

Mach Number Prediction for a Wind Tunnel Based on the CNN-LSTM-Attention Method

ZHAO Luping*, WU Kunyang

(College of Information Science and Engineering, Northeastern University, Shenyang 11081, China;

*Correspondence: zhaolp@ise.neu.edu.cn)

Abstract: The test section's Mach number in wind tunnel testing is a significant metric for evaluating system performance. The quality of the flow field in the wind tunnel is contingent upon the system's capacity to maintain stability across various working conditions. The process flow in wind tunnel testing is inherently complex, resulting in a system characterized by nonlinearity, time lag, and multiple working conditions. To implement the predictive control algorithm, a precise Mach number prediction model must be created. Therefore, this report studies the method for Mach number prediction modelling in wind tunnel flow fields with various working conditions. Firstly, this paper introduces a continuous transonic wind tunnel. The key physical quantities affecting the flow field of the wind tunnel are determined by analyzing its structure and blowing process. Secondly, considering the nonlinear and time-lag characteristics of the wind tunnel system, a CNN-LSTM model is employed to establish the Mach number prediction model by combining the 1D-CNN algorithm with the LSTM model, which has long and short-term memory functions. Then, the attention mechanism is incorporated into the CNN-LSTM prediction model to enable the model to focus more on data with greater information importance, thereby enhancing the model's training effectiveness. The application results ultimately demonstrate the efficacy of the proposed approach.

Keywords: Wind Tunnel Test, Mach Number Prediction; CNN-LSTM; Attention Mechanism

1 Introduction

A wind tunnel is a circular, tubular experimental apparatus designed to artificially create and manage airflow. It replicates the flow of air around an aircraft or object, quantifies the impact of airflow on the object, and allows for the observation of various physical phenomena^[1]. As an infrastructure for flight vehicle design, wind tunnels have supported the development of the aviation industry for many years. It is generally accepted in the field of aviation that safe and stable aircraft as well as other space vehicles cannot be

designed without wind tunnels. Many vehicles require hundreds of tests in wind tunnels to obtain the best shape.

Wind tunnel experiments are essential components of aircraft development efforts. They hold a critical role not only in advancing aviation and aerospace engineering but have also become indispensable in various domains, including transportation, architectural construction, and harnessing wind energy, owing to the evolution of industrial aerodynamics. This method of experimentation facilitates precise control of flow

conditions. Typically, in wind tunnel experiments, researchers securely position the model or physical object within the wind tunnel and subject it to repeated airflow, while data is collected via measuring and controlling instruments and equipment.

Currently, there are two mainstream wind tunnel prediction model building methods are mechanistic modeling^[2] and data-driven modeling^[3].

(1) Mechanistic modeling

In the early days when high-precision sensors were not available, most scholars used this method to combine aerodynamic principles and derive mathematical formulas to build predictive models of wind tunnel processes. The mechanistic modeling is highly explanatory, utilising physical relationships of the blowing process and accurately describing the working process of the wind tunnel system. This makes it more readable for those involved.

The Swedish T1500 (1.5m × 1.5m induced transonic) wind tunnel combines aerodynamic and control theories to study control strategies. The non-linear one-dimensional aerodynamic equations inform a numerical model of air circulation, from which a hydraulic servo valve model produces a numerical model of wind tunnel operation^[4].

The Dutch National Space Laboratory's continuous high-speed wind tunnel (HST) uses a PID controller to simulate air circulation with a Mach number of 0.8. However, this model only describes the linear system and fails to account for system changes or rapid shifts in frontal velocity, leading to major disturbances in the flow field^[5].

In the transonic NTF wind tunnel located in the United States, there exists a significant multisystem coupling in the control of variables. The controller design and analysis for multiple-input and multiple-output systems in this wind tunnel employ a transfer function model to describe the controlled object through a linear ordinary differential equation. However, it is difficult to obtain better control results utilizing controllers designed based on transfer functions for systems with time-varying characteristics

and severe nonlinearity^[6].

(2) Data-driven modeling

Data-driven modeling regards the entire system as a "black box" and employs field-collected data to establish the relationship between system inputs and outputs. At present, with the advent of numerous electronic sensors that offer a high degree of accuracy, it is increasingly possible to gather and evaluate data in real-time while the blowing process takes place. The wind tunnel system possesses a wealth of experimental data and the requisite conditions for data-driven modeling. Establishing the data-based wind tunnel Mach number model has become an inevitable requirement to improve its control accuracy. However, the research in the direction of data-driven modeling of wind tunnel systems has been carried out late, so the work for this area still needs to be continued by researchers.

In 1984, Manitius^[7] developed a data-based system identification theory model for controlling the second throat Mach number in a wind tunnel. The model predicted the Mach number only for the steady Mach number phase of the test and did not account for other phases.

In 2004, Zhao Shujun and colleagues developed an online model for identifying Mach numbers. The model was based on analyses of the flow field characteristics of induced-jet wind tunnels and used neural networks to predict Mach numbers under two distinct operational scenarios.

In 2011, Fengjing Shen simplified a complex wind tunnel system into a two-input, two-output system and used the least squares method for system identification. However, the data used in this model were obtained by step experiments, and only the effects of main exhaust valve displacement and standing chamber flow valve displacement on the Mach number were considered in the modeling process, and the effects of air source, angle of attack, and main regulator displacement on the Mach number were ignored.

In 2013, Dandois & Pamart^[8] used a Nonlinear Auto-Regressive with Exogenous variables (NARX)^[9] model with external inputs to identify the pressure

signal in a wind tunnel system.

To attain swift and precise control of the Mach number, the predictive control in the control strategy plays an important role, but also puts higher requirements on the predictive model. The prediction model must not only accurately capture the relationship between inputs and outputs but also align with the characteristics of the wind tunnel system in terms of nonlinearity and time lag. The choice of the model architecture should take into account both the wind tunnel system properties and the feasibility of the available models. Only by matching the two can a suitable model structure for the system be selected.

The partial least squares (PLS) model combines multivariate regression, principal component analysis, and correlation analysis. It addresses data co-linearity and small sample problems that traditional regression methods struggle with, making it suitable for process modeling and prediction^[10].

The BP neural network, proposed by American scientist Rumelhart in 1986, features a simple structure and stable working conditions. It is used in pattern recognition, image processing, and prediction. Although the BP neural network has powerful nonlinear processing capability, the model has no memory and feedback function and only relies on the prediction output and the actual output error of the current moment to adjust the model parameters, which is a typical static neural network. Moreover, the prediction time domain of this prediction model cannot be too long, otherwise it will lead to the phenomenon of error accumulation, thus reducing the model's prediction accuracy.

The CNN-LSTM-attention mechanism is a deep learning model designed for processing sequential data. The approach integrates the convolutional neural network (CNN), long short-term memory network (LSTM), and attention mechanism to enhance the feature extraction and modeling of sequential data^[11].

In this paper, first, the input data such as rotational speed, angle of attack and total pressure are analyzed using a CNN to capture features at different

scales and levels of abstraction so that the model can better understand the structural and semantic information of the wind tunnel data. Next, the feature sequences that have been extracted are inputted into the LSTM network, which is a type of recurrent neural network that successfully captures the time dependence of data sequences such as Mach number and angle of attack. This allows the LSTM to effectively identify the long-term patterns and contextual information in the wind tunnel sequence data and produce the corresponding hidden state representations. Then, an attention mechanism is introduced to assign varying weights to different portions of the input wind tunnel sequence. The objective is to direct the model's focus towards the vital information relevant to Mach number prediction.

In summary, CNN-LSTM-Attention is a powerful sequence modeling method, which can effectively extract the features and capture the temporal dependence of wind tunnel sequence data represented by rotational speed, angle of attack, and total pressure, and provides an effective solution for Mach number prediction tasks^[12].

2 Key Parameters and Test Conditions of Continuous Wind Tunnel

The focus of this study is a 0.6-meter transonic wind tunnel designed for the development of large aircraft. To enhance the accuracy of Mach number prediction, it is crucial to know enough about the current wind tunnel and fully analyze it in order to build a more accurate Mach number prediction model. This paper introduces the aerodynamic structure and operating processes of a wind tunnel, analyzes the key variables and characteristics, and summarizes the challenges in modeling Mach number for continuous wind tunnels. The paper also provides modeling ideas on this topic.

2.1 Key Parameters and Effects of Continuous Wind Tunnels

The flow field in a continuous wind tunnel is a complex, multivariable system. Based on the above

continuous wind tunnel aerodynamic structure and process flow analysis, it is clear that the compressor speed (S) is the primary control variable, while the total pressure of the stable section (P_0) and the model angle of attack (A_n) serve as disturbing quantities. The Mach number (Ma) of the test section is the most significant controlled variable in the flow field system, primarily derived from the stable section's total pressure (P_0) and the test section's static pressure (P_s). It is crucial to calculate Ma precisely and efficiently for accurate results.

In wind tunnel testing, the Mach number of the test section holds significant importance, where the aircraft model is carried to simulate the aircraft flying in the sky. For the time being, there is no accurate device to measure the Mach number, and it is necessary to be calculated indirectly. It is evident that the Mach number at a certain moment is coupled with the total and static pressures of the system, and a change in the total and static pressures will affect the Mach number at the next moment, so the static and total pressures are selected as input variables for the magnitude prediction of the Mach number in the wind tunnel system. Also, Mach number prediction as a continuous time series problem, the Mach number values of previous time series should be considered.

The angle of attack (A_n) is the angle between the aircraft model and the airflow velocity, as shown in Fig.1. Wind tunnel testing of the variable A_n requires the change of the aircraft model running angle after the Mach number becomes stability to ensure that the Mach number is stable within the required accuracy. Therefore, changing of A_n affects the value of Mach number.

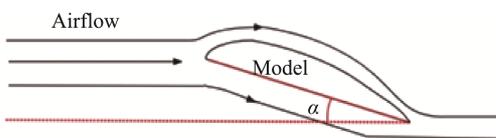


Fig.1 Diagram of Model Angle of Attack

The stability of Mach number is directly controlled by the size of the speed. Therefore, the compressor speed

is also considered as an input variable.

The wind tunnel analyzed in this paper operates mainly in the transonic section, which requires a very high Mach number control accuracy of $|\Delta Ma| \leq 0.001$.

In summary, the model constructed in this paper extracts compressor speed (S), total pressure (P_0), static pressure (P_s), model angle of attack (A_n) and Mach number (Ma) of previous time series as prediction inputs. The Mach number prediction results of corresponding moments are obtained through the processing of the constructed model.

2.2 Continuous Wind Tunnel Test Conditions

The wind tunnel necessitates numerous parameter settings prior to performing variable angle of attack tests. To commence, the Mach number must be specified for current operational conditions. Distinct Mach number values for setting result in varying gas flow rates in the wind tunnel. Consequently, alterations in the angle of attack of the model produce divergent impacts on the Mach number and the system characteristics. According to the different total pressure size, there are three forms of pressure: negative pressure, atmospheric pressure, and positive pressure. The total pressure setting for the negative pressure operation is 50KPa, which is also the lowest pressure at which the wind tunnel can be operated, and usually the total pressure changes less in this pressure state. The standard operating pressure is typically the normal pressure, with a total pressure setting of 100KPa, which is also the pressure utilized in the majority of wind tunnel tests. The total pressure setting for boosted operation is 150KPa. Abbreviated technical terms shall be explained when first used. Additionally, the wind tunnel system includes features such as blade angle, pilot slit, and opening/closing ratio. The conditions' differences bring changes to the flow field characteristics and produce new working conditions. Moreover, the Mach number experiences varying effects at different angle of attack speeds. During the continuous variable angle of attack test, the greater the angle of attack change speed, the more the Mach number experiences perturbation.

In this paper, three typical operating conditions were selected with adequate data accumulation, and the relevant data were used to establish a prediction model for the wind tunnel system's Mach numbers. The operating parameters of Mach numbers studied in this paper fall between 0.8 to 0.9 in the transonic range, where the pilot slit is set to 24mm and the opening/closing ratio is set to 1.5%: The total pressure is set at 100 units. The blade angle is primarily set to 76.5° and 56.3° , and the rate of change in angle of attack is set to $0.1^\circ/s$ and $0.2^\circ/s$, respectively. The working conditions are as outlined in Table 1.

3 Prediction Model Building for Mach Number

Based on the aforementioned research on the wind tunnel flow field modeling and the analysis of the aerodynamic structure of the continuous wind tunnel studied in this paper, this chapter aims to investigate the main factors impacting the Mach number. The primary focus is on predicting the Mach number during the modeling test phase under typical operating conditions, addressing the challenges in modeling the continuous wind tunnel system.

3.1 1D-CNN Model

Fig.2 illustrates the network architecture of 1D-CNN. Once the input signal passes through the input layer and reaches the convolutional layer, feature extraction occurs through one-dimensional convolutional kernels. Local regions of the input signal are convolved to extract different feature signals through the convolutional layer's different kernels^[13].

For the convolutional layer ($l-2$), the output can be expressed as:

$$x_j^{l-2} = f\left(\sum_{i=1}^M k_i^{l-3} * w_{ij}^{l-2} + b_j^i\right) \quad (1)$$

where, k denotes the convolution kernel, j denotes the number of convolution kernels, M denotes the number of channels of input x^{l-1} , b denotes the bias corresponding to the kernel, f denotes the activation function, and $*$ denotes the convolution operator.

The convolutional layer extracts feature signals, which are then passed to the pooling layer for feature dimensionality reduction, simplifying the computational complexity of the network. If the final pooling layer is the ($l-1$) layer, it passes its output as input to the fully connected layer, and the output of the fully connected layer is:

$$y_i^l = f(w^{l-1} \cdot x^{l-1} + b^{l-1}) \quad (2)$$

Table 1 Working Conditions of the Wind Tunnel Flow Field

Test Conditions	Mach Number	Total Pressure (KPa)	Angle of Attack Speed ($^\circ/s$)	Number of Samples
1	0.8	100	0.1	1556
2	0.85	100	0.1	1632
3	0.9	100	0.2	752

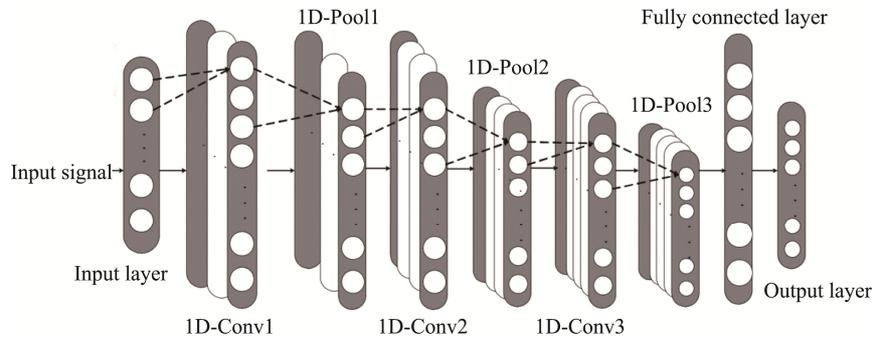


Fig.2 One-dimensional Convolution Neural Networks

where, w represents the weight and b represents the bias. Finally, the fully connected layer is connected to the softmax layer to do classification operation.

Error between the output layer and the expected result is passed back to the network, and the fully connected layer, the pooling layer and the convolutional layer errors are obtained in turn. Weights and thresholds are updated by calculating the error gradient until the error tolerance is satisfied to complete the training. The mean square error of the input vector a in the output layer is:

$$E_a = \sum_{i=1}^{N_l} (y_i^l - t_i^a)^2 \quad (3)$$

where y_i^l is the output finally of the network and t_i^a is the target output of vector a . In order to find the derivative of E_a , the incremental error should first be calculated as follows:

$$\Delta_k^l = \frac{\partial E}{\partial x_k^l} \quad (4)$$

The incremental error is utilized to modify the weights and biases of each neuron through gradient descent. Additionally, the incremental error is used to update the bias of the neuron and all the weights of the neurons connected to the previous layer of the neuron, which is defined as follows:

$$\frac{\partial E}{\partial w_{ik}^{l-1}} = \frac{\partial E}{\partial x_k^l} \cdot \frac{\partial x_k^l}{\partial w_{ik}^{l-1}} = \Delta_k^l y_i^{l-1} \quad (5)$$

$$\frac{\partial E}{\partial b_k^l} = \frac{\partial E}{\partial x_k^l} \cdot \frac{\partial x_k^l}{\partial b_k^l} = \Delta_k^l \cdot 1 = \Delta_k^l \quad (6)$$

The execution is then back-propagated from the fully connected layer to the pooling layer.

$$\frac{\partial E}{\partial s_k^l} = \Delta s_k^l = \sum_{i=1}^{N_{l+1}} \Delta_i^{l+1} \cdot w_{ki}^l \quad (7)$$

It can be obtained that Δs_k^l is provided by n_{l+1} neurons in layer $l+1$ to provide the incremental error between layers.

$$\Delta s_k^l = \sum_{i=1}^{N_{l+1}} \text{conv1Dz}(\Delta_i^{l+1}, \text{rev}(w_{ki}^l)) \quad (8)$$

where $\text{rev}(\cdot)$ denotes the inverse array and $\text{conv1Dz}(\cdot)$ denotes the full convolution.

Traditional deep learning models typically require a large number of samples, whereas 1D-CNN can achieve model training with limited samples. In addition, 1D-CNN can commence model training without any predefined transformation, such as manual feature extraction and feature selection. Lastly, the 1D

convolution of 1D-CNN's compact architecture configuration offers cost-effective features and straightforward hardware implementation, which are particularly suitable for Mach number prediction. Therefore, this paper utilizes a 1D-CNN as the foundational architecture to construct the model for predicting Mach numbers^[14].

3.2 LSTM Model

The Long Short-Term Memory (LSTM) is a neural network capable of retaining both long- and short-term information^[15]. Unlike traditional Recurrent Neural Network (RNN) nodes, the output of LSTM nodes is determined by the weights, bias, and activation function. The RNN has a chain structure where each time slice employs the same parameters, as shown in Fig.3.

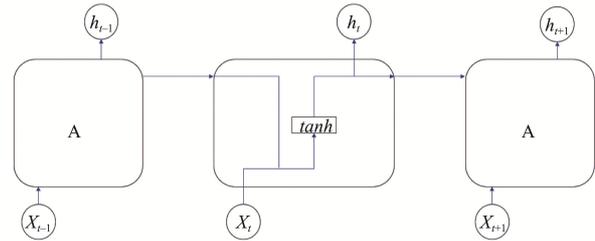


Fig.3 RNN Unit

The LSTM introduces the gate mechanism to control feature circulation and loss, effectively resolving the long-term dependency problem of RNNs. The LSTM consists of a sequence of LSTM units arranged in a chain structure, as shown in Fig.4:

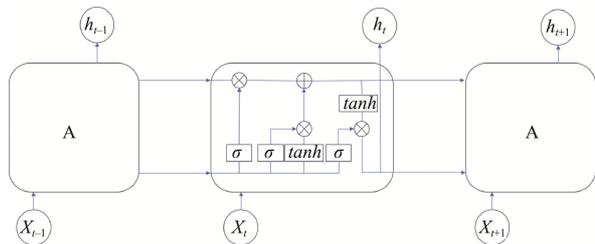


Fig.4 LSTM Unit

The following is a specific description of how it works:

1. The core of LSTM is the uppermost part of Fig.4, which is similar to the conveyor belt, and is generally called cell state. Among them:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (9)$$

where f_t is known as the forgetting gate and signifies

which features of C_{t-1} are utilized to compute C_t . f_t is a vector with each element in the range of $[0, 1]$. The activation function commonly used is the sigmoid function, and its output is a value within the interval of $[0, 1]$. Nevertheless, upon examining a trained LSTM, it is evident that the bulk of the gate values are extremely near 0 or 1, and the remainder are sporadic. Where the product gate mechanism, \otimes , is the most crucial aspect of the LSTM. The specific expression for h is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (10)$$

2. The LSTM's unit state update value, \tilde{C}_t , derives from x_t and the prior hidden node h_{t-1} via a neural network layer. Typically, the activation function for \tilde{C}_t is tanh. The specific expressions of \tilde{C}_t are as follows:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (11)$$

i_t is known as the input gate, and similar to f_t , it is a vector with components in the range $[0, 1]$, also computed by x_t and h_{t-1} via the sigmoid activation function. The specific expression for i_t is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (12)$$

i_t is used to control which features of \tilde{C}_t are used to update C_t , and is used in the same way as f_t . The expression for updating C_t is as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (13)$$

3. Finally, in order to compute the prediction \hat{y}_t and generate the complete input for the next time unit, it is necessary to compute the output h_t of the hidden node:

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (14)$$

$$h_t = o_t * \tanh(C_t) \quad (15)$$

h_t is acquired from the output gate o_t and the cell state C_t . The calculation for o_t follows the same process as f_t and i_t .

3.3 Attention Mechanism

The attention mechanism is a distinct construct that is integrated into machine learning models to autonomously learn and compute the weight of input data towards output data.

Based on the original framework, an attention module is added to predict the weight of each input data dimension during the current time step. This is done by utilizing both the input data and the historical output of the neurons. Then, based on this weight, the input data is weighted and summed to obtain the

current time step. This value is then fed to the neuron's last input, along with the historical output, enabling the neuron to compute the current time step's output. With the inclusion of the attention module, the model focuses more on essential data, resulting in an improved training effect.

In this paper, the Squeeze-and-Excitation (SE) module, Efficient Channel Attention (ECA) module, and Convolutional Block Attention Module (CBAM) are applied, and these modules are described below:

SE module: the purpose of the SE module is to extract more significant feature information by assigning varying weights to different image positions based on the channel domain perspective through a weight matrix^[16]. Fig.5 displays the algorithm flowchart of the SE module.

Step 1 Transformation: Given the input data X , the feature data U is generated by convolution operation.

Step 2 Squeeze: This step pools the feature map globally, generating a $1 \times 1 \times C$ vector, so that each channel can be represented by a value. The specific formula is as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (16)$$

Step 3 Excitation: This step is done by two fully connected layers to generate the required weight information by the weights W , where W is obtained by learning to show the feature relevance in the model. The specific formula is as follows:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = (W_2 \delta(W_1 z)) \quad (17)$$

Step 4 Scale: The third step generates the weight vector s to allocate weights to the feature data U to get the desired feature data, whose size is the same as the feature data U . The SE module does not change the small size of the feature data. The specific formula is as follows:

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c u_c \quad (18)$$

ECA module: the SE attention mechanism initially compresses the input feature map's channels, which unfortunately results in a reduction of dimensionality that impairs the dependencies learned between channels. Following this idea, the ECA attention mechanism avoids dimensionality reduction and efficiently implements local cross-channel interaction with 1-dimensional convolution to extract the dependencies between channels^[17]. The flowchart of the ECA attention mechanism is shown in Fig.6:

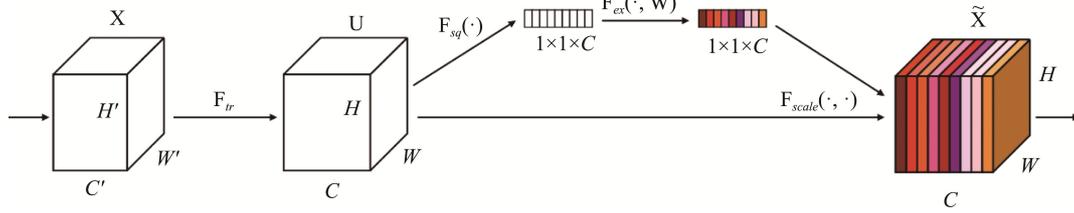


Fig.5 SE Module Algorithm Flow Chart

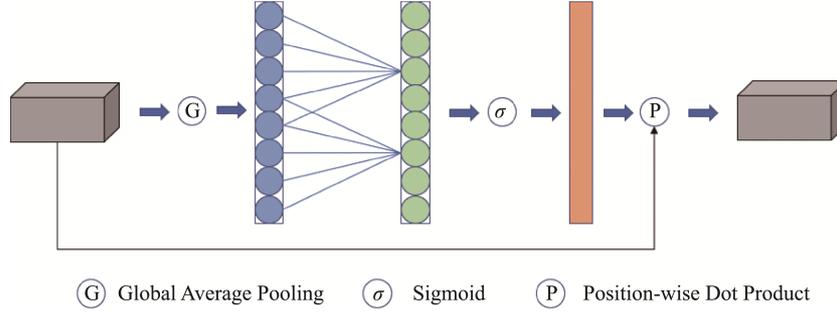


Fig.6 Flow Chart of ECA Attention Mechanism

The steps are as follows:

Step 1: Conduct a global average pooling operation on the input feature data;

Step 2: Perform a 1-dimensional convolution operation with a convolution kernel size of k and obtain the weights w of each channel after applying a Sigmoid activation function, as shown in Eq (19):

$$\omega = \sigma(C1D_k(y)) \quad (19)$$

Step 3: The corresponding elements of the original input feature data are multiplied with the weights to generate the final output feature data.

CBMA module: The Convolutional Block Attention Module represents the attention mechanism module of the convolutional module, which combines spatial and channel attention^[18].

The details of the channel attention module are illustrated in Fig.7:

The input features undergo global max pooling and global average pooling, and then pass through MLP,

respectively. The channel attention feature map and input feature map are multiplied element-wise to generate the required input features for the Spatial attention module. The channel attention mechanism is expressed below:

$$M_c(F) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (20)$$

The details of the spatial attention module are illustrated in Fig.8:

The feature data yielded by the Channel Attention module functions as the input feature data of this module. The procedure begins with executing global max pooling and global average pooling based on the channel, followed by applying a concept operation based on the channel to the two outcomes. The spatial attention feature is generated via sigmoid, and finally, the feature is multiplied by the input feature of the module to obtain the final generated feature. The spatial attention mechanism is expressed as follows:

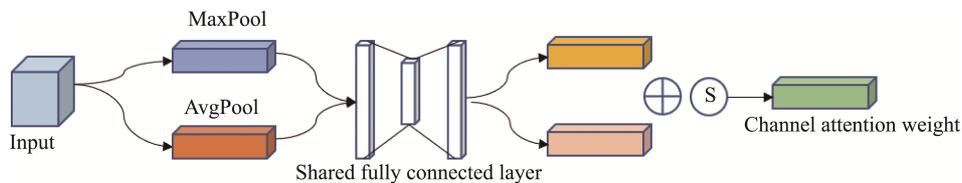


Fig.7 Channel Attention Module

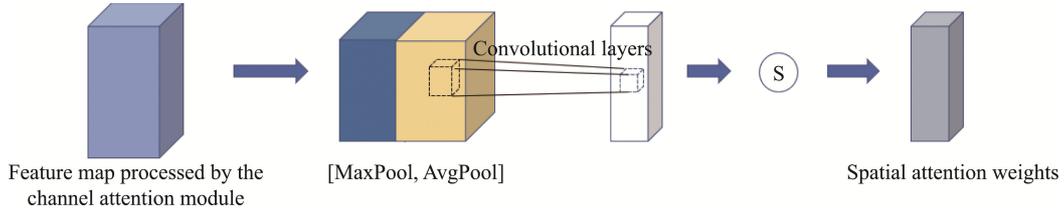


Fig.8 Spatial Attention Module

$$M_s(F) = \sigma \left(f^{7 \times 7}([F_{avg}^s; F_{max}^s]) \right) \quad (21)$$

3.4 CNN-LSTM-Attention Mach Number Prediction Model

Aiming at the operational characteristics of the wind tunnel system, this section mainly determines the model input and output according to the key physical quantities determined by the wind tunnel test process, extracts the features of the data by using the 1D-CNN algorithm, and then establishes a CNN-LSTM Mach number prediction model with LSTM model to predict and analyze the Mach number under standard test conditions and provide evaluation criteria for the model's performance in conjunction with the wind tunnel system's actual situation.

Based on the analysis of the continuous wind tunnel structure and key variables in Chapter 2, along with expert experience, the input parameters were selected as total pressure (P_0), rotational speed (S), and angle of attack (A_n), which play a significant role in determining the Mach number. Mach number (Ma) is considered the output parameter that needs to be predicted. Based on the working conditions outlined in Table 1, this paper examines Mach number variation across the speed of

sound within the range of 0.8 to 0.9. Additionally, the total pressure is set at 50KPa, 100KPa, and 150KPa.

The wind tunnel system normalizes its total pressure, speed, angle of attack, and Mach number, resulting in P_{0nor} , S_{nor} , A_{nnor} and Ma_{nor} . Fig.9 shows the Mach number CNN-LSTM-Attention prediction model:

The input layer defines the format of the input data as (batch size, time step, feature dimension). By default, set the batch size as 1, denote the time step as t , and the feature dimension is represented as n . A sample can be represented as a real sequence matrix $R_{t \times n}$, where the vector representation of the data at the i -th time step in matrix $R_{t \times n}$ is denoted as x_i .

The second layer is the CNN layer, extracting the spatial association between various feature values in the data, compensating for LSTM's failure to capture the data's spatial component, while the extracted features are still temporal in nature. The sample data enters the CNN layer and undergoes convolution, pooling, and node expansion (dimensionality reduction) operations in turn. For sequential data, this model takes one-dimensional convolution, and the convolution kernels are convolved in a single time-domain

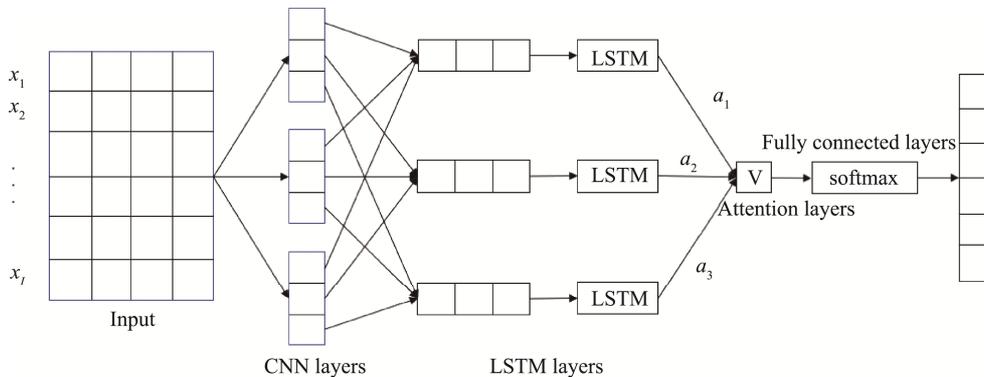


Fig.9 CNN-LSTM-Attention Prediction Model of Mach Number

direction only. The size of the convolutional kernels is set to k , and the number of kernels is r . Then $x_{i:i+k-1}$ is a real matrix from the i -th time step to the $(i+k-1)$ -th time step in $R_{t \times n}$ with a sliding step of 1. The weight matrix W_l is a $k \times n$ real matrix. The feature o_i can be obtained by calculating the sequence vector once for each k time steps, and it is calculated as below:

$$o_i = f(W_l \otimes x_{i:i+k-1} + b_1) \quad (22)$$

Where a nonlinear activation function f and bias b_1 (where $b_1 \in R$) are utilized. When a convolution kernel has extracted the sequence data of a sample, a $(t-k+1) \times 1$ shaped feature o , of the following form, is obtained:

$$o = [o_1, o_2, \dots, o_{t-k+1}]^T \quad (23)$$

The CNN has a total of r convolution kernels, so r feature maps will be obtained in the end. The convolution is then followed by a maximum pooling operation with a pooling size of 2 and a sliding step of 2 to get a feature matrix of $r \times [(t-k+1)/2] \times 1$ shapes.

These r feature maps are the features extracted by the CNN layer, which are downscaled into a real vector of length $r \times (t-k+1)/2$. This vector preserves the spatial relationship between various feature values from the sample data. Finally, the vector is passed into the LSTM layer for further processing.

The third layer, the LSTM, has a memory function to extract information about the temporal variation of the nonlinear data of the building cooling and heating loads. It introduces input gates, forgetting gates, and output gates, and adds candidate states, cellular states, and hidden states. The cell state stores long-term memories that can prevent gradient disappearance, while the hidden state stores short-term memories. This model uses a multi-layer LSTM, where the output from the previous LSTM layer serves as the input for the subsequent layer, passing down one layer at a time. The output from the final LSTM hidden layer proceeds to the attention layer for additional processing.

The fourth layer is the attention layer, which enhances the contribution of crucial time steps and minimizes prediction errors. Attention refers to a calculated sum of the output vectors from the last LSTM layer, weighted by their significance. The

output vector of the LSTM hidden layer functions as the input for the attention layer and is trained using a fully connected layer. The output of the fully connected layer is then normalized by the SoftMax function to determine the assigned weights for each hidden layer vector. The weight size denotes the significance of the hidden state at each time step for the predicted outcome. The training process for the weights is as follows:

$$S_i = \tanh(WH_i + b_i) \quad (24)$$

$$\alpha_i = \text{soft max}(S_i) \quad (25)$$

The trained weights are then used to compute the weighted average sum of the hidden layer output vectors, which is calculated as follows:

$$C_i = \sum_{i=0}^k \alpha_i H_i \quad (26)$$

Where H_i represents the output of the last LSTM hidden layer, S_i represents the score of each hidden layer output, α_i represents the weight coefficient, C_i represents the outcome after weighted summation, and SoftMax is the activation function.

The fifth layer, output layer, specifies the prediction time step o_t . The final output is the prediction result of step o_t .

3.5 Model Evaluation Metrics

For high-precision control systems of wind tunnel flow fields, modeling accuracy has a direct impact on the stability of subsequent control Mach numbers. In order to better assess the level of modeling, a suitable criterion is needed to analyze the model. Generally, the accuracy of Mach number predictions is evaluated by the root mean square error (RMSE), which quantifies the difference between predicted and actual values. A lower RMSE indicates a closer predicted value to the true value, thus a more precise model.

$$RMSE = \sqrt{\frac{1}{K} \sum_{k=1}^K (Ma(k) - \bar{Ma}(k))^2} \quad (27)$$

For wind tunnel systems, the ultimate goal of model prediction is to perform Mach number control. In general, the Mach number must remain stable around a set value, which is denoted as Ma_{set} . The model's evaluation entails both accuracy prediction and the impact on the control effect. To evaluate the model's control effect, the Mach number accuracy (A) and the

maximum deviation (MD) are usually also taken for evaluation.

(1) Mach number accuracy (A) measures the deviation between the set value and the average predicted value during prediction. The lower the A during testing, the better the model can represent the overall changes in the Mach number. The equation for Mach number accuracy is as follows:

$$A = \left| Ma_{set} - \frac{1}{K} \sum_{k=1}^K \bar{Ma}(k) \right| \quad (28)$$

(2) The maximum deviation (MD) represents the highest predicted deviation between actual and predicted values during the prediction process. As an important performance indicator in the Mach number control accuracy, the maximum deviation reflects whether stability requirements have been met and whether further improvement of the control scheme is needed. The maximum deviation formula is as follows, and the maximum deviation should not exceed 0.001 in value.

$$MD = \max |Ma(k) - \bar{Ma}(k)| \quad (29)$$

4 Simulation Experiments and Analysis

In this section, the proposed CNN-LSTM-Attention algorithm is applied to a wind tunnel process with the aim of predicting the Mach number. The common CNN-LSTM model is analyzed and compared with the PLS model and BP neural network model, based on three typical working conditions selected in the wind tunnel system.

4.1 Key Parameters of the Single Working Condition Mach Number Prediction Experiment Show

For the typical operating conditions in Table 1, 70% of the data for each test condition are chosen for training while 30% for testing. The prediction of Mach number is carried out using the CNN-LSTM-Attention model, and its accuracy is compared with both the PLS regression prediction model and the BP neural network model introduced in Section 2.1.

The following is the flow chart of the for CNN-LSTM-Attention models:

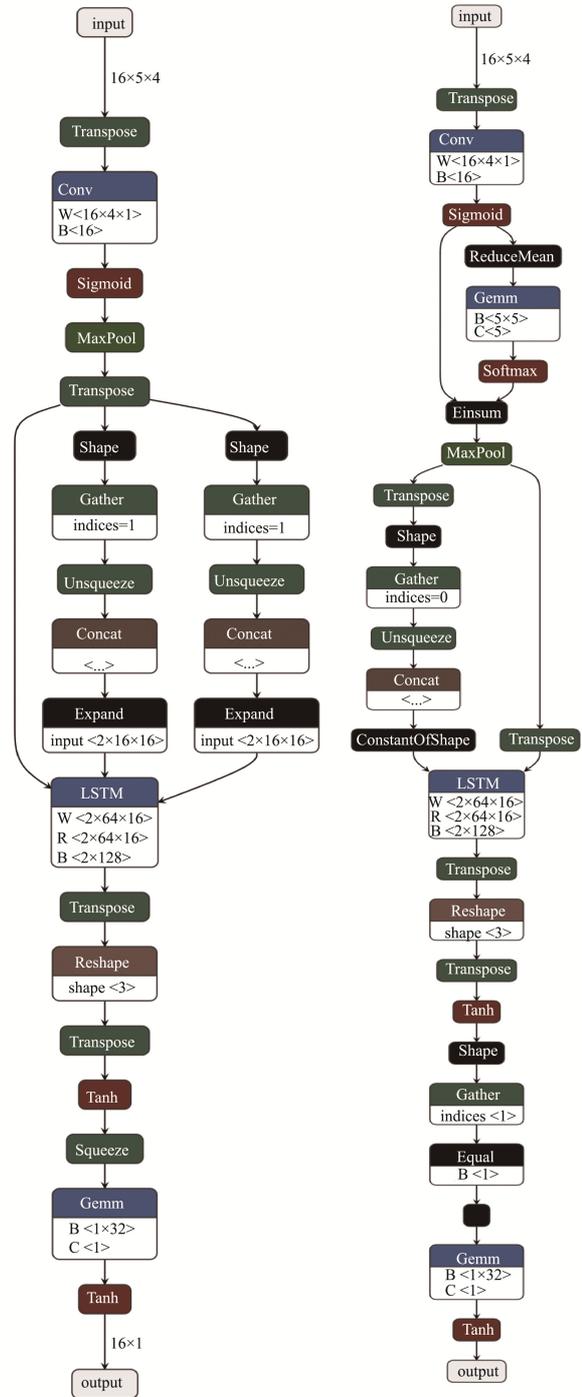


Fig.10 Flowchart Showing the Base Model (left) and SE Model

The following is an example of the Mach 0.9 working condition training process, showing the values of the key parameters:

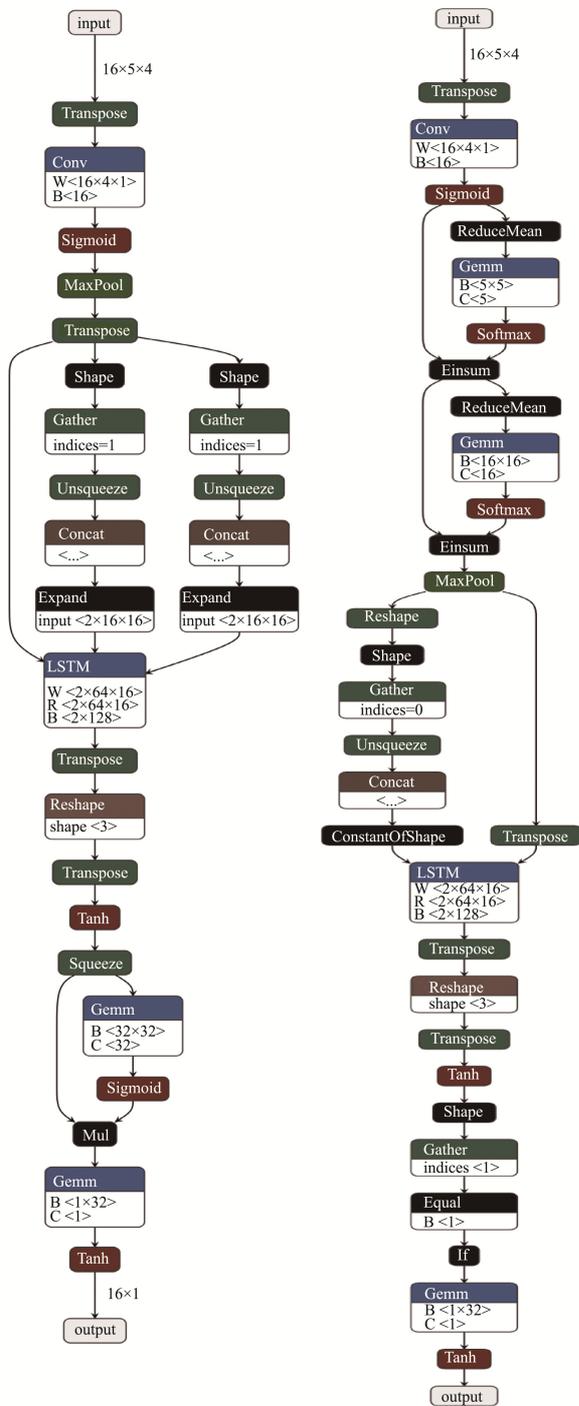


Fig.11 Flowchart Showing the ECA Model (left) and CBAM Model

Among them, *conv1d.weight* and *bias* represent the weights and biases obtained from training in 1D-CNN; the eight parameters involved in LSTM are the weights and biases in Eqs. (10), (11), (12), and (14),

respectively. *se_fc*, *hw_fc*, and *attn* represent the weights and biases of the corresponding attention mechanisms.

For the Base model the values of the key parameters are shown as below in Fig.12:

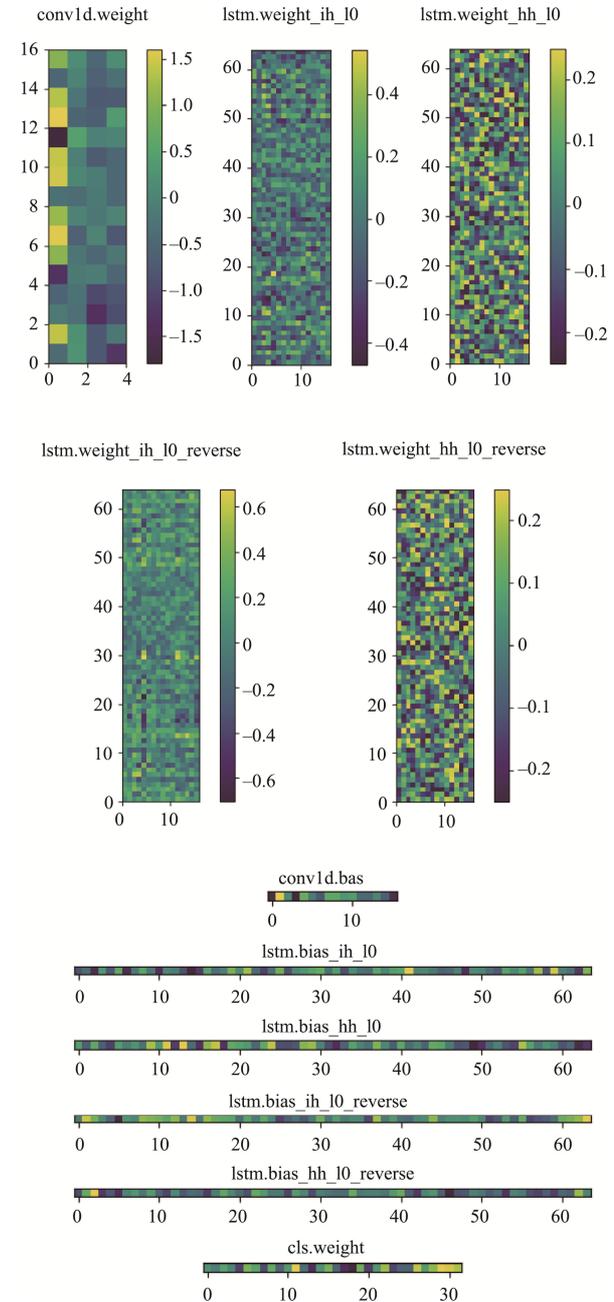


Fig.12 Values of the Key Parameters for the Base Model

For the SE model the values of the key parameters are shown as below in Fig.13:

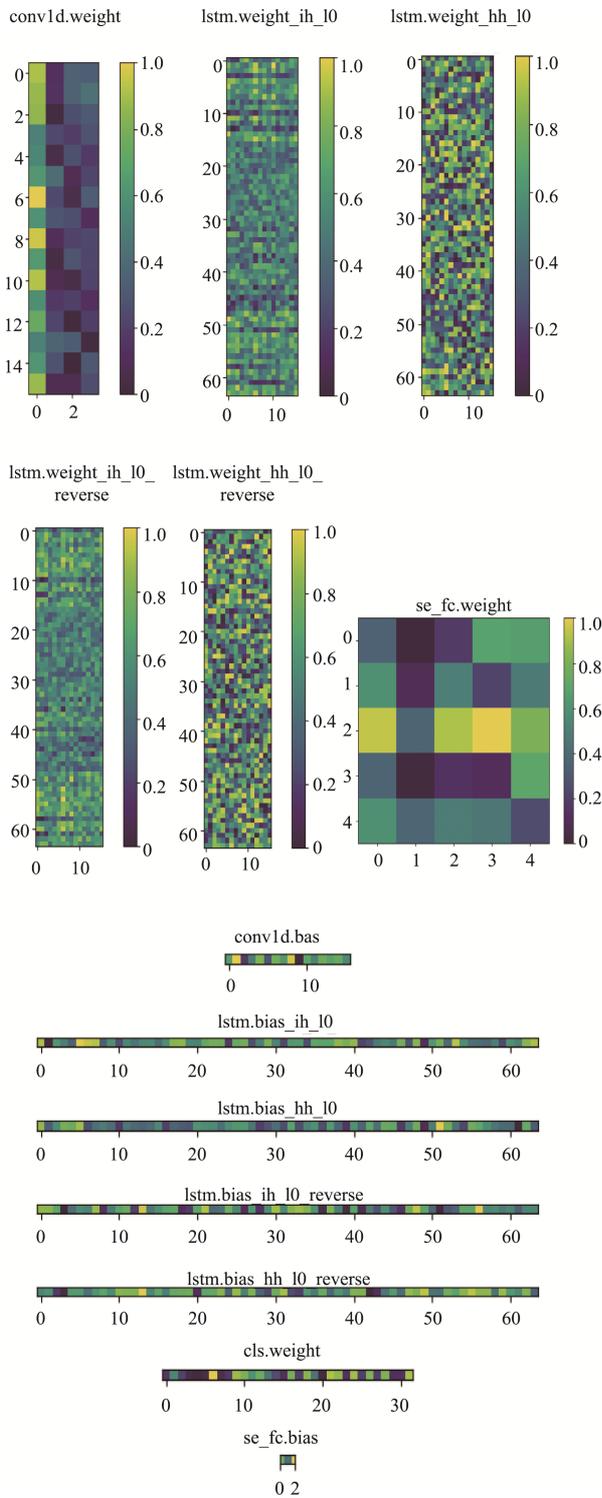


Fig.13 Values of the key Parameters for the SE Model

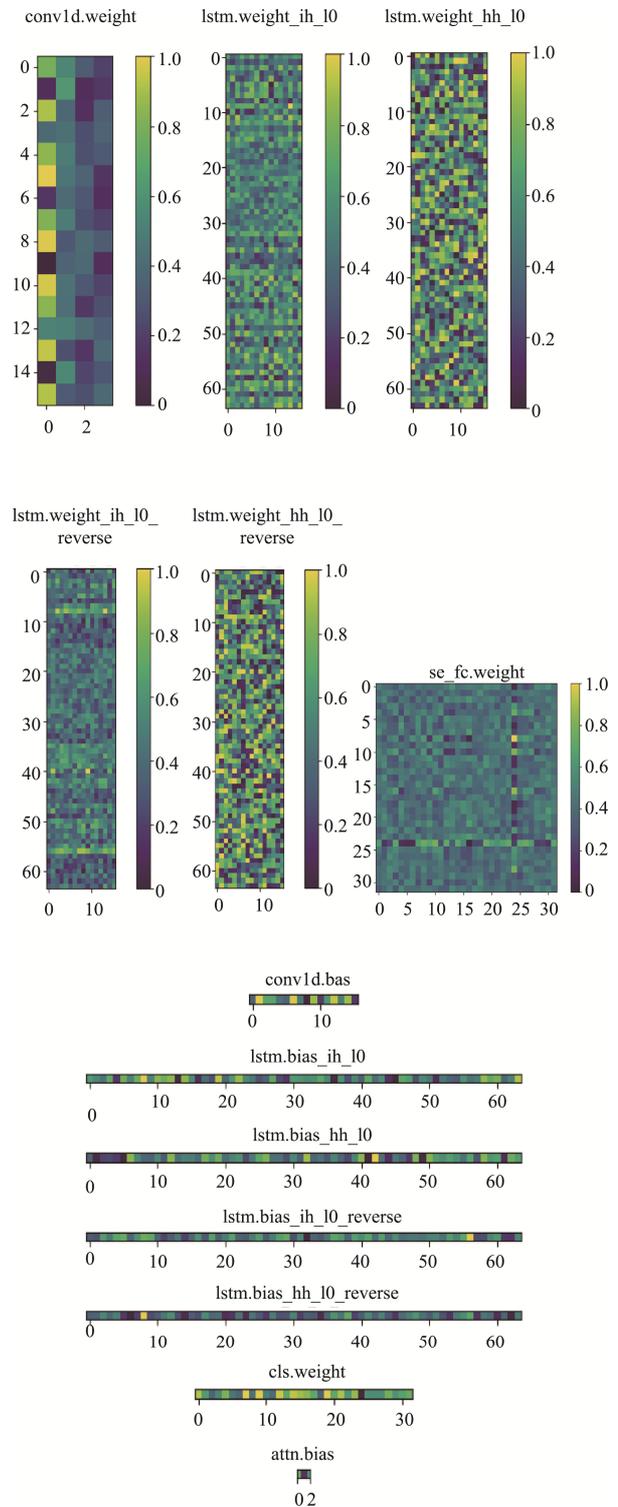


Fig.14 Values of the Key Parameters for the ECA Model

For the ECA model the values of the key parameters are shown as below in Fig.14:

For the CBAM model the values of the key parameters are shown as below in Fig.15:

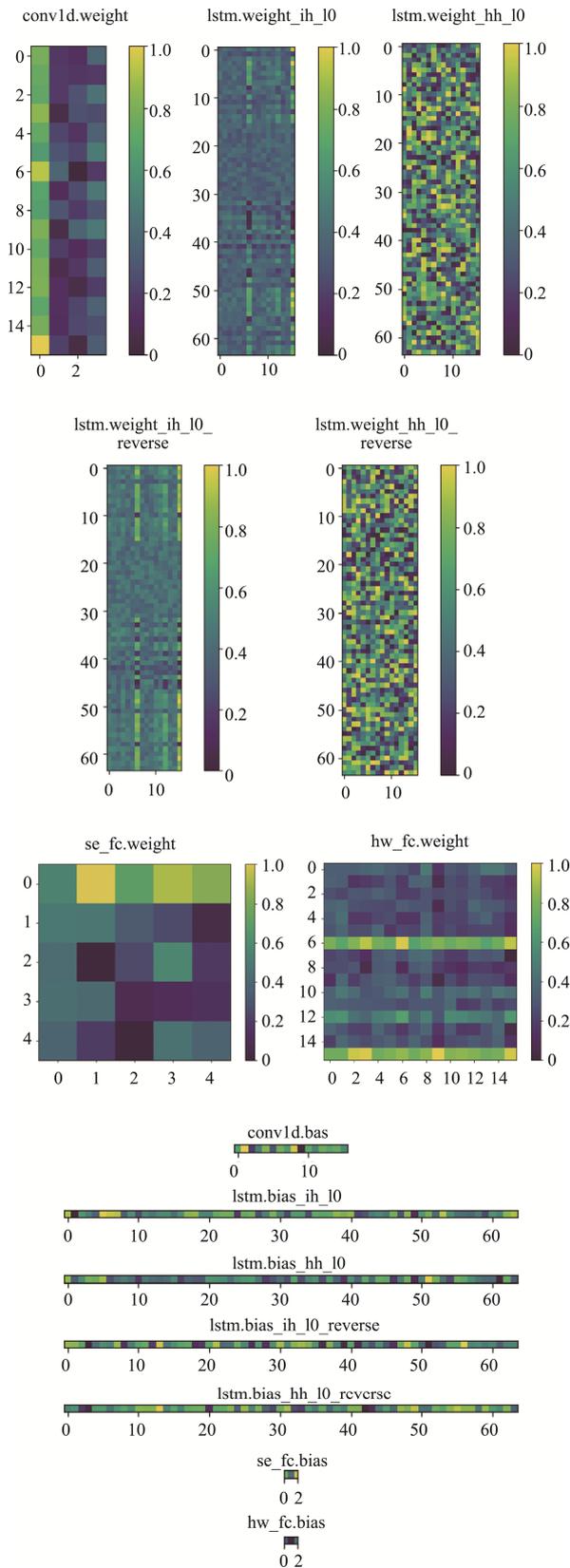


Fig.15 Values of the Key Parameters for the CBAM Model

4.2 Experimental and Comparative Analysis of Mach Number Prediction

According to the Mach number prediction results in Fig.16, the CNN-LSTM-Attention model more accurately predicts the trend of Mach number variation compared to the BP neural network and PLS methods, which also predict the trend with respect to actual Mach number. Across-sectional comparison of the CNN-LSTM-Attention model shows that the fluctuation of the training of the model becomes correspondingly smaller and closer to the actual Mach number after adding the attention mechanism; Meanwhile, the best performance is achieved on this typical working condition after adding the SE module, and the prediction curve fits the actual curve better, but fluctuates frequently, and the CBAM module is less volatile and the curve is more obedient.

To demonstrate the model's superiority, six modeling methods are used for Mach number prediction for each working condition. The model evaluation indexes for each test working condition are obtained, and the specific simulation results are shown in Table 2. For the sake of simplicity, the name base model is used instead of the name CNN-LSTM model.

By analyzing Table 2, the predicted RMSEs obtained from PLS modeling and BP neural network modeling as well as CNN-LSTM-Attention modeling are all smaller than 0.001. The predicted Mach number RMSEs from CNN-LSTM-Attention modeling are smaller than those from PLS modeling and BP neural network modeling under the typical working conditions in this paper. The RMSE indicates that the predicted Mach number in the CNN-LSTM-Attention modeling is in closer agreement with the actual Mach number. In summary, the proposed CNN-LSTM-Attention model offers more advantages over the PLS model and BP model. In a side-by-side comparison of the CNN-LSTM models, the training effect of the CNN-LSTM-Attention model surpasses that of the basic CNN-LSTM model. These findings suggest that the CNN-LSTM-Attention prediction model could potentially enhance the tracking and prediction capabilities of the Mach number; Meanwhile, the CNN-LSTM-SE model during the training process performs optimal.

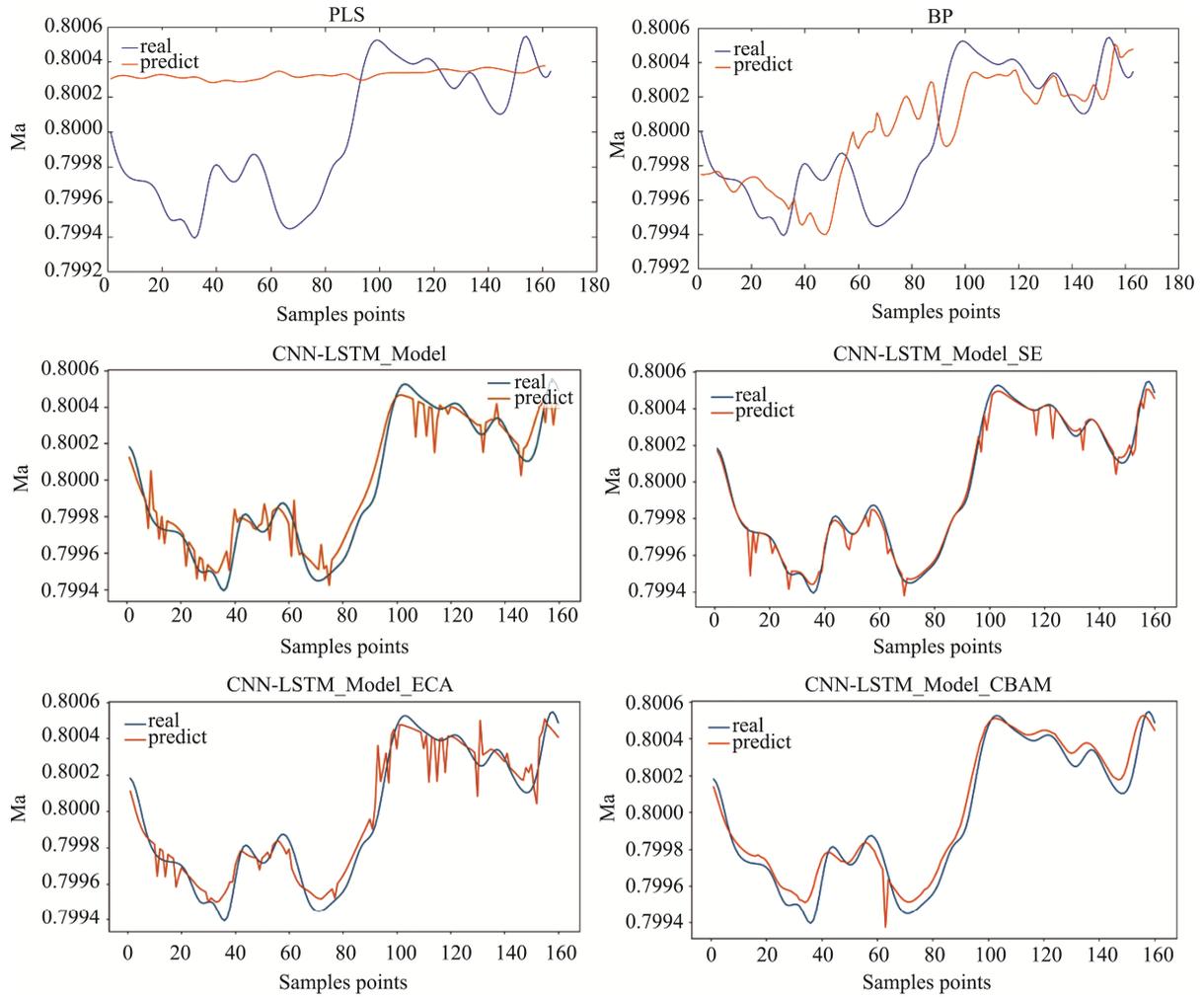


Fig.16 Prediction Curves of 0.8 Mach Number Working Conditions

Table 2 Evaluation Indexes of Mach Number Prediction Models for the 0.8 Mach Number Working Conditions

PLS Model			BP Model		
RMSE	A	MD	RMSE	A	MD
5.20e-04	1.80e-04	1.39e-03	3.00e-04	1.50e-04	8.10e-04
BASE Model			BASE-SE Model		
RMSE	A	MD	RMSE	A	MD
8.85e-05	1.52e-05	2.53e-04	5.34e-05	1.33e-05	4.35e-05
BASE-ECA Model			BASE-CBAM Model		
RMSE	A	MD	RMSE	A	MD
9.76e-05	0.95e-05	3.69e-04	7.19 e-05	4.73e-06	3.37e-04

The analysis of the experimental phenomena presented in Fig.16 shows that the SE attention mechanism performs better in terms of performance and efficiency at 0.8 experimental working conditions. It is characterized by fewer parameters, high flexibility, and achieved significant performance improvement in this experiment. Although the ECA and the CBAM atten-

tion mechanisms also show good performance in some cases, they do not outperform the SE attention mechanism relative to the SE attention mechanism in the typical operating conditions of this experiment. The following experiments are conducted for the other two typical working conditions and the experimental results are shown in Fig.17, Fig.18, Table 3 and Table 4.

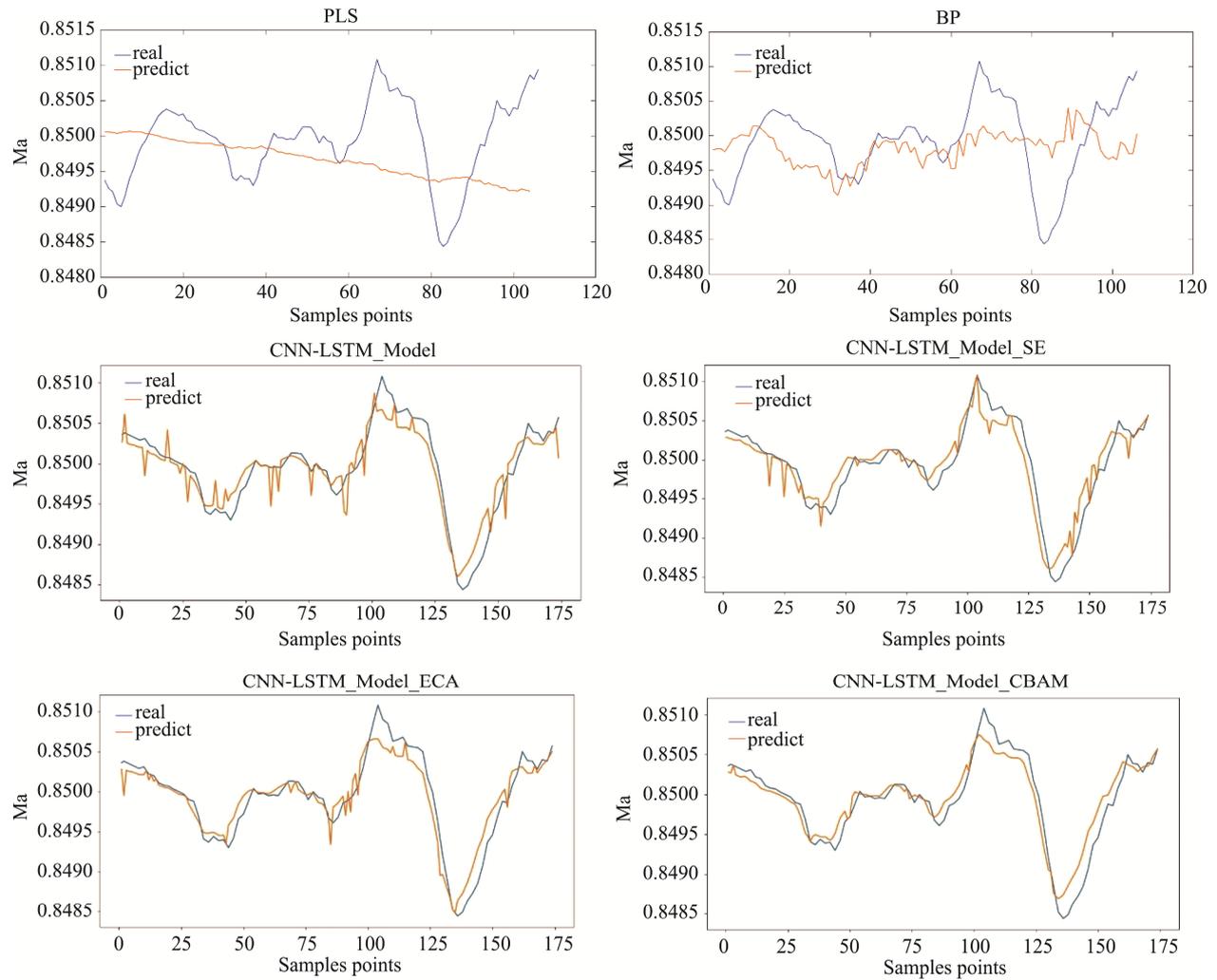


Fig.17 Prediction Curves of 0.85 Mach Number Working Conditions

Table 3 Evaluation Indexes of Mach Number Prediction Models for the 0.85 Mach Number Working Conditions

PLS Model			BP Model		
RMSE	A	MD	RMSE	A	MD
0.0007	0.0003	0.0016	0.0006	0.0002	0.0014
BASE Model			BASE-SE Model		
RMSE	A	MD	RMSE	A	MD
1.95e-04	7.73e-05	4.98e-04	1.91e-04	5.66e-05	4.77e-04
BASE-ECA Model			BASE-CBAM Model		
RMSE	A	MD	RMSE	A	MD
1.77e-04	6.88e-05	4.62e-04	1.71e-04	4.27e-05	4.67e-04

Comparing the experimental data obtained for the two new working conditions above shows, we can see these methods can be applied in different working conditions, but their effects may also vary in different working conditions due to their slightly different implementations and objectives. However, the

CNN-LSTM-Attention models all have significantly better results than the PLS and BP models, and they all fulfill the accuracy requirements for Mach number prediction. These results suggest that the CNN-LSTM-Attention models have better generalizability.

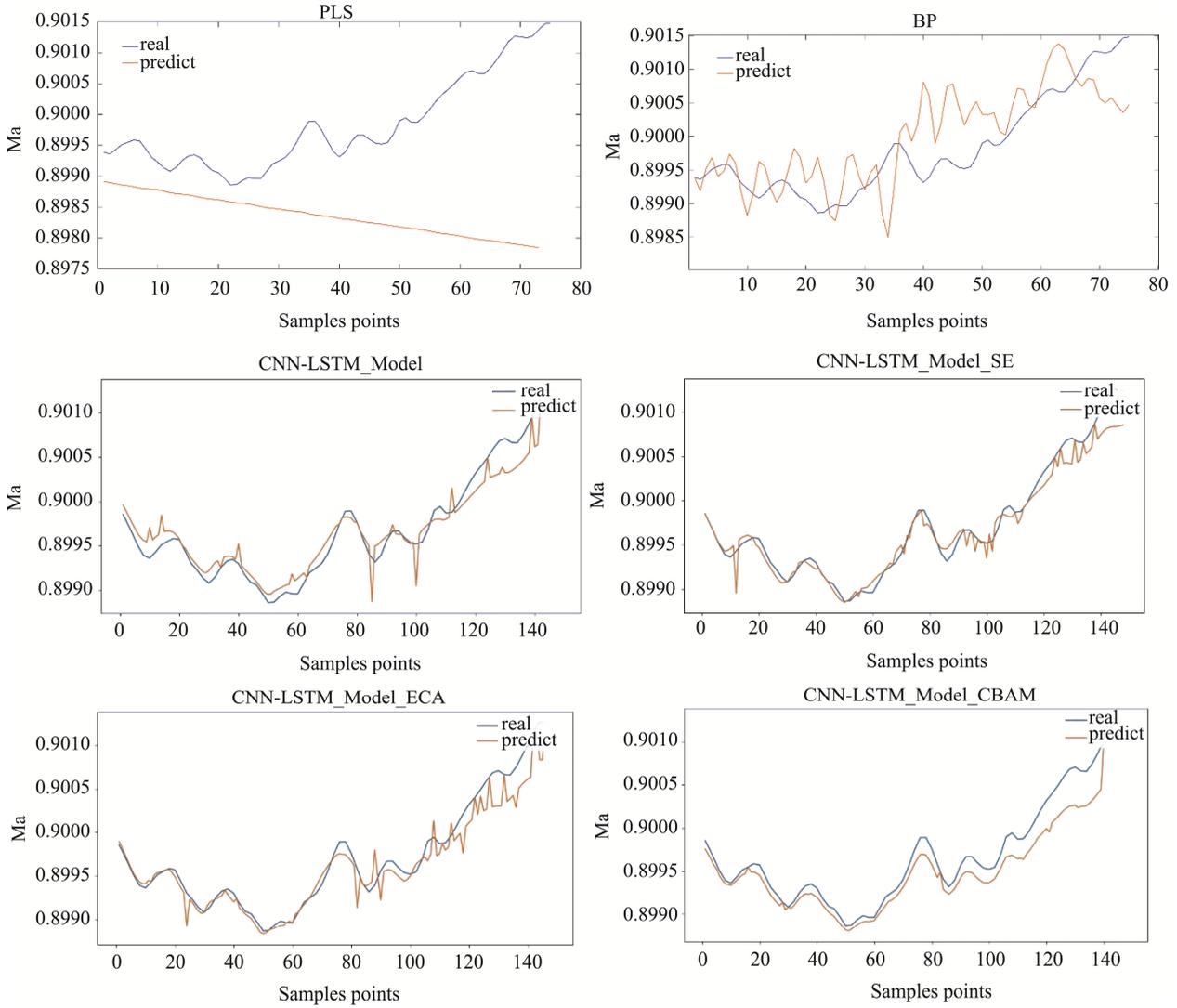


Fig.18 Prediction Curves of 0.9 Mach Number Working Conditions

Table 4 Evaluation Indexes of Mach Number Prediction Models for the 0.9 Mach Number Working Conditions

PLS Model			BP Model		
RMSE	A	MD	RMSE	A	MD
0.0017	0.0016	0.0036	0.0006	0.0007	0.0015
BASE Model			BASE-SE Model		
RMSE	A	MD	RMSE	A	MD
1.63e-04	2.81e-04	4.89e-04	1.38e-04	3.34e-04	4.77e-04
BASE-ECA Model			BASE-CBAM Model		
RMSE	A	MD	RMSE	A	MD
1.64e-04	3.58e-04	4.65e-04	1.93e-04	4.31e-04	4.98e-04

The SE attention mechanism focuses on the relationship among channels and adjusts their weights by learning their importance. The ECA attention

mechanism aims to enhance the computational efficiency of the SE attention mechanism, and is particularly suitable for data with many channels.

When the number of channels is small, the ECA attention mechanism may not improve the performance significantly because the relationships between channels are relatively simple. The CBAM attention mechanism utilizes cross-batch memory information to improve the feature representation of the model. Meanwhile, it assumes that the relationship between different samples is beneficial for feature representation.

Therefore, it can be concluded that the RMSE performance of the ECA model is inferior to that of the other two models because of the smaller input feature dimension. The CBAM model performs better in both the 0.8 and 0.85 working conditions where the data set is larger. The SE model has good performance in all three tasks and has better overall performance.

In conclusion, the effectiveness of SE, ECA and CBAM attention mechanisms may vary in different working conditions, depending on factors such as the feature relevance of the data, noise and redundancy, number of channels, computational resource constraints, batch size, and relevance of the memorized information. The application of these methods in different scenarios requires comprehensive consideration and experimental evaluation in the context of specific problems and data characteristics.

5 Conclusion

The purpose of this paper is to develop a Mach number prediction model suitable for the wind tunnel flow field controller, with a required Mach number prediction RMSE of less than 0.001.

First, the selection of the wind tunnel flow field model structure is analyzed based on the operational characteristics of the wind tunnel system. The CNN-LSTM-Attention model is finally selected as the basic structural framework according to the advantages and disadvantages of each model. The model accomplishes Mach number prediction under specific operating conditions, and is compared to the PLS model and BP neural network in simulation. According to the three evaluation indexes of the root mean square error, Mach number prediction accuracy and maximum

deviation, the prediction results from the proposed CNN-LSTM-Attention Mach number prediction model meet the accuracy index of wind tunnel flow field under three typical operating conditions and have good generalizability. Meanwhile, a cross-sectional comparison is conducted on the CNN-LSTM-Attention models, revealing the impact of SE, ECA, and CBAM attention mechanisms may vary under different working conditions, depending on the feature correlation, noise and redundancy of the data, the number of channels, and other factors. The application of these methods in different scenarios requires comprehensive consideration and experimental evaluation with specific problems and data characteristics.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (No. 61503069) and the Fundamental Research Funds for the Central Universities (N150404020).

References

- [1] Zhao L, Shao Y, Jia W. NARX-Elman based Mach number prediction and model migration of wind tunnel conditions. *Aerospace* 2023, 10, 498. <https://doi.org/10.3390/aerospace10060498>.
- [2] Piovesan T, Magrini A, Benini E. Accurate 2-D modelling of transonic compressor cascade aerodynamics. *Aerospace* 2019, 6, 57. <https://doi.org/10.3390/aerospace6050057>.
- [3] Liu D, Zhou J, Peng Y. Pognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression. *Microelectronics Reliability*, 2013, 53(6):832-839.
- [4] Long D F, Gladen K S. Development of a control system for an injector powered transonic wind tunnel, The 15th Aerodynamic Testing Conference, 1988, San Diego, CA. 436-445.
- [5] David M N. Wind tunnel computer control system and instruction, ISA, 1989:87-101.
- [6] Soeterboek R A M, Pels A F, Verbruggen H B, et al. A predictive controller for the Mach number in a transonic wind tunnel. *IEEE Control Systems Magazine*, 1991, 11(1): 63-72.

- [7] Manitius A. Feedback controllers for a wind tunnel model involving a delay: Analytical design and numerical simulation, *IEEE Transactions on Automatic Control*, 1984, 29(12): 1058-1068.
- [8] Dandois J, Pamart P Y. NARX modeling and extremum-seeking control of a separation, *Journal Aerospace Lab*, 2013, ALO6-06.
- [9] Leontaritis J, Billings S A. Input-output parametric models for nonlinear systems, Part I: deterministic nonlinear systems; Part II: stochastic nonlinear system. *International Journal of Control*, 1985, 41(1): 303-344.
- [10] Höskuldsson A. PLS regression methods. *Journal of chemometrics*, 1988, 2(3): 211-228.
- [11] Yang Y, Xiong Q, Wu C, et al. A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. *Environmental Science and Pollution Research*, 2021, 28(39): 55129-55139.
- [12] Chung W H, Gu Y H, Yoo S J. District heater load forecasting based on machine learning and parallel CNN-LSTM attention. *Energy*, 2022, 246: 123350.
- [13] Eren L, Ince T, Kiranyaz S. A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *Journal of Signal Processing Systems*, 2019, 91: 179-189.
- [14] Hochreiter S, and Schmidhuber J. Long short-term memory *Neural Computation* 9.8(1997):1735-1780.
- [15] Wijesinghe S, *Instrumentation*, 2020, 7(4): 25-39.
- [16] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020: 11534-11542.
- [17] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module//*Proceedings of the European conference on computer vision (ECCV)*. 2018: 3-19.
- [18] Qin Z, Zhang P, Wu F, et al. Fcanet: Frequency channel attention networks//*Proceedings of the IEEE/CVF international conference on computer vision*. 2021: 783-792.

Author Biographies



ZHAO Luping received the B.S. degree in automation from Northeastern University, Shenyang, China, in 2007, the M.S. degree in control theory from Northeastern University, Shenyang, China, in 2009, and the Ph.D. degree in control theory from the Hong Kong University of Science and Technology, Hongkong, China, in 2014. Her research interest includes industrial process control and industrial artificial intelligence, specifically, control theory, process modeling, model migration, process monitoring, variable prediction of complex industrial processes.

E-mail: zhaolp@ise.neu.edu.cn



WU Kunyang is currently enrolled in automation from Northeastern University, Shenyang, China. His main research interests include industrial artificial intelligence and variable prediction of complex industrial processes.

E-mail: 13998211089@163.com



Copyright: © 2023 by the authors. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).