Mission Research Corporation

Copy No. 4

MRC-R-1488
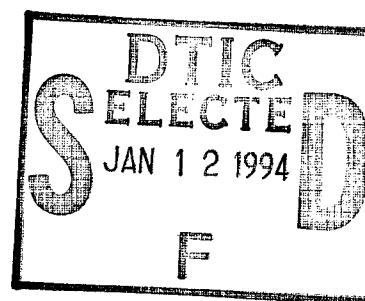
Technical Report

STOCHASTIC MODELING AND GLOBAL WARMING TREND EXTRACTION FOR OCEAN ACOUSTIC TRAVEL TIMES

Steven Bottone
Henry L. Gray
Wayne A. Woodward

DTIC
ELECTE
JAN 1 2 1994
F

January 1995

19950110 003

Copy No. _____ MRC-R-1488

<u>Technical Report</u>

STOCHASTIC MODELING AND GLOBAL WARMING TREND EXTRACTION FOR OCEAN ACOUSTIC TRAVEL TIMES

Steven Bottone
Henry L. Gray
Wayne A. Woodward

January 1995

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 the 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 6 Jan 1995	3. REPORT TYPE AND DATES COVERED Interim Technical Report

4. TITLE AND SUBTITLE
Stochastic Modeling And Global Warming Trend Extraction For Ocean Acoustic Travel Times

5. FUNDING NUMBERS
Contract MDA972-93-C-0021

6. AUTHOR(s)
Steven Bottone
Henry L. Gray[*]   Wayne A. Woodward[*]

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Mission Research Corporation
735 State Street, P. O. Drawer 719
Santa Barbara, CA 93102-0719

8. PERFORMING ORGANIZATION REPORT NUMBER
MRC-R-1488

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
Attn: Dr. Ralph Alewine/NMRO

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

[*] Southern Methodist University, Dallas, TX

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public distribution;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A possible indication of the existence of global climate warming is a negative trend for the travel time of an acoustic pulse along a fixed long path, or paths, in the ocean over a period of many years. The goal of this report is the development of methods specifically for determining the presence of a long term trend for climate change from a temporal sequence of measurements of acoustic propagation times. Robust statistical methods for determining whether a significant trend is present in a given set of time series data have been developed and, for illustration, applied to some specific traveltime time series generated by the MASIG and GFDL ocean models. In this report we consider line + noise and ARIMA statistical models. We show that if the time series are long enough, somewhat over 20 years, then series such as those simulated by the MASIG and GFDL models can be classified reliably as line + noise when this is the case. However, it is shown that the results are considerably different for the two ocean models under consideration and that these models can not currently be relied upon by themselves to predict global warming. Experimental data is most certainly needed, not only to measure global warming itself, but to help improve the ocean model themselves.

14. SUBJECT TERMS
Global Warming   Time Series Analysis
Ocean Acoustics   Statistical Models
Trend Extraction   Climate Change

15. NUMBER OF PAGES

16. PRICE CODE

17. Security CLASSIFICATION OF REPORT UNCLASSIFIED	18. Security CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. Security CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

SECURITY CLASSIFICATION OF THIS PAGE

CLASSIFIED BY:

DECLASSIFY ON:

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE   **PAGE**

LIST OF TABLES

0. EXECUTIVE SUMMARY

A possible indication of the existence of global climate warming is a negative trend for the travel time of an acoustic pulse along a fixed long path, or paths, in the ocean over a period of many years. A warmer ocean implies, on average, an increased speed of sound which in turn implies a decrease in this travel time. The use of long acoustic paths substantially reduces the variability of temperature from local ocean weather, potentially allowing the detection of underlying climate trends. If taken over a long enough period of time, this data may provide an indication that global warming is occurring, if, in fact, it is. The Acoustic Monitoring of Global Ocean Climate (AMGOC) experiment will measure acoustic travel times over a number ocean paths with the coupled goals of improvement of ocean models and detection of any existing warming signature.

The goal of this report is the development of methods specifically for determining the presence of a long term trend for climate change from a temporal sequence of measurements of acoustic propagation times. This traveltime time series from the AMGOC experiment are expected to contain substantial weather scale variability which can have both deterministic and stochastic features. These fluctuations will make it difficult to detect the underlying smaller climate changes. Current ocean and coupled ocean/atmosphere models cannot reproduce the details of ocean variability so it is not now possible to remove such details from data in a deterministic manner. However, with additional understanding of ocean properties it becomes possible to substantially improve predictions of weather scale variability, at least in a statistical sense. Consequently, improved ocean models resulting from assimilation of the AMGOC data will provide the future basis for improved extraction of climate trends from the data.

For current consideration, we present methods for extraction of possible warming signatures from the data that can be carried out without the explicit use of ocean models. In this report we develop techniques for use of statistical models for warming trend extraction which are based exclusively on the traveltime data. Even though such models are not precisely correct and only implicitly address the physics of the problem, it is often possible to capture many of the statistical properties of the time series by properly modeling the correlation structure of the process.

Robust statistical methods for determining whether a significant trend is present in a given set of time series data have been developed and, for illustration, applied to some specific traveltime time series generated by the MASIG and GFDL ocean models, both with and

1

without warming. A popular approach to the problem has been to fit a line through the data and use statistical methodology to determine if this line has significant slope. However, by assuming such a model one has automatically assumed that if there is a trend in the past then there will be a similar trend in the future. Note that, at least implicitly, a statistical model must be assumed in order to draw a conclusion about the trend implications of the data. Since this assumption may not be correct we have considered alternative models which allow the trend component to be of a more general nature. In this report we limit ourselves to two types of models, the *line + noise* model, just described, and the so-called ARIMA model. This allows us a contrast between models since the ARIMA model allows for "temporary" trends and only predicts the previously observed trend to continue if the estimated correlation of the future to the past is sufficiently strong. Since the two competing models lead to opposite inferences it is necessary to attempt to use the data to determine which of the models appears most likely.

Once the candidate statistical models are fit to the data, each model can be used to determine how often a time series similar to that given by the data, but of any length, will contain a significant trend. ARIMA models can give realizations which often have significant "random" trends, which would not be predicted to continue, but if the time series were not long enough it would be difficult to distinguish this from a "true" trend. It is therefore necessary to determine which model best describes a given time series. We describe methods of how, using the data alone, the data can be classified as line + noise or ARIMA. The reliability of these methods is contained in classification tables, which estimate the probability of classifying correctly and misclassifying a time series which is actually line + noise or ARIMA. We show that if the time series are long enough, somewhat over 20 years, then series such as those simulated by the MASIG and GFDL models can be classified reliably as line + noise when this is the case. However, it is shown that the results are considerably different for the two ocean models under consideration and that these models can not currently be relied upon by themselves to predict global warming. Experimental data is most certainly needed, not only to measure global warming itself, but to help improve the ocean model themselves.

1. INTRODUCTION

The release of carbon dioxide and other greenhouse gases into the atmosphere has been associated with changes in temperature of the earth's climate system, which includes the atmosphere and hydrosphere, whose most important component for climate change is the oceans. An increase in the average temperature of the climate system over time scales of many tens of years (whether due to increasing amounts of greenhouse gases or not) will be referred to as global warming. The oceans are a vast reservoir of heat and carbon, and as such, directly effect global climate changes. A direct measure of temperature changes in the ocean is therefore necessary to understand and predict global climate change and warming.

Acoustic thermometry of the world's oceans offers a method of measuring large-scale temperature changes in the ocean [Munk and Forbes, 1989; Spiesberger and Metzger, 1991]. Acoustic tomography was originally introduced by Munk and Wunsch [1979]. The idea behind acoustic thermometry is relatively simple. The speed of sound in the ocean increases about 4.6 meters per second for a one degree Celsius increase in ocean temperature. The travel time of an acoustic pulse transmitted from a source to a receiver in the ocean will therefore decrease in a warmer ocean. By transmitting acoustic signals over long distances (thousands of kilometers) an integrated measure of the temperature along that path will be obtained, with the benefit that local variations in the temperature due to, for example, mesoscale eddies, will automatically be averaged out.

It seems reasonable to conjecture that if global warming is occurring the temperature along any given acoustic path in the ocean will be increasing with time, which would imply a decreasing travel time. Early studies from various global climate models suggest that this may not be the case [Mikolajewicz et al., 1990b; Manabe et al., 1991]. Results from these models show that if increasing amounts of carbon dioxide are added to the global climate system, then the temperature at most points in the ocean will increase with time, however, there will be some ocean regions, such as the Antarctic and the northern Atlantic, where the temperature will actually decrease with time (remember these are model results). The ATOC (Acoustic Thermometry of Ocean Climate) experiment currently plans to have two sources, one near Hawaii and one near Point Sur, and various receivers giving a total of 14 paths, most in the northeast Pacific. In addition, the GAMOT (Global Acoustic Mapping of Ocean Temperature) experiment will include a number of surface suspended ocean receivers (SSAR). Model results indicate that warming, and, hence, decreasing travel times, are expected on these paths. As a first step in looking for a warming trend, it is

reasonable to consider an individual path and test the traveltime time series for that path for a trend that decreases with time, which is the approach that will be taken in this report. The above comments concerning predicted cooling on some paths should always be kept in mind. Future work will address extraction of warming trends from a simultaneous analysis of data on many paths.

This report will concern itself with methods to extract warming trends from time series consisting of travel time data from single ocean acoustic paths. Due to the current lack of experimental data, we will demonstrate our techniques on output from various climate and ocean models. This data will be discussed in the next section. Following a description of some theoretical aspects of trend extraction in time series analysis relevant to our purposes, we will apply these techniques to thoroughly investigate two specific time series generated by the MASIG (Mesoscale Air-Sea Interaction Group) ocean model and the GFDL (Geophysical Fluid Dynamics Lab) general circulation model. In the analysis we will show not only how to determine whether there is a significant trend in a given set of data, but how to choose which statistical model, from a select class, best describes the data. Different statistical models may give different forecasts as to whether a trend will be predicted to continue, so it is important to determine which model best describes the data.

4

2. DATA

Experimental data will be in the form of traveltime time series. The travel time is the time it takes for an acoustic pulse to travel from the acoustic source to the receiver (actually, there is a travel time between source and receiver for each acoustic mode, but we will ignore this aspect here). The ATOC experiment, for example, currently plans to have two sources, one near Hawaii and one near Point Sur, and various receivers giving a total of 14 paths, most in the northeast Pacific. Due to the current lack of experimental data—the experiment has been delayed for a number of reasons—it is necessary to demonstrate our methods of trend extraction on similar time series generated from ocean models. Since the main purpose of this report is describe our development of these methods, it is not necessary to use experimental data to demonstrate the usefulness of these techniques. Although the results we will obtain in this report are for simulated data only, they are of interest in their own right. Ocean models are in and of themselves an integral part of the ocean acoustics global warming experiment. It is only through these models, without excessive data, that a connection can be made between acoustic travel times, which is the form in which the data is obtained, and global ocean temperature, which is the quantity most closely related to global warming. It is therefore of great importance that the fidelity of the various climate and ocean models is continually checked and improved, and our methods and results will help to do this.

Ocean and climate models do not compute acoustic travel times directly. Usually the models compute various ocean parameters, such as temperature, salinity, and current velocity, as functions of grid point (latitude, longitude, and depth) and time. A traveltime time series along a given path can then be computed by first using an equation of state to calculate a sound speed from the temperature and salinity (the pressure is a function of depth) and then using an acoustic propagation code to compute the travel time of a pulse from one end of the path to the other through the given sound speed profile. If there were a close correspondence between an experimental traveltime time series and the corresponding time series produced by the ocean model, then one would have confidence in using the model to predict the future behavior of the ocean, in particular, the model could be used to determine the rate of global warming in the future. In actuality the correspondence will not be good enough to make such claims with much certainty and the experimental data will at first be used to aid in improving the models.

5

We will carry out a thorough trend extraction analysis on simulated data obtained from two models: 1) the MASIG model, and 2) the GFDL model. The MASIG (Mesoscale Air-Sea Interaction Group) model is a reduced gravity ocean model driven by COADS (Comprehensive Ocean-Atmosphere Data Set) winds, coupled to an equatorial model at its southern boundary [Pares-Sierra and O'Brien, 1989]. A twenty year simulation (for the years 1970–1990) of the northeast Pacific ocean was run and changes in acoustic travel times over 10 paths from Hawaii to various locations on the western shore of North America were computed. These 10 time series were provided by Jim O'Brien of Florida State University.

The GFDL (Geophysical Fluid Dynamics Lab) general circulation model is a coupled atmospheric-ocean-land surface model with global geography [Manabe et al., 1991]. This model is one of the current best predictors for global warming due to increased greenhouse gases, so the trends it predicts for acoustic travel times are the "state-of-the-art best guesses". We purchased from the NCDC (National Climatic Data Center) two tapes which contained output from the GFDL model for a set of simulations run in 1989. The data included temperature and salinity computed at points on an ocean grid at 12 depths for one hundred years. Output was stored every month, so all time series consisted of 1200 points. There were two separate runs: one was a control run in which constant levels of CO_2 were input to the system and the other run had increasing CO_2 levels of 1% per year. This warming run gives increasing ocean temperatures with time. Traveltime time series were obtained by integrating sound speed, obtained from temperature, salinity, and pressure (depth) using an appropriate equation of state, along the sound channel axis (sound speed minimum) along the path from source to receiver location.

In addition to the simulated data for which we have carried out extensive trend analyses we have also obtained further simulated data from two other models: 3) the Semtner-Chervin model, and 4) the Hamburg model. Although we do not include in this report time series analyses for these data sets, we provide here a description of the output for reference. Output from the Semtner-Chervin ocean general circulation model (OGCM) [Semtner and Chervin, 1988], which was sent to us by Dimitris Menemenlis at MIT, is in the form of time series of average temperature along 13 paths at 20 separate ocean depths over a time period of about 1350 days (3-4 years) given at 3 day intervals. Equivalent traveltime time series were approximated from these temperature time series by using the relation [see Spiesberger et al., 1992; Spiesberger et al., 1989]

$$\frac{\Delta T}{\overline{T}} \approx - k \, \Delta\Theta \, ,$$

where T is the equivalent traveltime, $\overline{T}$ is the time-averaged travel time for a given path, $\Delta T = T - \overline{T}$, $\Delta\Theta$ is the temperature change, and $k = 3.19 \times 10^{-3}$ / degree Celsius.

Output from a simulation using the Hamburg OGCM with forcing climatological boundary conditions by stochastic freshwater fluxes, was provided by Uwe Mikolajewicz [Mikolajewicz et al., 1990a]. The simulation covered a period of 3800 years and output (such as temperature and salinity as a function of position) was obtained at two year intervals. Acoustic travel times along 106 paths were computed by integrating the ocean sound speed along the sound channel axis. In this manner 106 traveltime time series covering 3800 years at 2 year intervals were obtained. In addition, a second set of time series along the same 106 paths was obtained by running the model with a greenhouse warming boundary condition over a 100 year period, the output being given at 2 year intervals. This latter case had no stochastic freshwater flux.

Before turning to an analysis of specific traveltime time series simulated by the MASIG and GFDL models, a description of the time series analysis techniques which will be used in the analysis is provided in the next section.

3. THEORY

In the remainder of this report we will be concerned with time series giving the travel time of an acoustic pulse along a single path from a source deep in the ocean to a distant receiver. If it is assumed that a time series is a stochastic process (this can also mean that the series has a deterministic component and a stochastic, noise component) then to make statements such as "there is a significant trend in this time series", requires some assumptions about the stochastic nature of the process. A statistical model must be chosen to describe the data, and only in the context of the model can questions of a probabilistic nature be posed and answered. Since one can never know what model (or physical process) generates the data, then one must attempt to choose the best approximate model from whatever list of choices is available and tractable. Even though such a model is not precisely correct and only implicitly addresses the physics of the problem, it is often possible to capture many of the statistical properties of the time series by properly modeling the correlation structure of the process. The resulting statistical model can then be used to economically generate realizations which are representative of the data one would expect to obtain from an experiment or appropriate OGCM. That is, realizations from the statistical model can be generated easily for statistical analysis, while generating similar data from the actual circulation model would be prohibitively time consuming. If the data comes from an actual experiment it may be impossible to retake the data in a statistically meaningful way.

3.1. Time Series Models

Let the time series for traveltime anomaly be given by the discrete stochastic process $\{X_t; t = 0, \pm 1, \ldots\}$, where X_t is a random variable giving the travel time as a function of time, t. The time unit is usually taken to be the time interval between successive readings. For the GFDL model, for example, the time unit is one month, so X_1 would be a random variable describing the travel time for month 1 ($t = 1$), X_2 the travel time for month 2, and so on. A realization of length n of the time series, X_t, is a set of real-valued outcomes which will be denoted $\{x_t; t = 1, \ldots, n\}$. Loosely speaking, a set of traveltime data, x_t, will be considered a realization from a traveltime time series, X_t.

Two different statistical models will be used to analyze the traveltime time series, line + correlated noise and ARIMA. The line + noise model is given by

$$X_t = a + bt + E_t, \tag{1}$$

8

where E_t is an ARMA (autoregressive-moving average) process [for a detailed account of ARMA processes see Box and Jenkins, 1976 or Gray et al., 1994]. An ARMA(p,q) process, $\{E_t\}$, is defined by

$$\phi(B)E_t = \theta(B)a_t, \tag{2}$$

where B is the backward shift operator given by $B^n X_t = X_{t-n}$,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \tag{3}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^q, \tag{4}$$

and a_t is discrete white noise with zero mean and variance σ_a^2, i.e., $E[a_t] = 0$ and $E\left[a_t^2\right] = \sigma_a^2$. The ARIMA($p,d,q$) (autoregressive integrated moving average) model, $\{X_t\}$, is defined by

$$\phi(B)(1-B)^d(X_t - \mu) = \theta(B)a_t, \tag{5}$$

where the parameter d is an integer, usually 1 or 2, μ is the mean of the process, i.e., $E[X_t] = \mu$, and the remaining parameters defined as above.

The characteristic equation of a general ARMA process (the ARIMA process defined by equation (5) is a special case) with autoregressive operator $\phi(B)$ is defined by $\phi(r) = 0$, where r is a complex number and

$$\phi(r) = 1 - \phi_1 r - \phi_2 r^2 - \cdots - \phi_p r^p. \tag{6}$$

An ARMA process is stationary (stable) if and only if the roots of the characteristic equation lie outside the unit circle. It turns out that models which have found much application in the physical sciences have roots that lie on (or outside) the unit circle. As can be seen in equation (5), ARIMA models have d roots of the characteristic equation on the unit circle. The autoregressive operator $\phi(B)$ can be factored into irreducible first and second-order factors. The roots associated with the irreducible second-order factor, $1 - \alpha_1 B - \alpha_2 B^2$, are complex conjugates whose absolute reciprocal is $\sqrt{-\alpha_2}$. The system frequency associated with this factor is

$$f = \frac{1}{2\pi} \cos^{-1}\left(\frac{\alpha_1}{2\sqrt{-\alpha_2}}\right), \tag{7}$$

9

where frequency is in cycles per unit time. For a first order factor, $1 - \alpha_1 B$, the absolute value of the reciprocal of the associated root is $|\alpha_1|$ while the associated frequency is $f = 0$ if $\alpha_1 > 0$ and $f = 1/2$ if $\alpha_1 < 0$. Factors associated with roots near the unit circle dominate the behavior of an ARMA model. Factor tables, as defined by Gray and Woodward [1986], are quite useful in understanding the behavior of an ARMA model.

The difference between the two statistical models, (1) and (5), may best be understood by considering the "best" forecast function. For the line + noise model, (1), the best eventual forecast is for the line to continue. For the ARIMA model, (5), if $d = 1$, the best eventual forecast is a constant, approximately equal to the last observation. However, because there is a root on the unit circle, there will be long random trends [see Gray et al., 1994], which may be significant, and for time series of moderate length it may be difficult to determine which model best fits the data and, therefore, difficult to determine if an apparent trend will continue or not. If $d = 2$ in model (5) then the best eventual forecast is a straight line primarily determined by the last $p + d$ sample points. In the sequel, we will determine whether a given time series is best modeled by line + AR (we will consider only noise given by a stationary AR process), (1), or ARIMA with $d = 1$, since ARIMA with $d = 2$ results in the same inference as a line + AR model, i.e., the conclusion that under the business as usual (BAU) scenario the current observed trend will continue.

Two types of questions can be asked given these two models: 1) Given data in the form of a traveltime time series and one of the models, what are the best fit parameters and does the model imply the existence of a significant warming trend? This question has been treated by Woodward and Gray [1993], which we follow here. 2) Which model, line + AR or ARIMA, actually fits the data better [Woodward and Gray, 1994]? The best fit parameters of a given model can be defined in various ways and the estimation of the parameters, while straightforward, must usually be done numerically [see Box and Jenkins, 1976]. If the estimate of the parameter b, in the line + AR model is significantly less than zero, then the line + AR model suggests that a warming trend exists and would be predicted to continue under the BAU assumption. Of course, the validity of such a prediction depends on how well the data has been modeled and whether or not the model remains valid in the future. If $d \geq 2$ in the ARIMA(p, d, q) model, then a trend would be predicted to continue, however, if $d = 1$, while there will be intervals of the series which show substantial "random" trends, such trends would *not* be predicted to continue for a prolonged period of time, even in the BAU case.

3.2. Testing for Trend in Time Series Data

Let us first consider question 1): is there a significant warming trend in the data? We wish to test a given time series for traveltime anomaly, $\{x_t; t = 1,...,n\}$, for the presence of a negative linear trend. The standard approach is to assume the data is a realization from the model given in equation (1), and test the hypothesis $b = 0$ against $b < 0$. If the hypothesis is rejected at an appropriate significance level, then it is generally accepted that a linear trend is present. If one uses a basic regression approach, then the least squares estimators for b and a in equation (1) are

$$\hat{b} = \frac{\sum_{t=1}^{n}(t - \bar{t})X_t}{\sum_{t=1}^{n}(t - \bar{t})^2} , \tag{8}$$

$$\hat{a} = \bar{X} - \hat{b}\bar{t}, \tag{9}$$

where

$$\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t, \tag{10}$$

$$\bar{t} = \frac{1}{n}\sum_{t=1}^{n} t = \frac{n+1}{2}. \tag{11}$$

Sample estimates of these quantities are obtained by substituting the sample data x_t into the appropriate equation. Under the usual regression assumptions that the residuals are independent and normally distributed with mean zero and variance σ^2, the estimated standard error of $\hat{b}$ is given by

$$\hat{SE}^{(1)}(\hat{b}) = \left[\frac{12\sum_{t=1}^{n}(X_t - \hat{a} - \hat{b}t)^2}{(n-2)n(n^2 - 1)} \right]^{1/2}. \tag{12}$$

Under these assumptions, the test statistic $\hat{r}^{(1)} = \hat{b} / \hat{SE}^{(1)}(\hat{b})$ is distributed as Student's t with $n - 2$ degrees of freedom when the null hypothesis H_0: $b = 0$ is true. For a realization $\hat{r}^{(1)}$ can then be compared with the value for which the Student's t distribution

11

with $n - 2$ degrees of freedom yields a given significance level, say 95%. The null hypothesis is rejected, and the line is said to be significant, if $\hat{r}^{(1)} < -t_{n-2}(0.05)$, which is the critical region for the test ($t_{n-2}(0.05)$ is approximately equal to 1.65 for large n, i.e., it is asymptotically normal).

If the independent errors assumption of the standard regression analysis is not made, then the above analysis must be extended. If the residuals E_t in equation (1) are assumed to be given by an autoregressive (AR) process ($q = 0$ in equation (2)), then Bloomfield and Nychka [1991] show that the standard error in this case is given by

$$SE^{(2)}(\hat{b}) = \left[2 \int_0^{1/2} W(f) S(f) \, df \right]^{1/2}, \tag{13}$$

where

$$W(f) = \left| \sum_{t=1}^{n} b_t e^{-2\pi i f t} \right|^2, \tag{14}$$

with

$$b_t = \frac{t - \bar{t}}{\sum_{t=1}^{n} (t - \bar{t})^2}, \tag{15}$$

and $S(f)$ denotes the spectrum of E_t. An estimator for this standard error, $\widehat{SE}^{(2)}(\hat{b})$, can be obtained by replacing $S(f)$ in equation (13) with an appropriate estimate, in our case we fit an autoregressive model to

$$\hat{E}_t = X_t - \hat{a} - \hat{b}t \tag{16}$$

and use the corresponding autoregressive spectral estimator to estimate $S(f)$. The test statistic

$$\hat{r}^{(2)} = \hat{b} / \widehat{SE}^{(2)}(\hat{b}) \tag{17}$$

can now be defined. We test the null hypothesis of no trend, H_0: $b = 0$. For a given time series, x_t, we reject the H_0 at the nominal $\alpha = .05$ level whenever $\hat{r}^{(2)}$ is less than -1.65.

For a given set of traveltime time series data, least squares estimates for a and b can be computed using of equations (8) and (9), and the residuals E_t computed using equation (16). An AR model can then be fit, using standard techniques, to these residuals. In order to determine the power of the test of the hypothesis H_0: $b = 0$, versus the alternative hypothesis H_1: $b < 0$, we analyze a large number of realizations, 100 say, of model (1) for a given set of parameters . The power of the test is the probability of rejecting the null hypothesis H_0 when H_0 is false (H_1 is true) and is conditional on b. We can estimate this probability by counting the proportion of those realizations for which H_0 is rejected, that is, those realizations for which a significant negative slope is found when $\hat{b} = b_1 < 0$. If there is reason to believe that the estimated AR model for the residuals represents the noise for any time, then this same test can be applied to time series of different length. It should be clear that for a given slope, b, realizations of greater length will produce a larger fraction of time series with significant slope. Simulations can be performed which determine the power of this test as a function of series length, and that series length for which the power has a desired value, say .8 for example.

It may be the case that the time series data is modeled better by the model defined by equation (5) (a method which uses the data itself to choose the better model will be discussed below). If $d = 1$ in equation (5) then the null hypothesis of no trend is technically true (the null hypothesis is also true for any stationary time series). Woodward and Gray [1993] have shown, however, that for such time series, for realizations of moderate length, say n around 100, a much larger proportion than the significance level will be rejected, for the null hypothesis test described above. That is, for such realizations, the test will lead one to infer the presence of a trend, when no such trend exists, much more often than the significance level states. This analysis therefore indicates that trend detection in a given set of time series data depends on which model is chosen for the time series. By choosing the best fit line + noise model one has automatically assumed that if there is a trend in the past, there will be a similar trend in the future. If the ARIMA with $d = 1$ is chosen, however, then long trends may be present in some span of the data, but this trend will not be predicted to continue (there will be similarly long spans in the future where the apparent trend will be in the opposite direction). Therefore, if a significant trend using the above test is found in a time series of moderate length, this does not necessarily mean that the best forecast is for the trend to continue, for if the data actually comes from an ARIMA model (or is better fit by an ARIMA model) then this trend is only apparent and the best forecast would be for the trend not to continue. Physically one may think of such "trends" as being triggered by natural events which will eventually subside and reverse the trend.

Unlike cycles, the times when these reversals will occur are not predictable. In the next section, where we analyze specific time series, we will fit both a line + noise and ARIMA model to the data, and compute the proportion of realizations which show significant trend when the true model is ARIMA. If this proportion is substantially larger than the significance level, this will be an indication that the false alarm rate (significant trend is detected when false) is large enough to be of concern. This brings us to the question 2): which model, line + AR or ARIMA, actually fits the data better.

3.3. Selecting a Model: The Bootstrap Procedure

To allow the data itself to choose which model, line + AR or ARIMA, is better, we employ a parametric bootstrap procedure, developed by Woodward and Gray [1994]. The technique is an extension of a procedure suggested by Tsay [1992] for model checking in the time series setting, which consists of assessing whether realizations from a fitted model are sufficiently similar to the observed realization. We utilize the parametric bootstrap to choose between the two competing models, along with a classification procedure to make this decision.

We wish to classify a given time series, $\{x_t; t = 1,...,n\}$, as belonging to either population (1), line + AR, or population (2), ARIMA. Our procedure will be first to model the given time series by model (1), equation (1), and model (2), equation (5), respectively. Many realizations, say 100, are generated from each model for the purpose of providing "training samples". Classification variables, which are potentially useful for discriminating between realizations from models (1) and (2), are calculated for each of the simulated realizations. These are treated as training samples from the populations of classification variables for the two models. The classification variables are also calculated on the original time series and the resulting "observation" is classified as being from model (1) or (2) using standard classification procedures.

Two diagnostic ratios are used as classification variables:

(a) The ratio $R_1 = \widehat{WNV}\,(\text{line} + \text{AR fit}) / \widehat{WNV}\,(\text{ARIMA fit})$, where WNV denotes the sample white noise variance,

(b) $R_2 = \sqrt{|\hat{b}|/\widehat{SE}^{(2)}(\hat{b})}$, with $\hat{b}$ as in equation (8) and $\widehat{SE}^{(2)}(\hat{b})$ as in equation (13).

For each time series to be classified, we perform two classifications, a univariate classification based on the white noise variance ratio alone, and a bivariate classification using Anderson's linear discriminant function [Anderson 1984], based on both of the above ratios. For each classification we generate $B = 100$ bootstrap realizations of the length of the original time series from each of the two models fit to the series. The classification variables R_1 and R_2 are calculated for each bootstrap realization and the resulting $2B$ observations are treated as training samples, B from model (1) and B from model (2). Under the assumption that the distribution of the test statistic array for both models is multivariate normal with a different mean for each model but the same covariance matrix, then Anderson's linear discriminant function is applicable. To estimate this function using the training samples, let $\mathbf{R}_j^{(i)}$, $j = 1, \ldots, n_i$, denote the training sample from model (i), $i = 1, 2$, where the vector $\mathbf{R}_j^{(i)} = (R_{1,j}^{(i)}, R_{2,j}^{(i)})'$ in the bivariate case. The mean for each model is estimated by

$$\overline{\mathbf{R}}^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{R}_j^{(i)}, \quad i = 1, 2, \tag{18}$$

and the pooled estimate of the covariance matrix is

$$\mathbf{S} = \frac{1}{n_1 + n_2 - 2} \left(\sum_{j=1}^{n_1} \left(\mathbf{R}_j^{(1)} - \overline{\mathbf{R}}^{(1)} \right)\left(\mathbf{R}_j^{(1)} - \overline{\mathbf{R}}^{(1)} \right)' + \sum_{j=1}^{n_2} \left(\mathbf{R}_j^{(2)} - \overline{\mathbf{R}}^{(2)} \right)\left(\mathbf{R}_j^{(2)} - \overline{\mathbf{R}}^{(2)} \right)' \right). \tag{19}$$

In our case $n_1 = n_2 = B$. The classification criterion is defined as

$$V(\mathbf{R}) = \left[\mathbf{R} - \frac{1}{2}\left(\overline{\mathbf{R}}^{(1)} + \overline{\mathbf{R}}^{(2)} \right) \right]' \mathbf{S}^{-1}\left(\overline{\mathbf{R}}^{(1)} - \overline{\mathbf{R}}^{(2)} \right), \tag{20}$$

where $\mathbf{R}$ is the observation vector of ratios for the original time series. We then classify the observation, $\mathbf{R}$, as being from model (1) if $V(\mathbf{R}) \geq 0$ and as being from model (2) if $V(\mathbf{R}) < 0$.

For the univariate classification which uses only the ratio of white noise variances as the test statistic, vector quantities such as $\mathbf{R}$ are replaced by the scalar R_1 in equations (18)–(20). The classification criterion (20) is then equivalent to

$$D(R_1) = \left| R_1 - \overline{R}_1^{(2)} \right| - \left| R_1 - \overline{R}_1^{(1)} \right|. \tag{21}$$

We classify the observation, R_1, as being from model (1) if $D(R_1) \geq 0$ and as being from model (2) if $D(R_1) < 0$. This amounts to saying that the observation R_1 is classified as belonging to model (1) if it is closer in absolute value to the average value of the white noise variance ratios for all the realizations from model (1), $\overline{R}_1^{(1)}$, than to those of model (2), and it belongs to model (2) when it is closer to $\overline{R}_1^{(2)}$.

Finally, we would like to estimate the probability of misclassification. Our classification procedures have assumed that a given time series has equal *a priori* probability of being from the population of model (1), line + AR (probability π_1), and model (2), ARIMA (probability π_2), and that an equal cost of misclassification has been assigned to each of the two types of misclassification: 1) classifying a time series as belonging to model (2) when it is actually from model (1), $C(2|1)$, and 2) classifying a time series as belonging to model (1) when it is actually from model (2), $C(1|2)$. Let the probability of the first type of misclassification be $P(2|1)$ and the second type of misclassification be $P(1|2)$. Our classification procedures are chosen to minimize

$$C(2|1)P(2|1)\pi_1 + C(1|2)P(1|2)\pi_2 \tag{22}$$

where we have assumed $\pi_1 = \pi_2 = 1/2$ and $C(2|1) = C(1|2)$. One can construct what is sometimes called a classification matrix, or "confusion matrix", which shows the two probabilities for misclassification as well as the probability of correctly classifying a time series from model (1) as model (1), $P(1|1)$, and the probability of correctly classifying a time series from model (2) as model (2), $P(2|2)$:

Actual Model

|		Line + AR	ARIMA
Predicted Model	Line + AR	$P(1|1)$	$P(1|2)$
	ARIMA	$P(2|1)$	$P(2|2)$

The probabilities in the above chart can be estimated in the following way. For a given time series, a line + AR model and ARIMA model are fit to the data. A large number of realizations (usually limited by computer time), say 100, are run for each model. For each realization of the actual model, B bootstrap realizations are simulated for both predicted models. Each realization of the actual model is then classified as line + AR or ARIMA using the classification procedures described above. The proportion of those actual line + noise realizations which are classified correctly as line + noise gives the probability $P(1|1)$,

the proportion of those actual line + AR realizations which are classified incorrectly as ARIMA gives the probability $P(2|1)$, the proportion of those actual ARIMA realizations which are classified correctly as ARIMA gives the probability $P(2|2)$, and the proportion of those actual ARIMA realizations which are classified incorrectly as line + AR gives the probability $P(1|2)$. These probabilities will be functions of series length. Once models are chosen to fit the data, realizations of any length may be simulated to estimate the probabilities as functions of series length.

4. ANALYSIS

Two traveltime time series were chosen from the available simulation data on which the time series analyses described in the previous section were thoroughly performed, one series from the MASIG model and one from the GFDL model. Each time series gives the traveltime anomaly (the travel time minus the average travel time) taken at monthly intervals along an acoustic path from Hawaii to San Diego. The distance along the path is approximately 4000 kilometers and the travel time is about 2600 seconds (43 minutes). The path is shown in figure 1.
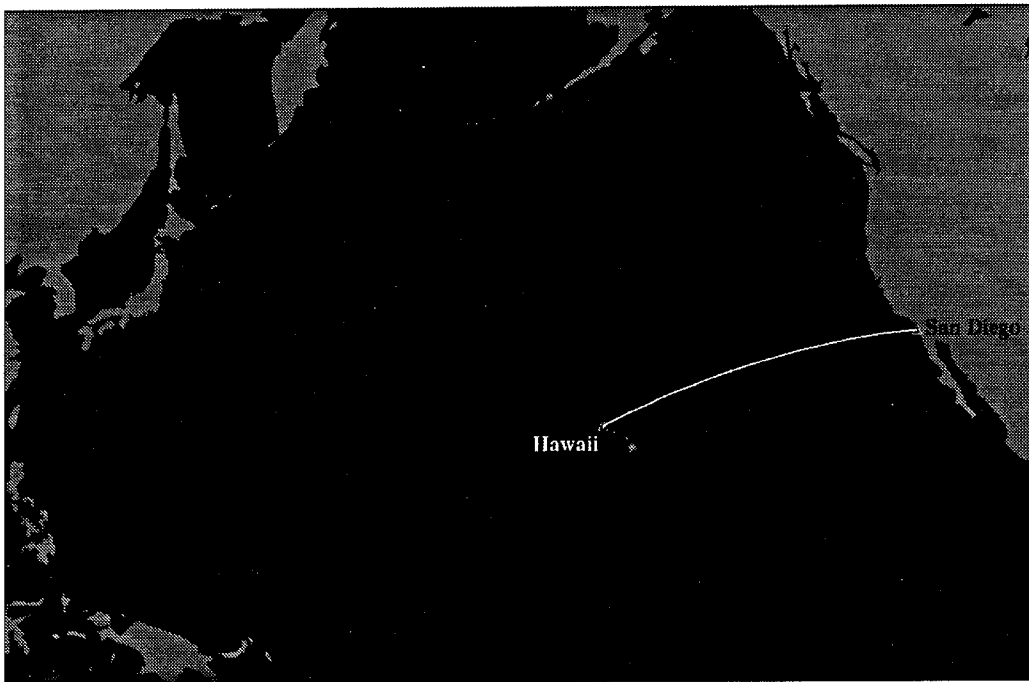
Figure 1. Acoustic path from Hawaii to San Diego.

4.1. MASIG Model

Figure 2 shows a plot of a time series for acoustic traveltime anomaly along a path from Hawaii to San Diego computed using output from the MASIG model. The time axis is given in years. The time series contained 1440 data points spaced at 5 day intervals. For purposes of convenience, a year is taken to be 360 days, so the entire time series represents 20 years. An average over 6 data points was then obtained, resulting in a time series with 240 points, the time interval being one month (30 days). As can be seen in the figure, the traveltime anomalies are between ± 2 seconds. This time series, which represents the model years 1970–1990, has no trend (the slope of the best fit straight line is close to zero).
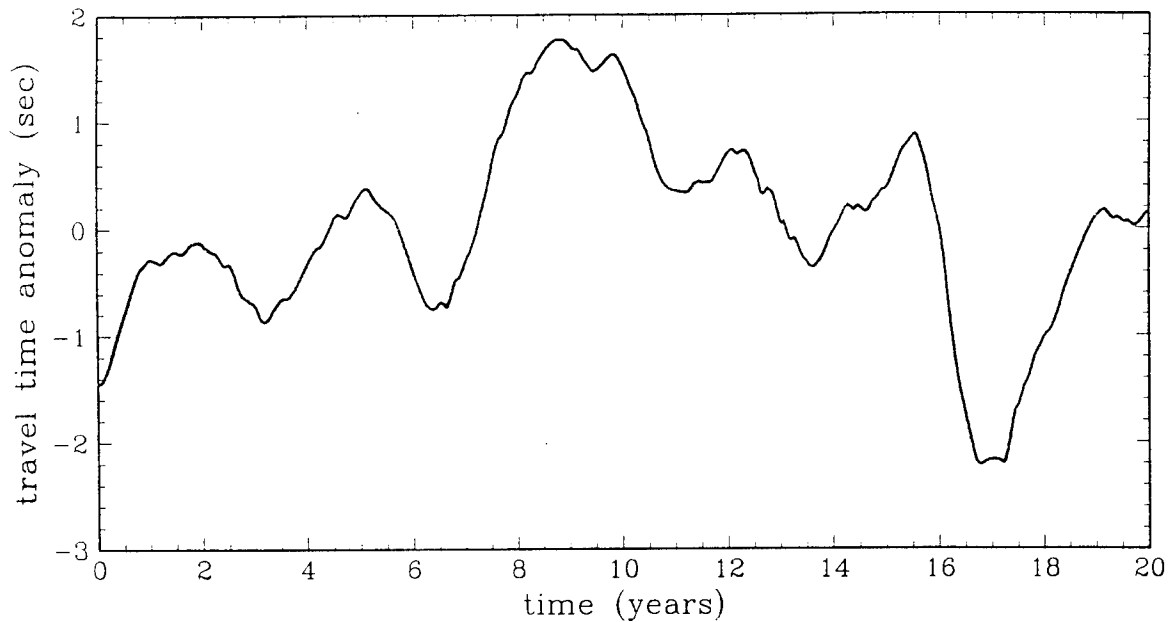
18

Figure 2. Time series for acoustic traveltime anomaly from the MASIG model.

If there were warming occurring during this period the time series would look similar to that shown in figure 2, except that there would be added to it a warming trend. In this context, a warming trend would be given by a negative slope (the best fit straight line through the data would have a negative slope). A slope of –0.10 seconds/year, a change of –2 seconds over 20 years, would correspond to an approximate increase in temperature along the path of 0.01 degrees Celsius per year, or 0.2° C increase over 20 years (a slope of –0.05 sec/yr would give half these values). For the following analysis, which is meant to demonstrate how, and with what reliability, significant trends can be extracted from time series data, we consider the time series given in figure 2 with lines of various slopes added to the original time series.

Lines with slopes of –0.05 sec/yr and –0.10 sec/yr were added to the time series given in figure 2. The resulting time series, along with the original time series, are plotted in figure 3. The time series with the added slope of –0.05 sec/yr is given by the dotted curve and the time series with the added slope of –0.10 sec/yr is given by the dashed curve (the original time series, marked "no trend", is plotted with the solid curve). The first question to be addressed is whether there is a significant trend, as a function of series length, in these times series. Each time series has 240 values (20 years). We test for trend in each of these time series for various portions of the series using the test described in the previous section. We consider series of length 5, 10, 15, and 20 years (the first 60 points, the first 120 points, etc.). For each series the test statistic $\hat{r}^{(2)}$, given in equation (17), is
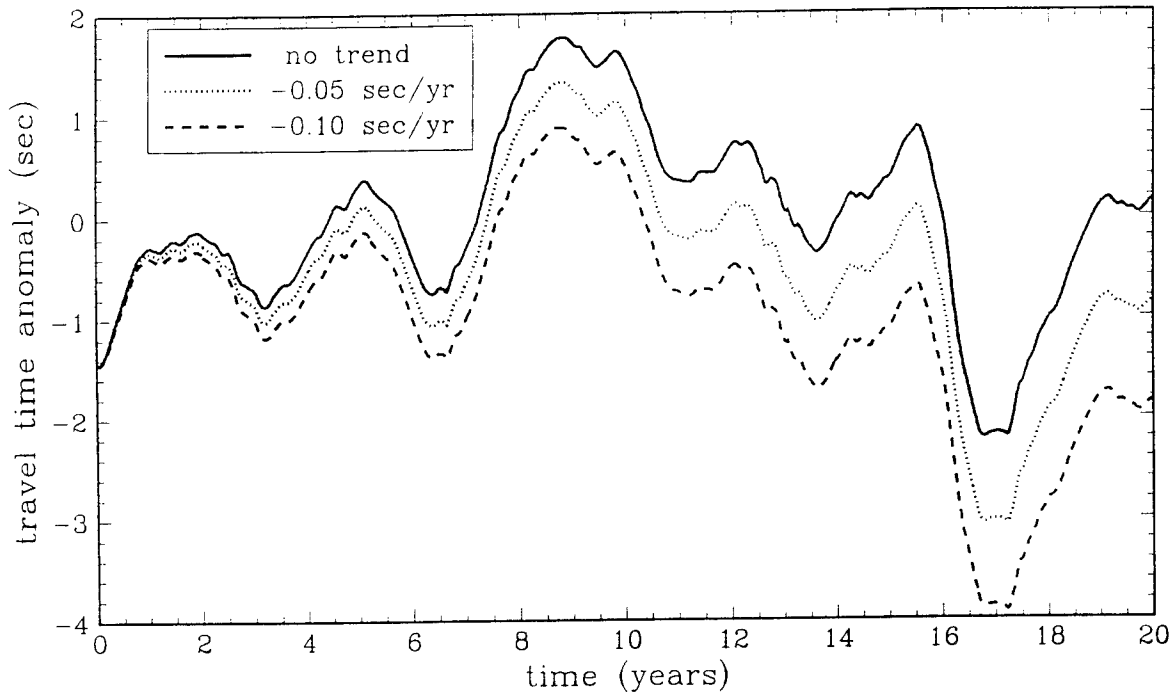
19

Figure 3. MASIG time series with trends added.

computed. If $\hat{r}^{(2)}$ is less than -1.65, we say there is a significant negative slope (which corresponds to a significant warming trend). Table 1 gives the value of $\hat{r}^{(2)}$ (and whether slope is significant) for the two slopes considered and the four series lengths.

	SLOPE	
Years	$-.05$ sec/yr	$-.10$ sec/yr
5	1.207 (Not Sig)	0.003 (Not Sig)
10	3.956 (Not Sig)	0.012 (Not Sig)
15	0.747 (Not Sig)	-0.001 (Not Sig)
20	-1.190 (Not Sig)	-2.217 (Significant)

Table 1. $\hat{r}^{(2)}$ (and significance) for time series from MASIG model.

As can be seen in the table, the only significant negative slope occurs for the slope of $-.10$ sec/yr and the full 20 years length. Figure 4 shows a plot of this time series with the significant best fit line superimposed (dashed line). Interestingly, there is a significant
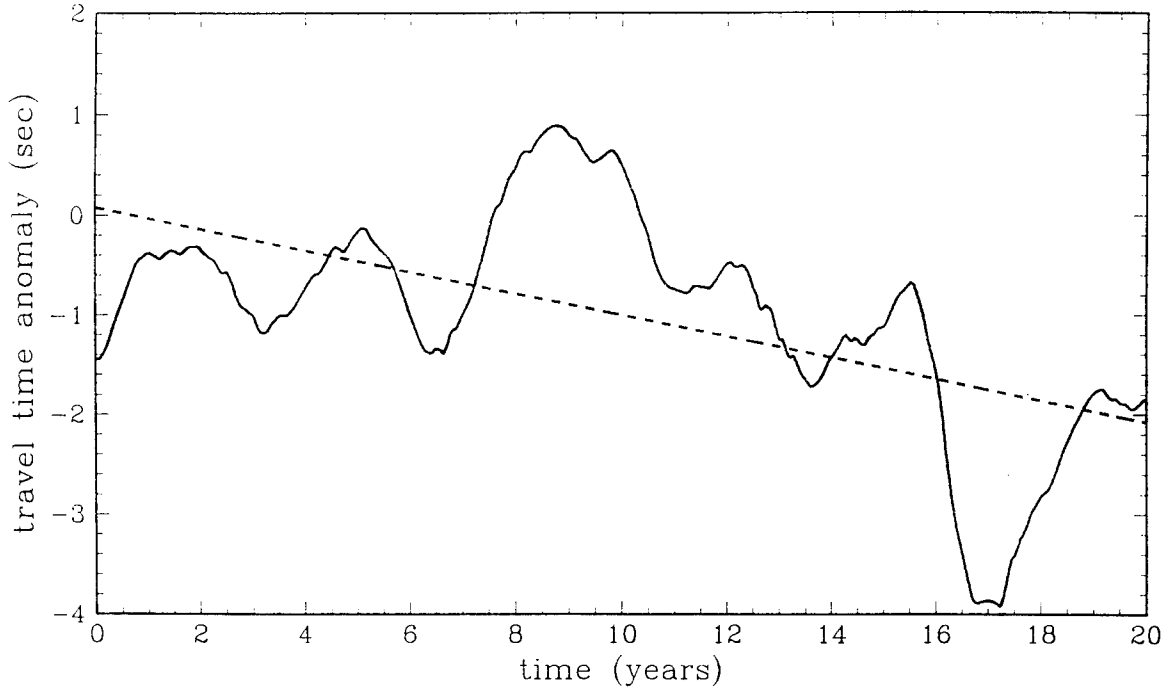
Figure 4. MASIG time series with significant trend.

positive slope in the first ten years of the data even when the slope of $-.05$ sec/yr is added to the data, which an examination of figure 3 will show is not surprising. Any portion of the original time series, with a line having any slope added to it, may be tested for trend in this manner.

In order to determine how often one can expect to find a significant trend in similar time series (we are assuming now that an experimental time series will have the same correlation structure as those in figure 3), which amounts to computing the power of the test for significance, we model the time series as line + AR as in equation (1). We model the original time series, with no slope added, as an AR process. The best fit model is an AR(10), given by

$$\left(1 - 2.1765B + 1.6926B^2 - .9443B^3 + .7730B^4 - .6793B^5 + .4348B^6 \right.$$
$$\left. -.0608B^7 + .2369B^8 - .4745B^9 + .2024B^{10}\right)E_t = a_t,$$
(23)

with white noise variance $\sigma_a^2 = .001196$. The factor table associated with the AR(10) operator given in equation (23) is shown in table 2.

Factor	Absolute Reciprocal of Root	Frequency (cycles/month)	Period (years)
$1 - 1.873B + .884B^2$	.940	.014	5.87
$1 - .768B + .818B^2$	.905	.180	0.46
$1 + .670B + .738B^2$	.859	.314	0.27
$1 - 1.540B + .635B^2$	.797	.041	2.02
$1 + 1.334B + .598B^2$	.773	.416	0.20

Table 2. Factor table associated with AR(10) model fit to MASIG data.

As can be seen in the factor table, the dominant frequency (the frequency associated with the root with absolute reciprocal closest to the unit circle) is 0.014 cycles per month, with a corresponding period of about 6 years. This period may be related to an ENSO (El Niño Southern Oscillation) cycle and is not surprising.

Realizations from line + AR(10) models of any length, with a given line, may now simulated. Figure 5 shows typical realization from the model

$$X_t = -0.0083t + E_t, \tag{24}$$

with E_t computed using equation (23). The slope -0.0083 sec/month is equal to -0.10 sec/yr. As can be seen, this typical realization has similar statistical properties as the dashed curve in figure 3. To compute the power of the test for trend, 100 realizations from the model given by equation (24) and 100 realizations from a similar model with a slope of -0.05 sec/yr (-0.00042 sec/month) with lengths of 5, 10, 15, 20, 25, and 30 years were simulated and each realization was tested for trend. The percentage of those realizations which had a significant trend are shown in table 3.

	SLOPE	
Years	−.05 sec/yr	−.10 sec/yr
5	32	40
10	29	41
15	23	56
20	38	79
25	55	94
30	66	100

Table 3. Percentage of realizations with significant slope for the MASIG model.

As would be expected, the table shows that it is more likely to detect a trend for a steeper slope and for a longer time series. As can be seen in the table, a slope of −.10 sec/yr is detectable about 80% of the time for time series of 20 years length. Note this is the slope that was added to the MASIG data and detected after 20 years, so that these statistical simulations appear quite consistent with the MASIG model. However, in this case, since
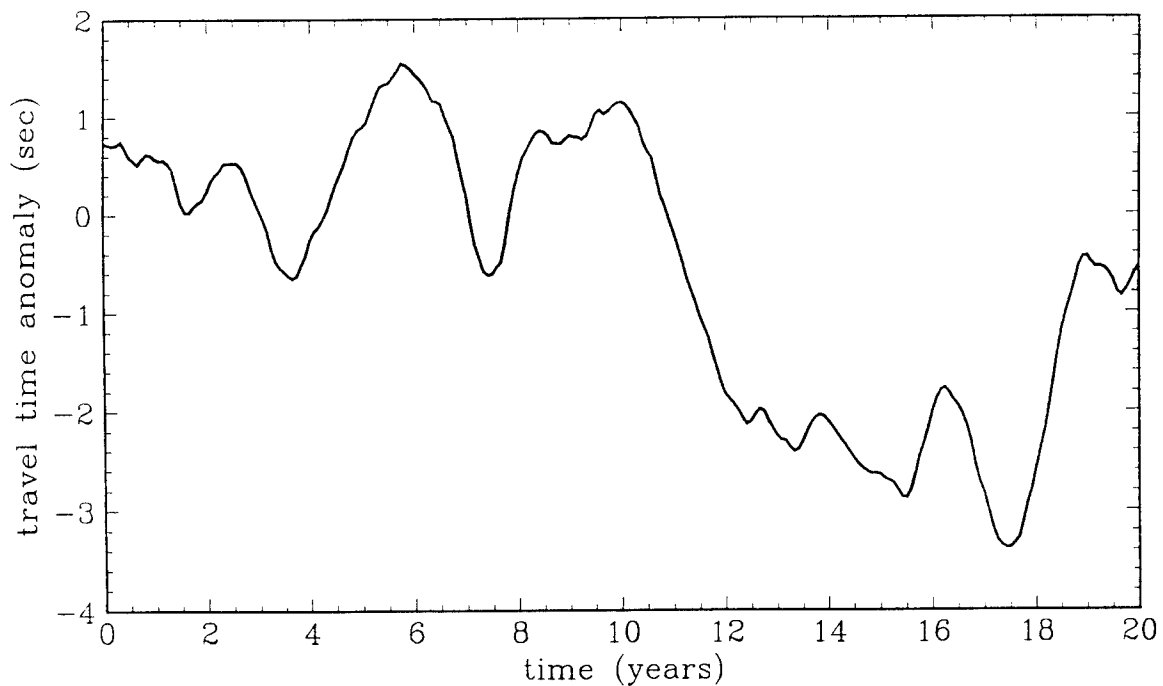

Figure 5. A typical realization from line + AR(10) model.

we can generate as many realizations as we like, we can see that for such data this slope would be detected within 25 to 30 years with virtual certainty. A slope of −.05 sec/yr is reasonably detectable (about 66%) after about 30 years.

The best forecast for the line + AR(10) model with a significant slope is for the trend to continue indefinitely. However, one does not know in advance the best model for the data. For example, an ARIMA model could produce similar realizations. We now fit the data to an ARIMA model. A line with slope −.10 sec/yr was then added to the original time series and the resulting best fit ARIMA model was 10th order with one unit root, i.e., it was an ARIMA(9,1,0), given by

$$(1 - B)\left(1 - 1.2060B + .5230B^2 - .4449B^3 + .3433B^4 - .3402B^5 \right.$$
$$\left. + .1009B^6 + .0471B^7 + .2706B^8 - .1800B^9\right)X_t = a_t, \tag{25}$$

with white noise variance $\sigma_a^2 = .001231$. The factor table associated with the ARIMA(9,1,0) operator given in equation (25) is shown in table 4.

Factor	Absolute Reciprocal of Root	Frequency (cycles/month)	Period (years)
$1 - B$	1.000	0	—
$1 - .778B + .815B^2$	.903	.179	0.47
$1 - 1.734B + .782B^2$	.884	.032	2.65
$1 + .661B + .727B^2$	.852	.313	0.27
$1 + 1.313B + .582B^2$	.762	.415	0.20
$1 - .668B$	.668	0	—

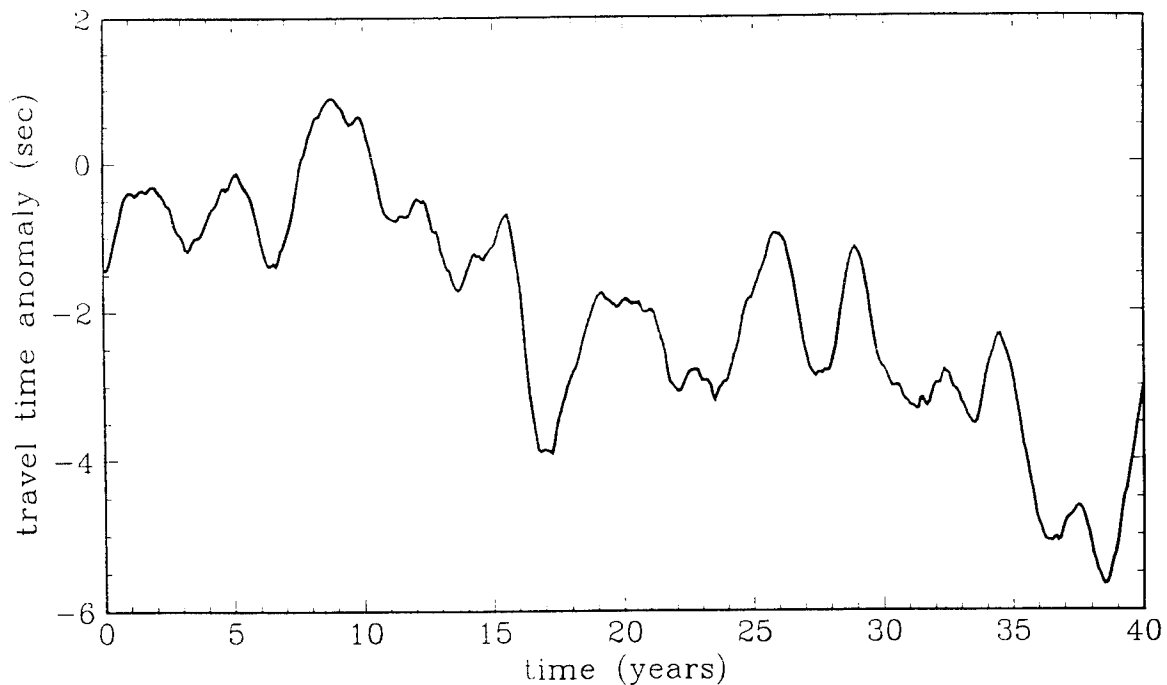*Table 4. Factor table associated with ARIMA(9,1,0) model fit to MASIG data.*

Figure 6. A typical realization from ARIMA(9,1,0) model.

This model will produce many realizations similar to the original time series + line but for which the trend would not be forecast to continue. Figure 6 shows a typical realization from this model. Even though there is a significant slope in this time series (when modeled as line + AR), as ARIMA(9,1,0) that trend would not continue for a sufficiently long realization length. 100 realizations from this model with a series length of 20 years (240 points) were simulated and the trend test found significant slope in 59% of the realizations, which is a substantial percentage for a model for which there is no true trend. Thus, clearly it is necessary to determine which model is more representative of the data.

The difference in the two models, line + AR(10) and ARIMA(9,1,0), is demonstrated by comparing figures 7 and 8, respectively. Each figure shows four realizations from each model, figure 7 showing realizations from the line + AR(10) model, with slope equal to –0.10 sec/yr, and figure 8 showing realizations from the ARIMA(9,1,0) model. The first 20 years of each plot is the original time series plus a line with slope –0.10 sec/yr and the second 20 years of each is a realization from the appropriate model. As can be seen in figure 7, each time series from the line + AR(10) model continues during the second 20 years with a slope equal to the slope for the first 20 years. In contrast, the second 20 years for the ARIMA(9,1,0) realizations have both positive and negative slopes, demonstrating that the negative slope in the first 20 years would not be forecast to continue if the ARIMA(9,1,0) was a suitable model.
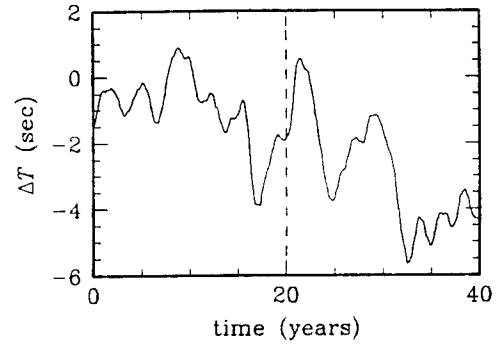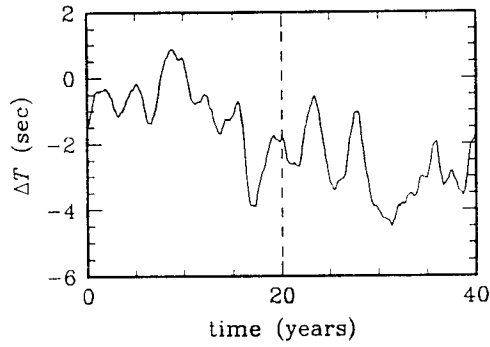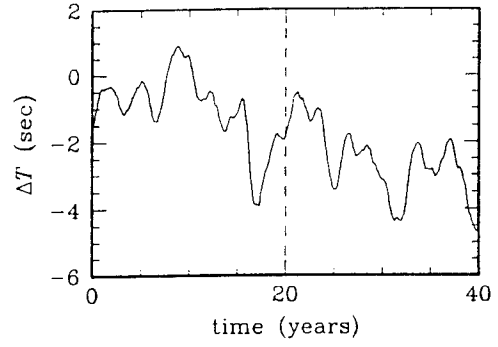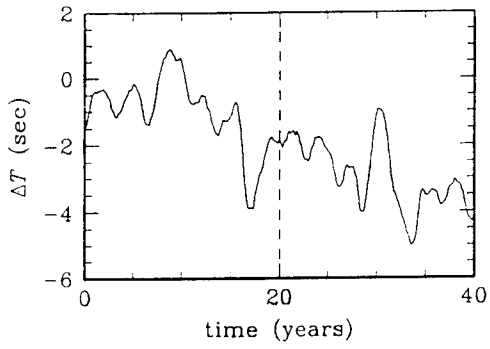
25

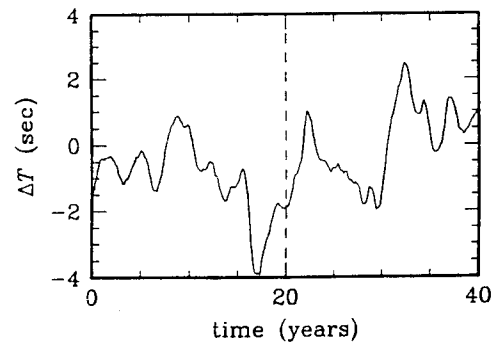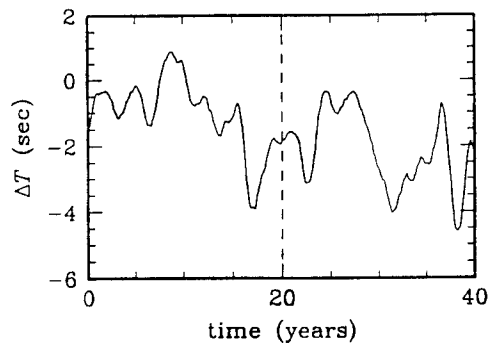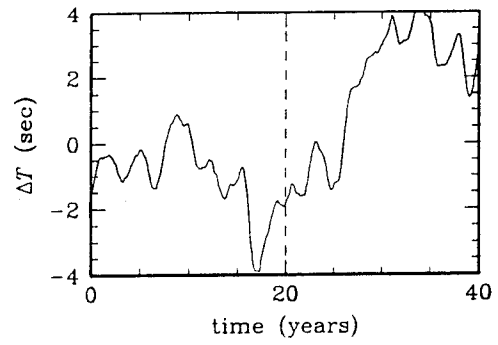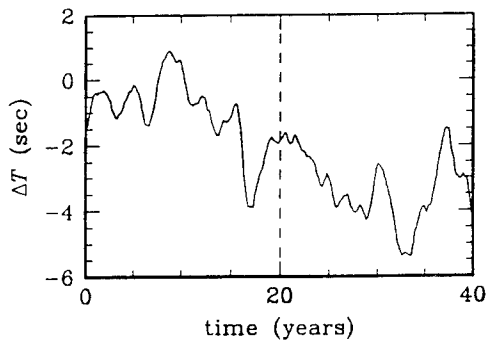*Figure 7. Realizations from line + AR(10) model.*

Figure 8. Realizations from ARIMA(9,1,0) model.

To determine if the data alone can determine which model best describes the MASIG time series, line + AR(10) or ARIMA(9,1,0), we apply the parametric bootstrap classification procedure described in the previous section. We consider only the series with the added −0.10 sec/yr slope and the full 20 years, since this was the only time series containing a significant trend (see table 1). 100 bootstrap realizations from the line + AR(10) model given in equation (24) and 100 bootstrap realizations from the ARIMA(9,1,0) model given by equation (25) were simulated and the appropriate test statistics were computed. The classification criteria given in equation (21) for the univariate white noise ratio classification and in equation (20) for the bivariate classification combining the white noise ratio and the square root of the t-statistic were then computed for the time series data. The result was that the univariate classification chose the ARIMA as the best model and the bivariate classification chose line + AR, indicating a close call.

To estimate the probability of misclassification and fill in the entries to the classification matrix, 20 realizations of each model were simulated. For each realization with a significant trend (there were 14 for the line + AR(10) model and 11 for the ARIMA(9,1,0) model, consistent with the statistics of table 2), 50 bootstrap realizations from each model were simulated (the number of realizations is limited by computer time constraints) and test statistics calculated. The classification matrix for the univariate white noise ratio classification is shown in table 5 and for the bivariate classification in table 6.

		Actual Model	
		Line + AR	ARIMA
Predicted Model	Line + AR	11/14	3/11
	ARIMA	3/14	8/11

Table 5. Classification matrix for the univariate white noise ratio classification for MASIG data.

Actual Model

|		Line + AR	ARIMA
Predicted Model	Line + AR	11/14	4/11
	ARIMA	3/14	7/11

Table 6. Classification matrix for the bivariate classification for MASIG data.

The tables show that for each classification the estimated probability of choosing line + AR when the time series actually is line + AR is almost 80%. Note that the white noise classification has a 20% chance of selecting ARIMA when the time series is line + AR. The probability of choosing ARIMA correctly is about 60–70%. It is not surprising that the two classification procedures do not agree here. Which of the two approaches is best is yet to be determined. However, the upshot of this analysis is that for data with the amount of variability as large as the MASIG data, 20 years may not be quite long enough to determine the proper model with very high confidence. It should be pointed out that these calculations assumed an equal cost of misclassification (see equation (22)). If there were a reason to assume it would be more costly to misclassify the line + AR (and there might be because one would falsely predict that global warming was not occurring) then the above results might change enough to more strongly choose line + AR as the more acceptable model for the MASIG data.

4.2. GFDL Model

Traveltime time series along the path from Hawaii to San Diego were obtained from the GFDL output by integrating sound speed, obtained from temperature, salinity, and pressure (depth) using an appropriate equation of state, along the sound channel axis (sound speed minimum). Two time series were obtained from the two separate runs: the control run in which constant levels of CO_2 were input to the system and the run with increasing CO_2 levels of 1% per year. Figure 9 shows the time series for traveltime anomaly obtained from the control run and figure 10 from the 1% run each plotted for 100 years (1200 monthly values). It is perhaps surprising that the control run shows a greater "cooling" trend (the best fit line has a positive slope of about 1 second over 100 years) than the 1% run has "warming" trend (slope of about –0.7 seconds over 100 years). The

appropriate interpretation of these two runs, according to Ron Stouffer at GFDL, is that the best prediction of warming is given by "subtracting" the control run from the 1% run. This, indeed, is the purpose for having a control run. An examination of the two plots indicates that the two time series are relatively equal for the first 30 years and then start to diverge following this "warm-up" period. The best measure of global warming for 20 year or so spans in situations where warming has been going on for many years should then come from the later portions of these time series. We therefore choose to analyze the last 500 months (42 years) of these time series. A best fit line through the last 500 points of the control time series was computed and this line was subtracted from the 1% series. The resulting time series, which has similar statistical character as the 1% series (point by point subtraction of the two time series would result in a time series with about twice the variance of the individual series) is plotted in figure 11. This is the time series which we will analyze in a similar way as the MASIG time series. The best fit line for this series has a slope of about −0.033 sec/yr.
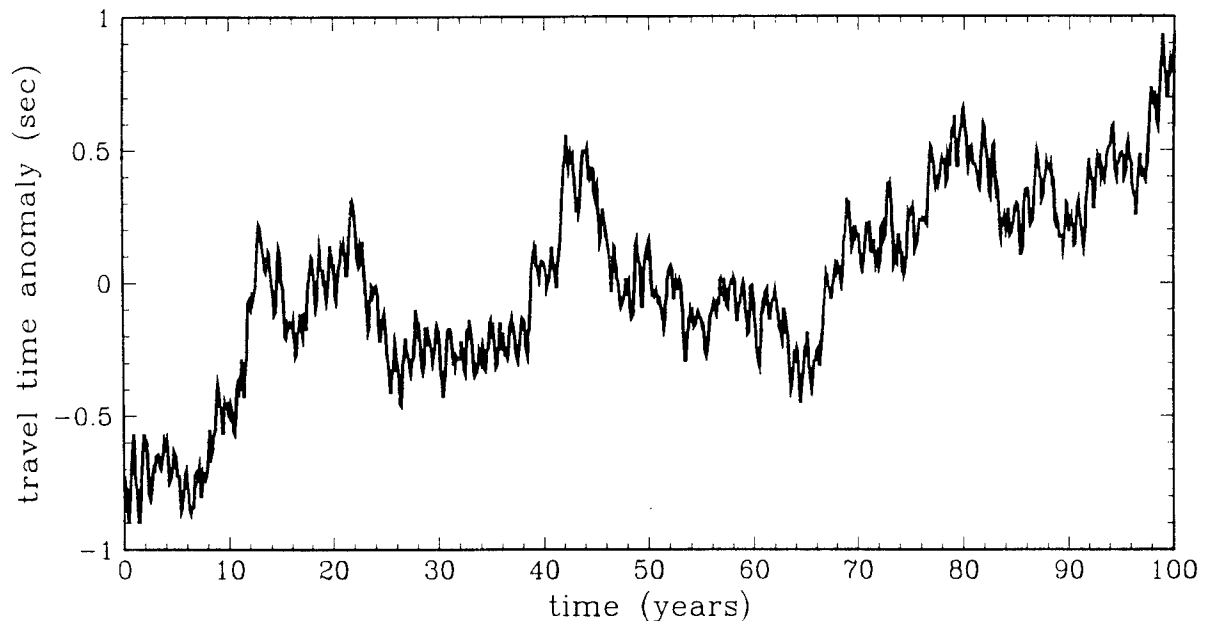
Figure 9. Time series of the control run from the GFDL model.
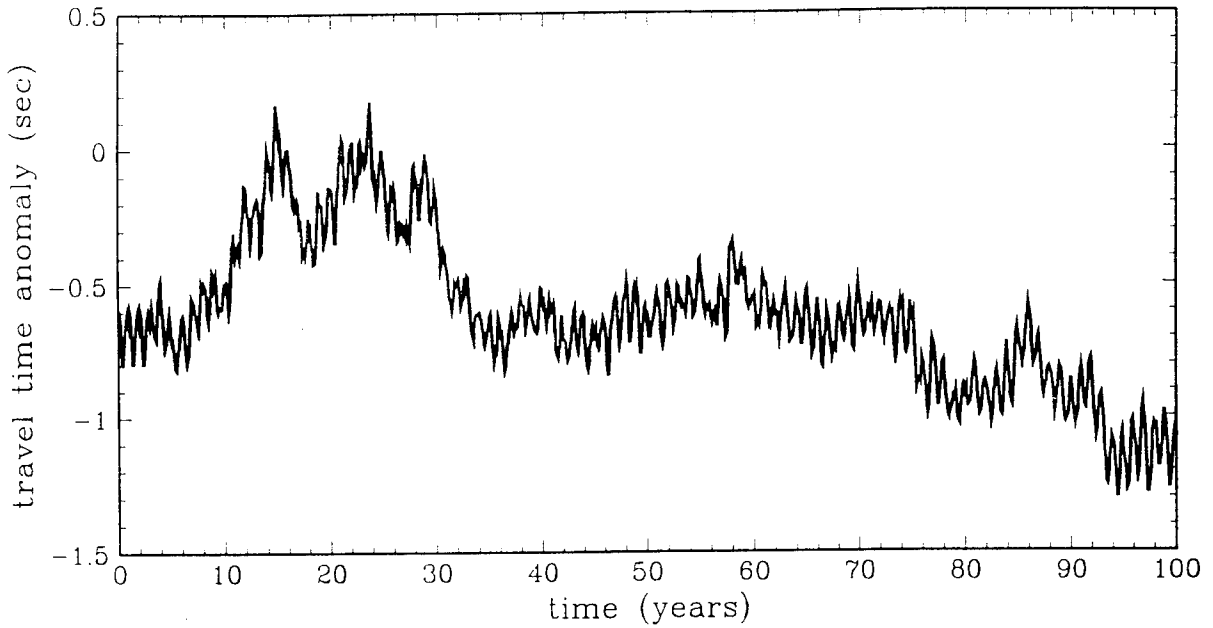
Figure 10. Time series of the warming run from the GFDL model.
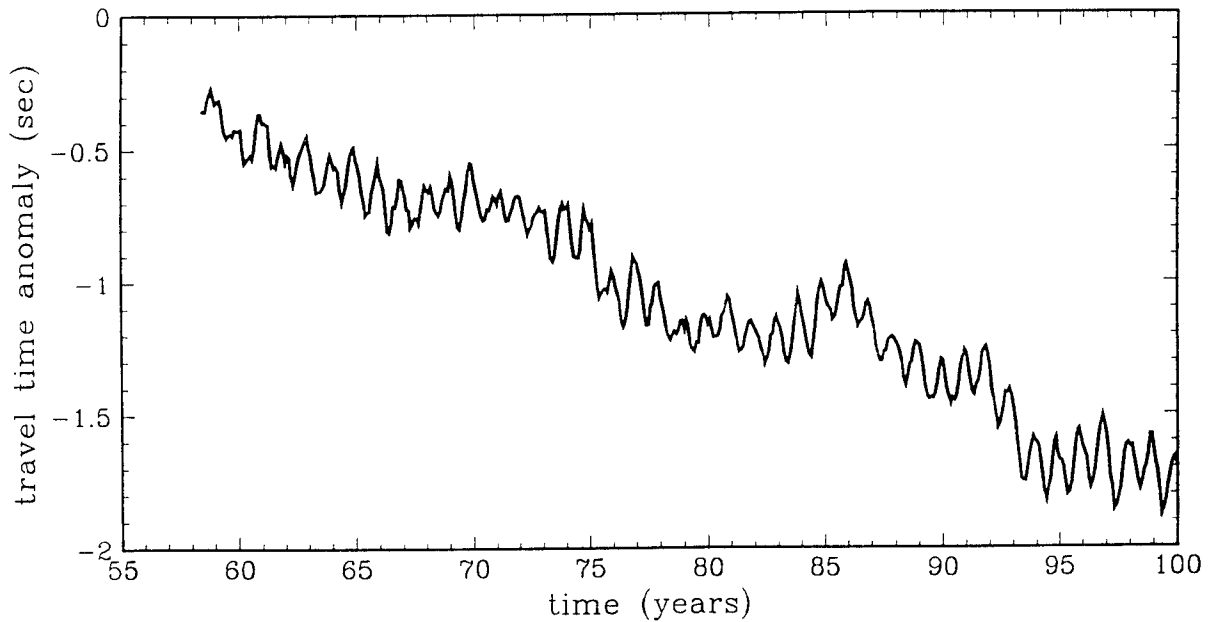
Figure 11. Resultant time series for acoustic traveltime anomaly from the GFDL model.

We divide the time series in figure 11 into 8 intervals with length 5 years (60 points), 4 intervals with 10 year length (120 points) and 2 intervals with 20 year length (240 points). For each resulting time series we test for the presence of a trend by computing $\hat{r}^{(2)}$ for each series and accept the time series as having significant trend (negative slope) if $\hat{r}^{(2)}$ is less than -1.65. The results are compiled in table 7.

30

Starting Month	5 Years	10 Years	20 Years
1	Significant	Significant	Significant
61	Significant		
121	Significant	Significant	
181	Significant		
241	Not Significant	Not Significant	Significant
301	Not Significant		
361	Significant	Significant	
421	Not Significant		

Table 7. Significance of trend for various intervals of the GFDL time series.

The first entry in the second column of the table shows that the time series consisting of the first 60 points (5 years) of the GFDL time series has a significant slope, the second 5 year interval, including points 61–120, also has significant slope and so on. In all 5 out of the 8 five year intervals have significant slope and 3 out 4 of the ten year intervals have significant slope. All intervals of length greater than or equal to 20 years had slopes which were significant.

The time series was then fit with a line, with slope equal to –0.033 sec/yr, and this line was subtracted from the time series. The residuals were modeled as an AR(13) process, given by

$$\left(1 - 1.3308B + .4678B^2 - .0636B^3 - .0687B^4 + .0910B^5 + .1755B^6 - .1864B^7 \right.$$
$$\left. + .0016B^8 + .0078B^9 - .1122B^{10} - .1075B^{11} + .0251B^{12} + .1395B^{13}\right)E_t = a_t, \tag{26}$$

with white noise variance $\sigma_a^2 = .000524$. The factor table associated with the AR(13) operator given in equation (26) is shown in table 8.

Factor	Absolute Reciprocal of Root	Frequency (cycles/month)	Period (months)
$1 - 1.790B + .978B^2$	.989	.084	11.9
$1 - .939B$	.939	0	—
$1 + .054B + .812B^2$	.901	.255	3.9
$1 - .898B$	.898	0	—
$1 - .795B + .795B^2$	.892	.177	5.7
$1 + 1.379B + .627B^2$	.792	.418	2.4
$1 + .841B + .568B^2$	.754	.344	2.9
$1 + .736B$	.736	.5	2

Table 8. Factor table associated with AR(13) model fit to GFDL residuals.

The dominant frequency is 0.084 cycles per month, with a corresponding period of 12 months, or 1 year. The yearly cycle is readily apparent in figure 11. The other dominant frequencies are simply harmonics of this lowest frequency. The best fit line + AR(13) model for the GFDL time series is, thus,

$$X_t = -0.0027t + E_t, \qquad (27)$$

with E_t computed using equation (26). The slope −0.0027 sec/month is equal to −0.033 sec/yr.

The best fit ARIMA model is 13th order with one unit root, i.e., it is an ARIMA(12,1,0), given by

$$(1 - B)\big(1 - .3440B + .1245B^2 + .0611B^3 - .0028B^4 + .0880B^5 + .2551B^6$$
$$+ .0583B^7 + .0566B^8 + .0676B^9 - .0473B^{10} - .1573B^{11} - .1230B^{12}\big)X_t = a_t, \qquad (28)$$

with white noise variance $\sigma_a^2 = .000535$. The factor table associated with the ARIMA(12,1,0) operator given in equation (28) is shown in table 9.

Factor	Absolute Reciprocal of Root	Frequency (cycles/month)	Period (months)
$1-B$	1.000	0	—
$1-1.790B+.977B^2$	.989	.084	11.9
$1+.062B+.807B^2$	.898	.256	3.9
$1-.792B+.783B^2$	.885	.176	5.7
$1-.829B$	.829	0	—
$1-1.357B+.605B^2$	.778	.419	2.4
$1+.849B+.553B^2$	.744	.347	2.9
$1+.718B$	.718	.5	2

Table 9. Factor table associated with ARIMA(12,1,0) model fit to GFDL time series.

The ARIMA(12,1,0) operator has a similar structure to the AR(13) as can be seen by comparing table 9 with table 8. The rather random long trends in the data are due to the unit root which somewhat dominates the behavior of the series. The next most dominant frequency has a corresponding period of 12 months (1 year) and the other frequencies are simply harmonics of this frequency.

100 realizations from the line + AR(13) model given by equation (27) and 100 realizations from the ARIMA(12,1,0) model given by equation (28) with lengths of 2.5, 5, 7.5, and 10 years were simulated and each realization was tested for trend. The percentage of those realizations which had a significant trend are shown in table 10.

Years	Line + AR(13)	ARIMA(12,1,0)
2.5	61	60
5.0	66	63
7.5	85	67
10.0	93	62

Table 10. Percentage of realizations with significant slope for the GFDL model.

33

As expected, it is more likely to observe a trend for the line + AR(13) model the longer time series, with virtual certainty obtained in series not much longer than 10 years (120 points). Interestingly, the chance of detecting a significant line in the ARIMA(12,1,0) model is nearly independent of series length for the lengths considered, although, if the length of the series was sufficiently long the chance of a significant line in the ARIMA would be negligible. However, for data of this length there is a very real chance that an ARIMA might look like a line + noise, as the table shows.

We now apply the bootstrap procedure to the GFDL time series to determine which model best describes the data. The time series was partitioned into shorter time series of length 5 years (60 points), 10 years (120 points), 20 years (240 points), and 40 years (480 points). 100 bootstrap realizations from each model were simulated and the various test statistics for the data were compared with the averages of those statistics from the bootstrap realizations using the classification criteria equations (20) and (21). Of the 8 five year time series, 5 had significant slope. The bivariate procedure classified all five time series as line + AR and the univariate procedure classified 3 out of the 5 time series as ARIMA. For the time series of 10 years length, the bivariate procedure classified 2 out of 3 as line + AR and the univariate procedure classified 2 out of 3 as ARIMA (3 of the 4 time series had significant slope). For the time series of 20 years length, the bivariate procedure classified 2 out of 2 as line + AR and the univariate procedure classified 1 out of 2 as line + AR. For the single time series of length 40 years, both procedures classified the time series as line + AR.

Table 11 compiles the classification statistics. For each model a number of realizations (limited by computer time constraints) are simulated and each realization is classified as either line + AR or ARIMA, using both the bivariate classification procedure and the univariate white noise ratio classification procedure. Time series of length 5, 10, 20, and 40 years were simulated. In the table, each entry gives the number of realizations classified as the correct model / the total number of realizations (percent correct). For example, the first entry for time series length 5 years, actual model line + AR(13) and bivariate classification is 12/18 (67%), which means that 12 out of the 18 (67%) realizations from the line + AR(13) model were classified correctly, using the bivariate classification (6 out of 18 (33%) were classified incorrectly as ARIMA).

Years	Line + AR(13)		ARIMA(12,1,0)	
	Bivariate	Univariate	Bivariate	Univariate
5	12/18 (67%)	22/88 (25%)	7/14 (50%)	52/68 (76%)
10	8/17 (47%)	28/110 (25%)	10/15 (67%)	52/67 (78%)
20	25/30 (83%)	91/230 (40%)	16/26 (62%)	60/78 (77%)
40	10/10 (100%)	94/110 (85%)	3/4 (75%)	34/49 (69%)

Table 11. Classification table for the GFDL time series.

It is clear from the table that if the time series is actually line + AR then the bivariate classification works much better than the univariate classification , and, in general, works better the longer the time series. Indeed, the univariate classification works poorly for short time series and these results indicate it should not be used in such cases. If the time series is actually generated by the ARIMA model, then the classification is not as sensitive to series length (this is to be expected) and the univariate classification works somewhat better, especially for shorter series. If the cost for misclassifying line + AR incorrectly is greater than misclassifying ARIMA (which is likely to be the case when attempting to predict global warming), then the bivariate classification should certainly be used, and, if so, the GFDL time series would be classified as line + AR and the warming trend would be predicted to continue.

5. CONCLUSION

Figure 12 plots a portion of the GFDL time series and the MASIG time series (with the steepest slope) together for comparison. The GFDL time series is the first 20 years of the time series from figure 11 and the MASIG time series is the dashed line of figure 3. The slope (of the best fit straight) of the GFDL time series is –0.033 sec/yr and the slope of the MASIG time series is –0.10 sec/yr, three times that of the GFDL time series. It is evident, even by a glance at figure 12, that due to the greater variability in the MASIG data, the "trend" is more apparent in the GFDL time series, even though the slope is much less (in absolute value). Indeed, our analysis of the previous section has demonstrated this. According to that analysis it would take about 20 years to detect a trend 80% of the time in a time series like the MASIG series (see table 3) but only about 7 years for the GFDL time series (table 9). It is most likely the case that experimental variability will be closer to that given by the MASIG simulated data (this data driven model was designed to reproduce variabilities seen during the years 1970–1990) and the actual warming trend closer to that predicted by the GFDL model. If the we add to the MASIG data a line with slope equal to –0.033 sec/yr (the GFDL prediction), then our analyses show that it would take nearly 40 years to detect a trend just 50% of the time.
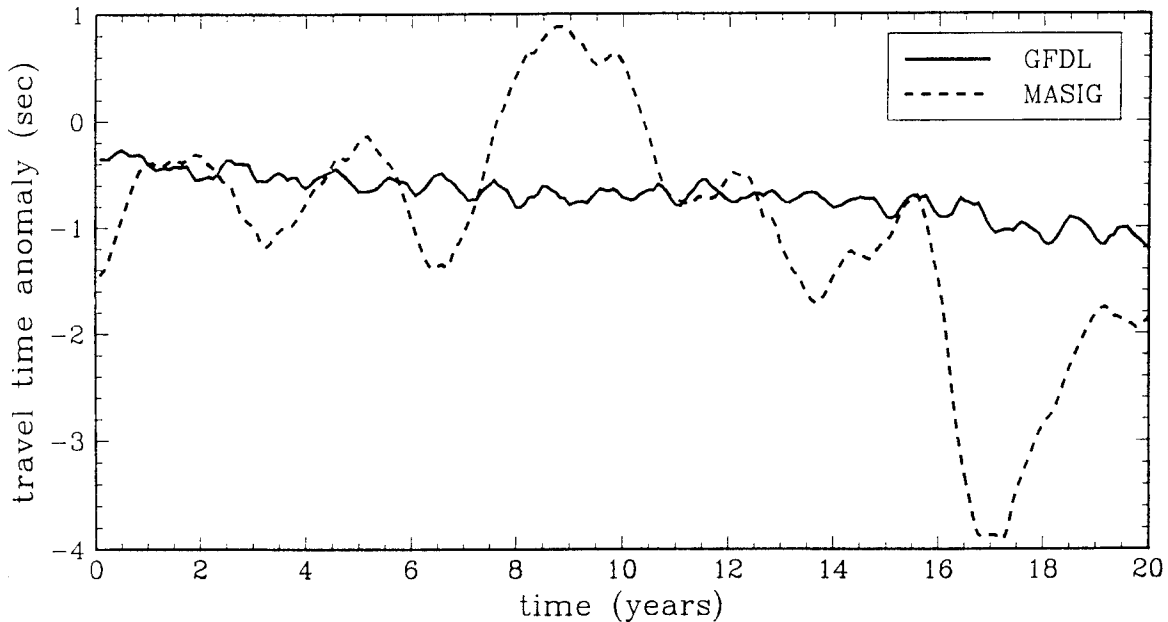

Figure 12. Comparison of time series from the GFDL and MASIG models.

This large discrepancy in the results for the two different models and the large amount of time it would take to detect a warming trend using *one path only* given the worse case of MASIG variability and GFDL trend leads to two conclusions. First, more work by the modelers must be done to improve the ocean and climate models. Experimental data acquisition of the Acoustic Monitoring of Global Ocean Climate program will be essential to make these improvements. Second, extension of our trend extraction techniques which uses time series data on many paths at once is necessary. It is reasonable to suppose that the length of time necessary to detect a trend in a set of traveltime time series will be reduced by considering many paths together. How much this time can be reduced is yet to be determined. Further research is therefore needed to see if such reduction can be achieved.

6. REFERENCES

Anderson, T. W., 1984: *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New York, 675 pages, second edition.

Bloomfield, P., and D. Nychka, 1991: Climate spectra and detecting climate change. Unpublished manuscript.

Box, G. E. P., and G. M. Jenkins, 1976: *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 575 pages, revised edition.

Gray, H. L., and W. A. Woodward, 1986: A new ARMA spectral estimator. *Journal of the American Statistical Association*, **81**, Number 396, 1100–1108.

Gray, H. L., W. A. Woodward, and N. F. Zhang, 1994: *Time Series Analysis*. In preparation.

Manabe, S., R. Stouffer, M. Spellman, and K. Bryan, 1991: Transient responses of a coupled ocean-atmosphere model to gradual changes of atmospheric CO_2. Part I: Annual mean response. *Journal of Climate*, **4**, 785–818.

Mikolajewicz, U., E. Maier-Reimer, and T. Barrett, 1990: Acoustic detection of greenhouse-induced climate changes in the presence of slow fluctuations of the thermohaline circulation. *Journal of Physical Oceanography*, **23**, 1099–1109.

Mikolajewicz, U., B. Santer and E. Maier-Reimer, 1990: Ocean response to greenhouse warming. *Nature*, **345**, 589–593.

Munk, W., and A. M. G. Forbes, 1989: Global ocean warming: an acoustic measure?, *Journal of Physical Oceanography*, **19**, 1765–1778.

Munk, W., and C. Wunsch, 1979: Ocean acoustic tomography: a scheme for large-scale monitoring, *Deep Sea Research*, **26**, 123–161.

Pares-Sierra, A. and J. O'Brien, 1989: The seasonal and internal variability of the California current system: A numerical model. *Journal of Geophysical Research*, **94**, 3159–3180.

Semtner, A. and R. Chervin, 1988: A simulation of the global ocean circulation with resolved eddies. *Journal of Geophysical Research*, **93**, 15,502–15,522.

Spiesberger, J. L., P. J. Bushong, K. Metzger, and T. G. Birdsall, 1989: Ocean acoustic tomography: estimating the acoustic travel time with phase. *Journal of Physical Oceanography*, **19**, 1073–1090.

Spiesberger, J. L., and K. Metzger, 1991: Basin-scale tomography: a new tool for studying weather and climate. *Journal of Geophysical Research*, **96**, 4869–4889.

Spiesberger, J. L., K. Metzger, and J. A. Furgerson, 1992: Listening for climatic temperature change in the northeast Pacific: 1983–1989. *Journal of the Acoustical Society of America*, **92** (1), 384–396.

Tsay, R. S., 1992: Model checking via parametric bootstraps in time series analysis. *Applied Statistics*, **41**, 1–15.

Woodward, W. and H. Gray, 1993: Global warming and the problem of testing for trend in time series data. *Journal of Climate,* **6**, 953-962.

Woodward, W. and H. Gray, 1994: Selecting a model for detecting the presence of a trend. Tentatively accepted by *Journal of Climate.*