COMPARISON TESTS FOR THE SPECTRA OF DEPENDENT MULTIVARIATE TIME SERIES

René A. Carmona, [1]
Department of Mathematics
University of California at Irvine

Andrea Wang [1]
Department of Statistics
University of California at Berkeley

August 4, 1994

[1]Partially supported by ONR N00014-91-1010

Abstract

Statistical test procedures are proposed to compare the spectra of multivariate time series. We are mostly concerned with the general case of dependent series. The study was motivated by physical oceanography problems for which daily satellite measurements of drifter positions are readily available. Within the mathematical models used for the velocity field at the surface of the ocean, the Lagrangian velocities (time series of the velocities of the drifters along their trajectories) are realizations of stationary stochastic processes having the same one-dimensional marginal distributions. The theory leads to the conjecture that they also should have the same spectra. The tools presented in this paper were intended to test this hypothesis on real data and on data obtained from numerical simulations of transport properties in a two dimensional incompressible Gaussian velocity field.

1 Introduction and Notations

We first describe the notations and the mathematical assumptions which are in force throughout the present study. We consider two d-dimensional finite time series $\{x^{(1)}(t);\ t=1,\cdots,n\}$ and $\{x^{(2)}(t);\ t=1,\cdots,n\}$. We shall use the notations $x^{(i)}(t)=[x_1^{(i)}(t),\cdots,x_d^{(i)}(t)]^t$ to emphasize the components in a standard vector notation, the superscript ^{t} being used for the vector transpose. We want to compare the statistics of the two series, and especially their spectra. It is mathematically convenient to assume that the time series are measurements from two $\mathbb{R}^d$-valued stochastic processes $\{X^{(1)}(t);\ t\in\mathbb{Z}\}$ and $\{X^{(2)}(t);\ t\in\mathbb{Z}\}$ which in turn are viewed as the marginals of a $2d$-dimensional $\mathbb{R}^{2d}$-valued stochastic process $\{X(t);\ t\in\mathbb{Z}\}$. We assume that this process is of order 2, that it is stationary (at least in the weak sense) and that its covariance is integrable. This guarantees the existence of a spectral density which we denote by $f(\omega)$. In fact:

$$f(\omega)=\begin{bmatrix} f_{1,1}(\omega) & f_{1,2}(\omega) \\ f_{2,1}(\omega) & f_{2,2}(\omega) \end{bmatrix}$$

where $f_{1,1}(\omega)$ is the $d\times d$ spectral density matrix of the process $\{X^{(1)}(t)\}$, $f_{2,2}(\omega)$ is the $d\times d$ spectral density matrix of $\{X^{(2)}(t)\}$ and $f_{1,2}(\omega)=f_{2,1}(\omega)^*$ is the $d\times d$ spectral coherence matrix between the processes $\{X^{(1)}(t)\}$ and $\{X^{(2)}(t)\}$. We use the notation * to denote the complex adjoint, i.e. transpose of the matrix whose entries are the complex conjugates of the entries of the original matrix. We consider the problem of the test of the hypothesis:

$$H_0: \qquad f_{1,1}(\omega)\equiv f_{2,2}(\omega)$$

against a general alternative. This problem has been considered in the univariate case $d=1$ when the time series are independent (in which case $f_{1,2}(\omega)\equiv f_{2,1}(\omega)\equiv 0$). See for example [4]. We are interested in the general case of (possibly) dependent multivariate time series. Both generalizations create mathematical difficulties which prevent us from extending ideas from the univariate case. In the case of independent time series, we were able to give an approximate test for which one can determine rigorously the asymptotic distribution of the test statistic. We describe this test procedure in Section 4. Since we are mostly interested in real and simulated data for which the independence assumption is not realistic, we consider the general case of dependent time series in Section 5. We derive another approximate test but unfortunately, we cannot determine rigorously the asymptotic distribution of the test statistic. We rely on Monte Carlo simulations (parametric bootstrap to be specific), to determine the approximate p-values of the tests. Numerical results are given in Section 6.

There are many practical situations in which the comparison of spectra is of importance. Some examples are mentioned in [4] and [5]. Our motivation comes from the analysis of the Lagrangian velocities of drifters at the surface of the ocean. For the sake of simplicity we regard these drifters as passive drifters floating at the surface of a 2-dimensional incompressible fluid. The positions of these drifters are recorded at regular time intervals (say every day), and their velocities are estimated from these measurements. We shall illustrate the properties of the test procedures on real data collected during 1991 and 1992 in the South Atlantic and Indian Oceans. Figure 1 gives the trajectories of the drifters in question.

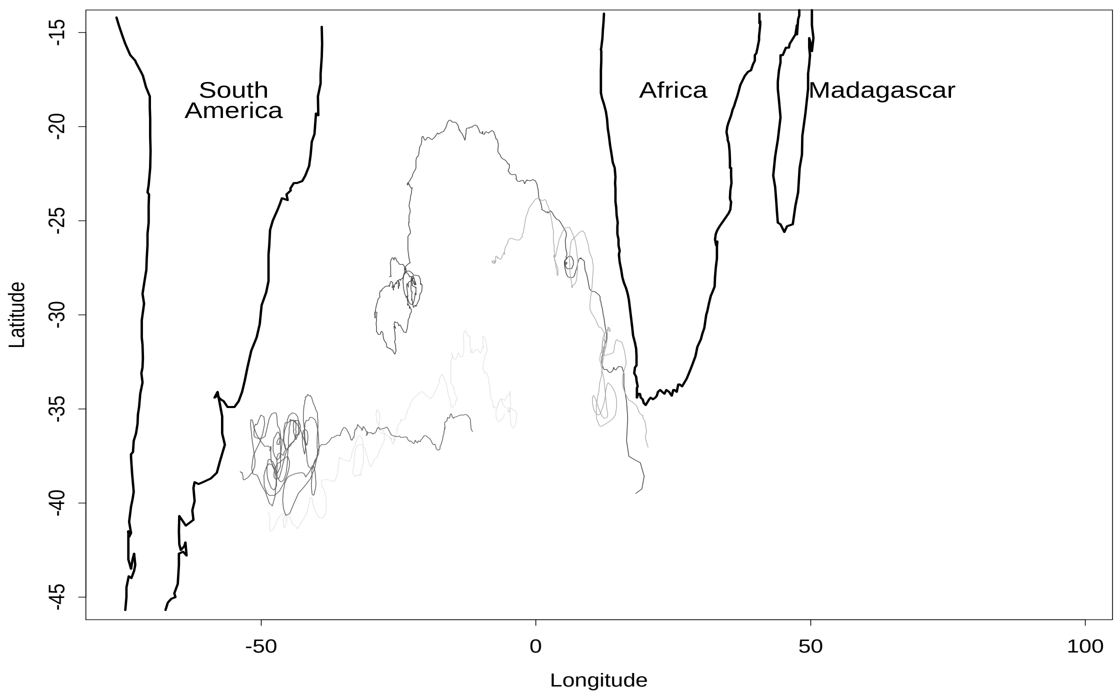

Figure 1: Plot of the daily positions of 4 of the drifters we use in this paper for the purpose of illustration.

There is a theoretical model for which the answer to the equality of spectra is known. We review the basics of this model in Section 2 where we also explain how numerical simulations of this models can be used to evaluate the power of our statistical procedures.

2 Motion in a Gaussian Velocity Field

This section is devoted to a short discussion of a standard mathematical model in fluid mechanics. Our interest in such a model is that, according to the theoretical results recalled below, the individual Lagrangian velocities have the same spectra. Consequently, numerical simulations should provide a testbed for our statistical procedures.

We consider a Gaussian random field $\{\vec{v}(t,x);\ t \geq 0, x \in \mathbb{R}^d\}$. $\vec{v}(t,x)$ represents the velocity field at time $t \geq 0$ at the location $x \in \mathbb{R}^d$. For this reason, $\vec{v}(t,x)$ is a d-dimensional vector ($d = 2$ in the application we have in mind). The trajectory $\{x(t)\}_t$ of a particle dropped at x_0 at time $t = 0$ is given by the solution of the ordinary differential equation:

$$dx(t) = \vec{v}(t, x(t))dt, \qquad\qquad x(0) = x_0.$$

The d-dimensional multivariate time series $\{\vec{v}(t, x(t));\ t \geq 0\}$ gives the velocity along the trajectory: it is called the Lagrangian velocity (as opposed to the Eulerian velocity $\{\vec{v}(t, x);\ t \geq 0, x \in \mathbb{R}^d\}$). An old result of Lumley says that, when the velocity field is homogeneous (in the sense that its distribution is invariant under the shifts of the space variable x) and incompressible (i.e. divergence free), then for each time t and initial position x_0, the random vector $\vec{v}(t, x(t))$ has the same distribution as $\vec{v}(t, 0)$. This result was extended in [6] by Port and Stone (see also [9]) who showed that if the velocity field is also stationary (i.e. its distribution is invariant under the time shifts) then all the time series $\{\vec{v}(t, x(t));\ t \geq 0\}$ are stationary.

So, in the framework of this model, two different initial positions x_0 and x'_0 give two trajectories $\{x(t) :\ t \geq 0\}$ and $\{x'(t) :\ t \geq 0\}$ which lead to two stationary d-dimensional multivariate time series $\{\vec{v}(t, x(t));\ t \geq 0\}$ and $\{\vec{v}(t, x'(t));\ t \geq 0\}$ which have the same one-dimensional marginal distributions in the sense that $\vec{v}(t, x(t))$ and $\vec{v}(t, x'(t))$ have the same distribution for each fixed time t.

As explained earlier, the following question is the subject of the present study: do the whole time series have the same distributions and in particular, do they have the same spectra? The theoretical results which we just recalled say that, if the divergence free Gaussian model applies for the velocity field, then the answer should be yes. Notice that, as in the case of real life drifter Lagrangian velocity time series, the time series from the mathematical model are multivariate and (strongly) dependent since they are integral curves of the same realization of the random velocity field.

Massively parallel simulations of motions in a Gaussian velocity field are reported in [3]. These simulations produce large numbers of bivariate trajectories $\{x(t)\}$ computed from the simulations of a Eulerian velocity field $\{\vec{v}(t, x);\ t \geq 0, x \in \mathbb{R}^d\}$ with a Kolmogorov spectrum. Figure 2 shows the trajectories of 4 of the simulated data produced by this simulation program. Since the purpose of this note is to derive tests for equality of spectra and since the above theoretical results show that any two Lagrangian velocities simulated from this model should indeed have the same spectra, we did not try to make our the results of our simulations look like real drifter trajectories. Some of these plots may not look very realistic because the input parameters of the program were chosen at random.
This feature is irrelevant to our use of these trajectories (see Section 6 below) to illustrate the statistical test procedures which we designed.

3 Set Up from the Spectral Theory of Time Series

We now return to the notations defined in the introduction. We use the standard periodogram to estimate the spectral density matrices. For each real number λ we define the quantity $d(\lambda)$ by:

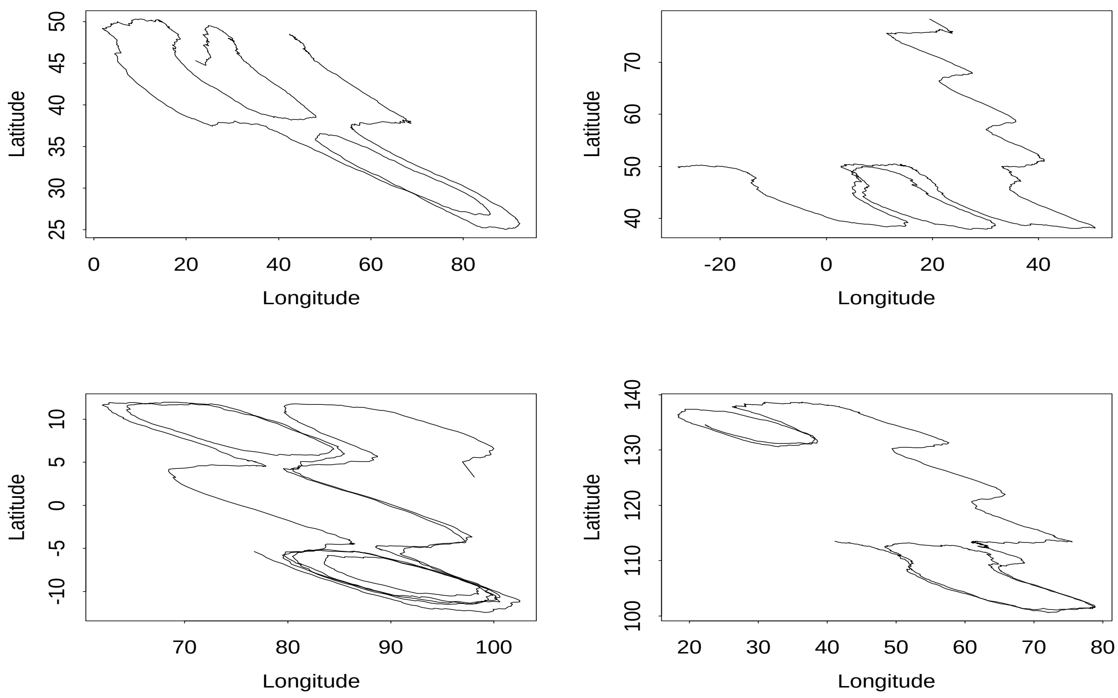

Figure 2: Plot of the trajectories of 4 of the simulation drifter data we use in this paper for the purpose of illustration.

$$d(\lambda) = \sum_{t=1}^{n} e^{i\lambda t} X(t)$$

and we denote by $d_1(\lambda)$ (resp. $d_2(\lambda)$) the d-dimensional vector formed by the first (resp. last) d components of $d(\lambda)$.

Without loss of generality we assume that the number of observations is a multiple of K, say $n = NK$ (we choose $K = 10$ in the applications described below), and we group them K by K to get estimates at N fundamental frequencies ω_j. Let us denote by $\{\lambda_{j,k}; k = 1, \cdots, K\}$ the set of K frequencies which we choose to converge to ω_j. In most applications the frequencies ω_j at which we want to estimate the spectra are regularly spaced and so are the $\lambda_{j,k}$. We consider the quantity $\hat{f}(\omega_j)$ defined by:

$$\hat{f}(\omega_j) = \frac{1}{K} \sum_{k=1}^{K} d(\lambda_{j,k}) d(\lambda_{j,k})^*$$

as an estimate of $f(\omega_j)$. $\hat{f}(\omega_j)$ is a $2d \times 2d$ complex matrix. In fact, it is a standard result from the spectral theory of multivariate time series that, asymptotically as $n \to \infty$ (while K remains fixed):

$$\hat{f}(\omega_1), \cdots\cdots, \hat{f}(\omega_N) \tag{1}$$

become independent with the complex Wishart distributions $W_d^C(K, f(\omega_1)), \cdots, W_d^C(K, f(\omega_N))$ with K degrees of freedom and the respective covariance matrices $f(\omega_1), \cdots, f(\omega_N)$. See for example [1]. Notice that the same construction starting from $d_1(\lambda)$ (resp. $d_2(\lambda)$) instead of $d(\lambda)$ shows that (in the same asymptotic regime):

$$\hat{f}_{1,1}(\omega_1), \cdots\cdots, \hat{f}_{1,1}(\omega_N) \tag{2}$$

and:

$$\hat{f}_{2,2}(\omega_1), \cdots\cdots, \hat{f}_{2,2}(\omega_N) \tag{3}$$

are realizations of independent complex Wishart distributions:

$$W_d^C(K, f_{1,1}(\omega_1)), \cdots\cdots, W_d^C(K, f_{1,1}(\omega_N))$$

and:

$$W_d^C(K, f_{2,2}(\omega_1)), \cdots\cdots, W_d^C(K, f_{2,2}(\omega_N))$$

respectively. The $d \times d$ lower left corner of the matrix $\hat{f}(\omega_j)$ will be denoted by $\hat{f}_{2,1}(\omega_j)$. It should be viewed as an estimate of the coherence $f_{2,1}(\omega_j)$ of the two time series at the frequency ω_j.

4 A Comparison Test in the Independent Case

We now consider the particular case of independent time series. Under this assumption $f_{2,1}(\omega) \equiv f_{1,2}(\omega) \equiv 0$ and the spectral density matrix $f(\omega)$ is block diagonal. This independence assumption is often realistic and the test procedures which we derive in this section are of independent interest. Under this assumption, the estimates $\hat{f}_{1,1}$ and $\hat{f}_{2,2}$ are constructed from independent samples. In particular the $d \times d$ matrices:

$$M_1(\omega_j) = f_{1,1}(\omega_j)^{-1/2}\hat{f}_{1,1}(\omega_j)f_{1,1}(\omega_j)^{-1/2}, \qquad j = 1, \cdots, N$$

and:

$$M_2(\omega_j) = f_{2,2}(\omega_j)^{-1/2}\hat{f}_{2,2}(\omega_j)f_{1,1}(\omega_j)^{-1/2}, \qquad j = 1, \cdots, N$$

form independent samples of independent matrices having the standard complex Wishart distribution $W_d^C(K, I)$ with K degrees of freedom. Notice that, under the null hypothesis $f_{1,1}(\omega_j) = f_{2,2}(\omega_j)$ and the distribution of the random number:

$$T(\omega_j) = \mathrm{trace}[M_1(\omega_j)M_2(\omega_j)^{-1}] \tag{4}$$

is free (i.e. independent of the particular spectra of the individual time series). This distribution, say $F_{d,K}$, is also independent of the frequency ω_j. In fact, $F_{d,K}$ is the distribution of the trace of a random matrix $M_1 M_2^{-1}$ provided that M_1 and M_2 are independent $d \times d$ matrices with the standard Wishart distribution $W_d^C(K, I)$. It is not easy to compute analytically the density of the distribution $F_{d,K}$, but it is very easy to determine its percentiles by Monte Carlo techniques. The value of T can be rewritten in the form:

$$\begin{aligned}
T(\omega_j) &= \text{trace}[M_1(\omega_j) M_2(\omega_j)^{-1}] \\
&= \text{trace}[f_0(\omega_j)^{-1/2} \hat{f}_{1,1}(\omega_j) \hat{f}_{2,2}(\omega_j)^{-1} f_0(\omega_j)^{1/2}] \\
&= \text{trace}[\hat{f}_{1,1}(\omega_j) \hat{f}_{2,2}(\omega_j)^{-1}]
\end{aligned}$$

provided we set $f_0(\omega_j)$ for the common value of the spectral matrices $f_{1,1}(\omega_j)$ and $f_{2,2}(\omega_j)$. This computation shows that the value of T can be computed without any knowledge of the unknown spectral matrices $f_{1,1}(\omega_j)$ and $f_{2,2}(\omega_j)$. Consequently, starting from a sample of size n from the stochastic process $\{X(t)\}$, we can compute

$$T(\omega_1), \cdots\cdots, T(\omega_N)$$

and the latter is an independent sample of size $N = n/K$ from the distribution $F_{d,K}$. Consequently, one can compute the one sample Kolmogorov-Smirnov distance statistic, say T_{KS}, from the empirical distribution of this sample to the theoretical distribution $F_{d,K}$ and determine the p-value of the test from the tables.

Remark 4.1 *It is possible to base the test on the one-sample Kolmogorov-Smirnov distance statistic D_{KS} for the sample formed by the random numbers:*

$$D(\omega_j) = det[M_1(\omega_j) M_2(\omega_j)^{-1}] \tag{5}$$

computed from the determinant instead of the trace. The same considerations apply and a test can be based on the values of this statistic. We do not report numerical results using this test procedure because they were not superior to the ones obtained from the T_{KS} statistic and most importantly, because the computations are much longer: computing the trace of a matrix is faster than computing its determinant.

5 A Test in the General Case

The tests derived in the preceding section are not appropriate for the analysis of Lagrangian spectra, whether they come from simulations of time evolutions in a random velocity field or real life drifter data. Indeed the independence assumption is not satisfied in these application and the results are poor. In order to build the dependence structure into the test procedures, we propose some modifications of the main ideas used so far.

Even though we do not assume any longer that $f_{2,1}(\omega) \equiv 0$, we nevertheless restrict ourselves to coherence structures which are constant with respect to the frequency variable ω. In other words our model includes the assumption of the existence of a complex $d \times d$ matrix F such that:

$$f_{2,2}^{-1/2}(\omega) f_{2,1}(\omega) f_{1,1}(\omega)^{-1/2} \equiv F \tag{6}$$

for all the frequencies ω. We suspect (but we were not able to prove) that this property is characteristic of linear models. Figure 3 below shows that this assumption is reasonable in the case of the Lagrangian velocities obtained from the drifter data considered in this study.

We propose to use the same statistic T_{KS} as before. But we need to be more careful in the computation of the p-value of the test. The first task is to estimate the spectral coherence F. We choose to use:

$$\hat{F} = \frac{1}{N} \sum_{j=1}^{N} f_{2,2}^{-1/2}(\omega_j) f_{2,1}(\omega_j) f_{1,1}(\omega_j)^{-1/2}.$$

Then we generate B (we used $B = 200$ in the numerical experiments which we report in the next section) independent samples $\{X^{(b)}(t);\ 1 \leq t \leq n\}$ for $b = 1, \cdots, B$ of a mean zero stationary Gaussian time series $\{X(t);\ 1 \leq t \leq n\}$ with spectral density matrix:

$$f(\omega) = \begin{bmatrix} I_d & F^* \\ F & I_d \end{bmatrix}$$

where we used the standard notation I_d for the $d \times d$ identity matrix. We compute the T_{KS} statistic for each of the B time series and form a histogram. This parametric bootstrap can then be used to compute the approximate p-value of the test based on the observed value of T_{KS}.

Remark 5.1 *The p-value is computed from Monte Carlo simulations of Gaussian time series. The Gaussian assumption can be regarded as restrictive at first, but is fully justified in the applications we have in mind. Indeed, it is well known that, at least for turbulent flows, the Lagrangian velocities series can be reasonably assumed to be normal. This assumption was first proposed by Kolmogorov in the case of high Reynolds numbers and it has been checked experimentally in many situations. Moreover, in the mathematical analysis of the transport properties of the Gaussian velocity fields, the result of Lumley which we recalled earlier states that the one time marginal distributions of the Lagrangian velocities are normal. Of course, this does not implies that the whole time series are normal, but it makes it reasonable to make such an assumption for a parametric bootstrap.*

Remark 5.2 *As before one can choose to use the determinant instead of the trace in the definition of the test statistic and the same conclusions apply as well.*

6 Numerical Results

Figure 3 is a matrix of plots. The (j, j')-th plot gives the values of the modulus of the estimate of the coherence of the j-th and j'-th components of $X(t) = [X_1(t), X_2(t)]^t$ for a specific choice of Lagrangian velocities $X_1(t)$ and $X_2(t)$ computed from a typical couple of drifters from our data set. In particular, the 4×4 upper left corner gives the plots of the entries of the estimate $\hat{f}_{1,1}(\omega_j)$ of the spectral matrix of the first Lagrangian velocity, the 4×4 lower right corner shows the entries of the estimate $\hat{f}_{2,2}(\omega_j)$ of the spectral matrix of the second Lagrangian velocity while the lower left 4×4 corner gives the plots of the modulus of the entries of the estimate $\hat{f}_{2,1}(\omega_j)$ of the spectral coherence between the two Lagrangian velocities. All the plots are on the same scale. One can see that the coherence is significantly different from zero. At the same time one can appreciate the validity of the assumption that this coherence is constant over the frequencies.

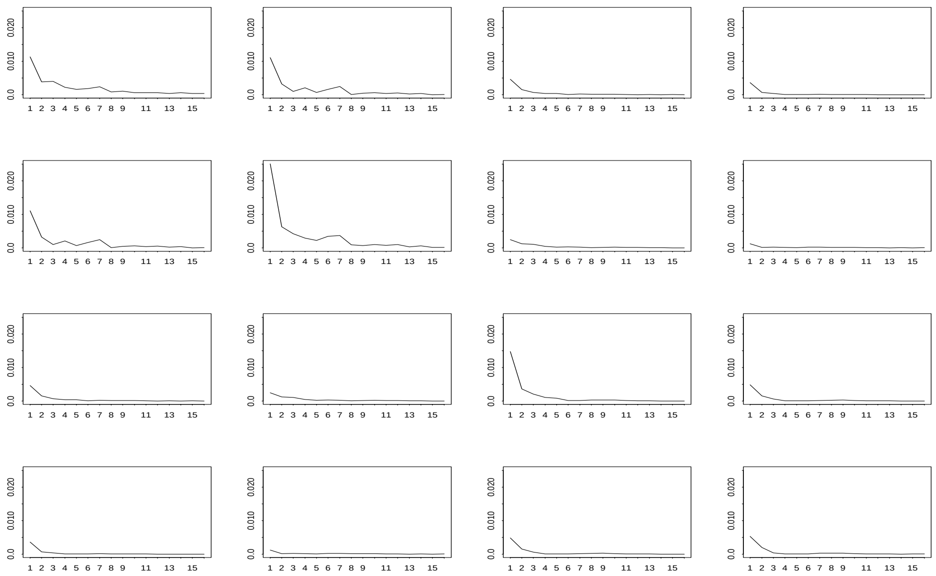

Figure 3: Plot of the modulus of the entries of the 4×4 matrix $\hat{f}(\omega_j)$ versus the frequency ω_j, for a typical couple of drifters. This plot was produced by the commands "`f <- spec.matrix(v80,v91)`" and "`plot.spec(f)`" described in the appendix.

6.1 Simulated Power

The power of the test based on the statistic T_{KS} was computed by Monte Carlo simulation for a certain number of alternatives. Since we suspect that our model is exact for linear processes, we computed the power of the tests (based on both statistics T_{KS} and D_{KS}) when the distribution of the processes were given by bivariate moving average processes ($MA_2(2)$ for short)of order 1 and 2.

Here is a typical example of the results one gets: We consider a 4 dimensional $MA_4(2)$ processes of the form:

$$X(t) = \epsilon(t) + A(1)\epsilon(t-1) + A(2)\epsilon(t-2)$$

for a $N(0, I_2)$ bivariate white noise sequence $\{\epsilon(t)\}_t$ and for the coefficient matrices:

$$A(1) = \begin{bmatrix} 1 & 0 & a & 0 \\ 0 & 1 & 0 & a \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad A(2) = \begin{bmatrix} 1 & 0 & b & 0 \\ 0 & 1 & 0 & b \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

The numbers a and b are parameters which we chave the freedom to choose. The following tables give, for each of the values of the parameters a and b, the probabilities (computed empirically from 50 runs) that the null hypothesis is accepted. In the present situation, the null hypothesis states that the bivariate time series formed by the first 2 components of the 4 dimensional time series $\{X(t)\}_t$ has the same spectrum as the bivariate time series formed by the last 2 components. Notice that these two bivariate time series are dependent.

	b=0.0	b=.2	b=,4	b=.6	b=.8	b=1.0
a=0.0	0	0	0	0	.1	.9
a=.2	0	0	0	0	0	.8
a=.4	0	0	0	0	0	0
a=.6	0	0	0	0	0	0
a=.8	0	0	0	0	0	0
a=1.0	1.0	.7	.1	0	0	0

Results for the test based on the statistic T_{KS} an the independence assumption.

	b=0.0	b=.2	b=.4	b=.6	b=.8	b=1.0
a=0.0	0	0	0	.08	.26	.6
a=.2	0	0	0	.02	.14	.6
a=.4	.02	0	0	0	0	.28
a=.6	.04	0	0	0	0	.1
a=.8	.18	.08	0	0	0	.01
a=1.0	.62	.6	.34	.18	.1	1.0

Results for the test based on the statistic T_{KS} computed without the independence assumption.

Notice that two bivariate times series have the same spectra in the case $a = b = 1$ but this was not picked up by the test based on the independence of the series.

Here is another example which shows that the general test should be preferred to the test based on the independence assumption. We use the same notation as above and the coefficient matrices:

$$A(1) = \begin{bmatrix} 1 & 0 & a & 0 \\ 0 & 1 & 0 & a \\ a & 0 & 1 & 0 \\ 0 & a & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad A(2) = \begin{bmatrix} 1 & 0 & b & 0 \\ 0 & 1 & 0 & b \\ b & 0 & 1 & 0 \\ 0 & b & 0 & 1 \end{bmatrix}.$$

In this case, the two bivariate time series always have the same spectra. They are independent when the numbers a and b are zero and the dependence increases with the values of these parameters. The results of runs made in the same conditions as above are contained in the following tables.

	b=0.0	b=.2	b=,4	b=.6	b=.8	b=1.0
a=0.0	.94	.98	.92	.96	.94	.92
a=.2	.92	.94	.96	.94	.98	..96
a=.4	.98	.96	.94	.98	.96	.98
a=.6	.76	.86	.76	.8	.84	.84
a=.8	.06	.08	.1	.18	.08	.08
a=1.0	0	0	0	0	0	0

Results for the test based on the statistic T_{KS} an the independence assumption.

	b=0.0	b=.2	b=.4	b=.6	b=.8	b=1.0
a=0.0	.98	1.0	.96	1	1	.98
a=.2	.94	.98	1	.96	.96	1
a=.4	.98	.98	.98	.96	.96	.96
a=.6	.98	.98	.98	.94	.96	1
a=.8	1	.9	.94	.96	.98	.94
a=1.0	1	1	1	1	1	1

Results for the test based on the statistic T_{KS} computed without the independence assumption.

One immediately sees that the dependent test performs better, especially when there is a strong dependence between the two time series. Indeed, this test continues to accept the null hypothesis while the test based on the independence of the time series see its level deteriorate in a drastic manner.

Remark 6.1 *We also investigated the influence of the window parameter K on the power of the tests. Our numerical simulations show that small values of K (by which we mean $3 \leq K \leq 6$ give a smaller power of detection than larger values (typically $8 \leq K \leq 10$).*

Acknowledgments:
We would like to thank Don Olson from the Marine Station in Miami for providing us with the drifter data used in the analysis.

Appendix

A Drifter Data Documentation

The drifter dataset consists of 19 files, and each file consists of following fields: days, latitude (degree), longitude (degree), surface temperature (degree Celsius), ease-west velocity (m/s), and north-south velocity (m/s). All data were collected from the South Atlantic and Indian Oceans. The file name begins with a two-digit number followed by ".vel", e.g., 80.vel. The two-digit numbers range from 80 to 99, with 94 skipped.

B S Function Documentation

This appendix contains documentation for the S functions developed during the study. It is organized into three categories: data structure/input, data display, and test procedures.

B.1 Data Structure/Input

When the data file is in table format like ours, it can be read in as a data frame via the S command `read.table`. The columns of the resulting data frame are the corresponding fields in the file. Appropriate names should be assigned to different columns for easy access. If the dataset contains more than one file, all the data frames can be combined into a list.

Although the following two functions were written specifically for the dataset we have in hand, users can use them as an example to read in and organize their own data files. Our functions require that each data frame at least consists of two fields named "latitude" and "longitude".

make.vel.frame

USAGE

```
make.vel.frame()
```

DESCRIPTION
Combines 19 data frames, which are the output of read.vel, into one data frame. This function should be called after the drifter data files have been read into individual data frames.

VALUE
The output is a data frame with seven fields (columns). They are called tag (the two-digit number associated with each data file), days, longitude, latitude, temperature, v1 and v2. 19 data frames are combined in increasing order of the numbers associated with the file names.

read.vel

USAGE

```
read.vel()
```

DESCRIPTION

Reads drifter data files into data frames and combines all 19 data frames into a list called "drifter."

VALUE

The output consists of 19 data frames with names starting with "v" followed by the two-digit number corresponding to the file name and a list named "drifter". Each data frame corresponds to one drifter file, and there are six fields (columns) in each each frame. They are called days, longitude, latitude, temperature, v1 and v2.

B.2 Data Display

drifter.world

USAGE

```
drifter.world()
```

DESCRIPTION

Plots part of the world that all the drifters travel, which include the South Atlantic and Indian Oceans.

plot.drifter

USAGE

```
plot.drifter(drifter.list, scale=F)
```

DESCRIPTION

Plots a list of drifters. If the option "scale" is TRUE, all the drifters are plotted with the same scale. The graphic window should be formatted according to the number of drifters before plotting. For example, use the S command "par(mfrow=c(4,5))" to divide the graphic window into a 4 by 5 rectangular array of plots, so that all 19 drifters can be viewed at one time.

ARGUMENTS

drifter.list: a list of data frames where each data frame corresponds to one drifter data and has at least two fields, longitude and latitude. It is like the output of read.vel.

scale: logical flag: if TRUE, all the drifters are plotted with the same scale.

plot.traj

USAGE

```
plot.traj(..., inter=F)
```

DESCRIPTION: Plots the trajectories of the given drifters and superimposes the parts of the drifters that intersect in time if the option "inter" is TRUE.

ARGUMENTS

...: any number of drifter data frames to be plotted. Each data frame consists of at least two fields named "latitude" and "longitude".

inter: logical flag: if TRUE, the parts of the drifters that intersect in time are superimposed on the trajectories. This option works properly only if each data frame has another field named "days" that record the time of the drifter data.

B.3 Test Procedures

dep.spec.test

USAGE

```
dep.spec.test(f, g, method=''determinant'')
```

DESCRIPTION: Tests if either the trace or the determinant sample computed from the spectral density matrix, f, has the same distribution as the corresponding sample computed from g. It returns the p-value of the Kolmogorov-Smirnov two-sample test.

ARGUMENTS

f: spectral density matrix estimate of the drifters.

g: spectral density matrix estimate of the simulated data with constant coherence.

method: a character string indicating the type of statistic used in the test, "determinant" or "trace".

det.stat

USAGE

```
det.stat(f)
```

DESCRIPTION: Computes the determinant statistic of the spectral density matrix f.

ARGUMENTS

f: spectral density matrix estimate of the drifters.

est.spectra.matrix

USAGE

```
est.spectral.matrix(x1, y1, x2, y2, K=10)
```

DESCRIPTION: Computes the 4 by 4 spectral density matrix estimate for the given vector-valued series.

ARGUMENTS

x1: a vector of numerical numbers.
y1: a vector of numerical numbers.
x2: a vector of numerical numbers.
y2: a vector of numerical numbers.
K: a positive number indicating the number of points to be grouped for the estimate at one frequency.

indep.spec.test

USAGE

```
indep.spec.test(f, K=10, method=''determinant'')
```

DESCRIPTION: Tests if either the trace or the determinant sample computed from f has the same distribution as the corresponding sample generated from Monte Carlo simulation. It returns the p-value of the Kolmogorov-Smirnov two-sample test.

ARGUMENTS

f: spectral density matrix estimate of the drifters.
K: the same positive number used when f is computed.
method: a character string indicating the type of statistic used in the test, "determinant" or "trace".

kstwo

USAGE

```
kstwo(sample1, sample2)
```

DESCRIPTION: Performs Kolmogorov-Smirnov two-sample test and returns the p-value.

ARGUMENTS

sample1: a numerical vector.
sample2: a numerical vector with the same length of sample1.

ma1.sim

USAGE

```
ma1.sim(N, a1=diag(4))
```

DESCRIPTION: Generates an N by 4 MA(1) series with coefficient matrix a_1.

ARGUMENTS

N: a positive number indicating the length of MA(1) series to be simulated.
a1: a 4 by 4 coefficient matrix for the MA(1) series.

ma2.sim

USAGE

```
ma2.sim(N, a1=diag(4), a2=diag(4))
```

DESCRIPTION: Generates an N by 4 MA(2) series with coefficient matrices a_1 and a_2.

ARGUMENTS

N: a positive number indicating the length of the MA(2) series to be simulated.
a1: a 4 by 4 coefficient matrix for the MA(2) series.
a2: a 4 by 4 coefficient matrix for the MA(2) series.

myqq

USAGE

```
myqq(sample1, sample2, xlab, ylab)
```

DESCRIPTION: Compares the distributions of two samples by `qqplot`. "xlab" and "ylab" are optional arguments.

ARGUMENTS

sample1: a numerical vector.
sample2: a numerical vector.
xlab: a character string.
ylab: a character string.

plot.coh

USAGE

```
plot.coh(f)
```

DESCRIPTION: Plots the coherency, both the real part and the imaginary part, computed from the spectral density matrix estimate, f, at all calculated frequencies.

ARGUMENTS

f: spectral density matrix estimate of the drifters.

plot.spec

USAGE

```
plot.spec(f)
```

DESCRIPTION: Plots the modulus of the 4 by 4 spectral density matrix estimate, f, at all calculated frequencies with the same scale in each panel.

ARGUMENTS

f: spectral density matrix estimate of the drifters.

sim.dep

USAGE

```
sim.dep(f, N=500)
```

DESCRIPTION: Simulates N by 4 data with the given spectral density matrix f.

ARGUMENTS

f: spectral density matrix estimate of the drifters.
N: a numerical number indicating the length of the simulated series.

spec.matrix

USAGE

```
spec.matrix(drifter1, drifter2, K=10)
```

DESCRIPTION: Computes the 4 by 4 spectral density matrix estimate for the given drifters.

ARGUMENTS

drifter1: a data frame containing at least two fields, longitude and latitude.
drifter2: a data frame containing at least two fields, longitude and latitude.
K: a positive number indicating the number of points to be grouped for the estimate at one frequency.

trace.stat

USAGE

```
trace.stat(f)
```

DESCRIPTION: Computes the trace statistic of the spectral density matrix f.

ARGUMENTS

f: spectral density matrix estimate of the drifters.

wishart.ratio

USAGE

```
wishart.ratio(N=100, r=2, K=10, method=''determinant'')
```

DESCRIPTION: Computes N trace statistics or determinant statistics of $W_1 W_2^{-1}$, where W_1 and W_2 are independent and have a Wishart distribution of dimension r and degrees of freedom K.

ARGUMENTS

N: a positive number indicating the size of the sample.
r: a positive number indicating the dimension of the Wishart distribution.
K: a positive number indicating the number of points to be grouped for the estimate at one frequency.
method: a character string indicating the type of statistic used in the test, "determinant" or "trace".

wishart.sample

USAGE

```
wishart.sample(N, method=''determinant'')
```

DESCRIPTION: Computes N trace statistics or determinant statistics of $W_1 W_2^{-1}$, where W_1 and W_2 are independent and have a Wishart distribution of dimension 4 and degrees of freedom K, ranging from 2 to 10.

ARGUMENTS

N: a positive number indicating the size of the sample.
method: a character string indicating the type of statistic used in the test, "determinant" or "trace".

References

[1] D.R. Brillinger (1973): The Frequency Analysis of Vector-valued Time Series. *Holt* New York, N.Y.

[2] R. Carmona (1994): Ornstein-Ulhenbeck Stochastic Flows. (in preparation)

[3] R. Carmona and S. Grishin (1994): Massively Parallel Simulations of Motions in a Gaussian Velocity Field. (preprint)

[4] D.S. Coates and P.J. Diggle (1986): Test for Comparing Two Estimated Spectral Densities. *J. of Time Series Analysis,* ***1***, 7-20.

[5] P.J.Diggle and N.I. Fisher (1991): Nonparametric Comparison of Cumulative Periodograms. *Appl. Statist.* ***40***, 423-434.

[6] S.C. Port and C.J. Stone (1976): Random Measures and their Applications to Motion in an Incompressible Fluid. *J. Appl. Prob.* ***13***, 498-506.

[7] M. Rosenblatt (1988): Stationary Sequences and Random Fields. *Birkhaüser*, Boston, MA.

[8] A.M. Yaglom (1987): Correlation Theory of Stationary and Related Random Functions. vol. I: Basic Results. Springer Verlag. New York, N.Y.

[9] C.L. Zirbel (1993): Stochastic Flows: Dispersion of a Mass Distribution and Lagrangian Observations of a Random Field. Ph. D. Princeton Univ.