

STATISTICAL CLASSIFICATION OF CHAOTIC SIGNALS

*Christophe Couvreur** *Cédric Flamme* *Marc PirLOT†*

* Service de Physique Générale

† Service de Mathématique et de Recherche Opérationnelle

Faculté Polytechnique de Mons, Rue de Houdain 9, B-7000 Mons, Belgium

E-mail: couvreur@thor.fpms.ac.be, pirlot@mathro.fpms.ac.be

ABSTRACT

The classification of chaotic signals generated by a low-dimensional deterministic models given a dictionary of possible model is considered. The proposed classification methods rely on the concept of “best predictor” of signal. A statistical interpretation of this concept based on the ergodic theory of chaotic system is presented. A sort of “bootstrapping” estimator of the statistical properties is introduced. The method is validated by numerical simulations. Directions for future research are suggested.

1. INTRODUCTION

Also known under the popular name of “chaos theory,” the theory of nonlinear dynamical system has been the subject of considerable advances over the past twenty years. Chaotic systems are deterministic dynamical systems with a small number of degrees of freedom whose behavior appears, in a sense, random and unpredictable. Most of the early research efforts on chaos theory have been the fact of mathematicians and physicists. Mathematical and computational tools for the characterization of chaos in theoretical and experimental systems have been developed [1, 2]. They have been used successfully in a variety of scientific fields. However, it is only recently that electrical engineers have started to look at possible “practical” applications of chaos theory [3, 4, 5].

From the signal processing viewpoint, the engineering applications of chaos that have been proposed so far can be broadly organized into four categories: characterization of chaotic signals [6], synthesis of chaotic signals with useful properties [7], filtering, prediction, or smoothing of chaotic signals observed in noise [8, 9], replacement of the stochastic model for the perturbing noise in a signal of interest by a chaotic model for parameter estimation and signal detection problems [10]. In the applications of the last category, taking advantage of the structure present in the chaotic “noise” generally led to improved performance in situations for which a low-dimensional chaotic noise model is better suited than a stochastic noise model.

In this paper we treat the following related problem: given a “random” signal of unknown origin and a dictionary of possible known chaotic models, find the model that corresponds to the signal. That is, we address the problem of classifying chaotic signals. Furthermore, the evaluation of the confidence level that can be associated with the classification result is also considered. A solution

Christophe Couvreur is a Research Assistant of the Belgian National Fund for Scientific Research (F.N.R.S.). He is also currently a Visiting Scholar with the Coordinated Science Laboratory of the University of Illinois at Urbana-Champaign.

based on powerful results of ergodic theory of dynamical systems [11] will be proposed.

This paper is organized as follows. In section 2, the ergodic theory of chaotic systems is briefly reviewed. The classification problem is formally stated in section 3, and a solution is proposed in section 4. Simulation results are presented in section 5. Concluding remarks are made and directions for future research are suggested in section 6.

2. CHAOS AND ERGODICITY

This section provides a brief informal review of the necessary theoretical concepts about chaotic systems and ergodicity. A more rigorous treatment can be found in the literature, e.g., in [11] or in [12].

2.1. Chaos and Discrete-Time Dynamical Systems

In our development, we consider discrete-time signals $y[n]$ with the following state-space description:

$$\mathbf{x}[n] = \mathbf{F}(\mathbf{x}[n-1]) \quad (1)$$

$$y[n] = h(\mathbf{x}[n]), \quad (2)$$

where $\mathbf{x}[n] \in \mathbf{R}^M$, $\mathbf{x}[0]$ denotes the initial conditions, $\mathbf{F}(\cdot)$ is a M -dimensional nonlinear transformation, and $h(\cdot)$ is a real-valued, possibly nonlinear, function of \mathbf{R}^M . This model can cover an extremely broad class of signals. For instance, by Taken’s embedding theorem [1] it is known that a state space vector representation $\mathbf{x}[n]$ can be constructed for many partially observed chaotic systems from a scalar time series $y[n]$ by the delay method: let

$$\mathbf{x}[n] = (y[n], y[n-1], \dots, y[n-M+1])',$$

with M adequately chosen, and let $h(\mathbf{x}) = (1, 0, \dots, 0)' \mathbf{x}$. In this case, the representation of a time series $y[n]$ in terms of the state and observation equations (1)–(2) is equivalent to the representation in terms of the (nonlinear) prediction equation

$$\begin{aligned} y[n] &= f(y[n-1], y[n-2], \dots, y[n-M]) \\ &= f(\mathbf{x}[n-1]) \end{aligned} \quad (3)$$

Chaos theory is concerned with the asymptotic behavior of a subclass of the dynamical systems defined by the state equation (1), *viz.* autonomous dissipating systems [2]. These systems exhibit a long-term behavior corresponding to the convergence of the state space trajectories $\mathbf{x}[n]$ to a subset A of the state space

called an *attractor*. Aperiodic (or chaotic) asymptotic behavior corresponds to *strange* attractors which have generally a fractal structure.

2.2. Statistical Interpretation and Ergodic Properties

A stochastic process $\{\mathbf{x}[n]\}$ can be associated to a deterministic recurrence \mathbf{F} by considering the initial condition $\mathbf{x}[0]$ as a random variable (r.v.) with a given distribution μ_0 . Let μ_n denote the distribution corresponding to the random variable $\mathbf{x}[n]$. Clearly, μ_n depends on μ_0 and on the mapping \mathbf{F} . The Frobenius-Perron operator \mathcal{F} associated with \mathbf{F} is defined as the operator in the space of probability distribution that maps μ_{n-1} to μ_n , that is,

$$\mu_n = \mathcal{F}\mu_{n-1}. \quad (4)$$

Under some regularity and smoothness conditions, the Frobenius-Perron operator \mathcal{F} will admit a fixed point μ called the *invariant distribution*,

$$\mu = \mathcal{F}\mu. \quad (5)$$

The invariant distribution is related to the attractor A by $\mu(A) = 1$. When μ_0 is an invariant distribution, it may be shown that the resulting stochastic process $\{\mathbf{x}[n]\}$ is stationary and, subject to certain constraints on \mathbf{F} , ergodic. In this case, Birkhoff's ergodic theorem can be used to establish the equivalence of time-average and ensemble-average for almost all trajectories, i.e.,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} g(\mathbf{F}^{(n)}(\mathbf{x}_0)) = \int g(\mathbf{x}) d\mu \quad \text{a.s.}(\mu), \quad (6)$$

where $\mathbf{F}^{(n)}$ denote the n -th iteration of the map \mathbf{F} and $g: \mathbf{R}^M \rightarrow \mathbf{R}$ is any integrable function. When μ_0 is not an invariant density, $\{\mathbf{x}[n]\}$ is no longer stationary but may still be ergodic.

Furthermore, because of the sensitivity to initial conditions of chaotic systems, it may be shown that the stochastic process $\{\mathbf{x}[n]\}$ possess an exponentially fast mixing property [12]. It follows that a central limit theorem can be stated for the time-average properties of functions of chaotic time series. That is, the time-averaged r.v.'s on the left hand side of (6) converges in law to a normal distributions when $N \rightarrow \infty$,

$$\frac{1}{N} \sum_{n=0}^{N-1} g(\mathbf{F}^{(n)}(\mathbf{x}_0)) \xrightarrow{L} \mathcal{N}(m, \sigma/\sqrt{N}), \quad (7)$$

with

$$m = \int g(\mathbf{x}) d\mu \quad (8)$$

and

$$\begin{aligned} \sigma^2 &= \int (g(\mathbf{x}) - m)^2 d\mu \\ &+ 2 \sum_{\ell=1}^{\infty} \int (g(\mathbf{x}) - m)(g(\mathbf{F}^{(\ell)}(\mathbf{x})) - m) d\mu. \end{aligned} \quad (9)$$

The exact conditions on \mathbf{F} under which (6) and (7) hold are quite technical and still a subject of active research in ergodic theory today. In this communication, we assume that the class of chaotic models considered possess the properties (6) and (7) without discussion.

3. PROBLEM STATEMENT

Let $\mathbf{y} = (y[1], \dots, y[N])$ be a length- N sample of a chaotic signal. Let $\Omega = \{\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_K\}$ be a set of K possible chaotic signal models for the dynamic of $y[n]$, or, alternately, let $\Omega = \{f_1, f_2, \dots, f_K\}$ be the set of K predictors (3) associated with the state-space models \mathbf{F}_k . The following problem is addressed: given the sample of $y[n]$, which of the models from Ω generated the signal?

The models in Ω can be obtained in one of the two following fashion. If a physical model of the process generating the signal is available, (1) and (2) are known explicitly. If only examples of the time series $y[n]$ are available, a nonlinear predictor f_k can be constructed [13].

At first glance, the classification problem may seem trivial. Indeed, let $\text{MSE}(\mathbf{F}_k)$ be the mean-squared prediction error associated with the model \mathbf{F}_k for the signal $y[n]$, i.e.,

$$\text{MSE}(\mathbf{F}_k) = \frac{1}{N-M} \sum_{n=M}^{N-1} (y[n+1] - f_k(\mathbf{x}[n]))^2, \quad (10)$$

and let \mathbf{F}_* (or f_*) denote the true dynamic of the observed system. Clearly, $\text{MSE}(\mathbf{F}_*) = 0$ since the systems considered are purely deterministic. One may thus suggest the following classification scheme:

1. compute $\text{MSE}(\mathbf{F}_k)$ for all the models $\mathbf{F}_k \in \Omega$
2. assign to the signal $y[n]$ the model whose MSE is equal to zero.

In practice, however, there will always be some model mismatch between the data and the models in the dictionary (i.e., $\mathbf{F}_* \notin \Omega$) and none of the MSE's will be strictly equal to zero. It might still be suggested to assign to the signal $y[n]$ the "best" or "closest" model from Ω . A possible definition of the "best" or "closest" model in Ω is the model that yields the smallest prediction error for the signal. The classification scheme would then become

1. compute $\text{MSE}(\mathbf{F}_k)$ for all the models $\mathbf{F}_k \in \Omega$
2. assign to the signal $y[n]$ the model whose MSE is the smallest, i.e., $\arg \min_{\Omega} \text{MSE}(\mathbf{F}_k)$.

Since the MSE's are empirical values computed from a chaotic signal, they will show some "random" variations from one realization to another. How will these variations affect the classification process? Is there a way to "quantify" the effect of the variations on the reliability of the classifications? It is possible to answer these questions by resorting to the statistical interpretation of chaotic systems of section 2.

4. PROPOSED SOLUTION

Let f_* denote the predictor associated with \mathbf{F}_* . We have

$$\begin{aligned} \text{MSE}(\mathbf{F}_k) &= \frac{1}{N-M} \sum_{n=M}^{N-1} (f_k(\mathbf{x}[n]) - y[n+1])^2 \\ &= \frac{1}{N-M} \sum_{n=M}^{N-1} g_k(\mathbf{x}[n]), \end{aligned} \quad (11)$$

where $g_k(\mathbf{x}) = (f_k(\mathbf{x}) - f_*(\mathbf{x}))^2$. Applying the ergodic theorem (6) to (11), we can deduce that, for large N , $\text{MSE}(\mathbf{F}_k)$ will tend to an asymptotic value defined by

$$\begin{aligned} \text{AMSE}(\mathbf{F}_k) &= \lim_{N \rightarrow \infty} \text{MSE}(\mathbf{F}_k) \\ &= \int g_k(\mathbf{x}) d\mu. \end{aligned} \quad (12)$$

Furthermore, by the central limit theorem (7), $\text{MSE}(\mathbf{F}_k)$ can be considered as a Gaussian r.v. with a $\mathcal{N}(m_k, \sigma_k/\sqrt{N-M})$ distribution where $m_k = \text{AMSE}(\mathbf{F}_k)$ and σ_k is given by (9).

Let us assume that the pairs (m_k, σ_k) , $k = 1, \dots, K$ are known. Given the set of empirical prediction errors $\text{MSE}(\mathbf{F}_k)$, deciding on a model for the signal $y[n]$ could then be viewed as a statistical decision problem and standard statistical tests could be applied (an example will be given in section 5).

Unfortunately, the pairs (m_k, σ_k) are not known in practice. Indeed, the values of (m_k, σ_k) depend on f_* which is unknown. Taking advantage of the ergodic properties of the models considered, it may be suggested using the empirical values of m_k and σ_k . First, $\text{MSE}(\mathbf{F}_k)$ itself can be viewed as an estimate of m_k . Note that this is equivalent to replacing the ensemble-average in (8) with the time-average. Similarly, σ_k could be obtained by replacing the ensemble-averages by the time-averages in (9) and truncating the convergent series to L terms, with $L \ll N - M$.

We propose an alternative approach. From the length- N sample \mathbf{y} , it is possible to compute a non-parametric predictor \tilde{f}_* using any adequate non-parametric technique (e.g., a neural network). This predictor can be viewed as an approximation of the true predictor f_* . Given an initial conditions \mathbf{x}_0 and the approximate predictor \tilde{f}_* , it is possible to construct a time series $\{y[n]\}$ of any desired length simply by applying recursively (3). Let $\tilde{\mathbf{y}} = (\tilde{y}[1], \tilde{y}[2], \dots, \tilde{y}[T])$ be a length T realization obtained in this way. Let \tilde{m}_k and $\tilde{\sigma}_k^2$ be defined by (8) and (9) where $g_k(\mathbf{x}) = (f_k(\mathbf{x}) - f_*(\mathbf{x}))^2$ has been replaced by $\tilde{g}_k(\mathbf{x}) = (f_k(\mathbf{x}) - \tilde{f}_*(\mathbf{x}))^2$. The easiest way to compute to evaluate \tilde{m}_k and $\tilde{\sigma}_k^2$ is to use $\tilde{\mathbf{y}}$ to compute time-averages. That is, to use

$$\tilde{m}_k = \frac{1}{T-M} \sum_{t=M}^{T-1} \tilde{g}_k(\tilde{\mathbf{x}}[t]) \quad (13)$$

and

$$\begin{aligned} \tilde{\sigma}_k^2 &= \frac{1}{T-M} (\tilde{g}_k(\tilde{\mathbf{x}}[t]) - \tilde{m}_k)^2 + 2 \sum_{\ell=1}^L \frac{1}{T-M-\ell} \\ &\times \sum_{t=M}^{T-\ell} (\tilde{g}_k(\tilde{\mathbf{x}}[t]) - \tilde{m}_k)(\tilde{g}_k(\tilde{\mathbf{x}}[t+\ell]) - \tilde{m}_k). \end{aligned} \quad (14)$$

Since T can be made arbitrarily large, it is possible to obtain values of \tilde{m}_k and $\tilde{\sigma}_k^2$ with any desired precision. For instance, L can be chosen greater than N .

If \tilde{f}_* is a good approximation of f_* , it can be hoped that \tilde{m}_k and $\tilde{\sigma}_k^2$ will be close to m_k and σ_k^2 and they can be used instead of the true values in the statistical models of the MSE's. In fact, the bias/variance trade-offs of \tilde{m}_k and $\tilde{\sigma}_k^2$ will depend only on the quality of the estimated predictor \tilde{f}_* . In our experiments, we have found that using this approach based on the construction of a predictor as an intermediate step before the evaluation of m_k and σ_k^2

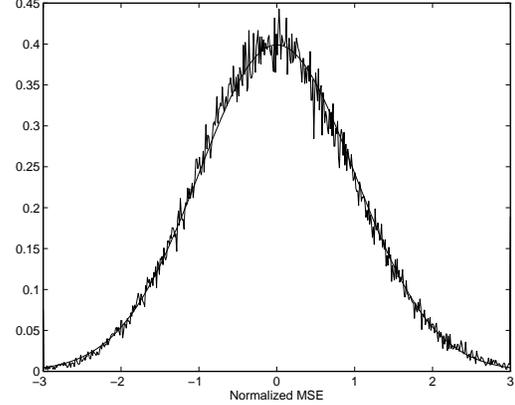


Figure 1: Normalized histogram of the MSE and $\mathcal{N}(0, 1)$ distribution.

yield better results than the direct evaluation by the empirical mean and variances.

Remark: The method that has been proposed for the estimation of the means m_k and variances σ_k^2 can be viewed as a kind of “bootstrap” estimation: the original sample \mathbf{y} is used to generate a larger sample $\tilde{\mathbf{y}}$ which is then used to estimate some parameters related to the original sample. This is especially obvious if “nearest-neighbors” types of predictors [14] are used for \tilde{f}_* .

5. PRELIMINARY RESULTS

A preliminary series of simulations has been conducted in order to validate the proposed approach. In these preliminary experiments, a standard 1-D chaotic models is used: the logistic equation. The logistic model is defined by the recurrence/predictor

$$f : [0, 1] \rightarrow [0, 1] : x[n+1] = \lambda x[n](1-x[n]), \quad (15)$$

for some $\lambda \in (0, 4)$. Note that for this simple model, it is not known if the ergodic theorem (6) and the central limit theorem (7) apply. However, this can be verified experimentally. All the simulations described below were conducted in MATLAB.

Let us assume that a time series $\{y[n]\}$ is generated by a logistic model f_* with $\lambda_* = 3.8$ and that it is attempted to predict this time series with a logistic model f with parameter $\lambda = 3.7$. Figure 1 gives the histogram of the distribution of the MSE for the prediction of samples of length $N = 1000$ of the signal. This histogram has been obtained as follows. First, 100 000 independent initial conditions $x[0]$ where drawn from a uniform distribution on $[0, 1]$. Each initial condition was then used with (15) to generate a signal $x[n]$ of length 1100 and the first 100 values of $x[n]$ were dropped in order to get rid of transients and to insure that the chaotic model is evolving “on” its attractor (i.e., that the distribution of the $x[n]$ for a given n is the stationary distribution μ). For each of the 100 000 samples of length 1000, $\text{MSE}(f)$ was evaluated. Finally, the histogram of figure 1 was computed. The distribution appears clearly Gaussian. This can be confirmed by a Kolmogoroff-Smirnoff test [15]. It can be further verified [15] that the values of the mean and the variance of the distribution of $\text{MSE}(f)$ are the ones given by (8) and (9) with $g(x) = [(\lambda - \lambda_*)x(1-x)]^2$. The ensemble-averages can be obtained by

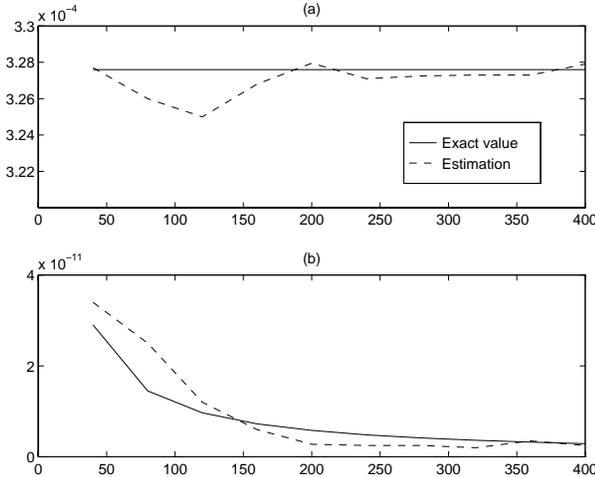


Figure 2: True and estimated MSE mean (a) and MSE variance (b).

Monte-Carlo integration using the samples of $x[1000]$ [15]. In our case, we get $m = 3.276 \times 10^{-4}$ and $\sqrt{N}\sigma = 3.407 \times 10^{-5}$. Note that with the simple model considered here, it is also possible to show that $m \propto (\lambda - \lambda_*)^2$ and $N\sigma^2 \propto (\lambda - \lambda_*)^4$. Again, these relations can be verified numerically.

Next, we experimented with the approach proposed in section 4. In this series of experiment, the true model was again $\lambda_* = 3.8$ and the “supposed” model was $\lambda = 3.7$. An approximate predictor \hat{f}_* was constructed using a Radial Basis Function (RBF) Network [16]. Figure 2 presents the values of the MSE mean estimator \hat{m}_k and of the MSE variance estimator $\hat{\sigma}_k^2$ obtained from (13) and (14) for a typical realization of $x[n]$, for several lengths N . The figure also includes the theoretical value of m_k and σ_k^2 obtained from (8) and (9). It can be verified that the proposed method yields a pretty good approximation of m_k and σ_k^2 .

The following simple application will illustrate the proposed method. Let us assume that a dictionary of two logistic models is available, $\Omega = \{\lambda_1 = 3.8, \lambda_2 = 3.7\}$. Let $(x[1], \dots, x[50])$ be a sample of generated from the logistic equation with $\lambda = 3.72$. Using the above method, we get: $\hat{m}_1 = 2.2 \times 10^{-4}$, $\hat{\sigma}_1 = 3 \times 10^{-6}$, $\hat{m}_2 = 1.4 \times 10^{-5}$, $\hat{\sigma}_2 = 1.9 \times 10^{-7}$. It is clear that with these values, the probability of classification error is negligible (can be evaluated at $< 0.005\%$).

Other results will be shown at the conference.

6. CONCLUDING REMARKS

Deterministic chaotic signals and stochastic processes have traditionally been viewed as alternative ways of looking at “randomness.” Thanks to advances in ergodic theory, it becomes now possible to adopt a unified view on both approaches. The statistical interpretation of chaotic properties allow standard statistical tools to be used with with low-dimensional deterministic chaotic signals. A simple example is the classification application that has been proposed here.

However, the approach described in this paper relies on a purely deterministic model of the signal. In practice, there will always

be some “randomness” in the signal that can not be explained by a purely deterministic model, even a chaotic one. For instance, there will always be some observation noise in (2). This will make the estimation of the dynamical model and the interpretation of its properties more difficult. Trying to include stochastic observation noise in the model would certainly improve its usability.

The effect of the approximation error between \hat{f}_* and f_* on the estimation of m_k and σ_k^2 also deserves further attention.

7. REFERENCES

- [1] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer-Verlag, Berlin, 1996.
- [2] J. Awrejcewicz, *Bifurcation and Chaos*, Springer-Verlag, Berlin, 1995.
- [3] Special Issue on Chaos in Nonlinear Electronic Circuits, *IEEE Trans. Circ. and Syst.-I*, vol. 40, no 10 & no 11, oct. & nov. 1993.
- [4] E. Ott, T. Sauer, and J. A. Yorke, *Coping with Chaos: Analysis of Chaotic Data and The Exploitation of Chaotic Systems*, John Wiley & Sons, New York, 1994.
- [5] A. V. Oppenheim, G. W. Wornel, S. H. Isabelle, and K. M. Cuomo, “Signal processing in the context of chaotic signals”, *Proc. ICASSP*, San Francisco, CA, March 1992.
- [6] K. Vibe and J.-M. Vesin, “On chaos detection methods”, *Int. J. Bifurcation and Chaos*, vol. 6, no. 3, pp. 529–543, 1996.
- [7] H. Dedieu, M. P. Kennedy, and M. Hasler, “Chaos shift keying: Modulation and demodulation of a chaotic carrier using self-synchronizing Chua’s circuits”, *IEEE Trans. Circ. and Syst.-II*, vol. 40, no. 10, pp. 634–642, Oct. 1993.
- [8] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and Th. Schreiber, “On noise reduction methods for chaotic data”, *Chaos*, vol. 3, no. 2, pp. 127–141, 1993.
- [9] S. M. Kay and V. Nagesha, “Methods for chaotic signal estimation”, *IEEE Trans. Sig. Proc.*, vol. 43, no. 8, pp. 2013–2016, Aug. 1995.
- [10] S. Haykin and X. B. Li, “Detection of signals in chaos”, *Proc. IEEE*, vol. 83, no. 1, pp. 95–122, Jan. 1995.
- [11] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, Springer-Verlag, New York, second edition, 1994.
- [12] J. L. Jensen, “Chaotic dynamical systems with a view towards statistics: A review”, in *Networks and chaos – Statistical Aspects and Probabilistic Aspects*, O. E. Barndorff-Nielsen, J. L. Jensen, and W. S. Kendall, Eds., chapter 4, pp. 201–250. Chapman & Hall, London, 1993.
- [13] M. Casdagli, “Nonlinear prediction of chaotic time series”, *Physica D*, vol. 35, pp. 335–356, 1989.
- [14] M. Casdagli, “Chaos and deterministic versus stochastic nonlinear modelling”, *J. Roy. Stat. Soc. B*, vol. 54, no. 2, pp. 303–328, 1991.
- [15] C. Flamme, “Classification des systèmes chaotiques: approche non paramétrique par réseaux de fonctions à base radiale”, Undergraduate thesis, Faculté Polytechnique de Mons, Mons, Belgium, Jan. 1997, in French.
- [16] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan, New York, 1994.