NON-LINEAR ANALYSIS ON EXXON MOBIL SHARE PRICES

S. TSAKONAS[*], L. MAGAFAS[**,***], M. HANIAS[**,***], and L. ZACHILAS[*]

Abstract. *ExxonMobil Corporation is an American multinational company engaged in energy development as well as oil and natural gas extraction. The purpose of this work is to examine if chaos models are suitable for modeling and forecasting financial time series data such as the company's share prices. At first a non-linear analysis is performed on the company's weekly share prices and later on a phase space reconstruction is carried out according to Takens' theorem, in order to fit a locally linear regression model and predict futures values.*

Keywords: *chaos, forecast, financial markets*

1. Introduction

Exxon Mobil is a company with a history of 125 years and activity in most countries of the world. It is one of the largest non-governmental energy organizations that produces 3% of the world's total oil and 2% of global energy. In this paper we attempt to forecast the company's future share prices. It is well known that financial return series are not characterized by the normal distribution [1]. Contrariwise, the distribution curve appears to be leptokurtic showing high peaks and long tails. For that reason we assume linear models not suitable for modeling financial time-series. Instead we apply methods that derive from chaos theory in order to extract information that will help predict future values. Original time series are shown in figure 1.

[*] Department of Economics, University of Thessaly.

[**] Department of Electrical Engineering, Eastern Macedonia and Thrace Institute of Technology.

[***] Complex Systems Laboratory, Eastern Macedonia and Thrace Institute of Technology.

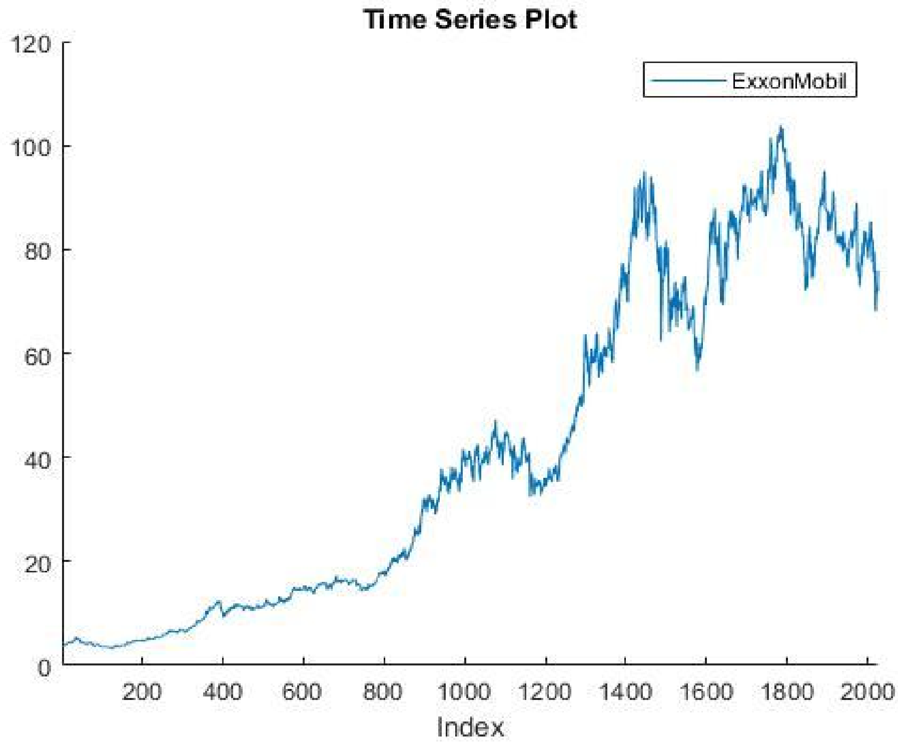

Figure 1. Time Series of Exxon Mobil index.

2. Data Preprocessing

The data values employed in this work consist of weekly closing share prices of EXXON MOBIL from 23/3/1980 until 27/1/2019 (2025 values total). Before proceeding with non-linear analysis we convert the data into stationary time-series by removing the long-term trend. For this purpose we use Standard Score (z-score) normalization that makes use of the function

$$f(x)=\frac{x-\bar{X}}{\sigma(X)}$$

where each data point x is normalized by subtracting a 10 week moving average and then dividing by a 10 week standard deviation. The normalized times series are shown in figure 2.

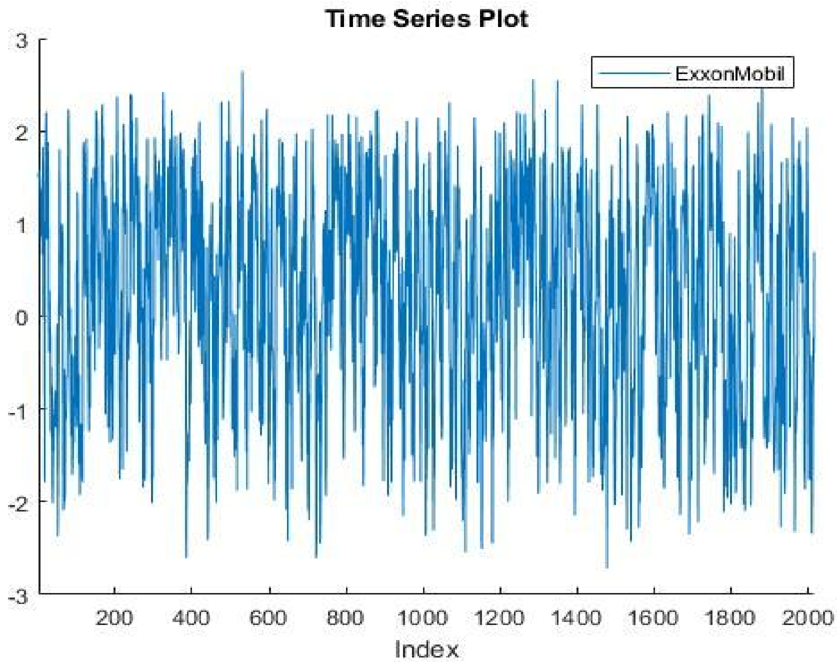

Figure 2. Normalized Time Series.

Augmented Dicker-Fuller statistical test is applied on the time series to check if a unit root is present [2]. The alternative of the null-hypothesis suggests that the time series are stationary.

Augmented Dickey-Fuller Test: Null Hypothesis: Exxon contains a unit root

Table0 1.
Test Parameters

Model	Test Statistic	Significance Level
AR	t1	0.05

Table0 2.
Test Results

Null Rejected	P-Value	Test Statistic	Critical Value
true	0.001	–18.314	–1.9416

The null hypothesis is rejected. The normalized data are stationary by mean.

3. Non-Linear Analysis

According to Takens' theorem [3], under specific conditions, a set of observed time series can be reconstructed into a multi-dimensional phase space. The first step is to determine the parameters of the embedding dimension m and time delay τ. The time delay parameter can be calculated by using the average mutual information method [4] and picking the first minimum of the equation below as the optimized value [5].

$$I(\tau) = \sum_{x_i, x_{i+\tau}} p(x_i, x_{i+\tau}) log_2 (\frac{p(x_i, x_{i+\tau})}{p(x_{i)} p(x_{i+\tau})})$$

where$p(x_{i)} p(x_{i+\tau})$ is the joint probability distribution.

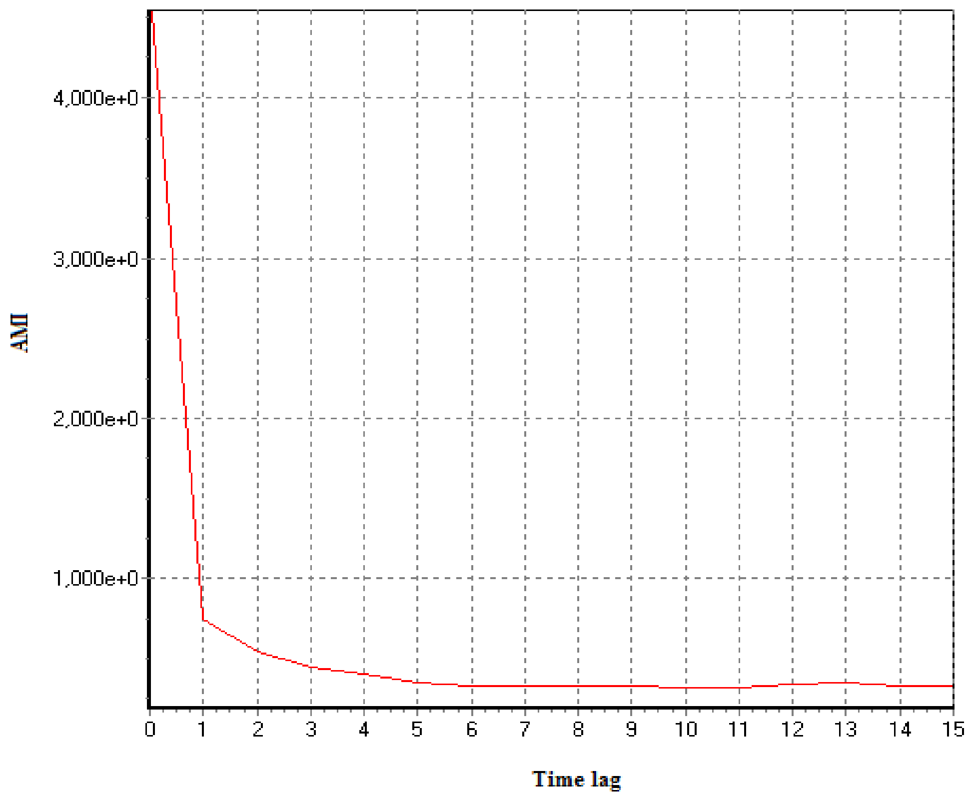

Figure 3. Average Mutual Information.

The first minimum was found at τ=10 as shown in figure 3.

Table 3.
Average Mutual Information Values

τ	I
1	0,740005
2	0,544039
3	0,448488
4	0,406377
5	0,342203
6	0,330536

τ	I
7	0,330454
8	0,330358
9	0,322788
10	0,317636
11	0,319775
12	0,332686

Before estimating the embedding dimension, first a time window needs to be determined that excludes all temporal correlations between neighboring vectors [6]. The lines of the Space-Time separation plot [7] as shown in figure 4, represent the constant probabilities vector pairs being closer to each other than a distance threshold r at a specific time separation Δt. A suitable time window value can be picked as the first maximum joint of all the contours. In our case, we pick a time window value of w=10.

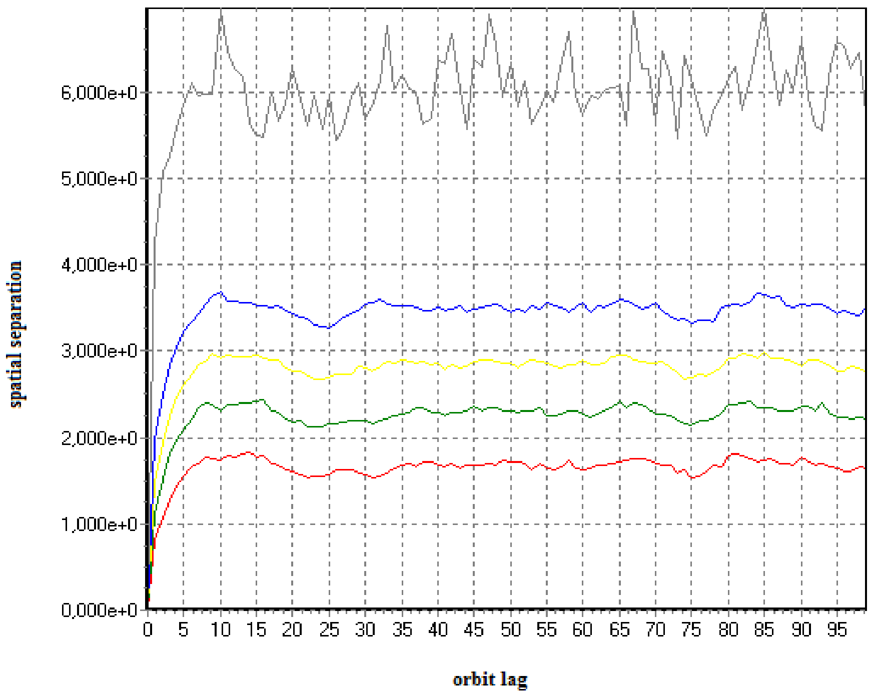

Figure 4. Space-Time Separation Plot.

Correlation Dimension

The algorithm introduced by Grassberger and Procaccia is employed to estimate the fractal dimension of the attractor [8]. The spatial correlations are measured by using the correlation integral adjusted by Theiler proposal.

$$C^m(r) = \frac{1}{(N-w)(N-w-1)} \sum_{i,j=1}^{N} H\,(r - \|X_j - X_i\|)$$

N is the data size

H is Heaviside step function

W is Theiler window

The correlation dimension is then calculated as

$D^m = \lim_{\varepsilon\to\infty} \frac{log C^m}{\log r}$.

The correlation dimension algorithm can also be used to distinguish between stochastic and deterministic chaotic dynamics. As the embedding dimension gradually increases, the more vectors are enclosed within a sphere of radius r within the phase space which results to the decline of the correlation dimension, until it reaches a saturation level as shown in figure 5. The saturation of the Correlation Dimension curve at a non-integer value is an indication of deterministic chaos.

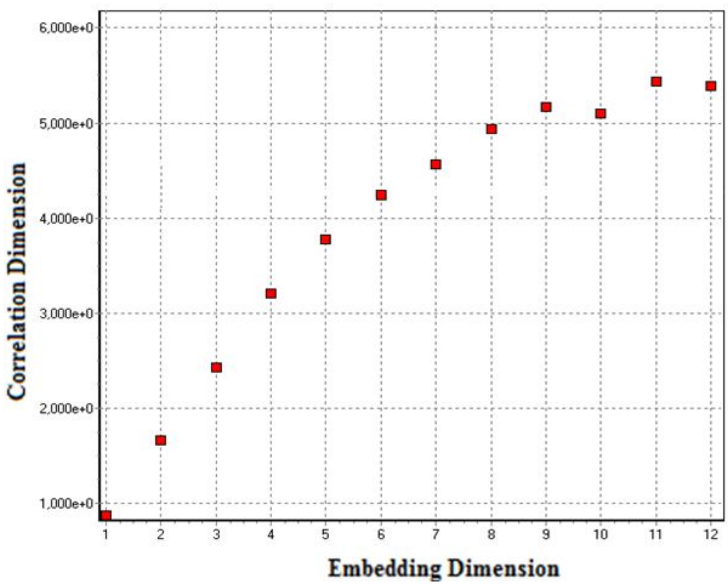

Figure 5. Embedding Dimension vs Correlation Dimension of Normalized Data.

The surrogate data method is also used as a significance test. The surrogate data are random data that maintain all the statistical properties (mean, standard deviation) of the normalized data. The correlation dimension of the surrogate data is expected to increase along with the embedding dimension, as shown in figure 6, since the white noise will expand and occupy all the space that is available to it.

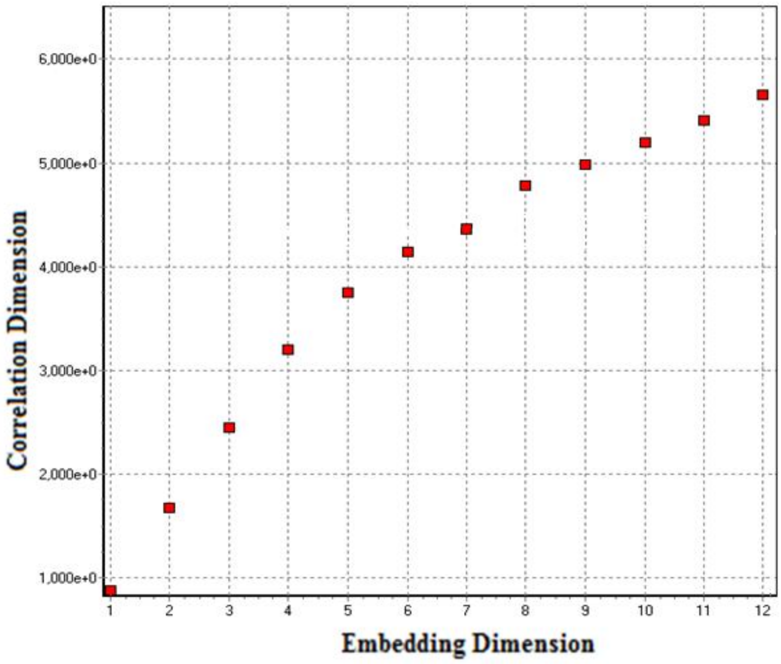

Figure 6. Embedding Dimension vs Correlation Dimension of Surrogate Data.

The estimation of the minimum embedding dimension makes use of the Global False Nearest Neighbors method, in order to identify vectors that may be falsely considered as neighbors due to low embedding dimension. This method calculates the percentage of all vectors that exceed a Euclidian distance threshold within the phase space [9] in different embedding dimensions. As a suitable embedding dimension for the phase space reconstruction, it is accepted the one where the percentage of false neighboring vectors reaches below a specific threshold or becomes equal to zero.

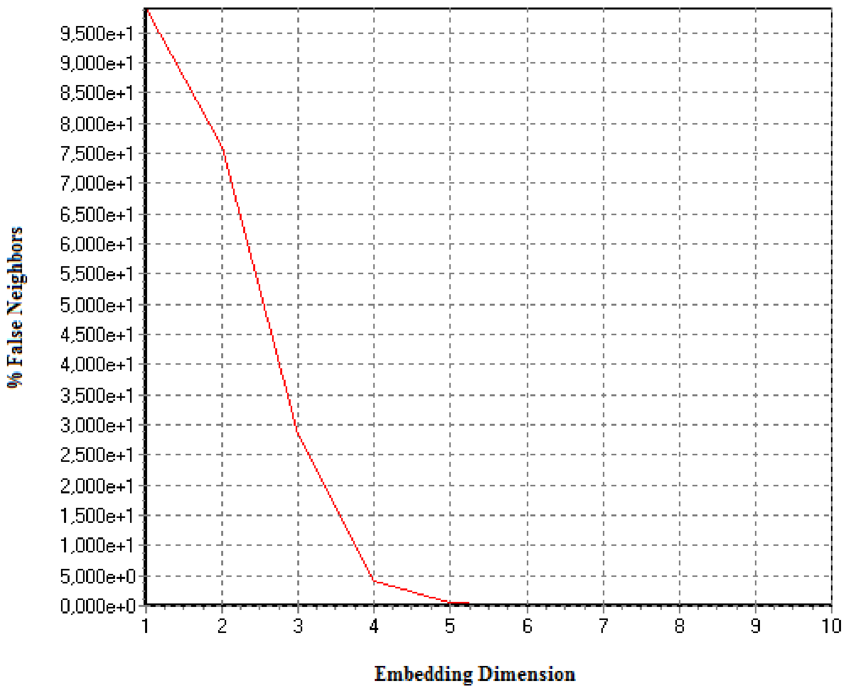

Figure 7. False Nearest Neighbors plot.

Table 4. *False Neighbors percentage*

1	2	3	4	5	6
99,00835	76,09603	28,34029	4,070981	0,417537	0

The percentage of false nearest neighbors zeroes at m = 6 as shown in figure 7.

4. Forecast

A local linear regression model is fitted on a set of vectors that consists the neighborhood around the prediction target using a weighted least squares method [10] in order to predict 31 steps ahead. A tri-cube weighting function is also used to add more weight to the vectors, closer to the prediction target. The number of neighboring vectors used for this multi-step forecast is k =50. Table 4 shows the comparison between the actual normalized values and the predicted.

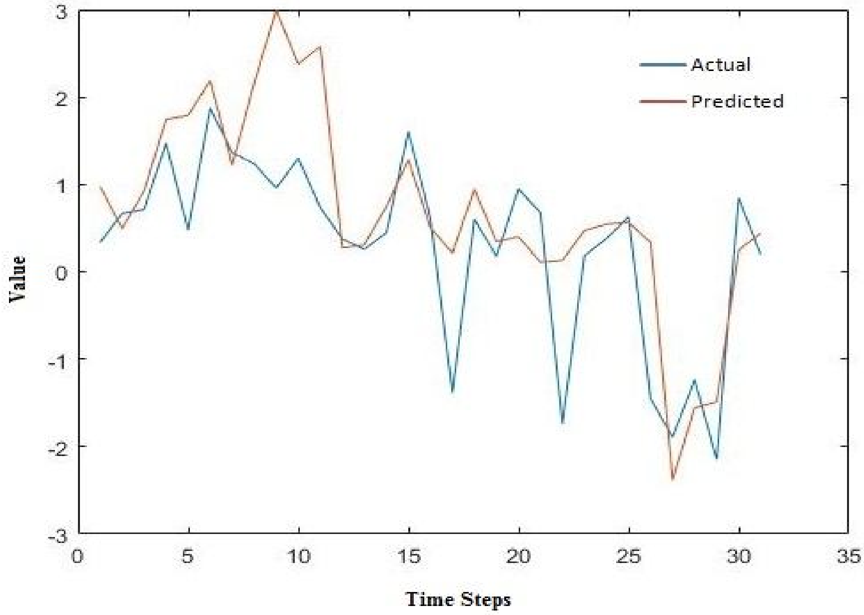

Figure 8. Forecast: Actual vs Predicted Values

The root mean square error is also calculated as a quality measurement of the fit of the model. RMSE = 0,8853281

Table 4. *Actual vs Predicted Values*

Date	Actual	Predicted
11/6/2006	–1,4422133	–1,1439838
18/6/2006	–1,4065127	0,3075367
25/6/2006	0,1771428	0,6705335
2/7/2006	0,9679974	0,3329669
9/7/2006	1,8310379	0,8178534
16/7/2006	1,1914324	1,542417
23/7/2006	1,7786311	1,1274198
30/7/2006	1,7008306	0,8731676
6/8/2006	1,5051942	0,7879081
13/8/2006	1,1080807	0,6334243
20/8/2006	1,1839473	0,8966334
27/8/2006	0,4742075	0,5806052
3/9/2006	0,1343791	0,878703
10/9/2006	–1,1875662	1,30698
17/9/2006	–1,735362	0,4529398
24/9/2006	0,2866289	0,1045211
1/10/2006	0,961713	0,4350697
8/10/2006	0,381822	0,2877829

15/10/2006	1,80498	0,4228656
22/10/2006	1,6213337	0,6447517
29/10/2006	1,6531937	0,4594402
5/11/2006	1,7964276	1,6767
12/11/2006	1,1099423	0,9498913
19/11/2006	0,7447284	0,3536119
26/11/2006	1,8271056	1,498103
3/12/2006	1,815113	1,1227393
10/12/2006	1,3875927	1,1035845
17/12/2006	0,6140974	0,6938194
24/12/2006	0,9530129	0,2878905
31/12/2006	0,7605571	0,2837299
7/1/2007	–1,1220316	0,5033767

5. Conclusions

In this paper, non-linear analysis has been performed on the weekly share prices of Exxon Mobil Corporation. First, the data were normalized, removing the trend and achieving stationarity. The non-linear properties were examined in order to understand the structural complexity of the system and reconstruct the phase space. The results of the correlation dimension calculation indicate noisy chaos. Local linear regression was used to predict 31 values ahead in the future. The forecasting results may be lacking precision since chaotic dynamics are not suitable for long term prediction due to sensitivity to the initial conditions. Future work could include further investigation of the non-linear properties and dynamics of the dataset and modification of the regression fitting in order to improve the precision of the forecasting results.

REFERENCES

[1] Peters, E. E. (1991), *Chaos and Order in the Capital Markets*: *A New View of Cycles, Prices, and Market Volatility.*

[2] Dickey, D. A. (1976), *Estimation and hypothesis testing in non-stationary time series*, Ph.D. dissertation (Iowa State University, Ames, IA).

[3] Takens, F. (1981), *Detecting Strange Attractors in Turbulence*, Lecture notes in Mathematics, Vol. 898, Springer, New York.

[4] Casdagli, M., Eubank, S., Farmer, J. D. and Gibson, J. (1991), *State Space Reconstruction in the Presence of Noise*, *Physica D*, 51, pp. 52-98.

[5] A. M. Fraser and H. L. Swinney (1986), *Independent coordinates for strange attractors from mutual information*, Phys Rev A Gen Phys, 33(2):1134-1140.

[6] Theiler (1986), *Spurious dimension from correlation algorithms applied to limited time-series data,* Phys Rev A Gen Phys, 34(3):2427-2432.

[7] A. Provenzale, L. A. Smith, R. Vio, and G. Murante (1992), *Distinguishing between low-dimensional dynamics and randomness in measured time series*. Phys D Nonlinear Phenom, 58(1-4):31-49
[8] Peter Grassberger, Itamar Procaccia (1983), *Measuring the strangeness of strange attractors*, Phys D Nonlinear Phenom, 9(1-2):189-208
[9] MB Kennel, Brown, and HD Abarbanel (1992), *Determining embedding dimension for phase-space reconstruction using a geometrical construction*. Phys. Rev., A, 45(6):3403-3411.
[10] Farmer, J. D. and Sidorowich, J. J. (1987), *Predicting chaotic time series*. Physical Review Letters, **59,** 845.