Electronic Research Archive

ERA, 31(10): 6197–6215.
DOI: 10.3934/era.2023314
Received: 27 July 2023
Revised: 04 September 2023
Accepted: 06 September 2023
Published: 18 September 2023

http://www.aimspress.com/journal/ERA

Research article

Empowering high-speed train positioning: Innovative paradigm for generating universal virtual positioning big data

Xiaoyu Zheng[1,2], Dewang Chen[1,2,*] and Liping Zhuang[3]

[1] School of Transportation, Fujian University of Technology, Fuzhou 350118, China
[2] Intelligent Transportation System Research Center, Fujian University of Technology, Fuzhou 350118, China
[3] School of Electronics, Electrical Engineering and Physics, Fujian University of Technology, Fuzhou 350118, China

* **Correspondence:** Email: dwchen@fjut.edu.cn.

Abstract: High-speed trains (HSTs) positioning is a critical technology that affects the safety and operational efficiency of trains. The unique operating environment of HSTs, coupled with the limitations of real data collection, poses challenges in obtaining large-scale and diverse positioning data. To tackle this problem, we introduce a comprehensive method for generating virtual position data for HSTs. Utilizing virtual simulation technology and expert expertise, this method constructs a HST operating simulation environment on the Unity 3D platform, effectively simulating a range of operating scenarios and complex scenes. Positioning data is collected using virtual sensors, while error characteristics are incorporated to emulate real data collection behavior. The contribution of this paper lies in providing abundant, reliable, controllable and diverse positioning data for HSTs, thereby offering novel insights and data support for the evaluation and optimization of positioning algorithms. This method is not only applicable to various routes and scenarios, but also delivers fresh perspectives on data generation for research in other domains, boasting a broad scope of application.

Keywords: intelligent transportation; high-speed railway; virtual train positioning data; big data; dataset generation

1. Introduction

High-speed railways play a vital role in modern urbanization and transportation network construction [1,2]. They not only transform people's travel modes and enhance transportation efficiency but also drive economic development and urbanization, making them an essential transportation tool in contemporary society [3,4].

The positioning technology of HSTs is a complex and critical task. Accurate train positioning data is of great significance for the safe operation of trains, vehicle condition monitoring, and transportation management [5]. However, traditional train positioning methods face many challenges in practical applications due to the unique operating environment of HSTs, such as high operating speeds, complex terrain and topography and variable weather conditions. Therefore, the development of efficient, accurate and practical train positioning methods becomes an important problem that needs to be solved in the field of railway transportation.

In train positioning research, the collection and use of large-scale and high-quality positioning datasets are crucial [6]. These datasets not only provide a foundation for the development and validation of train positioning algorithms but also serve as valuable reference materials for related research in the field of railway transportation [7,8]. However, due to the unique operating environment of HSTs and the complexity of data collection, there are limitations in terms of long acquisition cycles, high costs and single data types of train positioning datasets. Furthermore, real datasets may be limited by factors such as privacy and security, making it difficult for researchers to fully utilize these data for in-depth research and analysis.

Existing research mainly uses measurements or simulations to obtain operational data of HSTs [9–12]. The first method involves collecting data using hardware devices in actual engineering. Common high-speed train positioning methods include global satellite navigation systems, inertial navigation systems, and technologies related to ground equipment and communication signals. Michael [13] provided a dataset for various rail vehicle positioning experiments. The data were collected using the German Aerospace Center (DLR) research vehicle RailDriVE on a segment of the Braunschweig harbor railway. However, the experiment only involved repeated back and forth travel on a 1.2 km section at a maximum speed of 25 km/h, resulting in a relatively limited dataset. Winter [14] provided a dataset used for rail vehicle positioning experiments. It contains measurements of a 6-DOF IMU and two GNSS receivers. The sensors were mounted on a regular rail vehicle during a trip about 120 km from Chemnitz to Schwarzenberg and back. However, long acquisition cycles and high costs are noteworthy limitations of train positioning datasets. Another method is to construct a simulation model of the train to generate virtual train operation data. Cao et al. [15] generated simulated train operation data by combining classical train control models, discrete throttle settings, empty sections and segmented tunnel resistance, studying the optimization of HSTs running trajectories. Maksym et al. [16] collected a large amount of virtual data by simulating the running strategies of vehicles on different types of tracks and evaluated the performance of train operation. Yang et al. [17] introduced a distributed coupling simulation platform for the maglev transportation system and the vehicle-track-girder coupling model, taking into account the complex vehicle structure, the guideway structure, and the Proportional Integral Differential (PID) levitation control system. Yu et al. [18] studied the distributed data-driven event-triggered model-free adaptive iterative learning control of multiple HSTs under iteration-varying topologies, which breaks away from the dependence on train dynamics. Based on the above research, the use of computer technology to establish accurate

models and algorithms can simulate the operation of trains under different conditions, including different speeds, different loads and different route conditions. This helps to evaluate the performance, stability and safety of train operation, save time and costs and reduce the risks of actual testing. In addition, virtual data effectively avoids the problems of data missing and abnormalities in the actual data acquisition process, improving the efficiency of the data preprocessing process.

To compensate for the limitations of real datasets and promote the development of train positioning technology, virtual train data generation methods have become increasingly important. Through virtual data generation techniques, researchers can simulate the operation of HSTs under different conditions and generate a large number of virtual positioning data with different characteristics. These virtual data not only meet the needs of large-scale datasets but also cover a variety of operating scenarios, providing a wide range of test bases for the design and optimization of train positioning algorithms.

Therefore, we propose a method for generating large-scale virtual train positioning data for general-purpose HSTs. This method utilizes virtual simulation technology to simulate train operations in different scenarios of high-speed railways and generate virtual large-scale train operation data. Next, performance evaluation indicators are defined, and high-quality virtual data are selected from a large amount of virtual operation data. Finally, virtual sensors are used to collect train operation data at a certain sampling interval, efficiently obtaining a large, rich and specific dataset of HSTs positioning. Specifically, this paper contributes in the following ways:

(i) Construction of a highly customizable virtual train operation environment for high-speed railways by combining scene fusion technology with geographical information data. Train operating parameters are set based on expert experience, and a highly realistic virtual simulation system for HSTs is built using the Unity 3D engine.

(ii) Generation of highly continuous and extensively covered virtual train operation curves using the proposed model. Multiple evaluation indicators are defined for the operation curves, allowing the selection of high-performance curves from the abundant pool of virtual data.

(iii) Development of virtual sensors and their error characteristic models to simulate the collection of positioning data during virtual train operations. Two sets of HSTs positioning datasets, one impacted by positioning noise and the other under ideal conditions, are generated as adversarial samples to evaluate the accuracy and robustness of HSTs positioning algorithms.

2. Design of train operation and positioning simulation model

Constructing a dynamic model of HSTs is an integral aspect of designing a simulation to replicate the virtual operational behavior of trains. Furthermore, simulating the error characteristics of the sensors is crucial to ensure a close correspondence between the obtained positioning data of HSTs and the actual situation.

2.1. Dynamics model of HSTs

The HSTs operation model is a complex problem of nonlinear mechanics. During the train operation process, considering the relatively small number of vehicles in the train formation and the train length compared to the running distance, the train set can be treated as a single point mass model for analysis. The existing train dynamics models are typically described by the following system of

equations [19–21].

$$\begin{cases} \widehat{M}a = \varepsilon_v^a F(t,v,a) - \varepsilon_v^b B(t,v,a) - \widehat{M}(f_r(v) - f_g(s)) \\ a = dv/dt \\ v = ds/dt \\ f_r(v) = \alpha v^2 + \beta v + \gamma \\ f_g(s) = g\,sin(\xi(s)) \end{cases} \tag{1}$$

where s, v, a, t represent the train's position, velocity, acceleration/deceleration and current time; $\widehat{M}$ represents the total mass of the train; F, B represent the traction force and braking force of the train, both of which are related to t, v, a[22]; $\varepsilon_v^a, \varepsilon_v^b$ represent the relative traction and braking coefficients; $f_r(v)$ represents the Davis equation describing the air and friction resistance between the train's velocity v, where $\alpha, \beta, \gamma$are Davis coefficients[23]; $f_g(s)$ represents the resistance generated by the track gradient and $\xi(s)$ is the gradient at the position s.

2.2. Simulation of sensor error characteristics

The role of train sensors is to monitor and detect the operational status and environmental parameters of trains to ensure safe, efficient and punctual railway transport, as shown in Figure 1. We utilize speed sensors, acceleration sensors, and transponders to form a virtual multi-sensor module for collecting HSTs positioning data. Table 1 provides a technical analysis of the proposed sensors.

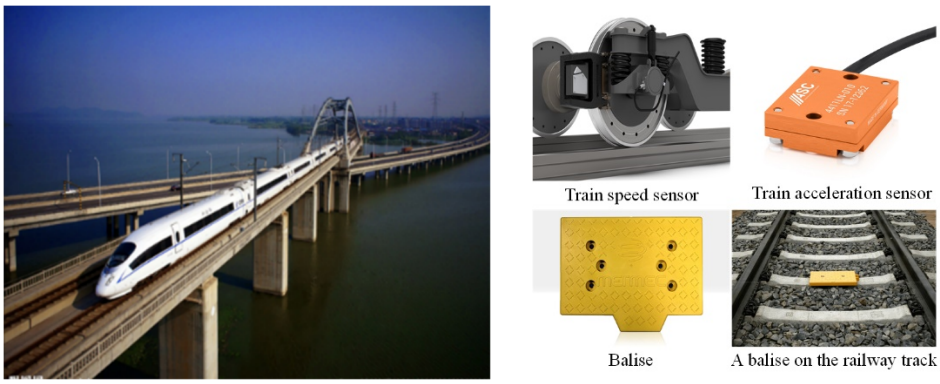


Figure 1. Schematic of sensors proposed.

1) Simulation of system errors

During the measurement process of train sensors, various errors may occur, which can affect the accuracy and reliability of the train positioning data. Common train sensor errors include measurement noise, nonlinearity, cross-interference and environmental influence. The measurement deviation caused by external environmental factors on the sensors is often approximated as zero-mean Gaussian white noise [24]. In this study, the train operation data is simulated with zero-mean Gaussian noise to represent the sensor system noise error, as shown in Eq (2).

$$X_i \sim N(\mu, \sigma^2) \tag{2}$$

where $X_i = (a_i, v_i, s_i)$ represents the vehicle state information at each sampling point, including acceleration, speed and distance traveled. μ is the average of $X_{i=1}, \ldots, X_{i=n-1,}\ X_{i=n}$, and σ^2 is the variance of $X_{i=1},\ \ldots,\ X_{i=n-1,}\ X_{i=n}$.

2) Simulation of wheel diameter error

As the operation time of a vehicle progresses, the wheels experience wear and tear. If the wheel diameter is set according to the factory parameters, it will inevitably introduce errors. Additionally, traction and braking during train operation accelerate the wear of the wheels. To simulate the wheel diameter error, a wheel diameter error coefficient γ is introduced for the virtual speed sensor and virtual acceleration sensor. Assuming the actual speed is represented as v_i^r, the measured speedv_i^mcan be expressed by Eq (3).

$$(v_{i-M+1}^m, \ldots, v_{i-1}^m, v_i^m) = (v_{i-M+1}^r, \ldots, v_{i-1}^r, v_i^r) * (1-\gamma) \tag{3}$$

where the variable γ represents the degree of wear on the train wheels, which ranges from 0 to 1.

Table 1. Analysis of sensor positioning technology.

Technique	Characteristic	Main advantages	Main disadvantages
Speed sensor	Continuously	1) Easy speed measurement and signal processing, less affected by the environment 2) High measurement accuracy at high speeds	1) Influenced by wheel counting errors and wheel wear 2) Positioning error accumulates with distance due to speed integration
Acceleration sensor	Continuously	1) All-weather and high autonomy 2) Can continuously receive multi-dimensional train operation information with high output frequency	1) Existence of drift and inherent errors 2) Errors accumulate with time, long-term accuracy requirements cannot be met
Balise	Point-based	1) Large amount of information transmission, provides absolute position information. 2) Less affected by the environment	1) High deployment quantity 2) High cost

3) Simulation of wheel idling and sliding errors

During the operation of HSTs, the wheels may experience idling and sliding due to factors such as the wheel-rail adhesion coefficient and the instantaneous acceleration of the train. When the traction of the train exceeds the maximum static friction force between the wheels and rails, the contact state between them is suspended [25]. The idling phenomenon of the train wheels mainly occurs during the acceleration stage when the train starts. The wheels will idle due to the excessive traction of the train. When idling occurs, the position for measuring idling can be expressed by Eq (4).

$$S_c = S_a + S_e \tag{4}$$

where S_c represents the position measurement value calculated by the train's sensors, S_a represents the actual position value of the train and S_e represents the position error.

The error of wheel idling in the train is simulated during the starting and acceleration phases by incorporating the idling ratio α. Assuming the actual speed is v_t^r, the measured speed v_t^m can be

expressed by Eq (5).

$$(v_{i-M+1}^{m}, \ldots, v_{i-1}^{m}, v_{i}^{m}) = (v_{i-M+1}^{r}, \ldots, v_{i-1}^{r}, v_{i}^{r}) * (1 + \alpha) \tag{5}$$

where α represents a value between 0 and 1, simulating the extent of wheel idling.

The phenomenon of wheel sliding in the train mainly occurs during the braking phase, where the excessive braking force causes relative motion between the train wheels and the tracks, resulting in the occurrence of sliding. When sliding occurs, the measurement position of the train can be represented by Eq (6).

$$S_c = S_a - S_e \tag{6}$$

where S_c represents the position measurement value calculated by the train's sensors, S_a represents the actual position value of the train and S_e represents the position error.

During the deceleration phase, the error in wheel sliding speed in the train is simulated by incorporating the sliding ratio β. Assuming the actual speed is v_i^r, the measured speed v_i^m can be expressed by Eq (7).

$$(v_{i-M+1}^{m}, \ldots, v_{i-1}^{m}, v_{i}^{m}) = (v_{i-M+1}^{r}, \ldots, v_{i-1}^{r}, v_{i}^{r}) * (1 - \beta) \tag{7}$$

where β represents a value between 0 and 1, simulating the extent of wheel slidding.

The balise is a correction device used in the train positioning process to correct cumulative errors and provide the train with absolute position at regular intervals distributed along the track. During the experiments, the HSTs accumulates errors as it runs on the track and the cumulative error at the end of the train's entire journey is estimated by Eq (8).

$$\Gamma = \sum_{i=1}^{n} s_i(t) - m_i(t) \tag{8}$$

where $\mathrm{s_i}(t)$ represents the train position measured when passing each transponder, while $m_i(t)$ represents the actual train position.

2.3. Building a virtual simulation model for trains based on unity 3D

In the research on HSTs positioning, the construction of a virtual simulation model is crucial for simulating the real train operating environment and generating realistic virtual positioning data, as shown in Figure 2. By utilizing the powerful capabilities and flexibility of the Unity 3D engine, we can build highly customizable and diverse virtual train operating environments, providing strong support for subsequent positioning algorithm validation and performance evaluation.

In this framework, the virtual HSTs model and virtual sensor model are connected to the Unity 3D virtual simulation module for model configuration. Furthermore, in the Unity 3D virtual simulation, the construction and configuration of the HSTs model, track and route model and environmental model are carried out sequentially. The HSTs behavior control module receives external parameter inputs to simulate the train's operating behavior. Finally, HSTs positioning data is acquired through virtual sensors.

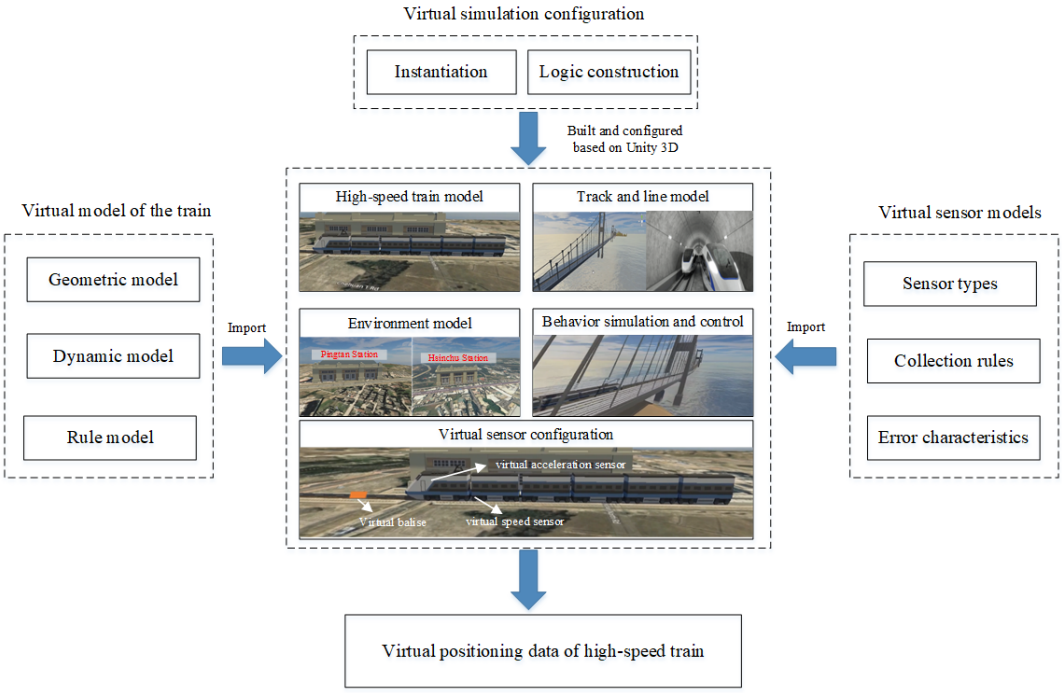


Figure 2. Framework of virtual simulation model.

3. Methodology for generating virtual HSTs positioning dataset

In this section, a virtual operational simulation model for HSTs is developed by integrating geometric, dynamic and environmental models. Leveraging the power of big data and virtual simulation technologies, a virtual positioning dataset for HSTs is generated. The detailed process is illustrated in Figure 3.

The virtual running curves serve as the foundation for generating virtual positioning data, which accurately depict the train's position and operational status within the virtual environment. In the virtual simulation environment, a large amount of virtual running data for HSTs is generated by simulating their operational processes, including acceleration, deceleration and turning. Subsequently, the virtual positioning data for the HSTs are simulated, considering the error characteristics observed during the data acquisition process of real-world positioning sensors.

During its operation, the high-speed train undergoes multiple processes of acceleration, deceleration and uniform motion. According to expert experience, the train line is divided into several sections while setting the maximum and minimum running speed limits for each section. Figure 4 shows the schematic diagram of the virtual operating curve. Before the n-th section, each section includes the process from constant speed operation to constant speed operation. The n-th section contains two constant speed changes and one constant speed driving process. In each operating section, specific acceleration limits $(a_i^{min} \sim a_i^{max})$ and speed limits $(v_i^{min} \sim v_i^{max})$ are set for the train. By setting the value interval within the acceleration and speed range, there are u_a type of acceleration value and u_v type of speed value in each section. When we combine the generation conditions of each section, we will get a large number of train operating curves.

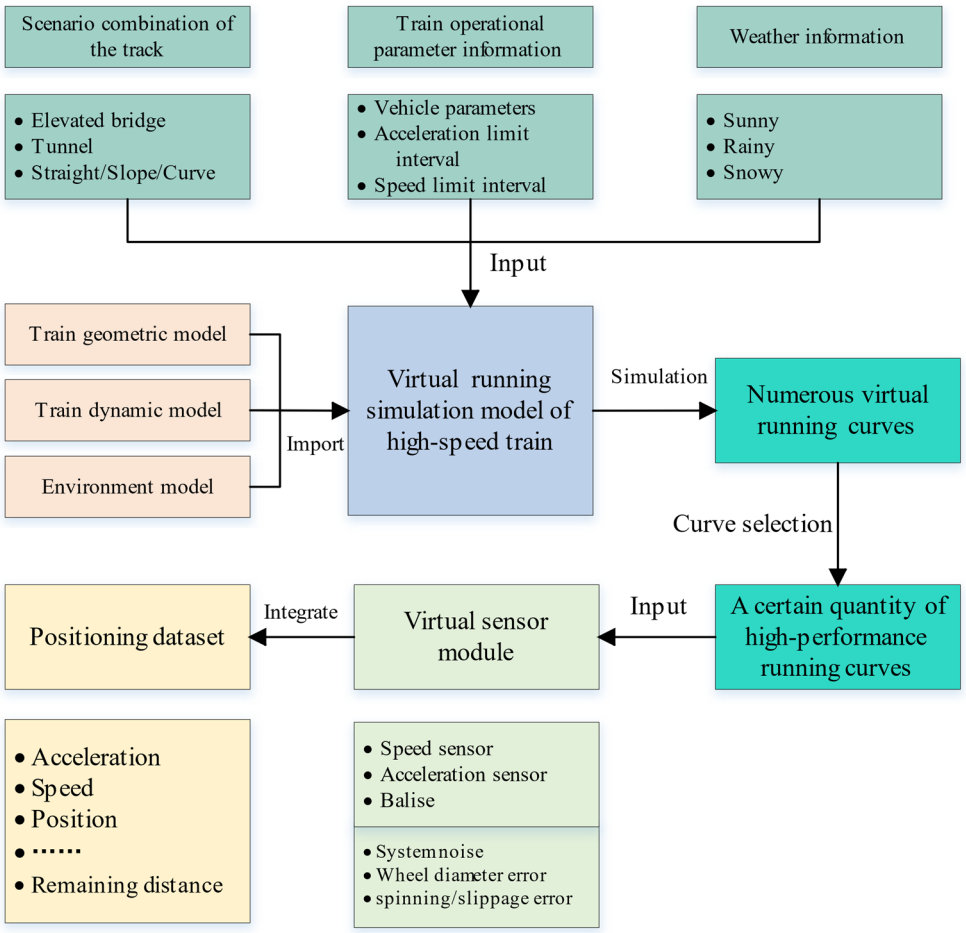


Figure 3. Flowchart of positioning data generation.

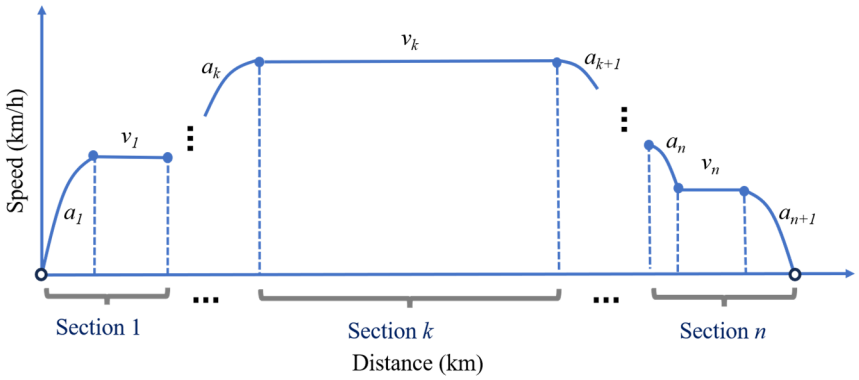


Figure 4. Illustration of HSTs operating curve based on interval speed limit.

The simulation of various operating scenarios for HSTs enables the acquisition of a diverse set of operating curve data. By considering both train energy consumption and passenger comfort, the performance of virtual train running curves can be comprehensively evaluated [26,27]. The objective of curves screening is to select the optimal train operation scheme for the train operation curve with a cycle of 1 s, and establish performance evaluation standards based on train comfort and energy

consumption. The virtual operating curve selection process is illustrated in Figure 5. First, the given target arrival time $T_0 = \{T_0^1, T_0^2, \ldots, T_0^m,\}$ is used to divide the actual arrival time dataset into m subsets $X = \{X_1, X_2, \ldots, X_m,\}$ based on the train travel curve. Then, the performance scores of all virtual running curves in each subset are calculated using evaluation indicators. Finally, the running curve with the highest performance score is selected within each subset.

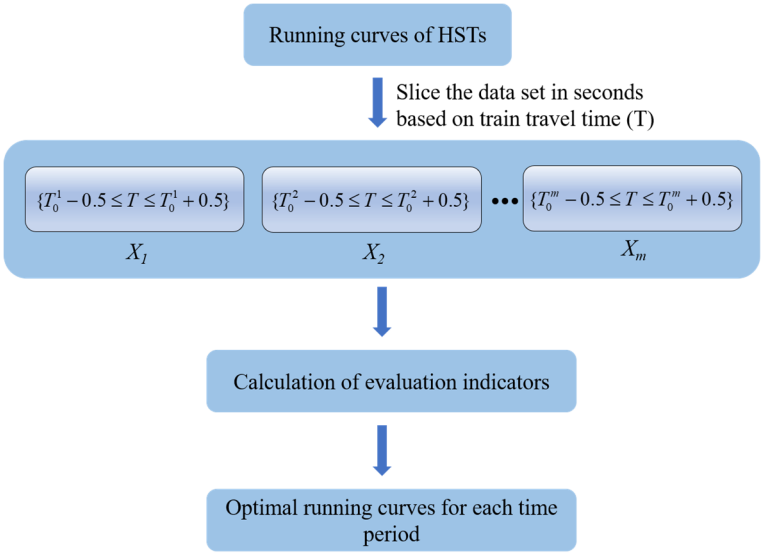


Figure 5. Virtual operating curve selection process.

4. Experimental simulation and analysis

4.1. Evaluation metrics

1) Train running energy consumption: ψ

The speed and running distance mainly influence the study of energy consumption of HSTs, and the energy consumption of trains is calculated to compare the advantages and disadvantages of train running curve performance by Eqs (9) and (10).

$$\psi = min(\sum_{i=1}^{n} K \times \frac{V_{avg}^2}{ln(X_i)} + C) \tag{9}$$

$$V_{avg} \approx \frac{X_i}{t_{var}+t_{uni}} \tag{10}$$

where ψ is the energy consumption, kJ/(t·km); V_{avg} is the average speed in the i-th running section; X_i is the distance of the i-th running section; K and C denote the constants related to the train; t_{var} is the variable speed time of the i-th running section; t_{uni} is the uniform speed time of the i-th running section.

2) Passenger comfort: η

Comfort is an important indicator to measure the feeling of passengers riding on board. Because the acceleration must be adjusted during HSTs according to different route conditions, the acceleration changes must be controlled within a specific range to protect the passengers' feelings. Therefore, the

comfort level is also an essential basis for measuring the performance of the train running curve by Eq (11).

$$\eta = \frac{1}{\sum_{i=1}^{n} \Delta|a|} \tag{11}$$

where $\Delta|a_i|$denotes the acceleration change value of the i-th segment.

3) The performance of running curves: ξ_{curve}

In order to select a small number of high-quality running curves from a large number of virtual train running curves that do not miss the running characteristics, we first classify each running curve in its unit moment. Next, the curve performance was evaluated by assigning a weighting of 50% each to ψ and η. Finally, the optimal train running curve is filtered at each running time. In addition, to avoid the magnitude influence of ψ and η, it is necessary to normalize these two indicators. The specific processing process is shown in Eqs (12) and (13).

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{12}$$

where x^* is the normalized index value, x is the index value before normalization, x_{min} is the minimum value of the index value and x_{max} is the maximum value of the index value.

$$\xi_{curve} = min(\sum_{T=t_0}^{t_n} 0.5 \times EC_{train} + 0.5 \times \delta_{comfort}) \tag{13}$$

where ξ_{curve} represents the quantified index with the ability to assess the performance of the running curve; $T = t_0, \ldots, t_n$ represents the range of the total time distribution of the train operation.

4) Data evaluation metrics for positioning

The data collection process for positioning involves periodically acquiring HSTs operation data with a sampling period of t. The collected data is then processed to simulate sensor error characteristics. To assess the performance of the virtual sensor model, metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), Maximum Error (ME) and Standard Deviation of Error (SDE) are used to analyze the data before and after introducing noise. The specific content can be found in Eqs (14)–(18).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} X_i^{meas} - X_i^{true} \tag{14}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (X_i^{meas} - X_i^{true})^2 \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (X_i^{meas} - X_i^{true})^2} \tag{16}$$

$$ME = max(|X_i^{meas} - X_i^{true}|) \tag{17}$$

$$SDE = std(X_i^{meas} - X_i^{true}) \tag{18}$$

where $X_t^{true} = (a_t, v_t, s_t)$ represents the true train positioning data, and $X_t^{meas} = (a_t, v_t, s_t)$ represents the train positioning data measured by the sensors.

4.2. Simulation description

We present a study on virtual simulation for cross-sea high-speed railways. By constructing highly realistic and diverse virtual operating environments, the operation of cross-sea high-speed railways is simulated, and a large-scale virtual positioning dataset is generated. In order to ensure safe and efficient train operation, the majority of the track consists of straight segments with some curved sections as required. The train's trajectory is constrained by the track, and the HSTs is assumed to operate mainly in the horizontal plane, disregarding pitch, roll and vertical velocity changes. To simulate the scenario effectively, the CRH3 type EMU is chosen as the representative train model. The virtual track covers a distance of approximately 150 km, with Station A as the starting point and Station B as the endpoint, featuring a cross-sea tunnel that spans 135 km in length. Three operating scenarios are designed to examine the train's performance at maximum speeds of 250, 350 and 400 km/h, as detailed in Table 2.

Table 2. Simulation scenarios for the HSTs.

Simulation scheme	Maximum speed (km/h)	Weather conditions	Division of track sections	Length (km)	Speed limit sections (km/h)	Acceleration limit sections (m/s²)
Scheme 1	250	Rain/Snowy Weather	1	35	(120, 150)	(0.2, 0.4)
			2	80	(220, 250)	(0.2, 0.4)
			3	35	(120, 150)	(-0.2, -0.4)
Scheme 2	350	Sunny Weather	1	15	(120, 150)	(0.2, 0.6)
			2	20	(200, 230)	(0.2, 0.6)
			3	80	(320, 350)	(0.2, 0.6)
			4	35	(220, 250)	(-0.2, -0.6)
Scheme 3	400	Sunny Weather	1	30	(170, 200)	(0.3, 0.9)
			2	80	(370, 400)	(0.3, 0.9)
			3	20	(280, 310)	(-0.3, -0.9)
			4	20	(200, 230)	(-0.3, -0.9)

Scheme 1 focuses on simulating the operation of HSTs under rainy/snowy conditions with a maximum operating speed of 250 km/h. The track is divided into three sections: the first one encompasses the downhill segment from Station A to the underwater tunnel, the second consists of the long-distance underwater tunnel where the train can reach its maximum speed, and the third entails the segment from the tunnel exit to Station B. The acceleration and speed ranges are determined for each section based on expert knowledge, with a speed interval of 5 km/h and an acceleration interval of 0.1 m/s².

Scheme 2 concentrates on simulating the operation of HSTs in clear weather conditions with a maximum operating speed of 350 km/h. The track is divided into four sections: the first 35 km is divided into two segments, ensuring continuous acceleration to enable the train to reach its maximum speed in the third section, which is the long-distance underwater tunnel. The acceleration and speed ranges are established for each section, with a speed interval of 5 km/h and an acceleration interval of 0.2 m/s².

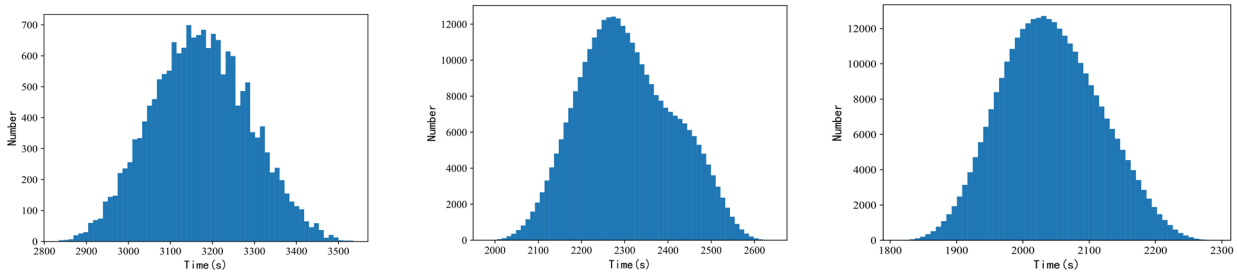
Scheme 3 also focuses on simulating HSTs operation in clear weather conditions with a maximum operating speed of 400 km/h. The track is divided into four sections: unlike schemes 1 and 2, the first section simulates the train's downhill travel at a speed limit of 200 km/h. The second section consists of the long-distance underwater tunnel where the train operates at its maximum speed. Finally, the

third and fourth sections simulate the train's deceleration in segments. The train's speed limit parameters are determined based on scheme 2.

4.3. Simulation results

4.3.1. Generation and selection of running curves

The three schemes generated a significant number of virtual operating curves for HSTs. As shown in Figure 6, the time distribution of the generated virtual operating curves follows a normal distribution and exhibits continuity in terms of train running curve time, with numerous operating curves corresponding to each second. To select representative and high-performance curves, it is necessary to utilize statistical indicators to analyze the optimal curve corresponding to each operating duration.



(a) Scheme 1 (b) Scheme 2 (c) Scheme 3

Figure 6. Distribution of running curve time.

The simulation data for Scheme 1 is presented in Table 3, for Scheme 2 in Table 4 and for Scheme 3 in Table 5. Scheme 1 generated 17,496 virtual running curves, with travel times ranging from 2873 seconds to 3590 seconds, encompassing a total of 718 consecutive time points. Following curve performance evaluation, 718 curves were selected. Scheme 2 generated 314,928 virtual running curves, with travel times ranging from 1982 seconds to 2642 seconds, encompassing a total of 661 consecutive time points. After curve performance evaluation, 661 curves were selected. Scheme 3 generated 314,928 virtual running curves, with travel times ranging from 1813 seconds to 2289 seconds, encompassing a total of 477 consecutive time points. After curve performance evaluation, 477 curves were selected. All three sets of schemes successfully filtered the optimal virtual running curves at each runtime, maintaining continuity before and after the selection process.

To better understand the performance differences between different operation schemes of HSTs, the average travel times for the three schemes were measured at 3222.7 seconds, 2304.8 seconds and 2042.0 seconds. These results show that as the maximum speed of the trains increases, the travel time can effectively decrease. However, it is important to note that average energy consumption rates and comfort levels also vary accordingly. Specifically, average energy consumption rates increase by 59.1 and 10.8% while average comfort levels decrease by 37.5 and 28.0%. These numbers indicate a trade-off between travel time, energy consumption and comfort levels.

Table 3. Simulation curve selection for scheme 1.

Indicator	Before selection	After selection
Runtime range (s)	2873–3590	2873–3590
Number of curves	17,496	718
Energy range (kJ/(t·km))	79.1–86.6	79.3–86.8
Comfort range	0.6–1.25	0.6–1.25
Average travel time (s)	3222.7	3232.0
Average energy (kJ/(t·km))	82.7	79.8
Average comfort	0.8	1.1

Table 4. Simulation curve selection for scheme 2.

Indicator	Before selection	After selection
Runtime range (s)	1982–2642	1982–2642
Number of curves	314,928	661
Energy range (kJ/(t·km))	124.5–138.9	124.5–137.2
Comfort range	0.3–1	0.3–1
Average travel time (s)	2304.8	2309.5
Average energy (kJ/(t·km))	131.6	128.8
Average comfort	0.5	0.6

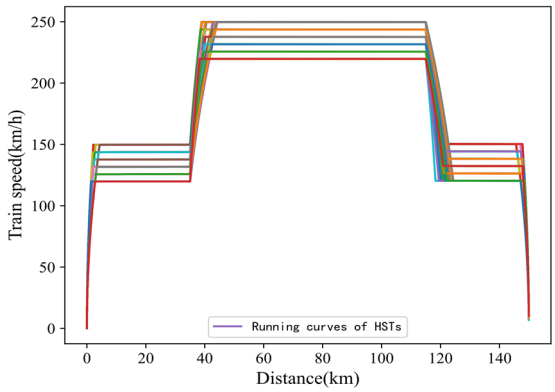
Table 5. Simulation curve selection for scheme 3.

Indicator	Before selection	After selection
Runtime range (s)	1813–2289	1813–2289
Number of curves	314,928	477
Energy range (kJ/(t·km))	137.9–153.8	138.0–152.3
Comfort range	0.25–0.67	0.25–0.67
Average travel time (s)	2042.0	2050.4
Average energy (kJ/(t·km))	145.8	142.4
Average comfort	0.36	0.45

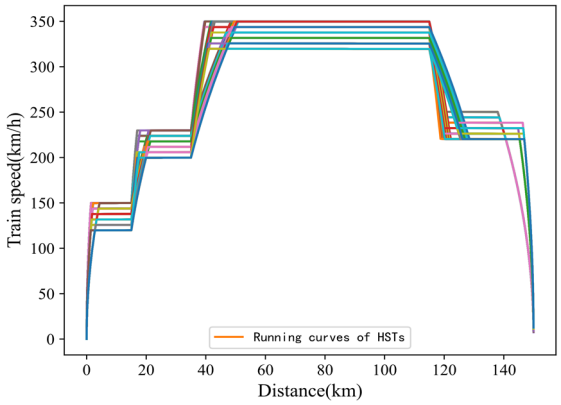
To further examine the impact of selection steps on the performance of the virtual train running curve, the average travel times of the three groups of train operation schemes after selection increased slightly by 0.3, 0.2 and 0.4% compared to before selection. This slight increase can be attributed to the removal of curves with high acceleration variation rates, which affects the overall travel time. However, in terms of energy consumption and comfort levels, the filtered operating curves show improvements compared to the pre-selection curves, with energy consumption optimized by 3.5, 2.1 and 2.3%, and comfort level improvements of 37.5, 20.0 and 25.0%. Figure 7 shows the speed-distance curves of virtual operation for HSTs after selection.

To highlight the advantages of the virtual data generation method in terms of data quantity and diversity, a comparison was made with existing methods [28]. The proposed method achieved significant breakthroughs in speed limits and section lengths, resulting in a 34% increase in the quantity of virtual data. Moreover, the method allows for extracting high-performance curves from numerous virtual running curves, providing greater flexibility to adjust the ideal virtual schemes and

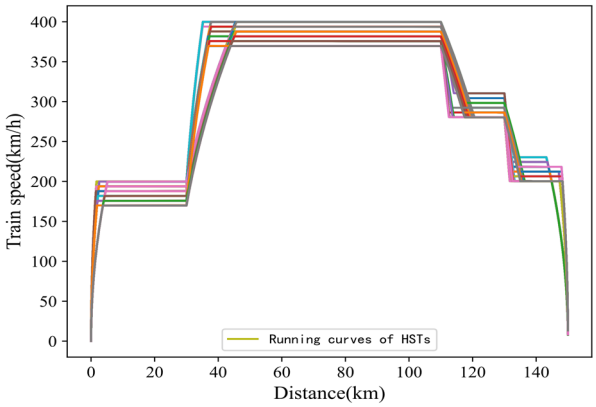
train speed limits according to actual needs. This demonstrates the superior performance of the proposed method compared to existing methods.



(a) Scheme 1 (local)



(b) Scheme 2 (local)

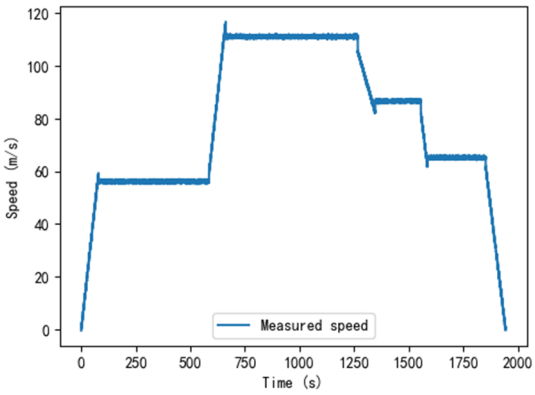


(c) Scheme 3 (local)

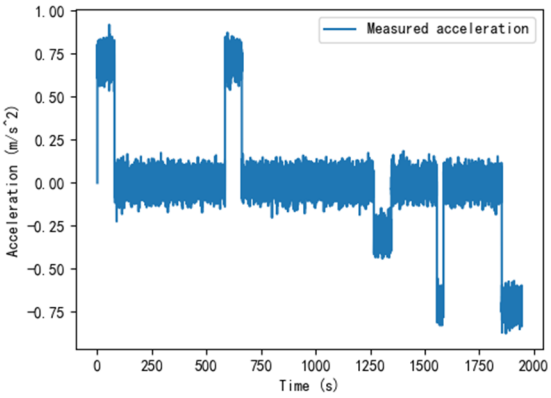
Figure 7. Speed-distance curves of virtual operation for HSTs after selection.

4.3.2. Generation of positioning data

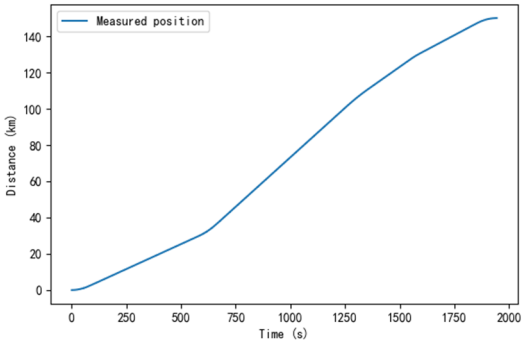
In this study, one of the HSTs operation curves from Scheme Three was selected to demonstrate the measurement process of positioning data using a sampling period of 0.1 seconds. To simulate realistic measurement errors, we introduced environmental noise, wheel diameter errors and idling/slidding errors. The environmental noise was set to follow a Gaussian distribution $X_i \sim N(0,0.05)$, the wheel diameter error coefficient γ was set to 0.02 and both the idling ratio α and slidding ratio β were set to 0.005.



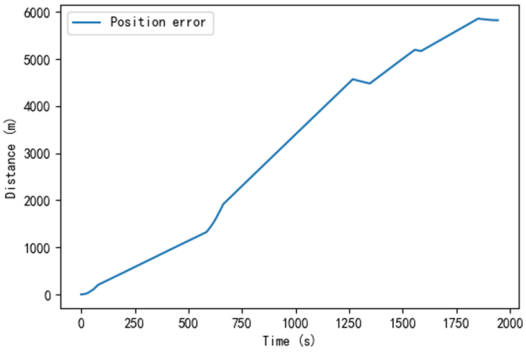
(a) Speed data at each time step



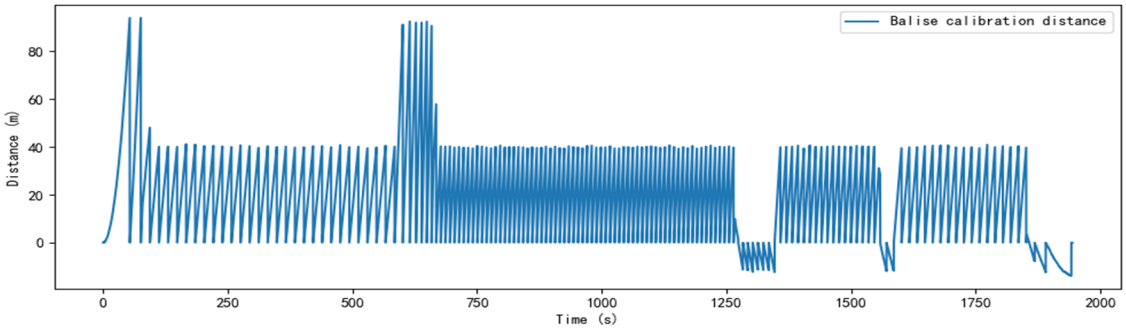
(b) Acceleration data at each time step



(c) Position data at each time step



(d) Accumulated error in train position



(e) Train position correction with balise

Figure 8. Results of HSTs positioning data.

The experimental results are shown in Figure 8. Figure 8(a) illustrates the relationship between acceleration and time during HSTs operation. Due to the introduction of environmental noise, the acceleration curve may exhibit fluctuations and jitter, reflecting the influence of noise interference on sensor data. Figure 8(b) shows the relationship between speed and time during HSTs operation. Due to the introduction of wheel diameter errors and idling/slidding errors, the speed curve may exhibit fluctuations or instability. Wheel diameter errors may cause speed measurement deviations, while idling/slidding errors may cause speed instability. Figure 8(c) reflects the difference between the actual

position and the measured position of the HSTs. Due to the introduction of error terms in the virtual model, including system noise, wheel diameter errors and idling/slidding errors, the measurement of the HSTs position gradually accumulates deviations. Figure 8(d) reflects the continuous accumulation of errors in measuring the position of the HSTs without the use of virtual balise. According to the current experimental parameters, the position error reaches 6 kilometers. Figure 8(e) reflects the correction of the train position after setting virtual balise at every kilometer. Position errors above zero indicate wheel idling during acceleration, influenced by wheel diameter errors and system errors, resulting in a measured train position greater than the actual position. During deceleration, position errors below zero indicate wheel sliding, which leads to a measured train position that is smaller than the actual position. These errors interact each other with wheel diameter errors and system errors, resulting in a lower error peak during deceleration compared to acceleration.

Table 6 evaluates the measurement results of the HSTs positioning data. Since only system errors were introduced to the acceleration sensor in the experiment, the values of MAE, RMSE and ME are low, indicating that the acceleration measurement results are less affected by errors. Regarding the evaluation of speed data, the values of MAE and RMSE are relatively small, indicating higher accuracy of the speed sensor's measurements. However, the value of ME is large, indicating significant deviations in speed measurements when the train's wheels slidding or idle. As for the evaluation metrics of position data, their values are all large. This is due to the combined influence of measurement errors in speed and acceleration sensors, leading to significant cumulative errors in the measurement data of the train’s position.

Table 6. Analysis of HSTs positioning data with noise.

Metric	Speed	Acceleration	Distance
MAE	1.84	0.04	20.30
MSE	4.54	0.0025	638.34
RMSE	2.13	0.05	25.26
ME	8.20	0.2	94.05
SDE	1.58	0.05	16.60

In conclusion, by analyzing the experimental results generated from the virtual positioning dataset, we can evaluate the performance of the virtual sensor error model and the accuracy of the generated positioning dataset. These results are of important reference value for optimizing the sensor error model and improving measurement accuracy, and provide useful data foundation for the design and testing of train operation and control systems.

5. Conclusions

To address the need for acquiring HSTs positioning data, we propose a method for generating virtual positioning big data for general HSTs. By combining virtual simulation technology with expert knowledge to set vehicle operation parameters, diverse route schemes and a large-scale dataset of virtual train positioning data are generated. During the data acquisition process, an error model for virtual sensors is introduced to simulate the error characteristics in data collection. By setting different noise models and error parameters, a HSTs positioning dataset with error characteristics is generated. Additionally, adversarial samples are constructed by comparing with ideal data to assess the accuracy

and robustness of positioning algorithms in future research. However, there are limitations in the experiment, such as considering only train energy consumption and passenger comfort as evaluation criteria during the virtual train trajectory selection process. Future research could define more performance metrics to comprehensively evaluate the performance of virtual train trajectories. Furthermore, introducing more types of virtual sensor noise can make the virtual train positioning data more closely resemble real-world data characteristics.

The proposed method of generating virtual positioning big data for HSTs based on virtual schemes and virtual sensors provides an innovative solution to the challenge of obtaining real data. By constructing a large-scale and diverse virtual positioning dataset, this approach provides strong support for research and development in HSTs positioning technology. In the future, we will further explore and optimize virtual data generation algorithms to be applicable to a wider range of train positioning schemes and promote the intelligent development of railway transportation.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 61976055), Third Batch of Innovative Star Talent Plan in Fujian Province (No. 003002), Special Fund for Education and Scientific Research of Fujian Provincial Department of Finance (No. GY-Z21001), and Scientifc Research Foundation of Fujian University of Technology (No. GY-Z22071).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Y. Ye, P. Huang, Y. Zhang, Deep learning-based fault diagnostic network of high-speed train secondary suspension systems for immunity to track irregularities and wheel wear, *Railway Eng. Sci.*, **30** (2022), 96–116. https://doi.org/10.1007/s40534-021-00252-z
2. V. Lauda, V. Novotný, Role of railway transport in green deal 2050 challenge-situation in the Czech Republic, *Promet-Traffic Transp.*, **34** (2022), 801–812. https://doi.org/10.7307/ptt.v34i5.4117
3. L. Zhang, Z. Wang, Q. Wang, J. Mo, J. Feng, K. Wang, The effect of wheel polygonal wear on temperature and vibration characteristics of a high-speed train braking system, *Mech. Syst. Signal. Process.*, **186** (2023), 109864. https://doi.org/10.1016/j.ymssp.2022.109864
4. I. A. Tasiu, Z. Liu, S. Wu, W. Yu, M. Barashi, J. O. Ojo, Review of recent control strategies for the traction converters in high-speed train, *IEEE Trans. Transp. Electrif.*, **8**(2022), 2311–2333. https://doi.org/10.1109/TTE.2022.3140470

5. J. Yang, J. Wang, Y. Zhao, Simulation of nonlinear characteristics of vertical vibration of railway freight wagon varying with train speed, *Electron. Res. Arch.*, **30** (2022), 4382–4400. https://doi.org/10.3934/era.2022222
6. H. Song, E. Schnieder, Availability and performance analysis of train-to-train data communication system, *IEEE Trans. Intell. Transp. Syst.*, **20** (2019), 2786–2795. https://doi.org/10.1109/TITS.2019.2914701
7. W. Li, O. P. Hilmola, J. Wu, Chinese high-speed railway: Efficiency comparison and the future, *Promet-Traffic Transp.*, **31** (2019), 693–702. https://doi.org/10.7307/ptt.v31i6.3220
8. F. Bădău, Railway interlockings—A review of the current state of railway safety technology in Europe, *Promet-Traffic Transp.*, **34** (2022), 443–454. https://doi.org/10.7307/ptt.v34i3.3992
9. H. Gu, T. Liu, Z. Jiang, Z. Guo, Experimental and simulation research on the aerodynamic effect on a train with a wind barrier in different lengths, *J. Wind Eng. Ind. Aerod.*, **214** (2021), 104644. https://doi.org/10.1016/j.jweia.2021.104644
10. H. Song, S. Gao, Y. Li, L. Liu, H. Dong, Train-centric communication based autonomous train control system, *IEEE Trans. Intell. Veh.*, **8** (2022), 721–731. https://doi.org/10.1109/TIV.2022.3192476
11. R. Kour, A. Patwardhan, A. Thaduri, R. Karim, A review on cybersecurity in railways, *Proc. Inst. Mech. Eng., Part F: J. Rail Rapid Transit*, **237** (2022), 3–20. https://doi.org/10.1177/09544097221089389
12. L. Zhang, Vibration analysis and multi-state feedback control of maglev vehicle-guideway coupling system, *Electron. Res. Arch.*, **30** (2022), 3887–3901. https://doi.org/10.3934/era.2022198
13. M. Roth, H. Winter, An open data set for rail vehicle positioning experiments, in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, (2020), 1–7. https://doi.org/ 10.1109/ITSC45102.2020.9294594
14. H. Winter, Rail vehicle positioning data set: Lucy, October 2018, Technische Universität Darmstadt, 2020. https://doi.org/10.25534/tudatalib-360
15. Y. Cao, Z. Zhang, F. Cheng, S. Su, Trajectory optimization for high-speed trains via a mixed integer linear programming approach, *IEEE Trans. Intell. Transp. Syst.*, **23** (2022), 17666–17676. https://doi.org/10.1109/TITS.2022.3155628
16. M. Spiryagin, Q. Wu, O. Polach, J. Thorburn, W. Chua, V. Apiryagin, et al., Problems, assumptions and solutions in locomotive design, traction and operational studies, *Railway Eng. Sci.*, **30** (2022), 265–288. https://doi.org/10.1007/s40534-021-00263-w
17. Y. Feng, C. Zhao, W. Zhai, L. Tong, X. Liang, Y. Shu, Dynamic performance of medium speed maglev train running over girders: field test and numerical simulation, *Int. J. Struct. Stab. Dyn.*, **23** (2023), 2350006. https://doi.org/10.1142/S0219455423500062
18. W. Yu, D. Huang, Q. Wang, L. Cai, Distributed event-triggered iterative learning control for multiple high-speed trains with switching topologies: A data-driven approach, *IEEE Trans. Intell. Transp. Syst.*, **2023** (2023). https://doi.org/10.1109/TITS.2023.3277452
19. J. Yin, C. Ning, T. Tang, Data-driven models for train control dynamics in high-speed railways: LAG-LSTM for train trajectory prediction, *Inf. Sci.*, **600** (2022), 377–400. https://doi.org/10.1016/j.ins.2022.04.004
20. Q. Wu, C. Cole, Computing schemes for longitudinal train dynamics: sequential, parallel and hybrid, *J. Comput. Nonlinear Dyn.*, **10** (2015), 064502. https://doi.org/10.1115/1.4029716

21. Q. Wu, C. Cole, S. Maksym, W. Yucang, W. Ma, C. Wei, Railway air brake model and parallel computing scheme, *J. Comput. Nonlinear Dyn.*, **12** (2017), 051017. https://doi.org/10.1115/1.4036421
22. K. Fadhloun, H. Rakha, A. Loulizi, A. Abdelkefi, Vehicle dynamics model for estimating typical vehicle accelerations, *Transp. Res. Rec.*, **2491** (2015), 61–71. https://doi.org/10.3141/2491-07
23. Y. Cheng, J. Yin, L. Yang, Robust energy-efficient train speed profile optimization in a scenario-based position–Time–Speed network, *Front. Eng. Manage.*, **8** (2021), 595–614. https://doi.org/10.1007/s42524-021-0173-1
24. M. Zhang, Z. Yang, J. Cheng, Speed and distance measurement algorithm of train control onboard equipment based on adaptive federated filter, *China Railway Sci.*, **43** (2022), 144–151.
25. H. Chen, T. Furuya, S. Fukagai, S. Saga, J. Ikoma, K. Kimura, et al., Wheel slip/slide and low adhesion caused by fallen leaves, *Wear*, **446** (2020), 203187. https://doi.org/10.1016/j.wear.2020.203187
26. D. Zhang. Y. Tang, Q. Peng, A novel approach for decreasing driving energy consumption during coasting and cruise for the railway vehicle, *Energy*, **263** (2023), 125615. https://doi.org/10.1016/j.energy.2022.125615
27. Y. Peng, Y. Lin, C. Fan, Q, Xu, D. Xu, S. Yi, et al., Passenger overall comfort in high-speed railway environments based on EEG: assessment and degradation mechanism, *Build. Environ.*, **210** (2022), 108711. https://doi.org/10.1016/j.buildenv.2021.108711
28. Y. Lu, D. Chen, Z. Zhao, Algorithm for automatically generating a large number of speed curves of subway trains based on AlphaZero, *Chin. J. Intell. Sci. Technol.*, **3** (2021), 179–184.

AIMS Press