



ELSEVIER

Physica A 253 (1998) 379–393

PHYSICA A

Metric adaption for analog forecasting

Klaus Fraedrich*, Bernd Rückert

Meteorologisches Institut, Universität Hamburg, Bundesstrasse 55, D-20146 Hamburg, Germany

Received 15 November 1997

Abstract

The performance of analog forecasts is sensitive to the selection procedure of analogs from the history of observed time series. A method is presented to iteratively reduce a user-defined forecast error measure by adapting suitable metric weights for the components of the reconstructed states to be selected. Applications of the adapted analog forecast scheme to time series generated by low-dimensional systems demonstrate successfully the potential of the proposed technique.

© 1998 Elsevier Science B.V. All rights reserved

Keywords: Analog forecasting; Metric adaption; Nonlinear time series

1. Introduction

Weather analogs and their utilization are a leitmotiv in meteorology [1–3]. They are used to estimate the atmospheric predictability, the dimension of the atmospheric phase space, and the number of weather regimes. Another application is the specification of surface weather through analogs associated with the predicted upper-level flow. Further, practical issues based on analog techniques are related to short-range and long-term weather forecasting.

In principle, forecasting future values of a time series utilizing a set of measurements is possible if patterns of these values have a one-to-one correspondence to states of an underlying dynamical system. The embedding theorem provides the theoretical background for analysing and forecasting dynamical systems utilizing observations. If patterns meet the conditions of the embedding theorem they are referred to as reconstructed states. This one-to-one correspondence offers techniques, arising then as inter- or extrapolation schemes in the reconstructed state space, to forecast the future evolution. The performance of different forecast schemes is generally compared by quantifying their forecast errors. The forecast error is a given or, with respect to a

* Corresponding author.

specific forecast problem, a particular user-defined measure of the deviation of the true value of an observed quantity (verification) from its corresponding forecast. Optimizing a parametric forecast model means adapting its parameters to achieve minimum forecast errors. If, for example, these parameters enter the model equations linearly and the error function is defined as the mean square error at a given lead time, the parameters are, in most cases, estimated by a linear least-square routine. However, if the parameters enter the forecast model in a nonlinear fashion, they need to be adapted by an iterative procedure starting from first guess values (of the parameters). Examples of iteration techniques are the gradient descent method minimizing nonlinear functions and the back-propagation rule [4] in neural (feedforward) nets [5].

The aim of this work is to describe a numerical algorithm that reduces a predefined forecast error measure for an analog forecast scheme applied to a preselected state space reconstruction. The idea of analog forecasting is very basic: assume the state \mathbf{s} , reconstructed from past observations, to be similar to the state \mathbf{r} , reconstructed from the actual observations in the same way, then the future evolution of the past state can be used to forecast the future of the actual state. Measuring the similarity of the states with the metric, $d(\mathbf{r}, \mathbf{s}) = \sum_i G_i (r_i - s_i)^2$, one is interested in an optimal choice for the metric coefficients, G_i , minimizing the predefined forecast error. Murray [6] has shown the potential to reduce the mean square forecast error (for a given lead time) by prescribing different metric coefficients using a state space reconstructed by the time-delay embedding of an observed univariate time series. This paper provides an extension by introducing the following three aspects: The reconstructed phase space may be spanned by more than one observed time series in delay coordinates; the error function is user-defined; and the optimized metric coefficients G_i are adapted by an iteration procedure.

Section 2 gives a brief review of the embedding theorem. Section 3 describes the basic analog forecast method and the iteration procedure to reduce the forecast error. This method is applied in Section 4 to forecast experiments using time series of some well-known low-dimensional nonlinear systems.

2. Embedding theorem (for multivariate time series)

Assume $\mathbf{z}(t)$ to be a state of a dynamical system evolving on a finite-dimensional attractor A . A k -variate time series is a table of measurements $o_j(t)$ (with $j = 1, \dots, k$), recorded at past times $t = 0, -1, -2, \dots$, representing the observation of the j th measurement function applied to the state $\mathbf{z}(t)$. The embedding theorem (for details see, for example, Ref. [7]) asserts that the *pattern* $\mathbf{r}(t)$

$$\begin{aligned} \mathbf{r}(t) &= (r_1(t), r_2(t), \dots, r_m(t)) \\ &\equiv (o_{j_1}(t - t_{j_1}), o_{j_2}(t - t_{j_2}), \dots, o_{j_m}(t - t_{j_m})) \end{aligned} \quad (1)$$

has (under generic conditions) a one-to-one correspondence to the state $\mathbf{z}(t)$ of the system, provided that at least $d_s = [2d_A + 1]$ of the m pairs (j_i, t_{j_i}) (with $j_i \in \{1, \dots, k\}$, $t_{j_i} \in \{0, 1, \dots\}$) are disjunct. Here d_s is referred to as a sufficient embedding dimension;

d_A is the finite box-counting dimension (or capacity) of the attractor A . Patterns $\mathbf{r}(t)$ having this one-to-one correspondence to the states $\mathbf{z}(t)$ are noted as reconstructed states. The important consequence for this prediction theory is the existence of the *reconstructed dynamics* \mathbf{F} , which describes the time evolution of the reconstructed states $\mathbf{r}(t)$ for lead time T as

$$\mathbf{r}(t + T) = \mathbf{F}^T(\mathbf{r}(t)), \quad T \geq 0. \quad (2)$$

Forecasting based on observed time series can now be defined as finding the images, $\mathbf{r}(t+T)$, for lead times $T = 1, 2, \dots$ of a given initial state $\mathbf{r}(t)$ in the reconstructed phase space or, equivalently, finding a model for the unknown reconstructed dynamics \mathbf{F} .

Here it should be noted that the embedding theorem gives only information on the required dimension of the embedding space, provided it is possible to derive an estimation of the attractor dimension d_A . Such estimate can be achieved using the Grassberger–Procaccia algorithm or some variant of it [8,9]. The embedding theorem does not cover the case of infinite-dimensional attractors where the one-to-one property of the system's states to any finite pattern definition is lost. All state space reconstructions based on empirical data are imperfect even in the case that the dimension of the reconstructed phase space exceeds the sufficient embedding dimension. This is due to the fact that any observation of a real valued quantity exhibits measurement errors. Thus, in general, imperfect state space reconstructions cannot be circumvented. One may hope that the chosen (m)-dimensional reconstruction (Eq. (1)) still provides the possibility for imperfect but meaningful predictions – at least for short lead times.

3. Adapted analog forecasting: basic method and metric iteration

Forecasting by analogs of past events is an implication of the Taylor expansion of the reconstructed dynamics (Eq. (2)): If $\mathbf{r}(t_1)$ is close to $\mathbf{r}(t_2)$ then $\mathbf{r}(t_1+T)$ should also be close to $\mathbf{r}(t_2+T)$ [10,11]. The construction of the analog forecast model proceeds in four steps: The first step (A) consists of the reconstruction of a state space and the definition of the forecast problem including the error measure; some guidance to suitable state space reconstructions can be found in Ref. [11]. The second step (B) describes the evaluation of the analog forecast procedure. Both steps together form the basic analog forecast. In the next step (C) the scheme is improved by a learning rule which then, in step (D), is used iteratively to provide an adapted and optimized metric.

Step A: The preselection of an (m)-dimensional state space reconstruction leads to a sequence $\mathbf{r}(t)$ of reconstructed states labeled by the times t of the corresponding observations (Eq. (1)). This sequence is divided into two parts: a learn set, LS , and a test set, TS . The learn set is used for the construction of the forecast model, the test set only to evaluate the model performance on independent data. It is measured by an error function E , which occurs as the sample mean of the *individual* forecast errors

$e(t)$ calculated for the test set:

$$E = \langle e(t) \rangle. \quad (3)$$

The individual error $e(t)$ is, in general, a user and application oriented definition. It measures the deviation of the true time evolution of the initial state, $\mathbf{r}(t)$, at lead time T , $\mathbf{r}(t+T)$, from its corresponding forecast, $\hat{\mathbf{r}}(t+T)$. For example, a convenient measure is the squared error of the first component of the reconstructed states, summed over the lead times $T = 1, 2, \dots, L$,

$$e(t) = \sum_{T=1}^L |r_1(t+T) - \hat{r}_1(t+T)|^2.$$

Step B: The lead time T analog forecast of the j th component of the initial state $\mathbf{r}(t)$ of the test set TS is

$$\hat{r}_j(t+T) \equiv r_j(a(t)+T) \quad \text{with} \quad a(t) \in LS,$$

where $\mathbf{r}(a(t))$ is the nearest neighbour or *analog* of the initial state $\mathbf{r}(t)$ with respect to the metric

$$d(t, t')^2 = \sum_{i=1}^m G_i \cdot (r_i(t) - r_i(t'))^2. \quad (4)$$

Here $d(t, t')$ measures the distance between the two reconstructed states $\mathbf{r}(t)$ and $\mathbf{r}(t')$. For given $\mathbf{r}(t)$ the distance $d(t, t')$ becomes minimal for the closest analog denoted by $\mathbf{r}(a(t))$. The metric coefficients G_i are positive reals. After the normalization of the reconstructed states the Euclidean metric, given by $G_i = 1.0$ (for $i = 1, \dots, m$), serves as a *first guess* for the metric coefficients.

Step C: The forecast error attached to the first guess can be reduced by searching for a better metric. This requires the introduction of a learning rule before an iteration scheme can be designed. To derive a rule for a proper modification of the metric coefficients, the first two nearest neighbours, $\mathbf{r}(a_1(t))$ and $\mathbf{r}(a_2(t))$, of the reference states $\mathbf{r}(t)$, are identified. All of them lie in the learn set LS . Now, it is possible to decide which of the two analogs, say $\mathbf{r}(a_b(t))$, gives a better result; that is, the smaller individual error $e(t)$. Applying this analysis to all reconstructed states of the learn set, LS , leads to a set of relative distances for the first analogs, $f_i(t) = r_i(t) - r_i(a_1(t))$, and another for the better ones, $b_i(t) = r_i(t) - r_i(a_b(t))$, for all $i = 1, \dots, m$ components. Since only the squares of these distances enter the metric, the ratios $q_i(t)$ are calculated, which transform the squared distances of the better nearest neighbours to those of the first ones; that is, $q_i(t) = \langle f_i^2(t) \rangle / \langle b_i^2(t) \rangle$, where $\langle \rangle$ denotes the arithmetical mean for all components i of the learn set. This ratio is used to modify the metric coefficients to favour the finding of those nearest neighbours whose scaling coincides with the scaling of the better nearest neighbours. This leads to a *learning rule* for calculating the new metric coefficients from their old values

$$G_i' \equiv G_i \cdot q_i(t).$$

A ratio larger (smaller) than unity, $q_i(t) > 1$ (< 1), indicates that the squared distances of the better analogs are typically smaller (larger) than those of the first nearest neighbours. In this case the metric coefficient is increased (decreased) by the effective ratio, $q_i(t)$.

Step D: To establish the metric adaption, the learning rule is used iteratively,

$$(G_i(n), E_{LS}(n)) \rightarrow (G_i(n+1), E_{LS}(n+1)), \quad i = 1, \dots, m$$

with the “ \rightarrow ”-arrow indicating a single iteration step.

The tuple $(G_i(0), E_{LS}(0))$ is associated with the first guess metric coefficients, $G_i(0)$, and the learn set error, $E_{LS}(0)$, of the first guess. This scheme should be iterated as long as the learn set error $E_{LS}(n)$ decreases with increasing number, n , of iterations steps. The following simple *stop criterion* is used: The iteration continues until step $n = n_{stop}$, where the relative learn set error reduction $\alpha(n)$ drops below a prescribed threshold α (say 3%) over the past n^* (say 10) iterations:

$$\alpha(n) = \frac{E_{LS}(n - n^*) - E_{LS}(n)}{E_{LS}(n - n^*)} \leq \alpha \quad \text{for } n = 0, 1, \dots, n_{stop}. \quad (5)$$

That is, at least $n^* = 10$ iterations need to be performed before this criterion can be satisfied. As the *solution* for the metric coefficients, we take those values $G_i = G_i(n_{sol})$ which correspond to the minimal learn set error. That is, the iteration step n_{sol} , whose metric coefficients define the solution, is found from the relation $E_{LS}(n_{sol}) \leq E_{LS}(n)$ for all $n = 0, 1, \dots, n_{stop}$.

The following technical comment is in order: Throughout the iteration procedure the nearest neighbours, $\mathbf{r}(a_1(t))$, and, $\mathbf{r}(a_2(t))$, of $\mathbf{r}(t)$ are selected from the learn set LS . To guarantee that these analogs are the nearest neighbours in the reconstructed *space* and not in *time* [12], it is necessary to add the restrictions of their independence: $|a_1(t) - a_2(t)| > \Delta T_{ex}$, $|t - a_1(t)| > \Delta T_{ex}$ and $|t - a_2(t)| > \Delta T_{ex}$, with the time interval ΔT_{ex} of the order of the correlation time.

4. Applications

The adapted forecast technique is subjected to forecast experiments based on the following time series from often cited low-dimensional dynamical systems. These low-order nonlinear models have been developed in various scientific disciplines ranging from mathematics via medicine and physics to meteorology. The Henon [13] system is a simple mapping of a two-dimensional plane defined by

$$\begin{aligned} x(t+1) &= 1 - ax(t)^2 + y(t), \\ y(t+1) &= bx(t). \end{aligned}$$

With $a = 1.4$ and $b = 0.3$, it exhibits the basic properties of a strange attractor. The example from medicine is a first-order nonlinear delay or Mackey–Glass [14] equation,

which describes the fluctuations in peripheral white blood cell counts in chronic granulocytic leukemia. It shows that simple models of physiological systems encounter chaotic regimes similar to those observed in human disease:

$$\dot{x}(t) = \frac{ax(t-s)}{1+x(t-s)^c} - bx(t).$$

The constants are $a=0.2$, $b=0.1$, $c=10.0$ and the delay time of $s=17.0$. Ikeda [15] analyses a plane-wave model of a bistable ring cavity and exhibits period-doubling cascades to chaos, which occurs at the parameter constellation $p=1.0$, $B=0.9$, $\kappa=0.4$ and $\alpha=6.0$. This map relates the field amplitude at the $(t+1)$ th cavity pass to that of a round trip earlier:

$$z(t+1) = p + Bz(t) \exp\left(i\kappa - i\alpha \frac{1}{1+|z(t)|^2}\right).$$

The Lorenz-63 system [16] describes a truncated model of convective rolls in a phase space spanned by the three variables (x, y, z) , which represent normalized spectral (trigonometric) amplitudes of a truncated Boussinesq fluid flow. They simulate the fields of the stream function and the temperature anomaly in a vertical plane. The first-order mode of the stream-function field is denoted by x . The related temperature field of the flow is composed of two modes; its first-order mode, y , shows a warm anomaly associated with upward motion (and vice versa); the second-order mode z is horizontally homogeneous and characterizes the time evolution of the vertical stratification. In meteorology, this model has been serving as a paradigm of the predictability problem:

$$\begin{aligned}\dot{x} &= \sigma(y - x), \\ \dot{y} &= -xz + rx - y, \\ \dot{z} &= xy - bz\end{aligned}$$

with $\sigma=10.0$, $b=\frac{8}{3}$ and $r=28.0$.

These four low-order models are used in the following to analyse the performance of the adapted scheme forecasting time series generated by these models. The analysis of the experiments consists of two steps: First, the analog forecast scheme is established, adapting the metric weights by the described iteration procedure in the learn set. Then, this adapted analog forecast model is applied to an independent verification in the test set.

In all examples N denotes the total length of the observed time series. The first-half of the time series is used as a learn set LS , the remaining part as a test set TS . The error function is defined by the average $E = \langle e(t) \rangle$ of the individual errors $e(t)$. The functional form of the individual forecast errors (reflecting the special task of the prediction) and the (m) -dimensional state space reconstructions are given below. For each example the results are presented in one figure containing two plots labelled (a) and (b): Plot (a) shows the weights of the metric coefficients, $G_i(n)$, for increasing

number of iterations $n = 0, 1, \dots, n_{stop}$, that is,

$$w_i(n) = \frac{G_i(n)}{\sum_{i=1}^m G_i(n)} \quad \text{for } i = 1, \dots, m.$$

Plot (b) displays the decrease of the learn set error $E_{LS}(n)$ during the adapting procedure (solid line). The small diamonds (\diamond) mark the weights which correspond to the solution, $w_i(n_{sol})$, and the attached learn set error $E_{LS}(n_{sol})$ found by the stop-criterion (Eq. (5)).

At each iteration step, n , the analog scheme is applied to forecast the independent test set TS using the available learn set metric $w_i(n)$ shown in plot (a). The dashed line in plot (b) displays the stepwise gain in performance which, at n_{sol} , leads to the test set error $E_{TS}(n_{sol})$ (indicated by a $*$) defined by the solution determined by the iteration procedure in the learn set.

The horizontally dotted line in plot (b) displays the test set forecast error, E_{ref} , of a *reference* forecast. This reference forecast model is given by an analog forecast scheme based on the Euclidean metric; it is the best performing of all these Euclidean analog forecasts which uses all or only one subset of the reconstructed state space components.

Henon (Fig. 1): The x -component of the Henon map is used as a univariate time series of length $N = 10\,000$. The state space reconstruction is a ($m = 6$)-dimensional time-delay embedding $r_j(t) \equiv x(t - j + 1)$ (for $j = 1, \dots, 6$). The error measure used in this prediction experiment is defined as follows. Only lead time ($T = 5$)-forecasts are accepted useful, which deviate less than 10% from their verification (in units of the learn set standard deviation), otherwise they are discarded. The individual error measure is formally expressed as $e(t) = 0$, if $\max_{T=1\dots 5} \{|r_1(t + T) - \hat{r}_1(t + T)|\} < 0.1$, or $e(t) = 1$, otherwise.

Mackey–Glass [14] (Fig. 2): A time series with length $N = 10\,000$ and sampling period $\tau_s = 10.0$ of the Mackey–Glass delay equation, is used in this example. The state space reconstruction is a ($m = 8$)-dimensional time-delay embedding, defined as $r_j(t) \equiv x(t - \Delta t(j - 1))$ (for $j = 1, \dots, 8$) with $\Delta t = \tau_s$. The task is to minimize the ordinary mean square error function for the lead time $T = 5$ (in units of τ_s) for the first component of the reconstructed states. Thus, the individual errors are given by $e(t) = |r_1(t + T) - \hat{r}_1(t + T)|^2$.

Ikeda (Fig. 3): The univariate time series chosen consists of the real parts of the Ikeda map with $N = 10\,000$. The state space reconstruction is a ($m = 8$)-dimensional time-delay embedding with the components $r_j(t) \equiv \text{Re}(z(t - j + 1))$ (for $j = 1, \dots, 8$). The individual error is defined as $e(t) = \sum_{T=1}^L |r_1(t + T) - \hat{r}_1(t + T)|^2$ with maximum lead time $L = 5$. It reflects the forecast task to minimize, on average, the sum of the squared errors, accumulated over the first $L = 5$ lead times for the first component of the reconstructed states.

Lorenz-63 (Fig. 4): The 3-variate time series comprises the (x, y, z) -components of the Lorenz system for $N = 10\,000$ with sampling period $\tau_s = 0.025$. The reconstructed state space is ($m = 6$)-dimensional with $\mathbf{r}(t) = (x(t), y(t), z(t), x(t - \Delta t), y(t - \Delta t), z(t - \Delta t))$; $\Delta t = 8 \cdot \tau_s$ is chosen to avoid high autocorrelations amongst the components of the

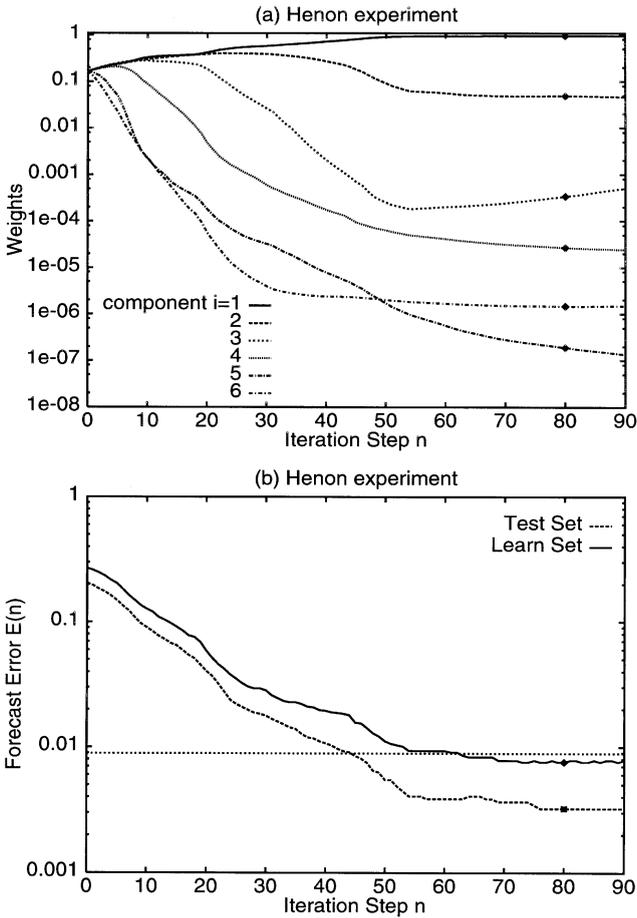


Fig. 1. The adapted analog forecast (AAF) scheme developing by metric iteration (n) for the Henon forecast experiment: (a) Metric weights $w_i(n)$ of the reconstructed state components $i=1, \dots$ and (b) learn set averaged forecast errors (solid line); the diamonds (\diamond) at $n_{sol}=80$ mark the AAF learn set solution. For each n panel (b) includes the test set forecast errors (dashed) based on the learn set metric $w_i(n)$ displayed in (a); the asterik (*) indicates the AAF test set error at the solution $n_{sol}=80$. The horizontally dotted line in (b) gives the test set forecast error of the best reference analog forecast based on the Euclidean metric.

reconstructed states. The functional expression for the individual error $e(t)$ is the same as in the Ikeda experiment, but with a maximum lead time $L=25$ (in units of τ_s).

The following results of the prediction experiments with the adapted analog forecast (AAF) scheme are discussed:

(1) In all cases the AAF-scheme produces a remarkable error reduction. The error reductions in the test and learn are more or less in phase throughout the iteration procedure. The forecast error calculated for the test set is smaller than the forecast error of the learn set, since all learn set states may be used as analogs to forecast the test set states, but not vice versa.

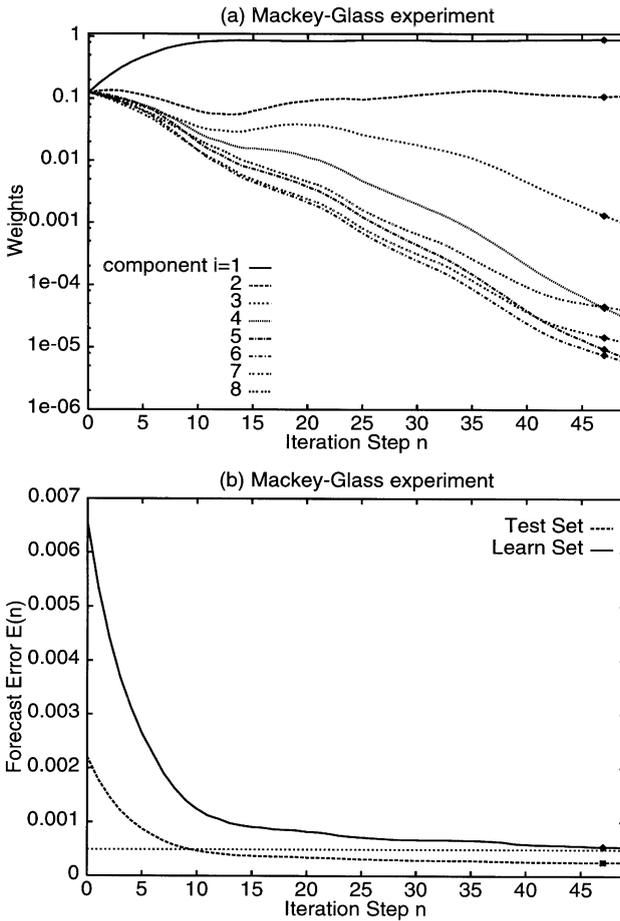


Fig. 2. Same as in Fig. 1, but for the Mackey–Glass forecast experiment.

(2) The skill of the AAF defines the relative gain in accuracy compared with the reference scheme

$$S = \frac{E_{ref} - E_{TS}(n_{sol})}{E_{ref}}.$$

The skill is about 64% for the Henon forecast experiment, 8% for Mackey–Glass, 11% for Ikeda, and 29% Lorenz-63.

(3) In the Ikeda experiment, the iteration passes an optimal (or suboptimal) solution without converging to it. This is due to the fact that the derivation of the new coefficients is based on a rule which modifies the metric coefficients without considering possible drawbacks of this modification. That is, a change of the metric coefficients by the learning rule may be caused by a few cases only (for which a better analog exists), but this change may be bad for all other states. The risk to miss an optimal solution

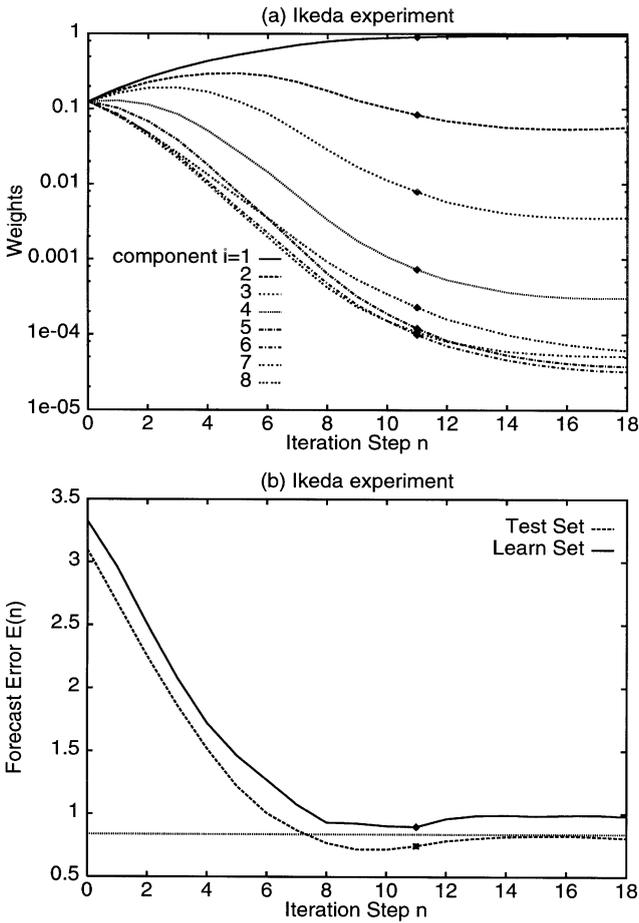


Fig. 3. Same as in Fig. 1, but for the Ikeda forecast experiment.

can be reduced by introducing a control set (a randomly chosen subsample of the learn set) for an interim forecast trial made during the iteration procedure. Rising errors in the control set would then lead to an immediate n_{stop} overruling the ($\alpha = 3\%$)-threshold criterion.

(4) The forecast examples show that some components of the reconstructed states are attached to very low weights by the adapting iteration scheme. It seems likely that at least the same forecast skill could be achieved setting these very low weights to zero. This leads to a state space reduction with improved forecast skill. A threshold for eliminating the low weighted components should itself be adaptable.

(5) Minimizing error measures, whose forecasts lead times lie beyond the limit of predictability generates more or less large fluctuations in the metric weights during iteration phase and no error reduction. Furthermore, the performance of the

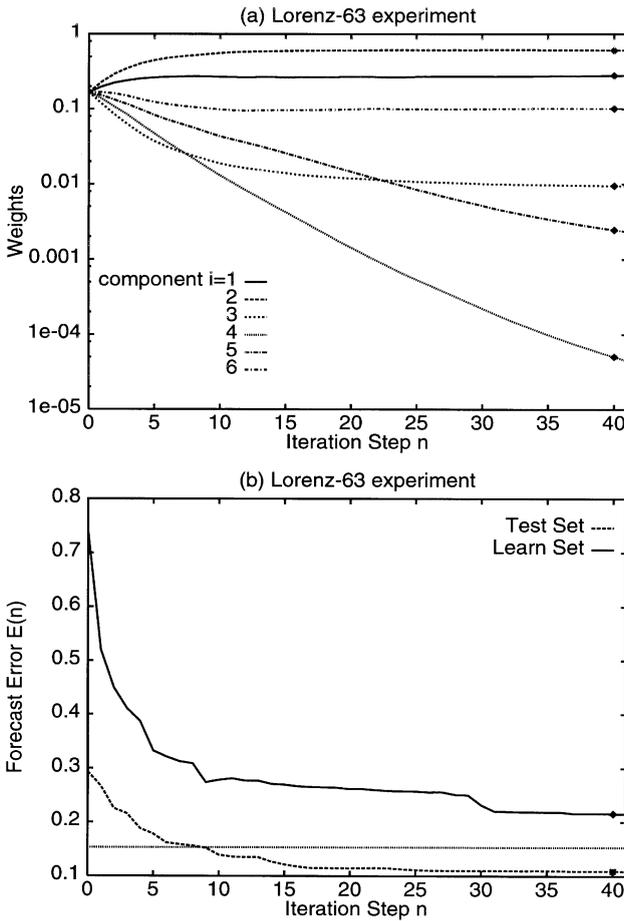


Fig. 4. Same as in Fig. 1, but for the Lorenz-63 forecast experiment.

AAF-scheme may depend on the first guess metric and on the state space reconstruction. For example, highly correlated time-delay components reduce the forecast accuracy (e.g. for $\Delta t = 1\tau_s$ in the Lorenz-63 example, not shown).

(6) Including off-diagonal weights to the metric is easily obtained replacing the similarity measure of states (Eq. (4)) by

$$d(t, t') = \sum_{i,j=1}^m G_{ij} |r_i(t) - r_i(t')| \cdot |r_j(t) - r_j(t')|. \tag{6}$$

The distance $d(t, t')$ between the two analog states $r(t)$ and $r(t')$ is determined by the metric coefficients, G_{ij} , which are positive reals; $G_{ij} = 1$ (for $i, j = 1, \dots, m$) serves as the first guess for the search of an optimal metric. Consequently, the learning rule

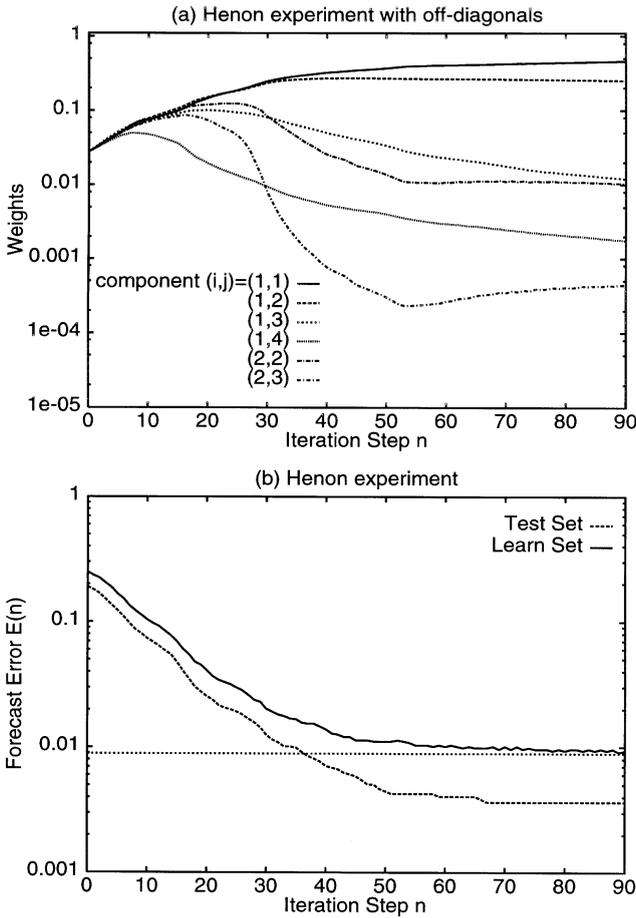


Fig. 5. As in Fig. 1, but including off-diagonal metric weights w_{ij} : (a) The first six diagonal or off-diagonal weights, w_{ij} , changing with iteration step n ; (b) the test set forecast error (dashed line) of the adaptive analog scheme based on the learn set metric, the best reference analog forecast error (dashed horizontal line) based on the Euclidean metric, and the learn set (solid line) forecast error evolving with the iteration step n .

(STEP-C) to compute the new metric coefficients from an old predecessor needs to be modified accordingly:

$$G'_{ij} = G_{ij} q_{ij}(t). \quad (7)$$

Again, the ratios for all i, j components are calculated by the individual scaling ratios, $q_{ij} = \langle f_{ij} \rangle / \langle b_{ij} \rangle$; they are defined as the arithmetic learn set sample mean distances of the first two nearest neighbours to the reference state $r(t)$; that is, the nearest $f_{ij} = |r_i(t) - r_i(t')| \cdot |r_j(t) - r_j(a_1(t'))|$ and the better one $b_{ij} = |r_i(t) - r_i(t')| \cdot |r_j(t) - r_j(a_b(t'))|$. The iterative adaption can now be performed following the description of STEP-D. This

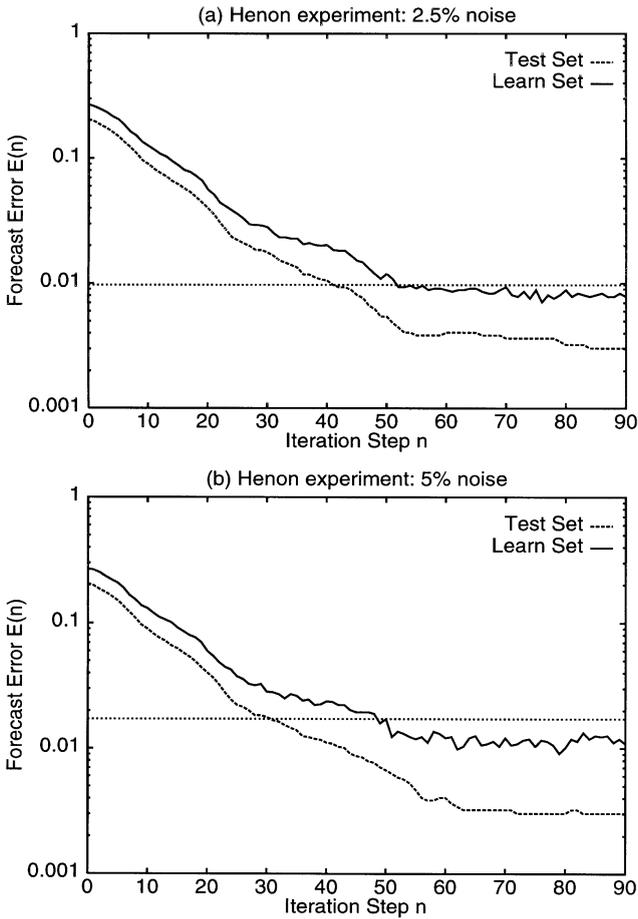


Fig. 6. As in Fig. 1b, but for adding white noise with an intensity of (a) 2.5% and (b) 5% of the system’s standard deviation. The test set forecast error (dashed line) of the adaptive analog scheme based on the learn set metric, the best reference analog forecast error (dashed horizontal line) based on the Euclidean metric, and the learn set (solid line) forecast error evolving with the iteration step n .

leads to the weights $w_{ij} = G_{ij} / \sum_{i,j=1}^m G_{ij}$. Application to the Henon experiment (Fig. 5) show that the introduction of off-diagonal elements into the adaptive analog scheme does not lead to considerable improvement of the forecast accuracy.

(7) The sensitivity of the scheme to adding white noise of varying intensity is analysed and the results are shown for the Henon system (Fig. 6). Two noise levels of increasing intensity of (a) 2.5% and (b) 5% of the system’s standard deviation do not reveal a considerably different forecast performance compared with the zero noise case (Fig. 1). Further increase of the noise level, the adaptive analog scheme attains the optimal metric weights at fewer iterations which, however, are associated

with larger forecast errors. Utilizing off-diagonal elements does not change these results considerably.

5. Discussion and conclusions

It is possible to rewrite the adapted metric with coefficients G_{ij} as a linear filter matrix \mathbf{B} acting on the reconstructed states $\mathbf{r}(t)$. The elements of \mathbf{B} are defined as $B_{ij} \equiv \sqrt{G_{ij}}$ ($i, j = 1, \dots, m$). Analog forecasting using the coefficients G_{ij} of the adapted metric may therefore also be interpreted as applying first the linear filter \mathbf{B} to the reconstructed states before analog forecasts are made based on the standard Euclidean metric. In this sense, the described technique of metric adaption fits to the theory of global linear filters; but the calculation of the filter components is based on local considerations and related to the defined forecast problem.

The nonlinear predictive scheme adapted here is a particular form of analog prediction which, for a time-delay embedding of dimension m , utilises the nearest neighbours in the time-delay coordinate phase space chosen from the “library” or history of the univariate observations. In this sense, it appears to be similar to the widely used simplex method of Sugihara and May [17] for the prediction of time series; the spectrum of applications ranges from ecology to economy. Instead of using a prescribed metric, however, the weights are determined by an adaption procedure which is linked to the forecast task.

Further improvement of the analog method may be achieved by recycling forecast errors. As the errors do indeed contain information about both the observed and the prediction system the error history can be expected to contribute in increasing the skill of the predictions if included in the forecast. In this sense analog predictions can be improved by an error recycling process which contains both observations and errors as predictors. The above approach is just one of many such procedures now being employed by Fraedrich and Leslie [18] with success to now-casting geophysical systems like short-term climate variability, hurricane tracks, etc. While the conventional approach to forecasting might be termed dualistic, because it treats predictions and observations as two totally different entities, this is a monistic approach grouping predictions and observations to a unified predictor set. In this sense prediction errors enlarge the phase space of a predictive model with both observations *and* errors characterizing the enlarged state vector.

In summarizing, this paper describes an application of the embedding theorem to forecasting future values of time series by an adapted analog forecast scheme utilizing a univariate or multivariate measured time series. It has been emphasized that optimal predictions are achieved only with respect to an error measure. Once the forecast errors are quantified, self-adapting iteration schemes emerge in a natural way leading to improved forecast accuracy for classes of forecast methods. This procedure is demonstrated for analog forecasting which appears to be the most intuitive way of nonlinear empirical prediction.

References

- [1] E.N. Lorenz, *J. Atmos. Sci.* 26 (1969) 636.
- [2] T.P. Barnett, R.W. Preisendorfer, *J. Atmos. Sci.* 35 (1978) 1771.
- [3] H.M. van den Dool, *Tellus A* 46 (1994) 314.
- [4] D.E. Rumelhart, G.E. Hinton, R.J. Williams, *Nature* 323 (1986) 533.
- [5] B. Widrow, M.A. Lehr, *Proc. IEEE*. 78 (1990) 1415.
- [6] D.B. Murray, *Physica D* 68 (1993) 318.
- [7] T. Sauer, J.A. Yorke, M. Casdagli, *J. Stat. Phys.* 65 (1991) 579.
- [8] P. Grassberger, I. Procaccia, *Physica D* 9 (1983) 189.
- [9] K. Fraedrich, R. Wang, *Physica D* 65 (1993) 373.
- [10] J.D. Farmer, J.J. Sidorowich, *Phys. Rev. Lett.* 59 (1987) 845.
- [11] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, L.S. Tsimring, *Rev. Mod. Phys.* 65 (1993) 1331.
- [12] J. Theiler, *Phys. Rev. A* 34 (1986) 2427.
- [13] M. Henon, *Commun. Math. Phys.* 50 (1976) 69.
- [14] M.C. Mackey, L. Glass, *Science* 197 (1977) 287.
- [15] K. Ikeda, *Opt. Commun.* 30 (1979) 257.
- [16] E.N. Lorenz, *J. Atmos. Sci.* 20 (1963) 130.
- [17] G. Sugihara, R. May, *Nature* 344 (1990) 734.
- [18] K. Fraedrich, L.M. Leslie, *Weather Forecasting* 12 (1998), in press.