Parsimonious description for predicting high-dimensional dynamics

Yoshito Hirata1,2,3, Tomoya Takeuchi2, Shunsuke Horai1, Hideyuki Suzuki2,3 & Kazuyuki Aihara1,2,3

When we observe a system, we often cannot observe all its variables and may have some of its limited measurements. Under such a circumstance, delay coordinates, vectors made of successive measurements, are useful to reconstruct the states of the whole system. Although the method of delay coordinates is theoretically supported for high-dimensional dynamical systems, practically there is a limitation because the calculation for higher-dimensional delay coordinates becomes more expensive. Here, we propose a parsimonious description of virtually infinite-dimensional delay coordinates by evaluating their distances with exponentially decaying weights. This description enables us to predict the future values of the measurements faster because we can reuse the calculated distances, and more accurately because the description naturally reduces the bias of the classical delay coordinates toward the stable directions. We demonstrate the proposed method with toy models of the atmosphere and real datasets related to renewable energy.

Nonlinear time series analysis1–3, or time series analysis based on dynamical systems theory4–6, has been developed intensively in the last 35 years. The most important result is the method of delay coordinates7–9: Suppose that we can observe a scalar time series $s(t)$ for $t = 1, 2, \ldots$ from a target system whose dimension is $m$. If we construct $d$-dimensional vectors $\{(s(t), s(t-1), \ldots, s(t-d+1))\}$, called delay coordinates, by using successive scalar measurements, then it is generally true that if $d > 2m$, the states $x(t-d+1)$ for the underlying dynamical system and the vectors with delay coordinates $\{(s(t), s(t-1), \ldots, s(t-d+1))\}$ are one-to-one on the attractor, or a set of states the trajectory of the underlying dynamical system is attracted after the transient. Although this statement is supported how large the dimension $m$ for the underlying dynamics is, practically the method of delay coordinates cannot be used for high-dimensional dynamics partly because the delay coordinates are distorted toward the stable directions10, and partly because the calculation of delay coordinates becomes expensive.

The core part of the proposed method came from the idea of weighted delay coordinates4. In the weighted delay coordinates, the distortion of high-dimensional delay coordinates toward the stable directions is avoided by reducing weights of the past observations exponentially. If we denote the decay rate by $\lambda$ ($0 < \lambda \leq 1$), the weighted delay coordinates can be written as $(s(t), \lambda s(t-1), \ldots, \lambda^{d-1}s(t-d+1))$. Berry et al.10 uses the weighted delay coordinates for dimension reduction.

Our idea is to virtually use infinite-dimensional weighted delay coordinates, namely $\tilde{s}(t) = (s(t), \lambda s(t-1), \lambda^2 s(t-2), \ldots)$, where we extend the time axis toward the minus infinity and assume $s(t) = 0$ for $t \leq 0$ for simplicity. As we will show in the Methods section, the distance between $\tilde{s}(t_1)$ and $\tilde{s}(t_2)$ in $L_1$ norm is calculated efficiently and successively. To make the distance converge, we

1Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan.
2Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan.
3CREST, JST, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. Correspondence and requests for materials should be addressed to Y.H. (email: yoshito@sat.t.u-tokyo.ac.jp)
just need to set \( 0 < \lambda < 1 \). Thus, by combining with the Lorenz’s method of analogues\(^{11,2}\), we construct a method of time series prediction (see the details in the Methods section).

First, we tested the proposed time series prediction on Lorenz’96 I model\(^{12,13}\), which is the minimum model of the atmosphere (see Supplementary Information for the details of the numerical experiment). When we compared the performance of the proposed method with that of the conventional 10-dimensional delay coordinates, we found that the proposed method tended to achieve the higher correlation coefficient (see Supplementary Information for the definition) between the prediction and the corresponding actual values up to 5 steps ahead (Fig. 1(a,c)). The proposed method was significantly better than the persistent prediction, where we let the current values be the prediction for the future values. For one of the 100 tested time series, we compared the speed of calculations. The proposed method only consumed 0.72 seconds, while the conventional 10-dimensional delay coordinates consumed 14.66 seconds. We used a laptop computer with Intel Core i7 CPU(3GHz) with 16.0GB memory. The programs were implemented in MATLAB. We did not use the recursive formula of equation (4) in Methods here. Thus, the proposed method is faster and more accurate in prediction than the conventional delay coordinates.

Second, we tested the proposed time series prediction on Lorenz’96 II model\(^{12,13}\). In the Lorenz’96 II model, there are two types of variables: the slow variables correspond to the upper-layer of the atmosphere and the fast variables correspond to the layer close to the surface of the earth (see Supplementary Information for the details of this numerical experiment). We found that the proposed method tended to have the greater correlation coefficient than the conventional 10-dimensional delay coordinates up to 10 steps ahead (Fig. 1(b,d)). The proposed method was significantly better than the persistent prediction, where we let the current values be the prediction for the future values. For one of the 100 tested time series, the computational time required for the proposed method was 0.80 seconds with the laptop computer, while that required for the conventional delay coordinates was 20.16 seconds; we did not use the recursive formula of equation (4) here too.

Third, we applied the proposed time series prediction to the sunshine duration at a single point of Fuchu, Japan (see Supplementary Information for the details of the numerical experiment). The result presented in Fig. 2 shows that the proposed method achieved the higher correlation coefficient than
the persistent prediction and the prediction using 1 day periodicity when the prediction steps were greater than or equal to 1.2 hours. To make the prediction for the 10 minutes dataset spanning 5 years of 2008–2012, it only took 3.8 hours with a desktop computer with 2.7 GHz 12-Core Intel Xenon E5 with 64 GB memory. Therefore, the proposed method can run in the real time.

We also applied the proposed time series prediction to the wind speed data at Fuchu, Japan (See the Supplementary Information for the details of this numerical experiment). We found that the proposed method achieved the greater correlation coefficient than the persistent prediction and the prediction using 1 day periodicity when the prediction steps were between 4 and 15.8 hours (Fig. S1). In addition, we needed 3.8 hours with the desktop computer to complete the prediction for the dataset between 2008 and 2012. Thus, for this wind speed dataset, the prediction can also be done in the real time.

By using the proposed method, we can circumvent problems of the current standard practice for obtaining delay coordinates, which is, for example, to decide the delay by the first minimum of mutual information and the embedding dimension by false nearest neighbors. Instead of choosing these two parameters, we need to choose the decay rate $\lambda$, by which the proposed method shows the robust performance as demonstrated in Fig. 3 and S2. Thus, the proposed method will make it easier to apply nonlinear time series prediction. We can even remove the choice of $\lambda$ by combing the proposed method with the expert advice algorithm (see Supplementary Information for the details). Thus, the proposed method is suitable for automating some time series prediction tasks.
The proposed method is robust against the observational noise. Even if we increase the noise level up to 10% of the standard deviation of the original time series, the correlation coefficient between the prediction using the proposed method and the actual values was significantly higher than that between the prediction using the conventional delay coordinates and the actual values (See Fig. 4). The proposed method naturally filters out observational noise when it is included in the measurements. Moreover, even if we evaluate the prediction with the root mean square errors, the proposed method is superior to the conventional delay coordinates for short-term predictions (Fig. 1(e,f)). Thus, our results could have the robustness to some extent in terms of ways for evaluating the prediction.

If we increased the size of database, then the prediction performance became better (Fig. S3). In addition, even if we used different numbers of neighbors for making the prediction, the performance was robust and did not change much (Fig. S4).

The lower the minimal Lyapunov exponent \( \sigma_1 \) for the underlying dynamics is, the higher the optimal \( \lambda \) is (Fig. S5). But, judging from the values of the optimal \( \lambda \) for the prediction, the reconstructed space was not reduced to the most stable direction because \( -\log \lambda > -\sigma_1 \). Therefore, the list of distances contained the information of more than the one-dimensional space. From this viewpoint, the proposed method provides a convenient description for the high-dimensional dynamics.

It is easy to further extend the proposed infinite-dimensional weighted delay coordinates to multivariate time series or point processes. Let \( W \) be a set of states. If we define the state at time \( t \) by \( w_t \in W \) and a distance function on these states by \( D: W \times W \rightarrow [0] \cup R^+ \), then the distance \( \tilde{D} \) between \( \tilde{w}_{t_1} = \{ w_t | t \leq t_1 \} \) and \( \tilde{w}_{t_2} = \{ w_t | t \leq t_2 \} \) on such infinite dimensional weighted delay coordinates can be defined as

\[
\tilde{D}(\tilde{w}_{t_1}, \tilde{w}_{t_2}) = \sum_{d=0}^{\infty} \lambda^d D(w_{t_1-d}, w_{t_2-d}).
\]

This distance, called the Fréchet product metric, satisfies the three conditions for the metric: (i) \( \tilde{D}(\tilde{w}_{t_1}, \tilde{w}_{t_2}) \geq 0 \) and \( \tilde{D}(\tilde{w}_{t_1}, \tilde{w}_{t_2}) = 0 \) if and only if \( \tilde{w}_{t_1} = \tilde{w}_{t_2} \); (ii) \( \tilde{D}(\tilde{w}_{t_1}, \tilde{w}_{t_2}) = \tilde{D}(\tilde{w}_{t_2}, \tilde{w}_{t_1}) \); (iii) \( \tilde{D}(\tilde{w}_{t_1}, \tilde{w}_{t_2}) \leq \tilde{D}(\tilde{w}_{t_1}, \tilde{w}_{t_1}) + \tilde{D}(\tilde{w}_{t_2}, \tilde{w}_{t_2}) \). Therefore, if a given time series is multidimensional, we may choose the Euclidean distance between times \( t_1 - d \) and \( t_2 - d \) as \( D \) to obtain \( \tilde{D} \). We will discuss this extension in our future communication.

The proposed description might also be useful in inferring a network structure. This is an open question and we are also interested in developing the method in this direction.

Comparing with the traditional delay coordinates, the proposed infinite-dimensional weighted delay coordinates can produce more accurate time series prediction faster. As there is an increasing demand for real-time prediction for a big dataset especially in the field of renewable energy such as photovoltaic and wind powers, we hope that the proposed method helps to introduce more renewable energy into the power grids so that we can reduce CO₂ emissions.

**Methods**

Suppose that a scalar time series \( \{ s(t) | t = 1, 2, \ldots \} \) is given successively. Denote by \( N \) the size of the database. Let \( \tilde{s}(t) = (s(t), \lambda s(t-1), \lambda^2 s(t-2), \ldots) \) be the infinite-dimensional weighted delay coordinates, where \( \lambda \) is a decay rate. We define a distance \( ||\tilde{s}(t_1) - \tilde{s}(t_2)||_1 \) between \( \tilde{s}(t_1) \) and \( \tilde{s}(t_2) \) as
\[ \| \mathbf{s}(t_1) - \mathbf{s}(t_2) \|_i = \sum_{d=0}^{\infty} \lambda^d |s(t_1 - d) - s(t_2 - d)|. \]  

(2)

We use $\lambda = 0.5$ if not mentioned.

This distance is convenient because we can reuse the previous calculations of distances to obtain distances for a pair of its future infinite-dimensional weighted delay coordinates. Namely, observe the following relation:

\[ \| \mathbf{s}(t_1) - \mathbf{s}(t_2) \|_i = |s(t_1) - s(t_2)| + \lambda \sum_{d=0}^{\infty} \lambda^d |s(t_1 - 1 - d) - s(t_2 - 1 - d)| \]

\[ = |s(t_1) - s(t_2)| + \lambda \sum_{d=0}^{\infty} \lambda^d |s(t_1 - 1) - s(t_2 - 1)|. \]

(3)

When we use the conventional $d$-dimensional delay coordinates ($\lambda = 1$), the similar recursive logic may be applied to simplify the calculation of distances as follows:

\[ \| \mathbf{s}(t_1) - \mathbf{s}(t_2) \|_i = |s(t_1) - s(t_2)| + \lambda \sum_{d=0}^{\infty} \lambda^d |s(t_1 - 1 - d) - s(t_2 - 1 - d)| + \lambda \| \mathbf{s}(t_1 - 1) - \mathbf{s}(t_2 - 1) \|_i. \]

(4)

We combine the relation of equation (3) with Lorenz’s method of analogues\(^{2,11}\). In the Lorenz’s method of analogues, we find close matches in the past and average their following points as prediction for the future. Let us construct prediction of up to $P$ steps ahead. Suppose that the current time is at time $t$ and we have the following datasets: a list of distances $\Delta_i = |s(t - i) - s(t - i - 1)|$ ($i = 1, \ldots, N$) and a list of the corresponding observed values $S^t_i = s(t - i)$ ($i = 1, \ldots, N$). First, we find the set $\{(i_1, \ldots, i_K) \mid \Delta_{i_1} \leq \ldots \leq \Delta_{i_K}\}$ of indices for the $K$ smallest distances among $\Delta_i$ ($1 \leq i \leq N$), and provide the $p$ steps ahead prediction for each $p$ ($0 < p \leq P$) by $\frac{1}{K} \sum_{i \in T(t,p)} S^t_{i-p+1}$. We used $K = 10$ if not mentioned.

Second, we update the list of distances by

\[ \Delta_{i+1}^t = \| \mathbf{s}(t) - \mathbf{s}(t - i) \|_i \]

\[ = \sum_{d=0}^{\infty} \lambda^d |s(t - d) - s(t - i - d)| \]

\[ = |s(t) - s(t - i)| + \lambda \sum_{d=0}^{\infty} \lambda^d |s(t - 1 - d) - s(t - i - 1 - d)| \]

\[ = |s(t) - s(t - i)| + \lambda \| \mathbf{s}(t - 1) - \mathbf{s}(t - i - 1) \|_i \]

\[ = |s(t) - s(t - i)| + \lambda \Delta_i^t. \]

(5)

We also update the list of the corresponding values by

\[ S_{i+1}^{t+1} = S_i^t \]

for $1 \leq i < N$ and

\[ S_{i+1}^t = s(t). \]

(6)

(7)

Therefore, to run the proposed method, each time we obtain a measurement, we only need to store $2N$ floating numbers, compare $N$ floating numbers, conduct $N$ additions, $N$ subtractions, and $N$ multiplications. Compare these numbers with the case using the traditional $d$-dimensional delay coordinates: If we do not use the recursive formula of equation (4), we need $dN$ comparisons for comparing element-wisely $N$ pairs of $d$ dimensional delay coordinates, $dN$ subtractions, and $(d - 1)N$ additions (The computation for finding nearest neighbors is in the same order).

References

1. Abarbanel, H. D. I. Analysis of Observed Chaotic Data (Springer-Verlag, 1996).
2. Kantz, H. & Schreiber T. Nonlinear Time Series Analysis (Cambridge University Press, 2003).
3. Small, M. Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance (World Scientific, 2005).
4. Alligood, K. T., Sauer, T. D. & Yorke, J. A. Chaos: An Introduction to Dynamical Systems (Springer, 1996).
5. Katok, A. & Hasselblatt, B. Introduction to the Modern Theory of Dynamical Systems (Cambridge University Press, 1996).
6. Strogatz, S. H. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering (Westview Press, 2014).
7. Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. Geometry from a time series. Phys. Rev. Lett. 45, 712–715 (1980).
8. Takens, F. Detecting strange attractors in turbulence. Lect. Notes Math. 898, 366–381 (1981).
9. Sauer, T., Yorke, J. A. & Casdagli, M. Embeddology. J. Stat. Phys. 65, 579–616 (1991).
et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

et al.: Hirata, Y. Parsimonious description for predicting high-dimensional

Acknowledgements

Author Contributions

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Hirata, Y. et al. Parsimonious description for predicting high-dimensional
dynamics. Sci. Rep. 5, 15736; doi: 10.1038/srep15736 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The
images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the
Creative Commons license, users will need to obtain permission from the license holder to reproduce
the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/