Fast time-series prediction using high-dimensional data: Evaluating confidence interval credibility

Yoshito Hirata
Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan
(Received 28 January 2014; published 29 May 2014)

I propose an index for evaluating the credibility of confidence intervals for future observables predicted from high-dimensional time-series data. The index evaluates the distance from the current state to the data manifold. I demonstrate the index with artificial datasets generated from the Lorenz'96 II model [Lorenz, in *Proceedings of the Seminar on Predictability*, Vol. 1 (ECMWF, Reading, UK, 1996), p. 1], the Lorenz'96 I model [Hansen and Smith, J. Atmos. Sci. **57**, 2859 (2000)], and the coupled map lattice, and a real dataset for the solar irradiation around Japan.

DOI: 10.1103/PhysRevE.89.052916 PACS number(s): 05.45.Tp, 05.45.Jn

I. INTRODUCTION

High-dimensional time-series data are ubiquitous because of recent developments in measurement techniques. Methods of extracting valuable information from such data are currently a key topic. An example in the field of renewable energy is the prediction of solar irradiation. Because solar irradiation fluctuates abruptly owing to weather conditions, its predicted values should be used to plan and control the outputs of other backup power plants. Such renewable energy outputs should be predicted online in the time scales from minutes to days because the weather conditions change the outputs in these time scales. In addition, such prediction should not consume much electricity, the production of which is a purpose of renewable energy. For applications in renewable energy, we recently proposed a method of time-series prediction that is online, spans multiple steps ahead [1], and is accompanied by confidence intervals [2], while conventional time-series prediction methods [3–8] in the context of nonlinear dynamics predict the future values but not confidence intervals, and often assume some family of mathematical models. However, a problem has arisen: The confidence intervals sometimes could capture the actual values observed in the future less often than the intended accuracy of the confidence intervals.

In this paper, I propose an index to distinguish cases where the credibility of the confidence intervals is high from cases where the credibility is low. My main idea is to determine how faithfully points in the database represent the current state. The remaining sections are organized as follows: In Sec. II, I review our previous study for predicting high-dimensional data based on past observations and discuss its problem. In Sec. III, I formulate the proposed index. In Sec. IV, I show examples of the proposed index. In Sec. V, I discuss the results and conclude the paper.

II. FAST TIME-SERIES PREDICTION FOR HIGH-DIMENSIONAL DATA

I review the method of online time-series prediction proposed in Refs. [1,2]. This method is an extension of the method of Kwasniok and Smith [9,10] by piecewise constant prediction [11]. In particular, Ref. [2] introduces online multistep time-series prediction accompanied by confidence intervals. A literature review [2,12,13] shows that more work is needed in this area.

A. The algorithm

Suppose that I observe a multivariate time series $x_t \in R^e$, where e is the dimension of my observations. Let $B \in (R^e)^{\beta\times(d+q)}$ be a database that is a matrix with β rows and $(d+q)$ columns, containing e-dimensional row vectors. The first d columns correspond to the past part of the database, and the following q columns correspond to the future part of the database.

During $t < \beta + d + q$, I feed the observed point of the time series to the database by $B(i,l) = x_t$ where $i + l = t + 1$, $1 \leqslant i \leqslant \beta$, $1 \leqslant l \leqslant (d+q)$.

When $t \geqslant \beta + d + q$, I first predict x_{t+p} ($p \in \{1,2,\ldots,q\}$): (i) I find the K nearest rows for the current state $(x_{t-d},x_{t-d+1},\ldots,x_{t-1})$ in terms of the Euclidean distance from the past part $B(:,1:d)$ of the database, where $B(:,1:d)$ is the first d columns of the matrix B. Let I_t be the set of indices for the K nearest rows. (ii) I construct the $100(1-1/K)\%$ confidence interval for the p steps ahead ($p \in \{1,2,\ldots,q\}$) of the jth element of the observation by

$$\left[\min_{i\in I_t}(B(i,d+p))_j,\ \max_{i\in I_t}(B(i,d+p))_j\right],$$

where $(B(i,d+p))_j$ corresponds to the jth element of the $(i,d+p)$ component of B. I also construct the $100(1-1/K)\%$ confidence interval for the sum of all the elements of the observation p steps ahead by

$$\left[\min_{i\in I_t}\sum_{j=1}^{c}(B(i,d+p))_j,\ \max_{i\in I_t}\sum_{j=1}^{c}(B(i,d+p))_j\right]$$

for each $p \in \{1,2,\ldots,q\}$.

Second, I attempt to update the database B: (i) I observe x_t. (ii) I use the K nearest rows for $(x_{t-q-d+1},x_{t-q-d+2},\ldots,x_{t-q})$ to predict x_{t-q+p}. Letting J_t be the index for the K nearest rows, I have

$$\hat{x}_{t-q+p} = \frac{1}{K}\sum_{i\in J_t} B(i,d+p)$$

for each $p \in \{1,2,\ldots,q\}$.

(iii) I construct a temporary database B' by randomly choosing a row k and inserting $B'(i,l) = B(i,l)$ for $i \neq k, l = 1,2,\ldots,d+q$ and $B'(i,l) = x_{t-q-d+l}$ for $i = k, l = 1,2,\ldots,d+q$. (iv) I predict $B(k,d+p)$ by finding the K nearest rows for $B(k,1:d)$ from $B'(:,1:d)$. Here $B(k,1:d)$

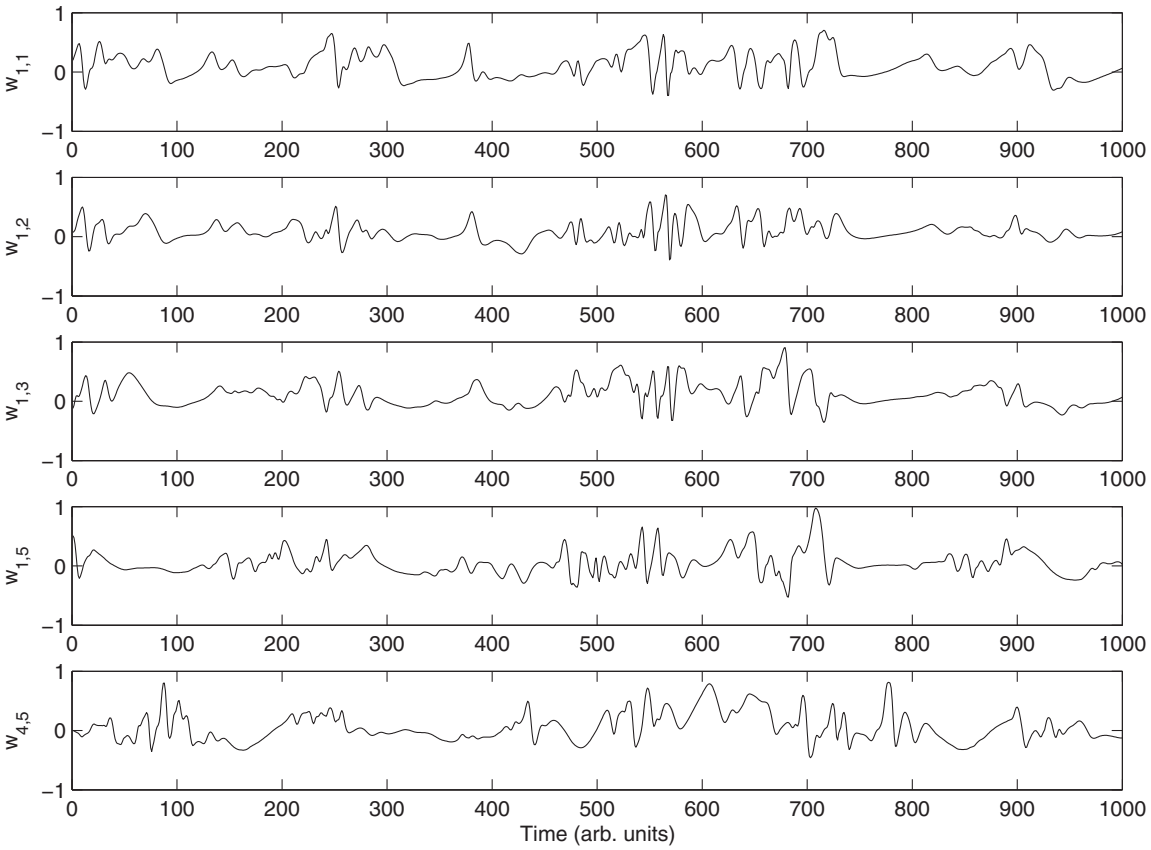

FIG. 1. A part of the time series of the Lorenz'96 II model.

is the kth row of B restricted between the first and dth columns, and $B'(:,1:d)$ is the first d columns of B'. Letting J'_t be the set of indices for the K nearest rows, the prediction for $B(k,d+p)$ can be obtained by

$$\hat{B}(k,d+p) = \frac{1}{K}\sum_{i\in J'_t} B(i,d+p).$$

(v) I compare the prediction errors of $\hat{x}_{t-q+p}$ and $\hat{B}(k,d+p)$ for $p \in \{1,2,\ldots,q\}$. If $\|\hat{x}_{t-q+p} - x_{t-q+p}\| > \|\hat{B}(k,d+p) - B(k,d+p)\|$ for more than half of $p \in \{1,2,\ldots,q\}$, I update B by B'. Otherwise, I do not update B.

B. Example: Lorenz'96 II

I evaluated the performance of the above algorithm using a time series generated from the Lorenz'96 II model [14,15], which is a toy model of the atmosphere. This model contains two types of variables: The first type, v_g, corresponds to the variables representing the upper layer, and the second type, $w_{g,h}$, corresponds to those representing the lower layer close to the surface of the Earth. The index g runs between 1 and $G = 40$, and the index h runs between 1 and $H = 5$. The model is represented as

$$\frac{dv_g}{dt} = v_{g-1}(v_{g+1} - v_{g-2}) - v_g + F - \frac{a_v c}{b}\sum_{h=1}^{H} w_{g,h},$$

$$\frac{dw_{g,h}}{dt} = cbw_{g,h+1}(w_{a,h-1} - w_{g,h+2}) - cw_{g,h} + \frac{a_w c}{b}v_g,$$

where I apply the following boundary conditions:

$$v_{G+g} = v_g,$$
$$w_{g,h+H} = w_{g+1,h},$$
$$w_{g,h-H} = w_{g-1,h}.$$

I used the following parameters: $F = 8$, $b = 10$, $c = 10$, $a_v = 1$, $a_w = 1$. I observed $w_{g,h}$ ($g \in \{1,2,3,4\}$, $h \in \{1,2,3,4,5\}$) every 0.01 time units to obtain a 20-dimensional time series that has 10 000 points. This time series can be regarded as a regional surface observation of the atmosphere. A part of the time series is shown in Fig. 1.

I set $\beta = 2000$, $d = 10$, $q = 20$, and $K = 25$ to obtain the $100(1 - 1/25) = 96\%$ confidence intervals for the sum of the 20 observables.

The result is presented in Fig. 2. Although I intended to obtain the 96% confidence intervals, the probabilities that the 96% confidence intervals contain the actual values were less than 94% throughout the prediction steps I tested, which were between 1 and 20.

III. PROPOSED METHOD

A possible reason that the 96% predicted confidence intervals had a lower probability of containing the actual values in the example of the Lorenz'96 II model could be that the dimension for the dynamics is too high, and the neighborhood constructed using the 25 nearest neighbors may sometimes not represent the current state well. In such a case, the current state

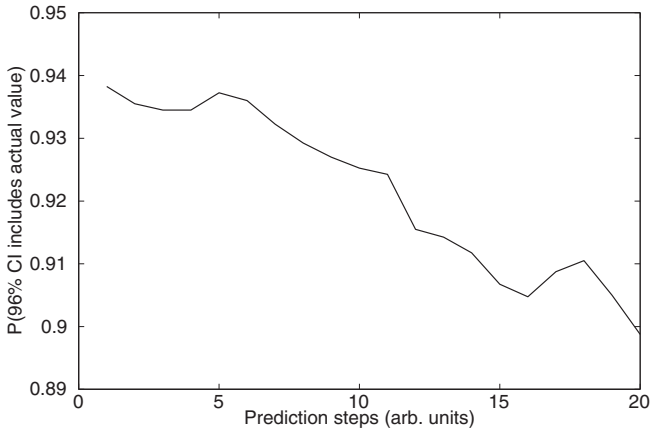

FIG. 2. Probability that the 96% confidence interval (CI) includes the actual value given the number of prediction steps, Lorenz'96 II model.

and the manifold spanned by the data points, which I call the data manifold, can be separated by a large distance, as shown in Fig. 3.

Therefore, I evaluate the distance between the current state and the data manifold. For this purpose, first I approximate the data manifold locally by a hyperplane (see Fig. 4). Let $I_t = \{I_{t,1}, I_{t,2}, \ldots, I_{t,K}\}$ be a set of indices for the K nearest neighbors for the current state $(x_{t-d}, x_{t-d+1}, \ldots, x_{t-1})$. I obtain the hyperplane such that $\{B(i,1:d), i \in I_t\}$ are close to it. Such a hyperplane can be written as

$$\sum_{i=1}^{d}\sum_{j=1}^{e}\alpha_{i,j}y_{i,j} = \gamma_t,$$

where $((y_{1,1}, y_{1,2}, \ldots, y_{1,e})\ (y_{2,1}, y_{2,2}, \ldots, y_{2,e}) \cdots (y_{d,1}, y_{d,2}, \ldots, y_{d,e}))$ is in the same space as the current state $(x_{t-d}, x_{t-d+1}, \ldots, x_{t-1})$. Although $\alpha_{i,j}$ depends on t, I drop t here for simplicity. In addition, I write

$$\alpha_t = (\alpha_{1,1}\alpha_{1,2}\cdots\alpha_{d,e})^T.$$

Here T represents the transpose of the matrix. There are $(de+1)$ parameters for the hyperplane. The problem of choosing the parameters can be formulated as the following

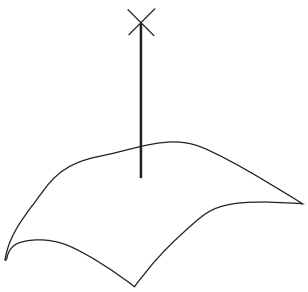

FIG. 3. Current state (cross at top) and data manifold (bottom surface). The quality of the representation of the current state can be evaluated using the distance between the current state and the data manifold (thick vertical line).

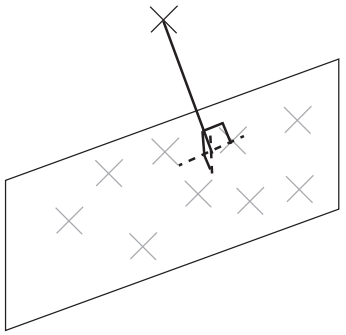

FIG. 4. Schematic diagram of current state (black cross), neighboring points (gray crosses), and hyperplane (parallelogram). Thick black line indicates distance between current state and hyperplane.

minimization problem:

$$\min_{\alpha_t}\left\|\begin{pmatrix}(B(I_{t,1},1))_1 & (B(I_{t,1},1))_2 & \cdots & (B(I_{t,1},d))_e\\ (B(I_{t,2},1))_1 & (B(I_{t,2},1))_2 & \cdots & (B(I_{t,2},d))_e\\ \vdots & \vdots & \ddots & \vdots\\ (B(I_{t,K},1))_1 & (B(I_{t,K},1))_2 & \cdots & (B(I_{t,K},d))_e\end{pmatrix} \times \begin{pmatrix}\alpha_{1,1}\\ \alpha_{1,2}\\ \vdots\\ \alpha_{d,e}\end{pmatrix} - \begin{pmatrix}\gamma_t\\ \gamma_t\\ \vdots\\ \gamma_t\end{pmatrix}\right\|^2. \quad (1)$$

I rewrite this as

$$\min_{\alpha_t}\|\breve{B}_t\alpha_t - \gamma_t\vec{1}\|^2.$$

Namely, $\vec{1}$ is the column vector such that all its elements are 1. However, the number of parameters is too large compared to the number K of neighboring points. Therefore, the problem of choosing the parameters is overdetermined.

To overcome this difficulty, I enforce two constraints. The first is that $\alpha_{i,j} = \gamma_t\bar{\alpha}_{i,j}$. Dividing Eq. (1) by γ_t^2, I obtain the following cost function:

$$\min_{\bar{\alpha}_t}\|\breve{B}_t\bar{\alpha}_t - \vec{1}\|^2. \quad (2)$$

The second constraint is the minimization of the norm for the coefficients. Namely, I simultaneously minimize the following cost function:

$$\min_{\bar{\alpha}_t}\sum_{i=1}^{d}\sum_{j=1}^{e}(\bar{\alpha}_{i,j})^2 = \min_{\bar{\alpha}_t}\|\bar{\alpha}_t\|^2. \quad (3)$$

Combining Eqs. (2) and (3) with a Lagrangian multiplier λ, I minimize the following cost function to obtain $\bar{\alpha}_{i,j}$:

$$\min_{\bar{\alpha}_t}\{\|\breve{B}_t\bar{\alpha}_t - \vec{1}\|^2 + \lambda\|\bar{\alpha}_t\|^2\}. \quad (4)$$

Solving Eq. (4) is easy. By partially differentiating Eq. (4) in terms of $\bar{\alpha}_t$, I have

$$\left\{\lambda E + \breve{B}_t^T\breve{B}_t\right\}\bar{\alpha}_t = \breve{B}_t\vec{1},$$

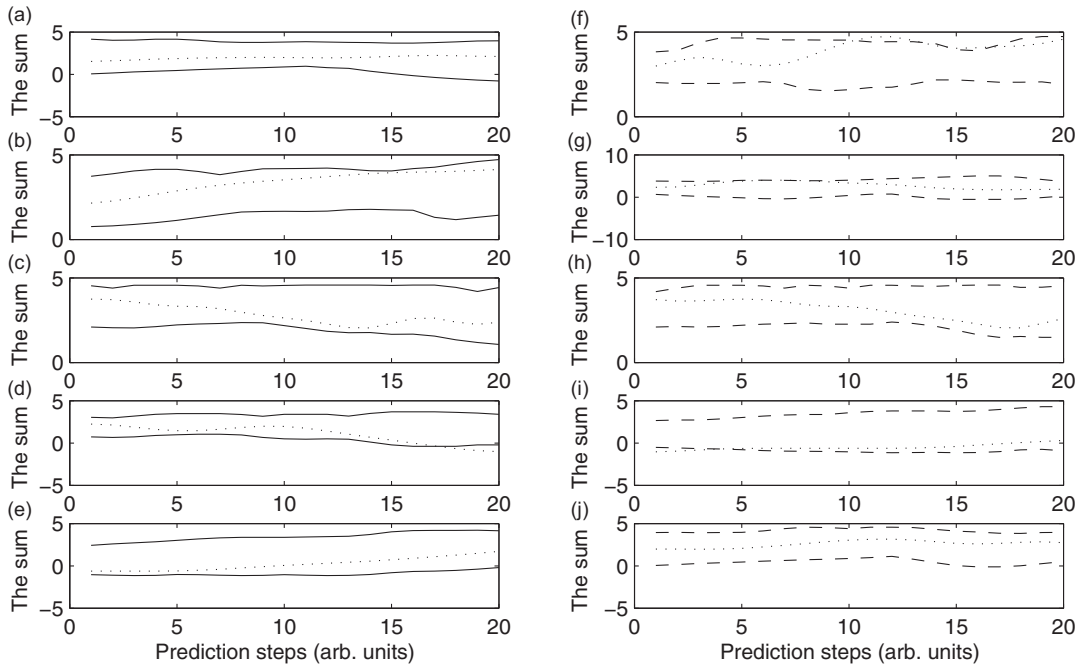

FIG. 5. Examples of predicted 96% confidence intervals for the sums of the observables for the dataset generated from the Lorenz'96 II model without observational noise. (a)–(e) $D_t < \bar{D}$; (f)–(j) $D_t \geqslant \bar{D}$. In each panel, solid lines or dashed lines show predicted 96% confidence intervals, and dotted line represents actual values.

where E is the identity matrix. By multiplying by $\{\lambda E + \breve{B}_t^T \breve{B}_t\}^{-1}$ from the left-hand side, I obtain

$$\bar{\alpha}_t = \left\{\lambda E + \breve{B}_t^T \breve{B}_t\right\}^{-1} \breve{B}_t \vec{1}.$$

Therefore, I have

$$\alpha_t = \gamma_t \left\{\lambda E + \breve{B}_t^T \breve{B}_t\right\}^{-1} \breve{B}_t \vec{1}.$$

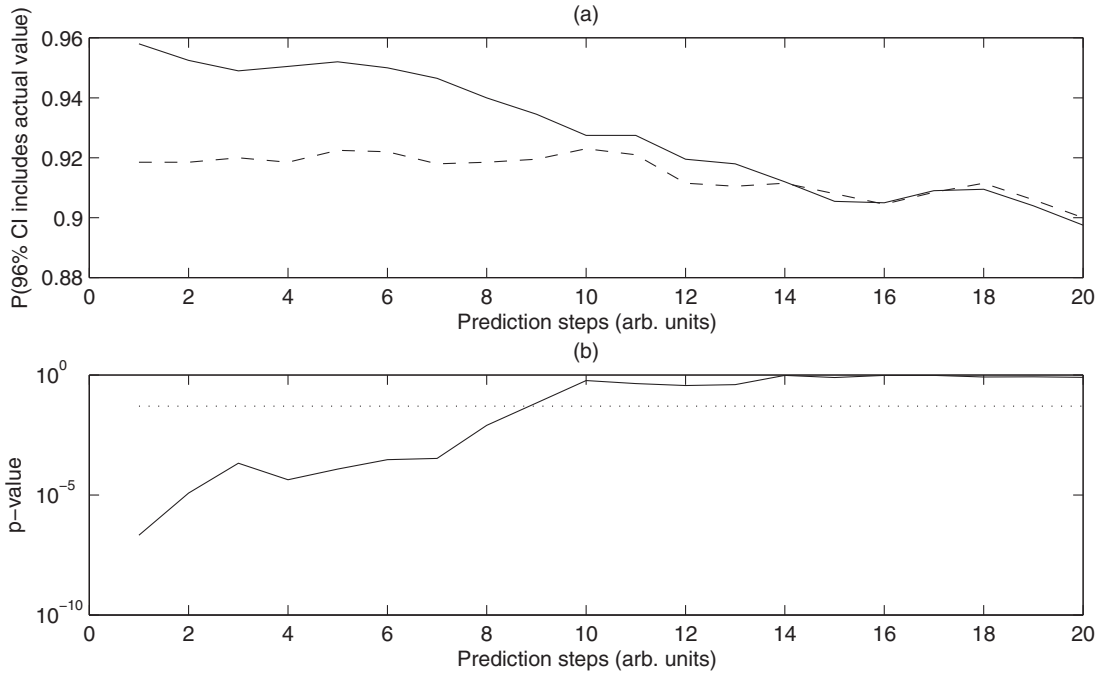

FIG. 6. Comparison of cases where the proposed index was small and large for the Lorenz'96 II model without noise ($\lambda = 1$). (a) Solid and dash-dotted lines correspond to $D_t < \bar{D}$ and $D_t \geqslant \bar{D}$, respectively; (b) shows the significance of the difference between these two cases. Solid line shows p-values, and dotted line represents line of $p = 0.05$.

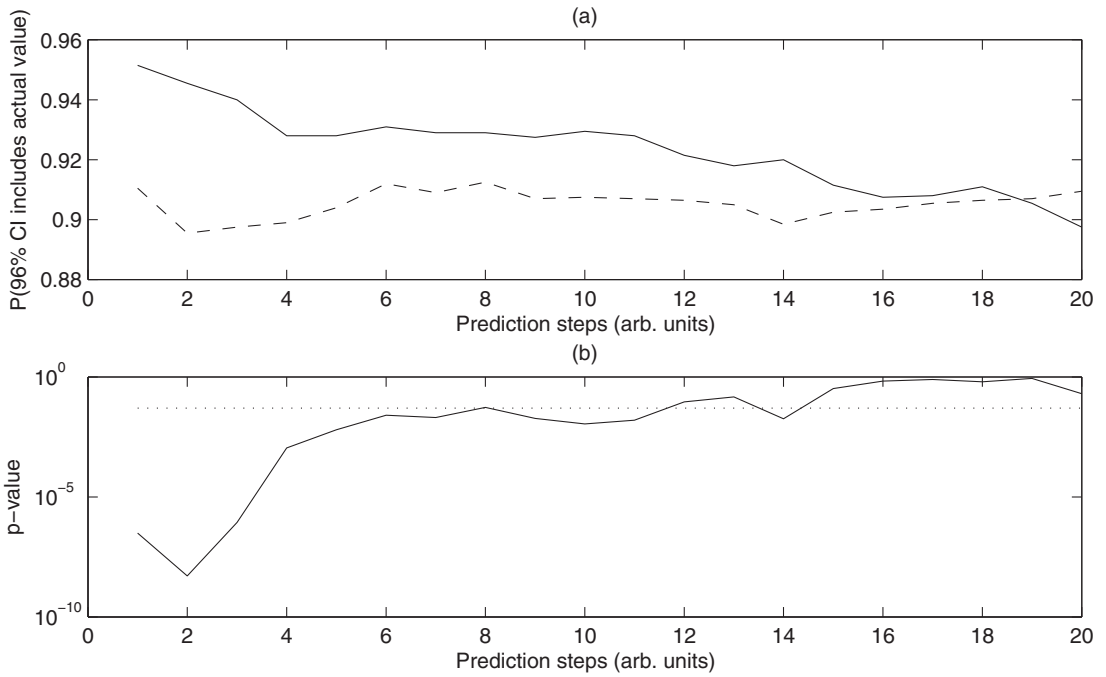

FIG. 7. Comparison of cases where the proposed index was small and large in time-series data generated from the Lorenz'96 II model contaminated by 5% observational noise. I set $\lambda = 1$. Lines have the same meaning as in Fig. 6.

For simplicity, I choose γ_t so that the norm of α_t becomes 1. Namely,

$$\gamma_t = \frac{1}{\left\| \left\{ \lambda E + \breve{B}_t^T \breve{B}_t \right\}^{-1} \breve{B}_t \vec{1} \right\|},$$

and

$$\alpha_t = \frac{\left\{ \lambda E + \breve{B}_t^T \breve{B}_t \right\}^{-1} \breve{B}_t \vec{1}}{\left\| \left\{ \lambda E + \breve{B}_t^T \breve{B}_t \right\}^{-1} \breve{B}_t \vec{1} \right\|}.$$

Because α_t represents the normal vector for the hyperplane whose norm is 1, the following D_t provides the distance between the hyperplane and the current state:

$$D_t = |(x_{t-q}\ x_{t-q+1}\ \cdots\ x_{t-1})\alpha_t - \gamma_t|.$$

I evaluate this distance D_t at the beginning of the predicted part of the time series and obtain the median value $\bar{D}$. In the following part of the time series, I compare the cases of $D_t < \bar{D}$ and $D_t \geqslant \bar{D}$. I will examine whether the probability that the obtained confidence interval contains the actual value is larger if $D_t < \bar{D}$.

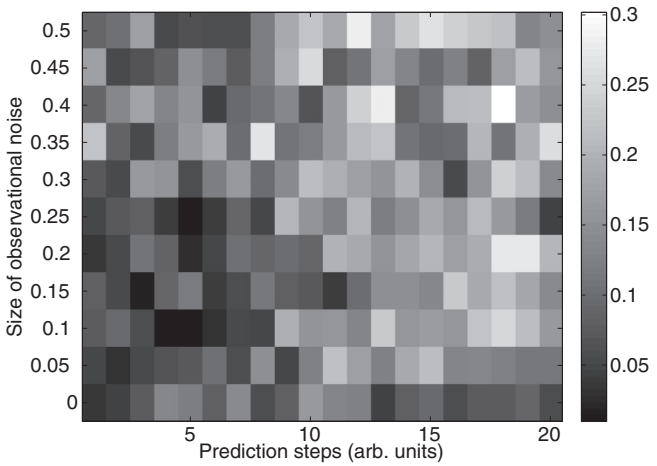

FIG. 8. The average significance level for the difference between $D_t < \bar{D}$ and $D_t \geqslant \bar{D}$ given a prediction step and the size of observational noise, in the case of the Lorenz'96 II model.

TABLE I. Dependence of 5% significance on sets of prediction steps and levels of observational noise, for the example of the Lorenz'96 II model. In this table, the number of joint sets for prediction steps and realizations is shown for each condition and each class of significance levels. I found that the condition that prediction steps is less than or equal to 10 is not independent of the 5% significance (p-value obtained by the one-sided Fisher's exact test using R package: 4.9×10^{-6}). I also found that the condition that the noise level is less than or equal to 0.25 is not independent of 5% significance (p-value: 5.3×10^{-3}).

Condition–significance level				
Prediction steps	Noise level	$\leqslant$0.05	$>$0.05	Total
$\leqslant$10	$\leqslant$0.25	446(74%)	154(26%)	600
	$>$0.25	393(66%)	207(35%)	600
$>$10	$\leqslant$0.25	346(69%)	154(31%)	500
	$>$0.25	301(60%)	199(40%)	500
Total		1486(68%)	714(32%)	2200

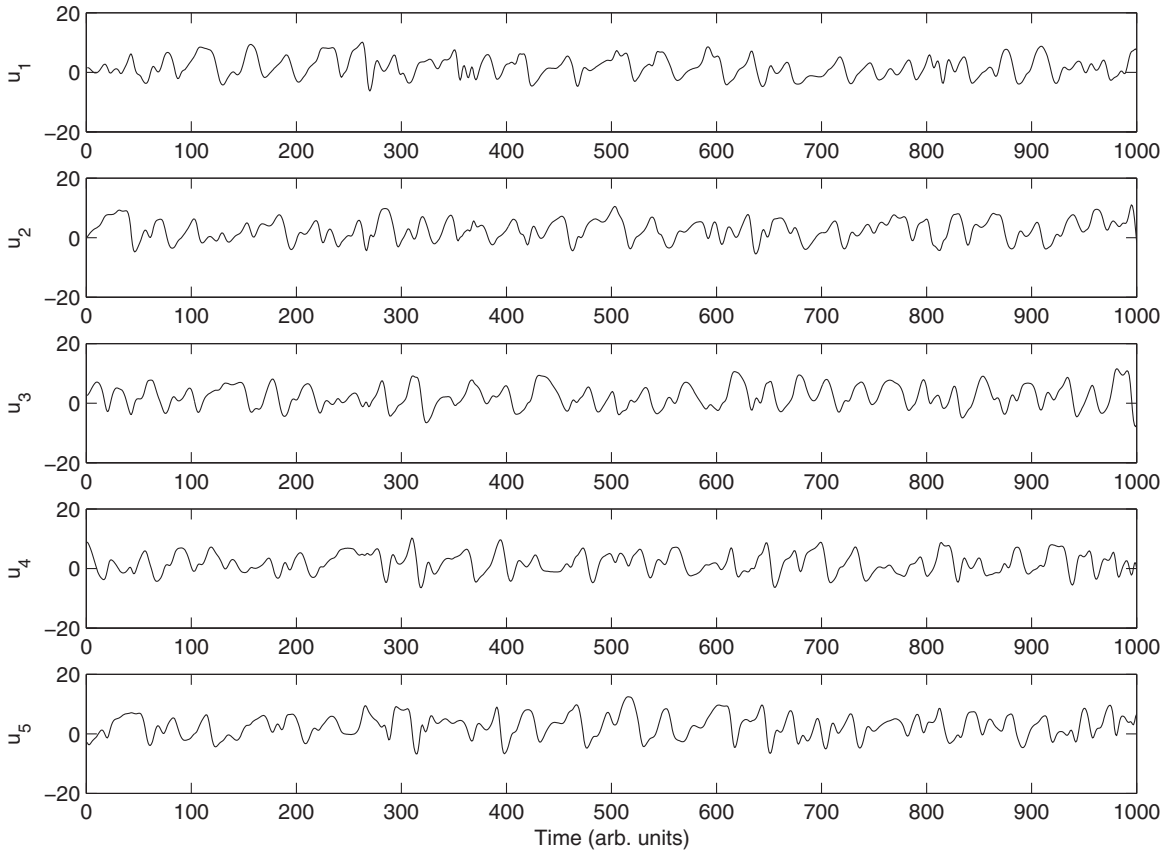

FIG. 9. A part of the time series for the Lorenz'96 I model.

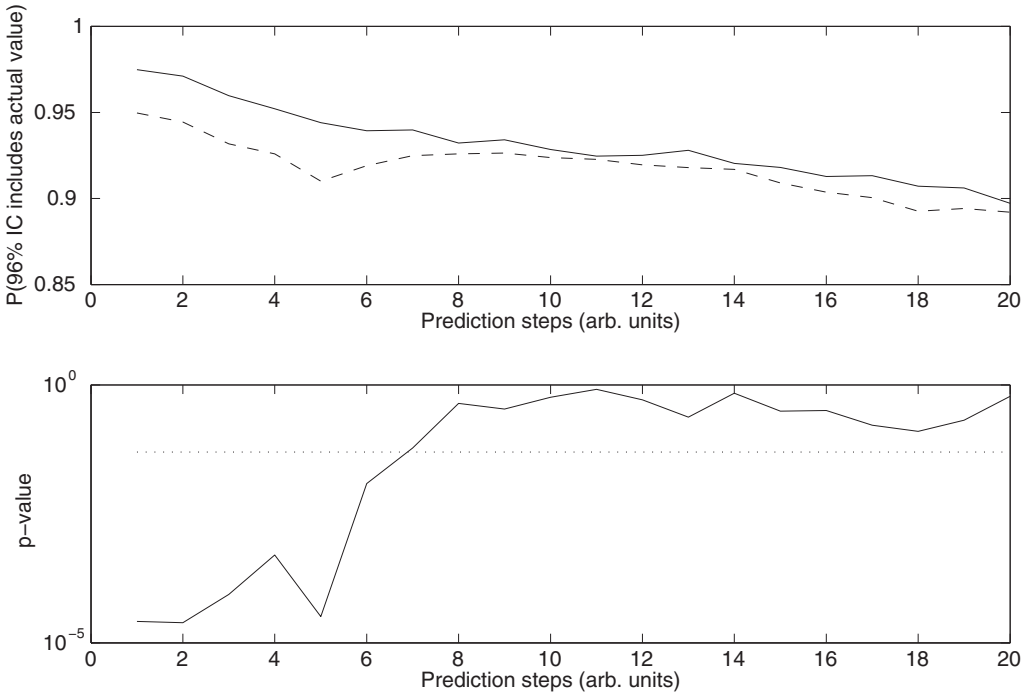

FIG. 10. Comparison between cases where the proposed index is low and high, in the case of the Lorenz'96 I model. See the caption of Fig. 6 to interpret the results.

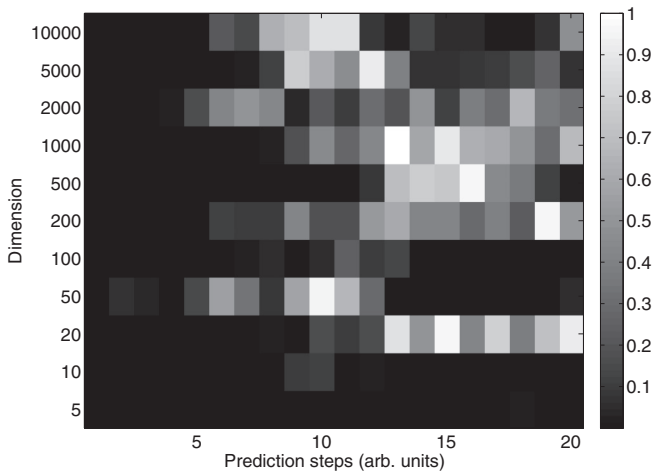

FIG. 11. The significance level for the difference between low and high D_t given the dimension of model and a prediction step, in the case of the Lorenz'96 I model.

IV. EXAMPLES

A. Lorenz'96 II without observational noise

First, I applied the method to the Lorenz'96 II model used in Sec. II. I applied the same conditions as in Sec. II B except that I used 2000 points starting at the 3001st point to obtain the median $\bar{D}$ of the proposed index and evaluated the prediction using 4000 points starting at the 5001st point.

Examples of the predicted 96% confidence intervals are shown in Fig. 5, and the results are summarized in Fig. 6. I found that the predicted confidence intervals tended to contain the actual values with a higher probability when $D_t < \bar{D}$ than when $D_t \geqslant \bar{D}$ if the prediction step was smaller than 14 [see Fig. 6(a)]. For instance, in the examples of Figs. 5(d), 5(g), and 5(i), the actual values were out of the ranges specified by the predicted 96% confidence intervals for at least a prediction step. The difference in the probabilities between $D_t < \bar{D}$ and $D_t \geqslant \bar{D}$ was statistically significant when the prediction step

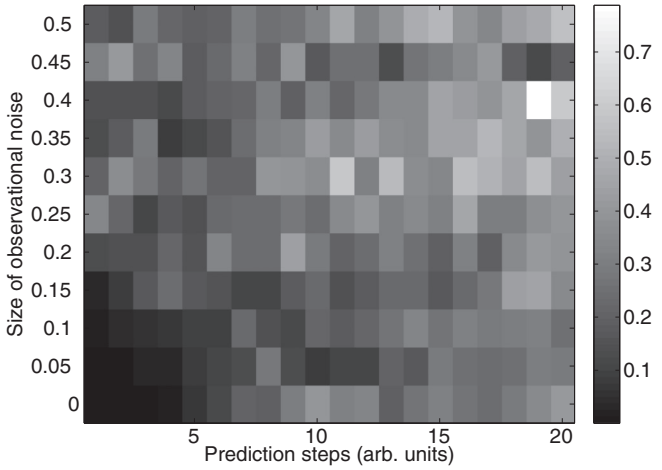

FIG. 12. The average significance level for the difference between cases where the proposed index is low and high given a prediction step and the size of observational noise, in the case of the Lorenz'96 I model.

TABLE II. Dependence of 5% significance on sets of prediction steps and levels of observational noise, for the example of the Lorenz'96 I model. See the caption of Table I to interpret this table. The condition that prediction steps $\leqslant 10$ is not statistically independent of the 5% significance (p-value $< 2.2 \times 10^{-16}$). In addition, the condition that noise level $\leqslant 0.25$ is neither statistically independent of the 5% significance (p-value $< 2.2 \times 10^{-16}$).

Condition–significance level				
Prediction steps	Noise level	$\leqslant 0.05$	>0.05	Total
$\leqslant 10$	$\leqslant 0.25$	337(56%)	263(43%)	600
	>0.25	181(30%)	419(70%)	600
>10	$\leqslant 0.25$	163(33%)	337(67%)	500
	>0.25	77(15%)	423(85%)	500
Total		758(34%)	1442(66%)	2200

was smaller than 9, with a significance level of 0.05 [the chi-square test [16]; see Fig. 6(b)]. It took 65 min to finish the prediction (2.4 GHz Intel Core 2 Duo CPUs with 4 GB of memory. I used the same computer for the other computations).

B. Lorenz'96 II with observational noise

I also tested the same dataset with an added 5% Gaussian observational noise. The other computational conditions were the same as in the previous subsection.

The results are presented in Fig. 7. Although the differences became smaller when the prediction steps were medium (between 4 and 8), the results were similar to those for the dataset without observational noise in that in short prediction ranges, the confidence intervals with smaller D_t exhibited a higher probability of containing the actual values within them. These results mean that the proposed index is robust against observational noise.

I also evaluated the influence of observational noise by varying the size of observational noise from 0% to 50% (see Fig. 8). I generated ten different realizations of time series using different initial conditions, added observational noise for each noise level, and averaged the significance level for the difference between $D_t < \bar{D}$ and $D_t \geqslant \bar{D}$ over the ten different realizations. I found that the p-value for $D_t < \bar{D}$ was likely to be smaller than that for $D_t \geqslant \bar{D}$ in the significance level of 5% if the size of observational noise was less than 30% or if the prediction step was less than or equal to ten steps (see Fig. 8 and Table I).

C. Lorenz'96 I without noise

In addition, I evaluated the performance for the proposed index with the Lorenz'96 I model [14,15]. The model is written as

$$\frac{du_g}{dt} = -u_{g-2}u_{g-1} + u_{g-1}u_{g+1} - u_g + F,$$

and

$$u_{g+G} = u_g,$$

where I set $F = 8$. I sampled u_g ($g = 1,2,\ldots,5$) every 0.05 unit times and generated a time series of length 10 000. A part of the time series is shown in Fig. 9. I predicted the sum

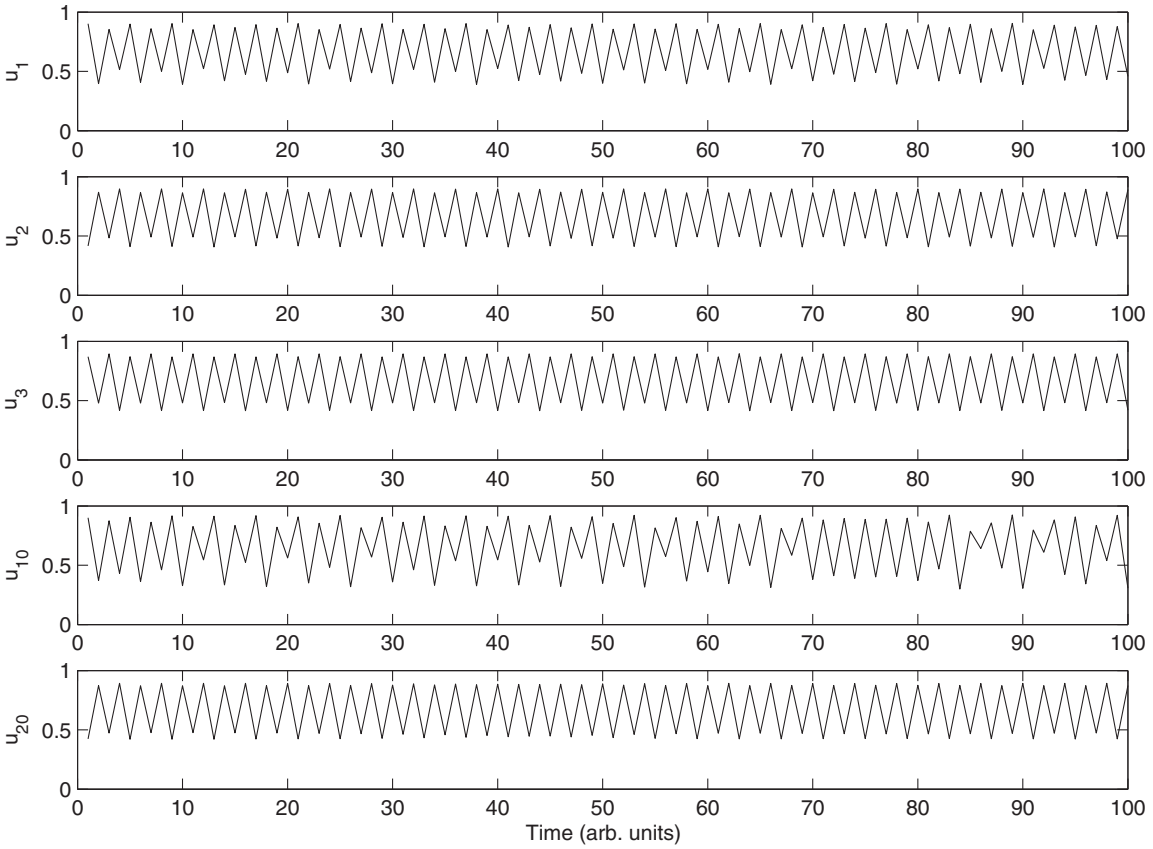

FIG. 13. A part of the time series for the coupled map lattice.

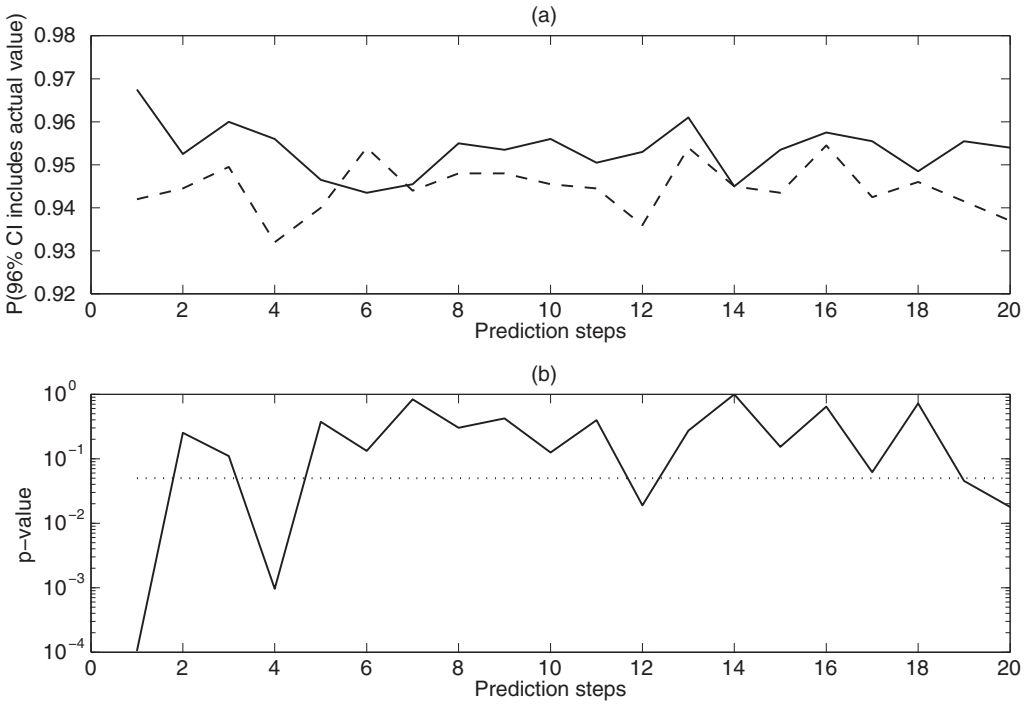

FIG. 14. Comparison for low and high D_t given a prediction step, in the case of the coupled map lattice. See the caption of Fig. 6 to interpret the results.

of $\sum_{g=1}^{5} u_g$. The rest setting was the same as the case of the Lorenz'96 II model.

When I used the $G = 40$-dimensional model, I obtained the results shown in Fig. 10. The probability that the 96% confidence interval contains the actual value was higher when $D_t < \bar{D}$ in the tested prediction range. Especially, the difference between $D_t < \bar{D}$ and $D_t \geqslant \bar{D}$ was significant when the prediction step was less than 7.

When I changed the dimensions to 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, and 10 000, I obtained the results shown in Fig. 11. These results mean that the proposed index worked for discriminating the cases where the prediction method discussed in Sec. II A was more effective when the prediction steps were small.

D. Lorenz'96 I with observational noise

I also evaluated the case where the observation of the Lorenz'96 I model was disturbed by observational noise. I set $G = 40$ and I generated ten different time series using different initial conditions. Moreover, I changed the size of observational noise from 0% to 50%, taking the average of significance level over ten different realizations of time series (see Fig. 12). I found the tendencies that the p-value for $D_t < \bar{D}$ was likely to be smaller than the p-value for $D_t \geqslant \bar{D}$ either if the prediction step was shorter or if the size of observational noise was smaller (see Fig. 12 and Table II).

E. Coupled map lattice without noise

Moreover, I used the coupled map lattice [17] for evaluating the proposed index. I used the following model:

$$\begin{aligned} u_n(t+1) &= (1-2\varepsilon)\{a u_n(t)[1-u_n(t)]\} \\ &\quad + (\varepsilon-\eta)\{a u_{n+1}(t)[1-u_{n+1}(t)]\} \\ &\quad + (\varepsilon+\eta)\{a u_{n-1}(t)[1-u_{n-1}(t)]\}, \\ u_{n+N}(t) &= u_n(t), \end{aligned}$$

where I set $N = 100$, $a = 3.8$, $\varepsilon = 0.05$, $\eta = 0.01$. I generated a time series with the length of 10 000 by observing $u_n(n = 1,2,\ldots,20)$. A part of the time series is shown in Fig. 13. I predicted their sum. The other conditions are the same as the case of the Lorenz'96 II model. The results are shown in Fig. 14. Although I found that the predicted 96% confidence interval tended to contain the actual value more frequently for the wide range of prediction steps for $D_t < \bar{D}$, this tendency was not statistically significant for most of the prediction steps.

F. Coupled map lattice with observational noise

I also evaluated the case where there is observational noise. I took the average over ten time series generated from different initial conditions. The results are shown in Fig. 15. For the coupled map lattice, when prediction steps were less than or equal to 10 and the noise level was less than or equal to 0.05, the probability that their difference is 5% significant was a chance level (Table III). In addition, the difference between $D_t < \bar{D}$ and $D_t \geqslant \bar{D}$ does not depend on either prediction steps or the size of observational noise (Table III).

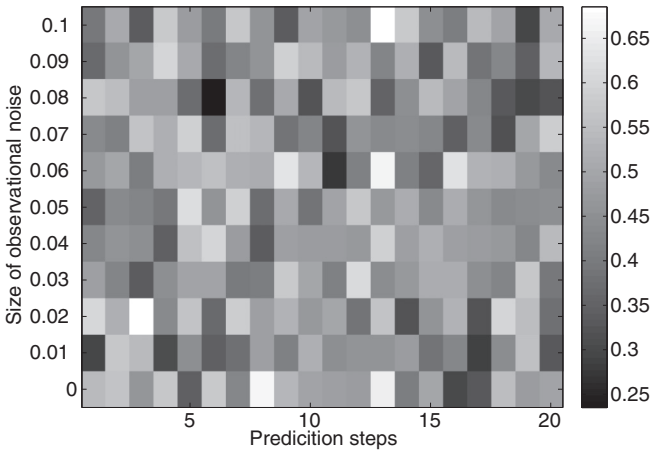

FIG. 15. The average significance level for the difference between low and high D_t cases given the size of observational noise and a prediction step, in the case of the coupled map lattice.

G. Solar irradiation around Japan

I applied the proposed index to solar irradiation data observed around Japan. The dataset was obtained from 48 weather stations across Japan and covered January 1, 2010 to December 31, 2012. Although the dataset was originally recorded every 10 min, I took the average over every hour. Therefore, the dataset used for the prediction was a 48-dimensional time series with a length of 26 304. A part of the time series is shown in Fig. 16.

I set $\beta = 9000$ to take into account the seasonality. I chose $d = 18$ so that the current state can identify the time within a day. I set $q = 36$ because this is the maximum target prediction step in a practical power grid system. In addition, I set $K = 25$ to obtain the 96% confidence intervals for the sum over the 48 dimensions. I used 4000 points starting at the 9001st point to obtain $\bar{D}$, and the following 13 000 points to evaluate the proposed method.

Examples of the predicted 96% confidence intervals are shown in Fig. 17. In addition, the results are summarized in Fig. 18. If $D_t < \bar{D}$, the confidence intervals tend to contain the actual values more frequently than if $D_t \geqslant \bar{D}$ when the

TABLE III. Dependence of 5% significance on prediction steps and levels of observational noise, for the example of coupled map lattice. See the caption of Table I to interpret this table. In this table, the condition that prediction steps $\leqslant 10$ is statistically independent of the 5% significance (p-value: 1.00). The condition that noise level $\leqslant 0.05$ is statistically independent of the 5% significance (p-value: 0.78).

Condition–significance level				
Prediction steps	Noise level	$\leqslant 0.05$	>0.05	Total
$\leqslant 10$	$\leqslant 0.25$	42(7%)	558(93%)	600
	>0.25	47(8%)	553(92%)	600
>10	$\leqslant 0.25$	27(5%)	473(95%)	500
	>0.25	55(11%)	445(89%)	500
Total		171(8%)	2029(92%)	2200

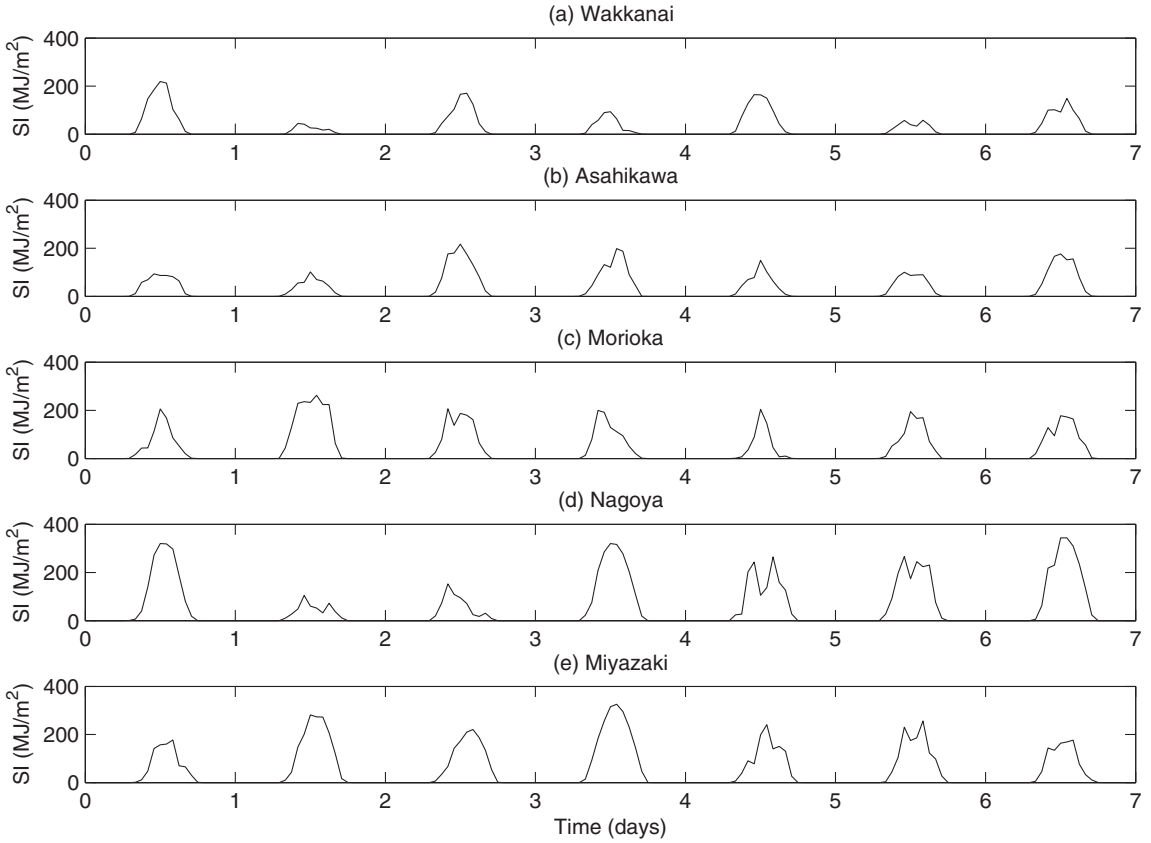

FIG. 16. A part of the time series for solar irradiation.

prediction steps were between 1 and 5 h, between 11 and 22 h, and for 35 and 36 h [chi-square test; see Fig. 18(b)]. For instance, in the examples in Figs. 17(b), 17(f), and 17(h), the actual value exceeded the confidence interval at the prediction steps for at least one prediction step. Moreover, if $D_t < \bar{D}$, the 96% confidence intervals were likely to contain

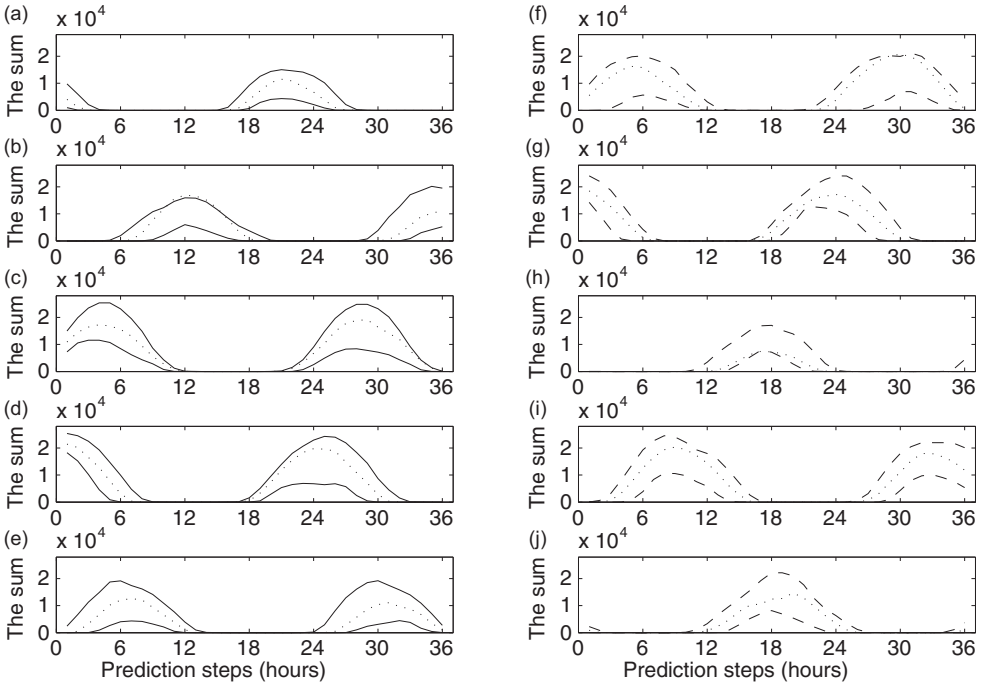

FIG. 17. Examples of predicted 96% confidence intervals for sums of solar irradiation observed at 48 weather stations. (a)–(e) $D_t < \bar{D}$; (f)–(j) $D_t \geqslant \bar{D}$. Lines have same meaning as in Fig. 5.

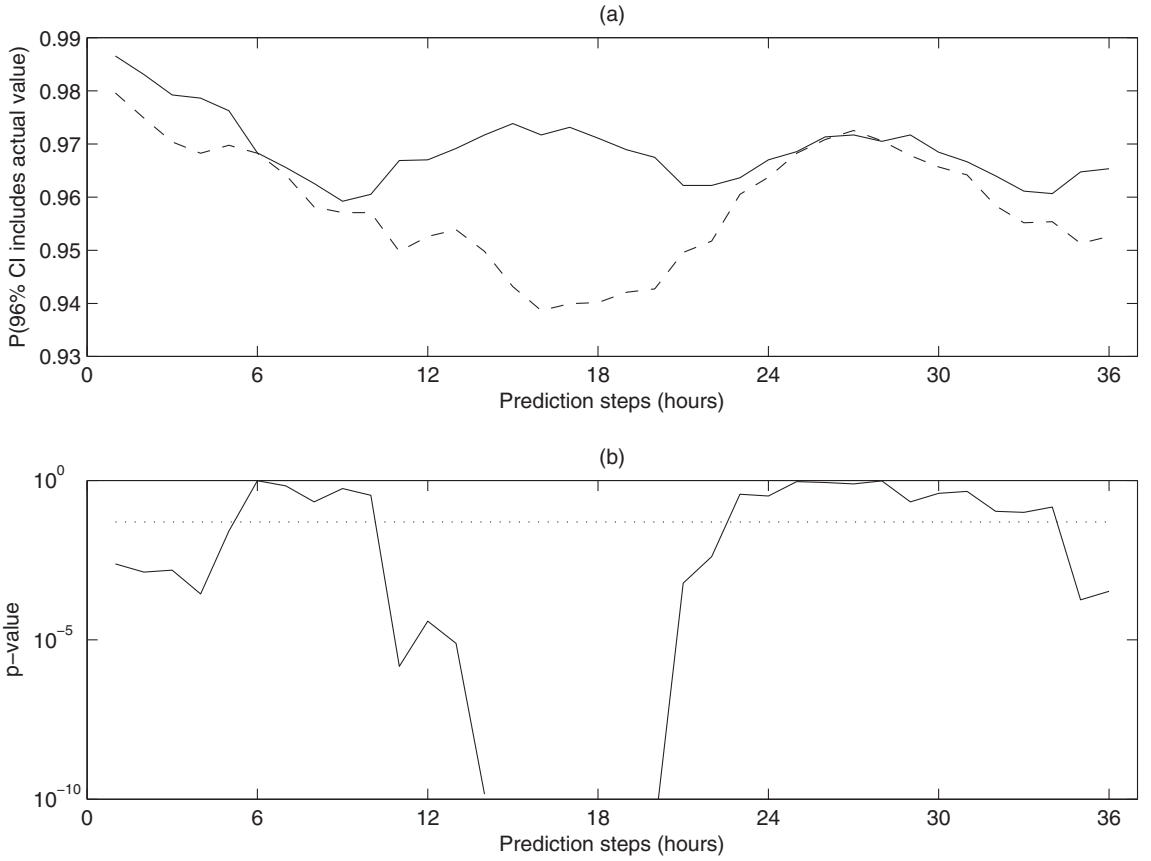

FIG. 18. Comparison of cases where the proposed index was small and large in real-time-series solar irradiation data. I used $\lambda = 1$ in this example. Lines have the same meaning as in Fig. 6.

the actual values with a probability of about 96% or higher [see Fig. 18(a)]. It took 1.4 days to finish the above prediction for the dataset covering 3 years.

V. DISCUSSION

I compared the proposed index with the local maximum Lyapunov exponent. Abarbanel *et al.* [18,19] proposed local Lyapunov exponents. I used the method of Kantz [20] to evaluate the maximum Lyapunov exponent locally by evaluating the maximum exponent using nine nearest neighbors and prediction steps of 2–4. The results for the Lorenz'96 II model, the Lorenz'96 I model, the coupled map lattice, and the solar irradiation data are shown in Figs. 19–22, respectively. These results show that the local maximum Lyapunov exponent and the proposed index show the different aspects for the predictability of a given time series. For instance, in the example of the Lorenz'96 II model, the proposed index identified the difference for the prediction skill when the prediction steps were less than 9 (Fig. 6), while the local maximum Lyapunov exponent identified the difference when the prediction steps were between 8 and 16 (Fig. 19). In the example of the Lorenz'96 I model, the difference in the prediction skill was identified by the proposed index for the prediction steps less than 7 (Fig. 10), while that by the local maximum Lyapunov exponent was for the prediction steps of less than 5 and between 13 and 14 (Fig. 20). Therefore, the local maximum Lyapunov exponent and the proposed index can possibly work complementarily. How to combine these two indices is a topic of future research.

Although the proposed index seems to be robust against the observational noise, I could not draw a general conclusion on how small the observational noise should be because the performance was different depending on the tested mathematical models (see Figs. 8, 12, and 15 and Tables I–III).

I examined the effects of λ. Figures 23 and 24 show the results for $\lambda = 0.1$ and $\lambda = 10$, respectively, for Lorenz'96 II without observational noise. These results are similar to those in Fig. 6, where $\lambda = 1$. Therefore, the proposed algorithm seems to be robust at least in the range of λ between 0.1 and 10.

As discussed in Ref. [2], the quality of the confidence intervals depends on the size of the database β and the prediction steps p. The appropriate selection of these parameters for a given set of experimental data is another topic for future research.

In the solar irradiation example, the 96% confidence intervals tended to contain the actual values with a probability of about 96% or higher when $D_t < \bar{D}$ [see Fig. 18(a)]. This information might be useful for planning when and how to start thermal power plants and hydroelectric power plants for backups because the proposed index can tell whether the predicted confidence intervals are reliable. This direction is also a topic of future research.

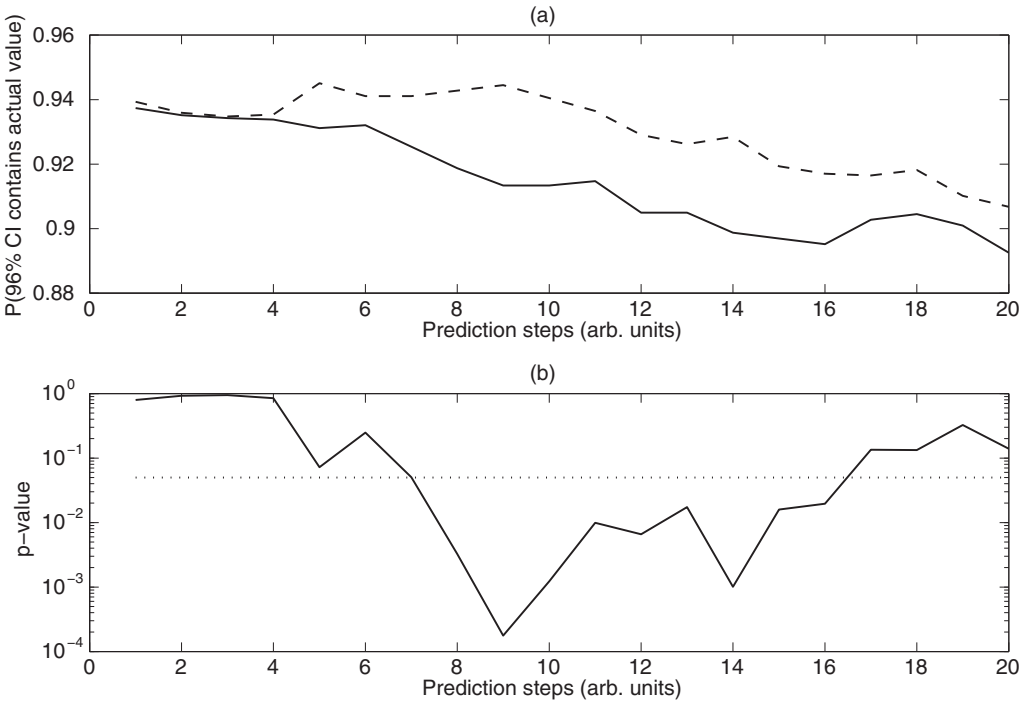

FIG. 19. Comparison of cases where local maximum Lyapunov exponent was small and large for the Lorenz'96 II model. (a) The probabilities that 96% confidence interval includes the actual value are shown for low (solid line) and high (dashed line) local maximum Lyapunov exponent, respectively. (b) The p-value for the difference given each prediction step is shown by the solid line. The dotted line shows the significance level of 5%.

A credibility measure for time-series prediction was previously proposed in Ref. [21]. The proposed measure is more appropriate from the viewpoint of nonlinear dynamics because the proposed index eventually uses the property that if the current state is far from its neighboring points in the database, they cannot shadow the current trajectory well (see Fig. 25 for

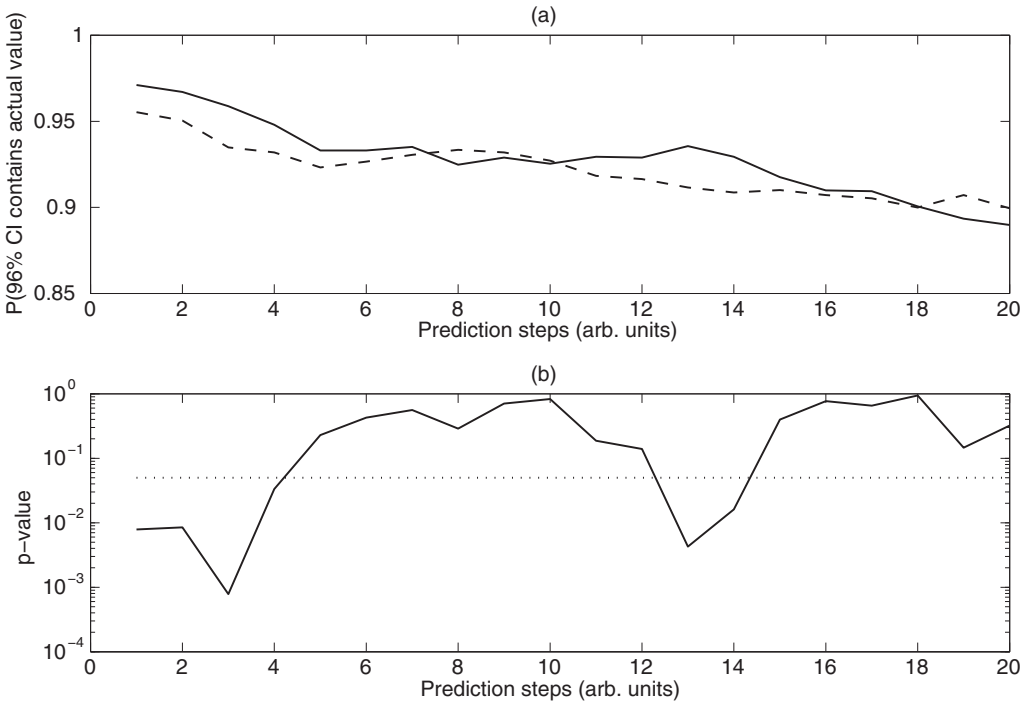

FIG. 20. Comparison of cases where the local maximum Lyapunov exponent was small and large for the Lorenz'96 I model. See the caption of Fig. 19 to interpret the results.

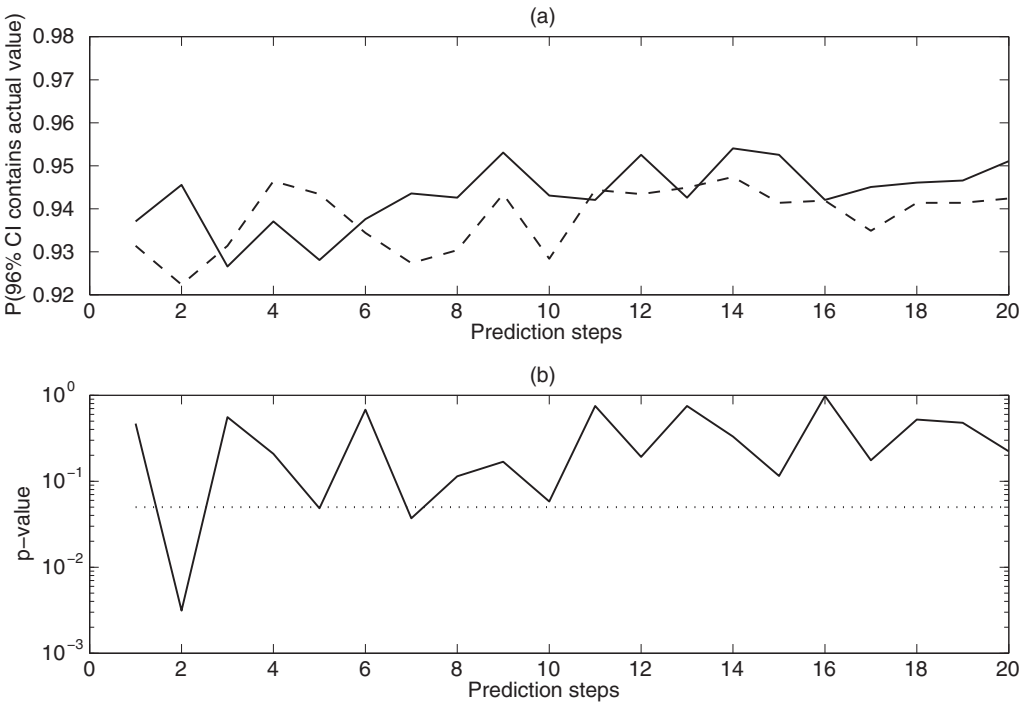

FIG. 21. Comparison of cases where the local maximum Lyapunov exponent was small and large for the coupled map lattice. See the caption of Fig. 19 to interpret the results.

a schematic diagram). Therefore, the proposed method can be applied even if one assumes that the underlying dynamics is nonlinear.

Although I proposed the index D_t for the prediction algorithm in Refs. [1,2], this index can be applied with another prediction algorithm because the index evaluates the difference between the current state and the known past data. In addition, the index can be applied to any time-series data in any field. My finding is consistent with the observation in Ref. [4] that a transient time series does not help much to construct a mathematical model for prediction.

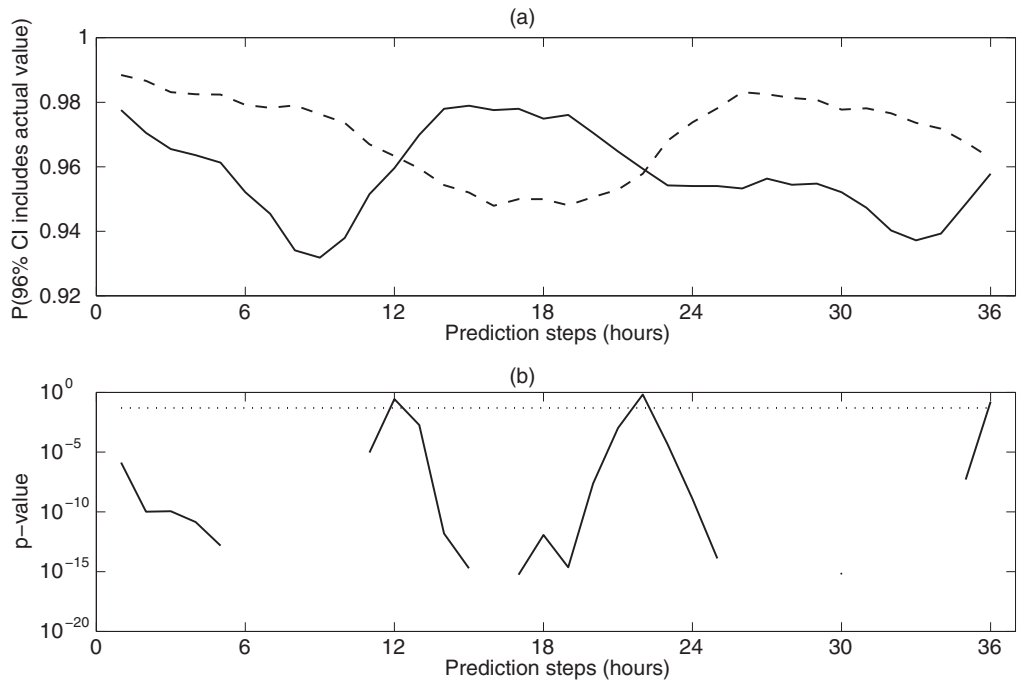

FIG. 22. Comparison of cases where the local maximum Lyaponov exponent was small and large for solar irradiation data. See the caption of Fig. 19 to interpret the results.

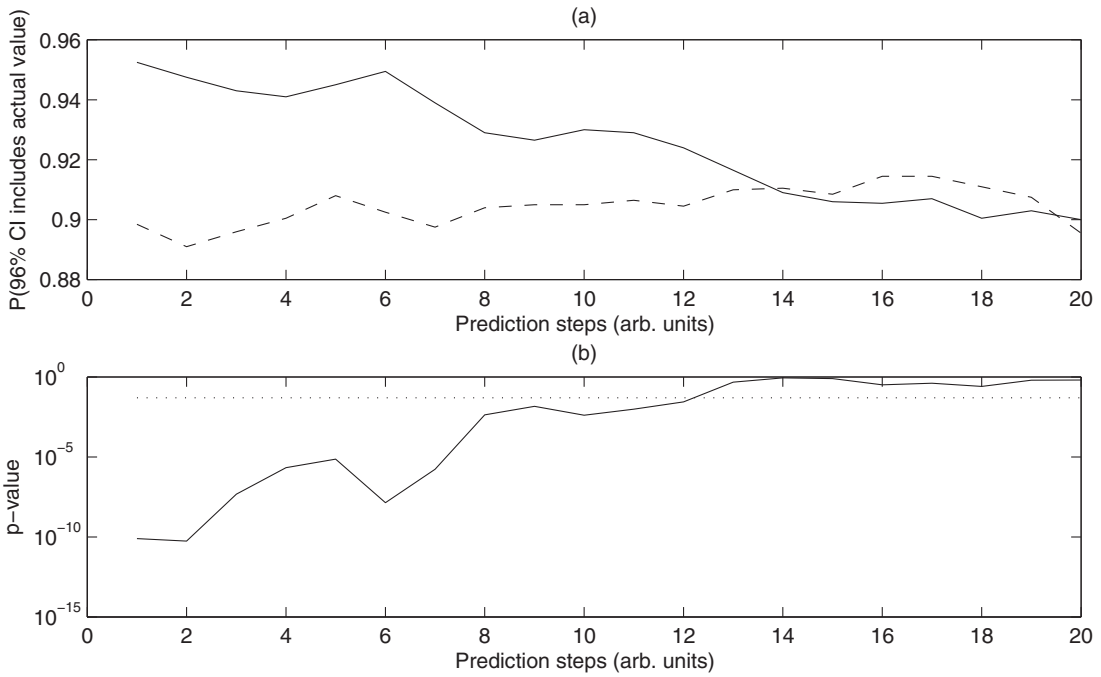

FIG. 23. Effect of λ. Results for the Lorenz'96 II model without observational noise when $\lambda = 0.1$. Lines have the same meaning as in Fig. 6.

I noticed the approach of data assimilation in the context of weather forecasting [22], but because its computational demands are huge, the proposed framework is more appropriate for predictions for shorter time scales such as minutes and hours.

After submitting the initial draft, I noticed the similarity between the current work with the work of Judd *et al.* [23], which was proposed in the context of weather forecasting. However, there is a clear distinction between the current work and the work of Judd *et al.* [23] because I evaluated the distance between the current state and the data manifold directly, while Judd *et al.* [23] evaluated the distance between the initial state and the attractor of the weather model by minimizing the indeterminism using the model and shadowing. This similarity also implies that I may be able to do better if I introduce some metric in the neighborhood of the current state in evaluating the

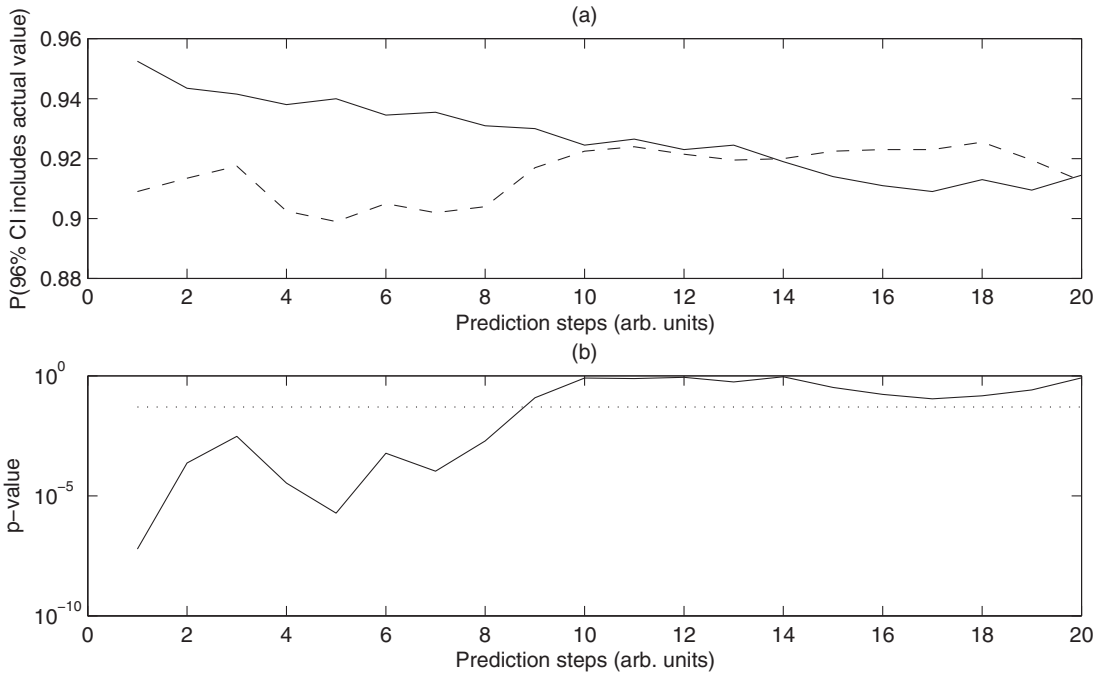

FIG. 24. Same as Fig. 23 except for $\lambda = 10$.

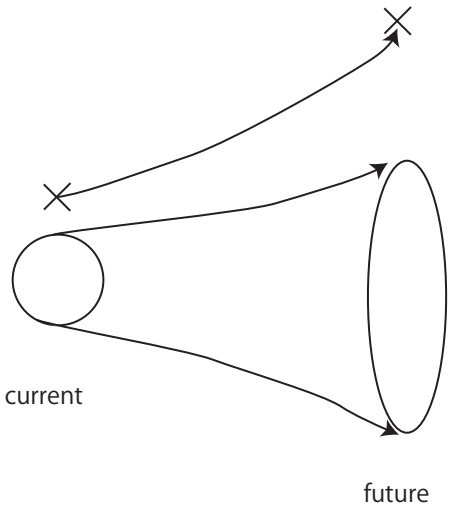

FIG. 25. Schematic diagram of current state (left cross), its neighbors (circle), future state for current state (right cross), and future states of neighbors (ellipse). If the current state is far from its neighbors, its future state will probably be far from those of the neighbors.

credibility for the predictability. But this direction is beyond the scope of this paper.

In sum, I proposed an index for evaluating the credibility of predicted confidence intervals. The index was constructed on the basis of the distance between the current state and the data manifold. Using the Lorenz'96 II model, the Lorenz'96 I model, and the coupled map lattice, and a real dataset for Japanese solar irradiation as examples, I demonstrated that the confidence intervals are more reliable when the distance is shorter. This index can provide information that is complementary to the local maximum Lyapunov exponent. Thus I hope that the proposed index facilitates the introduction of more renewable energy resources in power grid systems.

ACKNOWLEDGMENTS

I thank the Japan Meteorological Agency for providing the solar irradiation dataset used in this study. This research was supported by the Aihara Innovative Mathematical Modelling Project, Japanese Society for the Promotion of Science (JSPS), through its "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)," initiated by the Council for Science and Technology Policy (CSTP).

[1] Y. Hirata, T. Yamada, J. Takahashi, and H. Suzuki, Phys. Lett. A **376**, 3092 (2012).
[2] Y. Hirata, T. Yamada, J. Takahashi, K. Aihara, and H. Suzuki, Renew. Energy **67**, 35 (2014).
[3] K. Judd and M. Small, Physica D **136**, 31 (2000).
[4] B. P. Bezruchko, T. V. Dikanev, and D. A. Smirnov, Phys. Rev. E **64**, 036210 (2001).
[5] D. A. Smirnov, B. P. Bezruchko, and Y. P. Seleznev, Phys. Rev. E **65**, 026205 (2002).
[6] M. Small and C. K. Tse, Phys. Rev. E **66**, 066701 (2002).
[7] D. A. Smirnov, V. S. Vlaskin, and V. I. Ponomarenko, Phys. Lett. A **336**, 448 (2005).
[8] B. P. Bezruchko and D. A. Smirnov, *Extracting Knowledge from Time Series* (Springer, Heidelberg, Germany, 2010).
[9] F. Kwasniok and L. A. Smith, Phys. Rev. Lett. **92**, 164101 (2004).
[10] T. Yamada, J. Takahashi, and K. Aihara, J. Reliab. Eng. Assoc. Jpn. **28**, 489 (2006) (in Japanese).
[11] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 2004).
[12] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, *The State-of-the-Art in Short-Term Prediction of Wind Power: A Literature Overview*, 2nd ed., Specific targeted research project, Contract No. 038692 (ANEMOS.plus, January, 2011).
[13] K. Porter and J. Rogers, National Renewable Energy Laboratory Report No. NREL/SR-5500-54457, April 2012.
[14] E. N. Lorenz, in *Proceedings of the Seminar on Predictability*, Vol. 1 (ECMWF, Reading, UK, 1996), p. 1.
[15] J. A. Hansen and L. A. Smith, J. Atmos. Sci. **57**, 2859 (2000).
[16] M. H. DeGroot and M. J. Schervish, *Probability and Statistics* (Addison-Wesley, Boston, 2002).
[17] K. Kaneko, Prog. Theor. Phys. **72**, 480 (1984).
[18] H. D. I. Abarbanel, R. Brown, and M. B. Kennel, J. Nonlinear Sci. **1**, 175 (1991).
[19] H. D. I. Abarbanel, R. Brown, and M. B. Kennel, J. Nonlinear Sci. **2**, 343 (1992).
[20] H. Kantz, Phys. Lett. A **185**, 77 (1994).
[21] H. Shen, H. Hino, N. Murata, and S. Wakao, in *Proceedings of 2011 10th International Conference on Machine Learning and Applications*, Vol. 2 (IEEE, Piscataway, NJ, 2011), pp. 275–280.
[22] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge University Press, Cambridge, UK, 2003).
[23] K. Judd, C. A. Reynolds, T. E. Rosmond, and L. A. Smith, J. Atmos. Sci. **65**, 1749 (2008).