Transportation Research Part B 80 (2015) 82–106

Contents lists available at ScienceDirect

Transportation Research Part B

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

Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns

Marcello Montanino, Vincenzo Punzo *

Department of Civil, Environmental and Architectural Engineering, University of Naples Federico II, Via Claudio, 21, 80125 Napoli, Italy

ARTICLE INFO

Article history:
Received 16 January 2015
Received in revised form 18 June 2015
Accepted 19 June 2015
Available online 14 July 2015

Keywords:
NGSIM
Car-following
Lane-changing
Calibration
Validation
Traffic oscillations

ABSTRACT

This paper shows that the behavior of driver models, either individually or entangled in stochastic traffic simulation, is affected by the accuracy of empirical vehicle trajectories. To this aim, a "traffic-informed" methodology is proposed to restore physical and platoon integrity of trajectories in a finite time–space domain, and it is applied to one NGSIM I80 dataset. However, as the actual trajectories are unknown, it is not possible to verify directly whether the reconstructed trajectories are really "nearer" to the actual unknowns than the original measurements. Therefore, a simulation-based validation framework is proposed, that is also able to verify indirectly the efficacy of the reconstruction methodology. The framework exploits the main feature of NGSIM-like data that is the concurrent view of individual driving behaviors and emerging macroscopic traffic patterns. It allows showing that, at the scale of individual models, the accuracy of trajectories affects the distribution and the correlation structure of lane-changing model parameters (i.e. drivers heterogeneity), while it has very little impact on car-following calibration. At the scale of traffic simulation, when models interact in trace-driven simulation of the I80 scenario (multi-lane heterogeneous traffic), their ability to reproduce the observed macroscopic congested patterns is sensibly higher when model parameters from reconstructed trajectories are applied. These results are mainly due to lane changing, and are also the sought indirect validation of the proposed data reconstruction methodology.

1. Introduction

Research in traffic flow theory is encumbered with the inadequacy of traffic data. The ones available – either conventional aggregate data or small samples of trajectories – have been hindering the possibility to corroborate and validate microscopic theories until recently, and have driven research mainly toward theoretical analyses of models. Unavailability of detailed comprehensive data sided with the need for analytical tractability of models, motivating the adoption of ideal settings and assumptions such as single-lane homogeneous flow in car-following. For instance, extensive studies focused on properties of car-following models able to explain/cause traffic instabilities such as oscillatory patterns or hysteresis (see the research stream focusing on model's string instability, for instance: Igarashi et al., 2001; Gasser et al., 2004; Orosz et al., 2005, 2009, 2010; Orosz and Stepan, 2006; Kesting and Treiber, 2008; Treiber and Kesting, 2011; Ward and Wilson, 2011; Wilson and Ward, 2011). Accordingly, lane changing was treated as an external perturbation possibly triggering instability (Wilson, 2008). When moving from theoretical settings to simulation environments handling models interaction, such

* Corresponding author. Tel.: +39 081 7683948; fax: +39 081 7683946.
E-mail addresses: marcello.montanino@unina.it (M. Montanino), vincenzo.punzo@unina.it (V. Punzo).

http://dx.doi.org/10.1016/j.trb.2015.06.010
0191-2615/

as entangled car-following and lane-changing simulation, properties of 'standalone' models may not hold. A challenge in researching traffic flow theory, therefore, is to enrich the analysis of traffic models with the study of their characteristics and properties in stochastic simulation environments.

To this aim, detailed and comprehensive traffic data are vital. In this view, the recent acquisition of microscopic traffic data throughout the Next Generation SIMulation program (NGSIM, 2005a) has set new scene in researching the field, despite their well-known drawbacks (see Section 2). Indeed, availability of vehicle trajectories of a complete 'population' of drivers i.e. of all the vehicles in a finite time–space domain, for the first time ever permits to validate models at both the microscopic and macroscopic levels.

From a microscopic viewpoint, NGSIM data enable validating simulation models against vehicle kinematic quantities, such as distributions of speed, acceleration and inter-vehicle spacing. This is essential in many application contexts where the fidelity of results is tightly related to the accuracy of simulated vehicle dynamics, such as accelerations in the assessment of vehicle emissions (e.g. Song et al., 2015; Vieira da Rocha et al., 2015), or the distributions of maximum speeds in safety studies (e.g. Cunto and Saccomanno, 2008).

From another perspective, as macroscopic traffic patterns can be obtained from NGSIM-like[1] trajectories, such data may allow establishing the overdue link between individual driving behaviors and traffic phenomena (e.g. Chiabaut et al., 2010). In fact, they also allow validating microscopic traffic flow theories at the macroscopic level by ascertaining if microscopic models, calibrated against (disaggregate) data, are able to reproduce the macroscopic traffic pattern emerging from the same data. Recently, such validation of microscopic theory received considerable attention (e.g. Chen et al., 2012, 2014; Laval et al., 2014), though limited to qualitative analyses. In the physics community, in particular, validation studies aimed at reproducing qualitatively prototypical traffic oscillation patterns (Helbing et al., 2009; Treiber and Kesting, 2011), the so-called "stylized facts" (Treiber et al., 2010). To the best of our knowledge, only one study attempted to match field data quantitatively (i.e. Treiber and Kesting, 2012) though under ideal assumptions i.e. single-lane homogeneous flow. If multi-lane heterogeneous flow with lane-changing has to be simulated, instead, the ensemble of models is no longer tractable analytically, and can be validated macroscopically only through simulation. In fact, simulated traffic dynamics are the result of the unpredictable (i.e. stochastic) interaction of individual driver models, and there is no closed-form expression that could map microscopic inputs, such as the distribution of individual driver model parameters, into the resulting macroscopic dynamics.

Notwithstanding the significance of NGSIM-like data for traffic flow theory, highlighted also in previous considerations, the scarce accuracy of collected NGSIM trajectories highlighted in the field literature poses a serious threat on their use.

At the microscopic level, the presence of non-negligible errors in the positional data corrupts vehicle dynamics and resulting distributions of kinematic quantities. The problem was first pointed out by Thiemann et al. (2008). Thus, Punzo et al. (2011) proposed a methodology to quantify the degree of accuracy/bias in vehicle trajectory data, whose application to NGSIM datasets confirmed massive measurement errors and allowed unveiling their nature. To this aim, in particular, a traffic-based criterion, the so called "platoon consistency" was proposed (it refers to the physical consistency of inter-vehicle spacing with car-following dynamics). At the macroscopic level, instead, whether or not errors in individual trajectories distort macroscopic dynamics is unknown.

Apart from the direct impact of measurement errors on observed traffic micro/macro dynamics, such errors have also an indirect impact on traffic simulation outputs, when these data are used to calibrate models. To the best of our knowledge, only Ossen and Hoogendoorn (2008) investigated this topic, though limited to car-following models only. The study unveiled a big impact of synthetic injected errors on calibration results. Conversely, the impact of measurement errors on calibration of lane-changing models has never been explored.

Even more, the impact of such errors on results of a whole traffic simulation, when different calibrated[2] driver models interact in the same simulation environment, is unknown.

Overall, the significance of data accuracy has been long overlooked in the field literature. Although measurement errors could be prefigured to affect model calibration and simulation, very few studies applied some kind of filtering to NGSIM data before their use. According to the results of a search made on the Scopus database (Elsevier, 2014), among the journal papers that made use of NGSIM trajectories since 2005 (i.e. 42 papers), only 6 papers out of 25 on car-following and 4 papers out of 17 on lane-changing, applied some kind of data filtering. Furthermore, proposed filtering methods are generally inappropriate to handle the nature of such measurement errors, as explained in Section 2.

To fill these gaps this paper provides two main contributions: a "traffic-informed" methodology to reconstruct microscopic traffic data (Section 2), and a general validation framework for traffic simulation (Section 3).

To overcome limitations of existing filtering procedures, the proposed methodology consists in a 'reconstruction' of the measured trajectory. It operates a substitution of extremely biased fragments of trajectory with synthetic ones that are consistent with vehicle kinematics and traffic dynamics. Such reconstruction procedure is herein applied to an entire NGSIM dataset with the aim of restoring the physical consistency of trajectories at both microscopic and macroscopic levels.

However, as the actual trajectories are unknown, it is not possible to verify directly whether the reconstructed trajectories, though physically consistent, are "nearer" to the actual unknowns than those unfiltered. Therefore, the only feasible approach to verify the efficacy of the reconstruction methodology is indirect: to calibrate the individual microscopic models

[1] Trajectories of all the vehicles in a finite time–space domain.

[2] Measurement errors affect simulation results indirectly, that is by affecting the calibration of individual driver models.