Traffic State Estimation for Connected Vehicles Using the Second-Order Aw-Rascle-Zhang Traffic Model

Suyash C. Vishnoi, Sebastian A. Nugroho, *Member, IEEE*, Ahmad F. Taha, *Member, IEEE*, and Christian G. Claudel

***Abstract*— This paper addresses the problem of traffic state estimation (TSE) in the presence of heterogeneous sensors which include both fixed and moving sensors. Traditional fixed sensors are expensive and cannot be installed throughout the highway. Moving sensors such as Connected Vehicles (CVs) offer a relatively cheap alternative to measure traffic states across the network. Moving forward it is thus important to develop such models that effectively use the data from CVs. One such model is the nonlinear second-order Aw-Rascle-Zhang (ARZ) model which is a realistic traffic model, reliable for TSE and control. A state-space formulation is presented for the ARZ model considering junctions in the formulation which is important to model real highways with ramps. A Moving Horizon Estimation (MHE) implementation is presented for TSE using a linearized ARZ model. Various state-estimation methods used for TSE in the literature along with the presented approach are compared with regard to accuracy and computational tractability with the help of a numerical study using the VISSIM traffic simulation software. The impact of various strategies for querying CV data on the estimation performance is also considered. Several research questions are posed and addressed with a thorough analysis of the results.**

***Index Terms*— Traffic state estimation, highway traffic networks, second-order models, Aw-Rascle-Zhang model, moving horizon estimation, connected vehicles.**

I. Motivation and Paper Contributions

With the large number of vehicles overloading the transportation systems across the world, problems like congestion, accidents, and pollution have become common. As a remedy to such circumstances, control methods such as variable speed limits, ramp metering, route control and their combinations have become quite popular for instance see [1], [2], [3], and [4].

These methods require the knowledge of the system at all times to make them work effectively. A popular method for real-time monitoring of traffic systems is by means of traffic state estimation (TSE) using dynamic traffic models which provide a high-fidelity picture of the traffic spatio-temporally while utilizing data from sensors available throughout the highway. In general, more data results in better estimates of the system states. However, since fixed sensors like inductive loop detectors are quite expensive, they cannot be placed at short intervals throughout the highway. Connected vehicles (CVs) offer a potential solution to this problem by providing additional sources of data relatively free of cost [5]. Here, we assume that most of the communication between the CVs and the network operator will take place via existing cellular networks so there will be no additional costs of building connected highway infrastructure everywhere. As the proportion of CVs in the traffic rises, CVs will be able to provide useful data from across the system including both traffic density and speed. Thus, moving forward, it is imperative to develop such models that can utilize well different types of data from both fixed sensors and CVs to perform state estimation and control.

Traditionally, TSE is performed using first-order traffic models such as the Lighthill-Whitham-Richards (LWR) model [6], [7]. First-order models are simple to implement as they only have a single equation which is the conservation of vehicles to describe the traffic dynamics. They also have very few calibration parameters, making them a popular choice for state estimation. However, they only consider equilibrium traffic conditions, that is, the traffic density (number of vehicles per unit space expressed in vehicles per unit length for example, veh/km [8]) and traffic flux (number of vehicles that cross a given point per unit of time expressed in vehicles per unit time for example, veh/hr [8]) are assumed to follow a predefined relationship known as the *fundamental diagram*. This makes them unable to represent certain non-equilibrium traffic phenomena like capacity drop which are essential for the purpose of traffic control [9]. Thus, the use of these models in traffic control is considered less effective. Second-order traffic models such as the Payne-Whitham (PW) model [10], [11] and the Aw-Rascle-Zhang (ARZ) model [12], [13], on the other hand, can represent non-equilibrium traffic phenomena with the help of an additional equation to describe the traffic

Manuscript received 25 August 2023; revised 23 February 2024; accepted 3 June 2024. Date of publication 31 July 2024; date of current version 1 November 2024. This work was supported in part by Valero Energy Corporation; in part by the National Science Foundation (NSF) under Grant 1636154, Grant 1728629, Grant 2152928, Grant 1917056, and Grant 2152450; and in part by United States Department of Transportation (USDOT) Center for Advanced Multimodel Mobility Solutions and Education (CAMMSE). The Associate Editor for this article was S. Timotheou. *(Corresponding author: Suyash C. Vishnoi.)*

Suyash C. Vishnoi is with the School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: scvishnoi@gatech.edu).

Sebastian A. Nugroho is with the Cummins Technical Center, Cummins Inc., Columbus, IN 47201 USA (e-mail: sebastian.nugroho@cummins.com).

Ahmad F. Taha is with the Department of Civil and Environmental Engineering, Vanderbilt University, Nashville, TN 37235 USA (e-mail: ahmad.taha@vanderbilt.edu).

Christian G. Claudel is with the Department of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: christian.claudel@utexas.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/TITS.2024.3420438, provided by the authors.

Digital Object Identifier 10.1109/TITS.2024.3420438

dynamics. They are, therefore, considered more realistic than the first-order models. As a result, these models are not only good for state estimation but are also reliable for control. Additionally, second-order models provide a natural way to incorporate multiple sources of data as they consider both density and speed to be independent variables. In first-order models that only consider either the density or the speed as a variable at a time, any deviation of the speed from its equilibrium relationship must be considered a part of the modeling error. Thus, second-order models become a natural choice for state estimation using CVs. Note that while Lagrangian (vehicle-based) models of traffic exist [14] which are arguably more compatible with trajectory-based data from CVs, here we are using road density and average vehicle speed information obtained using both CVs and fixed detectors and not just relying on trajectory based information. Therefore, we have chosen an Eulerian (location-based) second-order model over a Lagrangian model besides the above reasons.

In light of the aforementioned discussion, the objective of this work is to develop a state-space representation of a reliable second-order traffic model and show the potential of CV data for TSE under different scenarios. Given this objective, in the following discussion, we present a literature review of traffic models used for TSE and studies utilizing heterogeneous sources of data, followed by a brief discussion on the estimation methods used.

The most popular model for TSE in the literature is the first-order LWR model. The simple form of the model with a minimal number of calibration parameters makes it an attractive option for large-scale implementation. Some works that implement a first-order model to perform state estimation using heterogeneous sensors include [15], [16], [17]. Readers can also refer to [18] for a comprehensive review of TSE literature involving first-order models. Due to the known limitations of first-order models, several studies have also undertaken state estimation using second-order models such as in [19], [20], and [21] and the references therein. Most of these studies use the second-order PW model implemented in the METANET [22], [23] framework. The PW model has well-known limitations [9] such as physical inconsistency under certain heterogeneous traffic conditions which make it unreliable. A significantly better model is the ARZ model which retains the benefits of second-order models without sacrificing the physical consistency of the first-order models. Despite this, there are very few studies in the literature that use the ARZ model for state estimation. The work in [24] develops a state-space formulation for the nonlinear ARZ model and performs state estimation using Extended Kalman Filter (EKF) considering both fixed and moving sensors. In [25], the authors propose a boundary observer for state estimation using a linearized ARZ model. The study in [26] uses Particle Filter (PF) for the estimation of traffic states using a modified ARZ model. However, it is worth noting that none of these papers considers junctions in the modeling. Modeling the traffic dynamics at junctions is essential to the modeling of traffic on real highways which consist of on-ramp and off-ramp connections. Therefore, unlike past studies, we formulate herein a state-space model for the nonlinear ARZ model considering junctions.

Note that the aforementioned studies using second-order models as well as the present work are different from studies like [27] which while do consider the speed to be an independent variable like the second-order models but consider it to be known everywhere and at all times using CV data. These have been categorized as *data-driven* methods by [18]. In the current work, unlike [27], we assume a bandwidth restriction on the data that can be transferred from the CVs to the network operator which forces data to be available only from a subset of all segments for estimation while the traffic data on other segments is considered unknown. Further, the impact of various strategies associated with the selection of the subset of segments to query CV data for estimation is also investigated. Besides [27], several other studies utilize heterogeneous data sources for TSE. Detailed reviews of the related literature can be found in [18], [28], and [29]. Most studies focus on data fusion methods to combine fixed and moving sensor data to achieve improved estimation performance. Some studies such as [24] focus on the impact of the penetration rate of CVs on TSE. To the best of the authors' knowledge, none of the studies investigate the impact of different approaches to query subsets of segments for estimation in a moving sensor setting.

A majority of the model-driven TSE literature either uses one of the Kalman Filter (KF) variants from among EKF, Unscented Kalman Filter (UKF), and Ensemble Kalman Filter (EnKF), or other methods like PF, and observers to perform state estimation, for instance, see [18], [30], and [31]. While these methods are computationally attractive, they have certain limitations with respect to TSE. The primary limitation is that they do not have an inherent way to deal with state constraints. Thus, it is possible that the estimates generated from these methods contain nonphysical values of certain states which can further cause the process model to collapse.

An estimation method that handles this limitation naturally, due to its optimization-based structure, is MHE. MHE has been explored extensively in the general state estimation literature, for instance in [32], [33], [34], and [35], but not so much in the TSE literature. In [36] and [37], the authors propose an MHE formulation for the estimation and control of large-scale highway networks using the Macroscopic fundamental diagram (MFD). MFD is a network-level traffic model and does not consider the variation in traffic density on individual stretches of the highway. Unlike [36] and [37], we investigate estimating the density throughout the highway stretch. The study in [38] presents an MHE formulation for traffic density estimation using the Asymmetric Cell Transmission model (ACTM). ACTM is based on the LWR model and therefore, has the drawbacks previously mentioned for first-order models. Moreover, the work in [38] does not consider moving sensors from CVs.

Besides the above approaches, a recent paradigm of TSE explores physics-informed deep learning (DL) [39], [40], [41]. These approaches aim to guide the training of DL-based models for TSE through physics-based traffic laws, such as those governing the first and second-order models discussed above. The traffic model parameters are automatically tuned as the DL model is trained, thus offering the determination of accurate traffic flow laws for a given scenario. However, these approaches suffer from several limitations, including a

lack of robustness to noisy data and the need for extensive tuning of training algorithm parameters for individual scenarios, limiting the models' applicability to real-world use cases. Interested readers are referred to [42] for a comprehensive survey on this paradigm of TSE. Compared to these, model-driven approaches such as the one presented in this work are favored for real-world applications due to their interpretability and computational advantage.

Given that, the main research gaps on this topic are a) the absence of a state-space formulation for a reliable second-order traffic model with junctions, b) the lack of exploration of MHE in the context of TSE and comparison with other state-estimation methods, and c) the absence of an investigative study on the impact of different strategies for querying data from CVs for TSE as opposed to fixed sensors. In what follows, we highlight the main contributions of this paper:

- We derive a nonlinear state-space formulation for the second-order ARZ model with junctions in the form of ramp connections. In that, we present the detailed dynamic equations of the model. This is a development over [24] which does not consider junctions in the formulation. The inclusion of junctions adds additional complexity to the model in terms of the nonlinearity which now comprises of minimum and piecewise functions in the model. Second-order traffic models are more realistic than first-order models like the LWR model as they can capture certain phenomena like capacity drop which are essential to control applications. The obtained state-space formulation can thus be used for state estimation as well as control purposes.
- We consider heterogeneous sensors including both fixed and moving sensors. The former consists of sensors like inductive loop detectors while the latter includes CVs. The state-space description is appended to include the measurement model which is also nonlinear thus resulting in a nonlinear input-output mapping of the system dynamics.
- We investigate the performance of various state estimation methods in terms of accuracy and computational tractability using the VISSIM traffic simulation software. As a departure from estimation based on KFs, PF, observers, and so on, we investigate MHE for TSE. MHE, unlike the other methods, naturally allows us to include constraints on the state variables making the problem more practical.
- The impact of moving sensors including CVs on the performance of TSE is studied under various scenarios including different frequencies of change in sensor positions, different sensor placement configurations, and different levels of measurement errors from varying penetration rate of CVs and sensor noise. The estimated states are examined qualitatively to understand the implication of moving sensors on TSE.

Paper's Notation: Let $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{R}^{p\times q}$ denote the set of natural numbers, real numbers, and real-valued column vectors with size n, and p-by-q real matrices respectively. $\mathbb{S}^m_{++}$ denotes the set of positive definite matrices. For any vector $z \in \mathbb{R}^n$, $\|z\|_2$ denotes its Euclidean norm, i.e. $\|z\|_2 = \sqrt{z^\top z}$, where $z^\top$ is the transpose of z. Tab. I provides the nomenclature utilized in this paper.

TABLE I

PAPER NOMENCLATURE: PARAMETER, VARIABLE, AND SET DEFINITIONS

Notation	Description
Ω	the set of highway segments on the stretched highway $\Omega = \{1, 2, \ldots, N\}$, $N := \|\Omega\|$
Ω_I	the set of highway segments with on-ramps $\Omega_I = \{1, 2, \ldots, N_I\}$, $N_I := \|\Omega_I\|$
Ω_O	the set of highway segments with off-ramps $\Omega_O = \{1, 2, \ldots, N_O\}$, $N_O := \|\Omega_O\|$
$\hat{\Omega}$	the set of on-ramps, $\hat{\Omega} = \{1, 2, \ldots, N_I\}$, $N_I = \|\hat{\Omega}\|$
$\check{\Omega}$	the set of off-ramps, $\check{\Omega} = \{1, 2, \ldots, N_O\}$, $N_O = \|\check{\Omega}\|$
T	duration of each time-step
l	length of each segment, on-ramp, and off-ramp
$\rho_i[k], \psi_i[k], w_i[k]$	traffic density, relative flow and driver characteristic for Segment $i \in \Omega$ at time kT, $k \in \mathbb{N}$
$q_i[k], \phi_i[k]$	traffic flow and relative flux from Segment $i \in \Omega$ into the next segment
$D_i[k], S_i[k]$	demand and supply functions for Segment $i \in \Omega$
$\hat{\rho}_i[k], \hat{\psi}_i[k], \hat{w}_i[k]$	traffic density, relative flow and driver characteristic for On-ramp $i \in \hat{\Omega}$ at time kT, $k \in \mathbb{N}$
$\hat{q}_i[k], \hat{\phi}_i[k]$	traffic flow and relative flux from On-ramp $i \in \hat{\Omega}$ into the attached highway segment
$\hat{D}_i[k], \hat{S}_i[k]$	demand and supply functions for On-ramp $i \in \hat{\Omega}$
$\check{\rho}_i[k], \check{\psi}_i[k], \check{w}_i[k]$	traffic density, relative flow and driver characteristic for Off-ramp $i \in \check{\Omega}$ at time kT, $k \in \mathbb{N}$
$\check{q}_i[k], \check{\phi}_i[k]$	traffic flow and relative flux from Off-ramp $i \in \check{\Omega}$
$\check{D}_i[k], \check{S}_i[k]$	demand and supply functions for Off-ramp $i \in \check{\Omega}$
$\bar{q}_i[k], \bar{\phi}_i[k]$	incoming traffic flow and relative flux for Segment $i \in \Omega$
$\bar{\check{q}}_i[k], \bar{\check{\phi}}_i[k]$	incoming traffic flow and traffic flux for Off-ramp $i \in \check{\Omega}$
$D_{in}[k], w_{in}[k]$	demand and driver characteristic of traffic wanting to enter Segment 1 of the highway
$\rho_{out}[k]$	traffic density downstream of Segment N of the highway
$\hat{D}_{in,i}[k], \hat{w}_{in,i}[k]$	demand and driver characteristic of traffic wanting to enter On-ramp $i \in \hat{\Omega}$
$\check{\rho}_{out,i}[k]$	traffic density downstream of Off-ramp $i \in \check{\Omega}$
$\beta_i[k]$	proportion of traffic entering from Segment $i \in \Omega$ into the next segment at an on-ramp junction, where $\beta_i[k] \in [0, 1]$
$\alpha_i[k]$	split ratio for the off-ramp attached to Segment $i \in \Omega$, where $\alpha_i[k] \in [0, 1]$
v_f	free-flow speed
ρ_m	maximum density
α	model parameter called relaxation time, where $\alpha \in \mathbb{R}_+$
γ	fundamental diagram parameter, where $\gamma \in \mathbb{R}_+$
$p(\rho)$	pressure function which takes traffic density ρ as input
$V_e(\rho)$	equilibrium traffic speed at traffic density ρ

II. NONLINEAR DISCRETE-TIME MODELING OF TRAFFIC NETWORKS WITH RAMPS

The objective of this section is to develop a state-space formulation for the nonlinear second-order ARZ model describing the evolution of traffic density on highways with ramps. The developed formulation is useful for several control theoretic purposes including state estimation and control of highway traffic.

A. The Aw-Rascle-Zhang Model

In this section, we present the modeling of traffic dynamics for a stretched highway connected with ramps. To that end, we use the second-order ARZ Model [12], [13] given by the following partial differential equations:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho v}{\partial d} = 0, \tag{1a}$$

$$\frac{\partial \rho\,(v + p\,(\rho))}{\partial t} + \frac{\partial \rho\,(v + p\,(\rho))\,v}{\partial d} = -\frac{\rho\,(v - V_e\,(\rho))}{\tau}, \tag{1b}$$

where t and d denote the time and distance; ρ is shorthand for $\rho(t,d)$ which denotes the traffic density (vehicles/distance), and v is shorthand for $v(t,d)$ which denotes the traffic speed (distance/time). Here, $p(\rho)$ is given by

$$p(\rho) = v_f \left(\frac{\rho}{\rho_m}\right)^{\gamma}, \tag{2}$$

and $V_e(\rho)$ is given by

$$V_e(\rho) = v_f \left(1 - \left(\frac{\rho}{\rho_m}\right)^{\gamma}\right). \tag{3}$$

In traffic literature, relationships like (3) are commonly called the *fundamental diagram*. The first PDE in the ARZ model ensures the conservation of vehicles which is also present in the first-order traffic models. The second PDE which ensures conservation of traffic momentum is unique to second-order models and accounts for the deviation of traffic from an equilibrium position. This equation makes the second-order models more realistic than the first-order models as it allows them to represent some non-equilibrium traffic phenomena such as capacity drop. As second-order models allow traffic flow to deviate from equilibrium, they also inherently allow traffic speed to deviate from the equilibrium speed which allows speed data to be incorporated independent of the density. With first-order models, any deviation of the speed from the equilibrium speed would have to be considered a part of the modeling error. Therefore, second-order models are more naturally suited to perform estimation using both density and speed data provided by the fixed sensors and CVs. The quantity $v+p(\rho)$ is also called the *driver characteristic* and is denoted by the variable $w(t,d)$. The expression $\rho(v+p(\rho))$ is also called the *relative flow* denoted by $\psi(t,d)$ which is essentially the difference between the actual flow and the equilibrium flow at any ρ. Notice that in (1), ρv is the flux of traffic (vehicles/time) which will be denoted by $q(t,d)$, while $\rho(v+p(\rho))v$ is the flux of relative flow (vehicles/time^2), also called the relative flux, which will be denoted by $\phi(t,d)$. Using the relative flow and the two flux, the ARZ model can simply be rewritten as

$$\frac{\partial \rho(t,d)}{\partial t} + \frac{\partial q(t,d)}{\partial d} = 0, \tag{4a}$$

$$\frac{\partial \psi(t,d)}{\partial t} + \frac{\partial \phi(t,d)}{\partial d} = -\frac{\psi(t,d)}{\tau} + \frac{v_f \rho(t,d)}{\tau}, \tag{4b}$$

which can be converted to a state-space equation with ρ and ψ as the states.

To represent this model as a series of difference, state-space equations, we discretize the ARZ Model (4) with respect to both space and time, also referred to as the Godunov scheme [43]. This allows us to divide the highway of length L into segments of equal length l and the traffic networks model to be represented by discrete-time equations. These segments form both the highway and the attached ramps. Throughout the paper, the segments forming the highway are referred to as mainline segments. We assume the highway is split into N mainline segments.

To ensure computational stability, the Courant-Friedrichs-Lewy condition (CFL) [44] given as $v_f T l^{-1} \leq 1$ has to be satisfied. Since each segment is of the same length l, then we have $\rho(t,d) = \rho(kT, il)$, where $i = 1, 2, \ldots, N$ represents the segment index, and $k \in \mathbb{N}$ represents the discrete-time index. For simplicity, we define $\rho(kT, il) := \rho_i[k]$. The other variables are also defined in the same way, namely $w_i[k]$, $\psi_i[k]$, $q_i[k]$, and $\phi_i[k]$. The expressions for the flux function $q_i[k]$ and $\phi_i[k]$ for Segment i depend on the arrangement of the segments before and after that segment. Mathematical expressions for the flux across different types of segment junctions and those for the traffic demand and supply functions needed to define the flux are omitted for brevity. Interested readers are referred to Appendix A of [45] for the same. Here, the demand of a segment denotes the traffic flux that wants to leave that segment while the supply of a segment denotes the traffic flux that can enter that segment.

B. State-Space Equations

The discrete-time traffic flow and relative flow conservation equations for any Segment $i \in \Omega$ can be written as

$$\rho_i[k+1] = \rho_i[k] + \frac{T}{l}(q_{i-1}[k] - q_i[k]), \tag{5a}$$

$$\psi_i[k+1] = \left(1 - \frac{1}{\tau}\right)\psi_i[k] + \frac{T}{l}(\phi_{i-1}[k] - \phi_i[k]) + \frac{v_f}{\tau}\rho_i[k] \tag{5b}$$

Similar equations can be written for ramp segments as well. Here, $q_i[k]$ and $\phi_i[k]$ take the expressions presented in Appendix A of [45] depending upon the arrangement of Segment i with respect to other segments. The state vector for this system can be defined as

$$\boldsymbol{x}[k] := [\rho_i[k]\psi_i[k] \ldots \hat{\rho}_j[k]\hat{\psi}_j[k] \ldots \check{\rho}_l[k]\check{\psi}_l[k] \ldots]^{\top} \in \mathbb{R}^{2(N+N_I+N_O)},$$

for which $i \in \Omega$, $j \in \hat{\Omega}$ and $l \in \check{\Omega}$. In this work, we assume that the demand and the driver characteristic upstream of the first mainline segment are known, that is, $D_0[k] = D_{in}[k]$ and $w_0[k] = w_{in}[k]$ and the density downstream of the last mainline segment is also assumed to be known, that is $\rho_{N+1}[k] = \rho_{out}[k]$. Similarly, the demand and driver characteristic upstream of the on-ramps and the density downstream of the off-ramps is also considered to be known. These values can be obtained using conventional detectors like the inductive loop detectors placed upstream of the input segments and downstream of the output segments of the highway. An approximate value of the demand can also be obtained using Origin-Destination flow matrices [46] if available for the given region. Then,

$$\boldsymbol{u}[k] := [D_{in}[k]w_{in}[k]\rho_{out}[k] \ldots \hat{D}_{in,j}[k]\hat{w}_{in,j}[k] \ldots \check{\rho}_{out,l}[k] \ldots]^{\top} \in \mathbb{R}^{3+2N_I+N_O},$$

where $j \in \hat{\Omega}$ and $l \in \check{\Omega}$.

The evolution of traffic density and relative flow described in (5) can be written in a compact state-space form as follows

$$\boxed{\boldsymbol{x}[k+1] = \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{G}\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}),} \tag{6}$$

where $\boldsymbol{A} \in \mathbb{R}^{n_x \times n_x}$ for $n_x := 2(N + N_I + N_O)$ represents the linear dynamics of the system, $\boldsymbol{f} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ for $n_u = 3 + 2N_I + N_O$ is a vector-valued function

representing nonlinearities in the state-space equation and $\boldsymbol{G} \in \mathbb{R}^{n_x \times n_x}$ is a matrix representing the distribution of nonlinearities.

The nonlinearities in $\boldsymbol{f}$ are in the form of a minimum of weighted nonlinear functions of the states and inputs. The structure of the above-mentioned matrices and functions is provided in Appendix B of [45]. Next, we discuss the measurement model for the ARZ model which is also nonlinear in nature.

C. Sensor Data and Measurement Model

We consider two types of sensors in this work, the first is fixed sensors like the inductive loop detectors, and the second is moving sensors which include CVs. This study assumes that it is possible to retrieve density and speed data from both types of sensors. Two loop detectors installed at opposite ends of a segment can be used to obtain the traffic density (using an approach similar to [47]) as well as the average speed of vehicles on the segment [48]. CVs are known to provide the current position and speed data for individual vehicles directly. The average speed of a segment can be assumed to be the average of the speed data provided by all the queried CVs in that segment similar to [49]. To obtain density data from CVs, we assume additional functionality including either spacing measurement equipment which is available as part of advanced driver assistance systems [50] or availability of vehicular ad-hoc networks (VANETs) which allow vehicles to communicate with each other in a neighborhood around the queried CV [51]. When assuming the latter it is important to note the limitation imposed by the communication range of the vehicles on the maximum cell length for traffic modeling. In the case of the former, while a cell length limitation may not be required, sufficient penetration of CVs is necessary on the segments that are queried for data. The data from the CVs is sent via cellular network to a network operator who performs any prior computation if necessary to convert the received information like the spacing data or neighborhood counts into density measurements before using them for state estimation. A measurement error can also be associated with the data at this point based on the available information on penetration rate and other factors.

Note that in this setting, both fixed sensors and CVs are assumed to provide similar data on the density and speed of traffic on segments. Traditionally, CVs are considered akin to floating cars which provide only trajectory information at high sampling rates and with a broader spatial coverage as compared to fixed sensors. However, with the increasing number of vehicles and devices capable of sending and receiving data over the internet allowing vehicles and objects to communicate with each other such as in the case of VANETS, it is reasonable to expect that CVs could provide data comparable to fixed sensors in quality and type but superior in spatial coverage allowing similar data retrieval over the entire road stretch rather than a few fixed segments. Also, CVs being multi-functional and mobile require a lower commitment than fixed sensors.

Figure 1 presents a schematic of the sensors' placement on the highway. Among the measurements, density $\rho_i[k]$ for any mainline segment $i \in \Omega$, and similarly for the ramps,

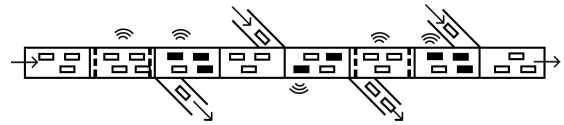

Fig. 1. Heterogenous sensors on the highway: fixed sensors represented by dashed lines across the highway and CVs represented by the solid black rectangles.

is directly a state and is used as it is, while the velocity $v_i[k]$ can be written in terms of the states as follows:

$$v_i[k] = \frac{\psi_i[k]}{\rho_i[k]} - p(\rho_i[k]).$$

We define a nonlinear vector-valued measurement function $\boldsymbol{h}(\boldsymbol{x}[k])$ which maps the state vector to a corresponding vector of all possible measurements in the system such that $\boldsymbol{h}(\boldsymbol{x}[k]) \in \mathbb{R}^{n_x}$. Note, that here the number of possible measurements is equal to the number of states in the system. For any mainline segment $i \in \Omega$, the corresponding measurements are denoted by h_{2i-1} and h_{2i}, where the subscripts represent the position of the element in the measurement function vector. These represent the density and speed of traffic on the segment, respectively. As mentioned above, these can be computed using the states corresponding to Segment i that is x_{2i-1} and x_{2i} representing the density and relative flow, respectively, and are defined as follows:

$$h_{2i-1}(\boldsymbol{x}[k]) = x_{2i-1}[k], \tag{7a}$$

$$h_{2i}(\boldsymbol{x}[k]) = \frac{x_{2i}[k]}{x_{2i-1}[k]} - p(x_{2i-1}[k]). \tag{7b}$$

The mapping corresponding to the ramp segments can also be defined similarly. Now, we can define the measurement vector $\boldsymbol{y}[k] \in n_p[k]$, where $n_p[k]$ is the number of available measurements from sensors at time k, as follows:

$$\boxed{\boldsymbol{y}[k] = \boldsymbol{C}[k]\boldsymbol{h}(\boldsymbol{x}[k]) + \boldsymbol{v}[k],}$$

where $\boldsymbol{C}[k] \in \mathbb{R}^{n_p[k] \times n_x}$ is the observation matrix at time k describing the availability of measurements from sensors. Note, that the observation matrix here is variable in time because of the measurements from CVs which allow data to be measured from different numbers and positions of segments with time. Here, $\boldsymbol{v}[k] \in \mathbb{R}^{n_v[k]}, n_v[k] = n_p[k]$ lumps all the measurement errors including the sensor noise into a single vector.

The above results are important as they allow us to perform state estimation for traffic systems using the second-order ARZ model. The state-space equation (6) can also be used for control purposes using control theoretic approaches from the literature. In the following section, we discuss a method for linearization of nonlinear functions which allows us to apply some linear state estimation methods to the otherwise nonlinear ARZ model.

D. Linear Model Approximation

The ARZ model specified in Section II-A is nonlinear due to the presence of the piecewise linear and nonlinear expressions in the traffic flux and relative flux terms. This prevents directly using some of the well-known and efficient

linear state estimation methods from the literature. However, it is still possible to apply linear state estimation methods to a linearized version of the ARZ model. Methods such as Taylor series expansion [52] can be used to obtain a good linear approximation of nonlinear functions about a suitable operating point. The detailed equations for linearization are omitted for brevity. The same can be found in Appendix C of [45].

Note that linearization is usually associated with reduced model accuracy and hence worse estimation performance as compared to using the nonlinear model when the model is an exact representation of the system. However, when the process is not exactly governed by the model dynamics, linearization may not necessarily result in a degradation of the estimation results. The latter is particularly relevant when using traffic flow models such as the ARZ model for estimation as they only focus on the aggregate behavior of traffic and do not capture the nuances of vehicle-to-vehicle interaction. A validation study against real-world traffic data similar to [53] is required to quantify the trade-off between any loss of accuracy due to linearization versus the reduced computational load of using linear state estimation compared to nonlinear estimation. Such an investigation is considered out of the scope of the present work which mainly focuses on the theory and examples of using CV data for estimation from the perspective of linear state estimation.

III. State Estimation Methods

In this section, we briefly discuss the different methods implemented in this work for TSE using the ARZ model.

A. Moving Horizon Estimation

MHE is an optimization-based state estimation method that uses measurement data in batches from the most recent time horizon along with a process model to determine the states of the system. It involves solving an optimization problem at every time step of the process with the objective of minimizing the deviation of the estimated states from the modeled states as well as from the measurement data. Being an optimization problem, it is possible to include additional constraints in the problem such as bounds on the state variables. Depending upon whether the model is linear or nonlinear, MHE is divided into linear MHE and nonlinear MHE, both of which have been well explored in the literature. While linear MHE only requires solving a linear program or a quadratic program (QP) and is generally fast and easy to solve using available solvers, nonlinear MHE involves solving a nonlinear optimization problem which is both time-consuming and difficult. Since TSE for control is required to be done in real-time, in practice it is not always possible to spend enough time in solving a nonlinear optimization problem. Therefore, in this paper, we implement a linear MHE approach on a linearized version of the process model.

Throughout the paper, N is used to denote the size of the horizon for optimization. For time steps up to N, that is, near the start of the process, the horizon size is kept equal to the number of time steps from the initial time up to that time. The decision variables for the MHE optimization problem at any time step k are the state vectors from step $k - N$ to k out of which the vector at step k is considered the final output for that step. The MHE algorithm implemented in this work has a similar objective function to [32] with three components. The first component is known as the *arrival cost* which serves to connect the decision variables of the current optimization problem with the estimates up to the previous time step. This effectively allows us to consider the impact of data prior to the current horizon in the estimation process. The second and third components are penalties on the deviation of the estimates from the measurement data and the modeled dynamics respectively. The notations μ, w_1 and w_2 are used to denote the weights specifying our relative confidence on the past data and past estimates, the current measurement data, and the process model, and can be set by the modeler accordingly. The goal of the problem is to minimize these errors over the decision vectors. The remaining implementation including a thorough description of the decision variables, the objective function, and the constraints is omitted from the main body of this article as it does not contribute directly to the results of this paper. Interested readers are referred to Appendix D and Appendix E of [45] for detailed implementation and notes on comparison with other existing MHE algorithms. Algebraic transformations allow us to write the problem as a convex QP which can be solved using readily available QP solvers like CPLEX or MATLAB's `quadprog` function. Next, we present a brief discussion on the usage of KFs for TSE.

B. Kalman Filter Variants and Limitations

KFs are quite popular when it comes to TSE. Since the traffic process models are nonlinear we cannot use the ordinary KF, instead, most works use variants of KF designed for nonlinear systems namely the EKF, UKF, and EnKF. There is ample literature available on the design of these filters and their application in TSE, see [18] for references. A common limitation of the KF variants is that they do not inherently allow bounds on the state estimates. Since traffic states can only take values from a particular range, this makes it difficult to apply the KF variants directly. Instead, some modifications are required such as manually restricting the states to within their bounds after the state estimate for any time step is obtained. Another limitation of the KF variants is that they assume all errors to be Gaussian. This assumption is not necessarily true in many cases including the traffic system which can result in potential errors in state estimation. MHE naturally overcomes both of these limitations.

In the following section, we discuss the implementation and results obtained by applying the above-mentioned estimation methods with the help of a numerical example.

IV. Numerical Study Using VISSIM

In this section, we apply the state estimation methods discussed above namely EKF, UKF, EnKF, and MHE, on a traffic simulation example generated in VISSIM micro-simulation software under both fixed and moving sensors to highlight their advantages and limitations with respect to TSE and investigate the performance of moving sensors as compared to fixed sensors.

All the simulations are carried out using MATLAB R2020a running on a 64-bit Windows 10 with 2.2GHz Intel^{R} Core^{TM}

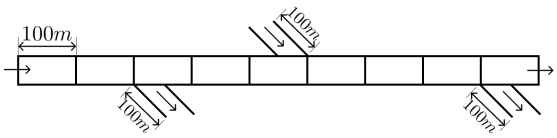

Fig. 2. Schematic diagram of the highway considered in this study.

i7-8750H CPU and 16GB of RAM. We use the `quadprog` function in MATLAB to solve the MHE optimization problem.

A. Numerical Study Objectives

The primary goal of this study is twofold- to test the performance of the state estimation methods discussed in Section III and to investigate the performance of moving sensors under various scenarios. In particular, we are interested in knowing the answers to the following questions:

- *Q1:* How does the number of fixed sensors on the highway impact the performance of the various estimation methods? Which method has the best estimation performance across different numbers of fixed segments?
- *Q2:* How does the state estimation performance of various methods vary with moving sensors? What is the impact of different frequencies of change in measurement positions on the estimation performance?
- *Q3:* Does the positional configuration of moving sensors impact state estimation performance?
- *Q4:* Which state estimation method is more robust to measurement errors? Do moving sensors impact estimation performance with different levels of data quality due to factors such as sensor noise and CV penetration rate?

Following is a description of the highway structure used for this study.

B. Highway Setup and VISSIM Simulation

In this study, we model the highway stretch as shown in Figure 2 consisting of one on-ramp and two off-ramps. An additional 100 m of highway stretch is modeled in VISSIM preceding the shown stretch. While we only perform state estimation on the latter 900 m and the attached ramps, this additional stretch of highway modeled in VISSIM provides us with the system inputs namely the demand upstream of Segment 1 and the upstream density and speed which are used to calculate the upstream driver characteristic. A similar 100 m stretch is modeled upstream of the on-ramp as well and serves the same purpose of providing the exact inputs. We set the following parameters for the Weidemann 99 car-following model in VISSIM: CC0 1.50 m, CC1 0.9 s, CC2 4.00 m, CC3 −8.00, CC4 −0.50, CC5 0.60, CC6 6.00, CC7 0.25 m/s^2, CC8 1.00 m/s^2, and CC9 1.50 m/s^2. The speed limit is set to 102 km/hr. Under the Godunov scheme, the highway and ramps are divided into segments of length 100 m each with a time-step value of 1 s, which satisfies the CFL condition. Thus, there are a total of 24 states in this highway system.

A traffic jam is introduced in the middle of the highway stretch to replicate a congested scenario which is more interesting for studying state estimation performance. In VISSIM, the jam is created with the help of a reduced speed decision area implemented on Segment 7 of the stretch. The reduction in traffic speed causes a reduction in flow creating a traffic jam that travels upstream on the highway. The jam dissipates once the speed of the reduced speed area is restored. The simulation scenario replicates the formation and dissipation of a traffic jam similar to that caused by an incident in the middle of the stretch. Note that the scenario considered in this work differs from those considered in previous studies such as in [24] where congestion travels upstream from the downstream end of the road stretch where estimation is performed. In that case, the source of the jam would be captured in the downstream supply conditions which are input to the system. In the scenario considered in this paper, the jam originates in the middle of the stretch. It is therefore not directly captured by any of the inputs and therefore the process model. The given scenario is arguably more difficult to estimate due to the absence of informative inputs to guide the process model.

The ARZ model parameters are selected to keep the simulated state trajectories from the macroscopic model as close to the VISSIM simulation as possible. The selected values are: $v = 102$ km/hr, $\rho_m = 345$ veh/km, $\tau = 20$, and $\gamma = 1.75$. As mentioned before, in this work we do not track individual vehicles, instead, we consider VANETs formed from CVs capable of measuring the density and speed of segments apart from fixed sensors. We consider a high penetration rate of CVs on the network such that we can query any desired segment for data. The only constraint we impose is bandwidth constraint on data transfer which limits the number of segments from which data can be obtained simultaneously.

C. Observability of the System

To determine the required minimum number and the corresponding placement of sensors, we perform a test of observability for our system using the concept of *Observability Gramian* for discrete-time systems [54]. The method is originally meant to determine the observability of linear systems. In this case, we use it to check the observability of the linearized ARZ model. The observability Gramian is defined as

$$\boldsymbol{W}_k = \sum_{m=0}^{\infty} (\tilde{\boldsymbol{A}}_k^\top)^m \tilde{\boldsymbol{C}}[k]^\top \tilde{\boldsymbol{C}}[k] \tilde{\boldsymbol{A}}_k^m,$$

where $\tilde{\boldsymbol{A}}_k$ is the coefficient matrix of the linearized state-space model and $\tilde{\boldsymbol{C}}[k]$ is the observation matrix of the linearized measurement model at time k around a suitable operating point. The system is considered observable if $\boldsymbol{W}_k$ is positive definite. In this case, since the model parameters change with time due to changing operating points of linearization, the Gramian changes with time as well. This can result in a change in the observability properties. To check if the system is observable for a given sensor placement, we calculate the Gramian for each time step over the duration of the simulation.

From this study, we find that to make the system observable, we need to at least sense the states on the last mainline segment and on all the off-ramps. Therefore, throughout the study, we keep fixed sensors on these segments. Any additional sensors are placed after these segments are populated with sensors. This is similar to the observations in [27] with respect to the observability of the model used in that paper. It appears

to be a common property of traffic models that the states of the output segments of the network (last mainline segment and off-ramps) need to be measured to ensure full-observability of the system. This is not surprising as traffic models share similar state-update equations and therefore have a similar structure of the state-space parameters which form the observability matrix. While the concept of observability can also be used to determine the optimal sensor placement for state estimation for any given number of sensors under certain conditions [55], here we only use it to determine a minimum number of sensors and their placement. In the following section, we discuss some nuances of implementing the aforementioned estimation methods in the current study.

D. Implementation of Estimation Methods

1) Evaluation Metrics: We use the *root mean squared error* (RMSE) and the *symmetric mean absolute percentage error* (SMAPE) [56] between the estimated and ground truth (simulated on VISSIM) density denoted by RMSE_ρ and SMAPE_ρ, respectively, and those between the estimated and ground truth speed denoted by RMSE_v and SMAPE_v, respectively to evaluate the performance of different methods. These metrics are defined as follows:

$$\text{RMSE}_\rho = \sqrt{\frac{1}{n_x t_f}\sum_{i=1}^{n_x}\sum_{k=1}^{t_f}(e_i^\rho[k])^2}, \tag{8}$$

$$\text{SMAPE}_\rho = \frac{100}{n_x t_f}\sum_{i=1}^{n_x}\sum_{k=1}^{t_f}\frac{|e_i^\rho[k]|}{\Xi_i^\rho[k]}, \tag{9}$$

$$\text{RMSE}_v = \sqrt{\frac{1}{n_x t_f}\sum_{i=1}^{n_x}\sum_{k=1}^{t_f}(e_i^v[k])^2}, \tag{10}$$

$$\text{SMAPE}_v = \frac{100}{n_x t_f}\sum_{i=1}^{n_x}\sum_{k=1}^{t_f}\frac{|e_i^v[k]|}{\Xi_i^v[k]}, \tag{11}$$

where $t_f = 500$ sec is the total time of simulation, $n_x = N + N_I + N_O$ is the total number of segments in the system, $e_i^\rho[k]$ and $e_i^v[k]$ denote the difference between the actual and estimated density and speed, respectively for the ith segment at time-step k, and $\Xi_i^\rho[k]$ and $\Xi_i^v[k]$ denote the sum of absolute values of the actual and estimated density and speed, respectively for the ith segment at time-step k. We do not consider the error in the relative flow states for evaluation since it is not directly relevant for traffic operators as compared to density and speed which are fundamental quantities in traffic. Note that, in this work, SMAPE is selected over the *mean absolute percentage error* (MAPE) [24] as MAPE has no upper bound and can give infinitely large values when the actual value is close to zero which is possible with traffic densities and speeds. The SMAPE on the other hand is bounded and can only assume values from 0% to 100%. While MAPE is easier to interpret than SMAPE, it is highly unstable in the present scenario and hence not preferable.

2) Parameter Tuning: In implementing KFs, three parameters need to be set in advance namely the estimate error covariance matrix ($\boldsymbol{P}$), the process noise covariance matrix ($\boldsymbol{Q}$), and the measurement noise covariance matrix ($\boldsymbol{R}$). In practical applications, these matrices are not known in advance or are difficult to get. In this paper, for all the KF variants, we use a process noise covariance matrix of the form $\boldsymbol{Q} = q\boldsymbol{I}_{n_x}$ where $q \in \mathbb{R}_+$ and $\boldsymbol{I}_{n_x}$ is an identity matrix of dimension n_x. Similarly, the measurement noise covariance matrix is set as $\boldsymbol{R} = r\boldsymbol{I}_{n_p[k]}$ with $r \in \mathbb{R}_+$ and $n_p[k]$ is the number of measured states at time k. The initial guess for the estimate noise covariance matrix is taken as $\boldsymbol{P} = 10^{-3}\boldsymbol{I}_{n_x}$. We set the values of r and q to 1 which is found to be sufficient. Marginally better results for the KFs can be obtained in each case by fine-tuning these matrices but it is avoided as in reality the real states are not known in advance. In general, algorithms requiring minimal tuning to achieve a reasonable quality of state estimates are desirable. For an estimation algorithm, it is important to determine a set of parameters robust to both traffic conditions and sensor placement. Here we focus on testing the algorithms with parameter values that yield reasonable state estimates across all scenarios, rather than fine-tuning parameters for individual cases.

Besides these values, there are also some method-specific parameters such as in UKF and EnKF. We find that fine-tuning the values of these parameters does not influence the performance of the methods considerably. For UKF, we set the following values: $\alpha = 0.1, \kappa = -4$, and $\beta = 2$, and for EnKF, we set the number of ensemble points to 100. These values are found to be sufficient for the respective methods. Interested readers can refer to [57] and [58] for interpretation of parameters and more detail on implementation of UKF and EnKF respectively.

For MHE, we set the values of the weights $\mu = 1, w_1 = 1$, $w_2 = 1$, and the horizon length $N = 4$. Just as with the KF parameters, fine-tuning MHE parameters is not a focus of this study, and the same parameter values are used in all tested scenarios without further tuning. In general, a large N is considered ideal as it allows the algorithm to track the system dynamics for a longer duration and also considers more data. However, this is not necessarily beneficial to estimation if the process model does not closely follow the real system states. In that case, particularly with a large weight w_2 on the process model error, the error in estimates can increase with increasing N as the error due to incorrect dynamics is amplified. To roughly tune N, we vary the horizon length from 1 to 10 with different numbers of fixed measurement segments. It is seen that the best N becomes smaller with an increasing number of measured segments. Also, a larger weight on the measurement error improves the results when there are more sensors. Both these observations are reasonable since the measurement data in this case is more accurate than the modeled states and so with sufficient data, increasing N only deteriorates the estimates by increasing the influence of the process model. Given a combination of these reasons, the aforementioned values are found reasonable for MHE.

3) Re-Scaling to Avoid Numerical Issues: The large difference in the order of magnitude of the two states, density, and relative flow, results in numerical issues in both the KFs as well as in MHE. This is handled by re-scaling the objective and constraints of the optimization problem in the case of MHE and by re-scaling the state vector in the case of KFs.

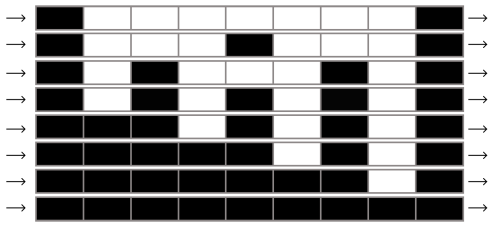

Fig. 3. Configurations for fixed sensor placement on the mainline segments. Black boxes depict segments with sensors and white boxes depict otherwise. Arrows indicate the direction of traffic. The ramp segments containing an additional 3 sensors are not presented in this figure. The top and bottom rows present the configuration with a total of 5 and 12 sensors in the system, respectively.

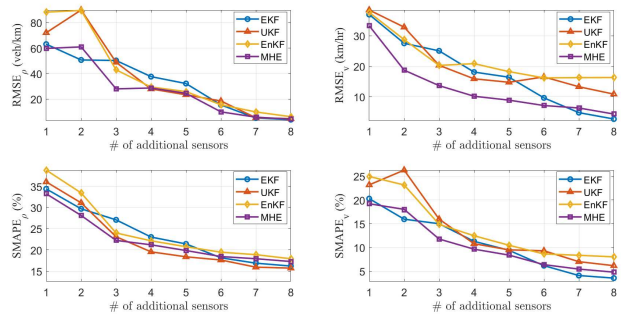

Fig. 4. RMSE_ρ [top left], RMSE_v [top right], SMAPE_ρ [bottom left], and SMAPE_v [bottom right] with different numbers of fixed sensors.

4) Applying External Bounds on States: The KFs sometimes run into the problem of producing non-physical states such as negative or extremely large densities and relative flows. This is an issue for the process model which includes terms like density raised to fractional power as in (2), which results in numerical issues and forces the estimation to stop. Therefore, it is important to bind the estimates from KFs to only the physical values of the states. In that, we project the obtained estimates in the case of EKF to a range with a lower bound of zero on all states, and an upper bound of ρ_m on the traffic densities and $\rho_m v_f$ on the relative flows. In the case of UKF, the sigma points are projected first followed by the obtained estimate. In the case of EnKF, the ensemble points are projected within specified bounds. This method of projecting vectors for EKF and UKF has been shown to fit in the KF theory mathematically and is among popular methods mentioned in [59].

We present the results of the study in the following section.

E. Results and Discussion

1) Comparison Under Fixed Sensor Positions: As sensors are indeed costly, it is imperative to determine which state estimation methods perform better with less number of sensors, and how the performance varies with the changing number of sensors. Herein, we test the effect of increasing the number of fixed sensors on the performance of the four estimation methods. We do not consider any CVs in this case. As discussed in Section IV-C, we have a minimum of three sensors, one on the last mainline segment and one each on the off-ramps. We also assume that there is always a sensor on the on-ramp. As we add more sensors we try to keep them well-distributed across the highway. The placement of the mainline sensors is depicted in Figure 3. No additional process noise is added to the state values generated from VISSIM while a zero mean uniform random noise with a standard deviation of 1 is added to the sensor measurements.

Figure 4 presents the plots of the evaluation metrics for each state estimation method. The x-axis presents the number of additional fixed sensors considered on the highway other than the sensors on the last mainline segment and ramps. Figure 4 shows that the performance of all the state estimation methods in terms of density and speed estimation improves with more additional sensors. Between methods, the RMSE_ρ and SAMPE_ρ at different numbers of segments appear comparable. A difference in performance in favor of MHE and EKF is observed in terms of RMSE_ρ at a small number of additional sensors but the difference diminishes as the number of sensors is increased. In terms of RMSE_v, MHE outperforms other methods at all numbers of additional sensors and is marginally outperformed by EKF at the two highest numbers of additional sensors. In terms of SAMPE_v, MHE and EKF perform comparably and better than other methods at a smaller number of additional sensors while all methods show a comparable performance at a higher number of sensors. While RMSE values are suitable for quantitative comparison between estimation methods, we need to compare the trajectories of the estimated states for a qualitative comparison. Figure 6 presents a 2-dimensional plot of the simulated and estimated density and speed evolution using MHE on all segments for the discussed scenario to provide a complete picture of the traffic evolution for the reader's reference. Figure 5 presents the simulated and estimated trajectories for the unmeasured segments for the case with 4 additional measured segments using MHE and EKF. The estimated trajectories obtained using UKF and EnKF are omitted from the plots in the main text to ensure clarity. The latter is presented in Appendix F of [45] for interested readers. The trajectories for the other cases of additional sensors are also omitted for brevity as they do not add value to the discussion provided in the context of the presented plots.

While EKF and MHE perform similarly for Segment 8 which does not have congestion, Figure 5e shows that EKF is not able to estimate the congested density on Segment 6. It does show a few spikes and a slight gradual increase in density but overall there is no significant congestion depicted by EKF. On the other hand, MHE follows the congested density more closely and also returns to the less congested ground truth condition once the congestion ends. In Figures 5a and 5c, MHE estimates congestion on the respective segments although the congestion is shown to start earlier than when it actually occurs. This is because the considered model inputs do not force congestion on any part of the stretch but one is observed in the measured data for middle segments from about 100 seconds into the simulation. The correction applied to the modeled states that are otherwise free-flowing to replicate the congestion in the measurement data causes congestion to be depicted earlier in the upstream segments corresponding to when it first occurs in the downstream segments rather than when it actually occurs on the upstream segments. Except for the time of the start of congestion in the estimated states on upstream segments, MHE is able to

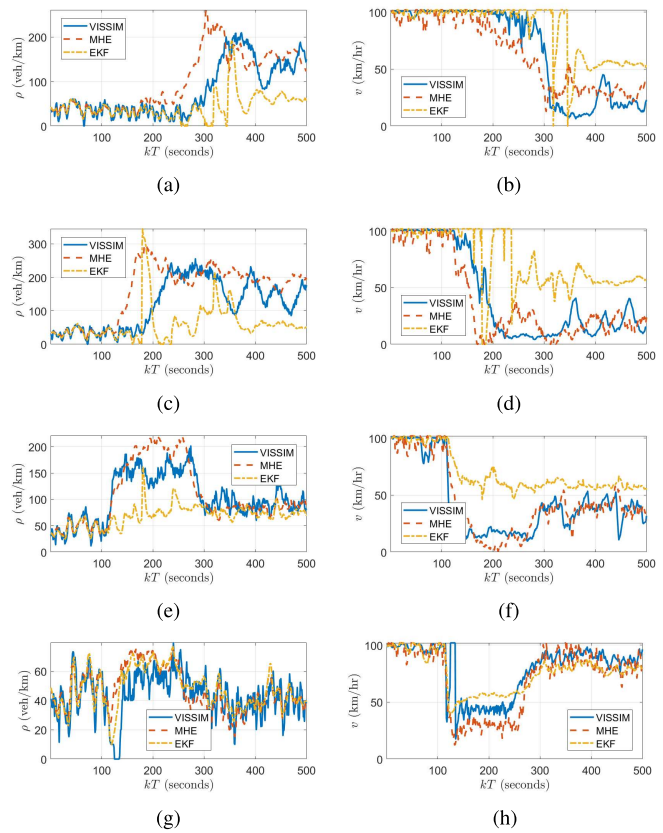

Fig. 5. Plots of simulated and estimated trajectories for densities [left] (a, c, e, g) and speeds [right] (b, d, f, h) in the presence of 4 additional fixed sensors. Rows of figures correspond to the unmeasured Segments 2, 4, 6, and 8 respectively.

TABLE II

COMPUTATIONAL TIME FOR STATE ESTIMATION PER TIME STEP (1 SEC) OF SIMULATION

Method	EKF	UKF	EnKF	MHE
Computation Time (sec)	0.002	0.006	0.016	0.075

replicate well the magnitude of the congestion in terms of both density and speed which EKF fails to do. The trajectory plots also explain the closeness of EKF and MHE in terms of RMSE_ρ and the significantly larger difference in RMSE_v. Both MHE and EKF observe a similar deviation from the actual density, with MHE depicting the congestion to start earlier while EKF not depicting or only partially depicting the congestion. As a result, both methods show a close RMSE_ρ. However, since MHE replicates the congestion while EKF does not, the former results in reduced speeds which are closer to the actual speeds than the higher ones estimated by EKF as a result of estimating lower densities. This results in a significantly smaller RMSE_v for MHE as compared to EKF. On the other hand, since EKF constantly overestimates the speeds, the denominator of SAMPE_v becomes large causing this metric to be close to its value for MHE despite a larger absolute error.

The average run times per time step of simulation for the methods are given in Table II. The run times include the time from when the data is received along with the information about the current observation matrix $\boldsymbol{C}[k]$ to when an estimate is produced. While the computation time of MHE is significantly higher than the KF variants, it is still only a fraction of the second and useful for real-time control. Moreover, the increased compute time can be justified by the improved estimation performance offered by MHE.

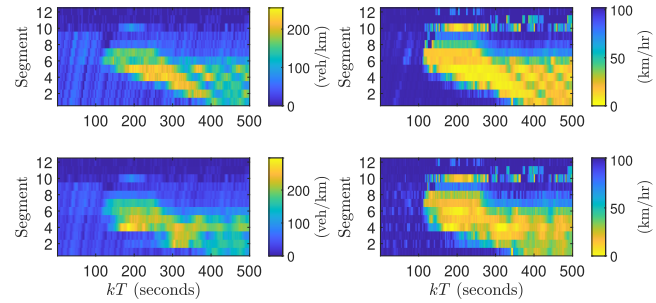

Fig. 6. Comparison of simulated and estimated densities [left] and speeds [right] obtained from MHE for the mainline segments (1 to 9), on-ramp segment (11), and off-ramps segments (11 & 12).

2) Effect of Moving Sensors: CVs can be used to measure traffic data from different segments over time giving more flexibility in terms of data collection than fixed sensors. Here, we test the impact of changing the segments from which measurements are obtained over time on the estimation performance of the considered state estimation methods. The frequency of change in measured segments is also varied and its impact on the estimation performance is analyzed. The last mainline segment and all ramp segments are assumed to have fixed sensors and CVs are used to get data from other segments. We assume that there is a sufficient penetration of CVs on the roadway to allow data collection from any segment on the stretch. However, we assume a restriction on the bandwidth for data transfer such that density and speed data collected using CVs can only be transferred from 3 segments at a time. A fixed bandwidth for data transfer in real-time is a realistic assumption however a stringent one of 3 segments is considered here in particular to clearly observe any benefit of covering different segments over time than fixed sensors which is difficult to observe if data is collected from several segments at all times. Hereafter, in the context of moving sensors, the term measured segment is used to refer to segments from which data is transferred and used for estimation rather than where data is collected (which is assumed to be all segments). Similarly, sensor position is used to refer to the position of a measured segment ignoring segments where sensors are present but data is not used for estimation. The initial sensor positions are the same as the third row from the top in Figure 3 with data being obtained from Segments $\{1, 3, 7\}$. The sensor positions are changed after a fixed duration of time. For this analysis, the duration is varied from indefinite (equivalent to fixed sensors) to 1 second (collecting data from a different set of segments every time step). A systematic update of sensor positions is utilized such that at every change the segments immediately following the current segments are selected. For instance, after Segments $\{1, 3, 7\}$, the positions are changed to Segments $\{2, 4, 8\}$. From Segment 8, the position is changed directly to Segment 1 skipping Segment 9 since Segment 9 already has a fixed sensor. So from Segments $\{2, 4, 8\}$, the positions are changed to Segments $\{3, 5, 1\}$, and so on after

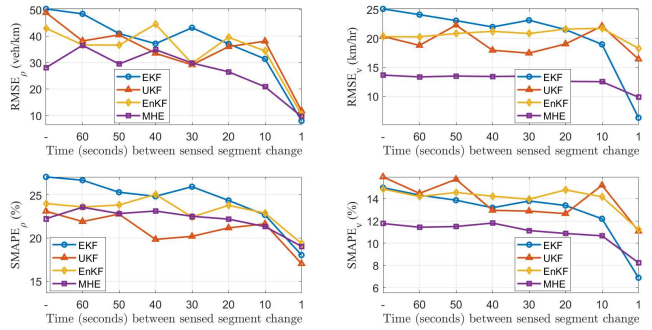

Fig. 7. RMSE$_\rho$ [top left], RMSE$_v$ [top right], SAMPE$_\rho$ [bottom left], SAMPE$_v$ [bottom right] with different duration between changes in sensor positions. The symbol '-' at the beginning of the x-axis represents the scenario with fixed sensor locations throughout the simulation.

the duration of change in each case. Figure 7 presents the plots of the evaluation metrics with increasing frequencies of changing the position of sensors for the four state estimation methods.

It is observed that overall for all estimation methods the value of both RMSE$_\rho$ and SAMPE$_\rho$ decreases with a decrease in the duration between consecutive changes in sensor positions. A majority of the improvement for all methods occurs with a change duration of fewer than 20 seconds and all methods converge in RMSE$_\rho$ at a change duration of 1 second. While overall there is an improvement with decreased duration between changes in sensor position, the trend is not completely monotonic as the error increases at few values of change duration. Since on some occasions, the process model may not be able to capture the traffic dynamics as well as on other occasions, the estimation error increases if the queried sensors end up not being at the location where the worse modeled behavior occurs at a given time. Given a total simulation duration of 500 seconds, certain values of change duration only result in a handful of position changes during the estimation period. In this case, the time in which data is not collected from segments whose dynamics are not captured well by the model may also be increased causing the error to increase, although marginally. With a higher frequency of sensor position changes, the error decreases monotonically, as the sensors send data from all the segments more frequently. The two metrics show a similar comparative trend between MHE, EKF, and EnKF while UKF shows a better performance than other methods in terms of SAMPE$_\rho$. This is primarily because UKF makes errors at higher actual density values compared to other methods and also relatively overestimates the densities (similar to the observations with EKF for SAMPE$_v$ in the case of fixed sensors) both of which lead to larger denominator values and a smaller percentage error. The corresponding trajectory plots are presented in Appendix F of [45]. As compared to density, the values of RMSE$_v$ and SAMPE$_v$ are less affected by the variation in frequency of change in sensor positions. Also, both metrics show a similar trend. In terms of speed, UKF, EnKF, and MHE only improve marginally compared to values with fixed sensor positions with the improvement observed at a change duration of 1 second. EKF, on the other hand, shows a bigger improvement outperforming other methods at the same duration between changes. This is similar to the observation in Figure 4 where EKF performs marginally

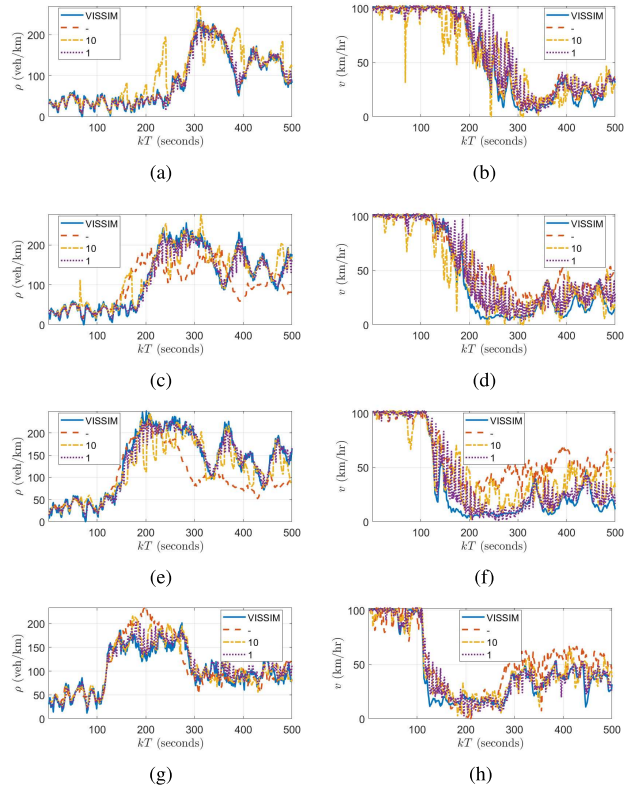

Fig. 8. Plots of simulated and estimated trajectories for densities [left] (a, c, e, g) and speeds [right] (b, d, f, h) in the presence of 3 additional measured segments with changing positions over time. Rows of figures correspond to Segments 3, 4, 5, and 6 respectively.

better than MHE when all segments are measured. Overall, MHE is observed to outperform other methods at all different durations between changes in sensor positions except when the positions are changed every 1 second. The simulated and estimated trajectories for density and speed using MHE for the cases with fixed sensors, sensor positions changing every 10 seconds and every 1 second are presented in Figure 8 to observe the qualitative improvement in estimation from using CVs as sensors.

Out of the plotted segments, Segment 3 is measured in the fixed sensor case therefore the estimated trajectory from fixed sensors overlaps well with the real trajectory. The other three plotted segments are unmeasured in the fixed sensor case. As expected, the moving sensors result in estimated densities and speeds that follow the real states more closely than the fixed sensor case except for Segment 3. Note that the trajectories estimated using the moving sensors show oscillations in both density and speed which are a result of the segment measurements becoming unavailable over regular intervals. A duration of 10 seconds between changes results in less frequent but larger oscillation as measurements are unavailable for more time steps allowing for larger deviations from the real state. The oscillations in speed estimates are more profound than those in density plots which can be attributed to speeds being obtained by a division of the density and relative flow states and thus being sensitive to changes in both. Since such oscillations in estimated states are generally undesirable, a smoothing filter such as a moving average filter may be applied to the estimated states from the moving sensors

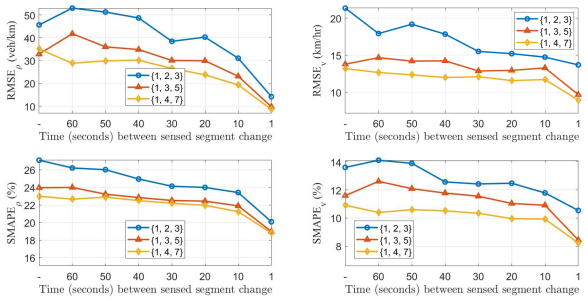

Fig. 9. RMSE$_\rho$ [top left], RMSE$_v$ [top right], SMAPE$_\rho$ [bottom left], and SMAPE$_v$ [bottom right] with different starting configurations of sensors and varying duration between changes in sensor positions using MHE. The symbol '$-$' at the beginning of the x-axis represents the scenario with fixed sensor locations for the starting configurations in the legend.

to make them more realistic and usable. Plots obtained by applying a moving average filter with values averaged over 15 time steps are presented in Appendix F of [45].

3) Impact of Segment Selection for Sensing: A limited budget leads to limited bandwidth for data transfer causing traffic measurement data to be available only for a few segments on the road at a time. Therefore, it is important to determine which segments to query for data to obtain the best state estimates. In this section, we test the impact of querying the same number of sensors placed with different spacing while changing the position of sensors. We consider 3 additional segments with data apart from the fixed sensors on the last mainline segment and all ramp segments. For this study, we only consider MHE as it is observed to perform the best with 3 additional sensors. We consider three scenarios with different spacing between measured segments such that their starting sensor positions are $\{1, 2, 3\}$, $\{1, 3, 5\}$, and $\{1, 4, 7\}$. The duration between the change in the position of segments is varied in the same way as in Section IV-E2. Figure 9 presents the plots for evaluation metrics for the different starting configurations of sensors (presented in the legend) with varying duration between changes in sensor positions presented on the x-axis.

There is an overall improvement in the estimation performance in terms of all metrics with decreasing duration between changes in sensor position as also observed in Section IV-E2. Between starting configurations, the configurations with more uniformly spaced measured segments namely $\{1, 3, 5\}$ and $\{1, 4, 7\}$ perform better than consecutive positions $\{1, 2, 3\}$ at all values of time between position changes. Between $\{1, 3, 5\}$ and $\{1, 4, 7\}$, the latter performs better. The difference in performance between the configurations decreases as the duration between position changes is reduced. Compared to the performance of the uniformly spaced starting configurations, the starting placement $\{1, 2, 3\}$ with an x-axis value of above 10 seconds performs worse than when the former is only fixed. This indicates that a more uniformly spaced positioning of sensors is always desirable and can even outperform when the same number of sensors is clustered even if moving to cover more segments over time.

4) Impact of measurement quality: As sensors are prone to faults, the sensor noise may change from its manufacturer-specified value for the sensor from time to time. At the same time, lower penetration rates of CVs in different segments can also result in reduced quality of data measurements. In such scenarios, a method more robust to measurement errors is considered more reliable. In this section, we check the impact of changing the measurement quality on the estimation performance of the four methods. We further investigate the impact of changing sensor positions on the estimation performance with different levels of error. The goal is to see if the performance improvement offered by moving sensors can offset the deterioration caused by measurement quality. We set the measurement error equal to a random noise drawn from a uniform distribution with bounds $[-1,\ 1]$, which is standardized and scaled to have zero mean and a standard deviation of $s \in \mathbb{R}$ which is varied to replicate different levels of error. The distribution of noise added to both density and speed measurement is kept the same. In general, the measurement error covariance matrix R for the KF variants is set according to the actual covariance of the noise which in this case is the diagonal matrix with all diagonal elements equal to s^2. However, in practice, it is difficult to know the distribution of noise if it is due to varying penetration rates or unexpected sensor faults. Therefore, in this case, we continue to use the value of R defined in Section IV-D2. For MHE, the objective weights and horizon length are also kept the same. Figure 10 presents the plots of RMSE$_\rho$ and RMSE$_v$ against increasing values of standard deviation s for the four state estimation methods. The plots for SMAPE are omitted for brevity as they show similar relative trends compared to the plots for RMSE and do not contribute to the discussion. The unit of s is the same as the measurements (veh/km for density and km/hr for speed) but is omitted from the plots as it represents the standard deviation for noise in both types of measurements. Three numbers of additional sensors are considered namely 3, 5, and 7 to present the trend in estimation error with increasing measurement noise for different numbers of sensors. For each number of sensors, we take 5 random seed values and average the metrics over the 5 seeds. From the plots, it appears that in all the cases, the estimation error increases with an increase in measurement noise s.

It is observed that the performance of all methods in general deteriorates with increasing noise in the measurements which is expected as the data becomes less reliable. The deterioration in terms of RMSE$_\rho$ is more prominent with the increasing number of sensors. Notice that the performance of the methods in terms of density does not change much up to $s = 20$ for the case with three additional sensors. However, the corresponding increase in RMSE$_v$ shows that the overall estimation performance does indeed deteriorate with increasing noise even at lower levels of noise. The RMSE$_\rho$ does not immediately increase for 3 additional sensors, because the error is already quite large to be sensitive to a small increase in noise. The large increase in the error for UKF is due to large jumps in the state trajectories that reach their bounds from time to time at higher noise possibly due to instability issues with UKF. For MHE, the percentage increase in the error in density and speed between the lowest and highest noise levels are 66.6%, 50.7%, 316.2% and 151.8%, 236.7%, 360.1% for 3, 5, and 7 additional sensors, respectively. Note that the deterioration in performance becomes more prominent in both density and speed with the increasing number of sensors

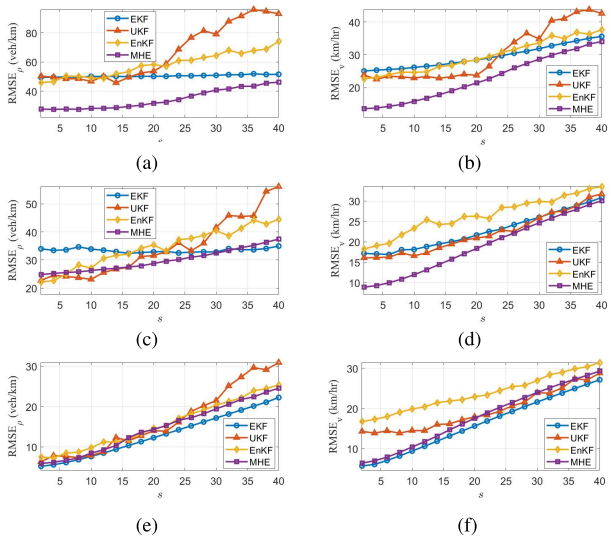

Fig. 10. RMSE_ρ [left] and RMSE_v [right] with changing s (standard deviation of added measurement noise) with 3, 5, and 7 additional fixed sensors in the three rows respectively.

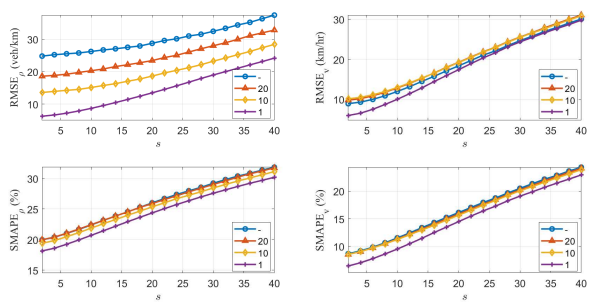

Fig. 11. RMSE_ρ [top left], RMSE_v [top right], SMAPE_ρ [bottom left], and SMAPE_v [bottom right] with changing s (standard deviation of added measurement noise) with 5 additional measured segments with changing positions using MHE.

as the performance becomes more directly associated with available data and thus more sensitive to measurement quality. Overall, UKF appears to be the least reliable in the presence of large measurement errors due to instability issues. Also as observed before, MHE performs better than other methods at smaller numbers of measured segments at lower error levels and while it is affected significantly by measurement error, its performance is still better or comparable to other methods with large noise. Next, we observe the impact of the changing sensor positions on the estimation performance in the presence of noise.

Figure 11 presents the variation in error values for estimation using MHE with changing levels of measurement noise in the presence of moving sensors similar to the setting in Section IV-E2. The legend presents the time (in seconds) between the change in sensor positions using the same position update logic as presented in Section IV-E2. We consider 5 additional sensors in the configuration presented in Figure 3. Noise is implemented in the same way as above. The plots for 3 and 7 additional sensors are omitted for brevity. As observed in Section IV-E2, the estimation performance in terms of RMSE_ρ improves with 20 seconds and further lower duration between changes in sensor positions while RMSE_v only shows a small improvement at a change duration of 1 second and performs similar to the case with fixed positions with a change duration of even 10 and 20 seconds. While the error at all frequencies of change in sensor position is observed to increase similarly with increasing noise, the RMSE_ρ is consistently smaller for higher frequencies. A change duration of 10 seconds at $s = 20$ performs equivalent to a change duration of 20 seconds at $s = 0$. In terms of RMSE_v, there is very little change/improvement with a change in position. Plots for SMAPE_ρ show a smaller difference in magnitude between the curves for the same values on the x-axis compared to RMSE_ρ. Qualitative examination of the trajectory plots indicates that a majority of the errors at longer change duration occur at higher densities leading to larger denominator values and hence smaller percentage errors. Also, for certain segments, the scenarios with longer change duration heavily overestimate the densities at lower actual density values which further leads to reduced percentage errors and hence smaller differences in curves. The trajectory plots can be found in Appendix F of [45]. Overall, the evaluation plots reiterate that there is merit in using CVs as moving sensors as the improvement in performance helps offset the deterioration caused by the measurement errors.

V. Conclusion and Future Work

From the previous analysis, we have some preliminary suggestions regarding the questions posed in Section IV-A which are as follows:

- *A1:* As expected, the performance of the state estimation methods is improved upon increasing the number of sensors in the system. The performance in terms of density is similar across methods while MHE outperforms other methods in terms of estimated speeds.
- *A2:* The performance of all estimation methods in general improves with moving sensors as sensors cover more segments on the highway over time as compared to the case with only fixed sensors. The variation in performance is non-monotonic at lower frequencies of change in sensor positions but a prominent improvement is observed at a duration of 20 seconds or lower. The improvement in density is more profound than the improvement in speeds.
- *A3:* More uniformly spaced segments result in a better state estimation performance than the same number of segments placed consecutively in the case of both fixed and moving sensors. Uniformly placed fixed sensors also outperform consecutively placed moving sensors up to a duration of 20 seconds between changes.
- *A4:* The performance of all methods worsens with worsening data quality that might result from lower penetration rates of CVs and higher sensor noise. UKF is the least robust out of all methods and shows abrupt increases in error compared to other methods with increasing noise. The impact of quality issues is more prominent in scenarios with more measured segments which show a deterioration in performance at lower levels of measurement noise as compared to scenarios with fewer measured segments. The performance improvement achieved by the use of moving sensors is able to offset the deterioration caused by measurement quality to a fair extent showing

the advantage of using moving sensors under adverse conditions.

To summarize, we present a state-space formulation for the nonlinear ARZ model while considering junctions in the form of ramp connections. Since the ARZ model is nonlinear, it is not possible to directly apply linear state estimation methods to it which are considered to be more efficient than nonlinear methods. We linearize the ARZ model using Taylor series approximation and use it to implement linear state estimation such as through linear MHE. We present the formulation for linear MHE which has not previously been used for TSE and show that it is a good choice for TSE compared to other popular methods namely EKF, UKF, and EnKF. We also show that the use of moving sensors is better for state estimation compared to fixed sensors and can offset the degradation in performance caused by reduced measurement quality resulting from sensor noise and lower penetration rates of CVs. Various strategies and variations in the selection of segments to obtain CV data are also investigated.

It is important to note that this study is constrained by a lack of access to real-world data corresponding to the given setting. Given the present theoretical demonstration of the benefits of using mobile CV data sources, a comprehensive investigation is required to assess the real-world capabilities of the approach. In particular, it is important to investigate the technical challenges in ingesting and fusing real-time data streams from both CVs and fixed sensors, examining their impact on estimation performance.

Future work will also consider the optimal placement of sensors considering CVs for TSE. Besides, while the performance of the ARZ model against the first-order LWR model has been studied in prior research [24] which claims the superiority of the former, some of the newer works [60] have suggested the possibility of the order of the model being less significant for TSE in the presence of sufficient data. Therefore, it would be interesting to carry out a detailed comparative study between the performance of the ARZ model and a first-order model under different scenarios specifically those depicting non-equilibrium conditions under different sensor placements.

REFERENCES

[1] G. Gomes and R. Horowitz, "Optimal freeway ramp metering using the asymmetric cell transmission model," *Transp. Res. C, Emerg. Technol.*, vol. 14, no. 4, pp. 244–262, Aug. 2006.

[2] R. C. Carlson, I. Papamichail, M. Papageorgiou, and A. Messmer, "Optimal motorway traffic flow control involving variable speed limits and ramp metering," *Transp. Sci.*, vol. 44, no. 2, pp. 238–253, May 2010.

[3] Y. Han, A. Hegyi, Y. Yuan, S. Hoogendoorn, M. Papageorgiou, and C. Roncoli, "Resolving freeway jam waves by discrete first-order model-based predictive control of variable speed limits," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 405–420, Apr. 2017.

[4] G. S. van de Weg, A. Hegyi, S. P. Hoogendoorn, and B. De Schutter, "Efficient freeway MPC by parameterization of ALINEA and a speed-limited area," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 16–29, Jan. 2019.

[5] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[6] M. J. Lighthill and G. B. Whitham, "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proc. Roy. Soc. London. A Math. Phys. Sci.*, vol. 229, no. 1178, pp. 317–345, 1955.

[7] P. I. Richards, "Shock waves on the highway," *Oper. Res.*, vol. 4, no. 1, pp. 42–51, Feb. 1956.

[8] F. Harrou, A. Zeroual, M. M. Hittawe, and Y. Sun, "Road traffic modeling," in *Road Traffic Modeling and Management*, F. Harrou, A. Zeroual, M. M. Hittawe, and Y. Sun, Eds., Amsterdam, The Netherlands: Elsevier, 2022, ch. 2, pp. 15–63.

[9] C. F. Daganzo, "Requiem for second-order fluid approximations of traffic flow," *Transp. Res. B, Methodol.*, vol. 29, no. 4, pp. 277–286, 1995.

[10] H. J. Payne, "Model of freeway traffic and control," *Math. Model Public Syst.*, vol. 1, pp. 51–61, Mar. 1971.

[11] G. B. Whitham, *Linear and Nonlinear Waves*, vol. 42. Hoboken, NJ, USA: Wiley, 2011.

[12] A. Aw and M. Rascle, "Resurrection of 'second order' models of traffic flow," *SIAM J. Appl. Math.*, vol. 60, no. 3, pp. 916–938, 2000.

[13] H. M. Zhang, "A non-equilibrium traffic model devoid of gas-like behavior," *Transp. Res. B, Methodol.*, vol. 36, no. 3, pp. 275–290, Mar. 2002.

[14] L. Leclercq et al., "The Lagrangian coordinates and what it means for first order traffic flow models," *Transp. Traffic Theory*, pp. 735–753, Jul. 2007. [Online]. Available: https://trid.trb.org/View/820161

[15] 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 *Proc. 47th IEEE Conf. Decis. Control*, Dec. 2008, pp. 5062–5068.

[16] M. Wright and R. Horowitz, "Fusing loop and GPS probe measurements to estimate freeway density," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3577–3590, Dec. 2016.

[17] A. Nantes, D. Ngoduy, A. Bhaskar, M. Miska, and E. Chung, "Real-time traffic state estimation in urban corridors from heterogeneous data," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 99–118, May 2016.

[18] 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, Apr. 2017.

[19] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: A general approach," *Transp. Res. B, Methodol.*, vol. 39, no. 2, pp. 141–167, Feb. 2005.

[20] M. Zhao et al., "Real-time freeway traffic state estimation with fixed and mobile sensing data," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–7.

[21] I. Agalliadis, M. Makridis, and A. Kouvelas, "Traffic estimation by fusing static and moving observations in highway networks," in *Proc. 20th Swiss Transp. Res. Conf. (STRC)*, 2020, pp. 1–5. [Online]. Available: https://www.research-collection.ethz.ch/handle/20.500.11850/419736

[22] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham, "Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool METANET," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 4, pp. 282–292, Dec. 2002.

[23] A. Messmer and M. Papageorgiou, "METANET: A macroscopic simulation program for motorway networks," *Traffic Eng. Control*, vol. 9, nos. 8–9, pp. 466–470, 1990.

[24] T. Seo and A. M. Bayen, "Traffic state estimation method with efficient data fusion based on the Aw-Rascle-Zhang model," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.

[25] H. Yu, A. M. Bayen, and M. Krstic, "Boundary observer for congested freeway traffic state estimation via Aw-Rascle-Zhang model," *IFAC-PapersOnLine*, vol. 52, no. 2, pp. 183–188, 2019.

[26] R. Wang, Y. Li, and D. B. Work, "Comparing traffic state estimators for mixed human and automated traffic flows," *Transp. Res. C, Emerg. Technol.*, vol. 78, pp. 95–110, May 2017.

[27] 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.

[28] J. Xing, W. Wu, Q. Cheng, and R. Liu, "Traffic state estimation of urban road networks by multi-source data fusion: Review and new insights," *Phys. A, Stat. Mech. Appl.*, vol. 595, Jun. 2022, Art. no. 127079.

[29] S. A. Kashinath et al., "Review of data fusion methods for real-time and multi-sensor traffic flow analysis," *IEEE Access*, vol. 9, pp. 51258–51276, 2021.

[30] S. A. Nugroho, A. F. Taha, and C. G. Claudel, "A control-theoretic approach for scalable and robust traffic density estimation using convex optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 64–78, Jan. 2021.

[31] S. C. Vishnoi, S. A. Nugroho, A. F. Taha, C. Claudel, and T. Banerjee, "Asymmetric cell transmission model-based, ramp-connected robust traffic density estimation under bounded disturbances," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2020, pp. 1197–1202.

[32] C. V. Rao, J. B. Rawlings, and J. H. Lee, "Constrained linear state estimation—A moving horizon approach," *Automatica*, vol. 37, no. 10, pp. 1619–1628, 2001.

[33] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE Trans. Autom. Control*, vol. 48, no. 2, pp. 246–258, Feb. 2003.

[34] Y. Wang, J. Yuan, S. Yu, Y. Hu, and H. Chen, "Resilient moving horizon estimation for cyber-physical systems under sensor attacks," in *Proc. 11th Asian Control Conf. (ASCC)*, Dec. 2017, pp. 2274–2279.

[35] A. Alessandri and M. Gaggero, "Fast moving horizon state estimation for discrete-time systems using single and multi iteration descent methods," *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4499–4511, Sep. 2017.

[36] I. I. Sirmatel and N. Geroliminis, "Nonlinear moving horizon estimation for large-scale urban road networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 12, pp. 4983–4994, Dec. 2020.

[37] I. I. Sirmatel and N. Geroliminis, "Model-based identification, estimation, and control for large-scale urban road networks," in *Proc. Eur. Control Conf. (ECC)*, May 2020, pp. 408–413.

[38] S. Timotheou, C. G. Panayiotou, and M. M. Polycarpou, "Moving horizon fault-tolerant traffic state estimation for the cell transmission model," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, Dec. 2015, pp. 3451–3456.

[39] A. J. Huang and S. Agarwal, "Physics informed deep learning for traffic state estimation," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6.

[40] C. Zhao and H. Yu, "Integrating PDE observer with deep learning for traffic state estimation," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2022, pp. 1964–1969.

[41] R. Shi, Z. Mo, K. Huang, X. Di, and Q. Du, "A physics-informed deep learning paradigm for traffic state and fundamental diagram estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11688–11698, Aug. 2022.

[42] X. Di, R. Shi, Z. Mo, and Y. Fu, "Physics-informed deep learning for traffic state estimation: A survey and the outlook," *Algorithms*, vol. 16, no. 6, p. 305, Jun. 2023.

[43] 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.

[44] R. Courant, K. Friedrichs, and H. Lewy, "On the partial difference equations of mathematical physics," *IBM J. Res. Develop.*, vol. 11, no. 2, pp. 215–234, Mar. 1967.

[45] S. C. Vishnoi, S. A. Nugroho, A. F. Taha, and C. G. Claudel, "Traffic state estimation for connected vehicles using the second-order Aw-Rascle-Zhang traffic model," 2022, *arXiv:2209.02848*.

[46] H. Yang and H. Rakha, "A novel approach for estimation of dynamic from static origin–destination matrices," *Transp. Lett.*, vol. 11, no. 4, pp. 219–228, Jul. 2019.

[47] C. Lee, J. Lee, and M. Kim, "Density measurement algorithm for freeway segment using two point detectors," *J. Adv. Transp.*, vol. 45, no. 3, pp. 207–218, Jul. 2011.

[48] L. A. Klein et al., *Traffic Detector Handbook*, vol. 1. McLean, VA, USA: Turner-Fairbank Highway Research Center, 2006.

[49] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation per lane in the presence of connected vehicles," *Transp. Res. B, Methodol.*, vol. 106, pp. 1–28, Dec. 2017.

[50] 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.

[51] S. Panichpapiboon and W. Pattara-atikom, "Evaluation of a neighbor-based vehicle density estimation scheme," in *Proc. 8th Int. Conf. ITS Telecommun.*, Oct. 2008, pp. 294–298.

[52] P. J. Moreno, B. Raj, and R. M. Stern, "A vector Taylor series approach for environment-independent speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Conf.*, vol. 2, May 1996, pp. 733–736.

[53] H. Yu, Q. Gan, A. Bayen, and M. Krstic, "PDE traffic observer validated on freeway data," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 3, pp. 1048–1060, May 2021.

[54] C.-T. Chen, *Linear System Theory and Design*. Philadelphia, PA, USA: Saunders, 1984.

[55] S. A. Nugroho, S. C. Vishnoi, A. F. Taha, C. G. Claudel, and T. Banerjee, "Where should traffic sensors be placed on highways?" *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 13026–13039, Aug. 2022.

[56] L. Li, X. Chen, and L. Zhang, "Multimodel ensemble for freeway traffic state estimations," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 3, pp. 1323–1336, Jun. 2014.

[57] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. IEEE Adapt. Syst. Signal Process., Commun., Control Symp.*, Oct. 2000, pp. 153–158.

[58] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean Dyn.*, vol. 53, no. 4, pp. 343–367, Nov. 2003.

[59] D. Simon, "Kalman filtering with state constraints: A survey of linear and nonlinear algorithms," *IET Control Theory Appl.*, vol. 4, no. 8, pp. 1303–1318, Aug. 2010.

[60] Y. Wang et al., "Real-time joint traffic state and model parameter estimation on freeways with fixed sensors and connected vehicles: State-of-the-art overview, methods, and case studies," *Transp. Res. C, Emerg. Technol.*, vol. 134, Jan. 2022, Art. no. 103444.

Suyash C. Vishnoi received the B.Tech. degree in civil engineering from IIT Roorkee in 2018 and the Ph.D. degree in transportation engineering from UT Austin in 2023. He is currently a Post-Doctoral Research Fellow with the School of Civil and Environmental Engineering, Georgia Institute of Technology. His research interests include simulation modeling, optimal control, and state estimation. He was a recipient of the Talbot S. Huff, Sr. Highway Engineering Graduate Fellowship from UT Austin.

Sebastian A. Nugroho (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Institut Teknologi Bandung, Indonesia, in 2012 and 2014, respectively, and the Ph.D. degree in electrical engineering from UT San Antonio, San Antonio, USA, in 2021. He was a Post-Doctoral Research Fellow with Michigan Power and Energy Laboratory, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA, from 2021 to 2022. He is currently a Technical Specialist for connected and intelligent systems research with the Cummins Technical Center, Research and Technology, Cummins Inc., Columbus, IN, USA. His primary research interests include control theory and engineering optimization with applications to energy systems and transportation networks.

Ahmad F. Taha (Member, IEEE) received the B.E. degree in electrical and computer engineering from the American University of Beirut, Beirut, Lebanon, in 2011, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2015. He is currently an Associate Professor with the Department of Civil and Environmental Engineering (and a secondary appointment with the Department of Electrical and Computer Engineering), Vanderbilt University, Nashville, TN, USA. Prior to joining Vanderbilt University, he was an Assistant Professor with The University of Texas at San Antonio, San Antonio, TX, USA. He is interested in understanding how complex cyber-physical systems (CPS) operate, behave, and misbehave. His research interests include optimization, control, and security of CPSs with applications to power, water, and transportation networks.

Christian G. Claudel received the M.S. degree in plasma physics from Ecole Normale Superieure de Lyon in 2004 and the Ph.D. degree in electrical engineering from UC Berkeley in 2010. He is currently an Associate Professor with the Department of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin. His research interests include the control and estimation of distributed parameter systems, wireless sensor networks, and unmanned aerial vehicles. He received the Leon Chua Award from UC Berkeley in 2010 for his work on the Mobile Millennium traffic monitoring system.