Sunspot Forecasting by Using Chaotic Time-series Analysis and NARX Network

Chuanjin Jiang
Electronic Circulation Research Center
Shanghai Business School, Shanghai, China
chuanjin@21cn.com

Fugen Song
Donghua University, Shanghai, China
fgsong@dhu.edu.cn

***Abstract*—Chaotic time-series is a dynamic nonlinear system whose features can not be fully reflected by Linear Regression Model or Static Neural Network. While Nonlinear Autoregressive with eXogenous input includes feedback of network output, therefore, it can better reflect the system's dynamic feature. Take annual active times of sunspot as an example, after verifying the chaos of sunspot time-series and calculating the series' embedding dimension and delay, we establish sunspot prediction model with NARX network. The result shows that compared with BP Network and ARIMA Model, NARX network can better predict the chaos.**

***Index Terms*—Chaos; Time-Series Prediction; NARX; Sunspot**

I. INTRODUCTION

Chaotic Time-Series Prediction is an important application and research hotspot of Chaos Theory, and is widely used in natural science and social science. Chen Ying [1] and Li Dongmei[2] established prediction model of chaotic time-series by using BP Neural Network and RBF Neural Network respectively, and made some achievements.

Both BP neural network and RBF network is static neural network. Using static network to set up model for dynamic chaos system actually turns dynamic time-series modeling into static dimension modeling. Static neural network is better when conducting one-step static prediction, but not ideal for medium-and-long-term dynamic prediction. Hann[3] indicates that when conducting short-term prediction to exchange rate chaotic time-series, static neural network is better than any other models, while when conducting medium-and-long-term prediction, static neural network is not even as good as simply linear regression model.

Nonlinear Autoregressive with Exogenous input includes feedback of network output, so it can better reflect the system's dynamic feature. This paper takes the sunspot number in 310 years from 1700 to 2009 as time-series sample, studies its internal chaos, calculates embedding dimension and delay, establishes model with NARX neural network and conducts prediction. The result shows that compared with BP network or traditional ARIMA model, NARX network is better in precision.

II. BUILDING THE FORECASTING MODEL

A. NARX Neural Network

NARX is a nonlinear discrete-time system and its mathematical expression is:

$$y(n+1) = f[y(n), y(n-1),...., y(n-d_y+1); \mathrm{u}(n), \mathrm{u}(n-1),...., \mathrm{u}(n-d_u+1; \mathrm{W}] = f[y(n); \mathrm{u}(n); \mathrm{W}] \quad (1)$$

Where $u(n)\in \mathrm{R}$; $y(n)\in \mathrm{R}$ respectively represent the input and output of the model at time t , $d_u \geq 1; d_y \geq 1\,(d_u \geq d_y)$ are memory vector input and memory vector output respectively, W is weight matrix, f is nonlinear function simulated by multilayer perception. Fig 1 shows structure chart of NARX network:

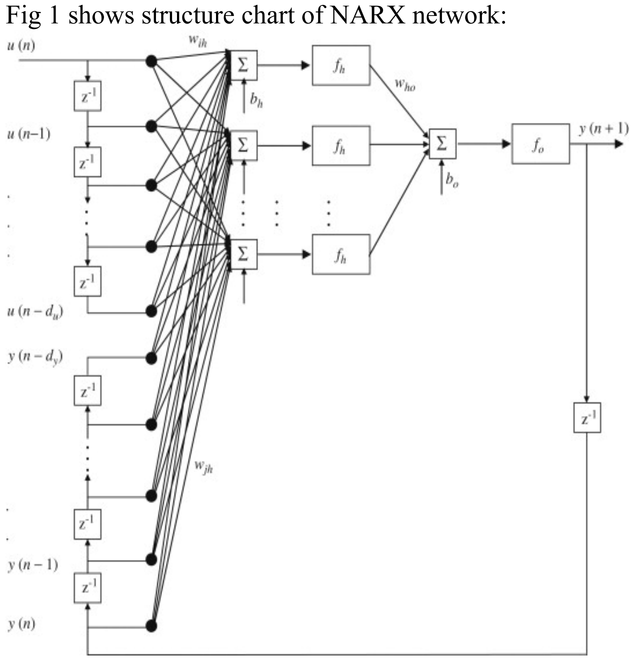

Figure 1. NARX Network

doi:10.4304/jcp.6.7.1424-1429

Usually NARX network can be trained and predicted through two modes:

Parallel(P) Mode: System's output regressor also includes the estimated value of output.

$$\hat{y}(t+1)=\hat{f}[y_{sp}(t),\mathrm{u}(t);\mathrm{W}]=\hat{f}[y(t),y(t-1),....,y(t-n_y+1);\ \mathrm{u}(t),\mathrm{u}(t-1),....,\mathrm{u}(t-n_u+1);\mathrm{W}] \tag{2}$$

Series-Parallel(SP) Mode: System's output regressor only includes the system's actual output value:

$$\hat{y}(t+1)=\hat{f}[y_p(t),\mathrm{u}(t);\mathrm{W}]=\hat{f}[\hat{y}(t),\hat{y}(t-1),....,\hat{y}(t-n_y+1);\ \mathrm{u}(t),\mathrm{u}(t-1),....,\mathrm{u}(t-n_u+1);\mathrm{W}] \tag{3}$$

Since estimated value's regressor is included in Mode *P*, during mult-steps medium-and-long-term prediction, Mode *P* can better show the system's dynamic feature with better predictive validity [5]. Therefore, Mode *P* is adopted in this paper for network training and prediction.

B. Phase Space Reconstruction of Chaotic Time Series

After certain changes, the pathway produced by chaotic system will finally make regular movements, so this pathway becomes a regular one, which is also called singular attractor. This kind of pathway, after stretching-and-folding-like treatment, becomes a time-related series, which presents features of chaos and complexity. Therefore, singular attractor can be recovered through Phase Space Reconstruction and this is also the initial purpose for raising the concept of Phase Space Reconstruction.

Since the driving factors of chaotic system have interactive influence to each other, the data points generated in time sequence are also related. It has been proven by Takens[7] that a proper embedding dimension d_E can be found to recover singular attractor in this embedding dimension space. Through Phase Space Reconstruction, the evolving law of singular attractor can be found, which is able to put the existing data into certain describable frame, and provide a new method and approach to study time series. Phase Space Reconstruction is an important step to analyze non-linear time series, and the reconstruction quality will have direct impact on model establishment and prediction.

Maguire [6] has proven that sunspot time-series satisfies features of chaotic time-series. Takens[7] gives us mathematical way to conduct phase space reconstruction on time-series:

$$\mathrm{u}(t)=[y(t),y(t-\tau),....,y(t-d_E-1)\tau)] \tag{4}$$

Among which d_E is embedding dimension, and τ is embedding delay. The compare of (1) and (4) shows $n_u=d_E$. Thus, the key of modeling NARX network of time-series is selecting embedding dimension and time delay.

C. Calculation of Embedding Dimension

Embedding dimension can be estimated through Cao method [8].

For input time series u(*t*), select τ; then for certain dimension *d*, vector serial $y_i(d)$ can be reconstructed. $y_i(d)$ indicates the i^{th} vector acquired when reconstructing dimension *d*. Define *a*(*i*,*d*):

$$a(i,d)=\frac{\| y_i(d+1)-y_{n(i,d)}(d+1)\|}{\| y_i(d)-y_{n(i,d)}(d)\|} \tag{5}$$

Here, ||···|| indicates certain measurement to Euclidean distance.

Then, define the average value of *a*(*i*,*d*):

$$E(d)=\frac{1}{N-d\tau}\sum_{i=1}^{N-d\tau}a(i,d) \tag{6}$$

According to the function above, E(*d*) only depends on dimension *d* and delayed time τ. Therefore, in order to study dimension change from *d* to *d*+1, define:

$$E_1(d)=E(d+1)/E(d) \tag{7}$$

If time serial is from certain attractor, when *d* is bigger than certain value of d_0, $E_1(d)$ will stop changing, and *d* is the smallest embedding dimension.

In addition, it is necessary to define another value, as it can be used to distinguish fixed signal and random signal. Make:

$$E^*(d)=\frac{1}{N-d\tau}\sum_{i=1}^{N-d\tau}|x_{i+d\tau}-x_{n(i,d)}+d\tau| \tag{8}$$

The meaning of *n(i,d)* is the same as discussed. Meanwhile, also define

$$E_2(d)=E^*(d+1)/E^*(d) \tag{9}$$

Signal with fixed structure and random signal can be clearly distinguished with E_2. For random time series, since both future value and previous value are independent, $E_2(d)$ always equals to 1 despite the value of *d*.

D. Selection of Delayed Time

Delay τ can be obtained through autocorrelation method. The first step is to make autocorrelation function C(τ)'s function diagram about time τ.

Takens[7] has proved that when there is no noise and the data accuracy of one dimensional time series is infinite, the selection of delayed time τ can be random. However, in fact, the actual testing time can not be infinite, and the noise is inevitable. Therefore, the selection of delayed time τ plays an important role during reconstruction process. The basic approach is to make y(*t*) and y(*t*+ τ) own certain independence while are not totally irrelevant to each other, so that they can be independent coordinates in phase space reconstruction.

Autocorrelation Function is a very mature method to acquire the value of delayed time, and its basic thought is to acquire the linear correlation between series. For chaotic time series, the autocorrelation function of time span $j\tau$ is:

$$R_{xx}(j\tau) = \frac{1}{n}\sum_{i=0}^{n-1} x_i x_{i+j\tau} \qquad (10)$$

According to this function, when fixing j, the autocorrelation function's image about delayed time τ (τ =1,2,) can be acquired. When the value of this function decreased to initial value, the delayed time τ acquired is the optimal delayed time value.

E. Lyapunov Index and predictable scale

Chaotic system is strongly sensitive to initial value. Even there's only a trivial change to initial conditions, the system's evolving pathway will deviate with previous pathway at exponential speed, and after certain period, covers the system's true status completely. Therefore, this indicates unpredictability of the system's long-term movement. When the attractor in chaotic system is partially unstable, the pathway finally falls on the same chaotic attractor in phase space, and the system's biggest Lyapunov index is bigger than zero (λ >0), it indicates chaotic attractor exists which can be used to measure the chaotic level. In addition, the inverse of the biggest Lyapunov index also shows the system's longest predictable time, beyond which the tolerance is out of control and covers the system's real status.

It is much simpler to calculate Lyapunov index with Small-data method[9] to avoid massive calculation amount.

After the phase space is reconstructed, find every point's nearest adjacent point on given pathway,

$$d_j(0) = \min_{\hat{j}} \| Y_j - Y_{\hat{j}} \|, |j - \hat{j}| > P \qquad (11)$$

For every point of phase space Y_j, calculate its distance after i discrete time step of the adjacent,

$$d_j(i) = \| Y_{j+i} - Y_{\hat{j}+i} \|$$
$$i = 1,2,\ldots\ldots,\min(M-j, M-\hat{j}) \qquad (12)$$

For every i, calculate average value of $\ln d_j(i)$:

$$y(i) = \frac{1}{q\Delta t}\sum_{j=1}^{q} \ln d_j(i) \qquad (13)$$

Among which q is the nonzero number of $d_j(i)$. Make fitting curve through least square method, and the curve's slop is the biggest Lyapunov index λ .

F. Neuron Quantity in Hidden Layer of NARX Network

Neuron quantity in two hidden layers of NARX neural network is also the parameter which needs to be set. Neuron in the first hidden layer N_1 and neuron in the second hidden layer N_2 will be determined via the formula below:

$$N_1 = 2d_E + 1, \quad N_2 = \sqrt{N_1} \qquad (14)$$

Among which N_2 is rounded downwards.

III. EMPIRICAL ANALYSIS

A. Data Analysis

The experimental data used in this paper 310 sunspot number from 1700 to 2009 as time-series. For easy calculation, the data are normalized with [0, 1] as Figure 2 indicates:

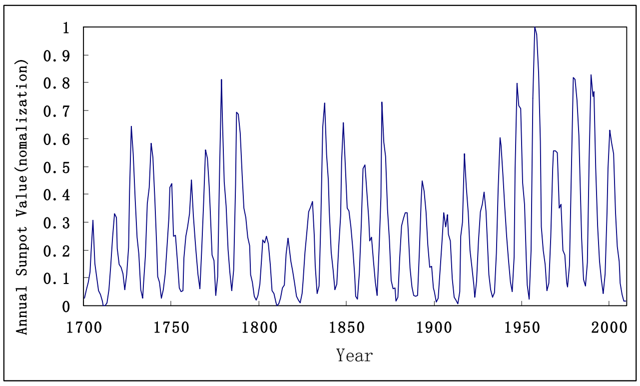

Figure 2. Sample data of Sunspot data

Observing Fingure 3, when d_E =10, $E_1(d)$ reaches saturation, and the value of $E_1(d)$ is also close to 1, therefore, the minimum embedding dimension of sunspot annul active value is 10, which is in accordance with the series' wave cycle in Fingure 2.

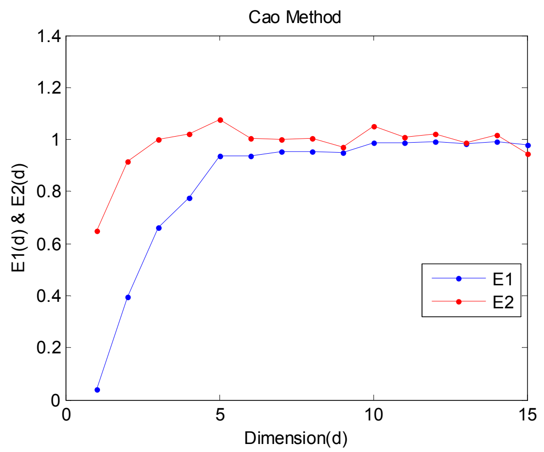

Figure 3. The value E1 and E2 of time series data

When substituting d_E =10 into (14), neutron quantity in two hidden layers can be calculated: N_1=21, N_2=4.

In Fingure 4, when τ=3, C(τ) reaches the first zero value, therefore the delay of time-series is 3.

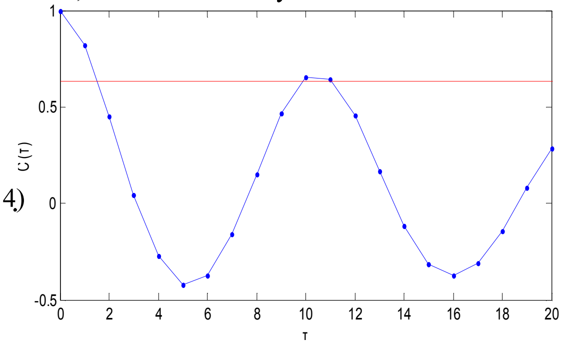

Figure 4. Embedding Delay by Autocorrelation Method

In Figure 5, a slop curve corresponds to theoretical value of Lyapunov index λ. Choose the longer linearity range in slope curve diagram to do straight line fitting, and the biggest Lyapunov index value in this range is 0.0931.

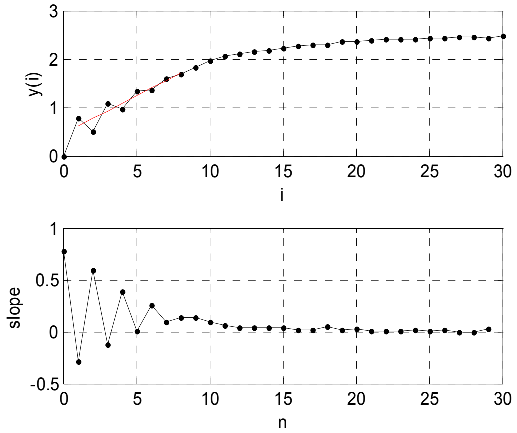

Figure 5. Lyapunov Index by Small-data Method

The predictable time of this chaotic time-series T=1/λ ≈10 years. If it exceeds 10 years, the system's prediction accuracy can not be controlled.

Therefore 310 sample data is divided into two parts. The first 300 sample data is training sample, and the last 10 is prediction validation sample.

In the first 300 training data samples, according to phase space reconstruction, the first sample that can be fit is the 31st data ($d_E \times \tau + 1 = 31$), i.e. data in the year of 1730. The whole fitting extent is [1730, 1999] with 270 sample data.

B. Result Analysis

Figure 6 shows the fitting result of using NARX neural network to establish model for sample. Figure 7 gives the error of NARX simulation from 1730 to 1999. It turns out that NARX fitting curve does not cover few maximal points which effectively avoids over fitting.

NARX, BP network and ARIMA (1,1,1) are used separately to establish model for first 300 sample data, and predict for the last 10 data. Table 1 and Figure 8 show the prediction result; Table 2 is error analysis.

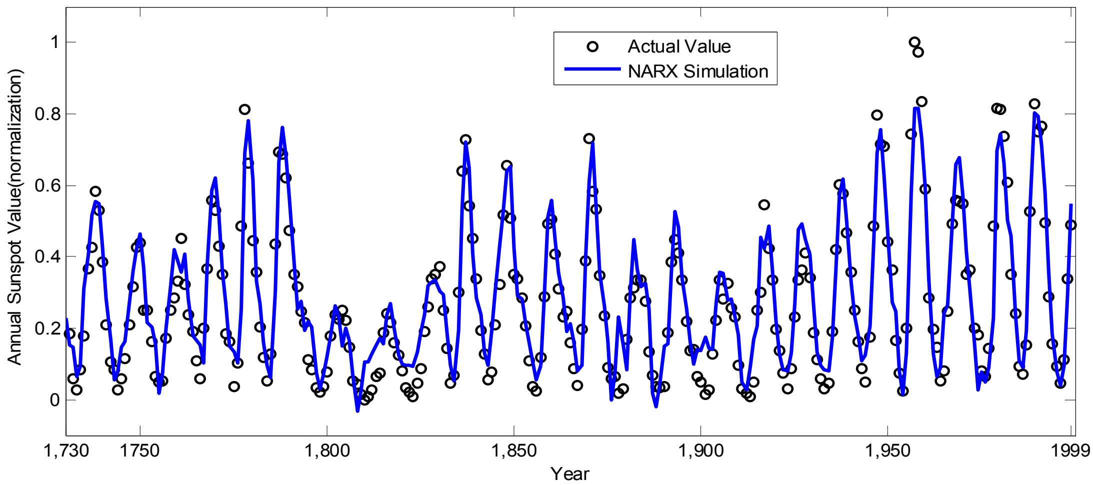

Figure 6. NARX fitting curve to the sample

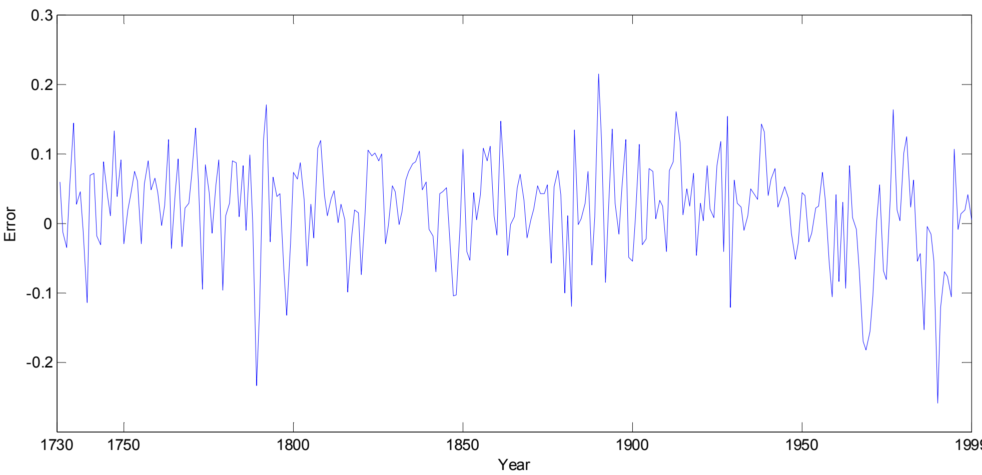

Figure 7. Error of NARX simulation from 1730 to 1999

TABLE I. PREDICTION OF NARX、BP AND ARIMA

Year	Actual Value	NARX Prediction	BP Prediction	Arima Prediction
2000	0.6288	0.6450	0.6807	0.5571
2001	0.5836	0.5825	0.6352	0.6998
2002	0.5468	0.5527	0.5707	0.5407
2003	0.3349	0.3447	0.3900	0.5333
2004	0.2124	0.2408	0.2816	0.2074
2005	0.1567	0.1636	0.1507	0.1641
2006	0.0799	0.1151	0.1090	0.1326
2007	0.0394	0.0766	0.0761	0.0378
2008	0.0152	0.0355	0.1185	0.0235
2009	0.0163	0.0067	0.2491	0.0038

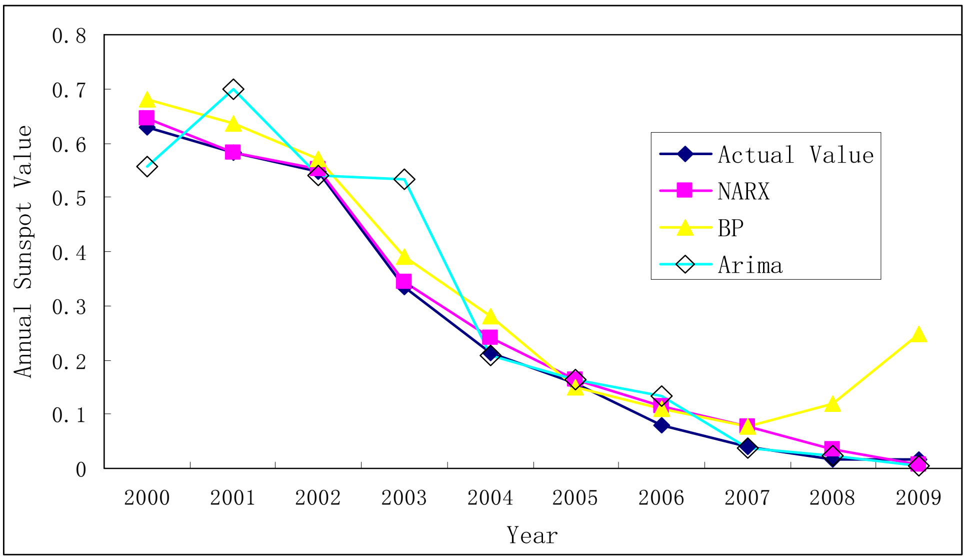

Figure 8. Prediction of NARX、BP and ARIMA

It is showed in Figure 8 that BP network has excellent prediction on first 7 years, but the accuracy for last 3 years is not ideal with big error; ARIMA is very unstable for first 4 years' prediction, but the prediction for last 6 years is very good; whereas, NARX network shows high prediction accuracy for whole 10 years.

TABLE II. AVERAGE ERROR OF NARX、BP AND ARIMA

	NARX	BP	ARIMA
MSE	0．0004	0.0081	0.0061
MAE	0.017	0.066	0.048
RMSE	0.006	0.028	0.025

The result in Table 2 also shows that NARX has the best prediction, ARIMA the second, and BP neural network the third.

IV. CONCLUSION

Since NARX network's input includes feedback of network output, NARX network can reflect system's dynamic feature vividly, therefore, can be used for chaotic time-series prediction. This paper takes sunspot chaotic time-series as an example to verify NARX network's ability in short-term and medium-term prediction. The result shows NARX's chaotic prediction validity is better than that of BP neural network and ARIMA model.

ACKNOWLEDGMENT

The authors wish to thank Wenguan Song and Zhang Jin.

REFERENCES

[1] C. Ying, P. Luo, "The Modeling and Forecasting of Chaotic Time Series Based on Neural Network", *Comuputer Engineering and Applictions*, 2005,(11)

[2] D. Li, Z. Wang, "Modeling and Multi-Step Prediction of Chaotic Time Series Based on RBF Neural Networks", *System Engineering and Electronic Technolgy*, 2002,(06)

[3] T. H. Hann, E. Steurer, "Much A do about Nothing? Exchange Rates Forecasting: Neural Networks vs. Linear Models Using Monthly and Weekly Data", *Neurocomputing*, 1996, 10:323-2339.

[4] I. J. Leontaritis, S. A. Billings, "Input–output parametric models for nonlinear systems", *International Journal of Control*, 1985, 41:303–344.

[5] T. Lin, B. G. Horne, P. Tino, C. L. Giles, “A delay damage model selection algorithm for NARX neural networks”, *IEEE Transactions on Signal Processing*, 1997,45(11), 2719-2730.

[6] L. P. Maguire, B. Roche, T. M. Mcginnity, “Predicting a chaotic time series using a fuzzy neural network”, *Information Sciences*, 1998, 112:125-136.

[7] F. Takens, “On the numerical determination of the dimension of an attractor”, *Dynamical System and Turbulence, Lecture Notes in Mathematics, Berlin: Springer-Verlag*, 1981, 898:230-241.

[8] L. Y. Cao, “Practical method for determining the minimum embedding dimension of a scalar time series”, *Physical D*, 1997, 110: 43-50.

[9] M. T. Rosenstein, J. J. Collins, L. C. J. De, “A practical method for calculating largest lyapunov exponents from small data sets”, *Physica D*, 1993, 65:117-134