ELSEVIER

Journal of Marine Systems 29 (2001) 33–50

JOURNAL OF MARINE SYSTEMS

www.elsevier.nl/locate/jmarsys

Prediction of particle trajectories in the Adriatic Sea using Lagrangian data assimilation

Sergio Castellari [a,1], Annalisa Griffa [a,b], Tamay M. Özgökmen [a], Pierre-Marie Poulain [c,*]

[a] *RSMAS, University of Miami, Coral Gables, FL, USA*
[b] *IOF–CNR, La Spezia, Italy*
[c] *Department of Oceanography, Naval Postgraduate School, Code OC / Pn, Monterey, CA, 93943-5000, USA*

Received 15 October 1999; accepted 16 October 2000

Abstract

The predictability of Lagrangian particle trajectories in the Adriatic Sea (a semi-enclosed sub-basin of the Mediterranean Sea) over a period of 1–2 weeks is investigated using three clusters consisting of 5–7 drifters. The analysis is conducted using a Gauss–Markov Lagrangian particle model, which relies on the estimate of climatological mean flow field, persistence of turbulence, and assimilation of velocity data from the surrounding drifters through a Kalman filtering technique.

The results are described using the data density N_R defined as the number of drifters within a distance on the order of the Rossby radius of deformation from the particle to be predicted. The clusters are inherently different with respect to this characteristic property with values ranging from $N_R < 0.5$ to $N_R \geq 2.0$ over the analysis period, depending on the initial launch pattern of the clusters and the dispersion processes. The results indicate that during the period when $N_R \geq 1$, the assimilation of surrounding drifter data leads to an improvement of predicted trajectories with respect to those based on advecting the drifters with the mean flow. When $N_R < 1$, the drifters are too far apart to exhibit correlated motion, and the assimilation method does not lead to an improvement. The effects of uncertainties in the mean flow field and initial release position are discussed. The results are also compared to simple estimates of particle location by calculating the center of mass of the cluster.

Keywords: Adriatic Sea; Lagrangian data; Assimilation; Prediction

1. Introduction

Our understanding of oceanic phenomena and our capability to predict them have strongly improved in the last decade, thanks to the joint use of models and

* Corresponding author. Tel.: +1-831-656-3318; fax: +1-831-656-2712.
E-mail addresses: castellari@ingv.it (S. Castellari), poulain@oc.nps.navy.mil (P.-M. Poulain).
[1] Present address: Istituto Nazionale di Geofisica e Vulcanologia (INGV), Bologna, Italy.

0924-7963/01/$ - see front matter
PII: S0924-7963(01)00008-2

data. However, a number of problems remain unresolved. An outstanding example is the study of transport and its predictability.

Predicting the motion of particle trajectories in the ocean is important at various scales and for various applications, including sea rescue, recovery of lost objects, pollution problems and tracking floating mines and fish larvae. It is also a difficult and delicate problem for at least two main reasons: (a) velocity errors tend to accumulate as errors of prediction of position; and (b) Lagrangian motion of particles is often inherently chaotic, even when the underlying Eulerian flow is regular (Aref, 1984; Samelson, 1996). This implies high sensitivity to the initial conditions and to details of the velocity field, making the prediction very delicate.

A first step in predicting particle motion can be taken by using statistical dispersion models (Falco et al., 2000). They provide information about the probability of finding particles in a given region, based on previous statistical information on the velocity field. Often though this information is not accurate enough for emergency problems like the ones mentioned above, when contemporary information on the state of the velocity fields are necessary. In these cases, on-line data, which help in reconstructing the velocity field at a given time, are crucial to reduce prediction errors.

In a recent study (Özgökmen et al., 2000; referred to as OGMP in hereafter), the use of Lagrangian data to improve prediction of particle motion has been explored. This work is focused on mesoscale prediction, and it assumes that particle motion is approximately two-dimensional (either at the surface or along isopycnals). A highly simplified model for particle motion has been used, based on the knowledge of the climatological mean field and on the assumption of Markovianity for the fluctuating field. The impact of assimilating Lagrangian data in the model to improve prediction of unobservable particles has been studied. Theoretical error estimates, depending on the data density, have been derived. The validity of this method has been tested by generating synthetic drifter data within the framework of an idealized double-gyre ocean model. The predictive capability has been found to increase as the surrounding data density increased, in good agreement with theoretical error estimates.

Even though the results of OGMP are strongly indicative of the validity of this method, the motivation for the present study comes from the fact that the data used in their study are relative to idealized open ocean situations and the synthetic drifter trajectories are smoother in space and time than real data. Therefore, some questions persist on the performance of this method for real data in coastal regions and in the presence of small space and time scales of variability. Another important distinction of the present study is that in OGMP the drifters were launched in a regular array simultaneously over the entire ocean basin (highly difficult to accomplish in reality) in order to maintain an approximately constant data density over the period of prediction. In real-life applications, however, drifters released close to one another tend to spread apart, hence the effective data density decreases over the prediction period. Therefore, the present study is a first step toward quantifying predictability of Lagrangian motion in the real ocean given a limited and irregular distribution of drifter data.

In this study, we test the results of OGMP using surface drifter data taken in the Adriatic Sea, a semi-enclosed basin of the Mediterranean Sea characterized by locally steep topography and high velocity variability. The drifter data used here are a subset of a comprehensive data set which spans the period between December 1994 and March 1996. This data set has been previously described and analyzed in Poulain (1999), Poulain and Zanasca (1999) and Falco et al. (2000) to study the characteristics of the Adriatic surface circulation and of its transports. In the present study, three clusters of drifters are used, each characterized by different values of initial data density. The trajectory of each particle within a cluster is predicted using the other ones, and the error is computed as the difference between the true and predicted positions. An estimate of the root mean square (r.m.s.) error for each cluster is obtained as an average of that of all cluster particles. The values of the r.m.s. error are related to data density and compared with the estimates of OGMP.

The work has two parallel goals.

(a) To apply the prediction model developed in the context of an idealized large-scale ocean model by OGMP to real drifter data characterized by small-scale variability. This will allow us to test the

validity of the results obtained in OGMP for realistic data. In particular, we are interested in quantifying the dependence of the prediction error on data density.

(b) To shed light on the scales of predictability of Lagrangian motion in the Adriatic basin. This is important for both practical applications and for the understanding of the intrinsic properties of the system.

The paper is organized as follows. In Section 2, a brief review of the methodology used in OGMP is given, together with a summary of the theoretical results on error estimates. The drifter data set is described in Section 3. The methodology followed in the analysis and the results are presented in Section 4. Summary and conclusions follow in Section 5.

2. Methodology for particle prediction

In this section, the general concept and algorithm of the prediction model are outlined. For a more detailed description, the reader is referred to OGMP.

2.1. General formulation of data assimilation using sequential estimation

The central idea of data assimilation using sequential estimation (Ghil and Malanotte-Rizzoli, 1991; Ide and Ghil, 1998; Robinson et al., 1998) can be expressed as follows. Consider a system, whose state is described by an N-vector $\boldsymbol{x}$. The best estimate of $\boldsymbol{x}$ at time t_j, $x^{\mathrm{a}}(t_j)$, is obtained by averaging information from a model describing the system behavior and from observations related to the system state:

$$x^{\mathrm{a}}(t_j) = x^{\mathrm{f}}(t_j) + \mathrm{K}(t_j)\left[\mathbf{y}(t_j) - \mathrm{h}_j\left(x^{\mathrm{f}}(t_j)\right)\right], \quad (1)$$

where x^{f} is the forecast estimate of $\boldsymbol{x}$ obtained using the model alone, $\boldsymbol{y}$ is the observation vector, h is the transition matrix between $\boldsymbol{x}$ and the observations, and K is a weighting matrix.

The evolution model for $\boldsymbol{x}$ can be expressed in general terms as a nonlinear stochastic equation, which includes the unresolved sub-grid scale parameterizations:

$$\mathrm{d}\boldsymbol{x} = g(\boldsymbol{x})\mathrm{d}t + \mathrm{d}w, \quad (2)$$

where $g(\boldsymbol{x})$ represents the deterministic part of the dynamics and w is a Wiener process whose increments have zero mean and covariance matrix Q.

The forecast estimate x^{f} is obtained by resolving the deterministic part of Eq. (2):

$$\mathrm{d}x^{\mathrm{f}}/\mathrm{d}t = g(x^{\mathrm{f}}). \quad (3)$$

2.2. Particle model

In our case, the vector $\boldsymbol{x}$ to be predicted is composed by the velocity V and the position r of the (unobservable) particle, $\boldsymbol{x} = (V, r)$, while the observation vector $\boldsymbol{y}$ is given by the velocities and positions of the (observed) Lagrangian drifters.

A simple form for the evolution model of the particle position and velocity is assumed, based on a decomposition of the velocity into mean and fluctuating components (Thomson, 1987). In the following, it is assumed for simplicity that the particle moves in two dimensions, and that the time scale and variance matrices for the fluctuating velocity are diagonal. Therefore, the equations for each component reduce to:

$$\mathrm{d}r = V(r)\mathrm{d}t = (U(r) + v)\mathrm{d}t, \quad (4)$$

$$\mathrm{d}v = -T^{-1}v\mathrm{d}t + \mathrm{d}w, \quad (5)$$

where $U(r)$ is the Eulerian climatological mean velocity field, assumed known, v is the fluctuation around the mean characterized by variance σ^2, T is the time scale, and $\mathrm{d}w$ has variance $2\mathrm{d}t\sigma^2/T$. Note that, since U is assumed to be known, only the fluctuating component v is predicted by the model.

Eq. (5) states that the fluctuation v following the particle is a linear Markov process with exponential autocovariance R_{L}. For each component, we have:

$$R_{\mathrm{L}}(\tau) = \langle v(t)v(t+\tau)\rangle = \sigma^2\exp(-\tau/T). \quad (6)$$

Physically, this means that the particle fluctuation has a certain persistence for time scales on the order of T, after which period the particle velocity tends to relax toward the climatology, represented by U. Hence, T is the "memory" time scale of the system, and information is lost for $t \gg T$.

Simple models like Eqs. (4) and (5), which are essentially statistical in nature and based on climatol-

ogy and persistence, have been successfully applied in a number of prediction problems (e.g., Bril, 1995). Alternatively, more sophisticated models incorporating more complete dynamics can be used, such as the output velocity fields from high resolution regional circulation models. The advantage of using Eqs. (4) and (5) is that this method is entirely portable (i.e., it does not assume an operational regional model in the area of interest), and it does not depend on the accuracy of the regional model (which is often still an open question in practical applications). However, the lack of dynamics in Eqs. (4) and (5) allows a forecasting period only on the order of T. This point will be investigated in Section 4 using results from the drifter data.

2.3. Algorithm implementation

A simple algorithm has been implemented in OGMP to predict particle velocity v and position r. The best estimate of v, called v^{a}, is computed using model (5) and following formulation (1). Since Eq. (5) is linear, the weighting function K can be easily computed using the Kalman filtering technique, which in this case coincides with the Optimal Interpolation technique. The position r^{a} is then predicted using v^{a} in Eq. (4).

The algorithm is implemented using some additional assumptions on the Eulerian statistics for computational simplicity. These assumptions can be relaxed without loss of generality. The velocity components are assumed to be uncorrelated and the Eulerian velocity covariance, B, is assumed to be a Gaussian function for each component:

$$B(r) = \exp\left(-r^2/2R^2\right) \tag{7}$$

where R is the Eulerian space scale. Under these assumptions, the algorithm can be written as:

$$v^{\mathrm{a}}(t_j) = \exp(\mathrm{d}t/T)\, v^{\mathrm{a}}(t_{j-1}) + \sum_{k=1}^{p} \mathrm{K}_k\left[v_k(t_j) - \exp(\mathrm{d}t/T)\, v_k(t_{j-1})\right], \tag{8}$$

$$r^{\mathrm{a}}(t_{j+1}) = r^{\mathrm{a}}(t_j) + \left[U\left(r^{\mathrm{a}}(t_j)\right) + v^{\mathrm{a}}(t_j)\right]\mathrm{d}t, \tag{9}$$

where p is the number of observed drifters, v_k is their velocity at discrete times denoted by t_j. The weighting matrix K is given by:

$$B_1\mathrm{K} = B_2 \tag{10}$$

where

$$(B_1)_{kl} = B\left(r_k(t_j) - r_l(t_j)\right),$$
$$(B_2)_k = B\left(r^{\mathrm{a}}(t_j) - r_k(t_j)\right) \tag{11}$$

2.4. Theoretical error analysis

First, some quantities used in the following analysis of results are defined. The absolute dispersion, S_{A}, defined as

$$S_{\mathrm{A}}(t) = \langle\left(r(t) - r(0)\right)^2\rangle^{1/2}, \tag{12}$$

provides a measure of the total distance traveled by the particles from their original starting position during the considered time, and it can be therefore considered as an upper bound for error for any reasonable prediction algorithm. The angle brackets denote ensemble averaging over all the drifters in the cluster considered.

Second, the forecast error, S_{f}, is defined as

$$S_{\mathrm{f}}(t) = \langle\left(r^{\mathrm{f}}(t) - r(t)\right)^2\rangle^{1/2}, \tag{13}$$

where r^{f} is the forecast position computed from Eqs. (4) and (5) with initial conditions $r(0)$, $v(0)$ and r is the position of the unobservable particle. As indicated in Eq. (3), r^{f} is computed using only the deterministic part of Eqs. (4) and (5). This error quantifies the forecasting skill of the simple model Eqs. (4)–(5), which is expected to quickly decrease after time scales of order T.

Next, S_{M} defined as

$$S_{\mathrm{M}}(t) = \left\langle\left(r(t) - \int_0^t U(r(s))\,\mathrm{d}s\right)^2\right\rangle^{1/2}, \tag{14}$$

is the error for trajectory prediction based only on the advection by the mean flow in Eq. (4), i.e., when no fluctuating component is calculated and utilized.

Finally, the r.m.s. prediction error with data assimilation, i.e., using the complete algorithm Eqs. (8)–(9), is defined as

$$S(t) = \langle\left(r^{\mathrm{a}}(t) - r(t)\right)^2\rangle^{1/2}. \tag{15}$$

In OGMP, analytical estimates for errors associated with the algorithm (8)–(9) are given, under the assumptions that v is statistically homogeneous with a typical r.m.s. value σ, and that U is approximately constant over a certain time and space interval. Particle density is measured in terms of N_R, the average number of observations within a circle of radius R centered at the particle position. It is shown in OGMP that

$$S(t) \approx \gamma S_M(t) \tag{16}$$

where γ is a proportionality constant that depends on data density. The values of γ have been computed analytically for the following values of N_R;

$$\gamma = 1 \text{ if } N_R = 0, \tag{17}$$

$$\gamma \approx 0.3 \text{ if } N_R = 1, \tag{18}$$

$$\gamma \approx \left(\frac{\ln N_R}{12 N_R} \left(1 - \frac{1}{N_R \sqrt{N_R}} \right) + \frac{1}{N_R \sqrt{N_R}} \right)^{1/2}$$

$$\text{if } 1/N_R \to 0. \tag{19}$$

These results indicate that observations at a lower data density than $N_R = 1$ do not improve the prediction obtained with the mean flow alone. Note from Eq. (18) that it is sufficient to have 1 observation per degree of freedom R, to improve substantially (by 70%) the prediction.

Note that S_M corresponds, for constant U, to the single particle dispersion relative to the mean. This implies that S grows first linearly with time

$$S(t) \approx \gamma S_M(t) \approx \gamma \sigma t \text{ if } t \ll T, \tag{20}$$

and then proportional to the square root of time,

$$S(t) \approx \gamma S_M(t) \approx \gamma \sigma \sqrt{Tt} \text{ if } t \gg T. \tag{21}$$

These theoretical error estimates have been tested in OGMP using synthetic drifters in a double-gyre simulation. The numerical results agree well with the estimates (16)–(19), even in regions where the mean flow is strongly inhomogeneous, suggesting that they can be used as general guidance in various dynamical regions.

3. The Adriatic drifter data set

The data set considered in this paper consists of three distinct clusters of drifters, corresponding to a total of 18 drifters. They were released in the eastern part of the Strait of Otranto, the southern entrance of the Adriatic Sea, at different seasons during 1995. We only consider the trajectories for the first 20 days after the drifters enter the basin. These data belong to a larger drifter data set spanning December 1994 to March 1996, that has been summarized in a data report (Poulain and Zanasca, 1998) and analyzed in previous papers (Poulain, 1999; Falco et al., 2000). The drifter position data were interpolated (using a "kriging" technique; Hansen and Poulain, 1996) and low-pass filtered (Poulain and Zanasca, 1998). Velocities were estimated by finite difference of the filtered positions and were subsampled, together with positions, at 6-h intervals. First, a quick review of the statistical results included in the latter papers will be provided, and will be used in the present study. The characteristics of the three clusters are then described in detail.

3.1. General statistics

The Adriatic Sea (Fig. 1a), is an elongated semi-enclosed basin connected to the Mediterranean Sea at its southern entrance through the Strait of Otranto. The topography is characterized by a deep center in the southern region reaching 1200 m depth, while the central and northern regions are shallower, around 200 m and less than 100 m, respectively. Here we concentrate mostly on the southern and central regions, where the drifter data coverage is greater.

The total data set in the Adriatic Sea, defined here as the area north of 40.4°N, consists of 41 drifters (Falco et al., 2000) with a mean half life time of about 100 days (Poulain and Zanasca, 1998). Ten drifters were launched randomly over the basin, and 31 were launched in the eastern Strait of Otranto south of 40.4°N and entered the basin with the inflow of Ionian waters through the Strait. The data set is extensive enough to provide significant information on the ensemble average properties of the flow during the measurements period, but it is not sufficient to resolve seasonality. Note that in the present study, only 18 of the 41 drifters from the data set are used. This is because the other ones are too sparse in space and time to be considered as

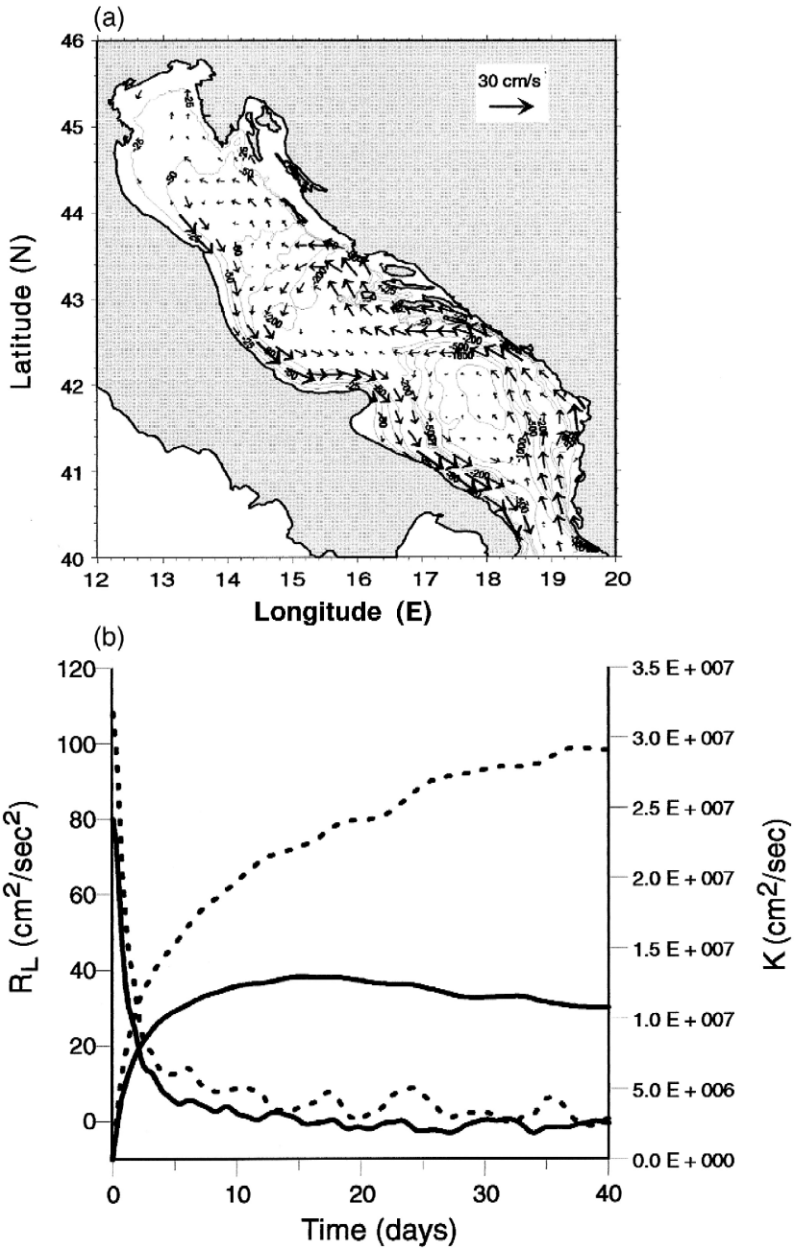

Fig. 1. (a) Large-scale mean flow obtained by the bicubic spline technique. (b) Autocovariances R_L and diffusivities K resulting from the flow decomposition. Solid (dashed) lines refer to the zonal (meridional) component (from Falco et al., 2000).

clusters and to be used in the present predictability study.

The mean flow over the whole drifter ensemble (December 1994–March 1996) has been computed using two different methodologies: binning over quarter-degree boxes (Poulain, 1999), and using an optimized spline method which minimize the energy in the residual velocity at scales longer than a week (Falco et al., 2000; Bauer et al., 1998). The results for the spline estimates, reproduced in Fig. 1a, are

qualitatively and quantitatively similar to the binning results of Poulain (1999). Subtidal fluctuations around the mean flow can be substantial, with a standard deviation as large as the mean (Poulain, 1999). Error estimates for the mean flow shown in Fig. 1a were estimated by Falco et al. (2000) to have a lower bound of 20%, corresponding to the simplifying assumption of homogeneous and stationary statistics. The depicted circulation is cyclonic, with recirculations in the southern and central regions, and with intensified currents along the boundaries, reaching values of approximately 30 cm/s. Historical (Artegiani et al., 1997) and satellite (Gacic et al., 1997) data suggest that the circulation is seasonally modulated, with a stronger and better defined mean flow during winter and less clear patterns during summer.

The Lagrangian fluctuation field has been analyzed by computing residual velocities along particle trajectories (obtained by subtracting the splined mean flow values at each time step), and then computing autocovariances and diffusivities from the residuals (Fig. 1b). The statistics computed in this way characterize the fluctuations in a bulk, without taking into account the spatial inhomogeneity, such as the fact that the fluctuations are more vigorous in the boundary regions than those in the interior (Poulain, 1999). Despite this limitation, the bulk values have been shown to effectively capture the variability (Falco et al., 2000), providing a good starting point for our analysis. The autocovariances show an exponential behavior, consistent with the model (5)–(6), and they suggest a value of roughly 1 day for the *e*-folding time scale. The two velocity components are slightly anisotropic, and the variance value is of the order of $\sigma^2 \approx 100$ cm^2/s^2, suggesting a typical r.m.s. value for the fluctuating velocity of the order of $\sigma \simeq 10$ cm/s.

Note that, as discussed in Falco et al. (2000), at closer inspection the autocovariances in Fig. 1b (especially the meridional one) do not appear to vanish exactly for $t > 10$ days. In other words, there is a deviation from the exponential behavior at time lags $\gg T$. This is likely to be the consequence of a large-scale or seasonal residual, which is not correctly removed. This is also reflected in the behavior of the diffusivity, $D = \int_0^\infty R_L(\tau)\mathrm{d}\tau$, which does not exactly converge toward a constant value (Fig. 1b). The presence of "tails" in autocovariances has important consequences for the estimate of the time scale T. This time scale can be estimated either as integral time scale defined as $T \approx \sigma^{-2}\int_0^\infty R_L(\tau)\mathrm{d}\tau = D/\sigma^2$ or as *e*-folding time scale of $R(\tau)$ (e.g., Griffa et al., 1995). For a perfect exponential autocovariance such as in Eq. (6), these two estimates give identical results, but in presence of tails they can differ significantly. In Falco et al. (2000), where long-term ($\gg 10$ days) dispersion processes are considered, T is estimated as integral time scale, in order to capture some of the long-term behavior. Values of T estimated in this way are ≈ 1.5 (3) days for the meridional (zonal) component. In the present paper, we are mostly interested in short-term prediction over a few days. Therefore, it appears more appropriate to estimate T as *e*-folding time scale, to capture the initial behavior. Estimates of T obtained in this way are smaller, suggesting values not exceeding 1 day, as mentioned before.

3.2. The clusters

The three clusters analyzed in this study were originally released at different times in the eastern Strait of Otranto, and are re-initialized after they cross 40.4°N and enter in the Adriatic basin. The position of the particles within each cluster are shown in Figs. 2a, 3a and 4a at time $t = 0$ (i.e., when they enter the basin), $t = 10$ days and $t = 20$ days. The main parameters characterizing the clusters are reported in Table 1.

As shown in OGMP and summarized in Section 2.4, the predictability of drifter trajectories is expected to depend on the effective data density N_R over the period of interest. Figs. 2b, 3b and 4b illustrate the average N_R for each cluster for the 20-day period. The individual drifters in each cluster have similar N_R so that the average value of N_R can

Table 1
Characteristics of the three clusters

Cluster	Number of drifters	Initial time	N_R range	σ [cm s^{-1}]
A	6	September 2, 1995	≥ 2	15
B	5	May 25, 1995	< 1	10
C	7	October 20, 1995	≤ 0.5	15

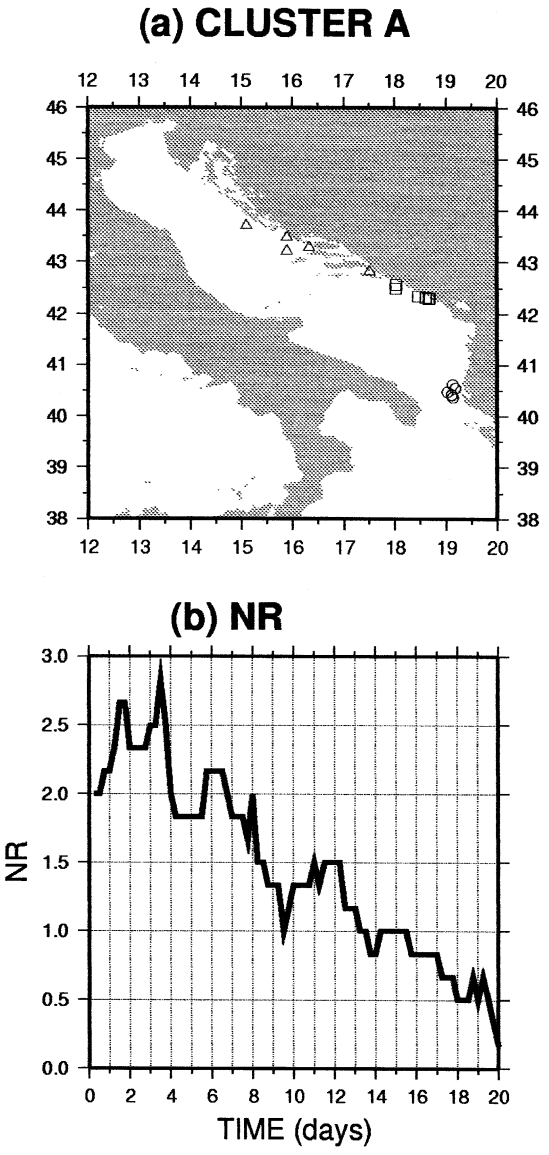

Fig. 2. (a) Positions of the drifters in cluster A at day 0 (circles), at day 10 (squares) and at day 20 (triangles). (b) Cluster-averaged N_R for 20 days.

be considered significant. Even though the initial values of N_R depend on the dynamics experienced by the drifters before they entered the basin, here they are considered as given, as if the drifters were actually released at $t = 0$ in the basin. It can be seen from Figs. 2b, 3b and 4b that these three clusters are characterized by different N_R at initial time, and therefore that they are expected to have different predictability characteristics. Cluster A is characterized by the highest value of initial N_R (≥ 2), cluster B has intermediate values ($N_R \leq 1.0$), and cluster C has the lowest values ($N_R < 0.5$). Therefore, if the results of OGMP remain valid, it is anticipated that assimilation of surrounding data will lead to signifi-

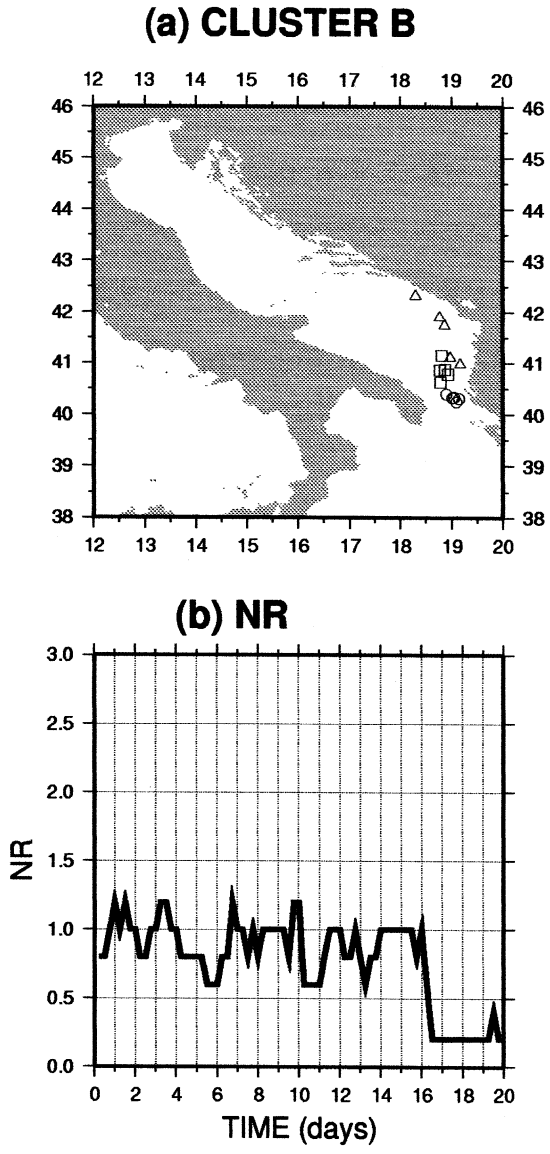

Fig. 3. (a) Positions of the drifters in cluster B at day 0 (circles), at day 10 (squares) and at day 20 (triangles). (b) Cluster-averaged N_R for 20 days.

cant improvement of predictability (with respect to mean flow prediction) for cluster A, marginal improvement for cluster B and no improvement for cluster C. Hence, these three clusters form a good data set for testing the validity of the ideas developed by OGCM.

Cluster A is composed of six drifters (Fig. 2a), starting on September 2. They appear to be embedded in the boundary current along the eastern side of the basin and move quickly to the north. The r.m.s. value of the fluctuation field with respect to the Eulerian velocity field depicted in Fig. 1a is $\sigma \approx 15$ cm/s. It is higher than the average basin value, $\sigma \approx 10$ cm/s, and it is also the highest of the three clusters. This is due to the fact that the drifters sample the boundary current in an especially intense period, at the end of summer/beginning of fall (Poulain, 1999; Falco et al., 2000). Cluster B is composed of five drifters (Fig. 3a), beginning on May 25. They move in the southern gyre with an r.m.s. velocity $\sigma \approx 10$ cm/s. With respect to cluster A, the drifter trajectories in cluster B are characterized by smaller scale fluctuations, probably because (a) they sample a region with less intense mean flow, and (b) the spring/summer season is characterized by smaller scale eddies dominating the circulation (Poulain, 1999). Finally, cluster C is composed of seven drifters (Fig. 4a), starting on October 20. The drifter trajectories cover different dynamical regions in the basin, from the interior to the boundary currents, and their r.m.s. fluctuation is $\sigma \approx 13$ cm/s.

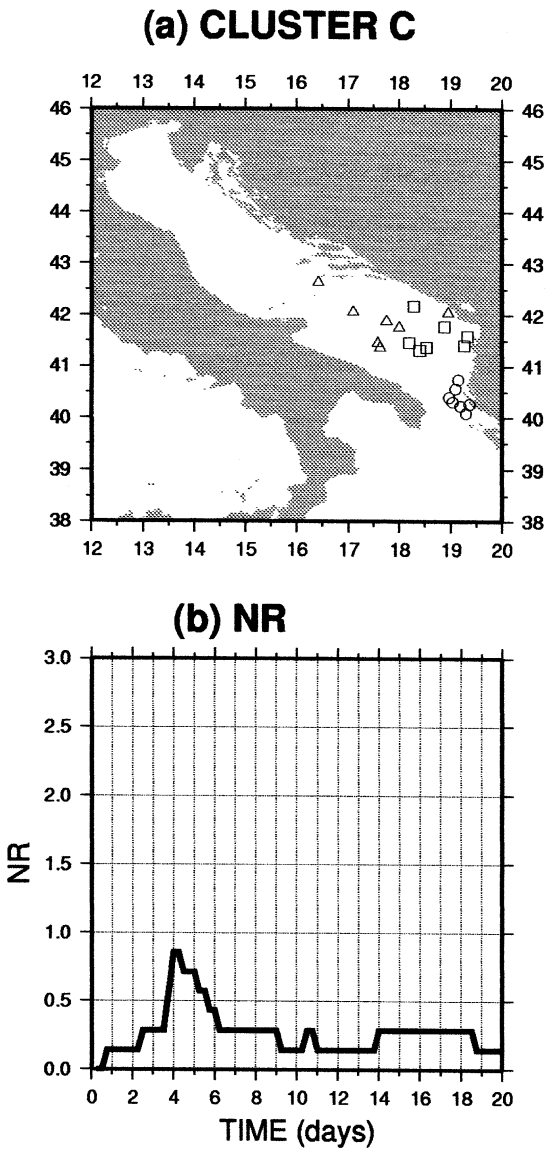

Fig. 4. (a) Positions of the drifters in cluster C at day 0 (circles), at day 10 (squares) and at day 20 (triangles). (b) Cluster-averaged N_R for 20 days.

Given the difference in pattern and seasonality of the three clusters, the values of absolute dispersion S_A are expected to be different and not directly comparable. However, the values of dispersion around the mean, S_M, are expected to be relatively similar and to depend mostly on σ. We will come back to this point in Section 4.

4. Results

4.1. Predictability analysis

The analysis is performed for the three clusters separately. First, the forecasting skill of the basic model (4)–(5) in absence of data is tested. Then, the impact of using data assimilation (8)–(9) in improving the predictability of particle trajectories as function of data density N_R is studied, checking the theoretical relationships (16)–(19).

The forecast of the model for each cluster is computed in the following way. The algorithm (8)–(9) is applied for each particle assuming that the initial position $r(0)$ and velocity $v(0)$ are known and that there are no surrounding data. The r^a position

corresponds to the forecast position r^f. The error of the forecast, S_f is computed according to Eq. (13). Time series of S_f are considered for the first few days only, since the forecasting skill of the simple model (4)–(5) is expected to quickly decrease after time scales of order T. Note that quantifying the forecasting skills of the model is important for practical applications, such as observable pollutants. If a drifter is released in a pollutant, and its velocity is computed by finite difference from its previous positions, its future position can be predicted using the procedure described above and the following results quantify the prediction accuracy.

The change of predictability in presence of data is studied by comparing the time series of S (r.m.s. prediction error with data assimilation), Eq. (15) and S_M (r.m.s. prediction error with mean flow advection only), Eq. (14), to check the validity of Eqs. (16)–(18). While computing the predicted trajectories with data assimilation, the initial velocity $v(0)$ is estimated from the surrounding data using Eq. (8). In the absence of data assimilation (i.e., for the computation of S_M), the initial turbulent velocity is taken as zero [$v(0) = 0$].

The parameters of the model, the mean flow field U, the time scale T and the space scale R are determined as follows. As guideline for the values of U and T, the results shown in Section 3 are used. For U, the estimate in Fig. 1a is used (Falco et al., 2000). For T, values of the order of 1 day are used, as suggested by the e-folding time scales of the autocovariances in Fig. 1b. Within the 1-day range, specific values are chosen (by trial and error) for each cluster application in order to optimize the assimilation results (see Table 2). Regarding R, values of the order of the Rossby radius of deformation are chosen, which is approximately 10 km in the southern Adriatic (Grilli and Pinardi, 1998). The parameters chosen for each cluster are nearly the same and they are listed in Table 2. The behavior of predictions for all clusters remain quite robust to variations on the order of 50% in T and R; the r.m.s. prediction error varies by less than 10% at 10 days and by 10–20% at 15 days when T or R are varied by 50% from the values listed in Table 2. Since drifter positions and velocities were subsampled at 6-h intervals, the time lag between two successive corrections in Eqs. (8)–(9) was set to 6 h. The total prediction time for each cluster is taken as the time for which the data density N_R remains roughly constant, which corresponds to about 1 week for cluster A and approximately to 2 weeks for clusters B and C (Figs. 2b, 3b and 4b). This is done in order to be able to compare the results with the analytical estimates (17)–(19), which are obtained under the assumption of constant N_R in OGMP. Drifters released in clusters spread apart in time such that the value of N_R for each particle (hence for the average) tends to decrease with time. For practical future applications, replenishing of the initial clusters with additional drifters at later times can be planned, in order to maintain $N_R > 1$.

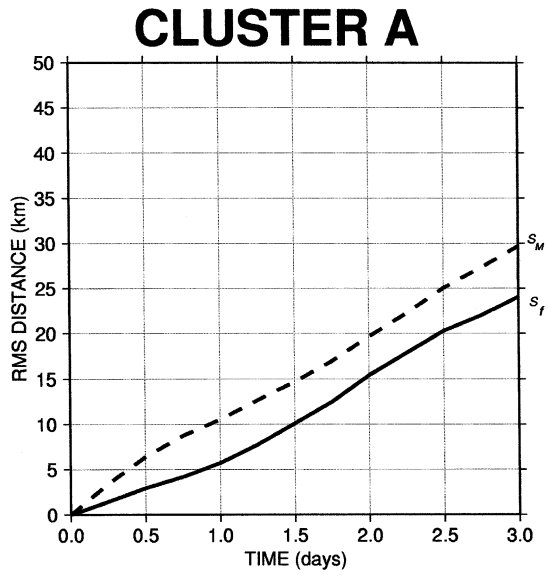

Fig. 5. Forecast r.m.s. error S_f (solid line) and mean flow prediction error S_M (dashed line) for cluster A from 0 to 3 days.

Table 2
Parameters used in the assimilation: T is the Lagrangian e-folding time scale and R is the Eulerian space scale

Cluster	T [day]	R [km]	Prediction duration [day]
A	0.4	10	7
B	0.4	10	15
C	0.8	10	13

4.2. Cluster A

In order to quantify the predictive skill of our basic model without data assimilation, the forecast error S_f is first calculated, and the results for cluster A are shown for the first 3 days in Fig. 5 (solid line). For comparison, r.m.s. prediction error obtained using only the mean flow, S_M, is also calculated (dashed line). Notice that S_M and S appear to grow linearly for $t > T$, rather than following the square root behavior suggested by Eq. (21). This is likely to be due to shear dispersion, which is present in the strongly sheared boundary currents, and which was not entirely removed by subtracting the Eulerian flow of Fig. 1a. Shear dispersion is not included in the theoretical estimate Eq. (21), which corresponds to $U \approx$ constant. The primary difference between S_f and S_M in Fig. 5 is that S_f incorporates the advantage of knowing the initial turbulent velocity, whereas prediction leading to S_M is conducted with zero initial turbulent velocity. The results indicate that the knowledge of the initial velocity reduces the prediction error over a time scale on the order of T, as expected. The values of S_f after 1 day are about 5 km, and then they increase quickly, reaching almost 15 km after 2 days. At later times, S_f is approximately parallel to S_M. The values of S_M can be approximated, at least initially, using $\gamma = 1$ in Eq. (20). The primary message in this experiment is that the basic model in the absence of data assimilation leads to improvement in predictive capability only during the initial period on the order of T. This appears to be true also for the other two clusters, B and C (not shown). For all three clusters, the values of S_f after the first day are similar, $S_f \approx 5$ km.

The impact of assimilating data is then studied for the first 7 days, corresponding to the period when $N_R \geq 2$. From the theoretical results (16)–(19) and

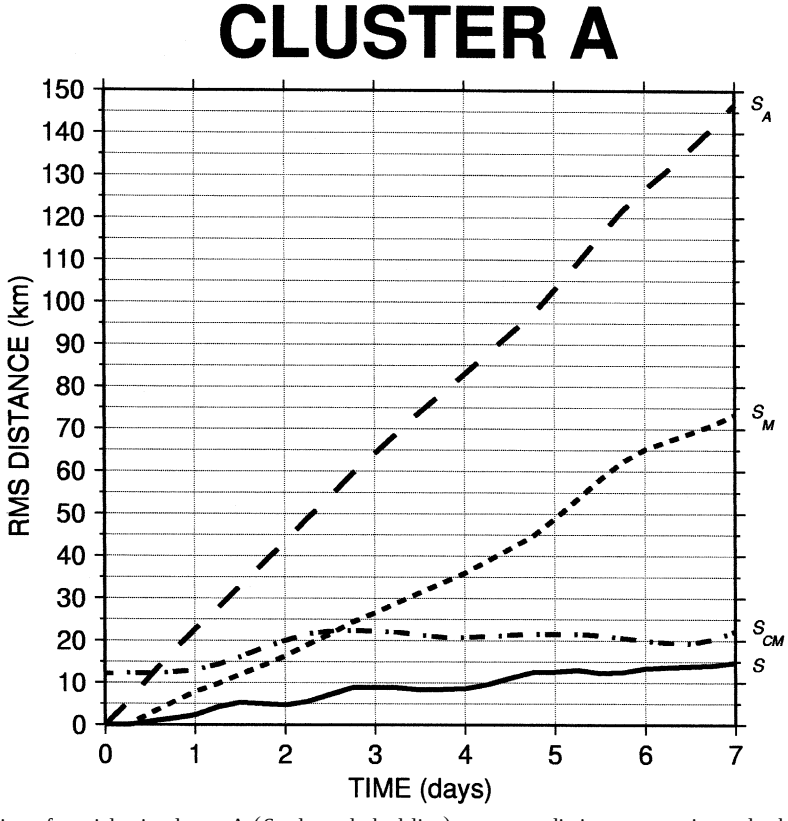

Fig. 6. Absolute dispersion of particles in cluster A (S_A, long dashed line), r.m.s. prediction errors using only the mean flow (S_M, short dashed line), Lagrangian data assimilation (S, solid line) and center of mass (S_{CM}, dashed–dotted line).

from the numerical results of OGMP, a significant reduction of the error S, of more than 70%, is expected with respect to S_M. For comparison, the absolute dispersion S_A is calculated, which can be considered as an upper bound to the S and S_M values. The results are illustrated in Fig. 6. The absolute dispersion S_A grows almost linearly and reaches approximately 145 km after 7 days. S_M, which corresponds to the prediction using only the mean flow and with no data assimilation, shows a consistent decrease with respect to S_A, reducing the error approximately by half, to ≈ 70 km, after the same amount of time. This is not surprising given that the region sampled by the drifters is characterized by a strong mean flow, and it is in good qualitative agreement with the results of OGMP. Notice also that the almost linear shape of S_M is suggestive of shear dispersion. Finally, the error of the prediction with data assimilation, S, is much smaller than S_M, reaching a maximum of only 17 km after 7 days. This error reduction is of approximately 75% with respect to S_M, as anticipated.

Two examples of true and predicted drifter trajectories for cluster A are shown in Fig. 7. As it can be seen, predictions with only the mean flow tend to significantly underestimate the velocity or the total displacement of the particle, while predictions with data assimilation overcome this problem and the predicted trajectory remains quite close to the true trajectory during the integration. The results for the other particles are qualitatively similar.

In Fig. 6, the errors for the data assimilation algorithm are compared with errors obtained for a different estimate of particle position. This is a simple "common sense" estimate, where the position of each particle is approximated by computing the center mass of the others. As the figure shows, the error of this estimate is consistently higher than S during first 7 days of integration. At later times (not shown), these two curves tend to converge towards each other. This happens when N_R is approximately zero and there is no useful information to assimilate. The growth of error for the center of mass estimate is tied to the properties of relative dispersion of the flow, characterized, for particle pairs, by the Lyapunov exponent (Lacorata et al., 2000; Piterbarg, 2001). In all the three clusters considered here, the center of mass error is initially higher than S and then it tends to converge. We can speculate, though, that for an optimal initial sampling, these two estimates and the respective errors could converge.

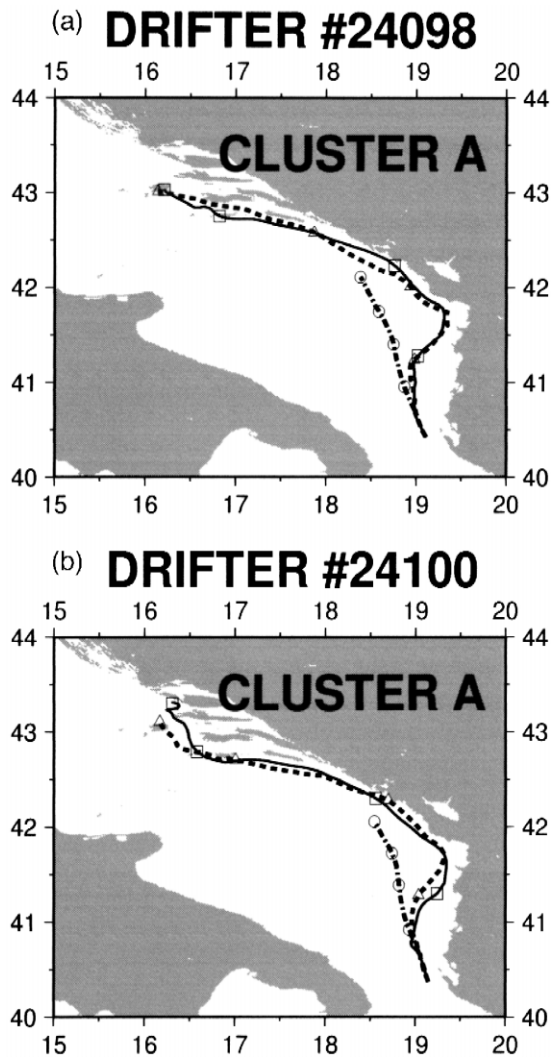

Fig. 7. True and simulated trajectories for drifter #24098 (upper panel) and #24100 (lower panel) of cluster A. Solid lines indicate the true trajectories (square symbols at 5-day intervals), predicted trajectories using only the mean flow information are shown with dashed–dotted lines (circle symbols at 5-day intervals), and dashed lines depict predicted trajectories using data assimilation (triangle symbols at 5-day intervals).

The effects of degrading the knowledge of the system are studied, as in OGMP, by considering two cases, which are of importance for practical applications: (a) the errors in the knowledge of the mean

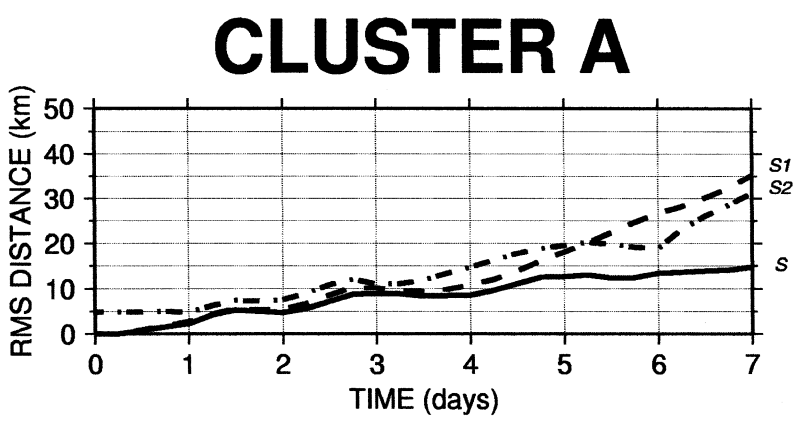

Fig. 8. The r.m.s. prediction errors S for drifters in cluster A using Lagrangian data assimilation with mean flow information and exact initial positions (solid line, same as in Fig. 7), zero mean flow and exact initial positions (S_1, dashed–dotted line), and with mean flow information but an error of 5 km on the initial positions (S_2, dashed line).

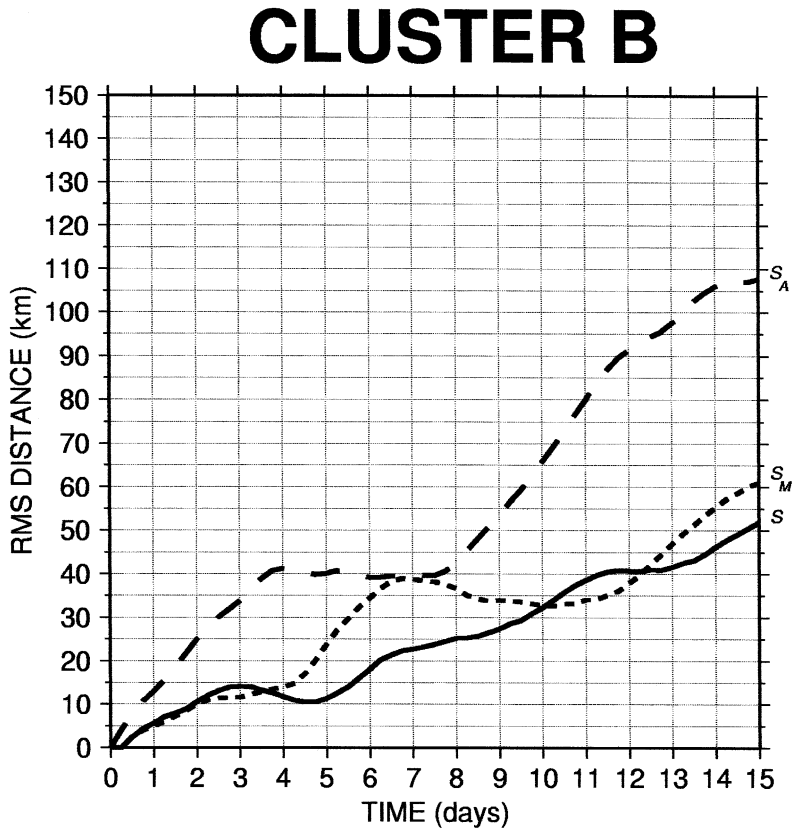

Fig. 9. Absolute dispersion of particles in cluster B (S_A, long dashed line), r.m.s. prediction errors using only the mean flow (S_M, short dashed line) and Lagrangian data assimilation (S, solid line).

flow and (b) the uncertainties regarding the initial position. First, it is assumed that the mean flow is unknown, and it is set to zero ($U = 0$) in Eq. (9). Second, it is assumed that there is an error in the initial position of the particle. Errors with a magnitude of 5 km (approximately half of the deformation radius) are introduced to the south, north, east and west of the exact initial positions, and then average r.m.s. prediction errors are computed. Similar experiments have been previously performed by OGMP with numerical drifters and they indicated that for $N_R > 1$ the predictive capability is recovered in both cases. The present results shown in Fig. 8 confirm the results of OGMP, demonstrating that the errors do not change significantly, at least for the first 5–6 days when N_R is sufficiently high.

An important issue to address is the significance of these results, given that the cluster consists of only 6 drifters (similar number of drifters for clusters B and C). To address this problem, first the variability in particle predictions is quantified for each cluster, computing sample standard deviations for S and S_M. The results from the standard deviation curves show that the difference between the curves in Fig. 6 is significant. Second, for short-term statistics, we have increased the number of degrees of freedom by a re-sampling strategy: the drifter trajectories, which are 7 days long for the predictability analysis in Fig. 6, were divided into segments 3.5 days long. Since 3.5 days is longer than the Lagrangian decorrelation time scale (≈ 1 day), each segment can be considered as a statistically independent trajectory, thereby increasing the number of trajectories available for the analysis to 12. When the 12 trajectories are used to compute the same quantities as in Fig. 6, the results remain qualitatively the same. Quantitatively, S_M and S tend to increase, of less than 20% with respect to those obtained without re-sampling along the trajectories, and the relative difference between S_M and S remains approximately the same.

4.3. Cluster B

As discussed in Section 4.2, the forecasting skills of cluster B are similar to those of cluster A. With regards to the assimilation, cluster B is not expected to show the same enhancement in predictability as for cluster A, given that $N_R < 1$. This is illustrated in Fig. 9, where the time series of S, S_M and S_A are shown for 15 days, corresponding to the total time when N_R is approximately constant (Fig. 3).

S_A grows slower than in cluster A, reaching a value of approximately 100 km after 15 days. The

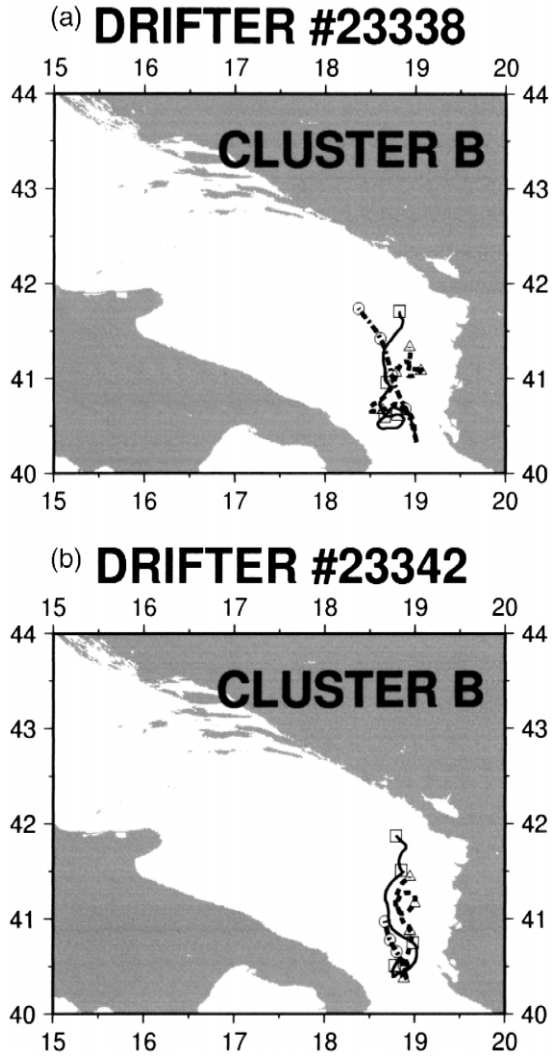

Fig. 10. True and simulated trajectories for drifter # 23338 (upper panel) and # 23342 (lower panel) of cluster B. Solid lines indicate the true trajectories (square symbols at 5-day intervals), predicted trajectories using only the mean flow information are shown with dashed–dotted lines (circle symbols at 5-day intervals), and dashed lines depict predicted trajectories using data assimilation (triangle symbols at 5-day intervals).

behavior is approximately linear, except for a plateau at approximately 4–8 days. This is probably related to the presence of a small, stationary eddy capturing the drifters (see Fig. 10), and it is expected to disappear with better statistics including more realizations. S_M show a decrease with respect to S_A, indicating the importance of the mean flow, and it reaches approximately 60 km after 15 days. S is generally smaller than S_M, even though the decrease is not completely significant given that the sample standard deviations overlap. When the re-sampling is applied, the results (not shown) are qualitatively similar, even though S appears clearly smaller than S_M even at initial times. We believe that the difference with and without re-sampling, is due to the fact that the drifter velocities, during their initial motion near the Strait of Otranto, are anomalously close to the mean flow, as shown also in Fig. 10. This is probably due to the fact that the mean current is particularly strong in the Strait and better defined than inside the basin, causing the values of S_M to be anomalously smaller.

Two examples of true and predicted trajectories are shown in Fig. 10. The improvement due to the data assimilation, even though not substantial as in cluster A, is noticeable especially in the tracking of the loops around the eddy.

4.4. Cluster C

The results for cluster C are shown in Fig. 11 for the first 15 days. N_R is very small (< 0.5), hence no improvement in predictability by assimilation of information from the surrounding drifters is expected. S_A grows almost as quickly as for cluster A, reaching approximately 210 km after 15 days. The mean

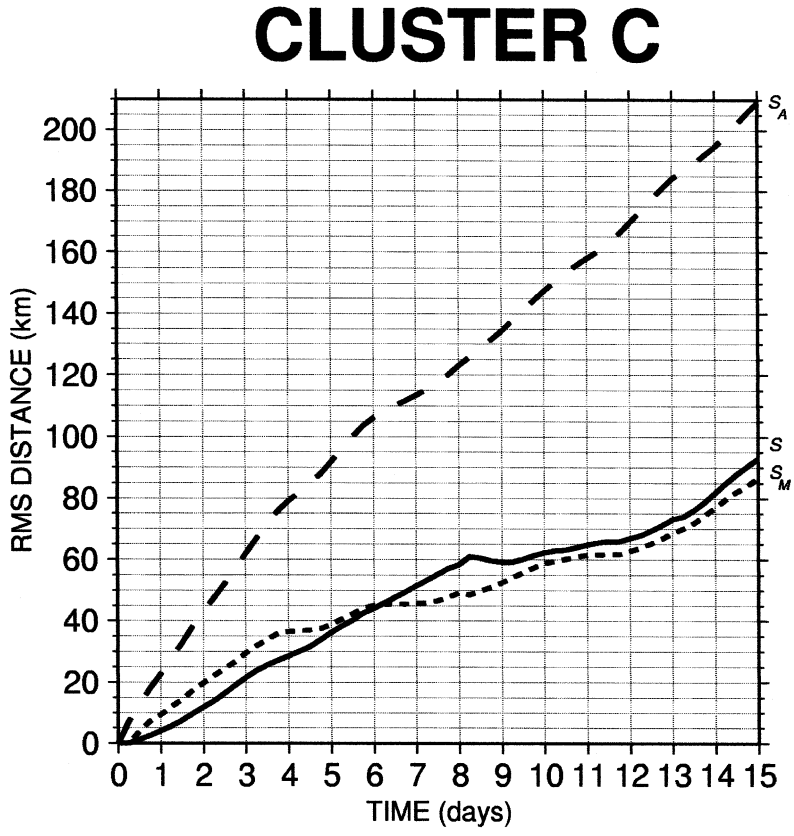

Fig. 11. Absolute dispersion of particles in cluster C (S_A, long dashed line), r.m.s. prediction errors using only the mean flow (S_M, short dashed line) and Lagrangian data assimilation (S, solid line).

flow knowledge reduces substantially the r.m.s. prediction error, as shown by S_M that reaches approximately 90 km after 15 days. The prediction with data assimilation, S, approximately coincides with S_M, showing that Lagrangian data assimilation is not effective, as expected. This is also shown by the two examples of true and predicted particle trajectories depicted in Fig. 12.

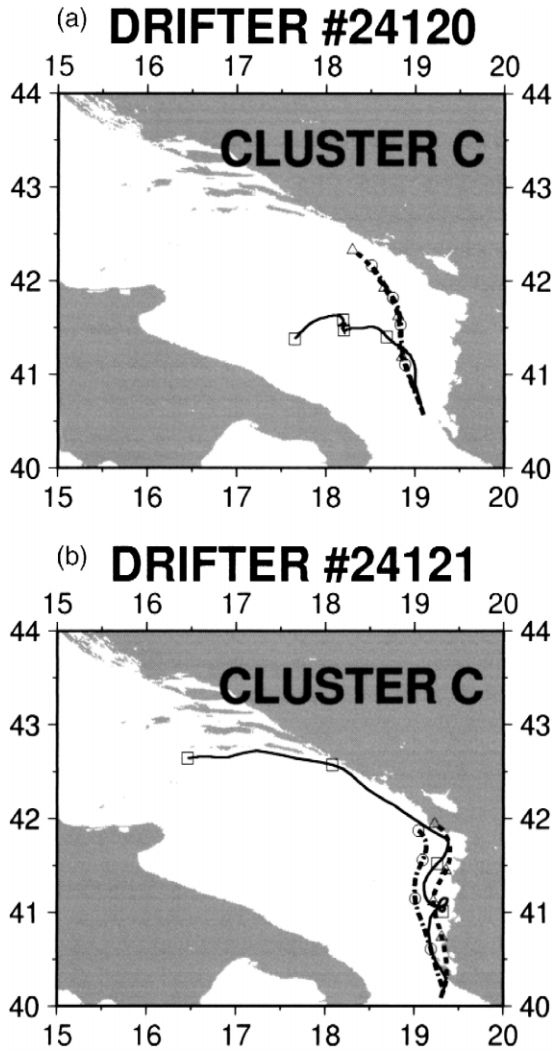

Fig. 12. True and simulated trajectories for drifter #24120 (upper panel) and #24121 (lower panel) of cluster C. Solid lines indicate the true trajectories (square symbols at 5-day intervals), predicted trajectories using only the mean flow information are shown with dashed–dotted lines (circle symbols at 5-day intervals), and dashed lines depict predicted trajectories using data assimilation (triangle symbols at 5-day intervals).

5. Summary and concluding remarks

The predictability of Lagrangian particle trajectories in the Adriatic Sea was explored using three clusters of drifters over a period of 1–2 weeks. The predictability scales were studied by employing a simple model, which relies on the assimilation of surrounding information from drifters to improve the estimation of the turbulent velocity component and hence the estimated trajectory. This model has been developed and previously tested by OGMP in the context of an idealized, large-scale ocean model. The objective of the present study is to apply this technique to real drifter data and to determine whether the conclusions of OGMP remain valid for this smaller scale, highly variable flow regime. Also, the study determined the specific scales of predictability for the Adriatic basin, providing useful information for future applications.

The drifter clusters used in this study are part of a larger data set collected in the Adriatic Sea between December 1994 and March 1996. Three clusters exhibiting different characteristics were selected. Cluster A consisted of 6 drifters, starting at the beginning of September 1995. These drifters remained in the eastern coastal boundary, in coherent and energetic currents. Cluster B was initialized in May 1995, and it was composed of 5 drifters that occupied the gyre interior. Cluster C consisted of 7 drifters, initialized in October 1995, which moved into both coastal and open sea regions. In particular, N_R, the average number of drifters within a circle of radius R ($= 10$ km) around the prediction particle, is different for these three clusters; for cluster A, $N_R \geq 2$ for 7 days, while for clusters B and C, $N_R < 1$ and $N_R < 0.5$, respectively, for 15 days. The results of OGMP, both analytical and numerical, showed that the data assimilation is effective only when $N_R \geq 1$. On this basis, it was expected that the assimilation is effective for cluster A, marginally effective for cluster B, and not effective for cluster C.

The OGMP algorithm was implemented using the mean flow estimated by Falco et al. (2000), and the values for the parameters T (Lagrangian e-folding

time scale) and R (Eulerian space scale) in the range of values suggested by the literature: $T \approx 1$ day, and $R \approx 10$ km. For each cluster, the prediction skills were quantified by calculating the r.m.s. position error averaged over all drifters in that cluster.

As a first step, the skill of the basic model (4)–(5) was determined, without including the data assimilation. This should be interpreted as the forecasting skill of the model, which is expected to last over a period comparable to T. The time scales for the three clusters are similar, on the order of 1 day, and the r.m.s. forecast error is approximately 5 km after 1 day. For times longer than 1 day, the forecast error increases quickly, reaching approximately 15 km after 2 days. This information is useful for practical applications, for example when forecasting the position of a pollutant, given its initial position and velocity. The model is of course very simple, based only on climatology and persistence, nevertheless its skills at initial times appear to be acceptable, even though they decay quickly due to the lack of sufficient dynamics. In the future, we may consider employing more complex models, in which the velocity field is provided by regional circulation models.

While the evolution model is simple, it is also ideally suited for the incorporation of drifter observations with minimal computational cost, complexity and also interference from the model. The impact of assimilating data has then been considered and it was shown that the main conclusions of OGMP remain valid for real drifter data: for cluster A, the r.m.s. prediction error with data assimilation, S, shows a decrease of more than 70% with respect to the prediction based only on the mean flow without data assimilation, S_M. S remains less than 20 km during the 7-day prediction. Also, the predictions appear to be robust even when the mean flow is assumed unknown, and when an error in the initial position is introduced, provided that the error is less than R. These results are again consistent with the numerical results of OGMP for $N_R > 1$. For cluster B, S decreases with respect to S_M, but the decrease is only marginally significant, while for cluster C no decrease is observed.

Finally, a "common sense" method of estimation was tested, where each trajectory in the cluster is approximated as the center of mass of the others. For this estimate, the error growth is expected to depend on the relative dispersion of the particles, characterized, for particle pairs, by the Lyapunov exponent (Piterbarg, 2001). For our clusters, this estimate is initially characterized by higher errors than the data assimilation estimate, converging to the same errors when $N_R \to 0$. Different results could be found, though, for different clusters where the initial sampling was optimized.

It is interesting to note that the prediction errors in this study are comparable to typical errors of forecasts through realistic numerical ocean models. Yang et al. (1999) used a model with high horizontal and vertical resolution to simulate directly drifters on the West Florida Shelf. They have shown a general agreement with the drifter observations but with an error increasing rapidly for first few days and hovering around 10 km during the first 10 days and finally reaching about 15 km after 12 days.

In the future, we intend to investigate the problem of optimal sampling for cluster deployments, in order to optimize resources and plan cluster replenishing. This study indicates that the results are highly dependent on the sampling adopted and that the response time scales are short, so that an optimization of the resources and of the methods is of great importance.

Acknowledgements

The authors acknowledge the support of the Office of Naval Research under grants N00014-97-1-0620, N00014-99-1-0049 and N00014-99-WR-30014. We thank E. Ryan and the three anonymous reviewers for constructive comments, which led to an improved manuscript.

References

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

Artegiani, A., Bregant, D., Paschini, E., Pinardi, N., Raicich, F., Russo, A., 1997. The Adriatic Sea general circulation: Part II. Baroclinic circulation structure. J. Phys. Oceanogr. 27, 1515–1532.

Bauer, S., Swenson, M.S., Griffa, A., Mariano, A.J., Owens, K., 1998. Eddy-mean flow decomposition and eddy-diffusivity estimates in the tropical Pacific Ocean: Part I. Methodology. J. Geophys. Res. 103 (C13), 30855–30871.

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

Falco, P., Griffa, A., Poulain, P.-M., Zambianchi, E., 2000. Transport properties in the Adriatic Sea deduced from drifter data. J. Phys. Oceanogr. 30 (8), 2055–2071.

Gacic, M., Marullo, S., Santoleri, R., Bergamasco, A., 1997. Analysis of seasonal and interannual variability of the sea surface temperature field in the Adriatic Sea from AVHRR data (1984–1992). J. Geophys. Res. 102 (C13), 22937–22946.

Ghil, M., Malanotte-Rizzoli, P., 1991. Data assimilation in meteorology and oceanography. Adv. Geophys. 33, 141–226.

Griffa, A., Owens, K., Piterbarg, L., Rozovski, B., 1995. Estimates of turbulence parameters from Lagrangian data using a stochastic particle model. J. Mar. Res. 53, 371–401.

Grilli, F., Pinardi, N., 1998. The computation of Rossby radii of deformation for the Mediterranean Sea. MTP News 6, 4.

Hansen, D.V., Poulain, P.-M., 1996. Quality control and interpolations of WOCE/TOGA drifter data. J. Atmos. Oceanic Tech. 13, 900–909.

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

Lacorata, G., Aurell, E., Vulpiani, A., 2000. Data analysis and modelling of Lagrangian drifters in the Adriatic Sea. Ann. Geophys., in press.

Özgökmen, T.M., Griffa, A., Mariano, A.J., Piterbarg, L.I., 2000. On the predictability of Lagrangian trajectories in the ocean. J. Atmos. Ocean. Tech. 17 (3), 366–383.

Piterbarg, L.I., 2001. The top Lyapunov exponent for a stochastic flow modeling the upper ocean turbulence. SIAM J. Appl. Math., in press.

Poulain, P.-M., 1999. Drifter observation of surface circulation in the Adriatic Sea between December 1994 and March 1996. J. Mar. Sys. 20, 231–253.

Poulain, P.-M., Zanasca, P., 1998. Drifter and oat observations in the Adriatic Sea (1994–1996)—Data Report. SACLANTCEN Memorandum, SM-340. SACLANT Undersea Research Centre, La Spezia, Italy, 46 pp.

Poulain, P.-M., Zanasca, P., 1999. Lagrangian measurements of surface currents in the northern and middle Adriatic Sea. In: Hopkins, T.S. (Ed.), The Adriatic Sea. Ecosystem Research Report No. 32, EUR 18834. European Commission, Brussels, pp. 107–115.

Robinson, R.R., Lermusiaux, P.F.J., Sloan III, N.Q. et al., 1998. Data assimilation. In: Brink, K.H., Robinson, A.R. (Eds.), The Sea. Wiley, pp. 541–594.

Samelson, R.M., 1996. Chaotic transport by mesoscale motions. In: Adler, A.J., Muller, P., Rozovskii, B. (Eds.), Stochastic Modelling in Physical Oceanography. Birkhäuser, pp. 423–438.

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., Weisberg, R.H., Niiler, P.P., Sturges, W., Johnson, W., 1999. Lagrangian circulation and forbidden zone on the West Florida Shelf. Cont. Shelf Res. 19, 1221–1245.