Journal of Transport Geography 51 (2016) 119–129

Contents lists available at ScienceDirect

Journal of Transport Geography

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

# An analysis of drivers route choice behaviour using GPS data and optimal alternatives

Wilner Ciscal-Terry [a,b], Mauro Dell'Amico [a], Natalia Selini Hadjidimitriou [a,b,*], Manuel Iori [a]

[a] Department of Sciences and Methods for Engineering (DISMI), University of Modena and Reggio Emilia, Via Amendola 2, Pad. Morselli, 42122 Reggio Emilia, Italy
[b] Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Viale del Risorgimento, 2, 40126 Bologna, Italy

ARTICLE INFO

*Article history:*
Received 31 December 2014
Received in revised form 6 October 2015
Accepted 7 December 2015
Available online xxxx

*Keywords:*
Route choice
Shortest path
Habitual travels
Cluster analysis
GPS
Trajectories

ABSTRACT

This work aims to study drivers' route choices using a dataset of low frequency GPS coordinates to identify travels' trajectories. The sample consists of 89 drivers who performed 42 thousand paths in the province of Reggio Emilia, in Italy, during the seventeen considered months. Four attributes that may be important for the driver are identified and four optimal alternative paths are created based on the selected objectives to evaluate route choice behaviour. The comparison between the characteristics of the paths allows to conclude that drivers select routes that are overall longer than their optimal alternatives but that allow for higher speeds. Furthermore the statistical analysis of drivers' route choices in macroareas evidences that drivers have different behaviours depending on the geography of the territory. Specifically, there is higher heterogeneity of route choices in the plain areas compared to the mountains. In the second part of this work, clusters of repetitive travels are identified and a Geographical Route Directness Index is proposed to identify the areas of the province where the deviation from the shortest alternative path is higher. The analysis shows that, among groups of repetitive travels, the value of the index is higher along the ring road of the city of Reggio Emilia and there is a strong negative correlation between the frequency the driver selects the longer alternative that allow for higher speed, and the number of additional kilometres the same driver has to travel.



## 1. Introduction

An ambition of the local authorities is to be able to predict the drivers' choices so that appropriate measures for dealing with traffic management or infrastructure upgrade can be adopted. Route choice behaviour has been studied in the past from several points of views to derive information on the driver performances and on the elements that determine the drivers' selection process. Lin et al. (2012) and Maag et al. (2010) propose a route choice behaviour-cognitive model to describe the drivers' decision making process. Mahmassani et al. (1990) modelled the interaction of route choices with several variables and found that geography, network condition and information on the traffic conditions had a strong influence on route choices.

GPS (Global Positioning System) coordinates can help to obtain more accurate information on drivers' route choices as evidenced by Schönfelder and Axhausen (2003) who underline the potential of GPS data for travel behaviour research, especially when studying the variability over time. Furthermore the diffused use of navigation systems for private route guidance has increased the availability of positioning data for research. For instance, the Borlänge GPS dataset described in Axhausen et al. (2004) involved 260 private and commercial cars for almost two years.

The analysis of behavioural patterns of travels using GPS trips, is usually built on the Activity-Based Approach that recognizes that transport is a derived demand and that there is a strong relation between travel behaviour and human activities (see Jones et al., 1990). Basically these studies use explanatory models trying to describe the observed behaviours from a sample of driver's routes. Moreover, simulation approaches aim to create a set of plausible alternatives that are as similar as possible to the selection made by a set of drivers. Our work differentiates from these studies in the methodology and in the approach since we aim to compare the observed drivers' behaviour to a set of criteria (minimize distance, travel time, number of turns and number of traffic lights).

We use the GPS tracks recorded during a test of an European project, involving 89 drivers monitored for 17 months using in-vehicle dataloggers which produced 42.000 routes. Several researches from the literature compare observed routes to shortest time routes, but the speed used to compute the shortest path not necessarily fit with the estimated speed. Shortest time paths are usually computed using the maximum allowed velocities on the road link. Indeed, the first innovation we propose is to use the observed data, which span almost all the

* Corresponding author at: Department of Sciences and Methods for Engineering (DISMI), University of Modena and Reggio Emilia, Via Amendola 2, Pad. Morselli, 42122 Reggio Emilia, Italy. Tel.: +39 522 52 22 28; fax: +39 522 52 22 30.
*E-mail addresses:* wilner.ciscalterry@unimore.it (W. Ciscal-Terry), dellamico@unimore.it (M. Dell'Amico), selini@unimore.it (N.S. Hadjidimitriou), manuel.iori@unimore.it (M. Iori).

http://dx.doi.org/10.1016/j.jtrangeo.2015.12.003
0966-6923/

territory of the test and cover it 24 h a day, to compute a good and realistic approximation of the travel speed. The second innovation is to use empirical GPS data without taking into account the drivers characteristics, but grouping the individuals by similar missions that we identify as moving from a given restricted area to another restricted area. The aim is to identify if there is a common behaviour for several individuals moving between two given zones. We believe that this information has great value for local authorities, since it describes the way most of the drivers moves on the territory. The knowledge on how drivers behave between zones of the province, will allow to support local authorities' decisions concerning traffic management. New network configurations could be conceived based on actual drivers routes choice behaviour or new information could be included in the traffic models to have more realistic simulation results.

The remainder of the paper is organized as follows. In Sections 3 and 4 we describe the dataset used and the methodology developed in this study. Specifically, we propose an iterative method to compute average travel velocities from observed data and we describe how the four optimal alternatives are computed. Section 5 presents the results of the analysis derived from the comparison between the characteristics of the observed and optimal alternative paths. Furthermore route choice behaviour in geographical macro areas is analysed using shortest path optimization. Finally, a clustering procedure of paths' origins and destinations is implemented to identify groups of habitual paths and a new index is proposed. A further analysis of the road links with the higher deviation from the optimal alternatives provides important information on drivers' behaviour. In this work, route, path and travel are used as synonyms.

## 2. Literature review

The first theoretical studies of traffic flows, go back to Wardrop and Whitehead (1952) which propose mathematical methods based on shortest paths and on statistical analysis. Wolfe (1956) improved upon these methods and modelled the traffic equilibrium as a *Quadratic Assignment Problem*, one of the hardest problems in combinatorial optimization (see, e.g., Burkard et al., 2009). These models are based on the assumption that there is a sort of cooperation between drivers, which can resemble the modern *Swarm Intelligence* (Beni et al., 1993), that induces the drivers to choose routes giving a local minimum for the system, in which no individual can reduce its travelling time without increasing the travelling time of other individuals. These models, however, do not consider that the driver choices may be affected by factor as familiarity with a route, avoiding congestion, avoiding traffic lights, directness of the route, which are not related to a global objective function, but to personal preferences. Papinski and Scott (2011) studied some of the factors affecting route choice by considering a sample of 31 individuals who stated that their objective is to minimize their home-to-work travel time and the analysis proved that they selected direct routes and routes that allow to avoid congestion.

Information on the paths followed by a driver can be collected either using a questionnaire or by automatically collecting information using GPS data. A survey allows to gather information on the purpose of the travel, the motivation of specific choices, and other factors that do not have an immediate dependence from position and time. The collection of GPS trajectories offers the possibility to deploy more accurate and complete information on travel characteristics such as the spatial location of the travel, the travel time or the timestamp. There is a reduced cost and increased accuracy, compared to surveys, but no direct access to personal motivations. Some works, see, e.g. Wolf et al. (2001) and Tsui and Shalaby (2006) are focused on the development of methodologies trying to infer from GPS trajectories additional information as the purpose of the trip. However these works cannot explain all cases, as for instance, when the trip terminates in the middle of a road or in unidentifiable location. Moreover the sample used is generally small (9 individuals in Tsui and Shalaby (2006) and 24 from the Borlänge dataset). Our work, instead, uses the entire cleaned dataset and we perform aggregate statistical analysis to understand the driver's behaviour. Specific driver's choices are recorded and observed, but do not affect the final results when we can prove that the result is statistically significant.

To observe routes the researcher have mainly adopted shortest distance and shortest time paths for comparison (see, e.g., Prato and Bekhor, 2006; Jan et al., 2000; Schüssler and Axhausen, 2009; Sun et al., 2014). Shortest paths with turn-penalty and stop lights-penalty have been also used, for instance by Bekhor et al. (2006) who evaluated frequent routes using a sample of 188 respondents, corresponding to 91 origin–destination pairs. Papinski et al. (2009) observed 31 drivers and performed a survey combined with GPS data collection to detect preferred routes. They found that respondents preferred the routes that allow to minimize time, the number of traffic lights or stop signs, to avoid traffic congestion and to maximize the directness of the route. Papinski and Scott (2011) give a resume of the route choice attributes used in the literature and propose a GIS-based toolkit for route choice analysis which consider up to 40 variables. It is worth noting that the variables are obtained as attributes of the shortest distance and shortest time paths. Using a sample of 237 observed routes and a pair test-t, the authors compared observed routes to the optimal alternatives and found that drivers choose paths that are significantly longer if compared to the optimal paths in terms of distance and travel time. Zhu and Levinson (2012) demonstrate using GPS coordinates that drivers do not select shortest paths but select routes basing on travel time. To compute shortest time paths they estimate average travel time on the road network. New methodologies to estimate average link travel time using probe vehicles are proposed by Hellinga et al. (2008) and Jenelius and Koutsopoulos (2013). While other studies are focused on the analysis of correlation of arc travel speeds (see Bernard et al., 2006).

Trip variability has been studied by Huff and Hanson (1986) who used a contingency table to compare different attributes such as trip purposes, mode, trip distance and arrival time and Pas (1983) who implemented a cluster analysis to identify groups of similar daily travel activities patterns. Pas and Sundar (1995) found high day-to-day variability in trip frequency that was almost the same for both home-based and non home-based trips.

Temporal changes within individuals and the factors that influence change are the subject of longitudinal studies (see, e.g., Fitzmaurice et al., 2008). In transportation studies, longitudinal analysis could be performed using repeated measures of GPS data. Few studies focus on temporal variability of individuals' travel activities. For instance, Pas and Koppelman (1987) studied the day to day variability of trips during five days and found that interpersonal variability was higher for individuals who do not have to carry on daily activities. Hanson and Huff (1982) detected high degrees of repetition on daily travel activities and applied clustering techniques to classify homogeneous groups in terms of travel behaviour using multi-day travel data (see Hanson and Huff, 1986). Hanson and Huff (1988) studied the systematic components of day-to-day variability of travel behaviour and found that a seven-week observation period was not enough to capture systematic variability. Buliung et al. (2000) analysed spatial repetitive location choices during a week by type of activity and found differences in spatial variability and transport mode choice between week-day and week-end. Thus they conclude that policies aimed to reduce week-days private travels, would be not effective for the week-ends. To analyse drivers' path choice variability, Jan et al. (2000) used the GPS coordinates of 216 drivers who performed 3000 trips. They compared drivers taking identical trips to make considerations on the seasonality of the departure time and on the working and holiday periods to improve traffic assignment models. They conclude that trips made by the same driver were consistent over time and trips made by different drivers had some deviations.

Schönfelder and Axhausen (2003) used the Mobidrive survey (see Axhausen et al., 2002) conducted in 1999 in two German cities during a period of 6 weeks on 361 drivers. They found that the 70% of trips covered 2–4 locations and 90% of trips were made to the same 8 locations. They also detected that drivers choose between a small number of