Research Article

Use of False Nearest Neighbours for Selecting Variables and Embedding Parameters for State Space Reconstruction

Anna Krakovská, Kristína Mezeiová, and Hana Budáčová

Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská Cesta 9, 842 19 Bratislava, Slovakia

Correspondence should be addressed to Anna Krakovská; krakovska@savba.sk

If data are generated by a system with a d-dimensional attractor, then Takens' theorem guarantees that reconstruction that is diffeomorphic to the original attractor can be built from the single time series in $(2d + 1)$-dimensional phase space. However, under certain conditions, reconstruction is possible even in a space of smaller dimension. This topic is very important because the size of the reconstruction space relates to the effectiveness of the whole subsequent analysis. In this paper, the false nearest neighbour (FNN) methods are revisited to estimate the optimum embedding parameters and the most appropriate observables for state space reconstruction. A modification of the false nearest neighbour method is introduced. The findings contribute to evidence that the length of the embedding time window (TW) is more important than the reconstruction delay time and the embedding dimension (ED) separately. Moreover, if several time series of the same system are observed, the choice of the one that is used for the reconstruction could also be critical. The results are demonstrated on two chaotic benchmark systems.

1. Introduction

State space reconstruction is usually an unavoidable step before the analysis of a time series in terms of dynamical systems theory. Suppose that we have data (a single time series) that was presumably generated by a d-dimensional deterministic dynamical system. Then, the usual choice for a reconstruction is a matrix of time shifts of one variable, as supported by Takens theorem from 1981 [1]. Alternate methods of reconstruction, such as derivatives or linearly independent coordinates found by principal component analysis, can be seen as transformations on time-shift vectors. By one of these embedding procedures, a new state space is created that is (in the sense of diffeomorphism) equivalent to the original state space. The reconstruction preserves relevant geometrical and dynamical invariants, such as the fractal dimensions of the attractor, the entropies, or the Lyapunov exponents (which measure the sensitivity to the initial conditions).

Reconstructing requires decision making regarding the size of the space of the reconstruction, the value of the time shifts between the coordinates, and another important—although often overlooked—aspect: which one or which combination of observables (if several of them are available) are to be used for the reconstruction?

Choice of the Time Delay. The time-delayed versions $[y(t), y(t - \tau), y(t - 2\tau), \ldots, y(t - 2m\tau)]$ of the known observable $y(t)$ form an embedding from the original m-dimensional manifold into R^{2m+1} (where $2m + 1$ is the embedding dimension and τ is the time lag between consecutive states) [1, 2]. Theoretically, for noise-free data of unlimited length, the existence of a diffeomorphism between the original attractor and the reconstructed image is guaranteed for almost any choice of delay and a sufficiently high embedding dimension. In practice, however, the experimental time series can be short and noisy. Then, the quality of the reconstruction can vary depending on the choices for the time delay and the embedding dimension. If the delay is too small, then each coordinate is almost the same, and the reconstructed trajectories resemble a line (the phenomenon known as redundancy). Geometrically, this

arrangement means that there are trajectory intersections at a small angle. In the case of noisy measurements, this circumstance makes the separation of trajectories impossible. On the other hand, if the delay is too large, then due to the sensitivity of the chaotic motion, the coordinates appear to be independent, and the reconstructed state portrait looks random or unnecessarily complicated (a phenomenon known as irrelevance). Such an extremely inappropriate choices for the delay can be detected at first sight in a 2-dimensional delayed plot.

To select the embedding parameters optimally, many competing approaches have been proposed. Most of them are based on heuristic reasoning rather than mathematically rigorous criteria. The simple idea is to unfold the reconstruction of the trajectories sufficiently to avoid self-crossing and extreme closeness of distinct parts. In particular, the delay that is used for reconstruction is often given by the first zero of the autocorrelation function or as the first minimum of the mutual information between the delayed components [3]. By using the first instead of the absolute minimum of the mutual information, the selection is biased toward small delays, to avoid irrelevance. The benefit of using the mutual information, as opposed to the autocorrelation function, is that the nonlinear character of the data is accounted for.

One method to avoid selecting the time delay is to use derivatives instead of delayed coordinates. In addition to the fact that this approach makes the embedding procedure delay-free, the derivative coordinates offer some further advantages. First, in some applications, they enable a clear physical interpretation. Moreover, the prediction results that are obtained in differential phase space could be better than in the time-delay phase space [4]. However, the largest problem is that the numerical estimate of the derivatives leads to errors and deteriorates quickly when calculating higher order derivatives. Any noise in the data would make the situation even worse [5].

Choice of the Embedding Dimension. In addition, to the time shift, you must choose a proper embedding dimension to be able to reconstruct the state portrait. The theorem of Whitney guarantees the possibility of embedding any m-dimensional smooth manifold into $(2m + 1)$-dimensional Euclidean space [6]. Sauer et al. generalised the theorem to fractal objects. They have proved that, under some conditions regarding periodic orbits and the measurement function, almost every C^1 map from the fractal A to R^d with $d > 2D_A$ forms an embedding, whereby D_A is the box-counting dimension of A [7]. This finding means that it is not the size m of the manifold of the original attractor that determines the minimal embedding dimension but only the fractal dimension D_A. However, even $2D_A$ represents only an upper limit—the embedding theorem does not rule out an embedding dimension that is lower than $2D_A$.

Sometimes, the required size of the reconstruction space can be smaller because of the less demanding goal of the investigation. For example, for the numerical calculation of the correlation dimension of the attractor A, any dimension above the box-counting dimension of A is sufficient [8]. Of course, such cases do not guarantee that the attractor

is mapped one-to-one; however, that is not necessary for dimension estimation.

On the other hand, if the objective is to model or predict the future behaviour, then self-intersections are unacceptable and a reconstruction of a dimension as high as $d > 2D_A$ might be needed. However, in favourable cases, embeddings into lower than $2D_A$-dimensional spaces could still exist. It is definitely worthwhile to explore such possibilities because, in practice, it is advantageous to construct embeddings of the lowest possible dimension (ideally of the original system's dimension). What are the favourable cases and how not to miss them are now a subject of research. For example, Cross and Gilmore contributed to the issue when they analysed differential mappings of the rotationally equivariant Lorenz dynamical system [9]. They showed that, while the differential reconstruction based on the x coordinate is an embedding of the attractor in three dimensions, it does not yield an embedding of the entire manifold; that is, the projection of the manifold into R^3 possesses singularities. However, it is possible to embed the manifold into a 3-dimensional twisted submanifold of R^4. Then, not only diffeomorphism invariants (as fractal dimensions or Lyapunov exponents) but also information about the mechanism responsible for generating the chaotic behaviour is preserved. The two objects are actually isotopic (smoothly deformable into each other) in R^4. Nonisotopic embeddings provide distinct representations of the original state space because one might not be deformed into another without self-intersection. For $m \geq 2$, any two embeddings of an m-manifold into R^{2m+1} are isotopic [10]. This result is known as an isotopy version of the strong Whitney embedding theorem. Moreover, Cross and Gilmore have shown that for 3-dimensional systems (if genus $g = 1$ and $g \geq 3$), all of the representations become equivalent for R^5 already [11]. This result is, however, limited to attractors that exist in a 3-dimensional manifold because the considered topological indices are restricted to three dimensions. Very little is known about lower than $(2m + 1)$-dimensional embeddings of dynamical systems with $m > 3$.

The choice of the minimal possible embedding dimension when the number of degrees of freedom of the original system is unknown and is not easy. A space of an undervalued dimensionality does not unfold the trajectories, while an unnecessarily large embedding space can result in overfitting. Typically, the search for the proper dimension is based on a step-by-step expansion of the reconstruction space while simultaneously following some proper diffeomorphism invariant that is expected to stay constant after reaching the sufficient embedding dimension. As examples of such invariants, the correlation dimension, largest Lyapunov exponent, predictability indices, or percentage of false nearest neighbours have previously been mentioned. In the long run, the various options have been superseded in practice by the false near neighbour test [12].

Choice of the Observable. When considering the Takens or Sauer theorem, the variables of the system are assumed to be in equal positions regarding their use for the state space reconstruction. For example, the theorems guarantee that a 5-dimensional delay reconstruction from any variable

(x, y, or z) of the Rössler system constitutes a diffeomorphism between the original manifold and the reconstructed image. However, in the case of the variable y, already a 3-dimensional differential reconstruction suffices for diffeomorphism [13, 14]. Due to computational and modeling reasons, we would like to know whether some variables lead to a diffeomorphism in a space of smaller dimension than others and to know how low the minimal possible embedding dimension is. To contribute to solving this problem, Letellier et al. defined some observability indices that enable ranking of the observables according to their effectiveness in the reconstruction process [13, 15–17].

Problems with the Standard Estimates of the Embedding Parameters. In practice, the most commonly used method for selecting the embedding parameters consists of the first minimum of the mutual information to estimate the time delay and the FNN test to find the sufficient embedding dimension.

It should be emphasised that the selection of the delay for reconstruction, which is based on the mutual information, holds for 2-dimensional embeddings but not necessarily for higher dimensional embeddings. Even in the 2-dimensional case, the criterion can be regarded as effective only for a time series that has a single, dominant periodicity or recurrence time. In that case, the suitable lag is approximately one-quarter of the dominant period, and this value is in good agreement with the minimum of the mutual information or the first zero of the autocorrelation function. However, the same delay time is often used regardless of the number of delay vectors that form the reconstruction, although some authors suggest lowering the delay time when increasing the dimension. They argue that the independent parameter that should be estimated is not the delay τ or the embedding dimension m separately but rather the whole embedding time window (TW), which is given as TW = $(m-1)\tau$ [18–22]. Despite all this, no standard procedure for estimating the time window has emerged yet and most researchers continue to use the same time delay regardless of the size of the reconstruction space.

In this work, we want to contribute to the debate about the importance of the time window and the possibility of using FNN methods as a tool for the optimal embedding parameters selection.

The paper is organised as follows.

In Section 2, a short review of the methods that use the idea of false nearest neighbours to estimate the parameters of the state space reconstruction is given. We also discuss the importance of the correct choice of observables, if several of them are at our disposal. Then, we describe the data that is used as benchmarks, and we introduce the three methods that were used for testing the data. In Section 2.3 a rank-based modification of false nearest neighbour method is proposed.

In Section 3, we present the results that regard the effects of the reconstruction on the evolution of the false nearest neighbours, on the estimates of the correlation dimension, and on the errors in the predictions.

Finally, the findings are discussed and summarised in Section 4.

2. Data and Methods

2.1. False Nearest Neighbours Algorithms. The false nearest neighbours method is the most popular tool for the selection of the minimal embedding dimension. This method is based on the assumption that two points that are near to each other in the sufficient embedding dimension m should remain close as the dimension increases. However, if the embedding dimension m is too small, then the points that are in reality far apart could seem to be neighbours (as a consequence of projecting into a space of smaller dimension). The various modifications of the method apply geometrical reasoning: one increases ED until the reconstructed image is unfolded. The method checks the neighbours in increasing embedding dimensions until it finds only a negligible number of false neighbours when going from dimension m to $m + 1$. This m is chosen as the lowest embedding dimension, which is presumed to give reconstruction without self-intersections.

In the case of clean deterministic data, we expect that the percentage of false neighbours will drop to zero when the proper dimension is reached. If the signal is too noisy, however, it could be that the method fails due to efforts to unfold the noise.

From several variants and modifications of the false neighbours algorithms [12, 23–26] we would describe the two most commonly used.

2.1.1. Kennel's Algorithm. The false nearest neighbours method, as introduced by Kennel et al., is an iterative process [12]. An m-dimensional state portrait is reconstructed by taking the time-delayed coordinates of the observed time series. The time delay is set as the first minimum of the mutual information function [3]. Then, the algorithm takes each point in the m-dimensional portrait and finds the distance $R(m)$ to its nearest neighbour and, afterward, the distance $R(m+1)$ between the two points in $m+1$ dimensions. If $\sqrt{(R^2(m+1) - R^2(m))/R^2(m)} > R_{\text{tol}}$ where R_{tol} is some threshold, then the neighbours are said to be false. One then repeats the process at higher dimensions, stopping when the proportion of false nearest neighbours becomes zero or sufficiently small and will remain so from then onward. For clean data with an infinite length, this criterion would be sufficient to determine the proper embedding dimension. However, noise with a limited amount of data would erroneously produce a finite embedding dimension. The problem turns out to be that even though two points are the nearest neighbours, they are not necessarily close to each other. Such points are considered to be false neighbours, and this arrangement is checked by the second criterion: $R(m + 1)/R_A > A_{\text{tol}}$, where R_A is an estimate of the attractor size and A_{tol} is the second threshold. The authors advocate using this pair of criteria jointly by declaring a nearest neighbour as false if either test fails. For data sets of similar size and complexity, as in our study, Kennel et al. recommend the next settings of the thresholds: $R_{\text{tol}} \approx 15$, $A_{\text{tol}} = 2$. For each dimension, the percentage of the false nearest neighbours is calculated. Eventually, the lowest possible dimension with no more false neighbours than

what we are prepared to tolerate is declared as the optimal embedding dimension.

2.1.2. Cao's Algorithm. One of the problems of the FNN method stems from the subjective choice of several parameters: Kennel's algorithm, for example, uses R_{tol} and A_{tol} to distinguish between true and false neighbours and another threshold parameter to determine when the fraction of FNN is sufficiently small (to allow the reconstruction space to be declared as sufficiently large). Unfortunately, for different thresholds of parameters, the algorithm could lead to different estimates of the optimal embedding dimension. To avoid this subjectivity, Cao introduced a modified algorithm that is sometimes called an averaged false neighbours method [24]. Instead of testing the neighbours to be false or not, Cao calculates how, on average, the distances between the nearest neighbours change after going from dimension m to $m + 1$. The dimension at which the change stops is taken as the proper embedding dimension, assuming that the trajectory is fully unfolded, and adding another dimensions does not change the average distance between the nearest neighbours. The main advantage of this method is that the number of subjectively chosen thresholds of parameters is reduced.

2.1.3. Comparison of the FNN Methods. To compare different FNN methods, Cellucci et al. [27] have tested them on the Rössler system and the Mackey-Glass equation. Five criteria for selecting embedding parameters have been applied to the observables of the systems. For the resulting combinations of embedding parameters, the largest Lyapunov exponent was calculated by using a procedure published by [28] and was compared against those that were determined by the more exhaustive analytically based calculations published by Benettin et al. [29]. The criterion that reproduced best the reference values of the Lyapunov exponents was considered to be the most successful. The best identification of the embedding dimension has been achieved with the method of Kennel [12], and the best value of the time shift has been found by using the mutual information.

In another comparative study, which was conducted by Letellier and his colleagues [30], three classical tests for whether a mapping is an embedding, depending on the geometric and dynamical measures, were compared with a fourth test, which depended on a topological measure (the Gauss linking number). The tests involved estimating the fraction of false near neighbours, the correlation dimension, and the largest Lyapunov exponent as a function of the embedding parameters. Finally, the topological test that was proposed by the authors was based on the idea that in regions where intersections of unstable periodic orbits occur and the linking numbers of the orbits change the mapping cannot be an embedding. For the testing examples, the authors used a periodically driven Takens-Bogdanov oscillator and a modification of the Malkus-Robbins equations, which were originally introduced to model the action of a self-exciting dynamo. Due to limitations of the topological test, a comparison of the methods could be performed only for mappings into three dimensions. The authors have found that the classical tests often fail to identify when the mapping is

an embedding. They have suggested that all claims for successful embeddings into three or higher dimensions that were based on geometric or dynamical methods should be treated with the greatest skepticism.

2.2. Suitability of Variables for Reconstruction. Another issue that is not satisfactorily resolved concerns the fact that if we have more observables from the same system, they do not appear to be equivalent with respect to the phase space reconstruction. It appears that different variables could contain different levels of information [13, 21]. For example, it is much easier to obtain a global model from the variable y of the Rössler system than from the variable z.

Moreover, Whitney's embedding theorem ensures us that a combination of different variables could form an embedding as well. For example, multichannel measurements are typical for neurology. In such cases, instead of time-lagged copies of a single variable, you can use several different simultaneously taken observables or you can yield a mixed time-delay and multivariate embedding.

As already mentioned in Introduction, it would be useful to have an index that enables a ranking of the observables according to their effectiveness in the reconstruction process. In control theory, the notion of observability is well defined for linear systems. The linear system is evaluated as either observable or not. If a system is observable, then from the system's outputs, it is possible to determine the behaviour of the entire system. If it is not observable, then the output data disallows us from estimating the states of the system completely. To check if a linear system with n states is observable, the rank of the so-called observability matrix is calculated. If it is equal to n, then the rows are linearly independent, the initial state can be recovered from a sequence of observations and inputs, and the system is observable in Kalman's sense [31].

Since 1998, to extend the theory of observability to nonlinear systems, Letellier et al. introduced several measures that rank the variables of the system according to their observability [13, 15–17]. The indices were derived for systems that have known equations, and they appear to be ranking the variables quite well. In [16], a procedure for comparing two observables of the same system without a need for the system equations is proposed. This time-series approach is based on the so-called omnidirectional nonlinear correlation functions, and it agrees relatively well with the earlier indices with respect to the observability order of some benchmark systems' variables.

In this paper, the so-called symbolic observability coefficient will be used for comparison purposes [17]. Its computation requires knowledge of the equations of the system, and it is based on the so-called fluency matrix, which emphasises constant and nonconstant elements of the Jacobian matrix; these elements correspond to linear and nonlinear terms in the vector field of the system. The symbolic observability coefficients are greater than one when the dimension of the reconstructed state space is too large. They allow choosing from the system equations the best variable or the best combination of variables for univariate (resp., multivariate) reconstruction. The observability coefficients provide

an upper limit for the size of the reconstruction space that is sometimes smaller than those provided by the Takens criterion. These computations indicate that the observability is more related to the couplings between dynamical variables than the dynamical regime itself. The authors also claim that the observability is related to the possibility of rewriting the system in a polynomial form while using only the chosen observable. Provided that the coordinate transformation is a global diffeomorphism in n-dimensional space and the original system is polynomial, the system can be rewritten under the form of an n-order ordinary differential equation in a polynomial form.

2.3. False First Nearest Neighbour (FFNN) Method. In this paragraph, let us introduce a new modification of false neighbour methods, which we then use to find the best embedding parameters. The basic idea is to use rank-based modification of FNN method and to create maps that visualise the evaluation of false neighbours for combinations of values of delay and embedding dimension.

When designing the algorithm, we intended to not leave room for subjective choices of thresholds in the method. Moreover, we also aimed to reduce another serious problem of the FNN methods, which is related to the phenomenon called the curse of dimensionality. When the dimensionality increases, the volume of the space grows so fast that soon the available data become sparse. To obtain statistically reliable results, the amount of data would need to grow exponentially with the dimensionality. Even with enormously large data sets, it is usually not recommended to use the algorithms in more than 10–15 dimensions. In [32], the authors explored the effect of increasing the dimensionality on the nearest neighbour problem. They showed that under a broad set of conditions, as the dimensionality increases, the distance to the nearest data point approaches the distance to the farthest data point. This arrangement is obviously a problem because it indicates poor discrimination ability, which arises from the fact that all of the distances between pairs of data elements appear to be very similar. In such cases, the use of rank-based measures can be considered because they appear to be less prone to the curse of dimensionality compared with to the primary distances from which the rankings are derived [33].

In this study, we use quite large data sets and low dimensions. Therefore, problems with dimensionality need not be critical, and theoretically, we could use the performance of the average distance of the nearest neighbours (the method of Cao) for creating the maps. Nevertheless, we suggest avoiding using the distances, and we prefer to rely only on counting the number of shared neighbours. In particular, we assess the performance of the first nearest neighbours as going from m-dimensional space to $(m + 1)$-dimensional, to obtain a secondary measure that is induced by the primary distance measure (Euclidean norm here).

Here is our modification of the false nearest neighbour algorithms.

(1) Take the observable of the system, and for combinations of time delays and embedding dimensions, form time-delay reconstructions.

(2) Both in m-dimensional and in $(m + 1)$-dimensional reconstructions, identify the closest point (the first neighbour in the Euclidean sense) to each point on the reconstructed trajectory.

(3) Quantify the rank-based measure of the FFNN method as a percentage of cases when the nearest neighbour to a point in m-dimensional space ceases to be the nearest neighbour of the same point in the space of one higher dimension.

(4) Visualise the results, for example, as a color-coded map to detect the best combinations of parameters for the state space reconstruction.

The best embedding parameters bring us as close as possible to the reconstruction for which the nearest neighbours remain nearest neighbours in larger state spaces. However, can we expect that the reconstruction that is chosen as optimal by the FFNN method is also the most suitable for purposes such as the estimation of Lyapunov exponents, modeling, or forecasting? Let us check this possibility at least for two applications—the estimation of the correlation dimension and a nonlinear one-point prediction. To take advantage of dynamical systems whose properties are quite well known, we are going to use the Rössler and the Lorenz system as the benchmark systems.

2.4. Rössler System. As the first test example, we use the Rössler system [34]:

$$\dot{x} = -y - z,$$
$$\dot{y} = x + ay, \qquad (1)$$
$$\dot{z} = b + z(x - c),$$

with parameters $a = 0.398, b = 2$, and $c = 4$ and the initial condition $[0, 0, 0.4]$. This system was integrated by means of the fourth order Runge-Kutta formula with integration step 0.02. The first 1500 points were discarded, and the next 100000 data points were saved.

Spectral analysis of any of the variables shows a peak, which suggests that one running around the attractor takes approximately 310 points. Then, one of the tips for the time delay for the reconstruction could be a quarter of that period, that is, 77. Finding the first minimum of the mutual information and the first zero crossing of the autocorrelation functions also suggest approximately the same value for all three variables. Therefore, the established way to proceed would be to use 77 as the time delay and to look for the minimum necessary embedding dimension by some nearest neighbours method.

For a long time, possible differences between the levels of observability of different variables were ignored. According to Takens theorem, 5-dimensional reconstruction from any of the variables ensures a diffeomorphism between the original phase space of the Rössler system and the reconstructed space. However, the investigation of the system shows that derivative reconstruction $(y, \dot{y}, \ddot{y})$ is globally diffeomorphic to the original state portrait in three dimensions already. The global model from observable y can be obtained with

TABLE 1: Mean values of the symbolic observability coefficient [17]. The coefficient is equal to 1 when the system is fully observable from a variable in 3-dimensional space.

	x	y	z
Rössler	0.88	1	0.44
Lorenz	0.89	0.46	0.35

relative ease. On the other hand, x and z need at least a 4-dimensional derivative reconstruction space, and especially the z variable is known to be a very problematic basis for reconstruction and modeling [14, 17, 35]. Moreover, other than derivative reconstruction might even need an additional dimension to fully unfold the attractor. The values of the symbolic observability degrees η confirm that y is the best and z is the worst observable of the Rössler system. The values of η can be found in Table 1.

2.5. Lorenz System. The second data set comes from the Lorenz system that has the well-known butterfly-like attractor [36]:

$$
\begin{aligned}
\dot{x} &= \sigma\left(y - x\right), \\
\dot{y} &= Rx - y - xz, \\
\dot{z} &= -bz + xz,
\end{aligned}
\tag{2}
$$

with parameters $\sigma = 10$, $R = 28$, and $b = 8/3$, and an initial condition $[0.3, 0.3, 0.3]$. The system was integrated by the fourth order Runge-Kutta method with an integration step of 0.005. The first 4000 points were removed, and the following 100000 points were saved.

The integration step was chosen to ensure that one round on the attractor responds to approximately 300 points on average. Nevertheless, the subsequent spectral analysis revealed no prominent frequencies in the data. In fact, it is not surprising for a chaotic time series. A typical trajectory of the Lorenz system stays on one wing of the attractor, circling from the inside to its peripheral border some time before it jumps to the other wing. Consequently, the autocorrelation function decays smoothly. The method of the time-delayed mutual information also refuses to provide a clear answer as to what value of delay to use. Finally, one would probably turn to the visual inspection of 2-dimensional plots to find the delay that sufficiently unfolds the trajectories. It appears that the result might be a time lag of approximately 37, which corresponds to a quarter of the average time that is spent on a single wind. We will show later whether this highly heuristic guess is usable.

In [15], the authors conclude that the nonequivalence between the observables of the Lorenz system has two basically different sources, which are the complexity of the coupling between the variables and the symmetry properties. Even for the variable x, however, the reconstruction needs at least a 4-dimensional space. We are addressing univariate reconstructions in this study, but allow us to briefly recall that multivariate reconstruction in 3-dimensional space is possible if the variable x, the derivative of x, and the variable z are used. The fact that the attractor is setwise symmetric

under a rotation around the z-axis plays a key role. As a consequence, the two wings cannot be distinguished by looking at the observable z [35]. The symbolic observability coefficients appear to give the correct observability order, indicating that z is the worst and x is the best observable. The values of the symbolic observability degrees η can be seen in Table 1.

2.6. Correlation Dimension Test. First, we are going to test the quality of the reconstruction by means of estimates of the correlation dimension (D_2). The valuation of the complexity and the fractal character of the studied chaotic process by the correlation dimension has been extensively used since 1983 when Grassberger and Procaccia proposed their computationally efficient approach to dimension estimation [37].

We are going to estimate the dimension of the Rössler and Lorenz attractor when reconstructed with different combinations of embedding parameters to find out how important the choices of m and τ are for the estimation of the correlation dimension.

The results that have been published so far are not entirely consistent; for example, the authors in [38] imply that it is the delay time itself, rather than the total observation window, that plays the most critical role in the determination of the correlation dimension. They measured the quality of the reconstruction by the length of the linear scaling region: for fixed values of the window, with a dimension of the space that is large enough to guarantee an embedding and a delay that is large enough to avoid the problems of autocorrelation, the length of the scaling region was found to be the largest for the smallest admissible value of the delay. On the other hand, in [20, 39, 40], it is shown that it is the embedding window that is crucial for estimating the correlation dimension. In [39], the results for three test cases (Rössler equations, Lorenz equations, and 3-dimensional irrational torus) lead to the warning that neither mutual information nor autocorrelation is consistently successful in identifying the optimal window.

In [21], the x and z variable of the Rössler system are given as an example of a good and a bad variable. He found that for measurements over the same epoch, the correlation dimension of the Rössler attractor was well estimated by the x-measurements but significantly underestimated by the z-measurements. The author claims that, geometrically, one could associate the proper time window with the mean orbital period, which can be estimated as the mean time between visiting a Poincare section or can be approximated by examining the oscillatory patterns in the data. In [19], on the other hand, the authors recommend choosing the time window as one-half of the critical window width.

To be able to take a position on whether the combinations of parameters that are promising according to the FFNN method are also optimal for the D_2 computation, we evaluated the error in the dimension estimates over the possible combinations. However, recall that even with a large data set, the detection of the plateau is not an easy task, and for a small amount of data, it is practically impossible [41, 42]. We used a computerised method for finding the linear region that includes use of the so called Theiler window and is

described in [43]. Then, we looked for the minimal difference between the correlation dimension estimate from the original three-dimensional state portrait and the estimates from the trajectories reconstructed from single coordinates.

2.7. Predictability Test. One of the most convincing arguments for a specific choice of embedding parameters is enabling the best possible modeling or forecasting of the data. Although the long-term prediction of a chaotic time series is not possible, for a short time ahead, novel nonlinear prediction methods are quite successful, especially when compared with predictability that is based on linear correlations in the data. The choice of a specific method depends on several aspects—how much data we have, what is the type and level of the noise, and so on. In this study, we have a long time series of clean artificial data, and we use a simple but effective prediction method for our testing purposes. The question is, will the parameters that are optimal for making predictions be consistent with those picked up by the FFNN method? To find out, the nonlinear method of one-point predictability of the time series is performed for various combinations of delays that are used for the reconstruction and dimensionality of the state space.

The idea of the forecasting method used here is to find historical data that is similar to the current situation and to assume that the system will react in the same way as in the past. This technique is generally known as the method of analogues [44]. To predict a follower of point X_n, the simplest version of the method of analogues finds its nearest neighbour X_i from the past states on the reconstructed trajectory and declares $X_{n+1} = X_{i+1}$. A modification that we have used in this study improves the simple version by utilising the direction in which the image of the found neighbour moves. Specifically, it means that we find the nearest neighbour X_i of the point X_n and declare $X_{n+1} = X_n + X_{i+1} - X_i$. In our examples, we made 620 (approximately two cycles) 1-point predictions. Each prediction was based on 93000 data points (approximately 300 preceding cycles).

To quantify the success of the prediction, the root mean squared error (RMSE) is often used. However, because the RMSE is scale-dependent, it is a good option to compare the forecasting of different methods for a specific observable. However, the RMSE is unable to compare how well one method forecasts different observables. To take the range of observed values into account, we used the normalized root mean squared error (NRMSE), namely, the RMSE divided by the standard deviation of the true time series during the prediction interval.

We evaluated the prediction error for different combinations of embedding parameters and compared the resulting maps with the maps that were obtained by the FFNN method.

3. Results

3.1. Cao's Algorithm. To demonstrate a traditional method of how the false nearest neighbours are used for the selection of the parameters of reconstruction, Cao's algorithm was tested for the variables of the Rössler system.
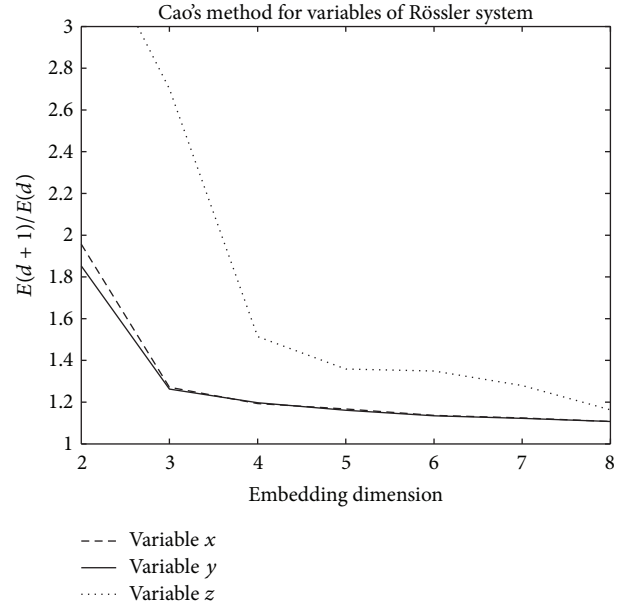

FIGURE 1: Cao's method [24] used for the variables of the Rössler system. The time delay is set to the value of 77. The figure shows the evolution of the average distance of the nearest neighbours for the variables x, y, and z.

In line with the arguments in Section 2.4, the time delay was set to the value of 77, which corresponds to the first minimum of the mutual information for all three variables. Then, the change in the average distance of the nearest neighbours was calculated for the embedding dimensions, up to 7. The results are presented in Figure 1.

To find the smallest sufficient embedding dimension, we must look for the dimension at which the average distance of the nearest neighbours stops growing. This criterion means that the monitored ratio $E(m + 1)/E(m)$ (the values on the vertical axis) falls below a certain threshold. For an unlimited amount of noise-free data, the ratio should settle to the value of 1. In practice, however, the threshold is chosen with respect to the noise level and other aspects of the data. However, as our test case demonstrates, the message of the resulting graph could be unclear even in a case in which there is a large amount of clean data.

Based on what we know about the Rössler system (see Section 2.4), we expect a result to tell us that in the case of y variable, a 3- or 4-dimensional embedding space will suffice, while for x and especially for z higher-dimensional space is necessary to fully unfold the attractor.

As Figure 1 shows, the results confirm the anticipated nonequivalence among the dynamical variables. However, the recommendations that regard what size of state space is sufficient for reconstruction differ substantially depending on the chosen threshold. For example, if the threshold is set to 1.4, then for the x and y variable the suggested embedding dimension equals 2, and the embedding dimension for the z induced state portrait is estimated as 4. However, if we consider a threshold of approximately 1.2, then the suggested

embedding dimension is 4 for x and y and approximately 7 for the z induced state portrait.

To summarise, despite the fact that we used a large amount of clean data, we were unable to unambiguously select the reconstruction parameters. Moreover, the choice of a fixed value of the delay can be a crucial problem if the hypothesis of importance of the embedding window is valid.

3.2. FFNN, D_2, Predictability. To select the best combinations for the delay and embedding dimension, we use the maps that are produced by the false first nearest neighbour method introduced in Section 2.3.

The FFNN algorithm was calculated for the Rössler and Lorenz system. The delay parameter was taken from the range 1 to 80, and the maximal dimension for which it was calculated was set to 7. The resulting maps show a color-coded dependence of the percentage of the false first nearest neighbours for combinations of the two parameters: the darker the colour, the fewer false neighbours remain after increasing the dimensionality of the space (Figures 2(a) and 3(a)). Looking at the map and recognising the dark bands reveal immediately that the time window is more relevant than either τ or m separately.

Our method is nonparametric as far as it is intended to provide the optimal size of the time window. If we are also interested in the minimal dimension for the reconstruction, then we must specify how many false neighbours we are prepared to tolerate. This threshold is the only parameter of the otherwise nonparametric method.

One of the main questions is whether the FFNN method leads to the same embedding parameters for different variables of the same system. Based on the maps, a time window of approximately 140 points appears to be optimal for the x-variable of the Rössler system, 156 points for the y-variable, and approximately 330 points for the z-variable (see Figure 2(a)). This finding means, for example, that if we decide to use the variable y for the reconstruction with embedding dimension 3, then a delay of 52 is an appropriate choice; in the case of an ED = 4, delay 39 should be taken and so on. The width of the time window is very close to the half of the cycle for the variables x and y and close to the length of one cycle for the z variable.

With respect to the inequivalence of the variables of the Rössler system, the observability coefficients η (see Table 1) sort the variables from the best to the worst, as follows: y, x, z. The results of FFNN confirmed that the z appears to be the less appropriate observable for the reconstruction, although there was not much visible difference between the variables x and y. However, the tests of the predictability and the correlation dimension show much more clearly, and in accordance with the values of observability indices, that some variables lead to significantly better results than others. In the case of the x variable of the Rössler system, the time window of approximately 140 points, as indicated by FFNN, appears to be optimal also for a one-point prediction. For D_2, the estimation time window of approximately 240 points should be used for the x-variable. Variable y (the variable with the highest observability degree) shows better predictability than

x, especially if a window of 125 or 205 points is used for reconstruction.

The estimation of the correlation dimension was best manageable by the y-variable. The huge black region in the corresponding map (image in the middle of Figure 2(c)) shows that there was a lot of successfully usable combinations of ED and τ. In contrast, the ability to estimate the correlation dimension from the z-variable was very limited. It was restricted to a few widths of the time window, such as 120 or 240. Even the slightest deviation from the applicable windows made the estimate impossible (right map of Figure 2(c)).

Next, let us make a quick comparison to the results for the Rössler system from Cao's method. Based on Figure 1 we can choose $\tau = 77$, ED = 8 for the reconstruction from the variable z or embedding parameters $\tau = 77$, ED = 4 for the variable y. In the former case, both the predictability and the accuracy of the D_2 estimate are very low (see Figure 2). The latter case (y variable, $\tau = 77$, ED = 4) is relatively applicable, although Figure 2 shows that, for an optimal prediction in 4-dimensional space, we should use a substantially lower time delay than 77.

For the Lorenz system, the observability coefficients sort the variables from the best to the worst, as follows: x, y, z. The same order was reflected by the tests that are presented in Figure 3. As discussed in Section 2.5, in 2-dimensional space, the trajectories appear to be best unfolded if a delay of approximately 37 points is used. This value indicates that there is an embedding window of approximately 75 points, which corresponds to a half of the average time (150 points) spent on a single wind. Our results show that TW = 75 is acceptable, although a one-point prediction was better for a slightly smaller window, while for an estimation of the correlation dimension from x or y, a slightly larger window is preferable. Estimation of the correlation dimension was most difficult for the z-variable, where the functional window had a width of approximately 190 points.

A comparison of the FFNN maps with the maps for predictability and D_2 shows that embedding parameters that are derived from the false nearest neighbours method are not necessarily the best choice for the specific data analysis. For example, let us look at the maps created for the x variable of the Rössler system. The darkest regions on the map for correlation dimension are not dark on the map for prediction. This is a reflection of the fact that estimates of D_2 from x variable were the most accurate for time window that was considerably wider than that optimal for the one point prediction of the x variable. However, this finding need not be true for different prediction methods. For example, Small and Tse claim that the best reconstruction is given by a relatively large embedding window and a constant lag of $\tau = 1$ [22]. However, they used an extremely simple local constant model to select the embedding window for which the model performed best.

4. Discussion and Conclusions

Much more attention should be given to the choice of parameters for the state space reconstruction because all of the tested nonlinear statistics suffer when the delay or
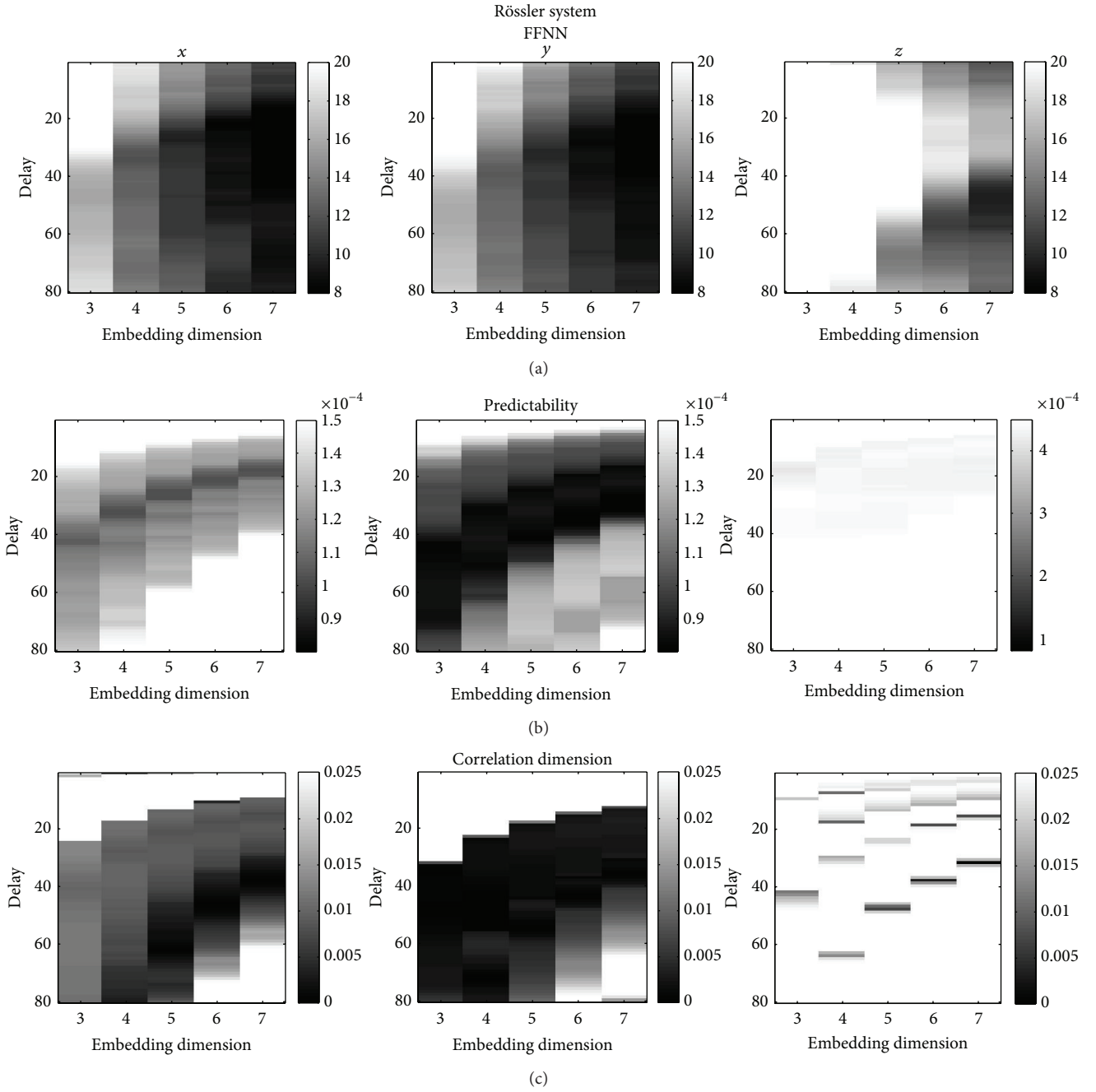
FIGURE 2: (a) Color-coded dependence of the percentage of the false first nearest neighbours for the variables x, y, and z of the Rössler system. The time delay is from 1 to 80, and the embedding dimension is from 3 to 7. (b) shows how the choice of the embedding parameters influences the one-step prediction error. (c) Precision of the estimates of the correlation dimension for combinations of embedding parameters.

the embedding dimension is chosen inappropriately. Until recently, the most common practice was to estimate the delay (usually as the minimum of the mutual information) in the first step. Then, the method of Kennel or Cao followed to estimate the minimal embedding dimension.

As we have shown, the above procedures do not enable us to use the idea of nearest neighbours to its full potential. The method that has been proposed in this paper was designed with the intention of getting the most out of

the false neighbours methods, to select the variable and the embedding window for a state space reconstruction.

To summarise, the outcomes of this study are the following.

(i) When creating a time-delay reconstruction, we cannot rely on a fixed value of the delay given by the mutual information or an autocorrelation function because this choice is justified only for 2-dimensional space, and even then, it could be far from optimal.
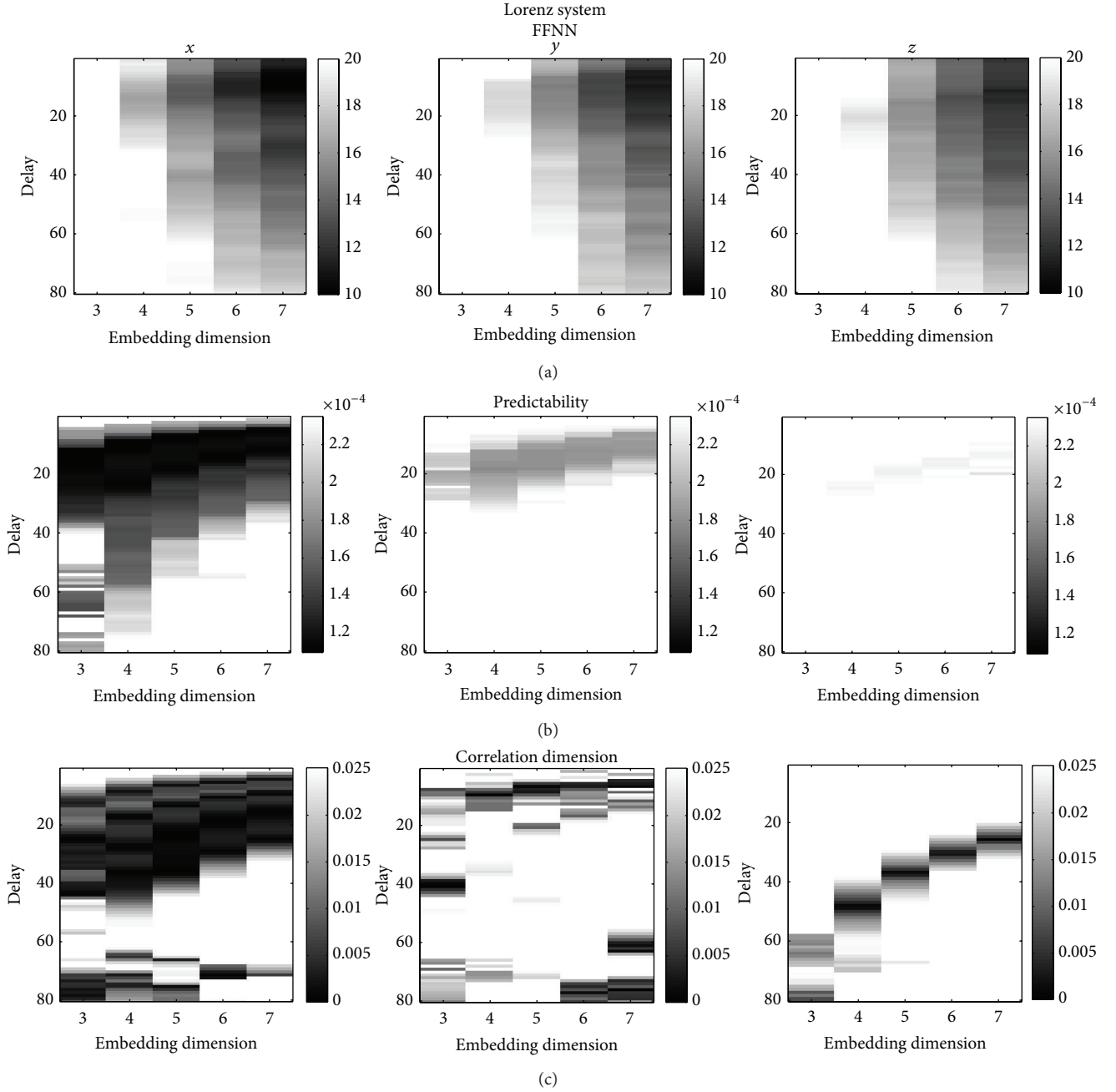
(a)

(b)

(c)

Figure 3: (a) Color-coded dependence of the percentage of the false first nearest neighbours for the variables x, y, and z of the Lorenz system. The time delay is from 1 to 80, and the embedding dimension is from 3 to 7. (b) shows how the choice of the embedding parameters influences the one-step prediction error. (c) Precision of the estimates of the correlation dimension for combinations of embedding parameters.

Moreover, the delay is difficult to find if the data are broadband and lacking any indication of periodicity. Instead, we use a map that is based on the false nearest neighbours idea to select the most appropriate combination of the delay and embedding dimension in one step. In this study, the earlier versions of the false neighbours methods, which use the Euclidean or other distances, were replaced by an almost nonparametric ranking based modification that we called false first nearest neighbour method (FFNN).

(ii) The resulting color-coded maps reveal the importance of the embedding window. However, as we demonstrated, the choice of the embedding window is problem dependent. The FFNN maps lead to parameters that ensure unfolding of data in a state space, but the same parameters do not seem to be necessarily optimal for purposes that follow the reconstruction. For example, the optimum window for the one-point prediction was typically substantially smaller than the optimum window for the computation of the correlation dimension.

(iii) Consequently, when looking for the correct embedding parameters, we should explore the parameter space while following some invariant, which is expected to indicate reaching the correct embedding window. The invariant should be chosen according to the purpose of the reconstruction. For example, when looking for the best embedding parameters for predictive modeling, one would minimise the prediction error over the possible combinations rather than trust the false nearest neighbour method without reservation. It appears to be worthwhile to prepare similar maps as we did or at least to try several combinations of embedding parameters to reveal the optimal embedding window for the specific task.

(iv) It turns out that in addition to the embedding window there is another crucial aspect of the reconstruction process, namely, the selection of the observable. Theoretically, the variable that is used for the reconstruction can be chosen arbitrarily. In practice, however, if several observables are available, then some of them or some combinations can be markedly better for reconstruction of the dynamics than others. This fact is not widely known. For testing purposes, we used two systems that have known equations and, hence, known symbolic observability degrees; thus, we had certain expectations regarding sorting the variables for better and worse. The results of the FFNN method, the D_2 estimates, and the predictability corresponded to the values of the symbolic observability degrees. The color-coded maps illustrated how dramatically the choice of variable can affect the ability to make predictions and estimate the complexity. In practical situations, when two or more scalar time series are recorded, it is definitely worthwhile to find the most appropriate observables for the reconstruction. The use of a combination of several observables can also be considered, but this alternative requires further study.

(v) Finally, recall that knowing the optimal TW does not mean that any larger than necessary ED (with an appropriately reduced delay) can be used equally well for the analysis of the underlying dynamics. With a limited amount of data, choosing too large of a space increases the redundancy and spoils the results. For example, the prediction benefits from having as many points as possible in the predicted neighbourhood, but the data become sparse in a higher embedding space. Consequently, using the optimal embedding window in spaces of lower dimensions leads to the best results.

Finally, let us recall that we have tested data under noise-free conditions. However, real time series are inevitably contaminated by noise, and it is likely to affect the value of the optimal embedding window even more than the number of data, sampling, type of application, or choice of observable. The noise amplification could cause a transition from a situation that can be investigated as approximately deterministic at least for short times to behaviour that appears to be random. Therefore, further studies are needed to investigate the robustness of the methods to noise.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, D. A. Rand and L. S. Young, Eds., pp. 366–381, Springer, Berlin, Germany, 1981.

[2] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.

[3] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.

[4] P. Xu, "Differential phase space reconstructed for chaotic time series," *Applied Mathematical Modelling*, vol. 33, no. 2, pp. 999–1013, 2009.

[5] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, 2nd edition, 2004.

[6] H. Whitney, "Differentiable manifolds," *The Annals of Mathematics*, vol. 37, no. 3, pp. 645–680, 1936.

[7] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *Journal of Statistical Physics*, vol. 65, no. 3-4, pp. 579–616, 1991.

[8] T. D. Sauer and J. A. Yorke, "Are the dimensions of a set and its image equal under typical smooth functions?" *Ergodic Theory and Dynamical Systems*, vol. 17, no. 4, pp. 941–956, 1997.

[9] D. J. Cross and R. Gilmore, "Differential embedding of the Lorenz attractor," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 81, no. 6, Article ID 066220, 2010.

[10] W.-T. Wu, "On the isotopy of C^r-manifolds of dimension n in euclidean (2n+ 1)-space," *Science Record (NS)*, vol. 2, pp. 271–275, 1958.

[11] D. J. Cross and R. Gilmore, "Representation theory for strange attractors," *Physical Review E*, vol. 80, no. 5, Article ID 056207, 2009.

[12] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, no. 6, pp. 3403–3411, 1992.

[13] C. Letellier, J. Maquet, L. Le Sceller, G. Gouesbet, and L. A. Aguirre, "On the non-equivalence of observables in phase-space reconstructions from recorded time series," *Journal of Physics A: Mathematical and General*, vol. 31, no. 39, pp. 7913–7927, 1998.

[14] L. A. Aguirre and C. Letellier, "Observability of multivariate differential embeddings," *Journal of Physics A: Mathematical and General*, vol. 38, no. 28, pp. 6311–6326, 2005.

[15] C. Letellier and L. A. Aguirre, "Investigating nonlinear dynamics from time series: the influence of symmetries and the choice of observables," *Chaos*, vol. 12, no. 3, pp. 549–558, 2002.

[16] L. A. Aguirre, S. B. Bastos, M. A. Alves, and C. Letellier, "Observability of nonlinear dynamics: normalized results and a time-series approach," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 18, no. 1, Article ID 013123, 2008.

[17] C. Letellier and L. A. Aguirre, "Symbolic observability coefficients for univariate and multivariate analysis," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 79, no. 6, Article ID 066210, 2009.

[18] D. S. Broomhead and G. P. King, "Extracting qualitative dynamics from experimental data," *Physica D: Nonlinear Phenomena*, vol. 20, no. 2-3, pp. 217–236, 1986.

[19] J. F. Gibson, J. D. Farmer, M. Casdagli, and S. Eubank, "An analytic approach to practical state space reconstruction," *Physica D. Nonlinear Phenomena*, vol. 57, no. 1-2, pp. 1–30, 1992.

[20] M. T. Rosenstein, J. J. Collins, and C. J. de Luca, "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica D: Nonlinear Phenomena*, vol. 73, no. 1-2, pp. 82–98, 1994.

[21] D. Kugiumtzis, "State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length," *Physica D: Nonlinear Phenomena*, vol. 95, no. 1, pp. 13–28, 1996.

[22] M. Small and C. K. Tse, "Optimal embedding parameters: a modelling paradigm," *Physica D: Nonlinear Phenomena*, vol. 194, no. 3-4, pp. 283–296, 2004.

[23] W. Liebert, K. Pawelzik, and H. G. Schuster, "Optimal embeddings of chaotic attractors from topological considerations," *Europhysics Letters*, vol. 14, no. 6, pp. 521–526, 1991.

[24] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1-2, pp. 43–50, 1997.

[25] R. Hegger and H. Kantz, "Improved false nearest neighbor method to detect determinism in time series data," *Physical Review E—Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 60, no. 4, pp. 4970–4973, 1999.

[26] M. B. Kennel and H. D. I. Abarbanel, "False neighbors and false strands: a reliable minimum embedding dimension algorithm," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 66, no. 2, Article ID 026209, 2002.

[27] C. J. Cellucci, A. M. Albano, and P. E. Rapp, "Comparative study of embedding methods," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 67, no. 6, 2003.

[28] J. Gao and Z. Zheng, "Direct dynamical test for deterministic chaos," *Europhysics Letters*, vol. 25, no. 7, pp. 485–490, 1994.

[29] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, "Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: theory," *Meccanica*, vol. 15, no. 1, pp. 9–20, 1980.

[30] C. Letellier, I. M. Moroz, and R. Gilmore, "Comparison of tests for embeddings," *Physical Review E*, vol. 78, no. 2, Article ID 026203, 2008.

[31] R. E. Kalman, "Mathematical description of linear dynamical systems," *Journal of the Society for Industrial & Applied Mathematics, Series A: Control*, vol. 1, no. 2, pp. 152–192, 1963.

[32] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is 'nearest neighbor' meaningful?" in *Database Theory—ICDT'99*, vol. 1540 of *Lecture Notes in Computer Science*, pp. 217–235, Springer, Berlin, Germany, 1999.

[33] M. E. Houle, H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "Can shared-neighbor distances defeat the curse of dimensionality?" in *Scientific and Statistical Database Management*, pp. 482–500, Springer, Berlin, Germany, 2010.

[34] O. E. Rössler, "An equation for continuous chaos," *Physics Letters A*, vol. 57, no. 5, pp. 397–398, 1976.

[35] L. A. Aguirre and C. Letellier, "Investigating observability properties from data in nonlinear dynamics," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 6, Article ID 066209, 2011.

[36] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[37] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena*, vol. 9, no. 1-2, pp. 189–208, 1983.

[38] Y.-C. Lai and D. Lerner, "Effective scaling regime for computing the correlation dimension from chaotic time series," *Physica D: Nonlinear Phenomena*, vol. 115, no. 1-2, pp. 1–18, 1998.

[39] J. M. Martinerie, A. M. Albano, A. I. Mees, and P. E. Rapp, "Mutual information, strange attractors, and the optimal estimation of dimension," *Physical Review A*, vol. 45, no. 10, pp. 7058–7064, 1992.

[40] H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamics, delay times, and embedding windows," *Physica D: Nonlinear Phenomena*, vol. 127, no. 1-2, pp. 48–60, 1999.

[41] J.-P. Eckmann and D. Ruelle, "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems," *Physica D: Nonlinear Phenomena*, vol. 56, no. 2-3, pp. 185–187, 1992.

[42] A. Krakovská, "Correlation dimension underestimation," *Acta Physica Slovaca*, vol. 45, no. 5, pp. 567–574, 1995.

[43] H. Budácová and S. Štolc, "Comparison of novel methods for correlation dimension estimation," in *Proceedings of the 9th International Conference on Measurement*, pp. 27–30, Smolenice, Slovakia, 2013.

[44] E. N. Lorenz, "Atmospheric predictability as revealed by naturally occurring analogues," *Journal of the Atmospheric Sciences*, vol. 26, no. 4, pp. 636–646, 1969.