JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2022MS003484

On Clustering of Floating Tracers in Random Velocity Fields

Jamie Meacham[1] and **Pavel Berloff**[1,2]

[1]Department of Mathematics, Imperial College London, London, UK, [2]Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia

Key Points:

- Expanding on other kinematic studies of clustering buoyant tracers, we include simple temporal structure in the carrying velocity fields
- We utilize global measures of clustering, as well as a cluster tracking algorithm, to compare the process quantitatively between models
- Our results show the sensitivity of clustering to distortions of time structure, which can inform future studies in dynamical contexts

Correspondence to:
J. Meacham,
jom20@ic.ac.uk

Citation:
Meacham, J., & Berloff, P. (2023). On clustering of floating tracers in random velocity fields. *Journal of Advances in Modeling Earth Systems*, *15*, e2022MS003484. https://doi.org/10.1029/2022MS003484

Received 25 OCT 2022
Accepted 8 MAY 2023

Abstract In this paper, we investigate the aggregation of a floating tracer into clusters. Motivated by observations of dense patches of buoyant material in the real ocean (e.g., microplastic pollutants, plankton, and sargassum), we develop an idealized model that can reproduce the clustering process. A stochastic, kinematic 2D velocity field is chosen to represent turbulent oceanic surface currents, with a weakly divergent component. Lagrangian particles are introduced and we track their concentrations. We differ from delta-correlated fields used in previous studies by including finite time correlations. Clustering in these fields can be compared to the traditional setting, through global measures and cluster detection algorithms. The enhanced velocity fields can be deformed using various interpolation methods. We can then investigate the sensitivity of clustering to the representation of temporal/spatial velocity structure to inform future studies of this phenomenon. We find coherency of time-correlated velocities leads to significantly faster rates of clustering, causing a larger number of longer lived/more populated clusters to form. Clustering is likely relevant to a host of biogeochemical processes of urgent interest, such as phytoplankton blooms and the ecological risk of microplastic pollutants. This work aims to establish an accurate basis for clustering simulations, to enable further exploration.

Plain Language Summary Floating material, such as microplastic waste and certain marine species, have a tendency to form dense patches on the ocean surface. Interest in this phenomenon is increasing, largely due to recent observations of the "garbage islands" of plastic pollution. We are attempting to develop greater understanding of the way these structures form by using simple models. These models give us the opportunity to assess the methodologies generally used in Lagrangian particle studies and how they impact numerical simulations of clustering floating material. They also allow us to try new methods of inter-comparison between models, which we have done to investigate the impact of common numerical approximations on the results of clustering simulations.

1. Introduction

1.1. Background

For our purposes, clustering refers to the tendency of floating tracers to aggregate, forming regions of enhanced concentration. Interest in this phenomenon is motivated by real life observations, primarily of pollutants and biological populations being advected by currents in the upper ocean (Cózar et al., 2014; Gower et al., 2006; Jordi et al., 2009). Additionally, clustering has been observed in ocean drifter data, in global accumulation zones (Maximenko et al., 2012), and on a smaller scale through dilation estimates (Rypina et al., 2022). Theoretical exploration intends to determine if this effect has meaningful implications for various ocean processes.

For example, buoyant microplastic pollution in the ocean are advected by currents, as are biological populations, such as phytoplankton. Since these pollutants represent a significant threat to said populations (Andrady, 2011), the potential for coextensive clustering could lead to greater risk by increasing reaction rates between them.

Passive tracers in an incompressible flow will not cluster due to the preservation of material volume. While non-divergent velocities can enhance existing tracer concentration gradients through mixing, they cannot increase the overall contrast. However, if a tracer is buoyant (such as microplastic pollution), its vertical motions are restricted and it will predominantly follow the horizontal flow. This leads to vertical motion in the ocean currents being experienced as surface convergence/divergence, which will cause the tracer to form clusters.

In the real ocean, potential sources of the weakly divergent flow include the ageostrophic or submesoscale velocities (Haza et al., 2016; Stepanov, Ryzhov, Berloff, & Koshel, 2020; Stepanov, Ryzhov, Zagumennov, et al., 2020). Observations of strong subduction at submesoscale fronts show the potential for strong surface convergence

(Archer et al., 2020). Other diverging contributions to the surface currents can be considered, such as Ekman currents or Stokes drift (Onink et al., 2019). Alternatively, considering particles with finite inertia allows for deviation of particle velocities from the background flow, which can aid in forming clusters (Baker et al., 2017; Liu et al., 2020; Monchaux et al., 2012). Whichever effects are considered, the important result is the presence of a nonzero potential component in tracer velocities.

Theoretical investigations of the clustering process have previously utilized kinematic velocity fields (Klyatskin, 2003; Koshel et al., 2019; Stepanov, Ryzhov, Berloff, & Koshel, 2020; Stepanov, Ryzhov, Zagumennov, et al., 2020; Wilson & Sawford, 1996). Often, Gaussian Random fields are generated with imposed spatial correlations, but delta-correlated in time. We adopt and extend this approach; moving away from delta-correlated fields by imposing a Langevin model in time. By retaining both models for time correlations, we study how the representation of time memory affects cluster formation. Furthermore, the flexibility of the kinematic framework provides the opportunity to explore sensitivity to time and spatial resolution of the velocity, as well as to several interpolation methods which are commonly used in more complex Lagrangian studies.

1.2. Statement of the Problem

Often in traditional Lagrangian studies, an offline advection is conducted. For reasons of computational/storage constraints or lack of available data, velocity fields may be time-averaged and interpolated, to the point where mesoscale and submesoscale flow features are no longer well resolved in time or space. Spatial resolutions of around 1/4° or lower are typical, as well as time resolutions of several days, with additional smoothing through time averaging (Blanke et al., 2002; Martinez et al., 2009; Onink et al., 2019; van Sebille et al., 2014). These simplifications are more than sufficient to preserve the structure of the transport pathways. However, there is no guarantee that the sensitivities of the exponential clustering process will be similar.

Spatial variations of submesoscale ocean currents are well constrained in both modeling and observational contexts (McWilliams, 2016; Rocha et al., 2016). Often, interest is focused on these currents as a function of the larger scale flow for the purposes of parametrization. As a result, their inherent time variability is often neglected. The limited understanding of time correlations has been addressed, and investigated in an observational setting by Callies et al. (2020), for a mooring array in the northeast Atlantic. Here, a well defined decorrelation time of $\sim$1 day was found at the horizontal scale of $\sim$10 km. Given that submesoscale currents will be a predominant source of surface convergence, sensitivity to time variations will necessitate resolving these decorrelation scales well. This is an achievable goal for future clustering simulations, but is also a much more stringent requirement than many other Lagrangian applications. In some contexts, choosing a typical decorrelation timescale may require a more comprehensive understanding of time variability than currently exists.

Using a kinematic model allows us to refine our methodology in a controlled setting. This is an essential initial step in creating a robust framework for further exploration of the clustering problem. In the immediate future, this will allow us to investigate, with confidence, the formation of clusters in dynamical velocity fields, reacting biological systems and when including inertial effects on particle velocity. Detecting a strong dependence of clustering on the representation of time correlations in this framework will indicate that this must be a consideration in more realistic clustering studies and can help motivate methodologies in future work.

Our velocity fields are constructed similarly to those in Koshel et al. (2019). Gaussian random fields can be constructed in Fourier space with imposed energy spectrum. This spectrum can be chosen to mimic the narrow band, isotropic structure from the classical picture of turbulence first proposed by Kolmogorov (1941). Kinematic fields such as these have often been used as a model for oceanic turbulence (Klyatskin, 2003; Koshel et al., 2019; Stepanov, Ryzhov, Berloff, & Koshel, 2020; Stepanov, Ryzhov, Zagumennov, et al., 2020; Wilson & Sawford, 1996). Traditionally, an assumption of delta-correlations in time is used to generate a time series of the velocity. We include the optional extension to allow for exponentially decaying temporal auto-correlations.

Despite the necessity of a potential component of the tracer velocity to cause clustering, the spatial and temporal properties of clusters also depend on the nondivergent component of the carrying flow. Our velocity fields therefore contain both solenoidal and weak potential components, which enhances the resemblance of the clustering process to that which occurs in the real ocean.

We consider floating tracers which follow the horizontal surface flow passively, so that the clustering observed is completely determined by the properties of the velocity field, and not dependent on any external particle properties. The equation for the concentration field takes a simple form along Lagrangian paths, allowing for exact integration. A Lagrangian approach means we resolve the concentration field precisely where there is clustering. The areas of enhanced concentration will be attractive sinks for Lagrangian paths, so will be sampled more often than the less dense voids. It also allows us to forgo the inclusion of unphysical damping (e.g., hyperviscosity), which is necessary in Eulerian models for draining out the forward cascade of enstrophy but may arbitrarily inhibit the clustering process (see Appendix A). The inclusion of weak compressibility leads to exponential clustering (Koshel et al., 2019), which is particularly unsuited to representation through Eulerian means, due to a persistent decrease in the dominant spatial scale of the concentration field.

Our framework also provides us the opportunity to investigate the effectiveness of interpolation methods for Lagrangian integration (see Appendix B). Our fields can be evaluated exactly, up to some modal truncation, at each particle position. This is a major advantage over dynamical studies, where the impacts of interpolation errors are rarely established.

Following Koshel et al. (2019), we can use Lagrangian concentration values, combined with statistical measures, to assess the global rate of clustering.

Cluster identification is a desirable tool for examining the local structure of a tracer concentration field. This requires a quantitative definition of the qualitative idea of a "cluster." In Baker et al. (2017) and Liu et al. (2020), methodologies to achieve this were developed and applied to the case of heavy inertial particles in dynamical turbulence. Voronoi diagrams were used to identify particles within clusters based on a simple statistical test. The resulting clusters can then be tracked over time and their statistics can be compiled. We have used these statistics for intercomparison between models, demonstrating the sensitivity of the clustering process to alterations in the temporal structure of the carrying flow. In particular, this approach provided a strong signal of "spurious clustering" when temporal correlations were underesolved.

The paper is organized as follows. In Section 2 we describe our kinematic flow model. In Section 3 we explain clustering properties in the focus. Section 4 contains the main results followed by summary and discussion in Section 5. Appendix A, B, and C can be found at the end of the paper.

2. Kinematic Flow Model

Here we establish the stochastic flow model used in our numerical experiments, both delta-correlated in time and with exponentially decaying auto-correlations. The model of tracer particles is explained, and description of the numerical implementation of each component of the model is also provided.

2.1. Velocity Fields

Our velocity fields are constructed in spectral space, with the imposition of homogeneity, isotropy and a chosen energy spectrum. The velocity is formed from a potential and solenoidal component, which are constructed separately:

$$\mathbf{u}(\mathbf{x},t) = \gamma \mathbf{u}_{\mathrm{p}}(\mathbf{x},t) + (1-\gamma)\mathbf{u}_{\mathrm{s}}(\mathbf{x},t), \tag{1}$$

$$\mathbf{u}_{\mathrm{p}} = (\partial_x \phi, \partial_y \phi),$$

$$\mathbf{u}_{\mathrm{s}} = (-\partial_y \psi, \partial_x \psi),$$

where $\gamma \in [0, 1]$ is a parameter controlling the relative strengths of these components. The functions $\phi(\mathbf{x}, t)$ and $\psi(\mathbf{x}, t)$ are the velocity potential and streamfunction of the potential and solenoidal components, respectively.

Similarly to Koshel et al. (2019), the spectral representation of both components is determined by four spectral coefficients: a_p, b_p, a_s and b_s.

$$\mathbf{u}_{\mathrm{s}}(\mathbf{x},t) = \Re\left\{ \int d^2\mathbf{k}\, [a_{\mathrm{s}}(\mathbf{k},t) + \mathrm{i}b_{\mathrm{s}}(\mathbf{k},t)]\, \frac{(k_y,-k_x)}{k} \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x}) \right\}, \tag{2}$$

$$\mathbf{u}_{\mathrm{p}}(\mathbf{x},t) = \Re\left\{\int d^2\mathbf{k}\left[a_{\mathrm{p}}(\mathbf{k},t) + \mathrm{i}b_{\mathrm{p}}(\mathbf{k},t)\right]\frac{(k_x,k_y)}{k}\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x})\right\}, \tag{3}$$

where $\mathbf{k}$, $k = |\mathbf{k}|$ is the 2-dimensional wavevector.

In Koshel et al. (2019), the coefficients a, b are taken to be delta-correlated (in time and spectral space) Gaussian random fields, which are independent of one another and all have zero mean:

$$\left\langle a_i(\mathbf{k},t)\, a_j\left(\mathbf{k}',t'\right)\right\rangle = t_0\,\sigma_{\mathbf{u}}^2 E(k)\,\delta_{ij}\,\delta\left(\mathbf{k}-\mathbf{k}'\right)\delta\left(t-t'\right), \qquad i,j\in\{\mathrm{s},\mathrm{p}\}, \tag{4}$$

$$\left\langle a_i(\mathbf{k},t)\, b_j\left(\mathbf{k}',t'\right)\right\rangle = 0, \tag{5}$$

where t_0 is a sampling rate when computing discrete realizations of the delta-correlated fields; and $\sigma_{\mathbf{u}}$ is a parameter controlling the standard deviation of the velocity. The function $E(k)$ is an idealized spectrum which we enforce to increase the similarity between our stochastic fields and turbulence in the upper ocean. The angled brackets $\langle\cdot\rangle$ represent the expected value of a quantity $\cdot$.

In this paper, we also allow for finite-time correlations by setting each wavevector coefficient to be a solution of a stochastic ODE forced by a delta-correlated, zero-mean Gaussian random field $\eta(\mathbf{k}, t)$:

$$\frac{d}{dt}a_i(\mathbf{k},t) = -\lambda\, a_i(\mathbf{k},t) + \eta_i(\mathbf{k},t), \tag{6}$$

$$\left\langle \eta_i(\mathbf{k},t)\eta_j\left(\mathbf{k}',t'\right)\right\rangle = \lambda\,\sigma_{\mathbf{u}}^2 E(k)\,\delta_{ij}\,\delta\left(\mathbf{k}-\mathbf{k}'\right)\delta\left(t-t'\right). \tag{7}$$

In this case, each coefficient at each point in spectral space is an independent Langevin process, exhibiting exponentially decaying memory. The parameter λ controls the memory timescale of the stochastic fields and is the exponential decay rate of temporal autocorrelations in the flow.

Both of these frameworks have been used and will be referred to as uncorrelated and correlated, respectively. These two cases allow for an evaluation of the exact impact of the inclusions of temporal correlations on clustering. For all velocity fields, the spectrum $E(k)$ is chosen to mimic the narrow band structure from the classical picture of Kolmogorov turbulence (Kolmogorov, 1941; Koshel et al., 2019):

$$E(k,l) = \frac{1}{2\pi}\frac{k^2l^4}{4}\exp\left(-\frac{k^2l^2}{2}\right). \tag{8}$$

This spectrum yields statistically stationary, spatially isotropic and homogeneous velocity fields, and has a single peak at the wavenumber $k = \sqrt{2}/l$, which sets the dominant scale of flow structures. An important benefit of using this spectrum is the analytical simplicity, which allows for the explicit calculation of many statistical properties.

2.2. Flow Statistics

All the stochastic fields in this paper are Gaussian, and we provide the point-wise mean and variances of these fields, to illustrate how the parameters we have chosen affect the behavior of the carrying velocity. Some illustrative plots are also included in Figure 1 for a specific set of parameter values, showing the statistics in detail.

Correlations are calculated as two-time auto-correlations, to elucidate the time-dependent memory of our two frameworks. If we were to consider two-point correlations, we would find that the length scales of correlations are entirely controlled by the parameter l, whereas the one-point statistics are independent of it.

Let us start with uncorrelated fields:

$$\langle\mathbf{u}(\mathbf{x},t)\rangle = \mathbf{0}, \tag{9}$$

$$\left\langle u_i(\mathbf{x},t)u_j\left(\mathbf{x},t'\right)\right\rangle = \frac{1}{2}\left[\gamma^2+(1-\gamma)^2\right]\sigma_{\mathbf{u}}^2\, t_0\,\delta_{ij}\delta\left(t-t'\right), \qquad i,j\in\{x,y\}. \tag{10}$$

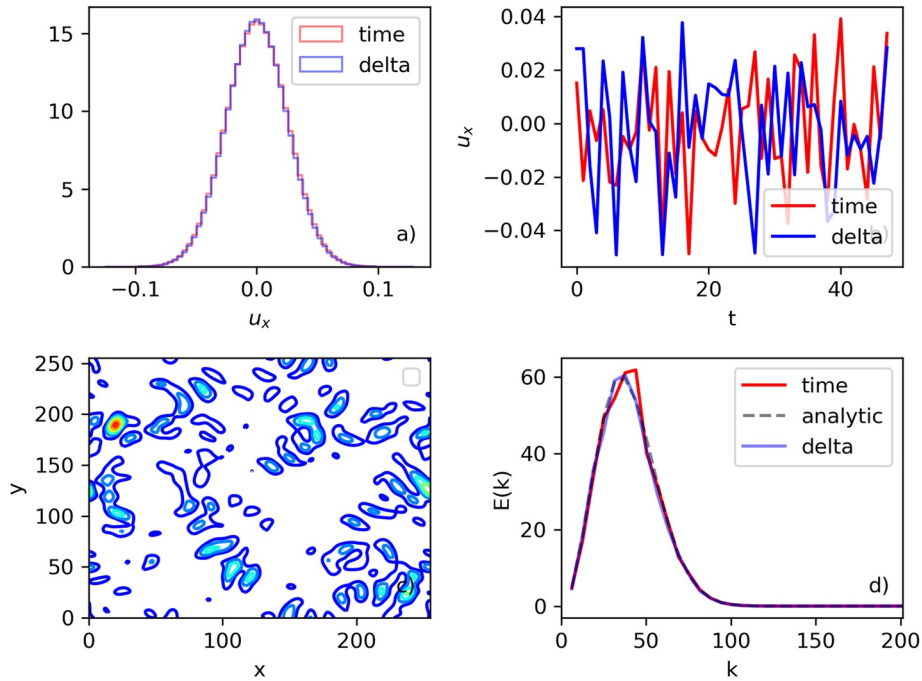

Figure 1. Plot (a) shows the similarity in velocity distributions between the time correlated and delta-correlated fields, whilst (b) shows two timeseries of the field realizations. (c) shows a typical distribution of energy density for the velocity, whilst (d) shows the imposed power spectrum, which is identical in both cases.

In practice, if we are sampling the uncorrelated field at a rate t_0, this means the instantaneous flow field will have velocity variance $\left[\gamma^2 + (1-\gamma)^2\right]\sigma_{\mathbf{u}}^2$. From this we can define an advective timescale for these flows

$$t_{\text{adv}} = \frac{l}{\left[\gamma^2 + (1-\gamma)^2\right]^{1/2}\sigma_{\mathbf{u}}}. \tag{11}$$

For the correlated fields, if we set $t < t'$ (without loss of generality), then:

$$\langle \mathbf{u}(\mathbf{x}, t)\rangle = \mathbf{0}, \tag{12}$$

$$\left\langle u_i(\mathbf{x}, t) u_j\left(\mathbf{x}, t'\right)\right\rangle = \frac{1}{2}\left[\gamma^2 + (1-\gamma)^2\right]\sigma_{\mathbf{u}}^2\,\delta_{ij}\,e^{-\lambda\left(t-t'\right)}\left(1 - e^{-2\lambda t}\right), \tag{13}$$

where the temporally correlated fields are initialized at rest. We release the tracer after the velocity field has equilibrated statistically, so that the instantaneous velocity field achieved variance $\left[\gamma^2 + (1-\gamma)^2\right]\sigma_{\mathbf{u}}^2$, as in the uncorrelated case. The advection timescale Equation 11 is therefore also identical to the uncorrelated case. Kinetic energy is an important characteristic of a carrying velocity field, and it's density, $\mathcal{E}$, can be defined in terms of the velocity $\mathbf{u} = (u, v)$ to characterize the instantaneous structure of our flow.

$$\mathcal{E} = \frac{1}{2}\left(u^2 + v^2\right), \tag{14}$$

In Figure 1, this quantities are shown for a snapshot of a typical realisation of a temporally correlated field. We can see clearly here the structure of these fields as a collection of small eddy-like flow features, analogous to a typical submesoscale flow from simulation or observation (McWilliams, 2016). For illustrative purposes, we also show the distribution of velocity values at a single point in the flow, a time series of those velocity values and the power spectra which are verified to match our desired spectrum $E(k)$.

The above fields are restricted to a doubly periodic domain, so they can be calculated in terms of their Fourier modes. Considering the first 512×512 modes (unless stated otherwise), the velocity is evaluated on a grid of the same resolution on $(0, 1) \times (0, 1)$. To solve Equation 6, a first-order Euler-Maruyama scheme is implemented.

We can further alter these fields by varying the number of modes used in the evaluation, the type of spatial interpolation method used, or by sampling the time-correlated velocity field at intervals greater than the decorrelation time and time interpolating. In Lagrangian studies, offline velocity fields are commonly utilised hence the interval between samples is generally larger than many fast ocean processes.

2.3. Tracers

In order to investigate clustering, the simplest form of tracer particles—purely passive—is introduced to the statistically stationary velocity fields defined above. For each particle, we solve:

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{u}(\mathbf{x}(t), t). \tag{15}$$

Additionally, to find the concentration at each particle location, we solve the equation:

$$\frac{d}{dt}C(t) = -C(t)\nabla \cdot \mathbf{u}(\mathbf{x}(t), t). \tag{16}$$

The form of this equation comes from expressing the Eulerian equation of tracer mass conservation in the frame of the Lagrangian particle:

$$\frac{\partial C}{\partial t} + \nabla \cdot (C\mathbf{u}) = 0. \tag{17}$$

Taking the discrete form of the velocity fields from Section 2.1, a standard fourth-order Runge-Kutta scheme is used to integrate the position of each particle forward in time. For each particle, we use linear spatial interpolation to find the velocity and divergence values at their location. This was chosen from a variety of interpolation methods (see Appendix B), as a compromise between accuracy and computation time.

Using gridded fields with interpolation allowed for the use of Fast Fourier Transforms and restricting ourselves to a limited resolution meant there was a smaller cost to introducing more tracer particles. This was important in resolving the detail of clusters, described in Section 3.2. The grid scale for the velocity was deliberately chosen so that the single peak of the energy spectrum Equation 8 at scale l was well resolved.

In these model runs, 25×10^4 particles were initialized on a square grid. A uniform starting distribution guarantees that all clustering observed is due to the action of our velocity field, since there are no initial concentration gradients.

3. Clustering Properties

The discussion of the dynamics of clustering is split into two sections. First we consider global measures of clustering which demonstrate the overall action of the divergent flow to cause increases in concentration of tracer. For these purposes, we make use of Lagrangian concentration values. Largely, we follow Koshel et al. (2019) and Klyatskin (2003) in defining and calculating these global measures. Second, we consider individual clusters of accumulated particles and how these evolve over time, utilizing a cluster detection algorithm. The methodology of which is largely based on Baker et al. (2017) and Liu et al. (2020).

3.1. Global Clustering Properties

The specific cluster mass and area allow us to keep track of how much of the material has clustered above a certain density and what proportion of the domain these clustered particles occupy. Here,

$$s\left(t, \overline{\rho}\right) = \frac{\int_D d^2\mathbf{x}_0\, \Theta\left(\rho(\mathbf{x}, t; \mathbf{x}_0) - \overline{\rho}\right)}{\int_D d^2\mathbf{x}_0 \rho(\mathbf{x}, t; \mathbf{x}_0)} \tag{18}$$

is the specific cluster area, and

$$m\left(t,\overline{\rho}\right)=\frac{\int_D d^2\mathbf{x}_0\,\rho(\mathbf{x},t;\mathbf{x}_0)\Theta\left(\rho(\mathbf{x},t;\mathbf{x}_0)-\overline{\rho}\right)}{\int_D d^2\mathbf{x}_0\rho(\mathbf{x},t;\mathbf{x}_0)} \tag{19}$$

is the specific cluster mass, where Θ is the Heaviside step function and $\overline{\rho}$ is some density threshold. We denote our domain ($[0,1]\times[0,1]$) as $\mathcal{D}$.

In the case of purely divergent ($\gamma=1$), delta-correlated in time velocity fields, an asymptotic theory for the specific cluster area and mass has been derived (Klyatskin, 2003), namely:

$$\left\langle s\left(t,\overline{\rho}\right)\right\rangle \approx\sqrt{\frac{\rho_0}{\pi\overline{\rho}D_p t}}\,e^{-D_p t/4}, \tag{20}$$

$$\left\langle m\left(t,\overline{\rho}\right)\right\rangle \approx 1-\sqrt{\frac{\overline{\rho}}{\pi\rho_0 D_p t}}\,e^{-D_p t/4}, \tag{21}$$

where ρ_0 is the initial uniform concentration of tracer ($\rho_0=1$ in all our simulations), and D_p is a diffusion parameter which is defined below. We expect from the form of these equations that $\left\langle s\left(t,\overline{\rho}\right)\right\rangle\to 0$ and $\left\langle m\left(t,\overline{\rho}\right)\right\rangle\to 1$ as $t\to\infty$, that is, eventually all the material is clustered in an infinitesimally small area. We also gain an expression for the diffusion parameter, D_p:

$$D_p=\int_0^\infty d\eta\,\langle(\nabla\cdot\mathbf{u}(\mathbf{x},t))(\nabla\cdot\mathbf{u}(\mathbf{x},t+\eta))\rangle. \tag{22}$$

The validity of the above asymptotics for the case where $\gamma\neq 1$ is demonstrated in Koshel et al. (2019) and the limiting behavior is found even for relatively small values of γ. By evaluating Equation 22 for both types of stochastic velocity field, it is found that:

$$D_p=\gamma^2\frac{\sigma_{\mathbf{u}}^2 t_0}{l^2} \tag{23}$$

for delta-correlated velocity fields, and

$$D_p=\gamma^2\frac{\sigma_{\mathbf{u}}^2}{l^2\lambda} \tag{24}$$

for temporally correlated velocity fields (see Appendix C for the details of the calculation). This motivates the understanding that, all other parameters being identical, the choice $t_0=1/\lambda$ actually represents the same flow, under a different approximate representation. In both frameworks, these timescales are a characteristic timescale of decorrelation for the flow and so this is an intuitive result.

From this diffusion parameter, we identify a diffusion timescale which is naturally related to the clustering process, $\tau=1/D_p$.

In our own results, we found that the cluster mass and areas (shown in Figure 4) could both be fit to simple exponential curves with great accuracy.

$$m(t,1)\approx 1-A\exp(-R_m t) \tag{25}$$

$$s(t,1)\approx B\exp(-R_s t) \tag{26}$$

From these curves we could then extract clustering rates R_m and R_s. Although the two rates could differ for s and m, it was found that the ratio between model runs was preserved. In Table 2, we quote R_m extracted from this fit.

3.2. Local Clustering Properties

The Voronoi diagram allows for an assessment of clustering of tracer particles (Baker et al., 2017). It is defined as a tesselation of the domain in which each cell corresponds to a unique particle in the set and all the interior points in the cell are closer to that particle than any other in the set. Generally, the smaller the area of the Voronoi cell of a particle is, the closer it is to its neighbors. Hence this is a measure of how much a particle has clustered.

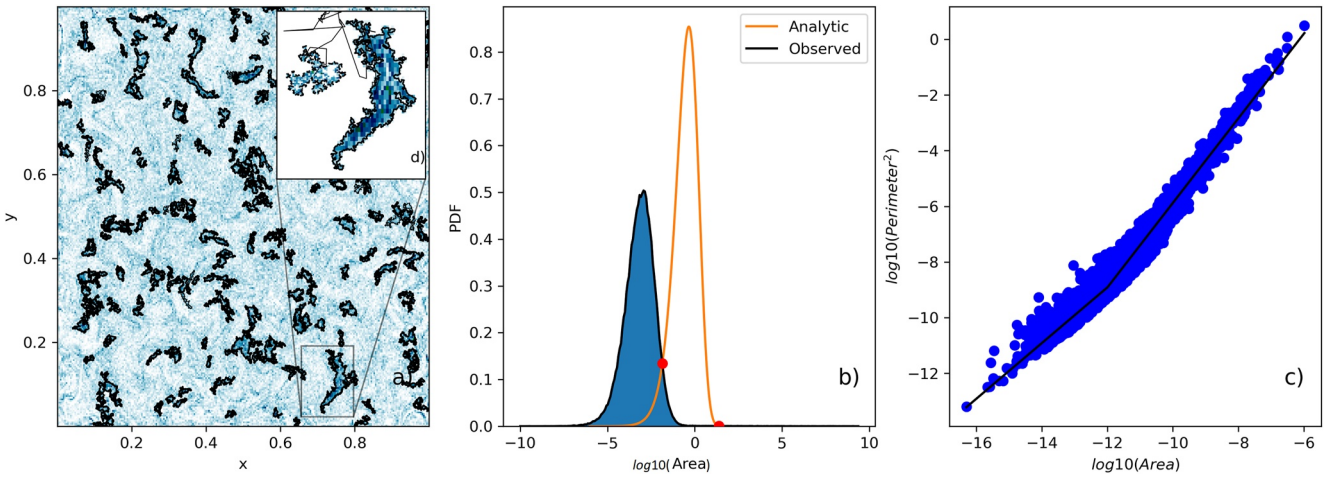

Figure 2. (a) Contains the boundaries of all the clusters detected at a single time step in a realization from the ensemble model B. To demonstrate the enhanced concentration inside these boundaries, an approximate concentration was found by a particle bin method. (b) Shows how we determine which particles are "clustered." In orange, we have the theoretical distribution of Voronoi cell areas for a collection of random and independently placed particles, in blue (with black boundary) is the observed distribution of cell areas from our clustering simulation. The first crossing point of these distributions is used as a cutoff tell if a particle belongs to the "clustered" tail of the observed distribution. (c) Illustrates how we separate out the "fractal"/"significant" clusters. We find where the relationship changes between area A and perimeter L from quadratic to sub-quadratic ($A \sim L^k$, $k < 2$) using a broken power-law fit to the ensemble of clusters at a given timestep. Any cluster in the sub-quadratic regime is considered a "significant" cluster and these are precisely the ones we track over time. (d) Is a magnified image of the largest cluster from (a). We also include the cluster as it was 40 timesteps previously (this is the smaller object), to show a typical evolution (the thin line is the trajectory of the cluster).

To determine what area counts as "small," we compare the distribution of normalized Voronoi cell areas in the advected tracer particle set with that which would be observed if particles were randomly, uniformly and independently positioned. Such Voronoi cell areas were found to be approximately gamma distributed by Ferenc and Néda (2007). We identify the clustered particles as those to the left of the crossing point between the two distributions, that is, those with the smallest Voronoi cell areas (see Figure 2b; for an example). The particles in this tail have clustered in a way that cannot be explained by some random effect, so we can say they have done so due to the action inherent in the model. The cells corresponding to these particles can then be combined, if adjacent, to form larger connected regions of clustered particles.

Following the method of Baker et al. (2017), we filter the set of connected regions to single out significantly large clusters. This is an important step because focusing on a subset of clusters makes tracking them over time more tractable. This recognizes the fact that clusters formed from a handful of particles are unlikely to be of physical interest. They contain a negligible amount of the total mass of tracer. They are also often spontaneous and short lived. The filter stipulates that we only retain clusters with a total area larger than a threshold. This threshold is determined using the perimeter/area scaling relationship of the ensemble of connected regions. Smaller clusters tend to be simple geometric shapes, hence we see $L^2 \sim A$, where L and A are the perimeter and area of a cluster respectively. For larger clusters, it is instead observed that $L^2 \sim A^k$, $k > 1$. This is because they have more complex "fractal" boundaries, as a result of the greater degrees of freedom in their boundary, as well as the chaotic filamentation undergone at the peak energy scale of the velocity. Using a broken power law fit to the L/A curve, we can determine the cluster size at which this regime change occurs, and that becomes the threshold for our filter (see an example in Figure 2c).

The conditions of the algorithm are generally applicable to any ensemble of particles, and also allow for tuning to a specific instance. The important point is that the algorithm consistently identifies regions that would naturally be considered clusters of material in a qualitative sense, whilst also providing a quantitative definition of what such a cluster is.

Our interest is also focused on persistent clusters which survive over time. We track clusters over time using a slight modification of the method employed by Liu et al. (2020). We take the instantaneous clusters from two adjacent timesteps (found using the algorithm described above) and sort them in descending size order. The clustering algorithm is applied every 100 simulation timesteps, so every $10t_0$. Then, we sequentially match up each cluster with the one from the subsequent timestep with the largest overlap in the particles it contains. Once a

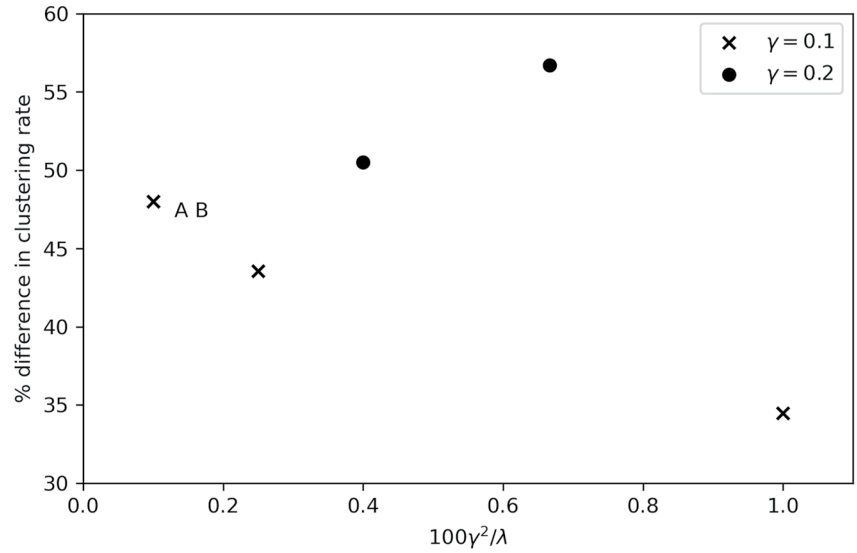

Figure 3. Percentage difference in the clustering rate between correlated and uncorrelated model runs for different values of γ and λ. $\sigma_u = 0.04$, $l = 0.04$. The percentage difference in clustering rate for the pair of models AB is labeled. The x-axis, $100\gamma^2/\lambda$, is motivated by the form of the asymptotic diffusion timescale D_p Equation 24. Correlated model runs had a faster clustering rate than their uncorrelated analog in every comparison.

cluster from the later timestep is matched with a, it cannot be matched with another cluster. Clusters at the initial timestep which cannot be paired are considered to have "died." This allows us to define cluster lifetimes based on when they are first and last detected. This method ensures:

1. Larger clusters survive with priority. For instance, if two clusters merge, we identify the result with the largest of its parent clusters.
2. We never discard a potential match between clusters. In previous research (e.g., Liu et al., 2020), clusters at adjacent timesteps were only linked if they shared a large majority of their constituent particles. This means we tend to observe clusters for longer times, since our definition of cluster survival is more relaxed. This is reasonable for our Lagrangian data because of the short timestep between applications of the algorithm.

4. Main Results

Use of the global cluster statistics verified that there was a significant deviation between clustering rates in comparable correlated and uncorrelated fields. For a variety of λ and two different values of γ, it was always found that clustering was faster in correlated fields (see Figure 3). With this established, a pair of models were selected for further exploration through use of the cluster detection algorithm. These models also acted as a basis for comparison of interpolation error, by computing additional runs with various interpolation schemes imposed (See Table 1 for a description of each model referenced in this paper).

Table 1
A Description of Each Model Referenced in This Paper

Model	Description
A	Uncorrelated in time. 15 run ensemble. 25×10^4 particles. 512×512 Fourier modes
B	As above but correlated in time.
T1	Correlated in time. Linear interpolation in time with sampling time $1.5t_0$
T2	As above, with sampling time $3.5t_0$
T3	As above, with sampling time $7.5t_0$
L1	Correlated in time. Low resolution for velocity field (64×64 Fourier modes).
L2	Same as above, with 32×32 Fourier modes (Correlation scale l poorly resolved).

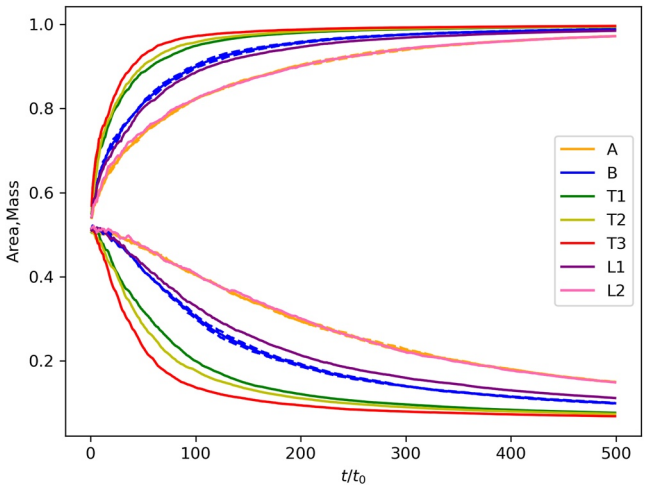

Figure 4. Cluster mass (increasing curves) and area (decreasing curves) for every model in table. For ensembles A and B, one standard deviation above/below the mean is shaded, showing the negligible ensemble variance.

4.1. Model Run Parameters

Our parameters are chosen to replicate the conditions where a delta-correlated velocity field would traditionally be considered appropriate (i.e., $t_{mem} \ll t_{adv}$). We also are interested in a velocity field with a weak potential component, to mimic the magnitude of surface divergences in the real ocean which are generally on the order of the Rossby number.

$$\sigma_{\mathbf{u}} = 0.04,\ l = 0.04,\ t_0 = 0.1,\ \gamma = 0.1 \tag{27}$$

With these chosen, we can now construct analogous velocity fields which are either correlated or uncorrelated in time. An ensemble of 15 model runs of each were completed to achieve good statistics.

Results from our investigation of interpolation methods (Appendix B) suggest that linear spatial interpolation to evaluate gridded velocity values at particle positions was sufficient to preserve accuracy in both Lagrangian paths and Lagrangian concentration values. This could in part be related to the simplicity of our idealized velocity field.

Each model was run for several diffusion timescales (as determined by Equations 23 and 24), to ensure that the particles achieved a well clustered state. This was verified by the cluster mass and area statistics, shown in Figure 4, as they tend to 1 and 0 respectively. In every case, we used a timestep of $dt = 0.01$ so that all timescales were resolved. For the time interpolated simulations, the velocity model was updated using this timestep, but was only communicated to the particle model once every time sample.

4.2. Model Results

4.2.1. A and B

Model runs A (uncorrelated) and B (correlated) are intended to supply a baseline for our clustering models. To this end we chose to maximize our grid resolution and number of particles subject to computational constraints. An ensemble of runs was collected, until there was confidence that the mean cluster mass and area curves were sufficiently smooth, and the distribution of cluster statistics (shown in Figure 5) was dense with observations.

In both global and local measures, clustering had a significantly different character in each field type. In the global measures (Figure 4), the rate of clustering was much faster, to the extent that there was no overlap between the ensembles. The clusters detected also had longer lifetimes on average in model run B by 13% and were also 14% larger, in terms of the number of particles contained.

The enhanced clustering in model run B can be explained by an increased number of extremely large, long lived and dense clusters. These have skewed the average cluster statistics, despite making up a small proportion of the total number of clusters.

If we integrate Equation 16 along Lagrangian paths, we find that the concentration of each particle satisfies:

$$C(\mathbf{x}(t;\mathbf{x}_0),t) = \exp\left(-\int_0^t d\tau\, \nabla\cdot\mathbf{u}(\mathbf{x}(\tau;\mathbf{x}_0),\tau)\right), \tag{28}$$

or:

$$C(t_2) = C(t_1)\exp\left(-\int_{t_1}^{t_2} d\tau\, [\nabla\cdot\mathbf{u}](\tau)\right), \tag{29}$$

where everything is being evaluated along Lagrangian paths. This shows that a particle being clustered between times t_1 and t_2 is equivalent to:

$$\int_{t_1}^{t_2} d\tau\, \nabla\cdot\mathbf{u}(\tau) < 0. \tag{30}$$

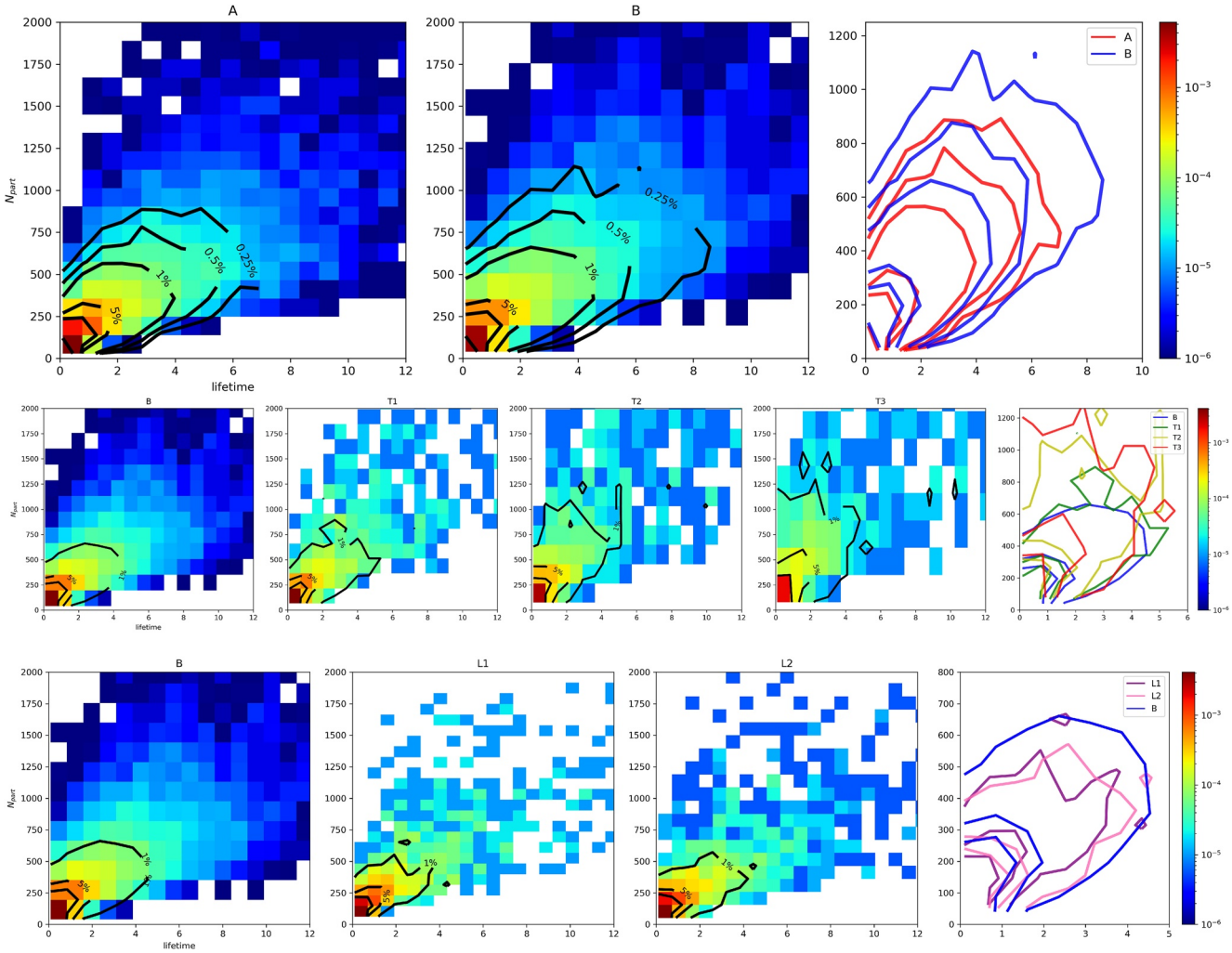

Figure 5. Histograms of clusters found for all model runs. The number of particles, N_{part}, is the maximum amount of tracer particles contained in a cluster over its lifetime. Lifetime is the time elapsed between the first and last detection of a cluster.

If we look at the timeseries of the Lagrangian divergence taken at a sampling rate t_0, we can compute a simple autocorrelation at a lag time of k decorrelation timescales along a Lagrangian path:

$$A_k = \frac{\sum_{i=0}^{N-k} (f_i - \bar{f})(f_{i+k} - \bar{f})}{\sum_{i=0}^{N} (f_i - \bar{f})(f_i - \bar{f})}, \tag{31}$$

where $f_n = \nabla \cdot \mathbf{u}(\mathbf{x}(t_n; \mathbf{x}_0), t_n)$, $t_n = nt_0$. Figure 6 shows the mean autocorrelation averaged over all 25×10^4 Lagrangian paths at lag times up to $15t_0$. It can be seen that non-zero autocorrelation persists at much longer range for the time-correlated simulation, even over several decorrelation times. It is also observed that divergences skew negative at all times along Lagrangian paths, and this is a direct result of the attractivity of convergent zones to the passive tracers. Combining these two facts, the enhanced clustering can be explained by the increased persistence of negative divergence along Lagrangian paths, which can only exist when there is non-trivial time history. This is precisely the effect which is neglected when a delta-correlated velocity field is employed.

4.2.2. T 1, 2, and 3

Time interpolation is common in Lagrangian studies, particularly when observed currents are used. Also, when simulated fields are outputted at regular time intervals, time interpolation is often favored over greater resolution due to storage capacity concerns. To investigate the impact of poor time resolution on the clustering process, we used linear time interpolation, with increasingly severe time windows. These were: $1.5t_0$, $3.5t_0$ and $7.5t_0$. In each

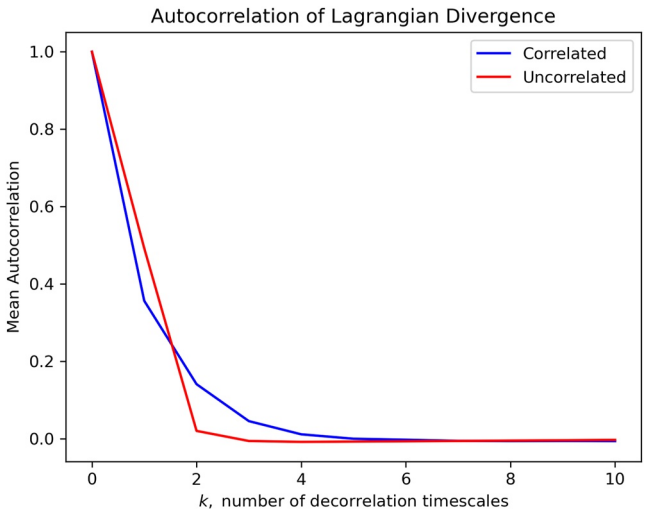

Figure 6. Shows the decay of divergence autocorrelation along Lagrangian paths, averaged over all tracer particles, for lag times of several decorrelation timescales, t_0. Correlated velocities have more persistent autocorrelations.

case, the exponential autocorrelations are not being properly resolved, and this is reflected in our clustering measures.

In Figure 4, there is a continuous increase in the speed of clustering as the time window is increased. Given the evidence from models A and B, that stronger time correlations increase the rate of clustering, it could be that linear interpolation is amplifying time correlations. A linear decrease in autocorrelation will beat an exponential decay and this causes spurious clustering. Furthermore, we see evidence of completely different statistics for significant clusters.

This can be demonstrated using cluster densities, defined as the ratio of the number of particles in a cluster to the area at largest extent. In the density distributions in Figure 7, it is evident that as the time sampling window increases, much denser objects are observed, and the ability to resolve the sharply peaked distribution is completely lost.

Spurious clustering is a big concern for studies of this type going forward, and we argue this is a natural result of looking in the "tail end" of distributions. By focusing on the smallest, densest regions of our tracer concentration, we amplify the effect of misrepresenting the structure of our flow field. A poorly chosen spatial or time resolution could lead to unphysical clustering. This is especially relevant to Lagrangian studies in observed currents, where both resolutions may be insufficient.

4.2.3. L 1 and 2

Since our narrow band energy spectrum will peak at the 1/l ∼ 25th wavenumber, we expect that the velocity field is well resolved using both 512 × 512 Fourier modes and 64 × 64 modes (model L1). Therefore we expect a marginal reduction in accuracy from model B to L1. Model run L2 was chosen specifically to poorly resolve the peak of the spectrum.

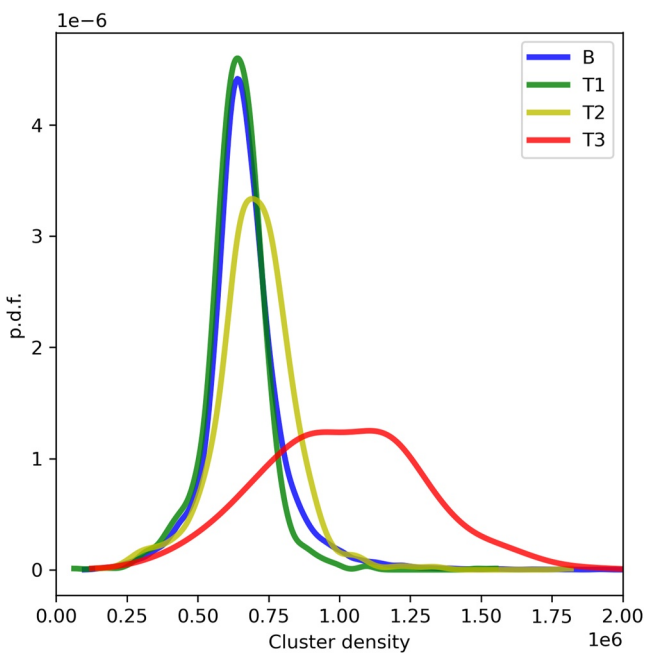

Figure 7. P.D.F.s of cluster density for B (correlated) versus correlated with time interpolated velocity. These were computed using kernel density estimation.

Somewhat surprisingly, we can argue that model L1 preserves most of the details of the clustering process. The cluster mass and area lie just outside the ensemble range from model run B, closer than any other model run included in this paper.

For model L2, there is a more obvious degradation of our results that is most apparent in the cluster mass and area (shown in Figure 4). These show evidence of a much slower process with the clustering rate approximately halved (see Table 2). However, we still see relatively good agreement in the cluster statistics (Figure 5 and Table 2).

A possible explanation of the quality of our low resolution models is the use of interpolation. For a simple idealized velocity field with a single peaked spectrum, it is possible that linear interpolation is reproducing the smaller scales very accurately. Furthermore, this accurate reconstruction is not guaranteed in more complex fields, which limits the extent to which we can interpret this result.

5. Summary and Discussion

We have investigated the effect of time correlations in carrying velocity fields on the clustering of passive tracers, employing recently developed cluster detection algorithms as well as other measures of the cluster formation rate.

Interest in this process is rapidly increasing due to several outstanding problems in ecology and biogeochemistry. Hence, the ability to accurately describe cluster formation is desirable. It is important to establish a firm

Table 2
Some Results From Applying the Cluster Identification Algorithm and Global Clustering Measures to Each Model Run

Model	Average lifetime	Longest lifetime	Average N_{part}	Clustering rate
A	1.39	22.0	268	0.26
B	1.60	24.2	312	0.47
T1	2.20	20.9	464	0.97
T2	2.70	21.0	657	1.04
T3	3.19	23.2	838	1.49
L1	1.57	17.0	301	0.49
L2	1.59	20.0	290	0.28

Note. The number of particles is taken from when each cluster contained the largest number of particles.

baseline for clustering simulations, with a thorough understanding of the appropriate sensitivities. With this motivation in mind, we have utilized kinematic velocity fields where we have fine control over their statistical characteristics. The dynamics of passive particles in these fields provide us with a simple, idealized model of clustering material in the upper ocean. This model has been investigated to improve confidence in future studies of this phenomenon, as well as to provide its own insight to the important factors influencing clustering.

The inclusion of time history was found to amplify the clustering process. Even though time correlations were short lived and a delta-correlated approximation would be considered appropriate, the model statistics were significantly altered. Globally, the rate of increase in concentration was higher. In terms of the individual clusters, increased coherency was detected. Larger, longer lived clusters were more prominent in the time correlated simulations. These can be explained as a result of the additional history of the divergence along the Lagrangian paths, which leads to a preference for clustered particles to further cluster.

Motivated by the standard methodology of Lagrangian studies, we have further investigated the impact of low resolution data, temporal and spatial interpolation on the clustering process.

Using the toolkit of methods developed, we have observed the dependence of clustering on the representation of time correlations. This can guide us in more complex models. Strong signals of spurious clustering were detected when using standard time interpolation techniques. The clusters that formed in error were large, but short lived; a potential candidate type to identify poor resolution in dynamical studies. Additionally, the Lagrangian concentration values are more sensitive to interpolation error than the traditional Lagrangian position variables, likely due to the exponential nature of cluster formation. Motivated by this, we propose the adoption of online velocity fields for use wherever possible, so that all time scales required to integrate the dynamics are also present in the Lagrangian simulation. Observational evidence suggests a dominant decorrelation timescale of approximately 1 day in the open ocean for partially divergent submesoscale currents. Given the sensitivity to temporal correlations we have detected, a half-day timestep for realistic clustering simulations appears to be a minimum requirement. However, in many applications, such as submesoscale resolving models and larger scale observations, the time characteristics of divergent surface currents are still largely unconstrained. As a result, robust investigation of this aspect of carrying flows will be an integral component of future clustering studies. Clustering rate calculations have proved to be an efficient and effective method of detecting spurious clustering. In a general setting, finding a time-interpolation window which doesn't cause large changes in clustering rate when altered is a good starting point for robust clustering simulations.

Undersampling the velocity field in space led to significantly slower clustering, and a loss of detected clusters similar to the delta correlated simulations. While less severe than the time interpolation errors, they still demonstrate the importance in the dynamics of spatio-temporal structure at the tail end of the concentration distribution.

With these methodologies and observations in mind, we can further explore clustering in various settings. The inclusion of time correlations will allow these kinematic fields to be used in the Maxey-Riley equations, which require the carrying velocity field to have a well defined time derivative. Furthermore, we can use the method of integrating the Lagrangian concentration equation to investigate the dynamics of simple interacting particle models, such as population models for plankton and pollutant, by adding source and sink terms to the right hand side. Expansion of our methodology to dynamical carrying velocity fields, such as Quasigeostrophic simulations, can also be considered to investigate a more physically constrained clustering process.

Appendix A: Lagrangian Preference Over Eulerian for Clustering Models

Due to clustering, the long-time locations of Lagrangian particles will be biased toward areas of higher concentration. When analyzing the ensemble of particles, this leads to greater resolution precisely in our areas of interest. To achieve similar resolution with an Eulerian method we must also increase resolution in the voids.

When particles are advected with intrinsic concentration, we must integrate $3N$ ODEs, where N is the number of particles. For an equivalent Eulerian simulation we will have as many equations as we have grid points, N_{grid}. In each case, we can choose the difficulty of the computations by varying these numbers. This is where the inherent bias becomes important; if we can choose $N < N_{\text{grid}}/3$ while retaining the resolution of the important clustered regions, then using Lagrangian methods provides a simple and clear advantage.

The motivations do not stop there. Lagrangian methods also allow for easier generalization to more complex forms of floating particle (e.g., particles with inertial effects), and are more suitable when we have lower resolution velocity data, through the use of spatial interpolation methods (see Appendix B).

On the Eulerian grid, our concentration field satisfies the P.D.E. Equation 17, with one major modification. For any basic Eulerian solver, we must introduce a diffusion or hyperdiffusion to regularize our solution. This would be true even in a non-divergent flow, but the introduction of the weakly potential flow makes it even more essential, since the presence of clustering creates an inevitable cascade to smaller and smaller scales (which becomes catastrophic below the grid scale). Lagrangian methods allow us to forgo this inclusion in a consistent, stable way.

$$\frac{\partial C}{\partial t} + \nabla \cdot (\mathbf{u}C) = \nu \nabla^2 C \tag{A1}$$

$$\frac{\partial C}{\partial t} + \nabla \cdot (\mathbf{u}C) = \mu \nabla^4 C \tag{A2}$$

In many applications, the introduction of a diffusion leads to a minor modification of the bulk behavior. This makes it an ideal way to regularize a fluid problem. However, we find that at it is precisely the areas of higher concentration that are most sensitive to the form and magnitude of the added diffusion, which is undesirable for our purposes. In Figure A1, we show the impact of said diffusivity on our assessment of clustering by solving Equation A1 for two appropriate values of ν which both stabilize our solution without significantly affecting the qualitative behavior. The 75th and 90th percentiles of the concentration field are much lower when we have higher diffusivity and there is also a significant effect on the shape and location of the high concentration regions (shown by the regions containing the 90th percentile and above concentration values). These extreme values of concentration are analogous to the clusters we have been detecting in our Lagrangian analysis. Given that the chosen parametrizations are entirely motivated by numerical stability after the imposition of a grid, we find this to be a significant weakness of the traditional Eulerian methods in clustering studies.

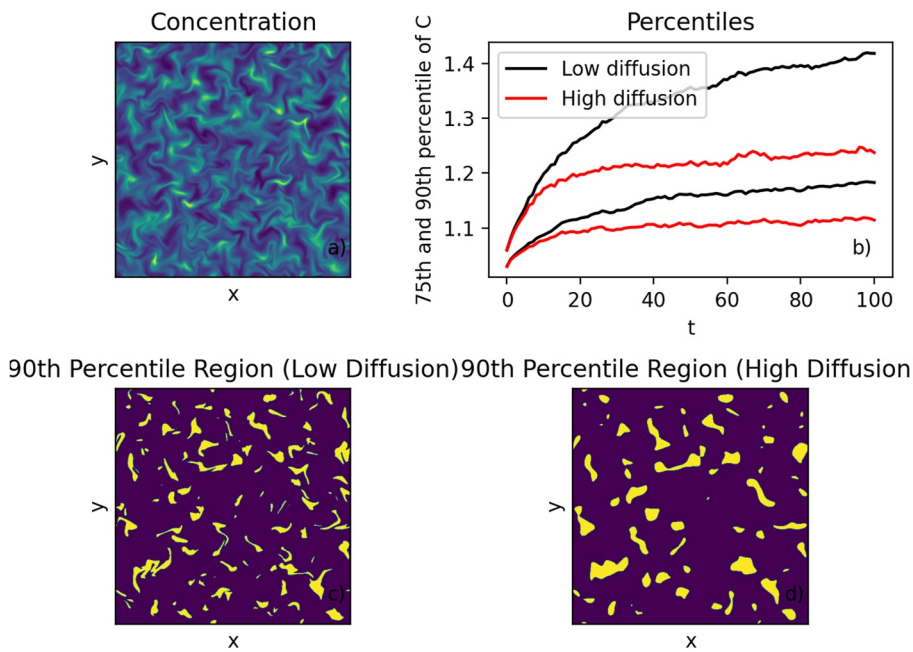

Figure A1. Plot (a) shows a typical solution from the Eulerian solver. (b) Shows the 75th and 90th percentile values of the concentration field for two values of hyperdiffusivity. (c) and (d) Shows how diffusivity alters the shape and character of clusters (the regions of the 90th percentile values).

Appendix B: Comparison of Interpolation Methods

When solving particle tracking problems, interpolation methods need to be implemented to evaluate field values at each particle position. Quantities are usually known on some subset of points, for example, on a regular grid. Hence, in most applications, we cannot investigate the impact of interpolation on our Lagrangian particles, since there is no exact expression for the particle velocity.

Our kinematic fields differ in this respect, since using their spectral form, we can evaluate the velocity field and its divergence at any point in the domain, up to some modal truncation. As a result, we can compare the behavior of our particles with each interpolation method directly to the exact case. This allows us to choose the right interpolation method for our purposes, and also demonstrates the quality of each method.

We used a nearest-neighbor, a bi-linear and a bi-cubic interpolation scheme to evaluate both our velocity fields and divergences, in order to solve Equations 15 and 16. We solved the equations for 250 particles initialized in a grid. We then assess the error in the particle locations and their concentration using the following measures:

$$\Delta x_{\text{interp}}(t) = \left\langle \left| \mathbf{x}_{\text{interp}}(t) - \mathbf{x}_{\text{exact}}(t) \right| \right\rangle, \tag{B1}$$

$$\%C_{\text{interp}}(t) = \left\langle \left| \frac{C_{\text{interp}}(t) - C_{\text{exact}}(t)}{C_{\text{exact}}(t)} \right| \right\rangle, \tag{B2}$$

where $\langle \cdot \rangle$ denotes averaging over all the particles.

Somewhat unsurprisingly, we found cubic performed better than linear, and linear performed better than nearest neighbor interpolation. The degree of improvement is incredibly significant (see Figure B1), with two orders of magnitude less error with each successively better method. This is most pronounced in the concentration values, where nearest neighbor interpolation leads to average errors as large as 20% over just one diffusion timescale, when linear and cubic methods produce insignificant error.

This result has implications for future use of the Lagrangian concentration integration method, which has shown promise in previous investigations of clustering (Koshel et al., 2019), but is evidently sensitive to our representation of the velocity field. This is of special concern when using fields from dynamical models, since the spatial resolution is often low for clustering applications.

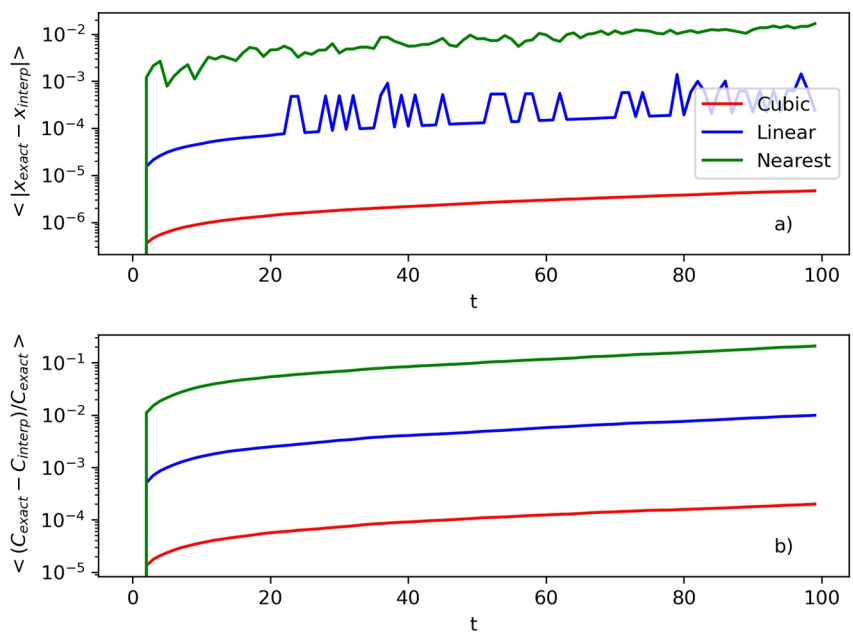

Figure B1. Tracking the interpolation error through time. All parameters were chosen to be identical to models A and B. Plot (a) shows the average error in the Lagrangian paths, and plot (b) shows the proportion $\left\langle \frac{C_{\text{interp}} - C_{\text{exact}}}{C_{\text{exact}}} \right\rangle$.

Appendix C: Derivation of the Diffusion Timescales

As mentioned in Section 3.1, the diffusion parameter of the clustering process can be estimated from an asymptotic solution through the integral Equation 22, leading to Equations 23 and 24 for delta-correlated and temporally correlated carrying velocity fields, respectively.

Starting with the decomposition of the velocity into potential and solenoidal parts Equation 1, we take the divergence:

$$\nabla \cdot \mathbf{u}(\mathbf{x}, t) = \gamma \nabla \cdot \mathbf{u}_{\mathrm{p}}(\mathbf{x}, t). \tag{C1}$$

Applying the divergence to the spectral representation of the potential component Equation 3 yields:

$$\nabla \cdot \mathbf{u}_{\mathrm{p}}(\mathbf{x}, t) = \Re\left\{ \int d^2\mathbf{k}\left(-b_{\mathrm{p}}(\mathbf{k}, t) + \mathrm{i}a_{\mathrm{p}}(\mathbf{k}, t)\right) k \exp(\mathrm{i}\mathbf{k} \cdot \mathbf{x}) \right\}. \tag{C2}$$

Expressing the real part of the quantity in the brackets:

$$\nabla \cdot \mathbf{u}_{\mathrm{p}}(\mathbf{x}, t) = \frac{1}{2} \int d^2\mathbf{k}\left[-\left(b_{\mathrm{p}}(\mathbf{k}, t) + b_{\mathrm{p}}(-\mathbf{k}, t)\right) + \mathrm{i}\left(a_{\mathrm{p}}(\mathbf{k}, t) - a_{\mathrm{p}}(-\mathbf{k}, t)\right)\right] k \exp(\mathrm{i}\mathbf{k} \cdot \mathbf{x}). \tag{C3}$$

Focusing on uncorrelated fields, we can derive Equation 23 using the statistical properties of the Gaussian random fields a_{p} and b_{p} in the uncorrelated case, namely:

$$\left\langle a(\mathbf{k}, t) a\left(\mathbf{k}', t + \eta\right) \right\rangle = t_0 \sigma_{\mathbf{u}}^2 E(k) \delta\left(\mathbf{k} - \mathbf{k}'\right) \delta(\eta), \tag{C4}$$

$$\left\langle b(\mathbf{k}, t) b\left(\mathbf{k}', t + \eta\right) \right\rangle = t_0 \sigma_{\mathbf{u}}^2 E(k) \delta\left(\mathbf{k} - \mathbf{k}'\right) \delta(\eta), \tag{C5}$$

$$\left\langle a_{\mathrm{p}}(\mathbf{k}, t) b_{\mathrm{p}}\left(\mathbf{k}', t'\right) \right\rangle = 0. \tag{C6}$$

Multiplying the two divergences in the integrand of Equation 22 and taking the expected value:

$$\langle \nabla \cdot \mathbf{u}(\mathbf{x}, t) \nabla \cdot \mathbf{u}(\mathbf{x}, t + \eta) \rangle = \gamma^2 \sigma_{\mathbf{u}}^2 t_0 \delta(\eta) \int d^2\mathbf{k} d^2\mathbf{k}' \delta\left(\mathbf{k} + \mathbf{k}'\right) k k' E(k) \exp\left[i\left(\mathbf{k} + \mathbf{k}'\right) \cdot \mathbf{x}\right], \tag{C7}$$

which further simplifies to give:

$$\langle \nabla \cdot \mathbf{u}(\mathbf{x}, t) \nabla \cdot \mathbf{u}(\mathbf{x}, t + \eta) \rangle = 2\pi\gamma^2 \sigma_{\mathbf{u}}^2 t_0 \delta(\eta) \int k^3 E(k)\, dk. \tag{C8}$$

Substituting the chosen spectrum Equation 8 and evaluating the integral yields:

$$\langle \nabla \cdot \mathbf{u}(\mathbf{x}, t) \nabla \cdot \mathbf{u}(\mathbf{x}, t + \eta) \rangle = 2\gamma^2 \frac{\sigma_{\mathbf{u}} t_0}{l^2} \delta(\eta). \tag{C9}$$

Using this as the integrand in Equation 22 produces:

$$D_p = 2\gamma^2 \frac{\sigma_{\mathbf{u}}^2 t_0}{l^2} \int_0^\infty d\eta\, \delta(\eta). \tag{C10}$$

Evaluating this integral recovers Equation 23. The typical diffusion timescale is then the inverse of the diffusion parameter $D_p = 1/\tau$.

We can perform a similar calculation for the correlated fields by solving the ODE Equation 6 to find the coefficients a_p and b_p. We get:

$$a_p(\mathbf{k}, t) = e^{-\lambda t} \int_0^\infty d\tau e^{\lambda \tau} \eta(\mathbf{k}, \tau), \tag{C11}$$

where the velocity field is initialized from the state of rest. From this we can further derive the statistics of a_p, making use of Equation 7 and the zero-mean property:

$$\left\langle a_p(\mathbf{k},t)a_p\left(\mathbf{k}',t'\right)\right\rangle = \frac{1}{2}\sigma_{\mathbf{u}}^2 e^{-\lambda(t-t')}\delta\left(\mathbf{k}-\mathbf{k}'\right), \tag{C12}$$

where $t' > t$, and we are in the asymptotic limit $t \to \infty$ (so the statistics are stationary). The result for $b_p(\mathbf{k}, t)$ is identical. Performing an analogous step to Equation C7 yields:

$$\langle \nabla\cdot\mathbf{u}(\mathbf{x},t)\nabla\cdot\mathbf{u}(\mathbf{x},t+\eta)\rangle = \frac{1}{2}\gamma^2\sigma_{\mathbf{u}}^2 e^{-\lambda\eta}\int d^2\mathbf{k}d^2\mathbf{k}'\delta\left(\mathbf{k}+\mathbf{k}'\right)kk'E(k)\exp\left[i\left(\mathbf{k}+\mathbf{k}'\right)\cdot\mathbf{x}\right]. \tag{C13}$$

This is mostly a repetition of the calculation for the delta-correlated fields. By simplifying the integral term and evaluating it in the same way as previously, we get:

$$\langle \nabla\cdot\mathbf{u}(\mathbf{x},t)\nabla\cdot\mathbf{u}(\mathbf{x},t+\eta)\rangle = \gamma^2\frac{\sigma_{\mathbf{u}}^2}{l^2}e^{-\lambda\eta}. \tag{C14}$$

Substituting this into Equation 22 results in

$$D_p = \gamma^2\frac{\sigma_{\mathbf{u}}^2}{l^2}\int_0^\infty d\eta\, e^{-\lambda\eta}, \tag{C15}$$

and evaluating this integral recovers Equation 24.

Data Availability Statement

Files containing all the detected clusters from every model run along with python scripts necessary to interpret them are available at https://zenodo.org/record/7248579 (Meacham, 2022b) and https://zenodo.org/record/7248624 (Meacham, 2022a). Figures were produced using Matplotlib 3.3.4.

Acknowledgments
Pavel Berloff was supported by the Leverhulme Trust Grant RPG-2019-024, by the NERC Grant NE/T002220/1 and by the Moscow Centre for Fundamental and Applied Mathematics (supported by the Agreement 075-15-2019-1624 with the Ministry of Education and Science of the Russian Federation). We would like to thank the reviewers for their detailed analysis of our manuscript, which benefited greatly from their input.

References

Andrady, A. L. (2011). Microplastics in the marine environment. *Marine Pollution Bulletin*, *62*(8), 1596–1605. https://doi.org/10.1016/j.marpolbul.2011.05.030

Archer, M., Schaeffer, A., Keating, S., Roughan, M., Holmes, R., & Siegelman, L. (2020). Observations of submesoscale variability and frontal subduction within the mesoscale eddy field of the Tasman Sea. *Journal of Physical Oceanography*, *50*(5), 1509–1529. https://doi.org/10.1175/JPO-D-19-0131.1

Baker, L., Frankel, A., Mani, A., & Coletti, F. (2017). Coherent clusters of inertial particles in homogeneous turbulence. *Journal of Fluid Mechanics*, *833*, 364–398. https://doi.org/10.1017/jfm.2017.700

Blanke, B., Speich, S., Madec, G., & Maugé, R. (2002). A global diagnostic of interior ocean ventilation. *Geophysical Research Letters*, *29*(8), 108-1–108-4. https://doi.org/10.1029/2001GL013727

Callies, J., Barkan, R., & Garabato, A. N. (2020). Time scales of submesoscale flow inferred from a mooring array. *Journal of Physical Oceanography*, *50*(4), 1065–1086. https://doi.org/10.1175/JPO-D-19-0254.1

Cózar, A., Echevarría, F., González-Gordillo, J. I., Irigoien, X., Úbeda, B., Hernández-León, S., et al. (2014). Plastic debris in the open ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), 10239–10244. https://doi.org/10.1073/pnas.1314705111

Ferenc, J.-S., & Néda, Z. (2007). On the size distribution of Poisson Voronoi cells. *Physica A: Statistical Mechanics and its Applications*, *385*(2), 518–526. https://doi.org/10.1016/j.physa.2007.07.063

Gower, J., Hu, C., Borstad, G., & King, S. (2006). Ocean color satellites show extensive lines of floating sargassum in the gulf of Mexico. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(12), 3619–3625. https://doi.org/10.1109/TGRS.2006.882258

Haza, A., Özgökmen, T., & Hogan, P. (2016). Impact of submesoscales on surface material distribution in a gulf of Mexico mesoscale eddy. *Ocean Modelling*, *107*, 28–47. https://doi.org/10.1016/j.ocemod.2016.10.002

Jordi, A., Basterretxea, G., & Anglès, S. (2009). Influence of ocean circulation on phytoplankton biomass distribution in the Balearic Sea: Study based on sea-viewing wide field-of-view sensor and altimetry satellite data. *Journal of Geophysical Research*, *114*(C11), C11005. https://doi.org/10.1029/2009jc005301

Klyatskin, V. (2003). Clustering and diffusion of particles and passive tracer density in random hydrodynamic flows. *Physics-Uspekhi*, *46*(7), 667–688. https://doi.org/10.1070/pu2003v046n07abeh001600

Kolmogorov, A. N. (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *The Proceedings of the USSR Academy of Sciences* (Vol. 30, pp. 301–305).

Koshel, K. V., Stepanov, D. V., Ryzhov, E. A., Berloff, P., & Klyatskin, V. I. (2019). Clustering of floating tracers in weakly divergent velocity fields. *Physical Review E—Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *100*(6), 063108. https://doi.org/10.1103/PhysRevE.100.063108

Liu, Y., Shen, L., Zamansky, R., & Coletti, F. (2020). Life and death of inertial particle clusters in turbulence. *Journal of Fluid Mechanics*, *902*, R1. https://doi.org/10.1017/jfm.2020.710

Martinez, E., Maamaatuaiahutapu, K., & Taillandier, V. (2009). Floating marine debris surface drift: Convergence and accumulation toward the south Pacific subtropical gyre. *Marine Pollution Bulletin*, *58*(9), 1347–1355. https://doi.org/10.1016/j.marpolbul.2009.04.022

Maximenko, N., Hafner, J., & Niiler, P. (2012). Pathways of marine debris derived from trajectories of Lagrangian drifters. *Marine Pollution Bulletin*, *65*(1), 51–62. (At-sea Detection of Derelict Fishing Gear). https://doi.org/10.1016/j.marpolbul.2011.04.016

McWilliams, J. C. (2016). Submesoscale currents in the ocean. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *472*(2189), 20160117. https://doi.org/10.1098/rspa.2016.0117

Meacham, J. (2022a). Post-processing scripts for clustering data. [Software]. https://doi.org/10.5281/zenodo.7248624

Meacham, J. (2022b). Clustering data for passive tracers in random velocity fields [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7248579

Monchaux, R., Bourgoin, M., & Cartellier, A. (2012). Analyzing preferential concentration and clustering of inertial particles in turbulence. *International Journal of Multiphase Flow*, *40*, 1–18. https://doi.org/10.1016/j.ijmultiphaseflow.2011.12.001

Onink, V., Wichmann, D., Delandmeter, P., & van Sebille, E. (2019). The role of Ekman currents, geostrophy, and Stokes drift in the accumulation of floating microplastic. *Journal of Geophysical Research: Oceans*, *124*(3), 1474–1490. https://doi.org/10.1029/2018jc014547

Rocha, C. B., Chereskin, T. K., Gille, S. T., & Menemenlis, D. (2016). Mesoscale to submesoscale wavenumber spectra in Drake Passage. *Journal of Physical Oceanography*, *46*(2), 601–620. https://doi.org/10.1175/JPO-D-15-0087.1

Rypina, I. I., Getscher, T., Pratt, L., & Ozgokmen, T. (2022). Applying dynamical systems techniques to real ocean drifters. *EGUsphere*, 1–42. https://doi.org/10.5194/egusphere-2022-250

Stepanov, D. V., Ryzhov, E. A., Berloff, P., & Koshel, K. V. (2020). Floating tracer clustering in divergent random flows modulated by an unsteady mesoscale ocean field. *Geophysical & Astrophysical Fluid Dynamics*, *114*(4–5), 690–714. https://doi.org/10.1080/03091929.2020.1786551

Stepanov, D. V., Ryzhov, E. A., Zagumennov, A. A., Berloff, P., & Koshel, K. V. (2020). Clustering of floating tracer due to mesoscale vortex and submesoscale fields. *Geophysical Research Letters*, *47*(3), e2019GL086504. https://doi.org/10.1029/2019gl086504

van Sebille, E., Sprintall, J., Schwarzkopf, F. U., Sen Gupta, A., Santoso, A., England, M. H., et al. (2014). Pacific-to-Indian ocean connectivity: Tasman leakage, Indonesian throughflow, and the role of ENSO. *Journal of Geophysical Research: Oceans*, *119*(2), 1365–1382. https://doi.org/10.1002/2013JC009525

Wilson, J. D., & Sawford, B. L. (1996). Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere. In J. R. Garratt & P. A. Taylor (Eds.), *Boundary-layer meteorology 25th anniversary volume, 1970–1995: Invited reviews and selected contributions to recognise ted Munn's contribution as editor over the past 25 years* (pp. 191–210). Springer Netherlands. https://doi.org/10.1007/978-94-017-0944-6_9