Blending of Floating Car Data and Point-Based Sensor Data to Deduce Operating Speeds under Different Traffic Flow Conditions

Giulia Del Serrone\(^1\)*, Giuseppe Cantisani\(^1\), Paolo Peluso\(^1\)

\(^1\)Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy

Abstract

Nowadays, smart mobility can rely on innovative tools for the knowledge of road system conditions, like operating speed data extracted from the so-called Floating Car Data (FCD). Probe vehicles in the traffic flow send to operation centres a large amount of travel information, collected through GPS detection systems, especially with regard to geolocation, date and time, direction and speed. As the sample deriving from these vehicles represents a tiny portion of the entire vehicular fleet, in this paper an analysis and a comparison with data obtained by point-based traffic sensors is proposed.

Therefore, the study analyses data collected by inductive loop detectors and microwave radar sensors, that provide information on the entire traffic flow in the time domain, in particular with the aim to identify free flow speed time bands. Afterwards, by means of the fusion between the results obtained from the data coming from these point-based control units and the ones coming from the probe vehicles, a comparison of the operating speeds in the two conditions of constrained and unconstrained traffic flow is performed.

Keywords: Floating car data; Point-based sensor data; Operating speeds; Traffic flow conditions.

1. Introduction

Monitoring, analysis, and management of a road network are complex activities, which a road operator must deal with, evaluating various factors related to the territorial context and the variety and complexity of operating conditions. The road system generally involves the use of a common infrastructure by different types of users (pedestrians, cars, buses, trucks, bicycles, etc.), who continuously interact with each other (Chang, Wang and Ioannou, 2007; Lewandowski et al., 2018). To control and manage the whole system, it is necessary to have reliable traffic data, both historical and in real time, from which the basic parameters for traffic analyses can be obtained: speed, density, and flow (Gitahi et al., 2020). In this paper, a methodology that elaborates and compares continuous operating speed profiles, in the two traffic flow conditions - constrained and unconstrained - is proposed, merging information from two different data sources: point-based control units and probe vehicles embedded in the traffic flow.

* Corresponding author: Giulia Del Serrone (giulia.delserrone@uniroma1.it)
Over the years, various traffic detection devices have been used in order to identify and describe the operational conditions of road mobility. Nowadays, the data sources are mainly of two types: static sensors and probe vehicles equipped with GPS sensors. Static devices are the most traditional traffic detection systems: among these manual counting, but also those that use more advanced technologies such as detection stations, automatic traffic counters, video cameras, radar and laser guns, and traffic sensors, such as microwave radar and acoustic sensors can be found (Ottesen and Krammes, 2000; Hashim, 2011; De Luca, Lamberti and Dell’Acqua, 2012; Dell’Acqua, 2012; Lobo, Rodrigues and Couto, 2013; Bassani et al., 2016; Cantisani, Serrone and Biagio, 2018; Cantisani, Del Serrone and Di Biagio, 2020). These devices are installed in specific sections of the road and sample the traffic flow that passes under them, but do not allow traffic monitoring along the entire development of the route.

In order to overcome this limit, innovative methods of collecting operating traffic data have been tested, in particular through the use of vehicles equipped with GPS devices capable of recording and sending information with a high sampling rate to a collection centre (Castro et al., 2008; Eboli et al., 2017; Ajmar et al., 2019; Astarita et al., 2020; J. Gitahi et al., 2020; Ma, Wei and Qian, 2021). This source provides high frequency and very closely spaced data useful to develop, in particular, actual operating speeds representations, through continuous profiles along the examined network (Talebpour and Mahmassani, 2016; Del Serrone, 2020).

Extensive collections of geo-referenced data from vehicles - known as Floating Car Data (FCD) if processed in real time, or Historical Car Data (HCD) if collected and analysed at different times - currently represent one of the most effective and low-cost traffic monitoring tools for the study and evaluation of road traffic conditions. In recent years, there is an increasing use of FCD / HCD thanks to the simplicity with which data relating to the position, speed, direction and time of acquisition are collected and sent anonymously to processing and analysis centres. Currently, vehicles equipped with geolocation and transmission systems have a penetration rate estimated around 2-5% of the entire traffic flow but, although this value is relatively low, several studies have shown the reliability and statistical representativeness of this sample, obtaining very positive results with determination coefficients $R^2$ higher than 0.9 (Zhao et al., 2009; Cantisani, Del Serrone and Peluso, 2022). On the other hand, the monitoring and analysis of the operational conditions of road traffic can be performed more completely and in greater detail by merging the information obtained from both survey methods: point-based sensors and probe vehicles (J. Gitahi et al., 2020; Croce et al., 2021). Consequently, the merge of the two data sources can allow to overcome some typical limitations of each methodology: the data obtained by the control units allow determining the volume and composition of traffic flows, information that is difficult to obtain directly from the FCD, while the latter allow to continuously follow some characteristic parameters of the vehicular motion.

The scientific literature also proposes several original approaches for the best and most extensive use of each data source. In particular, a study proposes innovative methods to estimate and reconstruct the traffic flow in transport networks: starting from travel time, estimated by the traffic conditions thanks to the crowd sourced data (FCD), the reconstruction of the traffic flows using the method of machine learning has been proposed (Li et al., 2021). Another study instead proposes algorithms of machine learning and deep learning to predict the flow of traffic, for example in a crossing, allowing adaptive control system, both with the remote control of the traffic lights and applying an
algorithm that regulates the timing according to the expected flow (Navarro-Espinoza et al., 2022). In addition, a traffic forecasting model has been proposed which can accurately predict the three key traffic variables: traffic density, flow and velocity; using FCD and a recurrent convolutional neural network that takes as input the three variables, estimated using FCD and data from induction loops (Mena-Oreja and Gozalvez, 2021). Finally, another study proposes to develop a multiperiod fundamental diagram for the traffic flow estimation from FCD considering the hysteresis phenomena, with the aim of improving the accuracy of flow estimation (Jiang et al., 2021).

Considering this background, the present study aims to develop a suitable method to identify the time slots (in the different days of the week) characterized by a condition of free-flow speed, by analysing the traffic data collected by inductive loop detectors and by microwave radar sensors. Subsequently, this information on the characteristic trend of traffic conditions will form the basis for the analysis of the HCD data sample, with the aim of filtering and selecting the speed data recorded along the same road infrastructure, dividing them into two subsets respectively referable to the two conditions of unconstrained and constrained flows.

2. Data and Methods

The proposed methodology has been applied along the Italian State road n.4, known as Via Salaria, in order to obtain the representation of the continuous profiles of the operating speeds (85% -ile of the speed distribution in the traffic flow examined). The operating speed $V_{85}$ is defined as the 85th percentile of the distribution of speeds selected by drivers in free-flow conditions on location of the road alignment. The estimation of $V_{85}$ on the geometric elements of the alignment make possible to associate every location of the alignment to a value of $V_{85}$ and to verify the design consistency (Pérez Zuriaga et al., 2010).

The free flow conditions (unconstrained) are characterized by desired speeds of road users, which are not bound by the presence of other vehicles or by traffic control and regulation devices, such as traffic lights, roundabouts or stop signs. The flow is instead constrained (by traffic) as the vehicle density increases, since the presence of other vehicles limits the manoeuvres of each user within his/her own traffic flow.

The constrained and unconstrained flows time bands determination (in the different days of the week) has been carried out through the analysis and processing of the data recorded by the control units placed along the examined road, that have been provided by the ANAS S.p.A. "Traffic Observatory". In particular, the survey stations considered in this study consist of microwave sensors and inductive loops, which record the information and send it to the “PANAMA” central Platform for Monitoring and Analysis (the acronym stands for: Piattaforma Anas per il Monitoraggio e l’Analisi), and are located as indicated in the following list (see also Figure 1):

- Station 64 - Road: S.S. 4, kilometre: 35.451, Municipality: Passo Corese, Province: Rieti, Region: Lazio;
- Station 66 - Road: S.S. 4, kilometre: 86.027, Municipality: Cittaducale, Province: Rieti, Region: Lazio;
- Station 67 - Road: S.S. 4, kilometre: 96.528, Municipality: Borgo Velino, Province: Rieti, Region: Lazio;
- Station 68 - Road: S.S. 4, kilometre: 133.400, Municipality: Amatrice, Province: Rieti, Region: Lazio;
The traffic surveys operated by the control units provide the following information:

- date and time of the beginning and end of the registration (aggregated every 5 minutes);
- lane in which each vehicle travels;
- travel direction of the vehicles -ascending (A) if directed according to the increasing kilometers, descending (D) otherwise-;
- number of samples recorded in 5 minutes;
- average vehicular speed;
- speed standard deviation;
- vehicle type;
- average length of vehicles;
- average distance between vehicles.

The data processing phase involved the implementation of calculation codes, to separate the records in the two travel directions and then filter and exclude data relating to overtaking situations. For each direction, the fundamental flow diagrams in the flow - average speed plane (F, S), shown in Figure 2, were represented, and the trends of vehicular flows over time (t, F), shown in Figure 3, were studied. The identification of the time bands constrained, and unconstrained flows has been performed distinguishing between the weekdays (i.e., from Monday to Friday) and the weekend (i.e., Saturday and Sunday).
The study of the fundamental diagram then continued by identifying the envelope of the stable outflow branch, by means of a cubic function whose inflection point was subsequently determined. Known the coordinates \((F^*, S^*)\) of this inflection point, the vehicular density in correspondence with it ("critical density") was defined by the relation \(d^* = \frac{F^*}{S^*}\), and was assumed as a threshold that separates the unconstrained from the constrained flow conditions. As shown in Figure 4, the points to the left of the ray connecting the origin of the axes \(O\) \((0,0)\) and the point \(d^*\) \((F^*, S^*)\) were considered representative of unconstrained conditions - points in orange - while those to the right of the ray - points in blue - have been associated with constrained conditions.

As shown in Figures 2 and 3, the representation with different colors of the recordings below and above the limit threshold identified, made it possible to also identify the time bands in which the unconstrained/constrained flow conditions occur. Analyzing in more detail the number of data recorded over time, it was noted that for all the control units located on the Via Salaria the amount of flows was very different from month to month: for this reason, the data were divided into three temporal subgroups with similar flows (January-February, October-November, and September-December) and were studied separately; nevertheless, despite the disparity in the flow rates, the characteristics time slots were the same for all months. From the study, it is clear that only the control units installed near the city of Rome (n. 1046 and n. 64) register high flow values and, therefore, only for them different time bands of constrained and unconstrained flows have been identified – see Table 1:

Table 1: Time bands of constrained traffic flow conditions on the SS4 in the proximity of the city of Rome.

<table>
<thead>
<tr>
<th>Control Unit</th>
<th>Weekdays</th>
<th>Weekend</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dir. AB</td>
<td>dir. BA</td>
</tr>
<tr>
<td>1046</td>
<td>14:00-20:30</td>
<td>6:00-9:00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>-</td>
<td>6:30-8:30</td>
</tr>
</tbody>
</table>

The sample of GPS data refers to the period between August 2016 and February 2017 and to a quadrilateral including the entire Lazio region, part of the Marche, Abruzzo and Umbria regions. The information deriving from the HCD briefly concerns the following Vehicle identification, Satellite service identifier, Vehicle type, Device identification,
Vehicle ignition date and time, Date and time of emission of the signal, Longitude (in WGS 84 coordinates), Latitude (in WGS 84 coordinates), Signal quality, Speed, Direction and Odometer.

The proposed methodology considers that, after some preliminary treatments, the information obtained from the analysis of the control units must be extended to the sample of HCD collected near each point-based sensor, i.e. one kilometre before and one kilometre after the control unit. At the beginning, the geometric reconstruction of the existing road was carried out starting from the georeferenced vertices of the road graph, following an original methodology developed by the authors (Cantisani and Del Serrone, 2021). Subsequently, by querying the database in the MySQL Workbench software (Oracle Corporation, 2016), only the data located in an area extending 15 meters to the right and left of the reconstructed road axis were extrapolated from the initial HCD sample. Finally, a map matching procedure was developed to associate the data relating to the monitored vehicles to the road layout (Quddus et al., 2007; Xi et al., 2007). The speed values were then plotted as a function of the distances and, in conclusion, an attempt was made to create a continuous reordered speed profile through the use of a cubic spline smoothing (Del Serrone, 2020), that performs a least squares fitting trying to minimize the error obtained in terms of distances between the data and the spline itself (Dyer and Dyer, 2001; Cantisani et al., 2004).

The reconstruction of the continuous speed profiles in the two conditions of constrained and unconstrained flows, and their overlapping representation on the same graph, made it possible to observe and compare the different trends of the operating conditions of traffic along the infrastructure.

3. Results

The developed methodology highlights the importance of being able to merge the information extracted from the different traffic monitoring systems and the usefulness of each dataset in relation to various assessments and analyses to be performed.

In this study, the extraction and comparison of the continuous speed profiles from the HCD, in the two conditions of constrained and unconstrained flow (red and green lines in the figures, respectively), has been made possible as a result of the processing of data collected by the control units. In fact, the time slots corresponding to these different traffic conditions, both for the case of weekdays and weekends, have been identified by examining the fundamental diagram and determining the critical density value d*. This information was then used to examine the HCD sample, analysing the GPS data emitted in an interval of 2 km across the control unit, as it was assumed that the traffic condition detected by the point-based sensors can be extrapolated within a limited distance from the examined point. Figure 4 shows the HCD, as a dense point cloud, detected around the 1046 control unit on weekdays. In detail, it can be seen how the HCD corresponding to unconstrained flow conditions (Figure 4.a) have a distribution that is more concentrated in the 60-80 km/h range. Instead, in the case of constrained flows (Figure 4.b) the point cloud is characterized by a greater dispersion of the data and by a greater point density in the lower area of the diagram (speed values less than 60 km/h). Figure 5 shows the same outcome regarding the composition and distribution of the point cloud, on weekends. In this case, however, a less dense point cloud is noted, since the HCD emitted during the weekends is less than in the case of weekdays.

In order to characterize and describe the trend of operating speeds with a continuous function, it was decided to use the smoothing cubic spline in order to interpolate the 85th
%-ile values of the vehicle speeds recorded (Transportation Officials., 2011; Tottadi and Mehar, 2022) and obtained by collecting the HCD. Furthermore, through an original methodology and the implementation of a specific calculation code (Cantisani and Del Serrone, 2021), the trend of the theoretical speeds was also obtained as a function of the curvilinear abscissa, in analogy to the theoretical model for the design speed of the new roads in Italy (Ministero delle Infrastrutture e dei Trasporti, 2001). Through the overlapping of the operating speed profile with the theoretical speed diagram it can be observed how, in both flow conditions (constrained and unconstrained), the trend of operating speeds always deviates from that of the theoretical model. In Figures 4.a e 5.a, referred to unconstrained flows, the non-coincidence between operating and design speeds is caused not only by traffic, but also by other factors, such as speed limits, driver behaviours, meteorological conditions, particular characteristics of the infrastructure. In Figures 4.b and 5.b, however, an even greater deviation between the two profiles can be observed, due to the amplification effect produced by the constrained flow conditions.

Figure 4: Operating speeds constrained and unconstrained – Control unit 1046 dir. AB SS4 – Weekdays

Figure 5: Operating speeds constrained and unconstrained – Control unit 1046 dir. AB SS4 – Weekend

Figure 6 compares the two operating speed profiles, superimposed on the curvature diagram. It can be observed that the unconstrained flow condition is always higher than the constrained one, with an almost constant offset approximately equal to 8-11 km/h. The trend of the two speed profiles, however, follows the same pattern, demonstrating how the operating conditions of vehicular flows are influenced not only by traffic, but
also by other aspects, such as the geometric characteristics of the road. In particular, it can be noted that in the section between the distances 6+250 and 6+750 km, characterized by a succession of tight curves, the operating speeds actually tend to reduce at the most critical points and then increase again.

Figure 6: Operating speeds for constrained and unconstrained – Control unit 1046 dir. AB SS4 –

4. Conclusions

The proposed methodology allows to highlight the excellent results that can be obtained from the combined use and merging of data from different sources able to survey and monitor road traffic. In fact, both information obtained from control units and probe vehicles have been used. The first one has the great advantage of detecting the entire traffic flow that crosses a certain point of the road, where the monitoring station is located. The last one, instead, allows following the motion of the vehicles both in the temporal and in the spatial domain, albeit referred to a relatively limited (but representative) observed sample.

The large amount of data collected by the control units has proven to be fundamental in identifying the time slots characterized by constrained and unconstrained traffic flows. This information was then used to analyse the HCD sample, to obtain the continuous profile of operating speeds, along a road section, in the different traffic conditions. The unconstrained flow condition has shown an almost constant offset, approximately equal to 8-11 km/h, higher than the constrained one, both following the same trend. The operating speed trends can finally be compared with predictive speed models, as well as used for other observations and evaluations regarding the congruence between the actual operating conditions and the technical characteristics of the infrastructure.

In conclusion, the proposed methodology can be applied for numerous analyses aimed at studying the actual evolution of the vehicular traffic, in real conditions and under various scenarios. Indeed, it combines the statistical observation of traffic flow conditions (allowed by the point-based control units) with the innovative use of a sample of data (the FCD / HCD) more limited in quantitative terms, but more extensive and complete in a spatial and temporal domain. Overall, from the merge of multiple traffic data sources it is, therefore, possible to obtain a large amount of information regarding the operating conditions of vehicular mobility, traffic safety, monitoring and analysis of road infrastructure and networks.
References


Gitahi, J et al. (2020) ‘Multi-sensor traffic data fusion for congestion detection and tracking’, International Archives of the Photogrammetry, Remote Sensing and Spatial


