

## Open-source dataset for control-oriented modelling in diesel engines

Jinghua ZHAO<sup>1,2,3</sup>, Sitong ZHOU<sup>1,4</sup>, Yunfeng HU<sup>1,2\*</sup>, Mingjun JU<sup>4</sup>,  
Ruixue REN<sup>4</sup> & Hong CHEN<sup>1,2</sup>

<sup>1</sup>State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130025, China;

<sup>2</sup>Department of Control Science and Engineering, Jilin University, Changchun 130025, China;

<sup>3</sup>Computer College, Jilin Normal University, Siping 136000, China;

<sup>4</sup>Changchun FAW Sihuan Engine Manufacture Co., Ltd, Changchun 130000, China

Received 18 December 2018/Accepted 21 January 2019/Published online 21 March 2019

**Citation** Zhao J H, Zhou S T, Hu Y F, et al. Open-source dataset for control-oriented modelling in diesel engines. *Sci China Inf Sci*, 2019, 62(7): 077201, <https://doi.org/10.1007/s11432-018-9783-x>

In this study, an experimental dataset of a diesel engine under the European transient cycle (ETC) is demonstrated and opened for academic research. The dataset with 7 dimensional features is collected from the advanced diesel engine manufactured by Changchun FAW Sihuan Engine Manufacture Co., Ltd.

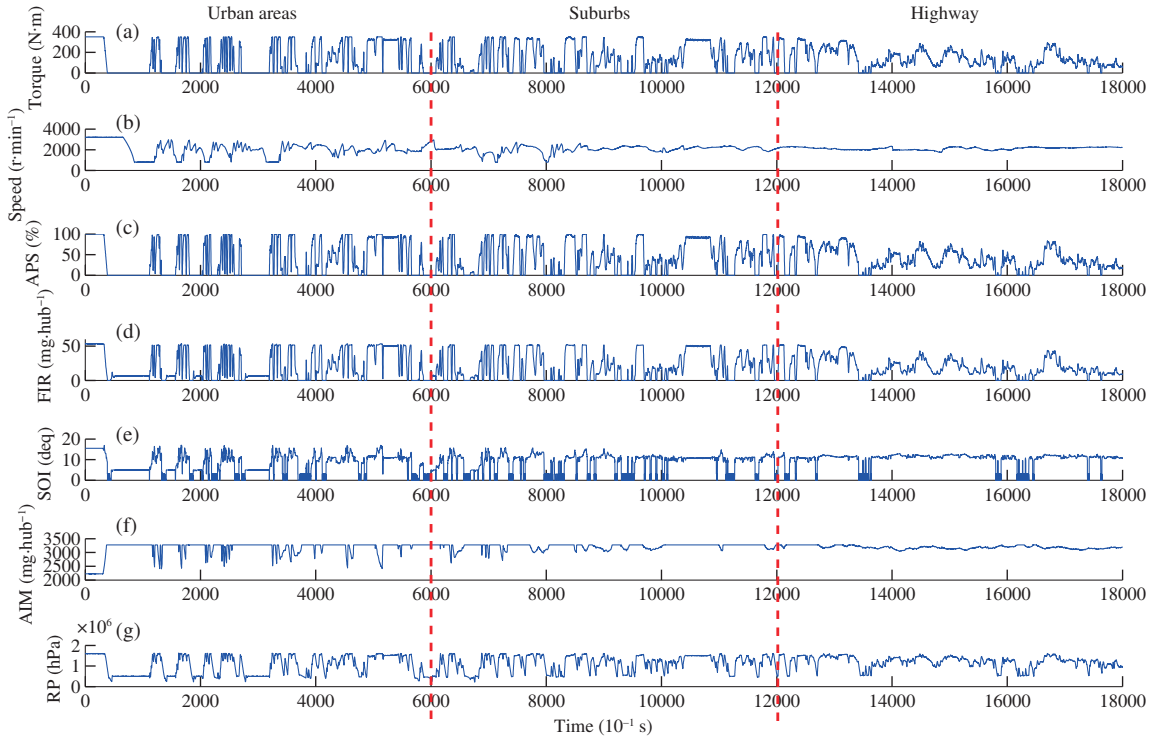
The experimental data of diesel engines is very important for system modeling and identification, and designing control system. Although some experimental datasets have been posted in published papers or websites, most of them are only collected under steady working conditions. On one hand, the collection of engine transient data is often limited in colleges owing to the cost of some sophisticated equipments such as transient electric dynamometer. On the other hand, the resource from the industry is also limited as the research and design departments of original equipment manufactures always keep the transient data in a confidential level and rarely disclose them.

In this study, the experimental data was gathered through test bench of a 2.7L Sihuan four-cylinder diesel engine via high-speed analog-to-digital acquisition system so that a number of fast changing signals can be measured in real time including engine speed, engine torque, fuel injection rate (FIR), and start of injection (SOI). Transient

engine load in ETC driving cycle can be realized by using the transient electric dynamometer which enables the data collection in wide range of transients. The duration of the dataset under the ETC cycle is 1800 s including three typical conditions, namely, urban areas, suburbs and highway, representing three types of low, medium and high speeds, respectively. The sampling period of the data is 0.1 s. The dataset is processed with 126000 sampling points and includes seven key engine parameters (each parameter has 18000 sampling points). The dataset stored with .mat formulation and can be directly read and post-processed in MATLAB software for the research use.

The test data is capable of supporting the data-driven modeling of diesel engine transients that improve the accuracy of look-up table-based and physics-based models. As the number of input variables increases, the complexity of the modelling process increases accordingly. Therefore, the original dataset may need to be down-sized according to different applications to balance the accuracy and complexity. Taking the engine torque modeling as an example, the modeled output is engine torque and the corresponding inputs are engine speed, acceleration pedal signal (APS), FIR, SOI, air intake mass (AIM) and rail pressure (RP), as shown in Figure 1. The considered model can

\* Corresponding author (email: huyf@jlu.edu.cn)



**Figure 1** (Color online) The open-source dataset. (a) Torque; (b) speed; (c) APS; (d) FIR; (e) SOI; (f) AIM; (g) RP.

be applied to some aspects such as estimating for torque and air-fuel ratio, and designing controller for fuel injection and rail pressure. Moreover, in addition to modeling, the data is potential to be utilized for engine identification and validation.

*Usage.* The policies and regulations about energy conservation and emission reduction are becoming increasingly stringent. In the future, vehicles will be tested under real-time road conditions. For simultaneously achieving contradictory high fuel efficiency and low emissions in diesel vehicles, estimators and control-oriented models for engine effective torque are usually indispensable. In this study, in order to demonstrate the functionality of the dataset, a case study for modelling the transient effective torque is proposed by using neural networks (NNs) method with the introduced open-source dataset.

The highway conditions are selected as a training dataset for identification, while the urban areas and suburbs conditions are selected as test data for validation. A suitable NN structure needs to be

determined to balance the conflicting requirements between accuracy and computation load. Thus, in this study three NN layers are employed, namely, input layer, hidden layer and output layer. The number of neurons  $I$  in the input layer varies between 4 and 16, whereas the number of neurons  $H$  in the hidden layer varies between 10 and 50, and the number of neurons  $O$  in the output layer is 1. The maximum number of training epochs is varied in the range of [3000, 6000] with a step of 500. Through testing, the NNs with the greatest density around the minimum error is selected (i.e.,  $I = 16$ ,  $H = 30$ ,  $O = 1$  and training epoch number = 6000). Overall, the mean absolute difference (MAD) is 13.2405 N·m, and the root mean squared error (RMSE) is 24.9043 N·m, compared between the measured value and predicted value. The results of the validation can satisfy the requirements of control-oriented modelling in diesel engines.

*Access methods.* **The .mat file of the open-source dataset** can be downloaded from: <http://www.ascl.jlu.edu.cn/vci/info/1010/1082.htm>.