# Data-driven Predictive Connected Cruise Control

Minghao Shen[1] and Gábor Orosz[1,2]

2023 IEEE Intelligent Vehicles Symposium (IV) | 979-8-3503-4691-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/IV55152.2023.10186677

***Abstract*— In this paper, we propose a data-driven predictive controller for connected automated vehicles (CAVs) traveling in mixed traffic consisting of both connected and non-connected vehicles. We assume a low penetration of connectivity, with only one connected vehicle in the downstream traffic. A model predictive controller is designed to integrate multiple specifications, including safety and energy efficiency, while accounting for the time delay in the longitudinal dynamics of the vehicle. A data-driven prediction method based on the behavioral theory of linear systems is proposed to model the relationship between the speeds of the distant connected vehicle and the vehicle immediately in front of the CAV. The proposed method is evaluated using real traffic data and demonstrates improved prediction accuracy and energy efficiency compared to model-based prediction methods.**

## I. Introduction

Adaptive cruise control (ACC) was invented in the 1990s and is now commonly used in passenger vehicles. Its purpose is to maintain a safe distance from the vehicle in front while driving at the desired speed. However, ACC relies on the movement of the vehicle in front, so it can be sensitive to changes in that vehicle's speed. Predicting the future movement of the preceding vehicle can be difficult because of limited knowledge of traffic conditions downstream.

With vehicle-to-vehicle (V2V) connectivity, connected automated vehicles can access information about other vehicles beyond their line of sight and potentially cooperate with them in traffic [1]–[3]. Assuming all vehicles in traffic are connected, extensive research has shown that controllers such as cooperative adaptive cruise control (CACC) [4]–[7] and platooning [8]–[10], have the potential to save energy and improve traffic efficiency. However, the adoption rate of connected automated vehicles is currently very low, and it may take decades to reach a significant level of penetration. Connected cruise control (CCC) is a control method that only requires a low level of cooperation (status-sharing) and has been shown via simulation and experiment to bring significant benefits in energy efficiency [11], even with a lean penetration of connected vehicles in traffic [12], [13].

Without V2V connectivity, there is often limited information about traffic, making it difficult to predict the future motion of the preceding vehicle. With V2V communication, integrated with traffic models, the additional information can potentially improve the prediction. On one hand, car-following models can be applied to simulate the vehicle chain in the traffic from the distant connected vehicle to the vehicle immediately in front. However, the number of vehicles between is typically unknown, which requires online estimation algorithms [14], [15]. On the other hand, continuum traffic models [16], [17] and/or neural network models [18] can be used for prediction, but these typically require offline training.

To address aforementioned issues of model-based prediction, we propose a data-driven prediction that directly captures the relationship between the distant connected vehicle and the vehicle immediately in front. The behavioral theory of linear systems provides theoretical foundation for our method. In our data-driven approach, we do not need to explicitly spell out the parameters of the model. Instead, we provide a representation of the trajectory space and predict future motions within that space [19]. Based on the behavioral theory, data-enabled predictive control (DeePC) was developed [20] and applied to quadcopter control [21] and to design a cooperative control of connected automated vehicles in mixed traffic [22], [23].

In this paper, we apply behavioral theory to predict the motion of the preceding vehicle based on the information of a distant connected vehicle, which is not required to be automated. Thus, the connected cruise controller we design may operate with lean penetration of connectivity and automation. Specifically, we propose a data-driven method to predict the motion of preceding vehicle based on the information of connected vehicles in the downstream traffic. The proposed method is evaluated with real human driving data. Compared to model-based methods, the proposed data-driven prediction achieves higher prediction accuracy and results in better energy efficiency, while circumventing the need for offline training and/or model calibration.

## II. Dynamics and Modeling

Consider a connected automated vehicle traveling in a chain of vehicles consisting of connected vehicles and non-connected vehicles. As shown in Fig. 1, the ego CAV has access to the position and speed of the vehicle 1 immediately in front via on-board sensors (like radar, camera or LIDAR). In addition, when there is a connected vehicle (vehicle $L$) in the downstream traffic, its position and velocity are also accessible to the ego vehicle through V2V communication. However, for the the non-connected vehicles driving between the vehicle 1 and vehicle $L$, the number of vehicles and their states are not available to the ego vehicle. We refer to these vehicles as *hidden vehicles*.

This work was supported by the Center for Connected and Automated Transportation, University of Michigan, through the U.S. Department of Transportation (DOT) under Grant 69A3551747105.

[1]M. Shen and G. Orosz are with the Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA {mhshen, orosz}@umich.edu

[2]G. Orosz is also with the Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, USA

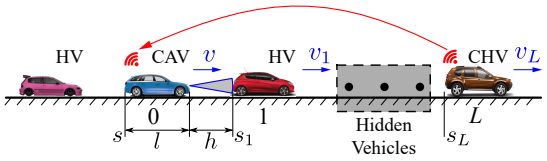


Fig. 1. Connected cruise control (CCC) strategy. The CAV (blue vehicle) has access to the position and velocity of the HV 1 (red) immediately in the front, and the distant CHV $L$ (green), but it does not have access to the states of non-connected hidden vehicles vehicles driving between vehicles 1 and $L$.

The longitudinal dynamics of the ego vehicle is given as

$$\begin{aligned} \dot{s} &= v, \\ \dot{v} &= -\frac{1}{m_{\text{eff}}}\left(mg\xi + kv^2\right) + \frac{T_{\text{w}}}{m_{\text{eff}}R}. \end{aligned} \tag{1}$$

Here the effective mass $m_{\text{eff}} = m + I/R^2$ incorporates the mass $m$, the moment of inertia $I$ of rotating elements, and the radius $R$ of wheels. In addition, $g$ denotes the gravitational acceleration, $\xi$ denotes the rolling resistance coefficient and $k$ denotes the air drag coefficient. The speed is controlled by the wheel torque $T_{\text{w}}$ generated by the engine/electric motors and the brakes. To highlight how the commanded acceleration $u$ influences the systems, we rewrite (1) as

$$\begin{aligned} \dot{s}(t) &= v(t), \\ \dot{v}(t) &= -f\big(v(t)\big) + \text{sat}\big(u(t-\sigma)\big). \end{aligned} \tag{2}$$

where

$$f(v) = \frac{1}{m_{\text{eff}}}\left(mg\xi + kv^2\right), \quad \text{sat}\big(u(t-\sigma)\big) = \frac{T_{\text{w}}}{m_{\text{eff}}R}. \tag{3}$$

Here we consider two additional physical effects: the time delay $\sigma$ from the powertrain, and the saturation resulted from the limited engine/motor torque and power and braking torque. More specifically, we consider the saturation function

$$\text{sat}(u) = \min\left\{m_1 v + b_1, m_2 v + b_2, \max\{u_{\min}, u\}\right\}, \tag{4}$$

as is shown in Fig. 2(a).

In order to follow the desired acceleration $a_{\text{d}}$, the control command

$$u = \tilde{f}(v) + a_{\text{d}}, \tag{5}$$

is applied, where the term $\tilde{f}(v)$ is used to compensate for the nonlinear physical effects $f(v)$. In this paper, we assume perfect compensation and focus on the design of desired acceleration $a_{\text{d}}$. The dynamics (2) can be simplified to

$$\begin{aligned} \dot{s}(t) &= v(t), \\ \dot{v}(t) &= a(t), \\ a(t) &= \text{sat}\big(a_{\text{d}}(t-\sigma)\big), \end{aligned} \tag{6}$$

Below we design the desired acceleration $a_{\text{d}}$ utilizing the information of vehicle 1 and vehicle $L$. Such control design is referred to as *connected cruise control (CCC)*.

## III. Data-driven Control Design

In this section, we introduce our data-driven control design method. We first introduce the framework of predictive connected cruise control (PCCC) design, and then introduce data-driven PCCC.

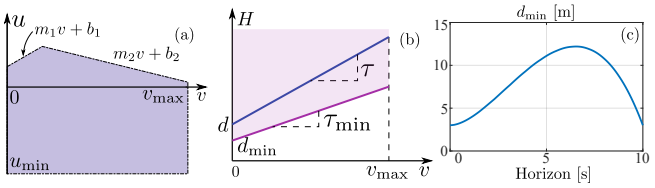


Fig. 2. Nonlinear functions in PCCC controller design. (a) Saturation function (4). (b) Range policy (9) and safety condition (10). (c) Time-varying $d_{\min}$ in safety constraint (10).

### A. Predictive Connected Cruise Control

Predictive connected cruise control (PCCC) [13] applies model predictive control (MPC) that combines controller design specifications and constraints into an optimization problem, and solves it in a receding horizon fashion. The optimization is usually formulated in discrete time. Thus, we discretize the longitudinal dynamics (6) with time step $\Delta t$ and let $\sigma = q\Delta t$. This results in

$$\begin{aligned} s(k+1) &= s(k) + \Delta t\ v(k) + \frac{1}{2}\Delta t^2 a(k), \\ v(k+1) &= v(k) + \Delta t\ a(k), \\ a(k) &= a_{\text{d}}(k-q). \end{aligned} \tag{7}$$

The position and velocity of the ego vehicle are combined into the state $x = [s\ v]^\top$, while $\hat{x}_1 = [\hat{s}_1\ \hat{v}_1]^\top$ denotes the state predictions the vehicle 1 immediately in front.

In this paper, we choose objective function

$$\ell(x, \hat{x}_1, a) = \big(\hat{h} - H(v)\big)^2 + q_a a^2, \tag{8}$$

where $\hat{h} = \hat{s}_1 - s - l$ is the predicted headway depending on the prediction of future motion of vehicle 1. The range policy

$$H(v) = \tau v + d, \tag{9}$$

characterizes the desired headway as a function of speed: we can accept short headway when speed is low, but large distance is preferred at high speed. The second term penalizes rapid acceleration or deceleration, to achieve better driving comfort and energy efficiency. The parameter $q_a$ balances the two objectives.

The safety constraint is given as

$$\hat{h} - H_{\min}(v) \geq 0, \quad H_{\min}(v) = \tau_{\min} v + d_{\min}, \tag{10}$$

where $H_{\min}(v)$ denotes the minimum safe distance for given velocity; see Fig. 2(b). We can choose $d_{\min}$ as a function of prediction horizon to compensate for prediction uncertainties [24]. As is shown in Fig. 2(c), $d_{\min}$ grows at the beginning since the uncertainty grows with the prediction horizon. After 6.5 [s], $d_{\min}$ decreases again assuming the receding horizon control can deal with the uncertainties in the following steps.

In summary, the optimization problem is formulated as

$$\begin{aligned}\min_{a(0|k),\dots,a(K_\mathrm{f}-1+q|k),\epsilon}\ & \sum_{i=0}^{K_\mathrm{f}}\Big(\hat{h}(i|k)-H\big(v(i|k)\big)\Big)^2 \\ & + q_\mathrm{a}\sum_{i=0}^{K_\mathrm{f}-1} a^2(i|k) + q_\epsilon\epsilon,\end{aligned}$$

s.t.

$$\left.\begin{aligned} s(i+1|k) &= s(i|k)+\Delta t\ v(i|k) \\ &\quad +\tfrac{1}{2}\Delta t^2\ a(i|k), \\ v(i+1|k) &= v(i|k)+\Delta t\ a(i|k), \end{aligned}\right\} i=0,\dots,K_\mathrm{f}-1,$$

$$\left.\begin{aligned} &\hat{h}(i|k)=\hat{s}_1(i|k)-s(i|k)-l, \\ &\hat{h}(i|k)-H_{\min}\big(v(i|k)\big)\geq -\epsilon, \\ &0\leq v(i|k)\leq v_{\max}, \\ &u_{\min}\leq a(i|k), \\ &a(i+q|k)\leq m_1 v(i+q|k)+b_1, \\ &a(i+q|k)\leq m_2 v(i+q|k)+b_2, \\ &a(i|k)=a_\mathrm{d}(k+i-q), \end{aligned}\right\} i=0,\dots,K_\mathrm{f},$$

$$s_1(0|k)=s_1(k),\quad s(0|k)=s(k),\quad v(0|k)=v(k), \tag{11}$$

where $x(i|k)$ means the predicted $x(k+i)$ based on the information available at time $k$. The equality constraints $x(i|k)=x(k)$, $s_1(i|k)=s_1(k)$ and $a(i|k)=a_\mathrm{d}(k+i-q)$ enforce timely updates of the prediction based on the observation data, which serves as an implicit feedback. In addition, $\epsilon$ is used to soften the safety constraint to ensure feasibility of the optimization problem. We choose the weight $q_\epsilon=10^6$ to put a heavy penalty on the violation of the safety constraint.

In PCCC, high quality of prediction $\hat{s}_1$ is crucial for the performance of the controller. With connectivity, the ego vehicle has access to more information beyond line of sight in the downstream traffic. Exploiting the correlation between $x_L=[s_L\ v_L]^\top$ and $x_1=[s_1\ v_1]^\top$, we can potentially make better prediction of the future motion of vehicle 1. However, the correlation between $x_L$ and $x_1$ is challenging to unveil due to the hidden vehicles: we do not know either the number of hidden vehicles $n_\mathrm{h}$, or the states of the hidden vehicles. We introduce a data-driven method to solve these problems.

### *B. Behavioral Theory of Linear Time Invariant Systems*

In this section, we provide background knowledge on behavioral theory of linear time invariant (LTI) systems. There are two representations of LTI systems [19]. On one hand, an LTI system can be characterized by the input-output relations, which is described by an autoregressive model with exogenous input (ARX):

$$y(k+1)=\sum_{i=0}^{K_\mathrm{p}}\phi_i y(k-i)+\sum_{i=0}^{K_\mathrm{p}}\psi_i u(k-i). \tag{12}$$

On the other hand, an LTI system can also be described by a state-space model:

$$\begin{aligned}\Sigma: x(k+1)&=Ax(k)+Bu(k),\\ y(k)&=Cx(k)+Du(k).\end{aligned} \tag{13}$$

The state variable $x$ is introduced to achieve the Markov property. The two representations can be transformed back and forth. Especially, the transform from ARX model to state-space form is widely studied as *realization* problem. The solution is to realization problem is not unique, and the state-space model that has the minimum state dimension is called *minimum realization*.

Classical system identification methods usually fit system parameters from data, and make predictions on the system behavior based on the fitted system parameters. On the other hand, the behavioral theory of LTI systems does not require explicitly writing down the system parameters, instead focuses on the subspace of all possible system trajectories.

Consider a minimal realization system (13) with state $x\in\mathbb{R}^{n_x}$, input $u\in\mathbb{R}^{n_u}$ and output $y\in\mathbb{R}^{n_y}$. The behavioral theory studies the space of system trajectories of length $K$:

$$\begin{aligned}\mathcal{W}_K(\Sigma)=&\big\{[u(0:K-1)^\top\ y(0:K-1)^\top]^\top \\ &|\{u,y\}\text{ are trajectories of (13)}\big\},\end{aligned} \tag{14}$$

where we introduced $u(0:K-1)=[u(0)\ \cdots\ u(K-1)]^\top$ and $y(0:K-1)$ is defined similarly. Given the data of a system trajectory $u^\mathrm{d}(0:T-1)$ and $y^\mathrm{d}(0:T-1)$ of length $T>K$, we define the Hankel matrix of order $K$ as

$$\mathcal{H}_K(u^\mathrm{d})=\begin{bmatrix} u^\mathrm{d}(0) & u^\mathrm{d}(1) & \cdots & u^\mathrm{d}(T-K+1)\\ u^\mathrm{d}(1) & u^\mathrm{d}(2) & \cdots & u^\mathrm{d}(T-K+2)\\ \vdots & \vdots & \ddots & \vdots\\ u^\mathrm{d}(K-1) & u^\mathrm{d}(K) & \cdots & u^\mathrm{d}(T)\end{bmatrix}, \tag{15}$$

and $\mathcal{H}_K(y^\mathrm{d})$ is defined similarly. Then, if the trajectory data is diverse enough, the space of system input/output trajectories can be represented with the range spaces of Hankel data matrices. The diversity of the input trajectory is described by the persistence of excitation.

**Definition 1** (Persistence of excitation [25])**.** The input trajectory $u^\mathrm{d}$ is persistent excitation of order $K$ if the Hankel matrix $\mathcal{H}_K(u^\mathrm{d})$ has full row rank.

The so-called Fundamental Lemma provides a representation of $\mathcal{W}_K(\Sigma)$.

**Lemma 1** (Fundamental Lemma [25], [26])**.** *If the data input trajectory $u^\mathrm{d}$ is persistent excitation of order $n_x+K$, then the space $\mathcal{W}_K(\Sigma)$ can be represented as the range space of the Hankel data matrix:*

$$\mathcal{W}_K(\Sigma)=\mathrm{range}\left(\begin{bmatrix}\mathcal{H}_K(u^\mathrm{d})\\ \mathcal{H}_K(y^\mathrm{d})\end{bmatrix}\right). \tag{16}$$

### *C. Data-driven Prediction*

In order to utilize the information from $x_L$ to predict $x_1$, as is shown in Fig. 3, one may train a car-following model offline, estimate the number of hidden vehicles $n_\mathrm{h}$ online, and simulate the vehicle chain [13], [15]. However, the prediction accuracy heavily relies on the quality of the model. In this paper, we drop the intermediate steps, and directly model the input-output relationship between $x_L$ and $x_1$. Specifically, we consider an LTI system that takes $x_L$ as input and outputs $x_1$; see the green box in Fig. 3,

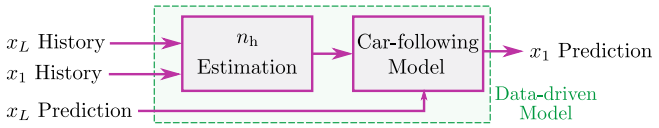


Fig. 3. Model-based and data-driven prediction. Model-based prediction estimates the number of hidden vehicles and simulates the vehicle chain using car-following models calibrated offline. Data-driven prediction captures the relationship between $x_L$ and $x_1$ directly from data.

For PCCC (11), we need to make predictions of $\hat{s}_1(i|k),\ i = 0, \dots, K_\mathrm{f}$ based on observations of input $x_L(j)$ and output $x_1(j),\ j = 0, \dots, k$. Due to the randomness of human driver behavior, the persistence of excitation is usually satisfied. Assume the length of initial condition is $K_\mathrm{p}$, then let $K = K_\mathrm{f} + K_\mathrm{p}$. By the Fundamental Lemma 1, there exists a vector $g \in \mathbb{R}^{k-K+1}$, such that

$$\begin{bmatrix} \mathcal{H}_K(x_L(0:k)) \\ \mathcal{H}_K(x_1(0:k)) \end{bmatrix} g = \begin{bmatrix} x_L(k-K_\mathrm{p}+1:k) \\ x_1(k-K_\mathrm{p}+1:k) \\ x_L(k+1:k+K_\mathrm{f}) \\ x_1(k+1:k+K_\mathrm{f}) \end{bmatrix}. \tag{17}$$

Based on the historical data $x_L(k-K_\mathrm{p}+1:k)$ and $x_1(k-K_\mathrm{p}+1:k)$ and the prediction of future trajectory of $x_L(k+1:k+K_\mathrm{f})$, we can derive all the possible future trajectory of $x_1(k+1:k+K_\mathrm{f})$. In practice, the system with input $x_L$ and output $x_1$ is not a perfect LTI system. Taking the modeling error into consideration, we construct the quadratic programming problem

$$\min_{g,e} \ \|e\|_2^2 + \lambda_g \|g\|_2^2, \tag{18}$$

$$\text{s.t.} \ \begin{bmatrix} \mathcal{H}_K(x_L(0:k)) \\ \mathcal{H}_K(x_1(0:k)) \end{bmatrix} g = \begin{bmatrix} x_L(k-K_\mathrm{p}+1:k) \\ x_1(k-K_\mathrm{p}+1:k) \\ x_L(k+1:k+K_\mathrm{f}) \\ x_1(k+1:k+K_\mathrm{f}) \end{bmatrix} + \begin{bmatrix} 0 \\ e \\ 0 \\ 0 \end{bmatrix},$$

where $\lambda_g$ is the regularization constant. Here we set $\lambda_g = 0.1$.

Since we do not have extra information about the vehicle $L$, one common choice is to assume its speed remains constant

$$v_L(j) = v_L(k), \quad j = k+1, \dots, k+K_\mathrm{f}. \tag{19}$$

On the other hand, one can also model the speed of vehicle $L$ as a linear autonomous system with state $x \in \mathbb{R}^{n_x}$, output $y \in \mathbb{R}^{n_y}$, and the minimum representation

$$\begin{aligned} \Sigma' : x(k+1) &= Ax(k), \\ y(k) &= Cx(k). \end{aligned} \tag{20}$$

The corresponding trajectory space is defined as

$$\mathcal{Y}_K(\Sigma') = \{y(0:K-1)^\top |\ y \text{ is a trajectory of (20)}\}. \tag{21}$$

Similar to Fundamental Lemma 1, the following lemma gives a data-driven representation of the trajectory space $\mathcal{Y}_K(\Sigma')$.

**Lemma 2.** *Let* (20) *be the minimum representation of the LTI autonomous system* $\Sigma'$. *Assume that the initial state* $x(0) \neq 0$ *is not an eigenvector or generalized eigenvector of matrix* $A$.

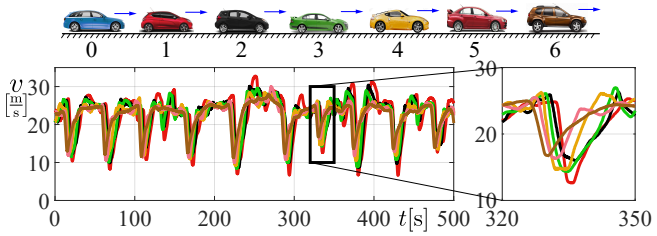


Fig. 4. Speed trajectories of human-driven vehicles 1-6 used in the paper.

*Given a trajectory of* (20) $\{y^\mathrm{d}(i)\}_{i=0}^{T-1}$, *with* $K \geq n_x$ *and* $T \geq K + n_x - 1$, *the trajectories in space* $\mathcal{Y}_K(\Sigma')$ *can be represented by*

$$\mathcal{Y}_K(\Sigma') = \operatorname{range}\left(\mathcal{H}_K(y^\mathrm{d})\right). \tag{22}$$

The proof is shown in the Appendix. Similar to (18), the future trajectory of $x_L$ can be predicted with quadratic programming

$$\min_{g,e} \ \|e\|_2^2 + \lambda_g \|g\|_2^2, \tag{23}$$

$$\text{s.t.} \ \mathcal{H}_K(x_L(0:k))g = \begin{bmatrix} x_L(k-K_\mathrm{p}+1:k) \\ x_L(k+1:k+K_\mathrm{f}) \end{bmatrix} + \begin{bmatrix} e \\ 0 \end{bmatrix}.$$

## IV. Simulations

In this section, we evaluate the proposed method with real driving data. We compare the prediction accuracy of the model-based method proposed in [13] and the data-driven method proposed in this paper. In addition, we compare the energy consumption of the corresponding PCCC designs.

We use data from a chain of 6 human-driven vehicles traveling on a highway [27], see the speed trajectories in Fig. 4. We place the CAV in the simulation in the position of vehicle 0, and assume that this does not affect the behavior of preceding vehicles. The ego vehicle follows vehicle 1 immediately in front, while also being connected to the distant vehicle 6 through V2V communication. The distant vehicle 6 frequently brakes and accelerates. These speed perturbations travel backward along the vehicle chain and, due to the string instability of human drivers, they grow when they reach vehicle 1.

Here we compare the prediction accuracy of the model-based method and the data-driven method. The detailed description of the model-based method was described in [13], and was briefly described in section III-A. Since we use intelligent driver model (IDM) [28] for prediction, we refer to it as *IDM* in the following text. We also use two variants of the data-driven method proposed in section III-C. In the first case, we assume that the speed of the distant vehicle 6 remains constant (19). We refer to this method as *Hankel const*. In the second case, we model $v_L$ as autoregressive process and apply the optimization (23). We refer to this method as *Hankel AR*.

As is shown in Fig. 5(a), when the CAV detects that the distant vehicle starts to brake, all methods predict that vehicle 1 will decelerate. In panel (b) the remote vehicle 6 starts to accelerate and the model-based method (incorrectly) predicts that the vehicle 1 accelerates immediately, while the

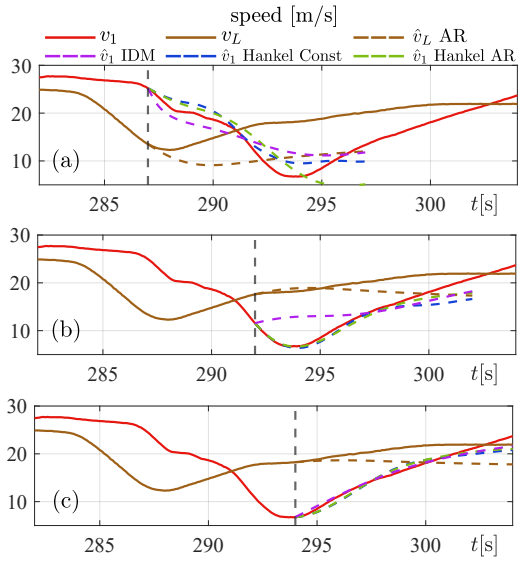


Fig. 5. Predictions of $v_1$ and $v_L$ at three different occasions.

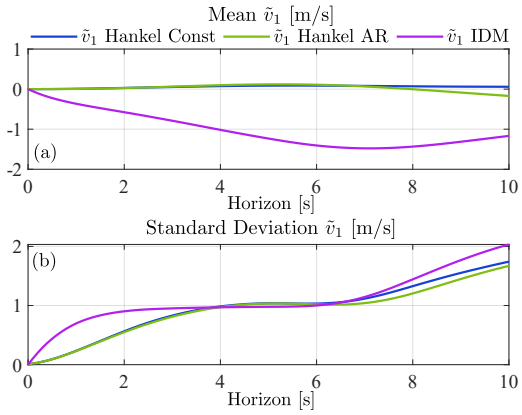


Fig. 6. The statistics of prediction error $\tilde{v}_1 = \hat{v}_1 - v_1$. (a) Mean of $\tilde{v}_1$. (b) Standard deviation of $\tilde{v}_1$.

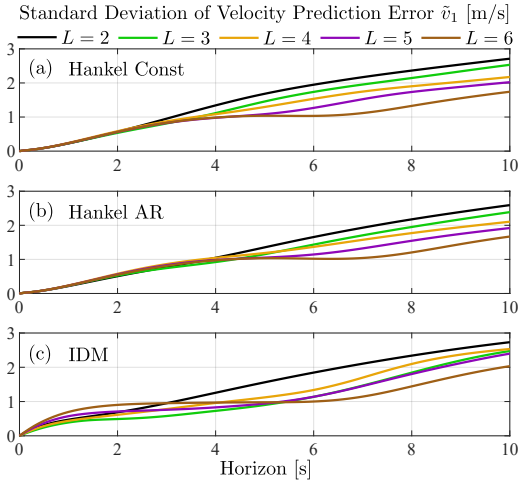


Fig. 7. Prediction error when using leading vehicles $L = 2, 3, 4, 5, 6$.

data driven methods capture that vehicle 1 first decelerates and then accelerates. Finally, in panel (c) all methods provide good predictions about the future behavior of vehicle 1.

Statistically, we can analyze the prediction errors in the 10 [s] prediction horizon. We compare the prediction of the speed of vehicle $\hat{v}_1$ with the observed data $v_1$. We compare all the simulation instances in 6 datasets. The prediction error $\tilde{v}_1 = \hat{v}_1 - v_1$ is shown in Fig. 6, all three methods can have accurate prediction for the near future: the mean of prediction error is close to 0 and the standard deviation is small. However, the model-based method suffers from great underestimation as the prediction horizon grows. In addition, the standard deviation of the prediction error using model-based method grows much faster compared to those using data-driven methods. For large horizon prediction, the mean of the prediction error is still near zero, but the prediction error shows a larger variance. One may observe a plateau in Fig. 6(b) up to about 6 seconds. This is a result of the perturbations imposed by the distant vehicle 6 takes about 6 seconds to reach vehicle 1. To further validate this, we evaluate the prediction error while using different remote vehicles, ranging from $L = 2$ to $L = 6$ in Fig. 7. Notice that as the vehicle becomes more distant, the plateau becomes more prolonged.

Due to the improvement of prediction accuracy, the data-driven prediction method results in better energy efficiency. The energy consumption per unit mass is defined as

$$w = \int_0^T v(t) g\big(\dot{v}(t) + f(v(t))\big) \mathrm{d}t, \tag{24}$$

where $g(x) = \max\{x, 0\}$ and $f$ is defined in (3). We simulate with model-based and data-driven prediction methods for 6 datasets. Due to better prediction accuracy, the data-driven method results in less braking and smoother speed trajectory. The resulting energy consumption show that data-driven prediction methods consume less energy compared to model-based method. On average, Hankel const method saves 8.5% energy compared to IDM method, and Hankel AR method saves 9.6% energy compared to IDM method.

## V. Conclusion

In this study, we introduced a data-driven predictive connected cruise controller for connected automated vehicles operating in mixed traffic environments comprising both connected and non-connected vehicles. The proposed controller utilized state information of remote vehicles beyond its line of sight. We captured the input-output relationship between the speed of the distant vehicle and the vehicle immediately in front utilizing the behavioral theory of linear systems. A model predictive controller, which considers safety and energy efficiency, was designed to utilize the prediction. The proposed method was evaluated using real human driver data and compared to model-based prediction methods. The results indicate that the data-driven method has lower variance compared to the model-based method. Additionally, the standard deviation of prediction error for different leading vehicles highlights the benefits of connection to distant vehicles beyond line of sight. This improvement in prediction leads to around 10% improvement in energy efficiency compared to a model-based predictor.

However, it should be noted that the data-driven prediction method is computationally intensive as it conducts system identification and prediction simultaneously, and requires a significant amount of data. Additionally, the entire theory is based on the behavioral theory of exact linear time-invariant systems, which may not accurately represent real-world traffic conditions. Future research will focus on increasing data-efficiency and incorporating uncertainties into analysis.

## APPENDIX

## PROOF OF LEMMA 2

*Proof.* For system (20) define the operator

$$\mathcal{O}_k(A, C) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^k \end{bmatrix}. \tag{25}$$

Without ambiguity, we simply write $\mathcal{O}_k$. Since (20) is the minimum representation, $(A, C)$ is observable and $A$ has full rank. Thus $\operatorname{rank}(\mathcal{O}_{n_x-1}) = n_x$. Since $K \geq n_x$, we have

$$\mathcal{Y}_K(\Sigma') = \operatorname{range}(\mathcal{O}_K) = n_x. \tag{26}$$

On the other hand, since $T > K$, we have

$$\operatorname{range}\left(\mathcal{H}_K(y^{\mathrm{d}})\right) = \left\{ \sum_{i=0}^{T-K} g_i \mathcal{O}_K A^i x(0) \,\middle|\, g_i \in \mathbb{R} \right\}. \tag{27}$$

Assume that $x(0)$ is not an eigenvector or generalized eigenvector of $A$, the vectors $x(0), Ax(0), \ldots, A^{n_x-1}x(0)$ are linearly independent. When $T - K \geq n_x - 1$,

$$\operatorname{range}\{x(0), Ax(0), \ldots, A^{T-K}x(0)\} = \mathbb{R}^{n_x}. \tag{28}$$

Therefore

$$\operatorname{range}\left(\mathcal{H}_K(y^{\mathrm{d}})\right) = \operatorname{range}(\mathcal{O}_K) = \mathcal{Y}_K(\Sigma'). \tag{29}$$

■

## REFERENCES


[1] G. Orosz, J. I. Ge, C. R. He, S. S. Avedisov, W. B. Qin, and L. Zhang, "Seeing beyond the line of site - controlling connected automated vehicles," *Mechanical Engineering*, vol. 139, pp. S8–S12, 12 2017.

[2] J. Guanetti, Y. Kim, and F. Borrelli, "Control of connected and automated vehicles: State of the art and future challenges," *Annual Reviews in Control*, vol. 45, pp. 18–40, 2018.

[3] T. Ersal, I. Kolmanovsky, N. Masoud, N. Ozay, J. Scruggs, R. Vasudevan, and G. Orosz, "Connected and automated road vehicles: state of the art and future challenges," *Vehicle System Dynamics*, vol. 58, no. 5, pp. 672–704, 2020.

[4] S. E. Shladover, C. Nowakowski, X.-Y. Lu, and R. Ferlis, "Cooperative adaptive cruise control: Definitions and operating concepts," *Transportation Research Record*, vol. 2489, no. 1, pp. 145–152, 2015.

[5] V. Milanés, S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe, and M. Nakamura, "Cooperative adaptive cruise control in real traffic situations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 296–305, 2014.

[6] Z. Wang, G. Wu, and M. J. Barth, "A review on cooperative adaptive cruise control (CACC) systems: Architectures, controls, and applications," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2884–2891, 2018.

[7] H. Liu, S. E. Shladover, X.-Y. Lu, and X. D. Kan, "Freeway vehicle fuel efficiency improvement via cooperative adaptive cruise control," *Journal of Intelligent Transportation Systems*, vol. 25, no. 6, pp. 574–586, 2021.

[8] Y. Zheng, S. E. Li, K. Li, F. Borrelli, and J. K. Hedrick, "Distributed model predictive control for heterogeneous vehicle platoons under unidirectional topologies," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 899–910, 2017.

[9] V. Turri, B. Besselink, and K. H. Johansson, "Cooperative look-ahead control for fuel-efficient and safe heavy-duty vehicle platooning," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 1, pp. 12–28, 2017.

[10] S. Feng, Y. Zhang, S. E. Li, Z. Cao, H. X. Liu, and L. Li, "String stability for vehicular platoon control: Definitions and analysis methods," *Annual Reviews in Control*, vol. 47, pp. 81–97, 2019.

[11] C. R. He, J. I. Ge, and G. Orosz, "Fuel efficient connected cruise control for heavy-duty trucks in real traffic," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 6, pp. 2474–2481, 2020.

[12] M. Shen, C. R. He, T. G. Molnar, A. H. Bell, and G. Orosz, "Energy-efficient connected cruise control with lean penetration of connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 4320–4332, 2023.

[13] M. Shen, R. A. Dollar, T. G. Molnar, C. R. He, A. Vahidi, and G. Orosz, "Energy-efficient reactive and predictive connected cruise control," *arXiv preprint arXiv:2210.04397*, 2022.

[14] L. Zhang and G. Orosz, "Beyond-line-of-sight identification by using vehicle-to-vehicle communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 6, pp. 1962–1972, 2018.

[15] R. A. Dollar, T. G. Molnár, A. Vahidi, and G. Orosz, "MPC-based connected cruise control with multiple human predecessors," in *American Control Conference (ACC)*, pp. 405–411, 2021.

[16] T. G. Molnár, X. A. Ji, S. Oh, D. Takács, M. Hopka, D. Upadhyay, M. V. Nieuwstadt, and G. Orosz, "On-board traffic prediction for connected vehicles with kalman filter," in *American Control Conference*, pp. 1036–1041, IEEE, 2022.

[17] H. Zhou, A. Zhou, T. Li, D. Chen, S. Peeta, and J. Laval, "Congestion-mitigating mpc design for adaptive cruise control based on newell's car following model: History outperforms prediction," *Transportation Research Part C*, vol. 142, p. 103801, 2022.

[18] S. Wong, L. Jiang, R. Walters, T. G. Molnár, G. Orosz, and R. Yu, "Traffic forecasting using vehicle-to-vehicle communication," in *3rd Conference on Learning for Dynamics and Control*, vol. 144 of *Proceedings of Machine Learning Research*, pp. 917–929, PMLR, 2021.

[19] I. Markovsky and F. Dörfler, "Behavioral systems theory in data-driven analysis, signal processing, and control," *Annual Reviews in Control*, vol. 52, pp. 42–64, 2021.

[20] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the deepc," in *European Control Conference (ECC)*, pp. 307–312, 2019.

[21] E. Elokda, J. Coulson, P. N. Beuchat, J. Lygeros, and F. Dörfler, "Data-enabled predictive control for quadcopters," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 18, pp. 8916–8936, 2021.

[22] J. Wang, Y. Zheng, Q. Xu, and K. Li, "Data-driven predictive control for connected and autonomous vehicles in mixed traffic," in *American Control Conference (ACC)*, pp. 4739–4745, 2022.

[23] J. Wang, Y. Zheng, K. Li, and Q. Xu, "DeeP-LCC: Data-enabled predictive leading cruise control in mixed traffic flow," *arXiv preprint arXiv:2203.10639*, 2022.

[24] T. Ard, R. A. Dollar, A. Vahidi, Y. Zhang, and D. Karbowski, "Microsimulation of energy and flow effects from optimal automated driving in mixed traffic," *Transportation Research Part C*, vol. 120, p. 102806, 2020.

[25] H. J. van Waarde, C. De Persis, M. K. Camlibel, and P. Tesi, "Willems' fundamental lemma for state-space systems and its extension to multiple datasets," *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 602–607, 2020.

[26] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. De Moor, "A note on persistency of excitation," *Systems & Control Letters*, vol. 54, no. 4, pp. 325–329, 2005.

[27] J. I. Ge, S. S. Avedisov, C. R. He, W. B. Qin, M. Sadeghpour, and G. Orosz, "Experimental validation of connected automated vehicle design among human-driven vehicles," *Transportation Research Part C*, vol. 91, pp. 335–352, 2018.

[28] A. Kesting, M. Treiber, and D. Helbing, "Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity," *Philosophical Transactions of the Royal Society A*, vol. 368, pp. 4585–4605, 2010.