2017

Exploration Of New Methods In Long Distance Transportation Data Collection And Tourism Travel In Vermont

Benjamin Kaufman

University of Vermont

Follow this and additional works at: https://scholarworks.uvm.edu/graddis

Part of the Place and Environment Commons, Sustainability Commons, and the Transportation Commons

Recommended Citation

Kaufman, Benjamin, "Exploration Of New Methods In Long Distance Transportation Data Collection And Tourism Travel In Vermont" (2017). Graduate College Dissertations and Theses. 809.

https://scholarworks.uvm.edu/graddis/809

This Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks @ UVM. It has been accepted for inclusion in Graduate College Dissertations and Theses by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.
EXPLORATION OF NEW METHODS IN LONG DISTANCE TRANSPORTATION DATA COLLECTION AND TOURISM TRAVEL IN VERMONT

A Thesis Presented

by

Benjamin Kaufman

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements for the Degree of Master of Science Specializing in Natural Resources

October, 2017

Defense Date: May 30, 2017

Thesis Examination Committee:

Gillian L Galford, Ph.D., Advisor
Richard A. Watts, Ph.D., Chairperson
Brian G. Voigt, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College
ABSTRACT

Human transportation patterns have continued to shift and increase in rate as technology has made travel between spatially disparate locations more feasible. These movements are responsible for approximately one third of global carbon emissions, and account for one half of Vermont’s greenhouse gas output. Modeling transportation behaviors is difficult due to changing travel patterns and issues of surveying human participants. Long distance travel patterns are especially difficult and have not received the attention that urban mobility has within the literature.

In this Masters thesis, I describe current methods of transportation data collection and propose new methods, as well as attempt to quantify the impact on Vermont’s roadways of the transportation-based tourism sector. In the first chapter of this thesis, I describe a GPS-based travel survey conducted over the course of one year, coupled with interview data of long distance trips undertaken by 10 participants. Long distance travel has historically been underrepresented in travel surveying due to its infrequency, resulting in decreased likelihood of capturing a long distance trip in a short travel study. By extracting points at intervals from the GPS dataset, it becomes possible to determine accuracy of trip matching between the two datasets with adjusted data collection methods. The second chapter examines transportation related to tourism in Vermont. As one of Vermont’s largest industry sectors, economic impact has been of particular interest to state planners. However, limited analyses of the transportation impacts of this sector are currently available. My research models route choice of drive through tourists, whom constitute 40% of visitors, attempting to begin quantifying tourist mileage and CO₂ emissions within the state.

Together, these studies expand knowledge on long distance transport data collection and the role of tourism in Vermont’s transportation mileage.
CITATIONS

Material from this thesis has been submitted for publication to the Journal of Transport and Land Use on October 31st, 2016 in the following form:

ACKNOWLEDGEMENTS

To my mother Stephanie, full of endless enthusiasm and belief in my abilities. My father, Steve, for reminding me to stay focused and put forth my best effort. My sister, Sonya, for always inspiring me to push my limits and exceed expectations. I am also grateful for my friends who have always been my support system and sounding board, entertaining my ideas and being my most honest critics.

With special mention to Ariana Cano, Javier Garcia-Bernardo, Michael Rothman, Alex Perkins, all the members of B4E, UVMRFC, and Aiken 312.

I am also thankful for the following University Staff: Gillian Galford, Richard Watts, Brian Voigt, and Lisa Aultman-Hall for their support, assistance, and the USDOT Funding for the Masters Research.
# TABLE OF CONTENTS

| CITATIONS | ii |
| ACKNOWLEDGEMENTS | iii |
| LIST OF FIGURES | vii |

## CHAPTER 1: OPTIMIZING MOBILE DEVICE GPS DATA COLLECTION TO CAPTURE LONG DISTANCE TRAVEL | 1 |

### 1.1. INTRODUCTION | 1 |

### 1.2. LITERATURE REVIEW | 3 |

#### 1.2.1. GPS Travel Surveys | 3 |

#### 1.2.2. Collection Devices | 5 |

#### 1.2.3. Cell Phone Usage | 6 |

#### 1.2.4. Passive Data Collection | 7 |

#### 1.2.5. Battery Issues | 8 |

#### 1.2.6. Travel Time Budgeting | 8 |

#### 1.2.7. Daily vs Long Distance Travel Surveying | 9 |

### 1.3 Methods and Data | 10 |

#### 1.3.1. Overview | 10 |

#### 1.3.2. Data Overview | 12 |

#### 1.3.3. Study Participants | 12 |

#### 1.3.4. Collection Application | 13 |
1.3.5. GPS Data Structure ................................................................. 13
1.3.6. Interview Data Collection ......................................................... 13
1.3.7. Interview Data Structure .......................................................... 14
1.3.8. Matching interview trips and machine-recorded trips ......................... 14
1.3.9. Finding the Best Times ............................................................... 16
1.3.10. Number of Possible Data Combinations ...................................... 17

1.4. Results ....................................................................................... 20
1.4.1. Number of Possible Data Combinations ........................................ 21
1.4.2. Accuracy Returned ..................................................................... 22
1.4.3. Hours of Importance .................................................................. 25
1.4.4. Missed Trips ............................................................................ 25
1.4.5. Long Distance Trip Mileage ....................................................... 26
1.4.6. Single Point Extraction Mileage .................................................. 26
1.4.7. Single-Point Hourly Accuracy .................................................... 26
1.4.8. Multi-point Method ................................................................... 27
1.4.9. Long Distance Travel Mileage Accuracy Formula ............................ 28
1.4.10. Multi-point Accuracy ............................................................... 28

1.5. Conclusions ............................................................................... 30
1.5.1. Lots of Data Does Not Return Perfect Accuracy .............................. 30

REFERENCES: ................................................................................. 33

CHAPTER 2: CANADIAN TOURIST TRIPS ON VERMONT .................... 36

ROADWAYS: ROUTE AND EMISSIONS PROJECTIONS ...................... 36
2.1. Introduction .................................................................................................................. 36
  2.1.1 Current State of Vermont Transportation................................................................. 36
  2.1.2. Vermont Tourism .................................................................................................... 38
  2.1.3. Vermont Emissions Goals ......................................................................................... 39
  2.1.4. Meeting Comprehensive Energy Plan Goals ............................................................ 41

2.2. Methods and Data ......................................................................................................... 44
  2.2.1. Modeling Drive-through Trips.................................................................................. 44
  2.2.2. Trip Generation ....................................................................................................... 47
  2.2.3. Trip Distribution ..................................................................................................... 49
  2.2.4. Mode Choice .......................................................................................................... 50
  2.2.5. Route Assignment .................................................................................................. 50
  2.2.6. Implementation ...................................................................................................... 51
  2.2.7. Emissions Calculations .......................................................................................... 54

2.3. Results .......................................................................................................................... 56
  2.3.1. Emissions Calculations .......................................................................................... 57

2.4. Conclusions .................................................................................................................. 57

2.5. Citations ......................................................................................................................... 61

REFERENCES ....................................................................................................................... 63
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.</td>
<td>Long Distance Travel Data Extraction Methodology</td>
<td>17</td>
</tr>
<tr>
<td>Figure 2.</td>
<td>The translucent circles represent buffers of 1, 2, 5, 10, 25, and 50 miles from the locations returned in interviews, Burlington, VT and Boston, MA. The black points are GPS locations extracted from the participants GPS data. Notice that the smallest buffer distances may work for certain locations, but will not return a correct trip for others.</td>
<td>20</td>
</tr>
<tr>
<td>Figure 3.</td>
<td>Graph of diminishing returns of additional points on accuracy of long distance travel data collection.</td>
<td>23</td>
</tr>
<tr>
<td>Figure 4.</td>
<td>Lack of correct data may be due to the true data function, or phones being off. It should be noted that while the city center provides a good reference point for participant travel, it does not directly represent user behavior during the study. Thus a buffer was applied to this centroid in the hopes of better capturing participant movement. As this buffer was expanded, greater results were returned.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 5.</td>
<td>The figure above and the table below demonstrate the gain in predictive ability of the methodology as more points are added. Noticeable benefits are found in the addition of a second point, increasing the accuracy of miles travelled by 11%. This change also demonstrates a doubling of the data required however. This trend of diminishing returns as more and more points are added remains consistent.</td>
<td>29</td>
</tr>
<tr>
<td>Figure 6.</td>
<td>The Four-Step Trip Model is used to model regional trip making by assigning mode and route choices based on the transportation network, demographic data, and trip generative and attractive ability of an area.</td>
<td>46</td>
</tr>
<tr>
<td>Figure 7.</td>
<td>Canadian-Vermont border crossings by both Americans and Canadians show similar travel patterns by length of stay throughout the year. However, volumes of travel by Canadians are about twice as high as Americans.</td>
<td>47</td>
</tr>
<tr>
<td>Figure 8.</td>
<td>Canadian vehicle border crossings are dominated by automobiles, which experience seasonal fluxes as tourism volumes increase in the summer months.</td>
<td>50</td>
</tr>
</tbody>
</table>
Figure 9: Canadian drive through tourist car volumes by road segment based on 2015 Canadian border crossing data allocated to 4 destinations along the Vermont road network. Largest volumes of travel are located along route I-89 before it exits Vermont.

Figure 10. Emissions calculations of automobiles by fuel type and kilometers travelled within Vermont by Canadian tourism drive through trips. Total emissions shown in grams of carbon dioxide equivalents. The vast majority of emissions are resultant from gasoline consumption.
CHAPTER 1: OPTIMIZING MOBILE DEVICE GPS DATA COLLECTION TO CAPTURE LONG DISTANCE TRAVEL

1.1. INTRODUCTION

The progression of travel survey data collection from traditional pen and paper surveys to random digit telephone interviews towards sophisticated, technology driven methods, has opened up new opportunities for research, planning, and improvement in quantity and accuracy of travel data. The use of GPS data streams collected passively by participants can be used to measure an individual’s movements more accurately and with fewer missed trips [1]. The reduction in respondent burden due to passive data collection allows for more participants to be involved in each study offering improved results [2]. Long distance travel (LDT) has been underrepresented in travel surveying due to limitations in the amount of time that a survey can take place and number of participants taking part in each study [3]. Additionally, participants have recall biases that often cause them to forget long distance trips and inaccurately estimate the distance that they traveled [4,5]. The ability to collect data from a greater number of subjects over a longer period of time will allow researchers to gain a better understanding of LDT behavior [3]. LDT, as defined by this study, is any trip from an origin to a destination that is greater than 50 miles (80km).

This study assesses the ability of a passive GPS check-in multiple times each day to accurately identify long distance trip origins and destinations by comparing GPS data streams to participant-reported long distance trips. By examining a 24-hour GPS
data set to determine how often and at what times a phone should collect GPS data points to collect LDT, it is possible to optimize the number of times, and at what times, a mobile application obtains a user’s location while limiting missed long distance trips.

In routine travel surveys conducted on mobile devices, higher levels of battery usage associated with large amounts of data collection have increased respondent burden. This battery drain can cause participants’ phones to die, reducing accuracy of the data collected and participant retention [6]. A detailed understanding of turn-by-turn movements is not needed in this study. Rather, more infrequently collected pieces of data can be used to model overall LDT origins, destinations, and trip rates, the data required for the first two steps of the planning model framework: trip generation and trip distribution.

By collecting entire streams of GPS data, we are able to experiment with different methods of data collection to determine optimal ways to collect long distance travel data. We are able to adjust the number of data points selected each day to identify LDT origins and destinations. More data points does not necessarily mean greater accuracy, as participants may remain stationary for large parts of the day, due to sleep or work. However, if data is captured too infrequently, day trips of short duration may be missed. To assess our methodologies, we compared data selected from our GPS streams to data from participant interviews, allowing us to determine how well each methodology was able to capture long distance trips, reduce incorrect trip
collection, report LDT mileage, and return accurate GPS locations for origin and destination points.

This paper first describes the previous developments in the field of long distance travel surveying and the use of mobile phones and GPS in data collection. Next it outlines the data collection methodology implemented in this survey, as well as the data cleaning and GPS location pairing algorithms. Finally, the accuracy and effectiveness of each algorithm is presented with its ability to represent long distance origins and destinations.

1.2. LITERATURE REVIEW

1.2.1. GPS Travel Surveys

GPS data collection devices have been used in a number of studies to track in-vehicle movements since their introduction to travel surveying in 1997 [14]. Stemming from this original collection, GPS proliferation into data collection methods for travel surveying has become common. Stopher et al recently produced an up to date review of the literature surrounding GPS-based travel surveying and processing, with 4 studies utilizing smart-phone devices in 14 different countries [13]. GPS data logs have historically been coupled with paper diaries, and later with portable electronic devices, to verify accuracy of the data collected. Computer assisted telephone interviews have commonly been used to collect data in addition to the personal diaries.

While there are drawbacks to vehicle based collection methods, such as the lack of ability to track non-vehicular travel, in-vehicle GPS collection devices are capable of
accurately collecting turn-by-turn travel in automobiles. Battery issues with in-vehicle GPS devices are limited as they draw their power from the vehicle itself. As the device is installed in the automobile, the weight of the device is a non-issue for participants.

Collection of data is easily limited to when the automobile is powered on, thus reducing the amount extraneous data collected during the study [16]. Issues arise when automating the detection of short duration trips, as well as preventing traffic lights and congested roadways from appearing as trip ends. Dwell time, use of stationary behavior for a period of time as an indicator of trip ends, was introduced as a way to further passive data collection, and parameters have been created to optimize the number of missed trips without obtaining false trip stops [7,8]. Dwell time thresholds with confidence ratings have been devised to detect trips based on length of time of non-movement. The longer the lack of movement, the more likely the stop location is a trip end.

GPS functions by contacting satellites to obtain a location through triangulation, locating the device on an X, Y, Z coordinate plane. As in-vehicle GPS systems use the car battery to obtain power, sometimes origins of trips are misreported due to the time it takes to acquire a “fix” on satellites [17] and correctly obtain origin location.

Further applications of GPS data recording were applied to underreporting of trips in household travel [9, 15]. These methods improve our ability to model actions of individuals who commonly underreport trips taken, such as high frequency travelers and long distance travelers, who are of specific interest in this study. Long distance
travel is an infrequent activity that requires large populations of study participants to accurately model.

Connecting GPS data to localized maps allows extrapolated trip ends to aid in reporting trip purposes based on information about origins and destinations [4, 15, 17]. These inferences decrease user burden by reducing the amount of data they must record for each trip, once again allowing for larger sample sizes and better understanding of long distance traveler behavior.

1.2.2. Collection Devices

Collection devices have become more advanced and many varieties of devices are now available to the researcher. Originally, collection was done solely in-vehicle [5,7] due to the limited battery capacity of GPS devices that required an external power source for long-term data collection [5]. On-person GPS devices have been used to track non-vehicle travel such as walking and biking trips [16]. Results demonstrated that the bulky, heavy devices were prone to being left at home on many trips, resulting in underreporting and errors in data [5, 10]. As devices became smaller, user burden issues moved from the carrying of the device to issues with charging the GPS device itself [15]. Additionally, the ownership of devices and delivery to users decreased the ability of researchers to have large sample sizes [36]. A better method was needed to collect data from many users across all modes without increasing user burden.
1.2.3. Cell Phone Usage

Enter the cell phone. An intelligent device that has been ubiquitously embraced by society, is in frequent communication with infrastructure, and has the ability to accurately and efficiently collect GPS data while already occupying space in much of society’s pockets and charging schedules. Many cellular positioning studies have now been completed to analyze the potential effectiveness of cellular devices and their many sensors for transportation surveying. Throughout the literature, two main cell phone sensors have been used to determine specific location of participants:

1. the phone’s cellular position from triangulated towers [2,11, 18]
2. the phone’s internal GPS system [4, 6, 7, 8, 9, 12, 15]

Other sensors have been used to determine highly specific movement characteristics, not necessarily location, which often aid in detection of movement speed or mode type. These sensors track minute movements through accelerometer, Bluetooth, and Wifi readings to differentiate between movement patterns associated with different modes. As these data are not topics evaluated in this study due to their meta-movement related nature, this paper will not investigate the use of these sensors. Extensive literature exists regarding the use of these auxiliary sensory data to detect travel mode types [19, 20, 21, 22].
1.2.4. Passive Data Collection

Travel data collection processes can be separated into two categories: user-flagged and passive. User flagged systems require the participant to initiate and terminate data collection during their trips, thereby determining trip starts and stops. This increases user burden and introduces human inaccuracy as participants may forget to initiate or end trips, even with prompting. Passive data collection requires no interaction with the participants, limiting user error and burden by automatically inferring trip ends based on movement characteristics [15,16]. Research into passive data collection has led to further development into intelligent collection methods. By determining a certain time frequency for collection, over (or under) collection of data can be avoided. Additionally, speed based collection allows for data to only be collected while the participant is in motion, reducing erroneous data collection, battery usage, and frequency of connection to a server while participants are stopped for long periods of time (Srinivasan, 2009).

The cost of each additional participant in travel surveying has been a major limiting factor in the expansion in survey size and longevity of these studies. Participant recall has been improved and time expenditure reduced by using GPS data streams of recorded movements overlaid on maps as a tool to aid participant memory [1,15]. Major reductions in the amount of time needed for participants to recall trips when prompted with a visual of their movements were found. This reduction could aid in reducing costs and increasing size of studies implementing GPS logs in the future.
1.2.5. Battery Issues

Collection of large amounts of data on personal cellular phones has increased the amount of battery power that the devices consume. Battery life issues are of major concern as smartphones have responsibilities besides their use as travel diaries [6, 16, 17]. As the device becomes useless as a GPS recorder if it is turned off or loses power, we aim to find a way to accurately represent long distance travel mileage using GPS data collection while minimizing device battery and data usage. GPS use on a smartphone can drain its battery in a matter of hours, more efficient activation of this sensor can improve battery lifespan [37].

1.2.6. Travel Time Budgeting

A travel time budget is the concept that people have a personal time allotment that they spend on travelling [38]. This has been found to be roughly 1 hour per day spent on travel, regardless of transportation infrastructures and across cultures [38]. With economic growth, people are able to afford to travel at faster speeds allowing them to reach more distant destinations within their allotted time budget [38]. Long distance is therefore relative, based on economic resources. With the adoption of motorized transportation, passenger miles traveled has grown exponentially [38]. In developed nations, automobile travel occupies the majority of passenger trips taken, and in developing nations cars are rapidly generating more and more miles each year.

In contrast to global trends, US urban and rural miles traveled by car have been decreasing in recent years [39], while air traffic is surging forward occupying both
greater passenger miles travelled, but also greater traffic volumes [38]. As the US experiences greater economic prosperity, travel time budgets have remained constant over time, prompting passengers to choose faster and faster means of travel [38].

1.2.7. Daily vs Long Distance Travel Surveying

The National Household Travel Survey (NHTS), conducted by the US Department of Transportation, uses versions of the question “what travelling did you do during our study period?” to obtain data. This works adequately for frequent trips such as urban travel where a forgotten trip does not drastically shift the results due to large trip volumes. Long distance travel occurs much less frequently, and therefore requires much larger population sizes in order to accurately collect data for short study periods such as those implemented in the NHTS. Each missed trip has a greater effect on the results of the study with regard to our understanding of LDT.

Shorter, urban trips take up a large number of the total trips taken, yet LDT occupies a large amount of the total mileage of our societal transport and contributes significantly to our Greenhouse gas emissions [23]. Long distance travel is hard to measure because it is uncommon on an individual level and participation is uneven across the population [34] as it is highly reliant on economic prosperity.

Thus, older methods of long distance travel data collection are not one-size-fits all. Virtually every LDT survey uses its own definition of long distance travel, making comparison of results difficult. New methods of collecting data on LDT can help
increase accuracy of the data collected in this field [25], and having consistency across studies can bolster collaborative efforts.

Numerous studies defined the construct of LDT differently. It was found that LDT was captured as >80km as the crow flies [25], >100km through the road network [26, 27, 28], a trip with an overnight stay [29], an excursion of greater than 3 hours [30], >300km [31], >50 miles [32], and > 100 miles as the crow flies [33]. This inconsistency may help create general guidelines for LDT research, but just as one study may not capture all travel effectively, one distance may not work in all instances to accurately capture LDT.

1.3 Methods and Data

1.3.1. Overview

In order to better understand long distance mobility, research methods in addition to the National Household Travel Survey are needed. This study implements a passive GPS data collection on participants’ mobile devices to collect 24-hour GPS location data.

These 24-hour logs are used to create LDT logs by extracting participant location at certain times of day, and then examining that GPS data for movement greater than the designated long distance threshold. A distance of 50 miles as the crow flies was selected by the research team.

This study implements periods of check-in between 1-24 hours. This was deemed as an appropriate length of time by the research team, as the data collected...
represents large movements, as opposed to turn-by-turn behavior found in traditional GPS surveys. This examination creates a LDT log; with each origin and destination GPS point local time and date it was collected.

At the end of the study period, participants were interviewed to obtain all long distance trips that they took. The use of the activity logger, Moves, which provides visuals of participant movements overlaid on Google Mapping software. Each trip’s origin, destination, and date that the trip was taken were recorded in a text format. These data were then converted into GPS coordinates to allow comparison to the data collected by the mobile application.

The long distance GPS data is then compared with the interview data. Each long distance trip extracted has an origin and destination, which are compared with the interview responses. A buffer of 24 hours was used to allow for trips originating or ending within a day’s time of the stated date in the interviews to be counted.

By testing many different times and number of check-ins per day, this study was able to isolate the minimum number and times of day a phone should passively collect GPS data to accurately collect LDT origins and destination, reduce battery drain, participant burden, and ensure privacy of those involved.

There are limitless possibilities when choosing times of day to analyze for potential accuracy in data collection. For this study, we decided to limit our data extraction to the beginning of each of the 24 hours of the day. Thus the maximum
number of unique data points that could be returned each day was 24. A minimum of one data point was tested to determine accuracy of the algorithm.

1.3.2. Data Overview

To process GPS data, Python 3 software was used, alongside programming library GeoPy. The most relevant steps when generating long distance travel trip logs are:

• Processing of data into 24-hour logs.
• Selection of data based on input variables.
• Comparison of long distance trips to true data collected using NGSA-II.

1.3.3. Study Participants

Participants were recruited to this study from the University of Vermont Transportation Research Center, and through personal networks of the researchers. The mobile phone application MOVES was installed on 12 IOS devices for 2-13 months to collect GPS data. At the end of this period, participants were contacted to conduct an exit interview and export their data. These interviews allowed for the collection of LDT logs that each participant completed during their time in the study. Of the 12 participants, 10 completed the exit interview and final data collection. Of the two missing participants, one was unable to meet during the available interview times and one had technical difficulties exporting their data.
1.3.4. Collection Application

This study selected a mobile device based GPS data collection app, *Moves*, which acts as an activity log and passively collects GPS data based on device movements. *Moves*’s ease of installation on participants’ devices, the ease of export of GPS data, and the application’s ability to collect data indefinitely without any participant interaction were all factors that made it a viable choice.

1.3.5. GPS Data Structure

For this study, GPS data was exported from *Moves* in JSON format and was analyzed using the coding language, Python. After export, it was converted into a travel log, with GPS coordinates and time stamps tracking each participant’s movements throughout their time in the study. GPS points are tagged as either a Place or a Movement dependent on the users behavior. These points make up a 24-hour log of each user’s location throughout their time in the study.

1.3.6. Interview Data Collection

At the culmination of the study, participants were contacted to complete closing interviews. At this time, participants were given detailed instructions on the process of exporting their GPS data for researchers to analyze. Participants were also told to bring any tools that might aid in their recollection of any Long Distance trips they took: personal calendars, any itineraries they had compiled, their email accounts, and the *Moves* application itself. They were informed that for the purpose of this study, a long
distance trip was defined as any movement from an origin to a destination that was greater than 50 miles. If a participant was unsure if a trip they took fell into that threshold, they were asked to report it and it would be checked by the researcher for accuracy at the culmination of the interview.

1.3.7. Interview Data Structure

During the interview, participants were asked to state the date that each trip took place, the origin location, and the destination location. Location City and State names were recorded for trips inside the US. For trips outside the US, the Country’s name and City name were recorded.

After the interviews were completed, the Moves GPS data was used to check that participants did not forget any trips. Researchers extracted movements of greater than 50 miles from the GPS data and any trips that were found, but went unmentioned, were presented to the participant to see if they had taken the trip, or if it was an incorrect GPS recording. If the participant confirmed the presence of a missed trip, that trip’s information was added to their interview data. Through this process, the researchers are confident that each long distance trip that participants took during the study was correctly recorded.

1.3.8. Matching interview trips and machine-recorded trips

A goal of this research project was to directly compare machine recorded GPS data with interview trips for the same travel. The post processing of data in order to
match these two trips was much more difficult than expected due to a number of reasons.

- Trips collected in participant interviews lack start times, length of time travelled, or distance between origin and destination.
- Interview trips lack exact origin or destination GPS locations; rather they only include city names, leading to lack of specificity.
- Trip start and end dates may vary between interview data and GPS data depending on what time of day the trip was taken and when the GPS point was selected from the travel log.

Travel distance measurements vary between GPS and Interview data. The GPS location recorded by the mobile device will measure the exact location selected for that trip. The Python library, GeoPy was used to match City and State/Country names returned by participants in interviews with city center GPS locations. With the interview data and the Moves data in GPS format, it became possible to test each method of GPS data extraction for accuracy.

The Moves location may be collected while at the final destination, or en route to that location. In the interview data, the final destination is simply the GPS city center of the name returned during the exit interview. Unless the participant is staying at that central GPS point, it is unlikely to return an exact distance for a trip. Additionally, participants may not know the nuances of their travel area. This may result in misspelled location names or incorrect city names due to uncertainty about ordinances.
Therefore during our matching process, if the returned GPS origin or destination is within 50 miles of the interview data, it is marked as a correctly identified trip.

During the study period, the participant’s phone may have been turned off during, at the start, or culmination of each trip. Thus a technique for matching all GPS trips with corresponding interview trips is difficult and returning 100% accuracy improbable. The data is only as good as the participants are at maintaining their phone battery.

1.3.9. Finding the Best Times

In order to determine the best times of day to collect data, and compare these points, this study implemented a multi-objective evolutionary algorithm (MOEA). The specific MOEA was a non-dominated generic sorting algorithm, NSGA-II, in the Python programming language.

This test had two objectives (Number of Times, What Time), with 24 possible dimensions (each hour of the day). Each possible solution, a candidate, is then compared to all other generated candidates based on its performance with regards to the objectives. Each candidate has two characteristics calculated for it: the number of times it performed worse than other candidates (domination), and the set of candidates that it dominates. Each candidate with domination counts of 0 are maintained as the top candidates, while the rest are removed from the remainder of the analyses.
1.3.10. Number of Possible Data Combinations

\[ \sum_{k=1}^{n} \frac{n!}{k! (n-k)!}, \text{ for } 0 \leq k \leq n \]

N is the number of possible times, 24.

\[ \sum_{k=1}^{24} \frac{24!}{k! (24-k)!} = 16,777,215 \text{ possible unique combinations} \]

With over 16 million possible choices, with multiple evaluation criteria, a manual examination was excluded.

Figure 1. Long Distance Travel Data Extraction Methodology
**Inputs:** JSON File: The participant’s JSON file of GPS data collected by *Moves* and exported during the exit interview is inputted into the model.

True Data: During each participant interview, an Excel file is generated with each date that the participant was involved in the study. This file is then populated with dated origins and destinations of long distance trips that the participant took during their involvement in the study. The City and State name are listed for each trip in the US. For each trip outside the US, City and Country name are used.

**Processing:** Trip Log: A 24-hour Trip Log is generated from the JSON File, with each GPS point collected during the study having a dated timestamp.

GPS City Center: The True Data city names are converted into City Center GPS points using the Python library, GeoPy.

**Extraction:** This process uses the following data collected from participants during their exit interviews, as well as parameters assigned by the researchers:

**Parameters:** Parameters are used to determine exactly what is extracted. A variable stating at what point(s) each day to extract data is implemented. These points are 1-24 hourly marks that return the next available data point for that subject.

**Comparison:** A pareto-optimal solution is one that optimizes possible outcomes in a multi-objective problem, finding the best overall solutions. In this case, maximizing the percentage of captured trips, while minimizing the data points used. Rather than attempt to test each possible solution by hand, evolutionary algorithms were employed to find the optimal time of day, and number of points used each day to
accurately retrieve LDT origins and destinations. 100 tests were run using the multi-objective evolutionary algorithm (MOEA), non-dominated sorting genetic algorithm II (NSGA-II) [40]. The accuracy of trip detection served as the factor of fitness for each candidate, allowing for it to dominate, or be dominated by each other candidate.

**Candidate Accuracy:** For each trip, if an origin or destination is present in both the True Data and the Trip Log it is marked as recorded. A variable buffer between the True origin and destination, and the Trip Log’s returned locations was used to determine how accurate the City Center GPS locations were at capturing actual travel locations. The Trip log sometimes varies by up to 24 hours compared with the True Data based on the time that the trip was taken, and when the data is collected based on the input parameters. If both the origin and destination are correct for a trip, that trip is counted as a correctly identified trip. The greater the percentage of trips returned by a candidate, the higher the score that it is assigned.

**Variation:** Variation within candidates is maintained by using evolution methods within offspring creation. Both crossover and mutation methods were implemented in this instance. Crossover divides successful candidates into pieces and combines these pieces with other successful candidates to generate offspring. Mutation creates slight adjustments within successful candidates to test for improved accuracy.

In this way, we were able to determine the maximum accuracy possible to obtain from the dataset. Additionally, we were able to find the best times of day to collect data; depending on how many points you were hoping to collect. Finally, we
were able to examine diminishing returns and suggest a number of points that optimize number of check-ins and accuracy and evaluate this process at determining precise trip ends.

**Figure 2.** The translucent circles represent buffers of 1, 2, 5, 10, 25, and 50 miles from the locations returned in interviews, Burlington, VT and Boston, MA. The black points are GPS locations extracted from the participants GPS data. Notice that the smallest buffer distances may work for certain locations, but will not return a correct trip for others.

1.4. Results

The goal of this project was to examine the ability of a passive mobile application to collect GPS data points for long distance travel origins and destinations. Participant GPS travel data was collected over a period of months, and then interviews were performed to determine LDT origins and destinations, which were turned into city
center GPS locations. Data was extracted from the GPS data set at the start of each hour in varying degrees to determine optimal times to check-in. Accuracy of this method was determined in two ways:

1. The ability for a non-generic sorting algorithm to determine optimal times, frequencies, and buffer distances to extract trip origins and destinations.
2. The ability for extracted data points to represent LDT mileage.

1.4.1. Number of Possible Data Combinations

Accuracy of each candidate was determined by examining the data for long distance trip origins and destinations collected. A candidate received a score of 100% if no long distance trips were missed.

A distance buffer was created surrounding each City Center GPS location to accommodate for the lack of exact location reporting in interviews. Evaluation was originally run using the ability of the test to return correct trip origins and destinations with 6 different distance buffers: 50, 25, 10, 5, 2, and 1-mile radii from city centers.

The maximum buffer implemented was 50 miles, the largest value allowable based on the minimum distance we allowed to categorize long distance travel. Any buffer larger than that might categorize the origin as the destination for trips nearing 50 miles.
1.4.2. Accuracy Returned

The ability for each distance buffer to return correct data decreased with buffer size. The largest buffer accurately matched GPS trip origins and destinations within 50 miles of the True Trip origins and destinations for 91% of the long distance trips made. Decreasing this buffer to 25 and 10 miles reduced the accuracy negligibly (91 and 89% respectively), but provided much more specific information about the users location. When this distance was reduced to a 5-mile radius from the city centroid, the algorithm was only able to accurately return 80% of the trips taken. With a 1-mile buffer, 55% of all trips were accurately returned.

At the maximum buffer distance, 50 miles, accuracy reached 92% of all long distance trips’ origins and destinations captured, using 8 data points. By decreasing this buffer to 25 miles, only 89% of trips’ origins and destinations were correctly accounted for, but with much greater detail as to where the participant was.
1.4.3. Importance of each Point

Each distance buffer expressed diminishing returns based on number of points used. The first optimal time for collection in each distance held between 49% and 26% of the correct data for that test. As the distance buffer reduced in size the first data point held less of the total accuracy. In the least specific test, using a 50-mile buffer, the first four data points held 92% of the correct data in the trial. In comparison, the 1-mile
buffer test required 11 data points to return 92% of the total data collection, with the first data point only returning 27% of the correct results.

**Figure 4.** Value of each additional point in total accuracy return by buffer size

Lack of correct data may be due to the true data function, or phones being off. It should be noted that while the city center provides a good reference point for participant travel, it does not directly represent user behavior during the study. Thus a buffer was applied to this centroid in the hopes of better capturing participant movement. As this buffer was expanded, greater results were returned.
1.4.3. Hours of Importance

The single hour of greatest importance for each test was inconsistent. Four of the buffer distances, 25, 10, 5, and 2 miles each found the most accuracy for a single point by extracting data at 9am. The hour of 3am tied with 9am for two-mile accuracy, and was the most accurate for one-mile accuracy. Each test returned an early morning hour and a late morning hour as the two points to be used for greatest accuracy. The first time a time after noon was used was for the three-point accuracy at five-mile buffers. Even when using five points, the majority of the data takes place in the morning, with only one extraction point coming from the after noon in any trial. Unsurprisingly, there was no consistent answer for the most accurate times of day across distance buffers.

While some might expect extraction throughout the day to return the best results, these tests show that this is not true. Using a 50-mile buffer, 80% of all trips can be caught using only three times, 4am, 7am, and 9am, all taking place before noon.

1.4.4. Missed Trips

While the 50-mile buffer returned 92% of all trips, that 8% is still unaccounted for. More research must be done to determine the characteristics of these missed trips. They could be due to lack of phones being turned on, incorrect recall in the interview phase, or error within the software.
1.4.5. Long Distance Trip Mileage

Testing was completed to see how extracting data throughout the day could be used to examine total mileage travelled. While the data represented at most, 92% of all trips taken, there remained the possibility that incorrect coding of trip ends could occur while still capturing the majority of the miles travelled.

1.4.6. Single Point Extraction Mileage

First, the ability to extract a single data point each day from each participant was examined to see if it was a feasible means of analysis. This test was run for each hour of each day. The results varied greatly, and seemed to mirror assumptions on traveler behavior, mainly that participants would be static during the nighttime hours, and have varying locations during the day. This test at best performed 100% accurately at estimating LDT. This successful test proved that the extraction of one single point each day was able to explain a majority (83%) of long distance travel with a fraction of the data usage.

1.4.7. Single-Point Hourly Accuracy

Time of day was shown to effect accuracy of the algorithm, with miles recorded ranging from 93-110% of the true distances. This wide range of 17% inaccuracy can be attributed to missed day trips in GPS logs, the reporting of intermediate locations’ GPS location instead of true trip ends, and the use of city center GPS points as Origin and
Destination points. Of the 24 hourly results, 4 hours were accurate within 1% of the total long distance travel.

The most accurate times of day to collect data were between 7-8am or 1-2pm, all reporting within 1% of the True Distance travelled by participants. During these times, it can be assumed based on the data that users had completed the majority of their traveling, returning an accurate amount of mileage for that trip’s distance without picking up any false mileage.

The night hours (from 9pm-3am) held the majority of mileage underreporting, with 7 continuous hours reporting 95% or less of the total miles traveled. This lack of mileage can be associated with missed mileage from day taken trips by participants. Over reporting of mileage occurred in spikes throughout the day, with the grossest mileage occurring between the hours of 9am-12pm, 3-4pm, and 6-7pm.

1.4.8. Multi-point Method

Using multiple strategically selected points throughout the day improved accuracy by increasing the data available for analysis. With the added points, it was predicted that missed day trips, and en route point selections would be averaged and their impact on total mileage reduced. When using multiple points, the points were evenly distributed throughout the day to prevent clustering and in an attempt to capture periodic movements. The time between points was dictated by the number of points being collected, the greater number of points, the fewer hours in between collection (for example, a 3 point test would collect data every 8 hours.
Trials were run using 2, 3, 4, 6, 8, 12, and 24 GPS points per day. From these points we were able to determine the impact of varying the number of GPS points as well as the times of collection throughout the day on the accuracy of the algorithm. For each of these trials, the segments’ starting and ending times were tested at each hourly interval. Then each segment calculated mileage individually, and these miles were averaged. The formula used to calculate the accuracy of each trial is below.

1.4.9. Long Distance Travel Mileage Accuracy Formula

\[
\frac{M_1 + M_2 \ldots + M_P}{P} \times \frac{100}{M_T}
\]

\(M_P\) is used as the mileage collected from each point. \(P\) represents the number of data points collected throughout the day and is used to find the average distance. \(M_T\) is the true mileage calculated based on the summation of all participants’ reported Long Distance Travel.

1.4.10. Multi-point Accuracy

Multi-point analysis improved in predictive ability as more points were added. The improvement from 1 point to 2 points resulted in increased accuracy of 11%, decreasing the maximum calculated distance to 103% from 110% and increased the minimum distance to 97% from 93% of the True Distance. As more points were added, accuracy continued to increase. 100% accuracy was first achieved when collecting 12 points per day.
The output from this algorithm is a list of origins and destinations and the associated long distance mileage for each long distance trip that a participant completed during the study. Accuracy of this algorithm is determined by the miles captured by GPS points compared to the true miles traveled as reported by participants. The distance between the highest and the lowest mileage returned are compared to the True mileage to determine the percent accuracy of the method.

Figure 5. High and Low Mileage Results

The figure above and the table below demonstrate the gain in predictive ability of the methodology as more points are added. Noticeable benefits are found in the
addition of a second point, increasing the accuracy of miles travelled by 11%. This change also demonstrates a doubling of the data required however. This trend of diminishing returns as more and more points are added remains consistent.

1.5. Conclusions

This study returned five major conclusions:

1. The collection of data a few times throughout the day allowed for capture of many of the participant’s long distance trips.

2. Accuracy of these collections is based around a characteristic that has not been analyzed yet - buffer distance.

3. Even with high levels of data collection, phones do no return 100% of long distance trips.

4. As you collect more data, accuracy does increase. To a limit.

5. The accuracy, and the amount of data collected, are dependent on how specific you want your responses to be.

1.5.1. Lots of Data Does Not Return Perfect Accuracy

In travel surveys, long distance travellers have a tendency to forget trips, and misrepresent mileage travelled. The purpose of this study is to evaluate the ability to capture LDT mileage with a reduced amount of data collected from a passive mobile phone GPS tracking application.
Even when collecting 24 data points a day, and using a distance buffer of 50 miles, 100% accuracy was never achieved for origin destination pairing. This could be due to several factors:

- Lack of phones being turned on to collect data throughout the study.
- Incorrect interview response data due to incorrect memories or lack of information by the participant.
- Error by the phone app in returning correct data.

Sometimes participants left their phone off, or it died, throughout the study period. If this occurred, it would be impossible for the phone to collect travel data. Participants sometimes provided responses that did not correctly locate their exact location; such as if someone visited a suburb of a city but was unaware that their destination was not within the city limits. Sometimes, the phone would place the participant in a location that they most likely did not travel to. This occurred most noticeably during flight.

This study found that it is possible to use only a single point per day (83% accuracy) to estimate LDT Behavior, greater accuracy in results were found as more data was made available. As more points per day were added, the accuracy increased and 100% accuracy was achieved when using 12 points per day.

Inaccuracy from using fewer points was caused by mileage accrued during missed day trips, improper Origin or Destination GPS points, and City Center GPS locations. This study found that there were consistencies in reported locations at night,
which could be examined further to better understand overnight locations as a detector of LDT. Future improvements to the study could involve the propensity for participants to travel certain distances based on their origin, mode of travel, or demographic profile.
REFERENCES:


32. United Kingdom Data Archive. Online Description of the United Kingdom NTS. www.data-archive.ac.uk/findingData/snDescription.asp?sn=5340


CHAPTER 2: CANADIAN TOURIST TRIPS ON VERMONT ROADWAYS: ROUTE AND EMISSIONS PROJECTIONS

Vermont hopes to reduce its transportation carbon footprint, but possesses limited information detailing regional travel through the state. Vermont is a rural state with high volumes of tourist travel that are invaluable in the Vermont economy supporting jobs and commerce. Vermont has a highly progressive agenda to reduce its carbon emissions. Transportation within the state produces nearly half of Vermont’s greenhouse gas (GHG) emissions annually. This study examines the movement of long-distance travel through the state from Canadian points of entry and its impact of greenhouse gas emissions. Due to availability of data detailing Canadian entrance into Vermont and study of common drive through tourist destinations, Canadian drive through trips were examined in this study. Using theory based on the gravity model, this study assigns tourist movements to specific segments along the road network and quantifies miles travelled and resultant carbon emissions. Common tourist routes were found along high volume corridors such as I-89 and I-93.

2.1. Introduction

2.1.1 Current State of Vermont Transportation

Creating a low emission transportation system in Vermont is difficult due to the rural nature of the state. Vermont is home to dispersed human populations due to historic land use practices such as farming and timber harvests. Large distances exist between individuals and population hubs creating difficulties in generating public
transport networks. One third of the GHG emissions within the U.S. come from transportation, yet in Vermont these behaviors represent nearly half (47%) of all GHG emissions with over 7 billion vehicle miles travelled (VMT) each year (Recchia, 2016a; Sears & Glitman, 2011a). While Vermont ranks 50th in total energy consumption and 43rd in per-capita carbon emissions of states in the United States, Vermont’s transportation emissions percentage compare poorly to the national average of 33% (U.S. Energy Information Administration, 2014).

Government priorities determine overall transportation systems by guiding funding for transport infrastructure. In Vermont, paving and maintenance account for nearly a third of all transportation dollars, bridges 14%, and non-single-occupancy-vehicle travel 8% (e.g., combined public transit, rail, cycling, and pedestrian travel) (Sears & Glitman, 2011b). Low levels of funding for non-single-occupancy-vehicle travel hinders the systems ability to develop reductions in carbon emissions. Vermont’s transportation revenue faces budget deficits expected “as high as $8 billion over the next 20 years” (Recchia, 2016b). While these budget constraints loom, “VTrans spending reflects a high priority on preserving and maintaining the existing transportation system across all modes” (Searles, Osborne, & Pierce, 2013). Vermont states its commitment to smart growth and efficient land use planning but still has a long way to go before achieving these goals.
2.1.2. Vermont Tourism

Vermont works hard to promote its brand, and attracts tourist drivers seeking activities such as bikers and hikers in the summers and skiers in the winters. Additionally, Vermont hosts 5 million Canadian and American drive-through visitors each year en route to other destinations. The majority of domestic visitation originates from drivers within a 300-500 mile driving range, including visitors travelling from Canada (Jones, 2015). Current tourism drive through and day trip mileage is estimated to be around 15% of all Vermont annual mileage based on fuel purchasing records (Jones, 2015).

While data on tourism transportation movements in Vermont is limited, the economic impacts of tourism are clear. Tourism is responsible for nearly 8% of Gross Domestic Product of Vermont, employing 30,000 Vermon ters, and providing $2.49 billion dollars of spending from out-of-state visitors annually (Jones, 2015). On average, tourists represent 1 out of every 10 people sleeping in Vermont each night (Jones, 2015). Vermont hosts 13 million visitors on 5 million trips each year, equivalent to more trips than the total trips served by all of Vermont’s annual passenger rail and public transit services combined (Jones, 2015; Recchia, 2016b). Each trip’s mileage on Vermont roadways causes GHG emissions that the state must account for when meeting state goals. This research attempts to determine the impact that Canadian drive through trips have on Vermont’s annual VMT and GHG emissions.
2.1.3. Vermont Emissions Goals

Vermont strives to extensively reduce its GHG emissions by 2050. Vermont law S.66 states emission reduction goals from a 1990 baseline of 25 percent on or before January 1, 2012; 50 percent on or before January 1, 2028; and 75 percent on or before January 1, 2050 (Lyons, Ayer, Cummings, MacDonald, & McCormack, 2017). These reductions will be made partially through a cap and trade program, allowing Vermont to purchase clean energy credits, but the majority of reductions must be found through reductions in energy use.

The Vermont Comprehensive Energy Plan has created more specific, sector-based goals to reduce GHG emissions. It aims to reduce emissions from transportation by 30% by 2025, employing a fleet that uses 10% renewable energy and 10% electric vehicles, and to reduce single occupancy vehicle commute trips by 20% by 2030 (Recchia, 2016b). Additional goals include shifting Vermont’s transport network from an automobile-based transportation system towards one where “transit, passenger rail, walking, biking, car sharing, and ridesharing… are a state priority” and (Recchia, 2016b).

In order to accomplish the goals of S.66 and the Comprehensive Energy Plan, a fleet transition to electric vehicles is poised to play a large role in the state’s future transportation network, which aims to host up to 600,000 electric vehicles on Vermont roadways by 2050 (105% of the registered 2017 vehicle fleet) (Recchia, 2016b; Report, 2017). In order to achieve these goals, a large-scale fleet transition is necessary,
facilitated by consumer incentives, marketing initiatives, and specific investment in workplace and public charging infrastructure (Recchia, 2016b).

These vehicles will require an increase in electricity generation to support their function instead of the current use of fossil fuels to power automobiles. S.66 states that for transportation, regulation of emissions will be determined based on where the fuel initially enters commerce in the state. Similarly, regulation of electricity generation within the state will be implemented at the generation site, thus ensuring that emissions from local generation are not ignored. For electricity generated outside of Vermont, regulation will be placed upon the utility that distributes it within the state. With these regulations in place, Vermont should be well equipped to determine the GHG emissions from its fossil fuel and electricity based transportation network. Tourism automobile transport is currently dominated by fossil fuel powered vehicles. A shift toward electric vehicles travelling into Vermont would help reduce emissions from tourism.

The Vermont Comprehensive Energy Plan commits over 20 pages of detailed industry analysis and projection to the state of the electric car market, while allocating only 10 pages to all alternative modes of travel. It fails to mention “tourism” even once in its entire 469-page manuscript. This lack of planning for a sustainable tourism transportation future could make Vermont’s carbon emission policy goals difficult to achieve moving forward.
2.1.4. Meeting Comprehensive Energy Plan Goals

Due to Vermont’s “urgent need to mitigate the global climate change that is resulting from greenhouse gas emissions while also advancing local environmental sustainability”, the Vermont Comprehensive Energy Plan was devised to outline specific actions to be taken to reduce Vermont’s carbon footprint (Recchia, 2016a). There are two ways to achieve this: changing where Vermonters go and/or changing how they get there. Noting the limited scope of the plan, only targeting Vermonters, it is of particular importance when thinking about transportation emissions to examine the full system, uncontained by state boundaries.

Smart growth practices are recommended throughout the state to impact land use patterns (Credit, 2007). Smart growth calls for compact, mixed-use development that reduces household travel distances as well as promote efficiencies in other energy sectors. Through strong community centers and rural landscapes, Vermont hopes to intentionally impact how its population moves within its town and city centers. These goals are represented in Act 250, Vermont’s Land Use and Development act, coupled with Act 183, a guide to local development. Act 250 encourages more dense settlement patterns, conducive to non-automobile transportation, through mixed-use development and regional planning processes. It was enacted to dictate land use planning across the state, rather than solely at the local level. Act 183 acts as the guide for smart growth development at the local level as a funding mechanism to support “high density, concentrated, mixed-use developments for growth centers” (Credit, 2007). It
specifically targets minimizing vehicle trips, as well as promoting non-automobile forms of transportation.

These adjusted land settlement patterns represent a long-term path to reducing carbon emissions and automobile dependency, specifically for state residents. Changing settlement patterns is a slow process that may not fulfill Vermont’s Comprehensive Energy Plan goals in the short term. These reductions will be difficult to find in modal shifts such as “the addition of improved transit, car sharing, biking, and walking infrastructure” without changes in land-use occurring first (Recchia, 2016b). Emissions reductions must be found within the current system if short-term goals are to be reached, alongside sustained land use change towards smart-growth to reach long-term goals.

Within the current Comprehensive Energy Plan, emissions reductions will be found in a shift toward electric cars powered by renewable energy generated within Vermont. Electric vehicles have no tailpipe emissions; rather they take on the emissions profile of the local grid, and Vermont boasts a highly renewable electricity production system. Through the proposed cap and trade program laid out in S.66, carbon emissions of electricity sent into Vermont for transportation are effectively accounted for, yet it fails to deal with the life cycle of the manufacturing of the vehicles. Electric vehicles have a large carbon footprint at point of manufacturing due to the process of battery production. This process will be take place outside of Vermont, yet Vermont has no means of accounting for emissions from the production of these vehicles, externalizing
emissions from the most carbon intensive part of developing this new transportation fleet.

To ensure that state efforts to create a public transit system are successful and efficient, Vermont Statute Title 24 Municipal and County Government Chapter 126 constitutes that public transit services be evaluated annually using fiscal and performance standards, with underperforming routes subject to “be reviewed to determine if the service is needed, and if alternate methods for providing the service might be more efficient and effective” (Sullivan, Kenyan, & Watts, 2013). Recent development of intercity bus services are currently running, providing trips between certain towns daily.

However, the current transportation system is already operating at a fiscal loss. If the state’s plan to change to electric vehicles comes to fruition, funding for roadways often found through gas taxation will decrease as drivers shift away from purchasing gasoline. With home charging technology, determining funding allocation to roads may no longer be based on fuel usage. Even with these decreases in revenue, demand for maintained roadways will remain, accompanied by an estimated $4.2 billion-$8.7 billion shortfall in funding between 2006-2021 (Sullivan et al., 2013). If Vermont hopes to maintain its transportation network, a new taxation system may need to be implemented in the near future.

Initiatives such as the Northeast CanAm Connections Trade Corridor Study or the Multi-State Zero Emission Vehicle Action Plan, and Quebec’s ratification of a Zero
Emission Vehicle Action Plan, are important steps toward seamless multimodal mobility options across the entire region (“Vermont Long Range Transportation Business Plan,” n.d.). This initiative is important, as this may signal that future tourism populations from Canada may be more likely to use electric vehicles.

2.2. Methods and Data

2.2.1. Modeling Drive-through Trips

Within transportation planning, modeling of travel behavior is commonly referred to as trip modeling. Trip modeling is commonly completed using origin and destination data, specified locations that trips begin and end. Trip-based modeling has two stages: development of traveler and network characteristics, and placing trips within the network. In the initial stage, traveler and network characteristics are created to understand demand for travel. In the second stage, this demand is placed on the transportation network to model movements. This process utilizes the gravity model, a method of dispersing trips throughout a network based on origin characteristics, destination characteristics, and costs of spatial separation (Cesario, 1973). In its original form, the gravity model follows two basic rules:

1) The larger a location is, the more pull it has on neighboring nodes.

2) More pull is exacted on closer nodes than farther ones.

The term “gravity” is used due to the similarities in its creation and derivation to Newton’s law of gravity, which calculated the forces generated between two objects based on their mass and the inverse square of their distance. Originally, when gravity
was modeled in Reilly’s Retail Law, the mass of an object was metaphorically compared to each city or town’s population size, while the distance between two points was measured along the road network (Reilly, 1931). The gravity model is now used to model everything from tourism demand to international trade flows (Carrere, 2006; Morley, Rosselló, & Santana-Gallego, 2014).

Gravity models are broken down into four-step models of transportation demand and analysis of transportation systems. Requiring extensive data collection, four-step transportation models are used to analyze study areas for transportation behavior. The study area is split into Traffic Analysis Zones (TAZs). Entry and exit of the area is monitored through external stations, locations that connects travelers whose trips take them outside the TAZs, that connect TAZs to the surrounding transportation network. All trip ends occur within the study TAZs or at an external station.

The four-step model (Figure 7) generalizes the trip-making decision process for individuals, determining overall travel behavior between origins and destinations. An origin is defined as the beginning location in a trip, and the destination is where the trip ends. The data required for this process usually originates from a household travel survey that provides socioeconomic, activity travel, and household vehicle data. The first step, trip generation, the number of trips that start in a TAZ, is used to model total travel volumes produced in an area. Factors that impact trip making such as origin population are entered into the model. The second step, trip distribution, allocates “generated” trips to destinations based on the attractiveness of destinations and travel
impedances. Travel time or cost to destination are examples of travel impedances that are commonly used. The third step, mode choice, allocates travelers to alternative modes based on expected volumes. The final step, route choice, assigns each trip on each mode to likely routes on the transportation network (McNally, 2008). Modeling trips that exit the TAZs is done by taking a percentage of productions and attractions based on counts at external stations. These percentages are often based on additional survey data used to determine trip purposes of these behaviors (Ruegg & Group, 1999).

The Four-Step Trip Model

Figure 6. The Four-Step Trip Model is used to model regional trip making by assigning mode and route choices based on the transportation network, demographic data, and trip generative and attractive ability of an area.
2.2.2. Trip Generation

Trip frequency was determined based on data extracted from road-based border-crossing points monitored by the Canadian government. Decades of data are available from Statistics Canada on traffic entering the U.S. at the Canadian border (Canada, 2017). Variables for each border crossing include vehicle type, length of stay, and country of citizenship. Figure 8 demonstrates aggregate car visitation to Vermont by Canadian Citizens, representing Canadian tourism travel. Of these visitors, a percentage are determined to stay within Vermont as tourists, and the remainder drive through Vermont to reach other destinations.

![Canadian Car Trips to Vermont by Length of Stay](image-url)

- Same day
- One night
- Two or more nights

47
Figure 7. Canadian-Vermont border crossings by both Americans and Canadians show similar travel patterns by length of stay throughout the year. However, volumes of travel by Canadians are about twice as high as Americans.

The road network from VCGI 2015 Road Centerlines was utilized to locate external stations, state border crossing points. Google Maps was used to generate possible routes to the largest population centers within the surrounding region. By visually examining routes through Vermont between Boston, Portland, and Montreal, external stations connecting Vermont’s to neighboring road networks were determined. Routes were generated using ArcGIS Network Analyst software, overlaid on a network dataset. The dataset was created using road centerlines of Vermont’s state highway system and US interstates within Vermont. This dataset was used to model routes taken by drivers. The volume of travelers allocated to each destination was determined based on data for expected drive through destinations as found by the Vermont Tourism
Report. Each trip was modeled using the same route to reach the destination and return to the origin.

The model isolates 14 external stations as entrances into the state along Vermont’s northern border (Figure XXXXXX), with 450,000 trips included in the model. Of these points, the northern I-89 entrance generates nearly 50% of all trips into Vermont. Montreal route 55 generates 20% of travellers as it turns into I-91. The remaining travel is split throughout the other 12 external stations along the Vermont-Canadian border representing minor roads, with each station receiving between 1-7% of all travel. Origin-destination pairs were isolated and examined for each of the 56 possible routes across Vermont. These routes represent the shortest distances that drive-through tourists are likely to take through Vermont.

2.2.3. Trip Distribution

Trip distribution uses known traveler behavior to allocate travel volumes to destinations. Traditionally, this is done through an iterative process based on demographic data collected in a household travel survey such as household size, number of vehicles owned, and income. Using initial distributions based on origin-destination surveys, these data were used to further adjust expected results. For this study, four destination external stations were selected along Vermont’s highway and interstate system, as I-89, I-91, I-93, and Vermont Route 142, based on common routes through Vermont, utilizing roads with high volumes of annual traffic and study results from the Vermont Tourism Report 2013.
2.2.4. Mode Choice

The Vermont transportation network offers routes for cars, trucks, cyclists, trains, and pedestrians. Canadian automobiles heavily dominate Canadian border crossings. Truck traffic remains consistent year round, while automobile crossing varies by time of year. This study focuses on automobile travel originating at these points. Automobiles make up 72% of all travel across these points (Figure 9).

![Canadian Vehicle Trips to Vermont by Vehicle Type](image)

**Figure 8:** Canadian vehicle border crossings are dominated by automobiles, which experience seasonal fluxes as tourism volumes increase in the summer months.

2.2.5. Route Assignment

An Origin-Destination matrix was created using trip generation data from each origin external station to each destination external station (see 2.2.3). The next step is to place these trips along the road network based on expected route choices. Based on the tangible benefits of travelling on faster routes such as time savings and navigational
ease, tourist routes were more heavily weighted toward high speed routes (Xiao, Yu, & Wang, 2012). Time-cost values were generated based on road’s speed limits, tourists were placed on the fastest route between origins and destinations after calculating speed over lengths of road segments included in each route.

Each route ends at an exit external station. Each of these external stations experience an average annual daily travel volume (AADT). The summed volume of the annual drivership on these roads represents all travel through these points. By examining each segment individually, it is possible to understand how travel tends to exit the state. Each segment was allocated a percentage of drive through traffic based on the percentage of total AADT that travels along it out of the 4 selected exit points. For example, the I-89 entrance with 200,000 travellers allocates 60% of its travel to the I-89 exit external station, 26% of its travel to the I-91 exit external station, 9% to the I-93 exit external station, and 4% of its travel to the VT 142 exit external station.

2.2.6. Implementation

External stations connecting Vermont to neighboring states and countries were isolated from the 2013 Vermont EmergencyE911_RDS dataset from VCGI. Within ArcMap, endpoints for each road segment were generated. Then a buffer of 50 meters was created from the edge of 2010 Vermont Town & County Boundaries dataset from VCGI. with the goal of covering all roads that cross the border. All road points found within this buffer were considered to be external stations connecting Vermont to other areas.
Road segments were classified by roadway type to generate a network dataset that could be used to determine routes between points. A closest facilities function in ArcMap’s Network Analyst extension was used to determine routing between each origin and each destination (14 origins, 4 destinations), returning 56 possible routes (Transportation, 2003). Each destination was assigned a volume of overall usage within the network based on annual drivers (Scott, 2013). The volume of travelers determined on each route were aggregated by road segment (Figure 10).
Figure 9: Canadian drive through tourist car volumes by road segment based on 2015 Canadian border crossing data allocated to 4 destinations along the Vermont road network. Largest volumes of travel are located along route I-89 before it exits Vermont.
2.2.7. Emissions Calculations

Carbon emissions from tourism transport were based on fuel types and average efficiencies of Canadian Vehicles. Fuel usage does not occur at a consistent rate so consumption rates have been classified as either city or highway driving within the light duty fleet. Emissions per liter of fuel are dependent on the type of fuel used, with diesel emitting at a higher rate than gasoline per liter, but traveling further with lower rates of fuel consumption.

Canada’s light duty vehicle fleet consists of 3% diesel vehicles, with the remainder gasoline powered (Natural Resources Canada, 2009). The 2009 Canadian Vehicle Survey listed Canadian fuel consumption of diesel and gasoline light duty vehicles as almost identical, at 10.7 and 10.6 liters consumed per 100 kilometers (Natural Resources Canada, 2009). The life cycle analysis software, GHGenius, was used to generate average carbon emissions rates per kilometer of travel. GHGenius was developed as an Excel spreadsheet program to calculate emissions from transportation in Canada, and has been used to examine impacts of shifting fleet composition on overall life cycle emissions. Consumption of fuel during vehicle operation on highways was reported at 5.83L/100km for diesel vehicles and 8.24L/100km for gasoline for Canadian vehicles in Quebec for the study year, 2016.

GHGenius calculates emissions based on driver behavior, specifically noting differences in fuel consumption of highway driving from city driving (Consultants, 2016). Due to the behavior associated with city driving, lower driving speeds with more
frequent starts and stops, city driving is much less efficient than highway driving per mile. As this study focuses on tourism traffic along highways, it utilizes 100% highway driving fuel consumption rates in its calculations. Thus the model assumes that these drive through tourists do not stop or deviate from their route while travelling.

After calculating mileage travelled by each type of vehicle and determining the number of liters of fuel consumed, total emissions were calculated. GHGenius lists Canadian diesel and gasoline CO₂-equivalent emissions of 2725.4 and 2219.2 grams per liter, respectively. Multiplying total liters used by these emissions rates allows for total CO₂ equivalent greenhouse gas emissions to be calculated.
Figure 10. Emissions calculations of automobiles by fuel type and kilometers travelled within Vermont by Canadian tourism drive through trips. Total emissions shown in grams of carbon dioxide equivalents. The vast majority of emissions are resultant from gasoline consumption.

2.3. Results

Tourism is known to play a large role in Vermont’s economy but tourism transportation within the state has not been previously quantified. Canadian drive-through trips are responsible for 0.7% of all miles travelled by car within Vermont annually. One drive-through trip is responsible for approximately 266 miles of travel within Vermont. Many of these routes overlap on road segments, and convergence increases as tourists move from north to south through the model. Summation of travel
volume on each road segment from each route is found in Figure 10. The majority of travellers travel on routes I-89 and I-91.

2.3.1. Emissions Calculations

In total, 6 million liters of gasoline and 131,000 liters of diesel were consumed by Canadian drive through tourists while traversing Vermont in 2015. These fuels combined were responsible for 13,673 metrics tons of CO₂-equivalent emissions within Vermont.

2.4. Conclusions

Vermont is a state that relies on tourism to support its economy. It also aims to reduce its carbon emissions drastically within the coming decades, with transportation generating nearly half of its total GHG emissions. Tourism requires transportation, yet Vermont has done very little to directly combat GHG emissions from this sector. In order to better understand tourism transportation in Vermont, additional research is required to map regional tourism alongside methods to promote growth while minimizing greenhouse gas emissions. Specifically, combatting a lack of knowledge of tourism origins, destinations, and trip rates would provide a basis for developing an argument for an adjusted transportation network.

Current data is limited, but it is estimated that 15% of all miles travelled on Vermont’s roadways are due to tourism travel. When attempting to isolate origins and
destinations of these trips the story becomes tangled making infrastructural suggestions difficult to generate. This study placed Canadian drive through trips on the Vermont road network based on known origins and prescribed destinations allocated by annual average drivership trends. Drive through trips were responsible for 0.7% of all miles travelled within the state, with each trip on average traversing 266 miles within Vermont. More drive through visitors enter Vermont each year than all public transit and rail passengers trips within the state combined (Sears & Glitman, 2011b).

Vermonters travel on average 221 miles each week (Sears & Glitman, 2011b). The Canadian drive through population is equal to more than a third of Vermont’s total population. With current tourist volumes, Canadians driving through the state are responsible for greenhouse gas emissions of more than a quarter of the Vermont population’s travel within a week annually. These 46 millions miles are just a small segment of the total tourism travel that occurs within Vermont, roughly equivalent to all of the drivers annually in Essex County. Other groups that must be accounted for are Canadians visiting Vermont, drivers from surrounding states, mileage of drivers from Vermont travelling within the state, as well as tourists flying to Vermont. This research acts should be viewed as a signal that these tourist behaviors do have impact that should be taken into consideration in future plans for the state. This fraction of Vermont total emissions is equivalent to all the greenhouse gas emissions of all of the oil burned for electricity within the state annually (En & Division, 2015).
Future tourism research should include an examination of tourist destinations within Vermont. Tourism trips within the state are responsible for more than half of all trips by Canadians into Vermont. Using expected volumes of travelers to specific locations; it is possible to improve the model to include travel within Vermont by Canadians. It would become possible to estimate the impact of additional tourist facilities on tourism traffic, and identify optimal locations for new tourist development, as well as quantify impact of select destinations on GHG emissions. American tourists from neighboring states are not included in this study, yet their travel behaviors impact the model. However, in order to include them in the model, extensive data collection would need to occur year round to account for fluctuations in travel volumes. There is no available information denoting Americans’ entrance into Vermont, as there are at the Canadian border.

Coupling transportation expenditures with local energy production is one of the driving forces of the Comprehensive Energy Plan. Funding for non-SOV travel needs to increase dramatically if other transport modes are to provide projected emissions reductions. If proven to have a large impact on Vermont’s emissions portfolio, specific actions must be taken to facilitate sustainable growth of the tourism industry. To facilitate tourist travel within Vermont without vastly contributing to emissions, Vermont must shift its funding towards promoting a regional transport network that will allow for tourists to move efficiently between common destinations. Furthermore, Vermont should utilize life-cycle analysis techniques when examining their carbon
emissions portfolio to ensure that carbon emissions aren’t externalized when making public infrastructure and program decisions that shape the landscape for years to come.

It has been shown that “emissions control policy is shaped by beliefs in technological progress and cultural values associated with technology”, rather than environmentalist beliefs of underlying behavior change (Maddison & Watts, 2011). The solutions expanded upon in the Vermont Comprehensive Energy Plan are primarily cornucopian in nature, stemming from an extremely narrow framing of the issue of GHG emissions leaving technological fixes as the only solution to the posed problems. Mitigating these emissions cannot be done by only targeting Vermonter, but must be viewed as a regional systemic issue with Canadians and drivers from other states impacting behavior within Vermont.
2.5. Citations

Canada, S. (2017). *Table 427-0001: Number of international travellers entering or returning to Canada, by type of transport monthly (persons).*


Credit, D. T. (2007). VERMONT ’ S ACT 183: SMART GROWTH TAKES ROOT IN zoning restrictions typically “maximum” density restrictions and serve to promote low- low-density urban sprawl has led to a shift in zoning in some development. Courts are likely to uphold the validity, 376(183).


REFERENCES


Canada, S. (2017). Table 427-0001: Number of international travellers entering or returning to Canada, by type of transport monthly (persons).


Credit, D. T. (2007). VERMONT ’ S ACT 183 : SMART GROWTH TAKES ROOT IN zoning restrictions typically “ maximum ” density restrictions and serve to promote low- low-density urban sprawl has led to a shift in zoning in some development . Courts are likely to uphold the validity , 376(183).


French National Travel Survey home page. www.cmh.acsdm2.ens.fr/enquetes/XML/lil-0081.xml


http://doi.org/10.1016/j.techsoc.2011.10.003


64
http://doi.org/10.1016/j.annals.2014.05.008
http://doi.org/10.1017/CBO9781107415324.004
Stopher, Peter, Camden FitzGerald, and Min Xu. "Assessing the accuracy of the Sydney Household Travel Survey with GPS." Transportation 34.6 (2007): 723-741.


