Monograph:

Mao, K.Z. and Billings, S.A. (1997) Multi-directional Model Validity Tests for Nonlinear System Identification. Research Report. ACSE Research Report 677 . Department of Automatic Control and Systems Engineering

Multi-directional Model Validity Tests for Nonlinear System Identification

K.Z.Mao S.A.Billings

Department of Automatic Control and Systems Engineering
University of Sheffield
Mappin Street, Sheffield S1 3JD
United Kingdom

Research Report No. 677

20 June 1997

University of Sheffield

Multi-directional Model Validity Tests for Nonlinear System Identification

K.Z.Mao S.A.Billings

Department of Automatic Control and Systems Engineering

University of Sheffield

Sheffield S1 3JD, UK

Abstract

New multi-directional model validation tests are derived to provide improved statistical validation test procedures for a wide class of nonlinear modelling methods.

1 Introduction

Model validation is an important procedure in any system identification study. If the model structure is correct and the estimated parameters are unbiased the residuals should form an independent random sequence and should be unpredictable from all past inputs, outputs and residuals. Based on this principle correlation test procedures consisting of tests using the autocorrelation function of the residuals and the cross-correlation function between the residuals and the inputs were developed (*e.g.* Box 1976, Bohlin 1978, Soderstrom and Stoica 1990). But these tests, which were originally developed for linear system model validation, are not sufficient to detect unmodelled nonlinear terms. To solve this problem nonlinear system model validation algorithms which include both classical and higher order correlation tests were developed (*e.g.* Billings and Voon 1983, 1986, Billings and Zhu 1994, 1995). These nonlinear model validation algorithms have been successfully applied in practical system identification including model validation of a turbocharged automotive diesel engine (Billings *et al* 1991), a distillation column (Sriniwas *et al* 1995), a parallel-tube heat exchanger (Thomson *et al* 1996) and others. But there are certain situations where these algorithms do not perform adequately. This means that even if predictable components remain in the residuals the correlation functions can, under certain conditions, still fall inside the 95% confidence bands. This problem has been previously observed in the higher order correlation tests and was referred to as the small value problem in Billings and Voon (1986). But no solution was proposed.

In this study several contrived examples are used to induce these problems and to study why they occur. The causes of the problem are then analysed and two multi-directional test algorithms are developed as solutions. The multi-directional test algorithms are shown to provide more reliable results than previously developed algorithms and simulated examples are used to demonstrate the application of the new tests.

2 An analysis of previously developed algorithms

2.1 Preliminaries

Consider a SISO dynamic nonlinear model

$$y(k) = \hat{F}[y(k-1),\ldots,y(k-n_y),u(k-1),\ldots,u(k-n_u),\epsilon(k-1),\ldots,\epsilon(k-n_\epsilon)]+\epsilon(k) \quad (1)$$

where $y(k)$, $u(k)$ and $\epsilon(k)$ denote the output, input and residual at time instant k, $\hat{F}$ is the estimated nonlinear function. Two algorithms are available to test the validity of this model.

Test (I) (Billings and Voon 1986)

$$\begin{array}{rcll} \Phi_{\epsilon\epsilon}(\tau) & = & \delta(\tau) & \text{for any } \tau \\ \Phi_{u\epsilon}(\tau) & = & 0 & \text{for any } \tau \\ \Phi_{(u^2)'\epsilon}(\tau) & = & 0 & \text{for any } \tau \\ \Phi_{(u^2)'\epsilon^2}(\tau) & = & 0 & \text{for any } \tau \\ \Phi_{\epsilon(\epsilon u)}(\tau) & = & 0 & \text{for any } \tau \geq 1 \end{array} \quad (2)$$

Test (II) (Billings and Zhu 1994)

$$\begin{array}{rcll} \Phi_{(\epsilon^2)'(y\epsilon)'}(\tau) & = & \kappa\,\delta(\tau) & \text{for any } \tau \\ \Phi_{(u^2)'(y\epsilon)'}(\tau) & = & 0 & \text{for any } \tau \end{array} \quad (3)$$

where $0 < \kappa < 1$. The dash $'$ in tests (2) and (3) denotes that the mean level has been removed from the corresponding signal.

Test (II) checks the correlation between the output and the residuals. This enhances the tests based on the residuals and inputs only, and only two tests are needed, even for multi-input and/or multi-output systems (Billings and Zhu 1995).

For all the tests above the normalised correlation functions Φ are computed based on the formula

$$\Phi_{vw}(\tau) = \frac{\sum_{k=1}^{N-\tau} v(k)w(k+\tau)}{\left[\left(\sum_{k=1}^{N-\tau} v^2(k)\right)\left(\sum_{k=1}^{N-\tau} w^2(k)\right)\right]^{0.5}} \quad (4)$$

Throughout the 95% confidence bands will be computed using the approximate formula $1.96/\sqrt{N}$ where N is the data length.

2.2 An example where Tests (I) and (II) can fail to detect unmodelled terms

Tests (I) and (II) are independent of the nonlinear system identification algorithm and the form of the model that is estimated. They have therefore been used by many authors over a wide range of model types including the nonlinear polynomial model, nonlinear rational model and neural networks, and they work well under most circumstances. Some typical successful examples include model validation of a turbocharged automotive diesel engine (Billings *et al* 1991), a distillation column (Sriniwas *et al* 1995), a parallel-tube heat exchanger (Thomson *et al* 1996) and others. However there are certain situations where the two tests can fail to adequately detect unmodelled terms. This is best illustrated by an example which has been designed to exagerate this deficiency.

Example 1

$$\begin{aligned} z(k) &= \sin[u(k-1)\pi] + \cos[u(k-2)\pi] \\ y(k) &= z(k) + e(k) \end{aligned} \tag{5}$$

where $z(k)$ and $y(k)$ denote the noise free output and the output measurement at time k $(k = 1, 2, \ldots, N)$. The input $\{u(k)\}$ is a uniformly distributed random sequence with amplititute ± 1. $\{e(k)\}$ is a normally distributed white noise sequence with zero mean and variance 0.25. A total of 400 data samples were generated and used in this identification study.

Nonlinear polynomial models were employed to fit the data. Initially, the degree of nonlinearity of the polynomial model was deliberately incorrectly set to 2, and the maximum lag of the input was set to 2. Estimation produced the following model

$$\begin{aligned} y(k) = \; & 0.7651 + 0.8887u(k-1) + 0.1169u(k-2) + 0.1017u^2(k-1) \\ & -2.2053u^2(k-2) + \epsilon(k) \end{aligned} \tag{6}$$

The results of tests (I) and (II) illustrated in Fig.1 suggest that the model in eqn (6) is valid. However maps constructed from the true model in eqn (5) and the identified model in eqn (6), illustrated in Fig.2, clearly show that the model eqn (6) does not capture the function underlying the data at all. Tests (I) and (II) therefore do not perform adequately to detect missing model terms in this example.

2.3 Why do tests (I) and (II) perform inadequately?

Consider the following nonlinear system

$$y(k) = f[u(k-1), u(k-2), \ldots, u(k-n_u)] + e(k)$$

where the disturbance $\{e(k)\}$ is an independently and identically distributed random sequence with zero mean and finite variance. Assuming that the identified model is

$$y(k) = \hat{f}[u(k-1), u(k-2), \ldots, u(k-m_u)] + \epsilon(k)$$

The residual $\epsilon(k)$ is

$$\epsilon(k) = f[\bullet] - \hat{f}[\bullet] + e(k) = \tilde{f}[u(k-1), u(k-2), \ldots, u(k-d)] + e(k)$$

where $d = \max\{n_u, m_u\}$. If f, $\hat{f}$ and $\tilde{f}$ are in the commonly used nonlinear polynomial form, $\epsilon(k)$ can be interpreted as

$$\epsilon(k) = \sum_{i=1}^{m} \tilde{a}_i u^{k_{i1}}(k-1) u^{k_{i2}}(k-2) \ldots u^{k_{id}}(k-d) \tag{7}$$

where $\tilde{a}_i$ is the parameter estimation bias of the i^{th} term, $k_{ij} \geq 0$, $j = 1, 2, \ldots, d$.

Consider the cross correlation function $\Phi_{u\epsilon}(\tau)$, $1 \leq \tau \leq d$.

$$\begin{aligned}
\Phi_{u\epsilon}(\tau) &= \frac{\sum_{k=1}^{N-\tau}[u(k)\epsilon(k+\tau)]}{\left[\left(\sum_{k=1}^{N-\tau} u^2(k)\right)\left(\sum_{k=1}^{N-\tau} \epsilon^2(k)\right)\right]^{0.5}} \\
&= \frac{\sum_{k=1}^{N-\tau}[u(k)\sum_{i=1}^{m} \tilde{a}_i u^{k_{i1}}(k+\tau-1) u^{k_{i2}}(k+\tau-2) \ldots u^{k_{id}}(k+\tau-d)]}{\left[\left(\sum_{k=1}^{N-\tau} u^2(k)\right)\left(\sum_{k=1}^{N-\tau} \epsilon^2(k)\right)\right]^{0.5}} \\
&= \frac{\sum_{k=1}^{N-\tau}[\sum_{i=1}^{m} \tilde{a}_i u^{q_{i1}}(k+\tau-1) u^{q_{i2}}(k+\tau-2) \ldots u^{q_{id}}(k+\tau-d)]}{\left[\left(\sum_{k=1}^{N-\tau} u^2(k)\right)\left(\sum_{k=1}^{N-\tau} \epsilon^2(k)\right)\right]^{0.5}}
\end{aligned} \tag{8}$$

If the input is an independent random sequence whose odd order moments are zero, eqn (8) is approximately equal to

$$\Phi_{u\epsilon}(\tau) \approx \frac{\sum_{k=1}^{N-\tau}[\sum_{i\in\Omega_1} \tilde{a}_i u^{q_{i1}}(k+\tau-1) u^{q_{i2}}(k+\tau-2) \ldots u^{q_{id}}(k+\tau-d)]}{\left[\left(\sum_{k=1}^{N-\tau} u^2(k)\right)\left(\sum_{k=1}^{N-\tau} \epsilon^2(k)\right)\right]^{0.5}} \tag{9}$$

where Ω_1 is a set of integers

$$\Omega_1 = \{i, \text{where } q_{i1}, q_{i2}, \ldots, q_{id} \text{ are all even numbers}\}$$

For a linear stable system the residual can be interpreted as $\epsilon(k) = \sum_{i=1}^{n} \tilde{a}_i u(k-i)$, and the cross correlation function $\Phi_{u\epsilon}(\tau)$ will depend on the term $\tilde{a}_\tau u(k-\tau)$ only. If this term is important the cross correlation function will have a relatively large value. Consequently the cross correlation test works in detecting unmodelled process terms like $u(k-\tau)$ in linear system model validation. But when the cross correlation test is extended to nonlinear systems the situation can be quite different because the size of $\Phi_{u\epsilon}(\tau)$ can be determined by more than one term as shown in eqn (9). Even if some important model terms are unmodelled $\Phi_{u\epsilon}(\tau)$ can be very small if

(i) $\tilde{a}_i$ is small, $i \in \Omega_1$ or

(ii) $\tilde{a}_i u^{q_{i1}}(k+\tau-1)u^{q_{i2}}(k+\tau-2)\ldots u^{q_{id}}(k+\tau-d)$ terms cancel with each other, $i \in \Omega_1$.

If the size of $\Phi_{u\epsilon}(\tau)$ is smaller than the confidence band some missed terms, which should be detected in theory, cannot be detected in practice. This is largely due to the higher order moments that contribute to the test and which are known to be data sensitive under certain conditions.

Consider the autocorrelation function $\Phi_{\epsilon\epsilon}(\tau)$, $1 \le \tau \le d$.

$$\begin{aligned}\Phi_{\epsilon\epsilon}(\tau) &= \frac{\sum_{k=1}^{N-\tau}[\epsilon(k)\epsilon(k+\tau)]}{\sum_{k=1}^{N-\tau}\epsilon^2(k)} \\ &= \frac{\sum_{k=1}^{N-\tau}[\sum_{i=1}^{m}\tilde{a}_i u^{k_{i1}}(k-1)\ldots u^{k_{id}}(k-d)\sum_{i=1}^{m}\tilde{a}_i u^{k_{i1}}(k+\tau-1)\ldots u^{k_{id}}(k+\tau-d)]}{\sum_{k=1}^{N-\tau}\epsilon^2(k)} \\ &= \frac{\sum_{k=1}^{N-\tau}[\sum_{i=1}^{M}\tilde{b}_i u^{p_{i1}}(k+\tau-1)u^{p_{i2}}(k+\tau-2)\ldots u^{p_{i(d+\tau)}}(k-d)]}{\sum_{k=1}^{N-\tau}\epsilon^2(k)} \qquad (10)\end{aligned}$$

where M is the number of terms contained in $\epsilon(k)\epsilon(k+\tau)$, $\tilde{b}_i$ is the product of any two $\tilde{a}_j$ and $\tilde{a}_l$, $j = 1, 2, \ldots, m$ and $l = 1, 2, \ldots, m$.

If the input is an independent random sequence with zero odd moments eqn (10) is approximately equal to

$$\Phi_{\epsilon\epsilon}(\tau) \approx \frac{\sum_{k=1}^{N-\tau}[\sum_{i\in\Omega_2}\tilde{b}_i u^{p_{i1}}(k+\tau-1)u^{p_{i2}}(k+\tau-2)\ldots u^{p_{i(d+\tau)}}(k-d)]}{\sum_{k=1}^{N-\tau}\epsilon^2(k)} \qquad (11)$$

where Ω_2 is a set of integers

$$\Omega_2 = \left\{i, \text{where } p_{i1}, p_{i2}, \ldots, p_{i(d+\tau)} \text{ are all even numbers}\right\}$$

Similar to the cross correlation function case the size of autocorrelation function can be smaller than the 95% confidence band even if predictable components remain in the residual if one or both of the following two cases occur(s)

(i) $\tilde{b}_i$ is small, where $i \in \Omega_2$ or

(ii) $\tilde{b}_i u^{p_{i1}}(k+\tau-1)u^{p_{i2}}(k+\tau-2)\ldots u^{p_{i(d+\tau)}}(k-d)$ terms cancel with each other, where $i \in \Omega_2$.

The above analysis reveals that cancellation or small values of $\tilde{a}_i$ $(i \in \Omega_1)$ or $\tilde{b}_i$ $(i \in \Omega_2)$ are potential factors that can cause incorrect results in both the autocorrelation and cross correlation tests. Whether these situations which can cause failure will occur in practice is system and input signal dependent but the possibility does exist. Consider an evaluation of Example 1.

In fact the model eqn (5) can be adequately approximated by a polynomial model of nonlinearity degree 4

$$\begin{aligned} y(k) &= 1.0653 + 2.7084u(k-1) - 4.8774u^2(k-2) - 2.8793u^3(k-1) \\ &\quad +2.9414u^4(k-2) + e(k) \end{aligned} \tag{12}$$

Comparing eqn (6) and eqn (12) yields

$$\begin{aligned} \epsilon(k) &= 0.3 + 1.8197u(k-1) - 0.1169u(k-2) - 0.1017u^2(k-1) - 2.672u^2(k-2) \\ &\quad -2.8793u^3(k-1) + 2.9414u^4(k-2) + e(k) \\ &= \tilde{a}_0 + \tilde{a}_1u(k-1) + \tilde{a}_2u(k-2) + \tilde{a}_3u^2(k-1) + \tilde{a}_4u^2(k-2) + \tilde{a}_5u^3(k-1) \\ &\quad +\tilde{a}_6u^4(k-2) + e(k) \end{aligned} \tag{13}$$

Consider initially the cross correlation function $\Phi_{u\epsilon}(1)$

$$\begin{aligned} \Phi_{u\epsilon}(1) &= \frac{\sum_{k=1}^{N-1} u(k)\epsilon(k+1)}{\left[\left(\sum_{k=1}^{N-1} u^2(k)\right)\left(\sum_{k=1}^{N-1} \epsilon^2(k)\right)\right]^{0.5}} \approx \frac{\tilde{a}_1 \sum_{k=1}^{N-1} u^2(k) + \tilde{a}_5 \sum_{k=1}^{N-1} u^4(k)}{\left[\left(\sum_{k=1}^{N-1} u^2(k)\right)\left(\sum_{k=1}^{N-1} \epsilon^2(k)\right)\right]^{0.5}} \\ &= \frac{237.5122 - 224.925}{\sqrt{130.5227 \times 201.5874}} = 0.0645 \end{aligned} \tag{14}$$

In this example the size of $\Phi_{u\epsilon}(1)$ is determined by two terms $\tilde{a}_1u(k)$ and $\tilde{a}_5u^3(k)$ in the residual $\epsilon(k+1)$. Correlations between these two terms and $u(k)$ have large values, *i.e.* 237.5122 and 224.925 respectively, but these two large values cancel as shown in eqn (14). The overall value falls inside the 95% confidence band ± 0.098, and the severe bias of $\hat{a}_1$ and the missing term $\tilde{a}_3u^3(k-1)$ were therefore not detected.

Consider the cross correlation function $\Phi_{u\epsilon}(2)$

$$\begin{aligned} \Phi_{u\epsilon}(2) &= \frac{\sum_{k=1}^{N-2} u(k)\epsilon(k+2)}{\left[\left(\sum_{k=1}^{N-2} u^2(k)\right)\left(\sum_{k=1}^{N-2} \epsilon^2(k)\right)\right]^{0.5}} \approx \frac{\tilde{a}_2 \sum_{k=1}^{N-2} u^2(k)}{\left[\left(\sum_{k=1}^{N-2} u^2(k)\right)\left(\sum_{k=1}^{N-2} \epsilon^2(k)\right)\right]^{0.5}} \\ &= \frac{-0.1169 \times 130.2011}{\sqrt{130.5227 \times 201.5874}} = -0.075 \end{aligned}$$

Although no cancellation is involved $\Phi_{u\epsilon}(2)$ is small because of the small value of $\tilde{a}_2$.

Consider the autocorrelation of the residual $\Phi_{\epsilon\epsilon}(1)$

$$\begin{aligned} \Phi_{\epsilon\epsilon}(1) &= \frac{\sum_{k=1}^{N-1} \epsilon(k)\epsilon(k+1)}{\sum_{k=1}^{N-1} \epsilon^2(k)} \\ &\approx \frac{(\tilde{a}_1\tilde{a}_2 + 2\tilde{a}_0\tilde{a}_3 + 2\tilde{a}_0\tilde{a}_4)\sum_{k=1}^{N-1} u^2(k) + (\tilde{a}_3\tilde{a}_4 + \tilde{a}_2\tilde{a}_5 + 2\tilde{a}_0\tilde{a}_6)\sum_{k=1}^{N-1} u^4(k)}{\sum_{k=1}^{N-1} \epsilon^2(k)} \\ &\quad + \frac{(\tilde{a}_3^2 + \tilde{a}_4^2)\sum_{k=1}^{N-1} u^2(k)u^2(k+1) + \tilde{a}_3\tilde{a}_6 \sum_{k=1}^{N-1} u^6(k) + \tilde{a}_3\tilde{a}_4 \sum_{k=1}^{N-2} u^2(k)u^2(k+2)}{\sum_{k=1}^{N-1} \epsilon^2(k)} \end{aligned}$$

$$
\begin{aligned}
& + \frac{\tilde{a}_4\tilde{a}_6\sum_{k=1}^{N-1}u^2(k)u^4(k+1)+\tilde{a}_3\tilde{a}_6\sum_{k=1}^{N-2}u^2(k+2)u^4(k)}{\sum_{k=1}^{N-1}\epsilon^2(k)} \\
& + \frac{\tilde{a}_4\tilde{a}_6\sum_{k=1}^{N-1}u^2(k+1)u^4(k)+\tilde{a}_6^2\sum_{k=1}^{N-1}u^4(k)u^4(k+1)+\tilde{a}_0^2(N-1)}{\sum_{k=1}^{N-1}\epsilon^2(k)} \\
= & \frac{-244.9838+185.62+302.0358-16.7268+11.1954-200.6318-7.4647}{199.291} \\
& + \frac{-194.5591+127.6718+35.91}{199.291} \\
= & -0.0097
\end{aligned}
$$

$\Phi_{\epsilon\epsilon}(1)$ is small and falls inside the 95% confidence band for the same reason as $\Phi_{u\epsilon}(1)$.

In theory if important terms are unmodelled correlation functions in tests (I) or (II) can have nonzero values. This is true. But the problem is that the size of these nonzero values are system and input signal dependent and can be smaller than the confidence bands. The results above for a contrived example reveal that the cancellation and small value of bias of some estimated parameters are the two factors that lead to incorrect results in the lower order correlation functions. In a similar manner these two factors can lead to incorrect results in higher order correlation tests as well. If all the five correlation tests in test (I) or both tests in test (II) provide incorrect results unmodelled terms will not be detected. Example 1 is just a simple instance of this but the causes summarised here are general and can happen in more complex systems.

3 Multi-directional model validity tests

3.1 The basic principle

In the previous section it was shown that the correlation function values can be model and input signal dependent and can be smaller than the confidence values even if predictable components are contained in the residuals. This conclusion is based on the assumption that the input sequence is random and independent. Consider what happens to the correlation functions if the input signal is deterministic and correlated. Assume that the residual is

$$\epsilon(k) = c_1\varphi_1(k) + c_2\varphi_2(k) + \ldots + c_m\varphi_m(k) + e(k) \tag{15}$$

where $\varphi_i(k)$ $(i = 1, 2, \ldots, m)$ are linear or nonlinear functions of the regression variables $u(k-1)$, …, $u(k-d)$. Consider the autocorrelation function of the residuals

$$
\begin{aligned}
\Phi_{\epsilon\epsilon}(\tau) &= \frac{\sum_{k=1}^{N-\tau}[\epsilon(k)\epsilon(k+\tau)]}{\sum_{k=1}^{N-\tau}\epsilon^2(k)} \\
&= \frac{\sum_{k=1}^{N-\tau}[c_1^2\varphi_1(k)\varphi_1(k+\tau)+\ldots+c_m^2\varphi_m(k)\varphi_m(k+\tau)+c_1c_m\varphi_1(k)\varphi_m(k+\tau)]}{\sum_{k=1}^{N-\tau}[c_1^2\varphi_1^2(k)+\ldots+c_m^2\varphi_m^2(k)+c_1c_m\varphi_1(k)\varphi_m(k)]}
\end{aligned}
$$

If the input signal is deterministic and correlated, the regressors φ_i $(i = 1, 2, \ldots, m)$ will be as well. For small τ we have

$$\begin{aligned}
\textstyle\sum_{k=1}^{N-\tau}[\varphi_1(k)\varphi_1(k+\tau)] &\approx \textstyle\sum_{k=1}^{N-\tau}[\varphi_1^2(k)] \\
&\vdots \\
\textstyle\sum_{k=1}^{N-\tau}[\varphi_1(k)\varphi_m(k+\tau)] &\approx \textstyle\sum_{k=1}^{N-\tau}[\varphi_1(k)\varphi_m(k)] \\
\textstyle\sum_{k=1}^{N-\tau}\varphi_m(k)\varphi_m(k+\tau) &\approx \textstyle\sum_{k=1}^{N-\tau}[\varphi_m^2(k)]
\end{aligned}$$

$\Phi_{\epsilon\epsilon}(\tau)$ should therefore have a relatively large value. Similarly the other tests in tests (I) and (II) can be enhanced as well. However random and uncorrelated input are often used by many authors in the literature, and once data samples are provided the input signal can not be changed at the model identification stage.

Assume that a predictable component, for example $u^2(k-\tau)$, is contained in the residual $\epsilon(k)$. The plot of $\epsilon(k)$ versus $u(k-\tau)$ will be a quadratic curve. This provides a clue. If the input sequence $\{u(k-\tau)\}$ is first sorted in terms of size rather than time of occurrence and the residual sequence is also sorted to correspond and to maintain time alignment the newly generated residual sequence will be deterministic and correlated if a linear or a nonlinear term $g[u(k-\tau)]$ is contained in the residual. Thus correlation tests based on the resorted sequences should be enhanced. Alternatively, if the identified model represents the system adequately the residual time series will approximately converge to an independent and identically distributed noise sequence whose statistical properties should not be affected by the sorting operation because the sorting order has no relationship with the noise itself. The correlation functions based on the resorted sequences should therefore meet tests (I) and (II). The above consideration motivates the development of multi-directional model validity tests.

3.2 Signal Resorting

For convenience in the following sections, regression variables $u(k-1)$, ..., $u(k-n_u)$ will be denoted as $x_1(k)$, ..., $x_{n_u}(k)$.

Discrete-time signals are usually sorted in terms of the time of occurrence. If the output $\{y(k), k = 1, 2, \ldots, N\}$ and the residual $\{\epsilon(k), k = 1, 2, \ldots, N\}$ are sorted in a new order rather than the time of occurrence, but maintaining alignment, new sequences will be produced. For example resorting the output sequence $\{y(k)\}$ in an order in which variable x_1 is ascending, yields a new output sequence

$$\{y_{x_1}(k), k = 1, 2, \ldots, N\} = \{y(k_{11}), y(k_{12}), \ldots, y(k_{1N})\}$$

where the positive integers $k_{1i} \in \{1, 2, \ldots, N\}$ $(i = 1, 2, \ldots, N)$ satisfy

$$x_1(k_{11}) \leq x_1(k_{12}) \leq \ldots \leq x_1(k_{1N})$$

Because this new sequence is produced by sorting the output in the ascending order of x_1, $\{y_{x_1}(k)\}$ will be referred to as the output distribution along the direction of x_1. In the same way the output distribution along the direction x_i $(i = 2, 3, \ldots, n_u)$ and the residual distribution along the direction x_i $(i = 1, 2, \ldots, n_u)$ can be obtained and are denoted by $\{y_{x_i}(k)\}$ and $\{\epsilon_{x_i}(k)\}$ respectively. Similarly the distribution of a signal $\{x_i(k), k = 1, 2, \ldots, N\}$ along its own direction is defined as

$$\left\{x_{i_{x_i}}(k), k = 1, 2, \ldots, N\right\} = \{x_i(k_{i1}), x_i(k_{i2}), \ldots, x_i(k_{iN})\}$$

where

$$x_i(k_{i1}) \leq x_i(k_{i2}) \leq \ldots \leq x_i(k_{iN})$$

3.3 Multi-directional model validity tests

In §3.1 it was shown that if a model adequately represents the true system the residual should be reduced to an independently and identically distributed random sequence whose statistical properties are not affected by the sorting operation. Consequently the resorted residual and input sequences should meet the following tests

Multi-directional test (I)

$$\begin{array}{rcll} \Phi_{\epsilon_\gamma \epsilon_\gamma}(\tau) & = & \delta(\tau) & \text{for any } \tau \\ \Phi_{u_\gamma \epsilon_\gamma}(\tau) & = & 0 & \text{for any } \tau \\ \Phi_{(u_\gamma^2)' \epsilon_\gamma}(\tau) & = & 0 & \text{for any } \tau \\ \Phi_{(u_\gamma^2)' \epsilon_\gamma^2}(\tau) & = & 0 & \text{for any } \tau \\ \Phi_{\epsilon_\gamma (u_\gamma \epsilon_\gamma)}(\tau) & = & 0 & \text{for any } \tau \geq 1 \end{array} \quad (16)$$

Multi-directional test (II)

$$\begin{array}{rcll} \Phi_{(\epsilon_\gamma^2)'(y_\gamma \epsilon_\gamma)'}(\tau) & = & \kappa \delta(\tau) & \text{for any } \tau \\ \Phi_{(u_\gamma^2)'(y_\gamma \epsilon_\gamma)'}(\tau) & = & 0 & \text{for any } \tau \end{array} \quad (17)$$

where $0 < \kappa < 1$, γ can be $x_1, x_2, \ldots, x_{n_u}$ in tests (16) and (17).

The procedure of applying the above two tests is summarised as follows.

(i) Check if the conditions in eqn (2) or eqn (3) hold. If the conditions are not satisfied the model is invalid. Otherwise set $i = 1$.

(ii) Generate a new sequence by sorting the sequence $\{x_i(k)\}$ in an ascending order

$$\left\{x_{i_{x_i}}(k), k = 1, 2, \ldots, N\right\} = \{x_i(k_{i1}), x_i(k_{i2}), \ldots, x_i(k_{iN})\}$$

where

$$x_i(k_{i1}) \leq x_i(k_{i2}) \leq \ldots \leq x_i(k_{iN})$$

(iii) Generate a new residual sequence by resorting the sequence $\{\epsilon(k), k = 1, 2, \ldots, N\}$ to correspond to the order in $\left\{x_{i_{x_i}}(k)\right\}$ in step (ii)

$$\{\epsilon_{x_i}(k), k = 1, 2, \ldots, N\} = \{\epsilon(k_{i1}), \epsilon(k_{i2}), \ldots, \epsilon(k_{iN})\}$$

and a new output sequence by resorting $\{y(k), k = 1, 2, \ldots, N\}$ in the same way as $\{\epsilon_{x_i}(k), k = 1, 2, \ldots, N\}$.

$$\{y_{x_i}(k), k = 1, 2, \ldots, N\} = \{y(k_{i1}), y(k_{i2}), \ldots, y(k_{iN})\}$$

(iv) Check if the conditions in eqn (16) or (17) hold. Set $i = i + 1$ and repeat steps (ii)-(iv) until $i = n_u$, the value of the maximum lag in the input.

(v) If conditions in eqns (2) or (3) and (16) or (17) are satisfied for all x_i the model is considered to be statistically valid, otherwise the model is not valid.

Consider the validity of model eqn (6) again. The results of the $u(k-1)$ sorted directional tests (I) and (II) are illustrated in Fig.3 and now clearly show that the model is not valid.

3.4 Compound multi-directional model validity tests

The multi-directional model validity tests in eqns (16) and (17) include tests in all possible directions. This may involve a large number of correlations which have to be inspected. Recently new nonlinear model validity test procedures, which significantly simplify and reduce the tests for multivariable systems were introduced by Billings and Zhu (1995). The basic idea of this algorithm is to put the residual and input sequences of different subsystems in a "condensed" residual and input sequence respectively, and then to check for correlations using the two "condensed" signals. This idea can be employed here to simplify the multi-directional validity tests.

Define following variables

$$\xi(k) = \epsilon_{x_1}(k) + \ldots + \epsilon_{x_{nu}}(k)$$

$$\xi^2(k) = \epsilon^2_{x_1}(k) + \ldots + \epsilon^2_{x_{nu}}(k)$$

$$v(k) = x_{1_{x_1}}(k) + \ldots + x_{nu_{x_{nu}}}(k)$$

$$v^2(k) = x^2_{1_{x_1}}(k) + \ldots + x^2_{nu_{x_{nu}}}(k)$$

$$\mu(k) = \epsilon_{x_1}(k)x_{1_{x_1}}(k) + \ldots + \epsilon_{x_{nu}}(k)x_{nu_{x_{nu}}}(k)$$

$$\eta(k) = y_{x_1}(k)\epsilon_{x_1}(k) + \ldots + y_{x_{nu}}(k)\epsilon_{x_{nu}}(k)$$

The compound multi-directional tests can then be defined as

Compound multi-directional test (I)

$$\begin{array}{rclr} \Phi_{\xi\xi}(\tau) & = & \delta(\tau) & \text{for any } \tau \\ \Phi_{\upsilon\xi}(\tau) & = & 0 & \text{for any } \tau \\ \Phi_{(\upsilon^2)'\xi}(\tau) & = & 0 & \text{for any } \tau \\ \Phi_{(\upsilon^2)'\xi^2}(\tau) & = & 0 & \text{for any } \tau \\ \Phi_{\xi\mu}(\tau) & = & 0 & \text{for any } \tau \geq 1 \end{array} \qquad (18)$$

Compound multi-directional test (II)

$$\begin{array}{rclr} \Phi_{(\xi^2)'\eta'}(\tau) & = & \kappa\,\delta(\tau) & \text{for any } \tau \\ \Phi_{(\upsilon^2)'\eta'}(\tau) & = & 0 & \text{for any } \tau \end{array} \qquad (19)$$

where $0 < \kappa < 1$.

The application of the compound multi-directional tests (I) and (II) is summarised as follows.

(i) Check if the conditions in eqn (2) or (3) hold. If the conditions are not met the model is invalid. Otherwise set $i = 1$.

(ii) Generate a new sequence by sorting the sequence $\{x_i(k)\}$ in an ascending order

$$\left\{x_{i_{x_i}}(k), k = 1, 2, \ldots, N\right\} = \{x_i(k_{i1}), x_i(k_{i2}), \ldots, x_i(k_{iN})\}$$

where

$$x_i(k_{i1}) \leq x_i(k_{i2}) \leq \ldots \leq x_i(k_{iN})$$

(iii) Generate a new residual sequence by resorting the sequence $\{\epsilon(k), k = 1, 2, \ldots, N\}$ to correspond to the order in step (ii)

$$\{\epsilon_{x_i}(k), k = 1, 2, \ldots, N\} = \{\epsilon(k_{i1}), \epsilon(k_{i2}), \ldots, \epsilon(k_{iN})\}$$

and a new output sequence by sorting the sequence $\{y(k), k = 1, 2, \ldots, N\}$ in the same way as $\{\epsilon_{x_i}(k), k = 1, 2, \ldots, N\}$

$$\{y_{x_i}(k), k = 1, 2, \ldots, N\} = \{y(k_{i1}), y(k_{i2}), \ldots, y(k_{iN})\}$$

(iv) Set $i = i + 1$ and repeat steps (ii)-(iii) until $i = n_u$, where n_u is the maximum lag in the input.

(v) Generate new sequences $\{\xi(k)\}$, $\{\xi^2(k)\}$, $\{\upsilon(k)\}$, $\{\upsilon^2(k)\}$, $\{\mu(k)\}$, $\{\eta(k)\}$ and check if the conditions in (18) or (19) hold. If the conditions in tests (2) or (3) and (18) or (19) hold the model is considered to be statistically valid otherwise the model is not valid.

4 Examples

To illustrate the efficiency of new multi-directional model validity tests two examples were selected as demonstrations.

Example 2

Consider the following nonlinear system

$$\begin{aligned} z(k) &= \exp(-|u(k-1)|) + 1.2\tanh[u(k-2)\pi] \\ y(k) &= z(k) + e(k) \end{aligned} \tag{20}$$

where $z(k)$ and $y(k)$ denote the noise free output and the output measurement at time k. The input $\{u(k)\}$ was a uniformly distributed random sequence with zero mean and amplititute ± 1, the disturbance $\{e(k)\}$ is a normally distributed white noise sequence with zero mean and variance 0.01. A total of 400 data samples were generated.

Nonlinear polynomial models were employed to approximate the system in eqn (20). Initially the nonlinear degree of a polynomial model was deliberately incorrectly set to 3, and the maximum lag of the input was set to 2. Identifying the system using a forward regression orthogonal algorithm (Billings and Chen 1989) yielded the following model

$$y(k) = 0.826 + 2.645u(k-2) - 0.567u^2(k-1) - 1.6037u^3(k-2) - 1.1124u^2(k-1)u(k-2) + \epsilon(k) \tag{21}$$

The results of both tests (I) and (II) indicated that the model in eqn (21) is valid. However maps constructed from the true model in eqn (20) and the identified model in eqn (21), illustrated in Fig.4, clear show that the model in eqn (21) is not valid. The model deficiency was successfully detected by the $u(k-2)$ directional tests (I) and (II) and the compound multi-directional tests (I) and (II). Both results were almost identical and only the compound tests are shown in Fig.5.

The nonlinearity degree was then increased to 5 to yield the following model

$$\begin{aligned} y(k) &= 0.891 + 3.1511u(k-2) - 1.2552u^2(k-1) - 4.1103u^2(k-2) \\ &\quad + 0.8225u^4(k-1) + 2.2538u^5(k-2) + \epsilon(k) \end{aligned} \tag{22}$$

The $u(k-2)$ directional test and the compound multi-directional test were now all inside the confidence bands indicating that the model is statistically valid. This is confirmed by the map constructed from model eqn (22) shown in Fig.6.

Example 3

Consider the following nonlinear system

$$y(k) = 0.2y(k-1) + 1.5u^2(k-1) + u^2(k-2) + u^3(k-3) + e(k-1) + e(k) \tag{23}$$

where the input $\{u(k)\}$ was a uniformly distributed random sequence with amplititute ± 1, $\{e(k)\}$ is a normally distributed white noise sequence with zero mean and variance 0.04. A total of 400 data samples were generated and used.

Initially the nonlinearity degree, the maximum lags of the input, output and noise were deliberately incorrectly set to 2, 2, 1, and 1 respectively. These parameters produced a polynomial model with 15 candidate terms. Identifying the system using a forward regression orthogonal algorithm (Billings and Chen 1989) produced the following model

$$\begin{aligned} y(k) &= 0.2566y(k-1) + 1.5143u^2(k-1) + 0.8916u^2(k-2) \\ &\quad +0.1282u(k-1)u(k-2) + 0.1233\epsilon(k-1) + \epsilon(k) \end{aligned} \tag{24}$$

The model in eqn (24) is not valid because term $u^3(k-3)$ was unmodelled. This was detected by tests (I) and (II), and also by the $u(k-3)$-directional tests (I) and (II), and by the compound multi-directional tests (I) and (II) shown in Fig.7.

The nonlinearity degree, the maximum lags in the input, output and noise were then set to 3, 3, 1 and 1 respectively. Reidentifying the system in eqn (23) yielded the following model

$$\begin{aligned} y(k) &= 0.2129y(k-1) + 1.4932u^2(k-1) + 1.0083u^2(k-2) + 0.989u^3(k-3) \\ &\quad +0.8303\epsilon(k-1) + \epsilon(k) \end{aligned} \tag{25}$$

The $u(k-3)$-directional test and the compound multi-directional test were all inside the confidence bands indicating that the model is statistically valid. These coincide with the fact that model eqn (25) has correct model structure and accurate parameter estimates.

5 Conclusion

Model validation is an important procedure in system identification. In this study new multi-directional model validity test algorithms have been developed. The basic principle of these tests is to resort the residual sequence according to the amplititute of the corresponding input variable so that any predictable components in the resorted residuals can be more readily detected. The simulated examples show that the tests based on the resorted input and residual provide more reliable results than previously developed algorithms.

6 Acknowledgement

SAB gratefully acknowledges that part of this work was supported by EPSRC.

References

[1] Billings, S.A. and W.S.F.Voon, 1983, Structure detection and model validity tests in the identification of nonlinear systems, *IEE Proceedings Control Theory and Applications*, 130 (4), pp193-199.

[2] Billings, S.A. and W.S.F.Voon, 1986, Correlation based model validity tests for nonlinear models, *International Journal of Control*, 44 (1), pp235-244.

[3] Billings, S.A., S.Chen and R.J.Backhouse, 1989, The identification of linear and nonlinear models of a turbocharged automotive diesel engine, *Mechanical Systems and Signal Processing*, 3 (2), pp123-142.

[4] Billings, S.A. and S.Chen, 1989, Extended model set, global data and threshold model identification of severely nonlinear systems *International Journal of Control*, 50 (5), pp1897-1923.

[5] Billings, S.A. and Q.M.Zhu, 1994, Nonlinear model validation using correlation tests, *International Journal of Control*, 60 (6), pp1107-1120.

[6] Billings, S.A. and Q.M.Zhu, 1995, Model Validation Tests for Multivariable Nonlinear Models Including Neural Networks, *International Journal of Control*, 62 (4), pp749-766.

[7] Bohlin, T., 1978, Maximum power validation of models without higher order fitting, *Automatica*, 7 (2), pp137-146.

[8] Box, G.E.P. and G.M.Jenjins, 1976, *Time series analysis forcasting and control* (San Francisco: Holden-Day).

[9] Soderstrom, T and P.Stoica, 1990, On covariance function test used in system identification, *Automatica*, 26 (2), pp125-133.

[10] Sriniwas, G.R., Y.Arkun, I.Chien and B.A.Ogunnaike, 1995, Nonlinear identification and control of a high-purity distillation column: a case study, *Journal of Process Control*, 5, 149-162.

[11] Thomson, M., S.P.Schooling and M.Soufian, 1996, The practical application of a nonlinear identification methodology, *Control Engineering Practice*, 4 (3), 295-306.

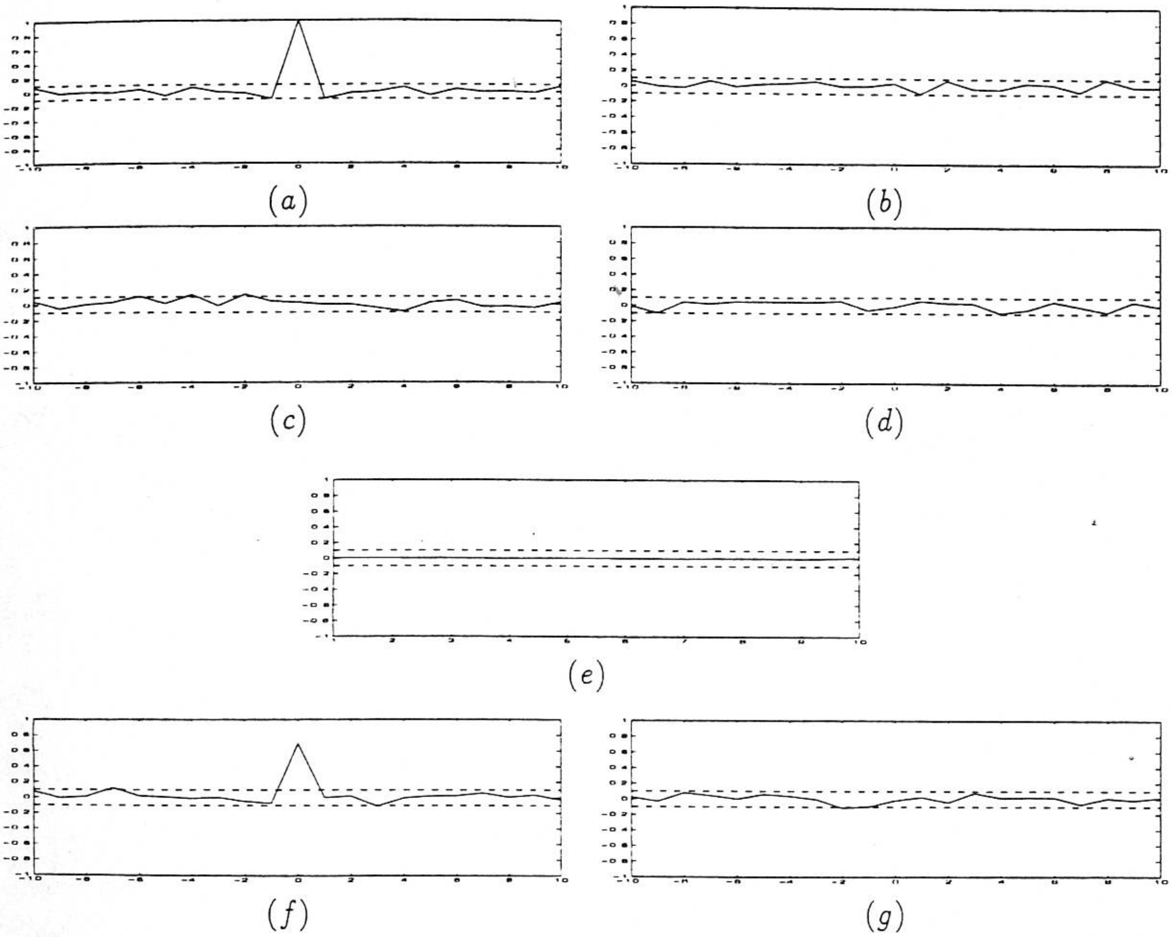

Figure 1: Results of tests (I) and (II) for model eqn (6) of Example 1 (a) $\Phi_{\epsilon\epsilon}$, (b) $\Phi_{u\epsilon}$, (c) $\Phi_{(u^2)'\epsilon}$, (d) $\Phi_{(u^2)'\epsilon^2}$, (e) $\Phi_{\epsilon(\epsilon u)}$, (f) $\Phi_{(\epsilon^2)'(y\epsilon)'}$, (g) $\Phi_{(u^2)'(y\epsilon)'}$

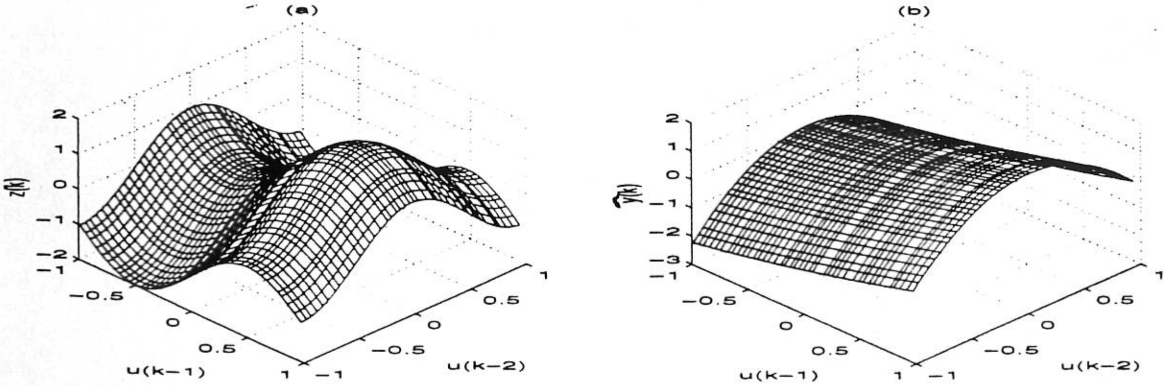

Figure 2: Maps constructed from the true model eqn (5) and the identified model eqn (6) for Example 1 ((a)-true (b)-identified)

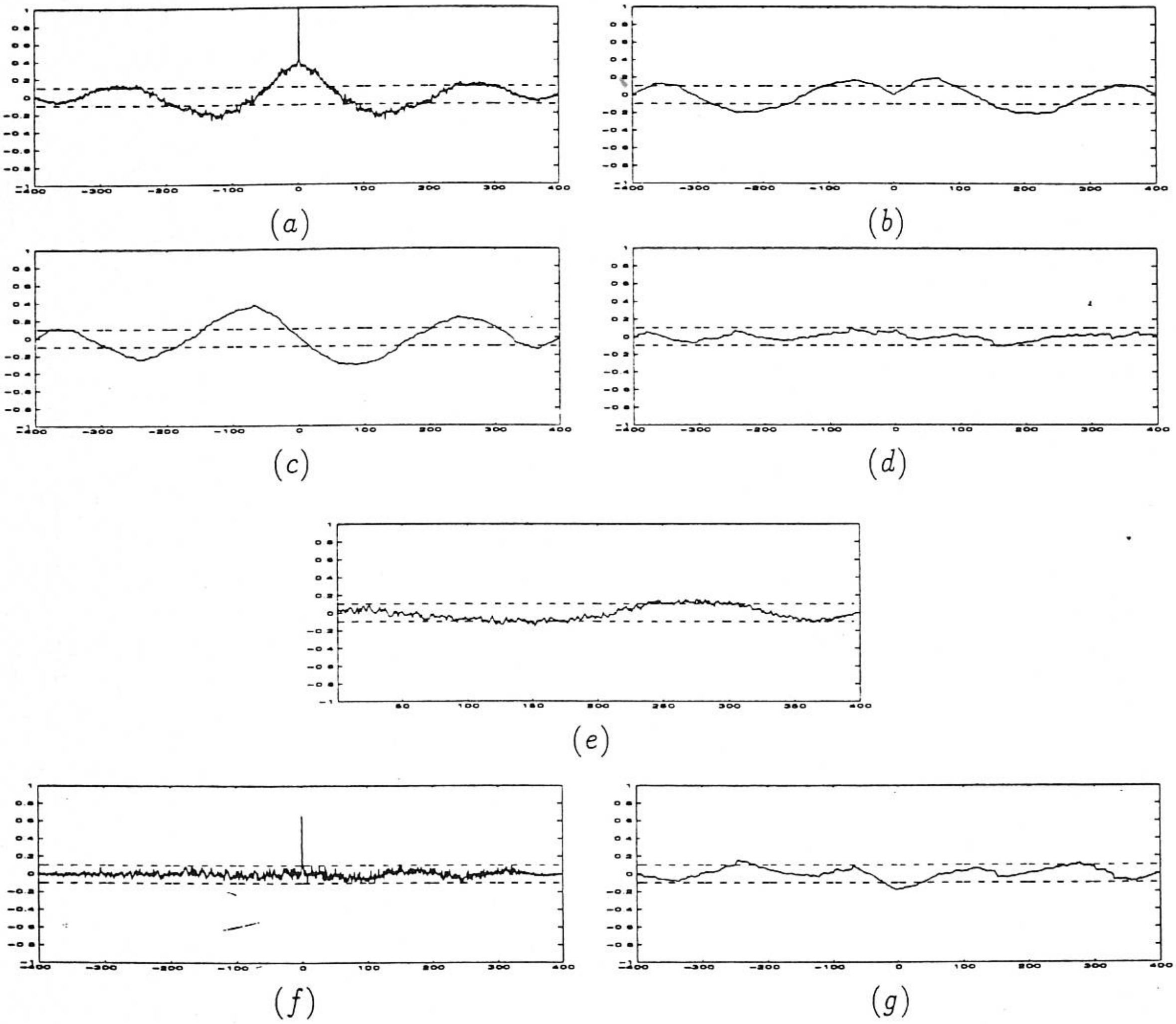

Figure 3: Results of the $u(k-1)$-directional tests (I) and (II) for model eqn (6) of Example 1 (a) $\Phi_{\epsilon_\gamma \epsilon_\gamma}$, (b) $\Phi_{u_\gamma \epsilon_\gamma}$, (c) $\Phi_{(u_\gamma^2)' \epsilon_\gamma}$, (d) $\Phi_{(u_\gamma^2)' \epsilon_\gamma^2}$, (e) $\Phi_{\epsilon_\gamma (\epsilon_\gamma u_\gamma)}$, (f) $\Phi_{(\epsilon_\gamma^2)'(y_\gamma \epsilon_\gamma)'}$, (g) $\Phi_{(u_\gamma^2)'(y_\gamma \epsilon_\gamma)'}$

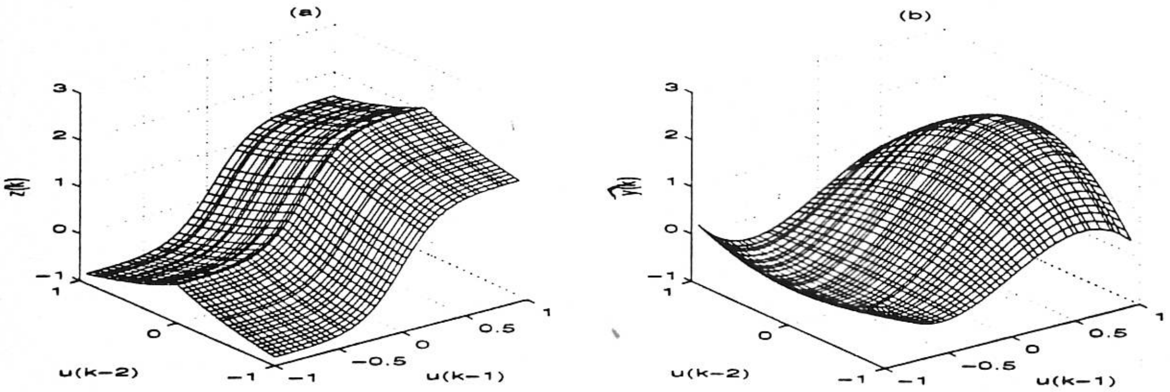

Figure 4: Maps constructed from the true model eqn (20) and the identified model eqn (21) for Example 2 ((a)-true (b)-identified)

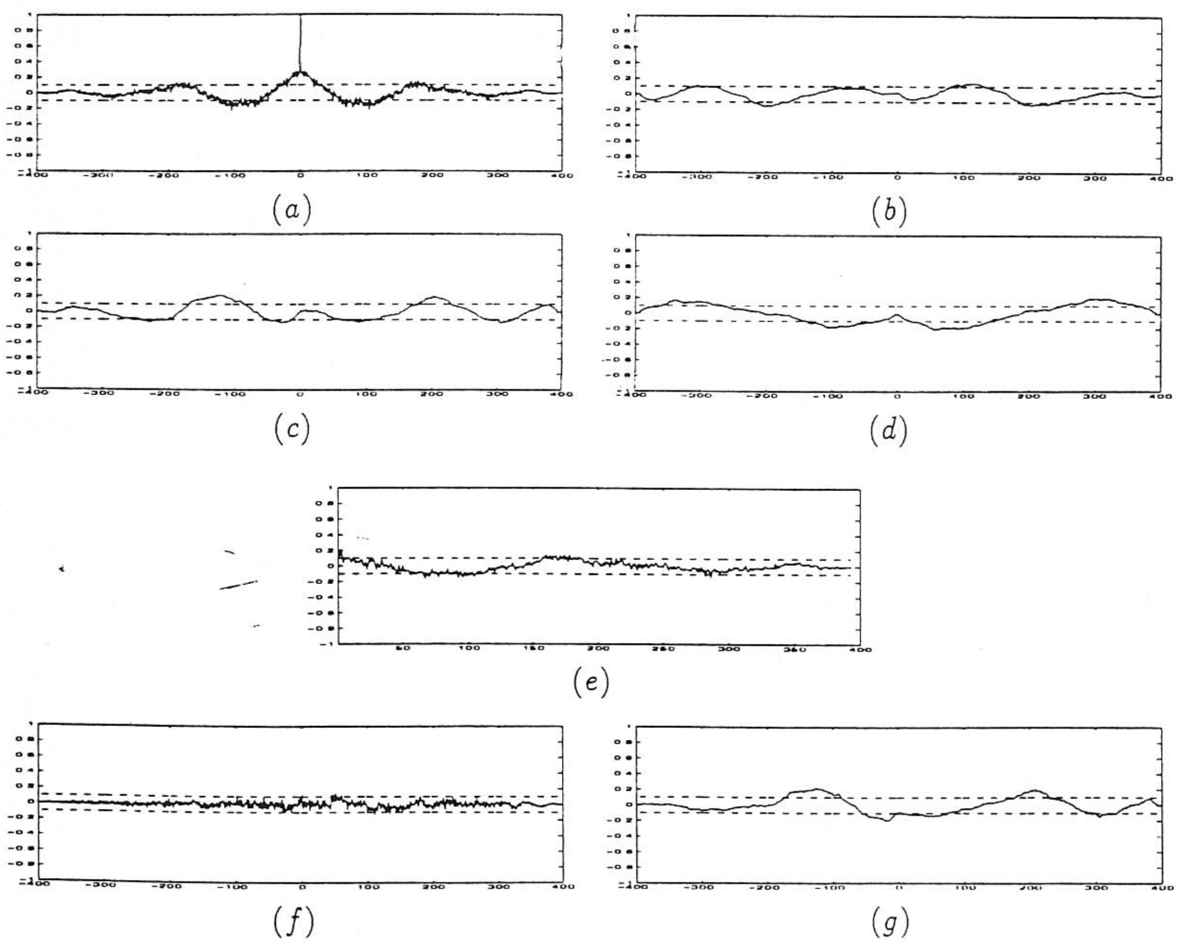

Figure 5: Results of the compound multi-directional tests (I) and (II) for model eqn (21) of Example 2 (a) $\Phi_{\xi\xi}$, (b) $\Phi_{\upsilon\xi}$, (c) $\Phi_{(\upsilon^2)'\xi}$, (d) $\Phi_{(\upsilon^2)'\xi^2}$, (e) $\Phi_{\xi\mu}$, (f) $\Phi_{(\xi^2)'\eta'}$, (h) $\Phi_{(\upsilon^2)'\eta'}$

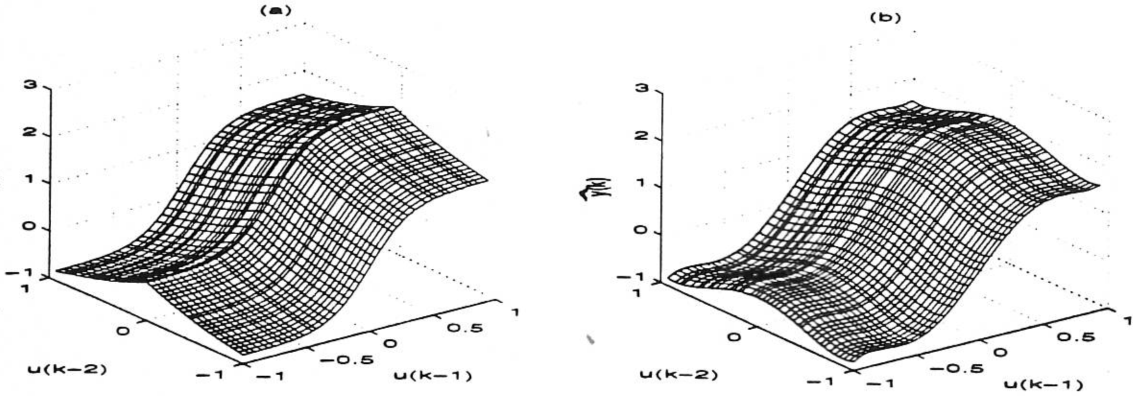

Figure 6: Maps constructed from the true model eqn (20) and the identified model eqn (22) for Example 2 ((a)-true (b)-identified)

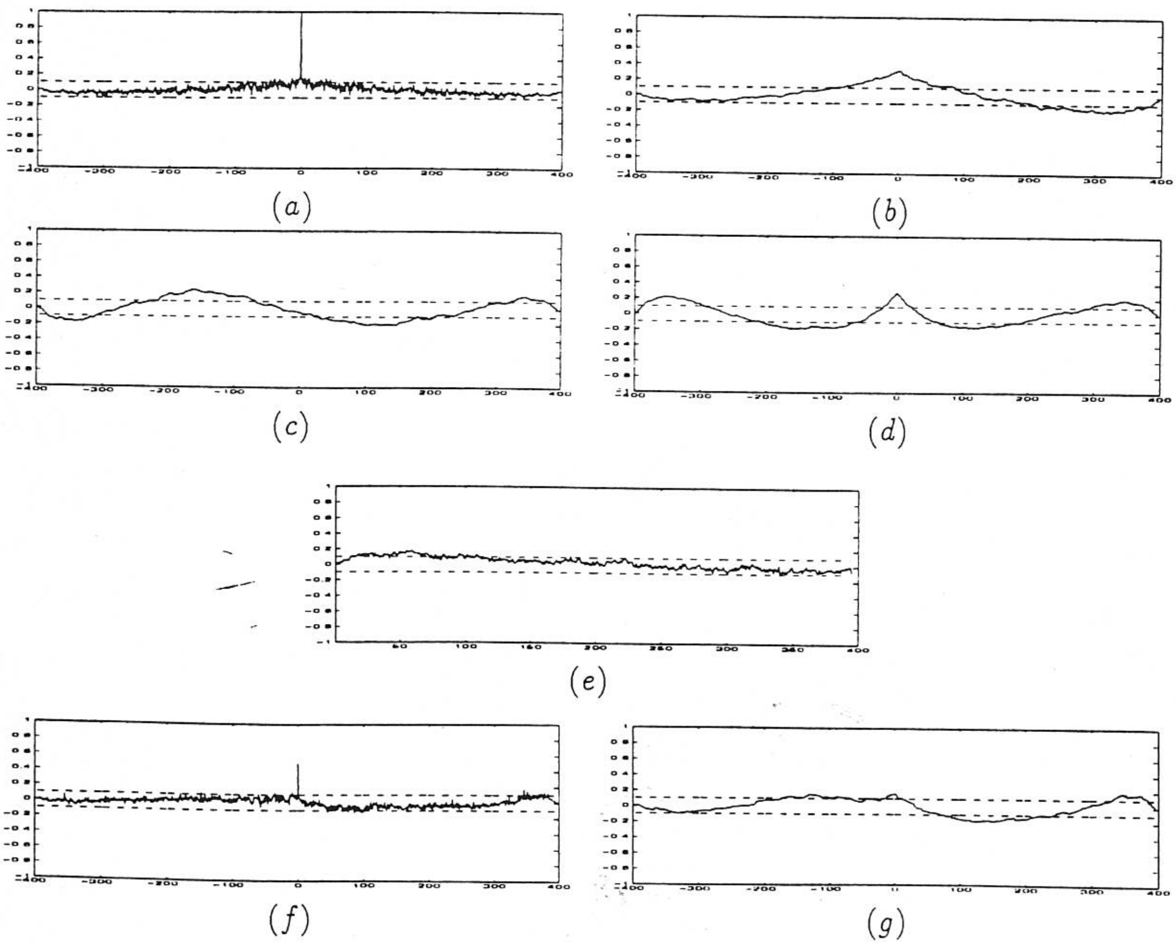

Figure 7: Results of the compound multi-directional tests (I) and (II) for model eqn (24) of Example 3 (a) $\Phi_{\xi\xi}$, (b) $\Phi_{\upsilon\xi}$, (c) $\Phi_{(\upsilon^2)'\xi}$, (d) $\Phi_{(\upsilon^2)'\xi^2}$, (e) $\Phi_{\xi\mu}$, (f) $\Phi_{(\xi^2)'\eta'}$, (g) $\Phi_{(\upsilon^2)'\eta'}$