TOWARDS INTERPRETABLE DAMAGE DETECTION BASED ON AERODYNAMIC PRESSURE MEASUREMENTS

Philip Franz
German Aerospace Center (DLR)
Institute for the Protection of Terrestrial Infrastructures
Rathausallee 12, 53757 St. Augustin, Germany
philip.franz@dlr.de

Max von Danwitz
German Aerospace Center (DLR)
Institute for the Protection of Terrestrial Infrastructures
Rathausallee 12, 53757 St. Augustin, Germany
max.vondanwitz@dlr.de

Gregory Duthé
Institute for Structural Engineering
ETH Zürich
Stefano-Franscini-Platz 5, 8093 Zürich, Switzerland
duthe@ibk.baug.ethz.ch

Alexander Popp
Institute for Mathematics and Computer-Based Simulations (IMCS)
University of the Bundeswehr Munich
Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
alexander.popp@unibw.de

Eleni Chatzi
Institute for Structural Engineering
ETH Zürich
Stefano-Franscini-Platz 5, 8093 Zürich, Switzerland
chatzi@ibk.baug.ethz.ch

October 5, 2026

ABSTRACT

The increasing flexibility of modern large wind turbine blades necessitates cost-efficient and reliable structural monitoring solutions. For this purpose, we propose to use aerodynamic pressure measurements obtained via Aerosense, a novel, non-intrusive and economical sensing system. In former work [Franz et al., 2025], we investigated the potential of aerodynamic pressure measurements for structural damage detection on elastic and aerodynamically loaded structures. An experimental campaign was conducted on a NACA 633418 airfoil mounted on a vertically vibrating cantilever beam within an open wind tunnel. Structural damage was introduced progressively through controlled saw cuts near the beam support. Aerodynamic pressure distributions were recorded under varying inflow conditions and structural states. Based on this data set, we developed a convolutional neural network to detect structural damage and classify its severity using only aerodynamic pressure signals. The results demonstrate that pressure measurements can effectively enable real-time detection and quantification of damage in elastic, beam-like structures subjected to mildly turbulent flow and varying operational conditions. Recognizing the limitations of pure black-box classification, in this study, we further incorporate physics-based insights and explainable machine learning methods to interpret how structural damage influences both the dynamic response and the aerodynamic pressure field. This leads to an enhanced damage detection pipeline, aiming to improve transparency, robustness, and physical consistency in data-driven monitoring of elastic, aerodynamically loaded structures.

Keywords integrated gradients · aerodynamic pressure distribution · CNN · damage detection · structural health monitoring · explainable AI

1 Introduction

Deep neural networks, such as convolutional or transformer architectures, have achieved remarkable success in pattern recognition tasks for multivariate time series, like classification, prediction or anomaly detection [Ismail Fawaz et al., 2019, Mohammadi Foumani et al., 2024, Kong et al., 2025, Zamanzadeh Darban et al., 2024]. This capability makes them particularly attractive for domains that rely on extracting information from dense, multi-sensor measurements, such as structural health monitoring (SHM) [Wang et al., 2024, Cha et al., 2024]. However, relying on opaque black-box models for high-stakes decisions in SHM is risky, as deep neural networks typically lack interpretability. Furthermore, it remains unclear whether these models generalize reliably to real-world environments where operating conditions differ from the training distribution. Compounding this, inferring states from sparse and noisy measurements can be ill-posed, admitting a distribution of plausible solutions rather than a unique answer [Vadeboncoeur et al., 2026]. To address the lack of generalization and interpretability, the field of scientific machine learning (SciML) [Baker et al., 2019, Quarteroni et al., 2025] or physics-informed machine learning has emerged [Karniadakis et al., 2021]. SciML integrates physical laws or domain-specific knowledge directly into the learning scheme, aiming to ensure adherence to the underlying physics, generalization beyond the training domain and to enhance the interpretability of the learned representations [Karniadakis et al., 2021, Cicirello, 2024]. However, SciML requires robust approaches and detailed physical insights and governing equations, which may not always be available for real-world systems. In such cases, explainable artificial intelligence (XAI) offers an alternative for increasing the transparency of black-box models. In our recent study [Franz et al., 2025], we suggested a convolutional neural network (CNN)-based approach to detect and rank structural damage based on aerodynamic pressure measurements (APM) over the section of an airfoil, acquired with the Aerosense system [Barber et al., 2022]. Despite its accuracy, the CNN's black-box nature limits its practical utility and prevents a thorough understanding of the underlying mechanism. Therefore, we employ an XAI method to enhance the interpretability of our approach and to identify damage-sensitive features within the APM. XAI methods can generally be categorized as ante-hoc or post-hoc, global or local, and model-specific or model-agnostic [Theissler et al., 2022]. For this study, we select Integrated Gradients (IG) [Sundararajan et al., 2017], a model-agnostic, post-hoc attribution method [Theissler et al., 2022], which is suited for multivariate time series and provides attributions across all channels. While IG is a local method providing only sample-specific explanations, subsequence or instance-based approaches such as XCM [Fauvel et al., 2021], TSProto [Bobek and Nalepa, 2025], or ProtoTSNet [Małkus et al., 2025], can offer global, data set-wide explanations via class-specific subsequences or prototypes. However, these methods require training new models, whereas our objective is to interpret the existing CNNs presented in [Franz et al., 2025]. The paper is organized as follows: Section 2 details the experimental setup, the previously trained CNNs and effects of damage on the aerodynamic pressure distribution (APD). Section 3 presents the IG method and baselines, analyzing the resulting attribution maps. Section 4 describes the evaluation and retraining of the CNNs based on insights from the attribution analysis, while Section 5 provides a physics-grounded discussion of the results. We conclude with a summary of the main findings in Section 6.

2 Experimental Setup & Effects Of Damage

2.1 Experimental setup

The experimental setup (Figure 1) consists of a NACA 633418 airfoil mounted on a flexible aluminum cantilever beam within the open test section of the EPFL Unsteady Flow Diagnostics Laboratory wind tunnel. The wind tunnel test section measures 40 cm×40 cm and operates up to 35 m/s, while the beam itself features a rectangular cross-section (1 cm × 4 cm). Artificial damage is introduced by transversely sawing the beam in z-direction near the support to reduce local stiffness; the severity is quantified by the cut length across five levels (0%, 12.5%, 25%, 37.5%, and 50% of the beam width). An additional damage scenario is considered by attaching a mass to the undamaged beam between the airfoil and the motor (see Figure 1). Periodic excitation is applied at the beam's tip via a motor-driven eccentric mass, primarily inducing flap-wise bending, although torsional motion also occurs. The Aerosense system [Barber et al., 2022] records the aerodynamic pressure distribution at 40 locations on the airfoil surface (only 37 were functioning correctly, see Figure 2) at the airfoil's mid-span with a sampling rate of 100 Hz. Ultimately, this cantilever setup serves as a proxy for a rotating wind turbine blade oscillating flap-wise due to gravitational loading. Table 1 summarizes the eight test series (TS) generated by combining the two angles of attack (AoAs), wind speeds V, and heaving frequencies f_h. Each TS comprises six structural states resp. damage classes: five crack lengths to simulate stiffness reduction and one added mass state (see Table 5). We conduct three 150 s runs per configuration: the first 15 s feature aerodynamic loading only, while the remaining 135 s involve simultaneous loading by aerodynamic and periodic forces. More details on the experimental setup, the data acquisition system and data normalization, can be found in Section 2 of [Franz et al., 2025].

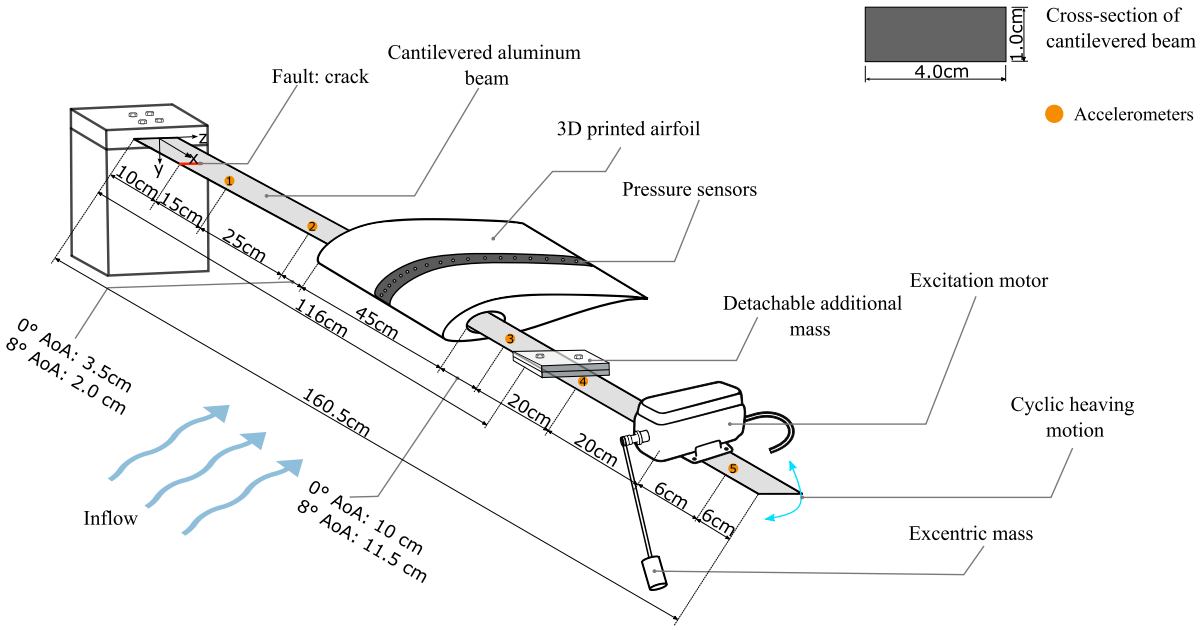


Figure 1: The experimental setup features an airfoil on a cantilever beam with an excitation motor and removable mass, placed in front of the outlet of the wind tunnel. The Aerosense sensor node records aerodynamic pressure on the airfoil, and accelerometers measure beam vibration. Beam dimensions are indicated in the upper right. Figure taken from [Franz et al., 2025].

2.2 Previous results with black-box approach

The damage detection and severity rating algorithm proposed in [Franz et al., 2025], follows the CNN architecture (see Figure 3) suggested by Wang et al. [2017], comprising three convolutional layers with batch normalization, a global average pooling layer, and a fully connected layer. To prepare the data, the first 40 s and last 10 s of each measurement are discarded to mitigate transient effects. From the remaining 100 s, 89 overlapping 1.5 s windows are extracted to capture approximately three vibration cycles per sample and minimize the influence of signal trends. Finally, each sample is z-score normalized and split into training, validation, and test sets, resulting in the algorithm shown in Figure 3. In [Franz et al., 2025], we separated the data for the classification by AoA, as the pressure distribution depends strongly on this variable. With six structural states across two frequencies and two wind speeds (see Table 1), each AoA data set comprises 72 measurements. We employ a non-random split strategy: two of the three measurements for each boundary condition are consistently assigned to training, while the third is reserved for testing to ensure evaluation on unseen data. This results in 48 experiments for training and 24 for testing. With the segmentation approach described above, this yields 4,272 training samples and 2,136 test samples. The training set is further divided into a 25% validation set (1068 samples) and a 75% training set (3204 samples). An overview of the classification results on the test set of each split is given in Table 2. The classification accuracy per split is the average accuracy over all balanced classes. Further details on the CNN approach, the split of the data set and the classification results are available in Section 5 of [Franz et al., 2025].

Table 1: Boundary conditions of the eight test series.

Test series	1	2	3	4	5	6	7	8
AoA [°]	0	0	0	0	8	8	8	8
approximate f_h [Hz]	1.0	1.0	1.9	1.9	1.0	1.0	1.9	1.9
v [m/s]	12	24	12	24	12	24	12	24

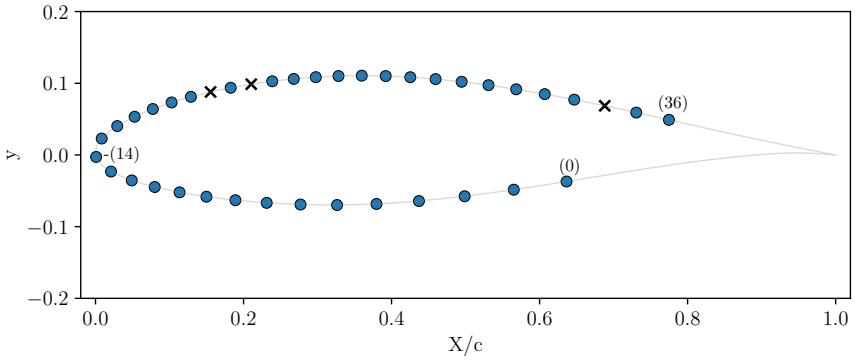


Figure 2: Distribution of barometers on the airfoil surface. The blue, circular markers indicate working sensors; the black crosses indicate malfunctioning sensors. The numbers of the sensors at the trailing edge and at the leading edge are indicated by the numbers in the brackets.

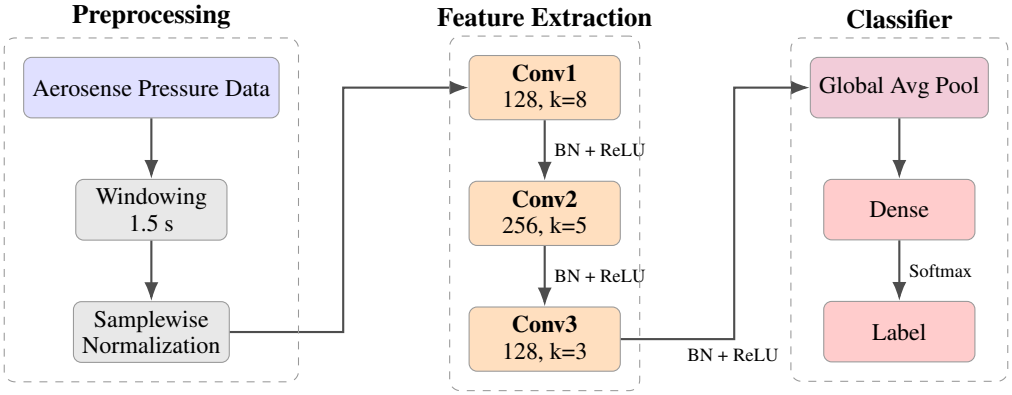


Figure 3: Preprocessing pipeline and CNN architecture for pressure signal classification.

2.3 Effects of damage on structural behavior and aerodynamic pressure field

The airfoil section is approximated as a spring-mass-damper system (Figure 4) with a vertical y and a rotational ϕ degree of freedom. The system is characterized by mass M, stiffness and damping coefficients K_y,D_y,K_ϕ and D_ϕ, inflow velocities V,V_{rel}, AoA α, and the aerodynamic lift and drag force, F_L and F_D. Stepwise cutting of the beam in the z-direction introduces artificial damage, influencing the dynamics as follows: First, the equilibrium position shifts downward. The cut reduces the local cross-sectional area and the flexural moment of inertia, resulting in a downward y-displacement of the cantilever beam that increases along the longitudinal axis. Additionally, the cut also alters the center of shear. Due to the eccentricity of the tip mass and airfoil, the beam exhibits a small initial twist; this torsional rotation ϕ increases with the cut length as the torsional moment of inertia is reduced at the site of damage. Secondly, as the crack length increases and the loading is not altered, the amplitudes of both flexural and torsional vibrations increase. Furthermore, these oscillations occur around an equilibrium position that becomes progressively more deflected and twisted. As discussed in Franz et al. [2025], the torsional vibrations are expected to be significantly smaller than the vertical vibrations; however, both magnitudes increase as the damage progresses. Regarding the impact of damage on the APD, the increased twist of the equilibrium state and larger torsional and vertical vibration amplitudes are expected to significantly influence the pressure distribution, whereas the increased inclination of the airfoil in the y-direction is unlikely to have a substantial effect. Specifically, airfoil twist and torsional vibrations alter the AoA. The twist induces a static offset relative to the undamaged state, while torsional vibrations increase AoA fluctuations. Additionally, vertical oscillations induce a time-varying velocity component $\dot{y}$ of the airfoil, resulting in corresponding variations in the AoA and APD. Moreover, torsional and flexural vibrations may enhance vortex shedding at the airfoil's trailing edge, further altering the APD. To estimate the vortex shedding frequency f_{vs}, we employ the Strouhal number $St = \frac{f_{vs}D}{V}$ [Modarres-Sadeghi, 2021]. Given the Reynolds regime of our experiments (see Section 3.3 of [Franz et al., 2025]), we assume $St \approx 0.2$ following Modarres-Sadeghi [2021], although an airfoil is not truly a bluff body. Using the airfoil chord length for D following Chang et al. [2022], the estimated shedding

Table 2: Test set accuracy for all splits for 0° AoA and 8° AoA. The classification accuracy given for a single split is the average of the classification accuracy over the six balanced classes of that split.

AoA	Split 1	Split 2	Split 3	Average
0°	82.6 %	98.5 %	93.7 %	91.6 %
8°	81.8 %	96.0 %	89.8 %	89.2 %

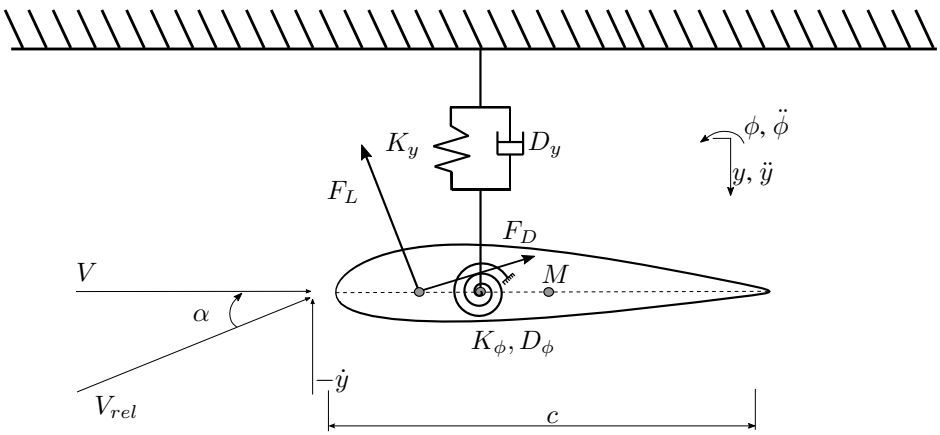


Figure 4: Schematic blade section with a flap-wise degree of freedom y and rotational degree of freedom ϕ. M denotes the system's mass, K_y and D_y denote stiffness and damping in flap-wise direction and K_ϕ and D_ϕ denote the rotational stiffness and damping. V describes the inflow velocity, V_{rel} describes the relative inflow velocity and α the AoA . The airfoil is subjected to lift and drag forces, denoted by F_L and F_D.

frequencies are $f_{vs} \approx \frac{St \times V}{D} = \frac{0.2 \times 12 \frac{m}{s}}{0.16m} = 15$ Hz for $V = 12$ m/s and $f_{vs} \approx 30$ Hz for $V = 24$ m/s. Analysis of signal spectra from TS 8, the highest loading case, with a 50% crack length revealed no significant energy bands or regular pulses at the predicted 15 Hz or 30 Hz frequencies (Figure 11). The spectra were computed via SciPy's Short-Time Fourier Transform (STFT) function [Virtanen et al., 2020] for sensor 35, using two resolutions; panel a) covers 0.5-50 Hz ($\Delta t = 1.0$ s, $\Delta f = 0.5$ Hz), while panel b) focuses on 10-35 Hz ($\Delta t = 0.5$ s, $\Delta f = 1.0$ Hz) to better detect short-duration shedding events. Aside from the already known band at ≈ 1.9 Hz, no consistent pulses were observed. Since TS 8 represented the most extreme conditions of the campaign, we conclude that significant vortex shedding did not occur in any of the experiments. Finally, a second type of vortex shedding, induced by the oscillations of the airfoil, may occur and intensify as structural damage increases. Since this phenomenon is governed by the motion of the airfoil, its frequency is expected to coincide with the vibration frequency. Thus, however, this second type of vortex shedding is difficult to isolate within the frequency spectra.

3 CNN Analysis Based On Integrated Gradients

3.1 Integrated gradients

To mitigate the opacity of the CNN previously described and to identify which input features drive classification into a specific damage class, we subsequently perform an IG analysis. IG is an attribution method for neural networks introduced in [Sundararajan et al., 2017]. We emphasize, however, that IG highlights sensitivity of the model output to input perturbations, rather than providing a causal or physically unique explanation. For a multivariate time series sample $x \in \mathbb{R}^{C \times T}$ (compare Figure 5), where $C = 37$ denotes the number of channels and $T = 150$ the number of time steps, IG evaluates the importance of an element $x_{c,t}$, with $c \in [1, C]$ (mapped to the noncontiguous set of working sensors 0 to 36, see Figure 2) and $t \in [1, T]$, relative to a baseline $x' \in \mathbb{R}^{C \times T}$, for the classification function $F(x) : \mathbb{R}^{C \times T} \to [0; 1]^m$ encoded by the fully trained CNN. As the data set comprises six classes, $m = 6$. For this purpose, x' is defined such that it excludes a specific signal property, e.g. the mean value of x, by modifying all elements $x'_{c,t}$ accordingly. Then a linear path in the sample space $\mathbb{R}^{C \times T}$ (see Figure 12) is defined between the baseline x' and the true sample x. Following Equation 1, the IG for element $x_{c,t}$ is then computed by determining the gradient of F for

each point on this path $x' + \gamma(x - x')$ with $\gamma \in [0; 1]$, with respect to the c, t-th dimension and aggregating these:

$$IG_{c,t}(x) := (x_{c,t} - x'_{c,t}) \cdot \int_{\gamma=0}^{1} \frac{\partial F(x' + \gamma(x - x'))}{\partial x_{c,t}} d\gamma, \tag{1}$$

Computing $IG_{c,t}$ for all elements in x yields the attribution map $IG(x) \in \mathbb{R}^{C \times T}$ for sample x and baseline x'. The actual computation of $IG(x)$ is conducted via an approximation based on Riemann integration, where the integral is replaced by a sum over adequately sized intervals alongside the path between baseline and input sample. The number of steps L along this path must be chosen large enough to allow for good approximation of the original integral. Sundararajan et al. recommend choosing $20 < L < 300$ [Sundararajan et al., 2017].

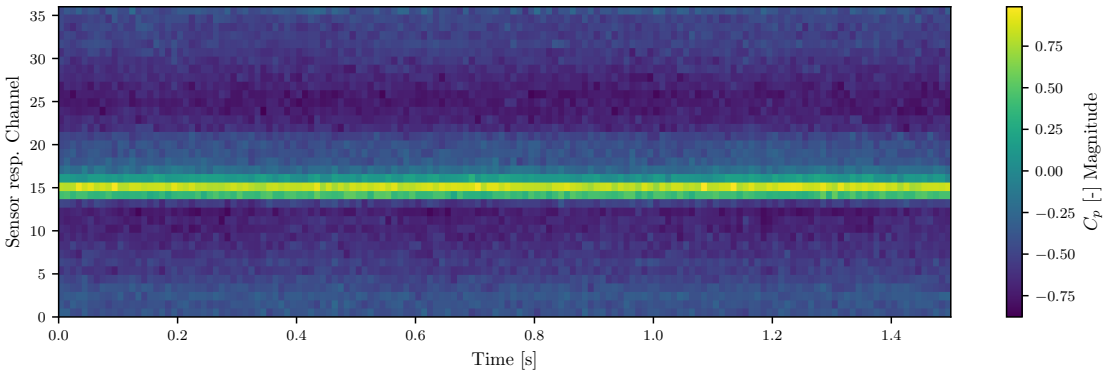


Figure 5: Exemplary multivariate time series sample (sample 14) $\in R^{37 \times 150}$ of the correctly classified samples of the validation set of split 1, 0° AoA.

3.2 Baselines and attribution analysis setup

For our analysis, we implement IG with TensorFlow[1] and choose the following baselines for our multivariate time series data:

1. Ambient Pressure Baseline (APB): The first baseline we employ consists of constant zeros across the entire sample. We adopt this baseline of an aerodynamic pressure coefficient $C_p = 0$ to emulate a scenario with no aerodynamic loading or vibration of the airfoil, thus only accounting for atmospheric pressure and neglecting sensor noise.
2. Temporal Variations Baseline (TVB): In this second baseline, we remove the offset of the signal in each channel, thereby eliminating the relative magnitude relationships between all channels while preserving only the temporal variations (see panel a) of Figure 6).
3. Channel-wise Mean Value Baseline (MVB): In this third baseline, the time-varying signal in each channel is replaced by its mean value. As a result, the sample becomes temporally constant across all channels (see Panel b) of Figure 6). This baseline eliminates the individual dynamics of each channel while preserving the relative magnitude relationships among them.

Subsequently, we compute attribution maps for correctly classified samples of the validation set to gain insight into the decision mechanism of the model. We restrict the attribution analysis to the validation sets of the two splits that yielded the lowest test set classification accuracy: split 1 for the 0° AoA data set and split 1 for the 8° AoA data set (see Table 2). For the 0° AoA validation set 1067 of 1068 samples were correctly classified and for 8° AoA validation set 1057 samples. As the signal windows are obtained by sliding a fixed-length window over the raw aerodynamic pressure signal, they are not phase-aligned. Consequently, attribution maps cannot be easily aggregated across all samples. The analysis that follows is therefore primarily qualitative and based on a small set of randomly chosen samples from each TS (see Table 1) and damage class rather than on the entire pool of correctly classified samples. For each sample in the validation set, we compute the corresponding attribution values with 200 steps along the linear path in the sample space. The resulting attribution maps contain positive and negative gradients, where positive gradients push the model towards the predicted class, negative gradients push against the predicted class. In addition, we compute

[1]Version 2.14.0 with Python 3.10.19

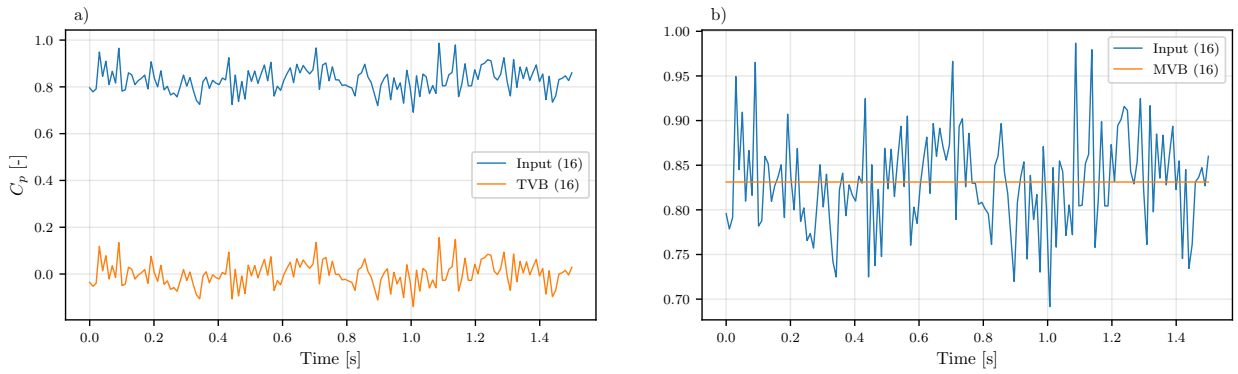


Figure 6: Visualizations of the TVB a) and MVB b) baselines for an exemplary sample and the signal of sensor 16. The APB, with the constant zero baseline is not displayed, as it consists of a constant line at zero for the duration of the sample.

the sum $\boldsymbol{c} \in \mathbb{R}^{37}$ of attributions per channel over the sample length (Equation 2), to evaluate which channels contribute most to the classification of the sample.

$$\boldsymbol{c} = \sum_{k=1}^{150} IG_k(x), \tag{2}$$

where $IG_k(x) \in \mathbb{R}^{37}$ describes the vector of attributions over all channels at the timestep k.

3.3 Attribution maps for the ambient pressure baseline

Figure 7 b) displays the attribution map using the APB for the randomly selected sample 14 of the correctly classified samples of the validation set of split 1. This sample was acquired in TS 1 (see Table 1) using an undamaged beam and is visualized in Figure 5. The attribution map covers the full 1.5 s duration of the signal and all 37 operational sensors. The color bar shown in panel a) is also applicable to panels b), d), e), and f). The attribution map is dominated by channels 14-16 which exhibit the largest gradients. Overall, significant IGs are concentrated in two bands between the sensors 9-17 and 23-29 and near zero elsewhere. This is also visible in panel c) that displays the sum of attributions over the sample duration (compare Equation 2), with sensors 14, 15, and 16 highlighted in orange as the largest. This indicates that the predictions of the 0° AoA CNN are primarily driven by these three adjacent sensors on the leading edge (compare Figure 2), with other channels contributing only weakly, e.g. by delivering an embedding for the information of the dominating channels. Panels d)-f) overlay the signals of sensors 14, 15, and 16 with their attribution scores, ordered by absolute sum (sensor 14 highest; 15 lowest). Sensors 15 and 16 exhibit three distinct episodes of mostly positive high-magnitude attributions, while sensor 14 displays four primarily negative episodes. Sensor 9 shows a similar temporal alignment to channels 15 and 16, whereas attributions for other sensors remain diffuse and lack distinctive patterns. These bursts in IG magnitude suggest that the 0° AoA CNN learned to respond to local signal morphology likely stemming from cyclic physical phenomena, rather than simple mean levels. However, these subsequences, potentially spanning multiple channels, remain physically difficult to interpret. These observations and conclusions hold across other inflow conditions (see Figure 13) and structural states (compare Figure 14). For 8° AoA, the attribution maps are similar across inflow conditions and damage states and exhibit minimal temporal variation (see Figures 15 and 16). Aside from slight patterns in TS 5 and 7, the attribution scores remain constant over the sample length, indicating that the CNN, trained on 8° AoA data, prioritizes static features over temporal dynamics. While leading edge sensors generally exhibit the highest gradients, some inflow conditions result in significant attributions across all sensors between 12 and 28. The violin plot in Figure 8 illustrates the summed attributions per channel (compare Equation 2) for all correctly classified validation samples of split 1 with 0° AoA, with red and black markers denoting the mean and median, respectively. Across all samples, sensors 14-17 near the leading edge exhibit the largest IGs and thus drive the prediction, albeit with the highest variance. Sensors 9-17 and 23-29 generally show IGs of relevant magnitude. The high variability and the bimodal distributions observed for sensors 23-36 likely stem from the diversity of inflow conditions and structural states, and to a lesser extent, from sample-to-sample variance. For sensors 7-10 and 22-27, gradients vary between positive and negative values suggesting these channels may suppress the predicted class or reflect sub-patterns in specific inflow or structural states. The remaining sensors with mean and median close to zero and a small spread are likely ignored by the model, indicating that they carry little class-discriminative information and may be noisy or redundant for the classification task. These findings also apply to

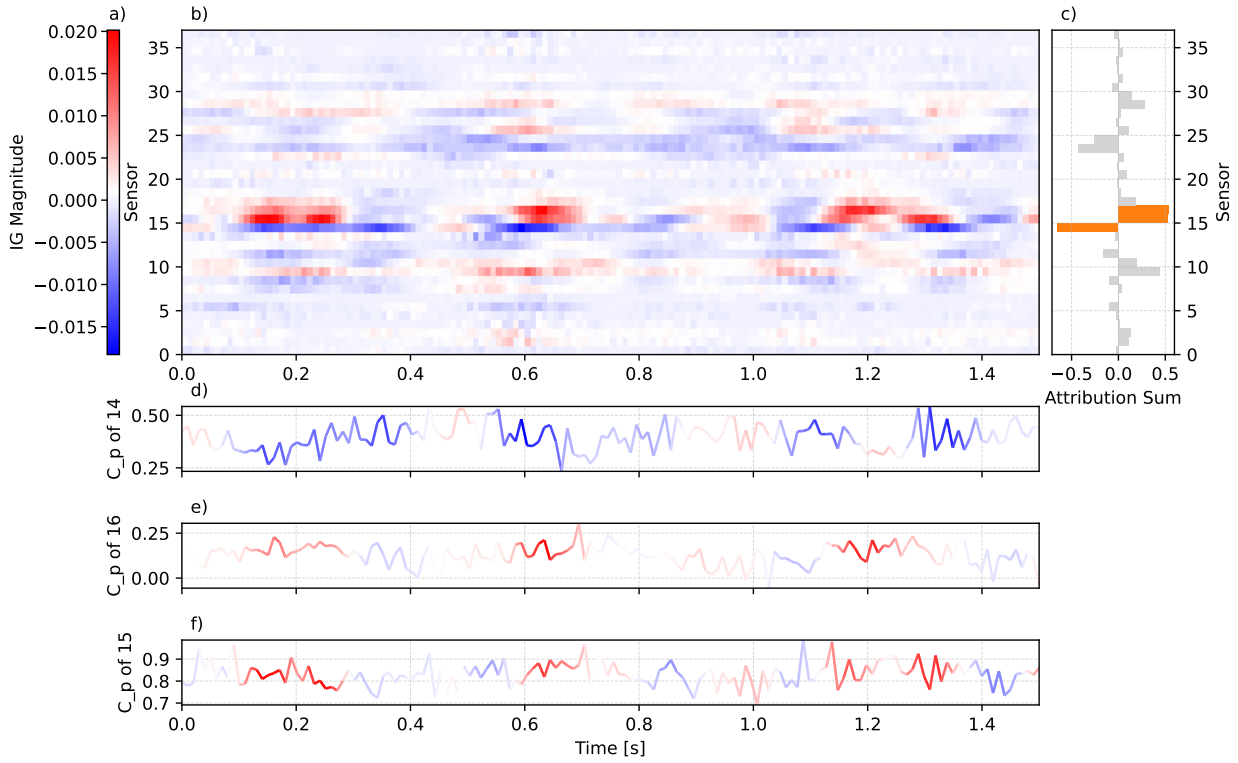


Figure 7: b) Attribution map for sample 14 (TS 1, damage class 0, split 1) using the APB. c) Sum of attributions per channel with the top three highlighted in orange. d)-f) Signals for the three highest-attribution channels, colored by attribution score. The colorbar in a) applies to b) and d)-f).

the 8° AoA attributions based on the APB, whose summed attribution distributions (Figure 17) closely resembles the 0° AoA attribution vector distribution, indicating that also the CNN trained on 8° AoA data focuses on the leading edge sensors.

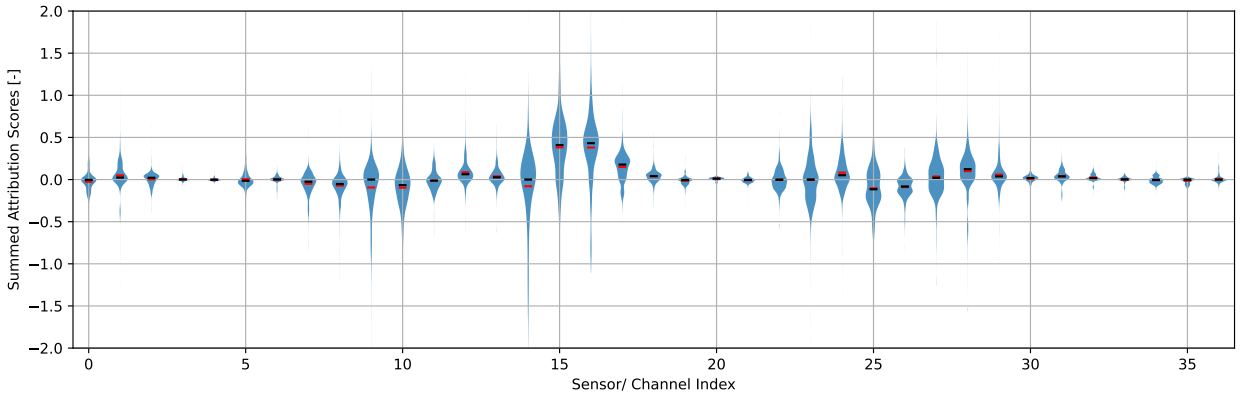


Figure 8: Violin plot presenting the distribution of the summed attribution vectors c (see Equation 2) for all samples of the correctly classified validation set (split 1, 0° AoA) using the APB. The red bar in each violin marks the mean value; the black bar the median.

3.4 Attribution maps for the temporal variations baseline

Figure 9 b) displays the attribution map based on the TVB for sample 14 of the correctly classified part of the validation set for split 1 and TS 1. Similar to the APB, the TVB attribution maps highlight sensors 9-17 and 23-29 as

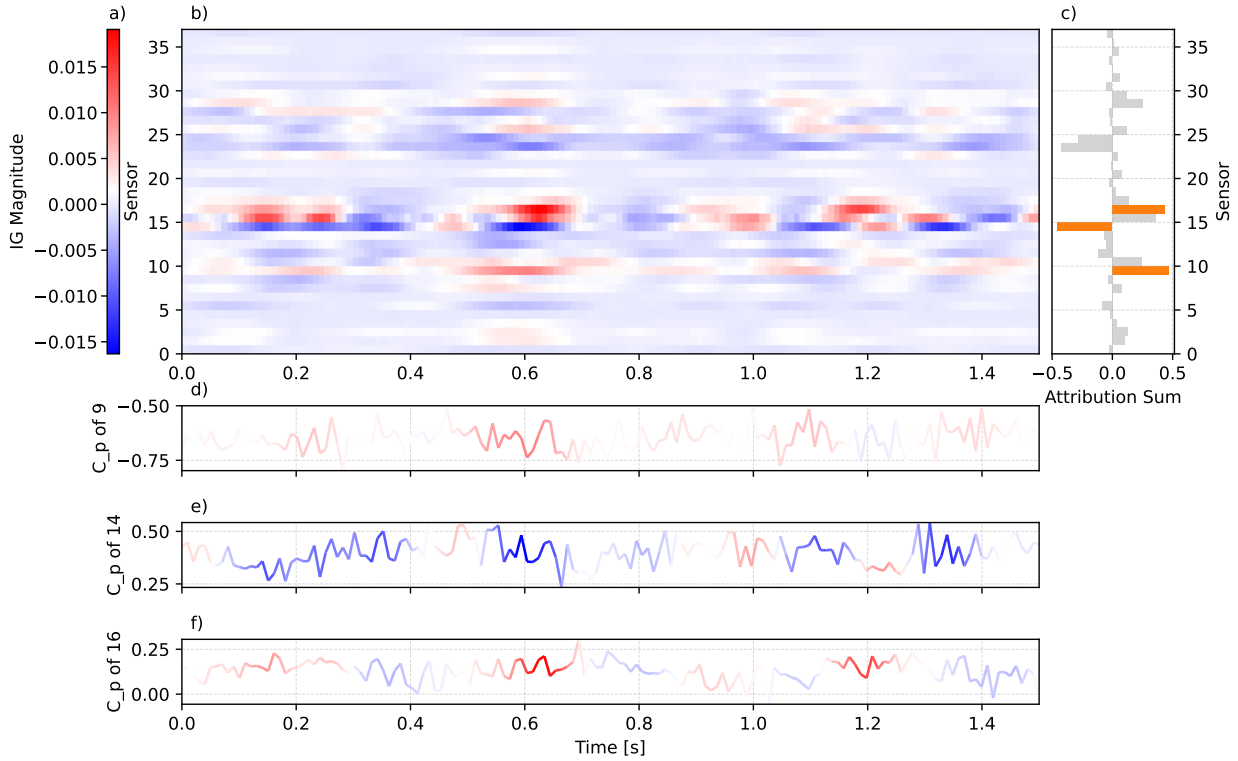


Figure 9: b) Attribution map for sample 14 (split 1, 0° AoA, damage class 0, TS 1) using the TVB. c) Sum of attributions per channel, with the top three highlighted in orange. d)-f) Signals for the three highest attribution channels, colored by attribution score. The colorbar in a) applies to b) and d)-f).

significant, a finding that remains consistent across different inflow conditions and AoAs, and damage states (compare Figures 18, 19, 21 and 22). For 0° AoA, the subsequences of IG magnitude in channels 14-16 mirror the APB, again suggesting the CNN trained on 0° AoA relies on recurring multivariate patterns rather than simple mean values, and remain physically difficult to interpret. Conversely, for 8° AoA (see Figure 21), the attribution maps exhibit fewer temporal variations and more constant sections for certain channels. This corroborates the indication from the APB that the 8° AoA CNN prioritizes static features or general signal characteristics over temporal dynamics. Evaluating the attribution vectors c (Equation 2) for correctly classified validation samples at 0° AoA via Figure 20 and at 8° AoA via Figure 23 reveals that, analogously to the APB, also for the TVB, both CNNs prioritize leading edge sensors (especially the channels 9-17 and 23-29), while trailing edge sensors (channels 0-8 and 30-36) receive negligible attributions, likely due to minimal contribution to the classification task or their smaller signal offsets.

3.5 Attribution maps for the mean value baseline

Figure 10 b) illustrates the attribution map using the MVB for the previously introduced sample 14. This map differs significantly from the previous attribution maps, as the IGs are more widely distributed across the entire sample. Furthermore, the concentration of high gradients in channels 14-16 is less pronounced; instead, the higher attribution scores are spread across channels 11-26. Summing the IGs over the sample length shows that channels 13, 16 and 17 exhibit the highest IG sums (see panel c)), although these sums range approximately from -0.1 to 0.25, which is lower than those observed for the APB and TVB. As before, panels d), e), and f) display measurement signals overlaid with attribution scores. The temporal structure of these attributions is scattered, highlighting individual time steps rather than subsequences, unlike the other baselines. Assuming that relevant physical phenomena occur with a frequency smaller than 50 Hz, these should, given our sampling frequency of f_s =100 Hz, manifest over time intervals rather than single time steps. Therefore, we attribute these scattered attributions to the baseline choice rather than the underlying

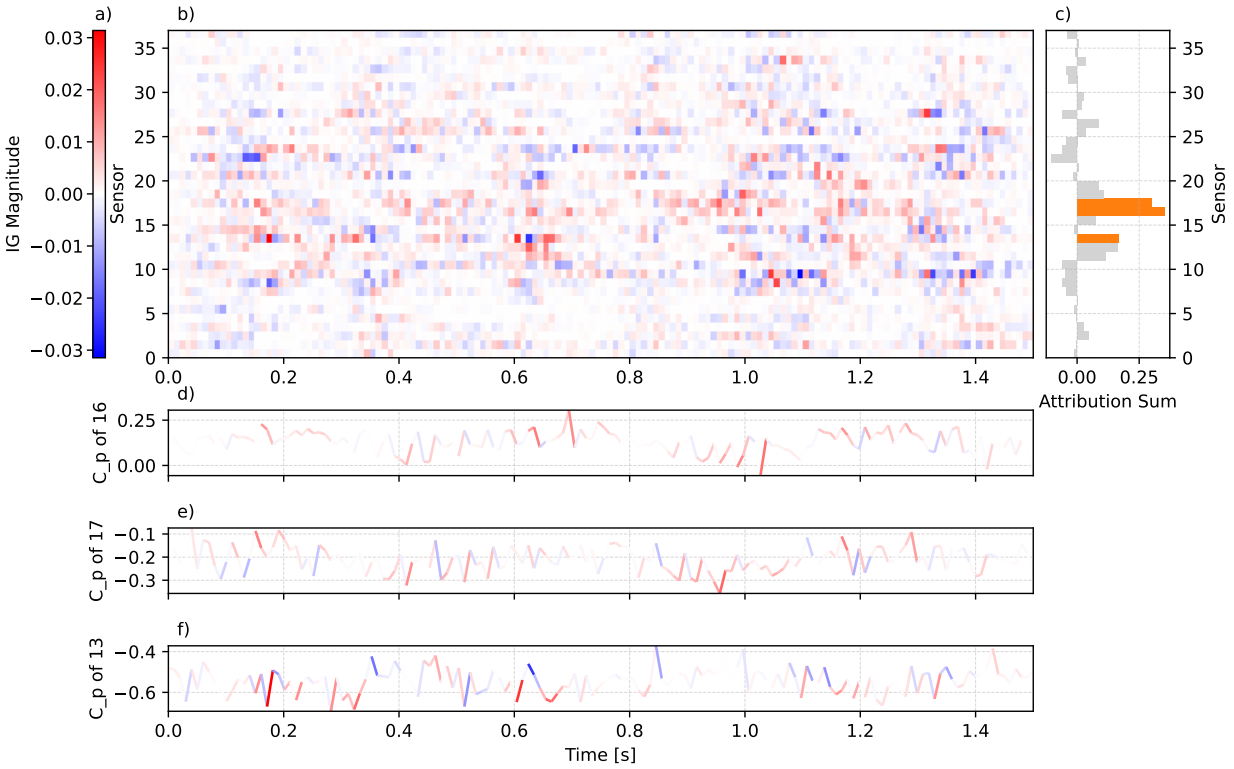


Figure 10: b) Attribution map for sample 14 (split 1, 0° AoA, damage class 0, TS 1) using the MVB. c) Sum of attributions per channel, with the top three channels highlighted in orange. d)-f) Signals for the three highest-attribution channels, color-coded by attribution score. The colorbar in a) applies to b) and d)-f).

physics. These findings persist across samples recorded under different inflow conditions (see Figure 24) and damage states (see Figure 25). The violin plot of the attribution vectors c in Figure 26 reveals notably lower attribution scores compared to the other baselines. While the attribution vectors c of the other baselines span approximately $[-2; 2]$, the MVB attributions lie mainly in the range of $[-0.5; 0.5]$. Also for the MVB, the channels 10-17, located around the leading edge, exhibit higher IG sums accompanied by considerable variability. As one can see in Figures 27-29, these findings also hold for the 8° AoA data and CNN. These findings confirm once more that both CNNs are highly sensitive to signals of the leading edge sensors. Overall, the MVB yields for both AoAs substantially lower attribution scores than the APB or TVB. This suggests that the MVB already captures significant class-discriminative information and hints that temporal variations contribute only marginally to the model's predictions.

4 Reevaluation Of The CNNs Based On Insights From The Attribution Analysis

4.1 Pre-trained models

After examining the attribution maps for the different baselines, we now quantitatively assess how the CNNs of [Franz et al., 2025], trained on the original, unaltered recordings, perform when the input is reduced to each baseline and thus lacking a significant signal property of the input samples. For the APB and TVB, the accuracy of the pre-trained models falls for each split to $1/6 \approx 16.7\%$, which is indistinguishable from random guessing in a six-class problem. In contrast, the mean value baseline yields considerably higher scores on all splits (see Table 3): about 50% for the 0° AoA and roughly 62% for the 8° AoA. These results suggest that the previously trained CNNs of [Franz et al., 2025] rely primarily on the relative magnitudes across sensors rather than on the temporal dynamics of the signals. This also fits the results obtained by the attribution analysis of the MVB: the scattered and very small attribution scores already hinted that the MVB might contain a significant amount of class-discriminative information. To verify this hypothesis we will subsequently retrain the CNNs on the baseline data.

Table 3: Classification accuracy on the test sets for the MVB with CNNs of [Franz et al., 2025] that were trained on the unaltered input samples. The results for the APB and TVB are omitted in this table as for these baselines the accuracy for each split is ≈16.67%.

	MVB			
	Split 1	Split 2	Split 3	Average
0° AoA	48.3%	53.0%	49.9%	50.4%
8° AoA	61.3%	60.2%	64.6%	62.0%

4.2 Retrained models

To assess information retention, we retrain the CNN of Section 2.2 on the TVB and a regularized MLP on the MVB, collapsed to mean value vectors $\in \mathbb{R}^{37}$. While an MLP was used to maintain comparability with the CNN, classical learners like random forests [Breiman, 2001] or gradient-boosting machines [Friedman, 2001] are also viable alternatives. The complete pipeline, including preprocessing and the MLP architecture, is illustrated in Figure 30 in the appendix. The preprocessing of the measurement data for the MLP is the same as for the CNN approach; only after computing the mean vectors an additional z-scoring is conducted (compare Figure 30). The MLP comprises four fully-connected layers and employs batch normalization and dropout layers to improve generalization and speed up training. The training protocol is similar to the CNN training and only slightly modified to enhance the stability of the training curve; no hyperparameter optimization was conducted. The full training details are provided in the Appendix 7.7. The classification performance of the CNN for the TVB and of the MLP for the MVB is subsequently reported in Table 4. As the classification problem is balanced, we present in Table 4 only the average accuracy over the six classes for each split of each AoA. The APB is omitted in this evaluation, as training a classifier on zero-only samples is not possible. Remarkably, the MLP using only mean values of the MVB yields a high classification accuracy, which is for each split larger than 70% and on average 86.5% resp. 83.7% for 0° and 8° AoA. The CNN using the TVB samples achieves a considerably lower overall accuracy for each split, with an average of 47.4% and 34.8%. Although it is likely that with more sophisticated neural networks and training approaches better classification results may be obtained, these results are a strong indicator that especially information encoded by the mean vector of a sample, most likely the inter-channel magnitude relationships, carry significant class-discriminative information.

Table 4: Classification accuracy on the MVB and TVB test sets of the retrained networks.

	MVB (MLP)				TVB (CNN)			
Split	1	2	3	Average	1	2	3	Average
0° AoA	74.3%	98.3%	86.8%	86.5%	45.3%	47.0%	50.0%	47.4 %
8° AoA	72.4%	93.8%	85.0%	83.7%	33.7%	36.5%	34.1%	34.8 %

5 Discussion

Across all investigated baselines, the CNNs of Franz et al. [2025] consistently prioritize sensors around the leading edge, particularly channels 14-16, regardless of inflow conditions or AoA. Furthermore, the attribution maps of the APB and TVB reveal that the CNN trained at 0° AoA, detects recurring temporal subsequences, whereas for the CNN trained on 8° AoA data, attribution scores become approximately constant, indicating a reliance on time-independent features of the 8° AoA model. In contrast, the MVB produces scattered IGs without clear temporal patterns and with magnitudes often significantly smaller than those of the APB/TVB for both AoAs. This suggests that the MVB already captures substantial class-discriminative information. When evaluated on the TVB, the accuracy of the pre-trained CNNs collapses to ≈16.7%, while the MVB retains accuracies of ≈50% (0° AoA) and 62% (8° AoA). Comparing a CNN retrained on TVB with an MLP trained on the MVB mean vectors further demonstrates this disparity: the MLP achieves 86.5% (0° AoA) and 83.7% (8° AoA), compared to 47.4% (0° AoA) and 34.8% (8° AoA) for the retrained CNN. Based on these results, we conclude that spatial sensor relationships, here reflected in the channel mean values, constitute the primary source of class-discriminative information. While temporal dynamics also contribute, as evidenced by the TVB accuracies and the superior performance of full-sample classification over mean value vectors, their influence appears secondary. The pronounced performance gap between the MVB and TVB suggests that temporal features are either intrinsically less informative or more difficult to extract, potentially due to signal noise or limitations of the employed neural network architecture. It should be noted that this conclusion may be influenced

by the representational capacity of the chosen architecture, which may not optimally capture temporal dependencies. Regarding these findings now in the context of the effects of damage on the structure and the APD, the classification accuracy of our CNNs is primarily driven by static, global features and, to a lesser extent, by temporal patterns. Given the absence of significant vortex shedding in this experimental campaign, these temporal variations are attributed to the oscillation amplitudes of the cantilever beam. Furthermore, the reliance on static features suggests shifts in the data across structural states, which likely stem from the increasing deformation of the equilibrium position for each state. However, the extent to which the static deformations of the cantilever beam alone account for the observed shifts in the APD remains an open question. Furthermore, the origin of the high-magnitude negative IGs in certain leading edge channels, particularly channel 14, cannot yet be explained from a physical perspective. While the attribution analysis identifies dominant feature patterns, it does not resolve class-specific differences. More structured XAI approaches, such as class-prototype-based methods, may therefore be required to disentangle state-dependent characteristics.

6 Conclusion & Outlook

In this paper, we employ IG to enhance the interpretability of the CNN-based structural damage detection approach presented in [Franz et al., 2025]. By analyzing the attribution maps from IG based on the APB, TVB, and MVB, we find that the CNNs focus on the signals of sensors located around the leading edge of the airfoil and rely primarily on static, inter-channel magnitude relationships, likely stemming from shifts in the equilibrium position of the structural system, rather than temporal dynamics. While temporal patterns do provide some discriminative information, as evidenced by the performance of the retrained network and the superiority of the CNNs of Franz et al. [2025] over the mean-value MLP, their role is secondary. The exact origin of these static shifts needs to be verified and the high-magnitude negative IGs in leading edge sensors remain unexplained, providing a direction for future experimental design. Future work should explore more sophisticated baselines, for example in the frequency domain, and more advanced, time-series-specific IG variants [Jang et al., 2025] to extract more targeted features. Additionally, an XAI approach using class-prototypes [Theissler et al., 2022] could reveal state-specific features by identifying class representative samples, although high dimensionality and noise remain significant challenges. Finally, developing an unsupervised approach for damage detection and ranking remains a primary objective.

Acknowledgements
The authors gratefully acknowledge the computational resources provided through the joint high-performance data analytics (HPDA) project “terrabyte” of the German Aerospace Center (DLR) and the Leibniz Supercomputing Center (LRZ). AP acknowledges support by dtec.bw - Digitalization and Technology Research Center of the Bundeswehr (project RISK.twin). dtec.bw is funded by the European Union - Next GenerationEU. GD and EC were partially supported by the French-Swiss project MISTERY funded by the French National Research Agency (ANR PRCI Grant No. 266157) and the Swiss National Science Foundation (Grant No. 200021L_212718).

Use of AI tools declaration
The authors declare they have used Artificial Intelligence (AI) tools in the creation of this article: In some paragraphs throughout the article, AI tools have been used to rephrase and improve sentences grammatically and language-wise. We emphasize that no content has been generated by AI tools.

References

Philip Franz, Imad Abdallah, Gregory Duthé, Julien Deparday, Ali Jafarabadi, Xudong Jian, Max von Danwitz, Alexander Popp, Sarah Barber, and Eleni Chatzi. On the potential of aerodynamic pressure measurements for structural damage detection. *Wind Energy Science*, 10(12):3001–3026, 2025. doi:10.5194/wes-10-3001-2025.

Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019. ISSN 1384-5810. doi:10.1007/s10618-019-00619-1.

Navid Mohammadi Foumani, Lynn Miller, Chang Wei Tan, Geoffrey I. Webb, Germain Forestier, and Mahsa Salehi. Deep learning for time series classification and extrinsic regression: A current survey. *ACM Comput. Surv.*, 56(9), 2024. ISSN 0360-0300. doi:10.1145/3649448.

Xiangjie Kong, Zhenghao Chen, Weiyao Liu, Kaili Ning, Lechao Zhang, Syauqie Muhammad Marier, Yichen Liu, Yuhao Chen, and Feng Xia. Deep learning for time series forecasting: a survey. *International Journal of Machine Learning and Cybernetics*, 16(7-8):5079–5112, 2025. ISSN 1868-8071. doi:10.1007/s13042-025-02560-w.

Zahra Zamanzadeh Darban, Geoffrey I. Webb, Shirui Pan, Charu Aggarwal, and Mahsa Salehi. Deep learning for time series anomaly detection: A survey. *ACM Comput. Surv.*, 57(1), 2024. ISSN 0360-0300. doi:10.1145/3691338.

Hao Wang, Baoli Wang, and Caixia Cui. Deep learning methods for vibration-based structural health monitoring: A review. *Iranian Journal of Science & Technology Transactions of Civil Engineering*, 48(4):1837–1859, 2024. doi:10.1007/s40996-023-01287-4.

Young-Jin Cha, Rahmat Ali, John Lewis, and Oral Büyüköztürk. Deep learning-based structural health monitoring. *Automation in Construction*, 161:105328, 2024. doi:10.1016/j.autcon.2024.105328.

Arnaud Vadeboncoeur, Gregory Duthé, Mark Girolami, and Eleni Chatzi. Geometric autoencoder priors for bayesian inversion: Learn first observe later. In *The Fourteenth International Conference on Learning Representations*, 2026. URL `https://openreview.net/forum?id=QMMlaI9VbR`.

Nathan Baker, Frank Alexander, Timo Bremer, Aric Hagberg, Yannis Kevrekidis, Habib Najm, Manish Parashar, Abani Patra, James Sethian, Stefan Wild, Karen Willcox, and Steven Lee. Workshop report on basic research needs for scientific machine learning: Core technologies for artificial intelligence, 2019.

Alfio Quarteroni, Paola Gervasio, and Francesco Regazzoni. Combining physics-based and data-driven models: advancing the frontiers of research with scientific machine learning. *Mathematical Models and Methods in Applied Sciences*, 35(04):905–1071, 2025. doi:10.1142/S0218202525500125.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. doi:10.1038/s42254-021-00314-5.

Alice Cicirello. Physics-enhanced machine learning: a position paper for dynamical systems investigations. *Journal of Physics: Conference Series*, 2909(1):012034, 2024. doi:10.1088/1742-6596/2909/1/012034.

Sarah Barber, Julien Deparday, Yuriy Marykovskiy, Eleni Chatzi, Imad Abdallah, Gregory Duthé, Michele Magno, Tommaso Polonelli, Raphael Fischer, and Hanna Müller. Development of a wireless, non-intrusive, mems-based pressure and acoustic measurement system for large-scale operating wind turbine blades. *Wind Energy Science*, 7(4): 1383–1398, 2022. doi:10.5194/wes-7-1383-2022.

Andreas Theissler, Francesco Spinnato, Udo Schlegel, and Riccardo Guidotti. Explainable ai for time series classification: A review, taxonomy and research directions. *IEEE Access*, 10:100700–100724, 2022. doi:10.1109/ACCESS.2022.3207765.

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 3319–3328. JMLR.org, 2017. URL `https://dl.acm.org/doi/10.5555/3305890.3306024`.

Kevin Fauvel, Tao Lin, Véronique Masson, Élisa Fromont, and Alexandre Termier. Xcm: An explainable convolutional neural network for multivariate time series classification. *Mathematics*, 9(23):3137, 2021. doi:10.3390/math9233137.

Szymon Bobek and Grzegorz J. Nalepa. Tsproto: Fusing deep feature extraction with interpretable glass-box surrogate model for explainable time-series classification. *Information Fusion*, 124:103357, 2025. doi:10.1016/j.inffus.2025.103357.

Bartłomiej Małkus, Szymon Bobek, and Grzegorz J. Nalepa. Prototsnet: Interpretable multivariate time series classification with prototypical parts *[preprint]*. 2025. doi:10.48550/arXiv.2511.02152.

Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1578–1585. IEEE, 2017. ISBN 978-1-5090-6182-2. doi:10.1109/IJCNN.2017.7966039.

Yahya Modarres-Sadeghi, editor. *Introduction to Fluid-Structure Interactions*. Springer International Publishing and Imprint Springer, Cham, 1st ed. 2021 edition, 2021. ISBN 978-3-030-85884-1. doi:10.1007/978-3-030-85884-1.

Jianlong Chang, Qingui Zhang, Liujing He, and Yi Zhou. Shedding vortex characteristics analysis of naca 0012 airfoil at low reynolds numbers. *Energy Reports*, 8:156–174, 2022. doi:10.1016/j.egyr.2022.01.149.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi:10.1038/s41592-019-0686-2.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 0885-6125. doi:10.1023/A:1010933404324.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5): 1189–1232, 2001. ISSN 00905364, 21688966. URL `http://www.jstor.org/stable/2699986`.

Hyeongwon Jang, Changhun Kim, and Eunho Yang. TIMING: Temporality-aware integrated gradients for time series explanation. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 26877–26895. PMLR, 13–19 Jul 2025. URL `https://proceedings.mlr.press/v267/jang25a.html`.

7 Appendix

7.1 Structural states & damage classes

Table 5: Mapping of structural states to damage classes from Franz et al. [2025].

Damage Class	0	1	2	3	4	5
Crack length in [% of beam-width]	0%	0% and added mass	12.5%	25%	37.5%	50%

7.2 Spectra of a pressure signal recorded near the trailing edge

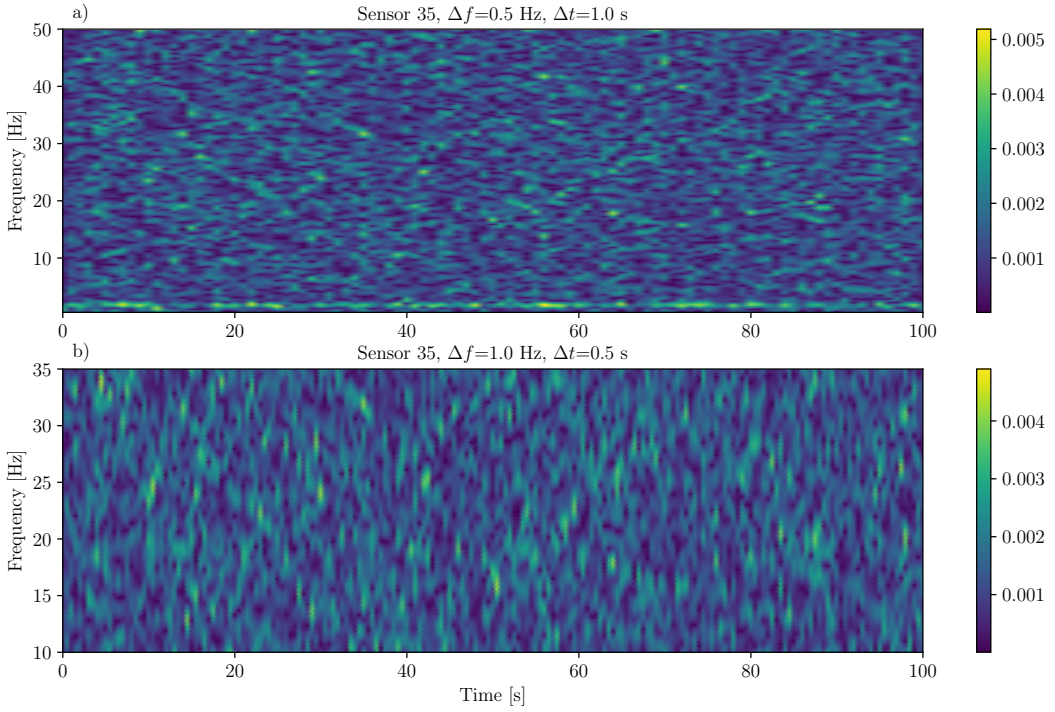


Figure 11: STFT spectra of sensor 35 (trailing edge) for TS 8 with a 50% beam-width crack. Panels a) and b) show different time-frequency resolutions for the ranges 0.5-50 Hz and 10-35 Hz, respectively. No prominent energy bands or regular pulses are detected at 15 Hz or 30 Hz.

7.3 Integration Path in the Sample Space for Integrated Gradients

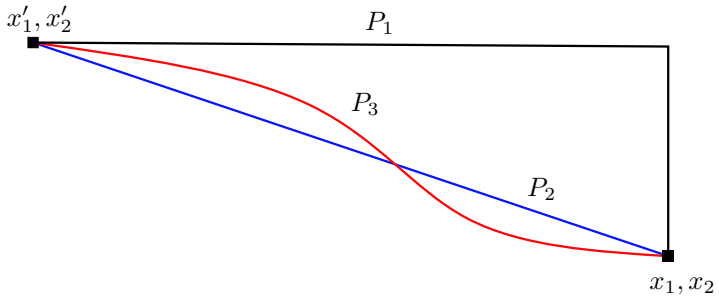


Figure 12: Example of possible integration paths in two dimensions between the baseline x' and the full input sample x. IG takes the linear integration path P_2 (blue). Image inspired by Sundararajan et al. [2017].

7.4 Plots for the ambient pressure baseline

7.4.1 Attribution maps for 0° AoA and varying inflow conditions

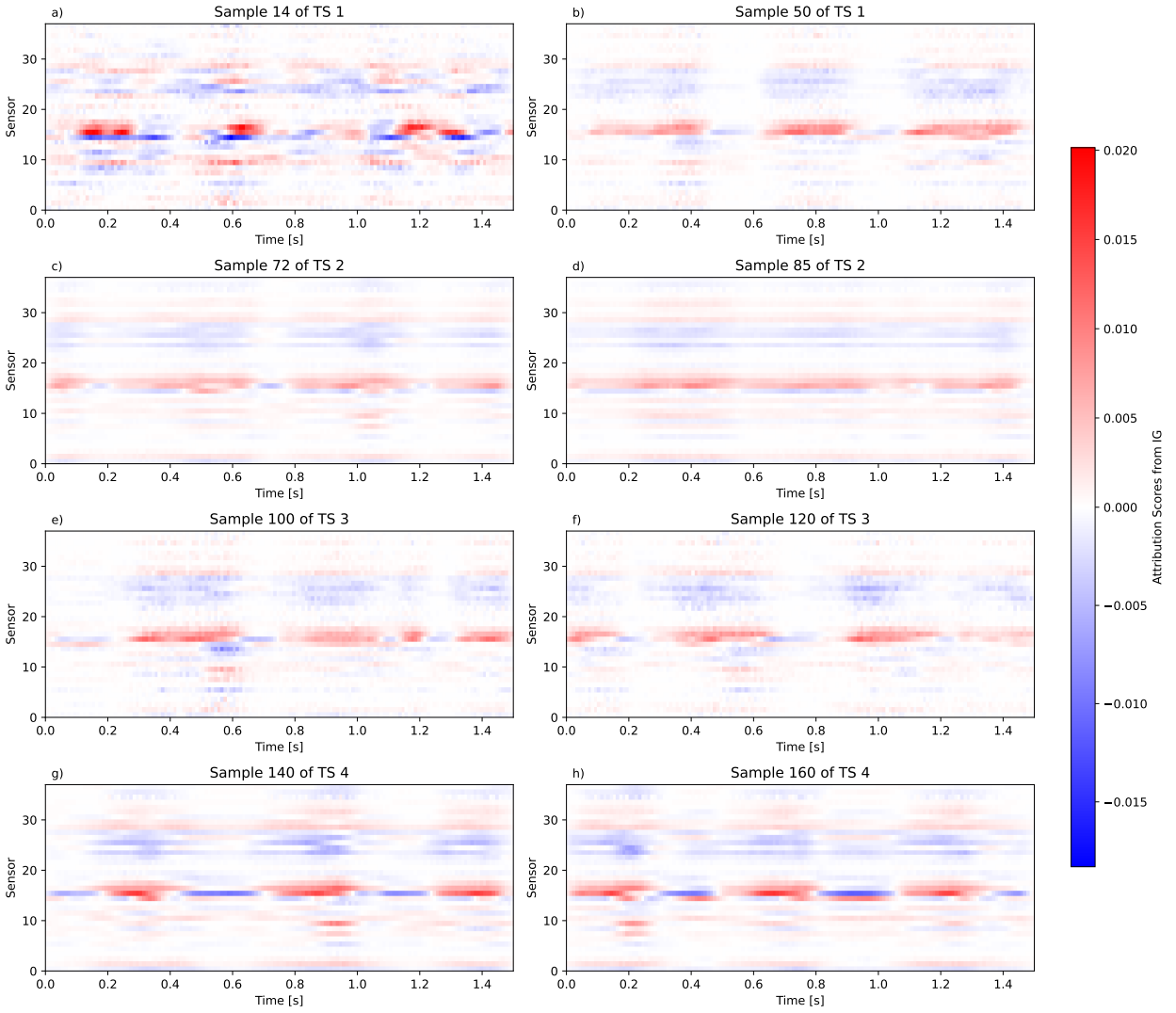


Figure 13: Overview of attribution maps based on the APB for different inflow conditions for damage class 0 and 0° AoA. Panels a) and b) belong to TS 1 (see Table 1), c) and d) to TS 2, e) and f) to TS 3 and g) and h) to TS 4.

7.4.2 Attribution maps for 0° AoA and varying structural states

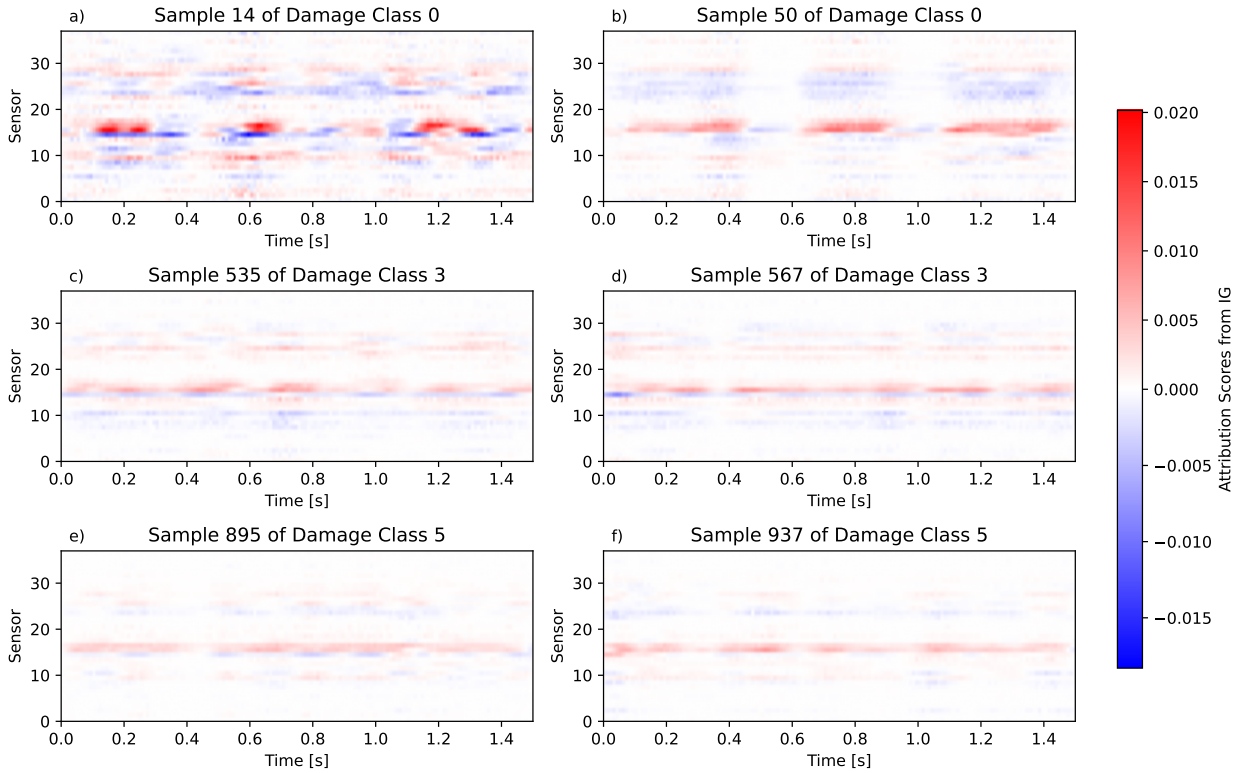


Figure 14: Overview of attribution maps based on the APB for different structural states for TS 1 and 0° AoA. Panels a) and b) belong to damage class 0 (no damage), c) and d) to damage class 3 (crack length of 25% of the beam width) and e) and f) to damage class 5 (crack length of 50% of beam width).

7.4.3 Attribution maps for 8° AoA and Varying Inflow Conditions

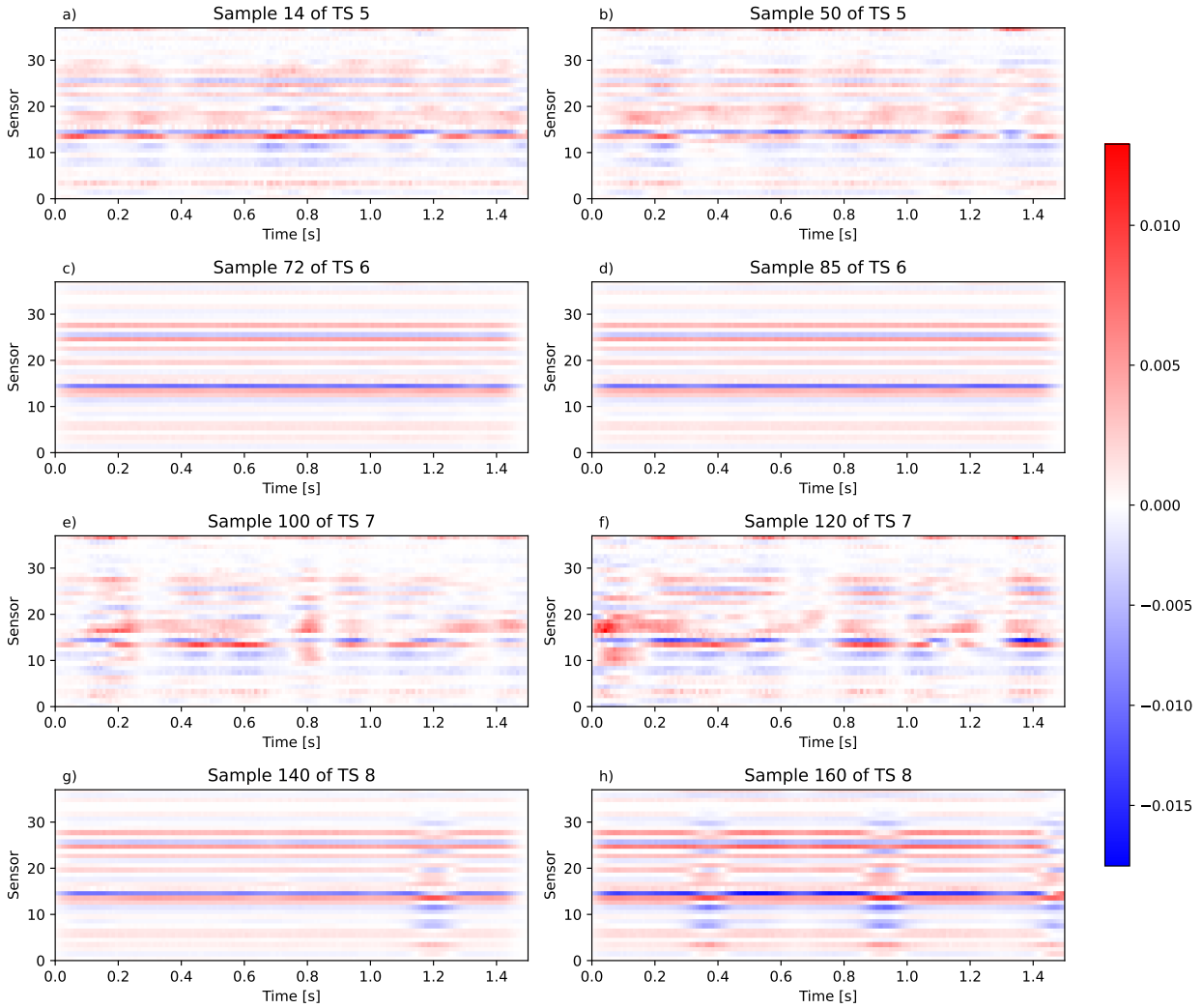


Figure 15: Overview of attribution maps based on the APB for different inflow conditions for damage class 0 and 8° AoA. Panels a) and b) belong to TS 5 (see Table 1), c) and d) to TS 6, e) and f) to TS 7 and g) and h) to TS 8.

7.4.4 Attribution maps for 8° AoA and Varying Structural States

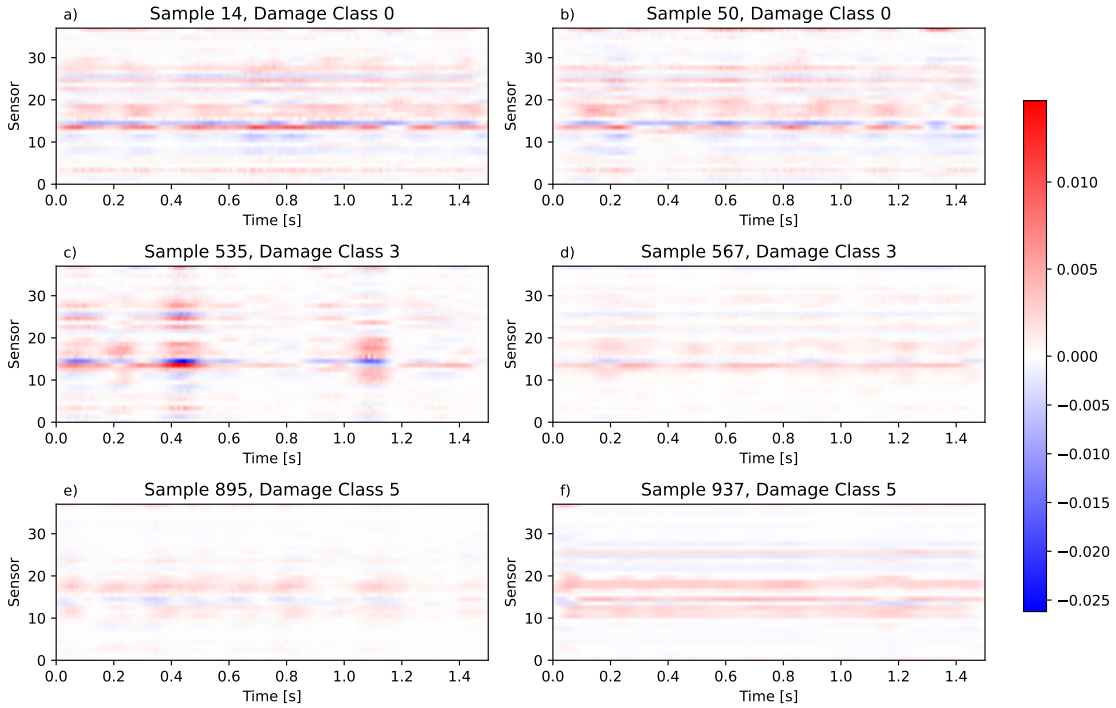


Figure 16: Overview of attribution maps based on the APB for different structural states for TS 5 and 8° AoA. Panels a) and b) belong to damage class 0 (no damage), c) and d) to damage class 3 (crack length of 25% of the beam width) and e) and f) to damage class 5 (crack length of 50% of beam width).

7.4.5 Distribution of summed attribution vectors for 8° AoA

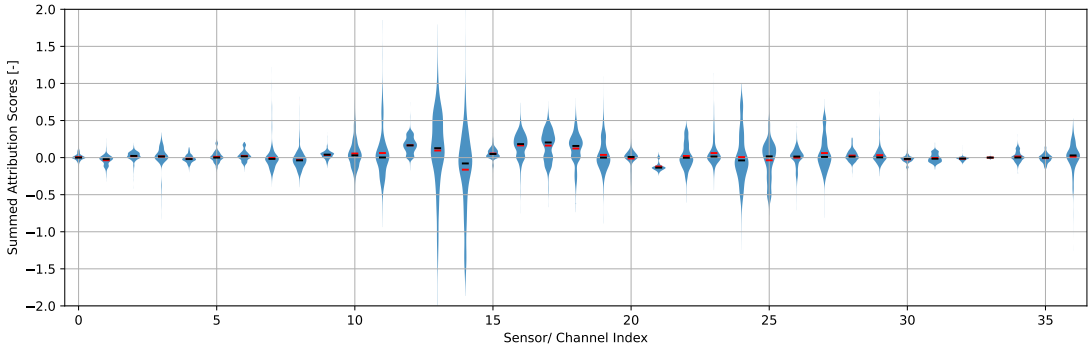


Figure 17: Distribution of summed attribution vectors for the APB and 8° AoA for all correctly classified samples of the validation set (red bar: mean value, black bar: median).

7.5 Plots for the temporal variations baseline

7.5.1 Attribution maps for 0° AoA and varying inflow conditions

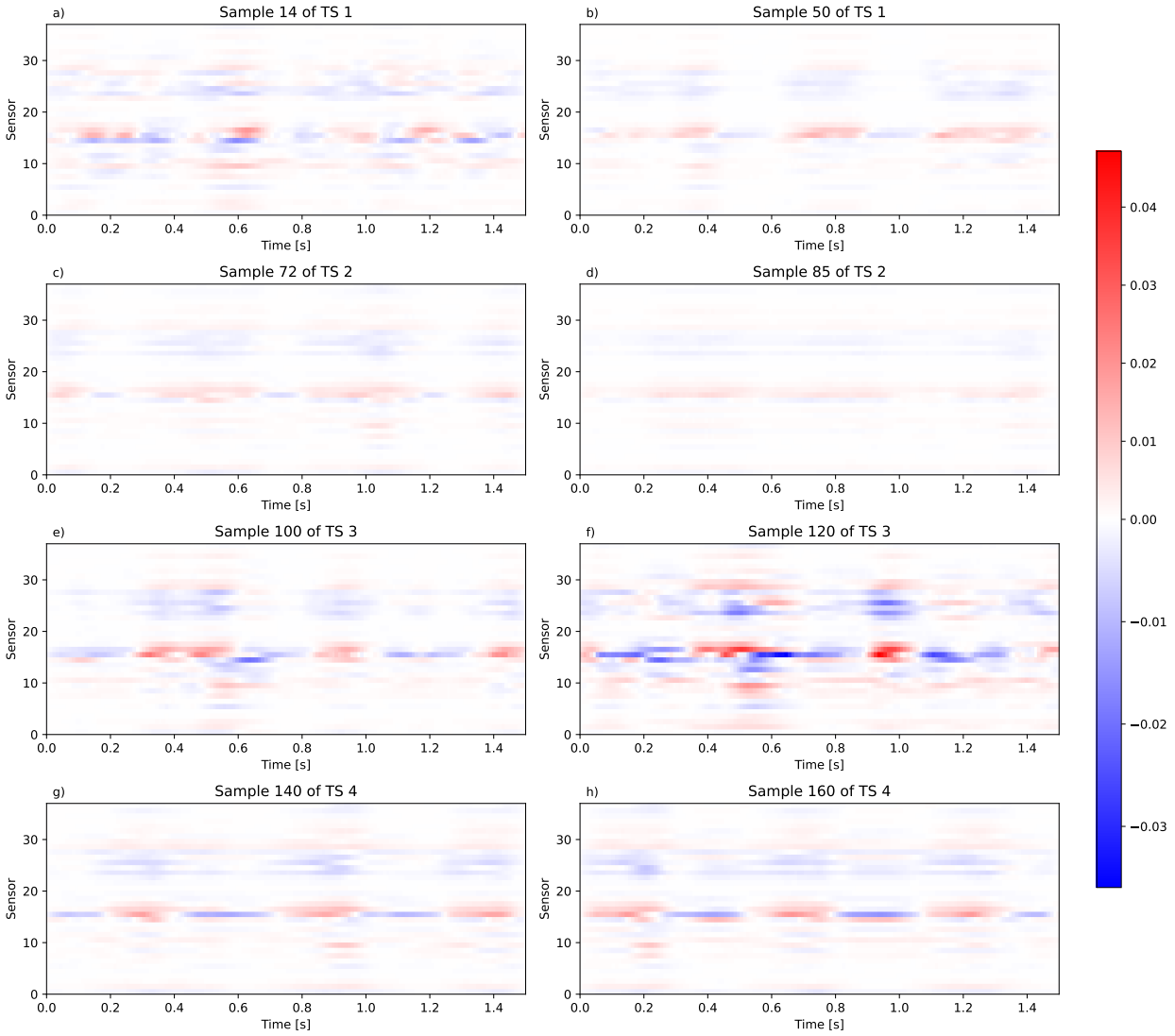


Figure 18: Overview of attribution maps based on the TVB for different inflow conditions for damage class 0 and 0° AoA. Panels a) and b) belong to TS 1 (see Table 1), c) and d) to TS 2, e) and f) to TS 3 and g) and h) to TS 4.

7.5.2 Attribution maps for 0° AoA and varying structural states

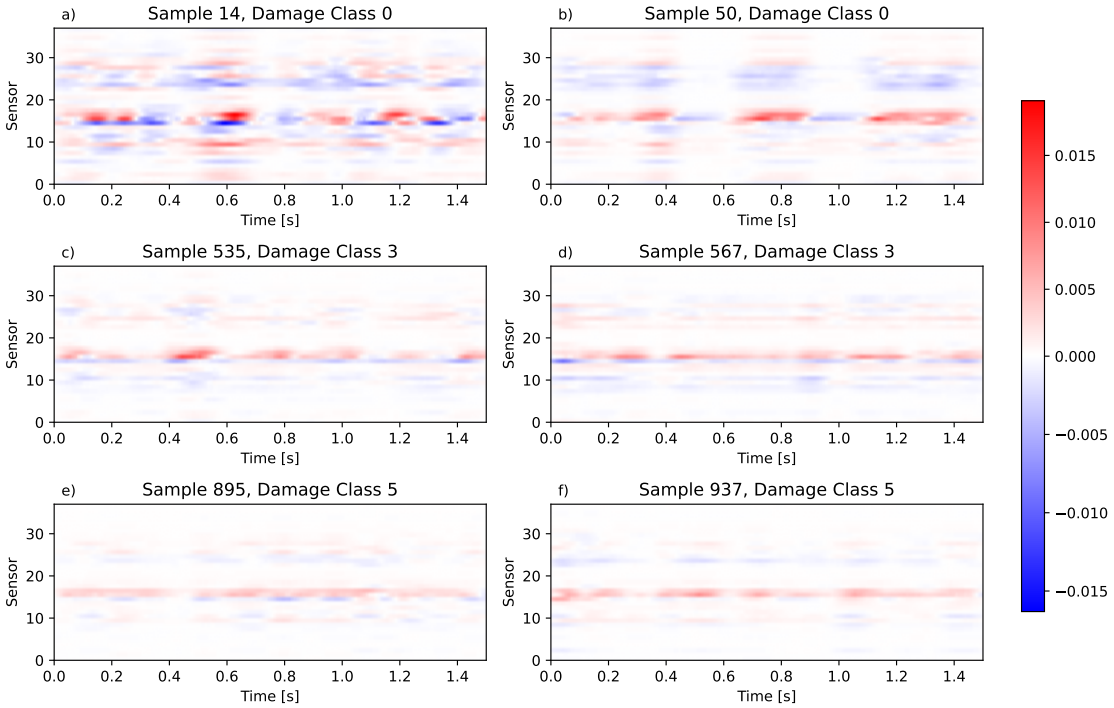


Figure 19: Overview of attribution maps based on the TVB for different structural states for TS 1 and 0° AoA. Panels a) and b) belong to damage class 0 (no damage), c) and d) to damage class 3 (crack length of 25% of the beam width) and e) and f) to damage class 5 (crack length of 50% of beam width).

7.5.3 Distribution of summed attribution vectors for 0° AoA

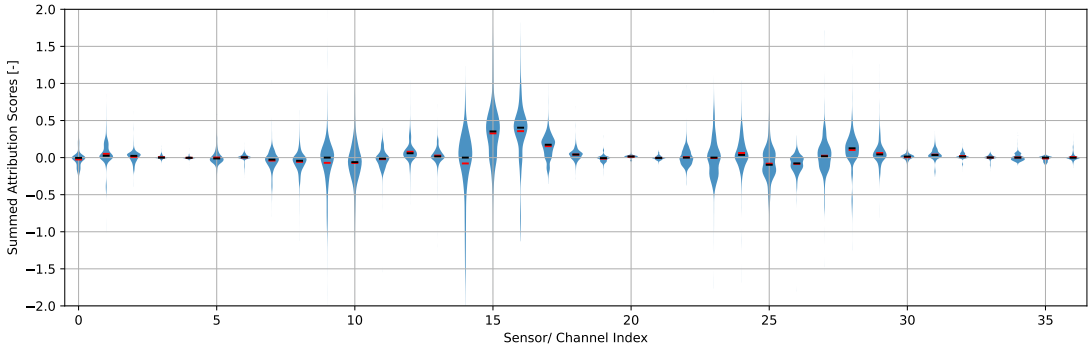


Figure 20: Distribution of summed attribution vectors for the TVB and 0° AoA, for all correctly classified samples of the validation set (red bar: mean value, black bar: median).

7.5.4 Attribution Maps for 8° AoA and varying inflow conditions

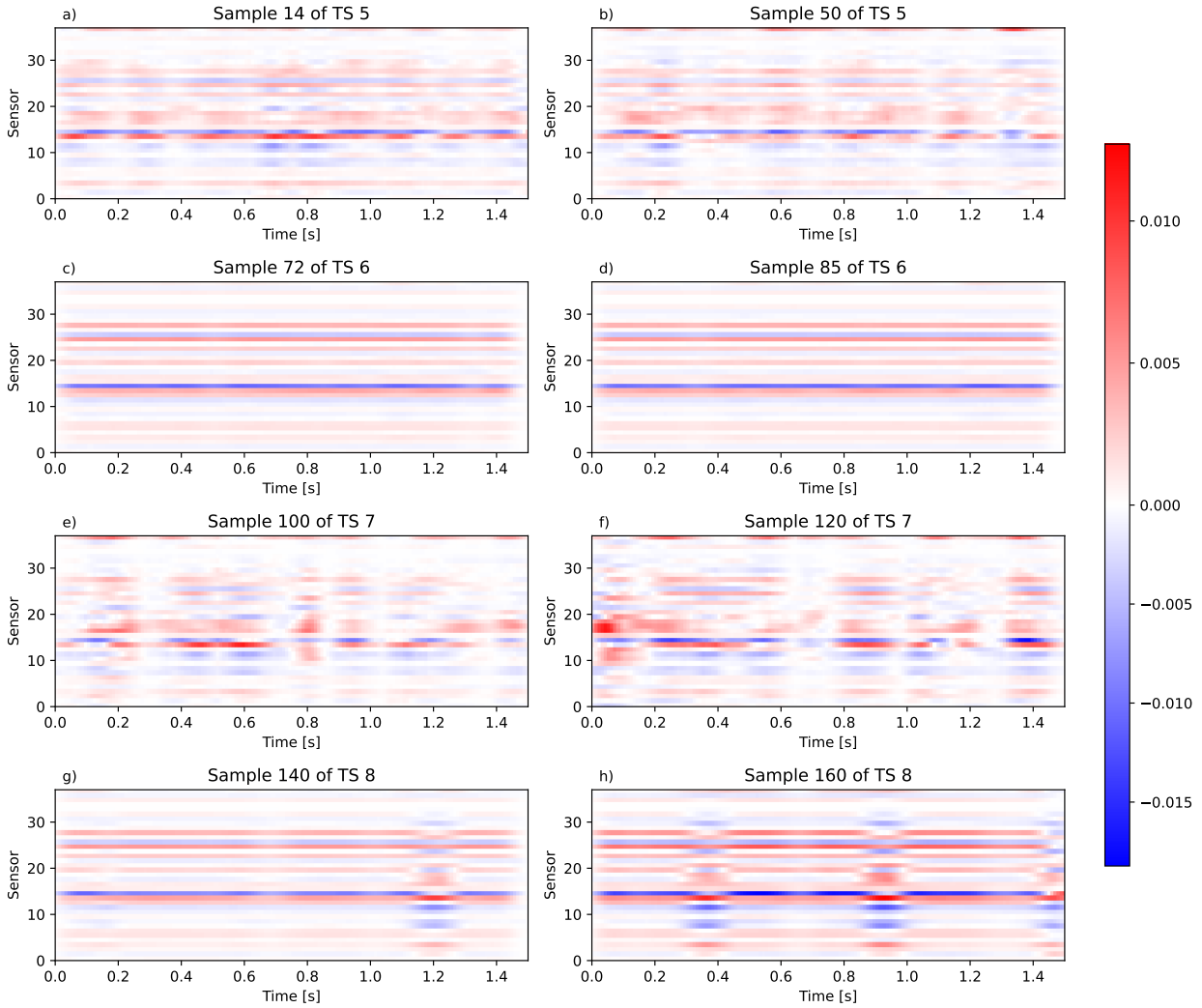


Figure 21: Overview of attribution maps based on the TVB for different inflow conditions for damage class 0 and 8° AoA. Panels a) and b) belong to TS 5 (see Table 1), c) and d) to TS 6, e) and f) to TS 7 and g) and h) to TS 8.

7.5.5 Attribution Maps for 8° AoA and varying structural states

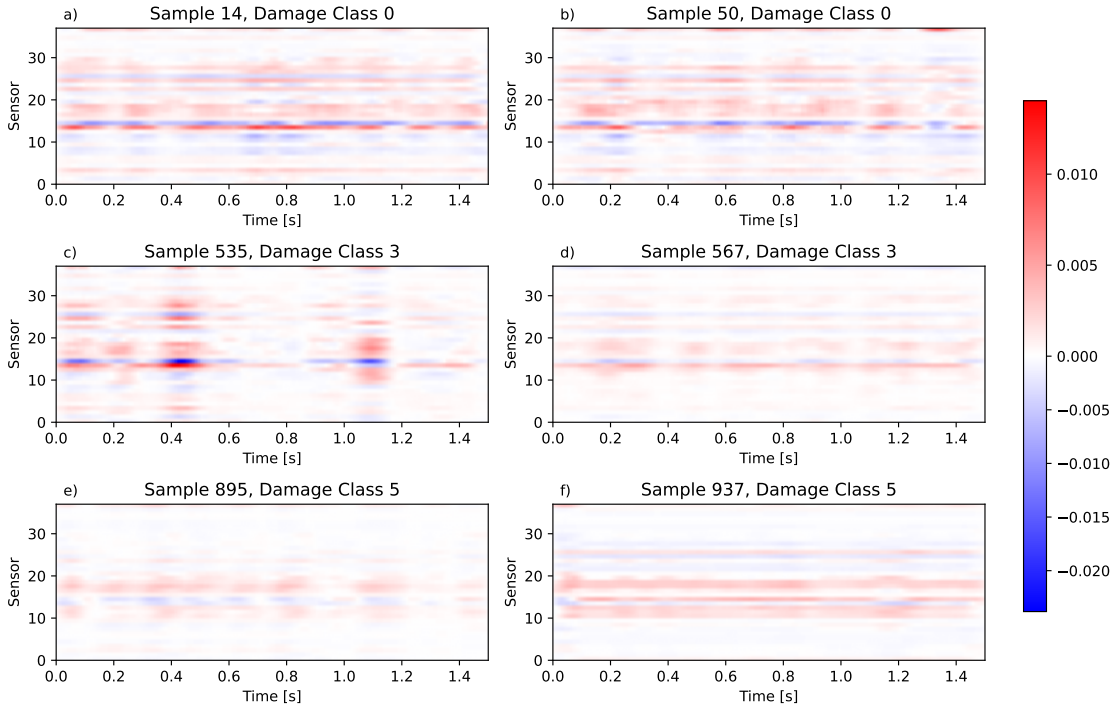


Figure 22: Overview of attribution maps based on the TVB for different structural states for TS 5 and 8° AoA. Panels a) and b) belong to damage class 0 (no damage), c) and d) to damage class 3 (crack length of 25% of the beam width) and e) and f) to damage class 5 (crack length of 50% of beam width).

7.5.6 Distribution of summed attribution vectors for 8° AoA

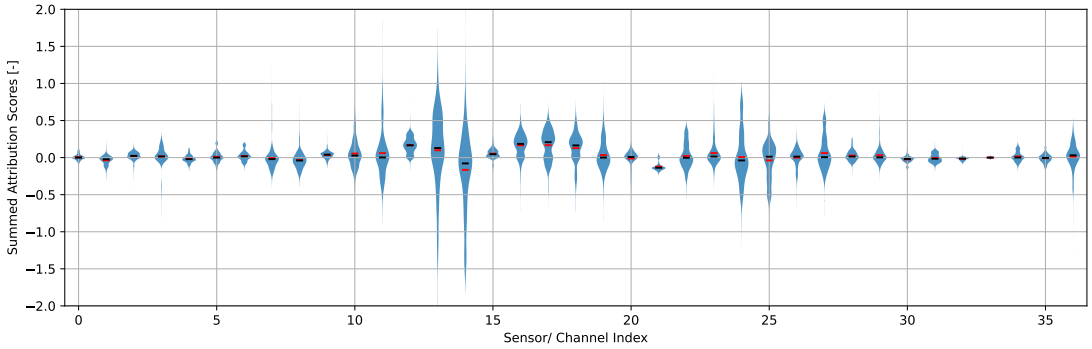


Figure 23: Distribution of summed attribution vectors for the TVB and 8° AoA, for all correctly classified samples of the validation set (red bar: mean value, black bar: median).

7.6 Plots for the mean value baseline

7.6.1 Attribution maps for 0° AoA and varying inflow conditions

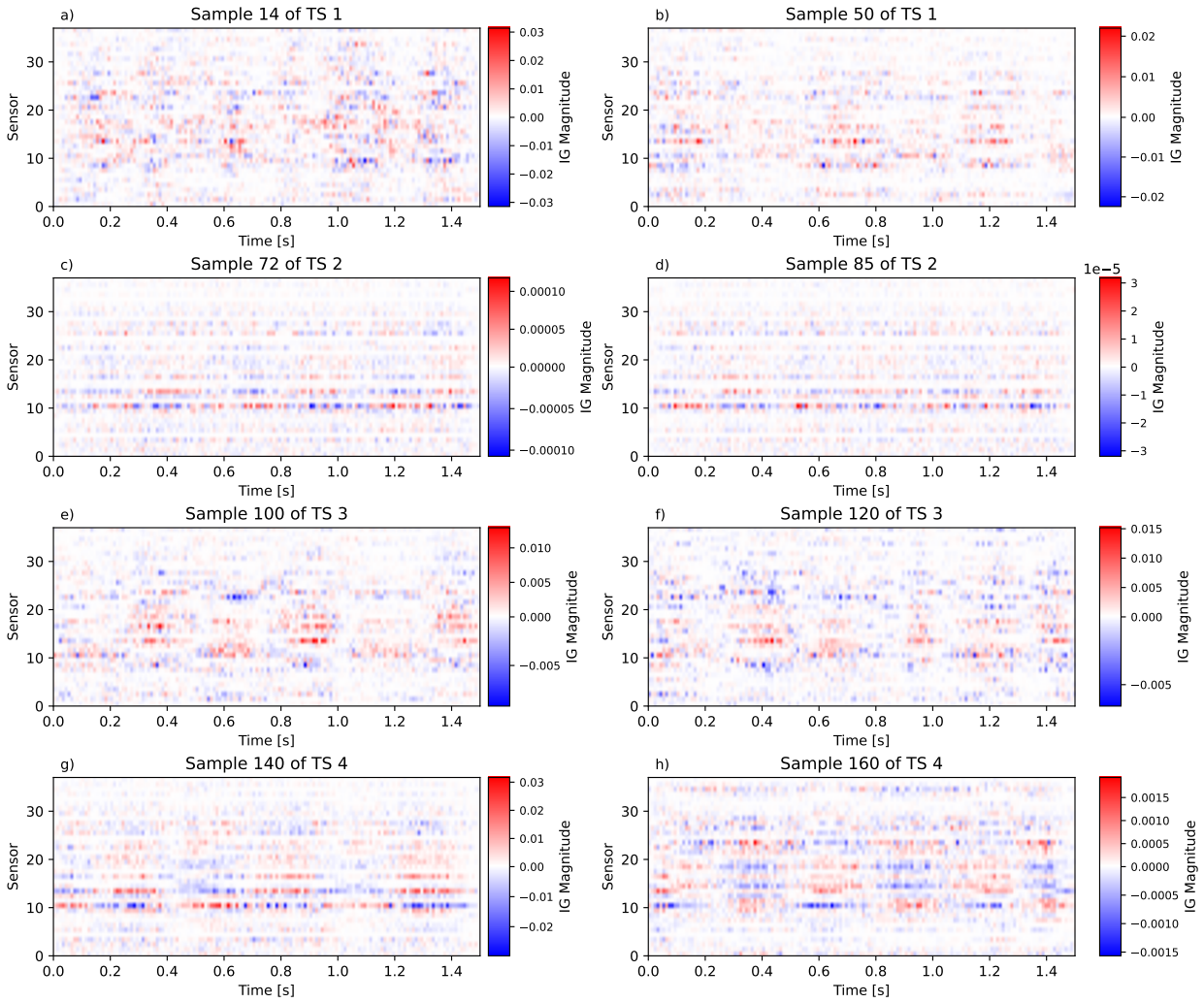


Figure 24: Overview of attribution maps based on the MVB for different inflow conditions for damage class 0 and 0° AoA. Panels a) and b) belong to TS 1 (see Table 1), c) and d) to TS 2, e) and f) to TS 3 and g) and h) to TS 4.

7.6.2 Attribution maps for 0° AoA and varying structural states

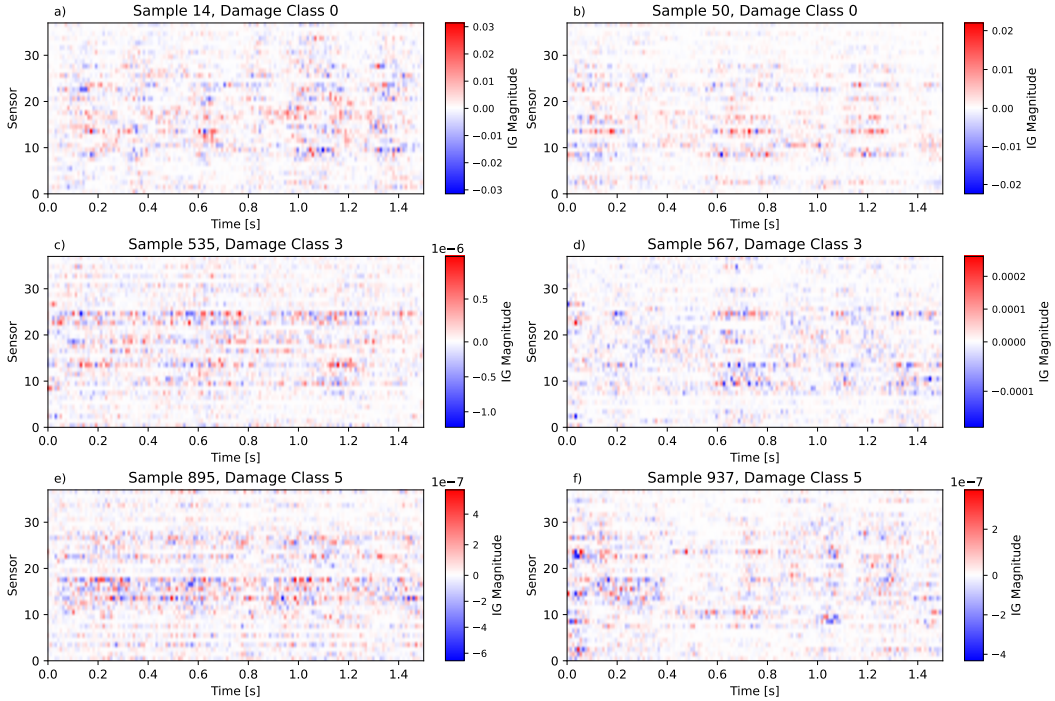


Figure 25: Overview of attribution maps based on the MVB for different structural states for TS 1 and 0° AoA. Panels a) and b) belong to damage class 0 (no damage), c) and d) to damage class 3 (crack length of 25% of the beam width) and e) and f) to damage class 5 (crack length of 50% of beam width).

7.6.3 Distribution of summed attribution vectors for 0° AoA

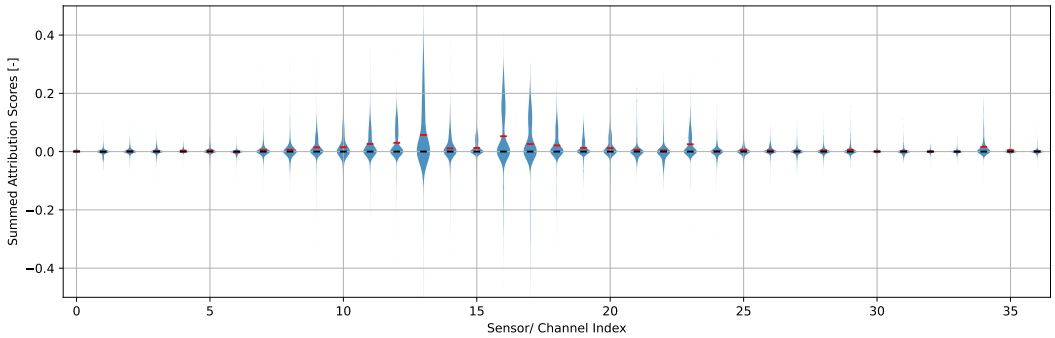


Figure 26: Distribution of summed attribution vectors for the MVB and 0° AoA, for all correctly classified samples of the validation set (red bar: mean value, black bar: median).

7.6.4 Attribution maps for 8° AoA and varying inflow conditions

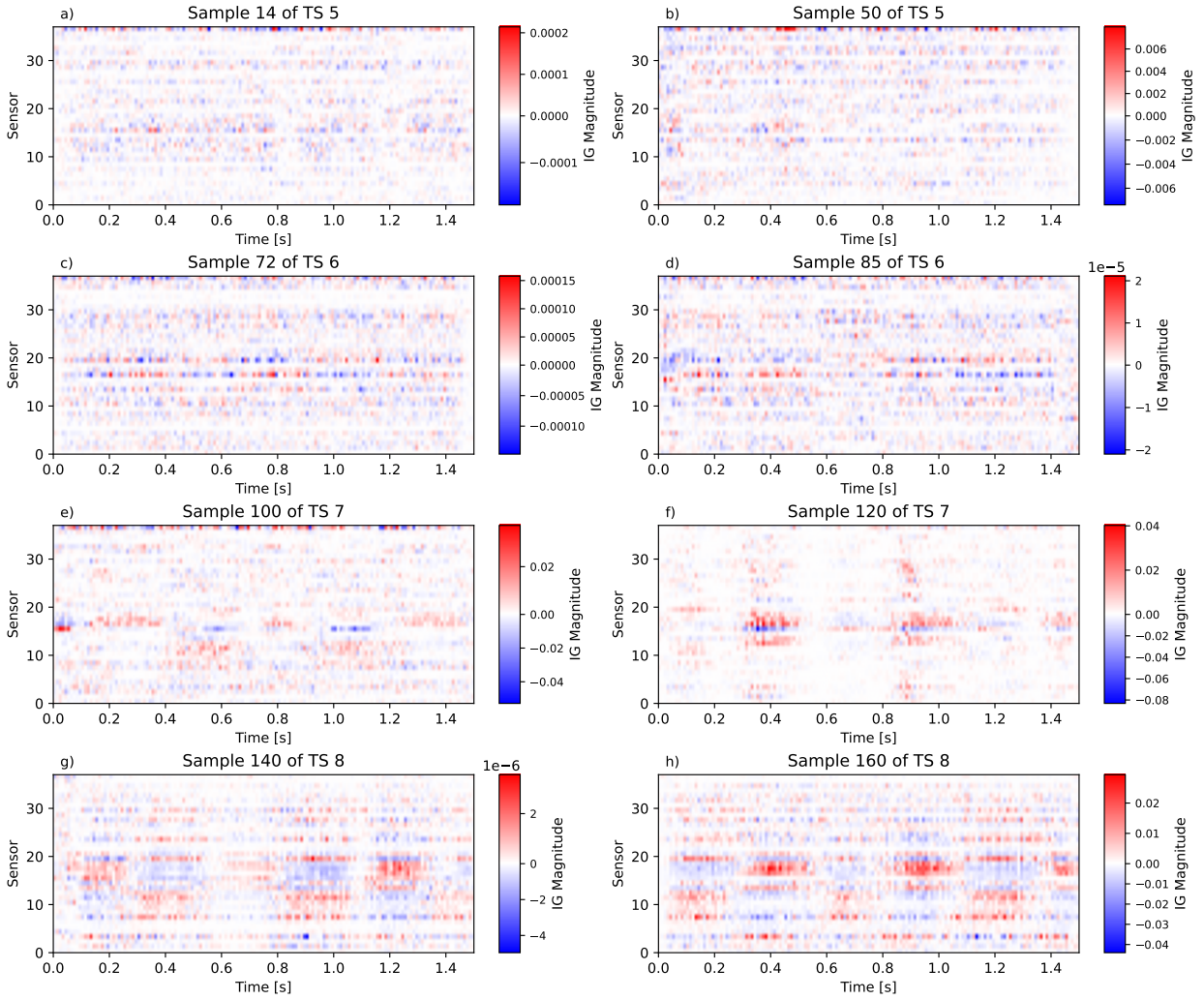


Figure 27: Overview of attribution maps based on the MVB for different inflow conditions for damage class 0 and 8° AoA. Panels a) and b) belong to TS 5 (see Table 1), c) and d) to TS 6, e) and f) to TS 7 and g) and h) to TS 8.

7.6.5 Attribution maps for 8° AoA and varying structural states

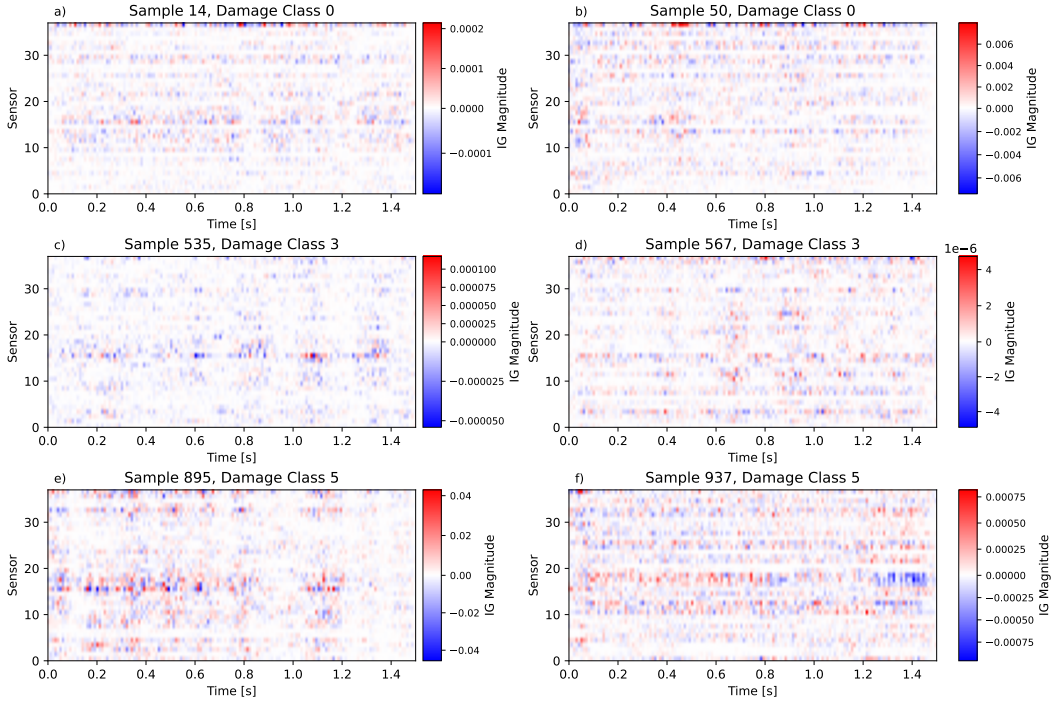


Figure 28: Overview of attribution maps based on the MVB for different structural states for TS 5 and 8° AoA. Panels a) and b) belong to damage class 0 (no damage), c) and d) to damage class 3 (crack length of 25% of the beam width) and e) and f) to damage class 5 (crack length of 50% of beam width).

7.6.6 Distribution of Summed Attribution Vectors for 8° AoA

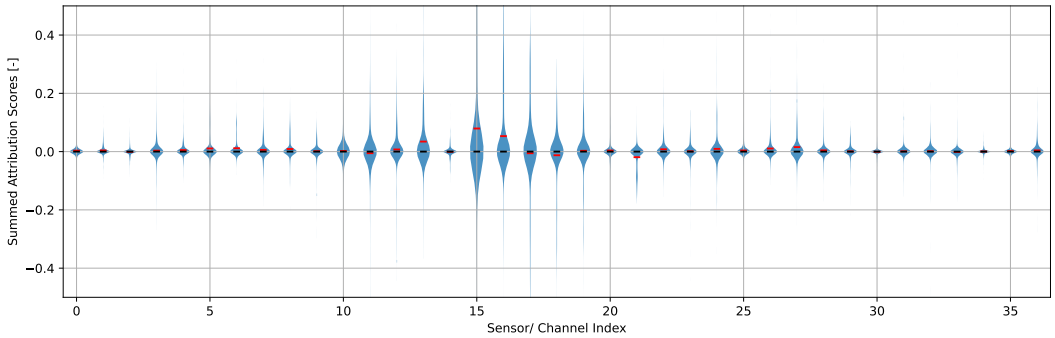


Figure 29: Distribution of summed attribution vectors for the MVB and 8° AoA, for all correctly classified samples of the validation set (red bar: mean value, black bar: median).

7.7 Training details of the MLP

Analogously to the CNN approach, the raw APM time series are first segmented into 1.5 s windows and normalized on a per-sample basis. For each window we compute the mean vector across time and subsequently the resulting vectors for the training, validation, and test set (splits identical to those used for the CNN) are z-normalized using the training set mean and standard deviation. The MLP comprises four fully-connected layers (widths: 128, 128, 64 and 6) that operate on the normalized mean vectors to produce the final class predictions. To improve generalization and speed up training, each hidden layer is followed by a batch normalization layer, which stabilizes the distribution of activations, and the input and the first two hidden layers are followed by dropout layers, which randomly deactivate a fraction of neurons. The hyperparameters were selected to obtain a stable training curve; no hyperparameter optimization was performed.

Table 6: Training hyperparameters for the MLP.

Parameter	**Value**
Optimizer	AdamW
Learning rate	1×10^{-3}
Weight decay	1×10^{-5}
Loss function	categorical crossentropy, with label smoothing (0.05)
Batch size	32
Maximum epochs	150
Callbacks on validation loss:	
ModelCheckpoint - save best only	
ReduceLROnPlateau - factor=0.5, patience=10, min lr=1e-5	
EarlyStopping - patience=12, restore best weights	

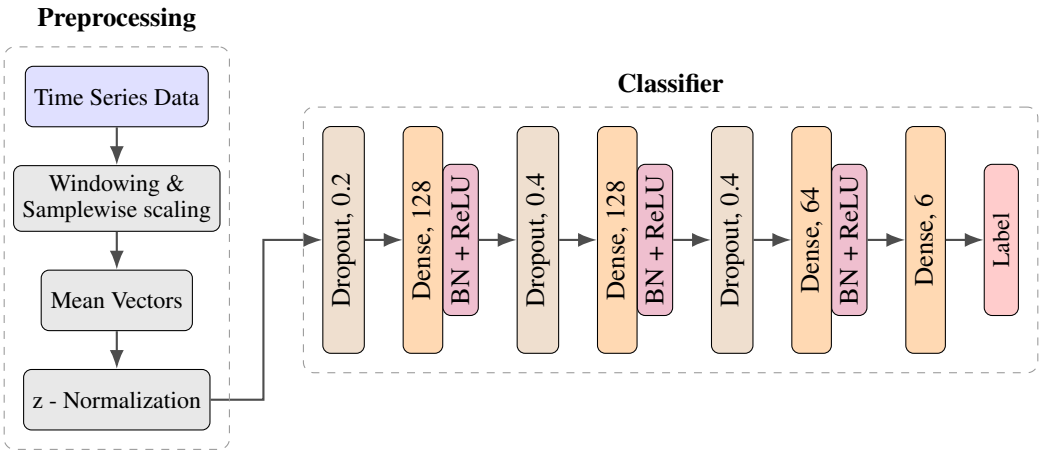


Figure 30: Preprocessing and structure for the mean vector MLP.