

BAYESIAN MACHINE LEARNING FOR THE PROGNOSIS OF COMBUSTION INSTABILITIES FROM NOISE

Ushnish Sengupta

Department of Engineering
University of Cambridge
Cambridge, Cambridgeshire CB21PZ
United Kingdom
Email: us271@cam.ac.uk

Carl E. Rasmussen

Computational and Biological Learning Lab
Department of Engineering
University of Cambridge
Cambridge, Cambridgeshire CB21PZ
United Kingdom
Email: cer54@cam.ac.uk

Matthew P. Juniper

Hopkinson Lab
Department of Engineering
University of Cambridge
Cambridge, Cambridgeshire CB21PZ
United Kingdom
Email: mpj1001@cam.ac.uk

ABSTRACT

Experiments are performed on a turbulent swirling flame placed inside a vertical tube whose fundamental acoustic mode becomes unstable at higher powers and equivalence ratios. The power, equivalence ratio, fuel composition and boundary condition of this tube are varied and, at each operating point, the combustion noise is recorded. In addition, short acoustic pulses at the fundamental frequency are supplied to the tube with a loudspeaker and the decay rates of subsequent acoustic oscillations are measured. This quantifies the linear stability of the system at every operating point. Using this data for training, we show that it is possible for a Bayesian ensemble of neural networks to predict the decay rate from a 300 millisecond sample of the (un-pulsed) combustion noise and therefore forecast impending thermoacoustic instabilities. We also show that it is possible to recover the equivalence ratio and power of the flame from these noise snippets, confirming our hypothesis that combustion noise indeed provides a fingerprint of the combustor's internal state. Furthermore, the Bayesian nature of our algorithm enables principled estimates of uncertainty in our predictions, a reassuring feature that prevents it from making overconfident extrapolations. We use the techniques of permutation importance and integrated gradients to understand which features in the combustion noise spectra are crucial for accurate predictions and how they might influence the prediction. This study serves as a first step towards establishing interpretable and Bayesian machine learning techniques as tools to discover informative relationships in combustor data and thereby build trustworthy, robust and reliable combustion diagnostics.

INTRODUCTION

Thermoacoustic instabilities, arising from the coupling between unsteady heat release rates and acoustic waves in combustors, are a persistent problem for gas turbine and rocket engine manufacturers. Heat release rate fluctuations at the flame create acoustic fluctuations, which then reflect off the boundaries, return to the flame and create more heat release fluctuations. This mechanism can set up a positive feedback loop—causing pressure fluctuations with progressively higher amplitudes and severe damage to the engine.

The phase lag between pressure and heat release rate fluctuations, which governs the thermoacoustic stability of a system, depends on acoustic, hydrodynamic and combustion mechanisms, which have different scaling behaviours. Accurate computational modeling is thus very challenging [1]. At the moment, designers accommodate thermoacoustic instabilities in their engines by avoiding unstable regions of the operating parameter space. This, however, conflicts with other design objectives such as reducing NO_x emissions by operating at leaner fuel-air ratios. The aim of this paper is to develop and test a machine learning algorithm that can learn how close a combustor is to instability and ensure its safe operation near unstable conditions.

Combustion noise as a diagnostic

The noise radiated by a turbulent combustor is generated by deterministic fluid dynamic phenomena such as unsteady dilatation due to fluctuating heat release rates or the acceleration of vorticity or entropy waves, and modified by acoustic reflections off boundaries [2]. We therefore expect pressure measurements

to contain some information about the state of the combustor. Inspired by Zelditch [3] who answered Kac’s whimsical question “Can you hear the shape of a drum?” [4] by proving that the eigenfrequencies of vibration determine the shape of an analytic and convex membrane uniquely, our study seeks to extract useful information about the state of a turbulent combustor from the power spectra of noise samples. This would have important practical implications. For example, it would allow noise to serve as an early warning prognostic for thermoacoustic instability or blowoff. It would also enable pressure measurements to validate the readings of other sensors such as flowmeters, making the system more robust to sensor failures. Pressure and vibration measurements are easily accessible in fielded combustors so it makes sense to use them as extensively as possible.

Historically, the motivation to understand and model combustion noise stems from the desire to reduce noise pollution, such as that from an aircraft [5] or a factory furnace [6]. Pioneering theoretical work by Lighthill in aeroacoustics [7] was extended to the analytical study of combustion noise by Strahle [8], who used Lighthill’s acoustic analogy to derive a formula relating the far-field acoustic perturbation to heat release rate fluctuations by treating the turbulent pre-mixed flame as an assembly of monopole sound sources. It was also noted [9] that noise can be generated by entropy inhomogeneities in regions of accelerating flow when combustion occurs in confined chambers. Aside from theoretical analysis, various empirical correlations that try to predict the overall noise level [10] or the spectral characteristics such as peak frequency, slope of rolloff in the high-frequency range, [11] etc. as a function of operating conditions were also obtained from experimental data. More recent studies have employed numerical simulations to predict combustion noise for open flames as well as complex geometries. A study by Ihme et al. [12] employs a model for predicting direct combustion noise generated by turbulent non-premixed flames where the Lighthill acoustic analogy is combined with a flamelet-based combustion model and incorporated into an LES simulation. Their predictions match well with experimental results although discrepancies were noted at high frequencies. The hybrid CHORUS method [5] predicts the noise output by performing LES of the combustion chamber, extracting the acoustic and entropy waves and then propagating these waves through the engine using analytical methods. Their results compare well with experiments.

The inverse problem of using noise to infer conditions inside the combustor is somewhat less well studied, although there has been a fair amount of research interest within the thermoacoustic community [1]. Simplifying the combustion noise generation process, Lieuwen [13] uses the decay rate of autocorrelation to determine the stability margin of a combustor. Several subsequent studies apply tools from nonlinear dynamics to combustion noise time-series and obtain useful precursors of instability. Gotoda et al. [14] employ the Wayland test for non-linear deter-

minism to show that when a system transitions to thermoacoustic instability, the combustion noise changes its character gradually from random and uncorrelated to completely deterministic. Similarly, Nair and colleagues [15] show the disappearance of the multifractal signature of combustion noise as it transitions to instability and note that measures such as the Hurst exponent can serve as an early warning of thermoacoustic instabilities. A follow-up study by Godavarthi et al [16] looks at measures derived from recurrence networks as instability precursors. Kobayashi et al [17] use a modified version of the permutation entropy to detect a precursor of the frequency-mode-shift in their staged aircraft engine model combustor before the amplification of pressure fluctuations. More recent work from the past year has also explored machine learning oriented approaches to the problem. Mondal et al [18] apply Hidden Markov Models to pressure timeseries for the early detection of instabilities in a Rijke tube, while Kobayashi et al [19] and Hachijo et al [20] combine Support Vector Machines with complex networks and statistical complexity measures, respectively, to do the same in a swirl-stabilized combustor.

The utilization of combustion noise for diagnostic purposes has not been limited solely to the forecasting of thermoacoustic instabilities. Acoustic precursors from combustion noise have been identified for lean blowout in a pre-mixed flame by Nair and Lieuwen [21] using the concentration of acoustic power in low-frequency bands, wavelet-filtered variance and thresholding techniques. Gotoda and co-workers [22] study the dynamics of pressure fluctuations near lean blowout using permutation entropy, fractal dimensions, and short-term predictability. Murayama et al. [23] have also used the weighted permutation entropy of combustion noise to develop precursors of blowout for their model gas turbine combustor.

The case for interpretable, Bayesian machine learning

A limitation of the approaches described above is that, by looking at the data through a handcrafted and predefined lens, one may miss other relevant information in the data. Machine learning techniques, on the other hand, find relevant functional relationships in the data without being influenced by researchers’ preconceptions. Deployed correctly, they also use all available information in the data. The downside of a purely data-driven approach, however, is that it is only applicable to the specific system that generated the data.

To address this problem of limited portability and the fact that acoustic emissions are an imperfect source of information, the Bayesian machine learning technique we employ in this study provides principled measures of uncertainty in our predictions. We start with an appropriately vague prior belief about what the output of our model should be and as we observe more data, we update this belief to obtain progressively tighter posterior distributions in accordance with Bayes’ rule. This ensures that

the model does not make overconfident predictions from out-of-distribution inputs which are entirely different from what it was trained on. These uncertainty estimates are particularly important when using a machine learning model in a critical device such as an aircraft engine. Bayesian machine learning techniques with correctly specified priors can also work with smaller amounts of data and are resistant to overfitting [24]. They can also be used in continual learning without catastrophic forgetting [25], which is particularly important if we want to keep learning from data throughout the operating lifetime of a device.

In this study, we use anchored ensembling, which is a simple and scalable way to train Bayesian neural networks [26]. First, we perform a set of experiments on a Rijke tube driven by a swirling premixed turbulent flame. The power, equivalence ratio, fuel composition and the exit area of the tube are all varied so that noise data can be collected over a wide range of operating parameters. The decay rates of oscillations provoked by acoustic pulses are measured. The thermoacoustic behaviour of the combustor ranges from very stable to almost unstable. The challenge for our neural network ensembles is then to predict the power, the equivalence ratio and the measured decay rate using only a 300 ms sample of the (un-forced) combustion noise as their input. We select this challenge because conditions in an engine can change rapidly and decisions should be based on only the most recent sensor data history. This is a high-dimensional regression problem to which our neural network ensembles are perfectly suited.

A common criticism of machine learning techniques is that they are black-box models that are completely opaque to the user. To remedy this, we have used a technique known as Integrated Gradients [27] to reveal features in the acoustic spectrum that drive the predictions from our Bayesian neural network ensembles.

EXPERIMENTAL SETUP

Figure 1 shows the experimental apparatus used in this study. An ordinary Bunsen burner is modified by attaching swirler vanes and a nozzle featuring a large central hole for the main flame and smaller surrounding holes for the pilot flames. A premixed mixture of methane and ethylene is used as fuel. This produces a noisy swirling premixed turbulent flame that is anchored over a wide range of operating conditions. The burner is placed inside a steel tube of length 800 mm and internal diameter 80 mm. Annular discs, with hole diameters 75, 65 and 55 mm, can be attached to the downstream end of the tube in order to change the acoustic boundary conditions.

The noise is recorded by a G.R.A.S. 26TK microphone placed near the bottom end of the tube. The raw pressure signal is sampled at 10000 Hz, which is considerably higher than the dominant frequencies in the typical noise spectra. Data acquisition is managed using a National Instruments BNC-2110 DAQ

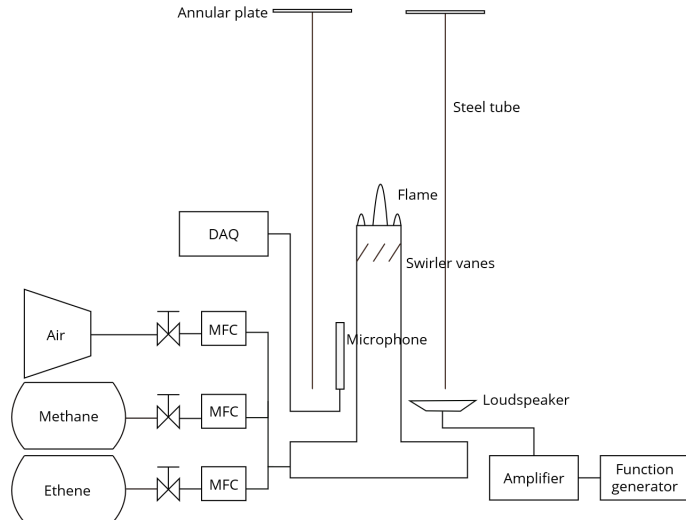


FIGURE 1. SCHEMATIC OF EXPERIMENTAL SETUP, CONSISTING OF A 1 KW TURBULENT SWIRL FLAME INSIDE A STEEL TUBE OF LENGTH 800 MM AND INTERNAL DIAMETER 80 MM

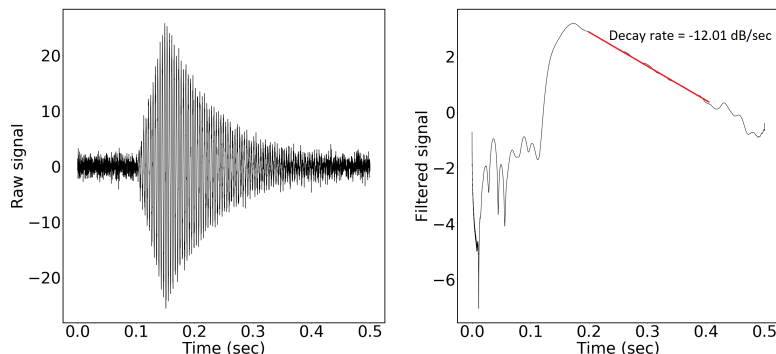


FIGURE 2. LEFT: TURBULENT COMBUSTOR EXCITED BY AN ACOUSTIC PULSE RIGHT: FILTERED SIGNAL WITH BEST-FIT LINE IN RED

device and the software LabVIEW. Flow rates for fuel and air are controlled using Bronkhorst EL-FLOW® Select flowmeters. A 70 W VISATON 3020 BG loudspeaker is placed near the base of the burner to supply acoustic pulses.

The system is operated at 900 different combinations of operating parameters which form a grid in our 4-dimensional operating parameter space (volumetric flow rate, equivalence ratio, fuel composition and outlet boundary condition). Experiments are performed at methane:ethene ratios of 3:4, 1:1 and 5:4 (v/v) and tube outlet diameters 80, 75 and 65 mm. For each of the 3 methane-ethene ratios and each of the 3 outlet boundary conditions, we methodically sweep through 100 different fuel and air mass flow rates. Figure 3 shows the 100 pairs of equivalence

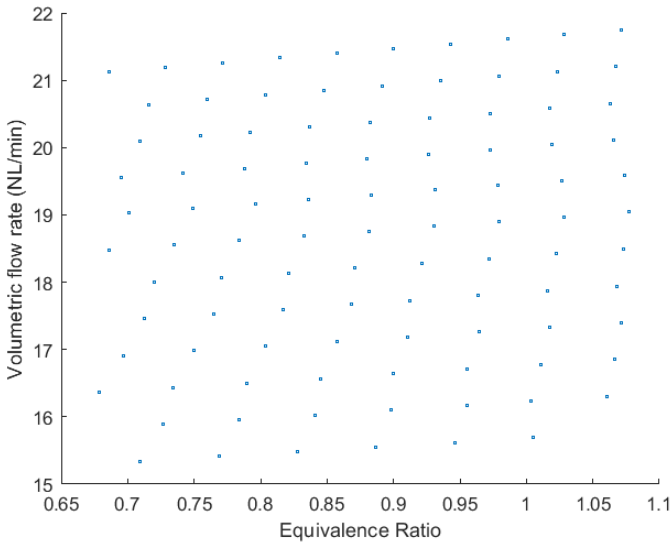


FIGURE 3. EQUIVALENCE RATIOS AND FLOW RATES OF DATAPOINTS FOR METHANE:ETHENE RATIO OF 3:4 AND OUTLET DIAMETER 80 MM

ratios and volumetric flow rates at which experiments are performed while the methane:ethene ratio and outlet diameter remained fixed at 0.75 and 80 mm, respectively. While this figure represents a 2D slice of the entire dataset—similar, but not identical, equivalence ratios and flow rates were achieved for the other fuel compositions and boundary conditions. These experiments mimic, in a laboratory setting, the multidimensional nature of the operating parameter space in a real jet engine where the boundary condition is typically dynamic and the engine controller has the authority to change multiple quantities such as fuel split, power, fuel inlet pressure, equivalence ratio, core speed and others. Any early warning signal needs to function over the whole range of operating parameters, not just when a single parameter is varied.

For each combination of operating parameters, the combustion noise is recorded and the decay rate of a 50 millisecond-long acoustic pulse at 230 Hz (the fundamental acoustic frequency of the system) is obtained. To extract the decay rate, the microphone signal is processed in a manner similar to Schumm et al [28]. First, a Butterworth filter, with a width of 20 Hz and centered at the excitation frequency 230 Hz, is used to filter out the undesired frequencies. Then a Hilbert transform is applied to obtain the instantaneous amplitude, $A(t)$, of the pressure signal. When the logarithm of the obtained amplitude is plotted against time, it is possible to identify a linear region corresponding to exponential decay. To isolate the linear region, we ignore 50 ms of data immediately after the ping. The noise floor is computed from the RMS value of the pre-pulse signal and the decaying signal is cut-off when it decays to twice this value. The slope of this region then corresponds to the decay rate of the oscillations.

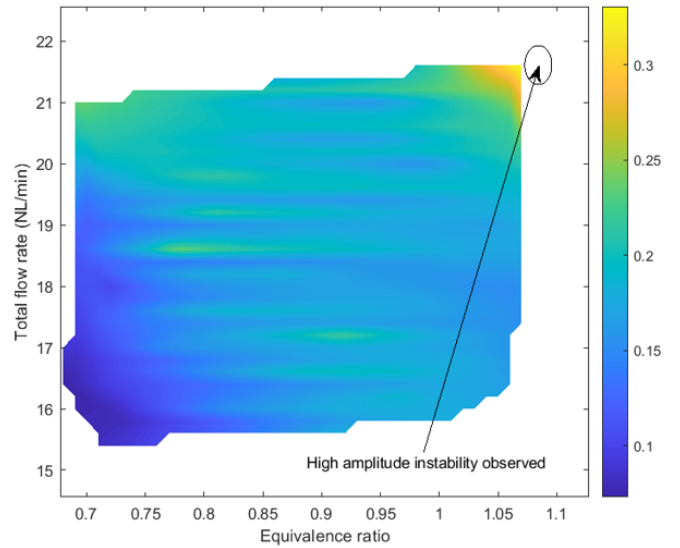


FIGURE 4. CONTOUR PLOT OF DECAY TIMESCALE (SECONDS) FOR METHANE:ETHENE RATIO OF 3:4 AND OUTLET DIAMETER 80 MM

We use the measured decay rate (or its negative inverse, the decay timescale) at each operating point as a proxy for the thermoacoustic stability at that point. In general, decay rates or timescales quantify the linear stability of a system, which is a necessary but not sufficient condition for global thermoacoustic stability. However, for the high-amplitude instability we want to avoid, the linear stability boundary is observed to characterize the onset quite well. In Figure 4, we show a plot of the decay rate as a function of flow rate and equivalence ratio, for the same boundary condition and fuel composition as in Figure 3. (outlet diameter = 80 mm, methane-ethene ratio = 0.75) We observe that the decay rate only reaches values close to 0 and decay timescales reach their highest values (0.35 seconds for this particular subset of the data) in the vicinity of the high-amplitude 230 Hz instabilities. This holds true for the other 8 combinations of boundary conditions and fuel compositions as well. Therefore, if a diagnostic tool is well-correlated with the decay rate or timescale, it will be able to warn us when we are too close to the instability.

STATISTICAL TOOLS

Precursors of thermoacoustic instability from the literature

Nair and colleagues [15] suggest that there is a loss of multifractality in combustion noise as combustors progress towards combustion instability, which is reflected in a decline of the signal's second-order generalized Hurst exponent H_2 prior to an instability. The generalized Hurst exponent H_q is defined by the

scaling of the q -th order moments of the signal. For a time series $g(t)$,

$$\frac{\langle |g(t+\tau) - g(t)|^q \rangle_t}{\langle |g(t)|^q \rangle_t} \sim \tau^{qH_q} \quad (1)$$

where $q > 0$, τ is the lag and the averaging, denoted by $\langle \cdot \rangle_t$ is over the window $t \gg \tau$. H_q can be computed simply from the above expression by applying a logarithmic transformation to both sides and obtaining the slope of the least squares fitted line for a suitable scaling window of lag times. For our study, we calculate the second-order generalized Hurst exponent using 1 second slices from the dynamic pressure sensor data (10000 data points) and a scaling window between 0.01 and 0.02 seconds, which corresponds to approximately two to four cycles of oscillations at combustion instability. Basic sanity checks have been performed and it is found that a synthetic Gaussian white noise signal has a H_2 close to 0.5 while a synthetic periodic signal at the instability frequency has a H_2 close to 0, as expected.

Lieuwen [13] derived an effective damping coefficient ξ_i for a thermoacoustic mode in terms of the decay rate ξ_i of the autocorrelation $C_i(\tau)$ of the i -th acoustic mode $\eta_i(t)$, assuming the combustor to be a second-order oscillator, the background noise to be spectrally flat and parametric disturbances to be absent:

$$C_i(\tau) = e^{-\omega_i \xi_i \tau} \left(\cos \left(\omega_i \tau \sqrt{1 - \xi_i^2} \right) + \frac{\xi_i}{\sqrt{1 - \xi_i^2}} \sin \left(\omega_i \tau \sqrt{1 - \xi_i^2} \right) \right) \quad (2)$$

The autocorrelation decay rate is calculated the same way as the decay rate of the acoustic pulses. A 1-second long sample of the raw pressure signal is first put through a Butterworth filter centered around the frequency of interest (here, 230 Hz) to obtain the signal $\eta_i(t)$. The autocorrelation $C_i(\tau)$ is then obtained as a function of the lag time τ and its envelope determined through the Hilbert transform. The slope of the least-squares fitted line through the logarithm of the autocorrelation amplitude then gives us the desired effective damping coefficient ξ_i . We expect this quantity to tend towards zero as our system approaches instability.

Bayesian neural network ensembles

The Bayesian approach to training neural networks [29] entails placing a sensible prior probability distribution over the parameters of the network and inferring the posterior distribution over parameters using the observed data and Bayes' rule. Training a Bayesian neural network can be technically challenging and computationally expensive. Exact inference is intractable and the gold-standard technique to integrate over the posterior,

Markov Chain Monte Carlo, can be inefficient. Researchers often resort to variational approximations of the true posterior [30], parametrizing the posterior and minimizing the Kullback-Leibler divergence between this variational distribution and the true posterior during training. However, while computationally cheap, mean-field variational inference too has its drawbacks such as not maintaining correlations between parameters. Recently, a different method to train Bayesian neural networks, based on ensembling, has been proposed. This new method is cheap, simple, scalable and yet manages to outperform variational inference in several uncertainty quantification benchmarks [26]. Consider a data set (\mathbf{x}_n, y_n) , where each data point consists of features $\mathbf{x}_n \in \mathbb{R}^D$ and output $y_n \in \mathbb{R}$. Define the likelihood for each data point as $p(y_n | \theta, \mathbf{x}_n, \sigma_\epsilon^2) = \mathcal{N}(y_n | \text{NN}(\mathbf{x}_n; \theta), \sigma_\epsilon^2)$, where NN is a neural network whose weights and biases form the latent variables θ while σ_ϵ^2 is the data noise. Define the prior on the weights and biases θ to be the standard normal $p(\theta) = \mathcal{N}(\theta | \mu_{\text{prior}}, \Sigma_{\text{prior}})$. The anchored ensembling algorithm then does the following:

1. The parameters $\theta_{0,j}$ of each j -th member of our neural network ensemble are initialized by drawing from the prior distribution $\mathcal{N}(\mu_{\text{prior}}, \Sigma_{\text{prior}})$.
2. Each ensemble member is trained ordinarily (e.g. using stochastic gradient descent) but with a slightly modified loss function that anchors the parameters to their initial values. The loss function for the j -th ensemble member is given by $\text{Loss}_{\text{anchor},j} = \frac{1}{N} \|\mathbf{y} - \hat{\mathbf{y}}_j\|_2^2 + \|\Sigma^{1/2}(\theta - \theta_{0,j})\|_2^2$, where the i -th diagonal element of Σ is the ratio of data noise to the prior variance of the i -th parameter.

Pearce et al. [26] prove that this procedure approximates the true posterior distribution for wide neural networks. In our study, we train an ensemble of 10 two-layer neural networks with 25 nodes in each layer and ReLU activation. The input to our network is the 1501-dimensional power spectrum obtained through a Fourier transform of the 300 ms noise sample. The outputs are the decay time scale (the negative inverse of the measured decay rate), the equivalence ratio, and power. Before training, all the input variables and outputs are normalized using a min-max strategy to lie between -1 and 1. Noise samples from 180 randomly chosen operating parameter combinations (20% of data) are held out in the test set for evaluating the performance of the model. 10-fold cross-validation is performed where 10 different models are trained using 10 random train-test splits. This ensures the stability of our algorithm's performance with respect to different train-test splits. Each ensemble member is trained using the stochastic gradient descent optimizer ADAM [31]. The tunable hyperparameters of our model such as the learning rate, the data noise and the number of nodes in each layer are optimized by minimizing the negative log-likelihood of data in a validation set.

Interpretation using Integrated Gradients and Permutation Importance

To attribute the predictions of our network ensemble to the input features we use the technique of integrated gradients [27]. This is a simple scalable method that does not need any instrumentation of the network and can be computed easily using a few calls to the gradient operation. For deep neural networks, gradients of the output with respect to the input is a natural analog of the linear regression coefficients. In a linear model, the coefficients characterize the relationship between input and output variables globally. In a nonlinear neural network, however, a gradient at a point merely characterizes the local relationship between a predictor variable and the output. The main idea behind integrated gradients is to compute the path integral of the gradient of outputs with respect to the inputs from a baseline input to the input at hand. For an image recognition algorithm, a completely black image could be a reasonable choice of baseline while for a regression problem like ours where the input variables were normalized to lie between -1 and 1 , an input of all zeros is a reasonable choice. We consider the straight-line path (in \mathbb{R}^n) from the baseline x' to the input x , and compute the gradients at all points along the path. The integrated gradient along the i -th dimension is then defined as follows.

$$\text{IntegratedGrads}_i = (x_i - x'_i) \int_{\alpha=0}^1 \frac{\delta F(x' + \alpha(x - x'))}{\delta x_i} d\alpha \quad (3)$$

This integral is computed numerically. Accumulating the gradients with a path integral ensures that we estimate the aggregate influence of each predictor variable over the outputs. We use an implementation of Integrated Gradients from the DeepExplain library [32].

Understanding network predictions via Integrated Gradients involves looking at individual examples and their attribution plots. To gain a global overview of the relative importance of input features, we also use the permutation importance technique [33]. To compute permutation feature importances, we shuffle the values of a set of features between samples in the dataset and measure the impact of randomizing them on the accuracy of a trained model. The average error of model predictions should increase significantly if important features are randomized in this way and the decrease in accuracy can therefore be understood as a measure of a feature’s criticality. For our data, we randomize contiguous 33 Hz blocks in the input spectra of the test data and measure the percentage increase of the Root Mean Squared Error.

The code and data used to produce the results in this paper are available as a Google Colab notebook in the Github repository <https://github.com/Ushnish-Sengupta/FYR>. The notebook can be run in the web browser and does not require the installation of any software.

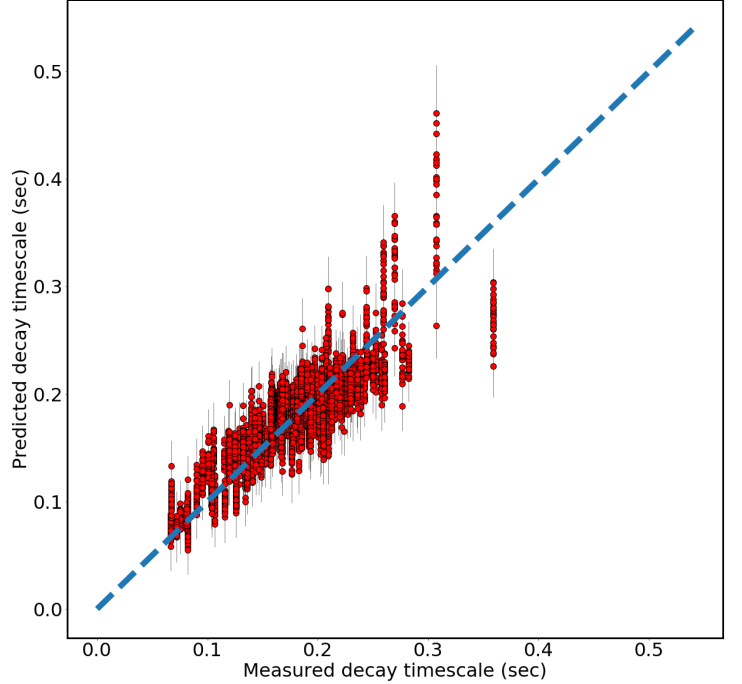


FIGURE 5. PLOT OF MEASURED DECAY TIMESCALES VS DECAY TIMESCALES PREDICTED BY THE NEURAL NETWORK ENSEMBLE ON THE TEST DATA

RESULTS

We train an ensemble of neural networks that takes the power spectrum of a noise sample as input and predicts the negative inverse of the measured decay rate (the decay timescale). Figure 5 shows the performance of the ensemble on the test dataset, where we observe that the decay timescales are predicted reasonably accurately. In the course of our 10-fold cross-validation, the root mean squared prediction error ranged from 0.023 seconds to 0.027 seconds, indicating that our algorithm is stable to variations in the training-test split. It is particularly interesting to note how the grey errorbars (± 1 S.D.) widen for the operating points closer to instability, corresponding to larger decay timescales. This is because there are comparatively fewer data points close to instability (only 13 operating conditions in the training dataset have a decay timescale exceeding 0.3 seconds) making the ensemble less certain about its predictions in that region. This demonstrates how principled uncertainties prevent blind overconfidence in our machine learning models. Nevertheless, even with the slightly larger uncertainties and prediction errors for those points, this algorithm can clearly indicate when the system approaches thermoacoustic instability.

We also trained our ensembles to recognize the equivalence ratio and burner power from a noise sample. Figures 6 and 7 show the measured values of these state variables plotted against the predictions of our ensembles and reveal an even more accu-

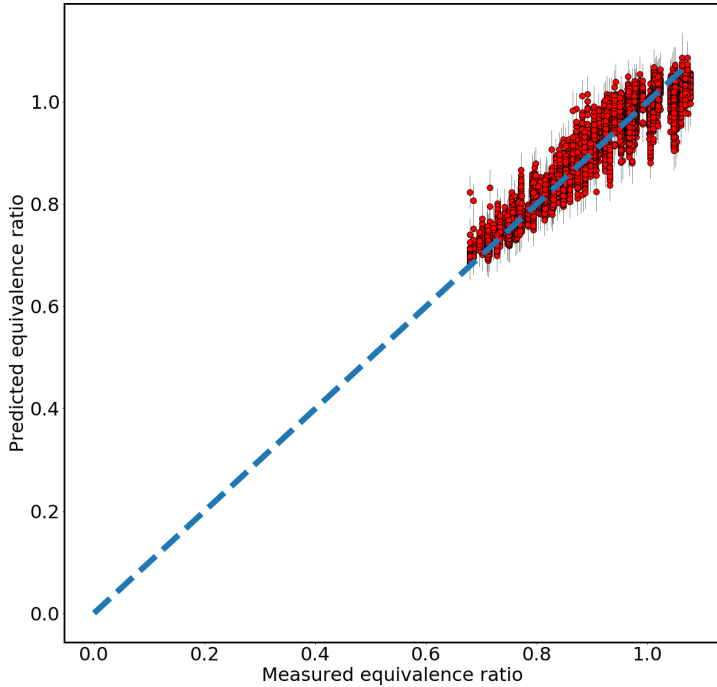


FIGURE 6. PLOT OF MEASURED EQUIVALENCE RATIO VS EQUIVALENCE RATIO PREDICTED BY THE NEURAL NETWORK ENSEMBLE ON THE TEST DATA

rate prediction. The root mean squared error in equivalence ratio prediction ranged between 0.031 to 0.033 ($\approx 3.5\%$), while the error for the power varied from 0.021 kW to 0.025 kW ($\approx 2\%$). The neural networks were thus able to predict these two important state variables quite accurately given a short noise sample. Each operating condition seem to have a unique acoustic signature which the machine learning algorithm can learn. In other words, one can indeed hear the state of a combustor.

For each ensemble of trained neural networks, we use the technique of integrated gradients to produce feature-level attribution plots, which can tell us how much a particular predictor influenced the prediction for a particular input. For example, the attribution plot in Figure 8 shows this technique applied to the decay rate prediction ensemble for an input with a decay timescale of 0.35 seconds. The most prominent feature is the large positive attributions for the frequency component around the fundamental, which is marked in Figure 8 as a grey line labeled 1f. The model has observed that an increased concentration of acoustic power around the fundamental frequency can be an indication that the system is close to instability. It is intriguing, however, that almost all of the predictor variables seem to make meaningful contributions to the final prediction and that if this information were removed, it would deteriorate the accuracy of predictions. This implies that for diagnostics to have higher predictive powers, the data should be considered in its entirety and

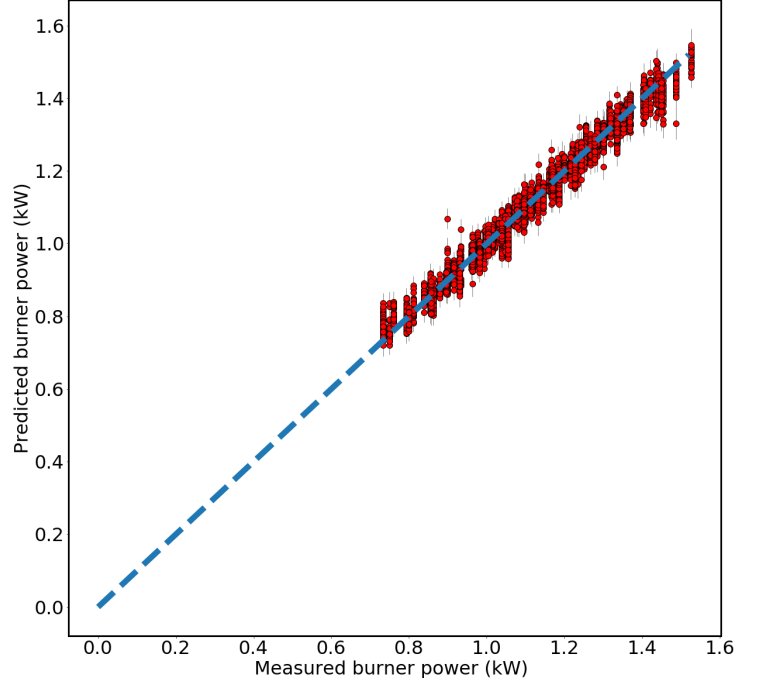


FIGURE 7. PLOT OF MEASURED BURNER POWER VS BURNER POWER PREDICTED BY THE NEURAL NETWORK ENSEMBLE ON THE TEST DATA

not examined through a pre-defined lens.

Our argument in favour of considering information from the entirety of the combustion noise spectrum is bolstered by the permutation feature importance plot (Figure 9). Here, a larger increase in the Root Mean Squared Error when a set of features is randomized indicates a strong dependence of the model on these features. We observe that the higher frequency portions of the input spectra have unique information that is independent of other parts of the spectra and randomizing them increases the RMSE on the test data by more than 100 percentage, in some cases. The same does not appear true for the lower frequency bands.

To compare our technique to those in the literature, the Hurst exponent and the decay of autocorrelation amplitude are computed for noise samples from each operating condition. While the Hurst exponent has a rough negative correlation with the measured decay rates, falling as the decay timescales grow, the relationship is very noisy (Figure 10). This means that if the Hurst exponent were to be used to forecast instabilities for our combustors, there would be many false alarms, which is clearly undesirable.

The same is found for the autocorrelation decay (see Figure 11), which also becomes close to zero, as expected, at the edge of instability. However, it also approaches zero for several operating points at which the combustor is very stable, making it somewhat unreliable as a prognostic for instability. While these

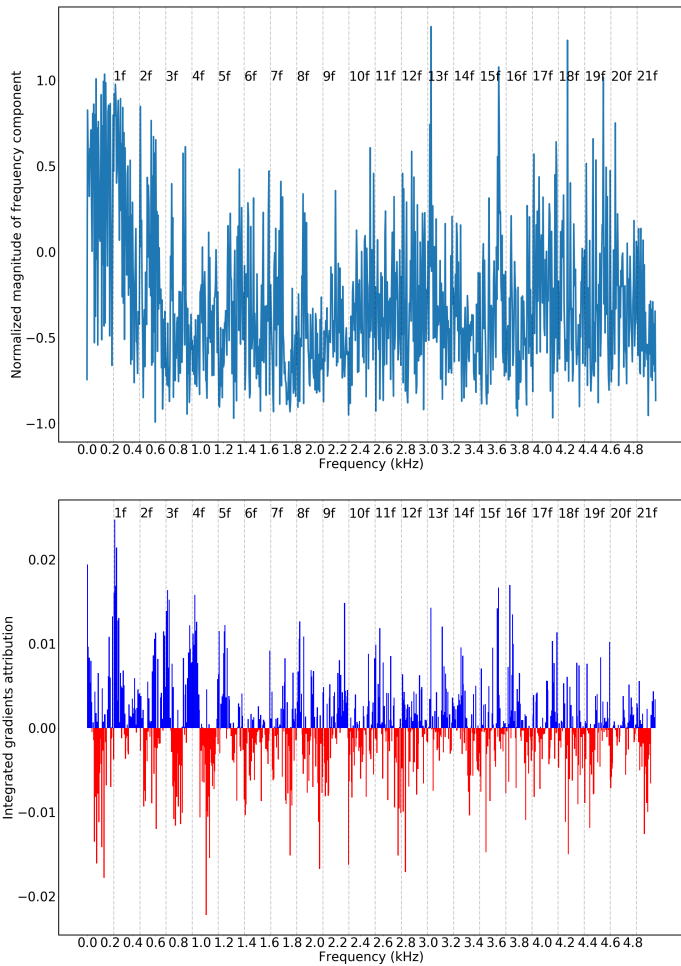


FIGURE 8. TOP: NORMALIZED POWER SPECTRUM OF A NOISE SAMPLE CORRESPONDING TO A 0.35 SECOND DECAY TIMESCALE (CLOSE TO INSTABILITY) BOTTOM: INTEGRATED GRADIENTS ATTRIBUTION PLOT FOR THIS POWER SPECTRUM

measures are attractive because they do not need to be trained on extensive experimental data, they also seem to have limited predictive power in our experiments.

CONCLUSIONS

We demonstrate that Bayesian ensembles of neural networks, a probabilistic machine learning algorithm, can be used to model relationships between measured combustion noise and the stability margins or operating conditions of a lab-scale turbulent combustor. We can estimate the decay rate of acoustic pulses, the power and the equivalence ratio from a single 300 millisecond sample of the combustor’s radiated noise over a wide range of operating conditions. Not only are these estimates rea-

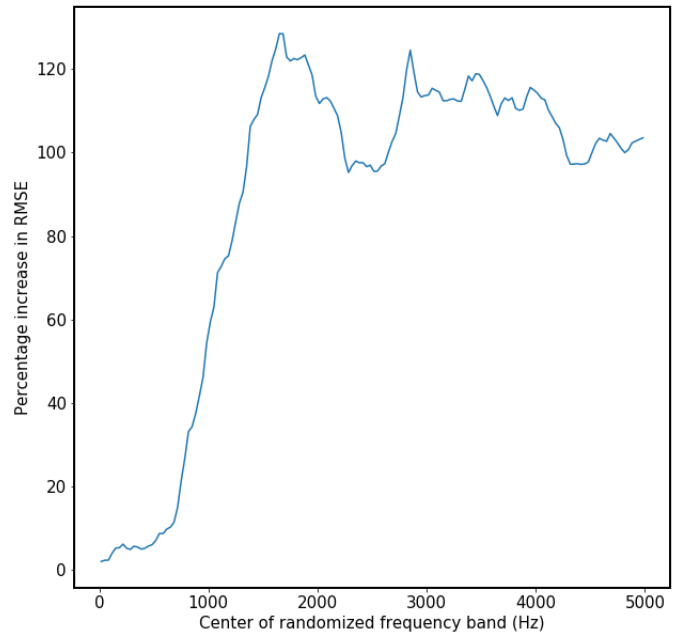


FIGURE 9. PERMUTATION FEATURE IMPORTANCE PLOT FOR DECAY TIMESCALE PREDICTION

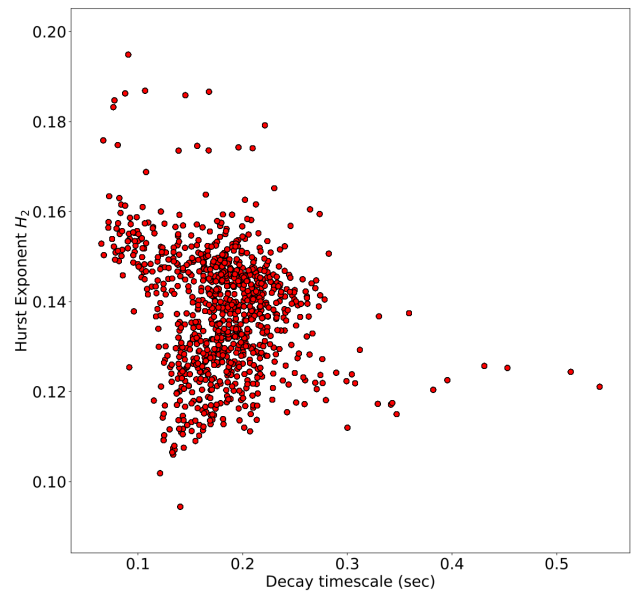


FIGURE 10. PLOT OF GENERALIZED HURST EXPONENT H_2 VS DECAY TIMESCALE

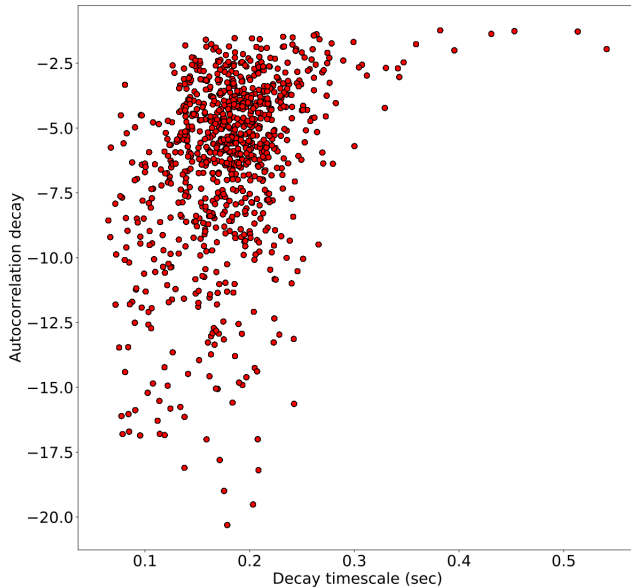


FIGURE 11. PLOT OF AUTOCORRELATION DECAY VS DECAY TIMESCALE

sonably accurate, but they contain principled estimates of uncertainty. This means that the model has the wisdom to “know what it doesn’t know” and this is reflected in higher uncertainties when a provided input is too different from those on which it has been trained. With Integrated Gradients, we discern which features in the input spectra drive the networks towards a particular prediction, making our technique interpretable. We compare our approach with two precursors of thermoacoustic instability from the literature: the Hurst exponent and the autocorrelation decay. While these broadly behave as expected, their relationship to the measured decay rates was noisy.

This study shows that a Bayesian ensemble of neural networks trained on a particular combustor is better at discerning the onset of thermoacoustic instability than are traditional measures based on the Hurst exponent and the autocorrelation decay. This shows that important information is lost when data is filtered through the traditional methods. The good agreement between predicted and measured values shown in Figures 5, 6 and 7 shows that each operating point has a distinctive noise, which can be learned by the ensemble. It is worth mentioning, of course, that this model is machine-specific and will not generalize to a different machine. In an industrial setting, however, this drawback is mitigated by the fact that there are only a few combustor designs and a lot of data for each combustor.

This work highlights the promise of building robust real-

time early warning systems for instabilities such as thermoacoustic oscillations using combustion noise data. It also shows how combustion noise can predict the power and equivalence ratio of the combustor and thus validate sensor measurements without any additional investment in hardware. We plan to build on this work and apply these tools to data from larger scale and more industrially relevant systems such as an annular combustor. We also realize that our Bayesian machine learning techniques may be used to fuse data from multiple sensors, not just a single pressure measurement, to build even more informative diagnostics. Finally, we seek to address the issue of machine-specificity in our machine learning models by using techniques from transfer learning to build machine-invariant diagnostic tools, which will generalize better to new devices.

ACKNOWLEDGMENT

We thank Mark Garner and Roy Slater for helping us manufacture the burner and the annular plates. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement number 766264.

REFERENCES

- [1] Juniper, M. P., and Sujith, R., 2018. “Sensitivity and Non-linearity of Thermoacoustic Oscillations”. *Annual Review of Fluid Mechanics*.
- [2] Dowling, A. P., and Mahmoudi, Y., 2015. “Combustion noise”. *Proceedings of the Combustion Institute*.
- [3] Zelditch, S., 2000. “Spectral determination of analytic bi-axisymmetric plane domains”. *Geometric and Functional Analysis*, **10**(3), sep, pp. 628–677.
- [4] Kac, M., 1966. “Can One Hear the Shape of a Drum?”. *The American Mathematical Monthly*.
- [5] Duran, I., Moreau, S., Nicoud, F., Livebardon, T., Bouty, E., and Poinsot, T., 2014. “Combustion noise in modern aero-engines”. *AerospaceLab Journal*.
- [6] Putnam, A. A., 2009. “Combustion Roar as Observed in Industrial Furnaces”. *Journal of Engineering for Power*.
- [7] Lighthill, M. J., 1961. “Sound generated aerodynamically”. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*.
- [8] Strahle, W. C., 1971. “On combustion generated noise”. *Journal of Fluid Mechanics*.
- [9] Marble, F. E., and Candel, S. M., 1977. “Acoustic disturbance from gas non-uniformities convected through a nozzle”. *Journal of Sound and Vibration*.
- [10] Strahle, W. C., and Shivashankara, B. N., 1975. “A rational correlation of combustion noise results from open turbulent premixed flames”. *Symposium (International) on Combustion*.

- [11] Abugov, D. I., and Obrezkov, O. I., 1978. “Acoustic noise in turbulent flames”. *Combustion, Explosion, and Shock Waves*.
- [12] Ihme, M., and Pitsch, H., 2012. “On the Generation of Direct Combustion Noise in Turbulent Non-Premixed Flames”. *International Journal of Aeroacoustics*, **11**(1), mar, pp. 25–78.
- [13] Lieuwen, T., 2005. “Online Combustor Stability Margin Assessment Using Dynamic Pressure Data”. *Journal of Engineering for Gas Turbines and Power*.
- [14] Gotoda, H., Nikimoto, H., Miyano, T., and Tachibana, S., 2011. “Dynamic properties of combustion instability in a lean premixed gas-turbine combustor”. *Chaos*.
- [15] Nair, V., and Sujith, R. I., 2014. “Multifractality in combustion noise: predicting an impending combustion instability”. *Journal of Fluid Mechanics*, **747**, may, pp. 635–655.
- [16] Godavarthi, V., Unni, V. R., Gopalakrishnan, E. A., and Sujith, R. I., 2017. “Recurrence networks to study dynamical transitions in a turbulent combustor”. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **27**(6), p. 063113.
- [17] Kobayashi, H., Gotoda, H., Tachibana, S., and Yoshida, S., 2017. “Detection of frequency-mode-shift during thermoacoustic combustion oscillations in a staged aircraft engine model combustor”. *Journal of Applied Physics*, **122**(22), p. 224904.
- [18] Mondal, S., Ghalyan, N. F., Ray, A., and Mukhopadhyay, A., 2019. “Early detection of thermoacoustic instabilities using hidden markov models”. *Combustion Science and Technology*, **191**(8), pp. 1309–1336.
- [19] Kobayashi, T., Murayama, S., Hachijo, T., and Gotoda, H., 2019. “Early detection of thermoacoustic combustion instability using a methodology combining complex networks and machine learning”. *Phys. Rev. Applied*, **11**, Jun, p. 064034.
- [20] Hachijo, T., Masuda, S., Kurosaka, T., and Gotoda, H., 2019. “Early detection of thermoacoustic combustion oscillations using a methodology combining statistical complexity and machine learning”. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **29**(10), p. 103123.
- [21] Nair, S., and Lieuwen, T., 2008. “Acoustic Detection of Blowout in Premixed Flames”. *Journal of Propulsion and Power*.
- [22] Gotoda, H., Amano, M., Miyano, T., Ikawa, T., Maki, K., and Tachibana, S., 2012. “Characterization of complexities in combustion instability in a lean premixed gas-turbine model combustor”. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **22**(4), p. 043128.
- [23] Murayama, S., Kaku, K., Funatsu, M., and Gotoda, H., 2019. “Characterization of dynamic behavior of combustion noise and detection of blowout in a laboratory-scale gas-turbine model combustor”. *Proceedings of the Combustion Institute*.
- [24] Ghahramani, I. M. Z., 2005. “A note on the evidence and Bayesian Occam’s razor”. *Gatsby Unit Technical Report*.
- [25] Li, H., Barnaghi, P., Enshaeifar, S., and Ganz, F., 2019. Continual learning using bayesian neural networks.
- [26] Pearce, T., Zaki, M., Brintrup, A., Anastassacos, N., and Neely, A., 2018. “Uncertainty in Neural Networks: Bayesian Ensembling”.
- [27] Sundararajan, M., Taly, A., and Yan, Q., 2017. “Axiomatic Attribution for Deep Networks”. In Proceedings of the 34th International Conference on Machine Learning.
- [28] Schumm, M., Berger, E., and Monkewitz, P. A., 1994. “Self-excited oscillations in the wake of two-dimensional bluff bodies and their control”. *Journal of Fluid Mechanics*.
- [29] Neal, R. M., 1996. *Bayesian Learning for Neural Networks*, Vol. 118 of *Lecture Notes in Statistics*. Springer New York, New York, NY.
- [30] Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D., 2015. “Weight Uncertainty in Neural Networks”.
- [31] Kingma, D. P., and Ba, J. L., 2015. “Adam: A method for stochastic gradient descent”. *ICLR: International Conference on Learning Representations*.
- [32] Ancona, M., Ceolini, E., Öztireli, C., and Gross, M., 2018. “Towards better understanding of gradient-based attribution methods for deep neural networks”.
- [33] Breiman, L., 2001. “Random forests”. *Machine Learning*, **45**(1), Oct, pp. 5–32.