USING GPS DATA FROM A SAMPLE OF PRIVATE CARS FOR MODELLING THE URBAN TRAFFIC

Sergio Mitrovich, ENEA – sergio.mitrovich@enea.it

Giuseppe Montesano, ENEA – giuseppe.montesano@enea.it

Francesco Pugliese, ENEA – francesco.pugliese@enea.it

Massimo Mancini, ENEA – massimo.mancini@enea.it

Gaetano Valenti, ENEA – gaetano.valenti@enea.it

ABSTRACT

This paper is focused on the development and use of the analytical and statistical methods needed for transforming raw GPS data, recorded from a large sample of private vehicles, into a source of information suitable as input for traffic models or for noise and air pollutant emission models, as well as for short-term travel time prediction models, and in general to improve the understanding of traffic and travel patterns existing in a city and in the wider surrounding region.
The first part of the paper describes the main steps of the data treatment process and the algorithms which can be used for different applications and in different operating conditions (i.e. sampling period, density of equipped vehicles, polling time intervals, etc.).
The second part of the paper reports about the results of a trial application: a full week test, related to Florence (a medium-large Italian city) province and surrounding urban area, is examined and discussed to demonstrate the practical potential of this new approach in understanding and modelling urban traffic performances.
Keywords: Floating Car Data, Traffic Analysis, Traffic Management

1. INTRODUCTION AND BACKGROUND

In the last years an increasing number of private cars have been using GPS based devices, like smart navigators or mobile communication systems, and moreover many car owners have installed on board "clear boxes" or other tracking systems required by insurance companies for incident monitoring and for e-safety and e-security services. Most of these devices are connected by means of telecommunication systems (GSM, GPRS, UMTS, etc.) to central control units, which provide to temporary store the data received. A large amount of row GPS data, recording the instant location and speed of such equipped vehicles, are now currently available and, besides the specific purposes for which such data are collected, they could be used for a number of other applications.

In fact, if such data are collected, map-matched, in case fused with other data, and treated by means of appropriate statistical techniques, they could be a reliable, powerful and cost-effective way to acquire accurate information about traffic from a wide-area road network.

This source of information can be used for real-time applications, like short-term forecast of road link travel times (very useful for smart navigators), but also for many other off-line tasks related to traffic and road network management. For instance, GPS data from a sample of private cars can be used to improve the knowledge of travel conditions and driver's behaviour, to estimate the O-D matrices needed in traffic simulation models, and to assess the impact of new transport policies and measures in term of energy consumption and pollutant emissions.

Nowadays the use of real-time Floating-Car Data (FCD), based on traces of GPS positions, is emerging as a reliable and cost-effective way to gather accurate traffic data for a wide-area road network. Unlike other traffic data collection techniques (Automated Vehicle Identification systems, video cameras, inductive loops, radar based sensors, etc.), floating cars act as moving sensors travelling in a traffic stream and do not require instrumentation to be set up on the roadway.

While increasing the understanding of individual travel behaviour, floating car technique can easily provide near real-time information on any part of large road networks and offers a viable way to complement fixed-point traffic sensors, such as cameras and loop detectors, involving high installation and maintenance cost.

The FCD technique is based on the exchange of information between a fleet of floating cars travelling on a road network and a central unit for storing data. The floating cars periodically send the recent accumulated data on their positions (latitude, longitude and altitude) and, optionally, instantaneous speed, whereas the central data system tracks the received floating car data along the travelled routes by matching the related trajectories data to the road network. The frequency of sending and reporting is usually determined by the required data resolution and the communication system between vehicles and the central unit.

The reliability of travel time estimates based on GPS data highly depends on the percentage of monitored floating cars participating in the traffic flow [1][2][3]; other factors affecting the reliability of travel time estimates, mainly for lower penetration of floating cars, are traffic conditions and road link capacities.

Different approaches have been proposed to use GPS data received from probe vehicles to predict short-term travel conditions, to detect incident or critical situations [7][8][9] and, finally, to determine Origin-Destination traffic flow patterns [10].However this last target is somewhat different from the others. In this case real-time data are needed only to monitor and detect random changes from a basic pattern, while to determine the basic O-D matrix we can use historical data. Therefore, together with the percentage of available floating cars, the polling time and the traffic condition, mostly depending from the day time, another important factor affecting the reliability of estimates is the duration of the observation period, during which data are stored. In other terms, we can say the reliability of estimated O-D matrices is mainly depending from the percentage of floating cars and from the overall amount of recorded GPS tracks.

2. AIM OF THE STUDY AND CONTENT OF THE PAPER

The authors are involved into a wide research and development project in the field of ITS, called Pegasus, co-funded by the Italian government within the research and development programme “Industria 2015”, which is now underway.
The Pegasus project is coordinated by the Octotelematics [11], which is the European company leader for development and deployment of Telematics for Insurance application, with approximately 700.000 On Board Units (OBU) installed and provides complete solutions from in-vehicle On Board Units up to Data Processing Centre for Pay As You Drive, Pay How You Drive, Pay Per Use insurance.
One of the main target of Pegasus is to study and to develop innovative applications based on the use of GPS data, recorded by the on board units (OBU) installed on private owned cars and transmitted to a central control unit, for modelling, monitoring and predicting the urban traffic, as well as for assessing its impact and for supplying suitable and useful information to system users.
Unlike previously proposed studies based on FCD techniques (mostly using data from taxi or bus fleets), the Pegasus project can exploit data from a large number of private cars, delivering historical and/or real-time information about the whole Italian road network and the major Italian metropolitan areas as well. As a matter of fact a large (and still growing) number of private cars (about 650.000 at the end of 2009) is now equipped by Octotelematics with a specific device (manufactured by the Metasystem company) covering a range of insurance-related applications and transmitting wirelessly GPS traces to a central unit.
Information about road traffic volumes, patterns and speed can be therefore estimated from the GPS traces transmitted and stored in the central unit managed by Octotelematics. Average traffic speed estimates, regularly updated at intervals of 3 minutes 24 hours a day, are already delivered to a number of mobility information service providers and to motorway/roadway operators.
This paper is focused on the development and set up of the analytical methods needed for transforming the raw GPS data, recorded from a large sample of private vehicles, into a source of information suitable as input for traffic simulation and prediction models, whose results can be used in turn to feed connected noise and air pollutant emission models, or to improve the understanding of traffic and travel patterns existing in a city and in the wider surrounding region.
In Italy the Origin-Destination matrices of systematic mobility, used as input for modelling and simulating road traffic in urban areas, are mainly derived from ISTAT (the Italian National Statistic system) data, taken during the periodic census, or from specific surveys, which are in general expensive, commissioned from time to time by municipalities. Generally speaking, in the decade between one census and the next there could be strong changes in travel demand pattern. For instance, taking also into account nowadays there is a general tendency towards an urban sprawl and an increase in connections with minor centres, there could be changes in the patterns of commuting from home (generators) to work or to other important locations (attractors), which are complex and difficult to predict. Furthermore, we can also have changes in the size of the catchment area of commuting for the different centres, which is a significant indicator of the role the different centres and the relationship patterns with external areas.

The research activity, which is reported in the present paper, concerns in particular the travel demand by private owned cars. Therefore it is not necessary to use a map-matching algorithm to locate GPS tracks into the road network graph, but we only need to locate GPS tracks into the O- D traffic zones of the road network graph zoning. This was easily made by means of a GIS system.

The objective of this study is to transform information provided by FCD into a daily travel diary for each of the monitored vehicle, as needed to characterize the mobility behaviour patterns (from a statistical point of view), and then to define a basic Origin-Destination matrix of systematic trips for commuting from home to work, to study or to other important destinations.

In order to be used for modelling the private car traffic in the study area, the basic O-D matrix of monitored cars should be also suitable to be extrapolated, in a simple way, to the whole of private cars and to be compared with other existing matrices to study changes and new trends in travel demand..

The first part of the paper outlines the preliminary steps of the data treatment process, the relevant algorithms and the main parameters, which can be used for different applications and in different operating and traffic conditions (i.e. sampling period, density of equipped vehicles, quality of GPS signal, polling time intervals, etc.).

The second part of the paper examines in details the data treatment process for a trial application and reports about the results: a full week test, related to the Florence (a medium-large Italian city) province and surrounding urban area, is examined and discussed to demonstrate the practical potential of this new approach in understanding, monitoring and predicting urban traffic performances.

3. THE SOURCE OF DATA AND THE STUDY AREA

This study uses data from the Octotelematics FCD system. The target of the system is basically insurance profiling, therefore the on board unit (OBU) design and operation is utterly different from other on board devices designed for anti-theft or fleet management services. One of the basic differences consists of the larger amount of data needed for a correct profiling: as a result at the time of writing there are over 20 million new records per day.

The OBU consists of a GPS receiver, a GPRS transmitter, a 3-axis accelerometer sensor, a battery pack, a mass memory, and a processor/RAM. The OBU stores GPS measurements and periodically transmits (on request or automatically) the recent accumulated measurements to the central data system. Transmission occurs every 100 Km travelled or every 12 minutes when the equipped car is running along predefined motorways or crossing city centres.

Each record of the car tracking data set includes the following information:

a) car identification number (ID),

b) Log of date and time in GMT (Greenwich Meridian Time),

c) Longitude and latitude of GPS track,

d) Speed of the vehicle from GPS,

e) course-angle at which the equipped car is travelling with reference to the North,

f) GPS status, indicating the measurement accuracy that depends on how many satellites are visible,

g) engine status ("starting" , "in motion" and "turning off"),

h) distance travelled from the previous record.

The area in which this analysis has been conducted is a rectangle of about 84 x 87 km (see Figure 1), mainly located in Tuscany (Italy), and covering the whole Provinces of Florence, Prato and Pistoia and part of the surrounding provinces of Pisa, Siena, Arezzo, Ravenna, Bologna, Modena e Forlì-Cesena. The area includes 148 small municipalities and a medium-large city: Florence, which has approximately 366.000 inhabitants and covers an area of 141 km2 [12]. Florence is not only the economic and administrative capital of the region Tuscany, but rather is an historical and artistic city well-known in all the world. The first FC data set used by the ENEA group, now working in Pegasus, for preliminary studies, carried out at the end of 2008 and at the begging of 2009 [16][17][18], was related to one day (Tuesday, March 4th 2008). The present study is based on an extended FCD set, covering all the first week of March 2008, of about 3.5 million records. On this date the penetration rate of floating vehicles was estimated by Octotelematics to be at about 1.3 percent. All the travels having an origin and/or a destination within the study area, as well as all the through trips, with both the origin and destination lying outside the study zone, have been considered and examined.

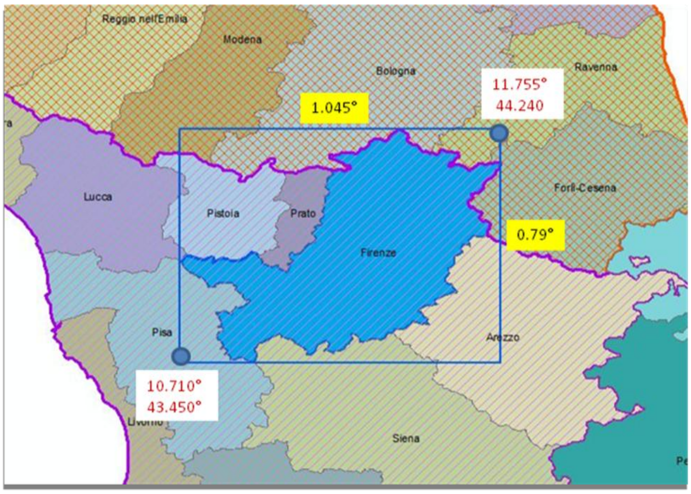

Figure 1 The study area

The study area is shown in figure 1 while the main freeways and highways crossing the study area (Highway A1, Highway 11, Freeway Florence-Pisa-Leghorn and Highway Florence-Siena) are represented in the map of Figure 2. The vehicle flows crossing the study area are estimated to be about 10% of the total number of vehicle circulating within the area.

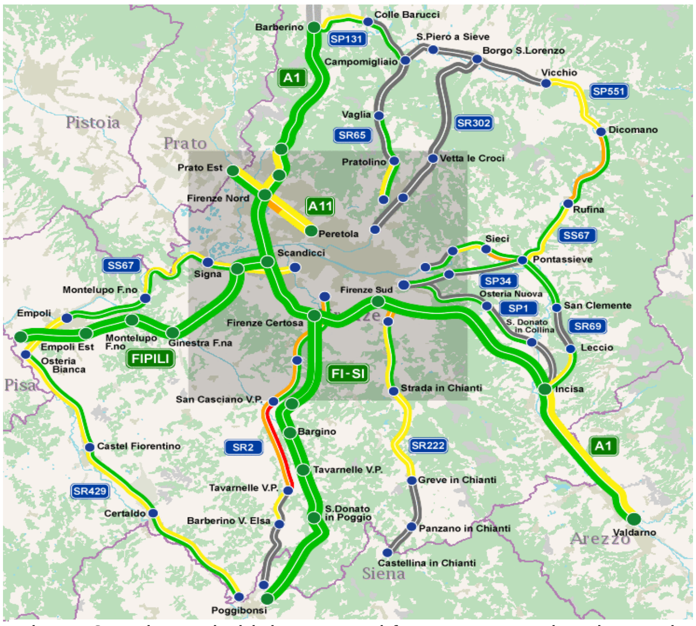

Figure 2 – The main highways and freeways crossing the study area

In Figure 3 the daily GPS positions, stored by all the floating vehicles travelling in the study area, are displayed on a map. The original monthly file (about 2 Giga bytes) including information about more than

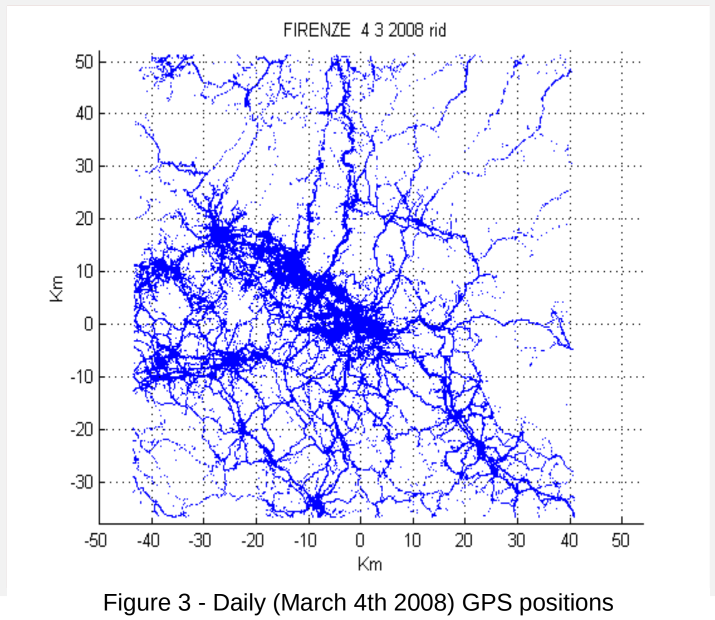

Figure 3 - Daily (March 4th 2008) GPS positions

17 millions GPS tracks was read by means of a suitable software and ordered by date and by vehicle (id-terminal). The analytical methods reported in the present paper were in particular applied to and tested with data records of one week (from March 3rd to 9th 2008). Depending from the location, data are transmitted from the on board unit (OBU) to the central control unit in two different ways: the first mode, used mainly in the urban areas, is based on the travelled distance (about every 2km), while the second one is base on a fixed sampling time (about every 30"). This is clearly shown below in Figure 4 and Figure 5.

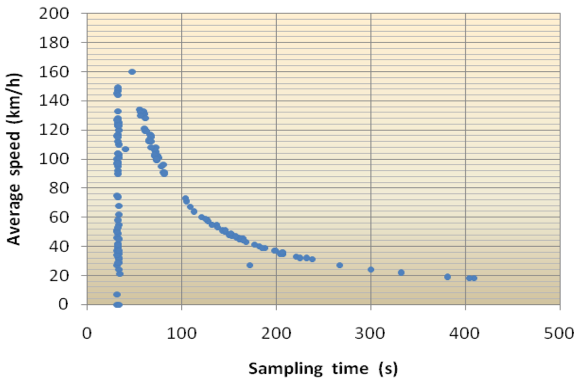

Figure 5 - Average speed vs. sampling time

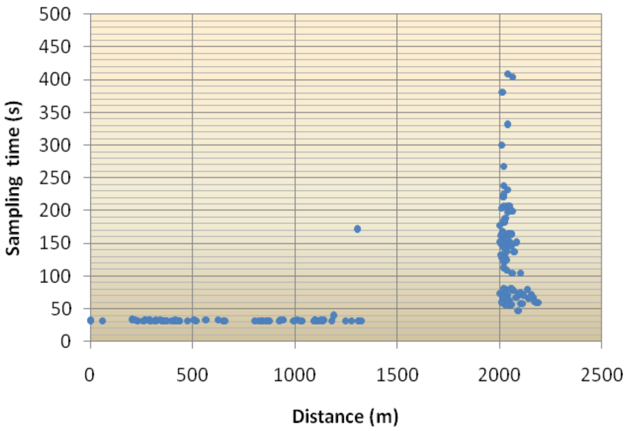

Figure 4 - Sampling time vs. travelled distance between two GPS points

The raw data are affected from a significant number of errors, mainly due to OBU malfunctioning and to transmission failures, and, in addition, some daily travel are interrupted because the vehicle is going temporally outside the study area. This last situation can have a big impact on the estimate of the main analyzed parameters (i.e.: the average speed). Therefore, before to apply the analytical methods, we had to carry out a first step of data pre-processing and cleaning up. This first step has also the aim of subdividing daily data received from each vehicle and ordering and organizing them as needed to facilitate the following steps of data processing.

4. DATA PRE-PROCESSING AND CLEANING

The aim of this first step of data pre-processing and cleaning step is to:

- transform the available coordinates of points (GPS tracks) from the international reference system WGS84 to the National Italian reference system (Gauss Boaga fuso Ovest), which is more suitable to compare data with a number of already available local analysis (i.e.: road network zoning);
- eliminate data coming from malfunctioning terminals and in general all data clearly affected by errors (i.e.: zero length displacements, travelled distance between two GPS points which is shorter than the geometrical distance, speed too high or too low, etc.);
- to rebuild the travel interrupted because the vehicle temporarily (more than 30') went outside the study area;
- store all data into a data base (SQL server 2005) as needed to obtain a daily travel diary for each monitored vehicle and to study and estimate the main parameters of private car transport demand;
- enter geo-referenced data in a GIS for the visualizing travels and other trip information;
- prepare data for feeding a software (i.e.:GPSbabel, Ploticus) for the visualization of single vehicle travels by means of graphs and plots.

For the first task (transforming the GPS point coordinates) we have used a software (TRASPUNTO) available on the Internet, developed by the Italian Ministry of the Environment within the POMA project (Progetto Operativo Multiregionale Ambiente).

The graph reported here below (Figure 6) give an idea of the quality of the sample of data used in the present paper. Pan=0 and Pan=2 are the starting and ending points, while Pan=1 indicate all the other of a trip.

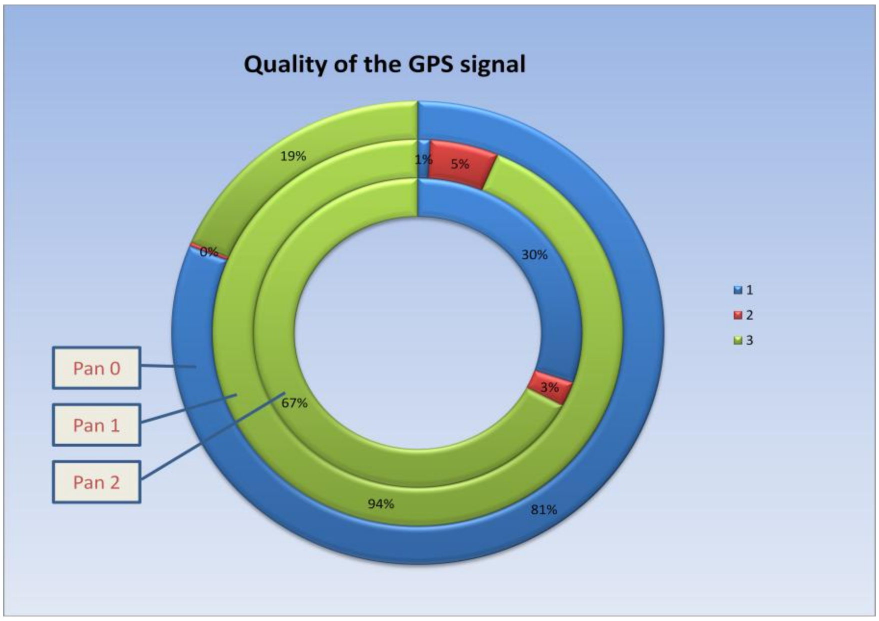

Figure 6 - Quality of the GPS signal for the sample used for this study

The flow of data pre-processing is summarized in the figure reported here below (Figure 7).

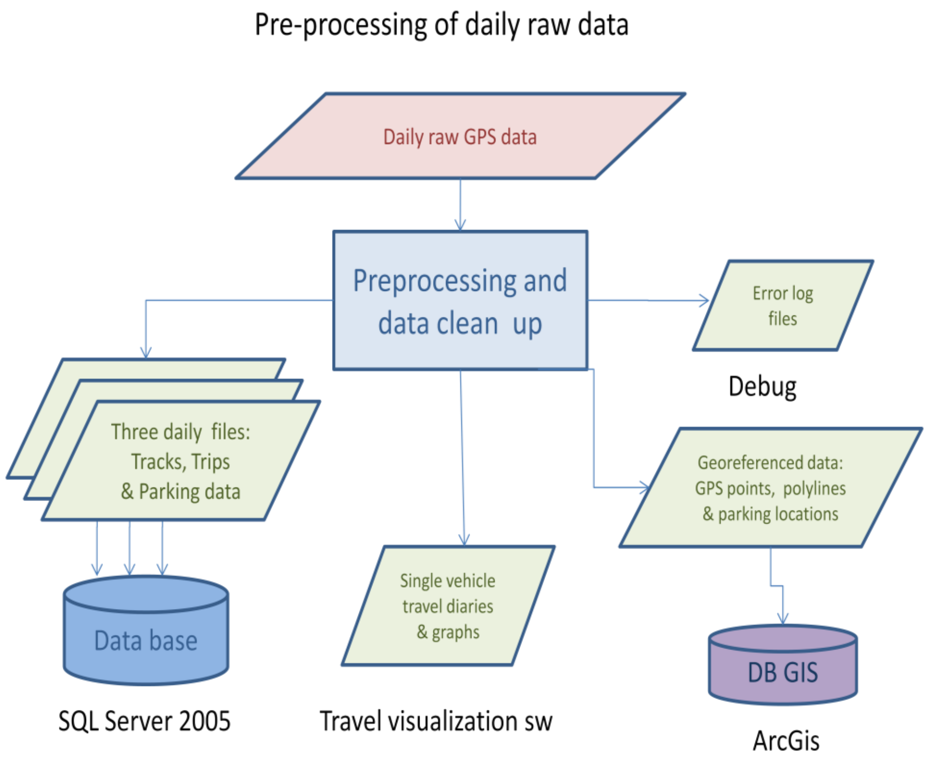

Figure 7 - Scheme of data pre-processing and cleaning

5. TRAFFIC INTENSITY AND AVERAGE SPEED

At the end of data pre-processing, three daily files are produced and stored in the DB: the information about each GPS track are memorized in the first one, while the information about trips (length, speed, location of starting and ending points, etc.) and stops are memorized respectively in the second and third file.

The data stored in the DB allow to calculate a great number of parameters concerning the circulation of monitored vehicles, which can be extrapolated to supply information about the whole private car mobility demand and traffic intensity in the road network. As matter of instance, it is possible to estimate the number and the average speed of vehicles circulating in a day, in a specific area and in a specific time interval, obtaining also a daily profile of such indicators, and the number of trips and their average length as well.

The work performed during this last year on these indicators is mainly an extension, based on a full week of data, of the case study (limited to the municipality of Florence) already carried out during the past year, which was based on one day (March 4th 2008) of data only.

Most of the results and of the traffic indicators estimated for the municipality of Florence have been confirmed and extended to the surrounding main municipality of the study area, while for some indicator the founded value are, as expected, lightly different. For instance the average length of trips, previously estimated as 4km for the town of Florence, is of about 17 km for the whole study area and the daily travelled distance, which was 14.5 km for Florence central area, is now about 47 km. However this strong increase is mainly due to commuters and to the traffic flows crossing the study area. The overall number of trips per day is ranging from 2.9 for the whole study area to 3.4 for the Florence central area.

The total number of monitored vehicles circulating in the study area during the analyzed week is about 20.000, while the same data within the municipality of Florence is of about 12.000.

Some statistics about the monitored vehicles circulating in the week from March 3rd to 9th 2008 in the Florence municipality are reported in the following table.

Table I – Monitored vehicles circulating in the week from March 3rd to 9th 2008 in the Florence municipality

	Mon 3	**Tue 4**	**Wed 5**	**Thu 6**	**Fri 5**	**Sat 8**	**Sun 9**
Number of records	73878	74866	77599	78510	84353	74179	54770
Number of tracked vehicles	4095	4021	4098	4237	4483	4338	3803
Number of travels	14189	14285	14750	14864	15924	14460	9837
Average no. of travels per single vehicle	3.46	3.55	3.60	3.51	3.55	3.33	2.59
Total travelled distance (km)	60901	61765	64999	65937	69958	64314	48968
Average duration of a travel (min)	12	13	13	13	13	12	11
Average covered distance of a travel (km)	4,3	4,3	4,4	4,4	4,3	4,4	5
Average speed (km/h)	20,9	20,1	20,3	19,5	20,2	22,4	25,4

In the graphs here below (Figure 8 and Figure 9) we have shown the number of vehicles circulating each day of the considered week and the percentage of vehicles circulating vs. the number of days for the whole study area (Florence and surrounding provinces).

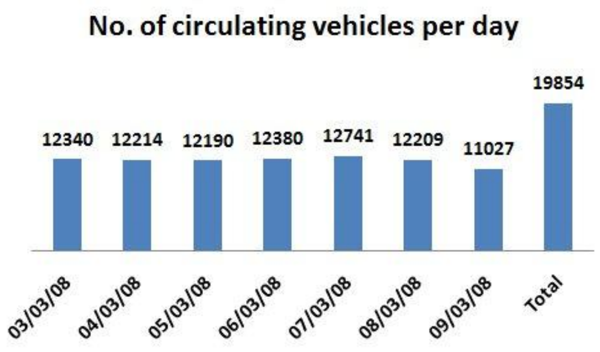

Figure 8 – Vehicles in the study area

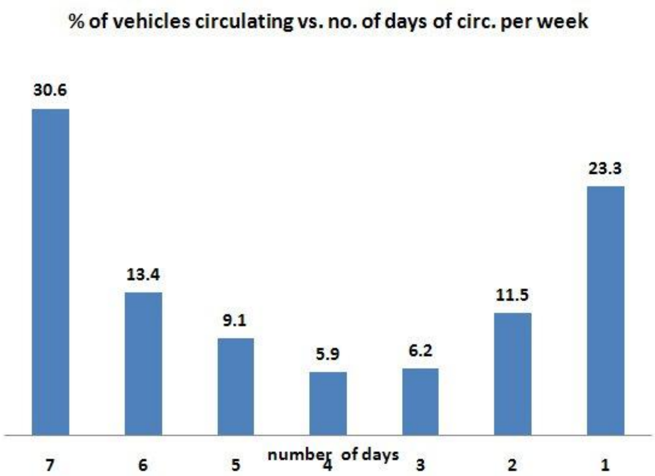

Figure 9 – Vehicle % vs. no. of days in a week

In the graph of Figure 10 we have reported the daily profile of the average speed of vehicle circulating in Florence central area during the different day of the analyzed week.

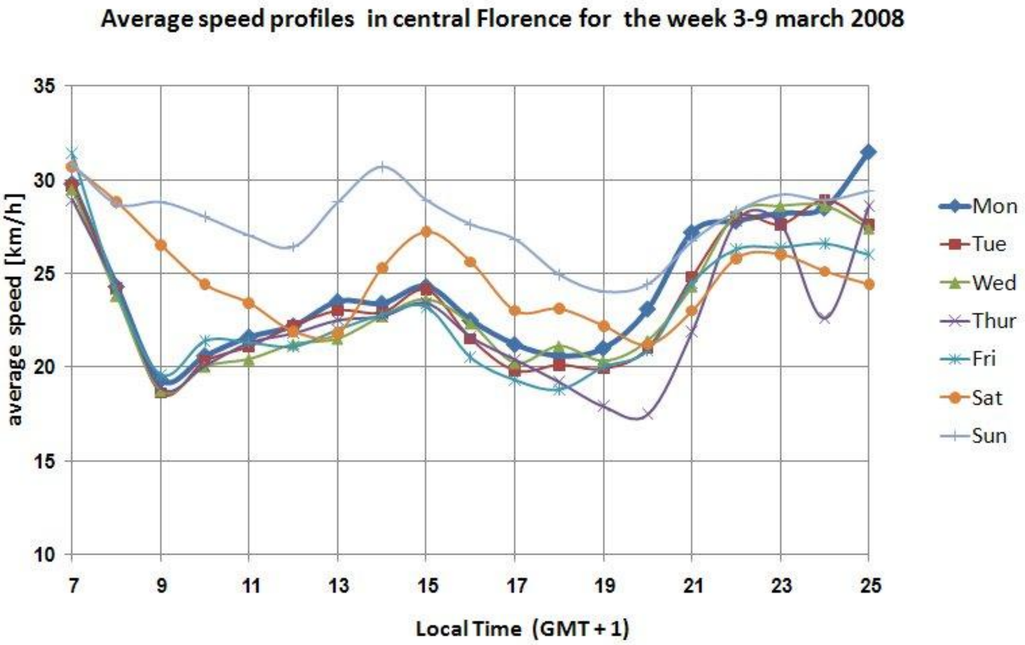

Figure 10 – Daily profile of average speed in Florence municipality (week March 3-9, 2008)

Finally we wish to point out that starting from the data stored in the above mentioned DB it is possible to calculate all the basic parameters (average speed profiles, average trip length, total number of circulating vehicle and number of departure for each time interval) needed by the well-known CORINAIR relationships to estimate the pollutant emission rate in the study area for each time interval. Of course, for this duty it also needed the knowledge of the local fleet composition (in term of vehicle emission categories) [26] and a suitable extrapolation from the monitored floating cars to the overall circulating fleet [27].

6. DEFINING THE BASIC O-D MATRIX

The focus of this study is to use data provided by floating cars to define a basic Origin-Destination matrix, relevant to systematic trips for commuting from home to work (also to study or to other important destinations), suitable for a traffic model.
While all the parameters and indicators mentioned in the above paragraph have been calculated using data provided by all circulating vehicles, we have now to create a subset of data to separate the systematic mobility from random mobility and to individuate the vehicles which have a repetitive behaviour during the week and then can supply useful information.
Starting from data stored into the above mentioned data base, it is possible to extract a daily travel diary for each of the monitored vehicle as needed to characterize, from a statistical point of view, their mobility behaviour pattern. However, to process the information relevant to many thousands of vehicles we need an automatic procedure implemented by a software.

6.1 Searching for "home"

When using information provided by the periodic census or by specific interviews and surveys to set up a new O-D matrix, the address of interviewed are obviously recorded and so the home locations of vehicle owners or users are well-known. On the contrary, in our case, because of a problem of privacy, we cannot ask for their home address. According to privacy rules, GPS data stored in our data base have been transmitted to us, throughout the Octotelematics central control unit, in an completely anonymous way and so they refer to a generic vehicle identifier ("id"code), which is not linked with any home location.

Therefore the first part of our procedure is aimed to identify the base location (what we have called "home") for each vehicle analyzing the behaviour of the vehicle itself. For this purpose we adopted the procedure reported here below.

The starting data source are the daily files where all GPS tracks are memorized. However this first analysis is based only on the first and the last record for each day and for each vehicle, and in particular on the panel status (starting = 0, underway = 1, stop = 2).

According to the daily starting and final panel status, the monitored vehicles can be subdivided in four categories: internal mobility, outgoing vehicles, incoming vehicles, transit or external commuters.

The vehicle of the fourth category are put aside and then used to analyse the crossing flows. The vehicles of the second and third categories, which have the initial starting point (outgoing) or the final stopping point (incoming) only, are considered in a subsequent step of the procedure, in order to test and to increase (or decrease) the likelihood of the detected vehicle base position.

In the first step of the procedure we use the first category of vehicles only.

All these vehicles were "filtered" on the basis of the geometrical distance between the daily first (panel =1) and last (panel= 2) GPS tracks and of the daily initial starting time (this because a starting point in the very early morning could be a departure from home or a departure from another location to go back home). The parameters used for this filtering operation must be calibrated taking into account the specific characteristics of the urban area, the parking availability and the behaviour of working people as well. In general, it must to be taken into consideration that our information is limited to parking positions.

In our case study we have excluded all vehicles having the above mentioned distance more than 1 km and we have taken all vehicles with distance less than 500 m. For distances between 500 and 1000 metres or initial starting times before 5 a.m. (local time), we made a limited use of such vehicles, as described for the vehicles of the above mentioned second and third categories.

At the end of this first step we have two positions (the first and last track of the day) for each of the vehicles, which passed through the filter, and for each day in which these vehicles are circulating.

The second step of the procedure is the comparison among all the positions referred to the same vehicle, but relevant to different days. The basic criteria for this comparison is that two or more parking positions refer to the same base location, if their distances from the geometric centre are less than a prefixed parameter (in the case study we have used 250 metres, and so the maximum distance between two parking points related to the same base location is 500 metres).

The comparison process starts calculating the geometrical centre of all points referred to the same vehicle and the distance of these points from the calculated centre. If this distance is less than a prefixed parameter, we assume to have found the searched "home" and we take the coordinates of the calculated centre as the position of the vehicle base. Otherwise the most distant point is eliminated, and a new geometrical centre is calculated. The process is then repeated until all remaining points satisfy the above mentioned condition or all points are eliminated (in this case the vehicle is excluded). Also the mentioned parameter must be calibrated in function of the characteristics of the area (in our case study we have taken 250 m.). For each of the vehicle base found , we have also calculated an index of likelihood, which is given by the number of points used to calculate the geometrical centre in the last iteration, divided by 2. Using a week of data, this index of likelihood can vary from 1 to 7.
The positions/vehicles which were partially excluded in the first step of the procedure, can now be used to add other points (having a distance from the geometrical centre less than 250 m.) in order to increase the index of likelihood of the found vehicle bases.
At the end of the procedure we should have found the base positions ("homes") for a large part of the vehicles owned or used by people resident in the study area. To each of the found vehicle bases we have also assigned an index of likelihood.

6.2 Searching for main destinations

Each vehicle contributing to the systematic mobility has at least one recurring destination. The second part of the procedure is aimed to identify the main recurring destination for each of the vehicles with an already indentified base location. For this purpose the data source are the above mentioned files containing daily information about all stops and parking times (stop coordinates, initial and final stopping times, parking duration, etc.). As mentioned above for the procedure for searching vehicle base locations, we cannot know the exact positions of destinations, but only the related parking positions. The importance of destinations depends from the number of time the destination is reached during a week and from the weekly overall time of parking there.
In order to avoid to confuse the vehicle bases with the destinations, in the first step we have eliminated from parking daily files all stops related to vehicle base locations (i.e. with a geometrical distance less than 500 m from the base). Then we can select, for each vehicle and for each day, the stops with the longest duration. The stops selected for the same vehicle in the different days are then compared each other. The comparison procedure is similar to the one mentioned above for comparing homes. However in this case to calculate the centre of the daily parking points the distances are weighted with the respective parking durations. Also in this case a destination is found when the distances of all points from the weighted centre are less than a prefixed parameter (in the case study we used 250 m.). If two parking locations are found to be referred to same destination their parking times are summed.
At the end of this step we should have found a main destination for almost all the vehicles of which the base positions is known. For each of the found main destinations we can also calculate the weekly frequency and the overall weekly parking time.

After the elimination from the starting daily files of all stops referred to the found first destination, this part of procedure can be repeated to find the second and third main destinations.

6.3 Frequency of trips between "home" and the main destinations

The aim of this third and last part of the procedure is to set for each vehicle the frequency of relationships, between the base locations and the main destinations, in the OD matrices for different time intervals. This can be made by analyzing the stored GPS data and calculating the average frequency of vehicle departures from their base locations and from the main destinations to return home. As first approximation, we can assume that in the morning OD matrices for the working days, this frequency value is 1. However, as we have seen in the previous point 5., not all the monitored vehicle are moving all the working days: in the case study only the 53.1% of monitored vehicles is moving at least 5 days. In addition we wish to define the frequency for the different time intervals during the day, which are normally considered by the different OD matrices (late morning, lunch time, afternoon, etc.).

This procedure uses, as data source, the mentioned daily trip files. In a first step all records concerning trips starting from the vehicle base locations are extracted for all the working days. The records are then subdivided into the different time intervals, considered by the OD matrices. In a second step, the records concerning the same time interval are put together for all the reference period and, for each vehicle, the total number of such trips is calculated and divided by the total number of working days in the reference period. The procedure concerning the departures from the main destinations is similar.

As result of the procedure, we will have a frequency for each vehicle and for each time interval, which can be introduced in the corresponding matrices.

6.4 Results of the application of the procedure to the case study

The procedure described above was successfully applied to the stored set of GPS data, transmitted by the on board units of the monitored cars circulating during the week from March 3rd to 9th 2008 in the mentioned study area (the province of Florence and surroundings), but also to a subset of such data concerning only the municipality of Florence. The whole study area includes more than one hundred municipalities. The municipality of Florence is subdivided in 12 macro-zones and more than 100 traffic zones.

The application of the procedure to the whole study area has allowed to identify the base locations of about 12.000 vehicles, out of the total number of about 20.000 vehicles circulating in the area during the reference period. Taking into account that about 2.000 cars are only crossing the study area, the identified base locations are about the 67% of the potential base locations.

The main destinations, which are of interest for the OD matrix of trips for commuting from home to work, concern about 10.500 vehicles.

This subset of monitored vehicles seems to be enough large to be a representative sample of the private car mobility in the study area, and seems to provide enough information to set up, by means of suitable extrapolations, a basic OD matrix of trips for commuting from home to work and to define the Florence catchment area.

However the quality of the OD matrix, resulting from the application of the above procedures, is still under examination, because at the moment we cannot know if the monitored vehicles are a sample biased or not.

The application of the procedure limited to the area of Florence municipality has identified the base location and the main destination of about 2000 of the circulating monitored vehicles. The total number of vehicles circulating in the Florence municipality in the reference period is about 12.000, but about the 53% of such vehicles are commuters coming from other municipalities. This fact point out the importance to select a suitable study area, which must cover the catchment area of the concerned urban centres, in order to minimize the commuters coming from external areas.

The results of the procedure application to the whole study area have allowed to verify the catchment area of Florence.

As matter of sample of the results of the above procedure application, three figures (Figure 11, Figure 12 and Figure 13) are reported here below to show the vehicle base locations and main destination within the Florence municipality and the study area.

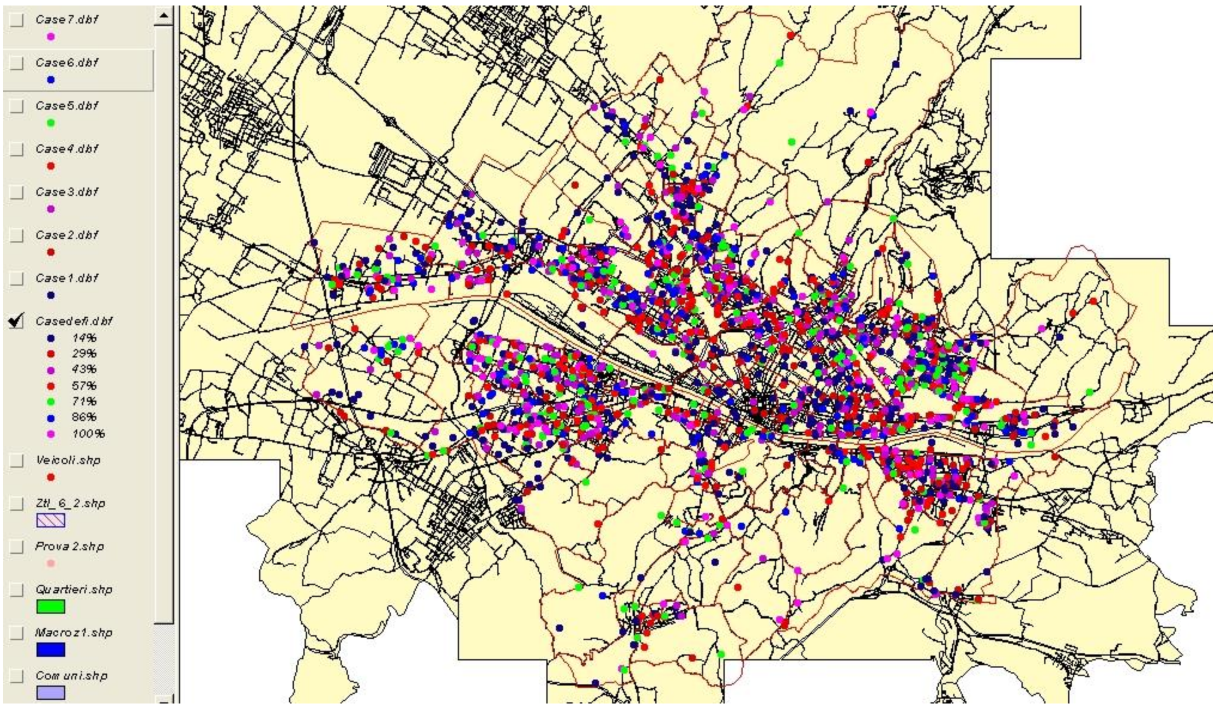

Figure 11 – Vehicle base locations in the Municipality of Florence: colours indicate the degree of likelihood

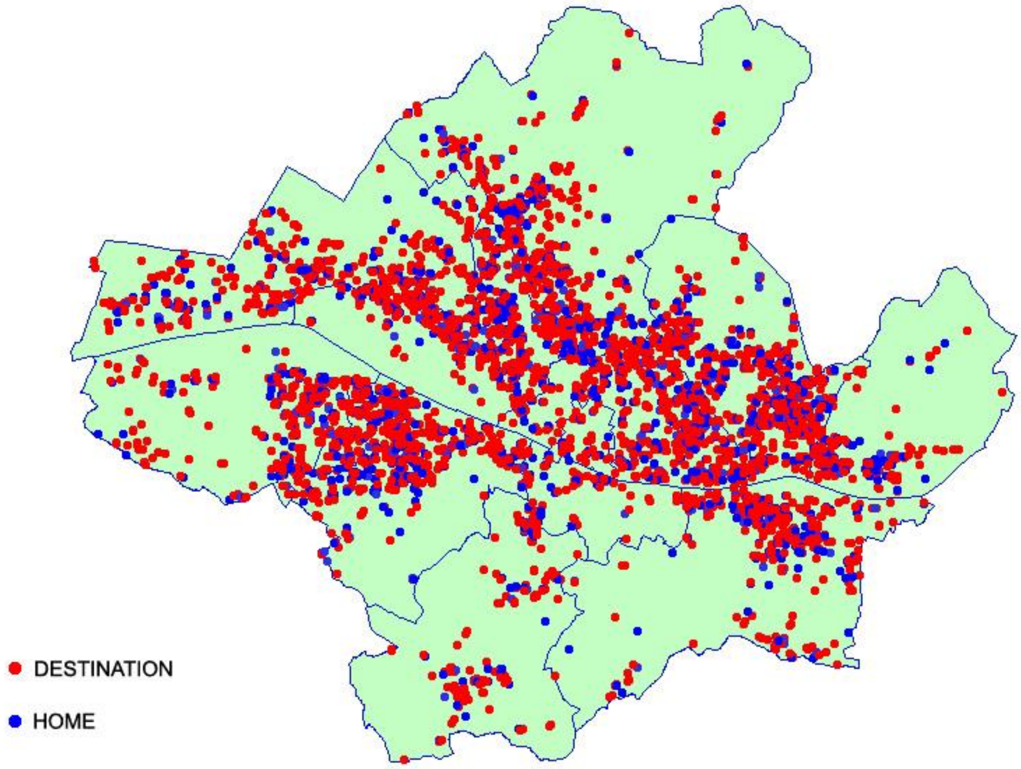

Figure 12 – Vehicle base locations and main destinations identified in the Municipality of Florence

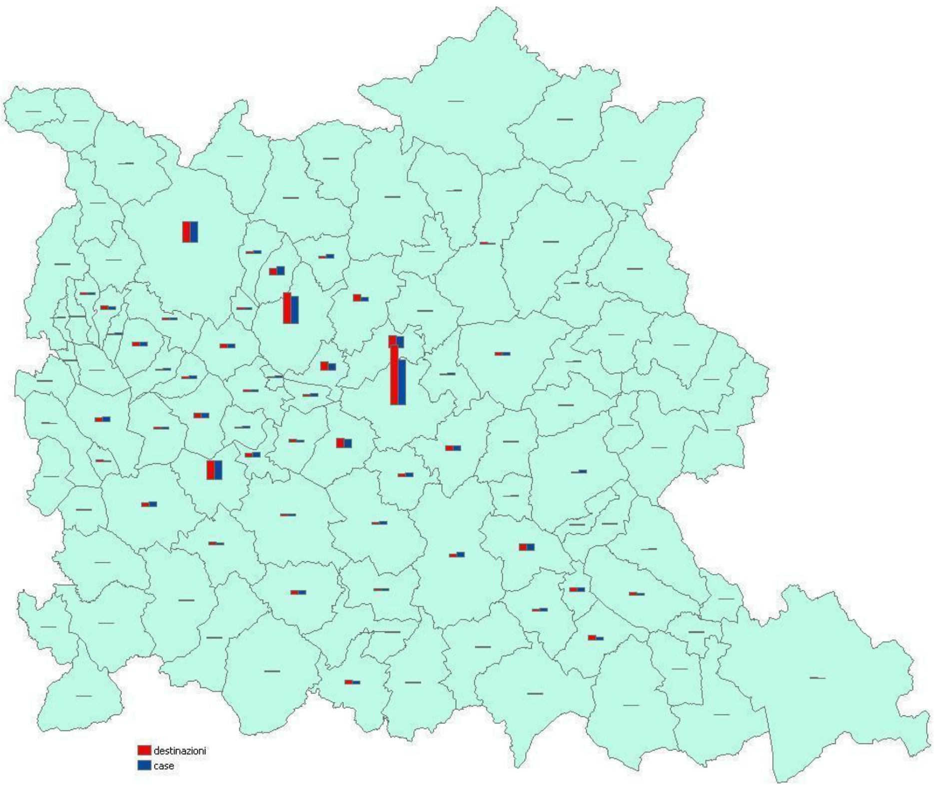

Figure 13 – Base locations (blue) and main destinations (red) localized in the Province of Florence

7. CONCLUSIONS

The paper shows the potentiality of FCD in the analysis, characterization and simulation of private car traffic in urban areas, illustrating a case study (the province of Florence), that is now underway within the Pegasus project, in which a full week of data concerning about 20.000 private cars was used.

Firstly we have described the method implemented for pre-processing and cleaning up the available GPS data and for their organization and storage.

Then we have shown some of the applications of the stored data to analyse and characterize the private car mobility demand and the traffic intensity in a road network and also to estimate the pollutant emission rate.

The focus of the paper was a new method proposed to set up a basic OD matrix of the systematic mobility by means of private cars, and to define as well the catchment area of an urban centre for commuting to work, from the stored FCD. The new procedure was described in detail, and we have demonstrated its feasibility implementing a software, which is now under calibration.

The preliminary results of the procedure application to the mentioned Florence case study are encouraging, but we still have to demonstrate that sample of FCD, used to produce the new basic OD matrix, is not biased and the matrix can be properly extrapolated to be used in a traffic model. The comparison between this new OD matrix and the matrices normally used for traffic simulation models will be the next step of our project.

ACKNOWLEDGES

The authors would like to thank the Octotelematics s.r.l. for kindly supplying the FCD used in this study.

REFERENCES

[1] Cheu, R. L., Xie, C. & Lee D. (2002). Probe Vehicles population and Sample Size for arterial speed Estimation, Computer-aided civil and infrastructure engineering, vol. 17(1), pp. 53-60.

[2] Hong, J., Zhang, X., Wei, Z., Li, L., Ren, Y. (2007). Spatial and Temporal Analysis of Probe Vehicle-based Sampling for Real-time Traffic Information System, Proceedings of the 2007 IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, June 13-15.

[3] Srinivasan, Karthik, Jovanis (1996). Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network, Transportation Research Record (1537), 15–22.

[4] Miwa, T., Tawada, Y., Yamamoto, T., Morikawa, T. (2004). En-Route Updating Methodology of Travel Time Prediction Using Accumulated Probe-Car Data, Proc. of the 11th ITS World Congress, Nagoya, Japan.

[5] Turksuma, S. (April 2000).The various uses of floating car data, Road Transport

Information and Control, 2000 Tenth Intl. Conference on, Conf. Publ. No. 472), pp.51—55.
[6] Yoon, J., Noble, B., Liu, M. (2007). Surface street traffic estimation, Proceedings of the 5th international conference on Mobile systems, applications and services, pp. 220 – 232.
[7] Kerner, B.S., Demir, C., Herrtwich, R.G., Klenov, S.L., Rehborn, H., Aleksic, M., Haug, A. (Sept. 2005). Traffic state detection with floating car data in road networks, Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems, pp. 44-49.B.S.,.
[8] Li, Y., McDonald, M. (Feb. 2005). Motorway incident detection using probe vehicles, Proceedings of the Institution of Civil Engineers Transport 158, Issue TR1, pages 11–15.
[9] Sethi, V., Bhandari, N., Koppleman, S., Schofer, J.L. (1995). Arterial incident detection using fixed detection and probe vehicle data, Transportation Research Part C, 3, No. 2, 99–112.
[10] Eisenman, S. M., List, G.F. (2004). Using Probe Data to Estimate OD Matrices, IEEE Intelligent Transportation Systems Conference, Washington. D.C., USA, October 36.
[11] Octotelematics s.r.l., www.octotelematics.it
[12] Demographic Report 2008, data from ISTAT, www.istat.it
[13] Comune di Firenze, http://www.comune.fi.it
[14] Firenze on Wikipedia , http://it.wikipedia.org/wiki/Firenze;
[15] Istituto Superiore di Formazione e Ricerca per i Trasporti, http://www.isfort.it
[16] Pugliese F. A., Montesano G., Valenti G.(2009). Urban traffic analysis from a large scale floating car data system, Bologna WIT Conference on ITS .
[17] de Fabritiis C., Ragona R., Valenti G. (2008). Traffic estimation and prediction based on real time floating car data. In: Proc. of the 11th International IEEE Conference on Intelligent Transportation Systems, pp. 197–203.
[18] G.Giuli, P.Zampetti (2009). Analisi delle distanze percorse in un giorno da veicoli privati in un'area di traffico urbano ed extra-urbano - ENEA Technical Report RT/2009/8/TER – ISSN/0393-3016
[19] G. Giuli, P. Zampetti (2008). Analisi ed elaborazione dei dati registrati a bordo di veicoli privaie: zona di Roma, Rapporto Tecnico ENEA, RT/2008/35/TER
[20] Travaglini D. (2004). Trasformazioni tra sistemi di coordinate: software disponibili, limiti e potenzialità.
[21] G.Paggi, A.Stoppini, L.Surace (1994). Tecniche per l'inserimento di rilievi GPS nella cartografia esistente. Bollettino ASIT, n.25/26.
[22] Baiocchi V., Crespi M., De Lorenzo C. (2002). Il problema della trasformazione di datum e di coordinate per applicazioni cartografiche: soluzioni informatiche e loro prestazioni, Documenti del territorio, anno XV, numero 49, pp. 11-18.
[23] http://www.centrointerregionale-gis.it/Rivista/Arretrati/49_2002/49_02.pdf
[24] A. Viviani, G. Alidori, L. Anglani, A. Conte, G. Dugheri, A. Petrucci (Nov. 2005). La mobilità a Firenze - Elaborazioni su dati provvisori al Censimento 2001 – Comune di Firenze.
[25] Sivaramakrishnan Srinivasan, Prabuddha Ghosh, Stacey Bricka, Chandra R.Bhat, (August 2005). Conversion of volunteer-collected GPS diary data into travel time

performance measures: algorithm for extracting travel-diary data from GPS streams and GPS-TDG software design, Technical report of the Center for Transportation Research, The University of Texas, Austin (U.S.A.)

[26] Ivana Cavar, Hrvoje Markovic, Hrvoje Gold (2006). GPS vehicles tracks data cleansing methodology, University of Zagreb, Croatia.

[27] http://www.aci.it

[28] Cetorelli G.,Grillo F.,Rambaldi S. (March 2009). Obiettivo:città senza traffico – La misurazione del problema e l'idea del *flexible congestion charge.* Published by Vision & Value S.r.l.