Case Studies on Transport Policy 4 (2016) 38–44

Contents lists available at ScienceDirect

Case Studies on Transport Policy

journal homepage: www.elsevier.com/locate/cstp

# Freight data collection using GPS and web-based surveys: Insights from US truck drivers' survey and perspectives for urban freight

Moshe E. Ben-Akiva[a], Tomer Toledo[b], Jorge Santos[a], Nathanael Cox[a], Fang Zhao[c], Yin Jin Lee[a], Vittorio Marzano[d,*]

[a] Intelligent Transportation Systems Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
[b] Technion, Israel Institute of Technology, Haifa, Israel
[c] Singapore-MIT Alliance for Research and Technology (SMART-MIT), Singapore
[d] University of Naples "Federico II", Department of Civil, Architectural and Environmental Engineering, Via Claudio 21, 80125 Napoli, Italy

ARTICLE INFO

*Article history:*
Received 9 January 2015
Received in revised form 21 October 2015
Accepted 10 November 2015
Available online 1 December 2015

*Keywords:*
GPS tracking
Intercity freight route choice
Urban freight modeling

ABSTRACT

This paper reports on tools, methods and experimental designs that have been developed to study the routing behavior and movement of trucks. The application of these capabilities is demonstrated with a case study on the route choices of North American intercity truck drivers', with a focus on the choice between tolled and free roads. An extension to the urban freight context, currently ongoing in Singapore, is briefly discussed, highlighting the challenges and main differences compared to the intercity case.


## 1. Introduction

Freight transport accounts for a considerable share of urban and intercity traffic, and the associated externalities. In the US, trucks carry the largest share of freight: in 2002, trucks moved 64% of freight by value, 58% by weight, and 32% by ton-miles (BTS, 2011). The movement of freight shipment tonnage is projected to increase by 65–70% by 2020 (FHWA, 2007); trucks are expected to haul 75% of the freight tonnage by 2020 (FHWA, 2005) and 68% of the value by 2040 (FHWA, 2011). The total annual highway miles driven by trucks increased by 109% between 1980 and 2008, a higher percentage increase than for other vehicle types. Similar trends have also been observed in the EU and other developed economies. The development of models and methods for planning and appraisal of freight transport systems is therefore a key priority. Current freight flow models are based on strong simplifying assumptions and weak behavioral foundations, which limit their explanatory power. A lack of data further limits their applicability. Thus, forecasts based on current models may be biased or imprecise.

There are three main dimensions of freight data collection: freight flows between the points of production and consumption (P–C); the logistics characteristics of shipments (e.g. shipment size, frequency of restocking, structure of the supply chain); and the transport characteristics of shipments (e.g. modes, routes). Freight flows are normally collected using costly and infrequent commodity flow surveys, providing a broad picture of national freight flows. These enable basic forecasting/planning, but do not include detailed information on the underlying logistics and transport choices. Thus, they do not support conversion of P–C flows into origin–destination (O–D) freight flows with the relevant characteristics of each leg of the logistics chain. The complete sequence of O–D flows corresponding to a P–C flows can only be traced by surveying producers and logistics operators, carriers and multimodal transport operators. This information provides useful insights into the supply chain structure (e.g. echelons and intermediate warehouses) and intermediate transport stops (e.g. transit points and intermodal terminals) for the various legs of the transport chain. Unfortunately, such data are not commonly available. When available, they are collected through traditional surveys, which have high costs and low response rates. Traditional surveys also tend to have non-representative samples. These stem from response biases resulting from respondents' short attention spans and limited ability to accurately recall information. In addition, traditional surveys fail in revealing the inter-relations and dependencies among the various entities involved in the freight industry that may influence their choices.

Improving freight data collection is needed to support the development of the next generation of freight transport models. Within the industry, there is already considerable penetration of

* Corresponding author.
*E-mail address:* vmarzano@unina.it (V. Marzano).

http://dx.doi.org/10.1016/j.cstp.2015.11.005

sensing devices, such as smartphones, GPS loggers and RFID tags. These provide an opportunity to leverage technology to unobtrusively collect a wealth of high-quality data, which could be complemented with information from other sources, such as shippers and carriers.

This paper presents an implementation of a next-generation freight data collection effort. It leverages GPS loggers, advanced sensing and communication technologies and machine learning architecture to deliver previously unobtainable data. These data reflect observed rather than stated information on the decisions of shippers and carriers.

The paper is structured as follows. Section 2 reviews the research on truck intercity route choices. Section 3 describes the data collection approach. Section 4 illustrates the results, showing the potential data quality and detail improvements that can be gained using this approach. Section 5 discusses the adaptation of a similar approach to collection of data on urban freight movements and then concludes.

## 2. Intercity truck route choices

Toll roads are an increasingly important part of the US road network, with 30–40% of new urban expressway mileage and about 150 new centerline miles expected per year (Perez and Lockwood 2006). Trucks make up a more significant percentage of toll road revenues than their traffic share suggests, because they typically pay higher tolls than cars. Standard and Poor's (2005) report that heavy trucks usually make up around 10% of traffic flow on toll roads, but 25% of the revenues. Accurate predictions of truck flows and the corresponding revenues are therefore crucial for toll road feasibility studies. Unfortunately, there is a record of significant biases and high variance in toll road forecasts (Bain 2009). One source of these errors is in truck drivers' route choice modelling, mainly resulting from the lack of relevant routing behavior observations.

Compared to passenger transportation, only limited work has been done on route choice in the trucking industry. Most truck route choice studies reported in the literature are based on stated preference (SP) data (e.g. Small et al., 1999; Kurri et al., 2000; Bolis and Maggi, 2001; Austroads, 2003; Hunt and Abraham, 2004; Danielis et al., 2005; Fowkes and Whiteing, 2006; Zhou et al., 2009; Wood, 2011; Toledo et al., 2013). SP studies present respondents with simplified hypothetical choice scenarios. The data collected may suffer from various biases and is generally considered less reliable compared to revealed preference (RP) data. Several studies (Jovicic, 1998; De Jong et al., 2004; Hess et al., 2014) use RP data in addition to SP data. These were collected using paper or computerized questionnaires in which respondents recorded their travel. These studies are limited in level of detail and accuracy in which the routes are reported, and in the number of observations that may be obtained from each respondent.

Large-scale data sets on truck route choice behavior have been obtained in varying ways. Hagino et al. (2010) used records of traffic permit applications submitted by drivers. Knorring et al. (2005) collected GPS traces using in-truck systems. However, both studies suffer from substantial limitations: they do not collect any information about the shipments or the drivers. Further, they only include a limited set of route attributes (e.g. travel times, distances and tolls and road characteristics, such as number of lanes) that may be directly derived from a map database. Knorring et al.'s study exemplifies the great potential of the use of GPS data, which is readily available in large quantities from in-truck navigation systems. However, it also shows the need to complement the location data with additional information related to the attributes of the trip and the constraints imposed by the shipment schedule and other factors.

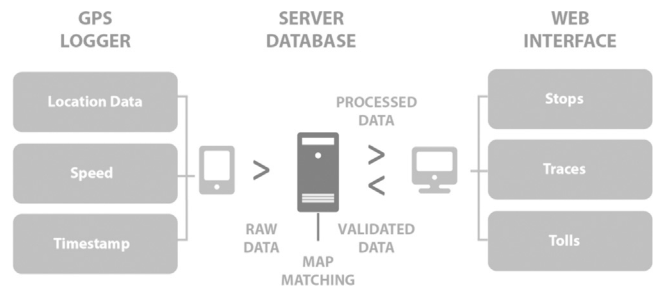


**Fig. 1.** Architecture for the truck' drivers route choice survey.

In the first phase of the current study, a traditional driver questionnaire with SP route choice questions was administered. The results of analysis of the collected data are reported in Sun et al. (2013) and Toledo et al. (2013). They show a wide variability in preferences towards toll roads and tolls, with route choices depending on multiple factors that include not only travel times and tolls, but also the probabilities and magnitudes of delays, toll bearing terms, driver compensation methods and shipment characteristics.

The next section describes the second phase of the study, which included a GPS-based RP survey using off-the-shelf GPS loggers to monitor all trips continuously and complemented by web-based prompted recall questionnaires.

## 3. Data collection system and methods

The architecture of the data collection system developed for this study is based on a combination of GPS loggers that were fitted in the participants' trucks with a web-based survey, as shown in Fig. 1. The location data collected by the GPS logger are transmitted in real-time to a backend server. The raw data is than processed to detect stops that the truck has made and to match the location observation to a GIS map database. The processed information is displayed to the participants in a web-interface. The participants are asked to validate the data presented to them and to respond to an additional prompted recall questionnaire.

The GPS loggers continuously collect data on the location and movement of the trucks and transmit this information through wireless networks to an application server. The GPS loggers did not require any professional installation and only need to be connected to the charger (cigarette lighter) in the truck cab. The logger used in the data collection described in Section 4 is a SANAV CT-24-D4F model with a backup battery (Fig. 2). The logger can collect location data, instantaneous speed and a timestamp. The reporting intervals can be set up to be either time intervals and/or based on minimum movement distance thresholds. At the end of the GPS data collection, participants were required to return the equipment, so that it could be re-used by subsequent participants.

At the backend server, algorithms are applied to match the observations to road segments on a GIS map database and to identify stops made by the drivers. On the matched route, tolling

**Fig. 2.** GPS logger used in the study.