

# A Hybrid Approach for Reconstruction of Transonic Buffet Aerodynamic Noise: Integrating Random Forest and Compressive Sensing Algorithm

Qiao Zhang<sup>1,2</sup>, Weiwei Zhang<sup>1,2\*</sup>

<sup>1</sup>School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, PR China <sup>2</sup>Intermational Joint Institute of Artificial Intelligence on Fluid Mechanics, Xi'an 710072, PR China

### **Abstract**

In response to the difficulty in obtaining high-precision aerodynamic noise data, this paper establishes a comprehensive standardization process for predicting the transonic buffet aerodynamic noise of RAE2822 airfoil. Firstly, a prior criterion is proposed based on flow correlation and the prediction accuracy of the Power Spectral Density (PSD) using the Random Forest (RF) algorithm. Subsequently, we determine whether the RF algorithm can be employed to directly obtain high-precision PSD results using this criterion. Successful PSD prediction is specifically determined when monitoring points simultaneously satisfy RMSE corr<0.05 and RMSE adj<0.05, the RMSE corr represents the Root Mean Square Error (RMSE) of the autocorrelation and the cross-correlation coefficients, and the RMSE adj indicates the RMSE of the PSD for the adjacent monitoring points. If not, we introduce an innovative approach by embedding the RF model into the Compressed Sensing algorithm reconstruction process (RF CS). This method efficiently achieves highprecision Overall Sound Pressure Level (SPL) and PSD reconstruction based on sparse sensor positions, demonstrating good robustness and generalization capabilities. Compared to the CS algorithm based on Proper Orthogonal Decomposition (POD\_CS), this method achieves high-precision PSD (SPL) reconstruction, with the RMSE has been reduced by a factor of 2 to 50 using 22 (9) sensor positions and 15 (12) basis functions, and the method does not exhibit phenomena such as high-frequency distortion or inflection point distortion.

**Keywords:** Random forest; Overall sound pressure level; Power spectral density; Flow correlation;

### 1. Introduction

Efficient and accurate prediction of aerodynamic noise is of paramount importance for flow control and noise reduction design [1-5]. Researchers have developed empirical and semi-empirical theoretical models by integrating wind tunnel experimental measurements with mathematical tools [6,7], such as the Rossiters' cavity aerodynamic noise prediction model [8]. Although this method is simple, intuitive, and efficient, its application is limited due to the lack of flexibility in accommodating arbitrary geometric shapes, flow conditions, and varied monitoring point positions. Numerical simulations, while relatively flexible and adjustable concerning geometric shapes and flow states, may encounter convergence issues even when meeting requirements for highprecision numerical schemes, low dissipation, and low dispersion [4]. Experimental measurements, on the other hand, offer an effective means of validating numerical simulations, aerodynamic noise theory, and optimizing aerodynamic noise reduction designs [9-13]. However, conducting experimental measurements demands a substantial number of dynamic pressure sensors, static pressure sensors, and a far-field microphone array [14]. Additionally, an acoustic wind tunnel necessitates specialized sound absorption treatments, leading to high acoustic measurement costs. While traditional methods of acquiring aerodynamic noise data may pose challenges, they have yielded a vast amount of multi-precision datasets. The introduction of Machine Learning (ML) [15-18] methods has provided technical support for obtaining high-precision aerodynamic noise. Tenney et al. [19] used the Deep Neural Network (DNN) to predict a complex non-axisymmetric jet noise.

Research shows that the DNN was able to predict directional far-field Sound Pressure Level (SPL) within ±0.75dB, and the absolute error of Overall Sound Pressure Level (OASPL) was maintained within 0.3dB [20]. Wu et al. [21] utilized a DNN to establish a multivariate nonlinear regression model between state variables such as Mach number (Ma), Angle of Attack (AOA), and the pressure coefficients or SPL. This investigation ensured that the maximum error in predicting the pressure coefficient remained within 1%, and the relative error of the maximum SPL for the first three-order flow-induced oscillation sound modes was controlled within 0.5%. However, it faced challenges in achieving accurate predictions of high-frequency characteristics and the generalization of flow states.

Building on this, researchers have focused on improving the prediction accuracy of SPL and SPL or Power Spectral Density (PSD) by considering both the number of samples and algorithm perspectives. Centracchio et al. [22] employed active metamodelling with the Artificial Neural Network (ANN) to predict jet noise under near-field conditions with under-expanded flow, incorporating an uncertainty function to estimate the reliability of the model itself. The uncertainty map revealed that the regions of the domain where additional experiments should be carried out are in correspondence with the Screech tone. Legendre et al. [23] also emphasized the critical importance of data quantity for predicting the SPL of multi-propeller drones. An adequate amount of data could yield SPL highly consistent with numerical simulations. However, the broadband trend of the prediction diverges from the experimental results due to the lack of sufficient computation of broadband noise.

Regarding the approach to supplementing data, Zhou et al. [24] devised a multi-fidelity ML model based on Active Learning (AL) and Transfer Learning (TL). The AL algorithm was designed to minimize the predictive error of the ML model at two far-field observer locations in order to intelligently select the necessary high-precision experimental measurements. Then, the TL algorithm iteratively enhanced the model by introducing a small number of high-fidelity data, resulting in a reduction of the Mean Squared Error (MSE) of SPL from 41.3 to 14.1. Arina et al. [25] observed that the Random Forest (RF) algorithm demonstrated significant advantages in predicting the SPL, enabling high-precision predictions of low intensity lobes. In contrast, the Artificial Neural Networks (ANN) could not achieve high-precision SPL predictions, even with a large amount of data.

Some researchers have adopted a different approach, aiming to achieve high-precision database expansion from an algorithmic perspective. In 2013, Huang et al. [26] first applied Compressive Sensing (CS) algorithms to the problem of duct acoustic mode reconstruction [27]. This method enables the reconstruction of the frequency characteristics of a linear-duct acoustic problem, reducing the required number of sensors (Nuquist Sampling Theorem) by approximately five times. Moreover, the study discovered that background noise interference could alter the sparsity of the whole measurements, and when the Signal-Noise Ratio (SNR) exceeds a certain range, it may lead to reconstruction distortion or even failure [28]. Building on this, Bu et al. [29] proposed an improved Compressive Sensing algorithm (CSM-II) from the perspective of enhancing sparsity. This algorithm can effectively suppress noise interference to a certain extent, enabling high-precision reconstruction of acoustic mode spectra under low SNR conditions.

To enhance the solution speed, Huang [30] developed the machine-learning-based mode detection method, which has been incorporated into the compressive sensing algorithm. Among these, the neural network architecture replaces the process of solving the measurement matrix between acoustic mode inputs and measurements. This approach efficiently obtains the amplitude of acoustic modes. Behn et al. [31] utilized the compressed sensing-based mode analysis to investigate the tonal and broadband sound generation and transmission in the ACAT1 fan stage. Additionally, Bu et al. [32] applied the Compressive Sensing algorithm to fan prediction and health management in aviation engines. This involved inferring possible incidents of changes in the fan noise mode spectrum using only a few acoustic sensors.

Researchers have conducted extensive studies on the prediction accuracy, data quantity, and ML algorithms for aerodynamic noise. However, there are still certain issues: (1) While the importance of data quantity for prediction accuracy is emphasized, there is a lack of quantitative analysis regarding the correlation between prediction accuracy and data quantity. (2) The mentioned

studies highlight the significant advantages of the RF algorithm in predicting SPL. They also propose using a combination of Compressive Sensing and neural networks to improve reconstruction accuracy. However, challenges such as insufficient accuracy and the inability to achieve generalization still exist.

Based on this, this paper proposes a comprehensive method for predicting high-precision aerodynamic noise parameters, such as PSD, SPL, and SPL. It develops an efficient and high-precision prediction approach for PSD and SPL using a Compressive Sensing algorithm based on the RF model (RF\_CS). Firstly, the correlation between data quantity and PSD prediction accuracy is quantitatively established based on flow correlations. Secondly, it is determined whether each monitoring point satisfies flow correlations. If flow correlations are satisfied, the RF algorithm can be directly used to predict PSD results. If not, the RF\_CS method is employed to accurately reconstruct PSD and SPL results, achieving a reduction in the number of acoustic sensors and lowering implementation costs without sacrificing data precision.

### 2. Model and method

### 2.1 Data set and results verification

Using the transonic buffeting aerodynamic noise of the RAE2822 airfoil as a test case, this paper employs the Delayed Detached Eddy Simulation (DDES) with the Spalart-Allmaras (S-A) turbulence model [33] to obtain the pressure distribution on the upper surface of the airfoil at various monitoring points. A total of 26 points are sampled evenly along  $x/c=-0.0018\sim1.14$ , as shown in Fig. 1. The entire computational process is divided into three steps.

- (1). Mesh Generation. The computational domain of the flow field consists of a combination of a semi-circle and a square. The semi-circle is centered at the leading edge of the airfoil, with a radius equal to 30 times the airfoil chord length (c). The square has lengths of 60c and 30c, forming the long and short sides, respectively. A C-type structured grid is employed, and details are illustrated in Fig. 2. The mesh scale and the distribution of nodes in each layer meet the requirements of the DDES method for the boundary layer grid [34].
- (2). Use the results obtained by the Reynolds-Averaged Navier-Stokes (RANS) model as the initial conditions for the DDES method. Due to the robustness and convergence of the S-A turbulence model in simulating transonic buffet, which can accurately capture the separated flow induced by shock wave boundary layer interaction, the S-A turbulence model is chosen for the DDES algorithm in numerical simulation [35]. Detailed parameters and formulas can be found in the reference [36].

The DDES model improves the expression of the shielding function compared to the DES model, with the shielding function fd described by Eq. (1) and Eq. (2).

$$\tilde{d}_{DDES} = d - f_d \max(0, d - C_{DES}\Delta), f_d = 1 - \tanh[(8r_d)^3]$$
 (1)

$$r_d = \frac{v_t + v}{\sqrt{U_{i,j}U_{i,j}}k^2d^2}$$
 (2)

In Eq. (1), Ui,j,vt represent the velocity gradient and turbulent viscosity coefficient, respectively. The shielding function fd ensures that the LES model does not prematurely enter the wall boundary layer region. When fd=0, the LES model is not activated, and in large separated flow regions, fd increases, gradually transitioning the DDES model to the LES model.

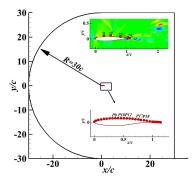
. For the various flow conditions obtained in step (2), the Welch method (using the Hamming window) is applied to the pressure signals at various monitoring points to obtain the PSD. Based on the distribution of PSD, the SPL distribution is determined using Eq. (3).

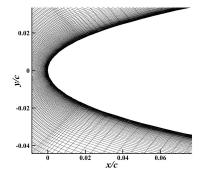
$$SPL = 10log_{10} \left( \frac{1}{p_0^2} \int_0^\infty PSD(f) df \right)$$
 (3)

Where  $p_0 = 2 \times 10^{-5} Pa$ ,  $\infty$  is the maximum frequency.

The Computational Fluid Dynamics (CFD) method in this study covers flow conditions with Mach

numbers (Ma) ranging from 0.71 to 0.76 and Angles Of Attack (AOA) ranging from 3.4° to 6.8°, totaling 67 different flow states, as illustrated in Fig. 3. For each flow condition, data are obtained from 26 monitoring points on the upper surface of the wing, as shown in the bottom right corner of Fig. 1. Accordingly, the complete datasets for SPL and PSD are obtained through steps (1)-(3), comprising a total of 67 sets of SPL samples and 1768 sets of PSD samples.





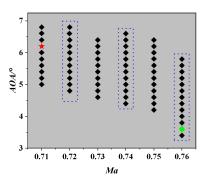
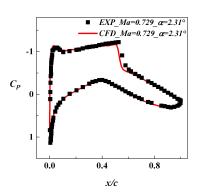


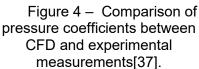
Figure 1 – The computational domain and the location of monitoring points.

Figure 2 – The mesh detail.

Figure 3 – The flow states.

To demonstrate the reliability of the numerical simulations, this paper compares and validates the numerical simulation in terms of steady, unsteady, and aerodynamic noise. Firstly, the pressure coefficients obtained from the RANS model in this study closely match the experimental measurements [37], confirming the accuracy of the CFD results, as shown in Fig. 4. Secondly, when comparing the numerical simulations of unsteady flow in this paper with the calculations of the transonic buffet onset boundary by Tian et al. [38], the relative error is found to be within 5%, validating the accuracy of the unsteady numerical simulations, as seen in Fig. 5. Additionally, by comparing the SPL at different monitoring points on the cavity floor of the C201 cavity with experimental measurements, the overall trend of the computed results aligns with the experimental data, providing basic validation of the accuracy of the aerodynamic noise acquisition results in this paper [39], as illustrated in Fig. 6.





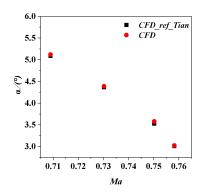


Figure 5 – Comparison of the RAE2822 buffeting onset boundary[38].

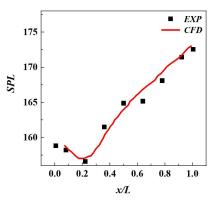


Figure 6 – Comparison of the SPL of the C201 cavities[39].

# 2.2 Introduction of Computational methods

Random Forest algorithm exhibits advantages in predicting both SPL and PSD, however, it does not consistently provide high-precision PSD predictions across all spatial intervals. Based on these findings, the study proposes a quantitative flow correlation prior criterion derived from PSD prediction results. In cases where predictive accuracy is inadequate, an efficient and high-precision PSD and SPL reconstruction algorithm based on CS is suggested. The study also conducts a

comparative analysis of the reconstruction capabilities between the POD basis function and the RF basis function, as illustrated in the upper part of Fig. 7. The specific workflow is as follows.

### Step 1: Dataset Construction

Obtain PSD for various monitoring points under different flow conditions and calculate the corresponding SPL for each flow condition.

### Step 2: Flow Correlation Analysis and Random Forest Prediction of PSD

Calculate the Root Mean Square Error (RMSE) of PSD results for adjacent monitoring points (RMSE\_adj). Utilize the RF algorithm to predict PSD and analyze the precision of PSD prediction (RMSE\_PSD). Evaluate the RMSE of the auto-correlation function at this point and the cross-correlation function between adjacent points (RMSE\_corr). Establish a criterion for the precision of RF algorithm predictions based on flow correlation. Refer to Section 2.2.3 for an introduction to the RF model.

# Step 3: Obtain the criterion for flow correlation

Perform auto-correlation and cross-correlation analyses on the PSD of each monitoring point to determine whether the criteria of RMSE\_corr < 0.05 and RMSE\_adj < 0.05 for flow correlation are met. If satisfied, it is considered that the flow phenomena at this point are highly correlated with neighboring points, and high-precision PSD can be directly obtained using the RF algorithm. Conversely, if the flow correlation criteria are not met, it is considered that the correlation of flow phenomena at this point with neighboring points is not high, and a CS algorithm is needed to reconstruct high-precision PSD results.

For points that satisfy the flow correlation criteria, the RF algorithm can be directly employed for high-precision PSD prediction, following the modeling process outlined in Section 2.2.3. For points that do not meet the flow correlation criteria, a CS algorithm is applied to reconstruct high-precision PSD results for the corresponding points, as illustrated in the lower part of Fig. 7. The specific workflow is outlined below.

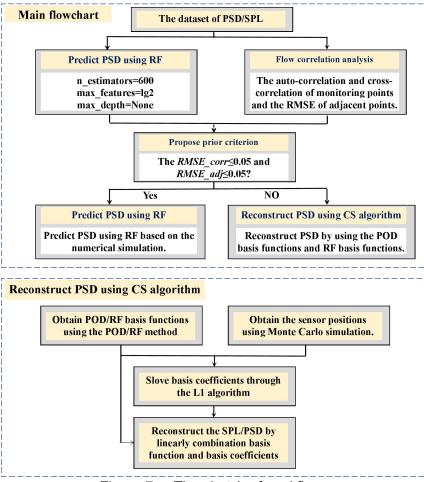


Figure 7 – The sketch of workflow.

## Step 4: Obtain basis functions

Considering the advantages of the RF algorithm in predicting PSD and SPL, this study contemplates using RF as the basis function for reconstructing high-precision PSD and SPL results, supplemented by a comparison with the Proper Orthogonal Decomposition (POD) basis function.

Step 5: Obtain sensor positions and corresponding PSD or SPL results

Monte Carlo simulation is employed for the selection of sparse sensor positions. By traversing combinations of sparse monitoring point positions through Monte Carlo simulation and using the RMSE between the reconstruction results and the original PSD results as a criterion, optimal sensor positions are determined.

Step 6: Use the L1 optimization algorithm to obtain basis coefficients

Utilize the L1 algorithm in conjunction with basis functions and the PSD or SPL results corresponding to the selected sparse monitoring points to solve for basis coefficients.

Step 7: Linearly combine basis functions and basis coefficients to reconstruct PSD and SPL

Linearly combine basis functions and basis coefficients to reconstruct PSD and SPL. Assess the reconstruction accuracy using the RMSE.

# 2.2.1 Compressed sensing method

Compressive sensing, also known as compressive sampling or sparse reconstruction, is a technique employed when the original signal exhibits sparsity in a specific domain. If the original signal is sparse in a certain domain, optimization algorithms [40-42] or greedy algorithms [43], among others, can be utilized to reconstruct the signal reliably using a small number of measurement points (sensor positions). The goal of CS algorithms is to reconstruct the original signal (x) from sparsely observed values (y), satisfying the equation  $y = \phi x$ , where  $\phi$  represents the measurement matrix. The original signal x can be sparsely represented by projecting it onto a particular feature space, such as Fourier basis (FFT), POD basis, and so on, which satisfies the equation  $x = \psi a$ . Based on this, the sparsely observed values y fulfill the equation  $y = \phi \psi x$ , as illustrated in Fig. 8.

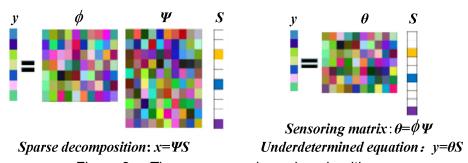


Figure 8 – The compressed sensing algorithm.

The selection of basis functions and basis coefficients is crucial in the CS algorithm. Huang [26] pointed out that for simple problems, the FFT basis function can be chosen, while for complex high-dimensional problems, the POD basis function, among others, should be selected. Considering the prediction of the PSD and the SPL in aerodynamic noise in this paper, and given the excellent predictive accuracy of the RF algorithm in this regard, this paper ingeniously incorporates the RF algorithm into the CS algorithm to achieve high-precision reconstruction of PSD and SPL. The POD basis function is used as a comparison. The basis coefficients  $a \in R^M$  (M is the number of basis functions) can be solved using optimization algorithms, as shown in Eq. (4).

$$\min_{a} \|\theta a - y\|_{0} \tag{4}$$

If  $P \le M$ , the CS algorithm is employed to solve Eq. (4), implying the use of L1 norm minimization to determine the basis coefficients.

$$\min_{a} \|a\|_{1}$$
s.t.  $y = \theta a$  (5)

If P > M, the basis coefficients can be solved by the minimization of the L2 norm.

$$\min_{a} ||a||_{2}$$
s.t.  $y = \theta a$  (6)

The PSD or SPL is obtained through the linear combination of basis functions and basis coefficients, as described in Eq.(7).

$$\widehat{x} = \psi a \tag{7}$$

Where  $\hat{x}$  represents the reconstruction result of PSD or SPL. When using POD basis functions,  $\psi$  corresponds to the POD basis functions. Otherwise,  $\psi$  denotes the RF basis functions. Detailed introductions to the POD and RF algorithms are provided in Sections 2.2.2 and 2.2.3, respectively.

# 2.2.2 Proper Orthogonal Decomposition

The POD method is commonly used in turbulence analysis [44, 45], image processing [46], structural dynamics analysis [47], and various applications involving other dimensionality reduction. Its main principle involves acquiring a set of orthogonal basis functions through matrix transformation and orthogonal decomposition of flow field samples.

To reduce the dimensionality of matrix decomposition and lower data processing costs, the Snapshot-POD method is frequently employed. First, the correlation matrix in Eq. (8) is defined.

$$C = C_P^T * C_P \tag{8}$$

Solve the eigenvalue of correlation matrix C.

$$CA^{[j]} = \lambda_j A^{[j]} \tag{9}$$

Here,  $A^{[j]}$  represents the j-th order basis coefficient,  $\lambda_j$  is the j-th eigenvalue, and the ratio of the sum of the first j eigenvalues to the total eigenvalues reflects the proportion of mode energy captured by the first j modes. The basis functions of POD (sparse dictionary set in CS algorithms) are defined as shown in Eq. (10).

$$\psi_j = \frac{1}{\sqrt{\lambda_j}} C_P * A^{[j]} \tag{10}$$

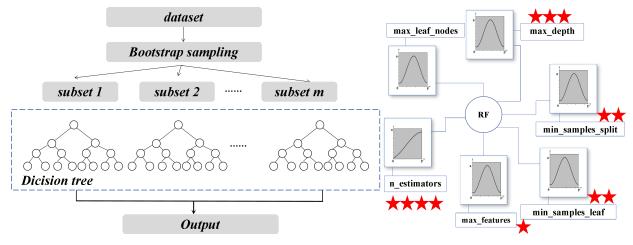
### 2.2.3 Random forest model

The RF algorithm is a parallel ensemble ML method introduced by Breiman [48]. In comparison to traditional neural network regression models, it employs an ensemble learning framework for parameter learning and model prediction. Through the Bootstrap sampling method, n samples are randomly selected with replacement from the training samples to obtain m subsets. A decision tree is trained individually for each subset, and the average of the predicted results from m decision trees serves as the output for the regression RF, as shown in Fig. 9.

In the process of predicting PSD, the state variables (Ma, AOA, f) are used as inputs, and PSD is the output. The training set and testing set samples are divided into 85% and 15%, respectively. Taking Ma=0.71, AOA=6.2° as an example, the PSD of 22 monitoring points is chosen as the training set, while the PSD for 4 monitoring points is used as the prediction set. Firstly, the PSD results are standardized, with the area enclosed by the PSD of each monitoring point as the normalization basis. Subsequently, a grid random search is conducted to determine the state variables for the RF, such as n\_estimators=600, max\_features=lg2, min\_samples\_split=2, max\_depth=None. The RMSE is used as the criterion for RF node splitting. Finally, an RF algorithm prediction model is established, and the error between the PSD obtained by the RF algorithm and the CFD method is compared.

In the SPL modeling process using the RF algorithm, the state variables (Ma and AOA) are used as inputs, and the SPL is the output. The training set and testing set follow a ratio of 0.85:0.15, where

57 SPL samples for different flow conditions are used as the training set, and data samples for SPL from 10 flow conditions are used as the testing set. A grid random search is employed to determine various parameters for the RF, such as n\_estimators=100, max\_features=lg2, min\_samples\_split=2, max\_depth=None. Finally, the RF algorithm is used to model and obtain the prediction error for SPL.



- (a) The random forest model.
- (b) The important network parameters.

Figure 9 – The random forest model.

# 3. The prior criterion of flow correlation

Using the RF algorithm to predict PSD, it was observed that not all monitoring points could yield high-precision PSD. However, the variability in PSD is primarily attributed to differences in flow phenomena. Therefore, the spatial variation in point locations can impact the prediction accuracy. In this study, the influence of spatial positions on prediction accuracy was quantitatively analyzed from the perspective of flow correlation [35]. The Pearson cross-correlation coefficient [49] was employed to analyze the correlation between adjacent monitoring points, as shown in Eq. (11).

$$\rho_{P_x, P_y} = \frac{\text{cov}(P_x, P_y)}{\sigma_{P_x} \sigma_{P_y}} = \frac{\sum_{i=1}^{n} (P_x - \overline{P}_x)(P_y - \overline{P}_y)}{\sqrt{\sum_{i=1}^{n} (P_x - \overline{P}_x)^2} \sqrt{\sum_{i=1}^{n} (P_y - \overline{P}_y)^2}}$$
(11)

Where  $ho_{P_x,P_y}$  represent the auto-correlation coefficient of PSD of monitoring points.

Taking Ma=0.71, AOA=6.2° as an example, the analysis involves examining the auto-correlation and cross-correlation coefficients of various monitoring points and quantitatively evaluating the influence of spatial spacing on prediction accuracy. The Eq. (12) is used as a metric to measure the errors in auto-correlation and cross-correlation coefficients.

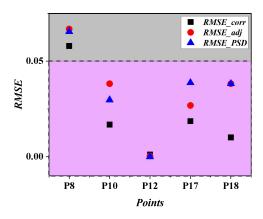
$$RMSE\_corr = \sqrt{\frac{1}{n} \sum_{t=-700}^{200} (auto\_corr - cross\_corr)^2}$$
 (12)

Where *auto\_corr* represents the auto-correlation coefficient, and *cross\_corr* represents the cross-correlation coefficient.

For the flow condition at Ma=0.71, AOA=6.2°, five monitoring points (P8-P18) are randomly selected. The RMSE of auto-correlation coefficients and cross-correlation coefficients at each point (RMSE\_corr), the RMSE between the PSD results for adjacent monitoring points (RMSE\_adj), and the RMSE between the PSD prediction results and CFD results (RMSE\_PSD) are calculated. As shown in Fig. 10, it is observed that RMSE\_PSD is highly correlated with RMSE\_corr, but it depends more on the similarity between the PSD of adjacent monitoring points. With the increase in RMSE\_corr, RMSE\_PSD also increases, but it consistently remains within the same order of magnitude as RMSE\_adj. For example, at P17, RMSE\_corr=8.25e-4, RMSE\_adj=3.87e-2, and

RMSE\_PSD=3.87e-2. However, when RMSE\_corr=4.93e-2 and RMSE\_adj=6.68e-2, the predicted error RMSE\_PSD=6.54e-2 at point P8.

The auto-correlation function, cross-correlation function, and their predicted results for this point are illustrated in Fig. 11. As observed from the figure, the prediction results of PSD exhibit significant deviations at f=13Hz. Subsequently, there is a rightward shift in the peak frequencies, causing substantial discrepancies in the range of f=100-200Hz, failing to meet the precision requirements for PSD prediction in this study. Based on this observation, a priori criteria for successful prediction PSD is proposed, considering both flow correlation, as RMSE\_corr and RMSE\_adj. Specifically, when the monitoring point satisfies both RMSE\_corr<0.05 and RMSE\_adj<0.05 simultaneously, ensuring RMSE\_PSD<0.05, it is considered a successful prediction.



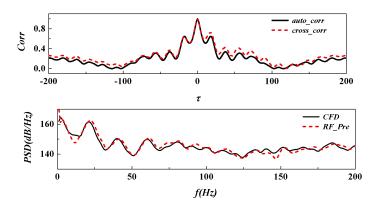


Figure 10 – The RMSE\_corr and RMSE\_adj of monitoring points at Ma=0.71, AOA=6.2°.

Figure 11 – The correlation characteristic and the prediction results at point 8.

Taking the flow condition with Ma=0.71, AOA=6.2° as an example, an analysis is conducted on all monitoring points on the upper surface of the wing and their PSD prediction results to examine the applicability of the prior criterion. The results of RMSE\_adj and RMSE\_corr for this flow condition are presented in Fig. 12. The pink region represents the area where the flow correlation criterion is satisfied, indicating that the RF algorithm can be directly used for prediction. As illustrated in Fig. 13, the predicted result at point P18 using the RF algorithm exhibits high agreement with CFD results, with RMSE\_PSD=3.82e-2. Conversely, the gray region in Fig. 12 signifies the area where the flow correlation criterion is not satisfied, and the RF algorithm cannot be directly applied for prediction. The predicted result at point P9, as shown in Fig. 13, reveals that only the first three peak frequencies are essentially consistent, and the high-frequency prediction accuracy is significantly insufficient, with a RMSE\_PSD=7.09e-2, exceeding 0.05. This validates the rationality of the flow correlation criterion proposed in this study.

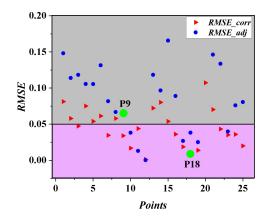


Figure 12 – The RMSE\_corr and RMSE adj of monitoring points at Ma=0.71,

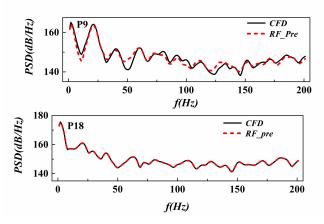


Figure 13 – The test of the criterion in flow correlation.

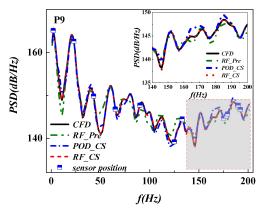
AOA=6.2°.

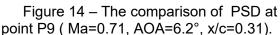
In summary, the flow correlation criteria, with RMSE\_corr < 0.05 and RMSE\_adj < 0.05, serve as pre-determinants to assess whether a monitoring point can use the RF algorithm for direct, high-precision PSD predictions. When the flow correlation prior criteria are satisfied, the RF algorithm is directly employed for modeling. Conversely, for situations not meeting these criteria, this study proposes the use of the CS method to reconstruct higher-precision PSD. The selection of basis functions in the CS algorithm is crucial for reconstruction effectiveness. Leveraging the advantages of the RF algorithm in predicting PSD and SPL, this paper integrates the RF model into the CS algorithm reconstruction process, providing comparisons with POD basis function reconstruction, as detailed in Section 4.

# 4. The reconstruction of frequency domain (PSD)

The analysis of flow correlation criteria reveals that, under the specified flow conditions (Ma=0.71, AOA=6.2°), only a few monitoring points can directly obtain high-precision PSD using the RF algorithm. The remaining points require additional monitoring points or alternative methods to achieve high-precision PSD, as illustrated in Fig. 12. Taking points P9 and P13 as examples, the PSD reconstruction accuracy using the CS algorithm was tested in this study. The POD CS method, employing the first 20 POD modes selected based on the mode energy percentage, and the RF CS method, using the first 15 RF basis functions, were both employed for PSD reconstruction, as shown by the blue dashed line and red lines in Fig. 14 and Fig. 15, respectively. Observations indicate that both POD CS and RF CS methods can effectively reconstruct highprecision PSD, improving the predictions of the RF algorithm (depicted by the deep green dashed lines in the figures), bringing the PSD predictions closer to the ground truth. However, the PSD results reconstructed using POD basis functions are prone to high-frequency distortion, especially above f=165Hz, as shown in the upper right corner of Fig. 14. Specifically, the POD CS method reduces the RF prediction error at point P9 from 0.063 to 0.0204, while the RF CS method further reduces the RF prediction error to 0.0030. This demonstrates the efficiency of the RF model as a basis for obtaining high-precision PSD results.

Using the RF algorithm for modeling and prediction at point P13 yields a prediction accuracy of 0.13. Introducing the CS algorithm aligns the prediction results closely with the CFD results, as shown in the upper right corner of Fig. 15. Specifically, the POD\_CS method reduces the PSD prediction error to 0.0088, and the RF\_CS method further reduces the PSD prediction error to 0.0013. This fully validates the feasibility of the CS algorithm in obtaining high-precision PSD results, highlighting the RF\_CS method in this study can reconstruct higher-precision PSD with fewer sensor positions and basis functions, making it more advantageous.





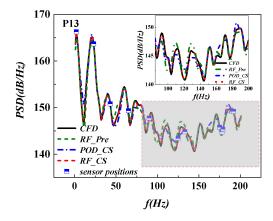


Figure 15 – The comparison of PSD at point P13 (Ma=0.71, AOA=6.2°, x/c=0.51).

Using Ma=0.76 and AOA=3.6° as an example, we further tested the generalization of the algorithm. The same sensor positions were selected (indicated by blue squares in Fig. 16 and Fig.17), and the same reconstruction method was used to obtain PSD. It is observed that both the POD CS

and RF\_CS methods significantly improve the RF prediction results, reducing the prediction error by about three orders of magnitude from 0.1314. Although the reconstruction accuracy of the two basis functions is similar, the RF\_CS method still achieves a smaller RMSE result (RMSE\_PSD=3.9719e-4), reducing the error by approximately half compared to the results using the POD\_CS method, as shown in the upper right corner of Fig. 16.

Maintaining the same sensor positions, the PSD reconstruction results for different monitoring points under the same state were tested, as shown in Fig. 17. It was found that with 22 sensor positions, high-precision PSD reconstruction results could be obtained using both the POD\_CS and RF\_CS methods. This indicates that with an adequate number of sensors, high-precision reconstruction of PSD results can be achieved without changing the sensor positions, and the RF\_CS method exhibits higher prediction accuracy with RMSE\_PSD=4.35e-4. Overall, this algorithm demonstrates generalization capabilities concerning flow states and spatial point positions, allowing for the acquisition of variable-state PSD results with an appropriate number of sensors at unchanged positions.

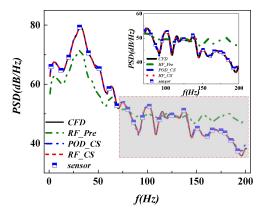


Figure 16 – The comparison of PSD results using different methods at point P1 (Ma=0.76, AOA=3.6°, x/c=-0.0018).

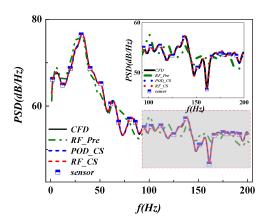


Figure 17 – The comparison of PSD using different methods at point P21 (Ma=0.76, AOA=3.6°, x/c=0.95).

Maintaining the same number of sensor positions and using the same method to obtain PSD results for Ma=0.71, AOA=6.2°, and Ma=0.76, AOA=3.6° flow conditions, the RMSE for each monitoring point is depicted in Fig. 18 and Fig. 19. The blue boxes represent the RMSE for POD\_CS reconstruction, while the red boxes represent the RMSE for RF\_CS reconstruction. It is evident that, for each Ma and AOA, the RF\_CS method achieves higher reconstruction accuracy. For Ma=0.71, AOA=6.2°, the average RMSE for RF\_CS reconstruction at each monitoring point is maintained at 0.0039 (RMSE\_PSD=0.0039). For Ma=0.73, α=3.6°, the average RMSE for RF\_CS reconstruction at each monitoring point is 0.000783 (RMSE\_PSD=0.000783). In both cases, RF\_CS reconstruction achieves an average RMSE one order of magnitude lower than that of POD\_CS reconstruction, fully demonstrating the superiority of incorporating the RF algorithm as the basis function.

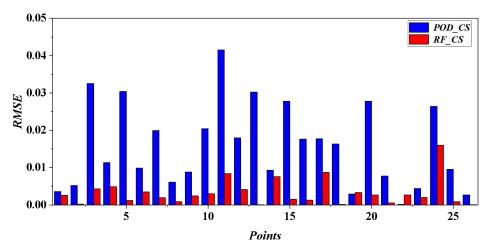


Figure 18 – The comparison of RMSE\_PSD results using POD\_CS and RF\_CS method at Ma=0.71, AOA=6.2°.

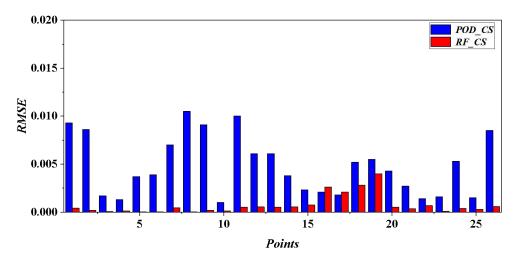


Figure 19 – The comparison of RMSE\_PSD results using POD\_CS and RF\_CS method at Ma=0.76, AOA=3.6°.

## 5. The reconstruction of spatial domain (SPL)

In the process of conducting experimental measurements for aerodynamic noise, a substantial number of static pressure and dynamic pressure sensors are required for near-field measurements. If sound source localization is involved, a far-field microphone array must also be arranged, resulting in high experimental measurement costs. To effectively control acoustic costs, the RF\_CS method can be employed for spatial reconstruction of the acoustic field. This not only reduces the number of required sensors but also serves as a reference for optimizing the arrangement of sensor positions in experimental measurements. Spatial reconstruction is similar to frequency-space reconstruction, and the error measurement indicator still utilizes the Eq. (13).

$$RMSE\_SPL = \sqrt{\frac{1}{N} \sum_{P=1}^{26} \left( SPL_{pre} - SPL \right)}$$
 (13)

Where  $SPL_{nre}$  and SPL denote the SPL obtained by reconstructed and calculated, respectively.

For the flow condition with Ma=0.71, AOA=6°, the reconstructed SPL using different methods is depicted in Fig. 20. The solid black line represents the CFD results, the blue dashed line represents the POD\_CS results, and the red dashed line represents the RF\_CS results. A comparison of the reconstruction accuracy of POD CS and RF CS method reveals that POD CS

exhibits significant errors at large turning points in SPL, such as near P5 and P23-P26, leading to substantial reconstruction errors for POD\_CS.

In the SPL reconstruction process, the RF\_CS method remains more advantageous for two main reasons. Firstly, using POD basis functions for reconstruction requires a higher number of basis functions and sensor positions. For instance, POD\_CS reconstruction necessitates 21 POD basis functions and 12 sensor positions, as indicated by the green pentagon in Fig. 20. In contrast, RF\_CS reconstruction only requires 18 basis functions and 9 sensor positions, represented by the blue square in Fig. 20. Secondly, the RF\_CS method achieves higher accuracy, with the RMSE\_SPL=0.0014 reconstructed by the POD\_CS method, while the RMSE\_SPL=3.2900e-04 reconstructed using the RF\_CS method. Comparing SPL prediction results under different flow conditions, the RMSE\_SPL=0.0079 reconstructed with the RF\_CS method is considerably lower than the RMSE\_SPL=0.0262 reconstructed with the POD\_CS method at Ma=0.72, AOA=5.2°, thereby validating the superior precision of the RF\_CS method in reconstruction.

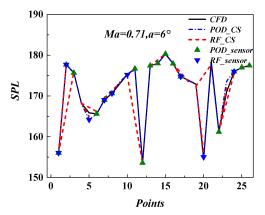


Figure 20 – The SPL of reconstruction using POD\_CS and RF\_CS method at Ma=0.71, AOA=6°.

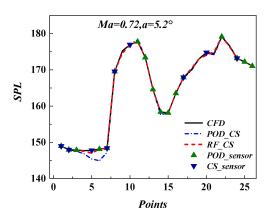


Figure 21– The SPL of reconstruction using POD\_CS and RF\_CS method at Ma=0.72, AOA=5.2.

Using the POD\_CS and RF\_CS methods to reconstruct the SPL distribution for the remaining flow states, we can obtain the corresponding RMSE\_SPL for each flow state. The RMSE for Ma=0.72, Ma=0.74, and Ma=0.76 at various AOA is shown in Fig. 22, with the flow conditions illustrated in Fig. 3. It is observed that, under the same Ma, as the AOA increases, the flow phenomena become more complex, leading to a slight decrease in reconstruction accuracy. Overall, the RMSE for POD\_CS reconstruction remains below 0.02, while the RMSE for RF\_CS reconstruction remains below 0.005. Specifically, for the Ma=0.72 flow condition, the average RMSE for POD\_CS reconstruction is 0.0049, and for RF\_CS reconstruction, it is 0.0011. For the Ma=0.74 flow condition, the average RMSE for POD\_CS reconstruction, it is 0.00139. For the Ma=0.76 flow condition, the average RMSE for POD\_CS reconstruction is 0.0068, and for RF\_CS reconstruction, it is 0.000943. This thoroughly validates the reconstruction accuracy of the RF\_CS method proposed in this paper, providing valuable insights for the distribution of surface sensors on the wing and effectively controlling measurement costs.

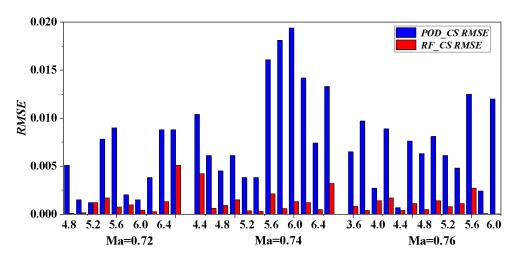


Figure 22 – The RMSE of reconstruction using POD\_CS and RF\_CS method.

### 6. Conclusion

This paper proposes a standardized process for obtaining high-precision Power Spectral Density (PSD) and Overall Sound Pressure Level (SPL) distributions. Firstly, a priori criterion for high-precision prediction of PSD results is established based on the flow correlation characteristic and Random Forest (RF) predictions. For monitoring points that meet this criterion, the RF can be directly applied for prediction. In cases where the flow correlation a priori criterion is not met, the paper introduces the RF-CS algorithm. This algorithm embeds the RF model into the compressive sensing reconstruction process, combining a small number of sensor positions to achieve high-precision reconstruction of PSD and SPL results. The main conclusions are as follows:

(1)Established a priori criterion based on flow correlation and RF prediction results. Specifically, when a monitoring point satisfies both RMSE\_corr < 0.05 and RMSE\_adj < 0.05, it ensures that RMSE\_PSD < 0.05. This indicates that using the RF algorithm can achieve high-precision prediction of PSD.

(2)For monitoring points that do not meet this criterion, the proposed RF-CS algorithm in this paper is employed to reconstruct high-precision PSD results. The results indicate that this algorithm can achieve higher precision PSD results with the same number of sensor positions, reducing the RMSE of predictions by 1-3 orders of magnitude. Comparing the POD\_CS and RF\_CS methods, the RF\_CS method has more advantages. It can use fewer sensor positions and basis functions to obtain higher precision PSD results, keeping the RMSE of reconstructed PSD results at the level of 10-3.

(3)The RF\_CS method can also be applied to the spatial reconstruction process of aerodynamic noise, achieving high-precision reconstruction of the SPL distribution at various monitoring points on the upper surface of the airfoil using approximately half of the monitoring points. Specifically, only 10 sensor positions are needed on the upper surface of the airfoil to obtain a complete and highly accurate SPL distribution, maintaining RMSE at the level of 10-3. In contrast, the mathematical foundation of the POD method relies on a linear relationship assumption, making it unable to achieve accurate PSD or SPL results without using a complete set of basis functions. This can lead to high-frequency distortion or inflection point distortion and even inaccurate reconstruction. This highlights the advantages of the RF\_CS method proposed in this paper.

While this paper has provided a comprehensive study of the POD\_CS and RF\_CS methods, there are still limitations in terms of the selection of sensor positions and experimental measurement support. Future research will explore different methods for selecting sensor positions and assess the applicability of the proposed approach with experimental measurement data.

## 7. Contact Author Email Address

The corresponding author email address is aeroelastic@nwpu.edu.cn (Weiwei Zhang) or you can

contact with the email with q zhang@mail.nwpu.edu.cn (Qiao Zhang).

# 8. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

### References

- [1] Chen L, Xu C, Lu X. Numerical investigation of the compressible flow past an aerofoil. J. Fluid Mech, Vol. 643, pp 97-126, 2010.
- [2] Schram C, Bennett G J. aerodynamic noise research in Europe: The CEAS-ASC report on 2022 highlights. J. Sound Vib, Vol. 568, pp 117895, 2023.
- [3] Nagy A B, Gareth J D. Bennett J, aerodynamic noise research in Europe: The CEAS-ASC report on 2020 & 2021 highlights. J. Sound Vib, Vol. 534, pp 117002, 2022.
- [4] Yang C, Arcondoulis E J G, Yang Y, Guo J, Maryami R, Bi C, Liu Y. Active control of airfoil turbulent boundary layer noise with trailing-edge blowing. J Acoust Soc Am, Vol. 153, No. 4, pp 2115, 2023.
- [5] Yang Y, Wang Y, Liu Y, Hu H, Li Z. Noise reduction and aerodynamics of isolated multi-copter rotors with serrated trailing edges during forward flight. J. Sound Vib, Vol. 489, pp 15688, 2021.
- [6] Goldstein M E, A unified approach to some recent developments in jet noise theory. Int. J Aeroacoust, Vol. 1, No. 1, pp 1-16, 2009.
- [7] Ahlefeldt T, Ernst D, Goudarzi A, Raumer H G, Spehr C. aerodynamic noise testing on a full aircraft model at high Reynolds numbers in the European Transonic Windtunnel. J. Sound Vib, Vol. 566, pp 17926, 2023.
- [8] Rossiter J E. Wind-Tunnel Experiments on the Flow over Rectangular Cavities at Subsonic and Transonic Speeds. R & M, Vol. 1412, pp 1-37,1966.
- [9] Farassat F, Casper J H. Towards an Airframe Noise Prediction Methodology: Survey of Current Approaches. in: 44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, 2006.
- [10]Liu X W, Hu Z W, Thompson D J, Jurdic V. Reduction of aerodynamic noise from square bars by introducing spanwise waviness. J. Sound Vib, Vol. 435, pp 323-349, 2018.
- [11]Sadawi L A, Chong T P, Kim J H. Aerodynamic noise reduction by plasma actuators for a flat plate with blunt trailing edge. J. Sound Vib, Vol. 439, pp. 173-193, 2019.
- [12]Li W, Nonomura T, Fujii K. Generation of Acoustic Disturbances in Supersonic Laminar Cavity Flows. J. Sound Vib, Vol. 20, pp 135-142,2015.
- [13]Fan Y, Vinuesa R, Gatti D, Schlatter P, Li W, Decomposition of the mean friction drag on an NACA4412 airfoil under uniform blowing/suction. J. Fluid Mech, Vol. 932, pp A31, 2022.
- [14] Yang D, Fan Z, Li J, Luo X, Jiang W. Suppression effect of reat-face angle of cavity on aerodynamic noise. J Exp Fluids, Vol. 24, No. 5, 24(5), pp 22-25, 2010.
- [15]L. Zhu, W. Zhang, J, Kou, Y. Liu. Machine learning methods for turbulence modeling in subsonic flows around airfoils. Phys. Fluids, Vol. 31, No. 015105, pp 1-14, 2019.
- [16] Chen W, Zhang W, Liu Y, Kou J. Accelerating the convergence of steady adjoint equations by dynamic mode decomposition. Struct Multidiscipl Optim, Vol. 31, No. 62, pp 747-756, 2020.
- [17]Zhu L, Zhang W, Sun X, Liu Y, Yuan X. Turbulence closure for high Reynolds number airfoil flows by deep neural networks. Aerosp Sci Technol, Vol.110, No. 106452, pp 1-18, 2020.
- [18]Bianco M J, Gerstoft P, Traer J, Ozanichet E, Roah M A, Gannot S, Deledale C A. Machine learning in acoustics: Theory and applications, J Acoust Soc Am, Vol. 146, No. 3590, pp 1-41, 2019.
- [19]Tenney A S, Glauser M N, Lewalle J. A Deep Learning Approach to Jet Noise Prediction. in: 2018 AIAA Aerospace Sciences Meeting, Kissimmee, Florida, 2018.

- [20]Tenney A S, Glauser M N, Ruscher C J, Berger Z P. Application of Artificial Neural Networks to Stochastic Estimation and Jet Noise Modeling. AIAA J, Vol. 58, No. 2, pp 647-658, 2020.
- [21]Wu J, Yang D, Zhang L, Gong T, Zhou F, Wang Y, Li Y. Investigation on artificial intelligence for the prediction of aerodynamic noise performances and controlling parameters optimization of aircraft. J Exp Fluids, Vol. 36, No. 3, pp 33-43, 2022.
- [22]Centracchio F, Meloni S, Kamliya K J, Azarpeyvand M, Roberto C, Umberto L. Under-expanded jet noise prediction using surrogate models based on artificial neural networks. in: 28th AIAA/CEAS aerodynamic noise 2022 Conference, Southampton, UK, 2022.
- [23]C. Legendre, F.A. Vincent, P. Athanasios, Y. Kitano, Y. Nakashima, W. Kobayash, G. Minorikawa. A machine learning-based methodology for computational aerodynamic noise predictions of multi-propeller drones. in: Inter-Noise and Noise-Con Congress and Conference Proceedings, Washington DC, 2949-3943, 2021.
- [24]Zhou B Y, Hanson L P, Pullin S F, Zang B, Hauth J, Huan X. A Data-Driven Approach for Enhancement of Propeller Performance Prediction. in: 28th AIAA/CEAS aerodynamic noise 2022 Conference, Southampton, UK, 2022.
- [25]Arina R, Ferrero A. Data-Driven aerodynamic noise Modelling: Trailing-Edge Noise. AIAA Aviation 2021 Forum, 2021.
- [26] Huang X, Compressive Sensing and Reconstruction in Measurements with an Aerospace Application. AIAA J, Vol. 51, No. 4, pp 1011-1016, 2013.
- [27]Bu H, Huang X, Zhang X. A conceptual study of utilizing compressive-sensing-based fan noise mode detection for aeroengine prognostic and health management. J Acoust Soc Am, Vol. 148, No. 2, pp 11063-1076, 2020.
- [28]Yu W, Huang X. Compressive sensing based spinning mode detections by in-duct microphone arrays. Measurement Science & Technology, Vol. 27, No. 055901, pp 1-12, 2016.
- [29]Bu H, Huang X, Zhang X. Compressive sensing method with enhanced sparsity for aeroengine duct mode detection. J Acoust Soc Am, Vol. 146, No. 1, pp EL39-EL44, 2019.
- [30] Huang X. A tutorial example of duct acoustics mode detections with machine-learning-based compressive sensing. J Acoust Soc Am. Vol, 146, No. 4, pp EL342-EL346, 2019.
- [31]Behn M, Tapken U. Investigation of Sound Generation and Transmission Effects Through the ACAT1 Fan Stage using Compressed Sensing-based Mode Analysis. in: 25th AIAA/CEAS aerodynamic noise Conference, Delft, Netherlands, 2019.
- [32]Bu H, Huang X, Zhang X. A conceptual study of utilizing compressive-sensing-based fan noise mode detection for aeroengine prognostic and health management. J Acoust Soc Am, Vol. 148, No. 2, pp 1063-1076, 2020.
- [33]Gao C, Zhang W, Kou J, Liu Y, Ye Z. Active control of transonic buffet flow. J. Fluid Mech, Vol. 824, pp 312-351, 2017.
- [34]Siggeirsson E M, Andersson N, Burak M O. Off design simulations of an S-shaped intermediate compressor duct: Experimental validation of DDES and RANS using G3D::Flow. AIAA Scitech 2020 Forum, Orlando, FL, 2020.
- [35]Zhang Q, Gao C, Zhou F, Yang D, Zhang W. Study on flow noise characteristic of transonic deep buffeting over an airfoil. Phys. Fluids, Vol. 35, pp. 046109, 2023.
- [36]Zhang Q, Gao C, Wang H, Zhang W, Yang D. Effects of bulb seal on slat flow dynamics and slat tones. Eur J Mech, Vol. 100, pp 124-140, 2023.
- [37]Cook P H, McDonald M A, Firmin M C P. Aerofoil RAE 2822: Pressure Distributions, and Boundary Layer and Wake Measurements. Experimental Data Base for Computer Program Assessment, AGARD Advisory Report ar 138, 1979.
- [38] Tian Y, Gao S, Liu P, Wang J. Transonic buffet control research with two types of shock control bump based on RAE2822 airfoil. Chinese J. Aeronaut, Vol. 30, No .5, pp 1681-1696, 2017.
- [39]Gao Q. Numerical Simulations and Physical Analyses of the Complex Turbulent Flow and Aerodynamic Noise in Cavities. CARDC, 2017.

- [40]Candes E J, Romberg J K, Tao T. Stable Signal Recovery from Incomplete and Inaccurate Measurements Via Regularized Orthogonal Matching Pursuit. IEEE Signal Process Mag, Vol. 59, No .8, pp 1207-1223, 2006.
- [41] Baraniuk R G. Compressive Sensing. IEEE Signal Process Mag, Vol. 24, pp 118-124, 2007.
- [42] Candes E J, Wakin M B. An Introduction To Compressive Sampling. IEEE Signal Process Mag, Vol. 25, No .2, pp 21-30, 2008.
- [43]Sahoo S K, and Makur A. Signal Recovery from Random Measurements via Extended Orthogonal Matching Pursuit. IEEE Trans. Signal Process, Vol. 63, No .10, pp 2572-2581, 2015.
- [44]Zhang W, Kou J, Liu Y. Prospect of artificial intelligence empowered fluid mechanics. Acta Aeronautica, Vol. 42, No . 4, pp 524689-524689, 2021.
- [45]Zhou B Y, Hanson L P, Pullin S F, Zang B, Hauth J. Huan X. A Data-Driven Approach for Enhancement of Propeller Performance Prediction. in: 28th AIAA/CEAS aerodynamic noise 2022 Conference, Southampton, UK, 2022.
- [46]Romberg J, Imaging via Compressive Sampling. IEEE Signal Process Mag, Vol. 25, No .2, pp 14-20, 2008.
- [47]Saito A, Kuno T. Data-driven experimental modal analysis by Dynamic Mode Decomposition. J. Sound Vib, Vol. 481, pp.115434, 2020.
- [48] Breimanl. Random forests, Machine learning, Vol. 45, No .1, pp 5–32, 2001.
- [49]Terracol M, Manoha E, Lemoine B. Investigation of the Unsteady Flow and Noise Generation in a Slat Cove. AIAA J, Vol. 54, No .2, pp 1-21, 2017.