Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Searching for New Benchmark Models for Detecting Nonlinearity in Data

Matias Waller *

* *Åland University of Applied Sciences, AX–22111 Mariehamn, Åland (Finland) (Tel. +358400536116; e-mail: matias.waller@ha.ax).*

Abstract: Detecting nonlinearity in experimental data is a key challenge within system identification. Based on two new models, some aspects of nonlinearity relevant for system identification are discussed. The models are used to illustrate differences between three different methods for data-based testing for nonlinearity in dynamical systems. The use of Fourier-based surrogate data is found to be the most reliable approach for detecting nonlinearity in data from dynamical systems.

Keywords: Nonlinear system identification, Statistical data analysis, Frequency domain identification

1. INTRODUCTION

Although all real systems are generally recognized as nonlinear, it is well-known that linear models often render sufficient descriptions. Since the concept of nonlinearity includes everything not linear it might further be impossible to provide general measures of nonlinearity. In practice, the application at hand will determine whether or not nonlinear aspects are of interest and in that case, which aspects are relevant. Although such apparent subjectivity can make general discussions on nonlinear identification difficult, the "purposefulness" is inherent for system identification as it is recognized that "the 'true system' is an esoteric entity that cannot be attained in practical modeling. We have to be content with partial descriptions that are purposeful for our applications" Ljung [1999]. Nevertheless, the task of finding "effective data-based nonlinearity tests for dynamical systems" is still acknowledged to be one of the key challenge in the field of system identification Ljung [2010].

What, then, can be considered as essential nonlinear features? In the field of nonlinear empirical modeling based on time series, it is not uncommon to validate a nonlinear model by comparing an array of nonlinear characteristics, e.g., Lyapunov exponents, fractal dimensions and topological properties, to the corresponding experimental estimates (see e.g. Bezruchko and Smirnov [2010]). This approach is not unproblematic: It is well-known that it can be difficult to obtain reliable estimates of such characteristics, especially from noisy data (see e.g. Barnard et al. [2001], Kantz and Schreiber [2004]). In addition, such measures might be hard to define for input-output systems and also of secondary interest in system identification.

In order to clarify what might be relevant from a system identification point of view the present paper discusses three different approaches for detecting nonlinearity.

(1) The first method has not, to the knowledge of the author, been presented as a method for detecting nonlinearity as such. Instead, it has been developed for the purpose of model validation wherein the residuals of a model are examined for remaining nonlinear correlations. This approach can, however, also be used for detecting nonlinearity by first fitting an extensive linear model to data and then examining the corresponding residuals for nonlinear correlations. Analyzing residuals for correlations is indeed a reliable tool for developing linear models, but the extension to nonlinear systems is not trivial. Within the framework of nonlinear modeling, correlations between different nonlinear terms have subsequently been suggested in Billings and Woon [1986], Billings and Zhu [1994, 1995], Mao and Billings [2000], Zhu et al. [2007]. This approach is attractive in that it directly addresses the following central question: Are there nonlinearities in the data captured by the model or not?

(2) The second method is part of an extensive methodology for identification in the frequency domain and is nicely presented in Pintelon and Schoukens [2012]. Using this approach, nonlinearity can be detected in data by, in essence, separating the estimate of the best linear approximation (BLA) of the data in the frequency domain from what is called "nonlinear distortion" in addition to the stochastic noise. The main objective of the approach can be regarded as finding the best linear approximation of the data. An estimate of how much of the discrepancy between model and data is due to, mainly static, nonlinear transformations is an additional bonus and can thus be used as a basis for detecting nonlinearity.

(3) The third method utilizes the idea that linear systems are characterized by auto- and cross-correlation functions and distributions. The methodology has recently been developed for system identification by the present author Waller [2012, 2014], and relies on the use of Fourier-based surrogate data for discriminating between data that can originate from linear time-invariant systems and data that can not,

Copyright © 2014 IFAC

i.e., the method enables detecting data that cannot adequately be described by linear models.

The three approaches differ slightly, both in terms of the theoretical framework and the main objectives. Still, they all try to answer the fundamental question: Can a better model be developed by considering nonlinear features? In the present paper, the different approaches to this question are compared using simulations of two new benchmark models.

The benchmark models are described in Section 2 followed by an overview of the methods for detecting nonlinearity in Section 3. Section 4 presents an assessment of the different approaches applied to the benchmark models followed by conclusions in Section 5 .

2. BENCHMARK MODELS

The current study focuses on data-based detection of nonlinearity in dynamical systems. Systems with static, nonlinear transformations of inputs and/or outputs are often employed in system identification, (see e.g. Pintelon and Schoukens [2012]). Nonlinear features that are inherently dynamical by nature, e.g., limit cycles, are seldom studied. Therefore, this paper focuses on models that are nonlinear in terms of their dynamic features. The first model uses the function $\sin(x)$ to saturate a state variable in the dynamic equations:

$$\begin{aligned} x(k) &= \sin(x(k-1)) + u(k-1) \\ y(k) &= x(k) + e(k) \end{aligned} \tag{1}$$

The fundamental question is at what magnitude of x the approximation $\sin(x) \approx x$ is inadequate in the presence of noise $e(k)$. Methods for detecting nonlinearity can be assessed by determining, e.g., a threshold for the norm of the input $u(k)$ that yields a positive test for nonlinearity.

The second model is inherently dynamic since it includes a chaotic implementation of the well-known logistic map. In the setting, the logistic map

$$z(k) = f(z(k-1)) = \mu z(k-1)(1 - z(k-1)) \tag{2}$$

with $\mu = 4$ is used as a basis for determining a state variable observed through a linear filter. In order to create more challenging tests, the logistic map can be repeated n times. Expressed in state space this is given by,

$$\begin{pmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_{n-1}(k) \\ x_n(k) \end{pmatrix} = \begin{pmatrix} \mu x_2(k-1)(1 - x_2(k-1)) \\ \mu x_3(k-1)(1 - x_3(k-1)) \\ \vdots \\ \mu x_n(k-1)(1 - x_n(k-1)) \\ \mu x_1(k-1)(1 - x_1(k-1)) \end{pmatrix} \tag{3}$$

If a state variable is then downsampled n times, e.g.,

$$w(k) = x_1(nk) \tag{4}$$

the resulting variable w encompasses all nonlinear features of the model and is combined with a linear filter to give the output,

$$y(k) = ay(k-1) + w(k-1) + bu(k-1) + e(k) \tag{5}$$

where $u(k)$ is the input and $e(k)$ is noise.

In order to illustrate the purpose of downsampling, the corresponding recurrence plots for $w(k)$ are provided in Fig. 1. In the figure, the cases $n = 1, \ldots, 4$ are illustrated and in order to make the plots more realistic, normally

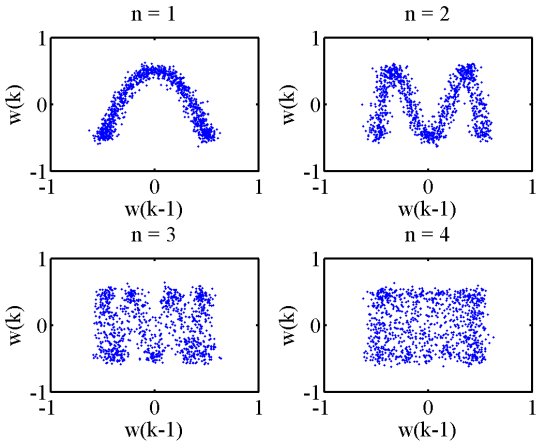

Fig. 1. Recurrence plots for $w(k)$ with noise for $n = 1, \ldots, 4$.

distributed noise with a standard deviation of 0.05 has been added to $w(k)$. In the upper left-hand corner, the well-known phase portrait for the logistic map is clearly visible, whereas any structure in phase space is difficult to distinguish in the lower right hand corner. In summary, the main idea behind the latter model is that increasingly challenging data sets can be obtained based on the logistic map by increasing n.

3. METHODOLOGIES FOR DETECTING NONLINEARITY

3.1 Nonlinear correlations

The estimate of the correlation function between two variables, α and β, for a data record of length N is denoted by $\delta^N_{\alpha\beta}(h)$. Furthermore, the difference between measured, $y(k)$, and predicted, $\hat{y}(k)$, output, defines the residual, i.e., the (one-step-ahead) prediction error

$$\varepsilon(k) = y(k) - \hat{y}(k) \tag{6}$$

In the linear case, the main focus is on the correlation between residuals and (past) inputs, $\delta^N_{\varepsilon u}(h)$, and the auto-correlation of the residuals, $\delta^N_{\varepsilon\varepsilon}(h)$. If the model manages to explain the underlying data, then $\delta^N_{\varepsilon u}(h)$ is small for all h while $\delta^N_{\varepsilon\varepsilon}(h)$ is small when $h \neq 0$. It is well-known, however, that these correlations only provide reliable estimates of linear dependencies. Therefore, a natural extension might be to study correlations between higher-order terms, e.g., $\delta^N_{\varepsilon^2(\varepsilon y)}(h)$, which is a theme developed in a number of articles Billings and Woon [1986], Billings and Zhu [1994, 1995]. It was noted, however, that nonlinear features may remain undetected by the approach of nonlinear correlations, e.g., even simple static nonlinear transformations of the inputs Mao and Billings [2000]. As a remedy, it was suggested to also consider re-ordered residuals, outputs and inputs based on the amplitude of the inputs for different delays. Still, even more powerful alternatives called combined omni-directional auto-correlation functions (ODACFs) of residuals and combined omni-directional cross-correlation functions (ODCCFs) between residuals and delayed inputs, are suggested in Zhu et al. [2007]. As the names indicate, the correlations are no longer between (simple) nonlinear terms and thus the tests

no longer indicate the types of nonlinearities present. However, as a tool for detecting nonlinearities not modeled, the approach appears to be the most powerful alternative using correlations. It is further evaluated in Section 4.

Examining nonlinear correlations can be viewed as a special case of using higher-order statistics for detecting nonlinearities (discussed in, e.g., Choudhury et al. [2008]). As illustrated in Section 4, even combined ODACFs and ODCCFs have limited power for detecting nonlinearities. The use of other higher-order statistics is likely to have similar limitations. In this paper, the use of higher-order statistics is illustrated with the use of combined omni-directional correlation functions in Section 4.

3.2 Nonlinear distortion in the frequency domain

The theoretical framework for detecting nonlinear distortion is based on Volterra-based descriptions thus excluding many nonlinear dynamical features such as hysteresis, limit cycles, etc. The method can be useful for a wide variety of systems but it can also be expected to fail for some nonlinearities. The main objective in this approach is to identify the best linear approximation of the system as well as to estimate the nonlinear distortion in the frequency domain. The approach relies on the discrete Fourier transform and thus employs assumptions regarding periodicity. For this reason, as well as for the objective of obtaining reliable estimates of the best linear approximation and nonlinear distortion in the spectrum, it is suggested that specific excitation signals be used, e.g., random phase multisines consisting of sinusoidal signals of different frequencies, amplitudes and phases Pintelon and Schoukens [2012]. The necessity of applying such special inputs as random phase multisines is a recognized limitation, and this should be noted in comparison to other approaches that do not require such special inputs. Although apparently a powerful tool for detecting static nonlinearities in the input, output or feedback terms, the suitability of this method for detecting dynamic nonlinearities is uncertain. These limitations are discussed in Section 4.

3.3 Fourier-based surrogate data

Despite some efforts to address the concept of nonlinearity in data with some *measure* for nonlinearity (such as the attempts made by, e.g., Small et al. [2001], Choudhury et al. [2008]) it seems that there is no general consensus regarding such a measure as discussed in Kantz and Schreiber [2004], Kugiumtzis [2008], Bezruchko and Smirnov [2010] among others. As an alternative to a measure of nonlinearity, it has been suggested within nonlinear time series analysis that the following question be considered instead: Can the data be adequately modeled with some linear time-invariant model with, possibly, non-Gaussian distribution of the signals? Within time series analysis, a well-established framework for answering this question in a statistical sense (i.e., with the possibility for a false answer corresponding to a chosen level of significance) relies on the use of Fourier-based surrogate data (see e.g. Theiler et al. [1992], Schreiber and Schmitz [2000], Kantz and Schreiber [2004]). The use of Fourier-based surrogate data within system identification and for model validation by residual analysis, including the MIMO case, is described in detail in Waller [2014]. The methodology presented in the article detects nonlinear predictability in multivariate data, nonlinearities that cannot be adequately captured by linear, stationary, models. For clarity, a brief summary of the methodology is provided below.

In order to assess nonlinear predictability of data, the original data is, by an appropriate measure, compared to a collection of Fourier-based surrogates. As a measure for nonlinear predictability, a version of the intuitively appealing nearest neighbor approach to prediction Lorenz [1969] is used. With this choice, an explicit fitting of a model, i.e., a parsimonious description of the data, can be avoided.

A general predictive description is given by

$$\hat{y}(k) = g(\varphi(k-1)) \tag{7}$$

where $g(\cdot)$ is a mapping (the predictive model) from the regressors, $\varphi(k-1)$, to the predicted output, $\hat{y}(k)$, at sampling k. Although choosing appropriate regressors is a challenge in system identification, a simple test naturally requires a simple approach. Consequently, the components of the regressors are chosen from the set of data, Z^N, i.e., the N observations of inputs and outputs. The regressor can thus be expressed

$$\begin{aligned}\varphi(k-1) = \ & (y(k-1), \cdots, y(k-p), \\ & u(k-L-1) \cdots u(k-L-m))\end{aligned} \tag{8}$$

where p is the order with respect to the output(s), m is the order with respect to the input(s) and L is the delay(s) from input(s) to output(s). Given p, m and L, regressors $\varphi(k-1)$ corresponding to the observations $y(k)$ are easily determined from Z^N for all k.

The nearest neighbor approach to prediction is based on finding the regressor $\varphi(l-1)$ which, in some sense, is closest to $\varphi(k-1)$, i.e.,

$$\|\varphi(k-1) - \varphi(l-1)\| < \epsilon_{\min} \tag{9}$$

where $\epsilon_{\min}$ is the smallest value for $\|\varphi(k-1) - \varphi(l-1)\|$ for all $l \in N$, $l \neq k$. The prediction for $y(k)$ is then given by

$$\hat{y}(k) = y(l) \tag{10}$$

In order to make the prediction less sensitive for specific noise characteristics, a collection of nearest neighbors can be used, resulting in the prediction Kantz and Schreiber [2004]

$$\hat{y}(k) = \frac{1}{|U_\epsilon(\varphi(k-1))|} \sum_{\varphi(l-1) \in U_\epsilon(\varphi(k-1))} y(l) \tag{11}$$

where $|U_\epsilon(\varphi(k-1))|$ denotes the number of regressors in the neighborhood $U_\epsilon(\varphi(k-1))$, i.e., the number of regressors satisfying the criterion $\|\varphi(k-1) - \varphi(l-1)\| < \epsilon$ for all $l \in N$, $l \neq k$. By varying ϵ, a suitable number of neighbors can be found. The extension of the simple predictive scheme to MIMO systems is trivial.

Surrogate data based on the Fourier transform yield data with a similar power spectrum and cross-spectrum as the original data. As such, the surrogates and the original data will result in similar models within the general family of linear models,

$$A(q)y(k) = \frac{B(q)}{F(q)}u(k-L) + \frac{C(q)}{D(q)}e(k). \tag{12}$$

The surrogate data will not, however, contain nonlinear relations. Thus, surrogate data can be used as a basis to test whether or not the data can be adequately described within the family of linear models of (12). Although the output y for a linear model will follow a Gaussian distribution if u and e are Gaussian, an observed non-Gaussian distribution does not necessarily imply that the data justifies a nonlinear dynamic model. An appealing extension to the Fourier surrogates is therefore to require that, in addition to the spectra, the distributions for u and y are similar for surrogates and original data. For this purpose, there are two widely spread algorithms resulting in what is known as the Amplitude Adjusted Fourier Transform (AAFT) surrogates presented in Theiler et al. [1992] and the Iterative AAFT (IAAFT) surrogates presented in Schreiber and Schmitz [1996].

Using surrogates for a statistical test for nonlinear predictability requires choosing the level of significance. The result of the test is binary, i.e., on the chosen level of significance the method either indicates a presence of nonlinearity or it does not. Therefore, the method will not give any guidance regarding the nature of the nonlinearity and is only a method for data-based testing for nonlinearity.

For the SISO case, the test for nonlinearity using surrogate data can be summarized by the following algorithm:

(1) Choose a level of significance, $(1-\alpha)$, for the discriminative test based on the probability of a false rejection, α.
(2) Use the set of outputs and inputs
$$\begin{pmatrix} y(1) & u(1) \\ \vdots & \vdots \\ y(N) & u(N) \end{pmatrix} \tag{13}$$
and generate a collection of surrogate data, i.e., $2/\alpha - 1$ sets for a two sided test: E.g., at least 39 surrogates for a minimal significance requirement of 95% are needed.
(3) Calculate predictions of the output using (11) and regressors
$$\begin{aligned}\varphi(k-1) = \ & (y(k-1), \cdots, y(k-m), \\ & u(k-L-1) \cdots u(k-L-m))\end{aligned} \tag{14}$$
where m can be varied depending on the dimension considered for the model to be identified. In addition, rough estimates of input-output delays are needed.
(4) Compare the variance of the prediction errors for the original set as well as for all surrogate sets. If the variance of the prediction error for the output is smaller (or larger) for the original set than for all surrogate sets, there is, on the chosen significance level, statistical evidence of nonlinear features in the original set.

Clearly, the method can be readily applied to model validation by substituting outputs y in Eq. (14) with residuals ε defined in Eq. (6). For the simulations presented in Section 4, the IAAFT surrogates are used.

Like the frequency domain approach for estimating nonlinear distortion, the approach of Fourier-based surrogates relies on the discrete Fourier transform and assumes periodicity of the registered signals. If this is not the case, e.g., periodic excitation signals are not used, it has been noted that an end-to-end mismatch, i.e., a "large" jump between the first and the last value of the data set as well as a jump between the initial and final direction in the data set, can lead to artifacts in the surrogate data. However, this can easily be avoided by choosing a suitable subset of the data Schreiber and Schmitz [2000].

4. SIMULATIONS

In order to evaluate the three methods for detecting nonlinearity, two measures are used.

(1) The simple model, Eq. (1), is used with a data length of $N = 5000$ and a standard deviation for the Gaussian noise e as 0.1. The norm of u is the measure for when nonlinearity is reliably detected, i.e., in ten independent simulations.
(2) The more challenging example, Eq. (5), is used with a standard deviation of 0.05 for the Gaussian noise e and the norm of u as 0.3. The logistic map is downsampled between two and four times, i.e., $n = 2, 3, 4$ and the measure for when nonlinearity is reliably detected is given by the required length of the data record $1000 \leq N \leq 100000$.

These measures are chosen based on the objectives of system identification: What can be modeled based on the *data*? Specifically for nonlinear identification, a key issue is the assessment of whether the data warrants nonlinear modeling, i.e., when is a linear model an insufficient description. In general, larger excitations are believed to reveal more nonlinear behavior, which is clearly the case for the model in Eq. (1): Data from the model in Eq. (1) should only motivate nonlinear modeling for larger variations in x, thus making the norm of u a simple and from a control perspective intuitive measure relevant for detecting nonlinearity.

In addition, it seems reasonable to assume that more complicated systems, e.g., higher-order chaotic maps, will require more data to be adequately modeled. Therefore, it is desirable that a method for detecting nonlinearity in data reflects this assumption (hence the choice for the second measure).

It can be noted that the three approaches are different with respect to the choice of inputs: The frequency domain approach requires random phase multisines as excitations, while the other two approaches can be applied with a variety of inputs. For the method for estimating nonlinear distortion in the frequency domain, the used norm of u is the root mean square (rms) value for the random phase multisines, while the the other approaches use the comparable standard deviation for normally distributed u.

For the model of Eq. (1), nonlinear distortion in the frequency domain is detected for $u_{\text{rms}} \approx 0.3$. Applying the other approaches, i.e., combined omni-directional correlation functions and IAAFT surrogates, nonlinearity is detected for $\text{std}(u) \approx 0.3$. With respect to detecting nonlinearity only, the methods thus seem to yield equivalent results for the model of Eq. (1).

The estimated frequency response function (G_{BLA}) and the nonlinear distortion (G_{NL}) using an rms of 0.5 for u are

illustrated in Fig. 2. For the relatively small excitations,

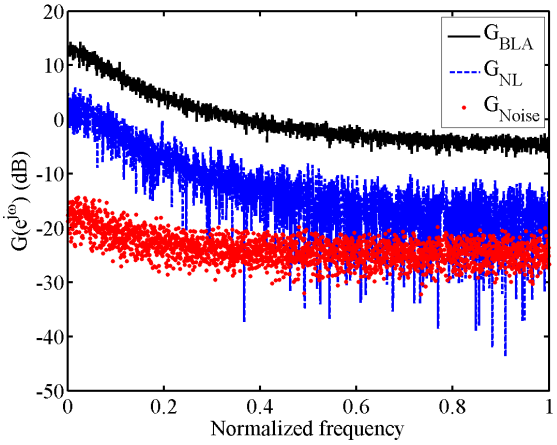

Fig. 2. Estimated frequency function (solid black) for the model of Eq. (1) with an rms of 0.5 for u, estimated nonlinear distortion (dashed blue) and estimated stochastic noise (dotted red).

the nonlinear distortion is, as can be expected, most apparent for lower frequencies. Indeed, for the model

$$\begin{aligned} x(k) &= -\sin(x(k-1)) + u(k-1) \\ y(k) &= x(k) + e(k) \end{aligned} \tag{15}$$

the nonlinear distortion is most apparent for higher frequencies when relatively small excitations are used. Interestingly, the nonlinear distortion will be more apparent over the whole spectrum with larger variations in u. A possibility to obtain further indications towards the nature of the nonlinearities might therefore be to perform several experiments with different amplitudes for the random phase multisines. As the figure also reveals, valuable information of the underlying system is obtained thus illustrating some other merits of the frequency domain approach. Even though the frequency domain approach apparently can be useful in many ways, it is also unreliable when faced with inherently dynamic nonlinearities as the next example will illustrate.

For the approach of nonlinear correlations, nonlinearity is most clearly detected by the combined ODCCF and illustrated in Fig. 3. In the figure, 95% confidence intervals for "small" correlations, i.e., approximately for $|\delta| < 1.94/\sqrt{N}$ Billings and Woon [1986], are included as dashed lines. Clearly, the remaining correlation is most apparent for $h = 2$.

For the more challenging example, Eq. (5), the method for estimating nonlinear distortion fails, even for the standard logistic map, i.e., $n = 1$. This case is also illustrated in in Fig. 4 and the nonlinear distortion cannot be distinguished from the stochastic noise. The empirical analysis thus confirms the theoretical foundation of the approach, i.e., it is unreliable for detecting dynamic nonlinearities. The results for all three approaches based on the simulations are summarized in Table 1.

As the table indicates, the frequency domain approach of estimating the nonlinear distortion clearly fails for the chaotic map. The approach of combined ODCF:s seems reliable for smaller n (i.e., low-order models) but struggles

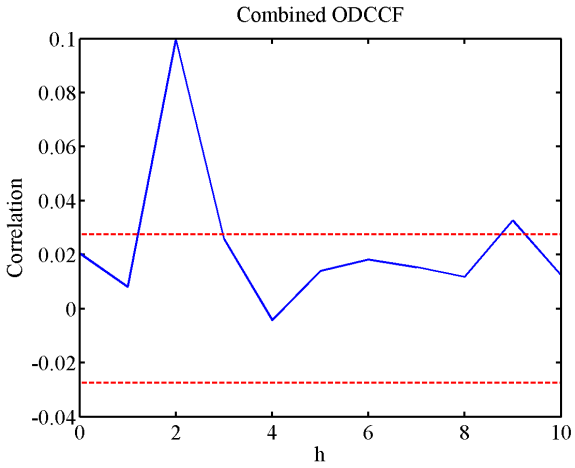

Fig. 3. Combined ODCCF (solid blue) for the model of Eq. (1) using a standard deviation of 0.3 for u. Approximate 95% confidence intervals for "small" correlations are included (dashed red). The remaining correlation is clearly most apparent for $h = 2$.

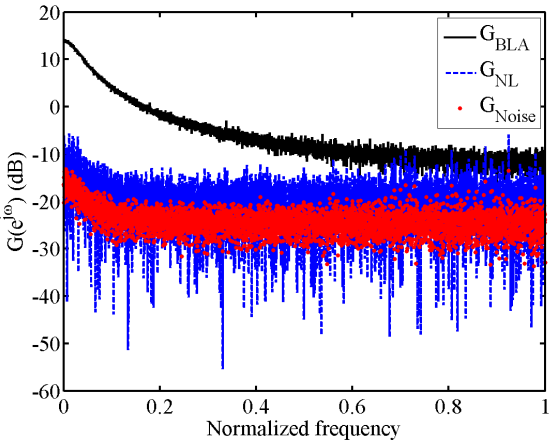

Fig. 4. Estimated frequency function (solid black) for the model of Eq. (5) with the logistic map repeated only once ($n = 1$) and an rms of 0.5 for u, estimated nonlinear distortion (dashed blue) and estimated stochastic noise (dotted red).

Table 1. Required length of data record ($1000 \leq N \leq 100000$) for detecting nonlinearity as a function of iterations (n) of the logistic map of Eq. (5).

Method	n		
	2	3	4
Nonlinear distortion	Fails	Fails	Fails
Combined ODCFs	$N = 1000$	$N = 50000$	Fails
IAAFT surrogates	$N = 1000$	$N = 5000$	$N = 100000$

for models of higher orders and fails when $n = 4$. For $n = 3$, the nonlinearity is only, and barely, detected by the combined ODACF for $h = 1$ which is illustrated in Fig. 5. The use of IAAFT surrogates, however, seems very reliable even for the higher-order models. It can be noted that in many cases it might not be realistic to use such long data segments, thus providing further support for the use of surrogate data. In addition, the results indicate that even

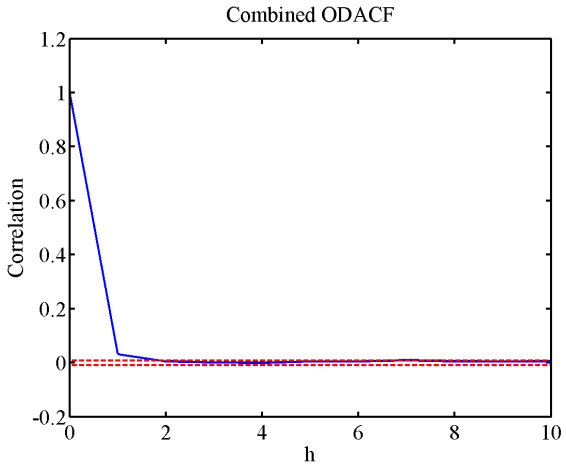

Fig. 5. Combined ODACF (solid blue) for the model of Eq. (5) with $n = 3$ and $N = 50000$. Approximate 95% confidence intervals for "small" correlations are included (dashed red). The remaining nonlinear correlation is barely detected for $h = 1$.

modestly complex systems, e.g., the logistic map repeated four times ($n = 4$), is not easily modeled based on data in the presence of noise.

In summary, the approach that uses Fourier-based surrogates clearly seems to be the most reliable alternative for a variety of systems if used only as a means for detecting nonlinearity in data from dynamical systems.

5. CONCLUSIONS

This paper introduces two nonlinear models designed for evaluating data-based testing for nonlinearity in dynamical systems. The models were used in an empirical analysis of three different approaches for testing for nonlinearity:

(1) The use of combined omni-directional correlation functions.
(2) To estimate nonlinear distortion in the frequency domain.
(3) The use of Fourier-based surrogate data.

The analysis shows that the different methods have different merits and seem comparable for easier models if used only for testing for nonlinearity. For more inherently dynamical nonlinearities, the approach of estimating nonlinear distortion in the frequency domain can clearly fail. For increasing complexity, e.g., higher-order chaotic maps, the approach using correlation functions is also unreliable and the approach of surrogate data is clearly superior to the other approaches. Similar results have also been noted for an array of examples found in the literature Waller [2014]. It should, however, be noted that these results are based solely on empirical analysis. Further study is required in order to develop a theoretical foundation that might provide greater insight, e.g., which classes or types of nonlinear models will be detected.

Still, based on empirical analysis the methodology that uses Fourier-based surrogates seems to be a very powerful data-based nonlinearity test for dynamical systems and can be seen to often outperform alternative approaches. As such, the method may be of great general value in the field of system identification.

REFERENCES

J. P. Barnard, C. Aldrich, and M. Gerber. Identification of dynamic process systems with surrogate data methods. *AIChE Journal*, 47:2064–2075, 2001.

B. P. Bezruchko and D. M. Smirnov. *Extracting Knowledge from Time Series: An Introduction to Nonlinear Empirical Modeling*. Springer-Verlag, Heidelberg, 2010.

S. A. Billings and W. S. F. Woon. Correlation based model validity tests for nonlinear models. *International Journal of Control*, 60:235–244, 1986.

S. A. Billings and Q. M. Zhu. Nonlinear model validation using correlation tests. *International Journal of Control*, 60:1107–1120, 1994.

S. A. Billings and Q. M. Zhu. Model validation test for multivariable nonlinear models including neural networks. *International Journal of Control*, 62:749–766, 1995.

M. A. A. S. Choudhury, S. L. Shah, and N. F. Thornhill. *Diagnosis of Process Nonlinearities and Valve Stiction: Data Driven Approaches*. Springer, Berlin, 2008.

H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, 2 edition, 2004.

D. Kugiumtzis. Evaluation of surrogate and bootstrap tests for nonlinearity in time series. *Studies in Nonlinear Dynamics & Econometrics*, 12, 2008.

L. Ljung. *System Identification—Theory for the User*. Prentice Hall, New Jersey, 2 edition, 1999.

Lennart Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34:1–12, 2010.

E. Lorenz. Atmospheric predictability as revealed by naturally occuring analogues. *Journal of Atmospheric Sciences*, 26:636–646, 1969.

K. Z. Mao and S. A. Billings. Multi-directional model validity tests for non-linear system identification. *International Journal of Control*, 73:132–143, 2000.

R. Pintelon and J. Schoukens. *System Identification—A Frequency Domain Approach*. John Wiley & Sons, New Jersey, 2 edition, 2012.

T. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Physical Review Letters*, 77:635–638, 1996.

T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142:346–382, 2000.

M. Small, K. Judd, and A. Mees. Testing time series for nonlinearity. *Statistics and computing*, 11:257–268, 2001.

J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. Testing for nonlinearity in time series: The method of surrogate data. *Physica D*, 58:77–94, 1992.

M. Waller. Using surrogate data for nonlinear identification: A case study. In *16th IFAC Symposium on System Identification*, pages 1581–1586, Brussels, 2012.

M. Waller. Data-based testing for nonlinearity in dynamical systems: The use of surrogate data. *Working title*, 2014.

Q. M. Zhu, L. F. Zhang, and A. Longden. Development of omni-directional correlation functions for nonlinear model validation. *Automatica*, 43:1519–1531, 2007.