Estimating the predictability of an oceanic time series using linear and nonlinear methods

G.-C. Yuan,[1,2,3,4] M. S. Lozier,[3] L. J. Pratt,[2] C. K. R. T. Jones,[1,5] and K. R. Helfrich[2]

Received 29 September 2003; revised 7 April 2004; accepted 5 May 2004; published 3 August 2004.

[1] This study establishes a series of tests to examine the relative utility of nonlinear time series analysis for oceanic data. The performance of linear autoregressive models and nonlinear delay coordinate embedding methods are compared for three numerical and two observational data sets. The two observational data sets are (1) an hourly near-bottom pressure time series from the South Atlantic Bight and (2) an hourly current-meter time series from the Middle Atlantic Bight (MAB). The nonlinear methods give significantly better predictions than the linear methods when the underlying dynamics have low dimensionality. When the dimensionality is high, the utility of nonlinear methods is limited by the length and quality of the time series. On the application side we mainly focus on the MAB data set. We find that the slope velocities are much less predictable than shelf velocities. Predictability on the slope after several hours is no better than the statistical mean. On the other hand, significant predictability of shelf velocities can be obtained for up to at least 12 hours. *INDEX TERMS:* 3220 Mathematical Geophysics: Nonlinear dynamics; 4528 Oceanography: Physical: Fronts and jets; 9325 Information Related to Geographic Region: Atlantic Ocean; *KEYWORDS:* predictability, delay coordinate embedding, shelf break

Citation: Yuan, G.-C., M. S. Lozier, L. J. Pratt, C. K. R. T. Jones, and K. R. Helfrich (2004), Estimating the predictability of an oceanic time series using linear and nonlinear methods, *J. Geophys. Res.*, *109*, C08002, doi:10.1029/2003JC002148.

1. Introduction

[2] Data available for even well-studied regions of the ocean are very sparse both temporally and spatially. For a host of applications, including those associated with navigation, resource management, weather forecasting, and national defense, we need, however, to be able to base predictions of changes in ocean properties on these data sets. Our work here is motivated by the desire to understand how much we can predict, and for how long, from such sparse data sets.

[3] Coastal oceanographers are interested in predicting the strength and location of a shelf break front over the course of a few days. We address the dynamics of the shelf break front in the Middle Atlantic Bight (MAB). Many past studies have focused on the temporal and spatial variability of this front [e.g., *Flagg and Beardsley*, 1978; *Gawarkiewicz*, 1991; *Lozier and Gawarkiewicz*, 2001; *Lozier et al.*, 2002; *Fratantoni et al.*, 2000] in an effort to understand the underlying dynamics. The focus of our investigation is to measure the predictability of the shelf break front, an analysis distinct from a description of its variability. We ascertain that the shelf velocities are more predictable than the corresponding slope velocities.

[4] Standard methods of prediction are based on linear methods in the analyses of time series [e.g., *Penland and Sardeshmukh*, 1995; *Emery and Thomson*, 1998], and the basic technique here uses an autoregressive model. The real system is, however, nonlinear, being driven by an underlying fluid flow. We apply therefore a nonlinear prediction method based on phase space reconstruction and ask the question as to whether this offers any improvement over the standard linear prediction technique. As will be explained in section 5, the utility of such an approach is limited by the fact that it requires data length to be longer than the duration of most existing data records. This is borne out in the main case we consider of prediction around the shelf break front of the Middle Atlantic Bight.

[5] Nonlinear techniques derive from recent advances in

tion. As a simple example: Imagine that an observed variable is a sinusoidal function of time. Variability can be quantified as the amplitude of the sinusoidal function. If one uses this sinusoidal function to predict future values, however, the prediction errors would always be zero. Predictability is thus associated with skill at uncovering the rules governing the evolution of a time series. Ideally, an understanding of the local dynamics produces those governing rules. However, since we do not yet understand the dominant physical mechanisms that govern the shelf and slope waters of the Middle Atlantic Bight, past observations can provide an approximation to the rules, allowing for a measure of predictability. In order to isolate the limits and potential of past observations to predict future states, this study assesses predictability based solely on measurements. The measurements that are used for this assessment are from time series collected in the region, as will be detailed in section 3.4.

[7] As a second consideration, aimed at understanding the usefulness of nonlinear techniques in geophysical prediction, we seek to obtain an idea of the cutoff at which nonlinear techniques cease to be useful. Two main variables are relevant here: length of time series available and dimension of the underlying system. In an observational data set the latter is most likely not known and cannot even be surmised. While the former is obviously known, a characteristic length for the time series from which the dynamics can be usefully deduced may not be known. Furthermore, it is important that the sampling rate of the data is not an inherent harmonic of the dynamics; otherwise, the subsequent analysis will be strongly subject to aliasing errors. To understand where, and why, nonlinear methods might be useful in a geophysical setting, we introduce a sequence of data sets with increasing complexity. These derive from a standard low-dimensional model, namely, the Lorenz system, a shallow water model, and an observational data set from the South Atlantic Bight.

[8] We are not attempting here to use information from the underlying physics in any way. Data assimilation has emerged as a key technique for prediction that incorporates data into models to improve predictability. The predictive capability for data assimilation–based schemes is, however, constrained not only by the quality and quantity of data but also by incomplete model physics, unknown initial and boundary conditions, and parameterizations. We are restricting ourselves here to the extreme case where only data are taken as a basis of prediction with the aim of addressing the question: What can be predicted purely from data, and under what circumstances?

[9] An overview of our approach is given in section 2. This approach is applied to the data sets presented in this function is linear, we can use a linear autoregressive (AR) model for prediction [e.g., see *Brockwell and Davis*, 1991; *Emery and Thomson*, 1998]. On the other hand, if we have reason to believe that the function is nonlinear, it is difficult to select a global prediction model. However, a globally nonlinear model can be approximated by a model that has simple local structures. To construct such local models, we use the delay embedding technique that has recently been developed [e.g., see *Ott et al.*, 1994; *Abarbanel*, 1996; *Kantz and Schreiber*, 1997]. Essentially, a delay vector is constructed by intelligently combining the current and past values in the time series, such that the evolution of the time series can be represented as a function of delay vectors. As an example of the basic concepts involved in the delay embedding technique, consider an (unforced) double pendulum system, in which the angles between the shafts of the two pendulums and the vertical are denoted by $\theta_1(t)$ and $\theta_2(t)$. The state of the system at any time t is uniquely specified by the two angles and the two angular velocities $d\theta_1/dt$ and $d\theta_2/dt$. All future states are uniquely determined once the present state is specified. Now suppose that we have a time series of just one of the variables, say, $\theta_1(t)$, spanning $0 < t < T$. We can construct a delay vector $[\theta_1(T), \theta_1(T-\tau), \cdots, \theta_1(T-(d-1)\tau)]$, where the time delay τ is some multiple of the sampling interval in the time series and the embedding dimension d is an integer. It follows from the embedding theorem [*Takens*, 1981; *Sauer et al.*, 1991] that for sufficiently large d the delay vectors uniquely specify the system state under a wide variety of conditions. This information can then be used for the prediction of future values. Details of this technique, such as the selection of the embedding dimension and the time delay, will be discussed in section 4.

[11] One of the objectives of this paper is to gain insight into the utility and limitations of both linear and nonlinear methods in the construction of prediction models for observed time series. To achieve this goal, we address the following two questions: (1) Do observed time series contain structures that cannot be explained by linear models? (2) Can nonlinear models provide better predictions than linear models? To answer the first question, we test whether the observed time series can be differentiated from other time series that are generated with the same linear properties, a technique called surrogate data testing [*Theiler et al.*, 1992; *Smith*, 1992; *Schreiber and Schmitz*, 2000]. The second question can be answered by applying both linear and nonlinear prediction models to the observed time series and then comparing their prediction errors. The methods used for the surrogate data testing and the computation of the prediction errors will be explained in section 4.

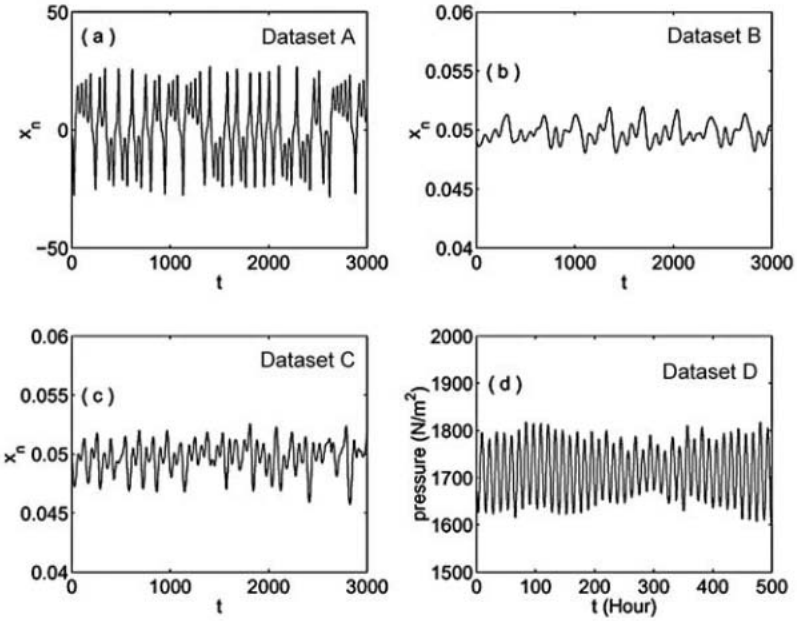


Figure 1. Representative time series from (a) data set A, (b) data set B, (c) data set C, and (d) data set D. In Figures 1a–1c, x_n is unitless.

primary goals is the application of these methods to time series from the Middle Atlantic Bight, considerable insight is gained from the application of these methods to less complicated systems and, for the Lorenz system and the shallow water model, systems that are known to be nonlinear.

[13] Whereas the numerical data are noise-free, the observational signals are inevitably corrupted by instrument and other errors. An important and practical question is how sensitive the prediction methods are to measurement noise. For the Lorenz system (which is the underlying model for data set A to be discussed in section 3.1) it has been shown that the prediction methods are still valid when the noise level is 45% (M. Dodson et al., Noise sensitivity of simple deterministic prediction model tested on Lorenz attractor data, available at http://gridlock.york.ac.uk/papers/noise.doc, 2004). For other systems the admissive noise level obviously can change depending on the details of the underlying dynamics. Statistical methods have been developed to incorporate the effects of observational noise and estimation error [*Casdagli et al.*, 1991; *Grassberger et al.*, 1993]. Typically, this requires some knowledge of some statistical properties of the noise. It is not a simple matter, however, to properly model the noise in the instrument: Is it

$$\frac{dy}{dt} = -xz + rx - y \tag{2}$$

$$\frac{dz}{dt} = xy - bz: \tag{3}$$

We numerically integrated these equations using the parameters $\sigma = 16$, $b = 4$, and $r = 45.92$ as was done by *Abarbanel* [1996] and a time step of D$t = 0.01$. The x coordinate was recorded after every time step. A time series of the x coordinate, containing 10,000 points, is referred to as data set A and is plotted in Figure 1a.

3.2. Shallow Water Model: Data Sets B and C

[15] In a study by *Helfrich and Pratt* [2003], flow in a bounded basin that is drained through a strait is modeled by the nondimensional, reduced gravity, shallow water momentum

$$\frac{@}{@} + \mathbf{u}\cdot\nabla\mathbf{u} + \mathbf{k}\times\mathbf{u} = -\nabla(h+b) + \mathbf{D} + \mathbf{M} \tag{4}$$

and continuity

10,000 data points. As r decreases, the time series changes from nearly steady to periodic and then to aperiodic. For sufficiently small r the time series appears irregular. In this paper, we consider two data sets, one obtained for $r = 0.003$ (data set B, Figure 1b) and one for $r = 0.001$ (data set C, Figure 1c). Although both time series are aperiodic, data set C appears more irregular than data set B. In section 5.1 we will show that data set B corresponds to a low-dimensional system, whereas data set C corresponds to a high-dimensional system.

3.3. South Atlantic Bight Data (Data Set D)

[16] In an effort to gain experience in the application of our methods to observational data, we chose to analyze a record of near-bottom pressure from the continental shelf off Georgia, collected as part of the South Atlantic Bight Synoptic Offshore Observational Network (SABSOON) [*Seim*, 2000]. We chose this time series for two main reasons. First, we believed the dynamics of this region would be less complicated than the dynamics associated with the shelf break front in the Middle Atlantic Bight. We anticipated that the dynamics of these shelf waters would be tidally dominated and thus primarily linear. In this sense, the South Atlantic Bight (SAB) time series was used as a learning tool for the application of our methods to the shelf break front. The second reason we chose this time series was its record length and the anticipation of even longer records as the SABSOON program continues its operations. In our study we used a time series of near-bottom pressure collected from May 2000 to October 2001. The raw data, recorded in 6-min intervals, were hourly averaged to create our data set D (Figure 1d). Since nonlinear models can be constructed from consecutive data points only, gaps in the training subseries were filled by using Fourier interpolation, where the Fourier coefficients were estimated by least squares fitting with a 2-day window leading each gap.

3.4. Middle Atlantic Bight Data (Data Set E)

[17] The main data set analyzed in this paper is a set of hourly averaged, high-quality current and temperature measurements collected during the Nantucket Shoals Flux Experiment (NSFE) from March 1979 to April 1980 [*Beardsley et al.*, 1985]. Six moorings were deployed across the continental shelf and upper slope, south of Nantucket, near 70°W. A schematic cross section of the array is shown in Figure 2. A thorough description of the current and temperature data, with a focus on the low-frequency variability, was given by *Beardsley et al.* [1985]. Briefly, they found the mean current to be along-shelf toward the west,

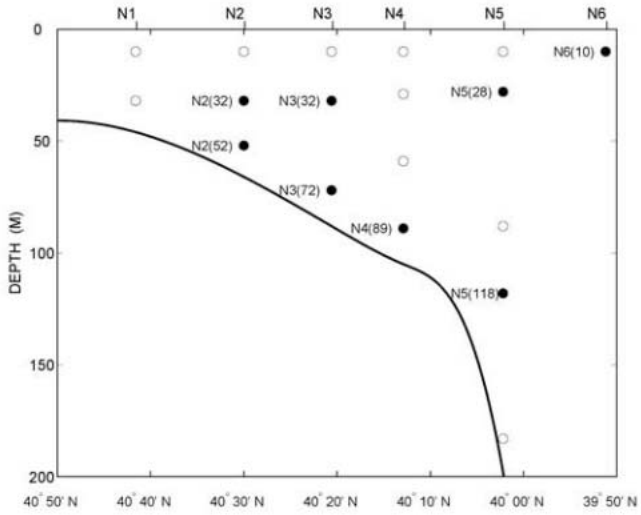


Figure 2. Schematic cross section of the NSFE79 moored array. The locations of the moored instruments are marked by open (short time series) or solid (long time series) circles. The water depth (in meters) appears in parentheses next to the mooring number for the long time series.

series is difficult because of the strong seasonality of the temperature record and the short length of the record relative to this dominant variability. Of the 18 velocity records available, 8 cover a period of more than 1 year. These eight relatively long records of both cross-shelf and along-shelf velocity (marked with solid circles in Figure 2) are investigated in this paper. The remaining records (at most about half of a year long) are of inadequate length for the application of our time series analysis. Four representative time series from the 16 selected for our study (eight cross-shelf and eight along-shelf) are shown in Figure 3. We should also note that two of the eight long records contain gaps of approximately half a month. These gaps were not filled because they are too wide. Special treatment of these gaps will be discussed in section 5. Although there are no other gaps in these selected MAB time series, we note that these time series were prepared from raw data that contained many short gaps [*Beardsley et al.*, 1985].

4. Methods

[19] In this section we give an overview of the fundamentals of linear and nonlinear prediction methods and comment on their utility and limitations. We also give an overview of surrogate data testing and detail how we

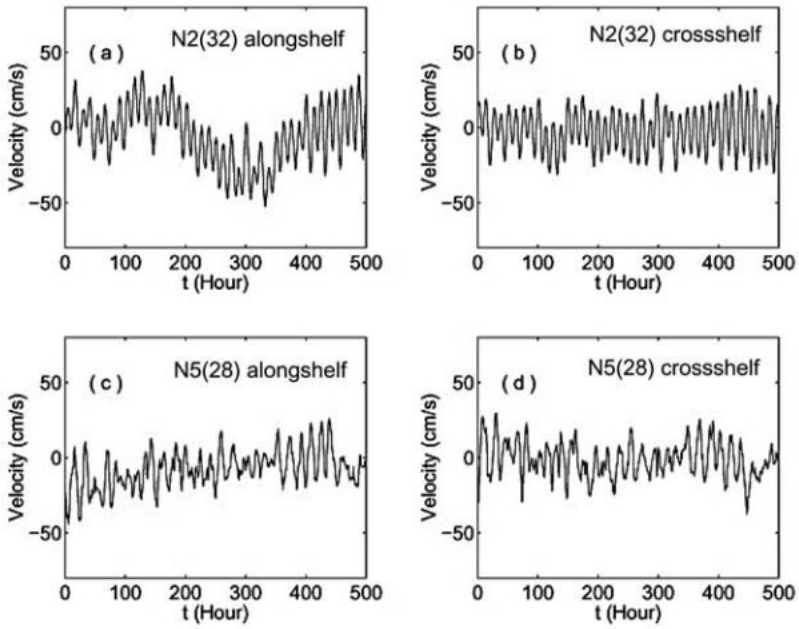


Figure 3. Representative MAB time series: (a) N2(32) along-shelf velocity, (b) N2(32) cross-shelf velocity, (c) N5(28) along-shelf velocity, and (d) N5(28) cross-shelf velocity.

where a_k, $k = 1, \ldots, p$ are constant coefficients and $\{w_n\}$ is a Gaussian white noise time series whose standard deviation is σ. The noise term $\{w_n\}$ comes from random input from the background. The parameters a_k are chosen such that the autocorrelation of the model matches with that for $\{x_n\}$. Such matching is assured by the solution of the associated Yule-Walker equations [*Brockwell and Davis*, 1991].

[21] A time series satisfying equation (6) is a Gaussian random process, and the parameters a_k are uniquely determined by the autocorrelation function. Another view is that these parameters are uniquely determined by the square of the Fourier amplitudes (i.e., the power spectrum), which can be converted from an autocorrelation by Fourier transform. A consequence is that the Fourier phases can be regarded as random. This important property is the basis for surrogate data testing to be discussed in section 4.2.

[22] Linear methods are useful because they are very well understood and are straightforward to implement. Additionally, unlike the nonlinear methods to be discussed in section 4.1, constructing linear models does not require large amounts of data. However, when the underlying dynamics are nonlinear, it is possible that the evolution of the time series deviates significantly from that predicted by

space reconstruction. For example, a possible realization of the ocean field can be considered as a state in phase space. If the dynamics are deterministic, then all future phase states are unambiguously determined by the present phase state. Furthermore, if the present phase state can be determined with small uncertainty, then at least for a short time the future phase states can be determined quite accurately. If a time series consists of only one variable $\{x_n\}$, then the phase state cannot be inversely determined. However, this task can be achieved by constructing delay vectors $\mathbf{X}_n = \{x_n, x_{n-\tau}, \cdots, x_{n-(d-1)\tau}\}$, which involve two parameters: the time lag τ and the embedding dimension d, provided d is sufficiently large, a property ensured by the embedding theorem [*Takens*, 1981; *Sauer et al.*, 1991]. To predict a future value of x_{n+s}, given the present delay vector $\mathbf{X}_n$, we search in the past record for its nearest neighbor, say, $\mathbf{X}_m$. The value of x_{n+s} can then be approximately predicted as x_{m+s}. Since observational data are generally contaminated with noise, the determination of the present state also involves uncertainty. Taking such uncertainty into account, prediction should be based not just on a single neighbor but rather on the average of an ensemble of close neighbors, each within a threshold level of uncertainty, resulting in a locally constant model.

involved in making such selections. If τ is too small, components in a delay vector are nearly identical, such that adding new components does not provide new information. On the other hand, if τ is too large, successive components are totally unrelated. Popular choices for τ include the first zero crossing of the autocorrelation function and the first minimum of mutual information function [*Fraser and Swinney*, 1986]. Typically, the effectiveness of a model is not highly sensitive to the choice of τ [*Kantz and Schreiber*, 1997].

[25] Delay vectors can be viewed as projections from the full phase space of realizable states to a d-dimensional space. The embedding dimension d reflects how many degrees of freedom are being modeled. It should be chosen sufficiently large so that different states are truly represented by distinct delay vectors. The existence of d can be proven mathematically for purely deterministic systems of finite degrees of freedom [*Takens*, 1981; *Sauer et al.*, 1991]. A too large choice of d, however, will introduce redundancy and thus will degrade the performance of the model. An appropriate choice of d can be identified by using the false nearest neighbor (FNN) technique [*Kennel et al.*, 1992]. To choose d, one first assumes a provisional value for d, then locates the nearest neighbor of each delay vector using the d-dimensional metric. Next, the same pairs of vectors are both extended by adding one more delay coordinate and are compared using the $d + 1$-dimensional metric. If they become far apart, this "nearest neighbor" is judged false. The ratio (or percentage) of false nearest neighbors that is computed (the FNN ratio) can be viewed as a measure of determinism. By increasing d, the FNN ratio will decrease. If a d can be found for which there are no more FNNs, then determinism is established. This d is used for constructing delay embedding models. However, such a d may not always exist in application. When the data are too short or too noisy, estimates of the FNN ratio may contain significant uncertainty. In fact, we will show that in some cases the FNN ratio may appear to increase with d. As a result, determinism may not be established. In such cases, we choose d to be the dimension for which the FNN ratio either appears to saturate or achieves a minimum.

[26] Nonlinear methods have a drawback when applied to high-dimensional systems since the length of data required for constructing nonlinear models increases rapidly with the dimension of the underlying dynamics. This requirement can be prohibitive for many observational time series, where data sufficiency is usually an issue. Unfortunately, it is difficult to determine how much data are needed for a given embedding dimension without strong simplifying assumptions. Different estimates have been given in the literature [e.g., see

surrogate data test contains the following steps. First, a null hypothesis (a potential explanation for the evolution of the time series) is specified. Second, a number of random realizations (called surrogate data) of the hypothesized process are generated numerically. Third, we select a discriminating statistic, denoted by η, and compute its value for the observed and surrogate time series. Finally, we determine whether the value for the observed data, η^o, is significantly different (in a statistical sense) from the surrogate data η_i^s, $i = 1, \cdots, m$. If it is, then the null hypothesis is rejected.

[28] The null hypothesis used in this paper is the following: The observed time series may not have a Gaussian distribution, but it can be derived from another time series originating from a linear Gaussian stochastic process by a nonlinear transform. The corresponding surrogate time series are generated by an iterative Fourier transform method [*Hegger et al.*, 1999; *Schreiber and Schmitz*, 2000]. Notice that a Gaussian stochastic process is uniquely determined by its Fourier amplitudes.

[29] After selecting η (discussed later in this section), we can compare η^o and η_i^s. If the difference is unlikely to be due to chance only, then the null hypothesis is rejected. For example, if η is the nonlinear prediction error and we want to show that η^o is less than η_i^s at a significance level of $\alpha = 0.05$, then at least $m = 1/\alpha = 20$ surrogate time series need to be generated. If η^o is less than η_i^s, for all $i = 1, \cdots, m$, then the null hypothesis is rejected because the probability that such an event occurs by chance is less than 0.05.

[30] While the above analysis can distinguish between the surrogate and observed time series, it does not provide information about the magnitude of the discrepancy between η^o and η_i^s. Such a quantification is necessary since the selection of a complex rather than a simple model is not worthwhile if the discrepancy is small. To quantify the relative magnitude of discrepancy, we select the following measure:

$$g = \left| \frac{1}{m} \sum_{i=1}^{m} \frac{\eta_i^s}{\eta^o} - 1 \right| \tag{7}$$

If g is much less than 1, the relative discrepancy can be considered negligible.

[31] We now discuss our choice of discriminating statistic η. Since our aim is to test whether nonlinear models contain extra predictable information compared with linear correlation, the discriminating statistic should reflect predictability. In particular, we consider the following two choices.

4.2.1. FNN Ratio (H_F)

[32] The FNN ratio discussed in section 4.1 reflects degrees of determinism and therefore serves as a good

where $\hat{x}_{n+s}$ is the predicted value for x_{n+s} and $\bar{x}$ is the mean of x_n estimated by $\bar{x} = (\Sigma x_n)/N$. In equation (8) the RMS magnitude of x_n is used as a normalizing factor; thus η_P is a dimensionless measure.

[34] Previous studies have shown that the statistic η_P is a powerful discriminating statistic for detecting nonlinearity [*Schreiber and Schmitz*, 2000]. However, η_F is also useful because it provides a guideline for the choice of the embedding dimension, which is important for the success of nonlinear prediction models.

[35] To avoid statistical bias, the delay embedding model used for computing η_F or η_P should not be optimized for the original data. Therefore a different choice of embedding parameters could potentially improve predictions. In section 4.3, when we discuss in detail how to quantify predictability, we will search for an optimal model among a number of candidates.

[36] What can we learn from a surrogate data test? We learn whether a time series contains predictable structures that cannot be modeled by linear methods, thereby gaining information about the utility of nonlinear methods. However, it may not determine whether the underlying system is nonlinear per se. In section 5 we present a time series that is nonlinear deterministic but is nevertheless indistinguishable from surrogate data. For this time series the use of nonlinear methods does not improve prediction skill because the available data are insufficient to uncover the deterministic dynamics. In fact, the results presented in section 5 suggest that surrogate data tests can detect low-dimensional nonlinear systems effectively, but not high-dimensional systems, unless, in theory, an extremely long time series is available.

[37] Application of these methods to time series with strong periodicity has drawbacks that have been previously noted [*Theiler et al.*, 1993]. In particular, positive surrogate data tests may result, not from nonlinearity, but from error induced by spectral estimation. In particular, this error may occur if the length of an interval used for spectral estimation is not a multiple of full cycles. In such a case, redundant frequencies are induced in the estimated spectrum, and, as a result, the surrogate time series appear noisier than the original data, even for linear systems. Such issues cause uncertainty for tests of the continental shelf data (data set D) and the MAB data set. However, we believe that such spectral errors compose only a small portion of prediction errors and thus do not seriously affect our results.

4.3. Prediction Errors

[38] To test the accuracy of a prediction method, we various orders and nonlinear models with various embedding parameters, as specified in section 5.

[39] As a final comment on our methods we note that we have purposely chosen not to employ a more advanced measure of predictability offered by dynamical systems theory, that is, the Lyapunov exponent [see, e.g., *Ott*, 1993], which measures the exponential growth rate of initial errors. Theoretically, this measure has the advantage in that it is a property of the underlying dynamical process. However, estimating Lyapunov exponents on the basis of time delay embedding is a very difficult task because of data limitations. Studies have shown that estimates are likely to be incorrect, even in relatively idealized situations [*Ott et al.*, 1994; *Kantz and Schreiber*, 1997].

5. Results

[40] In this section we apply surrogate data tests to data sets A–E. Additionally, we compute prediction errors for each of these data sets using linear and nonlinear methods, as discussed in section 4.

5.1. Testing Nonlinearity Using Surrogate Data

[41] First, consider data sets A through C, all of which are generated from nonlinear, deterministic systems. We test the null hypothesis that the time series can be derived from a linear Gaussian stochastic process via a possibly nonlinear transform. In Figures 4a–4c η_F is plotted as a function of d, with τ at a fixed value of 4. The original data are plotted with the thick curve, while the distribution of the surrogate data is shaded. For data sets A and B, η_F is initially large but drops to near zero as d increases to 3 or larger, reflecting the low dimensionality of data sets A and B. For the surrogates, η_F decreases also with increasing d, but the values are substantially higher than the original time series. Clearly, the null hypothesis can be rejected with a significance level of 0.05 for both data sets A and B. For data set C, η_F remains relatively high despite increasing d, suggesting that the underlying dynamics are highly dimensional. Additionally, η_F for the original data falls within the cloud of the surrogate data sets; thus the null hypothesis cannot be rejected for data set C. The increase in η_F past $d = 4$ for data set C is an artifact caused by data insufficiency.

[42] The results of the surrogate data tests using η_P as a discriminating statistic are shown in Figures 4d–4f, where η_P is computed by using a locally constant model. The parameters d and τ were each fixed at 4. The averaged prediction errors for data sets A, B, and C are plotted as a function of prediction time s in Figures 4d, 4e, and 4f,

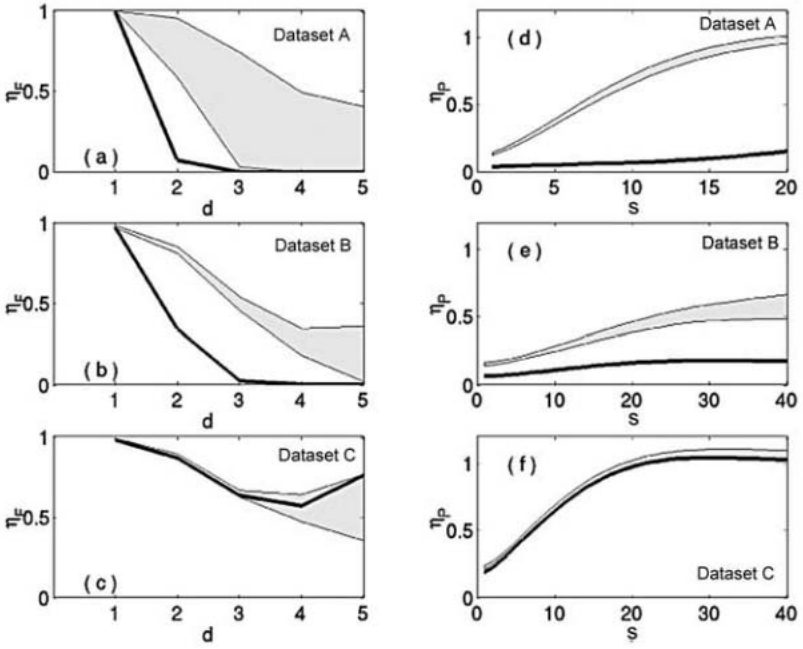


Figure 4. Surrogate data tests for data sets A–C: (a–c) η_F and (d–f) η_P. Values for the original data are shown as the thick solid curves, while the shaded region indicates the range of values for the surrogate data.

much less than 1, suggesting that the difference between the original and the surrogates is not substantial. As with the analysis of η_F and η_P, these results suggest that surrogate data tests can detect low-dimensional nonlinear systems effectively but not high-dimensional systems.

[44] The results of the surrogate data tests for the SAB shelf data are shown in Figure 5. The null hypothesis can be rejected using η_P as the discriminating statistic but not when using η_F. However, the evaluation of g_P (Table 1) shows that the difference between the original data and the surrogate data is quite small. Thus, though the cloud of surrogates in Figure 5b is distinct from the original data, the difference is small enough such that the use of a nonlinear model would not be advisable.

[45] The above results for the SAB data set (D) could suggest that the underlying dynamics are nonlinear and of high dimensionality, similar to data set C. However, unlike data set C, η_F approaches zero when the embedding dimension is 4 or 5 (Figure 5a), suggesting a system of relatively low dimension. Additionally, the prediction errors for data set D are quite small relative to data set C, cases the original and surrogate data cannot be distinguished. This assessment is confirmed by the small values for g_F and g_P (Table 1).

[47] As seen in Figures 6a and 6b, η_F is significantly different from zero in all four cases, suggesting that the embedding dimension is not sufficient to uncover any determinism. Increasing the embedding dimension does not reduce η_F significantly. In Figure 6b, η_F at $d = 5$ is actually greater than at $d = 4$. As discussed in section 5.1, this artifact is due to data insufficiency. When the prediction errors for the surrogate time series saturate (Figures 6c and 6d), the prediction errors for the observed data are much larger than those for data sets A and B, also an indication that the system is not low dimensional. We add a cautionary note here that these prediction errors calculated in this section are not entirely indicative of the predictability of the system for two reasons. First, these errors represent only in sample errors, when out-of-sample errors should be used to quantify predictability. Second, no effort has been made to optimize the prediction models. With a different choice of

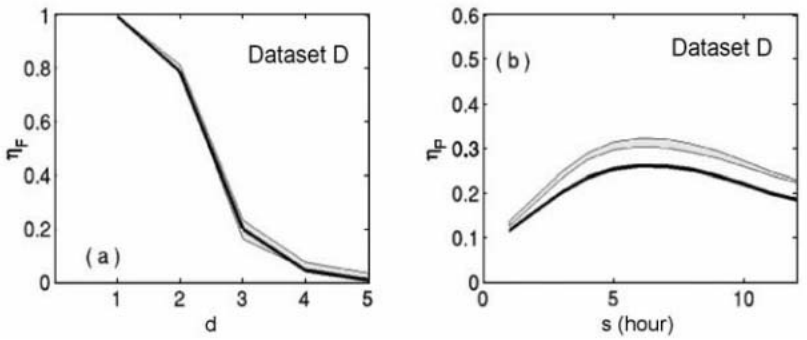


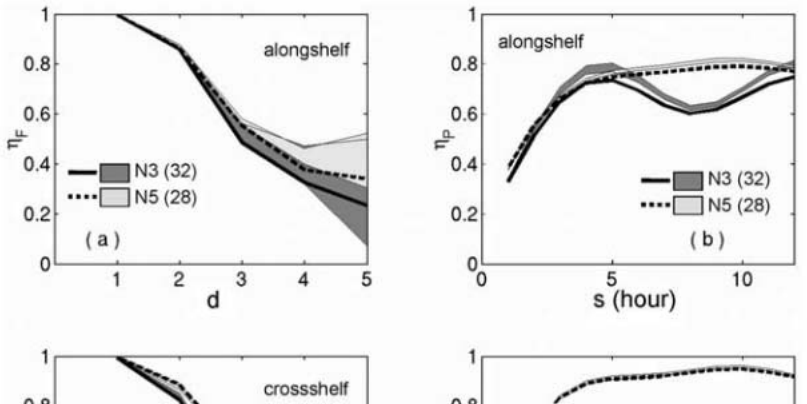
Figure 5. Surrogate data tests for data set D: (a) η_F and (b) η_P. Values for the original data are shown as thick solid curves, while the shaded region indicates the range of the surrogate data.

embedding parameters, prediction errors are likely to be reduced, as will be seen in section 5.2.

5.2. Predictability

[48] To quantify predictability, we divide a time series into a training subseries and a testing subseries, as explained in section 4.3. Linear and nonlinear models are constructed from the training subseries and are applied to the testing subseries. To test whether the prediction of nonlinear systems can be improved by using nonlinear models, we first consider data sets B and C. The order of AR models is increased from 1 until a further increase does not significantly improve prediction. This occurs when the order is about 6 in both cases. The minimum prediction error obtained from these 6 models is referred to as the linear prediction error. We vary the embedding parameters to optimize the nonlinear model. Our earlier computation of η_F (see Figure 4) suggests that the embedding dimension should not exceed 5. Therefore we apply locally linear models using embedding dimensions from 3 up to 5, with the time delay varying from 1 through 6, for a total of 18 models. The same set of embedding parameters was also used to construct locally constant models, but we find that locally linear models are more accurate than locally constant models. The minimum prediction error obtained from these models is referred to as the nonlinear prediction error.

[49] Linear and nonlinear prediction errors as a function of time for data sets B and C are shown in Figure 7. Nonlinear prediction errors are much smaller than linear prediction errors for data set B, suggesting that nonlinear prediction models are superior. On the other hand, nonlinear prediction errors are actually larger than linear prediction errors for data set C. Since an AR model can be viewed as a special case of locally linear models, the better performance of linear prediction errors for data set C indicates that none of the selected nonlinear models is



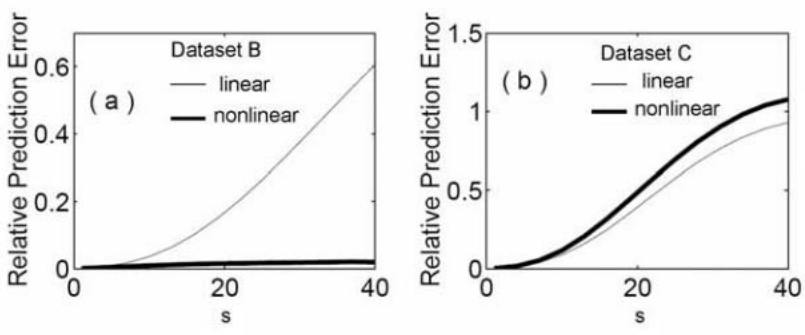
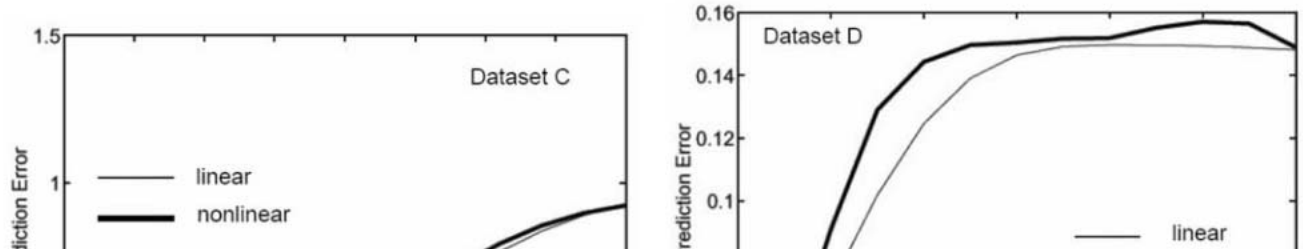


Figure 7. Relative prediction errors as a function of time for (a) data set B and (b) data set C. Linear errors are shown by the thin curves, and nonlinear errors are shown by the thick curves.

optimal. We note that the flexibility of nonlinear models makes it difficult to find the optimal one. Comparison of the two cases suggests that nonlinear models can improve prediction significantly if the system is low dimensional but not if the system is high dimensional. We believe this difference arises because the data are not sufficient for constructing high-dimensional models.

[50] How much data are needed for nonlinear methods to be effective? Although in principle, data set C can be extended indefinitely, computational resources limited the regeneration of data set C to a tenfold increase, or 100,000 data points. For this longer time series the FNN ratio ceases to decay at around dimension 7. (As discussed in section 5.1, it is an artifact that the FNN ratio does not drop further for higher embedding dimension.) Thus we use embedding dimension from 3 up to 7. The testing subseries is chosen to be the same as for the shorter time series. Compared with the shorter time series, the magnitude of nonlinear prediction errors is smaller (see Figure 8), while linear prediction errors maintain the same magnitude. Nonetheless, nonlinear predictions are still no more effective than linear predictions. Furthermore, we find that the longer time series is also indistinguishable from the surrogate data. These results further suggest that high-dimensional systems require very long duration data for the effectiveness of nonlinear methods.

[51] For data sets D (the SAB data) and E (the MAB data) we use the same method to select the maximum order of AR models, which turns out to be about 12. Thus in total we consider 12 different models. As before, our nonlinear model is a locally linear model using embedding dimensions from 3 up to 5 and time delays from 1 hour up to 12 hours, for a total of 36 models. For the predictability of data set D, a year-long training subseries is used, and the length of the test subseries is chosen to be 1 month. The resulting prediction errors are plotted against prediction time in Figure 9. Both the nonlinear and the linear errors are quite small; however, linear models seem to outperform nonlinear models, which, as discussed earlier, is due to the suboptimal behavior of the selected nonlinear models. The fact that nonlinear models fail to outperform linear models could be due to the high dimensionality of the SAB data set, as suggested for data set C. However, for data set C both linear and nonlinear prediction errors are quite large, while for data set D they are both small. Therefore a more feasible explanation for why nonlinear models do not outperform linear models for data set D (the SAB data) may be that the underlying dynamics are close to linear, as postulated in



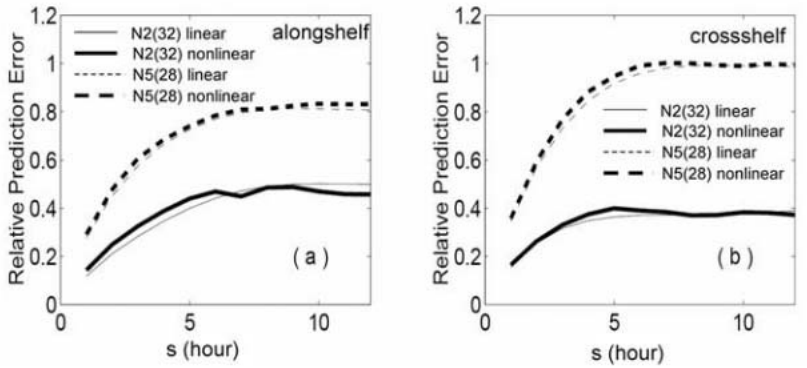


Figure 10. Relative prediction errors as a function of time for the N2(32) (solid curves) and N5(28) (dashed curves) time series: (a) along-shelf velocity and (b) cross-shelf velocity. Linear errors are shown by the thin curves, and nonlinear errors are shown by the thick curves.

section 5.1. We note here that our results show that the pressure field in this region of the SAB is remarkably predictable out to 12 hours, with errors on the order of 10% of the RMS, a result consistent with our expectation of a tidally dominated regime. Data set E (the MAB data) is shorter than data set D, so a shorter training subseries is used. We use an 11-month training subseries for the six time series containing year-long consecutive records. For the two time series with wide gaps (at N2) we use the uninterrupted segments before the gaps as the training subseries. The length of the test subseries is again chosen to be 1 month.

[52] Again, we use N2(32) and N5(28) to illustrate the typical time dependence of prediction errors for data set E (Figure 10). In all cases, prediction errors saturate at about 5 hours. Predictions are moderately good for N2(32) but rather poor after 5 hours for N5(28). Interestingly, cross-shelf velocity is more predictable than along-shelf velocity at N2(32), but the opposite is found at N5(28). This difference might be due to the presence of the strong along-slope frontal current at N5(28), which would facilitate along-slope predictions. In general, this analysis shows that nonlinear predictions fail to outperform linear predictions, rendering their use unnecessary for this data, as suggested in section 5.1 by the surrogate data tests.

[53] To examine the spatial structure of predictability, we show prediction errors at 1 hour, 3 hours, and 12 hours for all eight time series in Figure 11. In all cases, linear and nonlinear prediction errors have approximately the same magnitudes, again suggesting that there is no advantage in using nonlinear prediction models. Figures 11a and 11b

where the dynamics are dominated by a highly variable and unstable frontal current; therefore it is not surprising that prediction degenerates quickly for these time series. On the shelf the relative prediction errors are at most 0.5, suggesting there is significant predictability during this period, doubtless due to the dominance of the tides on the shelf, as was also noted for the SAB data set. On the other hand, on the slope the 12-hour prediction errors can be as large as the RMS magnitude of the velocity itself, which is no better than simply predicting the statistical mean. Hence predictability is lost almost completely. Prediction errors for cross-shelf and along-shelf components are roughly the same order of magnitude, yet the difference between the two has significant spatial structures. In particular, the cross-shelf velocity seems to be more predictable than the along-shelf components on the shelf, but the opposite is true on the slope. The greater predictability of the along-shelf component over the slope may be due to the downstream advective influence provided by the shelf break frontal current.

6. Summary

[54] This analysis was designed with a twofold purpose: first, to establish procedures that would evaluate the effectiveness of nonlinear methods in the analysis of oceanic time series and second, to determine the predictability of the flow field in the shelf break region of the Middle Atlantic Bight. Subsumed by the first and second goals was another, namely, to determine whether nonlinear methods could improve on prediction estimates for the shelf break velocity

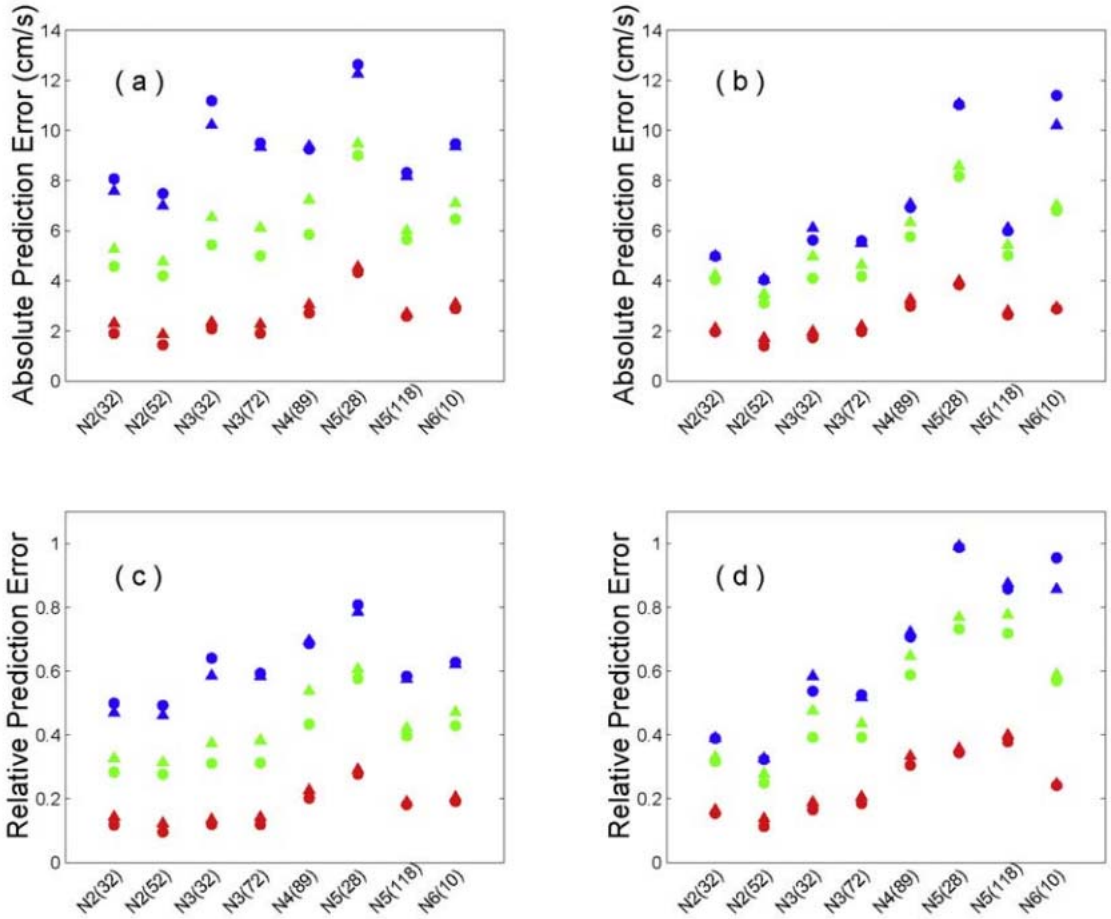


Figure 11. Spatial distribution of prediction errors for the MAB data: (a) absolute prediction errors for along-shelf velocity, (b) absolute prediction errors for cross-shelf velocity, (c) relative prediction errors for along-shelf velocity, and (d) relative prediction errors for cross-shelf velocity. Predictions are made for 1 hour (red), 3 hours (green), and 12 hours (blue). Triangles represent nonlinear errors, and circles represent linear errors.

predictability generally decreasing from the shelf to the slope waters. This decrease is attributed to the presence of a highly variable frontal current that is generally situated at the shelf break. For the shelf waters the predictability of the velocity field is reasonably good. Prediction errors after 12 hours are generally no larger than one half of the standard deviation of the observed velocity field. Prediction errors at the slope, however, are significantly larger. After 12 hours these errors approach the magnitude of the standard deviation of the observed field, calling into question the utility of low dimension characterizes the system. However, for high-dimensional systems the utility of nonlinear models is severely limited by the sufficiency and quality of data.

[57] This was demonstrated by our use of data set C, again generated from a nonlinear shallow water model but with parameters that yielded a more aperiodic, irregular time series. In fact, increasing the data length tenfold is still not enough for the nonlinear methods to be more effective. Our judgment as to whether nonlinear methods were useful for the analysis of these data sets was based on surrogate data

dynamics. This distinction between the data sets is primarily based on the analysis of the false nearest neighbors, which showed that the SABSOON data had a relatively low dimension, in contrast to the NSFE data for which a clear embedding dimension could not be established. Thus, while we believe the NSFE data is inherently nonlinear, the available data set was not sufficient for constructing the necessary high-dimensional model that could, possibly, have outperformed a linear model. To counter the difficulties imposed by data limitations, efforts to assess the underlying dynamics in this region should be encouraged. The application of model dynamics could place important constraints on the predicted fields.

[59] Finally, in order to pursue the question as to whether nonlinear methods can be useful in the analysis of ocean time series, future work could establish whether, indeed, the methods are limited by quality (i.e., noise) and record length. Given the obvious limits of observed time series, such work would more than likely need to focus on model-generated data. Additionally, to improve the predictability of the dynamics in the shelf break frontal region, future work could incorporate spatial information in the time series analysis. We note also that for the prediction of a shelf break front, marked by sharp thermal and haline contrasts, temperature and salinity fields may be more informative than velocity fields.

[60] **Acknowledgments.** We thank Harvey Seim for providing the shelf data (data set D) from the South Atlantic Bight and Kipp Shearman, Glen Gawarkiewicz, Lenny Smith, Steve Lentz, and Cecile Penland for helpful discussions. We are also grateful to the two anonymous referees whose comments improved the paper. This research was supported by Office of Naval Research grants N00014-01-1-0260, N00014-92-J-1481, and N10014-99-1-0258.

References

Abarbanel, H. D. I. (1996), *Analysis of Observed Chaotic Data*, 272 pp., Springer-Verlag, New York.

Beardsley, R. C., D. C. Chapman, K. H. Brink, S. R. Ramp, and R. Schlitz (1985), The Nantucket Shoals Flux Experiments (NSFE79). Part I: A basic description of the current and temperature variability, *J. Phys. Oceanogr.*, *15*, 713–748.

Brockwell, P. J., and R. A. Davis (1991), *Time Series: Theory and Methods*, 2nd ed., 577 pp., Springer-Verlag, New York.

Casdagli, M., S. Eubank, J. D. Farmer, and J. Gibson (1991), State space reconstruction in the presence of noise, *Physica D*, *51*, 52–98.

Eckmann, J.-P., and D. Ruelle (1992), Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems, *Physica D*, *56*, 185–187.

Emery, W. J., and R. E. Thomson (1998), Data analysis methods in physical oceanography, 634 pp., Pergamon, New York.

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

Flagg, C. N., and R. C. Beardsley (1978), On the stability of the shelf water/slope water front south of New England, *J. Geophys. Res.*, *83*, 4623–

Frison, T. W., H. D. I. Abarbanel, M. D. Earle, J. R. Schultz, and W. D. Scherer (1999), Chaos and predictability in ocean water levels, *J. Geophys. Res.*, *104*, 7935–7951.

Gawarkiewicz, G. (1991), Linear stability models of shelfbreak fronts, *J. Phys. Oceanogr.*, *21*, 471–488.

Grassberger, P., R. Hegger, H. Kantz, C. Schaffrath, and I. Schreiber (1993), On noise reduction methods for chaotic data, *Chaos*, *3*, 127–141.

Hegger, R., H. Kantz, and T. Schreiber (1999), Practical implementation of nonlinear time series methods: The TISEAN package, *Chaos*, *9*, 413–435.

Helfrich, K. R., and L. J. Pratt (2003), Rotating hydraulics and upstream basin circulation, *J. Phys. Oceanogr.*, *33*, 1651–1663.

Kantz, H., and T. Schreiber (1997), *Nonlinear Time Series Analysis*, 304 pp., Cambridge Univ. Press, New York.

Kennel, M. B., R. Brown, and H. D I. Abarbanel (1992), Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Phys. Rev. A*, *45*, 3403–3411.

Lozier, M. S., and G. Gawarkiewicz (2001), Cross-frontal exchange in the Middle Atlantic Bight as evidenced by surface drifters, *J. Phys. Oceanogr.*, *31*, 2498–2510.

Lozier, M. S., M. S. C. Reed, and G. G. Gawarkiewicz (2002), Instability of a shelfbreak front, *J. Phys. Oceanogr.*, *32*, 924–944.

Ott, E. (1993), *Chaos in Dynamical Systems*, 385 pp., Cambridge Univ. Press, New York.

Ott, E., T. Sauer, and J. A. Yorke (Eds.) (1994), *Coping With Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems*, 418 pp., John Wiley, Hoboken, N. J.

Penland, C., and P. D. Sardeshmukh (1995), The optimal growth of tropical sea surface temperature anomalies, *J. Clim.*, *8*, 1999–2024.

Procaccia, I. (1988), Weather systems: Complex or just complicated?, *Nature*, *333*, 498–499.

Sauer, T. (1993), Time series prediction using delay coordinate embedding, in *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, 643 pp., Addison-Wesley-Longman, Reading, Mass.

Sauer, T., J. A. Yorke, and M. Casdagli (1991), Embedology, *J. Stat. Phys.*, *65*, 579–616.

Schreiber, T., and A. Schmitz (2000), Surrogate time series, *Physica D*, *142*, 346–382.

Seim, H. E. (2000), Implementation of the South Atlantic Bight Offshore Observational Network, *Oceanography*, *13*(2), 18–23.

Smith, L. A. (1988), Intrinsic limits on dimension calculations, *Phys. Rev. A*, *133*, 283–288.

Smith, L. A. (1992), Identification and prediction of low dimensional dynamics, *Physica D*, *58*, 50–76.

Sugihara, G., and R. M. May (1990), Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, *344*, 734–740.

Takens, F. (1981), *Detecting Strange Attractors in Turbulence*, *Lect. Notes Math.*, vol. 898, Springer-Verlag, New York.

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer (1992), Testing for nonlinearity in time series: The method of surrogate data, *Physica D*, *58*, 77–94.

Theiler, J., P. S. Linsay, and D. M. Rubin (1993), Detecting nonlinearity in data with long coherence times, in *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, 643 pp., Addison-Wesley-Longman, Reading, Mass.

K. R. Helfrich and L. J. Pratt, Physical Oceanography Department, Woods Hole Oceanographic Institution, MS 21, Woods Hole, MA 02543, USA. (khelfrich@whoi.edu; lpratt@whoi.edu)