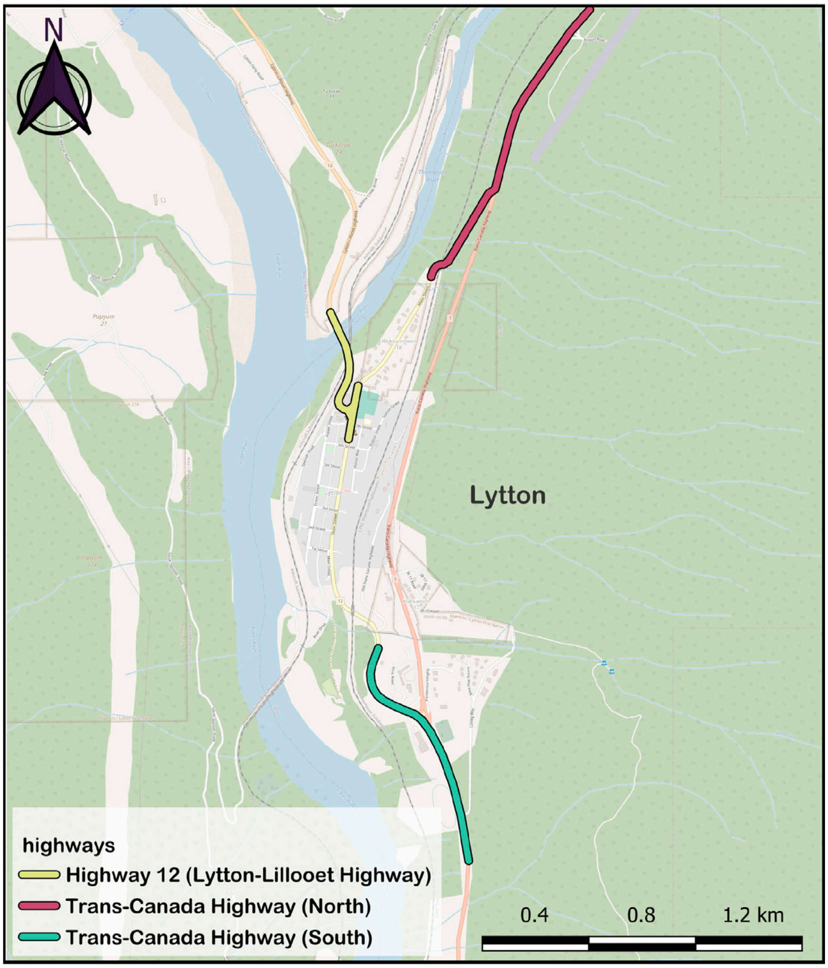

Fig. 5. Three main exit roads from Lytton. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2.1. Map matching

The map matching process is essential for accurately associating raw GPS data with specific road segments, enabling the estimation of traffic flow and speed on highways. As depicted in Fig. 6, the raw GPS data points appear scattered and unaligned with the road network due to inaccuracies and noise inherent in GPS measurements. To address this, we employ the proposed map matching algorithm, as discussed in Section 4.3, to align GPS records with the corresponding road segments.

The green line in Fig. 6 presents a sample of map-matched GPS trajectories overlaid on the road network. This visualization highlights the accuracy of the map matching process, as GPS records are now aligned with specific road segments. At the same time, the direction of the GPS records can be determined by performing map matching. By matching GPS records to road segments, we can accurately identify the routes chosen by individuals and calculate the number of vehicles passing each road during the wildfire.

5.2.2. Calculation of hourly flow on highways subsection

By aggregating individual map matched GPS trajectories, we can estimate the number of vehicles exiting from Lytton through specific highway segments per hour. This process involves summing the binary indicators denoting whether GPS points fall within the bounds of selected highway segments as mentioned in Section 4.3. Through this calculation, we obtain insights into variations in traffic volume over time, enabling us to identify peak periods of congestion and assess the overall flow of vehicles on critical evacuation routes.

Fig. 7 depicts a line graph illustrating the hourly flow of vehicles on the three selected highway segments over week 4 and week 5 of the study period (see Section 5.1). The graph showcases variations in traffic volume throughout the analyzed timeframe, highlighting fluctuations in flow patterns before and after the onset of the wildfire. By comparing flow data before and after the

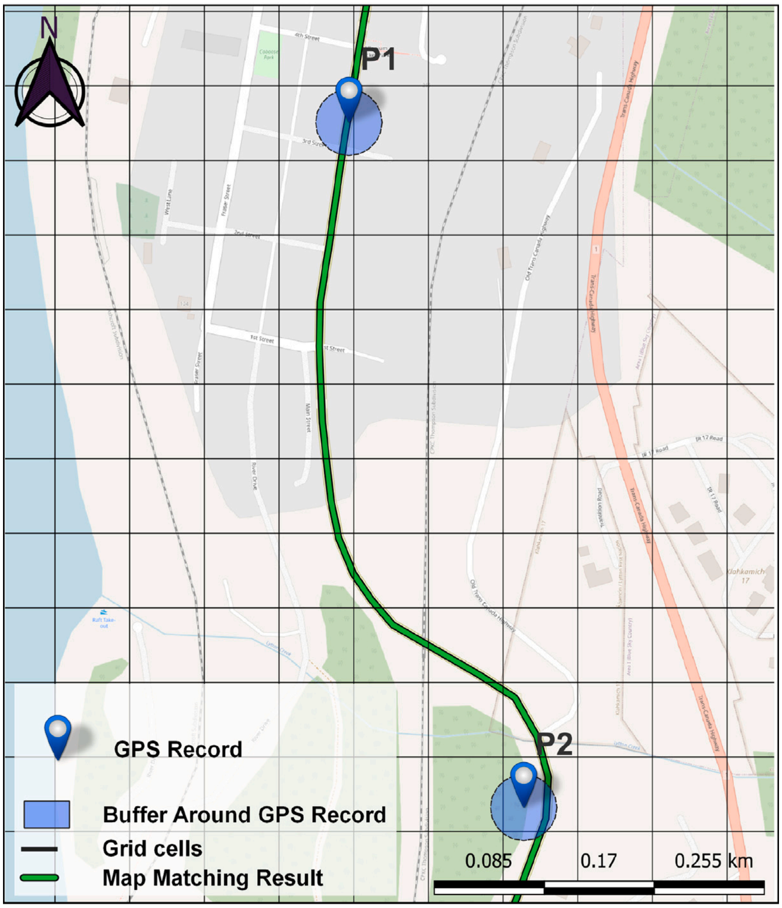

Fig. 6. Example GPS records from a single device before map matching (blue pins) and estimated trajectory on the road after map matching (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

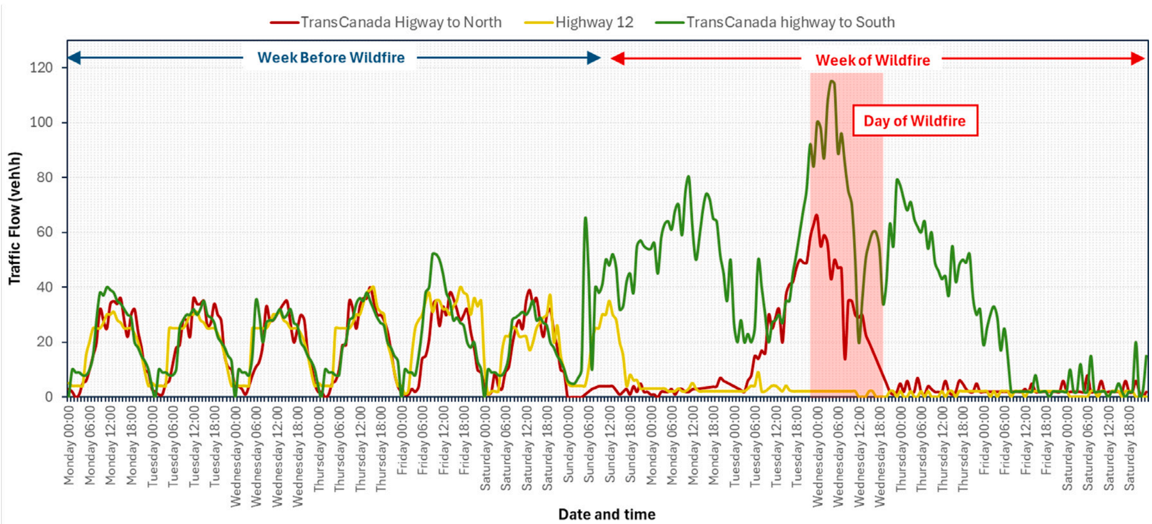

Fig. 7. Number of passing individuals (Flow) on each highway on an hourly basis. Green: Trans-Canada Highway (South), Red: Trans-Canada (North), Yellow: Highway 12. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
The RMSE and MAE of the model.

Highway	RMSE	MAE	R^2
Highway 12	2.459	1.865	0.524
Trans-Canada North	2.004	1.656	0.692
Trans-Canada South	2.383	1.905	0.793

wildfire, we can identify any significant changes in traffic volume and assess the impact of the wildfire event on evacuation routes. As can be seen, before Monday of week 5 which is two days before the wildfire reached the village, the number of GPS records followed a consistent pattern. However, as the wildfire approached, specifically nearing the wildfire day on Wednesday of week 5, there was a decrease in the number of records along Highway 12, while there was an increase in records along the Trans-Canada Highway both in the north and south. As can be seen, the rise in the number of people passing Trans-Canada from the south part of the village is by far more than the other two highways. This suggests that residents preferred the southern route for evacuation.

5.2.3. Synchronization of flow and speed data

Fig. 8 presents a series of scatter plots illustrating the relationship between hourly flow and average speed on three different highways: Highway 12, Trans-Canada (North), and Trans-Canada (South). Each data point represents a specific time interval, with flow values plotted on the x-axis and average speed values on the y-axis. The surrounding histograms show the distribution of speed and traffic flow for each highway.

Analysis of the data points reveals a strong and consistent negative correlation across all three highways, where higher flow corresponds to lower speeds. The linear regression lines clearly illustrate this trend, indicating that as traffic flow increases, traffic speed decreases. This relationship was modeled using linear regression rather than a quadratic curve, as the observed flow values did not reach levels where nonlinear congestion effects would dominate. The evacuation routes under analysis did not approach the capacity limits seen in high-volume urban highways, making the linear regression model appropriate for understanding traffic dynamics during this event.

Fig. 9 depicts a time series plot showing the synchronized flow and speed data over the analyzed period. The plot enables visualization of temporal trends and correlations between flow and speed on highways throughout the selected timeframe. As can be seen from Fig. 9, it is evident that there exists a negative correlation between flow and speed on highways. Specifically, during morning and evening rush hours on weekdays, characterized by higher traffic flow, the average speed tends to decrease significantly. This phenomenon is indicative of congestion and traffic congestion during peak commuting hours. Conversely, during weekends, a different pattern emerges, with lower traffic volumes and relatively higher average speeds compared to weekdays.

5.2.4. Training a linear regression model

In this section, we describe the process of training a linear regression model using the synchronized flow and speed data. During training, we fit a linear regression model to the training data, aiming to learn the relationship between flow and speed on highways. Table 2 presents the average RMSE, MAE, and R^2 values for all runs of the cross-validation for each of the three selected highways, computed based on the predictions generated by the trained linear regression model. The RMSE, MAE, and R^2 metrics provide measures of the model's predictive accuracy, with lower RMSE and MAE values and higher R^2 values indicating better performance. From the table, it is observed that the model achieves relatively low RMSE and MAE values and high R^2 values across all highways, indicating a good fit to the training data.

The validation and testing results based on Table 2 indicate that the trained linear regression model performs reasonably well in predicting traffic speeds on highways during normal conditions. The low RMSE and MAE values suggest that the model effectively captures the underlying relationship between flow and speed, demonstrating its utility for traffic prediction purposes. Furthermore, the consistency of performance across different highways highlights the robustness of the model across varying traffic scenarios.

However, it is important to address the validation methodology used. In this study, we employed nested cross-validation to ensure the robustness and reliability of our model. Nested cross-validation involves an outer loop to split the data into training and test sets, and an inner loop for cross-validation within the training set to tune the model parameters. This approach helps mitigate overfitting and provides a more unbiased evaluation of the model's performance on unseen data. The final performance metrics were calculated on the test set that the model had not seen during the training process.

This rigorous validation process confirms that the model's performance metrics, such as RMSE, MAE, and R^2 values, provide a realistic measure of its predictive accuracy. While the R^2 values indicate a moderate fit, the model's predictive capability is still useful for practical applications in traffic speed prediction during wildfire evacuations. Future work could explore more complex models and additional data sources to improve predictive accuracy further.

5.2.5. Speed prediction during wildfire

In this section, we discuss the application of the trained linear regression model to predict traffic speeds on highways during wildfire evacuation events. The model was trained on the relationship between typical traffic speeds (provided by HERE for Lytton) and the number of vehicles (traffic flow) on the road. The relationship between traffic flow and speed is modeled as a regression problem, as increased traffic flow generally results in reduced speeds, and vice versa.

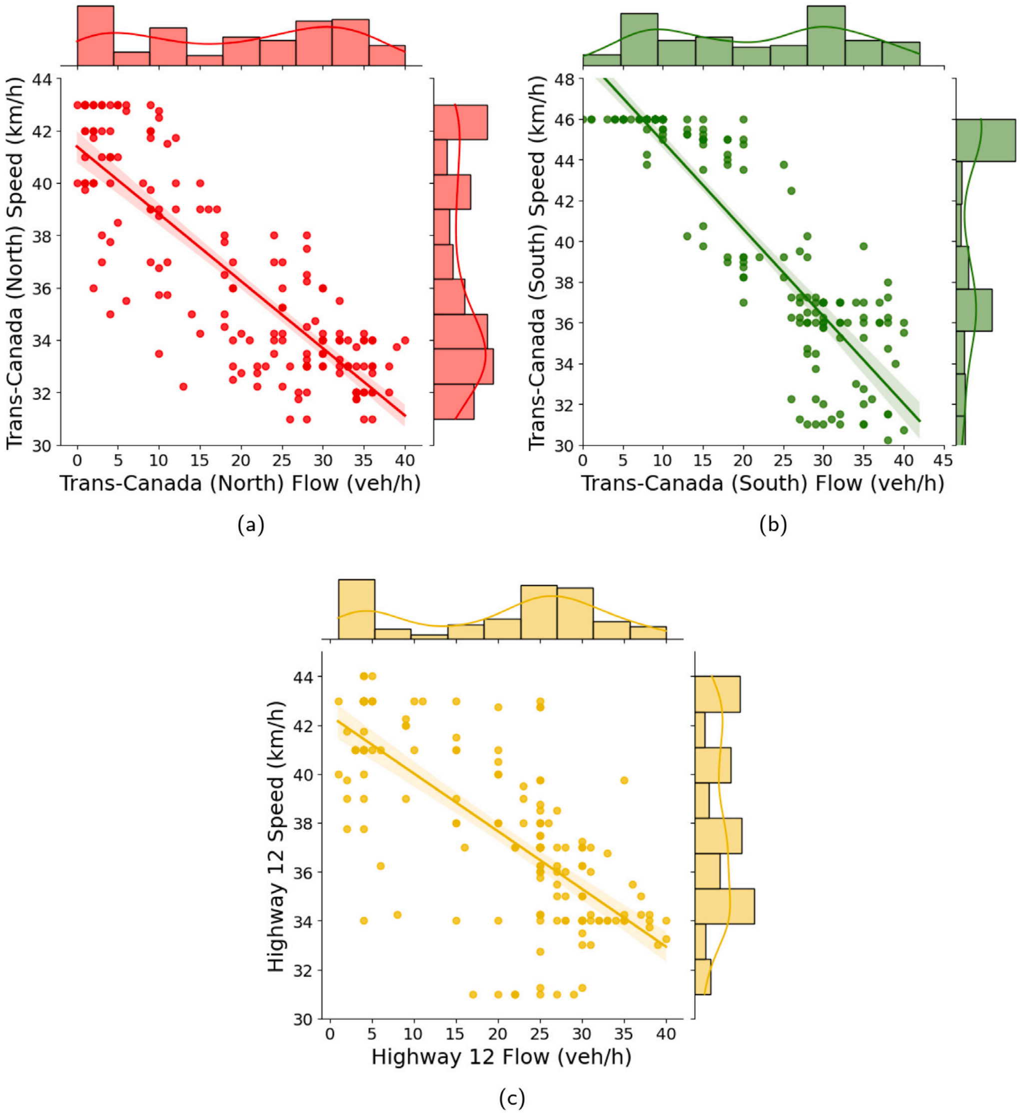

Fig. 8. Scatter plot illustrating the relationship between hourly flow and average speed on (a) Trans-Canada Highway (North), (b) Trans-Canada Highway (South), (c) Highway 12.

Due to the size of our dataset, a linear regression model was selected, as it provided the best performance for learning the relationship between traffic flow and speed. With the model trained on normal traffic conditions, we then used flow data collected during the wildfire week on the three selected highways to predict traffic speeds during the evacuation.

By inputting this flow data into the trained model, we generated predictions of traffic speeds during the wildfire, enabling us to anticipate potential bottlenecks and congestion points along evacuation routes. These predictions help identify when and where congestion is likely to occur, providing actionable insights for emergency management agencies.

Fig. 10 illustrates the predicted traffic speeds on these main roads during the wildfire evacuation. The graph shows that during the early hours of Wednesday, June 30, a significant number of evacuees chose the Trans-Canada Highway South over the other two exit routes. The traffic flow on the main road leading to the Trans-Canada Highway South was more than 100 vehicles per hour during this period, while the road leading to the Trans-Canada Highway North had a flow of around 60 vehicles per hour, and Highway 12 saw almost no traffic. As a result, the predicted traffic speed on the main road leading to the Trans-Canada Highway South was substantially lower compared to the other roads during these hours.

Additionally, around 4 AM on Wednesday, June 30, the model predicts a traffic jam on the main road leading to the Trans-Canada Highway South, with traffic speeds approaching zero due to the high volume of vehicles. This prediction highlights the importance

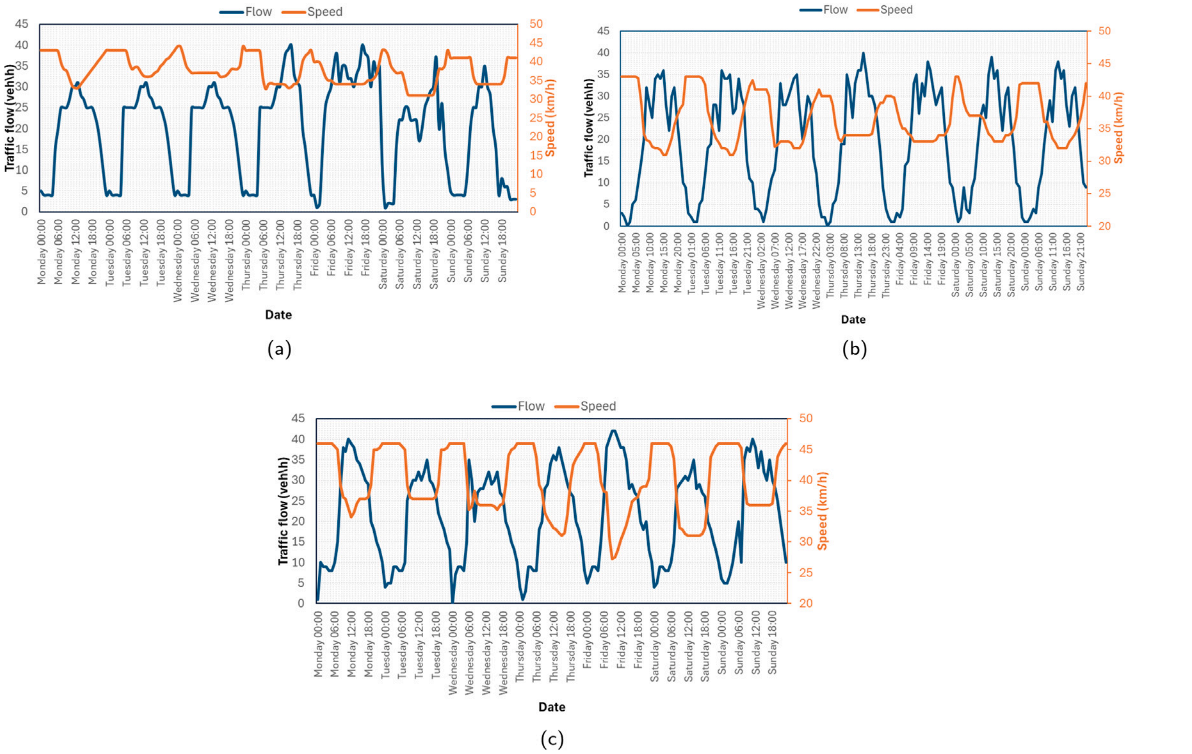

Fig. 9. Time series plot showing the synchronized flow and speed data over the analyzed period for highways (a) Highway 12, (b) Trans-Canada (North), and (c) Trans-Canada (South).

of preemptive traffic management, as congestion on these critical exit routes can significantly hinder the safe and timely movement of evacuees.

6. Discussion

The objective of this study is to examine the spatial and temporal behavior of residents in small communities such as Lytton during a wildfire event, utilizing GPS data obtained from mobile devices. Unlike many studies focusing on broader geographical regions with extensive datasets, this research concentrates specifically on the Lytton wildfire, aiming to address two primary inquiries: the timing of residents' evacuation and the preferred exit routes from the village. To address the first question, the study tracks the daily population count in Lytton, providing insights into the evacuation trends. The findings reveal the stay rates in Lytton stayed steady around 96% before the wildfire, dropping significantly during the wildfire weeks to as low as 23.5%. Even after the wildfire, rates continued to decline, hitting around 16%. These trends reflect the wildfire's impact on residency patterns and evacuation behavior. On June 30, the wildfire sparked a sharp increase in evacuations, reaching 66.4%. By July 1, the evacuation rate surged to 92%, showing a swift response to the escalating danger, with many residents evacuating during or right after the wildfire began. This indicates that the majority of residents remained in Lytton until the day of the wildfire, Wednesday, June 30, necessitating evacuation. This observation holds significant implications for evacuation planners, shedding light on the management of residents in compact areas such as Lytton.

Subsequently, the study investigates the evacuation routes and evaluates the proportion of residents utilizing each road and highway, along with assessing traffic flow dynamics. To achieve this objective, the research employs map matching methodology to ascertain the traffic volume on the main exit roads leading from Lytton to the highways, utilizing GPS data and hourly flow calculations on these road segments. Furthermore, typical speed data sourced from HERE maps are synchronized with the flow data to train a linear model predicting speed variations on each exit route. The analysis reveals that during the early hours of Wednesday, June 30, a substantial number of individuals opted for the main road leading to the TransCanada South highway over alternative routes. Notably, the flow on this road exceeded 100 vehicles per hour, contrasting with approximately 60 vehicles per hour on the road leading to TransCanada North and minimal flow on the road leading to Highway 12.

One possible reason for the lower usage of Highway 12 during the evacuation is the proximity of other active fires in the region. The McKay Creek Fire, which started near Highway 12 on June 29, 2021, likely influenced residents' decisions to avoid that route due to safety concerns. Similarly, the Sparks Lake Wildfire, which began on June 28, 2021, to the north of Lytton, may have also affected route choices, leading evacuees to favor the perceived safer routes, particularly the TransCanada South highway. Consequently, the

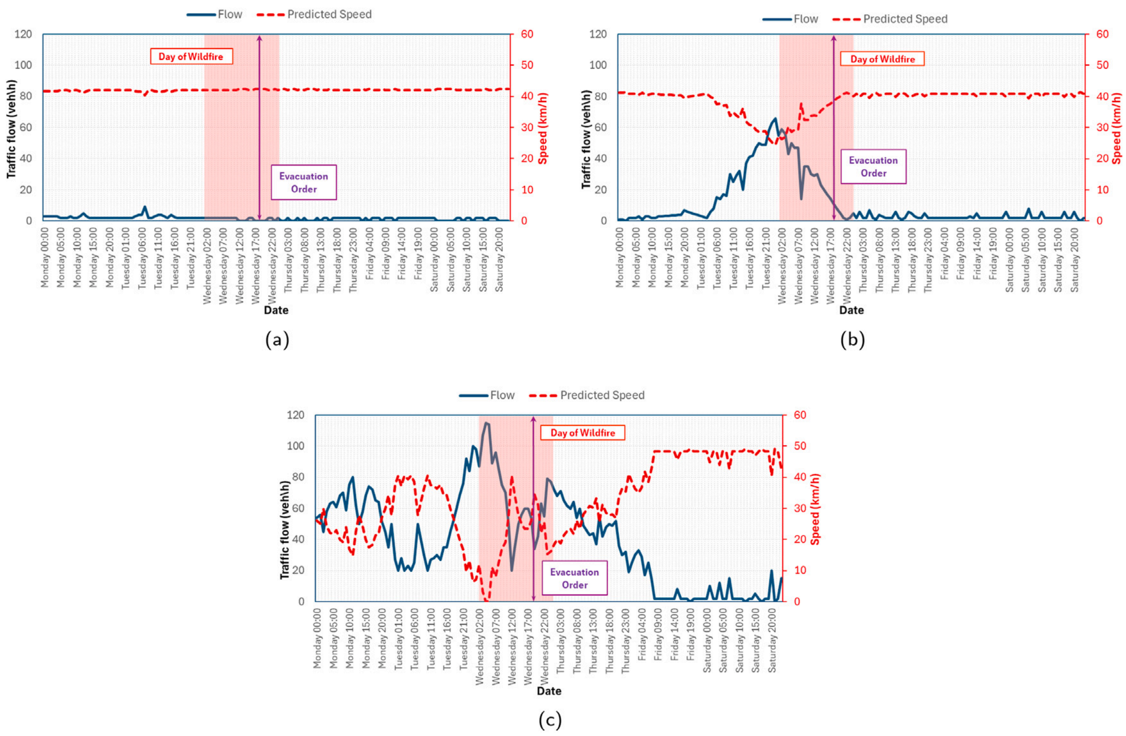

Fig. 10. Predicted speeds during wildfire evacuation events for highways (a) Highway 12, (b) Trans-Canada (North), and (c) Trans-Canada (South).

predicted traffic speed on the main road leading to the TransCanada South highway is anticipated to be considerably lower during these hours, with indications of a traffic congestion event around 4 AM.

Our model is capable of predicting traffic speeds for different traffic volumes on these exit roads, revealing the likelihood of congestion and bottlenecks, particularly on the main road to TransCanada South, where the flow exceeded 100 vehicles per hour. According to historical traffic data obtained from the BC Ministry of Transportation and Infrastructure's traffic database,[5] typical road capacity on these smaller exit roads is limited. While 100 vph may not cause significant congestion on larger highways, it can create bottlenecks on these local exit routes, contributing to the observed slowdown in traffic speed during the evacuation.

The predictive capacity of traffic speeds during wildfire evacuations holds significant implications for emergency management and transportation planning. By anticipating congested points along evacuation routes, authorities can implement preemptive measures to mitigate traffic issues, ensure evacuee safety, and expedite the evacuation process. These insights also help authorities issue more targeted evacuation orders, manage resources efficiently, and predict where congestion is likely to occur. The traffic model developed in this study enables adaptive traffic management strategies, such as opening additional lanes or providing real-time updates to evacuees. Additionally, the study highlights the need for improved public communication to encourage earlier evacuations and prevent last-minute traffic surges. Finally, these findings can inform community preparedness efforts and pre-evacuation drills, particularly by promoting the use of alternative routes, if possible, to better distribute traffic during emergencies.

6.1. Limitations and future work

This study focused on analyzing the challenges faced by the small community of Lytton, Western Canada, during a wildfire evacuation. Lytton had a population of 210 residents prior to the fire, and the small population and local geography posed significant constraints on obtaining accurate and frequent data. As a result, certain analyses commonly conducted in more densely populated areas were not feasible. For instance, despite employing map matching methodologies, the low data accuracy prevented us from distinguishing between multiple southbound roads, limiting our ability to analyze detailed route preferences among evacuees.

Another challenge was the limited frequency of GPS records for most unique device IDs, which hindered our ability to determine the home locations of residents and accurately track the timing of evacuations across different regions. This contrasts with other studies, such as the analysis of the Kincade wildfire in California, which had higher data resolution [34].

Additionally, predicting vehicle speeds during the wildfire week was limited by gaps in the GPS data. Missing values at different hours were addressed through interpolation methods to prepare the dataset for training. However, the size and quality of our data

5 https://tradas.th.gov.bc.ca/tradas.asp

restricted the machine learning models we could employ. Although linear regression is commonly used in similar research [51,52], our decision to use it was also influenced by the limitations of our dataset, which constrained the predictive power of more complex models.

Moreover, due to the lack of census data or demographic information about Lytton during the wildfire, we made assumptions about the evacuee population. Specifically, we assumed that all GPS data points corresponded to permanent residents of Lytton, as we did not have access to specific tourism or census data to account for the presence of visitors or tourists [53]. We assumed that all evacuees had access to vehicles and were residents of Lytton. This assumption may overlook vulnerable populations, such as older adults, underserved communities, or individuals without vehicles, who may have faced additional barriers during the evacuation. Future research with more comprehensive demographic data, such as census records, could offer deeper insights into the evacuation experiences of these populations.

The present dataset needs to be interpreted within the local context. While no demographic information about the specific dataset was available, general census data[6] can be used to describe this case study. For example, according to census data, 27% of the residents are over 65 years old. Future, more detailed analysis is needed to tie the present work to the specific community characteristics.

Despite these limitations, our study has the potential to provide real-time traffic predictions and departure time analysis. This could empower stakeholders such as emergency managers, first responders, and residents with timely information to optimize their evacuation plans, reduce the risk of congestion, and improve overall evacuation efficiency.

Looking ahead, future research should explore additional factors influencing traffic dynamics and evacuation behavior during wildfires, such as weather conditions, road infrastructure, and individual preferences. Incorporating advanced analytics techniques and real-time data streams could enhance the accuracy of traffic predictions and evacuation analyses, providing more comprehensive insights into emergency scenarios. Furthermore, the scalability of this methodology could be extended to other disaster contexts, such as hurricanes, floods, or earthquakes, to improve emergency preparedness and response efforts on a broader scale.

7. Conclusion

In this research, we provided a detailed departure time analysis of a wildfire evacuation from a small community, using the case of Lytton BC. This work offers insights into the evacuation behavior of residents during wildfire events, in particular on evacuation timing and route choice. By analyzing stay point detection data and daily separation of residents before and during the wildfire, we quantified the departure times of individuals from the village. This analysis provides information on the timing and magnitude of evacuations, enabling authorities to better understand evacuation patterns and optimize resource allocation and evacuation planning strategies accordingly.

While this study focused on a small rural community, the methodology and models developed here are scalable and can be adapted to larger, more complex communities. The principles of traffic flow analysis, stay point detection, and evacuation behavior modeling can be applied to larger datasets in urban areas, where population density and road network complexity are greater.

Additionally, we developed a comprehensive methodology for predicting traffic speeds on highways during wildfire evacuations using GPS data and machine learning techniques. Through a series of steps including GPS data preprocessing, calculation of hourly flow on highways, synchronization of flow and speed data, training a linear regression model, and speed prediction, we aimed to provide valuable insights into traffic dynamics during emergency scenarios. By leveraging real-time data and predictive modeling, we sought to enhance emergency response strategies and improve evacuation planning to mitigate the impact of wildfires on affected communities. Future work could explore the model's performance in larger settings, but the approach presented here remains relevant for a variety of community sizes, offering valuable insights into improving evacuation strategies across different scenarios.

CRediT authorship contribution statement

Bahareh Raei: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Max Kinateder:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Noureddine Bénichou:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Islam Gomaa:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Xin Wang:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Bruce Spencer from National Research Council Canada for his support and the National Research Council Canada, Canada and the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada for financial support as well as Quadrant for making the GPS data available.

[6] https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/search-recherche/results-resultats.cfm?Lang=E&SearchText=Lytton

Data availability

The authors do not have permission to share data.

References

[1] K.B. Marvel, Research reports: An open market for scientific verbiage? Nature 551 (7680) (2017) 300, http://dx.doi.org/10.1038/d41586-017-06031-3, URL https://www.nature.com/articles/d41586-017-06031-3.

[2] Y. Liu, J. Stanturf, S. Goodrick, Trends in global wildfire potential in a changing climate, Forest Ecol. Manag. 259 (4) (2010) 685–697, http://dx.doi.org/10.1016/j.foreco.2009.09.002, URL https://linkinghub.elsevier.com/retrieve/pii/S0378112709006148.

[3] S. McCaffrey, R. Wilson, A. Konar, Should I stay or should I go now? or should I wait and see? influences on wildfire evacuation decisions, Risk Anal. 38 (7) (2018) 1390–1404, http://dx.doi.org/10.1111/risa.12944, URL https://onlinelibrary.wiley.com/doi/10.1111/risa.12944.

[4] S. Erni, X. Wang, T. Swystun, S.W. Taylor, M.-A. Parisien, F.-N. Robinne, B. Eddy, J. Oliver, B. Armitage, M.D. Flannigan, Mapping wildfire hazard, vulnerability, and risk to Canadian communities, Int. J. Disaster Risk Reduct. 101 (2024) 104221.

[5] T.J. Cova, Y. Sun, X. Zhao, Y. Liu, E.D. Kuligowski, N. Janfeshanaraghi, R. Lovreglio, Destination unknown: Examining wildfire evacuee trips using GPS data, J. Transp. Geogr. 117 (2024) 103863.

[6] J. Cohen, A. Westhaver, An examination of the Lytton, British Columbia wildlandurban fire destruction: Summary report to the British Columbia FireSmart Committee, Institute for Catastrophic Loss Reduction, 2022, https://www.iclr.org/wpcontent/uploads/2022/05/Lytton-Wildfires.pdf.

[7] R. Lovreglio, E. Kuligowski, S. Gwynne, K. Strahan, A modelling framework for householder decision-making for wildfire emergencies, Int. J. Disaster Risk Reduct. 41 (2019) 101274, http://dx.doi.org/10.1016/j.ijdrr.2019.101274, URL https://linkinghub.elsevier.com/retrieve/pii/S221242091930041X.

[8] K. Zhang, H. Li, K. Torkkola, M. Gardner, Adaptive learning of semantic locations and routes, in: J. Hightower, B. Schiele, T. Strang (Eds.), Location- and Context-Awareness, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 193–210.

[9] X. Zhao, Y. Xu, R. Lovreglio, E. Kuligowski, D. Nilsson, T.J. Cova, A. Wu, X. Yan, Estimating wildfire evacuation decision and departure timing using large-scale GPS data, Transp. Res. Part D: Transp. Environ. 107 (2022) 103277, http://dx.doi.org/10.1016/j.trd.2022.103277, URL https://linkinghub.elsevier.com/retrieve/pii/S136192092200102X.

[10] E. Kuligowski, Evacuation decision-making and behavior in wildfires: Past research, current challenges and a future research agenda, Fire Saf. J. 120 (2021) 103129, http://dx.doi.org/10.1016/j.firesaf.2020.103129, URL https://linkinghub.elsevier.com/retrieve/pii/S0379711220302204.

[11] Whitefish Lake First Nation 459, A.C. Christianson, T.K. McGee, Wildfire evacuation experiences of band members of Whitefish Lake First Nation 459, Alberta, Canada, Nat. Hazards 98 (1) (2019) 9–29, http://dx.doi.org/10.1007/s11069-018-3556-9, URL http://link.springer.com/10.1007/s11069-018-3556-9.

[12] T. Toledo, I. Marom, E. Grimberg, S. Bekhor, Analysis of evacuation behavior in a wildfire event, Int. J. Disaster Risk Reduct. 31 (2018) 1366–1373, http://dx.doi.org/10.1016/j.ijdrr.2018.03.033, URL https://linkinghub.elsevier.com/retrieve/pii/S2212420918304059.

[13] A. Wu, X. Yan, E. Kuligowski, R. Lovreglio, D. Nilsson, T.J. Cova, Y. Xu, X. Zhao, Wildfire evacuation decision modeling using GPS data, Int. J. Disaster Risk Reduct. 83 (2022) 103373, http://dx.doi.org/10.1016/j.ijdrr.2022.103373, URL https://linkinghub.elsevier.com/retrieve/pii/S2212420922005921.

[14] A. Fathianpour, B. Evans, M.B. Jelodar, S. Wilkinson, Tsunami evacuation modelling via micro-simulation model, Progr. Disaster Sci. 17 (2023) 100277, http://dx.doi.org/10.1016/j.pdisas.2023.100277, URL https://linkinghub.elsevier.com/retrieve/pii/S2590061723000042.

[15] S. Ahmad, A. Ali, H.U. Ahmed, Y. Huang, P. Lu, Evaluating Traffic Operation Conditions during Wildfire Evacuation Using Connected Vehicles Data, Fire 6 (5) (2023) 184, http://dx.doi.org/10.3390/fire6050184, URL https://www.mdpi.com/2571-6255/6/5/184.

[16] L.M. Johnston, M.D. Flannigan, Mapping Canadian wildland fire interface areas, Int. J. Wildland Fire 27 (1) (2017) 1–14.

[17] C.M. Edgeley, T.B. Paveglio, Exploring influences on intended evacuation behaviors during wildfire: What roles for pre-fire actions and event-based cues? Int. J. Disaster Risk Reduct. 37 (2019) 101182.

[18] E.D. Kuligowski, X. Zhao, R. Lovreglio, N. Xu, K. Yang, A. Westbury, D. Nilsson, N. Brown, Modeling evacuation decisions in the 2019 Kincade fire in California, Saf. Sci. 146 (2022) 105541, http://dx.doi.org/10.1016/j.ssci.2021.105541, URL https://linkinghub.elsevier.com/retrieve/pii/S0925753521003842.

[19] N. Xu, R. Lovreglio, E.D. Kuligowski, T.J. Cova, D. Nilsson, X. Zhao, Predicting and assessing wildfire evacuation decision-making using machine learning: findings from the 2019 kincade fire, Fire Technol. 59 (2) (2023) 793–825, http://dx.doi.org/10.1007/s10694-023-01363-1, URL https://link.springer.com/10.1007/s10694-023-01363-1.

[20] D. Zheng, H. Wang, L. Hong, M. Zhou, W. Zhuang, Y. Hao, Experiment on evacuation behavior: Applying the sector grid model to a clear area, Int. J. Disaster Risk Reduct. 100 (2024) 104159.

[21] D. Li, A data-driven approach to improving evacuation time estimates during wildfires for communities with part-time residents in the wildland-urban interface, Int. J. Disaster Risk Reduct. 82 (2022) 103363.

[22] T. Yabe, Y. Sekimoto, K. Tsubouchi, S. Ikemoto, Cross-comparative analysis of evacuation behavior after earthquakes using mobile phone data, PLOS ONE 14 (2) (2019) e0211375, http://dx.doi.org/10.1371/journal.pone.0211375, URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0211375. Publisher: Public Library of Science.

[23] A. Muhammad, R. De Risi, F. De Luca, W. Kongko, N. Mori, T. Yasuda, K. Goda, Integrated tsunami risk framework considering agent-based evacuation modelling: The case of Saga, Kochi prefecture, Japan, Int. J. Disaster Risk Reduct. 101 (2024) 104193.

[24] B. Melendez, S. Ghanipoor Machiani, A. Nara, Modelling traffic during Lilac wildfire evacuation using cellular data, Transp. Res. Interdisc. Perspect. 9 (2021) 100335, http://dx.doi.org/10.1016/j.trip.2021.100335, URL https://linkinghub.elsevier.com/retrieve/pii/S2590198221000427.

[25] A. Rohaert, E.D. Kuligowski, A. Ardinge, J. Wahlqvist, S.M. Gwynne, A. Kimball, N. Bénichou, E. Ronchi, Traffic dynamics during the 2019 Kincade wildfire evacuation, Transp. Res. Part D: Transp. Environ. 116 (2023) 103610, http://dx.doi.org/10.1016/j.trd.2023.103610, URL https://linkinghub.elsevier.com/retrieve/pii/S136192092300007X.

[26] X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, B. Zhang, Urban traffic congestion estimation and prediction based on floating car trajectory data, Future Gener. Comput. Syst. 61 (2016) 97–107, http://dx.doi.org/10.1016/j.future.2015.11.013, URL https://linkinghub.elsevier.com/retrieve/pii/S0167739X15003611.

[27] Z. Wang, M. Lu, X. Yuan, J. Zhang, H. Van De Wetering, Visual traffic jam analysis based on trajectory data, IEEE Trans. Vis. Comput. Graphics 19 (12) (2013) 2159–2168, http://dx.doi.org/10.1109/TVCG.2013.228, URL https://ieeexplore.ieee.org/document/6634174/.

[28] N. Ferreira, J. Poco, H.T. Vo, J. Freire, C.T. Silva, Visual exploration of big spatio-temporal urban data: A study of New York city taxi trips, IEEE Trans. Vis. Comput. Graphics 19 (12) (2013) 2149–2158, http://dx.doi.org/10.1109/TVCG.2013.226, URL http://ieeexplore.ieee.org/document/6634127/.

[29] Z. Zhang, Y. Liu, Q. Tong, S. Guo, D. Li, Evacuation based on spatio-temporal resilience with variable traffic demand, J. Manag. Sci. Eng. 6 (1) (2021) 86–98, http://dx.doi.org/10.1016/j.jmse.2021.02.009, URL https://www.sciencedirect.com/science/article/pii/S2096232021000111.

[30] N. Polakis, A. Tsouchlaraki, Strategy for effective evacuation of pedestrians using GIS and GPS measurements, J. Geogr. Inf. Syst. 14 (5) (2022) 487–502, http://dx.doi.org/10.4236/jgis.2022.145027, URL http://www.scirp.org/Journal/Paperabs.aspx?paperid=120553. Number: 5 Publisher: Scientific Research Publishing.

[31] B. Yang, S. Sun, J. Li, X. Lin, Y. Tian, Traffic flow prediction using LSTM with feature enhancement, Neurocomputing 332 (2019) 320–327.

[32] P.K. Anyidoho, R.A. Davidson, T. Rambha, L.K. Nozick, Prediction of population behavior in hurricane evacuations, Transp. Res. Part A: Policy Pract. 159 (2022) 200–221.

[33] Y. Song, X. Yan, A method for formulizing disaster evacuation demand curves based on SI model, Int. J. Environ. Res. Public Health 13 (10) (2016) http://dx.doi.org/10.3390/ijerph13100986, URL https://www.mdpi.com/1660-4601/13/10/986.

[34] Y. Xu, X. Zhao, R. Lovreglio, E. Kuligowski, D. Nilsson, T.J. Cova, X. Yan, A highway vehicle routing dataset during the 2019 Kincade fire evacuation, Sci. Data 9 (1) (2022) 608, http://dx.doi.org/10.1038/s41597-022-01731-6, URL https://www.nature.com/articles/s41597-022-01731-6.

[35] J. Hu, P. Li, S. Bisnath, Enhancing smartphone precise point positioning to sub-meter accuracy in suburban environments: a new stochastic model and outlier diagnosis, GPS Solut. 28 (3) (2024) 1–11.

[36] J. Wolf, R. Guensler, W. Bachman, Elimination of the travel diary: An experiment to derive trip purpose from GPS travel data, 2001, in: Transportation Research Board 80th Annual Meeting, pp. 7–11.

[37] M.C. González, C.A. Hidalgo, A.-L. Barabási, Understanding individual human mobility patterns, Nature 453 (7196) (2008) 779–782, http://dx.doi.org/10.1038/nature06958, URL https://www.nature.com/articles/nature06958.

[38] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, L. Terveen, Discovering personally meaningful places: An interactive clustering approach, ACM Trans. Inf. Syst. 25 (3) (2007) 12, http://dx.doi.org/10.1145/1247715.1247718, URL https://dl.acm.org/doi/10.1145/1247715.1247718.

[39] A.T. Palma, V. Bogorny, B. Kuijpers, L.O. Alvares, A clustering-based approach for discovering interesting places in trajectories, in: Proceedings of the 2008 ACM Symposium on Applied Computing, ACM, Fortaleza, Ceara Brazil, 2008, pp. 863–868, http://dx.doi.org/10.1145/1363686.1363886, URL https://dl.acm.org/doi/10.1145/1363686.1363886.

[40] J.H. Kang, W. Welbourne, B. Stewart, G. Borriello, Extracting places from traces of locations, ACM SIGMOBILE Mob. Comput. Commun. Rev. 9 (3) (2005) 58–68, http://dx.doi.org/10.1145/1094549.1094558, URL https://dl.acm.org/doi/10.1145/1094549.1094558.

[41] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, W.-Y. Ma, Mining user similarity based on location history, in: Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, Irvine California, 2008, pp. 1–10, http://dx.doi.org/10.1145/1463434.1463477, URL https://dl.acm.org/doi/10.1145/1463434.1463477.

[42] D.H. Hu, C.-L. Wang, GPS-based location extraction and presence management for mobile instant messenger, in: T.-W. Kuo, E. Sha, M. Guo, L.T. Yang, Z. Shao (Eds.), Embedded and Ubiquitous Computing, in: Lecture Notes in Computer Science, vol. 4808, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 309–320, http://dx.doi.org/10.1007/978-3-540-77092-3_27, URL http://link.springer.com/10.1007/978-3-540-77092-3_27.

[43] P. Nurmi, S. Bhattacharya, Identifying meaningful places: The non-parametric way, in: J. Indulska, D.J. Patterson, T. Rodden, M. Ott (Eds.), Pervasive Computing, in: Lecture Notes in Computer Science, 5013, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 111–127, http://dx.doi.org/10.1007/978-3-540-79576-6_7, URL http://link.springer.com/10.1007/978-3-540-79576-6_7.

[44] C. Lee, G. Yoon, D. Han, A probabilistic place extraction algorithm based on a superstate model, IEEE Trans. Mob. Comput. 12 (5) (2013) 945–956, http://dx.doi.org/10.1109/TMC.2012.64, URL http://ieeexplore.ieee.org/document/6175020/.

[45] H. Mahmoud, N. Akkari, Shortest path calculation: A comparative study for location-based recommender system, in: 2016 World Symposium on Computer Applications & Research, WSCAR, 2016, pp. 1–5, http://dx.doi.org/10.1109/WSCAR.2016.16, URL https://ieeexplore.ieee.org/abstract/document/7791971?casa_token=D5vZAnd7ocMAAAAA:mjygqs-CDp5IVlY1V8DSa-0I8NZMuIM3p2ohzG0b9AWtE6bC5DQBVsuuNvsHwcsknO1-6jagiQ.

[46] C.C. Robusto, The cosine-haversine formula, Amer. Math. Monthly 64 (1) (1957) 38–40.

[47] R. Safarzadeh, X. Wang, Map matching on low sampling rate trajectories through deep inverse reinforcement learning and multi-intention modeling, Int. J. Geogr. Inf. Sci. (2024) 1–36.

[48] R.S. Ramhormozi, A. Mozhdehi, S. Kalantari, Y. Wang, S. Sun, X. Wang, Multi-task graph neural network for truck speed prediction under extreme weather conditions, in: Proceedings of the 30th International Conference on Advances in Geographic Information Systems, ACM, Seattle Washington, 2022, pp. 1–11, http://dx.doi.org/10.1145/3557915.3561029, URL https://dl.acm.org/doi/10.1145/3557915.3561029.

[49] J. Zhao, Y. Gao, Z. Yang, J. Li, Y. Feng, Z. Qin, Z. Bai, Truck traffic speed prediction under non-recurrent congestion: Based on optimized deep learning algorithms and GPS data, IEEE Access 7 (2019) 9116–9127.

[50] I. Alam, D.M. Farid, R.J.F. Rossetti, The prediction of traffic flow with regression analysis, in: A. Abraham, P. Dutta, J.K. Mandal, A. Bhattacharya, S. Dutta (Eds.), Emerging Technologies in Data Mining and Information Security, in: Advances in Intelligent Systems and Computing, vol. 813, Springer Singapore, Singapore, 2019, pp. 661–671, http://dx.doi.org/10.1007/978-981-13-1498-8_58, URL http://link.springer.com/10.1007/978-981-13-1498-8_58.

[51] X. Tian, Q. Zheng, Z. Yu, M. Yang, Y. Ding, A. Elhanashi, S. Saponara, K. Kpalma, A real-time vehicle speed prediction method based on a lightweight informer driven by big temporal data, Big Data Cogn. Comput. 7 (3) (2023) http://dx.doi.org/10.3390/bdcc7030131, URL https://www.mdpi.com/2504-2289/7/3/131.

[52] I.K. Singgih, M.L. Singgih, Regression machine learning models for the short-time prediction of genetic algorithm results in a vehicle routing problem, World Electr. Veh. J. 15 (7) (2024) 308.

[53] A. Labhiri, S. Vaiciulyte, E. Kuligowski, E. Ronchi, Evacuation decisions of tourists in wildfire scenarios, Int. J. Disaster Risk Reduct. 113 (2024) 104836, http://dx.doi.org/10.1016/j.ijdrr.2024.104836, URL https://www.sciencedirect.com/science/article/pii/S2212420924005983.