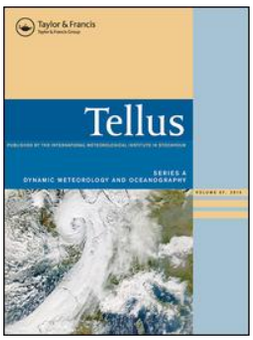

Tellus A: Dynamic Meteorology and Oceanography

ISSN: (Print) 1600-0870 (Online) Journal homepage: https://www.tandfonline.com/loi/zela20

Discriminating randomness from chaos with application to a weather time series

V. Cuomo, C. Serio, F. Crisciani & A. Ferraro

To cite this article: V. Cuomo, C. Serio, F. Crisciani & A. Ferraro (1994) Discriminating randomness from chaos with application to a weather time series, Tellus A: Dynamic Meteorology and Oceanography, 46:3, 299-313, DOI: 10.3402/tellusa.v46i3.15480

To link to this article: https://doi.org/10.3402/tellusa.v46i3.15480

Tellus (1994), 46A, 299-313
Printed in Belgium - all rights reserved

TELLUS
ISSN 0280-6495

Discriminating randomness from chaos with application to a weather time series

By V. CUOMO and C. SERIO*, *Dipartimento Ingegneria e Fisica dell'Ambiente Università della Basilicata, Potenza, Italy*; F. CRISCIANI and A. FERRARO, *Istituto Talassografico di Trieste, Trieste, Italy*

(Manuscript received 25 March 1993; in final form 27 July 1993)

ABSTRACT

A new method is presented which permits one to discriminate low-dimensional chaos from randomness. The method consists in fitting autoregressive processes to the data and forecasting future values of the system on the basis of the model selected. We distinguish between 2 possible forecasting techniques of a dynamical system given by experimental series of observations. The "global autoregressive approximation" views the observations as a realization of a stochastic process, whereas the "local autoregressive approximation" views the observations as the realizations of a truly deterministic process. A proper comparison between the predictive skills of the 2 techniques allows us to gain insight into distinguishing low-dimensional chaos from randomness. The procedure has been applied to a daily temperature time series recorded in Trieste (Italy) over the past 40 years (1950 to 1989). The analysis gives no evidence for low-dimensional chaos, the dynamics being compatible with a limit cycle blurred by red noise.

1. Introduction

Atmospheric phenomena are typically classed as either weather or climate. While the former refers to the behaviour of the atmosphere over a period ranging from few to several days and arising primarily from internal instabilities, the latter deals generally with the behaviour over a relatively long time not only of the atmosphere but of the entire earth system. According to Leith (1973, 1975, 1978) a climatic state is defined as a finite time average of a weather state. Such an average is subject to fluctuations of statistical nature arising from day-to-day weather events that are unpredictable on time scales of climatological concern.

Discussing the problem of the predictability of climate, Leith (1973, 1975, 1978) showed that the magnitude of the statistical error, i.e., the so-called climatic noise, affecting a given climatic state was related to persistence in the atmosphere. Persistence is another characteristic peculiar to weather events. In spectral terms it means that power spectra of meteorological time series are dominated by low frequency components. Leith suggested that many meteorological variables could be approximated by a Markov process and showed that the amplitude of the climatic noise decreases approximately as $(T/\tau)^{-1/2}$, where T is the averaging time and τ is an integral correlation time.

On meteorological time scales, quantitative analyses of the predictability of atmospheric phenomena had been already undertaken by Lorenz (1969) who estimated that the effective number of degrees of freedom was a few hundred in the global atmosphere. Thus, such findings were in agreement with the above stochastic view which assumes that irregular behaviour in nature results of necessity from the interaction of a large number of degrees of freedom. A stochastic approach was also taken by Hasselmann (1976) in formulating his stochastic models of climate.

However, according to the new ideas about the

* Corresponding author.

irregular variability of systems only a few degrees of freedom suffice, by interacting nonlinearly, to create deterministic chaos. According to such a view, despite of its complexity, the weather system, locally in time and space, may attain to a low-dimensional attractor. On such an attractor, we could describe the system on the basis of few variables, instead of the large set of the primitive ones. Therefore determining the presence of low-dimensional chaos from experimental time series is not a mere academic matter, but can have important dynamic implications.

In the past few years, many efforts have been made to determine the dimensionality of some underlying dynamical system for various atmospheric variables with rather surprising answers of less than 10 (e.g., Fraederich, 1986). However, more recently, Ruelle (1990) has shown that these low dimensions are likely to be an artifact of the finite lengths of the time series examined. Thus, the presence of low-dimensional chaos in weather time series is still an open problem.

The main problem we want to address in this paper is to distinguish between chaos and randomness, when *the only information about a given signal comes from the observations of the signal itself. It is explicitly assumed that the dynamic equations which govern the phenomenon are unknown.*

So far the way to diagnose chaos has been based on procedures aiming at estimating geometric invariants of the attractor or Lyapunov exponents. However, today there exists an extensive literature indicating that such parameters are quite difficult to estimate even when a long and accurate series of observations is available.

Conversely the method we have devised is not based on estimating dimensions, entropies or Lyapunov exponents. The key point of the procedure we have developed to distinguish between chaos and randomness is the ability of forecasting a given signal on the basis of its past events with either stochastic-based approaches or deterministic procedures.

The approach we propose consists in fitting autoregressive processes to the data, then forecasts can be produced on the basis of the model selected. We distinguish between two possible strategies of approximating a given time series with autoregressive processes. The *global autoregressive technique* views the signal as the realization of a stochastic process, the autoregressive model is fitted to all the data points at once and estimates of the autoregression coefficients are computed consistently with the above assumption of randomness. On the other hand, based on recent ideas about the problem of forecasting chaotic time series (Farmer and Sidorowich, 1987; Serio, 1992), the *local autoregressive technique* assumes the signal to be truly deterministic, and the autoregression coefficients, and consequently predictions, are obtained according to such an assumption.

The predictions obtained using the local autoregressive technique are based on finding local portions of a given time series in the past, which closely resemble the present, and basing forecasts on what happened immediately after these past events. This strategy works well for deterministic systems if their motions is on an attractor, regular or chaotic as well, and provided that the attractor dimension is not too high.

On the other hand, reflecting the infinite dimensionality of the system, the ability of making good predictions when a stochastic process is involved increases with the number of data points that are used to get estimates of the autoregression coefficients. Then, for such a case the global autoregressive technique is expected to be more adequate.

In practice, we compare the forecast error functions computed according to the two techniques. Because both techniques use the same mathematical tool (i.e., autoregressive models) to build predictors but different approaches to the estimation of parameters, differences between the above two functions are expected to be mostly due to the dynamical characteristics of the process.

A local approach to prediction was proposed by Lorenz (1969) in his aforementioned study about the predictability of the global atmosphere. However, the origin of the method goes back to the old idea of analog weather forecasting (e.g., Malone (1951)). Moreover, a forecasting deterministic technique was implicitely suggested by Eckmann and Ruelle (1985). Finally, the deterministic approach was discussed by Farmer and Sidorowich (1987) who clarified the interrelationships between the predicting procedure and the geometric and dynamical invariants of the underlying attractor. More recently, Serio (1992), based on the local linearization of dynamical systems, proposed to construct local predictors which were autoregressive processes.

In this paper we have concentrated on analysing the possible presence of low-dimensional chaos in one metereological time series of daily average ambient temperature. However the procedure could be applied to any time series. Therefore we shall discuss the mathematical background of such a procedure (Section 2), before applying it to the observations (Section 3). Conclusions are drawn in Section 4.

2. Mathematical background

2.1. The global autoregressive prediction

Let $x(t)$ denote a certain signal, or function of time. We shall assume that the signal is sampled at equal interval of time, say Δt, then writing $x(t=n\Delta t)=x(n)$, the sequence $\{x(n)\}$, $n=1, ..., N$ (N being the number of data) constitutes a discrete time series.

According to linear prediction a forecast $\hat{x}(n+1)$ of $x(n+1)$, standing at origin $n\Delta t$, is given by a linear function of current and previous observations, i.e.,

$$\hat{x}(n+1)=\phi_1 x(n)+ \cdots +\phi_p x(n+1-p), \qquad (1)$$

where the weights $\phi_1, ..., \phi_p$ are to be determined. Such a forecast function gives an optimal estimate of $x(n+1)$ (in a statistical sense) provided that the weights are determined in such a way that the series of the one step ahead forecast errors, i.e., the residual time series: $z(n+1)=x(n+1)-\hat{x}(n+1)$, $n=p, ..., N$, becomes uncorrelated (Box and Jenkins, 1976). Thus, considering the z's as a mere set of independent random variables or a white noise process, the above problem of linear prediction is equivalent to fit to the data an autoregressive process.

A pth-order autoregressive process (Box and Jenkins, 1976) can be expressed as:

$$x(n+1)=\sum_{j=1}^{p} \phi_j x(n+1-j)+z(n+1), \qquad (2)$$

with the parameters $\phi_1, ..., \phi_p$ being the autoregression coefficients. Also in (2) it is required that the z's are mutually independent and normally distributed with zero mean and finite variance, σ_z^2. The linear process (2) can be fitted to the data in any of several ways. For the work reported here we used the well-known Yule-Walker recursive method, while, when required, the optimal order, p_{opt}, was selected by the Akaike's Information Criterion (Akaike, 1974), but the other Akaike criterion, the Final Prediction Error (FPE), could have been used as well.

For $p=1$, eq. (2) defines a first order autoregressive process which is often called Markov process in much literature. Such a distinction is sometimes misleading, since process (2), when converted to a state space representation, satisfies a first order difference equation, i.e., a pth-order scalar autoregressive process becomes a Markov process in a space of higher dimension. The conversion can be made by introducing a p-dimensional state vector, $\boldsymbol{x}(n)=(x(n), ..., x(n+1-p))^{\mathrm{T}}$, with the superscript T indicating transpose. This state vector is governed by the state-transition equation: $\boldsymbol{x}(n+1)=\boldsymbol{A}\boldsymbol{x}(n)+\boldsymbol{z}(n+1)$, where $\boldsymbol{z}(n+1)=(z(n+1), 0, ..., 0)^{\mathrm{T}}$ and the p by p matrix $\boldsymbol{A}$ is in the so-called phase-canonical form with 1's along the underdiagonal line and the parameters $\phi_1, ..., \phi_p$ displayed on the top row:

$$\boldsymbol{A}=\begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \qquad (3)$$

Such a conversion to a state space is referred to as embedding in dynamical system literature.

Once the autoregression coefficients have been estimated, the l-step ahead forecast, $\hat{x}(n+l)$; $l \geqslant 1$, standing at origin n, can be obtained directly in terms of the difference equation (2) but with the residual part, z set equal to zero:

$$\hat{x}(n+l)=\sum_{j=1}^{p} \phi_j \tilde{x}(n+l-j), \qquad (4)$$

where

$$\tilde{x}(n+l-j)=\begin{cases} x(n+l-j), & j \geqslant l \\ \hat{x}(n+l-j), & j<l. \end{cases} \qquad (5)$$

We see that this form of the forecast function, $\hat{x}(n+l)$, contains observations which have already happened at origin n and observations which have not yet happened (when $l>p$, the forecast func-

tion contains only observations which have not yet happened). The unknown observations are replaced by their forecasts at origin n. It should be noted that the set of autoregression coefficients is left unchanged at each step l.

If we write down the model for the series $x(n)$ in the state space representation, it is possible to derive an equivalent expression for the forecast function, where the *initial state*: $(x(n), x(n-1), ..., x(n+1-p))$ is left unchanged at each step l and the coefficients are continuously updated. The scalar form of this forecast function reads:

$$\hat{x}(n+l) = \varphi_1(l)\, x(n) + \varphi_2(l)\, x(n-1) + \cdots + \varphi_p(l)\, x(n+1-p), \tag{6}$$

where the coefficients, $\{\varphi_1(l), \varphi_2(l), ..., \varphi_p(l)\}$, are the elements of the top row of the matrix

$$\boldsymbol{A}^l = \boldsymbol{A} \cdot \boldsymbol{A} \cdots \boldsymbol{A} \tag{7}$$

obtained by multiplying the transition matrix, in the phase-canonical form, l-times by itself. Of course $\varphi_j(1) = \phi_j$; $j = 1, ..., p$, and the iterative nature of the procedure for computing the other coefficients, $\varphi_j(l)$ $l > 1$, is quite evident from (7).

Finally, it is possible to show (e.g., Box and Jenkins, 1976) that the root mean square (r.m.s.) estimation error $\sigma_{\hat{x}}(l)$ of the l-step ahead forecast is given by

$$\sigma_{\hat{x}}^2(l) = \left\{1 + \sum_{j=1}^{l-1} \psi_j^2\right\} \sigma_z^2, \tag{8}$$

where the weights satisfy the difference equation:

$$\psi_j = \phi_1 \psi_{j-1} + \cdots + \phi_p \psi_{j-p} \tag{9}$$

The relation between the weights ψ and φ is simply: $\psi_j = \varphi_1(j)$. That is, the weight ψ_j is the first element of the top row of the matrix $\boldsymbol{A}^j$.

For convenience $\sigma_{\hat{x}}(l)$ can be normalized to the r.m.s. deviation of the data, σ_x forming the normalized forecast error function:

$$E(l) = \frac{\sigma_{\hat{x}}(l)}{\sigma_x}. \tag{10}$$

From now on we shall consider only normalized errors. The forecast error function, either normalized or not, has been used by many authors in connection with fitting linear stochastic processes to atmospheric variables (among many others, see Hasselmann and Barnett (1981), Privalsky (1983), Amato et al. (1989)).

Provided that the system attains to a stationary state, the effectiveness of autoregression coefficient estimation depends only on the size of the time series, the more data we use, the better, since the statistical error decreases as $N^{-1/2}$. Having N data points the best we can do is of best-fitting the model to all the data at once (hence the name global autoregressive technique) in order to minimize the statistical error affecting the autoregression coefficients.

The statistical procedure above should only be applied to data which have at least an approximately Gaussian distribution, whereas chaotic signals may have distributions that are highly non-Gaussian. However, nonlinear transformations can be considered which allow one to deal with data having an approximately Gaussian distribution (e.g., Katz, 1982, Amato et al., 1989).

2.2. *The local autoregressive prediction*

As opposed to a random process, a chaotic system evolves according to an unique path if it is prepared in the same initial state, since the system is governed by a deterministic state-transition equation:

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = F(\boldsymbol{x}(t)), \tag{11}$$

where the function F can be highly non-linear. The finite predictability, now, arises from the phenomenon of sensitive dependence on initial conditions.

Despite their different physical meaning, low-dimensional chaos and randomness are to some extent mathematically reconcilable. First of all, regardless of whether the function in (11) is linear or not, regular attractors, i.e., attractors which describe periodic systems (limit cycles) and quasi-periodic systems (tori) admit an exact autoregressive representation (e.g., Serio, 1992). That is, every one of the components, x_k, of the state vector, $\boldsymbol{x}$, evolves, on the attractor, according to the process (2) with $\sigma_z^2 = 0$, the order p being equal to $2\times$ the number of frequencies. As an example, the order of a pure sine wave (one frequency) is $p = 2$. If F in (11) is not linear and the

attractor is not regular, e.g., chaotic, we always can linearize locally the system by developing F in a Taylor series and retaining only the first order terms. Such a system, when the time is discretized, will evolve, approximately, according to a Markov state-transition equation: $\boldsymbol{x}(n) = \boldsymbol{A}\boldsymbol{x}(n-1)$ with the random part equal to zero. Again, the components of the state vector, $\boldsymbol{x}$ can be approximated by a loeal linear autoregressive representation:

$$x(n) \approx \sum_{j=1}^{p} \phi_j(n)\, x(n-j), \tag{12}$$

the order of autoregression being equal to the order or degrees of freedom of the system plus one. In (12), the notation $\phi_j(n)$ indicates that the autoregression coefficients, now, depend on time. Therefore if we try to estimate such coefficients by fitting the linear autoregressive process to all the data points at once, we will get only poor estimates (hence forecasts), since the global technique produces only one set of autoregression coefficients, instead of the time varying functions, $\phi_j(n)$, we need in order to get effective forecasts. Indeed, in such a case we should allow for time dependent autoregression coefficients. This can be done by localizing the estimation procedure itself.

Let $\boldsymbol{x}_o$ and $\boldsymbol{x}_o^*$ denote two different but close initial states, then the two paths, corresponding to the two initial conditions, will evolve remaining close until a certain time when the trajectories will diverge each other. Now let us suppose that the system evolves according to $\boldsymbol{x}_o$ but the evolution equations are not known, then a natural predictor of the future state of the system will be $\boldsymbol{x}^*(t)$, provided that the path corresponding to $\boldsymbol{x}_o^*$ is known. If the system is chaotic only limited term prediction will be possible, in general, since, by definition itself of chaos after a certain time the trajectories will diverge according to the exponential law:

$$\|\delta\boldsymbol{x}(t)\| = \|\delta\boldsymbol{x}_o\| \exp(\lambda_1 t), \tag{13}$$

where $\delta\boldsymbol{x}_o = \boldsymbol{x}_o^* - \boldsymbol{x}_o$, $\delta\boldsymbol{x}(t) = \boldsymbol{x}^*(t) - \boldsymbol{x}(t)$, λ_1 is the largest Lyapunov exponent of the system and $\|\cdot\|$ denotes the usual Euclidean norm:

$$\|\boldsymbol{x}_o^* - \boldsymbol{x}_o\| = \sqrt{(x_{o1}^* - x_{o1})^2 + \cdots + (x_{op}^* - x_{op})^2}, \tag{14}$$

with x_{ok}; $k = 1, ..., p$, the kth component of the vector $\boldsymbol{x}_o$ and so on.

The key point of the deterministic prediction is that the motion takes place in a limited region of the state space, therefore we expect that the system will return near to a generic initial state after a time sufficiently long. Thus if $\boldsymbol{x}_o$ denotes the current state we can search for an *analogue* $\boldsymbol{x}_o^*$, on which to base predictions, simply by looking at the past history of the system.

According to the local autoregressive approach, to get the l step ahead forecast, $\hat{x}(n+l)$; $l \geqslant 1$, starting at the origin n, we again construct a predictor which is an autoregressive filter:

$$\hat{x}(n+l) = \varphi_1(n, l)\, x(n) + \cdots + \varphi_p(n, l)\, x(n-p+1). \tag{15}$$

Here, x indicates a single observable of a certain state vector $\boldsymbol{x}$, and the notation $\varphi_j(n, l)$ for the l-step ahead autoregression coefficients indicates that such coefficients depend on n and l.

For a given origin n, once the autoregression coefficients have been estimated, one might compute the $\varphi_j(n, l)$'s directly from the canonical matrix as discussd in the previous section. However, the local estimation is based on few data compared to the global approach, therefore in order to smooth the forecast estimates we prefer to obtain, at each step l, the coefficients, $\varphi_j(n, l)$ directly from the data by the following procedure.

First, we consider k p-ple, $\boldsymbol{x}_p(m_i) = (x(m_i), ..., x(m_i - p + 1))$; $m_i < n$ and $i = 1, ..., k$, that minimize the Euclidean norm $\|\boldsymbol{x}_p(n) - \boldsymbol{x}_p(m_i)\|$. Then, we regard every p-ple $\boldsymbol{x}_p(m_i)$ as the input of the filter (15) and $x(m_i + l)$ as the output. Finally the coefficients $\varphi_1(n, l), ..., \varphi_p(n, l)$, are obtained by best-fitting the autoregressive model to the couples: $(\boldsymbol{x}_p(m_i), x(m + l)) \in \Re^p x \Re$, where $\Re$ denotes the real axis. It should be noted that, for a given origin n, the search for the k p-ples is done once, while the best-fitting procedure is performed l times, that is at each step l, in order to compute the $p \cdot l$ coefficients, $\varphi_j(n, l)$. The fit can be made in any way one likes. We did least-squares by singular-value decomposition (e.g., Press et al., 1986). When $k = p$, this is equivalent to linear interpolation which is notoriously unstable, therefore to ensure stability we chose $k = 2p$.

It must be noted that the effectiveness of the

local procedure rests, ultimately, on the presence in the time series of good analogues of the current state. This presence depends on the sample size, N, i.e., on the whole observing time period, on the Lyapunov exponents and on the dimension of the attractor. The larger such a dimension is, the larger N must be in order to observe naturally occurring analogues, so that only low dimensional chaos is likely to be identified analysing samples of finite size.

To get good predictions the minimal requirement $p \geqslant d$ must be fulfilled, where d is the dimension of the attractor. Since d is unknown *a priori*, we implemented the local prediction scheme starting with $p = 1$ and repeated it for increasing values up to a suitable upper bound p_u, then the optimal order p_{opt} was selected that minimized the one step ahead forecast error, $E(1)$. The procedure to compute $E(1)$, and more in general the function $E(l)$, is based on producing out-of-sample forecasts.

First, we divide the total number of data points, N, into N_1 and N_2 and consider N_1 as the time origin. The first N_1 points are used to compute the coefficients of the predictor according to the procedure discussed above. Then, the first out-of-sample forecasts for origin N_1, $\hat{x}(N_1 + l)$, are produced and compared with the data, $x(N_1 + l)$. Next, we move the time origin of one step and produce the second out-of-sample forecasts, $\hat{x}(N_1 + 1 + l)$, and so on. Finally the forecast error function is computed as:

$$E^2(l) = \frac{1}{\sigma_x^2} \frac{1}{N_2 - l} \sum_{n=N_1+1}^{N_1+N_2-l} (x(n+l) - \hat{x}(n+l))^2, \quad 1 \leqslant l \leqslant l_{\max} \tag{16}$$

with $l_{\max}$ the maximum lead time at which the forecasts are computed.

To summarize, the rationale for the selection of the optimal order is as follows. Let $x(n)$ a realization of the process for which the order is to be determined. We obtain $\varphi_1(n_o, l), ..., \varphi_p(n_o, l)$, $p = 1, ..., p_u$, $l = 1, ..., l_{\max}$ for each given origin n_o from $x(n)$ by least-squares. If $y(n)$ is another realization of the process, we can compute $\hat{y}(n_o + l)$ by using the previous observation in the series $y(n)$ and the coefficients from $x(n)$. Finally we determine $E(l) = \langle (y(n_o + l) - \hat{y}(n_o + l)^2 \rangle$ (with the angular brackets denoting expectation values) and adopt as the order of the process, p_{opt}, that p for which $E(1)$ is minimized.

Basically, this the same rationale as the one behind the final prediction error, FPE (Akaike, 1974), although we really use two realizations of the process whereas Akaike's FPE is estimated by the same realization as the one used to compute the autoregression coefficients. We prefer the out-of-sample estimation procedure since it is a direct check of the model, too. For the same reason and for consistency with the local approach, also in the case of the global approach, we use the out-of-sample technique to compute the function $E(l)$, instead of the theoretical relation (8). However, we found that the final results are not affected by using either (8) or the out-of-sample procedure to estimate $E(l)$. What is important in our analysis is not the particular fitting procedure itself but the logic behind it: that is either local or global.

Finally, we want to focus the attention on how well the local approximation works. This depends on various geometric parameters of the attractor, on the number of data points and on the signal-to-noise ratio as well. According to Farmer and Sidorowich (1987) the following empirical formula can be used for the normalized forecast error $E(l)$:

$$E(l) = E(1) \exp(2hl), \qquad E(1) = \frac{C \exp(2h)}{N^{2/d}}, \tag{17}$$

where h denotes the metric entropy (it is essentially the sum of all the positive Lyapunov exponents) and C is a constant to be determined, providing that we are in the limit of the so called small error, $E(l) \ll 1$ and that the signal-to-noise ratio, R, fulfils the condition

$$R = \frac{\sigma_x}{\sigma_{\text{noise}}} > N^{1/d} \tag{18}$$

The average spacing, Δ, of data points on the attractor is $\Delta \approx \sigma_x N^{-1/d}$, therefore relation (18) simply requires that the r.m.s. deviation of the noise is less than the typical spacing of the data. When condition (18) is not fulfilled, the forecast will be limited by noise. This does not mean that valuable forecasts are no longer possible. It only means that the noise may swamp out the geometrical characteristics of the attractor, therefore making quite difficult the estimation of the parameters of the attractor (i.e., Lyapunov exponents,

dimensions and so on). Conversely, the quality of a forecast is determined by the value of E, and $E \ll 1$ is still possible when condition (18) is not met. This a very important point, since it means that even in a case in which geometric invariants of the attractor cannot be derived because of noise, valuable predictions may be still possible.

2.3. *Metric entropy and autoregressive representation of chaotic signals*

There exists a close relation between the metric entropy of a chaotic process and the white noise variance σ_z^2 of an autoregressive process *globally* fitted to the observations of the process itself (Serio, 1992). Let $f: \Re^m \to \Re^m$ be a map with a chaotic attractor. The attractor is mapped into itself by f. Let $\boldsymbol{x}(n) = f^n(\boldsymbol{x}(0)); n > 1$ be a sequence of iterates under f lying on the attractor. Let us suppose that we do not know the map f but only the iterates $\boldsymbol{x}(n)$ from $n = 0$ up to a given value $n = n_0$. Then based on this *a priori* knowledge we want information about the next iterate $\boldsymbol{x}(n_0 + 1)$. It can be shown (Peterson, 1983) that the metric entropy *is a measure of our average uncertainty about where the map f moves the point* $\boldsymbol{x}(n_0)$.

Now, let $x(n)$ be a time seies of a single component of the vector $\boldsymbol{x}(n)$ belonging to the attractor. When such a series is best fitted with an autoregressive process we replace the non-linear deterministic map by a linear stochastic map. The linear map is only an approximation of the true non-linear map but it retains most of the characteristics of the underlying attractor; the more data points we have, the better. What is the average uncertainty about where the stochastic map will move the point $x(n)$? From eq. (8), we get immediately the solution, the answer is σ_z or its normalized version $E(1)$. Thus, there is a one-to-one relation between σ_z and h, the higher h is, the higher σ_z will be and vice versa. It must be stressed that it is not possible to have $\sigma_z = 0$ for a chaotic process, otherwise the process would not be chaotic. Indeed, when $\sigma_z = 0$ it means that exact predictions are possible with the linear map. This case corresponds to regular attractors (limit cycles and tori) which, in turn, correspond to signals which are periodic or quasi-periodic. It is well-known that such systems do not produce information, that is, they are characterized by $h = 0$. Conversely chaotic systems are characterized by a large production of information, the production rate of such information is then measured by h and reflected by σ_z. Since the larger σ_z is, the worst the forecasts based on the filter (2) are, the metric entropy poses clear limits to the ability of forecasting a chaotic signal with the global approach. This is because only poor forecasts are to be expected by globally fitting an autoregressive map to a chaotic system. In this case we should resort to a local procedure.

The considerations above are crucial to understand our discriminating procedure. In practice, it could happen that both techniques give poor forecasts, e.g., due to insufficient statistics or too large noise level. However, if the system is chaotic we expect the global approach to give the worst results.

2.4. *Numerical examples*

To illustrate the use of our procedure we shall apply it to some simple non-linear and linear dynamical systems. First, we discuss the well-known logistic map:

$$y(n+1) = A \cdot y(n)(1 - y(n)). \tag{19}$$

We consider $y(n)$ as the observable, while the measurements, $x(n)$, are modeled as having Gaussian noise $\eta(n)$ with variance σ_η^2 and zero mean:

$$x(n) = y(n) + \eta(n). \tag{20}$$

Selecting $A = 3.821$, the map is chaotic with the Lyapunov exponent $\lambda_1 = 0.44$, the metric entropy $h \approx \lambda_1$ and the dimension $d \approx 1$.

Fig. 1 illustrates the effects of forecasting a chaotic signal generated from the logistic map with both techniques (global and local). For the cases shown in Fig. 1, 5000 data points were generated and the first 3000 discarded to form a sequence of size 2000. Two levels of external noise were considered: $\sigma_\eta = 0$ and $\sigma_\eta = 0.1\sigma_x$, ($\sigma_x$ denoting the standard deviation of the data).

In the case of no external noise added, the local predictor error (lower solid line in Fig. 1) is extremely low even when the lead time, l becomes large. In this case we are in the small error regime, i.e., $E \ll 1$, and the typical spacing of the data is greater than the r.m.s. deviation of the noise, therefore the growth of the forecast error is exponential.

Conversely, the global predictor has an error which is comparable with the variance of the series

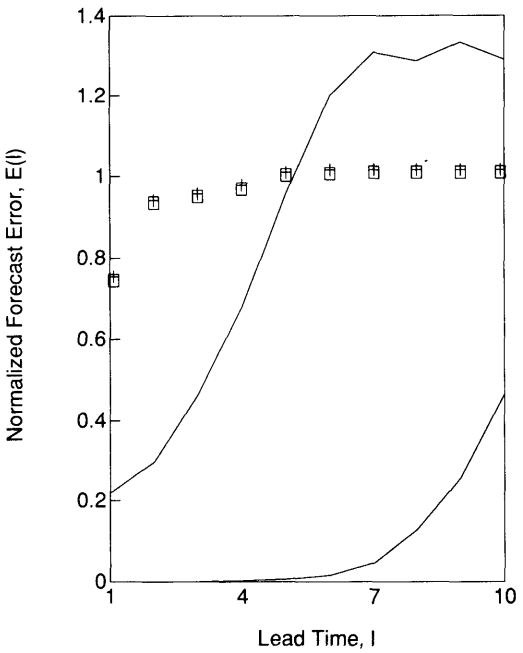

Fig. 1. Logistic map with $a = 3.821$. Figure shows the normalized forecast error function against the lead time l. The lower solid curve refers to the local predictor and 0% external noise added; the other solid line refers to the local predictor but for an error level of 10%. The two curves corresponding to the global predictor are drawn with open squares (0% error) and plus signs (10% error).

and it is clearly inferior to the local one. According to our discussion about the metric entropy (Subsection 2.3), this result is not surprising and finds its justification in the relatively large value of h.

The optimal order selected for the global predictor was $p_{opt} = 12$, while $p_{opt} = 2$ was obtained for the local predictor.

In the case of the noise level equal to 10%, the local forecast error function (upper solid line in Fig. 1) in about 4 steps becomes comparable with the variance of the series, but it is quite evident that the local predictor is still superior to the global one. In this case we are not in the small error limit and according to relation (18) σ_η is by far larger than the typical spacing of the data on the attractor, yet the deterministic nature of the signal is quite distinguishable.

The optimal order selected for the global predictor procedure was $p_{opt} = 12$, while again $p_{opt} = 2$ was obtained for the local predictor, the same as for the case with no external noise.

The logistic map permits one also to analyse the dependence of p_{opt} on the lead time, l. According to the procedure discussed above, p_{opt} is selected in correspondence of the order p for which the one step ahead forecast error exhibits a minimum. Someone could argue that p_{opt} might change in the case that the l-step ahead forecast error was minimized instead of the one step error. This could happen in the case $E(l) \approx 1$, since in this case the estimate would be dominated by the external noise. However, if we limit ourselves to case for which $E(l) < 1$ the dependence of p_{opt} on the order of the forecast error function is not crucial. As an example Fig. 2 shows the normalized 1-step, 2-step, 3-step ahead forecast error against the model order, p in the case where no external noise was added to the signal. All the curves exhibit a clear minimum at $p = 2$. The same results were obtained by considering the 10% noise level (Fig. 3).

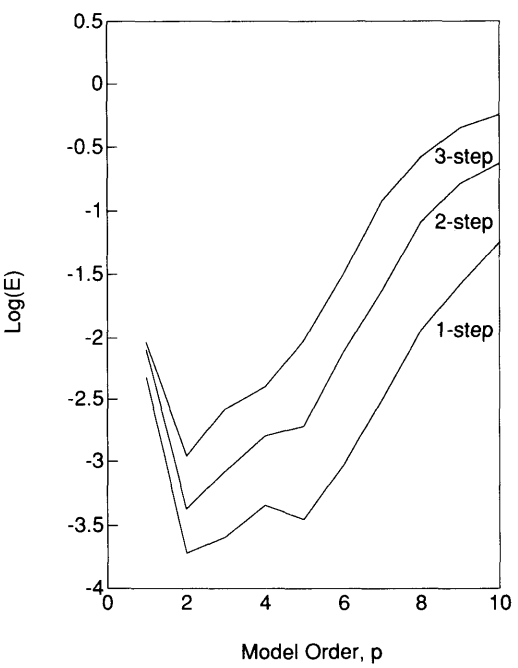

Fig. 2. Logistic map with $a = 3.821$; no external noise added. Figure shows the 1-step, 2-step and 3-step ahead normalized forecast error against the model order, p. The cases shown in figure refers to the local procedure.

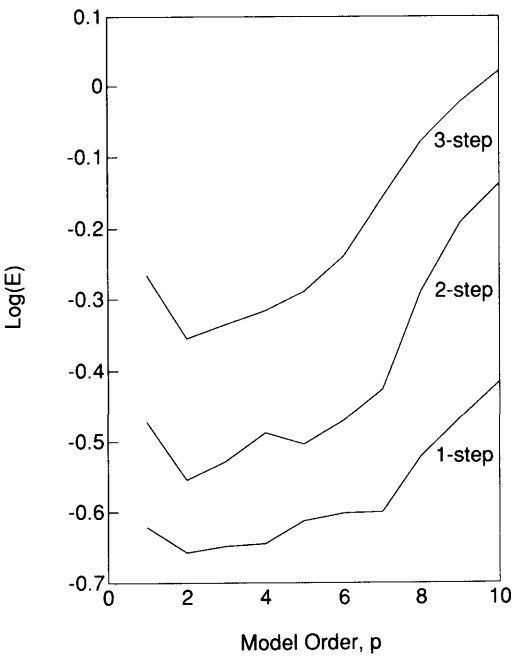

Fig. 3. As Fig. 2 but in the case of 10% external noise added.

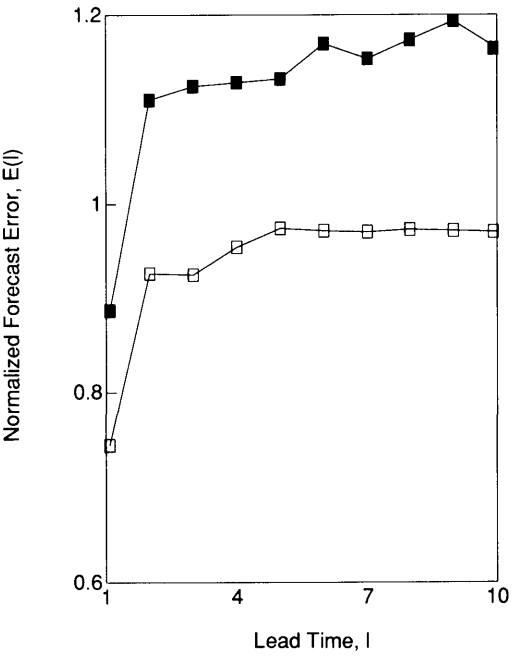

Fig. 4. AR(2) process corrupted with additive white noise (signal-to-noise ratio: 10). Figure shows the normalized forecast error function: global technique (open squares) and local technique (full squares).

We have applied the above procedure to a lot of artificial low-dimensional systems including the Henon map (Henon, 1976), the Lorenz equations (Lorenz, 1963), the Mackey-Glass delay-differential equation (Mackey and Glass, 1977). Such systems are characterized by dimension lower than 4 and metric entropy ranging from ≈ 0 to ≈ 2.5. We found that less than 10000 data points suffices to discriminate chaos from randomness providing that the error level is of order of (or less than) 10% of the r.m.s. deviation of the data. Similar results are implicitly contained also in the works of Farmer and Sidorowich (1987) and Casdagli (1989).

In contrast to the above example, now we consider an autoregressive model with $p = 2$ for the signal: $y(n) = 0.75y(n-1) - 0.5y(n-2) + z(n)$, where $z(n)$ is chosen to be a Gaussian white noise with zero mean and unit variance. The signal has zero mean and variance $\sigma_y^2 = 1.78$. As above the signal was corrupted with external noise having standard deviation equal to $0.1\sigma_y$. We point out that such a process is intrinsically characterized by a low predictability. Even if the signal had not been corrupted with external noise, the one step ahead forecast error would be as large as the standard deviation of the endogenous noise term. The latter is comparable with the standard deviation of the signal itself.

From Fig. 4 we see that the global predictor always gives better results picking up the stochastic nature of the signal. For such an example we found $p_{opt} = 2$ in the case of the global analysis, whereas $p_{opt} = 3$ was found with the local analysis.

3. Application to a meteorological time series

Now, we come back to our objective of analysing the structure of a long meteorological time series. The data we analysed consist of daily average values of air temperature recorded over the past 40 years at the Istituto Sperimentale Talassografico of Trieste, Italy. Air temperature is systematically recorded hourly on the basis of analogue data from thermographs and electronic thermometers which are daily calibrated by means of a standard reference thermometer (0.05 °C scale)

and a couple of maximum-minimum thermometers. The hourly data are then averaged to form daily values.

The overall variance of the series is, $\sigma_T^2 = 51.3°C^2$ so that the signal to observational error ratio, $R = \sigma_x/\sigma_{noise}$ is better than 140 : 1. The size of the time series is $N = 14610$. There are no missing data in the whole sequence. Fig. 5 shows a plot of the forty annual mean temperatures against the year. Also to naked eye the stationarity of this series is quite evident.

In the following we shall show further details about the stochastic models which were built in order to describe the series above. The first model that was built is of the type cyclostationary which describes the observations as the composition of cyclic and stochastic components. For a meteorological daily time series the obvious cyclic component is the annual cycle, the other one less obvious is related to the mean fluctuations (climatic variability) around the annual cycle. Following Madden (1976), (see also Trenberth 1984a, 1984b and reference therein) the annual cycle was computed by finding the mean value for each day of the year over all the forty years and then determining a smoothed annual cycle. This smoothed cycle was determined by developing the 365 daily mean values in Fourier series and retaining only the first few harmonics (the first in the case of the annual cycle; the first, the second and the third in the case of the mean fluctuations).

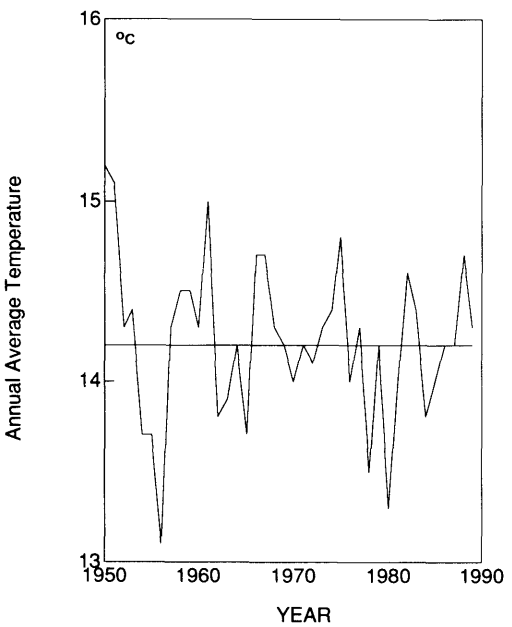

Fig. 5. Plot of the forty annual average temperature against the year. The horizontal line gives the overall mean.

Let $T(n)$ indicate the value of air temperature recorded at day n then the time series cyclostationary model is: $T(n) = m(n) + s(n)\, w(n)$; $n = 1, ..., N$, where $m(n)$ describes the smoothed annual cycle and $s(n)$ the behaviour of the daily standard deviation (climatic variability), $w(n)$ is a random variable. Both $m(n)$ and $s(n)$ are periodic functions with basic period equal to 365 days (one year).

The annual cycle explains more than 98 % of the overall variance, σ_T^2, of the observations, such a behaviour being typical of temperature time series in the Italian climate (Amato et al., 1989). The consequence is that the function $m(n)$ itself provides already a good model to forecast future values of the variable T. Standing at an origin n_o the forecast $\hat{T}(n_o + l)$ at l steps ahead ($l > 0$) is simply: $\hat{T}(n_o + l) = m(n_o + l)$, while an estimate of the forecast error is given by $s(n_o + l)$. Because the average value of the time series $s(n)$ is 2.6°C, the function $m(n)$ provides forecasts within a tolerance of about 2.6°C.

The term $w(n)$ is a stochastic process with zero mean and unit variance describing the day-to-day fluctuations. According to Leith's conjecture it was modelled by a Markov or red noise or first order autoregressive process: $w(n) = \rho w(n-1) + z(n)$, where ρ indicates the first autocorrelation coefficient and $z(n)$ is a Wiener-Levy process (white noise). To get an estimate $\hat{\rho}$ of ρ we must determine realizations of the process, $w(n)$. These were obtained from the time series $T(n)$ by removing the annual cycle: $w(n) = (T(n) - m(n))/s(n)$. The value of $\hat{\rho}$ we obtained was $\hat{\rho} = 0.81$. Fig. 6 shows an histogram of the noise term w; the empirical frequency distribution compares well with a Gaussian law which is also shown in the figure.

Based on the model above (hereafter referred to as the Leith model), we obtain the following forecast function:

$$\hat{T}(n+l) = m(n+l) + s(n+l)\, \hat{\rho}^l w(n), \qquad l > 0, \tag{21}$$

$$w(n) = \frac{T(n) - m(n)}{s(n)}.$$

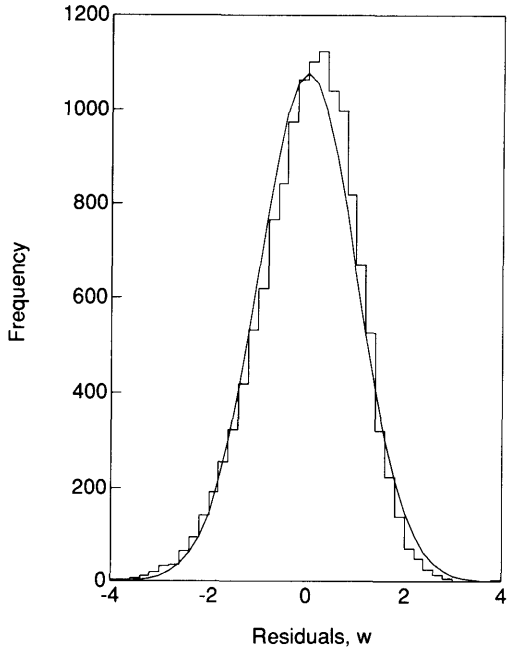

Fig. 6. Histogram of the noise term, w. The comparison with a Gaussian distribution function (solid line) with zero mean and unit standard deviation) is shown, too.

From (21), we see that, in the limit $l \to \infty$, $\hat{T}(n+l) \to m(n+l)$, that is the forecast converges to the value predicted by the annual cycle.

In the above model the annual cycle is regarded as a deterministic feature of the time series that is identical from year to year. A different approach allows for stochastic seasonalities and trends in the time series. According to such an approach the general autoregressive model (2) has to be fitted to the data directly rather than decomposing the series in trend seasonalties and random parts. The autoregressive filter is regarded as a device for transforming the highly dependent, and possibly nonstationary process $T(n)$, to a sequence of uncorrelated random variables $z(n)$, that is for transforming the process to white noise.

Using the well-known Yule-Walker recursive method, autoregressive processes of order, p, from 0 through 400 were fitted to the observations. The optimal order was selected by the Akaike's Information Criterion (Akaike, 1974). It turned out to be $p_{\text{opt}} = 164$.

The high optimal order, $p_{\text{opt}} = 164$, is due to the presence of strong harmonic components (i.e., the annual cycle) in the series of the observations. Harmonic components in additive noise give rise to a special case of an Autoregressive Moving Average (ARMA) process (Kay and Marple, 1981). Such a special process can be still modelled by an autoregressive process but now the order p becomes infinity, at least theoretically. Thus the high order observed is compatible with the structure of the series, confirming the sharp peak of the variance spectrum at the 12-monthly cycle. At this stage we want to point out that our objective is not to find a parsimonius representation of the observations, our objective is to analyse the deep structure of the observations and to distinguish between chaos and randomness. Once this task has been accomplished, the problem of finding a parsimonius representation of the data can be addressed as one thinks best.

Based on the autoregressive model fitted to the observations the forecast function reads:

$$(\hat{T}(n_o+l)-\bar{T}) = \sum_{j=1}^{p_{\text{opt}}} \hat{\phi}_j(\tilde{T}(n_o+l-j)-\bar{T}), \tag{22}$$

where $\bar{T}$ indicates the overall mean of the data set ($\bar{T} = 15.1$C), and with

$$\tilde{T}(n_o+l-j) = \begin{cases} T(n_o+l-j), & j \geqslant l, \\ \hat{T}(n_o+l-j), & j < l. \end{cases} \tag{23}$$

3.1. Results

For the work reported here the out-of-sample forecast technique was implemented with $N_1 = 37 \times 365$, i.e., the first 37 years, so that N_2 corresponds to the last 3 years of the 40 that were analysed. With such a choice the number of data points to estimate the function $E(l)$ was about 1000. Furthermore, no previous Gaussian transform was applied to the data, since the time series is well approximated by a Gaussian law.

The optimal order for the local predictor turned out to be $p_{\text{opt}} = 4$. At first, sight such a result would seem to indicate the presence of a chaotic attractor, the result being also in agreement with the findings of Fraedrich (1986). However the forecast error function, $E(l)$, computed for such a case ($p = 4$) is not consistent with the hypothesis of

low-dimensional chaos since the global predictor gives better results.

Fig. 7 shows a comparison among the different methods we used to forecast the observations. The forecast error function for predictions obtained on the basis of the annual cycle is included in such a figure, too. We see that the two global stochastic predictors converge to the annual cycle and nearly coincide, indicating that the series contains nothing but a strong harmonic component immersed in red noise. The local predictor gives results which are worse than the annual cycle already at the second day and the shape of the latter looks like the shape of the global predictors.

The forecast error functions were also derived for the time series $T(n) - m(n)$, i.e., for the residual series obtained by removing the annual cycle. The r.m.s. deviation of this time series is about 2.7°C while the signal-to-noise ratio is $R \approx 50$.

The results are shown in Fig. 8. The global forecast error function was obtained by fitting to the data a first order autoregressive process or

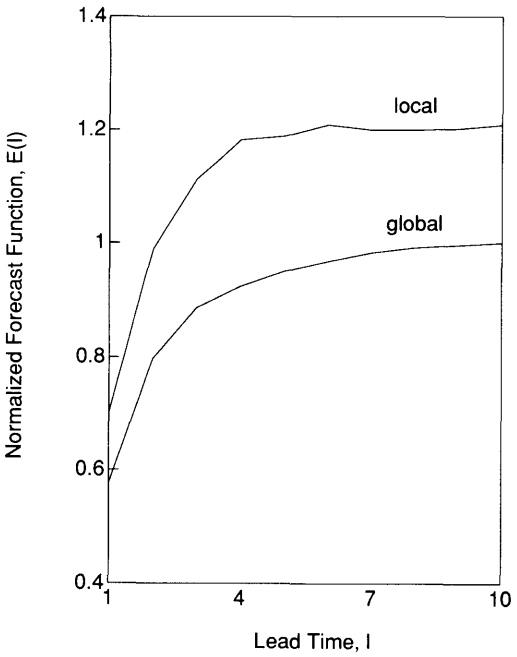

Fig. 8. Results of forecasting the residual time series.

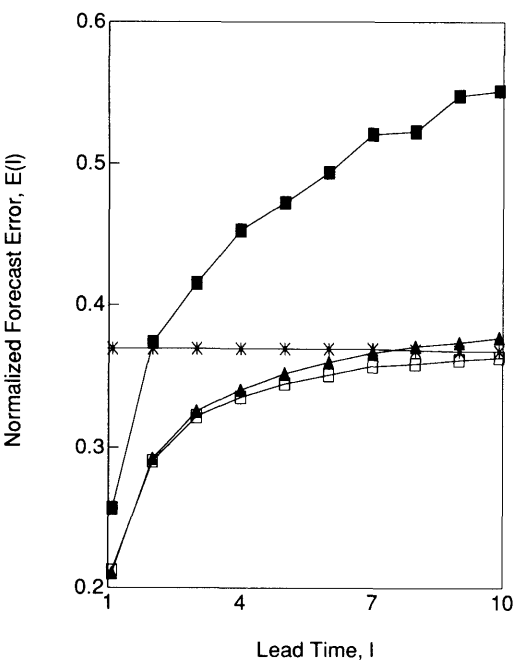

Fig. 7. Results of forecasting the daily average temperature time series; multi-year sample. Leith's model (open squares); global autoregressive model (full triangles); annual cycle (stars); local autoregressive model (full squares).

Markov process. Again, the global predictor is superior to the local one. The optimal order, p_{opt}, was found to be equal to 3 for both global and local analysis.

A possible criticism to our analysis above is that for studying weather processes one should not consider the highly non-linear longtime memory between season and years. A multi-year meteorological time series includes long range processes, the interannual variability and weather phenomena in summer and winter. A dynamical system with all such aspects has a large number of degree of freedom and, therefore, depends on a large number of independent variables, from which its stochastic nature as revealed by our analysis. Instead, the data set should be reduced to seasonal samples before applying the above procedure.

For such a reason an analysis using seasonal samples was also carried out. Predictions were obtained only for the summer and winter seasons and, as for the local predictor, the search for good analogues was limited to past summer seasons and the winter ones, respectively. The summer season was defined to begin on 1 May, while the winter season on 1 November, both lasting 120 days.

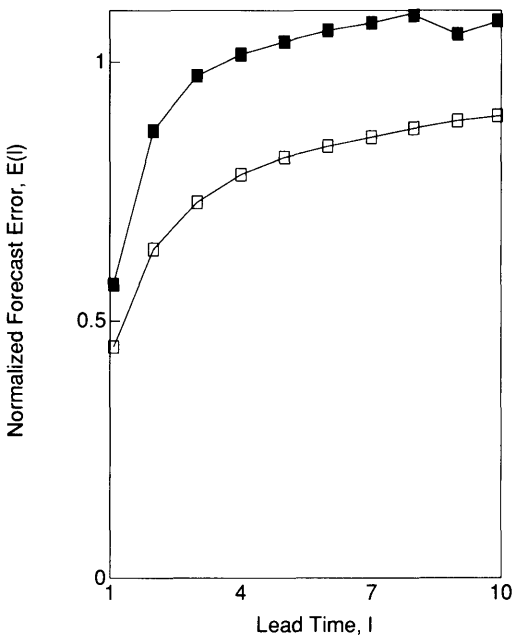

Fig. 9. Results of forecasting the daily temperature time series; seasonal samples (winter). Leith's model (open squares); local autoregressive model (full squares).

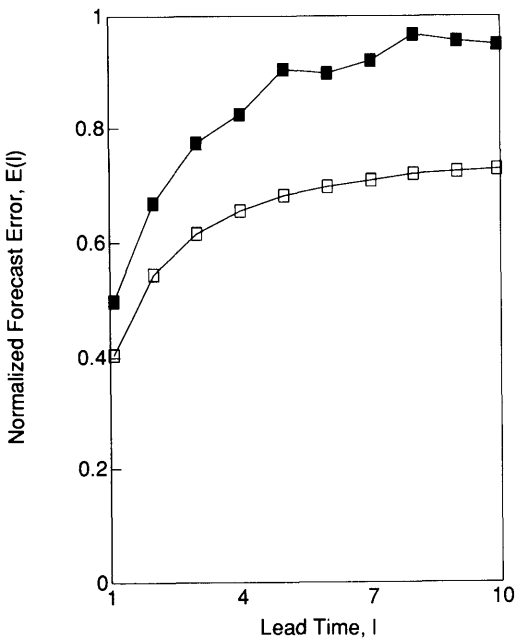

Fig. 10. Results of forecasting the daily temperature time series; seasonal samples (summer). Leith's model (open squares); local autoregressive model (full squares).

To ensure a reasonable convergence of the forecast error function, we chose N_1 equal to the first 35 years. As a consequence every forecast error function was obtained on the basis of 600 data points. For this exercise only one of the two global predictors was considered, the Leith model. Also, such a model was not again fitted to seasonal samples. The fit parameters were still the ones obtained for the whole series.

For both seasons $p_{opt} = 3$ was obtained. Figs. 9 (winter) and 10 (summer) summarize the results. We see that the global predictor always gets a better prediction than the assumption of low-dimensional chaos does. As expected, in winter season predictions are worse than in summer, due to the different climatic variability in the two periods.

4. Conclusions and discussion

A procedure for discriminating low-dimensional chaos from randomness has been discussed. The procedure is based on autoregressive processes and uses tools which are typical of the linear prediction theory. Using autoregressive predictors we obtain forecasts by means of two different strategies. In the global autoregressive approach, autoregressive models are globally fitted to all the data points at once and forecasts are obtained on the basis of the model selected. It has been shown that such a strategy works well with random processes.

In the local approach, the fitting procedure is localized, that is, the technique is based on finding local portions of a given time series in the past, that closely resemble the present and then basing forecasts on what happened immediately after these past events. It has been shown that such a strategy works well with chaotic systems.

The proper comparison between the forecast error functions obtained by both methods permits one to gain insight into the knowledge of the deep structure of the series.

The technique has been applied to one meteorological time series 40-year long (daily average air temperature) and the results can be summarized as follows.

(a) The local analysis indicates the possible presence of a low-dimensional chaos (3, 4) however,

(b) the lobal predictors are always superior to the local ones, whether we consider the whole time series or its seasonal samples,

(c) the analysis applied to the residual time series, i.e., to the series obtained by removing the annual cycle gives the same results.

In order to interpret properly these results the following remarks must be considered.

First of all, according to relation (18), the following relations should be met:

$$R > \begin{cases} 37 \cdot 365 = 13505, & d = 1 \\ 116, & d = 2 \\ 23, & d = 3 \\ 10, & d = 4, \end{cases} \tag{24}$$

in order to observe exponential growth in the forecast error function. The present value of the signal-to-noise ratio is $R \approx 140$, therefore if the dimension of the underlying attractor had been greater than 1 we would have observed an exponential growth in the local forecast error function. However, the absence of such a growth is still compatible with the presence of chaos assuming a very large value for the the greatest Lyapunov exponent. This might be compatible with the relatively large value observed for $E(1)$. A rough estimate of the order of the largest Lyapunov exponent can be obtained as follows. To fix the ideas let us suppose $d = 3$. Since the incertitude in a given initial state is of the order of the typical spacing of the data which is $\Delta \approx 0.3C$ we have that such an incertitude should grow at the first step according to $(\Delta/\sigma_x)\exp(\lambda_1) \approx E(1)$ from which $\lambda_1 \approx 1.8$. However, according to our discussion in Subsection 2.3, for such a value of λ_1 the global predictor should give extremely poor results. It has been shown in connection with the logistic map example that a value of $\lambda_1 \approx 0.44$ with $d \approx 1$ would suffice to cause the complete failure of the global predictor to give valuable forecasts. This a very important point to keep in mind. Our method is not based on observing the behaviour of the local forecast error function but on comparing it with the global one. A large value of the largest Lyapunov exponent means a large value of the metric entropy which measures the information entropy production rate. The more entropy a chaotic system produces, the more random the procees will appear to a global analysis, that is, the more unpredictable the process will be on the basis of autoregressive processes globally fitted to the data.

In the case $d = 1$, the signal to noise ratio would not be as great as we needed in order to observe exponential growth. However, the number of data points is large enough to observe $E(1) \ll 1$. But this is not the case. Again, the hypothesis of chaos is only compatible with a very large value of λ_1 (since in this case the accuracy in the initial state is limited by noise). Repeating the above computation with $\Delta \approx \sigma_{\text{noise}} \approx 0.05$, we obtain $\lambda_1 \approx 3.6$), but this is not confirmed by the behaviour of the global predictor.

As far as the seasonal sample and the residual time series are concerned the signal-to-noise ratio is lower than the one corresponding to the whole time series. However also in this case the hypothesis of chaos is only compatible with the presence of a large value of the largest Lyapunov exponent. However, again, this is not in agreement with the fact that the global predictors give always better results.

Furthermore, it should be noted that the failure of the local predictor cannot be explained on the basis of our linear fitting approach. The linearization of the approach is not crucial. Our discussion in Subsection 2.2 and the results shown in Subsection 2.3 point out that the key point to forecast successfully a chaotic system is the localization of the procedure. Farmer and Sidorowich (1987) point out that, in practice, predicting chaotic systems with non-linear maps (e.g., polynomials) does not give significant improvements over the results obtained with linear maps.

Moreover, the type of observations we have analysed must be definitely taken into account. A near surface temperature time series has been analysed. The observations could be influenced by many sources of variations which are quasi-independent of the dynamics of the atmosphere. Such sources may increase the dimensionality and the complexity of the time series therefore limiting the possibility of observing a low order chaos supposedly present in the atmosphere. The application of the technique to variables of the free atmosphere such as some large scale variables of the weather dynamics may give quite different results.

Finally, it must be also noted that both predictors agree about a low value of p_{opt}, providing that

the seasonal components have been removed from the time series. According to our simulations, $p_{opt}(\text{global}) \gg p_{opt}(\text{local})$ is expected for a chaotic signal. Conversely, when the underlying process is stochastic we expect $p_{opt}(\text{global}) \approx p_{opt}(\text{local})$. Indeed, providing that the signal is autoregressive, the global approach gives superior results over the local one, simply because it is statistically more efficient. It uses more data to get the estimation of the autoregression coefficients! In the limit $N \to \infty$ both methods should give similar results as far as the forecast error function is concerned. From Fig. 4, which illustrates the application of the procedure to an autoregressive process, it is quite clear than the two $E(l)$-curves have the same shape. Such a behaviour is also peculiar to the meteorological time series.

To sum up, the hypothesis of low-dimensional chaos in the time series at hand can be reasonable excluded. Conversely, our analysis supports the Leith conjecture about the Markov dynamics of the climatic noise. Regardless of whether the analysis were carried out on long-term period or short-term period (seasons), the predictability of the meteorological time series we analysed appears to be intrinsically limited.

REFERENCES

Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Trans. Autom. Control* **AC19**, 716–723.

Amato, U., Cuomo, V., Fontana, F. and Serio, C. 1989. Statistical predictability and parametric models of daily ambient temperature and solar irradiance: an analysis in the Italian climate. *J. Appl. Meteor.* **28**, 711–721.

Box, G. E. P. and Jenkins, G. M. 1976. *Time series analysis.* Holden-Day, S. Francisco, 525 pp.

Casdagli, M. 1989. Nonlinear prediction of chaotic time series. *Physica* **D35**, 335–356.

Eckmann, J. P. and Ruelle, D. 1985. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656.

Farmer, J. D. and Sidorowich, J. J. 1987. Predicting chaotic time series. *Rev. Phys. Lett.* **59**, 845–848.

Fraedrich, K. 1986. Estimating the dimensions of weather and climate attractors. *J. Atmos. Sci.* **43**, 419–432.

Hasselmann, K. 1976. Stochastic climate models, Part I. Theory. *Tellus A* **28**, 473–485.

Hasselmann, K. and Barnett, T. P. 1981. Techniques of linear prediction for system with periodic statistics. *J. Atmos. Sci.* **38**, 2275–2283.

Henon, M. 1976. A two-dimensional mapping with a strange attractor. *Commun. Math. Phys.* **50**, 69–79.

Katz, R. W. 1982. Statistical evaluation of climate experiments with general circulation models: A parametric time series modelling approach. *J. Atmos. Sci.* **39**, 1446–1455.

Kay, S. M. and Marple, S. L. Jr. 1981. Spectrum analysis. A modern perspective. *Proc. IEEE* **69**, 1380–1419.

Leith, C. E. 1973. The standard error of time-average estimates of climatic means. *J. Appl. Meteor.* **12**, 1066–1069.

Leith, C. E. 1975. The design of a statistical-dynamical climate model and statistical constraints on the predictability of climate. *The physical basis of climate and climate modeling.* GARP Pub. Ser. No. 16, WMO-ICSU, 137–141.

Leith, C. E. 1978. Predictability of climate. *Nature* **276**, 352–355.

Lorenz, E. N. 1963. Deterministic non-periodic flow. *J. Atmos. Sci.* **20**, 130–142.

Lorenz, E. N. 1969. Atmospheric predictability as revealed by naturally occuring analogues. *J. Atmos. Sci.* **26**, 636–646.

Madden, R. A. 1976. Estimates of the natural variability of time-averaged sea-level pressure. *Mon. Wea. Rev.* **104**, 942–952.

Mackey, M. C., Glass, L. 1977. Oscillation and chaos in physiological control systems. *Science* **197**, 287–298.

Malone, T. (ed.) 1951. *Compendium of meteorology.* Amer. Meteor. Soc., Boston.

Peterson, K. 1983. *Ergodic theory.* Cambridge University Press, Cambridge, 563 pp.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. 1986. *Numerical recipes.* Cambridge University Press, Cambridge, 818 pp.

Privalsky, V. E. 1983. Statistical predictability and spectra of air temperature over the northern Hemisphere. *Tellus* **A35**, 51–59.

Ruelle, D. 1990. The Claude Bernard lecture 1989. Deterministic chaos: the science and the fiction. *Proc. Roy. Soc. London* **A427**, 241–248.

Serio, C. 1992. Discriminating low-dimensional chaos from randomness: A parametric time series modelling approach. *Il Nuovo Cimento* **107B**, 681–702.

Trenberth, K. E. 1984a. Some effects of finite sample size and persistence on meteorological statistics, Part I. Autocorrelations. *Mon. Wea. Rew.* **112**, 2359–2368.

Trenberth, K. E. 1984. Some effects of finite sample size and persistence on meteorological statistics, Part II. Potential Predictability. *Mon. Wea. Rev.* **112**, 2369–2379.