AN ABSTRACT OF THE THESIS OF

THOMAS DANFORTH BURNETT for the M.S. in STATISTICS
(Name) (Degree) (Major)

Date thesis is presented May 15, 1964

Title PARAMETER ESTIMATION IN A STOCHASTIC MODEL OF OCEAN CURRENTS

Abstract approved [redacted]
(Major professor)

In 1962 a large amount of data was collected in an effort to establish the existence of subsurface ocean currents off the coast of Oregon. The presence of several phenomena, including tides and measurement errors not directly associated with the hypothetical currents, complicates the interpretation of the data and the estimation of current velocities.

This thesis advances a stochastic model to represent the sources of the data, and discusses alternative procedures for the estimation of parameters in the model.

The local motion of a small volume of water is, in the first approximation, assumed a sum of a linear drift current plus a periodic tidal current. The motion of any given particle of water is then a sum of the above mentioned currents plus a random motion due to the characteristic turbulence of the ocean's waters. This

random motion is shown similar to Brownian motion or more specifically the motion of the Wiener process.

From the properties of the Wiener process the joint probability distribution of the observed displacements of the parachute drogue is found to be Normal with a given variance-covariance structure.

By suitable linear transformation of the observed displacements, a new set of independent random variables may be obtained, where the variances due to random motion and measurement inaccuracies are combined in a known manner to form variances of the new set.

A Maximum Likelihood procedure for estimating the mean velocity of the currents is proposed, and shown to coincide with the methods of Least Squares in certain special cases.

PARAMETER ESTIMATION IN A STOCHASTIC MODEL OF OCEAN CURRENTS

by

THOMAS DANFORTH BURNETT

A THESIS

submitted to

OREGON STATE UNIVERSITY

in partial fulfillment of
the requirements for the
degree of

MASTER OF SCIENCE

June 1964

APPROVED:

[redacted]

Associate Professor of Statistics

[redacted]

Chairman of Department of Statistics

[redacted]

Dean of Graduate School

Date thesis is presented May 15, 1964

Typed by Illa W. Atwood

ACKNOWLEDGMENT

The author wishes to express his appreciation to Dr. Donald Guthrie for his help in the preparation of this thesis, and to Dr. June Pattullo for proposing the subject treated therein.

TABLE OF CONTENTS

PARAMETER ESTIMATION IN A STOCHASTIC MODEL OF OCEAN CURRENTS

INTRODUCTION

In recent years an ever increasing emphasis has been placed on the study of the physical properties of the ocean, and great advances have been made in the techniques and procedures used to obtain quantitative information about these physical phenomenon. Parallel to this, there has been considerable advancement in the field of statistics on the theory of Stochastic processes.

Because of the similarity of the mobile nature of the physical ocean to the basis for the theory of Stochastic processes, it seems only natural that this analytical approach be used to obtain a better understanding of the physics of the ocean.

This thesis deals with the problem of estimating the parameters of a model of ocean currents by statistical analysis of data obtained in the study of subsurface currents off the coast of Oregon.

The above mentioned data was obtained in 1962 during a study by Paul M. Maughan of the Oceanography Department, Oregon State University. The purpose of this study was to ascertain the velocity of the currents at various depths by indirect observation using a float type current measuring instrument, the parachute drogue.

The final estimate of the speed and direction of the current at any given depth was made using only the initial and final positions of the parachute drogue and the elapsed time between. Such variability existed in the path followed by the drogue and in the measurements of its position that there is some question as to the confidence with which this estimate of velocity can be utilized.

It is the aim of this thesis to make use of the observed path of the parachute drogue and its variations to obtain estimates of current velocity.

SOURCES OF THE DATA

The experiment conducted by Maughan involved the following equipment and procedure:[1]

Equipment

1. The Oregon State University Oceanography department's research ship, the *R. V. Acona*, equipped with a Loran positioning system as well as shipboard radar.
2. A reference buoy and sufficient line to anchor it to the bottom.
3. An adequate number of parachute drogues (hereafter referred to as drogues) to make possible the measurement of several currents at varying depths simultaneously. The drogues used consisted of three main sections: 1) a surface float or surface drogue equipped with a radar reflector and a blinking light, 2) a 28-foot parachute canopy rigged so that it would develop in the current, and weighted so that it would sink, and 3) a 5/32-inch cable connecting the parachute to the surface float and consequently suspending the parachute at the depth of interest.

---

[1] For a more detailed description of the experiment consult Maughan (7, p. 7-11).

Procedure

1. The reference buoy was anchored to the bottom in the area where current measurements were to be taken, and its position was fixed by Loran.
2. The drogues were launched in the immediate area of the reference buoy which allowed any movement of the drogues to be observed with relation to a semistationary position.
3. At regular intervals (if possible) the location of the drogue with respect to the reference buoy was determined by shipboard radar. Whenever possible this was done with the ship located at the reference buoy. However, in cases when the surface drogue was not visable from this position due to high seas or darkness, it was necessary to leave the reference position and hunt for the drogue. If it was found its position was fixed by Loran.

The obvious principle in this type of procedure is that the current at the depth of the parachute will carry the parachute with it. Consequently, the surface float, being fairly rigidly attached to the parachute, will follow. However it is also obvious that there can be large discrepancies between the measured distance and direction the surface float travels during any given time interval, and the distance and direction the current, at the depth of interest, travels during the same time interval.

These discrepancies stem from two sources: The failure of the parachute to follow the current exactly because of turbulence and friction, and the failure of the reference buoy and the surface float to stay fixed relative to the reference anchor and parachute respectively, due to the action of friction within the water and wind above the surface.

Fortunately, theoretical corrections can be made for the above mentioned discrepancies caused by friction and the wind. These corrections can be made from the velocity of the wind and of the various currents acting at different depths, at the time of the observation, on the surface float and reference buoy and their respective cables. It should be noted that except under abnormal conditions these corrections are small.

The plot of the adjusted positions, however, still appears to have considerable variations from the straight line flow that would be expected when considering the subsurface currents over short distances in the open ocean. Part of this variation can be accounted for by the periodic rotational effect of the tides. However, the magnitude of these effects is not known for the locality of interest, and corrections cannot be made. The remaining variation may be described as random fluctuation associated with the forces in the physical ocean.

A STOCHASTIC MODEL FOR THE POSITION OF THE DROGUE

A random phenomenon that arises through a process which is developing in time in a manner controlled by probabilistic laws is called a stochastic process (9, p. 7). A stochastic process is mathematically defined as a collection of random variables $\{\Phi(t), t \in T\}$.

In the present model the random variables $\{\Phi(t), t \in T\}$ are position vectors of the surface float. The surface float positions of interest are its positions at times $t_o, t_1, t_2, \ldots, t_n$ with $\Phi(t_i)$ being the position at time t_i.

The space in which the positions are located is the two-dimensional space of the surface of the ocean, therefore, the random variable $\underline{\Phi(t)}$ is a two-dimensional vector,

1) $$\underline{\Phi(t)} = \begin{pmatrix} X(t) \\ Y(t) \end{pmatrix}.$$

Let us define random variables

2) $$\Delta x_i = X(t_i) - X(t_{i-1}),$$

and

3) $$\Delta y_i = Y(t_i) - Y(t_{i-1}).$$

These random variables are the increments of change in position in the X and Y directions respectively from time t_{i-1} to t_i.

Let us also define

4) $$\Delta t_i = t_i - t_{i-1}$$

as the time interval associated with the random variables Δx_i and Δy_i.

Hereafter the random variables X(t) and Y(t) will be referred to as positions and the random variables Δx_i and Δy_i will be referred to as displacements.

In the first approximation the local motion of a small volume of water is the sum of a linear current $\underline{D}(t)$, and a periodic current $\underline{T}(t)$.

5) $$\underline{\Phi}(t) = \underline{D}(t) + \underline{T}(t),$$

where $\underline{\Phi}(t)$, $\underline{D}(t)$, and $\underline{T}(t)$ are two-dimensional vectors.

Assuming the current velocity is constant over the range of interest, the linear current, $\underline{D}(t)$, is a function of the current velocity, $\underline{\mu}$ and time;

6) $$\underline{D}(t) = t\,\underline{\mu}$$

with X and Y components,

7) $$D_x(t) = \mu_x t,$$
$$D_y(t) = \mu_y t.$$

The periodic current, T(t), is the sum of periodic effects caused by the attraction of the sun and the moon. Along the Pacific coast of the United States complicated tidal currents are found, representing a combination of several tidal effects corresponding to tides of different periods. In most instances a good approximation of the observed condition is obtained by superimposing the tidal effects of one single semidiurnal period of length 12 lunar hours and one single diurnal period of length 24 lunar hours (10, p. 573).

Defining $f_x^{(1)}(t)$ as the semidiurnal and $f_x^{(2)}(t)$ as the diurnal tidal components in the X direction and $f_y^{(1)}(t)$ and $f_y^{(2)}(t)$ similarly for the Y direction, and $\beta_x^{(1)}$, $\beta_x^{(2)}$, $\beta_y^{(1)}$ and $\beta_y^{(2)}$ as constants of magnitude, $T_x(t)$ and $T_y(t)$ can be expressed as,

8) $$T_x(t) = \beta_x^{(1)} f_x^{(1)}(t) + \beta_x^{(2)} f_x^{(2)}(t),$$

and

9) $$T_y(t) = \beta_y^{(1)} f_y^{(1)}(t) + \beta_y^{(2)} f_y^{(2)}(t).$$

Hence it is concluded that the mean increments of displacement from $t_{i=1}$ to t_i in the X and Y directions are,

10) $$E(\Delta x_i) = \mu_x \Delta t_i + \beta_x^{(1)}[f_x^{(1)}(t_i) - f_x^{(1)}(t_{i-1})] + \beta_x^{(2)}[f_x^{(2)}(t_i) - f_x^{(2)}(t_{i-1})],$$

and

11) $$E(\Delta y_i) = \mu_y \Delta t_i + \beta_y^{(1)}[f_y^{(1)}(t_i) - f_y^{(1)}(t_{i-1})] + \beta_y^{(2)}[f_y^{(2)}(t_i) - f_y^{(2)}(t_{i-1})],$$

respectively, where $E(\cdot)$ equals the expected value of $(\cdot)$.

In the theory of stochastic processes and its applications, a fundamental role is played by the Wiener process in that it provides a model for Brownian motion (9, p. 26-29). The notion of Brownian motion involves the ceaseless motion of a particle immersed in a gas or liquid. This motion is explained by continual bombardments or impacts upon the particle by the action of the force field of the surrounding medium.

Let $B(t)$ denote the displacement (from a starting point) after time t of a particle in Brownian motion. By definition let $B(0) = 0$. The displacement of a particle over a time interval (s, t) which is long compared to the time between impacts, can be regarded as the sum of a large number of small displacements caused by an equal number of impacts. Next, the assumption is made that the motion of the particle is due entirely to these very frequent and irregular impacts or forces from the surrounding medium. Mathematically this can be interpreted as saying the stochastic process, $B(t)$, has independent increments.

By the Central Limit Theorem, whenever a random variable Z is the sum of a large number of independent random variables, all of which have the same distribution with mean μ and finite variance σ^2, the distribution of $Z^* = (Z - EZ)/\sigma$ approaches the normal distribution as n, the sample size, increases without limit (5, p. 168).

It is therefore assumed that the random variable $B(t) - B(s)$ is normally distributed.[1] Furthermore, it can be assumed

[1] A more elegant presentation of necessary and sufficient conditions that $B(t) - B(s)$ be normally distributed is given by Theorem 7.1 of Doob (2, p. 420).

that $B(t+h) - B(s+h)$ is normally distributed for all $h \geq 0$ since it is supposed that the probability law associated with the observed particle displacements over any given time interval (s, t) should depend upon the length of the time interval, $t-s$, for $s < t$, and not upon the time at which the observation was begun. This foregoing assumption describes $B(t)$ as a process of stationary independent increments.

This leads to the definition of a Wiener process. A stochastic process $\{B(t), t \geq 0\}$ is said to be a Wiener process if:

(i) $\{B(t), t \geq 0\}$ has stationary independent increments;

(ii) For every $t > 0$, $B(t)$ is normally distributed;

(iii) For all $t > 0$, $E(B(t)) = 0$;

(iv) $B(0) = 0$.

Since $B(t)$ has independent increments, and $B(0) = 0$, to state the probability law of the stochastic process $B(t)$ it suffices to state the probability law of the increment $B(t) - B(s)$ for $s < t$. $B(t) - B(s)$ has been assumed to be normally distributed, and therefore $B(t)$ is normally distributed and its probability law is determined by its mean and variance. It follows that

12) $$E(B(t) - B(s)) = 0, \qquad s, t \geq 0,$$

and

$$\mathrm{Var}(B(t) - B(s)) = \sigma^2(t - s), \qquad s, t, \geq 0,$$

where σ^2 is some positive constant.

To illustrate how the problem at hand is mathematically similar to the Wiener process, the following approach is taken.

The fundamental premise of the notion of Brownian motion is that ceaseless irregular impacts act upon the particle in motion as a result of the force field of the surrounding medium. It is stated by Sverdrup, Johnson, and Fleming (10, p. 471), that the ocean currents are,

> ...characterised by numerous eddies of different dimensions by which small fluid masses are constantly carried into regions of different velocity. It is this completely irregular type of motion which is called turbulent flow.... The very character of the turbulent flow is such that rapid fluctuations of velocity take place in all localities, and no steady state motion exists if attention is paid to individual particles of the fluid.

The velocity field at any point in the ocean can then be represented as the vector sum of two different velocities: $\bar{v}$ representing the average velocity at the point in question during a long period of time, and v' representing the instantaneous turbulent velocity at that point.

Let a three-dimensional coordinate system be fixed with respect to the current moving at a velocity $\bar{v}$ with the (X, Y) plane parallel to the ocean surface. Consequently the coordinate system will be fixed with respect to the current in the locality of this point. If the parachute of the drogue is, at time $t = 0$, placed at the $(0, 0, 0)$ position of the coordinate system the force field acting on it is the same above mentioned force field caused by the turbulent nature of the water. This will cause the parachute to generally tend to follow the current while wandering in a random manner due to the irregular eddies.

Since the total force field is a result of numerous small eddies of presumed random direction and magnitude, the motion of the parachute with respect to the coordinate system is therefore similar to the motion described by the Wiener process.

It can also be expected that a similar condition of turbulence exists in the vicinity of the reference buoy and the surface drogue. As a result there will be a random irregular motion of the reference buoy with respect to its anchor and of the surface drogue with respect to the parachute.

The forces due to the turbulence of the water acting on the parachute, buoy, and drogue, however, are not the only forces acting on the system. Friction or drag, due to the motion of the reference buoy and drogue cables with respect to the surrounding water, cause variations in the buoy and drogue positions with respect to the actual position of the reference anchor and the parachute. Similarly the effect of the wind can cause the above described variations in buoy and drogue positions. As previously noted, theoretical corrections can be made to compensate for the bias introduced by these effects. Quite naturally, there will be errors in these corrections yet if average wind velocities and drag forces are used in the corrections, there will tend to be both over corrections and under corrections of varying magnitudes, and consequently, somewhat irregular or random errors.

It can therefore be seen that the actual horizontal distance between the initial position of the reference buoy (the position where the drogue was launched) and the water particle on which the coordinate system was fixed can vary considerably from the horizontal distance between the reference buoy and the surface drogue at the time of observation. This variation, a function of the many random variations as described above, is assumed to follow the probability law associated with Brownian motion and more specifically the Wiener process. This probability law is stated thusly: The corrected increments of change in position of the surface drogue with respect to the reference buoy are normally distributed with mean vectors

$$14)\quad E(\Delta x_i) = \mu_x \Delta t_i + \sum_{j=1}^{2} \beta_x^{(j)} [f_x^{(j)}(t_i) - f_x^{(j)}(t_{i-1})],$$

$$15)\quad E(\Delta y_i) = \mu_y \Delta t_i + \sum_{j=1}^{2} \beta_y^{(j)} [f_y^{(j)}(t_i) - f_y^{(j)}(t_{i-1})],$$

and variances

16) $$\mathrm{Var}(\Delta x_i) = \mathrm{Var}(\Delta y_i) = \sigma^2 \Delta t_i ,$$

where σ^2 is some constant inherent in the random fluctuations of the ocean.

INTRODUCTION OF MEASUREMENT ERROR

The following development concerns the joint probability distribution of the observed displacements of the surface drogue associated with the previously defined time intervals Δt_i, $i = 1, 2, \ldots, n$.

Let us associate the X and Y components of the position vector $\underline{\Phi(t)}$ with the E-W and N-S directions respectively where $E(\Delta x_i)$, $E(\Delta y_i)$, $Var(\Delta x_i)$ and $Var(\Delta y_i)$ are as defined in equations [14], [15] and [16].

In order to prevent unnecessary repetition, until otherwise noted the following analysis will concern only the displacements in the X or E-W direction realizing that the displacements in the Y or N-S direction would be treated similarly.

As defined $\Delta x_i = X(t_i) - X(t_{i-1})$ represents the actual E-W horizontal component of displacement and not the observed displacements. Let $\Delta x_i'$ be an observed displacement, or simply the i^{th} observation, and defined by

17) $$\Delta x'_i = [X(t_i) + \epsilon_i] - [X(t_{i-1}) + \epsilon_{i-1}]$$

where ϵ_i is the measurement error associated with fixing the i^{th} position of the surface drogue.

As an approximation, the distribution of the measurement error ϵ is assumed prescribed by the probability law,

18) $$\epsilon \sim NI(0, \tau^2),$$

where τ^2 is some constant inherent in the measurement equipment.

This assumption of normality is based on the supposition that the measurement error is the sum of many small random errors associated with the mechanical and electrical fluctuations of the measuring instrument. Then by the Central Limit Theorem the distribution of the measurement error approached the Normal distribution.

Assuming that the displacements and the measurement errors are statistically independent it is concluded that the joint distribution of the observations, Δx_i, $i = 1, 2, \ldots, n$, is Multivariate Normal with

19) $$E(\Delta x'_i) = \mu_x(\Delta t_i) + \sum_{j=1}^{2} \beta_x^{(j)} [f_x^{(j)}(t_i) - f_x^{(j)}(t_{i-1})],$$

and

20) $$\operatorname{cov}(\Delta x'_i, \Delta x'_j) = \begin{cases} \sigma^2 \Delta t_i + 2\tau^2, & i = j, \\ -\tau^2, & i = j \pm 1, \\ 0, & \text{otherwise.} \end{cases}$$

Using matrix notation the model can be expressed as,

21) $$\mathbb{\Delta}_x = \mathbb{X}\,\mathbb{\beta} + \mathbb{W}$$

where

$\mathbb{\Delta}_x$ is nx1, and is a column vector of observations,

$\mathbb{X}$ is nxp, and is a matrix of known constants,

$\mathbb{\beta}$ is px1, and is a column vector of unknown parameters, and

$\mathbb{W}$ is nx1, and is the observation error vector.

From the above it follows that

22) $$\mathbb{W} \sim N(0, \mathbb{\Sigma}),$$

where

23) $$\mathbb{\Sigma} = \sigma^2 \mathbb{T} + \tau^2 \mathbb{H}$$

with $\mathbb{T}$ being a diagonal matrix of Δt_i's and $\mathbb{H}$ being a matrix with components

(24) $$h_{ij} = 2\delta_{ij} - (\delta_{i,j-1} + \delta_{i,j+1}), \quad i, j = 1, 2, \ldots, n,$$

where δ_{ij} is the kronecker delta.

REDUCTION TO CANONICAL FORM

From the preceding discussion we see that the observations are correlated and that their variances are the sum of two variances $\Delta t_i \sigma^2$ and $2\tau^2$. The covariance of each displacement is $-\tau^2$ with its adjacent displacement and zero with all others. The existing covariances between observations can be removed, and the statistical problem reduced to a canonical form by use of the following theorem proved in (1, p. 341).

Theorem 1

Given $\mathbb{B}$ a positive semidefinite matrix and $\mathbb{A}$ a positive definite matrix there exists a nonsingular matrix $\mathbb{F}$ such that,

25) $$\mathbb{F}' \mathbb{B} \mathbb{F} = \mathbb{D},$$

and

26) $$\mathbb{F}' \mathbb{A} \mathbb{F} = \mathbb{I},$$

where $\mathbb{D}$ is a diagonal matrix of the roots of the characteristic equation

27) $$|\mathbb{B} - \lambda \mathbb{A}| = 0.$$

It has been shown in (3, p. 67) that the roots of the n^{th} degree equation,

28)
$$\begin{vmatrix} -\lambda & -1 & 0 & \dots & 0 \\ -1 & -\lambda & -1 & \dots & 0 \\ 0 & -1 & -\lambda & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -1 & \lambda \end{vmatrix} = 0$$

are

29)
$$\lambda_\nu = -2\cos\frac{\nu\pi}{n+1}, \quad \nu = 1, 2, \dots, n.$$

It follows directly that the n^{th} degree equation

30)
$$|\mathbb{H} - \lambda\mathbb{I}| = \begin{vmatrix} 2-\lambda & -1 & 0 & \dots & 0 \\ -1 & 2-\lambda & -1 & \dots & 0 \\ 0 & -1 & 2-\lambda & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -1 & 2-\lambda \end{vmatrix} = 0$$

has roots

31)
$$\lambda_\nu = 2 - 2\cos\frac{\nu\pi}{n+1} \quad \nu = 1, 2, \dots, n,$$

and that $\mathbb{H}$ is positive definite.

Therefore, since $\mathbb{T}$ is diagonal with positive values there exists a linear transformation which will produce a new set of random variables which are uncorrelated. Letting these transformed data be designated in matrix notation by the nx1 vector Δ_x^*, the transformed model can be expressed as,

32)
$$\Delta_x^* = \mathbb{X}^*\beta^* + \mathbb{W}^*,$$

where

33) $$W^{*} \quad NI(0, \Sigma^{*}),$$

with

34) $$\Sigma^{*} = \sigma^2 \mathbb{I} + \tau^2 \mathbb{D},$$

where

$$\mathbb{D} = \begin{bmatrix} \lambda_1 & 0 & . & . & . & 0 \\ 0 & \lambda_2 & . & . & . & 0 \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ 0 & 0 & . & . & . & \lambda_n \end{bmatrix}.$$

Although it has been established that there exists a transformation matrix $\mathbb{F}$ that will diagonalize the variance-covariance matrix of the observed displacements, the actual construction of such a matrix requires further analysis.

In the case when the Δt_i are equal

36) $$(\mathbb{H} - \lambda^{*}\mathbb{T}) = (\mathbb{H} - \lambda\, \mathbb{I})$$

where

37) $$\lambda = \lambda^{*}\Delta t = 2 - 2 \cos \frac{i\pi}{n+1}, \qquad i = 1, 2, \ldots n.$$

Therefore the equation of interest,

38) $$|\, \mathbb{H} - \lambda^{*}\, \mathbb{T} \,| = 0$$

In the general case when the Δt_i are not all equal the roots of equation [38] are the roots of the equation

40) $$\left| \mathbb{T}^{-\frac{1}{2}} \mathbb{H} \mathbb{T}^{-\frac{1}{2}} - \lambda \mathbb{I} \right| = 0.$$

In this case the roots may be found by numerical analysis with the aid of a computer.

Given the characteristic roots, we now exhibit the matrix $\mathbb{F}$ that produces the diagonal matrix of these roots by transformation of the $\mathbb{H}$ matrix.

Theorem 2

If $\mathbb{Z}$ is a nxn nonsingular, symmetrical matrix with distinct, nonzero characteristic roots λ_ν, $\nu = 1, 2, \ldots, n$, then the nxn matrix $\mathbb{P}$ whose columns are the n normalized characteristic vectors associated with the n characteristic roots of $|\mathbb{Z} - \lambda \mathbb{I}| = 0$ is orthogonal.

Proof

Let $\mathbb{X}_\nu$ and $\mathbb{X}_\eta$ be the characteristic vectors associated with the characteristic roots λ_ν and λ_η respectively.

By definition

41) $$\mathbb{Z} \mathbb{X}_\nu = \lambda_\nu \mathbb{X}_\nu .$$

Premultiplying this identity by $\mathbb{X}'_\eta$,

42) $$\mathbb{X}'_\eta \mathbb{Z} \mathbb{X}_\nu = \lambda_\nu \mathbb{X}'_\eta \mathbb{X}_\nu .$$

Also by definition

43) $$\mathbb{Z} \mathbb{X}_\eta = \lambda_\eta \mathbb{X}_\eta ,$$

and premultiplying by $\mathbb{X}'_\nu$,

44) $$\mathbb{X}'_\nu \mathbb{Z} \mathbb{X}_\eta = \lambda_\eta \mathbb{X}'_\nu \mathbb{X}_\eta .$$

Transposing both sides of equation [44]

45) $$\mathbb{X}'_\eta \mathbb{Z}' \mathbb{X}_\nu = \lambda_\eta \mathbb{X}'_\eta \mathbb{X}_\nu ,$$

and given that $\mathbb{Z}' = \mathbb{Z}$,

46) $$\mathbb{X}'_\eta \mathbb{Z}' \mathbb{X}_\nu = \mathbb{X}'_\eta \mathbb{Z} \mathbb{X}_\nu = \lambda_\eta \mathbb{X}'_\eta \mathbb{X}_\nu = \lambda_\nu \mathbb{X}'_\eta \mathbb{X}_\nu .$$

However, since

47) $$\lambda_\eta \neq \lambda_\nu , \quad \text{for } \eta \neq \nu ,$$

it follows

48) $$\mathbb{X}'_\eta \mathbb{X}_\nu = \begin{cases} 0, & \text{for } \eta \neq \nu , \\ 1, & \text{for } \eta = \nu . \end{cases}$$

Therefore the matrix $\mathbb{P}$ whose columns are the characteristic vectors $\mathbb{X}_\nu$, $\nu = 1, 2, \ldots, n$, is orthogonal.

By defining

49) $$\mathbb{H}^* = \mathbb{T}^{-\frac{1}{2}} \mathbb{H} \mathbb{T}^{-\frac{1}{2}} ,$$

with

50) $$\mathbb{T}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\Delta t_1} & 0 & . & . & . & 0 \\ 0 & \sqrt{\Delta t_2} & . & . & . & 0 \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ 0 & 0 & . & . & . & \sqrt{\Delta t_n} \end{bmatrix}$$

the characteristic equation

51) $$| \mathbb{H}^* - \lambda^* \mathbb{I} | = 0$$

is of the form considered in theorem 2. Hence the matrix $\mathbb{P}$ whose columns are the characteristic vectors associated with equation [51], and are defined by,

52) $$\mathbb{H}^* \mathbb{X}_\nu = \lambda^*_\nu \mathbb{X}_\nu ,$$

is an orthogonal matrix.

Letting

53) $$\mathbb{F} = \mathbb{T}^{-\frac{1}{2}} \mathbb{P} ,$$

it follows from

54) $$\mathbb{F}' \mathbb{T} \mathbb{F} = \mathbb{P}' \mathbb{T}^{-\frac{1}{2}} \mathbb{T} \mathbb{T}^{-\frac{1}{2}} \mathbb{P} = \mathbb{P}' \mathbb{I} \mathbb{P} = \mathbb{P}' \mathbb{P} = \mathbb{I}$$

that

55) $$\mathbb{F}' \mathbb{T} \mathbb{F} = \mathbb{I}.$$

Also it follows from

56) $$\mathbb{F}' \mathbb{H} \mathbb{F} = \mathbb{P}' \mathbb{T}^{-\frac{1}{2}} \mathbb{H} \mathbb{T}^{-\frac{1}{2}} \mathbb{P} = \mathbb{P}' \mathbb{H}^* \mathbb{P}$$

that

57) $$\mathbb{F}' \mathbb{H} \mathbb{F} = \mathbb{D} ,$$

since

$$58)\quad \mathbb{P}'\,\mathbb{H}^*\,\mathbb{P} = \begin{pmatrix} \mathbb{x}_1 \\ \mathbb{x}_2 \\ \vdots \\ \mathbb{x}_n \end{pmatrix} \mathbb{H}^*(\mathbb{x}_1, \mathbb{x}_2, \ldots \mathbb{x}_n) = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} = \mathbb{D}$$

from equation [45] and [48].

Therefore, $\mathbb{F}$ of equation [53] is the matrix that will transform the variance-covariance matrix $\mathbb{\Sigma}$ of equation [23] into canonical form.

Remembering that the above was concerned only with data in the X or E-W direction equal consideration must be given to the data in the Y or N-S direction.

Since the transformation matrix $\mathbb{F}$ was a function of the $\mathbb{H}$ and $\mathbb{T}$ matrices and the resulting characteristic roots, and since the $\mathbb{H}$ and $\mathbb{T}$ matrices are identical for both sets of data, the $\mathbb{F}$ matrix that is used to transform the E-W data will also be used to transform the N-S data.

This above presented consideration treated the case where the characteristic roots (equation [40]) are distinct. This will be the case when the Δt_i, $i = 1, 2, \ldots, n$, are constant and even in the majority of cases when they are not as can be seen from the equation of the roots.

If the situation arises that $|\mathbb{H} - \lambda^* \mathbb{T}| = 0$ has nondistinct roots, theorem 2 will not hold. However, the existence of a nonsingular transformation $\mathbb{F}$ that will reduce the variance-covariance matrix to canonical form is established by theorem 1. It is easily shown that the subspaces generated by equation [52] upon substitution of each of the distinct values of λ^*_ν are mutually orthogonal, and the development of explicit formulas for generation of a set of mutually orthogonal characteristic vectors is straightforward.

The ability to transform the displacement in such a way that the variance-covariance matrix is diagonal allows the probability density function of the transformed observations to be written as,

$$59)\quad f(\Delta x^*_1, \ldots, \Delta x^*_n;\ \Delta y^*_1, \ldots, \Delta y^*_n;\ \beta_1, \ldots, \beta_p)$$

$$= \frac{1}{(2\pi)^{n/2} \prod_{i=1}^{n} (\sigma^2 + \tau^2 d^*_{ii})^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n} \frac{(\Delta x^*_i - \sum_{j=1}^{p/2} \beta_j X^*_{ij})^2}{(\sigma^2 + \tau^2 d^*_{ii})} \right) - \frac{1}{2}\left(\sum_{i=1}^{n} \frac{(\Delta y^*_i - \sum_{j=p/2+1}^{p} \beta_j Y^*_{ij})^2}{(\sigma^2 + \tau^2 d^*_{ii})} \right) \right].$$

In the above equation Δx^*_i and Δy^*_i are the transformed observations, $\beta_1, \ldots, \beta_{p/2}$ are the unknown parameters in the E-W direction and $\beta_{p/2+1}, \ldots, \beta_p$ are the unknown parameters in the N-S direction. Also X^*_{ij} and Y^*_{ij} are the transformed known constants associated with the mean velocity of the current and the tidal effects in the E-W and the N-S directions respectively. d^*_{ii} is the

i^{th} characteristic root that is associated with the i^{th} observation.

Since the present problem includes only two periodic tidal fluctuation terms for each direction, p is equal to 6. From the model (equations [14] and [15]) β_1 and β_4 are μ_x and μ_y, and β_2, β_3, β_5 and β_6 are $\beta_x^{(1)}$, $\beta_x^{(2)}$, $\beta_y^{(1)}$ and $\beta_y^{(2)}$, respectively. The **X** matrix, (equation [21]), of known constants has elements

60)
$$
\begin{aligned}
X_{i1} &= \Delta t_i , \\
X_{i2} &= \cos(\omega_1 t_i - \phi_1) - \cos(\omega_1 t_{i-1} - \phi_1), \\
X_{i3} &= \cos(\omega_2 t_i - \phi_2) - \cos(\omega_2 t_{i-1} - \phi_2),
\end{aligned}
$$

where ω_1, ω_2, ϕ_1 and ϕ_2 are the periods and phase angles of the tidal components in the E-W direction. Similarly the counterpart of the matrix $\mathbb{X}$ for the data along the Y axis, the $\mathbb{Y}$ matrix, has elements,

61)
$$
\begin{aligned}
Y_{i1} &= \Delta t_i , \\
Y_{i2} &= \cos(\omega_3 t_i - \phi_3) - \cos(\omega_3 t_{i-1} - \phi_3), \\
Y_{i3} &= \cos(\omega_4 t_i - \phi_4) - \cos(\omega_4 t_{i-1} - \phi_4),
\end{aligned}
$$

where ω_3, ω_4, ϕ_3 and ϕ_4 are the periods and phase angles of the tidal components in the N-S direction. It will be noted that

$\omega_1 = \omega_3 = \frac{2\pi}{24}$ (if time is in lunar hours) $= \frac{2\pi}{24.84}$ (if time is in solar hours), and $\omega_2 = \omega_4 = \frac{2\pi}{12}$ (if time is in lunar hours) = $\frac{2\pi}{12.42}$ (if time is in solar hours).

To compute the phase angles comparison may be made with observed tidal currents off San Francisco Bay (10, p. 573). The phase angle ϕ_1 is that angle to compensate for the time difference between the time the drogues were launched and the time when the semidiurnal tidal component is a maximum in the E-W direction. Since the time interval between highest high tide and the maximum tidal component in this direction is known for the region off San Francisco Bay, by calculating the difference between highest high tide in the observation region and the time the drogues were launched, plus the difference between highest high tide at San Francisco and in the observation area, the phase angle can be obtained. Likewise ϕ_2, ϕ_3 and ϕ_4 are obtained for their respective directions and tidal periods.

Care must be taken to be consistent in the type of hour (lunar or solar) used in these computations and to use the value of ϕ in equation [63] that is consistent with the units that time is measured in.

The constants X^*_{ij} and Y^*_{ij} are then the ij elements of the transformed $\mathbb{X}$ and $\mathbb{Y}$ matrices defined by

62) $$\mathbb{X}^* = \mathbb{F}' \mathbb{X},$$

and

63) $$\mathbb{Y}^* = \mathbb{F}' \mathbb{Y}.$$

Also the transformed observations Δx_i and Δy_i are the i^{th} elements of the transformed vectors $\mathbb{\Delta}_x^*$ and $\mathbb{\Delta}_y^*$ defined by,

64) $$\mathbb{\Delta}_x^* = \mathbb{F}' \mathbb{\Delta}_x,$$

and

65) $$\mathbb{\Delta}_y^* = \mathbb{F}' \mathbb{\Delta}_y.$$

ESTIMATION OF PARAMETERS

The problem of making point estimates of $\hat{\theta}_j$, $j = 1, 2, \ldots, p$, of parameters θ_j deals with finding functions $\hat{\theta}_j = h_j(x_1, x_2, \ldots, x_n)$ of a random sample of n observations $x_1, x_2, \ldots, x_n$ from $f(x; \theta_1, \theta_2, \ldots \theta_p)$. The criterion for selecting any given set of functions is that the resulting parameter estimates be the "best" possible in that they are the closest values obtainable by known methods to the actual parameter value in terms of mean square error. This is to say that the estimates obtained from functions $\hat{\theta}_j = h_j(x_1, x_2, \ldots, x_n)$, $j = 1, 2, \ldots, p$, using a large number of random samples of n observations from $f(x; \theta_1, \theta_2, \ldots, \theta_p)$ are grouped closer to the actual value of the parameters than estimates obtained using any other set of functions $\theta_j = g_j(x_1, x_2, \ldots, x_n)$ from $f(x; \theta_1, \theta_2, \ldots, \theta_p)$.

By the transformation of the variance-covariance matrix to canonical form the estimation problem at hand has been reduced to a regression problem. This being the case, there are two basic methods by which estimates can be obtained: The method of Least Squares (3, p. 36-37) and the principle of Maximum Likelihood (8, p. 153-154).

The principal advantage of the method of Least Squares is that the frequency function of the "errors" need not be known. Yet in many cases unbiased and consistent estimates may be obtained, and under the conditions of statistical independence and equal variances, as prescribed by the Gauss Markoff theorem (6, p. 32), minimum variance estimates can be obtained.

If the observations are independent with equal variances which will be the case when

66) $$\lambda_i = \lambda(n; \Delta t_i, \ldots, \Delta t_n) \quad i = 1, 2, \ldots, n,$$

are constant, not only can minimum variance unbiased estimates be obtained but the variance of the estimates can be estimated.

The estimates and their variances can be obtained from the following formulas:

67) $$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{*\prime} \mathbf{X}^{*\prime})^{-1} \mathbf{X}^{*} \boldsymbol{\Delta}^{*},$$

and

68) $$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\sigma^2 + K\tau^2)(\mathbf{X}^{*\prime} \mathbf{X}^{*})^{-1}$$

where $\sigma^2 + K\tau^2$ is the common variance of the observations and $\mathbf{X}$ and $\boldsymbol{\Delta}$ are matrices as defined in equations [62] and [64].

It therefore follows that confidence intervals can be constructed for the estimates.

69) $$P_r\{|\hat{\beta}_j - \beta_j| < t_\alpha [\mathrm{Var}(\hat{\beta}_j)]^{\frac{1}{2}}\} = 1 - \alpha, \quad j = 1, 2, \ldots, p,$$

is the confidence statement associated with the confidence level

$1 - \alpha$, where $\mathrm{Var}(\hat{\beta}_j)$ is the jj element in the matrix in equation [68], and where $\sigma^2 + K\tau^2$ is the variance of the transformed random variables which can be found by general analysis of variance procedures.

The principle of Maximum Likelihood can be used as an estimation procedure when the observations do not have common variance as long as the density function is known.

There is no general argument that will show that the maximum-likelihood estimates are the best estimates possible for a given sample size. However, they do have at least three desirable large sample properties:

1. Consistency. An estimate $\hat{\theta}_n$ of a vector-valued parameter is said to be a consistent estimate if the probability that $|\hat{\theta}_n - \theta| > \epsilon$ converges to zero as n becomes large for every $\epsilon > 0$. $\hat{\theta}_n$ is the estimate of θ based on a large sample size of n.

2. Asymptotic normality. The distribution of $(\hat{\theta} - \theta)$ approaches a multivariate normal distribution with mean vector zero and variance-covariance matrix Γ where

70) $$\Gamma^{-1} = \left\{ -E\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \left(\ln f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_n) \right) \right] \right\}.$$

3. Asymptotic efficiency. The ratio of the mean square error

for any component of $\hat{\theta}$ to that of any other estimate of that component, is asymptotically no greater than unity.

It can be shown (6, p. 26) that if the observations are independent and normally distributed with common variance, the estimates obtained by the method of Least Squares are identical to the estimates obtained by the principle of Maximum Likelihood.

In this particular problem the probability density of the transformed observations (equation [59]) is known, and either the Least Squares or Maximum Likelihood procedure may be utilized. However, since the observations are generally not distributed with common variance (equation [34]), using the Least Squares procedure will not necessarily result in minimum variance estimates. The principle of Maximum Likelihood does, however, guarantee asymptotically efficient estimates.

Consequently, in this case it seems somewhat more desirable to employ the principle of Maximum Likelihood to obtain the estimates.

The actual procedure of maximizing the likelihood function $f(\Delta x_i^*, \ldots, \Delta x_n^*; \Delta y_i^*, \ldots, \Delta y_n^*; \beta_1, \ldots, \beta_p)$, involves the following set of $p + 2$ nonlinear equations with $p + 2$ unknowns:

71) $$\frac{\partial \ell n f}{\partial \sigma^2} = \frac{\partial \ell n f}{\partial \tau^2} = \frac{\partial \ell n f}{\partial \beta_1} = \frac{\partial \ell n f}{\partial \beta_2} = \ldots = \frac{\partial \ell n f}{\partial \beta_p} = 0$$

The solutions of these equations will be designated $\hat{\sigma}^2$, $\hat{\tau}^2$, $\hat{\beta}_1, \ldots, \hat{\beta}_p$, and are the maximum-likelihood estimates of the unknown parameters of equation [59]. The computations required to solve these equations are somewhat formidable and require a high-speed computer.

As it has been previously pointed out, maximum-likelihood estimates are asymptotically efficient with this asymptotic variance as stated in equation [70]. Consequently confidence statements can be developed concerning the estimates by using the estimates in equation [70] to estimate the asymptotic variance. In the situation where the least squares estimates are appropriate, equation [69] is used.

In any particular situation, equation [70] will be quite complex and solution will require the use of a computer. In practice this would be done at the same time that the estimates themselves are computed.

BIBLIOGRAPHY

1. Anderson, T. W. An introduction to multivariate statistical analysis. New York, John Wiley and Sons, Inc., 1958. 374 p.

2. Doob, J. L. Stochastic processes. New York, John Wiley and Sons, Inc., 1953. 654 p.

3. Graybill, Franklin A. An introduction to linear statistical models. New York, McGraw-Hill Book Company, Inc., 1961. 463 p.

4. Grenander, Ulf and Gabor Szegö. Toeplitz forms and their applications. Berkeley, University of California Press, 1958. 245 p.

5. Hodges, J. L. Jr. and E. L. Lehmann. Basic concepts of probability and statistics. San Francisco, Holden-Day, 1964. 375 p.

6. Kempthorne, Oscar. The design and analysis of experiment. New York, John Wiley and Sons, Inc., 1952. 631 p.

7. Maughan, Paul McAlpine. Observations and analysis of ocean currents above 250 meters off the Oregon coast. Master's thesis. Corvallis, Oregon State University, 1963. 49 numb. leaves.

8. Mood, Alexander McFarlane. Introduction to the theory of statistics. New York, McGraw-Hill Book Company, Inc., 1950. 433 p.

9. Parzen, Emanuel. Stochastic processes. San Francisco, Holden-Day, Inc., 1962. 324 p.

10. Sverdrup, H. U., and Martin W. Johnson and Richard H. Fleming. The oceans; their physics, chemistry, and general biology. Englewood Cliffs, N. J., Prentice-Hall, Inc., 1942. 1087 p.