Physica Scripta

PAPER

Recurrence Rate spectrograms for the classification of nonlinear and noisy signals

Thore Hertrampf[1,2,*]   **and Sebastian Oberst**[1]

1   University of Technology Sydney, Centre for Audio, Acoustics and Vibration, Sydney, Australia
2   Dynamics Group, Hamburg University of Technology, Hamburg, Germany
*   Author to whom any correspondence should be addressed.

E-mail: thore.hertrampf@tuhh.de and sebastian.oberst@uts.edu.au

Abstract

Time series analysis of real-world measurements is fundamental in natural sciences and engineering, and machine learning has been recently of great assistance especially for classification of signals and their understanding. Yet, the underlying system's nonlinear response behaviour is often neglected. Recurrence Plot (RP) based Fourier-spectra constructed through τ-Recurrence Rate (RR_τ) have shown the potential to reveal nonlinear traits otherwise hidden from conventional data processing. We report a so far disregarded eligibility for signal classification of nonlinear time series by training RESnet-50 on spectrogram images, which allows recurrence-spectra to outcompete conventional Fourier analysis. To exemplify its functioning, we employ a simple nonlinear physical flow of a continuous stirred tank reactor, able to exhibit exothermic, first order, irreversible, cubic autocatalytic chemical reactions, and a plethora of fast-slow dynamics. For dynamics with noise being ten times stronger than the signal, the classification accuracy was up to $\approx 75\%$ compared to $\approx 17\%$ for the periodogram. We show that an increase in entropy only detected by the RR_τ allows differentiation. This shows that RP power spectra, combined with off-the-shelf machine learning techniques, have the potential to significantly improve the detection of nonlinear and noise contaminated signals.

1. Introduction

Dynamical systems are often mathematically described by nonlinear differential equations, which are difficult to solve using analytical calculus, and commonly require simplification and approximation. Real-world problems related to engineering [1], biology [2], chemistry [3], physics [4], medicine [5] (neuroscience) or economy [6], regularly deal with nonlinear observations to which the underlying formulae are unknown, revealing the significance of nonlinear time series analysis [7, 8]. Often important signal figures are neglected or are assumed as noise, especially when only analyzing the Fourier transform[i] spectrum. In recent years, Machine Learning (ML) based techniques using Convolutional Neural Networks (CNNs)[ii] and their application to classifying acoustic or vibration data using periodogram techniques have found growing popularity [9–11].

     Recurrent, dynamic behaviour of an observable measured as time series of arbitrary dimension can be represented as matrix of recurrences. This is visualized using a two dimensional Recurrence Plot (RP) and quantified using Recurrence Quantification Analysis (RQA)[iii], not limiting the data to properties such as stationarity, linearity, or determinism [8, 12–14]. Embedding univariate time series may reveal higher order and often also higher-dimensional dynamics, both of which are often neglected in conventional signal processing but can be incorporated in a RP-based approach. Thus, linking RQA to linear data processing may contribute to bridging the gap in the analysis of nonlinear deterministic systems. The Fourier spectrum of a time series can be estimated from the RQA measure τ-Recurrence Rate (RR_τ) and is able to identify higher periodic orbits, which a classical spectrum estimation fails to achieve [15]. This technique has already been successfully applied to reveal unstable and stable periodic orbits of various continuous and discrete theoretical dynamical systems [16, 17]; to

characterize hyperspectral geological data [18], proteinsolvent interaction energy [19], and a phasor measurement unit in electrical power systems [20].

Despite potentially higher information content on recurrences present[3] in a RP-based spectrum, physical interpretation is difficult owing to the abstract meaning of power per frequency. Attempts have been made to overcome this issue [17]. However, the application of recurrence-based power spectra in combination with ML has never been attempted.

The current work aims to bridge this gap by employing recurrence-based power spectra to highly nonlinear, and strongly noise contaminated signals, to calculate spectrograms and further use this information for the classification of dynamics via CNN. The new approach is compared to the application of conventional periodograms using different dynamics and noise levels.

2. Materials and methods

2.1. Power spectral density estimation

Consider the discrete time series $\boldsymbol{x} \in \boldsymbol{R}^{N \times m}$ with N state vectors $\vec{x}_i$ of dimension $m = 1$. By application of Takens' embedding theorem[iv], the observable is expanded to a higher dimension through shifted copies of $\boldsymbol{x}$, recovering more of the dynamics [8, 13]. Embedding dimension m and time-delay τ_m are estimated using both, the global False-Nearest Neighbors[vi] [21, 22] and Average Mutual Information[v] [13, 23] algorithms. The embedded time series $\boldsymbol{x}_{\mathrm{emb}}$ has state vectors

$$\vec{x}_{i,\mathrm{emb}} = \{\vec{x}_i,\ \vec{x}_{i-\tau_m}, ..., \vec{x}_{i-(m-1)tau_m}\}. \tag{1}$$

Let the distance matrix $\mathbf{D} \in \boldsymbol{R}^{N \times N}$ be derived from $\boldsymbol{x}_{\mathrm{emb}}$ with arbitrary dimension m, holding pairwise distance information of all data points, i.e.

$$\mathbf{D}_{i,j} := \|\ \vec{x}_{i,\mathrm{emb}} - \vec{x}_{j,\mathrm{emb}}\ \|, \tag{2}$$

The distance matrix is symmetric by definition [13], since $\mathbf{D}_{i,j} = \mathbf{D}_{j,i}$. The recurrence matrix $\mathbf{R}$ is an abstraction of $\mathbf{D}$ by introducing the Heaviside function $\Theta(x)$[vii] [24] and a thresholding criterion ε[viii] [12, 13]

$$\mathbf{R}(\varepsilon) \in \boldsymbol{R}^{N \times N} := (\Theta(\varepsilon - \mathbf{D}_{i,j}))_{0 \leqslant i,j \leqslant N-1}. \tag{3}$$

RPs represent a binary visualization of $\mathbf{R}$, where $\mathbf{R}_{i,j} = 1$ is coloured black, and white otherwise. For ε-neighbourhoods of constant size, which are used throughout this work, RPs are mirrored along the line of identity ($i = j$) with $\mathbf{R}_{i,i} = 1$.[4]

For a univariate observable [22], the discrete Fourier transform[ix] [25] of $\boldsymbol{x}$ results in a discretized frequency spectrum

$$X[k] = \mathscr{F}_{\mathrm{discrete}}\{x\} = \sum_{i=0}^{N-1} \vec{x}_i\ e^{-j\frac{2\pi}{N}ki},\ k \in \{0,...,N-1\}, \tag{4}$$

with i indexing over all elements of $\boldsymbol{x}$, and k denoting the kth frequency bin, corresponding to $f_k = kf_n/N$, where $f_n = F_s/2$ is the Nyquist frequency[x] [26].

Due to the finite length of $\boldsymbol{x}$, its actual Power Spectral Density (PSD)[xi] can only be estimated (superscript ^) as the periodogram $\hat{S}[k] = (1/N)|X[k]|^2$. The Wiener-Khinchin theorem[xii] directly relates a signal's true (continuous) PSD $S(\Omega)$ and its auto-correlation function $r_{xx}(\kappa)$[xiii] as Fourier transform pair $S(\Omega) = \mathscr{F}(r_{xx}(\kappa))$ [15, 27], where κ is the continuous time shift between the original and a copy of the signal. For a stochastic process — e.g. discrete, finite signal $\boldsymbol{x}$ corrupted by Gaussian White Noise (GWN)[xiv] — $r_{xx}(\kappa)$ at discrete time shifts τ can be estimated as

$$\hat{r}_{xx}[\tau] = \begin{cases} \alpha \displaystyle\sum_{i=0}^{N-1-|\tau|} \overline{\vec{x}}[i]\vec{x}[i+\tau],\ |\tau| \leqslant (N-1), \\ 0,\ \text{otherwise,} \end{cases} \tag{5}$$

with $\overline{\vec{x}}$ denoting the complex conjugate of $\vec{x}$ [28] and the scaling factor α controlling the expected value $\boldsymbol{E}\{\hat{r}_{xx}[\tau]\}$. For $\alpha = 1/(N - |\tau|)$ the estimation is *unbiased*, i.e. the expected and true values are equal $\boldsymbol{E}\{\hat{r}_{xx}[\tau]\} = r_{xx}(\kappa = \tau)$. Digital signals are time-discrete. Therefore, to apply the Wiener-Khinchin theorem defined for continuous time, periodogram and auto-correlation function must be related such that the following condition holds

---

[3] assuming that all functions whith $f(t) = f(t + T)$ are a subset $f(t) \approx f(t + T)$

[4] The use of certain thresholding criteria with e.g. a fixed recurrence rate or fixed amount of neighbours renders the RP asymmetric, which is not considered in the current case.

$$\hat{S}[k] \approx \mathscr{F}_{\text{discrete}}(\hat{r}_{xx}[\tau]). \tag{6}$$

This is achieved by choosing $\alpha = 1/N$, resulting in a *biased* estimation $\boldsymbol{E}\{\hat{r}_{xx}[\tau]\} = (1 - |\tau|/N)r_{xx}(\kappa = \tau).$[5]

Extracted from the RP are a range of RQA measures. The Recurrence Rate (*RR*) represents the amount of recurrence points in a RP relative to its size, i.e.

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbf{R}_{i,j}, \tag{7}$$

depending on ε [13, 18], with instances of *RR* often expressed as percentages. Closely related, RR_{τ} is the probability of recurrence after time delay τ

$$\widehat{RR}_{\tau} = \frac{1}{N - |\tau|} \sum_{i=0}^{N-1-|\tau|} \mathbf{R}_{i,i+\tau}, \tag{8}$$

which is the ratio of *RR* on a diagonal at distance τ parallel to the line of identity [13] and the expected value $RR_{\tau} = \boldsymbol{E}\{\mathbf{R}_{i,i+\tau}\}$. Analogously to the auto-correlation function this estimate is unbiased, since

$$\boldsymbol{E}\{\widehat{RR}_{\tau}\} = \boldsymbol{E}\left\{ \frac{1}{N - |\tau|} \sum_{i=0}^{N-1-|\tau|} \mathbf{R}_{i,i+\tau} \right\}$$
$$= \frac{1}{N - |\tau|} \sum_{i=0}^{N-1-|\tau|} \boldsymbol{E}\{\mathbf{R}_{i,i+\tau}\} = RR_{\tau}. \tag{9}$$

Applying equation (6) to $\widehat{RR}_{\tau}$ instead of $\hat{r}_{xx}[\tau]$ may be used as PSD estimates in nonlinear dynamical systems [15–18].

Due to the scaling factor, unbiased auto-correlation functions are indefinite and have statistical and numerical disadvantages for practical applications in power spectral density estimations. They may result in negative values of power, contradicting physical interpretation and prohibiting a representation in logarithmic scale [26, 28–30]. This was observed by us also for unbiased RR_{τ} and the usage of a biased estimator (superscript *)

$$\widehat{RR}_{\tau}^* = \frac{1}{N} \sum_{i=0}^{N-1-|\tau|} \mathbf{R}_{i,i+\tau}, \tag{10}$$

with $\boldsymbol{E}\{\widehat{RR}_{\tau}^*[\tau]\} = (1 - |\tau|/N)RR_{\tau}$ is proposed in this paper for PSD estimation.[6]

2.2. Spectrogram

A sliding window of size N_w is moved across the time series $\boldsymbol{x}$ with overlap of η. An individual PSD is estimated for each window placement and thus the spectrogram [31] is constructed as a two-dimensional time-frequency representation of $\boldsymbol{x}$, with the first dimension denoting the window number and the second dimension the frequency bins f_k.

To display the spectrogram as a digital image, the power value of each frequency bin is window-wise mapped to an integer value in the range of [0, 1, 2,...,255], later used to represent the spectrogram as an 8-bit grey-scale[xv] image (cf. figure 2). This process is called contrast stretching [32]. Before stretching, power is converted to dB-scale[xvi] to increase the dynamic range.

2.3. The dynamical system: continuous stirred tank reactor

As benchmark system, a representative for chemical reactions of specimens inside an open, iso-thermal, continuous stirred tank reactor [33], found in industrial applications, well-mixed, and without diffusion, is used. The system shows exothermic, first-order, irreversible, cubic autocatalytic reactions, occurring in parallel, and exhibiting rich dynamic behaviour, following the period-doubling route to chaos [7]. Due to the higher complexity of two reactants involved, mixed mode oscillations (fast-slow dynamics) are present [34–37].

The system is explicitly described by three differential equations

$$\frac{dY}{dt} = 1 - Y - YDa_1 e^{\gamma_1 \frac{T}{1+T}}, \tag{11a}$$

---

[5] As a property of the discrete Fourier transform [26], the equality in equation (6) is attained from the auto-correlation function using the cyclic continuation of the underlying signal $\boldsymbol{x}$. However, the potential numerical error is assumed to be negligible.

[6] The positive definiteness of $\widehat{RR}_{\tau}^*$ has not been proven and where negative values of power (generally close to zero) were encountered, their absolute values were used.

Table 1. Parameter configurations to simulate the dynamical system inside a nonlinear chemical reactor [33]. A-D are limit cycles on the attractor; E and F are chaotic steady-state solutions. n_{sub} is the number of subharmonics in one period and T_{avg} the average subharmonic duration, averaged over 24 local maxima of T.

Configuration	β_2	Da$_1$	n_{sub}	T_{avg}
A	0.0155	11	1	0.0861 s
B	0.0157	11	2	0.0864 s
C	0.0159	11	4	0.0864 s
D	0.01595	11	8	0.0865 s
E	0.0161	11	∞	0.0866 s
F	0.0163	11	∞	0.0847 s

$$\frac{dZ}{dt} = 1 - Z - ZDa_2 e^{\gamma_2 \frac{T}{1+T}}, \qquad (11b)$$

$$\frac{dT}{dt} = \alpha(\beta_1 YDa_1 e^{\gamma_1 \frac{T}{1+T}} + \beta_2 ZDa_2 e^{\gamma_2 \frac{T}{1+T}} - T), \qquad (11c)$$

the first two representing the dimensionless concentrations Y and Z of reactants A and B (and their conversions into their products). The third equation represents the reactor temperature T. The time series of the latter are used in this analysis, in reference to the accessibility of measurements in real autocatalytic reactors [33, 38]. For arbitrary values of β_2 and Da_1, with $\alpha = 250$, $\beta_1 = 0.04$, $\gamma_1 = \gamma_2 = 25$ and $Da_2 = 2.75$, the differential equations provide one unique solution to the system's steady-state [39]. A number of parameter configurations resulting in different limit cycles on the attractor or chaotic behaviour are considered in table 1. Period doubling bifurcations occur between $A \rightarrow B \rightarrow C \rightarrow D$. In figure 1, phase portraits of all configurations with their embedding are given and the thresholds ε satisfying $RR = 30\%$ in the respective RPs are provided, which are applied throughout here. This was used to maximize the information content, contrary to application domains where a lower RR of e.g. 1% is recommended [40]. However, a state in phase space may be characterized by the trapping time and with an appropriately high threshold, this is encrypted in the RP as vertical structures [13].

One dimensionless time unit of the differential equations is hereafter assumed to be equal to 1 s. Time series of dimensionless T oscillations were computed using the function *ode45* of Matlab® (ver. R2023a) at a sampling frequency of $F_s = 1$ kHz and initial state equal to zero. The sampling time was 40 s and only the steady-state response was extracted by discarding the first 20 s of the time series. F_s was chosen with respect to the slow dynamics of the system and validated by comparison to the time series sampled at $F_{s,high} = 100$ kHz. Periods were aligned with regards to different convergence times to the steady-state solution. The The Root Mean Square Error (RMSE)[xvii] of temperature was less than 0.65% relative to the temperature range for all periodic configurations.
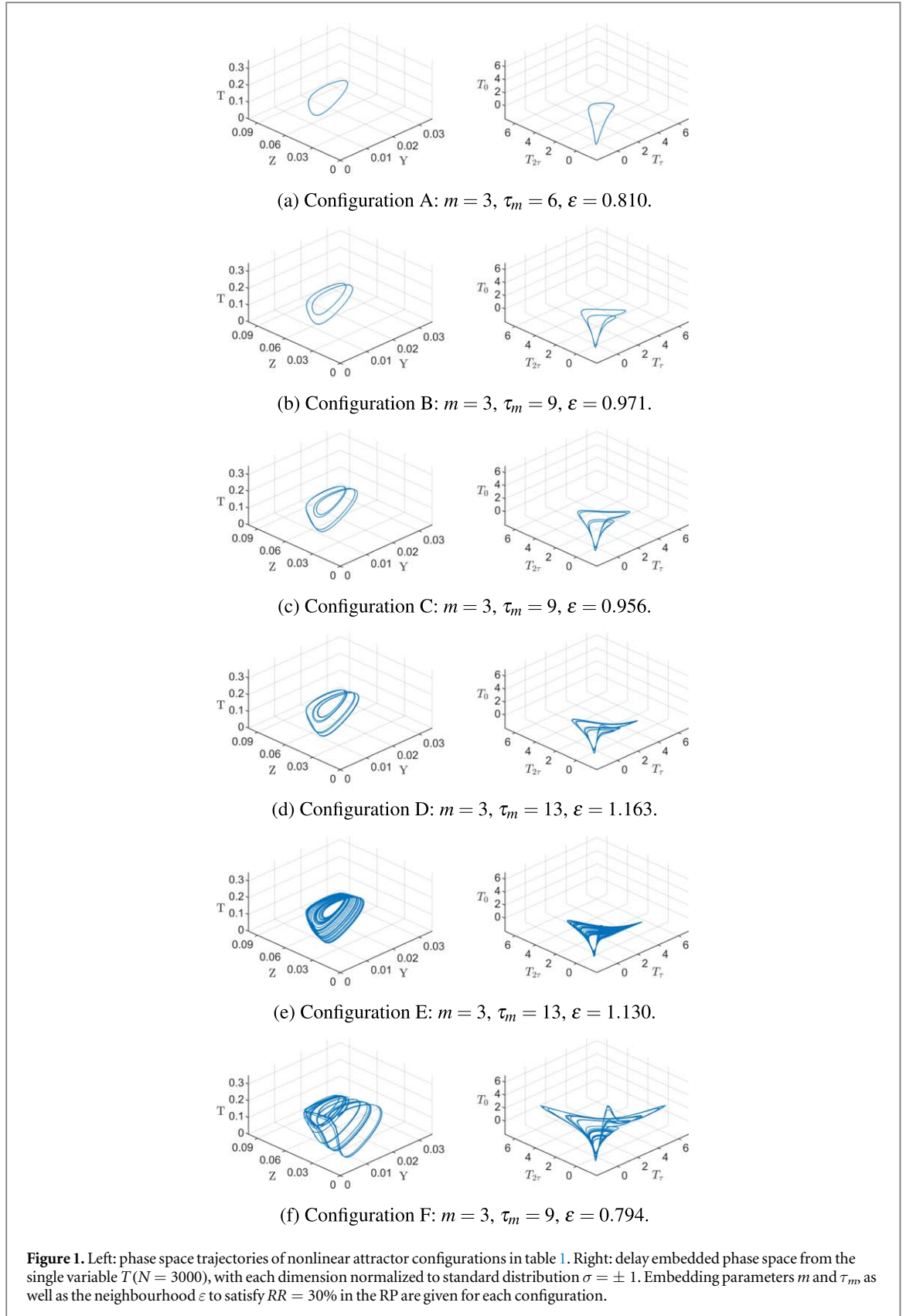
To simulate ambient conditions, computer generated signals are contaminated by additive WGN $\tilde{x} = x + \mathbf{n}$. Signal to Noise Ratio (SNR) is defined as the ratio of Root Mean Square (RMS)[xviii] of signal and noise, respectively

$$SNR = \frac{RMS(\boldsymbol{x})}{RMS(\mathbf{n})}. \qquad (12)$$

2.4. RESnet-50 as machine learning architecture

Within the domain of ML, CNNs have emerged as the most effective image recognition algorithms [41]. They are successfully used in signal detection when trained with spectrogram images [42, 43]. In this work, the pre-trained network *RESnet-50* [44] is incorporated in a transfer learning approach using the *Deep Network Designer* for Matlab. Overall accuracy of a network is evaluated as the percentage of correctly classified validation images, which are not part of the training set. The input image size for the CNN is 224×224 pixels. The network is altered, such that the last fully connected layer and classification layer have an output size of six, i.e. one probability for each signal class. The learning rate of the modified layers was increased tenfold to mitigate the disadvantage compared to pre-trained layers more quickly within the training process. As a solver stochastic gradient descent with momentum of 0.9 is used. During training, iterative optimization of classifiers is achieved through minimization of the cross-entropy loss function[xix] [45]

$$loss = -\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K} t_{lk} \ln y_{lk}, \qquad (13)$$

(a) Configuration A: $m = 3$, $\tau_m = 6$, $\varepsilon = 0.810$.

(b) Configuration B: $m = 3$, $\tau_m = 9$, $\varepsilon = 0.971$.

(c) Configuration C: $m = 3$, $\tau_m = 9$, $\varepsilon = 0.956$.

(d) Configuration D: $m = 3$, $\tau_m = 13$, $\varepsilon = 1.163$.

(e) Configuration E: $m = 3$, $\tau_m = 13$, $\varepsilon = 1.130$.

(f) Configuration F: $m = 3$, $\tau_m = 9$, $\varepsilon = 0.794$.

Figure 1. Left: phase space trajectories of nonlinear attractor configurations in table 1. Right: delay embedded phase space from the single variable T ($N = 3000$), with each dimension normalized to standard distribution $\sigma = \pm 1$. Embedding parameters m and τ_m, as well as the neighbourhood ε to satisfy $RR = 30\%$ in the RP are given for each configuration.

where L is the number of elements (training images), K is the number of classes and y_{lk} the probability output from the network, that element l belongs to class k. The binary value t_{lk} is one, if element l is of class k and zero otherwise.

Other training settings of the network are (i) a fixed learning rate of 0.001; (ii) the mini batch size of 12; (iii) the training duration of 10 epochs; (iv) a validation frequency of 30; (v) shuffling of training data after every epoch; (vi) all other settings according to the default. To reduce over-fitting issues, basic image augmentation

was used, namely *x*-translation up to ± 24 pixels. Other common augmentations (e.g. stretching, rotation) alter the information content of a spectrogram and were therefore not considered.

2.5. Power spectral entropy

Shannon entropy [46] measures the randomness of a probability distribution. In information theory, a discrete sequence of quantized values comprises a signal, which carries a certain amount of information measured by its entropy [47]. The entropy of a system's observation indicates how well it was captured and characterized [8]. By interpretation of a signal's PSD representation as a probability distribution of power per frequency, the power spectral entropy [48, 49] is obtained as

$$H(\hat{S}) = -\sum_{k=0}^{N} \hat{s}[k]\ln(\hat{s}[k]), \tag{14}$$

where $\hat{s}$ denotes the PSD estimation normalized to total power. The logarithmic base of equation (14) can be changed to $b = N$ by

$$\tilde{H}(\hat{S}) = \frac{H(\hat{S})}{\ln(b)}, \tag{15}$$

which gives a normalized value in the range [0, 1].

3. Results

3.1. Spectrogram classification using CNN

The $\widehat{RR}_\tau^*$-spectrum is beneficial for capturing higher-dimensional temporal dynamics of nonlinear time series. This shall be exploited by visualization of signal evolution as a spectrogram [31], then used to classify the observable. Training data is generated from steady-state time series $\boldsymbol{x}$ of system configurations A-F. Segments of $t = 16.36$ s are extracted, with a random shift between 0 to 1s, imitating a random time of measurement, as would be encountered in real-world data acquisitions.
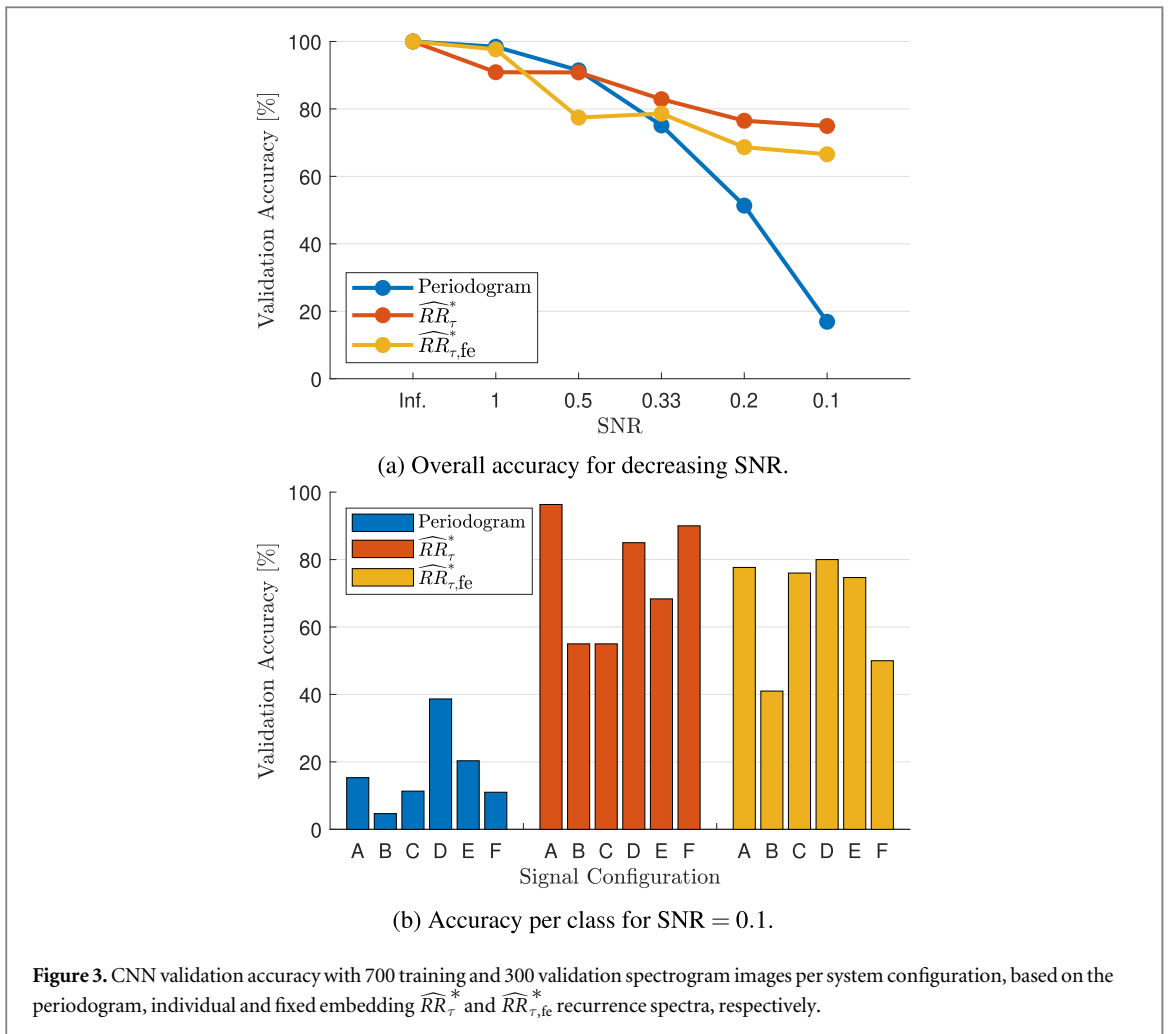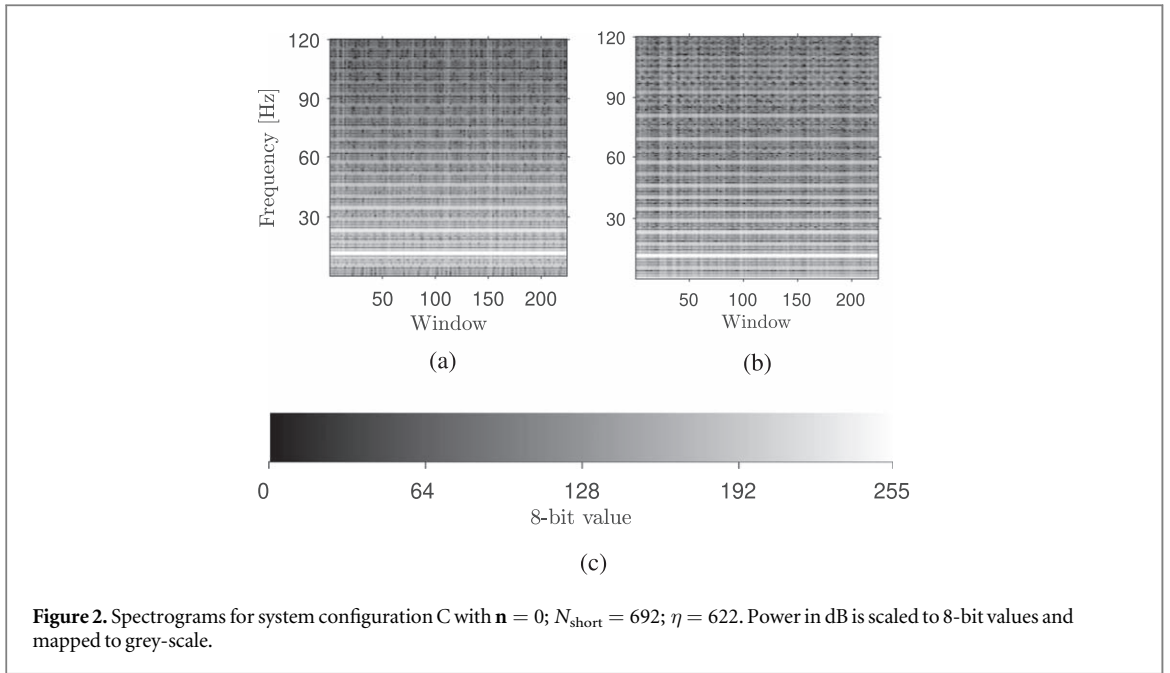
The signals are corrupted by additive WGN $\boldsymbol{n}$ to test the robustness of PSD estimation in challenging SNR scenarios. Signals A-F do not have equal power, hence the power of $\boldsymbol{n}$ is fixed relative to power of signal A equaling SNR = $\{\infty, 1, 0.5, 0.33, 0.2, 0.1\}$. This is applied to all signals rather than choosing power relative to each signal, to prevent the CNN from learning the different power in noise as a classifier. This is also in agreement with the assumption that the ambient noise is constant and independent of the signal.

Noisy time series $\tilde{\boldsymbol{x}}$ are normalized to standard distribution $\sigma = \pm 1$. For recurrence-based signal processing with $\widehat{RR}_\tau^*$, the parameters given in figure 1 are used to delay-embed the respective signals and construct their RPs. Additionally, the analysis is repeated using fixed embedding parameters $m = 3$ and $\tau_m = 9$ for all configurations, denoted as $\widehat{RR}_{\tau,\mathrm{fe}}^*$, with unchanged thresholds.

For window size $N_w = 692$ and overlap $\eta = 622 \approx 0.9 \times N_w$, spectrogram images are computed for each of the three techniques (periodogram, individual and fixed recurrence spectrum). The window was chosen to have the minimal size covering at least one period of steady-state signals A-D and the overlap facilitating 224 window shifts. The highest measurable frequency bin is $F_s/2 = 500$ hz, however, throughout all considered signals, more than 90% (>99% for periodogram) of total power is distributed within frequencies $f = (0, 120]$ hz, which is hence used to capture only the most important dynamics. The discrete Fourier transform resolution was interpolated to $f_\mathrm{res} = 120$ hz/224 by zero-padding [25] of r_{xx}, $\widehat{RR}_\tau^*$ and $\widehat{RR}_{\tau,\mathrm{fe}}^*$, respectively, to utilize all available 224 rows and columns of pixels in the image.

For every SNR scenario and signal configuration, 1000 random realizations of WGN were generated. Each of them is added to the respective signal and spectrogram images are computed for all techniques as previously described. Thus, for every SNR scenario, a set of 6000 images per processing method is obtained. Of each data set, 30% of images were separated for validation, leaving 4200 training images. Example spectrograms of the system in configuration C, without additive noise, are given in figure 2 and the final CNN accuracy is given in figure 3(a).

For signals without noise, 100% accuracy is achieved by all methods. Classification accuracy for the periodogram is directly related to SNR, as it decreases monotonically with increasing noise level. Up to SNR $\geqslant 0.5$, the accuracy is above 90% but then drops significantly, until eventually reaching an accuracy of 16.89% for SNR = 0.1, which is nearly identical to classification by chance (100%/6 = 16.67%). In contrast, the performance when recurrence-based training data is used is more consistent. Although the accuracy at SNR $\geqslant 0.5$ for $\widehat{RR}_\tau^*$ is not able to match that of the periodogram, the validation accuracy is only slowly decreasing and remains above 75% throughout all considered noise levels. For a fixed embedding $\widehat{RR}_{\tau,\mathrm{fe}}^*$, the

Figure 2. Spectrograms for system configuration C with $\mathbf{n} = 0$; $N_{short} = 692$; $\eta = 622$. Power in dB is scaled to 8-bit values and mapped to grey-scale.

(a) Overall accuracy for decreasing SNR.

(b) Accuracy per class for SNR = 0.1.

Figure 3. CNN validation accuracy with 700 training and 300 validation spectrogram images per system configuration, based on the periodogram, individual and fixed embedding $\widehat{RR}_\tau^*$ and $\widehat{RR}_{\tau,\mathrm{fe}}^*$ recurrence spectra, respectively.

accuracy at SNR = 1 is significantly improved compared to $\widehat{RR}_\tau^*$ and matches that of the periodogram. However, for SNR $\leqslant 0.5$ the accuracy is consistently ≈ 5 to 10% lower than for $\widehat{RR}_\tau^*$, suggesting that different embedding is beneficial but not solely responsible for classification.

Table 2. Kullback-Leibler divergence between pairs of reactor temperature time series of dynamic regimes given in table 1. For each column the average (Avg) and median (Med) divergence is given. This is used here as an estimation for each configuration, implying how well the time series is separable globally (within the set of configurations).

	A	B	C	D	E	F
A	0	.0024	.009	.0108	.0192	.0468
B	.0024	0	.0021	.0030	.0079	.0278
C	.0091	.0021	0	.0001	.0018	.0145
D	.0108	.0030	.0001	0	.0012	.0125
E	.0192	.0079	.0018	.0012	0	.0060
F	.0468	.0278	.0145	.0125	.0060	0
Avg	.0147	.0072	.0046	.0046	.0060	.0179
Med	.0099	.0027	.0020	.0021	.0039	.0135

Figure 3(b) gives the classification accuracy per signal for SNR = 0.1. From a dynamical point of view, dynamic regimes A-D are periodic and thus all of them do not generate any new information. Therefore an evenly distributed validation accuracy could be expected. Table 2 gives the Kullback-Leibler divergence$^{\{xx\}}$ [50, 51] for all combinations of configurations. The Kullback-Leibler is used here as a measure of how well one configuration is separable from the rest. A low divergence therefore is an indicator for a high probability of confusion during CNN classification. For each configuration, the average divergence among the set of signals is provided. In each case, the average divergence is strongly influenced by a relatively high divergence to the chaotic regime F. Therefore, to accentuate the low divergences in between limit cycle solutions, the median is expected to depict the probability of confusion better.

Across all 15 CNN training processes on noisy time series, B performs worst six times, C five times (once shared last with B), D four times and E once. This relates statistically with the configurations having the lowest median Kullback-Leibler divergence towards all other regimes, although the distribution is distorted, e.g. configuration B performs overall the worst but has higher median divergence than C and D. The reason for this distortion may lie in the relatively low number of only 15 training sets. Also, the estimation of the confusion probability is not complete and takes into account only the global median divergence rather than a local confusion probability to the least separable regimes.

In addition, the training behaviour of CNNs must also be considered, which aims at the minimization of a loss function. If no global solution, i.e. successful classification of all variables, is found, convergence to a local minimum will occur, favouring the validation accuracy of one or a few classes over other classes [52]. Which class this bias will fall on may differ with each training iteration, as it also depends on the exact images being considered at certain times during training while shuffled randomly. This is a widely studied problem within the ML research community, yet finding a solution here is beyond the scope of this study and requires optimising the training behaviour to become more evenly distributed.
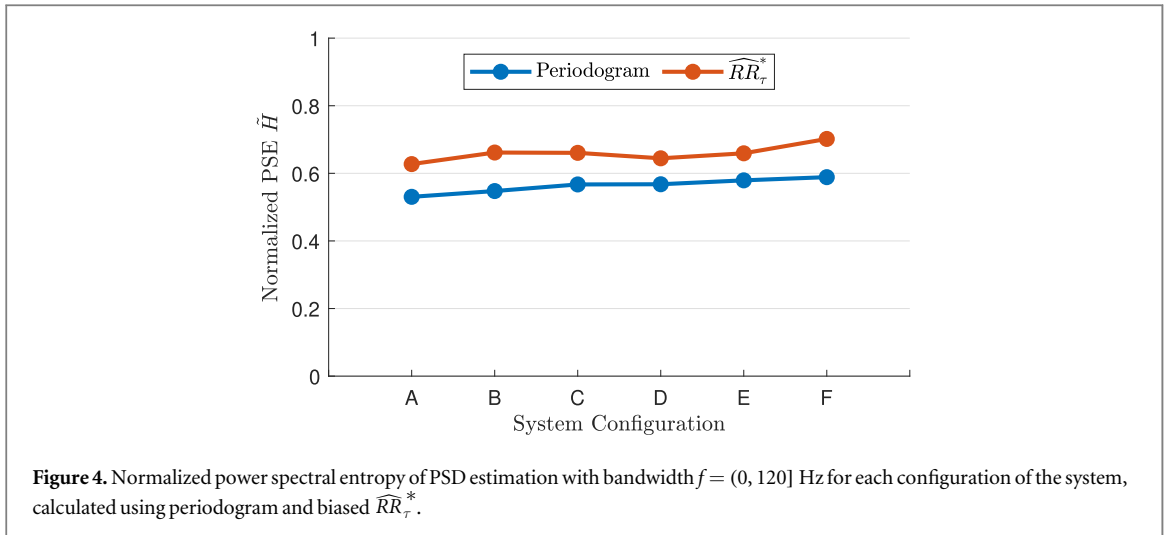
For SNR = 0.1 and periodogram training, it appears that the CNN has converged to a solution-favouring configuration D, which was therefore accurately classified in 38.67% of cases, whereas the other signals where classified correctly only 20% of times, or less. For $\widehat{RR}_\tau^*$ it stands out, that configurations A, D and F were correctly identified in more than 85% of cases. Signal A has a unique embedding delay compared to the other signals, which could be a reason for the accurate discrimination. However, for configurations D and F, the delay parameters are shared respectively by E, as well as B and C. This again confirms that the embedding alone is not responsible for successful classification.

Furthermore, although configurations D and E have not only common embeddings but also similar thresholds, with deviation of less than 3%, the chaotic behaviour of signal E was successfully discriminated even for high noise. For periodic signals B and C with neither unique embedding, nor unique threshold, the classification accuracy is significantly lower. In the classification, B and C were always correctly separated from the other signals but not from each other, i.e. false classification of B exclusively predicted a signal of class C and vice versa. The same is true for signal pairs D-E and A-F. This also occurred in the fixed embedding analysis, indicating that the classification is grouped into three separate problems, according to the similarity of threshold values. The overall validation accuracy of above 50% proves that within these groups successful discrimination is however possible.

3.2. Complexity analysis

Evaluating the complexity of a classification problem is useful in predicting the performance of a ML model [53]. Similarly, it can contribute to choice and validation of data pre-processing. Three categories are identified as

Figure 4. Normalized power spectral entropy of PSD estimation with bandwidth $f = (0, 120)$ Hz for each configuration of the system, calculated using periodogram and biased $\widehat{RR}_\tau^*$.

fundamental requirements to the underlying problem, namely existence of informative class features, i.e. maximization of information as input to the classification problem; homogeneity of topology and the distribution of manifolds of the input; and class linearity, i.e. the grade of separability of inputs belonging to different classes [53, 54]. In the following, complexity measures are used to validate the CNN classification results.

3.2.1. Information content
PSDs are calculated for signal segments with a size of $N_w = 8 T_{avg} F_s$ at random position. The factor of eight ensures that at least one period of all limit cycle solutions is included, while the windows across all signals have similar number of samples. The power spectral entropy, as a measure of information content, for the bandwidth $f = (0, 120)$ Hz is shown in figure 4. Recurrence-based spectra consistently generate significantly more information about the underling system than the periodogram throughout all configurations.

3.2.2. Consistency of PSD estimation
One period of a linear, periodic, stationary signal includes all information about the continuation of the signal. Hence, the same spectral density can be measured in a signal segment of length equal to the period, independent of time [25], which is not necessarily true for nonlinear signals. To test the robustness of PSD estimation against a delay in the time of measurement, signals A-F are segmented again by the rectangular window of size N_w, which is shifted in time incrementally for a total of N_w steps. For each realization the individual PSD is estimated using both, periodogram and $\widehat{RR}_\tau^*$. As in the spectrogram, power is then converted to dB for better dynamic range and re-scaled to [0, 100], allowing comparison between periodogram and recurrence-based spectra, which have different total power. For a statistical analysis of the homogeneity of PSD estimation, the average spectral density across all realizations is considered as control. RMSE between each realization and the control acts as an indicator of similarity [55]. The standard distribution of RMSE for all signals is presented in table 3(a).

There is no clear indication that one technique is advantageous in this respect as each of them shows lower average RMSE in three of the signals. The high variation of ± 3.08 for periodograms of signal F stands out being twice as high as for the others and could be related to the chaotic nature of the signal. Due to the complex periodic behaviour and qualitative differences between the signals, differences in RMSE between signals would be expected and exist for both techniques. Table 3(b) gives the results, when the experiment is repeated for noisy signals with SNR = 0.5. The distribution of RMSE in the periodogram is higher and now consistent at $\approx 10 \pm 1.5$ across all configurations, likely due to the noise substantially contributing to the PSD. In contrast, for $\widehat{RR}_\tau^*$ the average RMSE increased moderately compared to the signals without noise and now with the exception of A is always ≈ 12 to 30% lower than for the periodogram. Further, distinctions between signals are persistent and the variance in RMSE is high, indicating that the impact of noise varies with signal evolution and therefore features may not suppressed equally strong. Overall, this indicates a better resilience against noise, which is in good agreement with the improved classification accuracy of noisy signals, cf. [16, 18].

3.2.3. Uniqueness analysis
For classification, the representation of information has to be unique. Thus, Spearman's rank correlation $r_{rs}(x, y)$ [55] acts as a metric for the similarity of two signals' PSD. It returns a value in the range $[-1, 1]$, where $r_{rs} = 1$

Table 3. Distribution of RMSE between PSD estimation and mean PSD, as the window of length $N_w = 8T_{\mathrm{avg}}F_s$ of the respective configuration is incrementally shifted N_w times. Comparison of signals with and without noise.

	(a) SNR $= \infty$	
Configuration	Periodogram	$\widehat{RR}_\tau^*$
A	5.73 ± 1.64	5.47 ± 1.56
B	6.10 ± 1.57	5.74 ± 1.58
C	5.95 ± 1.45	6.17 ± 1.29
D	5.61 ± 1.17	6.68 ± 1.32
E	7.11 ± 1.59	7.74 ± 1.33
F	8.68 ± 3.08	7.85 ± 1.51

	(b) SNR $= 0.5$	
Configuration	Periodogram	$\widehat{RR}_\tau^*$
A	10.13 ± 1.51	10.39 ± 5.40
B	10.00 ± 1.40	7.09 ± 2.63
C	10.01 ± 1.33	8.81 ± 4.83
D	10.69 ± 1.54	7.36 ± 1.81
E	10.08 ± 1.46	7.32 ± 1.81
F	10.15 ± 1.53	8.00 ± 3.54
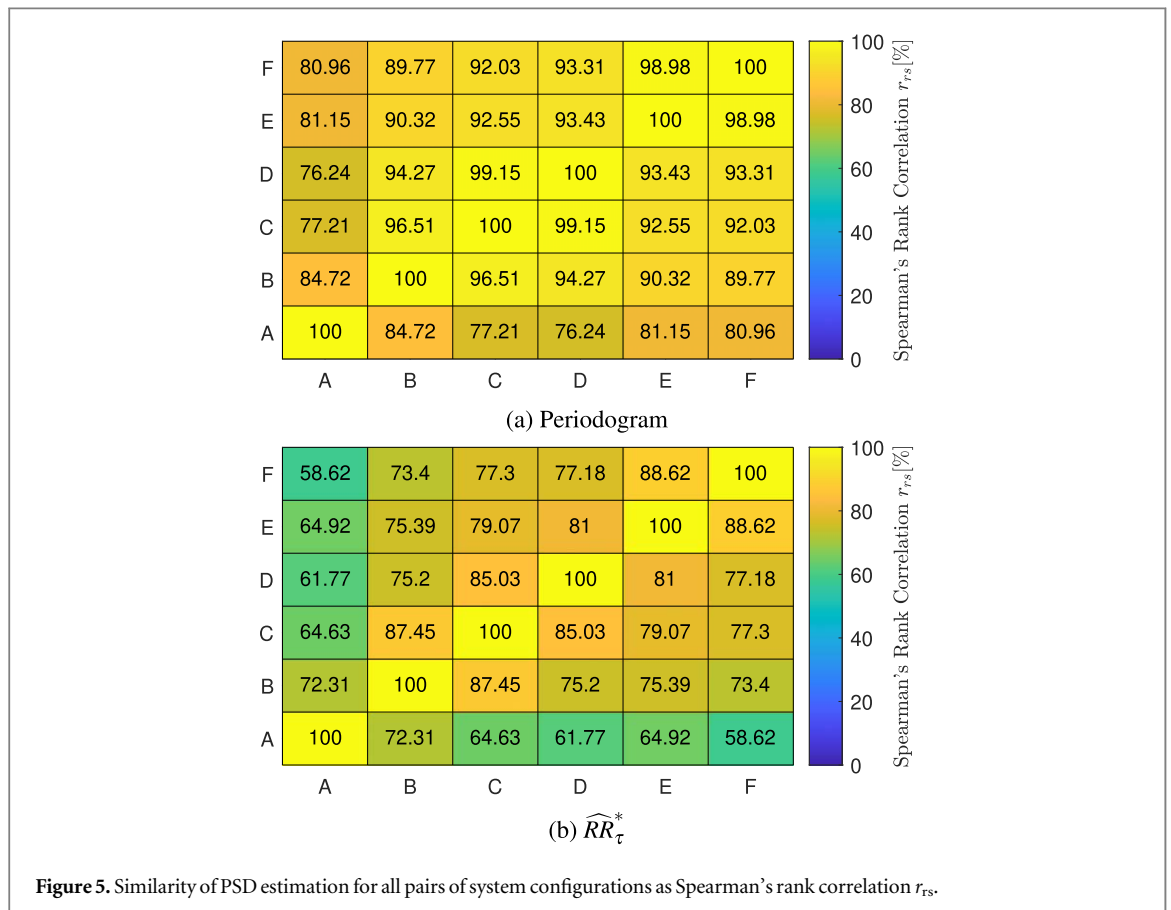
Figure 5. Similarity of PSD estimation for all pairs of system configurations as Spearman's rank correlation r_{rs}.

means identity, $r_{\mathrm{rs}} = -1$ indicates that x and y are opposites, and $r_{\mathrm{rs}} = 0$ denotes maximum disparity.[7] Figure 5 shows the correlation of control spectra (see above) between configurations A–F as a percentage.

Figure 5(a) displays the correlation of PSDs computed using the periodogram. The lowest correlation of $r_{\mathrm{rs}} = 76.24\%$ occurs between A and D. The spectra of pairs B-C; C-D; D-E; and E-F correlate significantly, each with 96.51%; 99.15%; 93.43%; and 98.98%, which translates to a low discriminative power between them.

---

[7] Equivalently, the mutual information [56] could be used and could be more suitable in case the r_{rs} does not work.

Figure 5(b) for $\widehat{RR}_\tau^{\,*}$ spectra shows much lower overall correlation between configurations. Signal A has a correlation of $r_{rs} \leqslant 72.31\%$ with all other signals and as low as $r_{rs} = 58.62\%$ for the pair A-F. This is in good agreement with the accurate classification of this signal by the CNN. Further, the correlation of all neighbouring pairs is below 90%, which is a significant improvement, compared to the periodogram, and shows that the $\widehat{RR}_\tau^{\,*}$ spectrum is more sensitive to weak characteristics of the signals.

4. Discussion

Classification of noisy, nonlinear data is a recurring and practically relevant, widespread issue which we address here using nonlinear time series analysis, especially RQA. A recurrence-based power spectrum, even though rarely studied, has been useful in a range of applications [15–18]. We therefore extended the PSD estimate to form the recurrence-based spectrogram using recurrence-based power spectra. We systematically estimated the PSD of periodogram and $\widehat{RR}_\tau^{\,*}$ using the time series of a model of a continuous stirred tank reactor [33]. The signal's complexity as well as its signal to noise level have been altered by considering higher order dynamics and adding noise to the signal. Different dynamic regimes following the period-doubling route to chaos, ranging from periodic to aperiodic/chaotic motion were classified using the deep CNN RESnet-50. For increasingly noisy conditions down to a SNR = 0.1, we showed that for the $\widehat{RR}_\tau^{\,*}$ method, both with signal dependent as well as fixed embedding, compared to the classical periodogram method a higher classification accuracy can be achieved.

The main improvement in validation accuracy for noisy time series achieved through the introduction and use of recurrence-based spectrograms is considered to result partly from the thresholding practice in recurrence plots, which offers a natural resilience against noise [13], and partly from the incorporation of higher order dynamics through attractor reconstruction, which adds more information to the PSD and therefore to the training images for the CNN to learn and classify.

By employing the power spectral entropy, we demonstrate that the recurrence-based power spectra contain more information; they are more homogeneous when measured at different times of the noisy signal, and they show stronger uniqueness, when compared to similar signals of the same attractor. Homogeneity is important because it increases ML trainability and efficiency and allows task specific network designs [57, 58]. Uniqueness is significant because it leads to higher discriminative power of features and fewer class overlaps reduce false predictions [53]. The information content in recurrence plots, underlying the RR_τ is dependent on a-priori information about the system and its embedding. The embedding delay τ_m has shown to have a strong influence on the information of the spectrogram and thus the classification accuracy. Determining optimal embedding in noisy signals is however a complex process and more research needs to be conducted to better understand optimality in light of different scenarios. The efficacy of equal embedding on all signals was demonstrated here though, again achieving excellent results, with a reduction in overall classification accuracy of only $\approx 10\%$. The threshold ε applied to RPs for the creation of recurrence-based spectrograms additionally supports the classification but is difficult to obtain from the noisy signal. While research on the effects of noise on the RP is ongoing [59], it was here shown, that a CNN trained on signals with optimal thresholding is capable of distinguishing between noisy signals processed with near identical threshold values, without any de-noising attempts on the RP.

While validation of the proposed technique on real-life data featuring various types of noise is yet outstanding and may further reveal the method's high potential in the future, the here considered scenario suggests already good applicability to numerous problems. Having extended the classification capability to very low SNR makes it suitable to the detection of weak signals. It is therefore expected to be well suited for e.g. classification of acoustic emissions [60] or vibratory insect signals [61], often being nonlinear due to the wave carrier material.

Oberst *et al* [18] filtered multiplicative noise with the GHKSS method [8] and then applied the recurrence-based power spectrum. However, it is yet unclear how phase noise and multi-operational coupling models of noise can be studied with the recurrence-based spectrograms. Also, further improvements may be achieved in low SNR scenarios combining feature extraction with using wavelet decomposition, either on the signal itself [62], or two-dimensional on the RP [63] before processing.

The best improvements over conventional signal processing were achieved when the optimal embedding and thresholds were known. Hence, application is highly recommended to problems where the observable of interest may be measured under laboratory conditions and later also classified during exposition to noise in its designated working environment. One area of such could be health monitoring of machine parts. Signature signals of e.g. faulty roll bearings [10] or friction parts [64] may be characterized in a test setup and then detected during active machine operation. Another applicable research field is acoustic source localization in anisotropic materials, which often faces the challenge of time difference of arrival measurements at nearby sensors. More

reliable classification and therefore timing of the nonlinear vibration response signals could lead to better approximation of the excitation's origin.

Using recurrence-based spectrograms in combination with more elaborate embedding algorithms and thresholding criteria could also lead to fully automatic classification capabilities without a-priori knowledge of the system for unsupervised ML in those applications.

Acknowledgments

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://zenodo.org/doi/10.5281/zenodo.8353983 [69].

Conflict of interest

The authors have no conflicts of interest to disclose.

Author contributions

Author 1: Conceptualization (support); Data curation (lead); Formal analysis (equal); Investigation (equal); Methodology (equal); Software; Validation (equal); Visualization (lead); Writing original draft (lead); Writing, review & editing (equal). **Author 2**: Conceptualization (lead); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (support); Resources; Funding acquisition; Writing original draft (support); Writing, review & editing (equal).

APPENDIX

In the previoussections, references to the APPENDIX, explaining theorems and concepts used here, are marked by the superscript $^{\{*\}}$.

Description of theorems and Concepts

(i) **Fourier transform:** Any signal in time may be equivalently represented as a function of the frequencies present within, i.e. its frequency spectrum. The conversion is known as Fourier transform [28].

(ii) **Convolutional neural network:** Artificial neural networks perform machine learning tasks through imitation of the way the human brain processes information. A subset of these are convolutional neural networks, which are specialized mostly for pattern recognition in images [65].

(iii) **Recurrence quantification analysis:** The analysis of nonlinear data is difficult as many linear signal processing methods are not applicable. Thus, looking at recurring states in time series gives insight into the underlying dynamics. This is the foundation of Recurrence Quantification Analysis [13].

(iv) **Taken's embedding theorem:** Given a univariate measurement of a deterministic system which is more than one dimensional, the theorem states, that the higher order dynamics of the attractor may be qualitatively reconstructed through delay embedding. In doing so, the invariants, being the signature of the dynamics, are preserved [66].

(v) **Average mutual information algorithm:** This algorithm is used to find the optimal delay τ_m for time series embedding [23]. The functions minimum occurs at the delay for which a time series and its copy are as independent from each other as possible.

(vi) **Global false-nearest neighbours algorithm:** Using an already determined time delay τ_m, a time series is embedded. The algorithm tests if neighbours of the time series remain neighbours after embedding. If not, the embedding dimension m is increased to minimize the false neighbours [21, 23].

(vii) **Heaviside function:** The function, defined as

$$\Theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geqslant 0, \end{cases}$$

maps negative inputs to zero and all other inputs to unity [24].

(viii) **Thresholding criterion:** The thresholding criterion is responsible for the appearance of a recurrence plot. It defines for each element of the distance matrix a threshold value, marking the state as recurrent or not [13]. While many criteria exist, in this work the threshold is fixed and thus equal for every element of the matrix.

(ix) **Discrete fourier transform:** For digital signals, the discrete Fourier transform is defined and can be efficiently computed using the fast Fourier transform algorithm [26, 67].

(x) **Nyquist frequency:** The Nyquist frequency is half of the sampling frequency and denotes the highest frequency that is observable in a discrete time series. If higher frequencies are present in the original signal [26].

(xi) **Power spectral density:** The power of each frequency component present in a signal is given as a function called the power spectral density. The function is independent of the frequencies phase, which makes it well suited for detecting certain signatures within the signal contents [28].

(xii) **Wiener-khinchin theorem:** The Wiener-Khinchin theorem states, that for stationary, statistical processes, the power spectral density may equivalently be derived from either the signals frequency representation or the Fourier transform of its auto-correlation function [15, 28].

(xiii) **Auto-correlation function:** The auto-correlation function may be regarded as the linear correlation measure of a signal with a shifted version of itself, i.e. how well they align. This is especially useful for the analysis of periodicities in a signal [28].

(xiv) **Gaussian white noise:** The noise signal has a Gaussian distribution in the time domain around its mean value of zero, while featuring a uniform power spectral density across its bandwidth [68].

(xv) **8-bit grey-scale:** In digital imaging, when the grey-scale colour of a pixel is given by a number in $\{0,...,2^8 - 1\}$, it may be represented by one Byte of storage. Each number has assigned a shade in between zero being white and $2^8 - 1 = 255$ displayed as black.

(xvi) **Decibel scale:** To emphasize differences and relations between frequencies, their power is often expressed logarithmically as decibel (dB) [28], with the conversion as follows

$$\hat{S}[f_k] = 10 \log_{10}(\hat{S}[f_k]) \, \text{dB}.$$

(xvii) **Root mean square error:** The error or deviation between two sets of numbers is measured as the square root of the arithmetic mean of the squared difference between the sets.

(xviii) **Root mean square:** For a set of numbers, the root mean square is defined as the square root of the arithmetic mean of the square of all values in the set.

(xix) **Loss function:** During the training of a neural network, the weights of all functional layers that the input is propagates through are iteratively adjusted. The loss function measures the performance of these weights and is zero only when the network's classification is optimal. Minimization of the loss function drives the training process [45].

(xx) **Kullback-leibler divergence:** Also known as relative entropy, it measures the difference of two probability distributions, i.e. the information of discrimination [50, 51]. A divergence of zero means equality of the distributions.

ORCID iDs

Thore Hertrampf ⓘ https://orcid.org/0009-0003-1327-4027
Sebastian Oberst ⓘ https://orcid.org/0000-0002-1388-2749

References

[1] Bongard J and Lipson H 2007 *Proceedings of the National Academy of Sciences of the United States of America Automated reverse engineering of nonlinear dynamical systems* 24 9943–8

[2] Mosconi F, Julou T, Desprat N, Sinha D K, Allemand J F, Croquette V and Bensimon D 2008 Some nonlinear challenges in biology *Nonlinearity* **21** T131–47

[3] Sagués F and Epstein I R 2003 Nonlinear chemical dynamics *Dalton Trans.* **7** 1201–17

[4] Wan Y, Seemann O, Tata S, Andriyash I A, Smartsev S, Kroupp E and Malka V 2022 Direct observation of relativistic broken plasma waves *Nat. Phys.* **10** 1186–90

[5] Wang W K *et al* 2022 A Systematic Review of Time Series Classification Techniques Used in Biomedical Applications *Sensors (Basel, Switzerland)* **22** 8016

[6] Barra S, Carta S M, Corriga A, Podda A S and Recupero D R 2020 Deep learning and time series-to-image encoding for financial forecasting *IEEE/CAA Journal of Automatica Sinica* **7** 683–92

[7] Schuster H G and Just W 2005 *Deterministic Chaos* (Weinheim: Wiley-VCH) (https://doi.org/10.1002/3527604804)

[8] Kantz H and Schreiber T 2010 *Nonlinear Time Series Analysis* (Cambridge: Cambridge University Press) (https://doi.org/10.1017/CBO9780511755798)

[9] Darwish A, Halkon B and Oberst S 2022 Non-Contact Vibro-Acoustic Object Recognition Using Laser Doppler Vibrometry and Convolutional Neural Networks *Sensors (Basel, Switzerland)* **22** 9360

[10] Shaw J and Wu B 2021 Prediction of Remaining Useful Life of Wind Turbine Shaft Bearings Using Machine Learning *J. Mar. Sci. Technol.* **29** 631–7

[11] Stowell D 2022 Computational bioacoustics with deep learning: a review and roadmap *PeerJ* **10** e13152

[12] Eckmann J P, Kamphorst S O and Ruelle D 1987 Recurrence Plots of Dynamical Systems *Europhysics Letters (EPL)* **4** 973–7

[13] MARWAN N, CARMENROMANO M, THIEL M and KURTHS J 2007 Recurrence plots for the analysis of complex systems *Phys. Rep.* **438** 237–329

[14] Goswami B 2019 A Brief Introduction to Nonlinear Time Series Analysis and Recurrence Plots *Vibration* **2** 332–68

[15] Zbilut J P and Marwan N 2008 The Wiener–Khinchin theorem and recurrence quantification *Phys. Lett. A* **372** 6622–6

[16] Oberst S, Marburg S and Hoffmann N 2017 Determining periodic orbits via nonlinear filtering and recurrence spectra in the presence of noise *Procedia Engineering* **199** 772–7

[17] Kraemer K H, Hellmann F, Anvari M, Kurths J and Marwan N 2022 Spike Spectra for Recurrences *Entropy (Basel, Switzerland)* **24** 1689

[18] Oberst S, Niven R, Lester D, Ord A, Hobbs B and Hoffmann N 2018 Detection of unstable periodic orbits in mineralising geological systems *Chaos: An Interdisciplinary Journal of Nonlinear Sciences* **28** 085711

[19] Karain W 2016 THz frequency spectrum of protein-solvent interaction energy using a recurrence plot-based Wiener-Khinchin method *Proteins* **84** 1549–57

[20] Arefi M and Chowdhury B 2018 *2018 IEEE Industry Applications Society Annual Meeting (IAS) Data-driven detection of sustained oscillations and frequency mode in power systems* 1–8

[21] Kennel M B, Brown R and Abarbanel H D 1992 Determining embedding dimension for phase-space reconstruction using a geometrical construction *Physical Review. A, Atomic, Molecular, and Optical Physics* **45** 3403–11

[22] Oberst S and Lai J 2015 A statistical approach to estimate the Lyapunov spectrum in disc brake squeal *J. Sound Vib.* **334** 120–35

[23] Wallot S and Mønster D 2018 Calculation of Average Mutual Information (AMI) and False-Nearest Neighbors (FNN) for the Estimation of Embedding Parameters of Multidimensional Time Series in Matlab *Frontiers in Psychology* **9**

[24] Zhang W and Zhou Y 2021 *Level-set functions and parametric functions The Feature-Driven Method for Structural Optimization* (Amsterdam: Elsevier) 9–46

[25] Sundararajan D D 2021 *Digital Signal Processing* (Cham: Springer International Publishing) (https://doi.org/10.1007/978-3-030-62368-5)

[26] Rao K D and Swamy M 2018 *Digital Signal Processing* (Singapore: Springer) (https://doi.org/10.1007/978-981-10-8081-4)

[27] Cohen L 1998 *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing The generalization of the Wiener-Khinchin theorem* 1577–80

[28] Castani F 2006 *Spectral Analysis* (London, UK: ISTE) (https://doi.org/10.1002/9780470612194)

[29] Zhang X D 2022 *Modern Signal Processing* (Berlin: De Gruyter) (https://doi.org/10.1515/9783110475562)

[30] Piet M T 2006 *Automatic Autocorrelation and Spectral Analysis* (London: Springer) (https://doi.org/10.1007/1-84628-329-9)

[31] Hlawatsch F and Auger F 2008 *Time-frequency Analysis: Concepts and Methods Digital Signal and Image Processing Series—DSP* (London; Hoboken, NJ: ISTE and Wiley) 978-1-84821-033-2

[32] Bredies K and Lorenz D 2018 *Mathematical Image Processing* (Cham: Springer International Publishing) (https://doi.org/10.1007/978-3-030-01458-2)

[33] Lynch D T, Wanke S E and Rogers T D 1982 Chaos in a continuous stirred tank reactor *Math. Modelling* **3** 103–16

[34] Bertram R and Rubin J E 2017 Multi-timescale systems and fast-slow analysis *Math. Biosci.* **287** 105–21

[35] Oberst S and Tuttle S 2018 Nonlinear dynamics of thin-walled elastic structures for applications in space *Mech. Syst. Sig. Process.* **110** 469–84

[36] Trzaska Z 2021 MMOs in Chemistry *Mixed Mode Oscillations (MMOs). Studies in Systems, Decision and Control* (Cham: Springer International Publishing) 71–86

[37] Lu S and Oberst S 2023 Recurrence-based reconstruction of dynamic pricing attractors *Nonlinear Dyn.* **111** 15263–78

[38] Abashar M and Judd M 1998 Synchronization of chaotic nonlinear oscillators: study of two coupled CSTRs *Chem. Eng. Sci.* **53** 3741–50

[39] Thompson J 1984 An introduction to nonlinear dynamics *Appl. Math. Modell.* **8** 157–68

[40] Zbilut J P, Zaldivar-Comenges J M and Strozzi F 2002 Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data *Phys. Lett. A* **297** 173–81

[41] Chen L, Li S, Bai Q, Yang J, Jiang S and Miao Y 2021 Review of Image Classification Algorithms Based on Convolutional Neural Networks *Remote Sensing* **13** 4712

[42] Costa Y M, Oliveira L S and Silla C N 2017 An evaluation of Convolutional Neural Networks for music classification using spectrograms *Appl. Soft Comput.* **52** 28–38

[43] Pham M T, Kim J M and Kim C H 2020 Accurate Bearing Fault Diagnosis under Variable Shaft Speed using Convolutional Neural Networks and Vibration Spectrogram *Applied Sciences* **10** 6385

[44] He K, Zhang X, Ren S and Sun J 2015 Deep residual learning for image recognition arXiv:10.48550/arXiv.1512.03385

[45] Bishop C M and Nasrabadi N M 2006 *Pattern Recognition and Machine Learning* (New York: Springer) 4978-0-387-31073-2

[46] Shannon C E 1948 A Mathematical Theory of Communication *Bell Syst. Tech. J.* **27** 379–423

[47] Chambert-Loir A 2022 *Information Theory* (Cham: Springer International Publishing) 144 (https://doi.org/10.1007/978-3-031-21561-2)

[48] Dai Y, Zhang H, Mao X and Shang P 2018 Complexity–entropy causality plane based on power spectral entropy for complex time series *Physica A* **509** 501–14

[49] Zhang A, Yang B and Huang L 2008 *2008 International Conference on BioMedical Engineering and Informatics (IEEE)Feature Extraction of EEG Signals Using Power Spectral Entropy* 435–9

[50] Kullback S and Leibler R A 1951 On Information and Sufficiency *The Annals of Mathematical Statistics* **22** 79–86

[51] Csiszar I 1975 I-Divergence Geometry of Probability Distributions and Minimization Problems *The Annals of Probability* **3** 146–58

[52] Nielsen M A 2015 *Neural Networks and Deep Learning* (San Francisco: Determination Press)

[53] Lorena A C, Garcia L P F, Lehmann J, Souto M C P and Ho T K 2020 How Complex Is Your Classification Problem? *ACM Comput. Surv.* **52** 1–34

[54] Basu M and Ho T K 2006 *Data Complexity in Pattern Recognition* (London: Springer) (https://doi.org/10.1007/978-1-84628-172-3)

[55] Yilmaz Ş, Vandermarliere E and Martens L 2017 Methods to Calculate Spectrum Similarity *Methods in molecular biology (Clifton, N.J.)* **1549** 75–100

[56] Oberst S, Lai J C S and Evans T A 2018 Key physical wood properties in termite foraging decisions *Journal of The Royal Society Interface* **15** 20180505

[57] Lo J E *et al* 2021 Data Homogeneity Effect in Deep Learning-Based Prediction of Type 1 Diabetic Retinopathy *Journal of diabetes research* **2021** 2751695

[58] Eitel F, Albrecht J P, Weygandt M, Paul F and Ritter K 2021 Patch individual filter layers in CNNs to harness the spatial homogeneity of neuroimaging data *Sci. Rep.* **11** 24447

[59] Wendi D and Marwan N 2018 Extended recurrence plot and quantification for noisy continuous dynamical systems *Chaos (Woodbury, N.Y.)* **28** 085722

[60] Sikorski W 2012 *Acoustic Emission* (London, UK: InTech) (https://doi.org/10.5772/2070)

[61] Oberst S, Baro E N, Lai J C S and Evans T A 2014 Quantifying ant activity using vibration measurements *PLoS One* **9** e90902

[62] Lang M, Guo H, Odegard J E, Burrus C S and Wells R O 1996 Noise reduction using an undecimated discrete wavelet transform *IEEE Signal Process Lett.* **3** 10–2

[63] Ghazali K H, Mansor M F, Mustafa M M and Hussain A 2007 *2007 5th Student Conference on Research and Development (IEEE)Feature extraction technique using discrete wavelet transform for image classification* 1–4

[64] Xing P, Zhu Y, Li G, Liu T, Gao H, Song Y and Zhang H 2023 A Detecting Method for "Weak" Friction-Induced Vibration Based on Cross-Correlation Analysis between Vibration and Sound Signals *Applied Sciences* **13** 7468

[65] O'Shea K and Nash R 2015 An Introduction to Convolutional Neural Networks arXiv:10.48550/arXiv.1511.08458

[66] Takens F 1981 Detecting strange attractors in turbulence *Dynamical Systems and Turbulence, Warwick 1980* (Berlin; Heidelberg: Springer) 366–81 978-3-540-38945-3

[67] Cooley J W and Tukey J W 1965 An algorithm for the machine calculation of complex Fourier series *Math. Comput.* **19** 297–301

[68] Balakrishnan A V and Mazumdar R R 2011 On Powers of Gaussian White Noise *IEEE Trans. Inf. Theory* **57** 7629–34

[69] Hertrampf T and Oberst S 2023 Supplementary Data: Recurrence Rate spectrograms for the classification of nonlinear and noisy signals *Zenodo*