The following publication Y. Zhou, E. Chung, M. E. Cholette and A. Bhaskar, "Real-Time Joint Estimation of Traffic States and Parameters Using Cell Transmission Model and Considering Capacity Drop," 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 2797-2804 is available at https://doi.org/10.1109/ITSC.2018.8569805

Real-Time Joint Estimation of Traffic States and Parameters Using Cell Transmission Model and Considering Capacity Drop

Yue Zhou, *Student Member, IEEE,* Edward Chung, Michael E. Cholette, and Ashish Bhaskar

***Abstract*—This paper contributes to an understudied category of traffic state estimation approaches, i.e. using a Godunov-type discrete traffic flow model (e.g. the Cell Transmission Model, CTM) to simultaneously estimate traffic flow parameters and traffic densities. Our main estimation algorithm is based on the CTM and the extended Kalman filter (EKF). Compared to previous studies, this study has two features. First, we take into account the effect of capacity drop, a factor that is largely ignored by previous studies in traffic state estimation. Second, a separate, supervisory observer capturing the capacity drop mode is attached to the main algorithm. Such a treatment enables the main estimation algorithm to more accurately switch between functions of free-flow regime and congested regime. It thus avoids mismatches between the applied models and the measurements, a common pitfall in conventional CTM-EKF approaches, hence can potentially enhance the quality of estimation. The proposed method was tested using micro-simulation data and showed a satisfactory performance in tracking variations of traffic flow parameters and estimating traffic densities in real time.**

Keywords—traffic state estimation, system identification, extended Kalman filter, Cell Transmission Model, Godunov scheme

I. Introduction

Traffic state estimation is essential to many Intelligent Transportation Systems (ITS) applications. Model-based traffic state estimation tries to infer vehicle densities in road sections based on a model of traffic flow dynamics and measurements of traffic flow variables such as flow rate, space-mean speed, and occupancy. A traffic flow model usually contains parameters such as free-flow speed and critical density. In reality, the values of these parameters are subject to external conditions such as weather, lighting, traffic compositions, etc., and thus is time-varying. As discussed recently in [10], proactive traffic control strategies attempt to prevent traffic breakdown at bottlenecks (and thus capacity drop), while reactive traffic control strategies allow the onset of congestion and then maximize the bottleneck discharge. Obviously, good knowledge of the values of traffic flow parameters such as free-flow speed and critical density in real time is important to model-based reactive traffic control strategies. Furthermore, we would like to point out that, even for proactive traffic control strategies, good knowledge of time variations of the traffic flow parameters is desirable, because it can help to mitigate "the latent possibility of capacity underutilization" or overutilization that fails the control [10]. Thus, even under the framework of a proactive control strategy, it can be useful to periodically allow a temporary occurrence of congestion, such that knowledge of the values of the traffic flow parameters can be updated in time.

In the rich literature of traffic state estimation, most have assumed the traffic flow parameters to be known and constant, e.g. [1-9]. In contrast, online calibration of traffic flow parameters is much less studied in literature. The seminal work of [15] is one earliest effort study that has jointly estimated traffic states and traffic flow parameters. The authors proposed a general approach based on extended Kalman filtering (EKF) under which free-flow speed, critical density, and etc., are estimated jointly with traffic densities in real time, by taking measurements of flow rates and space-mean speeds. However, the discrete traffic flow model employed in [15] does not conform to the Godunov scheme [16]. Godunov scheme is a numerical scheme that solves a conservation PDE with physically correct inter-cell boundary fluxes. The only two studies, as far as we have noticed, that have employed a Godunov-type discrete traffic flow model and treated the traffic flow parameters as time-varying unknown parameters are [11, 12]. However, neither study investigated their method's capability in tracking the free-flow speed and critical density, two important traffic flow parameters

TABLE I
SUMMARY OF REPRESENTATIVE TRAFFIC STATE ESTIMATION STUDIES

	Godunov-type traffic flow model	Non-Godunov-type traffic flow model
Traffic flow parameters treated as known and constant	[1-8]	[9]
Traffic flow parameters treated as unknown and time-varying	[11, 12]	[13-15]

According to whether a traffic state estimation study has jointly estimated traffic flow parameters in real time, and whether the study has employed a Godunov type traffic flow model, we have categorized representative existing literature

This work was supported by Queensland University of Technology.

Yue Zhou (y59.zhou@hdr.qut.edu.au) is with the School of Civil Engineering and Built Environment, Queensland University of Technology, Australia.

Edward Chung (edward.cs.chung@polyu.edu.hk) is with the Department of Electrical Engineering, Hong Kong Polytechnic University, Hong Kong, China.

Michael E. Cholette (michael.cholette@qut.edu.au) is with the School of Chemistry, Physics and Mechanical Engineering, Queensland University of Technology, Australia.

Ashish Bhaskar (ashish.bhaskar@qut.edu.au) is with the School of Civil Engineering and Built Environment, Queensland University of Technology, Australia.

in Table I. This article focuses on the lower-left category of Table I, considering the physical correctness of a Godunov type traffic flow model and the fact that traffic flow parameters are usually time-varying and cannot be perfectly known by offline methods. Existing studies in this category have the below shortcomings.

First, no previous studies in traffic state estimation, in all the four categories, have considered capacity drop. Although capacity drop has been intensively studied in empirical studies and traffic flow modeling and widely considered in designing traffic control strategies, they have been largely ignored by traffic state estimation studies. Since the capacity drop magnitude can be as high as 10% to 25% depending on locations [17, 18], ignorance of capacity drop in traffic state estimation could render the effect of capacity loss erroneously absorbed by the estimates of traffic flow parameters, hence leading to improper traffic control decisions.

Second, approaches belonging to the lower-left category can be vulnerable to poor initial estimates of critical density. Under the assumption of a piecewise-linear fundamental diagram and using a CTM-EKF estimation approach, the critical density will not become observable (hence cannot be corrected by measurements) until a congestion is present upstream of the bottleneck. For example, an underestimated (overestimated) initial critical density can cause the plant dynamics a premature (delayed) switch from free-flow regime to congested regime, while in reality the plant has not yet been congested (has already been congested for a while). This point will be discussed in detail in Section II. B. As a result, the quality of traffic state and parameter estimates can be potentially undermined. The more significant the initial bias is, the more significant the potential negative impact can be.

Third, existing studies in the lower-left category, e.g. [11, 12], did not investigate the methods' capability in tracking traffic flow parameters, as has been done by the lower-right category methods, e.g. [13-15]. Note that there exists a significant difference between the lower-right and lower-left categories' methods. That is, the lower-right category methods usually do not use a Godunov type discrete traffic flow model (e.g. CTM) and the plant dynamics model are usually continuous. Hence, conclusions on tracking capabilities drawn from the lower-right category do not necessarily apply to the lower-left category. One existing study in the lower-left category, [11] did test the tracking capability of their method, however, it was the so-called "capacity" that was tracked, rather than typically concerned traffic flow parameters, i.e. free-flow speed and critical density. Moreover, the method of [11] is not a generic extension of the general EKF approach of [15] from a non-Godunov type to a Godunov type model-based estimation. An investigation of the tracking capability of a generic extension of the general EKF approach of [15] from non-Godunov type to Godunov type is still missing from the literature.

This article tries to fill in the gaps and mitigate the shortcomings of the existing lower-left category methods identified above. The contributions are summarized in the following. First, we consider capacity drop in our estimation method. To our knowledge this is the first such effort in traffic state estimation, in all the four categories. Second, to alleviate negative impact of poor initial critical density estimate, we propose the idea of coupling a separate, supervisory capacity drop mode observer to the main estimation algorithm in order to more accurately determine times to switch between free-flow and congested regimes. The motivation behind the idea is to take advantage of the fact that in a road section consisting of a bottleneck, capacity drop is always associated with onset of a congestion. Third, we evaluate the tracking capability of the proposed method using CTM synthesized data and micro-simulation synthesized data, respectively, which offers insights into the general tracking capability of the lower-left category approaches in Table I.

The remainder of this paper is organized as follows. Section II introduces the conventional CTM-EKF estimation approach and points out a major deficiency associated with it. Section III introduces the proposed methodology, including a heuristic but effective capacity drop mode observer, and the modified CTM-EKF approach which takes advantages of the observer. Section IV validates the modified CTM-EKF approach by recovering true traffic states and traffic flow parameters synthesized using CTM. Section V presents an experiment using data generated by Aimsun, a commercial microscopic traffic simulation software. Section VI concludes the paper and points out future research.

II. The Conventional CTM-EKF Estimation Approach and Its Deficiency

A. Cell Transmission Model of a Lane-Drop Bottleneck Section

To fix ideas, from here on all the discussions are concerned with a lane-drop highway bottleneck section. But note that the methodology can be extended to other types of bottlenecks. It is assumed that the section is divided into N cells, where the lane drop happens at the last cell. Please refer to Fig. 1.

Fig. 1. Configuration of CTM modeling of a lane-drop bottleneck section.

We adopt a constant capacity drop proportion, and integrate it into the CTM in a way consistent with [19-21]. Traffic density updating equation of each cell is given by (1).

$$\rho_k^i = \rho_{k-1}^i + \frac{\Delta t}{\lambda_i \Delta x_i}\left(q_{k-1}^{i-1,i} - q_{k-1}^{i,i+1}\right) \quad (1)$$

In (1), k denotes simulation time step, i denotes the cell index, $i = 1,2,\ldots,N$, $q_{k-1}^{i-1,i}$ denotes the boundary flow between cell i-1 and i, Δt represents the simulation updating interval, Δx_i represents the length of cell i, and λ_i represents the number of lanes of cell i. The inter-cell boundary flow rates are determined by the minimum of the demand of the upstream cell and the supply of the downstream cell, i.e.

For cell $i = 1,2,\ldots,N$,

$$q_{k-1}^{i-1,i} = \min\{D_{k-1}^{i-1}, S_{k-1}^{i}\} \quad (2)$$

$$q_{k-1}^{i,i+1} = \min\{D_{k-1}^{i}, S_{k-1}^{i+1}\} \quad (3)$$

with the exceptions that

$$q_{k-1}^{0,1} = \min\{q_{k-1}^{in}, S_{k-1}^{1}\} \quad (4)$$

$$q_{k-1}^{N,N+1} = D_{k-1}^{N} \quad (5)$$

where q_{k-1}^{in} is known. Note that (5) is because of the fundamental assumption that cell N is the most restrictive bottleneck cell of the concerned section, and there is not a more restrictive bottleneck downstream of it. Such an assumption is common in similar studies. It is always possible to segment a highway into separate sections each of which contains a most restrictive bottleneck that is beyond the reach of congestion propagated from a further downstream bottleneck.

For cell i = 1,2,…,N-1, the demand and supply rates are given by (6) and (7), respectively.

$$D_{k-1}^{i} = \lambda_i v_{k-1}^{\mathrm{fr}} \min\{\rho_{k-1}^{i}, \rho_{k-1}^{\mathrm{cr}}\} \quad (6)$$

$$S_{k-1}^{i} = \lambda_i v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}} \min\left\{1, \frac{\rho^{\mathrm{jam}} - \rho_{k-1}^{i}}{\rho^{\mathrm{jam}} - \rho_{k-1}^{\mathrm{cr}}}\right\} \quad (7)$$

In (6) and (7), ρ_{k-1}^{i} is the density of cell i at time k-1, v_{k-1}^{fr} and ρ_{k-1}^{cr} are free-flow speed and critical density, respectively. In this study, these two parameters are treated as time-varying. ρ^{jam} is the jam density, treated as known and constant, because it is relatively easy to be estimated offline.

For cell N, its demand and supply functions are defined by (8) and (9), respectively.

$$D_{k-1}^{N} = \lambda_N v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{N} \quad (8)$$

$$S_{k-1}^{N} = \begin{cases} \lambda_N v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}}, & \rho_{k-1}^{N-1} < \frac{\lambda_N}{\lambda_{N-1}} \rho_{k-1}^{\mathrm{cr}} \\ \lambda_N v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}} (1-\theta), & \rho_{k-1}^{N-1} \geq \frac{\lambda_N}{\lambda_{N-1}} \rho_{k-1}^{\mathrm{cr}} \end{cases} \quad (9)$$

In (9), θ denotes the capacity drop proportion. In this paper, it is treated as a known and constant parameter, considering that in reality it can be easily estimated offline and is usually a constant. In future research, it may be desirable to also track its time variations along with other time-varying traffic flow parameters.

B. A Deficiency of the Conventional Nonlinear CTM-EKF Estimation Approach

To estimate traffic densities of each cell and traffic flow parameters in a recursive fashion, a state-space model consisting of a process model and a measurement model is needed. The process model describes plant dynamics, or state transition dynamics. The measurement model maps state variables to measurements. In this study, since we treat free-flow speed and critical density as time-varying and want to estimate their values in real time, we augment them into the state space which would otherwise only contain traffic densities. We model the transition dynamics of free-flow speed and critical density as random walks. Such treatments are consistent with [15].

If the free-flow speed and critical density are to be treated as state variables, both the process model and measurement model become nonlinear in state variables. This is in contrast to many previous traffic state estimation studies that have employed CTM, e.g. [1-8], where the traffic flow parameters were treated as known and constant, and hence the process and measurement models were both linear in state variables. As a result, in this study a nonlinear recursive estimator is needed to solve the nonlinear state-space model. The extended Kalman filter (EKF) is a natural choice, because it is straightforward to implement and is computationally more efficient than particle filters.

A general discrete-time state-space model composed of a nonlinear process model and a nonlinear measurement model with linear additions of noises are given as (10) to (13).

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \boldsymbol{\omega}_{k-1} \quad (10)$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \boldsymbol{\varepsilon}_k \quad (11)$$

$$\boldsymbol{\omega}_{k-1} \sim (0, \mathbf{Q}_{k-1}) \quad (12)$$

$$\boldsymbol{\varepsilon}_k \sim (0, \mathbf{R}_k) \quad (13)$$

In a CTM-EKF estimation approach that has augmented free-flow speed and critical density into the state space, the specific process model is given by (14) to (16).

$$\rho_k^i = \rho_{k-1}^i + \frac{\Delta t}{\lambda_i \Delta x_i}\left(q_{k-1}^{i-1,i} - q_{k-1}^{i,i+1}\right) + \omega_{k-1}^{\rho^i} \quad i = 1, 2, \ldots, \mathrm{N} \quad (14)$$

$$v_k^{\mathrm{fr}} = v_{k-1}^{\mathrm{fr}} + \omega_{k-1}^{v^{\mathrm{fr}}} \quad (15)$$

$$\rho_k^{\mathrm{cr}} = \rho_{k-1}^{\mathrm{cr}} + \omega_{k-1}^{\rho^{\mathrm{cr}}} \quad (16)$$

The state vector $\mathbf{x}_k$ is given by $\left[\rho_k^1\ \rho_k^2\ \ldots\ \rho_k^N\ v_k^{\mathrm{fr}}\ \rho_k^{\mathrm{cr}}\right]^T$. The input $\mathbf{u}_{k-1}$ is scalar, q_{k-1}^{in}, i.e. the in-flow to the concerned section (refer to (4)). The RHS of (14) to (16) without the noise terms collectively define $\mathbf{f}_{k-1}(\cdot)$ as in (10).

A key in the nonlinear CTM-EFK approach is to evaluate the time-varying Jacobian matrix of the process model at each time step. This requires determination of the specific functional form of the time-varying $\mathbf{f}_{k-1}(\cdot)$ at each step, from which the Jacobian matrix is derived. The time variation of $\mathbf{f}_{k-1}(\cdot)$ is due to the implicit switching nature of the boundary flow functions in (14). Hence the key is to correctly identify the functional form of the boundary flows at each time step.

In the following we explain why the conventional approach of determining the boundary flows as used by [12] can be problematic when the critical density is being estimated. Consider the boundary flow between cell N-1 and cell N, $q_{k-1}^{N-1,N}$. This boundary flow will always be the first to be influenced by a congestion because cell N is the bottleneck cell. As a result, it is always through this flow function the critical density first becomes observable. Conventionally, the functional form of $q_{k-1}^{N-1,N}$ is defined by the (17) to (19). Note that the purpose of (17) to (19) is to identify the functional forms of D_{k-1}^{N-1}, S_{k-1}^{N}, and $q_{k-1}^{N-1,N}$, not to calculate the values

as in simulation tasks.

$$D_{k-1}^{N-1} := \lambda_{N-1} v_{k-1}^{\mathrm{fr}} \min\{\rho_{k-1}^{N-1}, \rho_{k-1}^{\mathrm{cr}}\} \tag{17}$$

$$S_{k-1}^{N} := \begin{cases} \lambda_N v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}}, & \rho_{k-1}^{N-1} < \frac{\lambda_N}{\lambda_{N-1}} \rho_{k-1}^{\mathrm{cr}} \\ \lambda_N v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}} (1-\theta), & \rho_{k-1}^{N-1} \geq \frac{\lambda_N}{\lambda_{N-1}} \rho_{k-1}^{\mathrm{cr}} \end{cases} \tag{18}$$

$$q_{k-1}^{N-1,N} := \min\{D_{k-1}^{N-1}, S_{k-1}^{N}\} \tag{19}$$

Note that (17) is a shorthand for the following logic: If $\rho_{k-1}^{N-1} < \rho_{k-1}^{\mathrm{cr}}$, the functional form of D_{k-1}^{N-1} is $\lambda_1 v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{N-1}$; else it is $\lambda_1 v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}}$. Equation (19) is a shorthand for the following logic: If $D_{k-1}^{N-1} < S_{k-1}^{N}$, the functional form of $q_{k-1}^{N-1,N}$ is the same as D_{k-1}^{N-1}, else it is the same as S_{k-1}^{N}. Equation (18) has no ambiguous meaning.

Such an approach in defining flow function $q_{k-1}^{N-1,N}$ can be vulnerable to a biased initial estimate of critical density. For instance, suppose that the initial critical density is underestimated. Under a CTM-EKF estimation framework using a piecewise-linear fundamental diagram, the entire concerned road section will remain at free-flow regime and the critical density estimate will not be updated until the condition $\rho_{k-1}^{N-1} \geq \frac{\lambda_N}{\lambda_{N-1}} \rho_{k-1}^{\mathrm{cr}}$ is satisfied. As a result, the underestimated initial value of critical density will lead to a premature satisfaction of this condition, hence erroneously rendering a switch of S_{k-1}^{N} from its free-flow form to its congested form, and consequently an erroneous switch of the functional form of $q_{k-1}^{N-1,N}$, at some instant when in real-world it is still under the free-flow regime.

Due to similar reasons, a premature or delayed switching problem usually also exists in determining the functional forms of the Jacobian matrix of the measurement model.

III. Methodology

A. A Heuristic Capacity Drop Mode Observer

To overcome the deficiency of the conventional CTM-EKF estimation approach identified above, we introduce a separate method to estimate the presence (or lack thereof) of a capacity drop. The method is responsible for determining when all the flow functions should use their free-flow functional forms, and when the functional form of S_{k-1}^{N}, and consequently the functional form of $q_{k-1}^{N,N-1}$, should switch from free-flow regime to congested regime and when to switch back, hence avoiding the trap of premature and delayed switches. The motivation behind this idea is to take advantage of the fact that as long as there is a queue present in cell N-1, there is an active reduction in the boundary flow rate between cell N-1 and cell N, i.e. a capacity drop, which can be relatively easy to be detected in real world.

Clearly, design of such an observer can be a non-trivial task by itself. However, the primary focus of this paper is to demonstrate the effectiveness of a modified CTM-EKF approach which takes into account capacity drop and provided that capacity drop mode can be satisfactorily captured. Therefore, in this paper, we use a heuristic algorithm developed based on micro-simulation data generated by Aimsun. The algorithm is presented in Table II.

TABLE II
A Heuristic Capacity Drop Mode Observer

Initialization: capacity_drop_indicator(0) = 0

Input: q^{thre1}, q^{thre2} and Δ

Output: capacity_drop_indicator (0 := capacity drop off; 1:= capacity drop on)

At each time step *k*:

Use upstream bottleneck cell boundary flows measured at step *k*-2, *k*-1, and *k*, i.e. $m\_q_{k-2}^{N-1,N}$, $m\_q_{k-1}^{N-1,N}$ and $m\_q_{k}^{N-1,N}$

if capacity_drop_indicator(*k*-1) = 0 & any of the below three cases is satisfied:

- $m\_q_{k}^{N-1,N} - m\_q_{k-1}^{N-1,N} < 0$ & $m\_q_{k-1}^{N-1,N} - m\_q_{k-2}^{N-1,N} < 0$ & $m\_q_{k-2}^{N-1,N} - m\_q_{k-3}^{N-1,N} < 0$
- $m\_q_{k}^{N-1,N} - m\_q_{k-1}^{N-1,N} < 0$ & $m\_q_{k-1}^{N-1,N} - m\_q_{k-2}^{N-1,N} < 0$ & $m\_q_{k}^{N-1,N} - m\_q_{k-2}^{N-1,N} < q^{\mathrm{thre1}}$
- $m\_q_{k}^{N-1,N} - m\_q_{k-1}^{N-1,N} < q^{\mathrm{thre2}}$

  capacity_drop_indicator(*k*) = 1

else if capacity_drop_indicator(*k*-1) = 1 & $\frac{\left| m\_q_{k}^{N-1,N} - m\_q_{k-1}^{N-1,N} \right|}{m\_q_{k-1}^{N-1,N}} > \Delta$

  capacity_drop_indicator(*k*) = 0

else

  capacity_drop_indicator(*k*) = capacity_drop_indicator(*k*-1)

end

In Table II, q^{thre1}, q^{thre2} and Δ are constant parameters to be calibrated offline. Through simulation experiments, we found that, under the same traffic demand pattern, the observer with a set of well-calibrated parameters was able to effectively capture the capacity drop modes for different random replications of simulation. This algorithm can only work for an incident-free environment. A more sophisticated, model-based observer is currently under our investigation.

B. State-Space Model Taking Advantage of The Capacity Drop Mode

We take advantage of the capacity drop mode, which is indicated by the capacity_drop_indicator defined in Table II, to determine the functional form of the time-varying $\mathbf{f}_{k-1}(\cdot)$ at each step. The determination of the functional form of $\mathbf{f}_{k-1}(\cdot)$ is composed of two steps, shown in Table III and Table IV, respectively. Again, note that in Table III and Table IV all the equations with a $\min\{\cdot\}$ is a shorthand in the same sense as stated in the paragraph below (19).

Being consistent with [15], the measurement model takes in the flows and space-mean speeds as measured by detectors stationed at cell boundaries (except for the in-flow detectors), as shown in Fig. 1. Thus the measurement model consists of 2N equations. The first N equations map the state variables to measured boundary flows, and the second N equations map the state variables to measured boundary space-mean speeds, i.e.

$$\begin{bmatrix} \mathbf{m\_q}_k^{i,i+1} \\ \mathbf{m\_v}_k^{i,i+1} \end{bmatrix} = \begin{bmatrix} \mathbf{h1}_k(\mathbf{x}_k) \\ \mathbf{h2}_k(\mathbf{x}_k) \end{bmatrix} + \boldsymbol{\varepsilon}_k \tag{20}$$

To determine the functional forms of $\mathbf{h1}_k(\mathbf{x}_k)$ and

$\mathbf{h2}_k(\mathbf{x}_k)$ at each step, we also take advantage of the capacity_drop_indicator, in a way analogous to as shown by Table III and Table IV. Due to space limit, we do not present the detail in this paper.

TABLE III
DETERMINATION OF FORMS OF DEMAND AND SUPPLY FUNCTIONS

1:	**for** i=1,2,…,N-1
2:	**if** capacity_drop_indicator = 0
3:	$D_{k-1}^{i} := \lambda_i v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{i}$
4:	$S_{k-1}^{i} := \lambda_i v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}}$
5:	**else**
6:	$D_{k-1}^{i} := \lambda_i v_{k-1}^{\mathrm{fr}}\, min\{\rho_{k-1}^{i}, \rho_{k-1}^{cr}\}$
7:	$S_{k-1}^{i} := \lambda_i v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}} \min\left\{1, \frac{\rho^{\mathrm{jam}}-\rho_{k-1}^{i}}{\rho^{\mathrm{jam}}-\rho_{k-1}^{\mathrm{cr}}}\right\}$
8:	**end**
9:	**end**
10:	**for** i=N
11:	$D_{k-1}^{i} = \lambda_i v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{i}$
12:	**if** capacity_drop_indicator = 0
13:	$S_{k-1}^{i} := \lambda_i v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}}$
14:	**else**
15:	$S_{k-1}^{i} := \lambda_i v_{k-1}^{\mathrm{fr}} \rho_{k-1}^{\mathrm{cr}} (1-\theta)$
16:	**end**
17:	**end**

TABLE IV
DETERMINATION OF FUNCTIONAL FORMS OF BOUNDARY FLOWS

1:	**for** i=1
2:	**if** capacity_drop_indicator = 0
3:	$q_{k-1}^{i-1,i} := q_{k-1}^{in}$
4:	**else**
5:	$q_{k-1}^{i-1,i} := \min\{q_{k-1}^{in}, S_{k-1}^{i}\}$
6:	**end**
7:	**end**
8:	**for** i=1,2,…,N-2
9:	**if** capacity_drop_indicator = 0
10:	$q_{k-1}^{i,i+1} := D_{k-1}^{i}$
11:	**else**
12:	$q_{k-1}^{i,i+1} := \min\{D_{k-1}^{i}, S_{k-1}^{i+1}\}$
13:	**end**
14:	**end**
15:	**for** i=N-1
16:	**if** capacity_drop_indicator = 0
17:	$q_{k-1}^{i,i+1} := D_{k-1}^{i}$
18:	**else**
19:	$q_{k-1}^{i,i+1} := S_{k-1}^{i+1}$
20:	**end**
21:	**end**
22:	**for** i=N
23:	$q_{k-1}^{i,i+1} := D_{k-1}^{i}$
24:	**end**

We explain the advantages of the capacity_drop_indicator. From Table III and Table IV, it can be seen that when the capacity_drop_indicator is 0 (i.e. no capacity drop), all the N cells use free-flow regime functions. This will prevent a premature or delayed switch of the process model from free-flow functional forms to congested functional forms. As soon as the capacity_drop_indicator turns 1 (i.e. capacity drop is on), Line 15 of Table III and Line 19 of Table IV render the critical density observable instantly, i.e. being able to be corrected from the biased initial value by measurements.

At each time step, with the $\mathbf{f}_{k-1}(\cdot)$ and $\mathbf{h}_k(\cdot)$ identified using the above method, the Jacobian matrices of the process and measurement models, $\mathbf{F}_{k-1}$ and $\mathbf{H}_k$, respectively, can be derived. With these information available, the EKF algorithm can be applied, which is presented as Table V. Note that in line 9 of Table V, q_{k-1}^{in} is the known inflow rate from the upstream boundary of the concerned section, as introduced in Section II. A. In Table V, **P** denotes the covariance of the estimation error, and **K** denotes the Kalman gain. Meanings of the other notations in Table V have been defined previously.

TABLE V
THE DISCRETE-TIME EXTENDED KALMAN FILTER [22]

1:	**Initialization:**
2:	$\hat{\mathbf{x}}_0^+ = E(\mathbf{x_0})$
3:	$\mathbf{P}_0^+ = E[(\mathbf{x_0} - \hat{\mathbf{x}}_0^+)(\mathbf{x_0} - \hat{\mathbf{x}}_0^+)^T]$
4:	**for** k = 1,2,3,…
5:	(a) Jacobian matrix of the process model
6:	$\mathbf{F}_{k-1} = \frac{\partial \mathbf{f}_{k-1}}{\partial \mathbf{x}}\Big\|_{\hat{\mathbf{x}}_{k-1}^+}$
7:	(b) Time update
8:	$\mathbf{P}_k^- = \mathbf{F}_{k-1}\mathbf{P}_{k-1}^+\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}$
9:	$\hat{\mathbf{x}}_k^- = \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1}^+, q_{k-1}^{in})$
10:	(c) Jacobian matrix of the measurement model
11:	$\mathbf{H}_k = \frac{\partial \mathbf{h}_k}{\partial \mathbf{x}}\Big\|_{\hat{\mathbf{x}}_k^-}$
12:	(d) Measurement update
13:	$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1}$
14:	$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k[\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{x}}_k^-)]$
15:	$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\,\mathbf{H}_k)\mathbf{P}_k^-$
16:	**end**

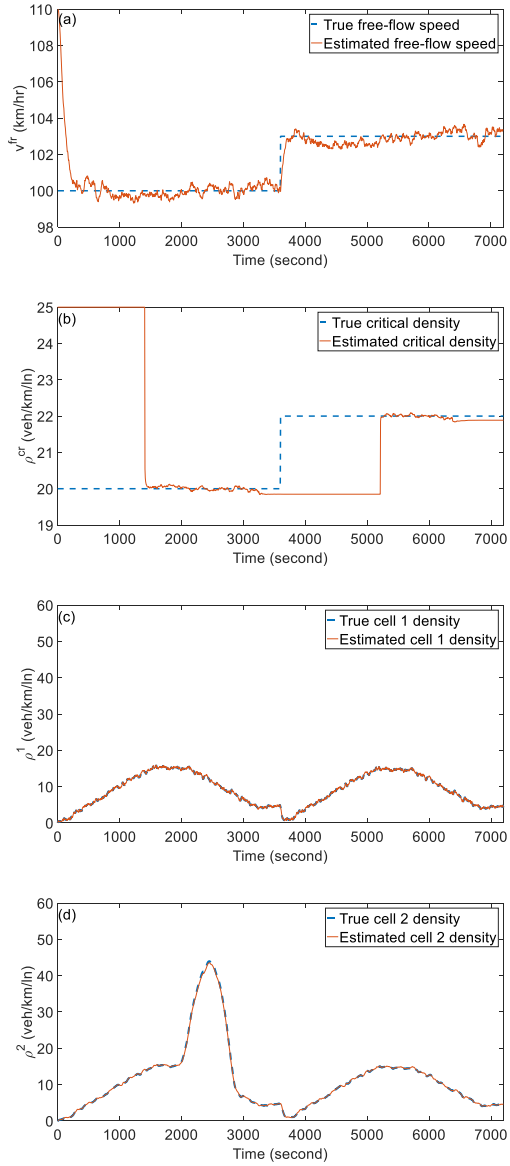
IV. VALIDATION OF METHODOLOGY

Due to dependence on linearization, EKF estimates could diverge or converge to wrong values when either or both of the process and measurement models are nonlinear in state variables [23]. In our case, both the process and measurement models are nonlinear. Thus, it is desired to first test whether results of our method converge to perfectly known true traffic states and parameters. Moreover, the dynamics switching logics in our state-space model has been simplified from those of the original CTM through taking advantage of the capacity_drop_indicator, hence, it is also desirable to examine whether our method can recover the true signals synthesized from the original CTM. For these considerations, we present a validation experiment in this section to demonstrate that the methodology is able to recover the parameters of a state-space simulation of a CTM.

Using the original CTM model as presented in Section II. A, we simulate the true signals. The simulation scenario is briefly described as follows. The simulation time is 2 hours, and the sampling interval is 5 seconds. The in-flow input is made such that two periods of congestion will be produced. The road section is divided into 5 cells, each cell being 700 m long. The first four cells consist of 3 lanes, and the last cell consists of 2 lanes. Parameters used for the simulation is as following: $v^{\mathrm{fr}} = 100$ km/hr for the 1st hour; $v^{\mathrm{fr}} = 102$ km/hr for the 2nd hour; $\rho^{\mathrm{cr}} = 20$ veh/km/ln for the 1st hour; $\rho^{\mathrm{cr}} = 22$ veh/km/ln for the 2nd hour; $\rho^{\mathrm{jam}} = 100$ veh/km/ln; $\theta = 10\%$. For the purpose of validation, in the truth simulation, the state variables are not corrupted by noise, while the measurements are corrupted by significant noise. This is in consistent with [15]. Since the capacity drop mode

is very easy to detect using CTM synthesized data, it is not meaningful to test it. Thus in this section we assume it is known. We will test the proposed capacity drop mode observer in the next section, where micro-simulation data are used.

The estimation results are presented by Fig. 2. Estimates of traffic densities of all the five cells match with the true values very well. Fig. 2 (a) shows that the estimator can track the abrupt change of the free-flow speed in a timely matter. Fig. 2 (b) shows that, as expected, the critical density is not observable when there is no congestion. However, as soon as a congestion is present, the estimator is able to converge to the true signal very quickly. This can be clearly seen by comparing Fig. 2 (b) against Fig. 2 (g). Particularly, note that the estimator is able to correct the purposely made large biases in the initial estimates of the traffic flow parameters in a timely manner.



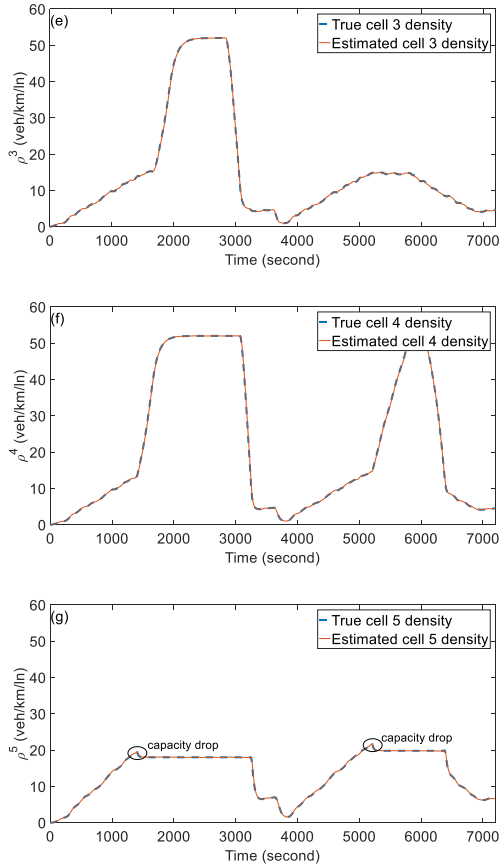


Fig. 2. (a) Free-flow speed; (b) Critical density; (c) Cell 1 density; (d) Cell 2 density; (e) Cell 3 density; (f) Cell 4 density; (g) Cell 5 density.

V. Experiment Using Micro-Simulation

To evaluate the effectiveness of the proposed method in an environment similar to real world, we test it using data generated from a commercial micro-simulation software, Aimsun.

The simulated highway section is divided into 5 cells, each cell being 400 meters long. Cell 1 through cell 4 each consists of 3 lanes, and cell 5 has 2 lanes. The simulation time is 90 minutes. One congestion period is produced. The "true" free-flow speed is specified by setting maximum-desired-driving-speed in micro-simulation. To be consistent with real-world applications, the measurement sampling time interval is set to be 30 seconds, but the estimation updating interval is set to be 10 seconds. The jam density is estimated offline to be 210 veh/km/ln. The "true" free-flow speed is 100 km/hr for 0 – 40 min, and 80 km/hr for the left of the simulation time. The capacity drop proportion is 23%, estimated offline.

Fig. 3 shows the measured boundary flow between cell 4 and cell 5, plotted against the capacity drop status captured by the capacity drop mode observer introduced earlier. Note that it is the boundary flow between the lane-dropped cell (i.e. cell 5 in this experiment) and its immediate upstream cell (i.e. cell 4 in this experiment) that can best reflect capacity drop. It can be seen from Fig. 3 that the captured capacity drop mode matches the truth satisfactorily. This was also visually verified by checking the simulation animation. Note that the

experiment has used measurements generated by a different random replication from those that have been used for calibrating the parameters of the capacity drop mode observer. As mentioned earlier, we have found that a set of well-calibrated parameters was sufficiently robust in capturing capacity drop mode under different random replications.

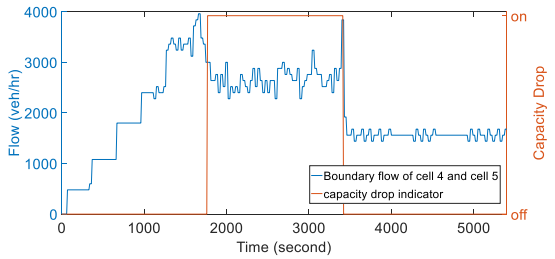
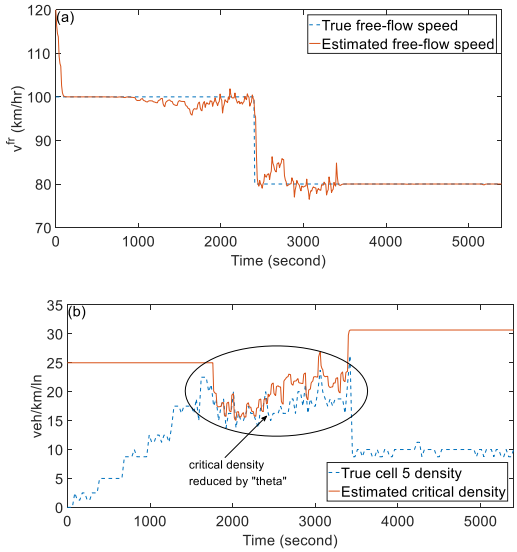


Fig. 3. Captured capacity drop mode vs. the boundary flow rate between cell 4 and cell 5.

Fig.4 (a) shows that the proposed method was able to track the time variations of free-flow speed. Particularly, we see that it was able to correct a very biased initial estimate and respond to an abrupt change in a timely manner.

Since it is difficult to specify critical densities in Aimsun, we were unable to directly simulate true signals for critical density. However, recognize such a fact, that cell 5 is the lane-dropped cell and no more restrictive bottleneck is downstream of it, so cell 5's density should be around critical density reduced by the capacity drop proportion, θ, when the capacity drop is active. Therefore, densities of cell 5 when its upstream is congested can provide reasonable inferences into the true critical densities. Thus in Fig. 4 (b), we plotted the estimated critical densities against the true cell 5 densities. As can be seen, the matching is satisfactory. This implies that the proposed method was able to unveil the true critical density from only traffic flow and speed measurements generated by a micro-simulation process for which we almost had no *a priori* knowledge at all of the critical density.



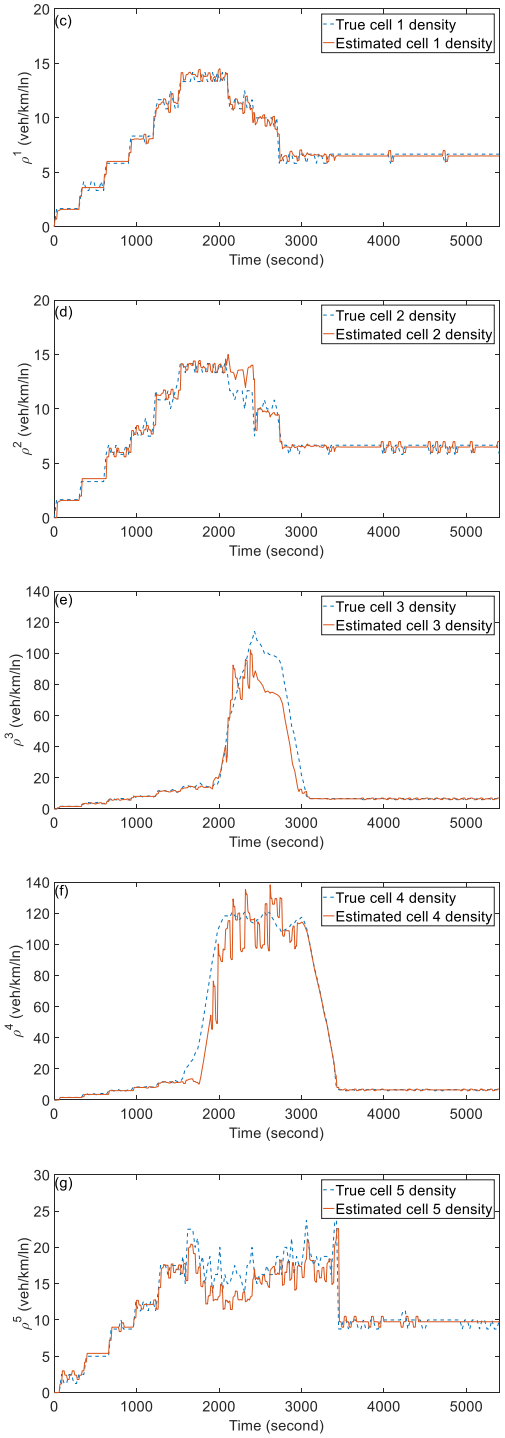


Fig. 4. (a) Free-flow speed; (b) Critical density; (c) Cell 1 density; (d) Cell 2 density; (e) Cell 3 density; (f) Cell 4 density; (g) Cell 5 density.

Fig. 4 (c) through (g) show the estimated densities versus the true densities of cell 1 to cell 5, respectively. Table VI presents the mean-absolute-percentage-errors (MAPE) of the estimated densities and the estimated free-flow speed, calculated against the true values. Note that here, ρ^i represents density for cell i. We see that the estimation results are reasonable.

TABLE VI
MEAN-ABSOLUTE-PERCENTAGE-ERRORS OF THE ESTIMATED DENSITIES AND FREE-FLOW SPEED

Quantity	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5	v^{fr}
MAPE (%)	4.61	6.66	9.90	10.18	11.05	0.92

VI. CONCLUSIONS AND FUTURE RESEARCH

This paper proposed a method to jointly estimate highway traffic state and traffic flow parameters using CTM and EKF, and considering capacity drop. The method takes advantage of a separate, supervisory observer which determines times to switch between plant dynamics of free-flow and congested regimes by capturing the on-and-off of the capacity drop. The proposed method has demonstrated satisfactory capability in tracking time variations of free-flow speed and critical density and estimating traffic densities in real time. Future research includes an evaluation of the proposed method by comparison against the conventional CTM-EKF approach, development of a more sophisticated, model-based capacity drop mode observer that can work under a wide variety of demand patterns and requires little offline calibration, and application of the proposed estimation method to traffic control.

REFERENCES

[1] D. B. Work, O.-P. Tossavainen, S. Blandin, A. M. Bayen, T. Iwuchukwu, and K. Tracton, "An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, 2008, pp. 5062-5068: IEEE.

[2] T. Seo, T. T. Tchrakian, S. Zhuk, and A. M. Bayen, "Filter comparison for estimation on discretized PDEs modeling traffic: Ensemble Kalman filter and Minimax filter," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*, 2016, pp. 3979-3984: IEEE.

[3] L. Mihaylova, R. Boel, and A. Hegyi, "Freeway traffic estimation within particle filtering framework," *Automatica*, vol. 43, no. 2, pp. 290-300, 2007.

[4] I.-C. Morărescu and C. Canudas-de-Wit, "Highway traffic model-based density estimation," in *American Control Conference (ACC), 2011*, 2011, pp. 2012-2017: IEEE.

[5] X. Sun, L. Muñoz, and R. Horowitz, "Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, 2003, vol. 6, pp. 6333-6338: IEEE.

[6] J. Thai, B. Prodhomme, and A. M. Bayen, "State estimation for the discretized LWR PDE using explicit polyhedral representations of the Godunov scheme," in *American Control Conference (ACC), 2013*, 2013, pp. 2428-2435: IEEE.

[7] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez, "Traffic density estimation with the cell transmission model," in *American Control Conference, 2003. Proceedings of the 2003*, 2003, vol. 5, pp. 3750-3755: IEEE.

[8] T. Seo and A. M. Bayen, "Traffic state estimation method with efficient data fusion based on the Aw-Rascle-Zhang model," in *Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on*, 2017, pp. 1-6: IEEE.

[9] C. Nanthawichit, T. Nakatsuji, and H. Suzuki, "Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1855, pp. 49-59, 2003.

[10] Y. Han and S. Ahn, "Stochastic modeling of breakdown at freeway merge bottleneck and traffic control method using connected automated vehicle," *Transportation Research Part B: Methodological*, vol. 107, pp. 146-166, 2018.

[11] C. M. Tampère and L. Immers, "An extended Kalman filter application for traffic state estimation using CTM with implicit mode switching and dynamic parameters," in *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, 2007, pp. 209-216: IEEE.

[12] A. Nantes, D. Ngoduy, A. Bhaskar, M. Miska, and E. Chung, "Real-time traffic state estimation in urban corridors from heterogeneous data," *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 99-118, 2016.

[13] K. Ozbay, I. Yasar, and P. Kachroo, "Development and Evaluation of On-Line Estimation Methods for Feedback Based Freeway Ramp Metering Strategy," in *in 85th TRB Annual Meeting*, 2006.

[14] K. Ozbay, I. Yasar, and P. Kachroo, "Improved online estimation methods for a feedback-based freeway ramp metering strategy," in *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, 2006, pp. 412-417: IEEE.

[15] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141-167, 2005.

[16] S. K. Godunov, "A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics," *Matematicheskii Sbornik*, vol. 89, no. 3, pp. 271-306, 1959.

[17] S. S. Kandala, "Analysis of freeway bottlenecks," Arizona State University, 2014.

[18] A. Srivastava and N. Geroliminis, "Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model," *Transportation Research Part C: Emerging Technologies*, vol. 30, pp. 161-177, 2013.

[19] M. Hadiuzzaman and T. Z. Qiu, "Cell transmission model based variable speed limit control for freeways," *Canadian Journal of Civil Engineering*, vol. 40, no. 1, pp. 46-56, 2013.

[20] Y. Zhang and P. A. Ioannou, "Combined variable speed limit and lane change control for highway traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1812-1823, 2017.

[21] H.-Y. Jin and W.-L. Jin, "Control of a lane-drop bottleneck through variable speed limits," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 568-584, 2015.

[22] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

[23] L. Perea, J. How, L. Breger, and P. Elosegui, "Nonlinearity in sensor fusion: divergence issues in ekf, modified truncated gsf, and ukf," in *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2007, p. 6514.