ORIGINAL PAPER   **Open Access**

Integrating mobility data sources to define and quantify a vehicle-level congestion indicator: an application for the city of Turin

Miriam Pirra[*] ⓘ and Marco Diana

Abstract

Purpose: Traffic congestion is a large-scale problem in urban areas all over the world that can lead to substantial costs for travellers and business operations. This paper focus on how to measure the way in which congestion selectively affects different traffic streams, with special emphasis on light duty vehicles travelling around a city.

Methods: The idea is to integrate a dataset collecting Global Positioning System (GPS) vehicle traces with road side data sources related to traffic conditions in a road network, which on the other hand usually lack focus on specific traffic streams. The core of the data integration method is the creation of a specific indicator focusing on the time lost in congestion. This is a Key Performance Indicator (KPI) of an urban network that is of paramount importance as a decision support tool for policy makers, also because it has an impact on other key issues such as air pollution, noise emissions, energy efficiency and health problems. Then, a method is proposed to quantify the congestion KPI in a highly disaggregated fashion (each single vehicle travelling on each single link or street segment).

Results: This KPI can be used to inform a wide range of policy actions within the transport sector, both from the viewpoint of a city and from that of an individual actor of the transport system, such as the operator of a fleet of vehicles for urban freight deliveries. Some preliminary examples of how the aggregation of the KPI at different scales can provide insights into the transport system are presented.

Keywords: Urban mobility, Data integration, GPS traces, Freight delivery logistics, Congestion, Passenger mobility, KPI

1 Introduction

At the European Union level, the annual cost of the congestion is estimated between € 146 and 243 billion (1 to 2% of the total GDP) [1]. Moreover, about 28% of greenhouse gas emissions are caused by transport, with 84% of these emissions coming from road transport, while more than 10% of the carbon dioxide emissions resulted from the urban road traffic [2]. Reducing congestion is therefore one of the primary goals of virtually any transport policy measure.

On the whole, data required for congestion management could be very costly in terms of work required for surveys to collect the data. Traditional traffic data sources include roadside devices such as magnetic loops, road tube counters, radar or, more recently, Bluetooth [3]. Since roadside measures usually cover only a tiny fraction of all roads in an urban area and given the need to forecast traffic conditions, traffic simulation models are usually implemented through different techniques [4]. The resulting traffic flow quantities (e.g. average speed, link occupancy, corridor density …) are then analysed to identify different road conditions, such as congestion, free-flow, etc.

Nowadays, the collection of traffic flow data can also be more massively done with the help of on board tracking devices, such as Global Positioning System (GPS) and mobile phone data from probe vehicles. These latest technologies were proven feasible and helpful in

* Correspondence: miriam.pirra@polito.it
DIATI - Department of Environment, Land and Infrastructure Engineering,
Politecnico di Torino Corso Duca degli Abruzzi, 24, I-10129 Torino, Italy

analysing traffic trends, which could result in cost-effective measures to mitigate traffic congestion, and they can even be exploited to derive travel demand (origin-destination pairs) patterns [5]. In particular, Floating Car Data (FCD), has gained momentum due to its lower cost and higher coverage [6], despite some reliability problems [7]. The main idea behind FCD is the collection of real-time traffic data by locating the vehicle via mobile phones or GPS over the entire road network. Usually, information on car location, speed and direction of travel are sent anonymously to a central processing centre. These data could then be processed to derive measures such as travel time or average speeds through road segments [8–10]. As GPS has become increasingly common, it is typically used to monitor fleet management services such as taxi drivers, delivery vans or trucks [11, 12]. Since these vehicles are expected to travel around the city for their duties, they could provide useful insights on the traffic trends, so that urban congestion information could be extracted [13, 14].

Traffic-related measures that were introduced in the previous two paragraphs show some complementary features. Roadside devices allow for a continuous monitoring of flows across a road section, thus enabling the study of how traffic conditions evolve over time. However, these monitoring activities are spatially bound to the road segments where the "hardware" is installed and properly working. While some technologies such as weight in motion can provide different figures by kind of vehicle, it is in general not possible to obtain more disaggregated information, for example to study which origin-destination patterns are more affected by congestion in a given area.

On the other hand, vehicle-based measures are completely tracking each vehicle from trip origins to destinations. However, fleet coverage is still fairly low in most countries, while GPS traces are collected for a variety of purposes by different stakeholders (e.g. transit fleets or freight delivery monitoring, car insurance data used for assistance in case of accidents, feedback from personal navigation devices and crowdsourcing applications). This leads to a fragmentation of available data, which compounds with both privacy and commercial confidentiality issues in making it difficult for public decision makers to gain access to all potentially available information.

The state of the art in this sector can therefore be seen as a transition from roadside to on board measures, which will hopefully provide a complete monitoring of traffic and congestion in a given area in the future. In this situation, it might be interesting to jointly consider both kinds of measures to exploit their strengths. This aspect has been rarely considered in past research work. In particular, in the present paper we compare vehicle-based measures, such as GPS traces for a specific link,

with the corresponding roadside traffic flow measures to better understand to which extent that specific vehicle is affected by congestion. Related existing studies generally focus in matching network links to GPS traces [15, 16]. The presented approach is however different, since the goal is not to join the spatial information between the datasets but rather to analyse travel times. It will therefore be shown in the methodological section that in such cases it is both sufficient and less computationally intensive to focus on the association between GPS recordings and the two nodes at the extreme of each arc.

This paper therefore proposes a method to quantify in a highly disaggregated fashion (each single vehicle travelling on each single link or street segment) the time lost in congestion, through the integration of different traffic-related data sources that are typically available by transport municipalities but that are seldom jointly exploited. The related Key Performance Indicator (KPI) is among those recommended at the European level thanks to the work done in projects as DISTILLATE or CONDUITS that have set a framework for a better evaluation of urban mobility conditions [17, 18]. The KPI proposed in the present paper looks at the so-called Travel Time Index extending the traffic information to all the time ranges of the day. Starting from the aforementioned background, the KPI presented in this paper can be used to inform a wide range of policy actions within the transport sector, both from the viewpoint of a city and from that of an individual actor of the transport system, such as the operator of a fleet of vehicles for urban freight deliveries. A vehicle level congestion KPI can in fact augment the visualization of congestion maps for more solid and tailored policy guidance [5].

The paper unfolds as follows. The following section includes the description of the two datasets collecting information about the mobility in the city of Turin on whom the whole methodology is based. Section 3 describes how the two datasets are integrated, while the definition of the KPI is presented hereafter. Finally, examples of visualization of the results are provided.

2 Datasets description
The work presented in this paper is part of the research carried out within the European H2020 SUITS project. SUITS (Supporting Urban Integrated Transport Systems: Transferable tools for authorities – http://suits-project. eu/) aims at increasing the capacity of local authorities to develop and implement sustainable, inclusive, integrated and accessible transport strategies, policies, practices and measures. Among the project activities, two datasets have been made available by the City Council of Turin, a city of about 900,000 inhabitants in the North-West of Italy.

The first dataset (Dataset1 throughout the paper) contains the traffic flows collected in the month of May 2017 on the main roads of the city network, which is made up of 5980 links (roads) of Turin and its surroundings. Moreover, latitude and longitude of all the nodes in the graph of the city road network are provided in association with Dataset1. Graph density is higher in the central area of the city, where virtually every street is represented through the graph, whereas only the main streets are included in the remainder of the metropolitan area. Information is provided with a hourly frequency for each arc and it contains: date, hour range (beginning from midnight until 11 pm), average flow (veh/h), travel time on the arc (sec), source of the data (measured on field through inductive loops or estimated by a traffic flow model). The traffic flow model used to estimate flow conditions for arcs where measures are not available was implemented by 5T, an in-house company of the Turin municipality and the Piedmont region that is in charge of the implementation of technological systems to support mobility solutions [19]. This dataset is very useful for both monitoring and planning purposes given the detailed information on the traffic in the city. However, its main limitations stand in the use of a combination of measured and modelled data and in the absence of, for example, a partition between heavy and light vehicles.

The second dataset (Dataset2 throughout the paper) includes the traces, i.e. GPS locations, of a fleet of logistic vehicles that are delivering goods inside the city. There are exactly 28 vehicles, for which the GPS traces that were recorded over a month period (29/04/2017–29/05/2017) are considered in this research. A special agreement between the City Council and the haulage companies operating these vehicles was concluded, granting easier access conditions to travel inside the limited traffic zone in the central part of the city against the disclosure of such GPS traces. The data available for each record include the position (latitude and longitude), the time and date of acquisition, the average speed, the direction (course) with a sampling frequency of approximately 10 sec, leading to 360,820 recordings in total. One of the main drawbacks of this dataset stands in its small sample size, since only 28 vehicles are tracked, thus not allowing to draw meaningful information related to general traffic flow conditions. This is a typical situation of many mid-sized cities, which might have access to the GPS traces of a limited number of vehicles that cannot constitute a statistically representative sample for the whole network, 24 h a day and 365 days a year. Moreover, the analysis of such kind of data usually would require quite elaborate post processing to derive general traffic flow measures.

The methodology presented in the paper is based on the integration of these two datasets. GPS data provided by

vehicle fleets of trucks or delivery vans, such as those collected in Dataset2, have usually been exploited to retrieve information on various aspects, such as freight performance measures [20], commercial vehicle tour activity [21], trucks routing behaviours [22] or delivery stops identification [23]. Although some examples of traffic data fusion coming from different sources could be found in literature [24–26], one of the main innovative features of the presented methodology lies in the procedure used to integrate a dataset coming from the road infrastructure side with one derived from road users. In fact, the present paper focuses on the GPS traces of the above mentioned freight operators, since related routes cover a wider portion of the whole network compared for example to the routes of public transport bus lines. However the data fusion methodology presented below can be used with any kind of GPS traces.

3 Methodology
3.1 Analysis of GPS traces
The first step of the methodology requests the temporal matching of the two datasets. Since traffic flow information in Dataset1 are on an hourly basis, GPS positions are consistently grouped according to the hour of their registration, even if this induced the loss of those traces spanning over two hourly intervals. Additionally, the analysis was restricted to the five working days in the week and GPS traces were available only for 23 out of 28 vehicles for these days.

The following point aims at spatially joining the vehicle GPS positions and the network arcs. This point is rather tricky since it implies the analysis of a large dataset (Dataset2). The goal is to assign those GPS positions to the different arcs of the network (Dataset1), checking that the vehicle driving direction is coherent with the arc direction. So, it is necessary to focus on a single vehicle (out of the 23 in the dataset) travelling around the city in a specific working day (out of the monthly observation period). At the same time, information on the various arcs of the city network from Dataset1 is considered, namely the latitude and longitude of the nodes at their extremes. Once a given arc has been selected, it is necessary to check whether the chosen vehicle has passed along that road in that day and in that hour range. This task is complicated by the fact that the precision of GPS recordings is limited and the vehicle trajectories can also transversally fluctuate of several meters, especially in multi lane streets. To cope with this, a round buffer with a radius that was tentatively set to 18 m is created around each of the two nodes identifying the selected arc and a check is done to verify if at least a GPS position is contained in these two areas. Mapping the vehicles at the nodes rather than along the arcs increases the chance to detect them for low sampling rates, since at intersections they spent usually more time,

in particular if the traffic signal is red, and because the positioning accuracy based on the satellites line of sight is generally higher.

In this stage of the analysis, it is fundamental to check whether the vehicle has really moved along the arc in its travel between the two nodes at the extremes. In fact, it could happen that a vehicle has been registered at node A and then at node B, but following a different route that does not correspond to the selected arc. An additional check is therefore made by comparing the course of each GPS recording, i.e. the driving direction of the vehicle, with the direction of the arc, i.e. the bearing from node A to node B. More precisely, if the root mean square error between all courses and the bearing is less than 50 degrees, then the vehicle is assumed travelling from A to B through the arc under consideration and without deviations. This threshold has been selected in order to avoid losing those course measures that are different from the arc bearing value simply because there is a bend along the road. These last steps of the approach, namely the association of the GPS recordings to the nodes and the investigation of the driving directions of the vehicles, allow avoiding the implementation of a more complex procedure that would require the association of all the traces to the different arcs, rather than to the nodes, of the network [15, 16].

Once GPS traces are assigned to arcs and it has been verified that the vehicle is travelling along the arc itself, it is necessary to estimate the related travel time. This value, called T_GPS in the following, is computed by selecting the last recording in the boundary around the origin node and the first recording registered in the boundary around the end node, and then computing the difference between the two corresponding timestamps. However, T_GPS also includes all intermediate stops between the two nodes. These stops could be either due to traffic conditions (congestion, traffic lights or yielding) or to service stops e.g. for deliveries. It is clearly important to distinguish between the two, in order to have a correct estimation of the time lost in congestion.

In this research, stops whose duration is shorter than 120 sec are considered as due to traffic conditions, since this is a typical maximum duration of a stop for yielding or due to the red phase of a traffic light, whereas service stop are normally longer than that. Clearly such hard threshold might lead to wrong classifications, since some service stops where the vehicle is parked in close proximity of the delivery point and one small package has to be delivered could take less than 2 min. On the other hand, vehicles might come to a complete stop due to congestion for more than 2 min. However this threshold seemed the best compromise to minimise such mis-classifications and was also adopted in previous research [21]. More specifically, the overall time interval for a
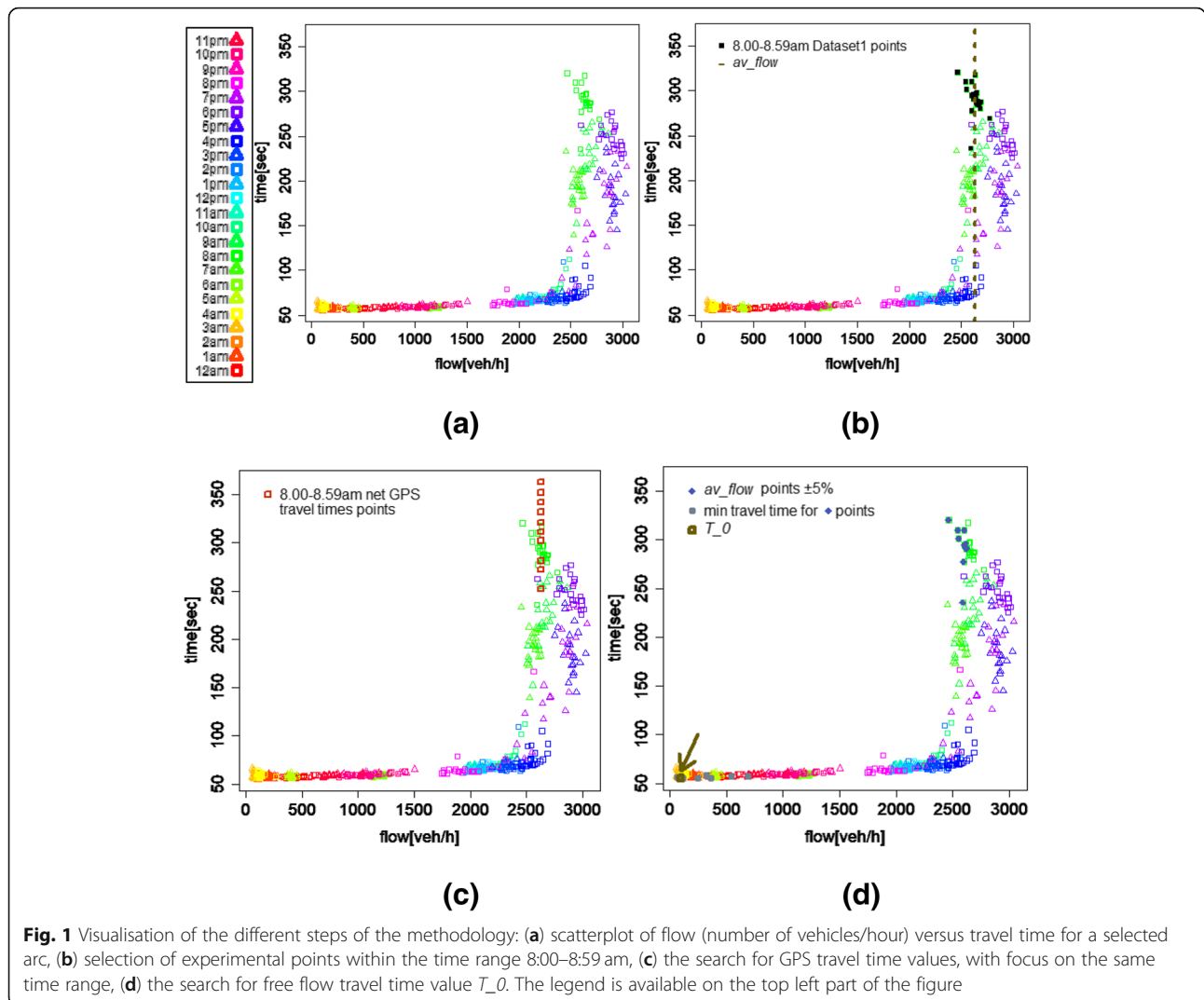
series of subsequent 0 speed recordings is calculated: if it is larger than 2 min, it is considered as a service stop.

The sum of the durations of all service stops for a given arc that are identified through such threshold is called T_GPS_ss. The final step is the computation of the net travel time, i.e. the travel time not considering service stops, given by the difference between T_GPS and T_GPS_ss. This analysis is done for all arcs in the network and it is repeated for each vehicle on every arc.

3.2 Combining the two datasets

As said previously, Dataset1 collects the flows and the travel times along the arcs on the Turin road network referring to different hour ranges and days of May 2017. Thanks to such information, it is possible to empirically evaluate the directional free-flow travel time on all arcs. For each arc, the relationship between the travel times in a certain hour range and the corresponding flow is visualized through a scatterplot of all the data available for the month of May 2017. Fig. 1(a) reports an example of such plot. Different colours and shapes represent data recorded in different hour ranges. As it can be seen, each hour range is represented by several points, since all the values registered for the different days of the month are represented.

This plot provides interesting information on the traffic trends characterizing such road. On the left side of the plot, it is possible to notice the hours of the day when a low number of vehicles are passing along this street (low values in x-axis), which corresponds to the lowest travel time values on the y-axis. For example, traffic flows between 4:00 am and 4:59 am are about 100 vehicles/hours while the corresponding travel time is around 60 sec (these points are marked as yellow squares in the scatterplot). This sounds reasonable since this arc is approximately 1 km of length and the maximum speed allowed is 50 km/h. On the other hand, this road is collecting all the traffic entering the city from the south and coming from a motorway, therefore prone to congestion in peak hours. The top right side of the graphic in Fig. 1(a) shows congested traffic conditions, characterized by high values of flows (more than 2000 veh/h) and longer travel times (more than 3 min). The shape of the interpolating function linking travel times and flows from Fig. 1(a) is clearly not linear and it could be represented through well know relationships such as the BPR (Bureau of Public Roads) formula that are customarily used in traffic assignment models [27]. However in the following we are not assuming any functional relationship between travel times and flows, which on the other hand perform quite poorly especially in urban settings. Pertinent measures such as saturation flow and free-flow

Fig. 1 Visualisation of the different steps of the methodology: (**a**) scatterplot of flow (number of vehicles/hour) versus travel time for a selected arc, (**b**) selection of experimental points within the time range 8:00–8:59 am, (**c**) the search for GPS travel time values, with focus on the same time range, (**d**) the search for free flow travel time value *T_0*. The legend is available on the top left part of the figure

travel times will be empirically derived from the GPS measures, as detailed in the following paragraphs.

Starting from the scatterplot in Fig. 1(a), those experimental points corresponding to the hour range in which GPS traces were recorded on the arc under consideration are selected. As an illustrative example, let us assume that one vehicle travelled through that arc between 8:00 am and 8:59 am. All points representing traffic flow conditions in the same time interval from Dataset1 are then represented through black squares in the plot of Fig. 1(b). Then, the average flow value *av_flow* of those selected points is computed and it is represented by a dashed bronze line in the same figure.

This average value is assumed to be the traffic flow that was experienced by the 23 vehicles that were monitored whenever they eventually travelled through the arc under consideration in that specific hour range, irrespective of the travel day and of the specific

vehicle. Assuming that this happened 10 times in the 1-month observation period, the corresponding "net GPS travel time values" (that were introduced in the previous subsection as the difference between *T_GPS* − *T_GPS_ss*) are retrieved. Fig. 1(c) then represents again the scatterplot with the addition of 10 experimental points from GPS traces (brown squares), having the same x-axis value (namely, *av_flow*) and the net GPS travel time values as y-axis value. It is interesting to notice that these net GPS travel time values are fairly consistent with those coming from Dataset1. In fact, y-axis values of brown squares in Fig. 1(c) have a mean value (310 sec) that is similar to the one found for y-axis values of black squares in Fig. 1(b) (291 sec), i.e. the one referring to the same hour range. As expected, net GPS travel times in Fig. 1(c) show a larger variation around the mean, since travel time values represented by the black squares in

Fig. 1(b) are averages over the entire flow for a specific time range and day.

The final step aims at estimating the free flow travel time value for the arc under consideration. Hence, all the flow measures from Dataset1 for the arc under consideration within a ± 5% interval around the *av_flow* value are selected (blue diamonds in Fig. 1(d)). Then, the corresponding days in which these measures were taken are selected (grey circles in Fig. 1(d)) and the minimum travel time among all those registered in those days, irrespective of the hour range, is computed. The latter is considered as the free flow travel time, namely *T_0* in the following (bronze square on the bottom left of Fig. 1(d), marked by the arrow), for that specific arc. The minimum observed travel time is not directly taken as the free-flow measure to avoid outliers. For example, if exogenous factors such as road works worsened traffic conditions for the largest part of the 1 month observation period, the resulting increase of travel times should not be considered as an effect of congestion. The free flow travel time is then computed for all arcs that where travelled by the fleet of vehicles.

3.3 Congestion KPI definition

The Key Performance Indicator related to traffic conditions stems from the previously defined quantities and it is consistent with the above reviewed European projects. The analysis can be done at the maximum level of disaggregation, i.e. for every single vehicle k travelling through a given arc j. The time lost in congestion can be defined as follows:

$$KPI_{j,k} = T\_0_j - \left(T\_GPS_{j,k} - T\_GPS\_ss_{j,k} \right)$$

where T_0_j is the free flow travel time of arc j, $T\_GPS_{j,k}$ is the total travel time of vehicle k moving along arc j and $T\_GPS\_ss_{j,k}$ is the service stopping time of vehicle k along arc j.

However, if the goal is to compare the results obtained for arcs with different length, it is necessary to have a common scale and proportionality in the final value. With this aim, the previous indicator can also be computed in relative rather than absolute terms as follows:

$$RKPI_{j,k} = \frac{T\_0_j - \left(T\_GPS_{j,k} - T\_GPS\_ss_{j,k} \right)}{T\_0_j}$$

Negative values of those indicators respectively represent absolute or relative measures of the time potentially lost in congestion when vehicle k travelled along arc j. Zero values indicate that the vehicle travelled at free flow speed. Positive values have also been found, since the reference value T_0_j is an average and it might well

be possible that vehicles are travelling faster than that, especially considering experienced drivers delivering goods in the city.

The level of detail considered in computing this KPI allows the decision maker to aggregate the results at different scales, according to the specific transport policy questions that need to be answered. Hence, bearing this in mind, it is interesting to highlight some ways to aggregate and visualize the results. Examples discussed in the following include a link-based aggregation to identify critical road segments, a zonal-based aggregation to check those areas with highest congestion problems, a vehicle types-based aggregation to check whether goods or passenger flows are most penalized etc. Both absolute and relative values of the KPI can be used in this aggregation step, according to the specific objectives.

4 Visualisation of the results and discussion

In the following some methods to extract and visualize results from the previously defined KPI are discussed.

A first example of arc-based aggregation is proposed in Fig. 2, where the minimum value of RKPI between all vehicles travelling over arc j, i.e. the maximum time loss due to congestion, is represented for different hourly intervals. Please note that those maps are different from the usual congestion maps that represent general traffic conditions for all vehicles, since they rather highlight the most problematic parts of the road network from the viewpoint of the considered logistic fleet.

Since the arcs are directional, in the following figures they are represented by arrows. Their different colours aim at depicting the different levels of criticality connected with the travelling along each arc: grey arrows indicate a travel time under uncongested traffic flow conditions, while a colour ranging from yellow to purple indicates an increasing travel time considering the worst case among the available GPS measures for that link. It can be noted that very small KPI values in Fig. 2 are normally reported only for very short arcs, for which both the boundary effects of what happens at nodes and other issues related to the lack of precision have a much stronger influence, given the fact that the proposed KPI is a relative rather than an absolute measure.

The maps proposed in Fig. 2 show a selection of time ranges of the day, namely two in the morning (8:00–8:59 am, 9:00–9:59 am) and two in the afternoon (5:00–5:59 pm, 6:00–6:59 pm), which represent two peaks given the considerable number of GPS traces found in the dataset. These maps have been selected to point out the changing in the direction in which the arcs were travelled. For example, Fig. 2(a) and Fig. 2(b) shows that the vehicles are moving mainly towards the city centre in the morning, mainly from 8:00 to 9:59 am, since they are going to deliver to the shops and they can do it only
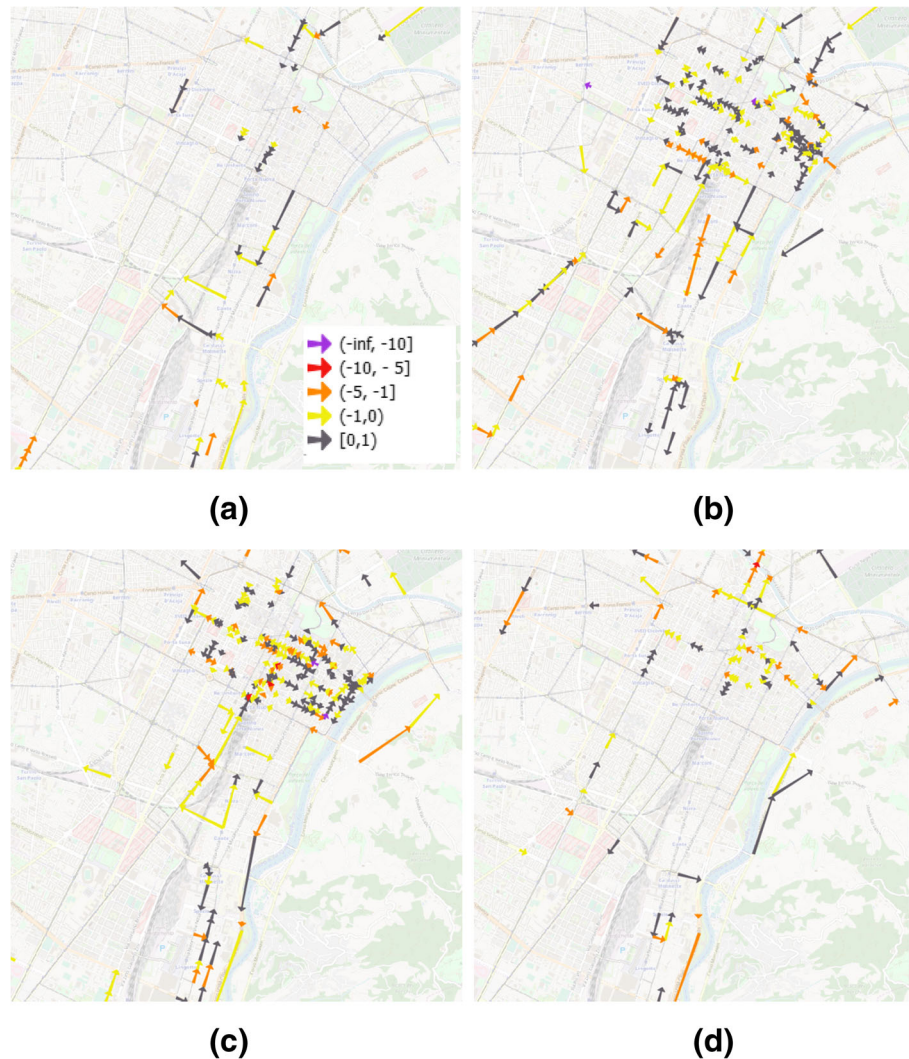
Fig. 2 Values of the worst $RKPI_{j,k}$ for the arcs of the city newtork in the time ranges 8:00–8:59 am (**a**), 9:00–9:59 am (**b**), 5:00–5:59 pm (**c**), 6:00–6:59 pm (**d**). Different colours of the arrows represent different values of the indicator according to the scale in figure (**a**)

when shops are open. A look at Fig. 2(c) and Fig. 2(d), instead, shows that when the logistic fleet is usually leaving the city during the late afternoon it is clashing with the afternoon peak hour. Thanks to these maps, it is possible to see where the vehicles lose time due to congestion, with the most critical arcs identified in red and purple. Not surprisingly, many arcs with the worst relative *KPI* value are very short and entering saturated street crossings, where the time lost for yielding and red traffic lights is considerable compared to the free-flow travel time.

Additional insights can be gained at a higher level of aggregation, namely considering the overall travel time trends on each arc, irrespective of the day and of the time range. To this effect, the following average indicator is computed for each arc *j*:

$$AKPI_j = \frac{\sum_{m=1}^{n} [T\_0_j - (T\_GPS_{j,m} - T\_GPS\_ss_{j,m})]}{n} \frac{1}{T\_0_j}$$

where all the differences between the free flow travel time T_0 and the net GPS travel time $(T\_GPS - T\_GPS\_ss)$ for all days and for all vehicles are summed and divided by the number of the measures (n) to obtain an average value. As done previously, the final indicator is then derived by dividing by the free flow travel time T_0 to allow a comparison among arcs of different length. Consistently with the visualization proposed in Fig. 2, the final results for the Turin road network can be observed in Fig. 3. Here, a zoom on the arcs in city centre is presented to show the most critical arcs in terms of relative
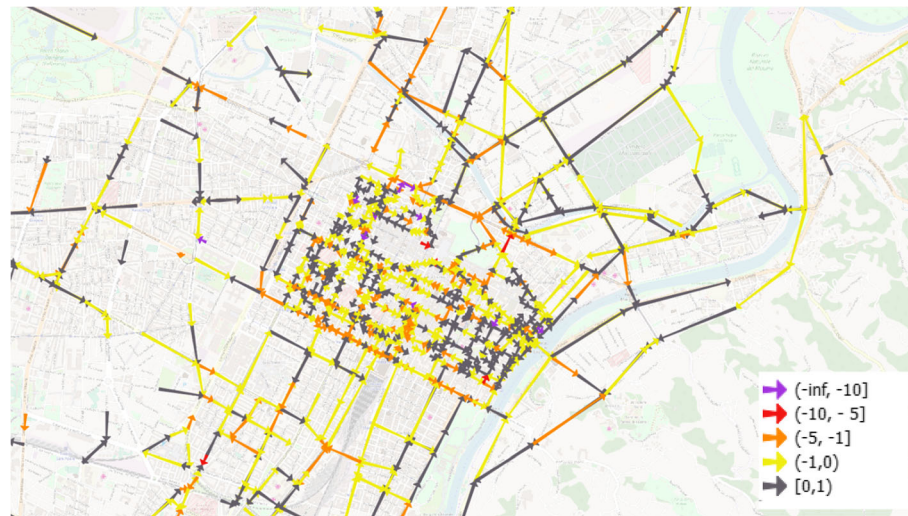
Fig. 3 Value of the mean indicator *AKPI_j* computed for all the arcs of the road network (zoom in the city center)

mean travel time lost in congestion by the fleet. As before, these are identified with purple, red and orange arrows. The observation of the map in Fig. 3 shows that not all the roads of the city network are represented. This limitation is due to the datasets analysed, since Dataset1 does not include all the streets of Turin while Dataset2 refers to a limited number of vehicles travelling around, which therefore are not likely to have travelled through all roads in the network. It could also happen that a vehicle is not properly assigned to a link if it is not detected at its extremes, due to a missing position in the buffer around that node. The ratio of the buffer has been selected as an average value that could assure to find all the possible passages of vehicles in all the crossings around the city that are, obviously, of different dimensions. A further refinement of the methodology could lead to identify different buffers amplitude according to the various kind of road crossing.

The analysis presented so far was mainly based on the collection of the "feedbacks" given by the whole logistic fleet travelling on the arcs composing the network. However, a further challenging point is the focus on the time loss affecting the path of a selection of those vehicles. In more details, it could be interesting to select those vehicles which are characterized by a low value of KPI on a considerable number of arcs, irrespective of the day and of the time range. This analysis is meaningful because many of the vehicles in the considered fleet travel a very similar delivery route every day within the city. An illustrative example with two vehicles of the fleet is here discussed. Fig. 4 (a) and (b) show all arcs travelled by vehicle 32 and 16, respectively, in May 2017. It is interesting to notice how these two vehicles are

delivering in different areas of Turin: both travel in the city centre but the first one leaves the city mainly from the northern part, while the latter from the southern. Thanks to these maps, it is clearly possible to identify those arcs (roads) where they mostly lost their time due to congestion. The most critical arcs are represented, as previously, by orange and red arrows.

5 Conclusions and future work

This paper has focused on how to measure the way in which congestion selectively affects a specific traffic stream through the creation of a congestion KPI defined after the integration of different data sources, namely vehicle-based GPS traces and road-based information of traffic flows. Preliminary examples of how the aggregation of the KPI at different scales can provide insights into the transport system are presented. Such methodology could help in potentially informing a wide range of policy actions, identifying, for example, the most critical arcs for given travel purposes (parcel services, commuting), the most congested areas in relation with specific user groups (if related metadata are associated with GPS traces) or the most congested lines in a public transport network. All these insights could be relevant for different stakeholders at level of city administrations, transport services operators and specific social groups. On a methodological viewpoint, future work could refine the KPI computation through both more sophisticated processes to identify service stops along the arc and the consideration of all GPS points, beyond those near the arc extremities, for a more precise representation of the vehicle trajectories.

Among the different actions to be developed thanks to the SUITS project activities, an important task is devoted
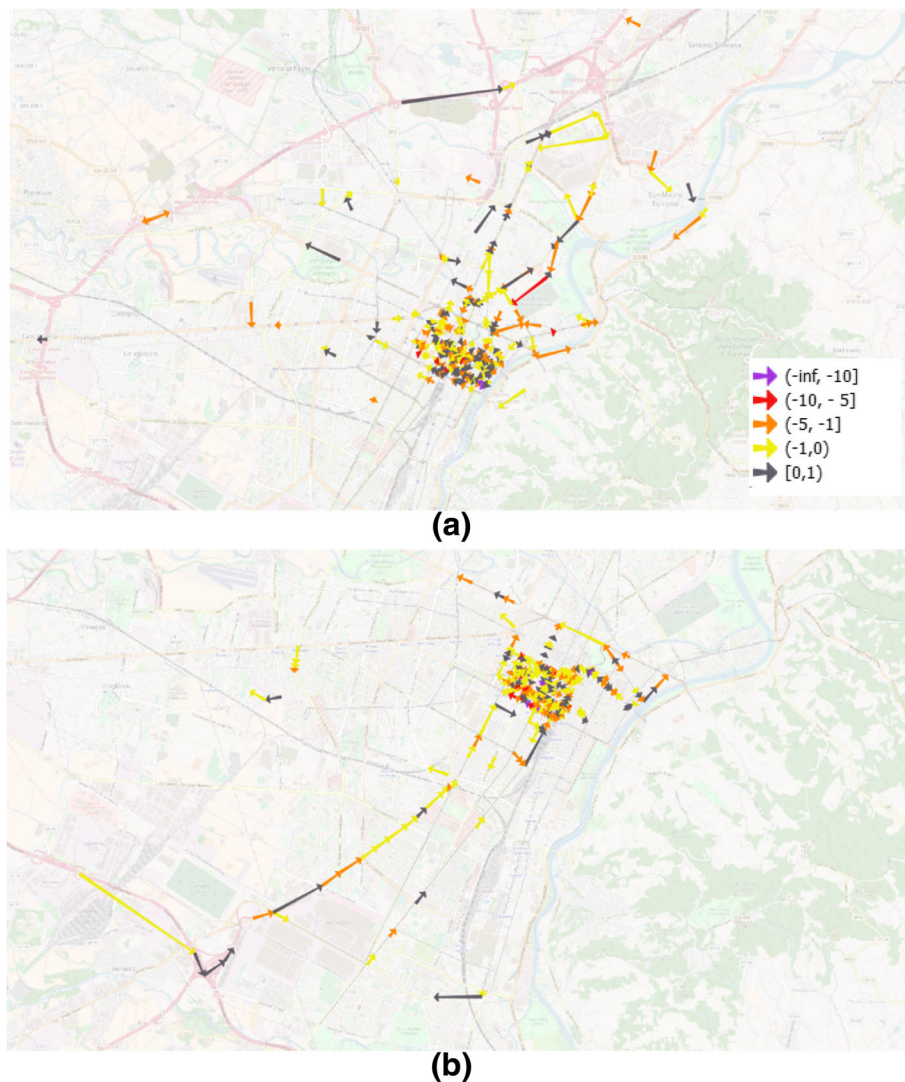
Fig. 4 Value of indicator for a selection of vehicle, namely vehicle 32 (**a**) and vehicle 16 (**b**)

to contribute to capacity building of urban planners and stakeholders through a better understanding of data collection, analysis and knowledge discovery methods to identify opportunities for improvement in urban transport efficiency and environmental impact. The methodology proposed in the current work could be exploited in the domain of the so-called Sustainable Urban Logistics Plan - SULP [28] and in the framework of commercial development and city logistic plan [29], providing useful information on the efficiency of the road network as perceived by a selected kind of users (freight deliveries). For instance, this KPI could be monitored in a certain range of time when specific actions are proposed at city level in order to evaluate the effectiveness of policies such as restricted access at roads or areas to certain kinds of traffic.

The data analysed focus on the tours of a selected number of vans (express courier) during a month of registration. This first round of results provided interesting feedbacks on the way the city congestion could affect urban freight deliveries and these findings could be refined adding the GPS recordings collected in other months and from other vehicles. Moreover, future works would request to further develop the method through the use of alternative datasets to meet as much as possible, for example, the local conditions of a typical mid-sized European city concerning the availability of data. The recording of GPS positions of both people and vehicles is nowadays more and more widespread and, in some cases, the related datasets are openly available. So, many works focus on the exploitation of GPS traces to study urban mobility analysing data from taxis [30], buses [31] or trucks [12]. The present paper contributes to this research field by proposing a rather flexible methodology that could work also with different datasets to

understand how congestion has a different impact on selected road users. Moreover, it is certainly relevant to explore the potentialities of this analysis method by finding other ways to aggregate indicators (e.g. by individual vehicle in a given day, to identify the worst arcs in terms of overall time lost, to identify most penalized vehicle according to the path it follows, to quantify daily or seasonal variations). Concerning the latter point, it is also possible to aggregate the congestion KPIs at the whole network level, leading to a Network Fundamental Diagram (NFD), or Macroscopic Fundamental Diagram (MFD) indicator, whose usefulness to manage congestion has already been proven in a number of applications (e.g. [32, 33]).

Abbreviations
BPR: Bureau of Public Roads; FCD: Floating Car Data; GDP: Gross Domestic Product; GPS: Global Positioning System; KPI: Key Performance Indicator; MDF: Macroscopic Fundamental Diagram; NDF: Network Fundamental Diagram

Authors' contributions
Study conception and design: M. Pirra, M. Diana; analysis and interpretation of results: M. Pirra, M. Diana; draft manuscript preparation: M. Pirra, M. Diana. All authors reviewed the results and approved the final version of the manuscript.

Availability of data and materials
The data that support the findings of this study are available from Turin Municipality through 5T srl but restrictions apply to the availability of these data, which were used under license for the current study, and so they are not publicly available. Data are however available from the authors upon reasonable request and with permission of Turin Municipality through 5T srl.

Competing interests
The authors declare that they have no competing interests.

References
1. DG MOVE (2014) *Update of the Handbook on External Costs of Transport*. https://ec.europa.eu/transport/sites/transport/files/handbook_on_external_costs_of_transport_2014_0.pdf. Accessed 28 June 2018.
2. European Environmental Agency (2013) *Understanding pollutant emissions from Europe's cities: Highlights from the Eu Air Implementation Pilot Project*. https://www.eea.europa.eu/publications/understanding-pollutants-emissions-from-europes/at_download/file. Accessed 10 July 2018.
3. Antoniou, C., Balakrishna, R., & Koutsopoulos, H. N. (2011). A synthesis of emerging data collection technologies and their impact on traffic management applications. *Eur Transp Res Rev, 3*, 139–148. https://doi.org/10.1007/s12544-011-0058-1.
4. Barceló, J. (2010). *Fundamentals of traffic simulation*. New York, USA: Springer-Verlag.
5. Toole, J. L., Colak, S., Sturt, B., et al. (2015). The path most traveled: Travel demand estimation using big data resources. *Transp Res Part C Emerg Technol, 58*, 162–177. https://doi.org/10.1016/j.trc.2015.04.022.
6. Altintasi, O., Tuydes-Yaman, H., & Tuncay, K. (2017). Detection of urban traffic patterns from Floating Car Data (FCD). *Transp Res Procedia, 22*, 382–391. https://doi.org/10.1016/j.trpro.2017.03.057.
7. Kong, X., Xu, Z., Shen, G., et al. (2016). Urban traffic congestion estimation and prediction based on floating car trajectory data. *Futur Gener Comput Syst, 61*, 97–107. https://doi.org/10.1016/j.future.2015.11.013.
8. Xu, L., Yue, Y., & Li, Q. (2013). Identifying urban traffic congestion pattern from historical Floating Car Data. *Procedia Soc Behav Sci, 96*, 2084–2095. https://doi.org/10.1016/j.sbspro.2013.08.235.
9. Sunderrajan, A., Viswanathan, V., Cai, W., & Knoll, A. (2016). Traffic state estimation using Floating Car Data. *Procedia Comput Sci, 80*, 2008–2018. https://doi.org/10.1016/j.procs.2016.05.521.
10. Fabrizi, V., & Ragona, R. (2014). A pattern matching approach to speed forecasting of traffic networks. *Eur Transp Res Rev, 6*, 333–342. https://doi.org/10.1007/s12544-014-0133-5.
11. Yang, C., Ma, X., & Daly, A. (2016). Demonstration of intelligent transport applications using freight transport GPS data. In *Transportation Research Board 95th Annual Meeting*. Washington D.C., USA. Retrieved from https://trid.trb.org/view/1394553.
12. Hess, S., Quddus, M., Rieser-Schüssler, N., & Daly, A. (2015). Developing advanced route choice models for heavy goods vehicles using GPS data. *Transp Res Part E Logist Transp Rev, 77*, 29–44. https://doi.org/10.1016/j.tre.2015.01.010.
13. Zhang, Y. C., Zuo, X. Q., Zhang, L. T., & Chen, Z. T. (2011). Traffic congestion detection based on GPS Floating-Car Data. *Procedia Eng, 15*, 5541–5546. https://doi.org/10.1016/j.proeng.2011.08.1028.
14. Bauer, D., & Tulic, M. (2018). Travel time predictions : Should one model speeds or travel times? *Eur Transp Res Rev, 10*, 1–12. https://doi.org/10.1186/s12544-018-0315-7.
15. Schuessler, N., & Axhausen, K. W. (2009). *Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT)*. https://pdfs.semanticscholar.org/7048/b08f9cc7216a371e3949a777d31c88d6a2fe.pdf. Accessed 5 Nov 2018.
16. Quddus, M., & Washington, S. (2015). Shortest path and vehicle trajectory aided map-matching for low frequency GPS data. *Transp Res Part C Emerg Technol*. https://doi.org/10.1016/j.trc.2015.02.017.
17. Marsden, G., Kelly, C., Snell, C., & Forrester, J. (2005). *Sustainable transport indicators: selection and use. Deliverable C1 of DISTILLATE project*, p 48. http://www.distillate.ac.uk/outputs/Deliverable%20C1%20Indicators%20specification%20v9.pdf. Accessed 31 July 2018.
18. Kaparias, I. & Bell, M.G.H. (2011). *Key Performance Indicators for traffic management and Intelligent Transport Systems. Deliverable 3.5 of CONDUITS project*, p 75. https://pdfs.semanticscholar.org/5453/8f8cf18c169d139390d09309adf4fdf8b4f7.pdf. Accessed 31 July 2018.
19. Arneodo, F., Gagliardi, G., & Botta, D. (2012). Traffic supervisor system on wide regional area. In *Proceedings of the 19th ITS World Congress*. Vienna.
20. Liao, C.-F. (2014). Generating reliable freight performance measures with truck GPS data. *Transp Res Rec J Transp Res Board, 2410*, 21–30. https://doi.org/10.3141/2410-03.
21. Greaves, S., & Figliozzi, M. (2008). Collecting commercial vehicle tour data with passive global positioning system technology: Issues and potential applications. *Transp Res Rec J Transp Res Board, 2049*, 1–18. https://doi.org/10.3141/2049-19.
22. Ben-Akiva, M. E., Toledo, T., Santos, J., et al. (2016). Freight data collection using GPS and web-based surveys: Insights from US truck drivers' survey and perspectives for urban freight. *Case Stud Transp Policy, 4*, 38–44. https://doi.org/10.1016/j.cstp.2015.11.005.
23. Yang, X., Sun, Z., Ban, X., & Holguín-Veras, J. (2014). Urban freight delivery stop identification with GPS data. *Transp Res Rec J Transp Res Board, 2411*, 55–61. https://doi.org/10.3141/2411-07.
24. Li, J., van Zuylen, H., & Wei, G. (2014). Diagnosing and interpolating loop detector data errors with probe vehicle data. *Transp Res Rec J Transp Res Board, 2423*, 61–67. https://doi.org/10.3141/2423-08.
25. Mazaré, P., Bayen, A. M., & Hall, S. D. (2012). Trade-offs between inductive loops and GPS probe vehicles for travel time estimation: A Mobile Century case study. In *Transportation Research Board 91st Annual Meeting*. Washington, D.C. Retrieved from https://trid.trb.org/view/1129929.

26. Qiu, T., Lu, X.-Y., Chow, A., & Shladover, S. (2010). Estimation of freeway traffic density with loop detector and probe vehicle data. *Transp Res Rec J Transp Res Board, 2178*, 21–29. https://doi.org/10.3141/2178-03.

27. Bureau of Public Roads (1964) Traffic Assignment Manual. Washington, D.C. Cited in Ortuzar, J. de D. and Willumsen, L. G. (2001). *Modelling Transport. (4th ed.) New York, John Wiley & Sons, Inc.* (p. 353). New York. p. 353

28. Ambrosino, G., Pettinelli, I., Freitas, C., & Sousa, C. (2015). *Sustainable Urban Logistics Plan (SULP) methodology for Small and Mid-sized European Towns: the IEE ENCLOSE project results.* https://pdfs.semanticscholar.org/6f6e/85e3 936f59eee580039fb97ac9c86bcba765.pdf?_ga=2.106007043.1683204 880.1565778895-1141325183.1557225211. Accessed 13 May 2019.

29. Cirianni, F., Panuccio, P., & Rindone, C. (2013). A comparison of urban planning systems between the UK and Italy: Commercial development and city logistic plan. *WIT Trans Built Environ, 130*, 785–797. https://doi.org/10.24 95/UT130631.

30. Lacombe, A. & Morency, C., (2016) Modeling taxi trip generation using GPS data: the Montreal case. In *Transportation Research Board 95th Annual Meeting, Washington, D.C., USA.* Retrieved from https://trid.trb.org/view/13 93633

31. Sakhare, R., Mathew, J. K., Avr, A., et al. (2018). Comparison of Bluetooth and Bus GPS Data for Estimating Arterial Travel Time and Trip Chaining. In *Transportation Research Board 97th Annual Meeting.* Washington, D.C. Retrieved from https://trid.trb.org/view/1495014.

32. Keyvan-Ekbatani, M., Papageorgiou, M., & Papamichail, I. (2013). Urban congestion gating control based on reduced operational network fundamental diagrams. *Transp Res Part C Emerg Technol, 33*, 74–87. https://doi.org/10.1016/j.trc.2013.04.010.

33. Alonso, B., Pòrtilla, Á. I., Musolino, G., et al. (2017). Network Fundamental Diagram (NFD) and traffic signal control: First empirical evidences from the city of Santander. *Transp Res Procedia, 27*, 27–34. https://doi.org/10.1016/j.trpro.2017.12.112.

Publisher's Note