

# Lagrangian tracking in stochastic fields with application to an ensemble of velocity fields in the Red Sea

Samah El Mohtar<sup>a,b</sup>, Ibrahim Hoteit<sup>b</sup>, Omar Knio<sup>b,d</sup>, Leila Issa<sup>c</sup>, Issam Lakkis<sup>\*,a</sup>

<sup>a</sup> Department of Mechanical Engineering, American University of Beirut, Beirut, Lebanon

<sup>b</sup> King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

<sup>c</sup> Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon

<sup>d</sup> Duke University, Durham, NC 27708, USA

## ARTICLE INFO

### Keywords:

Stochastic flow fields

Red sea

Lagrangian tracking

## ABSTRACT

Lagrangian tracking of passive tracers in a stochastic velocity field within a sequential ensemble data assimilation framework is challenging due to the exponential growth in the number of particles. This growth arises from describing the behavior of velocity over time as a set of possible combinations of the different realizations, before and after each assimilation cycle. This paper addresses the problem of efficiently advecting particles in stochastic flow fields, whose statistics are prescribed by an underlying ensemble, in a parallel computational framework (openMP). To this end, an efficient algorithm for forward and backward tracking of passive particles in stochastic flow-fields is presented. The algorithm, which employs higher order particle advection schemes, presents a mechanism for controlling the growth in the number of particles. The mechanism uses an adaptive binning procedure, while conserving the zeroth, first and second moments of probability (total probability, mean position, and variance). The adaptive binning process offers a tradeoff between speed and accuracy by limiting the number of particles to a desired maximum. To validate our method, we conducted various forward and backward particles tracking experiments within a realistic high-resolution ensemble assimilation setting of the Red Sea, focusing on the effect of the maximum number of particles, the time step, the variance of the ensemble, the travel time, the source location, and history of transport.

## 1. Introduction

Lagrangian particle tracking (L.P.T.) is a powerful tool for modeling transport phenomena in the ocean. Integrated with other tools, L.P.T. lies at the core of models that assess the quantity of pollutants affecting contaminated waters after catastrophic events, such as oil spills, and toxic beach littering (Nakashima et al., 2012; Cole et al., 2011). L.P.T. is also often used to locate the sources of polluting events such as tar ball deposition (Suneel et al., 2016) or to determine the site of an airplane crash. The application of the method extends to drifter design experiments, where one is interested in optimizing the location of drifter launch sites (Hernandez et al., 1995; Poje et al., 2002), and also to study ocean physics (e.g. water ventilation or diagnosing ocean mixing) and pathways of marine organisms (e.g. nutrients transport). A review of Lagrangian methods for large-scale ocean applications can be found in van Sebille et al. (2018).

Depending on the application, there are two modes of operation of L.P.T.: forward tracking investigates particles transport from known

sources (Brickman and Smith, 2002), while backward tracking is used to identify the source location, given observations at later times (Batchelder, 2006; Isobe et al., 2009). In both modes, a major challenge in this area is to design a robust algorithm for L.P.T. in the presence of uncertainties. Uncertainties can be categorized into two types: Eulerian and Lagrangian. Lagrangian uncertainties are those arising from processes affecting the particles themselves. Depending on what the particles are used to represent, such processes can range from diffusion and oil weathering (i.e., spreading, evaporation, emulsification, dispersion, etc) (Spaulding, 1988; Reed et al., 1999; Spaulding, 2017) to larval reactions (i.e., behavior, mortality and reproduction) (Batchelder, 2006). Eulerian uncertainties, introduced by the velocity field, affect the advection of particles. These uncertainties are due to uncertain input fields in ocean circulation models, such as initial conditions, atmospheric forcing, and poorly known modeling errors (e.g. National Research Council, 2012). In this work, an efficient algorithm for both forward and backward tracking of passive particles in the presence of Eulerian uncertainties is presented.

\* Corresponding author at: American University of Beirut, Department of Mechanical Engineering, Maroun Semaan Faculty of Engineering, Beirut, Lebanon.  
E-mail address: [il01@aub.edu.lb](mailto:il01@aub.edu.lb) (I. Lakkis).

**Nomenclature**

0	subscript: refers to initial conditions at $t = 0$
1	subscript: refers to post merging
$\alpha$	fraction of the assimilation interval
$\Delta t$	time step, $\Delta t = T_a/m$
$\Delta x$	x dimension of the grid cell or bin
$\Delta y$	y dimension of the grid cell or bin
$\Delta z$	z dimension of the grid cell or bin
$\sigma$	standard deviation
$\vec{u}$	velocity vector
$\vec{x}$	position vector
$a$	subscript: refers to assimilation
$b$	subscript: refers to bins
$i$	realization index at beginning of advection over a time period = $T_a$

$j$	realization index at end of advection over a time period = $T_a$
$k$	index of assimilation interval, $k = 1..n$
$l$	particle index
$m$	number of time steps per assimilation interval
$N$	number of particles
$n$	Number of assimilation intervals
$o$	order of x moment
$p$	probability
$q$	order of z moment
$R$	number of realizations
$r$	order of y moment
$S$	probability strength
$t$	time
$T_a$	assimilation interval

The most common approach to deal with Eulerian uncertainties is to advect the particles with a stochastic velocity field (Beaudoin et al., 2007; Guo et al., 2016). In particular, when particle tracking is performed in an ensemble data assimilation context, the flow-field is characterized by a set of velocity vectors (Hoteit et al., 2013). At each assimilation cycle, a set of (eventually random) realizations of this field are generated, and the new realizations do not necessarily have well defined correspondences to the realizations before or after the assimilation. As a result, the behavior of the velocity over time can only be described as a set of possible combinations of the different realizations, before and after each assimilation cycle (Höllt et al., 2015). Even with a deterministic initial condition, a particle can then have as many trajectories as the combination of all the previous ensemble realizations of all assimilation intervals. The number of particles will then grow exponentially with the number of assimilation intervals, making it impractical to track individual particles (Höllt et al., 2015). This motivates the need to wisely manage this type of growing data in a manner that makes it feasible to apply both forward and backward Lagrangian particle tracking methods. This paper addresses the problem of efficiently advecting particles with a random velocity field, whose statistics are prescribed by a set of realizations, in a parallel computational framework.

When the statistics of the velocity field are known, a Monte Carlo approach is often used and the uncertainty in the velocity field is generated based on a probability distribution. Previous work on particle tracking in such flow fields focused on developing algorithms to efficiently advect particles using parallel particle tracking (e.g. Beaudoin et al., 2007; Guo et al., 2013; Guo et al., 2016). Some of these algorithms (Beaudoin et al., 2007; Guo et al., 2016) employed stochastic integration schemes to numerically solve the advection equation. Most relevant to this article is the work in Guo et al. (2016) on using stochastic flow maps (SFM) to model uncertain transport. In Guo et al. (2016), advected particles were grouped in each grid cell, and at the next time step, new particles were launched from the center of the cell, following a Monte Carlo approach, which traces the particle stochastically in the uncertain unsteady flow field. Parallelization of the process was accomplished by decoupling the time dependencies in SFMs, thereby reducing the computational cost of the many runs required to efficiently trace the large number of seeded particles.

Our work is similar to Guo et al. (2016) in the following aspects: cost reduction, adaptivity, and binning. It differs as follows: (i) our method decouples the statistics from the tracking process. In this framework, the stochastic velocity field is first expressed as a sequence of independently sampled ensemble fields, which are then used for particle tracking, (ii) this makes our method suitable for particle tracking in an assimilation context, crucial for accurate simulation of the ocean flow-field (Hoteit et al., 2002), and (iii) cost reduction in our work is

attained by both invoking a parallel algorithm and controlling the growth in the number of particles. The latter is accomplished by binning and by using a spatially adaptive algorithm that controls the evolution of the uncertainties introduced by binning.

Our work builds on the approach proposed by Höllt et al. (2015) for particle tracking using ensemble flow fields. Höllt et al. (2015) introduced a way to keep the combination of path-lines manageable by binning the final positions of the particles. Particles with path-lines leading to the same bin are merged into one particle to be traced in the next assimilation interval. A probability map is then obtained at each assimilation step with the number of particles in a bin defining the probability assigned to it. The approach was combined with first order forward tracking and tested in the Red Sea using realistic ensembles simulated by an assimilative ocean general circulation model (OGCM). The test results illustrated the relevance of the method to efficiently estimate the probability maps of interest. One limitation of the binning suggested in Höllt et al. (2015) is that it collapses all particles onto the center of the bin they belong to, while conserving the total probability (zeroth moment). This approach not only increases the uncertainty, but it also does so in an unquantifiable and uncontrollable manner. The objective of this work is therefore to develop high order particle advection schemes in stochastic flow fields with statistics prescribed by an underlying ensemble, and to propose an elements growth control mechanism that uses an adaptive binning procedure, while conserving the zeroth, first and second moments of probability (total probability, mean position, and variance). The proposed adaptive binning procedure offers a tradeoff between speed and accuracy by limiting the number of particles to a desired maximum. The novelty of the method, manifested by a higher order advection scheme along with the adaptive and conservative binning procedure, will lead to improvements in cost, speed and accuracy, all essential to online tracking. The forward and backward algorithms presented in this work are implemented in a parallel computational framework using openMP.

The paper is structured as follows. Section 2 describes the ensemble of Red Sea flow-fields used in the numerical experiments. Section 3 presents the proposed methods for forward and backward tracking and the construction of the probability maps, including the integrations and binning schemes. The design and results of the numerical experiments are presented in Section 4. Finally, concluding remarks are offered in Section 5.

## 2. Ensemble dataset

The method is illustrated using an assimilated ensemble dataset of velocity fields of the Red Sea. The full description of the assimilation system from which the dataset was generated can be found in Toye et al. (2017). The system operates sequentially as cycles of

forecast-analysis steps, using the MIT ocean general circulation model (MITgcm) for forecasting the Red Sea circulation and an ensemble Kalman filter (EnKF) based on the Data Assimilation Research Testbed (DART) software for updating the forecast every time new observations become available (Hoteit et al., 2013). The system domain extends from 30°E to 50°E and from 10°N to 30°N, covering the Red Sea, the Gulf of Suez, the Gulf of Aqaba, and the Gulf of Aden. The system was configured on a spherical grid with a resolution of  $0.04^\circ \times 0.04^\circ$  resulting in a  $500 \times 500$  grid points and 50 z-vertical layers ranging from 4 m at the surface to 300 m near the bottom. The model bathymetry was extracted from the gridded General Bathymetric Chart of the Ocean (GEBCO), and the atmospheric forcing fields were obtained from the 6-hourly European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis. The open boundaries in the Gulf of Aden were prescribed from the Estimation of the Circulation and Climate of the Ocean (ECCO) consortium, and the initial ensemble selected from the outputs of a long model run without assimilation. The system assimilated combined along-track Sea Surface Height (SSH) observations from nine altimeter missions from the Radar Altimeter Database System (RADS) and mapped Sea Surface Temperature (SST) data from the 4 km Advanced Very High Resolution Radiometer (AVHRR) Pathfinder version 5 time series over  $T_a = 3$  days assimilation intervals. This assimilation system was integrated to generate an ensemble of 3-daily flow-fields for the month of January 2006, resulting in a dataset of  $n = 10$  sampled time steps, each consisting of  $N_e = 50$  realizations. In an EnKF, the realizations are independently sampled after every analysis step and do not therefore exhibit well defined correspondences to the realizations at the previous analysis steps (Hoteit et al., 2015; Höllt et al., 2015). As a result, the particles behavior over time can only be described as a set of possible combinations arising from advection from any realization of the ensemble at assimilation time  $t_a$  to any realization of the ensemble at assimilation time  $t_a + T_a$ , which characterizes the time interval separating these independently sampled ensembles.

### 3. Methods

#### 3.1. Stochastic velocity map

Given the source location  $\vec{x}_s$  and emission time  $t_0$ , the problem is to identify probable locations of a passive particle carried by a stochastic flow field, represented in terms of an ensemble of velocity fields, available at every assimilation step  $t_a$ ,

$$\vec{u}_i(\vec{x}_g, t_a), \quad t_a = kT_a, \quad k \in \mathbb{N}, \quad i = 1, \dots, N_e \quad (1)$$

where  $T_a$  is the assimilation frequency,  $\vec{x}_g$  denotes the grid coordinates, and  $N_e$  is the ensemble size. In this section, we build the velocity map that enables advection of passive tracers using high-order integration schemes, customized for such ensemble representation.

Implementation of high-order time integration schemes for particle advection requires the computation of the velocity at intermediate times,  $t = t_a + \alpha T_a$ , where  $0 \leq \alpha \leq 1$ . This is accomplished by using linear interpolation between the velocities at  $t = t_a$  and  $t = t_a + T_a$ . Since the ensemble velocity field at  $t = t_a + T_a$  is sampled independently from that at  $t_a$ , there are, for each realization  $i$  at  $t = t_a$ ,  $N_e$  possibilities for the velocity at the intermediate time  $t_a + \alpha T_a$ . These possibilities corresponds to  $N_e$  interpolations of the velocity fields, at the particle position  $\vec{x}$ , between the  $i$ th realization at  $t_a$  and all the realizations  $j = 1, \dots, N_e$  at  $t_a + T_a$ . The interpolated velocities, denoted  $\vec{u}_{ij}(\vec{x}, t_a + \alpha T_a)$ , are

$$\vec{u}_{ij}(\vec{x}, t_a + \alpha T_a) = (1 - \alpha) \vec{u}_i(\vec{x}, t_a) + \alpha \vec{u}_j(\vec{x}, t_a + T_a) \quad \text{for } j = 1, \dots, N_e. \quad (2)$$

Each realization is denoted by a single subscript, i.e.  $\vec{u}_i(\vec{x}, t_a)$  denotes realization  $i$  of the velocity at position  $\vec{x}$  and time  $t_a$ .

Thus, considering all the combinations arising from the ensemble

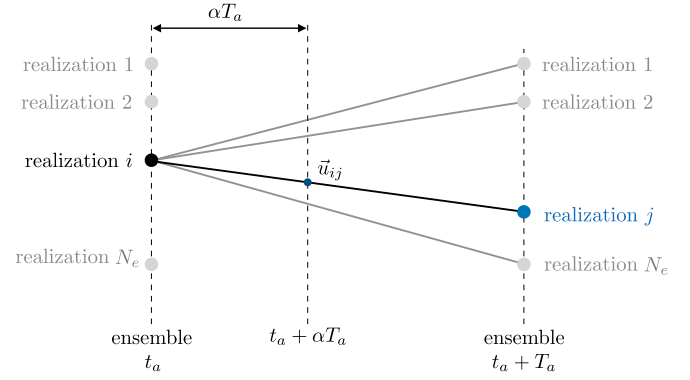


Fig. 1. A particle at time  $t_a$  is advected along  $N_e^2$  trajectories.

velocity fields at  $t_a$  and  $t_a + T_a$ , a particle initially located at  $\vec{x}^{t_a}$  at  $t = t_a$  will arrive at  $N_e^2$  equally probable destinations,  $\vec{x}_{ij}^{t_a + \Delta t}$ , at  $t_a + \Delta t$ , with  $0 < \Delta t < T_a$ , and  $N_e^2$  being the product of  $N_e$  equally probable velocities at  $t_a$  and  $N_e$  equally probable velocities at  $t_a + T_a$ , as depicted in Fig. 1.

#### 3.2. Higher-order time integration schemes

Here, we use the velocity map (2) to present high-order time integration schemes for advecting passive tracers in the stochastic field corresponding to the given ensemble representation (1). Since the ensemble is available at every assimilation step, we choose the integration time step,  $\Delta t$ , to be an integer fraction of  $T_a$ , i.e.  $\Delta t = T_a/m$ , where  $m$  is a positive integer. We then present the implementation of such schemes for advecting a particle in the interpolated field (1) from  $t = t_a$  to  $t = t_a + T_a = t_a + m\Delta t$ .

Starting from position  $\vec{x}^{t_a}$  at  $t = t_a$ , we consider the  $N_e^2$  equally probable paths expressed as:

$$\vec{x}_{ij}^{t_a + \Delta t} = \vec{x}_i^{t_a} + \int_{t_a}^{t_a + \Delta t} \vec{u}_{ij}(\vec{x}^s, s) ds, \quad 1 \leq i, j \leq N_e. \quad (3)$$

To showcase the simulation of (3) with higher-order time integration methods, we outline the implementation of the second-order Runge-Kutta scheme (RK2), with  $\Delta t = T_a/m$ . In this case, we have:

$$\vec{x}_{ij}^{t_a + \frac{\Delta t}{2}} = \vec{x}_i^{t_a} + \vec{u}_{ij}(\vec{x}_{ij}^{t_a}, t_a) \frac{\Delta t}{2}, \quad (4)$$

$$\vec{x}_{ij}^{t_a + \Delta t} = \vec{x}_{ij}^{t_a + \frac{\Delta t}{2}} + \vec{u}_{ij}(\vec{x}_{ij}^{t_a + \frac{\Delta t}{2}}, t_a + \frac{\Delta t}{2}) \Delta t, \quad (5)$$

for  $1 \leq i, j \leq N_e$ . Note that when  $t = t_a$ ,  $\vec{x}_{ij}^t = \vec{x}_i^{t_a}$  and  $\vec{u}_{ij}(\vec{x}_{ij}^t, t) = \vec{u}_i(\vec{x}_i^{t_a}, t_a)$ .

Generalization of the method to higher-order time integration methods, such as the fourth-order Runge-Kutta scheme (RK4), is straightforward. Aside from the additional cost of carrying out four sub-steps instead of two, and the associated cost of interpolating the velocity according to Eq. (2), the growth in the number of particles with time using the RK4 scheme is the same as that using the RK2 scheme. Specifically, if we start with one particle at a given location at  $t = 0$ , the number of particles grows as  $N_e^2, N_e^3, \dots, N_e^{n+1}$  at times  $T_a, 2T_a, \dots, nT_a$ .

#### 3.3. Binning

To control the exponential growth in the number of elements, a binning procedure is implemented. The procedure is in spirit similar to that presented in Höllt et al. (2015), where binning was employed to create probability maps that provide the probability of having a particle at a given location and time. In contrast to the procedure in Höllt et al. (2015), we propose a binning methodology that is spatially adaptive and that conserves the zeroth, first, and second moments of

probability. Specifically, the binning merges all particles that fall in the same bin and are associated with a particular realization, into a single particle that belongs to the same realization. We note at this point that preserving the realization identity of the particles, while binning, is necessary because the particle velocity associated with that particular realization is needed for the first step of the RK time integration scheme from  $kT_a$  to  $kT_a + \Delta t$ . Merging is performed in such a way that enables us to conserve the total probability, the mean position, and the variance.

To this end, we represent the probability field,  $p(\vec{x}, t)$ , as the convolution

$$p(\vec{x}, t) = \sum_{l=1}^N S_l \phi_{\vec{x}_l}(\vec{x} - \vec{x}_l), \quad (6)$$

where  $\phi_{\vec{x}_l}(\vec{x} - \vec{x}_l)$  is a basis function that describes the probability distribution due to element  $l$ , of strength (or weight)  $S_l$ , and located at  $\vec{x}_l$ . The basis function is radial (i.e. depends on  $r = |\vec{x} - \vec{x}_l|$ ) and integrates to unity, i.e.  $\int_0^\infty \phi(r) 4\pi r^2 dr = 1$ . In addition, the function and its derivatives vanish at infinity. In this work, we choose  $\phi$  to be the Gaussian distribution

$$\phi_{\vec{x}_l}(\vec{x}^* - \vec{x}_l^*) = \frac{1}{\pi^{3/2} \tilde{\sigma}_l^3} e^{-|\vec{x}^* - \vec{x}_l^*|^2}, \quad (7)$$

where  $\tilde{\sigma}_l^3 = (\sigma_x \sigma_y \sigma_z)_l$ , the superscript  $*$  denotes normalized quantities,  $x$  and  $x_l$  are normalized by  $(\sigma_x)_l$ ,  $y$  and  $y_l$  are normalized by  $(\sigma_y)_l$ , and  $z$  and  $z_l$  are normalized by  $(\sigma_z)_l$ . Note that in addition to satisfying all the requirements above, our choice of  $\phi$  allows for representation of the probability distribution associated with element  $l$  using three separate variances, i.e.  $(\sigma_x)_l^2$ ,  $(\sigma_y)_l^2$  and  $(\sigma_z)_l^2$ , characterizing the uncertainty in the position along the Cartesian coordinates  $x$ ,  $y$ ,  $z$ , respectively. Note also that in the limit  $(\sigma_x)_l$ ,  $(\sigma_y)_l$  and  $(\sigma_z)_l \rightarrow 0$ ,  $\phi(r) \rightarrow \delta(r)/4\pi r^2$ , where  $\delta(r)$  is the Dirac delta function.

Merging all elements,  $l$ , falling in the same bin and associated with a specific realization,  $i$ , into a single element, 1, while conserving the total probability, mean position, and variance is accomplished by satisfying the moment conditions

$$\sum_{l \in \text{bin}, l \in \text{realization } i} S_l \int \int \int_{-\infty}^{\infty} x'^o y'^r z'^q \phi_{\vec{x}_l}(\vec{x}'^* - \vec{x}_l^*) dx' dy' dz' = S_{1,i} \int \int \int_{-\infty}^{\infty} x'^o y'^r z'^q \phi_{\vec{x}_1}(\vec{x}'^* - \vec{x}_{1,i}^*) dx' dy' dz', \quad (8)$$

where  $o, r, q \in \mathbb{N}$  and the subscript 1 refers to the single particle after merging. For our choice of the basis function in (7), the moment conditions (8) yield the following system of equations,

$$\sum S_l = S_1, \quad (9)$$

$$\sum S_l x_l = S_1 x_1, \quad (10)$$

$$\sum S_l y_l = S_1 y_1, \quad (11)$$

$$\sum S_l z_l = S_1 z_1, \quad (12)$$

$$\sum S_l \left( x_l^2 + \frac{1}{2} (\sigma_x)_l^2 \right) = S_1 \left( x_1^2 + \frac{1}{2} (\sigma_x)_1^2 \right), \quad (13)$$

$$\sum S_l \left( y_l^2 + \frac{1}{2} (\sigma_y)_l^2 \right) = S_1 \left( y_1^2 + \frac{1}{2} (\sigma_y)_1^2 \right), \quad (14)$$

$$\sum S_l \left( z_l^2 + \frac{1}{2} (\sigma_z)_l^2 \right) = S_1 \left( z_1^2 + \frac{1}{2} (\sigma_z)_1^2 \right), \quad (15)$$

$$\sum S_l x_l y_l = S_1 x_1 y_1, \quad (16)$$

$$\sum S_l y_l z_l = S_1 y_1 z_1, \quad (17)$$

$$\sum S_l z_l x_l = S_1 z_1 x_1, \quad (18)$$

for  $o + r + q \leq 2$ . Note that the realization index  $i$  has been dropped to simplify the presentation. The strength of element 1 is obtained from Eq. (9). Its position  $\vec{x}_1$  is then determined from Eqs. (10)–(12). Eqs. (13)–(15) are then solved for the variances  $(\sigma_x)_1^2$ ,  $(\sigma_y)_1^2$  and  $(\sigma_z)_1^2$ , respectively. It can be then shown that once the zeroth and the first moments (Eqs. (9)–(12)) are satisfied, the cross moments Eqs. (16)–(18) are automatically satisfied by virtue of the radial symmetry of  $\phi$ . At the end of the merging procedure, the number of particles per bin cannot exceed the number of realizations,  $N_e$ .

Upon inspecting Eqs. (13)–(15), one can observe that although using larger bins produces a smaller number of particles, this will lead to an increase in the variance. For example, if all the particles before binning have zero variance, they are merged into a single particle with variance  $(\sigma_x)_1^2 = \frac{2 \sum S_l (x_l - x_1)^2}{S_1}$ ,  $(\sigma_y)_1^2 = \frac{2 \sum S_l (y_l - y_1)^2}{S_1}$ , and  $(\sigma_z)_1^2 = \frac{2 \sum S_l (z_l - z_1)^2}{S_1}$ . In this case, the sum of the  $x$ ,  $y$  and  $z$  variances is the post-merging variance

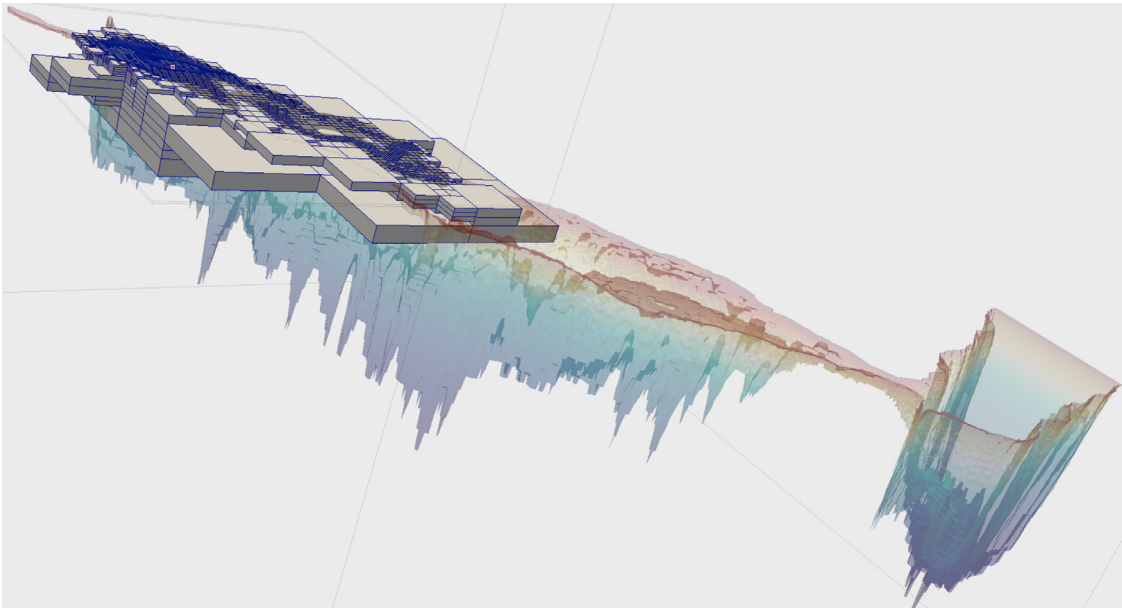


Fig. 2. An example of the adaptively sized bins.



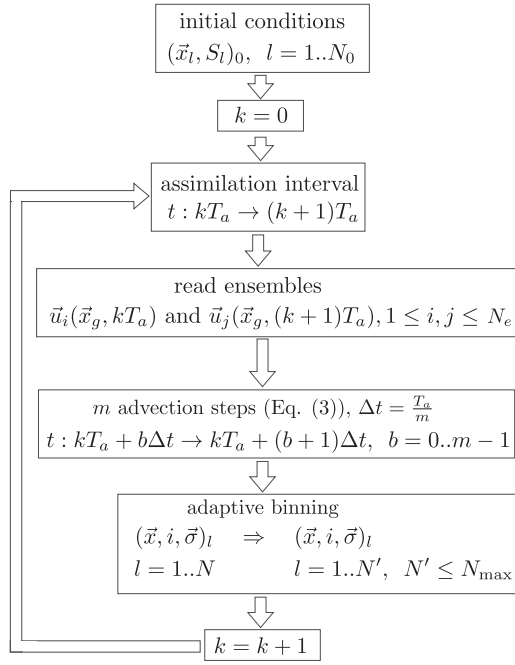


Fig. 3. Flowchart of the forward tracking algorithm.

$$\sigma_1^2 = \frac{2 \sum s_l |\vec{x}_l - \vec{x}_1|^2}{s_1}.$$

### 3.4. Adaptive binning and refinement

Spatially adaptive binning offers the possibility of controlling the variance by proper selection of the bin size. An example is shown in Fig. 2 (See also Fig. 11 for top views.). In the adaptive binning algorithm proposed here, the variance is chosen to be inversely proportional to the total probability strength in the bin,  $S_{\text{bin}} \equiv \sum_{i \in \text{bin}} S_i$ . Thus we seek to enforce  $S_{\text{bin}} \sigma_{\text{bin}}^2 = \text{constant}$ , where  $\sigma_{\text{bin}}^2$  is a measure of the average variance over all realizations after merging. This relationship between the bin size and the variance ensures that regions of high probability are also regions of low variance, and vice versa.

Starting with a rectangular box that contains all the particles, binning is accomplished by a recursive division of the bins using an oct-tree. Strict enforcement of the condition  $S_{\text{bin}} \sigma_{\text{bin}}^2 = \text{constant}$  requires solving the moments Eq. (8) every time a bin is divided (for all nodes at all levels of the oct-tree), which makes the procedure computationally expensive. To mitigate these costs, an alternative procedure is implemented in which a bin is divided into eight bins if its total probability is greater than a characteristic value,  $S^*$ . Due to the dependence of the post-merging variance on the distance between the weighted mean position and the positions of the elements to be merged (Eqs. (13)–(15)), dividing a bin into eight sub-bins limits the increase in variance by roughly a factor of four. Thus, the desired dependence of the variance on the probability is established.

### 3.5. Capping the number of particles

An attractive feature of both the fixed-size and adaptive binning procedures is that they allow capping the number of particles to a desired maximum,  $N_{\text{max}}$ . This feature not only guarantees that the code does not run out of memory, but also offers tradeoff between accuracy and speed. By choosing a smaller  $N_{\text{max}}$ , shorter simulation times can be achieved due to the smaller number of bins. The bin sizes are however larger in this case, which results in higher variances. Enforcing the condition  $N < N_{\text{max}}$  is accomplished by defining a characteristic bin probability  $S^*$ , selected so that the number of particles after binning satisfies  $N < N_{\text{max}}$ . This is done by iteratively running a series of virtual

binning processes, whereby at the end of each process,  $S^*$  is adjusted depending on the number of particles. In particular, if after iteration  $\kappa$ ,  $N_\kappa > N_{\text{max}}$ , we set for iteration  $\kappa + 1$ ,  $S_{\kappa+1}^* = \beta \frac{N_{\text{max}}}{N_\kappa} S_\kappa^*$ , where  $0 < \beta \leq 1$ . Note that virtual binning is a fast process because, except for the last iteration, it does not entail solving the moments equations.

### 3.6. Boundary conditions

In this section, we discuss the boundary conditions at the free-surface and at the beach. Motion of the free-surface may cause the particle to be advected above the top layer of the computational grid. In this case, the particle's velocity is obtained by extrapolating the velocities from the uppermost two layers.

When the flow field, over a given time-step ( $\Delta t$ ), advects a particle into the beach, the time step is recursively halved until the particle remains in water at the end of the time step or a minimum threshold time step ( $\Delta t_{\text{min}}$ ) is reached. Under the latter condition, a simple beaching model is implemented. If the flow over  $\Delta t_{\text{min}}$  pushes a particle inland, the point of intersection of the particle's trajectory with the coast is determined. The particle is then advected with a fraction of the time-step, sufficient for the particle to reach the coast. After that, the particle is advected for the remainder of the time step using the tangential component of the velocity at the shore (enforcing no-through flow boundary condition). An alternative boundary condition, that is also implemented in our model and used in the experiments presented in the results section, is to trap the particle at the coast.

Other strategies for handling these two boundary conditions are possible. For instance reflection at the free-surface can be treated as in North et al. (2011). The particle-beach interaction involves numerous uncertainties and variables that include the properties of the coast and particles (Neves et al., 2015; Samaras et al., 2014). The beaching model can be extended to combine the two aforementioned models by splitting the particle that reaches the coast into two particles. One particle is kept at the shore while the other is advected using the tangential component of the velocity. The probabilities of these two particles are based on the tendency of a particle to stay onshore, which is a function of the properties of the particles and the type of the coast.

### 3.7. The forward tracking algorithm

A flowchart of the forward tracking algorithm is presented in Fig. 3. In addition to reducing the number of particles by binning, further reduction in the number of particles is achieved by removing all the particles with strengths below a cutoff value.

### 3.8. Backward tracking method

In this section, we present a backward Lagrangian tracking algorithm for identification of probable sources at time  $t_0 - T_{\text{eo}}$ , based on observations at time  $t_0$ . As in the forward tracking algorithm, we assume the particles to be passive tracers that are transported by the velocity field. We also assume that the time elapsed,  $T_{\text{eo}}$ , from the instant,  $t_0$ , is known. The problem then reduces to solving for a particle's position at a previous time,  $\vec{x}(t - \Delta t)$ , the following Lagrangian advection equation

$$\frac{d\vec{x}}{dt} = \vec{u}(\vec{x}, t). \quad (19)$$

Integrating from  $t - \Delta t$  to  $t$  yields

$$\vec{x}(t - \Delta t) = \vec{x}(t) + \int_{t-\Delta t}^t -\vec{u}(\vec{x}, t') dt'. \quad (20)$$

The same procedure adopted for the forward tracking problem is adopted here, with the only difference being the use of the reverse velocity  $-\vec{u}(\vec{x}, t)$ . This backward procedure enables us to build

probability maps for the originating source, starting from observations at a given later time.

It should be noted, however, that although the velocity field for the backward tracking problem is the reverse of that of the forward problem, the variance continuously increases in both cases. This introduces intrinsic irreversibility in the problem that cannot be removed, but can at best be managed using the proposed adaptive binning algorithm.

#### 4. Results and discussion

Forward and backward experiments were conducted to validate our methods. These experiments are summarized in Table 1. Except for the experiment of Section 4.5, all the forward experiments are performed according to the following scenario: a particle carrying an initial strength of  $S_0 = 10^9$  is released from a fixed point source located at 37° latitude, 24° longitude and 2.5 m below sea surface in the Red Sea. This source is shown as a black square on each map. Next, particles are advected for a period specified by *Travel time* using a fourth order Runge–Kutta scheme (RK4) and a fixed time step  $\Delta t$ . In all the experiments, we chose  $\Delta t_{\min} = 60$  s. To investigate various sensitivity experiments and scenarios, the ensemble of flow fields ( $u_i^{(j)}, j = 1, 2, \dots$  and  $i = 1, \dots, N_e$ ) used for advection are obtained from the original ensemble ( ${}^o u_i^{(j)}, j = 1, 2, \dots$  and  $i = 1, \dots, N_e$ ) described in Section 2 by generating new members having the same average but a fraction,  $f^2$ , of the variance of the original dataset, as

$$u_i^{(j)}(\vec{x}_g, t_a) = f^2 \cdot {}^o u_i^{(j)}(\vec{x}_g, t_a) + (1 - f^2) \overline{{}^o u^{(j)}}(\vec{x}_g, t_a), \quad i = 1, \dots, N_e, \quad (21)$$

where  $\overline{{}^o u^{(j)}}$  is the average of the underlying ensemble. At the end of each forward simulation, a probability map is obtained according to Eq. (6). Note that the initial strength is selected to be sufficiently large to avoid reaching machine precision due to the exponential growth in the number of particles. In all the probability distribution figures, the actual probability is the reported value in the color bar divided by the initial strength. These values may be compared with a reference value of  $4.3 \times 10^{-5}$ , which is the value of the (actual) probability if the initial strength were distributed uniformly over the entire volume of the Red Sea. The backward tracking experiments are conducted as follows: once a forward simulation is executed over the desired *Travel time* and the probability map is computed, the location of the highest probability is chosen to be the location of the source for the backward run. All other parameters remain the same as those of the forward experiment. As outlined in Table 1, we assess the sensitivity of the proposed tracking system and its outputs to the following parameters: (i) the maximum number of particles  $N_{\max}$ , (ii) the time step  $\Delta t$ , (iii) the variance of the ensemble velocity field, (iv) the travel time, and (v) the spatial distribution of the ensemble velocity variance. We conclude the results section by showcasing adaptive binning along with the associated growth in the uncertainty in particles locations and commenting on the computational cost of the algorithm.

The method offers a tradeoff between accuracy and cost through the choice of the maximum number of elements ( $N_{\max}$ ) and the time step ( $\Delta t$ ). To reduce the computational cost and the simulation time, we chose the minimum possible  $N_{\max}$  and the maximum possible  $\Delta t$  without sacrificing accuracy, as discussed below.

##### 4.1. Sensitivity to the maximum number of particles

The impact of the maximum number of particles on the probability distribution is first investigated in forward/backward experiments as described in the first row of Table 1. The results for the forward experiments are displayed in the top row of Fig. 4 for  $N_{\max} = 0.5, 1, 50$  and 300 million, whereas those of the backward runs are displayed in the bottom row of the same figure. As can be seen, as  $N_{\max}$  increases from 1 million to 300 million, the probability maps obtained at the end

of the forward and forward/backward tracking simulations do not change appreciably. For smaller  $N_{\max}$ , loss in accuracy and resolution is expected. This can be observed for the case with  $N_{\max} = 0.5$  million particles, where the probability distribution is different in terms of shape and coverage.

Note that in the absence of binning, a particle is split into  $50^{10}$  particles after 27 days of advection; as discussed in Section 3.2. As  $N_{\max}$  continues to increase, the probability map is expected to converge toward the map for which no binning is performed. This is because a larger  $N_{\max}$  results in smaller bins, as discussed in the methods section.

For the backward tracking experiments (bottom row), one can notice two high probability zones; one along the coast and the other in the vicinity of the source. Due to the high variability characterizing the ensemble velocity field and the large coast length to surface area ratio, we expect to have a high probability coastal zone in most of the cases, namely because the beaching model traps the particles once they reach the coast.

The distance,  $\delta$ , between the location of highest probability in this region (green square in Fig. 4) and the source (black square) is a measure of the accuracy of the method. Sources of inaccuracy include the resolution of the computational grid of the velocity field, the large uncertainty in the ensemble velocity fields (the standard deviation is almost of the same order as the mean, as can be seen in Fig. 10), the simple beaching method adopted, and the smoothing introduced by the binning process which, for the forward/backward problem, takes place over  $2 \times 27$  days. Despite all of these challenges in this extreme scenario, the method was able to locate the source to within  $\delta \leq 45$  km, as can be seen in Fig. 4.

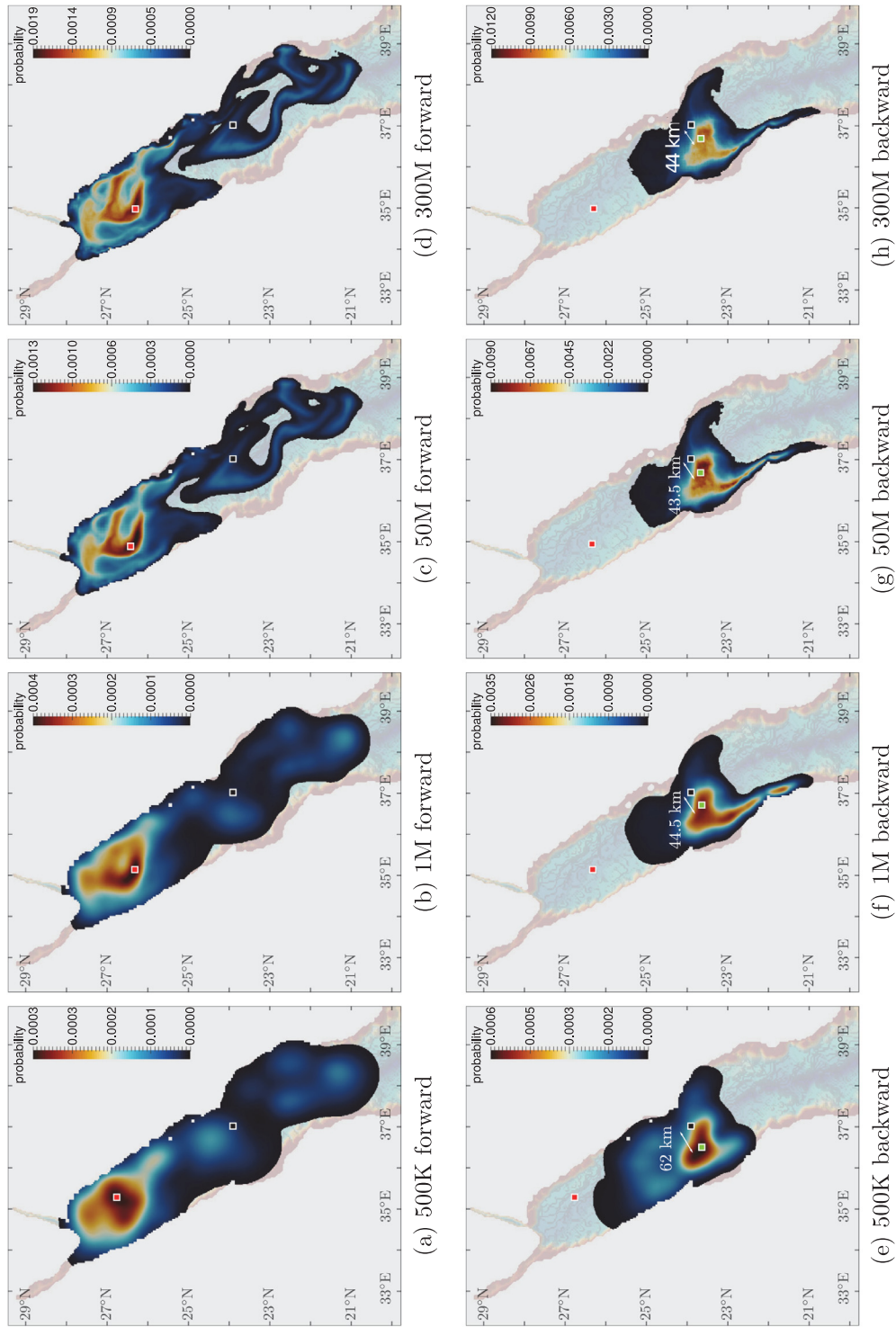
##### 4.2. Sensitivity to the time step

In this experiment, we study the sensitivity of the probability map,

**Table 1**

Summary of parameters used in forward (label F), and backward (label B), experiments. Seven experiment types are conducted focusing on the effect of the maximum number of particles, the time step, the variance of the ensemble, the travel time, the source location, and history of transport. Unless otherwise indicated, the source is located at 37° latitude, 24° longitude and 2.5 m below sea surface.

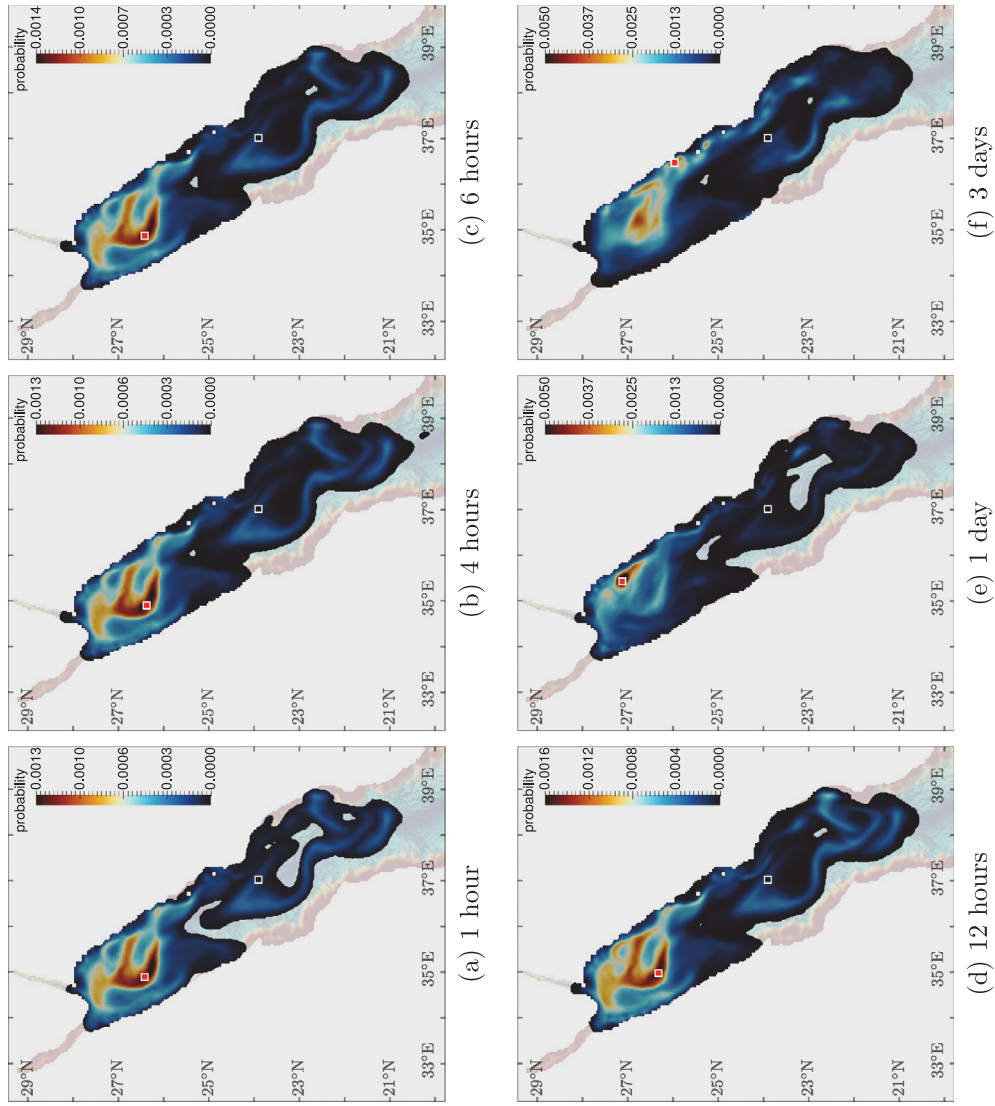
Experiment	Parameters				Tracking
	$f$	$N_{\max}$ (Million particles)	$\Delta t$	Travel time (days)	
Max. number of particles Fig. 4	1	0.5 1 50 300	1 hr	27	F/B
Time step Figs. 5 and 6	1	50	1 hr 4 hr 6 hr 12 hr 1 d 3 d	27	F
Variance (ensemble) Fig. 7	0.25 0.5 0.75 1	50	1 hr	27	F/B
Travel time Fig. 8	1	50	1 hr	3 6 12 18	F/B
Source location Fig. 9	1	50	1 hr	27	F/B
History of transport Fig. 11	0.75	50	1 hr	3 6 12 24	F



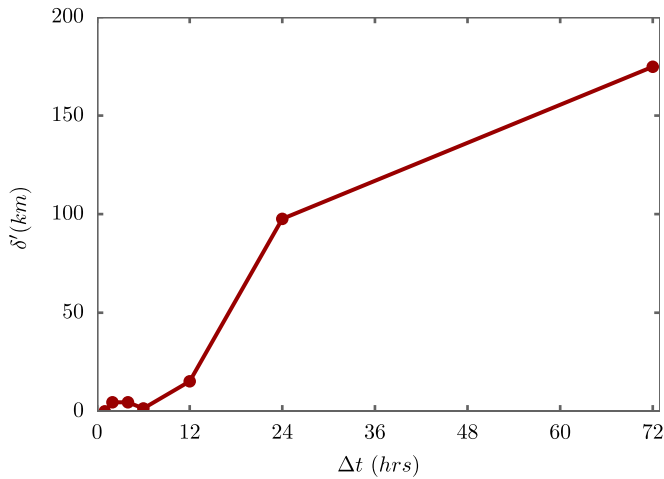
**Fig. 4.** Probability maps at  $t = 27$  days. Plots are generated for forward (top) and backward (bottom) tracking for different values of  $N_{\max}$  using an ensemble flow field of 100% the standard deviation of the initial dataset and a time step of 1 h. Black square: source. Red and green squares: locations of highest probability at the end of forward and backward tracking, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

obtained after 27 days of tracking, to the value of the time step used in the RK4 advection scheme. Simulations using  $\Delta t = 1, 4, 6, 12, 24$  and 72 h are carried out in the forward/backward experiments as outlined in the second row of Table 1, where the maximum number of elements is set to 50 million. The corresponding results are shown in Fig. 5. We notice that the probability map, and especially the regions of high probability, do not change much for  $\Delta t \leq 12$  h. As the time step gets

larger ( $\Delta t = 1$  and 3 days), the probability maps start to lose their similarity, and a remarkable shift in the position of the point of highest probability (red square on the map) is observed. This is confirmed by the plot in Fig. 6. The error in this plot is defined as the distance from the point of highest probability, obtained with a time step  $\Delta t$ , to that of highest probability obtained with  $\Delta t = 1$  h, considered as a surrogate of the true position of the particle. Note that for  $1 \text{ hr} \leq \Delta t \leq 6 \text{ h}$ , this



**Fig. 5.** Probability maps at  $t = 27$  days for forward tracking of particles (released from the same source, indicated by a black square). Plotted are the distributions obtained using different time steps, as indicated. In all cases,  $N_{\max} = 50$  M, and  $f = 1$ .



**Fig. 6.** Distance,  $\delta'$ , between the maximum probability point at  $t = 27$  days obtained using  $\Delta t = 1$  hr, and the maximum probability point using  $\Delta t \geq 1$  hr. Forward tracking from source ( $37^\circ$  latitude,  $24^\circ$  longitude and 2.5m below surface), with  $N_{\max} = 50$  M.

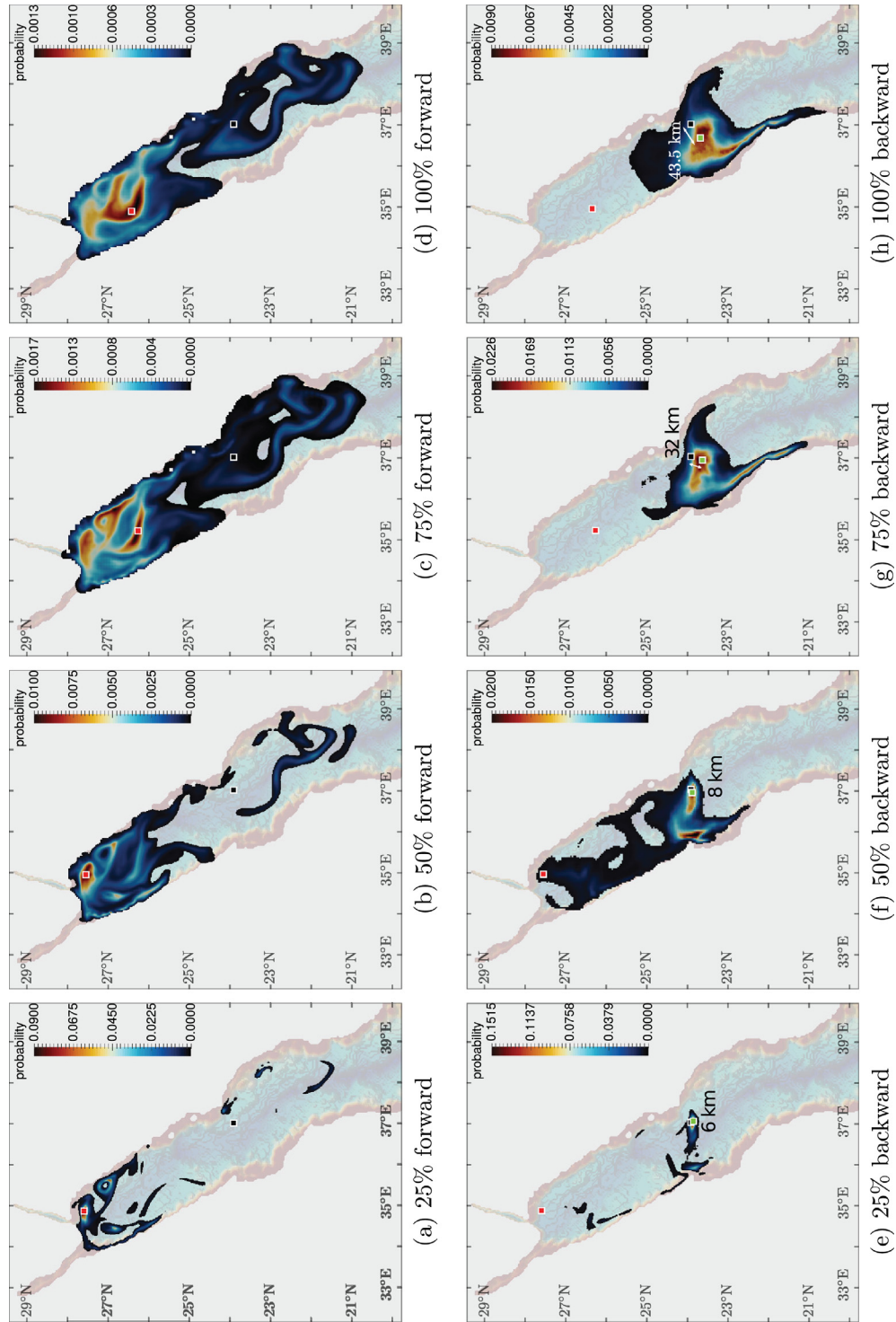
distance ranges between 1 km and 4.2 km, which is of the order of the grid cell size of the computational domain (4.1 km). To ensure good accuracy in terms of the time step size, we chose, conservatively,  $\Delta t = 1$  hr for all the experiments that follow.

#### 4.3. Sensitivity to the ensemble flow-field variance

The variance parameter,  $f^2$ , is introduced to control the uncertainty in the flow field. In this experiment, we study the effect of  $f$  on the ability of the system to recover the source. To this end, four forward/backward simulations (Table 1, third row) were carried out using ensemble flow fields of different standard deviations (25%, 50%, 75% and 100% of the standard deviation of the original dataset). The results are shown in Fig. 7 in terms of probability maps for the forward and backward simulations. For the forward simulations, we notice that for low variances, the region of high probability is more localized (Fig. 7 (a) and (b)). In fact, as the variance gets smaller, the point of highest probability gets closer to the position of a particle advected from the source, over the same travel time, using the ensemble average velocity. As the variance increases, the region of high probability is more spread out (Fig. 7 (c) and (d)).

For the backward simulations, the source of release is closest to the





**Fig. 7.** Probability maps for forward tracking (top row) and backward tracking (bottom row) experiments using values of  $f$  as indicated. In all cases,  $\Delta t = 1$  hr and  $N_{\max} = 50$  M. The sources location is marked by a black square.

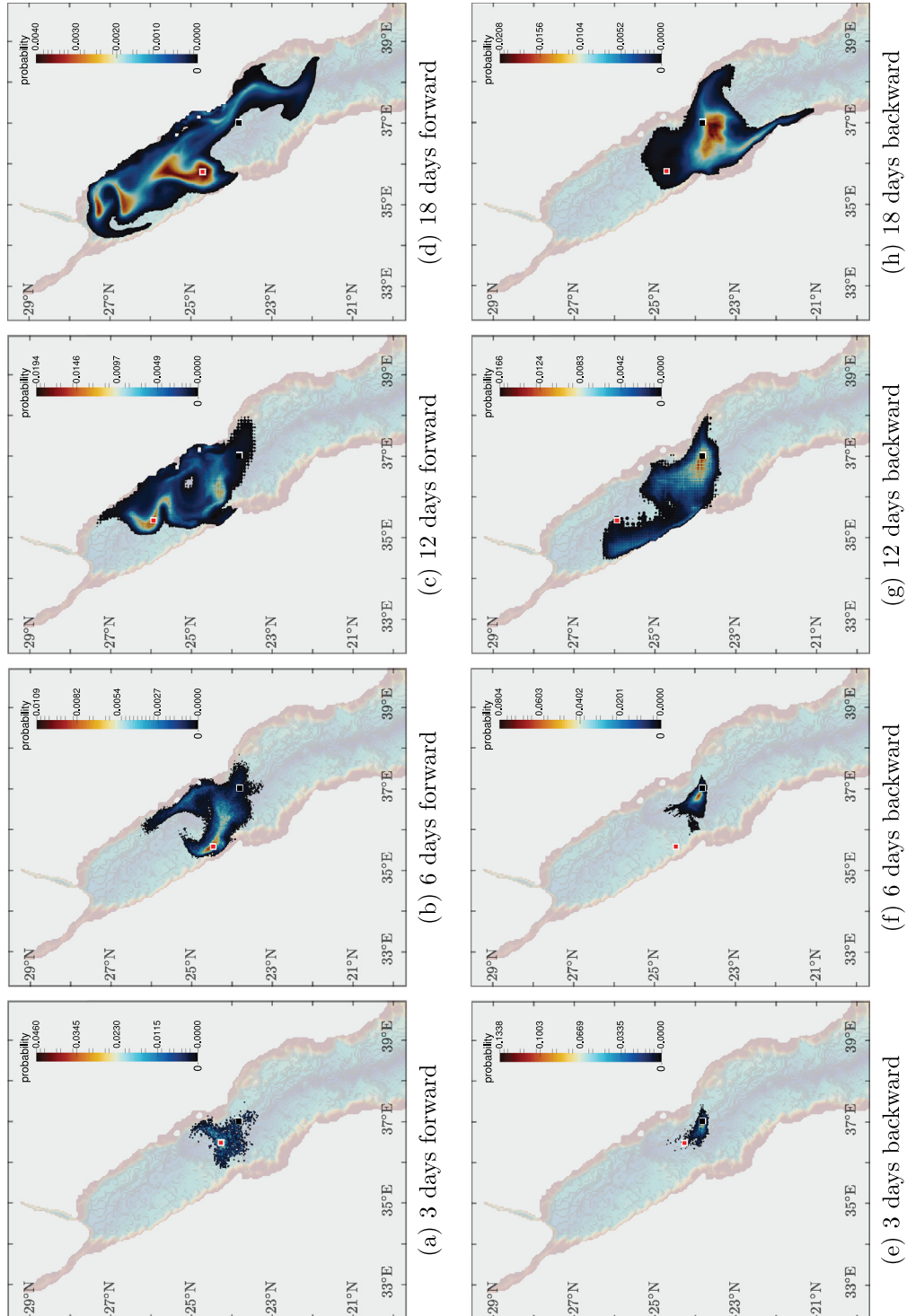
center of the region of high probability when the standard deviation of the ensemble velocity field is smallest (probability maps (e)). As the uncertainties in the ensemble velocity field grow, the center of the region of high probability moves farther away from the source (probability maps (h)). Thus, the source is predicted more accurately when the uncertainties in the ensemble velocity field are small. The distance between the recovered source (green square) and the actual source (black square) decreases from  $\sim 30$  km when  $f = 1$  to 6 km when  $f = 0.25$ , as seen in Fig. 7. As discussed earlier, binning also introduces

uncertainties in the particles positions that increase with the bin size. For the forward problem, binning takes place 9 times, whereas for the forward/backward problem, it takes place 18 times. The experiments show that while using the average velocity field is sufficient for predicting the source when the variance of the ensemble velocity field is small, this is not the case when the variance of the ensemble velocity field is large. Thus, when the variability is large, using the ensemble average may no longer be suitable for predicting the regions of highest probability.

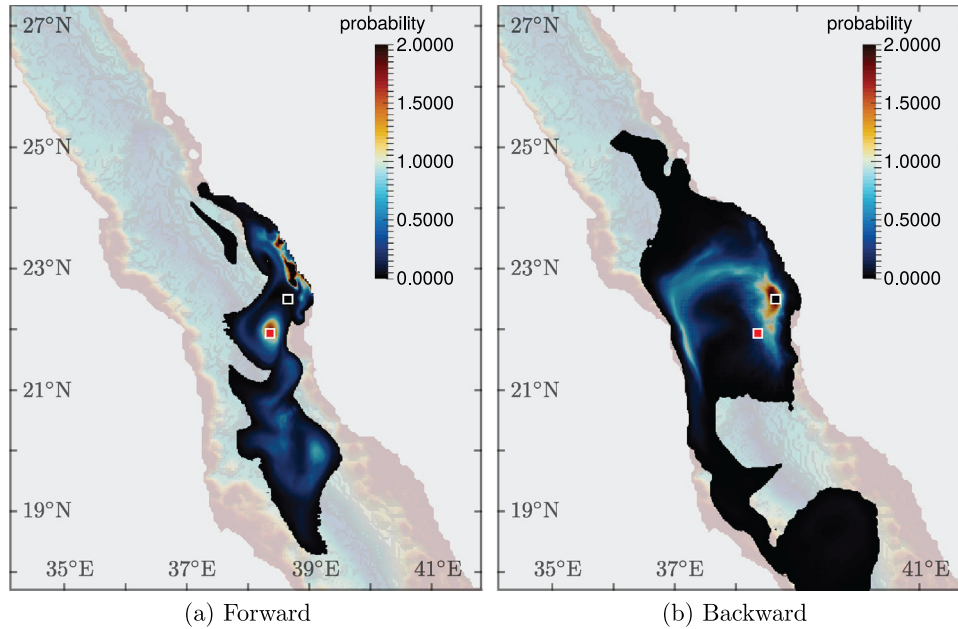
#### 4.4. Sensitivity to the travel time

In the experiments discussed above, the particles were advected forward (and backward) in time for 27 days, which corresponds to 9 assimilation cycles. We expect that the greater the travel time, the more difficult it is for the backward simulation to locate the source, due to the accumulation of the uncertainty introduced by binning, in addition to trapping at the beach. To illustrate this, we ran a series of forward/backward simulations using travel times of 3, 6, 12 and 18 days

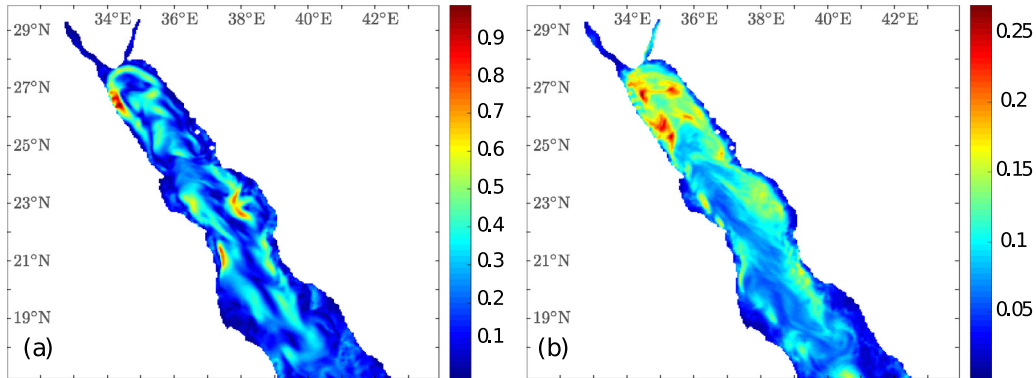
(Table 1, fourth row). The results are shown in Fig. 8. From the probability maps of the forward simulations (top row), we notice that, for short travel times, the region of high probability is more localized (Fig. 8(a–c)), whereas for larger travel times, it is more spread out (Fig. 8(d)). From the probability maps of the backward simulations (bottom row), and for short travel times, the source is either included or close to a region of high probability, which, again, is more localized (Fig. 8(e–g)). For larger travel times, the region of high probability is more spread out (Fig. 8(h)). These observations are consistent with the



**Fig. 8.** Probability distributions for forward (top row) and backward (bottom row) experiments. Plotted are results for different values of travel time, as indicated. For all cases,  $N_{\max} = 50$  M,  $\Delta t = 1$  hr, and  $f = 1$ , source is marked using a black square.



**Fig. 9.** Left: probability map for forward tracking of particles from a source (black square) near the coast of Yanbu. Right: probability map for backward tracking of particles from a point of high probability (red square) obtained from the forward simulation.  $\Delta t = 1$  hr,  $N_{\max} = 50$  M, and  $f = 1$ . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 10.** Maps of (a) average speed (m/s) and (b) standard deviation (m/s) of the ensemble velocity field at  $t_a = 18$  days.

conclusions made at the end of Section 4.1. For travel times of 3, 6, 12 and 18 days, the distances  $\delta$  between the source and recovered source are respectively 7, 23, 31, and 34 km.

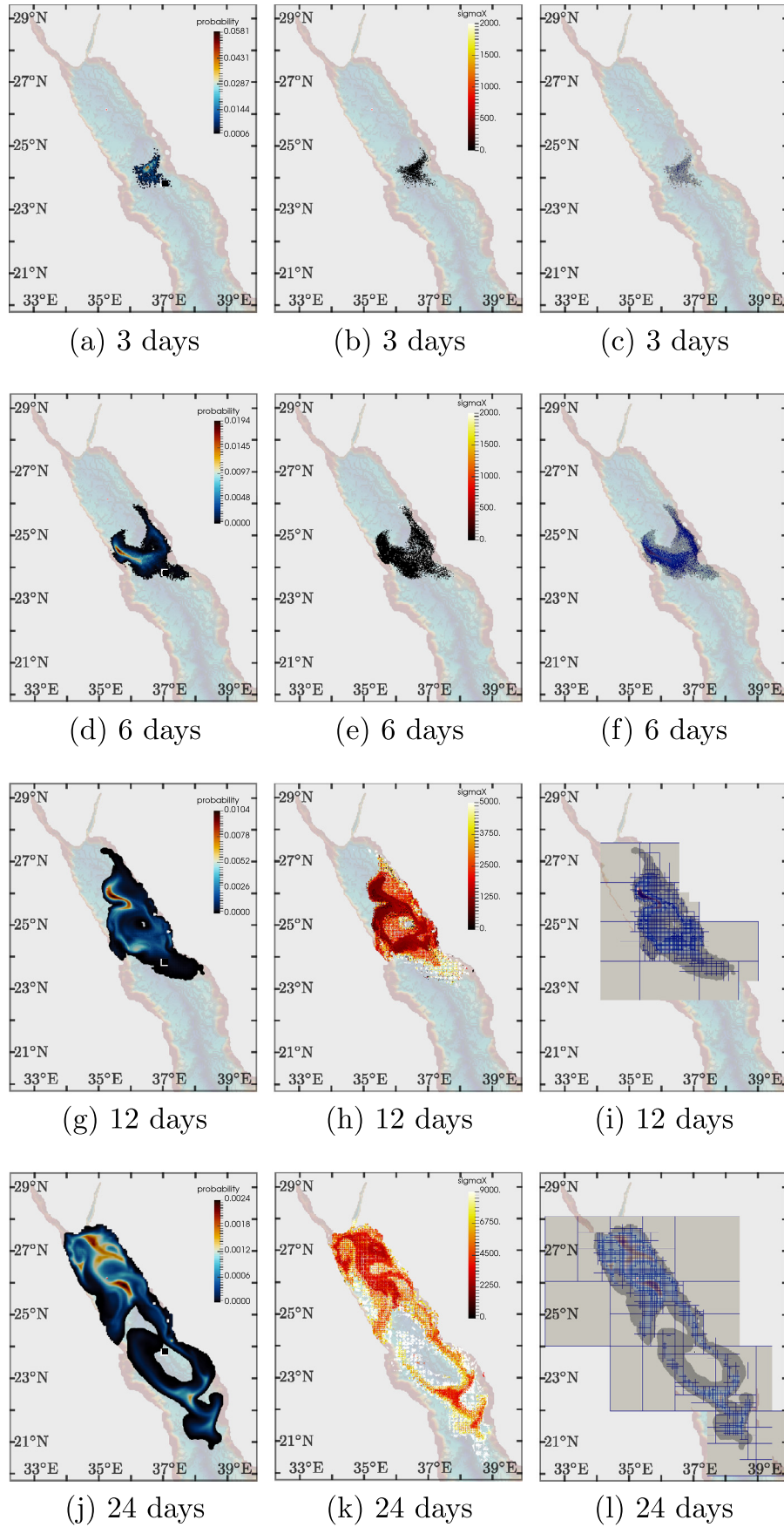
#### 4.5. Sensitivity to the source location

In all the experiments we presented so far, the particles were released from a fixed point source located at  $37^\circ$  latitude,  $24^\circ$  longitude and 2.5 m below sea surface in the Red Sea. The probability distributions at the end of the forward and forward/backward tracking period capture the impact of the spatio-temporal variability of the flow field on the particles trajectories. Subsequently, the accuracy of the source recovery is not only a function of the travel time, but also of the source location.

To illustrate the impact of the spatio-temporal variability of the ensemble flow field on the source location, we conducted a forward/backward runs with a source located at  $38^\circ 41' 50''$  longitude,  $22^\circ 35' 21''$  latitude and on the sea surface. The simulation parameters are listed in Table 1, row 6. The probability map of the forward simulation (Fig. 9 (a)) showed multiple regions of high probability, all of which, except one, are near the coast. This highlights the challenge posed by beaching in domains characterized by a high coastal length to surface area ratio,

such as the red sea. When we choose the location of highest probability in the region away from the coast (red square) as the source of a backward simulation, the backward simulation predicts a probability map with a single high probability region centered at the initial source of particle release (the black square). This prediction is more accurate than that of the forward/backward problem with the source located at  $37^\circ$  latitude,  $24^\circ$  longitude, and where the probability maps are shown in Fig. 7(d) and (h). Note that both cases have the same travel time (27 days), maximum number of particles (50 million), and ensemble velocity fields (using the standard deviation of the original set). This difference in the accuracy of source recovery may be attributed to the spatial variability of the uncertainties of the ensemble velocity field. If particles are advected to regions in which the uncertainties in the flow are high, the source recovery from the probability map of the backward simulation will be more uncertain. Fig. 10 shows the spatial distribution (at the surface) of the average and the standard deviation of the 50 members of the ensemble velocity field, for  $t_a = 6T_a = 18$  days. It can be seen that the in the forward problem with the source located at  $37^\circ$  latitude,  $24^\circ$  longitude, the particles travel in regions of higher variability (the northern part of the Red Sea) than those in the case with the source located at  $38^\circ 41' 50''$  longitude,  $22^\circ 35' 21''$ . The particles in the latter case travel through the central region following a large eddy





**Fig. 11.** Left: probability maps for the forward tracking problem at different travel times, as indicated. Center: corresponding maps of  $\sigma_x$ . Right: corresponding bin size distributions.  $\Delta t = 1$  hr,  $N_{\max} = 50$  M, and  $f = 1$ .



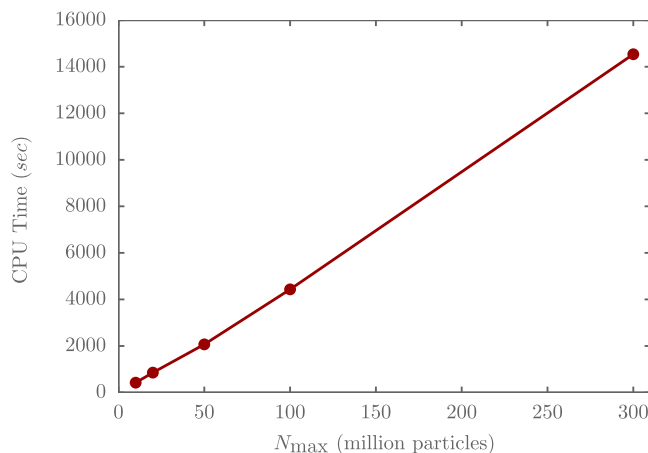


Fig. 12. Computational cost, measured in CPU seconds versus the maximum number of particles.

(Fig. 9) from the source (black square) to the point of high probability (red square) over a period of 27 days. By comparing Fig. 10(a) and (b), one can also see that there are regions where the standard deviation is as large as the average, making the prediction of the source location an even more challenging task when the particles trajectories pass through these regions.

#### 4.6. Adaptive binning and variance control

As mentioned in Section 3.4, adaptive binning endows the method with the ability to limit the increase in the variance of the position of merged particles, whenever the sum of the probability carried by those particles is high. To illustrate this feature, probability maps (Fig. 11(a), (d), (g), (j)) of a forward simulation were obtained at different times (3, 6, 12 and 24 days), along with maps showing the particles variance distribution (Fig. 11(b), (e), (h), (k)), and the bins distribution (Fig. 11(c), (f), (i), (l)). As can be seen in the figures, the regions of high probability are resolved using smaller bins. Accordingly, the variances ( $\sigma_x^2$ ,  $\sigma_y^2$ ,  $\sigma_z^2$ ) introduced by binning (see Section 3.3) are smaller in these regions. It can also be noticed that for short travel times (3 and 6 days) these variances are very small, reflecting the fact that very little binning takes place at these early stages, given that the number of particles has not yet been reached  $N_{\max}$ .

#### 4.7. Computational cost

Fig. 12 outlines the computational cost, measured in CPU seconds, versus the maximum number of particles for the forward experiment over a duration of 27 days. Other parameters are listed in the first row of Table 1. The simulations were carried using a parallelized version of the algorithm presented in Fig. 3 on a Intel Xeon E5-2680 v4 2.4 GHz Fourteen-Core (28 threads) LGA 2011-3 Processor. Parallelization of the method is implemented using OpenMP<sup>1</sup>, where each core (or thread) advects a fraction of the particles according to the algorithm shown in Fig. 3. The particles are equally distributed among the cores. As expected, the plot shows that the CPU time scales linearly with the number of particles.

## 5. Conclusions

We presented an efficient algorithm for advecting particles in stochastic flow fields, whose statistics are prescribed by an underlying ensemble. The algorithm, which employs high order particle advection

schemes, introduces an elements growth control mechanism based on an adaptive binning procedure, while conserving the zeroth, first and second moments of probability (total probability, mean position, and variance). The adaptive binning process offers a tradeoff between speed and accuracy by limiting the number of particles to a desired maximum. The forward and backward experiments, implemented in a parallel computational framework using openMP, demonstrated that the method is capable of recovering the source location using a relatively small number of particles ( $N_{\max} = 50$  M). Using the original ensemble velocities, where the standard deviation is as large as the mean in significant parts of the domain, and capping the number of particles to 50 million, the method succeeded in recovering the source location to within less than 40 km, for a travel time of 27 days. The distance separating the actual source and the recovered source (location of maximum probability in a forward/backward experiment) decreased from  $\sim 40$  km to  $\sim 6$  km when the standard deviation of the ensemble velocity is reduced by a factor of 4, which suggests the importance of reducing the uncertainty in the assimilated products. The accuracy of source recovery was also shown to increase for shorter travel times.

The novelty of the method, manifested by higher order advection scheme along with the adaptive and conservative binning procedure, will lead to improvements in cost, speed and accuracy, all essential to online tracking.

Future directions include extending the methodology to model specific species transport such as oil spills and incorporation of an advanced beaching model that takes into account the species properties and the beach characteristics.

The numerical implementation of the algorithm will be publicly available online in the near future.

## Acknowledgment

This work is partially supported by the University Research Board of the American University of Beirut, and by King Abdullah University of Science and Technology.

## References

- Batchelder, H., 2006. Forward-in-time-/backward-in-time-trajectory (fitt/bitt) modeling of particles and organisms in the coastal ocean. *J. Atmos. Oceanic Technol.* 23 (5), 727–741.
- Beaudoin, A., De Dreuz, J.-R., Erhel, J., 2007. An efficient parallel particle tracker for advection-diffusion simulations in heterogeneous porous media. *European Conference on Parallel Processing*. Springer, pp. 717–726.
- Brickman, D., Smith, P., 2002. Lagrangian stochastic modeling in coastal oceanography. *J. Atmos. Oceanic Technol.* 19 (1), 83–99.
- Cole, M., Lindeque, P., Halsband, C., Galloway, T.S., 2011. Microplastics as contaminants in the marine environment: a review. *Mar. Pollut. Bull.* 62 (12), 2588–2597.
- Guo, H., He, W., Seo, S., Shen, H.-W., Peterka, T., 2016. Extreme-Scale Stochastic Particle Tracing for Uncertain Unsteady Flow Analysis. Technical Report. Argonne National Laboratory (ANL).
- Guo, H., Yuan, X., Huang, J., Zhu, X., 2013. Coupled ensemble flow line advection and analysis. *IEEE Trans. Vis. Comput. Graph.* 19 (12), 2733–2742.
- Hernandez, F., Traon, P.-Y.L., Barth, N.H., 1995. Optimizing a drifter cast strategy with a genetic algorithm. *J. Atmos. Oceanic Technol.* 12 (2), 330–345.
- Höller, T., Hadwiger, M., Knio, O., Hoteit, I., 2015. Probability maps for the visualization of assimilation ensemble flow data. In: Middel, A., Rink, K., Weber, G.H. (Eds.), *Workshop on Visualisation in Environmental Sciences (EnvirVis)*. The Eurographics Association. <https://doi.org/10.2312/envirvis.20151090>.
- Hoteit, I., Hoar, T., Gopalakrishnan, G., Collins, N., Anderson, J., Cornuelle, B., Köhl, A., Heimbach, P., 2013. A MITgcm/DART ensemble analysis and prediction system with application to the Gulf of Mexico. *Dyn. Atmos. Oceans* 63, 1–23.
- Hoteit, I., Pham, D.T., Blum, J., 2002. A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical Pacific. *J. Mar. Sys.* 36, 101–127.
- Hoteit, I., Pham, D.-T., Gharamti, M.E., Luo, X., 2015. Mitigating observation perturbation sampling errors in the stochastic EnKF. *Mon. Wea. Rev.* 143 (7).
- Isobe, A., Kako, S., Chang, P., Matsuno, T., 2009. Two-way particle-tracking model for specifying sources of drifting objects: application to the east china sea shelf. *J. Atmos. Oceanic Technol.* 26 (8), 1672–1682.
- Nakashima, E., Isobe, A., Kako, S., Itai, T., Takahashi, S., 2012. Quantification of toxic metals derived from macroplastic litter on Ookushi Beach, Japan. *Environ. Sci. Technol.* 46 (18), 10099–10105. <https://doi.org/10.1021/es301362g>. PMID: 22916725

<sup>1</sup> <https://www.openmp.org/>.

- National Research Council, 2012. Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification. The National Academies Press, Washington, DC. <https://doi.org/10.17226/13395>.
- Neves, A.A.S., Pinardi, N., Martins, F., Janeiro, J., Samaras, A., Zodiatis, G., De Dominicis, M., 2015. Towards a common oil spill risk assessment framework—adapting ISO 31000 and addressing uncertainties. *J. Environ. Manage.* 159, 158–168.
- North, E.W., Adams, E., Schlag, Z.Z., Sherwood, C.R., He, R.R., Hyun, K.H.K., Socolofsky, S.A., 2011. Simulating oil droplet dispersal from the deepwater horizon spill with a lagrangian approach. *Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise*. pp. 217–226.
- Poje, A., Toner, M., Kirwan Jr, A., Jones, C., 2002. Drifter launch strategies based on lagrangian templates. *J. Phys. Oceanogr.* 32 (6), 1855–1869.
- Reed, M., Johansen, O., Brandvik, P., Daling, P., Lewis, A., Fiocco, R., Mackay, D., Prentki, R., 1999. Oil spill modeling towards the close of the 20th century: overview of the state of the art. *Spill Sci. Technol. Bull.* 5 (1), 3–16.
- Samaras, A.G., De Dominicis, M., Archetti, R., Lamberti, A., Pinardi, N., 2014. Towards improving the representation of beaching in oil spill models: a case study. *Mar. Pollut. Bull.* 88 (1), 91–101.
- Spaulding, M.L., 1988. A state-of-the-art review of oil spill trajectory and fate modeling. *Oil Chem. Pollut.* 4 (1), 39–55. [https://doi.org/10.1016/S0269-8579\(88\)80009-1](https://doi.org/10.1016/S0269-8579(88)80009-1). <http://www.sciencedirect.com/science/article/pii/S0269857988800091>
- Spaulding, M.L., 2017. State of the art review and future directions in oil spill modeling. *Mar. Pollut. Bull.* 115 (1), 7–19. <https://doi.org/10.1016/j.marpolbul.2017.01.001>. <http://www.sciencedirect.com/science/article/pii/S0025326X17300012>
- Suneel, V., Ciappa, A., Vethamony, P., 2016. Backtrack modeling to locate the origin of tar balls depositing along the west coast of India. *Sci. Total Environ.* 569–570, 31–39.
- Toye, H., Zhan, P., Gopalakrishnan, G., Kartadikaria, A.R., Huang, H., Knio, O., Hoteit, I., 2017. Ensemble data assimilation in the Red Sea: sensitivity to ensemble selection and atmospheric forcing. *Ocean Dyn.* 67 (7), 915–933. <https://doi.org/10.1007/s10236-017-1064-1>.
- van Sebille, E., Griffies, S.M., Abernathey, R., Adams, T.P., Berloff, P., Biastoch, A., Blanke, B., Chassignet, E.P., Cheng, Y., Cotter, C.J., Deleersnijder, E., Ds, K., Drake, H.F., Drijfhout, S., Gary, S.F., Heemink, A.W., Kjellsson, J., Koszalka, I.M., Lange, M., Lique, C., MacGilchrist, G.A., Marsh, R., Adame, C.G.M., McAdam, R., Nencioli, F., Paris, C.B., Piggott, M.D., Polton, J.A., Rhs, S., Shah, S.H., Thomas, M.D., Wang, J., Wolfram, P.J., Zanna, L., Zika, J.D., 2018. Lagrangian ocean analysis: fundamentals and practices. *Ocean Modell.* 121, 49–75. <https://doi.org/10.1016/j.ocemod.2017.11.008>. <http://www.sciencedirect.com/science/article/pii/S1463500317301853>