Parameter optimization of phase space reconstruction based on differential entropy

Yun Sun\textsuperscript{1,3}, Ying Wang\textsuperscript{1}, Chao Li\textsuperscript{1}, Xiangfei Meng\textsuperscript{1}, Yao Qi\textsuperscript{1} and Wen Guo\textsuperscript{2}

\textsuperscript{1}Air Force Engineering University Equipment Management and UAV Engineering College, Shaanxi Xi’an 710051, China
\textsuperscript{2}Information Engineering University, Zhengzhou 450000, China
\textsuperscript{3}Email: 1259006637@qq.com

Abstract. This paper presents a differential entropy optimization method for phase space reconstruction parameters. The degree of differential entropy can be used to describe the complexity of the system. By establishing the relationship between the differential entropy and the embedded dimension and the delay time, the objective function is obtained. The constraints of embedding dimension and delay time is established by describing the global features of the system through autocorrelation function. By using the improved particle swarm algorithm to solve the model, the minimum differential entropy is obtained and the best embedded dimension and delay time is obtained. Therefore, the reconstructed phase space not only maintains its independence but also maintains its dynamic characteristics. Finally, through the prediction of the annual data of sunspots and the simulation data from Lorenz system and, it is verified that the method can determine the appropriate embedding dimension and delay time.

1. Introduction
The generation of chaos can be considered to be a complex system in which the high-dimensional dynamic system is projected onto a two-dimensional space. If the "reduction" of the chaotic system to the high-dimensional space can reduce the complexity of the system theoretically, the original simple dynamic system can be obtained. This restoration process can be completed by way of phase space reconstruction. According to Takens’ theorem [1], the reconstruction of phase space mainly involves the determination of two parameters of embedding dimension and delay time. At present, the determination of two parameters can be roughly divided into two categories [2]: One class considers that there is no relationship between the two parameters and can be solved separately. The delay time is solved by the autocorrelation method and the mutual information method, and the embedding dimension is obtained by the G.P (Grassberger P) method, etc. One class considers the two parameters to be related to each other and can be solved jointly by C-C method and time window method [3, 4].

From the perspective of system theory, the elements that make up the system are related to each other [5]. If we neglect the links between elements, we will ignore certain important attributes of the system, and even make it difficult to form the system. This article believes that the joint solution of two parameters is more in line with the actual situation. Currently, in the process of joint solution of parameters, the description of the global features of the system is not yet sufficient. On the one hand, the constraints between two parameters are mainly established between adjacent coordinates, without considering the overall effect, and on the other hand, they are not used. Describe the global quantity to construct the objective function. Even if the relevant research involves the above two aspects, the two
are not considered together. In this paper, we try to construct a functional relationship between the quantity describing the complexity of the system and the two parameters of embedding dimension and delay time, and find the embedding corresponding to the minimum complexity within the acceptable range by finding the constraints describing the characteristics of the global system. The dimension and delay time are the best parameters [4].

Differential entropy can be used to describe the degree of system uncertainty by describing the degree of uncertainty of the system. The relationship between differential entropy and embedding dimension, delay time is used as an objective function, and the global feature of the system is described by an autocorrelation function to establish the constraints of the embedding dimension and the delay time, using the relevant optimization theory to obtain the minimum differential entropy, and getting the best embedding dimension and delay time. The solution of this optimal problem, due to the complex relationship of functions, is difficult to solve in the conventional way. In this paper, we consider the optimal particle swarm optimization algorithm to obtain the optimal value. Finally, the predictive simulation of the annual variation data of Lorenz system and sunspots is used to verify the rationality of this method [5, 6].

2. Differential entropy model

According to Takens' theorem [1], one-dimensional chaotic systems can establish the following phase space by embedding dimension \( m \) and delay time \( \tau \):

\[
X(t) = [x(t), x(x + \tau), \cdots, x(t + (m-1)\tau)]
\]

(1)

2.1. Establish objective function

The estimation of differential entropy is based on the description of the topological features of the system. In this paper, the Euclidean distance is used to describe the relationship with \( X(j) \) and its nearest neighbor \( X^{NN}(j) \):

\[
\rho(j)^2 = \|X(j) - X^{NN}(j)\|^2
\]

(2)

Based on the topological description of the above formula, the estimation of the differential entropy can be calculated by the following relationship:

\[
H(x) = \sum_{j=1}^{N} \ln[N \rho(j)] + \ln 2 + C_E
\]

(3)

Where, \( N \) represents the number of data, \( C_E \) represents Euler's constant.

The calculation of differential entropy has higher requirements on the robustness of data, but many original data are less robust to differential entropy calculation. In order to solve this problem, it is necessary to perform certain processing on the original data to obtain a series of alternative data for calculation. The alternative data generation methods are mainly divided into two types: traditional implementation methods and modern implementation methods. Traditional implementations [4, 5] mainly include: FT (Fourier Transform surrogate data) alternative data algorithm, AAFT (Amplitude Adjusted Fourier Transformed surrogate data) alternative data algorithm, IAAFT (Iterative Amplitude Adjusted Fourier Transformed surrogate data) alternative data algorithm, etc. However, the alternative data generated by the traditional method lacks randomness while taking less consideration of the dynamic characteristics of the original data. In order to overcome the limitations of traditional methods, the generation of alternative data in this paper is based on modern implementation [6]. This method is based on a regression graph and produces data that is consistent with the original data's probability distribution and nonlinear dynamic characteristics, with a high degree of matching on the waveform.
The original data \( x \) generated the replacement data \( x_{i,j} \) through the modern implementation method, and the corresponding differential entropy is denoted as \( H(x_{i,j}, m, \tau) \), thus, we can define the differential entropy rate as [6]:

\[
I(m, \tau) = \frac{H(x, m, \tau)}{\langle H(x_{i,j}, m, \tau) \rangle}
\]  

(4)

Where, \( \langle \cdot \rangle \) represents the average of the number of calculations, generally taking 5 or 10.

In the process of differential entropy calculation, the embedding dimension is more likely to be interfered with the delay time, affecting the subsequent results, and in the specific application, it is hoped that the embedding dimension will be as small as possible on the basis of ensuring the reconstruction effect [7]. Therefore, the penalty factor of embedding dimension is introduced in the differential entropy method. The final objective function can be expressed as:

\[
R_{ent}(m, \tau) = I(m, \tau)(1 + \frac{m\ln N}{N})
\]  

(5)

2.2. Establish constraints

This article establishes the model based on the system point of view and finds the corresponding relationship of phase space reconstruction related parameters through the overall description. Although the differential entropy can derive the objective function of the model from the overall uncertainty of the system, the delay time in the process of the specific study only considers the phase space reconstruction from the perspective of two adjacent coordinates, ignoring the overall effect. If the delay time \( \Gamma \) considered from the overall point of view may not be equal to the delay time \( \tau \) obtained only from the local consideration, the former is better than the latter because it considers more influence factors. From the overall point of view, relevant scholars have conducted corresponding research and proposed the concept of the embedded window length. Although this concept is still somewhat controversial, the following studies have found that this beneficial attempt is effective in solving the problem. In the following, we draw on this idea to propose constraints that need to be satisfied for the embedding dimension and delay time.

According to relevant literature analysis, from the perspective of the correlation between delay time and embedding dimension, the following relation between the embedded window length, delay time, and embedding dimension can be obtained [8]:

\[
\Gamma = (m-1)\tau
\]  

(6)

There are many methods to solve the selection of embedded time and delay time. Among them, an autocorrelation function has been proposed from a global perspective. The autocorrelation function of time series \( x_i = x(t + i\Gamma) \) (\( i = 1, 2, \ldots, N \)) can be obtained by referring to the relevant ideas of autocorrelation function as follows:

\[
c(\Gamma) = \frac{\sum_{k=0}^{N-1} [x(t + (k+1)\Gamma) - x_u][x(t + k\Gamma) - x_u]}{\sum_{k=0}^{N-1} [x(t + k\Gamma) - x_u]^2}
\]  

(7)

\[
x_u = \frac{1}{N-1} \sum_{k=0}^{N-1} x(t + k\Gamma)
\]

Where, \( x_u \) is the average of the time series.

The above functions describe the correlation of adjacent time series. For the reconstructed phase space, it is necessary to maintain the independence of each dimension as far as possible, and at the same time ensure that there is a certain connection. When \( |c(\Gamma)| = 0 \), there is no correlation between adjacent time series, indicating that the time delay between the time series exceeds the actual situation, in this case will lead to the loss of many information, the problem of the original sample size will
make heavy The phase space of the structure gets the wrong result. When \(|c(\Gamma)| = 1\), the correlation between adjacent time series is linear, indicating that the time delay between time series is too small, and overlapping in the reconstructed phase space leads to information redundancy, which adds unnecessary calculations. At the same time, it easily leads to wrong conclusions. Therefore, adjacent time series should maintain appropriate correlation, relative studies found that the value of the relationship should be guaranteed to meet \(0.5 < c(\Gamma) < 0.8\), in this interval can ensure that the adjacent time sequence is kept as independent as possible, but also there is a certain contact. To sum up, a differential entropy model with embedded dimension and delay time can be obtained as follows [9]:

\[
\min R_{ae}(m, \tau) = l(m, \tau)(1 + \frac{\ln N}{N}) \geq 0.5
\]

\[
\frac{\sum_{i=0}^{N-1} [x(t + (k + 1)(m - 1)\tau) - x_i] [x(t + k(m - 1)\tau) - x_i]}{\sum_{i=0}^{N-1} [x(t + k(m - 1)\tau) - x_i]^2} \leq 0.8
\]

\[
m, \tau \in \mathbb{R}^n
\]

3. Model solving

The differential entropy model established in the previous section has a large amount of computation when using the traditional method. Considering that the particle swarm algorithm [10] is simple and easy to use and has a fast convergence rate, the particle swarm algorithm is used to improve the above model. The particle swarm algorithm regards the solution of the objective function as a particle, and tracks the individual extreme value \(p_{best}(t)\) and the global extreme value \(g_{best}(t)\) in iterative iterations to update the state of the particle itself until it reaches the state that meets the condition, thereby obtaining the optimal solution of the objective function. The state of a particle includes both a velocity change and a position change, and the change can be defined by the following relational expression.

\[
v_i(t + 1) = w \times v_i(t) + c_1 \times r_1 \times [p_{best}(t) - x_i(t)] + c_2 \times r_2 \times [g_{best}(t) - x_i(t)]
\]

\[
x_i(t + 1) = x_i(t) + v_i(t)
\]

Where, \(c_1\) and \(c_2\) are learning factors, which are used to adjust the maximum step size of particle flight to individual extremes and global extremes; \(r_1\) and \(r_2\) are random numbers used to adjust the search range; \(w\) is the inertia weight, which is used to balance the global and local search capabilities.

According to the actual situation of the problem in this paper, we will improve the basic particle swarm algorithm to solve the problem in this paper. The adjustment of the inertia weight is essentially a change in the flight speed of the particle, and the change in the flight speed determines the change in the position of the particle. When the inertia weight is large, the individual particle has a larger flight speed, which can perform a global search better, but it is easy to fly over the optimal position; when the inertia weight is small, the individual particle has a smaller flying speed, and the particles are more inclined to local search, but the calculation speed is slow and easy to fall into local optimum. Therefore, the choice of inertia weight is very important for the global search and local search of particle balance. This article considers many documents based on the consideration that particles need to have a relatively large flight speed in order to achieve a better global search capability. From the global search to the later stages of the iteration, the particles need to have a relatively small flying speed in order to achieve a better local search capability. This article proposes the following design scheme:
\[ W = W_{\text{max}} - \frac{W_{\text{min}}}{1 + e^{11 \times ((w_{\text{max}} - w_{\text{min}}) - \frac{t}{t_{\text{max}}})}} \]  

(10)

Where, \( W_{\text{max}} \) represents the maximum value of inertia weight; \( W_{\text{min}} \) represents the minimum value of inertia weight.

The adjustment of the random number is essentially the adjustment of the global search ability. The randomness of the random number guarantees the global search ability of the algorithm. Therefore, the random number generation method has a greater impact on the improvement of the algorithm. Taking into account the good randomness of chaos, we can achieve the effect of traversing the whole number domain. We refer to the Logistics system to improve the corresponding algorithm.

\[ x_{k+1} = \mu x_k (1 - x_k) \]  

(11)

In this chaotic system, the chaotic domain is \((0,1)\), \(x_k \in (0,1)\). When \(3.6 \leq \mu \leq 4\), the system is in a chaotic state. Referring to this idea, we can get the random number generation algorithm as shown below.

\[
\begin{cases}
    r_i(t+1) = 3.8 \times r_i(t)(1-r_i(t)) \\
    r_i(t) \in (0,1) \quad i = 1,2
\end{cases}
\]  

(12)

The random number obtained by the above method can have a better global search capability and further optimize the particle swarm algorithm.

In order to ensure that the algorithm does not converge prematurely, the steady state of the algorithm needs to be broken during the specific calculation process. This article refers to the idea of crossover operator and mutation operator in the genetic algorithm to update the position and velocity of the particle to achieve further search effect.

In the calculation of the crossover operator, the corresponding selection probability is \( P_a \). Before the crossover operator selects the population, the population must be sorted according to the degree of fitness, and corresponding operations are performed on the sorted populations. The first \( N \times P_a \) sorted particles are regarded as parent \( i \), and they are recombined with their different parent \( j \). The position and velocity of the updated particles can be expressed as follows.

\[
x'_i(t) = P_a \times x_i(t) + (1 - P_a) \times x_j(t)
\]

\[
v'_i(t) = \frac{[v_i(t) + v_j(t)] \times |v_i(t)|}{\sqrt{v_i^2(t) + v_j^2(t)}}
\]  

(13)

In the calculation of the mutation operator, the corresponding mutation probability is \( P_c \). In the mutation operation, only the position of the particle is changed, and the flying speed of the particle is not changed. Treating \( N \times P_c \) particles as a parent, the position of the calculated child particles is:

\[ x'_i(t) = x_i(t) \times (1 + P_c) \]  

(14)

4. Example

The input vector required by the network prediction model is \( X(N \times m) \), and the output value is \( Y \), \( N \) is the total length of the time series. To guarantee the accuracy of network training, the network sample data must be normalized and preprocessed [11].

\[ X = \frac{(a-b) \times (x-x_{\text{min}}) + b}{x_{\text{max}} - x_{\text{min}}} \]  

(15)
Where: \( a, b \) are the upper and lower lines of the normalized interval, respectively, in this paper are \( 1, -1 \), that is, the normalized interval is \([-1, 1]\).

The method proposed in this paper uses MATLAB2012a to simulate the chaotic time series and obtain corresponding results for different situations. In order to compare the various prediction results uniformly, this paper uses two methods of calculating the error, namely the normalized variance and prediction accuracy. Their definitions are as follows [12]:

\[
E_{PA} = \frac{\sum_{i=1}^{T} (\hat{h}(t) - \hat{h}_m)(h(t) - h_m)}{(T - 1)\hat{\sigma}_h \sigma_h}
\]

(16)

Where: \( T \) denotes the number of predictions, \( \hat{h}(t) \) represents the actual value at the time (i.e., the predicted target value), \( h(t) \) represents the predicted value of the model at the time, \( \hat{h}_m \) represents the average of the true values, \( h_m \) represents the average of the model predicted values, and \( \hat{\sigma}_h \) represents the standard deviation of the true values \( \sigma_h \) represents the standard deviation of the model predictions [13, 14].

\[
\varepsilon^2(m,d) = \frac{\sum_{n=M+1}^{N} [\hat{x}_n - x_n]^2}{\sum_{n=M+1}^{N} [x_n - \bar{x}_n]^2}
\]

(17)

Where: \( x_n (n=1, 2, \ldots, L) \) represents the actual value and \( \hat{x}_n (n=1, 2, \ldots, L) \) represents the predicted value obtained from the predictive model [15].

In the ideal case, that is, all predicted values are equal to the true value, the above-mentioned various measurement error methods should be \( \varepsilon^2 = 0, E_{PA} = 1 \).

The Lorenz chaotic time series and sunspots are predicted separately. First, consider the Lorenz equation.

\[
\begin{align*}
\dot{x} &= \sigma(y - x), \\
\dot{y} &= rx - y - xz, \\
\dot{z} &= -bz + xy,
\end{align*}
\]

(18)

Figure 1. Chaotic attractors.

Where, \( \sigma = 16, b = 4, r = 45.92 \), this paper uses Runge-Kutta method to solve the equation, the step length is \( h = 0.01 \). The transient process was removed, Lorenz chaotic time series was obtained, 3000 pairs of input and output data after delay coordinate transformation were extracted, the former 2500 pairs of data were taken as experimental data, and the last 500 pairs of data were verification data, and
20 sets of substitute data were generated using the modern implementation method. According to the calculation of differential entropy, when the embedding dimension is \( m = 3 \) and the delay time is \( \tau = 10 \), the minimum value of differential entropy is \( R_{\text{ent}}(m, \tau) = 0.9238 \). The following figure 1-5 show the original attractors of the Lorenz system, the reconstructed attractors in the reconstructed phase space, and the prediction of the corresponding data. It can be seen from the graph that the attractor constructed in this way better maintains the corresponding dynamic characteristics.

\[ R_{\text{ent}}(m, \tau) = 0.9238 \]

Sunspot is a low-dimensional chaotic system. The data used in this paper comes from the SIRC (solar influences data analysis center), which mainly uses the annual average data of sunspots from 1700 to 1997. Among them, a total of 263 points were used as training samples from 1700 to 1962 and 20 points from 1963 to 1982 were used as independent test samples. In this paper \( R_{\text{ent}}(m, \tau) = 0.0892 \), the delay time is \( \tau = 1 \) and the embedding dimension is \( m = 5 \), and the model is predicted. The following shows the annual changes in sunspots and their corresponding predictions.

\[ R_{\text{ent}}(m, \tau) = 0.0892 \]
5. Conclusions
This paper establishes the relationship between differential entropy and embedding dimension and delay time as the objective function, describes the global features of the system by autocorrelation function, establishes the constraints of embedding dimension and delay time, and uses relevant optimization theories to obtain the minimum differential entropy. The best embedding dimension and delay time can be obtained. The solution of this optimal problem, due to the complex relationship of functions, is difficult to solve in the conventional way. In this paper, we consider the optimization of particle swarm optimization algorithm to obtain the optimal value.

Acknowledgments
This work is supported by the National Nature Science Foundations of China under Grant number 71601183.

References
[1] Takens F 1980 Detecting strange attractors in turbulence Dynamical Systems and Turbulence Warwick Springer Berlin Heidelberg 1981 366-381
[2] Kantz H and Schreiber T 1997 Nonlinear time series analysis Cambridge University press
[3] Eckmann J P and Ruelle D 1985 Ergodic theory of chaos and strange attractors Turbulence, Strange Attractors And Chaos 365-404
[4] Abarbanel H D I 1996 Analysis of Observed Chaotic Data Technometrics 39(3) 334-335
[5] Kim H S, Eykholt R and Salas J D 1999 Nonlinear dynamics, delay times, and embedding windows Physica D-nonlinear Phenomena 127(1–2) 48-60
[6] Li M W, Geng J, Han D F, et al. 2016 Ship motion prediction using dynamic seasonal Rv SVR with phase space reconstruction and the chaos adaptive efficient FOA Neurocomputing 174(4) 661-680
[7] Tang J, Liu F, Zhang W, et al. 2016 Exploring dynamic property of traffic flow time series in multi-states based on complex networks: Phase space reconstruction versus visibility graph Physica A Statistical Mechanics & Its Applications 450 635-648
[8] Men H, Sun B, Zhao X, et al. 2016 Dynamics of stainless steel corrosion based on the theory of phase space reconstruction and chaos Anti-Corrosion Methods and Materials 63(3) 214-225
[9] Fedotenkova M, Graben P B, Sleigh J W, et al. 2016 Time-Frequency Representations as Phase Space Reconstruction in Symbolic Recurrence Structure Analysis[C]// International Work-Conference on Time (Springer, Cham) 89-102
[10] Tang C, Guo S, Yanyu M O, et al. 2016 An Optimal Design of Structural Reliability Based on Particle Swarm Optimization-Sequential Quadratic Programming Algorithm Journal of Air Force Engineering University (Natural Science Edition) 17 (2) 107-111
[11] Zhang X, Lv M, Chen H, et al. 2018 Chaotic characteristics analysis of the sintering process system with unknown dynamic functions based on phase space reconstruction and chaotic invariables Nonlinear Dynamics 1-18
[12] Xi G U and Liao Z 2017 Short-term load forecasting based on phase space reconstruction and Gaussian process regression Power System Protection & Control 45(5) 73-79
[13] Hong M, Wang D, Wang Y, et al. 2016 Mid- and long-term runoff predictions by an improved phase-space reconstruction model Environmental Research 148 560-573
[14] Li D, Xie Q, Jin Q, et al. 2016 A sequential method using multiplicative extreme learning machine for epileptic seizure detection Neurocomputing 214 692-707
[15] Fan Z, Xu Y, Zuo W, et al. 2017 Modified principal component analysis: an integration of multiple similarity subspace models Transactions on Neural Networks & Learning Systems IEEE 25(8) 1538-1552