On the Predictability of Lagrangian Trajectories in the Ocean

TAMAY M. ÖZGÖKMEN, ANNALISA GRIFFA,* AND ARTHUR J. MARIANO

RSMAS/MPO, University of Miami, Miami, Florida

LEONID I. PITERBARG

Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, California

(Manuscript received 23 February 1999, in final form 25 February 1999)

ABSTRACT

The predictability of particle trajectories in oceanic flows is investigated in the context of a primitive equation, idealized, double-gyre ocean model. This study is motivated not only by the fact that this is an important conceptual problem but also by practical applications, such as searching for objects lost at sea, and ecological problems, such as the spreading of pollutants or fish larvae. The original aspect of this study is the use of Lagrangian drifter data to improve the accuracy of predicted trajectories. The prediction is performed by assimilating velocity data from the surrounding drifters into a Gauss–Markov model for particle motion. The assimilation is carried out using a simplified Kalman filter.

The performance of the prediction scheme is quantified as a function of a number of factors: 1) dynamically different flow regimes, such as interior gyre, western boundary current, and midlatitude jet regions; 2) density of drifter data used in assimilation; and 3) uncertainties in the knowledge of the mean flow field and the initial conditions. The data density is quantified by the number of data per degrees of freedom N_R, defined as the number of drifters within the typical Eulerian space scale from the prediction particle. The simulations indicate that the actual World Ocean Circulation Experiment sampling (1 particle/[5° × 5°] or $N_R \ll 1$) does not improve particle prediction, but predictions improve significantly when $N_R \gg 1$. For instance, a coverage of 1 particle/[1° × 1°] or $N_R \sim O(1)$ is already able to reduce the errors of about one-third or one-half. If the sampling resolution is increased to 1 particle/[0.5° × 0.5°] or 1 particle/[0.25° × 0.25°] or $N_R \gg 1$, reasonably accurate predictions (rms errors of less than 50 km) can be obtained for periods ranging from one week (western boundary current and midlatitude jet regions) to three months (interior gyre region). Even when the mean flow field and initial turbulent velocities are not known accurately, the information derived from the surrounding drifter data is shown to compensate when $N_R > 1$. Theoretical error estimates are derived that are based on the main statistical parameters of the flow field. Theoretical formulas show good agreement with the numerical results, and hence, they may serve as useful a priori estimates of Lagrangian prediction error for practical applications.

1. Introduction

The Lagrangian description of a flow field is a useful tool with which to explore oceanic circulation and transport. Logistic difficulties have delayed the application of Lagrangian methods to oceanographic field research programs. Stommel (1949) used aerial photography of floating paper sheets serving as Lagrangian drifters to observe oceanic turbulence. Lagrangian instruments such as drifting buoys were later designed to move with the currents so that their trajectories satisfactorily approximate the motion of fluid parcels. Satellite positioning revolutionized drifter technology, making possible cost-effective long-term global tracking. Lagrangian data have become increasingly more common in the last three decades, providing good coverage of extensive regions in the ocean (Davis 1991a).

Lagrangian observations can be analyzed in a variety of ways. In his seminal work, Taylor (1921) demonstrated that Lagrangian statistics can be used to estimate the dispersion of particles in homogeneous turbulent flow by relating the autocorrelation function of the Lagrangian velocity field to a diffusivity coefficient. This relationship has been verified within observation error for subsurface ocean flow using data from sound fixing and ranging (SOFAR) floats (Rossby et al. 1983). This relationship has been used extensively to provide a measure of diffusivity properties from oceanographic drifter data, with the purpose of studying the evolution of pas-

* Additional affiliation: IOF, Consiglio Nazionale Ricerche, La Spezia, Italy.

Corresponding author address: Dr. Tamay M. Özgökmen, RSMAS/MPO, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149-1098.
E-mail: tozgokmen@rsmas.miami.edu

sive tracer fields using the advection–diffusion equation (e.g., Colin de Verdiere 1983; Davis 1985; Figueroa and Olson 1989; Davis 1991b). This type of analysis has been extended also to eddy-resolving numerical ocean models (e.g., Böning and Cox 1988; Figueroa and Olson 1994). A conceptually equivalent approach is the use of stochastic models to describe the motion of particles (Griffa 1996). These models can be employed for ''direct'' applications, in which the dispersion of tracers is studied considering the behavior of a large number of tracer particles (e.g., Dutkiewicz et al. 1993). Alternatively, such models can be used for ''indirect'' applications to extract information about turbulence parameters from Lagrangian data (e.g., Griffa et al. 1995). Finally, Lagrangian data have also been assimilated into numerical models to improve the accuracy of simulated Eulerian flow fields (Carter 1989; Ishikawa et al. 1996).

In this paper, we focus on the use of Lagrangian data for predicting oceanic particle motion. The investigation of particle motion and transport is an important area of study, with a number of potential practical applications at very different scales, including searching for persons or valuable objects lost at sea, tracking floating mines, studying ecological problems such as the spreading of pollutants or fish larvae, and estimating mass transport for climate studies. As has been pointed out by a number of authors (e.g., Aref 1984; Samelson 1996), transport prediction is an intrinsically difficult problem because Lagrangian motion often exhibits chaotic behavior, even in regular and simple Eulerian flows. In the ocean, the combined effects of complex time dependence (Samelson 1992; Meyers 1994) and three-dimensional structure (Yang and Liu 1997) are likely to induce chaotic transport, especially in the more complex and turbulent regions such as the Gulf Stream Jet and extension. Chaos implies strong dependence on initial conditions, which are usually not known with great accuracy, so that the task of predicting particle motion is often extremely difficult. Some previous studies on particle motion in oceanic flows have focused on the characterization of trajectories using dynamical system tools, with special interest on exchanges between different flow regimes (e.g., Duan and Wiggins 1996; Samelson 1996).

The problem and approach considered here are conceptually different. We consider tracking and prediction of specified particles, given Lagrangian information of the surrounding flow field. Lagrangian data are the natural choice for predicting particle dispersion because they move with the flow. Therefore, from such data, information can be assimilated advantageously into the equations governing particle motion. This has been shown in a specific example by the preliminary study of Okamoto et al. (1994), in which the trajectories of three drifters in the Pacific Ocean have been simulated by assimilating position information from nearby drifters.

In the following, we assume that two types of information are available: 1) concurrent Lagrangian data and 2) some a priori knowledge of the large-scale mean flow, obtained for example, from historical data. In this setting, the knowledge of the mesoscale turbulent motion relies completely on the Lagrangian data. This allows us to isolate and to quantify the potential of Lagrangian data in improving the predictability of particle trajectories. The Lagrangian prediction experiments are designed to test the use of a simple stochastic model, Lagrangian data, and a simple Kalman filter for practical applications.

A comprehensive study of the methodology and its effectiveness is presented. The investigation is conducted in the context of an idealized primitive equation double-gyre ocean model. Numerical drifters are released in the flow field and are used to reconstruct the trajectories of the specified unknown particles. We investigate in detail the accuracy of the predictions as a function of a number of important factors: 1) the type of dynamically different flow regimes—the gyre interior, western boundary current, and midlatitude jet regions; 2) the density of drifter data used in the assimilation; and 3) uncertainties in the initial conditions and in the knowledge of the mean flow field. The results are discussed in terms of practical applications and conceptual implications for predictability properties.

The paper is organized as follows. The mathematical background and the Lagrangian prediction model are introduced in section 2. The methodology for the generation and assimilation of the Lagrangian data is discussed in section 3. The prediction results are presented and discussed in section 4, and section 5 contains a summary of this work and some remarks regarding future studies.

2. The prediction problem

a. Definitions and assumptions

We consider an Eulerian velocity field, $\mathbf{u}(t, \mathbf{r})$, in two dimensions $\mathbf{r} = (x, y)$, representing the flow at the ocean surface or along an isopycnal surface. Let

$$\mathbf{u}(t, \mathbf{r}) = \mathbf{U}(t, \mathbf{r}) + \mathbf{u}'(t, \mathbf{r}) \tag{1}$$

be a decomposition of $\mathbf{u}(t, \mathbf{r})$ into a mean circulation and a fluctuating component; that is, $\mathbf{U}$ is a deterministic field representing the large-scale, slowly varying component of the velocity field and $\mathbf{u}'(t, \mathbf{r})$ is a random (eddy) velocity field with zero mean $\langle \mathbf{u}'(t, \mathbf{r}) \rangle = 0$ and known covariance function

$$\mathbf{B}^{\mathbf{u}}(t_1, t_2; \mathbf{r}_1, \mathbf{r}_2) = \langle \mathbf{u}'(t_1, \mathbf{r}_1) \mathbf{u}'(t_2, \mathbf{r}_2)^{\mathrm{T}} \rangle. \tag{2}$$

The angle brackets denote ensemble averaging over all realizations of the velocity field, the superscript T stands for transposition, and the vectors are considered column vectors. Let $\mathbf{r}(t)$ denote the position of a particle released at the position $\mathbf{r}^0$ and driven by (1). The temporal evolution of $\mathbf{r}(t)$ is described by the equation

$$\frac{d\mathbf{r}}{dt} = \mathbf{u}(t, \mathbf{r}), \qquad \mathbf{r}(0) = \mathbf{r}^0. \tag{3}$$

Similarly, the Lagrangian velocity is

$$\mathbf{v}(t) = \mathbf{v}(t, \mathbf{r}^0) = \frac{d\mathbf{r}}{dt} = \mathbf{u}(t, \mathbf{r}(t)),$$

and let

$$\mathbf{v}(t) = \mathbf{U}(t, \mathbf{r}(t)) + \mathbf{v}'(t), \tag{4}$$

where $\mathbf{U}[t, \mathbf{r}(t)]$ is the Eulerian mean velocity at $\mathbf{r}(t)$, and $\mathbf{v}'(t)$ is the Lagrangian eddy velocity. In the following, we assume that $\mathbf{v}'(t)$ satisfies the following Langevin equation (Thomson 1987):

$$d\mathbf{v}' + \mathbf{T}^{-1}\mathbf{v}'\, dt = d\mathbf{w}(t), \tag{5}$$

where $\mathbf{T}$ is the relaxation time matrix and $\mathbf{w}(t)$ is a 2D Brownian motion with covariance matrix $\boldsymbol{\sigma}^2$. For simplicity, it is assumed that

$$\mathbf{T} = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix},$$

where T_1, T_2 are the zonal and meridional relaxation times, respectively.

Equation (5) implies that the Lagrangian eddy velocity is modeled as a linear Markov process characterized by an exponential autocovariance function. This type of model is supported by experimental and numerical results in fully developed turbulent flows (e.g., Yeung and Pope 1989), and it appears appropriate also for oceanic flows (Griffa 1996), especially at the sea surface, where exponential Lagrangian autocovariance functions are observed commonly from surface drifter data (e.g., Davis 1985; Swenson and Niiler 1996; Bauer et al. 1998). Even in situations in which the autocovariance function has a more complex structure—for example, in the presence of strong wave signatures—Eq. (5) can still be considered a first-order approximation valid for timescales of order T, which typically are between 3 and 10 days for ocean surface flows.

There are M equations of type (5) for M particles that can be discretized in time, yielding autoregressive equations of the first order (Box and Jenkins 1976):

$$\mathbf{v}_j'(n) = \alpha\mathbf{v}_j'(n-1) + \mathbf{w}_j(n), \qquad j = 1, \ldots, M,$$

where $\alpha = \exp\{-\mathbf{T}^{-1}\Delta t\}$; Δt is the time step; n is the number of time steps and will be used to denote time $n\Delta t$: for example, $\mathbf{v}_j'(n) = \mathbf{v}_j'(n\Delta t)$; and $\mathbf{w}_j(n)$ are white noises discrete in time. These noise terms, generally speaking, are correlated, and their mutual covariances are determined by the mean flow, statistics of the Eulerian velocity field, and initial positions. In general, this relation is very complicated and cannot be expressed in an explicit form. However, Lagrangian statistics can be related to the corresponding Eulerian statistics for a given set of particle positions. To formulate such a relation, let us introduce the conditional averaging operator $E_n\{\ \}$. Namely, $E_n\{\xi\}$ is defined for any random variable ξ as the average (or expectation) of ξ under the condition that all particle positions are known at all times 0, 1, 2, . . . , n. Then for each particle,

$$E_n\{\mathbf{v}_i(n)\} = \mathbf{U}(n, \mathbf{r}_i(n)), \tag{6}$$

where $\mathbf{v}_i(n)$ is the Lagrangian velocity of the ith particle at time n, and for any pair,

$$E_n\{\mathbf{v}_i'(n)\mathbf{v}_j'(n)^{\mathrm{T}}\} = \mathbf{B}^{\mathbf{u}}(n, n; \mathbf{r}_i(n), \mathbf{r}_j(n)). \tag{7}$$

Rigorously speaking, Eqs. (6) and (7) hold true only under strong limitations, such as an independence of positions and velocities, or homogeneity, stationarity, and white-noise assumption for velocity. In the general case Eqs. (6) and (7) should be viewed as a plausible hypothesis that is in good agreement with one's intuition and experience.

b. Prediction problem statement

We assume that $M \geq 1$ particles start at the same time ($n = 0$) from different positions $\mathbf{r}_j^0$, $j = 1, \ldots, M$. At each time $n > 0$, we observe the exact (or with a small error) positions of the first $p < M$ particles, called predictors. The general problem is to find a best prediction of the positions of the last $M - p$ unobserved particles (predictands) at each time, provided that the mean velocity field $\mathbf{U}(t, \mathbf{r})$ is known, as well as the Eulerian covariances and Lagrangian relaxation times. In this paper, we concentrate on the most simple situation of a single predictand, $p = M - 1$.

The discretized equations used to describe the system are

$$\mathbf{v}_j'(n) = \alpha\mathbf{v}_j'(n-1) + \mathbf{w}_j(n), \qquad j = 1, \ldots, M, \tag{8}$$

$$\mathbf{r}_j(n+1) = \mathbf{r}_j(n) + (\mathbf{U}(n, \mathbf{r}_j(n)) + \mathbf{v}_j'(n))\Delta t, \tag{9}$$

and the objective is to find a best prediction, $\mathbf{r}_M^a(n)$, of the position $\mathbf{r}_M(n)$ of the unobservable particle $j = M$ at any prescribed time $n = 1, 2, 3, \ldots$, given its initial position at time $n = 0$.

An accurate method to find the optimal prediction is to use the extended Kalman filtering (EKF) (e.g., Ghil and Malanotte-Rizzoli 1991; Bril 1995; Ide and Ghil 1998) applied to the nonlinear system [Eqs. (8)–(9)], assuming that the state vector has dimension $4M$ containing the positions of all M particles and their fluctuating velocities. A simpler—though less accurate—approach that allows a greater computational simplicity, a clear physical interpretation, and a theoretical error analysis is used here. It consists of finding the best estimate $\mathbf{v}_M^a(n)$ for $\mathbf{v}_M'(n)$, using Eq. (8) and the observations, and then predicting the position $\mathbf{r}_M^a(n+1)$ using $\mathbf{v}_M^a(n)$ in Eq. (9). Since Eq. (8) is linear, the best prediction can be calculated using the classical Kalman filter, using a reduced state vector with dimension $2M$, including only the fluctuating velocities. The observa-

tions in this approach are the fluctuation velocities observed by the $p = M - 1$ particles, computed as $\mathbf{v}_j'(n) = (\mathbf{r}_j(n) - \mathbf{r}_j(n-1))\Delta t^{-1} - \mathbf{U}(n, \mathbf{r}_j(n))$, $j = 1, \ldots, p$. An implementation of the full EKF for the complete system is planned for the future. Note that in the case of constant mean velocity, the two approaches are equivalent.

c. Kalman filter for Lagrangian fluctuation velocities

Let us introduce the state and noise vectors as follows:

$$\mathbf{V}' = \begin{pmatrix} \mathbf{v}_1' \\ \cdots \\ \mathbf{v}_{M-1}' \\ \mathbf{v}_M' \end{pmatrix}, \qquad \mathbf{w} = \begin{pmatrix} \mathbf{w}_1 \\ \cdots \\ \mathbf{w}_{M-1} \\ \mathbf{w}_M \end{pmatrix}.$$

Then, the dynamics of (8) can be written as

$$\mathbf{V}'(n) = \mathbf{A}\mathbf{V}'(n-1) + \mathbf{W}(n), \tag{10}$$

where $\mathbf{A}$ is the $2M \times 2M$ diagonal matrix with 2×2 blocks of α along the diagonal. Denote the conditional covariance matrix of the noise by $\mathbf{Q}(n)$; that is,

$$\mathbf{Q}(n) = E_n\{\mathbf{w}_i(n)\mathbf{w}_j(n)^{\mathrm{T}}\}. \tag{11}$$

Notice that, as shown in the appendix (section a), $\mathbf{Q}(n)$ is related in a very simple way to the Eulerian space covariance (2):

$$\mathbf{Q}(n) = (\mathbf{I}_{2M} - \mathbf{A})(\mathbf{B}_n^{\mathbf{u}}(\mathbf{r}_i(n), \mathbf{r}_j(n)))(\mathbf{I}_{2M} - \mathbf{A})^{\mathrm{T}}, \tag{12}$$

where $\mathbf{I}_{2M}$ is the $2M \times 2M$ unit matrix.

Now we introduce the observation vector $\mathbf{Y}$ via

$$\mathbf{Y} = \mathbf{H}\mathbf{V}' + \varepsilon\mathbf{b}, \tag{13}$$

where

$$\mathbf{H} = (\mathbf{I}_{2p} \quad \mathbf{0})$$

is the $2p \times 2M$ observation matrix, $\mathbf{0}$ is the $2p \times 2$ zero matrix, ε is the observation error level, and $\mathbf{b}$ is a white-noise process with zero mean and unit covariance matrix. Simply, $\mathbf{Y}$ consists of the velocities of the first M particles disturbed by a small noise. Then, the prediction with minimum mean square error is given by (e.g., Miller 1989; Jazwinskii 1970)

$$\mathbf{V}^a(n+1) = \mathbf{V}^f(n) + \mathbf{K}(\mathbf{Y}(n) - \mathbf{H}\mathbf{V}^f(n)),$$

where

$$\mathbf{V}^f(n) = \mathbf{A}\mathbf{V}^a(n-1), \qquad \mathbf{K} = \mathbf{P}^f\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \varepsilon^2\mathbf{I}_{2p})^{-1}.$$

The forecast error matrix

$$\mathbf{P}^f = E_n\{(\mathbf{V}^f(n) - \mathbf{V}'(n))(\mathbf{V}^f(n) - \mathbf{V}'(n))^{\mathrm{T}}\}$$

is computed by

$$\mathbf{P}^f(n+1) = \mathbf{A}\mathbf{P}^a(n)\mathbf{A}^{\mathrm{T}} + \mathbf{Q}(n),$$

and the analysis error covariance

$$\mathbf{P}^a = E_n\{(\mathbf{V}^a(n) - \mathbf{V}'(n))(\mathbf{V}^a(n) - \mathbf{V}'(n))^{\mathrm{T}})\}$$

is given by

$$\mathbf{P}^a(n) = (\mathbf{I}_{2M} - \mathbf{K}(n)\mathbf{H})\mathbf{P}^f(n).$$

Because of the very simple structure of the deterministic part of the dynamics and the observation matrix [see the appendix (section b)], the foregoing equations can be simplified drastically to obtain the following optimal prediction for the Mth particle:

$$\mathbf{v}_M^a(n) = \alpha\mathbf{v}_M^a(n-1) + \sum_{j=1}^{p} \mathbf{K}_j(\mathbf{v}_j'(n) - \alpha\mathbf{v}_j'(n-1)), \tag{14}$$

where $\mathbf{K}_j$ is the 2×2 matrix such that the weight "vector"

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_1 \\ \cdots \\ \mathbf{K}_p \end{pmatrix}$$

is the solution to

$$\mathbf{Q}_{11}\mathbf{K} = \mathbf{Q}_{21}. \tag{15}$$

The mean square error

$$\mathbf{S}(n) = E_n\{(\mathbf{v}_M^a(n) - \mathbf{v}_M'(n))(\mathbf{v}_M^a(n) - \mathbf{v}_M'(n))^{\mathrm{T}}\}$$

can be computed via

$$\mathbf{S}(n) = \mathbf{Q}_{22}(n) - \mathbf{Q}_{21}(n)\mathbf{Q}_{11}(n)^{-1}\mathbf{Q}_{12}(n) + \alpha\mathbf{S}(n-1)\alpha^{\mathrm{T}}. \tag{16}$$

The matrices $\mathbf{Q}_{11}$, $\mathbf{Q}_{12}$, $\mathbf{Q}_{21}$, and $\mathbf{Q}_{22}$ are specific blocks of matrix $\mathbf{Q}$ defined in the appendix (section b) and Eq. (20). In turn, $\mathbf{Q}$ is expressed explicitly in terms of the Eulerian covariances and Lagrangian relaxation times [Eq. (12)]. Thus, the prediction formulas are based on quantities estimated well from observations and/or synthetic data, such as numerical model output.

d. Prediction algorithm

A prediction algorithm is implemented by making some simplifying assumptions regarding the Eulerian statistics. This is done for computational simplicity, and as will be shown, these assumptions are justified by the numerical results. More accurate algorithms using more realistic statistics can be implemented for specific applications.

We assume that the velocity components are uncorrelated and that the Eulerian covariance can be parameterized using a Gaussian function for each component:

$$B(\mathbf{r}) = \exp\left\{-\frac{|\mathbf{r}|^2}{2R^2}\right\}, \tag{17}$$

where the parameter R is the Eulerian correlation velocity scale defined as $R = \sqrt{-B(\mathbf{0})/B''(\mathbf{0})}$.

The prediction is then carried out using the following two-step algorithm:

1) Calculate the turbulent Lagrangian velocity at time n from

$$\mathbf{v}_M^a(n) = \alpha \mathbf{v}_M^a(n-1) + \sum_{j=1}^{p} \mathbf{K}_j(\mathbf{v}_j'(n) - \alpha \mathbf{v}_j'(n-1)), \quad (18)$$

$$\mathbf{Q}_{11}\mathbf{K} = \mathbf{Q}_{21}, \quad (19)$$

where

$$(\mathbf{Q}_{11})_{ij} = B(\mathbf{r}_i(n) - \mathbf{r}_j(n)),$$
$$(\mathbf{Q}_{21})_i = B(\mathbf{r}_M^a(n) - \mathbf{r}_i(n)) \quad (20)$$

is defined by Eq. (12) and $B(\mathbf{r})$ is calculated from Eq. (17).

2) Calculate the predictand position at time $n + 1$ from

$$\mathbf{r}_M^a(n+1) = \mathbf{r}_M^a(n) + [\mathbf{U}(n, \mathbf{r}_M^a(n)) + \mathbf{v}_M^a(n)]\Delta t. \quad (21)$$

e. Error estimates

The error analysis for the prediction Eqs. (18)–(21) is presented in the appendix (section c). Here, we summarize the results of most practical interest. In particular, an estimate of the error for the predicted positions is obtained under some assumptions and in some limiting cases.

The fluctuating velocity field is assumed to be statistically homogeneous [i.e., $\mathbf{B}_n^{\mathbf{u}}(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{B}_n^{\mathbf{u}}(\mathbf{r}_1 - \mathbf{r}_2)$], the velocity components are uncorrelated, and the mean velocity field can be considered approximately constant over a certain time interval. We then define $\sigma_u^2 = B(0)$ as the velocity variance and N_R as the average number of predictors in the circle O_R of radius R centered at the predictand. Notice that N_R can be interpreted as the number of data per degrees of freedom. It can be shown that the position prediction error $s_x = \langle(\hat{x}_M(t^O) - x_M(t^O))^2\rangle^{1/2}$ in the x direction, where t^O is the observation time, grows first linearly in time,

$$s_x \sim \gamma t^O, \quad \text{if } t^O \ll T, \quad (22)$$

and then as the square root of time,

$$s_x \sim \gamma\sqrt{t^O T}, \quad \text{if } t^O \gg T. \quad (23)$$

The variable γ has different values depending on N_R:

$$\gamma^2 = \sigma_u^2 \quad \text{if } N_R = 0, \quad (24)$$

$$\gamma^2 = c\sigma_u^2 \quad \text{if } N_R = 1, \quad (25)$$

and $B(r) = \exp\{-r^2/2R^2\}$, where

$$c = \frac{7e-2}{12e} + \frac{\sqrt{\pi}}{3}\,\text{erf}(1) \sim 0.106,$$

$$\gamma^2 \sim \sigma_u^2\left(\frac{\ln N_R}{12N_R}\left(1 - \frac{1}{N_R\sqrt{N_R}}\right) + \frac{1}{N_R\sqrt{N_R}}\right) \quad \text{if } \frac{1}{N_R} \to 0. \quad (26)$$

These results show that the Lagrangian prediction errors are proportional to single-particle dispersion with a constant of proportionality that decreases with N_R. These error estimates can be considered general guidelines to be checked in the following with the numerical simulation results.

3. Generation and assimilation of the Lagrangian data

In this section, the numerical ocean model that is used to generate the Lagrangian drifter data is described and the experimental setup for trajectory predictions is discussed.

a. The numerical ocean model

The numerical model used in this study is the Miami Isopycnic Coordinate Ocean Model (Bleck and Boudra 1986). It is a primitive equation, isopycnic vertical coordinate model that can be viewed as a stack of shallow-water models, each consisting of a momentum and a continuity equation:

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{2}\nabla_\rho \mathbf{u}^2 + (\zeta + f)\mathbf{k} \times \mathbf{u} = -\nabla_\rho M + \theta\frac{\partial \boldsymbol{\tau}}{\partial z} + \nu h^{-1}\nabla_\rho \cdot h\nabla_\rho \mathbf{u} - \varepsilon \mathbf{u}, \quad (27)$$

$$\frac{\partial h}{\partial t} + \nabla_\rho \cdot (\mathbf{u}h) = 0, \quad (28)$$

where $\mathbf{u}$ is the Eulerian velocity field, $\zeta = v_x - u_y$ the vorticity, f the Coriolis frequency, $M = gz + p\theta$ the Montgomery potential, $\theta = \rho^{-1}$ the specific volume, $\boldsymbol{\tau}$ the wind stress vector, h the thickness of a layer of constant density, ν the lateral viscosity coefficient, and ε the bottom drag coefficient that acts only on the bottom layer. The subscript ρ indicates derivatives on surfaces of constant density. The layers communicate vertically through pressure forces. Horizontal velocities and vorticity are defined as mean layer properties. Pressure and geopotential are defined at the interfaces between layers. For more detail on the numerical model, the reader is referred to Bleck and Boudra (1986) and Bleck and Chassignet (1994).

For the purposes of this study, the ocean model is set up in an idealized configuration—the so-called box domain with a flat bottom and straight sidewalls. No-flow and free-slip conditions are applied along all lateral boundaries. The vertical density stratification is represented by three isopycnal layers. The parameters of the ocean model are listed in Table 1. The model is purely wind driven by a steady zonal wind stress of the form $\tau^x = -\tau_0 \sin(\pi y/L_y)$, where L_y is the meridional extent of the basin and y is the distance from the center of the domain. Hence, the wind forcing drives a classic double-gyre circulation representative of an anticylonic sub-

TABLE 1. Parameters of the numerical ocean model simulation.

Zonal basin size	$L_x = 3000$ km
Meridional basin size	$L_y = 3000$ km
Horizontal grid scale	$\Delta x = \Delta y = 20$ km
Coriolis parameters	$f_0 = 9.3 \times 10^{-5}$ s^{-1} $\beta = 2 \times 10^{-11}$ m^{-1} s^{-1}
Layer thicknesses	$h1 = 300$ m $h_2 = 700$ m $h_3 = 4000$ m
Reduced gravities	$g'_{12} = 0.016$ m s^{-2} $g'_{23} = 0.014$ m s^{-2}
Deformation radii	$R_{d1} = 39$ km $R_{d2} = 19$ km
Wind stress amplitude	$\tau_0 = 0.1$ N m^{-1}
Lateral viscosity	$\nu = 100$ m^2 s^{-1}
Bottom friction	$\epsilon = 10^{-7}$ s^{-1}
Time step	$\Delta t = 1200$ s

tropical gyre in the south and a cyclonic subpolar gyre in the north, separated by a midlatitude jet. The model equations are integrated for 15 years, starting from rest, until the layer kinetic and available potential energy levels reach a statistically steady state. Figure 1 shows the Eulerian mean field for the upper-layer circulation, averaged over 2 years (i.e., over years 15–17). Results from wind-driven models in a double-gyre configuration with idealized geometry are discussed extensively in the literature (e.g., Holland 1978; Chassignet and Gent 1991).

Figure 2 exhibits the instantaneous fields of the upper-layer thickness for the 3-month period during which the prediction experiments for the drifter trajectories are conducted. This figure indicates clearly the turbulent character of the model simulation. Not only does the midlatitude jet meander, generate mesoscale eddies, and interact with them, but the interior gyre circulation also shows high variability, implying that the drifter trajectories will exhibit a complex behavior in the entire domain.

b. Drifter trajectories

Particle trajectories are calculated according to a discretized version of Eq. (3), in which **u** is the velocity field computed by the numerical model (27). Equation (3) is integrated in time using a second-order Runge–Kutta method, and a bilinear interpolation scheme is employed for subgrid-scale positioning. The accuracy of these numerical methods is verified using various tests, and a time interval of 18 000 s is selected to advance Eq. (3) in time.

The particles are released in the top layer of the subtropical portion of the domain in a regular array, representing the motion of near-surface ocean drifters. Examples of drifter trajectories, released with a density of 1 particle/[1° × 1°], approximately 3 months after release, are depicted in Fig. 3. Dynamically different regions are apparent from this graph: 1) the western boundary current region, in which particles traverse to the north along the coast; 2) the midlatitude jet region, in which the western boundary current separates from the coast, meanders as a free jet, and induces high variability and eddy activity; and 3) the interior region, in which the trajectories still exhibit a complex behavior, even though velocities are much smaller than those in the western boundary current and the meandering jet region.

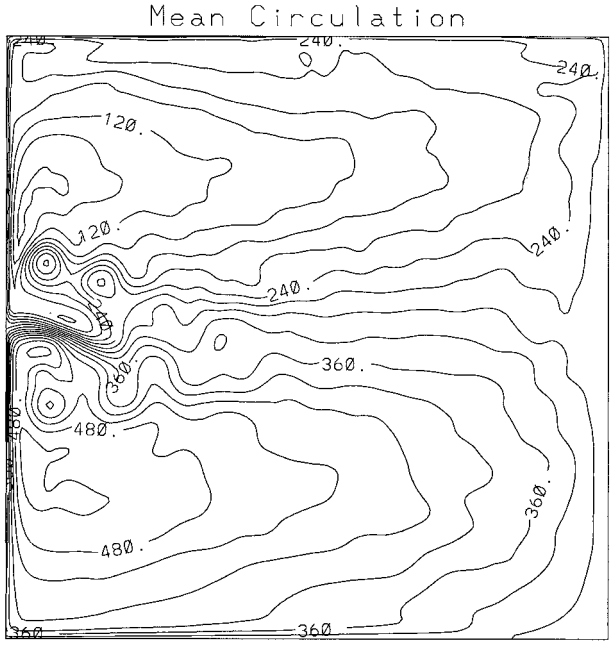

FIG. 1. The Eulerian mean field of the top layer thickness, h_1, averaged during years 15–17 (contour interval: 30 m).

c. Assimilation of drifter velocity

The prediction of particle trajectories presented in the following is implemented within quasi-homogeneous regions, in which the Eulerian and Lagrangian statistics can be described using simple functions characterized by a few parameters: the Lagrangian time scale T, the Eulerian space scale R, and the velocity variance σ_u^2. The partition of the subtropical velocity field into seven quasi-homogeneous regions is based on the eddy kinetic energy distribution of the Eulerian velocity field (Fig. 4). Regions 1 and 2 capture the midlatitude jet meander and ring region. Regions 4–6 specify the interior gyre regions. Region 3 is a transition zone between the midlatitude jet and interior regions. Region 7 represents the western boundary current.

The calculations have been conducted for all seven regions, and both the statistical and predictive behaviors indicate significant similarities between some of the regions. Regions 1 and 2 and regions 3–6 behave qualitatively similarly. For the purpose of a concise presentation of results, region 1 is selected to discuss the characteristics of the high-energy midlatitude jet area, and region 4 is chosen to represent the low-energy interior regions. In the rest of the paper, the discussion of the results therefore will be confined to regions 1, 4, and 7, which represent the three distinct regimes of the ocean

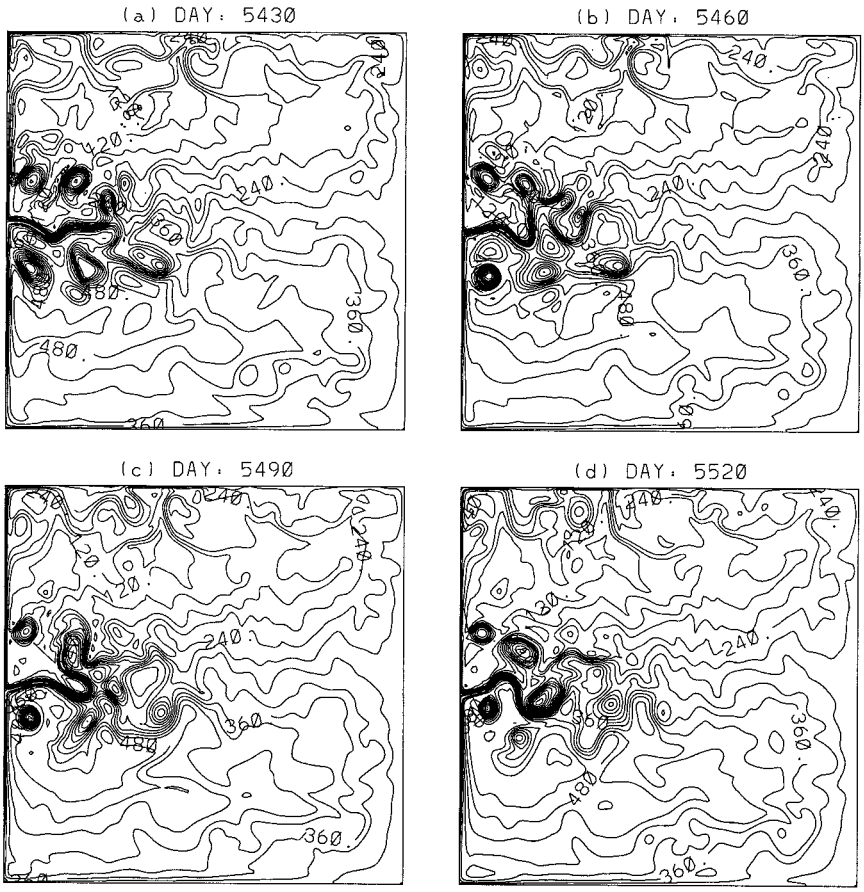

FIG. 2. The instantaneous fields of the upper-layer thickness (contour interval: 30 m) for the 3-month period during which the prediction experiments for the drifter trajectories are conducted. The flow field is highly turbulent and variable.

model dynamics. Examples of Lagrangian statistics in these three regions are shown in Fig. 5. The covariances of the fluctuation velocity $\mathbf{v}'$ are illustrated, where $\mathbf{v}'$ is computed using Eq. (4). The mean flow field $\mathbf{U}$ corresponds to the 2-yr time average shown in Fig. 1. The covariances exhibit a quasi-exponential decay at short timescales, in agreement with the simple first-order model (8) used for $\mathbf{v}'$. The only substantial departure from temporal exponential decay is shown in the meridional component of the western boundary region at lags longer than approximately 6 days. This lack of decay likely is due to strong advection in this region that is not represented by the mean flow. The process of forming climatological Eulerian averages weakens the strength of strong currents (see Mariano 1990).

The assimilation of the Lagrangian data is performed using Eqs. (18)–(21), with the parameters T and R set to the values shown in Table 2. These values are in the range suggested by the Lagrangian and Eulerian statistics in the regions. Preliminary tests performed by varying these values by $\pm 20\%$ have shown that the results of the prediction algorithm are not sensitive to the specific value used in that range. In the practical implementation of the assimilation scheme, only particles inside an influential data region (or a shadow zone) with a radius greater than R around the prediction particle are considered at each time step. The radius of the shadow zone is 100 km in the western boundary current region, and 200 km elsewhere. Predictions are carried out for 100 days in the low-energy interior regions and for about one week in the high-energy western boundary current and midlatitude jet regions. These timescales correspond to the approximate average periods that the particles spend in these regions.

For each prediction experiment, a set of M particles is launched homogeneously. One of the trajectories is considered unknown (predictand), and the other $p = M - 1$ trajectories are considered as known Lagrangian data (predictors). For the purpose of clarification, the true trajectory of the predictand drifter is not used in the assimilation; it is used only for evaluation purposes. The root-mean-square error, s, is computed as the distance between the predicted and the true position of the predictand, and ensemble averages are computed by considering different realizations of the predictand particles ($s(t) = \sqrt{\langle (x_M^a(t) - x_M(t))^2 + (y_M^a(t) - y_M(t))^2 \rangle}$).

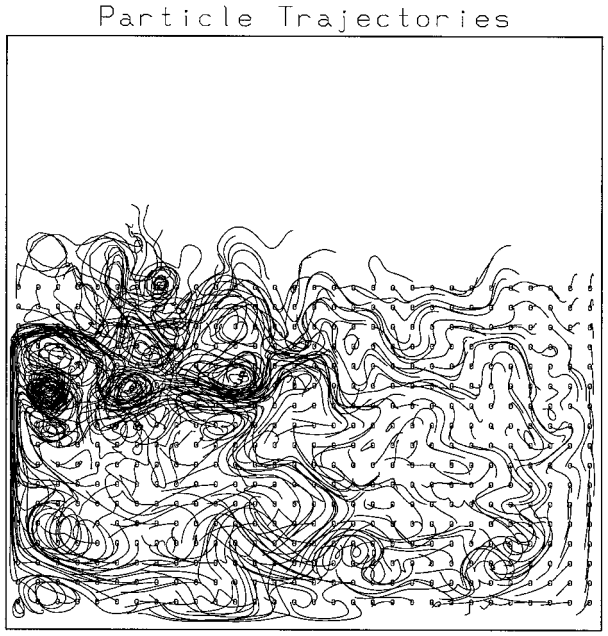

FIG. 3. Particle trajectories during 2 months of integration after being released into the top layer of the numerical ocean model. The particles shown here are launched with a data density of 1 particle/[1° × 1°]. The circles mark the launch positions.

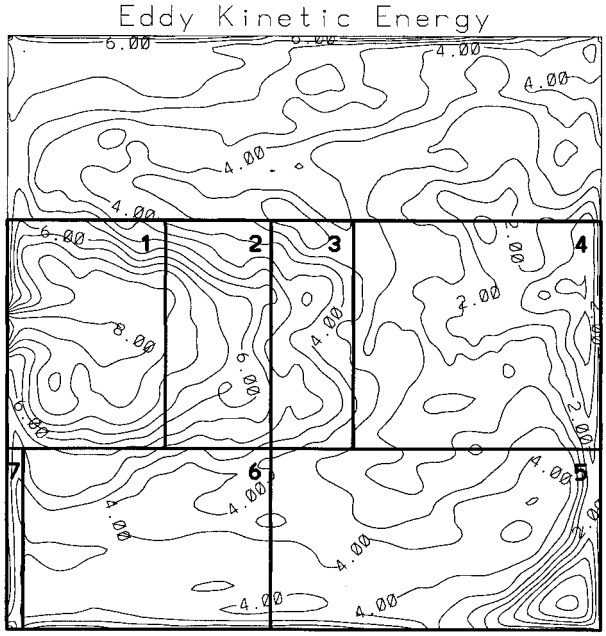

FIG. 4. Quasi-homogeneous regions 1–7 partitioned according to the eddy kinetic energy distribution [EKE = $(u'^2 + v'^2)/2$] in the numerical ocean model. Plotted is ln(EKE), where EKE is in $cm^2 s^{-2}$.

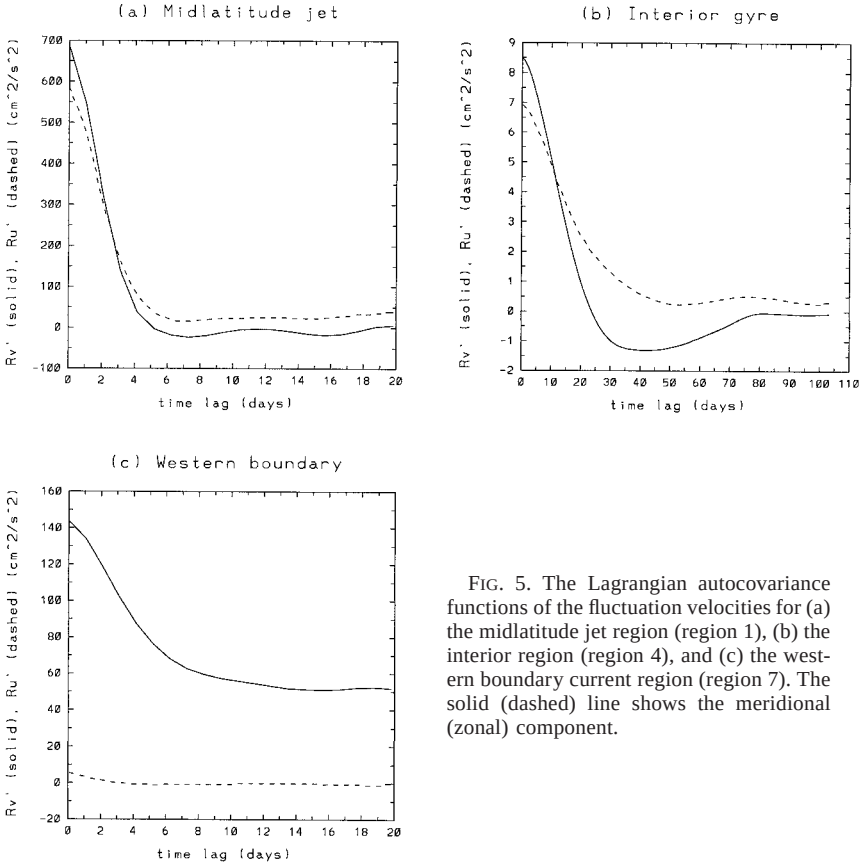

FIG. 5. The Lagrangian autocovariance functions of the fluctuation velocities for (a) the midlatitude jet region (region 1), (b) the interior region (region 4), and (c) the western boundary current region (region 7). The solid (dashed) line shows the meridional (zonal) component.

TABLE 2. Parameters of the prediction model and velocity variances in three dynamically different regions of the ocean model.

Region (Number)	Interior (4)	Western boundary (7)	Midlatitude jet (1)
Eulerian space scale, R (km)	70	70	35
Lagrangian timescale, T (days)	10	10	3
Shadow zone, r (km)	200	100	200
Number of realizations	50	15	50
Prediction period (days)	100	7	7
Velocity variance, σ_u^2 (cm s^{-2})	8	150	800

TABLE 3. Data density vs N_R for different Eulerian space scales R used in the prediction model.

Data density	N_R with $R = 70$ km (interior, western boundary)	N_R with $R = 35$ km (midlatitude jet)
0	0	0
1 particle/[5° × 5°]	0.06	0.015
1 particle/[2° × 2°]	0.39	0.1
1 particle/[1° × 1°]	1.54	0.39
1 particle/[0.5° × 0.5°]	6.16	1.54
1 particle/[0.25° × 0.25°]	24.64	6.16

Typically, 50 realizations are considered in each region except in the western boundary current, in which only 15 realizations are averaged because of its smaller area.

The performance of the prediction scheme is evaluated as function of a number of factors. First, the impact of data density is considered. Six different values for the density are chosen: zero, meaning that no Lagrangian data are available for assimilation, and the prediction is performed via advection by only the mean flow; 1 particle/[5° × 5°], which is a World Ocean Circulation Experiment (WOCE) sampling goal (D. B. Olson 1998, personal communication); and then increasingly higher data densities—1 particle/[2° × 2°], 1 particle/[1° × 1°], 1 particle/[0.5° × 0.5°], and finally 1 particle/[0.25° × 0.25°]. These densities correspond to a total number of approximately 18, 100, 500, 2000, and 8000 buoy releases, respectively, in the subtropical gyre. A more physically relevant measure of data density is N_R, which indicates the number of particles within a distance equal to the Eulerian space scale R from the prediction particle—that is, the number of data per degrees of freedom. The data densities and the corresponding N_R as a function of R are listed in Table 3 as reference for the following discussion of results.

Second, the effects of reducing the knowledge of the mean flow **U** and of the initial conditions are considered. These are important considerations for practical applications since such values are usually known only approximately. Two extreme situations are considered for the mean flow 1) The mean flow **U** is assumed to be the 2-year average shown in Fig. 1; that is, an accurate a priori knowledge of the large-scale climatological flow is assumed 2) The mean flow **U** is assumed to be zero; that is, no a priori knowledge of the large-scale flow is assumed. The impact of initial conditions is studied in a series of cases subject to gradually increasing levels of uncertainty. The first one is the ''ideal'' case, in which choice 1 is used for **U**, and accurate initial turbulent velocities and particle positions are employed. The second one is the case in which choice 2 is used for the mean flow, and no knowledge of the initial turbulent velocity of the predictand is considered. In this case, the initial velocity is evaluated directly from the surrounding particles. Finally, a case is considered in which errors are introduced in the initial position of the predictand.

4. Results

The performance of the prediction scheme is evaluated in the interior, western boundary, and meandering jet regions, testing the impact of varying data density and introducing errors in the knowledge of the mean velocity field and initial conditions. Theoretical error estimates are compared with the numerical results.

a. The interior region

Some examples demonstrating the effectiveness of Lagrangian data assimilation in improving the prediction of particle trajectories are shown in Fig. 6. The solid lines mark the true trajectories of selected particles released in the interior of the subtropical gyre during the 100 days following release. If only the mean flow is used to try to capture the particle motion, the predictions are quite poor (lines with triangles). If, on the other hand, the Lagrangian data assimilation scheme is employed to estimate the turbulent velocity component, the predicted trajectories improve considerably and gradually approach the true trajectories with increasing data density (lines with plus signs and circles).

Region 4, representative of interior regions, is characterized by a Lagrangian timescale $T = 10$ days, an Eulerian space scale $R = 70$ km, and a velocity variance $\sigma_u^2 = 8$ cm^2 s^{-2} (Table 2). Predictions are conducted for 100 days. Ensemble averages of prediction errors are shown in Fig. 7. Here the root-mean-square prediction errors are plotted as a function of time for different data densities and for different cases of uncertainty about the mean flow and initial conditions. Figure 7a illustrates the ideal case for which it is assumed that both mean flow and initial conditions are known accurately. The open circles mark the average distance traveled by the 50 particle realizations after release. Hence, no prediction is made for this case, and it is an upper bound for root-mean-square prediction error. Using the knowledge of a climatological mean flow only (no drifters—solid line), the error is somewhat reduced with respect to the total particle dispersion (i.e., no prediction), but far more substantial reductions in error are possible by employing the surrounding drifter data. The prediction error decreases consistently with increasing data density (an-

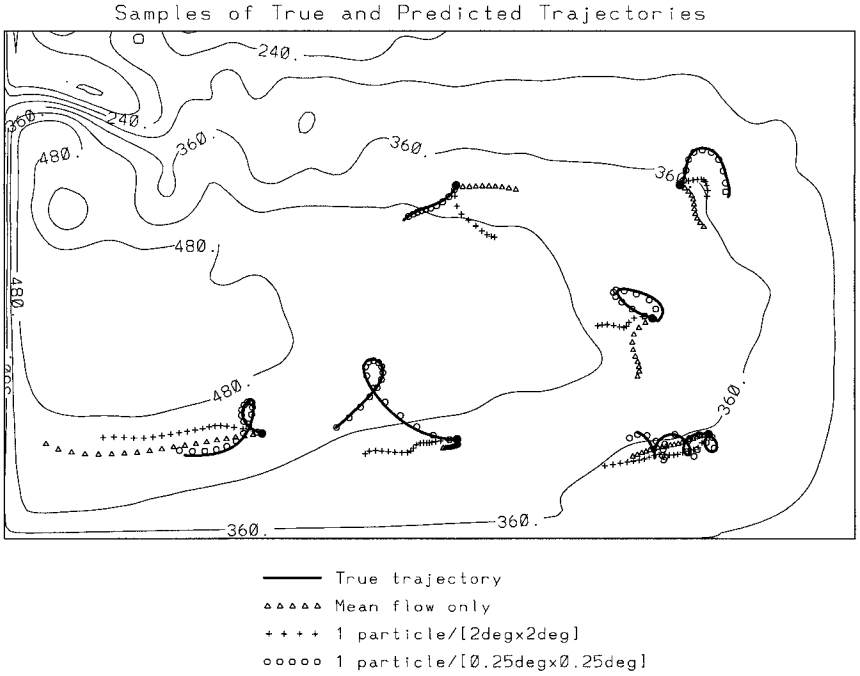

FIG. 6. Comparison of true and predicted trajectories for selected particles released in the gyre interior. Here the solid line indicates the true trajectory, the line composed of triangles indicates the prediction using the mean flow only, and the plus signs and circles show predicted trajectories with data densities of 1 particle/[2° × 2°] and 1 particle/[0.25° × 0.25°], respectively. Note that the accuracy of the predicted trajectories increases with increasing data density. The mean top-layer thickness field is plotted in the background (contour interval: 60 m).

notated lines). Quite accurate predictions are obtained especially for the higher data densities, in which cases the data resolution is of the order of R or higher [implying cases with data densities of 1 particle/[0.5° × 0.5°] ($N_R = 6.16$) and 1 particle/[0.25° × 0.25°] ($N_R = 24.64$); see Table 3].

For instance, the particles cover a distance of about 34 km (134 km) after 1 week (1 month) of integration, and the prediction made using only the mean flow reduces the error to approximately 23 km (90 km), while the error is about 1.5 km (11 km) with the use of the highest density data assimilation employed here (1 particle/[0.25° × 0.25°], $N_R = 24.64$) after the same amount of time.

The qualitative behavior of the prediction error for the ideal case is described well by the theoretical estimates [Eq. (22)]. In Table 4 three cases are shown, for data densities of zero ($N_R = 0$), 1 particle/[1° × 1°] ($N_R = 1.54$), and 1 particle/[0.25° × 0.25°] ($N_R = 24.64$), in which the theoretical rate of error growth is estimated from Eqs. (24), (25), and (26), respectively. Theoretical and numerical estimates of the prediction error are compared at 3, 7, and 30 days. The theoretical estimates show an excellent match to the numerical results for $N_R = 0$, and they compare quite favorably for the higher-density cases as well. Considering the large reduction in error with increasing data density, the theoretical formulas derived in section 2 appear to deliver very reasonable a priori estimates of prediction error as a function of the number of surrounding drifters based on only the main statistical parameters of the flow.

Figure 7b depicts the case in which no knowledge of the mean flow is assumed—that is, $\mathbf{U} = 0$ in Eqs. (4) and (21)—while the initial conditions are assumed to be known accurately. In this case, the error curve of the prediction made using only the mean flow coincides with the line that indicates the total dispersion of particles. The lack of mean-flow knowledge affects primarily those cases with low-density assimilation. As the data density increases, the prediction error decreases consistently, and the prediction capabilities of the cases employing data densities higher than or equal to 1 particle/[1° × 1°] (or for $N_R > 1$) remain nearly unchanged. Therefore, this graph indicates that the impact of the lack of knowledge of the mean flow can be compensated for by assimilating information from the surrounding drifters.

Next, a case is considered in which in addition to $\mathbf{U} = 0$ the initial turbulent velocity of the predictand is assumed also to be unknown and is estimated from the surrounding drifters. The results are shown in Fig. 7c. The greatest impact on prediction error is observed for low-data-density cases for the first few weeks. The cases with data density equal to or higher than 1 particle/[0.5°

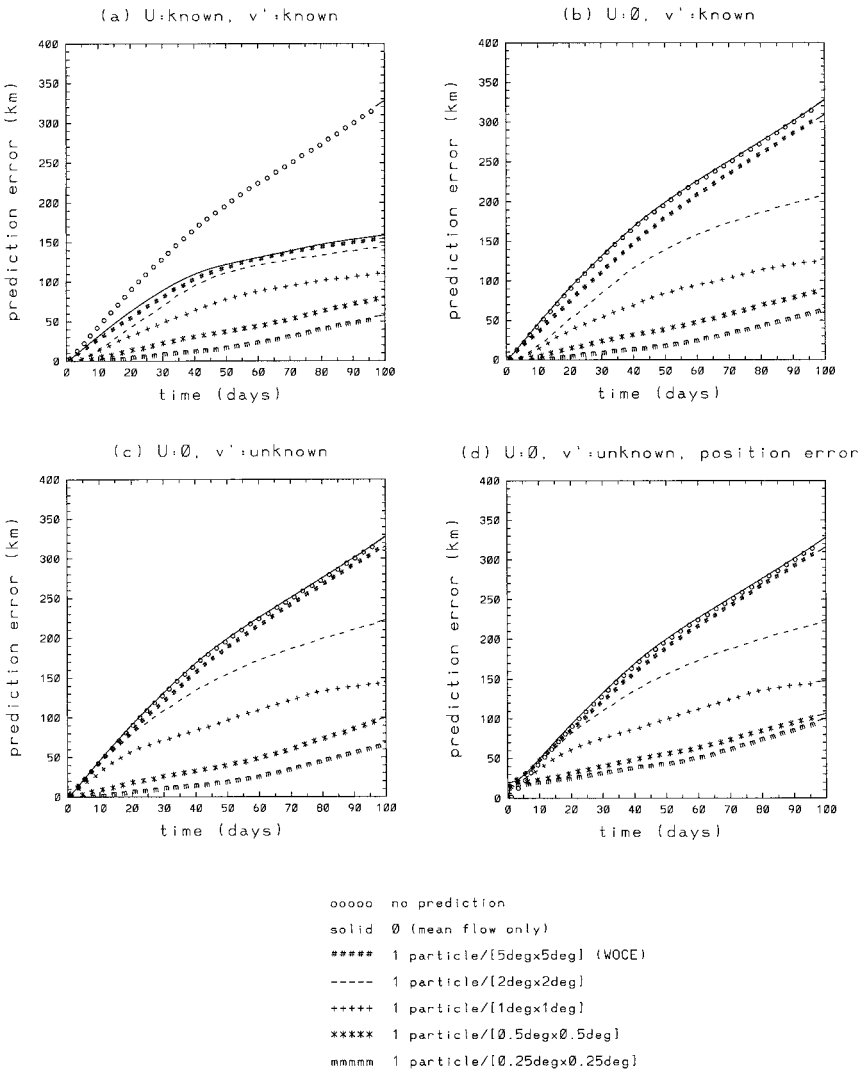

FIG. 7. Prediction (rms) error vs time as a function of data density in the gyre interior (region 4) for (a) the ideal case, with accurate mean flow and accurate initial conditions for the particle—that is, known initial position and turbulent velocity; (b) the case with zero mean flow and accurate initial conditions for the particle; (c) the case with zero mean flow and unknown initial turbulent velocities (estimated from the surrounding data); and (d) the case with zero mean flow, unknown initial turbulent velocities, and an initial position error of 20 km.

$\times$ 0.5°] (N_R = 6.16) remain nearly unchanged since these cases allow a good estimation of the initial turbulent velocity for the prediction particle. This result indicates that uncertainties regarding both the mean flow and initial turbulent velocities can be overcome with the use of the surrounding drifter data.

Finally, a further source of uncertainty is incorporated by considering that the initial position of the predictand

TABLE 4. Comparison of numerical results to theoretical error estimates in the interior region (region 4) for N_R corresponding to data densities of zero, 1 particle/[1° $\times$ 1°], and 1 particle/[0.25° $\times$ 0.25°] at 3 days, 1 week, and 1 month after release, respectively.

Error (km) vs time (days)	Upper bound	N_R = 0		N_R = 1.54		N_R = 24.64	
		Theoretical	Numerical	Theoretical	Numerical	Theoretical	Numerical
t = 3	14.9	7.8	9.7	2.6	1.8	1.1	0.4
t = 7	33.9	18.2	22.9	6.0	7.8	2.5	1.5
t = 30	133.9	78.0	89.9	25.7	50.0	10.6	11.3

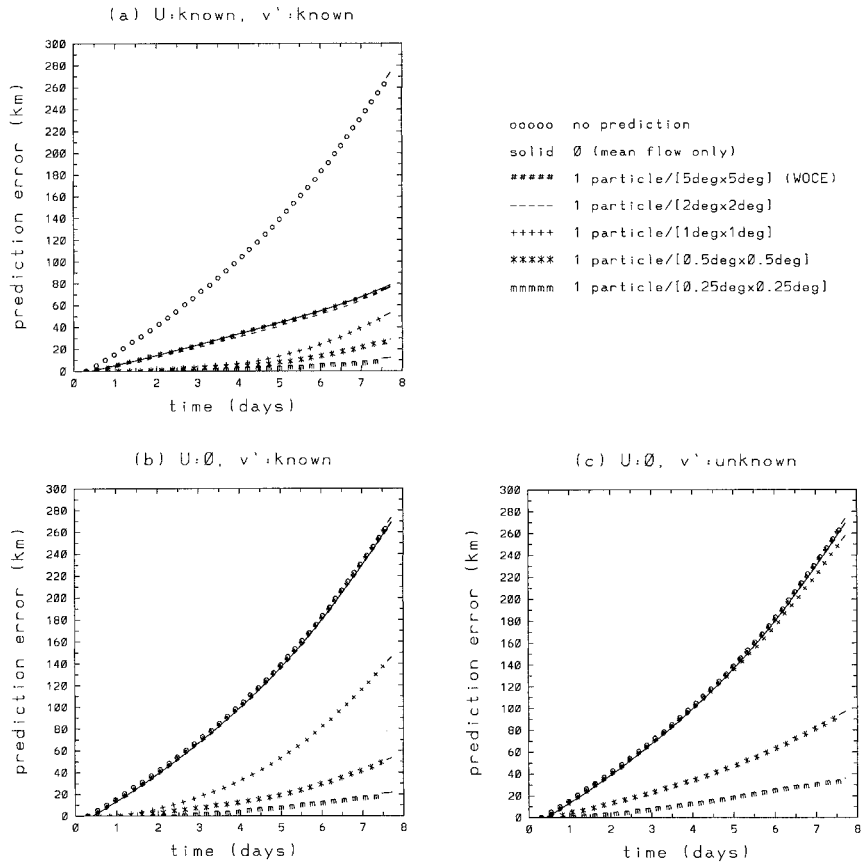

FIG. 8. Prediction (rms) error vs time as a function of data density in the western boundary current region (region 7) for (a) the ideal case with accurate mean flow and accurate initial conditions for the particle, (b) the case with zero mean flow and accurate initial conditions for the particle, and (c) the case with zero mean flow and unknown initial turbulent velocities (estimated from the surrounding data).

is known with an error Δr. Figure 7d shows the error curves for the average of four cases, with $\Delta r = 20$ km in the eastward, westward, northward, and southward directions. Comparing Figs. 7d and 7b, it appears that the effect of the initial error Δr for densities lower than 1 particle/[1° × 1°] ($N_R = 1.54$) is primarily on short timescales (first few days). From Fig. 7b it can be seen that an error of 20 km is reached after only a few days for low-density cases, but is reached on much longer timescales for the high-density cases. Since the initial position error cannot be recovered, its impact on higher-density predictions is more persistent and substantial. Considering also that the error curves of the two highest density cases almost coincide (Fig. 7d), this experiment indicates that there is a diminishing gain from using higher data densities in the case of an initial position error.

Finally, it should be noted from Fig. 7 that there is no significant improvement in predictive capability when a data density according to the WOCE requirement is employed with respect to the prediction when there is only the knowledge of the mean flow.

b. The western boundary current region

The western boundary current region (region 7) is characterized by a Lagrangian timescale $T = 10$ days, an Eulerian space scale $R = 70$ km, and a velocity variance $\sigma_u^2 = 150$ cm^2 s^{-2} (Table 2). The prediction errors are calculated for about 1 week, which corresponds approximately to T. The results are depicted in Fig. 8. In the ideal case (Fig. 8a) even the prediction using only the mean flow leads to significant improvement, indicating the importance of knowledge of the mean flow in this region. Predictions with data densities higher than or equal to 1 particle/[1° × 1°] ($N_R = 1.54$) yield very accurate results and exhibit almost no error for the first 2–3 days. For instance, the particles cover a distance of about 82 km (250 km) after 3 days (7 days) of integration, and the prediction made using only the

TABLE 5. Comparison of numerical results to theoretical error estimates in the western boundary region (region 7) for N_R corresponding to data densities of zero, 1 particle/[1° × 1°], and 1 particle/[0.25° × 0.25°] at 3 days and 1 week after release, respectively.

Error (km) vs time (days)	Upper bound	$N_R = 0$		$N_R = 1.54$		$N_R = 24.64$	
		Theoretical	Numerical	Theoretical	Numerical	Theoretical	Numerical
$t = 3$	81.9	31.8	27.1	10.5	6.2	4.3	3.0
$t = 7$	249.9	74.2	71.8	24.5	45.1	10.1	10.9

mean flow reduces the error to approximately 27 km (72 km), while the error is about 3 km (11 km) with the use of the highest density data assimilation employed here (1 particle/[0.25° × 0.25°], $N_R = 24.64$) after the same amount of time.

When the mean flow field is not known (Fig. 8b), there is no significant improvement in prediction capability for a data density less than 1 particle/[1° × 1°], showing the importance of the knowledge of the mean flow in this region. The predictions conducted with higher data densities remain approximately the same. High-data-density cases do quite well even when the initial turbulent velocities are not known accurately (Fig. 8c). This result appears to support the ability of the Lagrangian data assimilation algorithm to compensate for lack of knowledge of the mean flow and initial velocities from the drifter data alone, not only in the low-energy interior regions but also in the western boundary current region characterized by a strong mean flow and fluctuations, provided that a sufficiently high number of surrounding drifters are available. A priori estimates regarding the number of drifters can be obtained using the theoretical formulas (24)–(26). Such estimates compare favorably with the numerical results in the western boundary region as well (Table 5).

c. The midlatitude jet region

The midlatitude jet region (region 1) is characterized by a Lagrangian timescale $T = 3$ days, an Eulerian space scale $R = 35$ km, and a velocity variance $\sigma_u^2 = 800$ cm^2 s^{-2} (Table 2). This region is representative of the Gulf Stream ring and meander region. The midlatitude jet region is the most difficult one to tackle because of the high degree of variability and the existence of different flow regimes such as the jet, mesoscale eddies, and wind-driven circulation, all of which interact with each other. Consequently, the behavior of the particles is highly complicated, and frequently the particles are entrained from one flow regime into another. As a result, the variability in the predictions is higher. It can be observed that the real and predicted trajectories bifurcate while passing through high-gradient regions and end up in drastically different positions in a matter of a few days.

Prediction results for region 1 computed during 7 days are shown in Fig. 9. In the ideal case (Fig. 9a) the error plots indicate that good predictive capability is limited to the high-data-density cases (1 particle/[0.5° × 0.5°], $N_R = 1.54$, and 1 particle/[0.25° × 0.25°], $N_R = 6.16$) and is confined to the initial prediction times ($t < 4$ days $\approx T$). For instance, the prediction error can be reduced to about 15 km (67 km) using the highest data density of 1 particle/[0.25° × 0.25°] ($N_R = 6.16$) from a potential maximum of about 119 km (212 km) after 3 days (7 days), whereas using only the mean flow knowledge, it reduces to only approximately 92 km (178 km). The error increases approximately linearly in time with a slope that decreases with N_R. The theoretical estimates are found to be appropriate, as shown in Table 6. As can be seen from Tables 4–6, the error is larger in this region than are the errors in the interior and western boundary regions, primarily because the turbulent velocity variance σ_u^2 is larger and R is smaller, so that the same geographical data distribution results in a smaller N_R: for example, 6.16 versus 24.64 as in other cases for the highest data density of 1 particle/[0.25° × 0.25°]. For longer times, $t > T$, the error is found to increase further, with a greater slope than that predicted by Eqs. (24)–(26). This is likely because the region is characterized by a strong front dividing two different regimes, so that the dependence of the prediction error on the positions of the predictand particles and the surrounding ones is more fragile. In other words, the region is highly chaotic, and the present simple prediction scheme, which assumes statistical homogeneity, does not hold for prediction times longer than the Lagrangian integral timescale T.

When the mean flow and/or the initial turbulent velocities are not known accurately, the results obtained in the context of the other regions remain valid in the meandering region as well. The low-density cases deteriorate, whereas the high-density cases—1 particle/[0.5° × 0.5°] and 1 particle/[0.25° × 0.25°]—are not affected significantly (not shown).

A final experiment has been conducted considering that the climatological 2-year-averaged mean velocity field (Fig. 1) may not be a good estimate with which to advect a particle for approximately 1 week in a region characterized by a very strong eddy field. The 2-year climatological mean field is very smooth and almost completely lacks the signatures of the jet and the eddies. As a result, particle trajectories calculated using low-data-density assimilations (i.e., subject to stronger steering by the mean field) can cross the high-gradient zones of the jet and eddies, leading to high errors. In this

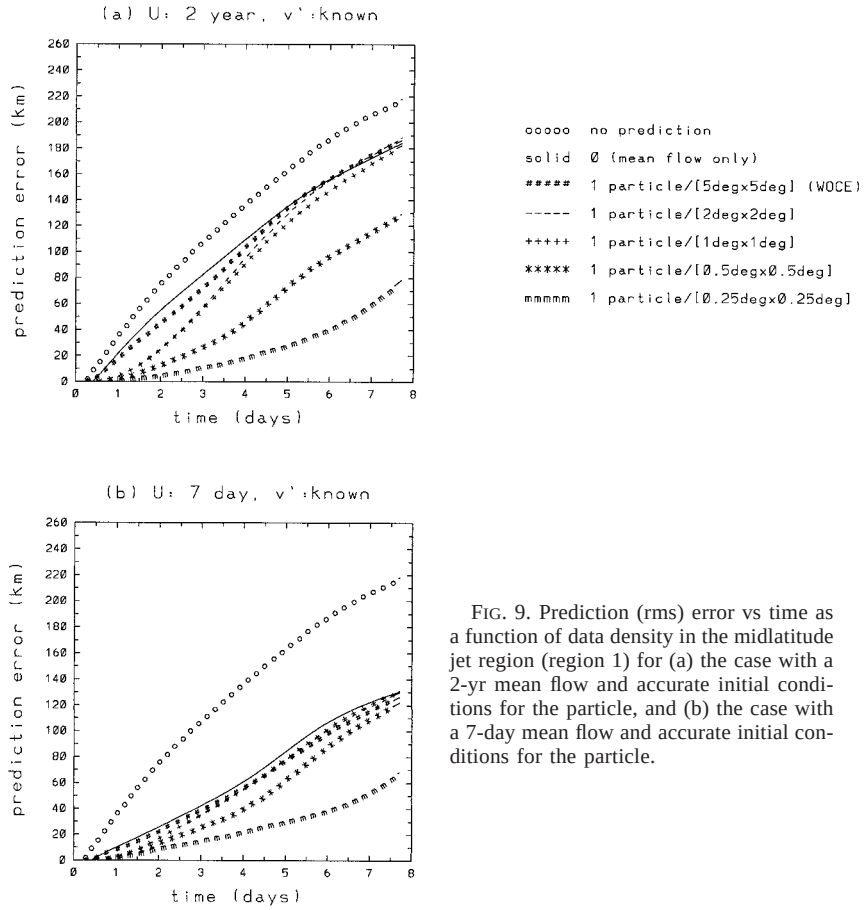

FIG. 9. Prediction (rms) error vs time as a function of data density in the midlatitude jet region (region 1) for (a) the case with a 2-yr mean flow and accurate initial conditions for the particle, and (b) the case with a 7-day mean flow and accurate initial conditions for the particle.

experiment, we assume that additional short-term information about the large-scale flow field is available from (e.g., satellite) observations, so that the ''mean'' field **U** used in Eqs. (18)–(21) is an average over a shorter time period. We have considered an average during 7 days, coinciding with the duration of prediction. In this case, **U** resembles the instantaneous field shown in Fig. 2a. The results are displayed in Fig. 9b, which indicates major improvement in the performance of low-density cases and little change for the high-density cases. Schneider (1998) investigated the use of different mean flows and the same stochastic model as ours for predicting surface drifters in the Gulf Stream region. The average prediction error for the best mean field case is found to be 37 km after 2 days. This case is somewhere between the cases plotted in Figs. 9a and 9b. There appears to be a reasonable agreement between our error estimates, approximately 20–50 km, from a model and the error estimate of 37 km for real data. These cases can be improved upon in the future by using a more complex algorithm that selects dynamically consistent information based on the knowledge of **U** or by using better estimates of **U** employing contour-based techniques (Mariano 1990). For example, knowledge of the jet position can be used to select drifters in the same frontal region as the predictand. This may lead to improvements in the midlatitude jet region characterized by strong gradients of the flow field.

5. Summary and concluding remarks

Predictability of particle trajectories in oceanic flows is explored for timescales on the order of a week to a

TABLE 6. Comparison of numerical results to theoretical error estimates in the midlatitude jet region (Region 1) for N_R corresponding to data densities of zero, 1 particle/[0.5° × 0.5°], and 1 particle/[0.25° × 0.25°], at 3 days and 1 week after release, respectively.

Error (km) vs time (days)	Upper bound	$N_R = 0$		$N_R = 1.54$		$N_R = 6.16$	
		Theoretical	Numerical	Theoretical	Numerical	Theoretical	Numerical
$t = 3$	119.0	73.2	91.6	24.2	31.9	21.5	14.6
$t = 7$	212.2	170.8	177.5	56.4	120.6	50.1	67.4

few months. This is a well-recognized, outstanding conceptual problem because errors accumulate in trajectory predictions and Lagrangian motion tends to be chaotic even when the Eulerian flow is smooth and regular. It is also relevant for practical applications, such as searching for objects lost at sea and ecological problems, such as the spreading of pollutants or fish larvae. The original aspect of this study is the use of surrounding velocity data provided by Lagrangian instruments to improve the accuracy of the predicted particle trajectories. The investigation is conducted in the context of a primitive equation isopycnic-coordinate numerical ocean model. Although the ocean model is idealized using a double-gyre, three-layer box configuration, it reproduces dynamically distinct regions representative of those in the real midlatitude ocean.

It is assumed that some information is known a priori about the large-scale slowly varying mean flow, while information about the mesoscale eddies is provided entirely by the Lagrangian data. The prediction is performed by assimilating the surrounding velocity data into a simplified equation that describes the particle motion. In this equation, the velocity field is decomposed into two components: the mean field **U** and the eddy field. The eddy velocity following particles is assumed to be a linear Markov process. That is, described by a first-order autoregressive process, AR(1), with an exponential covariance with a timescale T. This assumption appears to be acceptable for oceanic surface flows for timescales T on the order of 3–10 days. The assimilation is performed using Kalman filtering on the AR(1) process. The error covariance matrix is related in a very simple way to the Eulerian statistics, which are assumed to be known.

Theoretical estimates of the prediction error s of particle position are derived under some simplifying assumptions. These derivations demonstrate that s grows in time as in single particle dispersion multiplied by a factor that depends on the number of data per degrees of freedom N_R [Eq. (22), (24)–(26)].

Ensemble averages of prediction errors are calculated using numerical experiments in three fundamental regions—the subtropical gyre interior, the western boundary current, and the midlatitude jet regions—as a function of varying data density and degrees of uncertainty about the mean flow and initial conditions of the predictant particle. The results show that the prediction error increases in time but decreases consistently with increasing data density. The simulations indicate that the actual WOCE sampling (1 particle/[5° × 5°] or $N_R \ll 1$) does not improve particle prediction, but predictions improve significantly when $N_R \gg 1$. For instance, a coverage of 1 particle/[1° × 1°] or $N_R \sim O(1)$ is already able to reduce the errors of about one-third or one-half. If the sampling resolution is increased to 1 particle/[0.5° × 0.5°] or 1 particle/[0.25° × 0.25°] or $N_R \gg 1$, reasonably accurate predictions (rms errors of less than 50 km, can be obtained for periods ranging from 1 week (western boundary current and midlatitude jet regions) to 3 months (interior gyre region). Even if the mean flow field and initial turbulent velocities are not known accurately, the information derived from the surrounding drifter data can compensate, provided that $N_R > 1$. The theoretical estimates for prediction error are generally in good agreement with the numerical results, especially considering the drastic reduction in error with increasing data density, in all regions for initial times. Since these a priori estimates are based on only a few parameters that are generally known, they can be used as guidance in practical situations to assess error as a function of resources to be deployed. To summarize, this study improves oceanographer's ability to perform Lagrangian prediction, and at the same time, improves our understanding of the predictability limits in the real ocean.

The assimilation algorithm appears to be quite effective, in spite of its simplicity. For practical applications, improvements are foreseen, potentially involving relaxation of assumptions such as Gaussian Eulerian statistics, independence of velocity components, and isotropy. For complex regions exhibiting strong fronts such as a meandering jet, the Kalman filter technique can be generalized for the complete trajectory equation in order to obtain more accurate estimates. Also, since the large-scale field varies rapidly in these regions, the addition of other simultaneous datasets, such as satellite information about the instantaneous large-scale field, may help to improve the predictions.

As a last remark, we point out that a homogeneous sampling of drifters has been considered in this paper. This choice of sampling allows us to explore the role of data density and assimilation techniques while taking into account the Lagrangian nature of the data both in the formulation and in the error computation that assumes a homogeneous data distribution in space. For practical problems, however, such as tracking and predicting pollutants or biological organisms, as well as searching for objects lost at sea, different types of sampling can be envisioned to best exploit the specific properties of Lagrangian data, such as their tendency to move with the current. For example, an initial drifter cluster can be released around the initial position of the predictand particle, and additional drifters can then be released at later times to maintain the desired density around the prediction. This may allow minimization of the number of drifters necessary to obtain a preset prediction error. Sampling strategies will be studied in a future work, using the results presented herein as guidelines.

Acknowledgments. The authors greatly appreciate the support of the Office of Naval Research under Grants N00014-97-1-0620 and N00014-99-1-0049 (T. M. Özgökmen and A. Griffa), N00014-99-1-0042 (L. I. Pi-

terbarg), and N00014-95-1-0257 and N00014-99-1-0049 (A. J. Mariano). L. I. Piterbarg also acknowledges the support of the National Science Foundation Grant DMS-9709320.

APPENDIX

a. Relation between the noise covariance and the Eulerian covariance

Let us introduce the conditional Lagrangian covariances

$$\mathbf{B}^{\mathbf{v}}_{ij}(n) = E_n\{\mathbf{v}'_i(n)\mathbf{v}'_j(n)^{\mathrm{T}}\}$$

and set

$$\mathbf{B}^{\mathbf{v}}(n) = (\mathbf{B}^{\mathbf{v}}_{ij}(n)).$$

Thus, $\mathbf{B}^{\mathbf{v}}(n)$ is $2M \times 2M$ Lagrangian covariance matrix. From Eq. (10) one concludes that

$$\mathbf{Q}(n) = (\mathbf{I}_{2M} - \mathbf{A})\mathbf{B}^{\mathbf{v}}(n)(\mathbf{I}_{2M} - \mathbf{A})^{\mathrm{T}}.$$

Then from Eq. (7) it follows that

$$\mathbf{Q}(n) = (\mathbf{I}_{2M} - \mathbf{A})(\mathbf{B}^{\mathbf{u}}_n(\mathbf{r}_i(n), \mathbf{r}_j(n)))(\mathbf{I}_{2M} - \mathbf{A})^{\mathrm{T}},$$

where $\mathbf{B}^{\mathbf{u}}_n(\mathbf{r}_1, \mathbf{r}_2)$ is the Eulerian space covariance at time n defined in Eq. (2). Therefore, we have a very simple relation between the matrix $\mathbf{Q}$ playing a key role in the Kalman filtering and the Eulerian covariance matrices. This relation is based on the assumptions given in Eqs. (6) and (7).

b. Simplification of the prediction formulas

Let us represent the basic matrices in the block form:

$$\mathbf{P}^f = \begin{pmatrix} \mathbf{P}^f_{11} & \mathbf{P}^f_{12} \\ \mathbf{P}^f_{21} & \mathbf{P}^f_{22} \end{pmatrix}, \qquad \mathbf{P}^a = \begin{pmatrix} \mathbf{P}^a_{11} & \mathbf{P}^a_{12} \\ \mathbf{P}^a_{21} & \mathbf{P}^a_{22} \end{pmatrix},$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

where the upper-left, upper-right, lower-left, and lower-right matrices have dimensions $2p \times 2p$, $2p \times 2$, $2 \times 2p$, and 2×2, respectively. Assuming that $\varepsilon = 0$, after simple computations we get

$$\mathbf{K} = \begin{pmatrix} \mathbf{I}_{2p} \\ \mathbf{P}^f_{21}(\mathbf{P}^f_{11})^{-1} \end{pmatrix},$$

$$\mathbf{P}^a = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}^f_{22} - \mathbf{P}^f_{21}(\mathbf{P}^f_{11})^{-1}\mathbf{P}^f_{12} \end{pmatrix},$$

$$\mathbf{P}^f(n+1) = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} + \mathbf{A}_{22}\mathbf{P}^a_{22}(n)\mathbf{A}^{\mathrm{T}}_{22} \end{pmatrix}.$$

c. Error analysis

For the theoretical error analysis, we assume that the velocity components are uncorrelated and the fluctuation velocity field is statistically homogeneous: that is, $\mathbf{B}^{\mathbf{u}}_n(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{B}^{\mathbf{u}}_n(\mathbf{r}_1 - \mathbf{r}_2)$. We then immediately obtain from Eqs. (16) and (18) that the mean-square error for an individual component—zonal, for instance—is given by

$$s(n) = (1 - \alpha^2)(q(n) + \alpha^2 q(n-1) + \cdots + \alpha^{2(n-1)}q(1)), \tag{A1}$$

where

$$q(n) = B(\mathbf{0}) - \sum_{i,j=1}^{p} B(\mathbf{r}_M - \mathbf{r}_i)B^{-1}(\mathbf{r}_i - \mathbf{r}_j)B(\mathbf{r}_M - \mathbf{r}_j), \tag{A2}$$

B is the covariance function for the individual component, and $B^{-1}(\mathbf{r}_i - \mathbf{r}_j)$ is the (i, j) entry of the matrix inverse to $(B(\mathbf{r}_i - \mathbf{r}_j))$. By averaging Eq. (A1) through the ensemble, we obtain

$$\langle s(n)\rangle = (1 - \alpha^2)(\langle q(n)\rangle + \alpha^2\langle q(n-1)\rangle + \cdots + \alpha^{2(n-1)}\langle q(1)\rangle).$$

Under the stationarity assumption, $\langle q(k)\rangle$ is independent of k, and hence,

$$\langle s(n)\rangle = \langle q\rangle(1 - \alpha^{2n}).$$

Now we estimate $\langle q\rangle$ assuming a high density of control particles. Let us introduce the correlation velocity scale $R = \sqrt{-B(\mathbf{0})/B''(\mathbf{0})}$ and assume that the set of all particles form a Poisson process with intensity λ. We assume that

$$\lambda\pi R^2 \gg 1$$

or, equivalently,

$$N_R \gg 1,$$

where N_R is the mean number of particles in a disk of radius R. Let us take an arbitrary δ such that

$$\delta \ll 1, \tag{A3}$$

but

$$\lambda\pi\delta^2 R^2 \gg 1; \tag{A4}$$

that is, δ is small. However, the number of particles N_δ in a disk O_δ of radius $r = \delta R$ is still large. Now we take into account only particles that are inside the disk O_δ centered at the predictand. Obviously, $q = B(0)$ if there are no predictors in O_δ. Otherwise, we can take a Taylor expansion in Eq. (A2):

$$q \sim B(0) - \sum_{i,j=1}^{N_\delta} \left(B(0) + \frac{1}{2}B''(0)(\mathbf{r}_M - \mathbf{r}_i)^2 \right) \times \left(B(0) + \frac{1}{2}B''(0)(\mathbf{r}_i - \mathbf{r}_j)^2 \right)^{-1} \times \left(B(0) + \frac{1}{2}B''(0)(\mathbf{r}_M - \mathbf{r}_j)^2 \right). \quad \text{(A5)}$$

Using lemma 2.1 from Piterbarg (1998), we obtain

$$q \sim -B''(0)\left(\frac{1}{N_\delta} \sum_{i=1}^{N_\delta} (\mathbf{r}_M - \mathbf{r}_i)^2 - \frac{1}{2N_\delta^2} \sum_{i,j=1}^{N_\delta} (\mathbf{r}_i - \mathbf{r}_j)^2 \right). \quad \text{(A6)}$$

Hence,

$$\langle q \rangle = -B''(0)\langle (\mathbf{r}_i - \mathbf{r}_j)^2 \rangle \Pr\{N_\delta > 0\} + B(0) \Pr\{N_\delta = 0\},$$

where the angle brackets on the right-hand side mean the conditional expectation under the given number of particles. Given that the particles are distributed uniformly throughout the disk, we obtain

$$\langle (\mathbf{r}_i - \mathbf{r}_j)^2 \rangle = \frac{r^2}{9}.$$

Applying

$$\Pr\{N_\delta = 0\} = \exp\{-\langle N_\delta \rangle\},$$

we obtain after some algebra:

$$\gamma^2 \equiv \langle q \rangle = B(0)\left(\frac{\delta^2}{12}(1 - \exp\{-N_R\delta^2\}) + \exp\{-N_R\delta^2\} \right).$$

Now recall that δ is arbitrary, satisfying Eqs. (A3)–(A4). Let us take

$$\delta^2 = \frac{1.5 \ln N_R}{N_R},$$

and we immediately arrive at Eq. (26).

Using the independence of the velocity estimation errors on different steps and assuming a constant mean velocity field, we obtain for the position prediction error $s_x^2 = \langle (\hat{x}_M(t^O) - x_M(t^O))^2 \rangle$ in the x direction, where t^O is the observation time

$$s_x \sim \gamma t^O, \quad \text{if } t^O \ll T, \quad \text{and} \quad \text{(22)}$$

$$s_x \sim \gamma\sqrt{t^O T}, \quad \text{if } t^O \gg T. \quad \text{(23)}$$

It is also of practical interest to consider the case of a small number of predictors. If there are no particles in the circle O_R, then obviously

$$\gamma^2 = \sigma_u^2. \quad \text{(24)}$$

If there is only one predictor, then assuming that its position and the predictand's position are independent and uniformly distributed over O_R, we obtain in the isotropic case

$$\gamma^2 = \frac{4\sigma_u^2}{R^2} \int_0^R \int_0^R \left(B(0) - \frac{B(r_1 - r_2)^2}{B(0)} \right) r_1 r_2 \, dr_1 \, dr_2.$$

In particular, for $B(r) = \exp\{-r^2/2R^2\}$, we have Eq. (25).

REFERENCES

Aref, H., 1984: Stirring by chaotic advection. *J. Fluid Mech.,* **143,** 1–21.

Bauer, S., M. S. Swenson, A. Griffa, A. Mariano, and K. Owens, 1998: Eddy-mean flow decomposition and eddy-diffusivity estimates in the tropical pacific ocean. Part 1: Methodology. *J. Geophys. Res.*, **103,** 30 855–30 871.

Bleck, R., and D. B. Boudra, 1986: Wind-driven spin up in eddy-resolving ocean models formulated in isopycnic and isobaric coordinates. *J. Geophys. Res.,* **91,** 7611–7621.

——, and E. P. Chassignet, 1994: Simulating the oceanic circulation with isopycnic coordinate models. *The Oceans: Physiochemical Dynamics and Resources,* The Pennsylvania Academy of Science, 17–39.

Böning, C. W., and M. D. Cox, 1988: Particle dispersion and mixing of conservative properties in an eddy-resolving model. *J. Phys. Oceanogr.,* **18,** 320–338.

Box, G. E. P., and G. M. Jenkins, 1976: *Time Series Analysis: Forecasting and Control.* 2d ed. Holden-Day, 575 pp.

Bril, G., 1995: Forecasting hurricane tracks using the Kalman filter. *Environmetrics,* **6,** 7–16.

Carter, E. F., 1989: Assimilation of Lagrangian data into a numerical model. *Dyn. Atmos. Oceans,* **13,** 335–348.

Chassignet, E. P., and P. R. Gent, 1991: The influence of boundary conditions on midlatitude jet separation in ocean numerical models. *J. Phys. Oceanogr.,* **21,** 1290–1299.

Colin de Verdiere, A., 1983: Lagrangian eddy statistics from surface drifters in the eastern North Atlantic. *J. Mar. Res.,* **41,** 375–398.

Davis, R. E., 1985: Drifter observations of coastal surface currents during CODE: The statistical and dynamical views. *J. Geophys. Res.,* **90,** 4756–4772.

——, 1991a: Lagrangian ocean studies. *Annu. Rev. Fluid Mech.,* **23,** 43–64.

——, 1991b: Observing the general circulation with floats. *Deep-Sea Res.,* **38,** S531–S571.

Duan, J., and S. Wiggins, 1996: Fluid exchange across a meadering jet with quasiperiodic variability. *J. Phys. Oceanogr.,* **26,** 1176–1188.

Dutkiewicz, S., A. Griffa, and D. B. Olson, 1993: Particle diffusion in a meandering jet. *J. Geophys. Res.,* **98,** 16 487–16 500.

Figueroa, H. A., and D. B. Olson, 1989: Lagrangian statistics in the South Atlantic as derived from SOS and FGGE drifters. *J. Mar. Res.,* **47,** 525–546.

——, and ——, 1994: Eddy resolution versus eddy diffusion in a double gyre GCM. Part I: The Lagrangian and Eulerian description. *J. Phys. Oceanogr.,* **24,** 371–386.

Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances in Geophysics,* Vol. 33, Academic Press, 141–226.

Griffa, A., 1996: Applications of stochastic particle models to oceanographic problems. *Stochastic Modelling in Physical Oceanography,* R. J. Adler, P. Müller, and B. L. Rozovoskii, Eds., Birkhäuser, 114–140.

——, K. Owens, L. Piterbarg, and B. Rozovskii, 1995: Estimates of turbulence parameters from Lagrangian data using a stochastic particle model. *J. Mar. Res.,* **53,** 371–401.

Holland, W. R., 1978: The role of mesoscale eddies in the general circulation of the ocean. *J. Phys. Oceanogr.,* **8,** 363–392.

Ide, K., and M. Ghil, 1998: Extended Kalman filtering for vortex systems. Part I: Methodology and point vortices. *Dyn. Atmos. Oceans,* **27,** 301–332.

Ishikawa, Y. I., T. Awaji, and K. Akimoto, 1996: Successive correction of the mean sea surface height by the simultaneous assimilation of drifting buoy and altimetric data. *J. Phys. Oceanogr.,* **26,** 2381–2397.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.

Mariano, A. J., 1990: Contour analysis: A new approach for melding geophysical fields. *J. Atmos. Oceanic Technol.,* **7**, 285–295.

Meyers, S., 1994: Cross-frontal mixing in a meandering jet. *J. Phys. Oceanogr.,* **24,** 1641–1646.

Miller, R. N., 1989: Direct assimilation of altimetric differences using the Kalman filter. *DAO,* **13,** 317–334.

Okamoto, H., M. Kubota, and A. Shikauchi, 1994: Simulation and data assimilation of drifting buoy trajectories. *J. Fac. Mar. Sci. Technol.,* **37,** 13–28.

Piterbarg, L. I., 1998: Drift estimation for Brownian flows. *Stochastic Processes Appl.,* **79,** 132–149.

Rossby, H. T., S. C. Riser, and A. J. Mariano, 1983: The Western North Atlantic—A Lagrangian viewpoint. *Eddies in Marine Science,* A. R. Robinson, Ed., Springer-Verlag, 66–91.

Samelson, R. M., 1992: Fluid exchange across a meandering jet. *J. Phys. Oceanogr.,* **22,** 431–440.

——, 1996: Chaotic transport by mesoscale motions. *Stochastic Modelling in Physical Oceanography,* R. J. Adler, P. Müller, and B. L. Rozovoskii, Eds., Birkhäuser, 423–438.

Schneider, T., 1998: Lagrangian drifter models as search and rescue tools. M.S. thesis, Dept. of Meteorology and Physical Oceanography, University of Miami, 96 pp.

Stommel, H., 1949: Horizontal diffusion due to oceanic turbulence. *J. Mar. Res.,* **8,** 199–225.

Swenson, M. S., and P. P. Niiler, 1996: Statistical analysis of the surface circulation of the California Current. *J. Geophys. Res.,* **101,** 22 631–22 645.

Taylor, G. I., 1921: Diffusion by continuous movements. *Proc. London Math. Soc.,* **20,** 373–416.

Thomson, D. J., 1987: Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *J. Fluid Mech.,* **180,** 529–556.

Yang, H., and Z. Liu, 1997: The three-dimensional chaotic transport and the great ocean barrier. *J. Phys. Oceanogr.,* **27,** 1258–1273.

Yeung, P. K., and S. B. Pope, 1989: Lagrangian statistics from direct numerical simulations of isotropic turbulence. *J. Fluid Mech.,* **207,** 531–586.