Floating Car Data and Fuzzy Logic for classifying congestion indexes in the city of Shanghai

Maja Kalinic [a, *], Jukka M. Krisp[b]

[a] University of Augsburg, Augsburg, Germany, maja.kalinic@geo.uni-augsburg.de
[b] University of Augsburg, Augsburg, Germany, jukka.krisp@geo.uni-augsburg.de

* Corresponding author

Abstract: In this paper, we use Floating Car Data from the city of Shanghai and Fuzzy Inference model to detect congestion indexes throughout the city. We aim to investigate to which extent traffic congestion is severe during afternoon rush hour. Additionally, we compare our results to the ones obtained by calculating congestion indexes on conventional way. Although we do not argue that our model is the best measure of congestion, it does allow the mechanism to combine different measures and to incorporate the uncertainty in the individual measures so that the compound picture of congestion can be reproduced.

Keywords: Fuzzy logic, Floating Car Data, FCD, Traffic Congestion, Congestion Index

1. Introduction

Is there any congestion on my way to work today? How can I avoid it? If I cannot avoid it, how much time will I need to spend in it? These are just some of the questions coming from ever-growing demand of drivers. Such questions require traffic data to be accurate, reliable, timely and as complete as possible.

Throughout the years, collecting traffic data methods have been evolving considerably, which makes traffic information accessible worldwide. Conventional methods of gathering data (such as loop detectors) are still necessary but rather insufficient for obtaining good traffic information. Their restricted coverage and expensive costs of implementation and maintenance makes them less attractive than alternative methods.

One such alternative and cost – effective method is based on collecting data from "in-vehicle" devices through mobile phones or GPS and it is broadly known as Floating Car Data (FCD) acquisition method (Cohn and Bischoff, 2012). Floating Car Data (FCD) is an alternative and rather complement source of high quality data to the existing technologies. It is capable of improving safety, efficiency and reliability of the transportation system. As such, its role is becoming increasingly crucial in the development of new Intelligent Transportation Systems (ITS).

Very often, traffic data itself, as well as traffic related events come with ambiguity, uncertainty, vagueness and imprecision. Subjectivity judgement is present in many traffic phenomena such as route choice, mode of transportation, drivers' perception, established level of service, safety standards, defining criteria for alternative routing, etc. Therefore, existing deterministic and stochastic models for traffic handling cannot effectively deal with afore mentioned characteristics. Rather, we approach to these problems by using different fuzzy set theory techniques.

In this paper, we use FCD obtained from taxi vehicles in Shanghai. We build two input – one output fuzzy inference model in order to detect Congestion Indexes throughout the city during afternoon peak hours. We further discuss our findings in relation to results obtained from conventional methods of detecting congestion indexes.

2. Floating Car Data in Traffic Monitoring

The principle of FCD is to collect traffic data by locating the vehicle via mobile phones or GPS over the entire road network. In other words, vehicles act as sensors for the road network. Collected data (such as car location, speed and direction of travel) are sent anonymously to a central processing center and, if necessary, sent back to the drivers on the road, in form of useful information - status of traffic, less congested route, approximate time to be spent in congestion event, etc. (Stanica et al., 2013).

Sanwal and Walrand (1995) investigated upon opportunities to have a traffic monitoring system based on probe vehicle reports (position, speeds, or travel times), and concluded that they constitute a feasible source of traffic data. Yim and Cayford (2001) and Yim (2003) argue that if GPS equipped cell phones are widely used, they will become more attractive and realistic alternative for traffic monitoring. Zito et al. (1995) also investigated the use of GPS devices as a source of data for traffic monitoring. They performed tests to evaluate the accuracy

of the GPS as a source of velocity and acceleration data. The found accuracy level was good.

The main drawback of this technology is that its low penetration in the population is not sufficient to provide an exhaustive coverage of the transportation network. Nonetheless, Sanwal and Walrand (1995); Westerman et al. (1996); Yim and Cayford (2001) argue that data coming from no more than 5% of the total flow are sufficient to obtain acceptably accurate traffic information. Similarly Xiaowen et al. (2003) define confidence intervals for accuracy in terms of average speed and travel times across the links as a function of probe penetration. The authors conclude that 3% to 5% probe penetration is sufficient for confidence levels of 90% and above.

Moore et al. (2001); Schwarzenegger et al. (2008); Bertini and Tantiyanugulchai (2004) suggest the possibility of using dedicated fleets of vehicles equipped with GPS (such as delivery trucks, taxis or buses) to monitor traffic. Even though this type of traffic information acquisition faces challenges concerning coverage issues, penetration, biases due to operational constrains and vehicle travel patterns, we argue that it is still a viable source of data, particularly in large cities.

FCD has a wide range of applications: network performance analysis (e.g. network monitoring, before/after analysis, route choice); forecasting (traffic growth, origin – destination relations, emission modelling); road maintenance and safety analysis as well as location planning. In this particular example, we are interested how to use FCD to detect congestion indexes in urban areas.

3. Congestion Index measures

According to the existing research, several criteria should be taken into account when measuring congestion. Levinson and Lomax (1996) define congestion index as a measure of vehicle travel density on major roadways in an urban area. They further discuss that congestion index should measure congestion at a range of analysis level (a route, subarea or entire urban region) and in relation to a standard. It should provide a continuous range of values, be based on travel time data because travel time based measures can be used for multimodal analysis and adequately describe various magnitudes of congested traffic conditions. In addition, Lomax et al (1997) elaborate on issues in measuring congestion. Congestion measures have to reflect full range of road network performance, based on widely available data, and allow comparison across metropolitan areas.

Basic congestion measures are delay estimations. Delay is seen as an additional time spent in traffic in comparison to an acceptable or free - flow travel time. As a beginning of a delay threshold, Lindley (1987) use a volume to capacity (V/C) ratio of 0.77 (or the speed of 55 miles per hour (mph) corresponding to V/C ratio of 0.77). Schrank and Lomax (2005) use 60 mph for freeways and 35 mph for arterial roads as free-flow speed for comparison with congested speeds. Victoria Transport Policy Institute (2018) emphasize that some roadways have instruments that measure hourly traffic volumes and speeds. By averaging these counts, one can calculate average daily/hourly measure of traffic flow as well as average speed (Table 1). These estimates are further used as standards when compared with real world traffic data. Based upon results, one can detect and categorize five levels of congestion namely free flow, moderate, heavy, severe and extreme congestion.

Arterial road	Average Hourly Traffic Per Lane	Average Vehicle Speed (km/h)
Extreme	> 417	34
Severe	354 – 417	37
Heavy	292 – 354	43
Moderate	208 – 292	48
Free Flow	< 208	56

Table 1. Roadway Congestion Categories (original measures are expressed as Average Daily Traffic Per Lane and Average Vehicle Speed in miles per hour), based on VTPI (2018)

One of the mostly used congestion measures is Level Of Service (LOS) measure. The LOS of a facility is determined by traffic flow characteristics such as vehicle density, volume-to-capacity ratio, average speed and intersection delay, depending on facility type. The scale of LOS measure has six discrete classes ranging from A to F where A represents completely free flow and F extreme congestion (Röss et al., 1985).

Schrank et al. (1990) developed a Roadway congestion Index (RCI) as a measure of area - wide severity of congestion. The daily vehicle per mile per lane of the area is weighted by the type of the road (freeway of arterial) and compared with the total expected vehicles per mile in the area under congested conditions (as well weighted by the road type).

Lomax et al. (1997) developed Relative Delay Rate as a measure of flow quality relative to ideal or acceptable conditions. Relative Delay Rate is calculated as the ratio between difference of actual and acceptable travel time, divided by the acceptable travel time.

D'Este et al. (1999) and Taylor (1992) developed a measure called Congestion Index, as a ratio between difference of actual travel time for the section and free flow travel time, divided by the free flow travel time. Here, free flow travel time is time traveled obeying the speed limits or travel time measured in the field as density approaches zero (no congestion).

None of these measures considers that real world traffic information is not always precise and human perception of the ideal quality of flow is usually vague. Since both observations and measurements are approximate, any measure of congestion has to be associated with uncertainty regarding the accuracy of its representation of the real conditions. Real world conditions change

depending on the roadway section and traffic participant's experience and familiarity with the area. Stepwise approaches, such as LOS, can lead to a wrong impression that the measures are very well defined. Nonetheless, even a small change in the input sometimes can significantly change the outputs. As congestion is seen as a vague concept, one should include combination of conditions in order to model the "traffic participants feeling" of acceptable and good. Hence, the process of determining the degree of congestion has to involve imprecise quantities and subjective notion of acceptability, as well as to use judgement in the calculation and interpretation of the results (Aftabuzman, 2007). Therefore, we suggest including fuzzy measures. Since the fuzzy set theory recognizes the vague characteristics of traffic data, different fuzzy set theory techniques can be used to properly model traffic and transportation problems characterized by ambiguity, subjectivity and uncertainty.

4. Fuzzy Logic Theory and Fuzzy Inference System in Traffic events modelling

Fuzzy set theory was introduced by Zadeh (1965) as a means of representing and manipulating data that was not precise, but rather fuzzy. Zadeh successfully showed that vague logical statements enable the formation of algorithms that can use vague data to derive vague inferences. The aim was to mathematically represent uncertainty and vagueness and to provide formalized tools for dealing with the imprecision intrinsic to many problems (Fullér and Zimmermann, 1993).

Many traffic related problems are ambiguous, vague and characterized by subjectivity. It is hard to disregard the fact that subjective judgment is present in problems dealing with the choice of route, a driver's perceptions and reactions, an established level of service, etc. Both deterministic and stochastic models that have been developed to deal with traffic events are characterized by mathematics based on binary logic (Sarkar et al., 2012). Binary logic is, undoubtedly, the basis for the development of many scientific disciplines; however, it cannot deal effectively with traffic uncertainty, vagueness and ambiguity. Since the fuzzy set theory recognizes the vague boundary that exists in some sets, different fuzzy set theory techniques need to be used in order to properly model traffic events.

Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic (Fullér and Zimmermann, 1993). The process itself involves several phases: defining and fuzzifying input parameters, applying fuzzy rules and operators, applying implication and aggregation method, and defuzzification (if necessary).

Defining input parameters is a challenging task and involves both knowledge and experience in the specific filed (Klir and Yuan, 1995). Fuzzifying input parameters refers to assigning the crisp numerical values of selected inputs, through membership functions, into membership degrees of the fuzzy set. In order to project input variables onto outer space, one has to specify fuzzy rules and operators. Fuzzy if – then rules are specified based on previous data exploration and experience in specific traffic event. Linguistic variables of the fuzzy rule if – part are connected with AND/OR fuzzy operators, while then – part infers the conclusion out of if –part based on min operator. Then – parts often come as fuzzy sets themselves and need to be combined into a single fuzzy set. This step is called aggregation and it is followed by defuzzification – transformation of the fuzzy set into crisp values.

5. Floating Car Data (FCD) from Shanghai' Taxis

The inspected FCD set is the result of a survey on a taxi fleet in Shanghai with an average of 7120 frequently observed vehicles. This number represents the average for each hour. In total there are around 10,000 different taxi identifications (Keler and Krisp, 2016).

Original FCD dataset contains 15 selected days, between February 1st and March 1st, 2007. The data structure of the inspected data set is shown in the Table 2. Out of originally ten provided attributes, we select to work with only certain attributes – car ID, longitude, latitude, time and speed.

Field name	Details
Car ID	The unique ID of the car
Longitude	Longitude
Latitude	Latitude
Time	In form of yyyy-mm-dd
Speed	In km/h

Table 2. Data structure of the taxi FCD of Shanghai

While usually travel times are estimated from FCD, we use FCD for estimating traffic flows. Having the position (longitude and latitude) and identification of each taxi, we are able to reconstruct vehicle trajectories. Additionally, time component allows us to partition these trajectories depending on which part of the day they were recorded. We are specifically interested in the afternoon peak hours, which in Shanghai range from 4pm until 7pm. The data is further grouped based on proposed hourly distribution and new attributes are calculated – traffic flow and average speed. Traffic flow is obtained by counting the number of vehicles that pass certain cross – section (here city district 1 - 15) per unit of time (here one hour interval - 4pm, 5pm, 6pm, or 7pm).

Average speed is calculated as normal arithmetic average of the individual vehicle speeds in a time interval of one hour. Figure 1 shows the case study area and Shanghai's districts based on which new attributes – traffic flow and average speed are calculated.

Figure 1. Administrative districts of Shanghai, indexed with numbers from 1 to 15

We further compare calculated values of flow and speed within each district with the standard values proposed by VTPI (2018) for arterial roads. This conventional way of observing traffic behavior allows us to immediately see how congested Shanghai' districts are. Table 3 shows traffic flow and average speed obtained from Shanghai' taxi FCD, for all four hours of the afternoon peak hour (4pm, 5pm, 6pm and 7pm respectively).

	4pm		5pm		6pm		7pm	
District	Speed	Flow	Speed	Flow	Speed	Flow	Speed	Flow
1	15,5	853	15,2	832	15,2	864	15,6	886
2	25,6	753	24,7	133	24,9	143	23,6	140
3	24,9	1473	23,5	1437	24,1	1505	24,6	1533
4	22,2	753	19,2	748	20,5	796	21,9	791
5	17,2	1623	15,9	1565	16,4	1559	18,8	1624
6	20,5	846	17,3	817	18,3	822	21,2	849
7	21,8	793	19,6	793	20,6	794	23,0	857
8	29,4	72	27,8	68	27,4	64	29,3	72
9	22,6	950	21,6	951	21,4	993	22,6	1063
10	21,3	1106	19,4	1035	19,7	1020	23,1	1130
11	19,2	1369	17,4	1239	17,3	1237	22,1	1328
12	18,2	1382	16,7	1332	16,9	1285	19,7	1363
13	29,9	129	27,9	142	29,3	140	29,5	141
14	26,1	21	27,9	27	22,2	24	21,7	26
15	29,5	7	36,6	6	41,2	3	27,6	4

Table 3. Calculated traffic flow and average speed of the inspected taxi FCD of Shanghai, for the afternoon rush hours

6. Fuzzy Inference Model for detecting Congestion Index in districts of Shanghai

We use previously calculated traffic characteristics – traffic flow and average speed as the input variables for the fuzzy inference model. Figure 2 shows proposed fuzzy inference model with input variables, fuzzy input membership functions with assigned linguistic variables, specification of the fuzzy if – then rules, as well as output variable – Congestion Index fuzzified with five linguistic variables (CI 1, CI 2, CI 3, CI 4 and CI 5 respectively).
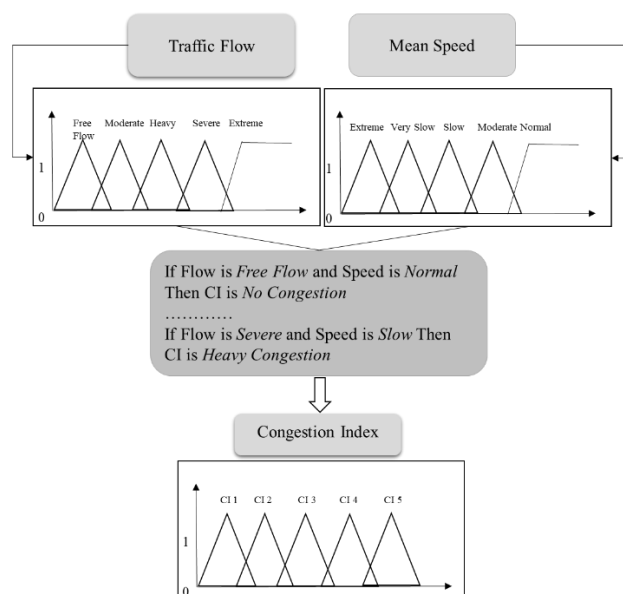

Figure 2. Fuzzy Inference model with two input variables – traffic flow and mean speed and one output variable – congestion index

We fuzzify our inputs by assigning them with five membership functions respectively. These neighboring membership functions overlap with each other by 20 - 50%, and are all of the same kind (triangular membership functions). They describe the middle range of the universe of discourse. In addition, two half-triangle membership functions represent the end of the domain of discourse respectively.

The input variable Flow is assigned with the following linguistic variables: *Free Flow, Moderate, Heavy, Severe* and *Extreme*. The input variable Speed is fuzzified as: *Normal, Moderate, Slow, Very Slow* and *Extremely Slow speed*. We further specify one output variable – Congestion Index by assigning as well five congestion levels (Table 4).

Congestion Index	Description
CI 1	No congestion
CI 2	Moderate Congestion
CI 3	Heavy Congestion
CI 4	Severe Congestion
CI 5	Extreme Congestion

Table 4. Interpretation of Fuzzy Inference Model Output Variable (Congestion Index)

Our model has 20 if –then rule combinations connected with AND operator meaning that minimum condition has to be met in order for rule to be fulfilled. We run the model to get the individual congestion indexes within all districts of Shanghai, for all four hours of rush hour event.

7. Analyses results

First results show us the distribution of traffic flow and accompanied average speed for each individual district of

Shanghai, during afternoon rush hour (Table 3). We can observe almost the same pattern in both flow and speed distribution within the entire period of time. In addition, we notice slightly different behavior of both variables in district 15, compared to the other districts. Significantly lower flow rate values are also observed in district 14.

Second results refer to the outputs of the proposed fuzzy inference model. Figure 3(a) shows the detected congestion index throughout the city at the beginning of rush hour, at 4pm. We observe extreme congestion in all districts of the city, except districts 14 and 15 where congestion shows to be no present and heavy respectively. Figure 3(b) shows detected congestion indexes in the rest of the peak hour interval, for all city districts as well. Even though similarity with the first rush hour (4pm) is obvious, we can still observe slight worsening of the traffic conditions in the district 14, where congestion goes from being heavy to severe (for the rest of the observed interval)
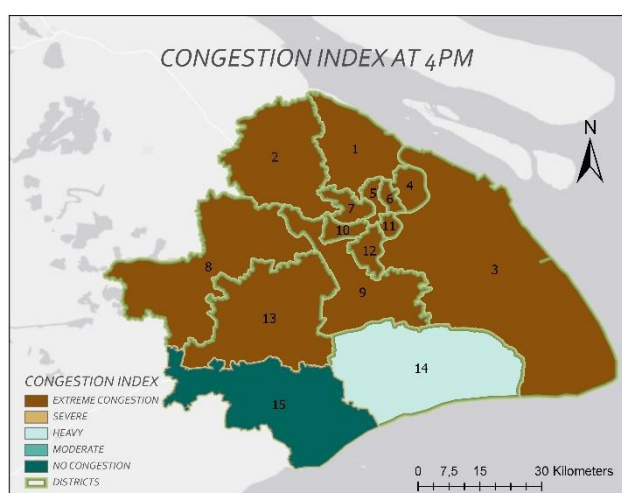
a)
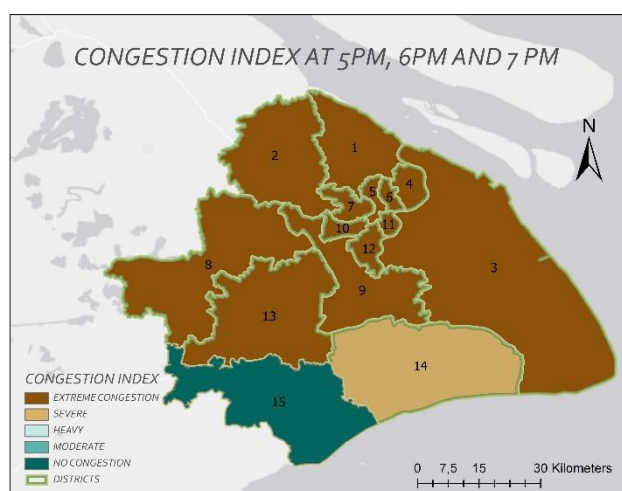
b)

Figure 3. Detected congestion indexes throughout the city of Shanghai, at the afternoon peak hours ((a) 4pm, (b) 5 - 7pm)

8. Discussion and Conclusion – feasibility of using FCD and fuzzy logic theory in detecting area – wide congestion indexes

We use FCD to obtain two variables – traffic flow and average speed. Literature findings suggest best ways of calculating travel times from FCD rather than some other elements, such as flow rates. Having that in mind, we define our own standard on how to count the number of vehicles passing a certain cross – section per time unit. In addition, we work with row FCD data meaning that we do not apply any map matching technique to match our data to existing road network, which leads to a lower accuracy of the reconstructed trajectories. On the other side, considering that penetration rate was far beyond suggested 5%, we believe that the achieved accuracy is sufficient. The good example of FCD underestimation would be district 15, where we observe no congestion or free flow most probably because of low FCD coverage in this area.

The second calculated variable is average speed. In general, speed is the distance covered per unit of time. In practice, average speed is measured by sampling vehicles in a given area over a period of time. It is important to distinguish time from space mean speed. By averaging the speeds of all vehicles in specific location interval, we get time mean speed. However, average speed measurements obtained this way are not accurate enough because instantaneous speeds averaged over several vehicles do not account for the difference in travel time for the vehicles that are traveling at different speeds over the same distance. The better approach would be to calculate space mean speed over specific interval, since space mean speed is seen as harmonic mean of the individual speeds.

Based on these considerations, our proposed fuzzy model is fed with input variables which could be further improved. That means that the model itself could show better performance, than it already has. We already stressed the importance of properly specifying input variables, as well as fuzzy if – then rules.

Nevertheless, our model was able to detect finer distribution of traffic congestion throughout the city, compared with the conventional method proposed by VTPI (2018). We observe severity of traffic in downtown area of Shanghai, as well as its surrounding districts, but also transition between heavy to severe congestion in district 14 (which was not obvious in the first results obtained through conventional approach). The inference process in our model is based on natural-language rules. These rules are consistent with the general drivers' or traffic participants' feeling of traffic. The proposed approach is simple to apply and follows common sense logic.

As depicted in Figure 3, our district areas of Shanghai are somewhat rough as they cover a rather large area. Utilizing smaller districts with the same method will reveal more detailed information on the congestion index. Within the scope of this paper, we tried to define and compute the input values for our fuzzy inference model – traffic flow

and speed using FCD. Taken together, we are convinced that there is a great potential in FCD as cost - effective and large scale in traffic monitoring.

In addition, proposed fuzzy inference method provides a mechanism that combines two congestion measures into a single composite measure of congestion (Congestion Index). We argue that this approach is appropriate because possible errors in collecting or pre – processing data cause the individual values to be imprecise and the implication of the values with respect to the severity of congestion is also ambiguous.

However, future work is needed to investigate how will model react on previously discussed changes in input variables, or including even some other measures as inputs. Additionally, further research is needed in the domain of fine – tuning fuzzy if – then rules based on study case area characteristics as well as the nature of the chosen input variables.

9. References

Aftabuzzaman, M. (2007). Measuring Traffic Congestion - A Critical Review. 30th Australasian Transport Research Forum. Available at: https://atrf.info/papers/2007/Aftabuzzaman.pdf. Accessed 30 November 2018

Bertini, R., Tantiyanugulchai, S. (2004). Transit buses as traffic probes: empirical evaluation using geo-location data. Transportation Research Record: Journal of the Transportation Research Board 1870, p. 35–45.

Cohn, N., Bischoff, H. (2012). Floating car data for transportation planning. Explorative study to technique and applications and sample properties of GPS data, Presentation slides, NATMEC 2012 Dallas.

D'Este, G.M., Zito, R. and Taylor, M.A.P. (1999). Using GPS to measure traffic system performance. Computer-Aided Civil and Infrastructure Engineering, 14(4), p. 255-265.

Fullér, R., Zimmermann, H. (1993). Fuzzy reasoning for solving fuzzy mathematical programming problems. Fuzzy Sets and Systems 60(3), p. 121–133. Available at: https://pdfs.seman-ticscholar.org/0ca9/bb36a2b7917f2652ac1062ad25c37d acc97b.pdf. Accessed 30 November 2018

Keler, A. and Krisp, J.M. (2016). Visual Analysis of Floating Taxi Data Based on Interconnected and Timestamped Area Selections. Progress in Cartography: EuroCarto 2015, p. 115-131.

Klir, G.J., Yuan, B. (1995). Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall PTR, 1995.

Levinson, H.S. and Lomax, T.J. (1996). Developing a travel time congestion index. Transportation Research Record: Journal of the Transportation Research Board 1564, p. 1-10.

Lindley, J.A. (1987). Urban freeway congestion: quantification of the problem and effectiveness of potential solutions. ITE journal, 57(1), p. 27-32.

Lomax, T., Turner, S., Shunk, G., Levinson, H.S., Pratt, R.H., Bay, P.N. and Douglas, G.B. (1997). Quantifying congestion, volume 1 & 2. NCHRP Report 398, Washington, DC: Transportation Research Board.

Moore, J., Cho, S., Basu, A. (2001). Use of Los Angeles Freeway Service Patrol Vehicles as Probe Vehicles. Technical report, Berkeley, CA.

Röss, R.P., Messer, C.J., Mschane, W.R., Fruin, J.J., Levinson, H.S., May, A.D. and Dudek, C.L. (1985). Highway Capacity Manual. Special Report 209, Washington, D.C.: Transportation Research Board.

Sanwal, K., Walrand, J., (1995). Vehicles as Probes. California PATH Working Paper UCB-ITS-PWP-95-11, Institute of Transportation Studies, University of California, Berkeley, CA.

Sarkar, A., Sahoo, G., Sahoo, U.C. (2012). Application of Fuzzy Logic in Transport Planing. International Journal on Soft Computing 3(2), p. 1 – 21.

Schrank, D., Turner, S. and Lomax, T. (1990). Estimation of urban roadway congestion in major urban areas: 1982-1987. Research Report 1131 - 2, Texas: Texas Transportation Institute.

Schrank, D. and Lomax, T. (2005). The 2005 Annual Urban Mobility Report, Texas: Texas Transportation Institute.

Schwarzenegger, A., Bonner, D.E., Kempton, W., Copp, R. (2008). State Highway Congestion Monitoring Program (HICOMP), Annual Data Compilation. Technical report, Caltrans, Sacramento, CA.

Stanica, R., Fiore, M., Malandrino, F. (2013). Offloading Floating Car Data. Proceedings of the 2013 IEEE 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM): p. 1-9.

Victoria Transport Policy Institute – VTPI (2018) Transportation Cost and Benefit Analysis II – Congestion Costs. Available at: http://www.vtpi.org/tca/tca0505.pdf Accessed: 08 December 2018

Westerman, M., Litjens, R., Linnartz, J.P. (1996). Integration of Probe Vehicle and Induction Loop Data – Estimation of Travel Times and Automatic Incident Detection. PATH Research Report UCB-ITS-PRR-96-13, Institute of Transportation Studies, University of California, Berkeley, CA.

Xiaowen, D., Martin, F., Robert, P. R., et al. (2003). A simulation evaluation of a real-time traffic information system using probe vehicles. Intelligent Transportation Systems, Proceedings 1, p. 475–480.

Yim, Y. (2003). The State of Cellular Probes. California PATH Working Paper UCB-ITS-PRR-2003-25, Institute of Transportation Studies, University of California, Berkeley, CA.

Yim, Y., Cayford, R. (2001). Investigation of Vehicles as Probes using Global Positioning System and Cellular Phone Tracking: Field Operational Test. California PATH Working Paper UCB-ITS-PWP-2001-9, Institute

of Transportation Studies, University of California, Berkeley, CA.

Zadeh, L.A. (1965). Fuzzy Sets. Information and Control 8, p. 338 – 353.

Zito, R., D'Este, G., Taylor, M. (1995). Global positioning systems in the time domain: how useful a tool for intelligent vehicle - highway systems? Transportation Research C 3 (4), p. 193–209.