Monograph:
Billings, S.A. and Zhu, Q.M. (1994) Model Variation Tests for Multivariable Nonlinear Models Including Neural Networks. Research Report. ACSE Research Report 484 . Department of Automatic Control and Systems Engineering

Model Validation Tests for Multivariable Nonlinear models Including Neural Networks

S.A. Billings and Q.M. Zhu

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

A fast and concise MIMO nonlinear model validity test procedure is derived, based on higher order correlation functions, to form a global to local hierarchical validation diagnosis of identified MIMO linear and nonlinear models. The new procedure is applied to four MIMO nonlinear system models including a neural network training example to demonstrate the effectiveness of the tests.

Research Report No 484

January 1994

Model Validation Tests for Multivariable Nonlinear Models Including Neural Networks

S.A. Billings and Q.M. Zhu

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

Abstract:

A fast and concise MIMO nonlinear model validity test procedure is derived, based on higher order correlation functions, to form a global to local hierarchical validation diagnosis of identified MIMO linear and nonlinear models. The new procedure is applied to four MIMO nonlinear system models including a neural network training example to demonstrate the effectiveness of the tests.

1.0 Introduction

Dynamic modelling is widely applied in engineering analysis and design, financial forecasting, weather prediction, studies of complex social phenomena and many other systems. Models of these systems can be developed using many different approaches ranging from analytical modelling procedures through least squares parameter estimation to neural network algorithms. But one fundamental issue which is relevant to all these cases is model validation. Once a model of a system has been determined it is important to test the validity of the model, to determine if the model is representative of the underlying system.

There are several approaches to model validation (Box and Jenkins 1976, Bohlin 1971, 1978, Soderstrom and Stoica 1990) but one of the most powerful methods is based on the concept that if the model is correct the residuals should be a completely random sequence. This is relatively easy to test if the system is assumed to be linear because the autocorrelation function of the residuals and the cross correlation function between the input and the residuals, or closely related methods, provide adequate tests. But these tests are not sufficient for nonlinear systems and higher order correlation functions have to be introduced in an attempt to detect all possible missing nonlinear terms in the residuals (Billings and Voon 1983, 1986, Subba Rao and Gabr 1984).

The majority of the tests have been developed for single-input single-output (SISO) models and have exploited the relationship between the model input and the residuals. While extensions to the multi-input multi-output (MIMO) case has either been given or is relatively straightforward (Subba Rao and Gabr 1984, Billings, Chen and Korenberg 1989), this often produces a large number of correlation plots which have to be inspected.

The present study is an attempt to overcome some of the above difficulties by developing tests which utilize the information in the inputs, residuals and the model outputs. Recent results derived for the single-input single-output case (Billings and Zhu 1994) which showed the advantages of using model outputs to develop new tests, are extended to the general class of nonlinear multi-input multi-output models. A new concise nonlinear MIMO model validity test procedure is derived. This is based on a global to local hierarchical diagnosis procedure where initial tests are used to test the overall model validity and more detailed focussed tests are only used if the model is found to be inadequate. The application of the new tests is demonstrated using several nonlinear multi-input multi-output examples including validating a neural network model.

2.0 Problem formulation

Consider the general MIMO model representation

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

where t (t=1, 2, ...) is a time index, $y(t)$, $u(t)$ and $\varepsilon(t)$ denote the output, input and residual vectors respectively, and $f(.)$ is the vector valued linear or nonlinear function so that

$$y(t) = \begin{bmatrix} y_1(t) \\ \dots \\ y_q(t) \end{bmatrix} \quad u(t) = \begin{bmatrix} u_1(t) \\ \dots \\ u_r(t) \end{bmatrix} \quad \varepsilon(t) = \begin{bmatrix} \varepsilon_1(t) \\ \dots \\ \varepsilon_q(t) \end{bmatrix} \quad f = \begin{bmatrix} f_1 \\ \dots \\ f_q \end{bmatrix} \tag{2.2}$$

where q is the number of outputs and r the number of model inputs. Define

$$y^{t-1} = \begin{bmatrix} y_1^{t-1} \\ \dots \\ y_q^{t-1} \end{bmatrix} \quad u^{t-1} = \begin{bmatrix} u_1^{t-1} \\ \dots \\ u_r^{t-1} \end{bmatrix} \quad \varepsilon^{t-1} = \begin{bmatrix} \varepsilon_1^{t-1} \\ \dots \\ \varepsilon_q^{t-1} \end{bmatrix} \tag{2.3}$$

where

$$y_i^{t-1} = [y_i(t-1), \dots, y_i(t-n_y)]$$
$$u_i^{t-1} = [u_i(t-1), \dots, u_i(t-n_u)]$$
$$\varepsilon_i^{t-1} = [\varepsilon_i(t-1), \dots, \varepsilon_i(t-n_\varepsilon)]$$

(2.4)

Notice that the form of the model can be very wide and can include the MIMO linear model, the Nonlinear AutoRegressive Moving Average with eXogenous input (NARMAX) model (Billings and Chen 1989), a neural network expansion etc.

The selection of one of two models to describe data can be formulated as a statistical hypothesis testing problem. Two hypotheses are always included in the model selection, the null hypothesis denoted by H_0 and the alternative hypothesis denoted by H_1. Usually in model validation H_0 is taken as the identified model.

Statistical model validity tests mainly consist of the following three steps (Kendall and Stuart 1967, Bohlin 1978). The first step is to form a parameter free statistic, which is a function of the available data such that the distribution of the statistic variable is known if the hypothesis H_0 is true. In this study the statistic variable is assigned as the residual vector or one step ahead prediction error vector of the model. This is defined from eqn (2.1) as

$$\begin{aligned}\varepsilon(t) &= y(t) - f(y^{t-1}, u^{t-1}, \varepsilon^{t-1}) \\ &= y(t) - \hat{y}(t)\end{aligned}$$

(2.5)

where $y(t)$ is the system output and $\hat{y}(t)$ is the model one step ahead predicted output. The hypothesis H_0 is then defined assuming the model is correct to give

$$\varepsilon(t) \equiv e(t)$$

(2.6)

where $e(t)$ is a totally random vector which is unpredictable from all linear and nonlinear combinations of past inputs and outputs.

The second step is to define a domain D^{α} such that prob $\{\varepsilon(t) \notin D^{\alpha} \mid H_0\}=\alpha$. In this study the domain is defined as

$$D^{\alpha} = \{\varepsilon | \phi < k_{\alpha}\}$$

(2.7)

where ϕ is the new test to be developed in the following sections and k_{α} is the decision value. The probability of incorrectly rejecting a correct model is α. Typically the 95% confidence limits $k_{\alpha}=k_{0.05}=\pm 1.96/\sqrt{N}$ (N is the data length) are used when ϕ is normally distributed.

The third step is to reject the hypothesis H_0 if $\varepsilon(t) \notin D^{\alpha}$. Different tests can be formulated by choosing different weight functions and different domains. From Bohlin (1978) all the tests will have the same risk α of rejecting a model when it is actually valid, however all the tests do not have the same probability of rejecting a model when it is not valid. Thus they are not equally efficient. In this study the system output is introduced as a new weight function to enhance the discriminatory performance.

3.0 Model validity tests

The model validity tests can be formulated as a statistical hypothesis testing problem based upon the three steps described in section 2.0.

3.1 The linear input/output model case

3.1.1 Correlation tests

Validity tests for the classical SISO linear model ($q=r=1$ $f(.)$ is a linear map) based on correlation functions can be summarized as

$$\phi_{\varepsilon_1\varepsilon_1}(\tau) = E[\varepsilon_1(t)\varepsilon_1(t+\tau)]$$
$$\phi_{u_1\varepsilon_1}(\tau) = E[u_1(t)\varepsilon_1(t+\tau)] \tag{3.1}$$

where $E[.]$ denotes the expectation operator. In practice normalized correlation functions are computed (Priestley 1981) from finite record length N as

$$\phi_{\varepsilon_1\varepsilon_1}(\tau) = \frac{\sum_{t=1}^{N}\varepsilon_1^o(t)\varepsilon_1^o(t+\tau)}{\sum_{t=1}^{N}(\varepsilon_1^o(t))^2} \qquad \phi_{u_1\varepsilon_1}(\tau) = \frac{\sum_{t=1}^{N}u_1^o(t)\varepsilon_1^o(t+\tau)}{\sqrt{\sum_{t=1}^{N}(u_1^o(t))^2\sum_{t=1}^{N}(\varepsilon_1^o(t))^2}} \tag{3.2}$$

where

$$\varepsilon_1^o(t) = \varepsilon_1(t) - \overline{\varepsilon_1} \qquad u_1^o(t) = u_1(t) - \overline{u_1}$$
$$\overline{\varepsilon_1} = \frac{1}{N}\sum_{t=1}^{N}\varepsilon_1(t) \qquad \overline{u_1} = \frac{1}{N}\sum_{t=1}^{N}u_1(t) \tag{3.3}$$

If the model is an adequate representation of the system then from eqn (2.6) $\varepsilon_1(t)$ should equal $e_1(t)$ and therefore ideally

$$\phi_{\varepsilon_1\varepsilon_1}(\tau) = \begin{cases} 1, & \tau = 0 \\ 0, & otherwise \end{cases}$$
$$\phi_{u_1\varepsilon_1}(\tau) = 0, \forall\tau \tag{3.4}$$

For large N the correlation function estimates given in eqn (3.2) are asymptotically normal with zero mean and finite variance (Box and Pierce 1970), the standard deviations are $1/\sqrt{N}$ and the 95% confidence limits are approximately $1.96/\sqrt{N}$.

New MIMO linear model validity test procedures can easily be developed from the above tests to form a global to local hierarchical diagnosis procedure. The global test checks for auto-correlations among all the submodel residuals and cross correlations among all the inputs and submodel residuals using

$$\phi_{\varsigma\varsigma}(\tau) = E[\varsigma(t)\varsigma(t+\tau)]$$
$$\phi_{\nu\varsigma}(\tau) = E[\nu(t)\varsigma(t+\tau)]$$

(3.5)

where

$$\varsigma(t) = \varepsilon_1(t) + \ldots + \varepsilon_q(t)$$
$$\nu(t) = u_1(t) + \ldots + u_r(t)$$

(3.6)

Under the null hypothesis H_0 that the MIMO model is valid ($\varepsilon(t)=e(t)$ from eqn (2.6)) then ideally eqn (3.5) becomes

$$\phi_{\varsigma\varsigma}(\tau) = \begin{cases} 1, \tau = 0 \\ 0, otherwise \end{cases}$$
$$\phi_{\nu\varsigma}(\tau) = 0, \forall\tau$$

(3.7)

These global tests can be localized to isolate inadequacies in individual model loops by computing the correlation matrices

$$\Phi_{\varepsilon\varepsilon}(\tau) = E[\varepsilon(t)\varepsilon^T(t+\tau)] = \begin{bmatrix} \phi_{\varepsilon_1\varepsilon_1}(\tau) & \ldots & \phi_{\varepsilon_1\varepsilon_q}(\tau) \\ \ldots & \ldots & \ldots \\ \phi_{\varepsilon_q\varepsilon_1}(\tau) & \ldots & \phi_{\varepsilon_q\varepsilon_q}(\tau) \end{bmatrix}$$

$$\Phi_{u\varepsilon}(\tau) = E[u(t)\varepsilon^T(t+\tau)] = \begin{bmatrix} \phi_{u_1\varepsilon_1}(\tau) & \ldots & \phi_{u_1\varepsilon_q}(\tau) \\ \ldots & \ldots & \ldots \\ \phi_{u_r\varepsilon_1}(\tau) & \ldots & \phi_{u_r\varepsilon_q}(\tau) \end{bmatrix}$$

(3.8)

where

$$\varepsilon(t) = [\varepsilon_1(t)\ldots\varepsilon_q(t)]^T$$
$$u(t) = [u_1(t)\ldots u_r(t)]^T$$

(3.9)

If the MIMO model is valid ($\varepsilon(t)=e(t)$ from eqn (2.6)) then ideally eqn (3.8) becomes

$$\Phi_{\varepsilon\varepsilon}(\tau) = \begin{cases} I_{q\times q} & ,\tau = 0 \\ 0_{q\times q} & ,otherwise \end{cases}$$
$$\Phi_{u\varepsilon}(\tau) = 0_{r\times q}, \forall\tau$$

(3.10)

where $I_{q\times q}$ is an identical matrix, $0_{q\times q}$ and $0_{r\times q}$ are zero matrices. The total number of model validity tests for a q output r input MIMO model is $q*r+q*q+2$. This can be reduced to $q*r+q*(1+q)/2+2$ because of the symmetry of the auto-correlation function matrix $\Phi_{\varepsilon\varepsilon}(\tau)$.

3.1.2 Chi-squared tests

An alternative to the correlation based approach Chi-squared (χ^2) tests can also be derived to yield a global test. Define

$$d_\varsigma = N\mu_\varsigma^T(\Gamma_\varsigma^T\Gamma_\varsigma)^{-1}\mu_\varsigma \qquad d_\nu = N\mu_\nu^T(\Gamma_\nu^T\Gamma_\nu)^{-1}\mu_\nu$$

(3.11)

where

$$\mu_\varsigma = \frac{1}{N}\sum_{t=1}^{N}\grave{\varsigma}(t)\,w_\varsigma(t)$$
$$\grave{\varsigma}(t) = [\varsigma(t-1),\varsigma(t-2)\ldots\varsigma(t-s)]^T$$
$$w_\varsigma(t) = \frac{\varsigma(t)-\bar{\varsigma}}{\sqrt{\sum_{t=1}^{N}(\varsigma(t)-\bar{\varsigma})^2}}$$
$$\frac{1}{N}\Gamma_\varsigma^T\Gamma_\varsigma = E[\mu_\varsigma\mu_\varsigma^T]$$

$$\mu_\nu = \frac{1}{N}\sum_{t=1}^{N} \hat{\varsigma}(t)\, w_\nu(t)$$

$$w_\nu(t) = \frac{\nu(t) - \bar{\nu}}{\sqrt{\sum_{t=1}^{N}(\nu(t) - \bar{\nu})^2}}$$

$$\frac{1}{N}\Gamma_\nu^T\Gamma_\nu = E[\mu_\nu\mu_\nu^T]$$

(3.12)

The tests in eqn (3.11) are asymptotically χ^2 distributed with s degrees of freedom where s is the dimension of vector $\hat{\varsigma}(t)$. The confidence bands of the tests are given by

$$d_\varsigma = k_{\gamma_\varsigma}(s) \qquad d_\nu = k_{\gamma_\nu}(s)$$

(3.13)

where $k_{\gamma_\varsigma}(s)$ and $k_{\gamma_\nu}(s)$ are the critical values of the χ^2 distribution with s degrees of freedom γ_ς and γ_ν are the significance levels for model acceptance regions. The local χ^2 test formulation has been developed in many publications Bohlin (1971).

3.2 The nonlinear input/output model case

3.2.1 Correlation tests

It is well known that linear model validity test methods based on eqn (3.1) sometimes fail to diagnose missing nonlinear model terms (Billings and Voon 1983) and new tests have to be considered to avoid this deficiency. These can either be higher order extensions of the tests in eqn (3.1) (Billings and Voon 1986) which are based on the input and residual vectors only or they can be extended to exploit the information in the system outputs as well (Billings and Zhu 1994). New global tests which check for correlations among all the sub-model input, output and residual vectors can be defined as

$$\phi_{\xi\eta}(\tau) = E[\xi(t)\eta(t+\tau)]$$

$$\phi_{\vartheta\eta}(\tau) = E[\vartheta(t)\eta(t+\tau)]$$

(3.14)

where $\xi(t)$, $\eta(t)$ and $\vartheta(t)$ are normalized variables given by

$$\xi(t) = \varepsilon_1^2(t) + \ldots + \varepsilon_q^2(t)$$
$$\eta(t) = y_1(t)\varepsilon_1(t) + \ldots + y_q(t)\varepsilon_q(t)$$
$$\vartheta(t) = u_1^2(t) + \ldots + u_r^2(t)$$

(3.15)

Ideally if every subsystem in the MIMO model is valid ($\varepsilon(t)=e(t)$ from eqn (2.6)) eqn (3.14) becomes

$$\phi_{\xi\eta}(\tau) = \begin{cases} k, \tau = 0 \\ 0, otherwise \end{cases}$$
$$\phi_{\vartheta\eta}(\tau) = 0, \forall\tau$$

(3.16)

where $k = \dfrac{\sqrt{\sum_{t=1}^{N} (\xi^o(t))^2}}{\sqrt{\sum_{t=1}^{N} (\eta^o(t))^2}}$ is a constant (Billings and Zhu 1994).

Notice that the output has been introduced to enhance the discriminatory performance compared to tests which traditionally have been based on only the inputs and residuals.

Localized tests can then be used to check the correlations between submodel residuals and outputs, submodel residuals, outputs and inputs to yield

$$\Phi_{\varepsilon^2\eta}(\tau) = E[\varepsilon^2(t)\eta^T(t+\tau)] = \begin{bmatrix} \phi_{\varepsilon_1^2\eta_1}(\tau) & \ldots & \phi_{\varepsilon_1^2\eta_q}(\tau) \\ \ldots & \ldots & \ldots \\ \phi_{\varepsilon_q^2\eta_1}(\tau) & \ldots & \phi_{\varepsilon_q^2\eta_q}(\tau) \end{bmatrix}$$

$$\Phi_{u^2\eta}(\tau) = E[u^2(t)\eta^T(t+\tau)] = \begin{bmatrix} \phi_{u_1^2\eta_1}(\tau) & \ldots & \phi_{u_1^2\eta_q}(\tau) \\ \ldots & \ldots & \ldots \\ \phi_{u_r^2\eta_1}(\tau) & \ldots & \phi_{u_r^2\eta_q}(\tau) \end{bmatrix}$$

(3.17)

where

$$\varepsilon^2(t) = [\varepsilon_1^2(t)\ldots\varepsilon_q^2(t)]^T$$
$$\eta(t) = [[y_1(t)\varepsilon_1(t)]\ldots[y_q(t)\varepsilon_q(t)]]^T$$
$$u^2(t) = [u_1^2(t)\ldots u_r^2(t)]^T$$

(3.18)

are squared residual, residual and output product, and squared input vectors respectively. Under the null hypothesis H_0 that the MIMO model is valid ($\varepsilon(t)=e(t)$ from eqn (2.6)) eqn (3.17) becomes

$$\Phi_{\varepsilon^2\eta}(\tau) = \begin{cases} k_{q\times q} & ,\tau = 0 \\ 0_{q\times q} & ,otherwise \end{cases}$$
$$\Phi_{u^2\eta}(\tau) = 0_{r\times q}, \forall\tau$$

(3.19)

where $0_{q\times q}$ and $0_{r\times q}$ are zero matrices and $k_{q\times q}$ is a diagonal matrix with constant

elements $k_{ii} = \dfrac{\sqrt{\sum_{t=1}^{N}(\varepsilon_i^{2^o}(t))^2}}{\sqrt{\sum_{t=1}^{N}(\eta_i^o(t))^2}}$. The total number of nonlinear model tests for

a q output r input MIMO model is $q*r+q*q+2$. The global tests in eqn (3.14) can therefore be used initially to indicate if the model is valid. Local test diagnoses can then be performed as required to determine which submodels are incorrect.

3.2.2 Chi-squared tests

Similarly an associated χ^2 test can also be derived based on the global tests

$$d_\xi = N\mu_\xi^T(\Gamma_\xi^T\Gamma_\xi)^{-1}\mu_\xi \qquad d_\vartheta = N\mu_\vartheta^T(\Gamma_\vartheta^T\Gamma_\vartheta)^{-1}\mu_\vartheta$$

(3.20)

where

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

$$\vec{\eta}(t) = [\eta(t-1), \eta(t-2) \ldots \eta(t-s)]^T$$

$$w_\xi(t) = \frac{\xi(t) - \bar{\xi}}{\sqrt{\sum_{t=1}^{N} (\xi(t) - \bar{\xi})^2}}$$

$$\frac{1}{N}\Gamma_\xi^T \Gamma_\xi = E[\mu_\xi \mu_\xi^T]$$

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

$$w_\vartheta(t) = \frac{\vartheta(t) - \bar{\vartheta}}{\sqrt{\sum_{t=1}^{N} (\vartheta(t) - \bar{\vartheta})^2}}$$

$$\frac{1}{N}\Gamma_\vartheta^T \Gamma_\vartheta = E[\mu_\vartheta \mu_\vartheta^T]$$

(3.21)

The tests in eqn (3.20) are asymptotically χ^2 distributed with s degrees of freedom where s is the dimension of vector $\vec{\eta}(t)$. The confidence band limit of the tests are given by

$$d_\xi = k_{\gamma_\xi}(s) \qquad d_\vartheta = k_{\gamma_\vartheta}(s)$$

(3.22)

where $k_{\gamma_\xi}(s)$ and $k_{\gamma_\vartheta}(s)$ are the critical values of the χ^2 distribution with s degrees of freedom γ_ξ and γ_ϑ are the significance levels for model acceptance regions. The local χ^2 test has been derived by Billings and Zhu (1994).

3.3 The nonlinear time series case

The nonlinear time series model is a subset of eqn (2.1) which is defined when all input signals are excluded. Validity tests for time series models are therefore just special cases of the tests above.

3.4 Test procedure

In summary a coarse to fine strategy can be implemented to test an identified model as follows

i) Apply the global tests $\phi_{\xi\eta}(\tau)$ and $\phi_{\vartheta\eta}(\tau)$ to check if the model valid.

Although $\phi_{\xi\eta}(\tau)$ and $\phi_{\vartheta\eta}(\tau)$ should detect all possible missing model terms it is often informative to apply the global linear tests $\phi_{\varsigma\varsigma}(\tau)$ and $\phi_{v\varsigma}(\tau)$ as well.

Stop if all the tests are satisfied otherwise go to the next step for local diagnosis of all submodels.

ii) Apply the local tests $\Phi_{\varepsilon^2\eta}(\tau)$ and $\Phi_{u^2\eta}(\tau)$ and determine which of the submodels are incorrect.

The localized linear tests $\Phi_{\varepsilon\varepsilon}(\tau)$ and $\Phi_{u\varepsilon}(\tau)$ can be computed if required.

3.5 Worked examples

To simply demonstrate how the tests perform two straightforward examples will be considered where the tests can be easily evaluated analytically.

Example W_1

To illustrate the validity test procedure consider a very simple MIMO ARMAX (AutoRegressive Moving Average with eXogenous input) model

$$\begin{aligned} y_1(t) &= u_1(t-1) + \varepsilon_1(t) \\ y_2(t) &= u_1(t-1) + \varepsilon_2(t) \end{aligned} \tag{3.23}$$

where $u_1(t)$ is an independent excitation sequence with zero mean and finite variance. Assume that the system has been incorrectly modeled or identified so the residuals are given by

$$\begin{aligned} \varepsilon_1(t) &= e_1(t) \\ \varepsilon_2(t) &= e_1(t-1) + e_2(t) \end{aligned} \tag{3.24}$$

where the noise sequences $e_1(t)$ and $e_2(t)$ are independent with zero mean and finite variance.

Note that all the above tests have been derived without making any assumptions on the form of the input. Most of the examples however assume the input is an independent sequence simply because this can cause offending model terms to average out to zero in the model validity tests and therefore represents a worst case example. The global test $\phi_{\varsigma\varsigma}(\tau)$ can be used to check for delayed noise terms like $e_1(t-j)$ in $\varepsilon(t)$. Similarly $\phi_{v\varsigma}(\tau)$

can be used to check for $u_1(t-j)$ terms in $\varepsilon(t)$. When the residuals include delayed outputs like $y_1(t-j)$ both $\phi_{\varsigma\varsigma}(\tau)$ and $\phi_{v\varsigma}(\tau)$ will give an indication that $\varepsilon(t)$ is correlated because $y_1(t-j)$ is auto-correlated and cross-correlated with the input. The local tests can be further used to determine which of the submodels are at fault.

Using simple algebraic operations the MIMO linear model global test of eqn (3.5) gives

$$\phi_{\varsigma\varsigma}(\tau) = \begin{cases} 1, \tau = 0 \\ \rho_1, \tau = 1 \\ 0, otherwise \end{cases}$$
$$\phi_{v\varsigma}(\tau) = 0, \forall\tau$$

(3.25)

where for this example ρ_1 is a constant. The value of this and all other constant values in the examples of this section are given in appendix A. Because $\phi_{\varsigma\varsigma}(\tau)$ is not zero for all $\tau \neq 0$ this indicates that the elements in the residual vectors are not all uncorrelated and the model is deficient in some way. This deficiency can be diagnosed by considering the submodel tests of eqn (3.8) which yield

$$\phi_{\varepsilon_1\varepsilon_1}(\tau) = \begin{cases} 1, \tau = 0 \\ 0, otherwise \end{cases} \qquad \phi_{\varepsilon_1\varepsilon_2}(\tau) = \begin{cases} \rho_2, \tau = 1 \\ 0, otherwise \end{cases}$$
$$\phi_{\varepsilon_2\varepsilon_2}(\tau) = \begin{cases} 1, \tau = 0 \\ 0, otherwise \end{cases}$$

$$\phi_{u_1\varepsilon_1}(\tau) = 0, \forall\tau \qquad \phi_{u_1\varepsilon_2}(\tau) = 0, \forall\tau$$

(3.26)

where for this example ρ_2 is a constant. The test $\phi_{\varepsilon_1\varepsilon_2}(1) = \rho_2$ suggests a delayed noise $\varepsilon_1(t-1)$ in the residual $\varepsilon_2(t)$.

To demonstrate that the nonlinear model validity tests can be applied to the linear model case consider the application of the tests in eqns (3.14) and (3.17) to the model eqn (3.23). The global test yields

$$\phi_{\xi\eta}(\tau) = \begin{cases} k, \tau = 0 \\ \rho_3, \tau = 1 \\ 0, otherwise \end{cases}$$
$$\phi_{\vartheta\eta}(\tau) = 0, \forall\tau$$

(3.27)

where for this example ρ_3 is a constant. Comparison with eqn (3.16) shows that the elements in the residual vector are not all uncorrelated. Diagnosing the submodels gives

$$\phi_{\varepsilon_1^2\eta_1}(\tau) = \begin{cases} k_{11}, \tau = 0 \\ 0, otherwise \end{cases} \qquad \phi_{\varepsilon_1^2\eta_2}(\tau) = \begin{cases} \rho_4, \tau = 1 \\ 0, otherwise \end{cases}$$

$$\phi_{\varepsilon_2^2\eta_1}(\tau) = 0, \forall\tau \qquad \phi_{\varepsilon_2^2\eta_2}(\tau) = \begin{cases} k_{22}, \tau = 0 \\ 0, otherwise \end{cases}$$

$$\phi_{u_1^2\eta_1}(\tau) = 0, \forall\tau \qquad \phi_{u_1^2\eta_2}(\tau) = 0, \forall\tau$$

(3.28)

where for this example ρ_4 is a constant (given in Appendix A). The test $\phi_{\varepsilon_1\eta_2}(\tau)$ suggests a delayed noise $e_1(t-1)$ in the residual $\varepsilon_2(t)$ because of $\phi_{\varepsilon_1\eta_2}(1) = \rho_4$.

Example W_2

Consider as a second example a MIMO NARMAX model

$$y_1(t) = u_1(t-1) + \varepsilon_1(t)$$
$$y_2(t) = u_1(t-1) + \varepsilon_2(t)$$

(3.29)

Assume a linear model identification algorithm has been used to identify the model and that nonlinear terms have been omitted to leave the residual terms

$$\varepsilon_1(t) = e_1(t-1)\,e_2(t-2) + e_1(t)$$
$$\varepsilon_2(t) = e_2(t)$$

(3.30)

where the input $u_1(t)$ is an independent excitation sequence with zero mean and finite variance, and the noise sequences $e_1(t)$ and $e_2(t)$ are independent with zero mean and finite variance. Applying the global MIMO linear model tests of eqn (3.5) to the residuals in eqn (3.29) gives

$$\phi_{\varsigma\varsigma}(\tau) = \begin{cases} 1, \tau = 0 \\ 0, otherwise \end{cases}$$
$$\phi_{v\varsigma}(\tau) = 0, \forall\tau$$

(3.31)

which shows that the linear tests fail to detect the omitted nonlinear terms embedded in the residuals.

Applying the MIMO nonlinear model global tests of eqn (3.14) to the residuals in eqn (3.29) gives

$$\phi_{\xi\eta}(\tau) = \begin{cases} k, \tau = 0 \\ \rho_5, \tau = 1 \\ \rho_6, \tau = 2 \\ 0, otherwise \end{cases}$$

$$\phi_{\vartheta\eta}(\tau) = 0, \forall\tau \tag{3.32}$$

where the constants ρ_5 and ρ_6 are given in Appendix A. Inspection of eqn (3.26) indicates that the residual vector is not uncorrelated because $\phi_{\xi\eta}(1) = \rho_5$ and $\phi_{\xi\eta}(2) = \rho_6$. Using the local tests to diagnose which submodels are causing the problem yields

$$\phi_{\varepsilon_1^2\eta_1}(\tau) = \begin{cases} k_{11}, \tau = 0 \\ \rho_7, \tau = 1 \\ 0, otherwise \end{cases} \qquad \phi_{\varepsilon_1^2\eta_2}(\tau) = 0, \forall\tau$$

$$\phi_{\varepsilon_2^2\eta_1}(\tau) = \begin{cases} \rho_8, \tau = 2 \\ 0, otherwise \end{cases} \qquad \phi_{\varepsilon_2^2\eta_2}(\tau) = \begin{cases} k_{22}, \tau = 0 \\ 0, otherwise \end{cases}$$

$$\phi_{u_1^2\eta_1}(\tau) = 0, \forall\tau \qquad \phi_{u_1^2\eta_2}(\tau) = 0, \forall\tau \tag{3.33}$$

where ρ_7 and ρ_8 are constants given in Appendix A. Eqn (3.33) indicates that the residual $\varepsilon_1(t)$ includes some delayed noise terms at τ=1, 2 because $\phi_{\varepsilon_1^2\eta_1}(1) = \rho_7$ and $\phi_{\varepsilon_2^2\eta_1}(2) = \rho_8$.

4.0 Simulated examples

Four simulated systems were selected to demonstrate the new MIMO nonlinear model validity test procedures. In each case the data sequences were of length 1000. The MIMO nonlinear model orthogonal identification algorithm (Billings, Chen and Korenberg 1989) which includes model term selection and associated parameter estimation was used in the study of examples one, two and four. The hybrid radial basis function identification algorithm (Chen, Billings and Grant 1992) which includes network structure selection and training was used in example three.

Example S_1

A simulated two input output nonlinear system was excited by two uniformly distributed uncorrelated input sequences (u_1 , u_2) with zero mean and variance 1.33. The noise sequences (e_1, e_2) were normally distributed uncorrelated sequences with zero mean and variance 0.36. The system model was correctly identified as

$$y_1(t) = 1.002u_1(t-1)u_2(t-1) + 1.043e_1(t-1)e_1(t-2) + e_1(t)$$
$$y_2(t) = 1.000u_2^2(t-1) + e_2(t) \tag{4.1}$$

and all the model validity tests were satisfied. The identified model was then deliberately constrained to be the incorrect deterministic model given by

$$y_1(t) = 1.002u_1(t-1)u_2(t-1) + \varepsilon_1(t)$$
$$y_2(t) = 1.000u_2^2(t-1) + \varepsilon_2(t) \tag{4.2}$$

with the residuals

$$\varepsilon_1(t) = 1.043e_1(t-1)e_1(t-2) + e_1(t)$$
$$\varepsilon_2(t) = e_2(t) \tag{4.3}$$

The corresponding model validity tests are shown in Figure 1. The global test $\phi_{\xi\eta}(\tau)$ fails since it is outside the 95% confidence limits at point τ=1. This suggests the possibility of some delayed noise terms included in the residuals $\varepsilon_1(t)$ and $\varepsilon_2(t)$. The valid global test $\phi_{\vartheta\eta}(\tau)$ is well within the 95% bands and indicates that there are no delayed inputs in the residuals. With a series of local tests the correlated nonlinear noise term $e_1(t\text{-}1)e_1(t\text{-}2)$ in $\varepsilon_1(t)$ was detected using $\phi_{\varepsilon_1^2(y_1\varepsilon_1)}(\tau)$ which is obviously outside the 95% confidence limits at the points τ=1, 2.

Example S_2

Consider the MIMO nonlinear time series model

$$y_1(t) = 0.428y_1(t-1) + 0.210y_1(t-2) - 0.489y_2(t-1)\varepsilon_1(t-2) + \varepsilon_1(t)$$
$$y_2(t) = 0.690y_2(t-1) - 0.295y_2(t-2) + 0.953\varepsilon_1(t-1)\varepsilon_1(t-2) + \varepsilon_2(t) \tag{4.4}$$

where the noise sequence (e_1, e_2) were normally distributed uncorrelated with zero mean and variance 1.0. When the term $\varepsilon_1(t-1)\varepsilon_1(t-2)$ was deliberately excluded from the model of eqn (4.4), estimation produced the identified model

$$y_1(t) = 0.428y_1(t-1) + 0.210y_1(t-2) - 0.489y_2(t-1)\varepsilon_1(t-2) + \varepsilon_1(t)$$
$$y_2(t) = 0.727y_2(t-1) - 0.333y_2(t-2) + \varepsilon_2(t) \tag{4.5}$$

The model validity tests are shown in Figure 2, the global test $\phi_{\xi\eta}(\tau)$ shows that the model is incorrect because of the value outside the 95% confidence limits at point τ=1. The missing term $\varepsilon_1(t\text{-}1)\varepsilon_1(t\text{-}2)$ was detected by the local test $\phi_{\varepsilon_1^2(y_2\varepsilon_2)}(\tau)$. Notice that the local test $\phi_{\varepsilon_2^2(y_1\varepsilon_1)}(\tau)$ is outside the confidence limits at points τ=-1, -2 because of the effect of coupled residuals and does not indicate omitted noncausal terms.

Example S_3

A hybrid radial basis function (rbf) neural network structure was used to identify system S_2 above.

The rbf network consisted of just one hidden layer, the input vector was defined as $[y_1(t-1) \quad y_1(t-2) \quad y_2(t-1) \quad y_2(t-2)]^T$, 60 centres using the thin-plate-spline function $v^2 * log(v)$ were used with two output nodes $y_1(t)$ and $y_2(t)$. The network was trained using the recursive hybrid algorithm (Chen, Billings and Grant 1992) with the following settings:

Initial covariance: 1000

Initial forgetting factor: 0.99

Rate factor: 0.98

Initial clustering gain: 0.6

Number of passes for training: 5

Initial hidden layer nodes were chosen from data.

The resultant total mean squared residual was 0.757 and all the model validity tests were satisfied indicating the network had been correctly trained to represent the system. When the node $y_1(t\text{-}2)$ was deliberately omitted from the network the model validity tests, illustrated in Figure 3, clearly showed that the trained network was an inadequate model of the system and the total mean squared residual increased to 0.890. In Figure 3 the global test $\phi_{\xi\eta}(\tau)$ shows that the model is incorrect because there are values outside the 95% confidence limits at points τ=1, 2 . The missing term $y_1(t\text{-}2)$ was detected by local test $\phi_{\varepsilon_2^2(y_1\varepsilon_1)}(\tau)$ which exhibits a value outside the confidence limit at τ=2. The coupling effect caused by the missing term was also detected by local tests $\phi_{\varepsilon_1^2(y_2\varepsilon_2)}(\tau)$ and $\phi_{\varepsilon_1^2(y_2\varepsilon_2)}(\tau)$ with values outside the confidence limits at τ=-1 and 1 respectively.

Example S_4

Consider the nonlinear model studied by Billings, Chen and Korenberg (1989)

$$\begin{aligned} y_1(t) &= 0.5y_1(t-1) + u_1(t-2) + 0.1y_2(t-1)u_1(t-1) \\ &\quad + 0.5e_1(t-1) + 0.2y_1(t-2)e_1(t-2) + e_1(t) \\ y_2(t) &= 0.9y_2(t-2) + u_2(t-1) + 0.2y_2(t-1)u_2(t-2) \\ &\quad + 0.5e_2(t-1) + 0.1y_2(t-1)e_1(t-2) + e_2(t) \end{aligned} \tag{4.6}$$

This system was simulated with the noise $e(t)=[e_1(t) \;\; e_2(t)]^T$ defined as an uncorrelated sequence with zero mean and covariance

$$COV[e(t)] = \begin{bmatrix} 0.04 & 0.0 \\ 0.0 & 0.04 \end{bmatrix} \tag{4.7}$$

$u_1(t)$ was a normally distributed uncorrelated sequence with zero mean and variance 1.0 and $u_2(t)$ was an uncorrelated sequence with a uniform distribution with zero mean and variance 1.0.

The parameter estimates obtained by correctly including all the terms in eqn (4.6) in the identified model are given in Table 1 and Figure 4 shows the global model validity tests because all the local tests were inside the confidence bands. All the results show that an acceptable model has been obtained. If any of the terms in the identified model given in Table 1 is omitted from the model the validity tests will display values outside the confidence limits. For example Figures 5 and 6 show the global model validity tests with the terms $y_2(t-1)u_1(t-1)$ in submodel one and $u_2(t-1)$ in submodel two deliberately excluded. In each case the model validity tests clearly indicate that the system was inadequately modelled.

Comparing the computational complexity with that presented in Billings, Chen and Korenberg (1989) the new procedure only requires $q*q+q*r+2$ tests compared to the previous $q*r+q*(1+q)/2+q*r*(1+q)/2+q*r*(1+r)/2+q*r*(1+q)*(1+r)/4$ tests. For example for a three input output MIMO nonlinear model the new method only requires 2 tests which can be expanded to 20 local tests to isolate the problem if required whereas the previous method required 87 tests.

Sub-system 1	Parameter estimates
$u_1(t-2)$	0.997
$y_1(t-1)$	0.499
$y_2(t-1)u_1(t-1)$	0.103
$\varepsilon_1(t-1)$	0.534
$y_1(t-2)\varepsilon_1(t-2)$	0.220
Sub-system 2	
$y_2(t-2)$	0.901
$u_2(t-1)$	1.000
$y_2(t-1)u_2(t-2)$	0.200
$\varepsilon_2(t-1)$	0.419
$y_2(t-1)\varepsilon_1(t-2)$	0.097

Table 1 Identified model from eqn (4.6)

5.0 Conclusions

New model validation procedures have been introduced for general multi-input multi-output nonlinear models. The tests can be implemented in a global to local diagnosis procedure and computed using the same number of tests as in case of MIMO linear models. Several different system representations including MIMO NARMAX, time series and neural network models have been tested to demonstrate the wide applicability of the new procedures.

6.0 Appendix A

The constant ρ_i, i=1, ..., 8 presented in section 3.3 are valued as below,

$$\rho_1 = \frac{\overline{e_1^2}}{2\overline{e_1^2} + \overline{e_2^2}} \qquad \rho_2 = \sqrt{\frac{\overline{e_1^2}}{\overline{e_1^2} + \overline{e_2^2}}}$$

$$\rho_3 = \frac{\overline{e_1^4} - (\overline{e_1^2})^2}{2\overline{e_1^4} - 4(\overline{e_1^2})^2 + \overline{e_2^4} - (\overline{e_2^2})^2 + 2\overline{e_1^2}\;\overline{e_2^2}}$$

$$\rho_4 = \sqrt{\frac{\overline{e_1^4} - (\overline{e_1^2})^2}{\overline{e_1^4} - (\overline{e_1^2})^2 + \overline{e_2^4} - (\overline{e_2^2})^2}}$$

$$\rho_5 = \frac{\overline{e_1^4}\;\overline{e_2^2} - (\overline{e_1^2})^2\;\overline{e_2^2}}{\overline{e_1^4}\;\overline{e_2^4} - (\overline{e_1^2})^2(\overline{e_2^2})^2 + 4(\overline{e_1^2})^2\overline{e_2^4} + \overline{e_1^4} - (\overline{e_1^2})^2 + \overline{e_2^4} - (\overline{e_2^2})^2}$$

$$\rho_6 = \frac{\overline{e_1^2}\;\overline{e_2^4} - \overline{e_1^2}(\overline{e_2^2})^2}{\overline{e_1^4}\;\overline{e_2^4} - (\overline{e_1^2})^2(\overline{e_2^2})^2 + 4(\overline{e_1^2})^2\overline{e_2^4} + \overline{e_1^4} - (\overline{e_1^2})^2 + \overline{e_2^4} - (\overline{e_2^2})^2}$$

$$\rho_7 = \frac{\overline{e_1^4}\;\overline{e_2^2} - (\overline{e_1^2})^2\;\overline{e_2^2}}{\overline{e_1^4}\;\overline{e_2^4} - (\overline{e_1^2})^2(\overline{e_2^2})^2 + 4(\overline{e_1^2})^2\overline{e_2^2} + \overline{e_1^4} - (\overline{e_1^2})^2}$$

$$\rho_8 = \frac{\overline{e_1^2}\;\overline{e_2^4} - \overline{e_1^2}\,(\overline{e_2^2})^2}{\sqrt{\overline{e_2^4} - (\overline{e_2^2})^2}\,\sqrt{\overline{e_1^4}\;\overline{e_2^4} - (\overline{e_1^2})^2\,(\overline{e_2^2})^2 + 4\,(\overline{e_1^2})^2\,\overline{e_2^2} + \overline{e_1^4} - (\overline{e_1^2})^2}}$$

Acknowledgment

The authors gratefully acknowledge that this work is supported by SERC under grant GR/H3528.6. We also wish to thank Dr G. L. Zheng for his help in running part of the hybrid radial basis function simulation.

References

Billings, S. A., Chen, S., 1989, Extended model set, global data and threshold model identification of severely nonlinear systems. *Int. J. Control*, **50**, 1897-1923.

Billings, S. A., Chen, S. and Korenberg, M. J., 1989, Identification of MIMO nonlinear systems using a forward regression orthogonal estimator. *Int. J. Control*, **49**, 2157-2189.

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., 1994, Nonlinear model validation using correlation tests. (to appear in *Int. J. Control*).

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.

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

Box, G. E. P. and Pierce, D. A., 1970, Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *J. of American Statistical Society*, **65**, 1509-1526.

Chen, S., Billings, S. A. and Grant, P. W., 1992, Recursive hybrid algorithm for nonlinear system identification using radial basis function network. *Int. J. Control*, **55**, 1051-1070.

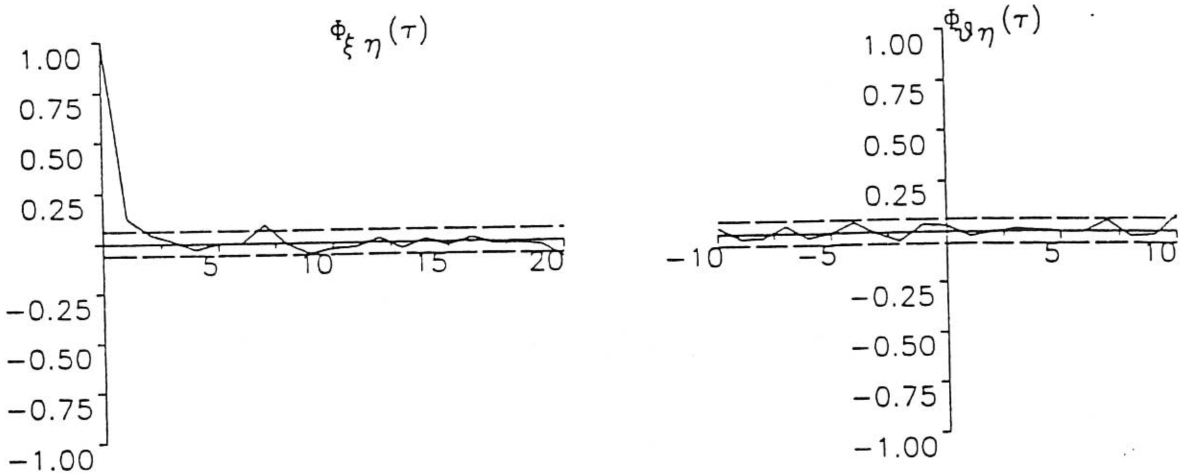
Kendall, M. G. and Stuart, A., 1967, *The advanced theory of statistics*, Vol. 2 (London: Griffin).

Kumar, K., 1986, On the identification of some bilinear time series models. *Journal of Time Series Analysis*, **7**, 117-122.

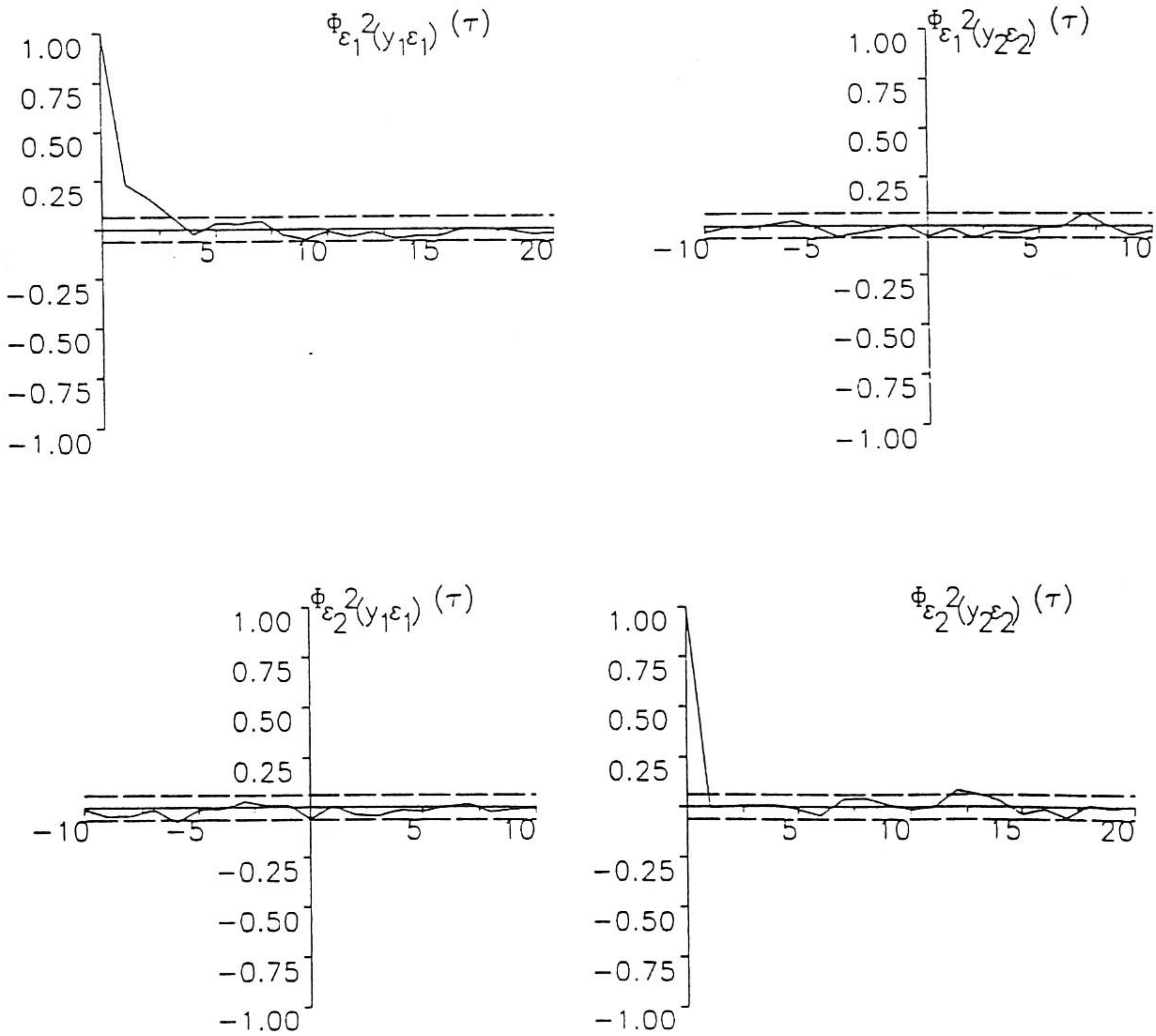
Priestley, M. B., 1981, *Spectral analysis and time series*, (New York: Academic Press).

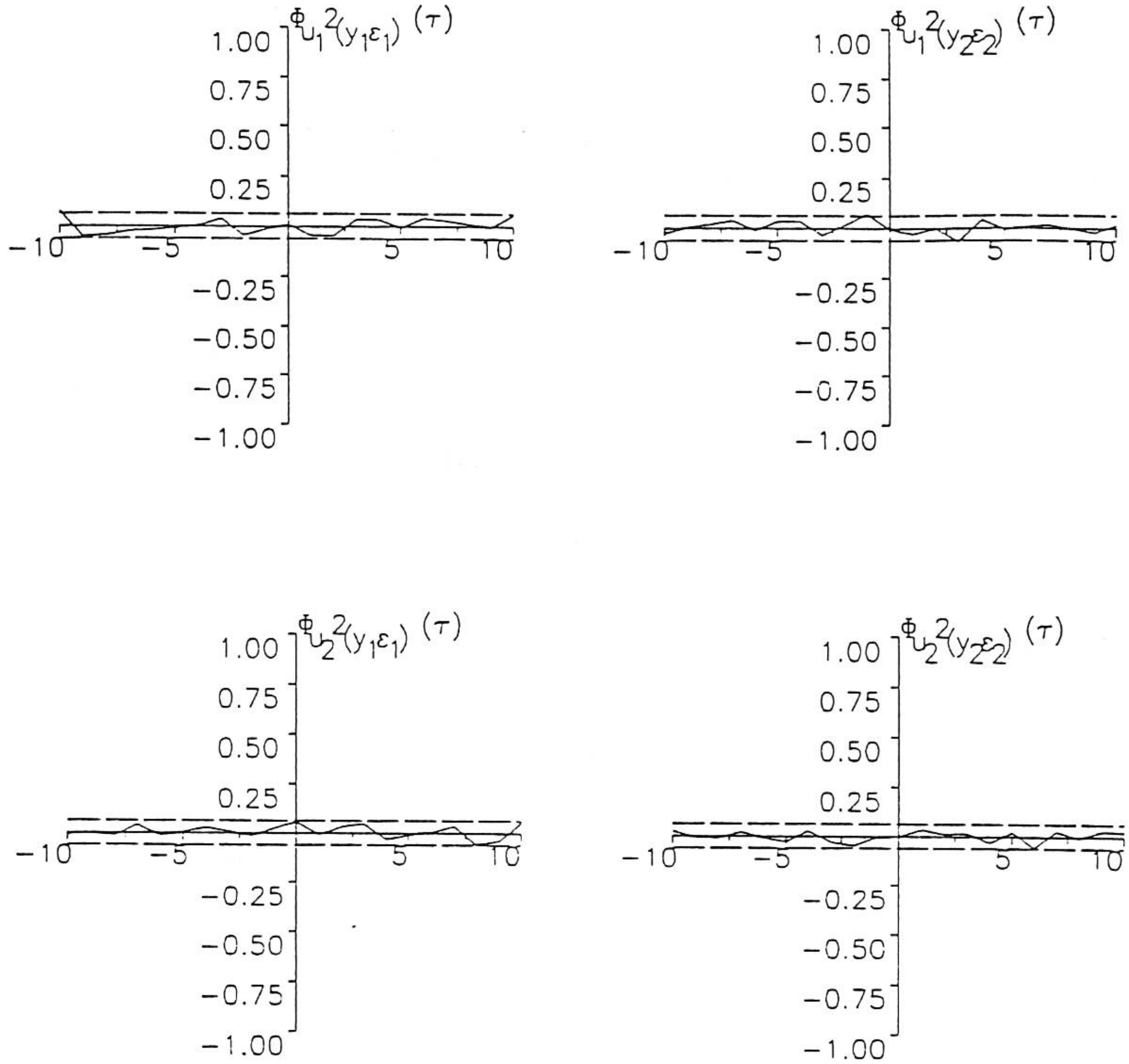
Subba Rao, T. and Gabr, M. M., 1984, *An introduction to bispectral analysis and bilinear time series models*, Lecture notes in statistics (New York: Springer-Verlag).

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



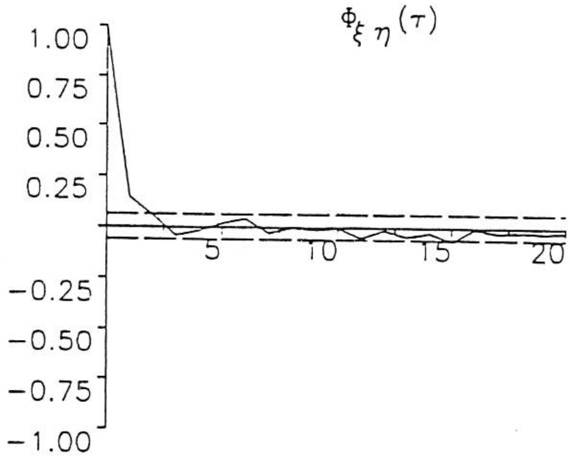
(a) Nonlinear global tests



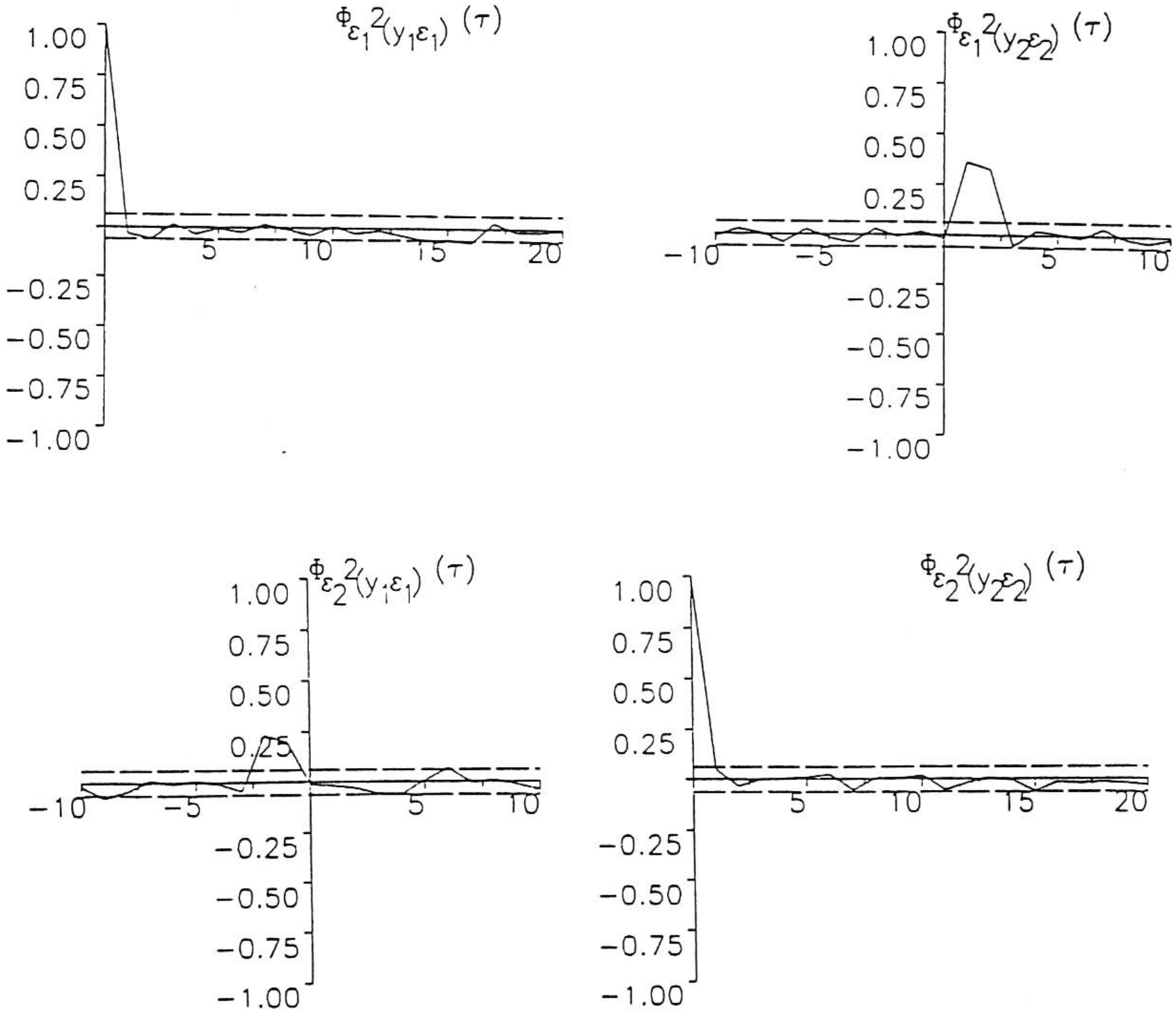


(b) Nonlinear local tests

Figure 1 Model validity tests for example S_1

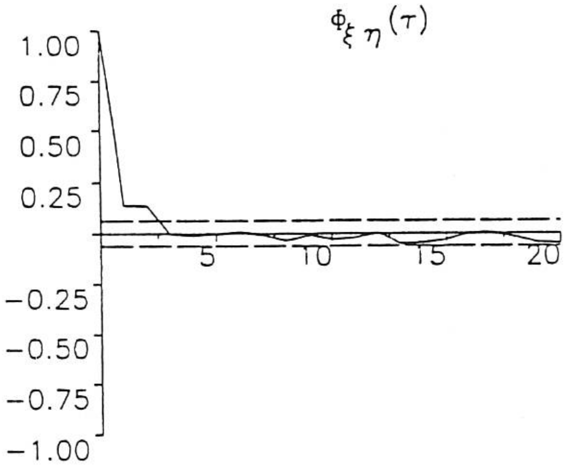


(a) Nonlinear global test

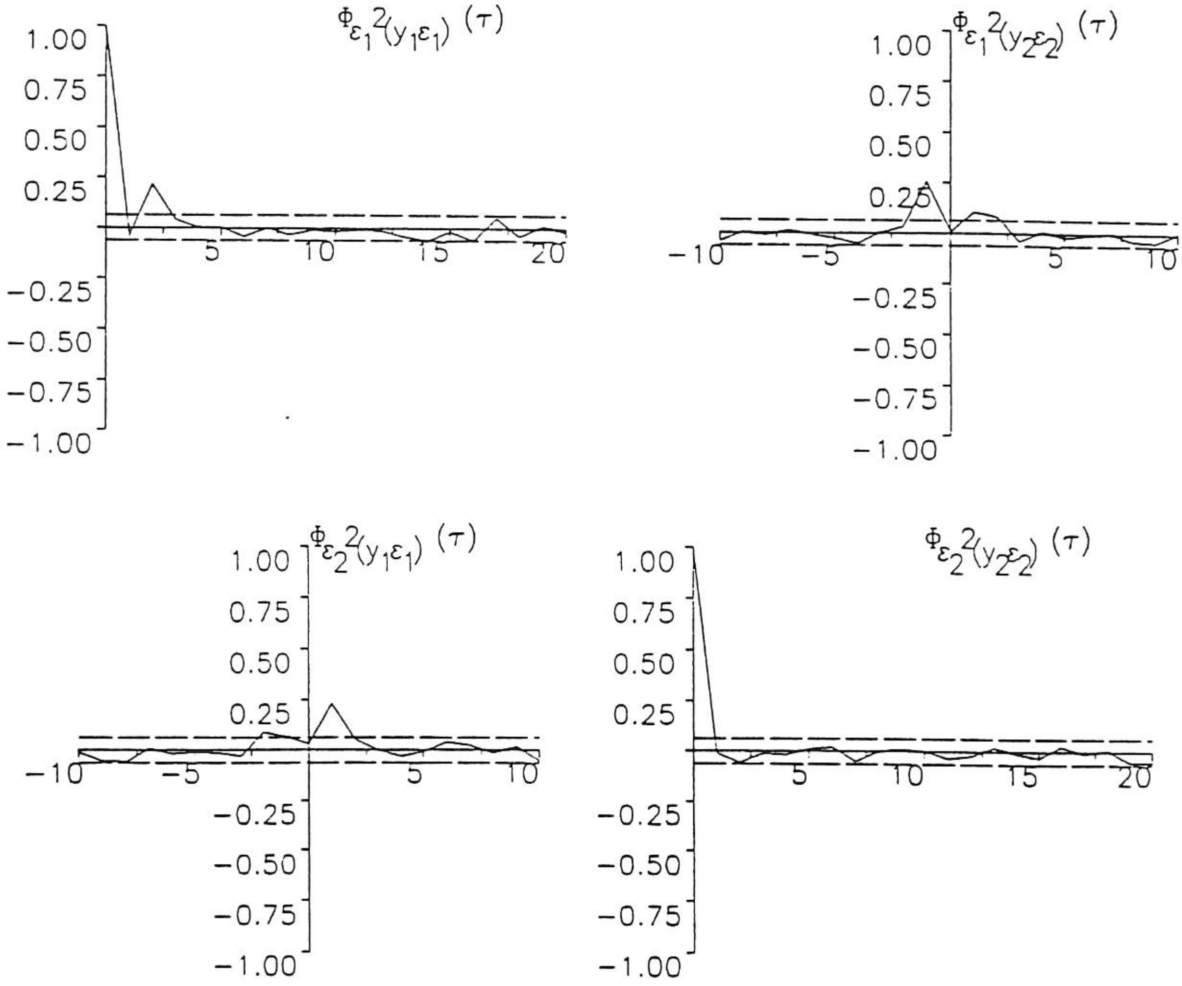


(b) Nonlinear local tests

Figure 2 Model validity tests for example S_2



(a) Nonlinear global test



(b) Nonlinear local tests

Figure 3 Model validity tests for example S_3

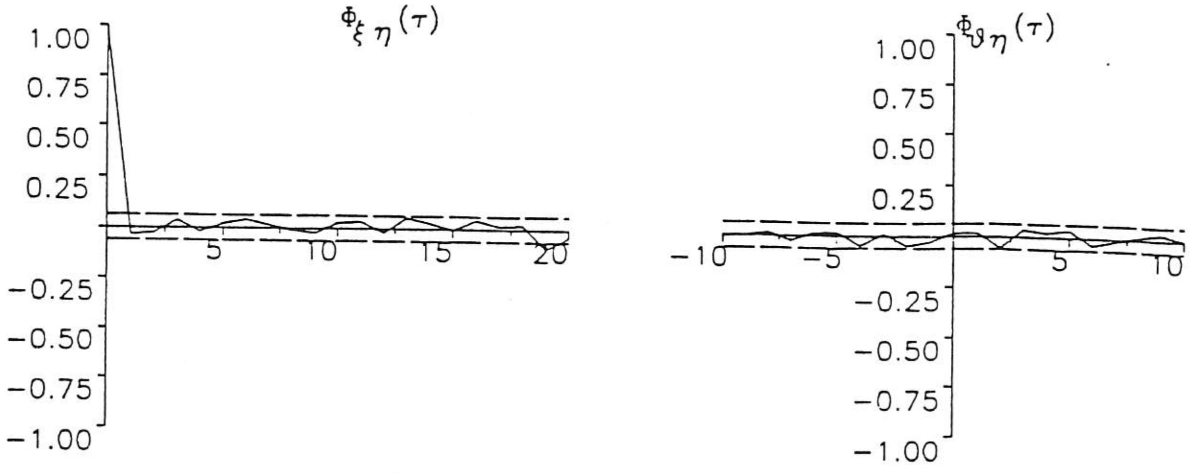


Figure 4 Global nonlinear model validity tests for example S_4

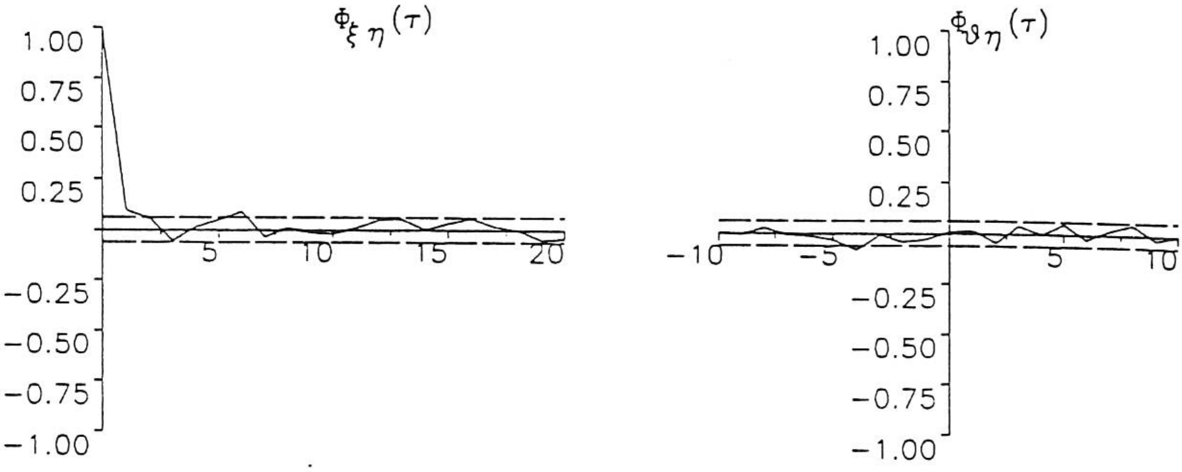


Figure 5 Global nonlinear tests for example S_4 with the term $y_2(t-1)\,u_1(t-1)$ omitted from the model

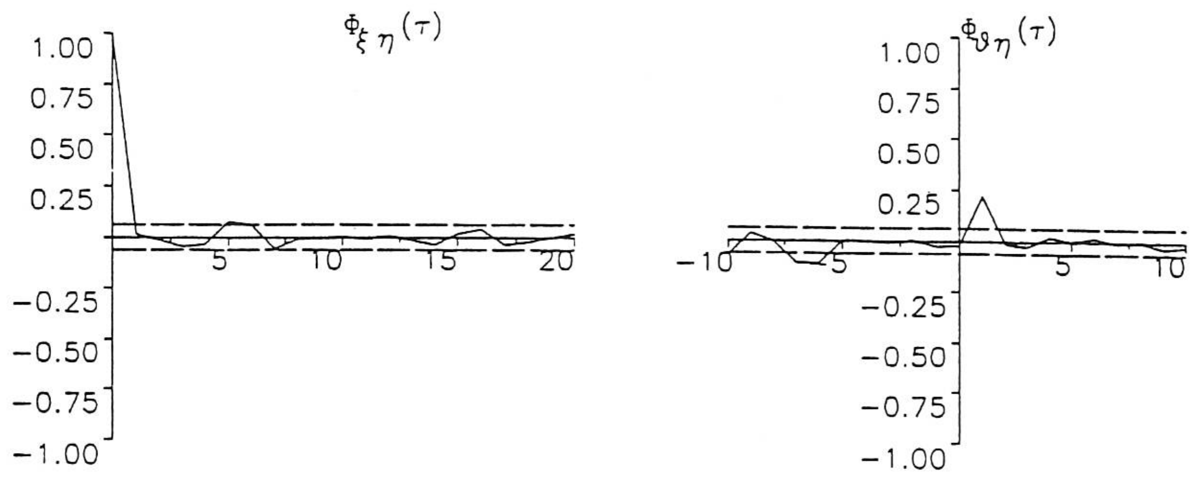


Figure 6 Global nonlinear tests for example S_4 with the term $u_2(t-1)$ omitted from the model