Test for Nonlinear Input Output Relations in SISO Systems by Preliminary Data Analysis

Knudsen, Torben

Publication date:
2000

Document Version
Også kaldet Forlagets PDF

Test for Nonlinear Input Output Relations in SISO Systems by Preliminary Data Analysis

Torben Knudsen

Abstract

This paper discusses and develops preliminary statistical tests for detecting nonlinearities in the deterministic part of SISO systems with noise. The most referenced method is unreliable for common noise processes as e.g. colored. Therefore two new methods based on superposition and sinus input are developed. They are much more robust, especially the sinus method which is reliable also for colored, heavy tailed or skew distributed noise.

Keywords: Nonlinearity test; preliminary test; identification; Subba Rao test; Robust methods;

1 Introduction

For some years there has been a substantial and increasing research on estimation of parameters in nonlinear dynamical models from experimental input output data. Deriving nonlinear models are much more complicated than deriving linear models. Consequently, if there is doubt about the existence or significance of nonlinear effects a good preliminary test for linearity would be very useful [5, p. 652]. Test of this kind exists [1]. They are based on methods developed for time series and uses third order correlations. These tests suffers from lack of robustness and do not have well defined statistical properties.

Therefore the objectives in this paper are: To develop tests based on basic properties for linear systems namely superposition and sinus input gives sinus output. And to compare the statistical properties for these test with the existing ones. The test should be kept robust and simple and applicable before a candidate model is found.

Below the problem is first discussed in more details. Then existing methods are explained. This is followed by the development of new methods. The statistical properties are then analyzed and finally a conclusion is drawn.

2 Choice of problem

It is important to consider the different choices that together specifies the problem. To that end some notation are first introduced. Assume N samples of input u and output y are measured from a SISO system and collected in vectors U and Y (1)–(2).

$$U \triangleq \begin{pmatrix} u(1) & \dots & u(N) \end{pmatrix}^T \tag{1}$$

$$Y \triangleq \begin{pmatrix} y(1) & \dots & y(N) \end{pmatrix}^T \tag{2}$$

To get a solvable problems it seems to be necessary to assume the noise $v(t)$ stochastic stationary (3e) and additive (3a). Of cause it could also be multiplicative and then a logarithmic transformation gives the additive structure. It is also necessary to assume the initial conditions known (in some sense) otherwise they would enter in (3a), assume therefore that the system is initially at a stationary state given by $u(1)$ which is obtained by (3c).

$$y(t) = f_t(u_1^t) + v(t)\ ,\ t \in \mathbb{N} \tag{3a}$$
$$u_1^t \triangleq \begin{pmatrix} u(1) & \dots & u(t) \end{pmatrix}^T \tag{3b}$$
$$u(t) = u(1) \quad \forall\, t < 1 \tag{3c}$$
$$u, v \text{ are independent} \tag{3d}$$
$$v \text{ stationary} \tag{3e}$$
$$\mathrm{E}(v(t)) = 0\ ,\ \mathrm{Var}(v(t)) = \sigma_v^2 \tag{3f}$$

The choices specifying the problem are the following:

- *Preliminary* analysis methods i.e. method based on data before a candidate model is found or *validation* methods based on a candidate model.
- Assumptions about u.
- Assumptions about v as e.g. white noise and normal distribution.
- Proper *statistical* test or tests based on *indicators* with no precise statistical properties.

In this paper the choices are a proper preliminary statistical test method whit as few assumptions on u and v as possible.

3 Existing methods

For time series this problem is well covered and there has been developed a number of methods starting with the Subba Rao (SR) method [10, sec. 5.3]. This method are based on the fact that third order moments of linear combinations of white noise e with symmetrical distribution f_e are zero.

$$v(t) = \sum_{i=0}^{\infty} a_i e(t-i)\ ,\ f_e(x) = f_e(-x) \Rightarrow \tag{4}$$
$$\mathrm{E}(v(t_1)v(t_2)v(t_3)) = 0 \quad \forall\, t_1, t_2, t_3 \in \mathbb{Z} \tag{5}$$

There also exist a well known statistical method of validation type where specific nonlinear terms can be tested [7, 8]. This test are based on correlating residuals ϵ with effects, if e.g. $\mathrm{E}(\epsilon(t)u(t)^2) \neq 0$ then $u(t)^2$ should be include in the model. This residual validation test can be generalized to test for missing effects of any type [2]. A very different validation test of indicator type is found in [6].

The literature covering preliminary methods for systems with measurable input is more limited. The reference text books on system identification [7, 8] only gives a brief discussion on the problem and some references. For purely deterministic systems 10 indicator methods are presented in [4]. The statistical test methods for noisy systems are based on the ideas from the Subba Rao test. In [1] the following is shown:

Assume

$$y(t) = f_t(u_1^t) + v(t) \tag{6a}$$
$$u(t) = u_c(t) + b\ ,\ b \neq 0 \tag{6b}$$
$$u_c, v \text{ independent} \tag{6c}$$
$$\mathrm{E}(v) = \mathrm{E}(u_c) = 0 \tag{6d}$$
$$\mathrm{E}(v^{2n+1}) = \mathrm{E}(u_c^{2n+1}) = 0 \quad \forall\, n \in \mathbb{N} \tag{6e}$$

$$v(t) \text{ is a linear stochastic process as defined by (4)} \tag{6f}$$

then

$$\mathrm{E}(y_c(t+\tau)y_c(t)^2) = 0 \quad \forall \tau \in \mathbb{Z}\,,\ y_c(t) = y(t) - \mathrm{E}(y(t)) \tag{7}$$

if and only if

$$f_t \text{ is linear in } u_1^t \tag{8}$$

The test suggested in [1] is then to reject linearity if the estimated correlation $|\hat{\phi}_{y_c y_c^2}(\tau)|$ corresponding to (7) is to large (9). The correlation estimate is defined in the usual way (10).

$$\text{reject linearity if } |\hat{\phi}_{y_c y_c^2}(\tau)| > \frac{1.96}{\sqrt{N}} \tag{9}$$

$$\hat{\phi}_{xy}(\tau) \triangleq \frac{\widehat{\mathrm{Cov}}_{xy}(\tau)}{\sqrt{\widehat{\mathrm{Cov}}_{xx}(0)\widehat{\mathrm{Cov}}_{yy}(0)}}\,, \tag{10}$$

$$\widehat{\mathrm{Cov}}_{xy}(\tau) \triangleq \begin{cases} \frac{1}{N}\sum_{t=1}^{N-\tau}(x(t)-\bar{x})(y(t+\tau)-\bar{y}) & ,\tau \geq 0 \\ \widehat{\mathrm{Cov}}_{yx}(-\tau) & ,\tau < 0 \end{cases}\,,\ \bar{x} \triangleq \frac{1}{N}\sum_{t=1}^{N} x(t) \tag{11}$$

Notice that the test (9) is only approximate because it requires the additional assumption that y_c and y_c^2 must be white noise [7, sec. 16.6] which normally is not the case. However, especially with a small number of samples it is difficult to obtain a more exact test.

It is not clear in [1] which τ to use and what to do if only one or few correlations falls outside the limits. In this paper $|\tau| \leq 12$ is used because this is the number used in the examples in [1]. To obtain a p-value it is chosen to use the χ^2 test (12)–(13) where $F^{-1}_{\chi^2(2\tau_m+1)}$ is the inverse distribution function for a χ^2 distribution with $2\tau_m + 1$ degrees of freedom. This test is consistent with (9) because it requires the same assumptions.

$$T = N\sum_{|\tau|\leq\tau_m} \hat{\phi}_{y_c y_c^2}(\tau)^2\,,\ \tau_m = 12 \tag{12}$$

$$p = 1 - F^{-1}_{\chi^2(2\tau_m+1)}(T) \tag{13}$$

The choice of input is only limited by (6e) according to [1]. This includes e.g. a symmetrical distributed random process which also is sufficient for (7) to hold. However, with a limited number off samples the test (12)–(13) is very sensitive to departures from the sample version of (6e) i.e. $\frac{1}{N}\sum_{t=1}^{N} u_c^{2n+1} = 0$. Consequently a sinus input with a integer number of periods is used in this paper which also is used for the examples in [1].

The main problem with the method above is that only the limiting conditions (7) is proven. Moreover there is no results concerning the statistical properties of the suggested test.

4 New methods

Clearly there are several reasons to develop new methods. As seen in (9) the input u does not explicitly enter the test at all, this is because it is based on a time series approach. The methods developed here should make more use of the input and the basic properties for linear systems because this will probably be more power full. The first method suggested is based on superposition and the second one is based on that sinus input gives sinus output for linear systems.

Non-Linearity test based on superposition

In short the method is to compare the difference between the response to a sum of two inputs and the sum of responses to the inputs with the variance on the noise.

The simplest solution is first to choose an appropriate input sequence U. Run this two times through the system and run $2U$ through the system. This produces the following signals where V_i is the noise sequence corresponding to U_i.

$$U_2 \triangleq U_1 \triangleq U \ , \ U_3 \triangleq 2U \tag{14}$$

$$Y_i = F(U_i) + V_i \ , \ i = 1,2,3 \tag{15}$$

Introduce also the differences (16) and (17) then the following basic theorem is easily obtained.

$$D_1 \triangleq Y_2 - Y_1 = V_2 - V_1 \tag{16}$$

$$D_2 \triangleq Y_3 - Y_2 - Y_1 = F(2U) - 2F(U) + V_3 - V_2 - V_1 \tag{17}$$

Theorem 1 (Superposition method with 3 sequences (SUP3)). *Assume the system is given by* (3) *and the sequences* V_1, V_2, V_3 *are independent, then:*

$$D_1, D_2 \textit{ are uncorrelated and independent if } V_i \textit{ are normal} \tag{18}$$

$$\frac{\mathrm{E}(D_1^T D_1)}{2N} = \sigma_v^2 \tag{19}$$

$$\frac{\mathrm{E}(D_2^T D_2)}{3N} \geq \sigma_v^2 \tag{20}$$

$$F(2U) = 2F(U) \Leftrightarrow \frac{\mathrm{E}(D_2^T D_2)}{3N} = \sigma_v^2 \tag{21}$$

Proof. (18) is proven by using independence for V_1, V_2, V_3, U and $\mathrm{E}(V_2 V_2^T) = \mathrm{E}(V_1 V_1^T)$ which gives

$$\mathrm{E}(D_1 D_2^T) = \mathrm{E}((V_2 - V_1)(F(2U) - 2F(U) + V_3 - V_2 - V_1)^T) = 0 \tag{22}$$

(20)–(21) are proved by

$$\mathrm{E}(D_2^T D_2) = \mathrm{E}((F(2U) - 2F(U))^T (F(2U) - 2F(U))) + N3\sigma_v^2 \tag{23}$$

Finally (19) follows directly from (16). □

The test for linearity is now performed by testing superposition (24) which reduces to comparing variances for the stochastic processes $d_1(t)$ and $\sqrt{\frac{2}{3}} d_2(t)$ (25) where the means are assumed zero and $d_i(t)$ are the elements in D_i. This variance test is discussed after the next section.

$$H_0 : F(2U) = 2F(U) \qquad H_0 : F(2U) \neq 2F(U) \tag{24}$$

$$H_0 : \mathrm{Var}\left(\sqrt{\frac{2}{3}} d_2\right) = \mathrm{Var}(d_1) \qquad H_1 : \mathrm{Var}\left(\sqrt{\frac{2}{3}} d_2\right) > \mathrm{Var}(d_1) \tag{25}$$

There is however at least one type of nonlinearities (57) and input which the above test is insensitive to. In this case the similar method below with 5 sequences can be used.

$$U_2 \triangleq U_1 \triangleq U \ , \ U_3 \ , \ U_4 \ , \ U_5 \triangleq U_3 + U_4 \tag{26}$$

$$Y_i = F(U_i) + V_i \ , \ i = 1,2,3,4,5 \tag{27}$$

$$D_1 \triangleq Y_2 - Y_1 = V_2 - V_1 \tag{28}$$

$$D_2 \triangleq Y_5 - Y_4 - Y_3 = F(U_5) - F(U_4) - F(U_3) + V_5 - V_4 - V_3 \tag{29}$$

Theorem 2 (Superposition method with 5 sequences (SUP5)). *Assume the system is given by* (3) *and the sequences* V_1, V_2, V_3, V_4, V_5 *are independent, then theorem 1 also applies to* D_1, D_2 *defined for five sequences* (28)–(29) *except that* D_1, D_2 *is always independent and the linearity condition in* (21) *is now* $F(U_5) = F(U_4) + F(U_3)$.

Proof. Similar to the proof for theorem 1 □

Non-Linearity test based on sinus input

The sinus response method is very similar to the superposition method above. Here the difference between the output and a estimated sinus is compared to the noise variance.

Assume that the input is now a sinus (30) with appropriate choices of a and ω. Notice that the system is also initialized by the sinus input which strictly speaking violates (3c), however it still serves the same purpose. Assume that $\omega = \frac{k2\pi}{N}\ ,\ k \in \mathbb{N}$ then N is a multiple of the period which makes it possible in practice to use one long sequence with $2N$ samples for U_1 and U_2. Below some of the previous notation are reused.

$$u(t) = a\sin(\omega t)\ ,\ t \in \mathbb{Z} \tag{30}$$

$$U_2 \triangleq U_1 \triangleq U \tag{31}$$

$$Y_i = F(U_i) + V_i\ ,\ i = 1, 2 \tag{32}$$

Introduce also the differences (33) and (37) below where $\widehat{Y}_i$ is a simple LS estimate. This then gives the next theorem.

$$D_1 \triangleq Y_2 - Y_1 = V_2 - V_1 \tag{33}$$

$$X \triangleq \begin{pmatrix} \sin(\omega) & \cos(\omega) \\ \sin(2\omega) & \cos(2\omega) \\ \vdots & \vdots \\ \sin(N\omega) & \cos(N\omega) \end{pmatrix} \tag{34}$$

$$\hat{\theta}_i \triangleq (X^T X)^{-1} X^T Y_i\ ,\ i = 1, 2 \tag{35}$$

$$\widehat{Y}_i \triangleq X\hat{\theta}_i\ ,\ i = 1, 2 \tag{36}$$

$$D_2 \triangleq (Y_2 - \widehat{Y}_2) + (Y_1 - \widehat{Y}_1) \tag{37}$$

Theorem 3 (Sinus input method (SIN)). *Assume the system is given by* (3) *except for* (3c) *whish is replaced by* (30) *and the sequences* V_1, V_2 *are independent, then:*

$$\frac{\mathrm{E}(D_1^T D_1)}{2N} = \sigma_v^2 \tag{38}$$

assymtotically for $N \to \infty$

$$\frac{\mathrm{E}(D_2^T D_2)}{2N} \geq \sigma_v^2 \tag{39}$$

and for a linear system

$$\frac{\mathrm{E}(D_2^T D_2)}{2N} = \sigma_v^2 \tag{40}$$

$$D_1, D_2 \textit{ are uncorrelated} \tag{41}$$

Remark 3.1. In the statistical tests it is convenient to have D_1 and D_2 uncorrelated which is the reason for the specific choice (37) of D_2.

Remark 3.2. It is also possible to estimate a common θ by the LS problem (42). This does not change the convergence properties in the theorem but it improves the estimate a little.

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X \\ X \end{bmatrix} \theta \tag{42}$$

Proof. If the system is linear then $f_t(u_1^t) = \alpha \sin(\omega t) + \beta \cos(\omega t)$ for some α, β and, because regressors are uncorrelated with noise, the LS estimate is consistent i.e. the following mean square (ms) convergence applies:

$$\hat{\theta}_i \to \begin{pmatrix} \alpha & \beta \end{pmatrix}^T \ , \ \hat{y}_i(t) \to f_t(u_1^t) \\ , \ y(t) - \hat{y}_i(t) \to v_i(t) \ , \ d_2(t) \to v_2(t) + v_1(t) \text{ (ms) for } N \to \infty \tag{43}$$

which proves (40) and (41). (38) and (39) are now straight forward. □

The test (46) following from theorem 3 is similar to (25) except no scaling is required.

Statistical test comparing two variances

The above developed linearity tests are both turned into test of equal variances for two stochastic processes which are zero mean under H_0. Adopting the notation from the sinus input method the problem is the following: Given data (44) and basic assumptions (45) find tests for (46).

$$d_i(t) \ , \ t = 1, \dots N \ , \ i = 1, 2 \tag{44}$$

$$d_i \text{ stationary } \ , \ \mathrm{E}(d_i(t)) = 0 \ , \ d_1(t_1), d_2(t_2) \text{ uncorrelated } \quad \forall \, t_1, t_2 \tag{45}$$

$$H_0 : \mathrm{Var}\,(d_2) = \mathrm{Var}(d_1) \qquad H_1 : \mathrm{Var}\,(d_2) > \mathrm{Var}(d_1) \tag{46}$$

In all cases d_i are linear combinations of independent system noise processes v_i. Thus under H_0 d_i's are all zero mean and with the same autocorrelation function as $v(t)$. Therefore the assumption about second order properties for the noise v_i also applies to d_i under H_0 which is very convenient.

To develop a proper statistical test a test statistic and its distribution under H_0 are required. The basic assumptions (45) are really minimal for this purpose. Below a number of test are presented with decreasing restrictive assumptions i.e. increasing robustness.

Normal white (NW): The noise samples $v_i(t)$ are normal and mutually independent with variance σ_v^2 ($\mathrm{NID}(0, \sigma_v^2)$) then the standard F test (48) and p-value (49) can be applied. When the sinus method with two parameters estimated is used with normal white noise it is more correct to use $F(N-2, N-2)$ in (48)–(49).

$$H_0 \Rightarrow d_i(t) \in \mathrm{NID}(0, \sigma_d^2) \wedge d_1, d_2 \text{ independent} \Rightarrow \tag{47}$$

$$T = \frac{\sum_{t=1}^N d_2(t)^2}{\sum_{t=1}^N d_1(t)^2} \in F(N, N) \Rightarrow \tag{48}$$

$$p = 1 - F^{-1}_{F(N,N)}(T) \tag{49}$$

Normal colored (NC): The noise v_i is a linear normal process (4) with a unknown correlation structure. In this case the results above does not hold. One way to obtain a more white sequences is the whitening method where d_i is filtered with the inverse of a estimated transfer function. This has several drawbacks, a transfer function with unknown structure must be estimated, if the structure does not match the real one the method does not work to well and finally under H_1 the difference between d_2 and d_1 will for some nonlinearities bee dominated by a low frequent component which the filter might remove and thereby spoiling the power of the test. The alternative approach used here is to find a correlation length n_r where correlation between $d_i(t)$ and $d_i(t+\tau)$ is small for $\tau \geq n_r$ and then resample every n_r sample and finally use the NW method on the decimated sequences. The correlation length n_r is estimated by (50) where the lag one autocorrelation $\hat{\rho}_1$ is the standard correlation coefficient estimate for d_1. This method is based on an AR(1) approximation. However it has proven robust in simulation experiments at the cost of reduced number of samples and thereby power.

$$|\hat{\rho}_1|^{n_r} = \rho_l \Leftrightarrow n_r = \frac{\log(\rho_l)}{\log(|\hat{\rho}_1|)} \ , \ \rho_l = 0.31 \tag{50}$$

The reason for the seemingly large choice $\rho_l = 0.31$ is that it corresponds to 5% of the standard deviation on $d_i(t+n_r)$ explained by $d_i(t)$ which turn out to be a suitable value.

White (W): The noise samples $v_i(t)$ is mutually independent with variance σ_v^2 and equal but unknown distribution ($\mathrm{ID}(0, \sigma_v^2)$). Then the non-parametric squared ranks test (51)–(54) is used with the normal approximation for the test statistic. d_1 and d_2 are ranked together and $R(d_2(t))$ are the ranks for d_2, see [3, sec. 5.3] for details.

$$T = \frac{S - \mu_S}{\sigma_S} \ , \ S = \sum_{t=1}^{N} R(d_2(t))^2 \tag{51}$$

$$\mu_S = \frac{N(2N+1)(4N+1)}{6} \ , \ \sigma_S = \sqrt{\frac{N^2(2N+1)(4N+1)(16N+11)}{180}} \tag{52}$$

$$F_{d_1}(x) = F_{d_2}(x) \ \forall x \wedge d_1, d_2 \text{ independent} \Rightarrow T \underset{N\to\infty}{\in} \mathrm{N}(0,1) \Rightarrow \tag{53}$$

$$p = 1 - F^{-1}_{\mathrm{N}(0,1)}(T) \tag{54}$$

Basic assumptions (B): The noise v_i has unknown correlation structure and distribution i.e. only the basic assumptions (45) applies. When $d_i(t)$ is not linear i.e. filtered white noise (4), correlation can not be removed by linear whitening filtering which makes the decimation procedure (NC) a must. After decimation the non-parametric method (W) can be used.

5 Non-linearity test procedures

At first glance your could expect that the three new methods SUP3, SUP5 and SIN can be combined with all four variance test methods. It is however difficult to make use of the non-parametric methods for non normal noise because of the necessary conditions (53) which the test is very sensitive to. The SUP methods must be excluded because the distribution of d_1 and d_2 are different if v_i is non normal. For the SIN method the distribution of d_1 and d_2 will be equal if the distribution of v_i are symmetric. This really only applies asymptotically because of the effect of the estimated sinus. Further the SIN method gives uncorrelated but dependent d_1 and d_2 for non normal noise and independence is the second conditions in (53). Fortunately this dependence improves rather than spoils the performance of the SIN method. It is actually possible to construct a SIN method using 4 sequences which complies with both conditions in (53) but its performance is inferior to the 2 sequences SIN method and it is therefore not included. Also the SUP3 method will be excluded from further analysis because it looses power in at least one case (57). This leaves the methods in table 1 for simulation analysis.

Method	Variance test method used	Description
SR		Suba-Rao method
SUP5	NW and NC	5 sequences superposition method for normal noise
SINN	NW and NC	SIN method for normal noise
SIN	W and B	SIN method for non normal symmetric noise

Table 1: Non linearity test method chosen for simulation analysis.

6 Statistical properties

Statistical tests can be compared by their power function i.e. the probability of rejecting H_0 as a function of the degree of violation of H_0 towards H_1. When H_0 is true the power is also known as false alarm probability and must be less or equal to α the level of significance and when ever H_1 is true the power, or detection probability, should be as high as possible. The only safe conclusion from the theory is that the power given H_0 and NW noise assumptions will equal α_0 for the SUP5 and SINN methods. Below the power is therefore estimated by simulation of different cases of H_0 and H_1 combined with all the different assumptions.

Examples and experimental conditions

Three examples will be considered for the deterministic part. A basic linear first order system (55) with gain one, q^{-1} is the backshift operator. Next only the input part is made nonlinear to give a Hammerstein type of nonlinearity (56) and finally the linear system is perturbed with a changing time constants depending on whether the output increases or decreases (57) as e.g. with direction dependent friction.

$$y_d(t) = \frac{b}{1+aq^{-1}}u(t-1)\ ,\ a=-0.9\ ,\ b=1+a \tag{55}$$

$$y_d(t) = \frac{b}{1+aq^{-1}}\tan^{-1}(ku(t-1))\ ,\ k=2 \tag{56}$$

$$\begin{aligned} y_d(t) &= -a(t)y_d(t-1)+b(t)u(t-1)\ , \\ a(t) &= \begin{cases} a_f = -(-a)^{k_\tau} & \text{if } y_d(t)\geq y_d(t-1) \\ a_s = -(-a)^{\frac{1}{k_\tau}} & \text{otherwise} \end{cases}\ ,\ b(t)=1+a(t)\ ,\ k_\tau = 1.25 \end{aligned} \tag{57}$$

The measured output follows (58) where there are six examples of noise from normal white (59) to non normal and correlated (64). The MA(10) process (60) is chosen in favor of AR(1) processes to test the robustness of the decimation method. The non normal cases (61)–(64) are with heavy tails as with "outliers" and skewed in the two last cases. The scaling results in equal standard deviation σ_v at 0.1. For the heavy tailed noise it was necessary to use a robustified LS [9, sec. 7.6] for estimating the sinus (35).

$$y(t) = y_d(t)+v(t) \tag{58}$$

$$v(t) = \sigma e(t)\ ,\ e(t)\in \text{NID}(0,1)\ ,\ \sigma=0.1 \tag{59}$$

$$v(t) = \sigma w(t)\ ,\ w(t)=M(q)e(t)\ ,\ M(q)=\frac{1+q^{-1}+\cdots+q^{-9}}{\sqrt{10}} \tag{60}$$

$$v(t) = \sigma\,\text{sign}(e(t))e(t)^4\kappa_1\ ,\ \kappa_1=\frac{1}{\sqrt{7\times5\times3}} \tag{61}$$

$$v(t) = \sigma\,\text{sign}(w(t))w(t)^4\kappa_1 \tag{62}$$

$$v(t) = \sigma(e(t)^2-1)\kappa_2\ ,\ \kappa_2=\frac{1}{\sqrt{2}} \tag{63}$$

$$v(t) = \sigma(w(t)^2 - 1)\kappa_2 \quad (64)$$

As suggested in [1] the input for the SR method is a sinus with a level different from zero. The amplitude and level are 0.5 to make it vary between 0 and 1 and thus give approximately 10% noise. The frequency equals the bandwidth of the deterministic system (55). Input to the SUP5 method is a modified PRBS signal [8, sec. 5.3] with the same level and RMS as the sinus and with a bandwidth as the system. The input to SIN and SINN has zero level and an amplitude given the same RMS as above. The total number of samples is set to 500 i.e. $1 \times 500, 5 \times 100$ and 2×250 for the SR, SUP5 and SIN methods respectively. The level of the test α_0 is set to the widely used 5% and the observed power can then be estimated as (65). Further it is reasonable to have $2\sigma_{\hat{\alpha}} = 0.01$ for the real power at 0.05 which is obtained with the number of simulations N_s at 1900.

$$\hat{\alpha} = \frac{\#\{p_i < 0.05\}}{N_s} \quad (65)$$

Results

For every combination of deterministic systems and noise 1900 simulations of 500 samples are performed i.e. $3 \times 6 \times 1900 = 34200$ simulations in total. For every simulation of 500 samples all 7 p-values corresponding to the methods in table 1 are calculated. Based on the 1900 p-values for each combination of deterministic system, noise and method the power is estimated by (65) which adds up to $3 \times 6 \times 7 = 126$ estimated powers. The results are presented in separate bar plots for each deterministic system. In the plots bars are grouped by method and noise process.

A carefull examination off all figures shows that the decimation procedure works very well. Only for non linear systems and heavy tailed correlated noise (b1 (62)) does the sinus methods without decimation perform better than the methods using decimation. In all other cases the methods using decimation are superior to the corresponding methods without decimation. Consequently the decimation based methods SUP5-NC, SINN-NC and SIN-B are in general preferable and we can consentrate on comparing these with the SR method.

In figure 1 it is seen that the existing SR method only works for normal white noise and the powers which in this case are false alarm probabilities are unacceptable high in all other cases especially the colored ones. The SUP5-NC method is a little better as it works also for normal colored noise. The SINN-NC method is perfectly reliable in all cases. The SIN-B method works for symmetrical distributed noise as expected. The reason for the zero false alarm probabilities for SIN-B is the special dependence between d_1 and d_2 arrising from heavy tailed noise, the full explanation is however to long and complicated to be included here.

Figure 2–3 shows detection probabilities for methods having false alarms less than 0.1. For the Hammerstein sytem the SR and SUP5-NC methods are more powerfull than the SINN-NC and SIN-B methods. The picture in figure 3 where the dynamical part is nonlinear is reversed in the sense that the SIN methods are very good in contrast to the poor performence of the SR and SUP5-NC methods.

7 Conclusion

This paper discusses and developes preliminary statistical tests for detecting nonlinearities in the deterministic part of SISO systems with noise. The most referenced method is based on third order correlations. It is shown to work for normal white noise but for other common noise processes e.g. colored, it is absolutely unreliable. Therefore two new methods based on basic principles are developed. One is based on superposition and the other on sinus input. They are much more robust, especially the sinus method whis is reliable also for colored, heavy tailed or skew distributed noise. All though there are cases where the

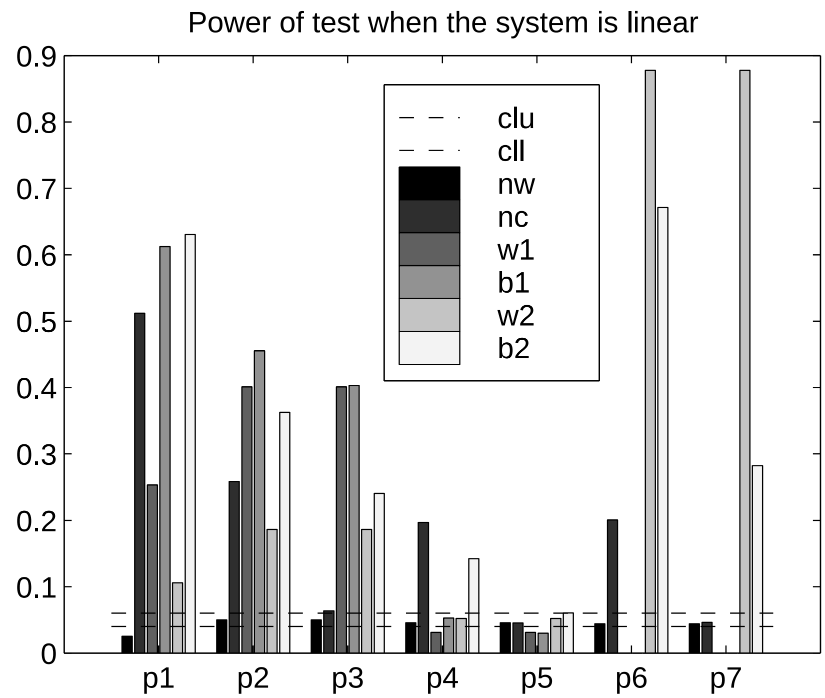

Figure 1: False alarm probabilities i.e. estimated power (65) when the system is linear. The 7 groups of bars corresponds to the methods in table 1. The 6 bars in each group corresponds to the noise cases. Notice that the 4 bars which can not be seen are because the corresponding $\hat{\alpha}$ are zero.

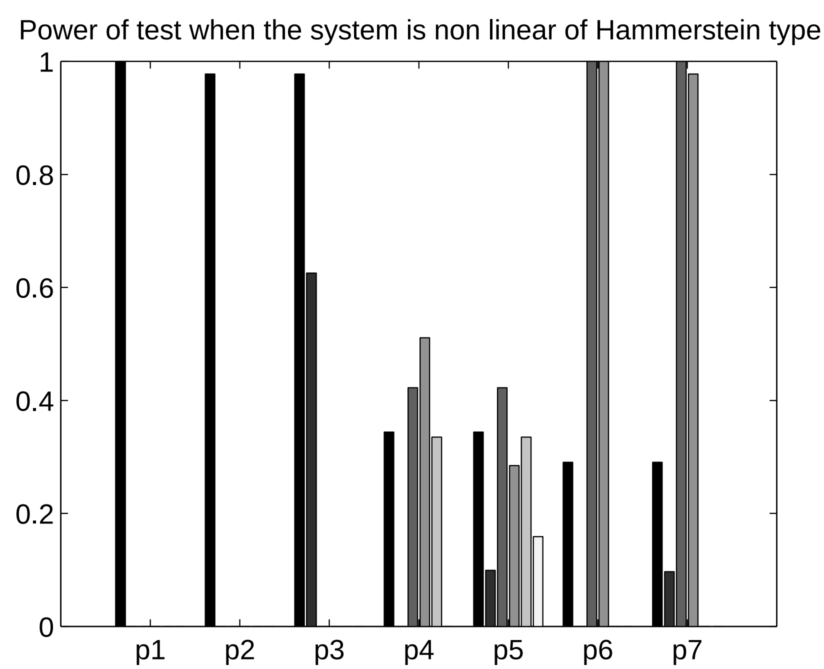

Figure 2: Detection probabilities i.e. estimated power when the system is non linear of the Hammerstein type (56). Bars corresponding to $\hat{\alpha} > 0.1$ when the system is linear (figure 1) are omitted.

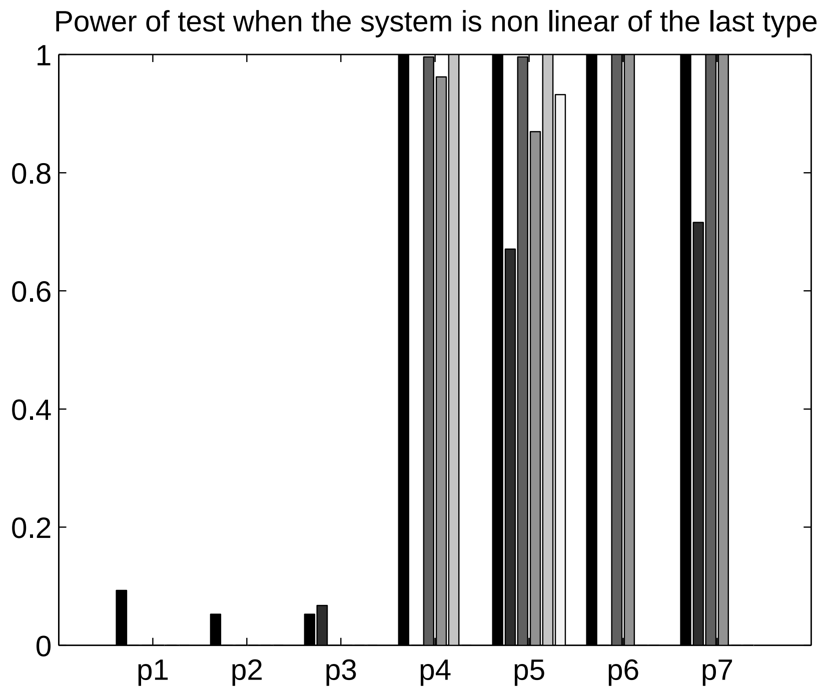

Figure 3: Detection probabilities i.e. estimated power when the system is non linear of the last type (57). Bars are omitted as in figure 2.

existing method has better detection capabilities compared to the new ones, an overall conclusions is that the new methods are better and especially the sinus method is a good choise in practise.

References

[1] S.A. Billings and W.S.F. Voon. Structure detection and model validity tests in the identification of nonlinear systems. *IEE Procedings D*, 130(4):193–199, July 1983.

[2] S.A. Billings and W.S.F. Voon. Correlation based model validity tests for non-linear models. *International Journal of Control*, 44(1):235–244, 1986.

[3] W. J. Conover. *Practical Nonparametric Statistics*. John Wiley & Sons, 1999.

[4] R. Haber. Nonlinearity tests for dynamic processes. In *7th IFAC Symposium on System Identification*, pages 409–413, 1985.

[5] R. Haber and H. Unbehauen. Structure identification of nonlinear dynamic systems - a survey on input/output approaches. *Automatica*, 26(4):651–677, 1990.

[6] Morten Knudsen. Determination of parameter estimation errors due to noise and undermodelling. In *IEEE Instrumentation and Measurement Technology Conference*, 1996.

[7] Lennart Ljung. *System Identification, Theory for the User*. Prentice-Hall information and system sciences series. Prentice-Hall, second edition, 1999.

[8] Torsten Söderström and Petre Stoica. *System Identification*. Prentice Hall International Series in System and Control Engineering. Prentice Hall, New York, 1989.

[9] Robert G. Staudte and Simon J. Sheather. *Robust Estimation and Testing*. John Wiley & Sons, Inc., 1990.

[10] Howell Tong. *Non-linear Time Series - A Dynamical System Approach*. Oxford Statistical Science Series. Clarendon Press Oxford, 1990.