Preprints of the
15th IFAC Symposium on System Identification
Saint-Malo, France, July 6-8, 2009

Identification of a Wiener-Hammerstein System Using the Polynomial Nonlinear State Space Approach

J. Paduart, L. Lauwers, R. Pintelon, and J. Schoukens

Vrije Universiteit Brussel, dept. ELEC, Pleinlaan 2, B-1050 Brussels, Belgium
e-mail: jpaduart@gmail.com

Abstract: In this paper, we apply the Polynomial NonLinear State Space (PNLSS) approach to model a nonlinear system with a Wiener-Hammerstein structure. To obtain good initial estimates, the best linear approximation of the system under test is first identified. Next, this linear model is extended to a polynomial nonlinear state space model to capture also the system's nonlinear behaviour. The identification procedure is applied to measurement data.
Keywords: System Identification, Wiener-Hammerstein Systems, Best Linear Approximation, State space models.

1. Introduction

Most real-life systems can be modelled quite well with a linear model, but even better results can be obtained using a nonlinear model. System identification methods (Söderström & Stoica, 1989; Ljung, 1999; Pintelon & Schoukens, 2001) allow to build models for nonlinear devices. An excellent starting point for nonlinear modelling is Sjöberg et al. (1995).

When considering the Device Under Test (DUT) as a black box, no information about the device's internal structure is utilized in the modelling process. The only available information about the system is given by its measured input(s) and output(s). Hence, black box modelling implies the use of a model structure that is as flexible as possible. Often, this flexibility results in a high number of parameters.

Numerous black box model structures are available when modelling a nonlinear system. Most of them are dedicated to Single Input, Single Output (SISO) systems. In order to cope with Multiple Input, Multiple Output (MIMO) systems, a suitable model structure is necessary. For such systems, it is important that the common dynamics, present in the different outputs of the DUT, are exploited in such a way that the number of model parameters is small. A model structure that fulfils this requirement in a natural way is the state space representation.

In this paper, an identification procedure is applied which uses polynomial nonlinear state space equations for the modelling of nonlinear systems with a Wiener-Hammerstein structure. First, a linear state space model is fit on the measured data, starting from the Best Linear Approximation of the DUT. Next, this linear model is extended to a polynomial nonlinear state space model which is optimized to capture also the system's nonlinear behaviour.

The structure of the paper is the following: First, the Best Linear Approximation of a nonlinear system is defined. Then, the state space framework and the Polynomial Nonlinear State Space model are presented. Next, the identification procedure is explained in detail. Finally, this identification method is applied to measurement data from a physical system.

2. Best Linear Approximation

The Best Linear Approximation of a nonlinear system is defined as the model that minimizes the mean square error between the true output and the modelled output for a given class of input signals, i.e.,

$$G_{\text{BLA}} = \arg\min_{G} E\{[y(t) - G(u(t))]^2\} \tag{1}$$

with $u(t)$ and $y(t)$ the input and output, respectively. The Best Linear Approximation is calculated as the solution of (1) by performing classical Frequency Response Function (FRF) measurements:

$$G_{\text{BLA}}(j\omega) = \frac{S_{yu}(j\omega)}{S_{uu}(j\omega)}, \tag{2}$$

where $S_{uu}(j\omega)$ is the auto-power spectrum of the input, and $S_{yu}(j\omega)$ the cross-power spectrum between the output and the input.

A nonlinear system can be modelled as the sum of a linear system $G_{\text{BLA}}(j\omega)$ and a noise source y_s. This noise source represents the unmodelled nonlinear contributions of the system. In other words, it represents that part of the output y that cannot be captured by the linear model $G_{\text{BLA}}(j\omega)$ (Schoukens et al., 2005; Enqvist & Ljung, 2005). These nonlinear contributions depend on the particular input realization, and for a stochastic input signal they exhibit a stochastic behaviour from realization to realization. In practice, the Best Linear Approximation can be estimated by averaging the measured FRFs for different input realizations.

Since G_{BLA} depends on the probability density function and the power spectral density of the input signal (Pintelon & Schoukens, 2001; Enqvist, 2005; Enqvist & Ljung, 2005), we restrict ourselves to the extended class of Gaussian excitation signals with a user-defined power spectrum. This class of Gaussian(-like) signals includes Gaussian noise and random phase multisines.

3. Nonlinear State Space Framework

The most natural way to represent systems with multiple inputs and outputs is to use the state space framework. An n_a-th order discrete-time state space model is generally expressed as

$$\begin{cases} x(t+1) = f(x(t), u(t)) \\ y(t) = g(x(t), u(t)) \end{cases} \quad (3)$$

with $u(t) \in \mathbb{R}^{n_u}$ the vector containing the n_u input values at discrete time instance t, and $y(t) \in \mathbb{R}^{n_y}$ the vector of the n_y outputs. The state vector $x(t) \in \mathbb{R}^{n_a}$ represents the memory of the system and includes the common dynamics present in the different outputs. The first equation of (3) is referred to as the state equation. It describes the evolution of the state as a function of the input and the previous state. The second equation of (3) is called the output equation. It relates the system output with the state and the input.

The state space representation is not unique. By means of a similarity transform, the model equations in (3) can be converted into a new model that exhibits exactly the same input/output behaviour. The similarity transform $x_T(t) = T^{-1}x(t)$ with an arbitrary non-singular square matrix T yields

$$\begin{cases} x_T(t+1) = T^{-1}f(Tx_T(t), u(t)) = f_T(x_T(t), u(t)) \\ y(t) = g(Tx_T(t), u(t)) = g_T(x_T(t), u(t)) \end{cases} \quad (4)$$

4. Polynomial Nonlinear State Space Model

Consider the general model in (3) and apply a functional expansion of the functions $f(.)$ and $g(.)$. In principle, various kinds of basis functions can be used for this purpose. We opted for a polynomial approach, resulting in the Polynomial NonLinear State Space (PNLSS) model which is defined as

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) + E\zeta(t) \\ y(t) = Cx(t) + Du(t) + F\eta(t) \end{cases} \quad (5)$$

The coefficients of the linear terms in $x(t)$ and $u(t)$ are given by the matrices $A \in \mathbb{R}^{n_a \times n_a}$ and $B \in \mathbb{R}^{n_a \times n_u}$ in the state equation, and $C \in \mathbb{R}^{n_y \times n_a}$ and $D \in \mathbb{R}^{n_y \times n_u}$ in the output equation. The vectors $\zeta(t) \in \mathbb{R}^{n_\zeta}$ and $\eta(t) \in \mathbb{R}^{n_\eta}$ contain nonlinear monomials in $x(t)$ and $u(t)$ of degree two up to a chosen degree p. The coefficients associated with these nonlinear terms are given by the matrices $E \in \mathbb{R}^{n_a \times n_\zeta}$ and $F \in \mathbb{R}^{n_y \times n_\eta}$. Note that the monomials of degree one are included in the linear part of the PNLSS model structure. The separation in a linear and a nonlinear part is of no importance for the behaviour of the model. However, this distinction will turn out to be very practical, since the first step of the identification procedure consists in estimating a linear model.

When a full polynomial expansion of (3) is carried out, all monomials up to degree p must be taken into account. First, $\xi(t)$ is defined as the concatenation of the state vector and the input vector:

$$\xi(t) = \left[x_1(t) \; \dots \; x_{n_a}(t) \; u_1(t) \; \dots \; u_{n_u}(t)\right]^T \quad (6)$$

As a consequence, the dimension of the vector $\xi(t)$ is given by $n = n_a + n_u$. Then, using the notation explained in Appendix A., $\zeta(t)$ and $\eta(t)$ in (5) are defined as

$$\zeta(t) = \eta(t) = \xi(t)_{\{p\}} \quad (7)$$

This corresponds to considering all the distinct nonlinear combinations of degree p, which is the default choice for the PNLSS model structure. The total number of parameters required by the model in (5) is given by

$$\left(\binom{n+p}{p} - 1\right)(n_a + n_y) = \left(\binom{n_a+n_u+p}{p} - 1\right)(n_a + n_y) \quad (8)$$

For high model orders, the number of parameters can become quite large; this is the so-called curse of dimensionality. If the size of the data set does not justify many parameters, the size of the vectors $\zeta(t)$ and $\eta(t)$ should be restricted. This can for instance be done by eliminating certain combinations, or by restricting them to a specific model class such as state affine models (see Section 6.C).

5. Identification Procedure

The identification procedure for the PNLSS model in (5) consists of three major steps. First, G_{BLA} of the DUT is determined nonparametrically in mean square sense. Then, a parametric linear model is estimated from the Best Linear Approximation using frequency domain subspace identification methods (McKelvey et al., 1996; Pintelon, 2002). This is followed by a nonlinear optimization of the linear model. The last step consists in estimating the full nonlinear model by using again a nonlinear search routine.

A. Best Linear Approximation

From the measured input/output data, the Best Linear Approximation $\hat{G}_{BLA}$ and its sample covariance $\hat{C}_G$ can be determined by performing classical FRF measurements. Explicit expressions for these quantities can be found in Paduart (2008), both for periodic and non periodic data. The reduction of the input/output data to a compact FRF and covariance form offers a number of advantages. First of all, the Signal-to-Noise Ratio (SNR) is enhanced. Secondly, it allows the user to select, in a straightforward way, a frequency band of interest. Finally, when periodic data are available, the measurement noise and the effect of the nonlinear behaviour can be separated (Schoukens et al., 1998).

B. Frequency Domain Subspace Identification

Next, the nonparametric estimate $\hat{G}_{BLA}(j\omega_k)$ is transformed into a parametric model. The goal is to estimate a linear, discrete-time state space model, taking into account the sample covariance matrix $\hat{C}_G(j\omega_k)$. The state space matrices ($ABCD$) can be retrieved by using the frequency domain subspace identification algorithm described in McKelvey et al. (1996) and by relying on the results presented in Pintelon (2002). In the latter, the stochastic properties of this algorithm are analysed for the case in which the sample covariance matrix is employed instead of the true covariance matrix, as is done in this paper.

C. Nonlinear Optimization of the Linear Model

The Weighted Least Squares (WLS) cost function V_{WLS} is defined as (McKelvey et al., 1996):

$$V_{WLS} = \sum_{k=1}^{F} \varepsilon^H(k) C_G^{-1}(k) \varepsilon(k), \quad (9)$$

where

$$\varepsilon(k, \theta) = \text{vec}(G_{SS}(A, B, C, D, j\omega_k) - G(j\omega_k)), \quad (10)$$

$$G_{SS}(A, B, C, D, j\omega_k) = C(z_k I_{n_a} - A)^{-1}B + D, \quad (11)$$

with $z_k = e^{j2\pi(k/N)}$, I_{n_a} the identity matrix of size n_a, and θ the vector containing all entries of A, B, C and D. This cost function V_{WLS} is a measure of the model quality. According to this measure, the frequency domain subspace algorithm generates reasonable model estimates. However, in practical applications V_{WLS} strongly depends on the dimension parameter r chosen in the identification procedure. A first action that is taken to improve the model estimates, is to apply the subspace algorithm for different values of r, for instance $r = n_a + 1, \ldots, 6n_a$, and to select the model that corresponds to the lowest V_{WLS}. Secondly, V_{WLS} is minimized with respect to all the linear parameters ($ABCD$). This nonlinear problem is solved using the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963). It requires the computation of the Jacobian of the model error $\varepsilon(k, \theta)$ with respect to the model parameters. From (10) and (11), the following expressions are calculated:

$$\begin{cases} \dfrac{\partial\varepsilon(k,\theta)}{\partial A_{ij}} = \text{vec}(C(z_k I_{n_a} - A)^{-1} I_{ij}^{n_a \times n_a}(z_k I_{n_a} - A)^{-1} B) \\ \dfrac{\partial\varepsilon(k,\theta)}{\partial B_{ij}} = \text{vec}(C(z_k I_{n_a} - A)^{-1} I_{ij}^{n_a \times n_u}) \\ \dfrac{\partial\varepsilon(k,\theta)}{\partial C_{ij}} = \text{vec}(I_{ij}^{n_y \times n_a}(z_k I_{n_a} - A)^{-1} B) \\ \dfrac{\partial\varepsilon(k,\theta)}{\partial D_{ij}} = \text{vec}\left(I_{ij}^{n_y \times n_u}\right) \end{cases} \tag{12}$$

where $I_{ij}^{m \times n} \in \mathbb{R}^{m \times n}$ is defined as a zero matrix with a single element equal to one at entry (i, j). Note that special measures need to be taken during the optimization since the state space representation is overparametrized (see Section H.). The subspace method in McKelvey et al. (1996) is used to generate a number of different linear models (as a function of r) which are then used as starting values for the nonlinear optimization procedure. In this way, there is a higher probability to end up in a global minimum of V_{WLS}, or at least in a good local minimum. Finally, the model that corresponds to the lowest cost function V_{WLS} is selected.

The estimation procedure is carried out for different model orders n_a. Next, a model selection criterion such as AIC (Akaike, 1974) or MDL (Rissanen, 1978) can be used to choose the best linear model.

D. Estimation of the Full Nonlinear Model

The last step in the identification process is to estimate the full nonlinear model

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) + E\zeta(t) \\ y(t) = Cx(t) + Du(t) + F\eta(t) + e(t) \end{cases} \tag{13}$$

with the initial state given by $x(1) = x_0$ (see Section F.), and where $e(t)$ is the output noise. In order to keep the parameter estimates unbiased, the input $u(t)$ is assumed to be noiseless, i.e., it is observed without any errors and independent of the output noise. The last step in the identification process is carried out in the frequency domain. This has the advantage that the noise information can be integrated nonparametrically in the cost function.

Consider the weighted least squares cost function

$$V_{\text{WLS}}(\theta) = \sum_{k=1}^{F} \varepsilon^H(k,\theta) W(k) \varepsilon(k,\theta), \tag{14}$$

with $W(k) \in \mathbb{C}^{n_y \times n_y}$ a user-chosen, frequency domain weighting matrix. Typically, this matrix is chosen equal to the inverse covariance matrix of the output $\hat{C}_Y^{-1}(k)$. By choosing $W(k)$ properly, it is also possible to put more weight in a frequency band of interest. When no covariance information is available and no specific weighting is required by the user, a constant weighting is employed ($W(k) = 1$, for $k = 1, \ldots, F$). Furthermore, the model error $\varepsilon(k, \theta) \in \mathbb{C}^{n_y}$ is defined as

$$\varepsilon(k,\theta) = Y(k,\theta) - Y(k), \tag{15}$$

where $Y(k, \theta)$ and $Y(k)$ are the DFT of the modelled and the measured output, respectively.

E. Calculation of the Jacobian

$V_{\text{WLS}}(\theta)$ is minimized with respect to the model parameters $\theta = [\text{vec}(A);\text{vec}(B);\text{vec}(C);\text{vec}(D);\text{vec}(E);\text{vec}(F)]$ via the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963). This requires the computation of the Jacobian $J(k, \theta)$ of the modelled output with respect to the model parameters:

$$J(k,\theta) = \frac{\partial\varepsilon(k,\theta)}{\partial\theta} = \frac{\partial Y(k,\theta)}{\partial\theta}. \tag{16}$$

Since it is impractical to calculate $Y(k, \theta)$ and $J(k, \theta)$ directly in the frequency domain, the calculations are performed in the time domain, followed by a DFT. It is known that the calculation of the Jacobian for a nonlinear state space model boils down to computing the output of another nonlinear model. The dynamics of this new nonlinear model are closely related to the dynamics of the original model (e.g., see Narendra Kumpati & Parthasarathy (1990) and Suykens et al. (1996)). For the model equations in (13), explicit expressions for the Jacobian are derived in Paduart (2008).

F. Initial Conditions

When computing the state sequence $x(t)$, the initial state x_0 of the model in (13) should be taken into account. For this, three possible approaches are distinguished. The simplest, but rather inefficient way, is to calculate the Jacobian for the full data set, and then to discard the first N_{trans} transient samples of both the Jacobian and the model error. However, in this way a part of the data is not used for the model estimation. The second method is only applicable when using periodic excitations. In this case, it suffices to calculate the Jacobian for several periods, and to select a period for which the transients become negligible. The last method, which is suitable for both periodic and non periodic excitations, is to estimate the initial conditions x_0 as if they were ordinary model parameters. Note that this corresponds to the estimation of an extra column in the state space matrix B and to an extra value 1 in the input vector.

G. Starting Values

The last obstacle before starting the nonlinear optimization is to choose good starting values for θ. The estimates obtained from the parametric linear state space model are used as initial values for the A, B, C, and D matrices. The other state space matrices E and F are initially set to zero. The idea of using the parametric BLA as the initial nonlinear model offers two important advantages. First of all, it guarantees that the estimated nonlinear model performs at least as good as the best linear model. Secondly, this principle results in a rough estimate of the model order n_a.

H. Similarity Transform

Another issue that needs to be addressed is the rank deficiency of the Jacobian, which is present due to the non-uniqueness of the state space representation. As mentioned before, the similarity transform $x_T(t) = T^{-1}x(t)$ leaves the input/output behaviour unaffected. The n_a^2 elements of the transformation matrix T can be chosen freely, under the condition that T is non singular. Consequently, the parameter space has at least n_a^2 unnecessary dimensions. This poses a problem for the gradient-based identification of the model parameters $\theta \in \mathbb{R}^{n_\theta}$: the Jacobian will not be of full rank and, hence, an infinite number of equivalent solutions will exist. To deal with the rank deficiency of the Jacobian when calculating the parameter update, a pseudo-inverse is used. This is achieved by a truncated Singular Value Decomposition (SVD) (Pintelon et al., 1999; Wills & Ninness, 2008; Golub & Van Loan, 1996); it allows a full parametrization of the model.

I. Overfitting and Validation

The nonlinear search should be pursued until the cost function in (14) stops decreasing. However, as it is often the case for model structures with many parameters, overfitting can occur during the nonlinear optimization. In order to avoid this effect, the so-called stopped search (Sjöberg et al., 1995) is used. The model quality of every estimated model is evaluated on a 'fresh' validation set, and then the model that achieves the best result is selected. This method is a form of implicit regularization, because unnecessary parameters are not explicitly removed.

6. EXPERIMENTAL RESULTS

The presented identification procedure is now applied to measurements from a physical system.

A. Description of the DUT

The DUT is an electronic nonlinear circuit with a Wiener-Hammerstein structure (see Fig. 1), designed by Gerd Vandersteen (Vandersteen, 1997). The system is composed of a static nonlinear block, sandwiched between two linear dynamic blocks. The first linear system is a third order Chebyshev low-pass filter with a 0.5 dB ripple and a pass band up to 4.4 kHz. The static nonlinearity is realized by two resistors and a diode. The second linear system is a third order inverse Chebyshev low-pass filter with a -40 dB stop band, starting at 5 kHz.

B. Description of the Experiments

The Wiener-Hammerstein system was excited with a filtered Gaussian excitation signal with a cut-off frequency of 10 kHz. The measurements were performed at a sampling frequency of 51.2 kHz. The noise sequence of $N = 188000$ samples is split in two parts: the estimation data ($N = 1, \ldots, 100000$) and the test data ($N = 100001, \ldots, 188000$). The estimation data is again split in two: the first part ($N = 1, \ldots, 80000$) is used for estimation purposes, the second part ($N = 80001, \ldots, 100000$) for validation and model selection purposes. The test data is used to benchmark the quality of the selected PNLSS model.

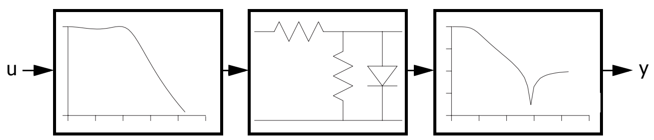

Fig. 1. Wiener-Hammerstein system.

C. Results

First, we calculate $\hat{G}_{BLA}(j\omega_k)$ and the total variance $\hat{\sigma}^2_{BLA}(k)$ (i.e., the combined effect of the nonlinear distortions and the measurement noise) according to the non periodic approach described in Paduart (2008). The BLA is plotted in Fig. 2 (solid black line), together with the total standard deviation $\hat{\sigma}_{BLA}(k)$ (dashed black line).

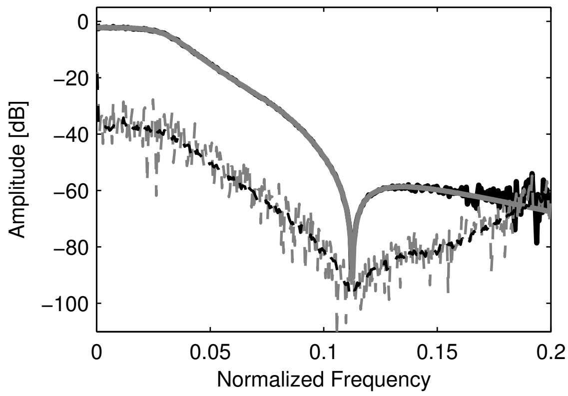

Fig. 2. BLA of the Wiener-Hammerstein circuit (solid black line); Total standard deviation (dashed black line); 6th order linear model (solid grey line); Model error (dashed grey line).

To obtain a parametric model for the BLA, linear models of various orders are estimated. For this, a subspace technique is used which is followed by a numeric optimization, both carried out in the frequency domain. The 6th order linear model yields the best result and is shown in Fig. 2 (solid grey line), together with the model error (dashed grey line).

Next, the 6th order linear model is validated using the test data set. In Fig. 3, the simulation error (grey) is plotted together with the measured output (black). The Root Mean Square Error (RMSE) is compared to the RMS level of the output: 56.2 mV versus 242.3 mV.

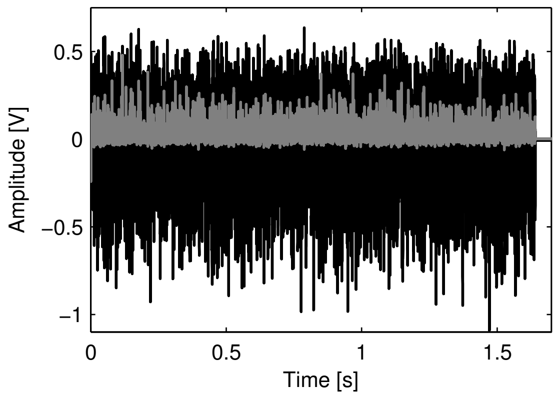

Fig. 3. Validation result for the 6th order linear model: measured output (black) and model simulation error (grey).

The linear models obtained in the previous step are now used as starting values to estimate a number of polynomial nonlinear state space models. We estimate models of 5th and 6th order. Two kinds of models are discussed: models that use all the nonlinear combinations of the states and the input ($\xi(t)_{\{3\}}$, "full"), and models that use only the nonlinear combinations of the states ($x(t)_{\{3\}}$, "states only"). We also verify whether the use of a linear ($\eta(t) = [\ ]$) or nonlinear ($\eta(t) = \xi(t)_{\{3\}}$ or $\eta(t) = x(t)_{\{3\}}$) output equation influences the results. Table I shows the RMSE results for the "full" PNLSS models using the validation data set (i.e., the second part of the estimation data set). When calculating the RMSE, the first 200 transient samples were discarded.

Table I: Validation results for the 'full' PNLSS models.

Model Order	PNLSS, 'full' $\zeta(t) = \eta(t) = \xi(t)_{\{3\}}$		PNLSS, 'full' $\zeta(t)=\xi(t)_{\{3\}}, \eta(t)=[\ ]$	
	Validation RMSE [mV]	Number of parameters	Validation RMSE [mV]	Number of parameters
n=5	1.93	473	2.03	396
n=6	0.42	797	0.44	685

In Table II, the validation results are shown for models that do not use the input in the nonlinear combinations. Since there are less nonlinear terms in these models, the number of required parameters is significantly lower. However, the RMSE values are always higher than the corresponding entries in Table I. When taking a closer look at Table I, we observe that the models with a nonlinear output equation always yield better results than the models with a linear output equation. The best PNLSS model is the 6th order model with a nonlinear output equation that uses all the nonlinear combinations of the states up to degree 3. It has a validation RMSE of 0.42 mV and contains 797 model parameters.

Table II: Validation results for the 'states only' PNLSS models.

Model Order	PNLSS, 'states only' $\zeta(t) = \eta(t) = x(t)_{\{3\}}$		PNLSS, 'states only' $\zeta(t)=x(t)_{\{3\}}, \eta(t)=[\ ]$	
	Validation RMSE [mV]	Number of parameters	Validation RMSE [mV]	Number of parameters
n=5	2.29	311	2.44	261
n=6	0.61	552	0.65	475

Next, we estimate state affine models (Sontag, 1979) of various orders and of degree 3 and 4. State affine models are the discrete-time counterpart of bilinear state space models. It has been shown that any continuous discrete-time system can be approximated arbitrarily well by these models (Fliess & Normand-Cyrot, 1982). The main difference with the standard PNLSS model is the absence of monomials with a degree higher than 1 in the states. The modelling results are given in Table III. We see that the RMS error diminishes smoothly as the model order increases. For state affine models, the best validation result is achieved by the 10th order model of degree 4, with a RMSE of 1.87 mV.

The quality of the best PNLSS model is now benchmarked on the test data set. The simulation error of this model is plotted in Fig. 4 (grey), together with the modelled output signal (black).

Table III: Validation results for some state affine models.

Model Order	State Affine degree 3		State Affine degree 4	
	Validation RMSE [mV]	Number of parameters	Validation RMSE [mV]	Number of parameters
n=5	7.90	83	7.71	119
n=6	4.66	111	3.73	160
n=7	3.64	143	3.00	207
n=8	3.37	179	2.58	260
n=9	2.89	219	2.06	319
n=10	2.67	263	1.87	384

Fig. 5 shows the spectra of the modelled output signal (black), the linear simulation error (light grey), and the nonlinear simulation error (dark grey). In the pass-band of the device, the nonlinear model pushes down the linear model error with about 30 dB. Beyond 15 kHz, no significant difference between the linear and the nonlinear model error can be observed.

Table IV shows the mean value, the standard deviation and the RMS value of the simulation error on both the full estimation and the test data set. The first 1000 transient samples were omitted from the calculation. Note that the nonlinear model achieves an RMS error which is 100 times smaller than the RMSE of the BLA (56.2 mV).

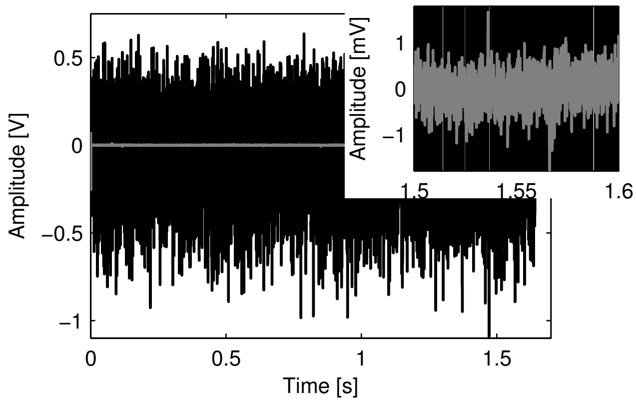

Fig. 4. Benchmark result for the best PNLSS model: modelled output (black) and model simulation error (grey).

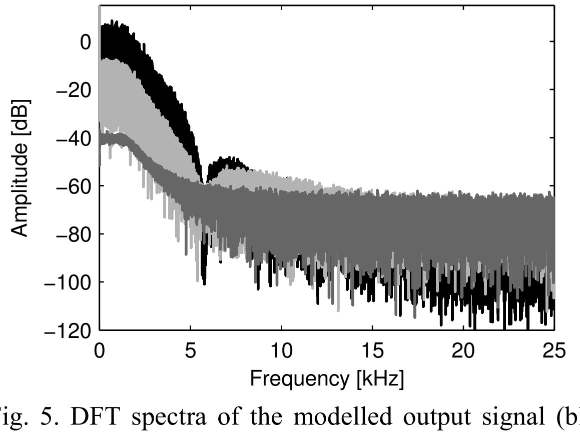

Fig. 5. DFT spectra of the modelled output signal (black); linear simulation error (light grey); and nonlinear simulation error (dark grey).

Table IV: Characteristics of the simulation error.

	estimation data [mV]	test data [mV]
mean	0.031	0.048
std	0.359	0.415
RMSE	0.360	0.418

7. CONCLUSION

In this paper, we successfully applied the Polynomial Nonlinear State Space approach to identify a nonlinear system with a Wiener-Hammerstein structure. The identification procedure is straightforward, and consists of three simple steps: (1) compute the Best Linear Approximation, (2) estimate a linear state space model, and (3) solve a standard nonlinear optimization problem. We achieved a significant model error reduction compared with the best linear model.

ACKNOWLEDGEMENT

This work was funded by the Methusalem grant of the Flemish Government (METH-1), and the Belgian program on interuniversity poles of attraction initiated by the Belgian State, Prime Minister's Office, Science Policy programming (IUAP-VI/4).

REFERENCES

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

Enqvist, M. (2005). *Linear models of nonlinear systems*. Ph.D. thesis, Linköping University, Sweden.

Enqvist, M., & Ljung, L. (2005). Linear approximations of nonlinear FIR systems for separable input processes. *Automatica*, 41(3), 459-473.

Fliess, M., & Normand-Cyrot, D. (1982). On the approximation of nonlinear systems by some simple state-space models. *Proc. of the IFAC Identification and Parameter Estimation Conference*, USA, pp. 511-514.

Golub, G. H., & Van Loan, C. F. (1996). *Matrix Computations* (3rd ed.). John Hopkins University Press, Baltimore.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, 2, 164-168.

Ljung, L. (1999). *System Identification: Theory for the User* (2nd ed.). Prentice Hall, Upper Saddle River, New Jersey.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, 11, 431-441.

McKelvey, T., Akçay, H., & Ljung, L. (1996). Subspace-based multivariable system identification from frequency response data. *IEEE Trans. Autom. Contr.*, 41(7), 960-979.

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

Paduart, J. (2008). *Identification of Nonlinear Systems Using Polynomial Nonlinear State Space Models.* Ph.D. thesis, Vrije Universiteit Brussel, Belgium. http://wwwtw.vub.ac.be/elec

Pintelon, R. (2002). Frequency-domain subspace system identification using non-parametric noise models. *Automatica*, 38(8), 1295-1311.

Pintelon, R., & Schoukens, J. (2001). *System Identification. A Frequency domain approach.* IEEE Press, New Jersey.

Pintelon, R., Schoukens, J., Vandersteen, G., & Rolain, Y. (1999). Identification of Invariants of (Over)parametrized Models: Finite Sample Results. *IEEE Trans. Autom. Contr.*, AC-44(5), 1073-1077.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14, 465-471.

Schoukens, J., Pintelon, R., Dobrowiecki, T., & Rolain, Y. (2005). Identification of linear systems with nonlinear distortions. *Automatica*, 41(3), 491-504.

Schoukens, J., Dobrowiecki, T., & Pintelon, R. (1998). Identification of linear systems in the presence of nonlinear distortions. A frequency domain approach. *IEEE Trans. Autom. Contr.*, 43(2), 176-190.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.-Y., Hjalmarsson, H., & Juditsky, A. (1995). Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12), 1691-1724.

Söderström, T., & Stoica, P. (1989). *System Identification*. Prentice Hall, Englewood Cliffs.

Sontag, E. D. (1979). Realization theory of discrete-time nonlinear systems: Part I: The bounded case. *IEEE Trans. on Circuits and Systems*, 26(5), 342-356.

Suykens, J. A. K., Vandewalle, J., & De Moor, B. (1996). *Artificial Neural Networks for Modelling and Control of Non-Linear Systems. and their application to control*. Kluwer Academic Publishers.

Vandersteen, G. (1997). *Identification of linear and nonlinear systems in an errors-in-variables least squares and total least squares framework.* Ph.D. thesis, Vrije Universiteit Brussel, Belgium.

Weisstein, E. W. *Multinomial Series*. From MathWorld - A Wolfram Web Resource. http://mathworld.wolfram.com/MultinomialSeries.html

Wills, A., & Ninness, B. (2008). On gradient-based search for multivariable system estimates. *IEEE Trans. Autom. Contr.,* 53(1), 298-306.

Appendix A. Notational Issue

In order to denote monomials in an uncomplicated way, the n-dimensional multi-index α is defined which contains the powers of a multivariable monomial:

$$\alpha = \left[\alpha_1\ \alpha_2\ \ldots\ \alpha_n\right], \tag{17}$$

with $\alpha_i \in \mathbb{N}$. A monomial composed of the components from the vector $\xi \in \mathbb{R}^n$ is then simply written as

$$\xi^{\alpha} = \prod_{i=1}^{n} \xi_i^{\alpha_i}, \tag{18}$$

where ξ_i is the i-th component of ξ. The total degree of the monomial is given by

$$|\alpha| = \sum_{i=1}^{n} \alpha_i, \tag{19}$$

and the factorial function of the multi-index α is defined as

$$\alpha! = \alpha_1!\alpha_2!\ldots\alpha_n!. \tag{20}$$

Furthermore, $\xi_{(p)}$ is defined as the column vector of all the distinct monomials of degree p (i.e., with multi-index $|\alpha| = p$) composed from the elements of vector ξ. The number of elements in vector $\xi_{(p)}$ is given by the following binomial coefficient:

$$\binom{n+p-1}{p}. \tag{21}$$

Finally, the vector $\xi_{\{p\}}$ is defined as the column vector containing all the monomials of degree two up to degree p.

The notations introduced above can now be used to express the multinomial expansion theorem (Weisstein). The multinomial expansion theorem gives an expression for the power of a sum, as a function of the powers of the terms:

$$\left(\sum_{i=1}^{n} \xi_i\right)^k = \sum_{|\alpha| = k} \frac{k!}{\alpha!}\xi^{\alpha} \tag{22}$$