ELSEVIER

3 November 1997

PHYSICS LETTERS A

Physics Letters A 235 (1997) 233–240

Local random analogue prediction of nonlinear processes

F. Paparella [a], A. Provenzale [a,1], L.A. Smith [b,2], C. Taricco [a], R. Vio [c]

[a] *Istituto di Cosmogeofisica del CNR, Turin, Italy*
[b] *Mathematical Institute, University of Oxford, Oxford, UK*
[c] *Dipartimento di Astronomia, Università di Padova, Padua, Italy*

Received 5 September 1996; revised manuscript received 2 July 1997; accepted for publication 23 July 1997
Communicated by A.P. Fordy

Abstract

Given that is not possible to predict the precise evolution of either stochastic processes or chaotic processes from observations, a data-based algorithm with minimal model-structure constraints is presented for generating stochastic series which are realistic, in that their long-term statistics reflect those of a process consistent with the observations. This approach employs random analogues, and complements that of deterministic nonlinear prediction which estimates an expected value. Contrasting these approaches clarifies the distinction between Lorenz's predictions of the first and second kind. Output from several nonlinear stochastic processes and observations of quasar 3C 345 are analysed; the synthetic time series have power spectra, amplitude distributions and intermittency properties similar to those of the observations. © 1997 Published by Elsevier Science B.V.

Keywords: Time series analysis; Nonlinear prediction; Dynamical reconstruction; Variability of astrophysical and geophysical systems; Stochastic systems; Deterministic systems

1. Introduction

Predicting the evolution of physical systems from observations is one of the most compelling challenges of modern time series analysis. Given an *observed* initial condition, exact prediction is rarely possible even for deterministic systems. This implies two distinct approaches within prediction: either to attempt a "best" estimate of the future, or to attempt a realistic trajectory which is consistent with the observations. Lorenz [1] labels these predictions of the first kind and second kind, respectively. By analogy, one may contrast these two approaches through the distinct aims of weather forecasting and climate studies. Weather forecasting attempts to simulate the precise trajectory the atmosphere will follow starting from the present set of observations (predictions of the first kind). Climate studies, on the other hand, are less concerned with matching the precise trajectory than with generating trajectories which are consistent with the dynamics of the atmosphere in the long run (predictions of the second kind). These goals will almost always diverge, since the "best" weather forecast is judged by some cost function (i.e. least square error) the minimum of which will not, in general, resemble a trajectory of the atmosphere (or even a trajectory of the model).

[1] E-mail: anto@icg.to.infn.it.
[2] E-mail: lenny@maths.ox.ac.uk.

PII S0375-9601(97)00607-5

Further complication arises depending on whether the process under study is itself deterministic or stochastic in nature. In general, predictions of the first kind aim at minimizing the difference between prediction and future observation for both deterministic and stochastic processes. When the initial condition is known with only finite accuracy (hereafter, an observation), deterministic prediction aims for the expected value of all trajectories consistent with the initial (uncertain) observation rather than any individual trajectory. For sufficiently large times, the predicted trajectory from an optimal (e.g. least square error) model will be constant and equal to the mean value of the observable; this trajectory is unphysical and the particular state forecast need never be observed. While a particular deterministic model may not behave in this way, the optimal [3] deterministic model will. In contrast, predictions of the second kind resemble a typical trajectory of the process in both cases; although defining "typical" requires some care in the stochastic case, since often any finite trajectory could be shadowed given some (very improbable) set of innovations.

Roughly paralleling the distinction between deterministic models and stochastic models, there are two broad literatures on data based modelling techniques. *Classical statistical approaches* construct a global stochastic model (see, e.g., Tong [2,3]) under the assumption that the dynamics are for the most part stochastic; there is usually no attempt to reproduce fine scale variations in the dynamics [4] (for an exception, see Priestley [5]). Alternatively, *deterministic dynamic reconstruction methods*, infer fine scale deterministic dynamics, assuming the complex behaviour is due to deterministic nonlinearity; extrapolation in time is thus transformed into interpolation in state-space (see Refs. [6–8] and references therein). Both approaches assume a stationary underlying process. Recent work combines aspects of both approaches, the SEQUIN model of Borland [9], for example, models both the deterministic and stochastic components of the dynamics explicitly.

In this Letter, we introduce a hybrid approach, with the aim of making predictions of the second kind for nonlinear, stochastic dynamical systems while minimizing restrictions arising from the choice of a particular model-class. Our goal is to generate a synthetic signal (or ensemble of signals) consistent with the initial condition and which reproduces the long-term statistics of "the" system (or process) that generated the data; of course in practice one can do no more than construct a model for one of the many processes which are consistent with a given finite data set. Through the random selection of a near neighbour in state-space (an analogue forecast), our algorithm exploits the deterministic nature of the process while incorporating variations in the local probability distribution function (PDF), thereby adhering to the stochastic nature of each observed trajectory. In deterministic systems, this PDF converges toward a δ-function, and the (correct) deterministic dynamics are recovered in the combined limits of data sets of infinite duration and zero observational uncertainty, although the equations of motion are never recovered explicitly.

Applications include (a) both forecasting and hindcasting, as in estimating missing observations, (b) simulation, which may for example, be used to generate stochastic input for large physical simulation models, and (c) generating stochastic surrogate data sets for use in time-series analysis [10,11]. In applications like gap filling, the aim is for a typical realization, *not* a least-square error solution which is the mean over all typical solutions.

The appeal of this approach is two-fold: its simplicity and the fact that reduces, and perhaps minimizes, the error due to mis-specification of the structure of the model. Early attempts at weather forecasting had this approach in mind; random analogue prediction (RAP) is most applicable in cases with a vast observational record and little understanding of either the process or the class of models which describe it. Forecast errors arise both from incorrectly estimating parameters of a given model and from having chosen a model whose structure is incapable of representing the underlying process. Errors due to model structure are, of course, inavoid-

[3] In the least square sense.

[4] By fine-scale we mean detailed structure in the surface defined by the expected future value as a function of location in state-space. Bi-linear, tri-linear,... models capture the large-scale structure while locally optimized local (point-wise) linear models, for example, attempt to capture the small scale structure (see Ref. [4] and the references therein).

able when dealing with physical observations, where the true nature of the process can never be known with certainty, and the observations always reflect the state of the system imperfectly. All attempts to determine explicit equations of motion from observations, whether deterministic (e.g. Refs. [6–8,12]), stochastic (e.g. Refs. [2,5,13]), hybrid (e.g. Ref. [9]), or based on large "physical simulations" (e.g. numerical weather prediction [14]) impose a model-space which restricts the processes (e.g. the functions) which can be fit. Random analogue prediction lifts this constraint, but at the hefty cost of requiring a span of observations which far exceeds the time required for the system to return near a typical point on its state-space trajectory. This constraint is shared by deterministic dynamical reconstruction methods; it was recognized by Lorenz [15], who also saw both the restrictions of analogue prediction and its applicability to inexact observations of either deterministic or stochastic processes.

2. A random analogue predictor

To obtain a local deterministic predictor [4,12,16], we transform the scalar signal s_i, $i = 1, \ldots N$, with sampling time Δt, into M-dimensional time-delay vectors, $\boldsymbol{x}_i = (s_i, s_{i-r}, \ldots, s_{i-(M-1)r})$, with delay time $\tau = r\Delta t$. For sufficiently large M, the trajectory of reconstructed vectors $\boldsymbol{x}_i$ reflects the true state-space evolution of chaotic systems [7,8,17]. Deterministic predictions assume that the trajectory of a given $\boldsymbol{x}_i$ will resemble that of "nearby" states and predict by interpolation [4,7,8,18]. Let $\boldsymbol{x}$ be the present state; to predict s', the value of the signal a time T hence, we determine the K nearest neighbours of $\boldsymbol{x}$, $\{\boldsymbol{x}^k, k = 1, \ldots, K\}$ (the superscript k does not refer to time-order) and denote the image of each vector $\boldsymbol{x}^k$ as $s^{k'}$. The predicted s' is then estimated from these K images. Specifically, for each $\boldsymbol{x}^k$ we define the scalar displacement $\Delta^k = s^{k'} - s^k$. s' is then computed as $s' = s + \Delta$ where Δ is determined by local interpolation. The predicted state vector $\boldsymbol{x}'$ is then obtained through delay coordinates. This approach works well for low-dimensional chaotic systems. It may be improved by reducing the weight given to near neighbours which are themselves neighbours in time, and by allowing K to vary with $\boldsymbol{x}$ [4,6].

How will this approach perform on a stochastic process? As a concrete example, we will consider the process

$$\mathrm{d}s(t) = [(\alpha - 0.5)\beta - s(t)]\,\mathrm{d}t + [2\beta s(t)]^{1/2}\,\mathrm{d}\epsilon, \tag{1}$$

where $\mathrm{d}\epsilon$ is an independent and identically distributed (IID) Gaussian random increment with zero mean and second order moment $\langle \mathrm{d}\epsilon \cdot \mathrm{d}\epsilon \rangle = 2\mathrm{d}t$. Process (1) defines a stationary, nonlinear first-order stochastic process; s is distributed exponentially with unit mean [13]. Numerical integration via local linearization [13] with $\Delta t = 0.01$ and $\alpha = \beta = 1$ reveals a power-law power spectrum and intermittent behaviour [19,20].

A deterministic algorithm applied to data from process (1), will predict an estimate of the mean value of the ensemble trajectories consistent with a given initial condition. As the prediction time increases, the mean value of this ensemble will approach a constant; while predicting this constant will minimize the prediction error, the dynamics of this predicted series are inconsistent with those of process (1). This has been illustrated for a global deterministic scheme applied to an AR process in Ref. [10].

Without the random term, process (1) is a first-order differential equation with a 1-D state-space. While the stochastic term formally leads to an infinite-dimensional state-space, local prediction in 1-D can completely exploit the deterministic component of the model (in absence of observational noise), and yield optimal [5] predictions of the second kind, provided an appropriate element of randomness is introduced. The approach we pursue is to alter the deterministic algorithm above by selecting one of the K neighbours at random. Thus, the displacement Δ applied to s is not determined by interpolation of the Δ^k, but rather by choosing one of them, where the probability of choosing each Δ^k is related to the distance $D^k = |\boldsymbol{x}^k - \boldsymbol{x}|$. In practice, these weights

[5] By optimal, we mean that the $P(s(t + \Delta t)|\,\boldsymbol{x}(t))$ defined by the model converge to those of the process in the limits defined in the following paragraph.

would ideally reflect observational uncertainty in the limit of large N; for simplicity, weights proportional to $1/D^k$ are used below. The value of M should reflect the dimensionality of a deterministic system or the order of a stochastic system, and the amount of data; it is treated as a free parameter when not known a priori.

Fig. 1a shows a 4096-point realisation of process (1), while Fig. 1b shows one random analogue prediction of the same length based on this data set for $M = 1$ and $K = 5$. Figs. 1c and 1d show the amplitude distribution and power spectrum of the observed series and an ensemble of synthetic series. The generalised fractal dimensions of the signals themselves (*not* their delay reconstructions) [21–23] are also similar. This close correspondence indicates that the purely data-based random analogue predictor generates realisations with statistics similar to the observed time series.

Different values of K in the range $2 < K < 10$ give similar results, and given our choice of weighting function, taking larger K has little impact. As N increases, the appropriate K will increase without bound, while $\lim_{N\to\infty} K/N = 0$. In this limit, the conditional probability distribution for $s(t + \Delta t)$ given $x(t)$ for the model will converge to that of the

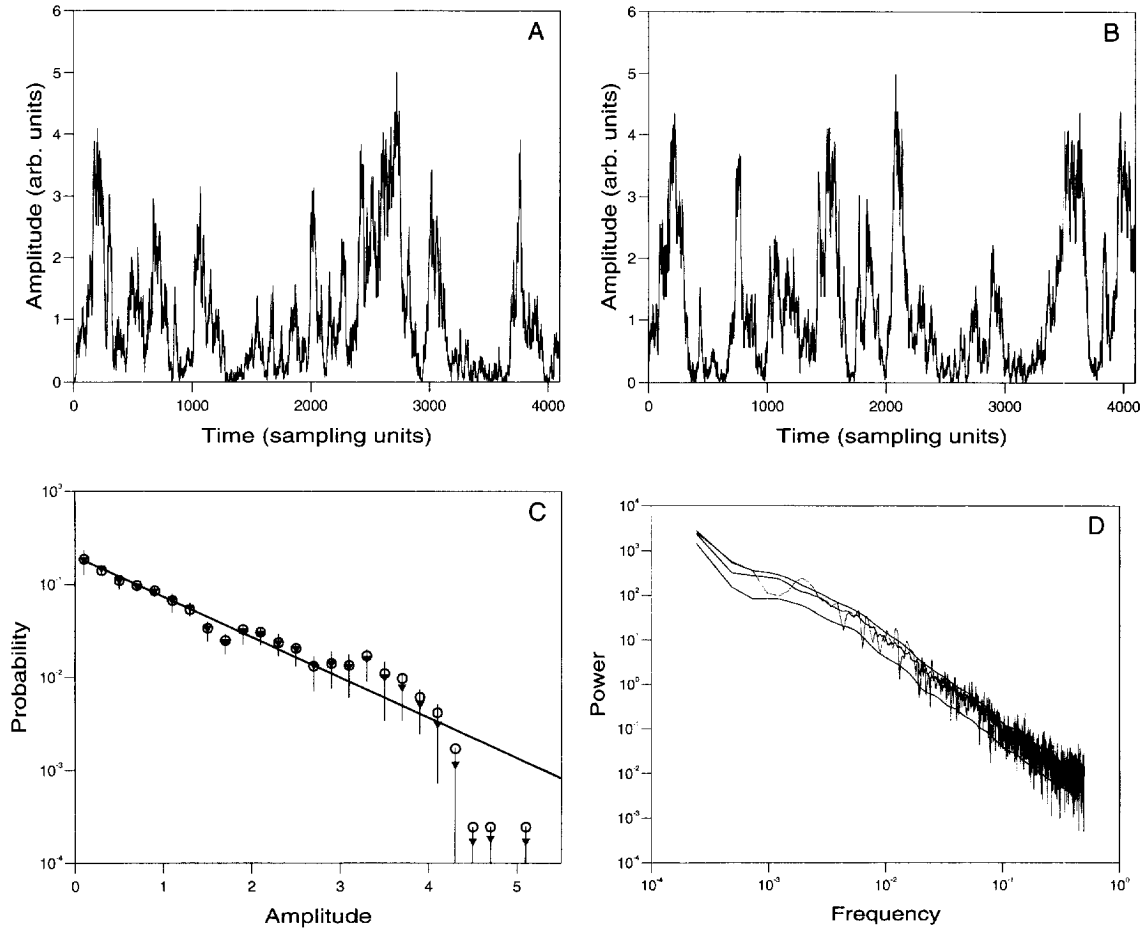

Fig. 1. (a) A 4096-point realisation of process (1). (b) One random analogue prediction ($M = 1$ and $K = 5$). (c) The probability distribution for the amplitude of the realisation (open circles), and of an ensemble of 20 RAP series (triangles). Error bars reflect 67% of the ensemble values, while the solid line indicates the theoretical distribution [21]. (d) Power spectrum of the original signal (grey line) and average spectrum from the ensemble (central solid line). The two outer lines again bound 67% of the results.

process for $M = 1$. For fixed N, the "optimal" value of K may vary with location in state-space, analogously to number of near neighbors used in local linear deterministic models [4]. In contrast, increasing M for fixed N degrades the results, consistent with the fact that process (1) is first order.

3. Two astrophysical examples

We now turn attention to two astrophysical examples, the Barnes sunspot model [24] and an observational dataset of the quasar 3C 345 [25]. Based on an ARMA (2, 2) model with nonlinear modifications to ensure that s remains positive and tends to increase more rapidly than it decreases, the Barnes model is

$$Z_i = \phi_1 Z_{i-1} + \phi_2 Z_{i-2} + a_i - \theta_1 a_{i-1} - \theta_2 a_{i-2}, \tag{2}$$

$$s_i = Z_i^2 + \alpha\left(Z_i^2 - Z_{i-1}^2\right)^2, \tag{3}$$

where $\phi_1 = 1.90693$, $\phi_2 = -0.98751$, $\theta_1 = 0.78512$, $\theta_2 = -0.40662$, $\alpha = 0.03$ and the a_i are IID Gaussian random variables with zero mean and $\sigma = 0.4$.

For $N = 4096$, good random analogues were obtained with $M = 4$. Fig. 2 shows results from both the original series and a random analogue model ($M = 4$, $r = 2$ and $K = 3$). Again, RAP generates time series with statistics similar to those of the

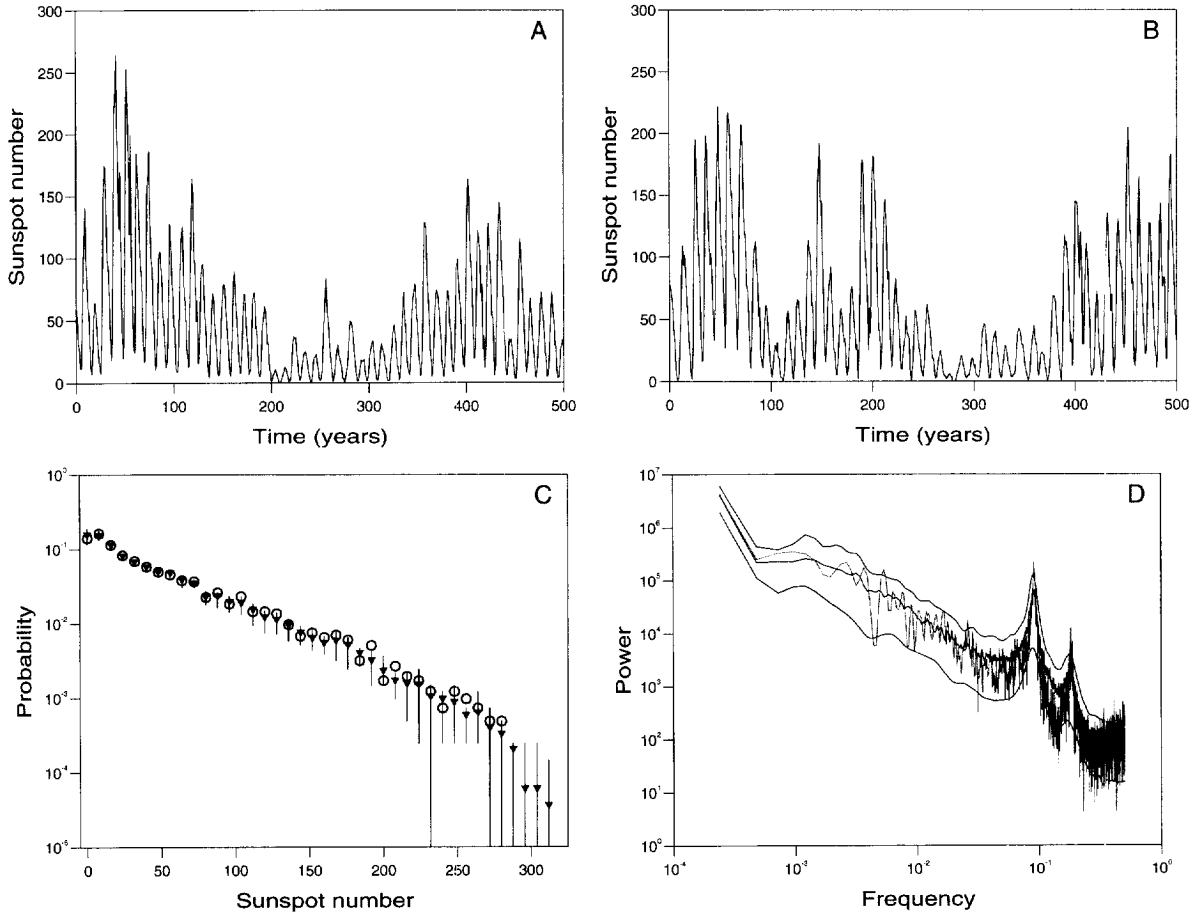

Fig. 2. (a) A signal generated by the Barnes model. (b) One random analogue prediction ($M = 4$, $K = 3$ and $r = 2$). Panels (c) and (d) are as in Fig. 1. Note that the peaks in the spectra of the RAP ensemble mean are lower and broader than that of the observed series.

process. As the Barnes model is based on an infinite order process, the best value of M will depend on the data density (i.e. N). We note, along with an anonymous referee, that the model peak at 11 years is both lower and broader than that of the process. While we do not know how to translate the shortcomings of RAP into frequency space, we point out that integrating over the peak, the model power typically falls within in the range of 0.4 to 0.8 that of the process. Further, we note that with a sampling length of $N = 4096$, the process itself exhibits fluctuations of 50% about its mean value. Hence we find the discrepancy tolerable.

Finally, we consider the 800-point light curve of the optically violent variable quasar 3C 345 [25] (Fig. 3a), which has been interpreted as consistent with a nonlinear stochastic process [20]. Typical of observed data, this series is much shorter than the numerical examples and contains several gaps. To apply the random analogue predictor, one-point gaps were filled by linear interpolation, while time-delay vectors requiring observations from longer gaps were omitted. Fig. 3 shows the results for the case $M = 2$, $r = 74$ (close to the first zero of the signal autocorrelation function) and $K = 3$. Again, satisfactory correspondence of statistical properties is obtained to

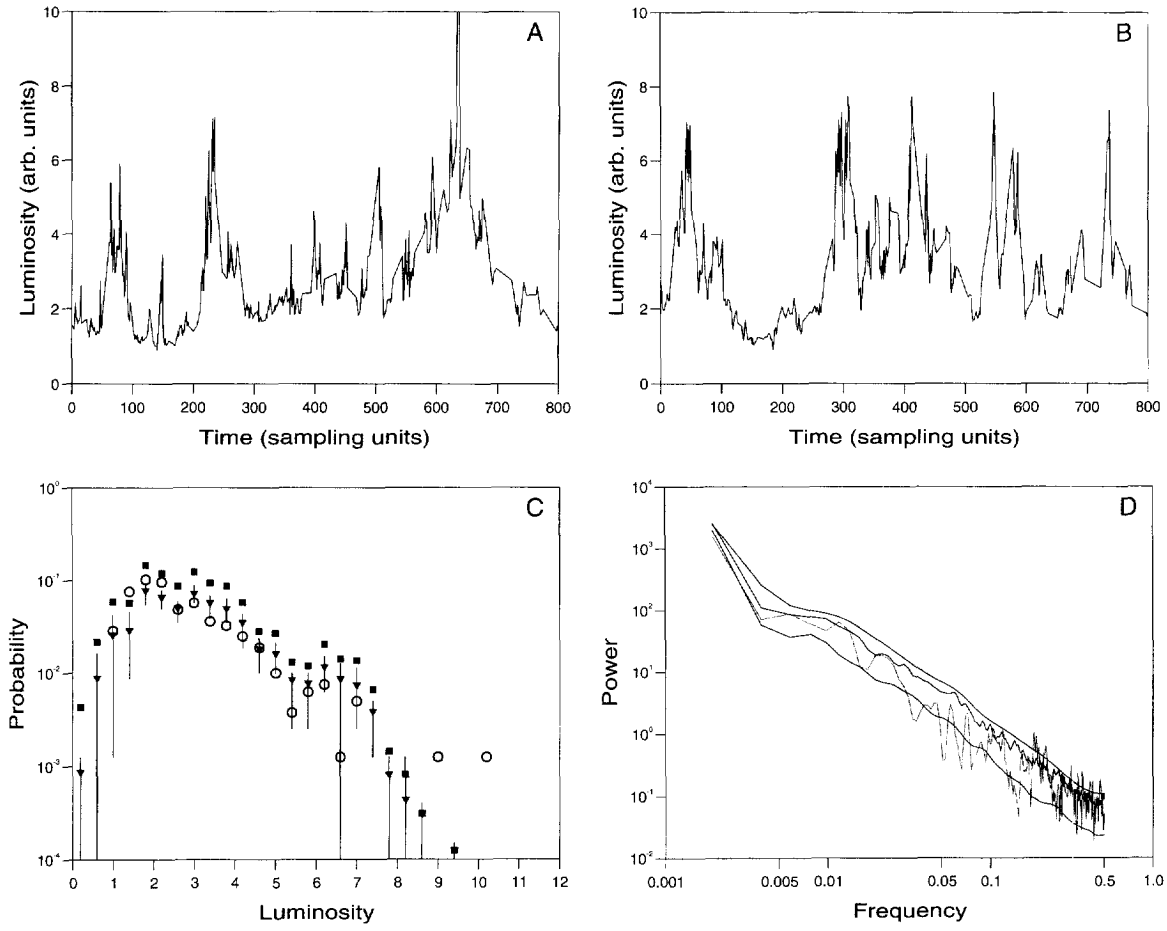

Fig. 3. (a) Light curve of the OVV 3C 345. For visualisation, gaps have been filled by linear interpolation. (b) One random analogue prediction ($M = 2$, $K = 3$ and $r = 74$). For consistency, each random analogue prediction is re-sampled as the real light curve. (c) The probability distribution for the amplitude of the realisation as in Fig. 1, here the filled squares show the mean result for the raw predictions (i.e. without imposing observational gaps). (d) Power spectra as in Fig. 1.

the extent to which they are known. This test was particularly stringent due to the extremely limited statistics available. Note that when the data density is low, trajectories become unrealistic when the predictor ventures into regions unexplored in the observations. To circumvent this, whenever the distance between the current state and its nearest neighbour in state-space exceeds half the diameter of the set, we back up 5 steps and continue with a different random seed, inspired by a similar technique introduced by Grassberger for tracing strange repellors. To illustrate the agreement between the observations and the simulations, gaps reflecting those in the observations have been introduced into the predicted data; the filled squares in Fig. 3c indicate the results when the full predicted signal is analysed.

4. Discussion and conclusions

The random analogue predictor provides time series of arbitrary length with statistical properties which resemble those of the observations, and thus a general approach to predictions of the second kind for low order, nonlinear stochastic processes. Two particular applications are (1) to estimate missing observations, as gaps are common in geophysical and astrophysical observations, and (2) to produce arbitrarily long series which, in turn, provide forcing functions for other models, as in atmospheric forcing of the ocean surface [26] or the impact on generating power plant simulations of electrical grid frequency fluctuations [27]. Work is now in progress on these issues, exploring more complex stochastic processes, and the use of ensemble prediction (see Refs. [14,28] and references therein).

The reconstruction approach to prediction translates the problem from extrapolation to interpolation; for purely data-based models, reliable results are limited to regions of state-space explored in the data set. In our case, this limitation is reflected in the inability to generate events which significantly exceed the range of the observations. For example, the observed distribution from a long realisation of process (1) will have a large amplitude tail which is not reproduced by models based on a short realisation. This simply reflects the difference between interpolation and extrapolation. No general method can both avoid spurious behaviours and realistically forecast extreme behaviours, without additional information. While information is often supplied by the functional form of a traditional model, it is difficult to justify the relevance of this restriction in regions where there are no observations.

The figures illustrate the extent to which RAP mimics the statistics of the observations. Clearly, our ultimate goal is to correctly approximate the conditional probability $P(\Delta \mid \boldsymbol{x})$ of the underlying process without assuming a functional form for either its mean as a function of location, or the (local) shape of the distribution. To the extent that we succeed, we obtain a better estimate of the process dynamics. As noted by Farmer and Sidorowich [16], local forecasting provides a test for determinism: if $P(\Delta \mid \boldsymbol{x})$ converges to a function of $\boldsymbol{x}$ for large data sets, the system is deterministic; contrasting ensemble predictions based on the models of Ref. [16] with those given in this Letter suggests a direct test for determinism in the presence of observational uncertainty. RAP has also been incorporated as a method for generating (stochastic) surrogate data [29] for Monte Carlo tests for deterministic chaos in observed time series. Here RAP provides a systematic approach to constructing what Juneja et al. [30] have called synthetic data in the context of turbulent fluid flow. If a given "chaos test" cannot identify the observed data set from within a collection of appropriately generated RAP surrogate series of the same length then the conclusion of "chaos" is questionable [10,11].

In short, RAP provides a data-based framework for analogue forecasting with minimal assumptions regarding model structure. As such, it provides a useful addition to the toolkit for the analysis of nonlinear systems.

Acknowledgement

This work has been partially supported by the EC contract EV5V-CT94-0503, "Variability of the North Atlantic Storm Track," and by CNR Project "Climate Variability and Predictability." LAS is supported by a Senior Research Fellowship at Pembroke College. We have benefited from discussions with C. Allen, M. Allen and P. McSharry, and with M. Brown of Nuclear Electric Limited on the inter-

pretation of grid frequency variations, an insightful review and additional references from an anonymous reviewer, and the comments of M. Muldoon.

References

[1] E.N. Lorenz, in: GARP Publication No. 16. Climate predictability, WMO (1975) pp. 132–136, App. 2.1.
[2] H. Tong, Non-linear time series analysis (Oxford Univ. Press, Oxford, 1990).
[3] H. Tong, Chaos and forecasting (World Scientific, Singapore, 1995).
[4] L.A. Smith, Phil. Trans. R. Soc. Lond. A 348 (1994) 371.
[5] M.B. Priestley, J. Time Series Anal. 1 (1980) 47.
[6] J.D. Farmer and J. Sidorowich, in: Evolution, learning, and cognition, ed. Y.C. Lee (World Scientific, Singapore, 1988) p. 277.
[7] P. Grassberger, T. Schreiber and C. Schaffrath, Int. J. Bifurc. Chaos 1 (1991) 521.
[8] T. Sauer, J.A. Yorke and M. Casdagli, J. Stat. Phys. 65 (1991) 579.
[9] L. Borland, Physica D 99 (1996) 175.
[10] L.A. Smith, Physica D 58 (1992) 50.
[11] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J.D. Farmer, Physica D 58 (1992) 77.
[12] G. Sugihara and R.M. May, Nature 344 (1990) 734.
[13] T. Ozaki, in: Handbook of statistics, Vol. 5, eds. E.J. Hannan and P.R. Krishnaiah (North-Holland, Amsterdam, 1985) pp. 25–83.
[14] T.N. Palmer, R. Buizza, F. Molteni, Y.-C. Chen and S. Corti, Phil. Trans. R. Soc. Lond. A 348 (1994) 459.
[15] E.N. Lorenz, J. Atmos. Sci. 20 (1963) 130.
[16] J.D. Farmer and J. Sidorowich, Phys. Rev. Lett. 59 (1987).
[17] F. Takens, in: Dynamical systems and turbulence, Vol. 898, eds. D. Rand and L.-S. Young, (Springer, Berlin, 1981) p. 366.
[18] A. Weigend and N. Gershenfeld, eds., SFI studies in complexity, Vol. 15. Predicting the future and understanding the past: a comparison of approaches (Addison-Wesley, Reading, MA, 1993).
[19] A. Provenzale, L.A. Smith, R. Vio and G. Murante, Physica D 58 (1992).
[20] R. Vio, S. Cristiani, O. Lessi and A. Provenzale, Astrophys. J. 518 (1992) 391.
[21] B.B. Mandelbrot, The fractal geometry of nature (Freeman, San Francisco, 1985).
[22] A. Provenzale, B. Villone, A. Babiano and R. Vio, Int. J. Bifurc. Chaos 3 (1993) 729.
[23] K.R. Sreenivasan and C. Meneveau, J. Fluid Mech. 173 (1986) 357.
[24] J.A. Barnes, H.H. Sargent and P.V. Tryon, in: The ancient sun, eds. R.O. Pepin, J.A. Eddy and R.B. Merrill (Pergamon, New York, 1980) pp. 159–163.
[25] R. Vio, S. Cristiani, O. Lessi and L. Salvadori, Astrophys. J. 380 (1991) 351.
[26] M.A. Balmaseda, D.L.T. Anderson and M.K. Davey, Tellus A 46 (1994) 497.
[27] M. Brown, private communication.
[28] L.A. Smith, in: Predictability, ed. T. Palmer (Reading, UK, 1996) ECMWF.
[29] L.A. Smith, in: Mixing transport and the environment, eds. K. Beven, P.C. Chatwin and J.H. Millbank (Wiley, London, 1994) pp. 383–399.
[30] A. Juneja, D.P. Lathrop, K.R. Sreenivasan and G. Stolovitzky, Phys. Rev. E 49 (1994) 5179.