2022 American Control Conference (ACC)
Atlanta, USA, June 8-10, 2022

On-Board Traffic Prediction for Connected Vehicles: Implementation and Experiments on Highways

Tamás G. Molnár, Xunbi A. Ji, Sanghoon Oh, Dénes Takács,
Michael Hopka, Devesh Upadhyay, Michiel Van Nieuwstadt, and Gábor Orosz

***Abstract*— An on-board traffic prediction algorithm is proposed for connected vehicles traveling on highways. The prediction is based on data received from other connected vehicles ahead in the traffic stream, leveraging the fact that a vehicle will enter the traffic that other vehicles ahead have already met. Our method includes traffic state estimation with Kalman filter and prediction via traffic flow models describing the propagation of congestion waves. The end result is an individualized speed preview in real time up to about half a minute for the connected vehicle executing prediction. Most importantly, the traffic prediction was successfully implemented on board of a real vehicle and predictions were tested in real traffic with experiments involving connected human-driven vehicles.**

I. INTRODUCTION AND MAIN CONCEPT

Traffic congestion is one of the major factors that reduces the efficiency of road transportation. To overcome congestions, a plethora of methods have been proposed in the literature for traffic control, vehicle control in congested traffic and the integration thereof. Vehicle control may significantly benefit from speed previews and predictions about the upcoming traffic. Previews allow more efficient operation of on-board systems, such as after treatment and engine control, and may improve the energy efficiency of driving [1], [2]. Hence, this paper focuses on traffic predictions that vehicles may use on board in real time to optimize their operation.

Traffic estimation and prediction have long been of interest in the literature. Initially, traffic estimation relied on Eulerian data from road locations, for example, by counting vehicles via loop detectors or cameras [3], [4], [5]. The underlying traffic models are Eulerian, formulated in road-fixed frame. The availability of GPS-capable devices made vehicle trajectory data also widely accessible [6], [7]. These Lagrangian data were fused into Eulerian models for traffic estimation. Notably, [8], [9], [10] used Kalman filtering to achieve this, which inspired our work. Recently, trajectory data of automated [11] and connected [12] vehicles were used for traffic reconstruction by Eulerian models. Lagrangian models formulated in the vehicle frame have also proved to be useful for traffic estimation with trajectory data [13], [14], while fusion of different kinds of data also exists [15], [16], [17].

To obtain traffic predictions, routing apps (such as Google Maps, Here Maps, TomTom and Waze) are the most popular tools. While forecasts from these providers are extremely useful, they are road location specific as opposed to vehicle specific, and they are typically delayed with a few minutes due to processing and update latencies which prohibit their use for real-time control. Vehicle control could enjoy further benefits if predictions were tailored to the needs of the controlled vehicle and were made on board in real time.

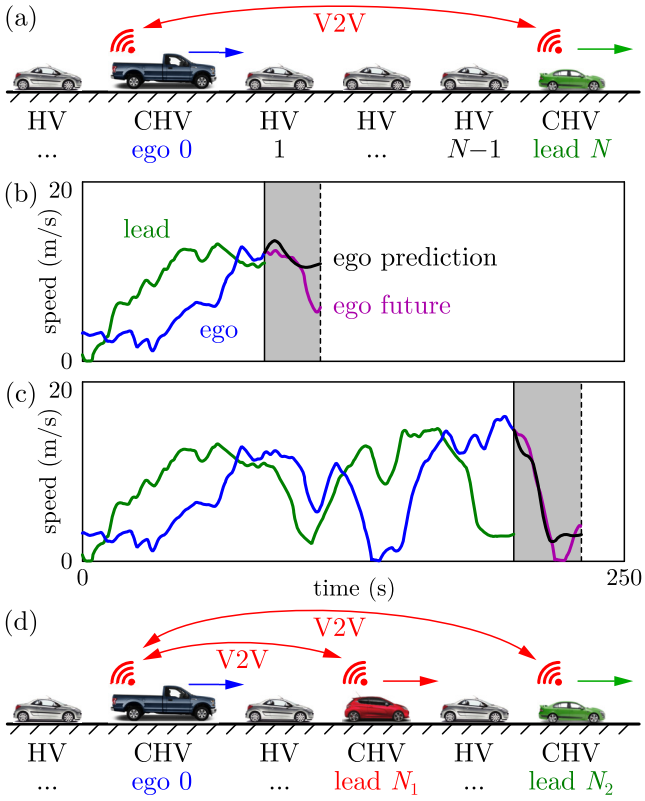

Fig. 1. (a) The concept of connectivity-based on-board traffic prediction where an ego vehicle (blue) predicts its future motion based on data received from a lead vehicle (green) via vehicle-to-vehicle (V2V) connectivity. (b)-(c) Speed previews (black) and the ego's future speed (purple) at two different time moments. (d) Multiple lead vehicles providing data for traffic prediction.

This work was supported by Ford Motor Co. by the University of Michigan's Center for Connected and Automated Transportation (CCAT) through the US DOT grant 69A3551747105, and by Dow (#227027AT).

T. G. Molnár is with the Department of Mechanical and Civil Engineering, California Institute of Technology, Pasadena, CA 91125, USA and was with the Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA; tmolnar@caltech.edu.

X. A. Ji, S. Oh, and G. Orosz are with the Department of Mechanical Engineering and G. Orosz is also with the Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, USA; {xunbij,osh,orosz}@umich.edu.

D. Takács is with the Department of Applied Mechanics, Budapest University of Technology and Economics, Budapest, Hungary; takacs@mm.bme.hu.

M. Hopka, D. Upadhyay and M. Van Nieuwstadt are with the Ford Research and Innovation Center, Dearborn, MI 48124, USA; {mhopka,dupadhya,mvannie1}@ford.com.

A potential technology to provide traffic data on board in real time (with a delay in the time scale of $0.1\,\mathrm{s}$) is vehicle-to-vehicle (V2V) or, in general, vehicle-to-everything (V2X) connectivity. This led to the concept of connectivity-based on-board traffic prediction in our previous work [18]. The concept is shown in Fig. 1(a). A vehicle whose future is of interest, called *ego vehicle*, is connected to another vehicle traveling ahead, called *lead vehicle*. Connectivity allows the lead to share its position and speed with the ego. As the ego meets the traffic the lead has already passed, we can assert that the ego's future motion can be predicted using data from the lead. This statement is valid for traffic jams and slow moving traffic under certain simplifying assumptions.

The envisioned speed previews are illustrated by Fig. 1(b)-(c) for two different time moments. The trajectories of the lead and ego vehicles (green and blue) up to present time (vertical line) are used to predict the ego's future trajectory (black). Ideally, the prediction — which requires not only instantaneous speed and position, but historical data — captures the ego vehicle's future motion (purple). Connectivity from multiple lead vehicles, as illustrated in Fig. 1(d), can improve the achievable prediction horizon and accuracy.

Building on this concept, this work makes two contributions. First, we propose a prediction method, wherein traffic state is estimated with Kalman filter. In [8], [9], [10] Kalman filter was used to estimate velocity fields by Eulerian models, whereas [13], [14] reconstructed trajectories for pairs of subsequent vehicles. We use Kalman filtering to estimate trajectories between two distant connected vehicles (lead and ego). Our second – and most important – contribution is implementing traffic prediction on connected human-driven vehicles in real highway experiments. For the first time, we provided individualized speed previews to an ego vehicle in real time, on board, in the time horizon of a minute.

The paper is outlined as follows. Section II shows experiments with connected human-driven vehicles on real highways, where trajectory data were collected and predictions were tested. Section III explains the prediction method. Section IV shows the outcome of traffic prediction using experimental data, with discussions about prediction horizon and accuracy. Section V concludes our work.

II. DATA COLLECTION

In our experiments, three passenger cars were driven on real highways to collect data and make on-board traffic predictions; see Fig. 1(d). The cars were equipped by V2X communication devices as shown in Fig. 2, which transmitted the GPS position and speed of vehicles every 0.1 s via basic safety messages (BSMs). Computers were used to record these data and tablets were connected wireless to the V2X devices so that passengers could monitor the positions of connected cars. The on-board prediction was implemented on the third vehicle, while all three cars were driven naturally without using the information from V2X units or the traffic prediction. The route of the experiment is shown in Fig. 3(a).

Our experimental setup enabled data collection at 10 Hz sampling frequency. Depending on the distance of the vehicles and other objects between them, the communication was sometimes interrupted and data packets were lost. In Fig. 3(b) the number of data packets sent (red) by the two lead vehicles and received (green) by the ego vehicle are shown against the distance between these vehicles. The corresponding packet delivery ratio (black) drops dramatically at about 250-300 m distance. Each vehicle recorded its own outgoing data packets too, which enabled us to overlap and post-process the data after the experiment despite packet losses. Our traffic prediction algorithm is validated via this aggregated dataset, while the technological limits of the V2X connectivity can be evaluated based on the packet delivery.

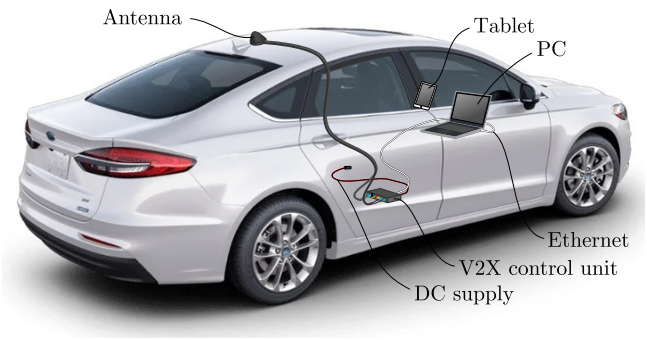

Fig. 2. Vehicle equipped with V2X communication device connected to an upper level computer for data collection and traffic prediction.

In Fig. 3(c)-(e), the recorded positions, speeds and distances of the three vehicles are plotted (with the color code of Fig. 1(d)) for 8 minutes of experiment in which heavy traffic congestion occurred. We use this dataset to test our traffic prediction algorithm. Continuous segments of the curves indicate uninterrupted communication between the ego and lead vehicles, whereas dashed segments relate to interrupted/poor communication. In Fig. 3(e), the distances between the ego and lead vehicles are plotted, showcasing another aspect of the distance over which packet losses occur. Below 250 m distance, stable connectivity was realized.

Observe in Fig. 3(c) that the ego vehicle overtook one of the lead vehicles at about 40 seconds (blue curve intersects the red one) since this leading vehicle was driven on a different lane of the highway. This explains why speed fluctuations in Fig. 3(d) do not correlate for all vehicles (red curve shows different behavior than others). In our experiment, drivers chose lanes based on their individual decisions. Since V2V packets did not contain lane information, we selected the useful lead vehicle data during post-processing intuitively, to avoid the use of uncorrelated data in our traffic prediction.

III. TRAFFIC PREDICTION METHOD

We separate the on-board traffic prediction problem into two steps: (i) traffic estimation and (ii) traffic prediction. *Traffic estimation* uses historical data about the lead and ego vehicles' trajectory to estimate the current state of traffic, including the number of vehicles between lead and ego and their positions and speeds. This relies on a nominal traffic model and Kalman filtering. The traffic state estimate is then

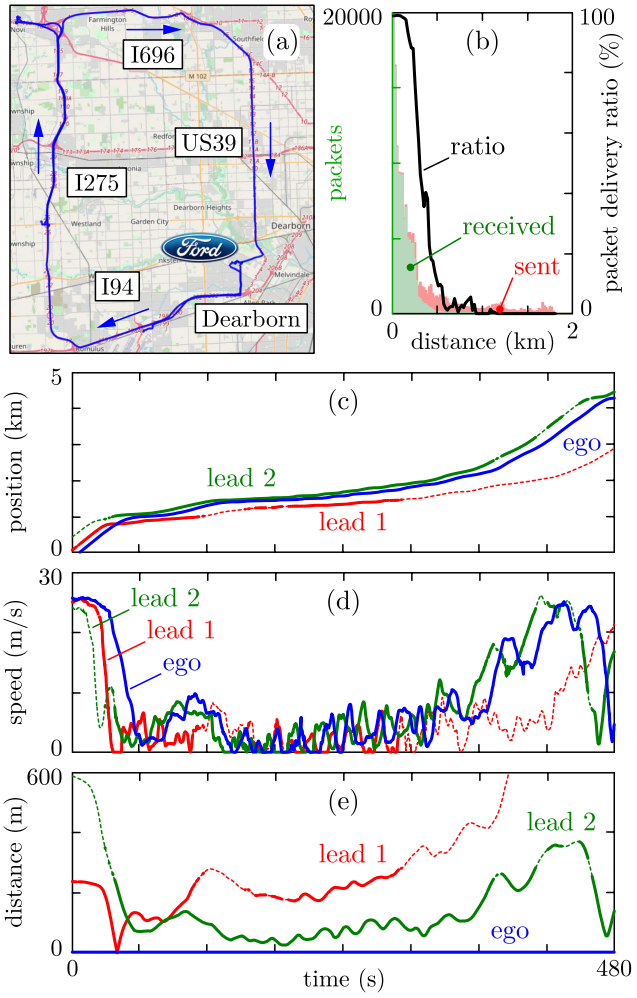

Fig. 3. (a) The route of the on-board traffic prediction experiment. (b) The number of data packets sent (red) and delivered (green) and the packet delivery ratio of V2V connectivity (black). (c)-(e) The position, speed and distance of the vehicles in the experiment with colors according to Fig. 1(d).

used as initial condition for *traffic prediction*, wherein the future motion of the ego vehicle (and the vehicles between lead and ego) is predicted based on a nominal model.

The framework is explained for the scenario in Fig. 1(a), where N vehicles follow the lead vehicle with the ego at the end. Vehicles are indexed from 0 to N starting at the ego with indices increasing in the direction of motion. Time is measured so that $t = 0$ when the prediction is made, i.e., traffic estimation is over $t \leq 0$ while prediction is for $t \geq 0$.

A. Traffic Estimation

We use the elementary car-following model of Newell [19] as nominal model for traffic estimation. Newell's model assumes that each vehicle copies the motion of its predecessor while keeping a time gap t_g and a standstill distance d_st. The position $X(n,t)$ of vehicle n as a function of time t is a copy of the position $X(n+1,t)$ of the preceding vehicle $n+1$, shifted in time by t_g and in space by d_st:

$$X(n,t) = X(n+1, t-t_\mathrm{g}) - d_\mathrm{st}. \tag{1}$$

By introducing the deviation $s(n,t) = X(n,t) - nd_\mathrm{st}$ from the standstill spacing, we transform out the spatial shift d_st:

$$s(n,t) = s(n+1, t-t_\mathrm{g}). \tag{2}$$

Similarly, the speed $v(n,t)$ of vehicle n is also a shifted copy of the speed $v(n+1,t)$:

$$v(n,t) = v(n+1, t-t_\mathrm{g}), \tag{3}$$

according to the time derivative of (1).

Newell's model captures the propagation of congestion waves upstream along the highway, where the wave speed is $w = d_\mathrm{st}/t_\mathrm{g}$. This relates to models that describe wave propagation by considering the traffic flow as continuum. In fact, (2) can also be written as $s(n,t) = s(0, t + nt_\mathrm{g})$, that is a solution to the linear Lighthill-Whitham-Richards model written in the Lagrangian (vehicle) frame [18], [20]:

$$\partial_t s(n,t) = \partial_n s(n,t)/t_\mathrm{g}. \tag{4}$$

As the solution to this continuum model exists for non-integer vehicle indices as well, we construct the following continuum equivalent of Newell's model (2)-(3):

$$\begin{aligned} s(n,t) &= s(n + \Delta n, t - \Delta t), \\ v(n,t) &= v(n + \Delta n, t - \Delta t), \end{aligned} \tag{5}$$

where $\Delta n > 0$ is a not necessarily integer shift in vehicle index while $\Delta t = t_\mathrm{g}\Delta n$ is the associated time shift. This continuum extension helps estimating the number of vehicles between lead and ego without restriction to integer numbers.

We perform traffic estimation in discrete time over the past $t \in [-T_\mathrm{p}, 0]$ by fixing Δt and using $t_k = k\Delta t$, $k \in \{-K_\mathrm{p} \dots, 0\}$, $T_\mathrm{p} = K_\mathrm{p}\Delta t$ with the corresponding discrete vehicle indices $n^\ell = \ell\Delta n$, $\ell \in \{0, \dots, L\}$, $N = L\Delta n$. The discretization of (5) leads to

$$\begin{aligned} s^\ell_{k+1} &= s^{\ell+1}_k, \\ v^\ell_{k+1} &= v^{\ell+1}_k. \end{aligned} \tag{6}$$

With the state defined as $\mathbf{x}^\ell_k = \begin{bmatrix} s^\ell_k & v^\ell_k \end{bmatrix}^\top \in \mathbb{R}^2$ we obtain

$$\mathbf{x}^\ell_{k+1} = \mathbf{a}^\ell \mathbf{x}^\ell_k + \mathbf{b}^\ell \mathbf{x}^{\ell+1}_k, \tag{7}$$

where $\mathbf{a}^\ell = \mathbf{0}$ and $\mathbf{b}^\ell = \mathbf{I}$ are 2×2 zero and identity matrices, respectively. Note that a wide class of linear(ized) car-following and Lagrangian continuum models can be written in the discretized form (7) with the appropriate coefficient matrices. Finally, considering $\ell \in \{0, \dots, L\}$ we get

$$\begin{bmatrix} \mathbf{x}^{L-1}_{k+1} \\ \mathbf{x}^{L-2}_{k+1} \\ \vdots \\ \mathbf{x}^0_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{a}^{L-1} & & & \\ \mathbf{b}^{L-2} & \mathbf{a}^{L-2} & & \\ & \ddots & \ddots & \\ & & \mathbf{b}^0 & \mathbf{a}^0 \end{bmatrix} \begin{bmatrix} \mathbf{x}^{L-1}_k \\ \mathbf{x}^{L-2}_k \\ \vdots \\ \mathbf{x}^0_k \end{bmatrix} + \begin{bmatrix} \mathbf{b}^{L-1} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \mathbf{x}^L_k, \tag{8}$$

where the coefficient matrix is lower bi-diagonal.

Here $\mathbf{x}^L_k = \begin{bmatrix} s^L_k & v^L_k \end{bmatrix}^\top = \begin{bmatrix} X^\mathrm{lead}_k - Nd_\mathrm{st} & v^\mathrm{lead}_k \end{bmatrix}^\top$ contains the lead vehicle's position X^lead_k and speed v^lead_k data, which are directly available from V2V communication. We consider the traffic estimation problem with input $\mathbf{u}_k = \mathbf{x}^L_k$. We intend to capture the position X^ego_k and speed v^ego_k

of the ego vehicle with our model, thus we consider the state $\mathbf{x}_k^0$ of the ego as output $\mathbf{y}_k = \mathbf{x}_k^0$. The states $\mathbf{x}_k^\ell$ of all follower vehicles $\ell \in \{0, \dots, L-1\}$ constitute the state $\mathbf{X}_k \in \mathbb{R}^{2(L+1)}$ of the traffic flow. Thus, the linear system model becomes

$$\begin{aligned} \mathbf{X}_{k+1} &= \mathbf{A}\mathbf{X}_k + \mathbf{B}\mathbf{u}_k + \boldsymbol{\xi}, \quad & \boldsymbol{\xi} &\sim \mathcal{N}(0, \mathbf{Q}), \\ \mathbf{y}_k &= \mathbf{C}\mathbf{X}_k + \boldsymbol{\eta}, & \boldsymbol{\eta} &\sim \mathcal{N}(0, \mathbf{R}), \end{aligned} \tag{9}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the coefficient matrices in (8), while $\mathbf{C}$ is a matrix whose last column block is $\mathbf{I}$ and all other column blocks are $\mathbf{0}$. Furthermore, we assume the system is subject to Gaussian process noise $\boldsymbol{\xi}$ and observation noise $\boldsymbol{\eta}$ with zero mean and covariance $\mathbf{Q}$ and $\mathbf{R}$, respectively

Finally, one needs the number L of trajectories between lead an ego, associated with $N = L\Delta n$ (not necessarily integer) number of vehicles. We estimate this as follows:

$$L = \frac{X^{\text{lead}}_{-K_\text{p}} - X^{\text{ego}}_{-K_\text{p}}}{(v^{\text{lead}}_{-K_\text{p}} + w)\Delta t}, \tag{10}$$

rounded to integer. Here the distance $X^{\text{lead}}_{-K_\text{p}} - X^{\text{ego}}_{-K_\text{p}}$ of the lead and ego vehicles at the start of traffic estimation is divided by $(v^{\text{lead}}_{-K_\text{p}} + w)$ to get the time it takes for a congestion wave to travel with speed w from the lead to the ego (if the lead travels at constant speed $v^{\text{lead}}_{-K_\text{p}}$). This yields the total time shift between lead and ego, and division by the time shift Δt between two subsequent trajectories gives the number L of trajectories. Inferring the number of vehicles between lead and ego with more sophisticated methods and as a function of time is a potential future research direction.

We use a standard Kalman filter to estimate the state $\mathbf{X}$, which consists of two steps. The first step is prediction that obtains *a priori* distribution and the second step is correction which gets *a posteriori* distribution. The prediction step is

$$\begin{aligned} \hat{\mathbf{X}}_{k|k-1} &= \mathbf{A}\hat{\mathbf{X}}_{k-1|k-1} + \mathbf{B}\mathbf{u}_{k-1}, \\ \mathbf{P}_{k|k-1} &= \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^\top + \mathbf{Q}, \end{aligned} \tag{11}$$

where subscript $k|k-1$ shows that the prediction is for time k based on estimations up to time $k-1$, and $\mathbf{P}$ is the covariance of the state estimate $\hat{\mathbf{X}}$. This prediction is corrected based on current measurement $\mathbf{y}_k^\text{m}$. Namely, the difference of the measured output $\mathbf{y}_k^\text{m} = \begin{bmatrix} X_k^{\text{ego}} & v_k^{\text{ego}} \end{bmatrix}^\top$ and the output $\mathbf{C}\hat{\mathbf{X}}_{k|k-1}$ produced by the nominal model is fed back using an optimally chosen Kalman gain $\mathbf{M}_k$:

$$\begin{aligned} \mathbf{M}_k &= \mathbf{P}_{k|k-1}\mathbf{C}^\top(\mathbf{C}\mathbf{P}_{k|k-1}\mathbf{C}^\top + \mathbf{R})^{-1}, \\ \hat{\mathbf{X}}_{k|k} &= \hat{\mathbf{X}}_{k|k-1} + \mathbf{M}_k(\mathbf{y}_k^\text{m} - \mathbf{C}\hat{\mathbf{X}}_{k|k-1}), \\ \mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{M}_k\mathbf{C})\mathbf{P}_{k|k-1}. \end{aligned} \tag{12}$$

These expressions can be computed efficiently given the structure of $\mathbf{C}$ (since all but one blocks of $\mathbf{C}$ are zero).

B. Traffic Prediction

The traffic prediction process simulates the nominal model further into the future $t \in [0, T_\text{f}]$ by evaluating (11) for $k \in \{0, \dots, K_\text{f}\}$, $T_\text{f} = K_\text{f}\Delta t$ with initial condition $\hat{\mathbf{X}}_0$. Since no data are available for the future, the correction steps (12) are no longer carried out. The covariance $\mathbf{P}$ is still calculated to evaluate the uncertainty of predictions.

The lead vehicle's past motion affects the ego vehicle's future motion up to a horizon $t \in [0, T_\text{h}]$, that is, $k \in \{0, \dots, K_\text{h}\}$, $T_\text{h} = K_\text{h}\Delta t$. This horizon is determined by the time over which the congestion wave of speed w propagates from the lead to the ego: K_h is the maximum value of k that satisfies

$$X_k^0 \leq X_0^L - wk\Delta t, \tag{13}$$

and it depends on the state of traffic. To perform simulations, we used constant speed future motion for the lead vehicle as input $\mathbf{u}_k$, and we limited the ego's predicted trajectory to the horizon $k \leq K_\text{h}$. Beyond this horizon ($k > K_\text{h}$) we disregarded predictions due to the bias caused by the constant speed assumption for the lead vehicle.

In the simulations we used $d_\text{st} = 10\,\text{m}$, $t_\text{g} = 1.67\,\text{s}$ (that yield $w = 6\,\text{m/s}$) and $\Delta t = 0.1\,\text{s}$. The duration of the traffic estimation was set to $T_\text{p} = (X_0^\text{lead} - X_0^\text{ego})/(v_0^\text{lead} + w)$ (rounded to integer multiples of Δt) that is the time shift between lead and ego at the moment of prediction; cf. (10). This keeps the amount of simulations reasonably low for real-time implementation. We forecasted over the maximum achievable horizon $T_\text{f} = T_\text{h}$. We specified the initial condition $\hat{\mathbf{X}}_{-K_\text{p}|-K_\text{p}}$ for (11) by linearly interpolating L number of states between the lead and ego vehicles' data. We set the initial covariance $\mathbf{P}_{-K_\text{p}|-K_\text{p}} = \mathbf{0}$. We chose the elements of $\mathbf{Q}$ and $\mathbf{R}$ such that $Q_{ij} = R_{ij} = 1$ if i, j are odd (covariance between two position terms), $Q_{ij} = R_{ij} = 0.1$ if i, j are even (covariance between two speed terms) and $Q_{ij} = R_{ij} = 0$ otherwise (covariance between a position and a speed term).

IV. APPLICATION AND RESULTS

We implemented the proposed traffic prediction algorithm on a vehicle and tested it in real traffic during the congestion of Fig. 3. The results below were obtained via the codes that were run in the experiment (apart from minor modifications). While the experiment included three vehicles, for simplicity we present predictions for a single lead-ego pair (green and blue in Figs. 1-3). We remark, however, that data from multiple lead vehicles may allow predictions for multiple lanes, or in case of a single lane they can be leveraged by providing multiple speed previews (one per lead vehicle) or a single combined prediction. A possible combination is to take the farthest lead vehicle's data as input $\mathbf{u}_k$ and include the other lead vehicles' data in the observation $\mathbf{y}_k$ when using the Kalman filter. Although this was also implemented, we omit further discussions on multiple lead vehicles.

Fig. 4(a)-(b) show the prediction at a selected time moment (vertical line at $t = 0$). Over $t \leq 0$ traffic estimation was conducted using historical data (green and blue). This resulted in the estimated trajectories s_k^ℓ and v_k^ℓ (gray). We plotted only those trajectories whose vehicle index n^ℓ is close to integer. The estimated state at $t = 0$ was used as initial condition for traffic prediction over $t \geq 0$, which produced the ego's predicted trajectory (see the expectation in black). The uncertainty of the prediction is indicated by

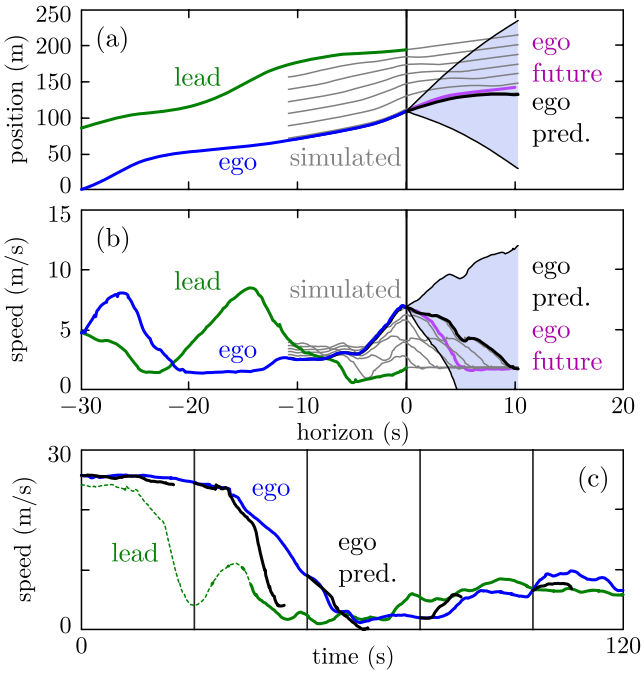

Fig. 4. (a)-(b) Illustration of traffic estimation and prediction at a selected time moment with position and speed plots. (c) Example of five predictions over a two minute time frame. Note that the prediction horizon changes over time as the distance and speeds of the vehicles change; cf. (13).

shading for one standard deviation (68% credible interval). This was obtained from those diagonal elements of $\mathbf{P}_{k|k}$ that correspond to the ego's state $\mathbf{x}_k^0$.

Fig. 4(c) illustrates how the prediction is repeated over time: at each time moment (vertical lines) a new prediction is plotted. While this example shows predictions at every 25 seconds, the algorithm is able to forecast with a higher update rate (within about a second including reading the data and plotting). The predictions (black) match the ego vehicle's actual trajectory (blue) qualitatively, and the quantitative accuracy is calculated below.

We evaluated the accuracy offline after the experiment, by including the data that were recorded by the lead vehicle but were unavailable to the ego due to packet losses. This separates the technology-related effects of imperfect communication from the methodology-related effects of prediction inaccuracy. Fig. 5 indicates the prediction error. A prediction was made (like in Fig. 4) at every second and the absolute error between prediction (black) and ground truth (purple) was calculated for each point along the horizon. This yields the error as a function of prediction time and horizon. The error of the Kalman filter algorithm is shown in Fig. 5(a). The color tells the magnitude of the error while the envelope of the colored area shows the achievable horizon found by (13).

The prediction horizon depends on the traffic state, including the distance D of lead and ego and the speed v of traffic; its rough estimate is $T_\mathrm{h} \approx D/(v+w)$. The horizon reached 20 seconds in this particular traffic jam. In general, the horizon is longer if the ego obtains data from farther ahead. Thus, the range of V2V connectivity (shown by Fig. 3(b)) limits the achievable horizon. By driving through multiple congestions, we observed that the method can provide predictions for up to 40-60 seconds with current V2V technology. The range and horizon may be extended by other means of connectivity, such as communication with the infrastructure.

The horizon has trade-off with accuracy: predictions farther into the future tend to be less accurate. This is shown by Fig. 5(a) where the color gets brighter towards top. The inset highlights this by zooming into the first 90 predictions. Here the ego vehicle reached the traffic jam and had a significant slowdown; cf. Fig. 3(d). Such large changes in the speed are especially important to predict and can potentially lead to high prediction errors. Using the data from connectivity, the Kalman filter algorithm manages to make predictions with good accuracy over most of the time and horizon.

Fig. 5(b) compares the proposed method to constant speed prediction over the same horizon. As this baseline does not use the lead vehicle's data from connectivity, it fails to predict the slowdown; see the high error in the inset. For low speeds and short horizons, however, it may achieve low error. Fig. 5(c) shows the case where the lead vehicle's data are used but state estimation relies on the nominal model without Kalman filter. Since it is no longer ensured that the estimated state stays close to the historical data, there is a discontinuity between the last available speed data and the speed predicted with the simple model (6). Thus, the prediction error is nonzero even at the smallest (zero) horizon.

We also evaluated the root mean square (RMS) prediction error. In Fig. 5(d), the RMS is calculated by taking the mean over the available horizon for each prediction time. The plot shows that connectivity helps us reduce the prediction error (green and blue) compared to constant speed prediction (gray), and the Kalman filter (green) further improves the results. The difference between the three methods is significant in the beginning (during a major slowdown) and at the end (during a speed up). In Fig. 5(e), the mean in the RMS is calculated over time and the result is plotted against the prediction horizon. As the horizon increases, the prediction error increases for all methods. The constant speed prediction (gray) has rapid error growth and may not be satisfactory for long horizons, while the information from connectivity reduces the error significantly (green and blue). Without Kalman filter (blue), the error is nonzero even at zero horizon, while the Kalman filter (green) ensures continuous speed preview and low errors, even when applied with a simple traffic flow model like (6).

V. CONCLUSIONS

We have proposed a traffic prediction method that allows connected vehicles on highways to forecast their future trajectories based on the motion of other connected vehicles ahead. We have decomposed the problem into traffic estimation using historical data and traffic prediction about the future, both relying on traffic flow models. The traffic state estimation was supported by Kalman filter. The prediction method has been implemented on a physical vehicle and onboard real-time predictions have been tested in real traffic via

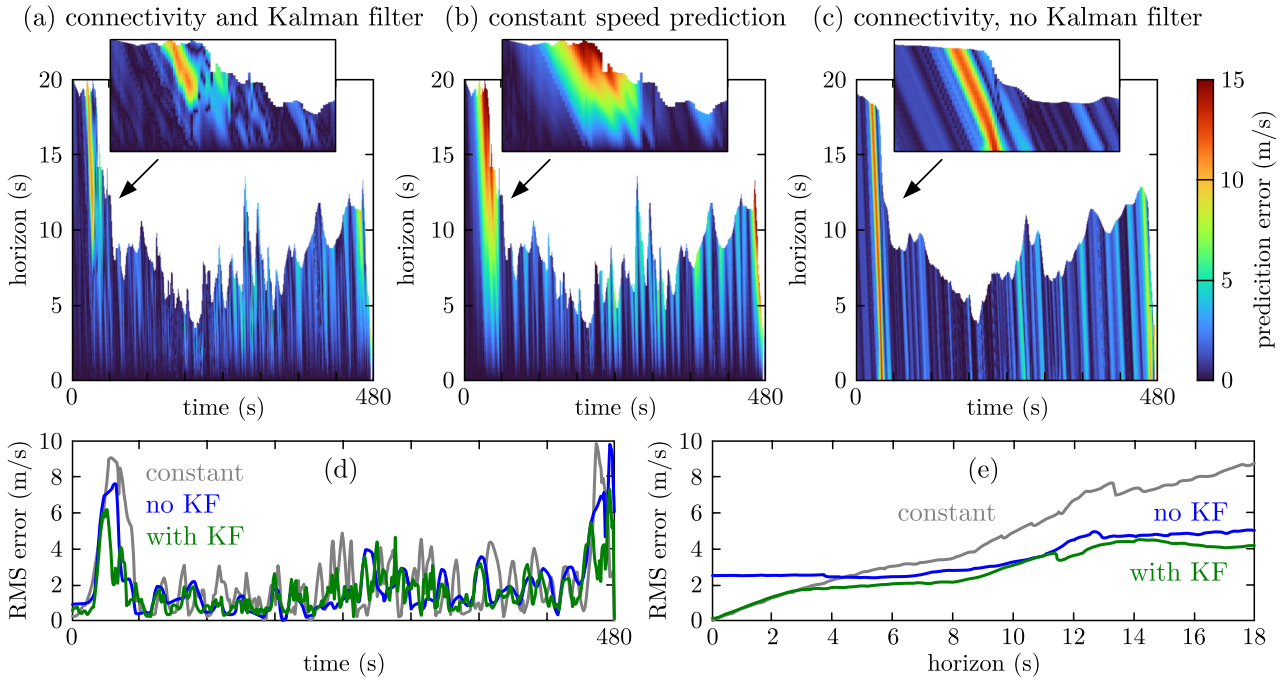

Fig. 5. (a)-(c) The absolute prediction error for various prediction times and horizons with three different methods: (a) using the lead vehicle data from connectivity with Kalman filter, (b) constant speed prediction, and (c) prediction without Kalman filter. The insets zoom into the first 90 predictions when there was a major slowdown upon reaching the traffic jam. (d)-(e) The RMS prediction error over the horizon and the prediction time.

experiments with connected human-driven vehicles. Future research directions include integrating our method with more sophisticated traffic models and supplementing it with data from vehicle-to-infrastructure communication.

References

[1] R. A. Dollar and A. Vahidi, "Efficient and collision-free anticipative cruise control in randomly mixed strings," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 4, pp. 439–452, 2018.

[2] M. A. S. Kamal, T. Hayakawa, and J. Imura, "Road-speed profile for enhanced perception of traffic conditions in a partially connected vehicle environment," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 6824–6837, 2018.

[3] G. Orosz, R. E. Wilson, and G. Stépán, "Traffic jams: dynamics and control," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4455–4479, 2010.

[4] B. Mehran, M. Kuwahara, and F. Naznin, "Implementing kinematic wave theory to reconstruct vehicle trajectories from fixed and probe sensor data," in *Proceedings of the 19th International Symposium on Transportation and Traffic Theory*, vol. 17, 2011, pp. 247–268.

[5] H. Yu, Q. Gan, A. Bayen, and M. Krstic, "PDE traffic observer validated on freeway data," *IEEE Transactions on Control Systems Technology*, pp. 1–13, 2020.

[6] D. B. Work and A. M. Bayen, "Impacts of the mobile internet on transportation cyberphysical systems: Traffic monitoring using smartphones," in *Proceedings of the National Workshop for Research on High-Confidence Transportation Cyber-Physical Systems: Automotive, Aviation and Rail*, Washington, DC, USA, 2008.

[7] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment," *Transportation Research Part C*, vol. 18, no. 4, pp. 568–583, 2010.

[8] 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 *Proceedings of the 47th IEEE Conference on Decision and Control*, Cancun, Mexico, 2008, pp. 5062–5068.

[9] D. B. Work, O.-P. Tossavainen, Q. Jacobson, and A. M. Bayen, "Lagrangian sensing: Traffic estimation with mobile devices," in *Proceedings of the American Control Conference*, St. Louis, MO, USA, 2009, pp. 1536–1543.

[10] J. C. Herrera and A. M. Bayen, "Incorporation of Lagrangian measurements in freeway traffic state estimation," *Transportation Research Part B*, vol. 44, no. 4, pp. 460–481, 2010.

[11] M. L. Delle Monache, T. Liard, B. Piccoli, R. Stern, and D. Work, "Traffic reconstruction using autonomous vehicles," *SIAM Journal on Applied Mathematics*, vol. 79, no. 5, pp. 1748–1767, 2019.

[12] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation per lane in the presence of connected vehicles," *Transportation Research Part B*, vol. 106, pp. 1–28, 2017.

[13] K. Chu, R. Saigal, and K. Saitou, "Stochastic Lagrangian traffic flow modeling and real-time traffic prediction," in *Proceedings of the IEEE International Conference on Automation Science and Engineering*, 2016, pp. 213–218.

[14] ——, "Real-time traffic prediction and probing strategy for Lagrangian traffic data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 497–506, 2019.

[15] Y. Yuan, J. W. C. van Lint, R. E. Wilson, F. van Wageningen-Kessels, and S. P. Hoogendoorn, "Real-time Lagrangian traffic state estimator for freeways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 59–70, 2012.

[16] Y. Yuan, H. Van Lint, F. Van Wageningen-Kessels, and S. Hoogendoorn, "Network-wide traffic state estimation using loop detector and floating car data," *Journal of Intelligent Transportation Systems*, vol. 18, no. 1, pp. 41–50, 2014.

[17] A. Duret and Y. Yuan, "Traffic state estimation based on Eulerian and Lagrangian observations in a mesoscopic modeling framework," *Transportation Research Part B*, vol. 101, pp. 51–71, 2017.

[18] T. G. Molnár, D. Upadhyay, M. Hopka, M. Van Nieuwstadt, and G. Orosz, "Delayed Lagrangian continuum models for on-board traffic prediction," *Transportation Research Part C*, vol. 123, p. 102991, 2021.

[19] G. F. Newell, "A simplified car-following theory: a lower order model," *Transportation Research Part B*, vol. 36, no. 3, pp. 195–205, 2002.

[20] L. Leclercq and J. Laval, "A multiclass car-following rule based on the LWR model," in *Traffic and Granular Flow '07 Part I*. Springer, 2009, pp. 151–160.