



X International Conference on Structural Dynamics, EURODYN 2017

Uncertainty Bounds on Higher-Order FRFs from Gaussian Process NARX Models

Keith Worden^{a,*}, Cecilia Surace^b, William Becker^c

^a*Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK*

^b*Politecnico di Torino, Department of Structural, Building and Geotechnical Engineering, Corso Duca degli Abruzzi, 24, 10129, Turin, Italy*

^c*European Commission, Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy*

Abstract

One of the most versatile and powerful algorithms for the identification of nonlinear dynamical systems is the NARMAX (Non-linear Auto-regressive Moving Average with eXogenous inputs) approach. The model represents the current output of a system by a nonlinear regression on past inputs and outputs and can also incorporate a nonlinear noise model in the most general case. In recent papers, one of the authors introduced a NARX (no noise model) formulation based on Gaussian Process (GP) regression and derived the corresponding expressions for Higher-order Frequency Response Functions (HFRFs). This paper extends the theory for the GP-NARX framework by providing a means of converting the GP prediction bounds in the time domain into bounds on the HFRFs. The approach is demonstrated on the Duffing oscillator.

© 2017 The Authors. Published by Elsevier Ltd.

Peer-review under responsibility of the organizing committee of EURODYN 2017.

Keywords: Structural dynamics; frequency response function; uncertainty; Gaussian process; Duffing oscillator

1. Introduction

NARMAX (Nonlinear Auto-Regressive Moving Average with eXogenous inputs) models are powerful tools for nonlinear system identification, which accommodate nonlinear discrete-time processes *and* noise models [1,2]. If, however, the noise process is assumed to be Gaussian, the simpler NARX model can be used. NARX models are of interest in system identification because, through a connection with the Volterra series, they allow the construction of Higher-order Frequency Response Functions (HFRFs) that allow one to visualise how different frequencies in the input to a nonlinear system interact in forming the output [3]. The HFRFs are obtained through an algorithm called *harmonic probing* - proved to be a simple extension of the long-held algorithm for differential equations [4,5].

NARX models predict the current value of the system output using a nonlinear function F of previous inputs and outputs, i.e.

$$y_i = F(y_{i-1}, \dots, y_{i-n_y}; x_{i-1}, \dots, x_{i-n_x}) \quad (1)$$

* Corresponding author

E-mail address: keith.worden@sheffield.ac.uk

The function F can be taken from any number of regression approaches: most commonly it is multinomial regression, but nonparametric model forms based on machine learning have also been adopted including Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) neural networks [6,7]. A recent addition to the literature of the NARX model was the discussion of the Gaussian Process (GP) NARX model in [8]. This model form allows a number of potential advantages over the existing common forms on the NARX model, including the generation of natural confidence intervals for model predictions. In [9], the expressions for the HFRFs for the GP-NARX model were explicitly stated, based on earlier work in [10].

In this paper the GP-NARX framework is taken a step further: to date, the uncertainty of the GP fit to the training data was only propagated as far as predictions in the time domain. Here, a Monte Carlo approach is presented which allows the uncertainty to be expressed on the HFRFs themselves.

2. Gaussian Process NARX models

2.1. Gaussian Processes

The basic premise of GPs is to perform inference over *functions* directly, as opposed to inference over *parameters* of functions. In short, a GP is a distribution over functions, which is conditioned on training data so that the most probable functions are the best fits to the data. This accounting of uncertainty is exploited in this paper to build a distribution over NARX models, and subsequently a distribution over HFRFs.

Let $X = [\underline{x}_1, \underline{x}_2 \dots \underline{x}_N]^T$ denote a matrix of multivariate training inputs, and y denote the corresponding vector of training outputs. The input vector for a testing point will be denoted by the column vector \underline{x}^* and the corresponding (unknown) output by y^* . A Gaussian process prior is formed by assuming a (Gaussian) distribution over *functions*,

$$f(\underline{x}) \sim \mathcal{GP}(m(\underline{x}), k(\underline{x}, \underline{x})) \quad (2)$$

where $m(\underline{x})$ is the *mean function* and $k(\underline{x}, \underline{x}')$ is a positive-definite *covariance function*.

One of the defining properties of the GP is that the density of a finite number of outputs from the process, both observed and unobserved, is multivariate normal. This property, combined with standard results for Gaussian distributions, can be used to condition unobserved points on observed training points: this mechanism effectively fits the GP to the training data.

Following a Bayesian approach, the prior mean is assumed to be zero (see [11] for a discussion). Assuming a Gaussian noise model with variance σ_n^2 , the joint distribution for training and testing values is,

$$\begin{pmatrix} y \\ y^* \end{pmatrix} \sim \mathcal{N}\left(\underline{0}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, \underline{x}^*) \\ K(\underline{x}^*, X) & K(\underline{x}^*, \underline{x}^*) + \sigma_n^2 \end{bmatrix}\right) \quad (3)$$

where $K(X, X)$ is a matrix whose i, j^{th} element is equal to $k(\underline{x}_i, \underline{x}_j)$. Similarly, $K(X, \underline{x}^*)$ is a column vector whose i^{th} element is equal to $k(\underline{x}_i, \underline{x}^*)$, and $K(\underline{x}^*, X)$ is the transpose of the same.

In order to make use of the above, it is necessary to re-arrange the joint distribution $p(y, y^*)$ into a conditional distribution $p(y^*|y)$. Using standard results for the conditional properties of a Gaussian reveals [11],

$$y^* \sim \mathcal{N}(m^*(\underline{x}^*), k^*(\underline{x}^*, \underline{x}^*)) \quad (4)$$

where,

$$m^*(\underline{x}^*) = K(\underline{x}^*, X)[K(X, X) + \sigma_n^2 I]^{-1} y \quad (5)$$

is the *posterior mean* of the GP and,

$$k^*(\underline{x}^*, \underline{x}') = k(\underline{x}^*, \underline{x}') - K(\underline{x}^*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, \underline{x}') \quad (6)$$

is the *posterior variance*.

Thus the GP model provides a posterior distribution for the unknown quantity y^* . The mean from equation (4) can then be used as a ‘best estimate’ for a regression problem, and the variance can also be used to define confidence intervals. The covariance function used here is the squared-exponential function, and its hyperparameters, along with the noise parameter σ_n^2 , can be readily found through maximum likelihood estimation (see [9]). For considerably more details on GPs than this short paper allows, see [11].

2.2. GP NARX Models

To apply GPs in the NARX framework, one simply performs a GP regression of the form in Equation (1). After fitting the GP, one way to make predictions is to compute *one step ahead* (OSA) predictions, which exclusively use the training data up to that time, i.e.

$$y_i^* = F(y_{i-1}, \dots, y_{i-n_y}; x_{i-1}, \dots, x_{i-n_x}) \tag{7}$$

This can be compared with predicted outputs. However, a more demanding test is to compute the *Model Predicted Output* (MPO), which uses predicted y^* values instead of observed y_t values. It is defined by,

$$y_i^* = F(y_{i-1}^*, \dots, y_{i-n_y}^*; x_{i-1}, \dots, x_{i-n_x}) \tag{8}$$

and this test can be conducted on testing data as well as training data, which is an important consideration in the more general context of machine learning.

In [9] it was noted that the confidence intervals provided by the GP when computing the MPO were very small because they did not accommodate the observed prediction errors. To fully account for the prediction uncertainty, during a prediction run, at each instant i the prediction y_i^* should be sampled from the distribution specified by the mean and covariance from the GP for that instant. One such run therefore generates a single realisation of the prediction process. In order to accumulate information about the distribution of predictions with state estimation taken into account, a Monte Carlo approach can be adopted, such that R runs are executed to build an estimation of the distribution over GP-NARX predictions.

3. Higher-order FRFs through harmonic probing

As described in [9], the response spectrum of a nonlinear system at an excitation frequency ω can be completely characterised by the Volterra series, such that,

$$Y(\omega) = Y_1(\omega) + Y_2(\omega) + Y_3(\omega) + \dots \tag{9}$$

where,

$$Y_1(\omega) = H_1(\omega)X(\omega) \tag{10}$$

$$Y_2(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega_1 H_2(\omega_1, \omega - \omega_1) X(\omega_1) X(\omega - \omega_1) \tag{11}$$

$$Y_3(\omega) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\omega_1 d\omega_2 H_3(\omega_1, \omega_2, \omega - \omega_1 - \omega_2) X(\omega_1) X(\omega_2) X(\omega - \omega_1 - \omega_2) \tag{12}$$

where the H_i are the *Higher-order Frequency Response Functions* (HFRFs). The interpretation of these quantities is well established: a description can be found in [3]. If the equations of motion are known for a system, the method of harmonic probing can be used in order to compute the HFRFs [4]. In the case of Gaussian process NARX models, the expressions related to harmonic probing were derived in [9]. Here they are briefly summarised.

The key to deriving these expressions is to state the GP in its parametric form, as an expansion in terms of basis functions fixed by the covariance kernel and the training data [11]. This gives the predicted output y^* corresponding to a new input \underline{x}^* as,

$$y^* = \sum_{i=1}^N a_i k(\underline{x}^*, \underline{x}_i) \tag{13}$$

where according to equation (5),

$$\underline{a} = [k(X, X) + \sigma_n^2 I]^{-1} \underline{y} \tag{14}$$

and this is fixed by the training data. In [9] it was shown that this parametric form, with some rearrangement, allows harmonic probing expressions to be derived. For H_1 this results in,

$$H_1(\omega) = \frac{\sum_{j=0}^{n_x} \beta_j e^{-ij\Delta t\omega}}{1 - \sum_{j=1}^{n_y} \alpha_j e^{-ij\Delta t\omega}} \quad (15)$$

where,

$$\alpha_j = \frac{\sigma_f^2}{\rho^2} \sum_{i=1}^{N-p} a_i \gamma_i v_{ij}, \quad \beta_j = \frac{\sigma_f^2}{\rho^2} \sum_{i=1}^{N-p} a_i \gamma_i u_{ij} \quad (16)$$

v_{ij} is the i, j th element of the matrix formed by the first n_y columns of X , and u_{ij} is the i, j th element of the matrix formed by the remaining $n_x + 1$ columns, and,

$$\gamma_j = \exp \left\{ -\frac{1}{2\rho^2} \left[\sum_{k=1}^{n_y} v_{jk}^2 + \sum_{m=0}^{n_x} u_{jm}^2 \right] \right\} \quad (17)$$

For H_2 , the expressions involve similar terms, but are lengthy and will not be stated here; they are given in full detail in [9].

In the GP-NARX framework to date, the potential of the Gaussian process has not been fully exploited, because the GP provides a stochastic fit to a set of training data, but so far this stochasticity has only been propagated as far as the time domain. Ideally one would like to propagate the uncertainty all the way to the end product of the exercise: the HFRFs.

To achieve this end, a Monte Carlo framework is adopted here. The best estimate of uncertainty in the time domain is represented by the Monte Carlo predictions in Figure 1. Importantly, each of the lines in this figure represents a draw from a single underlying Gaussian process, which has a single set of α_j and β_j coefficients. In order to propagate the uncertainty to the frequency domain, the approach here is to treat each Monte Carlo draw as a new set of training data, and fit a new GP-NARX model to each draw. Each new GP-NARX model comes with its own α_j and β_j , each of which can be used to generate a draw of the HFRFs. The resulting distributions over HFRFs can be used to build mean HFRFs and confidence intervals.

Let $y_{r,i}^*$, $i = 1, \dots, N$, $r = 1, \dots, R$ represent the GP prediction at x_i , corresponding to the r^{th} Monte Carlo draw. For each of the R Monte Carlo draws, a new GP-NARX model is fitted, such that,

$$y_{i,r}^{**} = F(y_{r,i-1}^*, \dots, y_{r,i-n_y}^*; x_{i-1}, \dots, x_{i-n_x}) \quad (18)$$

where $y_{i,r}^{**}$ represents the GP-NARX prediction corresponding to the r^{th} MC draw at x_i . Notice that the GP is fitted using the OSA approach, because the interest is only in obtaining the corresponding α_j and β_j parameters. To simplify matters, it is also assumed that the hyperparameters of each GP are all equal to the hyperparameters estimated in the original GP.

From here, the procedure is straightforward. For the r^{th} GP fit, the corresponding $\alpha_{j,r}$ and $\beta_{j,r}$ values can be obtained using the expressions in (16), via the parameters of the GP. These can then be translated into a draw from the HFRF, resulting in a total of R samples from the distribution of the HFRF. This easily yields statistics of the distribution, such as pointwise means, modes and confidence intervals.

4. Case study - an asymmetric Duffing oscillator

In order to illustrate the use of the GP NARX formulation, data simulated from a Duffing oscillator system will be used. In the asymmetric case when a quadratic stiffness is present, the relevant equation of motion is,

$$m\ddot{y} + c\dot{y} + ky + k_2y^2 + k_3y^3 = x(t) \quad (19)$$

Data were simulated by integrating the equation of motion using a fourth-order fixed step Runge-Kutta algorithm [12]. The parameters adopted were $m = 1$, $c = 20$, $k = 10^4$, $k_2 = 10^7$ and $k_3 = 5 \times 10^9$. The excitation used was

zero mean Gaussian random sequence with a standard deviation of 2.0. The time step used was $\Delta t = 0.001$ seconds corresponding to a sampling frequency of 1 kHz. Gaussian noise of 1% RMS of the signal was added to the response time data. The number of lags for the model used here were $n_y = 3$ and $n_x = 3$. The training data consisted of 1000 samples of input and output data.

Figure 1 shows the 50 realisations of the Monte Carlo draws from the GP-NARX model (due to space limitations, the OSA and MPO prediction plots are not shown here). Since these are OSA predictions for each realisation with a high density of training data, the GP fits almost perfectly to each data series. Collectively, there is a moderate degree of uncertainty in fitting the GP-NARX model to the underlying training data, as evidenced by the spread of fits. As noted in [9], the dominant contribution to uncertainty in the predictions is not the direct component from the parameter uncertainty, but the indirect component due to state estimation from the uncertain parameters.

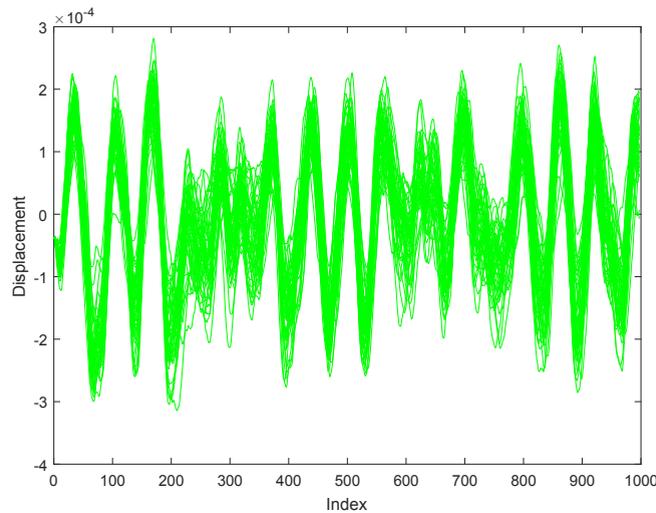


Fig. 1: MC realisations of predictions for GP NARX model of Duffing oscillator data.

In order to propagate the uncertainty in Figure 1 into the frequency domain, a GP-NARX model is fit to each of the 50 Monte Carlo draws, yielding draws of the HFRF. At each frequency, the mean and standard deviation can be estimated over the 50 draws. The result is in Figure 2 which plots the mean and 99% confidence intervals. It is evident that the uncertainty is actually quite narrow in the magnitude plot: although the time domain fits show some scatter, the underlying frequencies contained within each signal are quite similar. In the phase plot, the uncertainty is also quite small.

Figure 3(a) shows the mean H2 HFRF over the 50 draws, with the uncertainty (variance) at each point represented by the colour map. The response is flat at normalised frequencies above about 0.2, and the uncertainty is correspondingly very low. At the lowest frequencies the magnitude increases sharply and the uncertainty is also considerably higher. This is essentially a reflection of what is already visible in Figure 2, with the additional information that the interaction between the two frequencies is somewhat negligible. The corresponding phase plot in Figure 3(b) shows a similar picture: low uncertainty over the large range of frequencies explored, but higher uncertainty in the high-phase regions at the lowest frequencies.

In order to see the important structure in the HFRFs, it is often sufficient to plot only the leading diagonal i.e. $H_2(\omega, \omega)$: see Figure 4. This format also allows simple comparisons between the functions. This gives extra information over the H1 plots alone, because the interaction between ω_1 and ω_2 is clearly visible in the second peak of the plot. The confidence intervals, while wide, reveal that this second peak is not merely due to a spurious fit of the NARX model, because it is present in some way over all Monte Carlo iterations.

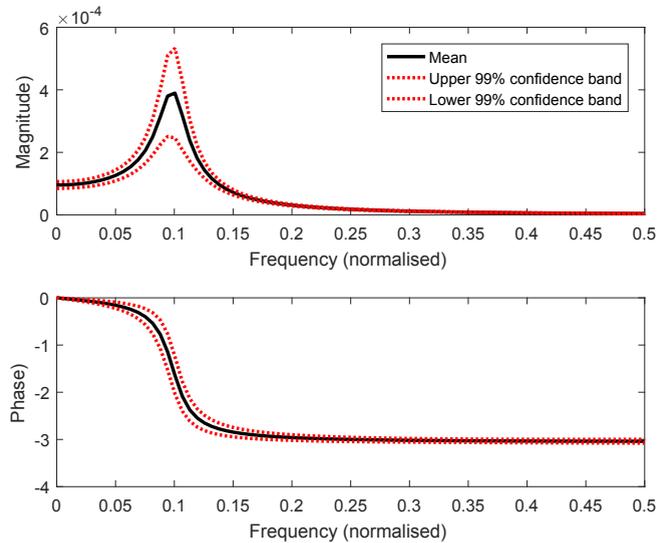


Fig. 2: H1 plotted with mean values over $R = 50$ draws from the GP, and 99% confidence intervals.

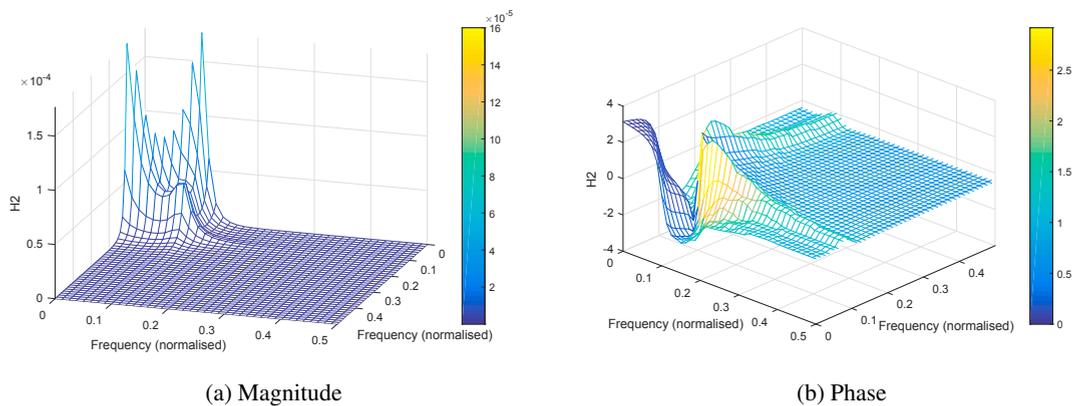


Fig. 3: H2 magnitude and phase plotted with mean values over $R = 50$ draws from the GP; colour map indicating standard deviation.

5. Conclusions

This article has continued and extended previous work relating to GP-NARX models, by fully propagating uncertainty into the frequency domain. Using a simple Monte Carlo approach, the draws of the GP-NARX fit to a set of training data were converted into draws of the HFRF by fitting a further GP to each set of points. This leads to the distribution of the HFRF, taking into account the uncertainty in the original GP fit, and can be used to generate mean or mode estimates, and confidence intervals on HFRFs.

Acknowledgements

KW would like to acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC) through grant reference number EP/K003836/2.

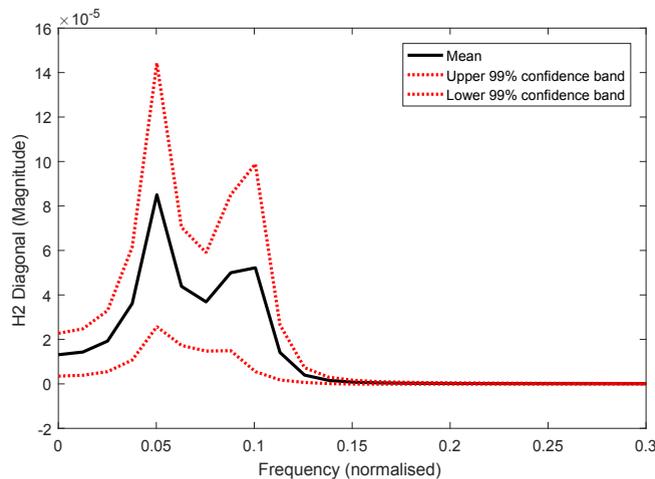


Fig. 4: H2 magnitude at $\omega_1 = \omega_2$: mean values over $R = 50$ draws from the GP and 99% confidence intervals.

References

- [1] I.J. Leontaritis, S.A. Billings, *Input-output parametric models for nonlinear systems, part I: deterministic nonlinear systems*, International Journal of Control, 41, (1985), pp.303-328.
- [2] I.J. Leontaritis, S.A. Billings, *Input-output parametric models for nonlinear systems, part II: stochastic nonlinear systems*, International Journal of Control, 41, (1985), pp.329-344.
- [3] K. Worden, G.R. Tomlinson, *Nonlinearity in Structural Dynamics: Detection, Modelling and Identification*, Institute of Physics, (2001).
- [4] E. Bedrosian, S.O. Rice, *The output properties of Volterra systems driven by harmonic and Gaussian inputs*, Proceedings IEEE, 59, (1971), pp.1688-1707.
- [5] S.A. Billings, K.M. Tsang, *Spectral analysis for nonlinear systems, part I: parametric non-linear spectral analysis*, Mechanical Systems and Signal Processing, 3, (1989), pp.319-339.
- [6] S.A. Billings, H.B. Jamaluddin, S. Chen, *Properties of neural networks with applications to modelling non-linear dynamical systems*, International Journal of Control, 55, (1992), pp.193-224.
- [7] S. Chen, S.A. Billings, C.F.N. Cowan, P.M. Grant, *Practical identification of NARMAX models using radial basis functions*, International Journal of Control, 52, (1990), pp.1327-1350.
- [8] K. Worden, J.J. Hensman, *Bayesian inference and nonlinear system identification*, In Proceedings of ICEDyn 2011 - 3rd International Conference on Engineering Dynamics, Tavire, Portugal, (2011).
- [9] K. Worden, G. Manson, E. J. Cross *Higher-order frequency response functions from Gaussian process NARX models*, In Proceedings of The 25th International Conference on Noise and Vibration Engineering, Leuven, Belgium, (2012).
- [10] J.E. Chance, K. Worden, G.R. Tomlinson, *Frequency domain analysis of NARX neural networks*, Journal of Sound and Vibration, 213, (1997), pp.915-941.
- [11] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning, (Adaptive Computation and Machine Learning)*, MIT Press, (2005).
- [12] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing, Third Edition*, Cambridge University Press, (2007).