



**HAL**  
open science

## Toward a cost-effective motorway traffic state estimation from sparse speed and GPS data

Zied Bouyahia, Hedi Haddad, Stéphane Derrode, Wojciech Pieczynski

### ► To cite this version:

Zied Bouyahia, Hedi Haddad, Stéphane Derrode, Wojciech Pieczynski. Toward a cost-effective motorway traffic state estimation from sparse speed and GPS data. *IEEE Access*, 2021, 9, pp.44631 - 44646. 10.1109/access.2021.3066422 . hal-03181894

**HAL Id: hal-03181894**

**<https://hal.science/hal-03181894>**

Submitted on 26 Mar 2021

**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.



Distributed under a Creative Commons Attribution 4.0 International License

Received February 15, 2021, accepted March 12, 2021, date of publication March 17, 2021, date of current version March 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3066422

# Toward a Cost-Effective Motorway Traffic State Estimation From Sparse Speed and GPS Data

ZIED BOUYAHIA<sup>1</sup>, HEDI HADDAD<sup>1</sup>, STÉPHANE DERRODE<sup>2</sup>,  
AND WOJCIECH PIECZYNSKI<sup>3</sup>

<sup>1</sup>College of Arts and Applied Sciences, Dhofar University, Salalah 211, Oman

<sup>2</sup>University of Lyon, CNRS, Central School of Lyon, LIRIS, CNRS UMR 5205, 69134 Ecully, France

<sup>3</sup>Polytechnic Institute of Paris, Telecom SudParis, SAMOVAR, CNRS UMR 5157, 91000 Évry, France

Corresponding author: Zied Bouyahia (zbouyahia@du.edu.om)

This work was supported by the Research Council (TRC) of the Sultanate of Oman through the Block Funding Program, TRC Block Funding Agreement, under Grant BFP/RGP/ICT/18/119.

**ABSTRACT** In this paper, we propose a new data-driven traffic state estimation model that estimates traffic flow based on average speed data only. The model is devised to implement a cost-effective framework that aggregates heterogeneous sources of vehicles' GPS and speed measurements to infer traffic flow using a novel triplet system called Conditionally Gaussian Observed Markov Fuzzy Switching Systems (CGOMFSM). Unlike its hard counterpart, CGOMFSM allows for a transient and gradual representation of traffic state transition and hence improves the estimation performance using a tractable scheme. The potential of the proposed model is illustrated through an application to the problem of traffic incident detection, particularly sporadic traffic congestion caused by unexpected road conditions. The performance of the proposed model is assessed using real traffic datasets from England highways. A simulation of traffic in the city of Salalah in Oman was conducted to evaluate the efficacy of the CGOMFSM-based traffic estimation and incident detection schemes with different penetration rates.

**INDEX TERMS** Traffic state estimation, intelligent transportation systems, traffic monitoring, conditionally Gaussian observed Markov fuzzy switching systems.

## I. INTRODUCTION

Traffic state estimation is of a paramount importance for the implementation of Intelligent Transportation Systems (ITSs) in smart cities of the future [1]. Traffic state estimation has been traditionally achieved using fixed sensors, such as inductive loops, radars and cameras, which allow for accurate observation and estimation of speed, flow and occupancy. The main disadvantage of these sensors is their high installation and maintenance costs, which doesn't allow them to be deployed at a large scale. They are instead deployed on major and selected road segments only, which limits their spatial coverage for traffic state estimation. Recent advances in Information and Communication Technologies, Ubiquitous Computing, Vehicular Networks, and Connected Vehicles have opened up new opportunities for smart cities to develop advanced ITSs that combine traditional fixed sensors with new ubiquitous sensing devices for city-wide traffic state

estimation, such as Cellular phones [2], GPS-equipped cars and devices [3], crowd-sourcing [4], VANETs [5], Connected Vehicles [6]–[9], and recently Vehicle–Infrastructure Integration (VII) technologies [10], [11]. While all these new sensing technologies address the coverage limitations of fixed sensors and allow for city-wide traffic state estimation, some of them present the other interesting advantage of providing infrastructure-less and cost-efficient solutions for traffic state estimation, making them particularly interesting for cities in developing countries that do not have road and sensing infrastructure of traffic monitoring. Probe Vehicles (PVs) [12]–[14] and Cell Phones [15] have been widely used for the implementation of infrastructure-less and cost-effective traffic state estimation solutions, mainly for travel time and signalized intersections' queue length estimation [14], [16]. But beside the penetration rate issue, these technologies cannot be used to directly estimate traffic flow. To address this problem, some research works tried to extend PVs with other data sources in order to collect traffic flow data, such as equipping PVs with spacing measurement equipment to count

The associate editor coordinating the review of this manuscript and approving it for publication was Michail Makridis<sup>1</sup>.

surrounding vehicles [17]–[20] and combining PV-data with traditional fixed sensory data (inductive loops, etc.) [21]–[23]. Nevertheless, traffic flow estimation based on mobile probe data only remains a challenging research problem that has been explored to a relatively limited extent in the state of the art. Recent works have used the fundamental diagram (FD) to estimate traffic flow or volume from PVs trajectory data [24]–[26], but these FD-based approaches need large-data calibration effort for every individual road. Other works have used Kalman filter [27], [28], Naive Bayes and shock-wave models [29], [30] to estimate traffic flow from PVs but are generally applicable only at signalized link approaches and intersections.

In this paper, we propose a new data-driven traffic state estimation approach that relies on infrastructure-less smart city sensing technologies and can therefore be used to support the implementation of traffic management services of smart cities in developing countries. We propose a model that can be used for traffic flow estimation on motorway segments based only on aggregated speed data collected from IoT-probe sources moving on these road segments. IoT-probe sources can be any individual connected devices such as private cars and drivers' smartphones, and not only Floating Cars or dedicated PVs. The model can be used to implement completely cost-less traffic control systems without installing or maintaining any road infrastructure. The increasing amount of vehicles and connected mobile devices in developing countries allows for maintaining a minimum penetration rate required for accurate traffic flow estimations. In this paper we only present the new proposed data-driven model of traffic state estimation [31] and we assume that the aggregated speed data are given as input. Issues related to the development of sensing architectures relying on aggregated GPS-Data sources and their use for the implementation of practical traffic state estimation solutions in the context of smart cities are out the scope of this paper and have been addressed in many previous research works, such as [32]–[34] and [3]. More recent insight about the use of GPS data for city-wide traffic state estimation and monitoring in general can be found in [35], [36].

The contributions of this work are as follows:

- A new traffic state estimation algorithm based on the Conditionally Gaussian Observed Markov Fuzzy Switching model (CGOMFSM) and which relies on an explicit representation of the dependence between the traffic flow, speed and state, represented as stochastic processes. The proposed algorithm is used to estimate traffic flow based on average speed data only.
- A new parameter estimation algorithm based on the CGOMFSM model.
- An application of the proposed algorithm to the problem of abnormal traffic event (congestion) detection.

The remainder of this paper is organized as follows: Section II presents an overview of the techniques and approaches used for infrastructure-less traffic flow estima-

tion in the state of the art, and depicts the proposed estimation framework. Section III details the triplet model used to represent the traffic state evolution over time along with the underlying traffic state estimation algorithm. Section IV presents the derivation of the fuzzy parametrization of the triplet model. Section V details the parameter estimation scheme adapted for the proposed fuzzy parametrization and presents an application of the scheme to the problem of sporadic traffic incident detection. Section VI introduces an extension of the proposed estimation algorithm for a one-step ahead prediction of traffic state. In Section VII, we report the results of the experimental study using both real datasets and simulated traffic data for traffic event detection applications.

## II. STATE-OF-THE-ART AND PROPOSED FRAMEWORK

### A. REVIEW OF UBIQUITOUS AND COST-EFFECTIVE TRAFFIC FLOW ESTIMATION

This section reviews the main state-of-the-art works related to the problem of ubiquitous and infrastructure-less traffic flow estimation, in terms of both sensing technologies and estimation models. Traffic flow estimation works using VANETs, Connected Vehicles and Vehicle-Infrastructure Integration are not considered given that these technologies require certain infrastructure to be deployed either on vehicles or on road sides or on both. An exhaustive review of traffic state estimation technologies and models can be found in [31].

GPS-data have been widely used for traffic state estimation purposes, collected either by PVs [3], [36] or smartphones [32], [34]. PVs are recognised as cost-efficient sources of traffic data, especially for traffic speed and travel time estimation, but their use for traffic flow or traffic density estimation is challenging. Some researchers equipped PVs with range measurement sensors, such as cameras, to have a local flow estimation of surrounding vehicles [18], [37], while other have combined PV data with data collected on the ground, such as loop detectors [22], [23]. There is a relatively limited number of works on estimating traffic flow or traffic density from PV data only. The Fundamental Diagram (FD) has been used to estimate traffic flow or volume from PV trajectory data [24]–[26] on highways and major arteries. The major drawback of the FD method is that it requires calibration for each individual road with a sufficiently large amount of traffic data [21]. Aside the well-studied issue of penetration rate [38], [39], the main problem of traditional PV-based methods concerns the nature of fleets—mostly commercial such as taxis and buses— which biases the estimation given that these types of fleets have their own specific driving and spatio-temporal mobility patterns that are different from private vehicles.

Cellular systems [40] have also been used to provide alternative methods to the cost and coverage limitations associated with infrastructure-based solutions, and some works have been proposed to estimate volumes of vehicles from anonymous cellular phone call data [41], [42]. These systems cannot generally provide fine-grained traffic volume

estimates and require long processing time which make them unsuitable for real-time estimation [15], [43].

Smartphone applications have also been widely used for traffic monitoring in general, including traffic state and congestion estimation, commonly deployed as a crowd-sourcing approach [4] where traffic data are collected from mobile applications that run on road users' mobile phones. By combining the computing capabilities and variety of sensors (GPS, accelerometers, microphones, cameras, etc.) of the smartphones, different types of traffic data can be collected and processed on the phone side before being sent to the server side for further processing. Several research and commercial projects have been proposed (a good review can be found in [44]). Other works have also explored the use of GPS data streams [33], [34], [45], [46]. Smartphone applications and GSP data streams have been mainly used for average speed and travel time estimation, but, to the best of our knowledge, not for traffic flow estimation.

With respect to estimation approaches, most of the existing works collect traffic volume data from fixed-location sensors. Only few works have addressed the problem of estimating traffic volume from mobile data sources. Different techniques have been used, such as statistical models [29], [30], [47], combination of Shockwave theory and Maximum likelihood estimation (MLE) [26], [48]–[50], Kalman filter [27], [28], [51], compressive sensing techniques [52], spatio-temporal correlation data-driven models [23], [47], [53], in addition to Flow Diagram-based models where travel speeds from probe vehicle data for each road are estimated and then converted to traffic volumes by exploiting the relationship between travel speeds and traffic volumes [54], [55]. A more complete review can be found in [31]. To the best of our knowledge, none of the previous research works has explored the use of Conditionally Gaussian Observed Markov Switching Model for traffic state estimation in general, and particularly for traffic flow estimation from aggregated mobile speed measures.

### B. PROPOSED FRAMEWORK

There exist many smart devices that allow for a quite accurate and timely measure of vehicles' speed. Smart speedometers are available at reasonable prices and operate on almost any type of vehicle. There is also a myriad of mobile applications that can perform as good as smart tachometers. In addition to the radar guns and speed traps used by law enforcement officers and mobile patrols, the number of smart vehicles with embedded smart speed measuring devices has been substantially increasing over the last few years.

Our framework exploits the speed data collected from sensing devices that correspond to a specific road segment within a certain time interval  $I_n = ]t_{n-1}, t_n]$ . We assume that traffic state is estimated in regular time intervals  $I_n = ]t_{n-1}, t_n]$ ,  $0 < n$ . The collected speed data are used to estimate the average velocity denoted hereafter as  $v_n^s$ , where  $s$  refers to the road segment and  $n$  to the period  $I_n$ . For GPS and speed data,

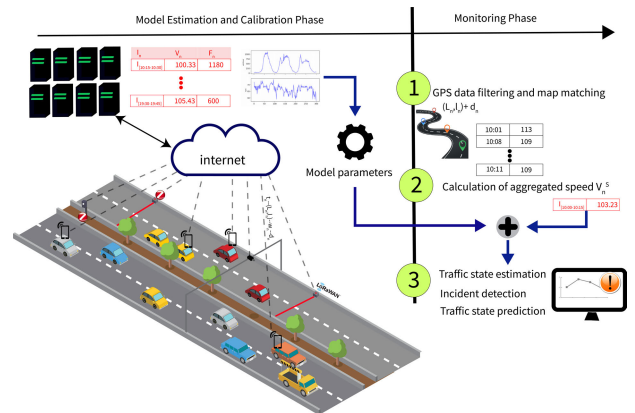


FIGURE 1. Architecture of the proposed framework.

we consider the following model:

$$\langle t_n, L_n, l_n, \omega_n, d_n \rangle,$$

where  $t_n$  refers to the timestamp,  $L_n$  to the longitude,  $l_n$  to the latitude,  $\omega_n$  to the speed and  $d_n$  to the direction of the vehicle movement during  $I_n$ . The GPS coordinates of the vehicle are mapped to the road segment using the map matching procedure described in [56]. It is worth mentioning that the issues related to network connectivity such as latency and those related to power consumption are not considered in the scope of this research. We mainly focus on the data-driven model and its ability to infer complete traffic state from a single traffic variable.

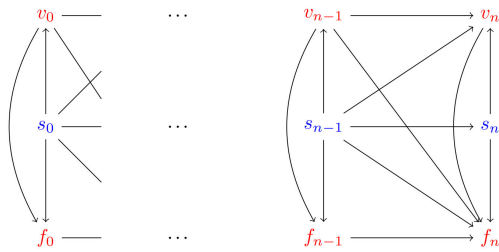
Once the location of a vehicle is matched to the road segment of interest, the vehicle speed is recorded throughout the journey. As soon as the period  $I_n$  is elapsed, the vehicle speed data are averaged to  $v_n$ . Whenever possible, traffic flow data during  $I_n$ , denoted as  $f_n$ , are recorded as a training set for the estimation module and used afterwards to infer traffic state using speed data only, as will be presented later. For the sake of simplification, we assume that connectivity related issues incurred by transmission lags are not considered. Hence, we assume that the time intervals  $I_n$  are sufficiently large so that communication delays do not impact the accuracy of the model. We distinguish between two phases (Figure 1):

- **Training phase:** During this phase, both flow and speed data are recorded on a regular basis spanning a certain period of time  $T$ . The corresponding  $N$  time intervals  $I_n$  are such that

$$\bigcup_{n=1}^N I_n = T.$$

The recorded speed data  $v_1^N = (v_1, \dots, v_N)$  and the flow data  $f_1^N = (f_1, \dots, f_N)$  are used to infer a set of parameters corresponding to the traffic condition. The duration  $T$  of this phase is determined by the convergence of the estimated state towards the observed one.

- **Monitoring phase:** The trained model is used during this phase to estimate traffic state, i.e. road occupancy



**FIGURE 2.** Dependence graph of the considered CGOMFSM during the training phase. Red labels correspond to observed variables while blue labels refer to hidden ones.

and traffic flow and fluency from the observed speed  $v_i$  only.

We propose to model the depicted framework by three stochastic processes  $F$ ,  $V$  and  $S$ , where  $F$  represents the flow variation,  $V$  the speed variation and  $S$  the traffic fluency. The difference between two consecutive switches  $S_n$  and  $S_{n+1}$  depicts the change of traffic condition. Let us note that the training phase can be conducted using a sample-based approach. Since we focus on motorways, we can select road segment samples based on common characteristics (basic section, on-ramp/off-ramp, speed limit, etc.). Ideally, the trained model could be directly transposed to road sections that share the same features.

The underlying dependency between traffic speed and flow, as well as the transient transition between traffic conditions over time, suggest the use of a model that (i) supports discrete-time discontinuities in a continuous fashion and (ii) represents the dependencies (supposed here linear) between the stochastic processes  $F_1^N$  and  $V_1^N$ .

To achieve a time-dependent representation of traffic state, we introduce a latent process denoted as  $S_1^N$  and referring to the traffic switch over time. Hence, the random variable  $s_n$  represents the traffic state during the time interval  $I_n$  during which the speed measure  $v_n$  has been recorded.

The problem can be formulated as follows: given  $P$  measures of average speed  $v_1^P$  and traffic flow  $f_1^P$  at a given road section spanning a time interval  $T = \bigcup_{n=0}^P I_n$ , where  $I_n$  are time periods of equal length, the goal is to determine a model that can be used to (i) explicitly represent the dependence between the stochastic processes  $F$  and  $V$  using an auxiliary process  $S$  and (ii) construct a tractable procedure to estimate the traffic state from the speed data  $V$  observed during a monitoring time interval  $\bigcup_{n=P+1}^M I_n$ ,  $M > P$ . To this purpose, we propose to exploit a recent fuzzy linear model called Conditionally Gaussian Observed Markov Fuzzy Switching Model (CGOMFSM).

### III. TRAFFIC STATE MODELING AND ESTIMATION USING CGOMFSM

First, let us briefly present the model we will deal with. The CGOMFSM is an extension of the classical Conditionally Gaussian Markov Switching Model (CGMSM) defined as

follows. Let us consider three discrete-time stochastic processes  $F_1^N$ ,  $V_1^N$ , and  $S_1^N$ , where, for each  $n \in [1, N]$ ,  $F_n$  is a (hidden) real-valued random variable,  $V_n$  is an (observed) real-valued random one and  $S_n$  is a (hidden) discrete-valued random variable with two states:  $S_n \in \Omega = \{0, 1\}$ .

*Definition 1:* Let us set  $Z_n = (F_n, V_n)^T$ ,  $T_n = (F_n, S_n, V_n)^T$  and assume the following:

- 1)  $T_1^N$  is Markov;
- 2)  $p(s_{n+1} | t_n) = p(s_{n+1} | s_n)$ , which implies that  $S_1^N$  is Markov;
- 3)  $Z_1^N = (Z_1, \dots, Z_N)$  is Gaussian conditionally on  $S_1^N$ .

The CGMSM is defined by  $p(t_1)$ , transitions  $p(s_{n+1} | s_n)$ , and the inductive relationship:  $\forall n = 1, \dots, N - 1$

$$Z_{n+1} = A_{n+1}(S_n^{n+1})Z_n + B_{n+1}(S_n^{n+1})W_{n+1} + N_{n+1}(S_n^{n+1}), \quad (1)$$

where

- $W_n = (U_n, R_n)^T$  with  $U_1, R_1, \dots, U_N, R_N$  Gaussian zero-mean independent vectors with identity covariance matrices;
- Matrices  $A_{n+1}(s_n^{n+1})$  and  $B_{n+1}(s_n^{n+1})$ :

$$A_{n+1}(s_n^{n+1}) = \begin{bmatrix} a_{n+1}^1(s_n^{n+1}) & a_{n+1}^2(s_n^{n+1}) \\ a_{n+1}^3(s_n^{n+1}) & a_{n+1}^4(s_n^{n+1}) \end{bmatrix},$$

$$B_{n+1}(s_n^{n+1}) = \begin{bmatrix} b_{n+1}^1(s_n^{n+1}) & b_{n+1}^2(s_n^{n+1}) \\ b_{n+1}^3(s_n^{n+1}) & b_{n+1}^4(s_n^{n+1}) \end{bmatrix};$$

- Means  $N_{n+1}(s_n^{n+1}) = (N_{n+1}^F(s_n^{n+1}), N_{n+1}^V(s_n^{n+1}))^T$  are given by

$$N_{n+1}(s_n^{n+1}) = M_{n+1}(s_{n+1}) - A_{n+1}(s_n^{n+1})M_n(s_n),$$

with

$$M_n(s_n) = \mathbb{E} \left[ \begin{pmatrix} F_n \\ V_n \end{pmatrix} \middle| s_n \right] = \begin{bmatrix} M_n^F(s_n) \\ M_n^V(s_n) \end{bmatrix}. \quad (2)$$

Recursive filtering is not tractable in the general CGMSM setting, but becomes tractable under the condition that, for each  $s_n^{n+1} \in \Omega^2$ ,

$$a_{n+1}^3(s_n^{n+1}) = 0. \quad (3)$$

This particular CGMSM, called ‘‘Conditionally Gaussian Observed Markov Switching Model’’ (CGOMSM) [57] allows for recursive optimal filtering even with a switching setup [58].

A fuzzy extension of the CGOMSM has been proposed and studied in [59] in which the switch process  $S_1^N$  is no longer assumed discrete but takes its values in the interval  $[0, 1]$  instead. Hence, the distribution of each  $S_n$  is defined by density  $h_n : [0, 1] \rightarrow \mathbb{R}$  w.r.t. the measure  $\nu = \delta_0 + \delta_1 + \mu_{]0,1[}$ , with  $\delta_0, \delta_1$  two Dirac’s distributions on 0, 1 respectively, and  $\mu_{]0,1[}$  a Lebesgue’s measure on  $]0, 1[$ . Thus, we have:

$$\int_0^1 p(s_n) d\nu(s_n) = p(0) + p(1) + \int_0^1 p(s_n) ds_n = 1, \quad (4)$$

and

$$\begin{aligned}
 p(s_{n+1}) &= \int_0^1 p(s_n)p(s_{n+1}|s_n)dv(s_n) \\
 &= p(s_n = 0)p(s_{n+1}|s_n = 0) \\
 &\quad + p(s_n = 1)p(s_{n+1}|s_n = 1) \\
 &\quad + \int_0^1 p(s_n)p(s_{n+1}|s_n)ds_n. \tag{5}
 \end{aligned}$$

The distribution of the fuzzy Markov chain  $S_1^N = (S_1, \dots, S_N)$  is defined by density  $p(s_1)$  and the conditional densities  $p(s_{n+1}|s_n)$ . All of them are densities on  $\Omega = [0, 1]$  w.r.t. the measure  $\nu = \delta_0 + \delta_1 + \mu_{]0,1[}$ . Figure 2 shows the dependence graph between the model processes.

**A. TRAFFIC STATE MODELING USING CGOMFSM**

We assume that the average speed data are measured and recorded on a regular basis, while the total carriageway flow data are not necessarily observable. The main rationale for considering such assumption is that it is easy to aggregate traffic speed from sampled measures while it is more difficult to determine the traffic volume unless necessary nay intrusive equipment has been installed.

The proposed traffic modeling framework is composed of two main stages:

- Stage 1: Traffic model parameter estimation. At this level, historic traffic data are utilized to infer the CGOMFSM model parameters and fit the fuzzy parametrization to a particular dataset. Let us denote by  $\{1, \dots, N\}$  the measurement time points for which both traffic volume and average speed data are available.  $Z_1^N = (F_1^N, V_1^N)$  data are used to estimate the CGOMFSM parameters and the associated switch process  $S_1^N$ .
- Stage 2: Traffic state estimation. For any time point  $n > N$ , only speed data are recorded from a variable number of road users. The aggregated speed data  $V_n$  are hence used to extrapolate the traffic volume  $F_n|V_1^n$ , the current switch  $S_n|V_1^n$  using a recursive filtering scheme and to predict the next measurement  $V_{n+1}|V_1^n, S_1^n$ . Within the CGOMFSM framework, let us denote by  $\Gamma_{Z_n, Z_{n+1}}(s_n^{n+1})$

the covariance matrix of the vector  $\begin{bmatrix} Z_n \\ Z_{n+1} \end{bmatrix}$ .

$$\Gamma_{Z_{n+1}}(s_n^{n+1}) = \begin{bmatrix} \Gamma(s_n^{n+1}) & \Sigma(s_n^{n+1}) \\ \Sigma(s_n^{n+1})^\top & \Gamma(s_n^{n+1}) \end{bmatrix}, \tag{6}$$

where  $\Gamma(s_n^{n+1}) = \begin{bmatrix} \Gamma_V(s_n^{n+1}) & b(s_n^{n+1}) \\ b(s_n^{n+1})^\top & \Gamma_F(s_n^{n+1}) \end{bmatrix}$

and  $\Sigma(s_n^{n+1}) = \begin{bmatrix} a(s_n^{n+1}) & d(s_n^{n+1}) \\ e(s_n^{n+1}) & c(s_n^{n+1}) \end{bmatrix}$

**B. TRAFFIC STATE ESTIMATION**

As mentioned earlier, the traffic state is assumed to be modeled by a continuous state-space  $\Omega = [0, 1]$  in which

the boundary conditions correspond to 0 for unoccupied road and to 1 for extremely congested. Depending on the estimated switch  $s_{n+1}$  (from the observed  $v_{n+1}$ ), and  $s_n$ , the unlikely transitions are detected. Let us denote by  $\Delta_n^{n+1} = |\mathbb{E}[s_{n+1}|v_1^n] - \mathbb{E}[s_n|v_1^n]|$  the random variable corresponding to the switch variation during the time interval  $I_n$ . The higher  $\Delta_n^{n+1}$ , the more likely an incident could have occurred. Normally, traffic conditions transition gradually over time unless a sporadic or routine incident has occurred and spurred a relatively abrupt switch. Using the assumption of CGOMFSM, the computation of the filter becomes tractable using the following equations:

$$\begin{aligned}
 p(s_n^{n+1} | v_1^{n+1}) &= \frac{p(s_n^{n+1}, v_{n+1} | v_1^n)}{p(v_{n+1} | v_1^n)} \\
 &= p(s_n | v_1^n) \frac{p(s_{n+1}, v_{n+1} | s_n, f_n, v_n)}{p(v_{n+1} | v_1^n)} \\
 &= p(s_n | v_1^n) p(s_{n+1} | s_n) \frac{p(v_{n+1} | s_n^{n+1}, v_n)}{p(v_{n+1} | v_1^n)}, \tag{7}
 \end{aligned}$$

with

$$\begin{aligned}
 p(v_{n+1} | v_1^n) &= \int_0^1 \int_0^1 p(s_n | v_1^n) p(s_{n+1} | s_n) \\
 &\quad p(v_{n+1} | s_n^{n+1}, v_n) d(v(s_n^{n+1})), \tag{8}
 \end{aligned}$$

and where

- $d(v(s_n^{n+1})) = d(v(s_n) \otimes v(s_{n+1}))$ ;
- $p(s_{n+1} | s_n)$  is given by the fuzzy parametrization (see section IV);
- according to the Gaussian assumptions and (1), (3),  $p(v_{n+1} | s_n^{n+1}, v_n)$  is Gaussian with mean  $a_{n+1}^4(s_n^{n+1})v_n$  and variance  $b_{n+1}^3(s_n^{n+1})b_{n+1}^3(s_n^{n+1})^\top + b_{n+1}^4(s_n^{n+1})b_{n+1}^4(s_n^{n+1})^\top$ .

Hence, the next switch is estimated using the equation:

$$\mathbb{E}[s_{n+1} | v_1^{n+1}] = \int_0^1 s_{n+1} p(s_{n+1} | v_1^{n+1}) dv(s_{n+1}),$$

where

$$p(s_{n+1} | v_1^{n+1}) = \frac{1}{p(v_{n+1} | v_1^n)} \int_0^1 p(s_n^{n+1}, v_{n+1} | v_1^n) dv(s_n) \tag{9}$$

Finally, the traffic state estimation algorithm runs as follows:

- Compute  $p(s_{n+1} | v_1^{n+1})$  using equations (7) and (8).
- Compute the estimated traffic state at time point  $n + 1$  using:

$$\begin{aligned}
 \mathbb{E}[Z_{n+1} | s_n^{n+1}, v_1^n] &= A_{n+1}(s_n^{n+1}) \begin{bmatrix} \mathbb{E}[F_n | s_n, v_1^n] \\ v_n \end{bmatrix} \\
 &\quad + N_{n+1}(s_n^{n+1}),
 \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E} \left[ \mathbf{Z}_{n+1} \mathbf{Z}_{n+1}^T \mid s_n^{n+1}, \mathbf{v}_1^n \right] \\ &= \mathbf{B}_{n+1}(s_n^{n+1}) \mathbf{B}_{n+1}(s_n^{n+1})^T \\ &+ \mathbf{A}_{n+1}(s_n^{n+1}) \begin{bmatrix} \mathbb{E} [F_n \mid s_n, \mathbf{v}_1^n] \\ v_n \end{bmatrix} \mathbf{A}_{n+1}(s_n^{n+1})^T \\ & \begin{bmatrix} \mathbb{E} [F_n \mid s_n, \mathbf{v}_1^n] \\ v_n \end{bmatrix}^T, \\ \text{Var} \left[ \mathbf{Z}_{n+1} \mid s_n^{n+1}, \mathbf{v}_1^n \right] \\ &= \begin{bmatrix} \alpha_{n+1}(s_n^{n+1}) & \beta_{n+1}(s_n^{n+1}) \\ \chi_{n+1}(s_n^{n+1}) & \delta_{n+1}(s_n^{n+1}) \end{bmatrix} \\ &= \mathbb{E} \left[ \mathbf{Z}_{n+1} \mathbf{Z}_{n+1}^T \mid s_n^{n+1}, \mathbf{v}_1^n \right] \\ &- \mathbb{E} \left[ \mathbf{Z}_{n+1} \mid s_n^{n+1}, \mathbf{v}_1^n \right] \mathbb{E} \left[ \mathbf{Z}_{n+1}^T \mid s_n^{n+1}, \mathbf{v}_1^n \right]. \end{aligned}$$

- Estimate the traffic flow  $F_{n+1}$  using:

$$\begin{aligned} & \mathbb{E} \left[ F_{n+1} \mid s_n^{n+1}, \mathbf{v}_1^{n+1} \right] \\ &= \mathbb{E} \left[ F_{n+1} \mid s_n^{n+1}, \mathbf{v}_1^n \right] \\ &+ \frac{\beta_{n+1}(s_n^{n+1})}{\delta_{n+1}(s_n^{n+1})} \left( v_{n+1} - \mathbb{E} \left[ V_{n+1} \mid s_n^{n+1}, \mathbf{v}_1^n \right] \right), \\ & \mathbb{E} \left[ F_{n+1} \mid s_{n+1}, \mathbf{v}_1^{n+1} \right] \\ &= \int_0^1 p \left( s_n \mid s_{n+1}, \mathbf{v}_1^{n+1} \right) \\ & \mathbb{E} \left[ F_{n+1} \mid s_n^{n+1}, \mathbf{v}_1^{n+1} \right] dv(s_n), \end{aligned} \tag{10}$$

- Finally, the switch gradient is calculated using  $p(s_n \mid \mathbf{v}_1^n)$  and  $p(s_{n+1} \mid \mathbf{v}_1^{n+1})$  using the following equation:

$$\begin{aligned} \Delta_n^{n+1} = & \left| \int \int_0^1 s_n p(s_n \mid \mathbf{v}_1^n) \right. \\ & \left. - s_{n+1} p(s_{n+1} \mid \mathbf{v}_1^{n+1}) \right| dv(s_n) dv(s_{n+1}) \end{aligned}$$

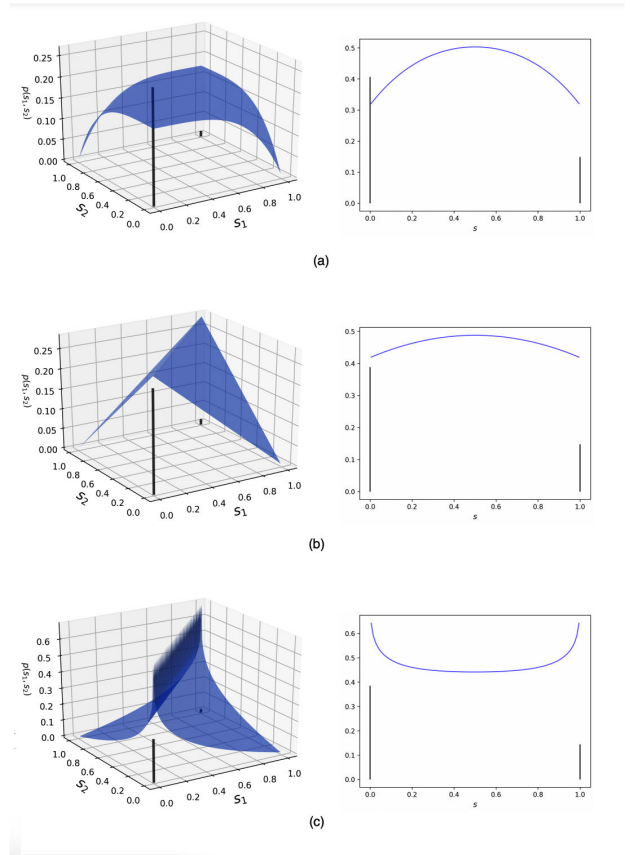
*Remark:* Integrating with respect to  $v$  is not always possible in a closed form. Hence, we use the following approximation:

$$\int_0^1 h(t) dt \approx h(0) + h(1) + \frac{1}{L} \sum_{i=1}^{L-1} h\left(\frac{i}{L}\right), \tag{11}$$

where  $L$  is the number of fuzzy levels. Since the set of integration here is compact, the approximation (11) remains tractable.

#### IV. PARAMETRIZATION OF THE FUZZY MARKOV CHAIN

In this section, we outline a possible parametrization of the density  $p(s_n^{n+1})$  of the Markov chain  $S_1^N$  suited to our targeted application. As we assume the model to be stationary,  $p(s_n^{n+1})$  is time-independent and hence  $p(s_n^{n+1}) \sim p(s_1^2)$ .



**FIGURE 3.** Three examples of joint density  $p(s_1^2)$  (left) as defined in eq. (12), with corresponding  $p(s_1)$  marginal (right) as defined in eq. (13). (a)  $m = 4$ ,  $\alpha_0 = 0.265$ ,  $\alpha_1 = 0.007$ ,  $\beta = 0$  and  $\eta = 0.175$ , (b)  $m = 1$ ,  $\alpha_0 = 0.248$ ,  $\alpha_1 = 0.008$ ,  $\beta = 0$  and  $\eta = 0.278$  and (c)  $m = 0.25$ ,  $\alpha_0 = 0.248$ ,  $\alpha_1 = 0.007$ ,  $\beta = 0$  and  $\eta = 0.684$ .

The density of  $P(s_1^2)$  w.r.t.  $v \otimes v$ —where  $v = \delta_0 + \delta_1 + \mu_{]0,1[}$ —is assumed of the following shape:

$$p(s_1^2) = \begin{cases} \alpha_0 & \text{if } s_1 = s_2 = 0, \\ \alpha_1 & \text{if } s_1 = s_2 = 1, \\ \beta & \text{if } (s_1, s_2) \in \{(0, 1), (1, 0)\}, \\ \eta (1 - |s_1 - s_2|)^m & \text{otherwise,} \end{cases} \tag{12}$$

with  $m > 0$ , and  $\int \int_0^1 p(s_1, s_2) dv(s_1) dv(s_2) = 1$ . This parametrization is defined by 5 parameters ( $m, \alpha_0, \alpha_1, \beta$  and  $\eta$ ). It is an extension of the parametrization studied in [59] and used for the estimation of buildings power consumption from outdoor temperatures.

Let us interpret this parametrization with respect to the targeted application. The probabilities that the traffic remains in the same boundary state are denoted by  $\alpha_0$  for very low traffic volume and by  $\alpha_1$  for traffic peaks. Another quite realistic assumption regarding the traffic dynamics behaviour with respect to the boundary conditions is the following: a direct transition from unoccupied road (state 0) to extremely congested traffic (state 1), and reversely, is fairly improbable when the number of discrete fuzzy levels  $L$  is sufficiently high

(typically when  $L > 5$ ). Therefore, we will set  $\beta = 0$  in the experimental part, yet this parameter will not be discarded for the sake of generality and for possible other application usage.

Parameter  $m$  allows to introduce very different behaviors for the Markov chain that governs the state of traffic. Figure 3 plots three possible shapes of the joint density  $p(s_1^2)$  for different values of  $m$ . When  $m = 1$ , the joint density is a piecewise linear shape and is proportional to the difference between  $s_1$  and  $s_2$ . When  $m$  is large,  $p(s_1^2)$  is almost constant within the diagonal. Setting  $m$  below 1 imposes a low probability on distant  $s_1^2$ , and the closer  $s_1$  to  $s_2$ , the higher  $p(s_1^2)$ .

By marginalizing (12), see calculations reported in Appendix, the density  $p(s_1)$  is given by

$$p(s_1) = \begin{cases} \alpha_0 + \beta + \eta \frac{m}{m+1} & \text{if } s_1 = 0, \\ \beta + \alpha_1 + \eta \frac{m}{m+1} & \text{if } s_1 = 1, \\ \eta \left( 3 - (s_1^m + (1 - s_1)^m) \right) & \\ - \frac{s_1^{m+1} + (1 - s_1)^{m+1}}{m+1} & \text{if } s_1 \in ]0, 1[. \end{cases} \quad (13)$$

Still from the appendix, we have

$$\eta = \frac{(m+1)(m+2)}{5m^2 + 11m} (1 - \alpha_0 - \alpha_1 - 2\beta) \quad (14)$$

Hence the joint a priori fuzzy density is only parametrized by four parameters  $m, \alpha_0, \alpha_1$  and  $\beta$ .

The density  $p(s_2 | s_1)$  is the ratio between the joint density (12) and the marginal density (13). We have to distinguish between different cases, according to the value of  $s_1$ :

$$p(s_2 | s_1 = 0) = \begin{cases} \frac{\alpha_0}{D_0} & \text{if } s_2 = 0, \\ \frac{\beta}{D_0} & \text{if } s_2 = 1, \\ \frac{\eta(1 - s_2^m)}{D_0} & \text{if } s_2 \in ]0, 1[. \end{cases} \quad (15)$$

$$p(s_2 | s_1 = 1) = \begin{cases} \frac{\beta}{D_1} & \text{if } s_2 = 0, \\ \frac{\alpha_1}{D_1} & \text{if } s_2 = 1, \\ \frac{\eta s_2^m}{D_1} & \text{if } s_2 \in ]0, 1[. \end{cases} \quad (16)$$

$$p(s_2 | s_1 \in ]0, 1[) = \begin{cases} \frac{(1 - s_1^m)}{D_{s_1}} & \text{if } s_2 = 0, \\ \frac{s_1^m}{D_{s_1}} & \text{if } s_2 = 1, \\ \frac{1 - |s_1 - s_2|^m}{D_{s_1}} & \text{if } s_2 \in ]0, 1[. \end{cases} \quad (17)$$

where we set

- $D_0 = \alpha_0 + \beta + \eta \frac{m}{m+1}$ ,
- $D_1 = \alpha_1 + \beta + \eta \frac{m}{m+1}$ , and
- $D_{s_1} = \eta \left( 3 - (s_1^m + (1 - s_1)^m) - \frac{s_1^{m+1} + (1 - s_1)^{m+1}}{m+1} \right)$ .

## V. PARAMETER ESTIMATION AND EVENT DETECTION

In this section we first explain how the parameters of the model can be estimated from training samples and then we present how the parameter estimation algorithm can be used to detect sporadic events from traffic data. The proposed parameter estimation algorithm is devised based on an iterative scheme in which each step  $q$  consists in estimating both the joint a priori density and the means and covariances of the CGOMFSM.

### A. PARAMETERS OF THE JOINT A PRIORI DENSITY

As specified in eq. (12), the density  $p(s_n, s_{n+1})$  is defined by four parameters  $\beta, \alpha_0, \alpha_1$  and  $m$ , recalling that the parameter  $\eta$  is deduced from them using (14). We consider a fuzzy parametrization preset to a fixed  $m$ . To fit the fuzzy parametrization to the data of interest, we use the following sampling-based procedure:

- 1) Simulate  $Q$  realizations of the process  $S_1^N$  conditionally to  $Z_1^N$  using the a posteriori distribution, i.e. using  $p(s_{n+1} | s_n, z_1^N)$ . We denote the  $q^{th}$  realization by  $(s_1^N)^{[q]}$ .
- 2) For each realization  $q \in [1, Q]$ ,
  - count the number of times we observe the sequence  $s_n^{n+1} = (0, 0)$  in  $(s_1^N)^{[q]}$ . Dividing by  $N - 1$  gives an estimation  $\alpha_0^{[q]}$  of  $\alpha_0$ .
  - count the number of times we observe the sequence  $s_n^{n+1} = (1, 1)$  in  $(s_1^N)^{[q]}$ . Dividing by  $N - 1$  gives an estimation  $\alpha_1^{[q]}$  of  $\alpha_1$ .
  - count the number of times we observe the sequence  $s_n^{n+1} \in \{(0, 1), (1, 0)\}$  in  $(s_1^N)^{[q]}$ . Dividing by  $N - 1$  gives an estimation  $\beta^{[q]}$  of  $\beta$ .
- 3) Average the  $Q$  estimations  $\alpha_0^{[q]}, \alpha_1^{[q]}$  and  $\beta^{[q]}$ , to get  $\alpha_0^{q+1}, \alpha_1^{q+1}$  and  $\beta^{q+1}$ . Deduce  $\eta^{q+1}$  using (14).

Figure 4 illustrates the trajectories of Markov chains ( $S$ ) simulated using the a priori distribution  $p(s_2 | s_1)$  and the a posteriori distribution  $p(s_{n+1}, s_n | z_1^N)$ .

### B. ESTIMATION OF THE MEANS AND COVARIANCES

The model parameters, i.e. means vectors and covariance matrices are estimated using a fuzzy C-means procedure. Let  $F_{max} = \max(F_1^N), F_{min} = \min(F_1^N), V_{max} = \max(V_1^N)$  and  $V_{min} = \min(V_1^N)$ .

- 1) The state space is divided into two sub-spaces: stable traffic and unstable traffic. Hence, we make an initial estimation  $\mu_k$  such that  $\forall k \in \{0, \dots, \lfloor \frac{L}{2} \rfloor\}, \mu_k =$

$$\begin{bmatrix} F_{min} + \frac{k}{L+1} (F_{max} - F_{min}) \\ V_{max} - \frac{k}{L+1} (V_{max} - V_{min}) \end{bmatrix}$$

and  $\forall k \in \{\lfloor \frac{L}{2} \rfloor + 1, \dots, L + 2\}, \mu_k =$

$$\begin{bmatrix} F_{max} - \frac{k}{L+1} (F_{max} - F_{min}) \\ V_{min} + \frac{k}{L+1} (V_{max} - V_{min}) \end{bmatrix} \text{ and initialize the}$$



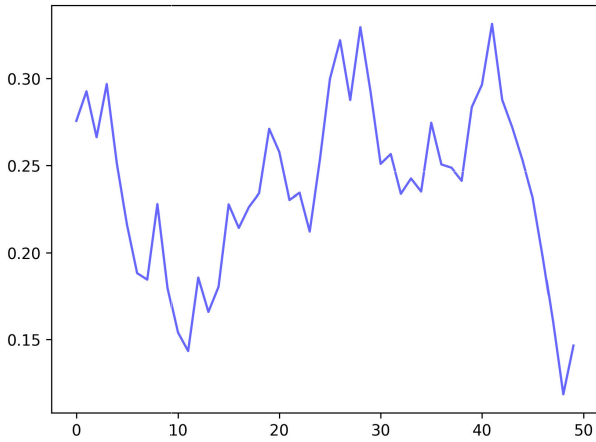


FIGURE 4. Example of trajectories of simulated Markov chains using the a priori fuzzy distribution with parameters  $m = 0.5$ ,  $\alpha_0 = 0.05$ ,  $\alpha_1 = 0.05$ ,  $\beta = 0$ .

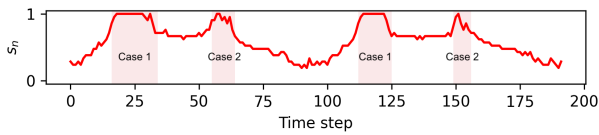


FIGURE 5. Examples of traffic anomalies observed by a steep increase of  $\Delta_n^{n+1}$  (Case 2) and constant high switches (Case 1) (red-shaded areas).

membership matrix  $\kappa^0 = [\kappa_{ij}]$ , such that  $\forall i \in 1, \dots, N$ ,  $\kappa_{ij}^0 = \begin{cases} 0 & \text{if } 0 < j < L + 1 \\ 1 & \text{otherwise} \end{cases}$

2) Calculate the means vectors  $\mu = \begin{bmatrix} \mu_0 \\ \vdots \\ \mu_L \end{bmatrix}$  using:

$$\mu_j = \frac{\sum_{i=1}^N (\kappa_{ij}^q)^{\log(L)} z_i}{\sum_{i=1}^N (\kappa_{ij}^q)^{\log(L)}} \quad (18)$$

3) Update the membership matrix  $\kappa^{q+1}$  using:

$$\kappa_{ij} = \frac{1}{\sum_{k=0}^{L+1} \left( \frac{\|z_i - \mu_j\|}{\|z_i - \mu_k\|} \right)^{\log(L) - 1}} \quad (19)$$

4) If  $\|\kappa^{q+1} - \kappa^q\| \leq \epsilon$ , STOP, else repeat step 2.

### C. ESTIMATION OF THE A POSTERIORI DISTRIBUTIONS $P(s_{n+1} | s_n, z_1^N)$

First, we calculate the a posteriori probabilities  $\psi_i(s_1^2) = P(s_n, s_{n+1} | z_1^N)$  from the membership matrix  $\kappa$  using  $\psi_i(s_1^2) = \kappa_{i, s_1} \kappa_{i+1, s_2}$

For each iteration  $q > 1$ :

- Compute  $p_{s_1^2} = P(s_1^2 | z_1^N) = \frac{1}{N} \sum_{i=1}^N \psi_i(s_1^2)$

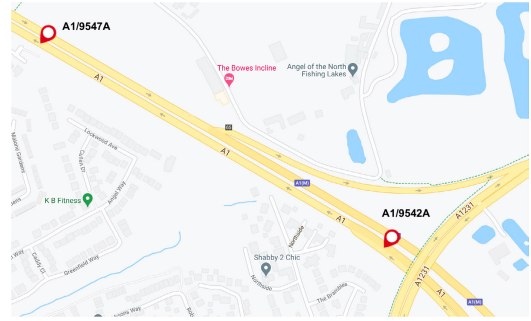


FIGURE 6. Locations of the considered monitoring sites A1/9547A and A1/9542A. Sites are pinpointed in red.

- Compute the model means and covariances using:

$$\mu_{s_1^2} = \frac{\sum_{i=1}^N \psi_i(s_1^2) \begin{bmatrix} z_i \\ z_{i+1} \end{bmatrix}}{\sum_{i=1}^N \psi_i(s_1^2)} \quad (20)$$

and

$$\Gamma_{s_1^2} = \frac{\sum_{i=1}^N \psi_i(s_1^2) \begin{bmatrix} z_i - \mu_{s_1} \\ z_{i+1} - \mu_{s_2} \end{bmatrix} \begin{bmatrix} z_i - \mu_{s_1} \\ z_{i+1} - \mu_{s_2} \end{bmatrix}^T}{\sum_{i=1}^N \psi_i(s_1^2)} \quad (21)$$

The matrices  $A_{n+1}(s_n^{n+1})$  and  $B_{n+1}(s_n^{n+1})$  are calculated using the following formulas:

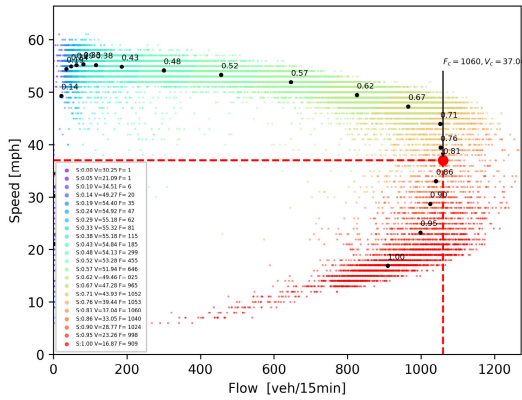
$$A_{n+1}(s_n^{n+1}) = \Sigma_{k,l}^{SF} (\Gamma_k^F)^{-1} \quad (22)$$

$$B_{n+1}(s_n^{n+1}) = \Gamma_l^S - A_{n+1}(s_n^{n+1}) \Sigma_{k,l}^{F,S} \quad (23)$$

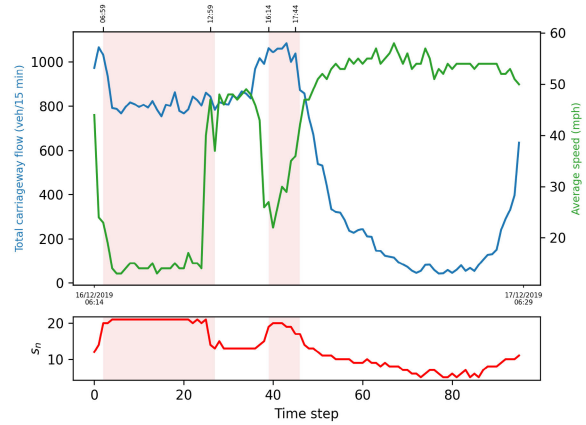
### D. EVENT AND ANOMALY DETECTION

Under normal conditions, traffic tends to seamlessly transition from one state to another. The proposed a priori densities for the fuzzy joint  $s_1^2$  are designed in a such a way that salient transitions are prevented. However, when a sporadic traffic incident occurs it causes the model to pass from a given switch to a relatively distant one. The severity of the incident can be intuitively estimated based on the gap between two consecutive switches: the higher the gap  $|s_{n+1} - s_n|$ , the more likely an event has occurred at time  $n + 1$ .

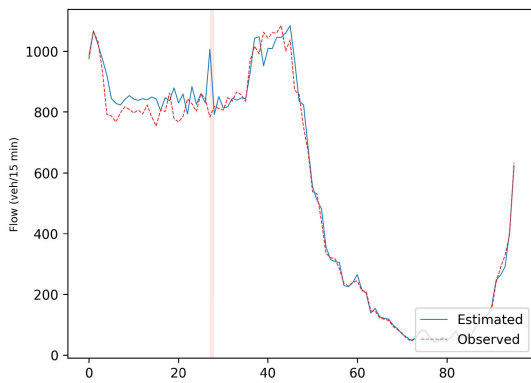
Routine events that occur on a regular basis correspond to traffic jam at peak hours. These events are detected when the estimated switch exceeds a specific threshold  $\tau$ . The threshold  $\tau$  is estimated from the jam density and speed and typically corresponds to 0.8. From a conceptual standpoint, we should distinguish between two different types of incidents: routine and sporadic. Using the proposed model, routine events can be detected may predicted when the estimated switch at time point  $n + 1$  verifies the following conditions: (i)  $s_{n+1} > \tau$  and (ii)  $|s_{n+1} - s_n| < \Delta$ , where  $\Delta$  is a preset threshold that depends on the number of discrete fuzzy levels. For example, if there exist  $L$  fuzzy levels, and if we wish that the model detects any consecutive switches such that the associated gap exceeds two levels,  $\Delta$  should be set to



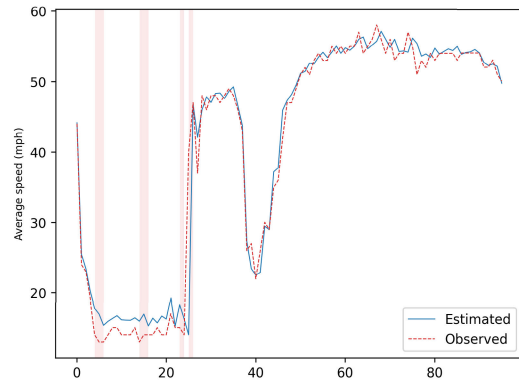
(a) Observation traffic points in a Flow-Speed diagram



(b) Traffic incident detection results from the observed speed data (Flow data are reported to support the detection results). Shaded areas correspond to the time interval during which the event occurs.



(c) Traffic flow estimation (shaded area corresponds to an estimation error above 25%).



(d) One step ahead prediction of of average speed (shaded area corresponds to a prediction error above 25%).

**FIGURE 7.** Traffic event detection and flow estimation using 20 fuzzy levels on site A1/9542A for the time period from 16/12/2019 06:14 until 17/12/2019 06:29.

$\frac{2}{L}$ . Sporadic events are in nature more complex to model and foresee. Nonetheless, they can be estimated based on fair hypotheses that correspond to real-world situations:

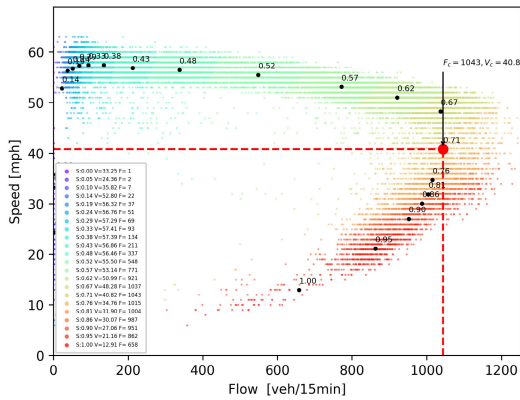
- Case 1: The traffic is heavy (which implies that the last estimated switch  $s_n > \tau$ ). The next switch  $s_{n+1}$  is lower yet still greater than  $\tau$ . Such case corresponds to an event related to one or multiple lanes that have been closed which resulted from a car collision or a broken down vehicle.
- Case 2: The traffic is in a road capacity flow state which typically corresponds to switches lower than  $\tau$  and higher than  $\frac{\tau}{2}$ . The estimated switch  $s_{n+1}$  is such that  $\Delta_n^{n+1} > \frac{2}{L+2}$ . This may indicate an unexpected road condition such as inclement weather, reduced visibility or a sudden road closure.

Figure 5 illustrates some examples of the traffic anomalies in which we distinguish between the two cases of traffic incidents by the shape of the plot of switches.

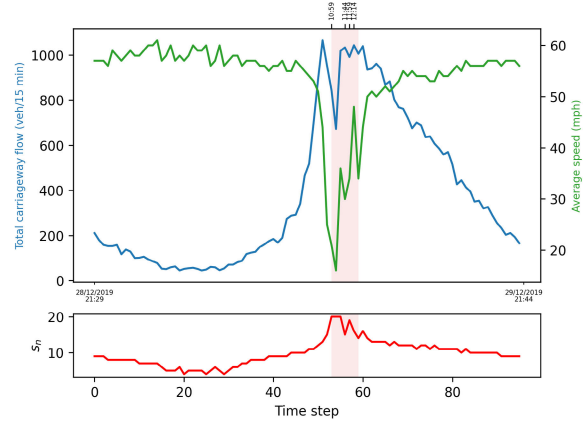
## VI. ONE-STEP AHEAD TRAFFIC STATE PREDICTION

To leverage the proposed CGOMFSM, we proposed a tractable scheme for the prediction of traffic data (speed and flow) at time index  $n + 1$  from the historic speed data only. In the stochastic framework considered here, it consists in estimating  $\mathbb{E}[Z_{n+1} | v_1^n]$ , which corresponds to the expected values of the traffic flow  $V_{n+1}$  and average speed  $F_{n+1}$  at time index  $n + 1$  from past observations  $v_1^n$ . Formally, this is equivalent to computing the following

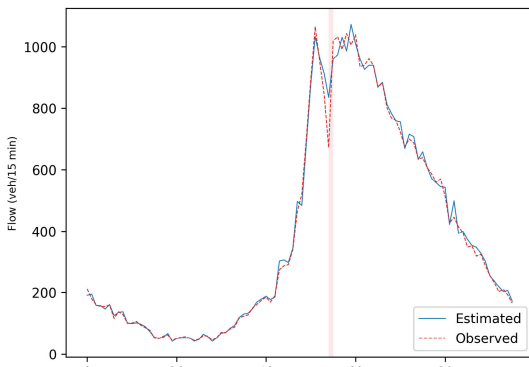
$$\begin{aligned} \mathbb{E}[Z_{n+1} | v_1^n] &= \int z_{n+1} p(z_{n+1} | v_1^n) dz_{n+1} \\ &= \int z_{n+1} \int \int_0^1 p(z_{n+1}, s_n^{n+1} | v_1^n) dv(s_n^{n+1}) dz_{n+1} \\ &= \int \int_0^1 \int z_{n+1} p(z_{n+1} | s_n^{n+1}, v_1^n) \\ &\quad dz_{n+1} p(s_n^{n+1} | v_1^n) dv(s_n^{n+1}) \\ &= \int \int_0^1 \mathbb{E}[Z_{n+1} | s_n^{n+1}, v_1^n] p(s_n^{n+1} | v_1^n) dv(s_n^{n+1}), \end{aligned}$$



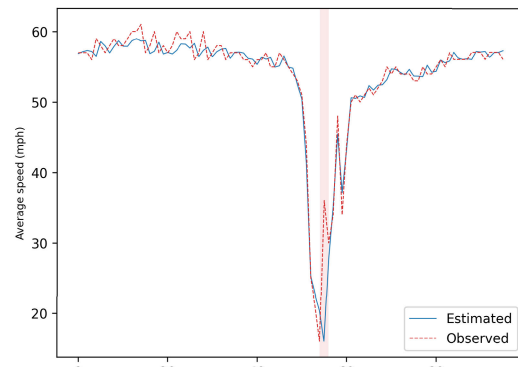
(a) Observation traffic points in a Flow-Speed diagram



(b) Traffic incident detection results from the observed speed data (Flow data are reported to support the detection results). Shaded areas correspond to the time interval during which the event occurs.



(c) Traffic flow estimation (shaded area corresponds to an estimation error above 25%).



(d) One step ahead prediction of of average speed (shaded area corresponds to a prediction error above 25%).

**FIGURE 8. Traffic event detection and flow estimation using 20 fuzzy levels on site A1/9547A for the time period from 28/12/2019 21:29 until 29/12/2019 21:44.**

denoting by  $d\nu(s_n^{n+1})$  the product measure  $d\nu(s_n) \otimes d\nu(s_{n+1})$ .

- The term  $\mathbb{E}[Z_{n+1}|s_n^{n+1}, v_1^n]$  can be calculated using (1) by

$$\mathbb{E}[Z_{n+1}|s_n^{n+1}, v_1^n] = A_{n+1}(s_n^{n+1})\mathbb{E}[Z_n|s_n^{n+1}, v_1^n] + N_{n+1}(s_n^{n+1}), \quad (24)$$

with

$$\mathbb{E}[Z_n|s_n^{n+1}, v_1^n] = \left[ \begin{array}{c} \mathbb{E}[F_n|s_n, v_1^n] \\ v_n \end{array} \right],$$

- The second term can be rewritten  $p(s_n^{n+1}|v_1^n) = p(s_{n+1}|s_n)p(s_n|v_1^n)$ , where  $p(s_n|v_1^n)$  are the filtered posterior probabilities of jumps, whose recursive calculation is detailed in (7) and (8).

Hence it is possible to predict both traffic flow  $\mathbb{E}[F_{n+1}|v_1^n]$  and average speed  $\mathbb{E}[V_{n+1}|v_1^n]$  in a recursive way by taking margins of (24).

### VII. EXPERIMENTAL RESULTS

The objective of this experimental study is twofold. First, we evaluate the efficacy of the proposed traffic flow esti-

mation model using real ground truth traffic data. England highways datasets have been used for this purpose. Second, we evaluate the effectiveness of the model under different penetration rates. A simulation-based study of the city of Salalah in the Sultanate of Oman has been used for this purpose.

#### A. EXPERIMENTS WITH REAL DATASETS

In this part of the experimental study, we focus on the validation of the proposed model using ground truth datasets. The goal is to measure the extent to which the fuzzy approximation of non-linear systems pertains to highways traffic data. For this purpose, we selected the England highways datasets publicly available through the Motorway Incident Detection and Automatic Signalling (MIDAS) system.<sup>1</sup> Each MIDAS site reports, on a regular basis of 15 minutes, the traffic volume that corresponds to different categories of vehicles depending on their lengths, as well as the average speed. For the sake of simplicity, we assume the following:

<sup>1</sup><http://webtris.highwaysengland.co.uk/>

**TABLE 1. One-step (15 minutes) ahead flow prediction results using CGOMFSM against baseline models. For each site, the MAPE and RMSE are reported in the first (%) and second row respectively.**

Site	$m$	$L = 0$	$L = 5$	$L = 10$	$L = 20$	$L = 30$	$L = 50$			
A1/9542A	$m < 1$	12.34%	10.56%	7.13%	5.98%	5.46%	5.39%			
		82.16	67.63	50.02	48.36	36.33	35.09			
	$m = 1$	12.85%	10.87%	7.64%	5.82%	5.61%	5.55%			
		89.95	68.21	54.83	51.79	44.6	41.06			
	$m > 1$	12.66%	10.42%	7.71%	5.90%	5.62%	5.33%			
		82.56	66.42	56.1	50.15	38.16	37.19	LSTM	ARIMA	RW
Average		12.62%	10.62%	7.49%	5.90%	5.56%	5.42%	7.82%	11.43%	13.83%
		84.89	67.42	53.65	50.10	39.70	37.78	58.23	72.33	90.73
A1/9547A	$m < 1$	13.22%	11.05%	8.14%	6.34%	6.29%	6.19%			
		90.05	82.65	64.08	51.11	49.52	47.17			
	$m = 1$	13.67%	11.03%	9.05%	7.13%	6.99%	6.31%			
		99.06	90.83	88.43	58.78	51.5	50.57			
	$m > 1$	13.27%	10.39%	7.95%	6.61%	6.27%	5.86%			
		90.31	81.28	78.9	51.92	48.26	46.55	LSTM	ARIMA	RW
Average		13.39%	10.82%	8.38%	6.69%	6.52%	6.12%	8.46%	11.98%	15.38%
		93.14	84.92	77.14	53.94	49.76	48.10	79.19	88.03	100.66

**TABLE 2. Traffic flow estimation from recorded average speed using CGOMFSM. For each site, the MAPE and RMSE are reported in the first (%) and second row respectively.**

	Site	$L = 0$	$L = 5$	$L = 10$	$L = 20$	$L = 30$	$L = 50$
$m < 1$	A1/9542A	10.92%	9.43%	6.86%	5.54%	5.25%	4.69%
		70.83	59.85	42.39	44.78	35.27	33.74
$m < 1$	A1/9547A	11.80%	10.42%	7.20%	5.87%	5.72%	5.38%
		79.69	69.45	55.24	44.83	43.44	39.31
$m = 1$	A1/9542A	12.60%	9.06%	6.76%	5.15%	5.01%	4.66%
		81.77	59.83	52.72	44.28	38.78	38.02
$m = 1$	A1/9547A	11.51%	10.02%	7.07%	5.67%	5.51%	5.03%
		76.44	62.07	52.43	48.22	36.69	35.42
$m > 1$	A1/9542A	12.66%	10.42%	7.71%	5.90%	5.62%	5.33%
		82.56	66.42	56.10	50.15	38.16	37.19
$m > 1$	A1/9547A	13.40%	11.12%	8.67%	6.81%	6.77%	6.27%
		94.83	88.59	80.48	55.03	50.19	48.41

- The traffic is homogeneous, i.e. we consider only one category of vehicles corresponding to 0-520 cm.
- The traffic is stationary, i.e. we discard traffic states that correspond to conspicuous outliers.

To evaluate the accuracy of the flow estimation, we considered the two error measures MAPE and RMSE defined as follows:

$$MAPE = \frac{1}{T} \sum_{i=1}^T \left| \frac{\hat{x}_i - x_i}{x_i} \right|$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (\hat{x}_i - x_i)^2},$$

where  $x_i$  represents the observed flow ( $x_i = f_i$ ) or the recorded speed ( $x_i = v_i$ ),  $\hat{x}_i$  represents the estimated flow ( $\hat{x}_i = \hat{f}_i$ ) or the predicted speed ( $\hat{x}_i = \hat{v}_i$ ) at time point  $i$ , and  $T$  is the observation interval.

Figure 6 shows two monitoring points situated at A1 highway corresponding to the sites A1/9547A and A1/9542A. Data were collected over a time interval spanning from 01/01/2019 00:14 to: 31/12/2019 23:44, corresponding to

27418 entries of which 26000 were used for the parameter estimation and the remainder (1418 entries) were used to test the estimation and the prediction algorithms. Figures 7(a) and 8(a) show the corresponding fundamental traffic flow speed diagrams where the different fuzzy switches centroids have been plotted to depict the distinct regimes. The results of the proposed incident detection scheme suggest that it is possible to infer a quantifiable traffic description using the observed speed data only. Figures 7(b) and 8(b) show the switches estimation and the traffic incident detection results from the recorded speed data. The shaded areas represent the interval of time during which the traffic incident might have been observed. Figures 7(c), 7(d) and 8(c), 7(d) show the results of traffic flow estimation from recorded speed and the 15 minutes ahead prediction of the average speed using a CGOMFSM with 20 fuzzy levels.

We studied the impact of the number of fuzzy levels on the estimation accuracy. We varied the number of fuzzy levels  $L$  between 0 (hard model) and 50 for different values of the parameter  $m$ . The results of these experiments were averaged and compared to LSTM, ARIMA with order (1, 0, 1) and seasonal order (1, 1, 0, 96) and random walk (RW) for traffic speed prediction. The results of this experiments are reported in Table 1. Similarly, we considered different values of fuzzy levels paired with values of the parameter  $m$  and we assessed the accuracy of traffic flow estimation. Table 2 reports the results of this experiment.

Unsurprisingly, the estimation is more accurate for larger fuzzy levels. However, the MAPE and the RMSE rates tend to stabilise for a number of fuzzy levels ( $L > 20$ ). Since a higher number of fuzzy levels incurs higher computation costs, we can consider that 20 fuzzy levels is a fair tradeoff between estimation accuracy and computational costs.

The parameter  $m$  was varied between 0.1 and 10 with a step of 0.1 for  $m < 1$  and a step of 1 for  $m > 1$ . We have noticed that the parameter  $m$  does not significantly impact the overall accuracy of flow estimation. The obtained results

**TABLE 3.** Impact of penetration rate on the traffic estimation accuracy.

$\rho$	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
MAPE	22.14%	17.01%	16.73%	12.12%	9.34%	9.22%	8.14%	7.83%	7.17%	6.10%

**TABLE 4.** Impact of the penetration rate on the event detection accuracy (Detection lag: time elapsed between the event occurrence and its eventual detection, false alarm: percentage of wrongly detected events out of the total of reported events and detection ratio: the percentage of correctly detected events).

$\rho$	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Detection lag (minutes)	15.2	15.1	13.2	10.9	7.3	4.5	4.2	4.3	4.2	4.1
False alarm (%)	28.7	29.2	27.6	25.3	22.1	17.5	16.7	16.3	16.4	16.3
Detection accuracy (%)	73.1	82.3	84.3	86.2	87.0	88.3	88.2	87.9	88.1	88.6

have not shown a direct correlation between the estimation and prediction accuracy and the parameter  $m$ . However the following conclusions were drawn:

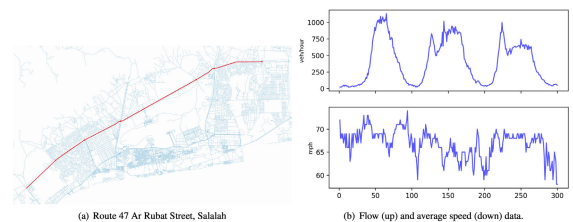
- For a large number of fuzzy levels ( $L > 20$ ),  $m > 1$  provided the best estimation accuracy results.
- For a large number of fuzzy levels ( $L > 20$ ),  $m = 1$  provided the best prediction accuracy results.
- Using the hard model ( $L = 0$ ), the best estimation and prediction results were obtained using  $m < 1$ .
- The worst prediction results were obtained with  $m > 1$  regardless of the number of fuzzy levels.
- In figure 7, the estimation and the prediction errors are relatively higher than those observed in figure 8. This phenomenon can be explained by the sensitivity of the model to high variances especially when the number of fuzzy levels is significantly high.

**B. SIMULATION-BASED EXPERIMENTS OF THE IMPACT OF THE PENETRATION RATE**

Traffic in the sultanate of Oman has been a significant concern due to the significant increase in the number of personal vehicles caused by the relatively limited public transportation service and to the surge of the number of trucks and utility vehicles due to the rapid growth of logistic-related activities across the country. The city of Salalah, the main city in Dhofar governorate, has been witnessing a substantial increase in the number of vehicles over the last years, ascribable to the diversification of the economical activities and the improvement of the level of living. Furthermore, during the tourist season of Monsoon, the city of Salalah witnesses a high influx of visitors from the other governorates and the bordering countries. These visitors travel using their private cars and significantly increase the traffic during the season. Figure 9 shows the road network of Salalah and highlights the stretch of the studied highway.

We used SUMO (Simulation of Urban MObility) to simulate traffic in the studied highways.

The objective of this study is to answer the following questions:



**FIGURE 9.** Stretch of the studied highway in Salalah, Sultanate of Oman (a) and traffic flow and average speed measures from simulations of traffic on Salalah highway (b).

- What is the optimal penetration rate that ensures a near-exact estimation of traffic data? The penetration rate is defined as the percentage of sensing devices reporting their speed. To answer this question, we varied the penetration rate  $\rho$  from 1% to 10% by a step of 1%. The results of this experiment are reported in table 3.
- Since average speed is collected on a regular basis every 15 minutes, if an incident occurs during the time interval  $[t_n, t_{n+1}]$ , how long would it take to detect the incident? If an event occurs during an interval  $I_n$  how many subsequent intervals will elapse before the event has been detected? The detection time lag results are reported in Table 4. Furthermore, we report the missed event and the false alarm ratios. To answer this question, we randomly simulated 1000 traffic incidents by blocking one lane at different locations of the highway at distinct times of the day.

*Remarks:*

- The penetration rate was simulated by a random sampling of vehicles crossing the segment of interest. For example, to obtain a penetration rate of 10%, the sampling scheme consists in reporting the data from one vehicle out of 10.
- The number of fuzzy levels considered in these series of experiments is set to 20. The rationale for this choice is that this number of fuzzy levels yielded accuracy results close to  $F = 50$  within significantly less time.
- For this experiment, we set  $m = 0.5$ .

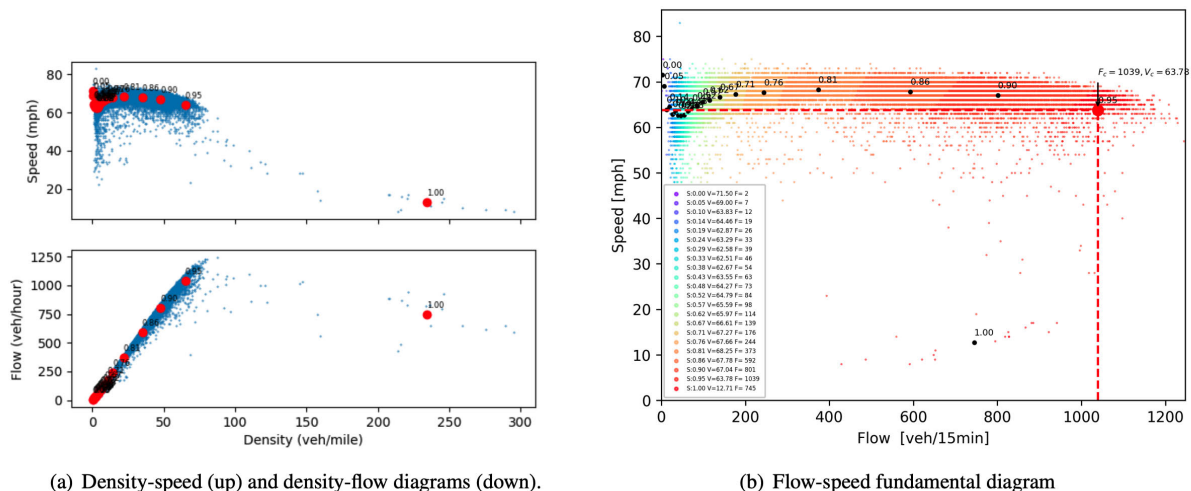


FIGURE 10. Traffic fundamental diagrams with the estimated fuzzy classes (red dots in (b) and labelled black dots in (c)).

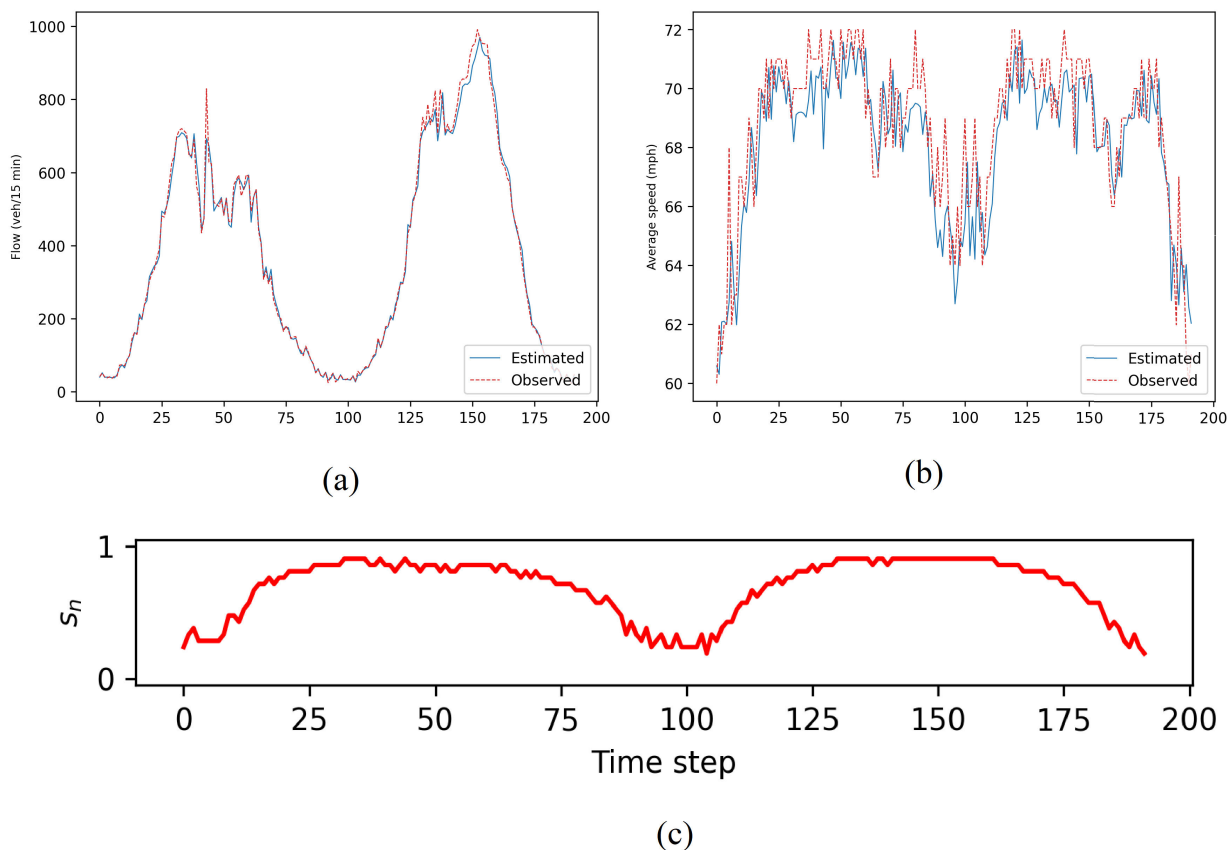


FIGURE 11. Simulated traffic flow estimation (a) and 15-minutes ahead prediction of traffic speed (b), estimated switches process (c).

VIII. CONCLUSION

In this paper, we proposed a cost-effective framework for integrated estimation and prediction of traffic state. The proposed framework relies on a triplet model explicitly fuzzy representing the variation of traffic over time called CGOMFSM. Such modeling allows to rapidly detect any

anomalies that result from an abrupt state change which likely refers to a road incident. The experimental study was conducted at two levels. First, we evaluated the efficacy of the CGOMFSM in terms of traffic flow estimation from aggregated average speed as well as its ability to predict the traffic state over a short time window. Then, we conducted

a simulation to assess the impact of the penetration rate (referring to the ratio of vehicles reporting their speed) on the estimation accuracy. The main results suggest that up to a sufficient penetration rate (around 10%), the proposed conditionally Gaussian observed Markov fuzzy switching model yields satisfactory results in terms of parameter estimation, speed prediction and traffic flow estimation. As the experimental results are very promising, the next research stage will focus on evaluating the performance of the proposed approach in real-world environments in collaboration with competent authorities.

## APPENDIX CALCULATION OF MARGINAL LAW OF $P_{s_1}$

The density  $p(s_1)$  corresponding to the distribution  $P_{s_1}$  is obtained by integrating  $p(s_1, s_2)$  over the interval  $[0, 1]$  w.r.t. the measure  $\nu$ :

$$\begin{aligned} p(s_1) &= \int_0^1 p(s_1, s_2) d\nu(s_2) \\ &= p(s_1, 0) + p(s_1, 1) + \underbrace{\eta \int_0^1 (1 - |s_1 - s_2|^m) ds_2}_{A(s_1)}, \end{aligned}$$

with

$$\begin{aligned} A(s_1) &= \int_0^{s_1} 1 - (s_1 - s_2)^m ds_2 + \int_{s_1}^1 1 - (s_2 - s_1)^m ds_2 \\ &= 1 - \frac{s_1^{m+1} + (1 - s_1)^{m+1}}{m+1}. \end{aligned} \quad (25)$$

Hence:

$$\begin{aligned} p(s_1 = 0) &= p(0, 0) + p(0, 1) + A(0) = \alpha_0 + \beta + \eta \frac{m}{m+1}, \\ p(s_1 = 1) &= p(1, 0) + p(1, 1) + A(1) = \beta + \alpha_1 + \eta \frac{m}{m+1}, \\ p(s_1 \in ]0, 1[) &= \eta (2 - (s_1^m + (1 - s_1)^m) + A(s_1)), \end{aligned}$$

To infer a relationship between the parameters  $\alpha_0$ ,  $\beta$ ,  $\alpha_1$  and  $\eta$ , we use the normalization condition:

$$\int_0^1 p(s_1) d\nu(s_1) = p(s_1 = 0) + p(s_1 = 1) + \int_0^1 p(s_1) ds_1 = 1.$$

Therefore, after a few calculations not reported here, we get

$$\alpha_0 + \alpha_1 + 2\beta + \eta \frac{m(5m+11)}{(m+1)(m+2)} = 1,$$

so that

$$\eta = \frac{(m+1)(m+2)}{m(5m+11)} (1 - \alpha_0 - \alpha_1 - 2\beta), \quad (26)$$

and the model is only parametrized by 4 parameters:  $m$ ,  $\alpha_0$ ,  $\alpha_1$  and  $\beta$ .

## REFERENCES

- [1] A. Allström, J. Barceló, J. Ekström, E. Grumert, D. Gundlegård, and C. Rydberg, "Traffic management for smart cities," in *Designing, Developing, and Facilitating Smart Cities*. Springer, 2017, pp. 211–240.
- [2] A. Janecek, D. Valerio, K. A. Hummel, F. Ricciato, and H. Hlavacs, "The cellular network as a sensor: From mobile phone data to real-time road traffic monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2551–2572, Oct. 2015.
- [3] W. Li, D. Nie, D. Wilkie, and M. C. Lin, "Citywide estimation of traffic dynamics via sparse GPS traces," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 3, pp. 100–113, Fall 2017.
- [4] H. Mai-Tan, H.-N. Pham-Nguyen, N. X. Long, and Q. T. Minh, "Mining urban traffic condition from crowd-sourced data," *Social Netw. Comput. Sci.*, vol. 1, no. 4, pp. 1–16, Jul. 2020.
- [5] M. Gramaglia, M. Calderon, and C. J. Bernardos, "Abeona monitored traffic: VANET-assisted cooperative traffic congestion forecasting," *IEEE Veh. Technol. Mag.*, vol. 9, no. 2, pp. 50–57, Jun. 2014.
- [6] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation with mixed connected and conventional vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3484–3497, Dec. 2016.
- [7] S. M. Khan, K. C. Dey, and M. Chowdhury, "Real-time traffic state estimation with connected vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1687–1699, Jul. 2017.
- [8] E. Papapanagiotou and F. Busch, "Extended observer for urban traffic control based on limited measurements from connected vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1664–1676, Apr. 2020.
- [9] Q. Guo, L. Li, and X. J. Ban, "Urban traffic signal control with connected and automated vehicles: A survey," *Transp. Res. C, Emerg. Technol.*, vol. 101, pp. 313–334, Apr. 2019.
- [10] Y. Ma, M. Chowdhury, A. Sadek, and M. Jeahani, "Real-time highway traffic condition assessment framework using vehicle–infrastructure integration (VII) with artificial intelligence (AI)," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 615–627, Dec. 2009.
- [11] Y.-E. Sun, H. Huang, S. Chen, Y. Zhou, K. Han, and W. Yang, "Privacy-preserving estimation of  $k$ -persistent traffic in vehicular cyber-physical systems," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8296–8309, Oct. 2019.
- [12] F. Zheng, S. E. Jabari, H. X. Liu, and D. Lin, "Traffic state estimation using stochastic Lagrangian dynamics," *Transp. Res. B, Methodol.*, vol. 115, pp. 143–165, Sep. 2018.
- [13] H. Yang, P. J. Jin, B. Ran, D. Yang, Z. Duan, and L. He, "Freeway traffic state estimation: A Lagrangian-space Kalman filter approach," *J. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 525–540, Nov. 2019.
- [14] C. N. Van Phu and N. Farhi, "Estimation of urban traffic state with probe vehicles," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 25, 2020, doi: 10.1109/TITS.2020.2975120.
- [15] N. Caceres, L. M. Romero, F. G. Benitez, and J. M. del Castillo, "Traffic flow estimation models using cellular phone data," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1430–1441, Sep. 2012.
- [16] Y. Zhao, J. Zheng, W. Wong, X. Wang, Y. Meng, and H. X. Liu, "Various methods for queue length and traffic volume estimation using probe vehicle trajectories," *Transp. Res. C, Emerg. Technol.*, vol. 107, pp. 70–91, Oct. 2019.
- [17] T. Seo, T. Kusakabe, and Y. Asakura, "Traffic state estimation with the advanced probe vehicles using data assimilation," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 824–830.
- [18] T. Seo and T. Kusakabe, "Probe vehicle-based traffic flow estimation method without fundamental diagram," *Transp. Res. Procedia*, vol. 9, pp. 149–163, Jan. 2015.
- [19] A. Sunderrajan, V. Viswanathan, W. Cai, and A. Knoll, "Traffic state estimation using floating car data," *Procedia Comput. Sci.*, vol. 80, pp. 2008–2018, Jan. 2016.
- [20] S. Messelodi, C. M. Modena, M. Zanin, F. G. B. De Natale, F. Granelli, E. Betterle, and A. Guarise, "Intelligent extended floating car data collection," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4213–4227, Apr. 2009.
- [21] X. Zhan, Y. Zheng, X. Yi, and S. V. Ukkusuri, "Citywide traffic volume estimation using trajectory data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 2, pp. 272–285, Feb. 2017.

- [22] J. Snowdon, O. Gkountouna, A. Züfle, and D. Pfoser, "Spatiotemporal traffic volume estimation model based on GPS samples," in *Proc. 5th Int. ACM SIGMOD Workshop Manag. Mining Enriched Geo-Spatial Data*, 2018, pp. 1–6.
- [23] C. Meng, X. Yi, L. Su, J. Gao, and Y. Zheng, "City-wide traffic volume inference with loop detector data and taxi trajectories," in *Proc. 25th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, Nov. 2017, pp. 1–10.
- [24] K. Anuar, F. Habtemichael, and M. Cetin, "Estimating traffic flow rate on freeways from probe vehicle data and fundamental diagram," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 2921–2926.
- [25] T. Seo, Y. Kawasaki, T. Kusakabe, and Y. Asakura, "Fundamental diagram estimation by using trajectories of probe vehicles," *Transp. Res. B, Methodol.*, vol. 122, pp. 40–56, Apr. 2019.
- [26] S. Wang, W. Huang, and H. K. Lo, "Traffic parameters estimation for signalized intersections based on combined shockwave analysis and Bayesian network," *Transp. Res. C, Emerg. Technol.*, vol. 104, pp. 22–37, Jul. 2019.
- [27] M. A. Aljamal, H. M. Abdelghaffar, and H. A. Rakha, "Developing a neural-Kalman filtering approach for estimating traffic stream density using probe vehicle data," *Sensors*, vol. 19, no. 19, p. 4325, Oct. 2019.
- [28] M. A. Aljamal, H. M. Abdelghaffar, and H. A. Rakha, "Real-time estimation of vehicle counts on signalized intersection approaches using probe vehicle data," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 21, 2020, doi: [10.1109/TITS.2020.2973954](https://doi.org/10.1109/TITS.2020.2973954).
- [29] M. Cetin and K. A. Anuar, "Using probe vehicle trajectories in stop-and-go waves for inferring unobserved vehicles," in *Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Jun. 2017, pp. 333–338.
- [30] B. Salahshour, R. V. Nezafat, and M. Cetin, "Estimation of unobserved vehicles in congested traffic from probe vehicle samples," in *Proc. Int. Conf. Transp. Develop., Innov. Sustain. Smart Mobility Smart Cities*. Reston, VA, USA: American Society of Civil Engineers, Aug. 2019, pp. 24–35.
- [31] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura, "Traffic state estimation on highway: A comprehensive survey," *Annu. Rev. Control*, vol. 43, pp. 128–151, Jan. 2017.
- [32] J. Martín, E. J. Khatib, P. Lázaro, and R. Barco, "Traffic monitoring via mobile device location," *Sensors*, vol. 19, no. 20, p. 4505, Oct. 2019.
- [33] Q. Ma, Z. Zou, and S. Ullah, "An approach to urban traffic condition estimation by aggregating GPS data," *Cluster Comput.*, vol. 22, no. 3, pp. 5421–5434, May 2019.
- [34] J.-D. Zhang, J. Xu, and S. S. Liao, "Aggregating and sampling methods for processing GPS data streams for traffic state estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1629–1641, Dec. 2013.
- [35] S. Kanarachos, S.-R.-G. Christopoulos, and A. Chronos, "Smartphones as an integrated platform for monitoring driver behaviour: The role of sensor fusion and connectivity," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 867–882, Oct. 2018.
- [36] J. Yu, M. E. J. Stettler, P. Angeloudis, S. Hu, and X. M. Chen, "Urban network-wide traffic speed estimation with massive ride-sourcing GPS traces," *Transp. Res. C, Emerg. Technol.*, vol. 112, pp. 136–152, Mar. 2020.
- [37] T. Seo, T. Kusakabe, and Y. Asakura, "Estimation of flow and density using probe vehicles with spacing measurement equipment," *Transp. Res. C, Emerg. Technol.*, vol. 53, pp. 134–150, Apr. 2015.
- [38] C. Bucknell and J. C. Herrera, "A trade-off analysis between penetration rate and sampling frequency of mobile sensors in traffic state estimation," *Transp. Res. C, Emerg. Technol.*, vol. 46, pp. 132–150, Sep. 2014.
- [39] A. D. Patire, M. Wright, B. Prodhomme, and A. M. Bayen, "How much GPS data do we need?" *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 325–342, Sep. 2015.
- [40] S. Li, G. Li, Y. Cheng, and B. Ran, "Urban arterial traffic status detection using cellular data without cellphone GPS information," *Transp. Res. C, Emerg. Technol.*, vol. 114, pp. 446–462, May 2020.
- [41] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 4, pp. 568–583, Aug. 2010.
- [42] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in Rome," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 141–151, Mar. 2011.
- [43] T. Derrmann, R. Frank, F. Viti, and T. Engel, "How road and mobile networks correlate: Estimating urban traffic using handovers," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 521–530, Feb. 2020.
- [44] J. Wahlström, I. Skog, and P. Händel, "Smartphone-based vehicle telematics: A ten-year anniversary," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2802–2825, Oct. 2017.
- [45] S. Geisler, C. Quix, S. Schiffer, and M. Jarke, "An evaluation framework for traffic information systems based on data streams," *Transp. Res. C, Emerg. Technol.*, vol. 23, pp. 29–55, Aug. 2012.
- [46] Q.-J. Kong, Q. Zhao, C. Wei, and Y. Liu, "Efficient traffic state estimation for large-scale urban road networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 398–407, Mar. 2013.
- [47] J. Zheng and H. X. Liu, "Estimating traffic volumes for signalized intersections using connected vehicle data," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 347–362, Jun. 2017.
- [48] X. Luo, B. Liu, P. J. Jin, Y. Cao, and W. Hu, "Arterial traffic flow estimation based on vehicle-to-cloud vehicle trajectory data considering multi-intersection interaction and coordination," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 6, pp. 68–83, Jun. 2019.
- [49] J. Yao, F. Li, K. Tang, and S. Jian, "Sampled trajectory data-driven method of cycle-based volume estimation for signalized intersections by hybridizing shockwave theory and probability distribution," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2615–2627, Jun. 2020.
- [50] K. Anuar and M. Cetin, "Estimating freeway traffic volume using shockwaves and probe vehicle trajectory data," *Transp. Res. Procedia*, vol. 22, pp. 183–192, Jan. 2017.
- [51] R. Chen and M. W. Levin, "Traffic state estimation based on Kalman filter technique using connected vehicle V2V basic safety messages," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 4380–4385.
- [52] Y. Cui, B. Jin, F. Zhang, B. Han, and D. Zhang, "Mining spatial-temporal correlation of sensory data for estimating traffic volumes on highways," in *Proc. 14th EAI Int. Conf. Mobile Ubiquitous Syst., Comput., Netw. Services*, 2018, pp. 343–352.
- [53] X. Tang, B. Gong, Y. Yu, H. Yao, Y. Li, H. Xie, and X. Wang, "Joint modeling of dense and incomplete trajectories for citywide traffic volume inference," in *Proc. World Wide Web Conf.*, 2019, pp. 1806–1817.
- [54] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2014, pp. 1027–1036.
- [55] Y.-C. Lai and S.-Y. Huang, "Accurate traffic flow estimation in urban roads with considering the traffic signals," in *Proc. Int. Conf. Internet Vehicles*. Kanazawa, Japan: Springer, Nov. 2017, pp. 41–52.
- [56] A. Millard-Ball, R. C. Hampshire, and R. R. Weinberger, "Map-matching poor-quality GPS data in urban environments: The pgMapMatch package," *Transp. Planning Technol.*, vol. 42, no. 6, pp. 539–553, Aug. 2019, doi: [10.1080/03081060.2019.1622249](https://doi.org/10.1080/03081060.2019.1622249).
- [57] S. Derrode and W. Pieczynski, "Exact fast computation of optimal filter in Gaussian switching linear systems," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 701–704, Jul. 2013.
- [58] N. Abbassi, D. Benboudjema, S. Derrode, and W. Pieczynski, "Optimal filter approximations in conditionally Gaussian pairwise Markov switching models," *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 1104–1109, Apr. 2015.
- [59] Z. Bouyahia, S. Derrode, and W. Pieczynski, "Filtering in Gaussian linear systems with fuzzy switches," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 8, pp. 1760–1770, Aug. 2020.



**ZIED BOUYAHIA** received the master's degree in protocols, networks, images and multimedia systems from the National School of Computer Science, University of Manouba, Tunisia, in 2007, and the Ph.D. degree in computer science from the University of Manouba, in 2013. In 2008, he joined the National School of Computer Science as an Assistant Professor. Since 2015, he has been with the Department of Computer Science, College of Arts and Applied Sciences, Dhofar University, Sultanate of Oman, as an Assistant Professor. His current research interests include smart mobility and intelligent transportation systems.





**HEDI HADDAD** received the M. B. A. degree and the Ph.D. degree in computer science from Laval University, Québec, QC, Canada, in 2002 and 2009, respectively. He is currently an Assistant Professor with the Department of Computer Science, Dhofar University, Salalah, Oman. His research interests include the development of spatio-temporal decision support systems using different disciplines, including multi-agent geosimulation, wireless sensor networks, and geographic information systems. Recently, he has been interested in spatially-aware smart applications for the IoT in the transportation domain.



**WOJCIECH PIECZYNSKI** received the Doctorat d'Etat degree in mathematical statistics from the Université Pierre et Marie Curie, Paris, France, in 1986. He is currently a Professor with the Telecom Sudparis, Institut Polytechnique de Paris, and the Head of the Communications, Image, and Information Processing (CITI) Department. His research interests include mathematical statistics, stochastic processes, and statistical signal processing.

...



**STÉPHANE DERRODE** was as an Associate Professor in signal and image processing first with the École Nationale Supérieure de Physique de Marseille, France, in 2001, and then with the École Centrale de Marseille, in 2005. He joined the École Centrale de Lyon, France, in 2014, where he is currently a Professor in computer sciences with the Mathematics and IT Department. His research interests include nonlinear, non-Gaussian state-space systems for time-series analysis, with applications to image segmentation and signal filtering.