
Article:
Vicente, B.A.H., James, S.S. and Anderson, S.R. (2021) Linear system identification versus physical modeling of lateral-longitudinal vehicle dynamics. IEEE Transactions on Control Systems Technology, 29 (3). pp. 1380-1387. ISSN: 1063-6536

https://doi.org/10.1109/tcst.2020.2994120

Linear system identification versus physical modelling of lateral-longitudinal vehicle dynamics

Bernardo A. Hernandez Vicente, Sebastian S. James and Sean R. Anderson

***Abstract*—Accurate physical modelling of vehicle dynamics requires extensive a-priori knowledge of the studied vehicle. In contrast, data-driven modelling approaches require only a set of data that is a good account of the vehicle's driving envelope. In this paper we compare, for the first time, the prediction capabilities of both approaches applied to a large-scale real world driving data set. The data set contains several cornering manoeuvres, acceleration and deceleration stages and was collected over public roads. Linear and nonlinear physical models were identified through nonlinear optimisation of their unknown parameters. Closed-form subspace identification methods were used to initialise the estimate of a linear state space model, and the initialisation was then refined through nonlinear optimisation. The optimised models were validated against 59 kilometres of independent driving data. The model fits, in the longitudinal velocity were 68.9% versus 80.2% for the nonlinear physical model and linear data driven (second order) model respectively, and in the yaw rate were 43.0% versus 63.5%. These results show that, for this vehicle, a simple linear data-driven model outperformed both linear and nonlinear physical models under real world driving conditions. This has important implications for control design approaches in autonomous vehicles.**

***Index Terms*—Nonlinear parameter estimation, system identification, vehicle dynamics, subspace identification, state space modelling.**

I. Introduction

The last two decades have seen a steep increase in research efforts to deploy (semi) autonomous ground vehicles both for private and commercial use [1]–[3]. An important part of these efforts has focused on devising mathematical models of vehicle dynamics for the purpose of control design [4]–[8]. Dynamic models for control design are often permitted to be relatively simple approximations because feedback control methods are highly tolerant of model error – a key advantage of using feedback. We suggest here that commonly used nonlinear physically-derived vehicle models are overcomplicated, difficult to tune appropriately, and unnecessary for the task of control design, and that simple linear models can capture the same basic dynamic behaviour, to a similar level of accuracy. This would be highly advantageous because linear control design methods could be confidently used in place of nonlinear methods, with key potential benefits in simplicity, robustness, stability guarantees and ease of implementation.

Physical models are usually a combination of Newton's second law and empirical understanding of certain driving conditions, which makes them easy to understand. Most physical models are nonlinear, due to the complexity of the process they attempt to simulate [9]. Linear physical models are assumed, in certain cases, to be accurate enough [10], and their validity on the extremes of the driving envelope can be boosted by allowing some of their parameters to vary in time [6], [11].

Given that even simplified physical models usually contain a high number of parameters, most approaches assume only a subset of them is unknown. In [12]–[15], for example, observer-based techniques are employed to estimate tyre forces and road friction coefficients, while in [16] related parameters are estimated through linearisation and least squares, and in [17] a particle-filtering approach is employed to obtain a Bayesian estimate of linear cornering stiffness values. In [18], [19], the focus is placed on identifying inertial parameters such as CoG position, while friction coefficients and suspension stiffness are assumed known. In the same context, usually only one main dynamic is studied, either longitudinal such as in [20]–[22] or lateral as in [6], [7], [23]–[25].

Whether linear or nonlinear, physically-derived models suffer from several shortcomings. They can only simulate the dynamics that they were derived for, hence only well defined experimental data is useful for the purpose of parameter estimation. It is usually complex to estimate appropriate value ranges for the model's unknown parameters without considerable a-priori knowledge of vehicle characteristics and in view of the bias introduced by modelling simplifications. Moreover, even linear models are usually nonlinear in the parameters, resulting in computationally expensive and time consuming estimation procedures.

Data-driven modelling techniques, on the other hand, result in models that best represent the entire set of available data. This allows for more flexibility in experiment design and/or data preprocessing. Moreover, the parameters of data-driven models do not necessarily represent any true physical quantity, hence rendering a-priori estimates unnecessary. Nevertheless, data-driven approaches are less abundant in the literature, including linear parameter varying methods to map steering angle to yaw rate and lateral acceleration [11], [26], linear transfer function and iterative learning methods with constant velocity assumptions [27], and subspace methods for longitudinal dynamics only [28], [29] and lateral dynamics only [30], [31].

It is clear that although several different modelling techniques have been employed to obtain physical and data-driven models of a vehicle's dynamics, much less attention has been given to their comparison. The focus of this paper, therefore, is to compare the prediction capabilities of nonlinear physical and data-driven models in both the velocity and displacement spaces. The models are identified and evaluated on a large-

Bernardo A. Hernandez Vicente, Sebastian S. James and Sean R. Anderson bahernandezv@gmail.com,{seb.james,s.anderson}@sheffield.ac.uk are with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, S1 3JD, UK.

scale real-world driving data set.

The rest of the paper is organized as follows. In Section II modelling assumptions are put forward and the selected model structures introduced. The identification methods are presented in Section III and Section IV briefly summarizes the data employed. Section V presents and compares the optimised models and their prediction capabilities. Finally, conclusions are drawn in Section VI.

II. Vehicle dynamics

A. Nonlinear physical modelling

In order to represent the nonlinear dynamics of the vehicle, a coupled (longitudinal-lateral) four-wheel model is used (Figure 1). We make a small steering angle assumption but only for the longitudinal dynamics. In practice this means that we decouple the longitudinal dynamics by neglecting the effect of the lateral forces on the longitudinal force equation, nevertheless we do observe the effect of the longitudinal forces on the lateral dynamics. We do this to reduce the complexity of the parameter estimation procedure, yet retaining some of the important coupling effects.

The variables l_f and l_r are the longitudinal distances between the vehicle's centre of gravity (CoG) and its front and rear axles respectively, while W is the axle track. The velocity of the CoG is described in the vehicle's body frame by v_x and v_y; the vehicle is also subject to a yaw rate $\dot{\psi}$ with corresponding yaw angle ψ. The steering angle of the front axle, δ, is assumed equal for both wheels. Each tyre is subject to longitudinal $F_{lon,ij}$ and lateral forces $F_{lat,ij}$, where $i = f(front), r(rear)$ and $j = l(left), r(right)$, and the vehicle's CoG is assumed to be subject to a dissipative force F_d. If the mass of the car is m and its yaw moment of inertia is I_z, then the vehicle's dynamics are governed by

$$m\left(\dot{v}_x - v_y\dot{\psi}\right) = F_{lon,fl} + F_{lon,fr} + F_{lon,rl} + F_{lon,rr} - F_d \tag{1a}$$

$$m\left(\dot{v}_y + v_x\dot{\psi}\right) = F_{y,fl} + F_{y,fr} + F_{y,rl} + F_{y,rr} \tag{1b}$$

$$\begin{aligned} I_z\ddot{\psi} = {} & \left(F_{y,fl} + F_{y,fr}\right) l_f - \left(F_{y,rl} + F_{y,rr}\right) l_r \\ & + \left(F_{x,fr} - F_{x,fl} + F_{x,rr} + F_{x,rl}\right)) W/2. \end{aligned} \tag{1c}$$

The geometric relation between the longitudinal/lateral forces and $F_{(x,y),ij}$ follows from Figure 1 and is omitted for brevity.

For simplicity, it is assumed throughout this paper that the longitudinal forces affecting the tyres are produced exclusively by the engine torque (F_e) and the brakes (F_b). This is due to the availability of such measurements and previous identification experiments performed on the data set [28]. The vehicle under study is front-wheel-drive, hence the engine torque is equally distributed amongst the two front wheels, whilst the brake force is assumed equally distributed among all tyres. The lateral forces are considered a function (to be defined) of the corresponding tyre side-slip angle, while tyre aligning torques [10] are neglected. Finally, the dissipative force F_d is assumed to be composed of aerodynamic drag, rolling friction and the mass component due to road slope (γ).

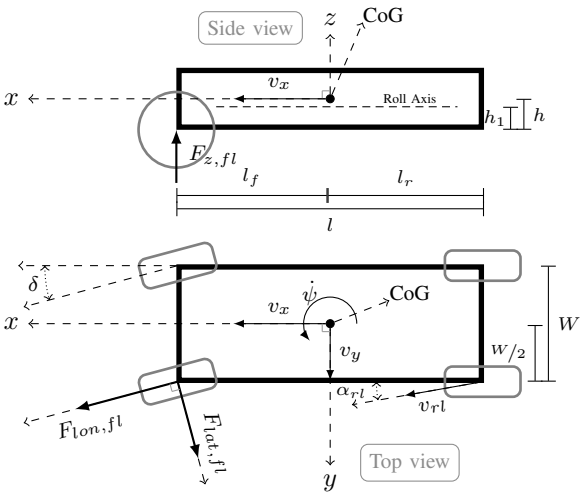

Figure 1. Diagram of the four-wheel car model.

1) Longitudinal dynamics: We employ the longitudinal model described in [28]. The force transmitted from the engine to the wheels is modelled by $F_e = \kappa_e T_e$, where T_e is the engine's delivered torque post gear-box and κ_e groups engine and transmission efficiencies and the driveline gear ratio. The brake force is defined in terms of the brake system pressure P_b by $F_b = \kappa_b P_b$. The dissipative force due to the road's slope is $F_s = mg\gamma$ where g is the gravitational constant and γ is the slope angle (assumed small). The friction force neglects road slope and hence is defined by $F_f = mg\mu$ where μ is the road-wheel static friction coefficient (assumed constant). Finally, the air drag force is modelled by a standard quadratic equation $F_D = \kappa_D v_x^2$. In this model κ_e, κ_D and μ are unknown.

2) Tyre lateral force model: The lateral force acting on the tyres depends on several factors including road condition, fabrication material, inflation pressure, etc. [4], [14]. If those parameters are fixed, however, the current magnitude of the lateral load varies only owing to the vertical load on the tyre and the tyre's side-slip angle. The latter is the angle formed between the the tyre's longitudinal symmetry plane and its velocity vector. Considering the CoG's velocity and the vehicle's yaw rate, the side-slip angle of each tyre can be easily defined

$$\alpha_{fl(r)} = \arctan\left(\frac{v_y + \dot{\psi} l_f}{v_x - (+)\dot{\psi} W/2}\right) - \delta \tag{2a}$$

$$\alpha_{rl(r)} = \arctan\left(\frac{v_y - \dot{\psi} l_r}{v_x - (+)\dot{\psi} W/2}\right). \tag{2b}$$

There exist several models that relate the side-slip angle to the tyre's lateral force [9], [10], [32]–[34]. Among them, a popular semi-empirical one is the so called Magic Formula [10]. This can be cast with several degrees of complexity by increasing the number of its parameters. It is, however, difficult to provide an a-priori estimate to most of these parameters without in-depth knowledge of the tyres and the road; henceforth we assume all tyres are identical and employ

a simplified version to avoid overfitting

$$F_{lat,ij} = F_{z,ij}\mu_y \sin\left(C \arctan\left(B\alpha_{ij}\right)\right), \tag{3}$$

where μ_y, C and B are unknown parameters.

3) Tyre vertical force model: In uniform-motion (zero acceleration), the vertical forces acting on each wheel depend only on the mass distribution of the vehicle; in accelerated motion, however, they also depend on the current acceleration experienced by the CoG. In this study we compute the vertical force acting over each wheel with a zero-order model assuming the front and rear axles are mechanically decoupled.

$$\begin{aligned} F_{z,fl(r)} &= {l_r mg}/{2l} - {hma_x}/{2l} \\ &\quad - (+){l_r h_1 ma_y}/{lW} + (-)\left[\bar{K}_{\phi,r}{h_2 ma_y}/{W}\right] \end{aligned} \tag{4a}$$

$$\begin{aligned} F_{z,rl(r)} &= {l_f mg}/{2l} + {hma_x}/{2l} \\ &\quad - (+){l_f h_1 ma_y}/{lW} + (-)\left[\bar{K}_{\phi,f}{h_2 ma_y}/{W}\right] \end{aligned} \tag{4b}$$

where the variables $\bar{K}_{\phi,f}$ and $\bar{K}_{\phi,r}$ represent the ratio of the axle stiffness, and they are unknown. We cast the model in terms of the ratios rather than the absolute values (also unknown) in order to improve the model's identifiability.

B. Linear physical modelling

A linear model of the vehicle's dynamic can be obtained from (1) through several simplification/assumption steps. First we assume small steering and wheel side-slip angles, and also $v_x \gg \psi^W/2$, which allows a simplification of (2) to

$$\alpha_{fl(r)} = \frac{v_y + \dot{\psi} l_f}{v_x} - \delta, \quad \alpha_{rl(r)} = \frac{v_y - \dot{\psi} l_r}{v_x},$$

and also of the geometric relationship between the longitudinal/lateral forces and $F_{(x,y),ij}$.

The second assumption is related to the tyre lateral force model. Since the side-slip angles are assumed small, the lateral force remains in the linear range of the Magic Formula, hence (3) is simplified to

$$F_{lat,ij} = C_i \alpha_{ij},$$

where C_i is referred to as the cornering stiffness [10]. This parameter is unknown and allowed to be different for each axle but not for each wheel in the axle, neglecting some of the effects of load transfer due to lateral acceleration in the linear range. Finally, the drag force term in the longitudinal dynamics is replaced by a linear function $F_D = \bar{\kappa}_D v_x$.

These simplifications result in a model that, except for the side-slip angle dependency on v_x, uncouples longitudinal and lateral dynamics. The model can be written in the standard state space form $\dot{x} = Ax + Bu + d$ with state vector $x = [v_y\ \dot{\psi}\ v_x]^\top$, input vector $u = [T_e\delta\ \ P_b\delta\ \ \delta\ \ T_e\ \ P_b\ \ \gamma]^\top$ matrices

$$A = \begin{bmatrix} {2(C_f+C_r)}/{mv_x} & {2(C_f l_f + C_r l_r)}/{mv_x} - v_x & 0 \\ {2(C_f l_f + C_r l_r)}/{I_z v_x} & {2\left(C_f l_f^2 + C_r l_r^2\right)}/{I_z v_x} & 0 \\ 0 & 0 & {\bar{\kappa}_D}/{m} \end{bmatrix} \tag{5a}$$

$$B = \begin{bmatrix} {\kappa_e}/{m} & -{\kappa_b}/{2m} & -{2C_f}/{m} & 0 & 0 & 0 \\ {\kappa_e l_f}/{I_z} & -{\kappa_b l_f}/{2I_z} & -{l_f C_f}/{I_z} & 0 & 0 & 0 \\ 0 & 0 & 0 & {\kappa_e}/{m} & -{\kappa_b}/{m} & -g \end{bmatrix}, \tag{5b}$$

and disturbance $d = -g\mu$ representing the constant friction.

The model described by (5) is nonlinear, since all the lateral entries of the state transition matrix depend on the longitudinal velocity of the vehicle, v_x. To obtain a linear time-invariant (LTI) physical model for comparison to the LTI data-driven model we treat v_x in these matrix entries as an unknown constant and estimate it along with the other parameters.

In certain (extreme) driving conditions, the performance of the linear model can degrade due to the omission of coupling between lateral and longitudinal dynamics. Therefore, we also estimate coupled versions of the LTI physical model. To do so, we directly modify (5) by allowing some of its coupling parameters to be non-zero. In order to provide a fair comparison with the nonlinear physical model, we consider the same one-way coupling scenario in which lateral dynamics are affected by longitudinal dynamics but not the other way around. In practice, this means that the state-transition matrix is modified by allowing the elements $A_{1,3}$ and $A_{2,3}$ to be non-zero, where subindices indicate (row, column) position.

C. Data-driven model

For the data-driven model, we employ a subspace system identification technique [35] to estimate an LTI state space model that best explains the data. The model has the following structure

$$\dot{x_d} = A_d x_d + B_d u_d \tag{6a}$$

$$y_d = C_d x_d + v, \tag{6b}$$

where u_d is the input, v is measurement noise and y_d is the measured output. In the physical modelling case, the matrices A and B are defined by the physics of the system, while C is the identity of appropriate dimension. In the subspace approach, on the other hand, input and output are chosen by the user depending on their understanding of the system's dynamics; however the size and contents of the model matrices are chosen by the algorithm to best simulate the output of the system given the inputs applied. Hence the state vector, and its dimension n_d, do not necessarily have any physical meaning.

For the data-driven model we chose $y_d = [v_x\ \ \dot{\psi}]^\top$ and $u_d = [T_e\ \ P_b\ \ \gamma\ \ \delta]^\top$, in order to match the inputs used in the linear physical model (5).

III. System identification procedure

A. System identification of physically-derived models

The estimation of the unknown parameters in the physically-derived models was performed through a least squares approach. The overall procedure is as follows:

i. Estimate the parameters of the independent nonlinear longitudinal model κ_e, κ_D and μ (done in [28]).
ii. Estimate the parameters of the linear longitudinal model $\bar{\kappa}_D$.
iii. Estimate the parameters of the coupled nonlinear lateral model μ_y, C, B, $\bar{K}_{\phi,f}$ and $\bar{K}_{\phi,r}$.
iv. Estimate the parameters of the uncoupled linear lateral model C_f, C_r and $\bar{v}_x$.

v. Estimate the parameters of the one-way coupled linear lateral model C_f, C_r, $\bar{v}_x$, $A_{1,3}$ and $A_{2,3}$.

Note that step (ii) does not re-estimate the parameters κ_e and μ to avoid biases due to linearization. Also note that, although coupling is allowed in step (v), the longitudinal parameters are not re-estimated to provide a fair midpoint for comparison with the nonlinear model.

In all of the cases listed above, the procedure to obtain an estimate of the parameters is the same and similar to that in [5]. Define θ_0 as the vector containing the true model parameters. The corresponding estimate $\hat{\theta}$ is obtained by solving the following optimisation problem

$$\mathbb{P} = \min_{\theta \in \Theta} \boldsymbol{J}(\theta), \tag{7}$$

The set Θ in (7) is particular to each estimation exercise and is defined by a set of box constraints imposed over each parameter. This is done to guarantee that the obtained parameters have a physical meaning. The cost function is defined as

$$\boldsymbol{J}(\theta) = \sum_{k=1}^{N} \left(h(t_k) - \hat{h}(t_k, \theta)\right)^{\top} \Upsilon \left(h(t_k) - \hat{h}(t_k, \theta)\right), \tag{8}$$

where $h(t_k)$ is a vector of measured outputs at time t_k and $\hat{h}(t_k, \theta)$ is the vector of corresponding simulated values at the same time instant given a vector of candidate parameters θ. The matrix Υ is a diagonal matrix composed of weights used to normalize the different variables in the output vector h. The definition of h depends, again, on the particular estimation being carried out. In the longitudinal case, $h = v_x$ and in the uncoupled and one-way coupled lateral cases $h = \dot{\psi}$.

B. Data-driven system identification

The identification of the data-driven model was performed in two stages. First, a closed-form subspace identification method was employed to initialise state space models of different orders (different number of states). The reader is referred to [35] for a detailed description of such methods. In the second stage, the parameters of the state space models were refined through a single implementation of (7) with $h = y_d$ and θ a vector containing all the parameters in the model matrices A_d, B_d and C_d. Note that, since the initialisation is performed in closed-form, the procedure to obtain the data-driven models does not require multiple starts.

C. Parameter estimation

Since no particular model is attached to measurement noise, all the physically-derived models are nonlinear in the parameters. It follows that the minimisation of the least squares functional is a nonlinear optimisation problem. In this case, we used a constrained interior-point method implemented in the *fmincon* function in MATLAB to solve (7). To avoid local minima of bad quality, each identification experiment was repeated 1000 times with different starting values within the constraint set Θ, except for the data-driven models where the parameters are estimated in a single instance of optimisation. The optimal vector of parameters amongst the 1000 results was chosen as the one with lowest cost function (8). Despite the 1000 starts it is not possible to guarantee that the chosen parameters are indeed the true vehicle parameters, nor the global optimum to our optimisation problem. This is one of the difficulties in estimating the parameters of a physically-derived model when there is little a-priori knowledge on the vehicle's characteristics.

Finally, in both modelling approaches the focus of the identification procedure was placed on simulation. This means that $\hat{h}$ was generated by simulating the entire length of the studied data set from initial conditions, as opposed to a prediction error approach in which measurements are used to correct simulations at each time step.

D. Evaluation metrics

The models were evaluated using a normalised fit metric, F [35],

$$F = 100\left(1 - {}^{||h-\hat{h}||_2}/_{||h-\bar{h}||_2}\right), \tag{9}$$

where $\bar{h}$ is the mean of measurements. A value of $F = 100\%$ indicates a perfect fit, whilst a value of $F = 0\%$ indicates a fit equivalent to the mean of the output data, and the F value becomes negative for poor fits.

We also performed a correlation analysis between the measured outputs and the linear and nonlinear simulated ouputs. This is characterised by the variance accounted for metric (VAF)

$$V = 100\left(1 - {}^{\mathrm{var}\left(h-\hat{h}\right)}/_{\mathrm{var}(h)}\right). \tag{10}$$

In this case, a value of $V = 100\%$ indicates that the model is able to explain the entire variance of the measurements. This, of course, is not expected since the model cannot account for the variance introduced by measurement noise.

IV. Experimental data

The experimental data was gathered in a Lancia Delta car during 108 km of driving on public roads in the Piemonte region in Italy. The known vehicle parameters used in the physical modelling are depicted in Table II. The route was designed to include a typical selection of extra-urban and urban roads, motorways, roundabouts and intersections, but the driving itself was unscripted since it had to accommodate for real-time traffic conditions. Figure 2 shows the driving route. A single lap around the route was 54 km long and it was circumnavigated twice. The longitude/latitude data was collected by a GPS, and the position of the car, with respect to a fixed set of coordinates, is computed by transforming the longitude/latitude data into distance using the equirectangular approximation.

The entire set of data was divided into training and validation portions as shown by the red line in Figure 2, with 2501 s of data for training and 2383 s for validation. The data was originally sampled at 100 Hz, however the power spectrum of the relevant measurements showed that over 99% of the power was concentrated below 1 Hz. In view of this the data was resampled at 20 Hz in order to decrease the computational load of the parameter estimation algorithms. Figure 3 shows the measurements employed to drive the identification procedure

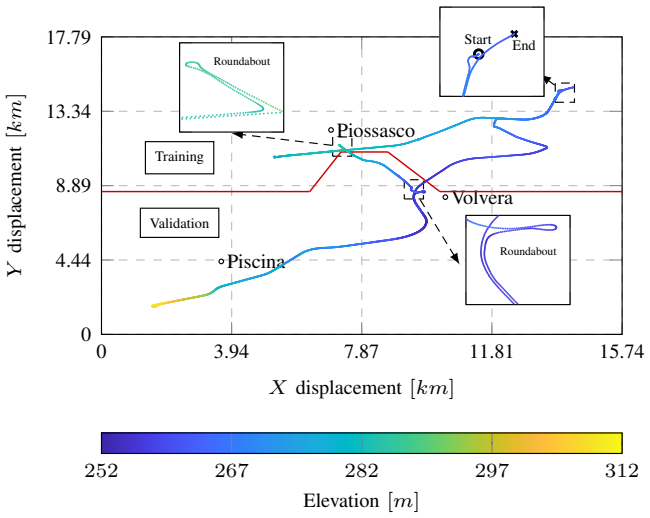

Figure 2. Driving route of the Lancia Delta - first lap: the red line indicates the split of the route between training data (upper portion) and validation data (lower portion).

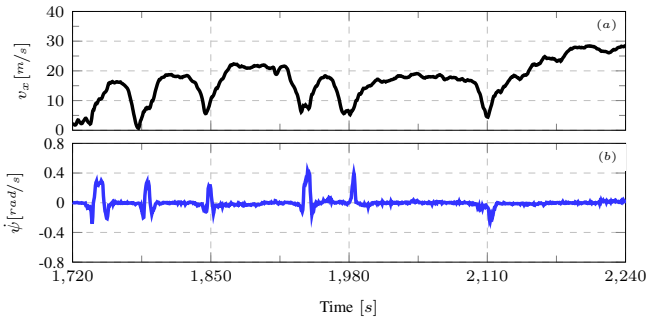

Figure 3. Output measurements acquired during the test route driving corresponding to 25% of the total training data set: (a) Longitudinal velocity, (b) Yaw rate.

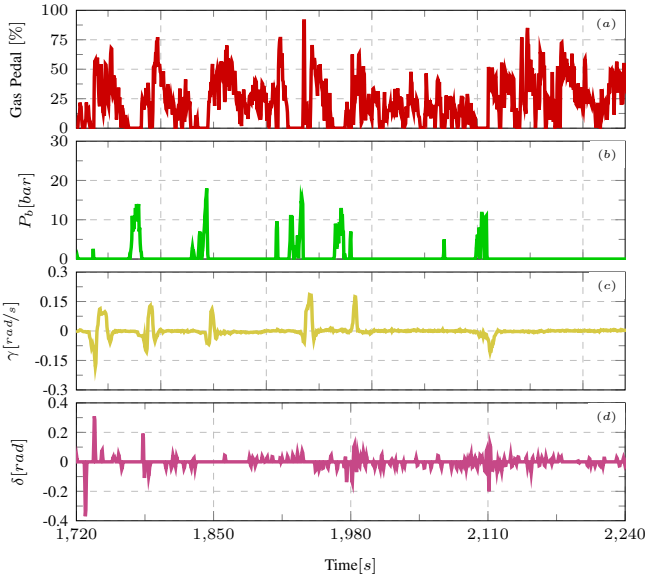

Figure 4. Input measurements acquired during the test route driving corresponding to 25% of the total training data set: (a) Gas pedal position, (b) Brake pressure, (c) Road slope angle, (d) Steering angle.

Table I
PERFORMANCE METRIC SUMMARY.

	Physical models			
	v_x (longitudinal)		$\dot{\psi}$ (lateral)	
	$F\%$	$V\%$	$F\%$	$V\%$
Nonlinear	68.9	93.7	43.0	66.7
Linear	3.4	66.2	62.5	86.1
	Data-driven models			
	v_x (longitudinal)		$\dot{\psi}$ (lateral)	
Order n	$F\%$	$V\%$	$F\%$	$V\%$
2	80.2	96.1	63.5	86.8
3	82.5	97.7	63.3	86.8
4	85.0	97.9	64.9	88.0

for a part of the training portion, while Figure 4 shows the corresponding inputs.[1]

V. Results and Discussion

For brevity we skip the training results. We place the focus on the validation data set to make sure that there is no overfitting in the data-driven models. Figures 5 to 9 show the validation data set alongside the simulation of the different estimated models, while Table I summarizes their performance metrics. Note that Figures 5 and 7 show only subsections of the entire validation data set for clarity. The values of the optimised parameters for the physically-derived and data-driven models can be found in the appendix in Tables III to V.

A. Longitudinal dynamics

Identification of physically-derived longitudinal models led to superior performance for the nonlinear physical model (fit 68.9%) when compared to the linearised physical model (fit 3.4%). Indeed, although the linear physical model is able to simulate the general trend in velocity, it overestimates it by 5-10 m/s. On the other hand, the simplest data-driven model ($n_d = 2$) outperforms both physically-derived models with a fit of 80.2%. Moreover, data-driven models of the longitudinal dynamics show, as expected, a general increase in performance as the order of the model grows, reaching 85.0% at $n_d = 4$ (see Table I).

A comparison between the validation data and the prediction of the several longitudinal models is shown in Figure 5 for one lap, whilst comparison over the full validation data is given in Figure 6 in the form of a correlation plot.

The data-driven model is also able to explain more of the measurement variances, with a variance accounted for that is 2.4% and 32% larger when compared to that of the nonlinear and linear physical models (respectively). Note that the nonlinear physical model presents a higher correlation in the high velocity range, however the data-driven model performs better throughout the entire driving envelope (see Figure 6). This can be explained by some of the modelling assumptions used in the physical nonlinear model such as the simplified drag and friction forces.

[1]For commercial confidentiality, we only report gas pedal and not engine torque

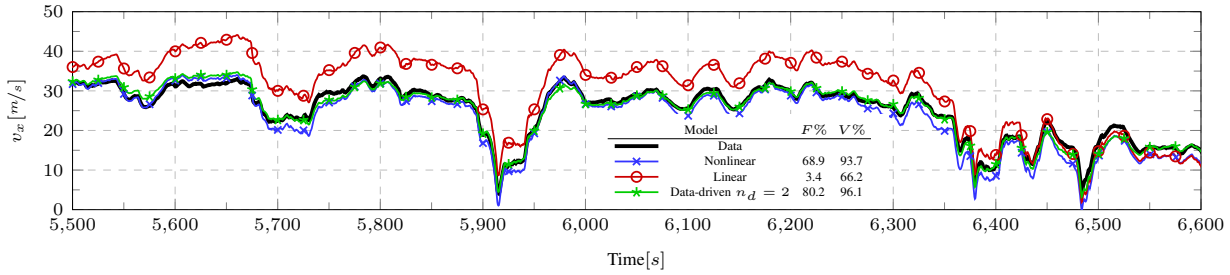

Figure 5. Validation in velocity v_x over half of the validation data (only the first tour around the validation section of the route is shown for clarity).

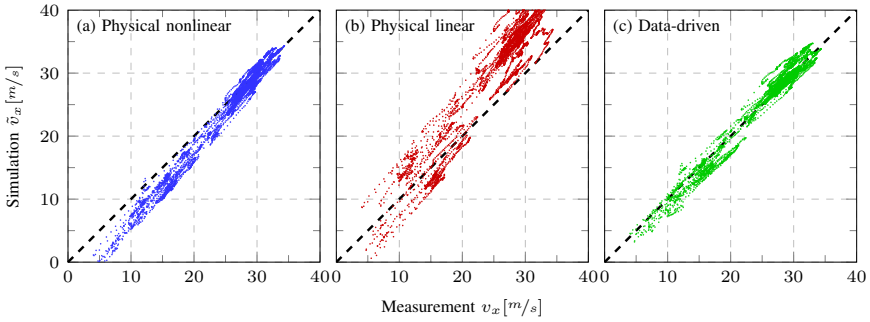

Figure 6. Correlation between measured and simulated velocity v_x over the entire validation data set.

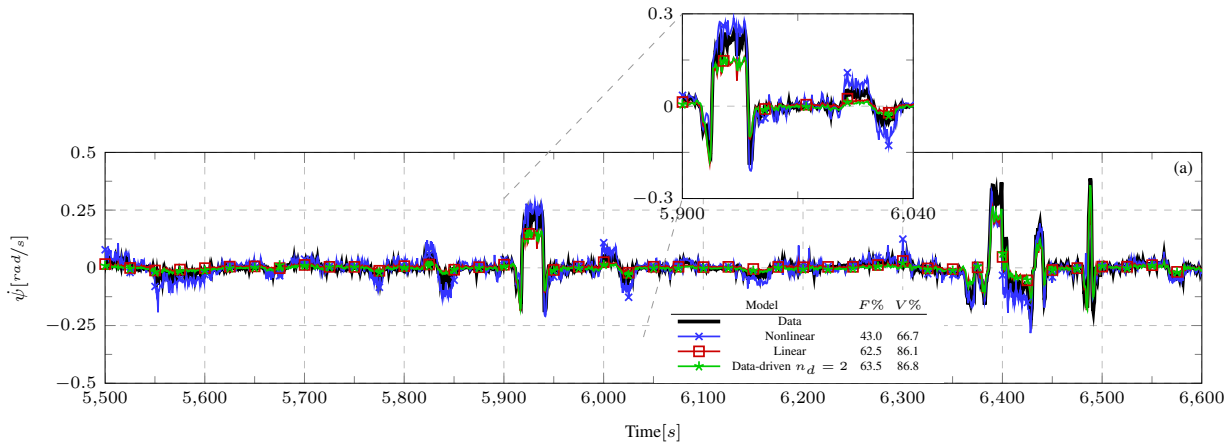

Figure 7. Validation in yaw rate $\dot{\psi}$ over half of the validation data (only the first tour around the validation section of the route is shown for clarity).

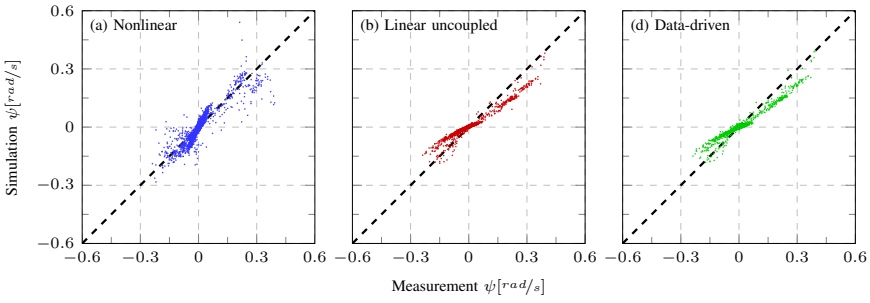

Figure 8. Correlation between measured and simulated yaw rate ψ over the entire validation data set.

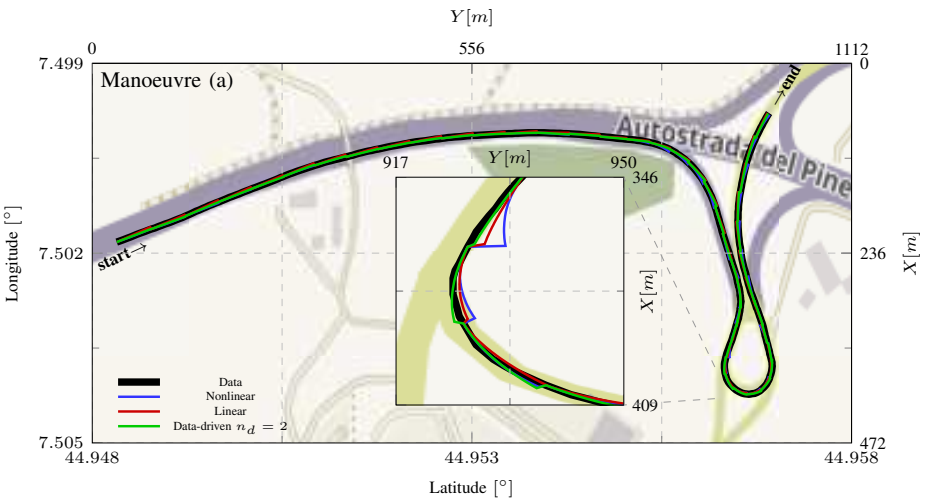

Figure 9. Model comparison in the displacement space with $T = 2$ s (prediction horizon of 40 samples) for an illustrative cornering manoeuvre.

B. Lateral dynamics

The estimation of physically-derived lateral models yielded an opposite result to that of the longitudinal case, with the linear physical model showing a considerably better performance (fit 62.5%) than the nonlinear physical model (fit 43.0%). The data-driven model, $n_d = 2$ (fit 63.5%) performed marginally better than the linear physical model. A comparison between the validation data and the prediction of the several lateral models is shown in Figure 7 for one lap.

The poor performance of the nonlinear physical model is attributed to its complexity. Indeed, although the nonlinear model is in principle a more accurate representation of the vehicle's lateral dynamics, finding the true values of its parameters without extensive *a-priori* knowledge of the vehicle's characteristics is a difficult computational task. Although we performed 1000 different optimisation attempts (with different parameter initialisations), there is no guarantee that the optimal values chosen are indeed an accurate approximation of the true physical values.

Both linear models (physically-derived and data-driven) present similar correlation to data throughout the driving envelope with variance accounted for metrics of 86.1% and 86.8% respectively. Comparison over the full validation data is given in Figure 6 in the form of a correlation plot. At low levels of yaw rate the linear models (physically-derived and data-driven) underestimate the yaw rate, while the nonlinear model overestimates it. This can be explained by the model structure. The nonlinear physical model simulates the lateral forces using the Magic formula, which given the optimised values of its parameters, presents a slope that is an order of magnitude higher than $C_{f,r}$ in the vicinity of the origin.

C. Performance in the displacement space

Finally, we studied the performance of the models in the displacement space and along prediction horizons usually employed by predictive controllers. The vehicle's displacement was simulated in a receding horizon fashion over a horizon of $T = 2$ s.

Figure 9 shows the results from the receding horizon simulation for a representative cornering manoeuvre within the validation data set. As expected given their fit metrics, the physically-derived models produced predictions with several instances in which the vehicle veered outside the driving lane, particularly during sharp turns. On the other hand, the data driven model was able to predict the vehicles displacement accurately enough to ensure the vehicle stayed inside the road. Note that this is true despite the fact that the data-driven model entirely neglects lateral velocity.

VI. Conclusions

In this paper we explored the potential of data-driven linear models for simulating longitudinal velocity and yaw rate of a ground 4-wheeled vehicle, and compared its performance with physical linear and nonlinear models. The results show that, under normal driving conditions, simple data-driven models are able to match and surpass the simulation capabilities of the physically-derived models, whilst being considerably simpler than the nonlinear physical model. In particular, a second-order state space model is able to predict velocity and yaw rate accurately enough such that, within the usual horizon lengths of model-based controllers, the predicted position of the vehicle stays close to its true position and remains inside the driving lane.

Acknowledgment

The authors would like to thank the EU for funding support through grant number 731593 (Dreams4Cars), as well as A. Saroldi and Centro Ricerche Fiat for their contribution in collecting the experimental data.

Appendix

Table II
Lancia Delta physical parameters.

Body dimensions		Roll axis		Inertia	
l_f	1.09 m	h	0.6 m	I_z	1260 kg/m^2
l_r	1.61 m	h_1	0.4 m	m	1550 kg
W	1.53 m	h_2	0.2 m		

Table III
PARAMETER ESTIMATES FOR THE PHYSICALLY-DERIVED LONGITUDINAL DYNAMIC MODELS.

	κ_e	μ	κ_D
Nonlinear	11.696	0.017	0.311
Linear	11.696	0.017	6.457

Table IV
PARAMETER ESTIMATES FOR THE PHYSICALLY-DERIVED LATERAL DYNAMIC MODELS.

	$\bar{K}_{\phi,f}$	$\bar{K}_{\phi,r}$	B	C	μ_y
Nonlinear	0.415	0.585	66.158	4.000	0.402
	C_f	C_r	$\tilde{v}_x$	A_{13}	A_{23}
Linear one-way coupled	-63719	-43321	5.450	1.074	-0.001

Table V
PARAMETER ESTIMATES FOR THE SECOND ORDER DATA-DRIVEN MODEL.

$$A = \begin{bmatrix} -401.072 & 328.759 \\ -481.589 & 394.758 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.038 & 1.323 & 63.921 & 2333.403 \\ 0.045 & 1.591 & 76.856 & 2801.894 \end{bmatrix}$$

$$C = \begin{bmatrix} 93.156 & -77.582 \\ 0.346 & -0.284 \end{bmatrix}$$

REFERENCES

[1] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A Survey of Motion Planning and Control Techniques for Self-Driving Urban Vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 527–538, 2016.

[2] M. Da Lio, F. Biral, E. Bertolazzi, M. Galvani, P. Bosetti, D. Windridge, A. Saroldi, and F. Tango, "Artificial co-drivers as a universal enabling technology for future intelligent vehicles and transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 244–263, 2015.

[3] R. Hussain and S. Zeadally, "Autonomous Cars: Research Results, Issues and Future Challenges," *IEEE Communications Surveys and Tutorials*, 2018.

[4] W. Sienel, "Estimation of the tire cornering stiffness and its application to active car steering," in *Proceedings of the 36th Conference on Decision and Control*, San Diego, CA, 1997, pp. 4744–4749.

[5] P. D. Filippi, M. Tanelli, M. Corno, S. M. Savaresi, and L. Fabbri, "Semi-Active Steering Damper Control in Two-Wheeled Vehicles," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 5, pp. 1003–1020, 2011.

[6] M. Akar and A. D. Dere, "A switching rollover controller coupled with closed-loop adaptive vehicle parameter identification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1579–1585, 2014.

[7] Y. F. Lian, Y. Zhao, L. L. Hu, and Y. T. Tian, "Cornering stiffness and sideslip angle estimation based on simplified lateral dynamic models for four-in-wheel-motor-driven electric vehicles with lateral tire force information," *International Journal of Automotive Technology*, vol. 16, no. 4, pp. 26–32, 2015.

[8] A. T. Nguyen, P. Chevrel, and F. Claveau, "LPV Static Output Feedback for Constrained Direct Tilt Control of Narrow Tilting Vehicles," *IEEE Transactions on Control Systems Technology*, pp. 1–10, 2018.

[9] M. Abe, *Vehicle Handling Dynamics: Theory and Application*, 1st ed. Oxford: Elsevier Ltd., 2009.

[10] H. B. Pacejka, *Tyre and Vehicle Dynamics*, 1st ed. Oxford: Elsevier Ltd., 2004.

[11] P. Gáspár, Z. Szabó, and J. Bokor, "A grey-box identification of an LPV vehicle model for observer-based side slip angle estimation," in *Proceedings of the 2007 American Control Conference*, 2007, pp. 2961–2966.

[12] S. Chakraborty, S. Sen, A. Sutradhar, and A. Sengupta, "Estimation of tire-road friction coefficient and frictional force for active vehicle safety system," in *Proceedings of the 2015 International Conferecen on Industrial Instrumentation and Control*. Pune: IEEE, 2015, pp. 674–679.

[13] M. Choi, J. J. Oh, and S. B. Choi, "Linearized Recursive Least Squares Methods for Real-Time Identification of Tire-Road Friction Coefficient," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 2906–2918, 2013.

[14] J.-o. Hahn, R. Rajamani, and L. Alexander, "GPS-Based Real-Time Identification of Tire Road Friction Coefficient," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 3, pp. 331–343, 2002.

[15] C. Ahn, H. Peng, and H. E. Tseng, "Robust estimation of road frictional coefficient," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 1, pp. 1–13, 2013.

[16] M. Tanelli, L. Piroddi, M. Piuri, and S. M. Savaresi, "Real-time identification of tire-road friction conditions," in *Proceedings of the 17th IEEE International Conference on Control Applications*, 2008, pp. 25–30.

[17] K. Berntorp and S. Di Cairano, "Tire-Stiffness and Vehicle-State Estimation Based on Noise-Adaptive Particle Filtering," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 3, pp. 1100–1114, 2018.

[18] S. Hong, C. Lee, F. Borrelli, and J. K. Hedrick, "A novel approach for vehicle inertial parameter identification using a dual kalman filter," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 151–161, 2015.

[19] B. L. Pence, H. K. Fathy, and J. L. Stein, "Recursive estimation for reduced-order state-space models using polynomial chaos theory applied to vehicle mass estimation," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 1, pp. 224–229, 2014.

[20] J. E. Alves Dias, G. A. Silva Pereira, and R. Martinez Palhares, "Longitudinal Model Identification and Velocity Control of an Autonomous Car," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 776–786, 2015.

[21] K. Li, J. Cao, and F. Yu, "Nonlinear tire-road friction control based on tire model parameter identification," *International Journal of Automotive Technology*, vol. 13, no. 7, pp. 1077–1088, 2012.

[22] R. Rajamani, G. Phanomchoeng, D. Piyabongkarn, and J. Y. Lew, "Algorithms for real-time estimation of individual wheel tire-road friction coefficients," *IEEE/ASME Transactions on Mechatronics*, vol. 17, no. 6, pp. 1183–1195, 2012.

[23] G. Reina and A. Messina, "Vehicle dynamics estimation via augmented Extended Kalman Filtering," *Measurement*, vol. 133, pp. 383–395, 2019.

[24] H. Abdellatif and B. Heimann, "Accurate modelling and identification of vehicle's nonlinear lateral dynamics," in *Proceedings of the 16th IFAC World Congress*, Prague, 2005, pp. 185–189.

[25] F. Cheli, E. Sabbioni, M. Pesce, and S. Melzi, "A methodology for vehicle sideslip angle identification: comparison with experimental data," *Vehicle System Dynamics*, vol. 45, no. 6, pp. 549–563, 2007.

[26] V. Cerone, D. Piga, and D. Regruto, "Set-membership LPV model identification of vehicle lateral dynamics," *Automatica*, vol. 47, no. 8, pp. 1794–1799, 2011.

[27] N. Liu and A. G. Alleyne, "Iterative learning identification applied to automated off-highway vehicle," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 1, pp. 331–337, 2014.

[28] S. James and S. R. Anderson, "Linear System Identification of Longitudinal Vehicle Dynamics Versus Nonlinear Physical Modelling," *2018 UKACC 12th International Conference on Control*, pp. 146–151, 2018.

[29] S. S. James, S. R. Anderson, and M. D. Lio, "Longitudinal vehicle dynamics: A comparison of physical and data-driven models under large-scale real-world driving conditions," *IEEE Access*, vol. 8, pp. 73 714–73 729, 2020.

[30] M. W. Choi, J. H. Ryu, H. S. Lee, K. S. Lee, and M. H. Lee, "Robust lateral controller design for an unmanned vehicle using a system identification method," *IEEE International Symposium on Industrial Electronics*, no. 6, pp. 1177–1182, 2008.

[31] S. M. Yoon, K. S. Lee, S. Y. Kim, J. H. Kang, and M. H. Lee, "Lateral Control of an UCT (Unmanned Container Transporter) Using Ultrasonic Satellite System and System Identification," in *2008 International Conference on Control, Automation and Systems*, vol. 0, no. 0. Seoul: IEEE, 2008, pp. 296–300.

[32] U. Kiencke and L. Nielsen, *Automotive Control Systems For Engine, Driveline, and Vehicle*, 2nd ed. Springer Berlin Heidelberg New York, 2005.

[33] C. Canudas-de Wit, P. Tsiotras, E. Velenis, M. Basset, and G. Gissinger, "Dynamic Friction Models for Road/Tire Longitudinal Interaction," *Vehicle System Dynamics*, vol. 39, no. 3, pp. 189–226, 2003.

[34] C. Canudas de Wit, H. Olsson, K. J. Aström, and P. Lischinsky, "A New Model for Control of Systems with Friction," *IEEE Transactions on automatic control*, vol. 40, no. 3, pp. 419–425, 1995.

[35] L. Ljung, *System Identification Theory for the user*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.