Determining the Ocean Circulation and Improving the Geoid from Satellite Altimetry

JOHN C. MARSHALL

Atmospheric Physics Group, Department of Physics, Imperial College of Science and Technology, London SW72BZ

(Manuscript received 23 July 1984, in final form 4 December 1984)

ABSTRACT

The combined problem of determining the ocean circulation and improving the geoid from satellite altimetry is formulated and studied. Minimum variance estimation is used to form optimum estimates of the ocean topography and the geoid. These estimates are a function of the altimeter observations, prior knowledge of the ocean circulation and prior knowledge of the geoid. Particular emphasis is placed on the use of a dynamical ocean model as a source of *a priori* oceanographic information capable of discriminating between geoid errors and ocean topography. The technique is illustrated in a simulation study of Gulf Stream variability, in which an ocean topography, degraded by noise representing the uncertainty in a gravimetric geoid, is reconstructed by assimilation into an ocean model. At the same time an improved estimate of the geoid is made.

1. Introduction

A satellite altimeter flying typically at a height of $\sim$1000 km, measures the distance between itself and the sea surface to centimeter accuracy. Given an independent measurement of the position of the satellite one can determine the sea surface h_o, relative to a suitable reference ellipsoid—see Fig. 1. The shape of this surface is of interest because it contains information about the geoid and the ocean circulation. If the ocean were at rest then its surface would be the equipotential defining the geoid which, due to inhomogeneities in the earth, undulates about the reference ellipsoid by an amount $h_g \sim 100$ meters. But because the ocean moves, the actual sea surface deviates from the geoid by an amount $h = h_o - h_g$, $\sim$1 meter. This small departure from the geoid interests the oceanographer for it contains information on tides and geostrophically balanced surface currents. The object of the present study is the identification of that part of the sea-surface elevation relative to the geoid associated with the surface geostrophic flow (which we will call the ocean topography, OT).

The limiting factor in the use of altimetry for ocean circulation studies will not be so much the determination of h_o (although extra-ordinary precision is required in the instrument and tracking) but rather the uncertainty in the shape of the geoid, for it appears that the errors in presently available geoid models can exceed the OT signal at dynamically interesting scales. Thus the ultimate objective of absolute surface velocity determination from altimetry must await more accurate geoid models.

Prior to the advent of satellite altimetry, the shape of the geoid could only be inferred from satellite orbit calculations and surface gravity data. The trajectory of a satellite is sensitive to the longest wavelengths of the gravity field, and so the errors in satellite geoids increase with degree. Such models cannot be extended with any reliability much beyond degree and order 20 $\times$ 20 (for a discussion of the implications of this for ocean dynamics see Wagner, 1981). But even if a reliable model of this resolution were available, it could only represent wavelengths down to 2000 km, leaving obscured many of the most interesting features of the ocean circulation such as boundary currents and jets. Fine resolution ($\sim$10 km) geoids have been developed for localized regions such as the northwest Atlantic, by incorporating surface gravity measurements (e.g., see Marsh and Chang, 1978). However, large gaps in the gravity coverage remain and global measurements of the earth's gravity field at fine resolution is not a realistic possibility. Until dedicated satellite gravity missions such as the Geopotential Research Mission (GRM) are underway (designed to give 10 cm accuracy down to wavelengths of 100–200 km), altimetry is the only realistic means of determining the fine structure of the geoid. Such altimetric geoids, though, cannot be used for oceanographic purposes because they lump the mean circulation in with the geoid. This is the central difficulty that must be overcome if the full potential of altimetry in ocean circulation studies is to be realized.

A way forward is suggested by the realization that the oceanography and geodesy are inseparably linked when viewed through the eyes of the altimeter, and should therefore be determined simultaneously. Although present geoids blurr the OT rather severely, the OT will come more into focus as the geoid

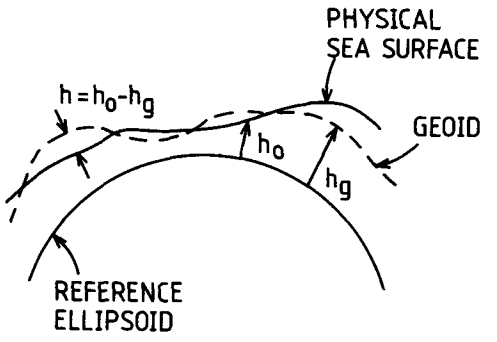

FIG. 1. Schematic diagram showing the sea surface and geoid relative to the reference ellipsoid: h_o, the physical sea-surface; h_g, the geoid and $h = h_o - h_g$, the departure of the sea-surface from the geoid.

becomes known at smaller and smaller scales. It seems reasonable to suppose that if the signal-to-noise ratio is not too small (regarding the OT as signal and the geoid error as noise) one could profitably exploit prior knowledge of the ocean circulation to help separate it from the residual geoid error, and in the process improve the geoid.

The present paper is a contribution to the study of this combined problem of determining the ocean circulation and correcting the geoid from satellite altimetry. It has been discussed in general terms before by Wunsch and Gaposchkin (1980) who couch it in the formalism of linear inverse theory. We use the same technique here, but approach it from a slightly different perspective, regarding the determination of the geoid and OT as an exercise in objective mapping. We adopt the hopefully familiar jargon that goes with "objective analysis," the meteorological manifestation of minimum variance techniques.

Wunsch and Gaposchkin (1980) limit the number of OTs consistent with altimetric data by combining them with hydrographic data. Instead, or perhaps in addition, we suggest that the assimilation of altimetric data into a dynamical ocean model could be a powerful way of discriminating between geoid errors and OT. The geoid is adjusted so that the implied OT is consistent with the prior dynamical constraints of the model.

Such assimilation techniques are also of interest because altimetric missions should not be seen in isolation, but only as part of larger observing campaigns using more conventional measuring techniques. Direct observations of ocean currents, density, surface stress, as well as surface elevation will be available but irregularly distributed in space and time, and each with their differing error characteristics. Perhaps the only way of making sense of this diverse information is through assimilation of data into ocean models which can dynamically link the variables.

In Section 2 we set out the under-determined problem of solving for the ocean circulation and improving the geoid from satellite altimetry. In Section 3 we describe how the dynamical information in an ocean model can be used to discriminate between the geoid and OT signal in the altimetric data. This idea is developed more quantitatively in Section 4 where minimum variance solutions for the OT and geoid are written down and studied. A practical illustration of the technique is presented in Sections 5 and 6 where, in a simulation study of Gulf Stream variability, the OT degraded by noise representing the uncertainty in a gravimetric geoid is reconstructed by assimilating altimetric data into an ocean model.

2. A reference problem

In this section the problem of determining the ocean circulation and improving the geoid from satellite altimetry is defined more precisely. By way of illustration frequent reference is made to a simulation study of Gulf Stream variability (developed further in Sections 5 and 6) in which a series of synoptic maps, degraded by noise representing uncertainty in the gravimetric geoid, is sampled to simulate an altimetric data set. From this noisy data we form an estimate of the true OT and correct for the geoid. Before proceeding it should be emphasised that the approach we adopt to the combined problem has general applicability. The focus on synoptic mapping should be regarded as a nominal one, chosen for illustrative purposes only in an interesting oceanographic context.

An eddy resolving ocean circulation model of Gulf Stream variability is used to generate a sequence of truth OTs at time-steps K on an $I \times J$ finite difference grid: $\mathbf{h}$. The $\mathbf{h}$ is a column vector of dimension $n = I \times J$ containing the OT at the grid points of the model at time K.

It is further supposed that the time variant geoid mapped onto the $I \times J$ grid is $\mathbf{h}_g$. The simulated height of the sea surface relative to the center of the earth is thus

$$\mathbf{h}_{oK} = \mathbf{h}_K + \mathbf{h}_{gK}. \tag{2.1}$$

Having chosen discrete values on an $I \times J$ grid at time K as a representation of $\mathbf{h}_o$, the truth field is sampled to simulate the spatial density and temporal frequency of the tracks laid down by a satellite altimeter in an exactly repeating orbit, thus forming a stream of observations of the sea-surface height $\mathbf{z}_o$:

$$\mathbf{z}_{oK} = \mathbf{A}\mathbf{h}_{oK} + \epsilon_{oK}. \tag{2.2}$$

The observation vectors $\mathbf{z}_o$ at time K have a dimension p, say, where $p \ll n$. The ϵ_o is the error vector in the altimetric measurement (instrumental and tracking). The permutation matrix $\mathbf{A}$ of dimension $(p \times n)$ consists of 0's and 1's and maps the field $\mathbf{h}_o$, onto the observations $\mathbf{z}_o$. It samples $\mathbf{h}_o$ simulating the regular spatial pattern of the (repeated) satellite tracks. For example, Fig. 2a shows an OT which in our

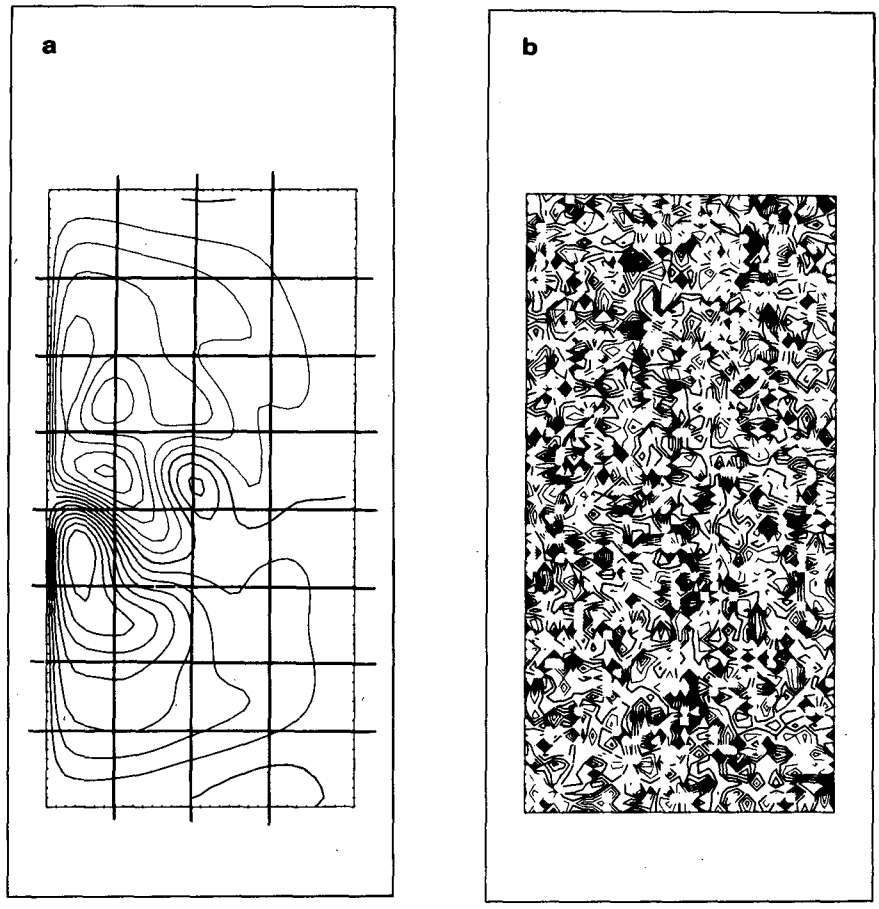

FIG. 2. (a) The six-month time mean truth OT, **h** (contour interval, C.I. = 6 cm) showing two counter-rotating wind-driven gyres in a 1000 × 2000 km basin (thick lines are positive contours, thin lines negative, and the very thick line the zero contour). Simulated altimetric tracks are indicated by the regular pattern of crossing lines, and give a spatial coverage appropriate to a satellite flying in a 7-day repeat orbit at an altitude of 1000 km. (b) The simulated error in a gravimetric geoid ϵ_g (C.I. = 15 cm) taken from a normal distribution with zero mean and standard deviation 30 cm.

simulation study is regularly sampled along the satellite tracks represented by the thick black lines. Equation (2.2) describes the situation in which at any time K there are incomplete and inaccurate measurements of the sea surface height. From these measurements we wish to form the best possible estimates of the OT and the geoid.

However, if the only information available is $\mathbf{z}_o$, then the problem as stated is under-constrained because there are an infinite number of $\mathbf{h}$ and $\mathbf{h}_g$ consistent with the observations: there are not (and never can be) enough measurements to determine $\mathbf{h}$ and $\mathbf{h}_g$ completely ($p \ll 2n$). Mathematically the problem is ill-posed, since there is no $\mathbf{H}$ such that $\mathbf{HA} = \mathbf{AH} = \mathbf{I}$, the identity matrix (the inverse of $\mathbf{A}$ does not exist because $p \neq n$). There is no alternative but to rely on prior statements about the nature of the solutions in deciding which ocean state and geoid is the most likely.

Accordingly (2.2) is augmented with *a priori* estimates (denoted by a caret) of the OT and geoid, $\hat{\mathbf{h}}(-)$, $\hat{\mathbf{h}}_g(-)$ of dimension n i.e., estimates of the geoid and OT that would have been made in the absence of altimetry:

$$\hat{\mathbf{h}}_g(-) = \mathbf{I}\mathbf{h}_g + \epsilon_g, \tag{2.3}$$

$$\hat{\mathbf{h}}(-) = \mathbf{I}\mathbf{h} + \epsilon_h. \tag{2.4}$$

where $(-)$ represents an estimate prior to the arrival of altimeter observations and ϵ_g, ϵ_h are the errors in these *a priori* estimates. For example, $\hat{\mathbf{h}}_g(-)$ could be a nonaltimetric geoid derived from ship gravity measurements, or a geoid from a future satellite gravity

mission. The $\hat{\mathbf{h}}(-)$ could be a dynamic topography derived from hydrographic measurements or a "first guess" provided by a dynamical ocean model.

The problem of determining the ocean circulation and improving the geoid from satellite altimetry can now be stated thus: given

(i) altimeter observations of the sea surface height along repeated tracks $\mathbf{z}_o$, Eq. (2.2) and

(ii) our prior knowledge of the geoid and ocean circulation, $\hat{\mathbf{h}}_g(-)$ and $\hat{\mathbf{h}}(-)$, Eqs. (2.3) and (2.4) respectively,

make an estimate of the OT over the $I \times J$ grid and improve the geoid.

This, in words, describes the general problem of solving for the ocean circulation and the geoid simultaneously from satellite altimetry. Although easy to state it is not obvious how to proceed both because of doubts about the most appropriate forms for $\hat{\mathbf{h}}(-)$ and $\hat{\mathbf{h}}_g(-)$, and once chosen uncertainty in the quantification of their errors. In practice most investigators have approached the oceanography and geodesy separately without exploiting prior knowledge of ocean circulation.

a. *Solving for* $\mathbf{h}$ *and* $\mathbf{h}_g$ *separately*

If we are content to form an estimate of the OT along the satellite tracks only, then the best available non-altimetric geoid can be subtracted from the sea surface height measured by altimetry to compute (solving for as many unknowns as there are measurements)

$$\begin{aligned}\hat{\mathbf{h}}^{\text{track}} &= \mathbf{z}_o - \mathbf{A}\hat{\mathbf{h}}_g(-) \\ &= \mathbf{A}(\mathbf{h} + \boldsymbol{\epsilon}_g) + \boldsymbol{\epsilon}_o \end{aligned} \tag{2.5}$$

This is the approach taken by Cheney and Marsh (1981) for example, who subtracted the GEM 8 gravimetric geoid from Seasat altimetric sea surface heights in the northwest Atlantic (for a review of the Seasat mission see Born *et al.*, 1979). The problem with Eq. (2.5) is that even with perfect altimetry $\boldsymbol{\epsilon}_o = 0$, the ocean signal is blurred by geoid noise $\boldsymbol{\epsilon}_g$. In the absence of additional information, it is impossible to decide which part of $\hat{\mathbf{h}}$ is ocean signal and which part is the geoid error.

An improved estimate of the geoid can be made by averaging $\mathbf{z}_o$ over many passes, ignoring the contribution from the mean circulation, and combining it with the gravimetric geoid

$$\left.\begin{matrix}\overline{\mathbf{z}_o^t} \\ \hat{\mathbf{h}}_g(-)\end{matrix}\right\} \begin{matrix}\text{to give a new improved} \\ \text{geoid } \hat{\mathbf{h}}_g(+),\end{matrix} \tag{2.6}$$

where (+) represents our estimate after the arrival of observations. Equation (2.6) is an over-constrained problem with enough information to define all the parameters ($n + p$ observations n unknowns) which can be approached by least-squares procedures. For example, Rapp (1983) uses a Seasat altimetric sea surface to fill in the fine-structure of a nonaltimetric geoid. However, the geoid computed as in (2.6), although much improved by altimetry, cannot be used for absolute geostrophic velocity determination [for example, to provide the geoid in (2.5)] because the mean circulation is lumped in with the geoid.

The approach represented by Eqs. (2.5) and (2.6) illustrate the fundamental limitation for ocean circulation purposes that will always be present if the geodesy and the oceanography are tackled separately.

b. *Solving for* $\mathbf{h}$ *and* $\mathbf{h}_g$ *simultaneously*

The drawback of the approach symbolized in Eqs. (2.5) and (2.6) is that the geodesy and oceanography are decoupled from one another because prior knowledge of the ocean circulation, $\hat{\mathbf{h}}(-)$, Eq. (2.4) is not exploited.

Suppose, however, that full use of this knowledge is made by forming "observations" of the OT and geoid

$$\left.\begin{aligned}\mathbf{z}_g &= \mathbf{z}_o - \mathbf{A}\hat{\mathbf{h}}(-) \\ \mathbf{z}_h &= \mathbf{z}_o - \mathbf{A}\hat{\mathbf{h}}_g(-)\end{aligned}\right\} \tag{2.7}$$

which are then combined with the prior estimates $\hat{\mathbf{h}}_g(-)$ and $\hat{\mathbf{h}}(-)$ thus:

$$\left.\begin{matrix}\hat{\mathbf{h}}_g(-) \\ \mathbf{z}_g\end{matrix}\right\} \begin{matrix}\text{to give a new improved} \\ \text{geoid } \hat{\mathbf{h}}_g(+),\end{matrix} \tag{2.8a}$$

$$\left.\begin{matrix}\hat{\mathbf{h}}(-) \\ \mathbf{z}_h\end{matrix}\right\} \begin{matrix}\text{to give a new improved} \\ \text{OT } \hat{\mathbf{h}}(+).\end{matrix} \tag{2.8b}$$

Now in contrast to Eqs. (2.5) and (2.6) prior knowledge of the ocean circulation is to put to use in the estimation of the ocean topography and the geoid. Note that Eqs. (2.8a) and (2.8b) reduce to Eqs. (2.6) and (2.5) respectively when $\hat{\mathbf{h}}(-)$ is set to zero in Eq. (2.7).

The use of $\hat{\mathbf{h}}(-)$ to constrain our estimates recognizes that altimetric information should not be considered in a vacuum. In addition to altimetry there are many independent oceanographic measurements and much understanding of the dynamics of ocean currents, their space and time scales. It is through $\hat{\mathbf{h}}(-)$ that such diverse information can be brought to bear in the analysis of the altimeter data.

Equations (2.3) and (2.4) are best regarded as "*a priori* constraint" equations or "virtual observations" which turn the under-determined problem (given $\mathbf{z}_o$ there are an infinite number of possible $\mathbf{h}$ and $\mathbf{h}_g$) into an over-determined one which can be tackled using standard statistical techniques. Writing these prior estimates as in Eqs. (2.3) and (2.4) shows explicitly that they can be thought of as observations, just as in Eq. (2.2). (Rodgers, 1976, has a particularly lucid discussion on the nature of *a priori* information in inverse problems when viewed from this perspec-

tive: see also Jackson, 1979). Indeed, all *a priori* statements (dynamical constraints, smoothness criteria etc.) can be treated as if they were observations by writing them down in the form Eq. (2.2). This prior knowledge of the geoid and ocean circulation restricts the class of admissible solutions consistent with the actual observations. Of course, the solutions to the new problem Eq. (2.8) can never be unique functions of the data, but will also depend on the nature of the *a priori* constraints.

The use of oceanographic information in conjunction with altimetry has been discussed before by several authors. Wunsch and Gaposchkin (1980) describe in general terms the problem of combining hydrography with geodesy and altimetry. The approach is implicit in the work of Cheney and Marsh (1981) who compare the relative merits of three geoid models by making use of *in situ* oceanographic measurements.

In the next section we discuss how a geoid error might be corrected for by adjusting it so that the implied OT is consistent with prior dynamical constraints contained in an ocean model.

3. Ocean models as a source of *a priori* information

The ocean model is a particularly valuable source of *a priori* information since it is a concise statement of our dynamical understanding of the ocean circulation. Of course this understanding is incomplete and so the models are far from perfect. But, to the extent that a particular model is a faithful representation of important aspects of the dynamics it will be a more or less useful analysis tool. A sensible and powerful direct application of models is to use them to provide $\hat{\mathbf{h}}(-)$ and thus assist in (i) the interpolation of data in space and time and (ii) the discrimination between geoid errors and ocean topography.

a. Interpolation

The interpolation problems in the ocean are so acute that all our prior knowledge of ocean dynamics must be brought to bear in the analysis of data if we are to make the best use of the data we have. The problem of constructing an analysis from incomplete data is familiar to meteorologists who handle their global data sets by assimilation into global dynamical models. Oceanographers have little similar experience, both because there are only limited operational requirements, but also because prior to the advent of altimetry there has been no observing system with sufficient time and space resolution. Altimetry, together with proposed oceanographic observing programs such as the World Ocean Climate Experiment (WOCE) and the Tropical Ocean-Global Atmosphere (TOGA) experiment is encouraging modellers to contemplate assimilation of data into ocean models.

One would suppose, for example, that a dynamical model of geostrophic eddy variability would be a useful tool in the analysis of altimeter data for synoptic mapping purposes, advecting information between satellite tracks and projecting information from one satellite pass to the next. In this way the past history of the data set can be used to limit the number of possible fields which fit the data, since the fields evolve according to dynamical laws. The best way of achieving this is to assimilate the observations into an ocean model and use the model to carry the estimate of the ocean topography:

$$\hat{\mathbf{h}}_{K+1} = \mathbf{L}\hat{\mathbf{h}}_K + \epsilon_m, \tag{3.1}$$

where $\mathbf{L}$ is a finite difference operator which advances the ocean topography estimate forward one time step, and ϵ_m represents dynamical and physical processes not described or resolved by the model. The following continuous updating strategy can be envisaged:

given an initial estimate of the state of the OT, $\hat{\mathbf{h}}_{K=o}$, Eq. (3.1) is integrated forward in time to provide an estimate of the state of the ocean which is imperfect because of imperfections in the initial state and the model physics. When observations $\mathbf{z}_h$ of $\mathbf{h}$ become available, they are combined with a "first guess" to the true field provided by the ocean model $\hat{\mathbf{h}}(-)$ to give a new improved estimate of ocean circulation $\hat{\mathbf{h}}(+)$, Eq. (2.8b). The ocean model is then integrated on from $\hat{\mathbf{h}}(+)$ to the next observation time. The procedure relies on there being some useful information in the first-guess field.

The most enlightening way of viewing the use of a model in this way is as a measuring instrument providing "virtual measurements" which are of the same nature as the real observations, but which differ only in their error characteristics.

b. Separation of geoid error from ocean topography

How can the dynamical information in an ocean model be used to separate the OT signal from other components such as geoid and instrument error?

First, it is worth reminding ourselves that no amount of mathematical analysis can make up for a basic lack of information. If there is no information on the error structures of $\hat{\mathbf{h}}(-)$ and $\hat{\mathbf{h}}_g(-)$ then they are not useful in constraining the solution; we must have knowledge of the amplitude and spatial scale of their errors. What is more, progress can only be made if the error in the geoid model and the error in the OT estimate reside at sufficiently different scales. Fortunately there is reason to suppose that this will be the case. On scales smaller than the height–height correlation scale of the OT (set by the baroclinic Rossby radius of deformation L_ρ = O(100 km) for the geostrophic eddy field, see for example Freeland and Gould, 1976) the error in the first guess, $\hat{\mathbf{h}}(-)$, is likely to be small because there is little structure in

the OT at these scales. But it is precisely at such small scales that most of the geoid error will be concentrated because future gravimetric satellite missions will probably always be limited to determining the global geoid on spatial scales of 100–200 km and larger. On the other hand, one might expect the error in a "first guess" provided by an ocean model to reside predominantly at scales greater than the correlation scale of the OT, where the geoid will be much better known. Thus at large scales we can envisage using the geodesy and altimetry to improve the ocean model, while at smaller scales where the altimetry and the ocean model are more reliable, they can be used to improve the poorly known geoid.

Although the accuracies of local gravimetric geoids are notoriously difficult to assess, it is questionable whether such a scale separation presently exists. The errors in the best available local gravimetric geoids (e.g., the 5′ GEM 8 northwest Atlantic geoid) probably encroach rather far on oceanographically interesting scales (see Cheney and Marsh, 1981). However, as our knowledge of the geoid improves and the errors are driven down to smaller and smaller scales, the OT signal will become more and more into focus. Our approach relies on there being some useful information in the geoid, i.e., the blurring of the OT by geoid errors is not too great.

Suppose, for example, that a gravimetric geoid of the northwest Atlantic is in error by several meters on a scale of several tens of kilometers over the Puerto Rico trench where the geoid has much small-scale variability. The difference between an altimetric surface and the geoid estimate, the observation of the OT, would thus contain a large geoid error. If this were assimilated into an ocean model it seems reasonable to suppose that the model would "realise" that a geoid error of such an amplitude and scale is unlikely to be a geostrophically balanced pressure gradient and could thus be used to compensate for it: there would be a systematic difference between the model's first guess $\hat{\mathbf{h}}(-)$ and the observation $\mathbf{z}_h$, and this difference could be attributed to a geoid error. In this way the ocean model may be used to help correct the geoid. In effect the geoid is adjusted so that the implied OT is consistent with the prior dynamical statements of the model.

Viewing the geoid error as a bias in the OT observations in this way, has close parallels with current research in meteorological data analysis. For example, Hollingsworth and Arpe (1981) map the mean difference between the 500 mb height field analysis and radiosonde observations and describe how over northern Europe it reflects national boundaries. They suggest that this bias is introduced by the inhomogeneity of radiosonde types over Europe. This is an interesting example of how a dynamical model can be used to assess the quality of the observing network, if the systematic errors in the model and the observations lie at different scales. In the same way, a dynamical ocean model has the potential of becoming an extremely useful tool in discriminating between OT signals and the bias introduced by geoid errors.

It should be clearly stated though, that there will certainly be oceanic regions with strong variability where there may be no such scale separation between geoid errors and systematic model errors. For example, the Gulf Stream flowing over the New England Seamounts shows strong topographic control and so here systematic model errors will almost certainly be *correlated* with the topography and hence the geoid, making any discrimination here between geoid errors and model errors problematical.

The rather qualitative notions discussed here will be made more quantitative in Section 4 where we write down the minimum variance solution of Eq. (2.8) (which weight the first guess and observation with respect to their error covariances) and study how its response as a function of scale is controlled by the choice of error covariance matrices. A practical demonstration of the approach will be given in Section 6.

4. Minimum variance estimation

a. Weighting with respect to error covariances

Standard statistical methods are available for combining independent measurements of a field. Here we use minimum variance estimation which is useful if we have information on, or are willing to make assertions concerning, the error structure of the measurements. There is an extensive bibliography on linear estimation theory but in the present context papers by Ghil *et al.* (1980), Jackson (1981) and particularly Rodgers (1976) were found to be helpful.

Given two independent measurements of a vector $\mathbf{x}$, $\mathbf{x}_1$ and $\mathbf{x}_2$ where

$$\mathbf{x}_1 = \mathbf{x} + \epsilon_1$$

$$\mathbf{x}_2 = \mathbf{x} + \epsilon_2$$

together with covariance information on their error structure

$$\mathbf{S}_1 = \mathbf{Exp}(\epsilon_1 \epsilon_1^{\mathrm{T}}); \quad \mathbf{S}_2 = \mathbf{Exp}(\epsilon_2 \epsilon_2^{\mathrm{T}}),$$

(where **Exp** is the expectation operator or the ensemble average, and T is the transpose) the optimum estimate in the sense of least-squares weights $\mathbf{x}_1$ and $\mathbf{x}_2$ inversely with their error covariance matrices:

$$\left.\begin{aligned} \hat{\mathbf{x}} &= \hat{\mathbf{S}}(\mathbf{S}_1^{-1}\mathbf{x}_1 + \mathbf{S}_2^{-1}\mathbf{x}_2) \\ \text{with error covariance} & \\ \hat{\mathbf{S}} &= (\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1})^{-1} \end{aligned}\right\}. \quad (4.1)$$

Eq. (4.1) is a straightforward generalization to vectors of the familiar combination of independent measurements of a scalar. It has been derived on numerous occasions and applied in many diverse fields. In the combined problem of Section 2 there are three pieces of information: altimeter measurements $\mathbf{z}_o$, a geoid model $\hat{\mathbf{h}}_g(-)$ and an ocean model $\hat{\mathbf{h}}(-)$, Eqs. (2.2), (2.3) and (2.4) respectively. From these we form:

two "measurements" of the geoid

the "first guess" $\hat{\mathbf{h}}_g(-)$

with error covariance $\mathbf{S}_{h_g}(-)$

and the "observation" $\mathbf{z}_g = \mathbf{z}_o - \mathbf{A}\hat{\mathbf{h}}(-)$

with error covariance $\mathbf{S}_{z_g} = \mathbf{A}\mathbf{S}_h(-)\mathbf{A}^T + \mathbf{S}_{z_o}$ (4.2)

and

two "measurements" of the OT

the "first guess" $\hat{\mathbf{h}}(-)$

with error covariance $\mathbf{S}_h(-)$

and the "observation" $\mathbf{z}_h = \mathbf{z}_o - \mathbf{A}\hat{\mathbf{h}}_g(-)$

with error covariance $\mathbf{S}_{z_h} = \mathbf{A}\mathbf{S}_{h_g}(-)\mathbf{A}^T + \mathbf{S}_{z_o}$. (4.3)

In Eqs. (4.2) and (4.3) it has been assumed that errors in the altimetric observations are uncorrelated both with errors in the geoid model and errors in the ocean model.

So, by analogy with Eq. (4.1) the optimum estimates can be written

$$\hat{\mathbf{h}}_g(+) = \mathbf{S}_{h_g}(+)(\mathbf{S}_{h_g}^{-1}(-)\hat{\mathbf{h}}_g(-) + \mathbf{A}^T\mathbf{S}_{z_g}^{-1}\mathbf{z}_g)$$

which has error covariance

$$\mathbf{S}_{h_g}(+) = (\mathbf{S}_{h_g}^{-1}(-) + \mathbf{A}^T\mathbf{S}_{z_g}^{-1}\mathbf{A})^{-1}$$

and

$$\hat{\mathbf{h}}(+) = \mathbf{S}_h(+)(\mathbf{S}_h^{-1}(-)\hat{\mathbf{h}}(-) + \mathbf{A}^T\mathbf{S}_{z_h}^{-1}\mathbf{z}_h)$$

which has error covariance

$$\mathbf{S}_h(+) = (\mathbf{S}_h^{-1}(-) + \mathbf{A}^T\mathbf{S}_{z_h}^{-1}\mathbf{A})^{-1} \quad (4.4)$$

Thus, given some prior knowledge of the ocean circulation and geoid and their uncertainty, when the altimeter measures the sea surface height (sum of geoid and ocean topography), Eq. (4.4) gives a new updated estimate of the OT and geoid, and a quantification of the error in these estimates.

Although writing the best estimates in the form Eq. (4.4) makes their relation to Eq. (4.1) explicit, it involves the inversion of large $n \times n$ matrices, and so is not suitable for practical implementation. However, it can be written in the computationally more economical form (e.g., see Rodgers, 1976).

(i) $\hat{\mathbf{h}}_g(+) = \hat{\mathbf{h}}_g(-) + \mathbf{K}_g[\mathbf{z}_g - \mathbf{A}\hat{\mathbf{h}}_g(-)]$

with error covariance

(ii) $\mathbf{S}_{h_g}(+) = (\mathbf{I} - \mathbf{K}_g\mathbf{A})\mathbf{S}_{h_g}(-)$

where

(iii) $\mathbf{K}_g = \mathbf{S}_{h_g}(-)\mathbf{A}^T[\mathbf{A}\mathbf{S}_{h_g}(-)\mathbf{A}^T + \mathbf{S}_{z_g}]^{-1}$ (4.5a)

and likewise for the OT estimate

(i) $\hat{\mathbf{h}}(+) = \hat{\mathbf{h}}(-) + \mathbf{K}_h[\mathbf{z}_h - \mathbf{A}\hat{\mathbf{h}}(-)]$

with error covariance

(ii) $\mathbf{S}_h(+) = (\mathbf{I} - \mathbf{K}_h\mathbf{A})\mathbf{S}_h(-)$

where

(iii) $\mathbf{K}_h = \mathbf{S}_h(-)\mathbf{A}^T[\mathbf{A}\mathbf{S}_h(-)\mathbf{A}^T + \mathbf{S}_{z_h}]^{-1}$. (4.5b)

Now, unlike Eq. (4.4), Eq. (4.5) only involves the inversion of much smaller $p \times p$ matrices.

So, for example, Eq. (4.5a) relates analysed deviations of the geoid from the "first guess" to observed deviations through a matrix of weighting coefficients $\mathbf{K}_g$ (sometimes called a Kalman-Bucy filter). The procedure is straightforward to implement. The weights to be given to each observation are computed from (iii) of Eq. (4.5a) (the covariances contain the information about the geometric configuration of the observations) and the best estimate and its error follow from (i) and (ii) of Eq. (4.5a).

Equation (4.5) is of the form in which mininum variance estimation techniques are applied in meteorological applications to form the analysis, where it is called "optimum interpolation" or "objective analysis" (e.g., see the review by Bengtsson, 1975, or, in an oceanographic context, Bretherton *et al.*, 1976); it is the "least-squares collocation" technique of physical geodesy (e.g., see Moritz, 1978); it is the "geophysical inverse theory" introduced by Wunsch (1978) to determine the ocean circulation from hydrographic data and used by Wunsch and Gaposchkin (1980) to combine geodetic, hydrographic and altimetric data.

The approach Eq. (4.5) is very powerful but it relies on knowledge of the errors in the various fields which are often poorly known. In practice the procedure is always suboptimum because there is no detailed information on the error structures. Nevertheless, it is always possible to put some error bars on the measurements, and Eq. (4.5) provides the framework through which we proceed.

b. Modeling of error covariances

The error covariances must be chosen with care because, as discussed in Section 4d, they determine the response of the analysis as a function of scale. The covariances will always be imperfectly known but in many applications this may not be severely limiting, because the performance of the analysis is not sensitive to their fine detail. There are a few

simple guidelines which are worth bearing in mind when choosing covariances.

First, it is important that the estimated covariance should be a possible covariance, i.e., that it be positive definite (no negative eigenvalues). If this is not so then merely by a change of basis the rms error in the transformed field can be made negative. Second, the estimated covariance should be an approximation to the true covariance, and it is reasonable to suppose that the covariance of the errors in the field reflects the covariance of the field itself. Third, we should attempt to reconcile the desire for optimality with that of simplicity in an acceptable way.

In the simulation study of Sections 5 and 6 we only attempt to model the gross or essential features of the probable error structure in a quasi-geostrophic ocean model, geoid model and the observations.

1) OCEAN MODEL

The error in the estimate carried by the ocean model, Eq. (3.1), will be smaller at observing times (satellite passes) and in data-rich areas (near the satellite tracks). Thus ideally the covariance $\mathbf{S}$ should be a function of space and time. In adopting the formalism Eq. (4.5) it is assumed that the space and time evolution of $\mathbf{S}$ can be modelled in a sufficiently realistic way.

We write

$$\mathbf{S}_h = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2},$$

where $\mathbf{C}$ is a time-independent correlation matrix which carries the information on the spatial structure of the errors in the estimate, and $\mathbf{D}$ is a time-dependent diagonal variance matrix containing the squares of the standard deviations of the estimate. This is the conventional separation of the covariance between time-dependence and space-dependence commonly adopted in optimum interpolation.

In view of the close dynamical similarity between the geostrophic eddy field in the atmosphere and oceans it seems appropriate for synoptic mapping purposes to adopt the form of the height–height correlation function most commonly used in meteorological applications—the Gaussian exponential function:

$$C_{ij} = e^{-1/2(r_{ij}/b)^2}, \tag{4.6}$$

where r_{ij} is the distance between points i, j and b an empirically determined coefficient fixing the scale of the height field errors. In (4.6) C_{ij} is supposedly independent of time. Equation (4.6) says that the correlation of the errors in $\hat{\mathbf{h}}$ between points a distance r apart decreases exponentially with separation, and is negligible at separations much larger than b. The correlation scale b of the errors depends on the scale of the features in $\mathbf{h}$, and so for the geostrophic eddy field b is set by the baroclinic Rossby radius of deformation, L_ρ = O (100 km).

It is further supposed that the variance is constant in space and time.

$$\mathbf{D} = \sigma_m^2\mathbf{I}. \tag{4.7}$$

Here no attempt is made to model the time evolution of the variances (in optimum interpolation the variance is often assumed to vary linearly with time). Equations (4.6) and (4.7) are our representation of the error structure in $\hat{\mathbf{h}}(-)$: they assume isotropy and homogeneity respectively.

2) GEOID MODEL

To apply Eq. (4.5) a quantification of the errors in existing geoid models is needed. To compute the geoid at one geographical position requires, in principle, knowledge of the gravity field over the entire surface of the earth. The error in the geoid will thus depend on the quality, quantity and distribution of the gravity measurements. The problem of constructing such a gravimetric geoid and quantifying its errors is discussed in, for example, Heiskanen and Moritz (1967) and Zlotnicki *et al.* (1982).

For the present simulation study, however, we adopt the simplest possible form for the error structure and write

$$\mathbf{S}_{h_g} = \sigma_g^2\mathbf{I}, \tag{4.8}$$

which assumes that the errors in the geoid model is a white-noise process: uncorrelated with a constant variance σ_g, and similarly for the

3) ALTIMETER OBSERVATIONS

$$\mathbf{S}_{z_o} = \sigma_o^2\mathbf{I} \tag{4.9}$$

No attempt will be made to justify the assumption of uncorrelated constant variance geoid and observational errors, except to say that Eqs. (4.8) and (4.9) are best regarded as "minimum of information" statements. In Section 4d they will allow us to rewrite Eq. (4.5) in a physically appealing form.

c. Eigenvector decomposition of the Gaussian correlation

The form of the error structure in $\hat{\mathbf{h}}(-)$ modeled by Eq. (4.6) can be most readily understood by finding the eigenvectors and eigenvalues of $\mathbf{C}$, i.e., by diagonalizing $\mathbf{C}$ (see Rodgers, 1976). Hollingsworth (1984), in particular, has investigated the error structure implied by the Gaussian exponential correlation in this way. Our discussion follows closely that of Hollingsworth.

As an example, Fig. 3 shows the eigenvalues and eigenvectors of the correlation matrix Eq. (4.6) appropriate to 11 equally spaced observations along a satellite track a distance Δ apart, with a correlation scale $b = 3\Delta$. For clarity only the four gravest modes ($\mathbf{e}_1$–$\mathbf{e}_4$) are shown. The gravest mode is the "mean,"

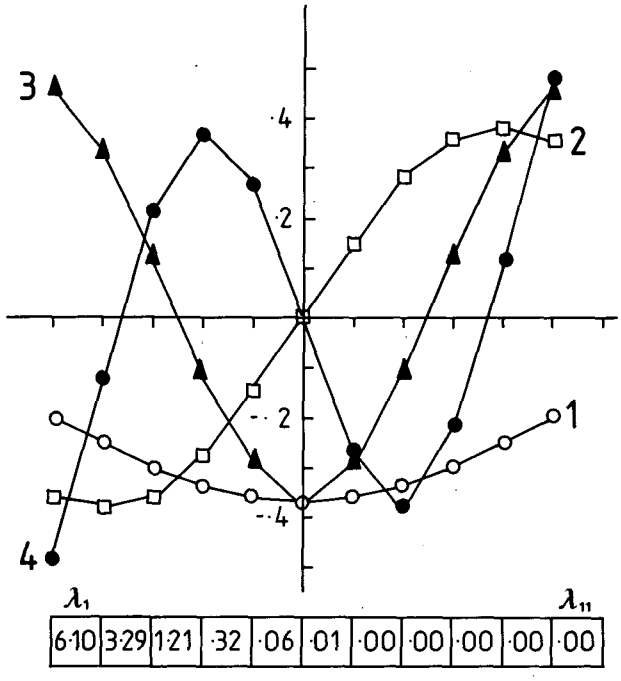

FIG. 3. The normalized eigenvectors and eigenvalues of the Gaussian exponential, Eq. (4.6), appropriate to 11 regularly spaced altimeter observations a distance Δ apart along a satellite track, with $b/\Delta = 3$. Only the four gravest eigenvectors, $\mathbf{e}_1$–$\mathbf{e}_4$, are shown and $\sum_{i=1}^{11} \lambda_i = 11$.

the next the "linear trend," etc. down to grid-scale waves at $\mathbf{e}_{11}$ (not shown). These eigenvectors form statistically independent error patterns (they are sometimes called the empirical orthogonal functions of the covariance $\mathbf{C}$): the estimate $\hat{\mathbf{h}}(-)$ is uncertain to the extent of adding on each eigenvector $\mathbf{e}_i$ with a random coefficient of variance λ_i. As can be seen from Fig. 3, the gravest modes are associated with the largest eigenvalues. Thus, in this case, the total variance = Trace of $\mathbf{C}$ = 11, the number of observations, (because each observation has unit variance and $\mathbf{C}$ is symmetric and positive definite) is explained by the first few modes.

If there are p observations, there are p eigenvectors, p eigenvalues and $\Sigma\, \lambda_i = p$. As $\Delta/b \to \infty$ (widely spaced observations with respect to b) then $\lambda_i \to 1$ and $\mathbf{e}_i^{\mathrm{T}} \to (0 \cdots, 1_i, \cdots 0)$: the observations are so widely spaced that they are independent of one another and the error is equally distributed amongst the eigenmodes.

In the case of a radar altimeter, however, giving a measurement every 10 km along its track, with a height–height correlation scale $b \sim 100$ km, $\Delta/b = 0.1$ suggesting that the limit $\Delta/b \to 0$ (closely spaced along-track observations with respect to b) is of more interest. As $\Delta/b \to 0$, $\lambda_1 \to p$, $\lambda_{i\neq 1} \to 0$ and $\mathbf{e}_1^{\mathrm{T}} \to (1, 1 \cdots 1, 1)p^{-1/2}$: scales smaller than b can be resolved, but almost all of the error is explained by the gravest mode.

Thus, the information carried by $\mathbf{C}$ on the error structure of $\hat{\mathbf{h}}(-)$ is made transparent by a study of its eigenvalues and eigenvectors. With linear, regularly spaced observations and a Gaussian correlation, there is a clear relationship between the size of the eigenvalue and the structure of its eigenvector. The eigenvector with the largest eigenvalue has the least structure and, as the size of the eigenvalue decreases, the structure of the corresponding eigenvector becomes more complex. With densely spaced observations relative to b, almost all of the variance is explained by the gravest modes. This ordering and weighting of the variance with scale implied by the Gaussian exponential is appropriate to our problem because we expect less uncertainty in $\hat{\mathbf{h}}(-)$ at scales much smaller than the correlation scale of the OT, because there is little structure in the OT at these scales. This seems physically plausible since the OT is essentially "rigid" over a correlation scale.

d. Eigenvector expansion

The choices of error covariances Eqs. (4.6) to (4.9) determine the response of our estimation procedure Eq. (4.4) as a function of scale. This response will be at its most understandable if the fields are transformed so that their errors become uncorrelated. So we specialize Eq. (4.4) to form an estimate along a track (i.e. put $\mathbf{A} = \mathbf{I}$) and project $\hat{\mathbf{h}}(-)$ onto the $\mathbf{e}_i$'s.

If ϵ_h has correlation $\mathbf{C}$ then $\mathbf{E}^{\mathrm{T}}\epsilon_h$ has correlation $\mathbf{E}^{\mathrm{T}}\mathbf{C}\mathbf{E} = \Lambda$ where

$$\Lambda = \begin{vmatrix} \lambda_1 & & 0 \\ & \ddots & \\ & \lambda_i & \\ & & \ddots \\ 0 & & \lambda_p \end{vmatrix}$$

is a diagonal matrix of eigenvalues of $\mathbf{C}$ (ordered in descending λ_i) if

$$\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_i, \cdots \mathbf{e}_p)$$

is the matrix of orthogonal normalized eigenvectors of $\mathbf{C}$. Thus, $\mathbf{E}$ forms the natural basis in which to expand $\hat{\mathbf{h}}^{\mathrm{track}}(\pm)$:

$$\left.\begin{aligned} \hat{\mathbf{h}}^{\mathrm{track}}(\pm) &= \mathbf{E}\mathbf{h}'(\pm) \\ \mathbf{S}'_h &= \sigma_m{}^2\Lambda \end{aligned}\right\}, \qquad (4.10a)$$

where the primed variables are the coefficients of the expansion of the fields in the basis $\mathbf{E}$ [remember the eigenvalues of $\mathbf{S}_h$ measure the error variance of each coefficient in the expansion of $\hat{\mathbf{h}}(-)$]. The errors in the primed basis are uncorrelated because Λ is diagonal.

Similarly, $\mathbf{z}_o$ and $\mathbf{h}_g$ can also be projected onto $\mathbf{E}$ and, because of the assumption of uncorrelated errors

in Eqs. (4.8) and (4.9), the errors in the transformed fields will remain uncorrelated under the orthogonal transformation ($\mathbf{E}^T\mathbf{E} = \mathbf{I}$):

$$\left.\begin{aligned}\mathbf{z}_o &= \mathbf{E}\mathbf{z}'_o\\ \mathbf{S}_{z_o} &= \sigma_o{}^2\mathbf{I}\end{aligned}\right\} \tag{4.10b}$$

and

$$\left.\begin{aligned}\hat{\mathbf{h}}_g{}^{\text{track}}(\pm) &= \mathbf{E}\mathbf{h}'_g(\pm)\\ \mathbf{S}'_{h_g} &= \sigma_g{}^2\mathbf{I}\end{aligned}\right\}. \tag{4.10c}$$

Substituting Eq. (4.10) into Eq. (4.5), the best estimates can be written in the primed basis in the very illuminating form:

$$h'_{gi}(+) = \frac{\gamma_o{}^2 + \lambda_i}{\gamma^2 + \lambda_i} h'_{gi}(-) + \frac{\gamma_g{}^2}{\gamma^2 + \lambda_i} z'_{gi} \tag{4.11a}$$

$$h'_i(+) = \frac{\gamma^2}{\gamma^2 + \lambda_i} h'_i(-) + \frac{\lambda_i}{\gamma^2 + \lambda_i} z'_{hi} \tag{4.11b}$$

where

$$\gamma_o{}^2 = \left(\frac{\sigma_o}{\sigma_m}\right)^2; \quad \gamma_g{}^2 = \left(\frac{\sigma_g}{\sigma_m}\right)^2$$

measure the noise in, respectively, the observations and the geoid model compared to the ocean model, and

$$\gamma^2 = \gamma_o{}^2 + \gamma_g{}^2.$$

Now, after projection onto the $\mathbf{e}_i$ the errors in each component are uncorrelated: each component i is independent and can be estimated independently, combining them as if they were scalars. Equation (4.11) shows clearly how $\gamma_o{}^2$ and $\gamma_g{}^2$ control the mix of first-guess $h(-)$ and observation z in the analysis of the geoid and OT. If the altimeter is very poor $\gamma_o{}^2$, $\gamma^2 \to \infty$ then we cannot distinguish the signal from zero and the best estimate reverts back to our prior estimate. The limit case in which geoid errors dominate over instrumental errors, however, is of more interest, $\gamma_o{}^2 \ll \gamma_g{}^2$ for then:

at large scales, $\lambda_i \gg \gamma_g{}^2$

$$\left.\begin{aligned}h'_{gi}(+) &\approx h'_{gi}(-)\\ h'_i(+) &\approx z'_{h_i} = h'_{oi} - h'_{gi}(-)\end{aligned}\right\} \tag{4.12a}$$

and the altimeter and the geoid are used to update the ocean model;

at small scales, $\lambda_i \ll \gamma_g{}^2$

$$\left.\begin{aligned}h'_i(+) &\approx h'_i(-)\\ h'_{gi}(+) &\approx z_{gi} = h'_{oi} - h'_i(-)\end{aligned}\right\} \tag{4.12b}$$

and the ocean model is used in conjunction with the altimetry to improve the geoid.

In Eq. (4.12) "small" and "large" here are measured relative to the scale exhibited by the eigenvector corresponding to $\lambda_i \approx \gamma_g{}^2$ and is thus a function of the height–height correlation scale of the ocean signal (which fixes $\mathbf{E}$ and Λ) and the relative accuracy of the geoid model and the ocean model (which fixes $\gamma_g{}^2$). So, for example, suppose there is equal uncertainty in the geoid model and ocean model $\gamma_g{}^2 = 1$ then in Fig. 3 the three gravest eigenvectors, $i = 1$, 2 and 3, of $\mathbf{z}_h$ will be associated with OT, whereas modes $i = 4$ to 11 in $\mathbf{z}_g$ will be associated with the fine structure in the geoid.

The eigenvector expansion is only valid under the assumption of uncorrelated errors in $\mathbf{z}_o$ and $\hat{\mathbf{h}}_g(-)$, but is nevertheless useful because it makes the response of the analysis transparent to errors at various scales and demonstrates how they are controlled by the error covariances.

The operation of the analysis is now very simple to understand. To analyze for the geoid and OT along a track using Eq. (4.11) we

i) form the difference fields $\mathbf{z}_h$ and $\mathbf{z}_g$,

ii) expand in the eigenvectors of the covariance matrix $\mathbf{S}_h(-)$, and then

iii) all structure in $\mathbf{z}_g$ corresponding to $\lambda_i < \gamma_g{}^2$ (small scales) is associated with the geoid

iv) all structure in $\mathbf{z}_h$ corresponding to $\lambda_i > \gamma_g{}^2$ (large scales) is associated with the OT.

It is instructive to note that in the case of perfect altimetry, then from Eq. (4.4) with $\mathbf{S}_{z_o} = 0$, our best estimates of $\mathbf{h}$ and $\mathbf{h}_g$ must add to give the true sea surface height, i.e.,

$$\hat{\mathbf{h}}_g(+) + \hat{\mathbf{h}}(+) = \mathbf{h}_o.$$

It then follows that

$$\hat{\mathbf{h}}(+) - \mathbf{z}_h = \hat{\mathbf{h}}_g(-) - \hat{\mathbf{h}}_g(+). \tag{4.13}$$

Thus our scheme associates the difference, or bias, between the best estimate of h and the observations of h, with a systematic error in the geoid. The relation (4.13) only requires that $\mathbf{S}_{z_o}$ be zero and so, unlike Eq. (4.11), is independent of particular statements about $\mathbf{S}_h(-)$ and $\mathbf{S}_{h_g}(-)$.

In this section we have quantified our notions regarding the separation of geoid errors from the ocean signal, and set out the theoretical framework in which it is proposed to tackle the combined problem of OT determination and geoid improvement from satellite altimetry. In the following sections a practical illustration of the method is presented using simulated OTs and geoid errors.

5. Simulation study

Here we briefly describe an eddy resolving circulation model of idealized time-dependent double gyres which is used to provide a truth circulation $\mathbf{h}$. This is degraded by a systematic (unchanging) error ϵ_g

representing uncertainty in a gravimetric geoid. We then proceed to separate out ϵ_g from **h** using the minimum variance estimation techniques developed in Section 4.

Our focus on Gulf Stream variability should be of particular interest because the eddy field is associated with pronounced sea surface tilts which have been shown to be detectable in Seasat records. Random errors in the altimeter instrument and systematic long wavelength uncertainties in tides and satellite orbits are probably not limiting here. Further, detailed gravimetric geoids exist for the Western North Atlantic whose errors do not appear to swamp the oceanographic signal, suggesting that absolute velocity determination may be a possibility (e.g., see Cheney and Marsh, 1981). Since the geostrophic eddy field is almost certainly a dynamical instability of the larger scale flow, rather well described by the conservation of potential vorticity on synoptic time-scales, powerful dynamical statements can be made.

a. *Ocean topography*

The ocean model is a homogeneous layer of fluid of depth H and density ρ confined to a basin of width L on a β-plane and governed by the barotropic vorticity equation. It is driven by an imposed wind-stress curl forcing F, and frictionally retarded D:

$$\frac{\partial}{\partial t}\nabla^2\psi + J(\psi, \nabla^2\psi + \beta y) = F - D, \qquad (5.1)$$

where J is the Jacobian, ψ the streamfunction for the horizontal flow, β the planetary vorticity gradient, t the time, x is east, and y north.

The deviation of the sea surface from the gravimetric geoid is $h = (f_o/g)\psi$ where f_o is the Coriolis parameter and g the acceleration due to gravity. In our simulation study Eq. (5.1) will be taken as a full and complete description of the time evolution of the ocean topography.

A finite-difference version of Eq. (5.1) is leap-frogged forward on a 33 × 65 grid from a known initial state in a double gyre configuration (see Fig. 2a).

$$\left.\begin{aligned} 0 < x < L \\ -L < y < L \end{aligned}\right\},$$

with

$$F = \frac{\tau}{\rho H}\sin\left(\frac{\pi y}{L}\right)$$

(for an ocean in which τ, the wind-stress = 10^{-1} N m^{-2}, $\rho = 10^3$ kg^{-3}, $\beta = 2 \times 10^{-11}$ s^{-1}, $H = 5 \times 10^2$ m and $L = 10^6$ m) to generate a reference six-month sequence of "synoptic" maps of the Gulf Stream and its recirculation. Figure 4 shows a series of height fields, **h** at eight-day intervals. The internal jet separating counter-rotating gyres is barotropically unstable and meanders along the zero wind-stress curl line occasionally forming cutoff rings. The rings have a height signal of ½ meter and a horizontal scale of 200 km. Smaller amplitude, larger-scale westward propagating Rossby basin modes are excited to the north and south which disturb the Sverdrup interiors. This is our "truth" circulation which will be sampled to simulate an altimetric data set.

Our choice of model deserves comment. The barotropic formulation Eq. (5.1) cannot, nor does it attempt to reproduce the plethora of space and time-scales that make up the oceanic variability. The absence of stratification means that $(u/\beta)^{1/2}$ rather than $L\rho$, sets the scale of the geostrophic eddy field which, as a result, tends to be too intense, and rather too broad in scale. For our present purposes, however, the formulation has several advantages over more

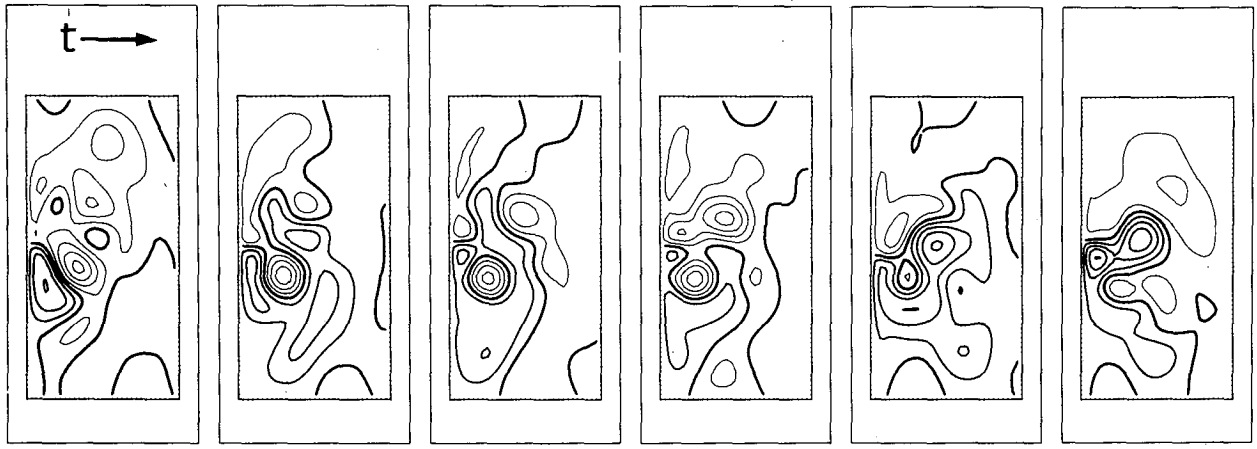

FIG. 4. A series of instantaneous OT. $h(x, y, t)$ (C.I. = 15 cm) at 8-day intervals showing the barotropically unstable internal jet separating counter-rotating gyres.

complex and perhaps realistic models. It has space and time scales in common with the oceanic geostrophic eddy field, and indeed can reproduce much of the complexity seen in the active upper layer of multilayer quasi-geostrophic models (for a fuller description of the present model and its use in an eddy-resolving mode, see Marshall 1984). Even more importantly, the model has only as many degrees of freedom as are consistent with altimetric data and so is fully constrained by it. Finally the barotropic model does not suffer from initialization problems and so can be rather straightforwardly used as an analysis tool to assimilate noisy data.

b. Geoid error

The amplitude of the error in a gravimetric geoid as a function of scale is modeled by assuming that the error spectrum is white from the smallest wavelengths resolvable by the numerical grid up to the scale of the miniature basin, i.e., from 30 km to 1000 km. So ϵ_g is randomly chosen from a normal distribution of zero mean and standard deviation σ_g

$$\left.\begin{aligned}\mathbf{Exp}(\epsilon_g) &= 0\\ \mathbf{S}_{hg} = \mathbf{Exp}(\epsilon_g \epsilon_g^{\mathrm{T}}) &= \sigma_g{}^2\mathbf{I}\end{aligned}\right\}. \tag{5.2}$$

Figure 2b shows ϵ_g for the case σ_g = 30 cm. This is a significant error giving slopes in excess of ½ meter in 100 kms, sufficient to completely swamp the oceanographic signal. It should be compared, for example, with the six-month mean "truth" circulation shown in Fig. 2a, in which the OT drops down by about ½ meter across the Gulf Stream front separating subpolar and subtropical gyres. A model of correlated geoid errors will be considered in (ii) of Section 6c.

c. Sampling strategy

An altimeter can never resolve the richness of scale shown in Fig. 2, and so the reference circulation and reference geoid error is sampled to crudely simulate the spatial and temporal coverage of a radar altimeter in a seven-day repeat orbit at an altitude of 1000 km, giving a track separation of 250 km (in midlatitudes). A seven-day repeat is chosen because it is found to be an optimum specification given the space and time scales of the simulated variability. For simplicity the satellite is constrained to pass over columns and rows of the grid marked by the thick black lines in Fig. 2a, sampling $\mathbf{h}$ and ϵ_g at each grid point along its track. We form "observations" of the OT and geoid thus:

$$\left.\begin{aligned}\mathbf{z}_h &= \mathbf{A}[\mathbf{h} + \epsilon_g - \hat{\epsilon}_g(-)]\\ \mathbf{z}_g &= \mathbf{A}[\mathbf{h} + \epsilon_g - \hat{\mathbf{h}}(-)]\end{aligned}\right\}, \tag{5.3}$$

where

$h = h(x, y, t)$ the "truth" OT taken from the ocean model
$\epsilon_g = \epsilon_g(x, y)$ the "truth" geoid error
$\hat{\epsilon}_g(-)$ a prior estimate of the geoid error and
$\hat{h}(-)$ a prior estimate of the OT.

Before proceeding it should be emphasized that Eq. (5.3) does not attempt to model the instrumental and tracking errors of an altimetric system. This choice is made recognising that it is the uncertainty in the geoid, rather than the precision of the altimetric system, which will ultimately limit the impact of altimetry in ocean circulation studies. However, the neglect of systematic orbital uncertainties in our error analysis is certainly a gross over-simplification. The orbit error can be regionally and temporally correlated due to gravity model errors and the idiosyncrasies of the tracking data collection campaigns—see Anderle and Hoskins, 1977. The character of this error is difficult to quantify but it will probably be on the order of a few thousand kilometers and greater, and so should not compromise the use of altimeter data for mapping on geostrophic eddy scales.

In the next section best estimates, $\hat{\mathbf{h}}(+)$ and $\hat{\mathbf{h}}_g(+)$ are computed from $\mathbf{z}_h$ and $\mathbf{z}_g$ using versions of Eq. (4.5).

6. Minimum variance solutions

We now present minimum variance solutions for $\hat{\mathbf{h}}_g(+)$ and $\hat{\mathbf{h}}(+)$, using various specifications of $\hat{\mathbf{h}}(-)$. Our purpose here is to demonstrate how dynamical information in a first guess $\hat{\mathbf{h}}(-)$ can be used to constrain our estimates helping to discriminate between geoid errors and the OT. In each calculation the same *a priori* covariance information, i.e., $\mathbf{S}_h$, $\mathbf{S}_{hg}$ and $\mathbf{S}_{z_o}$, is used but the dynamical content of $\hat{\mathbf{h}}(-)$ is changed to study its impact on the analysis. The success of our procedure can be quantified by computing the rms error between the estimates and the truth fields.

a. Computational details

The height–height correlation scale appropriate to the OT signal is found by fitting the Gaussian Eq. (4.6) to the correlation function computed from the model fields by adjusting b: b = 100 km is found to be appropriate. This is consistent dynamically with $(u/\beta)^{1/2}$ which determines the scale of the features in Fig. 4. With simulated data every 32 km along the track b/Δ = 3, the ratio used to calculate the eigenvector decomposition of the correlation matrix $\mathbf{C}$ shown in Fig. 3. The analysis is sensitive to the assumed value of b, but the eigenvector expansion Eq. (4.11) makes this dependence understandable by

showing explicitly how the correlation scale determines the spectral response of the analysis. In the calculations that follow, therefore, our chosen value of 100 km, close to an optimum specification, is kept fixed.

Rather than update the whole vector field simultaneously it is more convenient to estimate it separately at each interpolation point: Eq. (4.5) is specialized to analyse for a point rather than a vector. Precise details are not of interest except to say that the size and number of matrix inversions can be reduced by exploiting the symmetry of the satellite tracks and sensibly using correlation scale information. The correlation is the most natural way of setting the scale over which observations can be allowed to influence the estimate, and thus it can be used to limit the size of the matrices to be inverted. In computing the best estimate at a point, more distant observations have less influence than adjacent observations. It is found convenient to allow only those observations within a circle of radius $3b/2$ to influence the analysis at the origin of the circle. This means that along track observations can be used to update the OT in a swath $3b$, or 300 kms wide centred on each track. Typically 11 observations are combined to estimate for points along the tracks, and somewhat fewer observations for off-track points. Since the geometrical configuration of the observations remains fixed and no attempt is made to model the time evolution of the covariances, the weights need only be computed once. Finally, it should be noted that because the geoid errors are assumed to be uncorrelated (i.e. the correlation scale for $\mathbf{h}_g$ is zero, Eq. (5.2)) then it is only possible to estimate for the geoid *along* the track: the *a priori* assumption Eq. (5.2), irrespective of the true form of the geoid error, precludes an estimate of the geoid between the tracks.

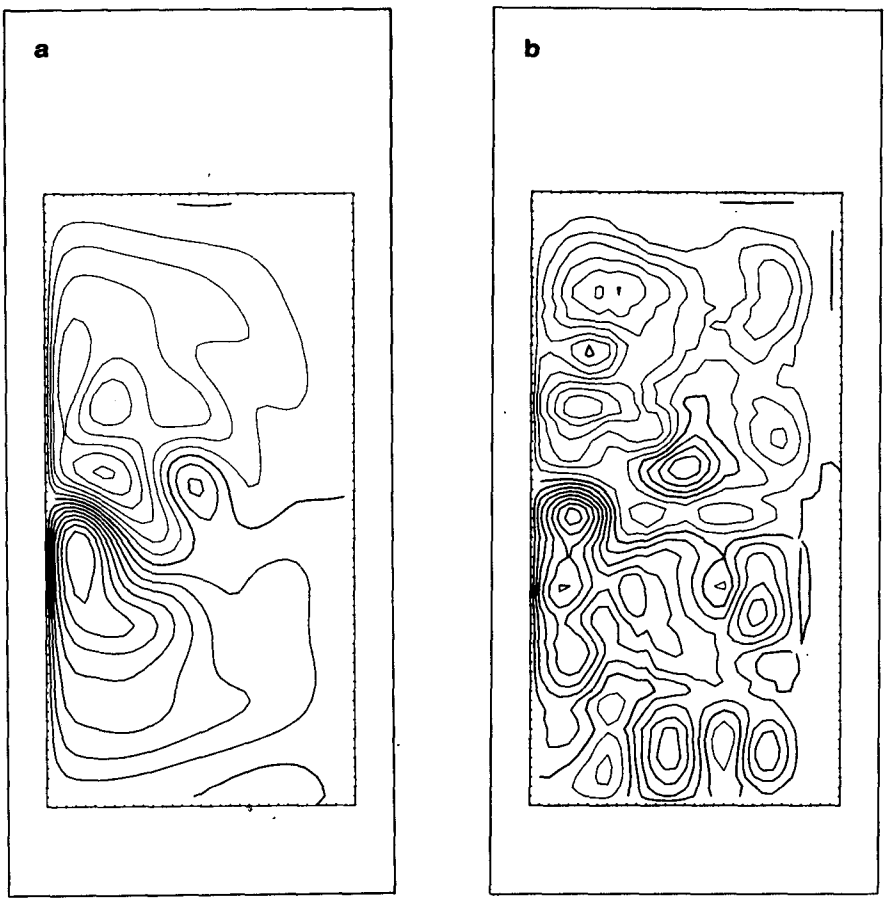

FIG. 5. (a) The truth six-month time-mean OT. (b) The $\hat{\mathbf{h}}(+)$ of the mean OT obtained by setting the first guess fields $\hat{\mathbf{h}}(-) = \hat{\epsilon}_g(-) = 0$ (the rms error with respect to the truth field is 12 cm). C.I. = 6 cms.

b. Zero "first guess" information

The first guess fields, $\hat{\mathbf{h}}(-)$ and $\hat{\epsilon}_g(-)$ are set to zero [thus decoupling (4.5a) and (4.5b)], and the observations Eq. (5.3)

$$\mathbf{z}_h = \mathbf{z}_g = \mathbf{A}(\mathbf{h} + \epsilon_g)$$

averaged over the six-month period of the altimetric record. The minimum variance solutions are computed from Eq. (4.5) supposing that there is equal uncertainty in the (zero) geoid and OT "first guess" fields. i.e., $\gamma_g^2 = 1$.

Figure 5b shows the minimum variance solution $\hat{\mathbf{h}}(+)$. A two-gyre structure is evident, although severely distorted by geoid errors. The rms error with respect to the truth circulation (Fig. 5a) is 12 cm. Evidently geoid errors along the tracks are smeared out over a correlation scale in the analysis resulting in significant height field errors. In our analysis with $b = 100$ kms, $b/\Delta = 3$, and $\gamma_g^2 = 1$ the scale separation between geoid errors and OT occurs at about 250 kms, the wavelength exhibited by the $\mathbf{e}_i$ with $\lambda_i \sim 1$ (ϵ_3 and ϵ_4 in Fig. 3). This scale is evident in Fig. 6 which plots the true geoid error ϵ_g and the residual error $\epsilon_g - \hat{\epsilon}_g$ along a selected track. Our estimates associate this residual with OT, which consequently suffers from errors as large as 20 cm on scales of 200–300 km.

Although γ_g^2 determines the cut-off above which structure in $\mathbf{z}_h$ is attributed to OT, the analysis is not overly sensitive to γ_g^2. The eigenvalues of $\mathbf{C}$ fall so rapidly towards zero (see Fig. 3) that γ_g^2 would have to be very small (great certainty in $\hat{\mathbf{h}}(-)$ relative to $\hat{\mathbf{h}}_g(-)$) if the OT analysis were to draw closely to $\mathbf{z}_h$ at small scales.

c. Assimilation into ocean models

A priori oceanographic information of a statistical nature was brought to bear in the analysis presented in Section 6b through the correlation scale b. But this is an understatement of our prior knowledge because the simulated data was generated by sampling a series of synoptic maps whose dynamics is governed by Eq.

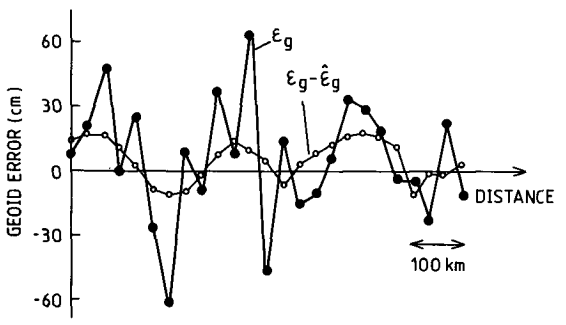

FIG. 6. The truth (uncorrelated) geoid error ϵ_g, and the difference field $\epsilon_g - \hat{\epsilon}_g$ (plotted along a selected track) obtained by setting $\hat{\mathbf{h}}(-) = \hat{\epsilon}_g(-) = 0$, as in Fig. 5(b).

(5.1). It would be sensible, therefore, to use Eq. (5.1) to assimilate the observations and thus aid in the analysis of the data.

The following continuous updating strategy is employed:

1) starting from initial estimates of the ocean circulation and geoid error $\hat{\mathbf{h}}(-)$ and $\hat{\mathbf{h}}_g(-)$, the ocean model is integrated forward in time providing an estimate of the state of the ocean $\hat{\mathbf{h}}(-)$ which is imperfect because of imperfections in the initial conditions and model physics.

2) when altimeter observations of the sea-surface height become available they are combined with $\hat{\mathbf{h}}(-)$ (current state of the ocean model) and $\hat{\mathbf{h}}_g(-)$ (current estimate of the geoid) to form the observations $\mathbf{z}_h$ and $\mathbf{z}_g$.

3) from these observations the best estimates of the geoid and the OT, $\hat{\mathbf{h}}_g(+)$ and $\hat{\mathbf{h}}(+)$ are computed from Eq. (4.5), using as *a priori* constraint information $\hat{\mathbf{h}}_g(-)$ and $\hat{\mathbf{h}}(-)$.

4) the ocean model is integrated on from $\hat{\mathbf{h}}(+)$ to the next observation time.

It should be noted that rather than apply Eq. (4.5) once to averaged observations as in Section 6b, it is applied successively as new observations become available. So, between observation times $\hat{\mathbf{h}}_g$ remains unchanged whilst $\hat{\mathbf{h}}$ evolves according to Eq. (5.1). At observation times a new estimate of the geoid error and OT is made. Hopefully the iterative procedure will converge so that as time progresses, better and better estimates can be made because the *a priori* statements Eq. (2.3) and (2.4) contain more and more useful information.

1) THE COMBINED PROBLEM

Figure 7 shows the rms error in the geoid and OT estimates plotted as a function of the number of satellite passes obtained as described above by assimilating the observations into the ocean model. The procedure is started by setting the initial geoid and OT estimates to zero. The rms error in these initial estimates is 30 cm for the geoid and 27 cm for OT. As before γ_g^2 is set to unity.

The errors have fallen dramatically after the first 7-day pass. This is almost entirely due to covariance information which immediately associates [Eq. (4.12)] the smallest spatial scales in the observations with geoid errors and large spatial scales with OT. From then on the ocean model provides a "first guess" which contains more and more information about the OT based on the dynamical content of the ocean model, its memory for past observations and the improving geoid. This slowly but surely pulls the errors down until after twenty 7-day passes they have been reduced to only 6 cm. Extended integrations beyond six months do not lead to further error

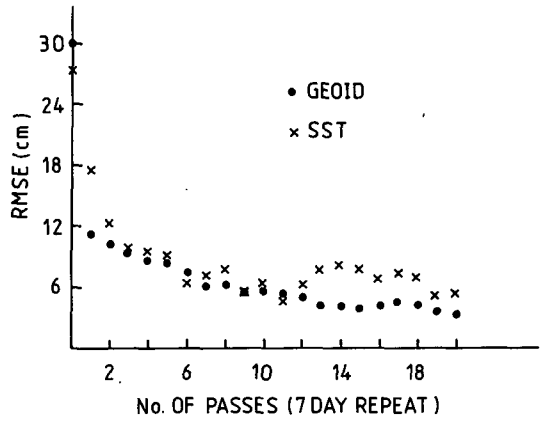

FIG. 7. The rms error in the OT and geoid estimates (for the case of uncorrelated geoid errors) plotted as a function of time, obtained by continuous assimilation of data into a perfect ocean model, and solving simultaneously for the OT and geoid error.

reductions. This lower limit on the achievable errors is a function of the amplitude and scale content of the geoid errors, and the spatial density and temporal frequency of the observations.

Figure 8a shows the six month mean OT estimate and should be compared to the truth circulation, Fig. 5a. It has a rms error of only 4.1 cm.

A sequence of geoid error estimates along the track selected in Fig. 6 is shown in Fig. 9. Small spatial scales in the geoid error are again quickly corrected for, but after twenty passes the errors have been greatly reduced and lie only at large scales. This residual error, at scales in common with the time-mean ocean circulation, cannot be separated from it. It should be noted that the driving of the geoid error to large scales implies that although initially spatially uncorrelated, the geoid errors soon become correlated as the integration proceeds, i.e., the correlation matrix

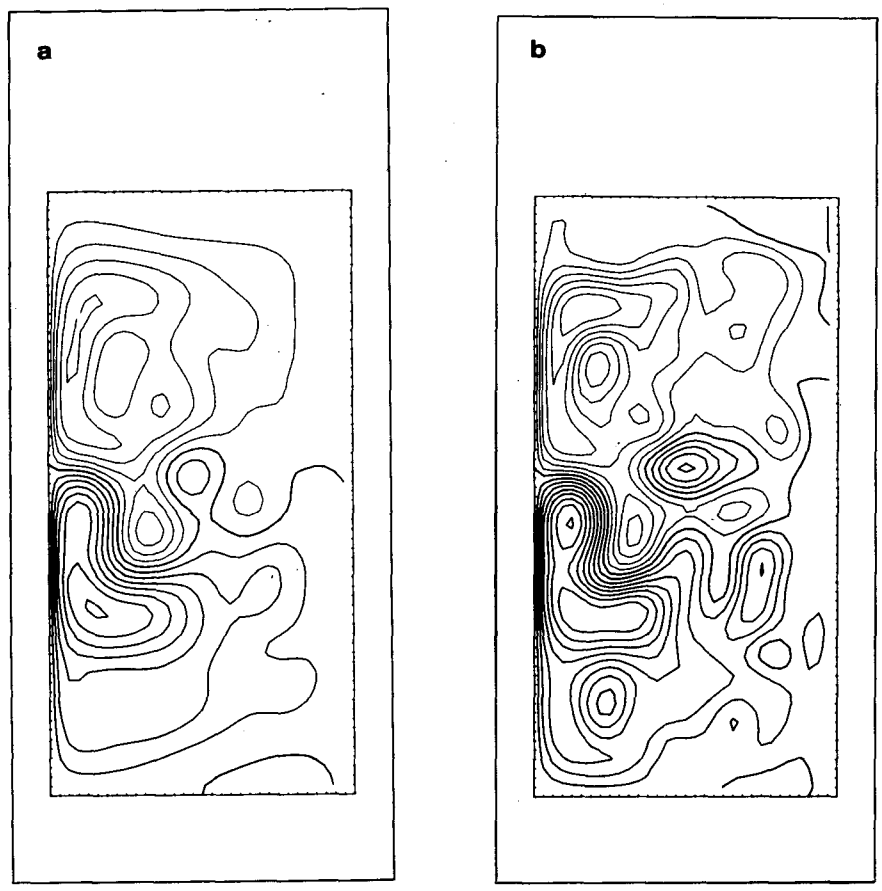

FIG. 8. "Best estimates" $\hat{h}(+)$ of the mean OT obtained as in Fig. 7 through assimilation of data into a perfect ocean model and (a) continuously updating the OT and geoid estimates (rms error = 4.1 cm) and (b) updating the OT estimate but setting $\hat{\epsilon}_g(+) = \hat{\epsilon}_g(-) = 0$ (rms error = 8.4 cm). C.I. = 6 cm.

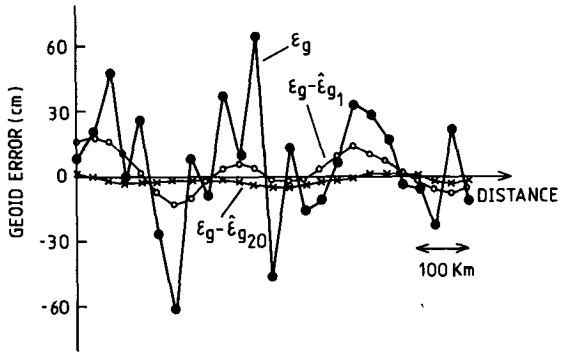

FIG. 9. The truth (uncorrelated) geoid error ϵ_g (plotted as in Fig. 6) and the difference fields $\epsilon_g - \hat{\epsilon}_{g1}$, $\epsilon_g - \hat{\epsilon}_{g20}$ where $\hat{\epsilon}_{g1}$, $\hat{\epsilon}_{g20}$ are the geoid estimates after the first and twentieth pass of the satellite respectively.

of the geoid errors becomes less and less diagonal. Although the minimum variance estimators allow one to calculate the new correlation in terms of the old, (ii) of Eq. (4.5a), no attempt has been made to incorporate this level of sophistication into the analysis. It is probable that the magnitude of the errors can be reduced further if, in the analysis, due account is taken of the change in the correlation scale of the geoid errors as the iteration proceeds i.e. by modeling the evolution of the geoid error covariance matrix using (ii) of Eq. (4.5).

The OT estimate obtained by assimilating the observations into the ocean model, but setting $\mathbf{h}_g(+) = 0$, Fig. 8b, demonstrates the efficacy of solving simultaneously for the geoid and OT. The same *a priori* oceanographic information is exploited as before, but this time no use is made of it to correct for the geoid. A comparison with Fig. 8a shows that the quality of the analysis is much reduced (the rms error is 8.4 cms). It has many small-scale features in common with Fig. 5b which are associated with geoid errors.

2) CORRELATED GEOID ERRORS

The random uncorrelated geoid error described in Section 5b is perhaps an unrealistic representation of geoid error structure and, moreover, a rather straightforward one to correct for. By merely averaging observations separated by a distance Δ over a wavelength l, random errors can be reduced by a factor $\sqrt{N}$, where $N = l/\Delta$ is the number of independent observations. So, if l = 300 km and Δ = 30 km, averaging over 10 observations could reduce the error by a factor of $\sqrt{10}$ (it is this spatial averaging of uncorrelated errors which accounts for the initial rapid fall of the errors in Fig. 7).

If more realistically, the geoid errors are correlated over oceanographically interesting scales $k^{-1} \sim b$, where k is the wavenumber, they will be correspondingly more difficult to separate out from the OT. For example, Fig. 10b shows a geoid error field generated using

$$\epsilon_g(x, y) = \sum_{n,m=2}^{20} a_{mn} \sin\left(\frac{2\pi}{n\Delta} x + \theta_{mn}\right) \sin\left(\frac{2\pi}{m\Delta} y + \phi_{mn}\right) \quad (6.1)$$

where θ_{mn} and ϕ_{mn} are randomly chosen phases in the range from $-\pi$ to π and a_{mn} are randomly chosen amplitude coefficients scaled so that the total rms error in the field is 30 cm. Thus the geoid error is of the same magnitude as in Fig. 2b but now correlated on scales up to 600 km.

Figure 11 shows the rms error as a function of the number of satellite passes for this correlated error. Although the initial geoid error covariance is not diagonal, the statistical weights are computed assuming that it is, Eq. (4.8). The rate of decrease of the errors with time is less rapid than for correlated errors (cf. Fig. 7) because now, as can be seen from Fig. 10b, there is large power in the geoid error at longer wavelengths. Covariance information has relatively less impact in the initial stages of the assimilation and, subsequently, the first guess information provided by the ocean model plays a relatively larger role in reducing the errors. The mean OT estimate, Fig. 10a, is somewhat distorted by residual geoid errors, but it has a rms error of only 6.6 cm. After twenty passes this residual error, plotted along our selected track in Fig. 12, lies at wavelengths as large as 500 km.

3) DEGRADED OCEAN MODEL

In Section 6c 1) and 2) the model which generated the data is used to assimilate it, and so there is danger of obtaining over-optimistic results because a perfect model overstates any realistic expectation of dynamical knowledge. Figure 13, therefore, presents two OT analyses formed by assimilating data into degraded models. In Fig. 13a (rms error 10.2 cm) the wind-stress curl forcing F has been removed from the model and, in Fig. 13b (rms error 9.2 cm), the nonlinear terms have been suppressed. These are both very severe degradations of the model, and indeed the quality of the analyses are much reduced.

Consistent with the absence of external forcing the analysis Fig. 13a is markedly in error in the forced Sverdrup interiors, whereas the nonlinear boundary currents and the inertial recirculation are somewhat better represented. With a linear forced model, on the other hand, Fig. 13b, the boundary currents and jets are poorly represented, for now they are controlled by frictional rather than inertial effects. As a consequence, they are too narrow $((u/\beta)^{1/2}$ which sets the scale of the boundary currents in Fig. 2a is absent in Fig. 13b).

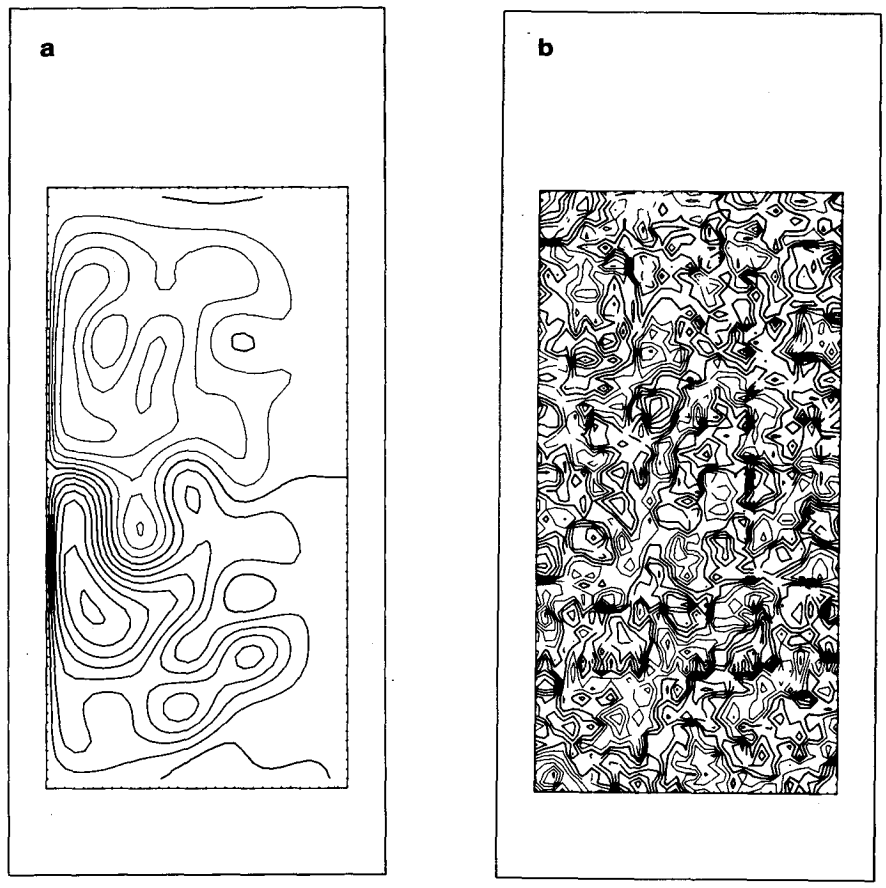

FIG. 10. (a) The $\hat{h}(+)$ of the mean OT obtained by assimilation into a perfect model but this time with the correlated geoid error shown in Fig. 10(b) (rms error = 6.6 cm, C.I. = 6 cm). (b) A simulated geoid error correlated on scales up to 600 kms with a standard deviation of 30 cms, computed using Eq. (6.1). C.I. = 15 cms.

Despite the rather crude dynamical information contained in the first guess, Fig. 13 still represents an improvement over Fig. 5b which used no first guess information.

7. Concluding remarks

Satellite altimetry offers the prospect of global coverage of surface geostrophic currents on synoptic time-scales, and as such could play a central role in future oceanographic observing campaigns. Perhaps the only way of making sense of this information and blending it in a dynamically consistent way with other data types, is by assimilation into dynamical ocean models in much the same way as meteorologists form their analyses. However, as soon as such a procedure is considered for altimetric data, one is faced with the prospect of dealing with large systematic errors introduced by the uncertainty in the geoid, which severely distorts the OT signal. Our philosophy here has been to bring to bear dynamical information about the OT, concisely stated in an ocean model to help separate the OT signal from the geoid.

This combined problem has been formulated generally, making the use of *a priori* information explicit and stressing that it is a necessary and integral part of the analysis. The tried and tested method of minimum variance estimation has been used to optimally combine this prior knowledge with the observations to form best estimates. The application of the minimum variance technique itself is relatively straightforward: the major difficulty arises from the uncertainty in the error covariances. The most pressing need is for a quantification of errors both in geoid models and (if they are going to be used to assimilate altimetric data) ocean models.

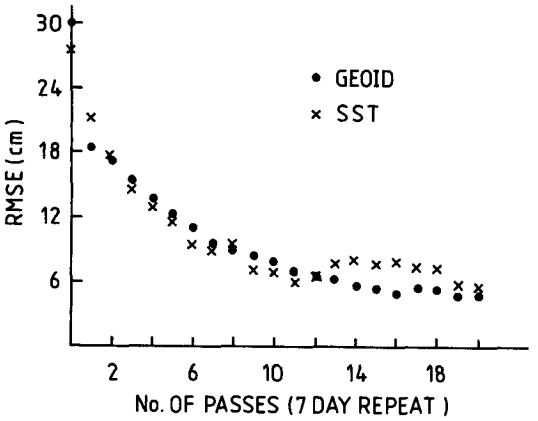

FIG. 11. The rms error in the OT and the geoid estimate (for the case of correlated geoid errors) plotted as a function of time obtained, as in Fig. 7, by continuous assimilation into a perfect ocean model.

In our simulation study these difficulties are somewhat ameliorated because there is a degree of control over the error structures and the model physics. In this preliminary study though we make no apology for using simulated instead of real data, and rather idealized error models. No attempt has been made to make definitive recommendations concerning the optimum specification of (for example) the ground-track or repeat-time strategies for synoptic mapping. Rather we have tried to set out the central difficulties as clearly as possible and indicate a sensible strategy with which to tackle them. Further simulation studies using more realistic error models certainly need to be carried out, but it is hoped that the numerical examples presented here provide a practical demonstration of how dynamical information contained in a first guess can be used to discriminate between geoid errors and OT, with a finesse which depends on the scale separation between the OT and the geoid error, and the dynamical content of the "first guess."

The success of our least-squares technique requires that the blurring of the OT by geoid errors should not be too severe, i.e., the geoid must have some information content in it. This is not surprising for it is difficult to see how any technique can be used to identify errors in a geoid model which reside at scales in common with the OT. If there is no scale separation between the OT and the geoid error, then we must await detailed geoid models for which there is. Then, expressed in its most optimistic form, we can envisage using altimetry in conjunction with a geoid model to update an ocean model with the height-field on large scales (~ few hundred km up to basin scale) whereas on small scales altimetry can be used in conjunction with the ocean model to fill in the fine structure of the geoid. This objective, though, can only be realized given the much improved geoid made possible by a dedicated gravity satellite mission such as the GRM. For the immediate future we must proceed in stages investigating regional problems, using local geoids and ocean models tailored to the prevailing local dynamics.

Given the problem at hand, the most appropriate form of dynamical representation must be investigated. For example, although a quasi-geostrophic ocean model might be appropriate for assimilating height-field information in midlatitudes, it would be a most inappropriate choice in the tropics since it cannot describe the gravity wave response which is an important component of the signal in equatorial oceans. (Some general considerations on the assimilation of data into equatorial ocean models can be found in Gill *et al.*, 1984). In certain circumstances a much more schematic representation of the dynamics may be acceptable. For example, a few Rossby normal modes may be adequate to capture much of the power in the geostrophic eddy field described by MODE (Mid-Ocean Dynamics Experiment—see Simmons *et al.*, 1977). Here one could project fields directly on to these modes and optimally estimate for their coefficients: we use a basis chosen on dynamical rather than statistical grounds.

It is sensible to commence such data assimilation studies using, as here, simple models with only as many degrees of freedom as are consistent with the data. But if surface information is to be extrapolated downwards the extent to which baroclinic models are constrained by height-field information alone needs to be investigated. One suspects that, since the pressure field is specified at the surface where velocities are the largest, the flow would be rather strongly constrained. Furthermore, on those space and time scales for which the oceanic response is first baroclinic mode, we might expect the thermocline depth to be correlated with the surface elevation. But detailed study using models with vertical resolution is required in order to quantify the confidence that can be placed in a given model.

Finally, it should be mentioned that in this paper attention has been focussed exclusively on absolute

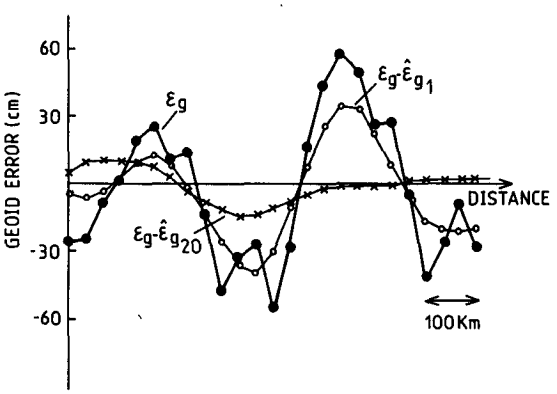

FIG. 12. The truth (correlated) geoid error ϵ_g plotted along a selected track, and the difference fields $\epsilon_g - \hat{\epsilon}_{g1}$, $\hat{\epsilon}_g - \hat{\epsilon}_{g20}$.

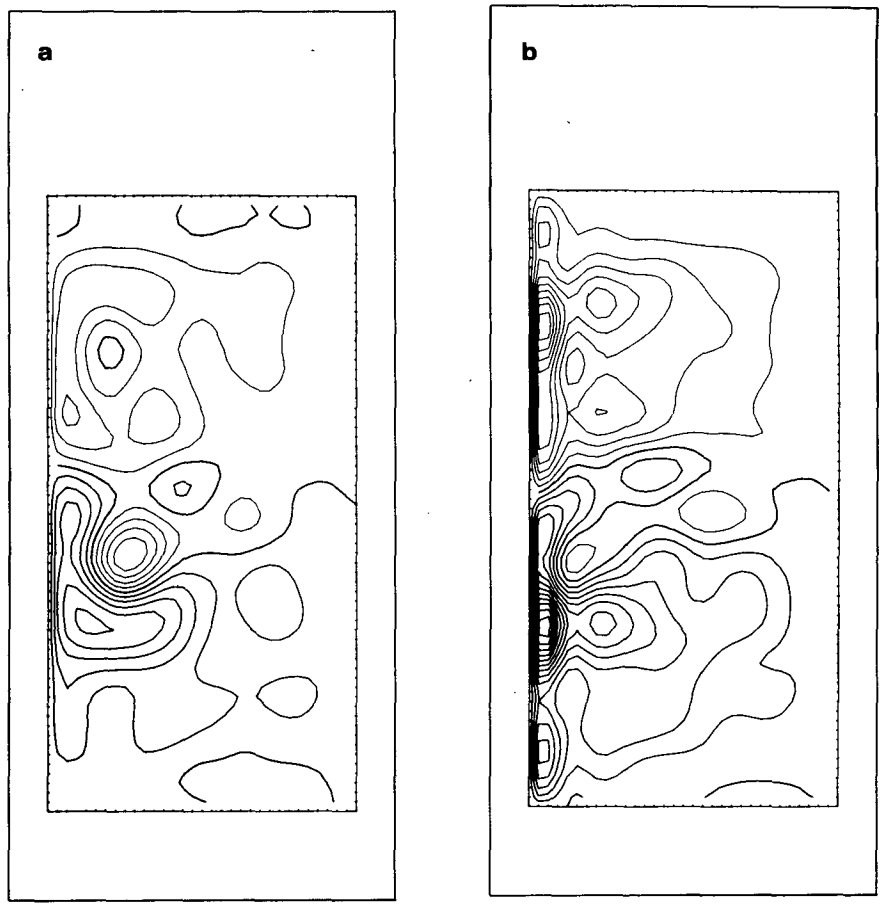

FIG. 13. The $\hat{h}(+)$ formed by continuous assimilation as in Fig. 8a (for the case of uncorrelated geoid errors) but using degraded models with (a) no wind-stress curl forcing (rms error = 10.2 cm) and (b) no nonlinear terms (rms error = 9.2 cm). C.I. = 6 cm.

velocity determination from altimetry. However, if one is only interested in the time-varying circulation, then a study of the difference of repeated passes suffices. On subtraction the geoid vanishes, greatly simplifying the analysis. Dynamical ocean models may have only limited application in the interpretation of height field *differences,* though, because ocean models integrate forward in terms of absolute velocity, i.e., surface elevation relative to the geoid. In many oceanographically interesting regions such as boundary currents and jets, the mean flow is strong and greatly distorts the planetary vorticity contours. Here the time-varying signal is embedded in and controlled by a mean potential vorticity field which departs markedly, but in an unknown way, from the planetary vorticity contours (latitude circles). Nevertheless, it must be said that in much of the ocean the mean flow is weak, and so a mean sea surface constructed from altimetric data alone is a rather good geoid which can be used as a reference to objectively map the variability using the least-squares technique outlined here.

Acknowledgments. I thank Dr. D. L. T. Anderson for his encouragement and interest during this study. I also acknowledge several short but extremely illuminating discussions on inverse theory with Drs. A. Hollingsworth and C. D. Rodgers. This work was supported by a NERC special topic award in Ocean Modelling and carried out in the Department of Atmospheric Physics, University of Oxford.

REFERENCES

Anderle, R. J., and R. L. Hoskins, 1977: Correlated errors in satellite altimetry geoids. *Geophys. Res. Lett.,* **4,** 421–424.

Bengtsson, L., 1975: 4-dimensional assimilation of meteorological observations. GARP No. 15, WMO.

Born, G. H., J. A. Dunne and D. B. Lame, 1979: Seasat mission overview. *Science,* **204,** 1405–1406.

Bretherton, F. P., R. E. Davis and C. B. Fandry, 1976: A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep Sea Res.,* **23,** 559–582.

Cheney, R. E., and J. G. Marsh, 1981: Oceanographic evaluation of geoid surfaces in the Western North Atlantic. *Oceanography from Space,* Plenum, 855–864.

Freeland, H. J., and W. J. Gould, 1976: Objective analysis of mesoscale ocean circulation features. *Deep Sea Res.,* **23,** 915–923.

Ghil, M., S. Cohn, J. Tavartzis, and E. Isaacson, 1981: Applications of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods,* Springer-Verlag, 139–224.

Gill, A. E. (Editor), 1984: Tropical Ocean-Global Atmosphere (TOGA) program report. WMO Rep.

Heiskanen, W. A., and H. Moritz, 1967: *Physical Geodesy,* Freeman.

Hollingsworth, A., 1984: Meteorological data analysis. ECMWF Lecture Note No. 2.2, Chap. 6.

——, and K. Arpe, 1982: Biases in the ECMWF data assimilation system. ECMWF Tech. Memo.

Jackson, D. D., 1979: The use of *a priori* data to resolve nonuniqueness in linear inversion. *Geophys. J. Roy. Astron. Soc.,* **57,** 137–157.

Marshall, J. C., 1984: Eddy mean-flow interaction in a barotropic ocean model. *Quart. J. Roy. Meteor. Soc.,* **110,** 573–590.

Marsh, J. G., and E. S. Chang, 1978: 5′ detailed gravimetric geoid in the northwestern Atlantic Ocean. *Mar. Geod.,* **1,** 253–261.

Moritz, H., 1978: Least squares collocation. *Rev. Geophys. Space Phys.,* **16,** 421–430.

Rapp, R. H., 1983: The determination of geoid undulations and gravity anomalies from SEASAT altimeter data. *J. Geophys. Res.,* **88,** 1552–1562.

Rodgers, C. D., 1976: Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. Space Phys.,* **14,** 609–624.

Simmons, V., and the MODE Group, 1977: The mid-ocean dynamics experiment. *Deep Sea Res.,* **25,** 859–910.

Wagner, C. A., 1983: The accuracy of the low-degree geopotential: implications for ocean dynamics. *J. Geophys. Res.,* **88,** 5083–5090.

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

——, and E. M. Gaposchkin, 1980: On using satellite altimetry to determine the general circulation of the oceans with application to geoid improvement. *Rev. Geophys. Space Phys.,* **18,** 725–745.

Zlotnicki, V., B. Parsons and C. Wunsch, 1982: The inverse problem of constructing a gravimetric geoid. *J. Geophys. Res.,* **87,** 1835–1845.