Operational Oceanography . Implementation at the European and Regional Scales
Proceedings of the second international Conference on EuroGOOS 11 - 13 March 1999, Rome, Italy
Elsevier.ISBN: 0-444-50391-9

Evolutionary computation for climate and ocean forecasting: "El Niño forecasting"

A. Álvarez [a,b,c] , P.Vélez [b,c] , A.Orfila [b,c] , G.Vizoso [b,c] and J. Tintoré [b,c]

[a] Departament of Physics, National Central University, Chung-Li, Taiwan, ROC
[b] Instituto Mediterráneo de Estudios Avanzados (CSIC-UIB) E-07071 Palma de Mallorca, Spain
[c] Departament de Física, Universitat de les Illes Balears, E-07071 Palma de Mallorca, Spain

ABSTRACT

It has been recently postulated that the irregular dynamics of El Niño-Southern Oscillation (ENSO) may embed a low order chaotic process{1,2}. If true, some aspects of the ENSO variability are predictable. Here, from observations of the annual average Sea Surface Temperature (SST) of the tropical Pacific Ocean (1949-1981) and using an evolutionary genetic analysis, we characterise the chaotic nature of ENSO. The mathematical model of the ENSO dynamics obtained is cross-validated with the observations for the next fifteen years (1982-1997).

The extracted ENSO dynamics show a chaotic behaviour characterised by a correlation dimension d=3.5, in agreement with numerical simulations{2}. An eight years forecast (1997-2005) of the ENSO shows a strong cooling (La Niña) during 1999 and a significant El Niño in 2001. Finally, a longer term forecast reveals a warming of about 0.2 Celsius per century that might be associated with very low frequency variability of ENSO.

1. INTRODUCTION

Traditionally, the forecasting of the climate system proceeds along the following line: deterministic equations of motion are derived from first principles, initial conditions are measured and the equations of motion are integrated forward in time. However, there are situations where a first principle model is unavailable or initial conditions are not accessible. In these situations, predictions can still be inferred from the time series of observations of the climate system.

The irregular behaviour of the climatic time records is assumed to result from complicated physics with many degrees of freedom, and is therefore analysed as a random process. However, in some cases, the apparent randomness may be due to a chaotic behaviour of nonlinear but deterministic interactions between few degrees of freedom. In such cases, it is possible to find, directly from the data, a dynamical model of the interactions and as a result, exploit the determinism to make short-term forecasts.

Operational Oceanography . Implementation at the European and Regional Scales
Proceedings of the second international Conference on EuroGOOS 11 - 13 March 1999, Rome, Italy
Elsevier.ISBN: 0-444-50391-9

In this paper, we present a new approach to climate and ocean forecasting based on evolutionary computation as well as its application to the forecast of "El Niño" phenomenon. In Section II, we briefly detail the physical basis of the method while Section III shows the results obtained with the method to predict future El Niño events.

2. BUILDING A DYNAMICAL MODEL FROM TIME SERIES

Building a dynamical model directly from the data involves the following two steps:

2.1.State space reconstruction

The goal of state space reconstruction is to use the immediate past behaviour of the time series to reconstruct the current state of the system. We assume that the time series $\{x(t_i)\}$, x_i=1,...,N is generated by a dynamical system,

$$\begin{aligned} \frac{ds}{dt} &= F(s(t)) \\ x(t) &= h(s(t)) \end{aligned} \qquad \textbf{(Eq.1)}$$

where s is a state of the phase space of the system, h a scalar-valued measurement function and F a nonlinear function. Our aim is to construct the state space of the dynamical system represented by Eq.(1). The past behaviour of the time series contains information about the present state. If a sample time τ is considered, this information can be represented as a delay vector of dimension **M** (embedding dimension)[3]:

$$\vec{x}(t) = (x(t),...,x(t-(m-1)\tau)) \qquad \textbf{(Eq.2)}$$

For adequate values of time delay τ and embedding dimension M, it has been proved that the map that takes the original state $S(t)$ to the delay vector $\vec{x}(t)$ preserves some dynamical invariants from the original dynamics: that is, the space of vectors $\vec{x}(t)$ shows the same properties as the original space state [4]. Even more, it can be also shown the existence of a smooth mapping :

$$P : R^m \rightarrow R$$

that satisfies [4],

$$\vec{x}(t) = P(x(t-\tau), x(t-2\tau),...,x(t-m\tau)) \qquad \textbf{(Eq.3)}$$

In short, if the mapping P is known, we will be in disposition to make predictions of the future behaviour of the system. In the next subsection we will describe a technique to estimate P.

2.2.Mapping estimation by evolutionary computation

We propose to employ an evolutionary computation approach to determine the mapping P involved in Eq.(3).

Operational Oceanography . Implementation at the European and Regional Scales
Proceedings of the second international Conference on EuroGOOS 11 - 13 March 1999, Rome, Italy
Elsevier.ISBN: 0-444-50391-9

Specifically, the evolutionary algorithm consists in a population of potential solutions of the problem (mathematical expressions that attempt to estimate the mapping *P*) that are subjected to an evolutionary process involving mutation, reproduction and survival of the strongest individuals characterised by those which best fit the environmental conditions, i.e. the data.

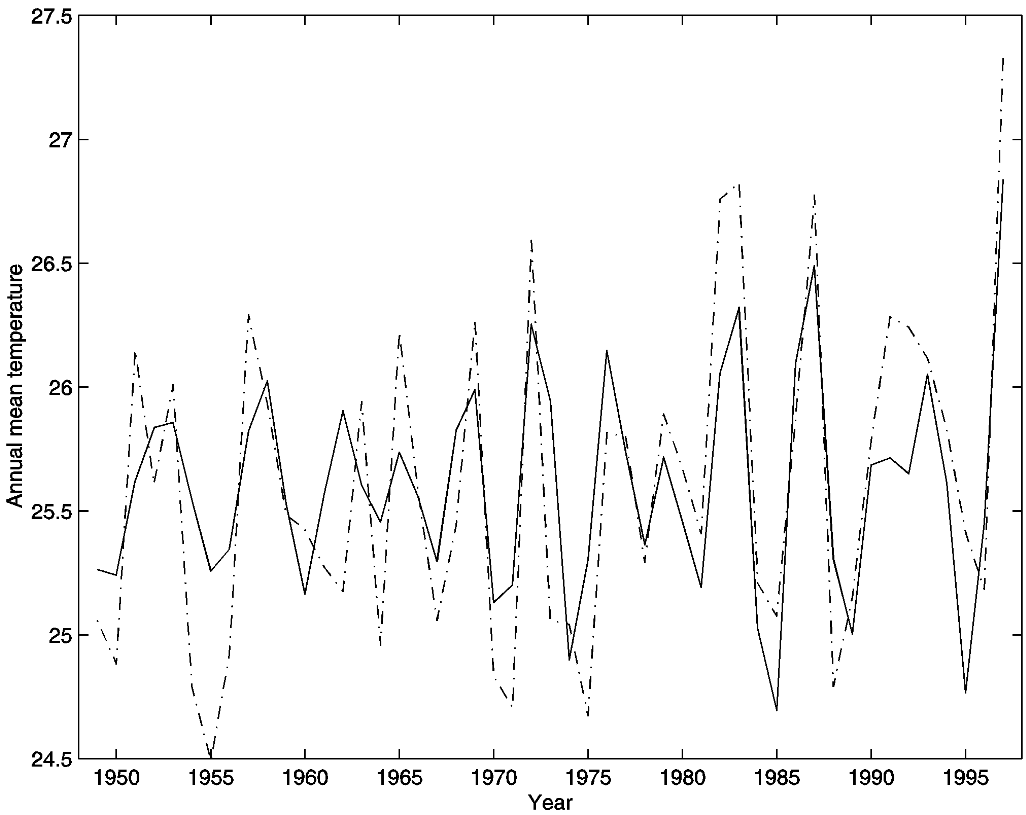
Evolutionary algorithms have demonstrated to be powerful methods in searching and optimisation{5,6,7}.

3. CLIMATIC APPLICATIONS: "EL NIÑO" FORECASTING

The above described procedure of building models from data has been applied to an annually and spatially averaged time series of temperature at the tropical Pacific ocean built from the Japan Meteorological Agency (JMA) Index. The index is a spatially averaged sea surface temperature (SST) anomalies over an area in the Pacific Ocean located between 4ºS-4ºN, 150º W-90º W. For the period from 1949 to present, the JMA-Index is based on observed data.

To discriminate signal and noise in the records, the original time series (dash-dot line in Fig.1) has been filtered employing the Monte Carlo Singular Spectral Analysis (MCSSA){8}. The filtered signal (solid line in Fig.1) accounts for the deterministic variability that represents over the 60% of the original variability.

The space state reconstruction is then applied to the filtered time series. The time delay τ is computed using an extended criterion that fixes the time delay as the time at which the first zero of the covariance function of the time series occurs{3}. In our specific case the first zero of the covariance function occurs at τ=1 yr.



[Fig.1] Real (dash-dot line) and filtered (solid line) temperature time series.

Operational Oceanography . Implementation at the European and Regional Scales
Proceedings of the second international Conference on EuroGOOS 11 - 13 March 1999, Rome, Italy
Elsevier.ISBN: 0-444-50391-9

The choice of the dimension M in which to embed the time series is not obvious. Vatuard *et al.* {9} suggest that the Singular Spectral Analysis is typically successful with an embedding dimension M=N/3 where N is the length of the time-series. As a trial we first consider the embedding dimension M=8. This embedding dimension allows to capture some basic dynamical behaviours such as self-sustained periodic oscillations, quasi-periodic oscillations or even chaotic oscillations.

Once the time delay τ and embedding dimension M are established, we proceed to estimate the mapping P in Eq.(3).

To find the mapping P, we divide the data set into two subsets, one ranging from 1949 to 1981 and the other from 1982 to 1997. The first subset is used as a training set in the symbolic regression method carried out by an evolutionary algorithm {7}. The evolutionary algorithm built to find the mapping P in Eq.(3) has been configured in such a way that the maximum number of symbols allowed for each tentative equation is 20. Each generation consists of a population of 120 equations. The initial population of equations is randomly generated. After 10000 generations, the symbolic expression that best maps the training set is:

$$T(t) = T(t-8) + \frac{T(t-5)(T(t-2) - T(t-6))}{T(t-4)}\left[\left(\frac{T(t-2)}{T(t-4)}\right)^2 - 2.650\right] \qquad \textbf{(Eq.4)}$$

We have studied the validity of Eq.(4) as representative of the ENSO dynamical behaviour contained in the time series, by cross-validating this mapping with the testing subset, the time series from 1982-1997. Explicitly, the cross-validation consists of a direct comparison between the prediction from Eq.(4) and the real data for this period of time. Such a cross-validation is the usual way to validate neural networks and genetic algorithms to discern over-fitting cases in the training set. Figure 2 presents the results from both the training and testing data sets.

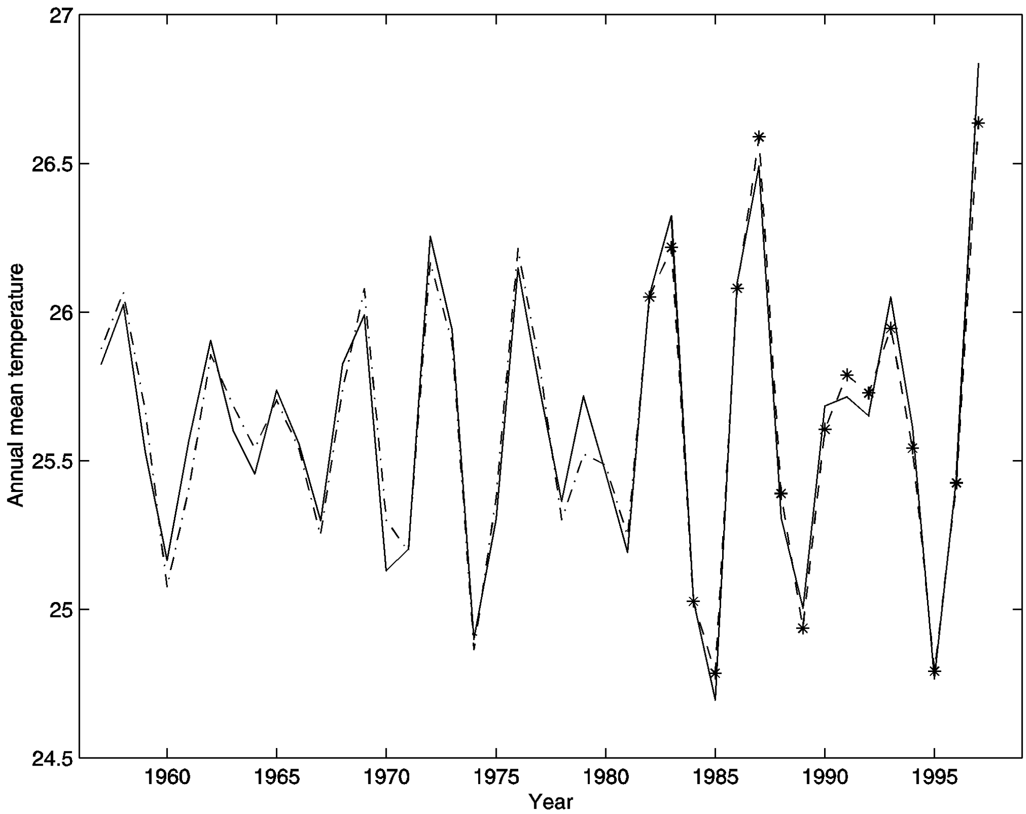


Fig.2. Filtered data (solid line), results from the training data set (dashed-dotted line), and results of the prediction of (Eq.4) in the testing data set.

Operational Oceanography . Implementation at the European and Regional Scales
Proceedings of the second international Conference on EuroGOOS 11 - 13 March 1999, Rome, Italy
Elsevier.ISBN: 0-444-50391-9

The prediction of Eq.(4) resulting from the training is represented by the stars, explaining 98.4% of the total variability of the filtered series.

The excellent agreement between model predictions and real data reveals the possible applicability of Eq.(4) as a short-term forecast model of ENSO events. Figure 3 shows the result obtained from iterating Eq.(4) from 1997-2005. We find the existence

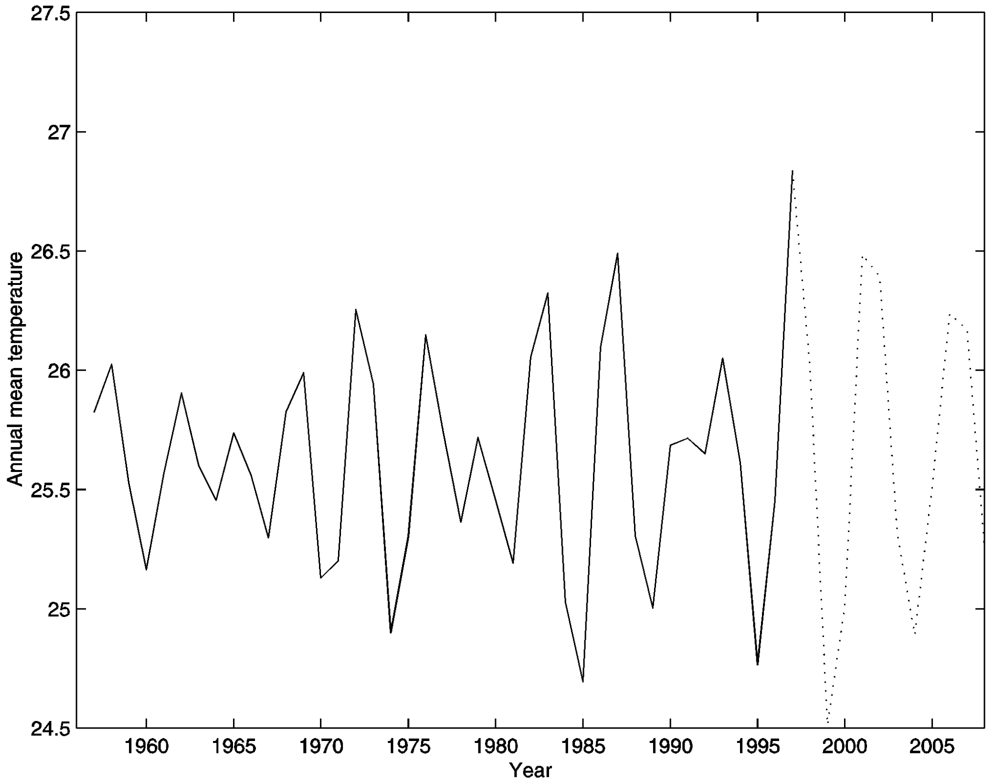


Fig.3. Forecast of the mean annual SST of the tropical Pacific Ocean for the next 8 years obtained from (Eq.4). Remarkable facts are the prediction of a strong cooling in 1999, the forecast of the next ENSO event in the year 2001 and its posible persitence during 2002.

of a strong cooling during 1999 as well as the prediction of an ENSO event in the year 2001. The SST anomalous warming trend will still persist during the year 2002. It is important to point out that Eq.(4) deals with 60% of the total SST variability. As a result, the magnitude of the predicted surface warming associated with ENSO are estimations from the deterministic part of the ENSO dynamics but still 40% of the total variability remains unexplained.

Based on the high forecasting performance obtained in the cross-validation set, we expect Eq.(4) to be an estimator of future ENSO events.

4. REFERENCES

1. Jin, F. F., Neelin, J. D., & Ghil, M. El Niño on the Devil's Staircase: Annual subharmonic steps to chaos. Science, 264, 70 (1994).
2. Tziperman, E., Stone, L., M. A. Cane, M. A. & Jarosh H. El Niño chaos: Overlapping of resonances between the seasonal cycle and the Pacific ocean-atmosphere oscillator. Science, 264, 72-74 (1994).
3. Abarbanel, H. D. I., Brown, R., Sidorowich, J. J. & Tsimring, L. Sh. The analysis of observed chaotic data in physical systems. Rev. Modern Phys., 65, 1331-1392 (1993).
4. Takens, F. Detecting strange attractors in turbulence. In Dynamical systems and turbulence. Springer-Verlag, Berlin (1991).
5. Goldberg, D. E., Genetic algorithms in search, optimization and machine learning. Addison-Wesley Longman, Inc., 412 pp. (1989).
6. K. R. Kozac, Genetic programming MIT Press, 745 pp. (1994).
7. Szpiro, G. G. Forecasting chaotic time series with genetic algorithms. Phys. Rev. E, 55, 2557-2568 (1997).
8. Allen, R. M. & Smith L. A. Monte CarloSSA: Detecting irregular oscillations in the presence of colored noise. J. Climate, 9, 3373-3404 (1996).
9. Vatuard R.,Yiou P. And Ghil M., Singular Spectrum Analysis: A toolkit for short, noisy and chaotic series. Physica D,58, 95-126 (1992) .