A Machine Learning Approach to the Identification of Dynamical Nonlinear Systems

Giorgio Biagetti, Paolo Crippa, Laura Falaschetti, Claudio Turchetti
DII – Department of Information Engineering
Università Politecnica delle Marche,
via Brecce Bianche, 12, I-60131 Ancona, Italy
{g.biagetti, p.crippa, l.falaschetti, c.turchetti}@univpm.it

***Abstract*—The aim of this paper is to present a general machine learning approach to the identification of nonlinear systems, using the observed input-output finite datasets. The approach is derived representing the input and output signals in the feature space by the principal component analysis (PCA), thus transforming the nonlinear time dependent identification problem to the regression of a nonlinear input-output function. To face this problem an effective machine learning technique based on particle-Bernstein polynomials has been used to model the input-output relationship that describes the system. The approach has been validated by identifying two real world nonlinear systems, in the fields of speech signals and nonlinear audio amplifiers.**

***Index Terms*—Machine learning, nonlinear systems, PCA, identification**

I. Introduction

Nonlinear system identification (NSI) refers to the problem of building a mathematical relation between input u and output y of an unknown dynamical system [1]–[7] . A large number of different approaches have been proposed in the literature over the last decades to face this problem. Among these the Lee-Shetzen method [8], [9] that identifies the Volterra kernels of nonlinear systems stimulated by random inputs with assigned statistics, is one of the most popular. To overcome calculation of multidimensional Volterra kernels a cascaded nonlinear identification model, with a static nonlinear element followed by a time-varying element (Hammerstein model [10]), and with a time-varying linear block followed by a static nonlinear element (Wiener model [11]), has been proposed. In the discrete-time domain one of the most successful approach for nonlinear system identification is the NARMAX model [12] (and its derivatives NARX [13] and NARMA [14]), in which the system is modelled in terms of a nonlinear functional expansion of lagged inputs, outputs and prediction errors. NARMAX models have shown to be very effective in many real-world applications [15]–[21], as they are powerful, efficient and unified representations of a wide variety of nonlinear systems. However a major difficulty in system identification using NARMAX model is selecting a model that is parsimonious in the number of parameters and represents the dynamics of the system adequately. Even though various methods, such as polynomials, multilayer perceptrons, wavelet ANNs and radial basis functions, have been used to build NARMAX models, the choice of an adequately model remains a bottleneck.

The aim of this paper is to present a general approach to the identification of nonlinear systems, that is based on principal component analysis (PCA) of input and output signals. In such a way the identification of a nonlinear input-output time-dependent transformation reduces to the regression of a nonlinear function, so that efficient machine learning techniques can be applied. The paper is organized as follows. Section II describes the mathematical framework of the proposed method. Section III reports an effective machine learning technique for the regression of nonlinear functions, recently suggested. Section IV presents the results of identifying some real world nonlinear systems. Finally, Section V discusses concluding remarks.

II. Transforming a Nonlinear Operator to a Nonlinear Function

Let us refer to an input-output generally nonlinear system, that transforms an input signal u to an output signal y, formally represented by an operator T such that

$$y = T(u),\ u \in \mathbb{R}^m,\ y \in \mathbb{R}^n \tag{1}$$

where both $u = (u(1), \dots, u(n))\,,\ y = (y(1), \dots, y(n))$ are real vectors both depending on time $t = 1, \dots, n$.

In this paper we will consider dynamical nonlinear systems alone, that is the case in which the operator T acts both on time t and magnitude of u, as the static case is not of interest. An example of such system is the well known autoregressive model given by

$$y(k+1) = h\left(y(k-1) \dots y(k-p), u(k) \dots u(k-q)\right). \tag{2}$$

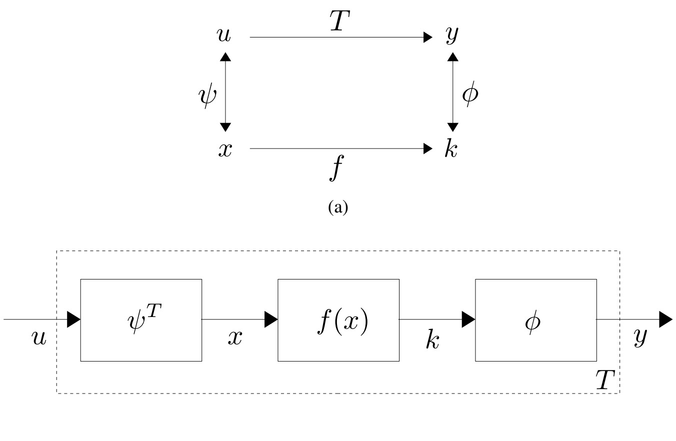
Assuming the input u belongs to a class of random signals with covariance matrix $R_{uu} = E\{u\,u^T\}$, thus R_{uu} can be decomposed as

$$R_{uu} = \psi\,\Lambda\,\psi^T,\ R_{uu} \in \mathbb{R}^{m\times m} \tag{3}$$

where ψ is an unitary matrix $(\psi^T\psi = \psi\psi^T = I)$ whose columns are the eigenvectors of R_{uu} and $\Lambda = diag(\lambda_1, \dots, \lambda_m)$ is the eigenvalue matrix. Due to the orthogonality property of ψ, the generic vector u can be written as a linear combination of the columns of ψ

$$u = \psi x \tag{4}$$

978-9-0827-9703-9/19/$31.00



(a)

(b)

Fig. 1: (a) Correspondance between u, y, x and k. (b) Schematic representation of (9).

being

$$x = \psi^T u, \ x \in \mathbb{R}^m \tag{5}$$

the vector of the coordinates on the basis ψ. (4) and (5) represent the well-known discrete Karhunen-Loève transform (DKLT) and its inverse, respectively [22]. Similarly, for y we have

$$y = \phi k \tag{6}$$

where ϕ is such that

$$R_{yy} = E\{yy^T\} = \phi \Gamma \phi^T \tag{7}$$

and $\Gamma \in \mathbb{R}^{n \times n}$ is a diagonal matrix. It is worth to notice that in the representations (4) and (6) the matrices ψ and ϕ depends on time as they are composed by eigenvectors, while x and k do not, as they represent coordinates (or features) in a basis. This is a well known property of DKLT transform that separates time from magnitude dependence. Combining (1), (4) and (6) it follows that: *i)* T establishes a correspondences, in general non biunivocal, between u and y; *ii)* ψ and ϕ establish two biunivocal correspondances between u and x, y and k respectively. As a consequence a correspondance between x and k exists, thus formally we have:

$$k = f(x) \tag{8}$$

where the function $f(\cdot)$ is in general nonlinear and independent of time as both x and k are. Fig. 1a gives an elucidation of the meaning of function $f(x)$ where u, y represents the input, output signals in time domain, while x, k in feature space. As you can see the nonlinear operator T in time domain reduce to a nonlinear function in feature space. This function reduces the input u to the output y since combining (5), (6) and (8) it results

$$y = \phi\, f(\psi^T u) = T(u) \tag{9}$$

where

$$T(\cdot) = \phi f\left(\psi^T(\cdot)\right) \quad . \tag{10}$$

A schematic representation of (9) is reported in Fig. 1b. Assume $u^{(i)}$, $i = 1, \ldots, N$ are realizations of u and $y^{(i)}$, $i = 1, \ldots, N$ the corresponding realizations of y satisfying (1). The estimation of the function $f(\cdot)$ in (8) can be viewed as a regression problem given the training set

$$\{\psi^T u^{(i)}, \, \phi^T y^{(i)}, \, i = 1, \ldots, N\} \tag{11}$$

which can be rewritten in matrix form as

$$X = U\Psi, \ K = Y\Phi \tag{12}$$

where U, Y, X, K are the data matrices of u, y, x, k, respectively. Finally, with this definitions in mind, the estimation of (10) reduces to the estimation of the function (8) given the data matrices X, K of x and k respectively. In order to reduce the dimensionality of the problem a PCA can be used [23]. To this end assume the eigenvalues in the matrix Λ are in descending order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, and the eigenvectors in ψ are in the corresponding order. Thus ψ can be partitioned as

$$\psi = (\psi_M \, \psi_\eta) \tag{13}$$

where the matrix $\psi_M \in \mathbb{R}^{m \times d}$ contains the d most significant eigenvectors and (4) can be rewritten according to this partition

$$u = \psi_d x_M + \psi_\eta x_\eta, \ x_M \in \mathbb{R}^d \quad . \tag{14}$$

Finally, by neglecting the residual term $\psi_\eta x_\eta$ corresponding to the least significant components, u can be approximated by

$$u \cong \psi_d x_M \tag{15}$$

and this yields to

$$x_M = \psi_d^T u \tag{16}$$

where the vector x_M of reduced dimension $d \ll m$ is used instead x. Having transformed the input u to the vector x_M, the identification of nonlinear system reduces to the regression of the function

$$\phi^T y = g(x_M), \ x_M \in \mathbb{R}^d \quad . \tag{17}$$

III. Regression by Particle Bernstein Polynomials

Recently an effective machine learning technique based on a set of new functions named particle-Bernstein polynomials (PBP) has been proposed for regression of input-output relationships [24]. Bernstein polynomials have the property that the coefficients are the values of the function to be approximated at points in a fixed grid, thus avoiding a time-consuming training stage. Thus this approach can fruitfully be used for solving the regression of function $f(\cdot)$ in (17). Following this method the estimate $f_m(z)$ of $f(\cdot)$ at the generic testing point z is given by

$$f(z) \cong f_m(z) = \frac{\displaystyle\sum_{j=1}^{N} f(x^{(i)}) k_\xi^m(x^{(j)})}{\displaystyle\sum_{j=1}^{N} k_\xi^m(x^{(j)})} \ , \tag{18}$$

where the set $\{x^{(j)}, f(x^{(j)}), j = 1, \ldots, N\}$ represents the training set that can be derived from the input-output set $\{u^{(i)}, y^{(i)},\ i = \ldots, N\}$ through (12), and $x_t, \xi_t,\ t = 1, \ldots, d$ are the coordinates of vectors $x = (x_1, \ldots, x_d)$ and $\xi = z/m = (\xi_1, \ldots, \xi_d)$ respectively. In (18) k_ξ^m is a polynomial of degree m defined as

$$k_\xi^m = \prod_{t=1}^{d} x_t^{\xi_t}(1 - x_t)^{m-\xi_t} \quad . \tag{19}$$

IV. Experimental Results

A. Validation of particle Bernstein polynomials approach for regression of nonlinear functions

The first experiment is addressed to the validation of particle Bernstein polynomials approach for the regression of a nonlinear function given an input-output data set. To this end let us refer to the problem of reconstructing a speech signal from the Mel-cepstral analysis, that is a central issue in the synthesis of speech [25], [26]. It is well known that the Mel filter bank defines a non invertible linear transform, so that reconstruction of speech signal is not guaranteed. In this experiment it will be shown that a nonlinear inverse transformation can be derived provided that some information on signal phase is used as input data. The experiment has been carried out on a speech signal extracted from an audio recording of a female Italian speaker. The signal of length 5800000 samples has been divided into a set of $N = 72500\,(N = 47838$ without silence periods) frames of length $n = 200$ samples with an overlap of 120 samples. From the spectrum of each frame 13 MFCC coefficients $c(j),\ j = 1, \ldots, 13$ and 24 phase values $\varphi(j),\ j = 1, \ldots, 24$ at the center frequencies of Mel filter banks have been derived. In order to reduce the input dimensionality only $l < 24$ phase components are taken into account, so that the input vector u is defined as

$$u = (c(1), \ldots c(13), \varphi(1), \ldots \varphi(l))^T \quad . \tag{20}$$

Thus a frame for testing is extracted from data set while the remaining frames are used for training the model (18). Fig. 2a compares the results achieved by PBP regression model (18) with data, for a polynomial order $m = 25$ and $l = 6$ phase components. Fig. 2b depicts the results achieved with $l = 10$ and $m = 15$, with reference to the same frame of Fig. 2a. PBPs do not require any time-consuming training phase, as only direct samples of the training set are used.

B. Identification of a vacuum tube audio amplifier

The second experiment refers to nonlinear audio system identification [5], [27] and aims at identifying the input-output dynamic nonlinear characteristic that is typical of vacuum tube audio amplifiers (VTAAs). To this end, a SPICE-level electrical simulation of a Fender Bandmaster 5E7 amplifier was perfomed to obtain input/output signal pairs. The circuit simulated is shown in Fig. 3. It is composed of two major stages: the input stage, which is essentially linear but presents an adjustable and strongly frequency-dependent gain, used for tone control, and the output power stage, which exhibits the VTAA typical soft-clipping and amplitude-dependent gain, as shown in Fig. 4. Clearly this is a time dependent nonlinearity as the input-output characteristic is strongly dependent on the input magnitude. For this experiment, a speech signal

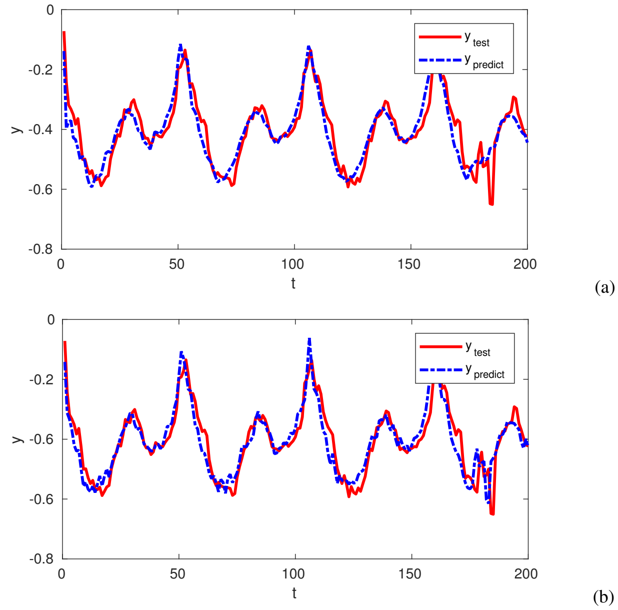


Fig. 2: Particle Bernstein: output y as a function of a discrete-time t. (a) features = 19, order = 25, $t_{\text{test}} = 0.029771\ s$; (b) features = 23, order = 15, $t_{\text{test}} = 0.034319\ s$.

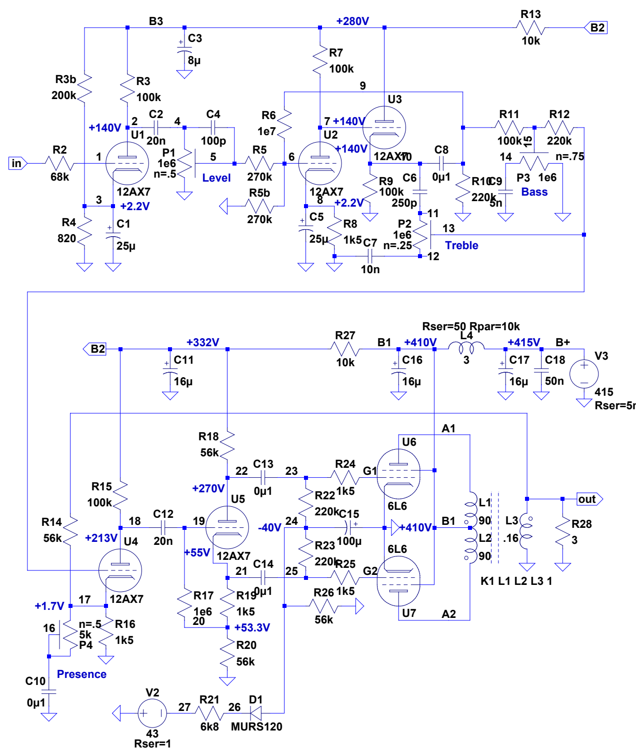


Fig. 3: Circuit diagram of the Fender Bandmaster 5E7 tube amplifier (restricted to a single channel) used for simulation.

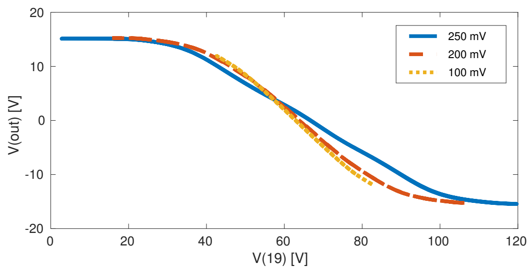


Fig. 4: Dynamic nonlinear behaviour of the Fender Bandmaster 5E7 tube amplifier: output voltage as a function of the grid voltage of tube U5 (node 19), for different amplitudes of the input signal (1 kHz sinusoid), showing decreasing gains as the VTAA enters its soft-clipping region.

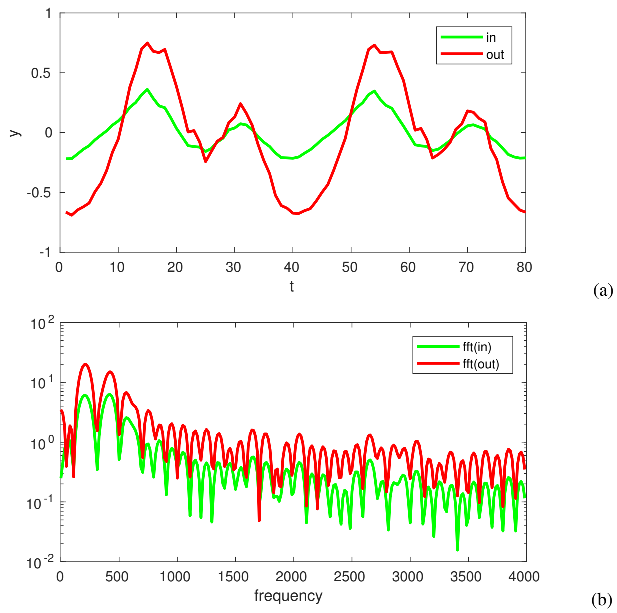


Fig. 5: A frame of the input/output signals - (a) time domain, (b) frequency domain.

extracted from an audiobook was used, at an amplitude of 400 mV. Fig. 5 shows a frame of the input/output signals used in the experiment and their spectra. Fig. 6 compares an output frame, extracted from training set, with the frame predicted by PBP regression model, in the time domain and frequency domain. Instead Fig. 7 compares an output frame, extracted from testing set (i.e. data not included in training set) with the frame achieved by the model (18), in the time domain and frequency domain, thus confirming (18) is able to satisfactorily model data lie outside the training set.

V. Conclusion

One of the main issues in system identification using current approaches is selecting a model that represents the dynamics of the system adequately and requires a reduced number of parameters. The approach presented in this paper does not require a suitable model to represent input-output data.

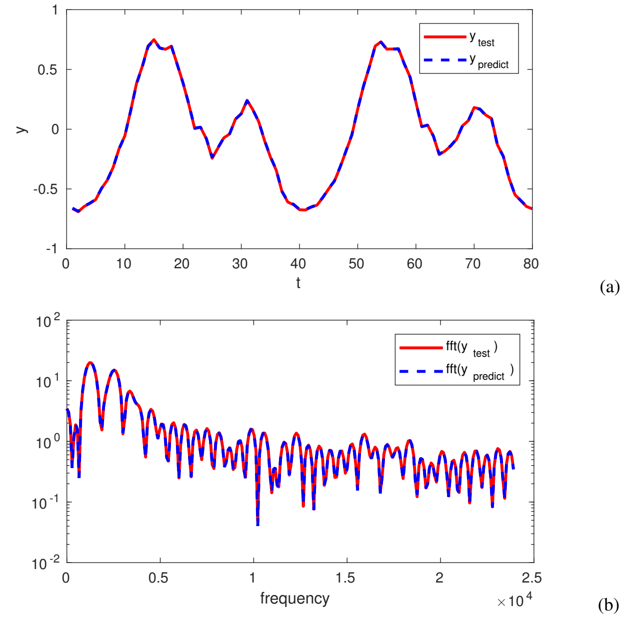


Fig. 6: Particle Bernstein (features $= 10$, order $= 100$), model validation - (a) time domain: output y as a function of a discrete-time t; $t_{\text{test}} = 0.037450\ s$, (b) frequency domain.

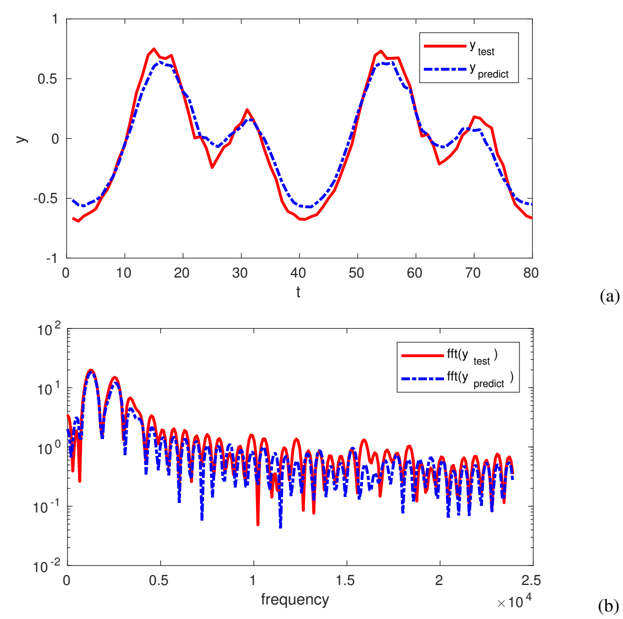


Fig. 7: Particle Bernstein (features $= 10$, order $= 100$) - (a) time domain: output y as a function of a discrete-time t; $t_{\text{test}} = 0.034963\ s$, (b) frequency domain.

Instead on the basis of PCA representation of input and output signals reduces to a nonlinear input-output static function. In this way it has been shown that using an effective machine learning technique, based on particle-Bernstein polynomials, the approach is able to identify real-world nonlinear dynamical systems.

References

[1] I. M. Yassin, M. N. Taib, and R. Adnan, "Recent advancements & methodologies in system identification: A review," *Scientific Research Journal*, vol. 1, no. 1, pp. 14–33, 2013.

[2] G. B. Giannakis and E. Serpedin, "A bibliography on nonlinear system identification," *Signal Process.*, vol. 81, no. 3, pp. 533–580, 2001.

[3] X. Hong, R. Mitchell, S. Chen, C. Harris, K. Li, and G. Irwin, "Model selection approaches for non-linear system identification: a review," *International Journal of Systems Science*, vol. 39, no. 10, pp. 925–946, 2008.

[4] J. D. Victor, "Analyzing receptive fields, classification images and functional images: challenges with opportunities for synergy," *Nature neuroscience*, vol. 8, no. 12, pp. 1651–1656, 2005.

[5] I. Mezghani-Marrakchi, G. Mahé, S. Djaziri-Larbi, M. Jaïdane, and M. T.-H. Alouane, "Nonlinear audio systems identification through audio input Gaussianization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 41–53, Jan 2014.

[6] G. Ramponi and G. L. Sicuranza, "Quadratic digital filters for image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 6, pp. 937–939, June 1988.

[7] K. Worden, *Nonlinearity in structural dynamics: detection, identification and modelling*. CRC Press, 2019.

[8] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proceedings of the IEEE*, vol. 69, no. 12, pp. 1557–1573, Dec 1981.

[9] M. Inagaki and H. Mochizuki, "Bilinear system identification by Volterra kernels estimation," *IEEE Transactions on Automatic Control*, vol. 29, no. 8, pp. 746–749, Aug 1984.

[10] X. Chen and H. Chen, "Recursive identification for MIMO Hammerstein systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 4, pp. 895–902, April 2011.

[11] L. A. Aguirre and C. Letellier, "Modeling nonlinear dynamics and chaos: a review," *Mathematical Problems in Engineering*, vol. 2009, 2009.

[12] S. Chen and S. A. Billings, "Representation of non-linear systems: the NARMAX model," *International Journal of Control*, vol. 49, no. 3, pp. 1012–1032, March 1989.

[13] L. Piroddi and M. Lovera, "NARX model identification with error filtering," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 2726–2731, 2008, 17th IFAC World Congress.

[14] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.

[15] N. A. Rahim, M. N. Taib, A. H. Adom, and M. A. A. Halim, "Nonlinear system identification for a DC motor using NARMAX model with regularization approach," in *International Conference on Control, Instrument and Mechatronics Engineering, CIM*, vol. 7, 2007.

[16] B. A. Amisigo, N. Van de Giesen, C. Rogers, W. E. I. Andah, and J. Friesen, "Monthly streamflow prediction in the Volta Basin of West Africa: A SISO NARMAX polynomial modelling," *Physics and Chemistry of the Earth*, vol. 33, no. 1-2, pp. 141–150, 2008.

[17] S. L. Kukreja, H. L. Galiana, and R. E. Kearney, "NARMAX representation and identification of ankle dynamics," *IEEE Transactions on Biomedical Engineering*, vol. 50, pp. 70–81, 2003.

[18] N. Chiras, C. Evans, and D. Rees, "Nonlinear gas turbine modeling using NARMAX structures," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 4, pp. 893–898, Aug 2001.

[19] R. Boynton, M. Balikhin, H.-L. Wei, and Z.-Q. Lang, "Applications of NARMAX in space weather," in *Machine Learning Techniques for Space Weather*. Elsevier, 2018, pp. 203 – 236.

[20] H.-L. Wei, "Sparse, interpretable and transparent predictive model identification for healthcare data analysis," in *Advances in Computational Intelligence*. Springer International Publishing, 2019, pp. 103–114.

[21] J. R. Ayala Solares, H.-L. Wei, and S. A. Billings, "A novel logistic-NARX model as a classifier for dynamic binary classification," *Neural Computing and Applications*, vol. 31, no. 1, pp. 11–25, Jan 2019.

[22] C. W. Therrien, *Discrete random signals and statistical signal processing*. Prentice Hall PTR, 1992.

[23] I. T. Jolliffe, "Principal component analysis," in *International encyclopedia of statistical science*. Springer, 2011, pp. 1094–1096.

[24] G. Biagetti, P. Crippa, L. Falaschetti, and C. Turchetti, "Machine learning regression based on particle Bernstein polynomials for nonlinear system identification," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, Sept 2017, pp. 1–6.

[25] Z. Ling, L. Deng, and D. Yu, "Modeling spectral envelopes using restricted Boltzmann machines and deep belief networks for statistical parametric speech synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2129–2139, Oct 2013.

[26] J. P. Cabral, K. Richmond, J. Yamagishi, and S. Renals, "Glottal spectral separation for speech synthesis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 2, pp. 195–208, April 2014.

[27] L. Tronchin, "The emulation of nonlinear time-invariant audio systems with memory by means of Volterra series," *J. Audio Eng. Soc*, vol. 60, no. 12, pp. 984–996, 2013.