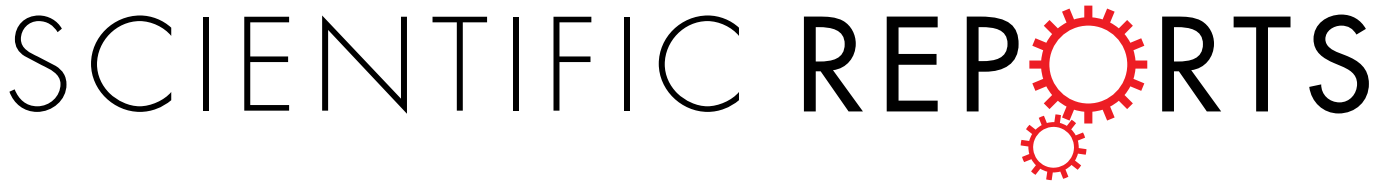

OPEN

Parsimonious description for predicting high-dimensional dynamics

Received: 22 May 2015
Accepted: 29 September 2015
Published: 29 October 2015

Yoshito Hirata[1,2,3], Tomoya Takeuchi[1], Shunsuke Horai[1], Hideyuki Suzuki[2,3] & Kazuyuki Aihara[1,2,3]

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 analysis[1–3], or time series analysis based on dynamical systems theory[4–6], has been developed intensively in the last 35 years. The most important result is the method of delay coordinates[7–9]: Suppose that we can observe a scalar time series $\{s(t)|t = 1, 2, \ldots\}$ 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 directions on the attractor[10], and partly because the calculation of delay coordinates becomes expensive.

The core part of the proposed method came from the idea of weighted delay coordinates[4]. 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 $\vec{s}(t) = (s(t), \lambda s(t-1), \lambda^2 s(t-2), \ldots)$, 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 $\vec{s}(t_1)$ and $\vec{s}(t_2)$ in L_1 norm is calculated efficiently and successively. To make the distance converge, we

[1]Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan. [2]Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan. [3]CREST, 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)

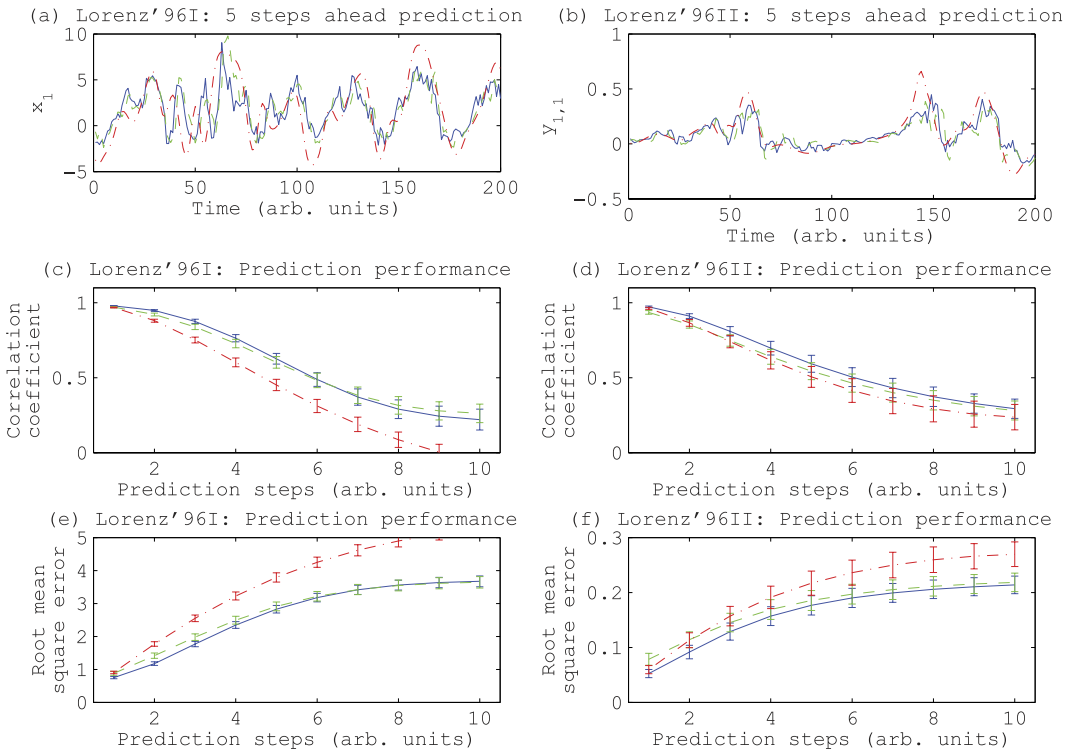

Figure 1. Prediction results on Lorenz'96 I model and Lorenz'96 II model. (**a**,**c**,**e**) correspond to the Lorenz'96 I model, while (**b**,**d**,**f**) correspond to the Lorenz'96 II model. (**a**) and (**b**) show examples of 5 steps ahead prediction, (**c**,**d**) show the correlation coefficients between the prediction and the actual value given a prediction step, and (**e**,**f**) show the root mean square errors given a prediction step. In panels (**a**,**b**), the blue solid line corresponds to the prediction by the proposed infinite-dimensional weighted delay coordinates, the green dashed line corresponds to the one by the 10-dimensional delay coordinates, and the red dash-dotted line, the actual values. In panels (**c**–**f**), the blue solid line and the green dashed line correspond to the proposed method and the conventional 10-dimensional delay coordinates, respectively, and the red dash-dotted line, the persistence prediction. The error bars show the standard deviations.

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)). In 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

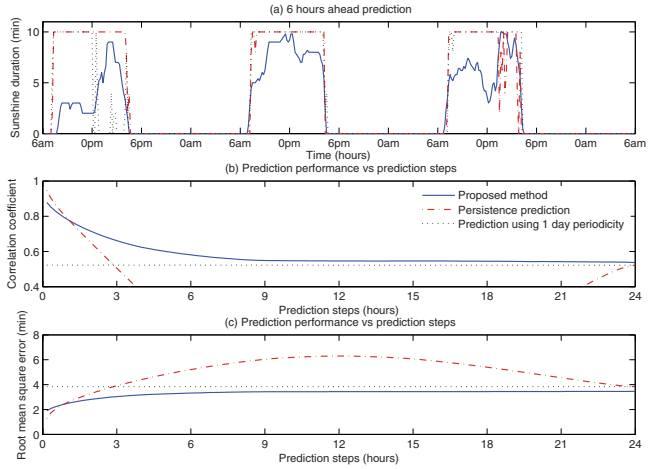

Figure 2. The prediction result on sunshine duration within 10 minutes at Fuchu, Japan. Panel (**a**) shows 6 hours ahead prediction (blue solid line), prediction using 1 day periodicity (black dotted line), and the actual values (red dash-dotted line). Panel (**b**) shows the correlation coefficients between the prediction and the actual observations. Panel (**c**) shows the root mean square errors between the prediction and actual observation. In panels (**b**,**c**), the blue solid line, the red dash-dotted line, and the black dotted line correspond to the proposed method, the persistence prediction, and the prediction using 1 day periodicity.

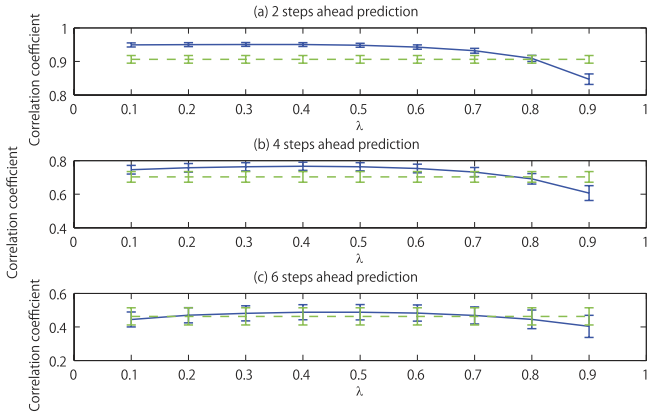

Figure 3. Dependence of the prediction performance on the parameter λ of the proposed method for Lorenz'96 I model. Panels (**a**–**c**) correspond to 2, 4, and 6 steps ahead predictions. In each panel, the blue solid line corresponds to the proposed method and the green dashed line corresponds to the conventional 10-dimensional delay coordinates, which do not depend on the choice of the parameter λ. Each error bar shows the standard deviation of the correlation coefficient between the prediction and the actual values.

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[2], which is, for example, to decide the delay by the first minimum of mutual information[14] and the embedding dimension by false nearest neighbors[15]. Instead of choosing these two parameters, we need to choose the decay rate λ, 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 λ by combing the proposed method with the expert advice algorithm[16–18] (see Supplementary Information for the details). Thus, the proposed method is suitable for automating some time series prediction tasks.

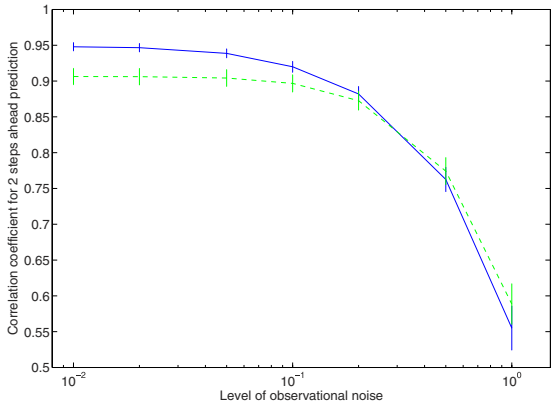

Figure 4. Dependence of the prediction performance on the level of the observational noise for Lorenz'96 I model. The blue solid line and the green dashed line correspond to the proposed method and the conventional 10 dimensional delay coordinates, respectively. The error bars show the standard deviations for the correlation coefficients between the prediction and the actual values.

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 σ_1 for the underlying dynamics is, the higher the optimal λ is (Fig. S5). But, judging from the values of the optimal λ for the prediction, the reconstructed space was not reduced to the most stable direction because $-\log_2\lambda > -\sigma_1$ [10]. 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[19–24] or point processes[25–29]. 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 $\mathrm{D}: \mathrm{W} \times \mathrm{W} \to \{0\} \cup R^+$, then the distance $\widetilde{D}$ between $\widetilde{w}_{t_1} = \{w_t | t \le t_1\}$ and $\widetilde{w}_{t_2} = \{w_t | t \le t_2\}$ on such infinite dimensional weighted delay coordinates can be defined as

$$\widetilde{D}\left(\widetilde{w}_{t_1}, \widetilde{w}_{t_2}\right) = \sum_{d=0}^{\infty} \lambda^d D\left(w_{t_1-d}, w_{t_2-d}\right). \tag{1}$$

This distance, called the Fréchet product metric[30], satisfies the three conditions for the metric: (i) $\widetilde{D}(\widetilde{w}_{t_1}, \widetilde{w}_{t_2}) \ge 0$ and $\widetilde{D}(\widetilde{w}_{t_1}, \widetilde{w}_{t_2}) = 0$ if and only if $\widetilde{w}_{t_1} = \widetilde{w}_{t_2}$; (ii) $\widetilde{D}(\widetilde{w}_{t_1}, \widetilde{w}_{t_2}) = \widetilde{D}(\widetilde{w}_{t_2}, \widetilde{w}_{t_1})$; (iii) $\widetilde{D}(\widetilde{w}_{t_1}, \widetilde{w}_{t_2}) \le \widetilde{D}(\widetilde{w}_{t_1}, \widetilde{w}_{t_3}) + \widetilde{D}(\widetilde{w}_{t_3}, \widetilde{w}_{t_2})$. 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 $\widetilde{D}$. 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 CO_2 emissions.

Methods

Suppose that a scalar time series $\{\mathrm{s(t)} | \mathrm{t} = 1, 2, \ldots\}$ is given successively. Denote by N the size of the database. Let $\vec{s}(t) = (s(t), \lambda s(t-1), \lambda^2 s(t-2), \ldots)$ be the infinite-dimensional weighted delay coordinates, where λ is a decay rate. We define a distance $\|\vec{s}(t_1) - \vec{s}(t_2)\|_{l_1}$ between $\vec{s}(t_1)$ and $\vec{s}(t_2)$ as

$$\|\vec{s}(t_1) - \vec{s}(t_2)\|_{l_1} = \sum_{d=0}^{\infty} \lambda^d |s(t_1 - d) - s(t_2 - d)|. \tag{2}$$

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:

$$\begin{aligned} \|\vec{s}(t_1) - \vec{s}(t_2)\|_{l_1} &= |s(t_1) - s(t_2)| + \sum_{d=1}^{\infty} \lambda^d |s(t_1 - d) - s(t_2 - d)| \\ &= |s(t_1) - s(t_2)| + \lambda \sum_{d=0}^{\infty} \lambda^d |s(t_1 - 1 - d) - s(t_2 - 1 - d)| \\ &= |s(t_1) - s(t_2)| + \lambda \|\vec{s}(t_1 - 1) - \vec{s}(t_2 - 1)\|_{l_1}. \end{aligned} \tag{3}$$

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:

$$\|\vec{s}(t_1) - \vec{s}(t_2)\|_{l_1} = |s(t_1) - s(t_2)| - |s(t_1 - d) - s(t_2 - d)| + \|\vec{s}(t_1 - 1) - \vec{s}(t_2 - 1)\|_{l_1}. \tag{4}$$

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 $\Delta_i^t = \|\vec{s}(t-1) - \vec{s}(t-i-1)\|_{l_1}$ (i = 1, …, N) and a list of the corresponding observed values $S_i^t = s(t-i)$ (i = 1, …, N). First, we find the set $T(t, p) = \{i_1, \ldots, i_K\}$ of indices for the K smallest distances among Δ_i^t ($P \le i \le N$), and provide the p steps ahead prediction for each p ($0 < p \le P$) by $\frac{1}{K}\sum_{i \in T(t,p)} S_{i-p+1}^t$. We used K = 10 if not mentioned. Second, we update the list of distances by

$$\begin{aligned} \Delta_i^{t+1} &= \|\vec{s}(t) - \vec{s}(t-i)\|_{l_1} \\ &= \sum_{d=0}^{\infty} \lambda^d |s(t-d) - s(t-i-d)| \\ &= |s(t) - s(t-i)| + \lambda \sum_{d=0}^{\infty} \lambda^d |s(t-1-d) - s(t-i-1-d)| \\ &= |s(t) - s(t-i)| + \lambda \|\vec{s}(t-1) - \vec{s}(t-i-1)\|_{l_1} \\ &= |s(t) - s(t-i)| + \lambda \Delta_i^t. \end{aligned} \tag{5}$$

We also update the list of the corresponding values by

$$S_{i+1}^{t+1} = S_i^t \tag{6}$$

for $1 \le i < N$ and

$$S_1^{t+1} = s(t). \tag{7}$$

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).

10. Berry, T., Cressman, J. R., Gregurić-Ferenček, Z. & Sauer, T. Time-scale separation from diffusion-mapped delay coordinates. *SIAM J. Appl. Dyn. Syst.* **12,** 618–649 (2013).
11. Lorenz, E. N. Atmospheric predictability as revealed by naturally occurring analogues. *J. Amtos. Sci.* **26,** 636–646 (1969).
12. Lorenz, E. N. Predictability: A problem partly solved. in *Proceedings of the Seminar on Predictability*, Vol. 1, pp.1–18 (ECMWF, 1996).
13. Hansen J. A. & Smith, L. A. The role of operational constraints in selecting supplementary observations. *J. Atmos. Sci.* **57,** 2859–2871 (2000).
14. Fraser, A. M. & Swinny, H. L. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33,** 1134-(1986).
15. Kennel, M. B., Brown, R. & Abarbanel, H. D. I.. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45,** 3403–3411 (1992).
16. Cesa-Bianchi, N. & Lugosi, G. *Prediction, Learning, and Games* (New York, Cambridge Univ. Press, 2006).
17. Chernov, A. & Zhdanov, F. Prediction with expert advice under discounted loss. *Proc. Of AIT 2010, Lecture Notes in Artificial Intelligence* **6331**, 255–269 (2010).
18. Morino, K., Hirata, Y., Tomioka, R., Kashima, H., Yamanishi, K., Hayashi, N., Egawa, S. & Aihara, K. Predicting disease progression from short biomarker series using expert advice algorithm. *Sci. Rep.* **5,** 8953 (2015).
19. Cao, L., Mees, A. & Judd, K. Dynamics from multivariate time series. *Physica D* **121,** 75–88 (1998).
20. Bocaletti, S., Valladares, D. L., Pecora, L. M., Geffert, H. P. & Carroll, T. Reconstructing embedding spaces of coupled dynamical systems from multivariate data. *Phys. Rev. E* **65,** 035204(R) (2002).
21. Garcia, S. P. & Almeida J. S. Multivariate phase space reconstruction by nearest neighbor embedding with different time delays. *Phys. Rev. E* **72,** 027205 (2005).
22. Hirata, Y., Suzuki, H. & Aihara, K. Reconstructing state spaces from multivariate data using variable delays. *Phys. Rev. E* **74,** 026202 (2006).
23. Pecora, L. M., Moniz, L., Nichols, J. & Carroll, T. L. A unified approach to attractor reconstruction. *Chaos* **17,** 013110 (2007).
24. Vlachos, I. & Kugiumtzis, D. Nonuniform state-space reconstruction and coupling detection. *Phys. Rev. E* **82,** 016207 (2010).
25. Victor, J. D. & Purpura, K. P. Metric-space analysis of spike trains: theory, algorithms and application. *Network* **8,** 127–164 (1997).
26. van Rossum, M. C. W. A novel spike distance. *Neural Comp.* **13,** 751–763 (2001).
27. Kreuz, T., Haas, J. S., Morelli, A., Abarbanel, H. D. I. & Politi, A. Measuring spike train synchrony. *J. Neurosci. Methods* **165,** 151–161 (2007).
28. Schoenberg, F. P. & Tranbarger, K. E. Description of earthquake aftershock sequences using prototype point patterns. *Environmetrics* **19,** 271–286 (2008).
29. Suzuki, S. Hirata, Y. & Aihara, K. Definition of distance for marked point process data and its application to recurrence plot-based analysis of exchange tick data of foreign currencies. *Int. J. Bifurcat. Chaos* **20,** 3699–3708 (2010).
30. Deza, M. M. & Deza, E. *Encyclopedia of distances*. 2nd edition (Springer, 2012).
31. Kantz, H. A robust method to estimate the maximal Lyapunov exponent of a time series. *Phys. Lett. A* **185,** 77–87 (1994).

Acknowledgements

We thank the Japan Meteorological Agency for providing us the datasets of sunshine duration and wind speed. The used datasets are commercially available from Japan Meteorological Business Support Center. This manuscript is partially based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This research is also partially supported by Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST).

Author Contributions

Y.H. and K.A. made the ground design of this study. Y.H., T.T., S.H. and K.A. designed the study of wind speed data. Y.H., H.S. and K.A. designed the study of sunshine duration data. Y. H. and T.T. created the method. Y.H. and T.T. wrote the programs and analyzed the datasets. All authors contributed to interpreting the results, and writing the manuscript.

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).