OPTIMAL SPACE-TIME INTERPOLATION
OF GAPPY FRONTAL POSITION DATA

Toshio M. Chin and Arthur J. Mariano
Department of Meteorology and Physical Oceanography, RSMAS, University of Miami

INTRODUCTION

The spatial and temporal variability of Gulf Stream meanders has been studied by many
including Watts and Johns (1982), Halliwell and Mooers (1979 and 1983), Olson et al.
(1983), and Cornillon (1986). The majority of these studies use the northern edge or north
wall, determined from the largest spatial gradient in advanced very high resolution
radiometer (AVHRR) data, as the Gulf Stream path indicator. The advantages of using
AVHRR data for locating the Gulf Stream are (i) the large contemporaneous spatial
coverage, (ii) the measurements have been collected daily since 1978, and (iii) the frontal
locations are the strongest signal in the data. The chief disadvantages are the amount of
processing (geometric corrections, cloud-screening/compositing, and manual digitizing of
frontal positions from images) required and that the satellite sensor cannot see through the
clouds. Consequently, there are large spatial (2-6 degrees) and temporal (3-6 days) gaps in
the Gulf Stream north wall position (GSNWP) data set. Mariano (1988 and 1990) devised
a new approach, termed contour analysis, for melding of oceanic data and for space-time
interpolation of gappy frontal data sets. The key elements of contour analysis are feature
matching and averaging in a coordinate system determined from the contour positions. In
applying his approach to the GSNWP, Mariano assumed a dominant one-dimensional east-
west phase speed in his algorithm. This assumption restricted the application of this
algorithm to other frontal data sets, such as the Brazil-Malvinus confluence (Garzoli et al.,
1992) where the north-south phase speeds are also important, and led to poor estimates of
the GSNWP when the north-south phase speed was significant.

The primary goal of this study is to develop an improved algorithm for space-time
interpolation of gappy frontal data sets. The major improvements are the inclusions of (i)
two-dimensional phase speed, (ii) a more autonomous algorithm, (iii) a better feature
matching algorithm, and (iv) the inclusion of a temporal smoothness constraint. The
space-time interpolator is formulated in the framework of probabilistic (Bayesian)
estimation. This report first reviews such an estimation theoretic framework and, in
particular, a Kalman filter-based interpolation algorithm. Then, feature detection and
matching algorithms are discussed, followed by presentation and discussion of some
preliminary results.
BACKGROUND

The approach described in this report is a two-step process: First, the locations of the sea surface temperature (SST) "edges" (gradient maximums) are detected and digitized by trained personnel at the University of Rhode Island (URI). Then, the longitude-latitude coordinates of the digitized points are interpolated by an autonomous computer program. This report describes this second step—a probabilistic approach to the development of a space-time interpolation algorithm.

The space-time interpolation problem is formulated as a quadratic optimization problem. Here, we review how the cost function can be optimized using a Bayesian estimation framework (with additive white Gaussian noise models) and how the solution can be obtained time-recursively using Kalman filters.

1. Space-only interpolation

We first discuss the problem of interpolating points digitized from a single frame of image, as this is the first step of our space-time interpolation algorithm. Let \((\tilde{x}_i, \tilde{y}_i), i = 1, 2, \ldots, m\) be the longitude-latitude coordinates of the digitized points. We assume, for the time being, that the latitudes \(y\) of the GSNWP can be described by a function of the longitudes \(x\) only, i.e., there exists a single-valued function \(y(x)\). This is a mathematically convenient description used in the previous studies of Gulf Stream variability, but it is not always appropriate for Gulf Stream meanders. The bi-variate formulation for multi-valued features, such as "S" and "Ω" shaped meanders, is discussed after analyzing the simpler single-valued case.

The function \(y(x)\) is interpolated based on the measurements \((\tilde{x}_i, \tilde{y}_i)\) by finding the function that optimizes

\[
\min_y \sum_{i=1}^{m} v_i \left| \tilde{y}_i - y(\tilde{x}_i) \right|^2 + \int_D \left[ \alpha_1 \left| \frac{\partial y}{\partial x} \right|^2 + \alpha_2 \left| \frac{\partial^2 y}{\partial x^2} \right|^2 \right] dx
\]

(1)

where \(v_i\) are the weights representing our confidence in the corresponding measurements. The two integral terms, weighted by the parameters \(\alpha_1\) and \(\alpha_2\), control continuity ("tension") and linearity ("smoothness") of the interpolated curve, respectively. This optimization approach finds applications in general geophysical interpolation and variational problems (e.g., Inoue, 1986).
2. Maximum likelihood estimation

To obtain a numerical solution of Eq. (1), the longitude is discretized as
\[ x = j \Delta x, \, j = 1, 2, \ldots, n. \]  
The interval \( \Delta x \) is chosen small enough for the discrete domain to include (within a reasonable quantization error) the measurements as \( \{ \tilde{x}_j \} \subset \{ x(j \Delta x) \} \),
which implies \( m < n \)—the number of points to be estimated is usually three to four times the number of data points. The corresponding latitudes are represented by an \( n \)-dimensional column vector \( y \) whose elements are \( y(j \Delta x), \, j \in [1, n] \), while the measurements of the latitudes are organized as an \( m \)-dimensional vector \( z \) whose elements are \( \tilde{y}_i, i \in [1, m] \). A discrete version of Eq. (1) is

\[
\min_y \alpha_1 \| S_1 y \|^2 + \alpha_2 \| S_2 y \|^2 + \| z - H y \|^2_M
\]  
(2)

where the vector-norms are weighted 2-norms, e.g., \( \| z - H y \|^2_M = (z - H y)^T M (z - H y) \). (The superscript \( T \) denotes matrix transpose.) The matrixes \( S_1 \) and \( S_2 \) are the first and second order difference operators, respectively, while \( M \) is a diagonal matrix whose diagonal elements are the measurement weights \( v_i, i \in [1, m] \). The \( m \times n \) matrix \( H \) is the data-estimate correspondence operator whose \((i,j)\)th element \( h_{ij} \) is defined as

\[
h_{ij} = \begin{cases} 
1 & \text{if} \, \tilde{x}_i = j \Delta x \\
0 & \text{if otherwise}.
\end{cases}
\]  
(3)

The process of determining the matrix \( H \)—the correspondence problem—is straightforward in this case where the latitudes are treated as a function of the longitudes. Some GSNWP features, such as an “S” shaped meander, can make the correspondence problem quite complex. Mariano (1990) showed that detecting and matching such features based on the sparse sets of data points are the key (and most difficult) components for a successful interpolation scheme. Our solution to the correspondence problem is presented in the next section.

The minimizing solution \( \hat{y} \) of Eq. (2) is exactly the maximum likelihood estimate \( y \) based on the observation equation

\[
\begin{bmatrix}
z \\
0 \\
0
\end{bmatrix} =
\begin{bmatrix}
H \\
S_1 \\
S_2
\end{bmatrix} y +
\begin{bmatrix}
v_H \\
v_1 \\
v_2
\end{bmatrix}
\]  
(4)
where the additive observation noise \( v_H, v_1, \) and \( v_2 \) are mutually independent zero-mean Gaussian random vectors with covariance \( M^{-1}, \alpha_1^{-1}I, \) and \( \alpha_2^{-1}I, \) respectively. The solution of this probabilistic estimation problem requires minimization described in Eq. (2) (Lewis, 1986); thus, the maximum likelihood formulation based on Eq. (4) constitutes a probabilistic interpretation of Eq. (2). An advantage of this probabilistic version is that the estimation error covariance can be computed, along with the estimate itself, allowing us to quantify confidence/uncertainty in the solution. For Eq. (4), the optimal estimate \( \hat{y} \) and estimation error covariance \( P \) are given by

\[
\hat{y} = L^{-1}H^TMz \\
P = L^{-1}
\]

where \( L = H^TMH + \alpha_1S_1^TS_1 + \alpha_2S_2^TS_2 \) is a sparse penta-diagonal matrix. Alternatively, the minimization problem Eq. (2) can also be reformulated as a Bayesian estimation problem in which the first two terms in Eq. (2) are interpreted as the prior statistics for the unknown \( y \) (Szeliski, 1989). Both the Bayesian and maximum likelihood formulations are equivalent when Gaussian noise models are used, as they yield the same solution.

In terms of selecting the parameters for the interpolation problem, the probabilistic formulation must be specified slightly more precisely than its variational counterpart: In Eq. (2) the weights \( \alpha_1, \alpha_2, \) and \( M \) are only required to be specified up to a multiplicative constant—only the ratios among the weights need to be controlled. The same parameters in the probabilistic formulation Eq. (4) play the roles of noise covariances whose values (not just the ratios among them) must exactly be given. This extra bit of precision is necessary for the computed \( P \) to be interpreted meaningfully as the estimation error covariance.

3. Time-extension and Kalman filtering

Equation (1) can be extended temporally to perform space-time interpolation for \( y(x,t) \) using an additional continuity constraint over time:

\[
\min_y \sum_{k=1}^{K} \sum_{i=1}^{m(k)} v_i(k) \left[ \bar{y}_i - y(\bar{x}_i(k), k\Delta t) \right]^2 + \int_0^T \int_D \left[ \alpha_1 \left( \frac{\partial}{\partial x} y \right)^2 + \alpha_2 \left( \frac{\partial^2}{\partial x^2} y \right)^2 + \beta_1 \left( \frac{\partial}{\partial t} y \right)^2 \right] dx \, dt
\]

where the time variable is discretized as \( t = k\Delta t, k = 1,2,\ldots,K \) and the variables associated with the measurements are indexed by \( k \). In the GSNWP estimation problem, \( \Delta t \) is two days. The parameter \( \beta_1 \) controls the strength of the temporal constraint.
A discrete and probabilistic interpretation of Eq. (7) can be obtained by supplementing Eq. (2) with an evolution equation (8) representing the time-continuity constraint. The result is a stochastic dynamic system indexed by the time variable $k$:

$$ y(k) = y(k-1) + w(k) $$

$$ \begin{bmatrix} z(k) \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} H(k) \\ S_1 \\ S_2 \end{bmatrix} \begin{bmatrix} y(k) \\ v_H(k) \\ v_1(k) \\ v_2(k) \end{bmatrix} $$

where $w(k)$ is a zero-mean Gaussian random vector with covariance $\beta^{-1}_i \mathbf{I}$. Representing the space-time interpolation problem as a dynamic system is attractive because the Kalman filtering algorithm (Gelb, 1974) allows computational efficiency (time-recursive computation) and flexibility (filtered, predicted, and smoothed estimates). Numerical solution of the space-time interpolation Eq. (7) is given by the smoothed estimate, which can be computed as a linear combination of forward and backward filtered estimates based on the system Eqs.(8,9): Let $(\hat{y}_f(k), P_f(k))$ be the estimate-covariance pair (the forward estimate) produced by the Kalman filter based on the system equations. Then, Eq. (8) is replaced by a backward dynamic equation $y(k) = y(k+1) + w(k+1)$ to compute the backward filtered estimates and covariances $(\hat{y}_b(k), P_b(k))$. The smoothed estimate-covariance pair $(\hat{y}(k), P(k))$ is given by

$$ \hat{y}(k) = P(k) \left[ P_f^{-1}(k) y_f(k) + P_b^{-1}(k) y_b(k) - H^T(k) M z(k) \right] $$

$$ P(k) = \left\{ P_f^{-1}(k) + P_b^{-1}(k) - H^T(k) M H(k) - \alpha_1 S_1^T S_1 - \alpha_2 S_2^T S_2 \right\}^{-1}. $$

Detailed derivations can be found in textbooks such as Lewis (1986) and Anderson and Moore (1979).

Figure 1a illustrates that the formulation Eq. (7) performs adequate interpolation for a simple ideal case in which $y$ is in fact a function of $x$. Here, for each integer value of $x \in [1,100]$ and $t \in [1,10]$, $y$, is computed as

$$ y = (1+u) \sin \left( \frac{x-2t}{20} \pi \right) \exp \left( \frac{x-5t}{100} \right) $$

where $u \in [0,0.2]$ is a uniformly distributed random number. The “measurements” are made by selecting 25 points along the curve for each $t$ (Fig. 1a). All measurements over the 10 time-frames are shown in Figure 1b by superposition. The interpolated curve (the
Figure 1. (a) An example of space-time interpolation using the formulation Eq. (7) is shown as the solid curve. The dotted curve is the "truth" while the circles are the "measurements" made in this particular time-frame. The dashed curve is a result obtained by adding the "temporal linearity" term (cf. Eq. (15)) into the formulation. (b) All the "measurements" superimposed over time.
solid line in Fig. 1a) estimated the crests of the waveform reasonably well. The parameters used were $M = I$, $\alpha_1 = 0.01$, $\alpha_2 = 1$, and $\beta_1 = 0.1$.

**FORMULATIONS**

1. **Bi-variate unknown**

Problems with uni-variate formulation (i.e., assuming that $y$ is a function of $x$) include inability of representing certain frequently occurring shapes of meanders (e.g., large "S" and "Ω" shapes) and inability to model uncertainty in the measurements of the longitudes $x$. The spatial domain of interpolation must be dynamic, rather than fixed, to correctly assimilate measurements in time under temporal movements of the GSNWP. A dynamic reference frame is crucial to GSNWP interpolation as smoothing over a fixed spatial grid will smear out meanders and other important shape features along the contours, as described by Mariano (1990) in a more general context of data melding. It is an adaptive ("object-oriented") reference frame similar in spirit to the Lagrangian frame. Unlike typical Lagrangian formulations, in which physical motion models are available, our problem must deal with phenomenologically characterized motions of the GSNWP contours, making the formulation challenging because of lack of accurate mathematical models.

We will convert Eq. (7) to a bi-variate formulation. Let $p(s, t) \equiv [x(s, t), y(s, t)]^T$ be the true contour location, where the spatial domain $s$ is the arc-length along the contour at a given $t$. We denote the points digitized from the $k^\text{th}$ SST image as $\tilde{p}_i(k), i \in [1, m(k)]$. The bi-variate version of Eq. (7) is

$$
\begin{align*}
\min_{p} \sum_{k=1}^{K} \sum_{i=1}^{m(k)} v_i(k) \left\| \tilde{p}_i(k) - p(s_i(k), k\Delta t) \right\|^2 \\
+ \int_0^T \int_C \left[ \alpha_1 \left\| \frac{\partial}{\partial s} p \right\|^2 + \alpha_2 \left\| \frac{\partial^2}{\partial s^2} p \right\|^2 + \beta_1 \left\| \frac{\partial}{\partial t} p \right\|^2 \right] ds dt.
\end{align*}
$$

(12)

This minimization is more complex than Eq. (7) because $s_i(k)$, the spatial coordinates (in terms of arc-length) of the digitized points, are unknown. Specifically, the origin of the spatial index $s$ is difficult to define, since there is no guarantee (even though it is a reasonable assumption for the Gulf Stream) that all contours pass through a given point (i.e., the origin) on the $x$-$y$ plane. Also, $s_i(k)$ must be determined concurrently as the contours are interpolated. The arc-length, in fact, cannot be specified exactly without knowing the contour $p(k)$ itself! A Kalman filter-based solution for Eq. (12) becomes an adaptive filtering/smoothing problem:

$$p(k) = p(k-1) + w(k)$$

(13)
where the components of the vector \( \mathbf{q}(k) \) are \( \tilde{p}_i(k), i \in [1,m(k)] \). Note that the data-estimate correspondence matrix \( \mathbf{H}(p(k),k) \) is now dependent on the state \( \mathbf{p}(k) \).

Clearly, Eq. (12) must be optimized adaptively: For each \( k \), either of \( s(k) \) and \( p(k) \) is estimated alternately using the best guess for the other, and this process is iterated for a fixed number of times or until an agreement between the two estimates is obtained within an accuracy parameter. Because of the gaps in the measurements, the estimates at the previous frame (i.e., \( \hat{p}(k-1) \)) are often the best estimates of the general shape of the contour at the current time. Thus, the problem of establishing correspondence can be approached by incrementally matching the best available estimate of the current contour based on the previous contour and that based on the spatially sparse measurements. This important feature matching problem will be addressed in the next section.

2. Imposing linearity over time

Once the data-estimate correspondence is established, it is straightforward to expand the dynamic system formulation Eqs (13,14) to incorporate various structural models for the GSNWP contours. For example, we can impose a linearity constraint over time by inserting an additional integrand term

\[
\beta^2 \left\| \frac{\partial^2}{\partial t^2} \mathbf{p} \right\|^2
\]

(15)


to Eq. (12). The corresponding change in the dynamic system is augmentation of the state vector; the dynamic equation is changed to

\[
\begin{bmatrix}
\mathbf{p}(k) \\
\mathbf{p}(k+1)
\end{bmatrix} =
\begin{bmatrix}
\mathbf{I} & 0 \\
\mathbf{I} & -2\mathbf{I}
\end{bmatrix}
\begin{bmatrix}
\mathbf{p}(k-1) \\
\mathbf{p}(k)
\end{bmatrix} +
\begin{bmatrix}
\mathbf{w}_1(k) \\
\mathbf{w}_2(k)
\end{bmatrix}
\]

(16)

where \( \mathbf{w}_1(k) \) and \( \mathbf{w}_2(k) \) are zero-mean Gaussian random vectors with covariance \( \beta_1^{-1}\mathbf{I} \) and \( \beta_2^{-1}\mathbf{I} \), respectively. Equation (16) can be written in a more attractive form which includes the local displacement \( \mathbf{d}(k) \equiv \mathbf{p}(k+1) - \mathbf{p}(k) \) as the extra component of the state vector. The estimates of the local displacement field are of interest in their own right for statistical characterization of Gulf Stream dynamics. The resulting reformulation consists of a modified dynamic equation and an additional row in the observation equation:
\[
\begin{bmatrix}
  p(k) \\
  d(k)
\end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p(k-1) \\
  d(k-1)
\end{bmatrix} + \begin{bmatrix} 0 \\
  w_2(k)
\end{bmatrix}
\]  

(17)

\[0 = d(k) + v_d(k)
\]  

(18)

where \(v_d(k) = -w_1(k+1)\). By replacing Eq. (13) with Eq. (17) and adding Eq. (18) to Eq. (14), we can jointly estimate the GSNWP \(p(k)\) and the local displacement \(d(k)\). This formulation is the same in spirit as the approach used by Mariano (1990), except that the formulation presented here uses two-dimensional (bi-variate) displacement vectors, instead of one-dimensional in the previous approach, and that the presented formulation is optimal in the least square sense.

The formulation based on Eqs. (17, 18) is applied to the uni-variate example in the previous section, i.e., a temporal linearity constraint (i.e., Eq. (15) imposed on \(v\) instead of \(p\)) is added to Eq. (7). The dashed line in Figure 1a shows one of the resulting interpolated curves. The figure shows that the curve has gained more “stiffness” and the crests of the waves are estimated more accurately with this extra constraint (dashed line) than without it (solid line). The parameter used for the constraint was \(\beta_2 = 0.1\).

FEATURE MATCHING

This section describes an approach to establish the data-estimate correspondence. For conciseness in discussion we discuss the filtering problem based on the dynamic system Eqs. (13, 14). As mentioned before, we adopt an adaptive filtering approach where best predictions of the GSNWP contour at a given time-frame \(k\) are used to estimate the positions, i.e., arc-length indexes \(s_j(k)\), of the measurements along the contour. Specifically, two rudimentary contours, one predicted ahead in time based on the estimated contour at \(k-1\) and the other interpolated only over space based on the measurements at \(k\), are “matched” for correspondence, allowing incorporation of the measurements to update the predicted GSNWP estimate. In another words, the matrix \(H(p(k), k)\) in Eq. (14) is evaluated as \(H(\hat{p}_f(k-1), k)\) in the forward filter and as \(H(\hat{p}_b(k+1), k)\) in the backward filter, where \(\hat{p}_f(k)\) and \(\hat{p}_b(k)\) represent the forward and backward filtered estimates, respectively. The two contours are matched hierarchically—using larger-scale “features” first and then smaller, more local, inflections of the curves.

1. Feature detection

Large bends, especially those at the apexes of the meanders, are the major features along the GSNWP contours. Although these features are always associated with relatively large values of curvature (second-order derivative along the arc), such local attributes alone are not necessarily useful in isolating large meanders among a variety of contour inflections
with much smaller magnitudes. In fact, the magnitudes of the inflections themselves can be used to identify the meander features more directly. These magnitudes are computed as the deviations from a progressively fine-scaled, piece-wise linear approximation of the contour shape. Specifically, consider a segment of the curve between two arbitrary points \( p(s_a) \) and \( p(s_b) \). Let the deviation \( \zeta(s, s_a, s_b) \) be the (perpendicular) distance from the point \( p(s) \) along the segment \( (s \in [s_a, s_b]) \) to the line connecting points \( p(s_a) \) and \( p(s_b) \), as shown in Figure 2. The points along the curve where large deviations occur are used to segment the curve into a piece-wise linear "skeleton", exemplified in Figure 3. Those points associated with large deviations are the nodes of the skeleton of the curve. The following is an iterative algorithm to compute the set of nodes, or node set, given the tolerance parameter \( \epsilon \) for the deviations:

1. Initialize the node set with the two end-points of the curve.
2. Let the number of nodes in the set be \( L \). Let the indexes of the nodes be \( s_\ell \) so that \( s_\ell < s_{(\ell+1)} \) for \( \ell = 1, 2, \ldots, (L - 1) \).
3. Find the maximum deviation \( d^* \) over the entire curve, i.e., for \( \ell = 1, 2, \dots, (L - 1) \),
   \[
   d^* = \max_{s \in s} \max_{\ell} \zeta(s, s_\ell, s_{(\ell+1)})
   \]
   Let \( s^* \) be the spatial index for the point where the maximum deviation occurs.
4. If \( d^* > \epsilon \), include \( s^* \) into the node set; then, go back to step 2 and repeat. Otherwise (\( d^* \leq \epsilon \)), stop.

The internal node points, \( p(s_2), p(s_3), \ldots, p(s_{L-1}) \), after the final iteration are referred to as the feature points.

2. Feature matching

Let us consider matching feature points from two curves. Each feature point is at the apex of a corner on the skeleton of a curve. A cost is assigned to each of possible matching pairs of feature points as a sum of the costs associated with the distance, angle, and direction of the corner. Let \( p(a) \) and \( p(b) \) be feature points from each of the two curves. Each feature point, say \( p(a) \), is a junction of two line segments of the skeleton; let the two unit vectors pointing along these line segments and originating in the feature point \( p(a) \) be \( u_{a1} \) and \( u_{a2} \). Let \( u_{b1} \) and \( u_{b2} \) be similarly defined unit vectors around the feature point \( p(b) \). Also, we measure the direction of a vector \( v \) as the angle \( \angle(v) \) in radians (in the longitude-latitude coordinate system). The costs are defined as
Figure 2. $\zeta(s, s_a$, or $s_b$) equals the length of the line segment $SC$, where points A, B, and S correspond to $p(s), p(s_a)$, and $(s_b)$, respectively.

Figure 3. The contour is segmented by the set of nodes \( \{s_1, s_2, \ldots \} \). The dashed line represents a skeleton for this contour.
1. Distance. \( C_1 = \| p_a - p_b \|^2 \)
   The distance between the pair of points.

2. Angle. \( C_2 = (|\angle(u_{ai}) - \angle(u_{a2})| - |\angle(u_{bi}) - \angle(u_{b2})|)^2 \)
   The absolute value of the difference between the angles of the corners associated with each of the two feature points.

3. Direction. \( C_3 = |\angle(u_{ai} \times u_{a2}) - \angle(u_{bi} \times u_{b2})|^2 \)
   The difference between the directions of the openings of the two corners, i.e., the directions of the vectors bisecting the angles.

We penalize large values of these cost functions more heavily (i.e., more than by a linear proportion) than relatively small values. This is achieved by post-distorting the cost by a piece-wise linear mapping function, such as that shown in Figure 4, which discounts smaller cost values and inflates larger values by multipliers (slopes in the figure) smaller and larger, respectively, than 1.

The pairs of feature points with smaller total costs (the sum of three post-processed cost functions) are considered to be matching pairs, with the following constraints:

- The total cost for any matching pair must be smaller than a specified value, which we will refer to as \( C_{max} \).
- A feature point cannot be matched to more than one other feature point.
- The line segments connecting matched feature points can never cross each other.

The last constraint reflects the structural integrity of the meanders (features): The GSNWP meanders can only appear and disappear; they cannot change their sequencing order along the contour.

To summarize, the number of the parameters to be specified for feature point matching is 10: \( C_{max} \) and the two multipliers and a threshold value (the slopes and “th” in Fig. 4 for each of the three cost functions \( C_1, C_2, \) and \( C_3 \).

3. Local matching

Once correspondence of major features is established, non-feature points can be matched by a simple proportional mapping, leading to a correspondence match of the two contours in their entireties. In Figure 5, for example, the pairs of points \((A, A')\) and \((B, B')\) represent matched feature points, and arc-length indexes \( s \) and \( s' \) along the two contour segments between the respective feature points are considered to be a matching pair if
Figure 4. A typical mapping function for postprocessing of the cost "C" (representing $C_1$, $C_2$, or $C_3$). The values smaller than the threshold "th" are discounted while values larger are inflated.

Figure 5. Mapping contour segment $AB$ to segment $A'B'$.
where $s_A$, $s'_A$, $s_B$, and $s'_B$ are indexes of the feature points.

Unfortunately, matched pairs of feature points are sometimes too sparse to be able to guide correspondence of the two contours reliably. Distance between adjacent feature points on a contour can be larger than the phenomenological length scale, a gap in measured points can occur between feature points, and some measurements do not contain any significant meander features.

To remedy this, we need a secondary method to register the indexes for two given contours without relying on feature identification and matching. One way of performing such a task is to deform one of the contours toward another using a variational formulation involving cost terms for structure of the deformed contour and for distances between points on two contours. Let $p_1(s)$ and $p_2(s)$ be the two contours to be matched and $\rho(s)$ be a deformation of $p_1(s)$. The deformed contour $\rho(s)$ inherits the indexes of $p_1(s)$; thus, by physically registering $\rho(s)$ onto $p_2(s)$, correspondence between the two index sets can be found. [Such a technique for contour registration is generically known as “snake” in computational vision (Kass et al., 1988)]. Specifically, we consider the optimization problem

$$
\min_\rho \int_{C_1} F(p_2, \rho) + \alpha_1 \left\| \frac{\partial}{\partial s} \rho \right\|^2 + \alpha_2 \left\| \frac{\partial^2}{\partial s^2} \rho \right\|^2 \\
+ \gamma_1 \left\| \rho - p_1 \right\|^2 + \gamma_1 \left\| \frac{\partial}{\partial s} (\rho - p_1) \right\|^2 + \gamma_2 \left\| \frac{\partial^2}{\partial s^2} (\rho - p_1) \right\|^2 ds
$$

(20)

where the “gravity” term $F(p_2, \rho)$ works to minimize the distances between points along $\rho(s)$ and $p_2(s)$ and is given by

$$
F(p_2, \rho) \equiv -\int_{C_2} \exp \left( -\frac{1}{2} \left\| \rho(s) - p_2(s') \right\|^2 \right) ds'.
$$

(21)

The domains $C_1$ and $C_2$ of the integrations are given by the contours $p_1$ and $p_2$, respectively. The three cost terms, with coefficients $\gamma_0$, $\gamma_1$, and $\gamma_2$ contain the shape of $\rho(s)$ from becoming radically different from that of $p_1(s)$. The minimizing $\rho(s)$ is given by the non-linear Euler-Lagrange equation.
\[ 2 \left[ (\alpha_2 + \gamma_2) \frac{\partial^4}{\partial s^4} - (\alpha_1 + \gamma_1) \frac{\partial^2}{\partial s^2} + \gamma_0 \right] \rho \\
-2 \left[ \gamma_2 \frac{\partial^4}{\partial s^4} - \gamma_1 \frac{\partial^2}{\partial s^2} + \gamma_0 \right] p_1 + \frac{\partial}{\partial \rho} F(p_2, \rho) = 0 \] (22)

which, since \( \rho \) is the only variable, can be written concisely as

\[ 2 \left[ (\alpha_2 + \gamma_2) \frac{\partial^4}{\partial s^4} - (\alpha_1 + \gamma_1) \frac{\partial^2}{\partial s^2} + \gamma_0 \right] \rho - \bar{p} + \frac{\partial}{\partial \rho} F(\rho) = 0 \] (23)

where

\[ \bar{p} = 2 \left[ \gamma_2 \frac{\partial^4}{\partial s^4} - \gamma_1 \frac{\partial^2}{\partial s^2} + \gamma_0 \right] p_1. \]

Given a parameter \( \kappa \), Eq. (23) can be solved iteratively (Kass et al., 1988) as

\[ 2 \left[ (\alpha_2 + \gamma_2) \frac{\partial^4}{\partial s^4} - (\alpha_1 + \gamma_1) \frac{\partial^2}{\partial s^2} + \gamma_0 + \kappa \right] \rho_\ell \]

\[ = \bar{p} + \kappa(\rho_\ell - \rho_{(\ell-1)}) \frac{\partial}{\partial \rho} F(\rho_{(\ell-1)}) \] (24)

which is equivalent to Eq. (23) if \( \rho_\ell \to \rho \) as \( \ell \to \infty \). Given \( \rho_{(\ell-1)} \), Eq. (24) can be solved simply by inversion of a linear differential operator as

\[ 2 \left[ (\alpha_2 + \gamma_2) \frac{\partial^4}{\partial s^4} - (\alpha_1 + \gamma_1) \frac{\partial^2}{\partial s^2} + \gamma_0 + \kappa \right] \rho_\ell \]

\[ = \bar{p} + \kappa(\rho_\ell - \rho_{(\ell-1)}) \frac{\partial}{\partial \rho} F(\rho_{(\ell-1)}) \] (25)

which we have implemented numerically. The iterations are initialized with \( \rho_0 = p_1 \). Graphically, as the iterations progress, the contour \( \rho_\ell(s) \) approaches \( p_2(s) \) in a structurally constrained manner (from which the name “snake” is derived). When \( \kappa \) is large convergence is slow; when it is small the solution becomes unstable. We have chosen a relatively small value of \( \kappa \) for the first few iterations and then a larger value of \( \kappa \) for the rest of the iterations to ensure convergence. We used a total of about 20 such iterations per solution of Eq. (22).
Figure 6. Examples of interpolated GSNWP contours (solid lines). The small circles are the digitized data points. The cross hairs along the 32°N lines are the standard deviations associated with the estimated contour points directly above them. The lengths of the two arms of each cross represent standard deviations in the estimates of the longitude and latitude associated with the estimated point.
RESULTS

Equation (12), along with the feature detection and matching scheme discussed in the previous section, has been used to interpolate 150 frames of data from the period April 1982 ~ February 1983. Figure 6 shows two of the interpolated GSNWP contours along with the digitized data points (small circles), indicating that the bi-variate formulation is able to reconstruct macroscopic features like the “S” and “Ω” shapes by interpolating data from nearby time-frames. The data points from nearby frames are shown in Figure 7. Also, the standard deviations (produced by the Kalman filter-based algorithm) in the longitude/latitude estimates of selected points are depicted in Figure 6 by the crosshairs (see the figure caption). As expected, the standard deviations are larger away from the data points and smaller near the data points.

The algorithm has been tested further by “hind-forecasting”: a particular frame of data points is removed, and the contour in that frame is then predicted by interpolation based only on data in other frames. Ideally, the predicted contour matches well with the actual data points which did not participate in the interpolation. (Note, however, that the digitized points in a given frame can sometimes misrepresent the true frontal location because of imaging noise, inconsistency among the personnel who perform the digitization task, etc.) Figure 8 shows the hind-forecasted contours of the same two frames as those in Figure 6, while Figure 9 (cf. Fig. 10) shows the hind-forecasts for another pair of frames. In these figures, the data points match fairly well with the hind-forecasts, and, in fact, the agreement between the data and hind-forecasts is observed generally throughout our test. There are, however, several inconsistent hind-forecasts, two of which are shown in Figure 11 (cf. Fig. 12). As indicated in the figure, a major flaw in these hind-forecasts is inability to resolve some fast movements of the meanders and to detect transformations of the meanders into rings. Obviously, simple smoothness constraints like those in Eq. (12) by themselves are not able to handle events such as formation of rings and are heavily dependent on the data to resolve such events.

DISCUSSION

Although the present-day pattern recognition and matching algorithms have yet to realize flexibility and sensitivity of trained personnel, major advantages of a mechanized system in GSNWP estimation are speed, objectivity, and consistency, which are important in high volume production of the estimates. Also, a probabilistic formulation, such as that presented in this report, yields a measure of confidence in the estimates in the form of the second order statistics to facilitate interpretation of the results. We feel that such a statistical interpretation will be enhanced if the uncertainty (noise variance) in each digitized data point is quantified by using a probabilistic edge-detection algorithm (e.g., Canny, 1986) on the SST images. A new edge detection algorithm using both spatial and temporal constraints is being tested by Cayula and Cornillon (cf. 1990) at URI. A symbiotic merging of such an edge detection with our interpolation algorithm should
Figure 7. The digitized data points from five frames centered around the two frames depicted on Figure 6. Each of two columns of five frames shows a time-sequence of the digitized data points, with the third frame being the frame from Figure 6.
Figure 8. Hind-forecasts for the two frames in Figure 6.
Figure 9. Two more examples from the hind-forecasting test.
Figure 10. The digitized data points from five frames centered around the two frames depicted on Figure 9. Each of two columns of five frames shows a time-sequence of the digitized data points, with the third frame being the frame from Figure 9.
Figure 11. Two cases where hind-forecasts have failed, due to temporal Gulf Stream dynamics unresolvable from this particular data sequence.
Figure 12. The digitized data points from five frames centered around the two frames depicted on Figure 11. Each of two columns of five frames shows a time-sequence of the digitized data points, with the third frame being the frame from Figure 11.
reduce the effect of inconsistencies in the initial frontal locations. We are also considering a higher order model for contour dynamics (Pratt and Stern, 1986) as an extension of the work presented in this report.

In the near future, all available digitized frontal locations in the Gulf Stream, Brazil-Malvinus confluence, and Kuroshio current systems will be interpolated. The spatial/temporal variability and phase speed distribution of the resulting complete frontal locations will be documented.

Acknowledgments. This work is supported by Office of Naval Research Random Field in Oceanography ARI under Grant N00014-91-J-1120. We would also like to thank Tong Lee and Peter Cornillon for sharing the GSNWP data set with us.

REFERENCES


