ELSEVIER

Lagrangian data in a high-resolution numerical simulation of the North Atlantic
II. On the pseudo-Eulerian averaging of Lagrangian data

Zulema D. Garraffo [a,*], Annalisa Griffa [a,b], Arthur J. Mariano [a], Eric P. Chassignet [a]

[a] *RSMAS / MPO, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA*
[b] *IOF / CNR, La Spezia, Italy*

Abstract

In this paper, the statistical properties of the mean flow reconstruction using Lagrangian data are studied, considering the classical "binning" approach based on space-time averaging of finite difference velocity estimates. The work is performed numerically, using as the test flow a solution from a high resolution MICOM simulation of the North Atlantic. A set of trajectories are computed, simulating the motion of surface drifters initially launched on a regular $1° \times 1°$ array, transmitting positions every $\Delta t = 12$ h, and analyzed over approximately 2 years of the simulation. The drifter distribution in time is influenced by the Ekman flow, resulting in maximum data concentration in the subtropical convergence regions and minimum concentration in the upwelling regions.

Pseudo-Eulerian averages U_{pE}, computed from Langrangian data, are compared to "true" Eulerian averages U_E, computed from grid point velocities inside $1° \times 1°$ bins for approximately 2 years. For the full Lagrangian data set (which is substantially larger than the WOCE requirement), $U_{pE} - U_E$ is on the order of 10–20 cm/s in regions of major ocean currents. These differences are usually not significant with respect to the sampling error, due to subgrid-scale variability and finite sampling, except in a few regions. Patterns of the magnitude of the differences between U_{pE} and U_E in these regions show that U_{pE} tends to underestimate (overestimate) the velocity in the eastern equatorial upwelling regime/South Equatorial current (western boundary currents). This study suggests that these under/overestimates by pseudo-Eulerian averaging of Lagrangian data are related to a bias due to mesoscale divergences, and result in nonzero correlations between instantaneous drifter concentration and velocity, $\hat{U}_B = \langle u'c' \rangle /C$ (Davis, 1998; Gent and McWilliams, 1990). In this framework, the overestimates (underestimates) are interpreted as due to preferential (reduced) sampling of high velocity regions by Lagrangian particles, due to convergent (divergent) phenomena. A similar phenomenon has been observed for real drifters and biological organisms. The overestimates are found to increase with sub-sampling in space and decrease with sub-sampling in time. For $\Delta t = 3$ days, we actually find underestimates, probably because instantaneous high velocities are smoothed and energetic drifters are not appropriately accounted for in the bins. Direct implications of the results for the analysis of real data, and directions for future work (in particular investigation of the bias) are discussed. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Lagrangian data; High resolution numerical simulation; North Atlantic

---

* Corresponding author. Tel.: +1-305-361-4882.
  E-mail address: z.garraffo@miami.edu (Z.D. Garraffo).

1. Introduction

Lagrangian data play an important role in our understanding of ocean circulation and dynamics. Lagrangian floats can provide extensive sampling in both the horizontal and the vertical, a unique feature among the presently available observing platforms. Satellite data provide excellent horizontal and temporal coverage but they are confined to the ocean surface, while hydrographic or current meter data can provide high vertical resolution information, but cannot easily achieve good horizontal coverage. Lagrangian data are an important component of WOCE, providing direct information on velocity, transport, and also (when having profiling capability) on vertical stratification. Here, we focus on the velocity information, and on how it can be used to estimate the mean Eulerian velocity field U.

In a number of papers (e.g., Rossby et al., 1983; Owens, 1991; Brügge, 1995; Davis, 1998), Lagrangian data are used to estimate ocean mean flows. The mean flow reconstruction, often indicated as "pseudo-Eulerian" averaging, is usually performed by averaging over certain spatial bins and over certain time periods, using all the available velocities computed from Lagrangian data by finite differencing of the buoy positions. While the "binning" reconstruction is technically very simple, it can be conceptually delicate and can lead to errors or biases with respect to the true Eulerian mean velocity field, as discussed by a number of authors (e.g., Davis, 1991; Mockett, 1999). These biases can be related to sampling and dynamical inhomogeneities in space and time, as well as to sampling resolution. The impact of such biases has been tested in previous studies for simple idealized flows, and it has been shown that sampling biases have the potential to contaminate pseudo-Eulerian estimates of the mean flow. This is an important issue, given the relevant role of Lagrangian data in our knowledge of the global ocean circulation.

In this paper, an analysis of the estimation errors for calculating mean Eulerian velocities from Lagrangian data, as a function of horizontal and temporal resolution of the Lagrangian data, is performed using a "realistic" numerical flow, i.e., a flow with sufficient geographical and dynamical complexity to be considered as a "controlled laboratory" of a real oceanic flow. The obvious advantage of using numerical data is that the true Eulerian mean is known, so that the pseudo-Eulerian averaging errors can be directly evaluated. At the same time, the flow is complex enough so that realistic issues and problems can be identified and studied. Our experimental Lagrangian data densities are, at a minimum, in the range proposed by WOCE, i.e., an instantaneous horizontal coverage of approximately one drifter per $5° \times 5°$ for several years. Pseudo-Eulerian averaging errors, as a function of space and time sampling, are estimated for simulated drifter trajectories in the tropical and North Atlantic.

The numerical flow that we consider is a high resolution $(1/12°)$ solution of the Miami Isopycnic Coordinate Model (MICOM) configured for the North Atlantic. Numerical drifters are launched in the surface mixed layer, with an initial resolution of 1 drifter per $1° \times 1°$, and their trajectories are calculated for a period of approximately 2 years. In a companion paper (Garraffo et al., 2001, hereafter G01), statistics from the numerical drifters are compared with those from real in situ drifters, showing that the numerical results capture most of the characteristics of the mean flow estimated from the data. More significant differences are found regarding the eddy field statistics, with the model underestimating eddy kinetic energy and overestimating Lagrangian velocity time scales. This is likely due mostly to the lack of high frequency forcing in the model and to the fact that there is no vertical shear in the bulk Kraus–Turner mixed layer (Chassignet et al., 2001).

In this study, only numerical drifters are considered, and the focus is on a comparison between pseudo-Eulerian and Eulerian estimates of the mean flow U. First, pseudo-Eulerian estimates of U are computed considering the complete Lagrangian data set, i.e., all of the data obtained from the drifters seeded at $1°$ resolution in longitude and latitude and "tracked" for approximately 2 years with a temporal resolution of $\Delta t = 12$ h for recording position data. The total number of these data approximately corresponds to an average coverage of $5° \times 5°$ for 50 years, definitely more than the WOCE requirement, providing therefore an upper limit to realistic resource availability. The pseudo-Eulerian mean velocity field estimated from these data is compared with the "true" Eulerian mean velocity field that is com-

puted directly from all of the model surface velocity data, and the nature of different sources of error is considered.

The effect of sub-sampling in space and time is then considered. In space, the number of drifters is decreased by decreasing the horizontal resolution in the initial drifter releases to 1 drifter per $5° \times 5°$. This last case can be considered as a lower limit for the WOCE resolution, corresponding to approximately two drifter-years per $5° \times 5°$ over the 2-year analysis period. In time, the temporal resolution is decreased from one particle position every 12 h to one particle position every 3 days. One update in position every 3 days is the sampling rate of many of the WOCE drifters.

The paper is organized as follows. In Section 2, the model and the Lagrangian data set are briefly described. In Section 3, the basic methodology and error analysis are discussed. A comparison between pseudo-Eulerian and Eulerian estimates and the effects of sub-sampling are presented in Section 4. A summary and a discussion are found in Section 5.

2. The numerical model and the Lagrangian data set

A very high resolution North Atlantic ocean simulation has been carried out for 20 years with the Miami Isopycnic Coordinate Ocean Model (MICOM; Bleck et al., 1992; Bleck and Chassignet, 1994). The computational domain is the North and Equatorial Atlantic Ocean basin from 28°S to 65°N, including the Caribbean Sea and the Gulf of Mexico. The bottom topography is derived from a digital terrain data set with 5′ latitude–longitude resolution (ETOPO5). The surface boundary conditions are based on the COADS monthly climatological data sets (Da Silva et al., 1994). The simulation allows estimation of the natural variability of the model. This is a necessary first step to later assess the impact of daily forcing (such a simulation is in progress). Open ocean boundaries are treated as closed, but are outfitted with 3° buffer zones in which temperature, T, and salinity, S, are linearly relaxed toward their seasonally varying climatological values (Levitus, 1982). The buffer zones restore the T and S fields to climatology in order to approx-

imately recover the vertical shear of the currents through geostrophic adjustment.

The horizontal grid (6 km on average) is defined on a Mercator projection with resolution given by $1/12° \times 1/12° \cos(\phi)$, where ϕ is the latitude. The vertical density structure is represented by 15 isopycnic layers, topped by a dynamically active Kraus–Turner surface mixed layer that exchanges mass and properties with the interior layers (see Bleck et al., 1989, for details). The vertical discretization was chosen to provide maximum resolution in the upper part of the ocean. The computational requirements for basin-scale ocean modeling at this resolution are extreme and demand the latest in high performance computing (Bleck et al., 1995). The model was spun up from the rest for a total of 20 years.

The high horizontal grid resolution drastically improved the model's behavior in comparison to that of previous coarse-resolution simulations. The major improvements are (a) a correct Gulf Stream separation and (b) higher eddy activity. These results support the view that an inertial boundary layer, which results from the fine resolution, is an important factor in the separation process (Özgökmen et al., 1997), and that resolution of the first Rossby radius of deformation is necessary for a correct representation of baroclinic instabilities. The simulated Gulf Stream is highly inertial, meanders strongly, and sheds several cyclonic and anti-cyclonic rings. Paiva et al. (1999), G01 and Chassignet et al. (2001) discuss the strengths and weaknesses of this simulation.

At the beginning of model year 14, a total of 25,000 numerical particles are launched at the surface (in the mixed layer) and at depths of 400, 1000, 1500 and 3000 m, in a regular $1° \times 1°$ grid (~ 5000 particles at each level). The trajectories and diagnostics are then computed for the 2-year period. The focus here is on the surface particles whose trajectories are calculated using the modeled surface velocity fields that are representative of the depth-averaged mixed layer, which varies both seasonally and with latitude. The mixed layer depth is generally between 20 and 100 m, but can occasionally reach 2000 m in water in the high latitudes during deep water formation events (Smith et al., 2000). The numerical particle advection scheme is second-order Runge–Kutta, with 16-point space interpolation in

the ocean interior and 4-point near the coasts. Particles are advected with a 2-h time step. Off-line tests with a frozen velocity field suggest that, for eddies and currents, errors due to the interpolation scheme are negligible at least for scales on the order of 100 days. The advection inside the first grid box from the coast is accomplished by interpolation, imposing zero velocity at the first point inside land. Consequently, in the first half-grid box from the coast, the Lagrangian particles have a slightly larger along coast velocity than that which would be obtained by linear interpolation from the model with no-slip boundary conditions. This choice was made because of its simplicity and because it minimized the loss of particles near the coast on the computational C-grid. Very few cases of "beaching" were observed with this boundary condition (less than 3%). Most of the particles that "beached" followed successive corners along the coast, suggesting errors in particle advec-

tion due to space interpolation and time extrapolation.

The total number of surface position data in our model ocean for the 2-year analysis period is about 3,600,000 drifter-days (Fig. 1a), or approximately 20 times larger than the near-surface drifter set for the same region as the model, and during the time period 1989–1998, archived at the WOCE/CLIVAR Drifter Data Assembly Center at NOAA-AOML, as discussed in G01. It is quite evident in Fig. 1a that there is preferential sampling over the 2-year time period even though the initial "release" of numerical drifters was uniform in space. The highest sampling densities (about 2000 drifter-days per $1° \times 1°$) are found in the subtropical convergence zones and the lowest densities in the equatorial divergence zones, due to wind-driven Ekman transport. A similar tendency can be seen also in the distribution of the observed in situ drifters (G01), even though the phenomenon is less
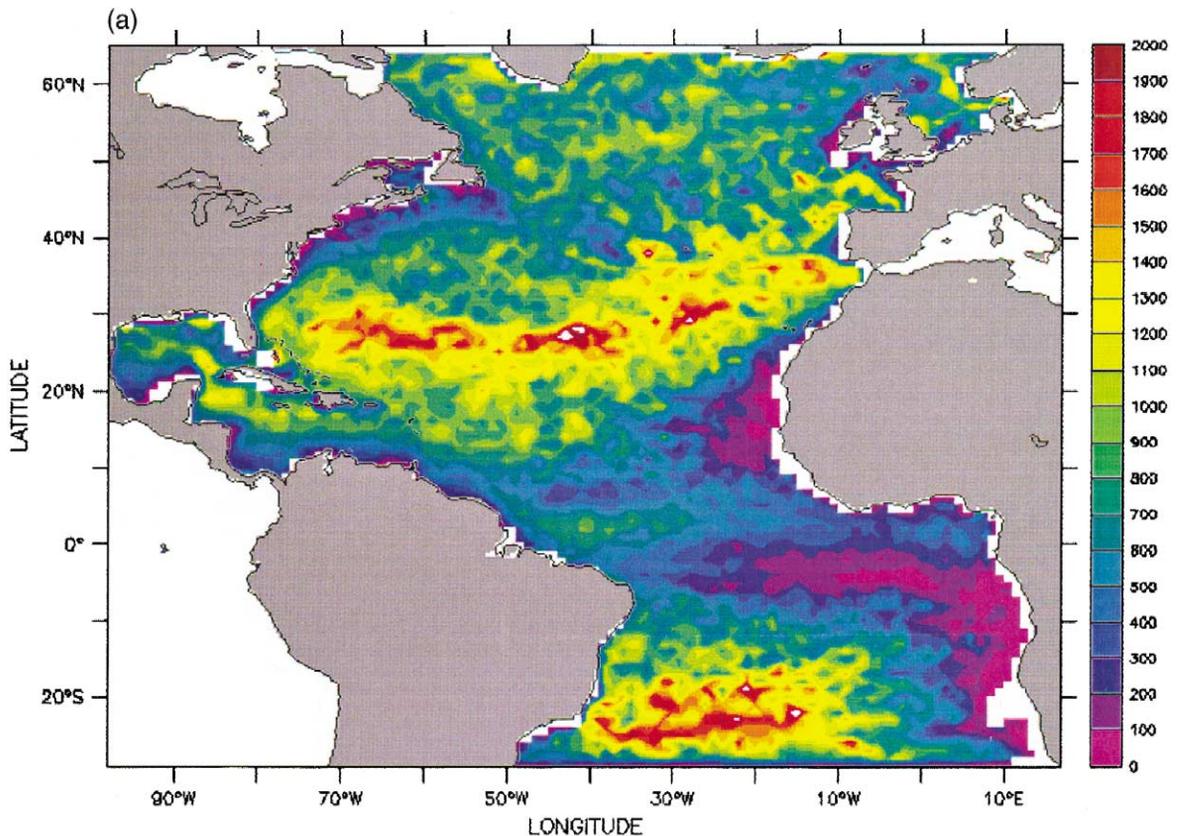

Fig. 1. Number of drifter-days/deg^2, for (a) all the numerical drifters (initial seeding on a $1° \times 1°$ grid); (b) initially seeded on a $5° \times 5°$ grid.
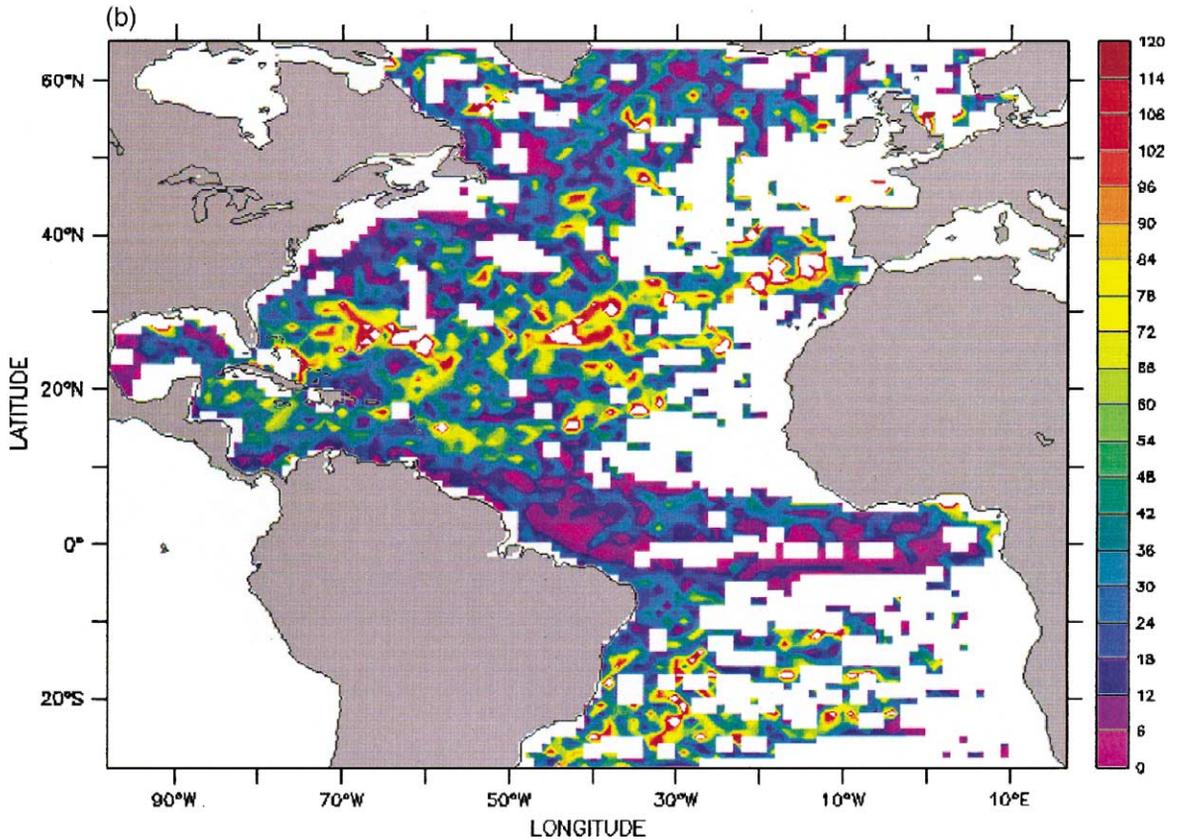
Fig. 1 (*continued*).

evident because the initial distribution was not homogeneous and the lifetime of the in situ drifters is considerably shorter (238 days on average, with many drifters living less than 200 days).

Fig. 1b shows the data density for the spatially subsampled (one drifter initially seeded every $5° \times 5°$) data set used in our analysis. The total number of position data is 144,000 drifter-days. In most of our analysis domain, the density does not exceed 100 drifter-days per $1° \times 1°$, with almost no observations in the eastern boundary upwelling regions and at subpolar latitudes.

3. Methodology and error estimates

The Eulerian, U_E, and pseudo-Eulerian, U_{pE}, mean velocity fields for the full data set are estimated as area-time averages computed over $1° \times 1°$ bins and

over the 2-year period. The U_E estimates are computed using all the velocity data u from all of the grid points in each bin during the 2-year period. The Eulerian "truth" data are instantaneous velocities measured once per day and are at $1/12°$ horizontal resolution u. The average number of grid points per bin is 181, so that the average total number of observation day per bin is approximately 130,300. The number of independent measurements is of course smaller, and depends on the time and space scales of the velocity field.

The pseudo-Eulerian average estimates U_{pE} are computed using the velocity estimates U_{pE} calculated from the Lagrangian trajectory positions by finite differencing. Velocities are computed at the midpoint between two successive particle locations. The temporal resolution of the position data is $\Delta t = 12$ h for the full data set, and 1 and 3 days for the temporal sub-sampled data sets. All of the position data for

drifter trajectories falling into each bin during the 2-year period are considered. Even though the initial sampling is uniform, the total data density in drifter-days/bin (Fig. 1) varies in space, as a consequence of the large scale Ekman divergence pattern, as discussed in G01. For the full data set, with an initial release at 1° horizontal resolution, the maximum number of drifter-days/bin is 2000, occurring in a restricted area in the center of the convergence region (Fig. 1a), while for the spatially sub-sampled data set at 5° horizontal resolution, the maximum data densities are around 120 drifter-days/bin (Fig. 1b).

In the following, the nature of the statistical errors expected to occur in the mean velocity estimates is discussed.

3.1. Estimation errors

The Eulerian estimator is an unbiased one so that no biases are expected in the estimates U_E. On the other hand, U_E is expected to be affected by sampling errors due to finite sampling and subscale variability. In the comparison with U_{pE} performed in Section 4, however, this error is neglected and U_E is considered as the "true" value of the average Eulerian velocity, since the number of data points in the Eulerian estimates is consistently much larger than the number of pseudo-Eulerian estimates.

For the pseudo-Eulerian estimates, the finite sampling error, expected to be larger because of the lower sampling density, is explicitly considered in the comparison. The root-mean-square error for each velocity component is given by (Riser and Rossby, 1983; Davis, 1991)

$$Er(U_{pE}) = \sigma/\sqrt{\left(N_{pE}^*\right)}, \tag{1}$$

where σ is the square root of the velocity variance and N_{pE}^* is the number of independent measurements. N_{pE}^* can be approximated as (Riser and Rossby, 1983; Owens, 1991)

$$N_{pE}^* \approx N_{pE}\Delta t/2T, \tag{2}$$

where N_{pE} is the total number of measurements, Δt is the sampling interval, and T is the Lagrangian integral time scale. In Eq. (2), for simplicity, only

the time correlation along trajectories is considered, while the space correlations of concomitant drifters are neglected.

In addition to sampling errors, bin-averaged Lagrangian estimates U_{pE} are also potentially affected by bias errors, so that, even in the limit of high data density, they do not necessarily converge to true Eulerian estimates. As discussed by Davis (1991), the bias error, or difference between the pseudo-Eulerian (tracer) and Eulerian velocity, can be expressed as

$$\hat{U}_B = U_{pE} - U_E = \langle u'c' \rangle/C, \tag{3}$$

where for each bin, $\langle \ \rangle$ indicates the ensemble-space average, $c' = c - C$, c is the instantaneous drifter concentration, and C is the average concentration. Eq. (3) states that since U_{pE} is computed as the flux of marked particles, nonzero correlations between particle concentration and velocity result in a bias.

There are several mechanisms that can lead to non-zero correlation in Eq. (3). Some of them are related to eddy dispersion, in the presence of nonuniform eddy diffusivity K or nonuniform concentration $C(x)$. Davis (1991) has investigated these mechanisms using a flux vs. gradient-history law (Davis, 1987) to parameterize the form of $\langle u'c' \rangle/C$. Two simple limiting forms to describe $\hat{U}_B$ have been derived, corresponding respectively to the known "diffusion bias" and "array bias".

Diffusion bias (e.g., Freeland et al., 1975), $\hat{U}_K$, is the limiting form of Eq. (3) that emphasizes the effects of spatial inhomogeneities in the diffusivity tensor K, and is obtained neglecting spatial variations of U_E, and of $\nabla \cdot K$

$$\hat{U}_B \approx \hat{U}_K = \nabla \cdot K. \tag{4}$$

Eq. (4) expresses the fact that, in deployments from a point, particles tend to move farther toward regions at high K, inducing a mean particle motion.

Array bias, $\hat{U}_C$, is the limiting form of Eq. (3) describing the effects of gradients of C, and is obtained considering $C(x)$ steady in time.

$$\hat{U}_B \approx \hat{U}_C = -K \cdot \nabla \ln(C) = -K \cdot \frac{\nabla C}{C}. \tag{5}$$

Eq. (5) represents a bias toward low data concentration regions, and it expresses the fact that particles

moving randomly tend to move away from their positions at any given time. As a consequence, if more particles are, on the average, concentrated in one region, it would appear that more particles move away from that region under the action of turbulent dispersion, inducing a down gradient bias in the mean estimate.

Notice that two limiting forms (4) and (5) do not strictly apply to our case, since the underlying assumptions are not exactly met. In particular, the large-scale concentration C is slowly varying in time in our results, because of the divergent nature of the large scale flow, in which case Eq. (5) would not be valid. Also, Eqs. (4) and (5) describe specifically the effects of eddy dispersion, while in our application dispersion processes are expected to be affected also by shear, i.e., by spatial variation of the mean flow $U_E(x)$ (e.g., Zambianchi and Griffa, 1994). Nevertheless, expressions (4) and (5) are of potential interest because they indicate bias mechanisms related to the eddy field structure, and they will be explicitly estimated and further investigated in the following.

Another mechanism potentially leading to nonzero correlations in Eq. (3) is related to mesoscale divergence processes in the flow. The instantaneous concentration c inside the averaging bin is expected to vary as a function of the local divergence. If u' is correlated to the divergence, this implies a preferential sampling and therefore a bias velocity. In order to express this effect in terms of observable variables, Davis (1998) considers the ideal case of a layer of fluid bounded by material surfaces and characterized by instantaneous thickness h, with average H, and instantaneous deviation from the average h'. If the layer is uniformly and densely (per area) sampled by drifters, it is expected that $c/C = h/H$. In this ideal case, Eq. (3) can be expressed as a "divergence" bias

$$\hat{U}_B = \langle u'c' \rangle / C \approx \hat{U}_D = \langle u'h' \rangle / H. \qquad (6)$$

In our case, the situation is not necessarily so simple, because the mixed layer cannot be considered as a bounded layer, since other nonconservative phenomena, such as turbulent mixing and surface mass flux, can occur. Also, the sampling is not as dense as assumed in Eq. (6). Despite this, as for the other limiting forms (4) and (5), Eq. (6) will be

considered in the following as illustrative of a specific mechanism to be investigated.

Notice that the component of the bias related to divergence process, $\hat{U}_D$, cannot be completely eliminated by appropriate sampling, as could be argued for the array bias Eq. (5). This bias expresses the conceptual difference between tracer advection velocity and average Eulerian velocity, and as noted by Davis et al. (1996) and Davis (1998), is also related to the eddy transport as parameterized by Gent and McWilliams (1990).

In addition to Eq. (3), other possible sources of bias errors exist, for example sources of error related to sample resolution. Mockett (1999) has shown, in a numerical study of a meandering jet, that the time sampling interval Δt plays an important role, and that estimates using data with large Δt tend to underestimate the mean velocity. As Δt increases, estimation variance decreases at the expense of temporal resolution and biased estimates of the mean. This is the well known (in the field of statistical optimization and estimation) variance-bias trade-off. This error is evaluated below, for $\Delta t = 3$ days.

Also, additional errors might arise from the numerical algorithm used for calculating the simulated particle trajectories. Simulated Lagrangian trajectories are, in fact, computed by interpolating Eulerian velocities u, and it is possible that the interpolation introduces biases in trajectory positions and in u_{pE}. We have checked for this source of error by considering an ensemble of random points in space and time, and by comparing u and u_{pE}, where u_{pE} is computed with $\Delta t = 12$ h. The values agree to within 1%, indicating that the numerical error is negligible.

3.2. A-priori estimates of error sizes

It is useful to have estimates of the expected errors to provide guidance in the interpretation of the results. In the following, the sampling errors, $Er(U_{pE})$ from Eq. (1), are evaluated for the full data set (initial seeding on the $1° \times 1°$ grid). Estimates for the bias errors (Eq. (3)) cannot be computed directly "a-priori", i.e., as functions of the average macroscopic variables describing the flow and the drifter distribution, σ, C, K. Only the approximations (4) and (5), related to eddy dispersion effects, can be

computed in terms of macroscopic quantities, and they will be computed in the following. An estimate of the approximation (6), related to mesoscale divergence effects, will be computed "a-posteriori" in Section 4.1.

The magnitude of the vector sampling error is illustrated in Fig. 2a and is computed by assuming no significant cross-correlation between velocity components, applying Eq. (1) to each velocity component. The sampling error depends on the relative size of σ, characterizing the subgrid-scale variability, and N_{pE}^* the number of independent measurements. The distribution of N_{pE}^* depends on the data density (Fig. 1a) and on the Lagrangian integral time scales, T_u and T_v. Lagrangian time scales have been estimated in G01 for the full data set, and the results are presented in Fig. 3, while details of the estimation method are given in Appendix A. Fig. 3a,b show that T has smaller values, O (3 days), close to

the western boundary and larger values, up to 15 days, close to the eastern boundary where the flow is unrealistically smooth (see G01 for more details). The distribution of the velocity variance is illustrated by r.m.s. eddy velocity/eddy kinetic energy in Fig. 4.

As shown in Fig. 2a, the sampling error is the highest in the equatorial currents, especially the South Equatorial Current, and in the North Brazil Current retroflection region, with values on the order of 10 cm/s. This is not surprising given that these regions are highly energetic and have low data density. Also the Gulf of Mexico, Caribbean, and Gulf Stream regions have high errors in the order of 5–6 cm/s. Energetic regions and upwelling regimes have the largest sampling errors in our numerical simulations. Sampling errors are significantly lower in the ocean interior, approximately 1 cm/s or less, due to low eddy variability and high data density.
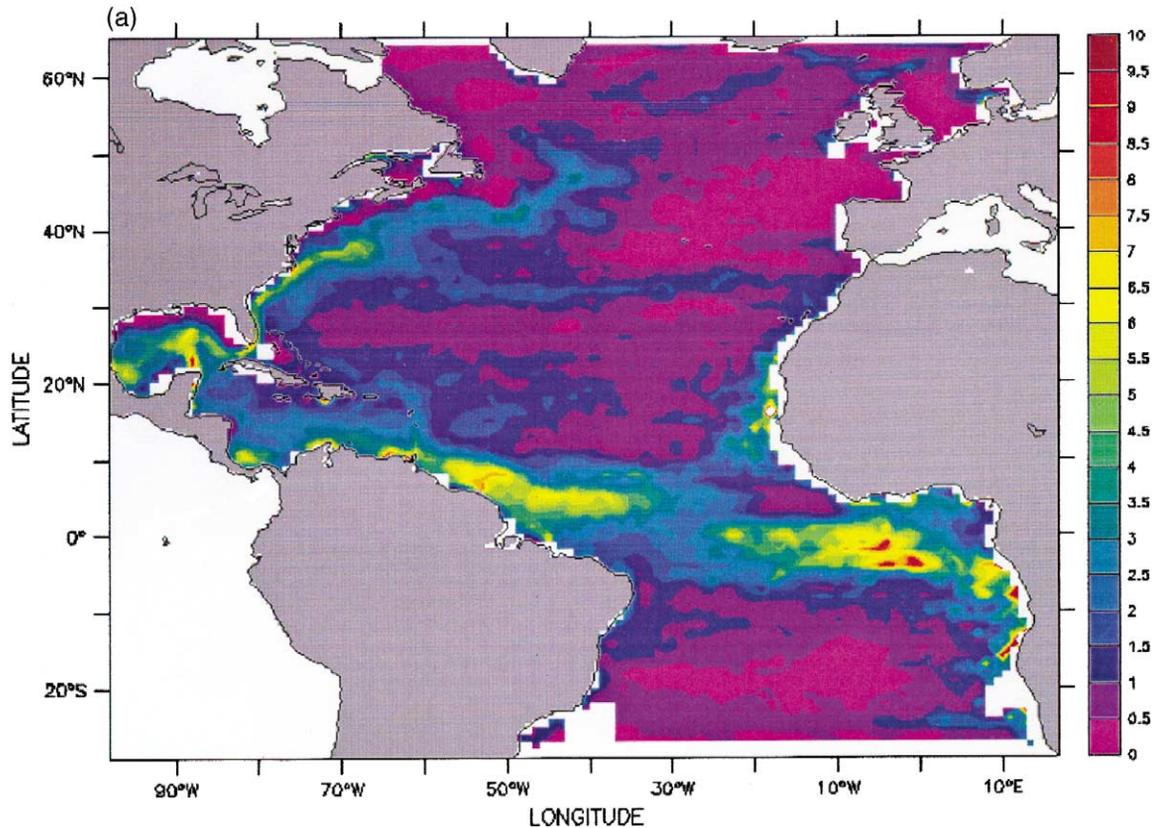

Fig. 2. (a) Magnitude of the vector sampling error (cm/s). (b) Diffusivity bias $\hat{U}_K$ (Eq. (4)). (c) Array bias $\hat{U}_C$ (Eq. (5)).

(b) $\hat{U}_K$

0.05 cm/s

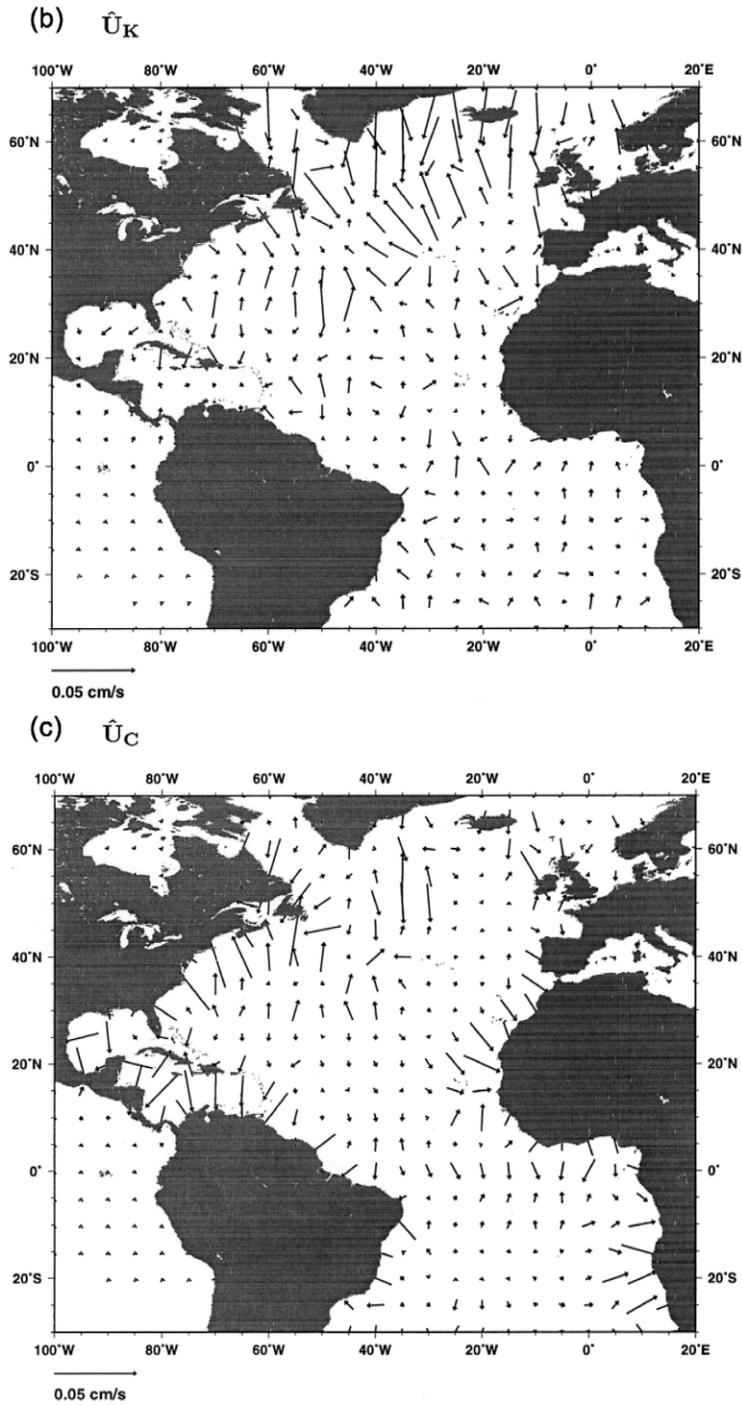(c) $\hat{U}_C$

0.05 cm/s

Fig. 2 (*continued*).

Estimates of the bias expressions (4) and (5) are shown in Fig. 2b,c. The diffusivity term K is computed using the Taylor (Taylor, 1921) parameterization $K = \sigma^2 T$, while the C and K derivatives are
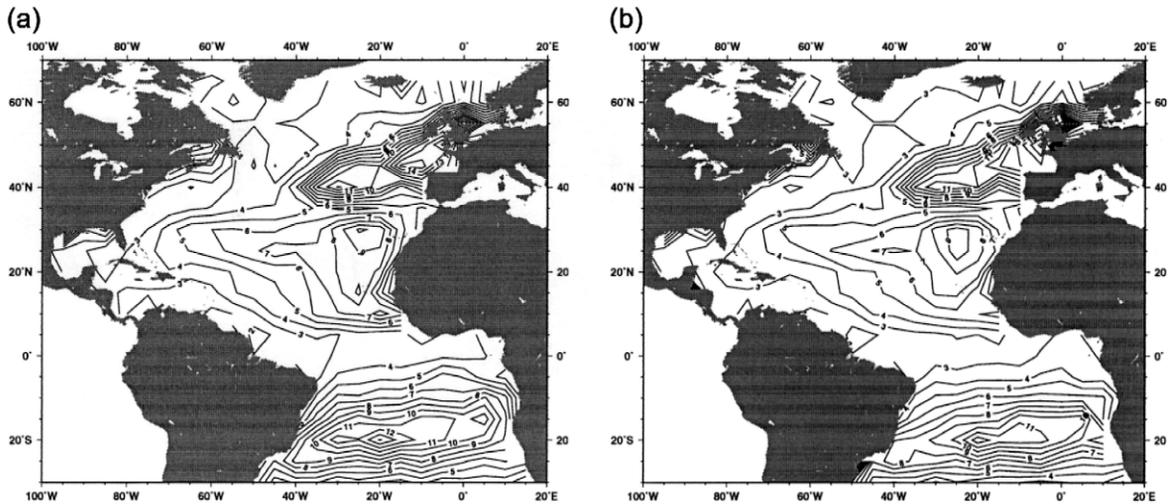
Fig. 3. Lagrangian velocity integral time scale in days for the (a) zonal and (b) meridional directions, for all model drifters in the region.

estimated using a least-squares bi-cubic smoothing spline fit to the fields (see G01). Notice that the estimates of K based on Taylor's parameterization probably underestimate the effective diffusivity values, especially in regions where the shear is not completely resolved by the mean flow estimates, and residual shear is present inside specific bins.

The diffusivity bias vector $\hat{U}_K$ (Eq. (4)) (Fig. 2b) shows a clear pattern of convergence along the major currents such as the Gulf Stream near 40°N, the North Atlantic Current east of Newfoundland and south of Greenland, and the South Equatorial Current. Maximum values are on the order of 5×10^{-2} cm/s, definitely smaller than the sampling error, especially in the major currents where the difference between Eulerian and pseudo-Eulerian mean speeds is greater than the bias by two orders of magnitude.

The array bias vector $\hat{U}_C$ (Eq. (5)) (Fig. 2c) shows a down gradient pattern with respect to C, indicating a subtropical divergence and equatorial convergence (opposite to the Ekman flow), as well as a convergence toward the boundary currents. An order of magnitude estimate for the region north of the subtropical gyre, which as a large concentration gradient, is 10^{-2} cm/s (with $K = 10^6$ cm^2/s). Thus, the bias errors for our simulations are expected to be much smaller than the error due to finite sampling, and this is seen to be the case in Fig. 2c, indicating values of the same order as in Fig. 2b.

Estimates of $\hat{U}_C$ are unreliable near land, since data density is usually lower in one degree bins containing land. This results in an increase in the numerator of Eq. (3) because of the larger gradient, and a decrease in the denominator due to the lower values of C. This leads to over-estimates of $\hat{U}_C$ near land.

Summarizing, the bias estimates (4) and (5) provide values consistently smaller than the sampling errors, especially in the strong current regions. This might be partially due to the underestimate of dispersion effects related to shear dispersion. There are only minor qualitative similarities between Fig. 2b and c. It may require significantly more simulated drifters whose trajectories are integrated over many more years before a quantitative comparison between the two bias errors is possible.

The sampling errors are expected to be larger for the subsampled data sets given the decreased number of data, while the bias errors are expected to have a similar pattern, given the similar patterns of C (Fig. 1a,b) and K. This expectation is confirmed below.

4. Comparison between eulerian and pseudo-Eulerian estimates

A comparison between Eulerian, U_E, and pseudo-Eulerian, U_{pE}, estimates of the mean velocity

field U is now presented. We also show the root mean square of the fluctuation velocity field. The fluctuation velocity in each bin is given by $u'_E(x,y,t) = u(x,y,t) - U_E(x,y)$ and $u'_{pE}(x,y,t) = u_{pE}(x,y,t) - U_{pE}(x,y)$ for Eulerian and pseudo-Eulerian estimates, respectively. The r.m.s. of the fluctuation velocity is related to the eddy kinetic energy EKE by $u'_{rmsE} = \sqrt{\langle u'^2_E \rangle + \langle v'^2_E \rangle} = \sqrt{2EKE_E}$ and $u'_{rmspE} = \sqrt{\langle u'^2_{pE} \rangle + \langle v'^2_{pE} \rangle} = \sqrt{2EKE_{pE}}$. The magnitude of the Eulerian mean flow, $|U_E|$, and of the eddy r.m.s. velocity, u'_{rmsE}, are shown in Fig. 4a and b. We show a double scale in all the figures, which allows us to identify the mean flow kinetic energy $(MKE_E = 0.5(U^2_E + V^2_E)$, $MKE_{pE} = 0.5(U^2_{pE} + V^2_{pE}))$ and the eddy kinetic energy. These fields are used as the reference "truth" fields for the analysis presented in this section. A quantitative comparison, using a statistical test, is performed between U_E and U_{pE}, to evaluate the significance of the differences with respect to the sampling error. The bias effects are investigated considering the patterns of the difference between estimated U_E and U_{pE}.

Results from the full data set (initial release of 1 drifters per $1° × 1°$, $\Delta t = 12$ h) are presented first. This is our "benchmark", in which the main results and the nature of the various errors are highlighted and discussed in detail. We consider the effects of subsampling, first in space and then in time.

4.1. Results for the full data set

The pseudo-Eulerian estimates for the full data set are shown in Fig. 4c,d for $|U_{pE}|$ (MKE_{pE}) and u'_{rmspE} (EKE_{pE}), respectively. By inspection, the fields appear very similar to the Eulerian fields in Fig. 4a,b. The only easily detectable difference is that the energy is slightly higher in the Eulerian estimates in the equatorial region, both in the mean and in eddy kinetic energy. Also, it appears that the maximum velocity in the strong western boundary currents is slightly higher in the pseudo-Eulerian estimates of the mean, especially in the Gulf of Mexico.

In order to explore more closely the differences in mean flow estimates, the difference $|U_{pE} - U_E|$ has been computed and is shown in Fig. 5a. As can be seen, differences in the order of 10 cm/s are present

in the strong current regions, reaching occasional maxima of about 20 cm/s. In the more quiescent regions, the differences are on the order of 1 cm/s or less.

The first question that we address is whether or not these differences can be considered significant given the sampling error, i.e., given the uncertainty due to finite sampling and subgrid-scale variability. As a first qualitative step, the difference between Eulerian and pseudo-Eulerian mean velocity estimates (Fig. 5a) is compared to the magnitude of the sampling error (Fig. 2a). The patterns appear similar, with maximum sampling errors on the order of 10 cm/s. Assuming a Gaussian distribution, this would imply a 95% confidence level of 20 cm/s, which is on the same order as the maximum differences in Fig. 5a. This suggests that the estimates cannot be considered significantly different in most of the cases. The evaluation above is a rough one, which considers only the average size of differences and does not properly take into account the vectorial nature of the variables and the spatial distribution of the differences and the sampling error.

A vectorial statistical test is therefore performed to evaluate the significance of the differences between the pseudo-Eulerian and Eulerian mean flow estimates, considering the Eulerian estimate U_E as "truth". The test is based on the statistic

$$\Lambda^2 = N(U_{pE} - U_E)^\dagger \mathbf{S}^{-1}(U_{pE} - U_E),$$

where † indicates a matrix transpose, $\mathbf{S}$ is the covariance matrix for the pseudo-Eulerian velocity, and N is the number of independent data points in a given bin, $N = N^*_{pE}$ of Eq. (2). The quantity

$$\frac{N-2}{2(N-1)}\Lambda^2,$$

has an F-distribution with degress of freedom 2 and $N - 2$ (Morrison, 1976). If the magnitude of the test statistic Λ^2 is greater than the value of

$$\frac{2(N-1)}{N-2}F^{0.05}_{2,N-2},$$

with $F^{0.05}_{2,N-2}$ being the critical value at 95% confidence for $(2, N-2)$ degrees of freedom, then the mean flow estimates are different at the 95% confidence level. The values of $F^{0.05}_{2,N-2}$ are interpolated

for a given N from the tables of Abramowitz and Stegun (1964).

The results, shown in Fig. 5b, indicate that the differences are indeed not significant at the 95% level over most of the basin. There are some isolated regions, though, especially near the Equator and the Gulf Stream region, where the differences appear significant, indicating that error sources other than the sampling uncertainty come into play.

Next, we investigate the presence of possible biases. Because of the sampling error contamination, they cannot be computed directly as differences be-

tween pseudo-Eulerian and Eulerian means in the 1° bins. As a consequence, an indirect indicator is considered, given by the pattern of the absolute value differences, $|U_{pE}| - |U_E|$. In the absence of biases, the spatial distribution of the values of $|U_{pE}| - |U_E|$ is not expected to have a definite sign. The presence of a predominant sign in a given region, then, is considered as a bias indicator.

The pattern of $|U_{pE}| - |U_E|$ is shown in Fig. 6a. It indicates that the pseudo-Eulerian averages tend to underestimate the mean flow in the eastern Southern Equatorial Current and in the tropical basin, while
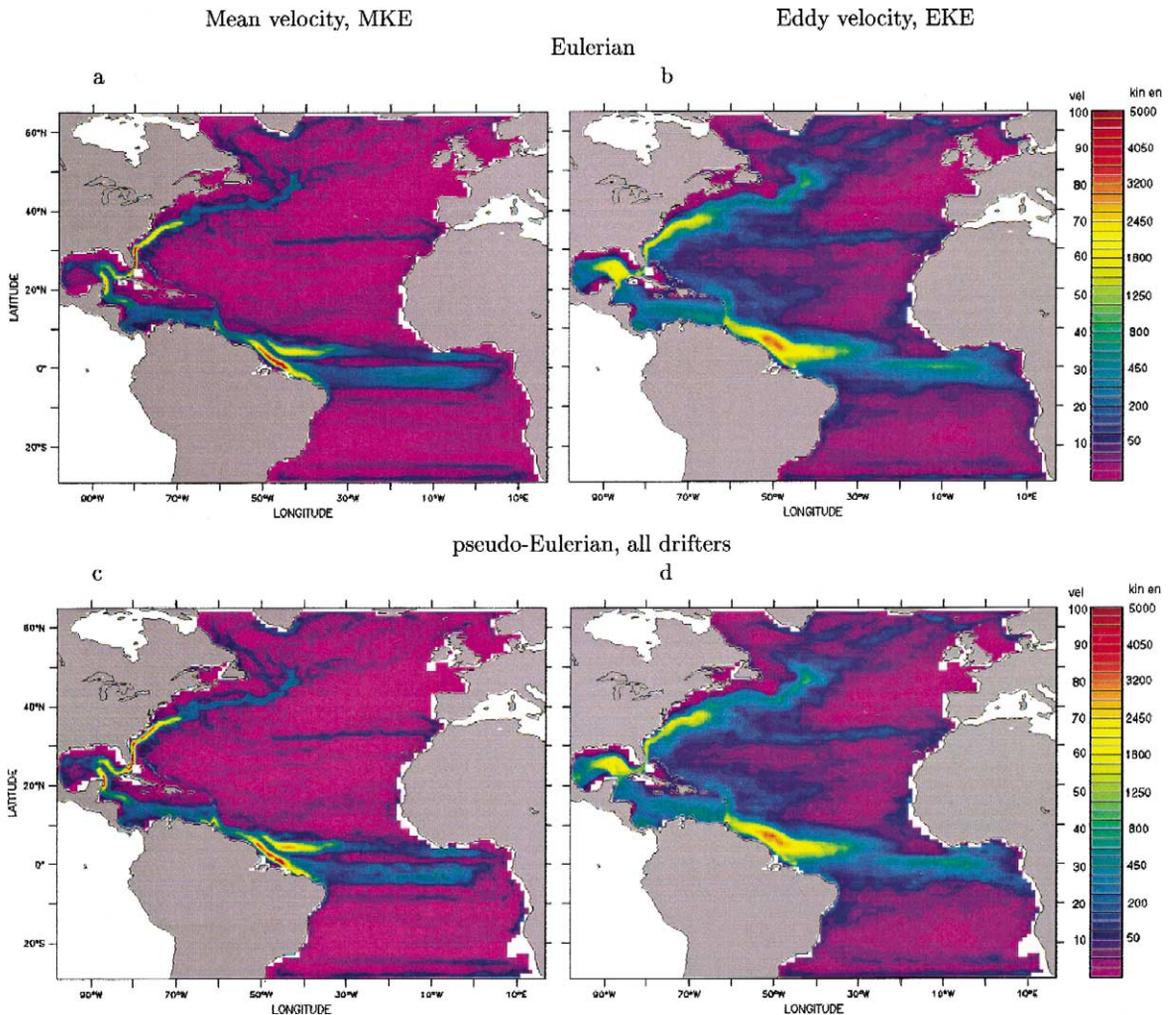

Fig. 4. Magnitude of the mean vector velocity and mean flow kinetic energy (left column, cm/s, and cm^2/s^2), and of the eddy vector velocity and eddy kinetic energy (right column, cm/s, and cm^2/s^2): (a,b) Eulerian; (c,d) pseudo-Eulerian, all drifters; (e,f) pseudo-Eulerian, initially seeded on a 5° × 5° grid; (g,h) pseudo-Eulerian, all drifters subsampled at 3-day interval.

Mean velocity, MKE

Eddy velocity, EKE

pseudo-Eulerian, space subsampled

e

f

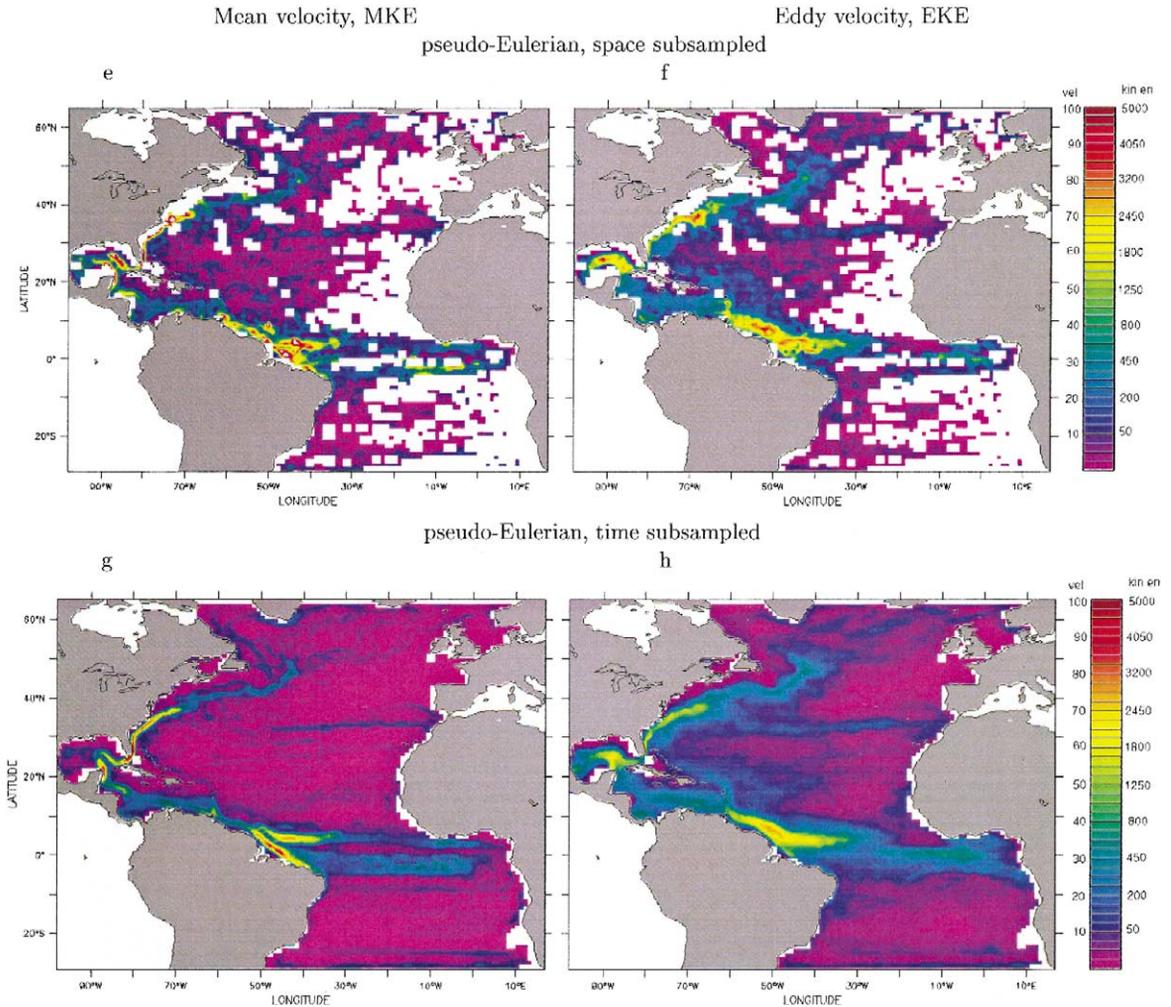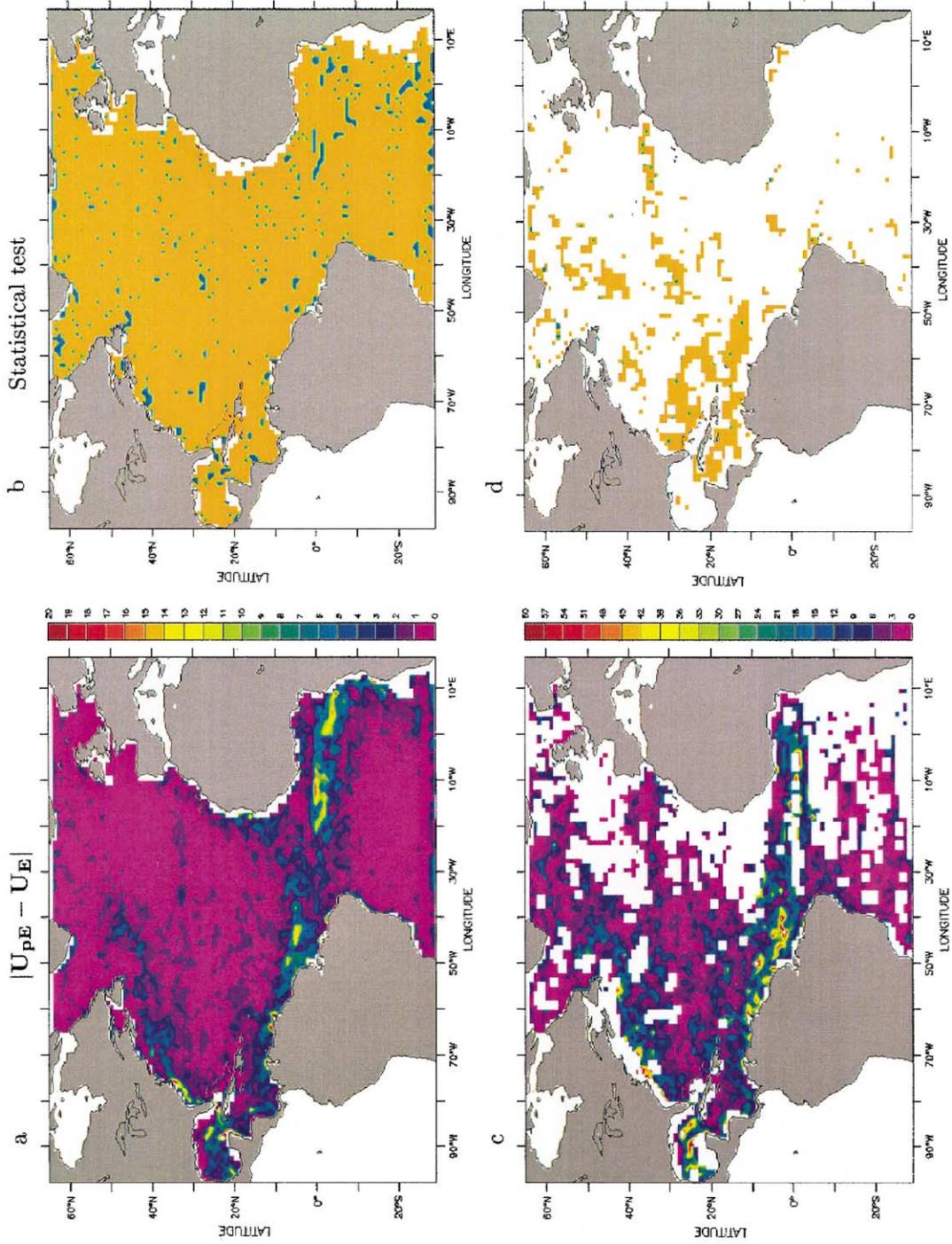pseudo-Eulerian, time subsampled
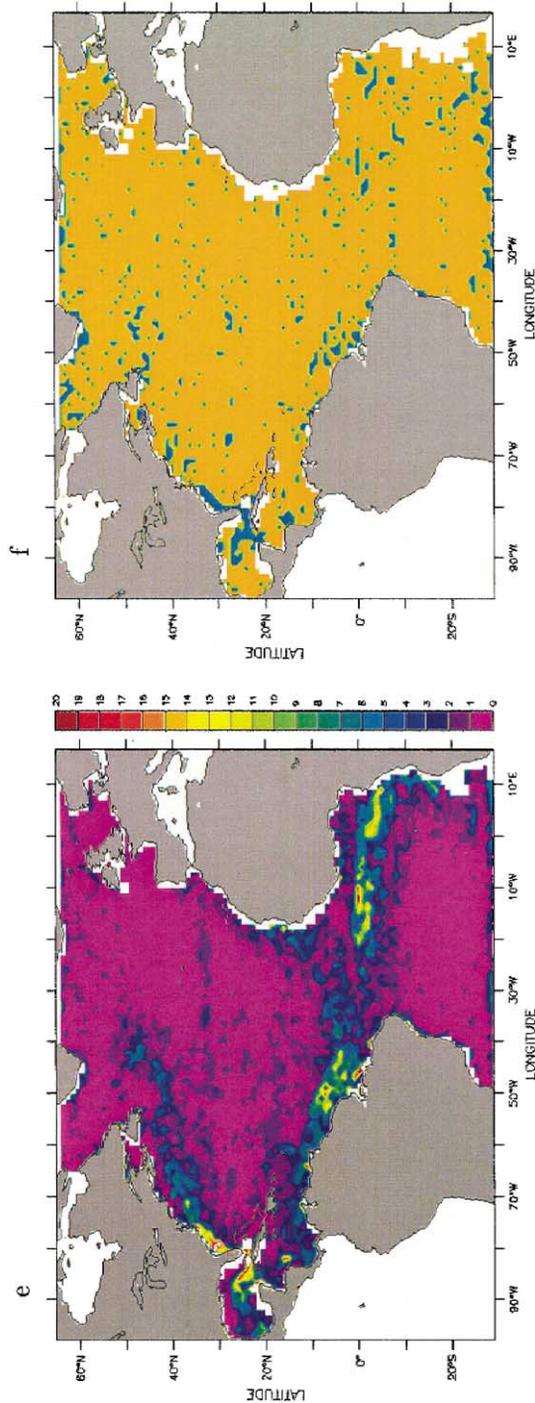
g

h

Fig. 4 (*continued*).

they tend to overestimate the mean flow in the western boundary currents, especially in the North Brazil Current and to a lesser degree in the Gulf Stream area.

Direct estimates of the biases, $(U_{pE} - U_E)$, have been computed for the full data set over 5° bins. In the 5° bins, in fact, sampling errors are significantly reduced relative to 1° bins (by approximately $1/5$, assuming N_{pE}^* is proportional to the bin area), so that bias patterns can be detected (this is not the case for the subsampled data sets, where sampling errors are too high to detect bias patterns even with 5° bins). The results are shown in Fig. 7a, indicating biases in the order of 2–3 cm/s (up to ≈ 5 cm/s), occurring

primarily in the equatorial region and in the western boundary. They provide a contribution against the current (negative bias) in the Equatorial system of currents, while in the western boundary and especially in the southwestward Brazil current retroflection they appear to be mostly in the same direction of the current (positive bias), in agreement with what was shown in Fig. 6.

The nature of these biases has been investigated, considering expression (3). A direct estimate of Eq. (3) using the simulated drifter is not feasible, even "a-posteriori", i.e., involving the full history of the flow and the drifters, because the instantaneous drifter concentration c is not high enough to significantly

estimate $\langle u'c'/C \rangle$. As a consequence, the bias effects have been studied using the approximations (4)–(6) which capture specific mechanisms as discussed in Section 3.

The effects of eddy dispersion depicted in Eqs. (4) and (5) (Fig. 2b,c) appear too small (by more than one order of magnitude) to justify the results. As discussed in Section 3, this might be partly related to neglecting the effects of shear dispersion. Not even the patterns, though, appear able to explain the observed results, except for an indication of the observed biases (Fig. 7a) in the Southern Equatorial Current in Fig. 2c.

The effects of mesoscale divergence processes have been investigated using Eq. (6). The right hand side of Eq. (6), $\langle u'h' \rangle/H$, is estimated by using the mixed layer thickness, and by performing a Eulerian average over 2 years and over 5° bins, from values first calculated in 1° bins. The results are shown in Fig. 7b. A direct comparison between Fig. 7a and b indicates a good qualitative agreement, even though the values in Fig. 7b are somewhat smaller (not exceeding $\approx$ 3 cm/s). The agreement is remarkable given that in our application, Eq. (6) includes contributions due to mass exchange between the mixed layer and interior layers, which do not influence drifter concentration. The negative bias in the equatorial region is clearly present in Fig. 7a. Also, significant contributions appear in the Brazil current region, mostly in the same direction as in Fig. 7a. Part of the differences in this area might be due to the sampling errors in Fig. 7a, particularly high in the Brazil Current because of high variability and sparse sampling. Also, a hint of positive bias contribution can be seen in the Gulf of Mexico and Gulf Stream regions, again in qualitative agreement with Fig. 7a. The most noticeable difference between Fig.

Fig. 5. Magnitude of the vector velocity difference with the Eulerian "truth", $|U_{pE} - U_E|$ (left, cm/s), and result for the statistical test on this quantity (right) for: (a,b) all drifters; (c,d) initial seeding on a $5° \times 5°$ grid; (e,f) subsampling at 3-day interval. Plotted ranges are 0–20, 0–80 and 0–20 cm/s for (a), (c) and (e), respectively. Blue regions in (b), (d) and (f) indicate velocity estimates which differ significantly (at 95% level) from the Eulerian "truth". The statistical test is done where the number of independent data is greater than 2.
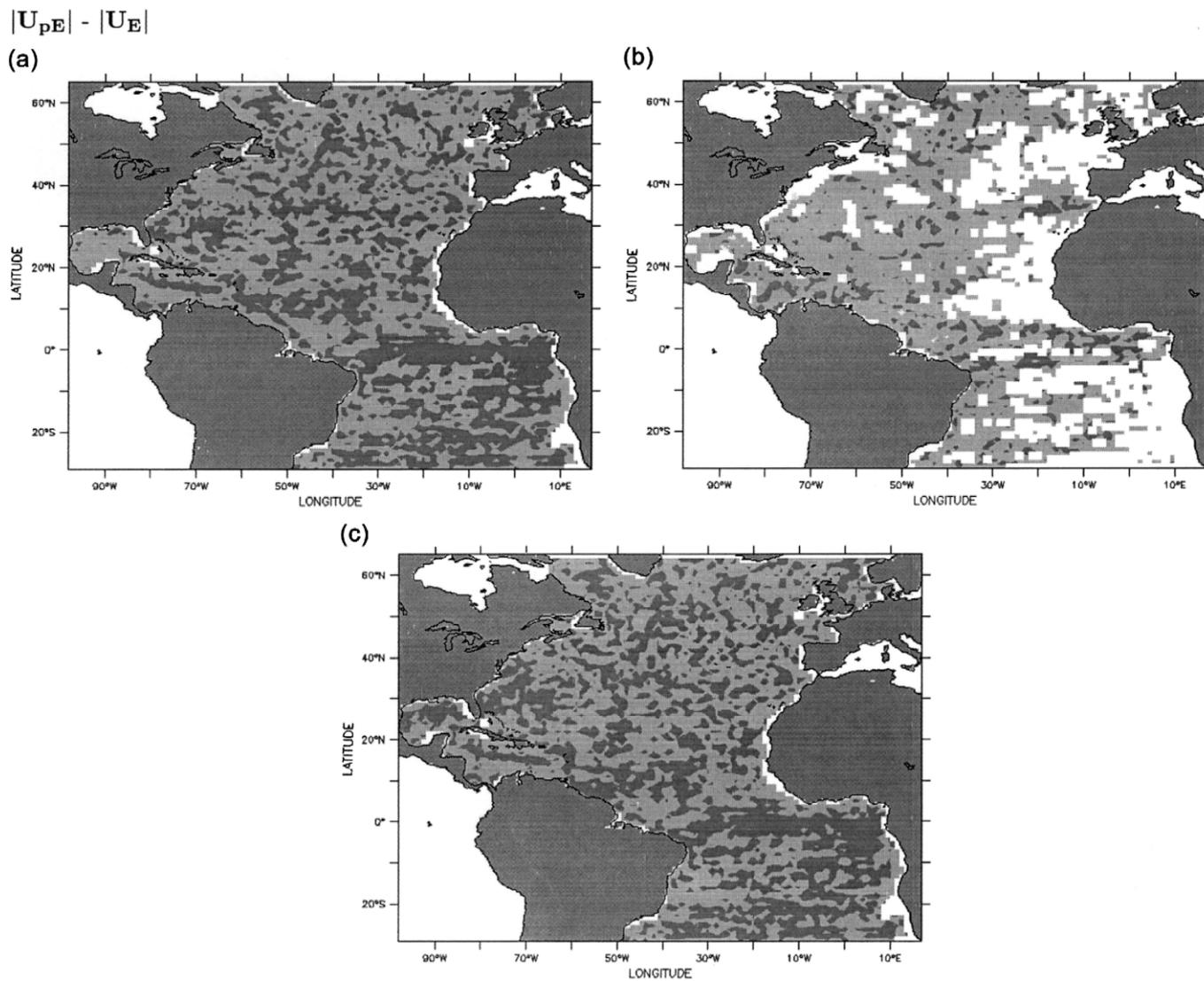
$|\mathbf{U}_{pE}| - |\mathbf{U}_E|$

Fig. 6. Pseudo-Eulerian velocity magnitude minus Eulerian velocity magnitude, $|U_{pE}| - |U_E|$, (cm/s) for: (a) all drifters, (b) initial seeding on a $5° \times 5°$ grid, (c) subsampling at 3-day interval. Dark (light) gray indicate negative (positive) values, with magnitude 0–50 cm/s.

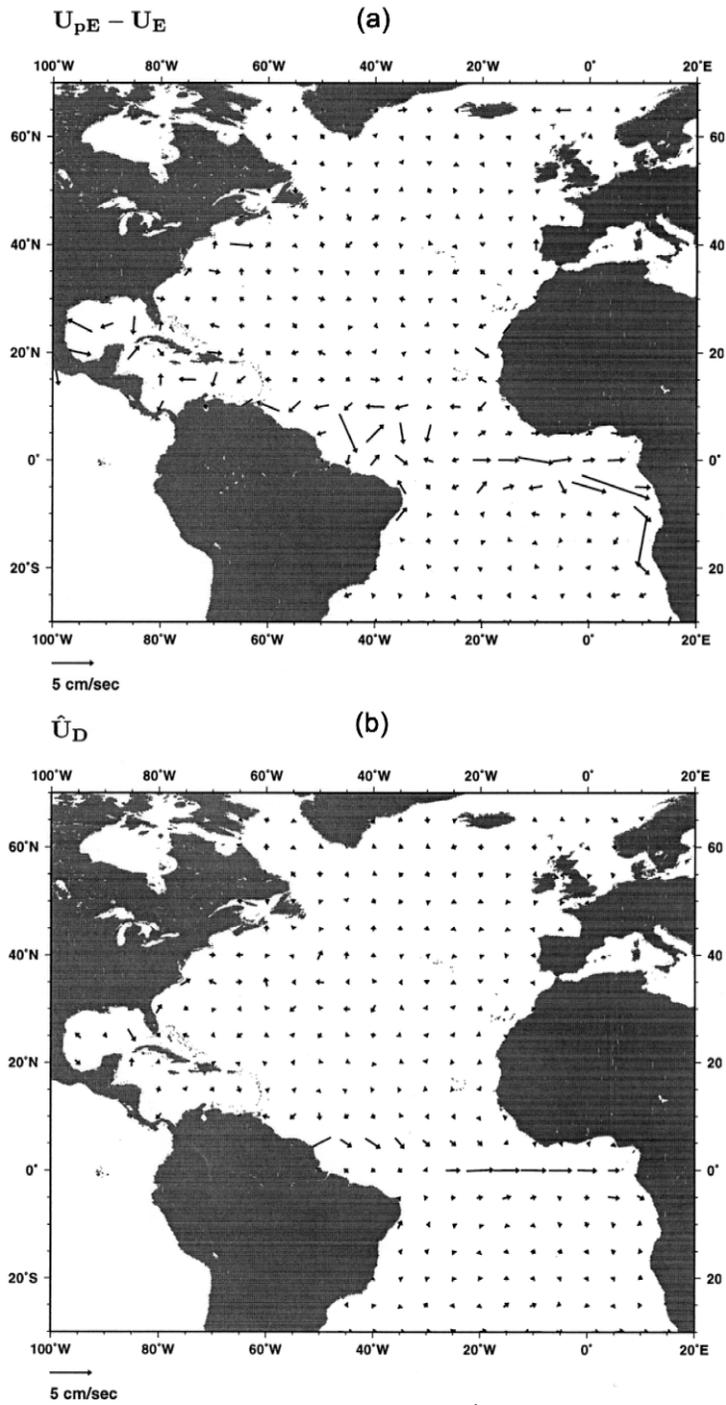$$U_{pE} - U_E \qquad \text{(a)}$$

$$\hat{U}_D \qquad \text{(b)}$$

Fig. 7. (a) Pseudo-Eulerian–Eulerian velocity, all drifters (5° × 5° averaged), (b) $\hat{U}_D = \langle u' h' \rangle / H$ computed from the Eulerian fields (same averaging).

7a and b occurs south of the Equator in the eastern boundary; the reason for this will be investigated in the future.

The results of Fig. 7 suggest that mesoscale divergence processes associated with mesoscale features like eddies and jets are mostly responsible for the observed biases. In the equatorial region, high velocity fluctuations appear to be undersampled by drifters, while they appear to be oversampled in the western boundary currents. This mechanism has been partially tested by considering the distribution of velocities inside some selected bins in the western boundary at random times. It has been found that the drifter trajectories indeed tend to occur in the regions of highest velocity inside the bins. A similar phenomenon has been observed also for in situ real drifters in the Kuroshio, comparing velocities along trajectories and velocity distribution from the altimeter (Uchida et al., 1998). Similarly, Olson and Backus (1985) have observed high concentrations of organisms in high velocity frontal eddy regions.

As a final aspect, we discuss how patterns of divergence at large scales also can lead to biases, when coupled with seasonality. This mechanism is expected to increase the underestimate in the equatorial region. It must be considered in fact that particles were initially deployed in this region during the fall season (model date September year 14), when the current is the weakest (maximum in the spring) and that the concentration c decreases with time because this is a large-scale divergence region. Consequently, the average flow appears biased toward the initial low velocity sampling, for which c has the highest values and u' has minimum magnitude. In order to investigate this seasonality effect in a divergent region, three mean velocity fields using different 1-year periods are computed, starting at model dates September year 14, April year 15, and September year 15 (not shown). When comparing the results, we first notice that in the equatorial region the velocity fields become progressively noisier from the first 1-year analysis period to the later ones, as can be expected given that the number of data progressively decreases. Despite this, the bias in the results initialized in September appears stronger than in the results initialized in April, suggesting an influence of seasonality. Similar seasonal effects on bias have not been observed in the western boundary region.

In many observational cases, a special type of potentially large error occurs, related to the binning of finite Lagrangian data with biased sampling (Smith and Parnes, 1994). If a bin contains low and high velocity regions (e.g., a narrow jet) and if there is no re-seeding, as in the launching of a single cluster of drifters at a single time, the low velocities are sampled a large number of times than the high velocities, because the drifters with low velocity stay longer in the bin. This difference can cause a serious velocity underestimate in bins that contain regions of high velocity gradient. This bias towards low velocities could be present in the eastern equatorial Atlantic, due to the presence of narrow currents and divergent flow (which causes the number of drifters entering the region to diminish with time).

4.2. Sub-sampling in space

The effects of sub-sampling are studies here by progressively decreasing the number of drifters, i.e., by considering only drifters initially launched with a horizontal resolution of 2°, 3°, and 5°. The total number of data, in drifter-days, decreases correspondingly by a factor of 4, 9 and 25, respectively. As noted in Section 2, the 5° case has 144,000 total drifter-days, less than the total number of historical drifter data presently available in the NOAA/AOML archive (220,000 drifter-days). The average value of 2 drifter-years per $5° \times 5°$ for the 5° case is lower than the recommended WOCE sampling. Note that as mentioned in Section 2, the data density concentration (Fig. 1b) is far from homogeneous, being dominated by large-scale divergence patterns. Maximum values of about 120 drifter-days/bin are reached in the convergence regions, while vast areas have only a few drifter-days coverage. By comparison, the coverage of the in situ drifters, with a maximum of about 400 drifter-days (see G01, Fig. 3), is not so fragmented, because of more concentrated initial conditions and the shorter life-time of the drifters. In the following, the bin size for averaging is maintained as $1° \times 1°$.

Examples of $|U_{pE}|$ (MKE$_{pE}$) and u'_{rmspE} (EKE$_{pE}$) estimates are shown in Fig. 4e,f, respectively, for the 5° initial resolution. The data sets at 2° and 3° initial resolution show a gradual transition between Fig. 4c,d and e,f. Despite the higher noise level, the

structure of the main currents is still well represented. At a more detailed level, though, the errors are obviously increased, and the pseudo-Eulerian mean flow estimate appears to be more energetic than both the $|U_E|$ and the $|U_{pE}|$ computed with the full data set (Fig. 4a,c). The underestimate in the eastern North Equatorial Current appears less evident.

The difference between U_E and U_{pE} is quantitatively shown in Fig. 5c. Differences in the order 40 cm/s commonly found in the regions of strong currents, reaching 60 cm/s in some locations, while in the interior the values are on the order of a few centimeter per second. On the average, the differences are approximately four times larger than in the case of the full data set (Fig. 5a). On the other hand, the sampling error is expected to increase by approximately five times, on average, given that the total number of data has decreased by 25 times (see Eqs. (1) and (2)). As a consequence, we can expect that most of the differences will not be significant, as for the full data set. The results of the test, shown in Fig. 5d partially confirm this hypothesis, even though there are extensive regions where the test cannot be applied because there are only two independent data points or less (N_{pE}^* of Eq. (2)), pointing out how sparse the spatially subsampled data set is. Statistical reliability can of course be increased by considering larger averaging bins (for example $5° \times 5°$). This is a common practice in practical applications with low data concentration, even though increasing binning size implies losing resolution, with detrimental effects especially in the strong current regions. Here, the $1° \times 1°$ binning is maintained because we are interested in the sub-sampling effects at a given resolution, and in the direct comparison with previous results.

Finally, the bias errors have been investigated considering the pattern of $|U_{pE}| - |U_E|$, which is shown for the $5° \times 5°$ initial resolution in Fig. 6b. As the number of drifters decreases from the full data set to the subsampled case (Fig. 6a,b), the observed overestimates in the western boundary currents become more evident. This was already shown by simple inspection of Fig. 4c,e and it is obvious in Fig. 6b. The overestimates are especially strong along the boundary currents, but they appear pervasive in the entire basin, while the underestimates in the eastern South Equatorial Current are less detectable, possibly for lack of data. The mechanism for increasing overestimates at decreasing spatial sampling is not clear yet, and it will be studied in detail in the future. First, the robustness of this result will be tested considering different initial realizations, even though the fact that the phenomenon is so pervasive might suggest that it is independent of specific initial condition realizations. Then, as mentioned in Section 4.1, the phenomenon will be linked with the bias Eq. (3), considering its sampling dependence. There is a possibility that the initial seeding of drifters at $5° \times 5°$ horizontal resolution has preferentially placed drifters in or near the strong currents, such as the Gulf Stream at 40°N or at 75°W, and that due to mesoscale convergence those drifters tend to stay in the high velocity core, sampling the larger velocities.

4.3. Subsampling in time

The effects of subsampling in time are studied by increasing the sampling interval Δt to 1 and 3 days, considering the data set with initial resolution of $1° \times 1°$. We recall that the integral time scales T are longer than 3 days over almost all of the model domain except for the western boundaries (see G01, Fig. 5), so that in the interior consecutive measurements can also be considered correlated in the subsampled data sets. There, the number of data points decreases by a factor of 2 and 6, respectively, but the number of independent measurements (Eq. (2)) does not change.

The estimates of $|U_{pE}|$ (MKE$_{pE}$) and u'_{rmspE} (EKE$_{pE}$) for $\Delta t = 3$ days are shown in Fig. 4g,h. The results for $\Delta t = 1$ day can be interpolated from the full data set case and the case discussed in this section, and are therefore not presented. Mean currents and kinetic energy appear qualitatively very similar to the full data set case. Quantitatively (Fig. 5e), the differences between U_E and U_{pE} appear more pronounced, especially near the Brazil Current, Gulf of Mexico, and Gulf Stream, with differences on the order of 15–20 cm/s. The pattern of sampling error is expected to be similar to the full data set case, since the number of independent measurements (Eq. (2)) is approximately the same except for the most energetic regions where drifters might cross the region in less that 3 days, and for regions with

the shortest Lagrangian time scales. If the sampling errors are approximately the same, it is not surprising that the test results for temporally subsampled data (Fig. 5f) show more significant differences than for the full data set, with larger values of $|U_{pE} - U_E|$. This indicates that the observed differences are due to other reasons than the sampling error, as discussed next.

The possible existence of biases is investigated considering the values of $|U_{pE}| - |U_E|$ in Fig. 6c. The underestimate in the eastern equatorial region, noticed in the full data set case (Fig. 6a), is evident here also. In contrast with what is seen in Fig. 6a, however, the western boundary regions appear characterized by underestimates rather than overestimates, especially in the regions of significant differences seen in Fig. 5f.

This result is likely to be due to the fact that increasing Δt results in decreasing the value of instantaneous high velocities, in agreement with the results of Mockett (1999), counterbalancing and over-riding the positive bias due to oversampling of high velocity fronts. Energetic drifters might not be approximately accounted for correctly, since they can cross the bin in less than 3 days, so that their velocity is not approximately registered inside a single bin. If both extremes of the 3-day sampled velocity are the average of the six values of the 12-h sampled velocity, and there is no error. Differences would only appear when the trajectory end points are not in the same bin. In this case, we attribute the 3-day mean velocity to a mid bin while in the 12-h case, velocities, are resolved over more than one bin. In a region where the trajectories curve and the magnitude of the velocity is approximately constant, as in the North Brazil eddies, Gulf of Mexico loop current, and Florida Current, the 3-day average velocity attributed to the middle bin is smaller than the 12-h velocity, and the location is displaced. For this effect to be present, the particles must cross the bins in less than the time sampling interval, and the interval has to be longer than the Lagrangian time scale. For intervals shorter than the Lagrangin time scale, successive velocities would essentially be equal in magnitude and direction, and there would be no significant velocity difference between the full and time-subsampled cases. As discussed also in Section 4.2 for the subsampling in space, results can be improved by considering larger averaging bins, but at the price of losing resolution (Figueroa and Olson, 1994).

It is important to emphasize that significant differences occur mostly in the western boundary, where the values of the Lagrangian time scales are on the order of (or less than) 3 days, i.e., the order of the sampling rate (Fig. 3). Since T is actually significantly shorter for real drifters than in the model (by about a factor of 2, G01), it can be speculated that subsampling effects will also be more relevant for real drifter analysis.

5. Summary and concluding remarks

In this paper, the reconstruction of the Eulerian mean flow using Lagrangian data and the effects of subsampling in space and time have been studied. A numerical flow, the solution of a high resolution MICOM simulation of the North Atlantic (Paiva et al., 1999, G01, Chassignet et al., 2001), is considered. A set of trajectories is computed, simulating the motion of surface drifters initially launched on a regular array of $1° \times 1°$, and results over a period of approximately 2 years are analyzed. The drifter distribution in time is influenced by the large scale Ekman divergence pattern, which results in a maximum data concentration in the divergent upwelling regions.

A comparison is carried out between estimates of the Eulerian mean flow U_E and of the pseudo-Eulerian mean flow U_{pE}, computed in $1° \times 1°$ bins. The Eulerian mean U_E is computed as a direct average of the instantaneous velocities of all grid points inside the bin for the 2-year period. U_{pE}, however, is computed by averaging in each bin all the available velocities along trajectories during the 2 years. The comparison is first performed using the full Lagrangian data set, with sampling intervals $\Delta t = 12$ h and drifters initially seeded at $1° \times 1°$ horizontal resolution, and then by subsampling the data in space and time.

For the full data set with total data corresponding to approximately 50 drifter-years per $5° \times 5°$ (substantially more than the WOCE requirement), differences on the order of 10 cm/s are found between U_E and U_{pE} in the strong current regions, reaching occasional maxima of about 20 cm/s. A statistical test

shows that these differences are usually not significant at the 95% confidence level, i.e., they are mostly due to the uncertainty related to subgrid-scale variability and finite sampling. Nevertheless, there are some areas, especially in the eastern equatorial region and in the western boundary currents, where the differences are significant, suggesting other sources of error.

The presence of biases is first investigated considering the pattern of the difference of the absolute values, $|U_{pE}| - |U_E|$. The calculations suggest that $|U_{pE}|$ underestimates $|U_E|$ in the eastern equatorial region, while it overestimates it in the western boundary currents in the eastern equatorial region, while it overestimates it in the western boundary currents and especially in the North Brazil Current. Direct bias estimates are also computed for the full data set with a 5° binning, confirming these results. The possibility that the observed biases are due to eddy dispersion processes has been tested, but the corresponding bias terms (Davis, 1991) are estimated to be too small to explain the observed values. Alternatively, it is suggested that the results are related to a bias induced by mesoscale divergence processes and expressed through the correlation between instantaneous velocity and mixed layer thickness, used as a proxy for particle concentration, $\hat{U}_D = \langle u'h' \rangle / H$ (Davis, 1998). Despite the fact that mixed layer thickness is influenced also by mass exchange with the interior layers, $\hat{U}_D$ appears able to explain most of the observed bias patterns. In the equatorial region, high velocity fluctuations appear undersampled by the drifters, while in the western boundary currents high velocity regions in eddies and fronts appear oversampled. A similar phenomenon has been observed also in in situ drifters in the Kuroshio (Uchida et al., 1998) and in biological organism concentration in eddies (Olson and Backus, 1985).

When subsampling in space is considered, it is observed that the errors increase, reaching approximately 40 cm/s with maxima of 70–80 cm/s in the 5° subsampling case (which is lower than the WOCE recommendation). Also in this case, the errors are mostly due to sampling uncertainty. The overestimate of the mean velocity U_{pE}, suggestive of a positive bias, appears to increase, becoming evident also in the interior of the basin. The reason for this

increase is not clear at this point, and is under investigation.

Subsampling in time slightly increases the errors in the major currents, which, for $\Delta t = 3$ days, become more clearly significant in the western boundary currents. This is probably due to the fact that Lagrangian time scales are shorter than Δt only in these highly energetic areas. U_{pE} tends to underestimate rather than overestimate the mean flow, in agreement with previous results (Mockett, 1999). This is probably because instantaneous high velocities are smoothed and energetic drifters are not appropriately accounted for, since they can cross the bin in less than 3 days. This induces a bias, which overrides the tendency to oversample high velocity fronts.

The present results are obtained for numerical flow and trajectories. An important question is how realistic are these results and how much can they be used as guidance for in situ data analysis. As discussed in G01, the numerical flow used here provides a realistic description of the main currents in the Atlantic Ocean, while it tends to underestimate the eddy kinetic enery in the interior and to overestimate the Lagrangian time scales T, probably due to the lack of high frequency winds in the model forcing and to the bulk mixed layer representation. Despite this, we recall that the sampling error (Eq. (1)) at given number of observations depends approximately on the product $\sigma^2 T$. As a consequence, the results for the error magnitude and significance discussed here for the various data sets can be used quite directly as guidance for in situ data. Possible differences can arise from the geographical data distribution. The numerical data distribution, in fact, is more clearly dependent on Ekman divergence (see G01), because the initial releases of in situ drifters are usually more concentrated and the drifter's life time is too short to be strongly influenced by large scale divergence.

Regarding the results on the under/overestimates of U_{pE} and their relationship to subsampling, we believe that they are of potential great interest in the analysis of real flows. In particular, attention is drawn to the possible relationship to the bias $\hat{U}_D$ related to the correlation between drifter concentration and velocity. As indicated by Davis (1998), this bias is likely to be related to the "eddy transport"

discussed and parameterized by Gent and McWilliams (1990), and is thought to represent the actual difference between tracer mean advection and Eulerian mean velocity. In the future, we intend to further investigate this matter, first testing the results quantitatively, by verifying their robustness with respect to numerical advection schemes, initial distributions, and characteristics of the mixed layer parameterization. At a more conceptual level, we also intend to investigate possible parameterizations of $\hat{U}_{\mathrm{D}}$ as a function of macroscopic variables and understand its dependence on sampling in space and time.

Acknowledgements

Appendix A. Lagrangian time scales

The Lagrangian velocity time scale $T = (T_u, T_v)$ is a purely Lagrangian quantity defined for the two velocity components as

$$T_u = \frac{1}{\sigma_u^2} \int_0^\infty R_{uu}(\tau)\mathrm{d}\tau,$$

and

$$T_v = \frac{1}{\sigma_v^2} \int_0^\infty R_{vv}(\tau)\mathrm{d}\tau.$$

R is the Lagrangian temporal auto-covariance function, computed for u and v following particles, and defined as

$$R_{uu}(\tau) = \langle (u(t) - \bar{u}(t))(u(t+\tau) - \bar{u}(t)) \rangle,$$
$$R_{vv}(\tau) = \langle (v(t) - \bar{v}(t))(v(t+\tau) - \bar{v}(t)) \rangle,$$

where the symbol $\langle \ \rangle$ indicates expected values.

The estimates of T, performed in G01 and shown in Fig. 3, are calculated by first dividing the drifter trajectories into segments whose first and last positions are 400 km apart. Four hundred kilometers is chosen as a compromise in order to have sufficient data to reliably calculate temporal covariance functions and also to resolve the spatial details of the eddy field. All "leftover" segments or short trajectories containing at least 50 positions are also used in the analysis. For each velocity component, R is estimated computing the expected value, $\langle \ \rangle$, as an arithmetic average over all data with lag τ. The average velocities, $\bar{u}(t)$, $\bar{v}(t)$, are given by a least-squares fit to a linear trend using 400 km long subsets of a drifter's trajectory. The variances of u and v, denoted by σ_u^2 and σ_v^2, are $R_{uu}(0)$ and $R_{vv}(0)$, respectively.

When R is calculated from a finite data set, a subjective choice for the limits of integration to compute T must be made. To avoid this problem and to gain statistical confidence, a function of a small number of parameters is fitted to each temporal auto-covariance function R,

$$R(\tau) = \sigma^2 (1 - \epsilon_n^2)\cos(2\pi\tau/P)\mathrm{e}^{-(\tau/\tau_e)^2},$$

where ϵ_n^2 is the normalized (by total variance σ^2) variance of subgrid scale processes and measurement variance for the data, P is the period that equals four times the zero-crossing scale of the covariance function, and τ_e is the e-folding scale or the turbulent time scale. This covariance function, based on a smoothed version of a second-order auto-regressive process, contains a wave component and a turbulent component, and produces excellent fits to the observed and simulated Lagrangian velocity covariance functions.

The three parameters (ϵ_n^2, P, and τ_e) are determined using the feature-based technique described in Mariano and Chin (1996), which finds the value of $R(\tau)$ at zero lag ($= \sigma^2(1 - \epsilon_n^2)$), finds the zero-crossing scale of $R(= P/4)$ by determining where $R(\tau)$ changes from positive to negative, and finds the e-folding scale τ_e from the first two parameters, an initial guess consisting of the lag at which $R(\tau)$ is $1/e$ of its initial value, with the best fit being determined in a least-squares sense.

The time scales, computed for each 400 km long subset of a drifter trajectory, are determined by

evaluating the integral of $R(\tau)$ (Gradshteyn and Ryzhik, 1980),

$$\int_0^\infty R(\tau)\mathrm{d}\tau = \sigma^2 \left(1 - \epsilon_n^2\right) \frac{\tau_\mathrm{e}}{2} \sqrt{\pi} \exp\left(\frac{-\pi^2 \tau_\mathrm{e}^2}{P^2}\right),$$

using the estimated three parameters and σ^2. Then, the integral time scale is

$$T = \left(1 - \epsilon_n^2\right) \frac{\tau_\mathrm{e}}{2} \sqrt{\pi} \exp\left(\frac{-\pi^2 \tau_\mathrm{e}^2}{P^2}\right).$$

Direct numerical integration of the temporal auto-covariance functions, using n temporal lags, requires estimating n parameters for the integration. Each of these n parameters is noisy and small changes in n can cause large changes in the value of the integral. This is especially true when the temporal auto-covariance function has large negative lobes. The integration of the Lagrangian velocity autocovariance function using three parameters leads to more robust estimates of the Lagrangian velocity time scales than does the numerical integration of n values $R(\tau)$.

The values of T are assigned to the midpoint of the drifter trajectory. A least-squares bi-cubic smoothing spline (Inoue, 1986) is then used to interpolate the estimates of T to a regular grid. The plots (Figs. 3 and 6 of G01) are contoured to a $5° \times 5°$ grid to enhance the large scale structure of the Lagrangian velocity time scales.

References

Abramowitz, M., Stegun, I.A., 1964. Handbook of Mathematical Functions, US Dept. of Commerce, National Bureau of Standards, Applied Mathematics. 55.

Bleck, R., Chassignet, E., 1994. Simulating the oceanic circulation with isopycnic-coordinate models. In: Majundar, S.K. (Ed.), The Oceans: Physical–Chemical Dynamics and Human Impact. pp. 17–39.

Bleck, R., Hanson, H.P., Hu, D., Kraus, E.B., 1989. Mixed layer/thermocline interaction in a three-dimensional isopycnic coordinate model. J. Phys. Oceanogr. 19, 1417–1439.

Bleck, R., Rooth, C., Hu, D., Smith, L.T., 1992. Salinity-driven thermocline transients in a wind- and thermohaline-forced isopycnic coordinate model of the North Atlantic. J. Phys. Oceanogr. 22, 1486–1505.

Bleck, R., Dean, S., O'Keefe, M., Sawdey, A., Numrich, R.W., 1995. A comparison of data-parallel and message-passing versions of the Miami Isopycnic Coordinate Ocean Model (MICOM). Parallel Comput. 21, 1695–1720.

Brügge, B., 1995. Near-surface circulation and kinetic energy in the central North-Atlantic from drifter data. J. Geophys. Res. 100 (20), 543–554.

Chassignet, E.P., Garraffo, Z.D., Smith, R.D., Hurlburt, H.E., 2001. High resolution Gulf Stream Modeling. Ocean Modelling, Submitted for publication.

Da Silva, A.M., Young, C.C., Levitus, S., 1994. Atlas of surface marine data 1994. Technical Report. National Oceanic and Atmospheric Administration.

Davis, R.E., 1987. Modeling eddy transport of passive tracers. J. Mar. Res. 45, 635–666.

Davis, R.E., 1991. Observing the general circulation with floats. Deep Sea Res. 38 (Suppl. 1), 531S–571S.

Davis, R.E., 1998. Preliminary results from directly measuring middepth circulation in the tropical and South Pacific. J. Geophys. Res. 103 (24), 624–639.

Davis, R.E., Killworth, P.D., Blundell, J.R., 1996. Comparison of Autonomous Lagrangian Circulation Explorer and fine resolution Antarctic model results in the South Atlantic. J. Geophys. Res. 101C, 855–884.

Figueroa, H.A., Olson, D.B., 1994. Eddy resolution versus eddy diffusion in a double-gyre GCM. Part 1: The Lagrangian and Eulerian description. J. Physical Oceanogr. 24 (2), 371–386.

Freeland, H.J., Rhines, P., Rossby, H.T., 1975. Statistical observations of trajectories of neutrally buoyant floats in the North Atlantic. J. Mar. Res. 33, 383–404.

Garraffo, Z.D., Mariano, A.J., Griffa, A., Veneziani, C., Chassignet, E.P., 2001. Lagrangian data in a high-resolution numerical simulation of the North Atlantic. I. Comparison with in situ drifter data. J. Mar. Syst. 29, 157–176.

Gent, P.R., McWilliams, J.C., 1990. Isopycnal mixing in ocean circulation models. J. Phys. Oceanogr. 20, 150–155.

Gradshteyn, I.S., Ryzhik, I.M., 1980. In: Jeffrey, J. (Ed.), Table of Integrals, Series, and Products. Academic Press, New York, Translation by Scripta Technica, 1160 pp.

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

Levitus, S., 1982. Climatological Atlas of the World Ocean. NOAA Professional Paper 13, US Dept. of Commerce, NOAA.

Mariano, A.J., Chin, T.M., 1996. Feature and contour based data analysis and assimilation in physical oceanography. In: Adler, Muller, A., Rozovskii, J. (Eds.), Stochastic Modelling in Physical Oceanography. Birkhauser, Boston, MA, pp. 311–342.

Mockett, C., 1999. Dispersion and reconstruction. Summer GFD Program, WHOI Technical Report 99-01, 166–188.

Morrison, D.F., 1976. Multivariate Statistical Methods. McGraw-Hill, New York.

Olson, D.B., Backus, R.H., 1985. The concentrating of organisms at fronts: a cold-water fish and a warm-core Gulf Stream ring. J. Mar. Res. 43, 113–137.

Owens, W.B., 1991. A statistical description of the mean circulation and eddy variability in the northwestern North Atlantic using SOFAR floats. Progr. Oceanogr. 28, 257–303.

Özgökmen, T.M., Chassignet, E.P., Paiva, A.M., 1997. Impact of wind forcing, bottom topography, and inertia on mid-latitude jet separation in a quasi-geostrophic model. J. Phys. Oceanogr. 27, 2460–2476.

Paiva, A.M., Hargrove, J.T., Chassignet, E.P., Bleck, R., 1999. Turbulent behavior of a fine mesh (1/12 deg) numerical simulation of the North Atlantic. J. Mar. Sys. 307–320.

Riser, S.C., Rossby, H.T., 1983. Quasi-Lagrangian structure and variability of the subtropical western North Atlantic circulation. J. Mar. Res. 41, 127–172.

Rossby, H.T., Riser, S.C., Mariano, A.J., 1983. The western North Atlantic—a Lagrantian viewpoint. In: Robinson, A.R. (Ed.), Eddies in Marine Science. Springer-Verlag, 609 pp.

Smith, Parnes, 1994. Mean Streets: the median of a size-biased sample and the population mean. The American Statistician 48 (2), 106–110.

Smith, L.T., Chassignet, E.P., Bleck, R., 2000. The impact of lateral boundary conditions and horizontal resolution on North Atlantic water mass transformations and pathways in an isopycnic coordinate ocean model. J. Phys. Oceanogr. 30, 137–159.

Taylor, G.I., 1921. Diffusion by continuous movements. Proc. London Math. Soc. 20, 196–212.

Uchida, H., Imawaki, S., Hu, J.H., 1998. Comparison of Kuroshio surface velocity derived from satellite altimeter and drifting buoy data. J. Oceanogr. 54, 115–122.

Zambianchi, E., Griffa, A., 1994. Effects of finite scales of turbulence on dispersion estimates. J. Mar. Res. 52, 129–148.