Alsger *et al.* (2018):

- “Different trip purposes show different sensitivities to the applied spatial and temporal attributes”

- “the results show 67% correct inference after applying spatial attributes, but the correct inference increases to 78% after applying temporal attributes”

An *et al.* (2015)

Problem is that one is often taken in expense of the other with spatial and temporal data An *et al.* (2015).

Data collection [for 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).

Arribas-Bel, 2017

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).

Arribas-Bel & Tranos, 2017:

Capturing space-time structures (Arribas-Bel & Tranos, 2017)

Attard *et al.*2016

Spatial and temporal information provided from this VGI can be integrated into city-level decision-making to help inform planning a variety of essential and non-essential services (Attard *et al.*, 2016). For example, if we knew that people tended to cycle to cafés during lunch breaks, policy could be implemented to introduce bike racks near the cafés.

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

Despite the potential to produce more VGI that can be used to generate insight into mobility within a city, there are many cities globally that have no form of formal research initiated within them (Attard *et al.*, 2016).

Attard *et al.* (2016) advocate the use of VGI in to study transport.

(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.

Badu-Marfo *et al.* 2019:

"The public release of such precise information, particularly location data such as place of residence, opens the risk of privacy violation" Badu-Marfo *et al.* (2019)

Bantis & Haworth, 2017:

“Combining such information [detailed GPS speed, acceleration, etc] with socio-demographic characteristics of travellers has the potential of offering a richer modelling framework that could facilitate better transportation mode detection using variables such as age and disability” [mention it has success in mode transport classification but not purpose] (Bantis & Haworth, 2017)

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

Batty, 2013:

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).

Bricka *et al.* 2015:

“GPS technology allows for comprehensive tracking and sharing of location and route information” (Bricka *et al.*, 2015)

Real time transport demand management (Bricka *et al.*, 2015)

Burini *et al.*, 2017

New nexus, between data science and VGI (Burini *et al.*, 2017)

Potentially to cross-analyze multisource *big data*” (Burini *et al.*, 2017)

Cheng *et al.* 2017

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)

Indeed, understanding the space and time structures in cities help us understand them better (Chen*g et al.* 2017).

“understanding the interactions within complex system such as a city is a prerequisite for predicting changes within it -> big data gives us an opportunity” Chen*g et al.* 2017

Big geographic data allows us to not only study the spatial and temporal interactions but also interactions of socio-economic factors [this is what this research aims to do] (Cheng *et al.*, 2017).

Eluru *et al.* 2012

Montreal has a relatively high share of transit ridership (for a North American city) Also multimodal public transport network (Eluru *et al.* 2012)

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

The city of Montreal itself is the largest city within Quebec and the second largest within Canada ()

Elwood *et al.*, 2012:

Where users can “opt to share” Geographic information (Elwood *et al.*, 2012)

Elwood *et al.* (2012) find that VGI gives us insight that other forms of data do not such as subjectivity which is tied to space (i.e. like purpose of trips).

Indeed, ﻿“users must opt in to share information on their activities” (Elwood *et al.*, 2012)

Difficult to falsify VGI (Elwood *et al*. 2012).

Ermugan *et al.* 2017:

The need for using an NL model rather than a standard multinomial logit model is tested by measuring inclusive values (IV) (McFadden, 1978).

A value of zero indicates complete correlation among unobserved components of the alternatives in a nest, while a value of one indicates absolute independence and makes the estimates similar to multinomial logit.

Flanagin & Metzger, 2008

More User generated content online (Flanagin & Metzger, 2008)

“several instances of VGI involve perceptual information that can only be reliably known and communicated by ‘‘locals’” (Flanagin & Metzger, 2008)

VGI credibiliity (Flanagin & Metzger, 2008)

Gong *et al.* 2014:

Detecting “accelerometer sensor, magnetometer sensor in Android-based smart-phones.” (Gong *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 (Feng and Timmermans, 2016) and travel mode (Shafique and Hato, 2015b; Zheng et al., 2008).”

Less hassle for users, (than traditional travel surveys), thus can reach larger audiences and for longer times in the background (Gong *et al.*, 2018)

Goodchild & Li, 2012:

OpenStreet Map contributions and crowdsourcing (Goodchild & Li, 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)

Goodchild, 2007

VGI, first described by (Goodchild, 2007),

“the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information” (Goodchild, 2007),

Whether that be geo-tags, georeferences, GPS, (Goodchild, 2007).

“humans as sensors” (Goodchild, 2007, pp. 218)

“The term citizen science is often used to describe communities or networks of citizens who act as observers in some domain of science” (Goodchild, 2007)

“VGI is defined as the subset of user- generated content (UGC) with a geographic reference (Goodchild 2007).”

Goodchild, 2013:

Lack of quality control (Goodchild, 2013)

“Virtually Impossible to create a representive sample in geographic space (Goodchild, 2013)

Hecht & Stephens, 2014:

VGI can be biased towards cities (Hecht & Stephens, 2014)

common VGI, ﻿Twitter, Flickr, and Foursquare (Hecht & Stephens, 2014)

“People ﻿in rural areas tend to use technology differently than people who live in cities.” (Hetch & Stephens, 2014). “﻿Foursquare check-ins are not public by default, but can be shared widely if a user connects her/his account to Twitter” (Hetch & Stephens, 2014)

Jahromi *et al.*, 2016

smart-phones being able to better record similar mobility behaviour as their carriers (Jahromi *et al.*, 2016).

Li *et al.* 2016:

Li *et al.* (2016) distinguish between two types of VGI participatory (conscious inclusion of their data) and opportunistic (unconscious) forms of VGI.

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

There are a number of drawback of studies using VGI,

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

Despite this, Li *et al.* (2016) find that use of VGI often “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)

VGI requires rethinking of geographical concepts (Elwood *et al.*, 2012). [With big geospatial data] validity or trust is traded for the velocity of information production (Li *et al.*, 2016)

Lin & Hsu, 2014:

Most “﻿human mobility behaviors follow a simple and reproducible pattern [in GPS].” Lin & Hsu (2014) -> go on about predictability and hence mode classification and hence opportunity

Jahromi *et al.*, 2016

Many previous studies have proposed and validated in limited geographical areas, such as campuses (Jahromi *et al.*, 2016)

Kim *et al.*, 2015:

Especially as the quality and breadth of travel surveys is not high (Kim *et al.*, 2015)

[VGI has huge potential for trip purpose classification as] Data for trip purpose has typically been collected using travel surveys which tend to emphasise behavioural patterns of people from certain socio-demographics and fail to be representative of the population as a whole (Kim *et al.*, 2015).

Data for trip purpose has typically been collected using travel surveys which tend to emphasise behavioural patterns of people from certain socio-demographics and fail to be representative of the population as a whole (Kim *et al.*, 2015).

Meng *et al.* 2019:

“In practice, people’s trip purposes are very important in understanding travel behaviors and estimating travel demands” (Meng *et al.*, 2019)

Miller & Goodchild 2014:

We can capture “ both mundane and unplanned events” Miller & Goodchild (2014)

Populations not samples (Miller & Goodchild (2014))

Miller & Goodchild (2014) becomes very problematic when we make generalisations about populations from inferred data (such as twitter)

Oliveria *et al.*, 2015:

This has affected probabilistic models over machine learning models (Oliveria et al., 2015)

Patterson & Fitzsimmons, 2017

MTL Trajet (Patterson & Fitzsimmons, 2017). This has fuelled a shift from tradition methods (i.e. travel surveys, phone surveys) to using more VGI sources (smart-phone apps)

Rashidi *et al*., 2017:

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

VGI, has also allowed study at larger geographic scales

(close to population level) with purpose and sentiment studies (twitter and geo-reffered social media) (Rashidi *et al*., 2017)

Robinson *et al.*, 2017:

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

Shi *et al.* 2018:

Partly due to Crowd-sourcing, mobile phones cost effective (Shi *et al.*, 2018).

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

Shi *et al.* (2018) greater authenticity issues with VGI than other data sources

Sioui*, et al*, 2012

Car sharing in Montreal – some trips always car (Sioui*, et al*, 2012)

Tu *et al.*, 2017:

Inferring information (Tu *et al.*, 2017). Through this, VGI can give us insight into processes occurring in space time that more traditionally collected information cannot (Elwood *et al.*, 2012).

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

Wu *et al.*, 2016

‘People as sensors’ (ref). “there is an opportunity for smart-phones to replace dedicated GPS devices” (Wu *et al.*, 2016)

Xie *et al.* 2016

VGI has gained an increasing amount of attention in the literature and allowed researchers to begin to study

* “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)

Broader classification eith larger data types of patterns at larger scales (Xie *et al.*, 2016)

Yazdizadeh *et al* 2019

Yazdizadeh *et al* (2019) find that models classifying transport mode techniques tend to be more applicable to at a wider range of spatial and temporal scales and are generally more accurate [predictability, less to identify].

By contrast, activity detection and itinerary inference have received less attention still (Montini et al., 2014; Gong et al., 2014; Zahabi et al., 2017), and the field is open to further research

Zhang & Cheng, 2019:

problem with SVM, Decision tree and Naïve Bayes is these such classifiers usually need hand-crafted features as input for training (Zhang & Cheng, 2019).

*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

“Montreal with its unique multimodal public transportation system consisting of bus, metro and commuter train offers multiple transit route alternatives to individuals commuting to downtown" Eluru *et al.* (2012)

an region containing around 4 million people (WPR, 2018; **Figure 3.1**).

### 2.3.1 MTL Trajet Project

The MTL Trajet project is a large scale mobile phone travel survey that has been run yearly around Oct-Nov since 2016 (Ville de Montreal, 2019). The project relies on participatory volunteered geographic information from its app. User friendly interface

Developed onto of the success of the DataMobile smart-phone Travel Survey example from 2014 carried out by Concordia University in Montreal (Patterson & Fitzsimmons, 2016) 10 November - 5 December 2014. Close to 900 people participated in the survey [Only around the univirsity]. Based in Montreal. 18 and over.

MTL2016:

* 5452 subjects
* GPS functionality is turned off and is only reactivated once the phone has moved 100 meters from the last GPS location (using the approximate location from Wi-Fi)
* Respondents were prompted to “validate” their movements between locations by providing information (travel mode and trip purpose) about﻿ their latest trip when phones were detected to have stopped (the application ran 24 hours a day)
* MTL ﻿Trajet provides block level spatial accuracy and is thus suitable for mapping subject locations to air pollution grids derived from existing exposure surfaces.

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

The MTL Trajet is a part of the Itinerum platform which is an app providing researchers a platform to develop their own spatial surveys (Yazdizadeh *et al.*, 2019) Now Itinerum Platform **Figure 2.2.**

See (Zahabi *et al.*, 2017) for datamobile analysis

MTL originally had personal locations (See Table 2.1), although have been removed for the data used in this report as this is available from Portail de Ouverte Donnes

The app employs geofencing

[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).” (**Figure 2.3**; Patterson *et al.*, 2019). Map-matching (**Figure 2.4**)

The casual links have been harder to prove in the literature

Owing to the