Data-Driven Substation Energy Minimization for Train Speed-Profile and Dwell-Time Optimization

Liu, Xiao; Tian, Zhongbei; Gao, Yuan; Jiang, Lin; Goverde, Rob M.P.

DOI
10.1109/TTE.2025.3575703

Publication date
2025

Document Version
Final published version

Published in
IEEE Transactions on Transportation Electrification

Citation (APA)
Liu, X., Tian, Z., Gao, Y., Jiang, L., & Goverde, R. M. P. (2025). Data-Driven Substation Energy Minimization for Train Speed-Profile and Dwell-Time Optimization. *IEEE Transactions on Transportation Electrification* , *11*(5), 11320-11331. https://doi.org/10.1109/TTE.2025.3575703

Data-Driven Substation Energy Minimization for Train Speed-Profile and Dwell-Time Optimization

Xiao Liu, Zhongbei Tian, *Member, IEEE*, Yuan Gao, *Member, IEEE*, Lin Jiang, *Member, IEEE*, and Rob M. P. Goverde, *Member, IEEE*

***Abstract*— As regenerative braking systems become more widespread in railways, rising attention is paid to collaborative train operations under optimized timetables to enhance regenerative braking efficiency. The effective usage of regenerative braking energy (RBE) is determined by the dynamic nature of the traction power supply network, driven by constant changes in train power and positions. Solving the power flow with multiple trains significantly, however, increases the computing time required to solve the optimization model. Most existing methods have to solve optimization problems neglecting the dynamic power flow analysis, which sacrifices the accuracy of regeneration efficiency. In order to address this challenge, we propose a data-driven model that emulates the power flow analysis and reduces the computational demands. Initially, data from both single and multitrain simulators are collected and stored in a database, from which critical information regarding train position, power, and substation power is extracted. A neural network is then used to develop a data-driven model that predicts the power of a substation in a power supply network based on train positions and powers. Case studies with Beijing Yizhuang Metro line data show that the calculation time of the data-driven model is 0.33% of the power flow simulation while keeping the accuracy above 99%. Based on this data-driven model, by optimizing train speed profile and dwell time, the energy supplied by substations can be reduced by up to 13% compared to traction optimization.**

***Index Terms*— Data-driven, energy saving, headway, railway energy supply, speed profile.**

I. Introduction

SUSTAINABLE railway technologies have gained significant attention in recent years due to the environmental and economic benefits associated with energy efficiency [1]. With railway energy consumption increasing, researchers have explored strategies for integrating renewable energy sources, such as wind power [2] and hydrogen [3], into the railway power supply system to reduce carbon emissions and enhance sustainability. Renewable energy sources depend on external generation and are, however, subject to weather-related variability, making them less predictable for real-time railway operations. In contrast, regenerative braking energy (RBE) offers an immediate, self-sustained energy source, recovering kinetic energy from braking trains and either reusing it within the system or storing it for later use [4]. Effectively harnessing RBE, however, requires accurate power flow modeling and efficient optimization of train operations, both of which pose significant challenges.

Existing studies have explored various methods to enhance RBE usage. For a single train, Scheepmaker and Goverde [4] compared the optimal speed profiles considering solely mechanical braking to those that account for both mechanical braking and RBE simultaneously. Yang et al. [5] developed an integer programming model to optimize train timetables by defining overlapping time between motoring and braking trains. Other works [6] extended these approaches to integrate train driving strategies and timetables, while onboard energy storage systems have been explored as another means to improve RBE recovery efficiency [7]. In spite of these efforts, practical challenges, especially related to power transmission losses, remain a critical concern that could undermine the effectiveness of these methods. In order to address this, Peña-Alcaraz et al. [8] proposed a power-saving factor to quantify the total amount of RBE transferred to the motoring trains. Su et al. [9] developed a distance-based RBE usage model. Pan et al. [14] enhanced RBE usage through an overlap current method, and Ning et al. [15] introduced a linear regression model to simplify the calculation process of substation energy consumption. These simplifications, however, do not fully account for the complexities of the power supply network. In order to achieve a more accurate representation of transmission losses and ensure the energy provided by braking trains is effectively used by motoring trains, a more comprehensive modeling of the entire power supply system is suggested [10].

Recognizing these limitations, Chen et al. [11] modeled the coupling mechanism between one train and two reversible substations and then adopted a pseudospectral method to achieve the optimal energy-efficient train control and schedule under

Received 28 January 2025; revised 18 March 2025 and 28 April 2025; accepted 26 May 2025. Date of publication 2 June 2025; date of current version 25 September 2025. This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) Project Connected and Coordinated Train Operation and Traction Power Supply Systems (COOPS) under Grant EP/Y003136/1 and in part by the State Key Laboratory of Power System Operation and Control under Project SKLD24KM21. *(Corresponding author: Zhongbei Tian.)*

Xiao Liu and Lin Jiang are with the Department of Electrical Engineering and Electronics, University of Liverpool, L69 7ZX Liverpool, U.K. (e-mail: xiao.liu@liverpool.ac.uk; ljiang@liverpool.ac.uk).

Zhongbei Tian is with the School of Engineering, University of Birmingham, B15 2TT Birmingham, U.K. (e-mail: z.tian@bham.ac.uk).

Yuan Gao is with the School of Engineering, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: yuan.gao@leicester.ac.uk).

Rob M. P. Goverde is with the Department of Transport and Planning, Delft University of Technology, 2600 AA Delft, The Netherlands (e-mail: r.m.p.goverde@tudelft.nl).

Digital Object Identifier 10.1109/TTE.2025.3575703

dc traction power supply system. Meanwhile, Zhang et al. [12] developed the operation section method to analyze the power flow of multiple trains and substations, thereby improving RBE usage efficiency by optimizing dwell times and headways. Both methods, however, struggle with nonlinear power supply components and cannot effectively manage undervoltage or overvoltage conditions in railway power networks. In order to address these issues, an iterative power flow analysis method has been introduced [13], offering a more accurate representation of dynamic power distribution and voltage fluctuations. Additionally, Pan et al. [14] extended this iterative framework by developing an integrated train speed profile and timetable optimization model for ac railway networks. In spite of these advancements, iterative power flow analysis remains computationally demanding, making it impractical for large-scale, real-time railway operations. Although specific times are not mentioned in these references, our preliminary experiments indicate that the computational demands are substantial. For instance, consider a metro line with two directions and 14 stations, supporting 20 running trains simultaneously. Approximately 40 h are required to solve this train schedule and speed profile optimization model.

Given the high computational cost of iterative power flow simulations, machine-learning (ML) techniques have emerged as a promising alternative for energy-efficient train control. Ning et al. [15] proposed a deep deterministic policy gradient approach for optimal train speed profiles, while Su et al. [16] employed a soft actor-critic method for energy-efficient train control. Extending these methods, an echo state neural network has been used for cooperative train control [17], and a multiagent cooperative actor-critic reinforcement learning strategy was introduced for multitrain scheduling [11]. While these ML applications have demonstrated efficiency improvements, they primarily focus on train movement optimization and largely ignore the complexities of railway power supply constraints and substation dynamics. A few studies have explored ML-based strategies for railway energy management, such as deep reinforcement learning for supercapacitor energy storage systems [18] and power quality regulation in railway traction power supply systems [19]. Existing ML research, however, does not integrate train dynamics, substation energy prediction, and railway power supply network modeling into a unified optimization framework.

Overall, while significant progress has been made in methods such as timetable adjustments and cooperative train control, existing approaches often suffer from two key limitations.

1) *Computational Inefficiency:* Iterative power flow simulations are time-consuming, making them impractical for real-time large-scale applications.
2) *Limited Integration of Train and Power Supply Models:* Most research focuses either on train control or power supply modeling, but not both simultaneously.

In order to overcome these challenges, we propose a novel data-driven framework that is the first to employ ML for predicting traction substation power demand in railway operations. Unlike previous ML methods that focus solely on train speed control or energy management, our approach integrates real-time substation power constraints, ensuring that train operations do not exceed power availability. This article makes the following key contributions.

1) *ML-Based Substation Power Prediction Model:* A neural network-based model is developed to emulate power flow simulations, reducing computation time to just 0.33% of traditional methods while achieving over 99% accuracy.
2) *Integrated Train-Power Supply Optimization:* By coupling the neural network-based model with GA, we optimize train speed profiles and dwell times to minimize substation energy supply under varying operational conditions.
3) *Real-World Validation:* The proposed method is evaluated using data from Beijing Yizhuang Metro line, demonstrating energy savings of up to 13% and flexibility under headway fluctuations.

This article outline is as follows. Section II introduces the fundamentals of train operation and substation energy calculations. Section III discusses the methodology of data-driven modeling and the optimization process using GA. Section IV presents case studies on the application of data-driven optimization in fixed and flexible headway scenarios. This article concludes with a summary of the findings in Section V.

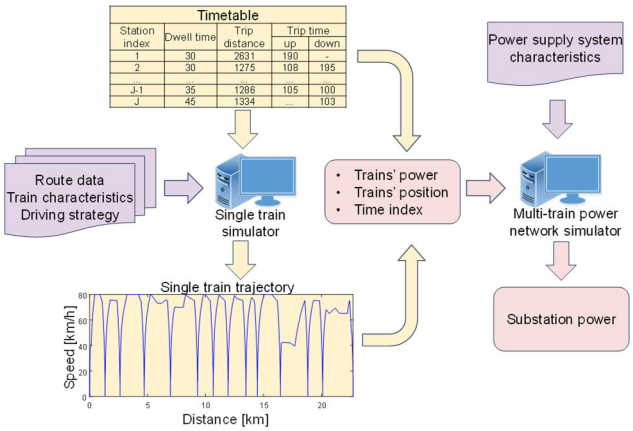

Fig. 1. Integration of train trajectory, timetable, and power supply system.

II. Fundamentals of Integrated Train Operation and Power Supply System

A. Architecture of Railway Energy System

The railway system is dynamic in real-time, i.e., the location and speed of a train are changing all the time, which leads to variations in the transmission resistance value and distribution. Additionally, train power levels may also vary sharply and fast. In order to assess the overall energy consumption of the power supply system, trains in the same line must all be taken into account. The operation of the multitrain power network will be introduced in Section II-C.

As illustrated in Fig. 1, train operation is shaped by factors such as route data, train characteristics, driving strategy, and the timetable. This information feeds into the single train model, determining the trajectory of a single train along the

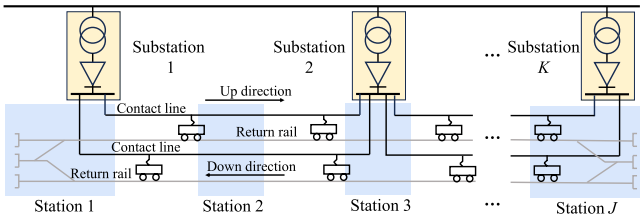

Fig. 2. Topology of a typical dc power network with multiple trains.

whole journey. Then, the single train trajectory, the dwell time, and headway in the timetable are used to generate the power and position of multiple trains at every time index. When this data is integrated into the multitrain power network model, the power supplied from all the substations can be derived. Here, we assume the power supplied by the substations can always meet the trains' requirements, since the substations are always designed according to the peak power [20]. In this research, the optimal train trajectory and dwell times that can minimize the substation energy for a given period will be identified.

B. Single Train Simulation

Assuming the train as a mass point, its movement can be expressed using the Newtonian equations [21]

$$(1+\lambda)M\frac{dv}{dt} = f(s) - Mg\tan(\theta(s)) - r(v) \tag{1}$$

$$dt = \frac{ds}{v} \tag{2}$$

where λ is the rotary allowance [22], M is the mass of the rolling stock, s is the position, v is the speed, t is the time, $f(s)$ is the tractive or braking force, g is the acceleration due to gravity, $\theta(s)$ is the angle of the route slope, $r(v) = A + Bv + C(v)^2$ is vehicle resistance where A, B, and C are Davis constants, which are determined by the attributes of the rolling stock [23].

The mechanical power of this train at position s can be represented by

$$p_{\text{mech}}(t) = f(s) \cdot v. \tag{3}$$

When the train is motoring, $p_{\text{mech}}(s)$ is positive, indicating the required power. Conversely, when the train is braking, $p_{\text{mech}}(s)$ is negative, converted into regenerative energy. Assuming the efficiency between mechanical power and electrical power is η, the electric power required by the train and the regenerative braking power derived by the train can be expressed as

$$p(t) = \begin{cases} \dfrac{p_{\text{mech}}(t)}{\eta}, & \text{if } p_{\text{mech}}(t) \geq 0 \\ p_{\text{mech}}(t) \cdot \eta & \text{if } p_{\text{mech}}(t) < 0. \end{cases} \tag{4}$$

C. Multitrain Power Network Simulation

A typical metro power network with multiple trains is depicted in Fig. 2. There are two parallel tracks designated for upward and downward operation, encompassing J stations and K rectifier substations. Stations 1 and J serve as the turnaround stations, allowing trains to switch directions. The train power supply system consists of contact lines and return rails. These lines form a closed circuit with trains running on them. Power is either transferred from substations to trains, or from trains using regenerative braking to those that are motoring.

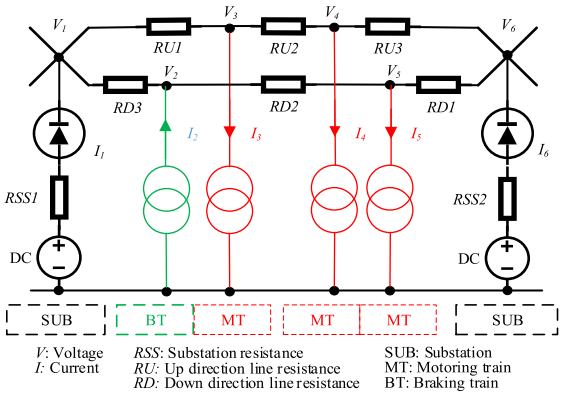

Fig. 3. DC traction power network equivalent circuit.

In order to analyze the power flow of the metro system, an equivalent circuit is constructed based on the topology of the power network and the distribution of trains. Fig. 3 shows the equivalent circuit for four trains and two substations. In this circuit, the substations are represented as ideal voltage sources, the motoring trains are modeled as loads, and the braking trains are modeled as power sources. Additionally, the contact lines and return rails are represented as resistive components to reflect their inherent electrical resistance characteristics. The movement of the trains affects the length of the transmission paths, which, in turn, changes the resistance values.

According to the Nodal Voltage Equation, $\boldsymbol{Y} \times \boldsymbol{V} = \boldsymbol{I}$, the relationship between the variables in Fig. 3 can be expressed with (5), as shown at the bottom of the next page. At each time t, the line resistance can be calculated according to the distance between the corresponding nodes. The power required or generated by the trains, denoted as $\boldsymbol{p}(t)$, is derived from the power profiles obtained through the simulation described in Section II-B. Our goal is to determine the current and voltage at each node so that we can calculate the power supplied by the substations.

In order to calculate the voltage and current of every node at time t, we employ an iterative method. First, $\boldsymbol{V}(t)$ is initialized with the no-load voltage, and the substation resistance, $\mathbf{RSS}(t)$, is set to a constant value. We define $\boldsymbol{V}_{\mathbf{tr}}(t)$ as the voltage vector of all trains. In the subsequent iteration, the train current, $\boldsymbol{I}_{\mathbf{tr}}(t)$, is calculated using (6). Similarly, $\boldsymbol{V}_{\mathbf{sub}}(t)$, representing the voltage of all the substations allows us to compute the substation current in the next iteration using (7). The current $\boldsymbol{I}_{\mathbf{tr}}(t)$ and $\boldsymbol{I}_{\mathbf{sub}}(t)$ are arranged in ascending order based on their physical positions along the track to form the total current $\boldsymbol{I}(t)$, as depicted in Fig. 3. Then, $\boldsymbol{I}(t)$ is used to derive the voltage of the same iteration as shown in (9). Equations (7)–(9) are repeated until $\boldsymbol{I}_{\mathbf{tr}}(t) \odot \boldsymbol{V}_{\mathbf{tr}}(t)$ of motoring trains is approximately equal to $\boldsymbol{p}(t)$. Here, $\odot$ is the

Hadamard product

$$\boldsymbol{V}^0(t) = \boldsymbol{V}_{\text{noload}}, \quad \mathbf{RSS}^0(t) = \boldsymbol{R}_{\text{intial}} \tag{6}$$

$$\boldsymbol{I}_{\mathbf{tr}}^{k+1}(t) = \boldsymbol{p}(t)/\boldsymbol{V}_{\mathbf{tr}}^{k}(t) \tag{7}$$

$$\boldsymbol{I}_{\mathbf{sub}}^{k+1}(t) = \boldsymbol{V}_{\mathbf{sub}}^{k}(t)/\mathbf{RSS}^{k}(t) \tag{8}$$

$$\boldsymbol{V}^{k+1}(t) = \boldsymbol{Y}^{-1}(t) \times \boldsymbol{I}^{k+1}(t). \tag{9}$$

Following this iteration process, the total effective regenerated braking power of the whole network, $P_{\text{reg}}(t)$, is calculated by:

$$P_{\text{reg}}(t) = \sum_{\boldsymbol{I}_{\mathbf{tr}}(t)<0} \boldsymbol{V}_{\mathbf{tr}}(t) \odot \boldsymbol{I}_{\mathbf{tr}}(t). \tag{10}$$

The efficiency of converting braking energy into RBE is expressed by

$$\eta_{\text{reg}} = \frac{\int_0^T P_{\text{reg}}(t)dt}{\int_0^T \left(\sum_{\boldsymbol{p}(t)<0} \boldsymbol{p}(t)\right)dt}. \tag{11}$$

The sum of all the substations' power and the energy supplied by the substations over a specific period T are given by

$$P_{\text{sub}}(t) = \sum \boldsymbol{V}_{\mathbf{sub}}(t) \odot \boldsymbol{I}_{\mathbf{sub}}(t) \tag{12}$$

$$E_{\text{sub}} = \int_0^T P_{\text{sub}}(t)dt. \tag{13}$$

III. Data-Driven Modeling and Optimization

The simulation models become complicated when integrating multiple trains into the power supply system. For a typical metro line, there are often more than ten substations, with roughly 20 trains operating in both directions simultaneously. In order to determine the power supplied by all the substations at any moment, it is necessary to analyze the power flow involving more than 30 components in the network. Such a simulation process results in significant computational demands. In order to address this issue, we propose a data-driven model to replace the multitrain power network simulator, thereby substantially reducing the computational cost of subsequent calculation, evaluation, and optimization. In this study, the term simulator refers to two in-house MATLAB-based tools developed by the authors.

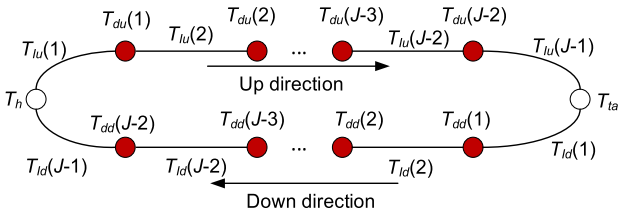

Fig. 4. Time distribution for a metro line.

A. Database Establishment

For a metro line with J stations, there will be $2(J-1)$ interstations for the up and down directions and $2(J-2)$ dwell times in the timetable. Assume the train starts from the station on the far left in Fig. 4 The time interval between the departures of two consecutive trains from this station is usually called the headway, T_h. Theoretically, T_h is fixed for a certain period. But it may fluctuate around the fixed values because of various unexpected disturbances. The time distribution for a single journey is illustrated in Fig. 4, where T_{du} and T_{dd} are the dwell times for the up and down directions, T_{Iu} and T_{Id} are the interstation running times in each direction. The station index in the up direction goes from left to right. At the station on the far right, the train will reverse direction. The duration that the train remains at this station is known as the turnaround time, denoted by T_{ta}. For the purposes of this research, T_{ta} is assumed to be a constant.

The power profiles are generated using the single-train motion simulator introduced in Section II-B. Here, we assume that the speed profile of the train at each interstation is composed of full power acceleration, cruising, coasting, and full braking. The cruising speed is the maximum allowable

$$\begin{bmatrix} -\frac{1}{RU1}-\frac{1}{RD3} & \frac{1}{RD3} & \frac{1}{RU1} & 0 & 0 & 0 \\ \frac{1}{RD3} & -\frac{1}{RD2}-\frac{1}{RD3} & 0 & 0 & \frac{1}{RD2} & 0 \\ \frac{1}{RU1} & 0 & -\frac{1}{RU1}-\frac{1}{RU2} & \frac{1}{RU2} & 0 & 0 \\ 0 & 0 & \frac{1}{RU2} & -\frac{1}{RU2}-\frac{1}{RU3} & 0 & \frac{1}{RU3} \\ 0 & \frac{1}{RD2} & 0 & 0 & -\frac{1}{RD2}-\frac{1}{RD1} & \frac{1}{RD1} \\ 0 & 0 & 0 & \frac{1}{RU3} & \frac{1}{RD1} & -\frac{1}{RU3}-\frac{1}{RD1} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \end{bmatrix} \tag{5}$$

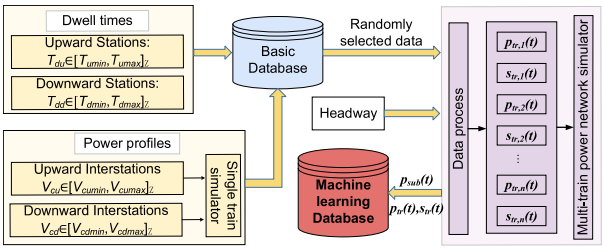

Fig. 5. Database establishment.

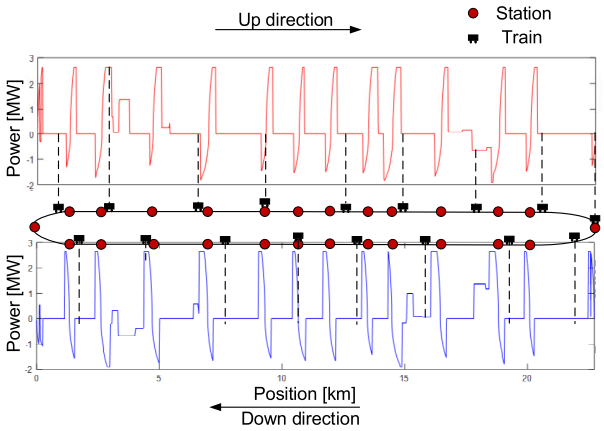

Fig. 6. Extension from single train to multitrain.

speed of this route; therefore, only by adjusting the coasting speed (the switching speed from coasting to braking) at each interstation, the power profiles for each interstation that can meet the time constraint can be obtained. These power profiles are subsequently stored in a database for future reference, as shown in Fig. 5. The coasting speed for an interstation in the up direction, V_{cu}, varies within an allowable range set by V_{cumin} and V_{cumax}, based on the permitted interstation running time. Similarly, the coasting speed in the down direction V_{cd} can range between V_{cdmin} and V_{cdmax}. The maximum and minimum dwell times at each station are determined according to the allowable range based on the prescribed times in the timetable.

After all the required data are saved in the basic database, we randomly select indexes corresponding to the power profiles and dwell times for a single train's entire route. The relevant data are then retrieved from the database. Assuming all the trains follow the same speed profiles, similar to other research on train trajectory and timetable optimization [14], [24], [25], the data obtained above can be combined with the headway information for the data process. This data process is explained in Fig. 6, where the power and position of each train at every given moment is pinpointed. After this data process, the power and position of all the trains at time t are input into the multitrain power supply simulator to obtain the substation power. The database for ML is built by executing the multitrain power network simulator repeatedly for different inputs. This database must be rebuilt for different railway networks. The data collection process, however, remains the same as described above.

The selection of train power and position as input features was based on their fundamental role in determining train dynamics and the overall power flow within the railway system. Train power inherently encapsulates both train

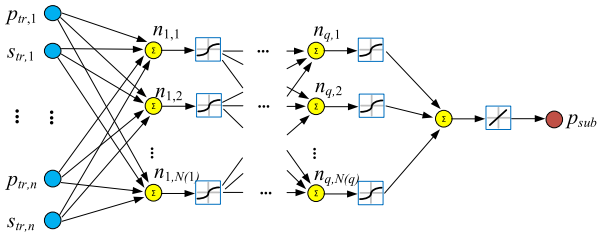

Fig. 7. Structure of the neural network.

acceleration and regenerative braking efficiency, as power consumption increases during acceleration and decreases (or is fed back into the grid) during braking. Explicit inclusion of acceleration and braking efficiency as separate features was, therefore, deemed unnecessary. Additionally, train position plays a crucial role in defining the topology of the railway power network. The relative positions of trains affect how power is distributed across substations and overhead lines, impacting system-wide energy flow. By combining train power and position, we can effectively model power demand and energy distribution throughout the network, making these two features sufficient for predicting substation power consumption.

B. Algorithms for ML

Neural networks, which are inspired by the structure of animal brains, have been widely used in railway research. Value regression based on neural networks has been used to predict the urban rail transit safety performance [26], predict track geometry degradation [27], model the energy consumption of electric metro trains [28], and more. These regression models based on a neural network have achieved good results. In this research, the neural network is, therefore, employed to train the regression model from the position and power of each train at a certain time to the substation power. Then, the substation energy can be calculated according to the substation power.

We consider the substation power regression model as transient, meaning it does not require the time variable t for $p_{tr,n}(t)$ and $s_{tr,n}(t)$. Consequently, both the power p_{tr} and position s_{tr}, as well as the count of trains n, can vary over time. The regression model can then be formulated as

$$\tilde{p}_{\mathrm{sub}} = f_p\left(p_{tr,1}, s_{tr,1}, p_{tr,2}, s_{tr,2}, \ldots p_{tr,n}, s_{tr,n}\right). \tag{14}$$

In order to approximate this function, we employ a feed-forward multilayer perceptron (MLP), as shown in Fig. 7. The MLP was chosen for its ability to effectively model complex nonlinear relationships while maintaining computational efficiency, which is crucial for large-scale simulations. The network comprises three main components.

1) Input Layer: This layer takes a vector x, formed by concatenating the power and position of each train at a given time step

$$x = \left[p_{tr,1}, s_{tr,1}, p_{tr,2}, s_{tr,2}, \ldots p_{tr,n}, s_{tr,n}\right]'. \tag{15}$$

2) Hidden Layers: The MLP contains q fully connected hidden layers. Each layer i includes $N(i)$ neurons. The output

of each hidden layer is computed by

$$y_i = \sigma_i\left(\theta_i . y^*_{i-1}\right) \tag{16}$$

where the parameter set $\theta_i = [W_i, b_i]$ consists of weights and biases, and the input vector is augmented as $y^*_{i-1} = [y_{i-1}, 1]^T$. For the first hidden layer, the input is defined as $y^*_0 = [x, 1]^T$.

3) Output Layer: The final output $\tilde{p}_{\text{sub}}$ is generated using an activation function

$$\tilde{p}_{\text{sub}} = \sigma_q\left(\theta_q \cdot y^*_{q-1}\right). \tag{17}$$

In order to train the network, we apply a mean-square error (MSE) loss function between the predicted and simulated substation power values

$$l_{\text{loss}} = \frac{1}{\text{NT}} \sum_{j=1}^{\text{NT}} \left(\tilde{p}_{\text{sub},j} - p_{\text{sub},j}\right)^2 \tag{18}$$

where NT is the total number of training samples, and $\tilde{p}_{\text{sub},j}$ and $p_{\text{sub},j}$ denote the predicted and simulated substation power values for the jth sample, respectively. The simulated power values are obtained from the simulation system described in Sections II-B and II-C.

C. Substation Energy Optimization Based on GA

Minimizing the substation energy, presented as f_E in (19) is the main objective of this research. The inputs of this function are the indexes of power profiles of every interstation and the dwell time at each station. These indexes are divided into four parts. The first part, $I_1 \sim I_{s1}$, includes the speed profile indexes of the up-direction interstations. The second part, $I_{s1+1} \sim I_{s2}$, are the dwell time indexes of the up-direction stations. $I_{s2+1} \sim I_{s3}$ are the speed profile indexes of the down-direction interstations. $I_{s3+1} \sim I_{s4}$ are the dwell time indexes of the down-direction stations. The indexes are positive integers constrained within the maximum number of indexes, as presented in (20)–(23). In order to avoid achieving energy savings simply by increasing the interstation running time (IRT), a penalty term PN has, however, been added. The objective function is presented in (19), where PN is calculated in (24) and (25). The threshold, denoted as PN_{th}, is established by calculating the total IRT in the timetable. For each individual in the GA, the total IRT, denoted as T_{ri}, is derived by the sum of $T(I)$, where $T(I)$ represents the IRT associated with the power profile index I. Meanwhile, the penalty coefficient, PN_{co}, is established through experimental trials, considering the magnitude of substation energy

$$\text{Min. } f_E\left(\begin{matrix} I_1, \ldots, I_{s1}, I_{s1+1}, \ldots, I_{s2}, \\ I_{s2+1}, \ldots, I_{s3}, I_{s3+1}, \ldots, I_{s4} \end{matrix}\right) + \text{PN} \tag{19}$$

$$\text{s.t. } 1 \le I_1, \ldots, I_{s1} \le IS1^{\max} \tag{20}$$

$$1 \le I_{s1+1}, \ldots, I_{s2} \le IS2^{\max} \tag{21}$$

$$1 \le I_{s2+1}, \ldots, I_{s3} \le IS3^{\max} \tag{22}$$

$$1 \le I_{s3+1}, \ldots, I_{s4} \le IS4^{\max} \tag{23}$$

$$\text{PN} = \begin{cases} (\text{PN}_{\text{th}} - T_{ri}) \cdot \text{PN}_{co}, & \text{if } \text{PN}_{\text{th}} - T_{ri} > 0 \\ 0, & \text{else} \end{cases} \tag{24}$$

$$T_{ri} = T(I_1) + \cdots + T(I_{s1}) + T(I_{s2+1}) + \cdots + T(I_{s3}). \tag{25}$$

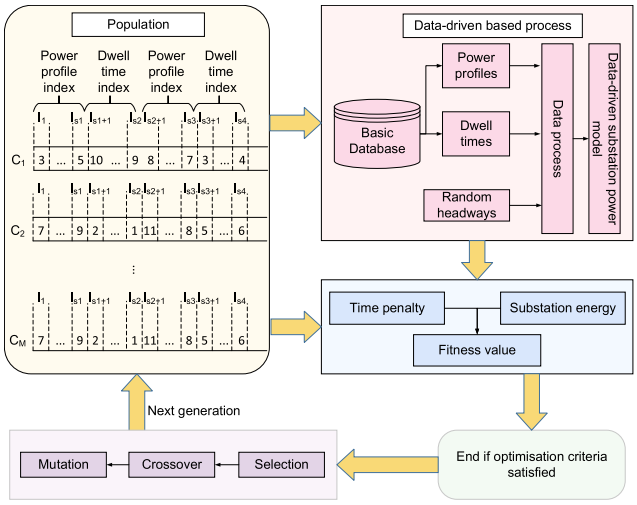

Fig. 8. Substation energy optimization with GA.

Based on the decision variables and the objective function, the GA is employed to obtain the minimum substation energy supply. Assume that there are M chromosomes in the initial population, the schematic GA optimization procedure is shown in Fig. 8. The indexes, $I_1 \sim I_{s4}$, are the genes of each chromosome. The substation energy f_E is calculated by the data-driven based process, where the power profiles and dwell times are obtained from the database according to the indexes. For every iteration, several sets of headways are randomly generated, with the number of sets equal to the number of individuals. Each set of headways assigns unique intervals to successive train pairs. These headways are uniformly distributed, and their range is determined before the optimization process to meet the specific needs of the research. Then, the data process is carried out to obtain the position and power for each train at each time, which would be the input of the data-driven substation power model obtained in Section III-B, and then derive the substation power. The substation energy can then be calculated with the substation powers. After the fitness value is calculated by using the substation energy and the IRT penalty, the optimization end requirement will be checked. If this is not the last generation, selection, crossover, and mutation will be applied to generate the next generation.

IV. Case Study

A. Training of Data-Driven Model

The data of Beijing Yizhuang metro line is employed to illustrate the performance of the data-driven method. The length of this metro line is 22.73 km, containing 14 stations and 12 rectifier substations. The maximum speed of the line is 80 km/h. Fig. 9 shows the distribution of stations and substations. The parameters of the traction system are presented in Table I. In this research, the passenger mass and the auxiliary power are assumed to be 0. These assumptions are made to investigate the energy-saving performance of the power supply system more effectively, isolating it from the influence of passenger flow and other environmental factors.

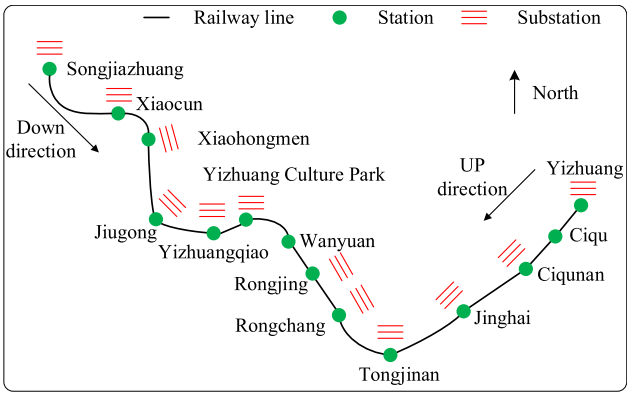

Fig. 9. Beijing Yizhuang metro line.

TABLE I
PARAMETERS OF THE TRACTION SYSTEM [24]

Category	Parameters	Values
Electrical parameters	Substation no-load voltage [V]	850
	Substation source resistance [Ω]	0.02
	Overvoltage limitation [V]	950
	Power supply type	DC
Mechanical parameters	Maximum service acceleration [m/s^2]	0.8
	Maximum service deceleration [m/s^2]	0.55
	Maximum tractive effort [kN]	160
	Maximum braking effort [kN]	160
	Train mass [tonnes]	199
	Maximum traction power [MW]	2.65
	Maximum braking power [MW]	2.65
Operational parameters	Maximum operation speed [km/h]	80

TABLE II
ALLOWABLE RANGE BASED ON THE CURRENT TIMETABLE

	Interstation time	Dwell time	headway
Permissible range	± 5 s	± 5 s	± 20 s

Similar assumptions are also present in other studies focusing on energy-efficient timetable optimization [9], [29], [30].

The database in this research is built based on the current timetable of the Beijing Yizhuang metro line, with the allowable ranges presented in Table II. Given that the fixed headway is 254 s, the actual headway can vary randomly within ±20 s.

The computer used for ML is equipped with an Intel Core i5-10210U processor, which has a base clock speed of 1.60 GHz and a boost speed of 2.11 GHz. It includes 16 GB of installed RAM (15.8 GB usable) and operates on a 64-bit system with an x64-based processor. The neural network incorporates two hidden layers, each equipped with ten neurons. The activation function used for these two hidden layers is the hyperbolic tangent sigmoid (tansig), which facilitates nonlinear transformations. For the output layer, the linear (purelin) activation function is selected, enabling the network to produce outputs across a continuous range. This configuration allows the model to effectively capture the dynamics of train movements and power levels.

During training, we employed the Levenberg–Marquardt algorithm (trainlm) with full-batch updates—meaning the entire training set was processed as a single batch each epoch—and set the number of epochs to 150. A total of 1 275 000 training samples were collected from the simulator. Before splitting, we used MATLAB's built-in random permutation function to shuffle the data, ensuring a random selection of subsets. We then adopted a holdout validation approach: the dataset was randomly divided into three parts, with 70% used for training, 15% for validation, and 15% reserved for testing. The training set was used to optimize the model parameters, while the test set was used to evaluate the model's generalization capability. We monitored the data distribution to confirm that the feature ranges were appropriate for the neural network. The model's performance was assessed using the MSE, as shown in Fig. 10. After about 50 epochs, the MSE of the substation power converged below 10^{-2}.

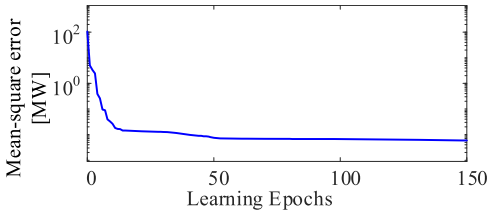

Fig. 10. MSE convergence over learning epochs (training time: ∼100 min).

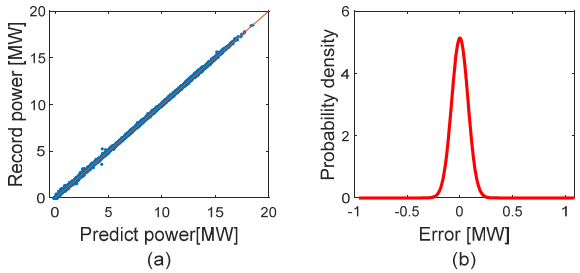

Fig. 11. Predicted substation power. (a) Predicted power versus recorded power. (b) Probability density of the predicted error.

B. Case 1 (fixed headway)

1) Accuracy Test: For a fixed headway of 254 s, the performance of the ML model can be seen in Fig. 11. For 50 000 sets of randomly selected speed profiles and dwell times, the predicted substation power with the data-driven model and the corresponding recorded power from the simulator is illustrated in Fig. 11(a). Each point on the plot represents a pair of predicted and corresponding recorded power values. It is clear that all the points cluster around the line $y = x$. The correlation coefficient between the predicted and recorded power can reach 0.999, indicating a strong agreement. Additionally, to analyze the prediction error—defined as the difference between predicted and recorded substation power—we generate 1000 evenly spaced points between the minimum and maximum values of this error. The probability density of these errors, modeled using a fit normal distribution, is calculated at each point. These calculations are represented in Fig. 11(b). From this figure, we can see that the probability density around 0 is the highest. Nearly all predicted values fall within an error margin of ±0.025 MW. While there are outliers

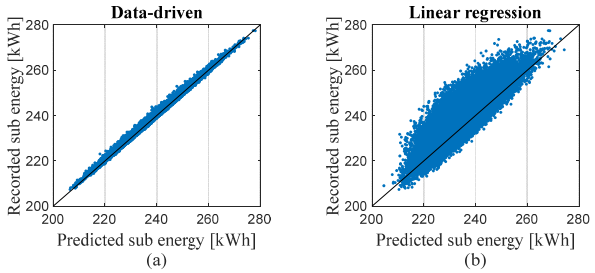

Fig. 12. Predicted substation energy compared with recorded energy. (a) Data-driven. (b) Linear regression [24].

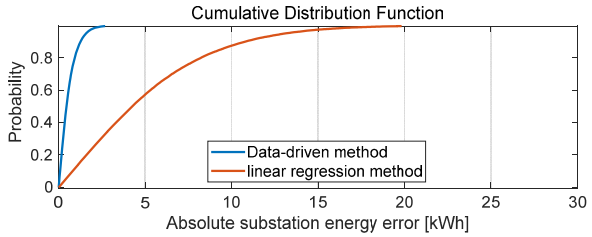

Fig. 13. Predicted substation power error cumulative distribution.

TABLE III
TIME CONSUMPTION COMPARISON

	Data-driven method	Linear regression method	Simulator
Time [s]	4233	180	1,250,000

Intel Core i5-10210U processor, 16GB RAM, and Windows 10 operating system

with errors ranging from [−0.94, −0.025] to [0.025, 1.09]; their frequency is minimal and can be overlooked.

The proposed data-driven model is used to calculate the energy consumption during a headway period and is then compared with the linear regression method. Fig. 12 shows the predicted power versus recorded power by the data-driven method and linear regression method in [24], respectively. In Fig. 12(a), the points all concentrate near the line of equality. In contrast, the points are more dispersed in (b). The correlation coefficient between recorded substation energy and predicted substation energy can reach 0.998 by using the data-driven method, which is much higher than that of 0.8615 by the linear regression method.

The cumulative error of the predicted substation energy with the data-driven method and with the regression method are compared in Fig. 13. The data-driven approach nearly guarantees an absolute error below 2.5 kWh, with a likelihood approaching 100%. In contrast, the linear regression method exhibits roughly a 95% probability of keeping the absolute error below 10 kWh.

2) Calculation Time and Optimization Results: The computational time required by the traditional simulation method, linear regression method, and data-driven method for 500 000 headway periods is detailed in Table III. All three methods are run on the same laptop independently. Because of the simulator's excessive time consumption, we adopt the average computation time of 2.5 s for each headway period as a basis for estimating the total time requirement. In contrast, the time required by the linear regression model and the data-driven model is significantly less. Notably, the linear regression model demands the least amount of time, while the data-driven model's time requirement is approximately 0.33% of that of the simulator.

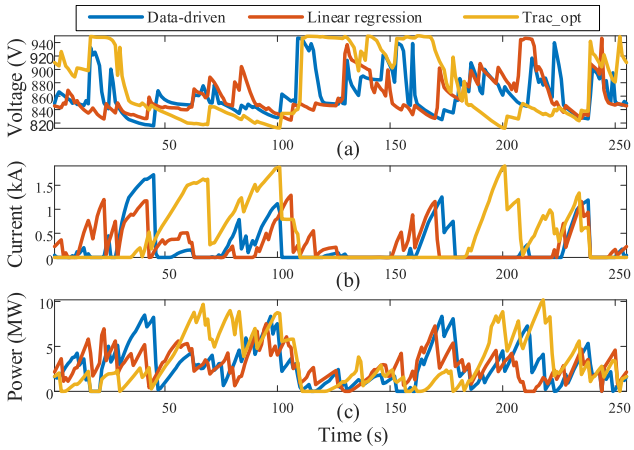

Fig. 14. (a) Voltage of the Rongjing substation. (b) Current of the Rongjing substation. (c) Power of all the substations.

Because of the enormous computational time required, the simulation method is impractical for optimizing speed profiles and dwell times. For the two alternative methods, their results are compared with those of the traction optimization method (Trac_opt) [21]. Trac_opt is a method that does not account for the power supply system or RBE. Instead, it minimizes energy consumption solely by optimizing train speed profiles to reduce traction energy.

In order to assess the electrical impact of these methods, we input the optimized speed profile and dwell time indices obtained from each approach into the power network simulator. The simulator then produced the voltage, current, and power curves shown in Fig. 14. Since the Beijing Yizhuang metro line has 12 substations, displaying all of them would make the figure overly complex and difficult to interpret. Only the voltage and current of the Rongjing substation are, therefore, shown to highlight the differences among the optimization results. Additionally, the total power of all the substations is summed at each time step for comparison. This figure illustrates that the peak voltage of the data-driven and linear regression methods is lower than that of the Trac_opt method [Fig. 14(a)]. Additionally, Trac_opt exhibits higher current peaks than the other two methods [Fig. 14(b)]. Further calculations show that both the data-driven and linear regression approaches reduce substation energy consumption and increase RBE by approximately 13% compared to Trac_opt [Fig. 14(c)].

In order to further examine train-level power dynamics, we use the optimized speed and dwell time indices from the data-driven method and simulate the power profiles of individual trains during a headway period. A total of 17 trains operated in this period, and their power consumption is shown in Fig. 15. For clarity, trains are separated by direction, and gray dashed lines mark station positions (S1–S14) to support analysis of interstation behavior. This representation

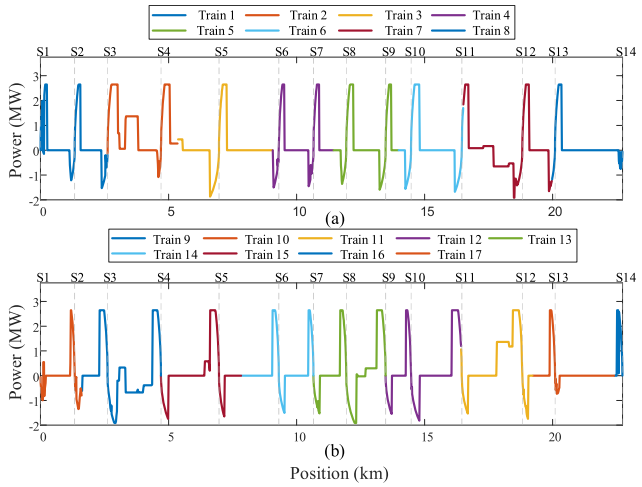

Fig. 15. Train power profiles along the track during a headway period. (a) Up direction. (b) Down direction.

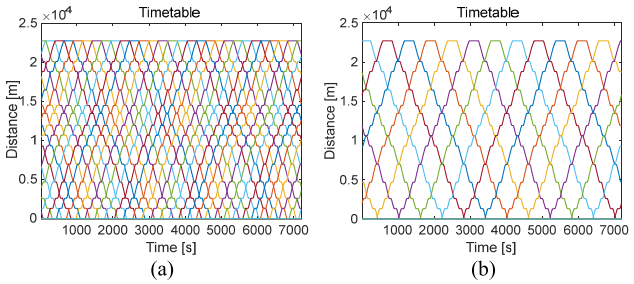

Fig. 16. Timetable with different headway basements. (a) Headway basement as 254 s. (b) Headway basement as 600 s.

provides both spatial and temporal insights into energy usage at the individual train level, complementing the system-wide assessment shown in Fig. 14.

C. Case 2 (flexible headway)

1) Headway Fluctuation Setting: For the Beijing Yizhuang metro line, the headways at the peak time on weekdays and weekends are about 254 and 600 s, respectively. In this case study, the energy-saving performance of DD_GA and Trac_opt are compared with headway fluctuating within [234, 274 s] and [580, 620 s]. The substation energy supply and the regenerative braking efficiency during two hours are calculated. The timetables during this period with headways randomly fluctuating within ±20 s are illustrated in Fig. 16. In order to evaluate the results with different headway settings, we must standardize the substation energy supply measurement. Using the timetables in Fig. 16, we estimate the number of complete journeys within a 2-h period, where a complete journey entails a single train traversing from the start station through both directions and returning to the start. These counts enable us to calculate the average energy supply per train for a journey. Specifically, Fig. 16(a) reveals 28 total complete journeys at 254 s headway, while Fig. 16(b) shows 12 at 600 s headway.

2) Headway Fluctuates Around 254 s: For the Beijing Yizhuang metro line, the headway at the peak time during weekdays is about 254 s. A headway time between 234 and 274 s for every two consecutive trains is, therefore, randomly generated. The predicted power with the data-driven model and

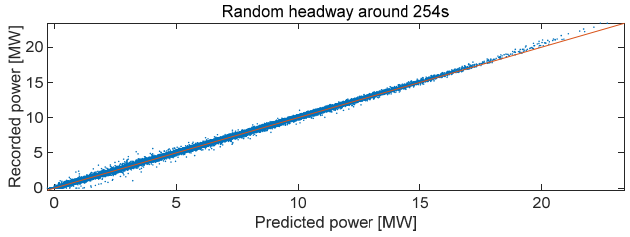

Fig. 17. Predicted substation power with random headways around 254 s.

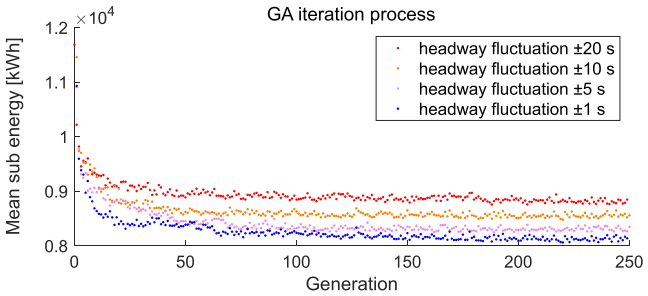

Fig. 18. Iteration process of GA with headway fluctuation around 254 s.

the recorded power by the simulator are shown in Fig. 17. The correlation coefficient between recorded substation power and predicted substation power can reach 0.993.

During the optimization process, the GA is configured with a single-point crossover at a rate of 0.7, complemented by a mutation rate set to 0.02 for each variable within an individual, and the selection of individuals is facilitated through the roulette wheel selection method. The maximum iteration number is 250. For the case when the headway randomly changes within ±20, ±10, ±5, and ±1 s, the iteration processes are shown in Fig. 18. The "mean subenergy" represents the average energy supply from substations, calculated across all individuals over 2-h intervals for each generation. For each headway fluctuation scope, the iterations do not converge on a specific value. This is because the objective is to minimize the mean substation energy rather than minimizing the substation energy of one particular individual in order to address the random headway of each individual in every generation. This figure, however, still shows a clear tendency of convergence. Additionally, when the headway fluctuation is constrained within 1 s, the convergence value is the lowest.

The optimization results of DD_GA are assessed against Trac_opt. Our analysis compares DD_GA with Trac_opt over two hours, focusing on substation energy and regenerative braking across random headways. In order to mitigate the impact of randomness on the outcomes, all results are obtained by averaging the calculations from ten sets of randomly generated headways. Considering that there are approximately 28 complete journeys during two hours, the results for one single train over a complete journey are then calculated.

The key findings are summarized in Table IV. Because of the introduction of PN in Section IV-C, the IRT of DD_GA consistently remains less than or equal to that of Trac_opt regardless of the headway changes. This makes sure that the energy saving of DD_GA is not caused by extending the

TABLE IV
ENERGY SAVING PERFORMANCE WITH HEADWAY FLUCTUATING AROUND 254 s

Headway fluctuation		±20 s	±10 s	±5 s	±1 s
Total *IRT* [s]	Trac_opt	3271	3271	3271	3271
	DD_GA	3271	3270	3271	3271
Substation energy [kWh]	Trac_opt	327.9	325.5	325.8	324.7
	DD_GA	322.8	307.3	303.2	300.0
	Difference	**-5.1**	**-18.2**	**-22.6**	**-24.7**
Traction energy [kWh]	Trac_opt	442.4	442.1	442.7	443.0
	DD_GA	444.3	440.3	440.8	441.8
	Difference	1.9	-1.8	-1.9	-1.2
Regenerative braking energy [kWh]	Trac_opt	145.7	147.9	148.5	149.4
	DD_GA	152.5	160.9	164.9	168.6
	Difference	**6.8**	**13.0**	**16.4**	**19.1**
Substation loss [kWh]	Trac_opt	11.4	11.2	11.3	10.8
	DD_GA	11.0	9.3	8.8	8.6
	Difference	-0.4	-1.8	-2.5	-2.2
Transmission loss [kWh]	Trac_opt	19.8	20.1	20.3	20.4
	DD_GA	20.0	18.6	18.5	18.2
	Difference	0.2	-1.5	-1.8	-2.2

running times. The impact of headway fluctuation on traction energy, substation loss, and transmission loss is quite small. With traction optimization, the changes in these three variables do not exceed 1 kWh. The differences between these variables in the DD_GA results and the corresponding traction optimization remain within 2.5 kWh for a single train over a complete journey. In contrast, when headway fluctuations are maintained within different ranges, the substation energy and RBE values can change significantly. Specifically, implementing stricter control over the variance in train headway can improve the energy efficiency of optimized train speed profiles and dwell times. When the headway fluctuations are constrained to ±1 s, DD_GA results in the most significant energy savings. Compared to the Trac_opt strategy, it can reduce substation energy by 24.7 kWh and increase RBE by 19.1 kWh.

The regenerative braking efficiency of Trac_opt and DD_GA over different headway fluctuations is illustrated in Fig. 19. The headway set index represents the index of ten random headways varying within the corresponding range. Data series 1–4 correspond to the optimization results of DD_GA, with the headway randomly varying within ±20, ±10, ±5, and ±1 s, respectively. Meanwhile, data series 5–9 correspond to the optimization results of Trac_opt, with the headway randomly varying within ±20, ±10, ±5, and ±1 s, respectively. Fig. 19 shows that DD_GA's regenerative efficiency, η_{reg}, consistently outperforms Trac_opt with headway fluctuations of 10, 5, and 1 s. With headway fluctuations of 20 s, DD_GA's η_{reg}, however, aligns closely with Trac_opt's results, indicating negligible energy savings for the optimal speed profile within this fluctuation range.

3) Headway Fluctuates Around 600 s: Since the headway at the peak time during weekends is about 600 s for Beijing Yizhuang metro line, the headway fluctuation between 580 and 620 s is studied. By employing a random generation of headway times within this interval for every pair of consecutive trains, we analyze the substation power prediction performance of the data-driven model. The outcomes of this

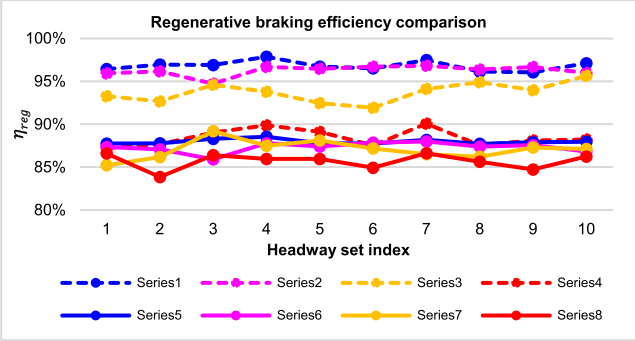

Fig. 19. Regenerative braking rate for each set of headway.

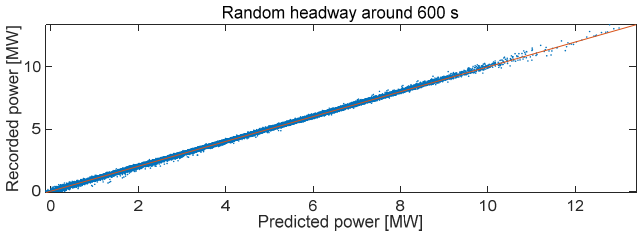

Fig. 20. Predicted substation power with random headways around 600 s.

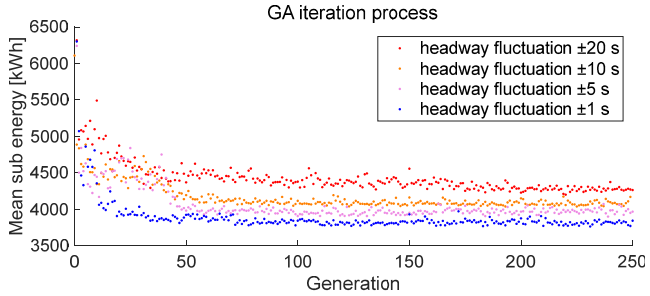

Fig. 21. Iteration process of GA with headway fluctuation around 600 s.

analysis are detailed in Fig. 20. The correlation coefficient between recorded substation power and predicted substation power can reach 0.998.

The parameter setting of the GA is the same as that in Section II). For the case when the headway randomly changes within ±20, ±10, ±5, and ±1 s, the iteration processes are shown in Fig. 21. Although the degree of dispersion of the points and the convergence values corresponding to different headway fluctuations are different, the convergence trends are the same as in Fig. 18.

The energy-saving performance of DD_GA and Trac_opt are compared in Table V. Similar to the results in Table IV, the IRT of DD_GA consistently remains less than or equal to that of Trac_opt regardless of the headway changes. Substation energy and RBE, moreover, vary significantly with different headway fluctuation constraints. Implementing stricter control over train headway fluctuations can enhance the energy efficiency of DD_GA's optimization results. Among the four headway fluctuation magnitudes, DD_GA saves the most energy when the headway fluctuations are limited to ±1 s. Specifically, it reduces the substation energy by 41.2 kWh

TABLE V
ENERGY SAVING PERFORMANCE WITH HEADWAY FLUCTUATES AROUND 600 s

Headway fluctuation		±20 s	±10 s	±5 s	±1 s
Total *IRT* [s]	Trac_opt	3271	3271	3271	3271
	DD_GA	3271	3269	3271	3271
Substation energy [kWh]	Trac_opt	361.3	354.9	353.3	355.6
	DD_GA	357.5	343.1	331.1	314.5
	Difference	**-3.8**	**-11.8**	**-22.2**	**-41.2**
Traction energy [kWh]	Trac_opt	436.8	436.7	436.8	436.9
	DD_GA	439.7	438.6	438.4	437.2
	Difference	2.9	1.9	1.6	0.3
Regenerative braking energy [kWh]	Trac_opt	104.3	110.9	113.2	110.9
	DD_GA	110.4	124.5	135.9	151.3
	Difference	**6.1**	**13.6**	**22.7**	**40.3**
Substation loss [kWh]	Trac_opt	9.6	9.4	9.6	9.6
	DD_GA	9.2	9.0	8.2	7.6
	Difference	-0.4	-0.4	-1.4	-2.1
Transmission loss [kWh]	Trac_opt	19.2	19.7	20.1	20.0
	DD_GA	19.1	20.1	20.4	20.9
	Difference	-0.2	0.4	0.3	0.9

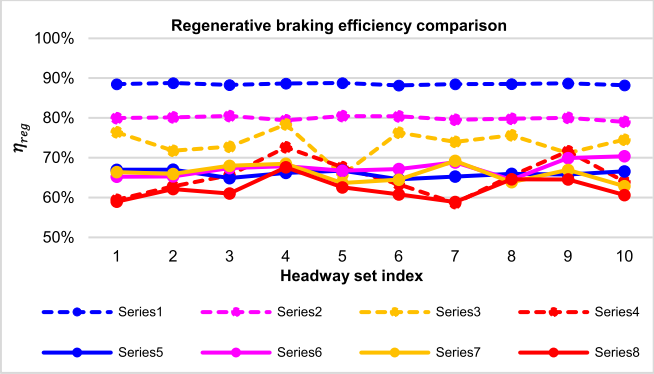

Fig. 22. Regenerative braking rate for each set of headway.

and increases the RBE by 40.3 kWh. Comparing Table V with Table IV, we can find that the substation energy supply for one single train over the whole journey has increased, which means that longer headway increases the average energy consumption for a single train. Additionally, when the headway is strictly controlled to less than 1 s, a baseline headway of 600 s achieves maximum energy savings that are almost twice as high as those with a 254-s headway. This indicates that longer headways, when tightly managed, offer substantially greater energy savings than shorter ones.

Consistent with the results above, when headway fluctuates by 1 s, DD_GA's regenerative braking efficiency remains stable at around 90%. This efficiency is about 10% higher than when the headway fluctuation is within 5 s, and significantly exceeds all results from Trac_opt as illustrated in Fig. 22. As the headway fluctuation range widens, however, DD_GA's η_{reg} becomes more erratic, as illustrated by the orange dashed line. Similar to Fig. 19, DD_GA's η_{reg} lies within that of Trac_opt when the headway fluctuation magnitude is 20 s. Comparing Figs. 19 and 22, it becomes evident that imposing stricter constraints on headway results in significantly higher energy savings, particularly when the baseline headway is longer.

V. CONCLUSION

This article proposes a data-driven substation power prediction model to simplify the power flow simulation of the railway power network and reduce the computational demands. In this research, a neural network is used to develop the data-driven model, which maps train position and power to substation power. This model is then integrated with GA to minimize substation energy. We conducted a comparative assessment of its time requirements and accuracy under both fixed and flexible headways. The results show that the calculation time of the data-driven model is only 0.33% of the power flow simulation while keeping the accuracy above 99%. For fixed headway, the energy supplied by the substations can be reduced by 13% compared to Trac_opt. The case study on headway fluctuating around 254 and 600 s shows that the greater the fluctuation in headway, the poorer the energy savings of the optimization results. When the fluctuation reaches 20 s, it is not possible to find a set of speed profiles and dwell times that achieve energy savings for all headways.

While the proposed model effectively predicts substation power and reduces computation time, it assumes stable operations with limited headway fluctuations. This may limit its performance in scenarios involving major delays, extended dwell times, or significant passenger flow variations. The current model also lacks real-time adaptability. Future work will focus on integrating real-time data and adaptive algorithms, such as reinforcement learning, to enhance its robustness and energy efficiency under dynamic operating conditions.

REFERENCES

[1] N. Ahsan, K. Hewage, F. Razi, S. A. Hussain, and R. Sadiq, "A critical review of sustainable rail technologies based on environmental, economic, social, and technical perspectives to achieve net zero emissions," *Renew. Sustain. Energy Rev.*, vol. 185, Oct. 2023, Art. no. 113621.

[2] C. Wu, W. Ochieng, K.-C. Pien, and W.-L. Shang, "Carbon-efficient timetable optimization for urban railway systems considering wind power consumption," *Appl. Energy*, vol. 388, Jun. 2025, Art. no. 125593.

[3] G. Meng, C. Wu, B. Zhang, F. Xue, and S. Lu, "Net hydrogen consumption minimization of fuel cell hybrid trains using a time-based co-optimization model," *Energies*, vol. 15, no. 8, p. 2891, Apr. 2022.

[4] G. M. Scheepmaker and R. M. P. Goverde, "Energy-efficient train control using nonlinear bounded regenerative braking," *Transp. Res. C, Emerg. Technol.*, vol. 121, Dec. 2020, Art. no. 102852.

[5] X. Yang, X. Li, Z. Gao, H. Wang, and T. Tang, "A cooperative scheduling model for timetable optimization in subway systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 438–447, Mar. 2013.

[6] X. Wang, T. Tang, S. Su, J. Yin, Z. Gao, and N. Lv, "An integrated energy-efficient train operation approach based on the space-time-speed network methodology," *Transp. Res. E, Logistics Transp. Rev.*, vol. 150, Jun. 2021, Art. no. 102323.

[7] C. Wu, S. Lu, Z. Tian, F. Xue, and L. Jiang, "Energy-efficient train control with onboard energy storage systems considering stochastic regenerative braking energy," *IEEE Trans. Transport. Electrific.*, vol. 11, no. 1, pp. 257–274, Feb. 2025, doi: 10.1109/TTE.2024.3389960.

[8] M. Peña-Alcaraz, A. Fernández, A. P. Cucala, A. Ramos, and R. R. Pecharromán, "Optimal underground timetable design based on power flow for maximizing the use of regenerative-braking energy," *Proc. Inst. Mech. Eng. F, J. Rail Rapid Transit*, vol. 226, no. 4, pp. 397–408, Jul. 2012.

[9] S. Su, X. Wang, Y. Cao, and J. Yin, "An energy-efficient train operation approach by integrating the metro timetabling and eco-driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4252–4268, Oct. 2020.

[10] J. Chen et al., "Power flow control-based regenerative braking energy utilization in AC electrified railways: Review and future trends," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 7, pp. 6345–6365, Jul. 2024.

[11] M. Chen, Q. Wang, P. Sun, and X. Feng, "Train control and schedule integrated optimization with reversible substations," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 1586–1600, Feb. 2023.

[12] Z. Zhang, H. Zhao, X. Yao, Z. Xing, and X. Liu, "Metro timetable optimization for improving regenerative braking energy utilization efficiency," *J. Cleaner Prod.*, vol. 450, Apr. 2024, Art. no. 141970.

[13] X. Tao, Q. Wang, M. Chen, P. Sun, and X. Feng, "Comprehensive optimization of energy consumption and network voltage stability of freight multitrain operation based on amMOPSO algorithm," *IEEE Trans. Transport. Electrific.*, vol. 10, no. 2, pp. 4074–4094, Jun. 2024.

[14] Z. Pan, M. Chen, S. Lu, Z. Tian, and Y. Liu, "Integrated timetable optimization for minimum total energy consumption of an AC railway system," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3641–3653, Apr. 2020.

[15] L. Ning, M. Zhou, Z. Hou, R. M. P. Goverde, F. Wang, and H. Dong, "Deep deterministic policy gradient for high-speed train trajectory optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11562–11574, Aug. 2022.

[16] S. Su, Q. Zhu, J. Liu, T. Tang, Q. Wei, and Y. Cao, "A data-driven iterative learning approach for optimizing the train control strategy," *IEEE Trans. Ind. Informat.*, vol. 19, no. 7, pp. 7885–7893, Jul. 2023.

[17] H. Liu, L. Yang, and H. Yang, "Cooperative optimal control of the following operation of high-speed trains," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 17744–17755, Oct. 2022, doi: 10.1109/TITS.2022.3163971.

[18] Z. Yang, F. Zhu, and F. Lin, "Deep-reinforcement-learning-based energy management strategy for supercapacitor energy storage systems in urban rail transit," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1150–1160, Feb. 2021, doi: 10.1109/TITS.2019.2963785.

[19] C. Xing, K. Li, and J. Su, "Thermal constrained energy optimization of railway cophase systems with ESS integration—An FRA-pruned DQN approach," *IEEE Trans. Transport. Electrific.*, vol. 9, no. 4, pp. 5122–5139, Dec. 2023, doi: 10.1109/TTE.2022.3218762.

[20] Q. Qin, T. Guo, F. Lin, and Z. Yang, "Energy transfer strategy for urban rail transit battery energy storage system to reduce peak power of traction substation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 11714–11724, Dec. 2019.

[21] Z. Tian, N. Zhao, S. Hillmansen, C. Roberts, T. Dowens, and C. Kerr, "SmartDrive: Traction energy optimization and applications in rail systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 7, pp. 2764–2773, Jul. 2019.

[22] C. Goodman, "Overview of electric railway systems and the calculation of train performance," in *Proc. IET Prof. Develop. Course Electr. Traction Syst.*, Jun. 2008, pp. 1–24.

[23] B. P. Rochard and F. Schmid, "A review of methods to measure and calculate train resistances," *Proc. Inst. Mech. Eng., F, J. Rail Rapid Transit*, vol. 214, no. 4, pp. 185–199, Jul. 2000.

[24] Z. Tian, P. Weston, N. Zhao, S. Hillmansen, C. Roberts, and L. Chen, "System energy optimisation strategies for metros with regeneration," *Transp. Res. C, Emerg. Technol.*, vol. 75, pp. 120–135, Feb. 2017.

[25] S. Yang, J. Wu, X. Yang, H. Sun, and Z. Gao, "Energy-efficient timetable and speed profile optimization with multi-phase speed limits: Theoretical analysis and application," *Appl. Math. Model.*, vol. 56, pp. 32–50, Apr. 2018.

[26] F. A. Awad, D. J. Graham, R. Singh, and L. AitBihiOuali, "Predicting urban rail transit safety via artificial neural networks," *Saf. Sci.*, vol. 167, Nov. 2023, Art. no. 106282.

[27] H. Khajehei, A. Ahmadi, I. Soleimanmeigouni, M. Haddadzade, A. Nissen, and M. J. L. Jebelli, "Prediction of track geometry degradation using artificial neural network: A case study," *Int. J. Rail Transp.*, vol. 10, no. 1, pp. 24–43, Jan. 2022.

[28] P. M. Fernández, P. S. Zuriaga, I. V. Sanchís, and R. I. Franco, "Neural networks for modelling the energy consumption of metro trains," *Proc. Inst. Mech. Eng., F, J. Rail Rapid Transit*, vol. 234, no. 7, pp. 722–733, Aug. 2020.

[29] Y. Wang, S. Zhu, A. D'Ariano, J. Yin, J. Miao, and L. Meng, "Energy-efficient timetabling and rolling stock circulation planning based on automatic train operation levels for metro lines," *Transp. Res. C, Emerg. Technol.*, vol. 129, Aug. 2021, Art. no. 103209.

[30] J. Liao, F. Zhang, S. Zhang, G. Yang, and C. Gong, "Energy-saving optimization strategy of multi-train metro timetable based on dual decision variables: A case study of Shanghai metro line one," *J. Rail Transp. Planning Manage.*, vol. 17, Mar. 2021, Art. no. 100234.

Xiao Liu received the B.S. and M.S. degrees from the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China, in 2011 and 2014, respectively. She is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K.

After spending six years as a Railway Signaling Engineer with China Railway Eryuan Engineering Group, University of Liverpool, her research interests include railway traction power system modeling, railway traction system modeling, energy-efficient train control, energy system optimization, and automatic train protection systems.

Zhongbei Tian (Member, IEEE) received the B.Eng. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2013, and the B.Eng. and Ph.D. degrees in electrical and electronic engineering from the University of Birmingham, Birmingham, U.K., in 2013 and 2017, respectively.

He is currently an Assistant Professor in transport energy systems with the Department of Electronic Electrical and Systems Engineering, University of Birmingham. His research interests include railway traction power system modeling and analysis, energy-efficient train control, energy system optimization, and sustainable transport energy systems integration and management.

Yuan Gao (Member, IEEE) received the B.S. degree in engineering from Dalian Maritime University, Dalian, China, in 2013, the M.S. degree in aeronautical engineering from Beihang University, Beijing, China, in 2017, and the Ph.D. degree in electrical and electronic engineering from the University of Nottingham, Nottingham, U.K., in 2021.

He is currently a Lecturer with the Electrical and Electronic Engineering, University of Leicester, Leicester, U.K. Prior to this, he was a Post-Doctoral Research Associate in Hybrid Autonomous Systems at the University of Bristol, Bristol, U.K. His research interests include motor drives, model predictive control, multiagent systems, autonomous systems, and machine learning-based design and control.

Lin Jiang (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1992 and 1996, respectively, and the Ph.D. degree in electrical engineering from the University of Liverpool, Liverpool, U.K., in 2001.

He is currently a Reader with the Department of Electrical Engineering and Electronics, University of Liverpool. His research interests include control and analysis of smart grids, renewable energy, and power electronics.

Rob M. P. Goverde (Member, IEEE) received the M.Sc. degree in mathematics from Utrecht University, Utrecht, The Netherlands, in 1993, the Engineering Doctorate (Eng.D.) degree in mathematical modeling and decision support from Delft University of Technology, Delft, The Netherlands, in 1996, and the Ph.D. degree in railway transport from TU Delft, Delft, in 2005.

He is currently a Professor of railway traffic management and operations and the Director of the Digital Rail Traffic Lab, TU Delft. His research interests include railway timetabling and traffic management, automatic train operation, train control, and railway signaling.