**Errors in dynamical fields inferred from oceanographic cruise data.
Part I: the impact of observation errors and the sampling distribution**

Damià Gomis⁽¹⁾ and Mike A. Pedder⁽²⁾

⁽¹⁾ Grup d'Oceanografia Interdisciplinar, IMEDEA (UIB-CSIC)
Campus UIB, 07071 Palma de Mallorca (Spain)
damia.gomis@uib.es

⁽²⁾ Department of Meteorology, University of Reading
2, Earley Gate, Reading RG6 6BB (UK)
m.a.pedder@reading.ac.uk

November 15, 2004

Corresponding author address: Damià Gomis, Dep. de Física,
Universitat de les Illes Balears.
07071 Palma de Mallorca, Spain
Tel.: +34 971 173236
Fax: +34 971 173426
e-mail: damia.gomis@uib.es

ABSTRACT

Diagnostic studies of ocean dynamics based on the analysis of oceanographic cruise data are usually quite sensitive to observation errors, to the station distribution and to the synopticity of the sampling. Here we present an error analysis of the first two sources (the third one is evaluated in Part II of this work). For observed variables and those linearly related to them, we use the Optimal Statistical Interpolation (OI) formulation. For variables which are not linearly related to observed variables (e.g., the vertical velocity), we carry out numerical experiments in a consistent way with OI statistics. Best results are obtained when some kind of scale selection or spatial filtering is applied in order to suppress small scales that cannot be properly resolved by the station distribution.

The formulation is first applied to a high resolution (SeaSoar) sampling aimed to the recovery of mesoscale features in a region of large spatial variability (noise-to-signal fraction of the order of 0.002). Fractional errors (rms error divided by the standard deviation of the field) are estimated in about 2% for dynamic height and between 4% and 20% for geostrophic vorticity and vertical velocity. For observed variables, observation errors and sampling limitations are shown to contribute in similar amounts to total errors. For derived variables, sampling errors are by far the dominant contribution. For less dense samplings (e.g., equally spaced CTD stations), fractional errors are about 6% for dynamic height and between 15% and 30% for geostrophic vorticity and vertical velocity. For this sampling strategy, errors of all variables are mostly associated with sampling limitations.

Key words: interpolation errors; sampling; oceanographic surveys; vertical motion; dynamical oceanography.

1. INTRODUCTION

Within the last 20 years, diagnostic studies of ocean dynamics have evolved along with observational and theoretical developments. Early studies used to focus on the description of the general circulation, basing mainly on the concept of water masses or, in the best case, estimating the geostrophic circulation from usually sparse observations of the mass field. Presently, studies are undertaken even at the mesoscale, taking advantage of fast sampling techniques such as towed undulating CTD probes (e.g., Pollard, 1986; Allen and Smeed, 1995) and ship-mounted Acoustic Doppler Current Profilers (e.g., Joyce, 1989; Gomis et al, 2001). Moreover, diagnostic studies are increasingly based on the analysis of dynamically relevant variables such as potential vorticity or the vertical velocity component. These dynamical fields can be obtained through the assimilation of observations in a numerical model (see for instance Nechaev and Yaremchuk, 1994; Morrow and De Mey, 1995). However, they are more often inferred from spatial distributions of observed variables assuming some kind of balance condition (see for instance Leach, 1987; Pollard and Regier, 1992).

The Quasi-Geostrophic (QG) theory provides a suitable theoretical framework for a diagnostic analysis of the sub-inertial and sub-tidal modes of 3D circulation in the ocean, and has therefore been used in a large variety of studies (Leach, 1987; Tintoré et al, 1991; Rudnick, 1996; Viúdez et al, 1996; among many others). In fact, the differences between QG results and those given by some more ‘exact’ theoretical framework such as the semi-geostrophic approximation (Pinot et al, 1996) or an iterative approximation to a primitive equation solution (Shearman et al, 2000), are usually small compared with the uncertainty associated with the recovery of observed fields from a set of scattered observations.

This uncertainty is mainly caused by the presence of errors in the observations, by the discrete nature of the sampling and by the lack of synopticity of the measurements. Although the impact of these three error sources is not fully independent, the departures of the retrieved fields with respect to the true fields will hereafter be referred to as ‘observational errors’, ‘sampling errors’ and synopticity errors’, respectively. They all have a negative impact on the interpolation of observations onto a regular grid. However, the impact is particularly relevant when spatial derivatives are computed by applying finite

difference methods to mapped fields, to the point that very unrealistic results can be obtained for high-order derived variables such as the vertical velocity.

One of the main motivations for this work is actually the link between vertical motion and biology in the ocean. It has been established, for instance, that the triggering of primary production is largely controlled by bio-geochemical fluxes associated with mesoscale vertical motion in the upper ocean (see for instance Fielding et al., 2001). The vertical velocity has also been related to the size spectrum of phytoplankton, which has important implications for the ocean carbon cycle (Rodriguez et al, 2001). Estimates of vertical motion at the mesoscale can be obtained via QG analysis of hydrographic data retrieved from oceanic surveys. However, in order to assess the reliability of the connection between vertical fluxes and the biological response, it is crucial to estimate the accuracy of the computed vertical velocity.

The main aim of this work is to produce a quantitative analysis of the errors involved in the computation of relevant dynamical variables (with particular attention to the vertical velocity). A second objective is to identify the main sources of analysis errors, in order to provide some guidance regarding the sampling strategy of oceanographic cruises. [Despite the obvious progress related to technical developments such as remote sensing, lagrangian profilers or tomography, oceanographic cruises still provide the most complete, 3D near-synoptic view of a particular ocean region.]

The formulation used to produce the error estimates follows the principles of Optimal Statistical Interpolation (hereafter referred simply as OI). OI was firstly used in atmospheric sciences (Gandin, 1963), but it has also been extensively used to interpolate oceanographic observations (see for example Bretherton et al., 1976; McWilliams et al., 1986; Chereskin and Trunnell, 1996). Regarding the estimation of analysis errors, the basic formulation is also given in the pioneering work of Bretherton et al. (1976). What is new in this work is the extension to variables that are not linearly related with observed variables (such as the vertical velocity). We also pay special attention to the concept of scale selection. This was also introduced by Bretherton et al. (1976), but in posterior works it has received less attention than, in our opinion, it deserves.

The methodology will be applied to the recovery of dynamic height, geostrophic velocity, geostrophic vorticity and vertical velocity fields inferred from an actual set of

profiles sampling a small region of the Western Mediterranean. The analysis presented in this work assumes that observations can be regarded as synoptic. Although this is not usually true in practice, the contribution of observational and sampling errors can be evaluated rather independently of synopticity errors, even in the case that the latter are significantly larger than the first. This can occur, for instance, in the presence of fast moving disturbances (see for instance Allen et al., 2001). Synopticity errors will be quantified in Part II of this work, where a method for error reduction is also proposed.

The structure of the paper is as follows. A brief review of OI formulation is first presented (section 2), focusing on the estimation of errors involved in the mapping of observed variables (e.g., dynamic height) and of errors involved in the computation of their spatial derivatives (e.g., geostrophic velocity and vorticity). Section 3 addresses the problem of variables that are not linearly related to observed variables, with particular attention to the vertical velocity. Section 4 deals with a key concept: the scale selection or smoothing of true fields. Results from applying the proposed formulation to a particular oceanographic survey are presented in section 5. They are extended to other case studies in a discussion section (section 6). Finally, conclusions are outlined in section 7.

2. THE OI ERROR FORMULATION

The principles of OI are described for instance in Bretherton et al. (1976). For simplicity, we consider only a 2D scalar field (such as dynamic height at constant depth), though similar principles apply to the analysis of 3D scalar data or vector field data.

The OI formulation is based on the idea that an observed field $\psi(\mathbf{r})$ can be represented as the sum of a large-scale ‘mean’ field $\Psi(\mathbf{r})$ and a smaller-scale ‘anomaly’ field $\psi'(\mathbf{r})$, where $\mathbf{r} = (x, y)$ denotes location on a horizontal plane. The key requirement is that the residual anomaly field behaves like a stationary, zero-mean random function of location, with variance $Var(\psi')$. In some applications (e.g., when OI is used to assimilate data into an operational forecasting model), the mean field is usually a forecast of the true field for the time of observations (referred to as ‘first guess’). For the analysis of cruise data, however, no first guess is usually available, and the mean field is typically modelled by a low-order polynomial surface. The polynomial coefficients can be estimated by

ordinary least-squares fitting or by a generalization of the OI minimization problem (e.g., Davies, 1985; Le Traon, 1990). The second option has clear advantages in undersampled regions or where the statistical mean field accounts for a significant part of the spatial structures. The analysis presented in this work focuses on the spatial structure of the anomaly field. This is, we neglect any error associated with the estimation of the mean field, since it is generally reasonable to assume that they are relatively small in most applications.

Conventionally, the observations ψ_i are assumed to be of the form

$$\psi_i = \Psi(\mathbf{r}_i) + \psi'(\mathbf{r}_i) + \varepsilon_i, \quad (1)$$

where ε_i is an ‘observation error’ that is assumed to be spatially uncorrelated. Given a reasonable estimate of $\Psi(\mathbf{r})$, the problem of estimating $\psi(\mathbf{r})$ at a grid point $\mathbf{r}_g = (x_g, y_g)$ then reduces to one of estimating the random anomaly $\psi'(\mathbf{r}_g)$. The ‘optimal’ solution is defined as that linear combination of observed anomalies which minimises (in a statistical sense) the deviations between the prediction and the true field. Assuming that the anomaly field variance $Var(\psi')$ is constant over the domain, the solution is given by

$$\hat{\psi}'(\mathbf{r}_g) = \underline{\rho}_g^T \underline{\Omega}_{oo}^{-1} \underline{\psi}'_o. \quad (2)$$

Here, $\underline{\psi}'_o$ is a data vector containing the observed anomalies $(\psi_i - \Psi(\mathbf{r}_i))$, $i=1,2,\dots,N$, $\underline{\Omega}_{oo}$ is a matrix of anomaly autocorrelations evaluated for all pairs of observation points, $\underline{\rho}_g$ is a vector of autocorrelations with i ’th entry equal to the autocorrelation between the anomaly variables sampled at $\mathbf{r}_g$ and $\mathbf{r}_i$, and superfix T denotes transposition.

Historical data are often too poor for the computation of a meaningful spatially-dependent lag-autocorrelation statistics. Hence, a common assumption is that lag-autocorrelation is homogeneous and isotropic, so that when averaged over the observed domain it can be approximated by some positive-definite function of separation distance $\rho(R_{ij})$, where $R_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ (see for instance Thiébaux and Pedder, 1987). This is used to generate the entries in $\underline{\rho}_g$ and the off-diagonal entries in $\underline{\Omega}_{oo}$. The diagonal entries in $\underline{\Omega}_{oo}$

are all equal to $(1 + \eta)$, where η is the so-called ‘noise-to-signal ratio’ of the observed anomalies $Var(\varepsilon)/Var(\psi')$.

It is worth noting that an isotropic correlation model can be a reasonable approach even in frontal regions, provided a suitable form for the mean field $\Psi(\mathbf{r})$ is used. If the latter is chosen so as to account for most of the anisotropy (e.g., modelling the front with a simple linear field), the anomaly field can reasonably fulfil the simplifying hypotheses.

2.1 Accuracy of mapped variables

A useful feature of OI is that it provides an a-priori measure of the statistical accuracy of mapped variables. In particular, the variance of the statistical error on the grid-point variable $\hat{\psi}(\mathbf{r}_g)$ is given by

$$Var(e_\psi(\mathbf{r}_g)) = Var(\psi') \times (1 - \underline{\rho}_g^T \underline{\Omega}_o^{-1} \underline{\rho}_g), \quad (3)$$

where $e_\psi(\mathbf{r}_g) = \{\hat{\psi}(\mathbf{r}_g) - \psi(\mathbf{r}_g)\}$ is the error of the mapped field evaluated at $\mathbf{r}_g$. Within the domain, $Var(e_\psi(\mathbf{r}_g))$ may vary from one analysis point to another, in response to changes in the spatial distribution of observations in the vicinity of $\mathbf{r}_g$ (though $Var(e_\psi(\mathbf{r}_g))$ can never exceed $Var(\psi')$).

For further applications, it will be necessary to know not only the error variance of $\hat{\psi}$ at every analysis point $\mathbf{r}_g$ (given by (3)), but also the spatial correlation of analysis errors. In order to obtain this information, the analysis can be represented by an M -vector $\underline{\hat{\psi}}$, whose entries contain the values of $\hat{\psi}(\mathbf{r})$ evaluated on the M nodes of a regular grid. Denoting the vector of analysis errors by $\underline{e}_\psi = \underline{\hat{\psi}} - \underline{\psi}$, the analysis-error covariance matrix for $\hat{\psi}$ is defined by $\underline{\Sigma}_\psi = E(\underline{e}_\psi \underline{e}_\psi^T)$, where E is the expectation operator. Using (2), this error covariance matrix is given by

$$\underline{\Sigma}_\psi = Var(\psi') \times (\underline{\Omega}_G - \underline{\Omega} \underline{\Omega}_o^{-1} \underline{\Omega}^T) \quad (4)$$

Here, $\underline{\Omega}_G$ is an $M \times M$ correlation matrix containing the anomaly correlation between grid point pairs. $\underline{\Omega}$ is an $M \times N$ correlation matrix with g 'th row equal to $\underline{\rho}_g^T$, as in (2).

The diagonal entries in $\underline{\Sigma}_{\psi}$ contain the values of $\text{Var}(e_{\psi}(\mathbf{r}))$ at the analysis points, while off-diagonal entries represent the error covariance between pairs of analysis points.

2.2 Accuracy of linearly related variables

The diagnostic analysis of a sampled feature often involves operating linearly on one or more mapped variable fields. For example, if $\hat{\psi}(x, y)$ is a mapped field of the geostrophic streamfunction (or dynamic height) on some constant depth surface, then variables such as geostrophic velocity and vorticity may be estimated by applying horizontal difference operators to the mapped data. The accuracy of an estimated derivative of $\psi(\mathbf{r}_g)$ depends not only on the magnitude of the error on $\hat{\psi}$ evaluated at $\mathbf{r}_g$, but also on the extent to which the latter correlates with errors on $\hat{\psi}$ evaluated on points in the vicinity of $\mathbf{r}_g$.

Let $\hat{\underline{\zeta}}$ be a data vector generated by operating linearly on $\hat{\underline{\psi}}$, so that $\hat{\underline{\zeta}} = \mathbf{H}\hat{\underline{\psi}}$, where $\mathbf{H}$ is an $M \times M$ matrix containing the coefficients of the difference operator. Defining the ‘true’ values of ζ as those obtained by applying the same operator to the true values of ψ , the associated vector of analysis errors on $\hat{\underline{\zeta}}$ will be given by $\underline{e}_{\zeta} = \hat{\underline{\zeta}} - \mathbf{H}\underline{\psi}$. And defining the analysis-error covariance matrix for $\hat{\underline{\zeta}}$ as $\underline{\Sigma}_{\zeta} = E(\underline{e}_{\zeta}\underline{e}_{\zeta}^T)$, where E is the expectation operator, it can be shown that $\underline{\Sigma}_{\zeta}$ and $\underline{\Sigma}_{\psi}$ are simply related by

$$\underline{\Sigma}_{\zeta} = \text{Var}(\psi') \times \mathbf{H}(\underline{\Omega}_G - \underline{\Omega}\underline{\Omega}_o^{-1}\underline{\Omega}^T)\mathbf{H}^T = \mathbf{H}\underline{\Sigma}_{\psi}\mathbf{H}^T \quad (5)$$

An alternative approach is to derive a continuous convolution of the linear operator $\mathbf{H}$ with the correlation model of $\underline{\Omega}_G$. In this way, $\underline{\Sigma}_{\zeta}$ (and the analysis of the linear variable itself) should be insensitive to the choice of the analysis grid.

2.3 Partition of analysis errors

With regard to the interpretation of analysis-error covariance maps, it can be useful to associate the error on the mapped variable $\hat{\psi}$ with two sources. A first one is the influence of observation errors. It is not difficult to show that for an analysis based on (2), the latter contributes to $\underline{\Sigma}_{\psi}$ by an amount given by

$$\underline{\underline{\Sigma}}_{\psi\eta} = \eta \text{Var}(\psi') \underline{\underline{\Omega}} \underline{\underline{\Omega}}^{-2} \underline{\underline{\Omega}}^T \quad (6)$$

This measures the ‘noise-response’ of the analysis system, which depends on the noise-to-signal ratio (it is obviously zero in the case where $\eta = 0$) but also on the station distribution (so that it cannot be regarded as fully independent of sampling errors).

The sampling contribution ($\underline{\underline{\Sigma}}_{\psi/s}$) is defined as what is left when the observational contribution $\underline{\underline{\Sigma}}_{\psi\eta}$ is subtracted from total error covariances $\underline{\underline{\Sigma}}_{\psi}$. It accounts for the error covariances that would be obtained if (2) was actually applied to ‘perfect’ observations sampling $\psi(\mathbf{r})$ without error. For a given correlation model and prescribed value of η , $\underline{\underline{\Sigma}}_{\psi/s}$ depends only on the spatial distribution of the observation points. In this sense it can be regarded as measuring the interpolation error associated with inadequate spatial sampling of all those spatial scales which influence the observations. The same kind of partition can be applied to the error-covariance matrix of linearly related variables (5). For non-linear variables, the partition between observational and sampling errors will also be possible, but it will require a different approach.

3. ERROR FORMULATION FOR NON-LINEAR OPERATORS

Of particular interest in this study is the accuracy of vertical motion diagnosed by solving a quasi-geostrophic omega equation. Following Holton (1992), a form of the omega equation suitable for the analysis of mesoscale features in the ocean can be written as

$$N^2 \nabla_h^2 w + f_o \frac{\partial^2 w}{\partial z^2} = f_o \frac{\partial}{\partial z} (\mathbf{V}_g \cdot \nabla_h \xi_g) + \frac{g}{f_o} \nabla_h^2 (\mathbf{V}_g \cdot \nabla_h \rho) , \quad (7)$$

where N is the basic-state buoyancy frequency, ξ_g is the geostrophic relative vorticity, ∇_h is the horizontal gradient operator, and other symbols have conventional meaning. It is not difficult to show that the forcing terms on the right-hand side of (7) can be expressed as the result of a non-linear operator applied onto dynamic height alone.

Mc Williams et al. (1986) addressed the problem of optimally estimating products of observed variables from products of observations. Because this involves fourth moments which are rarely known from observations, they were forced to assume quasi normality.

However, the analysis of w is further more complicate than a product of observed variables, as it also involves integrating (7) subject to appropriate boundary conditions. Therefore, we followed the common approach of estimating the forcing terms from a 2-D mapping of observed dynamic height data on a series of constant depth surfaces. Whichever approach is followed, it is clear that the method of predicting analysis errors for simple spatial derivatives cannot be applied to the problem of predicting errors for the forcing terms in (7) or for w itself. In order to estimate it consistently with the spatial statistics of the observed variables, we developed the following procedure.

Let us consider some parameter ζ which is to be diagnosed from a purely 2-D field, based on a discrete approximation to $\zeta = H\{\psi(x, y)\}$, where H here represents a non-linear operator. And let $\mathbf{L}\mathbf{L}^T = \underline{\underline{\Omega}}_G$ represent the Cholesky factorisation of the correlation matrix for the anomaly variable ψ' sampled on the analysis grid, as in (4). Then, one grid point realisation of $\underline{\psi'}$ is generated by

$$\underline{\psi'} = \mathbf{L}\underline{\delta} , \quad (8)$$

where $\underline{\delta}$ is a vector of random numbers drawn from a normal distribution with zero mean and variance equal to $Var(\psi')$, so that $E\{\underline{\psi'}\underline{\psi'}^T\} = \underline{\underline{\Omega}}_G$. The values of realisation ψ' at observation points can be obtained by simple bilinear interpolation between gridpoints, provided that the gridspace is relatively small compared with the correlation scale of ψ' . However, we preferred the alternative of applying the Cholesky factorisation to a correlation matrix associated with ψ' sampled both at gridpoints and observation points.

Observation errors can be simulated by sampling a random-normal distribution with zero mean and variance equal to $Var(\varepsilon)$ and added to the values of ψ' at observation points. Finally, the assumed contribution to $\psi(x, y)$ from the mean field $\Psi(\mathbf{r})$ must be added to the simulated anomalies when the diagnosed variables of interest depend on the mean field.

The full analysis system (i.e., interpolation of observations followed by the application of the non-linear operator onto the output) can then be applied to the simulated dynamic height observations. On the other hand, the non-linear operator can also be applied to the simulated, ‘true’ gridpoint dynamic height data. Total analysis-error variance

and covariance statistics can then be obtained by averaging the differences between both outputs over a very large number of realisations. Sampling errors can be evaluated in the same way, but without adding noise to the station data (so that it reflects errors associated with an interpolation applied to perfect observations). Finally, the observational contribution can be obtained simply as the difference between total and sampling errors.

In the particular case of (7), however, an evaluation of the forcing term from a given dynamic height field actually involves vertical as well as horizontal derivatives. This suggests that the simulated realisations of the observed dynamic height field should be based on a fully 3-D covariance model. Although feasible, such an approach is very expensive in numerical terms. Fortunately, a simpler approach is possible provided all profiles sample the same vertical range. In that case, it may be possible to approximate the observed profiles using a low-order EOF expansion, as demonstrated by Pedder and Gomis (1998). Thus, at horizontal location (x_i, y_i) the observed profile $\underline{\psi}_i$ is approximated by

$$\underline{\psi}_i[K] = \sum_{k=1}^K a_{k,i} \underline{\Phi}_k, \quad (9)$$

where $\underline{\Phi}_k$ is the k 'th leading EOF, $a_{k,i} = \underline{\psi}_i^T \underline{\Phi}_k$ is the value of the k 'th EOF amplitude at horizontal location (x_i, y_i) , and K is the number of leading EOFs included in the expansion. For each k , $a_{k,i}$ can now be treated as an 'observation' of a 2-D field $a_k(x, y)$, and a value of $a_k(\mathbf{r}_g)$ thus estimated by conventional OI using a 2-D model for the spatial statistics of a_k and its observation error. By analogy with (1), this approach assumes that the 'observed' amplitudes are of the form

$$a_{k,i} = A_k(\mathbf{r}_i) + a'_k(\mathbf{r}_i) + \varepsilon_{k,i}. \quad (10)$$

It follows that a simulation of the amplitude field associated with each EOF can be generated by the same method as used to simulate an observed 2-D field. If the leading EOF explains a large fraction of the observed spatial variance in the diagnosed parameters, then a simulation of only the leading EOF contribution may be sufficient for the purpose of estimating analysis accuracy.

It is worth noting that considering amplitude errors ($\varepsilon_{k,i}$) is equivalent to assume an error vertical correlation of the form of the EOF. In practice, instrumental errors usually

do have a vertically correlated component. For dynamic height, which is integrated from the reference level up to each pressure level, such systematic errors in the observed variables result in higher errors at upper levels than at lower levels. Although the error profiles do not necessarily have the same shape as the EOF profile, the considered approach is probably more realistic than considering no vertically correlated errors.

4. THE KEY CONCEPT OF SCALE SELECTION

In the model for observations (1), the sum of the mean and anomaly field represents the response of the measurement system to all scales of spatial variation in $\psi(\mathbf{r})$. Consequently, OI in principle provides a minimum variance estimate of all spatial variations in the observed field. In spectral terms, the Fourier transform of $\rho(R)$ should describe the distribution of measured variance with respect to spatial wavelength, down to some lower limit $\lambda_{\min}$ determined only by the physical characteristics of the sensors.

In practice, the assumed correlation models do not necessarily have contributions at all wavelengths present in the observed field. In particular, the smallest scale represented by simple models $\lambda_{\min}^{cor}$ can be significantly larger than $\lambda_{\min}$. Nevertheless, for most oceanographic survey experiments, $\lambda_{\min}^{cor}$ is still likely to be smaller than the smallest horizontal wavelength which can be resolved unambiguously (in terms of amplitude and phase) from the given spatial distribution of observations. Conventionally, the latter is considered to be of order $\lambda_N = 2 \times \Delta$ (though more conservative authors suggest to take $\lambda_N = 4 \times \Delta$), where Δ is the separation distance between adjacent observation points. Smaller wavelength features affecting the observations then contribute to analysis error on the resolved wavelength features with $\lambda \geq \lambda_N$. This aliasing effect is likely to be more significant at resolvable wavelengths not much larger than the Nyquist wavelength λ_N , as shown by Bretherton et al. (1976).

In general, the influence of aliasing errors needs not be considered if the only objective is to describe spatial features with reference to their shape. However, these effects may seriously affect the estimation of spatial derivatives of mapped fields, to the point that it might be difficult to attach much confidence to diagnosed features which

appear to be of dynamical significance. The reason is that difference operators amplify the contribution from high wavenumber components relative to those associated with low wavenumbers. Hence, it may be desirable to target the analysis at resolving only a subset of spatial scales, so that the smallest resolved wavelength is somewhat larger than λ_N .

In practice, the problem of scale deletion or ‘smoothing’ has been addressed in different ways. In many works, the noise-to-signal parameter η is subjectively increased until the interpolation output shows a suitable degree of smoothing. In principle, the ‘optimallity’ of the interpolation is lost in this way, since an increase in η is only suitable for suppressing purely white (spatially uncorrelated) noise. However, it must be admitted that treating small scales as spatially uncorrelated noise may not be a stronger assumption than, for instance, the homogeneity and isotropy of autocorrelation functions.

Perhaps more important regarding this work is the fact that if non resolvable scales are treated as spatially uncorrelated noise, the distinction between instrumental and sampling errors made in section 2.3 would no longer make sense. The reason is that the two contributions would mix up within the noise-to-signal ratio η . Therefore, we followed the method (already suggested by Bretherton et al., 1976) of applying a numerical low-pass filter to the estimated anomaly variables.

When this is applied to grid point values, the result can be represented by

$$\underline{\hat{\psi}}' = \mathbf{F} \underline{\hat{\psi}}' , \quad (11)$$

where $\underline{\hat{\psi}}'$ is the analysis generated on the M gridpoints using (2), and $\mathbf{F}$ is an $M \times M$ matrix whose rows contain the elements of the low-pass filter. As for the case of linearly derived variables, the smoothed estimation (11) can be made insensitive to the choice of the analysis grid by deriving a correlation model $\bar{\rho}(R)$ from the continuous convolution of the smoothing filter with $\rho(R)$ (see, for example, Pedder, 1993). In such case, the conventional OI solution (2) is replaced by

$$\hat{\psi}(\mathbf{r}_g) = \bar{\rho}_{-g}^T \Omega_o^{-1} \psi_o' . \quad (12)$$

Following this method, the cut-off wavelength of the spatial filter can be set in an objective way (e.g., basing on the station separation distance), in contrast to the option of subjectively increasing the noise-to-signal parameter.

A key feature regarding analysis errors is that since the analysis is targeted at resolving only a subset of the observed scales, the error in the ‘smoothed’ analysis should be defined with respect to the ‘smoothed’ truth $\underline{\bar{\psi}}' \equiv \mathbf{F}\underline{\psi}'$. The corresponding error covariance matrix for $\hat{\underline{\psi}}$ is then given by

$$\underline{\bar{\Sigma}}_{\underline{\psi}} = \text{Var}(\psi') \times \mathbf{F} \left(\underline{\bar{\Omega}}_G - \underline{\bar{\Omega}} \underline{\bar{\Omega}}_o^{-1} \underline{\bar{\Omega}}^T \right) \mathbf{F}^T, \quad (13)$$

which is of course analogous to (4) but with the smoothing filter $\mathbf{F}$ replacing the linear operator $\mathbf{H}$. When the filter is convoluted with the correlation functions as in (12), the corresponding analysis-error covariance matrix for $\hat{\underline{\psi}}'_o$ is given by

$$\underline{\bar{\Sigma}}_{\underline{\psi}} = \text{Var}(\psi') \times \left(\underline{\bar{\Omega}} - \underline{\bar{\Omega}} \underline{\bar{\Omega}}_o^{-1} \underline{\bar{\Omega}}^T \right), \quad (14)$$

where the elements of $\underline{\bar{\Omega}}$ are calculated from $\bar{\rho}(R)$ and the elements of $\underline{\bar{\Omega}}_g$ are calculated from a function $\bar{\bar{\rho}}(R)$ obtained by an additional convolution of the filter with $\bar{\rho}(R)$.

5. RESULTS

The formulation presented above has first been applied to an oceanographic survey carried out in the Alborán Sea (Western Mediterranean) onboard BIO *Hespérides*. A 100x80 km² domain was sampled along 11 N-S legs spaced 10 km using a towed undulating CTD probe (SeaSoar) for hydrographic observations and an Acoustic Doppler Current Profiler (ADCP) for velocity measurements. The continuous SeaSoar saw-tooth record was averaged over 8 m (in the vertical) x 2 km (along-track) cells, which resulted in 216 vertical profiles sampled at 50 levels (from 5 m to 397 m). A subset of 73 stations conforming a more or less regular 2D distribution (station separation of 10 km) will also be considered, in an attempt to simulate a more conventional CTD sampling (see Fig. 1). ADCP data have not been considered in this work.

5.1 Errors in interpolated fields of observed variables: dynamic height

The key observed variable for QG dynamics is dynamic height. In our particular case, it was referred to 301 m (any level below 250 m can be considered as a no-motion level in the Alborán Sea). The large-scale ‘mean’ field $\Psi(\mathbf{r})$ was first estimated at every level by fitting a 2D bilinear polynomial to the set of observations ψ_i . The observed anomalies $\psi'_i = \psi_i - \Psi(\mathbf{r}_i)$ were subsequently computed together with their variance $\text{Var}(\psi')$. The values obtained for the latter were about 38 (dyn cm)^2 at surface (5 m), 10 (dyn cm)^2 at 100 m and decreased rapidly to less than 0.1 (dyn cm)^2 below 200 m.

A first parameter required by the OI analysis is the noise-to-signal variance ratio η . Cross-validation experiments carried out in the region by Gomis et al. (2001) indicate that for dynamic height, this ratio depends on depth, ranging from order 0.001 at surface to order 0.01 below 200 m. For the variances given above, this translates in random noise variances of the order of 0.038 (dyn cm)^2 at surface and of 0.001 (dyn cm)^2 below 200 m (standard deviations of 0.195 and 0.031 dyn cm, respectively). The fact that noise is significantly higher at upper levels can be explained in two ways. A first one is that, as stated above, temperature and salinity sensor errors can be correlated within a single profile, in which case the impact on dynamic height would be directly related to the upwards integration distance from the reference level. A second reason could be the influence at upper levels of internal waves or near-inertial motions, which would be interpreted as random noise by cross-validation experiments.

For the calculations reported below, a compromise value of 0.002 was assumed for η . This value can be considered representative for intermediate levels of frontal regions such as the Alborán Sea, characterised by sharp density gradients (and therefore high anomaly field variances). In less baroclinic regions, the dynamic height field variance may be much smaller, whereas the observational noise might be expected to be about the same, implying a much larger η . Results for larger values of η will be considered later on.

For the anomaly correlation, an isotropic Gaussian model $\rho(R) = \exp(-R^2 / 2L^2)$ was assumed, with the characteristic scale L set to 15 km. The latter is based on estimates for this region obtained by Gomis et al. (2001). Unlike the noise-to-signal parameter,

which can be highly dependent on the study region, values of L between 15 and 25 km are common to most regions with significant mesoscale structures.

Figure 2a shows the result of solving (3) in the case of a dynamic height field analysis derived from the full station distribution shown in Fig. 1. Relatively large errors are well apparent near domain boundaries and data voids, where the interpolation accuracy is reduced due to the asymmetry of the observation distribution. Within the inner domain, the ratio of total rms errors to the standard deviation of the anomaly field (hereafter “fractional error”) varies between 2.0 and 2.5%, but approaches 6% in the data void close to the northern boundary. The contribution of observational noise (Fig. 2b), given by the diagonal elements of $\sum_{\psi\eta}$ in (6), is of the same order as total errors in the inner domain. This implies that the accuracy of the dynamic height analysis within the inner domain is limited mainly by the influence of observing errors. On the other hand, analysis accuracy near domain boundaries is limited mainly by the influence of sampling error (Fig. 2c).

Figure 3 shows dynamic height analysis errors in the case where scale selection has been applied. The error covariance was computed by solving (14) with ‘smoothed’ correlation functions obtained via a convolution of a normal error filter with the autocorrelation function $\rho(R)$. The filter scale was chosen to remove most of the spatial variance associated with wavelengths smaller than 40 km (four times the separation distance between sampling legs). This results in a substantial reduction in fractional errors within the inner domain, where they range from less than 1.5% to 2.5% (Fig. 3a). The influence of the data void is also significantly reduced. In the inner domain, observational errors are still the dominant contribution (about 1.5%, Fig. 3b), the sampling errors ranging from less than 0.5% to 2.0% (Fig. 3c).

Compared with the result of the ‘unfiltered’ analysis, it is clear that although sampling errors are reduced, the influence of the boundaries penetrates further into the domain. This is an inevitable consequence of applying a smoothing filter to the ‘discontinuous’ data distribution at a boundary, and obviously represents a disadvantage of the scale selection. However, this disadvantage should be more than offset by the resulting increase in overall analysis accuracy within the inner domain.

5.2 Errors in linearly related variables: geostrophic velocity and relative vorticity

Errors on geostrophic velocity and geostrophic relative vorticity fields derived from spatial derivatives of the dynamic height field can be both obtained from (5). In the absence of any scale selection, total fractional errors for geostrophic velocity range from 6 to 10% in the inner domain (not shown). For geostrophic vorticity they range from 9 to 12%, compared with 2 to 6% for dynamic height. The error distribution and the relative contributions from observational and sampling errors have similar shapes to those obtained for dynamic height. In the case of geostrophic velocity, the contribution from observation error is still the dominant one within the inner domain, whereas for geostrophic vorticity both contributions are more similar.

The fractional error on geostrophic velocity and relative vorticity derived after applying scale selection to the interpolated dynamic height field is significantly smaller: about 2.5% for the first and less than 4% for the second (not shown). However, the combined effect of the filtering and data distribution results in relatively large fractional errors (of the order of 20%) near the domain boundaries.

In order to better appreciate the significance of these results, it is useful to translate a fractional error into an rms error value on the diagnosed variable by multiplying by the standard deviation (std) of the ‘true’ anomaly field. Results for the particular survey of Fig. 1 are given in Table 1. For dynamic height and in absence of scale section, errors are of the order of 0.1 dyn cm in the upper 100 m, where the anomaly field std is of several dyn cm. Filtering out wavelengths shorter than 40 km reduces the errors to about a half, at the expenses of a very small reduction in the anomaly field std. This emphasizes the large contribution of small scales (i.e., those filtered out by the scale selection) to analysis errors, whereas these scales just account for a small fraction of the total field variance.

For geostrophic velocity, errors are of a few cm/s in absence of filtering and of less than 1 cm/s when scale selection is applied. These errors are at most of the same order than the accuracy of velocity measurements obtained with a ship-mounted ADCP. For geostrophic relative vorticity and in absence of filtering, errors are of the order of $0.2\text{--}0.3 \cdot 10^{-5} \text{ s}^{-1}$ (i.e., $0.02\text{--}0.03 \text{ f}$) in the upper 100 m, where the anomaly std is of the order of $2\text{--}4 \cdot 10^{-5} \text{ s}^{-1}$ ($0.2\text{--}0.4 \text{ f}$). When scale selection is applied, errors reduce to $0.005\text{--}0.01 \text{ f}$. However, it is worth noting that the fraction of variance accounted for by small scales is

larger than for dynamic height. This feature will be more apparent for the vertical velocity and is due to the downscaling resulting from the application of differencing operators onto observed fields.

5.3 Errors associated with non-linear operators: the vertical velocity

The methods summarized in section 3 were used to estimate the error on the QG vertical velocity (w). In the Alborán Sea, the first leading EOF (EOF₁) accounts for more than 90% of the observed dynamic height variance (Pedder and Gomis, 1998), so that the contribution of this single component can be considered a good approximation to the total field. It follows that a good approximation to the full 3D dynamic height field can be obtained by applying the OI analysis scheme to a single set of observed EOF₁ amplitudes. The horizontal scale of the corresponding amplitude anomaly correlation was estimated to be 15 km. The variance and noise-to-signal variance ratio of the observed amplitudes were estimated to be 380 (dyn cm)^2 and 0.00028 respectively.

250 realisations of the amplitude anomalies sampled on both the grid points and the station points were obtained using (8). The Cholesky factorisation was applied to a correlation matrix with entries associated with anomaly variables sampled at both the grid points and the observation points. [Results were actually rather insensitive to the number of realisations, provided this exceeded the order of 100.]

We firstly consider the case where the mean field $\Psi(\mathbf{r})$ is a simple constant, so that the solution of (7) depends only on the structure of the anomaly field. Figure 4 shows the relevant dynamical fields at an intermediate level (101 m) corresponding to one single ‘true’ realization. The downscaling of dominant structures in derived variables with respect to observed variables is well apparent.

This Figure also illustrates the sensitivity of the vertical velocity with respect to the imposed boundary conditions, a compulsory point to be checked since the error estimates shown later on will not consider this contribution. In practice, it is very common to set $w=0$ at the upper and lower boundaries, which is a reasonable assumption provided the vertical domain ranges from the ocean surface down to a meaningful reference level. Lateral boundary conditions are usually set in a more arbitrary way, common choices being setting $w=0$ or Neumann conditions. However, provided the horizontal scale of the

structures is smaller than the size of the domain, the ellipticity of (7) makes that the interior solution for w is relatively insensitive to the imposed conditions. This is demonstrated in Figs. 4d and 4e: the first was obtained imposing $w=0$ at the surface and bottom of the domain and Neumann conditions at the lateral boundaries; instead, Fig. 4e was obtained imposing typical positive and negative constant values of w at the eastern and southern boundaries (and Neumann conditions at the western and northern boundaries). Hence, the following error analysis may be considered to yield realistic error estimates except at the few grid points closer to the boundaries.

The error distribution at an intermediate level, where vertical velocities are most significant, is shown in Fig. 5a. The fractional error on w is in this case of the order of 5-6% right in the centre of the domain, but exceeds 20% between half a way between the centre of the domain and the lateral boundaries. A relevant feature is that, unlike for dynamic height, sampling errors account for most of total errors even in the inner domain, where it is about 5% compared with the 3% of the contribution from observational noise (not shown). This is another consequence of the commented downscaling of the dominant structures, which makes the field more sensitive to the sampling distribution. The vertical velocities simulated at intermediate levels are about $3.4 \cdot 10^{-4}$ m/s (of the order of 30 m/day), so that errors in the inner domain would typically be between 2 and 6 m/day (Table 1).

Figure 5b confirms once more the effectiveness of scale selection: the fractional error is reduced to less than 4% in the centre of the domain. However, the boundary effects derived from the filtering penetrate almost until the centre of the domain. Another worth noting point, advanced when commenting the results of linear derived variables, is that the vertical velocities associated with the non-filtered dynamic height scales only account for a small fraction of the total field variance. This could be a drawback when the discarded scales represent meaningful QG processes. In our case, however, filtering wavelengths shorter than 40 km implies neglecting structures of less than 10 km radius, which is a lower limit for the validity of QG theory in the region. The velocities simulated at intermediate levels are now less than 10^{-4} m/s (or 10 m/day) compared to the $3.4 \cdot 10^{-4}$ m/s of the unfiltered field (Table 1). Hence, typical errors would be between 0.5 and 2 m/day.

The above examples do not take into account the structure of the dynamic height mean field $\Psi(\mathbf{r})$, which has so far been treated as a simple constant, corresponding to a

zero mean geostrophic velocity. As noted earlier, variables estimated by applying linear operators to the interpolated dynamic height field should not be sensitive to the large scale trend if this is approximated by a low-order polynomial function of spatial coordinates. However, this does not necessarily hold for the forcing terms in (7), due to the contribution from advective terms whose magnitudes will depend partially on the value of the mean geostrophic velocity and its variation with depth. A more realistic error analysis should therefore take into account any large-scale trend which contributes significantly to the geostrophic velocity. In practice, it is probably reasonable to ignore any error in the mean geostrophic velocity itself, in which case the error analysis can proceed as before, but with the mean field added to the simulated anomaly data before solving (7).

We therefore repeated the simulations but now considering a mean field $\Psi(\mathbf{r})$ corresponding to a constant geostrophic velocity. It was obtained simply by fitting a bilinear trend to the Alborán Sea dynamic height data at each horizontal level. The associated geostrophic velocities were predominantly eastward, with speeds ranging from 62 cm/s at surface to 10 cm/s at 149 m depth (below 200 m the velocity associated with the mean field was negligible). Fractional error values obtained in this case did not differ significantly from the values reported for the case where the dynamic height mean field $\Psi(\mathbf{r})$ was a simple constant and are therefore not shown.

6. DISCUSSION

A key question regarding the above results is the extent to which they can be extrapolated to other regions and sampling patterns. The case-to-case differences will mainly depend on the homogeneity and density of the station distribution and on the magnitude of the noise-to-signal ratio (η). This dependence is examined in the following through a set of case studies; the error values reported for each case will always correspond to the inner domain and to analyses performed with scale separation.

Regarding the station distribution, it is clear from the examples above that any small data void will produce high errors in its vicinity. For regular distributions, errors will basically depend on the ratio between the correlation scale and the mean station separation. [The whole OI formulation could actually be set up in terms of relative distances, so that any survey with a similar ratio (e.g. 15 km correlation scale relative to 10 km station

separation) should result in similar fractional errors for observed variables and linearly derived variables (not necessarily for the vertical velocity, due to the non-linearity of dynamics).]

Because in practice, SeaSoar surveys are still less common than CTD surveys, we firstly considered a station distribution with the same station separation in both directions. Namely, we selected a subset of the almost continuous SeaSoar sampling in such a way that stations were also spaced 10 km along track (Fig. 1). The obtained fractional errors were about 4%, compared to the 1-2% obtained for the full SeaSoar sampling (compare Figs. 6a and 3a). For geostrophic velocity and relative vorticity (not shown), fractional errors were about 5% and 6% respectively, compared to the 2% and 3% obtained for the full SeaSoar sampling (see Table 2 for error values at a particular level). As expected, sampling errors are significantly higher for the CTD sampling than for the full SeaSoar sampling, now becoming the dominant contribution to total errors. This result is of practical significance, as it establishes a clear difference between the capabilities of SeaSoar and CTD surveys to resolve the mesoscale circulation (here represented by a dominant 15 km correlation scale).

The above analyses have all assumed a value of total field variance which is typical of the Alborán Sea region. Since the noise-to-signal parameter η is inversely proportional to the field variance (the variance of instrumental noise is likely to be similar for most oceanographic cruises), it turns out that η will be significantly larger in regions of lower spatial variability. As an example, we considered a noise-to-signal parameter $\eta = 0.02$ (i.e., corresponding to a region with a dynamic height anomaly variance one order of magnitude smaller than in the Alborán Sea). For the full SeaSoar sampling, fractional errors were about 3-4% for dynamic height (Fig. 6b), 5% for geostrophic velocity and 7% for relative vorticity, compared with the 1%, 2% and 3% obtained for $\eta = 0.002$ (see Table 2 for error values at a particular level). Regarding the partition between observational and sampling errors, the contribution of the first increases in regions of lower spatial variability, whereas sampling error remains approximately constant.

Finally, we considered the case of having a noise-to-signal parameter $\eta = 0.02$ and the subset of 73 stations taken as representative of a standard CTD sampling. The fractional errors obtained were about 6% for dynamic height (Fig. 6c), 9% for geostrophic

velocity and 13% for relative vorticity. At intermediate levels this means having errors of about 0.06 dyn cm, 0.5 cm/s and 0.007 f in regions where typical values for the true anomaly fields are 1 dyn cm, 6 cm/s and 0.05 f , respectively (Table 2).

Because this case is probably more representative of the majority of oceanographic cruises (both in terms of field variance and station distribution) than the SeaSoar sampling in the Alborán Sea, we also computed the fractional error for the case in which no scale selection is applied. Results indicate that when no scale selection is applied, fractional errors in the inner domain can be as high as 10% for dynamic height, 18% for geostrophic vorticity and 30% for geostrophic relative vorticity, in comparison to the 6%, 9% and 13% obtained when scale selection is applied. For the vertical velocity, errors are similar to those obtained for the geostrophic vorticity.

7. CONCLUSIONS

We have demonstrated methods for estimating the accuracy of dynamical field variables diagnosed from hydrographic survey data, when analysis accuracy is limited mainly by errors in the observations and by the sampling distribution (errors associated with non-synoptic sampling are evaluated in Part II of this work). The applied methodology assumes that actual fields exactly obey the simple correlation model used for the interpolation of observations, which is obviously not exactly the case in practice. Consequently, our analysis should be regarded as establishing realistic lower bounds for both, observational and sampling errors.

For parameters which are obtained by applying difference operators onto interpolated field data, results depend on both the difference scale and the spectral content of the interpolated data. The latter may be controlled by applying a suitable spatial filter to the observations, which has the effect of increasing the signal-to-noise content of the analysis while decreasing resolution. It is suggested that a reasonable compromise between resolution and accuracy can be achieved using a filter which removes most of the variance associated with wavelengths less than a few times the station separation measured along the direction of least dense sampling.

In relation to the important problem of diagnosing the mesoscale vorticity field or the associated QG vertical velocity, results suggest that the ‘best’ rms noise-to-signal

ratio that could be obtained from a high resolution (e.g., SeaSoar) sampling in a region of large spatial variability such as the Alborán Sea is of the order of 5%. Such a value might be achieved only within the interior of the survey area, which is least affected by errors associated with interpolation near the boundaries and uncertainty in the boundary conditions. For most cases, however (a typical CTD sampling in a region with spatial variability smaller than in the Alborán Sea), the ‘best’ rms noise-to-signal ratio that could be obtained (when scale selection is applied) is of the order of 10-15%. Such a level of accuracy should be adequate for determining the resolved part of a mesoscale circulation and associated fluxes, but might not be sufficient to test the validity of the QG approximation compared with other models such as a semi-geostrophic approximation (Pinot et al., 1996) or iterative models (Shearman et al., 2000). When no scale selection is applied, fractional error can be as high as 30%.

It has also been shown that for a high resolution sampling in a region of large spatial variability (e.g., when the noise-to-signal fraction is of the order of 0.002), sampling and observational errors can be of the same order. In most cases, however, the dominant contribution to total errors comes from sampling limitations. This means that for the recovery of mesoscale features, errors can only be reduced by increasing the density of the sampling. The latter, however, is usually achieved at the expenses of a loss of synopticity, so that the trade-off between these two error sources should be carefully evaluated. Other options would be using more than one ship and complementing in situ observations with the vertical projection of satellite data or with other in situ data. As an example of the latter, Gomis et al. (2001) applied a multivariate analysis merging dynamic height (from CTD) and velocity (ADCP) data. Although the latter are often of a limited accuracy, this is compensated by the increase in the number of observations entering the analysis.

ACKNOWLEDGMENTS

We thank Dr. Simón Ruiz for helping with the representation of results. We also want to thank Dr. Jean-Marie Beckers and Dr. Michel Rixen for their positive, in-depth reviews of the paper, which have surely contributed to improve its presentation. This work was inspired in the results of the EU OMEGA project (under contract MAS3-CT95-0001) and carried out in the

framework of the BIOMEGA project (REN2002-04044-C02-01/MAR) funded by the Spanish Marine Science and Technology Program.

REFERENCES

- Allen, J. T., D. A. Smeed, 1995: Potential vorticity and vertical velocity at the Iceland-Faeroes front. *J. Phys. Ocean.*, **26**, 2611-2634.
- Allen, J. T., D. A. Smeed, A. J. G. Nurser, J. W. Zhang, M. Rixen, 2001: Diagnosis of vertical velocities with the QG omega equation: an examination of the errors due to sampling strategy. *Deep-Sea Res.* **48**, 315-346.
- Bretherton, F. P., R. E. Davis, C. B. Fandry, 1976: A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep Sea Research I*, **23**, 559-582.
- Chereskin, T. K., M. Trunnell, 1996: Correlation scales, objective mapping and geostrophic flow in the California Current. *J. Geophys. Res.*, **101**, C10, 22619-22629.
- Davies, R. E., 1985: Objective mapping by least square fitting. *J. Geophys. Res.*, **90**, C3, 4773-4777.
- Fielding, S, N. Crisp, J. T. Allen, M. C. Hartman, B. Rabe, H. S. J. Roe, 2001: Mesoscale subduction at the Almería-Orán front. Part II: Biophysical interactions. *J. Mar. Sys.*, **30**, 287-304.
- Gandin, L. S., 1963: Objective analysis of meteorological fields. Transl. from russian by Israel Program for scientific translations, 1965. [NTIS No. TT65-50007]
- Gomis, D., S. Ruíz, M. A. Pedder, 2001: Diagnostic analysis of the 3D ageostrophic circulation from a Multivariate Spatial Interpolation of CTD and ADCP data. *Deep Sea Res. I*, **48**, 269-295.
- Holton, J. R., 1992: *An introduction to dynamic meteorology*. Academic Press, San Diego.
- Joyce, T. M., 1989: On in situ “calibration” of shipboard ADCPs. *J. Atmos. Ocean Tech.*, **6**, 169-172.
- Leach, H., 1987: The diagnosis of synoptic scale vertical motion in the seasonal thermocline. *Deep Sea Res. I*, **34**, 2005-2017.
- Le Traon, P. Y., 1990: A method for optimal analysis of fields with spatially variable mean. *J. Geophys. Res.*, **95**, C8, 13543-13547.

- Mc Williams, J. C., W. B. Owens, B. L. Hua, 1986: An objective analysis of the POLYMODE Local Dynamics experiment. Part I: General formalism and statistical model selection. *J. Phys. Ocean*, **16**, 453-504.
- Morrow, R., P. De Mey, 1995: Adjoint assimilation of altimetric, surface drifter, and hydrographic data in a quasi-geostrophic model of the Azores Current. *J. Geophys. Res.*, **100**, NO. C12, doi:10.1029/95JC02315.
- Nechaev, D. A., M. I. Yaremchuk, 1994: Conductivity-temperature-depth data assimilation into a three-dimensional quasigeostrophic open ocean model. *Dyn. Atmos. Ocean*, **21**, 137-165.
- Pedder, M. A., 1993: Interpolation and filtering of spatial observations using successive corrections and gaussian filters. *Mon. Wea. Rev.*, **121**, 2889-2902.
- Pedder, M. A., D. Gomis, 1998: Application of EOF Analysis to the spatial estimation of circulation features in the ocean sampled by high-resolution CTD samplings. *J. Atmos. Ocean Tech.*, **15**, 959-978.
- Pinot, J. M., J. Tintoré, D. P. Wang, 1996: A study of the omega equation for diagnosing vertical motions at ocean fronts. *J. Mar. Res.*, **54**, 239-259.
- Pollard, R. 1986. Frontal surveys with a towed profiling conductivity/temperature/depth measurements package (SeaSoar). *Nature*, **323**, 433-435.
- Pollard, R. T., L. Regier, 1992: Vorticity and vertical circulation at an ocean front. *J. Phys. Ocean.*, **22**, 609-625.
- Rodríguez, J., J. Tintoré, J. T. Allen, J. M. Blanco, D. Gomis, A. Reul, J. Ruiz, V. Rodríguez, F. Echevarría, F. Jiménez-Gómez, 2001: The role of mesoscale vertical motion in controlling the structure of phytoplankton in the ocean. *Nature*, **410**, 360-363.
- Rudnick, D. L., 1996: Intensive surveys of the Azores front, 2. Inferring the geostrophic and vertical velocity fields. *J. Geophys. Res.*, **101**, C7, 16291-16303.
- Shearman, K., J. A. Barth, J. S. Allen, R. L. Haney, 2000: Diagnosis of the three-dimensional circulation in mesoscale features with large Rossby number". *J. Phys. Ocean.*, **30**, 2687-2709.
- Thiébaux, H. J., M. A. Pedder, 1987: Spatial objective analysis with applications in atmospheric sciences. Academic Press, London.

- Tintoré, J., D. Gomis, S. Alonso, G. Parrilla, 1991: Mesoscale dynamics and vertical motions in the Alborán Sea. *J. Phys. Ocean.*, **21**, 811-823.
- Viúdez, A., J. Tintoré, R. L. Haney, 1996. Circulation in the Alborán Sea as determined by quasi-synoptic hydrographic observations. Part I: Three dimensional structure of the two anticyclonic gyres. *J. Phys. Ocean.*, **26**, 684-705.

FIGURE CAPTIONS

- Figure 1: map of the Omega-1 cruise in the Alborán Sea. Squares (both black and white) indicate actual stations (216) resulting from the SeaSoar sampling. Black squares indicate a subset of 73 stations that could be representative of a typical CTD sampling. Dots are grid points over which the fields were interpolated and the analysis errors were evaluated.
- Figure 2: standard deviation of analysis errors divided by the standard deviation of the anomaly field (“fractional error”) corresponding to an OI analysis of dynamic height observations at the whole set of stations of Fig. 1. a) Total errors; b) contribution of observational noise; c) contribution of the sampling. [The correlation model was a Gaussian function with characteristic scale $L=15$ km; the noise-to-signal parameter η was set to 0.002]
- Figure 3: as for Fig. 2, but for an OI analysis convoluted with a filter with cut-off wavelength $\lambda_o=40$ km.
- Figure 4: fields corresponding to one single ‘true’ realization (out of the set of 250) anomaly fields used to derive errors in the vertical velocity. a) Dynamic height (dyn cm); b) geostrophic velocity; c) geostrophic relative vorticity (10^{-5} s^{-1}); d) vertical velocity (10^{-3} m s^{-1}) obtained imposing $w=0$ at surface and at the reference level (300 m), and Neumann conditions at the lateral boundaries; e) as in d), but imposing $w=-0.3 \cdot 10^{-3} \text{ m/s}$ at the southern boundary and $w=0.3 \cdot 10^{-3} \text{ m/s}$ at the eastern boundary.
- Figure 5: fractional error distributions for vertical velocity (w). A constant trend (corresponding to zero geostrophic velocity) was added to each horizontal level of the dynamic height anomaly field prior to computing w . Boundary conditions were $w=0$ at surface and at the reference level (300 m) and $dw/dn=0$ at the lateral boundaries. a) Results with no scale selection applied; b) for an OI analysis convoluted with a filter with cut-off wavelength $\lambda_o=40$ km.
- Figure 6: fractional error distributions corresponding to the OI analysis of dynamic height observations for different station distributions and different values of the noise-to-signal parameter: a) for the subset of 73 stations shown in Fig.1 and $\eta=0.002$; b) for the whole set of stations of Fig. 1 but with $\eta=0.02$; c) for the subset of 73 stations and $\eta=0.02$. [For all cases the correlation model was a Gaussian with characteristic scale $L=15$ km convoluted with a filter with cut-off wavelength $\lambda_o=40$ km.]

TABLES

<i>Variable</i>	<i>Dynamic Height (dyn cm)</i>			<i>Geostr. Velocity (cm s⁻¹)</i>			<i>Geostr. Relat. Vorticity (10⁻⁵ s⁻¹)</i>			<i>Vertical Velocity (10⁻⁴ m s⁻¹)</i>		
<i>Level (m)</i>	<i>13</i>	<i>53</i>	<i>101</i>	<i>13</i>	<i>53</i>	<i>101</i>	<i>13</i>	<i>53</i>	<i>101</i>	<i>13</i>	<i>53</i>	<i>101</i>
<i>Actual anomaly std</i>	6.1	5.0	3.2	46.7	38.2	24.5	4.1	3.4	2.2	2.2	2.7	3.4
<i>Filtered anomaly std</i>	5.9	4.9	3.1	37.4	31.1	19.7	3.2	2.7	1.7	0.5	0.7	1.0
<i>Actual analysis rms errors</i>	0.12–0.36	0.10–0.30	0.06–0.18	2.8–4.7	2.3–3.8	1.5–2.4	0.37–0.49	0.31–0.41	0.20–0.26	0.13–0.44	0.16–0.54	0.20–0.68
<i>Filtered analysis rms errors</i>	0.06–0.15	0.05–0.12	0.03–0.08	0.7–3.7	0.6–3.1	0.4–2.0	0.10–0.45	0.07–0.38	0.05–0.24	0.02–0.12	0.03–0.17	0.04–0.27

Table 1: Rms errors at different levels and for the different variables. They all refer to the full station distribution of Fig. 1, but for two analysis options: with and without scale selection (cut-off wavelengths of 40 km). The two error values reflect the error range within the inner domain.

<i>Variable</i>	<i>DH (dyn cm)</i>	<i>GV (cm/s)</i>	<i>GRV (10⁻⁵ s⁻¹)</i>
<i>Alborán Sea ($\eta=0.002$)</i>			
<i>Filtered anomaly std</i>	3.1	19.7	1.7
<i>Full station distribution Rms</i>	0.03	0.4	0.05
<i>Regular station distribution Rms</i>	0.12	0.9	0.07
<i>Region with lower variability ($\eta=0.02$)</i>			
<i>Filtered anomaly std</i>	1.0	5.9	0.52
<i>Full station distribution Rms</i>	0.03	0.3	0.036
<i>Regular station distribution Rms</i>	0.06	0.5	0.068

Table 2: Rms errors at an intermediate level (101 m) for the different variables. All errors correspond to an analysis performed with scale selection (cut-off wavelength of 40 km), but for different station distributions and different values of the noise-to-signal ratio η . The quoted error value corresponds to the inner domain.

Figure 1. map of the Omega-1 cruise in the Alborán Sea. Squares (both black and white) indicate actual stations (216) resulting from the SeaSoar sampling. Black squares indicate a subset of 73 stations that could be representative of a typical CTD sampling. Dots are grid points over which the fields were interpolated and the analysis errors were evaluated.

Figure 4: fields corresponding to one single ‘true’ realization of the set of 250 used to derive errors in the vertical velocity field. a) Dynamic height (dyn cm); b) geostrophic velocity; c) geostrophic relative vorticity (10^{-5} s^{-1}); d) vertical velocity (10^{-3} m s^{-1}) obtained imposing $w=0$ at surface and at the reference level, and Neumann conditions at the lateral boundaries; e) as in d), but imposing $w=-0.3 \cdot 10^{-3} \text{ m/s}$ at the southern boundary and $w=0.3 \cdot 10^{-3} \text{ m/s}$ at the eastern boundary.

Figure 5: fractional error distributions for vertical velocity (w). A constant trend (corresponding to zero geostrophic velocity) was added to each horizontal level of the dynamic height anomaly field prior to computing w . Boundary conditions were $w=0$ at surface and at the reference level (300 m) and $dw/dn=0$ at the lateral boundaries. a) Results with no scale selection applied; b) for an OI analysis convoluted with a filter with cut-off wavelength $\lambda_0 = 40$ km

Figure 6: fractional error distributions corresponding to the OI analysis of dynamic height observations for different station distributions and different values of the noise-to-signal parameter: a) for the subset of 73 stations shown in Fig.1 and $\eta = 0.002$; b) for the whole set of stations of Fig. 1 but with $\eta = 0.02$; c) for the subset of 73 stations and $\eta = 0.02$. [For all cases the correlation model was a Gaussian with characteristic scale $L=15$ km convoluted with a filter with cut-off wavelength $\lambda_0=40$ km.]

d)