# Embedding of multidimensional time-dependent observations

Jakobus P. Barnard,* Chris Aldrich, and Marius Gerber
*Department of Chemical Engineering, University of Stellenbosch, Private Bag X1, Matieland, Stellenbosch, 7602 South Africa*
(Received 7 February 2000; revised manuscript received 4 December 2000; published 19 September 2001)

A method is proposed to reconstruct dynamic attractors by embedding of multivariate observations of dynamic nonlinear processes. The Takens embedding theory is combined with independent component analysis to transform the embedding into a vector space of linearly independent vectors (phase variables). The method is successfully tested against prediction of the unembedded state vector in two case studies of simulated chaotic processes.

DOI: 10.1103/PhysRevE.64.046201 PACS number(s): 05.45.Tp

Parametrization of nonlinear dynamic systems by embedding of observations in phase space is central to the analysis of nonlinear time series, and the treatment of one-dimensional systems in this way is well established in nonlinear system identification [1,2]. However, parametrization of dynamic systems by embedding of multidimensional observations has not been sufficiently formalized, despite the practical importance of these types of systems. One cannot always predict the time evolution of a system state from a single observed variable [3]. For example, the Lorenz system [4] has three state variables $x$, $y$, and $z$, but $\dot{x}=f(x,y)$, while $\dot{z}=f(x,y,z)$; thus one cannot properly predict $z$ only from $x$ or even $(x,y)$ observations.

Cao, Mees, and Judd [3] have proposed the embedding of all components of the multidimensional observations by using an optimal Takens embedding for each component, obtained by minimizing the average prediction error of a nearest neighbor, locally constant predictor. Unfortunately they did not indicate how to optimize the embedding lag, which is crucial in the reconstruction of a representative attractor on realistic systems, especially if noise is present in the observations. Individual embedding of each observation into phase space could lead to significant statistical dependence between some of the phase variables, resulting in an attractor that is not optimally reconstructed from the observations.

In this article we therefore propose a method of embedding multidimensional observations that avoids both linear approximations in finding embedding dimensions and potentially suboptimal embedding lags. With this approach, each observation component is treated as a one-dimensional time series and embedded individually to generate a subspace. The concatenation of these subspaces into a combined phase space is consequently found to form a first approximation of the attractor in $\Re^{\Lambda}=\Re^{m_1}\cup\Re^{m_2}\cup\cdots\cup\Re^{m_M}$. Finally, the embedding variables are linearly separated to form proper phase variables. This results in a reconstructed dynamic attractor defined by a set of phase variables, based on the observation space.

More specifically, each component of the $M$-dimensional observation (multidimensional time series) $\mathbf{Y}=[\mathbf{Y}_1,\mathbf{Y}_2,\mathbf{Y}_3,\ldots,\mathbf{Y}_N]^T$, is embedded using Takens embedding as the matrix $\mathbf{X}$, where $\mathbf{X}_{ij}=[y_{ij},y_{i-k_j,j},\ldots,y_{i-m_j(k_j-1),j}]$, $i=b_j,\ldots,N$; $j=1,\ldots,M$, and $b_j=\max[k_j(m_j-1)]-k_j(m_j-1)+1$.

The embedding lag is determined by the average mutual information (AMI) algorithm of Frazer and Swinney [5] and the embedding dimension by the false nearest neighbor (FNN) algorithm [6]. Next, we find the optimal linear separation of non-Gaussian phase variables in $\mathbf{X}$ by optimal projection of $\mathbf{X}$ of the observations, as follows:

$$\mathbf{S}=\mathbf{W}\mathbf{X}, \tag{1}$$

where $\mathbf{S}$ is the optimal projection of the original embedding and $\mathbf{W}$ the transformation (separating) matrix. The dimension of $\mathbf{S}$ may be lower than that of $\mathbf{X}$. Thus one may achieve optimal projection, reduction of dimensionality, as well as linear independence of the phase variables.

Applying Hyvärinen's method [7] to find $\mathbf{W}$, one has to maximize the negentropy $J_G$ of $\mathbf{X}$, which is equivalent to minimizing the mutual information among components of $\mathbf{X}$, under the constraint of linear decorrelation of the components, i.e., maximize

$$\sum_{i=1}^{M} J_G(\mathbf{w}_i) \quad \text{with respect to } \mathbf{w}_i \tag{2}$$

under the constraint,

$$E\{\mathbf{w}_k^T\mathbf{w}_j^T\}=\delta_{jk}, \tag{3}$$

where

$$J_G(\mathbf{w})=[E\{G(\mathbf{w}^T\mathbf{x})\}-E\{G(\nu)\}]^2 \tag{4}$$

with $G(\cdot)$ some sufficiently smooth, even, so-called contrast function that estimates the probability density function of an independent component and $\nu$ a standardized Gaussian variable. Each vector $\mathbf{w}_i$ is a row of matrix $\mathbf{W}$.

We compared models based on embedded observations with similar models when the observations are not embedded. Two case studies were considered, namely, a chaotic autocatalytic process and the familiar Lorenz system. When the observations were not embedded, a subset of the ob-

*Email address: jsteyl@ing.sun.ac.za

 

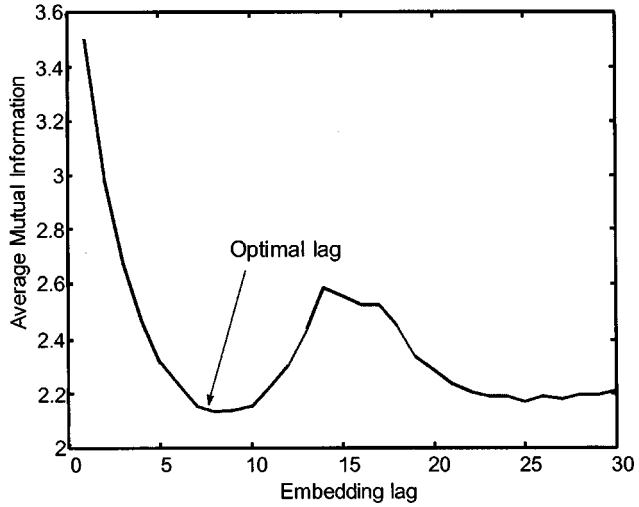


FIG. 1. Calculation of optimal embedding lag, shown here for $x$ component of autocatalytic process. Heuristically, optimal lag is set at first minimum.

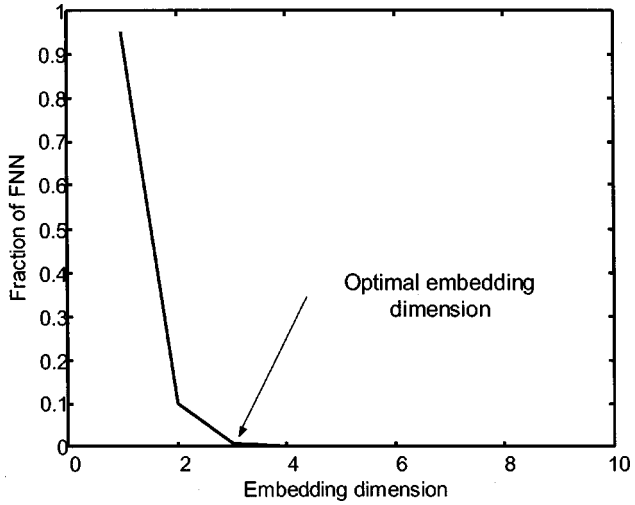


FIG. 2. Calculation of optimal embedding dimension, shown here for the $x$ component of the autocatalytic process.

served states was used directly to predict another subset of the observed states from the same system. In both test cases it is apparent from inspection of the state equations of the process [Eqs. (5) and (6) below] that the $Z$ state is dependent on all three state variables $(X,Y,Z)$, so that successful modeling of $Z$ from $X$ and $Y$ would constitute a proper test of the proposed embedding strategy.

The autocatalytic process previously considered by Lynch [8] can be expressed as a state space system by the following set of differential equations:

$$\frac{dX}{dt} = 1 - X - aXZ^2,$$

$$\frac{dY}{dt} = 1 - Y - bYZ^2,$$

$$\frac{dZ}{dt} = 1 - (1+c)Z + daXZ^2 + ebYZ^2, \tag{5}$$

where $X$, $Y$, and $Z$ are dimensionless concentrations and $a$, $b$, $c$, $d$, and $e$, dimensionless input concentrations. The process is chaotic with a well-defined attractor for specific ranges of two process control parameters, $d$ and $e$. For the settings $a = 18\,000$, $b = 400$, $c = 80$, $d = 1.5$, $e = 4.2$, and initial conditions $[0,0,0]^T$, the authors solved the set of equations using a fifth order Runge-Kutta numerical method over 100 simulated seconds. This gave approximately 10 000 points which were resampled with a constant sampling period of 0.01 s. The evolution of the three states $X,Y,Z$ was observed over the whole period of the simulation to form a three-dimensional set of observations (time series) of the system.

The first step of the proposed multidimensional embedding strategy was to embed the individual observation components. For each observation component, the lag $k = 7$ from AMI calculations (Fig. 1), and the embedding dimension $m = 3$ (Fig. 2) from FNN calculations (9-embedding space was obtained by columnwise concatenation of the $\Re^3$ subspaces that resulted from the individual embedding of each observation component.

Hyvärinen's algorithm [7] with a contrast function of the form $G(s) = (1/\alpha)\ln\cosh(\alpha s)$, with $\alpha$ some trivial constant, was used to project the embedding space to a proper state space. For this purpose, we used the MATLAB compatible software FASTICA [9]. The optimal contrast function was selected after qualitatively investigating the distribution of the observation components by plotting histograms of the components superimposed on a Gaussian distribution (Fig. 3). Guidelines for the selection of a contrast function according to the distribution of the data are available in the FASTICA software help files, as well as in [7].

A multiple-layer perceptron neural network with a single hidden layer consisting of 32 hyperbolic tangent nodes and a single-node linear output layer was fitted to the state space to predict $Z$ from the selected observations. For this purpose, we used the MATLAB Neural Network Toolbox V2.0. The network was trained with the Levenberg-Marquardt algorithm on the first 7000 observations, tested on observations 7001 to 8000, and validated against observations 8001 to

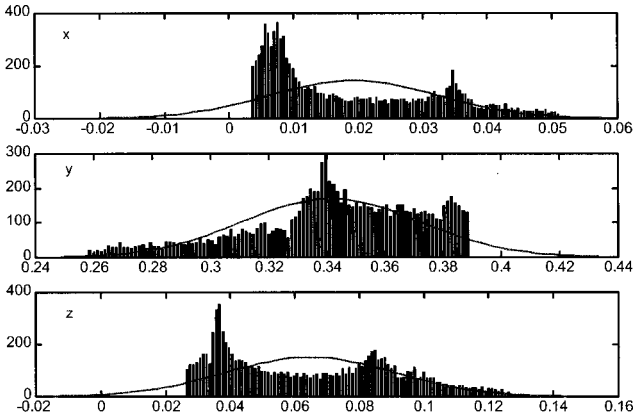


FIG. 3. Choosing the contrast function: Histograms of autocatalytic components superimposed on Gaussian distribution.

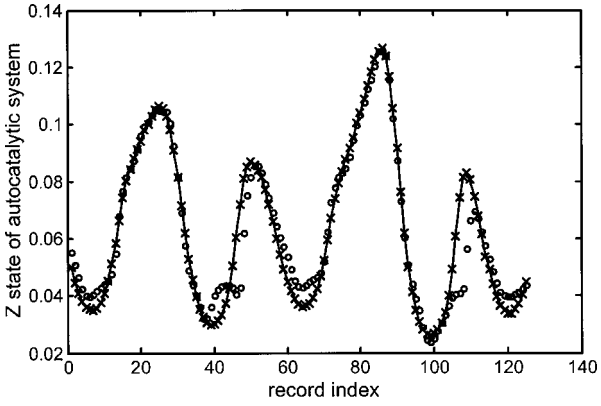


FIG. 4. Autocatalytic $Z$ state predicted from a multichannel embedding of $X$ and $Y$ states ($\times$ marker), as well as predicted directly from $X$ and $Y$ states ($\bigcirc$ marker) with neural network. Observed $Z$ state is represented by the line.

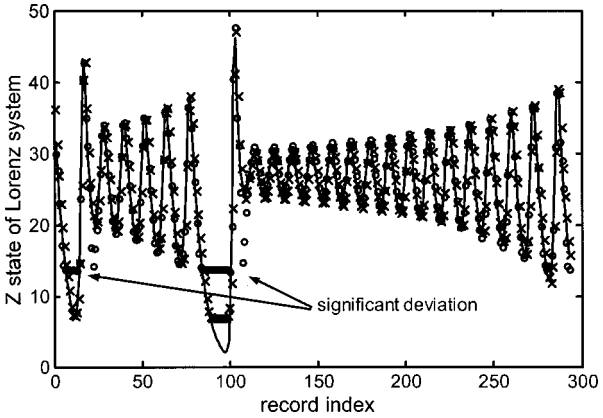


FIG. 5. Lorenz $Z$ state predicted from a multidimensional embedding of $X$ and $Y$ states ($\times$ marker), as well as predicted directly from $X$ and $Y$ states ($\bigcirc$ marker) with neural network. Observed $Z$ state is represented by the line.

9000. The number of hidden nodes was determined by comparing the model error on the test set against increasing network size. A minimum sum squared error was reached for 32 hidden nodes. For comparison of embedding strategies, the neural network model was also fitted directly to the same observations as were used for embedding, to predict the $Z$ state.

The same approach was followed with the Lorenz system, which is also a chaotic system and is described by the following set of differential equations:

$$\dot{x} = \sigma(y - x),$$

$$\dot{y} = \rho - y - xz,$$

$$\dot{z} = -\beta z + xy. \tag{6}$$

These equations were solved for the standard values of $\sigma = 10$, $\rho = 28$, and $\beta = \frac{8}{3}$ by using fifth order Runge-Kutta numerical integration over 50 simulated seconds. All three states were resampled at a constant 0.05 s sampling period.

As in the previous case, the first step was to embed the observed $X$ and $Y$ states. For each observed state, the lag $k = 3$ was found from AMI calculations, while the embedding dimension $m = 3$ was found from FNN calculations, resulting in an $\Re^6$ state space. The same multilayer perceptron neural network described above was fitted to the state space to predict the observed $Z$ state. For comparison of embedding strategies, the same model was fitted directly to the same $X$ and $Y$ observations used for embedding to predict the $Z$ state. Optimal projection of the initial embedding was based on the use of a Gaussian contrast function of the form $G(s) = -(1/\alpha)\exp(-\alpha s^2/2)$.

The model based on embedding gave significantly better prediction of the $Z$ state than the model fitted directly to the $X$ and $Y$ observations. For one-step ahead prediction of the autocatalytic system, the multilayer perceptron could predict the data with an accuracy characterized by a multiple coefficient of determination $R^2 = 0.995$ when the proposed embedding method was used, as opposed to an $R^2$ value of 0.928 when embedding was not used. Likewise, embedding of the Lorenz system enabled one-step ahead prediction with the multilayer perceptron characterized by an $R^2$ value of 0.991, as opposed to $R^2 = 0.880$ when embedding was not used. These results are shown in Figs. 4 and 5. Note, for example, in Fig. 5 that in the case of the Lorenz system the prediction of $Z$ in the region of crossover between the wings of the Lorenz attractor was markedly better using the multidimensional embedding method. This indicates that the proposed embedding technique can be particularly useful for inferring system states where operational measurement of these states is not feasible.

In conclusion we shall discuss the general applicability of our proposed system parametrization method, based on the embedding of multidimensional observations. The proposed method clearly works well with systems defined by ordinary differential equations, for which we can generate noiseless data. A valid question concerns the generalization of the parametrization method to cases involving noisy observations. A particular difficulty can be expected with the calculation of the embedding lag by the average mutual information statistic, which is susceptible to noise.

In this case the problem can be surmounted by embedding each observation component individually, with a default lag of $k = 1$ in a phase space of arbitrary large dimension. After concatenation of the subspaces, the combined phase space is reduced by projection onto the significant eigenvectors of the phase space covariance matrix [10]. Selection of the number of eigenvectors onto which to project is based on the variance collectively captured by the set of eigenvectors (at least 95% is recommended). Thereafter, optimal linear separation is obtained, as in this article, by applying Hyvärinen's method to the reduced phase space.

[1] H. D. Abarbanel, I. Z. Gills, C. Liu, and R. Roy, Phys. Rev. A **53**, 440 (1996).
[2] M. Small and K. Judd, Physica D **117**, 283 (1998).
[3] L. Cao, A. Mees, and K. Judd, Physica D **121**, 75 (1998).
[4] E. N. Lorenz, J. Atmos. Sci. **20**, 131 (1963).
[5] A. M. Frazer and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).
[6] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).
[7] A. Hyvärinen, IEEE Trans. Neural Netw. **10**, 626 (1999).
[8] D. T. Lynch, Chem. Eng. Sci. **47**, 347 (1991).
[9] Available on the Helsinki University of Technology website (http://www.cis.hut.fi/projects/ica/fastical).
[10] R. Vautard, P. Yiou, and M. Ghil, Physica D **58**, 95 (1992).