engineering proceedings

MDPI

Proceeding Paper

Nonlinear System Identification Using Taylor Polynomial Parameterization of SDP Model and ERR-Based Term Selection Algorithm †

Khoi Ho [1,2,3], Duc Hieu Tran [1,2,3], Le Trinh [1,3], Nhat Vu Nguyen [3,4], Duc Duy Nguyen [3,4] and Nguyen-Vu Truong [3,4,*]

[1] Institute of Information Technology, Hanoi 10000, Vietnam; khoih@ioit.ac.vn (K.H.); tranduchieu@gmail.com (D.H.T.)
[2] Faculty of Information Technology and Telecommunication Engineering, Graduate University of Science and Technology, Hanoi 10000, Vietnam
[3] Vietnam Academy of Science and Technology, Hanoi 10000, Vietnam
[4] Institute of Mechanics, Hanoi 10000, Vietnam
* Correspondence: tnvu@vast.gov.vn
† Presented at the 2026 International Conference on Energy, Infrastructure and Environmental Research (EIER 2026), Ho Chi Minh City, Vietnam, 15–17 July 2026.

Abstract

This paper presents an approach to nonlinear system identification using state-dependent parameter (SDP) models. Here, polynomial-based functional approximation is used for SDP parameterization, and the Error Reduction Ratio (ERR)-based term selection algorithm is exploited to deliver compactness and parsimony in the resulting final SDP model. The effectiveness of the proposed approach has been demonstrated through simulation results.

Keywords: SDP model; state-dependent parameter; ERR term selection

1. Introduction

In the midst of machine learning and deep learning golden age, the search for efficient and parsimonious representation of nonlinear dynamic engineering system has been still an open research topic [1–8]. While NARX (Nonlinear Autoregressive with Exogenous input) and its related variations have been thoroughly studied and investigated [3,4], the SDP model has received less attention in spite of being proven to be highly effective in the area of data-based modeling of environmental systems [5,9–12].

In its simplest form, the SDP model can be seen as a direct renovation of linear ARX models to represent a nonlinear system in which the associated model's parameters are no longer constant coefficients as in the ARX model but expressed in the form of nonlinear functions with respect to the system's states [5,9–13]. In this manner, the system's most dominant nonlinear dynamics are located and captured in the representation of SDPs.

Unlike "black box" machine learning algorithms, the SDP model provides a natural yet highly interpretable instrument for assessment and analysis of the system's nonlinearities in a simple mathematical formulation. The advantages rely on the following facts:

- Within this structure, the relationship between each SDP and relevant states, which corresponds to associated nonlinearities, would be visualized and conceptualized.
- Its underlying pseudo-linear ARX-like structure facilitates a computationally efficient parameter estimation scheme in contrast to training heavy machine learning/deep learning algorithms.

Check for updates

Academic Editor: Le Kim Hung

Published: 12 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Traditionally, SDP non-parametric estimation is obtained by exploiting recursive Kalman filter and fixed interval smoothing (FIS) algorithms [5,9–13]. An approach for such SDP parameterization is to use a wavelet network [10–13], which has been proven to be effective in a number of applications [11,12]. Despite its excellent capabilities for functional approximation, wavelet networks (as a multi-resolution analysis) can become highly complex and computationally expensive as the system's dimensionality increases [10,12,13].

To address the aforementioned computational bottleneck, a trade-off between accuracy and complexity needs to be made. The Taylor polynomial approximation presents an attractive alternative. Here, SDPs are expanded by a truncated Taylor series, and the associated SDP model is converted to a linear-in-the-parameters model in polynomial space (TP-SDP) which favors numerous efficient parameter estimation schemes.

Nevertheless, such TP-SDP models are often over-parameterized, and are more likely to lead to the danger of over-fitting. In order to obtain model parsimony, the ERR (Error Reduction Ratio)-based term selection algorithm is employed [3,4] to provide a quantifiable measure of the contribution of each model's term to the final model's prediction and generalization capability. In this manner, the resulting final model is compact and the associated parameter estimation is accomplished via the standard least squares algorithm.

The remainder of this paper is organized as follows. Section 2 presents the TP-SDP model, ERR-based model term selection algorithm and relevant orthogonal least squares (OLS) parameter estimation scheme. The simulation results are illustrated in Section 3, and finally, Section 4 concludes the paper.

2. Taylor Polynomial State-Dependent Parameter (TP-SDP) Model

2.1. TP-SDP Model

The formulation of the TP-SDP model starts with the following discrete time state-dependent parameter (SDP) model, i.e., [5,10,11]:

$$y(k) = \sum_{i=1}^{n_y} f_i[y(k-i)]y(k-i) + \sum_{i=0}^{n_u} g_i[u(k-i)]u(k-i) + e(k) \quad (1)$$

where $y(k)$,$u(k)$ denote the output–input sequence at index k; n_y,n_u regard the maximum lags at output and input terms; $f_i(x)$,$g_i(x)$ are the state-dependent parameters (SDPs) which are nonlinear functions of the state x; and $e(k)$ corresponds to the error term.

Exploiting the Taylor polynomial (to the order of m) to approximate $f_i(x)$ and $g_i(x)$ as $f_i(x) = \sum_{i=0}^{m} a_i x^i$ and $g_i(x) = \sum_{i=0}^{m} b_i x^i$, (1) can be converted into the following form:

$$y(k) = \sum_{i=1}^{n_y} \left\{ \sum_{j=0}^{m} a_{i,j} y(k-i)^{j+1} \right\} + \sum_{i=0}^{n_u} \left\{ \sum_{j=0}^{m} b_{i,j} u(k-i)^{j+1} \right\} + e(k) \quad (2)$$

This formulates the Taylor polynomial state-dependent parameter (TP-SDP) model. With the output–input data of N data length defined by $\{Y,U\}$: $Y = [y(1),\ldots,y(N)]^T$ and $U = [u(1),\ldots,u(N)]^T$, $E = [e(1),\ldots,e(N)]^T$, the following equation is obtained:

$$Y = A\theta_y + B\theta_u + E \quad (3)$$

$$A = [A_1, \ldots, A_N]^T$$

$$A_k = \begin{bmatrix} y(k-1) & \ldots & y(k-1)^{m+1} & \ldots & y\left(k-n_y\right) & \ldots & y\left(k-n_y\right)^{m+1} \end{bmatrix}^T$$

$$B = [B_1, \ldots, B_N]^T$$

$$B_k = \begin{bmatrix} u(k) & \ldots & u(k)^{m+1} & \ldots & u(k-n_u) & \ldots & y(k-n_u)^{m+1} \end{bmatrix}^T$$

$$\theta_y = \begin{bmatrix} a_{1,1} & \ldots & a_{1,m} & \ldots & a_{n_y,1} & \ldots & a_{n_y,m} \end{bmatrix}^T$$

https://doi.org/10.3390/engproc2026164005

$$\theta_u = \begin{bmatrix} b_{0,1} & \cdots & b_{0,m} & \cdots & b_{n_u,1} & \cdots & b_{n_u,m} \end{bmatrix}^T$$

Here, the sizes of A_k,B_k are $(m+1)n_y \times 1$ and $(m+1)(n_u+1) \times 1$; thus, the matrices A,B are $N \times [(m+1)n_y]$ and $N \times [(m+1)(n_u+1)]$, respectively, while the sizes of θ_y,θ_u are, respectively, $(m+1)n_y \times 1$ and $(m+1)(n_u+1) \times 1$.

Let $\theta = \left[{\theta_y}^T, {\theta_y}^T\right]^T$ and $P = [A,B]$, and (3) can be expressed in the linear-in-the-parameters format:

$$Y = P\theta + E \tag{4}$$

In this format, the parameter θ can be estimated as a standard least squares scheme to minimize E^TE, i.e., $\theta = \left[Y^TP\right]^{-1}Y^TY$. Here, the sizes of P and θ are, respectively, $N \times [(m+1)(n_y+n_u+1)]$ and $[(m+1)(n_y+n_u+1)] \times 1$.

Nevertheless, the model formulation as in (4) is often over-parameterized. In order to obtain sparseness in the model representation, the ERR-based term selection algorithm [3,4] is exploited to obtain sparseness and parsimony.

2.2. ERR-Based Term Selection Algorithm

The Error Reduction Ratio (ERR) provides a simple measure for the contribution of each column of the data matrix P to the output variance $\frac{1}{N}Y^TY$. By orthogonally decomposing P into

$$P = WT$$

(where T is a $N \times N$ upper triangular matrix and W is a $N \times [M = (m+1)(n_y+n_u)]$ matrix with orthogonal column w_1,...,w_M so that $W^TW = D = diag[d_1,...,d_M]$ with $d_i =< w_i, w_i >= \sum_{k=1}^{N} w_i(k)w_i(k))$, (4) can be expressed in the following form:

$$Y = (PT^{-1})(T\theta) + E = WZ + E \tag{5}$$

Here, the auxiliary parameter $Z = T\theta = [z_1, \ldots, z_M]^T = D^{-1}W^TY$. With the assumption that $e(k)$ is not correlated to the past outputs of the systems, the output variance is as shown:

$$\frac{1}{N}Y^TY = \frac{1}{N}\sum_{i=1}^{M} {z_i}^2 {w_i}^T w_i + \frac{1}{N}E^TE$$

ERR_i, which signifies the contribution of the particular term i to the output variance, is calculated as $ERR_i = \frac{{z_i}^2 {w_i}^T w_i}{Y^TY} \times 100\%$. The larger this value is, the bigger contribution it makes. As a result, ERR_i can be used as a criterion for the TP-SDP model term selection algorithm which is implemented in a forward regression manner in accordance with the decreasing order of ERR_i values:

- **Step 1**: Calculate the value ERR_i for the specified term i for $i = 1, 2, \ldots, M$.
- **Step 2**: Select M_0 terms which are associated with the decreasing order of ERR_i until $1 - \sum_i^{M_0} ERR_i < \delta$. Here, δ regards the terminated error tolerance. This step can be executed starting from $M_0 = 1$ until δ is achieved. The result of this step is a model which consists of M_0 terms and the associated data matrix is given by $P^0 = \left[P_1^0, P_2^0, \ldots, P_{M_0}^0\right]$, where the column P_i^0 corresponds to the selected term included in the M_0-terms final model. In this manner, the final parameter θ^0 of the M_0-terms model is estimated as a standard least squares solution, i.e., $\theta^0 = \left[Y^TP^0\right]^{-1}Y^TY$.

3. Simulation Results

In the present study, a data set of 296 pairs of input–output data points (as is depicted in Figure 1) from a gas furnace process [14], sampled at 9 s, is considered as the model building exercise. The input data $u(k)$ is the methane gas feed rate (cu.ft/min) and the output $y(k)$ is CO_2 concentration (% of output gas) from the gas furnace.

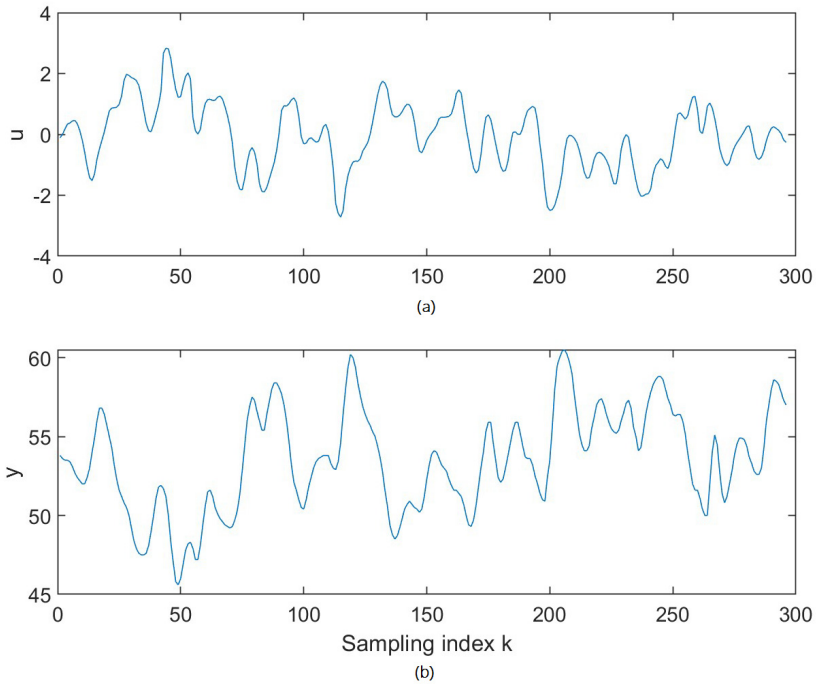

Figure 1. Gas furnace process data set: (**a**) input and (**b**) output.

Here, the input–output data is first demeaned and normalized into $[-1,1]$, i.e., $y_n = \frac{y-y_{mean}}{y_{max}}$ and $u_n = \frac{u-u_{min}}{u_{max}}$ ($y_{mean} = 53.5091$, $y_{max} = 60.5$, $u_{mean} = -0.0568$, $u_{max} = 2.834$). For the ease of representation, they are still designated as y and u.

The first 200 data points were used as the training data set and the remaining 96 data points were used for model validation. With $n_y = n_u = 2$ and $m = 2$, upon the ERR term selection algorithm terminated at $\rho = 0.4358\%$ *or* $\sum_i^{M_0} ERR_i = 99.5642\%$, the final TP-SDP, which consists of $M_0 = 5$ terms, is identified as

$$y(k) = \sum_{i=1}^{2} f_i[y(k-i)]y(k-i) + \sum_{i=0}^{1} g_i[u(k-i)]u(k-i) + e(k) \tag{6}$$

where

$$f_1(x) = 1.4826$$

$$f_2(x) = -0.5905 - 0.5139x^2$$

$$g_0(x) = 0.017$$

$$g_1(x) = -0.0325$$

The ERR values of each model's term are tabulated in Table 1, ranked from the most dominant to the least dominant. Figure 2 depicts the model's prediction over the training set (after being converted back to normal range, i.e., $y = y_n y_{max} + y_{mean}$). Its performance over the validation set is shown in Figure 3.

Table 1. The Error Reduction Ratio (ERR) of each model's term.

Rank	Model's Term	ERR (%)
1	$y(k-1)$	95.0541
2	$y(k-2)$	3.7912
3	$u(k)$	0.4311
4	$u(k-1)$	0.2762
5	$y(k-2)$	0.0115

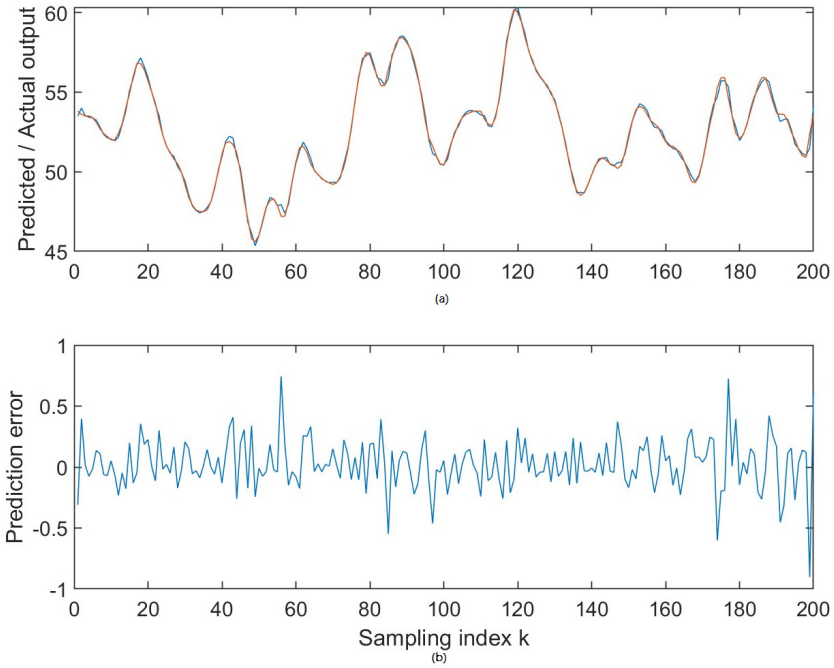

Figure 2. (**a**) The model's prediction (in blue) overlaid against the actual output (in red) and (**b**) the associated prediction error over the training set ($MSE = 0.0439$).

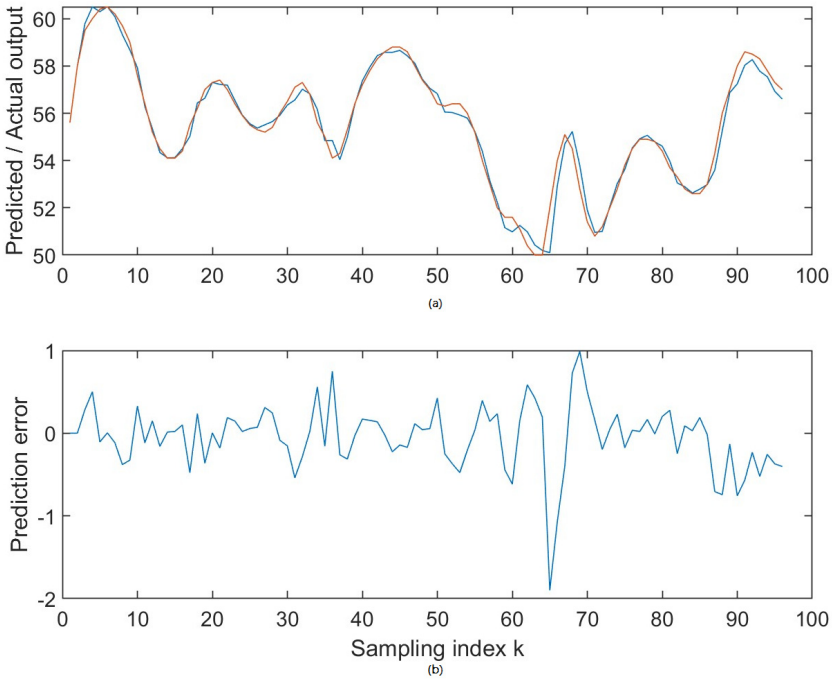

Figure 3. (**a**) The model's prediction (in blue) overlaid against the actual output (in red) and (**b**) the associated prediction error over the validation set ($MSE = 0.1551$).

In order to evaluate the model's generalization capability, (6) was simulated over the whole data's range by using the initial values of $y_n(1{:}2) = [0.0048\ 0.0015]$ or $y(1{:}2) = [53.8\ 53.6]$. The simulated results are displayed in Figure 4. This, in turn, implies the 5-term identified model excellently characterizes the input–output relationship of the gas furnace process under study.

https://doi.org/10.3390/engproc2026164005

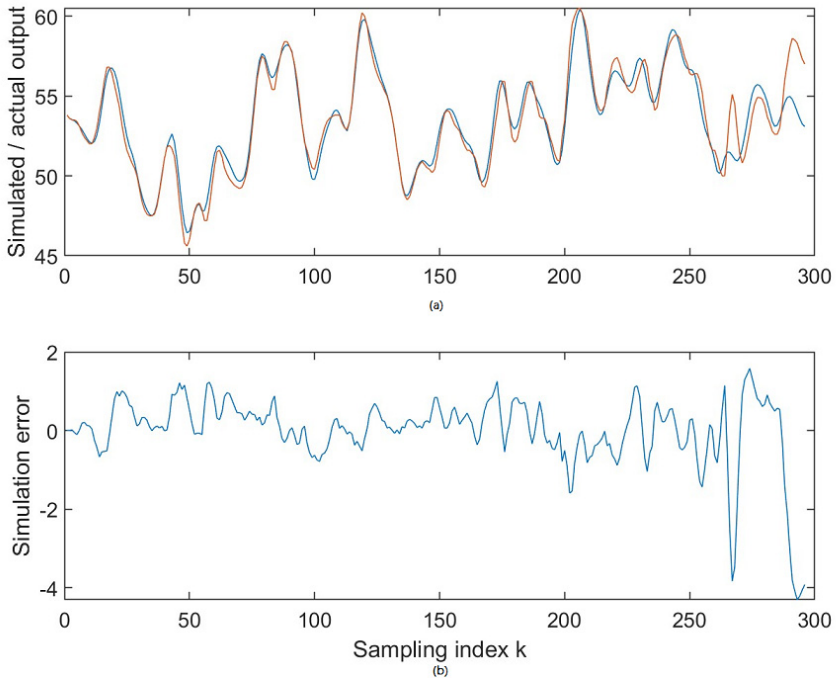

Figure 4. (**a**) The model's iterative output (in blue) overlaid against the actual output (in red) and (**b**) the associated prediction error over the whole data range.

The relevant SDPs and their relationship to the associated states are illustrated in Figure 5. Here, the system's nonlinearity is located at $y(k-2)$ and characterized by a parabolic function with negative coefficients.

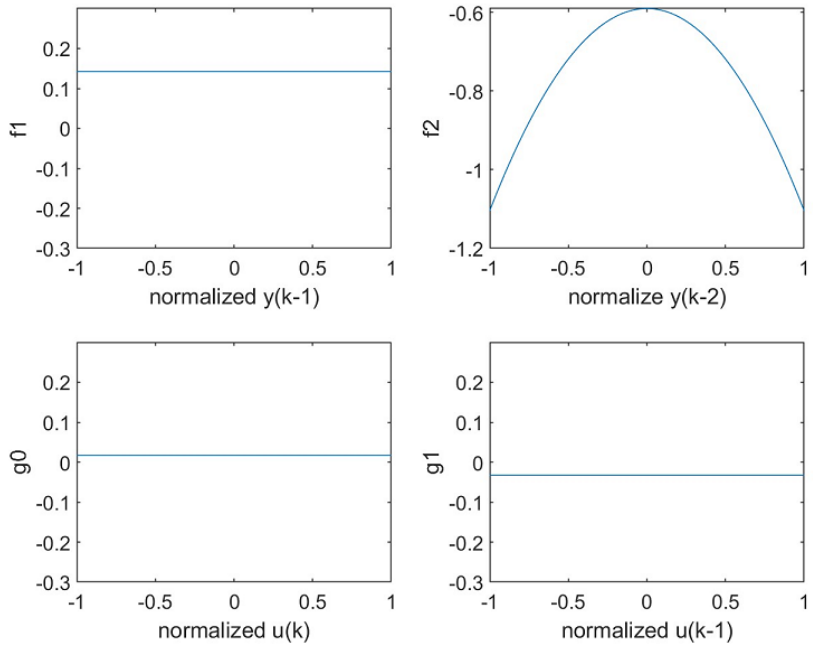

Figure 5. Relevant SDP location and characteristics of associated nonlinearities.

4. Conclusions

This paper has proposed a Taylor polynomial-based SDP model for nonlinear engineering system identification. The approach exploits the SDP structure to naturally locate and capture the system's nonlinearities while the Taylor polynomial is for SDP parameterization. The identification scheme incorporates an efficient ERR-based term selection algorithm to

achieve the final identified model's sparseness, avoiding the danger of over-fitting. The results obtained from the gas furnace process identification example demonstrate the merits and capability of such a TP-SDP modeling approach. The identified model is parsimonious and excellently characterizes the associated dynamics of the system under study. Future work includes identification under noise-corrupted input–output data, practical application, and extension to multiple-dimensional state dependencies.

Author Contributions: Conceptualization, K.H. and N.-V.T.; Methodology, K.H., D.H.T. and N.-V.T.; Software L.T., N.V.N. and D.D.N.; Formal analysis K.H.; Supervision N.-V.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hong, X.; Sharkey, P.M.; Warwick, K. Automatic nonlinear predictive model-construction algorithm using forward regression and the PRESS statistic. *IEE Proc.–Control Theory Appl.* **2005**, *150*, 245–254.
2. Hong, X.; Harris, C.J.; Chen, S.; Sharkey, P.M. Robust nonlinear model identification methods using forward regression. *IEEE Trans. Syst. Man. Cybern.–Part A Syst. Hum.* **2003**, *33*, 514–523. [CrossRef]
3. Wei, H.L.; Billings, S.A.; Liu, J. Term and variable selection for non-linear system identification. *Int. J. Control* **2004**, *77*, 86–110. [CrossRef]
4. Billings, S.A.; Wei, H.L. A new class of wavelet networks for nonlinear system identification. *IEEE Trans. Neural Netw.* **2005**, *16*, 862–874. [CrossRef]
5. Janot, A.; Young, P.C.; Gautier, M. Identification and control of electro-mechanical systems using state-dependent parameter estimation. *Int. J. Control* **2017**, *90*, 643–660. [CrossRef]
6. Noel, J.P.; Schoukens, J. Grey-box state space identification of nonlinear mechanical vibrations. *Int. J. Control* **2018**, *91*, 1118–1139. [CrossRef]
7. Chakraborty, R.; Jain, H.; Seo, G.S. A review of active probing-based system identification techniques with application in power systems. *Int. J. Electr. Power Energy Syst.* **2022**, *140*, 108008. [CrossRef]
8. Anastasio, D.; Raze, G.; Kerschen, G. Frequency-domain system identification of nonlinear structures using experimental continuation data. *J. Vib. Control* **2025**. [CrossRef]
9. Young, P.C. Data-based mechanistic modeling of engineering systems. *J. Vib. Control* **1998**, *4*, 5–28. [CrossRef]
10. Truong, N.V.; Wang, L.P.; Young, P.C. Nonlinear system modeling based on non-parametric identification and linear wavelet estimation of SDP models. *Int. J. Control* **2007**, *80*, 774–788. [CrossRef]
11. Truong, N.V.; Wang, L.P.; Huang, J.M. Nonlinear modeling of a magnetic bearing system using SDP model and linear wavelet parameterization. In Proceedings of the 2007 American Control Conference, New York, NY, USA, 9–13 July 2007; pp. 2254–2279.
12. Truong, N.V.; Wang, L.P.; Wong, K.C. Modeling and short-term forecasting of daily peak power demand in Victoria using two-dimensional wavelet based SDP model. *Int. J. Electr. Power Energy Syst.* **2008**, *30*, 511–518. [CrossRef]
13. Truong, N.V.; Wang, L.P. Nonlinear system identification using two-dimensional wavelet based SDP models. *Int. J. Syst. Sci.* **2009**, *40*, 1161–1180. [CrossRef]
14. Chen, S.; Hong, X.; Haris, C.J.; Wang, X. Identification of nonlinear system using generalized kernel models. *IEEE Trans. Control Syst. Technol.* **2005**, *13*, 401–411. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.