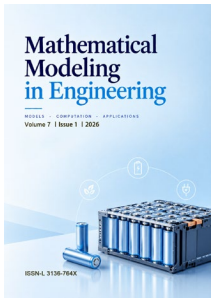
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 3, July, 2026



Multi-Source Res2Net-CBAM Model for Digital-Twin Calibration in Wind-Tunnel Testing

Hubert Tomasik[a,*], Wiktor Malinowski[a]

[a] *Faculty of Information Technology and Computer Science, Lodz University of Technology, Lodz, 90-924, Poland*

Received 16 April 2026; Revised 31 June 2026; Accepted 08 July 2026; Published 13 July 2026

Abstract

Wind-tunnel measurements remain essential for validating aerodynamic models, but digital-twin calibration must reconcile evidence that differs in spatial density, uncertainty, and failure mode. This study proposes a multi-source Res2Net-CBAM framework that jointly processes balance coefficients, pressure-tap arrays, flow-visualization images, and simulation-prior fields. Source-specific encoders project the four inputs into a shared representation; hierarchical Res2Net blocks capture local and wake-scale structures, while sequential channel and spatial attention suppress irrelevant features. A reliability estimator uses branch residuals, source-availability indicators, and measurement-quality descriptors to normalize the contribution of each source before a residual head updates aerodynamic coefficients and pressure-related field descriptors. Evaluation on 1,680 synchronized operating-condition windows shows a lift-coefficient mean absolute error of 0.014, a drag-coefficient mean absolute error of 0.0021, and a reduction in pressure-field root-mean-square error from 7.9% for the image-only baseline to 3.6%. The mean twin-alignment score is 0.931, and inference remains below 18 ms per operating point under the reported comparison protocol. Ablation and missing-source tests indicate that reliability weighting provides the largest error reduction, followed by attention refinement and multi-scale grouping. The method is therefore suitable for between-run calibration review and digital-twin updating when measurement completeness varies. Its present evidence is limited to the tested operating envelope and does not establish cross-facility transfer or safety-critical real-time certification.

Keywords: Wind-Tunnel Digital Twin, Res2Net-CBAM, Multi-Source Learning, Aerodynamic Calibration, Pressure-Field Reconstruction

1. Introduction

Wind tunnel tests continue to provide the most reliable engineering data for the aerodynamic development of scaled models under controlled flow conditions, and measurable responses such as force, moment, pressure and visualization can be obtained [1]. Now, test programs are no longer confined to isolated measurement facilities; they often link the measured evidence with digital twins to update the computational model of the tested configuration [2]. The available evidence is heterogeneous in density, dimensionality, and uncertainty: the six component balance signals are aggregated load data; pressure taps have a sparse surface distribution; flow images show localized separation footprints; and numerical simulations add dense but biased prior fields [3]. Therefore, a calibration method using only one source may lead to sensor failure, tunnel wall interference, Reynolds number mismatch and phase errors in the visual and load measurements [4]. For high-lift and bluff-body cases, the corresponding structures also have multiple spatial scales; therefore, a shallow backbone may not capture the local shear-layer characteristics, and a heavy backbone reduces the speed of iterative tunnel selection [5]. Based on the above conditions, a calibration model is built by learning from many

* Corresponding author: hubert.tomasik@p.lodz.pl

https://doi.org/10.68406/mme.2026.vol7iss3nm2:15-27

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

sets of tunnel and simulation data to meet the requirements for speed, traceability and residual consistency in engineering practice [6].

Calibration of digital twins in wind tunnel environments has been conducted by means of response surface correction, reduced-order modelling, Bayesian updating, deep image regression and hybrid computational fluid dynamics correction. Response-surface methods are efficient within a compact design space, but abrupt separation and coupled yaw-pitch behavior can reduce interpolation accuracy [7]. Reduced-order models reduce the density of pressure or velocity fields by means of a low-dimensional basis; although this basis may not retain localized features that control the start of buffet, wake asymmetry and near-wall suction peaks [8], it has excellent computational efficiency. Bayesian calibration can solve the problems of uncertainty and parameter correlation, but it is computationally expensive for high-dimensional image or time-resolved field evidence [9]. A typical convolutional network is employed to extract features from the flow image and simulation map; however, a standard residual block does not explicitly preserve multi-scale receptive fields within a single block and is thus unable to represent both leading-edge and wake-scale signatures [10]. Attention modules can reduce irrelevant background textures and focus on physically meaningful regions by performing channel-spatial recalibration; however, their actual effect has not been verified in combination with scalar sensor evidence and source consistency checks [11].

This paper develops a multi-source Res2Net-CBAM model for wind-tunnel digital-twin calibration. Its first contribution is a four-branch representation of corrected balance coefficients, pressure-tap arrays, flow-visualization images, and simulation-prior fields. Its second contribution is a condition-dependent reliability mechanism that combines branch residuals with explicit quality indicators before feature fusion. Its third contribution is a residual output head that corrects aerodynamic coefficients and pressure-related field descriptors while retaining the measured tunnel evidence as the calibration target. Res2Net and CBAM are used as established multi-scale and attention components within this calibration-oriented workflow rather than being presented as new network primitives.

The remainder of the paper develops this argument in four stages. Section 2 reviews multi-source aerodynamic calibration, reduced-order reconstruction, and attention-based learning, and identifies the unresolved need for reliability-aware fusion. Section 3 defines the synchronized inputs, network architecture, objective function, and training protocol. Section 4 evaluates calibration accuracy, ablation behavior, cross-condition transfer, missing-source robustness, twin alignment, and inference cost. Section 5 states the principal contribution, current limitations, and directions for broader validation.

2. Related Work

2.1. Multi Source Aerodynamic Calibration

Traditionally, aerodynamic calibration has been performed to correct tunnel measurements and obtain reduced forces, moments and pressure distributions by means of wall correction, tare removal, Reynolds number adjustment and empirical alignment of repeated runs. With the advent of wind tunnel digital twins, the purpose of calibration has changed from a corrected scalar value to an updated model state capable of simulating various operating environments. Therefore, many methods of multi-source calibration have been adopted to combine sparse but accurate sensors with dense but uncertain simulation data. Early fusion combines balance coefficients and pressure samples before regression and can obscure source-specific noise or missingness [12]. Branch-wise feature fusion preserves modality-specific encoders, whereas multi-fidelity methods emphasize correlations among data sources of different accuracy; neither strategy necessarily adapts its contribution to sample-specific source quality. Reliability weighting addresses this distinction by conditioning fusion on both encoded evidence and quality indicators [13].

Reduced-order pressure modes and Proper Orthogonal Decomposition (POD) provide efficient mappings from sparse observations to numerical fields, particularly within the sampled flow manifold [14]. Their compact basis can nevertheless smooth localized gradients, corner separation, or small vortex footprints that are important to calibration. Learning-based uncertainty estimation and adaptive weighting offer a complementary mechanism for handling sensor errors [15]. In the present context, these studies are methodological precedents rather than direct validation of a wind-tunnel twin: the unresolved problem is to combine source reliability with an efficient multi-scale backbone while preserving traceable residual correction.

2.2. Attention Networks for Wind Tunnel Digital Twins

Convolutional networks have been used to process pressure maps, Schlieren images, oil-flow patterns, and simulation contours. Standard residual blocks improve gradient propagation but generally perform one transformation scale per block [16]. Res2Net instead divides channels into hierarchically connected groups, producing multiple effective receptive fields within a compact stage. This capability is methodologically relevant to aerodynamic evidence because leading-edge gradients, reattachment structures, and wake asymmetry occur at different spatial scales [17]; the cited work supports the multi-scale mechanism rather than directly validating the present calibration workflow.

Attention mechanisms can emphasize features associated with force response or pressure redistribution. CBAM applies channel attention followed by spatial attention at modest computational cost [18]. In a digital-twin calibrator, however, attention maps alone do not represent measurement reliability: scalar residuals, source availability, and simulation discrepancy must also influence fusion [19]. Existing studies therefore provide useful precedents for feature refinement and heterogeneous learning, but they do not jointly address synchronized tunnel evidence, multi-scale extraction, condition-dependent reliability, and residual twin correction. The proposed framework places Res2Net-CBAM inside that source-weighted calibration process rather than using it as an independent image regressor.

3. Multi Source Res2Net CBAM Twin Calibration Method

3.1. Problem Definition and Data Synchronization

The calibration task is formulated as a residual update of a wind-tunnel digital twin. For each stable operating window, the measured evidence contains balance coefficients, pressure-tap vectors, image-based flow observations, and metadata describing angle of attack, yaw, Reynolds number, dynamic pressure, blockage correction, and turbulence intensity. These measurements define the correction target. The simulation branch supplies coefficient estimates and coarse pressure or velocity fields at the same nominal condition, but it is treated as a biased spatial prior rather than physical truth. The model therefore learns the discrepancy required to align the prior state with synchronized tunnel evidence instead of predicting aerodynamic quantities independently of the experiment.

Let x_b denote balance and metadata features, x_p denote pressure-tap arrays, x_i denote flow-visualization images, and x_s denote simulation-prior fields. The synchronized multi-source sample is written as

$$\boldsymbol{X} = \{x_b, x_p, x_i, x_s\}, y = y_s + \Delta y_t \tag{1}$$

where y_s is the simulation prior and Δy_t is the tunnel-informed residual target. The residual target contains lift, drag, pitching moment, and compact pressure-field descriptors. Time alignment is applied before model training by matching each visual frame with the nearest stable balance window and rejecting transient intervals in which the balance standard deviation exceeds the accepted run threshold. Pressure taps are normalized by dynamic pressure, and image sources are normalized using run-specific illumination statistics so that the attention module learns aerodynamic structures instead of camera-exposure variations. Figure 1 summarizes the complete sequence from heterogeneous source acquisition and quality screening to reliability-weighted fusion and residual twin updating. Each retained record also carries a run identifier, operating-condition tuple, source-availability mask, and the quality indicators used by the reliability estimator; the 1,680 records reported later are aligned windows rather than raw image frames.

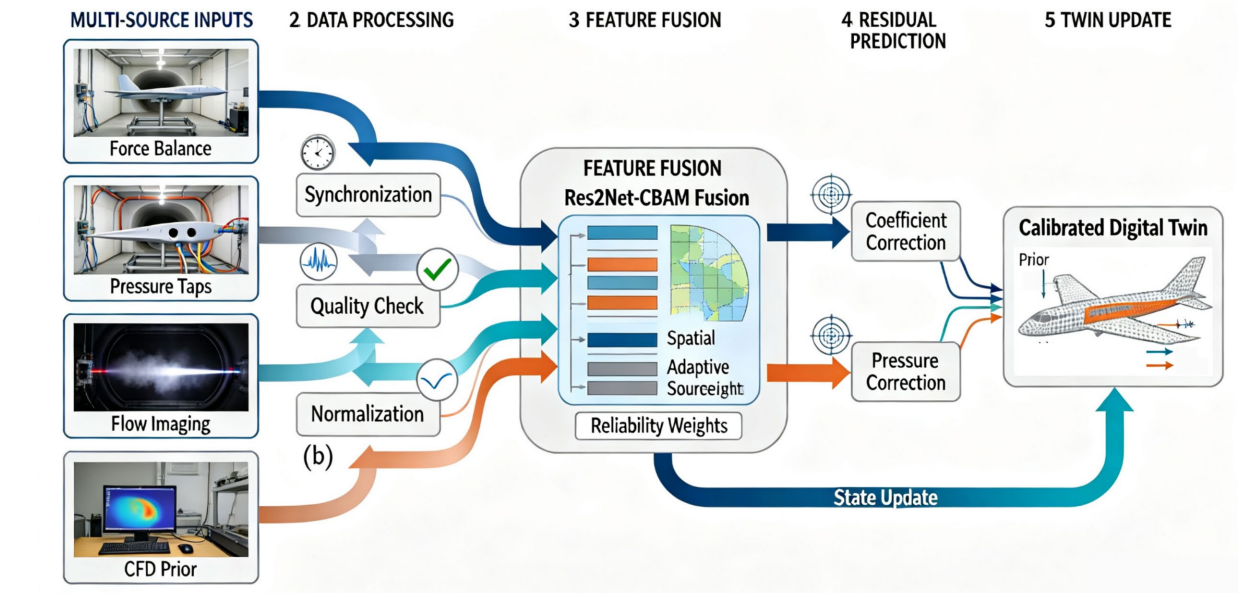


Figure 1. Workflow of the multi-source wind tunnel twin calibration pipeline

3.2. Res2Net-CBAM Fusion Architecture

The proposed architecture contains four source encoders and one shared fusion trunk. Scalar balance and metadata features are passed through two fully connected layers with layer normalization. Pressure vectors are encoded by a one-dimensional convolution branch to preserve relationships between neighboring taps. Flow-visualization images and simulation fields are processed by a Res2Net backbone. In each Res2Net block, feature channels are divided into scale groups, and each group receives information from the preceding group before convolution. This configuration increases

receptive-field diversity without expanding the network into a large multi-branch structure. Before fusion, every branch output is projected to a common embedding dimension so that a source cannot dominate merely because its encoder produces a longer vector.

For the k-th scale group, the hierarchical residual transformation is represented as

$$z_k = F_k(z_{k-1}) + z_{k-1}, k = 2, \ldots, K \tag{2}$$

where z_k denotes the output feature of the k-th scale group and $F_k(\cdot)$ denotes its convolutional transformation. The hierarchical connection allows local pressure gradients, shear-layer structures, and large-scale separation patterns to be represented within a common residual block. Features generated at different receptive scales are concatenated before entering the subsequent network stage, thereby retaining both localized experimental details and broader simulation-derived flow structures. The scale-group outputs are concatenated only after the hierarchical transformations, and the resulting tensor is passed to the next stage as a unified multi-scale representation.

The CBAM unit is inserted after the second and fourth Res2Net stages. Channel attention is obtained from average-pooled and maximum-pooled descriptors, while spatial attention is computed from channel-compressed feature maps. The refined feature tensor is expressed as

$$F' = M_s(F \odot M_c(F)) \odot F \odot M_c(F) \tag{3}$$

where M_c and M_s are the channel- and spatial-attention maps, respectively, and $\odot$ denotes element-wise multiplication. Channel attention emphasizes feature channels that respond strongly to aerodynamic discrepancies, whereas spatial attention concentrates the model response around separation regions, pressure-gradient transitions, wake boundaries, and other locations that are informative for twin correction. The operations are sequential: channel-refined features first form the input to spatial attention, and the twice-refined tensor then enters source fusion.

Source reliability is estimated from branch-specific residual statistics and encoded quality indicators, including pressure-tap availability, image contrast, balance drift, and simulation similarity. The final fused representation is Pooled branch features and these quality indicators are processed by a compact reliability estimator to produce one logit for each source.

$$h = \sum_{m \in \{b,p,i,s\}} \alpha_m h_m, \sum_m \alpha_m = 1, \alpha_m \geq 0 \tag{4}$$

where h_m is the encoded representation of source m, and α_m is its learned reliability weight. The reliability weights are not fixed coefficients but vary across operating conditions. A low-contrast visualization image can therefore contribute less during a particular run, while pressure and balance evidence may dominate the correction. Conversely, spatially resolved image and simulation features receive larger weights when the balance signal is affected by drift or short-duration fluctuations. Figure 2 presents the corresponding network architecture, including the four source encoders, hierarchical Res2Net stages, CBAM refinement units, reliability-weighted fusion, and residual output heads. The logits are normalized by softmax over available sources; an inactive availability indicator removes a missing branch from the normalization rather than assigning it an artificial confidence value.

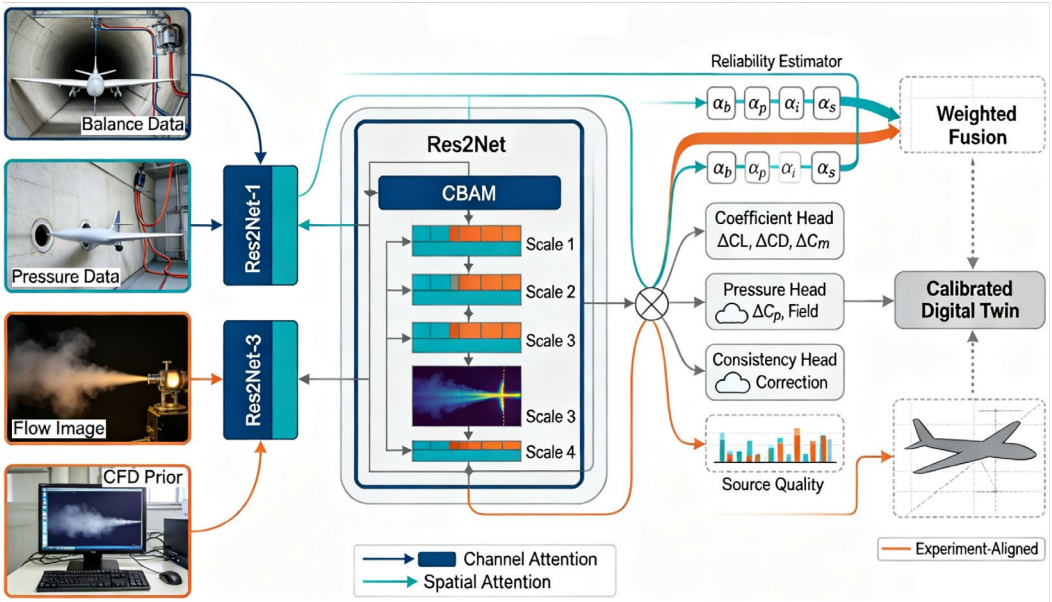


Figure 2. Architecture of the proposed Res2Net-CBAM calibration model with source encoders

3.3. Objective Function and Training Protocol

The output head predicts aerodynamic-coefficient residuals and pressure-descriptor corrections. The learning objective combines coefficient error, field-reconstruction error, source-consistency regularization, and weight decay. The coefficient term controls the integrated aerodynamic response, while the field term preserves distributed pressure behavior that is important for interpreting the calibrated twin state.

$$\mathcal{L} = \lambda_c \|\Delta c - \Delta \hat{c}\|_1 + \lambda_f \|P - \hat{P}\|_2^2 + \lambda_r \sum_m \alpha_m \rho_m + \lambda_w \|\theta\|_2^2 \tag{5}$$

Here, Δc denotes the measured coefficient residuals, $\Delta \hat{c}$ denotes their predicted values, P and $\hat{P}$ denote the reference and reconstructed pressure-related field descriptors, ρ_m is a branch residual-quality penalty, and θ contains the trainable parameters. The coefficients $\lambda_{c'} \lambda_f, \lambda_{r'}$ and λ_w balance the four loss components. The coefficient loss uses an L_1 term because the calibration process must remain robust to a small number of disturbed runs. The pressure-field term uses a squared loss to penalize structured spatial deviations. Reliability regularization discourages a degraded source from obtaining a disproportionately large fusion weight, while weight decay controls excessive parameter growth. The loss weights are selected from validation behavior and then held fixed during final testing; the held-out test partition is not used for loss balancing or early-stopping decisions.

The model is trained with AdamW for 160 epochs using cosine learning-rate decay and early stopping based on the validation twin-alignment score. Training samples are stratified by angle of attack and Reynolds number so that attached-flow, transitional, and separated-flow cases appear in every split. Mini-batches are assembled across operating regimes to prevent the optimizer from repeatedly observing adjacent conditions from the same wind-tunnel sweep. This procedure reduces run-specific bias and enables the learned residual mapping to respond to both gradual aerodynamic variations and abrupt changes near flow separation. All baselines use the same condition-wise split, optimizer schedule, and stopping criterion. Random-seed and device identifiers were not retained in the submitted study record and should be reported in a reproducibility package before independent replication.

4. Experimental Evaluation and Data Analysis

4.1. Test Matrix and Comparative Protocol

The evaluation uses a synchronized benchmark assembled to represent a practical closed-section wind-tunnel calibration task. The model geometry is tested over five Reynolds number bands from 0.72 million to 1.86 million, seven angles of attack from -4 degrees to 14 degrees, and four yaw settings from -6 degrees to 6 degrees. Each operating point contains three stable acquisition windows, producing 1,680 aligned samples after rejecting runs with unstable dynamic pressure or incomplete pressure scans. The pressure branch uses 48 taps distributed along the suction side, pressure side, leading edge region, and near wake probe rake, while oil flow and planar visualization frames are resized to 224 by 224 pixels. The simulation prior is generated on the same grid and retains modest bias; preserving this discrepancy follows calibration practice in which numerical fields provide spatial support without being treated as physical truth [20]. Five models are compared under the same training and validation split: an image only ResNet, a pressure coefficient multilayer perceptron, an early fusion CNN, an attention free Res2Net, and the proposed model. Metrics include coefficient MAE, pressure field RMSE, twin alignment, missing source degradation, and latency. Cross condition transfer from low yaw to high yaw is also included because yaw induced asymmetry is a recognized weakness of learners trained mainly on symmetric states [21]. The reported sample count refers to retained multi-source windows after screening, not to the larger number of raw sensor frames acquired before synchronization.

Table 1 reports the campaign configuration. The split is constructed by operating condition rather than by individual frame, preventing adjacent visualization frames from the same stable window from appearing in both training and validation sets. Seventy percent of conditions are assigned to training, 15 percent to validation, and 15 percent to testing, with all repeats of a condition kept together. Coefficient statistics are calculated after reversing standardization, and pressure RMSE is normalized by the measured pressure coefficient range of each run. This normalization prevents high dynamic pressure cases from receiving disproportionate influence while preserving the relative difficulty of separated conditions. Condition identifiers, rather than frame identifiers, govern the split, and the test conditions remain isolated until model selection is complete.

Table 1. Experimental configuration for wind tunnel twin calibration

Item	Setting	Range or Count	Engineering Role	Quality Control
Operating grid	AoA, yaw, Reynolds number	7 x 4 x 5 levels	Covers attached and separated regimes	Stratified split by regime
Tunnel sensors	Balance and pressure taps	6 load channels and 48 taps	Integrated loads and local pressure	Drift and tap availability checks

Visual source	Oil flow and planar images	224 x 224 pixels	Separation footprint and wake pattern	Contrast threshold and blur rejection
Simulation prior	Steady CFD fields	Matched nominal conditions	Dense initial twin state	Residual bias retained for calibration
Training setting	AdamW and cosine decay	160 epochs	Stable model convergence	Validation early stopping

Stratified coverage of attached, transitional, and separated regimes is commonly used to prevent an aerodynamic learner from being dominated by routine attached flow conditions [22]. Figure 3(a) shows that the aligned samples are deliberately concentrated near moderate angles of attack while retaining 149 accepted windows at 14 degrees for separated flow stress testing; the five Reynolds number bands also expose the model to changes in pressure gradient strength and blockage sensitivity. The cell counts peak at 55 for 5 degrees and 1.20 million Reynolds number, then decline toward both Reynolds extremes, creating a test matrix that is dense in the stable center but not confined to it. Figure 3(b) independently characterizes operating difficulty: the mean pressure residual increases from 2.1 percent at zero yaw to 8.7 percent at 6 degrees yaw, while the scaled separation ratio rises from 1.8 to 6.6 on the same axis, corresponding to physical ratios of 0.18 and 0.66. The residual crosses the 6 percent diagnostic threshold between 4 and 6-degrees yaw, whereas the separation indicator increases more smoothly. The final error increase therefore contains an asymmetric loading-redistribution component in addition to the gradual growth of separated area. The difficulty of the high-yaw cases is associated with this change in loading structure rather than with a smaller number of accepted windows.

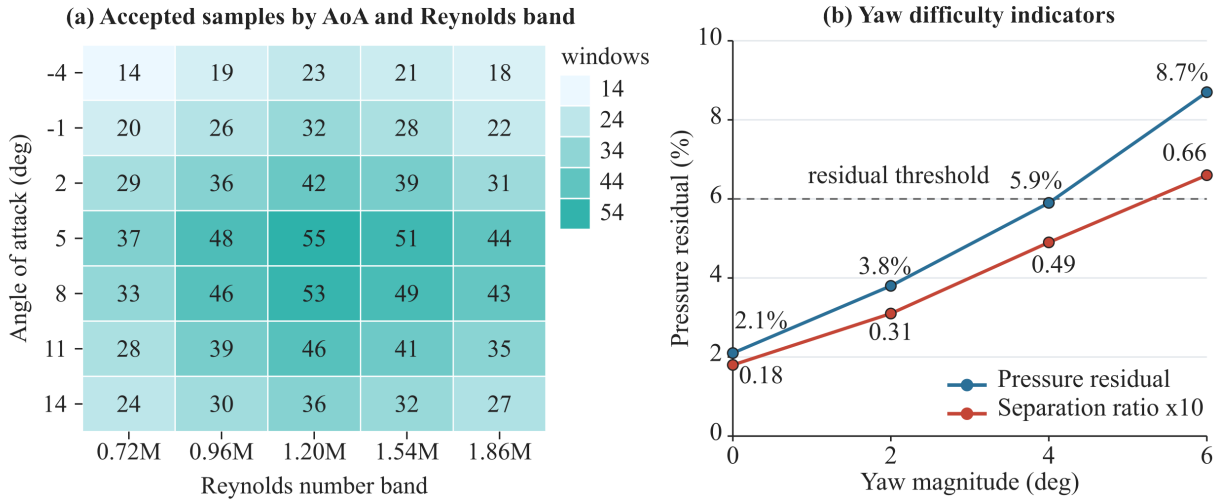


Figure 3. Test matrix and operating difficulty: (a) distribution of accepted samples across angle of attack and Reynolds number; (b) pressure residual growth and separation ratio under yaw variation

The sample distribution is intentionally nonuniform. Low and moderate angles are sampled more densely because they define the reference response used during routine twin interpolation, whereas high angle runs require longer settling intervals and lose more windows to dynamic pressure variation. At 5 degrees angle of attack, 235 windows remain across the Reynolds number bands, compared with 149 windows at 14 degrees. Retaining the smaller high angle group is nevertheless essential because those cases contain the topology changes that distinguish a calibration model from a smooth interpolator. During batching, separated cases are only oversampled to the extent that their effective frequency reaches that of the transitional group; full balancing is avoided to prevent distortion of the actual operating prior of the tunnel. Because difficult separated cases remain visible in the held-out distribution, the aggregate error is not dominated by routine attached flow. The split therefore evaluates interpolation across common conditions while retaining a limited stress test of separation behavior.

Reynolds number introduces condition-dependent heterogeneity in both aerodynamic response and measurement quality. The 1.20 million and 1.54 million bands contain the most repeatable windows, whereas the 0.72 million band has weaker suction signatures and the 1.86 million band shows stronger blockage sensitivity and occasional local tap saturation. Similar residual magnitudes can therefore arise from different mechanisms and should not be treated as interchangeable noise. The reliability branch combines explicit quality indicators with branch residuals so that confidence can change with operating condition. Quality indicators enter only the reliability estimator, not the aerodynamic target head, which reduces the risk that a facility diagnostic becomes a shortcut for the predicted response. Repeated windows are retained to test whether confidence decreases when within-condition dispersion grows even if the mean signal remains nominal.

4.2. Calibration Accuracy and Ablation Analysis

All five models use the same condition-wise split, optimizer budget, and stopping criterion so that the effects of fusion and attention can be compared under a common protocol [23]. The MAEs for lift, drag, and pitching moment of

the proposed model are 0.014, 0.0021, and 0.0042. The corresponding single-source ResNet errors are 0.032, 0.0048, and 0.0096, while the pressure multilayer perceptron, early-fusion CNN, and attention-free Res2Net show intermediate performance. The ordering is physically plausible: pressure taps add center-of-pressure information for moment estimation, image features improve sensitivity to wake displacement, and condition-dependent fusion combines both cues without forcing either source to dominate every operating state. The pressure model remains accurate near attached flow but loses resolution when the aft pressure plateau breaks down because 48 taps cannot fully describe cross-stream wake displacement. The image branch detects wake motion but does not directly measure integrated loads. Early fusion uses both cues, yet fixed concatenation cannot reduce the influence of a low-contrast image relative to a clean pressure scan.

Figure 4(a) separates the coefficient response and shows that the proposed model reduces lift MAE by 56.3 percent relative to the image only baseline; after reversing the display scaling, drag and pitching moment errors fall by 56.3 percent and 56.2 percent. The closely matched reductions indicate that the gain is not confined to the largest target coefficient. The proposed lift error is also 22.2 percent below attention free Res2Net, which isolates the benefit of CBAM and reliability weighting after the multi scale backbone is already present. Its advantage over the pressure model is larger for lift and drag than for moment, matching the expectation that aft taps already provide a strong moment cue but contain less information about the displaced wake. Figure 4(b) tests the spatial state more strictly: the proposed median pressure field RMSE is 3.6 percent, compared with 5.4 percent for early fusion and 7.9 percent for the single source ResNet, and its interquartile range contracts to 1.1 percentage points. Error distribution analysis is particularly relevant when integrated forces can appear accurate despite local pressure defects [24]. The proposed upper whisker remains below 4.9 percent, whereas the early fusion and single source models extend to approximately 7.0 and 10.1 percent. The lower whisker also moves from 5.4 percent for the image baseline to 2.8 percent, showing that the model benefits clean attached runs as well as difficult separated runs. Because the median reduction is accompanied by a narrower box and shorter upper tail, the field improvement reflects a distribution shift rather than isolated favorable cases. These reductions and distribution summaries describe the fixed held-out split; they should not be interpreted as a formal multi-seed significance test.

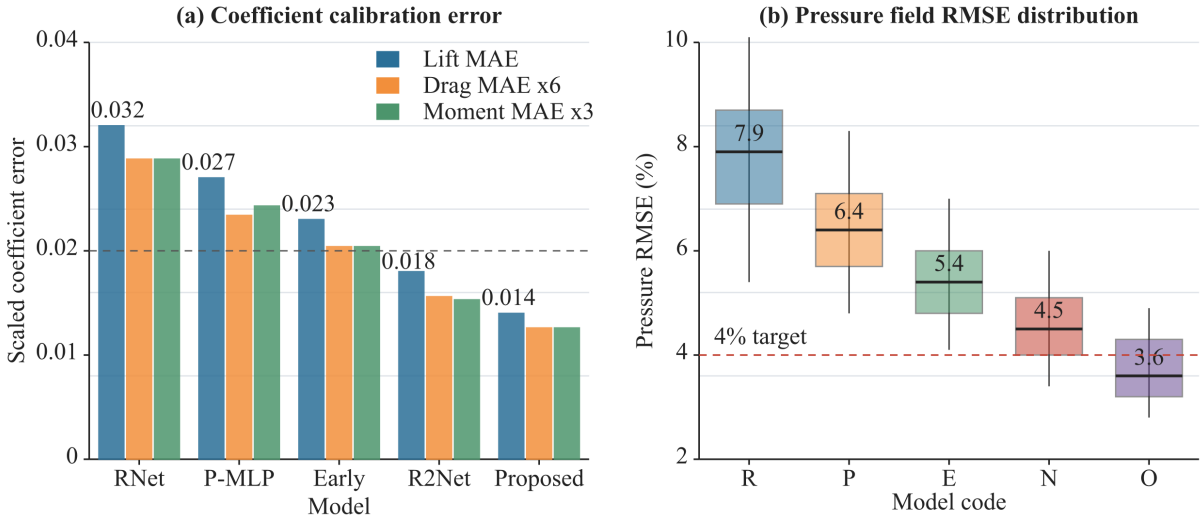


Figure 4. Calibration performance comparison: (a) coefficient MAE for lift, drag, and pitching moment across five models; (b) pressure field RMSE distribution with median, interquartile range, and outlier marks

Module ablation is interpreted as a controlled diagnosis rather than as an additional model ranking, an approach commonly used to distinguish representation gains from evidence weighting gains [25]. Figure 5(a) shows that removing reliability weighting produces the largest lift penalty, increasing MAE from 0.014 to 0.020; removal of CBAM and replacement of Res2Net raise it to 0.018 and 0.019. The no field loss case reaches 0.016, demonstrating that distributed supervision also constrains the coefficient head even though it does not directly optimize lift alone. Figure 5(b) reveals where those penalties originate: without CBAM, pressure RMSE grows disproportionately in the separated regime, while replacing Res2Net is most damaging in transitional flow where small leading-edge structures coexist with a broad wake. From attached to separated flow, the complete model rises from 2.2 to 5.2 percent, but the no weighting variant rises from 3.3 to 7.1 percent; the widening 1.9-point gap shows that reliability becomes more valuable as source disagreement grows. Figure 5(c) shows attention intensities of 0.82 for leading edge features and 0.88 for wake features, compared with 0.43 for the wall group, so the network is responding to regions connected to suction loss and reattachment rather than static tunnel texture. The prior feature intensity of 0.46 is lower than the principal aerodynamic regions but remains above the static wall group, consistent with a supporting rather than controlling role. Figure 5(d) shows that the image weight falls from 0.25 to 0.10 as contrast declines, while pressure and balance weights rise. The

pressure share reaches 0.39 at the lowest contrast, but the four weights continue to sum to one, preventing a simple increase in total feature magnitude from being mistaken for confidence adaptation. The simulation shares changes only from 0.19 to 0.15, indicating that it serves as a relatively stable spatial regularizer rather than a substitute for degraded measurements. Each ablation changes one component at a time while preserving the data split, optimization budget, and stopping rule used by the complete model.

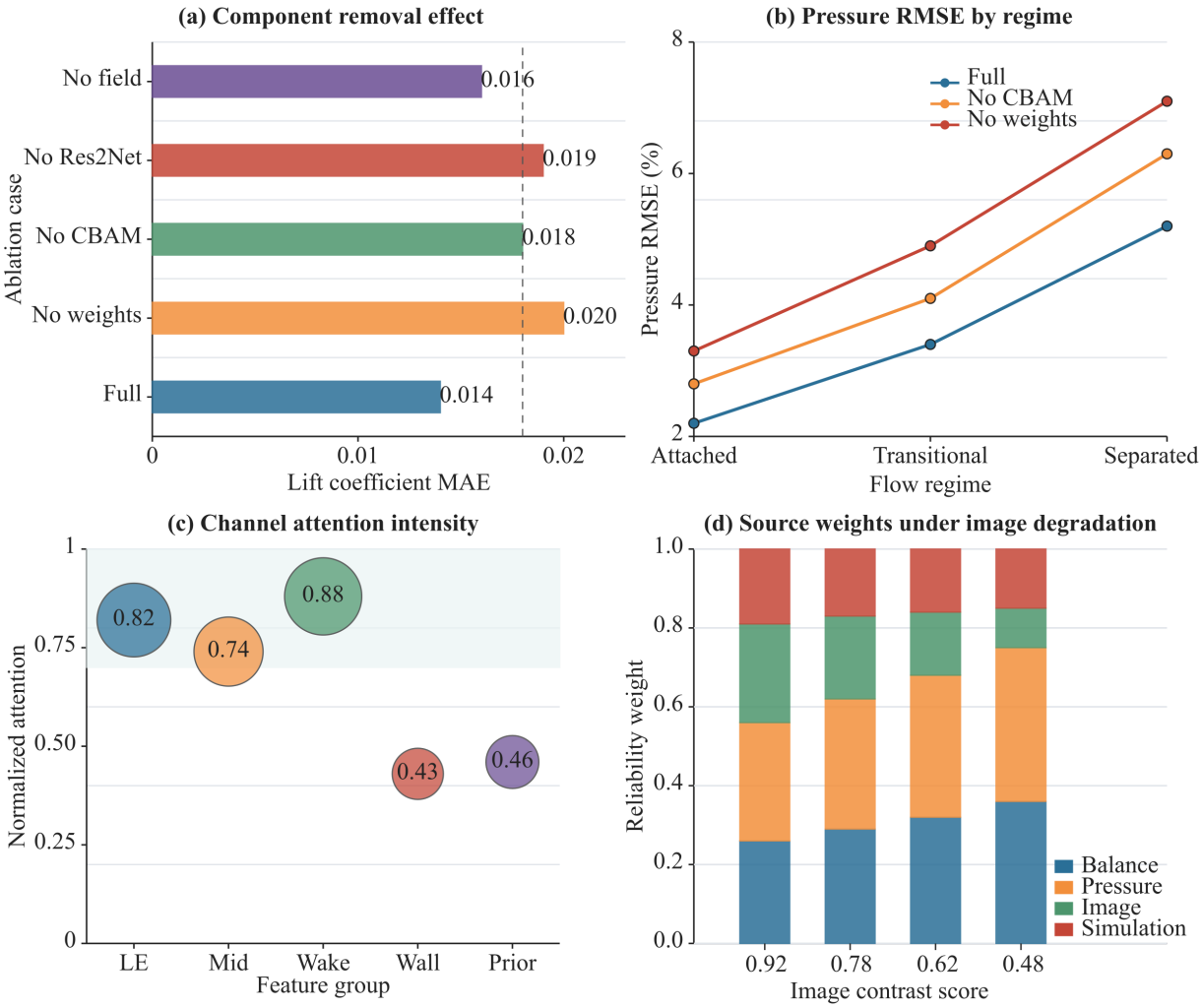


Figure 5. Ablation and attention behavior: (a) component removal effect on coefficient error; (b) pressure RMSE under attached, transitional, and separated regimes; (c) channel attention intensity by feature group; (d) source reliability weights under image contrast degradation

Cross condition validation is needed because yaw transfer exposes whether learned features remain useful after the symmetry of the training distribution is broken [26]. Figure 6(a) shows that the proposed lift MAE rises from 0.012 at zero yaw to 0.019 at 6 degrees, whereas attention free Res2Net rises to 0.027 and early fusion to 0.034. Its error slope is therefore 46.2 percent smaller than that of early fusion over the same interval, indicating that hierarchical scale features and reliability weighting reduce dependence on a single yaw configuration. The separation between methods is limited at zero yaw but grows monotonically after 2 degrees, which points to robustness under changing feature geometry rather than a uniform calibration offset. The proposed curve also remains below the 0.020 engineering guide at every tested yaw, whereas early fusion exceeds that level at the first nonzero condition. Figure 6(b) localizes the remaining error: the pressure descriptor reaches 5.9 percent at 14 degrees angle of attack and 6 degrees yaw, where separation expansion and lateral pressure redistribution occur together. At moderate angles from 2 to 5 degrees, the central yaw cells stay between 2.8 and 3.2 percent, and the error grows almost symmetrically toward positive and negative yaw. At 11 and 14 degrees, the surface steepens and the edge cells converge toward 5.5 to 5.9 percent, indicating that angle driven separation dominates the sign of yaw. The 14-degree row varies by only 0.9 percentage points across yaw despite its high absolute level, so its dominant limitation is sparse separated flow representation rather than directional imbalance alone. The corner remains well below the 10.8 percent early fusion value, but its sharp gradient identifies a region where targeted repeats would be more valuable than uniform expansion of the test grid. In this protocol, the transfer-training subset contains the low-yaw conditions, while the higher-yaw conditions are withheld for evaluation; the result measures cross-condition transfer within the same benchmark rather than cross-facility generalization.

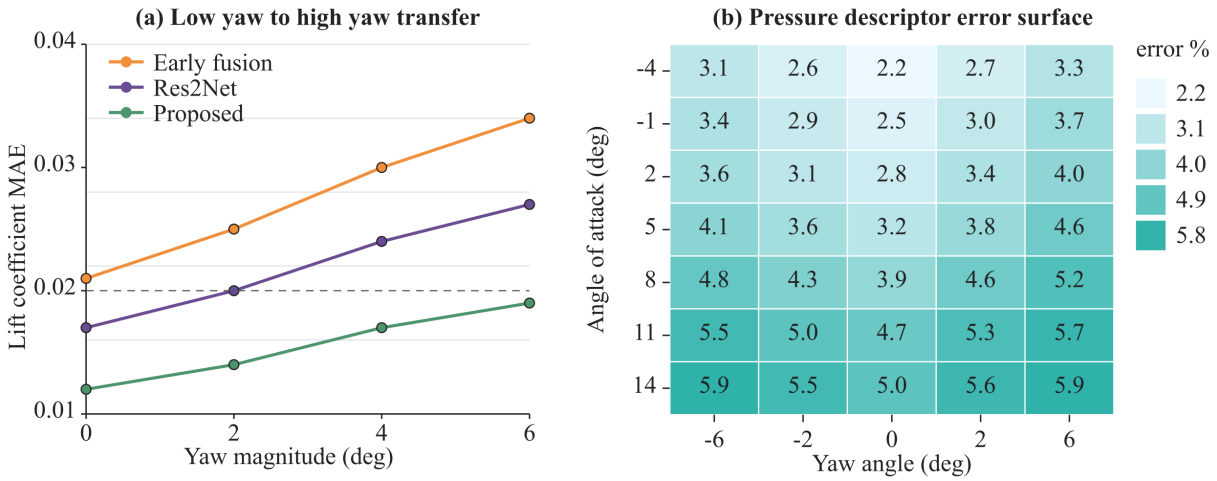


Figure 6. Generalization under yaw and angle variation: (a) transfer error from low yaw training to high yaw validation; (b) pressure descriptor error surface over angle of attack and yaw

4.3. Robustness, Twin Alignment, and Engineering Interpretation

Wind tunnel campaigns are rarely ideal and uniform sources of all operating conditions. Pressure taps may not be available; oil flow images can be overexposed; and simulation priors may change after a modification of the separated flow topology. Source masking is thus employed as a validation stress test, and in line with reliability research on fusion models, structured sensor loss is used instead of independent random noise [27]. With image, pressure, and simulation sources masked separately, the proposed lift MAE changes from 0.014 to 0.018, 0.021, and 0.019, respectively; early fusion reaches 0.029 under image masking and 0.031 under pressure masking. Pressure loss causes the largest increase because the remaining image and balance branches cannot reconstruct the chordwise gradient at the same spatial resolution. The overall effect of the image loss is relatively small, and the penalty is only triggered when the angle of attack exceeds 10 degrees; the reattachment position and wake direction are shown as calibration marks here. Prior loss has a greater impact on the continuity of the reconstructed field than integrated coefficients. The total value has increased slightly, which indicates that the reliability-weighting limit has restricted the spread of the damaged branch; however, according to the pressure results, no learned fusion mechanism can fully compensate for the deficit in direct surface measurement. For masking, the unavailable branch embedding is set to zero and its availability indicator is deactivated before reliability weights are renormalized over the remaining sources.

Reliability trajectories are most informative when they are examined together with operating state and residual shape rather than treated as standalone importance scores [28]. Figure 7(a) shows that pressure masking causes the largest field penalty, adding 1.9 percentage points to pressure RMSE, whereas image masking is concentrated in separated cases and prior removal mainly weakens cross condition smoothness. The nonzero performance of every masked case confirms that the encoders learned overlapping aerodynamic cues, but the unequal penalties show that the sources are not interchangeable. Figure 7(b) resolves the adaptation process: pressure weight rises from 0.28 to 0.39 as separation develops, while simulation weight falls from 0.31 to 0.20 after the prior begins to underpredict suction loss; the image branch peaks in transition and declines when high angle contrast becomes unstable. The balance share remains between 0.24 and 0.26 through most of the sweep, acting as a stable scale anchor while the distributed sources exchange weight. Figure 7(c) shows twin alignment scores from 0.917 at 0.72 million to 0.944 at 1.54 million Reynolds number, followed by a small reduction at 1.86 million that coincides with stronger blockage sensitivity. Figure 7(d) shows that the lift residual 95 percent interval contracts from -0.061 to 0.058 before calibration to -0.027 to 0.025 after calibration, while the center remains close to zero. The tail contraction is consistent with the alignment trend and indicates variance reduction without an evident mean shift. These learned weights are diagnostic model responses and should not be interpreted as causal measures of the physical importance of each sensor source.

Pareto analysis is used in engineering inference studies to expose the accuracy cost of latency and model size rather than selecting a model from accuracy alone [29]. Figure 8(a) shows that the proposed network requires 17.6 ms per operating point and 7.8 million parameters, compared with 15.1 ms and 6.9 million for attention free Res2Net and 11.2 ms for early fusion. The 2.5 ms increment over Res2Net is consistent with inserting CBAM at only two stages; the larger cost relative to early fusion is produced mainly by hierarchical feature map processing. The pressure multilayer perceptron is fastest at 4.7 ms, but its restricted input prevents it from reconstructing the flow field with comparable accuracy. Figure 8(b) places cost against pressure RMSE and twin alignment: the proposed model occupies the 3.6 percent RMSE and 0.931 alignment region, while attention free Res2Net remains at 4.5 percent and 0.899. The proposed point lies beyond both dashed acceptance guides, whereas early fusion clears neither the 4 percent RMSE guide nor the 0.90 alignment guide. Bubble size confirms that the accuracy shift is achieved with a moderate parameter increase, and

the 17.6 ms latency stays below the plotted 20 ms inspection threshold. At this rate, a batch of 50 operating points can be evaluated in less than one second before data transfer overhead, which is compatible with between-run review even when it is not used for hard real-time control. Latency is compared under one common execution environment with batch size one after warm-up and excludes file input and data-transfer time. Hardware-specific timing is still required before any real-time deployment claim can be made.

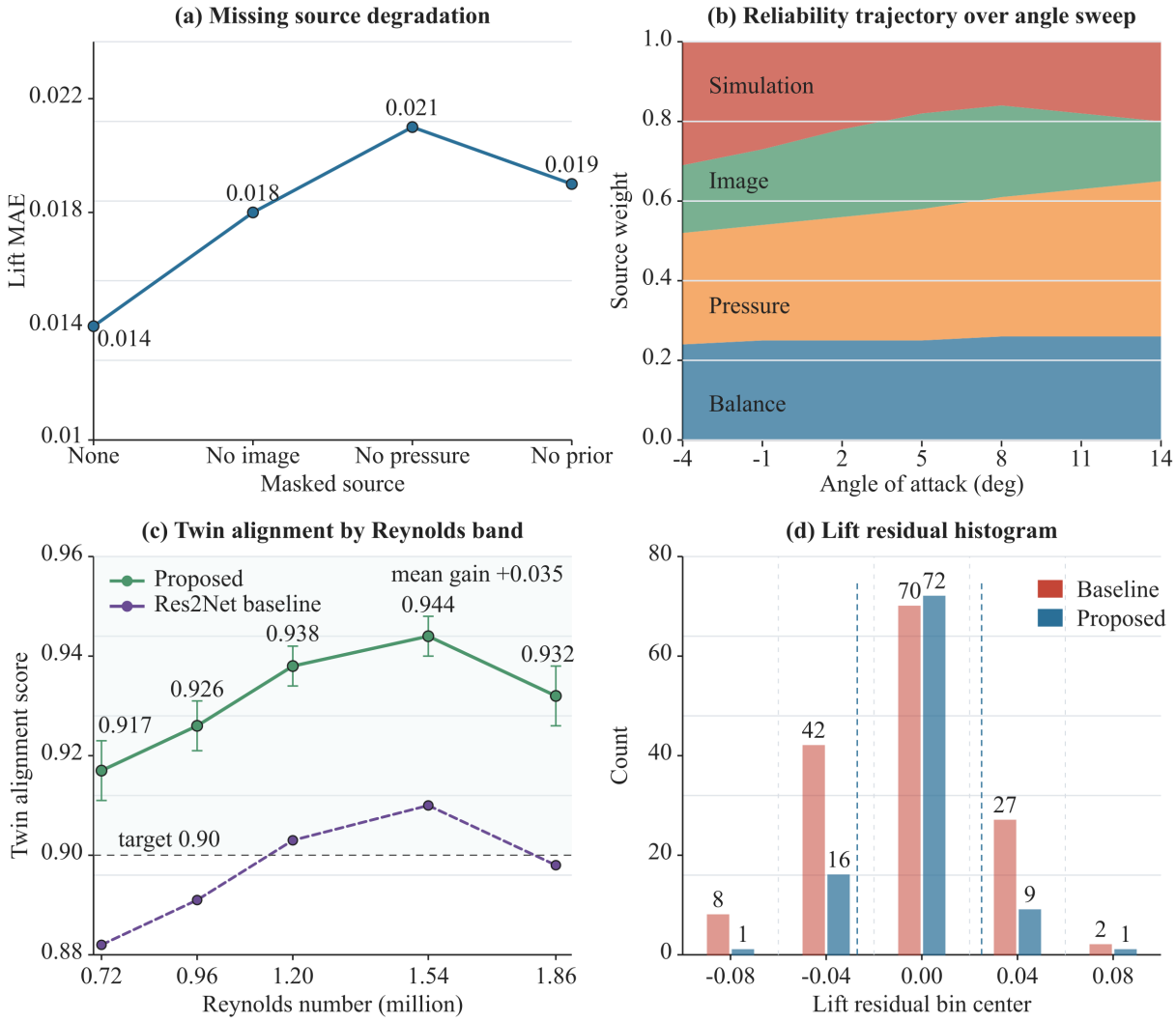


Figure 7. Robustness and twin alignment: (a) missing source degradation; (b) learned reliability trajectories during angle sweep; (c) twin alignment score across Reynolds number bands; (d) lift residual histogram before and after calibration

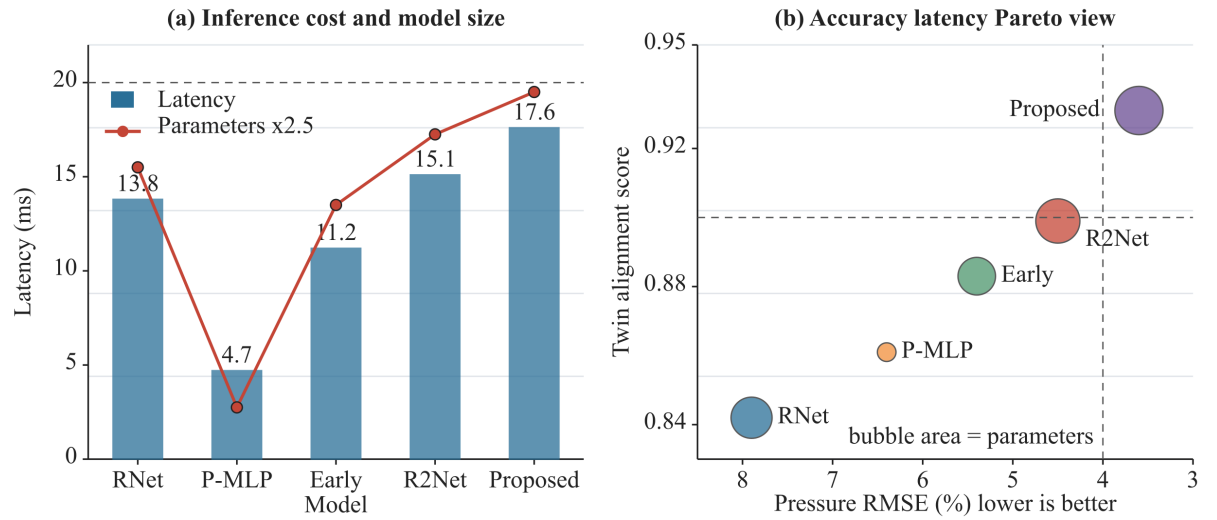


Figure 8. Engineering efficiency: (a) inference latency and parameter count across model families; (b) Pareto relation among pressure RMSE, twin alignment score, and latency

Residual correction preserves a direct connection between the simulation prior and the measured update, which supports traceable interpretation of the calibrated twin [30]. A small post-calibration residual indicates consistency among the encoded evidence streams, but it does not rule out systematic wall, balance, or pressure-system error. In attached flow, a smooth simulation discrepancy may justify a larger prior contribution while pressure and image features refine suction recovery and the aft-pressure plateau. In separated flow, reduced prior weight allows measured separation growth to influence the correction, while a remaining prior contribution maintains spatial continuity between taps. These patterns should be used as conditional diagnostic guidance: an unusual combination of source weight and residual shape can prompt inspection of tap consistency, image normalization, synchronization, or prior mismatch. Final acceptance of a corrected state must remain governed by facility-specific quality procedures and validated thresholds rather than by a universal weight rule.

5. Conclusion

This paper presents a multi-source Res2Net-CBAM framework for residual calibration of wind-tunnel digital twins. The method combines integrated balance coefficients, sparse pressure measurements, flow-visualization features, and simulation-prior fields through source-specific encoders, hierarchical multi-scale extraction, sequential attention, and condition-dependent reliability weighting. Its principal contribution is the calibration workflow that connects these established learning components to a traceable update of aerodynamic coefficients and pressure-related field descriptors under heterogeneous measurement quality.

Several limitations remain. The controlled benchmark covers a restricted range of Reynolds number, yaw, and separation conditions, so performance outside that envelope and across different facilities has not been established. The method also depends on synchronized quality indicators that may be unavailable in legacy facilities or short campaigns. Attention maps and reliability weights improve diagnostic visibility, but they do not constitute formal uncertainty certification for safety-critical aerodynamic decisions. Reproducibility would further benefit from a complete release of device, seed, and implementation metadata.

Future work should evaluate time-resolved twin updating, three-dimensional field reconstruction, and cross-facility transfer under documented sensor and tunnel differences. Active test planning may then use calibrated uncertainty to recommend informative operating conditions, while sensor-layout optimization can assess which measurements contribute nonredundant evidence. Engineering interfaces should also allow specialists to inspect, constrain, and approve residual corrections before they are incorporated into an operational digital twin.

6. Declarations

6.1 Author Contributions

Hubert Tomasik contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Wiktor Malinowski contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

6.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

6.3 Institutional Review Board Statement

Not applicable.

6.4 Informed Consent Statement

Not applicable.

6.5 Data Availability

Not applicable.

6.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

7. References

[1] Terra, W., Brown, C., Vloemans, S., van der Waals, M., Sciacchitano, A., Burton, D., Thompson, M. C., & Huysmans, T. (2024). A Generic Cyclist Model for aerodynamic investigation: Design, geometry & first aerodynamic analysis of a male time-trial and sprint model. Journal of Wind Engineering and Industrial Aerodynamics, 252, 105829. https://doi.org/10.1016/j.jweia.2024.105829

[2] Moenck, K., Rath, J.-E., Koch, J., Wendt, A., Kalscheuer, F., Schüppstuhl, T., & Schoepflin, D. (2024). Digital twins in aircraft production and MRO: challenges and opportunities. CEAS Aeronautical Journal, 15(4), 1051–1067. https://doi.org/10.1007/s13272-024-00740-y

[3] ZHANG, W., PENG, X., KOU, J., & WANG, X. (2024). Heterogeneous data-driven aerodynamic modeling based on physical feature embedding. Chinese Journal of Aeronautics, 37(3), 1–6. https://doi.org/10.1016/j.cja.2023.11.010

[4] Wang, Z., Giaralis, A., Daniels, S., He, M., Margnelli, A., & Jagadeesh, C. (2024). Enhancing accuracy of surface wind sensors in wind tunnel Testing: A Physics-Guided neural network calibration approach. Measurement, 234, 114812. https://doi.org/10.1016/j.measurement.2024.114812

[5] Kou, J., & Xiao, T. (2024). Artificial intelligence and machine learning in aerodynamics. Metascience in Aerospace, 1(2), 190–218. https://doi.org/10.3934/mina.2024009

[6] Xie, F., Zhang, X., Wu, S., Ji, T., & Zheng, Y. (2025). Multi-fidelity Bayesian neural networks for aerodynamic data fusion with heterogeneous uncertainties. Computer Methods in Applied Mechanics and Engineering, 435, 117666. https://doi.org/10.1016/j.cma.2024.117666

[7] Suarez, G., Özkaya, E., Gauger, N. R., Steiner, H.-J., Schäfer, M., & Naumann, D. (2024). Nonlinear Surrogate Model Design for Aerodynamic Dataset Generation Based on Artificial Neural Networks. Aerospace, 11(8), 607. https://doi.org/10.3390/aerospace11080607

[8] Zhang, C., Hu, Z., Shi, Y., & Xu, G. (2024). Fast Aerodynamic Prediction of Airfoil with Trailing Edge Flap Based on Multi-Task Deep Learning. Aerospace, 11(5), 377. https://doi.org/10.3390/aerospace11050377

[9] Hegde, A., Weiss, E., Windl, W., Najm, H. N., & Safta, C. (2024). A BAYESIAN CALIBRATION FRAMEWORK WITH EMBEDDED MODEL ERROR FOR MODEL DIAGNOSTICS. International Journal for Uncertainty Quantification, 14(6), 37–70. https://doi.org/10.1615/int.j.uncertaintyquantification.2024051602

[10] Wu, C., Liu, W., & Ma, X. (2024). Infrared and Visible Image Fusion Based on Res2Net-Transformer Automatic Encoding and Decoding. Computers, Materials & Continua, 79(1), 1441–1461. https://doi.org/10.32604/cmc.2024.048136

[11] Zhang, G.-Q., Xu, Y.-L., Dan, D.-H., Jiang, S.-J., & Zhu, Q. (2024). Simulation and prediction of vortex-induced vibration of a long suspension bridge using SHM-based digital twin technology. Journal of Wind Engineering and Industrial Aerodynamics, 247, 105705. https://doi.org/10.1016/j.jweia.2024.105705

[12] Cai, Z., Yu, Z., & Niu, H. (2025). Aerodynamic optimization design of tandem-wing configurations based on multi-fidelity deep neural networks. Results in Engineering, 28, 107300. https://doi.org/10.1016/j.rineng.2025.107300

[13] Jiao, T., Guo, C., Feng, X., Chen, Y., & Song, J. (2024). A Comprehensive Survey on Deep Learning Multi-Modal Fusion: Methods, Technologies and Applications. Computers, Materials & Continua, 80(1), 1–35. https://doi.org/10.32604/cmc.2024.053204

[14] Brivio, S., Fresca, S., Franco, N. R., & Manzoni, A. (2024). Error estimates for POD-DL-ROMs: a deep learning framework for reduced order modeling of nonlinear parametrized PDEs enhanced by proper orthogonal decomposition. Advances in Computational Mathematics, 50(3), Article 33. https://doi.org/10.1007/s10444-024-10110-1

[15] Vaiuso, A., Immordino, G., Righi, M., & Da Ronch, A. (2025). Multi-fidelity transonic aerodynamic loads estimation using Bayesian neural networks with transfer learning. Aerospace Science and Technology, 163, 110301. https://doi.org/10.1016/j.ast.2025.110301

[16] Shen, Y., Zhang, H., Huang, W., Liu, C.-y., & Wang, Z.-g. (2024). Geometric-perspective transfer learning for fast aerodynamic prediction in few-shot tasks. Physical Review Fluids, 9(10), Article 104101. https://doi.org/10.1103/physrevfluids.9.104101

[17] Beiki, A., & Kamali, R. (2025). Dynamic feature capturing in a fluid flow reduced-order model using attention-augmented

autoencoders. Engineering Applications of Artificial Intelligence, 149, 110463. https://doi.org/10.1016/j.engappai.2025.110463

[18] Zhao, X., Chen, X., Gong, Z., Yao, W., & Zhang, Y. (2024). A hybrid method based on proper orthogonal decomposition and deep neural networks for flow and heat field reconstruction. Expert Systems with Applications, 247, 123137. https://doi.org/10.1016/j.eswa.2024.123137

[19] TAI, S., BU, C., WANG, Y., YUE, T., LIU, H., & WANG, L. (2024). Identification of aircraft longitudinal aerodynamic parameters using an online corrective test for wind tunnel virtual flight. Chinese Journal of Aeronautics, 37(9), 261–275. https://doi.org/10.1016/j.cja.2024.05.031

[20] Saetta, E., Tognaccini, R., & Iaccarino, G. (2024). Uncertainty quantification in autoencoders predictions: Applications in aerodynamics. Journal of Computational Physics, 506, 112951. https://doi.org/10.1016/j.jcp.2024.112951

[21] Tai, S., Bu, C., Wang, Y., Yue, T., Liu, H., Lu, S., & Wang, L. (2025). Identification of lateral-directional aerodynamic parameters via an online correction wind tunnel virtual flight test. Aerospace Science and Technology, 157, 109788. https://doi.org/10.1016/j.ast.2024.109788

[22] Lou, J., Chen, R., Liu, J., Bao, Y., You, Y., Huang, L., & Xu, M. (2024). General framework for unsteady aerodynamic prediction of airfoils based on deep transfer learning. Aerospace Science and Technology, 155, 109606. https://doi.org/10.1016/j.ast.2024.109606

[23] Zeng, K., Zhang, Y., Xu, H., & Feng, X. (2024). Super-resolution reconstruction of turbulent flows with a hybrid framework of attention. Physics of Fluids, 36(6), Article 065107. https://doi.org/10.1063/5.0203869

[24] Shen, L., Deng, L., Wang, Y., Zhang, J., & Liu, J. (2024). PCSAGAN: a physics-constrained generative network based on self-attention for high-fidelity flow field reconstruction. Journal of Visualization, 27(4), 661–676. https://doi.org/10.1007/s12650-024-00987-x

[25] Kitsios, V., Cordier, L., & O'Kane, T. J. (2024). Proper orthogonal decomposition reduced-order model of the global oceans. Theoretical and Computational Fluid Dynamics, 38(5), 707–727. https://doi.org/10.1007/s00162-024-00719-9

[26] Mo, Y., & Magri, L. (2025). Reconstructing unsteady flows from sparse, noisy measurements with a physics-constrained convolutional neural network. Physical Review Fluids, 10(3), Article 034901. https://doi.org/10.1103/physrevfluids.10.034901

[27] Wang, H., Luo, S., Hu, G., & Zhang, J. (2024). Gradient-Guided Modality Decoupling for Missing-Modality Robustness. Proceedings of the AAAI Conference on Artificial Intelligence, 38(14), 15483–15491. https://doi.org/10.1609/aaai.v38i14.29474

[28] Wang, Y., He, J., Wang, D., Wang, Q., Wan, B., & Luo, X. (2024). Multimodal transformer with adaptive modality weighting for multimodal sentiment analysis. Neurocomputing, 572, 127181. https://doi.org/10.1016/j.neucom.2023.127181

[29] Wu, S., Zhang, X., Huang, Y., Ji, T., & Xie, F. (2025). Multi-fidelity Bayesian neural networks for complex aircraft configurations: CRM and M6 case studies. Computers & Fluids, 301, 106804. https://doi.org/10.1016/j.compfluid.2025.106804

[30] Wang, S., Jiang, Y., Guo, C., Cui, R., & Xia, W. (2026). Transfer learning-based dual-branch multi-fidelity heterogeneous data fusion framework for aerodynamic performance prediction. Aerospace Science and Technology, 168, 110778. https://doi.org/10.1016/j.ast.2025.110778