Probabilistic Prediction of Chaotic Time Series Using Similarity of Attractors and LOOCV Predictable Horizons for Obtaining Plausible Predictions

著者	Kurogi S., Toidani M., Shigematsu R., Matsuo K.
journal or publication title	Lecture Notes in Computer Science
volume	9491
page range	72-81
year	2015-12-09
URL	http://hdl.handle.net/10228/5935

doi: info:doi/10.1007/978-3-319-26555-1_9

Probabilistic Prediction of Chaotic Time Series Using Similarity of Attractors and LOOCV Predictable Horizons for Obtaining Plausible Predictions

Shuichi Kurogi(✉), Mitsuki Toidani, Ryosuke Shigematsu, and Kazuya Matsuo

Kyushu Institute of Technology, Tobata, Kitakyushu, Fukuoka 804-8550, Japan
{kuro,matsuo}@cntl.kyutech.ac.jp,
{toidani,shigematsu}@kurolabcntl.kyutech.ac.jp
http://kurolab.cntl.kyutech.ac.jp/

kurolab.cntl

Abstract. This paper presents a method for probabilistic prediction of chaotic time series. So far, we have developed several model selection methods for chaotic time series prediction, but the methods cannot estimate the predictable horizon of predicted time series. Instead of using model selection methods employing the estimation of mean square prediction error (MSE), we present a method to obtain a probabilistic prediction which provides a prediction of time series and the estimation of predictable horizon. The method obtains a set of plausible predictions by means of using the similarity of attractors of training time series and the time series predicted by a number of learning machines with different parameter values, and then obtains a smaller set of more plausible predictions with longer predictable horizons estimated by LOOCV (leave-one-out cross-validation) method. The effectiveness and the properties of the present method are shown by means of analyzing the result of numerical experiments.

Keywords: Probabilistic prediction · Attractors of chaotic time series · Leave-one-out cross-validation · Prediction of time series · Estimation of predictable horizon

1 Introduction

This paper presents a method of probabilistic prediction of chaotic time series. So far, we have developed several model selection methods for chaotic time series prediction [1,2]. The method in [1] uses moments of predictive deviation as ensemble diversity measures for model selection in time series prediction, and achieves better performance from the point of view of mean square prediction error (MSE) than the conventional holdout method. The method in [2] uses direct multi-step ahead (DMS) prediction to apply the out-of-bag (OOB) estimate of the MSE. However, both methods cannot be used for estimating the predictable horizons of predicted time series but for estimating the MSE, which is owing

S. Arik et al. (Eds.): ICONIP 2015, Part III, LNCS 9491, pp. 1–10, 2015.
DOI: 10.1007/978-3-319-26555-1_9

mainly to the fact that the MSE is affected by short term predictability and long term unpredictability of chaotic time series (see [2] for the analysis and [3] for properties of chaotic time series).

Instead of using model selection methods employing the estimation of the MSE, we present a method to obtain probabilistic prediction. Here, from [4], we can see that the probabilistic prediction has come to dominate the science of weather and climate forecasting. This is mainly because the theory of chaos at the heart of meteorology shows that for a simple set of nonlinear equations (or Lorenz's equations shown below) with initial conditions changed by minute perturbations, there is no longer a single deterministic solution and hence all forecasts must be treated as probabilistic. From another perspective, the forecast probabilities allow the user to take appropriate action within a proper understanding of the uncertainties. For probabilistic weather forecast, a number of ensemble methods have been employed and progressed, where the forecast uncertainties arised from perturbations to the initial conditions and model parameters are examined.

In this article, we try to utilize learning machines for probabilistic prediction, which indicates that forecast uncertainty arises from model uncertainty. We also try to obtain a set of plausible predictions by means of using the similarity of training time series and the time series predicted by a number of learning machines with different parameter values. Furthermore, we introduce LOOCV (leave-one-out cross-validation) method for estimating the predictable horizon to obtain a smaller set of more plausible predictions. We show the method of probabilistic prediction in Sect. 2, experimental results and analysis in Sect. 3, and the conclusion in Sect. 4.

2 Probabilistic Prediction of Chaotic Time Series

2.1 Point Prediction of Chaotic Time Series

Let $y_t (\in \mathbb{R})$ denote a chaotic time series for a discrete time $t = 0, 1, 2, \cdots$ satisfying

$$y_t = r(\boldsymbol{x}_t) + e(\boldsymbol{x}_t), \tag{1}$$

where $r(\boldsymbol{x}_t)$ is a nonlinear target function of a vector $\boldsymbol{x}_t = (y_{t-1}, y_{t-2}, \cdots, y_{t-k})^T$ for the embedding dimension k generated by the delay embedding from a chaotic differential dynamical system (see [3] for the theory of chaotic time series). Here, y_t is obtained not analytically but numerically, and then y_t involves an error $e(\boldsymbol{x}_t)$ owing to an executable finite calculation precision. This indicates that there are a number of plausible target functions $r(\boldsymbol{x}_t)$ with allowable error $e(\boldsymbol{x}_t)$. Furthermore, the time series y_t for numerical experiments shown below is one of the time series generated from the original chaotic dynamical system with a high precision, which we denote ground truth time series $y_t^{[\mathrm{gt}]}$ depending on the necessity in the context.

Let $y_{t:h} = y_t y_{t+1} \cdots y_{t+h-1}$ denote a time series with the initial time t and the horizon h. For a given and training time series $y_{t_\mathrm{g}:h_\mathrm{g}} (= y^{[\mathrm{train}]}_{t_\mathrm{g}:h_\mathrm{g}})$, we are supposed to predict succeeding time series $y_{t_\mathrm{p}:h_\mathrm{p}}$ for $t_\mathrm{p} \geq t_\mathrm{g} + h_\mathrm{g}$. Then, we make the training dataset $D^{[\mathrm{train}]} = \{(\boldsymbol{x}_t, y_t) \mid t \in I^{[\mathrm{train}]}\}$ for $I^{[\mathrm{train}]} = \{t \mid t_\mathrm{g} \leq t < t_\mathrm{g} + h_\mathrm{g}\}$ to train a learning machine. After the learning, the machine executes iterated prediction by

$$\hat{y}_t = f(\hat{\boldsymbol{x}}_t) \tag{2}$$

for $t = t_\mathrm{p}, t_{\mathrm{p}+1}, \cdots$, recursively, where $f(\hat{\boldsymbol{x}}_t)$ denotes the prediction function of $\hat{\boldsymbol{x}}_t = (x_{t1}, x_{t2}, \cdots, x_{tk})$ whose elements are given by

$$x_{tj} = \begin{cases} y_{t-j} \ (t - j < t_\mathrm{p}) \\ \hat{y}_{t-j} \ (t - j \geq t_\mathrm{p}). \end{cases} \tag{3}$$

Here, we suppose that y_t for $t < t_\mathrm{p}$ is known for making the prediction $\hat{y}_{t_p:h_p}$ as the initial state.

2.2 Probabilistic Prediction

For probabilistic prediction, we firstly make a number of predictions $\hat{y}_{t_p:h_p}$ or $y^{[\theta_m]}_{t_p:h_p}$ generated from (2) and (3) by means of learning machines with parameter values $\theta_m \in \Theta$, where Θ indicates the set of parameter values of learning machines. Here, we suppose that there are a number of plausible prediction functions $f(\cdot) = f^{[\theta_m]}(\cdot)$ for the chaotic time series, and we have to remove implausible ones. To have this done, we select the following set of plausible predictions $y^{[\theta_m]}_{t_p:h_p}$,

$$Y^{[S_\mathrm{th}]}_{t_p:h_p} = \left\{ y^{[\theta_m]}_{t_p,h_p} \;\middle|\; S\left(y^{[\theta_m]}_{t_p,h_p}, y_{t_g:h_g}\right) \geq S_\mathrm{th} \right\} \tag{4}$$

where

$$S\left(y^{[\theta_m]}_{t_p,h_p}, y^{[\mathrm{train}]}_{t_g:h_g}\right) \triangleq \frac{\sum_i \sum_j a^{[\theta_m]}_{ij} a^{[\mathrm{train}]}_{ij}}{\sqrt{\sum_i \sum_j \left(a^{[\theta_m]}_{ij}\right)^2} \sqrt{\sum_i \sum_j \left(a^{[\mathrm{train}]}_{ij}\right)^2}} \tag{5}$$

denotes the similarity of two-dimensional attractor (trajectory) distributions $a^{[\theta_m]}_{ij}$ and $a^{[\mathrm{train}]}_{ij}$ of time series $y^{[\theta_m]}_{t_p,h_p}$ and $y^{[\mathrm{train}]}_{t_g:h_g}$, respectively, and S_th is a threshold. Here, the two-dimensional attractor distribution, a_{ij}, of a time-series $y_{t:h}$ is given by

$$a_{ij} = \sum_{s=t}^{t+h-1} \mathbf{1}\left\{ \left\lfloor \frac{y_s - v_0}{\Delta_a} \right\rfloor = i \wedge \left\lfloor \frac{y_{s+1} - v_0}{\Delta_a} \right\rfloor = j \right\}, \tag{6}$$

where v_0 is a constant less than the minimum value of y_t for all time series and Δ_a indicates a resolution of the distribution. Furthermore, $\mathbf{1}\{z\}$ is an indicator function equal to 1 if z is true, and to 0 if z is false, and $\lfloor\cdot\rfloor$ indicates the floor function.

Next, we try to select predictions with longer predictable horizons from the plausible predictions $Y^{[S_{\rm th}]}_{t_p:h_p}$. To have this done, let us calculate the predictable horizon between two predictions $y^{[\theta_m]}_{t_p:h_p}$ and $y^{[\theta_n]}_{t_p:h_p}$ in $Y^{[S_{\rm th}]}_{t_p:h_p}$ given by

$$h\left(y^{[\theta_m]}_{t_p:h_p}, y^{[\theta]}_{t_p:h_p}\right) = \max\left\{h \mid \forall s \le h \le h_p; |y^{[\theta_m]}_{t_p+s} - y^{[\theta_n]}_{t_p+s}| \le e_y\right\} \qquad (7)$$

where e_y indicates the threshold of prediction error to determine the horizon. Then, we have the mean of $h\left(y^{[\theta_m]}_{t_p:h_p}, y^{[\theta_n]}_{t_p:h_p}\right)$ between an prediction $y^{[\theta_m]}_{t_p:h_p}$ and the other predictions $y^{[\theta_n]}_{t_p:h_p}$ in the set of plausible predictions, $Y^{[S_{\rm th}]}_{t_p:h_p}$, given by

$$\tilde{h}^{[\theta_m]}_{t_p} = \frac{1}{\left|Y^{[S_{\rm th}]}_{t_p:h_p}\right| - 1} \sum_{y^{[\theta_n]}_{t_p:h_p} \in Y^{[S_{\rm th}]}_{t_p:h_p} \backslash \{y^{[\theta_m]}_{t_p:h_p}\}} h\left(y^{[\theta_m]}_{t_p:h_p}, y^{[\theta_n]}_{t_p:h_p}\right). \qquad (8)$$

where $\left|Y^{[S_{\rm th}]}_{t_p:h_p}\right|$ denotes the number of elements in $Y^{[S_{\rm th}]}_{t_p:h_p}$. Note that the above method to obtain the horizon $\tilde{h}^{[\theta_m]}_{t_p}$ is based on the leave-one-out cross-validation (LOOCV), and the prediction with a longer horizon $\tilde{h}^{[\theta_m]}_{t_p}$ is considered to provide a "better" prediction in predicting all plausible predictions $y^{[\theta]}_{t_p:h_p} \in Y^{[S_{\rm th}]}_{t_p:h_p}$ on average. Thus, we sort the predictable horizons by their lengths as $\tilde{h}^{[\theta_{\sigma(i)}]}_{t_p} \ge \tilde{h}^{[\theta_{\sigma(i+1)}]}_{t_p}$, where $\sigma(i)$ denotes the order for $i = 1, 2, \cdots, |Y^{[S_{\rm th}]}_{t_p:h_p}|$. Let a subset of plausible predictions with longer predictable horizons be

$$Y^{[H_{\rm th}, S_{\rm th}]}_{t_p:h_p} = \left\{ y^{[\theta_{\sigma(i)}]}_{t_p:h_p} \;\middle|\; \frac{i}{|Y^{[S_{\rm th}]}_{t_p:h_p}|} \le H_{\rm th} \right\}, \qquad (9)$$

where the threshold $H_{\rm th}$ ($0 < H_{\rm th} \le 1$) indicates the ratio of the numbers of elements in $Y^{[H_{\rm th}, S_{\rm th}]}_{t_p:h_p}$ and $Y^{[S_{\rm th}]}_{t_p:h_p}$, or $H_{\rm th} = \left|Y^{[H_{\rm th}, S_{\rm th}]}_{t_p:h_p}\right| / \left|Y^{[S_{\rm th}]}_{t_p:h_p}\right|$. Now, we derive the probability $p_(y_t)$ of the prediction y_t for $t_p \le t < t_p + h_p$ as

$$p\left(v_i \le y_t < v_{i+1}\right) = \frac{1}{\left|Y^{[H_{\rm th},, S_{\rm th}]}_{t_p:h_p}\right|} \sum_{\theta \in \Theta^{[H_{\rm th},, S_{\rm th}]}} \mathbf{1}\left\{ \left\lfloor \frac{y^{[\theta]}_t - v_0}{\Delta_v} \right\rfloor = i \right\}, \qquad (10)$$

where $\Theta^{[H_{\rm th},, S_{\rm th}]}$ is the set of parameters θ of learning machines which have generated the time series $y^{[\theta]}_{t_p:h_p} \in Y^{[H_{\rm th},, S_{\rm th}]}_{t_p:h_p}$, and Δ_v is a constant representing the resolution of y_t, and $v_i = i\Delta_v + v_0$ for $i = 0, 1, 2, \cdots$. Note that the probability $p(y_t)$ depends on the threshold $H_{\rm th}$. Namely, the probability $p(v_i \le y_t \le v_{i+1})$ indicates how much the plausible predictions in $Y^{[H_{\rm th}, S_{\rm th}]}_{t_p:h_p}$ take the values in between

v_i and v_{i+1}, where $Y^{[H_{\rm th},S_{\rm th}]}_{t_p:h_p}$ consists of $H_{\rm th} \times |Y^{[S_{\rm th}]}_{t_p:h_p}|$ predictions with longer predictable horizons among all plausible predictions in $Y^{[S_{\rm th}]}_{t_p:h_p}$. As a special case for $H_{\rm th} = \left|Y^{[S_{\rm th}]}_{t_p:h_p}\right|^{-1}$, we have $Y^{[H_{\rm th},S_{\rm th}]}_{t_p:h_p}$ consisting of only one prediction $y^{[\theta_{\sigma(1)}]}_{t_p:h_p}$ with the longest predictable horizon $\tilde{h}^{[\theta]}_{t_p}$ which we assume the most plausible prediction and we call it representative prediction among plausible predictions. By means of the above probabilistic prediction, we hope that the representative prediction $y^{[\theta_{\sigma(1)}]}_{t_p:h_p}$ has a longer predictable horizon in predicting the ground truth time series $y^{[\rm gt]}_t$. Furthermore, we would like to tune $H_{\rm th}$ $\left(> \left|Y^{[S_{\rm th}]}_{t_p:h_p}\right|^{-1}\right)$ to provide more conservative and safe probabilistic prediction so that $p(y_t)$ may have a positive value for the ground truth $y_t = y^{[\rm gt]}_t$. Then, after the tuning of $H_{\rm th}$, we can provide an expected predictable horizon $\hat{h}^{[\theta_{\sigma(1)}]}_{t_p}$ of the representative prediction $y^{[\theta_{\sigma(1)}]}_{t_p:h_p}$ in predicting the ground truth time series $y^{[\rm gt]}_{t_p:h_p}$ by

$$\hat{h}^{[\theta_{\sigma(1)}]}_{t_p} = \max\left\{h \;\middle|\; \forall s \le h, \forall y^{[\theta]}_{t_p:h_p} \in Y^{[H_{\rm th},S_{\rm th}]}_{t_p:h_p}; |y^{[\theta_{\sigma(1)}]}_{t_p+s} - y^{[\theta]}_{t_p+s}| \le e_y\right\}. \tag{11}$$

3 Numerical Experiments and Analysis

3.1 Experimental Settings

We use the Lorenz time series, as shown in Fig. 1 and [2], obtained from the original differential dynamical system given by

$$\frac{dx_c}{dt_c} = -\sigma x_c + \sigma y_c, \quad \frac{dy_c}{dt_c} = -x_c z_c + r x_c - y_c, \quad \frac{dz_c}{dt_c} = x_c y_c - b z_c, \tag{12}$$

for $\sigma = 10$, $b = 8/3$, $r = 28$. Here, we use t_c for continuous time and t $(= 0, 1, 2, \cdots)$ for discrete time related by $t_c = tT$ with sampling time T. We have generated the time series $y(t) = x_c(tT)$ for $t = 1, 2, \cdots, 5,000$ from the initial state $(x_c(0), y_c(0), z_c(0)) = (-8, 8, 27)$ with $T = 25$ms via Runge-Kutta method with 128 bit precision of GMP (GNU multi-precision library). As a result of preliminary experiments as shown in [2], $y(t)$ for each duration of time less than 1,200 steps (= 30 s/25 ms) in Fig. 1, or $y_{t_0:1200}$ for each initial time $t_0 = 0, 1, 2, \cdots$ with initial state $(x(t_0), y(t_0), z(t_0))$, is supposed to be correct, while cumulative computational error may increase exponentially after the duration.

We use $y_{0:2000}$ for training a learning machine, and execute multistep prediction of $y_{t_{\rm p}:h_{\rm p}}$ with the initial input vector $\boldsymbol{x}_{t_{\rm p}} = (y(t_{\rm p}-1), \cdots, y(t_{\rm p}-k))$ for prediction start time $t_{\rm p} = 2000 + 100i$ $(i = 0, 1, 2, \cdots, 19)$. As a learning machine, we use CAN2 (see A and [5] for details), where the model complexity is the number of units, N, or the number of piecewise linear regions for approximating the target function $r(\cdot)$. We show the results with the embedding dimension $k = 8$ and we use the parameter of the learning machines as $\theta = N$. We set the thresholds as $S_{\rm th} = H_{\rm th} = 0.8$.

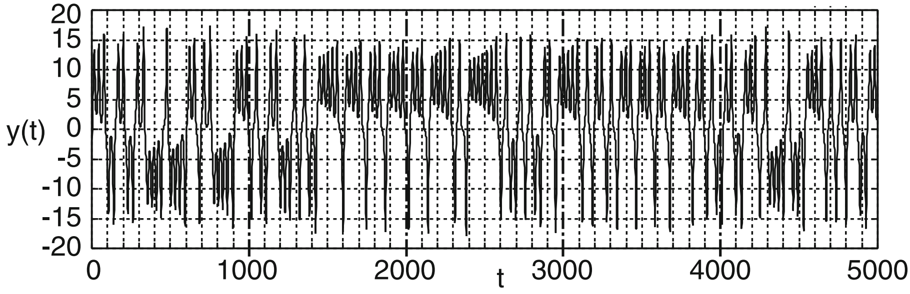

Fig. 1. Lorenz time series $y(t)$ for $t = 0, 1, 2, \cdots, 4999$, or the ground truth time series $y^{[\mathrm{gt}]}_{0:5000}$.

For probabilistic prediction, we firstly make a number of predictions $y^{[\theta]}_{t_p:h_p}$ by means of using CAN2s with different number of units as $\theta = N = 10, 12, \cdots, N_{\max}$, where we use $N_{\max} = 250 = 2000/k$, because CAN2 employs N piecewise linear regions and each piecewise linear region requires more than k independent data from $2000 - k$ data in the training dataset $D^{[\mathrm{train}]} = \{(\boldsymbol{x}_t, y_t) | t = k = 8, k+1, \cdots, 1999\}$.

3.2 Results and Analysis

In order to intuitively see how the present method works, we show examples of obtained representative prediction $y^{[\theta_{\sigma(1)}]}_{t_p:h_p}$ and the probability $p(y_t)$ in Fig. 2(a), (b) and (c). In order to see how the probability $p(y_t)$ is obtained, the superimposed plausible predictions $y^{[\theta]}_{t_p:h_p} \in Y^{[H_{\mathrm{th}}, S_{\mathrm{th}}]}_{t_p:h_p}$ are shown in Fig. 2(d), (e) and (f).

Now, let us examine the present method step by step. Firstly, in Fig. 3, we show the attractor distributions of (a) training and (b) predicted time series and (c) the similarity vs. prediction steps. From (c), we can see that the similarity of the attractors changes with the increase of prediction steps, and $h_p = 1000$ seems necessary for the convergence of the change. This indicates that the similarity have to be calculated after this period much bigger than predictable horizons whose mean values are less than 300 as shown in Fig. 4(a) explained below. This finding should be noted because usual methods for time series prediction (e.g. [1–4]) do not examine the property of the prediction after the predictable horizon so much.

Next, we have examined predictable horizons h, $\hat{h}$, $\tilde{h}$, h_S and h^* as shown in Fig. 4(a). Here, h indicates the predictable horizon achieved by the representative prediction, $\hat{h}$ the expected predictable horizon (see (11)), $\tilde{h}$ the LOOCV predictable horizon (see (8)), h_S the predictable horizon achieved by the learning machine which has generated the largest similarity of attractors, and h^* the longest predictable horizon among the horozons achieved by the learning machines for all N. We can see that h^* has achieved longer predictable horizons than others on average while h^* as well as h and h_S cannot be obtained until the

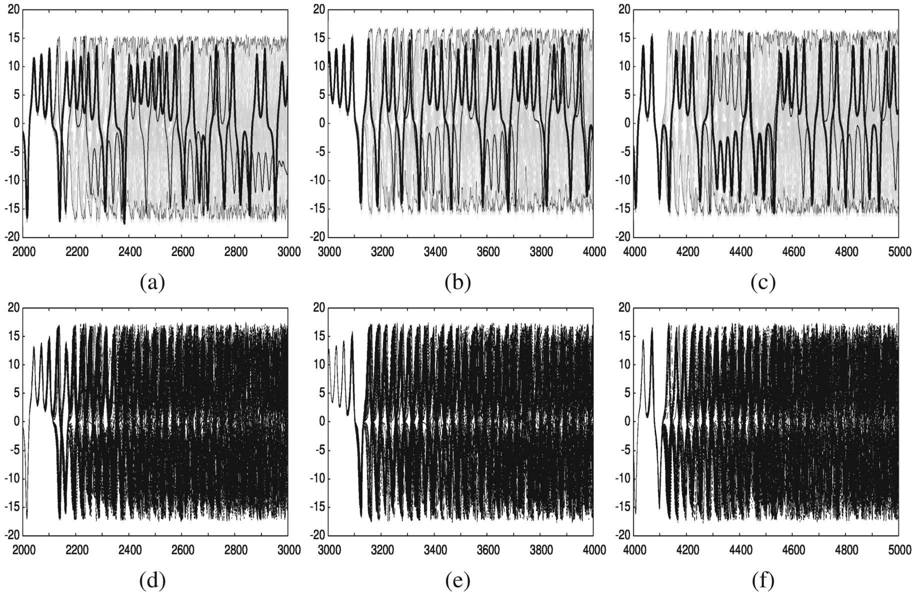

Fig. 2. (a), (b) and (c) show the grand truth time series $y_{t_p:h_p}^{[\mathrm{gt}]}$ (thickest line), representative prediction $y_{t_p:h_p}^{[\theta_{\sigma(1)}]}$ (2nd thickest line) and the probability $p(y_t)$ with positive values (gray area enveloped by thin lines) for $t_p = 2000$, 3000 and 4000, respectively, and $h_p = 1000$. By means of a close look at the difference between the thickest, the second thickest and thin lines growing greater than $e_y = 5$ in (a), (b) and (c), we can see that the pairs of expected and actual predictable horizons of the representative prediction $y_{t_p:h_p}^{[\theta_{\sigma(1)}]}$ are $(\hat{h}_{t_p}^{[\theta_{\sigma(1)}]}, h_{t_p}^{[\theta_{\sigma(1)}]}) = (102, 121)$, (92,234), (118,278) for $t_p = 2000$, 3000 and 4000, respectively. (d), (e) and (f) show superimposed plausible predictions $y_{t_p:h_p}^{[\theta]} \in Y_{t_p:h_p}^{[H_{\mathrm{th}},S_{\mathrm{th}}]}$ for $t_p = 2000$, 3000 and 4000, respectively.

ground truth is obtained. We can see that the estimated predictable horizon $\hat{h}$ is smaller than or equal to the actual predictable horizon h achieved by the representative prediction, and the mean is $\langle\hat{h}\rangle = 82$ [steps] while $\langle h\rangle = 156$ [steps]. This result of estimation can be said conservative and safe, and is obtained by using $H_{\mathrm{th}} = 0.8$. By means of using $H_{\mathrm{th}} = 0.2$, we have $\langle\hat{h}\rangle = 152$ [steps] nearer to $\langle h\rangle = 156$ [steps], but there are a number of $\hat{h}$ bigger than h. This indicates that we can tune H_{th} for the necessary degree of conservativeness in the estimation of predictable horizon.

Next, we can see that h has achieved better result than h_S. This indicates that the effectiveness of the LOOCV predictable horizon $\tilde{h}$ embedded in the present method. The relationship between $\tilde{h}$ and h for all plausible predictions $Y_{t_p:h_p}^{[S_{\mathrm{th}}]}$ at $t_p = 2000$ are shown in Fig. 4(b), where we can see that there are a number of data on or near the line $\tilde{h} = h$ while there are data far from the line. As a result, we have the correlation between $\tilde{h}$ and h being 0.197, which does not indicate

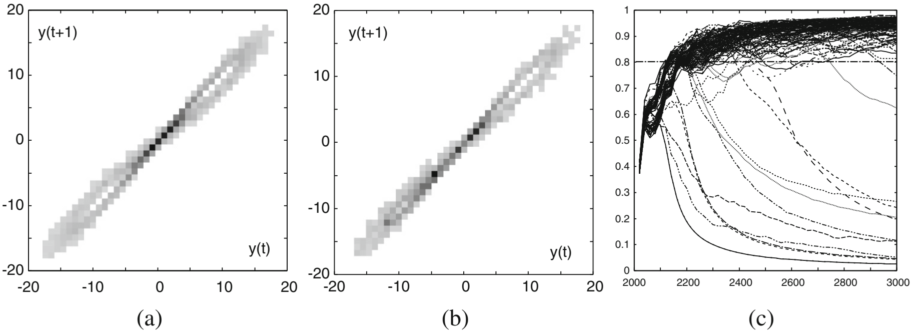

Fig. 3. (a) shows two dimensional training attractor distribution $a_{ij}^{[\text{train}]}$ and (b) shows the predicted distribution $a_{ij}^{[\theta]}$ for $\theta = N = 172$ at $t = 2999$ with $S\left(y_{t_p,h_p}^{[\theta]}, y_{t_g:h_g}^{[\text{train}]}\right) = S\left(y_{2000,1000}^{[\theta]}, y_{0:2000}^{[\text{train}]}\right) = 0.85$. Here, note that $N = 172$ has generated the representative prediction shown in Fig. 2(a). The resolution of the distributions is $\Delta_a = (v_{\max} - v_0)/40 = (18.5 - (-18.5))/40 = 0.925$. (c) shows the similarity $S\left(y_{2000,1000}^{[\theta]}, y_{0:2000}^{[\text{train}]}\right)$ vs. prediction steps for each $\theta = N = 10, 12, \cdots, 250$ for the increase of prediction steps. The predictions $y_{t_p:h_p}^{[\theta]} = y_{2000:1000}^{[\theta]}$ for $\theta = N =$10, 12, 14, 16, 20, 56, 58, 64, 132, 160, 162, 164 are removed because $S\left(y_{t_p,h_p}^{[\theta]}, y_{t_g:h_g}^{[\text{train}]}\right) < S_{\text{th}} = 0.8$.

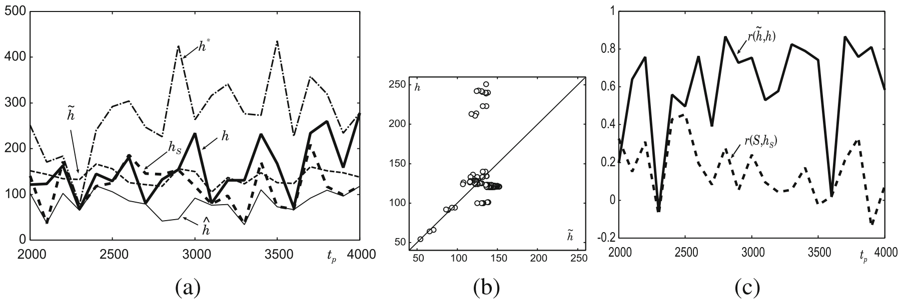

Fig. 4. (a) Predictable horizons h, $\hat{h}$, $\tilde{h}$, h_S and h^* (see the text for details) vs. t_p. The averages are $\langle h \rangle = 156$, $\langle \hat{h} \rangle = 82$, $\langle \tilde{h} \rangle = 139$, $\langle h_S \rangle = 118$ and $\langle h^* \rangle = 273$ [steps]. (b) shows the relationship between $\tilde{h}$ and h for plausible predictions $Y_{t_p:h_p}^{[S_{\text{th}}]}$ at $t_p = 2000$. (c) shows the correlations $r(\tilde{h}, h)$ and $r(S, h_S)$ vs. t_p.

high relationship. From Fig. 4(c), we can see that the correlation between $\tilde{h}$ and h, which we denote $r(\tilde{h}, h)$, is big for some t_p and not so big for other t_p. On the other hand, the correlation between the similarity $S = S\left(y_{t_p,h_p}^{[\theta]}, y_{t_g:h_g}^{[\text{train}]}\right)$ and the predictable horizon h_S derived from S, which we denote $r(S, h_S)$, is much smaller than $r(\tilde{h}, h)$. These results indicate that $\tilde{h}$ may be more effective than S for selecting the predictions achieving longer predictable horizons but it does not work for some predictions.

4 Conclusion

We have presented a method of probabilistic prediction of chaotic time series. The method obtains a set of plausible predictions by means of using the similarity of attractors of training and the predicted time series. It also obtains a smaller set of more plausible predictions with longer LOOCV predictable horizon. We have shown the effectiveness and properties of the method by means of analyzing the result of numerical experiments. We have also shown that the measure of LOOCV predictable horizon is more effective than the similarity measure to select longer actual predictive horizon, but it does not have worked for some predictions. We would like to examine and analyze it in detail and develop more reliable method in our future research studies.

Appendix

A CAN2

5

The CAN2 (competitive associative net 2) is an artificial neural net for learning efficient piecewise linear approximation of nonlinear function by means of the following schemes (See [?] for details): A single CAN2 has N units. The jth unit has a weight vector $\boldsymbol{w}_j \triangleq (w_{j1}, \cdots, w_{jk})^T \in \mathbb{R}^{k\times 1}$ and an associative matrix (or a row vector) $\boldsymbol{M}_j \triangleq (M_{j0}, M_{j1}, \cdots, M_{jk}) \in \mathbb{R}^{1\times(k+1)}$ for $j \in I^N \triangleq \{1, 2, \cdots, N\}$. The CAN2 after learning the training dataset $D^n = \{(\boldsymbol{x}_i, y_i) | y_i = r(\boldsymbol{x}_i) + e_i, i \in I^n\}$ approximates the target function $r(\boldsymbol{x}_i)$ by $\widehat{y}_i = \widetilde{y}_{c(i)} = \boldsymbol{M}_{c(i)}\widetilde{\boldsymbol{x}}_i$, where $\widetilde{\boldsymbol{x}}_i \triangleq (1, \boldsymbol{x}_i^T)^T \in \mathbb{R}^{(k+1)\times 1}$ denotes the (extended) input vector to the CAN2, and $\widetilde{y}_{c(i)} = \boldsymbol{M}_{c(i)}\widetilde{\boldsymbol{x}}_i$ is the output value of the $c(i)$th unit of the CAN2. The index $c(i)$ indicates the unit who has the weight vector $\boldsymbol{w}_{c(i)}$ closest to the input vector $\boldsymbol{x}_i$, or $c(i) \triangleq \underset{j\in I^N}{\operatorname{argmin}} \|\boldsymbol{x}_i - \boldsymbol{w}_j\|$. The above function approximation partitions the input space $V \in \mathbb{R}^k$ into the Voronoi (or Dirichlet) regions $V_j \triangleq \{\boldsymbol{x} \mid j = \underset{i\in I^N}{\operatorname{argmin}} \|\boldsymbol{x} - \boldsymbol{w}_i\|\}$ for $j \in I^N$, and performs piecewise linear prediction for the function $r(\boldsymbol{x})$. AQ1

References

1. Kurogi, S., Ono, K., Nishida, T.: Experimental analysis of moments of predictive deviations as ensemble diversity measures for model selection in time series prediction. In: Lee, M., Hirose, A., Hou, Z.-G., Kil, R.M. (eds.) ICONIP 2013, Part III. LNCS, vol. 8228, pp. 557–565. Springer, Heidelberg (2013)
2. Kurogi, S., Shigematsu, R., Ono, K.: Properties of direct multi-step ahead prediction of chaotic time series and out-of-bag estimate for model selection. In: Loo, C.K., Yap, K.S., Wong, K.W., Teoh, A., Huang, K. (eds.) ICONIP 2014, Part II. LNCS, vol. 8835, pp. 421–428. Springer, Heidelberg (2014)
3. Aihara, K.: Theories and Applications of Chaotic Time Series Analysis. Sangyo Tosho, Tokyo (2000)

4. Slingo, J., Palmer, T.: Uncertainty in weather and climate prediction. Phil. Trans. R. Soc. A **369**, 4751–4767 (2011)
5. Kurogi, S., Sawa, M., Tanaka, S.: Competitive associative nets and cross-validation for estimating predictive uncertainty on regression problems. In: Quiñonero-Candela, J., Dagan, I., Magnini, B., d'Alché-Buc, F. (eds.) MLCW 2005. LNCS (LNAI), vol. 3944, pp. 78–94. Springer, Heidelberg (2006)

Author Queries

Chapter 9

Query Refs.	**Details Required**	**Author's response**
AQ1	Please provide the respective reference citation in the sentence starting with "The CAN2 (competitive associative net 2) ...."	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	· · · under matter to remain	
Insert in text the matter indicated in the margin		New matter followed by or
Delete	through single character, rule or underline or through all characters to be deleted	or
Substitute character or substitute part of one or more word(s)	through letter or through characters	new character or new characters
Change to italics	under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	,
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	through character or where required	
Reduce space between characters or words	between characters or words affected	