


ORIGINAL PAPER

Open Access



# Georeferenced X (formerly twitter) data as a proxy of mobility behaviour: case study of Norway

María Díez-Gutiérrez<sup>1\*</sup> , Sahar Babri<sup>1</sup>, Erlend Dahl<sup>1</sup> and Olav Kåre Malmin<sup>1</sup>

## Abstract

Georeferenced messages on social media represent a powerful data source to gain a different perspective for estimating mobility behaviour, which is still mainly based on travel surveys. These data are openly available, yet few studies have explored their potential. This paper assesses the feasibility of large-scale Twitter data as a proxy of human mobility behaviour to complement traditional travel surveys, and for calibration and validation of transport models. Almost 12 million Tweets from more than 90,000 users were further analysed to detect the trip patterns at municipality level in Norway from 2012 to 2022. Results showed that the mobility patterns changed between 2014 and 2019 for the travel survey, as for 2019 most of the reported trips were short and concentrated in the densely populated areas of the country, where most respondents lived, triggering a lack of information for certain areas. In contrast, Twitter data presented a more stable data source along both years with similar population distribution and average trip length. Although Twitter data have limitations in relation to the socio-demographic information of the users, it could complement the travel survey given the broader spatial and temporal distribution of this large-scale data.

**Keywords** Mobility patterns, OD trips, Social media, National travel surveys, Twitter data

## 1 Introduction

Global climate change is one of the world's greatest challenges to achieve sustainable goals, where transportation accounts for a fourth of the total CO<sub>2</sub> emissions [2], therefore transforming this industry plays a significant role. Diverse transport policies need to be assessed for their effectiveness, sustainability, and feasibility before being implemented in real-life. Estimating mobility behaviour is crucial for developing tools and simulation environments which enable decision-makers to properly assess the policies in advance to their implementation. Nevertheless, transport user behaviour is complex and could rapidly adapt to different trends, such as the pandemic [24].

Traditional methods for capturing these behaviours are travel surveys and travel diaries, which face several challenges and potential data inaccuracies which can impact the performance of transport models [38], as data might not reflect the real behaviour of the population [82]. These methods have major challenges such as low response and completion rate [79, 93], or underestimation of short trips [13, 94]. Some of the reasons might be associated with the survey duration, forgetfulness of respondents, selective omission of some trips, or incorrect understanding of trip or activity definitions [12, 38, 79, 94], as well as difficulties to reach the potential respondents and their unwillingness to participate, which could be linked to the increasing screening of marketing calls [86].

New technologies and other data sources may not only improve the quality and quantity of the data collected by traditional travel survey methods, but also increase the variety of the data, as well as provide data over a longer

\*Correspondence:

María Díez-Gutiérrez  
maria.diez@sintef.no

<sup>1</sup> Mobility Department, SINTEF Community, S.P. Andersens vei 3,  
7031 Trondheim, Norway



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

period [16, 96]. Social media platforms such as Twitter and Instagram have emerged as significant sources of location-based information due to their geo-tagging capabilities for user posts. Although this data is presently available for research, its future accessibility is uncertain due to proprietary business strategies and general data protection regulations. Additionally, the data relies on user-generated posts, which may not represent all demographic groups and can be challenging to use for identifying individual trips. Nonetheless, Twitter has become a popular dataset for researchers, being a source for behaviour analysis, opinion mining, trend tracking [92], and sentiment analysis [32]. Twitter has previously been used to analyse population density, for example estimating the population distribution throughout the day [20] or in relation to the land use [30, 47]. Luo et al. [59] related the spatiotemporal features to demographic information, and Shelton et al. [72] observed the socio-spatial inequalities within users. Twitter data was also used to complement other data sources to predict large-scale human mobility [83].

Several countries are currently encountering difficulties with their traditional travel surveys, and there has been limited research on the use of Twitter data as a complementary source of information. This paper aims to bridge this gap by exploring the potential feasibility and reliability of leveraging large-scale Twitter data to analyse human mobility patterns. Over 12 million geo-referenced Tweets from Norway, spanning from 2012 to 2022, were analysed to estimate origin–destination (OD) trips between municipalities. These estimates were then compared with data from traditional Norwegian travel surveys to assess the potential of Twitter data as a supplementary resource for traditional travel surveys, contributing with valuable insights into the integration of social media data with traditional transportation research methodologies.

A more detailed literature review is included in Sect. 2. The methodology, including data sources, processing, and analysing is described in Sect. 3. The results are gathered in Sect. 4, followed by the discussion and implications in Sect. 5. Finally, a short summary with the main conclusions is in Sect. 6.

## 2 Literature review

In the last years, the drawbacks of traditional travel surveys have become more evident, highlighting the need to seek alternative data sources for understanding travel behaviour [71]. Traditional surveys, typically based on self-reported data gathered through telephone or computer-assisted interviews, encounter numerous issues such as decreasing response rates, recall bias, and high costs [82]. Integrating other data sources with traditional

surveys can enhance the overall quality and depth of travel behaviour research, leading to better-informed transportation planning and policy decisions, but the challenge lies in digitalizing and consolidating data from multiple sources for effective exploitation. Liu et al. [52, 53] emphasised that big data must be carefully used due to challenges related to unrepresentativeness, inconsistency, and unreliability. Several big data sources could be further explored, Li et al. [50] divided in three main types depending on the generation of the data, from transactions, from devices, or from users.

Transactions data, which could be web search data, or bank transactions, is limited for travel behaviour due to privacy policies. However, electronic fare payment systems in public transportation might be used for estimating travel behaviour [5]. An assumption needed is that a single card is used by a unique person, which might not be the case, this was only tested by Chu and Chapleau [18]. Hussain et al. [43] used these data for estimating OD matrixes, although some shortcomings were stated that could be overcome with the integration of several data sources. Unlike traditional travel surveys, it is challenged to obtain sociodemographic data from transactions data, as well as reliable data on travel demand for different transport modes. Nevertheless, it could be explored the suitability for model calibration.

There are several device data that has been explored to complement traditional travel surveys, such as data generated by Global Positioning System satellites (GPS), sensors, smartphones, or mobile roaming data.

GPS-based surveys might have a potential to replace or supplement traditional methods [16], which may enable large scale surveys at lower cost [9], and provide a more flexible method to capture rapid behaviour changes [38]. GPS-positioning captured with higher precision spatial–temporal movements of travellers [13], which reduce underreporting short trips as in traditional travel diaries [64, 73]. Rasmussen et al. [69] derived with high accuracy other trip attributes such as trip purpose. Nevertheless, people may forget to carry the device and there might be signal losses due to obstructions between the device and the satellites, in undergrounds for instance. Moreover, the travel attributes depend to a larger extent on the post-processing [76], although the potential assumptions might be validated by recall surveys [54].

Other data types are automatic vehicle location, automatic passenger counting and traffic counts. Automatic vehicle location uses GPS to record the position of the vehicles in the network in real time [58]. The vehicles tracked could be from private owners, companies with freight vehicles, to public transport companies. This data is not linked to a person, however, Chapleau et al. [15] highlighted that it could complement traditional travel

surveys, especially in the context of estimating public transportation demand. Within this setting, automatic passenger counting is also a data source to be considered, this sensor technology counts the number of passengers, mainly boarding and alighting at each stop. Nonetheless, sociodemographic information is also disregarded [31], thus these data could be more useful for model calibration purposes rather than as complement to traditional travel surveys.

Traffic counts data refer to the number of vehicles, which could be length aggregated, passing a point during a specified period of time. The technology to capture these data could be sensors or video recordings. This data has several limitations when compared to travel surveys, as it does not refer to a unique person and the complete trip patterns are not known, as data is only associated to a point. Despite numerous studies have concentrated on optimizing the placement of traffic counts to estimate origin–destination (OD) patterns, such as Fu et al. [28], the topic is still open to further research. Nevertheless, its used for model calibration is widely recognised [42].

Bluetooth technology, by fixed or mobile sensors, might be used to get insights into travel behaviour, some examples are Mei et al. [61] who used these data for estimating travel times, Abedi et al. [1] that estimated space-temporal movements of cyclist and pedestrians in combination to other data sources, or Yang and Wu [97] who estimated travel mode, although presenting some limitations. The lack of personal information, the trip patterns, and the transport mode reduce its used for travel demand estimations or in combination to travel surveys, being more significant for specific model calibrations.

Today's technology allows collecting travel data using smartphones. A rich data set can be derived and computed from multiple built-in sensors, such as motion sensors (accelerometer, gravity sensor and gyroscope), environmental sensors (barometer, photometer and thermometer) and position sensors (GPS and magnetometer) [6]. The smartphone applications may be divided in two main types, active or passive. The former requires that the user interacts, i.e. to select the start and end of the trips, as well as the mode, and potentially the purpose. A limitation is that users forget to activate and deactivate the application before and after the trip [96], as well as being more time consuming. The passive application does not require interaction, as it runs in the background of the phone. A set of algorithms automatically detect trips and modes [9, 16]. Ferrer and Ruiz [26] detected travel modes by using raw accelerometer data, with over 89% match for all modes. Alexander et al. [3, 4] showed representative results for daily origin–destination matrices by purpose. An advantage is that it is possible to overcome loss of GPS-signal in urban or indoor areas with use of

accelerometer sensor or connection to Wi-Fi access points [26, 51], but as a result of an increased sensor usage, high battery power consumption is a disadvantage [51, 96]. Discussions at different countries are being held to assess the potential of replacing the data collection of traditional travel surveys by smartphone applications, however there are concerns related to lack of standardization and reproducibility [7]. Additionally, passive tracking of people's behaviour also introduces privacy concerns that may set restrictions for the survey design. There is also scepticism among certain type of people to participate which might lead to bias representativity [81].

Another example of data source is cellular network signalling which might provide more information in terms of sample variety and duration. By using CDR-data there is no need for users to do anything, thus it is the most battery efficient method. It is based on cell tower triangulation from call detail records (CDR) from any telephone type [34]. Previous research showed the possibility to identify movements (origin–destination) [27, 85], transport modes with a precision between 80 and 97% [96], and activities [10] to serve directly into transport models. Bachir et al. [8] estimated travel mode and OD trips, which were validated by traditional travel surveys. On the other hand, Šulíková et al. [80] explored this data source to complement data from the Slovakian traditional travel survey for transport modelling purposes, however several challenges disregard this option. Individual trips cannot be tracked according to the European General Data Protection Regulation (GDPR), in addition at national level some countries might have more strict rules, in Norway each mobile ID are renamed every day, making very difficult to detect work or residential locations [23]. Moreover, the location precision depends on the tower density, being less suitable for rural areas [51].

In relation to user generated data, this could be online photo data, or online textual data. Peoples' movements can be extracted from their photo post on Flickr [11], this platform is especially interesting for tourist behaviour [95], including visiting places, crowded areas, or trajectories [22, 40, 55, 99]. Instagram was also used to identify the most visited places [35], however this data is not openly available longer which reduces their research interest. Similarly, Panoramio was scarcely explored when it was active [44].

Georeferenced Tweets represent a powerful and high-quality data source to gain a new perspective for estimating mobility patterns [36], being also valuable for continuous monitoring and trend detections [98]. These could be place-referenced or coordinate-referenced Tweets, the former represents different levels at municipality, city or town, or neighbourhood, whilst the latter could have a precision down to 5 m under open sky

[89]. The granularity of the Tweets allows to observe not only residence and work locations but also visited places or specific routes [47]. Lenormand et al. [49] compared the spatiotemporal distribution of people and individual mobility patterns using data from Twitter, cell phones, and census and concluded that the three data sources are feasibly interchangeable. Some work by Kurkcu et al. [46] and Lee et al. [48] comparing the mobility patterns to travel surveys also confirmed their similarities. Some of the previous research on mobility patterns from Twitter data estimated: origin–destination mobility flows [29], Jiajun [52, 53], next position in human trajectories [21], traffic events [68], preferred visited places [45], tourist flows [19], mobility patterns and dynamics in retail locations [57], mobility patterns between residence locations and public spaces in a medium-size city [70], differences of mobility patterns between visitors and residents [56], commuting patterns [60, 65], mobility dynamics before and after the pandemic [41, 74, 98].

Recent literature highlights the potential of georeferenced Tweets for analysing travel behaviour, yet few studies have investigated this data as a complement to traditional travel surveys. This paper aims to address this gap by exploring the feasibility of integrating user-generated data with conventional travel surveys to enhance the estimation of travel behaviour and improve the development of transport models.

### 3 Data

The geographical focus of this research was Norway. Although Twitter data is globally accessible, data pertaining to national travel surveys is restricted due to general data protection regulations. For this study, data from the Norwegian Travel Survey was made available. In this section, the two datasets used in this research are further described, namely, the National Travel Survey is Sect. 3.1, and the Twitter data in Sect. 3.2, including the data collection and cleaning processing.

#### 3.1 National travel survey

The transport pattern data in Norway is mainly collected through computer assisted telephone interviews (CATI). The survey is distributed among a representative population sample in terms of sociodemographic features every 4 years. Since 1985, the response rate has dropped from 77 to 20% [37]. Wilson [93] found similar decreasing response rates in the traditional data collection methods in other national household surveys. The total number of respondents for the national travel survey for 2014 and 2019 were 61,314 and 88,548, respectively Hjorthol et al. [37], Grue et al. [33]. The next Norwegian national travel survey will probably be performed on CATI and computer-assisted web interviews (CAWI), introducing

new challenges. This makes it more important that participants understand the definitions by written explanations, as there is no interaction with the interviewer. Despite that, Christiansen et al. [17] found an increase in the short trips reported with this method. However, the low response rate, amongst other limitations, may not be overcome using this method.

The information collected through the travel survey is divided into 8 sections: (1) residence location, (2) access to different transport modes, (3) job/study information, (4) short trips, (5) long trips, (6) commuting trips, (7) family structure and home options regarding parking and public transport availability, and (8) sociodemographic information.

Each respondent must describe all the short trips (less than 100 km) performed the day before to the interview, including origin, destination, purpose, transport mode, number of people travelling together, access to car, and public transport card. In addition, the frequency of the weekly use of the different transport modes for the season. For long trips (over 100 km or to and from out of the country), each respondent states the number of these trips for the last 30 days. The most recent long trip is described with more details, including day of the week, purpose, transport mode, origin, destination, number of people travelling together, number of days overnighing, type of accommodation, payer of the trip, frequency of long trips due to work, and some characteristics of these trips [33, 37].

#### 3.2 Twitter data

The Twitter streaming Application Programming Interface (API) [87] was used to collect all georeferenced Tweets posted in Norway from 2012 (the earliest available Tweets in the API) to 2022. This period was selected in order to obtain information from several days to compensate for the potential spatial sparsity of the sample [39].

The total dataset from January 2012 to December 2022 consisted of 12,727,651 Tweets generated by 224,096 unique users. The characteristics of the extracted data are as shown in Table 1.

Nevertheless, some of the Tweets did not represent actual people using this social media properly. As already identified in the literature, repeated Tweets might be spam [36, 91]. The number of Tweets per user could also be an indication of potential fake accounts. Lansley and Longley [47] considered as a maximum 3000 Tweets in 1 year, whilst 1000 Tweets in 2 years was set as limit in Osorio-Arjona and García-Palomares [65]. Simultaneously, a minimum number of Tweets was also considered in previous studies, although the limits vary into a great extent, 2,5 Tweets per day [70] or 5 Tweets during

**Table 1** Characteristics of extracted data from the Twitter streaming API

User	ID
	Username
	Residence location
	Biographical description
	Public metrics (followers, following, Tweets count)
Tweet	ID
	Language
	Text
	Creation time
Spacio-temporal info	ID (city level)
	Bounding box coordinates
	Exact coordinates

a period of 2 years [65]. In relation to spatial information, users that did not move [65] or Tweets with uncertain coordinates [47] were also removed from previous studies focusing on identifying mobility patterns.

The data cleaning process for our dataset is summarized in Table 2. After the process, users who still had more than 1500 Tweets per year (less than 0.1% of the total) were closely observed to identify potential fake accounts which could trigger unrealistic mobility patterns, however, the accounts were real users.

Each tweet has a limited number of characters. Thus, users might need to post multiple consecutive Tweets to express their thoughts. The final dataset consisted of 92,785 users.

#### 4 Methodology

This section describes the processing of the Twitter data, including the estimation of the location of the tweets at municipality level (Sect. 4.1), the trip patterns (Sect. 4.2), and the demographic information of the Twitter users (Sect. 4.3). Additionally, the residence location of the Twitter users is validated against the data from the National Travel Survey (Sect. 4.4).

#### 4.1 User location

Coordinate-referenced Tweets provide the location of the user at the time the tweet is posted. There are two types of locations on Tweets: exact coordinates, or bounding box coordinates, i.e., the tweet was posted within the borders of a polygon area. Both types are represented in Fig. 1. Tweets with exact coordinates were around 60% until the change in the Twitter's policy on sharing spatial information in 2015 [14], then around 10% after that.

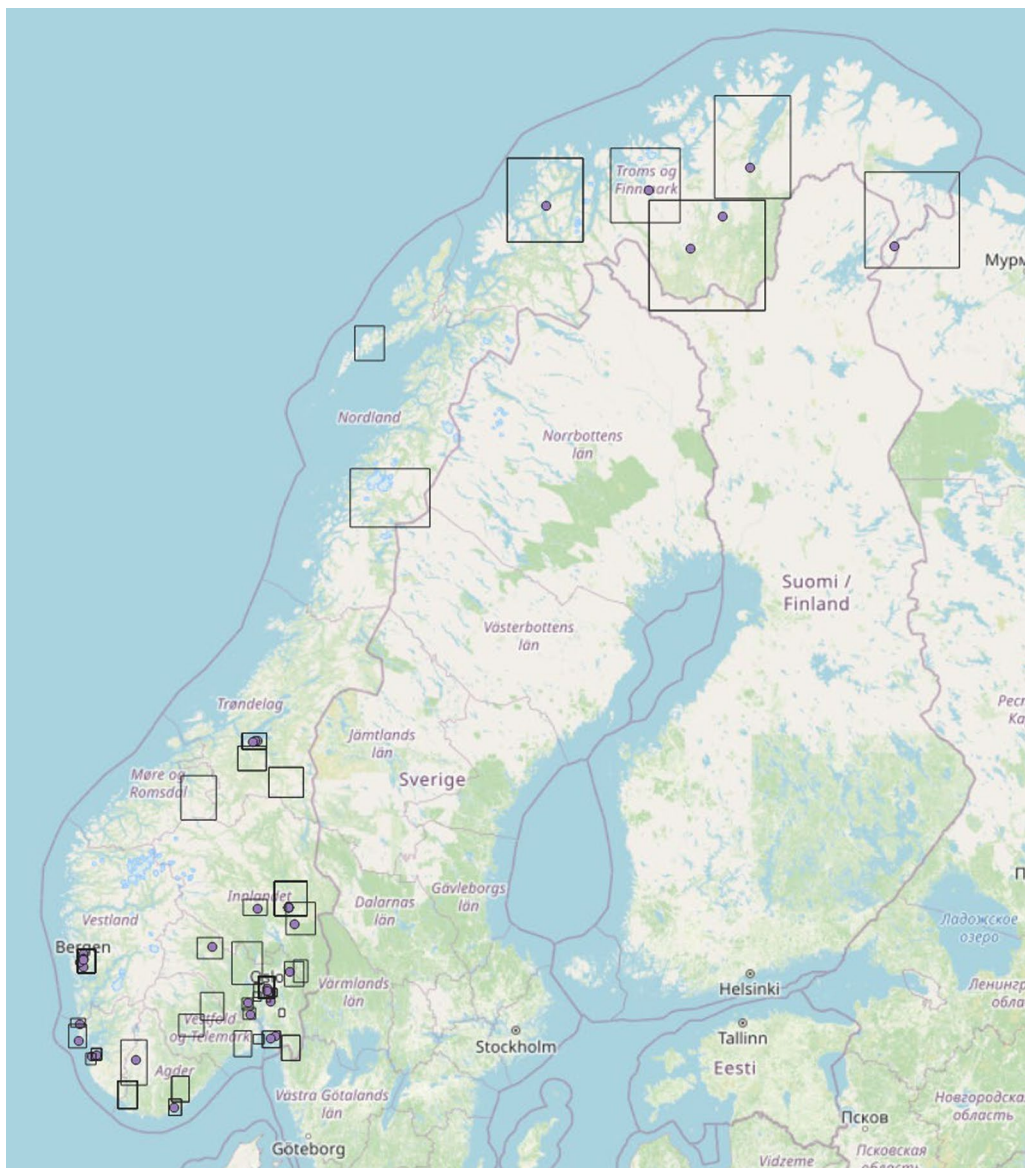
Tweets' locations were assigned to the different municipality borders for further estimation of the mobility patterns. Tweets with exact coordinates were mapped to the municipality that contained them. In the bounding box cases, each nearby municipality (m) was given a match score for a given bounding box (bb) ( $\text{score} = \text{intersection}(\text{bb}, \text{m}) / \text{union}(\text{bb}, \text{m})$ ), then the municipality with the highest score was picked. Using this score instead of a simple overlap test solved some challenges for concave municipality shapes and even for municipalities that surrounds others.

A random sample was taken to verify the matching between the bounding box and the municipality. This was possible as some Tweets were also place-referenced, which could be a town or a city, within a municipality. Tweets with an accuracy of less than 0.3 out of 1 were disregarded for the estimation of the mobility patterns, corresponding to 1.5% of the sample.

There are several studies applying similar but slightly different methods to detect the origin and destination of the trips. Some studies use the frequency counts to identify home or work location of the user, e.g. the most frequent tweet location as 'home' and second most frequent location as 'work' [60] or a combination of frequency and temporal (day and night) filtering [4, 67]. In this study, we defined the night as between 21:00 and 07:00, and identified residence location at municipality level as the location of most of the Tweets during the nights and weekdays i.e. between 21:00 and 07:00 from Monday to Thursday. This was estimated by each year, as some users might have changed their residence. Only users who did

**Table 2** Data cleaning process in relation to number of Tweets and users (2012–2022)

Data cleaning stage	Number of Tweets	Number of users
0. Initial database	12.727.651	224.096
1. Tweets without coordinates	12.704.759	224.096
2. Tweets with uncertain coordinates	12.503.686	224.096
3. Manually identified bots	12.235.661	224.088
4. Users with less than 5 Tweets	12.006.844	93.286
5. Users posting always identical messages	12.006.457	93.271
6. Users posting always from the same location	11.943.537	92.785



**Fig. 1** Example of georeferenced Tweets, exact and bounding box coordinate (background map source: OpenStreetMap)

at least one trip in the observed year and posted Tweets during the night period were considered. The residence location was estimated for 84% and 76% of the users for the years 2014 and 2019, respectively.

#### 4.2 Trip patterns

In the Twitter data, there is no explicit information about trip patterns, therefore some assumptions were taken.

Using a well-known trip-extraction procedure [67, 84], two subsequent Tweets from the same user were considered a trip if they were posted from two different municipalities and were within a given time limit. A person might have started the real trip at one municipality but

not posted a tweet until passing another, which would give a bias starting point, the same could happen at the destination, however, the likelihood of this was assumed to be low. In this study, the time limit was set to 12 h to allow the long trips that are possible within Norway. Terroso-Saenz et al. [84] assumed this limit to be 24 h in Spain.

The average number of trips per user per day was estimated as the total number of trips per studied period by the total number of users travelling per day in such period. Only Twitter users that travelled were included in the estimation, i.e. non-trips users were disregarded. The average trip distance was estimated based on the

distance between the centroids of the origin and destination municipalities per trip.

### 4.3 Demographic information

In addition, the biography description of each user was analysed to identify the gender of the user. Although, there are methods in the literature utilizing deep neural networks or traditional machine learning to identify demographic aspects of Twitter users, these methods require semantic analysis of the tweet texts [75, 90], which is beyond the scope of this paper. Using a word detection method in the bio description of the users and a manual quality check, the gender of about six percent of the users was identified (within this share 56% were men and 44% were women). Yet, the lack of more detailed demographic information is a challenge to ensure the representativity. Nevertheless, the use of Twitter in Norway is spread among the age groups: 41% 18–29 years old, 31% 30–39 years old, 31% 40–49 years old, 27% 50–59 years old, and 12% more than 60 years old, with about 1.1 million Twitter users in 2021 [78].

### 4.4 Validation

To validate the sample, the residence location of the users to the population census and to the stated residence location in the travel survey for the years 2014 and 2019 was compared. This method was previously recognised in some studies [66, 70], whilst other studies compare the density distribution of Tweets to population to assess their validity [25], Jiajun [52, 53].

To make the population distribution comparable, the respondents or the users in each municipality were

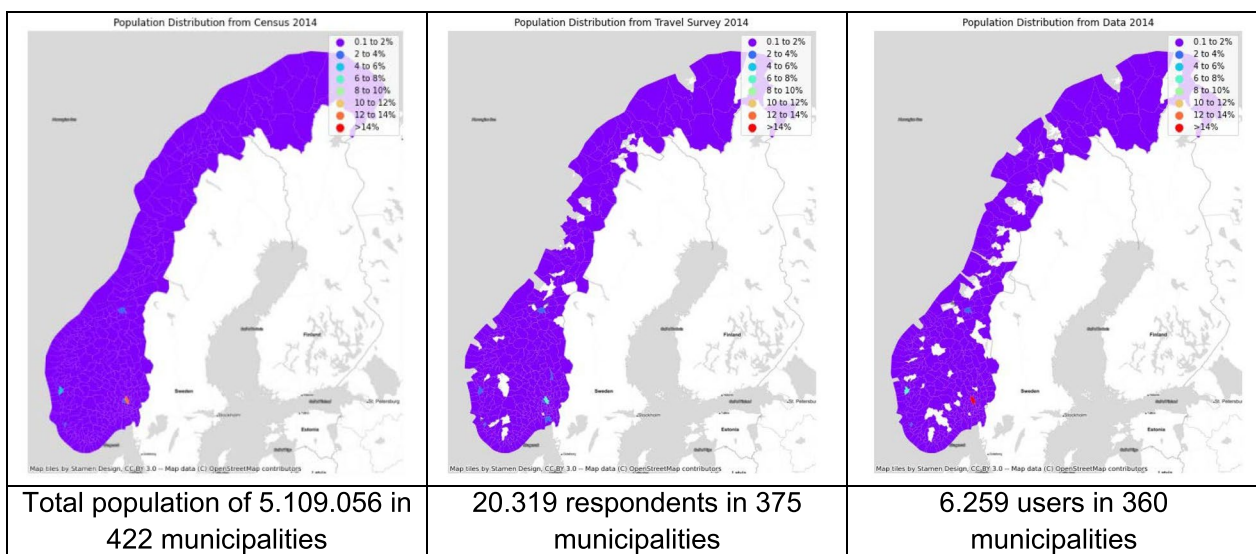
divided by the total number of respondents or by the total number of users. Figures 2 and 3 represent these distributions, as well as the sample number and the number of municipalities included.

Validation estimations show that the residence location distribution of the users resembles the population distribution from the census, data for 2014 covered 360 out of 422 municipalities, where more than 97% of the population lives. In 2019, there were less Twitter data, probably due to the sharing information policy [88], but still the estimated residence locations included 301 municipalities, where more than 92% of the census population lives, covering more territory than the national travel survey.

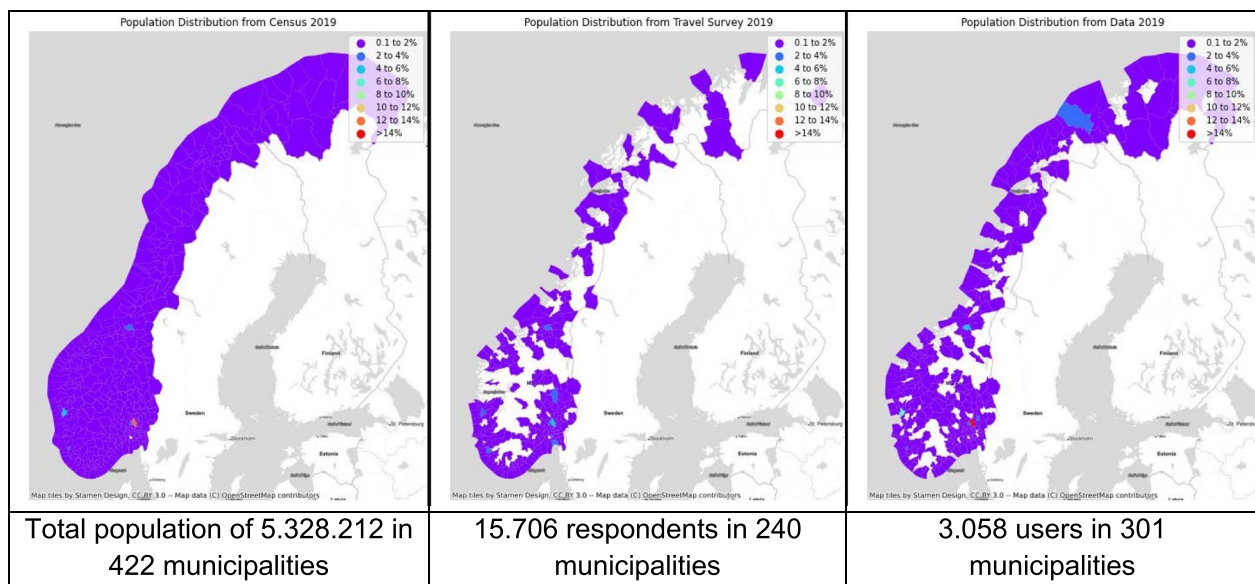
## 5 Results

In this paper the origin–destination (OD) trips between municipalities from the travel survey and from the Twitter data for the years 2014 and 2019 were estimated to assess if social media data could complement traditional data collection methods. It is important to acknowledge the disparity in trip definitions between the two data sources, rendering a direct comparison inappropriate.

Trips from the travel survey were already reported in the data by respondents, both short and long trips between different municipalities were considered. Trips from the Twitter data were estimated as previously described in Sect. 4.1. The number of OD trips between municipalities, the number of persons performing these trips, and the average distance of the trips, are described in Table 3. Figures 4 and 5 visually represent the OD trips and the distance distribution with respect to the number of trips for 2014 and 2019 respectively.



**Fig. 2** Population distribution in 2014 of **a** census, **b** Travel survey, **c** Twitter (background maps source: OpenStreetMap)



**Fig. 3** Population distribution in 2019 of **a** census, **b** Travel survey, **c** Twitter (background maps source: OpenStreetMap)

**Table 3** Number of persons, OD trips and average distance for both datasets

	2014		2019	
	Travel survey	Twitter	Travel survey	Twitter
Number of respondents/users	20,317	7428	15,721	4003
Number of OD trips	56,079	45,725	32,898	23,947
Average distance (km)	121	83	53	84

In 2014, almost 7500 Twitter users did more than 45,000 OD trips, with an average trip length of 83 km. In the travel survey, the reported trips were not evenly distributed along the year, resulting in less information for some months. Even if not directly comparison should be made, the mobility behaviour between the large cities was similar, although there was a slight underrepresentation of trips between the capital and municipalities 50–300 km towards the west in the Twitter data.

In 2019, more than 4000 Twitter users performed more than 23,000 OD trips, with an average trip length of 84 km. Even if less trips were represented, compared to 2014, the mobility patterns remained similar. On the contrary, the mobility behaviour reported in the travel survey was different between 2019 and 2014 as most of the trips were shorter, i.e. the average distance was less than 60 km and concentrated in the densely populated areas of the country, where most respondents lived.

Twitter data was further explored for the whole dataset 2012–2022. In Fig. 6 the average number of OD trips between municipalities per user per day are displayed. The average was around 1.5 trips per user per day, the

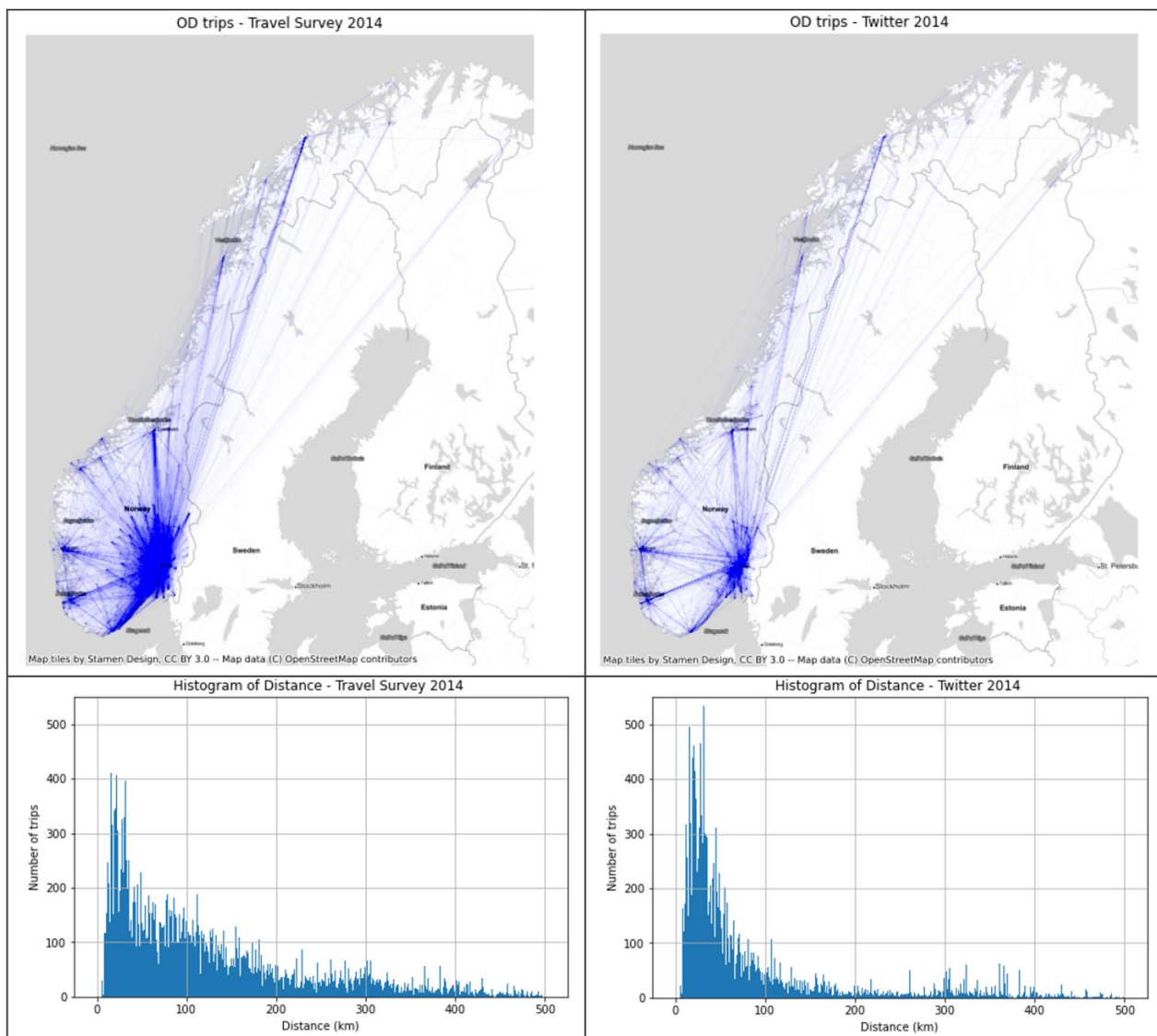
yearly variations within the studied period were lower than 5%, presenting a steady data source, although unable to correlate this metric to external trends.

In Fig. 6 the yearly distribution of the average trip distance is also shown, in this case there were three trend changes, which could be related to four time periods. (1) Prior to 2013, the trip distance remained relatively constant. However, due to the limited data spanning only two years, statistical significance could not be determined.

(2) Between 2013 and 2017, there was a notable upward trend in the average trip distance, rising from 76 to 109 km, indicating a growth of over 40%. This increase could potentially be attributed to a shift in trip destinations to municipalities located farther away. (3) Conversely, from 2017 to 2020, an opposite trend emerged, with a decline observed in average trip distances during these years. (4) Starting from 2020, the trend of decreasing trip distances persisted, with the lowest values along the period, although the decreasing trend had a less pronounced decline compared to previous periods.

In Fig. 7 the monthly distribution is represented for the studied periods. In general, the average trip





**Fig. 4** OD trips and histogram of distance for Travel Survey and Twitter data for 2014 (background maps source: OpenStreetMap)

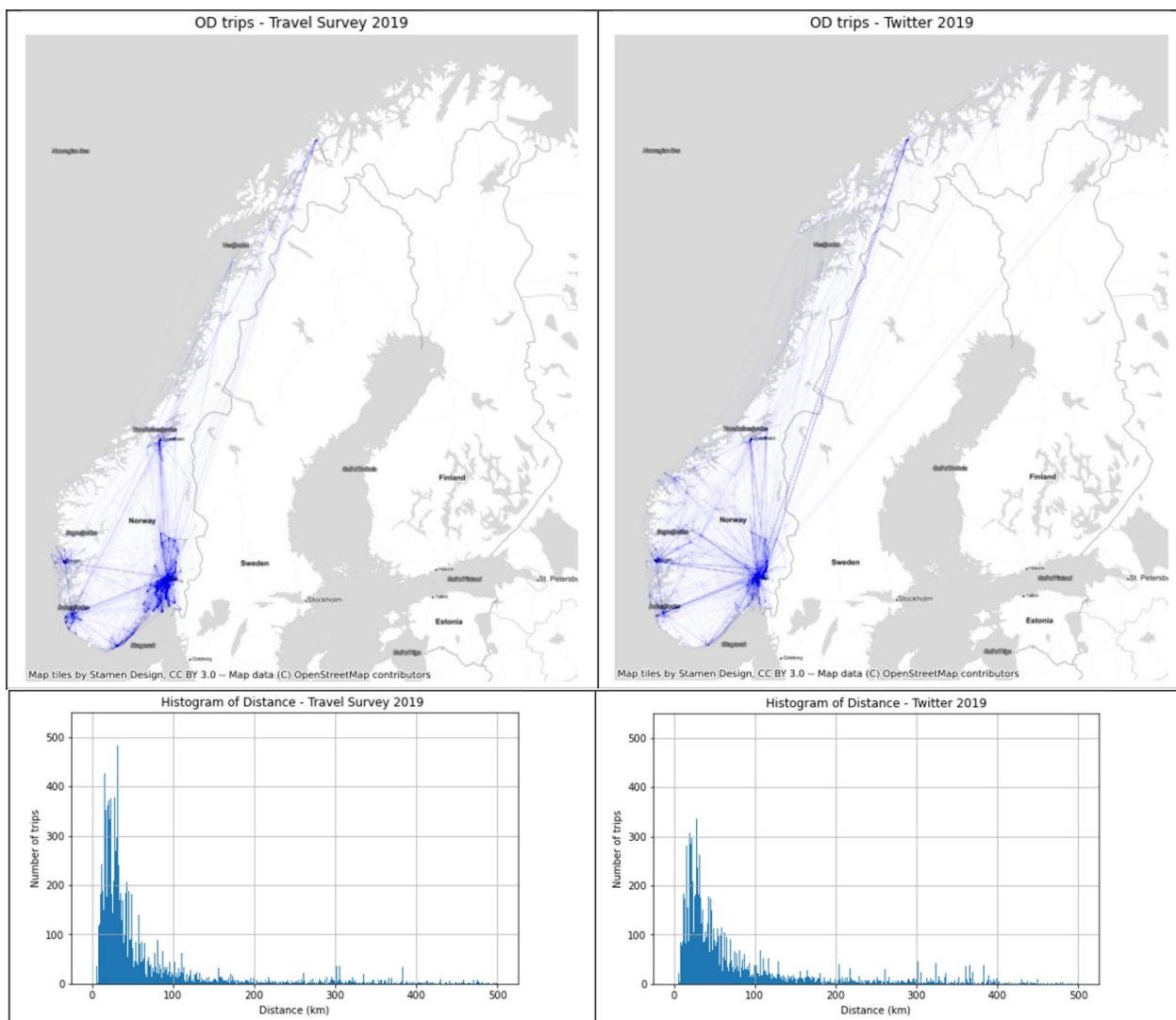
distance is slightly larger in the winter (January–March) and summer (July–September) periods, which could be associated to vacation periods and more trips to cabin areas.

The temporal data distribution allowed to detect a significant trend change in September 2017, associated to a significant increase in the average trip distance, although the number of OD trips and unique users for the same months in previous years were similar, when taking the spatial distribution into account, several origins and destinations were more popular in Nordland municipality, shown in Fig. 8 with yellow borders, where the Lofoten islands are situated among other touristic areas.

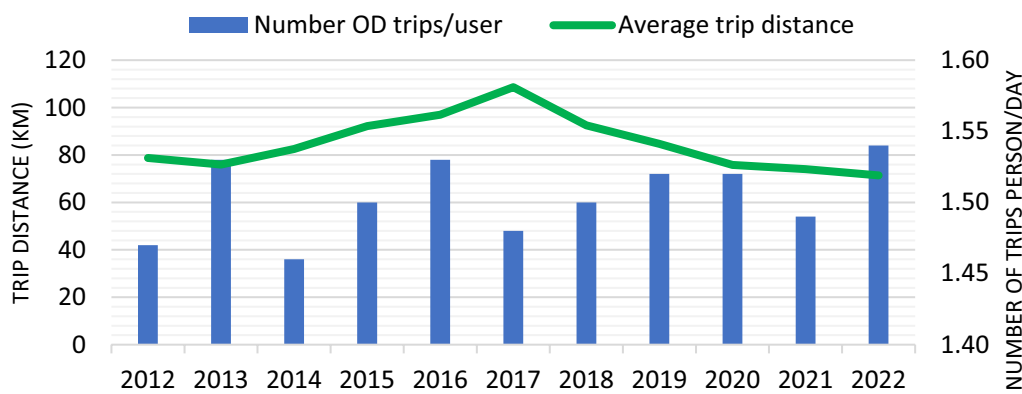
## 6 Discussion

Although this study only used Tweets from Norway, the data is available worldwide, and could be relevant for any other country or region. The number of georeferenced Tweets in Norway for a ten year period was only six times more than the Tweets in Australia for 1 week [63], and less than double than for 9 months in Spain [84], which is among the 20 leading countries based on Twitter users, still with ten and seven time less users than United States and Japan, respectively [77]. This emphasises the relevance for this data source in other countries.

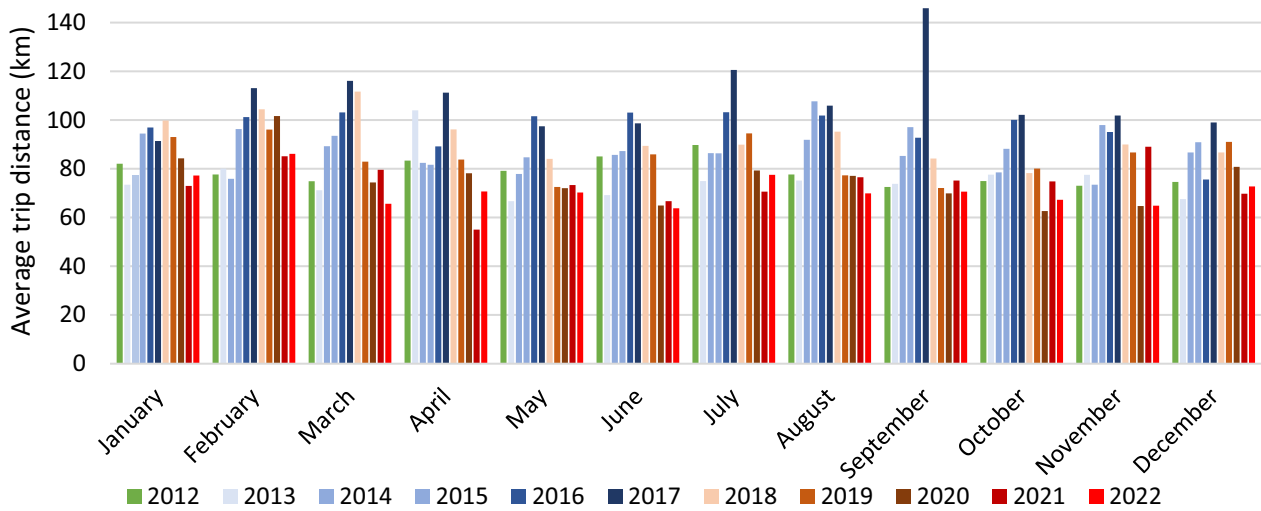
A potential limitation of this study is the number of georeferenced Tweets, as an average of 1.5 trips between municipalities per user and per day, might be



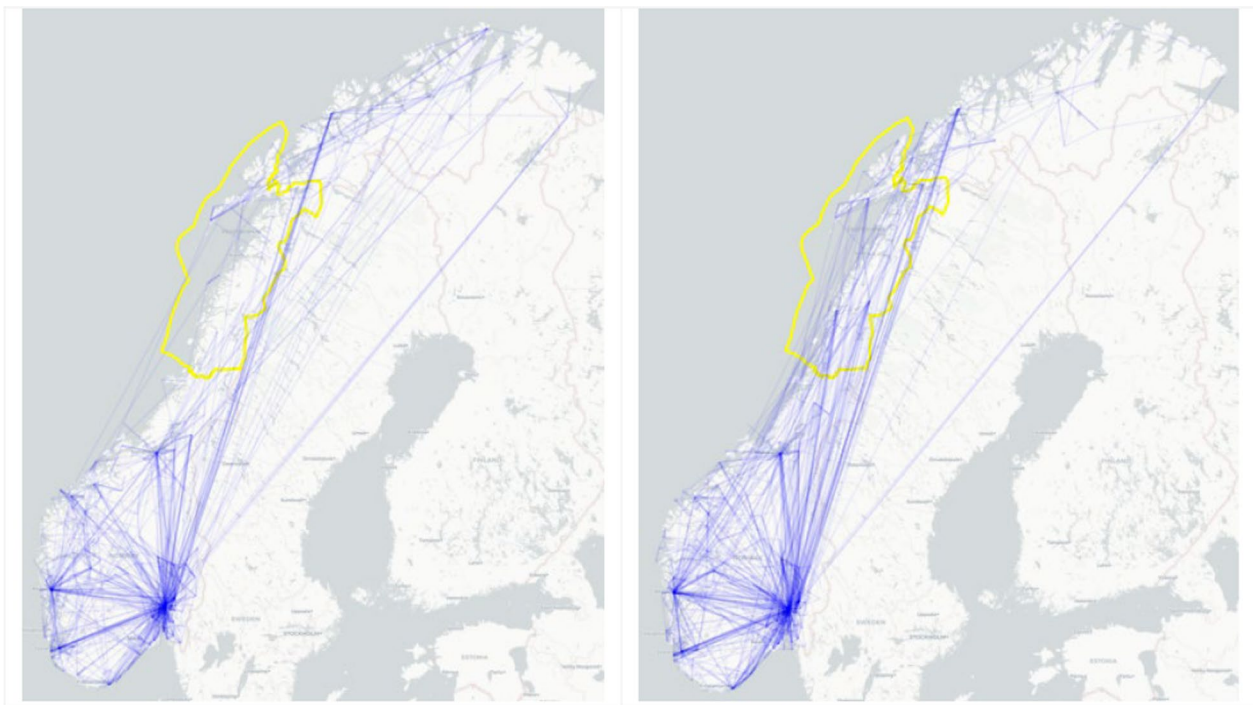
**Fig. 5** OD trips and histogram of distance for Travel Survey and Twitter data for 2019 (background maps source: OpenStreetMap)



**Fig. 6** Average trip distance and average number of OD trips per user per day (2012–2022)



**Fig. 7** Number of OD trips per user and average trip distance (2012–2022)



**Fig. 8** OD trips for September 2015 (left) and September 2017 (right) (background maps source: OpenStreetMap)

low compared to other data sources, which may challenge the trend detection. Despite that, a slight reduction of the number of trips and the average trip length was detected from 2020, when the pandemic restrictions started. Zhong et al. [98] also detected that in London users were making fewer trips, although these were longer. In the United States, Twitter data was also

used as a source to detect mobility changes, although different trends were found between states [41].

The movements in this study were limited to movements between municipalities, this aggregation might limit the full exploitation of the data, although few tweets were georeferenced with exact coordinates. Recent research is investigating how to estimate coordinates of

Tweets without geographically identified data, which could expand the data sample at any location [62], as well as allowing more detailed analyses. Thus, further work towards a finer spatial distribution is desired, which could also improve the detection of the trip purpose. Commuting trips could also be further explored as in McNeill et al. [60] and Osorio-Arjona and García-Palomares [65], given that the residence location was estimated, however several commuting trips are within the same municipality.

Twitter data presented a more stable data source than the national travel surveys along both years with similar population distribution and average trip length. Some previous work comparing the mobility patterns to travel surveys confirmed their similitudes, for New York city [46], and for California, where similarities in spatial distributions and trip lengths were also detected [48]. However, the latent mobility behaviour could not be captured in Spain [84]. Nevertheless, the behaviour could not be associated to different sociodemographic groups of population. Simple analyses were used to detect the user gender, resulting in a similar share than the travel survey, although further research should focus on expanding both the estimated sample and to other features, such as age.

The Norwegian travel survey is conducted quadrennially, covering varying time periods in each cycle. The 2014 survey collected data from January to October, with limited responses post-August, whereas the 2019 survey encompassed the entire year. Additionally, respondents only report trips from the previous day, resulting in the absence of panel data. Consequently, capturing and analysing longitudinal trend changes becomes challenging. National travel surveys reflect the mobility patterns of the population residing in the country, i.e., mobility pattern of the non-residents is never included. As a result, in some areas the real traffic volumes generated by the people movements differ far from those reflected in these surveys, especially in tourist areas. Social media data could potentially improve the representation of the non-resident's mobility as well as serving a complementary role for national travel surveys for residents' travel behaviour. It could also aid in uncovering trends through sentiment analysis of the post contents.

## 7 Conclusion

This paper contributes with the assessment of the feasibility of integrating user-generated data with conventional travel surveys to enhance the estimation of travel behaviour and improve the development of transport models. Twitter data presented a broad and stable spatial and temporal distribution of users' movements despite having limitations in relation to the socio-demographic

information of the users, compared to the travel survey. In addition, the availability of these data in real time could serve as a tool to detect trend changes, as consequence of diverse policies or other events at micro or macro level, such as recessions, pandemics, or wars.

Further work should concentrate on reducing the spatial location of the data from municipality level to spatial units corresponding to the transport models, as well as on improving the detection of the socio-economic characteristics of the users. These will ensure representativity and provide more detailed information towards a potential data fusion.

Integrating user-generated data with traditional travel surveys has significant policy implications. This approach enhances data accuracy and granularity, providing policy-makers with more precise insights into travel behaviour, and rapid detection of trend changes. This integration aids in further development of transport models to evaluate policy impacts and targeted interventions and might be a cost-effective alternative to complement traditional surveys, allowing for more frequent and updated data collection.

### Acknowledgements

Not applicable.

### Author contributions

María: conceptualization, methodology, formal analysis, validation, writing. Sahar: conceptualization, methodology, formal analysis, writing. Erlend: formal analysis. Olav Kåre: conceptualization, methodology, supervision.

### Funding

Open access funding provided by SINTEF. This research was partially financed by the Norwegian Public Road Administration.

### Availability of data and materials

The dataset supporting the conclusions of this article are not public available due to confidential agreements with the data source owners, the Norwegian Public Road Administration (travel survey data), and Twitter (social media data).

### Declarations

### Competing interests

Not applicable.

Received: 11 April 2024 Accepted: 3 September 2024

Published online: 11 September 2024

### References

1. Abedi, N., Bhaskar, A., Chung, E., & Miska, M. (2015). Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses. *Transportation Research Part C: Emerging Technologies*, 60, 124–141. <https://doi.org/10.1016/j.trc.2015.08.010>
2. Akimoto, K. (2023). Assessment of road transportation measures for global net-zero emissions considering comprehensive energy systems. *IATSS Research*, 47, 196–203. <https://doi.org/10.1016/j.iatssr.2023.02.005>

3. Alexander, L., Jiang, S., Murga, M., & Gonz, M. C. (2015). Validation of origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C*. <https://doi.org/10.1016/j.trc.2015.02.018>
4. Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58, 240–250. <https://doi.org/10.1016/j.trc.2015.02.018>
5. Ali, A., Kim, J., & Lee, S. (2016). Travel behavior analysis using smart card data. *KSCÉ Journal of Civil Engineering*, 20, 1532–1539. <https://doi.org/10.1007/s12205-015-1694-0>
6. Android, C.C.A., 2016. Sensors overview [WWW Document].
7. Azoulay, B., & Patterson, Z. (2024). Towards the standardization of reporting in smartphone travel surveys: The development and application of the Smartphone Survey Reporting Guidelines (SSRGs). *Transportation Research Procedia*, 76, 574–585. <https://doi.org/10.1016/j.trpro.2023.12.078>
8. Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M., & Puchinger, J. (2019). Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 101, 254–275. <https://doi.org/10.1016/j.trc.2019.02.013>
9. Barbeau, S. J., Labrador, M. A., Georggi, N. L., Winters, P. L., & Perez, R. A. (2009). TRAC-IT: Software architecture supporting simultaneous travel behavior data collection and real-time location-based services for GPS-enabled mobile phones. In *Transportation research board 88th annual meeting* (Vol. 21).
10. Bassolas, A., Ramasco, J. J., Herranz, R., & Cantú-Ros, O. G. (2019). Mobile phone records to feed activity-based travel demand models: MATSim for studying a cordon toll policy in Barcelona. *Transportation Research Part A: Policy and Practice*, 121, 56–74. <https://doi.org/10.1016/j.tra.2018.12.024>
11. Belcastro, L., Marozzo, F., & Perrella, E. (2021). Automatic detection of user trajectories from social media posts. *Expert Systems with Applications*, 186, 115733. <https://doi.org/10.1016/j.eswa.2021.115733>
12. Bhat, C. R. (2015). Workshop synthesis: Conducting Travel surveys using portable devices-challenges and research needs. *Transportation Research Procedia*, 11, 199–205. <https://doi.org/10.1016/j.trpro.2015.12.017>
13. Bohte, W., & Maat, K. (2008). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, 17, 285–297. <https://doi.org/10.1016/j.trc.2008.11.004>
14. Cao, J., Hochmair, H. H., & Basheeh, F. (2022). The effect of Twitter app policy changes on the sharing of spatial information through twitter users. *Geographies*, 2, 549–562. <https://doi.org/10.3390/geographies2030033>
15. Chapleau, R., Trépanier, M., & Chu, K. K. (2008). The ultimate survey for transit planning: complete information with smart card data and GIS. In *International conference on survey methods in transport: Harmonization and data comparability*.
16. Chen, C., Gong, H., Lawson, C., & Bialostozky, E. (2010). Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transportation Research Part A: Policy and Practice*, 44, 830–840. <https://doi.org/10.1016/j.tra.2010.08.004>
17. Christiansen, P., Engebretsen, Ø., & Hjorthol, R. (2015). Nasjonal reisevaneundersøkelse på telefon eller web? TØI rapport 1426/2015.
18. Chu, K. K. A., & Chapleau, R. (2008). Enriching archived smart card transaction data for transit demand modeling. *Transportation Research Record*, 2063, 63–72. <https://doi.org/10.3141/2063-08>
19. Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*, 57, 295–310. <https://doi.org/10.1016/j.tourman.2016.06.013>
20. Ciuccarelli, P., Lupi, G., & Simeone, L. (2014). *Visualizing the Data City*. Springer Briefs in Applied Sciences and Technology. Springer.
21. Comito, C. (2018). Human mobility prediction through Twitter. *Procedia Computer Science*, 134, 129–136. <https://doi.org/10.1016/j.procs.2018.07.153>
22. Domènech, A., Gutiérrez, A., & Anton Clavé, S. (2020). Cruise passengers' spatial behaviour and expenditure levels at destination. *Tourism Planning & Development*, 17, 17–36. <https://doi.org/10.1080/21568316.2019.1566169>
23. Dypvik Landmark, A., Arnesen, P., Södersten, C.-J., & Hjelkrem, O. A. (2021). Mobile phone data in transportation research: Methods for benchmarking against other data sources. *Transportation*, 48, 2883–2905. <https://doi.org/10.1007/s11116-020-10151-7>
24. Echaniz, E., Rodríguez, A., Cordera, R., Benavente, J., Alonso, B., & Sañudo, R. (2021). Behavioural changes in transport and future repercussions of the COVID-19 outbreak in Spain. *Transport Policy*, 111, 38–52. <https://doi.org/10.1016/j.tranpol.2021.07.011>
25. Fan, J., & Stewart, K. (2021). Understanding collective human movement dynamics during large-scale events using big geosocial data analytics. *Computers, Environment and Urban Systems*, 87, 101605. <https://doi.org/10.1016/j.compenvurbysys.2021.101605>
26. Ferrer, S., & Ruiz, T. (2014). Travel behavior characterization using raw accelerometer data collected from smartphones. *Procedia—Social and Behavioral Sciences*, 160, 140–149. <https://doi.org/10.1016/j.sbspro.2014.12.125>
27. Forghani, M., Karimipour, F., & Claramunt, C. (2020). From cellular positioning data to trajectories: Steps towards a more accurate mobility exploration. *Transportation Research Part C: Emerging Technologies*, 117, 102666. <https://doi.org/10.1016/j.trc.2020.102666>
28. Fu, H., Lam, W. H. K., Shao, H., Xu, X. P., Lo, H. P., Chen, B. Y., Sze, N. N., & Sumalee, A. (2019). Optimization of traffic count locations for estimation of travel demands with covariance between origin-destination flows. *Transportation Research Part C: Emerging Technologies*, 108, 49–73. <https://doi.org/10.1016/j.trc.2019.09.004>
29. Gao, S., Yang, J.-A., Yan, B., Hu, Y., Janowicz, K., & McKenzie, G. D. (2014). Detecting origin-destination mobility flows from geotagged Tweets in greater Los Angeles area. In *Conference: Eighth international conference on geographic information science (GIScience'14)*.
30. García-Palomares, J. C., Salas-Olmedo, M. H., Moya-Gómez, B., Condeço-Melhorado, A., & Gutiérrez, J. (2018). City dynamics through Twitter: Relationships between land use and spatiotemporal demographics. *Cities*, 72, 310–319. <https://doi.org/10.1016/j.cities.2017.09.007>
31. Ge, L., Sarhani, M., Voß, S., & Xie, L. (2021). Review of transit data sources: Potentials. *Challenges and Complementarity. Sustainability*, 13, 11450. <https://doi.org/10.3390/su132011450>
32. Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys*, 49, 1–41. <https://doi.org/10.1145/2938640>
33. Grue, B., Landa-Mata, I., & Flotve, B. L. (2019). Den nasjonale reisevaneundersøkelsen 2018/19 (Nøkkellrapport No. TØI rapport 1835/2021). Oslo.
34. Gundlegård, D., Rydergren, C., Breyer, N., & Rajna, B. (2016). Travel demand estimation and network assignment based on cellular network data. *Computer Communications*. <https://doi.org/10.1016/j.comcom.2016.04.015>
35. Gunter, U., & Önder, I. (2021). An exploratory analysis of geotagged photos from instagram for residents of and visitors to Vienna. *Journal of Hospitality & Tourism Research*, 45, 373–398. <https://doi.org/10.1177/1096348020963689>
36. Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41, 260–271. <https://doi.org/10.1080/15230406.2014.890072>
37. Hjorthol, R., Engebretsen, Ø., & Uteng, T. P. (2014). Den nasjonale reisevaneundersøkelsen 2013/14—nøkkellrapport. TØI rapport 1383/2014.
38. Hong, S., Zhao, F., Livshits, V., Gershenfeld, S., Santos, J., & Ben-Akiva, M. (2021). Insights on data quality from a large-scale application of smartphone-based travel survey technology in the Phoenix metropolitan area, Arizona, USA. *Transportation Research Part A: Policy and Practice*, 154, 413–429. <https://doi.org/10.1016/j.tra.2021.10.002>
39. Huang, Q., & Wong, D. W. S. (2015). Modeling and visualizing regular human mobility patterns with uncertainty: An example using Twitter data. *Annals of the Association of American Geographers*, 105, 1179–1197. <https://doi.org/10.1080/00045608.2015.1081120>
40. Huang, R. (2023). Analyzing national parks visitor activities using geotagged social media photos. *Journal of Environmental Management*, 330, 117191. <https://doi.org/10.1016/j.jenvman.2022.117191>
41. Huang, X., Li, Z., Jiang, Y., Li, X., & Porter, D. (2020). Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLoS ONE*, 15, e0241957. <https://doi.org/10.1371/journal.pone.0241957>

42. Huntsinger, L. F. (2022). Traffic count data for travel model validation in Transportation systems planning chapter. In *Highway engineering* (2nd ed.).
43. Hussain, E., Bhaskar, A., & Chung, E. (2021). Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. *Transportation Research Part C: Emerging Technologies*, 125, 103044. <https://doi.org/10.1016/j.trc.2021.103044>
44. Jiang, K., Yin, H., Wang, P., & Yu, N. (2013). Learning from contextual information of geo-tagged web photos to rank personalized tourism attractions. *Neurocomputing*, 119, 17–25. <https://doi.org/10.1016/j.neucom.2012.02.049>
45. Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., & Newth, D. (2015). Understanding human mobility from Twitter. *PLoS ONE*, 10, e0131469. <https://doi.org/10.1371/journal.pone.0131469>
46. Kurkcu, A., Ozbay, K., & Morgul, E. F. (2016). Evaluating the usability of geolocated twitter as a tool for human activity and mobility patterns: A case study for nyc. In *Transportation research board's 95th annual meeting* (pp. 1–20).
47. Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems*, 58, 85–96. <https://doi.org/10.1016/j.compenvurbsys.2016.04.002>
48. Lee, J. H., Davis, A., McBridge, E., & Goulias, K. G. (2017). Exploring social media data for travel demand analysis: A comparison of Twitter, household travel survey, and synthetic population data in California. In *Transportation research board 96th annual meeting*, 2017-1-8 to 2017-1-12.
49. Lenormand, M., Picornell, M., Cantú-Ros, O. G., Tugores, A., Louail, T., Herranz, R., Barthelemy, M., Frias-Martínez, E., & Ramasco, J. J. (2014). Cross-checking different sources of mobility information. *PLoS ONE*, 9, e105184. <https://doi.org/10.1371/journal.pone.0105184>
50. Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301–323. <https://doi.org/10.1016/j.tourman.2018.03.009>
51. Liu, H., Chen, C., & Fan, Y. (2016). Apps and battery efficient technologies for smartphone-based travel data collection—State of the art. In *95th Annual meeting of transportation research board* (pp. 16–6184). <https://doi.org/10.1017/CBO9781107415324.004>
52. Liu, J., Li, J., Li, W., & Wu, J. (2015). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 134–142. <https://doi.org/10.1016/j.isprsjprs.2015.11.006>
53. Liu, Jiajun, Zhao, K., Khan, S., Cameron, M., & Jurdak, R. (2015). Multi-scale population and mobility estimation with geo-tagged Tweets. In *Presented at the 2015 31st IEEE international conference on data engineering workshops (ICDEW)* (pp. 83–86). IEEE. <https://doi.org/10.1109/ICDEW.2015.7129551>
54. Liu, J., Zheng, H., Feng, T., Yuan, S., & Lu, H. (2013). Post-processing procedures for passive GPS based travel survey. *Procedia—Social and Behavioral Sciences*, 96, 310–319. <https://doi.org/10.1016/j.sbspro.2013.08.038>
55. Liu, L., Zhang, Y., Ma, Z., & Wang, H. (2023). An analysis on the spatiotemporal behavior of inbound tourists in Jiaodong Peninsula based on Flickr geotagged photos. *International Journal of Applied Earth Observation and Geoinformation*, 120, 103349. <https://doi.org/10.1016/j.jag.2023.103349>
56. Liu, Q., Wang, Z., & Ye, X. (2018). Comparing mobility patterns between residents and visitors using geo-tagged social media data. *Transactions in GIS*, 22, 1372–1389. <https://doi.org/10.1111/tgis.12478>
57. Lloyd, A., & Cheshire, J. (2017). Deriving retail centre locations and catchments from geo-tagged Twitter data. *Computers, Environment and Urban Systems*, 61, 108–118. <https://doi.org/10.1016/j.compenvurbsys.2016.09.006>
58. Lobo, A. X. (1998). A review of automatic vehicle location technology and its real-time applications. *Transport Reviews*, 18, 165–191. <https://doi.org/10.1080/01441649808717009>
59. Luo, F., Cao, G., Mulligan, K., & Li, X. (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*, 70, 11–25. <https://doi.org/10.1016/j.apgeog.2016.03.001>
60. McNeill, G., Bright, J., & Hale, S. A. (2017). Estimating local commuting patterns from geolocated Twitter data. *EPJ Data Sci.*, 6, 24. <https://doi.org/10.1140/epjds/s13688-017-0120-x>
61. Mei, Z., Wang, D., & Chen, J. (2012). Investigation with bluetooth sensors of bicycle travel time estimation on a short corridor. *International Journal of Distributed Sensor Networks*, 8, 303521. <https://doi.org/10.1155/2012/303521>
62. Muñoz-Dueñas, P., Martínez-Comesaña, M., Martínez-Torres, J., & Bastos-Costas, G. (2023). Estimating mobility of tourists. New Twitter-based procedure. *Heliyon*, 9, e13718. <https://doi.org/10.1016/j.heliyon.2023.e13718>
63. Nguyen, T., Gupta, S., Raman, J., Bellomo, R., & Venkatesh, S. (2020). Geolocated Twitter-based population mobility in Victoria, Australia, during the staged COVID-19 restrictions. *Critical Care and Resuscitation*. <https://doi.org/10.51893/2020.4.SC1>
64. Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., & Maurer, P. (2014). Supporting large-scale travel surveys with smartphones—A practical approach. *Transportation Research Part C: Emerging Technologies*, 43, 212–221. <https://doi.org/10.1016/j.trc.2013.11.005>
65. Osorio-Arjona, J., & García-Palomares, J. C. (2019). Social media and urban mobility: Using twitter to calculate home-work travel matrices. *Cities*, 89, 268–280. <https://doi.org/10.1016/j.cities.2019.03.006>
66. Picornell, M., Ruiz, T., Lenormand, M., Ramasco, J. J., Dubernet, T., & Frias-Martínez, E. (2015). Exploring the potential of phone call data to characterize the relationship between social network and travel behavior. *Transportation*, 42, 647–668. <https://doi.org/10.1007/s11116-015-9594-1>
67. Pourebrahim, N., Sultana, S., Niakanlahiji, A., & Thill, J.-C. (2019). Trip distribution modeling with Twitter data. *Computers, Environment and Urban Systems*, 77, 101354. <https://doi.org/10.1016/j.compenvurbsys.2019.101354>
68. Ribeiro, S. S., Davis, C. A., Oliveira, D. R. R., Meira, W., Gonçalves, T. S., & Pappa, G. L. (2012). Traffic observatory: a system to detect and locate traffic events and conditions using Twitter. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks. Presented at the SIGSPATIAL'12: SIGSPATIAL 2012 International Conference on Advances in Geographic Information Systems, ACM, Redondo Beach California* (pp. 5–11). <https://doi.org/10.1145/2442796.2442800>
69. Rasmussen, T. K., Ingvarsson, J. B., Halldórsdóttir, K., & Nielsen, O. A. (2015). Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area. *Computers, Environment and Urban Systems*, 54, 301–313. <https://doi.org/10.1016/j.compenvurbsys.2015.04.001>
70. Salas-Olmedo, M. H., & Rojas Quezada, C. (2017). The use of public spaces in a medium-sized city: From Twitter data to mobility patterns. *Journal of Maps*, 13, 40–45. <https://doi.org/10.1080/17445647.2017.1305302>
71. Sarmiento, I., González-Calderón, C., Córdoba, J., & Díaz, C. (2013). Important aspects to consider for household travel surveys in developing countries. *Transportation Research Record*, 2394, 128–136. <https://doi.org/10.3141/2394-16>
72. Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142, 198–211. <https://doi.org/10.1016/j.landurbplan.2015.02.020>
73. Shen, L., & Stopher, P. R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport Reviews*, 34, 316–334. <https://doi.org/10.1080/01441647.2014.903530>
74. Shende, S., Bhaduri, E., & Goswami, A. K. (2023). Analyzing changes in travel patterns due to Covid-19 using Twitter data in India. *Case Studies on Transport Policy*, 12, 100992. <https://doi.org/10.1016/j.cstp.2023.100992>
75. Simaki, V., Mporas, I., & Megalooikonomou, V. (2018). Age identification of Twitter users: Classification methods and sociolinguistic analysis. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science* (pp. 385–395). Springer. [https://doi.org/10.1007/978-3-319-75487-1\\_30](https://doi.org/10.1007/978-3-319-75487-1_30)
76. Srinivasan, S., Bricka, S., & Bhat, C. (2009). Methodology for converting GPS navigational streams to the travel-diary data format. <https://doi.org/10.1017/CBO9781107415324.004>
77. Statista. (2024). Leading countries based on number of X (formerly Twitter) users as of January 2024.
78. Statista. (2021). Share of twitter users in Norway as of 3rd quarter 2021, by age group. Social Media User-Generated Content. <https://www.statista.com/statistics/585035/twitter-users-in-norway-by-age-group/>
79. Stopher, P. R., & Greaves, S. P. (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, 41, 367–381. <https://doi.org/10.1016/j.tra.2006.09.005>

80. Šulíková, S., Vanya, P., Kováč, L., Buc, D., & Farkaš, R. (2024). Use of comprehensive datasets to estimate the Slovak National Transportation Model. *Transportation Research Procedia*, 78, 546–553. <https://doi.org/10.1016/j.trpro.2024.02.068>
81. Svaboe, G. B. A., Blekesaune, A., & Tørset, T. (2023). Understanding skepticism of smartphones in travel behavior research: A qualitative approach. *Transportation Research Interdisciplinary Perspectives*, 22, 100935. <https://doi.org/10.1016/j.trip.2023.100935>
82. Svaboe, G. B. A., Tørset, T., & Lohne, J. (2024). The decline of the Norwegian national travel survey empire. *Transportation Research Procedia*, 76, 246–257. <https://doi.org/10.1016/j.trpro.2023.12.052>
83. Terroso-Saenz, F., Flores, R., & Muñoz, A. (2022). Human mobility forecasting with region-based flows and geotagged Twitter data. *Expert Systems with Applications*, 203, 117477. <https://doi.org/10.1016/j.eswa.2022.117477>
84. Terroso-Saenz, F., Muñoz, A., Arcas, F., & Curado, M. (2022). An analysis of twitter as a relevant human mobility proxy: A comparative approach in Spain during the COVID-19 pandemic. *Geoinformatica*, 26, 677–706. <https://doi.org/10.1007/s10707-021-00460-z>
85. Tsumura, Y., Asada, Y., Kanasugi, H., Arai, A., Shibasaki, R., Kawaguchi, H., & Yamada, K. (2022). Examining potentials and practical constraints of mobile phone data for improving transport planning in developing countries. *Asian Transport Studies*, 8, 100043. <https://doi.org/10.1016/j.eastsj.2021.100043>
86. Tuckel, P., & O'Neill, H. (2002). The vanishing respondent in telephone surveys. *Journal of Advertising Research*, 42, 26–37.
87. Twitter. (2022). Twitter API. Developer platform. <https://developer.twitter.com/en/docs/twitter-api>
88. Twitter. (2018). Updates to our Terms of Service and Privacy Policy [WWW Document]. <https://help.twitter.com/en/rules-and-policies/update-privacy-policy>
89. van Diggelen, F., & Enge, P. (2015). The world's first GPS MOOC and worldwide laboratory using smartphones. In *Proceedings of the 28th international technical meeting of the satellite division of the Institute of Navigation (ION GNSS+ 2015), Tampa, Florida* (pp. 361–369).
90. Vashisth, P., & Meehan, K. (2020). Gender classification using Twitter text data. In *2020 31st Irish signals and systems conference (ISSC). Presented at the 2020 31st Irish signals and systems conference (ISSC)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ISSC49989.2020.9180161>
91. Wang, A. H. (2010). Don't follow me: Spam detection in Twitter. In *International conference on security and cryptography (SECRYPT). Proceedings of the 2010 international conference* (pp. 1–10).
92. Williams, S. A., Terras, M. M., & Warwick, C. (2013). What do people study when they study Twitter? Classifying Twitter related academic papers. *Journal of Documentation*, 69, 384–410. <https://doi.org/10.1108/JD-03-2012-0027>
93. Wilson, J. (2004). Measuring personal travel and goods movement: A review of the Bureau of Transportation Statistics' Surveys. In *TRB special report* (pp. 28–31).
94. Wolf, J. (2004). Applications of new technologies in travel surveys. In *Travel survey methods—Standards and future directions* (pp. 531–544).
95. Wong, E., Law, R., & Li, G. (2017). Reviewing geotagging research in tourism. In R. Schegg & B. Stangl (Eds.), *Information and communication technologies in tourism 2017* (pp. 43–58). Springer. [https://doi.org/10.1007/978-3-319-51168-9\\_4](https://doi.org/10.1007/978-3-319-51168-9_4)
96. Xiao, G., Juan, Z., & Zhang, C. (2015). Travel mode detection based on GPS track data and Bayesian networks. *Computers, Environment and Urban Systems*, 54, 14–22. <https://doi.org/10.1016/j.compenvurbsys.2015.05.005>
97. Yang, S., & Wu, Y.-J. (2018). Travel mode identification using bluetooth technology. *Journal of Intelligent Transportation Systems*, 22, 407–421. <https://doi.org/10.1080/15472450.2017.1384698>
98. Zhong, C., Morphet, R., & Yoshida, M. (2023). Twitter mobility dynamics during the COVID-19 pandemic: A case study of London. *PLoS ONE*, 18, e0284902. <https://doi.org/10.1371/journal.pone.0284902>
99. Zhou, X., Xu, C., & Kimmons, B. (2015). Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Computers, Environment and Urban Systems*, 54, 144–153. <https://doi.org/10.1016/j.compenvurbsys.2015.07.006>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.