



HAL
open science

Real-Time Traffic Data Smoothing from GPS Sparse Measures Using Fuzzy Switching Linear Models

Zied Bouyahia, Hedi Haddad, Nafaâ Jabeur, Stéphane Derrode

► **To cite this version:**

Zied Bouyahia, Hedi Haddad, Nafaâ Jabeur, Stéphane Derrode. Real-Time Traffic Data Smoothing from GPS Sparse Measures Using Fuzzy Switching Linear Models. The 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017), Jul 2017, Leuven, Belgium. pp.143-150, 10.1016/j.procs.2017.06.136 . hal-01572285

HAL Id: hal-01572285

<https://hal.science/hal-01572285>

Submitted on 6 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Real-Time Traffic Data Smoothing from GPS Sparse Measures Using Fuzzy Switching Linear Models

Zied Bouyahia^{a,*}, Hedi Haddad^a, Nafaâ Jabeur^b, Stéphane Derrode^c

^aDhofar University, P.O. Box 2509, P.C 211, Salalah, Sultanate of Oman

^bGerman University of Technology, P.O. Box 1816, Athaibah, PC 130, Sultanate of Oman

^cÉcole Centrale de Lyon, LIRIS, CNRS UMR 5205, Ecully 69130, France

Abstract

Traffic is one of the urban phenomena that have been attracting substantial interest in different scientific and industrial communities since many decades. Indeed, traffic congestions can have severe negative effects on people's safety, daily activities and quality of life, resulting into economical, environmental and health burden for both governments and organizations. Traffic monitoring has become a hot multi-disciplinary research topic that aims to minimize traffic's negative effects by developing intelligent techniques for accurate traffic states' estimation, control and prediction. In this paper, we propose a novel algorithm for traffic state estimation from GPS data and using fuzzy switching linear models. The use of fuzzy switches allows the representation of intermediate traffic states, which provides more accurate traffic estimation compared to the traditional hard switching models, and consequently enables making better proactive and in-time decisions. The proposed algorithm has been tested on open traffic datasets collected in England, 2014. The results of the experiments are promising, with a maximum absolute relative error equal to 9.04%.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Traffic state estimation, Gaussian linear models, fuzzy switching system.

1. Introduction

Traffic is one of the urban phenomena that have been attracting substantial interest in different scientific communities for many decades. Indeed, traffic congestions can have severe negative effects on people's safety, daily activities and quality of life, resulting into economical, environmental and health burden for both governments and organizations. Traffic monitoring has become a hot multi-disciplinary research topic that aims to minimize traffic's negative effects by developing intelligent techniques for accurate traffic states estimation, control and prediction. Moreover, recent progress in information and communication technologies has dramatically changed traffic data acquisition techniques, leading to a switch from the traditional so-called *Eulerian* perspective to the *Lagrangian* one¹. In the former perspective, fixed sensing equipments (e.g., loop detectors, video and radar cameras) are installed in fixed and pre-selected road parcels (segments) and / or intersections, and are used to collect traffic data in those particular segments / intersections (e.g., traffic flow and traffic density). This perspective is infrastructure-intensive and allows for traffic estimation only at certain locations of the transportation network. In the *Lagrangian* perspective, traffic data are collected by mobile sensors that are moving within the transportation network, which allows for covering the whole transportation network. The *Lagrangian* perspective can be implemented either in controlled or uncontrolled mode. In the controlled mode, dedicated vehicle probes are equipped with GPS receivers and communication links, and are exclusively used to move within the transportation network and collect traffic data. Where and when traffic data is collected is somehow controlled by probes' trajectories, but accurate results require a minimum penetration rate². On the other hand, uncontrolled mode is based on the use of mobile phones of road' users as sensors to collect traffic data. This mode takes benefit of the progress in the mobile communication technology to provide cheaper data collection solutions that do not need additional infrastructure's investment and maintenance costs, compared to the Eulerian and controlled Lagrangian approaches. However, relying on drivers' mobile phones to collect data raises one main challenge: the data collection process is opportunistic, in the sense that it is not possible to control where and when data will be collected in the transportation network. This depends both on the availability of the communication network (network coverage) and the presence of connected smart phones (switched-on) in space and time, leading to incomplete traffic data because of the gaps in the acquisition process. Moreover, due to the heterogeneity of measurement sources (in terms of accuracy, connectivity or redundant readings), the collected data are also noisy. Hence, a filtering pre-processing step is necessary to both remove noise and fill in the data gaps by estimating the missing traffic data.

* Corresponding author. Tel.: +(968) 23237215; fax: +(968) 23237201.
E-mail address: zbouyahia@du.edu.om

In this paper, we address the problem of data incompleteness, and we propose a solution for real-time, short-term macroscopic traffic data estimation and prediction in the context of mobile sensor-based traffic data acquisition (Figure 1). In this solution, current traffic speed and density in every road parcel g_i - denoted by $d(g_i, t)$ and $s(g_i, t)$, respectively- are sensed by drivers' smartphones and sent to a central Intelligent Transportation System (ITS). The ITS stores the collected data in a Traffic Data Repository (TDR) and updates the parameters used for the estimation and prediction model. Then, the ITS calculates the predicted parcel's density and speed (respectively denoted by $d(g_i, t+1)$ and $s(g_i, t+1)$) in Figure 1) and sends them back to end-users' smartphones, along with the file containing current updated estimation parameters (denoted by Θ_t). The ITS also uses the updated parameters to estimate the missing traffic data and to fill the gaps in the TDR. In case there is no internet connection, the prediction can be carried out on the end-users' mobile phones, but using the last estimation parameters received from the ITS for the corresponding road parcel (denoted by Θ_{t-1}). The main contribution of our solution is the use of a Fuzzy Switching Linear Model to implement a novel smoothing algorithm for traffic data estimation / prediction. Indeed, several filtering algorithms have been proposed in the literature, including Kalman filtering³, ARIMA⁴, SARIMA⁵ and particle filtering⁶ approaches. To the best of our knowledge, fuzzy switching linear models have not been used in the existing filtering approaches, which attribute two main advantages to our algorithm. In the one hand, the algorithm is linear and has an efficient computation time. Hence, it is suitable for real-time applications, and it can be run on end-users' smartphones if the internet connection is not available. On the other hand, the use of fuzzy switching techniques leads to more precise prediction results, given that it is possible to differentiate between more traffic states (e.g., slow, medium and heavy congestion), compared to the two common states in the classic switching models (congested vs. not congested).

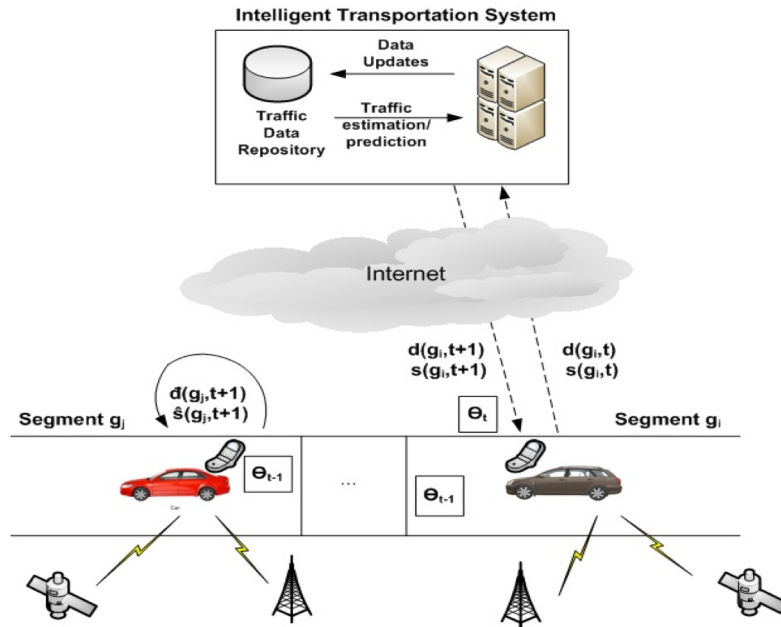


Figure 1: An architecture for smartphone-based real-time traffic data estimation and prediction.

In this paper we only present our proposed estimation algorithm, and the paper is structured as follows. In Section 2 we introduce macroscopic traffic flow modeling concepts. In Section 3 we outline the problem of traffic data acquisition using mobile probes. The estimation algorithm is detailed in Section 4, while an experimental study is presented in Section 5. Section 6 concludes the paper and outlines our future work.

2. Macroscopic traffic flow model

The definition of a macroscopic model for road traffic is an important prerequisite for the implementation of efficient control and monitoring strategies. At a macroscopic level, traffic behaves as a complex and non-linear system governed by the underlying interactions between vehicles depending on unpredictable individual human reactions. Several mathematical models have been proposed to represent the dynamics of traffic state. One of the most studied models is the Cell Transmission Model (CTM) which predicts traffic state on a road lane from observed traffic parameters - such as flow and density - at a finite set of spatial and temporal points⁷. This model is implemented by subdividing the road lane into N homogeneous parcels (or segments) called cells $c_i, 0 \leq i \leq N$ (Figure 2). The length of each cell corresponds to the distance driven during the observation period by free flow traffic. Traffic state is evaluated in each cell on short term basis with respect to three evolving traffic variables:

- Traffic average speed, denoted s_n^i , is the mean measured speed in cell c_i at time stamp n .
- Traffic average density, denoted d_n^i , is the number of vehicles in cell c_i at time stamp n .
- Traffic flow, denoted q_n^i , refers to the number of vehicles transferred from cell c_i to cell c_{i+1} at time stamp n .

We use the following notations to describe the parameters of a traffic model:

- $F_i(n)$: is the outflow of the cell $(i - 1)$;
- $A_i(n)$: is the flow that can be absorbed by the cell i ;
- $Q_{M,i}$: is the maximum flow from cell $(i - 1)$ to cell i .

Lets $s_{f,i}$ be the free flow speed, $d_i(n)$ the average vehicle density between times n and $n + 1$, $d_{j,i}$ the maximum density allowed in cell i (congestion density) and $w_{c,i}$ the backward congestion wave propagation speed. Then⁷:

$$F_i(n) = s_{f,i} d_i(n);$$

$$A_i(n) = w_{c,i} (d_{j,i} - d_i(n));$$

$$d_i(n) = \min\{F_i(n), A_i(n), Q_{M,i}\}$$

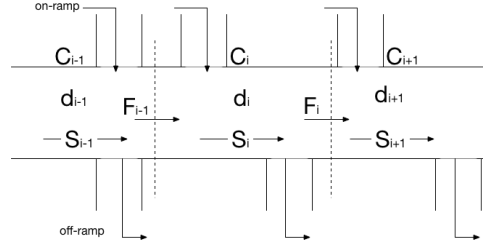


Figure 2: Cell transmission model for traffic flow on a highway.

The CTM is depicted by the so-called Model Fundamental Diagram (MFD) which ties together the traffic variables (speed, density and flow). Figure 3(a) shows the shape of a MFD. Point O corresponds to the zero-density and zero-flow state while the point J refers to the zero-flow and jam density denoted as d_{jam} . Point M refers to the maximum-flow which corresponds to density d_{max} . The slope of tangent line $[Ov]$ gives the mean free flow speed s_f . The slope of the line (MJ) corresponds to the backward congestion wave propagation speed. The speed-density diagram (Figure 3(b)) also shows the relationship between speed and density. One common assumption in macroscopic traffic modelling consists in considering linear relationship between density and speed. However, other non-linear functions can be used in several practical contexts. Due to the intrinsic correlation of traffic variables as depicted by the MFD, the traffic flow state can be explicitly estimated by two free variables d_n^i and s_n^i .

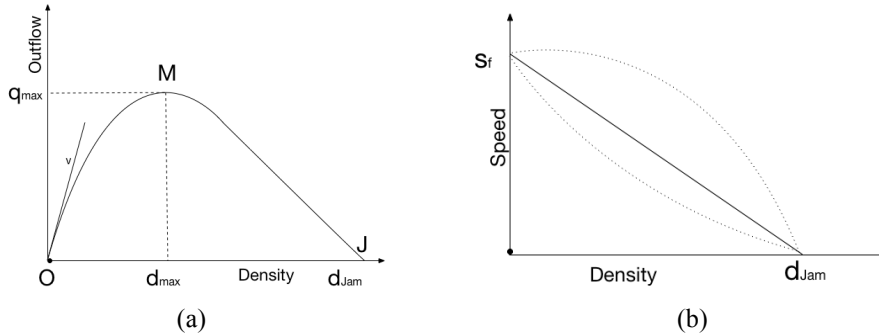


Figure 3: Model fundamental diagrams: flow density diagram (a) and speed-density diagram (b).

3. Traffic data collection using mobile probes

Instantaneous traffic data (i.e., traffic speed and traffic density) are collected from end-users' mobile devices embedding GPS sensors. Travelers' locations are mapped to roads network, providing an estimation of traffic density. Travelers' average speed is also recorded periodically on a short-term basis, allowing for a quite accurate estimation of traffic flow. The main drawback of relying on end-users' locations only to estimate traffic flow is that many users can share the same vehicle which biases the accuracy of the estimated density. Likewise, measured average speed is not reliable since some drivers can slow down for various reasons such as vehicle breakdown. An efficient approach to improve traffic data acquisition's accuracy consists in aggregating both measures (i.e., traffic velocity and traffic density). The implementation of this approach requires a centralized framework in which a database is used to collect the aforementioned GPS data on a short-term basis (e.g., every 5-15 minutes).

Table 1: Illustrative example of captured traffic data from users' GPS devices.

id	IMEI	Time	Longitude	Latitude	Speed
1	35377860998765	3/3/2017 4:00:00	17.025081	54.114485	70
2	35377860998765	3/3/2017 4:01:00	17.025019	54.116405	73
3	35377860990006	3/3/2017 4:01:00	17.025019	54.116405	74
4	33889946845688	3/3/2017 4:02:00	17.025019	54.116335	68
5	33556367688698	3/3/2017 4:02:00	17.025019	54.116337	73

From the example in Table 1, we can derive traffic data at each time point. For example, on 3/3/2017 at 4:01:00 and location (17.025019, 54.116405), the recorded density is 2 vehicles and the average speed is 73.5 km/h. The acquired GPS coordinates are mapped into the roads network which allows a real-time data acquisition of traffic velocity and traffic density in each road segment. Two temporal series, $S_1^N = (S_1, \dots, S_N)$ and $D_1^N = (D_1, \dots, D_N)$, serve to keep a record of N instantaneous speed and density measures, respectively.

Since the GPS floating sensors are not always connected, we expect several gaps in data acquisition for either traffic density or traffic average speed. Moreover, due to heterogeneity of measurement sources (in terms of accuracy, connectivity or redundant readings), we assume that the acquired data to be noisy and incomplete. Hence, a filtering/predicting algorithm is necessary not only to remove noise but also to fill in gaps in acquired data by estimating the traffic state and traffic patterns. In the following Section, we present our proposed filtering algorithm and scheme.

4. Traffic data filtering scheme

4.1. Gaussian observed fuzzy switching Markov models

Let $S_1^N = (S_1, \dots, S_N)$ and $D_1^N = (D_1, \dots, D_N)$ be two random sequences taking their values in \mathbb{R}^m and \mathbb{R}^q respectively. Let $R_1^N = (R_1, \dots, R_N)$ be an auxiliary process taking its values in \mathbb{N} and representing the model switches. We assume that S_1^N is a hidden process referring hereafter to traffic speed and D_1^N is an observed process which corresponds to traffic data. Let us also assume that $T_1^N = (S_1^N, R_1^N, D_1^N)$ is a stationary Markov chain and that $p(r_{n+1}|s_n, d_n, r_n) = p(r_{n+1}|r_n)$.

The pairwise process $Z_1^N = (S_1^N, D_1^N)$ follows a linear system called conditionally Markov switching hidden linear model (CMSHLM) defined as follows¹⁰:

$$\begin{bmatrix} S_{n+1} \\ D_{n+1} \end{bmatrix} = \underbrace{\begin{bmatrix} A_1(r_n^{n+1}) & A_2(r_n^{n+1}) \\ A_3(r_n^{n+1}) & A_4(r_n^{n+1}) \end{bmatrix}}_{A(r_n^{n+1})} \begin{bmatrix} S_n \\ D_n \end{bmatrix} + \underbrace{\begin{bmatrix} B_1(r_n^{n+1}) & B_2(r_n^{n+1}) \\ B_3(r_n^{n+1}) & B_4(r_n^{n+1}) \end{bmatrix}}_{B(r_n^{n+1})} \begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix},$$

where $A(r_n^{n+1}), B(r_n^{n+1})$ are two matrices specifying the model parameters and U_{n+1} and V_{n+1} are Gaussian unit variance white noise vectors.

The smoothing problem consists in the estimation of the hidden speed process S_1^N and the switches sequence R_1^N from the observed density process D_1^N only. However, fast smoothing is not workable in CMSHLM without setting assumptions on the parameters sets $A(r_n^{n+1})$ and $B(r_n^{n+1})$ ¹¹. Several CMSHLM variants have been derived to allow fast smoothing such as conditionally Gaussian observed Markov switching model (CGOMSM) in which $A_3(r_n^{n+1}) = 0$. In classical CGOMSMs, we assume that each switch r_n belongs to a finite set of K switches denoted as $\Omega = \{1, \dots, K\}$ allowing for modelling dynamic regime switching processes with a different set of parameters for each switch to depict nonlinear dynamic patterns. The main drawback of this hard switching model is that it is not adapted for several practical contexts since it does not take into account the transient parameters switching. To overcome the discontinuity of CGOMSM, a fuzzy approach has been proposed by Bouyahia et al.⁸, in which it is assumed that each switch is a mixture of more than one hard component in Ω . The corresponding model is called conditionally Gaussian observed switching Fuzzy Markov model (CGOFMSM).

For the sake of simplicity, we will consider that the set of hard components is limited to two jumps, i.e., $\Omega = \{0, 1\}$. Under this assumption, each switch r_n takes its values in $[0, 1]$. If we denote by ε_n^0 (resp. ε_n^1) the contribution of the hard switch 0 (resp. 1) of the switch r_n with $(\varepsilon_n^0, \varepsilon_n^1) \in [0, 1]^2$, we can simply write –without loss of generality– $r_n = \varepsilon_n$ where $\varepsilon_n = \varepsilon_n^1 = 1 - \varepsilon_n^0$.

The parameters of the interpolated fuzzy model are expressed using a bilinear function as follows:

$$\begin{aligned} A(\varepsilon_1^2) &= [(1 - \varepsilon_1)A(0,0) + \varepsilon_1A(1,0)](1 - \varepsilon_2) + [(1 - \varepsilon_1)A(0,1) + \varepsilon_1A(1,1)]\varepsilon_2 \\ B(\varepsilon_1^2) &= [(1 - \varepsilon_1)B(0,0) + \varepsilon_1B(1,0)](1 - \varepsilon_2) + [(1 - \varepsilon_1)B(0,1) + \varepsilon_1B(1,1)]\varepsilon_2 \end{aligned}$$

4.2. Parameters estimation

Parameter estimation is achieved using the Expectation-Maximization (EM) algorithm. The algorithm consists in a sequence of Q iterations aiming at maximizing the likelihood of the observed sequence during time interval $[1, N]$ ⁹.

Initially, a K-means clustering method is used to assign each pair (s_i^*, d_i^*) to one of the $2 + F$ clusters. Let $\varphi_i^0(l) = 1$, if (s_i^*, d_i^*) belongs to the cluster l and 0 otherwise $\forall 1 \leq l \leq 2 + F, 1 \leq i \leq N$. Let us also set $\forall 1 \leq i \leq M - 1, \psi_i^0(l, k) = \varphi_i^0(l) \times \varphi_{i+1}^0(k)$.

For each iteration $q > 1$, we successively perform the following items:

(i) Compute $p_{kl}^q = \frac{1}{M-1} \sum_{i=1}^{M-1} \psi_i^{q-1}(k, l)$; $\Lambda_{kl}^q = \frac{\sum_{i=1}^{M-1} z_i^{i+1T} \psi_i^{q-1}(k, l)}{\sum_{i=1}^{M-1} \psi_i^{q-1}(k, l)}$; $\Gamma_{kl}^q = \frac{\sum_{i=1}^{M-1} (z_i^{i+1T} - \Lambda_{kl}^q) (z_i^{i+1T} - \Lambda_{kl}^q)^T \psi_i^{q-1}(k, l)}{\sum_{i=1}^{M-1} \psi_i^{q-1}(k, l)}$

(ii) Compute the forward-backward probabilities defined as follows:

$$\alpha_1(k) = \int_0^1 p_{kl}^q p(s_1^*, d_1^* | r_1 = l) dv(r_1); \alpha_{i+1}(l) = \int_0^1 \alpha_i(k) \delta_i^q(k, l) \gamma(k, l) dv(\alpha_i(k))$$

$$\beta_M(l) = 1; \beta_i(k) = \int_0^1 \beta_{i+1}(l) \delta_i^q(k, l) \gamma(k, l) dv(\beta_{i+1}(l))$$

Where $z_i^{i+1} = (s_i^*, d_i^*, s_{i+1}^*, d_{i+1}^*)^T$, $\delta_i^q(k, l) = p(s_{i+1}^*, d_{i+1}^* | r_i = k, r_{i+1} = l, s_i^*, d_i^*)$ and $\gamma(k, l) = p_{kl}^q / \sum_{i=1}^{2+F} p_{kl}^q$.

(iii) Compute $\psi_i^q(k, l) = \frac{\alpha_i(k) \delta_i^q(k, l) \beta_{i+1}(l) \gamma(k, l)}{\sum_{k^*, l^*} \alpha_i(k^*) \delta_i^q(k^*, l^*) \beta_{i+1}(l^*) \gamma(k^*, l^*)}$.

(iv) Let $\Gamma_{kl} = \begin{bmatrix} \Gamma_k^{x,x} & \Sigma_{kl}^{x,y} \\ \Sigma_{kl}^{y,x} & \Gamma_l^{y,y} \end{bmatrix}$, compute matrices $A(r_n^{n+1})$ and $B(r_n^{n+1})$ using the following formulae:

$$A(r_n^{n+1}) = \Sigma_{kl}^{y,x} \Gamma_k^{x,x^{-1}}; B(r_n^{n+1}) = \Gamma_l^{y,y} - A(r_n^{n+1}) \Sigma_{kl}^{x,y}.$$

Remark: It is not always possible to derive a closed form formula for the integration with respect to $\nu(r_n)$. Hence, we will use the following approximation: $\int_0^1 h(t) dv(t) \approx h(0) + \frac{1}{F} \sum_{j=1}^{F-1} h\left(\frac{j}{F}\right) + h(1)$. This approximation relies on a finite set of F discrete fuzzy levels (see Figure 5). Using this approximation scheme, fast smoothing is achieved with a set of parameters estimated for $2 + F$ discrete components. Furthermore, we do not have an efficient computational scheme to determine the optimal number of fuzzy levels F . However, we underline that the larger is F , the more accurate are smoothing results and the higher is the computational cost.

4.3. Fast smoothing in CGOFMSM

Smoothing within the framework of CGOFMSM consists at first place in computing the probability $p(r_{n+1} | d_1^N)$, the mean vector $\mathbb{E}[S_{n+1} | r_{n+1}, d_1^N]$, and the variance matrix $\mathbb{E}[S_{n+1} S_{n+1}^T | r_{n+1}, d_1^N]$ from $p(r_n | d_1^N)$, $\mathbb{E}[S_n | r_{n+1}, d_1^N]$ and $\mathbb{E}[S_n S_n^T | r_{n+1}, d_1^N]$. Since (R_1^N, D_1^N) is assumed to be a Markov chain, it is possible to achieve the computation of $p(r_n^{n+1} | d_1^{n+1})$. Indeed, we have:

$$p(r_n^{n+1} | d_1^{n+1}) = \frac{p(r_n^{n+1}, d_{n+1} | d_1^n)}{p(d_{n+1} | d_1^n)} = \frac{p(r_n | d_1^n) p(r_{n+1}, d_{n+1} | r_n, d_n, v_n)}{p(d_{n+1} | d_1^n)} = \frac{p(r_n | d_1^n) p(r_{n+1} | r_n) p(d_{n+1} | d_n, r_n^{n+1})}{p(d_{n+1} | d_1^n)} \quad (1)$$

Moreover, using the fuzzy model, we can write:

$$p(d_{n+1} | d_1^n) = \int_0^1 \int_0^1 p(r_n | d_1^n) p(r_{n+1} | r_n) p(d_{n+1} | r_n^{n+1}, d_n) d(\nu(r_n) \otimes \nu(r_{n+1}))$$

$p(r_{n+1} | r_n)$ is given and $p(d_{n+1} | d_n, r_n^{n+1})$ follows a Gaussian distribution with mean $A_4(r_n^{n+1}) d_n$ and variance $B_3(r_n^{n+1}) B_3(r_n^{n+1})^T + B_4(r_n^{n+1}) B_4(r_n^{n+1})^T$. Then eq. (1) gives:

$$p(r_{n+1} | d_1^{n+1}) = \frac{1}{p(d_{n+1} | d_1^n)} \int_0^1 p(r_n, r_{n+1}, d_{n+1} | d_1^n) dv(r_n) \quad (2)$$

Besides,

$$p(r_n | r_{n+1}, y_1^{n+1}) = \frac{p(r_n | r_{n+1}, y_1^{n+1})}{p(r_{n+1} | y_1^{n+1})} \quad (3)$$

Optimal fuzzy switching fast smoother (OFSFS) runs as follows:

(i) compute $p(r_{n+1} | y_1^{n+1})$ (4)

(ii) compute $\mathbb{E}[Z_{n+1} | r_n^{n+1}, d_1^n] = A(r_n^{n+1}) \begin{bmatrix} \mathbb{E}[S_n | r_n^{n+1}, d_1^n] \\ d_n \end{bmatrix}$ and

$$\mathbb{E}[Z_{n+1} Z_{n+1}^T | r_n^{n+1}, d_1^n] = B(r_n^{n+1}) B(r_n^{n+1})^T + A(r_n^{n+1}) \begin{bmatrix} \mathbb{E}[S_n | r_n^{n+1}, d_1^n] \\ d_n \end{bmatrix} \left[A(r_n^{n+1}) \begin{bmatrix} \mathbb{E}[S_n | r_n^{n+1}, d_1^n] \\ d_n \end{bmatrix} \right]^T$$

$$\text{Let } \text{Var}[Z_{n+1} | r_n^{n+1}, y_1^n] = \begin{bmatrix} \omega(r_n^{n+1}) & \xi(r_n^{n+1}) \\ \chi(r_n^{n+1}) & \rho(r_n^{n+1}) \end{bmatrix} = \mathbb{E}[Z_{n+1} Z_{n+1}^T | r_n^{n+1}, d_1^n] - \mathbb{E}[Z_{n+1} | r_n^{n+1}, d_1^n] \mathbb{E}[Z_{n+1} | r_n^{n+1}, d_1^n]^T$$

- (iii) compute $\mathbb{E}[S_{n+1}|r_n^{n+1}, d_1^{n+1}] = \mathbb{E}[S_{n+1}|r_n^{n+1}, d_1^n] + \xi(r_n^{n+1})\chi(r_n^{n+1})^{-1}[d_{n+1} - \mathbb{E}[D_{n+1}|r_n^{n+1}, d_1^n]]$
and $\mathbb{E}[S_{n+1}S_{n+1}^T|r_n^{n+1}, d_1^{n+1}] = \omega(r_n^{n+1}) - \xi(r_n^{n+1})\chi(r_n^{n+1})^{-1}\rho(r_n^{n+1}) + \mathbb{E}[S_n S_n^T|r_n^{n+1}, d_1^{n+1}]$
- (iv) compute $\mathbb{E}[S_{n+1}|r_{n+1}, d_1^{n+1}] = \int_0^1 \mathbb{E}[S_{n+1}|r_n^{n+1}, d_1^{n+1}]p(r_n^{n+1}|d_1^{n+1})d\nu(r_n)$ and $\mathbb{E}[S_{n+1}S_{n+1}^T|r_{n+1}, d_1^{n+1}] = \int_0^1 \mathbb{E}[S_{n+1}S_{n+1}^T|r_n^{n+1}, d_1^{n+1}]p(r_n^{n+1}|d_1^{n+1})d\nu(r_n)$

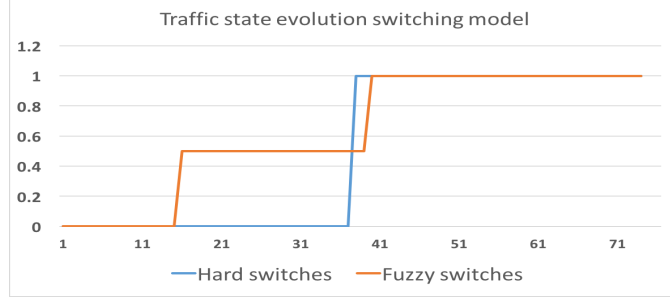


Figure 4: Example of hard and fuzzy switching traffic state models (level 1 corresponds to traffic congestion and 0 represents free flow traffic).

4.4. Qualitative interpretation of smoothed data

One novelty in our work consists in assigning to each time interval a qualitative description of traffic data. In fact, the proposed switching regime model provides an easily interpretable traffic flow state. For example, the use of 2 hard switches corresponds to two traffic states (traffic congestion and free traffic). The use of fuzzy switches allows a representation of transient traffic states, which does not only provide more efficient estimation of traffic data but also enables proactive measures for decision makers. For example, a gradual transition from free traffic to traffic bottleneck can be observed by a sequence of transitional states during a lapse of time long enough to avoid irretrievable congestion state. Figure 5 shows an example of traffic state evolution switching model. The use of steep switching model does not allow a progressive representation of traffic state evolution.

5. Experimental Study

5.1. Experimental setup

In order to evaluate the efficiency and the accuracy of our proposed fuzzy fast traffic data estimator, we used online highway traffic data captured on several segments of England express ways from January 1st 2014 to January 31st 2014. The data sets provide the average density and speed with a temporal frequency of 15 minutes. These data have not been collected by mobile sensors, but according to the Eulerian perspective. However, we used them to evaluate the performance of the proposed algorithm on real traffic data. The same algorithm can be applied if traffic density is derived from GPS data captured from end-users' smartphones or on-board car GPS sensors. The objective is to estimate traffic velocity from traffic density. Traffic data from January 1st 2014 to January 20th 2014 were extracted to serve as historical data for parameters estimation. Traffic density data from January 21st 2014 to January 30th 2014 are used to estimate the equivalent traffic speed data. The accuracy of the estimation algorithm is evaluated using the mean absolute percentage error defined as follows:

$$m = \frac{1}{N} \sum_{n=1}^N \left| \frac{\hat{s}_n - s_n}{s_n} \right|,$$

where \hat{s}_n is the estimated traffic velocity and s_n is the real one.

Let us denote by m_H and m_F the mean absolute percentage error of hard smoother and fuzzy smoother respectively. Table 2 reports the results. We notice that our smoothing algorithm yields encouraging results with an absolute relative error which does not exceed 9.04% (Site 105). Figure 5 illustrates the trajectories of real data together with estimated process. Unsurprisingly, we notice that the fuzzy smoothed data are more accurate than those estimated by the hard estimator. Our smoother is also able to detect highway traffic bottlenecks from vehicle density fluctuations. Figure 6 shows the trajectories of estimated switches depicting the variation of traffic speed. In order to evaluate the ability of our proposed model to withstand unpredictable fluctuations of traffic data (density and speed), we focused on measures corresponding to irregular traffic pattern. We noticed that our fuzzy smoother detected congested states more accurately than the hard counterpart.

Table 2: Mean relative error of speed from 1/2/14 to 4/2/2014 (450 samples).

	Site 104	Site 1048	Site 1049	Site 105	Site 1050
$m_{F=5}(\%)$	4.15	5.45	5.62	6.03	5.11

$m_{F=1}(\%)$	6.40	6.7	6.22	7.16	6.34
$m_H(\%)$	6.97	7.33	8.03	9.04	7.08

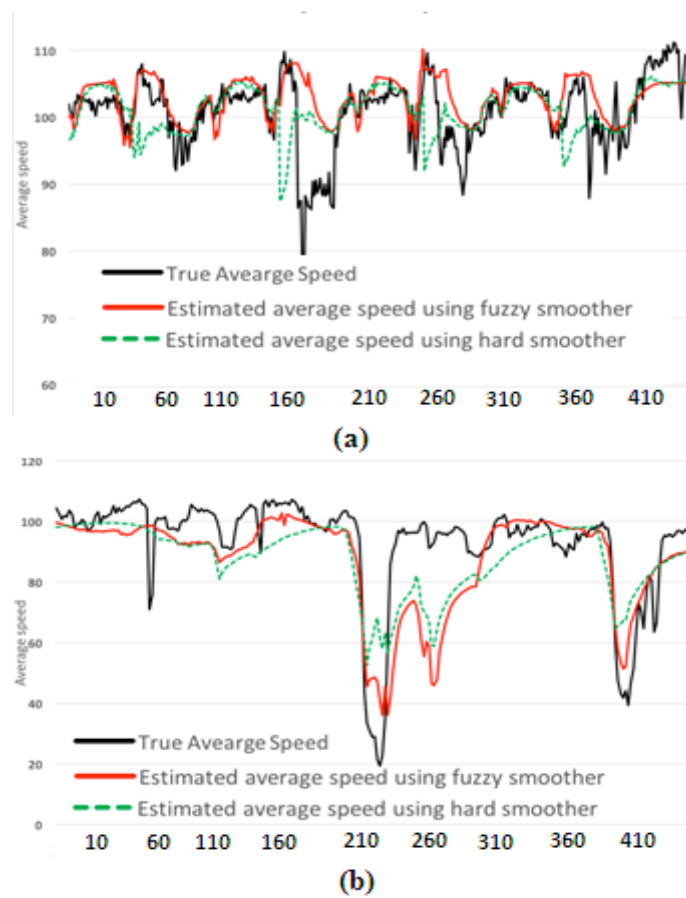


Figure 5: Traffic data estimation results for Site 1048(a) and Site 105 (b) datasets. True traffic average speed (black, solid), estimated traffic speed using fuzzy smoother (red, solid) and estimated data using hard smoother (green, dashed).

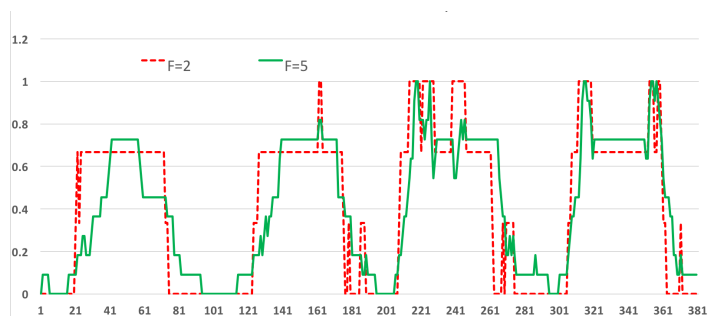


Figure 6: Estimated fuzzy switches using hard smoother and 6 levels fuzzy smoother for traffic speed estimation using CGOFMSM on 2014 Site 105.

6. Conclusion and future work

In this paper, we proposed a novel algorithm for traffic state data estimation from GPS data and using fuzzy switching linear models. The use of fuzzy switches allows the representation of intermediate traffic states, which provides more accurate estimation of traffic data compared to the classical hard switching models, and consequently gives the ability to propose better proactive and in-time decisions. The proposed algorithm is unsupervised and performs in reasonable time which motivates its usability in. The algorithm has been tested on open traffic datasets collected in England, 2014. The results of the experiments are promising, with a

maximum absolute relative error of about 9%. However, more experiments are required in order to evaluate the performance of the algorithm with mobile probes, either using end-users' smartphones or GPS-equipped vehicles.

References

1. Allström A., Barcelo J., Ekstrom J., Grumert E., Gundlegard D. and Rydergren C. (2016). Traffic Management for Smart Cities. Designing, Developing, and Facilitating Smart Cities-Urban Design to IoT Solutions, Angelakis et al. (Eds.), pp. 211-240, Springer, ISBN 978-3-319-44924-1_11
2. Herrera J.C. and Bayen A.M. (2010). Incorporation of Lagrangian measurements in freeway traffic state estimation, *Transportation Research Part B: Methodological*, Vol. 44(4), pp. 460–481, May 2010
3. Work D. B., Tossavainen O.-P., Blandin S., Bayen A. M., Iwuchukwu T. and Tracton K. (2008). An ensemble kalman filtering approach to highway traffic estimation using gps enabled mobile devices, *Proceedings of the 47th IEEE Conference on Decision and Control (CDC '08) December 2008 Cancun, Mexico* IEEE5062506810.1109/cdc.2008.47390162-s2.0-62949103909
4. Gavirangaswamy, V. B., Gupta, G., Gupta, A., & Agrawal, R. (2013). Assessment of ARIMA-based prediction techniques for road-traffic volume. In *Proceedings of the 5th International Conference on Management of Emergent Digital EcoSystems, MEDES 2013*, pp. 246-251, DOI: 10.1145/2536146.2536176
5. Lee S. and Fambro D.(1999). Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting, *Transp. Res. Rec.*, vol. 1678, pp. 179–188, 1999.
6. Hawes M., Amer H.M. and Mihaylova L. (2016). Traffic State Estimation via a Particle Filter with Compressive Sensing and Historical Traffic Data. 19th international Conference on Information Fusion (FUSION),05-08 July 2016, Hidelberg, Germany. IEEE. ISBN 978-0-9964-574-8
7. Daganzo C.F. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory, *Transportation Research Part B: Methodological*, Vol. 28(4), August 1994, pp. 269-287
8. Bouyahia Z., Derrode S. and Pieczynski W. (2016). An exact smoother in a fuzzy jump Markov switching model, in 6th Int. Workshop on Representation, analysis and recognition of shape and motion From Imaging data, Tunisia, Tunisia, 2016.
9. Gorynin I., Derrode S., Monfrini E., and Pieczynski W. (2015) Exact Fast Smoothing in Switching Models with Application to Stochastic Volatility, *EUSIPCO, Nice, France*, pp. 924-928, 2015.
10. Pieczynski W. (2011). Exact filtering in conditionally Markov switching hidden linear models, *Comptes Rendus Mathematique*, vol. 349, no. 9, pp.587–590, 2011.
11. Derrode S. and Pieczynski W. (2013). Exact fast computation of optimal filter in Gaussian switching linear systems, *IEEE Signal Processing Letters*, Vol. 20(7), pp. 701–704, 2013.
12. <http://tris.highwaysengland.co.uk/detail/trafficflowdata>, last access on April 6th, 2017