ASSESSMENT OF A TRAFFIC STATES ESTIMATION METHOD USING GPS-ENABLED CELL PHONES

Gabriel Hiribarren, Pontificia Universidad Católica de Chile, ghiribar@uc.cl

Juan Carlos Herrera, Pontificia Universidad Católica de Chile, jch@ing.puc.cl

ABSTRACT

GPS-enabled phones have opened a new way of collecting traffic data. However, better methodologies are needed to leverage this information. On this study a new methodology to estimate traffic states on arterials based on LWR is presented and assessed. Preliminary analysis suggests that this methodology works fine both at congested and uncongested situations, even for low penetration rates (1%). The methodology is also able to detect and correctly reproduce queue spillovers appearing on the road section.

Keywords: Traffic state estimation, gps-enabled cell phones, arterial roads.

INTRODUCTION

Estimating traffic states is fundamental for operational management of urban road networks, as it allows the corresponding agencies to know where and when congestion arises. This knowledge opens the possibility to develop short term transport policies and to implement real time solutions (i.e. changing traffic lights program or advanced traveller information systems).

Currently traffic state estimation is mainly done with information provided by loop detectors and cameras, which provide density and traffic flow rates at specific points. The increasing penetration of GPS-enabled cell phones has opened a new opportunity of collecting valuable real-time information of traffic conditions, providing data such as position, velocity and acceleration of the vehicle (Herrera & Bayen 2010). Based on current trends of mobile technology, the future of this technology looks promising.

Compared to the traditional fixed data collection method, this technology is cheaper because need no dedicated infrastructure, and meets the real-time requirement of UTMS (universal traffic management system) (Wang et al, 2007). It has the potential of providing a good spatial-temporal coverage of the transportation network and useful data given a minimum penetration rate in the population (Herrera et al 2010).

Over the last few years several studies have explored the use of GPS-enabled cell phones to estimate traffic states on highways (Ygnace 2000, Bar-Gera 2007, Krause et al 2008, Herrera & Bayen 2010), achieving good results with low penetration rates at different congestion

levels. Traffic state estimation on arterials have been less explored, where the existence of traffics lights add more complexity to the analysis.
The first studies about traffic state estimation on arterial roads simply proposed to estimate a traffic state for a link by averaging instant speeds or average speeds on the corresponding link (Cheu et al 2002, Zhang et al 2007). They obtained good results at 5% penetration rate for nearly 94% of their estimations. Tao et al (2012) proposed a similar model, including the application of Map Matching, Kalman Filtering and Data Screening to reconstruct trajectories from the GPS data, yielding average errors of 2.2 km/h for velocity estimation for a 7% penetration rate. Other studies used methodologies based on Markov chains, logistic regressions and STARMA models (Herring et al, 2010a; Herring et al, 2010b; Feng et al, 2012), obtaining better results than first type of models. In particular, Feng et al (2012) were able to estimate travel times with less than 2% error. However, the model does not work at oversaturated situations.

None of the studies mentioned before included traffic theory in its methodologies. If included, we believe that traffic dynamics can be better captured by the model. Another important feature of all the studies presented before is that they work with penetration rates over 5%, which currently could be realistic for developed countries but not so much in developing countries.

This article present and assess a methodology for traffic state estimation on arterial roads using GPS-enabled cell phones. The methodology is based on kinematic wave theory and this study explores the advantages and disadvantages of this new approach on traffic states estimation.

The rest of this article is organised as follows. The second section describes the proposed methodology and the computational algorithm. Then the third section explains how the methodology is assessed. Section four presents the results and it interpretation and finally the last section states the main conclusions obtained from this study and future work.

PROPOSED METHODOLOGY

The proposed methodology is based on the Kinematic Wave Theory (Lighthill & Witham, 1955 and Richards, 1956). This theory treats traffic states changes as waves which are moving on the road. The methodology tries to reconstruct these waves in order to determine when and where traffic states are changing. In other words, the methodology tries to construct a density/velocity field by identifying when and where conditions change according to the kinematic waves.
The methodology uses the fact that a wave will not change its speed or disappear until it crosses with another wave. Therefore, the methodology determine first all the waves and then it tries to find where and when waves are crossing to each other. Once this is done, the new resulting wave is constructed (if exist).

Data requirements

a. Fundamental Diagram

The fundamental diagram for the section under analysis is assumed to be known. The relationship between density k and velocity v is approximated by a hyperbolic-linear velocity function (Figure 1b), which yields a relationship between flow q and density k as shown in Figure 1a. These forms allow us to completely characterize a traffic state only by knowing its velocity.

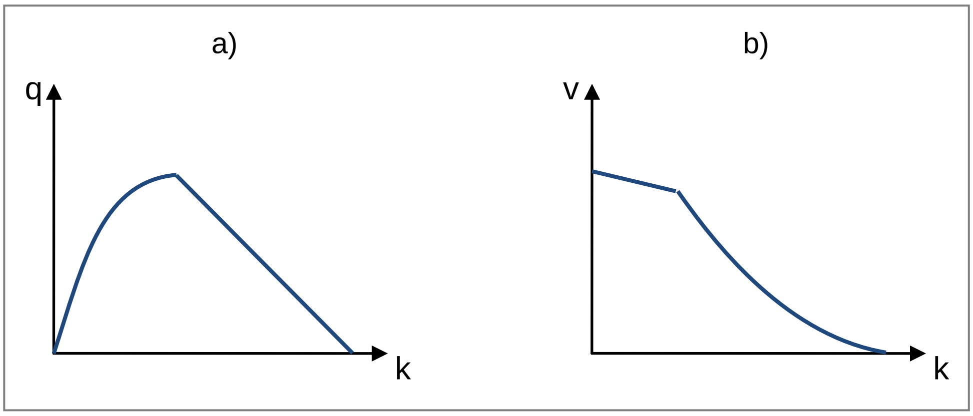

Figure 1 – Fundamental Diagram

a. Traffic Lights

As a first step, the methodology requires the knowledge of the location and timings of all traffic lights. This information is relevant because waves are generated at these locations when the lights change.

b. Trajectories

In terms of the GPS data, the methodology uses the last trajectory of an equipped vehicle that has travelled the entire section under study. Each trajectory is defined by a list of time – position points which are assumed to have no error and are obtained at a good resolution (1 second). The trajectory is used to recover the waves the vehicle encountered while travelling. As expected, having the information of the last trajectory is enough for our purposes. That is, the selected trajectory acts as a left boundary condition of the space-time area under analysis.

Reconstructing waves

The first step is to identify all possible waves based on the trajectory and traffic light programs.

Waves derived from the trajectory

The trajectory is piecewise linearized in order to determine when the speed changes (assuming instantaneous acceleration and deceleration) and what are the speeds before and after this point. Because of the functional form of the fundamental diagram, knowing the speed provides enough information to compute flow and density. Therefore, the wave speed (u) can be computed as follows:

$$u = \frac{q_1 - q_2}{k_1 - k_2} \qquad (1)$$

Subscripts 1 and 2 identify the states on both sides of the wave, and q is flow and k is density.

Waves derived from traffic lights program

When a traffic light changes, waves are generated. We now discuss how to reconstruct them.

a. Change from green to red light

In general (when the flow is different from zero) two waves will arise when light changes from green to red. The first one corresponds to the last vehicle crossing this location before the red starts. The second wave indicates how the queue is growing and can be obtained knowing the arrival traffic state and the fact that in the queue there is no flow and the density is equal to jam density. If these two states are known, the wave speed can be computed using equation 1. The arrival traffic state can be obtained from the last wave crossing the stop line. If this wave is traveling downstream (upstream), the traffic state when red light starts is equal to the one upstream (downstream) the wave. If no wave has crossed the stop line before the start of the red light, it is assumed that the traffic state is the same as the one observed by the trajectory when it crossed this location (which can be easily obtained with the fundamental diagram and the vehicle speed).

If the flow is equal to zero, two situations may explain it: there is a queue spill over from downstream or no demand arriving from upstream. In any case, no real wave arises but a pseudo-wave is generated with speed equal to zero to manage the situations in which the queue downstream disappear or when vehicles start arriving during the red light (at this point only appear the wave corresponding to queue growth).

b. Change from red to green light

When red light ends the vehicles at the queue start to move downstream, discharging at capacity (i.e. the traffic state with the highest flow). This vehicle corresponds to a wave. Another wave travels upstream as the vehicles discharge from the queue. The speed of this wave can be computed using equation 1 and knowing that the traffic state transition is from jam density to capacity.

In case that there is no queue at the end of the red light or the queue downstream still exist, no waves will appear.

c. Crossing of two (or more) waves

Once all the waves derived from the trajectory and the traffic lights are identified, the conflict of two (or more) waves crossing to each other needs to be solved. When different waves cross to each other (i.e. they converge at the same location and at the same time), they disappear and a new resultant wave arises. The speed of the resultant wave can be computed with equation 1 knowing the traffic states that this wave will divide. Since the speed of the waves converging and crossing to each other is known, these states are known as well. They correspond to the traffic state downstream the upper incident wave and the traffic state upstream the lower incident wave (an incident wave of a crossing is defined as one of the waves that cross or converge in a given point).

Iterative algorithm

The methodology generates all the waves by checking chronologically all the locations where two or more waves converge. This generates an iterative process in which one location is analysed at each iteration. The iterations must be done at chronological order in order to avoid incongruences. If there are two different intersections at the same time, the order in which they are processed is irrelevant because the waves involved in each one must be different.

Before the iteration process begins, some previous steps should be done:

1. Determine the last trajectory of an equipped vehicle that has travelled the complete section, and piecewise linearize it.
2. Recover the waves the trajectory has crossed and add them to the *active waves* set.
3. Find all the possible crossing among waves and add them to the *active crossings set*.
4. Add the points where all traffic lights change to the *active crossings* set.

The status of the algorithm after the initialization process can be seen at figure 2 part a, where the linearized trajectory is shown in blue, red light is in red, the *active waves* correspond to the orange dashed lines, and active intersections are the green dots.

Then, at each iteration five steps should be done:

1. Select a point from the *active crossings* set (chronologically).
2. For each incident wave:
    i. Add the location of the crossing as its end point.
    ii. Remove it from the *active waves* set.

    iii. Remove from *active crossings* all the future crossings of this wave with other active waves.

3. Obtain the new resultant wave and add it to the *active waves* set.
4. Add to *active crossings* set the crossing between the resultant wave and any other wave in the *active waves* set.
5. Remove the crossing point under analysis from *active crossings* set.

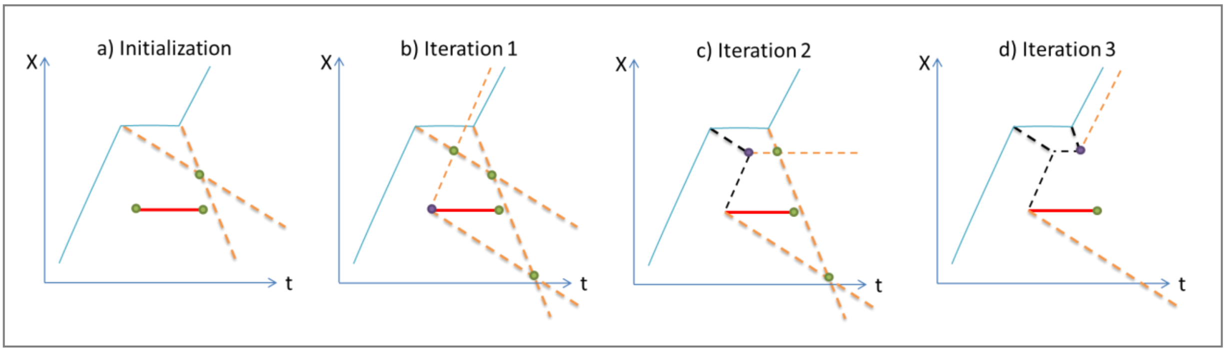

Figure 2 – Algorithm iterations a) Initialization b) Iteration1 c) Iteration 2 d) Iteration 3

For example, in figure 2 part b the crossing point under analysis is in purple. It corresponds to the beginning of the red light. It has no incident waves so we skip step 2. In this case there are two resultant waves that should be added to the *active waves* set and they generate two new possible crossings with other active waves that should be added to *active crossing* set.
For the next iteration, we should notice that the previous crossing point is not in the *active crossings* set anymore and the crossing point under analysis (purple circle in Figure 2 c) corresponds to the convergence of two waves. These incident waves finish at this crossing point, are removed from the *active waves* set (its colour changes to black) and future crossings are removed from the *active crossings* set. Then, the resultant wave (with velocity zero in this case) is added to the *active waves* set. A new crossing point between this resultant wave and an active wave is added to the *active crossings* set).
The next iteration corresponds to a crossing point between a pair of waves again. They are cut and removed from the *active waves* set and one future crossing point is removed from the *active crossings* set. Then, the resultant wave has no future crossings with other active waves and is added to the *active waves* set.

The iterative process ends when there are no more crossing points in the *active crossings* set. This always happens because we are only interested in what happens in the area between the trajectory (on the left) and the current time (on the right), and the boundaries on the section.

Finally, it is important to store all the waves that were removed from the *active waves* set because they will be needed when reconstructing the density field.

EVALUATION

The proposed methodology was evaluated using trajectories obtained with AIMSUN micro-simulator, which allow the perfect knowledge of ground truth.

To allow a systematic analysis and to avoid the impacts of geometry, a symmetrical network was constructed (see figure 3). The section under analysis has three lanes -as many important arterial roads- and crosses with other arterial as well as other minor streets. The distance between intersections is 125 meters. The section under analysis has a length of 1.9 km and includes 12 intersections.

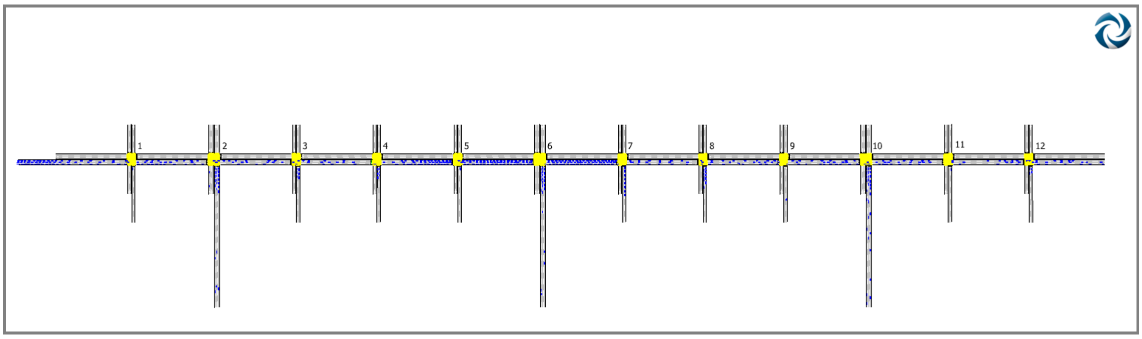

Figure 3 – AIMSUN network

The main road demand is 1875 veh/h, the crossing arterial demand is 1350 veh/h each and minor streets demand is 300 veh/h each, except for those before an arterial whose demand is just 75 veh/h. There also exist turnings in-to and out-of the main road ranging from 10 to 20 per cent of the total flow. The simulation runs for 1.5 hours. After half an hour of simulation the demand is amplified by a factor of 1.4 in order to represent a rush hour and to try the methodology under more congested situations. The traffic light programs are fixed and not coordinated.

To calibrate the fundamental diagram there was empirically calculated a capacity flow of 5645 veh/hr at a 140 veh/km density, and a jam density of 466 veh/km.

Vehicle trajectories where extracted from AIMSUN using the Surrogate Safety Assessment Model (SSAM) and one percent of them were selected randomly as equipped vehicles. It is important to notice that only vehicles that travelled the entire section were chosen. The reason for doing so is that at a real application, section should have a length of nearly three “blocks” or the distance between two arterial roads that crosses the arterial under analysis, so most of the vehicles will cross the full section. In this case a longer section is used in order to do more analysis at the same simulation.

Using all the trajectories, ground truth density field were obtained based on the generalized definitions proposed by Edie (1963). These definitions allow us to compute flow, density and spatial mean speed for a region in the time-space domain based on the distance and time travelled by all the vehicles in the region and the size of the region.

Finally, the estimates obtained with the proposed methodology are compared to ground truth. Qualitative (visual inspection) analysis are performed to assess the methodology, as presented in the next section.

All the methodology and the analysis of the results were coded in C#. The execution times are less than 2 seconds at a computer with an Intel Core I5 @3.0GHz and 7 Gb RAM. That shows the feasibility of using this methodology at real time applications.

RESULTS

Congested situation

Figure X shows the ground truth and the estimated density field. Redish colors correspond to higher densities, while greenish colors correspond to lower densities. In particular, red color represents jam density and green color represent empty zones (flow and density equal to zero). The vehicle trajectory used by the methodology is remarked with a black solid line in both fields.

Noticed that the estimated field have only a three possible states: zero vehicles (density equal to zero), critical density when vehicles are discharging at capacity after a queue, and jam density.

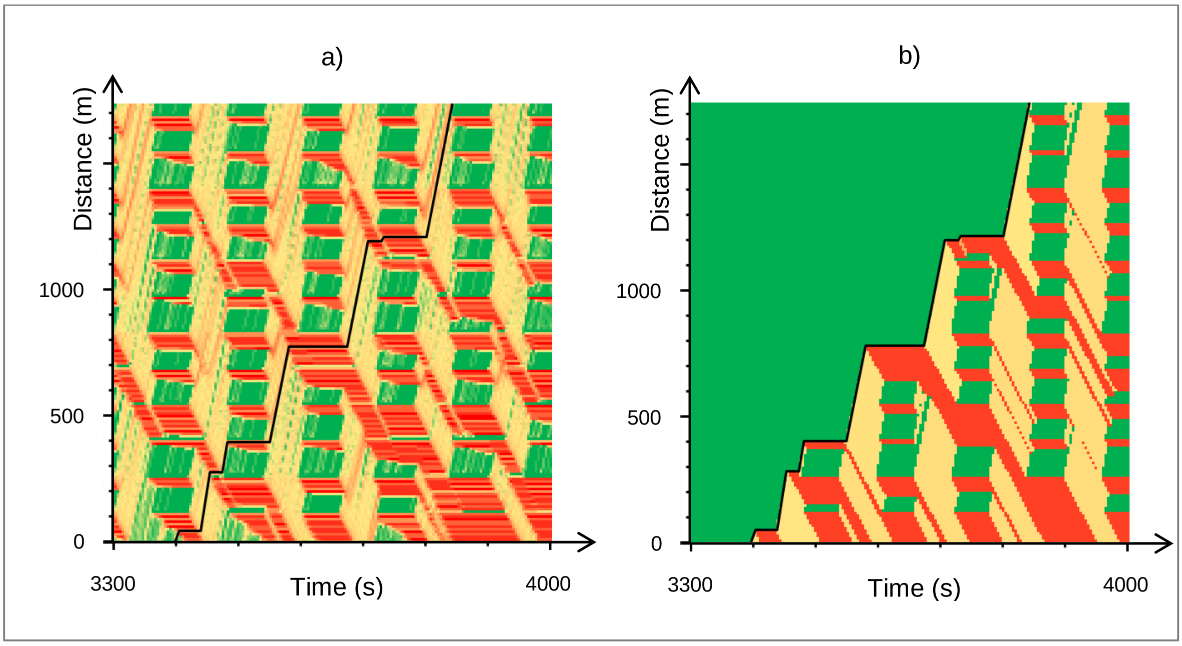

Figure 4 – Density fields under congestion a) Ground truth b) Estimated

Interestingly, the estimated field is able to capture queue spillovers, specially those that were crossed by the reference vehicle. This phenomenon is likely to occur in arterials, and impacts neighboring streets rapidly. Therefore, it is important to identify its occurrence.

Figures 5 and 6 present the density profile of the entire section under analysis at two different times. The horizontal axis is distance (in meters) with respect to the beginning of the section (i.e. vehicles flow from left to right), and the vertical axis is density (in veh/km). Figure 5 corresponds to time 4000 when all traffic light were on red, and figure 6 correspond to time 3940 when all traffic light were on green.
In figure 5, the peaks represent the presence of queues at traffic lights. As iexpected, in both curves the peaks end at the same point, which correspond to the location of the stop line. On the other hand, the beginning of the peaks is most of the times the same. That is, although the methodology is able to detect the presence of queues, there exist some error in terms of the queue length estimates.

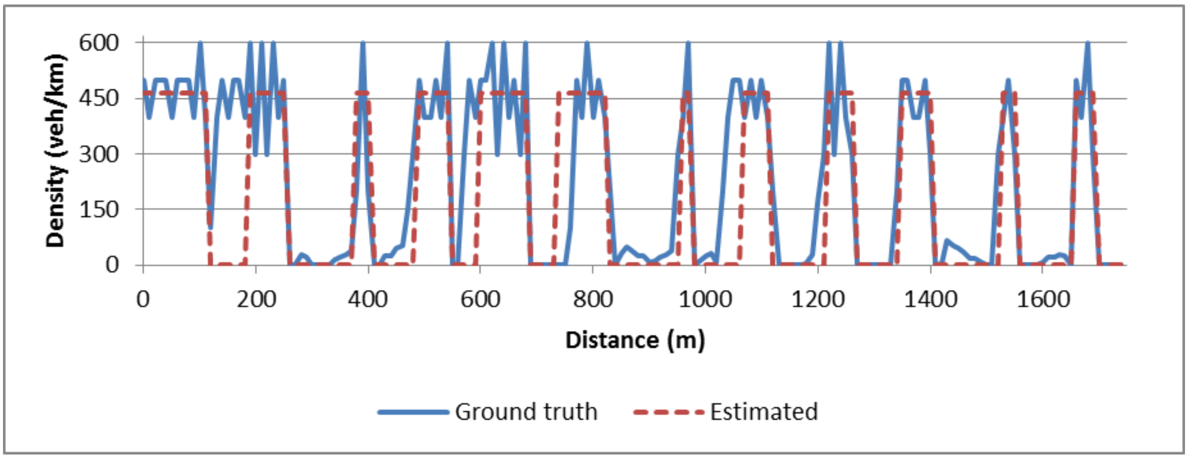

Figure 5 – Density profile at time 4000

When all the lights are on green (Figure 6), the methodology predicts critical density most of the time and detects some queues. The discrepancies between the actual queue length and the estimated queue length are evident in this case.

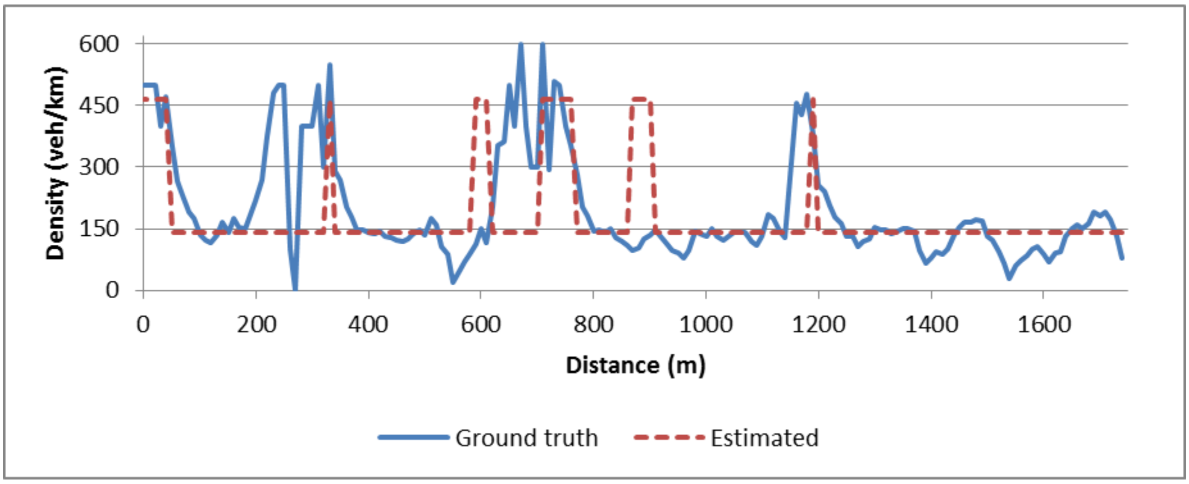

Figure 6 – Density profile at time 3940

Uncongested situation

Figure 7 shows the ground truth and the estimated density field. The trajectory used by the methodology is shown by the black solid line. Long queues or queue spill overs are not observed in this situation. However, the estimated field shows some of them, especially at the upstream zone. The reason for this is that the methodology does not have upstream boundary conditions and is assuming that vehicles are coming at a constant rate. In reality, this is usually not the case because of the presence of other upstream traffic lights . Since the methodology works with GPS information, it is impossible to trully know this boundary condition and this is one of the methodology limitations. However, the methodology can be used at many consecutive road sections, so this problem will arise only at the upstream end of the first section. In any case, the boundary condition could be obtained using other type of sensor (e.g. loop detector).

At the right side of the estimated field, some short queues do not disappear (points I, II and III in the graph). This problem was observed at most of the fields constructed during the analysis. This phenomenon is explained both by the Fundamental Diagram assumed and the LWR theory.

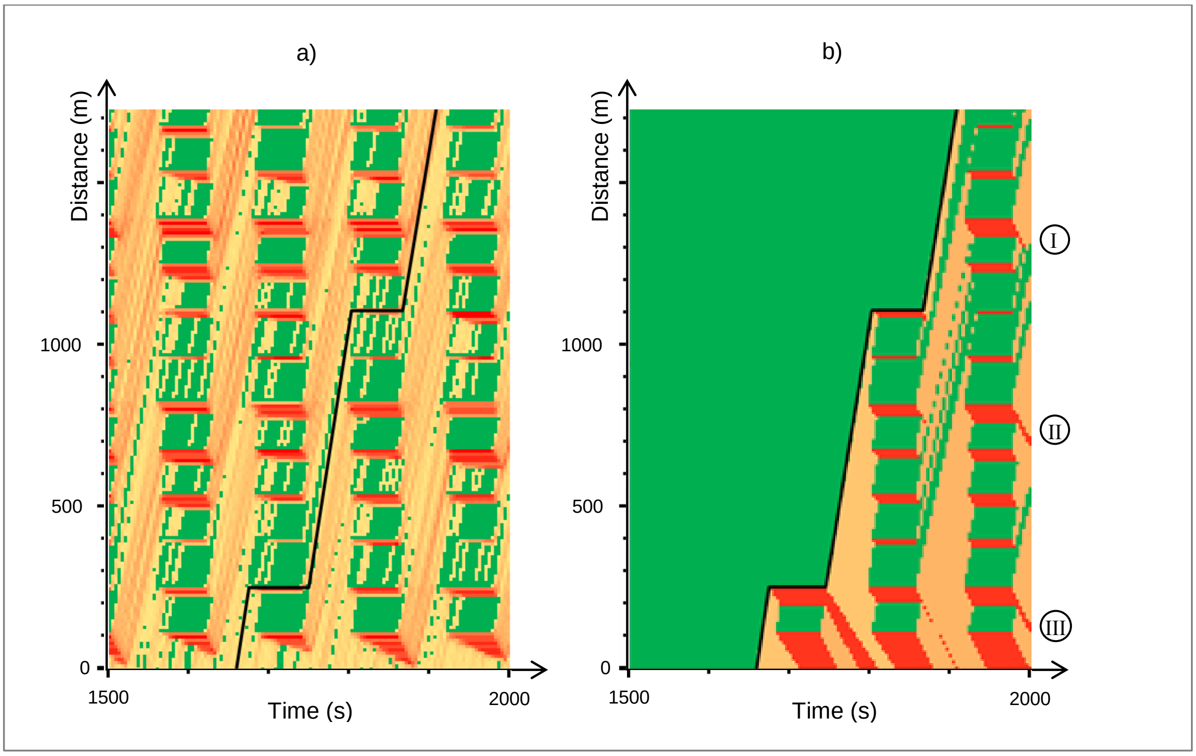

Figure 7 – Density fields without congestion

As before, two density profiles are shown at figure 8 and 9. Figure 8 shows the actual and the estimated density at time 1968 . All the queues are detected, and the length of most of them is correct.

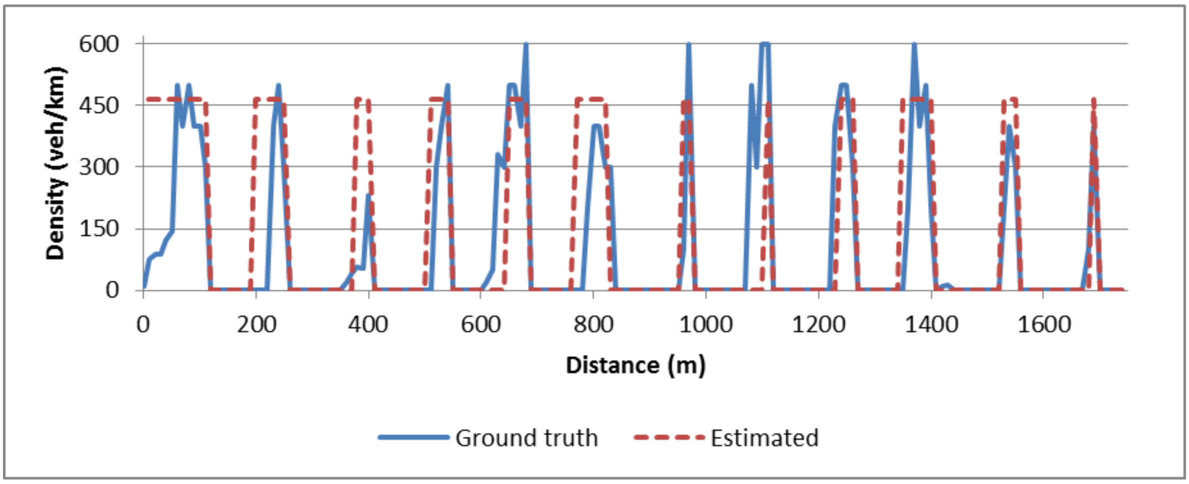

Figure 8 – Density profile at time 1968

The density profile in Figure 9 corresponds to time 2000 . In general, the density tends to be overestimated which is a consecuence of the lack of boundary conditions, resulting in an overestimation of the incoming demand. The queue observed around location 620m is detected, but its state and length are not correctly estimated. The methodology reflects correctly the places where there are no vehicles.

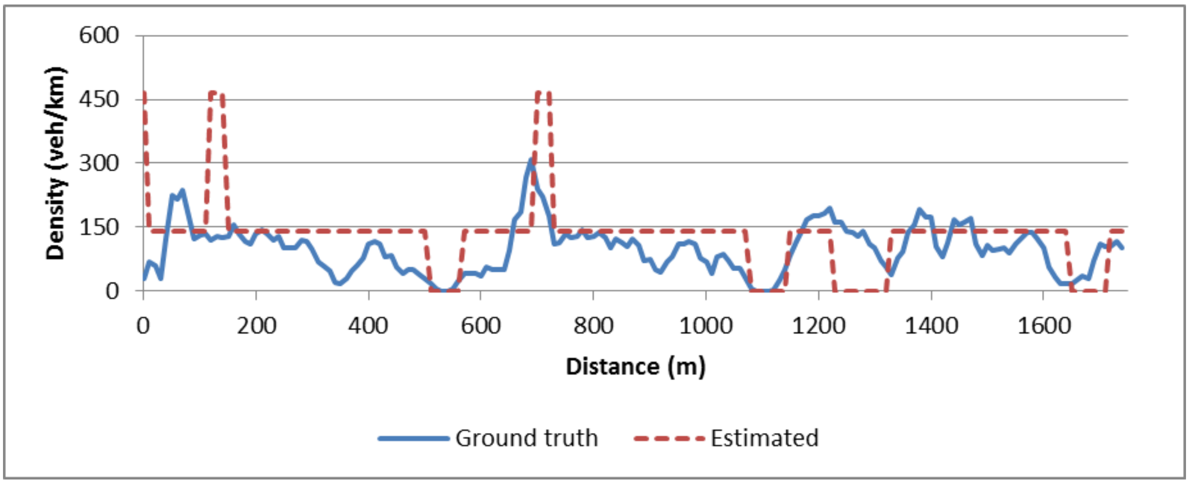

Figure 9 – Density profile at time 2000

CONCLUSIONS

Preliminary analysis suggests that the methodology proposed for traffic state estimation on arterials is promising. The results are encouraging, considering that only 1% penetration rate was tried, which is much less than most of the penetration rates found in the literature (around 5%). This constitutes an important feature of the methodology, as penetration rate is the main limitation of this type of information source, and could be explained by the use of traffic flow theory in the methodology.

The methodology was able to detect relevant traffic phenomenon, such as queue spillovers. Its correct identification is critical because of their impact on the surrounding network.

For the scenario studied, the lack of an appropriate upstream boundary condition generates an overestimation of the demand which is especially problematic during non-congested situations. The effect of this problem tends to disappear at more congested situations, which is exactly when traffic states estimation are more useful.

By visual inspection, the methodology seems to create density fields that in general fit well with the ground truth. However, density profiles show that there is still room for improvement.

The computational time required for the estimation is less than two seconds. Therefore, the methodology can be used in real time applications.

Further and deeper study must be done at assessing the methodology. Quantitative indicators would allow a better assessment of the methodology. In particular, it would be interesting to evaluate the impact on the results of using even smaller penetration rates than the one used here. Also the methodology can be used to predict future traffic states, in order to predict travel times through the arterial section. This information is extremely relevant both for drivers and transportation agencies. Finally the methodology should be evaluated on a real life situation.

BIBLIOGRAPHY

Bar-Gera, H., 2007. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. Transportation Research Part C 15 (6), 380–391.

Cheu, R. L., Xie, C., & Lee, D.-H. (2002). Probe vehicle population and sample size for arterial speed estimation. *Computer-Aided Civil and Infrastructure Engineering*, *17*, 53–60.

Edie, L. (1963). Discussion of traffic stream measurements and definitions. *Proc. 2nd Int. Symp. Theory Traffic Flow, Paris, France*, 139–154

Feng, Y., Hourdos, J., Davis, G., & Collins, M. (2012). Vehicle Probe Based Real-Time Traffic Monitoring on Urban Roadway Networks

Herrera, J. C., & Bayen, A. M. (2010). Incorporation of Lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B: Methodological*, *44*(4), 460–481.

Herrera, J. C., Work, D. B., Herring, R., Ban, X. (Jeff), Jacobson, Q., & Bayen, A. M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, *18*(4), 568–583.

Herring, R., Hofleitner, A., Abbeel, P., & Bayen, A. (2010 a). Estimating arterial traffic conditions using sparse probe data. *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems* (pp. 929–936). IEEE.

Herring, R., Hofleitner, A., Amin, S., Nasr, T. A., Khalek, A. A., Abbeel, P., & Bayen, A. (2010 b). Using mobile phones to forecast arterial traffic through statistical learning. *89th Transportation Research Board Annual Meeting, Washington DC.*

Krause, A., Horvitz, E., Kansal, A., Zhao, F., 2008. Toward community sensing. In: ACM/IEEE International Conference on Information Processing in Sensor Netwroks (IPSN), St. Louis, MO.

Lighthill, M. J., & Whitham, G. B. (1955 a). On kinematic waves I . Flood movement in long rivers. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *229*(1178), 281–316.

Lighthill, M. J., & Whitham, G. B. (1955 b). On kinematic waves II . A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *229*(1178), 317–345.

Richards, P. I. . (1956). Shock Waves on the Highway. *Operations Research*, *4*(1), 42–51.

Tao, S., Manolopoulos, V., Rodriguez, S., & Rusu, A. (2012). Real-Time Urban Traffic State Estimation with A-GPS Mobile Phones as Probes. *Journal of Transportation Technologies*, *02*(01), 22–31.

WANG, L., ZHANG, H., YU, G., & FAN, Y. (2007). Study of Probe Sample Size Model in Probe Vehicle Technology. *Journal of Transportation Systems Engineering and Information Technology*, *7*(5), 31–36.

Ygnace, J-L., Drane, C., Yim, Y., de Lacvivier, R., 2000. Travel Time Estimation on the San Francisco Bay Area Network Using Cellular Phones as Probes. California PATH Working Paper UCB-ITS-PWP-2000-18, Institute of Transportation Studies, University of California, Berkeley.

ZHANG, W., XU, J., & WANG, H. (2007). Urban traffic situation calculation methods based on probe vehicle data. *Journal of Transportation Systems Engineering and Information Technology*, *7*(1), 43–49.