On the interpretation and practice of dynamical differences between Hammerstein and Wiener models

L.A. Aguirre, M.C.S. Coelho and M.V. Corrêa

Abstract: It is suggested that the differences between the Hammerstein and Wiener models be interpreted and understood in terms of the system eigenvalues. In particular, it is shown that the Wiener representation should be preferred when the system dynamics vary with the operating point. Conversely, when only the system gain varies with the operating point, Hammerstein models generally outperform the Wiener representation. The paper also points out connections between such models and the more general non-linear autoregressive model with exogenous inputs (NARX) polynomial representation. From a practical control engineering point of view, the results presented seem to be more helpful than other ways of distinguishing between such model types. The main ideas are illustrated by means of three examples that use simulated and measured data.

1 Introduction

The identification of block-oriented models has attracted great attention in the last three decades or so. Such models are composed of two blocks, one is linear and dynamical, the other is non-linear and static. The various types of models differ in how such blocks are connected to each other. This paper is concerned with the Hammerstein and Wiener models which are illustrated in Fig. 1.

The interest in block-oriented models seems to have been rekindled lately because of their usefulness in simple yet apparently efficient control schemes [1, 2]. Regardless of being useful in such applications, block-oriented models are adequate to approximate a large class of non-linear processes.

In order to take advantage of block-oriented models in practical situations it is necessary to develop tools that would 1) enable the user to obtain such models directly from dynamical data, and also to 2) help choose which type of representation is preferable given the features of the process being modelled.

Many techniques have been put forward to handle the first aforementioned issue. An early paper [3] has discussed the identification of block-oriented models using non-parametric techniques, but more recently the emphasis seems to lay on parametric methods. Many methods assume that some aspect of the system is known a priori, such as the gain [4] or the structure of the linear dynamical block [5]. The method discussed in [6] does not assume prior knowledge but does assume that the input and output signals are sampled at different rates and that the output is oversampled. An interesting procedure has been recently suggested by which the static non-linearity is estimated from the same set of dynamical data used for identifying the dynamical linear model [7]. Frequently, the precise structure of the non-linearity is not assumed known but it is required that the non-linearity admits a polynomial representation [6, 8].

Although there are many available methods to identify block-oriented models from data, it seems that the judicious choice between the Hammerstein and Wiener models is an issue greatly neglected in practical situations. For instance, in the aforementioned works, no concrete justification is given to support the choice of the model type.

The differences between the Hammerstein and Wiener models have been studied and procedures to distinguish between them have been developed, see [9] for a review and [10] for a detailed description of parallel and multivariable systems. However, it is noticed that although the differences are mathematically shown, for instance in terms of Volterra kernels or non-linear correlation functions, the interpretation from an engineering point of view is far from obvious. On the other hand, the tests that have been suggested to distinguish between these models usually require specific inputs which are not always achievable in practice. (Another common situation is when a system is to be analysed and modelled from an existing set of data which does not usually attain the specific features required and no new tests are possible.) Considering such reasons, it is not hard to see why in practice no justification is given for choosing a Hammerstein or a Wiener model for a given plant.

The main concern of this paper is to show that an important dynamical difference between the Hammerstein and Wiener models is reflected on how the eigenvalues of such systems vary with the operating point. Because this interpretation is based on basic concepts of control engineering, it is believed that it will furnish a means by which practitioners will be able to readily decide between a Hammerstein or Wiener model without having to resort to difficult-to-estimate functions, such as Volterra kernels. The new interpretation has direct bearing in the practical identification of real systems using such block-oriented

IEE Proceedings online no. 20045152

doi: 10.1049/ip-cta:20045152

Paper received 18th August 2004. Originally published online 20th June 2005

L.A. Aguirre and M.C.S. Coelho are with the Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, M.G., Brazil

M.V. Corrêa is with the Curso de Engenharia Elétrica, UNILESTE-MG, Av. Tancredo Neves 3500, 35170-056 Cel. Fabriciano, M.G., Brazil

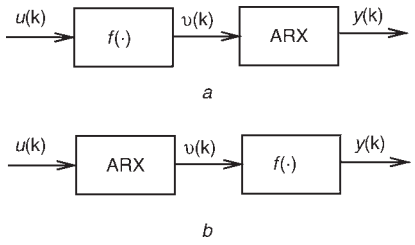

Fig. 1 *Block representations of models*
a Hammerstein
b Wiener

models. As part of the analysis presented in this paper, it is shown how the Hammerstein and Wiener models relate to the more general non-linear autoregressive model with exogenous inputs (NARX) polynomial representation. In particular it is shown under which conditions a given NARX polynomial model can be represented as a Hammerstein and/or a Wiener model. This, of course, has theoretical value in itself but not only that. In many practical instances NARX polynomial models are already available [11, 12] and to derive block-oriented counterparts from them could be useful.

2 Background

Since the early days of non-linear identification it has been pointed out that the block-oriented models are particular cases of the more general NARX model that can be written as [13]

$$y(k) = F^{\ell}[y(k-1),\ldots,y(k-n_y),u(k-d)\ldots u(k-n_u),e(k)] \tag{1}$$

where $u(k)$ and $y(k)$ are, respectively, the input and output signals and $e(k)$ accounts for uncertainties, possible noise and unmodelled dynamics. In this paper $F^{\ell}[\cdot]$ is assumed to be a polynomial-type function with non-linearity degree $\ell \in \mathbb{Z}^+$.

In particular, if the non-linearity $f(\cdot)$ only acts on the inputs of the ARX model, the result is a Hammerstein model. Omitting the noise terms for the sake of clarity, we have

$$\begin{aligned} y(k) = {} & a_1 y(k-1) + \cdots + a_{n_y} y(k-n_y) \\ & + f[u(k-d)\ldots u(k-n_u)] \end{aligned} \tag{2}$$

In the case of the Wiener model, the non-linearity acts on both output and input of the ARX model. Moreover, the inverse non-linearity acts on past values of the output. Mathematically we have

$$\begin{aligned} y(k) = {} & f[a_1 f^{-1}(y(k-1)) + \cdots + a_{n_y} f^{-1}(y(k-n_y)) \\ & + b_0 u(k-d) + \cdots + b_{n_u} u(k-n_u)] \end{aligned} \tag{3}$$

in which case it has been assumed that f is invertible as in [4, 13].

Once the static function $f(\cdot)$ is known, the identification of Hammerstein models is straightforward. From the knowledge of the static function, the internal signal can be readily produced, thus $v(k) = f(u(k))$. Subsequently, an autoregressive with an exogenous input (ARX) model relating the output $y(k)$ to the internal signal $v(k)$ can be obtained using standard linear techniques.

The static function $f(\cdot)$ between the internal signal and the output is also the static relationship between the input and the output but for a constant factor, which is the gain of the ARX model. If the DC gain of such a model is assumed unity, then in the steady state $\bar{u}(k) = \bar{v}(k)$ and $f(\cdot)$ is in fact the static function between input and output. This function is used to produce a graph of $\bar{y} \times \bar{u} = \bar{v}$ and linear regression techniques can then be applied to obtain a function $\bar{v} = g(\bar{y})$. Of course, $g(\cdot)$ is an estimate of $f^{-1}(\cdot)$ in the range of values considered.

Equation (1) can be expanded as the summation of terms with degrees of non-linearity in the range $[1 \quad \ell]$. Each $(p+m)$th-order term can contain a pth-order factor in $y(k-n_i)$ and an mth-order factor in $u(k-n_i)$ and is multiplied by a coefficient $c_{p,m}(n_1,\ldots,n_m)$ as follows.

$$\begin{aligned} y(k) = {} & \sum_{m=0}^{\ell} \sum_{p=0}^{\ell-m} \sum_{n_1,n_m}^{n_y,n_u} c_{p,m}(n_1,\ldots,n_m) \\ & \times \prod_{i=1}^{p} y(k-n_i) \prod_{i=1}^{m} u(k-n_i) + e(k) \end{aligned} \tag{4}$$

where

$$\sum_{n_1,n_m}^{n_y,n_u} \equiv \sum_{n_1=1}^{n_y} \sum_{n_2=1}^{n_y} \cdots \sum_{n_m=1}^{n_u} \tag{5}$$

and the upper limit is n_y if the summation refers to factors in $y(k-n_i)$ or n_u for factors in $u(k-n_i)$. The moving average part of the model is used during parameter estimation to reduce bias and it is not taken into account in the analysis. The model structure can be chosen using orthogonal techniques [14, 15]. Suppose the model is asymptotically stable and is excited by a constant input, then in the steady state $\bar{y} = y(k-1) = y(k-2) = \cdots = y(k-n_y)$, $\bar{u} = u(k-1) = u(k-2) = \cdots = u(k-n_u)$ and equation (4) can be rewritten as

$$\bar{y} = \sum_{m=0}^{\ell} \sum_{p=0}^{\ell-m} \sum_{n_1,n_m}^{n_y,n_u} c_{p,m}(n_1,\ldots,n_m) \bar{y}^p \bar{u}^m \tag{6}$$

The solution of (6) will yield the fixed points of model (4) for the particular value of the input being used. The next two definitions will be important in the remainder of the paper.

Definition 2.1: [16] The constants $\sum_{n_1,n_m}^{n_y,n_u} c_{p,m}(n_1,\ldots,n_m)$ in (6) are the coefficients of the term clusters $\Omega_{y^p u^m}$, which contain terms of the form $y^p(k-i)u^m(k-j)$ for $m+p \le \ell$, where i and j are any time lags. Such coefficients are called *cluster coefficients* and are represented as $\Sigma_{y^p u^m}$.

Definition 2.2: [17] The set of all possible terms of the form $y^p(k-d)u^m(k-j)$ for $m+p \le \ell$ is called a *d-cluster* and is represented as $\Omega_{y_d^p u^m}$. The sum of all the respective coefficients is referred to as the *d-coefficient* and is represented as $\Sigma_{y_d^p u^m}$.

In words, a term cluster is a set of terms of the same type and the respective cluster coefficient is the summation of the coefficients of all the terms of the corresponding cluster. Hence, terms of the same cluster explain the same type of non-linearity. Moreover, d-clusters are subsets of term clusters that do depend on the lag of the *output* terms. For instance, the terms $y(k-1)u(k-1), y(k-1)u(k-3)$ and $y(k-2)u(k-2)$ all pertain to the term cluster Ω_{yu}. On the other hand, the first two terms are members of the d-cluster $\Omega_{y_1 u}$ whereas the last term pertains to d-cluster $\Omega_{y_2 u}$.

3 Analysis of block-oriented models

3.1 Hammerstein models

In the Hammerstein model, the internal signal is given by

$$v(k-i) = f^{\ell}(u(k-i)), \quad i = 1, 2, \ldots, n_u \tag{7}$$

where f^{ℓ} is the static non-linearity of the model and is assumed to have a polynomial approximation of degree ℓ. In the case where the model has a pure time delay equal to d, then $i = d, d+1, \ldots, n_u$ and the maximum lag in the input variable for this case might be larger than for (7). The linear block is a standard ARX model of the form

$$y(k) = \sum_{j=1}^{n_y} \theta_j y(k-j) + \sum_{i=1}^{n_v} \sigma_i v(k-i) \tag{8}$$

where θ_j and σ_i are model parameters. Substituting (7) in (8) yields

$$y(k) = \sum_{j=1}^{n_y} \theta_j y(k-j) + \sum_{i=1}^{n_u} \sigma_i f^{\ell}(u(k-i)) \tag{9}$$

The local stability of a NARX polynomial model can be assessed around an operating point $(\bar{\boldsymbol{u}}, \bar{y})$ verifying the eigenvalues λ of the jacobian matrix $\boldsymbol{D_f}$ evaluated at the operating point. The eigenvalues are the roots of the equation $\det|\lambda I - \boldsymbol{D_f}| = 0$, that can be expressed in expanded form as

$$\lambda^{n_y} - \Delta_1 \lambda^{n_y - 1} - \cdots - \Delta_{n_y - 1}\lambda - \Delta_{n_y} = 0 \tag{10}$$

where

$$\Delta_1 = \left.\frac{\partial y(k)}{\partial y(k-1)}\right|_{(\bar{\boldsymbol{u}}, \bar{y})}, \ldots, \Delta_{n_y} = \left.\frac{\partial y(k)}{\partial y(k-n_y)}\right|_{(\bar{\boldsymbol{u}}, \bar{y})}$$

It can be readily verified that for (9) $\Delta_i = \theta, i = 1, 2, \ldots, n_y$. In other words, the eigenvalues at a given operating point are the roots of

$$\lambda^{n_y} - \theta_1 \lambda^{n_y - 1} - \cdots - \theta_{n_y - 1}\lambda - \theta_{n_y} = 0. \tag{11}$$

Because (11) is a polynomial with constant coefficients, its roots are also constant. This points to the well-known fact that the dynamics of Hammerstein models do not depend on the operating point. As will be seen in the next Section, this turns out to be an important difference with respect to Wiener models.

3.2 Wiener models

In the Wiener model, the static non-linearity operates on the internal signal to produce the output, that is $y(k) = f^{\ell}(v(k))$. It is assumed that f^{ℓ} has an inverse over the operating range of interest and that such an inverse accepts a polynomial representation of degree ℓ_1 which shall be denoted by g^{ℓ_1}, that is $g^{\ell_1} \approx (\hat{f}^{\ell})^{-1}$ and $v(k) = g^{\ell_1}(y(k))$.

The dynamical model is

$$v(k) = \sum_{j=1}^{n_v} \theta_j v(k-j) + \sum_{i=1}^{n_u} \sigma_i u(k-i) \tag{12}$$

After substitution the following holds

$$y(k) = f^{\ell}\left(\sum_{j=1}^{n_v} \theta_j v(k-j) + \sum_{i=1}^{n_u} \sigma_i u(k-i)\right) \tag{13}$$

or in terms of input and output signals

$$y(k) = f^{\ell}\left(\sum_{j=1}^{n_y} \theta_j g^{\ell_1}(y(k-j)) + \sum_{i=1}^{n_u} \sigma_i u(k-i)\right) \tag{14}$$

The eigenvalues of a Wiener model linearised around an operating point $(\bar{\boldsymbol{u}}, \bar{y})$ are also given by the solutions of (10) with

$$\begin{aligned}
\Delta_i &= \left.\frac{\partial y(k)}{\partial y(k-i)}\right|_{(\bar{\boldsymbol{u}}, \bar{y})} \\
&= \left.\frac{\partial f^{\ell}(x)}{\partial x} \times \frac{\partial x}{\partial y(k-i)}\right|_{(\bar{\boldsymbol{u}}, \bar{y})} \\
&= \left.\frac{\partial f^{\ell}(x)}{\partial x}\right|_{(\bar{\boldsymbol{u}}, \bar{y})} \times \theta_i \left.\frac{\partial g^{\ell_1}(z)}{\partial z}\frac{\partial z}{\partial y(k-i)}\right|_{(\bar{y})} \\
&= \ell f^{\ell-1}(x)|_{(\bar{\boldsymbol{u}}, \bar{y})} \times \theta_i \ell_1 g^{\ell_1 - 1}(z)|_{(\bar{y})} \\
&= \ell f^{\ell-1}(\Sigma_y g^{\ell_1}(\bar{y}) + \Sigma_u \bar{\boldsymbol{u}}) \times \theta_i \ell_1 g^{\ell_1 - 1}(\bar{y})
\end{aligned} \tag{15}$$

where x stands for the argument of $f(\cdot)$ and z for the argument of $g(\cdot)$ as seen in (14), and

$$\Sigma_y = \sum_{j=1}^{n_y} \theta_j, \quad \Sigma_u = \sum_{i=1}^{n_u} \sigma_i \tag{16}$$

As can be seen from (15), the coefficients Δ_i and therefore the model eigenvalues do depend on $(\bar{\boldsymbol{u}}, \bar{y})$, the operating point.

3.3 Connections with NARX polynomials

The static non-linearities in Hammerstein and Wiener models of a same process should be, in principle, the same provided the DC gain of the respective linear blocks are equal. In this respect, there is no loss of generality if such DC gains are taken to be unity [4].

Hammerstein and Wiener models cannot exhibit output multiplicity, i.e. a given fixed input value, $\bar{\boldsymbol{u}}$ will produce one single output in the steady state, $\bar{y}$. Hammerstein models, on the other hand, may display input multiplicity, i.e. more than one input value can yield the same output in the steady state if f^{ℓ} is not invertible. In principle, the same applies to Wiener models in the more general setting $y(k) = f^{\ell}[v(k)]$, but it is often required to write such models as a NARX polynomial (see (3) or (14)), in which case it is required that f^{ℓ} be invertible, and then no input multiplicities are allowed.

It is well known that while Hammerstein and Wiener models have identical steady-state performances, the related dynamics can be quite different. This can be readily explained by observing that the eigenvalues of the Jacobian matrix of a Hammerstein model are constant and therefore do not depend on the operating point, see (11). On the other hand, the eigenvalues of a Wiener model Jacobian matrix depend on the operating point at which the Jacobian is evaluated, as can be seen from (15). This remark should prove helpful in choosing between Hammerstein and Wiener models.

As for the relations of Hammerstein and Wiener models with NARX polynomials the following remarks are helpful.

1. Since the presence of any cluster of the form $\Omega_{y^p u^m}, p > 1, \forall m$ will yield a model with output multiplicity [18], then no NARX polynomial models with clusters of that type can be reduced to Hammerstein or Wiener models.
2. NARX polynomials for which all non-linear clusters are of the form $\Omega_{u^m}, \forall m$ can be written as a Hammerstein model

only if the lags of a given multinomial are the same, as for example in $u(k-1)u(k-1)$.

3. A NARX polynomial with terms from cross-clusters of the type $\Omega_{yu^m}, \forall m$ has eigenvalues that depend on the operating point [19], like in a Wiener model. Thus, NARX models with such clusters would suggest a Wiener model for the system. The terms in such clusters can have different time lags, but in the case of Ω_{u^m} clusters the same restriction on time lags for Hammerstein models applies.

4. The static function of a NARX polynomial with terms from cross-clusters of the type $\Omega_{y^p u^m}, \forall m$ and $p = 1$ is a rational function whereas if $p = 0$ the static function becomes polynomial.

5. Item 3 can be used to help decide between a Hammerstein or a Wiener model, if a NARX polynomial is available. Even if the lags of a NARX polynomial model do not permit direct derivation of Hammerstein and Wiener analogues according to items 2 and 3, if the only difficulty are the lags then good approximate block-oriented models should exist.

4 Results

The main point of this paper is to discuss important dynamical differences between Hammerstein and Wiener models in the context of practical system identification. The main results, therefore, are applicable to any procedure employed to identify such models. Some details on the particular procedure followed in this paper are provided in the Appendix.

For the sake of comparison, the root mean square error (RMSE) was computed for the identified models

$$\text{RMSE} = \frac{\sqrt{\sum_{k=1}^{N}(y(k) - \hat{y}(k))^2}}{\sqrt{\sum_{k=1}^{N}(y(k) - \bar{y})^2}} \tag{17}$$

where $\hat{y}(k)$ is the *free-run* predicted signal, $\bar{y}$ is the average value of the measured signal $y(k)$ over the estimation data window.

4.1 Example 1

This example considers a non-linear first-principle model of a polymerisation reaction taking place in a jacketed continuous stirred-tank reactor (CSTR) [20]. The reaction is the free-radical polymerisation of methyl methacrylate (MMA) with azo-bis-isobutyro-nitrile (AIBN) as initiator and toluene as solvent. The volumetric flow rate of the initiator was considered as the model input, $u(k)$, and the number-average molecular weight as the output $y(k)$. The equations, parameter values and sampling time were taken following [20] to produce the data presented in Fig. 2.

From the first half of the set of data in Fig. 2, the following static function $\bar{y} = f(\bar{u})$ was obtained (see Appendix for details)

$$\begin{aligned}\hat{\bar{v}} = &-0.164 \times 10^6 \bar{u}^3 + 0.947 \times 10^6 \bar{u}^2 \\ &- 1.492 \times 10^6 \bar{u} + 1.390 \times 10^6\end{aligned} \tag{18}$$

where $\bar{y}$ was replaced by $\bar{v}$. It is important to appreciate the reason for this change. For Hammerstein models, the output of such a function is the intermediate signal $v(k) = \hat{f}[u(k)]$. That is the reason for using v as the output of the static model (18). The estimated and original static non-linearities are compared in Fig. 3. It is worth pointing out that the best fit was attained within the range $2.2 \times 10^4 < y(k) < 2.6 \times 10^4$ where 60% of the (dynamical) data are contained. A better fit outside this range could be attained by means of weighted least squares imposing a greater weight outside the range.

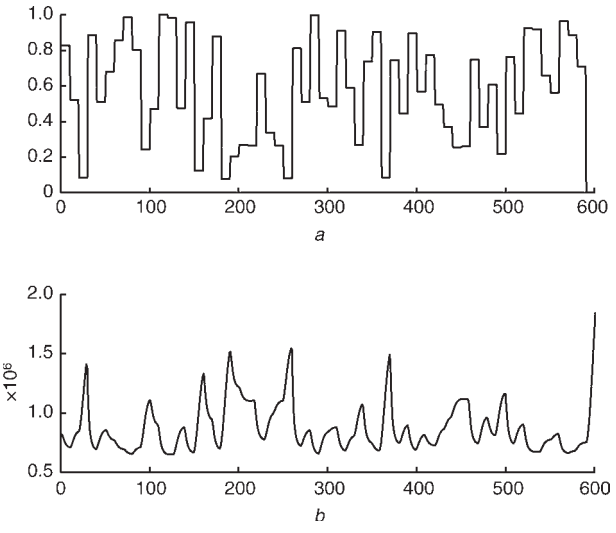

Fig. 2 *Normalised identification data. Both input and output sequences were divided by 0.03006*

a Input
b Output

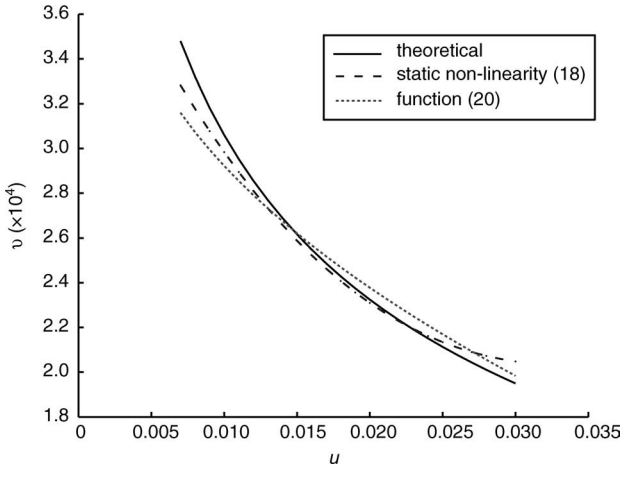

Fig. 3 *Comparison of static functions: theoretical, static non-linearity (18) and function (20) that is used as an inverse function in the context of Wiener modelling. The RMSE values of the curve fits are RMSE = 0.020 and RMSE = 0.030, respectively*

Alternatively, the auxiliary model could be made more non-linear not only by increasing the degree of non-linearity of polynomial (18) but by choosing a different representation, such as exponential model or a network model rather than a polynomial.

Hence, using (18) and input data $u(k)$ the intermediate signal $\hat{v}(k)$ can be produced. Such a signal is the input to the linear dynamical model. Then, with the data set $Z^N = [\,\hat{v}(k) \quad y(k)\,]$ (first half of the set of data in Fig. 2) and standard techniques [21] the ARX part of Hammerstein model was obtained

$$\begin{aligned}y(k) = &\, 2.248y(k-1) - 1.800y(k-2) \\ &+ 0.5098y(k-3) + 0.01721\hat{v}(k-2) \\ &+ 0.01444\hat{v}(k-3) + 0.01074\hat{v}(k-1)\end{aligned} \tag{19}$$

where the order $n_y = 3$ was selected by pole-zero analysis [22]. The free-run-prediction of this model is compared to validation data in Fig. 4.

To obtain a Wiener model, a function $\bar{v} = g^{\ell_1}(\bar{y})$ must be estimated which is an inverse of (18). Hence using (18) to

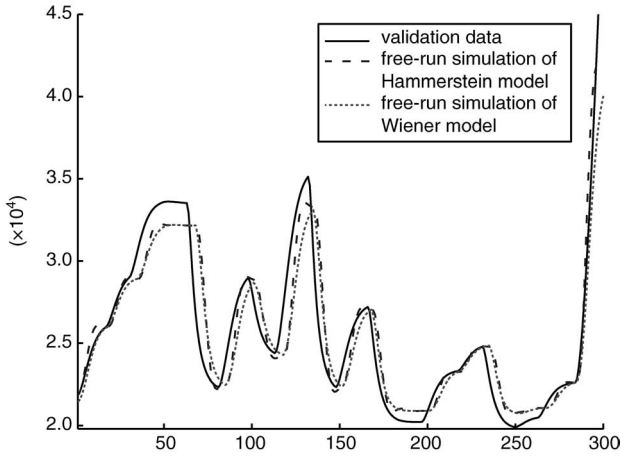

Fig. 4 *Validation data and free-run simulation of Hammerstein model (18) and (19) for which RMSE = 0.0354; and free-run simulation of Wiener model (18) and (21) for which RMSE = 0.0387*

produce data and performing the change of variables $\bar{v} \rightarrow \bar{y}, \bar{u} \rightarrow \bar{v}$, and employing standard numerical regression techniques, the following equation was obtained (see Fig. 3):

$$\hat{\bar{v}} = 2.23 \times 10^{-12}\bar{y}^2 - 5.77 \times 10^{-6}\bar{y} + 3.834 \qquad (20)$$

From input data $u(k)$ and intermediate variable $v(k)$ obtained passing the output signal $y(k)$ through (20), the following ARX model was determined

$$\begin{aligned}\hat{v}(k) = {} & 2.187\hat{v}(k-1) - 1.622v(k-2) \\ & + 0.4098\hat{v}(k-3) + 0.03179u(k-1) \\ & - 0.8892 \times 10^{-2}u(k-2) + 0.215 \times 10^{-2}u(k-3)\end{aligned} \qquad (21)$$

where the order of this model was selected as for the Hammerstein case. The result of the free-run-prediction of the Wiener model is presented in Fig. 4.

4.2 Example 2

The set-up used in this example consists of a small electrical heater. The temperature of the heater is the output, $y(k)$, and was measured by a thermocouple. The sampling time was $T_s = 6$s. Both steady-state and dynamical tests were performed but only the data measured during the dynamical test, shown in Fig. 5, were used in the identification. It is worth pointing out that both the static non-linear and the dynamical linear blocks are estimated exclusively from these data.

The steady-state relationship between the electrical power, which is the input $u(k)$, and the final temperature does not present two possible states for the same input. This, as discussed in the Appendix, means that all clusters of the form $\Omega_{y^p u^m}$ for $p > 1$ irrespective of m should be left out of the model. The choice of which clusters to include in the model that will be used for static function estimation is not critical as long as the main guidelines discussed in Section 3.3 and in the Appendix are taken into account. One possible approximation to the static function is simply

$$\bar{v} = 0.0566\bar{u}^2 \qquad (22)$$

and is compared to the measured one in Fig. 6. It is pointed out that the actual static non-linearity shown in Fig. 6 was measured only for the sake of comparison and that (22) was estimated from the set of dynamical data.

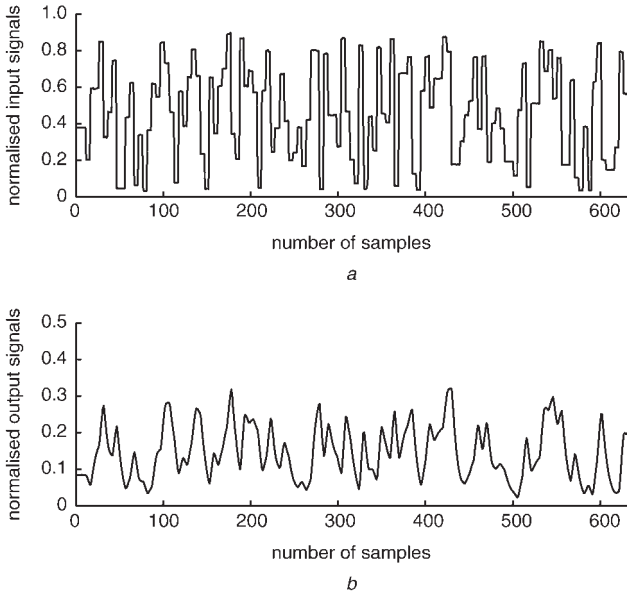

Fig. 5 *First part of the dynamical test used as identification data. The second part was used for validation. Both model blocks are estimated from these measured data*

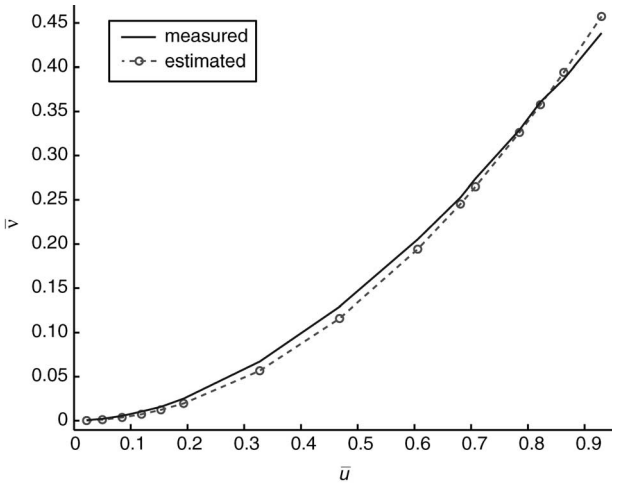

Fig. 6 *Static functions: measured (only used for the sake of comparison, not for identification) and estimated, RMSE = 0.054, see (22)*

Proceeding as in the previous example, the following model was obtained

$$\begin{aligned}y(k) = {} & 1.1729y(k-1) + 0.0887\hat{v}(k-1) \\ & - 0.2750y(k-2) + 0.0127\hat{v}(k-2)\end{aligned} \qquad (23)$$

The free-run simulation of model (23) with (22) is compared to the measured data in Fig. 7. In the case of a Wiener model, the static function is the same and (22) can be rewritten as

$$\bar{y} = 0.0566\bar{v}^2 \qquad (24)$$

Clearly, this expression cannot be used analytically to obtain the intermediary signal $v(k)$ from measured data. To work around this problem, $\bar{v}$ is varied in the same range of the input, $u(k)$, i.e. $0 \leq \bar{v} \leq 1$ and the corresponding values of $\bar{y}$ are calculated by $\bar{y} = 0.0566\bar{v}^2$. The next step is to use standard numerical regression techniques to fit a function $\bar{v} = g^{\ell_1}(\bar{y})$, as was done in Example 4.1. For the present example such a function is

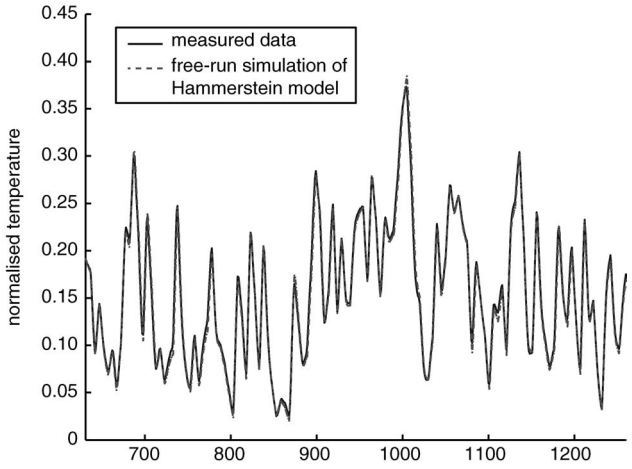

Fig. 7 *Validation data set: measured data set and free-run simulation of Hammerstein model composed by (22) and (23). RMSE = 0.093*

$$\hat{\bar{v}} = 21501.422\bar{y}^7 - 34353.954\bar{y}^6 + 22160.557\bar{y}^5 - 7405.473\bar{y}^4 + 1373.354\bar{y}^3 - 143.372\bar{y}^2 + 10.034\bar{y} + 0.040 \quad (25)$$

where the degree of such a polynomial was chosen based on the standard correlation index. The performance of this regression model is shown in Fig. 8.

As before, using $Z^N = [u(k)\ \hat{v}(k)]$ the following model was found:

$$\hat{v}(k) = 2.2823\hat{v}(k-1) + 0.0702u(k-1) - 1.6677\hat{v}(k-2) + 0.3838\hat{v}(k-3) - 0.0684u(k-2) \quad (26)$$

The measured output and the free-run simulation of the Wiener model (24) and (26) are compared in Fig. 9.

The Hammerstein model clearly outperforms the Wiener model. This was somewhat expected since previous results revealed that this system has eigenvalues that are not sensitive to the operating point [17, 19]. Then, as mentioned in Section 3.3, the Hammerstein model is more adequate to represent this kind of system.

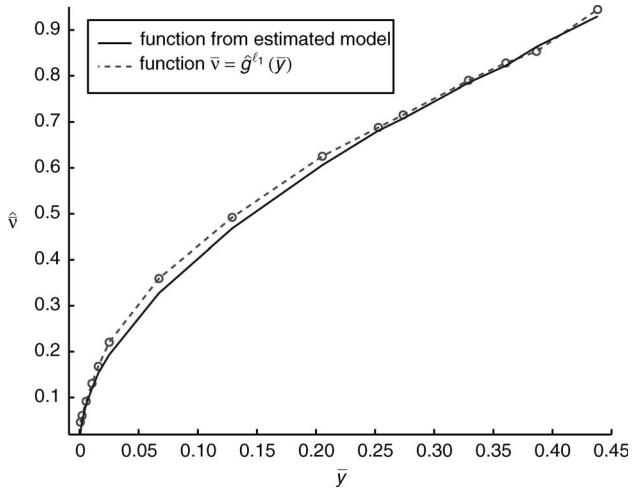

Fig. 8 *Plot of inverse static non-linearity and its estimated model $\hat{\bar{v}} = \hat{g}^{\ell_1}(\bar{y})$. The function marked (—) was taken from the estimated model (24) and plotted with the dependent variable on the x-axis (indicated by $\bar{y}$ in the figure). The function $\hat{\bar{v}} = \hat{g}^{\ell_1}(\bar{y})$, see (25), is estimated by standard curve fitting techniques, RMSE = 0.056*

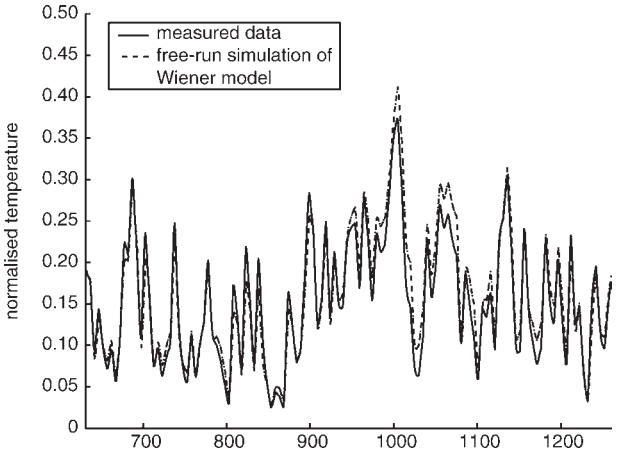

Fig. 9 *Validation data set: measured data and free-run simulation of Wiener model composed by (24) and (26). RMSE = 0.2819*

4.3 Example 3

To emphasize the last result in Example 2, we considered another lab thermal process that has eigenvalues that do depend on the operating point according to the following equation [17]:

$$\lambda(\bar{u}) = \frac{0.9122 - 0.6029 \times 10^{-2}\bar{u}}{1 - 0.493 \times 10^{-3}\bar{u}} \quad (27)$$

The best estimated Hammerstein model (in terms of RMSE) is

$$\hat{\bar{v}} = \frac{0.6185\bar{u}}{0.0878 + 5.5353 \times 10^{-3}\bar{u}} \quad (28)$$

$$y(k) = -1.307y(k-1) + 0.4247y(k-2) + 0.1772\hat{v}(k-1) - 0.05896\hat{v}(k-2) \quad (29)$$

The best estimated Wiener model (in terms of RMSE) is

$$\hat{v}(k) = 0.7621\hat{v}(k-1) + 0.4247\hat{v}(k-2) + 0.2317u(k-1) + 0.007258u(k-2), \quad (30)$$

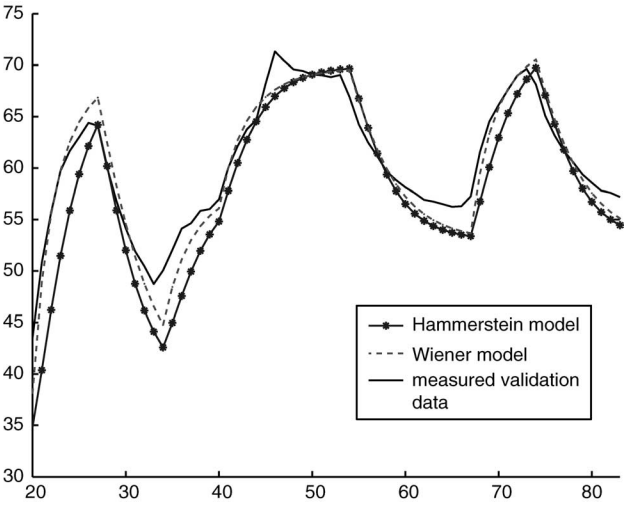

Fig. 10 *Free-run-prediction of Hammerstein model (28) (RMSE = 0.176), Wiener model (30) (RMSE = 0.080) and measured validation data from a thermal process system whose eigenvalues depend on the operating point. The peak occurring at sample 46 is an artifact*

and (28) with $\bar{v}$ replacing $\bar{u}$ and $\bar{y}$ replacing $\bar{v}$. The free-run-predictions of both models are presented in Fig. 10 and reveal that the Wiener model is superior to its Hammerstein counterpart. This was expected since, based on previous analyses, it is known that the eigenvalues of this process do vary with the operating point, see (27).

5 Discussion and conclusions

Example 1 illustrates an interesting point. On the one hand, the dynamical performance of both the Hammerstein and Wiener models is very similar as can be verified from Fig. 4. This would lead to the conclusion that the process does not have eigenvalues that vary with the operating point. If this were the case, the Wiener model would be expected to outperform the Hammerstein counterpart. On the other hand, however, it is crucial to see that the quality of fit of the inverse function (20), used in the context of Wiener modelling, is 50% worse than for the static function (18). That means that in the case of Wiener modelling the intermediate signal is worse than for Hammerstein modelling because the former will use (20) and the latter (18) to produce such signals. Nonetheless, the overall performance is somewhat the same. This means that the Wiener model did compensate for the loss in using (20). In fact, using prior knowledge to fit static functions $f(\cdot)$ and $g^{\ell_1}(\cdot)$ of comparable accuracy and subsequently using such functions to obtain new block oriented models, the identified Wiener model has an RMSE which is 11% smaller than that of the Hammerstein counterpart. It should also be mentioned that even if the structure of the linear dynamic blocks of the Hammerstein and Wiener models are not constrained to have the same structure, the best Wiener model still outperforms the best Hammerstein counterpart for this process. All this evidence together with the results developed in Section 3 would suggest that the process does have eigenvalues that vary with the operating point. In fact, a careful examination of the Jacobian matrix of the model used [20] will confirm such a hypothesis.

Concerning Example 2, it is important to notice that whereas the static function is the same and that the accuracy of the estimated inverse (25), see Fig. 8, is comparable to the estimated static function (24), see Fig. 6, the performance of the Hammerstein model is clearly better. This confirms, as widely accepted, that both representations are not completely equivalent.

Thus, based on the results in Section 2 and on the fact that the Hammerstein model performed better than the Wiener model in Example 2, it can be concluded that for this system the pole locations do not vary with the operating point. Indeed, this fact has been recently established by independent means [19].

Finally, in Example 3 the Wiener model clearly outperformed the Hammerstein counterpart, hence indicating a process with eigenvalues that vary with the operating point. For such a system this had been previously verified [17], hence confirming that the eigenvalues do vary with the operating point. One might ask how can this be, given that the poles of the Wiener model are clearly constant, since (26) does not change. The point to notice is that the dynamic output of the dynamical block in the case of Wiener models after passing through the static non-linearity will give the overall impression of varying dynamics, as can be seen from (10) and (15).

An important aspect of the present paper was to establish connections between block-oriented (Hammerstein and Wiener) models and NARX models. One of the advantages of this analysis is that it helps understand and interpret some of the main differences between the Hammerstein and the Wiener representation. Also, such analysis enables one to objectively use any available NARX polynomial model or even the clusters present in the model to make a dynamically-based choice between Hammerstein and Wiener representations. It has been shown that in cases for which the eigenvalues of the process vary with the operating point, the Wiener representation should be preferred. On the other hand, Hammerstein models usually outperform Wiener models in fixed-dynamic processes. This interpretation of the dynamical differences between such models seems to be a helpful piece of information in practice. This result and the whole procedure has been confirmed and illustrated by means of three numerical examples.

6 References

1 Fruzzetti, K.P., Palazoglu, A., and McDonald, K.: 'Nonlinear model predictive control using Hammerstein models', *J. Process Control*, 1997, **7**, (1), pp. 31–41
2 Lee, J., Cho, W., and Edgar, T.F.: 'Control system design based on a nonlinear first-order plus time delay model', *J. Process Control*, 1997, **7**, (1), pp. 65–73
3 Billings, S.A., and Fakhouri, S.Y.: 'Identification of systems containing linear dynamic and static nonlinear elements', *Automatica*, 1982, **12**, (1), pp. 15–26
4 Pearson, R.K., and Pottmann, M.: 'Gray-box identification of block-oriented nonlinear models', *J. Process Control*, 2000, **10**, pp. 301–315
5 Schröder, D., Hintz, C., and Rau, M.: 'Intelligent modeling, observation, and control for nonlinear systems', *IEEE Trans. Mechatronics*, 2001, **6**, (2), pp. 122–131
6 Bai, E.W.: 'A blind approach to the Hammerstein-Wiener model identification', *Automatica*, 2002, **38**, pp. 967–979
7 Abonyi, J., Babuska, R., Ayala-Botto, A., Szeifert, F., and Nagy, L.: 'Identification and control of nonlinear systems using fuzzy Hammerstein models', *Ind. Eng. Chem. Res.*, 2000, **39**, pp. 4302–4314
8 Celka, P., Bershad, N.J., and Vesin, J.: 'Stochastic gradient identification of polynomial Wiener systems: Analysis and application', *IEEE Trans. Signal Process.*, 2001, **49**, (2), pp. 301–313
9 Haber, R., and Unbehauen, H.: 'Structure identification of nonlinear dynamic systems — a survey on input/output approaches', *Automatica*, 1990, **26**, (4), pp. 651–677
10 Chen, H.W.: 'Modeling and identification of parallel nonlinear systems: Structural classification and parameter estimation', *Proc. IEEE*, 1995, **83**, (1), pp. 39–66
11 Palumbo, P., and Piroddi, L.: 'Seismic behavious of buttress dams: Non-linear modelling of a damaged buttress based on ARX/NARX models', *J. Sound Vib.*, 2001, **239**, (3), pp. 405–422
12 Chiras, N., Evans, C., and Rees, D.: 'Nonlinear gas turbine modeling using NARMAX structures', *IEEE Trans. Instrum. Meas.*, 2001, **50**, (4), pp. 893–898
13 Billings, S.A., and Leontaritis, I.J.: 'Identification of nonlinear systems using parametric estimation techniques'. Proc. IEE Conf. Control and its Applications, Warwick, 1981, pp. 183–187
14 Korenberg, M.J., Billings, S.A., Liu, Y.P., and McIlroy, P.J.: 'Orthogonal parameter estimation algorithm for nonlinear stochastic systems', *Int. J. Control*, 1988, **48**, (1), pp. 193–210
15 Kortmann, M., and Unbenhauen, H.: 'Structure detection in the identification of nonlinear systems', *Trait. Signal*, 1988, **22**, (5), pp. 5–25
16 Aguirre, L.A., and Billings, S.A.: 'Improved structure selection for nonlinear models based on term clustering', *Int. J. Control*, 1995, **62**, (3), pp. 569–587
17 Corrêa, M.V.: 'Grey-box identification of nonlinear systems using rational and polynomial model representations'. PhD Thesis, PPGEE, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, 2001, (In Portoguese)
18 Aguirre, L.A., and Mendes, E.M.: 'Nonlinear polynomial models: Structure, term clusters and fixed points', *Int. J. Bifurcation Chaos*, 1996, **6**, (2), pp. 279–294
19 Aguirre, L.A., Corrêa, M., and Cassini, C.C.S.: 'Nonlinearities in NARX polynomial models: Representation and estimation', *IEE Proc. D, Control Theory Appl.*, 2002, **149**, (4), pp. 343–348
20 Doyle, F.J., III, Ogunnaike, B.A., and Pearson, R.K.: 'Nonlinear model-based control using second-order Volterra models', *Automatica*, 1995, **31**, (5), pp. 697–714
21 Ljung, L.: 'System identification, theory for the user' (Prentice Hall, New Jersey, 1999, 2nd edn.)
22 van den Boom, A.J.W., and van den Enden, A.W.M.: 'The determination of the orders of process and noise dynamics', *Automatica*, 1974, **10**, pp. 245–256
23 Aguirre, L.A., and Souza, A.V.P.: 'An algorithm for estimating fixed points of dynamical systems from time series', *Int. J. Bifurcation Chaos*, 1998, **8**, (11), pp. 2203–2213

24 Visala, A., Pitkänen, H., and Halme, A.: 'Modeling of chromatographic separation process with Wiener-MLP representation', *J. Process Control*, 2001, **11**, pp. 443–458
25 Wright, R.A., and Kravaris, C.: 'On-line identification and nonlinear control of an industrial pH process', *J. Process Control*, 2001, **11**, pp. 361–374
26 Mendes, E.M.A.M., and Billings, S.A.: 'An alternative solution to the model structure selection problem', *IEEE Trans. Man Cyber. - Part A*, 2001, **36**, (21), pp. 597–608
27 Piroddi, L., and Spinelli, W.: 'An identification algorithm for polynomial NARX models based on simulation error minimization', *Int. J. Control*, 2003, **76**, (17), pp. 1767–1781

7 Appendix

The procedure used in the examples is a byproduct of this paper and follows from (4) and (6). In words, a general NARX polynomial of the form (4) is estimated from dynamical data and subsequently the static function is retrieved from conventional steady-state analysis of the estimated model (6).

A few practical remarks are in order.

1. A closed-form expression can be derived to directly compute the static function from the cluster coefficients of a NARX polynomial [19]:

$$f(\cdot) = \frac{\bar{y}}{\bar{u}} = \frac{\Sigma_0/\bar{u} + \Sigma_u + \sum_{m=2}^{\ell} \Sigma_{u^m} \bar{u}^{m-1}}{1 - \Sigma_y - \sum_{m=1}^{\ell-1} \sum_{p=1}^{\ell-m} \Sigma_{y^p u^m} \bar{y}^{p-1} \bar{u}^m - \sum_{p=2}^{\ell} \Sigma_{y^p} \bar{y}^{p-1}} \quad (31)$$

2. In the present step the only concern is with the steady-sate performance. Consequently the structure of the model used is not critical. Nonetheless, (31) shows that not every possible cluster should be included in this model. For instance, clusters $\Omega_{y^p u^m}, p > 1, \forall m$ should not be included to avoid output multiplicity and clusters of the type $\Omega_{yu^m}, \forall m$ should be left out if a polynomial form for $f(\cdot)$ is desired. With this in mind, two or three terms of each valid cluster are initially included in this auxiliary model. Cluster analysis can be performed to see whether a certain cluster can be discarded or not. This procedure has proved successful in estimating fixed-points from dynamical data [23]. This model is auxiliary in the sense that it will be only used to obtain an estimate of the static non-linearity $\hat{f}(\cdot)$.

3. Because of the way it is estimated, the static function is quite robust to noise. In order to illustrate that, consider Fig. 11 where it can be seen that the estimated static function does not deviate greatly even for relatively low values of the signal-to-noise ratio.

4. Since the static function is only required to map the measured signals $u(k)$ and $y(k)$ to the intermediate variable $v(k)$, it is not necessary that $f(\cdot)$ be estimated as a polynomial. In fact, other mathematical structures such as neural networks could be used to estimate $f(\cdot)$ [24] or even fuzzy models [7] with the additional advantage of being able to cope with stronger non-polynomial non-linearities. Moreover, if basis-functions that are more adequate to represent the steady state of a certain process are known [25] such functions can be used. Strictly speaking, this would render the procedure mildly grey-box because it would use some prior knowledge. However, the key point in the suggested procedure is that such an estimate can be achieved from dynamical data, so long as the range of $u(k)$ is sufficiently wide to reveal the non-linearity of the process. If this is not the case, the steady-state behaviour of the process should be estimated by conventional means.

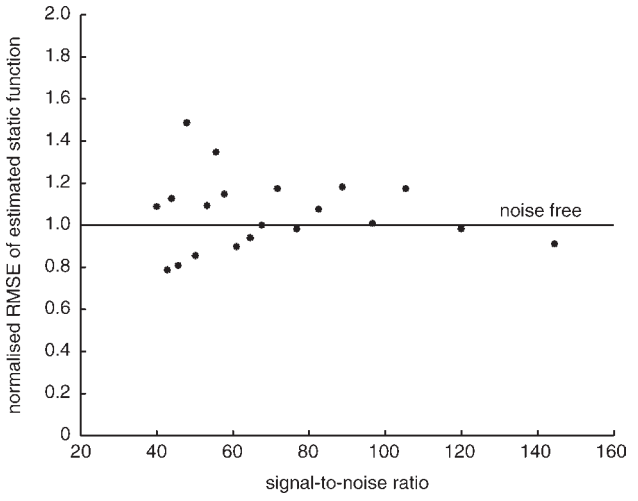

Fig. 11 *Normalised RMSE of the estimated static function. Normalisation was performed with respect to the noise-free value RMSE = 0.0236. The signal-to-noise ratio was computed as 20 times the logarithm of the ratio of signal variance over noise variance*

The procedure followed consists of two steps. First, the static non-linearity is estimated from the dynamical data. In this paper, this has been accomplished by means of an auxiliary model with polynomial structure for which the static function can be obtained analytically from the auxiliary model. However, such a model can be of any other type, as for instance a network, but in such cases the static function might only be accessible via simulation and not analytically. Of course, if a NARX polynomial model is already available then the static non-linearity can be directly extracted from such a model. On the other hand, the use of the auxiliary model does not require sophisticated structure selection techniques usually required for building parsimonious NARX polynomial models [26, 27]. The second step is a standard linear identification exercise that uses a sequence of measured data either input (for Wiener models) or output (for Hammerstein models) and a sequence of synthetic data obtained using the static non-linear function estimated in the first step.

This procedure is therefore parametric, does not require prior knowledge, input and output signals are assumed to be sampled at the same rate (a common situation in the process industry) and the static non-linearity is not restricted to a polynomial type (actually, polynomial and rational functions are possible).