Monograph:

Billings, S.A. and Zhu, Q.M. (1993) Nonlinear Model Validation Using Correlation Tests. Research Report. ACSE Research Report 463 . Department of Automatic Control and Systems Engineering

Nonlinear Model Validation Using Correlation Tests

S.A. Billings and Q.M. Zhu

Department of Automatic Control and Systems Engineering,
University of Sheffield, Sheffield S1 4DU, UK

Abstract:

New higher order correlation tests which use model residuals combined with system inputs and outputs are presented to check the validity of a general class of nonlinear models. The new method is illustrated by testing both simple and complex nonlinear system models.

Research Report No. 463

July 1993

Nonlinear Model Validation Using Correlation Tests

S.A. Billings and Q.M. Zhu

Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 4DU, UK

Abstract:

New higher order correlation tests which use model residuals combined with system inputs and outputs are presented to check the validity of a general class of nonlinear models. The new method is illustrated by testing both simple and complex nonlinear system models.

1.0 Introduction

Model validation is an important step in system identification since this is often the final check on the goodness of fit of any identified model. While linear models provide concise and efficient approximations to a wide range of linear systems they fail to adequately describe nonlinear systems. Nonlinear models are therefore increasingly used to approximate a wide variety of systems with complex dynamics. Model validation can also be divided into two main areas, linear model validation and nonlinear model validation.

A number of methods have been developed for linear model validation. Correlation based validation involves computing correlation functions composed of model residuals and system inputs and testing if these lie within preset confidence intervals. Bohlin(1971), Box and Jenkins (1976) and Soderstrom and Stoica (1990) studied the auto-correlation function (ACF) of the residuals and the cross-correlation function (CCF) between inputs and residuals and Pearson (1900), Bohlin (1971), and Baglivo, Olivier and Pagano (1992) investigated Chi-Squared (χ^2) tests. Recently Cressie and Read (1989) reviewed the literature on goodness of fit testing using the χ^2 test. Model comparison based validation involves applying statistical tests to compare models pairwise and to select the best model with the minimum or maximum statistic value. Representative approaches are the F test (Wadsworth and Bryan 1974) and the Akaike Information Criterion (AIC) (Akaike 1974).

Unfortunately nonlinear model validation it is not as straightforward as linear model validation. Several of the methods developed for linear models have obvious drawbacks for nonlinear model validation. For example Bohlin's ACF and CCF tests can not diagnose all nonlinear terms in the residuals (Billings and Voon 1983), and the χ^2 test and model comparison based tests can involve a combinational explosion because of the enormous number of possible terms in nonlinear models. Billings and Voon (1983, 1986), Leontaritis and Billings (1987) introduced higher order correlation functions and an extension to the χ^2 tests to overcome some of these problems.

Correlation based model validity tests have an advantage compared with model comparison based methods because it is possible to diagnose directly if an identified model is adequate or not without testing all the possible model sets. Model comparison methods however may involve testing over the vast combinations of models which are possible when the system is nonlinear and complex. But traditional correlation based model validity tests can sometimes exhibit reduced diagnosis power.

In the present study new tests designed to enhance the power of correlation based tests while simplifying the computations for nonlinear models are derived based on higher order correlation functions composed of residuals, inputs and particularly outputs are introduced. The characteristics of the new tests and the relationship to previous tests are investigated and the application of the tests to both simple and complex nonlinear system models is demonstrated.

2.0 Correlation tests using residuals and inputs

In order to understand the development of correlation based model validity tests and the derivation of new tests, linear model validation is considered initially and then progressively nonlinear model validation issues are introduced. One of the new tests is presented based on two higher order correlation functions, a higher order residual auto-correlation function and a higher order cross-correlation function between the residuals and input.

A SISO (Single Input and Single Output) linear discrete time model can be expressed as

$$y(t) = f_l(y^{t-1}, u^{t-1}, \varepsilon^{t-1}) + \varepsilon(t) \tag{2.1}$$

where t $(t=1, 2, ...)$ is a time index and

$$\begin{aligned} y^{t-1} &= [y(t-1), ..., y(t-r)] \\ u^{t-1} &= [u(t-1), ..., u(t-r)] \\ \varepsilon^{t-1} &= [\varepsilon(t-1), ..., \varepsilon(t-r)] \end{aligned} \tag{2.2}$$

are output, input and residual vectors respectively with delayed elements from 1 to r. $f_l(.)$ is a linear function which satisfies superposition and homogeneity

$$\begin{aligned} f_l(y^{t-1} + y^{t-2}, u^{t-1} + u^{t-2}, \varepsilon^{t-1} + \varepsilon^{t-2}) &= f_l(y^{t-1}, u^{t-1}, \varepsilon^{t-1}) \\ &+ f_l(y^{t-2}, u^{t-2}, \varepsilon^{t-2}) \\ f_l(\alpha y^{t-1}, \alpha u^{t-1}, \alpha\varepsilon^{t-1}) &= \alpha f_l(y^{t-1}, u^{t-1}, \varepsilon^{t-1}) \end{aligned} \tag{2.3}$$

A typical parametric realization of eqn (2.2) is the ARMAX (AutoRegressive Moving Average with eXogenous input) model

$$y(t) = \sum_{j=1}^{r} [\alpha_j y(t-j) + \beta_j u(t-j) + \lambda_j \varepsilon(t-j)] + \varepsilon(t)$$

(2.4)

Ideally the residual $\varepsilon(t)$ should be reduced to an uncorrelated sequence denoted by *e(t)* with zero mean and finite variance. Therefore correlation based model validity tests are used to check if

$$\varepsilon(t) \approx e(t)$$

(2.5)

This can be done by testing if all the correlation functions are within the preset confidence intervals. When eqn (2.5) is true Bohlin's test (1971) shows that

$$\phi_{\varepsilon\varepsilon}(\tau) = \frac{\sum_{t=1}^{N-\tau} (\varepsilon(t) - \bar{\varepsilon})(\varepsilon(t-\tau) - \bar{\varepsilon})}{\sum_{t=1}^{N} (\varepsilon(t) - \bar{\varepsilon})^2} = \begin{cases} 1, \tau = 0 \\ 0, \textit{otherwise} \end{cases}$$

$$\phi_{u\varepsilon}(\tau) = \frac{\sum_{t=1}^{N-\tau} (u(t) - \bar{u})(\varepsilon(t-\tau) - \bar{\varepsilon})}{\sqrt{\left(\sum_{t=1}^{N} (u(t) - \bar{u})^2\right)\left(\sum_{t=1}^{N} (\varepsilon(t) - \bar{\varepsilon})^2\right)}} = 0, \forall\tau$$

(2.6)

where $\phi_{\varepsilon\varepsilon}(\tau)$ and $\phi_{u\varepsilon}(\tau)$ are the normalised residual auto-correlation function and cross correlation function between the input and residuals respectively. The overbar denotes the time average operation to give

$$\bar{\varepsilon} = \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t)$$

$$\bar{u} = \frac{1}{N}\sum_{t=1}^{N} u(t)$$

(2.7)

For large N the correlation function estimates given in eqn (2.6) are asymptotically normal with zero mean and finite variance, the standard deviations are $1/\sqrt{N}$ and the *95%* confidence limits are therefore approximately $1.95/\sqrt{N}$. To illustrate the method consider three typical residuals from an ARMAX model as examples

$$\begin{aligned} \varepsilon_1(t) &= e(t-1) + e(t) \\ \varepsilon_2(t) &= u(t-1) + e(t) \\ \varepsilon_3(t) &= y(t-1) + e(t) \end{aligned} \tag{2.8}$$

where the input $u(t)$ is an uncorrelated persistently exciting sequence with zero mean and finite variance and the noise $e(t)$ is defined above. $\phi_{\varepsilon\varepsilon}(\tau)$ can be used to check for delayed noise terms like $e(t-j)$ in $\varepsilon(t)$. Similarly $\phi_{u\varepsilon}(\tau)$ can be used to check for $u(t-j)$ terms in $\varepsilon(t)$. When the residuals include delayed outputs like $y(t-j)$ both $\phi_{\varepsilon\varepsilon}(\tau)$ and $\phi_{u\varepsilon}(\tau)$ will give an indication that $\varepsilon(t)$ is correlated because $y(t-1)$ is auto-correlated and cross-correlated with the input.

Using simple algebraic operations, the first example in eqn (2.8) to gives

$$\begin{aligned} \phi_{\varepsilon_1\varepsilon_1}(\tau) &= \begin{cases} 1, \tau = 0 \\ \rho, \tau = 1 \\ 0, \textit{otherwise} \end{cases} \\ \phi_{u\varepsilon_1}(\tau) &= 0, \forall\tau \end{aligned} \tag{2.9}$$

where $0 < |\rho| < 1$ is a constant. For the second example

$$\begin{aligned} \phi_{\varepsilon_2\varepsilon_2}(\tau) &= \begin{cases} 1, \tau = 0 \\ 0, \textit{otherwise} \end{cases} \\ \phi_{u\varepsilon_2}(\tau) &= \begin{cases} \rho, \tau = 1 \\ 0, \textit{otherwise} \end{cases} \end{aligned} \tag{2.10}$$

In summary Bohlin's and Box and Jenkins' approach is to determine whether there is evidence of an inadequate model and also to suggest ways in which the model may be modified or improved.

For nonlinear models the validity tests are not as simple as in the linear case because nonlinear terms can exist in the residuals. Consider a SISO nonlinear discrete model

$$y(t) = f_n(y^{t-1}, u^{t-1}, \varepsilon^{t-1}) + \varepsilon(t) \tag{2.11}$$

where the output, input and residual vectors have been defined in eqn (2.2). However the nonlinear function $f_n(.)$ will not in general satisfy the superposition and homogeneity principle. A typical parametric expression for eqn (2.11) is the polynomial NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous input) model (Leontaritis and Billings 1985)

$$y(t) = \sum_{j=1}^{N} \alpha_j p_j(t) + \varepsilon(t)$$

(2.12)

where $p(t)$ denotes nonlinear terms such as $p_1(t)=y(t-1)u(t-1)$, $p_2(t)=u^2(t)\varepsilon(t-1)$ and $p_3(t)=\varepsilon^3(t-3)$.

Simple nonlinear residuals may take the form of terms such as

$$\begin{aligned}
\varepsilon_1(t) &= e(t-2)\,e(t-5) + e(t) \\
\varepsilon_2(t) &= u(t-1)\,u(t-3) + e(t) \\
\varepsilon_3(t) &= u(t-1)\,e(t-2) + e(t) \\
\varepsilon_4(t) &= y(t-1)\,e(t-2) + e(t)
\end{aligned}$$

(2.13)

and so on.

The simple ACF and CCF tests are now no longer sufficient (Billings and Voon 1983) and new tests have to be developed. One approach would be to use multidimensional correlation functions such as $\phi_{\varepsilon\varepsilon\varepsilon}(\tau_1,\tau_2)$,$\phi_{uu\varepsilon}(\tau_1,\tau_2)$ and $\phi_{u\varepsilon\varepsilon}(\tau_1,\tau_2)$ to check for $\varepsilon_1(t)$, $\varepsilon_2(t)$, and $\varepsilon_3(t)$ respectively in eqn (2.13). This approach however involves two dimensional correlations and this causes an enormous increase in the computations. This could be extended to a 3-D correlation function $\phi_{\varepsilon\varepsilon\varepsilon\varepsilon}(\tau_1,\tau_2,\tau_3)$ for the case

$$\varepsilon(t) = e(t-1)\,e(t-2)\,e(t-5) + e(t)$$

(2.14)

but this is clearly unrealistic in practice. Alternatively an n dimensional correlation function can be projected into a single index higher order correlation function with n points. This approach leads to the introduction of two higher order correlation functions

$$\phi_{\varepsilon^2\varepsilon^2}(\tau) = \frac{\sum_{t=1}^{N-\tau} (\varepsilon^2(t) - \overline{\varepsilon^2})(\varepsilon^2(t-\tau) - \overline{\varepsilon^2})}{\sum_{t=1}^{N} (\varepsilon^2(t) - \overline{\varepsilon^2})^2}$$

$$\phi_{u^2\varepsilon^2}(\tau) = \frac{\sum_{t=1}^{N-\tau}(u^2(t)-\overline{u^2})(\varepsilon^2(t-\tau)-\overline{\varepsilon^2})}{\sqrt{\left(\sum_{t=1}^{N}(u^2(t)-\overline{u^2})^2\right)\left(\sum_{t=1}^{N}(\varepsilon^2(t)-\overline{\varepsilon^2})^2\right)}}$$

(2.15)

where

$$\overline{\varepsilon^2} = \frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t)$$

$$\overline{u^2} = \frac{1}{N}\sum_{t=1}^{N}u^2(t)$$

(2.16)

$\phi_{\varepsilon^2\varepsilon^2}(\tau)$ can be used to detect delayed nonlinear noise terms like $e(t-2)e^3(t-5)$ and $\phi_{u^2\varepsilon^2}(\tau)$ can be used to detect delayed nonlinear input terms like $u(t-1)u(t-2)$. Both $\phi_{\varepsilon^2\varepsilon^2}(\tau)$ and $\phi_{u^2\varepsilon^2}(\tau)$ should detect other terms such as $u(t-1)e(t-1)$, $y(t-1)u(t-1)$ and so on in the residuals. When the residuals $\varepsilon(t)$ are reduced to a sequence which is uncorrelated with all linear and nonlinear combinations of past inputs and outputs then ideally

$$\phi_{\varepsilon^2\varepsilon^2}(\tau) = \begin{cases} 1, \tau = 1 \\ 0, \textit{otherwise} \end{cases}$$

$$\phi_{u^2\varepsilon^2}(\tau) = 0, \forall\tau$$

(2.17)

This follows because for large N and assuming ergodicity

$$\phi_{\varepsilon^2\varepsilon^2}(\tau) = E[\varepsilon^{2^o}(t)\,\varepsilon^{2^o}(t-\tau)]$$

$$\phi_{u^2\varepsilon^2}(\tau) = E[u^{2^o}(t)\,\varepsilon^{2^o}(t-\tau)]$$

(2.18)

where $E[.]$ is the expectation operator and

$$\varepsilon^{2^o}(t) = \frac{\varepsilon^2(t)-E[\varepsilon^2(t)]}{\sqrt{E[[\varepsilon^2(t)-E[\varepsilon^2(t)]]^2]}}$$

$$u^{2^o}(t) = \frac{u^2(t)-E[u^2(t)]}{\sqrt{E[[u^2(t)-E[u^2(t)]]^2]}}$$

(2.19)

are normalized. When $\varepsilon(t)=e(t)$ these tests yield

$$\phi_{\varepsilon^2\varepsilon^2}(\tau) = E[\varepsilon^{2^\circ}(t)\,\varepsilon^{2^\circ}(t-\tau)]$$

$$= E[e^{2^\circ}(t)\,e^{2^\circ}(t-\tau)]$$

$$= \begin{cases} E[(e^{2^\circ}(t))^2] = 1, \tau = 0 \\ E[e^{2^\circ}(t)]\,E[e^{2^\circ}(t-\tau)] = 0, otherwise \end{cases}$$

$$\phi_{u^2\varepsilon^2}(\tau) = E[u^{2^\circ}(t)\,\varepsilon^{2^\circ}(t-\tau)]$$

$$= E[u^{2^\circ}(t)]\,E[e^{2^\circ}(t-\tau)] = 0, \forall\tau$$

(2.20)

To illustrate these tests consider the first two examples in eqn (2.13). For the first case

$$\phi_{\varepsilon_1^2\varepsilon_1^2}(\tau) = \begin{cases} 1, \tau = 0 \\ \rho_1, \tau = 2 \\ \rho_2, \tau = 5 \\ 0, otherwise \end{cases}$$

$$\phi_{u^2\varepsilon_1^2}(\tau) = 0, \forall\tau$$

(2.21)

where $0 < |\rho_1| < 1$ and $0 < |\rho_2| < 1$ are constants. For the second example

$$\phi_{\varepsilon_2^2\varepsilon_2^2}(\tau) = \begin{cases} 1, \tau = 0 \\ 0, otherwise \end{cases}$$

$$\phi_{u^2\varepsilon_2^2}(\tau) = \begin{cases} 1, \tau = 0 \\ \rho_1, \tau = 1 \\ \rho_2, \tau = 3 \\ 0, otherwise \end{cases}$$

(2.22)

The tests correctly identify the missing terms in both cases since either $\phi_{\varepsilon^2\varepsilon^2}(\tau)\neq 0, \tau\neq 0$ or $\phi_{u^2\varepsilon^2}(\tau)\neq 0, \forall\ \tau$. The two tests can also be used to diagnose omitted linear model terms in the residuals Applying the tests to the first two linear residual examples of eqn (2.8) gives (EQ 1.1)

$$\phi_{\varepsilon_1^2\varepsilon_1^2}(\tau) = \begin{cases} 1, \tau = 0 \\ \rho_1, \tau = 1 \\ 0, otherwise \end{cases}$$
$$\phi_{u^2\xi_1^2}(\tau) = 0, \forall\tau$$

(1.2)

and for the second example

$$\phi_{\varepsilon_2^2\varepsilon_2^2}(\tau) = \begin{cases} 1, \tau = 0 \\ 0, otherwise \end{cases}$$
$$\phi_{u^2\varepsilon_2^2}(\tau) = \begin{cases} \rho_1, \tau = 1 \\ 0, otherwise \end{cases}$$

(1.3)

The disadvantage of these tests is that the higher order correlation functions can sometimes exhibit less power when the noise and input variances are small because the fourth and higher moments become small. Billings and Voon (1986) used a combination of five tests to try to increase the discriminatory power. An alternative solution may be to introduce delayed output terms to form more powerful higher order correlation functions while maintaining the computational simplicity. This is presented in the following section.

3.0 Correlation tests using residuals, inputs and outputs

Two new tests based on higher order cross correlation functions between the output and residuals and between the output, residuals and input, are presented to enhance the power of correlation based model validity tests. These two tests are

$$\phi_{\alpha\varepsilon^2}(\tau) = \frac{\sum_{t=1}^{N-\tau} (\alpha(t) - \bar{\alpha})(\varepsilon^2(t-\tau) - \overline{\varepsilon^2})}{\sqrt{\left(\sum_{t=1}^{N} (\alpha(t) - \bar{\alpha})^2\right)\left(\sum_{t=1}^{N} (\varepsilon^2(t) - \overline{\varepsilon^2})^2\right)}}$$

$$\phi_{\alpha u^2}(\tau) = \frac{\sum_{t=1}^{N-\tau} (\alpha(t) - \bar{\alpha})(u^2(t-\tau) - \overline{u^2})}{\sqrt{\left(\sum_{t=1}^{N} (\alpha(t) - \bar{\alpha})^2\right)\left(\sum_{t=1}^{N} (u^2(t) - \overline{u^2})^2\right)}}$$

(3.1)

where

$$\alpha(t) = y(t)\varepsilon(t)$$
$$\bar{\alpha} = \overline{y\varepsilon} = \frac{1}{N}\sum_{t=1}^{N} y(t)\varepsilon(t)$$

(3.2)

In the ideal case where the residuals are zero mean and uncorrelated with all linear and nonlinear combinations of past inputs and outputs these tests yield

$$\phi_{\alpha\varepsilon^2}(\tau) = \begin{cases} k_2, \tau = 0 \\ 0, \textit{otherwise} \end{cases}$$
$$\phi_{\alpha u^2}(\tau) = 0, \forall\tau$$

(3.3)

Where k_2 is a constant to be defined in eqn (3.6) and these results can be proved by considering the output of eqn (2.11)

$$y(t) = f_n(y^{t-1}, u^{t-1}, \varepsilon^{t-1}) + \varepsilon(t)$$
$$= \hat{y}(t) + \varepsilon(t)$$

(3.4)

which consists of the one step ahead predicted output plus residual. Therefore eqn (3.1) can be written as

$$\phi_{\alpha\varepsilon^2}(\tau) = \phi_{(y\varepsilon)\varepsilon^2}(\tau) = k_1\phi_{(\hat{y}\varepsilon)\varepsilon^2}(\tau) + k_2\phi_{\varepsilon^2\varepsilon^2}(\tau)$$
$$\phi_{\alpha u^2}(\tau) = \phi_{(y\varepsilon)u^2}(\tau) = k_1\phi_{(\hat{y}\varepsilon)u^2}(\tau) + k_2\phi_{u^2\varepsilon^2}(\tau)$$

(3.5)

where

$$k_1 = \frac{\sqrt{\sum_{t=1}^{N}(\hat{y}(t)\varepsilon(t) - \overline{\hat{y}\varepsilon})^2}}{\sqrt{\sum_{t=1}^{N}(y(t)\varepsilon(t) - \overline{y\varepsilon})^2}} \qquad k_2 = \frac{\sqrt{\sum_{t=1}^{N}(\varepsilon^2(t) - \overline{\varepsilon^2})^2}}{\sqrt{\sum_{t=1}^{N}(y(t)\varepsilon(t) - \overline{y\varepsilon})^2}}$$

(3.6)

In the ideal case where the model is unbiased such that $\varepsilon(t)$ is reduced to a zero mean uncorrelated sequence

$$\phi_{(\hat{y}\varepsilon)\varepsilon^2}(\tau) = 0, \forall\tau$$
$$\phi_{(\hat{y}\varepsilon)u^2}(\tau) = 0, \forall\tau$$

(3.7)

and therefore eqn (3.5) becomes

$$\phi_{\alpha\varepsilon^2}(\tau) = k_2 \phi_{\varepsilon^2\varepsilon^2}(\tau)$$
$$\phi_{\alpha u^2}(\tau) = k_2 \phi_{u^2\varepsilon^2}(\tau)$$

(3.8)

which are the same, except for the constant k_2, as the correlation functions presented with only residuals and inputs in eqn (2.15).

If the model is inadequate $\phi_{(\hat{y}\varepsilon)\varepsilon^2}(\tau)$ and $\phi_{(\hat{y}\varepsilon)u^2}(\tau)$ will test the correlation between the one step ahead predicted output from the identified model and the residual. These enhance the tests based on residuals and inputs only. It should be noticed that unlike traditional normalised ACF test where $\phi_{\varepsilon\varepsilon}(0) = 1$ is not effected by the amplitude of the residual, the new test $\phi_{(y\varepsilon)\varepsilon^2}(0) / k_2$ is affected by the residual to produce $\phi_{(y\varepsilon)\varepsilon^2}(0) / k_2 = 1$ in the ideal case when the one step ahead prediction and the residual are uncorrelated but can be $| \phi_{(y\varepsilon)\varepsilon^2}(0) / k_2 | < 1$ otherwise.

For large N the correlation function estimates given in eqn (3.1) are still asymptotically normal with zero mean and finite variance from the assumption of the central limit theorem (Bowker and Lieberman 1972) and the standard deviations are $1/\sqrt{N}$ and the 95% confidence limits are therefore approximately $1.95/ \sqrt{N}$.

An associated χ^2 test can also be developed. Define

$$w'_{\varepsilon}(t) = \frac{\varepsilon^2(t) - \overline{\varepsilon^2}}{\sqrt{\sum_{t=1}^{N} (\varepsilon^2(t) - \overline{\varepsilon^2})^2}}$$

(3.9)

and

$$\mu_{\varepsilon} = \frac{1}{N} \sum_{t=1}^{N} \vec{\alpha}(t)\, w_{\varepsilon}(t)$$

(3.10)

where

$$\vec{\alpha}(t) = [\alpha(t-1), \alpha(t-2), \ldots, \alpha(t-s)]^T$$
$$\alpha(t) = y(t)\,\varepsilon(t)$$
$$\frac{1}{N} \sum_{t=1}^{N} \alpha^2(t) = \Gamma^2$$

(3.11)

Assuming all odd order moments of the random variable $\varepsilon(t)$ are zero gives

$$E[\mu_\varepsilon] = \frac{1}{N}E\left[\sum_{t=1}^{N}\vec{\alpha}(t)\,w_\varepsilon(t)\right] = \frac{1}{N}\sum_{t=1}^{N}E[\vec{\alpha}(t)]\,E[w_\varepsilon(t)] = 0_{s\times 1}$$

$$E[\mu_\varepsilon\mu_\varepsilon^T] = \frac{1}{N}\Gamma^2 I_{s\times s}$$

(3.12)

From the central limit theorem (Bowker and Liebernan 1972) the random vector is asymptotically normal with zero mean and variance given by eqn (3.12). The random vector μ_ε can be normalized as

$$\varsigma = \frac{\sqrt{N}}{\Gamma}\mu_\varepsilon I_{s\times s}$$

(3.13)

which is asymptotically zero mean with unit variance. Then the variable

$$d_\varepsilon = \varsigma_\varepsilon^T\varsigma_\varepsilon = \frac{N\mu_\varepsilon^T\mu_\varepsilon}{\Gamma^2}$$

(3.14)

is asymptotically χ^2 distributed with s degrees of freedom where s is the dimension of the vector $\alpha(t)$. This statistic d_ε provides an alternative basis for nonlinear model validation. The confidence interval of d_ε is given by

$$d_\varepsilon < k_\gamma(s)$$

(3.15)

where $k_\gamma(s)$ is the critical value of the χ^2 distribution with s degrees of freedom and γ is the significance level for the model acceptance region. Similarly a χ^2 statistic for the input test in the residuals can be developed by assuming the worst case where the input is a random sequence with odd moments all zero.

$$d_u < k_\gamma(s)$$

(3.16)

where

$$d_u = \varsigma_u^T\varsigma_u = \frac{N\mu_u^T\mu_u}{\Gamma^2}$$

$$\mu_u = \frac{1}{N}\sum_{t=1}^{N}\vec{\alpha}(t)\,w_u(t)$$

$$w_u(t) = \frac{u^2(t) - \overline{u^2}}{\sqrt{\sum_{t=1}^{N} \left(u^2(t) - \overline{u^2}\right)^2}}$$

(3.17)

The previous χ^2 test (Leontaritis and Billings 1985) for nonlinear models suffered from the necessity to test several possible missing model terms before any confidence that the model had been properly validated could be established. The new test gives a rule to choose test terms $y(t)\varepsilon(t)$, $u^2(t)$, $\varepsilon^2(t)$ only.

4.0 Applications

Three simulated systems were selected to demonstrate the new model validity test methods. Each data sequence was of length *1000*, and the input for the first two simulated systems was a uniformly distributed uncorrelated sequence with zero mean and variance *1.33* and the noise sequence was an uncorrelated normally distributed sequence with zero mean and variance *0.36*.

Example one

This simulated system consisted of the model

$$y(t) = u(t-1) + e(t-2)e(t-5) + e(t)$$

(4.1)

Assuming that an inadequate model has been estimated so that the residuals become

$$\varepsilon(t) = e(t-2)e(t-5) + e(t)$$

(4.2)

Figure 1(a) shows that the results obtained from using the simple linear ACF and CCF tests give a false indication that the model has been properly identified because $\phi_{\varepsilon\varepsilon}(\tau)=\delta(\tau)$ and $\phi_{u\varepsilon}(\tau)=0$. Figure 1(b) however shows that the results obtained using the new tests $\phi_{(y\varepsilon)\varepsilon^2}(\tau)$ and $\phi_{(y\varepsilon)u^2}(\tau)$ give a correct indication that the model has been incorrectly identified. Close inspection shows that $\phi_{(y\varepsilon)\varepsilon^2}(\tau)$ is outside the confidence intervals at two points $\tau=2$ and 5.

Example two

Consider the model

$$y(t) = u(t-1) + u(t-2)e(t-5) + e(t)$$

(4.3)

and assume a residual of the form

$$\varepsilon(t) = u(t-2)e(t-5) + e(t)$$

(4.4)

Once again the linear ACF and CCF tests Figure 2(a) fail to detect the nonlinear residuals. The new tests $\phi_{(y\varepsilon)\varepsilon^2}(\tau)$ and $\phi_{(y\varepsilon)u^2}(\tau)$ in Figure 2(b) are plotted to show that $\phi_{(y\varepsilon)\varepsilon^2}(\tau)$ and $\phi_{(y\varepsilon)u^2}(\tau)$ are outside confidence intervals at $\tau = 5, 2$ respectively.

Example three

Consider the nonlinear rational model

$$y(t) = \frac{y(t-1) + u(t-1) + y(t-1)u(t-1) + y(t-1)e(t-1)}{1 + y^2(t-1) + u(t-1)e(t-1)} + e(t) \tag{4.5}$$

which is of a form which often appears in the chemical and related fields (Ford, Titterington and Kitsos 1989, Dimitrov and Kamenski 1991) and was used by Narendra and Parthasarathy (1990) to study the approximation of severe nonlinearities using neural networks.

Simulated output data were generated using the model of eqn (4.5) with uncorrelated uniformly distributed (zero mean and variance of *1.0)* input sequence and uncorrelated normally distributed (zero mean and variance of *0.01*) noise sequence. Only the input and output data were available for identification. The resultant identification using the nonlinear rational model identification algorithm (Zhu and Billings 1993, Billings and Zhu 1993) which included model term selection and associated parameter estimation produced the results in Table 1. Figure 3 (a,b,c) show the measured output, the one step ahead model predictions and residuals respectively and Figure 4 (a,b) show the linear ACF and CCF tests and the new tests respectively.

Terms Selected	**Parameters estimated**
Numerator	
u(t-1)	*0.996*
y(t-1)	*1.016*
y(t-1)u(t-1)	*0.980*
y(t-1)e(t-1)	*1.015*
Denominator	
y2(t-1)	*0.998*
u(t-1)e(t-1)	*0.925*
Noise Variance	*0.011*

Table 1

5.0 Conclusions

The important problem of validating nonlinear models has been investigated and two new correlation based tests have been proposed. The new tests are based on correlation functions defined in terms of the system outputs and these appear to provide improved discriminatory performance compared with earlier tests based purely on residuals and inputs. Although only simple nonlinear polynomial and rational model examples have been used as illustrations, largely to enable the results to be related back to the omitted model terms, the new tests should be applicable to a much wider class of nonlinear models including neural networks.

Acknowledgment

The authors gratefully acknowledge that this work is supported by SERC under grant GR/H3528.6.

References

Akaike, H., 1974, A new look at the statistical model identification. *IEEE Trans. Autom Control*, **AC-19**, 716-723.

Baglivo, J., Olivier, D., and Pagano, M., 1992, Methods for exact goodness of fit tests. *Journal of the American Statistical Association*, **87**, 464-469.

Billings, S. A. and I. J. Leontaritis, 1981, Identification of nonlinear systems using parameter estimation techniques. *IEE Conference Proceedings on Control and its applications*, 183-187. University of Warwick, Coventry, UK.

Billings, S. A. and Voon, W. S. F., 1983, Structure detection and model validity tests in the identification of nonlinear systems. *IEE Proceedings, Pt. D*, **130**, 193-199.

Billings, S. A. and Voon, W. S. F., 1986, Correlation based model validity tests for nonlinear models. *Int. J. Control*, **44**, 235-244.

Billings, S. A. and Zhu, Q. M., 1993, Structure detection algorithm for nonlinear rational models. *Int. J. Control* (to be published).

Bohlin, T., 1971, On the problem of ambiguities in maximum likelihood identification. *Automatica*, **7**, 199-210.

Bohlin, T, 1978, Maximum power validation of models without higher order fitting. *Automatica*, **14**, 137-146.

Bowker, A. H. and Lieberman, G. J., 1972, *Engineering statistics* (New Jersey: Prentice-Hall, Inc).

Box, G. E. P. and Jenkins, G. M., 1976, *Time series analysis forecasting and control* (San Francisco: Holden-Day).

Cressie, N. and Read, T. R. C., 1989, Person's X and the loglikelihood ratio statistic G : a comparative review. *International Statistical Review*, **57**, 19-43.

Dimitrov, S. D. and Kamenski, D. I., 1991, A parameter estimation method for rational functions. *Computers Chem. Engng.*, **15**, 657-662.

Ford, I., Titterington, D. M., and Kitsos, C. P., 1989, Recent advances in nonlinear experimental design. *Technomatrics*, **31**, 49-60.

Leontaritis, I. J. and Billings, S. A., 1985, Input-output parametric models for nonlinear systems, Part I and II. *I.J. Control*, **41**, 303-328, 329-344.

Leontaritis, I. J., and Billings, S. A., 1987, Model selection and validation methods for nonlinear systems. *Int. J. Control*, **45**, 311-341.

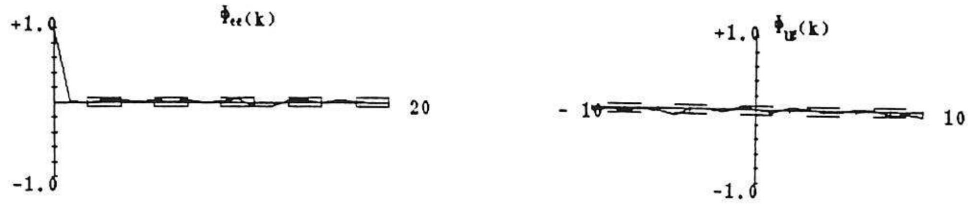
Narendra, K. S. and K. Parthasarathy, K., 1990, Identification and control of dynamical systems using neural networks. *IEEE Trans. on neural networks*, **1**, 4-27.

Pearson, K., 1900, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophy Magazine*, **50**, 157-172.

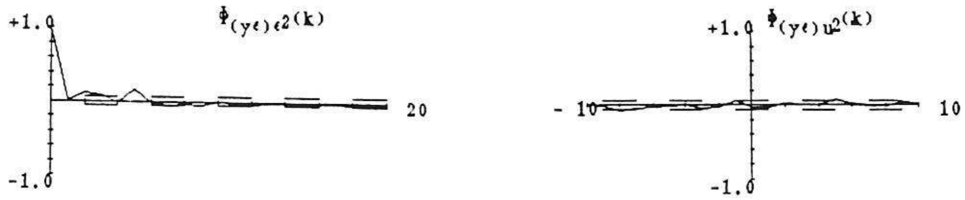
Soderstrom, T. and Stoica, P., 1990, On covariance function tests used in system identification. *Automatica*, **26**, 125-133.

Wadsworth, G. P. and Bryan, J. G., 1974, *Applications of probability and random variables*. (New York: McGraw-Hill).

Zhu, Q. M. and S.A. Billings, S. A., 1993, Parameter estimation for stochastic nonlinear rational models. *Int. J. Control*, **57**, 309-333.

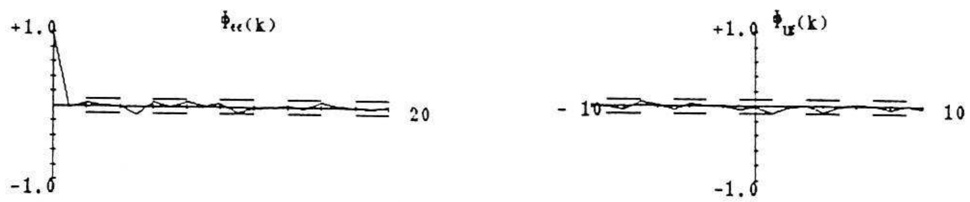


(a) ACF and CCF tests

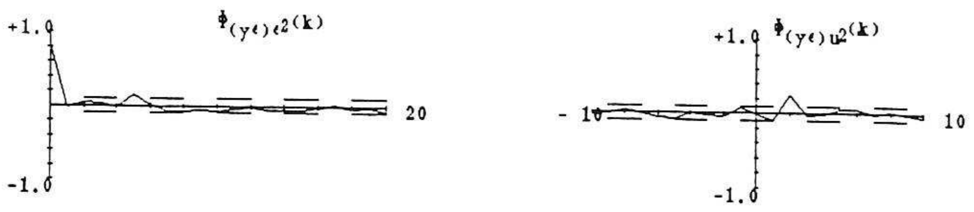


(b) New tests

Figure 1 Validation tests for example one

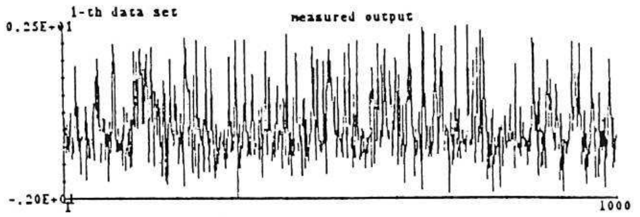


(a) ACF and CCF tests

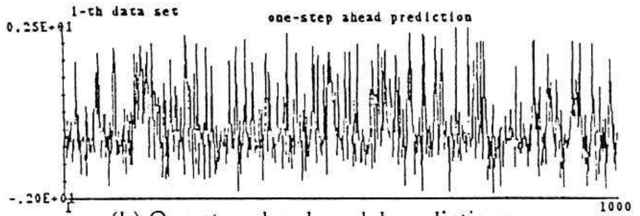


(b) New tests

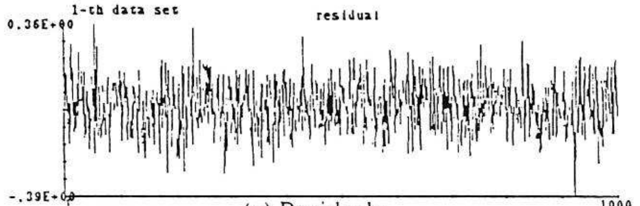
Figure 2 Validation tests for example two



(a) Measured outputs

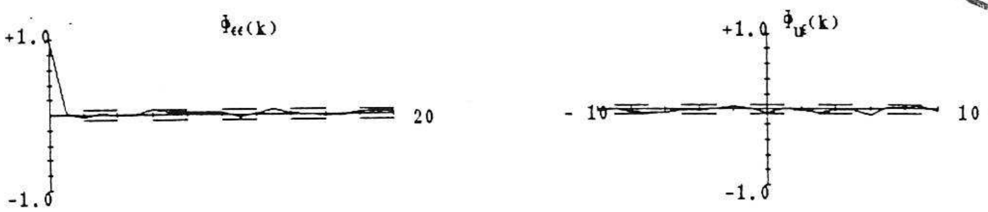


(b) One step ahead model predictions

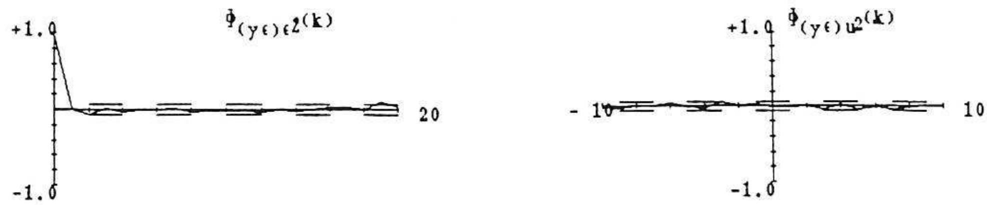


(c) Residuals

Figure 3 System outputs, model predictions and residuals



(a) ACF and CCF tests



(b) New tests

Figure 4 Validation tests for example three