Satellite based forecasting of sea surface temperature in the Tuscan Archipelago

A. ALVAREZ[†], A. ORFILA[‡] and J. SELLSCHOPP[†]

[†]SACLANT Undersea Research Centre, Viale San Bartolomeo,
19138 La Spezia, Italy

[‡]Instituto Mediterraneo de Estudios Avanzados-IMEDEA, CSIC-UIB,
07071 Palma de Mallorca, Spain

(Received 11 May 2001; in final form 13 November 2001)

Abstract. The system described employs a nonlinear forecasting technique based on a combination of genetic algorithms and empirical orthogonal function (EOF) analysis. The genetic algorithm identifies the equations that best describe the behaviour of the different temporal orthogonal functions in the EOF decomposition and therefore, enables global forecasting of future time variability. The method is applied to obtain a one-month ahead forecast of the monthly mean space-time variability of the sea surface temperature (SST) of the Tuscan Archipelago, northwest coast of Italy. The system performance has been validated comparing forecast fields with real satellite observations. Results indicate that the system provides better predictions than those based on climatology. Future research is oriented to make the system applicable to military operations, environmental control and fisheries activities.

1. Introduction

The Tuscan Archipelago consists of a group of islands located on the continental platform of the northwest coast of Italy, figure 1. The boundaries of this oceanographic region are defined by the Tyrrhenian Sea to the south, the mainland to the east, the island of Corsica to the west and the Ligurian Sea to the north. Each of these boundary areas is characterized by well defined oceanographic and atmospheric processes that influence the local oceanographic conditions of the area. The archipelago constitutes a transition oceanographic region with a particular spatiotemporal variability (Astraldi *et al.* 1990, 1993). Previous studies reveal that a common oceanographic feature is a northward flow of Tyrrhenian waters through the region. This flow is induced by the dynamic links between the Liguro-Provençal basin and the Tyrrhenian sea. Besides this large-scale circulation pattern, quasi-permanent mesoscale structures are present in the southern boundary of the archipelago. The most remarkable feature in this boundary is the existence of a cyclonic eddy off the Strait of Bonifacio (Krivosheya 1983). Most studies indicate that the origin of this gyre is a consequence of the year-round jet of wind blowing eastward through the Strait of Bonifacio (Moskalenko 1983). The gyre undergoes significant seasonal changes (Astraldi and Gasparini 1994). In summer, the cyclonic gyre is zonally oriented with

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2002 Taylor & Francis Ltd
<http://www.tandf.co.uk/journals>
DOI: 10.1080/01431160210144615

Figure 1. Geographic location of the Tuscan Archipelago.

intrusions on the corresponding shelf of the Italian and Corsican coasts. In winter, the gyre is stretched in a meridian line in the western part of the basin. At the northern boundary of the archipelago, the outflow of Tyrrhenian waters converges with the West Corsica Current generating an active frontal area (Astraldi *et al.* 1994). The front is associated with meanders and small eddies, mainly anticyclonic, which induce smaller scaled parcels and frontal intrusion in the archipelago. This eddy

activity resides in this region throughout the year. Besides its oceanographic interest, the Tuscan Archipelago represents an important part of the Ligurian Cetacean Sanctuary. For this reason, the characterization and prediction of the space-time variability of this marine environment constitutes an essential part of its environmental policy (Astraldi *et al.* 1995, Carniel *et al.* 2001).

Numerical ocean models are the most common tools for ocean prediction. The numerical approach forecasts oceanic evolution by integrating forward in time the equations of ocean dynamics. In general, numerical ocean models require continuous updating with real data in order to correct the deviations between model results and reality due to the nonlinear nature of the ocean. This data requirement is rarely available. Research ships allow reasonable spatial coverage but are limited by the relatively low frequency with which they can be deployed. On the other hand, measurements from coastal stations and buoys have high temporal resolutions but often no spatial dimensions. Satellite imagery constitutes the only way to monitor space-time oceanographic variability at high spatial and temporal resolutions simultaneously. Sea surface temperature, ocean colour, sea surface level winds, waves and, to a certain extent, currents can be measured in real time and employed to nowcast present ocean states (Vazquez-Cuervo *et al.* 1996, Scott 1997, Garcia-Gorriz and Vazquez-Cuervo 1999, Cole and Villacastin 2000). The nowcasting of satellite data could be extrapolated forward in time by an adequate forecast of the satellite information. This goal could be achieved by developing ocean forecasting systems based on physical and mathematical approaches different to those used in traditional ocean modelling. One of these approaches could consist of forecasting satellite-observed data by computing the mathematical model that best explains the behaviour of a given time series of satellite observations (Alvarez *et al.* 2000a, b).

This article describes a satellite based ocean forecasting system and its application to forecast the surface ocean conditions in the Tuscan Archipelago.

2. Methodology

Forecasts of satellite data are obtained in three major phases (figure 2): encoding satellite observed data, noise reduction and prediction of time variability.

2.1. Encoding satellite observed data

In principle, prediction of the satellite observed fields implies determination of the time evolution of each pixel. As this is unrealistic, an encoding procedure is required in order to reduce the complexity of the problem. A convenient way to encode satellite images into a small set of numbers is the use of the Empirical Orthogonal Function (EOF) technique (Preisendorfer 1988). This technique has been widely used in the analysis of Advanced Very High Resolution Radiometer (AVHRR) Sea Surface Temperature (SST) images to study the space-time variability associated with frontal, eddy, and jet structures in the ocean (Lagerloef and Bernstein 1988). Briefly, the EOF technique decomposes space- and time-distributed satellite data into modes ranked by their temporal variance. A set of spatial modes and associated mode amplitude functions are obtained. The spatial modes provide information on the spatial structures while the amplitude functions describe their dynamics. Temporal variance EOF modes were employed by Alvarez *et al.* (2000a) to predict SST patterns in the Alboran Sea. In this work, EOF modes have been ranked by their spatial variance. The space and time axes are reversed relative to the EOF decomposition of temporal variance: the amplitude functions become spatial

Figure 2. Flow chart of the satellite based forecasting system.

image patterns and the eigenvectors become dimensionless time series describing the time histories of the modes. While predictability is not affected by which variance (temporal or spatial) is decomposed (Alvarez 2001), spatial variance decomposition is more suitable to the extraction and prediction of small scale features (Lagerloef and Bernstein 1988). Besides, encoding with spatial variance EOF modes facilitates

the extraction and modelling of the deterministic part of the associated time variability. In the present approach, the space-time variability in the time series of satellite data can be explained as a combination of a small number of time dependent EOFs and their respective spatial amplitude functions.

2.2. Noise reduction

Each one-variable time-series EOF, will be contaminated by noise from the measurement process. In addition to the measurement noise, the EOFs of small variance, which are neglected in the previously described processes, introduce a stochastic component into the data. Noisy data affects prediction performance as the predictor attempts to predict the noise (i.e. find a dynamical law of a random effect) at the expense of predicting the true underlying dependence. Filtering to isolate the deterministic signal in the data from stochastic noise, should be applied (Broomhead and Lowe 1987). An adequate approach to remove noise without losing a significant portion of the deterministic signal is the Singular Spectral Analysis (SSA) or data adaptive approach (Penland *et al.* 1991, Casdagli *et al.* 1992). Briefly, for the j -EOF we form a trajectory matrix $\mathbf{X}$ the rows of which contain w -dimensional vectors of the form $(A_j(t - \Delta t), A_j(t - 2\Delta t), \dots, A_j(t - w\Delta t))$, $(A_j(t - 2\Delta t), A_j(t - 3\Delta t), \dots, A_j(t - (w+1)\Delta t))$, when Δt is the time step between observations and w the window of the filter. The covariance matrix, $\mathbf{X}^T \mathbf{X}$ is computed and diagonalized to obtain a set of eigenvectors which are the orthogonal singular vectors. The corresponding eigenvalues are the average root-mean-square projection of the w -dimensional vectors that constitute the trajectory matrix. This procedure can be useful in the presence of noise; discarding eigenvalues the magnitude of which is below the noise level can reduce noise, thus distinguishing the deterministic part of the signal from the noise. A new filtered time series can be reconstructed considering only the eigenvectors with eigenvalues above the noise level.

2.3. Prediction of time variability

The aim is to obtain a mathematical model for each representative EOF that using past values of the time series predicts the future ones. Explicitly, given a deterministic time series $\{x(t_i)\}$, $t_i = 1 \dots N$, Takens' theorem (Takens 1981) establishes the existence of a smooth map $P: R^m \rightarrow R$ satisfying:

$$x(t) = P(x(t-\tau), x(t-2\tau), \dots, x(t-m\tau)) \quad (1)$$

where m is called the embedding dimension obtained from a state-space reconstruction of the time series and τ is a time lag unit (Casdagli *et al.* 1992). During the past decade, various techniques have been developed to accomplish the task of approximating the mapping $P(\cdot)$ defined in equation (1). In this study, we have employed the genetic algorithm developed by Alvarez *et al.* (2001) to accomplish the task of approximating the mapping $P(\cdot)$. Genetic algorithms were introduced by Holland (1975), and have been used in such diverse areas as engineering (Goldberg 1989), biology (Lucasius *et al.* 1991), acoustics (Gerstoft 1995) and economics (Bauer 1994), among others. In our case, for the filtered j -EOF $\{\tilde{A}_j(t_i)\}$, $t_i = 1 \dots N$ a set of N_{pop} candidate equations (the population) for P is randomly generated. These equations (individuals) are of the form of equation (1) and their right hand sides are stored in the computer as sets of character strings with a general form:

$$S = ((A \otimes B) \otimes (C \otimes D)) \quad (2)$$

where A , B , C , and D are either real numbers (i.e. parameters of the function) or input variables, $(\hat{A}_j(t - \Delta t), \hat{A}_j(t - 2\Delta t), \dots, \hat{A}_j(t - m\Delta t))$, and $\otimes$ stands for one of the four basic arithmetic operators ($+$, $-$, $\times$, and $/$). Other mathematical constructs are conceivable, for instance, the log and exp functions (with one argument). For simplicity in this work, we limit ourselves to the four arithmetic operators. Operators and arguments are assigned at random, with numerical values chosen from a finite set of real numbers, uniformly distributed in $[-Z, Z]$ where Z is a real number. In this study a specific value of $Z = 10$ was employed. Notice that limitations on allowed real arguments and operators considerably reduce the dimensionality of the searching space, facilitating the optimization process carried out by the genetic algorithm. No substantial differences have been found in the results if representation schemes more complex than equation (2) are considered to initialize the population. In general, mathematical structures more complex than the initial ones are generated as the evolution proceeds from the initial population.

A criterion that measures how well the equation strings perform on a training set of the data is defined as:

$$\Delta_i^2 = \sum_{t=L+1}^T (\hat{A}_j(t) - P^i(\hat{A}_j(t - \Delta t), \hat{A}_j(t - 2\Delta t), \dots, \hat{A}_j(t - m\Delta t)))^2 \quad (3)$$

where P^i represents the i -equation string of the population, $L = m\Delta t$ and T is the total length of the j -EOF $\hat{A}_j(t)$. This criterion, the fitness to the data, establishes the strength of each individual. The strongest individuals, equation strings with best fits, are then selected to exchange parts of the character strings between them (reproduction and crossover) while the individuals less fitted to the data are discarded. Reproduction starts by ranking the individuals in descending order of their strength and discarding the $N_{\text{pop}}/2$ weakest ones. The remaining members of the population are organized into $N_{\text{pop}}/4$ pairs of the form [individuals 1&2], [individuals 3&4]... The proper reproduction involves information crossover between the individuals constituting a pair. The first two offspring are identical to their parents. The two other offspring are formed as recombinations of their parents. The recombination process is defined by the following steps:

- Self-contained parts of the functional relations $P(\cdot)$ are randomly chosen. These self-contained parts are either a randomly chosen datum or real number, or a randomly chosen opening parenthesis and all the contents of the string until the corresponding closing parenthesis. As an example, let us assume the following parents' individuals:

$$\begin{aligned} \text{Parent 1: } & \frac{(A+B)}{C} \\ \text{Parent 2: } & A' - (B' * C') \end{aligned} \quad (4)$$

where $A, B, C, \dots, A', B', \dots$ are either real numbers or input variables. In the first step of the recombination process, self-consistent parts of the parents' individuals are selected (marked in bold). As indicated, these selected parts can be either a real number or datum like C in *Parent 1*, or a mathematical consistent part of the functional relation like the term $(B' * C')$ in *Parent 2*.

- Self-contained parts are then interchanged between the parents' individuals originating new functional relations. In the proposed example, the two offspring obtained by crossover of *Parent 1* and *Parent 2* would be of the form:

$$\begin{aligned} \text{Offspring 3: } & \frac{(A+B)}{(\mathbf{B'} * C')} \\ \text{Offspring 4: } & \mathbf{A'} - C \end{aligned} \quad (5)$$

where bold parts indicate the exchanged self-consistent elements.

Notice that the population size, N_{pop} , is kept constant by the recombination process. The functional structures in the offspring generated by crossover can be more complex than those of the parents' individuals. However, the total number of characters, operators plus arguments, in each functional expression is generally bounded to avoid the generation of very long functional expressions.

Finally a small percentage of the equation strings' most basic elements, single operators and variables, are mutated at random. As a result of this process a new set of equations (generation) with better fitness properties is obtained. The process is repeated a large number of times, to improve the fitness of the evolving population. At the end, the strongest individual represents the best functional relation found for the given amplitude function. Future values of the amplitude function can be obtained from these functional relations. As a byproduct, the genetic algorithm can indicate the functional form that underlies the data (Alvarez *et al.* 2001).

Prediction of the satellite observed field is achieved by adding the most relevant spatial modes previously multiplied by their corresponding forecast EOFs.

3. Data

A time series of 94 monthly averaged SST images of the Tuscan Archipelago, from March 1993 to December 2000, has been considered. Each monthly image is composed based on the daily maximum images using the average for every single pixel's position. The monthly composition normally consists of approximately 160 AVHRR passes. Each image is constituted by 207 pixels $\times$ 311 pixels corresponding to a spatial resolution of 1.1 km. Several tests ensure that SST values are derived only for cloud free water surfaces. All pixels flagged as cloud are excluded from further processing. Details on cloud cleaning and image processing can be found at <http://isis.dlr.de>. The data set is an AVHRR MCSST product from the German Aerospace Research Center-DLR. The monthly satellite images from March 1993 to December 1998, a training set, was employed to determine the mathematical laws governing this oceanographic area and the prediction functions. Validation of the prediction system was carried out using the satellite data from January 1999 to December 2000, the validation set. The predictive relations found in the training set (March 1993–December 1998) are employed to obtain one-month ahead predictions of the mean SST and surface current fields in the validation set (January 1999–December 2000). During this period, predictions are validated with the corresponding real observations.

4. Results

In order to study the variance associated with front, eddy and jet structures of the oceanographic region, EOF modes were ranked by their spatial variance. Amplitude functions are then given by spatial image patterns and the eigenvectors

Figure 3. (a) First, (b) second and (c) third spatial modes.

are dimensionless time series describing the time histories of the modes. The first spatial mode, figure 3(a), accounts for 49 per cent of the spatial variability. This dominant pattern shows a front and eddy structures in the north and southern boundaries respectively. The northern front, aligned north-south, separates the warmer waters of the continental shelf in the eastern part from the colder waters of the Ligurian sea. The location of the front over the continental slope, suggests that its origin could be related with the flow in the area of the Western Corsica Current. The eddy structure is located in the southern boundary off the Strait of Bonifacio.

Unlike the northern front, cold water masses are direct consequences of the year-round jet of wind blowing eastward through the Strait of Bonifacio. Strong temperature gradients are induced by the low water temperatures achieved at the core of the eddy, where upwelling is present. The second spatial mode figure 3(b), representing 12 per cent of the spatial variability, is associated with the influence of the river Arno outflow to the north and the zonally oriented modulation of the eddy structure to the south. The third spatial mode figure 3(c) accounts for 7 per cent of the spatial variability and describes north-south subbasin temperature gradients.

Only the first three spatial structures have been considered dynamically significant. Following Alvarez *et al.* (2000a, b), the dynamical significance of these modes was indicated by the predictability of the corresponding time variability. Explicitly, the criterion to identify the amount of determinism for the time variability contained in each EOF, was based on a nonlinear prediction approach (Tsonis and Elsner 1991). Each time EOF associated to a determine spatial mode, is rebuilt using only a certain number of eigenvalues obtained from a SSA decomposition. Then, the genetic algorithm is employed to find the equation that best fits the data in the training set (March 1993–December 1998). The solution equation is then validated in the validation set (January 1999–December 2000), previously unknown by the algorithm. If the forecast performance is high (higher than 65 per cent of agreement between observations and predicted values), the rebuilt signal is considered to be mainly deterministic. A new time series is rebuilt from the original considering a larger number of eigenvalues and the previous process is repeated. The procedure is stopped when the inclusion of new eigenvalues from the SSA decomposition deteriorates the forecasting skill. In this case, the new included eigenvalues introduce unpredictable components of the signal, complicating the searching process of the genetic algorithm that attempts to find a mathematical relation also for the unpredictable part of the signal. This is often translated into decreasing performance of the algorithm. A bigger number of noisy samples would be required to achieve the same performance than in the noiseless case. The extra amount of data depends on the variance of the added noise (Magdon-Ismael *et al.* 1998). Following the above described method, only the first three EOFs were dynamically significant. A total of six SSA modes were required to rebuild each EOF satisfying the predictability criterion. Unlike the case studied by Alvarez *et al.* (2000a), the present approach subtracts to the data a varying spatial mean with strong seasonal component in order to identify the variance associated with front, eddy and jet structures. Thus, predictions of the mean spatial temperature of the region are also required as part of the SST forecast. The predictable part of the time series of the spatial mean component was obtained similarly to the EOF cases. The genetic algorithm has been trained using the first 70 points, the remaining 24 points are used to compute the forecast skill. The genetic algorithm has been configured in such a way that $m=6$ for the first and second EOF and $m=8$ for the third one. The maximum number of characters allowed for each tentative equation is 30. It was found that the performance of the genetic algorithm is degraded if the number of symbols is bigger than 30. This feature is induced by the high dimensionality of the searching space that difficulties the search of the algorithm. Each generation consists of a population of $N_{\text{pop}}=120$ randomly generated equations. A total of 10 000 generations involving $1.2 \cdot 10^6$ equations, have been considered for each case.

Figure 4(a, b, c and d) shows the eigenvectors describing the time histories of the first three spatial modes and the evolution of the spatial mean, as well as the

46

Figure 4. EOFs associated to the (a) first, (b) second and (c) third spatial modes and (d) time evolution of the spatial mean. Solid line represents the raw data while the dashed-dotted line indicates the predictable part of the variability found by the genetic algorithm. In the validation set corresponding to 1999–2000, in the figure months 71 to 94, one of every two points of the raw (predicted) data are marked with crosses (circles).

deterministic part identified by the above described procedure. It was found that 67 per cent, 92 per cent and 73 per cent of the total time variability (seasonal plus nonseasonal) were predictable for the first, second and third EOFs, respectively. For the spatial mean component, an accuracy of 99.07 per cent was obtained. Forecasting equations obtained from the genetic algorithm are:

$$A_1(t) = A_{s1}(t) + 2\delta_1(t - \Delta t) - 1.23\delta_1(t - 2\Delta t) - \delta_1(t - 6\Delta t) \left(0.23 - \frac{0.62\delta_1(t - 6\Delta t)}{10^{-4}} \right) + 1.34 \quad (6)$$

$$A_2(t) = A_{s2}(t) + \delta_2(t - 4\Delta t) - (1.75\delta_2(t - \Delta t) + \delta_2(t - 2\Delta t)) + \delta_2(t - 6\Delta t)\delta_2(t - 5\Delta t)(22\delta_2(t - 6\Delta t) + 2.25) + \frac{\delta_2(t - 3\Delta t)}{2\delta_2(t - 5\Delta t)} + 0.39 \quad (7)$$

$$A_3(t) = A_{s3}(t) + \delta_3(t - \Delta t) - \delta_3(t - 2\Delta t) \\ + 0.28(\delta_3(t - 3\Delta t) + 9.06 \delta_3(t - \Delta t)(\delta_3(t - \Delta t) - \delta_3(t - 7\Delta t))) \quad (8)$$

$$A_M(t) = A_{sM}(t) + \delta_M(t - \Delta t) + \delta_M(t - 3\Delta t) - \delta_M(t - 2\Delta t) \\ + \frac{2.81(\delta_M(t - 4\Delta t) - \delta_M(t - 2\Delta t) + \delta_M(t - \Delta t)) - 2.09 \delta_M(t - 3\Delta t)}{-3.64 + \delta_M(t - 4\Delta t)\delta_M(t - 2\Delta t)(\delta_M(t - 3\Delta t) - 2.13)} \quad (9)$$

for the first, second and third EOFs and the spatial mean, respectively. In equations (6)–(9), $\Delta t = 1$ month, $A_{s1(s2,s3,sM)}$ is the seasonal component of the first (second, third, spatial mean) EOF, computed as the average of the monthly SST during the 7 years and $\delta_{1(2,3,M)}$ is the fluctuation amplitude around the seasonal value. Thus, the prediction is carried out on the fluctuating field.

In order to compare the performance of the employed genetic algorithm a linear autoregressive model and a standard backpropagation neural network were applied to the same prediction problem. The orders of the autoregressive models, found using the Akaike information criterion (Konyayev 1990), were six for the first EOF, second EOF and the spatial mean, and eight for the third EOF. The neural network was constituted by three layers with six neurons in the hidden layer, and one neuron in the output layer. The learning rate of the backpropagation algorithm was fixed at $\eta = 0.1$. The forecast skill of each predictor is measured by the explained variance in the validation set:

$$R_\delta^2 = 1 - \frac{\sum_{t=1}^N (\delta(t) - \delta'(t))^2}{\sum_{t=1}^N (\delta(t) - \bar{\delta})^2} \quad (10)$$

where $N = 24$ is the number of points in the validation set (1999–2000), $\delta(t)$ is the observed value of the nonseasonal variability, $\delta'(t)$ is the predicted value and $\bar{\delta}$ is the mean value of the variable $\delta(t)$ in the validation set. Values of R_δ^2 close to one indicate good performance of the prediction system, while values of R_δ^2 close to or less than zero describe systems with poor prediction performance. $R_\delta^2 = 1$ if predictions are exact. Table 1 shows the explained variance of the non-seasonal part of the considered EOFs and the spatial mean, predicted by the different approaches. For the first EOF, comparison between explained variances indicates that the dynamical evolution of the non-seasonal variability is mostly linear. Only the neural network is able to capture part of the nonlinear behaviour, predicting 96 per cent of the non-seasonal time variability. The genetic approach can only describe the linear part of the variability. Unlike the first EOF, the genetic algorithm performs better than the linear model and the neural network for the second and third EOFs. Specifically,

Table 1. Explained variances in the validation set (1999–2000) obtained with different prediction approaches: the genetic algorithm, an autoregressive model and a backpropagation neural network.

Method	$R_{\delta_1}^2$	$R_{\delta_2}^2$	$R_{\delta_3}^2$	$R_{\delta_m}^2$
Genetic algorithm	0.89	0.96	0.4	0.82
Autoregressive	0.89	0.79	−0.82	−0.41
Backpropagation neural network	0.96	0.81	0.07	0.77

the non-seasonal variability contained in the third EOF seems to be characterized by a highly nonlinear behaviour. This fact is supported by the poor performance found with the linear model. The genetic algorithm also shows the best performance in predicting the fluctuations around the seasonal value of the spatial mean. It was found that the error in the predicted SST field is very sensitive to the accuracy of the prediction of the spatial mean temperature in the region. This is because an error in the prediction of the mean spatial value will constitute a constant offset in all sea points. Thus, the genetic algorithm provides the best overall performance. Notice that an optimum algorithm could be built taking the best predictor for each EOF. While this approach would slightly improve the performance of the system, it would also complicate its operational application and versatility. Because versatility and exportability are important scopes of the present work, the genetic algorithm (equations (6)–(9)) will be considered as the only predictor.

Climatological SST fields are often considered as first estimates of the unknown future values of the SST in a determined region. This approximation is commonly employed in numerical modelling of the ocean. Thus, comparison between predictions obtained with the approach proposed in this work and climatological SST fields, would provide an estimation of the information gain, if any, of the developed forecast method with respect to the standard procedure. Figure 5(a) compares the spatial average error in the SST obtained in the one-month ahead forecast carried out by the developed system during the years 1999–2000, with the error obtained if monthly climatological SST fields were employed as forecast fields. Results clearly indicate the improvement achieved by the satellite based forecasting procedure. The performance of the proposed system is significantly better during the periods from January to April and from August to October. These time intervals correspond to the existence of well developed oceanographic structures and forcing in the area. The SST field can be highly variable from year to year, as demonstrated by the deviation from the climatological value, but controlled by well defined dynamical processes. Part of the dynamical variability is captured by the proposed procedure and in consequence better forecasting skill is achieved. Figure 5(b) compares the spatial mean errors from the forecasting system and climatology when averaged for each month during the 7-year period. Results indicate that the satellite based ocean forecasting system is expected on average to provide better estimations of the SST than climatology for all months. Notice that for some periods, up to 0.6°C of reduction in the spatial average error is obtained with the proposed prediction method. The best improvement with respect to climatology is obtained in August 2000. Figures 6(a), (b) and (c) show the recorded monthly mean SST field corresponding to August 2000, the one-month ahead forecast field obtained from the satellite based forecasting system and the climatological SST of the region. The forecast field is in better agreement with the observed field than climatology. Specifically, the forecast field successfully reproduces the observed range of temperatures; around 26°C in the continental

Figure 5. (a) Comparison of the spatial average error in the SST obtained in the one-month ahead forecast carried out by the developed system during the years 1999–2000 (line with circles), with the error obtained if monthly climatological SST fields were employed as forecast fields (line with crosses). (b) Spatial mean errors from the forecasting system and climatology when averaged for each month during the considered 7-year period (same notation as (a)).

(a)

(b)

Figure 6. (a) Monthly mean surface temperature recorded during August 2000. (b) One month ahead forecast for the same period and (c) climatological sea surface temperature field.

platform decreasing until the 23–24° C in the core of the eddy structure off the Strait of Bonifacio. The temperature range of the climatological estimation is $\approx 2^\circ$ C colder.

5. Discussion

Fisheries, marine commerce, environmental control and military operations are among the activities that would greatly profit from the knowledge of future oceanographic conditions. Unfortunately, the strong spacetime variability and spatial extent of the ocean make ocean prediction a difficult task. Numerical modelling is the most common approach to forecast the ocean. If ocean volumetric data are available in real-time, ocean numerical models show good performance for short prediction times. Ocean measurements are sparse in time and space and so, continuous data support to numerical models is rarely achieved. On the other hand, the performance of numerical models is quickly degraded if only surface data is provided. Besides the numerical modelling approach, the ocean can be empirically modelled by expressing time evolving measurements in a suitable functional form. This standard practise in one-dimensional systems, has been recently generalized to chaotic two-dimensional systems (Lopez *et al.* 2000). The present work has shown an approach to obtain empirical predictive models of the ocean surface evolution from satellite data. The forecasting system has been applied and validated in an ocean area of remarkable oceanographic and environmental relevance. Results indicate that, at least part of the space-time ocean variability could be accurately forecast using satellite data. Specifically, forecasts of the monthly mean SST conditions more accurate than those inferred by climatology, have been obtained. The forecast fields already constitute valuable information to improve the atmosphere–ocean heat exchange fluxes in the ocean models, or to determine the evolution of frontal regions where an important part of marine mammal populations are concentrated.

Satellite based ocean forecasting does not require big computational resources and input-updating data are easily accessible, providing an operational profile to the prediction system. On the other hand, not all ocean variables and space-time scales can be predicted from a satellite based forecasting system. Predicted ocean variables are limited to fields measured by satellite and those fields inferred from

them. It is expected that the system performance degrades as the spatial and time scales are reduced. This is due to the increase of the stochastic component in the small and fast turbulent scales. At these scales, stochastic modelling could be more appropriate. Present research is focused on determining the limit spatiotemporal scales of the described forecasting system.

Acknowledgments

The work has been supported by the North Atlantic Treaty Organization and the EVK3-CT-2000-00028 European Project.

References

- ALVAREZ, A., 2001, Comparison of temporal versus spatial variance EOF decomposition in the performance of satellite based ocean forecasting systems: a study case. *J. Atmos. Ocean. Tech.*, submitted.
- ALVAREZ, A., LOPEZ, C., RIERA, M., HERNANDEZ-GARCIA, E., and TINTORE, J., 2000a, Forecasting the SST space-time variability of the Alboran Sea with genetic algorithms. *Geophysical Research Letters*, **27**, 2709–2712.
- ALVAREZ, A., ORFILA, A., and SELLSCHOPP, J., 2000b, A satellite based ocean forecasting system to support naval operations in crisis situations. NATO SACLANTCEN Report SM-375, Italy.
- ALVAREZ, A., ORFILA, A., and TINTORE, J., 2001, Darwin—an evolutionary program for nonlinear modelling of chaotic time series. *Computer Physics Communications*, **136**, 334–349.
- ASTRALDI, M., and GASPARINI, G. P., 1994, The seasonal characteristics of the circulation in the Tyrrhenian Sea. In *Seasonal and Interannual Variability of the Western Mediterranean Sea*, edited by P. E. La Violette (Washington: AGU), pp. 115–134.
- ASTRALDI, M., BACCIOLA, D., BORGHINI, M., DELL'AMICO, F., GALLI, C., GASPARINI, G. P., LAZZONI, E., NERI, P. L., and RASO, G., 1993, Caratteristiche stagionali delle masse d'acqua nell'Arcipelago Toscano. In *Arcipelago Toscano: Studio oceanografico, sedimentologico, geochimico e biologico*, edited by S. Furia (La Spezia: ENEA), pp. 3–7.
- ASTRALDI, M., BIANCHI, C. N., GASPARINI, G. P., and MORRI, C., 1995, Climatic fluctuations, current variability and marine species distribution: a case study in the Ligurian Sea (north-west Mediterranean). *Oceanologica Acta*, **18**, 139–149.
- ASTRALDI, M., GASPARINI, G. P., and MANZELLA, G. M. R., 1990, Temporal variability of currents in the Eastern Ligurian Sea. *Journal of Geophysical Research*, **95**, 1515–1522.
- ASTRALDI, M., GASPARINI, G. P., and SPARNOCCHIA, S., 1994, The seasonal and interannual variability in the Ligurian-Provençal basin. In *Seasonal and Interannual Variability of the Western Mediterranean Sea*, edited by P. E. La Violette (Washington: AGU), pp. 93–114.
- BAUER, J., 1994, *Genetic Algorithms and Investment Strategies* (New York: Wiley).
- BROOMHEAD, D. S., and LOWE, D., 1987, Extracting qualitative dynamics from experimental data. *Physica D*, **20**, 217–224.
- CARNIEL, S., KANTHA, L. H., FRANCHI, P., BERGAMASCO, A., and UMGIESSER, G., 2001, A real-time nowcast-forecast experiment in the Ligurian Sea during GOATS/MEANS 2000: Preliminary results. *Preprint*.
- CASDAGLI, M., DESJARDINS, D., EUBANK, S., FARMER, J. D., GIBSON, J., THEILER, J., and HUNTER, N., 1992, Nonlinear modeling of chaotic time series: theory and applications. In *Applied Chaos*, edited by J. H. Kim and J. Stringer (New York: John Wiley & Son), pp. 335–379.
- COLE, J., and VILLACASTIN, C., 2000, Sea surface temperature variability in the northern Benguela upwelling system, and implications for fisheries research. *International Journal of Remote Sensing*, **21**, 643–655.
- GARCIA-GORRIZ, E., and VAZQUEZ-CUERO, J., 1999, Ocean-atmosphere coupling in the Mediterranean Sea from TOPEX/POSEIDON, ERS1 and AVHRR data. *International Journal of Remote Sensing*, **20**, 2127–2147.

- GERSTOFT, P., 1995, Inversion of acoustic data using a combination of genetic algorithms and the Gauss-Newton approach. *Journal of the Acoustic Society of America*, **97**, 2181-2190.
- GOLDBERG, D. E., 1989, *Genetic Algorithms in Search, Optimization and Machine Learning* (Reading: Addison-Wesley).
- HOLLAND, J. H., 1975, *Adaptation in Natural and Artificial Systems* (Ann Arbor: University of Michigan).
- KONYAEV, K. V., 1990, *Spectral Analysis of Physical Oceanographic Data* (Rotterdam: A. A. Balkema).
- KRIVOSHEVA, V. G., 1983, Water circulation and structure in the Tyrrhenian Sea. *Oceanology*, **23**, 166-171.
- LAGERLOEF, G. S. E., and BERNSTEIN, R. L., 1988, Empirical orthogonal function analysis of advanced very high resolution radiometer surface temperature patterns in Santa Barbara channel. *Journal of Geophysical Research*, **93**, 6863-6873.
- LOPEZ, C., ALVAREZ, A., and HERNANDEZ-GARCIA, E., 2000, Forecasting confined spatio-temporal chaos with genetic algorithms. *Physical Review Letters*, **85**, 2300-2303.
- LUCASIS, C. B., BLOMMERS, M. J. J., BUYDEMS, L. M. C., and KATEMAN, G., 1991, A genetic algorithm for conformational analysis of DNA. In *Handbook of Genetic Algorithms*, edited by L. Davies (New York: Van Nostrand Reinhold), pp. 251-281.
- MAGDON-ISMAIL, M., NICHOLSON, A., and ABU-MOSTAFA, Y. S., 1998, Financial market: very noisy information processing. *Proceedings of the IEEE*, **86**, 2184-2195.
- MOSKALENKO, L. V., 1983, Seasonal variability of integrated wind-driven circulation in the Tyrrhenian Sea. *Oceanology*, **23**, 405-409.
- PENLAND, C., GILL, G., and WEICKMANN, K. M., 1991, Adaptive filtering and maximum entropy spectrum with application to changes in atmospheric angular momentum. *Journal of Geophysical Research*, **96**, 22 659-22 671.
- PREISENDORFER, R. W., 1988, *Principal Component Analysis in Meteorology and Oceanography* (Amsterdam: Elsevier).
- SCOTT, J. C., 1997, Satellite remote sensing in rapid environmental assessment: the optimum use of available data. In *Rapid Environmental Assessment*, edited by E. Pouliquen, A. D. Kirwan and R. T. Pearson (La Spezia: NATO-SACLANTCRN), pp. 65-72.
- TAKENS, F., 1981, Detecting strange attractors in fluid turbulence. In *Dynamical Systems and Turbulence* (Berlin: Springer).
- TSOIS, A. A., and ELSNER, J. B., 1992, Nonlinear prediction as a way of distinguishing chaos from random fractal sequences. *Nature*, **358**, 217-220.
- VAZQUEZ-CUERO, J., FONT, J., and MARTINEZ-BENJAMIN, J. J., 1996, Observations on the circulation in the Alboran Sea using ERS1 altimetry and sea-surface temperature data. *Journal of Physical Oceanography*, **26**, 1426-1439.