Journal of Computational Mathematics and Data Science 16 (2025) 100122

Contents lists available at ScienceDirect

Journal of Computational Mathematics and Data Science

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

Capturing patterns and radical changes in long-distance mobility by Flickr data

Anton Galich

Department of Mobility Spaces, German Aerospace Center, Institute of Transport Research, Rudower Chaussee 7, 12489 Berlin, Germany

ARTICLE INFO

Dataset link: https://github.com/antonlich/flickr_longdistance_mobility.git

Keywords:
Flickr data
Long-distance mobility
Travel patterns

ABSTRACT

In contrast to daily travel behaviour, long-distance mobility constitutes a poorly understood area in transport research. Only few national household travel surveys include sections on long-distance travel and these usually focus on the trip to the destination without gathering information about mobility behaviour at the destination. Other sources of data on mobility are either restricted to the national level such as cell phone data or to specific modes of transport such as international flight statistics or floating car data. In addition, the outbreak of the COVID-19 pandemic in 2020 has illustrated how difficult it is to grasp abrupt changes in mobility behaviour.

Against this background this paper investigates the potential of Flickr data for capturing patterns and radical changes in long-distance mobility. Flickr is a social media online platform allowing its users to upload photos and to comment on their own and other users' photos. It is mainly used for sharing holiday and travel experiences. The results show that Flickr constitutes a viable source of data for capturing patterns and radical changes in long-distance mobility. The distribution of the travel distances, the travel destinations as well as reduction of the mileage of all holiday trips in 2020 in comparison to 2019 due to the pandemic calculated on the basis of the Flickr data is very similar to the same indicators determined on the basis of a national household travel survey, official passenger flight statistics, and other official transportation statistics.

1. Introduction

Long-distance mobility constitutes a poorly understood area in transport research. While daily travel behaviour is captured quite well by large national household travel survey in many countries and is often quite predictable [1], long-distance mobility, defined as trips of at least 100 Kilometres, is often not surveyed at all or only by a few questions on the trip to the destination [2]. However, it is known that long-distance mobility accounts for a considerable share of the overall passenger transport demand. In Germany, for instances, long-distance trips account for 46% of the kilometres travelled by the residential population [3]. Also on the European level around 50% of all passenger kilometres travelled come from long-distance trips [4].

In addition to long-distance mobility, it is also difficult to capture sudden changes in mobility by large national household travel surveys as these usually need a lot of preparation. In contrast, radical changes in mobility can occur from one day to the other as the outbreak of the COVID19-pandemic illustrated. Consequently, many researchers rushed to find data sources that could help

Abbreviations: AM, ante meridiem (before noon); API, Application Programming Interface; COVID19, Coronavirus disease; DBSCAN, Density-based spatial clustering of applications with noise; i.e., id est (that is to say); Km, Kilometres; MiD, Mobilität in Deutschland (Mobility in Germany); PM, post meridiem (after midday); UK, United Kingdom of Great Britain and Northern Ireland; USA, United States of America

E-mail address: anton.galich@dlr.de.

https://doi.org/10.1016/j.jcmds.2025.100122
Received 6 March 2025; Received in revised form 10 June 2025; Accepted 22 June 2025

revealing the tremendous impact that the outbreak of the pandemic had on people's mobility so that a rich literature emerged on this topic in the past years [5–7]. Yet, the outbreak of the COVID19-pandemic constitutes just one example of an event that drastically changed mobility patterns from one day to the other. The attacks that occurred on the 11th of September in 2001 on the World Trade Centre and the Pentagon in the United States of America denote another example in which global mobility was changed from one day to the other [8]. In addition, there are various other events or developments that had a less immediate but nevertheless quite huge impact on global mobility such as the oil crises in the 80s or the emergence of affordable air travel in the 90s and early 2000s [9]. Finally, there are reoccurring events that can immediately impact mobility when they occur but their occurrence is either known well in advance such as in the case of the Olympic Games [10] or national holidays or their occurrence is more or less expected in certain times of the year such as in the case of specific weather conditions as snow and ice [11].

Social media platforms constitute a promising source of data for analysing changes in mobility and have already been used successfully to study other societal phenomena such as political polarization [12]. Therefore, various researchers have analysed information from social media platforms in order to evaluate its usefulness for transport research in recent years. Several researchers have, for instances, used geo-referenced Twitter data to analyse daily mobility behaviour [13,14], to predict next morning traffic [15], or to detect traffic incidents [16,17]. Others have used geo-referenced photo uploads from the Flickr platform to identify popular tourist attractions in cities [18,19], estimate visitor trajectories in world heritage cities [20], analyse international travel behaviour [21–23], and to predict national air travel [24].

Flickr is a social media online platform allowing its users to upload photos and to comment on their own and other users' photos. It is mainly used for sharing holiday and travel experiences. Against this background, various studies have evaluated the potential of Flickr data for land use attribution by comparing machine-generated and user-generated Flickr photo tags to the land use and land cover annotations of the authoritative data set provided by public administration [25]. Yet, while geotagged Flickr photos provide a lot of useful information, the sociodemographic background of the people contributing this data might systematically differ from the one of the average citizen and also the photos taken might focus on rather scenic parts of the landscape and not provide a good overall coverage of a certain area [26]. On the other hand, these potential biases in the Flickr data can help to reveal the most attractive and visited parts of forested areas used for recreational purposes and thus provide a useful source of data for specific areas of research where other information is difficult to obtain [27]. It is also noteworthy that different social media platforms might provide more reliable information for specific research questions and certain places than other. While, for instances, Instagram provides more reliable data than Flickr and Twitter on the monthly visitor patterns of national parks in South Africa and Finland, the differences between the different sources of data are considerably more evident in South Africa than in Finland [28].

Thus, various scholars have evaluated the potential of Flickr data for analysing different aspects in the area of transport research. However, none of these papers looks at Flickr users from a specific country and compares their travel behaviour with the one captured by long-distance sections in large national household travel surveys. Neither does any of the analyses conducted shed light on whether Flickr data can be used to capture radical changes in long-distance mobility. This paper contributes to closing this gap in the research and demonstrates the potential of Flickr data for capturing patterns and radical changes in long-distance mobility. For this purpose, key indicators such as trip length, travel destinations, and changes in mobility due to the COVID19 pandemic are calculated on the basis of the Flickr data of users living in Germany and compared to the same indicators derived from a large national household travel survey, official passenger flight statistics, and other official transportation statistics. As the analysis of long-distance mobility is relevant in many countries, the paper provides a methodology and a source code that can easily be adapted to conduct the same analyses in any other country.

The remainder of this paper is structured as follows. This brief introduction is followed by a chapter in which the Flickr data basis and the methods used for analysing it are explained in more detail. Subsequently, the main empirical results are presented. Finally, the most important insights from this study are summarized in the discussion conclusions section.

2. Data and methods

First the data accessed via the public Flickr Application Programming Interface (hereinafter "API") are illustrated. Subsequently, the methods applied for processing and analysing the data are outlined before the data sources used for the comparison with key indicators calculated on the basis of the Flickr Data are described.

2.1. Flickr data

Uploaded photos to the Flickr platform include information on the date and time when the photo was taken and uploaded as well as the latitude and longitude coordinates of the geographic location in which it was taken. These information as well as data on the user such as his or her hometown, city and country of residence, and occupation are available via a public API provided by Flickr.

This API was used to download all geographic locations of the photos taken between 00:00AM 01.01.2017 and 12:00PM 31.12.2021 by users residing in Germany. For the further analysis, only those users who uploaded at least two geo-references photos were kept in the data set as at least two different locations are needed to calculate the distances travelled in the destination region. This resulted in 9029 geo-referenced photos of 518 users. The distribution of these locations over the globe is illustrated in Fig. 1.

The documentation of the Flickr API does not provide any information on whether the data provided actually includes all photos of users residing in Germany or only a sample. However, as 9029 photos constitutes a rather small number for a 5-year-period, it can be assumed that the data retrieved only illustrates a small subsample of all photos uploaded by users from Germany from 2017 to 2021. Nevertheless, the data provided can be processed to reveal more information about long-distance mobility of Flickr users as illustrated in the following subchapter.

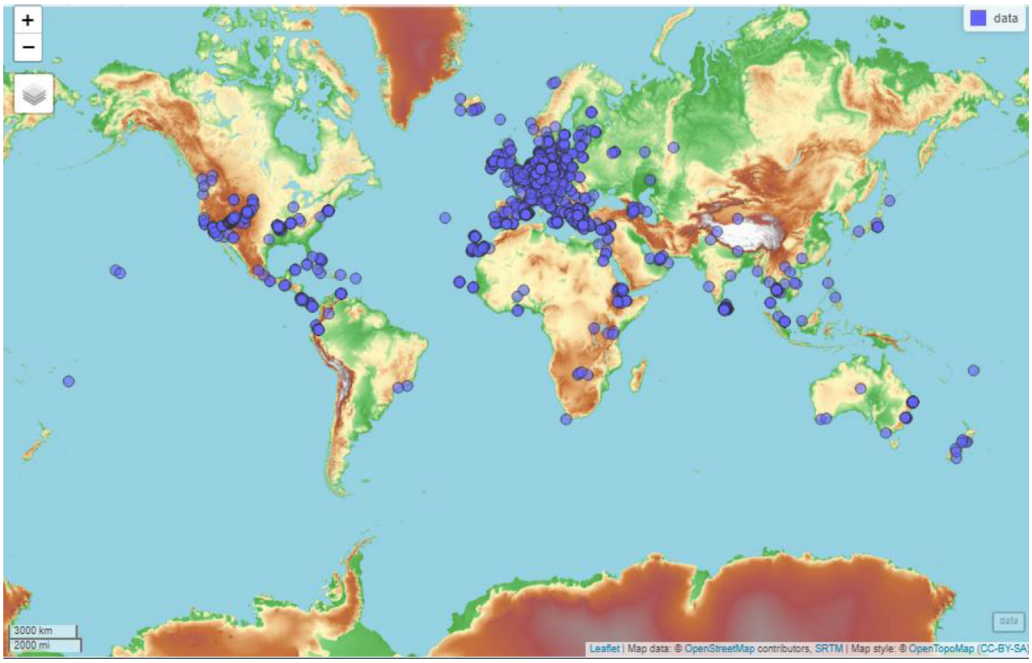

Fig. 1. Geo-referenced locations of the photos taken between 00:00am 01.01.2017 and 12:00pm 31.12.2021 and uploaded by Flickr users residing in Germany. The blue dots show where the photos were taken that the Flickr users uploaded to the platform.

2.2. Methods

The raw Flickr data basically constitutes a list of locations with a timestamp (date and time) for each user. Therefore, various data processing steps had to be taken in order extract trips undertaken from the place of residence to travel destinations as well as the trips undertaken at the travel destination within a holiday trip. For this purpose, the spatial and temporal dimension of the data were considered and various rules were set up as explained in more detail in the following paragraphs.

First, the latitude and longitude coordinates of the place of residence of the Flickr users had to be determined on the basis of the city names that the data provided. This was done by feeding the city names to the Nomatim search engine[1] and extracting the centroid coordinates for each city. In the next step it was determined which are of the coordinates extracted fall within Germany's borders. Those users with a place of residence outside of Germany were deleted from the data set as users residing in Germany were needed for a valid comparison with the other sources of data outlined in more detail in subchapter 2.3. Hence the centroids of the cities provided lying within Germany's borders were taken as the coordinates of the place of residence of the Flickr users and used as the basis for the later calculation of distance measures to the travel destinations.

After determining the coordinates of the places of residence, the locations of the photos uploaded were spatially clustered with a density-based spatial clustering of applications with Noise (DBSCAN) algorithm. Finding the right parameters for algorithm was a bit of a trial and error procedure with visual inspections of various clusters until a satisfactory solution was found. Fig. 2 illustrates the results of this procedure on the basis of selected clusters.

In the next step it had to be determined which clusters at least partly overlap with the region of the place of residence and thus cannot be considered as long-distance trips. For this purpose, a radius of 100 km was drawn around the places of residence. The threshold of 100 km to distinguish long-distance trips from everyday mobility was derived from the scholarly literature. While generally no common definition of long-distance trips exists [29] and it is possible to define long-distance trips by different criteria such as the duration of the stay, the distance to the destination [30] or the purpose of the trip [31], many studies use the threshold of 100 km distance to the destination to distinguish long-distance trips from everyday mobility [3,4,30].

The spatial clusters of the travel destinations were also transformed into geographical objects depending on the number of locations in each cluster. A radius of 100 km was drawn around the clusters that contained only one location. Clusters containing two or three locations were transformed into bounding boxes on the basis of the minimum and maximum latitude and longitude values. Finally, clusters containing four or more locations were transformed into polygons with the outmost locations determining the outer boundaries of each polygon. In a next step it was checked for each of these geographic objects containing the destinations whether there is an overlap with the 100 km radius around the place of residence. Those clusters for which an overlap was found were removed from the data set as 100 km was assumed to be threshold between long-distance and daily or local mobility.

[1] Nomatim search engine based on OpenStreetMap data: https://nominatim.openstreetmap.org/ui/search.html.

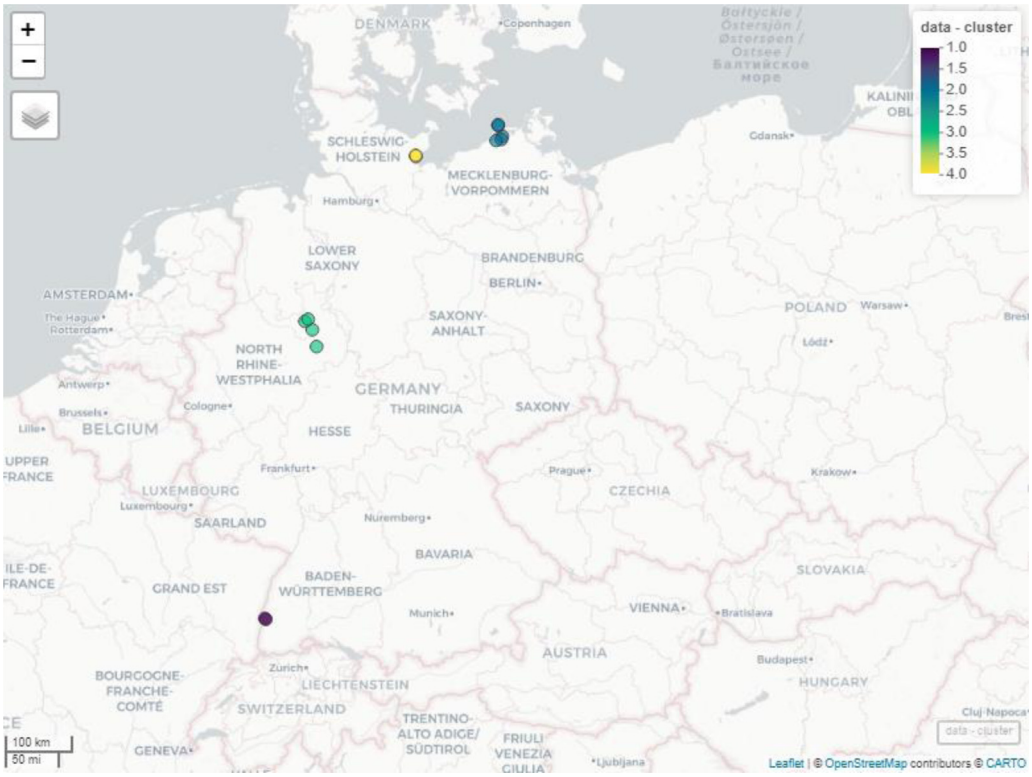

Fig. 2. Illustration of the spatially clustered locations of the photos uploaded by one user. The dots denote where the photos that the user uploaded were taken. The colour of each dot indicate to which cluster of dots is what assigned to in the clustering process.

The next step of the data preparation focused on the temporal dimension in order to ensure that all locations of a spatial cluster actually were part of the same holiday trip. For this purpose, first the mean distance (Haversine, as the crow flies) of each cluster from the place of residence was calculated as the mean of the distances of each location of the cluster to the place of residence. Thereafter, the following two assumptions were made and implemented:

(1) If the mean distance of a cluster from the place of residence is less than or equal to 750 km and the time span between two temporally consecutive locations of the cluster is larger than 30 days, then the locations before and after this time span are regrouped into two separate clusters

(2) If the mean distance of a cluster from the place of residence is larger than 750 km and the time span between two temporally consecutive locations of the cluster is larger than 90 days, then the locations before and after this time span are regrouped into two separate clusters

The resulting spatial and temporal clusters were treated as individual holiday trips in the further analysis. For the calculation of travel statistics to different destinations, the locations of all clusters were matched with country borders and each cluster was assigned that country and continent which most of its locations fell into. Furthermore, the distance between the centroid and the most distant location of each cluster was calculated and treated as the radius of movement at the travel destination.

The data preparation steps were necessary in order to calculate key indicators on the basis of the Flickr data which can be compared to the same indicators calculated on the basis of more established sources of data which are presented in the following subchapter.

2.3. Data sources used for the comparison with the Flickr data

Three different sources of data were used for the comparison with the key indicators calculated on the basis of the Flickr data. First of all, the most recent version of Germany's largest national household travel survey "Mobilität in Deutschland 2017" (Mobility in Germany 2017,[2] hereinafter "MiD 2017") was used for a comparison of the distances travelled. The MiD 2017 was conducted from May 2016 to September 2017 with 316,361 individuals providing information on their daily mobility via 960,619 trips described in travel diaries. In addition, the MiD 2017 includes a subsample of 20,454 individuals who answered questions about trip length,

[2] More information about the Mobility in Germany 2017 survey can be found here: https://www.mobilitaet-in-deutschland.de/archive/index.html.

destination etc. on 38,905 long distance trips.[3] The holiday trips of this subsample (which also includes business trips) will be used for a comparison of the distances travelled calculated on the basis of the Flickr data as it is assumed that the large majority of the photos uploaded on Flickr was taken on holiday trips.

The air traffic statistics provided by the Federal Statistical Office of Germany constitute another important source of data for the comparison with the Flickr data. More specifically, the number of passengers boarding an airplane at any airport in Germany per year and country of destination is used.[4] For the comparison both the countries of destination in the Flickr data and in the official air traffic statistics are aggregated into three different regions of the world: Germany, Europe, Rest of the world. These categories are mutually exclusive, *i.e.* people travelling within Germany are not included in the categories Europe or Rest of the world and people travelling to another European country and not included in the category Rest of the world.

Finally, for the comparison of the changes in long-distance mobility from 2019 to 2020 due to the COVID19-pandemic, the annual compendium "Verkehr in Zahlen" (hereinafter "Transport in Figures") provided by the German Federal Ministry for Digital and Transport was used. "Transport in Figures"[5] constitutes the central reference for statistics on transport in Germany. Among other statistics, it outlines key mobility indicators of passenger transport for the entire population of Germany with a focus on the trips made, the kilometres travelled, the modal split, and the trip purposes per year. In this study the decline in the kilometres travelled from 2019 to 2020 on holiday trips was used for the comparison with the Flickr data.

3. Results

The presentation of the results begins with the comparisons of the key indicators calculated on the basis of the Flickr data and the three more established sources of data. Thereafter, some insights gained from the Flickr data on the travel behaviour in the destination region are outlined.

3.1. Comparison of the distances travelled with Germany's largest national household travel survey

To ensure a as valid as possible comparison of the distances travelled to the destinations in the Flickr data with the long-distance trips in the MiD 2017, the distances travelled in the Flickr data were grouped into the same categories as in the MiD 2017. Furthermore, as the MiD 2017 was conducted in 2016 and 2017, on the distances travelled in 2017 were selected in the Flickr data. In the MiD 2017 only the holiday trips were selected as it is assumed that also the Flickr data mainly stem from holiday trips. Fig. 3 illustrates the results of the comparison

The distribution of the travel distances in Flickr data is very similar to the one of the holiday trips of the long-distance data set of the MiD 2017. In fact, only trips of a length of 2000 km or more have a twice as high share in the Flickr data (25%) in comparison to the MiD 2017 (13%). There could be various explanations for this. One could, for instances, assume that people who upload photos on Flickr are on average younger and have a higher degree of education and a higher salary than the average population in Germany and are thus also more eager to travel to far away destinations. Another assumption could be that trips shorter than 100 kilometres are more frequently conducted and the locations are better known to people so that they do not take new pictures each time. That could explain the lower share of trips shorter than 100 kilometres in the Flickr data. However, without any data on the motivation behind a trip and the actual sociodemographic background of the Flickr users these assumptions remain pure speculation.

Nonetheless, the distribution of the distances travelled to the destinations is remarkably similar in the Flickr data and the MiD 2017.

3.2. Comparison of the destination regions with official air traffic statistics

For the comparison of the destinations in the Flickr data and the official air traffic statistics the data for the years 2017, 2018, and 2019 were selected from both sources of data. For the Flickr data this means that only those trips derived from photos taken in 2017, 2018, or 2019 were selected. For the air traffic statistics from the Federal Statistical Office of Germany this means that the number of people boarding an airplane at any airport in Germany in 2017, 2018, or 2019 was selected. In addition, in both sources of data the destination countries were grouped into the mutually exclusive categories of Germany, Europe, Outside of Europe. Fig. 4 outlines the results of the comparison.

As can be seen, the travel destinations of the Flickr data in 2017 are quite similar to the ones in the official passenger flight statistics. However, in 2018 and 2019 the differences in the shares of trips within Germany and trips to other European countries in the Flickr data and the official passenger flight statistics increase considerably. These differences might be explained by the Flickr data also containing travel conducted by car or public transport, *i.e.* shorter trips within Germany which are less often conducted by airplane.

[3] The actual data collected in the Mobility in Germany 2017 survey can be requested here: https://daten.clearingstelle-verkehr.de/279/.

[4] The Office of Statistics of Germany sorts the data it provides by numbers. Statistics on air traffic can generally be found by the number 46421. The actual data used in this study is marked with the number 46421-0007 and can be retrieved here: https://www-genesis.destatis.de/genesis//online?operation=table&code=46421-0007&bypass=true&levelindex=0&levelid=1692946459814#abreadcrumb.

[5] More information about the Transport in Figures statistics as well as the actual annual reports can be accessed on the webpage of the Federal Ministry for Digital and Transport of Germany: https://bmdv.bund.de/SharedDocs/DE/Artikel/G/verkehr-in-zahlen.html.

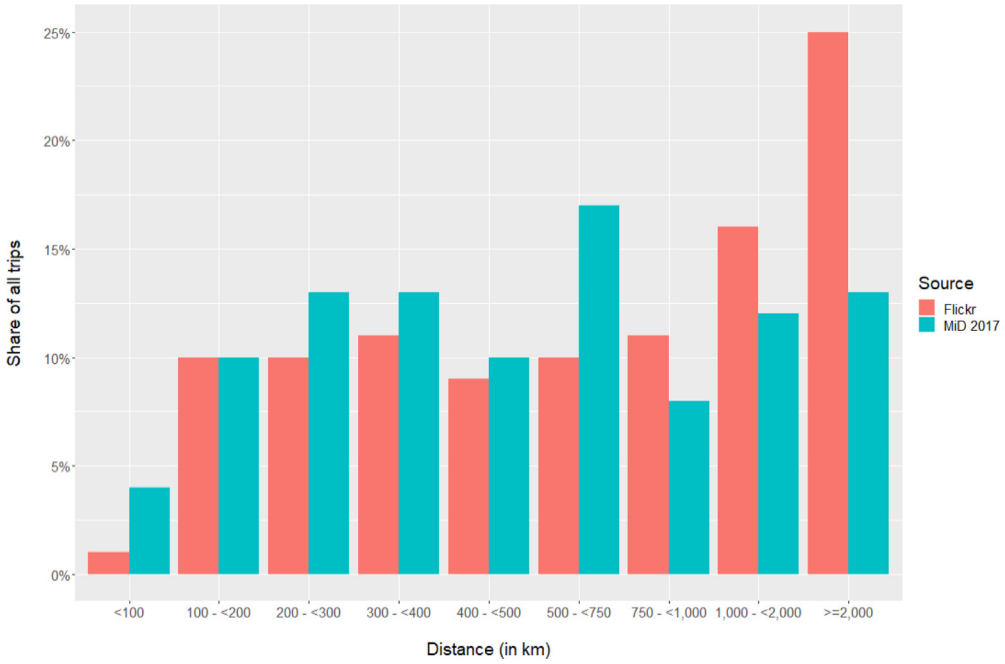

Fig. 3. Distances travelled in the MiD 2017 and the Flickr data. The red bars illustrate the share of all trips grouped by specific distance classes calculated on the basis of the Flickr data. These are compared to the turquoise bars which represent the share of all trips grouped by specific distance classes calculated on the basis of the largest national household travel survey in Germany.

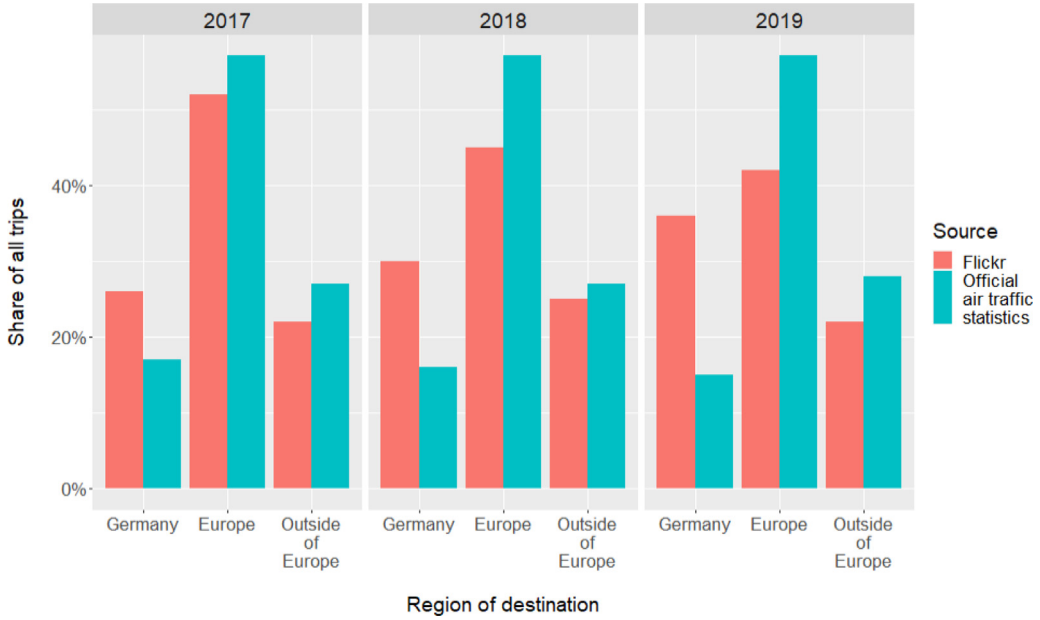

Fig. 4. Share of trips per destination region and year. The red bars outline the share of all trips calculated on the basis of the Flickr data that took place within Germany, within Europe or went outside of Europe in the years 2017, 2018, and 2019. The turquoise bars outline the equivalent share calculated on the basis of official air traffic statistics.

3.3. Comparison of the changes in the kilometres travelled from 2019 to 2020 due to the COVID19-pandemic

The outbreak of the COVID19-pandemic in 2020 dramatically altered people's travel behaviour all over the globe. In a nutshell, due to lockdowns, closed borders etc. the number of the trips made and the kilometres travelled decreased sharply in many countries. To analyse whether the changes are also adequately reflected in the Flickr data, a comparison with the changes in travel behaviour

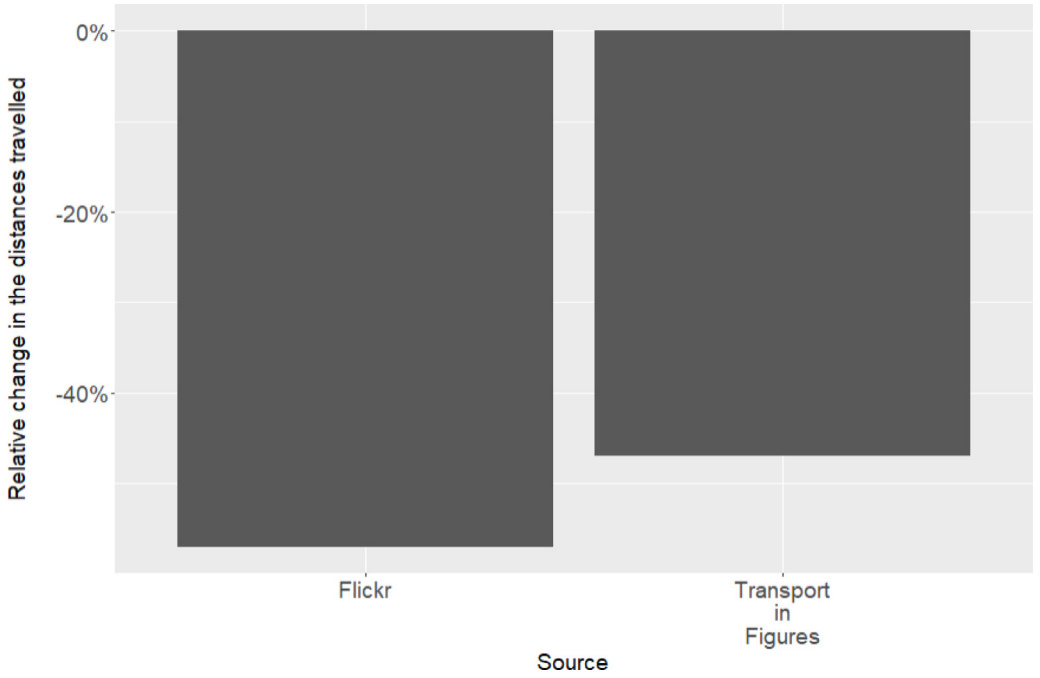

Fig. 5. Relative change in the distances travelled from 2019 to 2020. The left bar shows the relative decline in the distances travelled from 2019 to 2020 calculated on the basis of the Flickr data, while the right bar shows the equivalent taken from a national report on mobility in Germany.

Fig. 6. Locations of the photos taken in 2019 from 01 April onwards. Each blue dot represents one photo.

outlined in the "Transport in Figures" report 2020/2021 of the Federal Ministry for Transport and Digital of Germany was made. For this purpose, the changes in the overall distances travelled from 2019 to 2020 was calculated on the basis of the Flickr data and contrasted with the changes in the kilometres travelled on holiday trips outlined in "Transport in Figures". Fig. 5 illustrates the relative changes in the distances travelled from 2019 to 2020 in the Flickr data and on holiday trips in "Transport in Figures".

The reduction of the mileage of all holiday trips in 2020 in comparison to 2019 due to the pandemic amounts to 57% in the Flickr data which is relatively close to the 47% outlined in the statistics of the Federal Ministry for Digital and Transport. The larger reduction in the Flickr data might rely on the longer distances travelled outlined in subchapter 3.1, *i.e.* if Flickr users travel longer distances than the average German population in non-pandemic years, then also a higher reduction in the distances travelled in a pandemic year makes sense.

The reduction in the distances travelled can also be seen in Fig. 6 and Fig. 7 which outline the locations of the uploaded photos which were taken in 2019 and 2020. Both figures only show the locations of the photos taken after the 1st of April as the beginning of the pandemic was officially declared in March 2020 in Germany.

Fig. 7. Locations of the photos taken in 2020 from 01 April onwards. Each blue dot represents one photo.

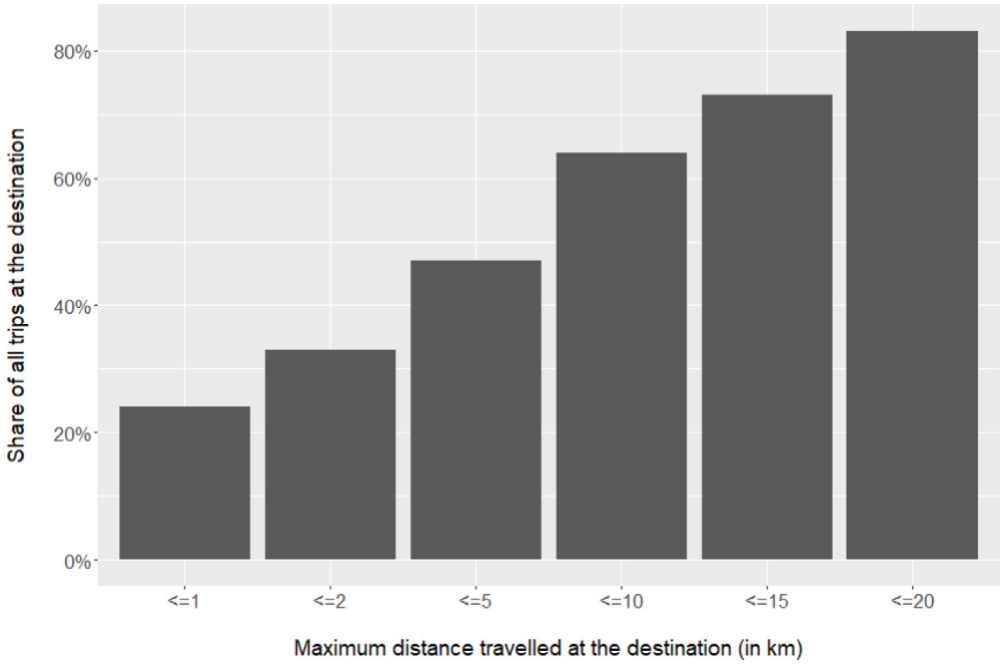

Fig. 8. Share of all trips by maximum distance travelled at the destination. Each bar shows the share of all trips at the destination grouped by distance classes in Kilometres.

3.4. Travel behaviour at the destination

As other sources of data usually do not cover people's travel behaviour at their holiday destination, the Flickr data was used to calculate the radius of movement at the travel destination. For this purpose, the distance (Haversine, as the crow flies) from the centroid of the cluster of locations at the destination clusters to the outmost location was calculated. This distance was treated as the radius of movement at the travel destination. Fig. 8 outlines the distribution of the radius of movement at the destination across all trips in the Flickr data.

The radius of movement is shorter than 16 km at 73% of the destinations in the Flickr data. In addition, the Pearson correlation coefficient was calculated between the maximum distance travelled at the destination and the distance travelled to the destination. The resulting value of 0.11 indicates that there is no or only a small correlation between the radius of movement at the destination and the distance of the place of residence to the travel destination. Hence much of the mobility at the destination could be conducted by sustainable modes of transport such as walking or cycling with bikes or e-bikes even at far away destinations.

4. Discussion

This study focused on investigating if Flickr constitutes a useful source of data for analysing long-distance trips and abrupt changes in mobility. In so doing, the results outlined above contribute to diverse strands of the scholarly literature as illustrated in more detail in the following paragraphs.

First, it was shown that the distribution of the travel distances and travel destinations in the Flickr data is very similar to the one in large national household travel surveys and official statistics which confirms the findings of various other studies. [22], for instances, used geo-tagged Flickr photos to analyse international travel to the UK. For this purpose, users who took at least one photo located in the UK and at least one photo located outside the UK were extracted and their country of residence was determined by analysing in which country they spent most of their time in the previous 12 months [22]. After this data preparation the number of visitors coming to the UK from 28 different countries in the Flickr data was compared to the official estimates of the UK Office for National Statistics and a significant correlation was found with a quite high Pearson's r of 0.86 [22].

The approach taken in this study differs from the one of [22] in that not incoming international travel to Germany was analysed but rather outgoing international trips of people whose place of residents lies within Germany. Yet, the results point in the same direction as the findings of [22]. The destinations of the long-distance trips of people residing in Germany calculated on the basis of the Flickr data are very similar to the ones derived from official air travel statistics and from the long-distance trips in Germany's largest national household travel survey. Thus, although applying slightly different research approaches, both the study at hand and the one of [22] demonstrate that Flickr constitutes a viable source of data for analysing international travel.

In a similar vein, the findings of this study support the results of [23] who undertook a similar approach as [22] and expanded it to several countries. In fact, [23] analysed geo-tagged Flickr photos to detect the place of residence of the users and their trips abroad. On this basis the number of international visitors for Canada, France, Germany, Italy, Japan, the UK, and the USA is estimated and compared to the official statistics on visitors from abroad of each of these countries [23]. In so doing, significant correlations between the number of visitors from abroad derived from the Flickr data and official statistics were found for each of the seven countries with Pearson's correlation coefficient ranging from 0.693 for France to 0.903 for Germany [23].

In contrast to [23] the study at hand focused on outgoing instead of incoming international trips. Based on the Flickr data of people residing in Germany, the share of long-distance trips to destinations in Germany, other countries in Europe, and other countries in the rest of the world was calculated and compared to the official air travel statistics for the years 2017, 2018, and 2019. The results were relatively similar in spite of the Flickr data generally showing higher shares of long-distance trips to destinations in Germany than the official air travel statistics. However, this difference could rely on the Flickr data also including trips conducted by car, bus, or train which are more often used for long-distance trips to other destinations in Germany than to destinations abroad. Therefore, both the findings of the study at hand as well as the ones of [23] indicate that Flickr constitutes a viable source of data for the analysis of long-distance trips.

The only other study found that also used Flickr data to calculate the distances travelled by the users was conducted by [21]. In fact, the radius of gyration of Flickr users residing in the USA, the UK, Spain, Germany, Italy, France, Canada, Australia, Brazil, Japan, the Netherlands, and China was calculated [21]. However, in contrast to the study at hand, [21] did not compare the distances travelled to the long-distance trips of a large national household travel survey or any official statistics. Instead, [21] found a high correlation with an average Pearson's correlation coefficient of 0.78 between the number of Flickr users residing in the USA, the UK, Spain, Italy, France, Canada, Australia, Brazil, Japan, or the Netherlands and the number of international departures from the respective country. Consequently, the results of [21] are not directly comparable to the findings of the study at hand but the outcomes of both studies provide evidence for the validity of the Flickr data.

Furthermore, the results of this study supplement the findings of Beiró, Panisson, Tizzoni and Cattuto [24] who used geo-referenced Flickr photos to analyse both domestic travel of residents of the USA. In fact, Beiró, Panisson, Tizzoni and Cattuto [24] fed the travel flows derived from the Flickr data it into a gravity model based on traffic analysis zones and then compared the resulting origin destination matrices of the base gravity model and the one refined by the Flickr data with official air travel and commuting statistics in the USA. In so doing, it is shown that the Flickr data improves the gravity models in both predicting air travel and commuter trips in the USA Beiró, Panisson, Tizzoni and Cattuto [24]. In addition, also the Flickr data alone shows a high correlation with actual air travel and commuter travel in the USA Beiró, Panisson, Tizzoni and Cattuto [24].

In contrast to Beiró, Panisson, Tizzoni and Cattuto [24], the study at hand does not analyse the long-distance trips between different traffic analysis zones in Germany but rather looks at the share of the overall long-distance trips of people residing in Germany to destinations within Germany and to other regions of the world and at the distances travelled. It was shown that the distances travelled on long-distance trips calculated on the basis of the Flickr data are relatively similar to the distances travelled on the long-distance trips of the largest national household travel survey in Germany. Furthermore, the shares of the overall trips to destinations within and outside of Germany in the Flickr data are close to the shares of the official air travel statistics. Thus, the findings of the study at hand supplement the results of Beiró, Panisson, Tizzoni and Cattuto [24] and both studies provide proof of the validity of Flickr data for analysing mobility patterns.

By demonstrating the usefulness of Flickr data for analysing long-distance mobility, the results of this study also supplement the findings of other researchers using Flickr data for other purposes. [27] showed that Flickr data is useful in analysing domestic recreational activities in forest areas. [25] demonstrated that Flickr data can generally be used for classifying land use and land cover, although it works better for recreational areas such as public parcs than for agricultural lands. Similar findings were brought about by [26] who demonstrated that geo-tagged Flickr photos tend to have a focus on scenic landscapes that are easily accessible by transport infrastructure, while other areas might be less well captured. Also, [28] point out that Flickr data works particularly well

for analysing visitor patterns to popular parks in Finland and South Africa. This study supplements the findings of these researchers by showing that Flickr data can also be used for analysing long-distance trips to both domestic and international destinations.

In addition, this study with its focus the long-distance trips of people residing in Germany supplements various other studies that highlighted mobility patterns at tourist destinations based on Flickr data. [18] combined Flickr data and OpenStreetMap information on Node tags to develop a method to identify tourist attractions in cities. [20] used geo-tagged photos from Flickr to analyse the travel trajectories of tourists in the city of Toledo in Spain. [19] focused on developing a general approach for identifying popular tourist routes in any given area based on Flickr data. In contrast, the study at hand analysed the mobility patterns of Flickr users residing in Germany on their trips to their long-distance destinations and shows that Flickr data is not only useful for analysing mobility at tourist destinations but also for highlighting the long-distance trips undertaken to reach these destinations.

This study has also shown that Flickr data captured the drop in the overall mileage travelled from 2019 to 2020 due to the outbreak of the COVID19-pandemic quite well. However, the analysis was performed on all annual basis because the data accessed via the official API was too scarce to allow for a more fine-grained temporal investigation. Therefore, the data used in this study would not allow analysing the impact on mobility of many other disruptive events pointed out in the scholarly literature such as severe weather, the closure of infrastructure, strikes in public transport [11], or large events such as the Olympic Games [10]. In addition, the Flickr data used in this study would also not allow for an investigation of the impact of the COVID19-pandemic on mobility during the different stages of the pandemic within a year as conducted by [6]. Thus, situating the results of this study in the scholarly literature on the impact of disruptive events on mobility helps revealing the limitations of the Flickr data that were accessed via the official API.

Nonetheless it has been shown that Flickr can constitute a viable source of data to shed light on the less well research area of long-distance mobility. While other researchers suggested combining transport statistics and survey data [3], proposed new survey designs [2,30,31], developed a new travel demand model [4], or conducted discrete choice experiments [29] in order to highlight long-distance mobility, the results of this study have pointed out that the distances of the long-distance trips calculated on the basis of the Flickr data are relatively similar to the distances travelled on the long-distance trips of the largest national household travel survey in Germany. Thus, one of the contributions of the study at hand to the scholarly literature is the insight that social media data should be considered as another viable source of data for the analysis of long-distance mobility.

5. Conclusion

This study has demonstrated that Flickr constitutes a viable source of data for capturing patterns and radical changes in long-distance mobility. Therefore, Flickr data can be used as a supplementary data source in the analysis of long-distance mobility which is often only poorly covered in national household travel surveys. Flickr data not only allows to analyse the long-distance trips from the residents of a particular country to international destinations but also to capture abrupt changes in long-distance trips caused by unforeseen events such as the outbreak of the COVID19-pandemic. Hence Flickr data can be a meaningful complement to other data sources in transport research such as large national household travel surveys.

In addition, analysing the Flickr data can provide useful information on the travel behaviour at the destination which usually is not covered by other sources of data and this in general not that well understood, yet. This study has found that the radius of movement is shorter than 16 km at 75% of the destinations in the Flickr data and does not correlate with the distance of the place of residence to the travel destination. Hence much of the mobility at the destination could be conducted by sustainable modes of transport such as walking or cycling with bikes or e-bikes even at far away destinations. This can be useful information for both travel agencies in the countries of origin as well as tourist boards, public administration, and public transport companies in the destination regions.

Therefore, future research should further explore the potential of Flickr data as a supplementary source of data in the analysis of long-distance mobility. Among others, future studies should use Flickr data to investigate changes in long-distance mobility over a longer period of time. If Flickr remains a popular social media platform and keeps its public API, analysis of changes in long-distance mobility in the future could cover a time span of 20 or 30 years and show how events such as global economic crises or wars impact long-distance mobility and how long this impact lasts. In addition, future studies should use Flickr data to compare long-distance trips of the residents of different countries and to analyse their mobility at their destination in more detail. In so doing, not only potential differences in long-distance trips between the residents of different countries could be revealed but also potential differences in their mobility at their destinations. In particular the latter constitutes a so far less researched area and thus there is a lot of room for future studies to deploy Flickr or other social media data to shed light on mobility patterns at the destinations of long-distance trips.

For this purpose, it would be desirable if the companies providing social media data via a public API such as Flickr, X (the former Twitter), or Facebook could be a bit more transparent with regard to the actual data provided. In particular, it would be helpful if it was outlined weather the data of all users or only of a sample of all users is made available and, if so, how the sample was drawn and what share of users it represents. An improved transparency would help in increasing the usage of social media data in transport research as a complimentary source of data to shed light on certain aspects which are not well covered by more traditional sources of data.

CRediT authorship contribution statement

Anton Galich: Study conception and design, Data collection, Analysis and interpretation of results, Draft manuscript preparation.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

No funding was received for the development of this article.

Declaration of competing interest

The author(s) declare(s) that they have no competing interests.

Acknowledgement

I would like to thank my former head of department, Christian Winkler, for mentioning Flickr as a potential source of data to me and for supporting my work in exploring new sources of data and methods of analysis.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jcmds.2025.100122.

Availability of data and material

This R code used for this project is stored in an open access repository on Github (https://github.com/antonlich/flickr_longdistance_mobility.git). Further usage and development is more than welcome.

References

[1] Jusup M, Holme P, Kanazawa K, Takayasu M, Romić I, Wang Z, Geček S, Lipić T, Podobnik B, Wang L, et al. Social physics. Phys Rep 2022;948:1–148. http://dx.doi.org/10.1016/j.physrep.2021.10.005.
[2] Christensen L. Considerations on how to conduct a survey about long-distance travel with reduced memory effect. Transp Res Procedia 2024;76:534–44. http://dx.doi.org/10.1016/j.trpro.2023.12.074.
[3] Schulz A, Magdolen M, Chlond B, Kuhnimhof T, Köhler K. Identifying and understanding long-distance travel demand by combining official transport statistics and survey data. Transp Res Procedia 2024;76:458–69. http://dx.doi.org/10.1016/j.trpro.2023.12.068.
[4] Rich J, Mabit SL. A long-distance travel demand model for Europe. Eur J Transp Infrastruct Res 2012;12. http://dx.doi.org/10.18757/ejtir.2012.12.1.2946.
[5] Zhang J, Hayashi Y. Research frontier of COVID-19 and passenger transport: A focus on policymaking. Transp Policy 2022;119:78–88. http://dx.doi.org/10.1016/j.tranpol.2022.02.014.
[6] Kellermann R, Conde DSivizaca, Rößler D, Kliewer N, Dienel H-L. Mobility in pandemic times: Exploring changes and long-term effects of COVID-19 on urban mobility behavior. Transp Res Interdiscip Perspect 2022;15:100668. http://dx.doi.org/10.1016/j.trip.2022.100668.
[7] De-Toledo KPawluk, O'Hern S, Koppel S. I saw a fox in Melbourne CBD: Urban travel behaviour changes during COVID-19 and beyond. Transp Res Part F: Traffic Psychol Behav 2024;100:115–32. http://dx.doi.org/10.1016/j.trf.2023.11.010.
[8] Clark DE, McGibany JM, Myers A. The effects of 9/11 on the airline travel industry. In: Morgan MJ, editor. The impact of 9/11 on business and economics: the business of terror. New York: Palgrave Macmillan US; 2009, p. 75–86.
[9] Van Cranenburgh S, Chorus C, Van Wee B. Substantial changes and their impact on mobility: A typology and an overview of the literature. Transp Rev 2012;32:569–97. http://dx.doi.org/10.1080/01441647.2012.706836.
[10] Parkes SD, Jopson A, Marsden G. Understanding travel behaviour change during mega-events: Lessons from the London 2012 games. Transp Res Part A: Policy Pr 2016;92:104–19. http://dx.doi.org/10.1016/j.tra.2016.07.006.
[11] Greg M, Jillian A, Chatterton T, Docherty I, Faulconbridge J, Murray L, Roby H, Shires J. Studying disruptive events: Innovations in behaviour, opportunities for lower carbon transport policy? Transp Policy 2020;94:89–101. http://dx.doi.org/10.1016/j.tranpol.2020.04.008.
[12] Chica M, Perc M, Santos FC. Success-driven opinion formation determines social tensions. IScience 2024;27:109254. http://dx.doi.org/10.1016/j.isci.2024.109254.
[13] Rashidi TH, Abbasi A, Maghrebi M, Hasan S, Waller TS. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. Transp Res Part C: Emerg Technol 2017;75:197–211. http://dx.doi.org/10.1016/j.trc.2016.12.008.
[14] van Eggermond M, Chen H, Erath A, Cebrian M. Investigating the potential of social network data for transport demand models. 2015.
[15] Yao W, Qian S. From Twitter to traffic predictor: Next-day morning traffic prediction using social media data. Transp Res Part C: Emerg Technol 2021;124:102938. http://dx.doi.org/10.1016/j.trc.2020.102938.
[16] Gu Y, Qian Z, Chen F. From Twitter to detector: Real-time traffic incident detection using social media data. Transp Res Part C: Emerg Technol 2016;67:321–42. http://dx.doi.org/10.1016/j.trc.2016.02.011.
[17] Xu S, Li S, Wen R. Sensing and detecting traffic events using geosocial media data: A review. Comput Environ Urban Syst 2018;72:146–60. http://dx.doi.org/10.1016/j.compenvurbsys.2018.06.006.
[18] Vaziri F, Nanni M, Matwin S, Pedreschi D. Discovering tourist attractions of cities using flickr and OpenStreetMap data. In: Proceedings of the advances in tourism, technology and smart systems, Singapore, 2020. 2020, p. 231–41.
[19] Spyrou E, Sofianos I, Mylonas P.

[20] Domènech A, Mohino I, Moya-Gómez B. Using Flickr geotagged photos to estimate visitor trajectories in world heritage cities. Int J Geo- Inf 2020;9:646. http://dx.doi.org/10.3390/ijgi9110646.
[21] Yuan Y, Medel M. Characterizing international travel behavior from geotagged photos: A case study of flickr. PloS One 2016;11:e0154885. http://dx.doi.org/10.1371/journal.pone.0154885.
[22] Barchiesi D, Moat HS, Alis C, Bishop S, Preis T. Quantifying international travel flows using Flickr. Plos One 2015;10:e0128470. http://dx.doi.org/10.1371/journal.pone.0128470.
[23] Preis T, Botta F, Moat H. Sensing global tourism numbers with millions of publicly shared online photographs. Environ Plan A 2020;52:471–7. http://dx.doi.org/10.1177/0308518X19872772.
[24] Beiró M, Panisson A, Tizzoni M, Cattuto C. Predicting human mobility through the assimilation of social media traces into mobility models. 2016, http://dx.doi.org/10.1140/epjds/s13688-016-0092-2, 5, 2193-1127.
[25] Yan Y, Schultz M, Zipf A. An exploratory analysis of usability of flickr tags for land use/land cover attribution. Geo- Spat Inf Sci 2019;22:21–2. http://dx.doi.org/10.1080/10095020.2018.1560044.
[26] Daymond T, Andrew ME, Kobryn HT. Crowdsourcing social values data: Flickr and public participation GIS provide different perspectives of ecosystem services in a remote coastal region. Ecosyst Serv 2023;64:101566. http://dx.doi.org/10.1016/j.ecoser.2023.101566.
[27] Ciesielski M, Stereńczak K. Using Flickr data and selected environmental characteristics to analyse the temporal and spatial distribution of activities in forest areas. For Policy Econ 2021;129:102509. http://dx.doi.org/10.1016/j.forpol.2021.102509.
[28] Tenkanen H, Minin EDi, Heikinheimo V, Hausmann A, Herbst M, Kajala L, Toivonen T. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Sci Rep 2017;7:17615. http://dx.doi.org/10.1038/s41598-017-18007-4.
[29] Van Acker V, Kessels R, Palhazi Cuervo D, Lannoo S, Witlox F. Preferences for long-distance coach transport: Evidence from a discrete choice experiment. Transp Res Part A: Policy Pr 2020;132:759–79. http://dx.doi.org/10.1016/j.tra.2019.11.028.
[30] Frei A, Kuhnimhof TG, Axhausen KW. Long-distance travel in europe today: Experiences with a new survey. In: Proceedings of the transportation research board 89th annual meeting, washington DC, United states. 2010.
[31] von Behren S, Puhe M, Magdolen M. Social aspects of long-distance travel - a study of two survey designs. Transp Res Procedia 2024;76:361–72. http://dx.doi.org/10.1016/j.trpro.2023.12.061.