DTIC FILE COPY

Reprinted from JOURNAL OF ATMOSPHERIC AND OCEANIC TECHNOLOGY, Vol. 6, No. 6, December 1989
American Meteorological Society

AD-A217 520

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

DTIC
ELECTE
JAN 31 1990
S D
cD

Sampling Characteristics of Satellite Orbits

CARL WUNSCH

Center for Meteorology and Physical Oceanography, Department of Earth, Atmospheric and Planetary Sciences, Cambridge, Massachusetts

(Manuscript received 12 September 1988, in final form 28 March 1989)

ABSTRACT

The irregular space–time sampling of any finite region by an orbiting satellite raises difficult questions as to which frequencies and wavenumbers can be determined and which will alias into others. Conventional sampling theorems must be extended to account for both irregular data distributions and observational noise—the sampling irregularity making the system much more susceptible to noise than in regularly sampled cases. The problem is formulated here in terms of least-squares and applied to spacecraft in 10-day and 17-day repeating orbits. The "diamond" pattern laid down spatially in such repeating orbits means that either repeat period adequately samples the spatial variables, but the slow overall temporal coverage in the 17-day pattern leads to much greater uncertainty than in the shorter repeat cycle. The result is *not* definitive and it is not concluded that a 10-day orbit repeat is the most appropriate one. A major conclusion however, is that different orbital choices have potentially quite different sampling characteristics which need to be analyzed in terms of the spectral characteristics of the moving sea surface. Conclusions drawn from model assimilation studies need to be placed in a context of the reality of their spectral content via-à-vis the ocean.

Reprints, (hdle
N00014-85-J-1241

1. Introduction

Subsatellite tracks lay out a geographically and temporally complex sampling pattern on the sea surface. As with any measurement system, understanding of its capabilities and discussion of orbital choice necessitates at least a semiquantitative determination of its sampling characteristics. For an instrument making point measurements, uniformly spaced in time, one normally determines the Nyquist frequency from the known characteristics of the physical process being measured, and the physical characteristics of the measurement system. Usually there is some alias-band of frequencies which one cannot hope to observe properly, and one designs the system with that in mind. A similar determination—finding which frequencies and wavenumbers are determinable, and which are aliased, from a given orbit—is an essential first step before one can discuss alternative choices. The immediate motivation is the need specifically to understand altimetric sampling, but the results apply equally well to scatterometer, or infrared, or in general, to any other instrument sampling three dimensions in a nonregular manner.

Most conventional sampling theorems (e.g., Freeman 1965) are derived for uniformly spaced data. Although the satellite case is much more complex, it is helpful to recapitulate the most basic and important of all cases. Consider an even number, N, of observations $f(n\Delta q)$, $n = 0$ to $N-1$ spaced equally in a continuous coordinate q, at intervals Δq (q is deliberately made abstract; it can be space or time). Then (e.g., Hamming 1973), the expansion functions

$$\left.\begin{aligned} &1, \cos\left[\frac{2\pi q}{L}\right], \cos\left[\frac{2\pi}{L}2q\right], \ldots, \cos\left[\frac{2\pi}{L}\frac{N}{2}q\right], \\ &\sin\left[\frac{2\pi q}{L}\right], \sin\left[\frac{2\pi}{L}2q\right], \ldots, \sin\left[\frac{2\pi}{L}\left(\frac{N}{2}-1\right)q\right], \end{aligned}\right\} \tag{1}$$

where $L = N\Delta q$, are a complete basis for the samples:

$$f(n\Delta q) = \frac{a_0}{2} + \sum_{r=1}^{N/2} a_r \cos\frac{2\pi r(n\Delta q)}{L} + \sum_{r=1}^{(N/2)-1} b_r \sin\frac{2\pi r(n\Delta q)}{L}, \quad n = 0 \text{ to } N-1 \tag{2a}$$

exactly, if

$$Na_k = \sum_{n=0}^{N-1} f(n\Delta q)\cos\left(\frac{2\pi}{L}kn\Delta q\right),$$

$$Nb_k = \sum_{0}^{N-1} f(n\Delta q)\sin\left(\frac{2\pi}{L}kn\Delta q\right). \tag{2b}$$

(The normalization is $2N$ for $k = 0, N$.) Any frequencies, ω, outside the "baseband," $|\omega| \leq 2\pi/2\Delta q$, alias into the set of frequencies,

Corresponding author address: Dr. Carl Wunsch, Dept. of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.

© 1989 American Meteorological Society

90 01 30 010

$$\omega_{\text{alias}} = \omega \pm \frac{2\pi j}{\Delta q}, \quad j = 0, 1 \cdots, \tag{3}$$

involved in (1).

This "Shannon–Whittaker sampling theorem" is sometimes best viewed from a converse point of view: that if $|f(\omega)| \equiv 0$ for $|\omega| \geq (2\pi/2\Delta q)$, (a "band-limited" function) then if there are an infinite number of observations, $f(q)$ can be reconstructed perfectly from the interpolation formula

$$f(q) = \sum_{n=-\infty}^{\infty} f(n\Delta q) \frac{\sin\left(\frac{\pi q}{\Delta q} - n\pi\right)}{\left(\frac{\pi q}{\Delta q} - n\pi\right)}. \tag{4}$$

If the band-limiting is violated, one can still use the formula (4), but the reconstructed continuous function $f(q)$ does not coincide with the correct function except at the observation points themselves. High frequencies alias (i.e., appear as lower frequencies) with potentially disastrous consequences when one goes to interpret the data, as depicted in Fig. 1.

The derivations of the aliasing formula and the interpolation rule (4) are elementary for equally spaced data (e.g., Bracewell 1978). But subsatellite data locations while not randomly distributed, are quite irregular, and the normal derivations will not apply. Furthermore, the presence of noise in the data is not easily accommodated by the conventional result. We need a different approach, which we will try to elucidate by first applying it to the known case just described.

2. Sampling theorems by least squares

If we attempt a least-squares fit of the basis functions (1) to $f(q)$ in the form

$$\min: \sum_n \left[f(n\Delta q) - \sum_k a_k \cos\left(\frac{2\pi k}{L} n\Delta q\right) - \sum_k b_k \sin\left(\frac{2\pi k}{L} n\Delta q\right)\right]^2, \tag{5}$$

it is well known that the conventional Fourier coefficients are obtained. The least-squares approach would not normally be used for the equally spaced case, because it fails to take advantage of the enormous computational efficiency made possible by the orthogonality of the basis functions and fast transform techniques. But least-squares has the advantage of generalizing to the cases we need to examine.

Even though the fit (5) is exact, with a minimum of zero if the full basis set (1) is used, and ordinary numerical Fourier transforms do a fine job on finding the a_k, b_k, let us rewrite (5) as: Find the best (in the minimum mean-square residual sense) solution to Eq. (2a), which we rewrite as

$$\mathbf{A}\alpha = \beta, \tag{6}$$

where

$$\mathbf{A} = \begin{Bmatrix} 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 1 & \cos\left[\frac{2\pi}{L}\Delta q\right] & \cdots & \cos\left[\frac{2\pi}{L}\frac{N}{2}\Delta q\right] & \sin\left[\frac{2\pi}{L}\Delta q\right] & \cdots & \sin\left[\frac{2\pi}{L}\left(\frac{N}{2}-1\right)\Delta q\right] \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \cos\left[\frac{2\pi}{L}(N-1)\Delta q\right] & \cdots & \cdots & \sin\left[\frac{2\pi}{L}(N-1)\Delta q\right] & \cdots & \sin\left[\frac{2\pi}{L}\left(\frac{N}{2}-1\right)(N-1)\Delta q\right] \end{Bmatrix}$$

$$\alpha = \begin{bmatrix} a_0 \\ \vdots \\ a_{N/2} \\ b_1 \\ \vdots \\ b_{(N/2)-1} \end{bmatrix}, \qquad \beta = \begin{bmatrix} f[0] \\ f[\Delta q] \\ f[2\Delta q] \\ \vdots \\ f[(N-1)\Delta q] \end{bmatrix}.$$

We solve (6) using the singular value decomposition (SVD); for example, see Lawson and Hanson (1974) or Wunsch (1978) and write the solution as

$$\hat{\alpha} = \sum_{i=1}^{k} \frac{\mathbf{u}_i^T \beta}{\lambda_i} \mathbf{v}_i + \sum_{i=k+1}^{N} \gamma_i \mathbf{v}_i, \tag{7}$$

with $k \leq N$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the column vectors of the singular value decomposition of $\mathbf{A}$ making up the columns of $\mathbf{U}$ and $\mathbf{V}$ respectively, where

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{V}^T. \tag{8}$$

The λ_i are the "singular values," the elements of the diagonal matrix Λ. We write $\hat{\alpha}$ to distinguish the computed solution from the true one. The γ_i are constants.

[If the summations over a_k and b_k are carried to the upper limits of (2a) the problem as posed in (6) is fully determined and the SVD machinery is unnecessary. One of its great advantages however, is that this conventional situation is included as a special case of the much more general problems encountered below.] We adopt the usual convention that rows and columns of

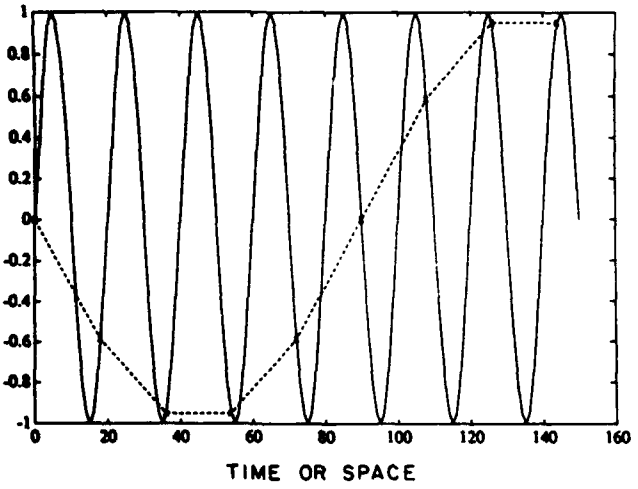


FIG. 1. The classical aliasing effect of uniform sampling too slow to distinguish a frequency above the Nyquist frequency from one below, as in expression (1). The interpolated function agrees with the original function at the sample points, and is smooth in between. But any estimate of the derivatives of the original function made from such a grossly undersampled situation will be erroneous.

the diagonal matrix Λ which are all zeroes are dropped, as are the corresponding columns of the U and V matrices [i.e., the null-space, if present, is dropped from the representation (8)]. Expression (7) is completely general, the second terms on the right representing the null-space: the components with unknown and indeterminate coefficients γ_i. The present case is special because the singular values λ_i are all about the same size, and there are $k = N$ of them, just the number necessary to solve the problem. Figure 2a displays the singular values of the present case.

A part of the SVD machinery is the so-called resolution matrix $\mathbf{V}\mathbf{V}^T$. We cannot provide a full discussion of the interpretation of this operator here (see Wiggins 1972; Wunsch 1978), but the following may help those unfamiliar with it.

If the null-space of the solution vanishes, then all elements of the solution are completely determined and denoted "fully resolved." If a null-space exists, it is a straightforward matter to show that the relationship between the correct solution α and the one determined, $\hat{\alpha}$, is

$$\hat{\alpha} = \mathbf{V}\mathbf{V}^T\alpha. \tag{9}$$

Equation (9) is readily interpreted as showing that in the general case, the solution estimated is a linear combination (an average) of the elements of the true solution. If element i of the true solution is fully resolved, then the corresponding row i of $\mathbf{V}\mathbf{V}^T$ is the ith row of the identity matrix. It is an elementary demonstration that in the fully resolved case, the entire $\mathbf{V}\mathbf{V}^T$ matrix reduces to the identity, as it must. In the opposite extreme, when the data contain no information about element i of the solution, the ith row of $\mathbf{V}\mathbf{V}^T$ is all zeroes.

In the present case everything is fully determined, and $\mathbf{V}\mathbf{V}^T$ is the identity. The problem was solved slightly differently from the way we have posed it, though. In particular, because the computer coding was slightly simpler, and because it is illustrative of a more general machinery, the basis functions employed included the sine of both zero frequency, and the Nyquist frequency—neither of which belongs to the basis set (1). Figure 2b displays the diagonal elements of $\mathbf{V}\mathbf{V}^T$

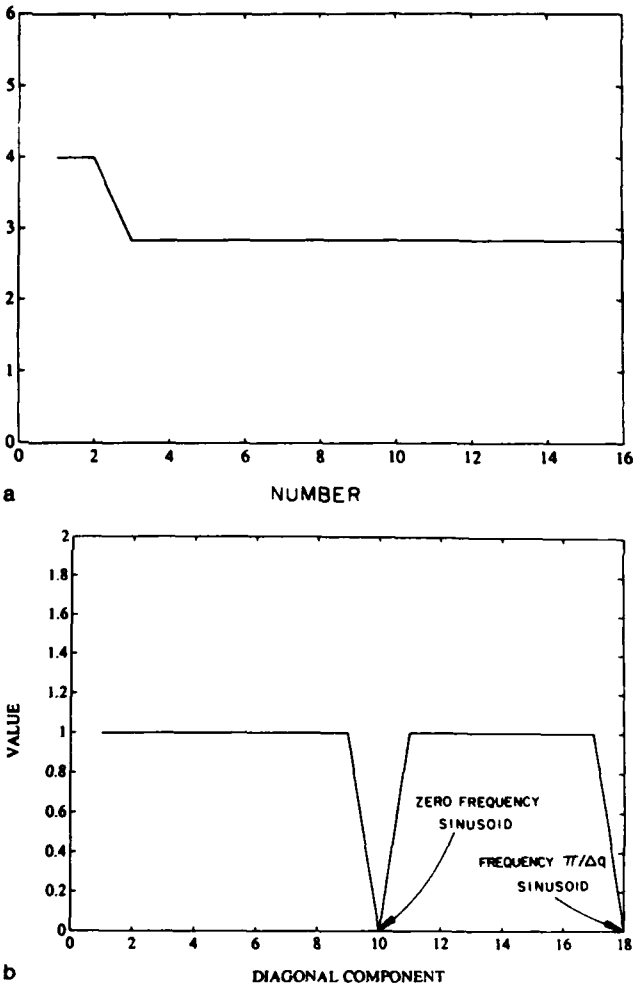


FIG. 2. (a) Singular values of the fit of 16 points to their Fourier series by least-squares. Two of the values are slightly higher than the others because the mean square value of unity and the cosine of the Nyquist frequency are 2× those of a cosine or sine. This phenomenon could be removed by appropriate column scaling, but it was not thought worthwhile to do so. (b) Diagonal component of the resolution matrix for the same fit as in Panel (a). All values are unity (full resolution) except for two that vanish. These latter two are an artifact deriving from carrying two extra, useless functions in the fit, notably the sine of zero frequency and the sine of the Nyquist frequency [neither belongs to the set in Eq. (1)]. One of the great virtues of the SVD is its ability to automatically handle this indeterminancy (we had 16 equations in 18 unknowns) which was done solely because the computer coding was marginally simpler. The handling is automatic in that one is told that no information is available about these two coefficients, and if one solved for them, the SVD would produce a best estimate of zero.

for the case actually run; one sees that all of them are unity (the remaining elements in each of the corresponding rows are zeroes), with the exception of the two corresponding to the two extra sinusoids—for which the elements are zeroes—showing that the data distribution we had contains no information whatever about these two basis elements. An important property of the SVD method is its ability to carry two unnecessary elements and to inform the user that they are unnecessary. Later on, we will see more interesting cases.

All least-squares solutions contain a priori assumptions about noise and solution statistics (Wunsch 1989) and the particular ones leading to (7) are: uniform variance in the solution elements, and zero covariance between solution elements; also, any noise in the observation is uncorrelated from one measurement to another, and is of constant variance.

It is instructive, before moving on, to examine modifications of the present example.

(i) Suppose we try to resolve frequencies outside the fundamental baseband $|\omega| \leq (2\pi/2\Delta q)$? That is, suppose we seek to find the Fourier coefficients out to frequencies = 2 × Nyquist frequency = $(2\pi/L)\,N$?

The problem is now underdetermined, because we are looking for twice as many Fourier coefficients as data points, but as will be seen, underdetermination has to be faced later anyway. The diagonal of the resolution matrix $\mathbf{V}\mathbf{V}^T$ for this case is displayed in Fig. 3a and a typical row in (3b). The resolution of the sines of frequencies 0, $\pi/\Delta q$ and $2\pi/\Delta q$ is zero as expected. The resolution of the amplitude of the cosine of frequency $\pi/\Delta q$ remains unity, but none of the other coefficients is now fully resolved. The diagonal element, which was unity before, is now reduced to ½, and ½ appears in the row at corresponding frequency $\omega + (2\pi/\Delta q)$ which is precisely the alias frequency (3). The form (9) thus states that the information available is only adequate to determine the average of the ith Fourier coefficient in the baseband and that of its higher frequency alias in the form $\hat{\alpha}_i = \frac{1}{2}(\alpha_i + \alpha_{\text{alias of } i})$. If the baseband frequency were truly absent and the higher frequency present, then analysis would nonetheless produce a nonzero value within the baseband. In particular, the coefficient of the cosine of zero frequency, the mean, is determined as the sum of the true coefficient of that frequency and that at frequency $2\pi/\Delta q$—one of its aliases. The rows of $\mathbf{V}\mathbf{V}^T$ thus produce an explicit statement of all the frequencies contributing to the estimates of Fourier amplitudes at any given frequency. In the present uniformly distributed noise-free case, the aliases are simple to understand; when we get to irregular noisy data, the dependences become very complex.

Practitioners of inverse methods will recognize that we could give a larger a priori weight to the lower frequencies if we chose, thus assigning all of the amplitude to it. Assignment of such a priori weights is outside the normal scope of sampling theorems, however, which treat all frequencies on an equal footing.

(ii) Now revert to consideration of the baseband alone, but take one of the original uniformly spaced samples, and remove it, replacing it by a sample at an irregular interval (see Fig. 4). Freeman (1965) shows how to prove that the irregularly spaced sample is a complete substitute for the missing regular one. But he deals only with the case of perfect observations, and I wish to demonstrate the effect of noise.

We replace one sample $f(m\Delta q)$ by a sample at an irregularly spaced point $f(t_m)$, and the corresponding missing equation in (2a) or (6) is replaced by

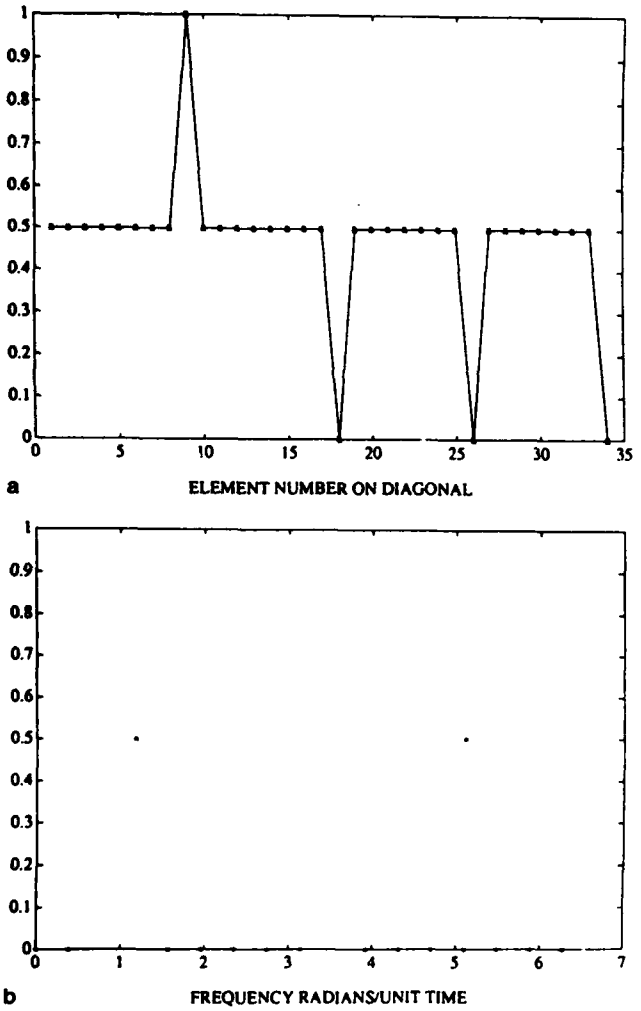


FIG. 3. (a) Diagonal of the resolution matrix of regularly spaced, aliased case. Apart from the cosine of frequency $\pi/\Delta q$ all nonzero resolutions have been reduced to 0.5. (b) One of the rows of the resolution matrix corresponding to 3b, showing that the cosine coefficient at that particular frequency is determined only in linear combination with the cosine of the alias frequency (elements corresponding to the sine parts have been suppressed; they all vanish). The result is an inability to distinguish the frequency from its alias (without provision of further information).

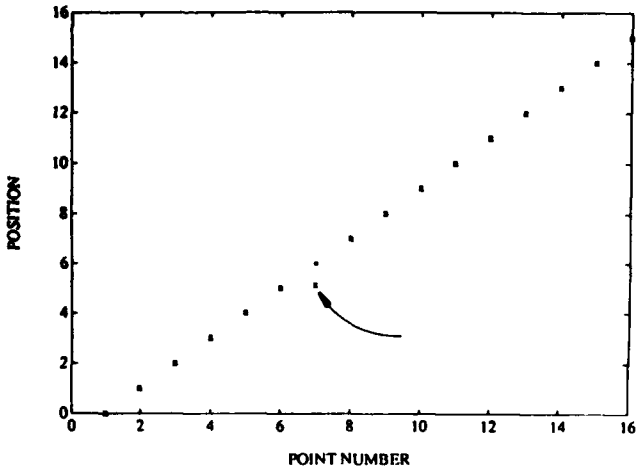


FIG. 4. Point number 7 has been displaced from its "regularly spaced" location to a position much closer to its neighboring point 6.

$$f(t_m) = \frac{a_0}{2} + \sum_{r=1}^{N/2} a_r \cos\frac{2\pi r}{L} t_m + \sum_{r=1}^{(N/2)-1} b_r \sin\frac{2\pi r}{L} t_m.$$

Formally, there are still N equations in N unknowns and this observation is the basis of the conclusion (formalized by Freeman 1965), that one does not need regularly spaced data to preserve the perfect reconstruction in (4). We persist however, in solving it by the SVD; the solution is still then of the form (7). Use of the SVD demonstrates the real effect of the irregular spacing: one of the singular values (displayed in Fig. 5) has become very small.

Because the covariance of the solution is

$$\langle(\hat{\alpha} - \langle\hat{\alpha}\rangle)(\hat{\alpha} - \langle\hat{\alpha}\rangle)^T\rangle = \sigma^2 \mathbf{V}\mathbf{\Lambda}^{-2}\mathbf{V}^T, \quad (10)$$

where σ^2 is the observation noise variance (for example, see Wunsch 1978), the effect of the irregularly spaced observation is to reduce the effective rank in the presence of noise. As the spacing becomes more irregular, this noise susceptibility grows. If the effective rank is significantly reduced, some of the Fourier coefficients will become indeterminate, or effective aliases of others, or simultaneously both of these things.

As the number of irrregularly spaced points increases, and as the spacing between any pair of points is reduced, the singular values decline. Thus the limiting case, when *all* the samples are squeezed arbitrarily close together in an arbitrarily small interval remains a hypothetical sampling strategy for perfect data, but is ruled out in practice because the slightest observational noise would render arbitrarily large the variance of the result and the numerical values as useless. (The limit is tricky in any case; see Jerri 1977.)

When we move to two and three dimensions, the SVD remains the analysis tool of choice because in a precise sense it provides all possible information concerning the solution (for example, a review of its properties may be found in Wunsch 1978). But the computational load grows very rapidly, and the amount of information is so large as to be overwhelming, and indeed more than we need for present purposes.

3. Tapered least-squares approach

In any estimation problem, there are generally two types of errors: (i) those owing to wholly missing information, and (ii), those owing to noise in the observations (a third error, often the most important, is model failure; but it is not an issue here). It is (i) which appeared in section 2, when we sought twice as many Fourier components as we were entitled to estimate with the available data, and is the error owing to a "failure to resolve." When computed by an SVD, the unresolved components are assigned zero amplitude; i.e., the null space coefficients are set to zero. If the measurements are not perfect, the noise gives rise to an uncertainty in those components which have been determined. This latter noise is what appears as the solution uncertainty in conventional overdetermined least-squares. The advantage of the SVD is that it is able to compute these errors separately (and because their origins and cures are quite different, a separation is often extremely helpful); the main disadvantage is the large computational load involved.

Conventional "tapered least-squares" combines these errors into a single error estimate which is more easily affordable. The result is obtained most simply from the Gauss–Markov theorem, in the restricted form in sections 5–7 of Liebelt (1967). To use this result, we must however, specify a prior estimate of

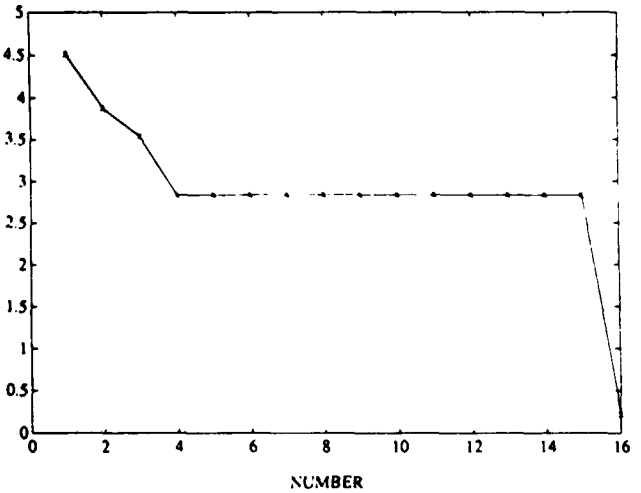


FIG. 5. Singular values of Fourier fit to data distributed as in Fig. 4; singular values change from those displayed in Fig. 2a, with one of them becoming very small—thus making the result vulnerable to noise.

the variance of the solution to (6) relative to the noise in the observations.

Suppose that the observational noise in the ith observation is n_i and has an "a priori covariance":

$$\langle n_i n_j \rangle = \sigma^2 \delta_{ij},$$

with $\langle \alpha_i \alpha_j \rangle = \Delta^2 \delta_{ij}$ as the a priori covariances of the Fourier coefficients; then we have

$$\hat{\alpha} = \left(\mathbf{A}^T \mathbf{A} + \frac{\sigma^2}{\Delta^2} \mathbf{I} \right)^{-1} \mathbf{A}^T \beta \tag{11a}$$

$$\mathbf{E} = \sigma^2 \left[\mathbf{A}^T \mathbf{A} + \frac{\sigma^2}{\Delta^2} \mathbf{I} \right]^{-1}, \tag{11b}$$

or

$$\hat{\alpha} = \mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T + \frac{\sigma^2}{\Delta^2} \mathbf{I} \right)^{-1} \beta \tag{12a}$$

$$\mathbf{E} = \Delta^2 \left[\mathbf{I} - \mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T + \frac{\sigma^2}{\Delta^2} \mathbf{I} \right)^{-1} \mathbf{A} \right], \tag{12b}$$

where $\mathbf{E} = \langle (\hat{\alpha} - \langle \alpha \rangle)(\hat{\alpha} - \langle \alpha \rangle)^T \rangle$.

Expressions (11) and (12) are algebraically identical; a choice between them is made for numerical efficiency depending upon whether there are more observations than unknowns [use (11)], or more unknowns than observations [use (12)]—choosing so as to minimize the dimension of the matrix to be inverted.

To demonstrate how E combines the overall mean-square errors, we display in Fig. 6a the result for the one irregularly spaced point in Fig. 2, but where now, we have asserted that $(\sigma^2/\Delta^2) = 0.1$. Figure 6 shows the normalized variance of the cosine coefficients for the classical case (the covariance matrix is diagonal). For the case of one (now noisy) irregularly spaced observation, the variance increases as shown in the same figure. There are off-diagonal components in this covariance matrix (Fig. 6b). For this particular pair of frequencies, there is a negative correlation between the estimation errors of each. As the number of data spacing irregularities increases, the variance generally grows monotonically, with some wavenumbers being more sensitive than others, dependent upon the precise nature of the irregularity. A smaller noise to signal ratio of 10^{-4} will be our standard value for the calculations below; it describes the measurement noise relative to the power in the Fourier coefficients, a quantity which for white noise decreases with the number of Fourier coefficients—by Parseval's theorem. The value is arbitrary, but is chosen so as to strongly influence the weakest components of the available information.

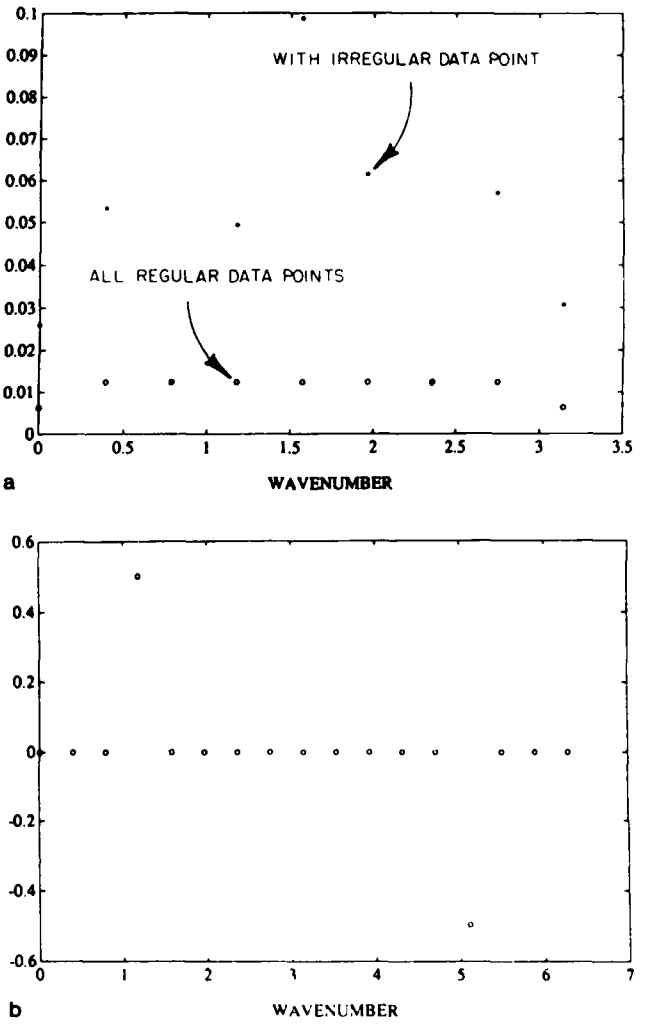


FIG. 6. (a) Diagonal component of E matrix, for one irregular point, cosine components only, compared to that of completely regular conventional case. The presence of one irregularly spaced point greatly increases the noise susceptibility. (b) One row of E matrix for conventional aliased case, showing the alias as a strong covariance between the error in a particular frequency component and the corresponding alias frequency.

4. Two dimensions

Consider a regular two-dimensional grid. A complete set of basis functions is now composed of functions of the form

$$1, \cos\left(\frac{2\pi}{L_q} nq\right) \cos\left(\frac{2\pi}{L_r} mr\right),$$

$$\sin\left(\frac{2\pi}{L_q} nq\right) \cos\left(\frac{2\pi}{L_r} mr\right), \ldots,$$

where L_q, L_r are the two appropriate length scales, q, r are the corresponding physical positions and m, n are integers. The two-dimensional sampling theorem of Petersen and Middleton (1962) shows that wavenumbers lying outside the baseband $|\omega_q| < (\pi/\Delta q)$, $|\omega_r| < (\pi/\Delta r)$ will be aliased into it.

It is possible to repeat essentially all the experiments done above in one dimension. It might appear that the two-dimensional case is of great interest for the orbital

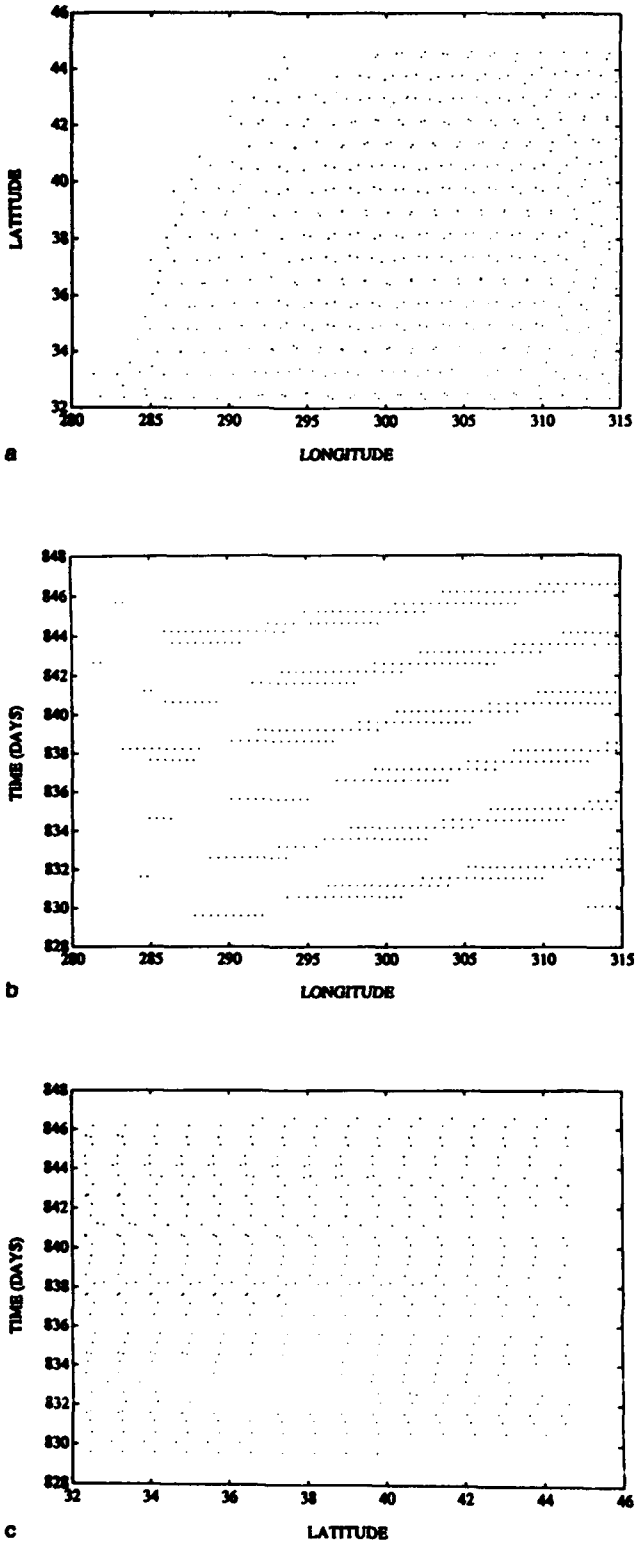


FIG. 7. (a) Grid pattern in part of the western North Atlantic showing the distribution of data obtained during repeat cycle 10 of GEOSAT. The northwest corner is blank owing to the presence of the eastern United States and Canada, and some of the descending arcs had no useful data. Data were averaged along-track to intervals of 105 km. (b) Time/longitude sampling for the data distribution in Fig. 8a. Rapid bursts are obtained in single arcs. (c) Time/latitude sampling for Panel (a).

sampling issue as it relates to study of the mean sea surface. But consider that the mean surface is often computed by first forming a set of mean arcs, each as an average over many repeat cycles. In a 17 day repeat, such as that for GEOSAT, any motions with time scales shorter than the Nyquist period of 34 days will alias directly into apparent periods longer than 34 days, including zero frequency, unless removed by filtering first. The potential for distorting estimates of both the mean and the low frequency variability is evident, especially in regions of strong western boundary currents with their intense short-time scale variability.

We will thus turn directly to the full three-dimensional case.

5. Three dimensions

a. Real GEOSAT

Consider the data distribution displayed in Fig. 7a from one of the repeat cycles of GEOSAT in the North Atlantic. The data are laid down in a complex space time structure. To gain some insight into the problem, we show in Fig. 7b the longitude structure versus time, and in Fig. 7c the latitude structure against time. In this not untypical region, the data arrive in bursts

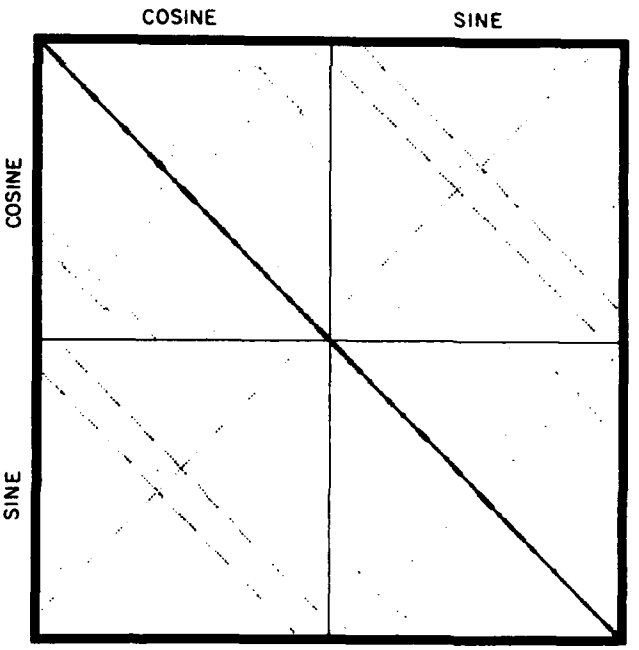


FIG. 8. Contours of the rows of the full covariance matrix for the sampling distribution of Fig. 7 and the assumptions given in the text. Only values with magnitudes lying between 0.3 and 1.0 are contoured. No negative values less than −0.2 occur. An element lying along the diagonal has a significant correlation of the error of its estimate with that of another Fourier component if a contour appears elsewhere in its row. A full display and discussion of this matrix is beyond the scope of the present paper; it does show that the major issue is not aliasing (at least within the restricted wavenumber/frequency band used here) but rather uncertainty owing to noise and the irregular sampling distribution.

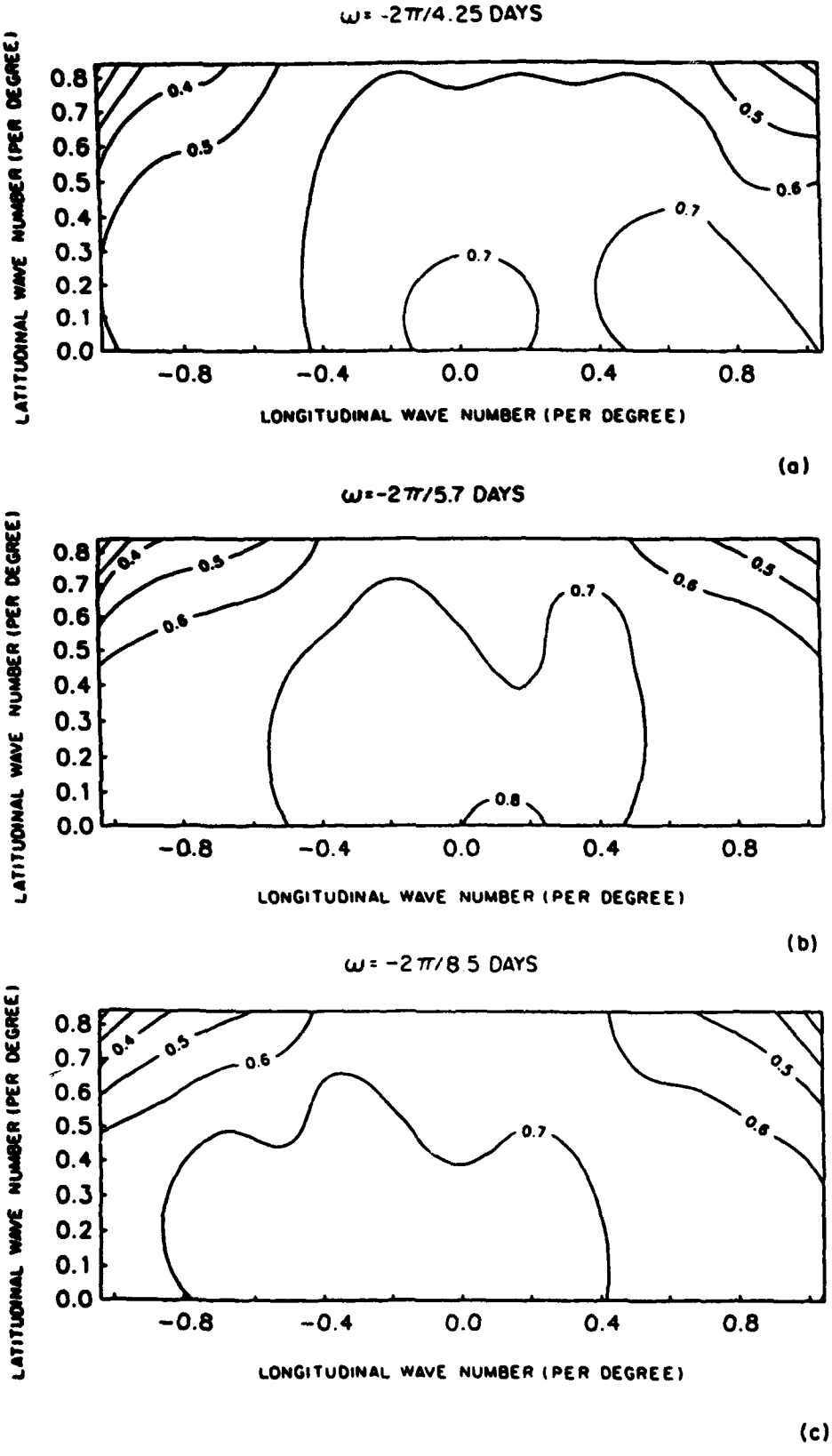
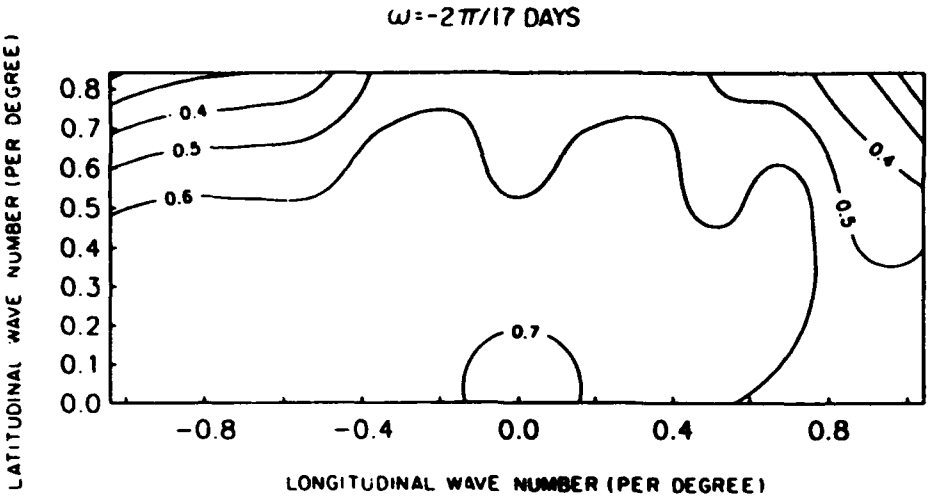
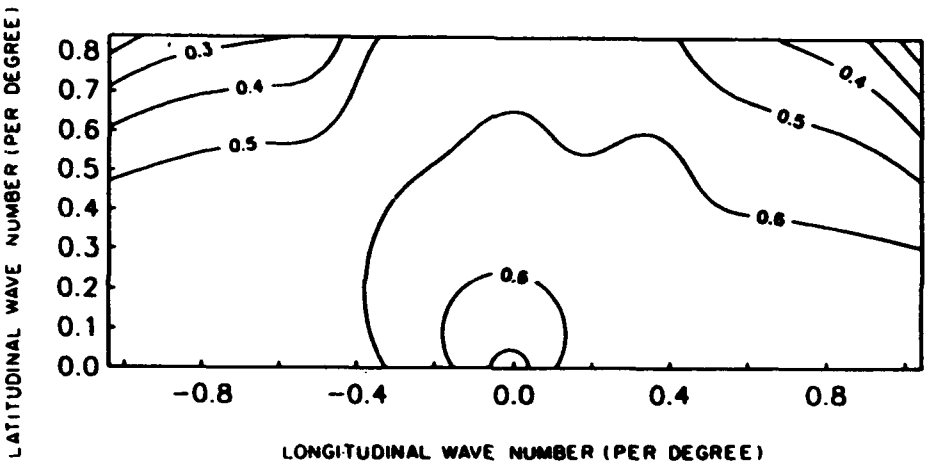


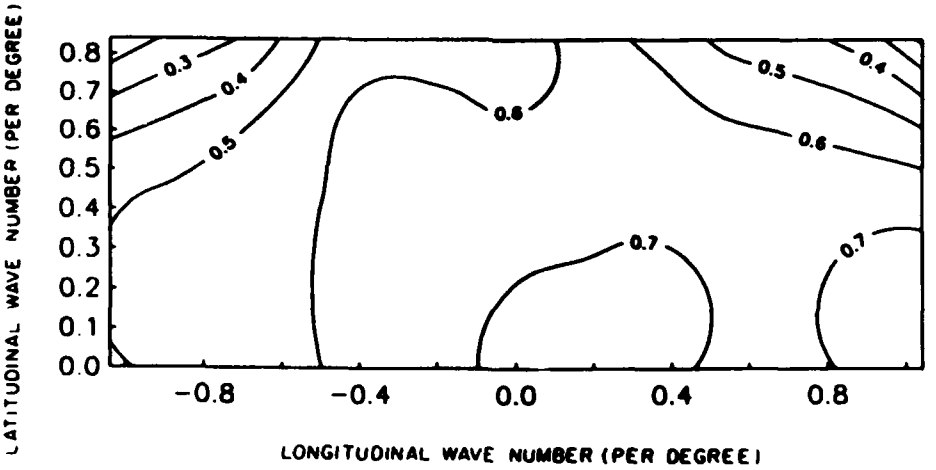
FIG. 9. (a) Estimate of the variance of the cosine coefficients of the three-dimensional Fourier transform of data distributed as in Fig. 7, at frequency $\omega = -2\pi/(4.25 \text{ day})$. Shown as radians/° of latitude against radians/° of longitude. Values are fraction of solution variance; thus formal error of any cosine coefficient would be $\pm\sqrt{\Delta^2}$ · value shown. These figures are subregions of the diagonal of Fig. 8 at fixed frequency. Panels (b)–(f) as in (a) but for frequency shown.



(d)



(e)

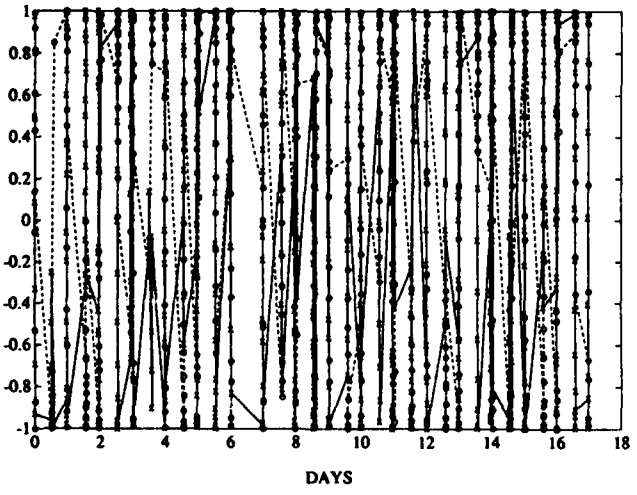


(f)

FIG. 9. (*Continued*)

closely spaced in time with a sequence of quasi-regular gaps built around two basic intervals: the orbital period, which gives the time to return to the area in the next orbit, and the longer interval when complete orbits lie outside the region under study. (The case shown is a real one, and some data are missing and some of the region is land-covered.)

The closely spaced bursts have a fundamental sampling interval of one second, the entire along-track burst occupying O(1 minute) for regions the size of that shown in Fig. 7a (the spacecraft moves at roughly 7 km s^{-1}). These closely spaced bursts will behave like the closely spaced, irregular data discussed in section 2, and one can anticipate that use of these data to determine many of the time dependent components will be very sensitive to slight noise in the observations.



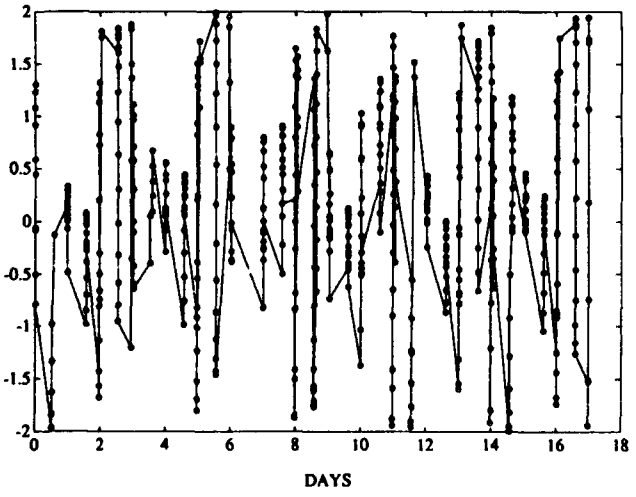


FIG. 10. Two pure cosines, (upper)

$$s_1 = \cos\left(\frac{-2\pi}{36^c}x + \frac{2\pi}{15^\circ}y - \frac{2\pi}{17d}t\right);$$

$$s_2 = \cos\left(\frac{-2\pi}{18^\circ}x + \frac{2\pi}{15^\circ}y - \frac{2\pi}{8.5d}t\right),$$

and their sum (lower) as they would be seen in time if actually sampled as shown in Fig. 7. The distribution of data is very far from the regular one that normally leads to a robust separation of two such distinct cosines through conventional Fourier fits. As seen in the text, even though separation still remains a theoretical possibility, the amplitude estimates are noise sensitive.

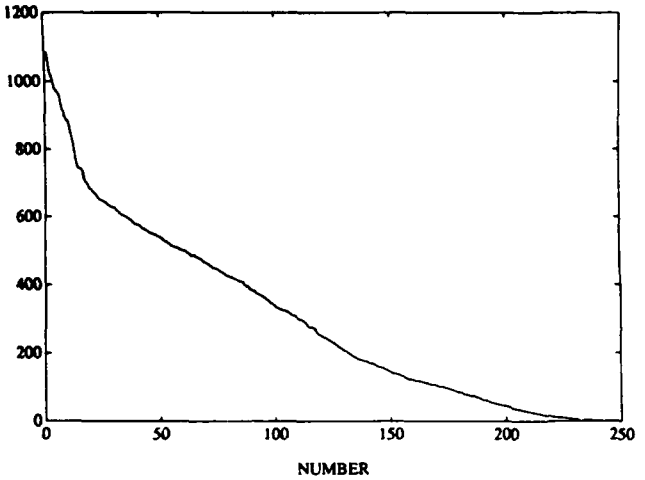


FIG. 11. Singular values (squared) underlying the solution depicted in Fig. 9; there are a significant number of very small values, whose effects are suppressed by the a priori noise estimate.

The region depicted in Fig. 7 occupies an area approximately 36° of longitude by 15° of latitude. To explore the sampling properties, we sought a Fourier fit: $k = (-2\pi/6^\circ)$ to $(2\pi/6^\circ)$ in steps of $(2\pi/36^\circ)$, $l = 0$ to $(2\pi/6^\circ)$ in steps of $(2\pi/15^\circ)$, at frequencies 0, $\pm(2\pi/4.25)$, $\pm(2\pi/5.7)$ $\pm$ $(2\pi/8.5)$, $\pm(2\pi/17.0)$ days. The spatial resolution is dictated by the zonal and meridional dimensions of the region. This region of frequency/wavenumber space is far smaller than can hypothetically be fit given a basic sampling interval of 7 km along track. The additional portion of the frequency/wavenumber space which is hypothetically determinable can be thought of as being suppressed by a priori assumption, and the analysis that follows must be thought of, for that reason, as somewhat optimistic. (If the suppressed region has nonzero amplitudes the uncertainty will be larger than we are going to find.) To make the volume of numbers manageable, the along track data was averaged over a distance of approximately 210 km to give a distribution with half the data of Fig. 7a, leaving 249 data points in about 650 unknown Fourier coefficients (a number of indeterminate Fourier coefficients were carried along for bookkeeping simplicity). This formal underdeterminism could be removed by reducing the averaging interval, thus yielding more observation points than the Fourier fit unknowns; but for the reasons already alluded to, nothing much is gained by doing so, the extra points being very closely spaced, and hence corresponding to very small singular values.

The error covariance matrix needs to be displayed in two wavenumber dimensions and one frequency dimension, which is difficult. In the present case, there are few major sidelobes to the covariance (little correlation of errors in the different coefficients) and the major part of the results can be obtained from the diagonal elements alone. (But there *are* some significant sidelobes and for complete understanding, one needs to examine the complete 702 by 702 resolution matrix. An attempt at displaying it is found in Fig. 8.) Figure 9 displays contours of the solution variance at several of the frequencies listed above, for the cosine coefficients, as a function of longitudinal and latitudinal wavenumber (radians/degree). The convention is such that unit variance means that the resulting estimate is completely indeterminate, and 0 that it is fully determined with no uncertainty (this latter limit is impossible to reach in the presence of measurement noise).

If a variance exceeding 0.5, that is 50% of the solution variance, is arbitrarily defined as beyond the limit of useful information, certain conclusions can be drawn. Consider Fig. 9e, the variance of the Fourier coefficients at 0 frequency.

The spatial mean [(0,0) wavenumber], and the highest wavenumbers are the only ones which can be determined with any confidence under the assumptions we have made. Because of the irregular data distribution, and the assumption of noise in the observations, the system must rely on comparatively poorly determined data differences to distinguish these wavenumbers at zero frequency from those at other frequencies, making the result statistically uncertain. Figure 10 displays the way two pure cosines would be sampled in practice. The extremely irregular way in which these two very regular waves are sampled is readily apparent, and suggests the difficulty the Fourier expansion has in separating them from each other and all other waves present. The noise level can be reduced somewhat by using a second 17-day repeat, but the gain would not necessarily be a factor of 2 reduction in variance: the wavenumber domain would not increase, but the number of frequencies requiring resolution would double from the initial one cycle per 17 days, to one cycle per 34 days, and thereby require precisely twice as many Fourier coefficients to calculate as well.

The results just described are insensitive to the assumption about the signal to noise ratio; when the noise variance was raised to 10% of the Fourier coefficient variance, the diagonals of the error covariance changed by only a few percent. The square of the singular values of the coefficient matrix A are displayed in Fig. 11; there is a long tail of small singular values. Evidently the 0.1% noise variance is sufficient to suppress most of these small values and hence the major part of the solution variance. Owing to the small slope, the result will be insensitive to the precise noise variance. Most of the solution uncertainty derives from the "missing" data owing to the irregular distribution and the noise is only a secondary problem.

Some physical insight into the results can be gained by considering a plane wave of wavenumber (k, l) as it would be sampled along a single, repeating arc. Suppose the wave period were exactly 17 days, so that aliasing in time would make a sampled wave of apparently infinite period. If (k, l) is such that there is spatial structure along the arc, then the Fourier analysis will "assign" the wave to the *band* of wavenumbers having the same projection onto the arc. So for example, if the wavenumber vector is orthogonal to the arc, the wave will appear to have both zero frequency and zero wavenumber and the along-arc spatial averaging cannot remove the alias into the overall (space and time) mean. Thus one anticipates a tendency for waves with crests oriented along the ascending or descending arcs to alias into zero wavenumber. The story, of course, is not this simple, because we do have a sequence of arcs, spaced out in time over the 17 days both ascending and descending, and a 17 day wave need not necessarily alias at all. In general, however, waves with orientation perpendicular to the ascending or descending arcs definitely tend to be less well determined than those tending to lie along them. The situation is potentially very serious when the usual simple averages along a single arc are used to generate a "mean" sea surface.

b. Hypothetical 17-day repeat

The results of the last sections represent a real case, the actual coverage from a real mission, involving irregular coverage not only because of dropped data, but because of the region of interest is partly land. For studying the open ocean, one might prefer to see the results for a more regular area, and under the optimistic

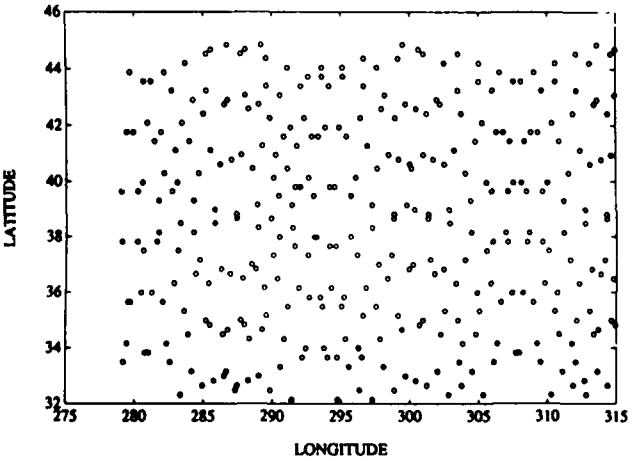


FIG. 12. Data distribution from a hypothetical perfect GEOSAT in a region of no land coverage. Data distribution points are shown at a time interval of 0.055 × orbit period, where the orbit period is about 101 minutes.

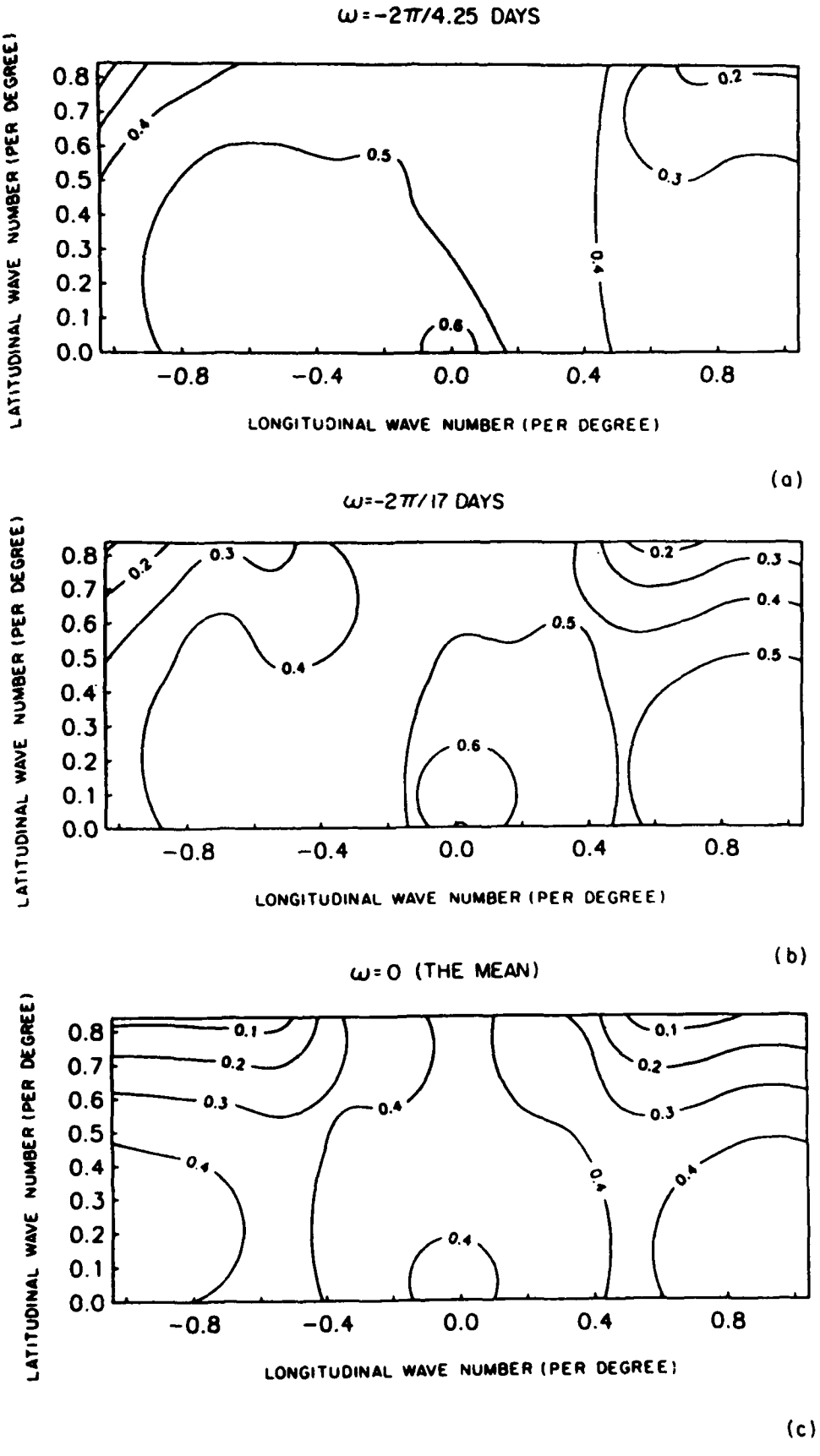


FIG. 13. Same as Fig. 9, except for the hypothetical data distribution of Fig. 12. Results are very similar to those for Fig. 9, except for an overall small reduction in variance owing to the presence of more data points with perfect coverage.

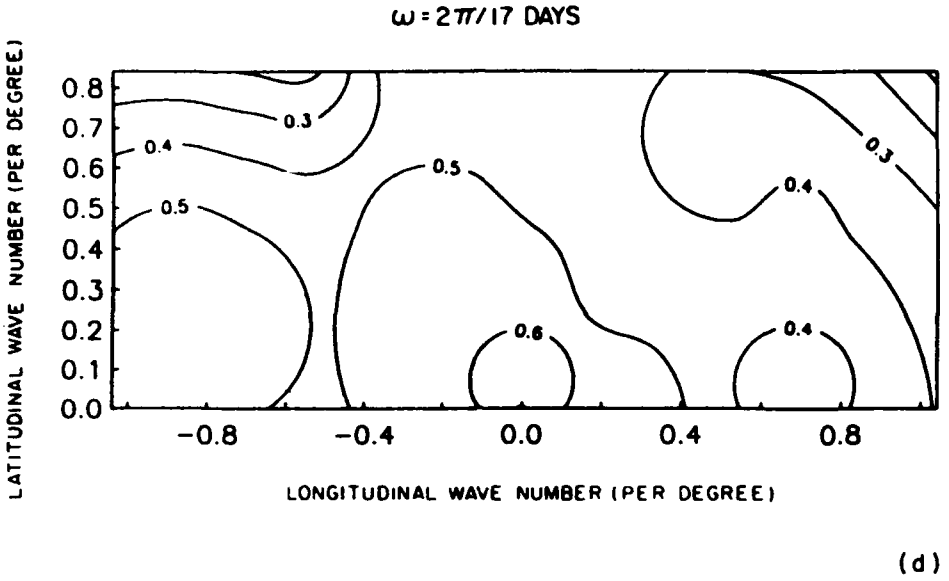


FIG. 13. (*Continued*)

assumption that no data are missing. We start with GEOSAT in a wholly oceanic region with the coverage computed hypothetically, from knowledge of the GEOSAT orbit alone, as depicted in Fig. 12. The same 10^4 to 1 signal to noise ratio is assumed to apply and the samples were placed at intervals of 0.055 orbital periods as depicted.

Some of the result is shown in Fig. 13 for a subset of all the periods which were determined. The qualitative features seen for the real GEOSAT data remain unchanged; the overall variance is somewhat reduced, principally because the absence of land in the region, and the assumption of good data everywhere, means that more data points were available for determining the same number of Fourier coefficients.

c. *Hypothetical 10-day repeat*

The planning orbit for the TOPEX/POSEIDON mission is a 10-day repeat with an orbital inclination of 63.1°. The 10-day period was adopted as a compromise to minimize temporal aliasing while at the same time producing a dense enough ground-track to observe most of the spatial structure of the large-scale ocean surface topography. What are the sampling characteristics?

The 10-day repeat occurs in about 130 orbits, and thus during one repeat cycle, fewer observation points are laid down within the given area than in the 17-day repeat which requires about 245 orbits. If the number of Fourier coefficients were held fixed, there would be an automatic relative increase in solution variance owing simply to the greater underdeterminism. Thus to obtain a direct comparison of the utility of the 10-day repeat with the 17-day repeat, the 10-day repeat was continued over the same number of orbits (245) as the 17-day repeat, or approximately 18 days. The data distribution is as shown in Fig. 14. The frequency-wavenumber domain was kept the same as in Fig. 13, and the new result is displayed in Fig. 15.

Generally, the overall variance is reduced—i.e., the more frequent temporal coverage more than compensates for the less dense surface coverage. The spatial domain, regarded as statistically homogeneous, is well covered in either case—because of the diamond-pattern of ascending or descending passes, all longitudinal and

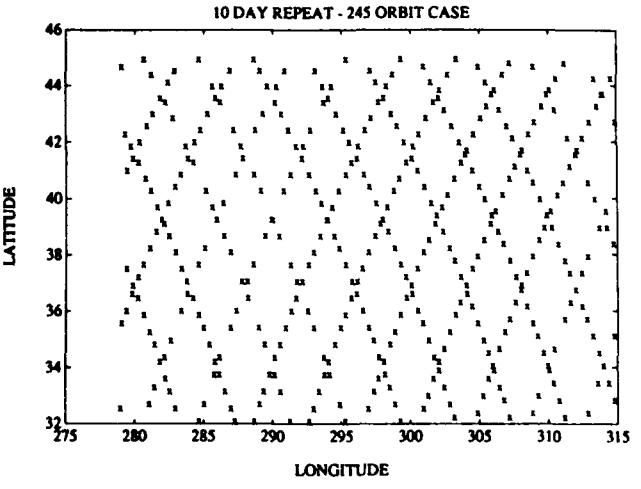


FIG. 14. Data coverage in a hypothetical TOPEX/POSEIDON orbit, where all data are obtained, and there is no land coverage. Assumed period was 10 days, and the coverage shown is that after 18 days, to make the number of data points approximately the same as in Figs. 12 and 13. Again, time interval is 0.055 × orbit period, where the orbit period is about 112 minutes.

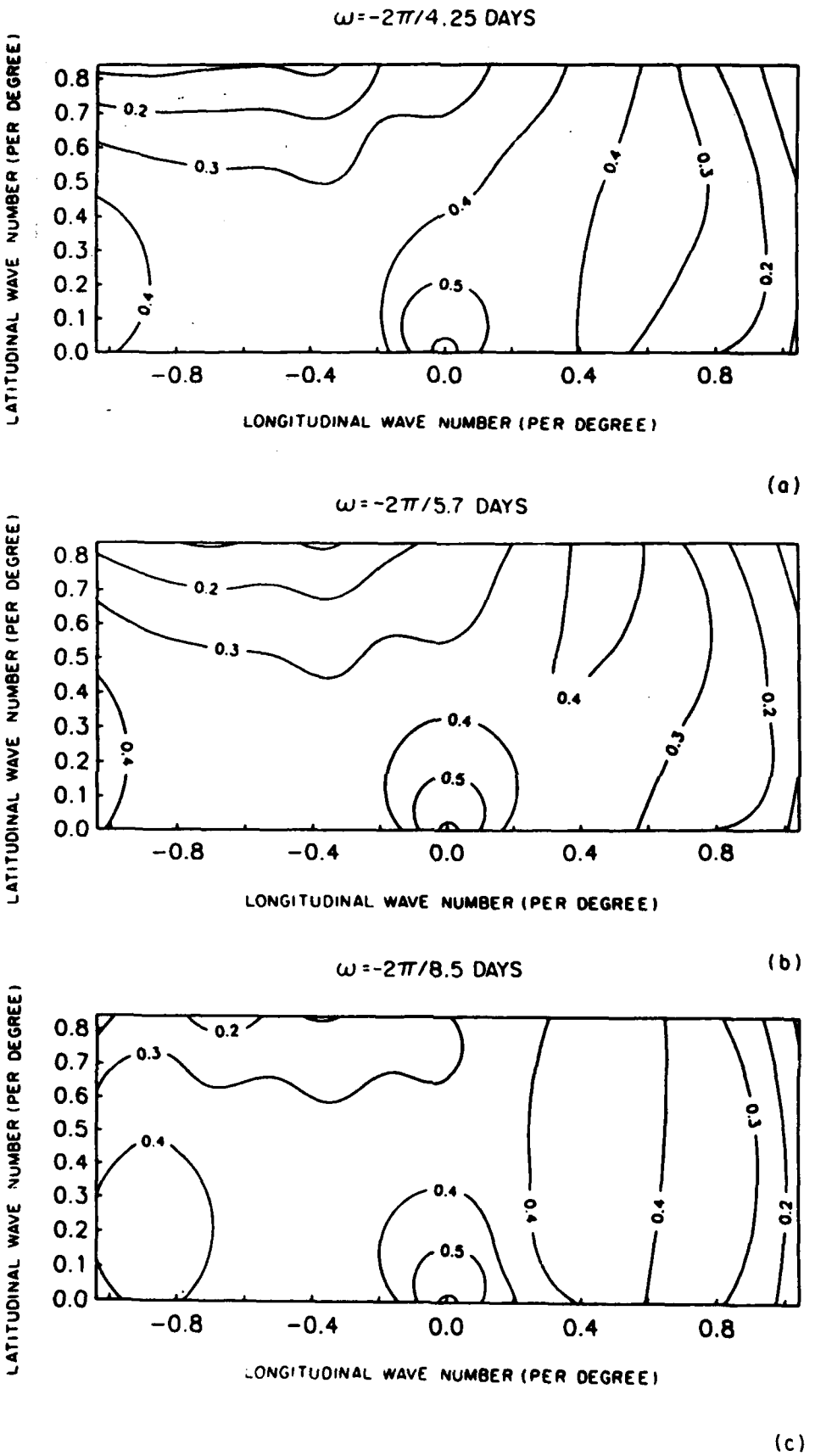
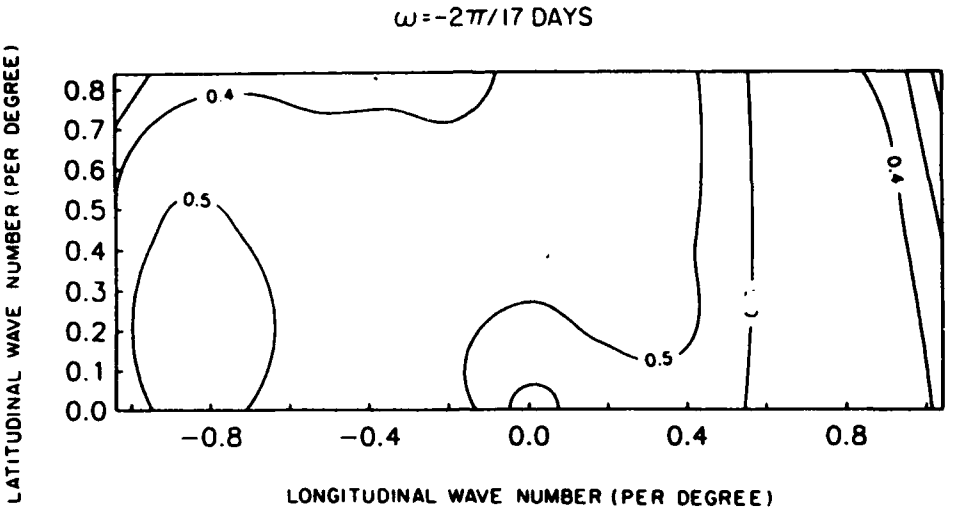
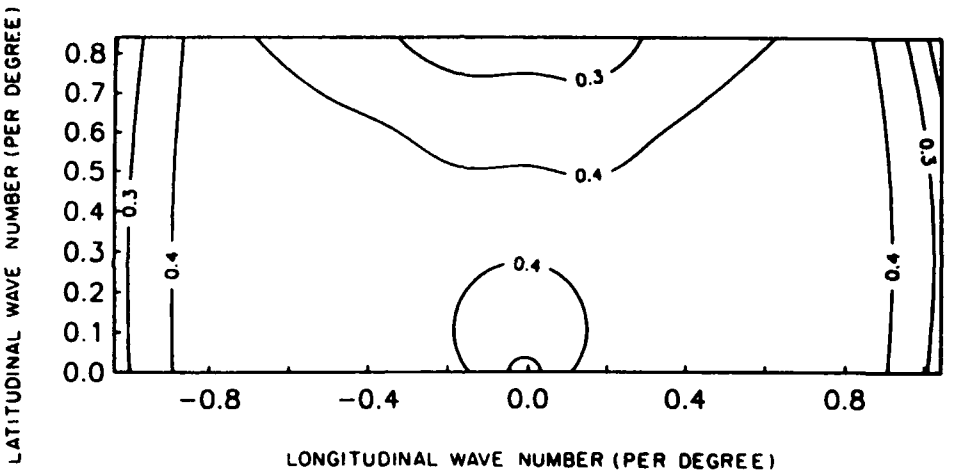


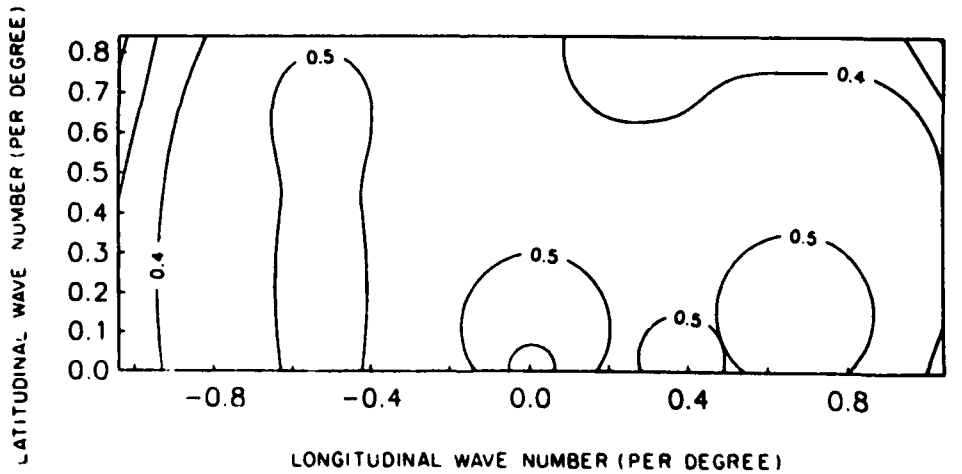
FIG. 15. Cosine coefficient error variance for hypothetical TOPEX/POSEIDON 10-day orbit coverage as shown in Fig. 14. The more frequent temporal coverage greatly reduces the unusable regions (defined as 0.5 times the solution variance) considerably compared to the 17-day repeat.



(d)



(e)



(f)

FIG. 15. (*Continued*)

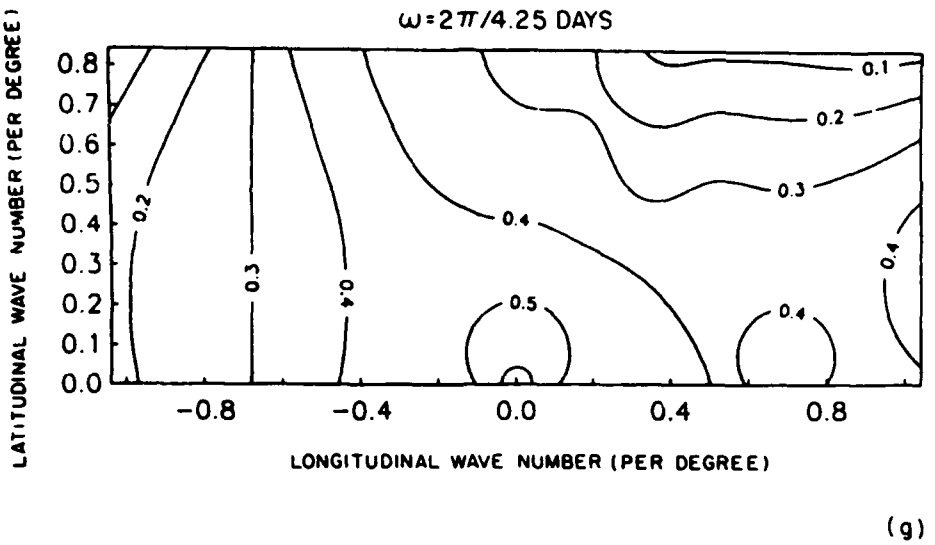


FIG. 15. (*Continued*)

latitudinal scales are observed, and the denser spatial coverage from the 17-day repeat is inadequate compensation for the lost time coverage. Thus the 10-day repeat produces more stable estimates.

6. Final remarks

A major problem is the great difficulty in displaying the complex interdependencies of the uncertainties of frequency/wavenumber composition in three dimensions. The information contained in the full covariance matrices (or the separate resolution and covariance matrices of the SVD) needs to be used in any quantitative application of satellite data. In particular, failure at least to be aware of aliases in a specific analysis can lead to grossly misleading results and should be a source of concern to anyone using such data.

The estimation procedure employed here can be modified simply to take account of the actual nonwhite nature of the observational noise—particularly the strong correlations introduced by the orbital error. But because there is reasonable expectation that for TOPEX/POSEIDON these errors will be reduced to a fraction of what they are today for GEOSAT, we postpone this extra analysis until such time as a clearer specification of the true error can be made.

Within its restricted terms of reference, this study suggests that the 10-day repeat is preferable to a 17-day repeat because of the reduced solution error variances associated with the former orbit. But the main purpose of this paper is less to draw a firm conclusion about spacecraft mission design than to set a context for the discussion.

As with classical sampling theorems, the assumption was that all frequencies and wavenumbers are, a priori, equally likely to exist. If one can demonstrate, or is willing to assume, that entire bands in frequency/wavenumber space are empty or nearly so, the conclusions drawn here could be totally reversed.

A number of studies are underway for future altimetric mission, using data assimilation into dynamical models (but it should be reemphasized that the sampling issues raised here would apply to any measurement made from an orbiting spacecraft). For any particular model, it is possible that a long repeat cycle with its denser spatial coverage does a better job of keeping the model computation closer to the "truth." One would have to presume that in such a model many of the frequencies and wavenumbers which cause indeterminacy problems in the present study are absent. They could be absent for a number of reasons—because in a nearly linear model the forcing fields do not contain those bands, or in a nonlinear model because the nonlinearities do not fill those bands, or both. As long as one is confident that the ocean has the same spectral properties as the model, then experiment design based on the model would be most appropriate. But the dangers inherent in such a strategy will be obvious. Spacecraft are global measuring devices, and demonstrating frequency/wavenumber consistency of a model with the global ocean would be a formidable and probably impossible task at the present time, given what is known about the world ocean, and the limitations on global models. With any new instrument, the normal sampling strategy is a conservative one—that is it errs on the side of oversampling—because sampling theorems tend to be rather unforgiving. Determining whether the 10-day orbit is really a more conservative choice than a longer period repeat will require more elaborate studies than the one produced here (if nothing

else, much greater computing power must be brought to bear on the problem). But the language of the discussion must be that of sampling theory.

Acknowledgments. Supported in part by the Jet Propulsion Laboratory, Contract 958125, the National Aeronautics and Space Administration Contract NAGW-1048, and by the Office of Naval Research Contract M00014-86-K-0751. I am grateful to Dr. V. Zlotnicki for making the orbit computation code available to me. I thank the referees for detailed and careful reviews, but especially the one who said there was too much tutorial material, and the one who said there was not enough—leading me to think about the issue.

REFERENCES

Bracewell, R. N., 1978: *The Fourier Transform and Its Applications,* second ed. McGraw-Hill, 444 pp.

Freeman, H., 1965: *Discrete-Time Systems. An Introduction to the Theory.* John Wiley, 241 pp.

Hamming, R. W., 1973: *Numerical Methods for Scientists and Engineers.* Dover, 721 pp.

Jerri, A. J., 1977: The Shannon sampling theorem—its various extensions and applications: A tutorial review. *Proc. IEEE,* **65,** 1565–1596.

Lawson, C. L., and R. J. Hanson, 1974: *Solving Least-Squares Problems.* Prentice-Hall, 340 pp.

Liebelt, P. B., 1967: *An Introduction to Optimal Estimation.* Addison-Wesley, 273 pp.

Petersen, D. P., and D. Middleton, 1962: Sampling and reconstruction of wavenumber-limited functions in *N*-dimensional Euclidean space. *Inform. and Control,* **5,** 279–323.

Wiggins, R. A., 1972: The general linear inverse problem: Implication of surface waves and free oscillations for earth structure. *Rev. Geophys. Space Phys.,* **10,** 251–285.

Wunsch, C., 1978: The North Atlantic general circulation west of 50°W determined by inverse methods. *Rev. Geophys. Space Phys.,* **16,** 583–620.

——, 1989: Tracer inverse problems. *School on The Ocean Circulation and Geochemical Tracer Transport,* Les Houches, France, D. L. T. Anderson and J. Willebrand, Eds., North-Holland, 1989, in press.

Accesion For
NTIS CRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1 | 21