Big Data Analysis to Measure Delays of Long-haul Truck Trips

Kevin Gingerich (Corresponding Author)
University of Windsor
Cross-Border Institute
Department of Civil and Environmental Engineering
401 Sunset Ave.
Windsor, ON N9B 3P4
Phone: 519-253-3000 x 3737
gingerik@uwindsor.ca

Hanna Maoh
University of Windsor
Cross-Border Institute
Department of Civil and Environmental Engineering
401 Sunset Ave.
Windsor, ON N9B 3P4
maohhf@uwindsor.ca

William Anderson
University of Windsor
Cross-Border Institute
Department of Political Science
401 Sunset Ave.
Windsor, ON N9B 3P4
bander@uwindsor.ca

Submitted to TAC for **presentation** at the 2016 conference

Acknowledgements

The authors would like to acknowledge Transport Canada for financial contributions and loaning a temporary dataset containing the GPS data used in this analysis.

Introduction

Transportation delays are a commonplace occurrence for road users where congestion can cause substantial stress and loss in time for passenger car users. On the other hand, delays have a direct impact on a region's economic performance due to the added costs incurred by firms. This burden is amplified when delays are unpredictable, resulting in an uncertain arrival time for truck deliveries. The delays associated with truck deliveries can be attributed to a number of factors including congestion on major highways, custom clearance at border facilities, or unexpected events such as road accidents or inclement weather. To understand the nature of truck delays in North America, it is important to determine the time spent delayed in congested highway traffic and at the border if a truck is exporting goods between Canada and the US. Analysis of delays could help firms employ certain strategies (e.g. time-shifting) to optimize their trip duration given the time sensitive nature of goods movement.

This paper seeks to address the prevalence of delays with an emphasis on long-haul truck trips in North America. The analysis makes use of big data that represent the movements of over 56,000 Canadian owned trucks in the year 2013. These movements are captured with approximately 1.1 billion GPS pings. These pings track the geographic position of trucks at different points in time. Numerical algorithms are devised to identify the trips conducted by the trucks and calculate their travel time from the origin to destination. International trips crossing the Canada-US border at several major crossing locations in Ontario are examined first to determine the delays occurring at the border and the relative impact of this delay on the total trip. Next, the total delays observed for truck trips in the dataset are then calculated and compared against the optimal travel time in the absence of delays. Expected delays will be analyzed for different periods of the day based on the average travel time for trips between a given origin and destination to capture the impact of peak and off peak travel on truck movements. Finally, unexpected delays make up the remaining portion of delay that exceeds an average travel time. The results from the analysis will contribute to the ongoing efforts to improve freight fluidity in Canada.

Data Processing/Big Data

A precursor to the analysis of truck movements in this paper was the processing of raw GPS data into identifiable trips. A *trip* in this context represents a single leg journey observed of a truck from one location to another. Excluded from this analysis are *trip chaining* events where trucks connect multiple legs together, and *truck tours* (also known as milk runs) where the truck returns to a start location after performing multiple stops.

The raw GPS data exists as a dataset of individual pings revealing the location (latitude/longitude) of a vehicle at a given point in time while also providing anonymous identifiers for the truck and corresponding carrier. While not included with our dataset, GPS based datasets may also provide: a dilution of precision (DOP) measure as an indication of spatial accuracy; the speed of the vehicle; engine characteristics if the GPS unit is connected to the vehicle's electronic system; and the weight of the vehicle to determine full vehicles/empty backhauls and vehicle emissions. Moreover, some GPS systems utilize *dead reckoning* algorithms to impute the location of a truck if the GPS signal becomes too weak due to urban canyons or a minimal number of connected satellites.

Connecting the individual GPS points for a given truck together sequentially based on the known time stamps provides a method of observing the movements of individual vehicles.

In this analysis, we utilize a portion of a dataset of GPS pings covering the month of July, 2013. For this time period, the data includes 31.1 million pings corresponding to 30,000 individual trucks and 580 carriers. While the movements of the trucks occur across both Canada and the US, each of the carriers in the dataset is a Canadian owned company, therefore the movements do not include the behaviour of US based vehicles.

To identify the trip ends associated with the start and end of an individual trip, an algorithm was developed to calculate the dwell time of a vehicle. The dwell time provides a measure of the length of time the vehicle was stopped at a single location. Potential methods of calculating this dwell time include flags for engine start/stop or a speed threshold such as a vehicle travelling less than 5 km/hr. The former method was not available due to the lack of engine information while the latter method can present false positive readings if the slow travel movements are caused by congestion.

As an alternative to the previous two approaches, we utilized a tethered distance approach whereby the location of an initial ping (ordered sequentially by time) sets the tethered location. Subsequent pings within a specified radius of the tethered location lead to an accumulation of the elapsed time between pings as the dwell time of the vehicle. When a subsequent ping exits the radius limit of the tethered location, the dwell time resets to 0 and the new location of the tether is moved to the current ping location. We use a radius threshold of 250 meters with our dataset and a dwell time of 15 minutes or greater to denote an important stop event. The 15 minute dwell time is larger than some urban deliveries but acceptable in our usage to determine inter-regional and international trips. The larger dwell time also helps remove the potential of false positives due to congestion. The busiest corridor of Highway 401 in Toronto, Ontario was analyzed to confirm that false positives due to congestion are not occurring.

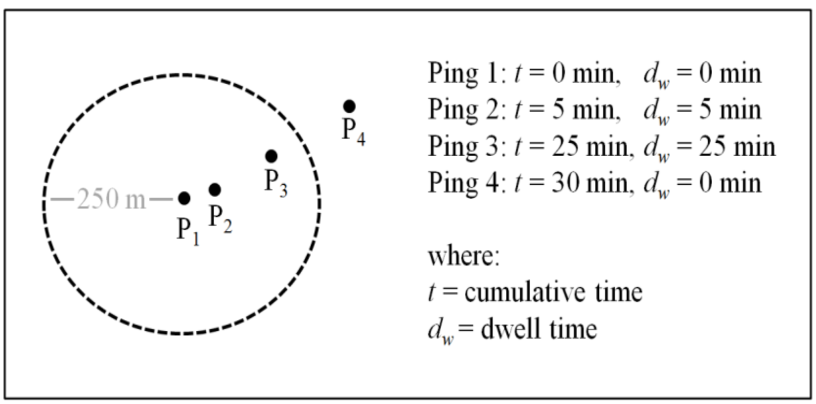

Figure 1: Distance based dwell time calculation
Source: Gingerich et al. (2016a)

An expanded description of the process applied to obtain trips from the GPS pings is provided in Gingerich et al. (2016b). The initial trips for the single month of July total 221,800. While these trips cover both the east coast and west coast and as far south as the US-Mexico border, the highest concentration of trips exists in Ontario. In particular, the Peel region observes the largest quantity of trips which is expected based on the high concentration of freight activity that occurs in that area. Finally, the travel time for any given trip i ($t_{i,travel}$) was calculated by subtracting the dwell time belonging to any intermediate stops ($t_{i,dwell}$) from the total elapsed time occurring between the trips start and trip end ($t_{i,elapsed}$).

$$t_{i,travel} = t_{i,elapsed} - t_{i,dwell}$$

<u>Border Delays</u>

The trips derived from GPS data allow us to observe a vehicle for the entire duration of the trip. This affords a valuable opportunity to observe delays at a single point in time for a trip as well as its relative impact on the total trip. For example, cross-border delays can be measured as a proportion of the overall travel time. The crossing time for an international trip at the Canada-U.S. border is measured here by utilizing virtual perimeters (*geo-fences*) that surrounds the international crossing and the corresponding customs plazas for both countries. The geofences for the border crossings are provided in Figure 2. Since each trip is derived from individual pings, an interpolation of the crossing time is necessary to estimate when the truck crossed the geofence perimeter. The outer zone in the Figure 2 ensures that the outside pings used in the interpolation are not located too far away from the site.

Based on the geofences in Figure 2, the crossing time at the border crossing includes the time necessary to make the crossing and the amount of time spent waiting at the customs inspection points located within the plaza. In this analysis, three border crossings are examined – (1) Ambassador Bridge connecting Windsor ON, and Detroit MI (2) Blue Water Bridge connecting Sarnia ON and Port Huron, MI (3) Peace Bridge between Fort Erie, ON and Buffalo, NY.

In the month of July, 2013, the observed total crossings in both directions at the Ambassador Bridge (AMB), Blue Water Bridge (BWB), and Peace Bridge (PCB) included 14,479, 6,187, and 8,387 respectively. While this forms a substantial dataset of truck crossing events, it is only a sample of the full population. According to the Public Border Operators Association (PBOA, 2016), the total truck crossings for the same month in 2013 included 175,258 (AMB), 125,182 (BWB), and 104,510 (PCB). Maps showing the origin of trips crossing the Ambassador Bridge are provided in Figure 3. These maps include a full year of data for 2013, but the pattern remains very similar from month to month.

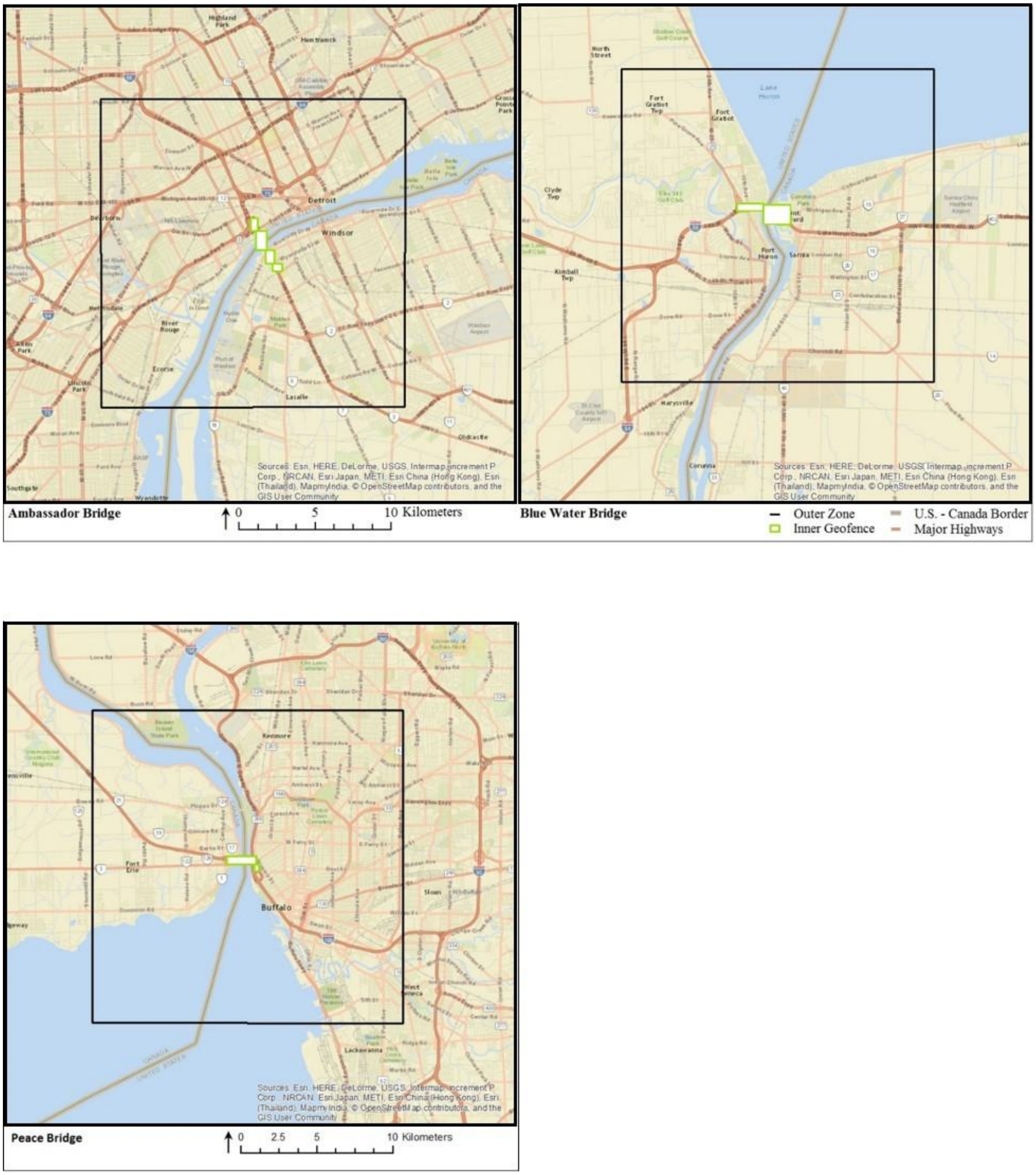

Figure 2: Border geofences (*Source: Adapted from Gingerich et al. 2016b*)

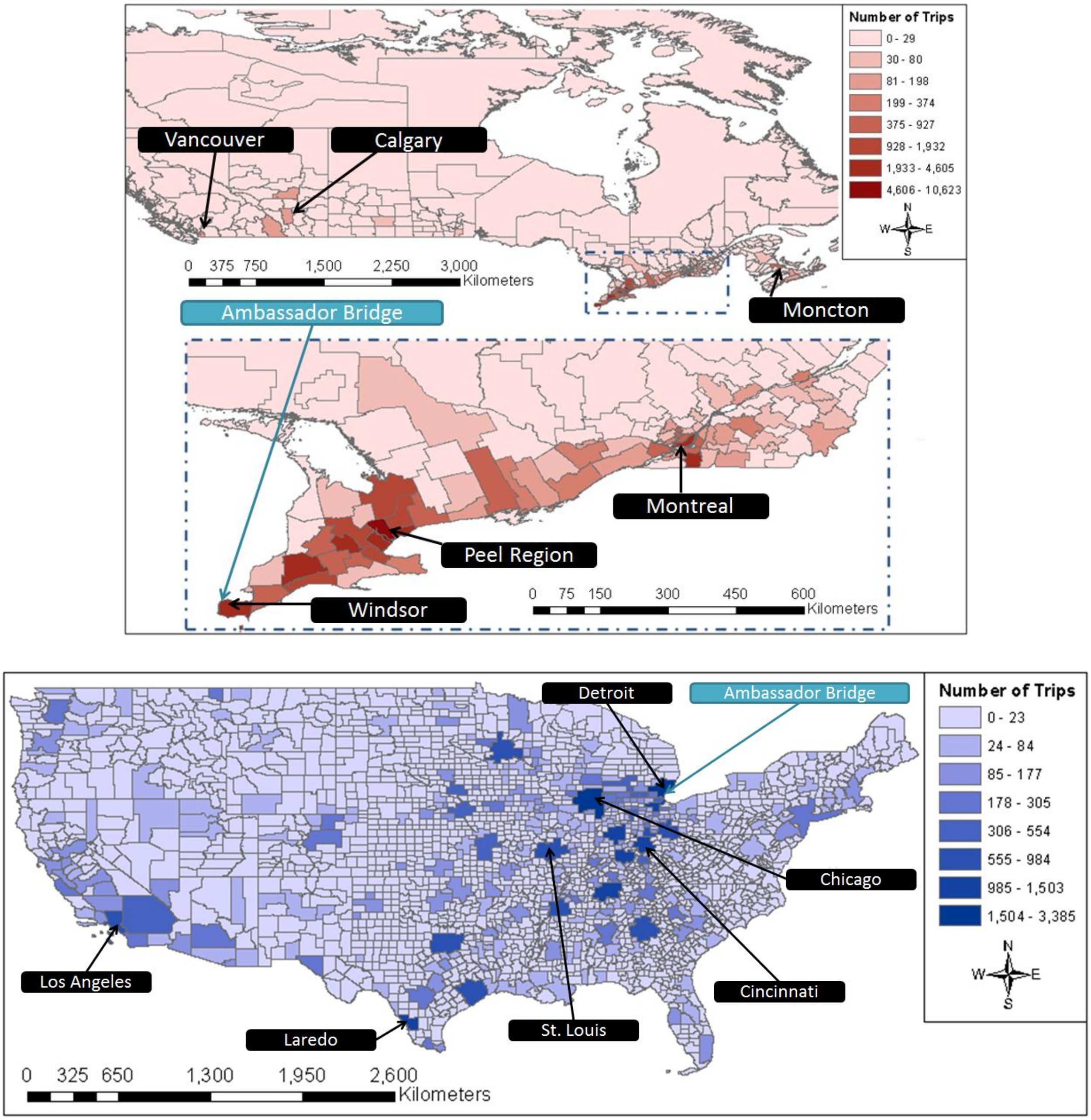

Figure 3: GPS trips crossing the Ambassador Bridge (full year - 2013) in Canada (top) and the U.S. (bottom) (*Source: Gingerich et al. 2016b*)

Statistics on the border crossing times are given in Table 1. The median border crossing time at the AMB, BWB, and PCB based on the observed crossing times includes 13.6, 11.3, and 12.3 minutes for vehicles headed to Canada[1]. Moreover, the 95 percentile crossing time to measure the spread of crossing events is 42.3, 48.7, and 47.6 minutes for trucks headed to Canada[1]. These statistics are indicative of border crossings – a relatively short average crossing time but the potential for a much larger wait. A late arrival of products caused by delays can have a high cost associated with them. Moreover, since the cost of late shipments is typically much higher than the cost of an early shipment, carriers may anticipate a crossing time such as the 95 percentile (Anderson and Coates, 2010).

Table 1: Border crossing times

	Ambassador Bridge		**Blue Water Bridge**		**Peace Bridge**	
	CAN bound	**USA bound**	**CAN bound**	**USA bound**	**CAN bound**	**USA bound**
Median	13.6	14.3	11.3	13.1	12.3	22.0
Average	17.6	18.9	16.8	18.2	17.6	27.1
95 Perc.	42.3	48.3	48.7	48.0	47.6	69.2

The actual delay at the border is then calculated by identifying the minimum travel time necessary to cross the border (3.35 min AMB, 4.61 min BWB, and 2.58 min PCB) and subtracting this value from each crossing time. However, the observed border delays are only a portion of the typical travel time for a truck. For a short distance trip, it is expected that the delay is a relatively large portion of the total trip. Conversely, a long distance trip will typically observe a much smaller overall impact from border delays as a proportion of the total trip time. From the GPS dataset used in this paper, the border crossing times were connected to the full trips to provide information on the observed impact of delays on the entire trip. This process resulted in 10,316 trips that utilized the Ambassador Bridge, 4,001 trips that utilized the Blue Water Bridge, and 3,001 trips that utilized the Peace Bridge[2].

Statistics on the relative impact of a border delay on a trip is based on the travel time for a given trip and its corresponding border delay. For Canadian bound trips, the delay represents an average of 2% (AMB), 2.5% (BWB), and 3.2% (PCB) of the total trip travel time. Due to higher delays for U.S. bound trips, the average increases to 2.5% (AMB), 3.6% (BWB), and 4.4% (PCB). A cross-section of the number of trips with a corresponding proportion of delay is shown in Figure 4 for Canadian bound and U.S. bound trips. As seen in those figures, the majority border delays only make up 0-5% of the travel time. However, a noticeable quantity of trips exhibit border delays encompassing 5-10% of their travel time.

While the proportion of trips with delays accounting for over 10% of trips is minimal (>15% for the U.S. bound traffic), the substantial border delay relative to the total trip will have a much larger impact on their costs. This is particularly true if their extensive delays lead to additional fees associated with late deliveries. Moreover, while many of the trips did not see substantial border delays, the *potential* for delays observed by a small number of trips likely caused all carriers to

[1] These values ignore crossing times above 90 minutes by assuming they resulted in secondary inspections
[2] The final trip counts are less than the number of total crossing events since there is not necessarily enough information in the GPS data to form a complete trip.

provide an extra buffer time (or hold extra inventory) that increases the indirect costs arising from uncertain border delays. While the actual costs of these delays cannot be easily estimated with the GPS data alone, Brown and Anderson (2015) estimate the extra ad valorem costs (as a percentage of the total value of goods) for carriers conducting cross-border trade between Canada and the U.S. varying on average between 0.4% to 0.9% (though this result only includes direct transportation costs).

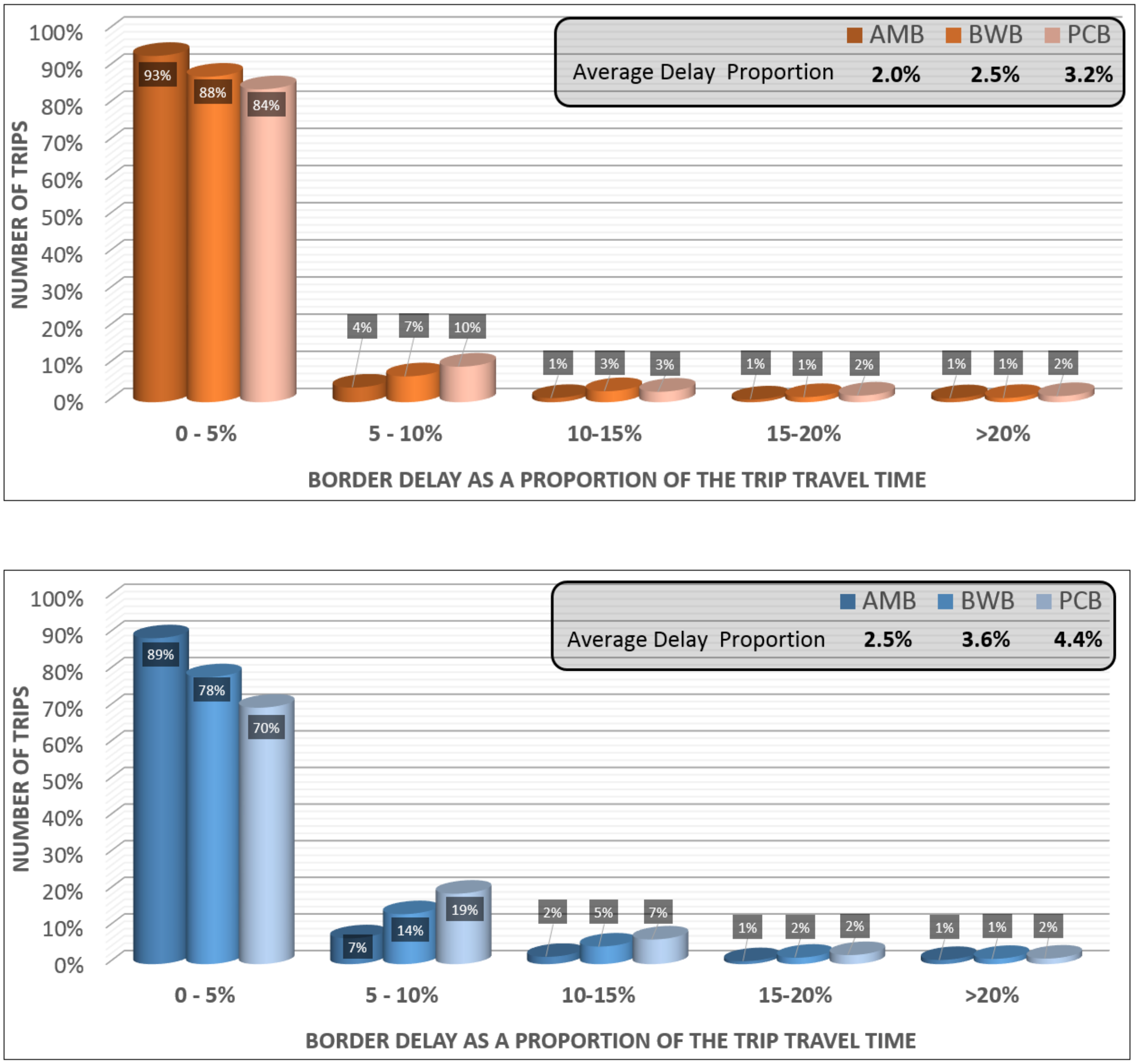

Figure 4: Border delay as a proportion of the trip travel time for Canada bound (top) and US bound (bottom) trips

Overall Trip Delays

Apart from border delays, the road system exhibits delays from other sources including commuter congestion on weekdays, construction activities, collisions, and extreme weather phenomenon. Delays caused by daily work commutes represent time that commuters expect to spend in congested traffic based on a given hour of the day with increasing delays during the morning and afternoon peak periods. While not commuters themselves, freight movements are similarly impacted by the temporal patterns of commuter traffic due to the shared nature of most major roads. Long-term construction projects also result in an expected additional travel time. By contrast, events such as traffic collisions and weather events are more difficult to predict at any given time. Their effects on travel time lead to generally unexpected delays.

Trips derived from GPS data in this analysis were organized by origin-destination paired zones. These zones represent census divisions in Canada and MSA zones in the U.S. (and counties filling in gaps between MSA zones). The time observed for a trip can be separated into three categories based on Figure 5. The first segment represents the proportion of the trip where no delay occurs. This was originally measured based on the minimum travel time identified for any trip between the origin and destination zones. However, the area and spatial configuration of a zone erroneously biased the results (i.e. a trip to the close side of a zone will take less time than a trip to the far end of a zone). A revised method was introduced to determine the maximum average speed resulting from any trip for a given OD zone pair and utilizing this speed to determine the minimum possible travel time.

A trip that exceeds the minimum travel time accrues time attributed to delays. Expected delays begin to accrue as a trip exceeds the minimum travel time. The maximum value for expected delays is set to the average travel time for trips between a given origin-destination zone pair for a given period of the day. This enables the measure to account for the expectation arising from peak period travel. The trips are categorized based on two hour stretches of the day (i.e. midnight to 2 AM, 2 AM to 4 AM, etc). This provides a larger opportunity for trips to be observed compared to hourly categorization.

As a trip exceeds the average travel time for the given origin-destination pair and time of day, it begins to accrue unexpected delays. As a result of these definitions, a trip can consist of (1) free flow travel (no delay), (2) free flow travel plus expected delays, or (3) free flow travel plus expected and unexpected delays.

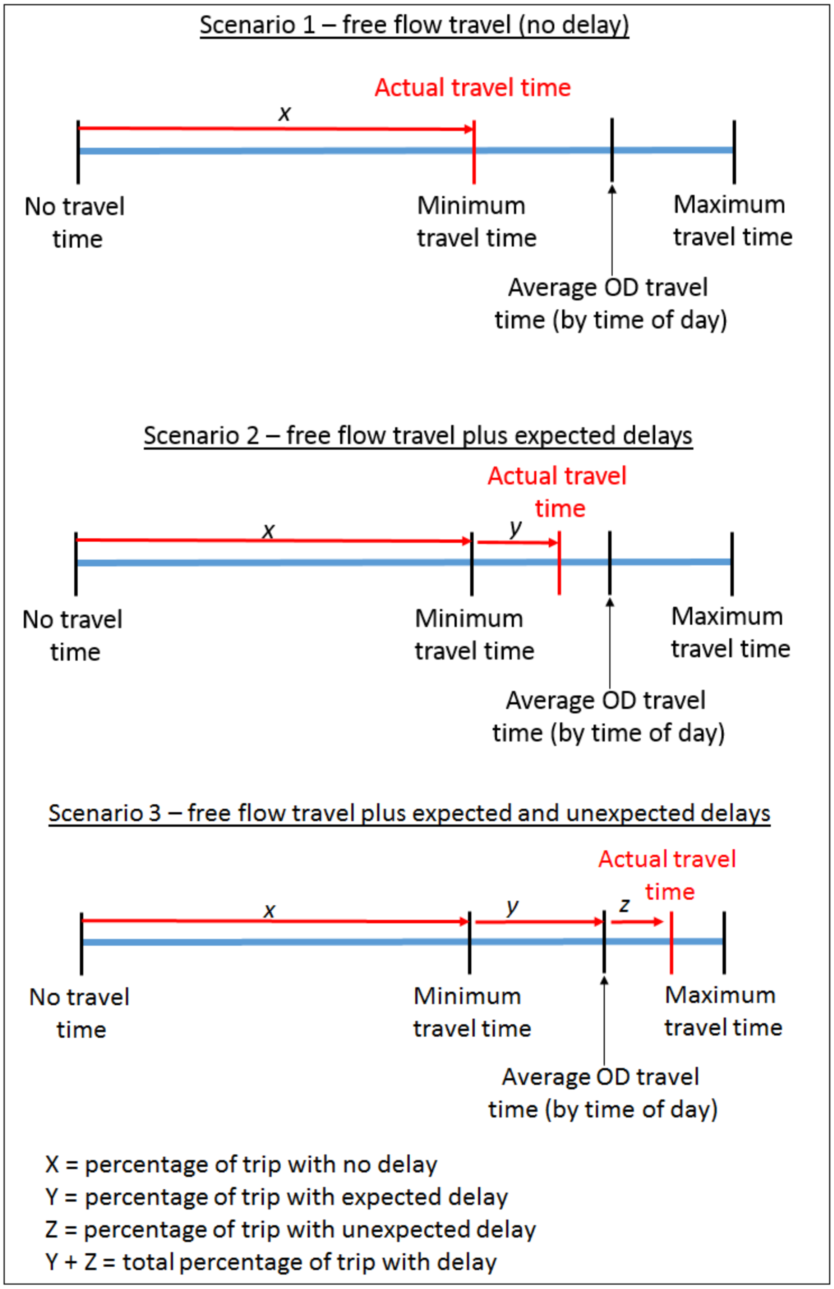

Figure 5: Delay as a proportion of full trips

The full trip dataset from the one month of GPS data resulted in 221,807 trips. However, due to the necessity for a suitable size of trips of any given origin-destination pair to assess the three definitions of delay above, only OD pairs with at least 40 trips were utilized. This resulted in a smaller dataset of 83,654 trips belonging to 756 OD pairs. The average proportion of travel time from these trips associated to free flow travel, expected delays, and unexpected delays is 75%, 19%, and 6%. Therefore on average, 25% of the travel time for the trips in the GPS dataset was caused by some form of delay.

Maps showing the proportion of trip delays are averaged for the zones based on the origin of trips and destination of trips in Figure 6. A ranking of the zones based on average of the origin and destination values for a given zone was tabulated to determine the zones with the highest and lowest delays based on 148 zones with enough data available. Overall, the average delay values for individual zones ranges from 10% to 34%. Two outliers with average delay proportions above 50% were removed since they were also inconsistent between origin and destination results.

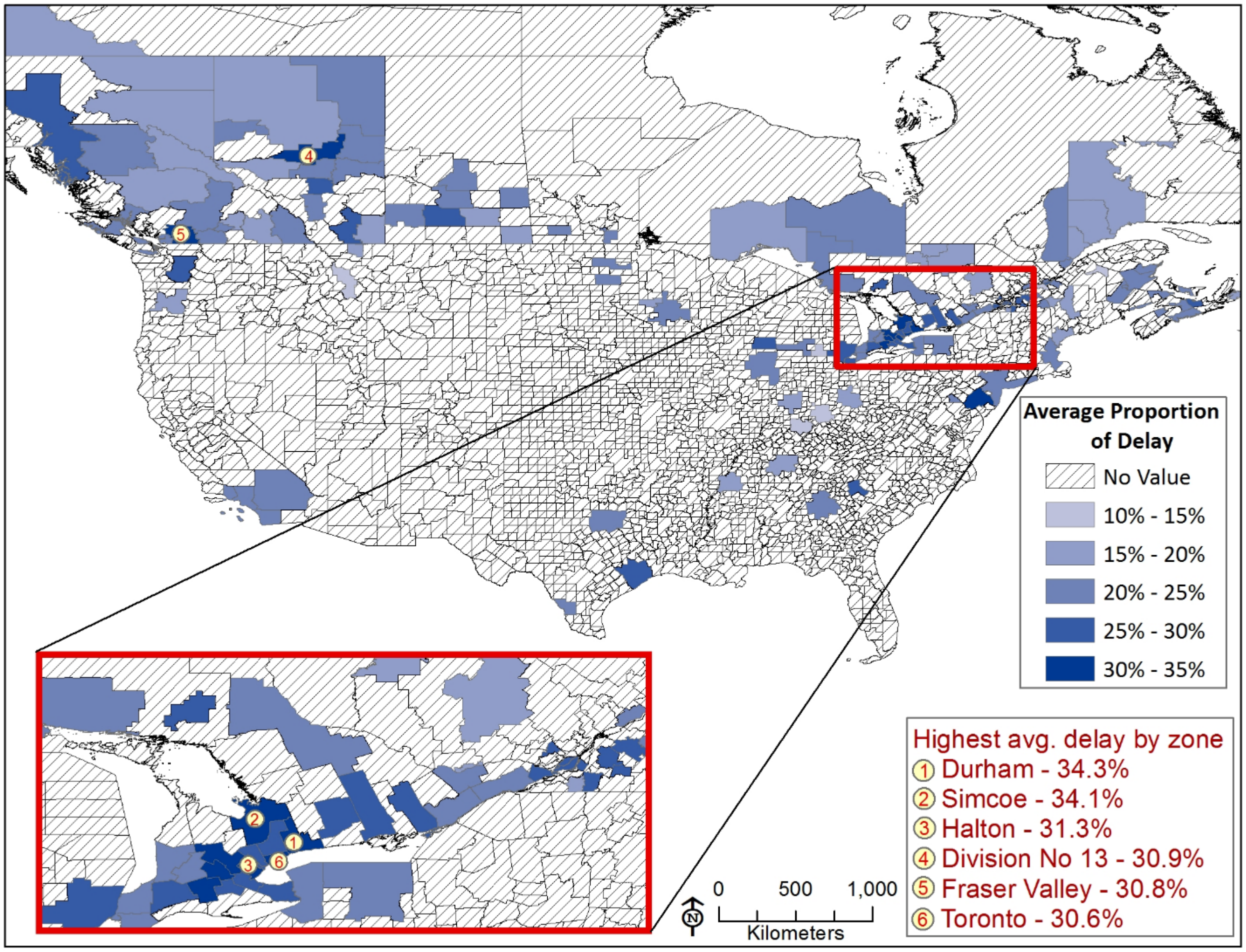

Figure 6: Average trip delay proportion by zone (averaged based on results from origin/destination end points)

Conclusions

This analysis provides an overview of the delays based on the origin/destination zones for trips. Mining GPS data observing the movements of trucks allowed for the calculation of delays across an entire trip. These delays were first applied to several border crossings to determine the effect of delays on the total trip travel time. The results found that border delays at three crossings resulted an average of 2% (AMB), 2.5% (BWB), and 3.2% (PCB) of the total trip travel time.

The overall delays for truck trips were also deconstructed into the proportion of a trip with no delay (75% average), expected delay (19% average), and unexpected delay (6% average). The results were averaged for given origin and destination zones to provide a visual example of the results. The advantage of this approach is to provide a starting point to determine areas with larger travel delays. However, the origin/destination zones for trips are not necessarily where the trip delays occur since they may appear at any point along the routes between the origin and destination.

The next steps in this analysis are therefore to look at origin/destination pairs with large delays, and examine the entire routes to determine where issues with delay are occurring most frequently. The advantage of the GPS data is that the movement of the vehicle along the routes are provided, lending itself to a more microscopic analysis of routes. For example, the speeds of trucks in the GPS dataset along segments of Highway 401 in Ontario from the Michigan border to the Quebec border are shown in Figure 7. This figure highlights the effect of traffic congestion approaching/within Toronto on Highway 401 (in the middle of the figure). A similar approach for the routes of any origin/destination pairs could be utilized to determine where delays are the most problematic.

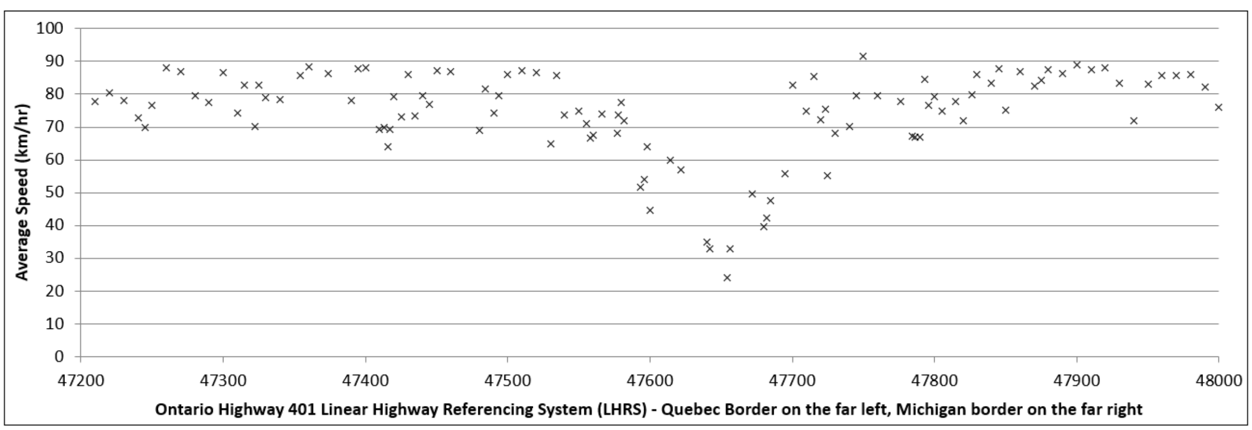

Figure 7: Average truck speeds derived from GPS data along Ontario's Highway 401

<u>References</u>

Anderson, W., Coates, A. (2010) Delays and uncertainty in freight movements at Canada-US border crossings, Proceedings of the *45th annual Canadian Transport Research Forum (CTRF) Conference*.

Brown, M., Anderson, W. (2015) How thick is the border: the relative cost of Canadian domestic and cross-border truck-borne trade, 2004-2009, *Journal of Transport Geography*, vol. 42, pp. 10-21.

Gingerich, K., Maoh, H., Anderson, W. (2016a) Classifying the purpose of stopped truck events: an application of entropy to GPS data, *Transportation Research Part C*, vol. 64, pp. 17-27.

Gingerich, K., Maoh, H., Anderson, W. (2016b) Characterization of international origin-destination truck movements across major Ontario-Michigan border crossings, *Transportation Research Record*, vol. 2547, pp. 1-10.

Public Border Operators Association – PBOA. (2016) *Traffic Data* [online] Available at http://www.publicborderoperators.org/index.php/traffic. Accessed on May 2, 2016.