Transportation Research Interdisciplinary Perspectives 28 (2024) 101261

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives

Using crowd-sourced traffic data and open-source tools for urban congestion analysis

Khaula Alkaabi [a,*], Mohsin Raza [b], Esra Qasemi [c], Hafsah Alderei [c], Mazoun Alderei [c], Sharina Almheiri [c]

[a] *Department of Geography and Urban Sustainability, Emirates Center for Mobility Research, United Arab Emirates University, Al Ain City, P.O. Box 15551, United Arab Emirates*
[b] *Transport Planning and Management, The Urban Unit Private Limited, Lahore, Pakistan*
[c] *Department of Geography and Urban Sustainability, United Arab Emirates University, Al Ain City, P.O. Box 15551, United Arab Emirates*

ARTICLE INFO

Keywords:
Traffic congestion
Crowdsourced data
Open-source tools
Middle east
HERE
Google Maps

ABSTRACT

Traffic congestion in urban areas poses significant challenges to city dwellers and consultants advising government. This study explores innovative methods to monitor and control traffic congestion, focusing on Al Ain city in the United Arab Emirates. Using the R Programming language and harnessing crowdsourced traffic information from HERE and Google Maps, the research delves into spatial data analysis. The methodology employed in this study builds on the previously applied congestion modeling methods for cities like Windsor, Toronto, and New York. The study focuses on Al Ain, addressing the scarcity of crowdsourced information-based congestion modeling research in the Middle East. The study details how to obtain and deploy crowdsourced traffic data, speed and jam factors, for a comprehensive visualization of the urban traffic congestion. For example, in the case of Al Ain, analysis showed an average traffic speed of 43 km per hour in Al Ain, where infrastructure could otherwise allow an average traffic speed of up to 51 km per hour under free flow conditions. The study findings highlight how traffic conditions, rather than speed limits, cause traffic flow disruptions in the city, which can inform traffic regulations. The study's high-confidence real-time data emphasizes the reliability of crowdsourced traffic flow data. This research demonstrates the applicability of open-source traffic information for congestion modeling in the UAE, and establishes a replicable methodology for other urban areas worldwide, contributing significantly to the modeling methods.

1. Introduction

Understanding urban transport congestion is crucial due to its implications for the economy, environment, public, and society. Significant productivity losses are associated with transport delays (Sciomachen & Stecca, 2023). The social costs of congestion include stress and lower quality of life for urban residents, decreased leisure time, and hindered social interactions (Thomson & Bull, 2002). The public health challenges associated with congestion include increased concentrations of pollutants such as nitrogen oxide and carbon monoxide (Alkaabi & Abuelgasim, 2022; Tonne et al., 2008). Therefore, metropolitan areas with a population exceeding 200,000 in the United States mandate a congestion management system (Hudson, 2002). However, congestion modeling and control practices have historically relied on expensive and labor-intensive methods, which are not only costly but also dependent on extrapolation and assumptions for a citywide analysis. Thus, the costs and errors associated with congestion management studies make it challenging for professionals, advising policymakers and public representatives, to justify investing in traffic studies for congestion management.

Despite calls for a shift towards electronic tools and crowdsourced traffic information more than two decades ago (Benz & Ogden, 1996; Quiroga, 2000), the current transportation engineering practices around measuring, monitoring, and addressing transport congestion continue to rely on manual, video-based, or sensor-based traffic counts (Noor et al., 2021). Manual traffic counting methods have their roots in traffic flow theories and computer simulation techniques developed around the mid to late 20th century (Drew, 1968; Gerlough & Huber, 1976). The prevailing state in engineering practice suggests a notable oversight by traffic engineers regarding the transportation ramifications of

* Corresponding author.
E-mail address: khaula.alkaabi@uaeu.ac.ae (K. Alkaabi).

https://doi.org/10.1016/j.trip.2024.101261
Received 23 November 2023; Received in revised form 1 August 2024; Accepted 20 October 2024
Available online 5 November 2024

technological advancements in the early 2000s. Improved GIS tools and high smartphone penetration rates could help collect and integrate geospatial data with crowdsourced traffic speeds, volumes, and incidents. Global providers of navigation services, such as Google, *HERE*, Mapbox, and TomTom, can now monitor congestion at high-resolution levels. Hence, there is a growing share of navigation service providers in congestion management programs. A case in point is Google's Greenlight, which was recently assigned to enhance traffic signal performance at 70 intersections across 12 cities worldwide (Mattias, 2023).

This study attempts to fill this gap by explaining traffic flow data extraction methods for congestion modelling using open-source tools. It will specifically appeal to UAE transportation engineers and policymakers since previous regional studies (Alkaabi, 2023; Nair et al., 2019) did not elaborate on traffic data collection methods. Nair et al. (2019) focused on data analysis using Google Maps API, while Alkaabi (2023) identified accident hotspots with GIS but did not address data collection. Therefore, this study uniquely emphasizes traffic data collection in Al Ain city. It study employs a use-case approach to demonstrate how to obtain urban traffic flow data for Al Ain city in the United Arab Emirates (UAE) and investigates the capabilities and limitations of open-source technologies as an alternative to traditional modeling practices. The study might add to the congestion monitoring and management methods as the extant literature focuses on applications and the architecture of open-source tools rather than traffic-data-obtaining techniques (Droj et al., 2021; Lovelace, 2021). This lack of guidance makes it difficult for scholars and practitioners to repeat the experiments. This study can serve as a reference for traffic data collection and, thus, promote wide-scale adoption of the previously mentioned tools and contribute towards informed decision-making on congestion management.

2. Literature review

The gap between the available technology and transportation engineering practices around congestion management is significant. However, the literature is not at par either. There are no comprehensive guides to help traffic engineers and scholars access georeferenced crowdsourced traffic data freely available for exploratory and small-scale studies. Various scholars who accessed and used the data for their studies focused more on the use cases rather than collection methodologies. For example, Naiudomthum et al. (2022) suggested using the spatiotemporal traffic distribution via Google Maps API to model traffic emissions in Bangkok. Cuervo et al. (2022) used open-source traffic congestion models to find the impact of congestion on car accessibility to tertiary emergency care services in Cali, Colombia. Schirck-Matthews et al. (2022) used Google Maps and MapQuest to compare route preferences of commute and sports cyclists between Miami-Dade and North Holland. Ibad et al. (2020) demonstrated the usefulness of Google Maps in transport policymaking for Bander Lamburg, Indonesia. However, these studies did not elaborate on obtaining traffic data, required tools and expertise, methodology limitations, obtained traffic information, or compatibility issues with the data.

Crowdsource technologies offer free and paid access to traffic information and analysis of spatial data. However, the focus of the study is to describe methods that can help access traffic flow data free of cost, which may be appropriate for many scholars and transport engineers with limited traffic data needs. There are many service providers in this regard; however, the study focuses on HERE and Google Maps for traffic speeds and congestion modeling and R for scripting code to access these services via application program interfaces (APIs). An overview of related R packages is available in the relevant literature but is summarized in Table 1 for quick reference (Marty, 2023; Unterfinger, 2023).

The GoogleTraffic package establishes an interface with the Google Maps API, enabling the acquisition of traffic congestion modeling data presented as a georeferenced raster (Marty, 2023). Each pixel's color in the raster represents a congestion level on a four-point scale, where 1 indicates minimal congestion (free-flow condition) denoted by green, and 4 signifies a traffic jam depicted by dark red (Fig. 1). However, unlike hereR, GoogleTraffic lacks detailed traffic flow data, limiting the ability to precisely identify the causes of traffic congestion. Additionally,

Table 1
Selected tools and their roles in crowdsourced congestion modelling.

Tools	Functions	Comments
HERE	Provide traffic flow data.	Accessing the service without charge necessitates the generation of an API key, and notably, it does not require the provision of any financial details such as credit card information. The basic version permits up to 1000 free requests per day, with a maximum allowable rate of 5 requests per second.
Google Maps	Provide congestion levels.	Obtaining a Google Maps API key is a cost-free process, albeit it necessitates the creation of a billing account and the submission of financial (credit card) details. It provides a welcome credit, albeit with limited validity, applicable for utilization of its cloud service, which should be adequate for the majority of city-wide transport analysis studies.
hereR	R package used to create an interface with various APIs including Geocoder; Routing, and Traffic	Helps access real-time (or typical) traffic speeds and congestion modelling data for selected regions (HERE, 2023).
GoogleTraffic	R package used to create an interface with Google Maps	Helps access congestion modeling data from GoogleMaps.
R	A programming tool serves as a platform to integrate the aforementioned packages and script the development of interfaces with crowdsource traffic information providers such as *HERE*, Google Maps, Mapquest, etc.	Free to use but requires considerable scripting expertise and has a steep learning curve.

Fig. 1. Example raster produced using GoogleTraffic showing traffic within Washington, DC (Marty, 2023).

the output from the GoogleMaps package is in raster format rather than a vector, posing challenges for seamless integration with other geospatial data, such as land use, for further analysis.

This study is designed to address the following research questions (RQs) and achieve corresponding research objectives:

RQ1: What are the key open-source urban transport congestion modeling tools, and what key computational and human resources are needed to utilize them effectively?

Objectives:

- Review and compile a list of potential open-source traffic data crowd-sources, programming tools, and program application packages (APIs).
- Identify the modeling and capacity requirements necessary for professionals to effectively utilize these tools for congestion measurement.

RQ2: To what extent can these tools model real-time and typical levels of urban transport congestion?

Objectives:

- Measure traffic speed and related parameters using crowd-sourced traffic information services and open-source tools to create a congestion model of Al Ain city (UAE) for a real-world application.
- Analyze the usability of output file formats for further analysis, such as georeferencing and overlaying on road network and land-use data.

RQ3: What are the implications of these tools for transport engineers advising government on social, public health, environment, and economic development sectors?

Objectives:

- Analyze how newer tools can empower resource-constrained professionals advising government on policy interventions related to social interactions, air quality, sustainable transport, and businesses.

RQ4: What are the limitations of urban transport congestion measurement tools?

Objectives:

- Analyze limitations of the crowdsource data providers in practical applications, considering factors like data availability, model accuracy, and usability in diverse urban contexts.

3. Methodology

This section has been divided into sub-sections for ease of understanding. These subsections describe the study area, the variables of interest (collected data), and their definitions.

3.1. Study area and data extraction

To showcase the capabilities of open-source tools, traffic flow information, encompassing speeds and jam factor, was gathered for Al Ain, United Arab Emirates (UAE). The UAE was selected as a test case to assess the traffic data coverage of the chosen crowdsourced technologies in the Middle East. Al Ain, the fourth-largest city in the UAE, is home to over 700,000 people. Traffic data collection utilized Google Maps and HERE technology services. Virtual bounding boxes were established around Al Ain city to limit traffic information collection to the city's geographic extent. Traffic speed, a widely acknowledged parameter for modeling traffic congestion (Noor et al., 2021; Unterfinger, 2023), was chosen as the key indicator for traffic flow conditions. Note that in this paper, congestion modeling implies congestion visualizations obtained

Variable	Description	Unit
Speed	The average speed, capped by the speed limit, that current traffic is travelling.	meters per sec
***speed**dUncapped*	The average speed, not capped by the speed limit, that current traffic is travelling.	meters per sec
Freeflow	The free flow speed speed is a reference speed provided to indicate the speed on the segment at which vehicles should be considered to be able to travel without impediment.	meters per sec
jamFactor	The number between 0.0 and 10.0 indicating the expected quality of travel. When there is a road closure, the jamFactor will be 10.0. As the value approaches 10.0 the quality of travel is getting worse.	
Confidence	The number between 0.0 and 1.0 indicates the proportion of real time data included in the speed calculation. 0.7 < confidence <= 1.0 indicates real time speeds 0.5 < confidence <= 0.7 indicates historical speeds 0.0 < confidence <= 0.5 indicates speed limit This field can be used to identify whether the data for a location is derived from real time probe sources or historical information only. All confidence data 0.71 and above is based on real-time information, where a confidence value of 0.75 or greater indicates high confidence real-time information. A confidence value equal to 0.70 means that the data is derived from historical data only.	
Traversability	"open" / "closed" / "reversibleNotRoutable" Indicates whether this roadway can be driven.	

Fig. 2. Description of congestion modelling variables (Durrani, 2023).

by plotting traffic flow/jam factor data on the road network.

Fig. 2 delineates various forms of traffic speeds and other variables measured in this study. The "Speed" variable indicates the average traffic speed constrained by the speed limit, while "SpeedUncapped" reflects the average traffic speed not restricted by traffic speed limits. If the speed is lower than "SpeedUncapped," it implies that the speed limits are overly restrictive, contributing to congestion. Conversely, if both these variables are the same but traffic is still congested (Speed < Freeflow), the cause of the congestion might be higher traffic volumes, reducing traffic speed compared to the so-called free-flow speeds, where vehicle speeds are unaffected by surrounding traffic. These variables can be obtained using the hereR package, which creates an interface with the HERE Traffic API. Further context and definitions are available elsewhere (Durrani, 2023; Unterfinger, 2023).

3.2. Software packages and resource requirements

The analysis process involved installing R, Rtools, RStudio, and various R packages as outlined in the appended code (Appendix A and Appendix B). The necessary R packages, specifically hereR, and osmdata, were loaded to acquire and map crowd-sourced traffic information data, focusing on Al Ain city in the United Arab Emirates. An area of interest was defined by a bounding box (xmin: 51.49798, ymin: 22.63162, xmax: 56.01813, ymax: 25.25262) encompassing the urban area and surrounding outskirts of Al Ain city. The osmdata::getbb() function was employed to retrieve the road network for Al Ain City based on the specified bounding box. The flow() function, provided by the hereR package, was utilized to obtain traffic data. The corresponding script (R code) is shown in *Appendix A*, and this process can be scaled to obtain hourly and daily traffic information using GitHub Action, as demonstrated in Appendix B.

A concise list of open-source software and packages essential for traffic analysis is provided in Table 1, which offers ample exploration potential in most transport congestion modeling applications. These tools are not resource-intensive, and standard personal computers suffice for computational needs. R itself requires minimal storage (hardly 200 MBs), with additional storage needed for allied packages like Rtools (around three GBs). The storage of traffic information data on the cloud (Google Maps and HERE in this case) alleviates the computational burden on end-users. Spreadsheet tools such as Excel® and GIS tools like OpenStreetMaps and ArcGIS® can aid in interpreting, analyzing, and visualizing the spatial and temporal aspects of obtained traffic flow information.

4. Results

To enhance clarity, this section has been organized into subsections corresponding to the research questions and stated objectives of the study. These subsections explain various aspects of traffic flow in Al Ain, including its spatial distribution, descriptive statistics, land use, and temporal aspects of traffic flow.

4.1. Traffic situation in Al Ain

Table 2 summarizes the traffic situation in Al Ain based on traffic data obtained on October 29, 2023, at 09:30 pm. The table displays ten out of more than 1400 traffic speed observations made by the crowd-sourced data tool (HERE) at the specified instant. Upon further analysis, the average traffic speed on various roads in Al Ain was found to be 11.9 m per second or 43 km per hour. This is the capped speed, theoretically affected by speed limits. However, the uncapped speed was also found to be the same, indicating that the posted speed limits did not impact traffic flow rates. Nevertheless, the average free-flow speed was determined to be 14.28 m per second, equivalent to 51 km per hour, which represents the speed that the road infrastructure could accommodate without traffic conflicts. Thus, traffic speeds, on average, were reduced by 8 km per hour due to high traffic volumes. The confidence level in all instances being equal to or greater than 0.7 indicated that the traffic condition was real-time, not an average of historical values.

4.2. Spatial analysis

To visualize the extracted traffic flow data, an interactive map was generated using the mapview package. The mapview() function was employed to explore traffic flow patterns in the city, and popup tables were added to provide detailed information about key traffic parameters, including free flow speed, average speed, jam factor, and confidence level. Fig. 3 displays a visualization of traffic speed based on the obtained data for Al Ain City. The corresponding R script is shown in Appendix A. As anticipated, traffic speeds increase along the city outskirts, while the city center exhibits lower traffic speeds. This visualization is created in the R environment using one of the available base maps. Other default base map options for visualizations include Dark Matter and Positron (CartoDB), Esri World Imagery, and OpenTopoMap.

Fig. 4 demonstrates how the Dark Matter background can assist end-users in focusing on traffic speed data while providing sufficient spatial context to comprehend low-speed roads. The left image illustrates average traffic speeds on the Al Ain Road Network observed on October 29, 2023, at 9:30 pm, while the right image illustrates traffic speeds on November 06, 2023, at 10:16 am. Overall, the average traffic speeds on most roads, especially those on the city outskirts, were higher on November 06 during the daytime compared to October 29 in the evening.

(L) October 29, 2023, at 09:30 pm (R) November 06, 2023, at 10:16 am.

In addition to plotting travel speeds, plotting the Jam Factor can aid in identifying traffic congestion spots more easily (Fig. 5). As the jam factor increases, travel speeds decrease. However, it may be a bit counterintuitive for some users that the same color (yellow) indicates higher speeds in speed plots while signifying congestion in jam factor

Table 2
Traffic condition on the road network in Al Ain, as of October 29, 2023, at 09:30 pm.

Obs.	speed m/s	speed_uncapped	Freeflow	Jam_factor	Confidence	Traversability
1	10.00	10.00	12.50	2.1	0.7	Open
2	6.11	6.11	9.72	3.9	0.7	Open
3	20.00	20.00	19.17	0	0.92	Open
4	16.67	16.67	17.50	0.6	0.79	Open
5	12.78	12.78	17.78	2.3	0.83	Open
6	4.72	4.72	8.33	2.9	0.7	Open
7	12.22	12.22	14.72	1.8	0.83	Open
8	8.61	8.61	10.56	2	0.7	Open
9	12.22	12.22	15.83	1.3	0.72	Open
10	16.11	16.11	16.67	0.5	0.78	Open
Averages	11.94	11.94	14.28			

*Speeds are reported in meters per second in HERE output.

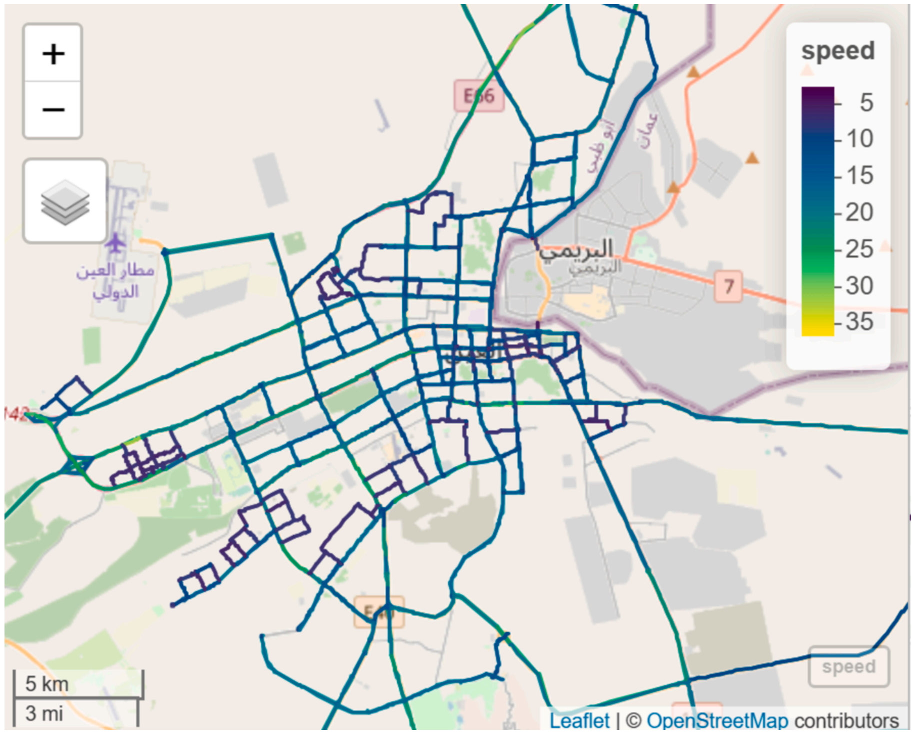

Fig. 3. Spatial Distribution of Traffic Condition in Al Ain (October 29, 2023, at 09:30 pm).

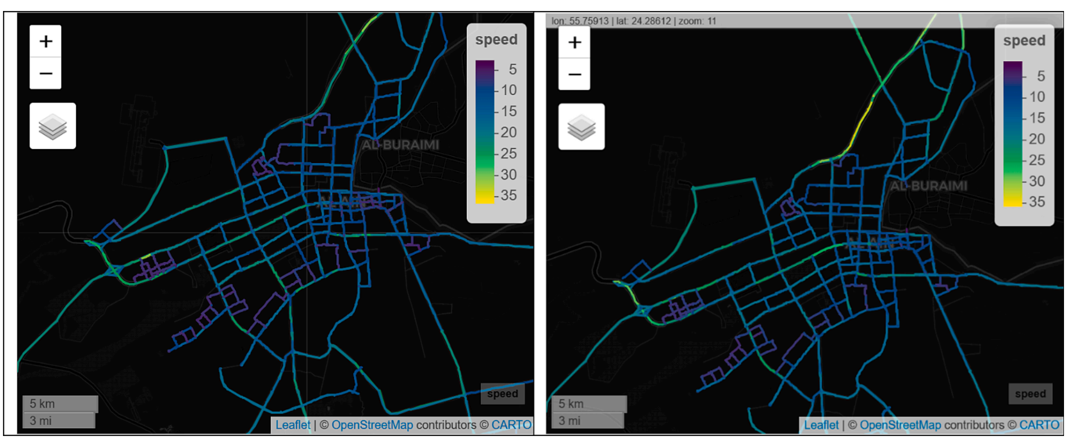

Fig. 4. Dak Matter Background to compare traffic speeds at two different instants.

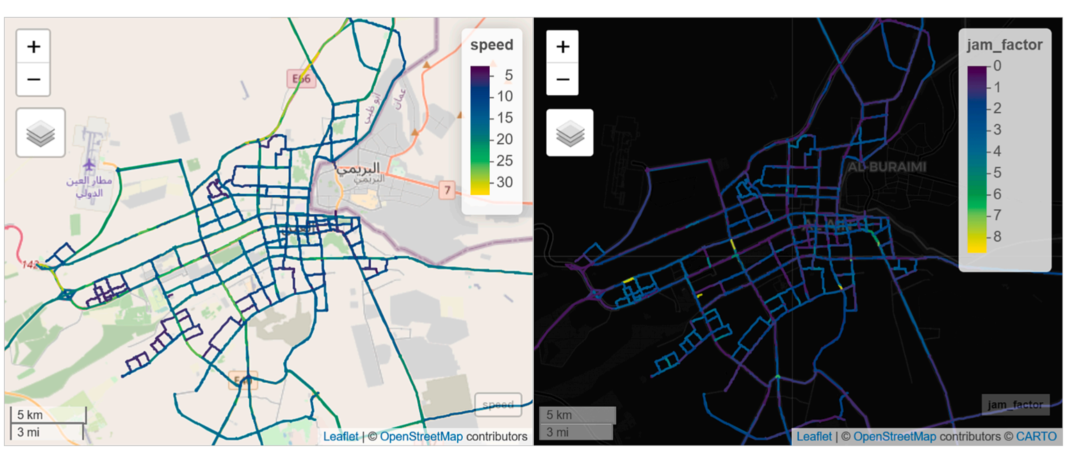

Fig. 5. Traffic Speed and Jam Factor on Al Ain Road Network, November 09 at 06:00 am.

plots.

While there could have been workarounds to change the representative colors in the R environment, it was recognized that this might become too cumbersome for readers with no prior experience working with R scripts. Those readers might prefer working with more familiar file extensions such as comma-separated values (csv) and shapefiles (shp). Therefore, data was exported as csv and shape files using the write () and st_write() functions, respectively. The former is a base R function, while the latter is an sf-package function. Off-the-shelf spreadsheet readers and statistical packages were employed to process csv files for further statistical analysis, such as creating speed histograms and scatterplots. Similarly, GIS tools were used to integrate traffic data with spatial information, such as key landmarks (schools, hotels, and shopping centers in Al Ain).

4.3. Descriptive statistics

Fig. 6a presents a histogram of observed traffic speeds for a randomly selected instant on a working day in Al Ain. The speed distribution curve exhibits unimodal positively (right) skewed distribution, implying that most of the observed speeds were higher than the mean. The data skewed left, suggesting that most motorists were driving at speeds higher than the mean (peak of the graph). Fig. 6b illustrates how the histogram would look if traffic speeds were not affected by nearby vehicles. In this case, the histogram shifts slightly towards the center, indicating the possibility of a more symmetrical distribution of traffic speeds. Interestingly, it also suggests the possibility of a bimodal distribution where two distinct traffic groups are possible if traffic speeds are not affected by congestion. The higher peak, occurring on the left side of the graph, possibly represents peak traffic speed in urban areas (8 m/s or 29 km/h), where there are more vehicles (higher frequency) but lower traffic speeds. The lower peak on the right side indicates the central tendency of higher driving speeds (around 16.3 m/s or 58.9 km/h), possibly on the peripheral roads in Al Ain.

To graphically summarize traffic speed statistics and depict median, quartiles, and outliers, Box and Whisker plots of traffic speeds were created (Fig. 7). The observed median speed, based on an analysis of 1483 speed observations in Al Ain, is 11.2 m/s or approximately 40 km/h. The interquartile range is 8.2 m/s to 16.2 m/s (30 km/h to 58 km/h), indicating driving speeds of the middle 50 % of observations. The 85th percentile speed is around 17 m/s or 61 km/h, which appears to be within the driving speed limits, assuming that the observations belong to vehicles on dual carriageways where speeds of up to 80 km/h are usually allowed. Speed limits in Al Ain differ depending on road type and area, warranting a spatial analysis of speeds based on corresponding area and road type. Outliers represent people driving much faster than their peers. The highest speed observed is 38.5 m/s or 139 km/h, which can be legal, for example, on Abu Dhabi-Al Ain Road. These findings might be of interest to law enforcement agencies.

To identify observations (vehicles) driving at a specific speed and later find their location, a scatterplot of speeds was developed (Fig. 8a). In the plot, each observation is represented by a point identified by speed and a unique ID. Slow-moving vehicles are fewer and randomly spread across the observations, which implies they are dispersed across the city. The concentration of points increases as the speeds rise to a middle level, and then there are fewer vehicles driving within high-speed bands. Fig. 8b shows a scatterplot of the Jam Factor. Most observations are concentrated around smaller jam factor values, indicating a reasonable level of service for most road users. As we move towards higher jam factor values, the concentration of points decreases, with less than 20 observations in the high-value range indicated by red points.

4.4. Congestion and land-use

To illustrate the possibility of integrating traffic information with spatial data such as key landmarks, Fig. 9 maps the crowdsource data with schools, cultural facilities, and shopping malls. Congestion factors are particularly high around the clusters of these traffic attractors. For example, in the Flaj-Al Hazza District of Al Ain City, there is a cluster of schools, shown with indigo-colored balls, where the surrounding road network is highly congested. Similarly, congestion can be observed around other schools and malls. Including more landmarks (hospitals, factories, community centers, etc.) and using broader land-use classifications (commercial, residential, industrial, etc.) can help explain more variations in congestion levels.

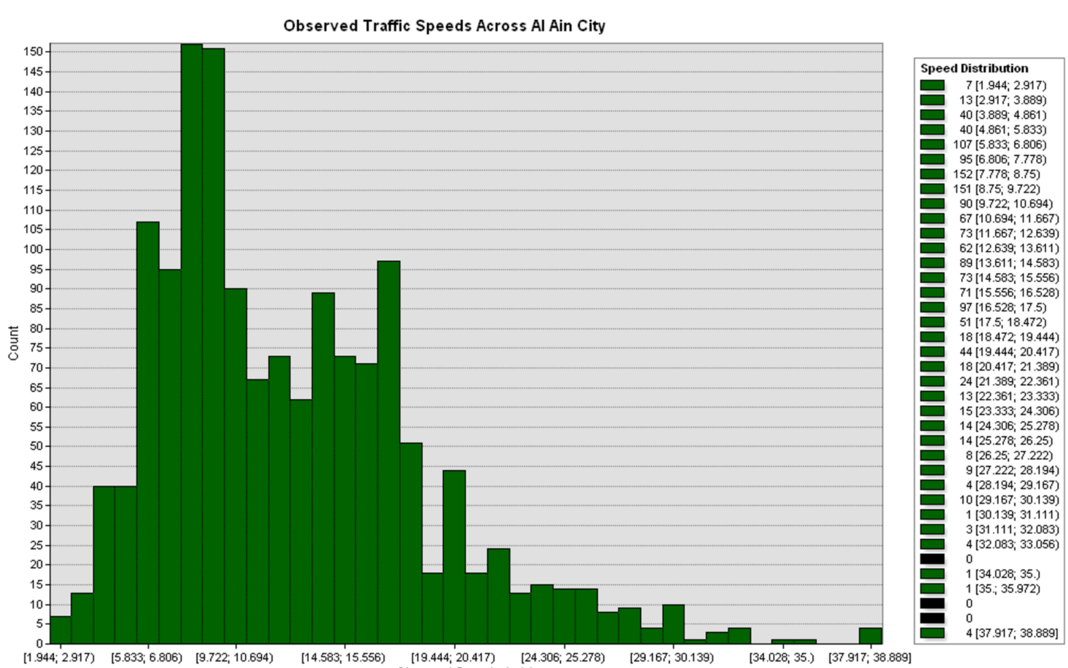

Fig. 6a. Observed traffic speeds in Al Ain on a randomly selected instant on a working day.

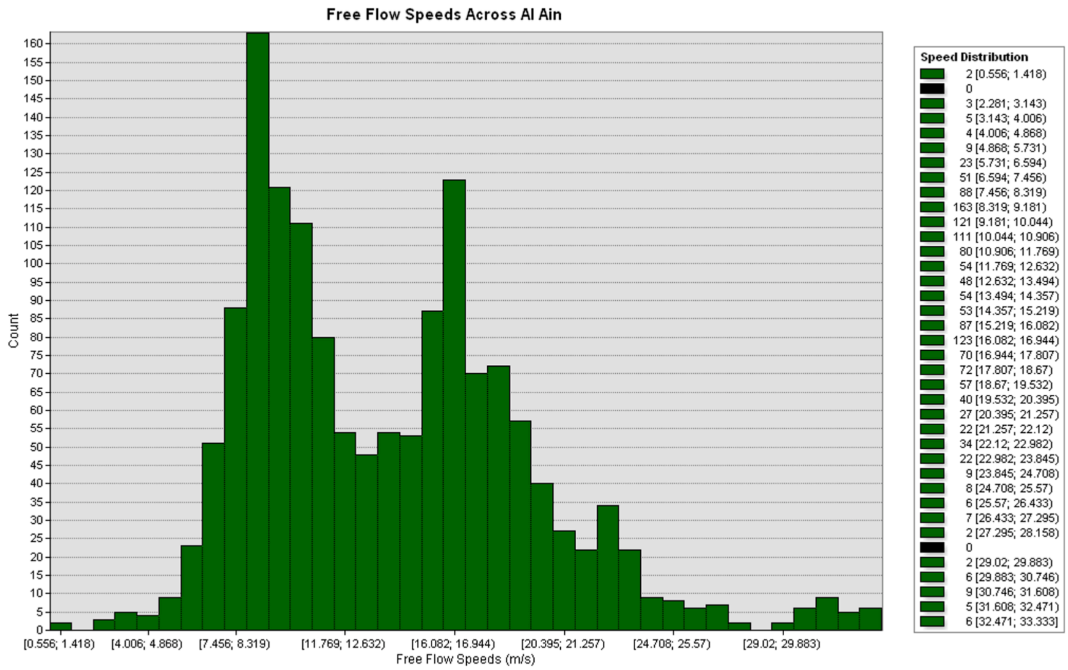

Fig. 6b. Estimated Free flow traffic speeds in Al Ain on a randomly selected instant on a working day.

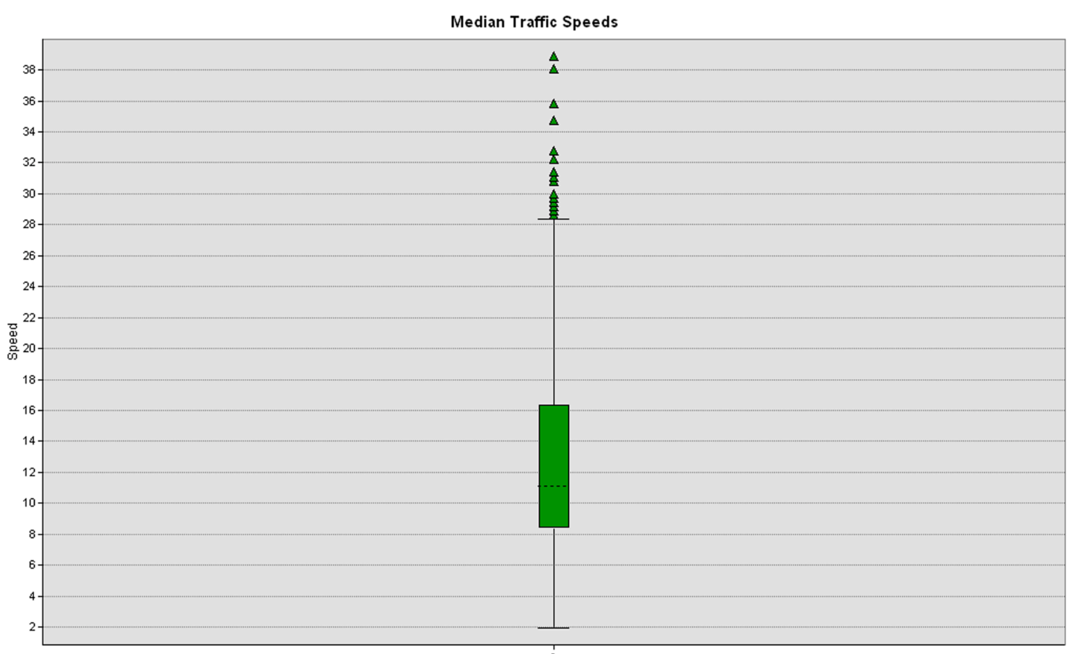

Fig. 7. Box and Whisker Plot of traffic speed at a randomly selected instant on a typical working day.

4.5. *Temporal analysis*

So far, the discussion has focused on the spatial aspects of traffic congestion; however, temporal distributions are equally important to explore. To determine when exactly congestion starts to build and how long it takes for traffic to recede to normal operation, the road network traffic was captured around school opening time (06:00 am-08:00 am) when students commute from home to school. Google Maps service was used to model typical traffic congestion levels around the school area in Al Ain. The service offers free unlimited access to both live and typical traffic congestion levels (from 6:00 am to 10:00 pm at 5-minute intervals). Fig. 10 shows temporal variations in congestion levels at half-hourly intervals (increasing from left to right). Traffic starts building up around 06:30 am, peaking at around 07:30 am before returning to lower levels by 08:00 am. A drawback of using the Google Maps service, however, is that traffic congestion can only be viewed on selected maps, and the authors could not find a way to download traffic flow information as a csv or shape file (vector tiles).

5. Discussions and conclusions

The downloaded shapefile of traffic speed and other flow parameters, such as jam factors, can assist transport engineers in correlating the impact of congestion on various aspects of everyday life, including the

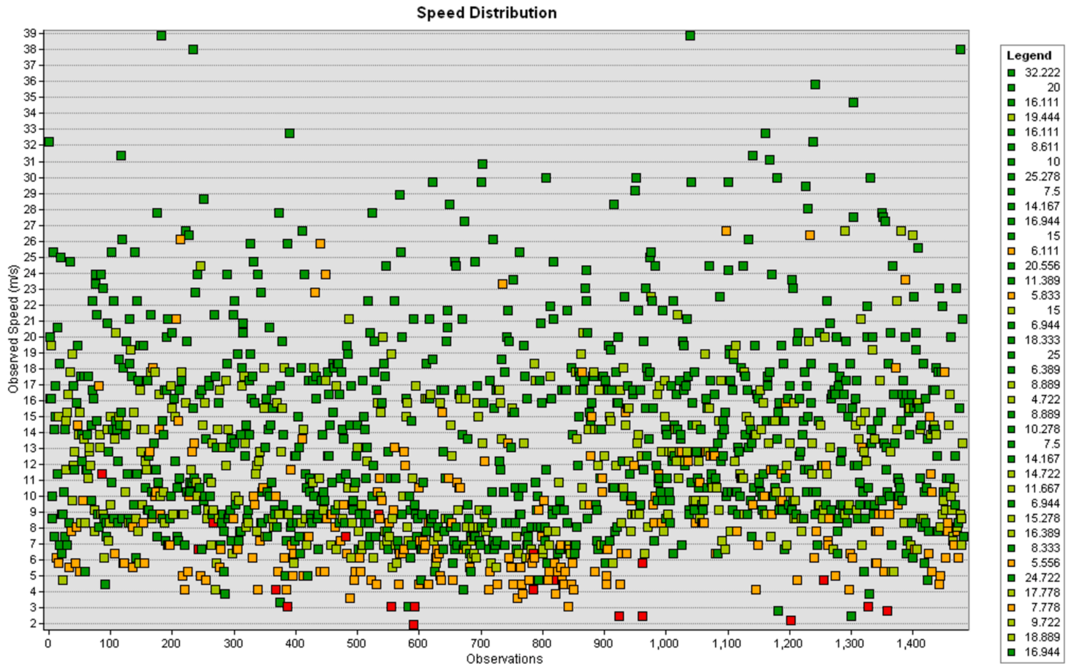

Fig. 8a. Scatterplot of Speed.

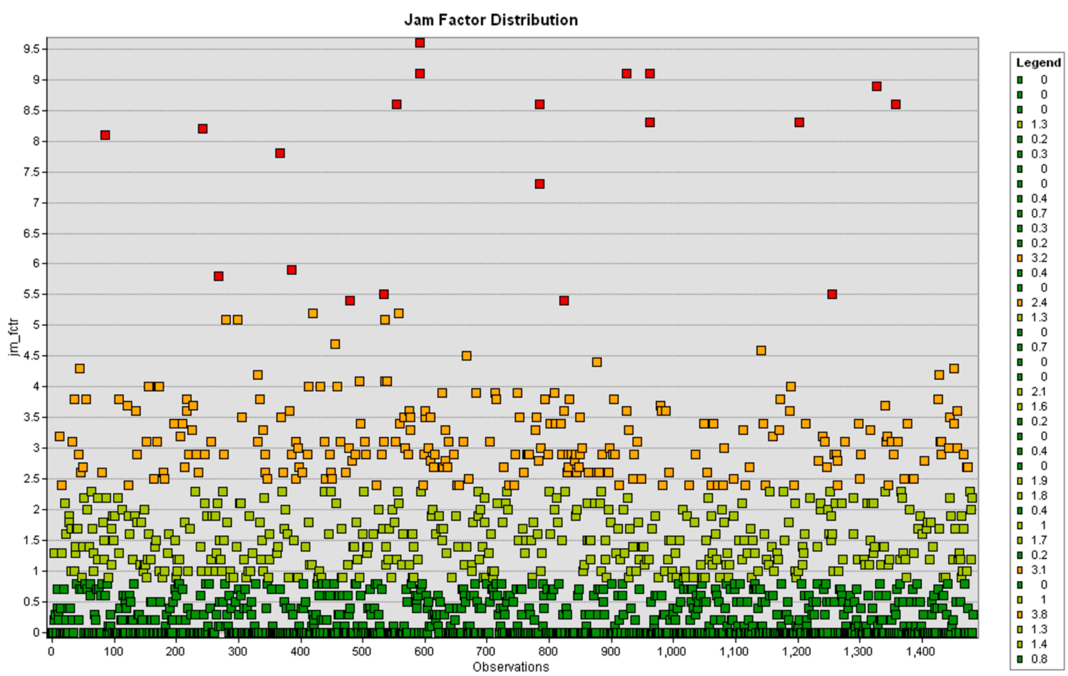

Fig. 8b. Scatterplot of Jam Factor.

environment, businesses, social interactions, and trade. For example, in Al Ain, there are four ground stations for measuring air pollution: Al-Tawia, Zakher, Al Ain Islamic Institute, and Al Ain Street. These stations measure levels of O_3, CO, $PM_{2.5}$, PM_{10}, and various other pollutants. Additionally, there are mobile pollution measurement units. This arrangement is analogous to the one used previously to establish the relationship between pollution and transport congestion elsewhere (Kumar & Goel, 2016). Similarly, geospatial data on social interactions, health, trade, and businesses may be integrated with crowdsourced data. However, a detailed analysis of these underlying relationships is beyond the scope of this study and has been reserved for future research in this field.

While crowd-sourced traffic data and related open-source tools offer various advantages for congestion modeling, the following limitations observed during this study merit attention:

- The data did not include information on traffic volumes, limiting its usability to applications that do not require traffic counts. This limitation affects its usefulness for various transportation engineering studies, such as determining peak hour traffic volumes, average daily traffic (ADT), and saturation flow rates. It's important to note that the variable of interest can vary by service provider, and this study covered only one provider (HERE). Additionally, this study explored only freely available crowd-sourced data, and based on the

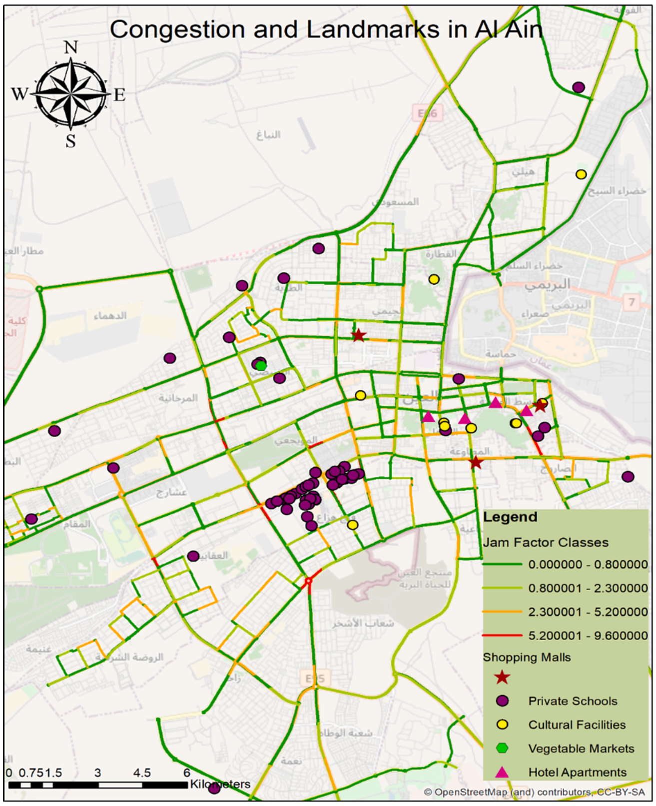

Fig. 9. Spatial Distribution of traffic congestion around key landmark clusters in Al Ain.

obtained data, it was not possible to comment on what paid crowd-sourced traffic information from the same provider could offer.

- The data did not include information regarding road infrastructure, such as road width, number of lanes, lane width, and lane configuration. This absence makes it difficult to calculate road capacity and, consequently, the level of service at desired nodes of the transport network.
- There was no information regarding traffic signal phasing plans, making it particularly difficult to analyze congestion at signalized intersections. As a result, it was not possible to comment on the intersection control type, whether it's signalized, stop sign, yield sign, or another type.
- The data does not include any information on vehicle classification, making it difficult to calculate related parameters such as load spectra for a given vehicle class and axle group type.
- Existing transportation engineering standards and guides are still based on traditional traffic counting methods. Therefore, it remains challenging to justify the use of crowd-sourced traffic data for transport studies until new guides and standards are published.

The skills required for utilizing open-source tools for congestion modeling include a general understanding of traffic flow dynamics, proficiency in working with developer tools such as APIs and scripting, and knowledge of R scripting and its coding environment. Additionally, the end user should be adept at analyzing output traffic data using spreadsheets and GIS tools. In essence, employing these open-source tools necessitates either a multidisciplinary individual or a team of professionals with expertise in these tools, particularly in R programming. This study aims to facilitate such professionals/teams by streamlining the learning curve associated with these tools.

While the outlined requirements may seem demanding, a comparison with the requisites of traditional congestion modeling methods provides a clearer perspective. Traditional methods typically involve field surveys conducted by semi-skilled personnel to count traffic and measure speeds. For instance, a standard four-legged intersection would, on average, require 12 surveyors (one each for monitoring left-turning, through, and right-turning vehicles for every approach). The number multiplies with every additional intersection if traffic data is collected on the same day (often avoided due to budgetary and workforce constraints). Therefore, many studies rely on typical weekday or weekend traffic counts, assuming that traffic patterns remain constant within these classes. Compiling and analyzing collected data typically involves the use of proprietary transport modeling tools, requiring skilled human resources and incurring commercial usage fees (although open-source alternatives are also available to supplement analysis).

In summary, utilizing crowd-sourced traffic data demands fewer but more skilled human resources with expertise in scripting and traffic engineering. Traditional methods, on the other hand, are more labor-intensive but may not require coding expertise; instead, they demand proficiency in operating transport modeling tools. It is crucial to note that traditional and crowd-sourced data-based congestion models differ not only in terms of resource requirements but also in the variable of interest. Crowd-source data offers a set of predefined variables (as observed in this study), while traditional methods can be more easily tailored to measure specific variables as needed.

To facilitate transport engineers in embracing crowdsourced-based methods, this study not only creates a congestion modeling test case but also provides a concise summary of the skills and resources required for tool deployment. A comparative analysis is conducted, juxtaposing the skill and resource requirements of creating a similar congestion model using traditional methods, thereby elucidating the advantages or disadvantages of the new tools. The study further delineates how the presented model can be utilized to comprehend the impact of traffic congestion on social interactions, air quality, sustainable transport, and businesses. In this endeavor, the study draws on existing literature that establishes correlations between transportation and these factors (Droj et al., 2021; Sciomachen & Stecca, 2023; Thomson & Bull, 2002; Tonne et al., 2008).

To assist end-users in setting realistic expectations for crowdsourced data, the study includes some of the key limitations associated with the obtained crowdsourced traffic data for congestion modeling and beyond (other transportation engineering applications). Lastly, for the sake of reproducibility, the complete R script used to obtain congestion modeling data for Al Ain city is attached as Annexure A. It is worth noting that detailed R scripts for obtaining traffic data for Windsor, Toronto, and New York are available elsewhere (Durrani, 2023; Marty, 2023).

The study's findings contribute to utilizing crowd-sourced traffic information data and open-source technologies for informing policy decisions on urban transport congestion. Through careful spatial analyses and integration of crowd-sourced traffic data with open-source tools like R, the study provides insights for professionals advising government on traffic congestion. Figs. 3, 4, and 9 depict the spatial distribution of traffic conditions in Al Ain city, offering nuanced perspectives on the capabilities of the congestion modeling tools. This information equips transport engineers with a clear understanding of how open-source tools could help model urban transport congestion and visualize its implications for daily life. The generated shapefile facilitates correlating congestion with air pollution, social interactions, and trade, serving as a practical resource for assessing the impact of traffic on urban planning. In essence, the study's findings contribute data-driven knowledge to empower professionals in making informed decisions and proposing effective strategies to address urban transport congestion challenges.

The research's academic significance lies in its methodological rigor, demonstrating the application of open-source tools, particularly R and associated packages, in retrieving and analyzing crowd-sourced traffic data. The detailed R and GitHub Action scripts provided in Appendices A and B attest to the transparency and reproducibility of the research

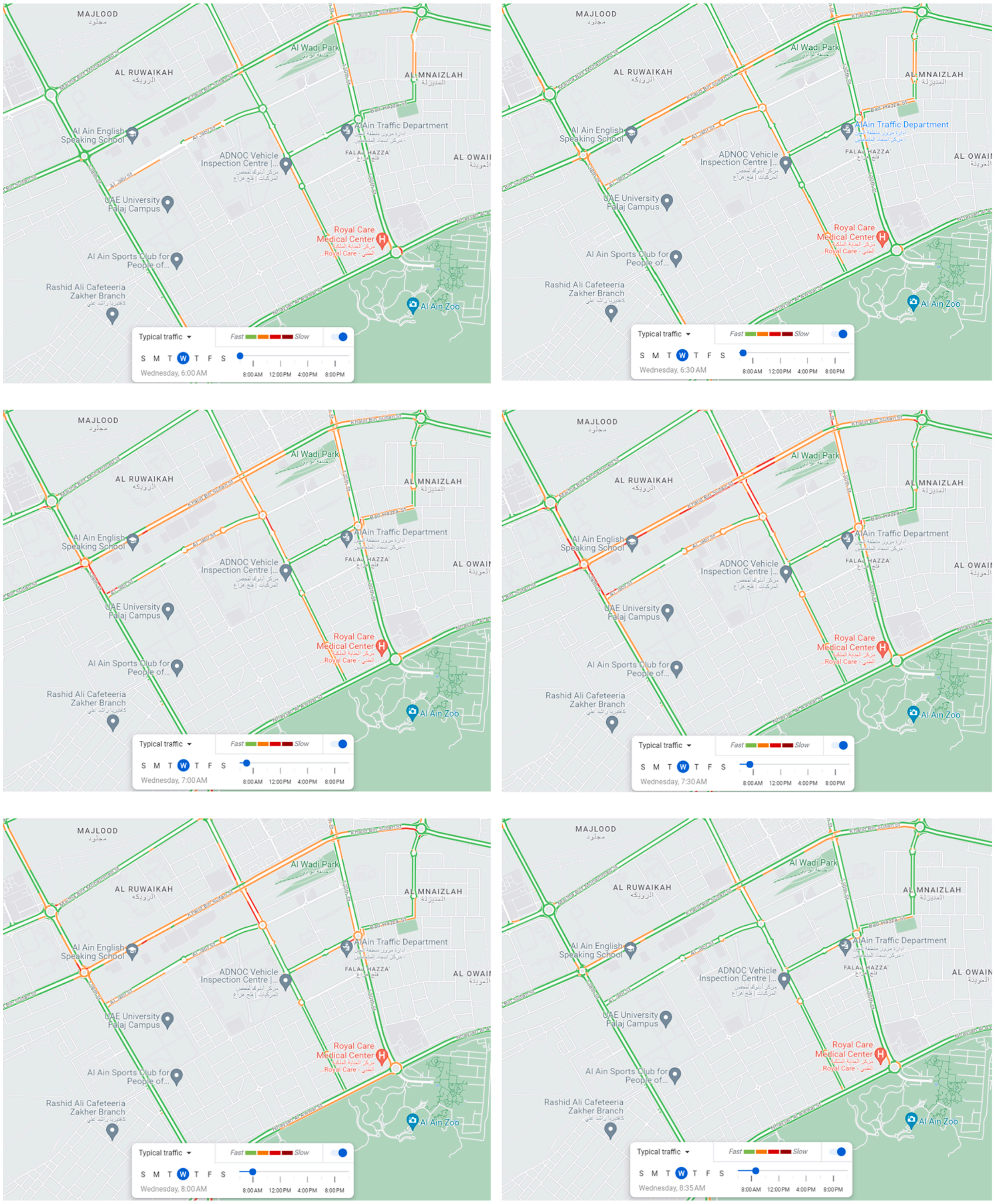

Fig. 10. Traffic congestion modelling using Google Maps Services.

methodology. This approach contributes to the scholarly discourse on the application of advanced tools in the field of transportation studies. The research transparently outlines the limitations associated with crowd-sourced traffic data. Notable limitations are the absence of information on traffic volumes, road infrastructure details, traffic signal phasing plans, and vehicle classification data (at least for the freely accessible traffic flow information). Acknowledging these constraints is crucial for advancing the application of crowd-sourced data in transportation engineering studies.

Privacy concerns surrounding crowd-sourced traffic data are paramount, especially considering the sensitive nature of transportation information and its potential implications for individual privacy. Therefore, it is essential for policymakers to thoroughly address the ethical considerations associated with the collection and utilization of crowd-sourced data. This entails discussing privacy-preserving measures and ensuring compliance with relevant regulations to uphold responsible data-handling practices. By doing so, the policymakers can contribute to fostering trust among stakeholders and mitigating potential privacy risks associated with the use of crowd-sourced traffic data.

The study is limited in its scope to exploring only the freely available data sets and is focused on the selected information providers only, while various other navigation service providers meticulously monitor and record traffic flow information. Therefore, this study lays the groundwork for future investigations, outlining avenues for more nuanced and comprehensive studies. Suggestions include exploring paid crowd-sourced data, integrating road infrastructure details, analyzing traffic signal phasing, enhancing vehicle classification data, and aligning transportation engineering standards with evolving technologies. Another limitation of the study is its focus on data collection, neglecting the perspective of technology providers who are the primary collectors of crowd-sourced traffic flow information. Future work should explore the methodologies employed by these technology companies to provide further insights into the usability of the traffic flow information discussed in this study.

CRediT authorship contribution statement

Khaula Alkaabi: Writing – review & editing, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization. **Mohsin Raza:** Writing – review & editing, Formal analysis. **Esra Qasemi:** Investigation. **Hafsah Alderei:** Investigation. **Mazoun Alderei:** Investigation. **Sharina Almheiri:** Investigation.

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

This research was funded by the UAE University's Research and Sponsored Projects Office (SURE Plus Grant # G00004434, 2023; and Strategic Research Program - Emirates Center for Mobility Research G00004542, 2024).

The authors thank HERE, R, Google, and other data providers and developers for offering adequate free access to their software and services, which enabled this study. The authors also express their gratitude to scholars, particularly Durrani (2023), for extensive blogging on the use of open-source tools.

Appendix A

R-Script to Obtain Traffic Flow Data of Al Ain Using hereR Package.

```
# Load required libraries
install.packages(c("hereR", "osmdata"))
library(hereR)
library(osmdata)
# Define the city and bounding box for Al Ain, UAE
city <- "Al Ain, UAE"
bbox <- getbb(city)
# Fetch geospatial data for Al Ain city
cdn <- osmdata::getbb(city, format_out = "sf_polygon")
# Obtain traffic flow data for Al Ain
flows_Al_Ain <- flow(aoi = cdn)
# Visualize traffic flow data on a map
library(mapview)
m <- mapview(flows_Al_Ain,
      popup = popupTable(flows_Al_Ain,
          zcol = c("free_flow", "speed", "jam_factor", "confidence")),
       zcol = "speed",
      layer.name = "speed"
)
M
```

Appendix B

GitHub Actions Worflow to Automate Traffic Data Collection for an Entire Day

To scale the workflow and obtain traffic data for Al Ain city at 15-minute intervals throughout an entire day using GitHub Actions, a workflow script needs to be created. GitHub Actions allows to automate tasks and workflows directly within your GitHub repository. Here is an example of how to set up a workflow to fetch traffic data for Al Ain city at the specified intervals:

1. Preparing the R Script

First, an R script is needed to perform the data extraction and save the data. Let us assume that the R script is named traffic_data_extraction.R and contains the necessary code to obtain traffic data for Al Ain city, similar to the code provided in Appendix A.

2. Creating a GitHub workflow File

A YAML file (e.g., workflow.yml) in the.github/workflows/ directory of a (scholar's personal) GitHub repository is created. This YAML file defines the workflow and its schedule. For example, the following workflow is triggered every 15 min (as per the cron schedule) and runs the traffic_data_extraction.R script. After extracting data, it commits the changes back to the repository.

```
name: Traffic Data Extraction
on:
  schedule:
    – cron: '*/15 * * * *' # Runs the workflow every 15 min
jobs:
  extract_data:
    runs-on: ubuntu-latest
      steps:
        – name: Checkout code
        uses: actions/checkout@v2
        – name: Set up R
        uses: r-lib/actions/setup-r@v1
        with:
          r-version: 4.3.0
        – name: Install R packages
        run: |
          R –e 'install.packages(c("hereR", "osmdata", "mapview"))'
        – name: Run R Script
        run: Rscript.github/scripts/traffic_data_extraction.R # Path to your R script
        – name: Commit and push changes
        run: |
          git config user.name "${{ github.actor }}"
          git config user.email "${{ github.actor }}@users.noreply.github.com"
          git add.
          git commit –m "Update traffic data"
          git push
```

3. Directory structure

The following directory structure needs to be ensured.

```
–.github
  – workflows
    – workflow.yml
–.github
  – scripts
    – traffic_data_extraction.R
```

Data availability

Data will be made available on request.

References

Alkaabi, K., 2023. Identification of hotspot areas for traffic accidents and analyzing drivers' behaviors and road accidents. Transp. Res. Interdiscip. Perspect. 22, 100929. https://doi.org/10.1016/j.trip.2023.100929.

Alkaabi, K., Abuelgasim, A., 2022. Comparative Analysis of Pollutant Levels during Lockdowns Across Different Land Use over the Emirate of Abu Dhabi, United Arab

Emirate. Arab World Geographer 24 (3), 205–220. https://meridian.allenpress.com/awg/article/24/3/205/484475/Comparative-Analysis-of-Pollutant-Levels-during.

Benz, R.J., Ogden, M.A., 1996. Development and benefits of computer-aided travel time data collection. Transp. Res. Rec. 1551 (1), 1–7. https://doi.org/10.1177/0361198196155100101.

Cuervo, L.G., Martinez-Herrera, E., Osorio, L., Hatcher-Roberts, J., Cuervo, D., Bula, M. O., Pinilla, L.F., Piquero, F., Jaramillo, C., 2022. Dynamic accessibility by car to tertiary care emergency services in Cali, Colombia, in 2020: Cross-sectional equity analyses using travel time big data from a Google API. BMJ Open 12 (9), e062178.

Drew, D. R. (1968). *Traffic flow theory and control* (467 pp). Article 467 pp. https://trid.trb.org/view/115219.

Droj, G., Droj, L., Badea, A., 2021. GIS-Based Survey over the Public Transport Strategy: An Instrument for Economic and Sustainable Urban Traffic Planning. ISPRS Int. J. Geo Inf. 11, 16. https://doi.org/10.3390/ijgi11010016.

Durrani, U. (2023, September 23). *Getting and plotting real-time traffic data*. Dr U. https://umairdurrani.com/posts/herer_traffic/herer_traffic.

Gerlough, D.L., Huber, M.J., 1976. Traffic flow theory. HS-006 783 Transport. Res. Board Spec. Rep. 165. https://trid.trb.org/view/37152.

HERE. (2023). *HERE Traffic API v6—Developer Guide*. https://www.here.com/docs/bundle/traffic-api-developer-guide-v6/page/topics/coverage-information.html.

Hudson, J.G., 2002. Improving the Congestion Management System (CMS) Process. Texas Transportation Institute The Texas A&M University System College Station.

Ibad, M.Z., Sulistyorini, R., Rahmah, C., 2020. Google Traffic sebagai Masukan Kebijakan Transportasi Perkotaan (Studi Kasus: Kota Bandar Lampung). TATALOKA 22 (3), 409–417. https://doi.org/10.14710/tataloka.22.3.409-417.

Kumar, P., Goel, A., 2016. Concentration dynamics of coarse and fine particulate matter at and around signalised traffic intersections. Environ. Sci. Processes Impacts 18 (9), 1220–1235. https://doi.org/10.1039/C6EM00215C.

Lovelace, R., 2021. Open source tools for geographic analysis in transport planning. J. Geogr. Syst. 23 (4), 547–578. https://doi.org/10.1007/s10109-020-00342-2.

Marty, R. (2023). *Googletraffic: Google Traffic. R package version 0.1.3* [Computer software]. https://dime-worldbank.github.io/googletraffic/.

Mattias, Y. (2023, October 10). Project Green Light's work to reduce urban emissions using AI. *Google*. https://blog.google/outreach-initiatives/sustainability/google-ai-reduce-greenhouse-emissions-project-greenlight/.

Nair, D.J., Gilles, F., Chand, S., Saxena, N., Dixit, V., 2019. Characterizing multicity urban traffic conditions using crowdsourced data. PLoS One 14 (3), e0212845.

Naiudomthum, S., Winijkul, E., Sirisubtawee, S., 2022. Near real-time spatial and temporal distribution of traffic emissions in bangkok using google maps application program interface. Atmos. 13 (11), Article 11. https://doi.org/10.3390/atmos13111803.

Noor, M.A., Ashrafi, S., Fattah, M.A., Morshed, S.R., Rahman, S., 2021. Assessment of traffic congestion scenario at the CBD areas in a developing city: In the context of Khulna City, Bangladesh. Transport. Res. Interdiscipl. Perspect. 11, 100435. https://doi.org/10.1016/j.trip.2021.100435.

Quiroga, C.A., 2000. Performance measures and data requirements for congestion management systems. Transp. Res. Part C Emerging Technol. 8 (1), 287–306. https://doi.org/10.1016/S0968-090X(00)00008-5.

Schirck-Matthews, A., Hochmair, H.H., Strelnikova, D., Juhász, L., 2022. Bicycle trips in Endomondo, Google Maps, and MapQuest: A comparison between South Florida and North Holland. Transp. Lett. 1–13. https://doi.org/10.1080/19427867.2022.2050494.

Sciomachen, A., Stecca, G., 2023. Forwarding containers to dry ports in congested logistic networks. Transp. Res. Interdiscip. Perspect. 20, 100846. https://doi.org/10.1016/j.trip.2023.100846.

Thomson, I., Bull, A., 2002. Urban traffic congestion: Its economic and social causes and consequences. CEPAL Rev. 2002 (76), 105–116. https://doi.org/10.18356/f43aec5d-en.

Tonne, C., Beevers, S., Armstrong, B., Kelly, F., Wilkinson, P., 2008. Air pollution and mortality benefits of the London Congestion Charge: Spatial and socioeconomic inequalities. Occup. Environ. Med. 65 (9), 620–627. https://doi.org/10.1136/oem.2007.036533.

Unterfinger, M. (2023). *hereR: 'sf'-Based Interface to the "HERE" REST APIs.* [Computer software]. https://munterfi.github.io/hereR/, https://github.com/munterfi/hereR/.