

Data-driven rapid prediction model for aerodynamic force of high-speed train with arbitrary streamlined head

Dawei Chen, Zhenxu Sun, Shuanbao Yao, Shengfeng Xu, Bo Yin, Dilong Guo, Guowei Yang & Sansan Ding

To cite this article: Dawei Chen, Zhenxu Sun, Shuanbao Yao, Shengfeng Xu, Bo Yin, Dilong Guo, Guowei Yang & Sansan Ding (2022) Data-driven rapid prediction model for aerodynamic force of high-speed train with arbitrary streamlined head, Engineering Applications of Computational Fluid Mechanics, 16:1, 2191-2206, DOI: [10.1080/19942060.2022.2136758](https://doi.org/10.1080/19942060.2022.2136758)

To link to this article: <https://doi.org/10.1080/19942060.2022.2136758>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 16 Nov 2022.



Submit your article to this journal [↗](#)



Article views: 328



View related articles [↗](#)



View Crossmark data [↗](#)

Data-driven rapid prediction model for aerodynamic force of high-speed train with arbitrary streamlined head

Dawei Chen^a, Zhenxu Sun^b, Shuanbao Yao^a, Shengfeng Xu^b, Bo Yin^b, Dilong Guo^b, Guowei Yang^b and Sansan Ding^a

^aCRRC Qingdao Sifang Co. Ltd., Qingdao, People's Republic of China; ^bKey Laboratory for Mechanics in Fluid Solid Coupling Systems, Institute of Mechanics, Chinese Academy of Sciences, Beijing, People's Republic of China

ABSTRACT

Due to the complicated geometric shape, it's difficult to precisely obtain the aerodynamic force of high-speed trains. Taking numerical and experimental data as the training data, the present work proposed a data-driven rapid prediction model to solve this problem, which utilized the Support Vector Machine (SVM) model to construct a nonlinear implicit mapping between design variables and aerodynamic forces of high-speed train. Within this framework, it is a key issue to achieve the consistency and auto-extraction of design variables for any given streamlined shape. A general parameterization method for the streamlined shape which adopted the idea of step-by-step modeling has been proposed. Taking aerodynamic drag as the prediction objective, the effectiveness of the model was verified. The results show that the proposed model can be successfully used for performance evaluation of high-speed trains. Keeping a comparable prediction accuracy with numerical simulations, the efficiency of the rapid prediction model can be improved by more than 90%. With the enrichment of data for the training set, the prediction accuracy of the rapid prediction model can be continuously improved. Current study provides a new approach for aerodynamic evaluation of high-speed trains and can be beneficial to corresponding engineering design departments.

ARTICLE HISTORY

Received 15 August 2022
Accepted 8 October 2022

KEYWORDS

Aerodynamic force; inverse design; high-speed train; SVM; numerical simulation; wind tunnel test

1. Introduction

As a near-ground rail transport tool with large slender ratio, the high-speed train usually experiences complex three-dimensional turbulent flow at high Reynolds number (Baker, 2010; Raghunathan et al., 2002; Schetz, 2001). The complex geometric shapes of key exposed components such as pantographs, bogies, and windshields have a great impact on the aerodynamic performance of trains. During the engineering design process, it's unbearable to perform high-fidelity numerical simulations for practical high-speed trains with multiple carriages. Therefore, most researches on the flow field characteristics of high-speed trains commonly consider simplified shapes (Hemida & Baker, 2010), for instance, ignoring the influence of bogies, pantographs and windshields, reducing the number of carriages (Wang et al., 2008), and scaling down the size of the model. Hemida and Baker (2010) adopted the Large Eddy Simulation (LES) method to investigate the influence of the streamlined shape and the deflection angle of crosswind on the flow field structures around the train for an extremely simplified scaled

model. Wang et al. (2008) utilized the Lattice Boltzmann Method (LBM) to analyze the flow characteristics of the simplified three-carriage high-speed train model. Yao et al. (2013) adopted the DES method to investigate the wake characteristics of a simplified high-speed train model with three carriages. Yang et al. (2012) adopted the RANS method to evaluate the aerodynamic drag of the simplified eight-carriage high-speed train model. Sima et al. (2008) validated the accuracy of numerical approaches with the wind tunnel experimental data from a scaled bogie model. Wang et al. (2017) adopted a simplified two-carriage high-speed train model to compare and analyze the simulation accuracy of different turbulence models. Catanzaro et al. (2010) adopted simplified models and wind tunnel test data to verify the accuracy of numerical calculations.

Although the main flow characteristics of the flow field around the high-speed train can be obtained through simplified models, the impacts of key components like bogies on the aerodynamic loads such as aerodynamic drag, aerodynamic lift, and aerodynamic noise of the

CONTACT Zhenxu Sun  sunzhenxu@imech.ac.cn

train, which are critical indicators for the shape design of high-speed train, cannot be accurately obtained. The shape smoothing design of high-speed train mainly aims at key components such as bogies, pantographs and inter-connection parts (Yang et al., 2012). A reasonable smoothing design will greatly reduce the aerodynamic drag and aerodynamic noise of the high-speed train, which can significantly improve the environmental adaptability of the train. Meanwhile, strong nonlinear relationship exists between the smoothing design of key components and factors such as train marshaling and streamlined shape, which requires investigation of the influences caused by various factors on the aerodynamic performance of trains during the engineering design process. Therefore, with the rapid development of high-speed train aerodynamic design technology and massively parallel computing technology, aerodynamic prediction methods aiming at the real-train scale, real carriages and complex model will become the new trend for train aerodynamics.

The reduced-order model can be adopted to investigate the coherent vortex structures of the wake field behind the bluff body at high Reynolds number (Östth et al., 2014). Muld et al. (2012) adopted different reduced-order models to investigate the flow characteristics of the wake field of high-speed train. However, literature using reduced-order model to predict aerodynamic force of high-speed train has not been found. With the continuous deepening of machine-learning applications, different types of machine-learning models are gradually adopted to predict aerodynamic force (Wang et al., 2015; Zhang et al., 2019; Zhu & Wang, 2019), so as to reduce the computational cost. In the field of aerodynamic shape optimization design, machine-learning models are more widely used as surrogate models. Shuanbao et al. (2014) and Sun et al. (2020) applied different machine-learning methods as surrogate models to the aerodynamic drag reduction design of high-speed train heads. Ku et al. (2010) and Yang et al. (2022) carried out the multi-objective aerodynamic optimization design of the streamlined shape of high-speed train with the use of Kriging model. Munoz-Paniagua and García (2020) adopted feedforward neural networks to develop single-objective aerodynamic optimization designs for the streamlined shape of high-speed train. The existing methods of predicting aerodynamic force applying machine-learning models can establish the nonlinear mapping between design variables and aerodynamic force. However, they mainly base on determined geometric design variables. Without determined design variables, prediction of aerodynamic force seems to be impossible.

Streamlined head of high-speed train usually contains complex three-dimensional surfaces and owns various topological structures. To rapidly predict the aerodynamic force of high-speed train based on machine-learning methods, it is necessary to first propose a parametrization method to extract design variables. Using block-by-block modeling method, Munoz-Paniagua and García (2020) implemented the three-dimensional parametric design of the high-speed train head shape with quadratic Bezier curves. Shuanbao et al. (2014) and Sun et al. (2020) have implemented the three-dimensional parametric design for the new streamlined shape and the existing streamlined shape of the high-speed train by adopting vehicle modeling function (VMF) method and local shape function method, respectively. Yang et al. (2022) achieved three-dimensional surface deformation of high-speed train by adopting free form deformation (FFD) method. However, these parametrization methods usually lack of the generalization capability. That's to say, they cannot be adopted to generate any given streamlined shape.

Aiming at precisely and efficiently predicting aerodynamic force of high-speed train with arbitrary streamlined shape, a data-driven rapid prediction model was proposed in this study, which could make full use of existing data from numerical simulations and wind tunnel tests. With use of this model, a nonlinear implicit mapping between design variables and aerodynamic forces of high-speed train could be constructed. More importantly, within this framework, it is a key issue to achieve the consistency and auto-extraction of design variables for any given streamlined shape. A general parameterization method for the streamlined shape which adopted the idea of step-by-step modeling has been proposed, so that the rapid extraction of the values of design variables could be realized.

The remainder of the paper is organized as follows: the general parametrization method and the inverse design of the streamlined shape are exhibited in Section 2. Combining these two methods, any arbitrary streamlined head could be represented by a series of design variables, which work as the input for the rapid prediction model. Section 3 gives out the detailed construction process of the rapid prediction model, which includes the approaches concerning the training data acquisition, the design space of the design variables, and the utilization of SVM model. Taking the aerodynamic drag coefficient as the prediction objective, results and discussion with regard to the rapid prediction model is carried out in Section 4. Finally, Section 5 concludes the research.

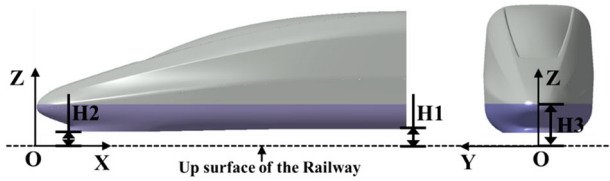


Figure 1. Definition of coordinate system.

2. Rapid extraction of design variables of streamlined head

As mentioned above, for any given streamlined head of high-speed train, it is of primary significance to obtain its design variables, so that they can be taken as the input for the rapid prediction model. It is required to sufficiently widen the design space so as to take as many streamlined shapes as possible into consideration. As a result, a more generalized parametrization method for the streamlined head has been proposed firstly. It is worth mentioning that we mainly focus on the parametrization of the streamlined head in present study, since the streamlined head shape of the leading and trailing car takes a major contribution in aerodynamic forces and the design of a brand new streamlined shape is the most conspicuous part when designing a new high-speed train. Once the parametrization method has been constructed, the inverse design idea has been adopted together with optimization method so that the design variables of any given streamlined shape can be approximated.

2.1. Parametrization methods for streamlined head

2.1.1. Location of key control profiles

The streamlined shape is a three dimensional free-form surface, and the parametric expression of the key control profiles is determined by the definition of the coordinate system, as shown in Figure 1. The x-, y- and z-axes are the length, width and height directions, respectively. It should be mentioned that the origin of the coordinate locates at the tip of the nose for x-axis, at the half width of the nose for y-axis and on the top of the rail surface for z-axis. The distance between the bottom of the train head and rail surface is H1, and the distance between the bottom of the cowcatcher and rail surface is H2. Taking the height H3 above the rail surface as the dividing point, the train head can be divided into upper and lower parts, in which parametric design are conducted separately. In order to facilitate the parametric design, the original streamlined shape is normalized along the x direction, so that the length of the streamlined part is 1 m, while the y direction and z direction are scaled proportionally.

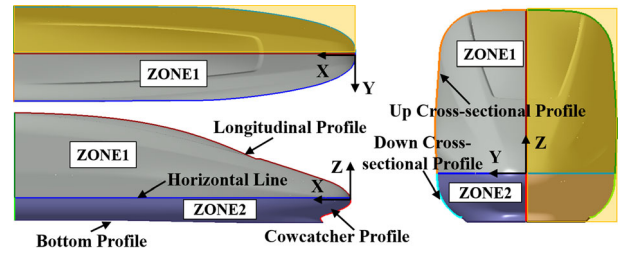


Figure 2. Key control profiles for the streamlined head.

The streamlined shape of the high-speed train is symmetrically distributed along the y direction, thus in order to reduce the complexity of parametrization, current study only carries out the parametrization design on half of the head, as shown in Figure 2. The topological structure of the head shape is determined by key control profiles. The key control profiles include cross-sectional profiles in zone 1 and zone 2, longitudinal profile, horizontal profile, bottom profile, and cowcatcher profile. The position of each profile is shown in Figure 2.

2.1.2. Definition of key control profiles

(I) cross-sectional profile. Any kind of profiles can be obtained by the non-uniform rational B-spline method by properly setting the number, coordinates, and corresponding weights of the control points. It is one of the most commonly used methods for parametric design of geometric shapes. The key control profiles of high-speed train head are rich and changeable. In order to fit the high-speed train head more accurately, NURBS curve method is chosen for parametric design of different types of control profiles.

The rational polynomial expression of a NURBS curve of degree k is:

$$p(u) = \frac{\sum_{i=0}^n w_i d_i N_{i,k}(u)}{\sum_{i=0}^n w_i N_{i,k}(u)} \quad (1)$$

where w_i is the weight factor, d_i is the coordinate vector of the control vertex, and the basis function $N_{i,k}(u)$ is determined by the recursive formulas (2) and (3):

$$N_{i,0}(u) = \begin{cases} 1, & u \in [u_i, u_{i+1}] \\ 0, & u \notin [u_i, u_{i+1}] \end{cases} \quad (2)$$

$$N_{i,k}(u) = \frac{u - u_i}{u_{i+k} - u_i} N_{i,k-1}(u) + \frac{u_{i+k+1} - u}{u_{i+k+1} - u_{i+1}} N_{i+1,k-1}(u) \quad (3)$$

where u_i is the coordinate of the node, which is related to the corresponding control vertex.

The cross-sectional profiles of the train body are generally composed of multiple straight lines and arcs, but

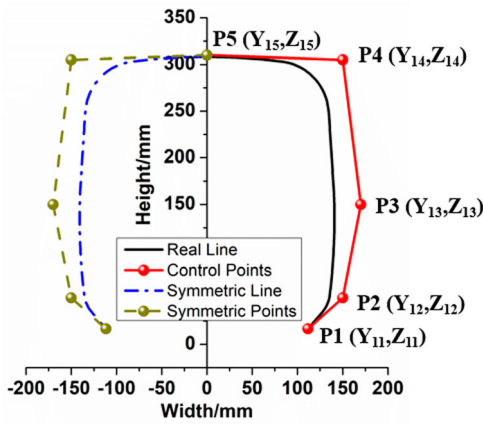
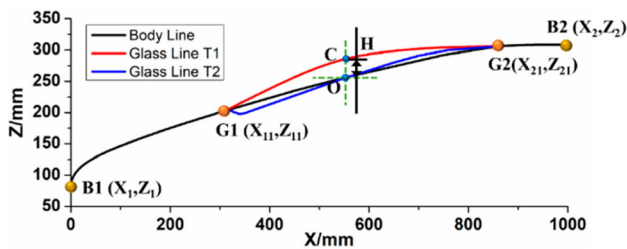


Figure 3. Control points of the cross-sectional profile of the train body.

for different models, combinations and sizes vary greatly. The NURBS curve is used to fit the cross-sectional profile of the train body, as shown in Figure 3. To ensure the continuous curvature at the center point of the curve, a symmetrical arrangement of control points is employed. Five control points P1 ~ P5 are arranged on one side of the train. Taking the z-axis as the axis of symmetry, P1 ~ P4 are symmetrized to the non-parametric side. The z coordinate of P1 is fixed, the y coordinate of P5 is 0, while the other coordinate parameters are design variables, with a total number of 8.

(II) Longitudinal profile. The longitudinal profile of the streamlined head includes the main profile and additional window profile, as shown in Figure 4. For different head shapes, the window profile can be quite different, as shown in Figure 4(a) with the convex shape and the front end downward chamfered shape. To ensure the consistency of the parametric design expression, the NURBS curve is employed to fit the longitudinal line, as shown in Figure 4(b), with a total of five control parameters. The x coordinate of P1 is 0, the z coordinate of P1 is H3, the z coordinate of P5 is the same as the z coordinate of P5



(a) Type of glass line

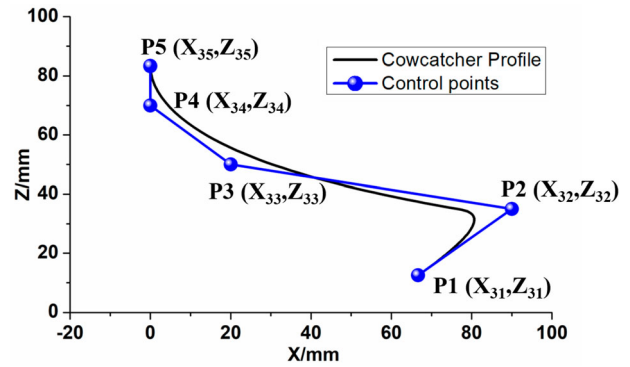
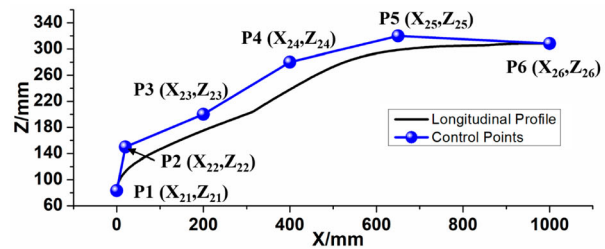


Figure 5. Control points of the longitudinal profile of the cowcatcher.

of the train body cross-sectional profile, and the coordinates of other control points are design variables, that is, 9 design variables in total.

(III) Profile of the Cowcatcher. The cowcatcher of high-speed train usually owns complicated shape, and has a great influence on the flow characteristics beneath the streamlined head. As stated previously, the NURBS curve was adopted to fit the longitudinal profile of the cowcatcher, as shown in Figure 5, five control points are adopted, in which P5 is the same point as P1 of the longitudinal profile, and the coordinates of the other four control points are design variables, with a total number of 8.

(IV) Horizontal profile. The horizontal profile of high-speed train streamlined head is the dividing curve between zone 1 and zone 2. To reduce the number of design variables during parametric design, the height H3 is defined as a fixed value. To ensure the shape of longitudinal beam structure of the train body is fixed, a straight line is positioned in the area close to the straight section of the train body, and the remaining part is a free curve, as shown in Figure 6. Here we did not carry out parametric



(b) Control points of longitudinal profile

Figure 4. Schematic diagram of parameterization of longitudinal line.

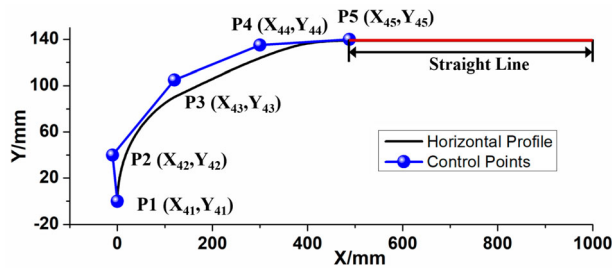


Figure 6. Control points of horizontal profile.

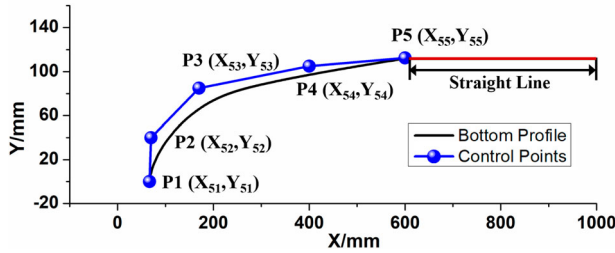


Figure 7. Control points of bottom profile.

design for the straight-line section but did adopt NURBS curve to fit the free curve. There are totally 5 control points, of which P1 is the same point as P1 of the longitudinal line, and the y -coordinate of P5 is determined by the cross-sectional profile. The coordinates of the remaining control points are design variables, that is, seven design variables in total.

(V) bottom profile. The bottom profile of the head shape is similar to the horizontal profile. Similarly, parametrical design was not carried out for the straight-line section and free curve was fit by the NURBS curve. There are totally five control points, of which P1 is the same point as P1 of the cowcatcher profile, the y -coordinate of P5 is determined by the cross-sectional profile, and the coordinates of the remaining control points are design variables, that is, seven design variables in total, as shown in Figure 7. Note that affected by the function of the cowcatcher, the z -coordinate of P1 is H2, the z -coordinate of P5 is H1, and the z -coordinate of the free curve is obtained by linear interpolation of H1 and H2.

2.1.3. Interpolation of the spatial surface

The surface of the streamlined head is a three-dimensional free-form surface. It was discretized into spatial grids, and the surface shape was fitted by coordinate interpolation of spatial grids. Polynomial functions were utilized to interpolate for the spatial curved surface. When dealing with connection places where two surfaces meet, quadratic polynomial functions are

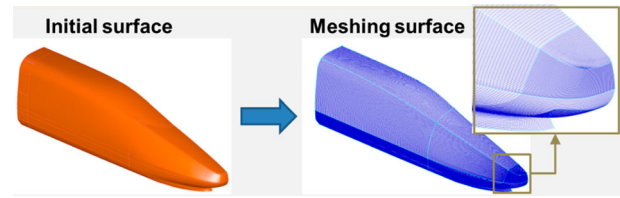


Figure 8. Discrete grid of spatial surface.

usually adopted so that the curvature could be kept the same at the connection parts, which could meet the smooth transition requirement for engineering design. The discrete surface is shown in Figure 8. The grid points are ordered structural surface grids.

To reduce the number of design parameters, this paper took the key control profiles as the boundaries, and adopted different surface fitting methods for different areas according to the area division in Figure 6. Taking x -coordinate of the surface point the same as that of horizontal profile, the interpolation formulas for the y coordinate and z coordinate take the forms as:

$$y(x) = y_{min}(x) + y_{max}(x) \frac{i-1}{n-1} \Delta y \quad (4)$$

$$z(x, y) = z_{min}(x, y) + z_{max}(x, y) \times \left(\frac{y - y_{min}(x)}{y_{max}(x) - y_{min}(x)} \right)^m \quad (5)$$

In formula (4), i is the serial number of the i^{th} point, n is the number of discrete points on the train body section curve, Δy is the increment of the y -coordinate between adjacent points. Quadratic parabolic interpolation method was adopted in ZONE1 to control the continuity of the curvature at the $y = 0$ symmetry plane, where the corresponding value of m in formula (5) is 2. Linear interpolation method was adopted in ZONE2, where $m = 1$.

The parametric design is the key to the inverse design of the streamlined head. It should be emphasizing that the inverse design accuracy of the streamlined shape relies heavily on careful specification of the parametric design method. Poor parametric method could lead to significant reduction of the prediction accuracy of the aerodynamic performance of high-speed train. The parametric method proposed in present study takes into account the characteristics of different topological structures of the head types to the greatest extent. The key design variables for the parametric design of the high-speed train streamlined shape are given in Table 1, with a total number of 39.

Table 1. Key design variables.

Curve type	Design variables								
	Y_{11}	Y_{12}	Z_{12}	Y_{13}	Z_{13}	Y_{14}	Z_{14}	Z_{15}	
Cross-sectional profile	Y_{11}	Y_{12}	Z_{12}	Y_{13}	Z_{13}	Y_{14}	Z_{14}	Z_{15}	
Longitudinal profile	X_{22}	Z_{22}	X_{23}	Z_{23}	X_{24}	Z_{24}	X_{25}	Z_{25}	X_{26}
Cowcatcher profile	X_{31}	Z_{31}	X_{32}	Z_{32}	X_{33}	Z_{33}	X_{34}	Z_{34}	
Horizontal profile	X_{42}	Y_{42}	X_{43}	Y_{43}	X_{44}	Y_{44}	X_{45}		
Bottom profile	X_{52}	Y_{52}	X_{53}	Y_{53}	X_{54}	Y_{54}	X_{55}		

2.2. Inverse design of the streamlined head

2.2.1. Inverse design process

The inverse design of the head shape of high-speed train is to obtain the values of the design variables according to the three-dimensional geometric data of the existing head shape, then input the values into the head shape parametric design model to implement the reconstruction of the three-dimensional shape. The specific process for the inverse design of a streamlined head is shown in Figure 9:

- 1) Determine the streamlined head that needs to be inversely designed, and perform grid discretization on it.
- 2) Use the self-developed data-processing code to automatically obtain the discrete data of each profile according to their location characteristics.
- 3) Adopt PSO algorithm (Kennedy & Eberhart, 1995) to optimize and obtain the optimal value of each design variable for the key control profiles by taking minimizing the average fitting error as the optimization goal.

As seen in Figure 9, the single-objective PSO algorithm was adopted herein for the inverse design of the streamlined head. The specific coefficients for PSO are listed as follows: the population of particle swarm is 200, the total number of iterations is 500, the value of acceleration factor is 2, the inertia factor gradually changes from 1.2 to 0.8 as the number of iterations increases, and the maximum flight speed of the particles: the value for the cowcatcher is 2 and the value for the other key control profiles is 5.

The inverse design objective of each control profile is the average error between the inverse design profile and the target profile, and the objective function is shown below:

$$f_r = \frac{1}{n} \sum_{i=1}^n d_i \quad (6)$$

Where f_r is the value of the objective function in inverse design, n is the number of discrete points of the control profile, and d_i is the minimum distance between each discrete point and the target line.

2.2.2. Validation of the inverse design process

Two aspects should be carefully paid attention to when dealing with the inverse design of a three-dimensional geometry, which are the inverse design of the two-dimensional key profiles and the final three-dimensional shape. In this section, four different types of streamlined shape models, as shown in Figure 10, are employed to illustrate the effect of current inverse design method, which are named TEST1, TEST2, TEST3 and TEST4, respectively. They shared the same cross-section shape of the carriage body except TEST2. Meanwhile, the main window profile of the driver's cab for TEST2 was a downward chamfered shape at the front end.

It should be mentioned that these four streamlined heads are also under investigation in wind tunnel tests and numerical validations, which will be discussed later.

(I) inverse design of two-dimensional profiles. The accuracy of the inverse design is mainly determined by the fitting accuracy of key control profiles. The topological structure of the cross-sectional profile, the longitudinal profile, and the cowcatcher profile are completely different. Different streamlined heads usually own quite different topological structures for the longitudinal profiles and the cowcatcher profiles. To verify the accuracy of inverse design for two-dimensional profiles, inverse design for different types of control profiles has been conducted. The five key profiles, cross-sectional profile, longitudinal profile, the cowcatcher profile, horizontal profile and bottom profile, are respectively named L1, L2, L3, L4 and L5 for simplicity. The convergence curve of the fitness of the control profiles with the number of iterations is shown in Figure 11. It can be observed that after 500 iterations, the fitness of each control profile converges, indicating that PSO can rapidly obtain the values of design variables for each profile. More specifically, the average error of the inverse design of L2 is the largest, and the average errors of the inverse design of L3, L4 and L5 are basically the same, but they are significantly smaller than the errors of L1 and L2. Thus, it can be concluded that sudden change in the curvature of the profile could propose a challenge to the fitting by NURBS curve method. The sharper the curvature of the profile changes, the lower the fitting accuracy becomes.

The inverse design results of the key profiles of TEST1 and TEST3 are presented in Figure 12. The topological structures of the key profiles of two streamlined shapes differ notably. Obviously, the current method is capable of accurately fitting profiles with completely different topological structures, and only a certain error exists in the locations where the curvature changes sharply. To reduce computational cost during the construction of aerodynamic prediction model for high-speed trains,

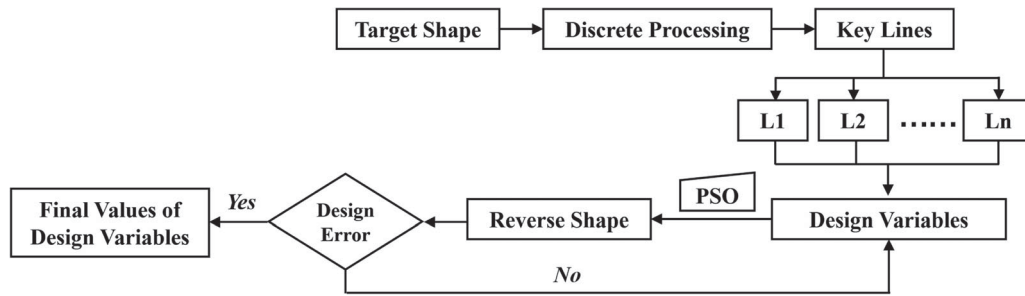


Figure 9. Flow chart for the inverse design of the streamlined head.

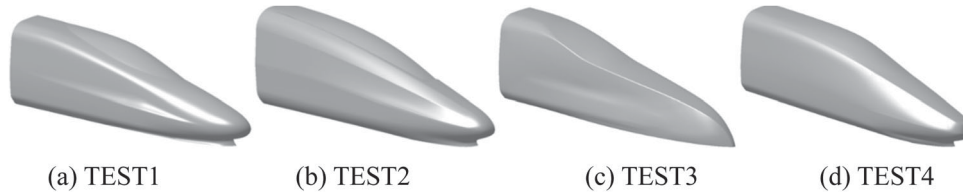


Figure 10. Different head types for wind tunnel test.

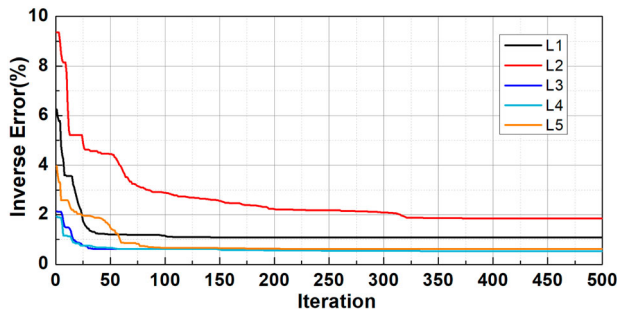


Figure 11. Convergence curve of the fitness of the control profiles.

under the premise that the inverse design error meets the requirements, the number of control points of the key profiles should be reduced as much as possible, thereby reducing the number of design variables.

In order to facilitate the comparative analysis of inverse design error of head shapes in different sizes, the head length is used as the characteristic length to normalize the three-dimensional shape of train head. After normalization, the streamlined head length is kept

1000 mm. Maximum and average inverse design errors of the key control profiles of four different head shapes is shown in Table 2. It can be found that the absolute average errors of the key control profiles of four different head shapes are within 2 mm, and the relative error (absolute error/head length) is less than 0.2%. Although the length and width of the longitudinal profile, the horizontal profile and the bottom profile are relatively large, the corresponding absolute errors are no more than 9 mm, and the relative errors are no more than 1%. Meanwhile, the absolute errors of cross-sectional profile and cowcatcher profile are less than 5 mm, and the relative errors are less than 0.5%. These errors are basically consistent with EN14067-6-2018(EN, 2010) for the manufacturing error requirements of wind tunnel test models. Therefore, it can be concluded that such errors have negligible effect on the results of rapid calculations.

(II) inverse design of the streamlined head. The two-dimensional key control profile determines the topological structure of the three-dimensional shape of the

Table 2. Maximum and average errors for the inverse design of two-dimensional profiles

		Cross-sectional Profile	Longitudinal Profile	Cowcatcher Profile	Horizontal Profile	Bottom Profile
TEST1	Max Error/mm	4.17	1.73	0.68	1.72	8.03
	Ave Error/mm	1.15	1.28	0.50	0.54	0.82
TEST2	Max Error /mm	2.88	7.31	2.23	3.44	1.80
	Ave Error /mm	1.53	1.18	0.93	0.90	0.36
TEST3	Max Error /mm	1.91	8.58	0.43	7.42	0.17
	Ave Error /mm	0.92	1.69	0.34	1.64	0.03
TEST4	Max Error /mm	1.64	8.64	1.45	3.01	0.86
	Ave Error /mm	1.08	1.85	0.61	0.52	0.61

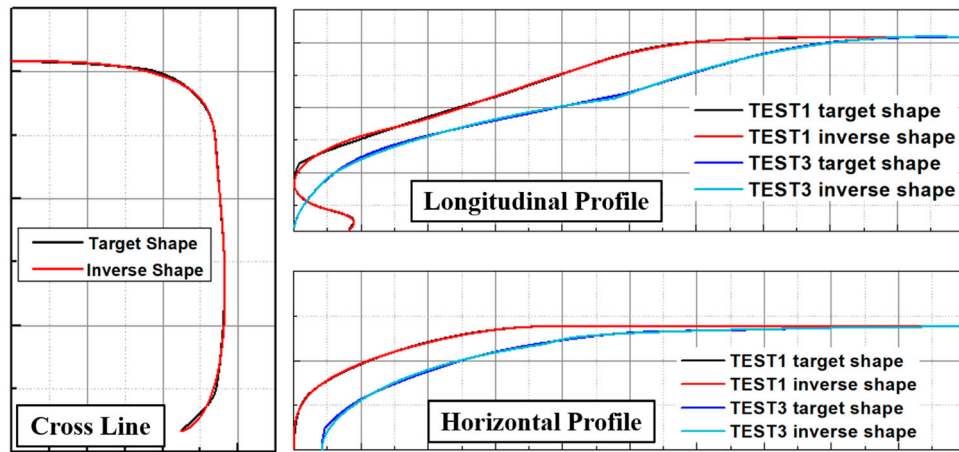


Figure 12. Inverse design results of different key profiles.

Table 3. Maximum and average errors of inverse design shape.

	TEST1	TEST2	TEST3	TEST4
Max error/mm	8.43	8.75	9.83	9.67
Ave error/mm	1.82	2.36	2.74	2.31

high-speed train head, and the surface interpolation method determines the fitting accuracy of the three-dimensional surface. Figure 13 shows the inverse design results of four different topological structures. It can be seen that the inversely designed shape obtained based on the interpolation of the two-dimensional key control profiles is basically consistent with the target shape, indicating that the inverse design method can relatively well realize the inverse design of the high-speed train head, which is also the basic condition for the implementation of the proposed rapid prediction model.

Similar to the error analysis of the inverse design of two-dimensional profiles, the maximum and average errors of the surface fitting of four streamlined shapes are shown in Table 5. It can be found that the fitting error of the curved surface is larger than that of the two-dimensional profiles, and the location of the largest error is mainly distributed in the area with sharp curvature changes. Thus, in order to ensure that the average error of the curved surface is small, the interpolation method was locally smoothed in the area where the curvature changes sharply. As shown in Table 3, the relative value of the maximum fitting error of the curved surface is less than 1%, while the relative value of the average fitting error is less than 0.5%, which both satisfy the requirements of EN14067-6-2018 (EN, 2010) for the manufacturing errors of wind tunnel test models. Therefore, the current inverse design method for high-speed train streamlined shape can be used for rapid prediction of aerodynamic forces.

Table 4. Mesh configurations.

	Boundary layers	Growth ratio	Minimum grid size	Number of grids
Coarse mesh	50	10	10mm	35.15M
Fine mesh	30	10	1.5mm	68.35M

Table 5. Aerodynamic drag coefficients of wind tunnel test and two meshing methods.

	TEST1	TEST2	TEST3	TEST4
Wind tunnel test (Gao et al., 2019; Shuanbao et al., 2014)	0.3261	0.3482	0.3477	0.3234
Wind tunnel test (Gao et al., 2019)	/	0.3549	/	/
Coarse mesh	0.3676	/	/	/
Fine mesh	0.3261	0.3442	0.3483	0.3279

3. Construction of the rapid prediction model

There are more than ten design indicators such as aerodynamic drag, aerodynamic noise and pressure wave in terms of aerodynamic design of high-speed train, whereas the proposed data-driven rapid aerodynamic prediction model does not have limitations on the type of design indicators. Without loss of generality, the aerodynamic drag coefficient of three-carriage train was taken as the objective to illustrate the effectiveness of the proposed method.

Rapid prediction model for aerodynamic force is an implicit function of the shape design variables and aerodynamic force of high-speed train. It is difficult to obtain the implicit function by analytical method because of the obvious nonlinear relationship between design variables and the objectives as well as the interaction between design variables. Consequently, we adopted the SVM model to describe the implicit relationship between the

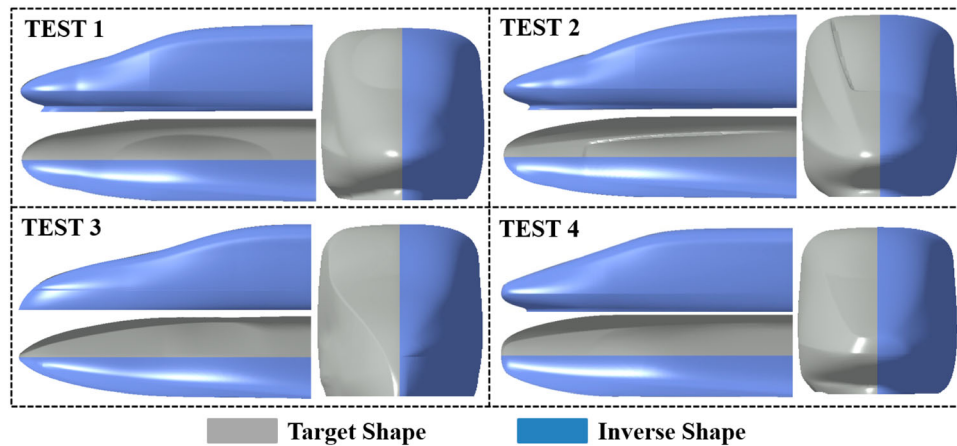


Figure 13. Inverse design results of different topological structures.

shape design variables and aerodynamic force in current study. To build up the SVM model, we need to find approaches for the collection of training data and determine the design space for the design variables, which are discussed below.

3.1. Training data acquisition

As the basis for the construction of the rapid prediction model, wind tunnel test and numerical simulation play an important role in providing sufficient initial data. In this study, similar high-speed train models (the term “similar” means models with the same train body, bogies, and windshields but with different streamlined shapes) have been employed for both wind tunnel test and numerical simulation. Details about wind tunnel tests and numerical simulations are presented in this section.

3.1.1. Wind tunnel test

The wind tunnel tests provide a direct way to evaluate aerodynamic loads of high-speed trains, which are used for aerodynamic prediction and comparison of different configurations during the initial design stage of new high-speed trains. Meanwhile, the experimental data from wind tunnel tests can also be utilized to validate the numerical algorithms. In this study, experimental data are adopted both for the buildup of initial sample set and numerical validation as well. For the sake of illustration, wind tunnel tests in Mianyang, Sichuan are exhibited. The experiments were completed in a large low-speed wind tunnel with an open test section of $8\text{ m} \times 6\text{ m}$. As shown in Figure 14, the test model was a three-carriage scaled train model with a ratio of 1:8, and the length, height, and width of the model are 9.75, 0.4375, 0.4225 m, respectively. During the test, the length of the model was parallel to the incoming flow, each carriage was independently supported, and a force balance was installed

at the geometric center of each carriage to evaluate the aerodynamic drag. The velocity of the incoming flow was 60 m/s. Taking the model height as the reference length, the Reynolds number is 1.71×10^6 .

Besides the aforementioned four wind-tunnel streamlined models, experimental data from other train models are also included during the construction of the rapid prediction model. The inverse design provides the specific value of each design variable of the train model, while the wind tunnel tests give out the aerodynamic performance.

3.1.2. Numerical simulation

Numerical simulation plays an important role in evaluation of the aerodynamic performance of high-speed trains, as long as its credibility has been validated. In order to provide sufficient initial data for the construction of the rapid prediction model, samples from numerical simulations have been extensively utilized.

This study uses wind tunnel test data as benchmark data to validate the accuracy of numerical approaches. The aerodynamic forces of different streamlined shapes are obtained by numerical simulations. To ensure the comparability of numerical models, the computational model and setup used in the simulations are consistent with wind tunnel tests.

Figure 15 shows the geometric model of the high-speed train. It is a three-carriage model with the scaled ratio of 1:8 including bogies, windshields, roadbeds, and tracks. The impacts of other components, such as pantographs, on the aerodynamic performance of the train are not considered.

The computational domain is identical to the wind tunnel test section. As shown in Figure 16, the height H from the top of the train body to the rail surface is taken as the characteristic length. The distance between the inlet boundary and the nose tip of the leading train



Figure 14. Wind tunnel test.

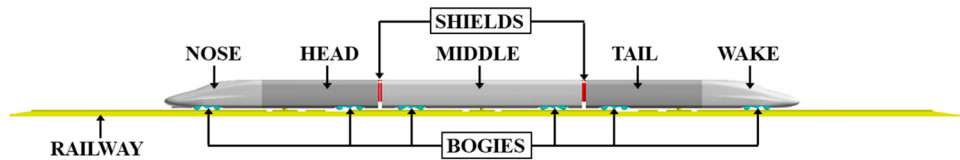


Figure 15. Geometric model of three-carriage high-speed train.

is $6.5H$, the distance between the outlet boundary and the trailing nose is $11H$, and the height and width of the computational domain are $13H$ and $17.3H$, respectively.

The incompressible steady RANS solver is employed for aerodynamic calculation. The second-order upwind scheme is adopted for convection term and second-order central difference scheme for viscous term. The k - ω SST model is used for turbulence closure. The boundary conditions of the numerical simulations are consistent with wind tunnel test. The velocity inlet condition with the incoming flow of 60 m/s is employed for the inlet boundary. The outlet boundary is set with a zero-pressure outlet. The lateral sides, top, ground, roadbed and track of the computational domain are non-slip walls.

The Cartesian grids are used for spatial discretization through the commercial software STAR-CCM+ 13.04. Prism layer grids are arranged along the train body, track, and ground. To reduce the amount of grids in the boundary layer area, the standard wall function is adopted, so that the value of y^+ for the height of the first cell near the wall can be kept around 30–50. To better capture the flow details around the train, several refined regions are placed, such as the irregular areas of the train body, the bottom space of the train, and the wake field, as shown in Figure 17. The minimum grid size of the refined zone is 12 mm , and the total amount of grids is approximately 68 million.

To validate the rationality of meshing and investigate the influence of the grid size on the numerical results, coarse mesh and fine mesh are both generated for model TEST1, as shown in Table 4. Among them, the configuration of the fine mesh is consistent with Figure 17.

Aerodynamic drag (Li et al., 2021) is nondimensionalized as follows,

$$C_d = F_d / 0.5\rho V^2 S \quad (7)$$

where F_d is the aerodynamic drag, ρ is air density, S is the maximum cross-section area of the train, and V is the speed of the train, whereas for wind tunnel test, V represents the speed of incoming flow. To ensure the comparability of the data, this paper takes the value of S as 0.175 m^2 .

Aerodynamic force of the model train was obtained through steady methods for wind tunnel test, therefore time-averaged value is adopted from numerical simulations. EN 14067-6-2018(EN, 2010) requires that when the same standard model is tested in different wind tunnels or at different times in the same wind tunnel, the average error of the test results should be less than 10%, and the maximum error should be less than 15%.

TEST2 model was tested in 2014 and 2019 respectively in the $8 \text{ m} \times 6 \text{ m}$ wind tunnel in Mianyang, Sichuan Province, China as shown in Table 2. The test error for aerodynamic drag is 1.92%, meeting the accuracy requirement of 14067-6-2018(EN, 2010). The aerodynamic drag coefficients of the three-carriage train model corresponding to the wind tunnel test and the two meshing methods are given in Table 5. It can be seen from Table 2 that the aerodynamic drag coefficient of the whole train calculated by the coarse mesh is obviously too large, while the aerodynamic drag coefficients of the four head shapes calculated by the fine grid are basically consistent with the wind tunnel data. The prediction error of TEST2 is the largest, but only with the value of 1.2%, which still

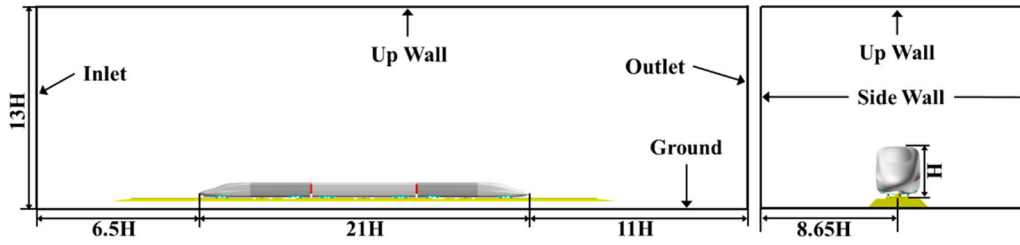


Figure 16. Computational domain.

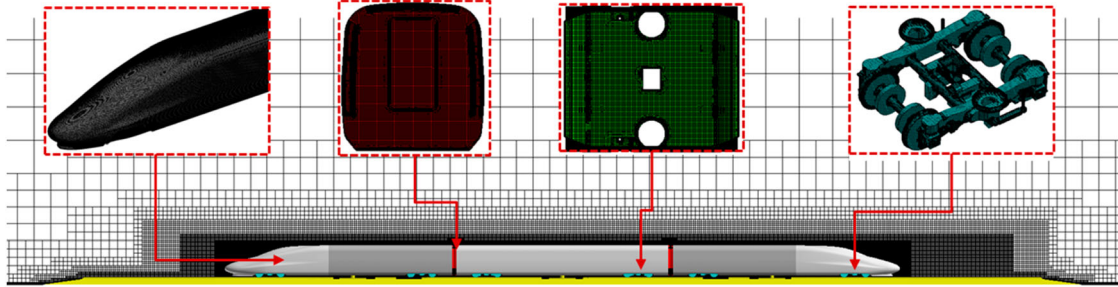


Figure 17. The spatial grids and local surface mesh.

meets the accuracy requirement of engineering applications. Consequently, the fine mesh configuration will be adopted for further numerical simulations.

3.2. Design space

The streamlined shape of the high-speed train is designed variously to adapt to different operating conditions. In order to cover as many topologic types of streamlined shapes as possible, all the design variables in Table 3 are used to construct the rapid prediction model. To facilitate the determination of the design space, the head length is taken as the characteristic length for the three-dimensional shape. The head length after scaling is 1. Thus, the design space is scaled down to the space of unit 1. The scaling formulas of design parameters are shown below:

$$x'_{ij} = \frac{x_{ij}}{L} \quad (8)$$

$$y'_{ij} = \frac{y_{ij}}{L} \quad (9)$$

$$z'_{ij} = \frac{z_{ij}}{L} \quad (10)$$

where L is the head length.

The design variables of the parametric design of high-speed train are coordinates of the control points on the control profiles. Along with the change of the cross-sectional shape, the positions of each control point will move correspondingly. However, if the design space of each design variable adopts the absolute coordinate value, the curvature of the control profile could easily vary too much or even interfere with each other. Therefore, this

Table 6. Design space of benchmark variables.

Benchmark variables	L	$W/2$	H
Value range/m	[0.2,1.75]	[0.1,0.1875]	[0.2,0.375]
Actual size/m	[2.4,21]	[1.2,2.25]	[2.4,4.5]

study adopts the relative coordinate value to obtain the design space.

Head length L , body width W and body height H are selected as reference lengths. Among them, L and H correspond to the design variable x_{26} and z_{15} , while W has no corresponding design variable. Design space of these three variables is determined using absolute coordinate value, through appropriately expanding the value range of the three benchmark variables on the basis of the actual size of the existing high-speed trains, as shown in Table 6.

Design space of other variables takes the proportional values of these three benchmark variables. Among them, L is the benchmark parameter for x-coordinate of design variables, $W/2$ for y-coordinate and H for z-coordinates. The design space of each design variable is shown in Table 7.

3.3. Support vector machines model

Support Vector Machines (SVM) model (Vapnik, 1998) benefits from good generalization ability, nonlinear processing ability and high dimensional processing ability. SVM takes training error as the constraint condition and confidence range minimization as the optimization objective to solve a convex quadratic optimization problem with linear constraints. For nonlinear regression

Table 7. Design space.

Parameter	Value range	Parameter	Value range	Parameter	Value range	Parameter	Value range
y_{11}	[0.5,1.0]	x_{23}	[0.2,0.4]	z_{32}	[0.3,0.5]	y_{44}	[0.4,1.2]
y_{12}	[0.5,1.2]	z_{23}	[0.1,0.7]	x_{33}	[-0.1,0.2]	x_{45}	[0.3,1.0]
z_{12}	[0.05,0.3]	x_{24}	[0.4,0.6]	z_{33}	[0.5,0.8]	x_{52}	[0.0,0.2]
y_{13}	[0.6,1.5]	z_{24}	[0.2,1.0]	x_{34}	[-0.1,0.2]	y_{52}	[0.05,0.4]
z_{13}	[0.3,0.85]	x_{25}	[0.6,0.9]	z_{34}	[0.8,1.0]	x_{53}	[0.2,0.5]
y_{14}	[0.6,1.5]	z_{25}	[0.8,1.0]	x_{42}	[0.0,0.2]	y_{53}	[0.1,1.0]
z_{14}	[0.85,1.0]	x_{26}	[0.2,1.75]	y_{42}	[0.05,0.4]	x_{54}	[0.5,1.0]
z_{15}	[0.2,0.375]	x_{31}	[-0.1,0.2]	x_{43}	[0.2,0.5]	y_{54}	[0.4,1.2]
x_{22}	[0.0,0.2]	z_{31}	[0.1,0.3]	y_{43}	[0.1,0.8]	x_{55}	[0.3,1.0]
z_{22}	[0.05,0.5]	x_{32}	[-0.1,0.2]	x_{44}	[0.5,1.0]		

problem, SVM first conducts a nonlinear mapping to map the vector to a high-dimensional space, and then performs linear regression in the feature space, so as to achieve the effect of nonlinear regression in the original space. In order to solve the problem that it is difficult to directly calculate the optimal hyperplane in the feature space due to the rapid increase of dimensions in the process of mapping from low-dimensional space to high-dimensional feature space, SVM introduces kernel function to transform the problem into calculation in the input space.

This study adopts the algorithm ϵ -TSVR (ϵ -twin support vector regression) proposed by Shao et al. (2013) compared with standard SVM algorithm, ϵ -TSVR owns higher prediction ability and requires less training time. For nonlinear regression problems, the original problems of ϵ -TSVR algorithm are described as below:

$$\begin{aligned} \min_{u_1, b_1, \xi} & \frac{1}{2}c_3(u_1^T u_1 + b_1^2) + \frac{1}{2}\xi^T \xi^* + c_1 e^T \xi, \\ \text{s.t.} & Y - (K(A, A^T)u_1 + eb_1) \geq -\varepsilon_1 e - \xi, \xi \geq 0, \\ & Y - (K(A, A^T)u_1 + eb_1) = \xi^* \end{aligned} \tag{14}$$

$$\begin{aligned} \min_{u_2, b_2, \eta} & \frac{1}{2}c_4(u_2^T u_2 + b_2^2) + \frac{1}{2}\eta^T \eta^* + c_2 e^T \eta, \\ \text{s.t.} & (K(A, A^T)u_1 + eb_1) - Y \geq -\varepsilon_2 e - \eta, \eta \geq 0, \\ & (K(A, A^T)u_2 + eb_2) - Y = \eta^* \end{aligned} \tag{15}$$

Where $c_1, c_2, c_3, c_4, \varepsilon_1$ and ε_2 are the coefficients greater than 0; u_1 and u_2 are real vectors; b_1 and b_2 are real coefficients; ξ, ξ^*, η and η^* are relaxation vectors; $K(A, A^T)$ is the kernel function which can have many expressions. The Gaussian kernel function is adopted, as shown below, where σ is a coefficient called width factor:

$$\exp\left(-\frac{1}{2\sigma^2}\|x - x_i\|^2\right) \tag{16}$$

For the problem with m dimensions and n training samples, the corresponding A is an $m \times n$ matrix, where

A_i is the i^{th} training sample. $Y = (y_1, y_2, \dots, y_m)$ is the response value of the training samples. e is unit vector. By constructing Lagrange function and using Karush-Kuhn-Tucker complementary condition, the dual problem of formulas (14) and (15) can be obtained respectively, as shown in formula (17) and (18). Specific derivation can be found in literature [15].

$$\begin{aligned} \max_{\alpha} & -\frac{1}{2}\alpha^T H(H^T H + c_3 I)^{-1} H^T \alpha^T \\ & - (e^T \varepsilon_1 + Y^T)\alpha + Y^T H(H^T H + c_3 I)^{-1} H^T \alpha \end{aligned} \tag{17}$$

$$\text{s.t. } 0 \leq \alpha \leq c_1 e. \tag{18}$$

where

$$H = [K(A, A^T)e] \tag{19}$$

$$v_1 = (H^T H + c_3 I)^{-1} H^T (Y - \alpha) \tag{20}$$

$$v_2 = (H^T H + c_4 I)^{-1} H^T (Y + \gamma) \tag{21}$$

where $v_1 = [u_1^T b_1]^T, v_2 = [u_2^T b_2]^T$. By solving formula (17) and (18), u_1, u_2, b_1 and b_2 can be derived, and then the predicted value of SVM model α can be obtained through formula (22).

$$\begin{aligned} f(x) &= \frac{1}{2}(f_1(x) + f_2(x)) \\ &= \frac{1}{2}(u_1^T + u_2^T)K(A, x) + \frac{1}{2}(b_1 + b_2) \end{aligned} \tag{22}$$

After determining the training sample points, the free coefficients of the SVM model are $c_1, c_2, c_3, c_4, \varepsilon_1, \varepsilon_2$ and σ , which affect significantly on the generalization ability of SVM model. However, there is no theoretical basis for rigorous calculation of these coefficients. The cross validation method and particle swarm optimization algorithm are combined herein to determine these free coefficients, and to simplify this problem, we set $c_1 = c_2, c_3 = c_4$. Therefore, only five coefficients need to be determined. For a set of given free coefficients, it is necessary to solve convex quadratic optimization problems (17)

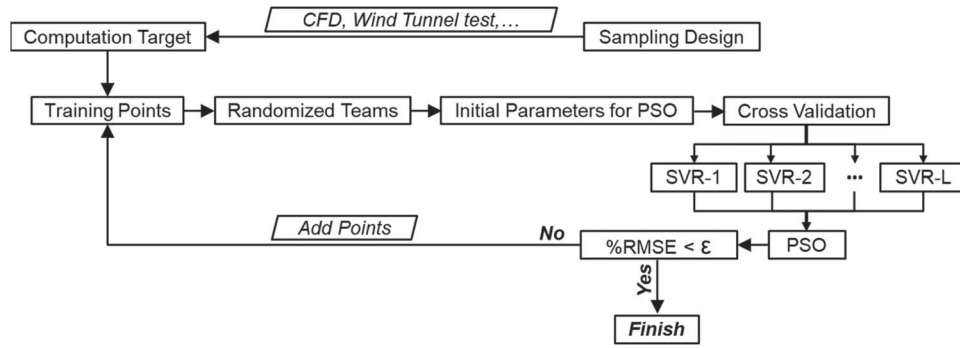


Figure 18. Construction flow chart of the SVM model.

and (18) twice. Shao et al. (2013) introduced the over-relaxation iteration method, which is also adopted when solving (17) and (18), to improve the training efficiency of the model.

Figure 18 shows the flow chart of optimizing free coefficients of the SVM model. The whole process is as follows:

- 1) For a given training sample set, determine the number of sampling groups, randomly group each training sample, and ensure that the number of training samples in each group is the same.
- 2) Determine the initial coefficients of PSO, such as the number of particle swarms, the number of iterations. The number of particles and the number of iterations have a great influence on the optimization efficiency, and should neither be too large nor too small.
- 3) Select a group of training samples sequentially as the test samples and use the other groups of training samples to construct the sub-SVM model, then obtain the prediction error of the test samples. Subsequently, the fitness function of the PSO algorithm can be calculated using formula (23).

$$fit = \frac{1}{l} \sum_{i=1}^l \%RMSE_i \quad (23)$$

where l is the number of sampling groups; $\%RMSE_i$ is the prediction error of the i^{th} test group, which gets the expression as:

$$\%RMSE = 100 \sqrt{\frac{\frac{1}{n_s} \sum_{i=1}^{n_s} (y_i - y_i^{(p)})^2}{\frac{1}{n_s} \sum_{i=1}^{n_s} y_i}} \quad (24)$$

In equation (24), y_i is the true value, $y_i^{(p)}$ is the predicted value of the SVM model, and n_s is the number of test samples.

- (1) Obtain the optimal value of free coefficients after iteration. When using SVM to predict the target value, the average of the predicted values of each sub-SVM model is used as the final predicted value.

4. Results and discussion

When carrying out aerodynamic shape design of high-speed trains, a variety of aerodynamic shapes are generally designed, and then through the comparative analysis of numerical simulations, some shapes with better aerodynamic performance are selected for wind tunnel tests. Therefore, a large amount of numerical data and wind tunnel test data will be generated during the research and development of high-speed trains, and these data can be used as initial training data for the rapid prediction model. Compared with numerical simulations, wind tunnel tests suffer from higher cost and longer experimental period. Numerical simulations have become the most important analysis method in aerodynamic design of high-speed trains, and the corresponding data formed is far more than that of wind tunnel tests. However, affected by the accuracy of numerical simulations, wind tunnel test is the most important method to validate numerical simulations, and it is also indispensable in engineering design. When constructing the repaid prediction model, data from numerical simulations and wind tunnel tests are both adopted as training data, and the former is far more than the latter.

The effectiveness of the rapid prediction model is validated by taking the aerodynamic drag coefficient as the objective in current study. The distribution of the drag coefficient of the initial sample set is shown in Figure 19. The design variables for all the samples are the same, which are the same as those in Table 1. The design objective is obtained either from numerical simulations, as the black dots show, or from wind tunnel tests, as the blue dots show in Figure 19. It can be seen that samples from the numerical simulations are uniformly distributed while that from wind tunnel tests are distributed

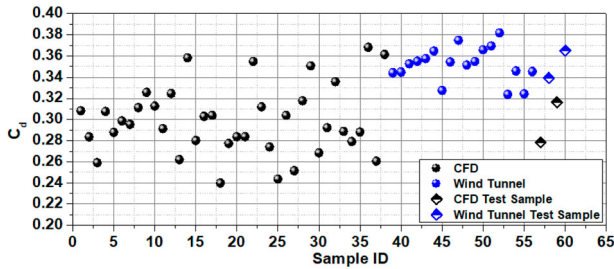


Figure 19. Target value distribution of the sample set.

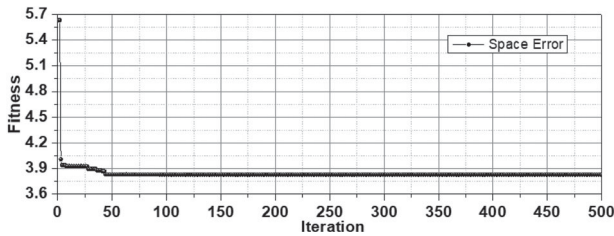


Figure 20. Convergence curve of fitness.

in area with larger value ranging from 0.32 to 0.39, which is because samples tested in wind tunnels are designed specifically under engineering design requirements with plenty engineering constraints such as vehicle gauge, cab space, and head length.

As observed in Figure 19, two samples from numerical simulations and two samples from wind tunnel tests are randomly selected from the initial sample set as test samples, which are marked with the diamond shape. The remaining 38 numerical samples are taken as initial training samples. They are divided into 38 groups, each with 1 sample point. The formula (23) is adopted as the objective function, and the SVM model is trained through the cross-validation method. The training process is shown in Figure 18. The PSO algorithm is adopted for the training process. The basic setting of PSO for the construction of the SVM model is as follows: the number of particle swarms is 200, the number of iterations is 500, the acceleration factor value is 2, and the inertia factor gradually changes from 1.2 to 0.8 as the number of iterations increases; the maximum flight speed of the particles is 0.1. The convergence curve of fitness with the number of iterations is shown in Figure 20, in which it can be observed the average calculation error can rapidly converge to a stable value. When applying the rapid prediction model into engineering applications, the iterations can be properly reduced to improve the training efficiency.

In order to analyze the effect of training set on the prediction accuracy of SVM, taking two wind tunnel test samples as a group, a total of eight groups of samples were divided. These samples were added to the training

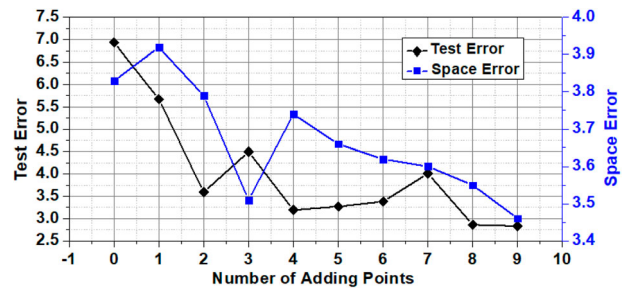


Figure 21. Prediction error of SVM.

set group by group. During this process, we still take 1 sample as 1 group and use the cross-validation method to train the SVM model, and the parameter setting of PSO remains unchanged. The average prediction error of the SVM model in the design space and the average prediction error of the test samples are shown in Figure 21. It can be observed that as the number of training samples increase, the average prediction error in the design space and the average prediction error of the test samples both gradually decrease. After adding points for 9 times, the average prediction error in the design space is reduced to 3.46%, and the average prediction error of the test samples is reduced to 2.83%, which are much smaller than the test error requirements of EN 14067-6-2018(EN, 2010), indicating the prediction accuracy of the proposed model will gradually improve as the number of training samples increases. It's worth mentioning that with the continuous development of high-speed trains, a large number of numerical calculations as well as wind tunnel test data will be accumulated in the engineering design process. Combining these data with the rapid prediction model will offer a promising alternative for accurate evaluation of aerodynamic forces of high-speed trains.

Table 8 shows the comparison of results of 4 test samples calculated by the rapid prediction model, wind tunnel tests and numerical simulations. The relative error of T1 and T2 refers to the error between the rapid prediction results and the CFD simulation results while the relative error of T3 and T4 refers to the error between the rapid prediction results and the wind tunnel test results. Note that the maximum relative error is 4.49%, and the minimum is 0.96%, both less than 5%, which is 50% less than the average error required by EN 14067-6-2018(EN, 2010). Compared with the wind tunnel test results, the relative error of CFD calculation of T4 is 3.12%, which is equivalent to the CFD calculation error of Gao et al. (2019).

Taking the wind tunnel test data as benchmark data, there is little difference between CFD calculation error and that of rapid prediction model, indicating that when the training samples reaches a certain number, the

Table 8. Comparison of results from three different approaches.

	T1	T2	T3	T4
Wind Tunnel Test	/	/	0.3394	0.3650
Numerical Simulation	0.2783	0.3163	0.3375	0.3536
Rapid Prediction	0.2849	0.3021	0.3485	0.3685
Relative Error	2.37%	4.49%	2.61%	0.96%

accuracy of the rapid prediction model is basically the same as the CFD simulation.

Apart from the prediction accuracy, the time cost is another important issue that needs to be paid attention to. Current study adopts the three-carriage high-speed train as the train model. To obtain the aerodynamic drag coefficient of one training sample, the model preparation time for CFD is about 4 h, and the simulation time is 6 h with 128 CPUs used for parallel computation. As for wind tunnel test, to obtain the aerodynamic drag coefficient of 1 training sample, 360 h needs to be spent on designing and processing wind tunnel test model and 3 h needs to be spent on experimental preparation and testing. However, as long as the construction of the rapid prediction model has been completed, when calculating the aerodynamic force of high-speed train, only the surface meshing of the head shape is required, which costs only 0.5 h, and the aerodynamic prediction is at practically no additional cost. Therefore, compared with numerical simulation and wind tunnel test, the rapid prediction model proposed in this study can significantly increase the prediction efficiency by 90% without reducing prediction accuracy.

It should be emphasizing that with regard to the proposed rapid prediction model, we can make full use of both the existing and the upcoming experimental and numerical data. Consequently, the sampling data set will be enriched continuously, to make the model more and more accurate.

5. Conclusion

The emphasis of this study is to present a framework of data-driven rapid prediction model and demonstrate its promising potential in predicting aerodynamic forces of high-speed trains. Using the proposed model, the prediction efficiency and accuracy could be both achieved. As demonstrated in this work, we adopted the idea of step-by-step modeling, and proposed a general three-dimensional parameterization method for head shape. Combining with the inverse design concept, the rapid extraction of the values of the design variables were realized. Using data from numerical simulations and wind tunnel tests as the initial training data, and adopting the SVM model to construct a nonlinear implicit function between design variables and aerodynamic forces of

high-speed train, the data-driven rapid prediction model was finally proposed. Taking aerodynamic drag as the prediction objective, the effectiveness of the model was verified. When the number of training samples reaches a certain amount, the accuracy of the rapid prediction model can be basically the same as numerical simulations. Remarkably, although only aerodynamic drag is used to verify the effectiveness of the prediction model, by changing the objective of the training samples, it can be directly applied to the rapid prediction of other aerodynamic indicators such as aerodynamic lift, aerodynamic noise, and tunnel pressure waves.

The proposed data-driven rapid prediction model can be used for performance evaluation of engineering design and aerodynamic optimization of high-speed trains. Compared with wind tunnel tests and numerical simulations, the prediction efficiency is improved by more than 90%, which significantly shortens the evaluation period of the aerodynamic force of high-speed trains. More notably, with the continuous enrichment of wind tunnel test data and numerical simulation data, the prediction accuracy of the rapid prediction model will be continuously improved, which makes this method more promising and suitable for companies and research institutions that have long been engaged in the development and design of high-speed trains to evaluate and optimize the aerodynamic force of high-speed trains.

One possible limitation of the current study is that compared with wind tunnel tests and numerical simulations, the data-driven rapid prediction model can only obtain the aerodynamic force of the high-speed train and is difficult to obtain the flow details around high-speed trains. Besides, with the change in the shape of windshields, pantographs and bogies, it is necessary to take more detailed consideration about the influences of these factors on the accuracy of the rapid prediction model, which is a subject that we will further work on. Continuously enriching the information of training samples and continuously optimizing the modeling ideas of the rapid prediction model will also be the future target to be carried out. Meanwhile, considering different aerodynamic loads such as aerodynamic lift and tunnel pressure wave, future work will be exerted to extend the rapid prediction model to more application scenarios.

Acknowledgments

This work was supported by Youth Innovation Promotion Association CAS (2019020).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Youth Innovation Promotion Association CAS (Grant Number: 2019020).

References

- Baker, C. (2010). The flow around high speed trains. *Journal of Wind Engineering and Industrial Aerodynamics*, 98(6-7), 277–298. <https://doi.org/10.1016/j.jweia.2009.11.002>
- Catanzaro, C., Cheli, F., Rocchi, D., Schito, P., & Tomasini, G. (2010). *High-speed train crosswind analysis: CFD study and validation with wind-tunnel tests*. International conference on engineering conferences international, 12–17 September, Potsdam, Germany.
- EN, B. (2010). 14067-6. *Railway applications— aerodynamics part 6: Requirements and test procedures for cross wind assessment*. British Standard Institute.
- Gao, G., Li, F., He, K., Wang, J., Zhang, J., & Miao, X. (2019). Investigation of bogie positions on the aerodynamic drag and near wake structure of a high-speed train. *Journal of Wind Engineering and Industrial Aerodynamics*, 185, 41–53. <https://doi.org/10.1016/j.jweia.2018.10.012>
- Hemida, H., & Baker, C. (2010). Large-eddy simulation of the flow around a freight wagon subjected to a crosswind. *Computers & Fluids*, 39(10), 1944–1956. <https://doi.org/10.1016/j.compfluid.2010.06.026>
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. Proceedings of ICNN'95-international conference on neural networks.
- Ku, Y.-C., Park, H.-I., Kwak, M.-H., & Lee, D.-H. (2010). Multi-objective optimization of high-speed train nose shape using the vehicle modeling function. 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition.
- Li, T., Dai, Z., Yu, M., & Zhang, W. (2021). Numerical investigation on the aerodynamic resistances of double-unit trains with different gap lengths. *Engineering Applications of Computational Fluid Mechanics*, 15(1), 549–560. <https://doi.org/10.1080/19942060.2021.1895321>
- Muld, T. W., Efraimsson, G., & Henningson, D. S. (2012). Flow structures around a high-speed train extracted using proper orthogonal decomposition and dynamic mode decomposition. *Computers & Fluids*, 57, 87–97. <https://doi.org/10.1016/j.compfluid.2011.12.012>
- Munoz-Paniagua, J., & García, J. (2020). Aerodynamic drag optimization of a high-speed train. *Journal of Wind Engineering and Industrial Aerodynamics*, 204, 104215. <https://doi.org/10.1016/j.jweia.2020.104215>
- Östh, J., Noack, B. R., Krajnović, S., Barros, D., & Borée, J. (2014). On the need for a nonlinear subscale turbulence term in POD models as exemplified for a high-reynolds-number flow over an ahmed body. *Journal of Fluid Mechanics*, 747, 518–544. <https://doi.org/10.1017/jfm.2014.168>
- Raghunathan, R. S., Kim, H.-D., & Setoguchi, T. (2002). Aerodynamics of high-speed railway train. *Progress in Aerospace Sciences*, 38(6-7), 469–514. [https://doi.org/10.1016/S0376-0421\(02\)00029-5](https://doi.org/10.1016/S0376-0421(02)00029-5)
- Schetz, J. A. (2001). Aerodynamics of high-speed trains. *Annual Review of Fluid Mechanics*, 33(1), 371–414. <https://doi.org/10.1146/annurev.fluid.33.1.371>
- Shao, Y.-H., Zhang, C.-H., Yang, Z.-M., Jing, L., & Deng, N.-Y. (2013). An ϵ -twin support vector machine for regression. *Neural Computing and Applications*, 23(1), 175–185. <https://doi.org/10.1007/s00521-012-0924-3>
- Shuanbao, Y., Dilong, G., Zhenxu, S., Guowei, Y., & Dawei, C. (2014). Optimization design for aerodynamic elements of high speed trains. *Computers & Fluids*, 95, 56–73. <https://doi.org/10.1016/j.compfluid.2014.02.018>
- Sima, M., Gurr, A., & Orellano, A. (2008). Validation of CFD for the flow under a train with 1: 7 scale wind tunnel measurements. BBAA VI International Colloquium on: Bluff bodies aerodynamics and applications, Milano, July 20–24, 66–79.
- Sun, Z., Yao, S., & Yang, G. (2020). Research on aerodynamic optimization of high-speed train's slipstream. *Engineering Applications of Computational Fluid Mechanics*, 14(1), 1106–1127. <https://doi.org/10.1080/19942060.2020.1810128>
- Vapnik, V. (1998). *Statistical learning theory* New York. Wiley. 1(2), 3.
- Wang, Q., Qian, W., & He, K. (2015). Unsteady aerodynamic modeling at high angles of attack using support vector machines. *Chinese Journal of Aeronautics*, 28(3), 659–668. <https://doi.org/10.1016/j.cja.2015.03.010>
- Wang, S., Bell, J. R., Burton, D., Herbst, A. H., Sheridan, J., & Thompson, M. C. (2017). The performance of different turbulence models (URANS, SAS and DES) for predicting high-speed train slipstream. *Journal of Wind Engineering and Industrial Aerodynamics*, 165, 46–57. <https://doi.org/10.1016/j.jweia.2017.03.001>
- Wang, Y., Wang, Y., An, Y., & Chen, Y. (2008). Aerodynamic simulation of high-speed trains based on the lattice boltzmann method (LBM). *Science in China Series E: Technological Sciences*, 51(6), 773–783. <https://doi.org/10.1007/s11431-008-0063-3>
- Yang, G., Guo, D., Yao, S., & Liu, C. (2012). Aerodynamic design for China new high-speed trains. *Science China Technological Sciences*, 55(7), 1923–1928. <https://doi.org/10.1007/s11431-012-4863-0>
- Yang, Y., He, Z., Shi, Z., & Xiong, X. (2022). Multi-Objective aerodynamic optimization of a high-speed train head shape based on an optimal kriging model. *Journal of Applied Fluid Mechanics*, 15(3), 803–813. <https://doi.org/10.47176/jafm.15.03.33183>
- Yao, S.-B., Sun, Z.-X., Guo, D.-L., Chen, D.-W., & Yang, G.-W. (2013). Numerical study on wake characteristics of high-speed trains. *Acta Mechanica Sinica*, 29(6), 811–822. <https://doi.org/10.1007/s10409-013-0077-3>
- Zhang, Z., Song, X.-d., Ye, S.-r., Wang, Y.-w., Huang, C.-g., An, Y.-r., & Chen, Y.-s. (2019). Application of deep learning method to reynolds stress models of channel flow based on reduced-order modeling of DNS data. *Journal of Hydrodynamics*, 31(1), 58–65. <https://doi.org/10.1007/s42241-018-0156-9>
- Zhu, S., & Wang, Y. (2019). Scaled sequential threshold least-squares (S2TLS) algorithm for sparse regression modeling and flight load prediction. *Aerospace Science and Technology*, 85, 514–528. <https://doi.org/10.1016/j.ast.2018.12.038>