Mechanical Systems and Signal Processing 104 (2018) 188–223

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

On the confidence bounds of Gaussian process NARX models and their higher-order frequency response functions

K. Worden [a,*], W.E. Becker [b], T.J. Rogers [a], E.J. Cross [a]

[a] Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom
[b] European Commission, Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra, VA, Italy

ARTICLE INFO

Article history:
Received 20 May 2017
Received in revised form 16 September 2017
Accepted 23 September 2017

Keywords:
Nonlinear system identification
Gaussian processes
NARX models
Wave loading
Morison's equation

ABSTRACT

One of the most powerful and versatile system identification frameworks of the last three decades is the NARMAX/NARX[1] approach, which is based on a nonlinear discrete-time representation. Recent advances in machine learning have motivated new functional forms for the NARX model, including one based on Gaussian processes (GPs), which is the focus of this paper. Because of their nonparametric form, NARX models can only provide physical insight through their frequency-domain connection to Higher-order Frequency Response Functions (HFRFS). Because of the desirable properties of the GP-NARX form (no structure detection needed, natural confidence intervals), the analytical derivation of the HFRFs for the model is presented here for the first time. Furthermore, an algorithm for propagating uncertainty from the GP into the HFRF estimates is presented. A valuable by-product of the latter algorithm is a new test for nonlinearity, capable of detecting the presence of odd and even system nonlinearities. The new results are illustrated via two case studies; the first is based on simulation of an asymmetric Duffing oscillator. The second case study presents a validation of the new theory in the area of wave force prediction on offshore structures. This problem is one that has been considered by some of the authors before; the current paper takes the opportunity to highlight and correct a number of weaknesses of the original study in the light of modern best practice in machine learning.

1. Introduction

Over the last 30 years, one of the most versatile and enduring time series models used for nonlinear system identification has been the NARMAX (Nonlinear Auto-Regressive Moving Average with eXogenous inputs) model. The NARMAX model was introduced in 1985 [1,2] and has been the subject of constant interest and development since. (A comprehensive monograph on the theory and applications of the model recently appeared in [3]). The most general model form accommodates nonlinear discrete-time process *and* noise models. However, if the noise process can be assumed to be white Gaussian, the simpler NARX model can be adopted, and this will be the focus of this paper. The NARX model assumes a form whereby the current value of the system output is predicted using a nonlinear function F of previous inputs and outputs, i.e.

$$y_i = F(y_{i-1}, \dots, y_{i-n_y}; x_i, \dots, x_{i-n_x+1}) + \epsilon_i \qquad (1)$$

* Corresponding author.
E-mail address: k.worden@sheffield.ac.uk (K. Worden).

[1] Nonlinear Auto-Regressive Moving Average with eXogenous inputs; NARX if the moving average noise model is omitted.

https://doi.org/10.1016/j.ymssp.2017.09.032
0888-3270/© 2017 The Authors. Published by Elsevier Ltd.

where the *residual sequence* ϵ_i is white Gaussian. The number of output (resp. input) lags is denoted n_y (resp. n_x). This formulation of a NARX model differs a little from the original [1,2], in that it allows the use of the present input x_i.

The earliest and still most common form of the NARX model adopts a multivariate polynomial (multinomial) expansion basis for the function F and learns the expansion coefficients/parameters by using linear (but advanced) least-squares methods. However, in general, *any* expansion basis which satisfies a universal approximation property can be used. This observation has led to various nonparametric NARX model forms based on machine learning, including Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) neural networks [4,5]. Most recently, variations on the NARX structure based on Polynomial Chaos Expansions (PCE-NARX) have emerged [6]. By utilising parameters that are random variables themselves, the PCE-NARX models are able to account for uncertainty in a natural way. The nonpolynomial variants of the NARX model have at least one very attractive feature, in that they bypass (or rather, usually ignore) the *structure detection* problem. One can think of the problem of establishing a 'traditional' NARX (or NARMAX) model in terms of two steps. The first step is *structure detection* i.e. determining which multinomial terms should be included in the model; the second is establishing the expansion parameters for the included terms i.e. *parameter estimation*. The nonparametric NARX models simply include all expansion terms consistent with certain *hyperparameters* of the model form e.g. number of nodes per layer in an MLP neural network. One then need only concern oneself with issues of including too many terms – leading to overfitting of models – and these issues can usually be addressed in a principled manner in a machine learning context [7].

Another fairly recent addition to the family of nonparametric NARX variants is one based on *Gaussian Processes* [8]. The GP-NARX[2] model form allows a number of potential advantages over the previously mentioned NARX variants, including a Bayesian framework encompassing the generation of natural confidence intervals for model predictions. (This accommodates uncertainty in a different way to the PCE-NARX models referred to earlier.) There is no intention here, to present a survey of the literature regarding GP-NARX/GP-AR models, the curious reader is instead directed towards the comparatively recent [10].

The main weakness of the nonparametric models is arguably in their inability to provide physical insight into the systems and processes under investigation – the expansion coefficients have no direct physical meaning. Of course, this problem was also present when polynomial NARX forms were used, but was overcome to a great extent by passing from the time-domain models to a *frequency-domain* representation. One of the most interesting and useful features of the NARX model is that, through a natural connection with the Volterra series [11], it allows 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 [12]. These objects are the natural nonlinear extension of the linear concept of a Frequency Response Function, which allows direct visualisation of resonant behaviour of dynamic systems. In the case of the polynomial form of the NARX model, the method for determining the HFRFs – the *harmonic probing* algorithm – proved to be a simple extension of the long-held algorithm for differential equations [13,14]. In the case of the neural network forms (MLP and RBF) of the NARX model, the harmonic probing algorithm could also provide closed form expressions for the HFRFS at the expense of a little more complicated algebra [15]. Other neural network structures – like the Time-Delayed Neural Network [16] – allow the calculation of Volterra *kernels* directly, rather than the HFRFs.

The main objective of the current paper is to provide a calculation for the HFRFs associated with the GP-NARX model. This is taken further, and an algorithm is also provided which can estimate confidence bounds on the HFRFs, based on the predictive uncertainty inherent in the GP algorithm. A useful by-product of the analysis is a new test for nonlinearity which can detect the presence of odd and even characteristics. The theory is validated via two case studies; the first is based on numerically simulated data from an asymmetric Duffing oscillator. The second case study develops a new application of the GP-NARX model in the area of *wave loading* on offshore structures. In the context of wave loading, it has long been known that the standard equation – Morison's equation – for the prediction of fluid loading forces on slender members [17], is inadequate outside a fairly narrow regime of wave conditions. There have been many attempts to improve on Morison's equation over the years, including one by the current first author, together with collaborators, based on polynomial NARMAX/NARX models [18]. The current paper updates that methodology, using the GP-NARX model. The advantages of the new formulation – including natural confidence bounds for predictions – are demonstrated, and the paper takes the opportunity to highlight and correct a number of weaknesses of the original study in the light of modern best practice in machine learning.

The layout of the paper is as follows: Section Two will provide a short summary of the relevant Gaussian process theory and how one can use it to define a NARX model. Section Three introduces a case study and shows how the GP-NARX model is applied in the context of a nonlinear Single-Degree-of-Freedom (SDOF) system. Section Four very briefly discusses the basic principles of the Volterra series and how it leads to the definition of HFRFs. Section Five presents the derivation of the HFRFs for the GP-NARX model, which are then computed for the case study system in Section Six. The new approach to computing uncertainty bounds for HFRFs is presented in Section Seven. The application of all the aforementioned theory to the wave loading problem for the Christchurch Bay Tower is presented and discussed in Section Eight. Finally, conclusions are presented.

[2] Within the machine learning community, the term GP-AR is sometimes used. This name makes complete sense in terms of the fact that the models are auto-regressive Gaussian processes; however it misses the fact that the terms AR or ARX in the time series literature usually refer to linear models. The term GP-NARX is preferred here as it indicates that the GP models are typically nonlinear. By appropriate choice of the GP covariance function, one could fit linear GP-AR models and the algorithm would then essentially be Bayesian linear regression [9].