Transportation Research Part C 19 (2011) 1243–1262

Contents lists available at ScienceDirect

Transportation Research Part C

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

On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data

Vincenzo Punzo [a,*], Maria Teresa Borzacchiello [b,1], Biagio Ciuffo [b,2]

[a] *Dipartimento di Ingegneria dei Trasporti, Università degli Studi di Napoli Federico II, Via Claudio, 21, 80125 Napoli, Italy*
[b] *Institute for the Environment and Sustainability, European Commission – Joint Research Centre, Via E. Fermi, 2749, 21027 Ispra (VA), Italy*

A R T I C L E I N F O

Article history:
Received 29 July 2009
Received in revised form 23 December 2010
Accepted 23 December 2010

Keywords:
Vehicle trajectory data
Data accuracy
Analysis of error
Delta method
Trajectory consistency
Spectral analysis
Traffic flow theory
NGSIM

A B S T R A C T

Trajectories drawn in a common reference system by all the vehicles on a road are the ultimate empirical data to investigate traffic dynamics. The vast amount of such data made freely available by the Next Generation SIMulation (NGSIM) program is therefore opening up new horizons in studying traffic flow theory. Yet the quality of trajectory data and its impact on the reliability of related studies was a vastly underestimated problem in the traffic literature even before the availability of NGSIM data. The absence of established methods to assess data accuracy and even of a common understanding of the problem makes it hard to speak of reproducibility of experiments and objective comparison of results, in particular in a research field where the complexity of human behaviour is an intrinsic challenge to the scientific method. Therefore this paper intends to design quantitative methods to inspect trajectory data. To this aim first the structure of the error on point measurements and its propagation on the space travelled are investigated. Analytical evidence of the bias propagated in the vehicle trajectory functions and a related consistency requirement are given. Literature on estimation/filtering techniques is then reviewed in light of this requirement and a number of error statistics suitable to inspect trajectory data are proposed. The designed methodology, involving jerk analysis, consistency analysis and spectral analysis, is then applied to the complete set of NGSIM databases.

1. Introduction

Trajectories drawn in a common reference system by all the vehicles on a road are the ultimate empirical data to investigate traffic dynamics. With such knowledge we can derive whatever information is sought about the physics of traffic phenomena on a road, including dynamics like car-following, lane-changing or gap-acceptance. This also explains the great interest that such data and related measurement techniques have attracted in the field of traffic flow theory (see e.g. Wu et al., 2003; Toledo et al., 2009).

Obtainable measurements of the trajectory of a vehicle consist of discrete observations of its position, equally spaced in time, that is the series of vehicle coordinates in the two- or three-dimensional space (coordinates refer to a specific point of the vehicle, e.g. the front bumper centre).

Remote-sensing and object tracking from video or photo cameras has been the main technique applied so far to gather such precious data. Cameras have been attached to aerial platforms as in Kometani and Sasaki (1959), Treiterer and Myers

* Corresponding author. Tel.: +39 081 7683948; fax: +39 081 7683946.
E-mail addresses: vinpunzo@unina.it (V. Punzo), maria-teresa.borzacchiello@jrc.ec.europa.eu (M.T. Borzacchiello), biagio.ciuffo@jrc.ec.europa.eu (B. Ciuffo).
[1] Tel.: +39 0332 785506; fax: +39 0332 786325.
[2] Tel.: +39 0332 786782; fax: +39 0332 785236.

0968-090X/$ - see front matter
doi:10.1016/j.trc.2010.12.007

(1974), Xing (1995), Hoogendoorn et al. (2004), or mounted at fixed and elevated locations, as in Ozaki (1993) and Coifman et al. (1998). The latter layout, in particular, allows vehicle trajectories to be collected only along a limited stretch of road. Generally, traffic flow is not disturbed nor influenced at all by the measurement operations following this approach.

Coordinates of a vehicle moving on a road can also be obtained by applying Global Positioning System (GPS) technology. Given the high expected accuracy in detecting positions (up to millimetres in the differential kinematic mode), the very precise timing of the system and the high sampling frequency of the measures, GPS technology is currently expected to provide the most accurate trajectory measurements. However, only traffic dynamics within a limited number of equipped vehicles can be observed, and not their interactions with the surrounding traffic – see e.g. the car-following experiments on vehicle platoons carried out on test track by Gurusinghe et al. (2002), and on real roads by Punzo and Simonelli (2005) – and only experimental, rather than empirical data, can be gathered, i.e. the driving behaviour of participants in the experiment is likely to be somewhat influenced by the experiment itself.

Other technologies allow detailed information to be gathered on vehicle interactions, even though they do not provide the trajectories of individual vehicles in a common reference space and generally apply to a single pair of vehicles. The very first experiments of such a data collection scheme were carried out using wire-linked vehicle pairs on test tracks and real roads in the pioneering works on car-following by the General Motors research team (see Chandler et al., 1958; Herman and Potts, 1959) where the spacing and relative speed between two vehicles and the follower's speed and acceleration were measured.

More recently, similar experiments were carried out by means of up-to-date technologies. For example, by equipping a vehicle with an optical speedometer and microwave radar providing relative spacing and speed, Wu et al. (2003), monitored the distance keeping behaviour between the test vehicle, driven by a test driver, and a preceding vehicle unaware of the experiment for long stretches of a route, in variable environments and traffic conditions.

All these experiments led to considerable improvements in the knowledge of traffic phenomena. Yet the limited amount of available data, their very different accuracy, their dishomogeneity as regards traffic conditions, road layouts and the extension of the observed time–space domain, as well as the very limited sharing of such data within the community of researchers, have not allowed a common understanding on many aspects of the traffic flow theory to be established. For instance, model calibrations and comparisons against different sets of measurements often led to contrasting results in car-following models, as pointed out in Brackstone and McDonald (1999).

Therefore, in 2002, the US Department of Transportation – Federal Highway Administration, started the Next Generation SIMulation (NGSIM) program, aimed at the theoretical development of new behavioural models and at collecting detailed trajectory datasets to estimate the model parameters and validate them (US Department of Transportation – FHWA, 2008a). Major innovations are the "open source" approach in developing the new behavioural models and free internet access to the vast amount of data gathered within the program. Data consist of vehicle trajectories, wide-area detector data and supporting data for researching driver behaviour. Vehicle trajectory data, in particular, are collected following the first approach mentioned: several synchronized video cameras, mounted on top of high buildings adjacent to the roadway, record vehicles passing through the study area (see e.g. the study area of the I-80 survey in Fig. 1). Post-processing of images finally gives vehicle positions on the road section every tenth of a second. In Table 1 the length of the road sections covered and the number of cameras per NGSIM survey site is reported.[3]

The free availability to the research community of detailed NGSIM data is opening up new horizons in the study of traffic flow theory and in its applications. Many studies have already benefited from using the trajectory data to perform experimental analyses and support theoretical findings by means of calibration and validation against such data. However, of more than 30 studies which used NGSIM data between 2007 and 2008 for different purposes, only four studies raised the issue of the quality of trajectory data (Duret et al., 2008; Hamdar and Mahmassani, 2008; Herrera and Bayen, 2008; Thiemann et al., 2008) and only two applied filtering methods in order to discard measurement errors (Hamdar and Mahmassani, 2008; Thiemann et al., 2008).

Yet depending on the aim of the study, the potential impact of measurement errors on the results can be significant. For example, Ossen and Hoogendhoorn (2008) verified the effects on car-following model calibration of adding, to vehicle displacements, errors from different distributions accounting for various combinations of random and systematic components. They basically pointed out that measurement errors yield a considerable bias in the calibration results.

In general, the dynamics under study could be partially hidden by the effect of measurement errors, and comparison of results of studies making use of different sets of data, with unknown accuracy, can be really hazardous (as well as uncertain their reliability).

Therefore the first necessary step to tackle such a problem is to define a methodology to inspect and quantify trajectory data accuracy. This is also needed to be able to develop more suitable trajectory estimation techniques. With this aim, in Section 2, the problem of estimating trajectory from discrete observations is analysed. First the structure of the error on point measurements and its propagation on the space travelled are investigated. In particular, under the specific assumption of independent errors on point measurements, evidence is given that the error propagated into the space travelled turns out to be a non-negative non-decreasing function of the number of observations. A consistency criterion for trajectories of following vehicles is therefore given. Literature on estimation/filtering techniques is thus reviewed in light of this consistency

[3] Note that the oldest surveys of the NGSIM program, namely "prototype" and "sample" datasets, are excluded from Table 1 as well as from the following analyses due to their lower accuracy.