# Chapter 2. Literature Review

## 2.1. Mobility studies based on purpose

Although there is a wealth of literature on mode-classification there exists a lack of pertinent literature to the study of how people travel for around a city for certain activities. Arguably this owes to a lack of data set available to examine this topic.

Relatively steady stream of research

Gong *et al.* (2014) reviews the rule-based, probabilistic and machine learning techniques used

Wolf *et al.* (2001) -> 13 respondents with GPS loggers

Bohte & Matt (2009) -> within 50 m of a POI that POI will be assigned to the trip. 13 categories of purpose

Oliveria *et al* (2014) -> 12 categories of purpose. Decision Tree

Kim *et al.* (2016) -> 16 categories of purpose. Decision Tree. Age, gender, POI

A number of studies employ contextual data (POI)

Xiao *et al* (2016) -> activity duration important. 200 m buffer around POI. Less accurate about shopping

Ermugun *et al.* (2017) -> ﻿real-time trip purpose prediction

Montini *et al.* (2014) -> 8 different purposes

Zhu *et al.* (2014) -> ﻿linear kernel SVM. “﻿we illustrate the importance of urban space characterization and temporal features for inferring travel purposes”

Alsger *et al.* (2019) -> ﻿The results show an overall 67% correct inference after applying spatial attributes, but the correct inference increases to 78% after applying temporal attributes. Work and Home 92 and 96%. More accurate with space and time

﻿clustering cannot capture the full complexity of travel patterns (Kim et al., 2014).\

Gong *et al.* (2018) ﻿“the RF and classification tree methods have already proven to

be better than some of the other supervised machine learning methods for the identification of trip purpose”. This study trains on seasons and compares the importance of season to purpose [\*Seasonality]. “﻿Training and test sets should not be populated with data from distinct seasons. 2) Weather features had a negative effect on the identification accuracy while GIS fea- tures had a positive effect”

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

An overview of classification models used in the literature along with their accuracyis shown in **Table 2.1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Author(s)* | *Predictor variables* | *Location and dates* | *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% (92% and 96% for work and home) |
| *Machine Learning Methods: Artificial Neural Networks* | | | | |
| Xiao *et al.* (2016) | Land use; POI | Shanghai, China,  ﻿2013-2015. | 321 (﻿7,039 trips) | 96.5% (99% for work and education) |
| *Machine Learning Methods: Random Forest Classification* | | | | |
| Kim *et al.* (2015) | POI; Socio-demographics | Singapore, 2013 | 793 (﻿7,856 trips) | 75.5% |
| Ermagun *et al.* (2017) | Nearby POI;  Socio-demographics; Temporal Features;  Travel Mode | Minesota & Iowa, USA, 2010-2012 | (﻿58,503 trips) | 64% (79% for Education and Shopping) |
| Gong *et al.* (2018) | POI; Socio-demographics; Temporal Features; Trip Duration; Weather | ﻿Hakodate City, Japan December 2012 and October 2013 | (﻿9981 trips) | n/a |
| Yadizadeh *et al.* (2019) | Socio-demographics; Personal Locations; Temporal Features; Land Use; Foursquare check-ins | Montreal, Canada (MTL Trajet 2016) between Oct-Nov 2016 | ﻿6845 (﻿131,777 trips) | 71% |
| *Machine Learning Methods: Support Vector Machines* | | | | |
| Lu *et al*. (2013) | Land Use; POI | Minesota , USA 2010 | (2238 trips) | 80.6% |
| 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.

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)

See Yazdizadeh *et al.* (2019) for references about activity/purpose detection [this study uses RF: ﻿71% for purpose, but 87% for mode and 81% travel iterinary]. - [for purpose pred using RF] → age and mode and distance between destination and individual’s home, later it is land use. Health and work better at predicting 87, 78 -> less good at shopping and returning home

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

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

\* Montini *et al.* (2014) -> “﻿The analysis was based on GPS tracks and accel- erometer data collected by 156 participants who took part in a 1-week travel survey in Switzerland that was completed in 2012”. “most Machine learning approaches are mostly decision trees”. \*\*most important features → mode, duration, distance to work and home, start time and cluster of occurence per day (Socio-demographic less important [link to other studies and mode classification studies]\*\* (Montini et al. 2014). Most important features are ﻿transport modes used before and after the activity

Machine learning -> Machine learning based approaches for activity recognition can automate some of these task (Kim *et al*. 2015)

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.

Meng *et al.* (2019) ﻿trajectories, POIs, and social media messages to infer purpose. Also POI popularity. [Example of inclusion in this dissertation: Meng *et al.* go one further by looking at the popularity of POIs-> another bit of interesting contextual information]

Space-time investigations in general

individuals’ mobility is found to be highly regular (Lin & Hsu, 2014)

[Write about] the challenging of overlaying the socio-economic data with routes (use Shi *et al.*, 2018 as a reference that it ignore the uncertainty)

Yazdizadeh *et al.*, 2019 -> study about forecasting mode and purpose

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)

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)

“Getting a deeper understanding of human mobility is a prerequisite for a broad range of possible studies on smart cities and related research areas”. (Xie *et al.*, 2016)

\*\*(Dubos-golain *et al.*, 2017) -> Results suggest that the variability in transit use is correlated with spatial location, weather and line purpose.

\*\*Problems with training and testing on different parts of the year -> ﻿However, the “feasibility and effects of choosing these data from dif- ferent periods of the year are still unknown” Gong *et al.* (2018)

*General Transport mode detection*

A significant amount of literature exists for transport-mode detection. Something which is of prime concern to companies utilising spatial information derived from GPS data. Determining transport mode through the use of deep-neural networks such as those with convolutional layers. These networks .

Dabiri & Heaslip (2018) raw GPS to mode

Bantis & Haworth (2017) socio-demographics and how you travel. Environmental and social factors affect the way you travel. Although data used in this study is not

Striving to include meaning to space (POI)

Bantis & Haworth (2017) overlay of GPS tracks and underlying LSOA socio-demographic information

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

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

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

Zhang *et al.* (2019) look at the ﻿relationship between passengers’ movement patterns and social-demographics by using smart card (SC) data with a household survey. Exploring] ‘how’ (including ‘when’ and ‘where’), ‘who’ and ‘why’ travel in public transit

Maybe classify mode and where people are going? (after Bantis & Haworth, 2017)

## 2.2 Research Using Volunteered Geographic Information

‘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 space-time events (i.e. Traffic Jam, Crowds, THINK OF MORE etc.).

Using twitter and sentiment analysis (similar to purpose)

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

(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).

Mobile phones as sensors

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)

Liu & Cheng (2018) conjoins socio-economic data to transit patterns to interpret behaviour

using temporal clusters from LDA and temporal words (Liu & Cheng, 2018)

Liu & Cheng (2018) Looks at who constitutes each temporal cluster (which socio-economic groups)

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

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

*City research*

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).

Understanding the interactions within complex system such as a city is a prerequisite for predicting changes within it… big data offers us an opportunity to study this (Cheng *et al.*, 2017)

Flows generate change immediately whereas the ultimate locational redistribution takes longer to work itself out. [In reality] this process of working out is implicit and the ultimate equilibrium that occurs is a product of both fast and slow processes with no explicit time scale. (Batty, 2013)

\* Jahromi *et al.* (2016) try to simulate GPS movement/mobility with purpose so that infer about interactions of people with a city and its services

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)

*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

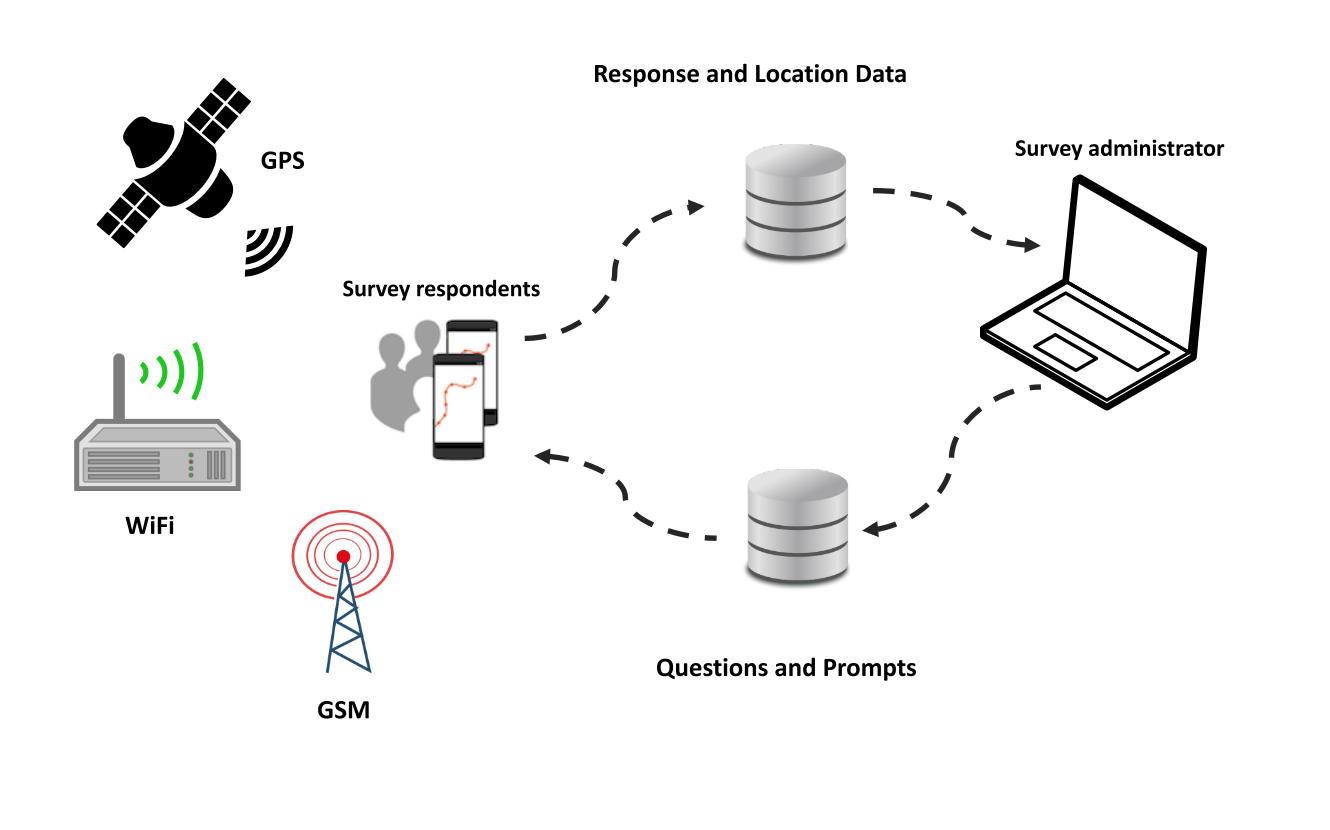
### 3.2.1 MTL Trajet & Similar Survey Projects

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)

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

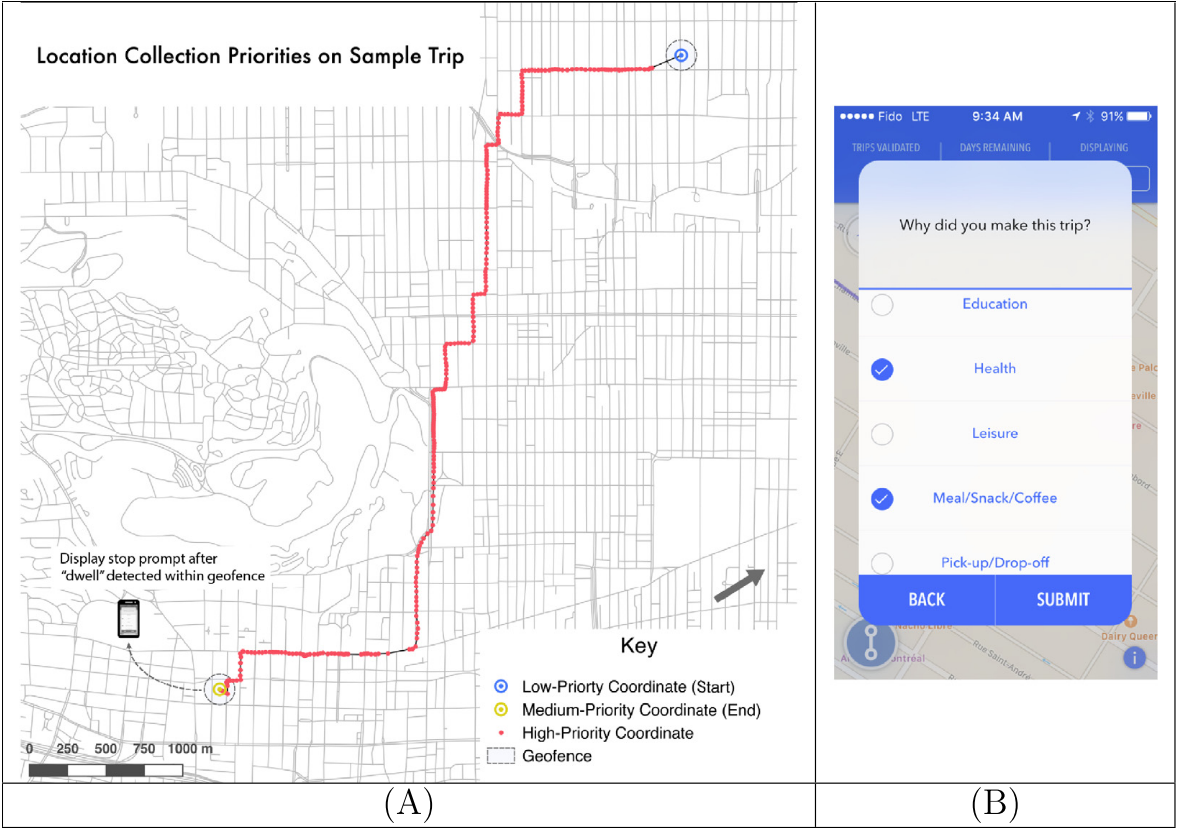
Itinerum Platform



**Figure 2.1** 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).



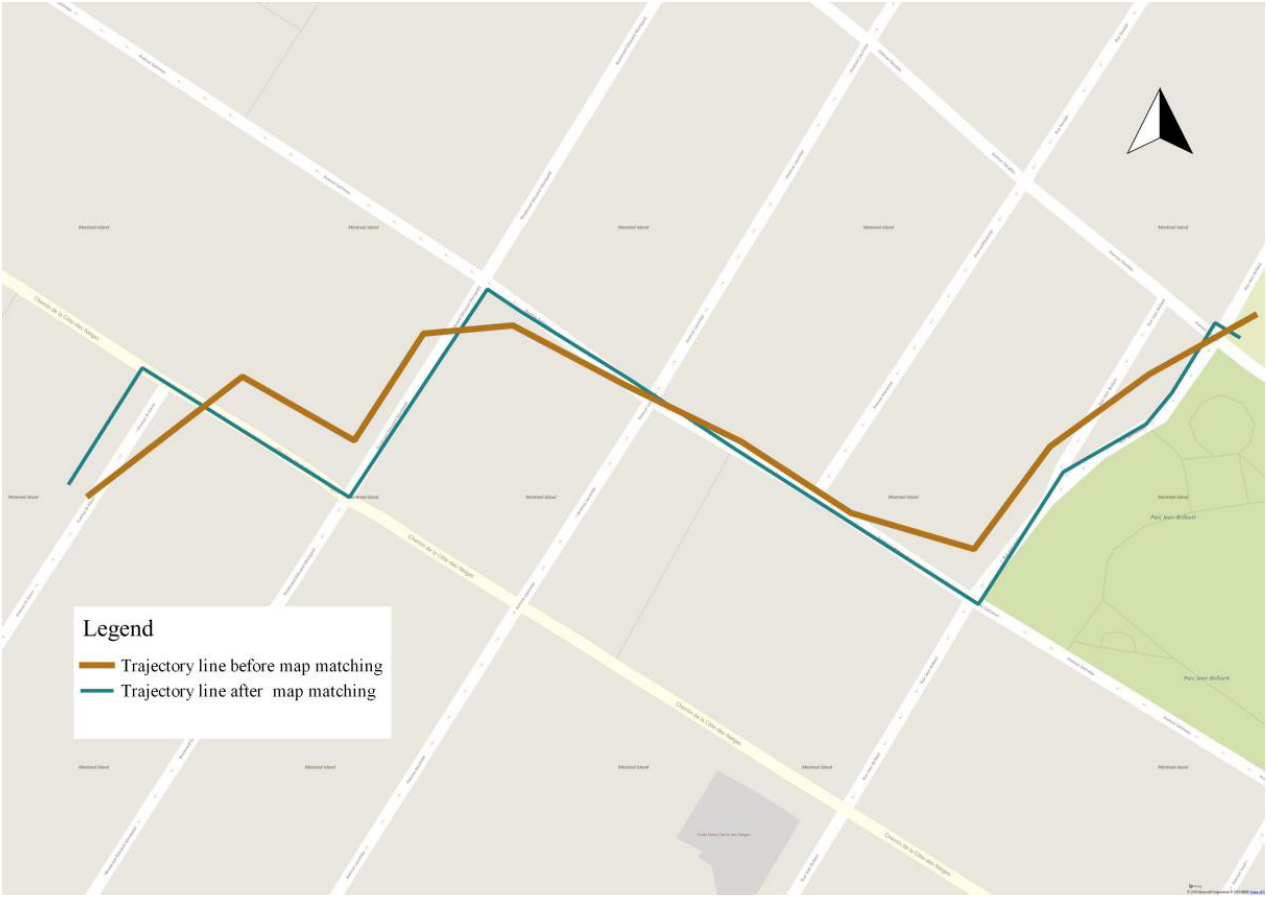
**Figure 2.2** Example of GPS trace with location collection priorities within an Itinerum platform app (A); Example of the on screen prompt after an Itinerum platform app stops recording movement (B) (Source: Patterson *et al.*, 2019).

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

## 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)\*