Motorway speed pattern identification from floating vehicle data for freight applications

A. Pascale [a], F. Deflorio [b], M. Nicoli [a,*], B. Dalla Chiara [b], M. Pedroli [c]

[a] *Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Milano, Italy*
[b] *Politecnico di Torino, Dept. DIATI, Engineering – Transport Systems, Torino, Italy*
[c] *W.A.Y. srl, Torino, Italy*

1. Introduction

Monitoring technologies based on probe vehicles have been emerging, in recent years, both as self-working solutions and in cooperation with infrastructure-based systems (Yoon et al., 2007; Kerner et al., 2005; De Fabritiis et al., 2008; Wei et al., 2007). Floating car data (FCD) systems are based on a set of probe vehicles equipped with satellite positioning and wireless connectivity – such as GSM/GPRS, UMTS/HSPA and LTE – to periodically send position-speed data to a central unit at a control room. Data are aggregated and processed by the central unit to draw the traffic information needed for the specific application. Applications range from traffic monitoring and forecasting, travel time estimation, construction of historical database to enable identification of anomalies or incident detection, fleet management, and dynamic navigation based on real-time traffic conditions (Yang, 2005; Sethi et al., 1995; Treiber et al., 2010).

---

* Corresponding author.
[1] More details can be found on the official website of the European Commission for Mobility and Transport, http://ec.europa.eu/transport/road_safety/specialist/knowledge/esave/esafety_measures_known_safety_effects/black_boxes_in_vehicle_data_recorders_en.htm, and of the European Parliament, http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+PV+20140225+ITEM-013+DOC+XML+V0//EN. For eCall see Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010, on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport, 6.8.2010, and Commission Delegated Regulation (EU) No 305/2013, of 26 November 2012, supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the harmonised provision for an interoperable EU-wide eCall.

For many services in transport systems, reliable speed data are useful to estimate or predict the travel time along a route, which is a sequence of road segments where traffic conditions usually change over time. In fleet management operations, for example, useful information for both operators and end users is the travel time for on-time pickup and delivery operations, if the freight distribution service is managed to comply with time constraints. A reliable baseline for speed data along the road network should be exploited to meet the time requirements. Furthermore, any event that may cause a delay in traffic is important to be detected in order to give update information on the speed observed along the route. In this context, the computation of typical speed patterns characterizing the traffic behavior over the various road sections and the detection of anomalies causing relevant delays are useful tools for the fleet management.

During the last years, the interest in FCD has been sensibly growing thanks to the diffusion of Global Positioning System (GPS) navigators – enhanced in Europe with the European Geostationary Navigation Overlay Service (EGNOS) - and location-enabled smartphones, as well as to the impressive surge of location-based services. An increasing number of dedicated companies/agencies is now dealing with data collection for fleet management, accident data recording, and vehicle insurance. The growing number of monitored vehicles, together with the extended connectivity provided by new communication systems (connecting vehicles to the central unit and also vehicles to vehicles) are making FCD a leading, sometimes even consolidated, solution for traffic monitoring.

One of the main problems, however, is the reliability of data collection which may be limited by the local penetration rate. In probe systems, sampling of traffic parameters is non-uniform and also time-varying due to the probe mobility. Resolution depends on a combination of factors including number of probes, traffic demand patterns, traffic conditions and road features (Kwon et al., 2007; Vandenberghe et al., 2012; Herrera et al., 2010; Rahmani et al., 2013; Fangfang et al., 2013). In such a com-plex scenario, data processing represents the key engine for integrating the sparsely sampled data into accurate and reliable traffic information, overcoming as much as possible the limits due to low penetration rate and/or non-uniform sampling.

In this paper we focus on processing of floating truck data (FTD) collected by a fleet of heavy-duty vehicles over the Italian A4 motorway, connecting Turin to Venice. We propose a procedure for the computation of mean speed patterns characterizing the typical traffic conditions along the road, to be used for construction of historical database and for detection of anomalous traffic events. GPS position-time measurements provided *locally* and *instantaneously* by single vehicles are processed to estimate the *mean* speed that results from aggregated vehicles in each road segment and each timeslot. Different methods based on the study of traffic daily patterns have been proposed in the literature (Rakha et al., 1995; Wild, 1997; Chrobok et al., 2004; Chung, 2003; Kerper et al., 2011) using data coming from either probe vehicles or loops. The analysis considered in this paper, based on probe vehicles, is particularly challenging as the available dataset relies mainly on truck measurements, it has moderate dimensions (3 months) and low penetration rate (estimated around 0.25% in capacity condition), resulting in a fragmented observation of the velocity field. To overcome this limit, we propose to estimate the typical velocity profile in each road segment using a clustering procedure that exploits data from probes in different segments, relying on the fact that speed profiles usually share common features over large sections of the road. A preliminary analysis is carried out to select reliable vehicle data, match the data samples to the road segments and compute the mean speed. A clustering procedure based on the Ward's method (Weijermsrs et al., 2005) is then applied for aggregating the road segments into homogeneous classes of speed trends and compute the typical speed profile associated to each class of segments. We also propose to exploit these typical profiles as reference for detection of anomalous traffic conditions and unusual events causing relevant delays.

Our main aim is to ascertain if typical speed patterns can be identified, even in scenarios with a fragmented observation of the velocity field due to low penetration rate, relying on the fact that speed profiles often share common features over several road segments. This basic idea allows the computation of a reliable set of speed profiles using a conventional clustering approach as the Ward's method. Other methods could be tested as well, but the evaluation of their performance is out of the scope of this paper, as the focus here is on the overall procedure not on the specific clustering component.

Even though the analysis is based on truck measurements and thus restricted to working days where heavy traffic is allowed (week-ends or holidays are excluded), the proposed method can be easily applied to extended datasets including also other vehicle-type data, if available from other sources. The analysis is validated using information broadcasted by the national provider of road traffic information, named CCISS (*Centro di Coordinamento Informazioni sulla Sicurezza Stradale*), by collecting all the anomalous events registered by CCISS over the considered motorway. The comparison to CCISS data shows that the anomalies recognized using the reference profiles correspond to real congestion events, confirming the reli-ability of the proposed approach.

2. Data analysis: method and procedures

The analysis herein considered is based on FTD provided by a system designed and managed by the Italian company W.A.Y. (Torino, Italy). Data are collected by an operation center which has got one of the most extended databases in Italy in this field; it receives signals from various fleets equipped with on-board devices, for a total number of more than 13,000 probe vehicles. Each road is divided in segments and data from any vehicle are mapped to a road segment only if the vehicle is localized within the Italian motorway network or an important highway, such as a ring road. The scenario considered in this paper refers to the motorways of north Italy as shown in Fig. 1. A preliminary analysis of these data has been presented in Pascale et al. (2013).

Fig. 1. Map of north Italy motorways and look on A4 (Turin–Venice) highway, section Milan–Brescia. Circles denote the vehicle velocity samples matched to highway segments.

Objective of this work is to derive from vehicle-based information the space-based information concerning the historical speed trends along defined roads on the map. As shown in Fig. 2, the process starts from speed samples collected at given time instants and locations by the on-board devices and forwarded to the central unit. Speed samples are used together with map information to compute daily velocity trends over the road network.

For a complete characterization of the speed pattern, a daily speed profile has to be computed for each road segment. To this aim, taking into account the limited amount of measurements available per each segment and timeslot, we propose an aggregation and filtering process based on clustering of speed data: segments are clustered in few classes with similar speed behavior and the speed profile of each class is computed by averaging the data collected in all segments of the class. An example of velocity map obtained by this process for the road section in Fig. 1 is shown on the bottom of Fig. 2. Once the typical speed profiles have been computed, these profiles are used as a reference for identification of anomalous events.

The main steps of the proposed processing methodology are summarized in Fig. 3. The procedure starts with the computation of a preliminary speed profile for each road segment, referred to as the *segment speed profile* (steps A,B,C,D). A classification procedure (step E) is then conducted to recognize the typical daily trends (referred to as the *typical speed profiles*) and to identify anomalous traffic behaviors (step F). Details of the processing steps are given in the following subsection.

2.1. Data provided by the FCD system

Each motorway is divided into a set of segments, with length $100 \text{ m} \leqslant l_S \leqslant 1300 \text{ m}$ (on average 500 m), and indexed in the direction of traffic as $s = 1,2,\ldots,N_S$. Temporal sampling ranges from 20 s to 3 min. Probe vehicles send to the central unit an array of data composed by: time stamp t_m [data, hh:mm:ss], vehicle ID [#], GPS position [lat, long], instantaneous velocity [km/h], incremental distance covered by the vehicle [km]. At the central unit, raw vehicle data are associated to the closest segment on the geographical map based on the computation of the distances between the GPS vehicle position and the segments. Velocities after map matching, v [km/h], are shown in Fig. 1 for the motorways of north Italy. For the subsequent analysis we focus on the section Brescia → Milan (BS → MI) of the motorway A4, as highlighted in the box, covering a total number of $N_S = 166$ segments. This road section has a length of more than 90 km and provides a challenging testing scenario thanks to the variegated and highly time-varying traffic behaviors that are observed along the way, ranging from suburban areas to congested urban sections in the area of Milan.

For the analysis we consider two different datasets collected over the motorway A4 BS → MI: the first obtained by a set of 5327 truck paths during the months of February, March and May 2011; the second one collected by a set of 3882 truck paths in the month of June 2011. The former set is used in this section to develop the processing method and in Section 3 to compute the typical speed profiles; the latter is used in Section 4 for validation on specific applications.

2.2. Processing steps for computation of speed profiles

We focus our attention on the first dataset. The dataset was obtained starting from data of vehicle positions and velocities after the map-matching procedure. From the 3-months data we extracted a subset by selecting 63 working days and performing a consistency check. To give an idea of sample size, Fig. 4 shows the number of monitored vehicles for all segments during Mondays; data are aggregated over 1 h. The overall penetration rate has been estimated to be approximately 0.25% in capacity condition.
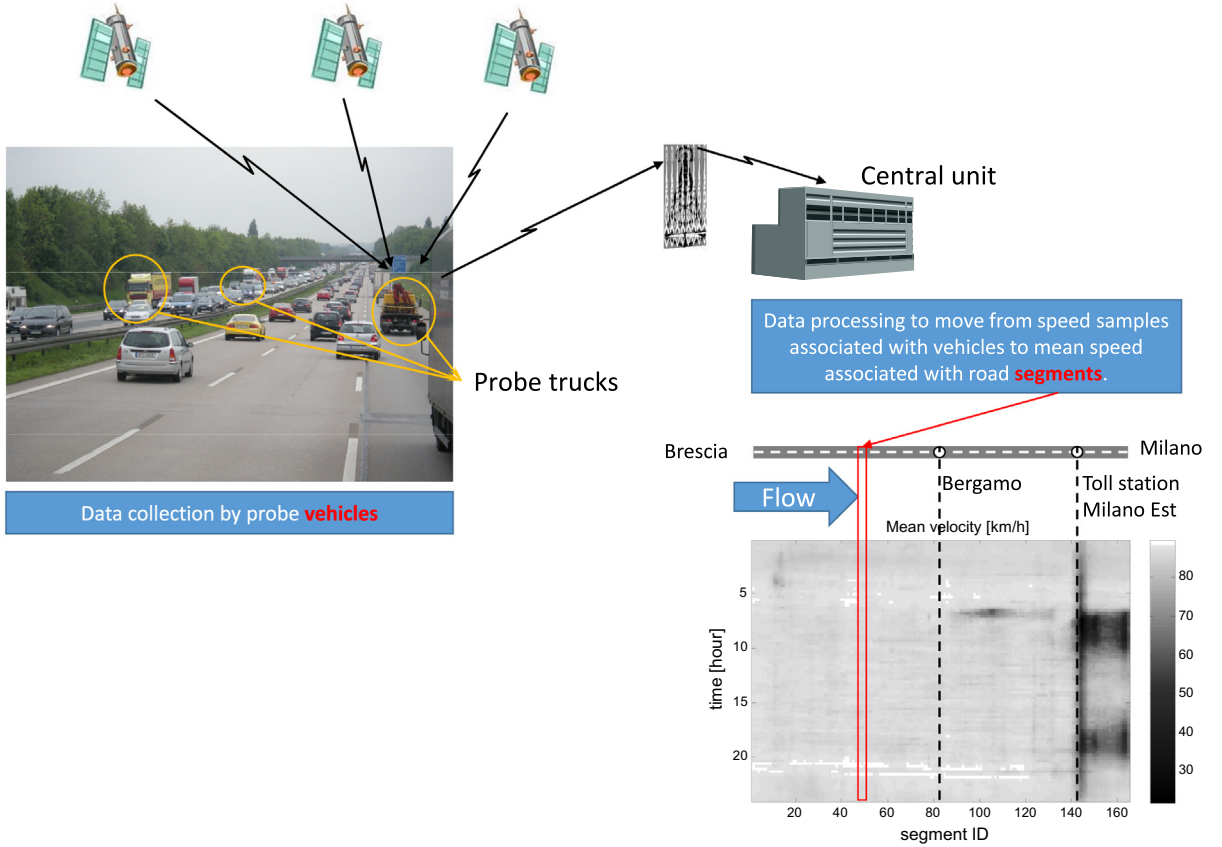
Fig. 2. Scheme describing the stream of data from probe vehicles to the central unit, which processes the data in order to extract the speed map information for the selected road section A4 Brescia → Milan.

For the computation of speed profiles associated with segments, we use only data collected by vehicles that have been active for a time window of at least 30 min, in order to exclude unreliable data from vehicles with either sporadic or non-constant transmission rate. First processing steps are those described in blocks A and B of Fig. 3. We compute the veloc-ity sample of vehicle i at time t_m as the ratio between the incremental distance covered from the last measurement, $\Delta s_{i,m}$, and the time interval, $\Delta t_{i,m} = t_m - t_{m-1}$, elapsed from the last measurement: $v_{i,m} = \Delta s_{i,m}/\Delta t_{i,m}$. We do not use instantaneous veloc-ity from GPS data as it is not averaged over $\Delta t_{i,m}$ and thus it may be less reliable for our analysis. The time interval $\Delta t_{i,m}$ in the considered FCD system can vary over time and from vehicle to vehicle; typically it is $20s \leqslant \Delta t_{i,m} \leqslant 3$ min. The evaluated velocity is assigned to the two segments associated with the measurement (i.e., the segments matched with the GPS posi-tions at times t_{m-1} and t_m) and also to the intermediate ones that the vehicle passed during the time $\Delta t_{i,m}$. This step prevents that fast vehicles generate less samples per segment than slow vehicles.

As regards vehicle filtering (blocks C in Fig. 3), we divide vehicles into classes based on their maximum detected speed, v_i^{max}, using ranges of 10 km/h, in a period of 2 months along the same highway. The large extension of the observed time period ensures a reliable estimation of the free flow speed for every detected vehicle.

Recalling that most of the probes of the considered system are trucks, to avoid biased results due to outliers with faster or slower vehicles (e.g., few cars belonging to the probe fleet), we exclude from the analysis vehicles that have not a homoge-neous behavior. We thus focus our attention on vehicles with 80 km/h $\leqslant v_i^{max} \leqslant 100$ km/h, which represent the two classes with more vehicles in our dataset.

Next step (block D in Fig. 3) is the computation of an average speed vs. time profile for each segment. Speed samples pro-vided by the vehicles are associated with the segments by map matching. Then, speed samples are aggregated in timeslots and averaged. The time interval used for aggregation is $T = 15$ min or $T = 30$ min, depending on the specific analysis (different analyses will be carried out throughout the paper, as explained in next section). We observe that in case of low traffic vol-ume, the sample size obtained with $T = 15$ min could be inadequate for a proper data analysis, but this scenario is not rel-evant for our applications as in free flow conditions the traffic speed is usually well known.

We indicate the average velocity profile on segment s at day d by the $N_T \times 1$ vector $\mathbf{v}_s^d$, which collects the N_T aggregated velocity samples for the day (e.g., for a 24 h period, $N_T = 96$ for $T = 15$ min and $N_T = 48$ for $T = 30$ min). The overall velocity map for day d, representing the velocity versus time and segment, is obtained by collecting the N_S profiles for all the
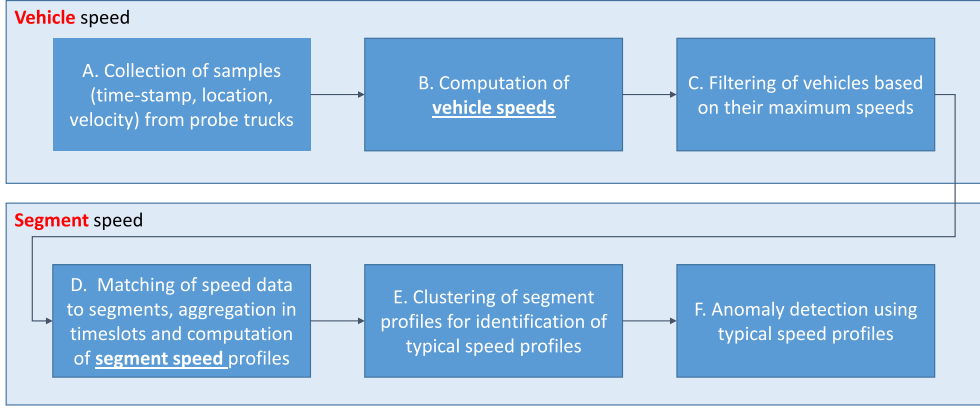
Fig. 3. Main steps of the processing method for the evaluation of typical speed profiles and the identification of anomalies.
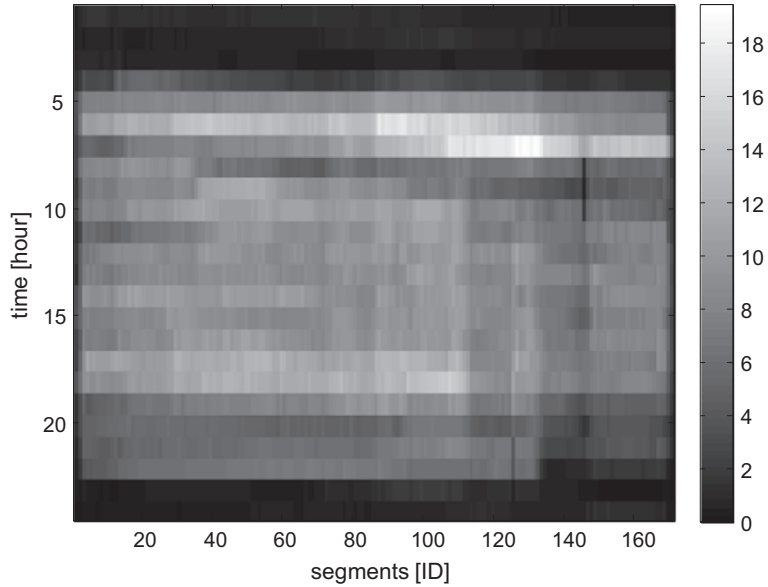

Fig. 4. Average number of monitored vehicles on Mondays.

segments into a $N_T \times N_S$ matrix $\mathbf{V}^d$. Examples of speed maps are shown - averaged over the days – on the bottom of Fig. 2 and in Fig. 5 for the two classes of vehicles 80–90 km/h and 90–100 km/h. In case of poor data, 2D linear interpolation is used to obtain a complete map.

The last step for the speed pattern computation is the clustering of profiles (step E in Fig. 3). This operation aims at aggregating the N_S velocity profiles associated to the N_S road segments into $N_C \leqslant N_S$ classes or clusters, each collecting segments with similar traffic behavior. The typical speed profile associated to class c, for $c = 1, \ldots, N_C$, is obtained as the average of all the segment profiles $\mathbf{v}_s^d$ belonging to that class.

For clustering, we adopt the Ward's method (Weijermsrs et al., 2005) which aggregates the speed profiles in homogeneous classes by constructing a tree based on a "bottom up" approach. Different clustering methods can be found in the literature (Gelbard et al., 2007). Here we select the Ward's one as it is a hierarchical approach that does not require the knowledge of the number of clusters in advance as it follows directly from the clustering process. At the beginning, a set of N_S clusters is defined, one for each profile $\mathbf{v}_s^d$. Then, clusters are paired on the basis of the minimum inner square distance metric. Pairing of clusters is repeated, until all profiles are enclosed in one single class. The number of clusters is decided putting a threshold on the desired inner squared distance. Details of the procedure can be found in Pascale et al. (2013).

Daily speed profiles obtained by the above procedure are presented and discussed in next section. The study is carried out considering speed data averaged over the days and also day-specific data, to analyse both the mean daily traffic behavior and the possible variations over the days. We make a first analysis using 24-h data, then we focus the attention on an interesting and smaller time period in the morning hours (5–12 am) where the speed variability is relevant.
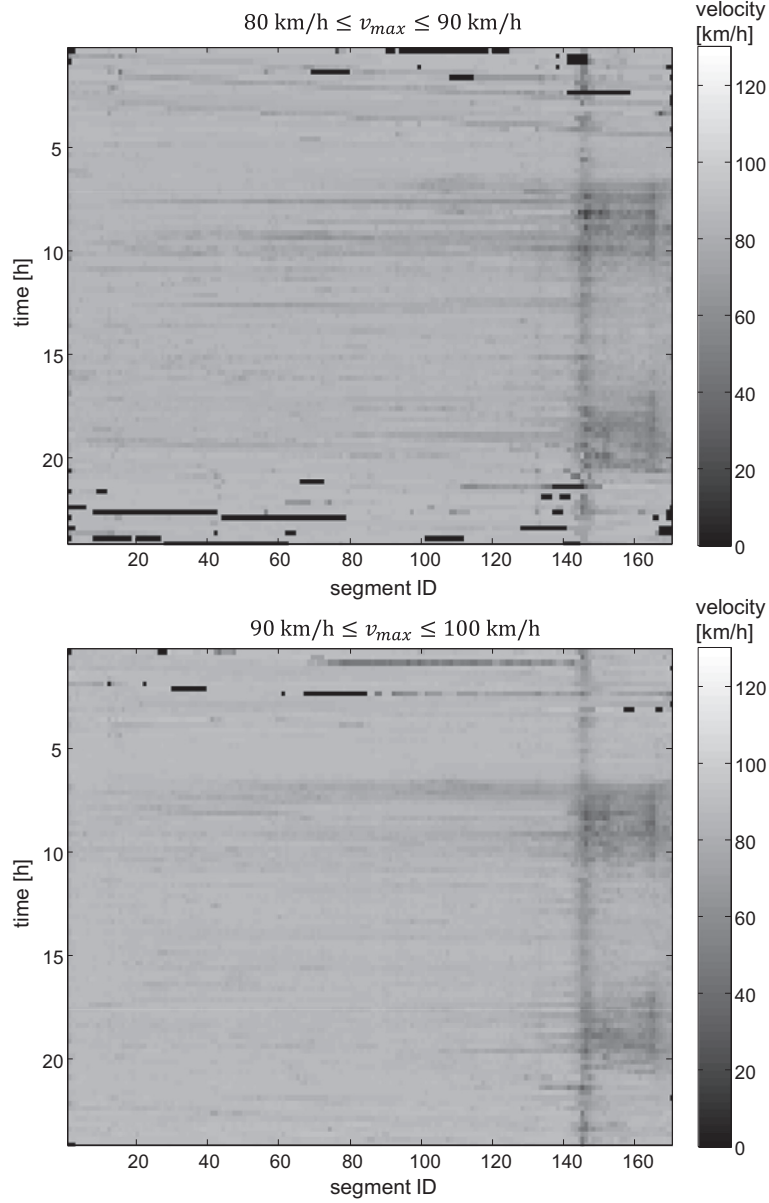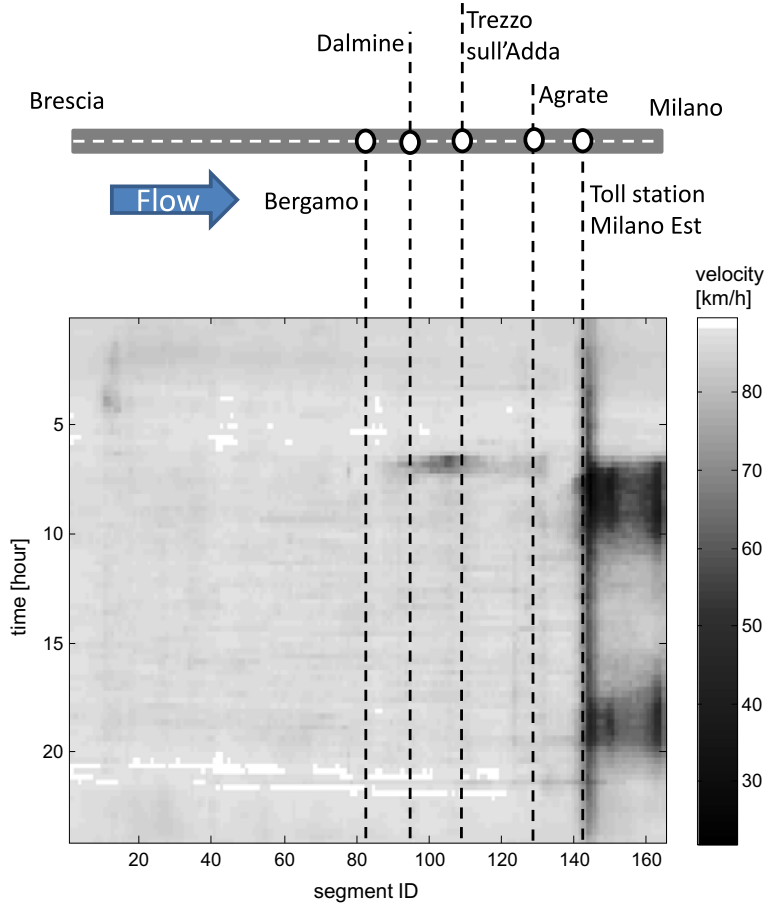
Fig. 5. Velocity maps [km/h] obtained by averaging over days and aggregating with $T = 15$ min the data from vehicles in the classes 80–90 km/h and 90–100 km/h. Black holes indicate lack of data.

3. Analysis of speed profiles

Segment speed patterns are analyzed in the following using two levels of data aggregation. In the first one, see Section 3.1, the speed data are averaged over the 3-month observation period ($N_D = 63$ days excluding week-ends and holidays), yielding a two-dimensional (2D) map $\mathbf{V}_{2D}$ of velocity vs. day time and segment. This 2D dataset is the average of the maps $\mathbf{V}^d$ over the N_D days: $\mathbf{V}_{2D} = \sum_{d=1}^{N_D} \mathbf{V}^d / N_D$. For this first analysis, a relatively high temporal resolution is used for the aggregation, $T = 15$ min (leading to $N_T = 96$ samples), since an adequate sample size can be assured by the observation over several days. The result-ing map $\mathbf{V}_{2D}$, with dimensions 96×166, is shown in Fig. 6. On the top of the figure a scheme shows the main geographical points of the related motorway section.

The velocity map $\mathbf{V}_{2D}$ is used as input of the clustering procedure in order to recognize the typical "macro" traffic behaviors that characterize the selected motorway and compute the related speed profiles. Some of these patterns can be easily recognized also by visual inspection of the speed map in Fig. 6, which shows that groups of road segments share similar behaviors. In particular, the segments in the first section of the road (segments 1–90 and 130–142) are all in free flow con-

Fig. 6. 2D velocity map [km/h] obtained by aggregating with $T = 15$ min and averaging over 63 days, using data from vehicles with 80 km/h $\leqslant v_{max} \leqslant 100$ km/h.

ditions apart from an area (segments 91–129) where a decrease in speed can be observed during morning peak hours due to congestion. On the other hand, the segments after the toll station (segments 148–166) correspond to the north ring of the city of Milan and are highly congested, especially during morning and afternoon rush hours. Finally, a systematic decrease of the speed can be observed – at any hour – in a third area (segments 143–147), close to the toll station, where vehicles slow down before approaching the tollgate. These main traffic behaviors will be automatically recognized by the clustering procedure which aggregates segments with similar characteristics and evaluates the typical speed profiles as the average of the aggregated data. The procedure will also be applied to the normalized 2D map obtained by normalizing the speeds to the segment free-flow speed, in order to catch similar time-dependent behaviors for segments with a speed offset in time series.

The second analysis, presented in Section 3.2, extends the study by recognizing typical traffic behaviors also over different days. In fact, traffic profiles may change not only from segment to segment, but also from day to day. As an example, for the selected A4 motorway BS → MI, the traffic observed during Monday morning is different from other week days due to people that commute toward Milano for the week; furthermore, anomalous events (e.g., accidents) may cause different behaviors in specific days. A detailed analysis is thus carried out to study these patterns, by applying the clustering procedure directly on the three-dimensional (3D) velocity dataset that collects the daily speed profiles associated with the 166 segments and all the 63 days: $\mathbf{V}_{3D} = \left\{ \mathbf{V}^d \right\}_{d=1}^{N_D}$. In other words, the dimension "days" is added to the 2D map in Fig. 6, yielding the $N_T \times 166 \times 63$ speed map $\mathbf{V}_{3D}$ shown in Fig. 7. Since for the number of available observations per day is moderate, a larger timeslot is used in this case for aggregation, $T = 30$ min (leading to $N_T = 48$ time samples per day). The analysis is focused on a time window that covers the morning rush hours.

3.1. Segment clustering on 2D speed dataset

The clustering procedure applied to the velocity map averaged over the days, $\mathbf{V}_{2D}$, brings to the identification of 7 clusters, as depicted in Fig. 8. The speed profile associated to each cluster – cluster profile or typical profile – is obtained by averaging
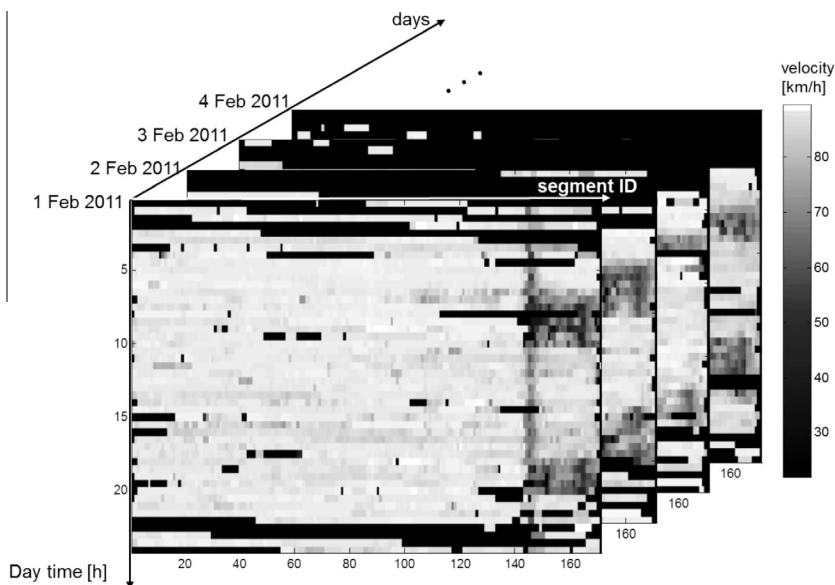
Fig. 7. 3D velocity map [km/h] obtained by aggregating on periods of T = 30 min.

the data within each cluster. The seven typical speed profiles are shown in Fig. 9, over the 24 h, together with the standard deviation range.

This segment classification gives a pretty clear and simple characterization of the traffic conditions over the motorway, with a small number of distinctive patterns, as depicted in Fig. 9. We can refer to the scheme on the top of Fig. 6 to understand the geographical meaning of the classification output. A number of regions can be recognized. Clusters 3 and 4 characterize the two free flow areas in the map, in particular cluster 4 indicates area where a lower speed is observed during rush hours in the morning. Clusters 1, 6 and 7 describe the area close to the toll station where vehicles slow down. Finally the north ring of Milan is associated with clusters 2 and 5 where congestions are observed during rush hours.

Now we apply the clustering procedure to the 2D velocity map obtained by normalizing over the free flow speed and averaging over the 3-month period. The procedure is applied to a smaller time period (5–12 am) including the rush hours and four clusters are selected. Fig. 10 shows the identified clusters and Fig. 11 the corresponding cluster speed profiles. A better description of the different traffic behaviors can be observed in this case, as the four main regions previously described can be more easily recognized. Speed profiles are clearly identified, since their overlapping is negligible, and the standard deviation is smaller than the one of the disaggregated case.

3.2. Segment clustering on 3D speed dataset

The clustering procedure is here applied to the 3D dataset $\mathbf{V}_{3D}$ in Fig. 7, limiting the analysis to the morning hours from 5 to 12 am. The resulting classes and the associated typical profiles are in Fig. 12 and Fig. 13, respectively.

Fig. 12 shows the classification of data (segments *and* days) in 5 clusters. The cluster speed profiles and the related standard deviation ranges are in Fig. 13. The standard deviation observed in each cluster is larger in this case as the profiles are not averaged over days and a wider variety of traffic behaviors is observed. In spite of this, it is interesting to observe that clusters are roughly preserved over different days, i.e. the classification is almost the same along the axis of days, apart from some local fluctuations and anomalous events. The space description of the motorway reveals a wider background zone where cluster 1 can be applied, and a reduced zone classified as cluster 2, which has a profile with only a short speed decrease of 25% at 7 am. The other three clusters describe heavy congestion phenomena which are mainly located at the final part of the motorway (around segment 143) where the toll station operations and suburban trips to Milan are relevant. How-ever, in Fig. 13 it is possible to note that cluster 5, which has a speed reduction greater than 30% between 7 and 9 am, is assigned also to other segments in sporadic days. These cases are likely to be related to anomalous events, as they fall in areas normally assigned to clusters 1–2.

We use the CCISS[2] data, published on the web for describing traffic events on the A4 motorway MI → BS, to verify the anomalous events discussed above. Four real traffic events detected by CCISS can be actually associated to these anomalies, as reported in Fig. 13:
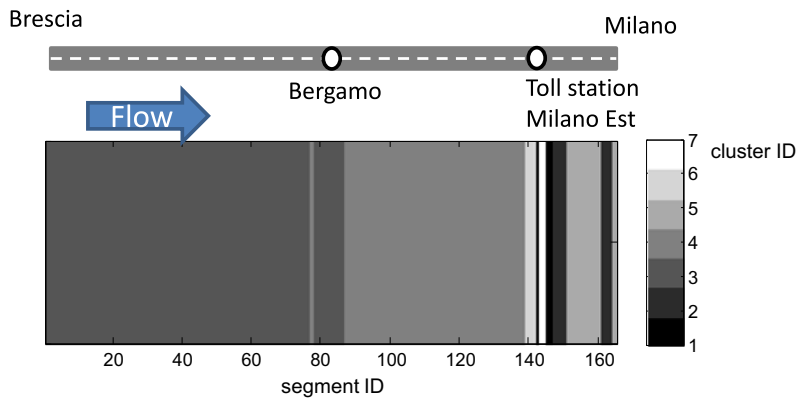
Fig. 8. Motorway segment classification by clustering of 2D velocity data.
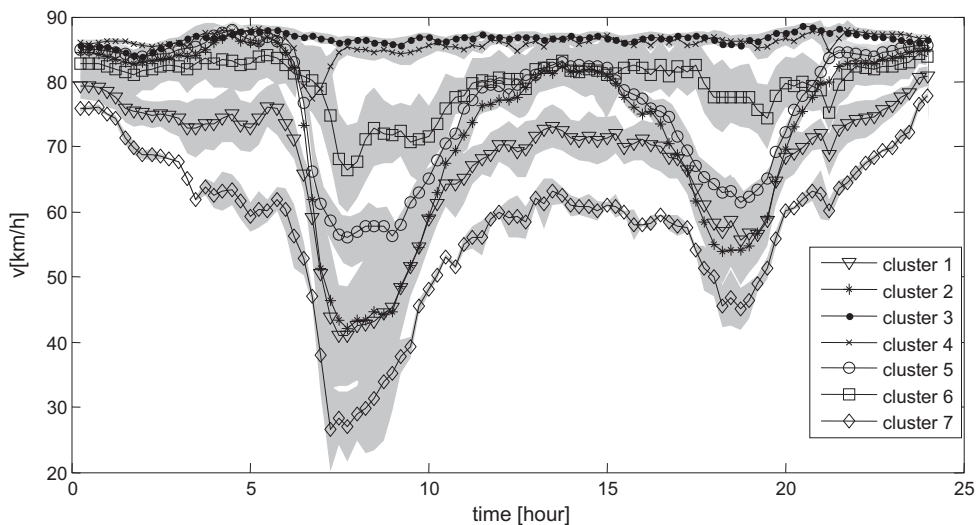

Fig. 9. Cluster speed profiles [km/h] associated with the 7 clusters in Fig. 8. Shaded areas represent the standard deviation range.

1. Wide jam between Palazzolo (km 195.2) and Rovato (km 203.4) due to an accident, at 6:43, on February 3rd 2011.
2. Wide jam – 8 km long – between Seriate (km 181.1) and Capriate (km 162.7) due to an accident, at 6:20, on February 25th 2011.
3. Accident between Seriate (km 181.1) and Bergamo (km 174.5), at 6:36, on March 10th 2011.
4. Accident between Rovato (km 203.4) and Palazzolo (km 195.2), at 8:43, on March 10th 2011.

We can conclude that by analysing single-day speed time series, the typical speed patterns of the motorway can be better identified and it is also possible to recognize anomalous events.

4. Segment classification and its use for road traffic applications

One of the most relevant traffic engineering applications, based on the synthetic information displayed in the previous section, is the easy identification of critical zones along the motorway where the traffic behavior needs to be better observed, e.g. by an infrastructure-based monitoring system with fixed detectors. Indeed, in the analyzed scenario only a part of the motorway reveals a variable speed over time (clusters 2 to 4 in Fig. 11) while for many segments (more than 80) the speed behavior of traffic is almost constant and does not need to be further observed or modeled (clusters 1, associated to free flow). Since fixed installations for traffic monitoring usually require not negligible resources (Pascale et al., 2012), a useful support for sensor location can be derived by this type of analysis based on floating vehicle data.

Another possible application of segment speed profile classification in transport modeling is large scale network building, where reference values are usually needed to calibrate cost functions and vehicle speed, in order to estimate road
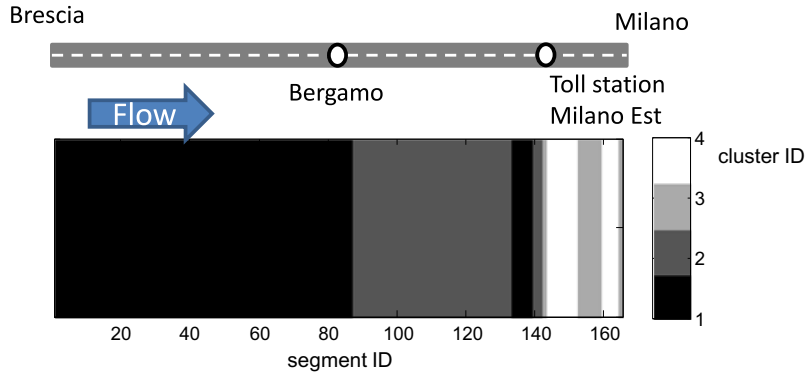
Fig. 10. Motorway segment classification by clustering of normalized 2D velocity data on morning hours 5–12 am.
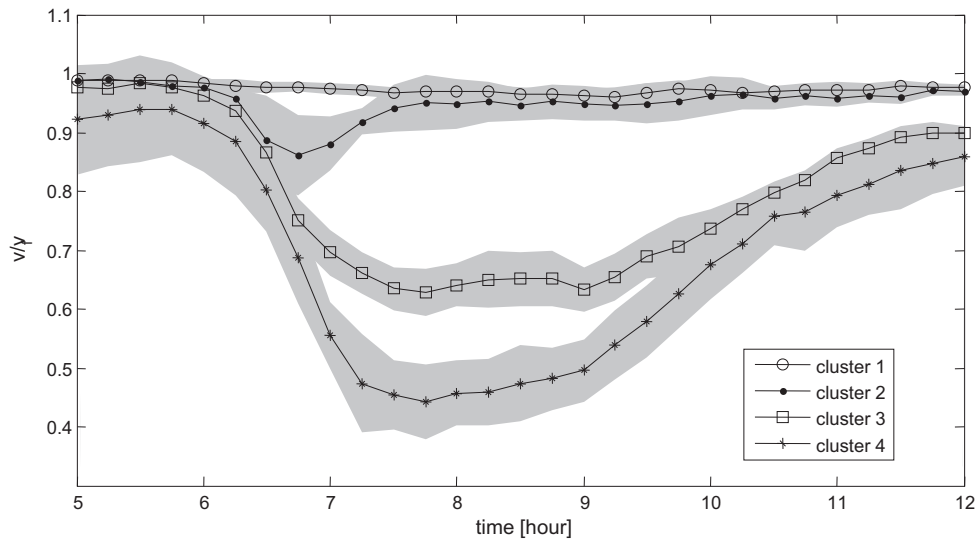
Fig. 11. Cluster speed profiles (normalized speed) associated with the 4 clusters in Fig. 10. Shaded areas represent the standard deviation range.

performance on average or in most frequent cases for the different elements of the road network. Also in medium/short term traffic control applications, information about speed along the road may be useful for many reasons, such as travel-time estimation for driver information or vehicle routing operations in freight delivery, where the *delivery time* can be estimated on the base of segment speed identification and updated in case of *anomaly detection*.

Two of the above mentioned applications, anomaly detection and travel-time estimation, are discussed more in detail in the following subsections.

4.1. Anomaly detection using cluster profiles

We consider here the anomaly detection step enclosed in block F of Fig. 3. The cluster speed profiles provided by the clustering method represent the typical behavior of speed on the road in terms of mean value and variance. Thereby, the comparison between an incoming speed profile and the typical one for the same segment enables the automatic recognition of anomalous situations. As required in detection theory, we need to define a threshold for this process.

We focus our attention on the sequence of segments from $s = 1$ to $s = 120$ where we can recognize essentially two clusters, indexed as 1 and 2 in Fig. 12. We begin our analysis by comparing - in terms of Euclidean distance – the typical normalized speed profile associated to cluster 1 shown in Fig. 13, with the normalized segment speed profiles of all days. Fig. 14 shows the Euclidean distance of the profiles on days of March, April and May. If we compare Fig. 14 with Fig. 12 we can observe that the four anomalies highlighted as red circles in Fig. 12 can be recognized when the distance is above 1 for at least 10 consecutive segments. If we analyse cluster 2 we observe a similar behavior. Automatic anomaly detection, thereby, could be performed by joint threshold detection over a set of consecutive segments.
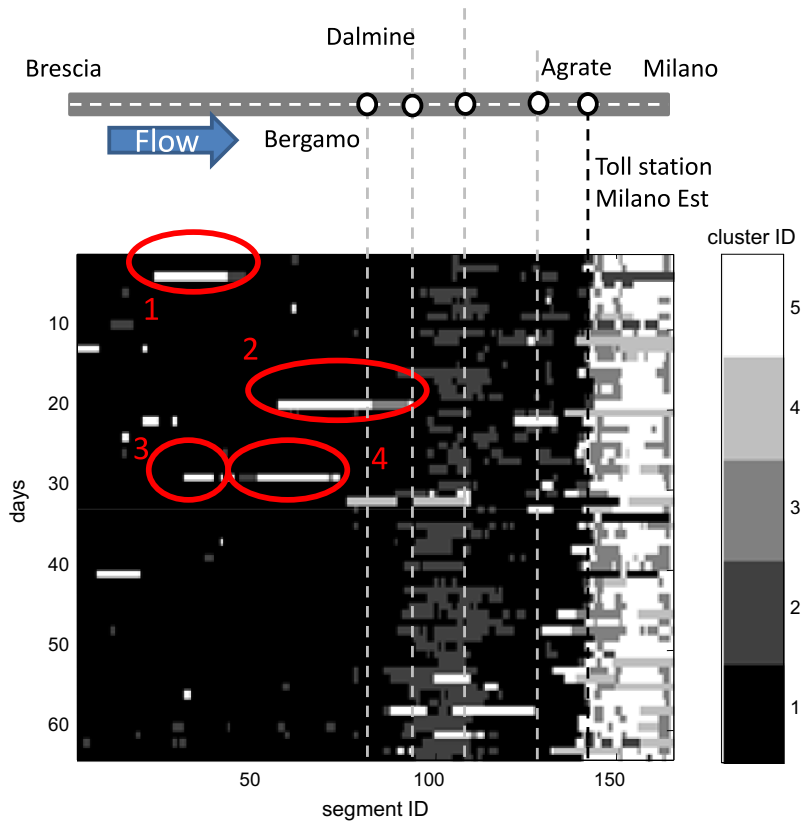
Fig. 12. Segment-day classification by clustering of 3D velocity data on morning hours 5–12 am. Red circles indicate anomalies manually verified using CCISS data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
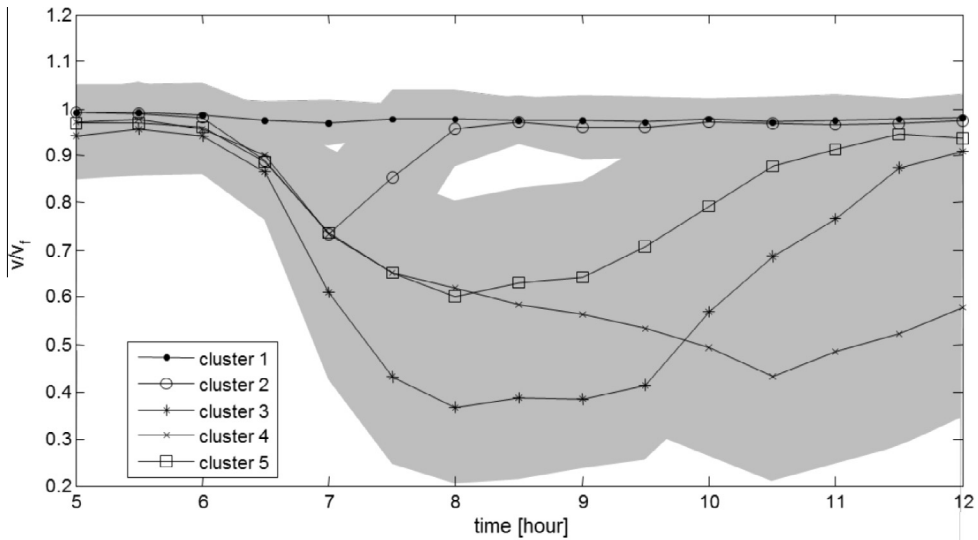

Fig. 13. Cluster speed profiles associated with the 5 clusters in Fig. 12. Shaded areas represent the standard deviation range.

For validating the proposed detection approach, we use the second dataset collected in June (as described in Section 2.1). In Fig. 15 the distance between each speed profile coming from the validation dataset and the cluster-1 typical profile is shown. Gray areas represent weekends and holidays that are not taken into account in the analysis, while the red square
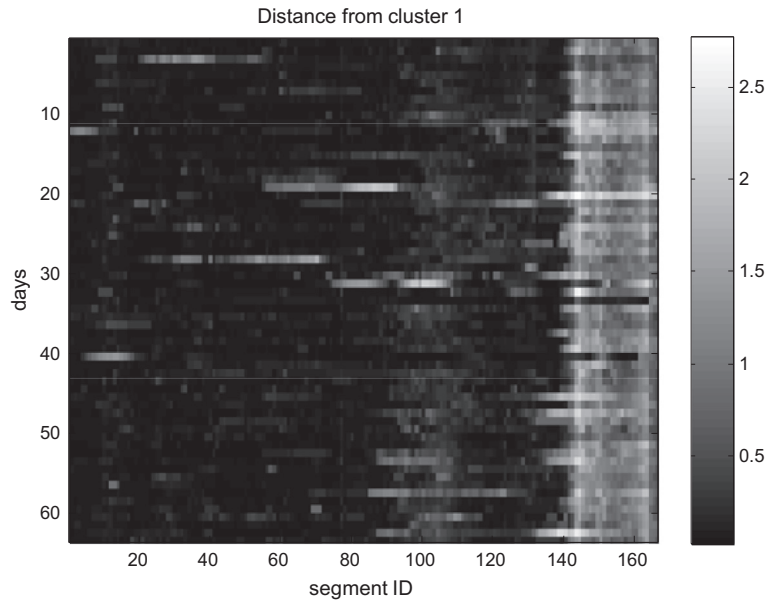
Fig. 14. Euclidean distance between the segment speed profiles collected on March, April and May and the cluster-1 profile in Fig. 13.
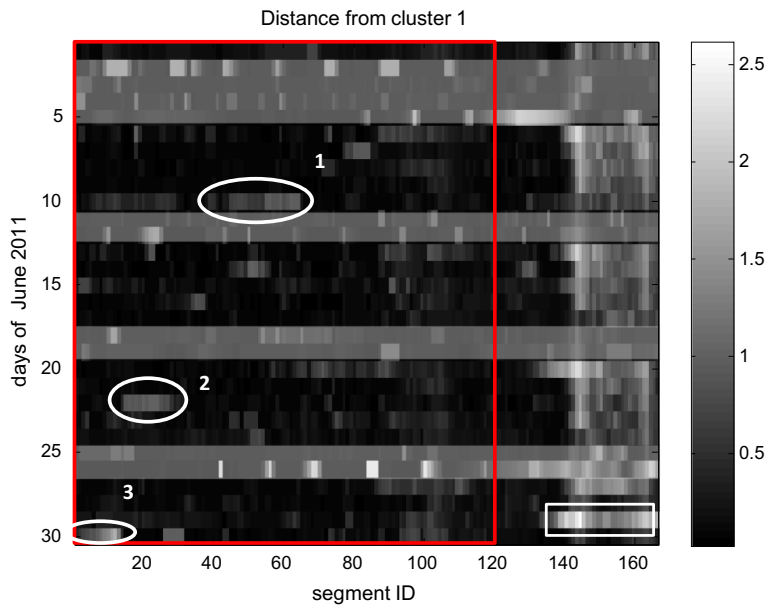
Fig. 15. Euclidean distance between the speed profiles collected on June and the cluster-1 profile in Fig. 13.

marks the area in which the detection process gives positive results. The three white circles indicate the sequence of segments that in a specific day obey our conditions. By cross-checking these results with the CCISS data, we find that the three detected areas can be related to the following events registered by CCISS:

Event 1. Jam of 3 km length due to a vehicle broken down between Grumello (km 189.7) and Seriate (km 181.1), at 8:08 on June 10th 2011.
Event 2. Two causes can be related to this event:
    a. Accident between Brescia Ovest (km 217.3) and Ospitaletto (km 208.4), at 11:12 on June 22th 2011.
    b. Roadwork between Brescia Ovest (km 217.3) and Ospitaletto (km 208.4), at 10:46 on June 22th 2011.
Event 3. Two causes can be related to this event:
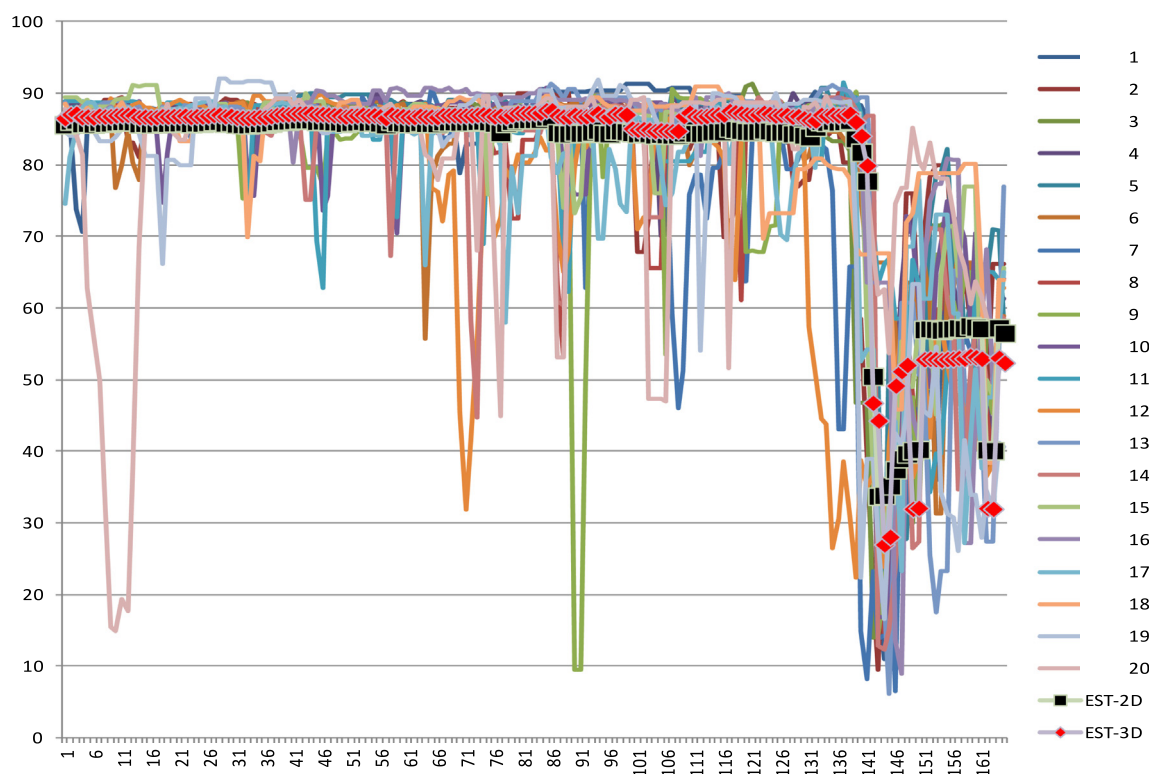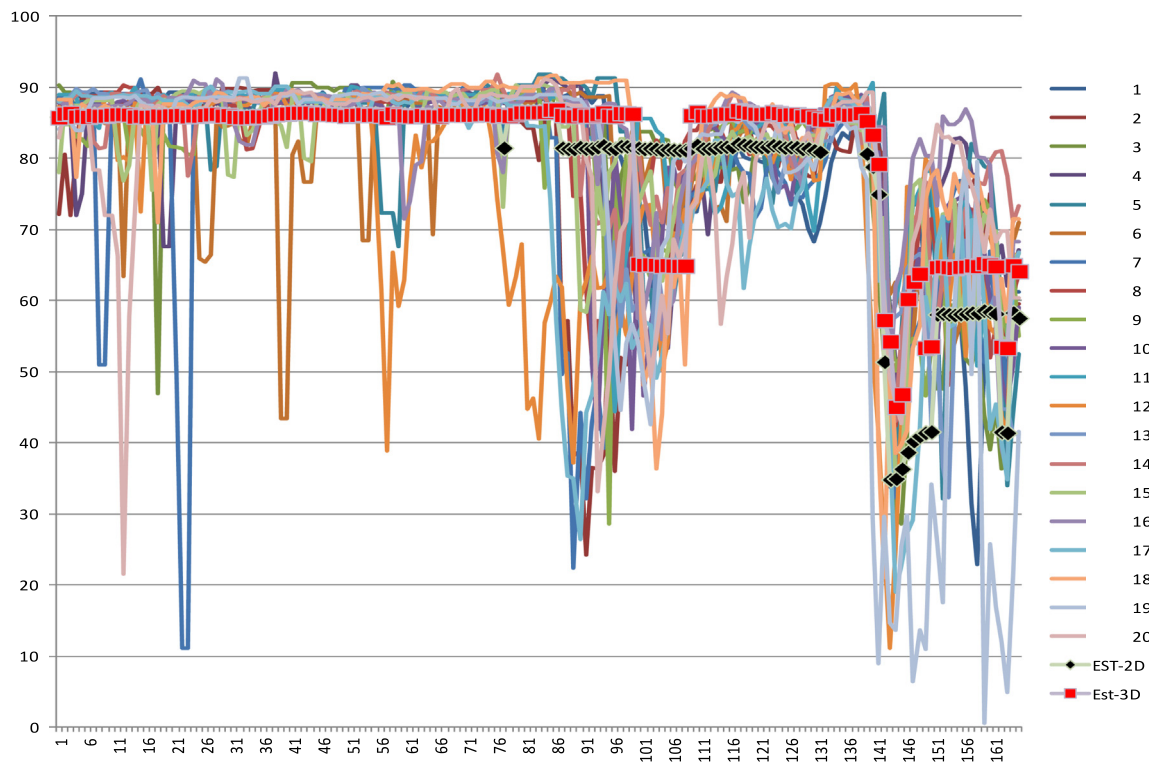
Fig. 16. Estimated speed [km/h] (clustering on 2D and 3D data) on segments [Id.] vs. observed speed for 20 days in June. Results are reported for 7 am (top) and 8 am (bottom).
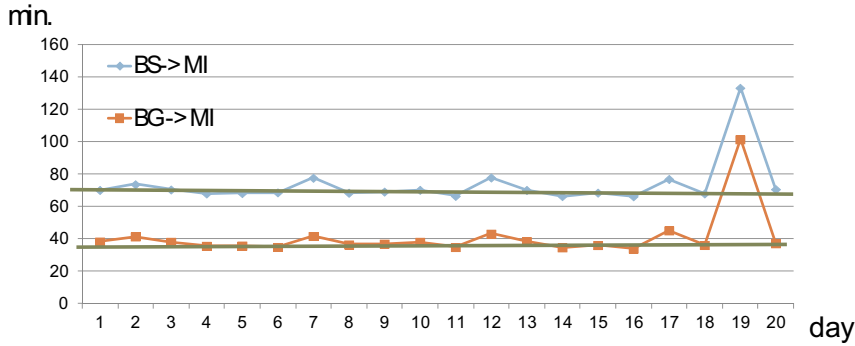
Fig. 17. Estimated vs. observed travel time for 20 days in June.

a. Jam due to a vehicle broken down between Brescia Ovest (km 217.3) and Ospitaletto (km 208.4), at 6:20 on June 30th 2011.
b. Vehicle burning between Ospitaletto (km 208,4) and Rovato (km 203.4) at 5:51 on June 30th 2011.

From the above results we can conclude that the anomalies recognised by the detection procedure *correspond to real congestion events*. This confirms that the proposed methodology can actually be useful for providing reference profiles for anomalous event detection.

4.2. Speed and travel time estimation for freight delivery operations

In many applications in transport systems, speed information is used to predict the travel time along a route, which is composed by a number of links where traffic conditions frequently change over time. In fleet management operations, for example, a reliable estimate of a vehicle travel time for delivery and pickup operations can be useful if the service is established to comply time constraints. To this aim, in the following sections two tests are reported on the estimation of speed and travel time data, for selected days in June 2011, in the A4 motorway BS → MI. Estimates are obtained using the typical speed profiles provided by the clustering results based on 63 working days in February, March and May 2011.

The comparison between the actual speed and the corresponding estimate obtained using the typical profiles is shown in Fig. 16, over 20 days of June, in each of the A4 segments, at 7 am (top figure) and 8 am (bottom figure). Estimates are drawn from the typical profiles resulting from the 2D and 3D velocity datasets as described in Sections 3.1 and 3.2. In the 2D case, the speed estimate is read directly from the cluster profile associated to the segment. In the 3D case, for any segment the speed pattern of the most frequent cluster observed over the 20 days has been chosen for the two selected time instants. It is possible to note the difference in the two time slices, where the congestion phenomena observed in segments around the 100th at 7 am move mainly in the downstream part of the motorway (after the 140th segment) at 8 am. If we look at the diagram at 7 am, the congestion is confirmed by the low speed values around the 100th segment (approx. 65 km/h for 3D and 80 km/h for 2D clustered data). On the other hand, we can notice in the 8 am plot that the speed at the 100th segment is increased to the free flow value, while in the downstream segments (after the 140th), the speed is lower than the one observed at 7 am (e.g. the 3D clustered data show that the speed of 65 km/h observed at 7 am is never reached at 8 am). The 3D map clustering, although also the 2D map clustering captures the global trend, better predicts the location of these phenomena.

Since the use of speed data is mainly for travel time estimation in practical applications, in the following we assess the accuracy of the estimate of the travel time for a journey along the A4 motorway using the typical speed profiles for estimation. The peak hours are selected to perform the test. For the time slice 7 am, from Brescia to Milan (BS → MI), the journey duration is estimated in 71 min, if the 2D clustered data is used, and 68 min if the speed pattern of the most frequent cluster is selected from the 3D clustered data, while 62 min is the time estimation assuming the speed value in free flow conditions. For the same time slice, the journey duration from Bergamo to Milan (BG → MI) is estimated as equal to 38 min, based on the 2D clustered data, and 36 min using the most frequent cluster of the 3D clustered data, while only 31 min is the estimation in free flow conditions.

The comparison of the travel time estimates with the actual parameters observed in the 20 days of June is reported in Fig. 17. The estimate is shown to be accurate for all days, apart from the day 19th where a relevant error is observed on both BS → MI and MI → BS journeys. This error is due to a number of anomalous events, as recorded by CCISS on this day. Looking at Fig. 15, we can observe that all these phenomena are related to segments located after the toll station as highlighted by a square on the bottom-right section of the figure. Also, for the days 2, 7, 12 and 17, a systematic difference occurs, but this can be explained observing that on Monday traffic flow in the morning hours is usually higher than on other days.

5. Conclusions

In this paper a method for the identification of typical traffic patterns has been proposed and validated with a motorway test case, by using data from truck probes as well as data collected on occurred accidents. A main result is that in the selected A4 motorway (Turin–Venice), within the Brescia to Milan section of approximately 90 km length, a small set (4–5) of typical speed patterns is enough to describe the average traveling speed over the morning hours. The proposed method is able to efficiently characterize a variety of traffic behaviors observed over the entire road section.

The analysis on speed data has been used as starting point to outline a procedure for anomaly detection. The typical profiles computed using data collected over three months have been used as reference to recognize anomalies occurred over the selected motorway route. Results have been verified using information on the anomalous events registered by the national provider of road traffic information (CCISS) over the considered road section.

Speed information can be nowadays used for different aims, such as:

- transport planning, when reference traffic values are needed to estimate the road performance on average or in most frequent cases, referred in the case to heavy-duty vehicles;
- medium and short-term traffic applications, where speed data along the road are needed for travel-time estimation or truck routing operations;
- traffic management, as historical speed time series can be used as baseline for travel time estimation, navigation, fleet management (time windows), when other data are not available;
- traffic flow modeling, where historical speed time series can be used for validation of the models and critical segments need to be selected along the road for a detailed traffic data collection.

In the next future, after the compulsory introduction in Europe of black boxes on new vehicles planned since October 2015, namely for eCall, the analysis we provided might be gradually enriched even for reconstructing overall accidents, for providing instantaneous risk analyses in flows both for primary and secondary accidents, for updating traffic data in real-time, for allowing optimal shift from traction to propulsion for electric-ICE hybrid vehicles, for collecting images associated with any event from the front of the vehicle, for transmitting these information through vehicle-to-infrastructure communications.

Some critical points need to be addressed as future work. Although the assessments on real traffic event data showed already usable and promising results for anomaly detection, the size of the dataset adopted for this first study was at last moderate and might be extended for a more accurate analysis, including further data over time and road segments. Moreover, to enhance the reliability of the estimates, further sources of traffic information, such as flow or density, could be useful to integrate the speed information, by merging or comparing data with those collected through devices on or close to the infrastructure. A further development is also the investigation of different clustering methods to explore if relevant effects on classification occur for the specific application examined in this paper.

The proposed procedure is planned to be used for the creation of a database that will integrate historical speed information – possibly with data from other types of probes or fixed detection systems as those used by CCISS – for supporting applications such as travel-time estimation, fleet management, navigation and automatic incident detection.

References

Chrobok, R., Kaumann, O., Wahle, J., Schreckenberg, M., 2004. Different methods of traffic forecast based on real data. Eur. J. Oper. Res. 155 (3), 558–568.

Chung, E., 2003. Classification of traffic pattern. In: Proc. 10th World Congress on Intelligent Transport Systems. Madrid, Spain.

De Fabritiis, C., Ragona, R., Valenti, G., 2008. Traffic estimation and prediction based on real time floating car data. In: Proc. 11th International IEEE Conference on Intelligent Transportation Systems (ITSC), Beijing, China, pp. 197–203.

Fangfang, Z., Van Zuylen, H., 2013. Urban link travel time estimation based on sparse probe vehicle data. Transp. Res. Part C: Emerg. Technol. 31, 145–157, June 2013.

Gelbard, R., Goldman, O., Spiegler, I., 2007. Investigating diversity of clustering methods: an empirical comparison. Data Knowledge Eng. 63 (1), 155–166.

Herrera, J.C., Work, D.B., Herring, R., Ban, X., Jacobson, Q., Bayen, A.M., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment. Transp. Res. Part C: Emerg. Technol. 18 (4), 568–583. August 2010.

Kerner, B.S., Demir, C., Herrtwich, R.G., Klenov, S.L., Rehborn, H., Aleksic, M., Haug, A., 2005. Traffic state detection with floating car data in road networks. In: Proc. 8th International IEEE Conference on Intelligent Transportation Systems (ITSC), Vienna, Austria, pp. 700–705.

Kerper, M., Wewetzer, C., Trompeter, H., Kiess, W., Mauve Kerper, M., 2011. Driving more efficiently – the use of inter-vehicle communication to predict a future velocity profile. In: Proc. 73rd IEEE Vehicular Technology Conference (VTC Spring), Budapest, Hungary, pp. 1–5.

Kwon, J., Petty, K., Varaiya, P., 2007. Probe vehicle runs or loop detectors?: effect of detector spacing and sample size on accuracy of freeway congestion monitoring. Transp. Res. Rec.: J. Transp. Res. Board 2012 (1), 57–63.

Pascale, A., Nicoli, M., Deflorio, F., Dalla Chiara, B., Spagnolini, U., 2012. Wireless sensor networks for traffic management and road safety. IET Intell. Transp. Syst. 6 (1), 67–77, Mar. 2012.

Pascale, A., Deflorio, F., Nicoli, M., Dalla Chiara, B., Pedroli, M., 2013. Typical speed profiles on motorways from floating vehicle data. In: Environment, Land Use and Transportation Systems. Selected Papers, Franco Angeli, Scientific Seminar SIDT, Venice, 6 October 2011, VL. 1797.42, pp. 272–286.

Rahmani, M., Koutsopoulos, H.N., 2013. Path inference from sparse floating car data for urban networks. Transp. Res. Part C: Emerg. Technol. 30, 41–54, May 2013.

Rakha, H., Van Aerde, M., 1995. Statistical analysis of day-to-day variations in real-time traffic flow data. Transp. Res. Rec., 26–34.

Sethi, V., Bhandari, N., Koppelman, F.S., Schofer, J.L., 1995. Arterial incident detection using fixed detector and probe vehicle data. Transp. Res. Part C: Emerg. Technol. 3 (2), 99–112.

Treiber, M., Kesting, A., Wilson, R.E., 2010. Reconstructing the Traffic State by Fusion of Heterogeneous Data. Computer-Aided Civil and Infrastructure Engineering. Wiley.

Vandenberghe, W., Vanhauwaert, E., Verbrugge, S., Moerman, I., Demeester, P., 2012. Feasibility of expanding traffic monitoring systems with floating car data technology. IET Intell. Transp. Syst. 6 (4), 347–354, December 2012.

Wei, C.H., Lee, Y., 2007. Development of freeway travel time forecasting models by integrating different sources of traffic data. IEEE Trans. Vehic. Technol. 56 (6), 3682–3694.

Weijermsrs, W., van Berkum, E., 2005. Analyzing highway flow patterns using cluster analysis. In: Proc. 8th International IEEE Conference on Intelligent Transportation Systems (ITSC), Vienna, Austria, pp. 831–836.

Wild, D., 1997. Short-term forecasting based on a transformation and classification of traffic volume time series. Int. J. Forecast. 13 (1), 63–72.

Yang, J.S., 2005. Travel time prediction using the GPS test vehicle and kalman filtering techniques. In: Proc. IEEE American Control Conference, Portland, OR, USA, pp. 2128–2133.

Yoon, J., Noble, B., Liu M., 2007. Surface street traffic estimation. In: Proc. 5th ACM International Conference on Mobile Systems, Applications and Services (MobiSys). San Juan, Puerto Rico, USA, pp. 220–232.