Approximating the Discrete State Equation from Chaotic noisy data

A. I. Fernández
Data Mining Department,
Altran SDB,
Ramírez de Arellano 15,
Madrid 28043, Spain.
ifernan@altransdb.com

L. Sánchez
Computing Department,
University of Oviedo,
Edif. Dep., Campus de Viesques,
33204 Asturias, Spain.
luciano@lsi.uniovi.es

J. J. Navarro
Data Mining Department,
Altran SDB,
Ramírez de Arellano 15,
Madrid 28043, Spain.

Abstract

Due to the inherent complexity of many real word systems, the application of traditional system identification techniques is limited. In such instance, more sophisticated modeling approaches are required. In this work, a method for building an *hybrid GA-P predictor* from physical system measurements of *chaotic nature* is proposed. Our ultimate goal is to build a reliable input flow predictor of blast furnace gas.

The main objective is to approximate the state equation and the initial state of the dynamical system. This way, the discrete equations have physical meaning and may readily be interpreted.

In this method, modeling is applied in order to minimize the deviation of predictions from *noise-free data*. It is important to notice that we talk about noise-free data instead of data cleaned by a filtering process.

Keywords: Chaos, GA-P, State Equation, noisy data.

1 Introduction.

In previous works, we developed a methodology to identify *predictors without physical interpretation* from chaotic physical system measurements [2] [3]. Fuzzy modeling and neural network techniques were applied to raw data in order to learn the *embedding function* [14] that reconstructs the underlying dynamics of the system.

This paper focuses on the reconstruction of the state equations which provide us a method for obtaining *models with physical interpretation*. Owing to this interpretability, it's possible to know dynamical properties of the physical system, such us the state space dimension, or to spot incorrect chaos detection, due to noise influence on the numerical algorithms of the test. Furthermore, the dimension of this kind of models, m, is less than the embedding dimension, D_E, [14]

$$2 * m + 1 \leq D_E,$$

so noise robustness is better.

This paper will proceed as follows. Firstly, we will explain both phases of the identification: characterization and model construction.Secondly, we will validate the methodology with numerical results and it will be applied to a real case, the blast furnace gas supply prediction. Finally, we will end with some conclusions of the work.

2 System Identification.

This method is indicated for obtaining the *state equations* of the system from *chaotic observations*, so it is applicable to know the dynamics properties of real physical systems.

2.1 System Characterization.

We applied the chaos-based test, developed by Haykin and Xiao Bo Li [7], to system classification. This test is preferred to others [9] [11] [12] etc. because provide us additional information about how far into the future a reliable prediction can be made. Because chaotic systems are sensitive to initial conditions, it is not possible to make a reliable prediction for any horizon length [1] [2].

The numerical algorithms used are affected by both, noise presence and filtering techniques. In order to minimize these effects, this test must be applied to *raw data* measured by the sensor, without any previous filtering that spoil the chaotic underlying dynamics [3] [4]. Therefore, a possitive test result could be obtained from a non chaotic system. If dynamics are reconstructed by a model with physical interpretation, this mistake will be detected, see Section 3.1.2.

2.2 Model Construction.

Model construction is an optimization problem. If we consider all the models which present a defined structure, we have a set of possible solutions for the problem, the search space. If we define a function that measure the goodness of each point of the space, the fitness function, the problem consists on finding the point that minimizes or maximizes the value of the fitness function.

Practical results also show us that learning from *raw signals* is preferred to *cleaned signals* [3].

2.2.1 Model Structure.

After the definition of the chaos test, the next problem is to find an appropriate model structure for chaos dynamics. It is accepted that when the *embedding dimension*, D_E, is chosen, the dynamics of a chaotic system can be reconstructed using the deterministic model [14]

$$y(n+D_E) = f(y(n), y(n+1), \ldots, y(n+(D_E-1))). \quad (1)$$

Since we are using vectors derived from observations, *time delayed* coordinates are used. The time delay value, D_T, is chosen by the Fraser and Swinney algorithm [5] by minimal mutual information method.

Now, we know the existence of a nonlinear model for our data

$$\widetilde{y}(n+D_E D_T) = \widetilde{f}(y(n), y(n+D_T), \ldots, y(n+(D_E-1)D_T)), \quad (2)$$

whose input size range defines a family of predictors [2]

$$\widetilde{D}_c \leq D_E \leq 2\widetilde{D}_c + 1, \quad (3)$$

where $\widetilde{D}_c$ is the estimation of the correlation dimension, which can be obtained from a time series by applying the Grassberger-Procaccia algorithm [6].

On the other hand, a chaotic system is an autonomous dynamical system. Therefore, it can be defined

$$\begin{array}{l} \frac{d\,\vec{x}}{d\,t} = f(\vec{x}) \\ y = g(\vec{x})\ . \end{array}$$

$\vec{x} \in \Re^m$ is the state vector of a system of dimension m, $y \in \mathbb{R}$ is the system output, $f : \Re^m \to \Re^m$ is the *state function* and $g : \Re^m \to \mathbb{R}$ is the *output function.*

Although the real system can be continuous, time series $\{y\}$ are obtained by sampling the output y

$$k \mapsto y_k = g(\vec{x}),$$

making a *discrete* system, whose state equation is

$$\begin{array}{l} \vec{x}_{k+1} = f(\vec{x}_k) \\ y_{k+1} = g(\vec{x}_{k+1}). \end{array} \quad (4)$$

It is also possible to rewrite the state equation as a family of m equations

$$\begin{array}{l} x_{1\,k+1} = f_1(x_{1\,k}, \ldots, x_{m\,k}) \\ x_{2\,k+1} = f_2(x_{1\,k}, \ldots, x_{m\,k}) \\ \ldots \\ x_{m\,k+1} = f_m(x_{1\,k}, \ldots, x_{m\,k}) \\ y_k = g(x_{1\,k}, \ldots, x_{m\,k}). \end{array} \quad (5)$$

We need of a learning technique that obtain these equations from physical observations $\{y_k^*\}$.

The model will have the same invariants of the dynamics as the physical system and the deviation of the artificial time series generated by the iterated evaluation of the model from *noise-free* data will be minimum. As the model will have physical interpretation, it will have the advantages enumerated in the introduction of the paper, see Section 1.

2.2.2 Learning technique.

In order to minimize the deviation of predictions from the noise-free signal, the train set is obtained by an uniform sampling of the time series measured by the sensor, without any previous filtering [4].

The learning of chaos dynamics presents some particular problems: Firstly, physical system observations assure us noise presence and filtering techniques spoil the underlying dynamics [15]. Secondly, the fitness function associated to a predictive chaotic model presents a lot of local minima [10]. So, the optimization technique must be chosen in order to obtain *a solution representation* that handles non-linear analytical expressions and a *searching algorithm* that:

1. can use non derivable fitness functions, even non continuous, as complex as needed to eficiently explore the search space,
2. with a good generalization capacity to manipulate noisy data that can not be filtered.

We use a GA-P learning algorithm [8] hybridated with the Simplex method by Nelder and Mead [13]. We are going to explain the *representation* and the defined *fitness function*, which exhibits the conditions we explained above.

Solution representation. The *representation* of the individual that models the solution is shown in Figure 1.

$$
\begin{aligned}
x_{1\,k+1} &= f_1(\vec{x}_k) = c_1 + (c_1 - x_{1\,k}) + c_4\\
x_{2\,k+1} &= f_2(\vec{x}_k) = x_{3\,k}\\
x_{3\,k+1} &= f_3(\vec{x}_k) = x_{2\,k} * c_2\\
y_{k+1} &= g(\vec{x}_{k+1}) = (c_1 + (c_1 - x_1) + c_4) * (x_{3\,k}) * (x_{2\,k} * c_2).
\end{aligned}
\quad (6)
$$

The *state equations* are codified as a GA-P typical individual [8]. This representation consists on a symbolic part represented by a tree, GP part, and a numeric part, the constant set that can appear in the expression, GA part. The GA-P individual described is completed in our method with a vector of m components, that codifies the *initial state*.

Using this representation we can handle all the elements of the state equation structure:

- each *state function* f_i is representd by a subtree of the root,
- the *current vector state components*, $\vec{x}_k$, are tree leaves,
- the *next vector state components* , $\vec{x}_{k+1}$, will be obtained by the evaluation of root subtrees f_i, i. e., $x_{i\,k+1} = f_i(\vec{x}_k)$,
- the *scalar output* $k+1$ will be calculated by evaluating the whole tree, i. e., by evaluating nodes labeled g over the components $y_{k+1} = g(\vec{x}_{k+1}) = g(f_1(\vec{x}_k), \ldots, f_m(\vec{x}_k))$,
- and the *intial state* $\vec{x}_0$ is a vector of m components calculated by the Simplex method [13] in order to minimize the deviation of the artificial time series generated by the model using iterative prediction from system observations.

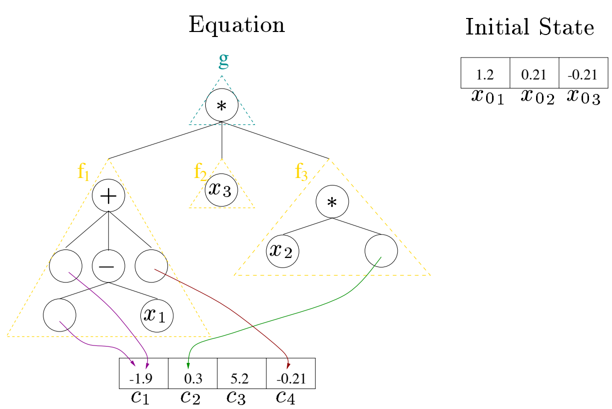


Figure 1: Individual that codifies Equations 6.

Fitness Function Nonlineear behaviour, especially chaos, causes similar models to produce very different approximation errors, given that a highly irregular error surface hides the global minimum and many good local minima to the search algorithm. In order to learn the analytical model expression of the *state function* avoiding local minima, a measurement of the deviation of two dynamics invariants estimations of the model from the invariant estimations of the physical system are included in the fitness fucntion. This two invariants are the correlation dimension, D_c, and the largest lyapunov exponent, λ_1. Therefore, in order to obtain a measure of the error of the model in the iterated prediction of an artificial time series, we have included in the fitness function the expression

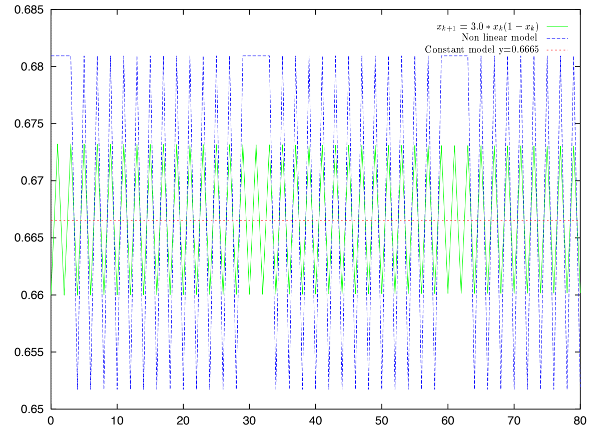


Figure 2: Models constructed from the same training set, the logistic map $x_{k+1} = 3.0 * x_k(1 - x_k), x_0 = 0.5$. The search for the optimum converges prematurely when the fitness function does not include the invariant measurement because the prediction error is lower.

$$
D_{inv} = \sqrt{(1 - \frac{\widetilde{D}_{c\,sol}}{\widetilde{D}_{c\,sis}})^2 + (1 - \frac{\widetilde{\lambda}_{1\,sol}}{\widetilde{\lambda}_{1\,sis}})^2}, \quad (7)
$$

where $\widetilde{D}_{c\,sis}$ is the estimation obtained from physical system observations and $\widetilde{D}_{c\,sol}$ is the value estimated from the artificial time series generated by the model.

As we mentioned, the numerical algorithms for the invariant estimation are affected by noise present in the observations [6] [16]. Whether the objective is to obtain a model of a noise-free physical system, we allow some degree of tolerance in the invariant error. The objective is to reduce the invariant error under the value

$$
D_{inv\,umbral} = \text{máx}\{0.2, \sqrt{\widetilde{D}_{c\,sis}^{\,2} + \widetilde{\lambda}_{1\,sis}^{\,2}}\}. \quad (8)
$$

The minimization of this value avoids the premature convergence to local minima and improves the efficiency of the search. This criterium discriminates between models with a good prediction error but very different dynamics, and characterizes some local minima, as can be seen in Figure 2.

Apart from the invariant measurement, the fitness function includes two additional criteria for optimizing. The best model of those with similar invariants is the one that produces the lowest iterated prediction error

$$RMSE = \sqrt{\frac{1}{k * NTrain} \sum_{i=1}^{NTrain} \sum_{j=1}^{k} [(n_1 * \widetilde{y}_{ij} + n_2) \ - y_{ij}^*]^2}. \tag{9}$$

The vector $(y_{i1}^*, \ldots, y_{ik}^*)$ is the training vector obtained from system measurements. Vector components $(n_1 * \widetilde{y}_{i1} + n_2, \ldots, n_1 * \widetilde{y}_{ik} + n_2)$ are values of the artificial time series generated by the model. Again, in order to avoid noise effects, we calculate the value of the prediction error introduced by noise level [4]

$$\begin{aligned} RMSE_{noise} &= \sqrt{\tfrac{1}{M} \textstyle\sum_{i=1}^{M} (y_i^* - \widetilde{y}_i)^2} \\ &= \sqrt{\tfrac{1}{M} E_{noise}} \\ &= \sqrt{\tfrac{1}{M} \textstyle\sum_{i=1}^{M} (y_i^*)^2 * \% noise}. \end{aligned} \tag{10}$$

Finally, when the invariants are similar and the $RMSE$ value is under the limit determined by the noise, the best model is the one that represents the simplest expression, i.e., the state equation codified by the tree with the minimum number of nodes.

The complete definition of the multiobjective fitness function, the evaluation of the each possible solution of the searching space is as follows.

Maximum priority:	$D_{inv} < D_{inv_{umbral}}$
Medium priority:	$RMSE \leq RMSE_{noise}$
Minimum priority:	minimun number of tree nodes

3 Numerical Experiments.

3.1 Method Validation.

As imprecision is inherent to sensors, noise-free measurements are not available. Therefore, the method will be validated using theoretical time series with added noise.

3.1.1 Chaotic system.

Several trials were conducted using the Henon map. A theoretical time series is necessary because model evaluation requires the measurement of the deviation of predictions from *noise-free* data. These information can not be obtained from physical systems.

State equations for the Henon map, the theoretical system, are

$$\begin{aligned} x_{n+1} &= y_n + 1 - 1.4 * x_n^2 \\ y_{n+1} &= 0.3 * x_n. \end{aligned}$$

This is a chaotic system of dimension two. The level of added noise is 10%. Time series with and without noise are represented in Figure 3.

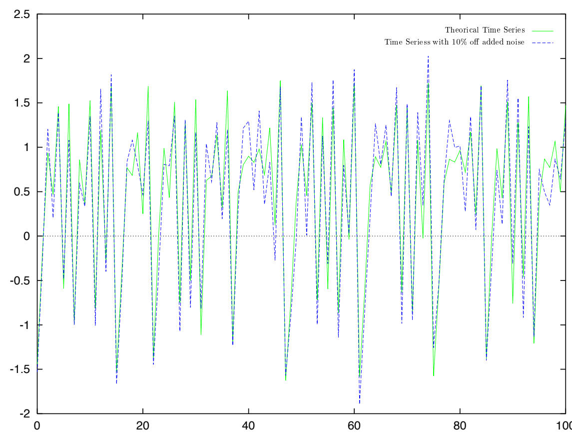


Figure 3: Henon time series with a 10% of noise level.

With a population of 200 individuals during 201 generations, with probablities of 80%, 80%, 4% and 1% for the genetic operators GA crossing, GP crossing, reproduction and mutation respectively, and tounaments with 6 individuals with replacement as selection method, the GA-P algorithm hybridated with Simplex constructs the state equation:

$$\begin{aligned} \Delta x_1 &= 0.322 * x_2 + 0.922 \\ \Delta x_2 &= 0.89 * x_1 - 1.45 * x_2^2 \\ \Delta x_3 &= 1.07 \\ \Delta y &= 3.6 * \Delta x_1 - 3.016, \end{aligned}$$

a *two dimensional model* of chaotic nature. The results are summarized at Table 1.

We obtained a chaotic model with a state space of two dimensions. The invariant estimations obtained from predictions are more accurate than estimations from observations.

3.1.2 Non Chaotic system.

We use a linear system defined by the sequence

$$\{3.6, -3.1, 3.6, -3.1 \ldots\}, \tag{11}$$

that is generated by the linear system

$$x_{k+1} = -x_k + 0.5. \tag{12}$$

With a 10% of added noise, the observed system is represented in Figure 5(a) and Figure 5(b).

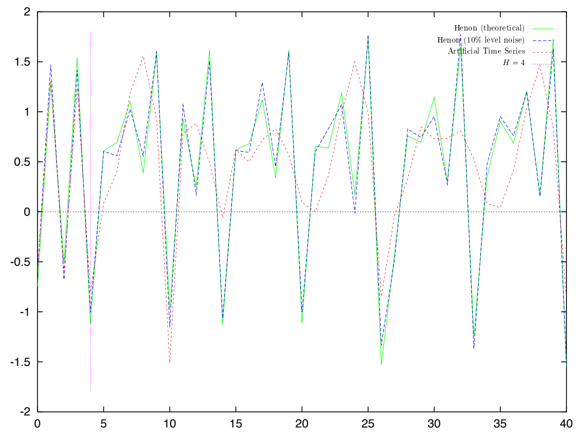
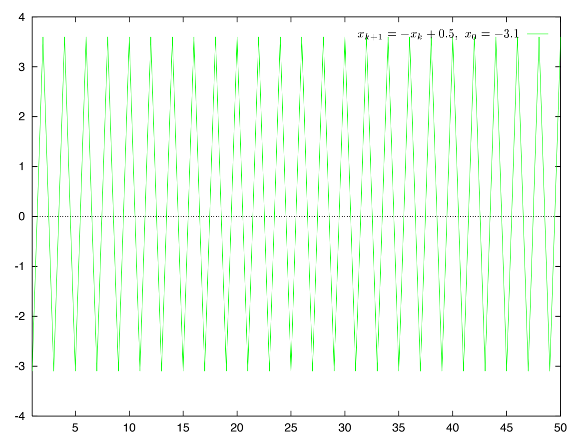
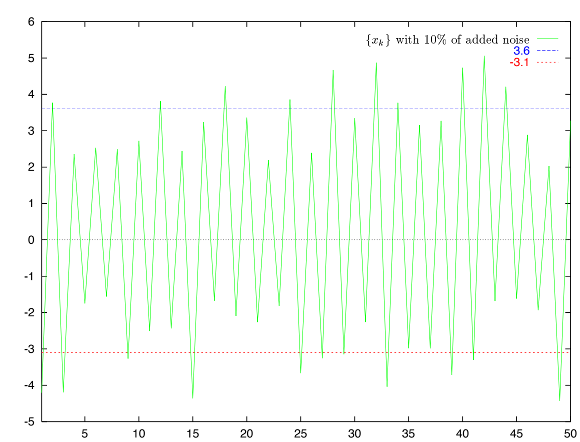


Figure 4: Prediction of the genetic model and the theoretical Time Series of the Henon map.



(a) Without noise



(b) With noise

Figure 5: Measurements of the system to model.

The result of the Haykin test is a possitive chaotic detection due to the noise. Therefore, the proposed method is applied, and results are summarized at Table 2.

We can conclude that the initial characterization was wrong, because the model is linear.

3.2 Practical Case: Input flow predictor of blast furnace gas.

The method is now applied for modelling of a real physical system.

The blast furnace-gas and CO-gas produces by a metallurgical plant is burnt in the thermal plant's gas boilers of an electricity company. The metallurgical plant will be reconstructed, resulting in a doubling of the available blast furnace gas. Because of these changes, the flow to the thermal plant will be increased and the boilers have a limited capability to follow the fluctuations. A predictive control is really interesting to achieve a proper operation. Therefore some anticipation to the future supply is needed. The Figure 6 shows input flow measurements.

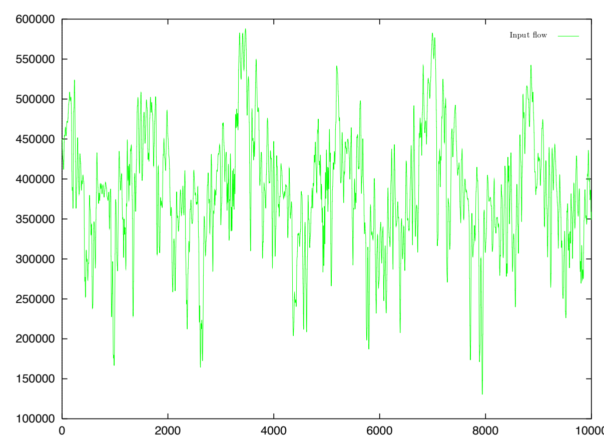


Figure 6: Input supply of blast furnace gas. Sampling time = 1 minute.

The state equations obtained for this system are

$$\begin{aligned}
\Delta x_1 &= -2.9 \\
\Delta x_2 &= -2.9 - 2.1 * x_2 \\
\Delta x_3 &= 0.98 \\
\Delta x_4 &= 0.67 \\
\Delta x_5 &= 3.48 + 0.27 * x_2 - 0.98(x_2^2 + 1) + 15.7 * x_2^5 - 22.82 * x_2^4 \\
\Delta y &= 20025 * (-1.25 + \Delta x_2 + \Delta x_5) + 41746.
\end{aligned} \tag{13}$$

Results are summarized in Table 3.

4 Conclusions.

As intended, the obtained model presents physical interpretation, as it is defined by the state equations. In

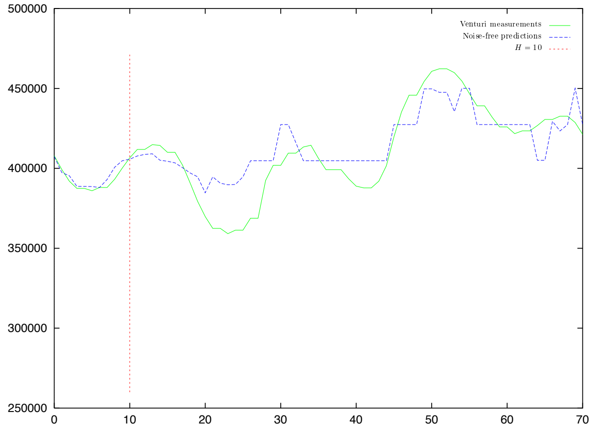


Figure 7: Input flow prediction and observed data.

addition, the dimension of the models is lower than the obtained in embedding function models, so they are more noise insensitive. Finally, using the state equations we can detect wrong system characterizations.

The best invariant estimations are obtained by applying numerical algorithms to the artificial time series generated by the model.

The deviation of predictions from noise-free data is lower than the deviation of noisy data, so predictions are closer to ideal data.

References

[1] M. Casdagli, "Nonlinear prediction of Chaotic time series", *Physica D*, Vol. 35, 1989, pág. 335-356.

[2] A. I. Fernández, Luciano Sanchez, "Prediction of blast furnace gas supply in Chaos", *Proc. of NOLTA'97*, Hawaii, EEUU, pp. 445-448.

[3] A. I. Fernández, Luciano Sanchez, "A Method for Prediction in Chaos by Fuzzy Modeling", *Proc. of IPMU'99*, Paris, France, pp. 403-410.

[4] A. I. Fernández, "Chaotic dymanics system identification from noisy time series by GA-P algorithms", Computing Department PhD. Thesis, June 1999.

[5] A. M. Fraser, H. Swinney, "Independent coordinates for strange attractors", *Physical Review A*, Vol. 33, 1986, pp. 1134-1140.

[6] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors", *Physica D*, Vol. 9, 1983, pp. 189-208.

[7] S. Haykin, Xiao BO LI, "Detection of Signals in Chaos", *Proceedings of the IEEE*, Vol. 83, No. 1, 1995, pp. 95-122.

[8] L. M. Howard, D. J. D'Angelo, "The GA-P: A Genetic Algorithm and Genetic Programming Hybrid", *IEEE Expert/Intelligent Systems & Their Applications*, Vol. 10, No. 3, 1995, pág. 11-15.

[9] D. T. Kaplan, L. Glass, "Direct Test for Determinism in a Time Series", *Physical Review Letters*, Vol. 68, No. 4, 1992, pp. 427-430.

[10] B. Kosko, "Fuzzy Engineering", Prentice Hall, 1997.

[11] W. Marzocchi, F. Mulargia, G. Gonzato, "Detecting low-dimensional chaos in time series of finite length generated from discrete parameter processes", *Physica D*, Vol. 90, 1995, pp. 31-39.

[12] P. V. Mcdonough, J. P. Noonan, G. R. Hall, "A new Chaos Detector", *Computers Elect. Engng.*, Vol. 21, No. 6, 1995, pp. 417-431.

[13] J. A. Nelder, R. Mead, *Computer Journal*, Vol. 7, 1965, pág. 308-313.

[14] F. Takens, "Detecting strange attractor in turbulence", *Dynamical Systems and Turbulence*, Warwick 1980, D. A. Rand, L.S. Young, Lectures Notes in Math, Vol. 898, 1981, pp. 366-381.

[15] Anastasios A. Tsonis, *Chaos, From Theory to Applications*, Plenum Press, 1992.

[16] A. Wolf, J. B. Swift, H. L. Swinney, J. A. Vastano, "Determining Lyapunov exponents from a Time Series", *Physica D 16*,1985, pág. 285-317.

	Theoretical system	Model
Expression	$x_{k+1} = y_k + 1 - 1.4 * x_k^2$ $y_{k+1} = 0.3 * x_k$	$x_{1k+1} = 0.322 * x_{2k} + 0.922$ $x_{2k+1} = 0.89 * x_{1k} - 1.4 * x_2^2$ $y_{k+1} = 3.6 * x_{1k+1} - 3.016$
Dimension	2	2
Linear	No	No
Chaotic	Yes	Yes
$\widetilde{D_c}$	1.3 theoretical time series 2.9 noise 10%	1.9
$\widetilde{\lambda_1}$	0.5 theoretical time series 0.35 noise 10%	0.37
Mean	0.36 theoretical 0.36 noise 10%	0.46
Variance	1.01 theoretical 1.02 noise 10%	0.9
$\sigma(\widetilde{f}_N)$	$\sqrt{\frac{1}{NTrain*0.9}\sum_{i=1}^{NTrain}(y_i^* - \widetilde{y}_i)^2}$ $\sqrt{\frac{1}{NTrain*0.9}\sum_{i=1}^{NTrain}(y_i - \widetilde{y}_i)^2}$	0.90 0.86
$\sigma(\widetilde{f}_N^{(15)})$	$\sqrt{\frac{1}{NTrain*k*0.9}\sum_{i=1}^{NTrain}\sum_{j=1}^{k}(y_{ij}^* - \widetilde{y}_{ij})^2}$ $\sqrt{\frac{1}{NTrain*k*0.9}\sum_{i=1}^{NTrain}\sum_{j=1}^{k}(y_{ij} - \widetilde{y}_{ij})^2}$	0.87 0.84

Table 1: Model constructed from observations: Henon time series with a 10% of noise level.

	Theoretical system	Model
Expression	$x_{k+1} = y_k + 1 - 1.4 * x_k^2$ $y_{k+1} = 0.3 * x_k$	$x_{1k+1} = 0.322 * x_{2k} + 0.922$ $x_{2k+1} = 0.89 * x_{1k} - 1.4 * x_2^2$ $y_{k+1} = 3.6 * x_{1k+1} - 3.016$
Dimension	2	2
Linear	No	No
Chaotic	Yes	Yes
$\widetilde{D_c}$	1.3 theoretical time series 2.9 noise 10%	1.9
$\widetilde{\lambda_1}$	0.5 theoretical time series 0.35 noise 10%	0.37
Mean	0.36 theoretical 0.36 noise 10%	0.46
Std. Dev.	1.01 theoretical 1.02 noise 10%	0.9
$\sigma(\widetilde{f}_N)$	$\sqrt{\frac{1}{NTrain*0.9}\sum_{i=1}^{NTrain}(y_i^* - \widetilde{y}_i)^2}$ $\sqrt{\frac{1}{NTrain*0.9}\sum_{i=1}^{NTrain}(y_i - \widetilde{y}_i)^2}$	0.90 0.86
$\sigma(\widetilde{f}_N^{(15)})$	$\sqrt{\frac{1}{NTrain*k*0.9}\sum_{i=1}^{NTrain}\sum_{j=1}^{k}(y_{ij}^* - \widetilde{y}_{ij})^2}$ $\sqrt{\frac{1}{NTrain*k*0.9}\sum_{i=1}^{NTrain}\sum_{j=1}^{k}(y_{ij} - \widetilde{y}_{ij})^2}$	0.87 0.84

Table 2: Model constructed from observations of a periodic system.

	Physical measurements	Model
Expression	unknown	see equation 13
Dimension	unknown	2
Linear	No	No
Dissipative	Yes	Yes
$\widetilde{D_c}$	3.6	1.3
$\widetilde{\lambda_1}$	0.015	0.005
Mean	360815	42706
Std. Dev.	$6.28 * 10^9$	$1.4 * 10^9$
$\sigma(\widetilde{f}_N)$	$\sqrt{\frac{1}{NTrain*1.4*10^9} \sum_{i=1}^{NTrain} (y_i^* - \widetilde{y}_i)^2}$	1.141
$\sigma(\widetilde{f}_N^{(15)})$	$\sqrt{\frac{1}{NTrain*k*1.4*10^9} \sum_{i=1}^{NTrain} \sum_{j=1}^{k} (y_{ij}^* - \widetilde{y}_{ij})^2}$	1.146

Table 3: Model constructed from input supply of blast furnace gas.