COMPARING FLOATING CAR DATA AND CARSHARING GPS DATA FOR TRAVEL TIME ASSESSMENT

Pierre Loustau, M.Sc.A., École Polytechnique de Montréal

Vincent Grasset, M.Sc.A., École Polytechnique de Montréal

Catherine Morency, Ph.D., Assistant professor, École Polytechnique de Montréal

Martin Trépanier, Ph.D., Associate professor, École Polytechnique de Montréal

INTRODUCTION

With the increasing problems of traffic congestion in urban areas, transportation planners need better tools to assess its evolution through the years. Travel time measures on road networks provide key information to identify critical spots of congestion and evaluate the scale of this phenomenon across the area. Many technologies are currently available to measure travel time at specific locations (license plate matching, loop detectors, bluetooth device matching). Traditionally, at larger scale, travel time is estimated with the help of mandated floating cars (probe vehicle) that run through the road network. Nowadays, ad-hoc –or randomly collected—data can be obtained from GPS devices aboard individual cars and commercial vehicles. However, each of these ways of collecting data suffers from limitations; cumulating the information they provide seems like a relevant approach and this starts by comparing the measures they can output.

This paper presents a comparison of travel time measure between two data collection means for the same locations in an urban area: GPS and floating cars. GPS data come from a carsharing company operating in the Montreal, Canada area. GPS traces are collected anonymously in some of the vehicles, without any path supervision (members were doing their usual trips). On another hand, a specialized firm operated floating car data on fixed routes. The Quebec Ministry of Transportation process has collected this data through the years in the context of a travel time evaluation. For modelling purposes, floating car routes were fragmented in one kilometre road links. Incidentally, some carsharing vehicles equipped with GPS devices were driven on parts of the floating cars routes. On these road links, GPS traces coming from carsharing vehicles are superimposed to floating car data and statistical analysis can be conducted.

This study is part of a wider project that tries to assess the evaluation of congestion in the Montreal area. It aims at enhancing the understanding of traffic congestion through the development of new indicators and modelling tools.

The paper is organised as follows. First, some background information will be provided with respect to concepts related to traffic analysis and research conducted in the field. Then, the databases used will be described, along with the case study and comparison methodology.

The results will then be presented and discussed and perspectives will be provided based on these results.

BACKGROUND

Multiple studies have been conducted on the negative impacts of congestion and it is an ongoing research issue. Besides, the current sustainability issues divert many studies towards the indirect impacts of congestion, namely environmental impacts (Nesamani et al. 2005), equity and accessibility issues or even the impacts of congestion on activity rhythms of households and people, for instance the reduction in the number of shopping trips (Schmöcker et al. 2006). At the same time, multiple studies are being conducted specifically on the assessment of the sustainability level of various transportation projects. Litman (2008) proposes a series of indicators allowing a better understanding of the role of transportation infrastructures, projects and politics in order to reorient the current planning process. He actually recommends that every transport indicator includes economical, social and environmental impacts. Indicators such as travel times between home and work location and delays due to congestion are identified as economical indicators where a reduction is perceived as a positive contribution to sustainability.

The need to better understand the problems related to traffic, coupled with the increasing availability of information technologies, resulted in an increasing availability of various types of travel data. Increasing availability and performance of intelligent transportation systems have actually contributed to this state. Many researches have recently focused on data processing and analysis challenges and the definition and estimation of relevant indicators to describe the traffic conditions. In the USA, the works from the Texas Transportation Institute (Schrank and Lomax 2003, 2005, 2007), propose various indicators describing the state of traffic in the main areas. For instance, they observe that, during the last twenty-five years, congestion has increased. Actually, household car ownership combined to urban sprawl, have contributed to the growth in recurrent congestion. Figure 1 shows the evolution of congestion (average annual delay for every person using motorized travel in the peak periods in 439 urban areas in the USA) from 1982 to 2007 in relation with the scale of the area. Globally, the average delay is 36 hours and there are 9 urban areas with delays higher the 50 hours.

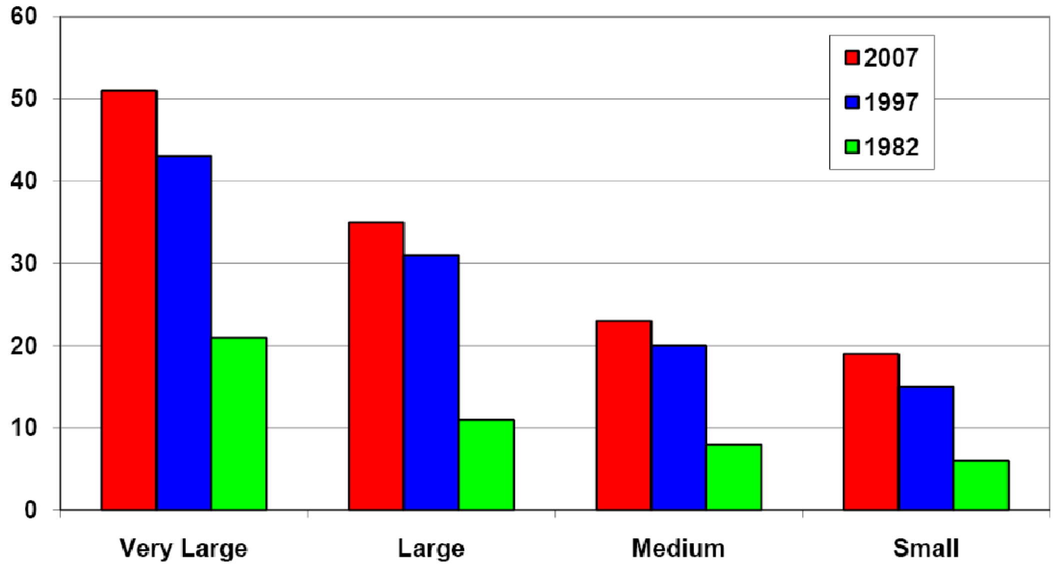

Figure 1. Evolution of congestion in the USA in relation to the scale of the area (Schrank et Lomax, 2009)

Challenged with the problems of congestion and sustainability, the institutions are less and less heading towards the reduction of congestion but are rather aiming to increase the reliability of travel times on their infrastructures. This new concern is leading to the development of new indicators and consequently the need for new and better data to assess both mean and variability of travel times in varying conditions (day of the week, period of the day, road conditions, etc.).

In the same spirit, FHWA (2004) puts a lot of effort into the development and estimation of traffic congestion indicators through its Mobility Monitoring Program that relies on electronic devices available in some thirty areas. Another interesting contribution is the one from Pennsylvania DOT (Szekeres and Heckman, 2005) that implemented a Congestion Management System for statewide transportation planning.

OBJECTIVES

This research is part of a project aiming to define and estimated travel time reliability indicators on the highway network. One part of the project also aims at evaluating the potentialities offered by various travel time data collection technologies to provide relevant and continuous data to enrich or replace current approaches. GPS traces are one of these technologies. Currently, the Quebec Ministry of Transportation relies on floating cars data to assess travel times on their highway networks. These kinds of data are relevant but costly and, due to sample size and structure, they are often insufficient to assess the impact of various attributes on travel times (month of the year, day of the week, road conditions, weather, etc.).

Using the floating car method, travel time data were collected from 1998 to 2004 on the main highways of the Montreal region. These travel time data were systematically split into onekilometer segments and significant simulation models were developed on the basis of travel time frequency distributions. Using these models, it is possible to estimated mean travel time on various highway segments as well as the variability of these travel times. This paper builds on previous modelling work using floating cars data and (see Loustau et al., 2010) and seeks to assess the value of GPS traces coming from shared cars to enrich the currently available travel time data or eventually replace them. The first step is hence to compare travel time distributions obtain using the two data collection approaches to see if they typically fit. This is the main objective of the work presented in this paper. More specifically, this analysis aims at estimating, for some highway segments sampled using the two approaches, the mean and variability of travel times. This will allow to assess the relevance of ad-hoc GPS traces, not collected for measurement purposes, to provide continuous information on travel times, as part of a continuous traffic monitoring program.

INFORMATION SYSTEM

This section brings some definitions relevant to this study, and exposes the two different datasets that were used: floating car data and GPS traces.

Definitions

Traffic congestion is easy to observe because it reflects the saturation state of a road or a network. The accumulation of vehicles and the increase in queue length are good signs of congestion, but let us define the concept more precisely. The congestion is defined as an excess volume of traffic on a portion of a road or traffic lane given the capacity of the road. This phenomenon can be due to the following seven main causes (Schrank et Lomax, 2005):

- Saturation of the road because of insufficient capacity: the capacity is function of the number of lanes, lane width, presence of intersections, curves, bottlenecks, etc.
- Accidents of the road that reduce the capacity by blocking or perturbing lanes.
- Work zones along the road. The capacity of the road can be temporally reduced because of lane shifts, reduced width and lower speed limits.
- Weather conditions can influence the capacity because of lower speeds or block lanes caused by low visibility and changes in road surface.
- Traffic lights and other traffic control devices can be badly designed or their malfunction can cause capacity reduction, they can also cause instability in the network (like railroad crossings, for example).
- Special events like shows, sporting games, and commercial sales can momentarily increase the volume on a road.
- Daily variations in transportation demand can also cause such perturbations.

These factors can interact between each other, and they can also be additive. Accidents could be caused by bad weather, for example.

The reliability of travel time is typically linked to its variability. The seven preceding factors will influence this variability. We will, in this paper, define two types of traffic congestion:

- The recurrent congestion is related to the normal use of the road network and will be likely caused by variations in daily demand.
- The non-recurrent congestion is due to random and unpredictable events like accidents, sporting events.

The recurrent congestion is observed during peak hours. It is usually well known by the commuters and easier to predict. However, non-recurrent congestion can happen more often on some road links; commuter will then consider both recurrent and non-recurrent congestions when they choose their road itinerary to destination. Sometimes, a commuter will choose a longer road but with a more reliable travel time, instead of taking the shorting path and fall in non-recurrent congestion. Travel time reliability can really influence the choice of the commuters.

A typical indicator of travel time variability is the coefficient of variation (ratio between standard deviation and mean). But there is also the Travel Ratio Index:

$$TRI = \frac{T95 - T50}{T50}$$

where:

- T95 is the time of the 95th percentile of the observed distribution
- T50 is the median time in the distribution

Floating car data

From 1998 to 2004, the Ministry of Transportation of Quebec has collected travel time dat using the floating car method. Floating cars were used to measure the travel time on 55 different routes in the Greater Montreal Area (see Figure 2), at different times of the day (mainly peak periods), days of week and seasons. A total of 29,229 measurements are found in the database, covering the 7 year-period: 51 months of observation, no observation in 2003. The 55 routes cover a total of 800 km, mainly highways. For each observation, a series of time and location stamps are available, based on the floating car odometer inked to a precise recording device. Therefore, the travel time can be measured at very good level along the route.

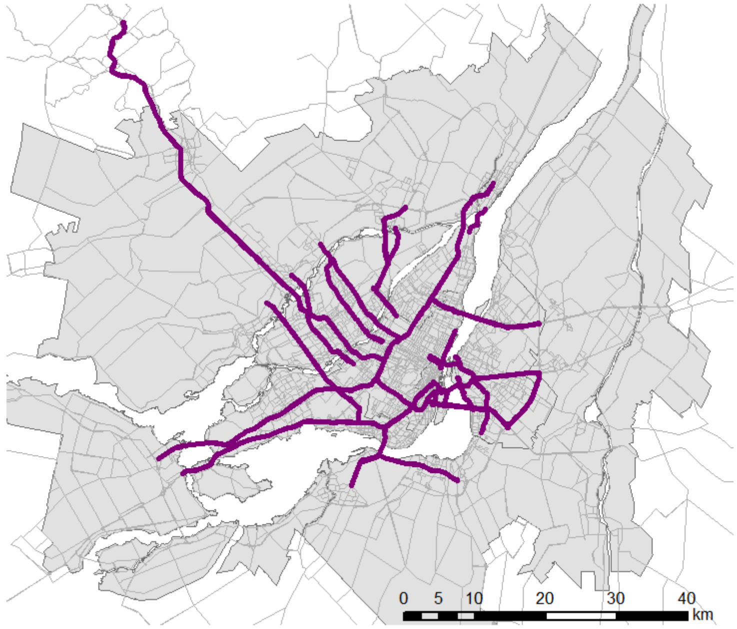

Figure 2. Map of the floating car routes

The raw dataset has been used in several preceding studies (Tremblay et Robitaille, 2005; Robitaille et Nguyen, 2003, Babin et Gourvil, 2007; Babin et al. 2007; Babin et al. 2006). Recently, a new methodology has been developed to better analyze the data; routes were segmented in one-kilometre segments. Each segment can be characterized with its own travel time (Loustau et al 2009). This allows refining the analysis: frequency distributions of travel time were calculated for individual segments. Using a clustering technique, segments were then aggregated based on the similarity of frequency distributions of travel times, for AM and PM peak periods (Loustau et al., 2010). Models combining three log-normal laws were developed for each cluster of segments and used for simulation.

GPS data

The second dataset is composed of GPS traces collected aboard vehicles from the Communauto carsharing company in Montreal. This can be qualified as indirect data collection compared to the floating car direct collection, because in the case of carsharing, the vehicles are not bounded to given routes. But the carsharing users will sometimes travel on the same road links that are found in the floating car routes, so we have here a base of comparison for the two types of data collection.

The database contains traces collected in 53 carsharing vehicles between March 2006 and March 2008. This data is randomly collected and therefore is not bounded by any sampling plan. Only the road segments with high volume of data will be retained in the analysis. The on-board GPS in carsharing vehicles is not primarily made for travel time measurement: the interval of time between location points can vary from 6 to 287 seconds, so interpolation is made necessary for some observations. As a matter of fact, because of communication costs, the carsharing company has reduced the number of sampling points through the years. For this reason, the observation of verified localized phenomenon with this dataset could be biased.

Today, more than 400 cars are equipped with GPS devices; hence, the probability of observing cars travelling on critical sections of the network is higher than for the period of time for which data were available at the time of the research.

Examples of frequency distributions of travel time on one-kilometre segments are presented in the next figure (Figure 3for five segments belonging to the route that is used for demonstration purposes. These distributions exemplify some types of frequency distributions that can be derived from GPS traces and observed in the area. For instance, segment 20 shows a concentrated distribution with a high peak at around 35-40 seconds per kilometre (≈100 km/h) revealing that this segment often provides free-flow travel conditions. Segment 30 is however more congested (peak at around 60-65 seconds per km – 57 km/h) and less reliable since the distribution if more dispersed.

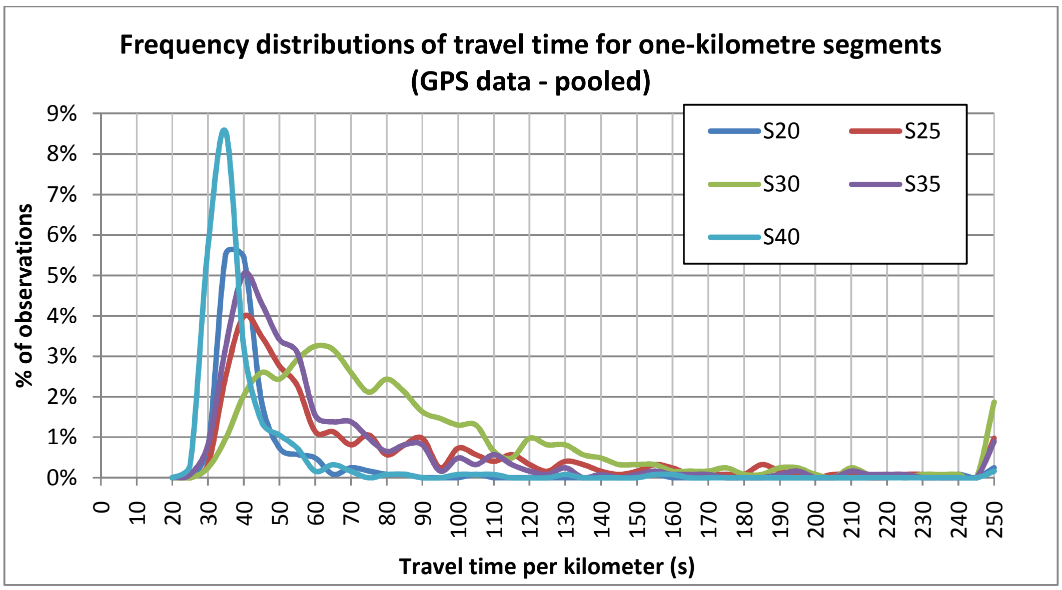

Figure 3. Examples of frequency distributions of travel times for one-kilometre segments of A-40 (westbound) – pooled GPS data

ANALYSIS

To point out the fact that it is clearly useful to compare the two kinds of datasets, an example is exposed using the Autoroute 40 that crosses the Island of Montreal from east to west. This circuit is 68 km long, and consequently is composed of 68 segments. This highway crosses four main highways, respectively the A13, A15 (twice) and A25 (Figure 4). As can be seen on the map, highway 15 merges into the A-40 for some 2.5 km before splitting back into separate highways. This combination of flows often translates into important congestion at many periods of the day and in both directions.

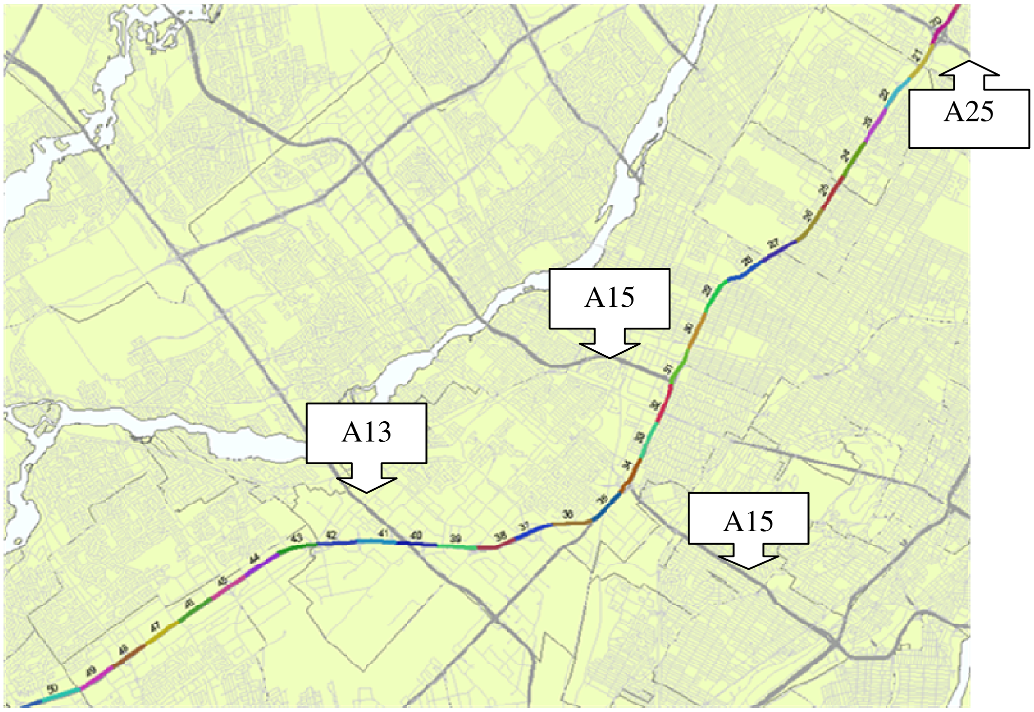

Figure 4. Highway 40 from east to west.

Sample size consideration

In order to assure accurate results, the size of the sample is important. GPS data are available from 2006 to 2008. However, some months of observation do not provide a sufficient sample and will hence not be considered. Figure 5 summarises the available sample, by month of observations. Only months with sufficient sample sizes will be used in the main analysis.

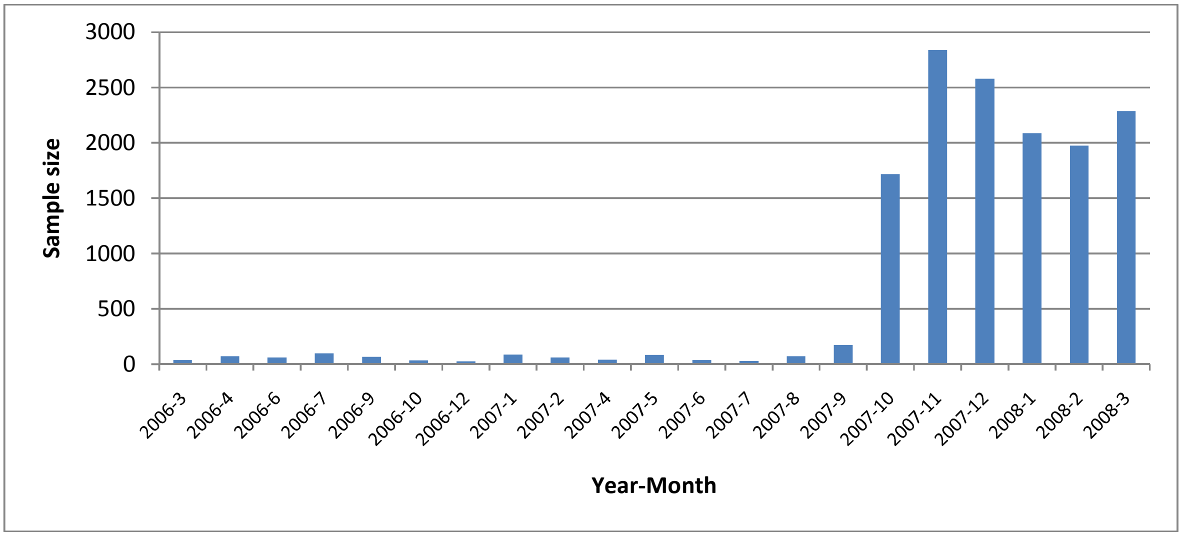

Figure 5 Sample size per month of observation

Means and variation of travel times

The first step of the process is the comparison between indicators estimated using the two datasets. For this first analysis, all available data are pooled: floating cars combining data from 1998 to 2004 and GPS traces combining data from 2006 to 2008.

Figure 6 first presents the mean travel time per segment (one-kilometre segment) estimated using both datasets. It is obvious that both datasets capture the same zones of congestion (segments with high travel times: 23 to 35 with a peak in the A-40-A15 merge zone). Values are also quite similar, in the western part of the route (segment 40 and subsequent). Higher differences are observed in the eastern part of the route but the overall pattern is still similar. At this point, it is hard to know if differences are due to technology or evolution of traffic conditions over time. Figure 7 presents the coefficient of variation of travel time on the same segments, again using both datasets. Some segments are similarly characterised by both datasets and the overall trend, from east to west, seem similar. Actually, as a first indicator of comparability, the correlation coefficients between results from the two datasets are respectively 0.83 and 0.43 for the mean and coefficient of variation. Mean estimations are hence quite correlated and the correlation is also relevant for the coefficient of variation.

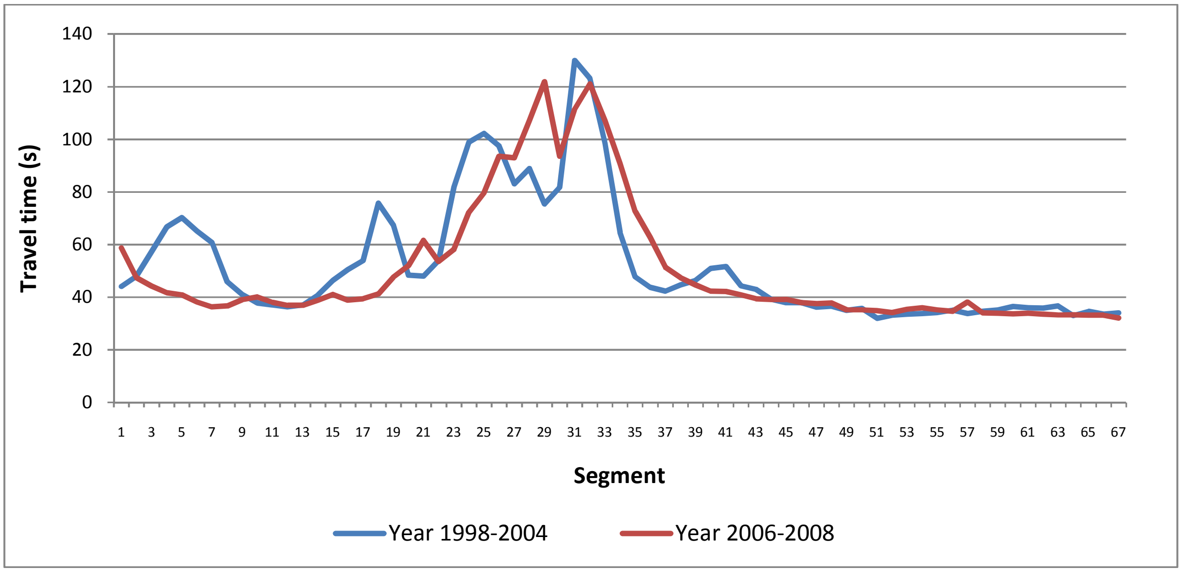

Figure 6. Means travel times estimation using the two datasets: GPS vs floating cars

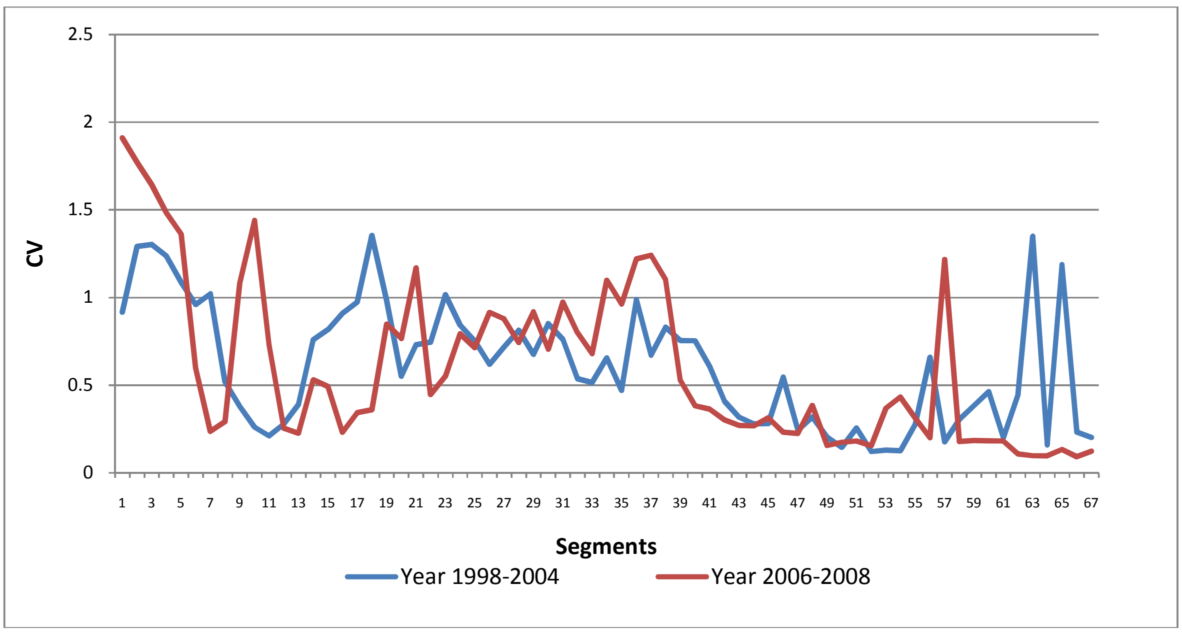

Figure 7. Travel times coefficients of variation of estimation using the two datasets: GPS vs floating cars

Difference between years

Since estimations of mean travel times seem to be consistent between the two technologies, we are using them in sequence to observe the evolution of this attribute throughout the years. As can be seen in Figure 8, showing mean travel times for the eight years of observation, it is clear that the segments which were subjected to congestion (especially segments 23-25 and 31-33) present, for years 2007 and 2008, high means of travel times.

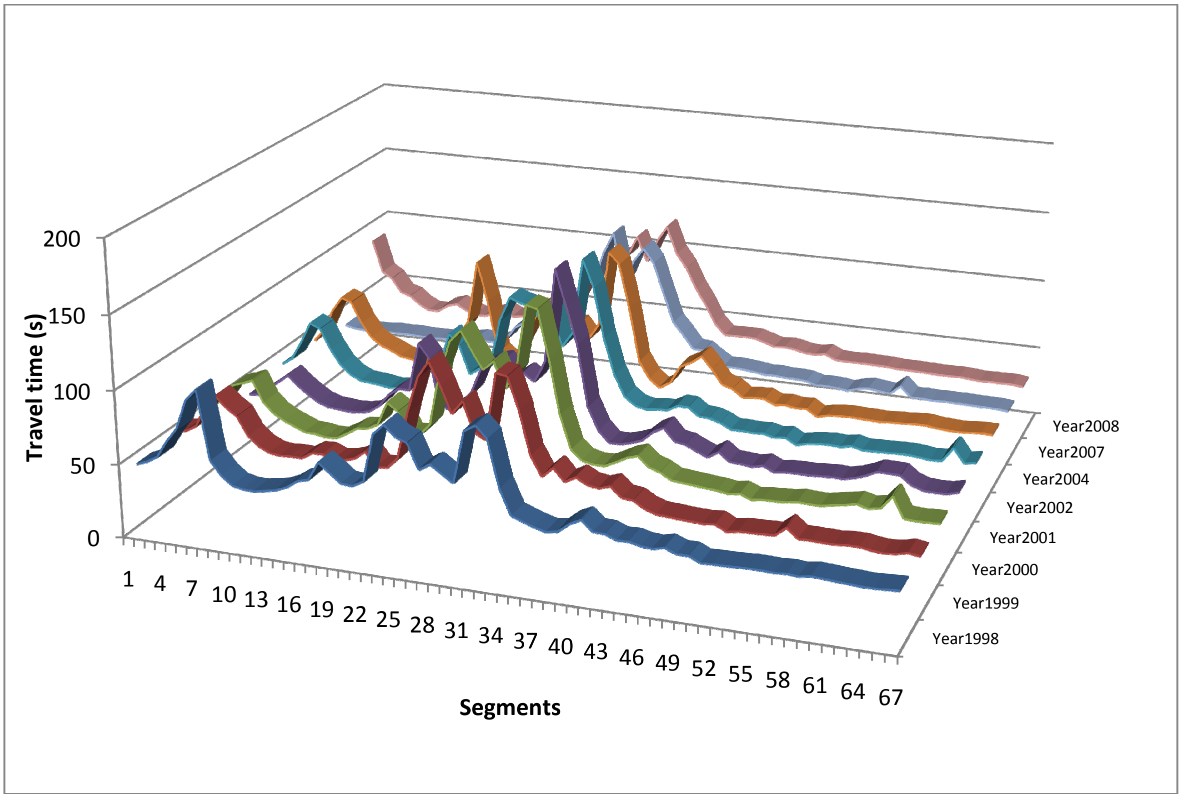

Figure 8. Means of travel times per year

One of the most interesting differences between data is for the section between segment 25 and 30. Actually, we have a clear confirmation from the Figure 9 of previously seen phenomenon: whereas for the past years, the traffic was more fluid, now, there seems to be a gradation of congestion between the two picks of congestion, an observation that could reveal saturation of the following segments and overflow to these segments.

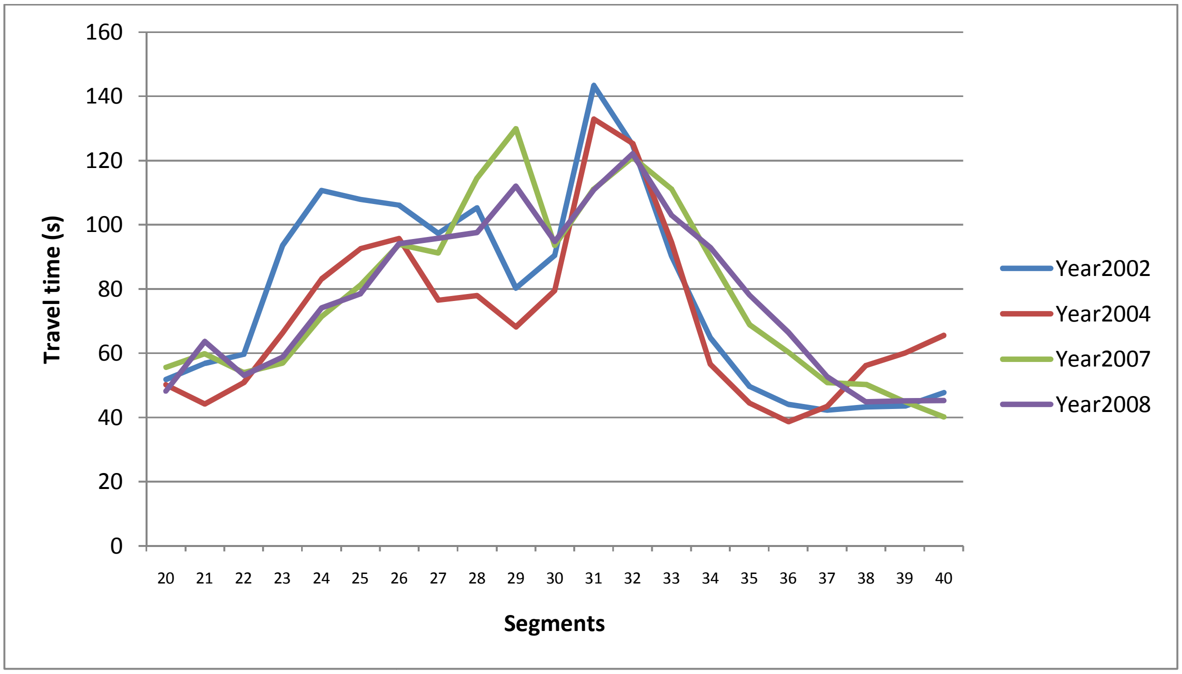

Figure 9. Means of travel time of segments 20 to 40

This observation is also confirmed by the two following figures presenting coefficients of variation of travel times, on each segment, for the eight years of observation. We can see that the traffic had become less reliable form year to year. Even if this evolution is not especially important in terms of number, the traffic can be unbearable for a daily user. It is quite clear for segments 20 to 35, as shown in Figure 11.

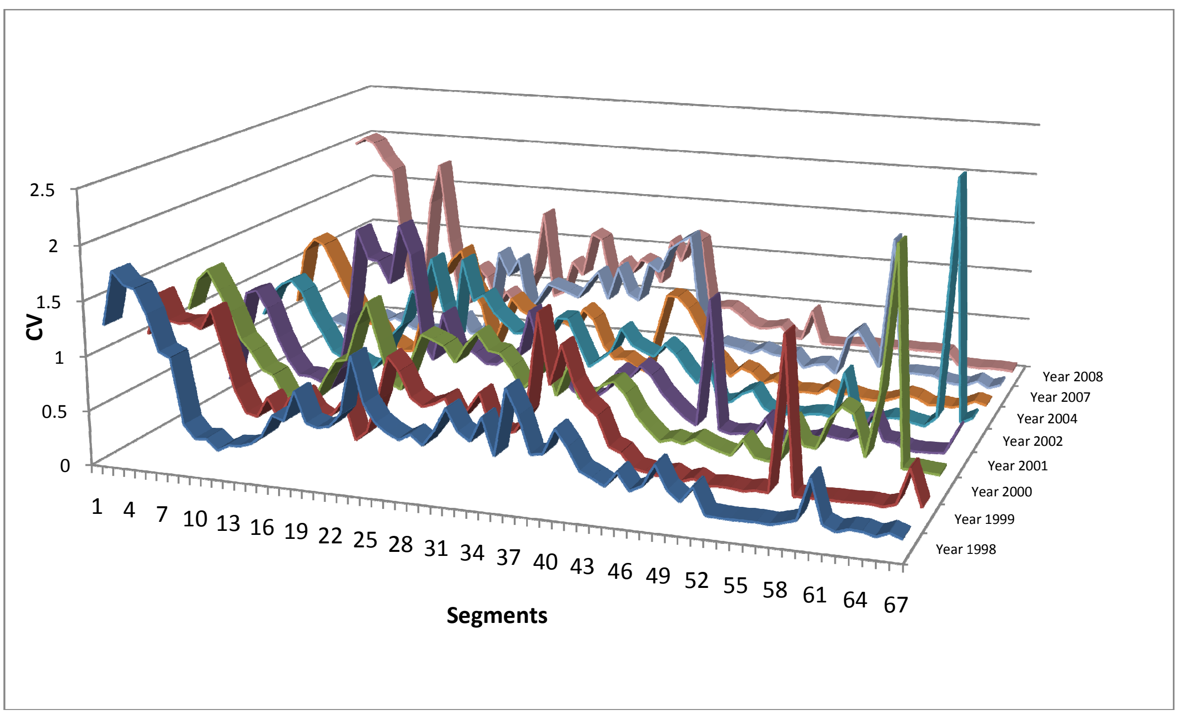

Figure 10. Coefficient of variation of travel times per year

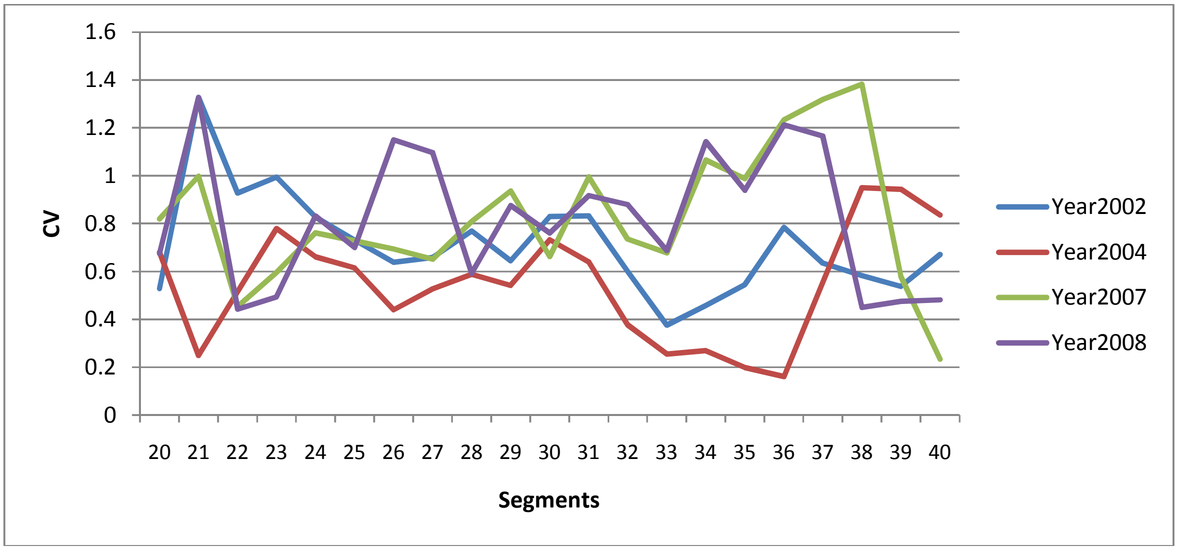

Figure 11. Coefficient of variation of segments 20 to 40

Expansion factor for means

In order to compare the years, it is possible to use an expansion factor which is defined in the next formula:

$$E_{year\ i-year\ i+1} = \frac{mean\ of\ travel\ times_{year\ i+1}}{mean\ of\ travel\ times_{year\ i}}$$

From this factor, there is a confirmation of a degradation of travel times for year 2007 and 2008. Obviously, the first segments of the circuit seem subjected to an enormous degradation. Several explanations are possible, considering the fact that they correspond to a bridge area:

- Problem due to one or both of the technologies (measurement errors)
- Highway maintenance
- Weather conditions

The second and third reasons, which need a confirmation from the Ministère des Transports du Québec, are clearly the most likely.

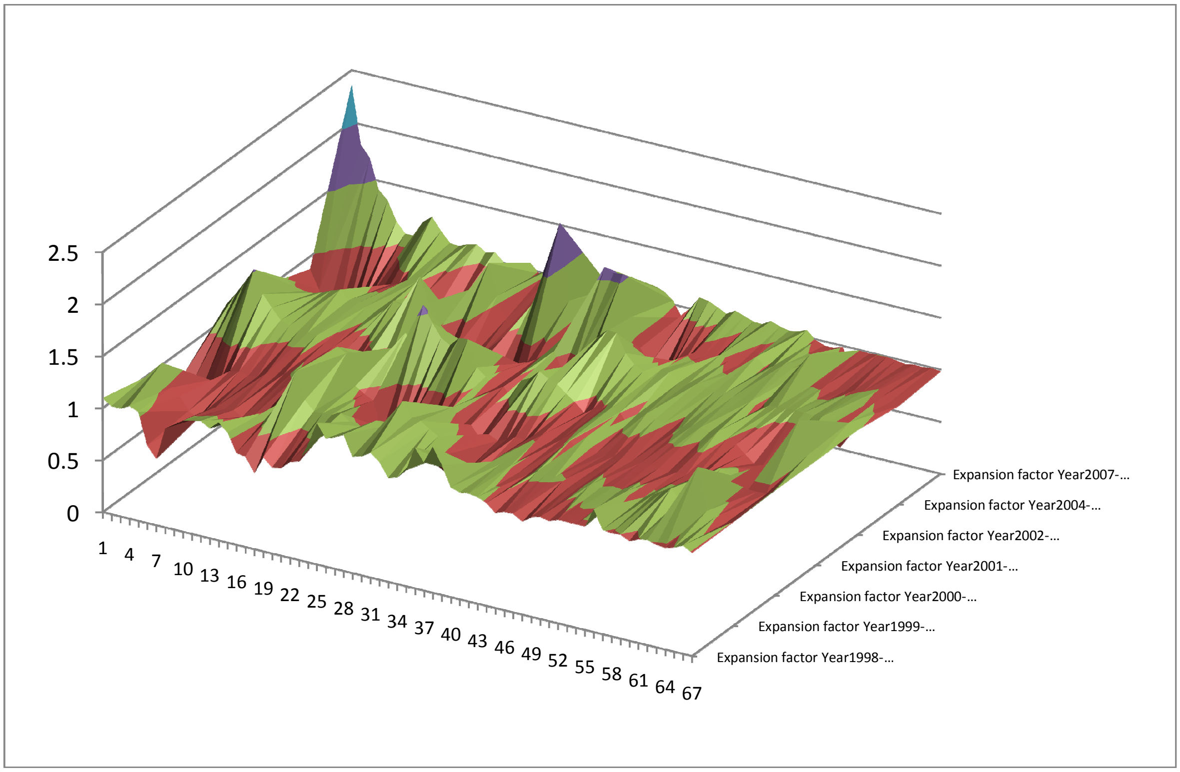

Figure 12. Expansion factors of travel times for years 1998 to 2008

CONCLUSION

This paper has presented a study aimed to the comparison of two different ways of measuring travel times on a freeway segment. In previous studies, floating car data, which are collected expressly for this task, has proven their benefit and their acuity on some routes. However, this exercise of data collection is too expensive to be viable on a continuous basis. Therefore, available GPS traces can be used to measure travel times over a network, if enough observations are available. In this study, we were able to estimate travel time on high volume segments by looking at specific GPS traces collected in carsharing vehicles. These results are promising and show that the measurement of travel time at high scale could be done with larger GPS datasets.

Of course, the analyses that were done here are partial and suffer from some limitations. Floating car data are limited to specific routes and cover an old period (1998-2004). Data quality is variable between the years, and not all periods are covered. GPS data are only available since 2006. In some cases location sampling is scarce and could suffer from interpolation. In addition, carsharing users are likely to avoid congestion and will not use the service as much on peak hours, limiting the number of observations. Hopefully, the increasing number of shared cars equipped with GPS traces will reduce this effect and increase the coverage of the area throughout the days and years. Still, other sources of GPS traces may have to be found to complete the dataset.

The study reveals that these two sources of data are complementary and can benefit one from each other. This brings new research perspectives for the measurement of travel time:

- Try to generalize the analysis to several routes to see if the differences between the datasets are systematic;
- Use more recent GPS traces, hoping to achieve better levels of resolution. Other sources of GPS data could also be envisioned.
- Try to add data from other sources of travel time collection technologies, like Bluetooth device matching, series of loop detectors, cell phone data, etc.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the collaborators from Communauto inc., the Company providing carsharing services in the Quebec Province, notably Benoît Robert and M. Marco Viviani. They also wish to acknowledge the Transportation system modelling service of the Quebec Ministry of Transportation for the continuous support on this topic and financing of the project. Pierre Tremblay and his team have contributed to this work. Authors also acknowledge the contributions of NSERC and FQRNT.

REFERENCES

Babin, A., Fournier, P., Gourvil, L. (2006). Base complète des relevés-tronçons des voitures flottantes du MTQ, Élaboration d'une base de données relative à la consommation de carburant et à l'émission de polluants à partir des informations des voitures flottantes, Ministère des transports du Québec (Service de la modélisation des systèmes de transport), Décembre 2006.

Babin, A., Gourvil, L. (2007). La base de données sur les relevés-tronçons des voitures flottantes du MTQ – Document technique, Ministère des transports du Québec (Service de la modélisation des systèmes de transport), Janvier 2007.

Babin, A., Gourvil, P., Fournier, P. (2007). Relevés des voitures flottantes du MTQ de 1998 à 2004. Exploitation par le Service de la modélisation des systèmes de transport de la Direction de la Planification du MTQ, Ministère des transports du Québec (Service de la modélisation des systèmes de transport), Octobre 2007.

FHWA (2004). Lessons Learned: Monitoring Highway Congestion and Reliability Using Archived Traffic Detector Data, Report n° FHWA-HOP-05-003, October 2004 http://www.ops.fhwa.dot.gov/publications/lessons_learned/index.htm.

Litman, Todd (2008). Well Measured. Developing Indicators for Comprehensive and Sustainable Transport Planning, Report, Victoria Transport Policy Institute, 47 pages (http://www.vtpi.org/wellmeas.pdf, page consultée en mai 2008).

Loustau, P., Morency, C., Trépanier, M., Gourvil, L. (2009). Portrait de la fiabilité des temps de parcours sur le réseau autoroutier montréalais, présenté au 44e congrès de l'association québécoise du transport et des routes, Montréal, Québec.

Loustau, P., Morency, C., Trépanier, M., Gourvil, L. (2010). Travel Time Reliability on a Highway Network: Some Estimations using Floating Cars Data, upcoming in Transportation Letters.

Loustau, P., Morency,C., Trépanier, C., Gourvil. L. (2010). Measuring, Describing and Modeling Travel Time Reliability, presented at the 89th Annual Meeting of the Transportation Research Board, Washington

Nesamania, K.S., Chub, L., McNallyc, M.G., Jayakrishnan, R. (2005). Estimation of vehicular emissions by capturing traffic variations, Center for Activity Systems Analysis. Paper UCI-ITS-AS-WP-05-2. http://repositories.cdlib.org/itsirvine/casa/UCI-ITS-AS-WP-05-2

Robitaille, M., Nguyen, T. (2003). Évaluation de la congestion « De la théorie à la pratique » : Réseau routier de l'agglomération de Montréal, communication présentée au Congrès 2003 de l'Association des Transports du Canada, St John's – Terre-Neuve, http://www.tac-atc.ca/English/pdf/conf2003/robita_f.pdf.

Schmöcker, J.-D., Fonzone, A, Quddus, M., Bell, M.G.H. (2006). Changes in the frequency of shopping trips in response to a congestion charge, Transport Policy, Volume 13, pp. 217–228.

Schrank, D., Lomax, T. (2003). The 2003 Urban Mobility Report, Texas Transportation Institute, http://mobility.tamu.edu, May 2003.

Schrank, D., Lomax, T. (2005). The 2005 Urban Mobility Report, Texas Transportation Institute, http://mobility.tamu.edu, May 2005.

Schrank, D., Lomax, T. (2007). The 2007 Urban Mobility Report, Texas Transportation Institute, http://tti.tamu.edu/documents/mobility_report_2007_wappx.pdf (page consultée en mai 2008), September 2007.

Szekeres, D., Heckman, M. (2005). Statewide Congestion Measurement Study – Final Report, Pennsylvania Department of Transportation, Contract N° 030419, Juillet 2005.

Tremblay, G. et M. Robitaille (2005). Analyse des temps de parcours sur le réseau routier de la grande région de Montréal. Rapport final du Groupe Conseil GENIVAR au ministère des Transports du Québec, Laval-Mille-Îles, région de Montréal, 124 p. + 2 annexes.