Contour Analysis: A New Approach for Melding Geophysical Fields

ARTHUR J. MARIANO*

Harvard University, Cambridge, Massachusetts

(Manuscript received 10 March 1989, in final form 22 September 1989)

ABSTRACT

This paper introduces a new approach, contour analysis, for combining independent estimates of a geophysical field to produce a single realization incorporating data from all sources. Contour analysis divides the field estimates into contours and then calculates the average position of each contour: it is shown that this method avoids the smearing and weakening of important dynamical features which can result from the traditional method of simply averaging data values at fixed geographical coordinates. The paper begins with an overview of how contour points are determined and averaged; this is followed by an illustrative example showing the basic steps one must go through to automate the process on a computer. The remainder of the paper presents comparisons between the traditional approach and the contour analysis approach over a wide range of data types: combining numerical model fields, averaging one-dimensional curves, space–time interpolation of Gulf Stream north wall positions and an application to data assimilation in the Gulf Stream ring and meander region.

1. Introduction

The melding (i.e., combining, blending or averaging) of different analyses of geophysical fields is an important component of a data analysis system. For example, suppose that a dynamical model produces a forecast of the streamfunction field at 100 m depth and a statistical analysis of in situ data produces another estimate of the 100 m streamfunction field. What is the "optimal" method for melding the two 100 m streamfunction field estimates and producing one 100 m streamfunction field estimate that is more accurate than both initial estimates? Let $\psi_1(x, y)$ and $\psi_2(x, y)$ denote the two fields, which have corresponding estimation variance fields $e_1^2(x, y)$ and $e_2^2(x, y)$, respectively. The classic "optimal" estimator (e.g., Gelb 1986; Ghil et al. 1981) for melding these fields is

$$\hat{\psi} = \frac{(e_2^2\psi_1 + e_1^2\psi_2)}{(e_1^2 + e_2^2)}. \tag{1.1}$$

If ψ_1 and ψ_2 are independent estimates, the variance field of (1.1) is

$$\hat{e}^2 = \frac{e_1^2 e_2^2}{(e_1^2 + e_2^2)}. \tag{1.2}$$

If ψ_1 and ψ_2 are unbiased estimates, $\hat{\psi}$ is also unbiased and has minimum variance with respect to all linear combinations of ψ_1 and ψ_2. The estimation error of our melded field, $\hat{e}$, is less than the estimation error of both initial fields, e_1 and e_2.

To understand the difficulties in this approach, consider the following example which illustrates some undesirable and well-known side effects of the classic estimator, $\hat{\psi}$ (1.1) Suppose, as it is shown in Figs. 1a and 1b, that each 100 m streamfunction field consists of a circular ring and a front. The frontal location differences between the two fields may be due to inadequate data or model resolution, incorrect model physics, or simplifying dynamical/statistical assumptions in the analyses. Assume that the estimation variance of each field is the same, $e_1^2 = e_2^2 =$ constant. Hence the melded field (Fig. 1c) is an arithmetic average of the two fields at each point in the domain. The melded front is weakened and the melded ring is no longer circular and is also of low amplitude (Fig. 1c). Estimate (1.1) smears the signatures of important dynamical features that are present in each initial analysis.

Investigators studying coherent but nonstationary features have found it necessary to devise alternate methods of melding, often by transforming the data into a *local* coordinate system which moves with the feature. Informal variants of the contour analysis approach have been used previously by several authors to preserve features in data that would be otherwise smeared by traditional approaches. For example, Halkin and Rossby (1985) showed that Gulf Stream velocity and temperature transects averaged in a "stream" coordinate system, defined by placing the or-

* Present Address: RSMAS/MPO, University of Miami.

Corresponding author address: Dr. Arthur J. Mariano, RSMAS/MPO, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149-1098.

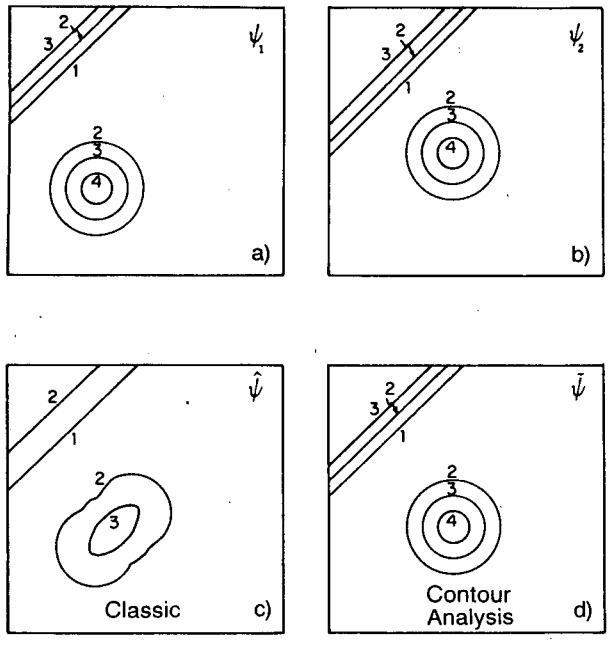

FIG. 1. (a) Streamfunction field one in nondimensional units. (b) Streamfunction field two. (c) The classic estimator's (1.1) streamfunction field from melding fields one and two. (d) Contour analysis' (2.1) streamfunction field from melding fields one and two.

igin of each transect at the location of maximum transport for that transect, produced a remarkably coherent Gulf Stream structure relative to Eulerian estimates smeared by spatial meandering. The Harvard "feature model" approach (Robinson et al. 1987) uses frontal locations determined from a variety of in situ and remotely-sensed data and analytical models of Gulf Stream rings and cross-stream structure for initializing numerical models (see section 4b for further details). Mariano and Rossby (1989) have shown that estimating the terms in the potential vorticity equation in a local space–time coordinate system defined by the centroid of SOFAR float clusters, i.e., a Lagrangian viewpoint, yields more accurate estimates than traditional Eulerian field estimates.

The contour analysis method presented in this paper was developed in order to obtain a formal procedure for determining the coordinate transformation that will eliminate the smearing associated with the classic "optimal" estimator (e.g., Fig. 1d shows the resulting melded field for the example above using the contour analysis estimator introduced in the next section). In contrast to the studies mentioned above which pick a point for a coordinate origin, contour analysis considers the whole field for the coordinate transformation.

As with any melding or averaging method, mathematical rigor does not guarantee physical significance and caution must be exercised when applying the results. Just as an Eulerian average of a meandering current would not be an appropriate basis for an instability model, there are bound to be situations in which the results of the contour analysis approach are not applicable. It is up to the researcher to devise appropriate tests to determine whether the observed quantities are being handled in a reasonable way. The examples presented below were chosen to indicate the variety of data sources to which the contour analysis method can be applied. The criteria for the success of the method is that the resulting field resembled that of the initial fields (in terms of spatial gradients and feature shape) to a greater degree than the result of the classic estimator.

This manuscript is organized as follows. First, in section 2, the contour analysis approach is described within the context of melding complete, similar two-dimensional fields. Similar fields are defined to be fields of the same type, i.e., the same variable on the same surface, and that the different field estimates to be combined are not radically different from each other. In section 3 the contour analysis approach is extended to the melding of one-dimensional curves, and in section 4 the approach is applied to the melding of spatially gappy and/or time variable fields. Examples utilizing the basic contour analysis approach for space–time interpolation of Gulf Stream north wall positions and data assimilation in the Gulf Stream system are presented. The advantages and disadvantages of the contour analysis approach are discussed in section 5 and a formal derivation proving that the contour analysis approach (2.1) does not smear simple fronts and circular eddies is given in the Appendix.

2. Contour analysis

The fundamental difference between classic estimators such as (1.1) and the contour analysis approach is that contour analysis averages the position of the field values, rather than the field values at each point. Thus, the first step is to apply a contouring program to each of the estimates of the geophysical field to convert each data field into a series of contours. A pattern recognition program must then be used to determine which contours from the individual field estimates should be paired together: the position data from the representations of the contour are then melded to find the average position of the contour.

Consider the melding of the two fields shown in Figs. 1a and 1b. Each contour from the first field is represented by an ordered set of x and y positions,

$$c_1(j) = \{x_{1k}(j), y_{1k}(j): k = 1, 2, \cdots, n_1(j)\},$$

where $c_1(j)$ is read as the jth contour of field 1 and $n_1(j)$ is the number of position data points determined by the resolution of the first field estimate and the contouring program. Each of the j contours from the first field must be combined with the corresponding contours from the second field,

$$c_2(j) = \{x_{2k}(j), y_{2k}(j): k = 1, 2, \cdots, n_2(j)\}.$$

The $n_i(j)$ points must be connected, so the values of $c_i(j)$ are not necessarily distinct. For example, for the fields in Fig. 1, $c_i(1) = 3$, $c_i(2) = 2$, $c_i(3) = 1$, $c_i(4) = 2$, $c_i(5) = 3$ and $c_i(6) = 4$ for $i = 1, 2$.

In general, $n_1(j) \neq n_2(j)$ for all j. Also, in order to find the average location of the jth contour it is necessary to find points on each contour which correspond: they must be at a similar position in the geophysical feature. The practical solution is to interpolate the contour position data so that the Kth point for both contours is located at the same fraction of the total respective arc length along each contour. The arc length S for each contour j is calculated by

$$S_i(j) = \sum_{k=2}^{n_i(j)} (\Delta x_{ik}^2 + \Delta y_{ik}^2)^{1/2};$$

$$\Delta x_{ik} = (x_{ik}(j) - x_{i(k-1)}(j)),$$

$$\Delta y_{ik} = (y_{ik}(j) - y_{i(k-1)}(j)),$$

where $i = 1, 2$ for the field estimates 1 and 2.

Corresponding contours $c_1(j)$ and $c_2(j)$ are both divided into $N(j)$ segments and the position of the segment nodes:

$$\{X_{iK}(j), Y_{iK}(j): K = 0, 1, \cdots, N(j)\},$$

are found by using cubic splines to interpolate both the latitude and the longitude to the same fraction of the arclength along both contours. For each pair of contours, four cubic spline interpolations with arclength as the independent variable are necessary: one cubic spline interpolation for each of the two latitude datasets and one cubic spline interpolation for each of the two longitude datasets. The choice of arclength as the independent variable permits melding of multivalued features.

The number $N(j)$, which determines the resolution $\Delta S_i(j) = S_i(j)/N(j)$ of the interpolated contours is picked empirically and depends on $n_1(j)$, $n_2(j)$, resolution requirements, and computational speed.

The melded contour field can then be calculated for m contour pairs by

$$\tilde{\psi}(X_K(j), Y_K(j)) = \tilde{\psi}\left(\sum_{i=1}^{2} w_{iK}(j)X_{iK}(j), \sum_{i=1}^{2} w_{iK}(j)Y_{iK}(j)\right),$$

$$j = 1, 2, \cdots, m \quad \text{and} \quad K = 0, 1, 2, \cdots, N(j). \tag{2.1}$$

Although the contour analysis approach has been described for melding two field estimates, the procedure is not limited: if more than two fields are to be melded, the index i used above would simply range from 1 to the total number of estimates.

The weights, $w_{iK}(j)$, are picked for the specific problem to be solved and their sum must equal one to prevent biases. *Optimal* weights analogous to (1.2) for melding two fields are

$$w_{1K}(j) = \frac{e_2^2(X_{2K}(j), Y_{2K}(j))}{(e_1^2 + e_2^2)},$$

$$w_{2K}(j) = \frac{e_2^1(X_{1K}(j), Y_{1K}(j))}{(e_1^2 + e_2^2)}. \tag{2.2}$$

The estimation variance field of (2.1) with weights given by (2.2) is given by (1.2) evaluated at the position of the melded contours. Figure 1d is a plot of the analytic solution for the first example using equal weights, $w_{1K}(j) = w_{2K}(j) = 0.5$. Other weights are given in the examples presented below and the estimation variance field associated with these weights are most easily calculated at the melded segment nodes using the estimation variance of the segment nodes of the initial contours.

Under certain circumstances, preprocessing steps will be necessary if there is not a one-to-one correspondence between contours of the initial fields. For instance, the initial fields might differ in magnitude. If the streamfunction field shown in Fig. 1b has a range of streamfunction values from 0 to 3 instead of 1 to 4, we would remove a mean from each field before melding, apply (2.1) to the deviations from the mean and add a weighted average of the means to the melded field of deviations. In the more general case, the initial fields would be detrended with a spatially dependent surface, the contour analysis approach would be used for melding the deviations from that surface, and a weighted average of the trend surfaces would be added to the melded deviations.

If there is still not a one-to-one correspondence between contours after detrending, other preprocessing steps should be tried. These might include: (i) changing the number of contours by changing the contour interval, (ii) normalizing by the range of contours for a given field after removing a trend, (iii) extrapolating one or both of the fields, and (iv) discarding a short contour(s) that can not be matched. If the minimum and maximum contours still do not match between the fields, one possible approach would be to average the maximum (or minimum) ψ values together and assign this average value to the position of the maximum (or minimum) melded contour. If these methods all fail, either the fields are dissimilar and should probably not be combined or a set of special selection rules for matching contours must be defined (examples are given in section 5).

a. Contour analysis of numerical fields

This example highlights important computational considerations for automating contour analysis. The steps in this automated contour analysis algorithm consist of: (i) running each initial field through a contouring program and storing the contour values and their positions, and the value of an incremental counter; (ii) sorting these first by contour value and then by incremental counter value; (iii) matching up similar

contours of identical magnitude using a pattern recognition algorithm (this step is necessary because of equivalued contours); (iv) dividing each contour into N segments using a one-dimensional (1-D) cubic spline for interpolating both the latitude versus arc length and longitude versus arc length series; (v) averaging the segment nodes one contour pair at a time; (vi) gridding the melded field with two-dimensional (2-D) bi-cubic splines; and (vii) (optional step) applying a Shapiro filter (Shapiro 1970) to remove noise with a period of two grid points.

Figures 2a and 2b show two realizations of Harvard Quasi-geostrophic Open Ocean Model (Robinson and Walstad 1987) output streamfunction fields (31 × 31) at 100 m depth that differ due to differing initial conditions. The melded field which results from using the classic estimator (1.1) yields a very smeared representation of reality (Fig. 2c). It would be difficult to claim that the field shown in Fig. 2c is an "optimal" estimate of the 100 m streamfunction field in the Gulf Stream. Contour analysis (Fig. 2d) produces an output more similar to the data from which it was calculated; the Gulf Stream maintains a uniform cross-stream structure and the rings are not smeared in strength. Also, the ring positions in Fig. 2d are the averaged positions of the ring positions in 2a and 2b and the shape of the Gulf stream front is the average of the frontal shapes shown in Figs. 2a and 2b. Generalization of this example to melding numerical model output from various initial conditions implies that contour analysis is ideal for averaging the results of Monte Carlo simulations.

The following pattern recognition algorithm was chosen for its simplicity and other algorithms should

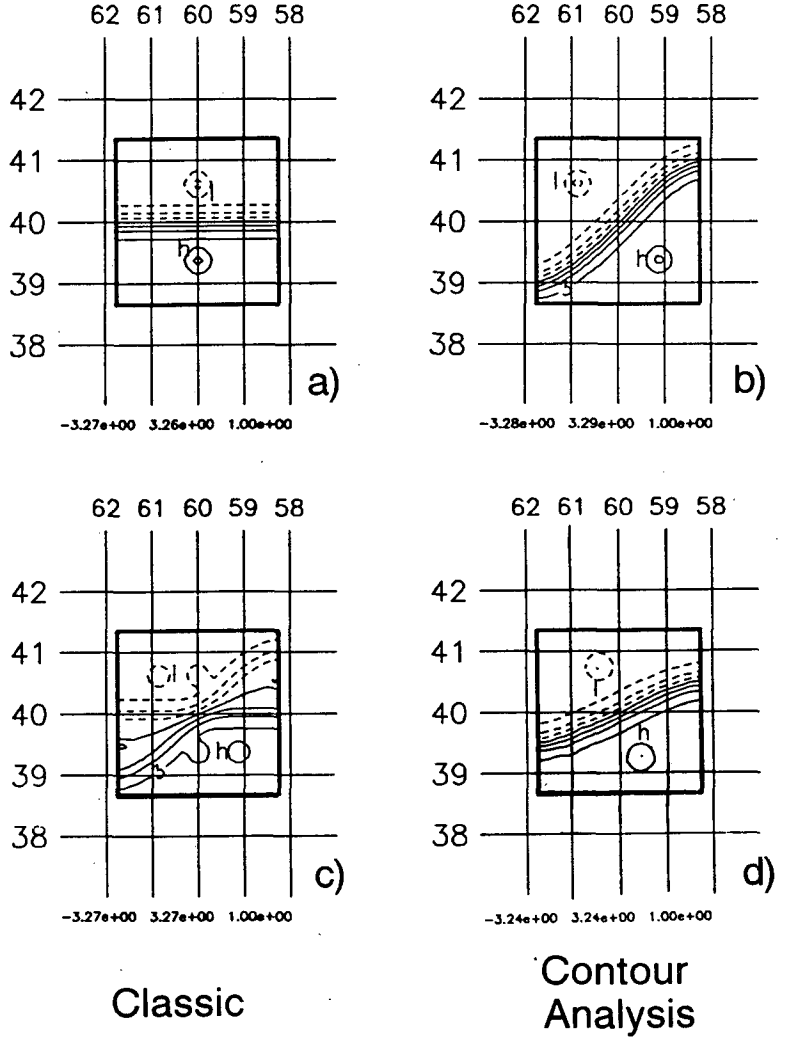

FIG. 2. (a) Harvard open ocean quasigeostrophic model output streamfunction field one. The three numbers below each plot are minimum streamfunction value, maximum streamfunction value and contour interval. (b) Streamfunction field two. (c) The melded field using the classic estimator (1.1). (d) The melded field using contour analysis (2.1).

be considered for different contour analysis applications. Pattern recognition was done by grouping contours by their contour value, fitting a polar coordinate system about (x_{ijo}, y_{ijo}), the center of mass of each contour j in the group from the ith estimate and calculating the difference between the smallest ($\theta_{ij}^{\min}$) and largest ($\theta_{ij}^{\max}$) angle for each contour and sorting SM, the similarity metric,

$$\mathrm{SM} = d_1*(\theta_{ij}^{\max} - \theta_{ij}^{\min}) + d_2*(x_{ijo} + y_{ijo}),$$

where d_1 and d_2 are the weighting for the shape factor and the position factor, respectively and were determined empirically (for this example, $d_1 = 1.0$ and $d_2 = 0.02$). Similar contours are matched by their relative placing in this sorting.

To reinterpolate the melded contour positions onto a regular grid, finite-element least-square bi-cubic splines with variable smoothness (ρ) and tension (τ) parameters are used. These splines use binary subdivision and the B-Spline basis for computational speed (Inoue 1986). Analogous one-dimensional cubic splines are used for segmenting the contour position data. The tension parameter can vary from no tension ($\tau = 0$) to maximum tension ($\tau = 1$). The splines converge to a mean value as the smoothness parameter approaches zero ($0 < \rho < 10^{-4}$) and the splines solve the fitting problem for large values of ρ ($>10^3$). Each data point, in general, is assigned a different error value for weighted least-square fitting of (bi-)cubic splines in each finite element (i.e., between the knots). The contour interval (0.04 for this example) and the 1-D and 2-D spline parameters (number of finite elements, tension, smoothness and data variance) are important computational parameters and were determined empirically by a trial and error search thru parameter space. The splines should be tense with large values of ρ and a large number of finite elements (see Table 1a which lists the spline parameters used to generate Fig. 2d). In other words, the splines should be chosen so that they solve the fitting problem within data error with no over-shooting.

Due to the sparsity of contours near the boundary, the 2-D bi-cubic splines must extrapolate over data-void regions. Two approaches have been used successfully by the author to deal with this problem. The first uses 2-D bi-cubic splines with unit tension (i.e., maximum tension) to keep the melded field from having spurious oscillations near the sparse boundary.

Another approach is to treat the boundary as a "variable contour." In general, the boundary values have the largest error, so a simple smoothing technique such as the following should be used. For each field, the boundary is segmented into segments consisting of typically ten points. The ten ψ values in each segment are averaged and the mean value, $\bar{\psi}$, is assigned to the midpoint of the segment. Then in the spirit of the contour analysis philosophy, corresponding $\bar{\psi}$ values have their locations averaged to obtain the melded boundary. Identification of corresponding $\bar{\psi}$ values is more complex than the determination of similar contours in the interior since the $\bar{\psi}$ values of each field will usually not be equal in magnitude. Computationally, similar values are defined by the $\bar{\psi}$ values being of approximately equal magnitude. The set of $\bar{\psi}$ for each field are ranked and are paired together by their rank. Melding consists of averaging the $\bar{\psi}$ values together as well as the segment midpoints. An ad hoc test to guarantee that the $\bar{\psi}$ values paired together actually corresponds is that the melded position must be on the boundary (i.e., $\bar{\psi}$ values must be from the same boundary). The author believes that this is a reasonable approach for datasets which consist only of boundary data.

TABLE 1. Spline parameters.

ρ	τ	Std. dev.	Number of finite elements
(a) Spline parameters for field melding			
One-dimensional spline parameters			
3000.	1.0	.01	5.0 × number of data pts.
Two-dimensional spline parameters			
2000.	1.0	.05	60, 60
(b) Range of one-dimensional parameters for GSNWP			
2000–5000	.03–.9	.03–.1	.2–1.2 × number of data pts.
(c) One-dimensional parameters for B–V frequency			
5000	.9	0.1	9 or 10 × number of data pts.

3. Contour analysis of one-dimensional curves

The next example shows how the contour analysis approach can be used for combining different estimates of a curve. Shown in Figs. 3a and 3b are two Brunt–Väisälä (B–V) frequency versus depth curves. The important differences between the curves are the strength and vertical location of the local maxima. Figure 3c shows the smearing that results from the usual oceanographic practice of averaging the B–V frequencies at each depth. Following the approach for boundary points, these curves can be treated as variable contours which are a function of depth. In field examples discussed in section 2, contours are a function of two spatial dimensions $\psi = \psi(x, y)$ and are connected, now the variable contour is a function of one spatial dimension $\psi = \psi(z)$ and the positions of a certain ψ value form a disjoint set. Ideally, the approach should be to average the coordinate of each ψ value following the guidelines above. For example, the coordinate of the first occurrence of $\psi = \psi_o$ from both curves should be averaged, the coordinate of the second occurrence of $\psi = \psi_o$ from both curves should be averaged and so on for all occurrences of $\psi = \psi_o$. This averaging is done for all ψ values. There are two problems with this ideal approach. The first is that there is probably not a one-

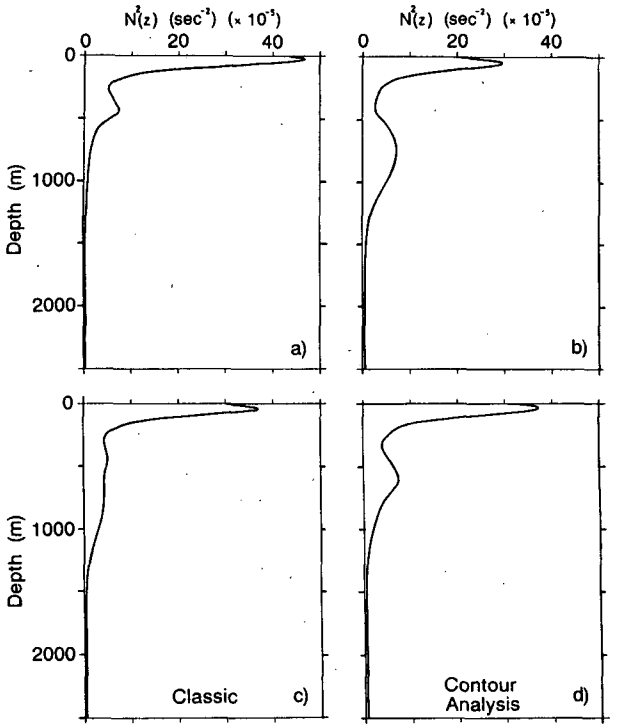

FIG. 3. (a) Brunt–Väisälä (B–V) frequency versus depth curve from averaged slope water data. (b) B–V frequency versus depth curve from averaged Sargasso Sea data. (c) The results of averaging (a) and (b) together using the standard approach. (d) The results of averaging (a) and (b) together using contour analysis.

to-one correspondence between the ψ values from each curve. Second, this approach is computationally cumbersome. The first problem could be remedied with the proper coordinate dependent mean and/or normalization as suggested in the last section. The second problem and experience with the variable boundary contour motivated the following approach for curves. Each curve is broken into segments; with break points located at local extrema in the curve and divided into $N + 1$ points. For the z values, a linear interpolation is sufficient while for the ψ values, a one-dimensional spline interpolation is necessary. These are averaged as before, but now since the contours are variable, the contour value must be averaged too. The resulting curve may be smoothed with a one-dimensional spline.

For the example shown in Fig. 3, the B–V curves were broken into four segments; the first data point to the first local maximum, the first data point after the first local maximum to the local minimum, the first data point after the local minimum to the second local maximum and the first data point after the second local maximum to the last point. For each segment, $N + 1$ evenly spaced depth values as well as B–V frequency values are averaged. The resulting contour analysis curve is shown in Fig. 3d. Again, the contour analysis approach produces a profile whose shape is more similar to the initial analyses than the standard averaging procedure. Moreover, local maxima and vertical gradients, which are of extreme dynamical importance for determining necessary conditions for stability problems and for providing the vertical restoring force, are not smeared.

4. Contour analysis of incomplete data and time-varying fields

In this section we extend the basic contour analysis algorithm for gappy and time-varying data. Two examples are given which use time series of advanced very high resolution radiometers (AVHRR) remotely sensed sea surface temperature (SST) data, which are gappy due to cloud cover. The first example involves space–time interpolation of Gulf Stream North Wall positions (GSNWP). The second example uses frontal locations determined from SST for data assimilation in the Harvard model (Robinson et al. 1987).

a. Space–time interpolation of Gulf Stream north wall positions

The processing of the GSNWP data is described by Cornillon (1985). The initial processing is done by the remote sensing group at the Rosenstiel School of Marine and Atmospheric Science, University of Miami. The data are mapped to a common coordinate system and composited into two-day groups at the University of Rhode Island (URI). The compositing consists of keeping the warmest pixel out of approximately ten satellite passes at each point. The northern edge of the Gulf Stream was hand digitized by Craig Gilman from URI. Due to clouds, there are many spatial and temporal gaps in this dataset; a subsample of this large dataset is shown in Figs. 4 and 5. The goal is to fill in these space–time gaps.

The space–time interpolation of the GSNWP is not a trivial interpolation problem since longitude versus latitude is often multivalued. Contour analysis deals with the multivaluedness by forming two single-valued space series for a given composite; latitude versus arc length and longitude versus arc length. The process of filling a spatial gap consists of first searching into the past (before the interpolation time) and then into the future (after the interpolation time) to find the nearest time where data exists in the region of the data gap. Figure 4 shows an example of this procedure. To interpolate the gappy region in Fig. 4b between 66° and 58°W, data from the same longitudinal range from Figs. 4a and 4c are used. Latitudes from the past and future space series are averaged together, as well as the longitudes. For this example, the weights are inversely proportional to the amount of time between the present gap and the time of the past and future data and are normalized such that their sum is one. Thus, the four day old past data from Fig. 4a is averaged with a weight of 1/3 and the two-day future data from Fig. 4c is averaged with a weight of 2/3.

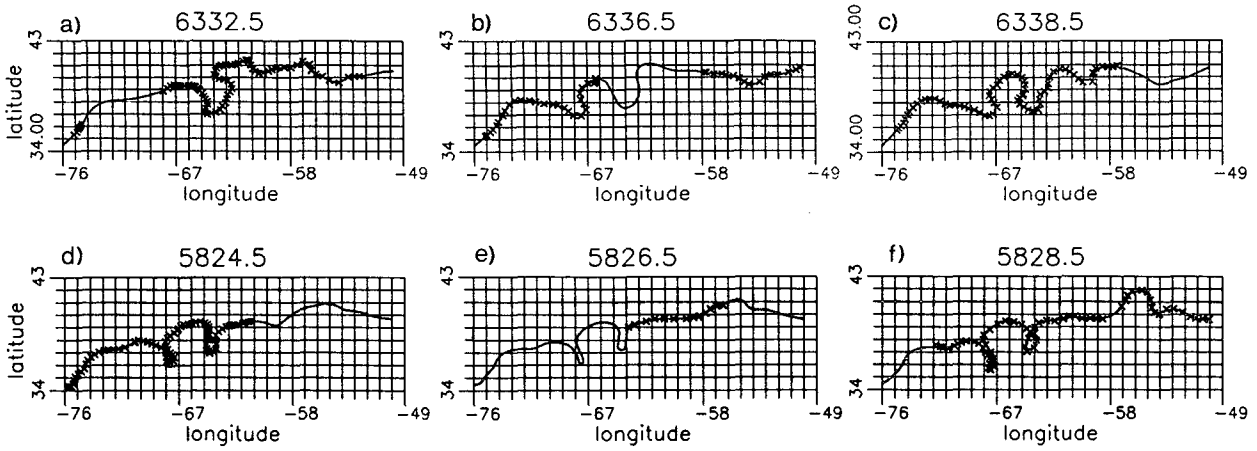

FIG. 4. The Xs are Gulf Stream north wall positions (GSNWP) from two day composites of satellite IR data digitized by Gilman (URI). The contour analysis fit is given by the solid line. Truncated Julian day is atop of the longitude/latitude grid. (a)–(c) Data from two day composites 6332.5 and 6338.5 are used for interpolating the data gap on day 6336.5. (d)–(f) Another example of interpolating multivalued features using contour analysis.

The basic contour analysis algorithm is used except for one major modification: contours must be propagated to a common analysis time. A phase speed is simply calculated by averaging the propagation of the local maximum and local minimum between the past and future time and assuming (for computational ease) the east–west propagation speed is dominant. The latitudes are propagated forward for past data and backward for future data using this local (in space and time) phase speed. The averaged and propagated latitude and longitude values are paired together by marching up the arc lengths.

Using this procedure, 824 continuous two-day composites of GSNWP were created for the period April 1982 to December 1986. Mariano (1988) presents a complete set of these interpolations as well as important computational details. Table 1b lists the range of parameters used for the 1-D cubic spline interpolations of latitude versus arc length and longitude versus arc length. As expected, one set of parameters did not consistently produce the best interpolation due to the rather large variations in data density and stream configurations from composite to composite.

It is evident in Fig. 4 and in the 824 interpolated GSNWP that the contour analysis approach is well suited for space–time interpolation of multivalued features (Mariano 1988). Careful inspection of the 824 GSNWP by the author reveals that the contour analysis approach produced realistic interpolations 99% of the time. The eight apparently poor interpolations (an example is shown in Fig. 5a) generally occurred at times immediately following ring births or strong ring–stream interactions or when the sparsity of data resulted in a poor estimate of the phase speed or a large north–south phase speed. Future algorithms could use both a two-dimensional phase speed and a better method for calculating phase speed. For instance, pieces of the front could be propagated in time using higher order dynamics such as contour dynamics (e.g. Pratt 1985) or matching up regions of similar curvature instead of simple phase estimates. The simple phase speed calculated is however sufficient for most of the interpolations.

Since the contour analysis approach is a purely statistical technique, i.e. no dynamics, it can not interpolate over temporal gaps longer than the phenomenological time scale. The time series of interpolated GSNWP are spatially smoother during data sparse periods (Mariano 1988). (This deficiency in contour analysis can be remedied if contour analysis output is used as input to a dynamical model.) The following evaluation was performed to quantify the amount of smoothing. Twenty-five 2-day composites were picked at random and all the data was removed from these 25 two-day composites. The contour analysis approach outlined above was used to interpolate the data void two-day composite for those 25 times. Figures 5b–d illustrate the three worst interpolations with the data superimposed. Most of the interpolated curve is within 10 km, the approximate data error, of the original digitized GSNWP data. The largest errors (50–80 km) are found in the meander amplitudes in the vicinity of 58°W in each curve. The apparent negative bias in amplitude results from a combination of phase speed error and deficient input data, i.e., the large amplitude meanders are not adequately represented by the data surrounding these estimation times. The interpolation error is dependent on both the distribution of the input data and the nature of the synoptic dynamical events. The interpolation error is less than 10 km for a straight and noninteracting stream configuration and good data coverage and is on the order of 100 km for a strong ring–stream interaction and poor data coverage.

The procedure for GSNWP can be extended for in-

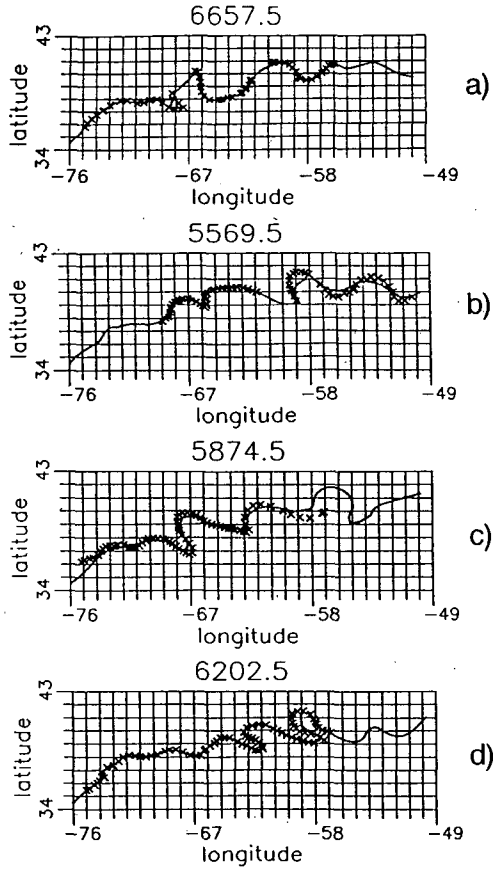

FIG. 5. As in (4) but showing poor interpolations. (a) Contour analysis produces an unrealistic interpolation of a meander which is about to form a warm core ring. Due to clouds, the closest input data are from 6 days before and 8 days after the interpolation time and do not contain a signature of the large meander. (b)–(d) Three of the worst (in the humble opinion of the author) interpolations from the simulation described in the text. None of the input data shown here was used for the interpolation. These poor interpolations are probably due to the combination of inadequate input data and the assumption of a one-dimensional east–west phase speed.

terpolation of time series of gappy fields. Similar contours are tracked throughout the time series. An interactive graphics package would make this tracking practical. Each set of similar contours are interpolated using the same space–time interpolation procedure that was used for GSNWP. After all the similar contour sets are interpolated, each resulting field is regularly gridded and smoothed using 2-D bi-cubic splines as described above for the model output fields.

b. A data assimilation example

Weekly forecasts of the Gulf Stream ring and meander region (see Fig. 6) using the multilevel Harvard open-ocean quasi-geostrophic (QG) model have been routinely calculated during the Gulfcasting program (Robinson et al. 1987). The available input data consists of NOAA IR maps, in situ (A)XBTs, floats and/or satellite altimetric data and are not sufficient for direct calculation of the streamfunction fields at each model level needed for initializing the QG model.

These data are used in an initialization procedure based on a feature model approach: analytical models of warm core rings, cold core rings, and the Gulf Stream front are used to represent the full three-dimensional streamfunction fields given only surface frontal locations. These primary fields are put into a motionless background ocean and thus the secondary or background field generated by the previous "Gulfcast" is not used. As a result of this procedure, information about the fields outside the domain of the initialization features is lost from one Gulfcast prediction to the next. Thus, the QG model must spend part of its integration time in spinning up the initial field to a QG state. To solve this problem, a generalization of the contour analysis procedure was developed with M. A. Spall (now at NCAR) and A. R. Robinson (Harvard) to incorporate both the model generated background field from the previous forecast and new frontal location information.

For each field, a curvilinear coordinate system (τ, η) is defined by each frontal feature. For the Gulf Stream, the τ coordinate is defined by the geographical location of the Gulf Stream axis or north wall: the τ coordinate is given by the arc length along the current. The η coordinate is perpendicular to the τ coordinate at every point: it measures the distance between the background field and the stream axis (Fig. 6b). A polar coordinate system is used for each ring, where τ, η are the azimuthal and radial coordinates, respectively.

Using the notation of section 2, let the model generated field be the first field and the new satellite frontal segment information be the second field. Frontal information from field two, e.g., the center position and radius of each ring and the location of the Gulf Stream, is assimilated into field one. Assimilation consists of moving the frontal features from field one to the new locations of the frontal features from field two by averaging the τ coordinate using (2.1) with two sets of weights. If new information is available, $w_{1K}(j)$ is zero and $w_{2K}(j)$ is one. If there is no new information or the front did not move, $w_{1K}(j)$ is one and $w_{2K}(j)$ is zero. In essence, the points where $\eta = 0$ in field one are mapped to a new position determined by the position of the points where $\eta = 0$ in field two. The rest of the points, $\eta \neq 0$, are mapped from field one to field two by the mapping determined by the $\eta = 0$ points. In other words, the τ coordinate axis for each frontal segment is mapped from field one to field two and the points perpendicular to each τ coordinate axis are mapped by exactly the same transformation.

Figure 6c shows a first attempt to use this approach in the Gulf Stream ring and meander region. The major differences between the previous Gulfcast (Fig. 6a) and the reinitialization (Fig. 6b) are the strength of the ring–stream interaction at 69° and at 57°W, the shape of the deep socklike meander between 64° and 60°W, the stream at 53°W is straighter in Fig. 6b and a small

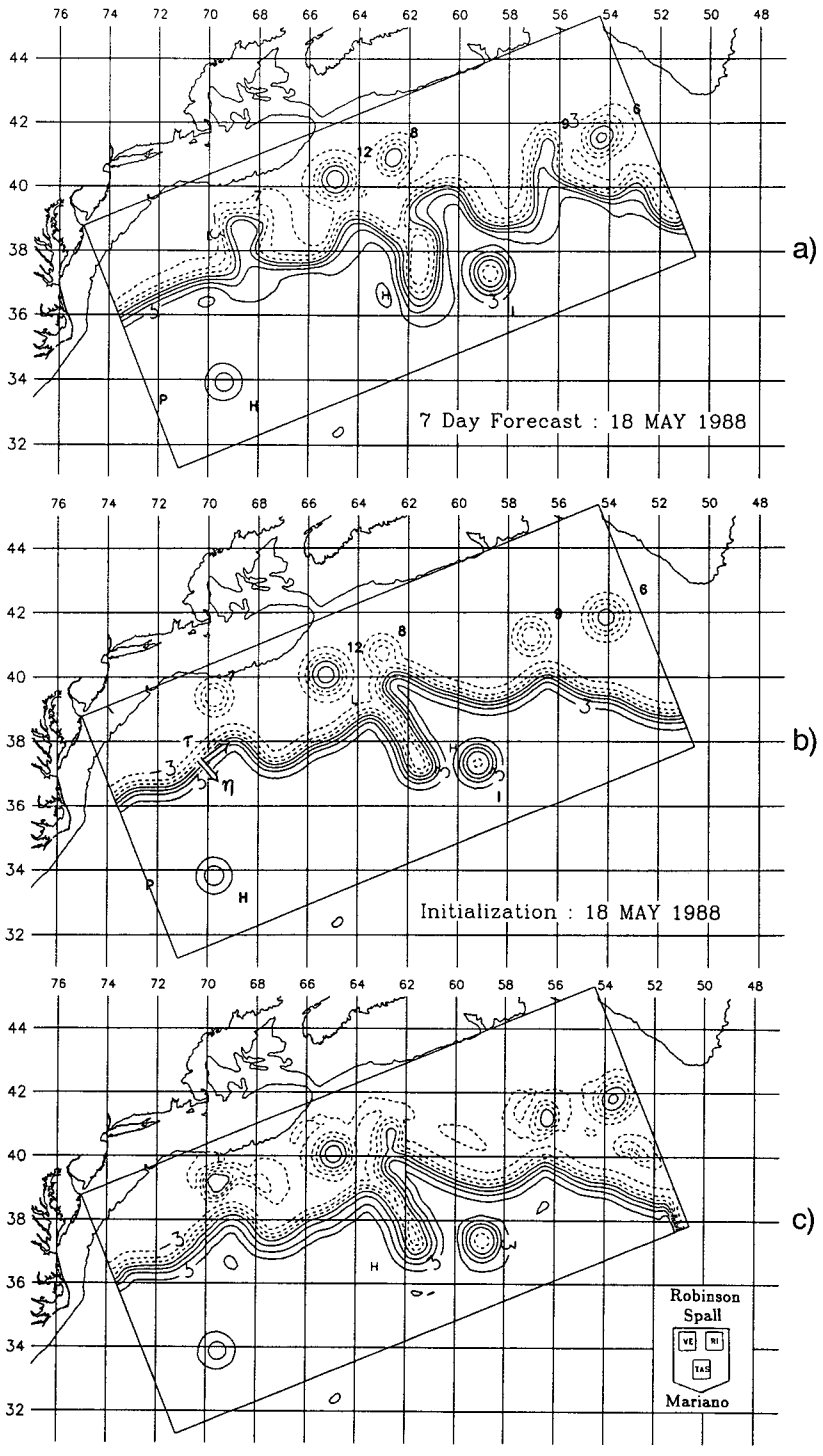

FIG. 6. (a) A seven day forecast of 100 m streamfunction from GULFCAST (Robinson et al. 1987). (b) The 100 m feature model streamfunction field for initializing next week's forecast. The (τ, η) curvilinear coordinate system described in the text is superimposed. (c) The new 100 m streamfunction for continuing next week's forecast. Contour analysis is used to extract the best new information available in (b) and the background fields of (a).

meander has propagated into the region near 73°W. Also, the rings in Fig. 6c result from moving the forecasted rings (Fig. 6a); hence, the shape of the rings are more realistic than the rings in the reinitialization (Fig. 6b) and the rings are closer to a quasi-geostrophic equilibrium state. Through the use of contour analysis,

the strongest data signal, e.g., the frontal locations consisting of stream axis or north wall location and ring size and location, is assimilated and model generated background fields are used for reinitialization.

5. Discussion

Why does contour analysis yield more physically realistic fields than the fields produced using the classic *optimal* estimators? To answer this question, one must first recall the basic premise of estimators. Properties of estimators are based on theoretical expectation operators working on an infinite ensemble of realizations from a population and are optimal only in an average sense. Thus, an individual estimate calculated from equation (1.1) using just two realizations from an infinite population of fields will not be unbiased with minimum variance every time, but only in an average sense. If our initial estimates, ψ_1 and ψ_2, are biased, even averaging an infinite ensemble of realizations will not remove the biases.

The error in geophysical analyses are usually caused by: 1) an unknown barotropic component, 2) inaccurate phase speeds, 3) wrong e-folding scales and zero-crossings of correlation functions, 4) biased instruments and/or algorithms, 5) navigation error in in situ and remotely sensed data. These errors are large scale and exhibit spatially coherent biases. Also, small-scale random errors are usually removed by dynamical/statistical filters, leaving mostly bias error. These bias errors affect the position of contours by moving the contours coherently in space. Contour analysis prevents smearing by operating in the coordinate system most affected by large-scale spatially coherent biases; a curvilinear coordinate system given by the contour coordinates.

Reliable implementation of the contour analysis approach is constrained by pattern recognition and matching limitations. Pattern recognition is used for matching similar contours (e.g., the 2-D field example), identifying features for phase speed calculations (e.g., the GSNWP example), selecting local significant extremum (e.g., the 1-D curve example) and for selecting and inserting data patches when melding gappy data (e.g., the GSNWP example). The selection rules for matching contours can be enhanced by a priori knowledge of the dynamical fields and the data. For instance, Fig. 7 illustrates two cases in the Gulf Stream ring and meander region that a set of selection rules must encompass. The first (Fig. 7a) case results from contours changing their topology after a ring birth. The algorithm needs to match up two contours from field 1 and one contour from field 2 such that the circular part of each field are matched and the straight front segments are matched. The second case (Fig. 7b) is due to data differences, such as rings losing their thermal signatures but not their altimetric signal. For this case, the present algorithm melds one ring (WCRA1 and WCRA2) from each field together and does not move the other ring (CCR2). The extent of practical applications of the contour analysis approach will ultimately depend on advances in pattern recognition algorithms.

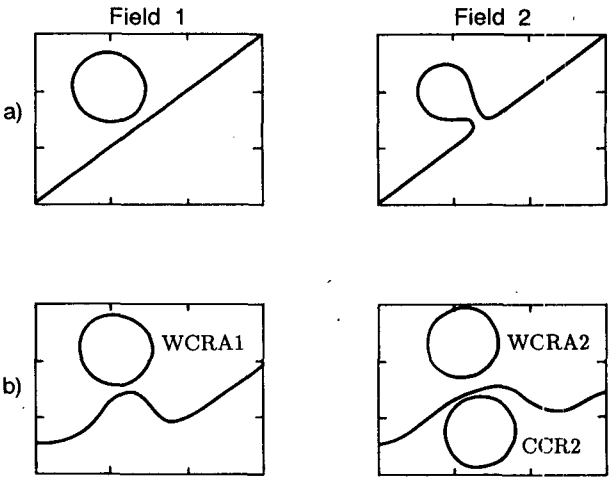

FIG. 7. (a) Two different estimates of a fixed contour value in the Gulf Stream ring and meander region. (b) Two different estimates of feature location in the Gulf Stream ring and meander region.

This study does not prove that an "optimal" analysis system consisting of an observational network and realistic dynamical/statistical models with full data assimilation capabilities based on a "feature" coordinate system is the best approach. The purpose of this study has been to show through several practical examples that the contour analysis approach is worth pursuing.

Acknowledgments. This study was generously supported by grants from the Office of Naval Research (N00014-84-C-0461) and the Institute for Naval Oceanography (S8766) and by the University Research Initiative Program through the office of Naval Research under Contract N00014-86-K-0751. Final preparation of this manuscript was supported by the Rosenstiel Fellowship and used the computer facilities of the University of Miami Remote Sensing Group. Contour analysis was motivated by Allan R. Robinson and benefited by discussions with Carlos Lozano, Michael Spall, Leonard Walstad, Ralph Milliff, Bill Johns, Jerry Miller, and Don Olson and by having good reviewers. Special thanks to Michael Spall for the software used in section 4b and his direct input of ideas used in sections 4a and 4b and to Dennis McGillicuddy for his help with section 3 and Fig. 3. Daniel Halkin, my fellow deaddy researcher, deserves special praise for unsmearing the original manuscript and melding it into its present shape.

APPENDIX

First we prove that (2.1) does not smear circular rings. Suppose that for a contour of ring one, there exists points, $Z_1 = X_1 + iY_1$, on a circle of radius R_1 and centered at C_1, viz.

$$|Z_1 - C_1| = R_1,$$

where $|A|$ is the magnitude of A. Similarly for the same contour of ring two, there exists points Z_2 such that

$$|Z_2 - C_2| = R_2.$$

The estimated ring E should be given by

$$|Z_E - C_E| = R_E, \tag{A.1a}$$

where

$$C_E = \frac{(e_2^2 C_1 + e_1^2 C_2)}{(e_1^2 + e_2^2)}, \tag{A.1b}$$

$$R_E = \frac{(e_2^2 R_1 + e_1^2 R_2)}{(e_1^2 + e_2^2)}. \tag{A.1c}$$

Equations (A.1b) and (A.1c) are the conditions for nonsmearing, the center position of the estimated ring is a weighted average of the positions of rings one and two and the strength of the ring is a weighted average of the strength of rings one and two. Strength has been defined to be the distant from the center to our chosen contour, the closer the contour, the stronger the ring.

Z_E is given by estimator (2.1) and weights (2.2), viz.

$$Z_E = \frac{(e_2^2 Z_1 + e_1^2 Z_2)}{(e_1^2 + e_2^2)}. \tag{A.2}$$

By substituting (A.2), (A.1b) and (A.1c) into (A.1a), it must be shown that

$$\left|\frac{(e_2^2 Z_1 + e_1^2 Z_2)}{(e_1^2 + e_2^2)} - \frac{(e_2^2 C_1 + e_1^2 C_2)}{(e_1^2 + e_2^2)}\right| = \frac{(e_2^2 R_1 + e_1^2 R_2)}{(e_1^2 + e_2^2)}, \tag{A.3}$$

is true for some values of Z_1 and Z_2. Rewriting (A.3) as

$$|e_2^2(Z_1 - C_1) + e_1^2(Z_2 - C_2)| = e_2^2|Z_1 - C_1| + e_1^2|Z_2 - C_2|,$$

shows that for the equality to hold, argument $(Z_1 - C_1)$ *must equal* argument $(Z_2 - C_2)$. In other words, the points Z_1 and Z_2 must start at the same place on the circle and sweep out equal angles, which is exactly the contour analysis procedure. To complete the proof, apply this proof for one contour to each contour of the circular rings.

The proof for fronts mimics the proof for rings. We show that for two straight lines, one from each analysis, contour analysis must be used to prevent smearing. Apply the argument presented next to each straight line for the full proof. Straight line from front one is written as

$$Z_1 = Z_{a1} + \beta_1(Z_{b1} - Z_{a1}), \quad 0 \leq \beta_1 < 1,$$

and straight line from front two as

$$Z_2 = Z_{a2} + \beta_2(Z_{b2} - Z_{a2}), \quad 0 \leq \beta_2 \leq 1;$$

Z_{ai} is the first point of line i, Z_{bi} is the last point of line i and Z_i are the points on line i. As β_i varies, the Z_i values between Z_{ai} and Z_{bi} are generated. Letting $\beta_1 = \beta_2 = \beta$ and averaging all the Z_1 and Z_2 points with the optimal weights (2.2) yields

$$e_2^2 Z_1 + e_1^2 Z_2 = e_2^2 Z_{a1} + e_1^2 Z_{a2} + \beta\{(e_2^2 Z_{b1} + e_1^2 Z_{b2}) - (e_2^2 Z_{a1} + e_1^2 Z_{a2})\}. \tag{A.4}$$

Equation (A.4) is a straight line whose endpoints are a weighted average of the original endpoints. It is the assumption of $\beta_1 = \beta_2$ that makes this possible. This assumption is equivalent to normalizing each line by its arc length and marching along at equal increments and averaging. So the position of the melded contour is the required result, a weighted average of the original contour positions. This will be true for all the contours in the front, hence the position of the melded front will be a weighted average of the original frontal positions.

REFERENCES

Cornillon, P., 1985: Gulf stream envelope and mean path between 75°W and 58°W. *Gulf Stream Workshop Proceedings,* D. R. Watts, Ed., II.182–II.194.

Gelb, A. (Editor), 1986: *Applied Optimal Estimation,* M.I.T. Press, 374 pp.

Halkin, D., and H. T. Rossby, 1985: The structure and transport of the Gulf Stream at 73°W. *J. Phys. Oceanogr.,* **15,** 1439–1452.

Inoue, H., 1986: A least-square smooth fitting for irregularly spaced data: Finite-element approach using the cubic B-spline basis. *Geophysics,* **51**(11), 2051–2066.

Mariano, A. J., 1988: Space-time interpolation of Gulf Stream north wall positions. Harvard Open Ocean Model Reports, (29), *Reports in Meteorology and Oceanography,* Harvard University.

——, and H. T. Rossby, 1989: The lagrangian potential vorticity balance during POLYMODE. *J. Phys. Oceanogr.,* **19**(11), 927–939.

Pratt, L. J., 1985: Theoretical and numerical studies of Gulf Stream ring formation. *Gulf Stream Workshop Proceedings,* D. R. Watts, Ed., II.468–II.478.

Robinson, A. R., and L. J. Walstad, 1987: The Harvard Open Ocean Model: Calibration and application to dynamical process, forecasting and data assimilation. *J. Appl. Numer. Math.,* **3**(1–2), 89–131.

——, M. A. Spall, W. G. Leslie, L. J. Walstad and D. J. McGillicuddy, 1987: Gulfcasting: Dynamical forecast experiments for Gulf Stream rings and meanders—November 1985–June 1986. Harvard Open Ocean Model Reports, 22, *Reports in Meteorology and Oceanography,* Harvard University.

Shapiro, 1970: Smoothing, filtering and boundary effects. *Rev. Geophys. Space Phys.,* **8**(2), 359–387.