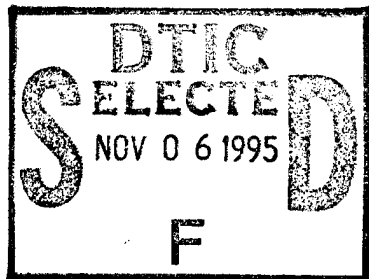NPS-OR-95-010

NAVAL POSTGRADUATE SCHOOL
Monterey, California

MODELING LONG TERM DEPENDENCE,
NONLINEARITY AND PERIODIC PHENOMENA
IN SEA SURFACE TEMPERATURES
using MARS

Peter A. W. Lewis
Bonnie K. Ray

19951102 073   September 1995

Approved for public release; distribution is unlimited.

DTIC QUALITY INSPECTED 8

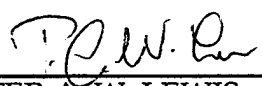NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943-5000

Rear Admiral M. J. Evans
Superintendent

Richard Elster
Provost

This report was prepared for and funded by the Naval Postgraduate School.

Reproduction of all or part of this report is authorized.
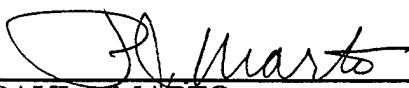
This report was prepared by:

_____
PETER A. W. LEWIS
Distinguished Professor of
Operations Research

_____
BONNIE K. RAY
Assistant Professor of Mathematics
New Jersey Institute of Technology

Reviewed by:

Released by:

_____
FRANK PETHO
Acting Chairman
Department of Operations Research

_____
PAUL J. MARTO
Dean of Research

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 29 Sept 95	3. REPORT TYPE AND DATES COVERED Technical

4. TITLE AND SUBTITLE
Modeling Long Term Dependence, Nonlinearity and Periodic Phenomena in Sea Surface Temperatures using MARS

5. FUNDING NUMBERS

6. AUTHOR(S)
Peter A. W. Lewis and Bonnie K. Ray

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-OR-95-010

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

We study a time series of 20 years of daily sea surface temperatures (SSTs) measured off the California coast. The SSTs exhibit quite complicated features, such as effects on many different time scales, nonlinear effects, and long-range dependence. We use a modified MARS algorithm to obtain univariate adaptive spline threshold autoregressive (ASTAR) models for the SSTs and discuss practical modeling issues, such as handling of cycles and long-range dependence. We approximate a nonlinear long memory model by allowing very long autoregressive terms in the ASTAR model. This large order ASTAR model is better predictively and descriptively than any other univariate model explored. Use of other concurrent predictor time series, in particular, categorical predictor variables such as wind direction, to extend the threshold autoregressive model for the SSTs to semi-multivariate adaptive spline threshold autoregressive (SMASTAR) models for the SSTs is also discussed. It is shown that SMASTAR modeling, with an added categorical time-of-year predictor, can also be used to model nonlinear structure in the data which is changing with time-of-year. Models for the SSTs are evaluated using out-of-sample forecast RMSEs, residual diagnostics, model skeletons, and sample functions of simulated series. Computational issues, such as choice of parameters in the MARS algorithm, in particular the span parameter, are discussed in an appendix.

14. SUBJECT TERMS
ASTAR; forecasting; long memory; MARS; nonlinear time series

15. NUMBER OF PAGES
38

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

MODELING LONG TERM DEPENDENCE, NONLINEARITY AND PERIODIC PHENOMENA IN SEA SURFACE TEMPERATURES using MARS

Peter A. W. Lewis
Dept. of Operations Research
Naval Postgraduate School
Monterey, CA 93943

Bonnie K. Ray*
Dept. of Mathematics and
Center for Applied Math and Statistics
New Jersey Institute of Technology
Newark, NJ 07102

September 29, 1995

Abstract

We study a time series of 20 years of daily sea surface temperatures (SSTs) measured off the California coast. The SSTs exhibit quite complicated features, such as effects on many different time scales, nonlinear effects, and long-range dependence. We use a modified MARS algorithm to obtain univariate adaptive spline threshold autoregressive (ASTAR) models for the SSTs and discuss practical modeling issues, such as handling of cycles and long-range dependence. We approximate a nonlinear long memory model by allowing very long autoregressive terms in the ASTAR model. This large order ASTAR model is better predictively and descriptively than any other univariate model explored. Use of other concurrent predictor time series, in particular, categorical predictor variables such as wind direction, to extend the threshold autoregressive model for the SSTs to semi-multivariate adaptive spline threshold autoregressive (SMASTAR) models for the SSTs is also discussed. It is shown that SMASTAR modeling, with an added categorical time-of-year predictor, can also be used to model nonlinear structure in the data which is changing with time-of-year. Models for the SSTs are evaluated using out-of-sample forecast RMSEs, residual diagnostics, model skeletons, and sample functions of simulated series. Computational issues, such as choice of parameters in the MARS algorithm, in particular the span parameter, are discussed in an appendix.

KEYWORDS: ASTAR; forecasting; long memory; MARS; nonlinear time series

1

1 Introduction and Data Description

Physical systems, such as Sea Surface temperatures (SSTs), cloud cover, wind velocities, wind speeds, and water salinity exhibit very complex and interconnected behaviors which are almost impossible to capture using standard linear time series models. In this paper, we analyze features of a very long series of daily SSTs using both linear models and nonlinear adaptive spline threshold autoregressive (ASTAR) models obtained using a modified Multivariate Adaptive Regression Splines (MARS) algorithm (Friedman, 1991a, 1991b, 1993). We also investigate models for the SSTs which incorporate shorter, covariate series of wind velocities and wind directions, as well as models which include a covariate categorical time series whose levels represent different times of the year. Section 2 gives a brief description of the use of MARS with time series. The remainder of this Introductory section discusses some physical features of the SST data in more detail.

Figure 1 shows the logarithm of the SSTs at Granite Canyon, on the coast of California approximately 30km south of Monterey, over the period March 1, 1971 to November 9, 1993. SSTs over the period March 1, 1971 to Feb. 28, 1991 are used for model estimation, while part of the additional data up to November 9, 1993 (620 days) is used for validation of the predictive ability of the derived models. The logarithm of the data is used to stabilize the variance of the SSTs (see Section 3).

The series of logged SSTs shown in Figure 1 exhibits cycles on several different time scales. A clear drop in temperatures is seen every spring, corresponding to the coastal upwelling. (The drop does not occur at precisely the same time every year). The cyclic time-of-year effect is the dominant feature in the data; note, however, that the temperatures do not increase in the summer and decrease in the winter. Following the spring transition, the log temperatures remain low and rise after the summer. It is well known (Hidaka, 1954; see also Breaker and Lewis, 1988) that this effect is predominantly due to the wind direction.

In addition to time-of-year effects, however, there is a longer cyclic effect of the periodic warming due to the El Nino phenomenon. This phenomenon is reflected in the higher temperatures seen in Figure 1 in 1972, 1976, 1983, and 1987, and has a quasi-period of about four to five years. An 18.6 year cycle in the SSTs has also been investigated (Royer, 1991).

A least-squares fit of a sinusoid with period of one year to the first 20 years of the data is shown in Figure 1. It is clearly not a complete fit of the yearly cycle (additional harmonics could be used), however it approximates the majority of the yearly effect. The futility of trying to capture these cyclic effects completely with a deterministic term can be seen from the top plot of Figure 1: the later data (March 1, 1991 and beyond) shows a massive and broad El Nino warming in progress.

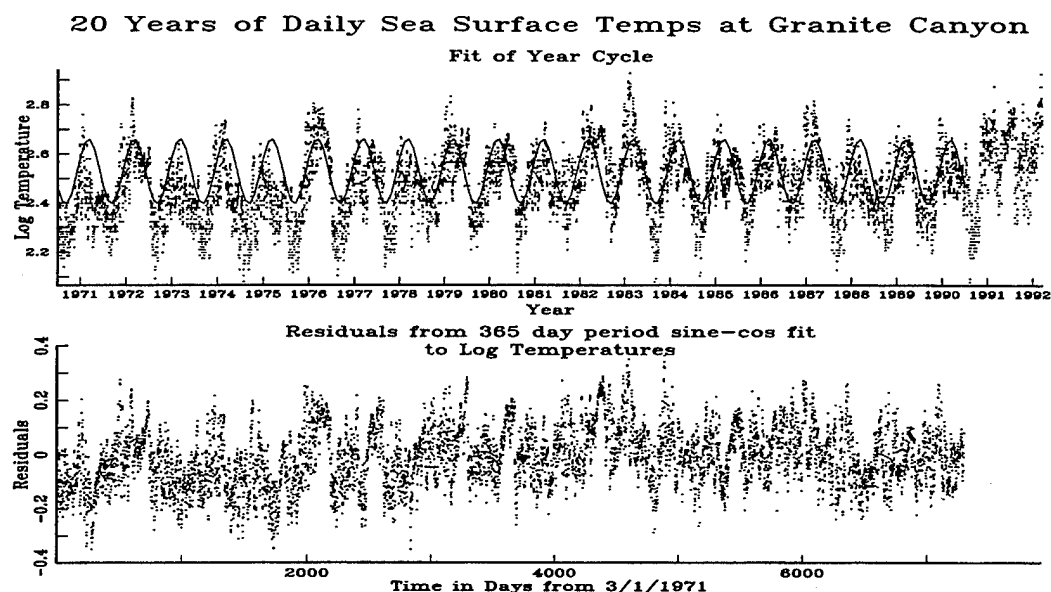**20 Years of Daily Sea Surface Temps at Granite Canyon**

Figure 1: The top plot shows the logarithm of the raw sea surface temperatures from March 1, 1971 to November 9, 1993. Only the data up to Feb. 28, 1991 is used for model estimation; a fitted sinusoid with a one year period is also shown for this period. The data beyond February 28, 1991 shows an extraordinarily broad and high El Nino effect which extended well into 1995. The lower plot shows the residuals from fitting a one year sinusoidal cycle to the data.

Therefore fitting a year cycle to the entire series would produce a sinusoid having a higher level than the sinusoid fitted using only the first 20 years of data, giving a fit to the data in the earlier years even worse than the fit shown in the top plot of Figure 1. In fact, a global linear trend term could be incorporated into the model, but this statistical artifice is clearly not acceptable to oceanographers as either a descriptive or predictive device. One knows that there may be a very long term cycle in the data, but continued warming at a rate predicted by a straight line fit would be unlikely. In Section 3.1, we discuss methods of handling the regular, yearly cyclic effects in the data when using MARS. Informally, the device of transforming the data to stabilize its variance and then subtracting an additive mean implies that the remaining data is stationary. This assumption will be examined in more detail in Section 4.

The top panel of Figure 2 gives another view of the logarithm of the data, showing more clearly the massive temperature drop in the spring, particularly following the El Nino warming in 1972. The bottom panel shows commonly used displays of the series, as either a smoothed series, or as a monthly average. This latter (low-pass filtered) series is the series which is usually studied by oceanographers.

Nonlinearities in the SSTs show up in several other ways. One is that, when a linear model is fit

3

20 Years of Daily Sea Surface Temperatures at Granite Canyon

Log of Raw Data
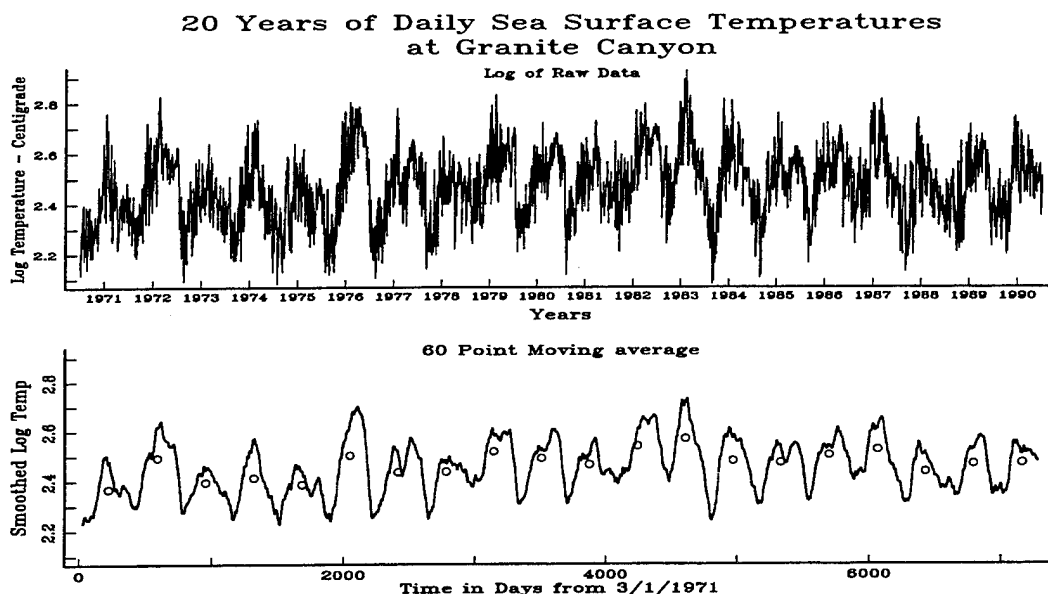
60 Point Moving average

Figure 2: The top plot again shows the logarithm of raw sea surface temperature using connected lines. The temperature drop in the spring is more clearly seen than in Figure 1. The bottom panel shows the monthly average of the series and the series smoothed using a 60-point moving average filter.

to the time series (Section 3.2), the residuals are uncorrelated but the squares of the residuals are still correlated (Granger and Anderson, 1978). Also, fitted ASTAR models (Section 3.3) clearly show the existence of thresholds, on either side of which the behavior of the series is different.

Additionally, the data appears to have long-term dependence (showing up as very long cycles of different lengths), as can be seen in the behavior of spectra of the time series near zero frequency. Figure 3 shows a plot of the log periodogram of the first 20 years of logged SSTs (depicted as circles) vs. the log of the frequency at a set of initial Fourier frequencies. The clear downward slope of the log periodogram is characteristic of long-range dependent processes (Mandelbrot and Van Ness, 1968). The logs of averaged adjacent periodogram ordinates (shown as x's) may be used to obtain a more robust straight line fit to this initial part of the spectral density, since logged periodogram ordinates are distributed as log exponential random variables, which are strongly negatively skewed. The component of the spectrum at a period of one year is omitted from the straight line fit.

We analyze the amount of long-range dependence in the SSTs and attempt to model it using both a linear long memory model and an approximating nonlinear ASTAR model with lags of up to 5 years. The different models obtained using lags of up to 5 years are evaluated in Section 3.4 on the basis of out-of-sample forecast root-mean-squared-errors (RMSEs) at different steps, residual diagnostics, model skeletons, and stability of simulated series. Note that there is no guarantee that

4

Initial part of periodogram of daily sea surface temperatures at Granite Canyon -- 20 years
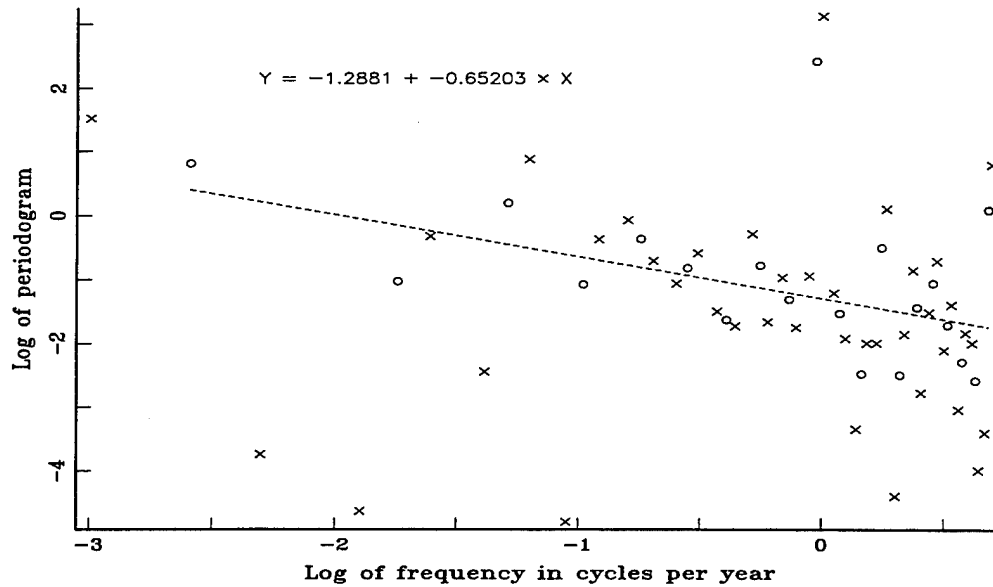
$$Y = -1.2881 + -0.65203 \times X$$

Figure 3: Plot of the logarithm of the periodogram of 20 years of logged SSTs at Granite Canyon vs. the logarithm of an initial set of Fourier frequencies. The circles depict logarithms of each periodogram ordinate. The x's are logarithms of averages of adjacent periodogram points. The latter, omitting the periodogram ordinate corresponding to the one year frequency, are used in the straight line regression for the slope of the log spectrum near zero frequency.

the models fitted to the data will be stable. This aspect of MARS modeling is briefly discussed in Section 3.4.3.

A much more subtle cyclic effect in the SSTs is an approximately 47 day oscillation (Breaker and Lewis, 1988). Previous attempts to elucidate the nature of this oscillation by, for example, complex demodulation, have proved fruitless. However it is believed that this oscillation may be related to the effect of the wind on SSTs. Wind speed and wind directions are two of the most important meteorological variables affecting SSTss and reliable observations of this data are more available today than in the past. For Granite Canyon, wind speed and direction data were available from January 1, 1985 to March 31, 1991. Wind speeds, which are continuous valued variables, can be incorporated into a Semi-multivariate ASTAR (SMASTAR) model in a fairly straightforward manner (Lewis and Stevens, 1992). However, wind direction, which is a circular variable, is better treated as a categorical variable. Wind direction at Granite Canyon is measured in positive degrees from 0 to 360 and reported as the direction in which the wind is blowing. We have coded the wind direction (WD) as a categorical variable using the following categories:

5

EFFECT OF WIND DIRECTION ON LOGARITHM OF SEA SURFACE TEMPERATURES

Temperature lags Wind Direction by 1 day   Temperature lags Wind Direction by 6 days

Temperature lags Wind Direction by 11 days   Temperature lags Wind Direction by 16 days
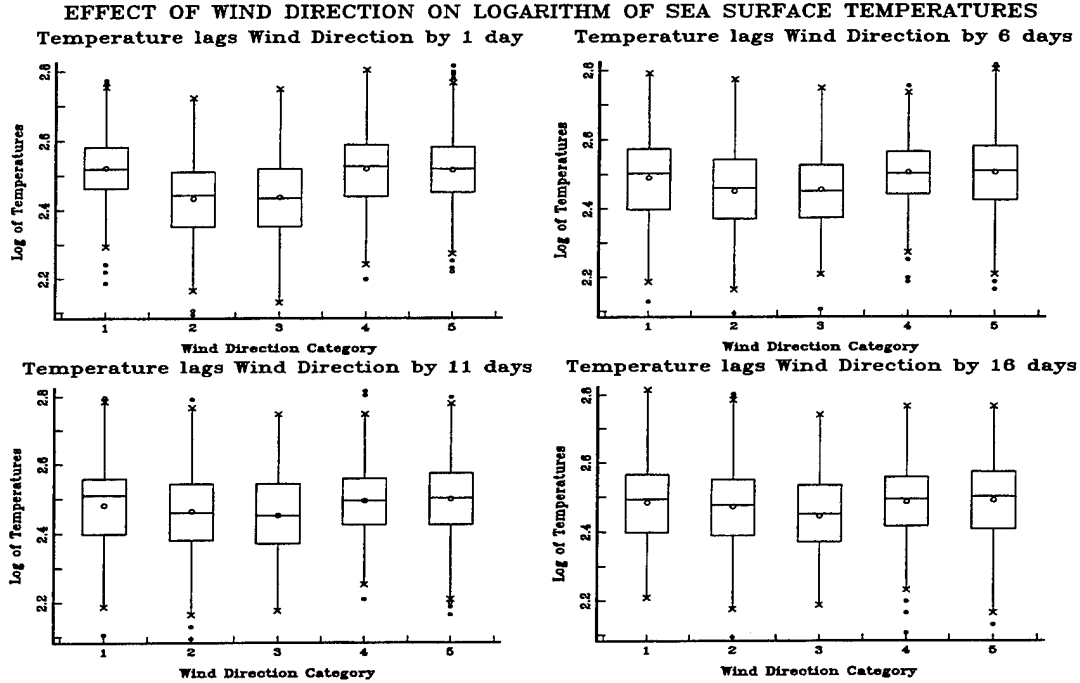
Figure 4: The effect of wind direction on the SSTs at Granite Canyon. Note that there are only five years and 90 days of data involved, from January 1, 1985, to March 31, 1991.

- 1=East: $0° \leq WD \leq 45°$ or $315° < WD \leq 360°$,

- 2=North: $45° < WD \leq 135°$,

- 3=West: $135° < WD \leq 225°$,

- 4=South: $225° < WD \leq 315°$.

Days in which no wind or only light airs were reported received a code of 5. (This data collection was not automated and the data was spotty and has been filled in parts. Collection started on January 1, 1985, and lasted for five years and ninety days.)

The effect of wind direction on SSTs is clearly shown in Figure 4. Using the 5 years and 90 days of data, the figure depicts the range of temperatures reported for each wind category when temperature lags wind direction by 1, 6, 11, or 16 days. Figure 4 shows clearly that the SSTs depend on the direction in which the wind is blowing; they tend to be lower when the wind blows from the Northwest (Categories 2 and 3), and this dependence is strongest at lag 1. It is clearly necessary to bring in this aspect of the physical situation for a complete model of the SSTs. In fact, oceanographers agree that the spring transition is driven by the shift in the wind direction to the Northwest in the spring. Thus a good descriptive model should show this nonlinear effect on SSTs.

In Section 4, we describe this type of extended analysis of the data using a SMASTAR model having wind speed as a covariate time series. We also use wind direction as a categorical valued covariate time series to obtain a CASTAR model — a semi-multivariate model in which the covariate time series may be categorical. The resulting CASTAR analysis shows that the oscillation seen by Breaker and Lewis (1988) is only present – as a 49-day autoregressive component – when the wind blows from the Northwest. In fact, the resulting CASTAR model for the SSTs at Grand Canyon given in Section 4.1 is very satisfying in that it gives an explicit model of the interaction of the three time series which is in accordance with behavior known to oceanographers.

Previous attempts at modeling the SSTs using MARS have been reported in Stevens (1991), Lewis, Ray, and Stevens (1993), and Lewis and Ray (1993). The analysis presented in this paper differs from earlier work in several ways. First, we investigate the ability of MARS to model seasonal effects. We find that explicitly subtracting a fitted sinusoidal cycle from the data before using MARS gives the best predictive models. Second, we incorporate lagged values of up to 5 years in the ASTAR models in an attempt to capture long-range dependent behavior. The ASTAR models of Lewis and Ray (1993) included only lags up to 365 and 50 days, respectively. It was clear from the ASTAR analysis that lags out to 365 days were not sufficient to account for long-range dependence or the El Nino warming. (In the CASTAR model given in Section 4.1, where wind speed and wind direction were introduced into the model, the 50 day model was retained because the data was not extensive enough to support the extension of the dependence to five years.) Third, we illustrate the inclusion of a covariate categorical time series whose levels represent different times of the year into the model to describe a cyclic nonstationary in the SSTs; the resulting models are such that the entire nonlinear structure of the SSTs, as opposed to only its mean value, changes with the time of year. They are related to the periodic autoregressive models used in hydrology (see, e.g., McLeod, 1994) but are more general. In fact, they are periodic autoregressive threshold models.

Section 5 summarizes our findings and gives directions for future research. Computational issues related to the MARS algorithm for time series, such as choice of model span and number of interactions to allow, are discussed in an Appendix.

2   Using MARS for Time Series Modeling

Threshold models (models with partition points) are a class of nonlinear models that emerge naturally as a result of changing physical behavior. Within the domain of the predictor variables, different model forms are necessary to capture changing relationships among the predictor and response variables. Tong (1983, 1990) provides one threshold modeling methodology for this behavior, Threshold Autoregression (TAR), which identifies piecewise linear pieces of nonlinear functions

over disjoint subregions of the domain D of the time series $\{x_\tau\}$, i.e., the identification of linear autoregressive models within each disjoint subregion of the domain.

As a simple example, take the first order (lag one) case with one fixed threshold, x_c, which has been studied by Petrucelli and Woolford (1984):

$$X_t = \mu + \rho_1(X_{t-1} - x_c)^+ + \rho_2(x_c - X_{t-1})^+ \qquad (1)$$

Here the $+$ indicates that the term is zero unless the quantity in the parentheses is positive. If $\rho_1 = \rho_2$ we have an ordinary first order linear autoregressive process.

TAR modeling methodology has tremendous power and flexibility for modeling of many times series. However, unless Tong's TAR methodology is constrained to be continuous, it creates disjoint subregion autoregressive models that are discontinuous at subregion boundaries. Nor is its implementation systematic.

By letting the predictor variables for the τth value in a time series $\{x_\tau\}$ be $x_{\tau-1}, x_{\tau-2}, \ldots, x_{\tau-p}$, and combining these predictor variables into a linear additive function, one gets the well known linear AR(p) time series models. Analogously, using the Multivariate Adaptive Regression Splines (MARS) methodology, a recent method for *nonlinear* regression modeling due to Friedman (1991), to model the effect of $x_{\tau-1}, x_{\tau-2}, \ldots, x_{\tau-p}$, one can still obtain autoregressive models. *However, these models, called Adaptive Spline Threshold Autoregressive (ASTAR) models (Lewis and Stevens, 1991) can be nonlinear models in the sense that the lagged predictor variables can have threshold terms, in the form of truncated spline functions (Friedman, 1991a) and can also interact with the nonlinear terms of other lagged predictor variables.* With MARS, by letting the predictor variables be lagged values of a time series, one overcomes the limitations of Tong's approach and admits a more general class of continuous nonlinear threshold models than permitted in TAR models.

Numerous simulation studies have been conducted to evaluate the ability of the ASTAR methodology to identify and evaluate simple linear and nonlinear times series models (Stevens, 1991; Lewis and Stevens, 1991). The use of various model selection criteria has also been examined (Stevens, 1991). The Schwarz-Rissanen criterion (SC), a time series model selection criterion based on stochastic complexity analysis, was found by Stevens (1991) and Lewis and Stevens (1991) to be appropriate when using MARS in a time series setting. In particular, it gives far more parsimonious models than are obtained with the Generalized Cross Validation criterion used by Friedman (Lewis and Ray, 1993). We use the SC criterion to select all models investigated in this paper.

To provide a framework for both the univariate (ASTAR) and the semi-multivariate time series models (SMASTAR) discussed in the following Sections, suppose that for $\tau = 1, 2, \ldots, N$, $\{Y_\tau\}$

8

and $\{Z_\tau\}$ denote the input time series and $\{X_\tau\}$ the output time series for a time series system we wish to model. Following Lewis, Ray and Stevens (1993) and using the notation $\|$ of Tong, Thanoon and Gudmundsson (1985) to separate the predictor variables of each different time series, we can describe X_τ with the semi-multivariate time series regression model

$$X_t = f\left(1 \parallel X_{t-1}, X_{t-2}, \ldots, X_{t-d_1} \parallel Y_t, Y_{t-1}, \ldots, Y_{t-d_2} \parallel Z_t, Z_{t-1}, \ldots, Z_{t-d_3}\right) + \epsilon_\tau, \qquad (2)$$

where 1 denotes a model constant and the maximum lags d_1, d_2, and d_3 are not necessarily equivalent. Also, Y_t and Z_t, the current values of the predictive time series, may or may not be included in (2), depending on the time series system. An example of this general form is given by Equation 1, which has only X_{t-1} as input and only one threshold.

The form of the function $f(.)$ in Equation 2 is determined by the data, and the nature of the inputs *as specified by the user*. Thus the input to the MAR program (say, Z_{t-1}) may be specified to be

1. an ordinal variable with no restrictions, so that it may have thresholds and interactions with other variables;

2. the same as above, but with no interactions, i.e., additive;

3. an ordinal variable, with no thresholds (linear) and no interactions;

4. a categorical variable, with no restrictions, or a categorical variable, with no interactions with other variables.

Categorical data input is new to the latest version of MARS (MARS 3.6A), and will be discussed in Section 4 in relation to CASTAR models.

3 Modeling SSTs using ASTAR and ARFIMA models

It is well known that the SSTs exhibit high, positive-valued autocorrelations at lags between one and, approximately, forty, and that the autocorrelation function does not seem to be modeled by the usual linear ARMA(p,q) models (Breaker, Lewis and Orav, 1988). In this section, we focus on modeling possible long-term dependence, in conjunction with yearly periodic effects, using linear ARFIMA models and nonlinear MARS-based ASTAR models.

3.1 Handling Cyclic Effects using MARS

In examining the daily SST data, the variance of the average monthly temperature over the years was found to increase as the mean of the average monthly temperature increased. To stabilize this

variance, we take the natural logarithm of the data and work with the logged values throughout the analysis. Note, however, that transformations must be used with care, partly because they make the results less interpretable by users, but also because in nonlinear time series they induce complications.

The dominant effect in the SSTs is the one-year cycle, thus it is appropriate to consider how to handle it before proceeding. Note that it is truly a fixed effect, unlike the cycle in the Wolfer sunspot data, which was analyzed using MARS in Lewis and Stevens (1991). That it is a true cycle can be shown by computing the amplitude of the yearly spectral component for the first year, the next two years, the next three years, *etc.*. A regression analysis shows that the amplitude increases linearly in n, the number of days.

We consider here three ways to handle this cyclic effect.

1. Use the general form of the univariate ASTAR model in Equation 1 with an extensive range of lagged $X'_{t-j}s$ allowed to come into the model, with thresholds and interactions between lagged values. This general nonlinear model can generate cyclic processes; in Lewis and Stevens (1991) this procedure gave an excellent model for the Wolfer sunspot series, with the threshold autoregressive model having a limit cycle. Of course even a linear, first order autoregressive equation with parameter close to one will generate long 'cycles', but the cycles will vary in length.

2. An alternative is to subtract from X_t a fitted sinusoidal function and proceed as above. Thus let

$$S_t = \mu + \alpha \cos(2\pi t/365) + \beta \sin(2\pi t/365) \tag{3}$$

and let

$$X'_t = X_t - S_t. \tag{4}$$

MARS is then used to model the transformed process, X'_t. *Although it is shown below that this gives possibly the best predictive model for the Granite Canyon SSTs, it is awkward and restrictive as an explanatory model.* To see this, consider Equation 1 in the variable X'_t and transform it back to the variable $exp(X_t)$. The threshold on $exp(X_t)$ will be cyclic! Moreover, in practical terms, this implies that the nonlinearity — the multiplier ρ_2 changing to ρ_1 — occurs when the variable gets higher than a "time of year level" rather than when the variable reaches an absolute level.

3. One could also use both lagged values of the SSTs and the one-year cosine and sine terms as covariate time series in Equation 2. This is a more general model than the first. However when this was investigated for the SSTs, the sine and cosine terms seldom came into the final model, even with 20 years of data, and the resulting predictions were extremely poor.

In fact, MARS seems very poor at detecting a fixed cyclic effect with a long period. As a simple test, we generated series from a first-order Gaussian autoregressive process with an added sinusoid and used MARS with a sinusoidal covariate and one lagged predictor to obtain strictly linear models for the series. Other MARS parameters were set commensurate with those used to model the Granite Canyon data. In 50 replications, the resulting MARS model only once included a sinusoid, however the estimates of ρ were better than those obtained by the usual device of fitting a cycle by least squares, subtracting the cycle, and estimating ρ from the residuals.

Consequently, for all the univariate MARS models discussed in the following sections, a yearly cycle was removed from the logged data prior to modeling and predicting, as in Equations 3 and 4. As noted above, this may not be the best procedure from an explanatory or descriptive viewpoint, but it does give the best predictors, as we discuss in Section 3.4.1.

This same variance stabilizing and cyclic detrending was applied to the long memory fractionally integrated ARMA models – ARFIMA – discussed in the next subsection. Since they are linear models, the detrending does not give rise to the problems which occur in nonlinear models.

3.2   Handling Long Memory using ARFIMA models

The existence of long-term dependence in physical systems in the form of very long cycles and apparently shifting mean levels has been the subject of much study in the last 20 years (*e.g.* McLeod and Hipel, 1978; Hosking, 1984; Haslett and Raftery, 1989). By long-term dependence, we mean that the dependence between observations k time periods apart decays at rate $k^{-\alpha}$, $0 < \alpha < 1$. Equivalently, the spectrum of the process at (low) frequency α is proportional to $\omega^{-\alpha}$ and thus is unbounded at zero frequency. Figure 3 shows that the Granite Canyon SSTs have a sample spectrum which appears to behave like that of a long-memory process. A similar behavior has been seen in 60 years of daily SSTs recorded at Pacific Grove, as well as other set of SSTs (Mendelson, 1990 and personal communication).

Several authors have proposed *linear* long-term time series models which describe the slowly decaying structure of such series (Mandelbrot and Van Ness, 1968; Hosking, 1981; Jonas, 1983). A commonly used linear model for describing long-term dependence is the fractionally integrated ARMA (ARFIMA) model of Granger and Joyeux (1980) and Hosking (1981), a generalization of Box-Jenkins ARIMA models in which the degree of integration, d, is allowed to take fractional

11

values. The fractional integration operation $(1-B)^d$ is defined via the binomial expansion $(1-B)^d = \sum_{i=0}^{\infty} \binom{d}{i}(-B)^i$, where B denotes the backwards differencing operator. If $d > 0$ the resulting time series has long memory.

We use a linear ARFIMA model to describe the low frequency components of the 20 years of daily SSTs at Granite Canyon, after first removing the one year cycle. Note that removing the cycles does not change the characteristics of the residual data (Yajima, 1988). Because of the length of the data, we analyze it using a two-step procedure. The long-memory parameter, d, is estimated first, using the periodogram spectral regression method of Geweke and Porter-Hudak (1983) with number of regressors equal to $m = [\sqrt{n}] = 85$. There has been some criticism of the method, which has trouble distinguishing between long memory and large autoregressive or moving-average components in small samples (Agiakloglou, Davis, and Wohar, 1993). However, given the length of our series, the method is justified. For the 20 years of Granite Canyon SSTs, we obtain an estimated d of 0.37, with standard error equal to 0.076. This is consistent with the slope of the spectral density for the SSTs shown in Figure 3.

This estimate of d is used to approximately fractionally "difference" the SST data using the infinite autoregressive representation of the fractional differencing operator, truncated at lag 500. The sample autocorrelation and partial autocorrelation function of the fractionally differenced data indicated that it contained short memory ARMA components. The differenced data appears to have the structure of an AR(1) process, with estimated coefficient equal to .606 (standard error of 0.009). Thus there is relatively strong correlation between SSTs from one day to the next, as well as a significant amount of long-range dependence in the temperatures. The result of this procedure is an ARFIMA(1,d,0) model for the logged and detrended SSTs.

The ACF of the residuals from the fitted ARFIMA$(1, d, 0)$ model (Fig 5, top), show little correlation. However, given the nonlinear nature of the data already discussed, it seems highly unlikely that the linear ARFIMA model adequately describes the data. Examining the ACF for the squared residuals, as suggested by Granger and Anderson (1978), shows (Fig 5, bottom) that the squared residuals retain some correlation. This is an indication of nonlinearity in the data not captured by the fitted ARFIMA(1,d,0) model.

3.3   Handling Long Memory using MARS

Threshold autoregressive models such as those generated by MARS can exhibit wide ranging modes of behavior. Lewis and Stevens (1991) fitted an ASTAR model to the Wolfer sunspot numbers which was found to have a limit cycle. As for long memory, it has been conjectured (Tong; discussion of

Best ARFIMA MODEL

Autocorrelation function for residuals

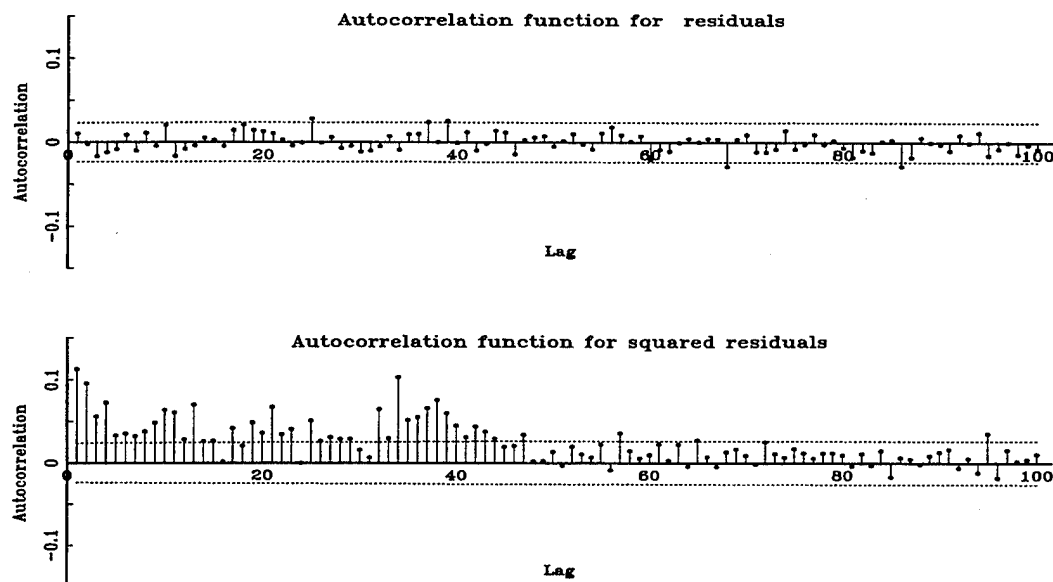

Autocorrelation function for squared residuals

Figure 5: The top panel shows the autocorrelation function for the residuals from the fitted ARFIMA(1,d,0) process and the Granite Canyon SSTs. The bottom panel is the autocorrelation function for the squared residuals. There is clearly autocorrelation remaining in the squared residuals. The dotted lines indicate upper and lower bounds on the estimated ACF assuming the residuals are white noise.

Haslett and Raftery, 1989, and personal communication) that even a first order TAR model may exhibit this behavior, although no proof has yet been derived.

An invertible long memory $ARFIMA(p, d, q)$ process has an infinite autoregressive representation. Thus it is reasonable to try and approximate long-term effects using an autoregressive model with many lags. For a linear fractional noise process (an $ARFIMA(0, d, 0)$ process), Ray (1993) finds that an approximating $AR(p)$ model performs well, in terms of forecast error variance, when used to forecast values of the fractional process at long lead times. Thus for nonlinear ASTAR models, we attempt to incorporate long-term dependence by allowing the number of lagged predictor values of the series to be very large. In Lewis and Ray (1993) lags out to 365 days were used, but the models could not capture, in particular, the El Nino effect. In this paper, we apply MARS to model 20 years of the Granite Canyon data, (with one-year cycle removed) using all lagged SSTs up to lag 100, and every fifth lag thereafter up to lag 1925, representing lags up to approximately 5 years and 3 months. (Using all lags to 1925 is computationally infeasible; the limited grid here represents a ten fold increase over the computation times in Lewis and Ray (1993)).

We allowed the SPAN parameter, which basically acts as a smoothness parameter, and the

13

maximum number of interactions, which determines the complexity of the model, to vary in order to investigate the effect these parameters have on the final model. A more detailed discussion of the effect of these parameters is given in the Appendix. We present the results for models having SPAN 1 with no allowable interactions between terms, and for models allowing up to three interactions (MI=3). Models having SPAN 25 and SPAN 50 with MI=3 allowable interactions are also discussed.

Following Equations 3 and 4, let the residual process be

$$X_t = \ln(S_t) - 2.467 - 0.10\sin(2\pi t/365) - .05\cos(2\pi t/365), \tag{5}$$

where S_t is the SST at day t. Using SPAN=1 and no allowable interactions (MI=1) when applying the MARS algorithm to X_t, the resultant model has the following form:

Nonlinear Long Memory Model for 20 Years of SSTs at Granite Canyon
SPAN 1; no interactions allowed (MI=1)

$$\widehat{X}_t = \left\{ \begin{array}{ll} -0.088 & +0.994(X_{t-1} + 0.116)_+ - 0.838(-0.116 - X_{t-1})_+ \\ & -0.131(X_{t-2} + 0.340)_+ + 0.035(X_{t-8} + 0.172)_+ \\ & +0.035(X_{t-17} + 0.340)_+ \end{array} \right. \tag{6}$$

Given that the minimum value of X_t over the 20 years is -0.340, we see that lags 2 and 17 always enter the model (no interior threshold). The threshold in the lag one term in the model also indicates that SST is strongly correlated with the previous day's SST, with temperature increasing (coefficient +0.994) if the temperature the previous day was more than -0.116 and decreasing (coefficient -0.838) if it was less than -0.116. The model does not contain terms with lags greater than 17, as would be expected for a long range dependent process. This suggests that the long memory behaviour may only be observed in conjunction with other variables, i.e., interactively.

When 3 interaction levels are allowed (MI=3), however, the resulting model for X_t has the following form:

Nonlinear Long Memory Model for 20 Years of SSTs at Granite Canyon
SPAN 1; order 3 interactions allowed (MI=3)

$$
\widehat{X}_t = \begin{cases}
-0.078 & +1.005(X_{t-1}+0.116)_+ - 0.816(-0.116-X_{t-1})_+ \\
& -0.120(X_{t-2}+0.340)_+ \\[6pt]
& +6.689(X_{t-2}+0.340)_+(X_{t-1415}+0.282)_+ \\[6pt]
& +0.332(X_{t-2}+0.340)_+(0.351-X_{t-3})_+(X_{t-8}+0.172)_+ \\
& -0.924(X_{t-2}+0.340)_+(0.045-X_{t-17})_+(X_{t-67}+0.158)_+ \\
& -40.594(X_{t-2}+0.340)_+(X_{t-17}-0.045)_+(-0.221-X_{t-1425})_+ \\
& -672.872(X_{t-2}+0.340)_+(X_{t-435}-0.071)_+(-0.282-X_{t-1415})_+
\end{cases} \tag{7}
$$

We see that the first order terms are very similar to those of the previous model; however, lags 8 and 17, as well as higher order terms, enter the model interactively. The existence of long lag terms (up to 1425, or slightly less than four years) in the model reinforces the postulation of long-range dependence in SSTs. Terms with lags of 8 and 17 probably reflect the fact that the average time between storm fronts in the vicinity of Granite Canyon in the winter is about 8 days. The last three-way interaction, having coefficient -672.872, causes this model to be unstable (see Section 3.4.3)

With SPAN=25, the resulting model for X_t has the following form:

Nonlinear Long Memory Model for 20 Years of SSTs at Granite Canyon

SPAN 25; order 3 interactions allowed (MI=3)

$$
\widehat{X}_t = \begin{cases}
-0.073 & +1.017(X_{t-1}+0.116)_+ - 0.856(-0.116-X_{t-1})_+ \\
& -0.115(X_{t-2}+0.340)_+ \\[6pt]
& -20.607(X_{t-1}+0.116)_+(X_{t-11}+0.050)_+(-0.202-X_{t-1635})_+ \\
& -68.653(X_{t-1}+0.116)_+(X_{t-33}-0.197)_+(0.095-X_{t-37})_+ \\
& -4.917(X_{t-1}+0.116)_+((0.095-X_{t-37})_+X_{t-122}-0.055)_+ \\
& -34.884(X_{t-1}+0.116)_+(X_{t-850}+0.340)_+(-0.222-X_{t-1395})_+ \\
& +14.159(X_{t-2}+0.340)_+(X_{t-8}+0.172)_+(X_{t-1395}+0.213)_+ \\
& -0.989(X_{t-2}+0.340)_+(-0.045-X_{t-17})_+(X_{t-67}+0.158)_+
\end{cases} \tag{8}
$$

Again, the first order terms are very similar to those of the previous models, however, this model contains no interactions of level 2. The model resulting from setting SPAN=50 was very similar to the above model, but contained only 3 terms of interaction level 3. Two of the 3 interaction level 3 terms were identical to terms 6 and 9 of the above model. Thus it appears that using SPAN=50 results in a "smoother" model. Also there are no three-way interactions with very large coefficients, as in the previous model, and the model can be shown (empirically) to be stable.

At this time, we do not know of a TAR model that is also truly long-range dependent, as opposed to an approximating model. Other types of nonlinear models having long-range dependence, such as

15

fractionally integrated models with ARCH errors or random solutions to the Burgers' equation, have been discussed by Robinson (1991), Rosenblatt (1987) and Taqqu (1987). In the next section, we compare the fitted ARFIMA model, as well as other linear models for the SSTs, with approximating long-range dependent ASTAR models.

3.4 Comparison of Nonlinear and Linear Models

In addition to the linear ARFIMA model, we fit linear time series models to the SSTs using standard ARMA modeling techniques. Using the SC criterion and the APL2 AGSS package, an AR(8) model was selected for the logged and detrended SSTs from AR(p) models of order $p = 1, 20$. This fitting can also be done using MARS, with entering terms restricted to be linear and having no interactions. Using MARS (with SC criterion), a linear model containing lags 1, 2, 8, and 17 was selected. In the following sections, we compare the estimated MARS linear and nonlinear models on the basis of their forecasts, residuals, and model skeletons.

3.4.1 Out-of-sample Forecasts

Predicted SSTs are used as input to large-scale weather models, so the predictive capability of the models is extremely important. Figure 6 shows the Root Mean Square Errors (RMSE) of predicted k-step ahead forecasts up to 600 days ahead, beginning 3/1/91, for the linear and nonlinear models discussed above. For comparison, we also include the RMSE obtained if the last observed value in the time series is used for prediction. The RMSEs were obtained using a moving forecast origin, forecasting k-steps ahead, and taking the square root of the average squared difference between forecast and actual value. We see that the RMSEs do not increase monotonically in k, which is another indication of the nonlinearity of the data (Casdagli, 1992). In fact, accuracy of nonlinear forecasts is extremely dependent on the region over which the forecasts are made. We extend our forecasts out to 600 days in order to include more than one year in the forecast region.

We see that all models are competitive for small forecast steps. This is not suprising, given the very high correlation between adjacent SSTs. However, the situation is different at larger forecast steps. The MARS model with SPAN=1 and MI=3 levels of interaction appears to be the best predictive ASTAR model. The SPAN 25 and SPAN 50 models are competitive at steps > 200. The linear MARS model performs better than the linear ARFIMA model at all lags. This may be because the ARFIMA$(1, d, 0)$ model is not flexible enough in terms of its correlation structure. A more complicated ARMA part of the ARFIMA model may be needed to adequately capture the short-range dependence in SSTs. The AR(8) model appears to be just as good as the ARFIMA model at steps > 200.

16

RMSE OF K–STEP FORECASTS FOR GRANITE CANYON SSTs
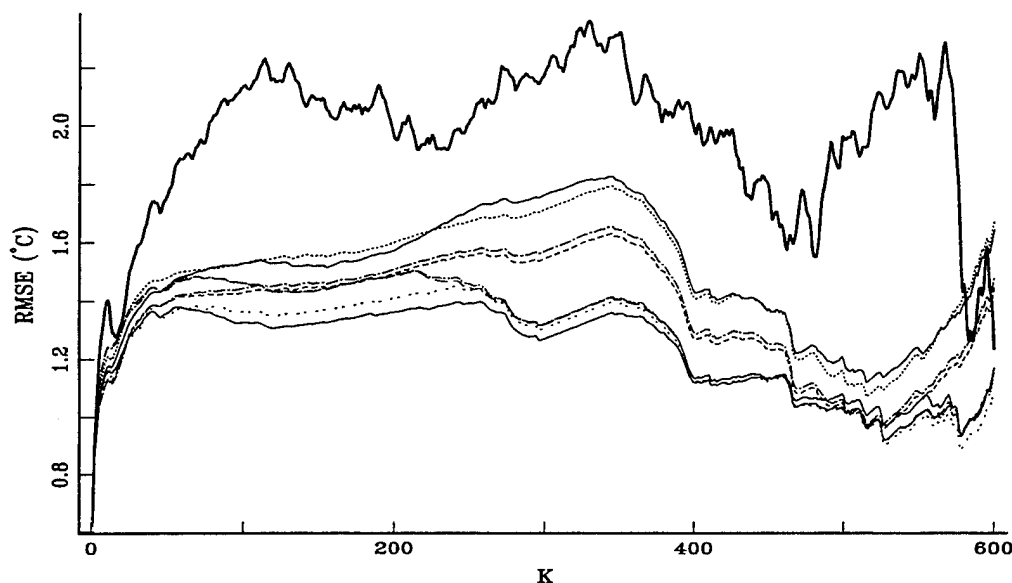### Period beginning 3/1/91

K

Figure 6: The Root Mean Square Errors (RMSE) of k-step ahead forecasts for the Granite Canyon SSTs. The forecast period begins on March 1, 1991, and extends for 600 days. To differentiate the forecasts of different models, examine lag 300. The topmost solid curve is the forecast using the SSTs at the end of the observation period, *i.e.* the last observed value. The next lightly solid curve is the ARFIMA(1,d,0) forecast, with the AR(8) forecast curve immediately below (depicted as dot-dot). The next (dot-dash) curve is the forecast for the Threshold MARS (SPAN 1) model, with the 'dash-dash' curve for the Linear MARS (SPAN 1) forecast below that. Continuing downward, the next group of curves includes the Interactive MARS forecasts for different spans. The SPAN 25 and SPAN 50 are almost indistinguishable here (almost solid line) while the SPAN 1 case (dot-dot space) is lower than these two. Finally, a solid forecast curve (obtained using methods of Section 4) is shown.

The solid curve having the lowest forecast errors in the range of 50 to 400 steps-ahead is a model which uses a categorical series to indicate time-of-year effects. This model will be discussed in Section 4.
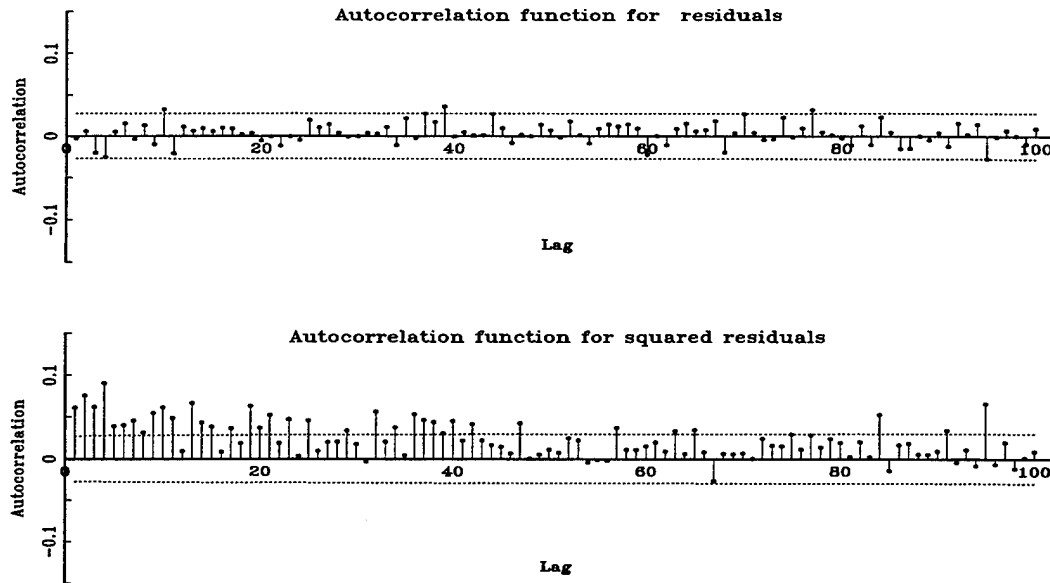
3.4.2 Residual Diagnostics

In Figure 8 we show, in the top panel, the autocorrelation function for the residuals from the fitted threshold MARS model with SPAN 1 to the twenty years of Granite Canyon SSTs. The dotted lines indicate upper and lower bounds on the estimated ACF assuming the residuals are white noise. Clearly the (linear) residuals are uncorrelated, however, this does not mean that they are necessarily independent. In fact, the bottom panel shows the autocorrelation function for the

17

RMSE OF K–STEP FORECASTS FOR GRANITE CANYON SSTs
Period beginning 3/1/91
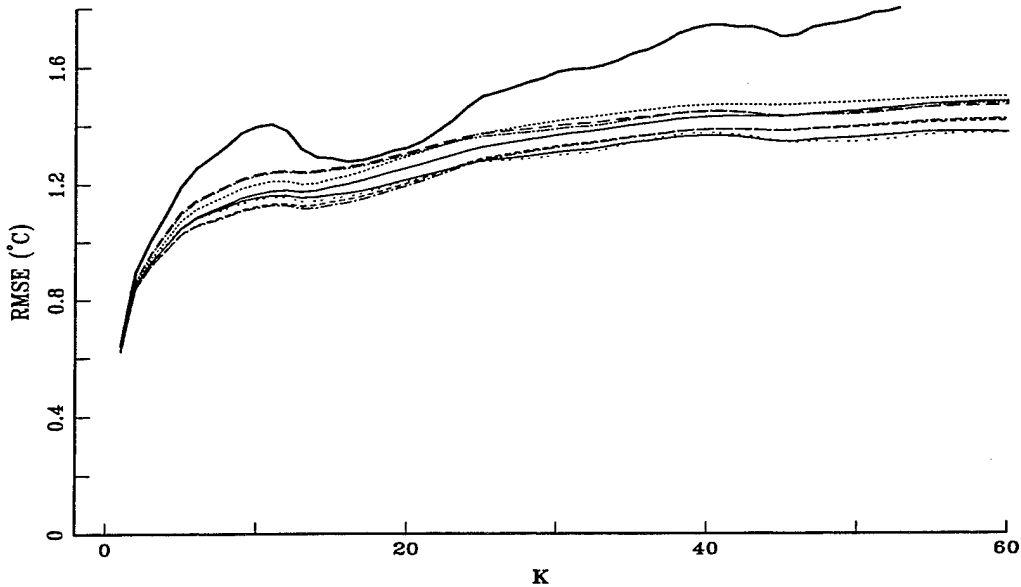
Figure 7: This is an expansion of Figure 6, showing the RMSE of forecasts up to 60 steps ahead more distinctly.

squared residuals, and they are clearly not uncorrelated.

The autocorrelation functions of Figure 8 are similar to those of Figure 5 for the fitted ARFIMA(1,d,0) process, and both suggest the kind of pattern seen in ARCH processes (see, Tong, 1991, p. 115), $i.e$ processes having variance which changes conditionally as a function of lagged values of the squared process. However, before investigating an ARCH model for the SSTs, it is necessary to examine some of the assumptions made in the analysis of the SST data.

Specifically, we have assumed that the dominant year cycle manifests itself both in the mean value and the standard deviation of the process. No fitted ASTAR model showed a limit cycle of 365 days, and therefore the cyclic effect on the mean value was taken out (approximately) with an additive sinusoidal component for the logarithm of the data, the logarithm being an (approximate) variance stabilizing transform.

However the whole structure of the process might be changing with time of year, and since the data is so extensive, this can be examined. (Note that in a threshold model the structure changes with the level of the process; no cyclic time thresholds have been postulated.) Figure 9 shows the mean, the standard deviation, and the Lag 1 Serial Correlation for each of the 365 days in the year. The plots are formed by averaging across the twenty years of logged data and then smoothing a small amount along the time axis. The mean value of the raw logged data (upper panel) is,

Best MARS model; no categorical season

Autocorrelation function for residuals

Autocorrelation function for squared residuals

Figure 8: The top panel shows the autocorrelation function for the residuals from the fitted threshold MARS model with SPAN 1 and the Granite Canyon SSTs. The bottom panel is the autocorrelation function for the squared residuals.

as we know, time-of-year dependent. The standard deviation is considerably more time-of-year independent than without the log transform, but the rather high lag one serial correlation changes with time-of-year and not always in the same way that the mean is changing. Given the multiplier 1.005 of the X_{t-1} in Equation 8 when $X_{t-1} > -0.116$, the size of ρ_1 is not surprising. However it does not remain high whenever X_{t-1} is high. We are, of course, neglecting interactive terms in the equation.

To cope with this clear nonstationarity in the data, a categorical time of year MARS model (CASTAR) is considered in Section 4.

3.4.3 Model Skeletons and Simulated Periodograms

It is possible for nonlinear models to generate processes having limit cycles, or chaotic structure. To check for limit cycles in the SSTs, we generated model skeletons (Tong, 1990, p. 96) using the nonlinear MARS models. The behavior of the sample path of the skeleton is dependent on starting values. No limit cycles were found in any of the models.

We also looked at the log periodogram versus log frequency for series generated from the nonlinear models, *i.e.* from Equations such as (8) driven by simulated white noise, to see if the models

19

LOG SEA SURFACE TEMPERATURES
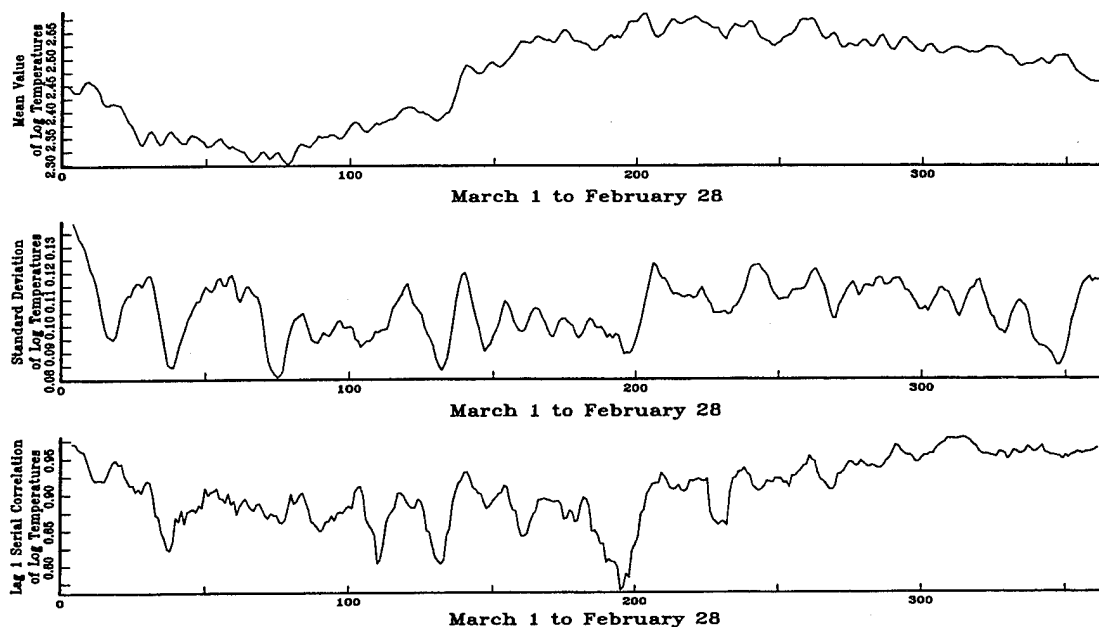
Figure 9: By averaging across the twenty years of SSTs and smoothing along the time axis, one can obtain a rough picture of the stability of the structure of the SSTswith time-of-year. The graphs show that the mean and standard deviationof the process are fairly constant over time, but that the first serial correlation coefficient is definitely changing with the time-of-year.

adequately captured the long-range dependence behavior of the series. This is a very difficult thing to do in any precise way via simulation; all we can conclude by examining a few very long simulated time series is that the simulated series appear to behave like long-range dependent processes.

The simulation study did yield other interesting results, however, which show how tricky an empirical investigation of this sort can be. In simulating values using the SPAN 1 model, we found that the estimated model was unstable. One can see the cause of this in Equation 7. The multiplier -672.872 of the three-way interaction on the last line is so large that it ultimately builds up an unstable oscillation with period 435, 1415 or 2 days. The interaction is clearly quite complex. One can say then that the models with small SPAN (little smoothing) are ultimately unstable, but are generally the best long term predictors. The difference in predictive power, however, as can be seen from Figure 6, is not great.

4   SMASTAR and CASTAR Models

A recent implementation of the MARS algorithm (Friedman, 1991) allows for categorical predictor variables as input to MARS. Adapting this implementation to time series implies that if one had n

categorical predictor variables and each was constrained in the MARS algorithm to enter linearly and without interactions, then each value of each of the categorical predictor variables would contribute a possibly different additive constant to the model for x_t at time t. If only interactions among the categorical variables were allowed, then more complex patterns of dependence of x_t, in the sense of the additive model constant, can occur. If interactions between categorical and ordinal variables are allowed, then a different autoregressive threshold model in the ordinal variables is obtained for each different combination of the categorical variables.

We use this added facility of MARS now to obtain semimultivariate models of SSTs, i.e., the SSTs may be functions of previous values of the SSTs, or of the present or past wind directions or speeds. We also aply it to handling the modeling of the cyclic nonstationarity found in the SSTs.

4.1   Modeling SSTs using Wind Speeds and Directions

We use the categorical variable implementation of MARS to allow lagged wind directions (WD_t) as predictors of logged SST (X_t) at Granite Canyon over the five year period 1/1/86 to 12/31/90, using 10 lags of wind direction, 10 lags of the logarithm of ($1 +$ wind speed), namely WS_t', and 50 lags of logged SSTs. The SPAN parameter in the MARS model is set equal to 25. The resulting CASTAR model, which follows, makes explicit the relationship seen in Figure 4 between the wind direction and the SSTs.

<div align="center">

Model for 5 Years of logged SSTs using

using Wind Direction and Log of (1 + Wind Speed)

</div>

$$\widehat{X}_t = \begin{cases} 2.192 & +0.878(X_{t-1} - 2.13)_+ \\ & +1.616(2.22 - X_{t-34})_+ \\ \\ & +0.013(WS_{t-1}' - 1.10)_+ I(WD_{t-1} \in \{1,2\}) \\ & -0.035(WS_{t-1}' - 1.10)_+ I(WD_{t-1} \in \{2,3\}) \\ \\ & -.499(X_{t-1} - 2.13)_+(2.75 - X_{t-7})_+(2.68 - X_{t-17})_+ \\ \\ & -0.584(2.27 - X_{t-34})_+(WS_{t-1}' - 1.10)_+ I(WD_{t-1} \in \{2,3\}) \\ & -0.517(X_{t-49} - 2.510)_+(WS_{t-1}' - 3.00)_+ I(WD_{t-1} \in \{2,3\}) \\ & +4.665(2.51 - X_{t-49})_+(2.26 - X_{t-24})_+ I(WD_{t-1} \in \{2,3\}) \end{cases} \tag{9}$$

From the model we see that:

1. The logged SSTs, X_t, range in value from 2.13 to 2.87 over this five year period and the first, second, and fifth lines in Equation 9 give terms in the model which are purely functions of

lagged SSTs. The first term — a lag one autoregressive term — is always present since the threshold of 2.13 is the minimum value of the SSTsseries. The first order term on the second line of the equation involves lag 34 and only adds to $\widehat{X}_t$ when X_{34} is less than 2.22, which is rare, given that the minimum value of the SSTss is 2.13. The three-way interaction in line five occurs infrequently. However the lags 1, 7 and 17 are familiar from models derived earlier in the paper.

2. The change in $\widehat{X}_t$ when the wind blows from the Northwest that is explicit in lines 3 and 4 of the equation for the model agrees with the relationship seen in Figure 4. Splitting the two terms into three by separating out directions 1,2 and 3, we see that direction 1 cause a slight increase in the overall level (+0.013), a decrease (0.013 - 0.035= - 0.022) for wind speed from direction 2, and a decrease (-0.035) for direction 3.

3. Another interesting feature model for the log of the wind speed is shown in the last two terms, both of which involve the term X_{t-49}. Only one of these terms is present at a time, depending on whether X_{t-49} is greater than or less than 2.51 and whether the wind direction is in category 2 or 3. When X_{t-49} is larger than 2.51 and the windspeed is greater than 3.00 knots, we have a coupling between $\widehat{X}_t$ and X_{t-49}. This is the oscillation which, in Section 1, we noted was observed by Breaker and Lewis (1988) for several sets of SSTs, including the one at Granite Canyon. Attempts to elucidate that oscillation using linear methods such as complex demodulation were fruitless. However the last two terms of the model show that the "oscillation" is present only when the wind direction is from the Northwest ($WD_{t-1} \in \{2,3\}$), and increases as lag one wind speed increases. This relationship between wind speeds 49 days apart is very nonlinear.

4. Another feature of this CASTAR model is that the log(1 +wind speed) thresholds which, we emphasize, are selected automatically by the MARS algorithm, have a clear meteorological interpretation. A transformed wind speed threshold of 1.10 knots translates into 1.031 m/sec, below which it is well known that wind speeds have little effect on SSTs. In the CASTAR model, the terms on lines 3, 4, and 6 of the equation only enter the model if wind speeds are above this threshold. Looking at term 7, a transformed wind speed greater than 3.00 knots, which translates into 9.82 m/sec, acts to lower SSTs when the wind is blowing from the NW and the logged temperature 49 days ago was grater than 2.510.

Note that in deriving Equation 9 we did not subtract a time of year cycle from either the transformed wind speeds or the transformed SSTs. This is because, as remarked before, subtracting the cycle

from the data destroys the relationship of the thresholds to the physical measurements. This may be better to do if one were making forecasts from the model. But the model as it stands is superb as a descriptive tool. In addition, since the lag one correlation in the data is so high, little will be gained for accuracy of prediction by going to the semimultivariate model.

4.2 Modeling Cyclic Nonstationary in the SSTs using Seasonal Categorical Variables

The residual plots in Section 3 (Figure 5 and Figure 8) indicated that the ARFIMA and ASTAR models do not provide a complete fit to the SST data in that the square of the residuals is not an *uncorrelated* sequence. We postulated that this result could be explained by the fact that the one year cycle manifested itself in the SSTs not only in the first-order characteristics of the time series, *i.e.* the mean, but also in the whole probabilistic structure of the process. In particular, Figure 9 showed a yearly variation in the lag one serial correlation of the process.

Using a categorical covariate process which tags the SSTs with time-of-year information, we can extend the ASTAR and SMASTAR modeling to encompass this type of cyclic non-stationarity. Specifically, let $C(t)$ have value 1,2,3,4,5, or 6 depending on whether t corresponds to the first two months of the year, the second two months of the year, *etc.*. Since the SST series begins on March 1, 1971, the covariate sequence in the input to MARS starts with 61 2's for March/April, continues with 61 3's for May/June, *etc.*. The coding for six contiguous time periods in the year is quite arbitrary, and is chosen in hopes of capturing the main nonstationary effects. The MARS algorithm is applied allowing 365 lags for this covariate time series. As before, the main predictor series is the logged SST data with first yearly harmonic removed and lags up to 1925 allowed in order to model long-range dependent behavior. The SPAN parameter was taken to be 5 to avoid instability, and the maximum number of interactions allowed was MI=3. The resulting CASTAR model follows:

Model for 20 Years of log SST data

with year harmonic removed

using a Time-of-year categorical variable

23

Best MARS model;   categorical season

Autocorrelation function for residuals

Autocorrelation function for squared residuals
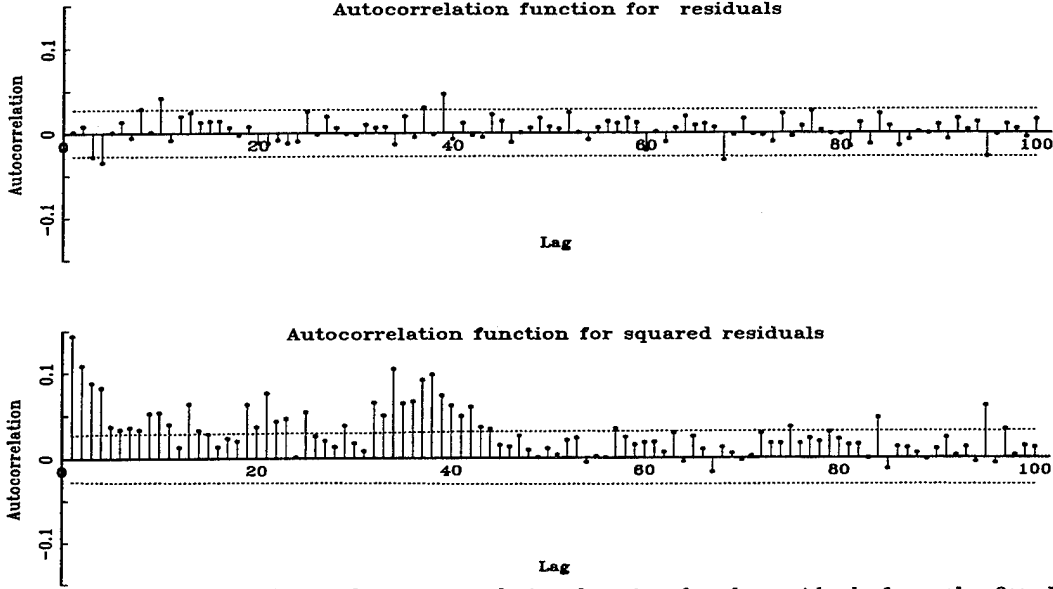
Figure 10: The top panel shows the autocorrelation function for the residuals from the fitted threshold MARS model with SPAN 5 and a categorical time-of-year series covariate, to the Granite Canyon SSTs. The bottom panel is the autocorrelation function for the squared residuals.

$$\widehat{X}_t = \begin{cases} -.0731 & -0.859(-0.116 - X_{t-1})_+ \\ & +1.006(X_{t-1} + 0.116)_+ \\ & -0.107(X_{t-2} + 0.340)_+ \\ \\ & -0.221(X_{t-2} + 0.340)_+(0.045 - X_{t-17})_+ \\ \\ & -.576(X_{t-1} + 0.116)_+(X_{t-650} + 0.013)_+(C_{t-234} \in \{2,6\}) \\ & -8.197(X_{t-1} + 0.116)_+(X_{t-1250} - 0.199)_+(C_{t-234} \in \{2,6\}) \\ & -117.251(X_{t-2} + 0.340)_+(-0.172 - X_{t-8})_+(X_{t-605} - 0.128)_+ \\ & -40.193(X_{t-2} + 0.340)_+(X_{t-17} - 0.045)_+(-0.222 - X_{t-1425})_+ \end{cases} \tag{10}$$

One sees that different equations for X_t are obtained depending on whether $t - 234$, the lag of the categorical covariate sequence C_t as it appears in lines 5 and 6 of the equation, is in the March/April or November/December time periods, or is in the remainder of the year. This is equivalent to t being between March 12 and May 11 or between July 9 and September 7. In this case, we see from line 5 of the equation that $\widehat{X}_t$ is decreased by an additional 0.576 if X_{t-1} is greater than -0.116 and X_{t-650} is greater than -0.013. Also, from line 6 of the equation, $\widehat{X}_t$ is decreased by an additional 8.197 if X_{t-1} is greater than -0.116 and X_{t-1250} is greater than 0.199. Thus there is selective time-of-year dependence in the model for the data.

In Figure 6 the forecast characteristics of this model are given by the solid curve lying below all

24

the other curves for k in the range 100 to 300. Thus there is some gain, using the model given in Equation 10, in predictive capability in this range, although Figure 7 shows that none of the models outpredicts the others for k less than about 60. The squared residuals for this $\widehat{X_t}$ process generated with a categorical seasonal covariate process are depicted in Figure 10, and show slightly less nonzero correlation at low lags than do the squared residuals whose autocorrelation function is shown in Figure 8. They are, however, undoubtedly a long way from being an uncorrelated sequence. A further refinement which might improve this is discussed in the next section.

5   Discussion and Directions for Future Research

We have demonstrated how the MARS algorithm can be used to obtain ASTAR, SMSTAR, and CASTAR models for a set of SSTs having complex nonlinear and cyclic effects. The ASTAR models are able to model the nonlinear effects in the data and approximate long-range dependence characteristics. Although none of the skeletons produced from the ASTAR models for SSTs indicate a limit cycle, as was found for the sunspot data (Lewis and Stevens, 1991), the number of cycles present (20) may be too small to induce this behavior. We believe that a model containing this kind of internal cycle would be better than one in which a one-year harmonic must be subtracted from the data for adequate fitting. However, using this artifice, the ASTAR models perform better at predicting at long forecast horizons than the linear AR or ARFIMA models.

In terms of descriptive aspects, the SMASTAR model including wind direction is clearly interpretable from an oceanographic standpoint. However, the SMASTAR models suffer from a difficulty in that predictions for more than one step ahead are difficult. For example in Equation (9), if one wanted to predict the sea surface temperature two steps ahead, one needs to have a prediction for the one step ahead wind speed. One answer (Lewis and Ray, 1993) is to derive a semi-multivariate model for the wind velocities, as well as for the SSTs, using MARS. In order to have a *complete* bivariate model, however, one needs to postulate or empirically derive a simple model for the two resultant error sequences. Alternatively, an extension of MARS to the multivariate case, analogous to multivariate linear regression modeling, could be entertained for nonlinear multivariate time series modeling. This is an area for future research.

Finally in Section 4.2 we showed that a CASTAR model can be used to take into account yearly nonstationarity in the SSTs data. Again we note that the choice of the six-valued categorical covariate term in the model given at Equation 10 is quite arbitrary – we could just as easily have used a twelve-valued covariate to capture effects which are changing in the order of a month. What we are trying to do here is to come up, in the simplest case, with a periodically stationary model

$$X_t = \begin{cases} X_t^1 & t(mod365) \le t_c \\ X_t^2 & t(mod365) > t_c \end{cases} \qquad (11)$$

where t_c is a (time) threshold. Such a two level model might be plausible in the case of the SSTs, given the abrupt change in SSTs temperature each year with the wind-driven spring temperature transition.

In the spirit of MARS, one might want to estimate the transition point from the data itself, and more generally explore different degrees of differentiation (numbers of time thresholds) via the data. One way to do this is to have C_t take on 365 values, one for each day, and use only the lag one realization of this covariate process in the MARS run. However, this is not doable in the present MARS algorithm because the maximum number of values a categorical variable may have is 28.

Additional research into the properties of such models and the estimation and identification of parameters is needed. Of course this model is just a generalization of the periodically staionary models of Gladyev (1961); see McLeod (1994) for details. In fact preliminary runs on *monthly* data such as that given in McLeod (1994) with MARS are very promising.

Appendix: Choice of Parameters in the MARS Algorithm

When choosing the form of the input of *ordinal* predictors, the choice should generally be no restrictions. This is because it is extremely difficult to outguess the very non-linear forms which can be generated by the MARS algorithm. The price of course, for the generality, is computing time, since the MARS algorithm uses an exhaustive search and is thus computer intensive. For example, the fits to the SSTs in Lewis and Ray (1993) used lagged SSTs of lags up to 365, but they found that this was not enough to capture the long-range dependence or the El Nino effect. Thus in this paper we used much longer lags. However, computing times increase approximately tenfold.

There are a number of other parameters which affect the performance of MARS. One is the SPAN, which is a smoothing parameter. The default value of 1 will generally result in "overfitting" the fine details of the process and will often give unstable results. Using values near 50 may result in too much smoothing, however. For example, the original work on modeling sunspot data using with MARS (Lewis and Stevens, 1991) used SPAN=50. Setting SPAN=5 produces limit cycle models having root-mean-squared prediction error approximately 50% smaller than that obtained using the SPAN=50 model. *An initial value of 5 is therefore recommended.*

It is seldom wise to choose MI, the maximum interaction level, to be greater than 3, *i.e.,* up to 3-variable interactions allowed, as one then gets unstable and uninterpretable models, as well as ex-

cessive computing times. (MI=1 means additive modeling, or equivalently, main effects only). And as remarked above, the Schwarz-Rissanen criterion has been found to produce more interpretable and more parsimonious models than the GCV for time series. A stark example of this is given for the Granite Canyon SSTs in Lewis and Ray (1993). There may be further refinements possible with other goodness of fit criteria.

The modified MARS3.6 FORTRAN program may be obtained from the authors. Note that the MARS3.6 program is a subroutine called from a user supplied driver program; we distribute the MARSDRV FORTRAN program for this. Its input is a regression matrix derived from the input time series by subroutine MARSBLD. Thus if one wants to regress a time series on 20 lagged variables, the matrix has columns which are different, but highly similar pieces of the time series. This is clearly wasteful, particularly since the size of the data matrix is the biggest limitation on the use of MARS for large time series. We hope to rewrite MARS to obviate this problem, but it is a large undertaking.

The runs given in this paper were run on an Amdahl 5995 Model 700 mainframe running under the MVS/ESA and VM/XA operating systems. Typical runs took four hours of CPU time and 500 megabytes of memory and 250 megabytes of temporary storage.

The SST data and related series may be obtained from the authors upon request.

6   References

Agiakloglou, C., Newbold, P., and Wohar, M. (1993), "Bias in an estimator of the fractional difference parameter," *Journal of Time Series Analysis*, 14, 235–246.

Breaker, L. C. and Lewis, P. A. W. (1988), "A 40-50 day oscillation in sea-surface temperature along the Central California coast," *Estuarine, Coastal and Shelf Science*, 26, 395–408.

Breaker, L. C., Lewis, P. A. W., and Orav, E. J. (1983), "Analysis of a 12-Year Record of Sea-surface Temperatures off Pt. Sur, California," Report NPS55-83-018, Department of Operations Research, Naval Postgraduate School.

Casdagli, M. (1992), "Nonlinear forecasting, chaos, and statistics," in *Modeling Complex Phenomena*, ed. L. Lam and V. Naroditsky, New York: Springer-Verlag.

Friedman, J. H. (1991a), "Estimating Functions of Mixed Ordinal and Categorical Variables using Adaptive Splines," Technical Report 108, Department of Statistics, Stanford University.

Friedman, J. H. (1991b), "Multivariate adaptive regression splines," *Annals of Statistics*, 19, 1–142.

Friedman, J. H. (1993), "Fast MARS," Technical Report 110, Department of Statistics, Stanford University.

Geweke, J. and Porter-Hudak, S. (1983), "The estimation and application of long memory time

series models," *Journal of Time Series Analysis*, 4, 221–238.

Gladyev,, E. G. (1961), "Periodically correlated random sequences," *Sov. Math. Dokl.*, 2, 385–388.

Granger, C. W. and Anderson, A. P. (1978), "An introduction to bilinear time series models," *Vandenhoeck and Ruprecht*, 1–94.

Granger, C. W. J. and Joyeux, R. (1980), "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis*, 1, 15–29.

Haslett, J. and Raftery, A. E. (1989), "Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource," *Appl. Statist.*, 38, 1–50.

Hidaka, K. (1954), "A contribution to the theory of upwelling and coastal currents", *Transactions of the American Geophysical Union*, 35, 431–444.

Hosking. J. R. M. (1981), "Fractional differencing," *Biometrika*, 68, 165–176.

Hosking, J. R. M. (1984), "Modeling persistence in hydrological time series using fractional differencing," *Water Resources Research*, 20, 1898–1908.

Jonas, A. B. (1983), *Persistent Memory Random Processes*. Ph.D. thesis, Harvard University.

Lewis, P. A. W., Ray, B. K., and Stevens, J. (1993), "Modelling Time Series using Multivariate Adaptive Regression Splines (MARS)," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, ed. A. Weigend and N. Gershenfeld, Santa Fe Institute:Addison-Wesley.

Lewis, P. A. W. and Ray, B. K. (1993), "Nonlinear Modelling of Multivariate and Categorical Time Series using Multivariate Adaptive Regression Splines," in *Dimension Estimation and Models*, ed. H. Tong, Singapore:World Scientific.

Lewis, P. A. W. and Stevens, J. G. (1991), "Nonlinear modeling of time series using multivariate adaptive regression splines (MARS)," *Journal of the American Statistical Association*, 87, 864–877.

Lewis, P. A. W. and Stevens, J. G. (1992), "Semi-multivariate modeling of time series using adaptive spline threshold autoregression (ASTAR)," preprint.

Mandelbrot, B. B. and Van Ness, J. W. (1968), "Fractional Brownian motions, fractional noises, and applications," *SIAM Review*, 10, 422–437.

McLeod, A.I. (1994), "Diagnostic Checking of Periodic Autoregression Models with Application," *Journal of Time Series Analysis*, 15, 221–233.

McLeod, A. I. and Hipel, K. W. (1978), "Preservation of the rescaled adjusted range. 1. A reassessment of the Hurst phenomenon," *Water Resources Research*, 14, 491–508.

Mendelson, R. (1980), "Relating Fisheries to the environment in the Gulf of Guinea: information, causality and long-term memory," in *Les Pecheries Ouest-Africaines: Variabilite, instabilite et changement,* ed. P. Cury and C. L. Roy, OSTROM, Paris, 446–465.

Petrucelli, J.D. and Woolford, S.W. "A threshold AR(1) model. Journal Applied Probability," 21, 207–286.

Ray, B. K. (1993), "Modeling long memory processes for optimal long-range prediction," *Journal of Time Series Analysis*, 14, 511–526.

Rosenblatt, M. (1987), "Scale renormalization and random solutions of the Burgers equation," *Journal of Applied Probability*, 24, 328–338.

Robinson, P. M. (1991), "Time series with strong dependence," preprint.

Royer, T. C. (1991), "High-Latitude Oceanic Variability Associated with the 18.6-Year Nodal Tide," *Journal of Geophysical Research*, 98, 4639–4644.

Stevens, J. (1991), *An Investigation of Multivariate Adaptive Regression Splines for Modeling and Analysis of Univariate and Semi-multivariate Time Series Systems*, Ph.D. thesis, Naval Postgraduate School, Monterey, CA.

Taqqu, M. (1987), "Random processes with long-range dependence and high variability," *Journal of Geophysical Research*, 92, 9683–9686.

Tong, H., Thanoon, B. and Gudmundsson, G. (1985), " Threshold time series modeling of two Icelandic riverflow systems," *Water Resources Bulletin,* 21, 651–660.

Tong, H. (1990), *Nonlinear Time Series*, New York:Oxford University Press.

Tong, H. (1983), *Threshold Models in Non-linear Time Series Analysis*, New York:Springer-Verlag.

Yajima, Y. (1988) "On estimation of a regression model with long-memory stationary errors," *Annals of Statistics*, 16, 791-807.

DISTRIBUTION LIST