# Chapter 2. Literature Review

## 2.1. Mobility studies based on purpose

Less purpose classification studies

Although a wealth of literature exists regarding the classification of transport mode derived from GPS traces, investigation into the transport purpose in geolocated movement data has received far less attention (Yazdizadeh *et al.*, 2019). One reasoning behind this is that these studies require additional information from the user i.e. it becomes impossible to determine the purpose of a trip purely based on GPS, as opposed to transport mode from GPS, which is just challenging (Gong *et al.*, 2014). Correspondingly, Yazdizadeh *et al.* (2019) finds that the accuracy of transport mode-classification model is often much higher than for transport purpose-classification in the literature.

Gong *et al.* (2014) -> ﻿“trip purpose could not be derived from the GPS raw data directly without further data processing or other assisted information”

Of the ones that exist 3 types. Present table and say about accuracy disparity

Of the existing literature, Gong *et al.* (2014) characterises trip purpose classification models into three categories: rule-based, probabilistic and machine learning models. In the last few years, studies have begun to employ models in the trip purpose classification that are more abundant i.e. Gong *et al.* (2018) finds that the use of machine learning methods and especially random forest and other classification tree models to have become more popular and most effective.

However, despite widespread convergence/similarity of the types of models used for this classification, a range of explanatory identifiers have been used within these models to help their classification accuracy (ref). As shown in **Table 2.1**, a variety of temporal and spatial predictor variables have been used to try to structure both space and time within the models. This table shows that between 43-96.6% accuracy have been found within the models, but it must be noted that the sample sizes varied dramatically.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Author(s)* | *Predictor variables* | *Location and year* | *Respondents (number of trips)* | *Overall classification accuracy* |
| *Rule-Based Methods* | | | | |
| Bohte & Matt (2009) | POI; Personal Locations | Netherlands, 2007 | ﻿1104 (﻿33,686 trips) | 43% |
| Alsger *et al.* (2018) | Activity Duration; Land Use; Temporal Features; Trip Frequency | ﻿Queensland, Australia,  2009-2012 | (﻿65,000 trips) | 78% |
| Probabilistic Methods: Multinomial Logit Models | | | | |
| Oliveria *et al.* (2014) | Duration; Mode; Land use; Personal locations proximity | Georgia, USA 2011 | 1,354 (﻿10,512 trips) | 70% |
| *Machine Learning Methods: Artificial Neural Networks* | | | | |
| Xiao *et al.* (2016) | Land use; POI | Shanghai, China,  ﻿2013-2015. | 321 (﻿7,039 trips) | 96.5% |
| *Machine Learning Methods: Random Forest and Decision Tree Models* | | | | |
| Montini *et al.* (2014) | Land Use; Personal Locations; Socio-demographics; Temporal Features | Zurich, Switzerland, 2012 | 156 (6,938) | 80% |
| Kim *et al.* (2015) | POI; Socio-demographics | Singapore, 2013 | 793 (﻿7,856 trips) | 75.5% |
| Ermagun *et al.* (2017) | POI;  Socio-demographics; Temporal Features;  Travel Mode | Minnesota & Iowa, USA, 2010-2012 | (﻿58,503 trips) | 64% |
| Yadizadeh *et al.* (2019) | Socio-demographics; Personal Locations; Temporal Features; Land Use; Foursquare check-ins | Montreal, Canada, 2016 | ﻿6845 (﻿131,777 trips) | 71% |
| *Machine Learning Methods: Support Vector Machines* | | | | |
| Zhu *et al*. (2014) | Foursquare check-ins; Socio-demographics; Temporal Info | Washington State, USA Spring 2014 | ﻿10,372 (﻿87,600 trips) | 75% |

**Table 2.1** Overview of classification models used in the literature to predict trip purpose.

About Models:

It must be noted that the multi-label vs binary classification

Ermagun *et al.* (2017) selection of binary variables

Kim *et al.* (2015) [a study looking at best performance of RF] ensemble better than individual purpose. “﻿The model using more training data shows better classification performance.”

Oliveria *et al.* (2015) -> ﻿“calculates the probability of each trip purpose based on household, temporal, etc…”

Zhu et al -> multi-class one vs all for each

Montini et al. and other RF -> ensemble method

Xiao *et al.* (2016) -> “﻿Consequently, 340 samples are randomly selected for each trip pur- pose, and the remaining samples constitute the test dataset.” “multi-calss i.e. 6 output neurons”

Space and Time predictors

Notably, each of the studies listed is for a different number of trips at a different time and location. Gong *et al.* (2018) find seasonality to particularly affect the model accuracy when predicting purpose as people tend to change travel patterns to account for weather.

Aslger *et al.* (2018) looks to divide temporal and spatial components of the data before modelling and suggest the importance of including both.

Temporal Features including time of day, day of week, rush hour etc.

Montini *et al.* (2014) Clustering of locations to reduce data and end points

It has been a challenge to include underlying socio-demographics and POI with GPS route data (use Shi *et al.*, 2018 as a reference that it ignore the uncertainty). Montini et al. (2014) find socio-demographic factors to be less important. Oppositely, ref finds that age and etc.

Although there is widespread use of contextual information within models such as including distances to points of interest (POI) have been employed in the literature to improve the prediction accuracy. Meng *et al.* (2019) go one further by looking at the popularity of POIs-> another bit of interesting contextual information]

(Montini et al. 2014). Most important features are transport modes used before and after the activity”

In terms of transport mode, however, socio-demographics have been found to be important…

Bantis & Haworth (2017) socio-demographics and how you travel. Environmental and social factors affect the way you travel.

Xie *et al.* (2016) -> people travel in distinct patterns broadly based on socio-economic groups, also other classes i.e. returners and explorers (different levels of variance of travel)

Inferring employment status (Zhang & Cheng, n.d.)

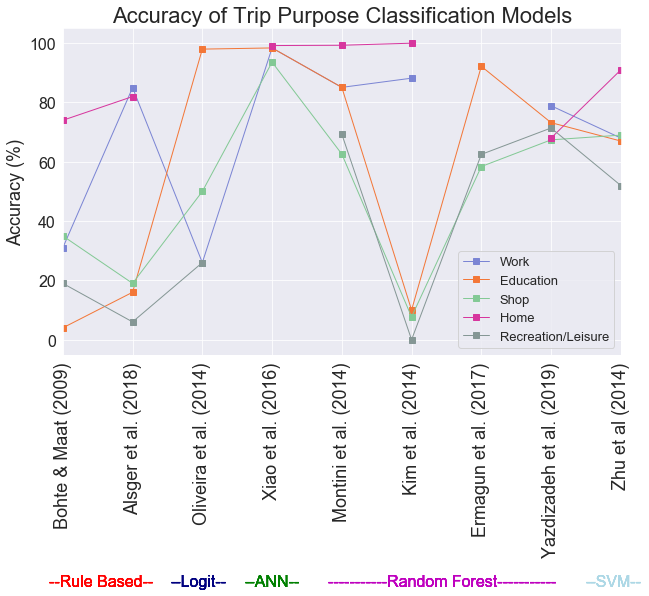
Xie *et al.* (2016) studied mobility in Beijing, finding distinct trends relating to socio-demographics (i.e. younger and employed moving further than older and unemployed ­– even weather has impact). Further Zhang & Cheng (n.d.) …

Indeed, finances can restrict a persons travel directly and indirectly (maybe old geography reference)

Semanjski *et al* (2017) use land use to indicate accuracy of classification (more accurate in rural areas)

Disparity between purpose

There is also, a significant disparity in the accuracy of the classification models based on individual purposes (ref;ref;ref). **Figure 2.1** shows that the broad classification work and education in most studies has been found to be most easily identifiable in models. It has been suggested that this is as work and education tend to be the most temporally and spatially structured (ref; ref). [As opposed to leisure and health, etc.]. individuals’ mobility is found to be highly regular (Lin & Hsu, 2014)



**Figure 2.1** Comparison of trip purpose classification model accuracy within the literature (ANN=Artificial Neural Network; SVM=Support Vector Machine)

Regarding how data is collected for these surveys.

The is also significant problem with the cross-comparibility of the data generated for these studies as it is typically collected using targeted travel surveys and proxy information i.e. grouped by areas and as such remains specific to the region and time-period the data were collected in (ref;ref). It is only in recent years, with the explosion of VGI collected from smartphones and mobile app surveys that has meant increasingly research using data these sources has become more wide spread. The quality of travel surveys is not high.

Kim *et al.* (2015) Smartphones and interactive web interfaces have emerged as an attractive alternative to conventional travel survey

Sentences:

The casual links have been harder to prove in the literature

Owing to the…

Of note to this study is Yazdizadeh *et al.* (2019) MTL Trajet 2016

To Add:

* Something about space and time profiles
* Something about different number of purposes
* One major limitation to the classification is that we do not know about its predictor variables (i.e. mode)

## 2.2 Research Using Volunteered Geographic Information

About VGI in the literature

The general use of VGI in studies has been

‘People as sensors’. Li *et al.* (2016) distinguish between participatory (conscious inclusion of their data i.e. like MTL Trajet) and opportunistic (unconscious) forms of VGI.

Can see people can experience space-time events (i.e. Traffic Jam, Crowds, THINK OF MORE etc.). Both fast and slow dynamics/flows have an impact on changes in a variable across space, accounting for these is essential in modelling [i.e. differencing to remove long-term trend] (Batty, 2013).

Tu *et al.* (2017) look at activities relating to mobile phone GPS tracts using social media data, check-ins and a hidden Markov model

Use

Governments pushing for greater capacity of transport networks rather than efficiency (Attard *et al.*, 2016).

high quality traveller information for public transport is undoubtedly crucial for Government’s transport policies (Lyons & Harman, 2002)

Issues

problems of representativeness in VGI (Li *et al.*, 2016; ref; ref)

(Attard et al., 2016 referencing Lyons & Harman, 2002). → 1. There are issues with trust over the information provided, 2. Lifestyle changes are opportunities for travel behaviour change, 3a. People have very poor judgement of cost and time when travelling by car with control over their journeys being seen as important. 3b. Public transport in contrast, is seen as difficult as information is sought from unfamiliar and uncertain sources.

VGI requires rethinking of geographical concepts (Elwood *et al.*, 2012).

Many studies have shown that voluntary contributions by individuals follow a frequency distribution with a long tail, with a few individuals making large numbers of contributions (Goodchild & Li, 2012)

*City research*

Understanding mobility through mobile phone has kicked off (Zhao *et al.*, 2019)

Insight into which activities occur on which days and times (similar to Zhang & Cheng, 2019). -> (lead onto Batty, 2013)

*On Neighbourhood effect ->* People often traverse neighbourhoods and boundaries throughout one day (Kwan, 2018)

Research carried out by Zhang & Cheng (2019) discover expected difference in the profiles of people travelling within London based on their employment status. In general, finding regularity in full-time transport patterns compared with those who are un-employed. While, this information is of use to transport authorities, there is still a lack of investigation into more of the local impacts of transport.

Other:

* Using twitter and sentiment analysis (similar to purpose)

### 3.2.1 MTL Trajet & Similar Survey Projects

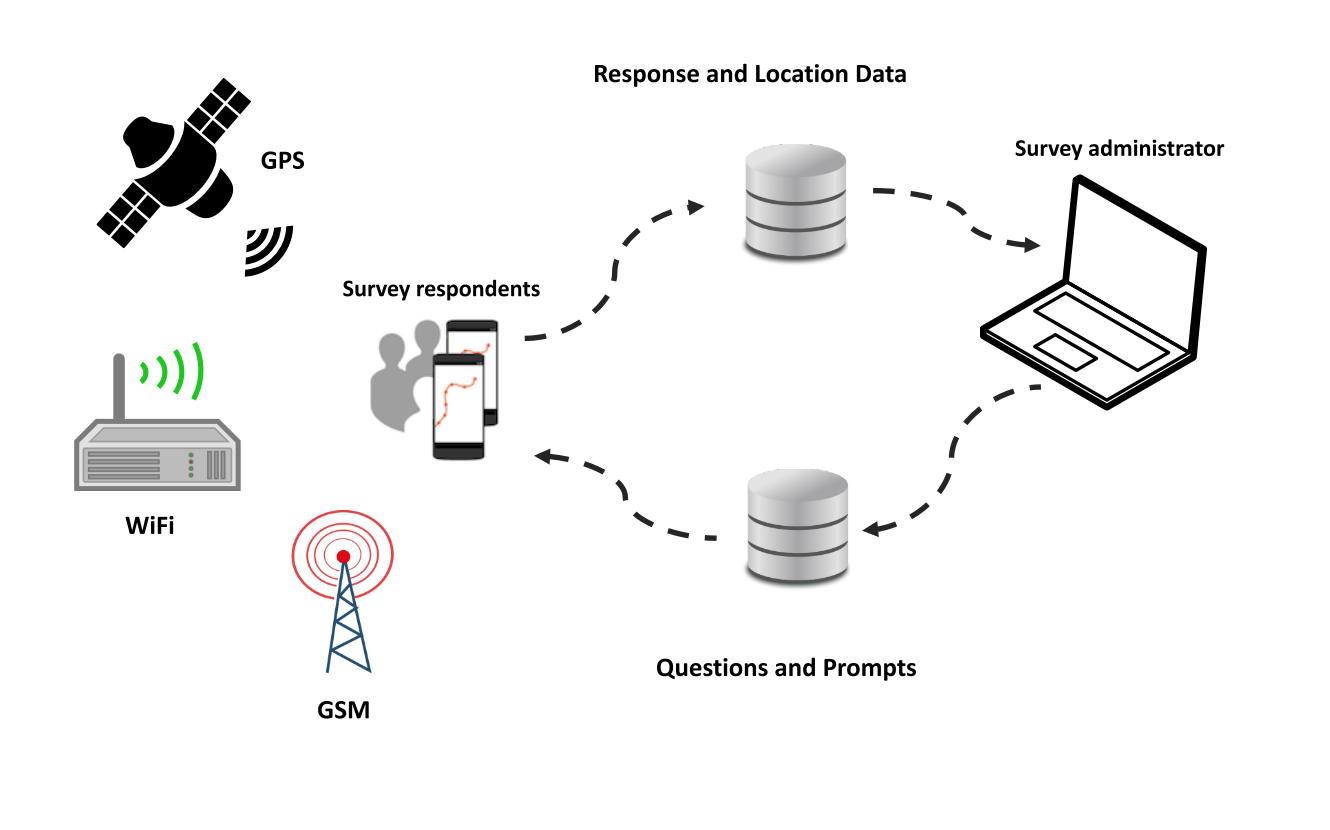
Data collection [or survey data] driven by data availability or convenience of data collection rather than by domain knowledge, theory, or insight into the process(es) of interest. (An *et al.*, 2015). Many previous studies have proposed and validated in limited geographical areas, such as campuses (Jahromi *et al.*, 2016)

previous studies have mainly concerned themselves on one aspect of trips (e.g. mode) at a time [but new technology (and the Itinerum platform) has given researchers new opportunity to study purpose] (Yazdizadeh *et al.*, 2019)

Fallah-Shorshani (2018) uses MTL to look at pollution exposure

DataMobile Smartphone Travel Survey example from 2014 carried out by Concordia University in Montreal (Patterson & Fitzsimmons) 10 November - 5 December 2014. Close to 900 people participated in the survey

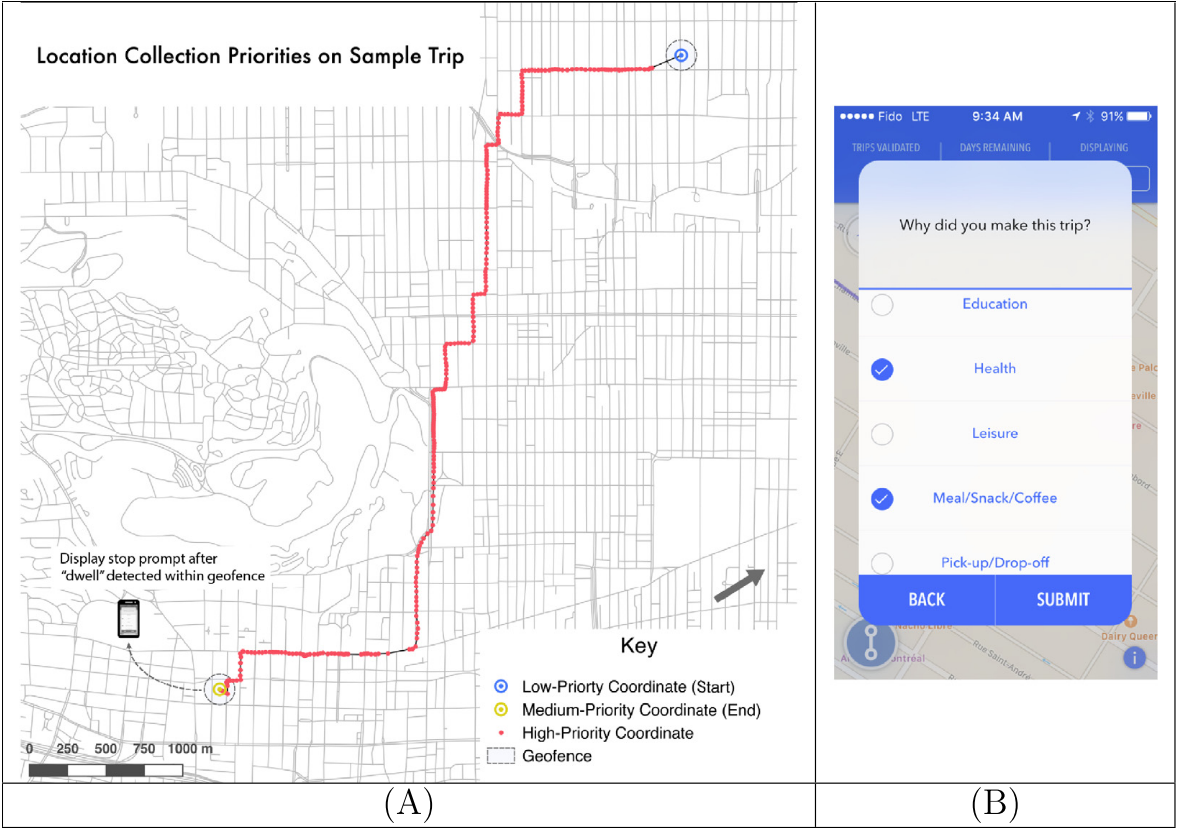
Itinerum Platform



**Figure 2.2** Framework for survey apps developed using the Itinerum platform (Source: Patterson & Fitzsimmons, 2017).

[About the app]:

“﻿Central to the application is a long-running location logging service that requests and monitors changes to location in a battery-efficient way” (Patterson *et al.*, 2019). ‘﻿If the device leaves the geofence at any time, it switches to high-priority mode to request a new point and establish the new ‘‘rolling’’ geofence (see Fig. 4).” (Patterson *et al.*, 2019).



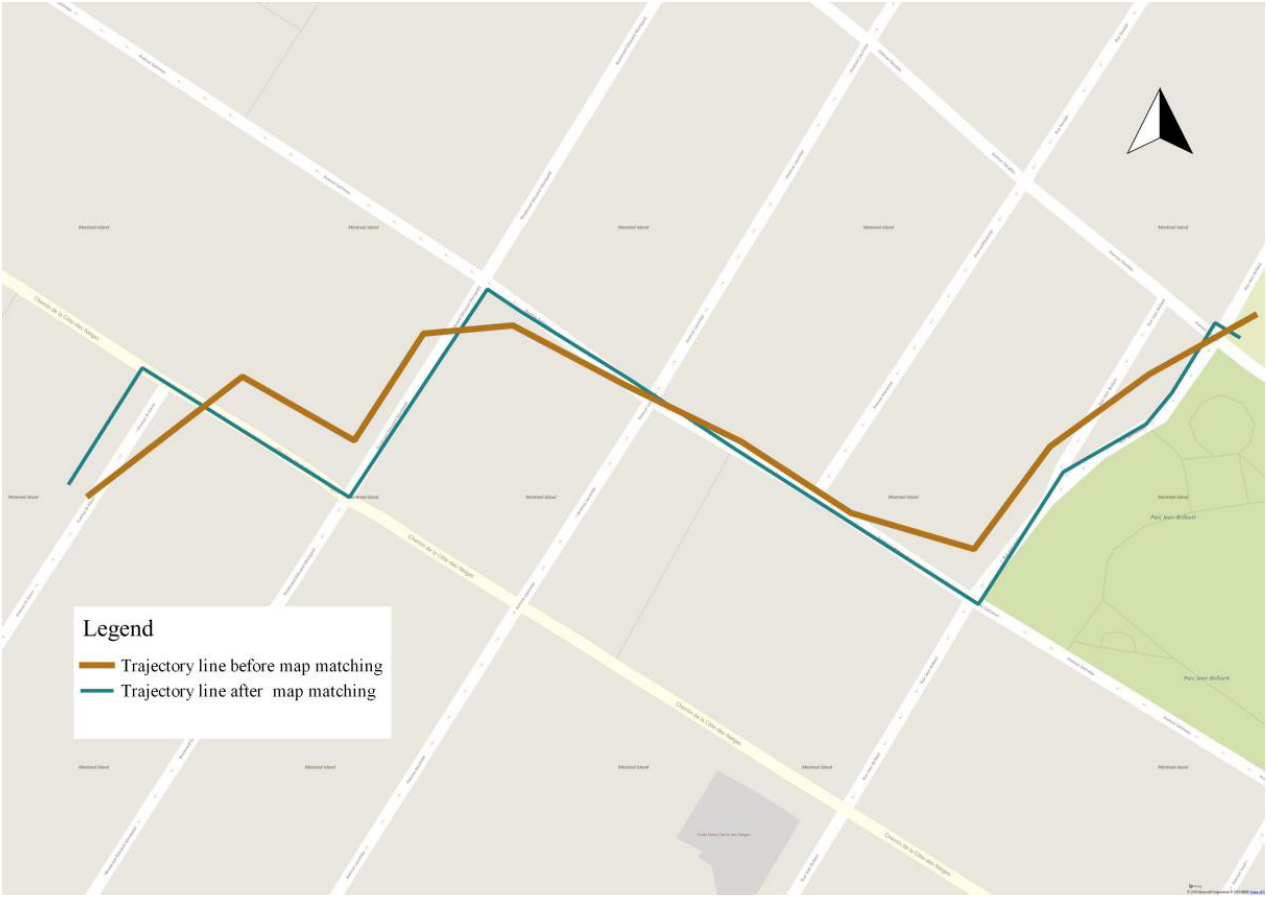
**Figure 2.3** Example of GPS trace with location collection priorities within an Itinerum platform app (A); Example of the on screen prompt after an Itinerum platform app stops recording movement (B) (Source: Patterson *et al.*, 2019).

Aim of the MTL Trajet is to better understand travel (MTL Trajet, 2017)

MTL Trajet is a supplement to the O-D survey in Montreal – also will get younger people involved (MTL Trajet, 2019)

MTL Trajet (2017) results were used to discover how many people went on trips and couldn’t find a parking space (moving round and round in circles downtown). Also, for the redevelopment and planning of the needs of the road network – the study is being used so that the redevelopment meets the needs of people (drivers, cyclists, etc.). Also, in alternative route analysis, where there are road works. Finally, the study helped with identifying common bike routes. Also, survey helped with identifying safety on roads .

The 2018 edition of the MTL Trajet study will take place from September 24 to October 28, 2018 (MTL Trajet, 2019).



**Figure 2.4** Example of ‘Map Matching’done by the Open Source Routing Machine when processing the raw GPS trace from user devices (Source: Hamouni, 2018).

*Montreal:*

Chevalier *et al.* (2018) the island of Montreal (Île de Montréal) largest of the Hochelaga Archipelago near the confluence of two rivers.

WPR (2019) -> 1.75 million people as of 2016. The island in total has an estimated population of 1.95 million. “The city proper has a population density of 4,517 people per square kilometer (11,701 residents per square mile)”. “By 2030, the Greater Montreal Area is expected to grow to 5.275 million”

The city of Montreal, located within the Greater Montreal region is the largest city in Quebec with a population size of … . It is of particular interest to city transport research due to its unique road and public transport networks. There are a total of

[Map of the Island Montreal within the Greater Montreal Region and then Quebec]

Eluru *et al.* (2012) survey to look at public transport and how people move around Montreal. Trying to encourage it

## 2.3. Metrics and Space-time Investigation

Including space and time in models is a challenging concept…

*\*\*Need part about representing space in models*

machine learning is the unable to directly handle spatio-temporal structure (Cheng *et al.*, 2011)

*\*\*Need part about representing time and space-time in models*

*Transport*

Transportation is a classic social dilemma where individually rational behaviour (being mobile) leads to collectively irrational outcomes such as congestion (Miller, 2013)

\* Nevertheless, some trips will always be car-dependent, due to their nature or their spatio-temporal location (Sioui *et al.*, 2012)

The highlights of the Origin-Destination Survey 2018 will be released in the fall of 2019.

Twitter for geographic understanding of purpose of movement -> “﻿extracting information from social media to track and analyse human movements” (Rashidi *et al*., 2017)

*Network Research*

The understanding of human mobility in an urban space has become crucial to optimize the network management (Jahromi *et al.*, 2016)

Size of space-time neighbourhoods change (Cheng *et al.*, 2014)

*Space-Time*

ST-KDE and space-time decomposition to compute space-time methodologies (Hohl *et al.*, 2016).

Spatio-temporal modeling has always been of interest to researchers in geographical information science (GIS) (Ren *et al.*, 2019).

CNN was not optimal for modeling the spatial patterns of the road because the authors did not consider the road network topology (i.e. direction) (Ren *et al.*, 2019 in their literature review)

[Problem with time sampling] -> ﻿different patterns can be observed for different temporal resolutions (Zhao *et al.*, 2019)

STARIMA and ANN (Cheng & Wang, 2011)

Transport Forecasting and modelling (Yue & Yeh, 2008; Cheng & Wang, 2011)

network and topological models not well suited for handling geospatial big data (Li *et al.*, 2016)

Paradigm shift (Brunsdon, 2015)

Space–time analysis seeks to understand when and where (and sometimes why) things occur. (An *et al.*, 2015).

there is not a close coupling between big data and space-time methods used to analyse them (An *et al.*, 2015)

Spatial Weights (Anselin & Rey, 2014)

Determining threshold of time and space (Adepeju & Evans, 2018)

MTUP (Cheng & Adepeju, 2014)

(Ren *et al.*, 2019)\* Need to account for local trends with models, a lot of papers use models that don’t consider local, but still do space-time analytics (i.e. with locally-weighted CNN layers on a network)

PYSAL

*Lead onto class identification in transport*

Significant class-imbalance exists in the MTL Trajet data

NY Taxi-cab movements OR Uber-movements

*Big Data*

[On big data] However, they are often collected on a fairly ad-hoc basis when compared with traditional data sources, and usually must be repurposed to fulfil research objectives (Li *et al.*, 2016)

[On visualising big data maps] New computational and technical paradigms for cartography are accompanying the rise of geospatial big data (Robinson *et al.*, 2017)

Ultimately, the value of SBD (Spatial big data analysis) relies on uncertainty handling (Shi *et al.*, 2018)

Mathematical models being employed without regard of space, often including problems that are inherently tied to spatial considerations (O'Sullivan & Manson, 2015)

*ABM:*

ABM a great tool but ABMs could be very data demanding and sometimes too complex without offering much additional insight (An *et al.*, 2015).

*Machine Learning:*

methodologies originally developed to analyse small data and are not necessarily equipped to tackle some of the distinctive features of newer sources (Gorman, 2013; Arribas-Bel, 2017).

However, methods to assess quality assurance of the geospatial data still relevant on big geospatial data as they still describe the same processes (Li *et al.*, 2016)

Big geospatial data visualisations do not always scale well – because they can become messy (Li *et al.*, 2016)

Li *et al.* (2016) challenges in dealing with big geospatial data, reviews if those traditional methods still useful for data

﻿ML methods are generally effective in tackling nonlinearity in spatial data (Li *et al.,* 2016)

Spatial statistics is well suited to handle big data. It offers capabilities to summarize the data, and express measures of variation and uncertainty. (Li *et al.*, 2016)

Machine Learning are often applied in detecting transportation mode (Gong *et al.*, 2014)

[With big geospatial data] validity or trust is traded for the velocity of information production (Li *et al.*, 2016)

Space-Time CNN -> CNN-based models employ a grid map to represent spatial data which is unsuitable for road- network-based data. (Ren *et al.*, 2019)\*