A method for identifying causality in the response of nonlinear dynamical systems

Joseph Massingham[1], **Dr Ole Nielsen**[1,2], **Dr Tore Butlin**[1]
[1] Department of Engineering, University of Cambridge; [2] Bose Corporation

Abstract

Predicting the response of nonlinear dynamical systems subject to random, broadband excitation is important across a range of scientific disciplines, such as structural dynamics and neuroscience. Building data-driven models requires experimental measurements of the system input and output, but it can be difficult to determine whether inaccuracies in the model stem from modelling errors or noise. This influences the investment of resources into building more accurate models. There are currently no solutions to this problem in the absence of a complete benchmark model. This paper presents a novel method to identify the causal relationship between the input and output of a system in the presence of output noise. Using this method, researchers could collect experimental input and output data for a nonlinear dynamical system, and identify how much of the output is caused by the input as a function of frequency.

1 Introduction

Nonlinear dynamical systems are observed across a broad range of scientific disciplines [11], from structural dynamics [19] to neuroscience [2]. Modelling their behaviour offers meaningful insights, enables predictions, and allows for control to generate desirable outcomes. Accurate models generally involve experimental measurements, e.g. to carry out model updating for physics-based models, or to train data-driven models. However, it can be difficult to determine whether prediction errors are due to additional noise or inaccuracies in the model. Consider a nonlinear dynamical system subject to broadband random excitation, $x(t)$, which is especially prevalent in structural dynamics. The output of the system is measured to be $y_n(t)$, where

$$y_n(t) = y(t) + \epsilon_n(t) = \mathcal{M}\{x(\tau)_{\tau \leq t}\} + \epsilon_n(t) \, . \tag{1}$$

Here $\mathcal{M}\{\cdot\}$ represents the true system, $\epsilon_n(t)$ is additive noise and $y(t)$ is the component of $y_n(t)$ that is caused by $x(t)$. The noise ϵ_n is assumed to be due to additional signal sources rather than measurement noise, hence there is no noise on the input. Measuring the causal relationship between x and y_n (i.e. the relative noise level) is important in determining the feasibility of a project before further investment in data collection and computational resources. This can be quantified using a nonlinear coherence metric, which is also equal to the linear coherence between $y(t)$ and $y_n(t)$ [7], but cannot be calculated directly due to the unavailability of $y(t)$. As a result, only a lower bound on the nonlinear coherence can be estimated using a model prediction:

$$y_z(t) = \mathcal{G}\{x(t)\} = \mathcal{M}\{x(\tau)_{\tau \leq t}\} + \epsilon_z(t) \, , \tag{2}$$

where $\mathcal{G}\{\cdot\}$ is the model and ϵ_z represents model errors caused by the inability of $\mathcal{G}\{\cdot\}$ to fully capture $\mathcal{M}\{\cdot\}$. Determining the true nonlinear coherence is challenging because it is difficult to distinguish between $\epsilon_n(t)$ and $\epsilon_z(t)$, which both appear as random noise.

Despite its wide ranging applications, this work is specifically motivated by feedforward

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

active noise reduction (ANR), where x is used as a reference signal to reduce the target signal, $y_n(t)$. In order to reduce the unwanted noise, a model must be developed to accurately predict $y_n(t)$ using $x(t)$. The maximum noise reduction that can be achieved is quantified by the causal relationship between $x(t)$ and $y_n(t)$. For linear systems, such as those in sound cancelling earbuds and headsets, this is quantified using the linear coherence metric, and the level of noise reduction that can be achieved, in decibels, is equal to $\Gamma_{XY}(f) = 10\log_{10}\left(1-\gamma^2_{XY}\right)$. However, ANR has recently begun implementation in cars, where the transmission pathway from x to y_n is highly nonlinear [3, 13]. For nonlinear systems such as this, the nonlinear coherence would be used to quantify the level of noise reduction that is possible. This is crucial in evaluating the benefits of investing resources into developing a high fidelity model of the car to enhance ANR performance. It is not currently possible to calculate the nonlinear coherence without investing the necessary resources into building the best model possible.

This works aims to estimate the nonlinear coherence between $x(t)$ and $y_n(t)$ for nonlinear dynamical systems, with relatively small quantities of data and an incomplete model. These systems can be described by the general class of second order ordinary differential equation (ODE) $\mathcal{N}\left(y,\dot{y},\ddot{y}\right) = x$, where $\mathcal{N}\left(y,\dot{y},\ddot{y}\right)$ represents an arbitrary nonlinear function. The most common form of ODE of this form seen in practice is:

$$\frac{1}{\omega_n^2}\ddot{y} + \frac{2\zeta}{\omega_n}\dot{y} + y + \mathcal{N}\left(y,\dot{y}\right) = x + \dot{x} + \ddot{x}\,. \tag{3}$$

Providing measurements are taken across nonlinearities (which are often localised), this can be extended to systems with multiple degrees-of-freedom (extension to multiple degrees-of-freedom is part of ongoing work). Nonlinear dynamical systems of this type are common in many scientific disciplines and there has been significant work in modelling their response using data-based machine learning methods [18, 8, 20, 9]. However, it is extremely difficult using feedforward modelling techniques due to the exponentially growing number of nonlinear monomial terms which contribute to the output, which can be formulated using the Volterra series [16]. It is especially challenging for systems with long memory and for random broadband inputs. In this paper, the structure of the equation is exploited to estimate the nonlinear coherence for systems of this type, with an incomplete model. To the best of the author's knowledge, there are currently no other methods that achieve this.

2 Related Methods

3 Related Work

The presented method for calculating the level of noise present and inferring the maximum performance of feedforward models is novel. However, there are two broad areas of research in the literature that overlap with, and contribute to, the goals of this paper.

Control Within control, there are methods to estimate the level of noise and model error. The Kalman filter [6], and its extensions, aim to estimate the true state of a system by optimally combining the predictions of a state space model with noisy measurements. A comparison of the model error and noise level is used to balance the two signals optimally. However, when applied to the systems described in this paper, the noise levels are not observable because it is not possible to distinguish the contributions from the two sources. Prior knowledge of the relative noise level is therefore required, which is not available here. However, a similar method is utilised in this paper.

Nonlinear Coherence Granger causality is a method developed to determine whether one signal can be used to predict another [5]. A simple linear auto-regressive model is built to predict the next value in the output signal using past values of itself and the input signal. An F-test is then used to evaluate whether a simpler model predicts the target signal just as well as a more complex model. This is extended to model nonlinear relationships between signals in the extended Granger causality [1]. Mutual information is another method for determining the causal relationship between input and output signals [17]. It represents the reduction in uncertainty in one signal when the other is known. This is applicable when the relationship between the two signals is nonlinear. Mutual information between two variables X and Y is the information that provided by X about Y.

In both of these methodologies, to determine the link between the input and output the correlation between high-dimensional time series inputs and the output must be established. If this link can be established, then it is possible to build a simple forward model using classical methods such as the implicit Wiener series [4]. This paper focuses on the cases where these correlations cannot be established using classical methods. Taking the linear coherence between the response of the system and the prediction of a CNN is effectively a nonlinear coherence metric and will outperform classical methods such as mutual information. A CNN is therefore used as the benchmark for this purpose.

4 Methodology

4.1 Architecture

Consider the signals $x(t)$, $y_n(t)$ and $y_z(t)$. Here $x(t)$ is the input to the system, $y_n(t)$ is a noisy observation and $y_z(t)$ is the prediction of a forward model. The noise is represented by $\epsilon_n(t)$, which is assumed to be zero mean but otherwise can be of an arbitrary spectra and the true output is denoted $y(t)$. The time series data is split into N frames (input-output pairs), each of length M samples. In the following sections, lowercase denotes the time domain and uppercase denotes the frequency domain. In the i^{th} frame:

$$Y_z^{(i)}(f) = Y^{(i)}(f) + \mathcal{E}_z^{(i)}(f) \qquad\qquad Y_n^{(i)}(f) = Y^{(i)}(f) + \mathcal{E}_n^{(i)}(f)\ , \tag{4}$$

where $\mathcal{E}_n^{(i)}$ is assumed to be uncorrelated with $Y^{(i)}$, whereas $\mathcal{E}_z^{(i)}$ is correlated with $Y^{(i)}$.

A key observation of the structure of Eq. 3 is that whilst it is difficult to learn a mapping from x to y in the 'forward' direction, it is relatively simple to map from y to x in the 'reverse' direction, even for highly nonlinear systems with long memory. This is because there is no nonlinear mixing of terms in time in the reverse direction, and so there is not an infinite Volterra series [16] as there is in the forward direction. The input, x, can therefore be used as a reference signal to infer the true y. The signals y_z and y_n can be combined optimally to find the best estimation of y, and therefore x, based on their relative noise levels. Consider a linear combination of $Y_n^{(i)}$ and $Y_z^{(i)}$ at each frequency:

$$\hat{Y}^{(i)} = Y_n^{(i)} K + Y_z^{(i)} \left(1 - K\right)\ , \tag{5}$$

where $0 \leq K(f) \leq 1$. An inverse Fourier transform is applied to $\hat{Y}$ which is then used to train a machine learning model to predict x in the time domain. Figure 1 illustrates the proposed architecture. In this paper, $\mathcal{F}_\theta$ is a 1D convolutional neural network (CNN). Due to the simplicity of the reverse mapping from y to x, a relatively small CNN can be constructed, even when the system memory from x to y is relatively long. It is hypothesised that the optimal value of K will yield the optimal

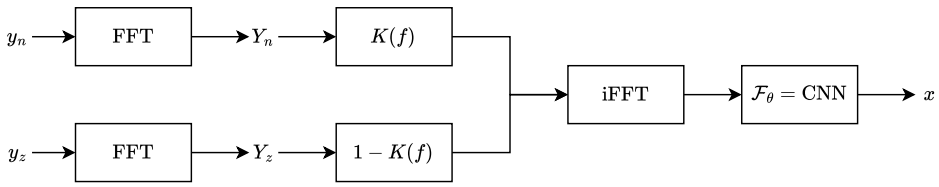

Figure 1: Architecture. The fast Fourier transform (FFT) of the signals y_z and y_n is taken, before the two signals are combined at each frequency point using the parameter $K(f)$. The inverse fast Fourier transform (iFFT) is then applied to the resultant signal, which is then mapped to x using a CNN.

estimate of the noise-free output Y. This optimum K corresponds to minimising:

$$J = E\left[(\hat{Y} - Y)(\hat{Y} - Y)^*\right], \tag{6}$$

because the optimal estimation of Y will also yield the best estimate of X. Minimising this expectation with respect to K, by setting $\frac{\partial J}{\partial K} = 0$, yields:

$$K = \frac{1}{1 + \frac{E[\mathcal{E}_n \mathcal{E}_n^*]}{E[\mathcal{E}_z \mathcal{E}_z^*]}}\ . \tag{7}$$

Therefore if the parameters of the model, θ and K, are trained simultaneously, the architecture will implicitly estimate the ratio of errors in order to optimally combine the two signals. This is used to calculate the nonlinear coherence. The nonlinear coherence between x and y_n is equal to the linear coherence between y and y_n and it can be shown that:

$$\gamma^2_{YY_n}(f) \equiv \frac{|S_{YY_n}|^2}{S_{YY}S_{Y_nY_n}} = \frac{E\left[YY^*\right]}{E\left[Y_nY_n^*\right]}\,, \tag{8}$$

where the notation $S_{AB} = E\left[AB^*\right]$ is the power spectrum density (PSD) between signals A and B at frequency f. The above estimation of K can be used to reverse engineer an estimation of $E\left[YY^*\right]$:

$$E\left[YY^*\right] \approx KE\left[Y_nY_n^*\right] - (1-K)\,E\left[Y_zY_z^*\right] + 2\,(1-K)\,\Re\{E\left[Y_nY_z^*\right]\}\,, \tag{9}$$

which can be calculated directly from the data using the inferred value of K. This is then substituted into Eq. 8 to estimate the nonlinear coherence. See appendix A.1 for a more detailed derivation.

5 Experimental Results

The method was evaluated on both a simulated and an experimental system, each with different forms of nonlinearities to demonstrate the applicability of the method. For each system, the robustness of the method was tested for different levels of noise. The noise level can be quantified by the nonlinear coherence: values close to 1.0 indicate low noise and values close to 0 indicate high noise. The predictions of the presented method were made using just 10 frames of data.

To benchmark the performance of the algorithm, the predictions of the nonlinear coherence are compared against those of the forward model, denoted $\mathrm{Co}\,(Y_z, Y_n)$, and a linearisation of the system, denoted $\mathrm{Co}\,(X, Y_n)$. The forward model, described in Section 4.1, was chosen to be a 1D temporal convolutional neural network with dilations, as described by van den Oord et al. [15]. The prediction of this forward model was used as y_z to demonstrate the improved estimation of the nonlinear coherence that can be calculated using the method presented in this paper. The generated y_z is therefore used both to provide a baseline and also to estimate the nonlinear coherence. It is likely that better forward models could be trained both now and in the future for the examples shown; the aim is to demonstrate how the presented method utilises the forward model to improve the estimation of the nonlinear coherence, and so it is not benchmarked against the forward prediction in the traditional sense and the selection of a forward model is somewhat arbitrary.

5.1 Polynomial stiffness

Systems with quadratic and cubic nonlinearities appear in many mechanical [12, 19] and biological systems [10], and y^2 and y^3 are the next terms in the Taylor expansion of the nonlinearity of any system. This example therefore has wide ranging real world applications. The following dimensionless nonlinear oscillator was simulated to generate 1000 frames of length 6000:

$$\ddot{y} + 2\zeta\dot{y} + y + \alpha_2 y^2 + \alpha_3 y^3 = x\,. \tag{10}$$

The input, x, was bandlimited random noise with a fixed rms τ. The spectrum of x was effectively flat in the range of interest. The hyperparameters were set as defined in Appendix B.1.

The linearised response of the system captured 88 % of the total system response and a CNN (kernel width=20, hidden layers = 5, features=10, trained for 1000 epochs with a learning rate of 0.01) captured 94% of the response using 10 frames of data. Without the method presented in this paper, with just 10 frames of data it would be impossible to determine how much of the unmodelled response is due to noise and how much is due to nonlinearities. Figure 2 shows the true nonlinear coherence (solid black line), the prediction of the CNN in the forward direction (blue dotted line) and the linear coherence between x and y_n (solid green line). The more noise that was added, the lower the true nonlinear coherence because the causal relationship between x and y_n is weaker. The predicted nonlinear coherence was calculated using 10 frames of data, which is represented by the red dashed line. The method presented in this paper provides an excellent prediction of the nonlinear coherence.

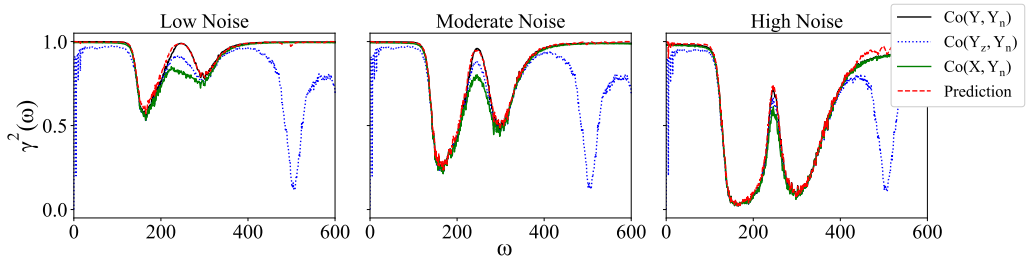

Figure 2: Polynomial stiffness case study.

5.2 Experiment

The method was then evaluated on an experimental dataset consisting of 360 frames of length 6000. A cantilever beam was excited through a nonlinear connection using a bandlimited random input. The nonlinear connection consisted of magnets (the cantilever is steel) and a rubber tip which caused rattling. This nonlinear connection is complex and has not been characterised. The set up was considered to be effectively noiseless and then three levels of additional noise were introduced by post-processing to represent uncorrelated noise in the system. Two minutes of data (10 frames) was used to generate the forward prediction and then predict the nonlinear coherence.

Given the nonlinearity in the system and small quantity of data captured, in the noise-free case only 45 % of the response was captured by the linearised component of the response, and a CNN (kernel width=10, hidden layers = 5, features=5) captured just 74 % due to the complexity of the system response.

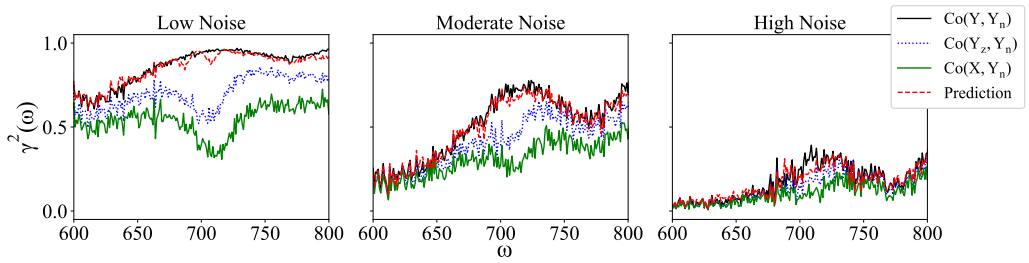

Figure 3: Experimental case study.

Figure 3 shows the predicted and true nonlinear coherence for three levels of noise. For all three levels of noise, the predicted nonlinear coherence was an excellent estimation of the true nonlinear coherence. In the low noise case, it demonstrates that significant improvements in the prediction could be achieved by collecting more resources and building larger models. Conversely, in the high noise case, the potential improvements are relatively small compared with the overall noise level.

6 Conclusions

This paper presents a novel method for calculating the nonlinear coherence for a nonlinear dynamical system in the presence of noise. A set of parameters are learnt to optimally combine an output prediction, calculated using an available model, with noisy measurements of the output to predict the input to the system. The nonlinear coherence is calculated using these parameters and is used as a metric of causality. Previously, the only way to estimate this was to model the system and assume the unmodelled component of y to be noise. The presented method improves upon this and is able to estimate the nonlinear coherence with excellent accuracy using a relatively small quantity of data. This would allow for researchers to collect experimental input and output data for a particular nonlinear dynamical system, and identify how much of the output is caused by the input across the frequency range, without developing expensive models.

Acknowledgements

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Partnership (DTP) PhD studentship. The authors would also like to thank Bose Corporation for their additional support and Professor Robin Langley for the many helpful discussions.

Reproduciblity Statement

The code and datasets required to reproduce the results in this paper are shared at https://github.com/obj22/repo-1. The hyperparameter settings for the simulations are provided in the appendix.

References

[1] Yonghong Chen, Govindan Rangarajan, Jianfeng Feng, and Mingzhou Ding. Analyzing multiple nonlinear time series with extended granger causality. *Physics Letters A*, 324:26–35, 04 2004. doi: 10.1016/j.physleta.2004.02.032.

[2] Fernando Corinto and Alessandro Torcini. *Nonlinear dynamics in computational neuroscience*. Springer, 2019. ISBN 9783319710471.

[3] M. De Brett. *Implications of nonlinear suspension behaviour for feedforward control of road noise in cars*. PhD thesis, University of Cambridge, 2020.

[4] M. Franz and B. Schölkopf. Implicit wiener series – part i: Cross-correlation vs. regression in reproducing kernel hilbert spaces. Technical report, Max Planck Institute, 2003.

[5] C.W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969. doi: 10.2307/1912791.

[6] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[7] Steven M. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

[8] S. Liang, Q. Zhu, and M. Ishitobi. Identification of duffing's equation with dynamic recurrent neural network. In *Advances in Neural Networks – ISNN 2005*, pp. 454–459. Springer Berlin Heidelberg, 2005. ISBN 9783540320678.

[9] Joseph Massingham, Ole Nielsen, and Tore Butlin. An efficient method for generalised wiener series estimation of nonlinear systems using gaussian processes. *Mechanical Systems and Signal Processing*, 209:111095, 2024. ISSN 0888-3270. doi: https://doi.org/10.1016/j.ymssp.2023.111095. URL `https://www.sciencedirect.com/science/article/pii/S0888327023010038`.

[10] L. Miller. Structural dynamics and resonance in plants with nonlinear stiffness. *Journal of Theoretical Biology*, 234(4):511–524, 2005. ISSN 0022-5193. doi: https://doi.org/10.1016/j.jtbi.2004.12.004. URL `https://www.sciencedirect.com/science/article/pii/S0022519304006083`.

[11] Erik Mosekilde. *Topics in Nonlinear Dynamics*. World Scientific, 1997. doi: 10.1142/3194. URL `https://www.worldscientific.com/doi/abs/10.1142/3194`.

[12] A.H. Nayfeh and D.T. Mook. *Nonlinear Oscillations*. Wiley, 1979. ISBN 9780471035558.

[13] Chi Oh, Kang-duck Ih, Hae Lee, and Jin Jeon. Development of an active road noise control system. *ATZ worldwide*, 117, 01 2015. doi: 10.1007/s38311-015-0151-5.

[14] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, third edition, 2007. ISBN 9780521880688.

[15] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alexander Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. In *Arxiv*, 2016. URL `https://arxiv.org/abs/1609.03499`.

[16] V. Volterra. *Theory of functionals and of integral and integro-differential equations*. Dover Publications, 1959.

[17] Janett Walters-Williams and Yan Li. Estimation of mutual information: A survey. In Peng Wen, Yuefeng Li, Lech Polkowski, Yiyu Yao, Shusaku Tsumoto, and Guoyin Wang (eds.), *Rough Sets and Knowledge Technology*, pp. 389–396, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-02962-2.

[18] W Wang, Y Lai, and C Grebogi. Data based identification and prediction of nonlinear and complex dynamical systems. *Physics Reports*, 644:1–76, 2016. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2016.06.004.

[19] K. Worden and G. Tomlinson. *Nonlinearity in structural dynamics : detection, identification, and modelling*. Institute of Physics, 2001. ISBN 0750303565.

[20] K. Worden, W. Becker, T. Rogers, and E. Cross. On the confidence bounds of gaussian process narx models and their higher-order frequency response functions. *Mechanical Systems and Signal Processing*, 104:188–223, 2018. ISSN 0888-3270. doi: https://doi.org/10.1016/j.ymssp.2017.09.032.

A Architecture

A.1 Derivation

The architecture is driven towards de-noising $\hat{Y}$ by optimally combining Y_n and Y_z.

$$J = E\left[(\hat{Y} - Y)(\hat{Y} - Y)^*\right] \tag{11}$$

$$= E\left[(Y_n K + Y_z\left(1 - K\right) - Y)(Y_n K + Y_z\left(1 - K\right) - Y)^*\right] \tag{12}$$

$$= K^2 E\left[\mathcal{E}_n \mathcal{E}_n^*\right] + (1 - K)^2 E\left[\mathcal{E}_z \mathcal{E}_z^*\right]. \tag{13}$$

Then, differentiating with respect to K gives

$$\frac{dJ}{dK} = 2KE\left[\mathcal{E}_n \mathcal{E}_n^*\right] - 2(1 - K)E\left[\mathcal{E}_z \mathcal{E}_z^*\right] = 0\,. \tag{14}$$

Therefore rearranging for K gives

$$K = \frac{1}{1 + \frac{E[\mathcal{E}_n \mathcal{E}_n^*]}{E[\mathcal{E}_z \mathcal{E}_z^*]}}\,. \tag{15}$$

The nonlinear coherence between x and y_n is equal to the linear coherence between y and y_n:

$$\gamma_{YY_n}^2 = \frac{|S_{YY_n}|^2}{S_{YY} S_{Y_n Y_n}}\,, \tag{16}$$

where $S_{Z_1 Z_2} = E\left[Z_1 Z_2^*\right]$ represents the power spectrum density (PSD) between signals Z_1 and Z_2 at frequency f. Assuming $\mathcal{E}_n$ and Y are uncorrelated, the following can be written.

$$S_{YY_n} = E\left[YY_n^*\right] = E\left[YY^*\right] \tag{17}$$

$$S_{Y_n Y_n} = E\left[Y_n Y_n^*\right] \tag{18}$$

$$S_{YY} = E\left[YY^*\right]\,. \tag{19}$$

Therefore

$$\gamma_{YY_n}^2 = \frac{E\left[YY^*\right]}{E\left[Y_n Y_n^*\right]}\,. \tag{20}$$

Note that $E\left[Y_n Y_n^*\right]$ is not simplified further because this quantity can be measured from data. An estimation of $E\left[YY^*\right]$ is required. The estimation of K, using the architecture described in Section

4.1, can be used to reverse engineer an estimation of $E[YY^*]$. The ratio between the error in each signal is given as

$$C \equiv \frac{E[\mathcal{E}_z\mathcal{E}_z^*]}{E[\mathcal{E}_n\mathcal{E}_n^*]} = \frac{K}{1-K} \ . \tag{21}$$

Therefore

$$\begin{aligned} CE[\mathcal{E}_n\mathcal{E}_n^*] &= E[\mathcal{E}_z\mathcal{E}_z^*] && (22)\\ &= E[(Y-Y_z)(Y-Y_z)^*] && (23)\\ &= E[YY^*] - 2\Re\{E[YY_z^*]\} + E[Y_zY_z^*] && (24) \end{aligned}$$

In addition to this equation, three quantities can be calculated directly from the data: $E[Y_nY_n^*]$, $E[Y_zY_z^*]$ and $E[Y_zY_n^*]$. Noting that $\mathcal{E}_z$ is correlated with Y because it includes model error, two equations can then be written:

$$\begin{aligned} E[Y_nY_n^*] &= E[YY^*] + E[\mathcal{E}_n\mathcal{E}_n^*] && (25)\\ E[Y_zY_n^*] &= E[Y_zY^*] \ . && (26) \end{aligned}$$

Combining Eq. 24 and Eq. 25 to eliminate $E[\mathcal{E}_n\mathcal{E}_n^*]$ gives

$$C\left(E[Y_nY_n^*] - E[YY^*]\right) = E[YY^*] - 2\Re\{E[YY_z^*]\} + E[Y_zY_z^*] \ . \tag{27}$$

Solving for $E[YY^*]$ using Eq. 26 and Eq. 27 gives

$$E[YY^*] \approx \frac{CE[Y_nY_n^*] - E[Y_zY_z^*] + 2\Re\{E[Y_nY_z^*]\}}{C+1} \ , \tag{28}$$

which can be calculated directly from the data using the inferred value of K. This is then substituted into Eq. 20 to estimate the nonlinear coherence.

B Dataset

The dataset is split into training, validation and test sets. 10 frames were used for training, 10 frames were used for validation, and remaining frames were used for the test set. A batch size of 1 frame was used for training using the Adam optimiser.

B.1 Polynomial Stiffness Case Study

The dataset contains 1000 frames of length 6000 samples. The parameters were set as: $\zeta = 4.5$, $\alpha_1 = 5 \times 10^3$, $\alpha_2 = -10$ and $\alpha_3 = 3 \times 10^3$. The input, x, was bandlimited noise with an rms of $\tau = 1.0 \times 10^3$. Fourth order Runge-Kutta numerical integration was used in order to calculate the benchmark true y using x [14]. A step of 0.0025 was used. Figure 4 shows the power spectrum density (PSD) for the system response and the linearised response.

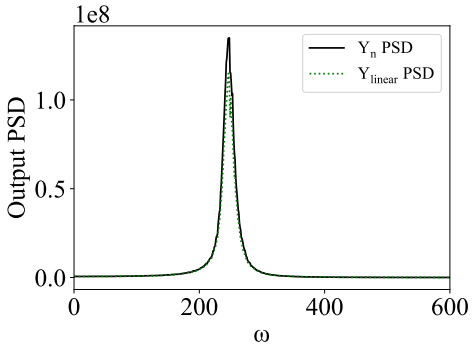

Figure 4: Plot of the power spectrum density of Y_n and Y_z for the polynomial stiffness case study with no noise added.

B.2 Experiment Case Study

The dataset contains 180 frames of data of length 6000. A sampling frequency of 500 Hz was used to record the data. Figure 5 shows the power spectrum density (PSD) for the system response and the linearised response.

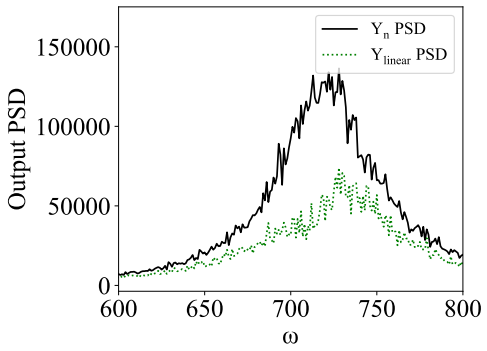

Figure 5: Plot of the power spectrum density of Y_n and Y_{linear} for the experimental data with no noise added.

C Computational Resources

The device used to run the code in this paper had the following specifications:
Processor: 12th Gen Intel(R) Core(TM) i7-12700KF 3.61 GHz
RAM: 64.0 GB
GPU: NVIDIA GeForce RTX 3090