

Nonlinear Dynamic Global-Local Retaining Projection-Based Anomaly Detection of Mach Number in Wind Tunnel Flow Field

Luping Zhao^{1,*} , Shihan Gao¹ 

¹Northeastern University  – College of Information Science and Engineering – Shenyang – China.

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

ABSTRACT

Mach number anomaly detection is crucial in wind tunnel experiments. Moreover, there is a need for an anomaly detection model that can detect new samples based solely on historical normal data. This paper proposes an anomaly detection model for wind tunnel flow field Mach numbers using the nonlinear dynamic global-local retaining projection method (NDGLPP). This paper first analyzes the key process variables that affect the Mach number. Then, an iterative GLPP (IGLPP) combined with implicit polynomial expansion is employed to process the selected process variables, thereby extracting both global and local features along with their nonlinear relationships. Next, the corresponding statistical metrics, squared prediction error (SPE) and T^2 , are calculated, and the control limits are determined through the cumulative distribution function (CDF). Finally, the constructed model is applied for anomaly detection. To validate the effectiveness of the model, a comparative analysis is conducted using principal component analysis (PCA), GLPP, and other methods. The experimental results indicate that the NDGLPP-based anomaly detection model for wind tunnel flow field Mach numbers not only achieves higher accuracy in detecting abnormal values but also effectively balances the capture of both global and local features, further confirming its effectiveness and superiority.

Keywords: Wind tunnel; Mach number; Anomaly detection; Nonlinear dynamics; Global and local features; Feature extraction.

INTRODUCTION

Wind tunnels are experimental devices that simulate airflow environments to study the aerodynamic performance of objects. Since Reynolds first introduced the concepts of laminar flow and turbulence through circular pipe flow experiments in the late 19th century, wind tunnel technology has gradually become a core method in aerodynamic research (Li and Caracoglia 2020; Zhao *et al.* 2021). Today, wind tunnel technology is widely applied in various fields, including aerospace engineering (Marinho *et al.* 2009), automotive engineering, and architectural design, playing a crucial role in aircraft aerodynamic performance tests (Wang *et al.* 2023) and bridge wind resistance assessments (Chen and Thiele 2023). In recent years, significant breakthroughs have been achieved in wind tunnel technology. For instance, in China, the JF22 hypersonic wind tunnel can simulate supersonic airflow up to Mach 30, serving as a key support for hypersonic vehicle development. Furthermore, the introduction of intelligent detection techniques and virtual wind tunnel simulations (Fu *et al.* 2022) has further enhanced experimental efficiency and data accuracy.

Wind tunnel experiments often require the collection of multidimensional parameters such as airflow characteristics, model responses, and environmental conditions, among others, to reflect the characteristics of the test flow field and the corresponding model responses. Among these parameters, the Mach number is a key metric that represents the ratio of airflow velocity to the

Received: Apr. 01, 2025 | **Accepted:** Jul. 09, 2025

Peer Review History: Single Blind Peer Review.

Section editor: Paulo Renato Silva 



speed of sound, and its accuracy directly determines the reliability of critical outcomes such as aerodynamic load predictions and boundary layer transition analysis (Qiu *et al.* 2024). However, in practical wind tunnel experiments, the measured data of the Mach number and other related variables are often affected by various factors, primarily including:

- Equipment errors – Systematic errors introduced by insufficient sensor accuracy or installation deviations (e.g., improper arrangement of seven-hole probes) (Yao *et al.* 2023). Inadequate sensor calibration or drift after prolonged use can also lead to measurement bias.
- Environmental interference – Fluctuations in environmental parameters such as temperature and pressure both inside and outside the wind tunnel directly affect the state of the flow field (Silva *et al.* 2009), thereby interfering with the measurement of most parameters.
- Flow field disturbances – Factors such as the geometry of the wind tunnel nozzle, model installation, wall effects (Gunasekaran *et al.* 2020), and the interaction between trailing shocks and the boundary layer often cause significant local disturbances and unsteadiness in the flow field (De Gregorio 2012; Richter and Rosemann 2013; Sabnis and Babinsky 2019), leading to measurement deviations in the parameters.

Affected by the above factors, measurement deviations in aerodynamic parameters (such as Mach number, total pressure, and static pressure) may lead to local data analysis failures in mild cases or lead to misjudgments about the overall wind tunnel model in severe cases. Therefore, developing efficient Mach number anomaly detection methods is crucial for improving wind tunnel data quality and making timely adjustments.

For the wind tunnel experimental system, there exists a strongly coupled nonlinear dynamic relationship between the Mach number and parameters such as static pressure and total pressure, while also displaying both global and local characteristics. When experimental conditions (e.g., angle of attack, compressor speed) are actively controlled, the Mach number generally follows a deterministic trend along the control direction. At the same time, due to various disturbances, it inevitably exhibits nonstationary random fluctuations at the local level. The coupling effect between this deterministic trend and random disturbances results in the evolution trajectory of the Mach number being both globally predictable and locally characterized by irreversible transient fluctuations.

Due to the high cost of wind tunnel experiments and the scarcity of abnormal data, there is an urgent need to develop a model that relies solely on normal data to effectively determine whether new samples are anomalous. This approach not only reduces experimental costs but also enhances the accuracy and applicability of anomaly detection. In recent years, various methods have been proposed to address nonlinear challenges in Mach number anomaly detection. Wang and Mao (2019), based on a robust regression model, utilized the advantages of Gaussian processes in nonlinear regression to build an anomaly detection framework that integrates Gaussian regression and Gaussian classification, significantly improving the accuracy of anomaly parameter identification. Zhao *et al.* (2020) improved Mach number prediction through regression-based anomaly detection, while also developing a dedicated detector to enhance the identification of Mach number anomaly data. To address the challenges of multimodal wind tunnel systems, Yuan and Zhao (2022) proposed combining the mean partial least squares (PLS) model with kernel PLS, which not only eliminated noise interference but also highlighted the key role of anomaly detection in capturing nonlinear flow field characteristics, while Guo *et al.* (2023) further applied the statistics and control limits on the basis of the prediction model to the monitoring of the wind tunnel process, and achieved a more satisfactory result. Additionally, Zhao *et al.* (2021) proposed an anomaly detection method based on clustering, which generates diverse training subsets through clustering and improves detector performance by using a hybrid criterion to determine the optimal clustering algorithm and the number of clusters.

Although the above methods have made some progress in revealing the nonlinear relationships between the Mach number and other parameters, they still fall short in simultaneously reflecting both the global and local characteristics of the data. In contrast, significant progress has been made in this regard in the field of chemical engineering.

The principal component analysis (PCA), as a classical method in multivariate statistical process monitoring (MSPM), effectively extracts the overall structural features of data (Ayesha *et al.* 2020), yet it often falls short in revealing subtle local correlations among variables. In contrast, locality preserving projection (LPP) focuses on extracting local characteristics (Ayesha *et al.* 2020; Li *et al.* 2019); however, it exhibits limitations in reflecting global information. These shortcomings have prompted

researchers to explore novel approaches that concurrently account for both global and local structures. Tong and Yan (2014) proposed the multiple manifold projection (MMP) method, which characterizes the global and local structures of the data by constructing global and local graphs. Li *et al.* (2024) combined the advantages of PCA and LPP, and proposed a global-local structure analysis (GLSA) fault detection method by constructing a biobjective function. Although these methods can capture the global and local characteristics of process variables, it is difficult to capture the nonlinear relationships between them (Jiang and Yan 2015). Therefore, to simultaneously capture the nonlinear relationships between parameters, Lee *et al.* (2004) extended PCA to the nonlinear domain and proposed Kernel PCA. Zhang *et al.* (2017) used enhanced kernel entropy component analysis for nonlinear monitoring, and Luo *et al.* (2016) designed a kernel global-local structure preservation algorithm as a nonlinear extension. However, these methods primarily rely on kernel mapping to linearize the nonlinear relationships between variables in a high-dimensional space. The choice of kernel function determines the dimensionality of the mapped space, often introducing excessive redundancy, which significantly increases computational complexity during subsequent dimensionality reduction. This can even lead to cost escalation when the sample size is small, making it particularly unfavorable for high-dimensional industrial processes (Zhao *et al.* 2019).

To more accurately reveal the nonlinear relationships between the Mach number and other parameters, as well as the global and local characteristics of the data in wind tunnel experiment process monitoring, and to overcome the issue of increased computational complexity in kernel-based methods when extracting nonlinear features, this paper adopts the nonlinear dynamic global-local retaining projection method (NDGLPP) algorithm for process monitoring of Mach number in wind tunnel flow fields. On one hand, a time window is used to eliminate autocorrelation among process variables, and iterative GLPP (IGLPP) is applied for dimensionality reduction to preserve the global and local features of dynamic data. On the other hand, an implicit polynomial mapping is constructed to handle the nonlinear characteristics of the process variables. This results in the construction of the NDGLPP model and the establishment of corresponding statistical metrics for detection.

Compared with the detection methods based on deep learning, NDGLPP can better preserve the manifold structure of the data, achieve nonlinear separation of abnormal samples, and avoid the dependence of deep learning on a large amount of labeled data. It has more stable generalization performance on small and medium-sized datasets. Although NDGLPP is not as flexible as deep models in end-to-end feature learning, its overall framework is simple, and the parameters are controllable. It has an irreplaceable advantage in some specific practical scenarios.

Furthermore, the advantage of using this method for Mach number process monitoring in wind tunnel experiments lies in:

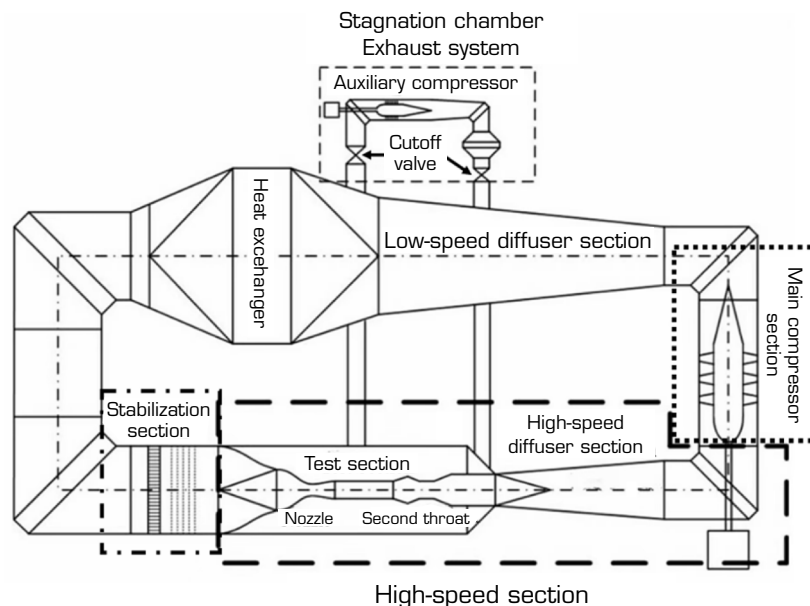
- Simultaneously preserving global and local features – This method not only captures the overall trend of the data during the dimensionality reduction process but also effectively retains the local random fluctuation characteristics. As a result, it constructs a statistical model that is both predictive and capable of capturing irreversible transient fluctuations, making anomaly detection more robust and reliable.
- Strengthened nonlinear mapping capability – By constructing polynomial mappings, high-order nonlinear mappings are realized, more precisely capturing the complex nonlinear coupling relationships between aerodynamic parameters, leading to higher adaptability and accuracy.
- Effectively reduced computational complexity – By constructing a similarity matrix, the method avoids the extensive computation of the full sample kernel matrix. It also employs a multilayer dimensionality reduction strategy to map the data into a lower-dimensional space, reducing the computational burden. At the same time, the inverse projection efficiently extracts nonlinear features in the low-dimensional space, preventing the increase in computational complexity caused by redundant information in high-dimensional mappings.

The rest of this paper is organized as follows: first, the structural composition and basic operation principle of continuous wind tunnels are briefly introduced, and the control mode of Mach number in wind tunnels and its influencing factors are analyzed. Next, the aerodynamic parameters in the wind tunnel are analyzed to determine the process variables related to Mach number, and the NDGLPP model is established. Subsequently, abnormal data is introduced into the model for testing, and the results are analyzed and compared with other models to draw conclusions. Finally, the paper is summarized, and future improvement directions are discussed.



Continuous wind tunnel system

Wind tunnels can be classified into continuous wind tunnels and intermittent wind tunnels based on the method of generating airflow. The layout of a continuous transonic wind tunnel is shown in Fig. 1, which primarily consists of a stabilization section, test section, second throat, reentry section, high-speed diffuser section, compressor, low-speed diffuser section, heat exchanger, and corner sections. During the operation of a continuous wind tunnel, a precise control system is required to dynamically regulate its state to achieve the desired Mach number in the test section. Specifically, the total temperature and total pressure in the stabilization section must remain constant, while adjustments to the main compressor speed, optimization of the second throat regulation, and control of the stagnation chamber exhaust rate are used to fine-tune the airflow parameters, ensuring the test section meets the required flow conditions.



Source: Elaborated by the authors.

Figure 1. Structure diagram of continuous wind tunnel.

Key parameters related to the main compressor include the mass flow rate within the loop and the total pressure loss throughout the system. To maintain the total pressure inside the wind tunnel, pressurization measures are implemented in the upstream section of the stabilization zone, or a vacuum extraction system is used to compensate for potential gas leakage at the compressor and wind tunnel seals. Since leakage is generally minimal, the internal mass flow rate of the wind tunnel can be assumed to remain stable. In supersonic conditions, the velocity at the nozzle throat equals the local speed of sound, meaning the mass flow rate within the wind tunnel is primarily determined by the nozzle geometry and is not affected by model positioning or second throat adjustments. Although the stagnation chamber exhaust system reduces the mass flow in the test section, it does not significantly impact the mass flow rate at the main compressor inlet.

This study focuses on a 0.6 m continuous wind tunnel in Shenyang, analyzing its experimental data and developing an NDGLPP anomaly detection model to identify and detect anomalies in wind tunnel operation.

METHODOLOGY

Analyzing aerodynamic parameters to determine process variables

Based on isentropic flow theory, during an ideal isentropic process, the static pressure, static temperature, and static density of a gas are related to the corresponding total pressure, total temperature, and total density by specific relationships. Continuous wind

tunnels employ precision design and strict control to ensure that the fluid experiences minimal heat exchange and irreversible losses during flow, allowing the process to be approximated as isentropic (Issa and Lakkis 2014). During the experiment, total pressure (P_t) represents the maximum pressure that the gas could reach in the absence of flow losses and is measured and recorded in the stable section of the wind tunnel system, while static pressure (P_s) represents the actual pressure of the fluid in the test section and is measured and recorded there. The ratio of the two not only reflects the energy conversion in the flow field but also serves as an important parameter for determining the Mach number (Ma).

For continuous wind tunnels, the relationship between the Mach number, total pressure, and static pressure can be expressed as:

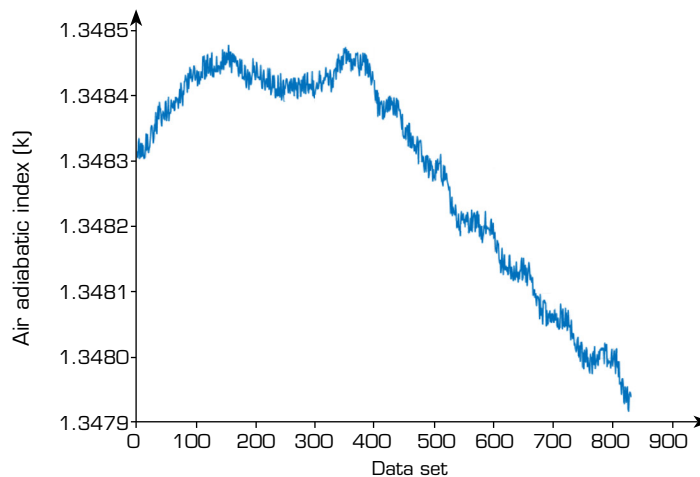
$$Ma = \sqrt{\frac{2}{k-1} \left[\left(\frac{P_t}{P_s} \right)^{\frac{k-1}{k}} - 1 \right]} \quad (1)$$

Where k is the adiabatic index of air. According to thermodynamic theory, its value depends on the molecular structure of the gas and the equation of state. Selecting Condition 4 from Table 1 as the dataset, the distribution of k values in the dataset is obtained using Eq. 1, as shown in Fig. 2. The variation rates of the maximum and minimum values relative to the mean do not exceed 0.03% (maximum variation: 0.014196%, minimum variation: 0.027209%).

Table 1. Details of different operating conditions of the 0.6 m continuous wind tunnel.

Operating conditions	Mach number range	Angle of attack variation (°)	Number of samples
1	0.3	-3 to 4	1,064
2	0.3	-4 to 0	499
3	0.8	-2 to 2	817
4	0.8	-4 to 9	829

Source: Elaborated by the authors.



Source: Elaborated by the authors.

Figure 2. Distribution of k values.

Therefore, although the exact value of k cannot be directly measured in wind tunnel experiments, its variation is minimal, allowing it to be considered approximately constant. Meanwhile, despite the fact that wind tunnel conditions do not fully meet ideal assumptions, the measured Mach numbers still approximately conform to the above Eq. 1. Based on this, this paper divides the experimental data into a training set and a test set and defines the error function as follows:

$$Error = \sum_{i=1}^n \left\{ Ma_i - \sqrt{\frac{2}{\hat{k}-1} \left[\left(P_{t_i} / P_{s_i} \right)^{\frac{\hat{k}-1}{\hat{k}}} - 1 \right]} \right\}^2 \quad (2)$$

Where $\hat{k}$ represents the fitted adiabatic index of air. By minimizing this error function, the value of $\hat{k}$ that best aligns the theoretical predictions with the actual data in the training set can be obtained.

Since the variation of k is small, it can be approximated as constant in the test set with the value $\hat{k}$. Next, the obtained value of $\hat{k}$ is substituted into Eq. 1, and the predicted Mach number $\bar{Ma}$ is calculated using the total pressure and static pressure data from the test set. The difference between the predicted Mach number $\bar{Ma}$ and the actual Mach number Ma is quantified by the mean squared error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (\bar{Ma} - Ma)^2 \quad (3)$$

From Table 2, it can be observed that the calculated MSE is consistently less than 10^{-8} , indicating that the predicted Mach number is very close to the actual Mach number.

Table 2. MSE of predicted and actual Mach numbers under different operating conditions.

Dataset	1	2	3	4
MSE	8.2554×10^{-12}	3.3251×10^{-11}	2.6901×10^{-11}	5.347×10^{-9}

Source: Elaborated by the authors.

In summary, without relying on the air adiabatic index k , the Mach number can be accurately estimated from historical data and the total pressure P_t and static pressure P_s from the test set. This result demonstrates that by analyzing the nonlinear relationship between P_t , P_s , and Ma , and constructing the corresponding data-driven model, it is possible not only to predict the Mach number but also to identify anomalous data, thereby enhancing the stability and reliability of the system.

Establishment of the NDGLPP model

Based on the previous analysis, P_t , P_s , and Ma are selected as input variables. The original input matrix X is constructed as:

$$X = [P_t \quad P_s \quad Ma] \quad (4)$$

Considering that in wind tunnel experiments, when the experimental conditions are actively controlled, P_t , P_s , and Ma exhibit an overall trend and significant temporal autocorrelation, using data from a single moment alone often fails to capture the overall temporal characteristics. The sliding window technique, by continuously sampling local data, captures short-term dynamic variations while revealing the overall trend of the data. Therefore, a sliding time window is used to construct the dynamic extended matrix. The dynamic extended matrix X_d is constructed as follows by concatenating the standardized data from consecutive time steps into a long vector to form a new sample, as shown in Eq. 5. This extended data representation not only contains information from the current time step but also incorporates the dynamic characteristics of the previous $h-1$ time steps.

$$X_d = \begin{bmatrix} X_h^T & X_{h-1}^T & \cdots & X_1^T \\ X_{h+1}^T & X_h^T & \cdots & X_2^T \\ \vdots & \vdots & \ddots & \vdots \\ X_n^T & X_{n-1}^T & \cdots & X_{n-h+1}^T \end{bmatrix}, X_d \in R^{(n-h+1) \times d_0 h} \quad (5)$$

Here, d_0 represents the original feature dimension, $d_0 h$ is the dimension of the dynamic extended matrix, and n is the number of standardized original samples. This approach integrates information from consecutive time steps into a single sample, allowing the subsequent process to effectively capture the trends in temporal dynamics.

After constructing the dynamic extended matrix, the IGLPP is first applied for dimensionality reduction, obtaining representative low-dimensional feature representations and the cumulative projection matrix. Then, implicit polynomial mapping is utilized to perform nonlinear feature expansion on the reduced-dimensional data, enabling the reconstructed data to exhibit multi-scale composite features. Finally, statistical measures are constructed to quantitatively analyze the reconstructed data, achieving multidimensional anomaly perception of the data distribution and laying the foundation for subsequent outlier detection.

Iterative GLPP dimensionality reduction

The key to dimensionality reduction lies in constructing an appropriate projection matrix that transforms high-dimensional data into a lower-dimensional space while preserving essential features. Traditional methods, such as PCA, retain global statistical characteristics but overlook local neighborhood structures. In contrast, LPP emphasizes local structure preservation but may lead to the loss of global distribution information. GLPP integrates the advantages of both within a unified framework, maintaining the orthogonality of PCA while avoiding potential singularity issues in LPP.

The objective functions of LPP and PCA are as follows (Zhang *et al.* 2011; Zhan *et al.* 2017):

$$J(a)_{LPP} = \min_a \sum_{i=1}^n (y_i - y_j)^2 \omega_{ij} = \min_a a^T L' a \quad (6)$$

$$J(a)_{PCA} = \max_a \sum_{i=1}^n (y_i - \bar{y})^2 = \max_a a^T C a$$

Where y_i, y_j are the mapped vectors, $\bar{y} = \frac{1}{n} \sum y_i$, L' is the local Laplacian matrix, C is the covariance matrix of the dynamic input matrix, a is the projection vector, and ω_{ij} is the element in the similarity matrix W :

$$C = \frac{1}{m-1} X_d^T X_d$$

$$L = D - W$$

$$L' = Z L Z^T \quad (7)$$

$$\omega_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2\mu_i}\right), & \text{if } j \in KNN(i) \text{ and } j \neq i \\ 0 & \end{cases} \quad (8)$$

Otherwise, $KNN(i)$ represents the k -nearest neighbors set of sample i , i.e., the average squared distance from i to its k -nearest neighbors. When k is too small, the neighborhood becomes disconnected, leading to fragmented local structures; when k is too large, locality is lost, and the method approximates a global method. D is the $n \times n$ diagonal matrix.

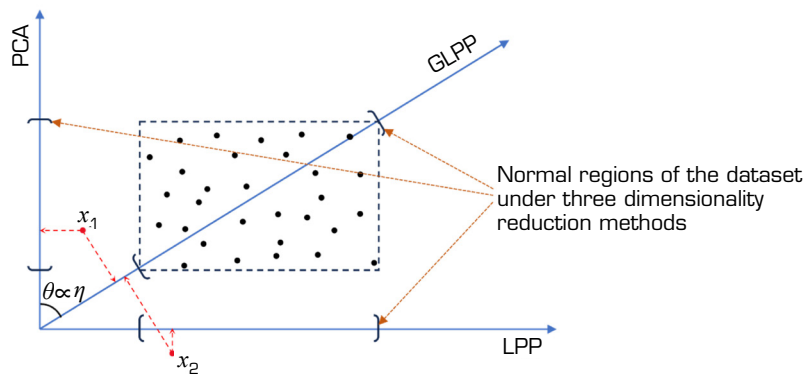
GLPP combines the manifold structure preservation capability of LPP with the global variance maximization property of PCA. The balancing factor η harmonizes the two aspects, achieving a collaborative optimization of both global distribution properties and local geometric constraints. This allows for the preservation of the overall data distribution while effectively capturing the nonlinear structural features in the local neighborhood. The objective function of GLPP is as follows (Zhang *et al.* 2011):

$$\begin{aligned}
 J(a)_{GLPP} &= \eta J(a)_{PCA} - (1 - \eta) J(a)_{LPP} \\
 &= \eta \max_a \sum_{i=1}^n (y_i - \bar{y})^2 - (1 - \eta) \min_a \sum_{i=1}^n (y_i - y_j)^2 \omega_{ij} \\
 &= \eta \max_a a^T C a - (1 - \eta) \min_a a^T L' a \\
 &= \max_a \{ \eta a^T C a - (1 - \eta) a^T L' a \} \\
 &= \max_a a^T [\eta C - (1 - \eta) L'] a \\
 &= \max_a a^T M a
 \end{aligned} \tag{9}$$

To determine the value of the balancing factor η , a method proposed in reference ²² is as follows:

$$\eta = \frac{\rho(L')}{\rho(C) + \rho(L')} \tag{10}$$

Where $\rho(L')$ and $\rho(C)$ represent the spectral radii of the corresponding matrices, which are the absolute values of the maximum eigenvalues of the matrices. The spectral radius focuses more on the differences in data along the principal component direction, maximizing the variance retention in the principal component space. At the same time, the parameter η is an adaptive parameter. As shown in Fig. 3, it can change according to different process data. It allows the feature extraction process to dynamically optimize based on the statistical characteristics of the input data, achieving an intelligent balance of global and local information, and demonstrating significant flexibility in various industrial process monitoring scenarios.



Source: Elaborated by the authors.

Figure 3. Comparison of GLPP, PCA, and LPP in 2D data dimensionality reduction.

For $M = \eta C - (1 - \eta) L'$, the matrix is positive definite, and a is the unit projection vector. Introducing a constant λ , the Lagrangian function is constructed as follows:

$$L(a, \lambda) = a^T M a - \lambda (a^T a - 1) \quad (11)$$

The function is differentiated with respect to a , and setting the derivative equal to zero, the critical conditions are obtained as follows:

$$\frac{\partial L}{\partial a} = 2Ma - 2\lambda a = 0 \Rightarrow Ma = \lambda a \quad (12)$$

Equation 12 indicates that λ and a are the eigenvalue and eigenvector of M , respectively.

In the characteristic equation $Ma = \lambda a$, the value of the objective function corresponding to the eigenvector a is $a^T M a = a^T (\lambda a) = \lambda$. Therefore, solving for the maximum value of the objective function in Eq. 9 is equivalent to finding the maximum eigenvalue of the matrix M , and the corresponding eigenvector a is the solution to the objective function.

By solving for the first d largest eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_d$, and their corresponding eigenvectors $a_1, a_2, \dots, a_d$, the projection matrix $A = [a_1 \ a_2 \ \dots \ a_d]$ is constructed. After several iterations, multiple layers of projection matrices A_i ($i = 1, 2, \dots, k$) are generated, where the dimensionality reduction space after the k -th iteration is:

$$Z^{(k)} = Z^{(k-1)} A_k \quad (13)$$

Where $Z^{(0)} = X_d$.

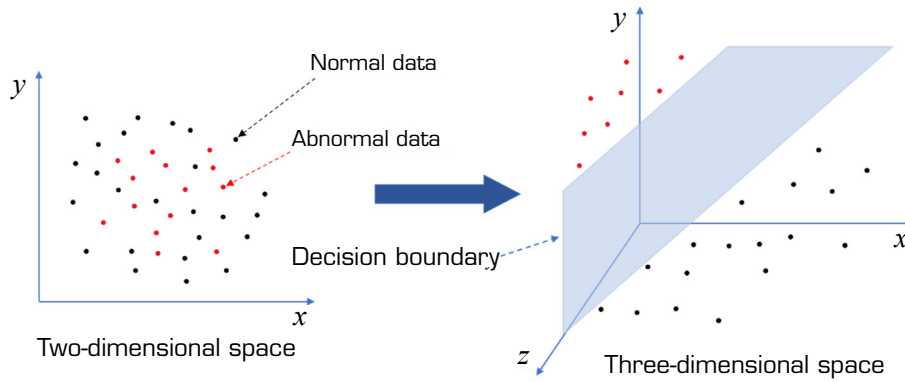
Compared to a single iteration of GLPP, which tends to lose high-order nonlinear features, multiple iterations of GLPP reconstruct the neighborhood graph at each iteration based on the current low-dimensional space $Z^{(k)}$, gradually optimizing the local structure representation. For example, the first iteration captures the linear coupling between pressure and Mach number, while subsequent iterations identify nonlinear relationships such as $(\Delta M \cdot \nabla P)$. Meanwhile, during the iteration process, the value of η is dynamically adjusted based on the number of iterations, which further helps balance the retention of both global and local features.

The final reduced-dimensional mapping space is obtained as follows:

$$Z = X_d A_{total}, \quad A_{total} = \prod_{i=1}^K A_i \quad (14)$$

Implicit polynomial mapping

While this method provides interpretable nonlinear terms, as shown in Fig. 4, it suffers from the dimensionality explosion problem when applied to high-dimensional inputs. Additionally, the fixed polynomial terms are unable to capture the time-varying coupling characteristics between Mach number, static pressure, and total pressure in wind tunnel experiments (e.g. $Ma^2 \propto (P_t / P_s)^{\frac{k-1}{k}}$). In contrast, implicit polynomial mapping does not require the construction of all second-order or higher-order cross terms. It not only effectively avoids dimensionality explosion and reduces computational complexity (for quadratic polynomial expansion, the computational complexity is reduced from the explicit $O(d^2)$ to $O(dm)$, where m is the dimension after dimensionality



Source: Elaborated by the authors.

Figure 4. Polynomial mapping diagram.

reduction, and typically $m \ll d$), but also automatically captures the inherent complex nonlinear relationships within the data, better adapting to the time-varying coupling characteristics between Mach number, static pressure, and total pressure. To this end, this paper adopts implicit mapping based on cascaded graph projection. For each projection matrix A_i generated after every iteration of GLPP, although it is essentially a linear transformation, the cumulative projection matrix A_{total} obtained by multiplying the projection matrices in sequence retains local structural information in each dimensionality reduction. Moreover, under certain conditions, the multi-layer linear combination can be approximated as a nonlinear mapping of the original data.

In describing nonlinear features, for quadratic polynomial mapping, the second-order cross terms are explicitly constructed to represent the nonlinear relationships. For a given input $X \in R^d$, its quadratic polynomial mapping can be expressed as:

$$\phi(x) = [x_1, x_2, \dots, x_d, x_1^2, x_2^2, \dots, x_d^2, x_1x_2, x_1x_3, \dots, x_{d-1}x_d] \in R^{\frac{d(d+3)}{2}} \quad (15)$$

For the cumulative projection matrix $A_{total} \in R^{d \times m}$, the conventional inverse matrix does not exist. Therefore, the pseudo-inverse operation is used to establish the inverse mapping relationship, and the pseudo-inverse matrix is given as follows:

$$A_{pinv} = (A_{total})^\dagger \quad (16)$$

The pseudo-inverse operation combines the Moore-Penrose generalized inverse with manifold learning theory. It ensures the reversibility of the dimensionality reduction process, effectively solving the irreversibility problem of information loss during dimensionality reduction, while enhancing the model's expressive power through implicit feature space transformations.

Using the pseudo-inverse matrix, the dimensionality-reduced data is mapped back to the original space, and the reconstructed data is as follows:

$$Z_{re} = ZA_{pinv} = XA_{total}A_{pinv} \quad (17)$$

The IGLPP combined with implicit polynomial mapping achieves the extraction of both local and global features through dual projections. The cumulative projection matrix A_{total} accurately captures the local manifold structure of the data, while the pseudo-inverse matrix A_{pinv} reconstructs the global topological structure of the data, ensuring the integrity of the overall data

structure. The reconstructed data fully incorporates both global and local features of the original data, forming a multi-scale composite feature representation.

Construction of the statistical metrics prediction error (SPE) and T^2

The SPE statistic reflects the magnitude of the error outside the low-dimensional subspace, i.e., the difference between the original data and its low-dimensional representation. It indicates the details and noise in the data that are not captured. When the SPE value is large, it suggests that there is a significant difference between the sample and the model subspace, and the sample may be an outlier or anomaly. The calculation formula is as follows:

$$SPE = \sum_{i=1}^d (x_i - \hat{x}_i)^2 = \|X_d - Z_{re}\|^2 \quad (18)$$

The T^2 statistic is used to measure the distribution of the sample within the subspace of the model and to determine whether the sample is abnormal in the model. If the T^2 value of a certain sample is too high, it indicates that the sample deviates from the overall distribution in terms of the main features, and there may be an anomaly.

The calculation formula is as follows:

$$T^2 = z_i \Sigma_z^{-1} z_i^T = \text{diag}(Z_{re} \Sigma_{Z_{re}}^{-1} Z_{re}^T) \quad (19)$$

Where $\Sigma_{Z_{re}}$ is the covariance matrix of Z_{re} .

Based on the obtained SPE and T^2 statistics, the cumulative distribution function (CDF) is first constructed using kernel density estimation (KDE). This calculates the probability that the sample data $\{x_i\}^n$ (where x_i represents the SPE and T^2 values of each sample) is less than or equal to a certain value. The CDF constructed using KDE is as follows:

$$F(x) = \frac{1}{n} \sum_{i=1}^n \Phi\left(\frac{x - x_i}{h}\right) \quad (20)$$

Where h is the bandwidth, which controls the smoothness of the density estimation, and $\Phi(\cdot)$ represents the CDF of the standard normal distribution:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (21)$$

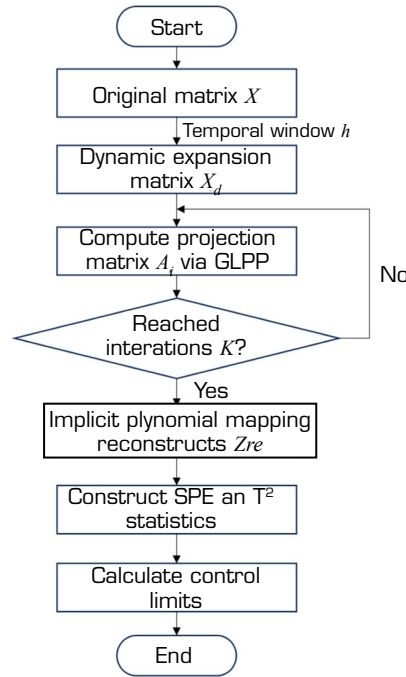
Next, interpolate the CDF and compute its inverse to determine the control limit x_{limit} corresponding to the $1 - \alpha$ confidence level:

$$\begin{aligned} F(x_{limit}) &= P(X < x_{limit}) = 1 - \alpha \\ \Rightarrow x_{limit} &= F^{-1}(1 - \alpha) \end{aligned} \quad (22)$$

If the SPE or T^2 statistic of a new sample exceeds the corresponding control limit, the sample is identified as an outlier.

SPE captures the error information that the model fails to explain, while T^2 reflects anomalies within the model's explained subspace. Together, they provide a more comprehensive way to detect anomalies. For a new sample, if either the SPE or T^2 value exceeds the respective control limit, the sample is considered anomalous.

The overall algorithm flow of NDGLPP is shown in Fig. 5, and the corresponding pseudo-code implementation is presented in Fig. 6.



Source: Elaborated by the authors.

Figure 5. Flowchart of model establishment for NDGLPP.

NDGLPP constructs a dynamic expansion matrix, a similarity matrix, and a local Laplacian matrix, and combines the covariance matrix to calculate parameter η and builds a fusion matrix M . Through eigenvalue decomposition, a projection matrix is obtained to achieve dimensionality reduction. Then, the cumulative projection matrix is used for pseudo-inverse operation to reconstruct the data, calculate the corresponding SPE and T^2 statistics, and obtain the 99% control limit using KDE and the interpolation method for anomaly detection. This method possesses the capabilities of dynamic characteristic capture, manifold structure preservation, and nonlinear processing. It also enhances the reliability of anomaly detection through the dual control limits of SPE and T^2 and applies to scenarios such as industrial process monitoring.

Outlier detection for new samples

Due to the scarcity of historical anomaly data from wind tunnels, to simulate the anomalies that may occur when the measurements of static pressure, total pressure, and Mach number are subjected to multiple disturbances in wind tunnel experiments, this paper constructs representative anomalous data by randomly selecting one or two of the three features on top of the normal data and applying random offsets of plus or minus three times the standard deviation to them, which are treated as if they have measurement errors:

$$x_{new}^i = x^i \pm 3\sigma \quad (23)$$

$$X_{new} = \begin{bmatrix} P_{t_{new}} & P_{s_{new}} & Ma_{new} \end{bmatrix} \quad (24)$$

Algorithm 1: Pseudocode of the NDGLPP algorithm

Input: Original matrix $X \in R^{n \times D}$
Output: Projection matrix A_i , Reduced-dimensional matrix Z ,
Reconstructed matrix Z_{re} , Control limit SPE_{limit}, T^2_{limit}

```

1 begin
2   Step 1: Build Dynamic Expansion Matrix
3   //h: the window size
4    $X \leftarrow \text{Standardize } X$  // z-score for standardization
5    $X_d \in R^{(n-h+1) \times (D-h)} \leftarrow \text{Build dynamic expansion matrix } (X, h)$ 
6   Step 2: IGLPP Dimensionality Reduction
7   //d: the target dimension
8    $Z \leftarrow X_d$ 
9   for  $i = 1$  to  $K$  do
10     $W \leftarrow \text{Build similarity matrix } (Z, k)$  // k-nearest neighbor
11     $D \leftarrow \text{diag}(\text{sum}(W, 2))$ 
12     $L \leftarrow D - W$  // the laplacian matrix
13     $C \leftarrow \text{cov}(Z) + \epsilon \cdot I$  //  $\epsilon$ : the regularization term
14     $L' \leftarrow Z^T L Z$ 
15     $\rho_C \leftarrow \max(|\text{eig}(C)|)$  // the spectral radii
16     $\rho_{L'} \leftarrow \max(|\text{eig}(L')|)$ 
17     $\eta \leftarrow \rho_{L'} / (\rho_C + \rho_{L'})$  // If  $\rho_C + \rho_{L'} = 0$ , then  $\eta = 0.5$ 
18     $M \leftarrow \eta \cdot C - (1 - \eta) \cdot L'$  // the integration matrix
19     $[A, \sim] \leftarrow \text{eigs}(M)$  // obtain eigenvalues and eigenvectors
20     $A_i \leftarrow A(d)$  // the projection matrix
21     $Z \leftarrow Z \cdot A_i$  // reduced-dimensional projection
22     $\text{pro\_mats}\{i\} \leftarrow A_i$  // record the projection matrix
23  end for
24  Step 3: Implicit Polynomial Mapping
25   $A_{total} \leftarrow \text{pro\_mats}\{1\}$  // the cumulative projection matrix
26  for  $j = 2$  to  $K$  do
27     $A_{total} \leftarrow A_{total} \cdot \text{pro\_mats}\{j\}$ 
28  end for
29   $A_{pinv} \leftarrow \text{pinv}(A_{total})$  // the pseudo-inverse matrix
30   $Z_{re} \leftarrow Z \cdot A_{pinv}$  // the reconstructed matrix
31  Step 4: Calculation of SPE and  $T^2$  control limits
32   $SPE \leftarrow \text{sum}(X_d - Z_{re})^2$  // squared prediction error
33   $T^2 \leftarrow \text{diag}(Z_{re} \cdot \text{inv}(\text{cov}(Z_{re})) \cdot Z_{re}^T)$  //  $T^2$  statistical metrics
34   $[f_{SPE}, x_{i_{SPE}}] \leftarrow \text{ksdensity}(SPE, \text{cdf})$  // KDE CDF
35   $[f_{T^2}, x_{i_{T^2}}] \leftarrow \text{ksdensity}(T^2, \text{cdf})$ 
36   $SPE_{limit} \leftarrow \text{interp1}(f_{SPE}, x_{i_{SPE}}, 0.99)$  // 99% control limits
37   $T^2_{limit} \leftarrow \text{interp1}(f_{T^2}, x_{i_{T^2}}, 0.99)$ 
38 end

```

Source: Elaborated by the authors.

Figure 6. Pseudocode of the NDGLPP algorithm.

This method not only effectively simulates data anomalies caused by factors such as environmental, equipment, or operational errors during wind tunnel experiments, but also provides a rich set of test samples for the outlier detection model. This, in turn, helps improve the model's accuracy and robustness in real-world applications, further ensuring the stability and reliability of the entire system.

For a new sample X_{new} , the first step is to transform it into a dynamically expanded matrix $X_{d_{new}}$ to ensure consistency in the data structure.

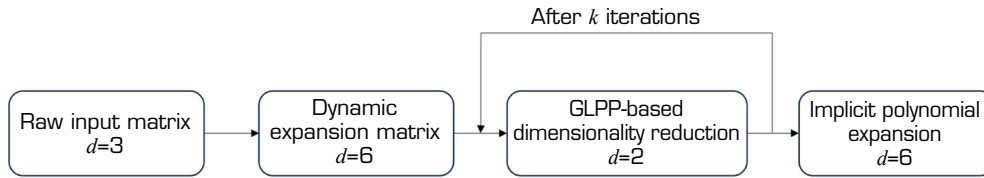
Next, using the cumulative projection matrix A_{total} learned from the training data during the IGLPP process, dimensionality reduction is applied to the new sample, preserving the local manifold structure of the data while removing redundant information. This allows the new sample to be compactly and effectively represented in the low-dimensional space:

$$Z_{new} = X_{d_{new}} A_{total} \quad (25)$$

Subsequently, implicit polynomial mapping is applied to the reduced data to generate reconstructed data $Z_{re_{new}}$, recovering its latent features in the high-dimensional space:

$$Z_{re_new} = Z_{new} A_{pinv} = Z_{new} (A_{total})^{\dagger} \quad (26)$$

In this process, the data undergoes a transformation from high-dimensional to low-dimensional and back to high-dimensional, as shown in Fig. 7. This nonlinear mapping helps capture complex nonlinear relationships, providing more accurate representations for subsequent statistical calculations and outlier detection.



Source: Elaborated by the authors.

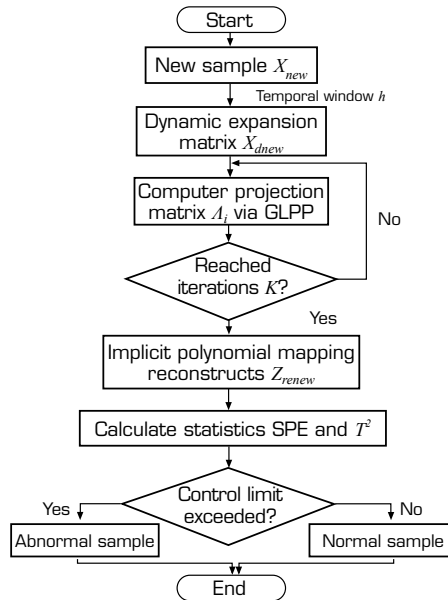
Figure 7. Diagram of data dimensionality change.

After reconstructing the new sample, the SPE and T^2 statistics for Z_{re_new} are calculated. These calculated SPE and T^2 values are then compared with the control limits determined by the training data, effectively distinguishing normal samples from anomalous ones:

$$\left. \begin{array}{l} SPE_i > SPE_{limit} \\ \text{or} \\ T^2_i > T^2_{limit} \end{array} \right\} \Rightarrow x_{i_new} \text{ is an outlier} \quad (27)$$

If the statistics for the new sample exceed the respective control limits, it suggests the presence of a potential anomaly, requiring further analysis or appropriate actions.

Figure 8 shows the flowchart of the new sample detection process, to intuitively understand the execution order and logical relationship of each step, so as to provide strong support for the subsequent process optimization and troubleshooting.



Source: Elaborated by the authors.

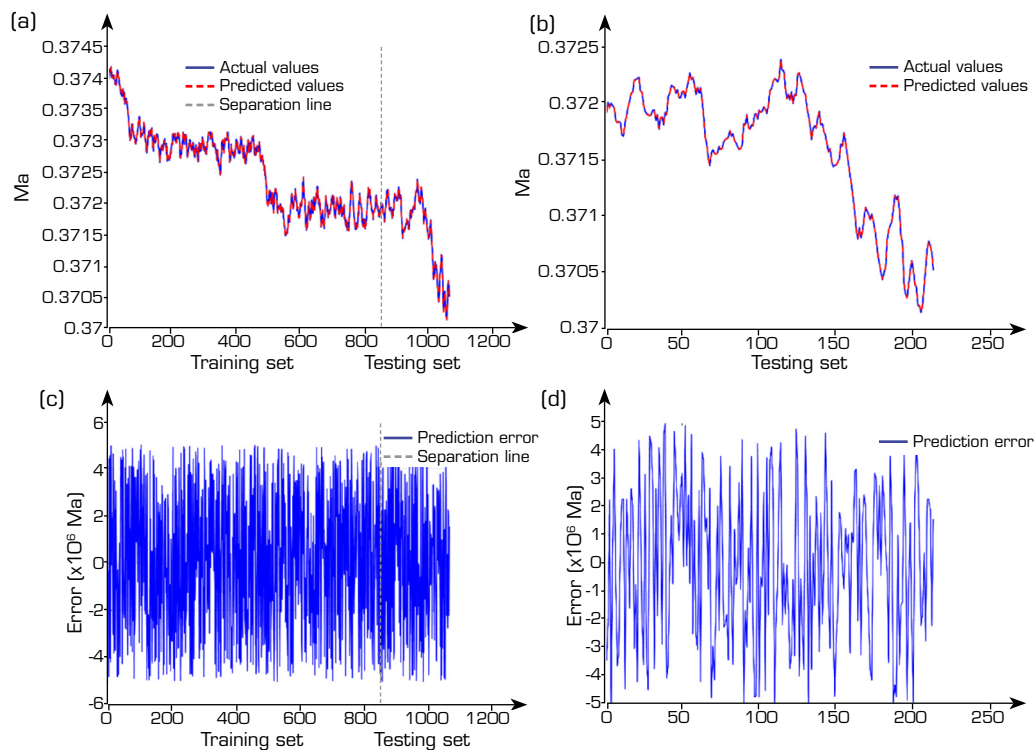
Figure 8. New sample detection flowchart.

Illustration

Comparison between predicted and actual Mach numbers

Using a 0.6 m continuous wind tunnel in Shenyang as an example, data from four different operating conditions are selected to compare and analyze the predicted Mach number with the actual Mach number. Due to confidentiality requirements, it is not appropriate to present this data openly in the text. This clarifies the situation. The characteristics of the four conditions are shown in Table 1.

For each dataset under different operating conditions, 80% of the data is selected as the training set, and the remaining data is used as the test set. First, the optimal air adiabatic index $\hat{k}$ is solved by minimizing Eq. 2. Then, the value of $\hat{k}$ is substituted into Eq. 1, along with the P_t and P_s from the test set to calculate the predicted Mach number $\hat{Ma}$ for the test set. Comparisons between predicted and measured Mach numbers for each operating condition are shown in Figs. 9–12.

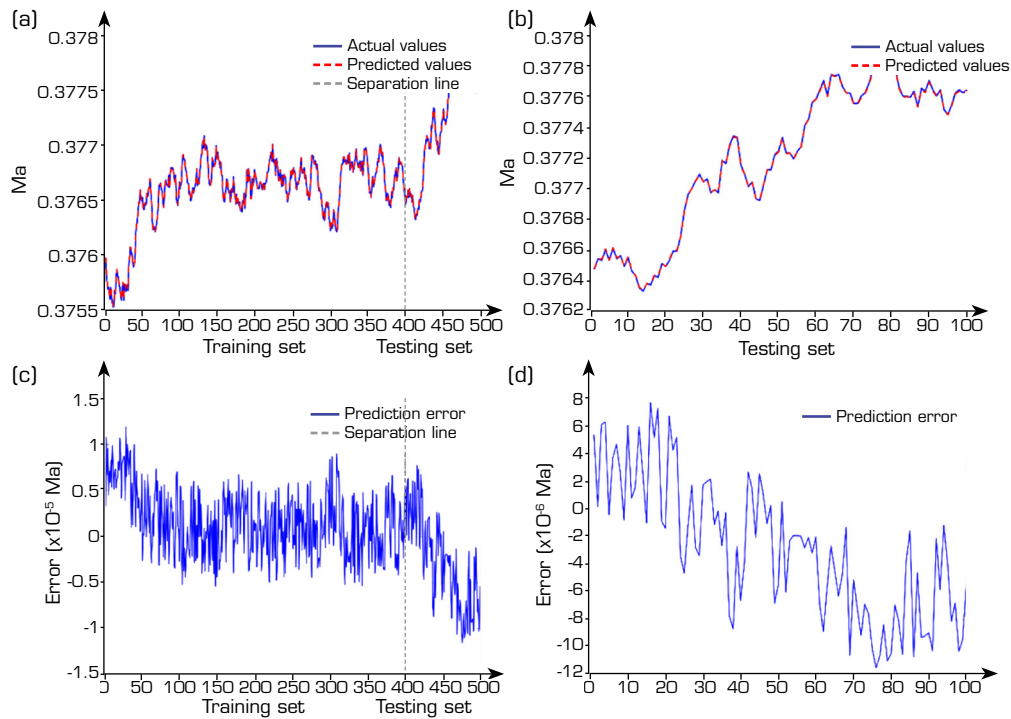


Source: Elaborated by the authors.

Figure 9. Comparison of predicted and actual Mach numbers under operating condition 1. (a) Comparison of predicted Mach number and measured Mach number under the full data set; (b) Comparison between predicted Mach number and measured Mach number in the test set; (c) Error distribution under the full dataset; (d) Error distribution under the test set.

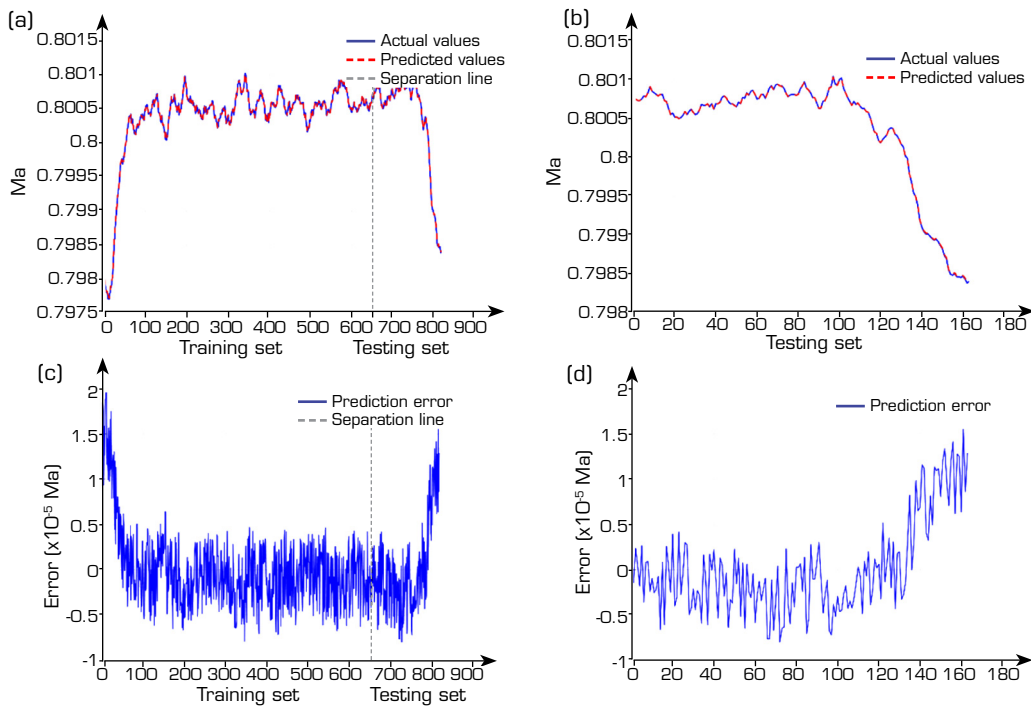
For different operating conditions, the predicted Mach numbers and actual measured values from the test set are substituted into Eq. 3 to calculate the MSE. The MSE values for each operating condition are then summarized in Table 2.

From Table 2, it can be observed that the MSE between the predicted Mach number and the actual Mach number across all operating conditions is extremely low, with the values ranging approximately from 10^{-12} to 10^{-9} . Among these, the MSE for operating condition 4 is significantly higher than the others, which may indicate the presence of higher data noise or more complex operating conditions in this scenario. Therefore, compared to the other conditions, operating condition 4 may require more attention to outlier detection. As a result, in the abnormal monitoring section that follows, operating condition 4 is selected as the dataset for outlier detection, and new samples in this condition will undergo outlier detection.



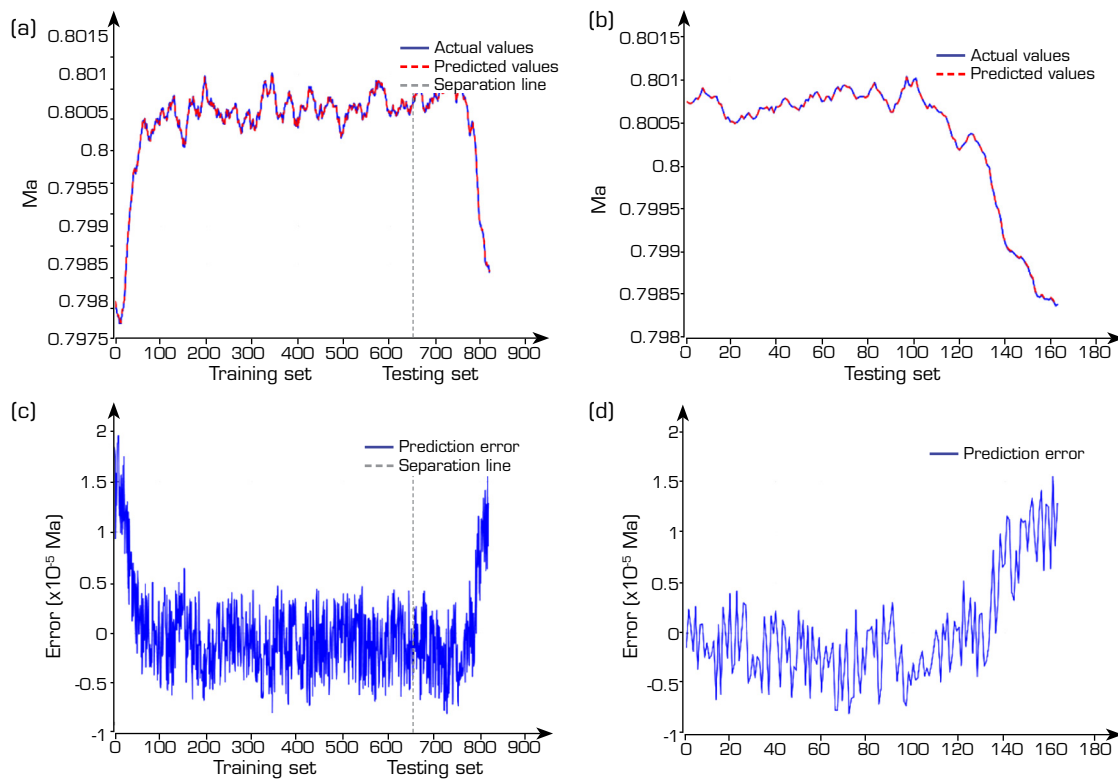
Source: Elaborated by the authors.

Figure 10. Comparison of predicted and actual Mach numbers under operating condition 2. (a) Comparison of predicted Mach number and measured Mach number under the full data set; (b) Comparison between predicted Mach number and measured Mach number in the test set; (c) Error distribution under the full dataset; (d) Error distribution under the test set.



Source: Elaborated by the authors.

Figure 11. Comparison of predicted and actual Mach numbers under operating condition 3. (a) Comparison of predicted Mach number and measured Mach number under the full data set; (b) Comparison between predicted Mach number and measured Mach number in the test set; (c) Error distribution under the full dataset; (d) Error distribution under the test set.

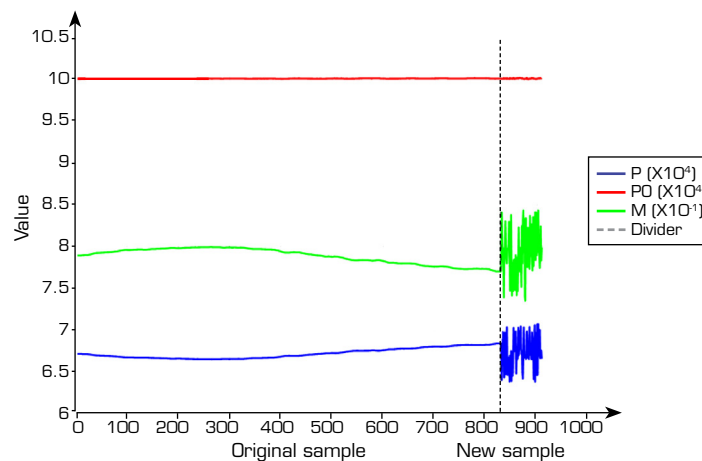


Source: Elaborated by the authors.

Figure 12. Comparison of predicted and actual Mach numbers under operating condition 4. (a) Comparison of predicted Mach number and measured Mach number under the full data set; (b) Comparison between predicted Mach number and measured Mach number in the test set; (c) Error distribution under the full dataset; (d) Error distribution under the test set.

Abnormal detection results and comparative analysis

In operating condition 4, 80 data samples were randomly selected as the basis. For each sample, one to two features were randomly selected, and measurement errors were modeled using Eq. 23 to generate anomalous data samples. The generated data is shown in Fig. 13.



Source: Elaborated by the authors.

Figure 13. Distribution of original and outlier data.

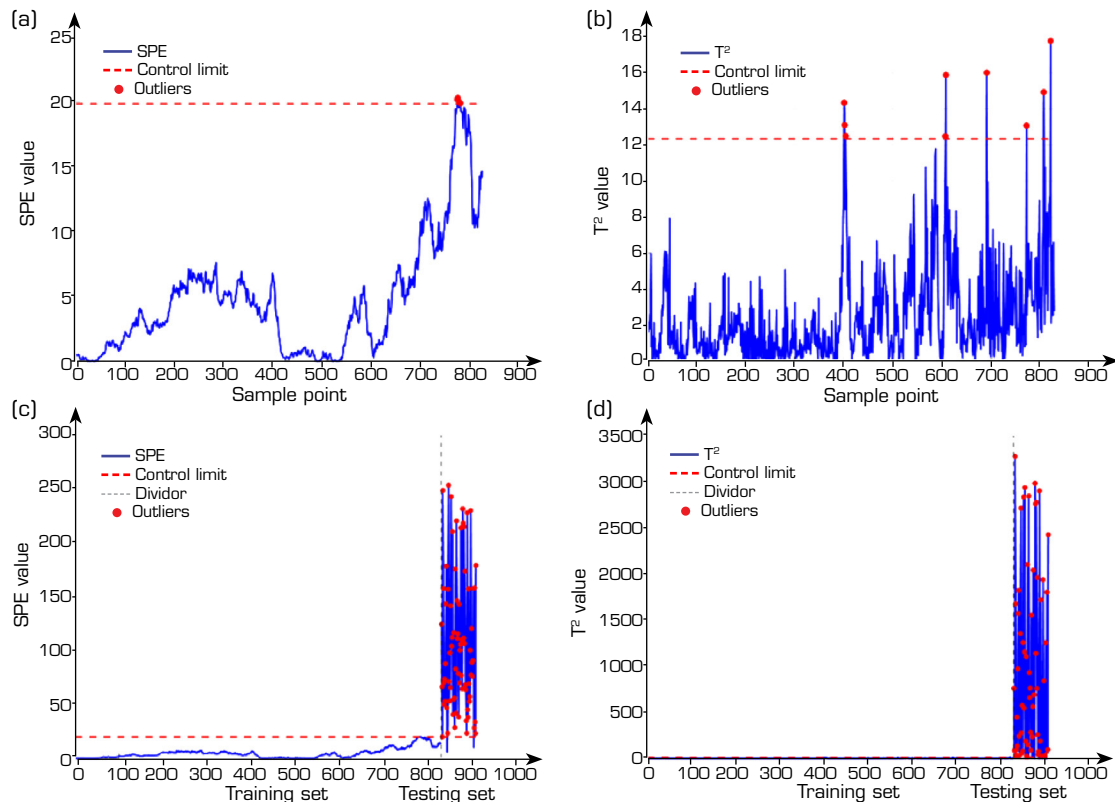
Subsequently, based on the steps in Fig. 8, the SPE and T^2 statistics are computed for the generated abnormal data and compared with the corresponding control limits to detect anomalies.

To provide a more intuitive presentation of the model parameters and analysis results, the following summary Table 3 includes the relevant model parameters and detection outcomes. In addition, Fig. 14 presents a set of four plots showing the distributions of SPE and T^2 statistics for both the training data and 80 outlier samples using the NDGLPP anomaly detection model.

Table 3. NDGLPP model parameters and anomaly detection results.

Parameter	Value
Temporal window h	2
Neighboring number K	5
Iterations number k	5
Dimension after IGLPP dimensionality reduction	2
Dimension after implicit polynomial mapping	6
Count of outlier samples N_{new}	80 (79 after dynamic expansion)
SPE control limit	19.8220
T^2 control limit	12.3398
SPE accuracy	0.9747
T^2 accuracy	0.7848

Source: Elaborated by the authors.



Source: Elaborated by the authors.

Figure 14. NDGLPP model establishment and outlier detection. (a) The changes in the SPE statistics on the training set; (b) The change of the T^2 statistic on the training set; (c) The changes in the SPE statistics on the complete dataset; (d) The changes in the T^2 statistics on the complete dataset.

To evaluate the performance of the NDGLPP anomaly detection model, it is compared with the GLPP, PCA, nonlinear dynamic PCA (NDPCA), and dynamic LPP (DLPP) models. Figures 15–18 show the distribution of SPE and T^2 statistics for

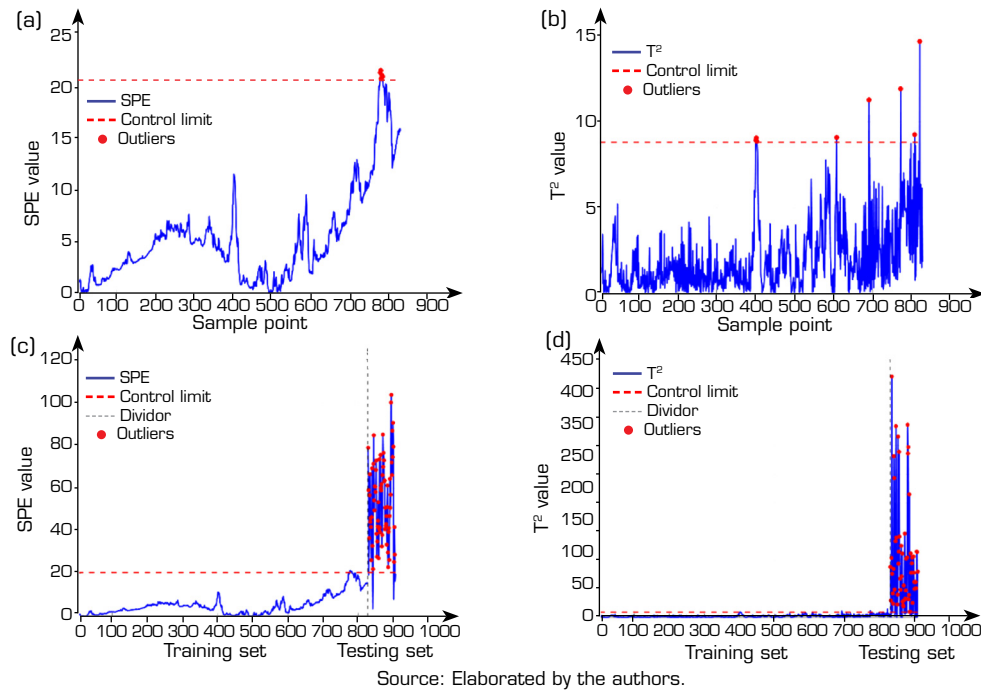


Figure 15. GLPP model establishment and outlier detection. (a) The changes in the SPE statistics on the training set; (b) The change of the T^2 statistic on the training set; (c) The changes in the SPE statistics on the complete dataset; (d) The changes in the T^2 statistics on the complete dataset.

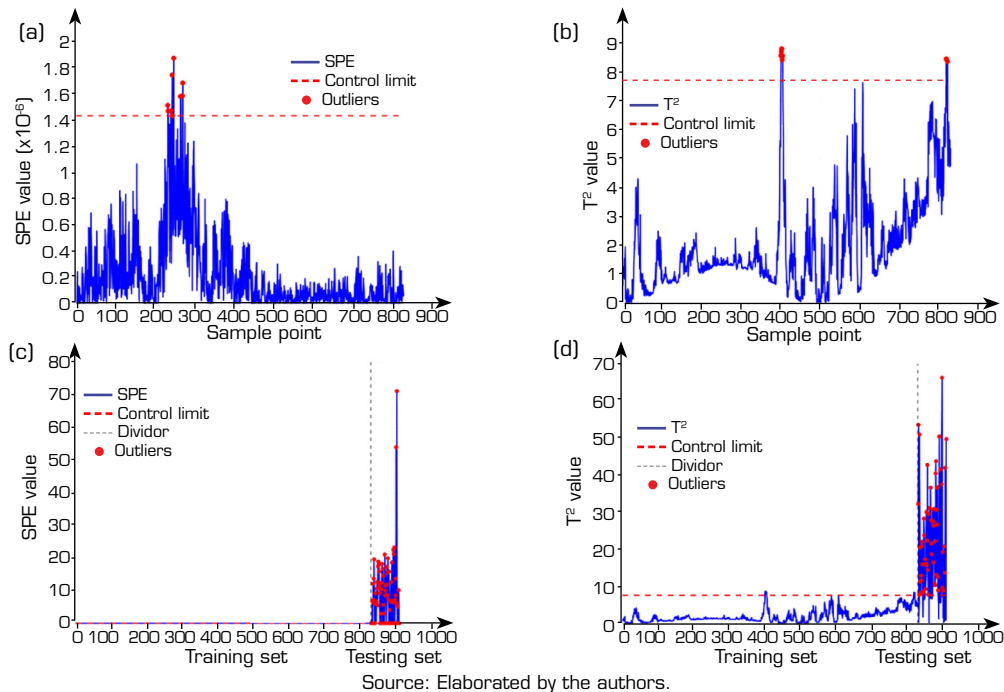
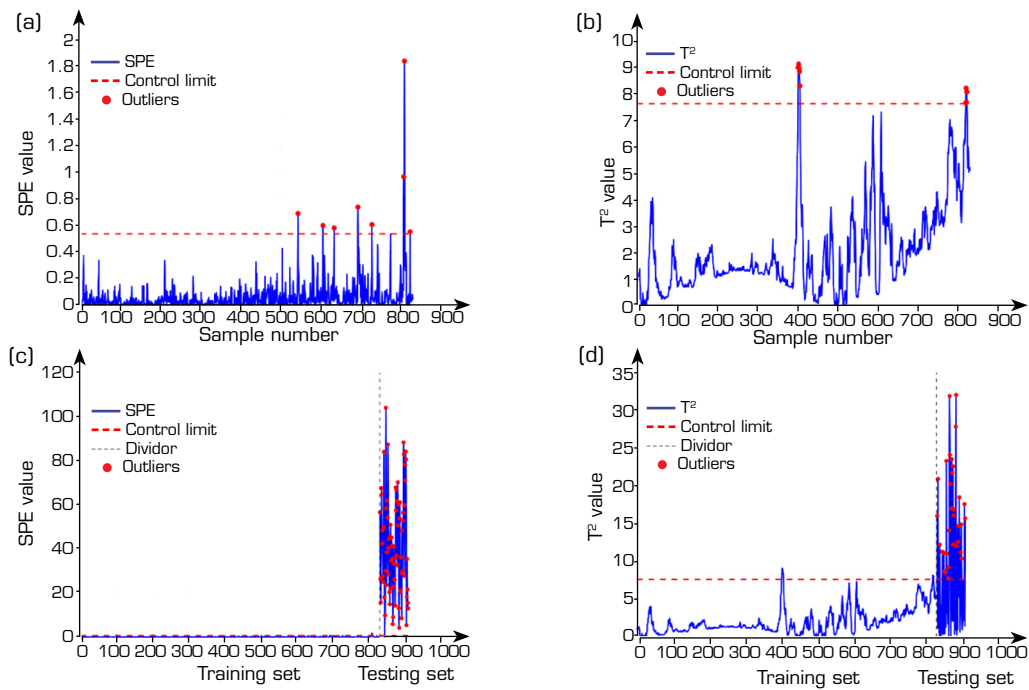
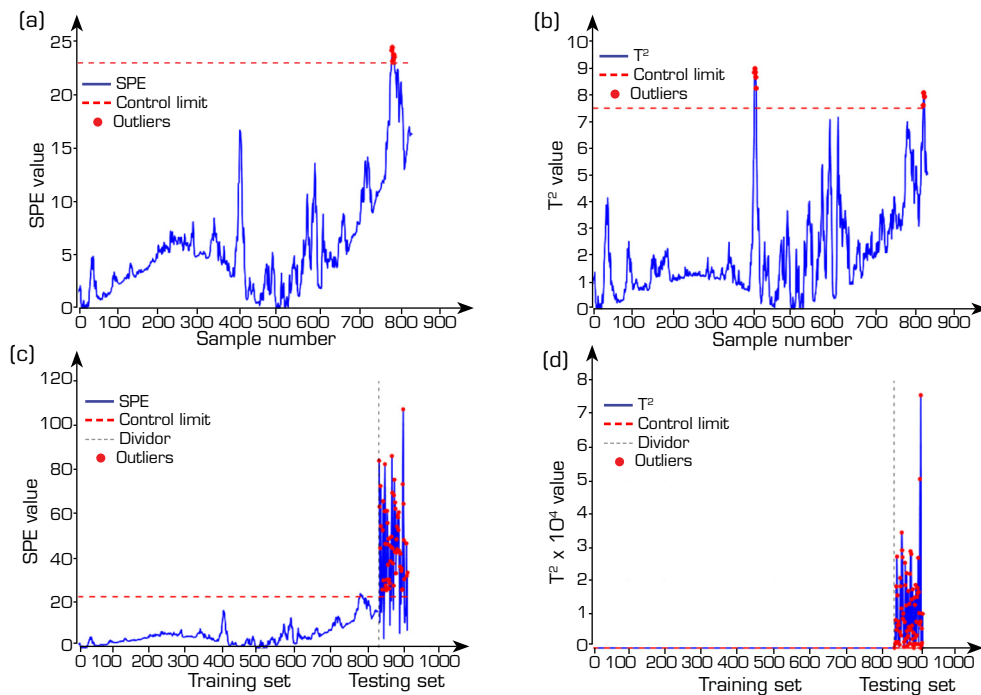


Figure 16. PCA model establishment and outlier detection. (a) The changes in the SPE statistics on the training set; (b) The change of the T^2 statistic on the training set; (c) The changes in the SPE statistics on the complete dataset; (d) The changes in the T^2 statistics on the complete dataset.



Source: Elaborated by the authors.

Figure 17. NDPCA model establishment and outlier detection. (a) The changes in the SPE statistics on the training set; (b) The change of the T^2 statistic on the training set; (c) The changes in the SPE statistics on the complete dataset; (d) The changes in the T^2 statistics on the complete dataset.



Source: Elaborated by the authors.

Figure 18. DLPP model establishment and outlier detection. (a) The changes in the SPE statistics on the training set; (b) The change of the T^2 statistic on the training set; (c) The changes in the SPE statistics on the complete dataset; (d) The changes in the T^2 statistics on the complete dataset.

different models on training data and 80 outlier samples, allowing for a visual comparison of the effectiveness of different methods in outlier detection. This comparison provides an in-depth analysis of the detection capabilities and applicability of each model.

Based on the analysis of the statistical distributions from the NDGLPP, GLPP, PCA, NDPCA, and DLPP models, the outliers detected by each model were identified. To more accurately assess the performance of different models in the outlier detection task, this study compares the accuracy of each model in identifying outlier data. The results are summarized in Table 4.

Table 4. Comparison of different models on training and test data.

Model	Statistic	Training set accuracy	Test set accuracy
NDGLPP	SPE	0.9952	0.9747
	T^2	0.9891	0.7848
GLPP	SPE	0.9928	0.9367
	T^2	0.9903	0.7342
PCA	SPE	0.9891	0.900
	T^2	0.9916	0.7125
NDPCA	SPE	0.9900	0.9870
	T^2	0.989	0.4180
DLPP	SPE	0.9915	0.7722
	T^2	0.9903	0.9494

Source: Elaborated by the authors.

The results indicate significant differences in anomaly detection performance among different models on the test set. NDGLPP achieves notably higher accuracy in both SPE and T^2 statistics compared to GLPP and PCA. Although it shows slight shortcomings in one statistic compared to DLPP and NDPCA, it exhibits a substantial advantage in the other, demonstrating a well-balanced performance across both metrics. This highlights NDGLPP's ability to effectively capture both global and local features. Consequently, NDGLPP not only fits the training data well but also maintains excellent generalization on the test set, proving its superiority and practicality in real-world anomaly detection applications.

Additionally, all models achieve nearly perfect accuracy on the training set, indicating their strong ability to fit the training data. However, accuracy drops across all models on the test set, with NDPCA showing the most significant decline in T^2 statistic accuracy, suggesting weaker generalization capability. In contrast, NDGLPP demonstrates more stable overall performance on the test set, particularly excelling in SPE statistics. This indicates that NDGLPP is more effective in identifying anomalies in real-world applications, offering greater robustness and adaptability.

Overall, NDGLPP demonstrates strong anomaly detection performance in both SPE and T^2 statistics, accurately distinguishing between normal and anomalous data. This validates its advantages in anomaly detection tasks.

CONCLUSION

To address the challenge of anomaly detection in a 0.6 m continuous transonic wind tunnel experiment, where historical anomaly data is scarce and normal operating samples dominate, this paper proposes an anomaly detection modeling strategy based on dynamic nonlinear feature extraction. An NDGLPP-based anomaly detection model for the wind tunnel flow field Mach number was established, and its performance was compared with that of PCA, GLPP, NDPCA, and DLPP. The results indicate that this method not only achieves high detection accuracy but also effectively addresses the challenge of limited anomaly data in engineering applications. Once the model is established, it can be applied for real-time monitoring and inspection of wind tunnel equipment, enabling early warnings and timely maintenance, thereby providing a reliable technical safeguard for the safe and stable operation of wind tunnel experiments.



Future work will further validate the model's real-time response performance and stability in practical applications. Additionally, efforts will be made to explore its efficient integration with existing equipment management systems, providing a solid data foundation and theoretical support for intelligent and preventive maintenance.

CONFLICT OF INTEREST

Nothing to declare.

AUTHORS' CONTRIBUTION

Conceptualization: Zhao L and Gao S; **Methodology:** Zhao L and Gao S; **Software:** Gao S; **Validation:** Gao S; **Formal analysis:** Zhao L and Gao S; **Investigation:** Gao S; **Resources:** Zhao L; **Data Curation:** Gao S; **Writing - Original Draft:** Gao S; **Writing - Review & Editing:** Zhao L; **Visualization:** Gao S; **Supervision:** Zhao L; **Project administration:** Zhao L; **Funding acquisition:** Zhao L; **Final approval:** Zhao L.

DATA AVAILABILITY STATEMENT

All data were provided by Aviation Technology Key Laboratory of Aerodynamics Research in High-Speed and High Reynolds, Shenyang 110819, China.

FUNDING

National Natural Science Foundation of China 

Grant No: 61503069

Fundamental Research Funds for the Central Universities

Grant No: N2404010

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

Ayesha S, Hanif MK, Talib R (2020) Overview and comparative study of dimensionality reduction techniques for high-dimensional data. *Inform Fusion* 59:44-58. <https://doi.org/10.1016/j.inffus.2020.01.005>

Chen C, Thiele K (2023) Wind tunnel tests on the unsteady galloping of a bridge deck with open cross section in turbulent flow. *J Wind Eng Ind Aerod* 233. <https://doi.org/10.1016/j.jweia.2022.105293>

De Gregorio F (2012) Flow field characterization and interactional aerodynamics analysis of a complete helicopter. *Aerosp Sci Technol* 19(1):19-36. <https://doi.org/10.1016/j.ast.2011.11.002>

- Fu Y, Zhao ZK, Ma XL (2022) Simulation calculation of aerodynamic performance of Mira model based on virtual wind tunnel. *Proc Spie* 12259. <https://doi.org/10.1117/12.2638771>
- Gunasekaran H, Thangaraj T, Jana T, Kaushik M (2020) Effects of wall ventilation on the shock-wave/viscous-layer interactions in a Mach 2.2 intake. *Processes* 8(2). <https://doi.org/10.3390/pr8020208>
- Guo J, Zhang R, Cui XC, Ma WT, Zhao LP (2023) Multimode wind tunnel flow field system monitoring based on KPLS. *Processes* 11(1). <https://doi.org/10.3390/pr11010178>
- Issa L, Lakkis I (2014) Reduced-order modeling of low Mach number unsteady microchannel flows. *J Fluid Eng-T Asme* 136(5). <https://doi.org/10.1115/1.4026199>
- Jiang QC, Yan XF (2015) Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA. *J Process Contr* 32:38-50. <https://doi.org/10.1016/j.jprocont.2015.04.014>
- Marinho LPB, Avelar AC, Fisch G, Roballo ST, Souza LE, Gielow R, Girardi RM (2009) Studies using wind tunnel to simulate the atmospheric boundary layer at the Alcântara Space Center. *J Aerosp Technol Manag* 1(1):91-98. Available in <https://jatm.com.br/jatm/article/view/11>
- Lee JM, Yoo CK, Choi SW, Vanrolleghem PA, Lee IB (2004) Nonlinear process monitoring using kernel principal component analysis. *Chem Eng Sci* 59(1):223-234. <https://doi.org/10.1016/j.ces.2003.09.012>
- Li JJ, Jing MM, Lu K, Zhu L, Shen HT (2019) Locality preserving joint transfer for domain adaptation. *IEEE T Image Process* 28(12):6103-6115. <https://doi.org/10.1109/Tip.2019.2924174>
- Li SN, Caracoglia L (2020) Wind tunnel experimental variability of aerodynamic loads for wind turbine blades. *J Phys Conf Ser* 1452. <https://doi.org/10.1088/1742-6596/1452/1/012054>
- Li Z, Tan C, Zhang SM, Dong F (2024) Global-local structure analysis-based status monitoring for oil-gas-water flow. *IEEE Sens J* 24(24):41610-41622. <https://doi.org/10.1109/Jsen.2024.3487176>
- Luo LJ, Sao SY, Mao JF, Tang D (2016) Nonlinear process monitoring based on kernel global-local preserving projections. *J Process Contr* 38:11-21. <https://doi.org/10.1016/j.jprocont.2015.12.005>
- Qiu HT, Shi MT, Zhu YD, Lee C (2024) Boundary layer transition of hypersonic flow over a delta wing. *J Fluid Mech* 980. <https://doi.org/10.1017/jfm.2024.42>
- Richter K, Rosemann H (2013) Numerical simulation of wind tunnel wall effects on the transonic flow around an airfoil model. *Notes Numer Fluid Me* 121:525-532. https://doi.org/10.1007/978-3-642-35680-3_62
- Sabnis K, Babinsky H (2019) Nozzle geometry effects on corner boundary layers in supersonic wind tunnels. *AIAA J* 57(8):3620-3623. <https://doi.org/10.2514/1.J058310>
- Silva MG, Gamarra VO, Koldaev V (2009) Control of Reynolds number in a high speed wind tunnel. *J Aerosp Technol Manag* 1(1):69-77. Available in <https://jatm.com.br/jatm/article/view/9/148>
- Tong CD, Yan XF (2014) Statistical process monitoring based on a multi-manifold projection algorithm. *Chemometr Intell Lab* 130:20-28. <https://doi.org/10.1016/j.chemolab.2013.09.006>
- Wang B, Mao ZZ (2019) Integrating Mach number prediction with outlier detection for wind tunnel systems. *J Aerospace Eng* 32(5). [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0001041](https://doi.org/10.1061/(ASCE)AS.1943-5525.0001041)
- Wang YG, Jian ST, Wang ZG, Fan R, Han P, Lyu Y (2023) An intelligent algorithm for aerodynamic parameters calibration of wind tunnel experiment at a high angle of attack. *Int J Aerospace Eng*. <https://doi.org/10.1155/2023/3093526>



- Yao T, Zhou SD, Wang M, Liao QX, Zhu JP (2023) Theoretical calibration and experimental test of hemispherical seven-hole probe under 3-D flow field in wind tunnel. *IEEE T Instrum Meas* 72. <https://doi.org/10.1109/TIM.2023.3301890>
- Yuan P, Zhao LP (2022) KPLS-Based Mach number prediction for multi-mode wind tunnel flow system. *Processes* 10(9). <https://doi.org/10.3390/pr10091718>
- Zhan C, Li S, Yang Y (2017) Enhanced fault detection based on ensemble global-local preserving projections with quantitative global-local structure analysis. *Ind Eng Chem Res* 56(38):10743-10755. <https://doi.org/10.1021/acs.iecr.7b01642>
- Zhang HL, Qi YS, Wang L, Gao XJ, Wang XC (2017) Fault detection and diagnosis of chemical process using enhanced KECA. *Chemometr Intell Lab* 161:61-69. <https://doi.org/10.1016/j.chemolab.2016.12.013>
- Zhang MG, Ge ZQ, Song ZH, Fu RW (2011) Global-local structure analysis model and its application for fault detection and identification. *Ind Eng Chem Res* 50(11):6837-6848. <https://doi.org/10.1021/ie102564d>
- Zhao HT, Lai ZH, Chen YD (2019) Global-and-local-structure-based neural network for fault detection. *Neural Networks* 118:43-53. <https://doi.org/10.1016/j.neunet.2019.05.022>
- Zhao HY, Yuan P, Mao ZZ, Wang BA (2021) Improving identification of wind tunnel systems using clustering-based ensemble outlier detection model. *J Aerospace Eng* 34(3). [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0001235](https://doi.org/10.1061/(ASCE)AS.1943-5525.0001235)
- Zhao HY, Yu D, Wang Y, Wang B (2020) Enhancing the prediction of Mach number in wind tunnel with a regression-based outlier detection framework. *IEEE Access* 9:27668-27677. <https://doi.org/10.1109/Access.2021.3058096>
- Zhao Z, Zhao CC, Pei ZC, Chen WH (2021) Design and implementation of a low-speed wind tunnel system. Paper presented 2021 16th Conference on Industrial Electronics and Applications. IEEE Industrial Electronics Chapter of Singapore, Southwest Jiaotong University, and IEEE Singapore Section; Chengdu, China. <https://doi.org/10.1109/Icica51954.2021.9516145>