# Chapter 5. Discussion

## 5.1 Summary and Implications of Findings

Overall, we see that the results

Clustering:

Requires post-hoc information, to be implemented we have to use a progressive clustering algorithm where the clustering is re-evaluated every so often (if we wanted it to be real time) Otherwise, can look at information backwards. I.e. KMeans can cluster new points in space, LDA new points in time, space-time can cluster emergence of time-space clusters.

### 5.1.1 Main Research Question: Can we effectively classify trip purpose

Purpose classification:

\* Which clustering technique performs best

Model is built on Montreal, may have a completely different result for other place, although we can infer spatial-temporal trends from the results, this may ‘frozen’ in time and space

Although, it must not be forgotten that this study primarily focusses in on Montréal and this may not be transferred to other cities (Ergodoic and Ecological Fallacy). Indeed, it is erroneous to assume that what is examined in across this covered region Montreal at the time of the study period can at all be scaled up to Montreal at a different point in time (i.e. to Winter or 5 years in the future or past), let alone to another city. It is easier to assume instead that is useful information for studying a network of interconnected movement.

Sometimes the movement patterns that result are understandable or explainable, like birds migrating south for the winter, but often times they are not obvious (Murray et al. 2012)

As with all data-driven geography, “﻿A challenge is how to identify the niches to which monitored population data can be applied with reasonable generality.” Miller & Goodchild (2014)

\*\*Bias-variance trade off -> <https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229>

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

On omitted-variable bias (OVB): “occurs when a statistical model leaves out one or more relevant variables” (i.e. purpose?)

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

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

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)

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

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

(fractal) emergence in patterns of travel (Li *et al.*, 2016)

Insight into those spatial and temporal trends can improve the performance of Intelligent Transportation Systems (ITS). (Taayab *et al.*, 2014)

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)

Other models:

CANT DO LSTM or RNN as time is not regular

Can do CNN because of grid

CNN-LSTM using videos of each trip could prove important

Space-Time:

Primarily the movement of people is of concern to time-space analysis.

Paradigm shift (Brunsdon, 2015)

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

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)

### 5.1.2 Sub-Question: Which indicators were the most useful?

Spatial disparity/diversity in the mis-classified. It could be argued that understanding where this occurs across space may help improve the classification accuracy and the understanding of the general processes.

LDA clusters were very effective in seperating the data classes (figure with LDA PCA)

### 5.1.3 Sub-Question: Which models performed the best?

The multi-class models

## 5.2 Uncertainty

Correlation doesn’t mean causality, especially with the Class-imbalance in the trip purposes

\*\*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) [i.e. can’t apply to other parts of the year]

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

We are still missing real information (~20%)

Furthermore, in order to be able to predict transport demand or traffic, not only are real-time data required but also historic data. (Li *et al.*, 2016 -> find another ref, but basically trying to say that historical is needed as well).

Schwanen, T. (2018) -> many forms of uncertainty that cannot be dealt with using better techniques [i.e. with VGI and general space time]

Weather important (Xie *et al.*, 2016), shown to decrease and more precip in later study. We know this affects transport mode, and transport mode may affect transport purpose (ref).

We discover both directional dependence and indendence in the data, thus we may see different things in different directions (anistrophy and isotrophy; see <https://en.wikipedia.org/wiki/Anisotropy>)

Problem ﻿-> “development of explicit, formal, and computable representations of geographic knowledge” (Miller & Goodchild, 2014)

“People ﻿in rural areas tend to use technology differently than people who live in cities.” (Hetch & Stephens, 2014)

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

Data has been routed (Patterson & Fitzsimmons, 2017b; Ville de Montréal, 2019)

Although, it must not be forgotten that this study primarily focusses in on Montréal and this may not be transferred to other cities (Ergodoic and Ecological Fallacy).

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

No consideration of space-time metrics -> clusters -> but harder to include

Land use

As shown in **Figures 3.3** and **Table 3.2**, the land use categories are fairly unbalanced with the majority of land use being residential and employment (27% + 18%, respectively).

## 5.3 Further Research

[Better modelling] “there are uses of machine learning methods that have been extended to account for the limitations of working with spatio-temporal data (such as the integration of convolutional neural networks and LSTM methods (Shi *et al*., 2015; Yu *et al.*, 2017; Han *et al.*, 2019)“. Using a better model that accounts for space-time (CNN-LSTM) -> which you input a video of trips

demographic shifts [through gentrification] creates changing travel demands and employment rate

Dabiri, S., & Heaslip, K. (2018) use CNN for mode classification

Liu *et al.* (2016) predicting the next step with mode -> maybe applied to purpose

\* Jahromi *et al.* (2016) try to simulate GPS movement/mobility that infer about interactions of people with a city and its services [Mention about ABM and simulating interactions -> could act as a scaled up version]. The understanding of human mobility in an urban space has become crucial to optimize the network management (Jahromi *et al.*, 2016). ABM a great tool but ABMs could be very data demanding and sometimes too complex without offering much additional insight (An *et al.*, 2015).

Train ML method to look for uncertainty and outliers (Shi *et al.* 2018). Deep learning requires huge datasets (Shi *et al.*, 2018)

Visualisation and Interactivity:

Can we make big geospatial data analysis and visualisation available to an end-user through interactivity? (maybe Li *et al.*, 2016) -> currently not

Videos of change over time may be needed for space-time investigation

# Conclusion:

Model is built on Montreal, may have a completely different result for other place, although we can infer spatial-temporal trends from the results, this may ‘frozen’ in time and space.

This study attempts to break away from its data-driven approach and provide a more qualitative investigation

Contextual information important -> moving in the future towards more comprehension of travel purpose

A form of spatial analysis which needs a lot more attention, and more surveys to be carried out.

It is observed that this study only builds upon one time period and one city

This study attempts to break away from its data-driven approach to provide more context

* \*Trip sentiment

EXTRA:

## 2.3. Representation of space and time in mobility modelling

Difficulty in space, time and space-time metrics in models

1. Models can’t handle

In general, representing space and time and machine learning models has been a difficult notion. Spatio-temporal information is of prime interest to researchers of GIS (Ren *et al*, 2019), yet it has been difficult to account for the combination of space and time structure cannot be accounted for (Cheng *et al.*, 2011).

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

The broader use of space and time metrics in models has been a challenging concept.

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

“Space–time analysis seeks to understand when and where (and sometimes why) things occur.” 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).

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

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

Significant class-imbalance exists in the MTL Trajet data

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

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)\*

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)

1. Computationally expensive

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

Network topology and space-time weight matrix

CNN was not optimal for modelling 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)

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

1. Things Change

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

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

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

MTUP (Cheng & Adepeju, 2014) and MAUP (Openshaw, 1983)

(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)

Within structures such as cities things change and develop over time in either a state of equilibrium or dis-equilibrium (Batty, 2013)