

Geophysical Research Letters[®]

RESEARCH LETTER

10.1029/2023GL107900

Unraveling the Non-Homogeneous Dispersion Processes in Ocean and Coastal Circulations Using a Clustering Approach



Key Points:

- We formulated and applied a clustering algorithm to classify oceanographic dispersion processes starting from Lagrangian trajectories
- A new analytic model for the autocorrelation functions is proposed which well describes loopers and non-loopers particles behaviors
- We identified three characteristic time scales to distinguish complex inhomogeneous dispersion processes typical of ocean circulations

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Stocchino,
alessandro.stocchino@polyu.edu.hk

Citation:

Lagomarsino-Oneto, D., De Leo, A., Stocchino, A., & Cucco, A. (2024). Unraveling the non-homogeneous dispersion processes in ocean and coastal circulations using a clustering approach. *Geophysical Research Letters*, 51, e2023GL107900. <https://doi.org/10.1029/2023GL107900>

Received 17 DEC 2023

Accepted 22 JAN 2024

Author Contributions:

Conceptualization: D. Lagomarsino-Oneto, A. De Leo, A. Stocchino, A. Cucco

Data curation: D. Lagomarsino-Oneto, A. Stocchino

Formal analysis: D. Lagomarsino-Oneto, A. De Leo, A. Stocchino, A. Cucco

Funding acquisition: A. Stocchino

Investigation: A. Stocchino

Methodology: D. Lagomarsino-Oneto, A. Cucco

Resources: A. Stocchino

Supervision: A. Stocchino

Validation: A. Stocchino, A. Cucco

Visualization: A. De Leo

Writing – original draft:

D. Lagomarsino-Oneto, A. Stocchino

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

D. Lagomarsino-Oneto¹ , A. De Leo² , A. Stocchino^{2,3} , and A. Cucco⁴ 

¹Institute of Marine Sciences (ISMAR), National Research Council (CNR), Lerici, Italy, ²Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, ³State Key Laboratory of Marine Pollution, City University of Hong Kong, Hong Kong, Hong Kong SAR, ⁴National Research Council (CNR), Institute for the Study of Anthropic Impacts and Sustainability in Marine Environment (IAS), Oristano, Italy

Abstract Dispersion processes in environmental flows have been traditionally studied under the strong assumption of homogeneous, isotropic and stationary turbulence. To overcome this limitation, we propose a new approach that combines autocorrelation analysis of simulated Lagrangian trajectories together with unsupervised clustering. To test the approach, we consider several dynamic scenarios around a coastal gulf, subject to different forcing, in order to compare our method with other approaches. Lagrangian trajectories forced by the varying coastal circulation exhibited different behaviors, looping and non-looping paths, and produced a variety of Lagrangian autocorrelation functions. Our approach proves to be able to reveal spatio-temporal variations in ocean dispersion processes without any a priori knowledge of the character of the trajectories. Clusters based on the autocorrelation functions are associated to different inhomogeneous dispersion processes. Finally, we propose a new stochastic model capable of predicting the different forms of autocorrelations.

Plain Language Summary Ocean and coastal circulations develop in complex domains, especially along the shorelines, and the resulting flow is turbulent in character and inherits the inhomogeneities from the generating forces. When we come to study how these chaotic circulations transport mass, we must expect that the associated dispersion is equally turbulent and high variable in time and space. Observations of particle paths taught us how the trajectories could be complicated, often showing looping behaviors generated by different mechanisms. Despite this complexity, many available studies on ocean and coastal dispersion rely on considering the process as homogeneous (no variations in space) and, applying different spatial and temporal averages, try to grasp the overall picture of the dispersion. We propose a new approach that combines the fundamentals of the dispersion theories with an automated algorithm for clustering. We show that this approach is able to retain the highly inhomogeneous character of the ocean dispersion, at the same time, showing the physical link between the circulations and its ability to transport mass.

1. Introduction

The ability of ocean and coastal circulations to transport heat and mass has always been considered of paramount importance due to the impact on atmospheric climate (Holmes et al., 2019) and the distribution of nutrients essential for sustaining ocean life (Lu & Gan, 2023). Mass and heat transport are strongly influenced by the local dispersion processes, which in turn are controlled by ocean and coastal circulations. Material dispersion by environmental flows have been traditionally studied following Taylor's theory (Taylor, 1921). He was inspired by observing that turbulence on geophysical scales spreads heat and substances similarly to what happens in molecular diffusion. Following this idea, he discovered the connection between the time evolution of particle dispersion and the velocity autocorrelations of the underlying flow. This also implies the existence, for large enough times, of an asymptotic regime with an associated turbulent diffusion coefficient (Eddy Diffusivity).

Although these fascinating results had been proven to be effective in certain real contexts (LaCasce, 2008), the assumption of homogeneous, isotropic and stationary turbulence is often not correct in geophysics. Therefore, several methods have been proposed to overcome this limitation, usually detecting homogeneous subsets where Taylor's hypotheses are restored. Examples of methodologies, aimed to obtain spatial distributions of the diffusivities, can be found in Davis (1991), later adopted by Jakobsen et al. (2003). Moreover, coherent structures are

Writing – review & editing: A. De Leo,
A. Cucco

known to affect dispersal statistics. Indeed, ocean eddies can trap particles for periods of the order of their typical life-time (Provenzale, 1999), showing loops in Lagrangian trajectories (loopers). In Richardson (1993), the importance of coherent eddies in setting the oceanic kinetic energy budget was demonstrated by classifying float trajectories with two or more consecutive loops in the same direction. Afterward several studies aimed to make this classification more systematic.

In particular, Veneziani et al. (2004) studied the loopers distribution in North Atlantic mesoscale turbulence by thresholding a scalar quantity, called spin, to automatically separate trajectories. They were able to identify geographical areas dominated either by looper or non-loopers, hence quantifying dispersal on statistically homogeneous regions. This approach was inspired by thinking Lagrangian velocities as realizations of a linear stochastic process with a deterministic term that corresponds to a uniform circular motion. A similar strategy was developed in Rupolo (2007), where trajectories were classified introducing a new dimensionless parameter, defined as the ratio between the acceleration and velocity time scales, that was inspired by a model in which deterministic circular motions are substituted by adding accelerations in the stochastic process. This strategy allowed a worldwide evaluation of dispersal properties in the main oceanic current systems. Furthermore, Griffa et al. (2008) focused the attention on dynamics in the upper ocean following the same approach of Veneziani et al. (2004) to analyze data from drifter measurements. The same data set was exploited by Lumpkin (2016) to propose a more robust criterion for loopers detection. Instead of considering chunks of trajectories of a fixed duration (as done in previous works), they split each trajectory into intervals where spin persisted keeping the same sign, then judging as loopers those intervals whose duration was at least twice the period of the circular motion associated to the corresponding spin value. Other works addressed loopers detection from purely geometric perspectives. Dong et al. (2011) presented a methodology to identify trajectories which cross themselves in at least two points, looping in the same direction and thresholding on the mean period of rotation. Methods proposed in Lilly and Olhede (2009) and Lilly et al. (2011) were designed to separate from background flow the motion due to waves and vortices. A valuable feature of their approach was to allow a description of trajectories that went beyond the simplifying assumption of near-circular loops. The effect of eddies on tracer dispersion has been also largely studied in the Eulerian framework. Particularly in Haigh et al. (2020, 2021a, 2021b) a quasi-geostrophic ocean motion was simulated to systematically investigate the effect of small-scale structures on Eddy Diffusivity fields of tensor form. Finally, recent advances in Deep Learning and Artificial Intelligence have been introduced also in this field. For example, Xu et al. (2019) borrowed state-of-art techniques from Computer Vision to detect surface eddies from SSH data outperforming more traditional algorithms. Moreover, a clustering approach has been introduced by I. M. Koszalka and LaCasce (2010) on Lagrangian drifter data. In the latter study, the binning technique used by Davis (1991) was replaced by the use of a clustering algorithm of the drifter-derived velocities.

In our study we go back to Taylor's statement that, for a homogeneous environment, the dispersal process is driven by the velocity correlation. We apply unsupervised Machine Learning to automatically identify clusters of trajectories with similar velocity autocorrelations. While being based on a solid theoretical result this method is fully data-driven, hence overcoming limitations related with any a priori modeling assumption about the underlying processes and avoiding technical issues like tuning thresholds. The method is compared with previous strategies on changing dynamic scenarios around a coastal gulf (in terms of both circulation and wind). Our approach can automatically depict spatio-temporal variations in ocean dispersion, recovering also the already known phenomenology, particularly loopers. Finally, to interpret our results we propose a stochastic model that describes well all different profiles of autocorrelation identified by the algorithm. Discussing the meaning of model parameters also gives the opportunity for a critical review of quantities which have been traditionally used to characterize dispersion.

2. Material and Methods

2.1. The Study Area and the Circulation Model

For the present analysis, we selected as study area the Gulf of Oristano (Italy), located on the west coast of Sardinia Island (approximately between 8.48 E and 8.68 E and 39.78 N and 39.98 N, Italy). It is a shallow water semi-enclosed bay of about 150 km² (Figure S1 in Supporting Information S1) with a 9 km wide opening into the open sea (Western Mediterranean Sea). We refer to the Supporting Information S1 for further details.

In the past decades various oceanographic studies have been carried out to investigate the water circulation in this stretch of coast (Coppa et al., 2013, 2019; Cucco, Sinerchia, Lefrançois, et al., 2012; Cucco, Sinerchia, Ribotti,

et al., 2012; Cucco et al., 2006, 2016; De Falco et al., 2008; Farina et al., 2018; Ferrarin & Umgiesser, 2005; Magni et al., 2008; Umgiesser et al., 2014). In most of the applications SHYFEM (System of Hydrodynamic Finite Element Model) (Umgiesser et al., 2004), a 3D-oceanographic model based on the finite element method, has been successfully applied to investigate the main hydrodynamics and the transport time scales in the Gulf and surrounding coastal areas (Coppa et al., 2013, 2019; Cucco, Sinerchia, Lefrançois, et al., 2012; Cucco et al., 2006, 2016; Farina et al., 2018).

In this study, we adopted SHYFEM to reproduce the surface water circulation inside and outside the Gulf during a whole meteorological year including the main atmospheric and oceanographic forcing. The same model setup adopted in Farina et al. (2018) was used to reproduce the main surface transport path in the area. Numerical simulations were performed to reproduce the surface water circulation during the years 2017 and 2018. The model outputs, consisting of the horizontal components of the current velocity computed at 2.5 m depth and at hourly frequency, were resampled on a 100 m regular mesh. Only the data obtained from the second-year model run were used for the analysis in order to avoid any disturbance generated by the initial conditions of the model.

2.2. Clustering Technique and Its Application to Lagrangian Dispersion

Clustering is a branch of unsupervised learning that aims at finding groups (clusters) of nearby points in a given data set. A cluster is then meaningful if distances between its points are significantly smaller than distances between its points and those belonging to other clusters. Specifically, K-means divides the data space into K clusters producing for each a corresponding prototype μ_k (or in other words the centroid of the cluster) and a label indicating to which cluster each sample data belongs. Mathematically this problem can be stated as follows: given a data set of N points $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ in a D -dimensional Euclidean space with norm $\|\dots\|$, solve the optimization problem of finding $\arg \min_{(\mu_k, r_{nk})} J$, with

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|\mathbf{x}_n - \mu_k\|^2 \quad (1)$$

where r_{nk} is 1 if the n th sample is assigned to the k th cluster, identified by the prototype μ_k , and 0 otherwise. In other words, we seek for a pair (μ_k, r_{nk}) that minimizes the sum of the squares of the distances of each point to its associated prototype. It can be seen easily that for a fixed set of prototypes the following definition of r_{nk} is optimal:

$$r_{nk} = \begin{cases} 1 & \text{if } k = \operatorname{argmin}_j \|\mathbf{x}_n - \mu_j\|^2 \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Moreover, for a fixed partition r_{nk} of a given data set, a minimum of the cost function J is given by

$$\mu_k = \frac{\sum_n r_{nk} \mathbf{x}_n}{\sum_n r_{nk}}, \quad (3)$$

which means that each optimal prototype is in fact the centroid of all points in the corresponding cluster. Equations 2 and 3 represent a solution of the K-means problem, for more details on the algorithm see for example, Bishop and Nasrabadi (2006).

A key point of our method, and the main difference compared to previous studies (I. Koszalka et al., 2011; I. M. Koszalka & LaCasce, 2010), is to apply clustering to Lagrangian velocity autocorrelations of the i th Lagrangian velocity component u_{L_i} , defined following Taylor (1921) as:

$$\mathcal{R}_{ii}(\tau) = \frac{\rho_{L_i}(\tau)}{\sqrt{\rho_{L_i}(0)^2}}, \quad \rho_{L_i}(\tau) = \langle u_{L_i}(t) u_{L_i}(t + \tau) \rangle. \quad (4)$$

where the brackets $\langle \cdot \rangle$ in $\rho_{L_i}(\tau)$ indicate an average over the entire duration of each trajectory. Moreover, the integral Lagrangian time scale along the i th direction can be then evaluated as the time integral of Equation 4.

Four test cases were performed varying the main parameters of the Lagrangian simulations, see Table S1 in Supporting Information S1. Hereinafter, we only show the results for the Test 1, and we refer to the SI document for comparison among the test cases.

Specifically the components of a sample vector \mathbf{x}_n are values of either $\mathcal{R}_{uu}(\tau)$ or $\mathcal{R}_{vv}(\tau)$ of a given Lagrangian trajectory, at each time lag τ considered. Two separate clusterings have been performed for \mathcal{R}_{uu} and \mathcal{R}_{vv} . The clustering algorithm was preliminary tested to find the optimal number of clusters, see SI document for the details.

2.3. Stochastic Model for Surface Tracers

Velocity autocorrelation profiles returned by our machine learning approach are physically interpretable. However, some of them cannot be described through models which were used in previous homogeneous and stationary contexts (Lumpkin, 2016; Veneziani et al., 2004).

To make a step toward a physical representation of this heterogeneous behavior, we propose a stochastic model for the velocity of our tracers, extending the model proposed in Veneziani et al. (2004, 2005). Here the two components of the Lagrangian velocity are described as a bivariate linear stochastic process with the form:

$$(u, v) = (\mathcal{U} + U, \mathcal{V} + V) \quad (5)$$

$$\left\{ \begin{array}{l} d\mathcal{U} = -\frac{1}{T_{\mathcal{U}}}dt - \Omega \mathcal{V} dt + \lambda_{\mathcal{U}} d\xi_{\mathcal{U}} \\ d\mathcal{V} = -\frac{1}{T_{\mathcal{V}}}dt + \Omega \mathcal{U} dt + \lambda_{\mathcal{V}} d\xi_{\mathcal{V}} \\ dU = -\frac{1}{T_U}U dt + \Lambda_U dW_U \\ dV = -\frac{1}{T_V}V dt + \Lambda_V dW_V \end{array} \right. \quad (6)$$

where the equations for \mathcal{U} and \mathcal{V} in the system (6) are the same velocities defined in the model presented in Veneziani et al. (2004), which was already able to produce looping trajectories, for example, Here we add two more independent one-dimensional processes U and V , with exponential decaying autocorrelation functions to each velocity component. Assuming all decaying time scales equal to T and sorting all constants, the whole process generates trajectories associated with velocity autocorrelations of the form:

$$\mathcal{R}_{uu}(\tau) = \mathcal{R}_{vv}(\tau) = \frac{e^{-\tau/T}(\cos \Omega \tau + C)}{1 + C}. \quad (7)$$

It is worth observing that Ω is the constant angular frequency associated to a circular contribution to particle motion, resulting by the coupling terms of first and second equation in system (6). This parameter has been usually called spin and estimated as:

$$\Omega = \frac{\langle u' dv' - v' du' \rangle}{2\Delta t E_c}, \quad (8)$$

where (u', v') are the Lagrangian residual velocities, E_c the kinetic energy, brackets denote average along a trajectory and Δt is the numerical time step. The spin or its associated rotational period $P = 2\pi/\Omega$ has been used in several studies to characterize the looping frequency (or the looping period) of the drifters' trajectories (Griffa et al., 2008; Lumpkin, 2016; Veneziani et al., 2004). In our analysis we mostly use Ω to indicate its instantaneous value (i.e., without applying the bracket average), and its average will be explicitly indicated. The constant C in Equation 7 plays the role of a coupling constant, tuning the relative importance of the two additional processes with respect to the initial two-dimensional model.

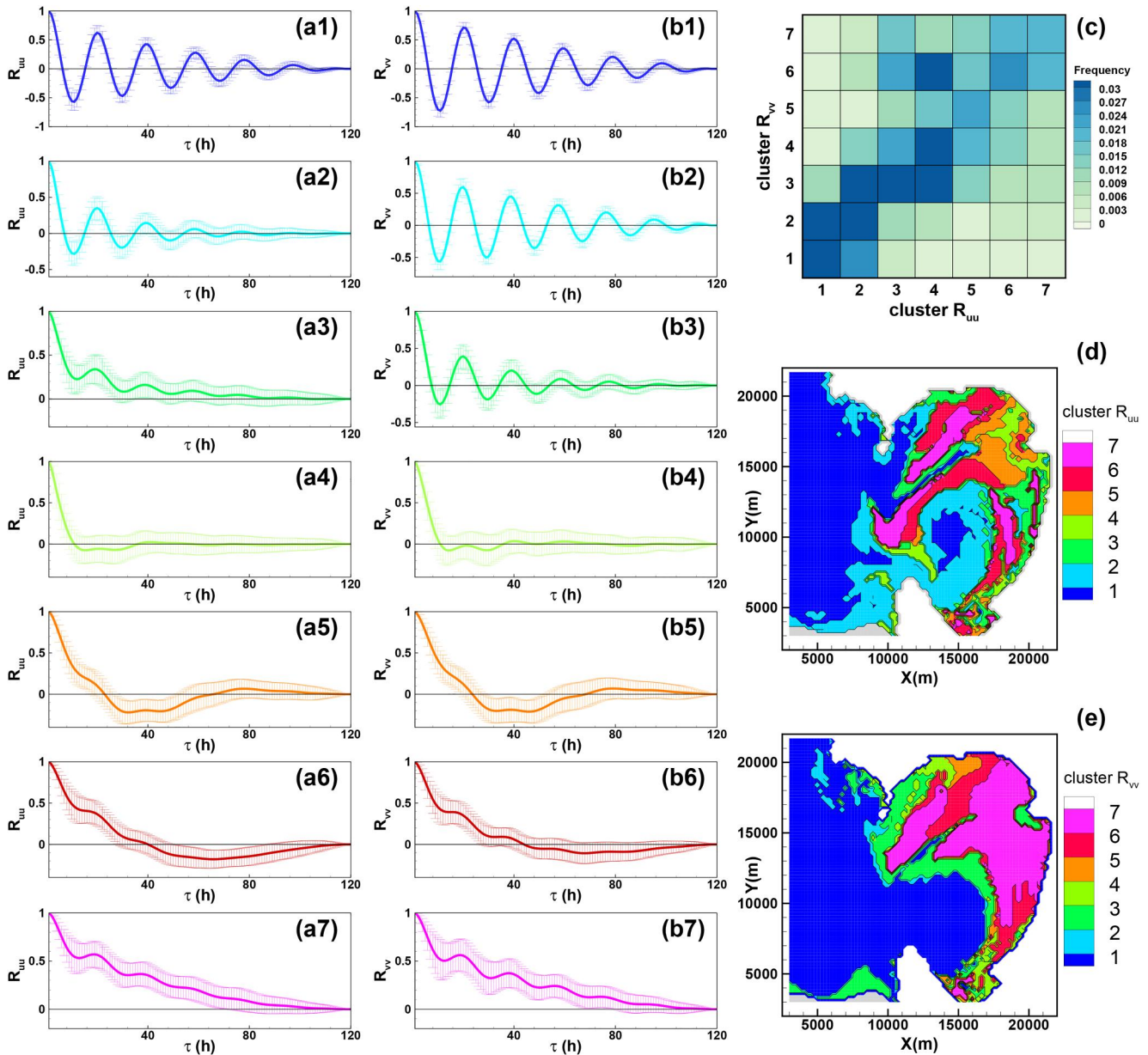


Figure 1. (a1-7, b1-7): representative autocorrelation profiles of the 7 clusters for \mathcal{R}_{uu} and \mathcal{R}_{vv} , respectively. The error bars indicate one standard deviation with respect to the mean profile. (c): occurrence matrix (relative frequency of appearance for each pair of \mathcal{R}_{uu} and \mathcal{R}_{vv} clusters). (d, e) Geographical pattern of the clusters based on \mathcal{R}_{uu} and \mathcal{R}_{vv} , respectively. The results used for panels (d, e) correspond to a single simulation, sequence 26.

3. Results and Discussion

As a first result our method retrieves, in a fully data-driven way, typical behaviors that were already identified in previous works. In particular, profiles 1 and 2 for \mathcal{R}_{uu} (Figures 1a1 and 1a2) and profiles 1, 2, and 3 for \mathcal{R}_{vv} (Figures 1b1–1b3) are immediately associated with the well-known looping trajectories. Exponentially and nearly exponentially decaying autocorrelation functions can be also recognized, for example, in cluster 7 (Figures 1a7 and 1b7). Other clusters may be related either to peculiar regimes or to transitional cases, as discussed in the following paragraphs.

The joint occurrence of \mathcal{R}_{uu} and \mathcal{R}_{vv} can be also evaluated, see the SI document for the details on the computation. The joint occurrence matrix, built on the entire 72 simulated initial conditions across the year (Figure 1c), shows the highest frequencies in diagonal and nearly diagonal entries, indicating that similar profiles tend to occur

simultaneously in both components. The strong diagonality observed indicates that the dispersion process tends to be isotropic, while preserving a strong non-homogeneity in space and time.

Furthermore, the assignment of each simulated trajectory to a cluster provides spatial patterns of the transport process across the studied domain (Figures 1d and 1e). Such patterns consist of evolving geographical patches that can be reviewed as homogeneous in terms of dispersal behavior. Note that the functions shown in the panels (a1-a7), (b1-b7) and (c) are the output obtained by clustering on the whole set of simulations of 120hr duration, whereas panel (d) and (e) are the results of a specific simulation with a duration of 120hr. In the Supporting Information S1 we provided the movie showing the time evolution of the clusters for the entire simulated year based either on \mathcal{R}_{uu} or \mathcal{R}_{vv} , and the corresponding joint occurrence matrix (clustermovie.mp4 (Movie S1)). It can be observed how the dispersion process is characterized by relatively strong inhomogeneities in space and time, which might be ascribed to the shoreline complexity and the variability of the circulation forcing.

A remarkable capability of our approach is that it allows disentangling of complex scenarios that may occur under different dynamical conditions. Panel (a) in Figure 2 shows the abundance of both \mathcal{R}_{uu} and \mathcal{R}_{vv} clusters for each non-overlapping time sequence (see Supporting Information S1 for details). The complexity of our case study immediately emerges, with transitions from looping to non-looping dominated scenarios and vice versa. Spatial patterns can be used to identify which geographical areas are subject to specific dispersion regimes and to associate such regimes to physical processes that play a role in determining the local circulation: in particular, wind has been shown to be the main driven factor in this area (Cucco et al., 2016). Figure 2 shows different scenarios generated during four selected time sequences in which wind behaves differently in terms of either direction or intensity (Figures 2b1, 2c1, 2d1, and 2e1). In previous works, the identification of loopers was based on thresholding absolute values of the spin parameter Ω (Griffa et al., 2008; Lumpkin, 2016). Here, we find that spatial patterns of high spin (Figures 2b2, 2c2, 2d2, and 2e2) are in agreement with those of simultaneously oscillating \mathcal{R}_{uu} and \mathcal{R}_{vv} profiles. However, defining loopers in terms of Ω , which gave good results in homogeneous cases, might be more sensitive to the choice of a threshold in our study area. Indeed, here the kinetic energy (that appears in the definition of Ω in Equation 8) might be not uniformly distributed, being sometimes subject to strong gradients (Figures 2b3, 2c3, 2d3, and 2e3). A good property of our approach based on profiles of autocorrelation function is to be more robust against this possible issue. The relation between spin and kinetic energy for each cluster is further supported by the scatter plots reported in the Supporting Information S1 (Figure S7 and S8). Moreover, any method based only on spin does not allow to identify further regimes like those depicted in Figures 2e4 and 2e5 where profiles which exhibit negative lobes in the autocorrelation (Figures 1a5 and 1b5) are dominant in the gulf. Negative lobes are often associated to non-Brownian diffusive regimes, both sub-diffusive and super-diffusive regimes (Berloff et al., 2002; De Leo et al., 2022). In particular, sub-diffusion is associated to a pronounced negative lobe, whereas a negative lobe followed by a positive lobe is linked to super-diffusive processes. Examples of these anomalous dispersion regimes have been observed in the transport due to oceanic gyres (Berloff et al., 2002; Veneziani et al., 2004) and in coastal regions as well (De Leo et al., 2022; Enrile et al., 2019). The general result that arises from our analysis is that strategies successfully used to characterize dispersion in homogeneous cases are not sufficient when applied to more complex scenarios. Our proposed method aims at tackling this lack of generalization in a fully data-driven manner.

If we now consider the functional form for the autocorrelations expressed by Equation 7, we can compute the distributions of the three parameters (T , P and C) by best-fitting each \mathcal{R}_{uu} and \mathcal{R}_{vv} with Equation 7, see Supporting Information S1 for the details of the nonlinear fitting.

The T and P time scales also appear in the form previously used for velocity autocorrelations (Lumpkin, 2016; Veneziani et al., 2004):

$$\mathcal{R}_{uu}(\tau) = \mathcal{R}_{vv}(\tau) = e^{-\tau/T} \cos(2\pi\tau/P). \quad (9)$$

However, this formulation cannot describe profiles such as those in panel (a3) of Figure 1. On the contrary, the formulation suggested by Rupolo (2007) is able to describe autocorrelation functions with a single pronounced negative lobe depending on the relative values of the velocity decorrelation time compared to the acceleration decorrelation time. However, the latter failed to describe strongly looping autocorrelation functions.

The new parameter C must be greater than or equal to zero and establishes the relative importance of looping and non-looping components in the stochastic process. Hence, cases described by Equation 9 are retrieved for $C = 0$,

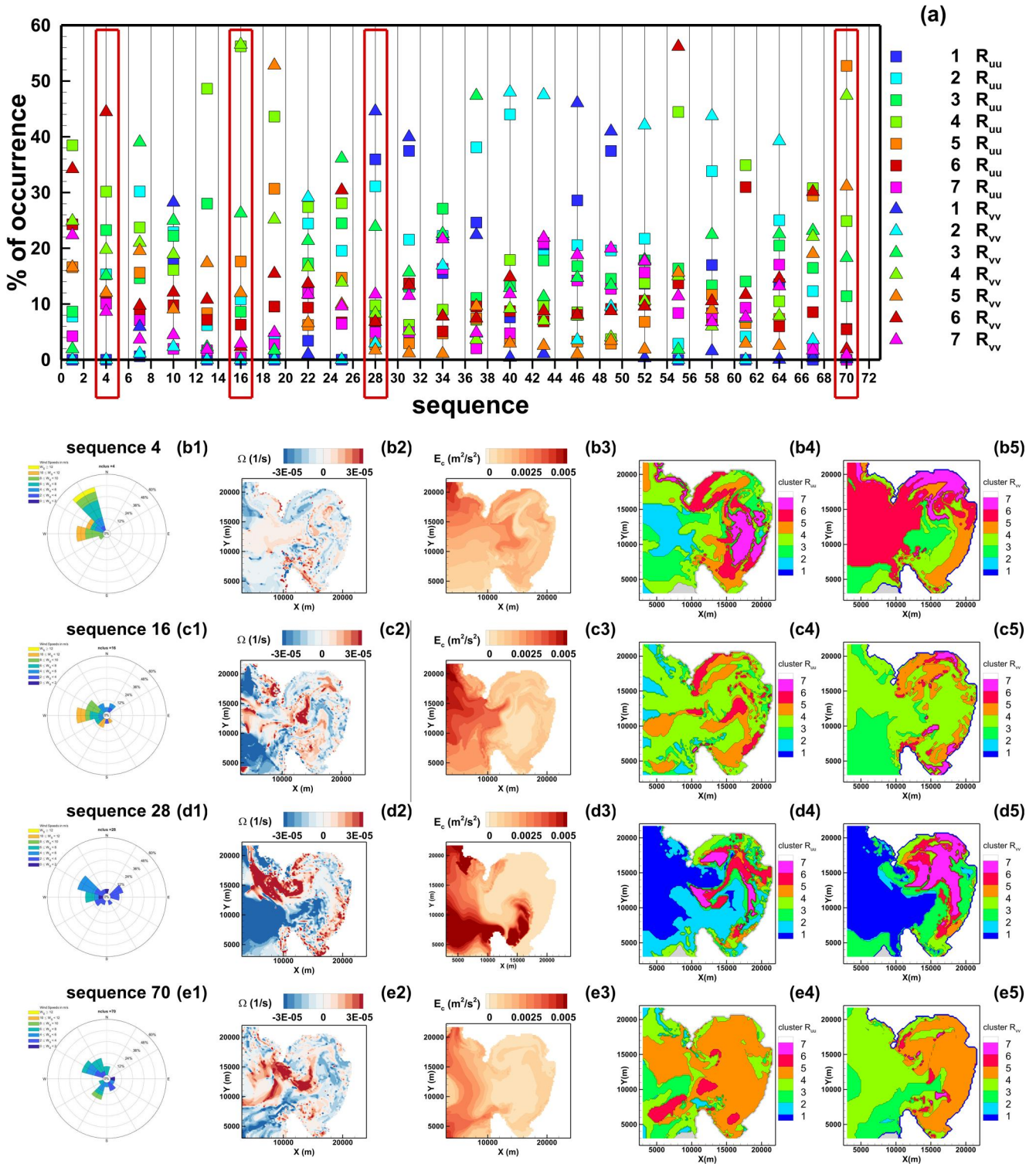


Figure 2. (a) Percentage of occurrence of the different clusters shown for 24 out of 72 simulations. Solid square indicates the clusters based on R_{uu} and the solid triangles show the clusters of R_{vv} . The red rectangles indicate the four sequences shown in the other panels, namely sequence 4 (b1-b5), 16 (c1-c5), 26 (d1-d5), and 70 (e1-e5). The four panels, from left to right, for every sequence (row) show the wind rose associated with the 120 hr simulation, the spatial distribution of the sequence averaged spin, the spatial distribution of the sequence averaged kinetic energy, the spatial distribution of the clusters for R_{uu} and R_{vv} .

values of $0 < C \ll 1$ are mixed cases dominated by loopers, and the weight of the exponentially decaying part grows as long as C increases. It can be also observed that this latter limit case of $C \gg 1$ can be obtained by the particular condition $C = P = 0$, although in practice this combination never emerges from our best fits (Figures 3e and 3h).

Combined with the data-driven approach proposed here, the estimation of parameters describing emergent profiles provides links to physical quantities traditionally used in dispersal studies and allows their change to be tracked across clusters. We have already mentioned the spin Ω as a successful quantity for loopers detection. In the ideal case when looping particles, deprived of their mean motion, undergo a uniform circular motion with radius r and period P , the latter turns out to be exactly $P = 2\pi/\Omega$. Estimates of P_x and P_y reveal a well-defined value around 20 hr for strong loopers (Figures 3e and 3h), which gives about $0.9 \cdot 10^{-4} \text{ s}^{-1}$ for the magnitude of Ω and roughly corresponds to the most frequent value for cluster 1 in \mathcal{R}_{uu} (Figure 3c). However, while under the ideal assumption of uniform circular motion the spin does not depend on time, in our case evaluating Ω locally in time yields very different values, even close to zero where the trajectory has low curvature. This produces distributions of Ω which do not show typical emptying close to zero that was reported in previous works instead (Lumpkin, 2016). On the contrary, other clusters show narrower spin distributions placed around zero, as expected. Spin analysis also reveals skewness in clusters that exhibit looping behavior, suggesting that in our case study, rotating particles tend to prefer a direction of rotation. This fact could be associated with some coherent process happening in the area like inertial oscillations which have a period that is compatible with observed values of P_x and P_y . Indeed, for looping trajectories these two parameters show very concentrated distributions of values around 20 hr that could be associated to typical periods of inertial gravity waves.

Similar to the definition of Taylor (1921)'s integral time scale, by simple integration of Equation 7, we can derive the formula:

$$T_L = \left(\frac{1}{1+C} \right) \left[\frac{T}{1 + \frac{4\pi^2 T^2}{P^2}} + CT \right], \quad (10)$$

that relates T , P , and C with T_L as defined by Taylor. Under the assumption of homogeneous turbulence, this time scale also represents the time separation that has to be significantly overcome to have uncorrelated Lagrangian velocities, thus indicating a transition from a ballistic to a diffusive regime in particle dispersion. However, it is well known that if Taylor's assumptions are not fulfilled, this association with the establishment of a diffusive regime could be not reached. This is also the case of looping particles, where lobes in the autocorrelation of opposite sign largely cancel their contribution to the integral, yielding very low values (e.g., substituting the values of $T = 40$ hr, $P = 20$ hr and $C = 0$ in Equation 10 gives $T_L \sim 15$ min). For this reason, in presence of oscillations, the Lagrangian time scale sometimes has been alternatively defined as the first zero of the autocorrelation (Lumpkin, 2016; Pope, 2000), which for $C = 0$ is $P/4$. Although this latter definition can be more effective for certain applications, it does not apply to positive valued autocorrelation functions and it is sometimes replaced considering the first lag time at which it drops below a given threshold (Azevedo et al., 2017).

What emerges clearly from our analysis is the role of three distinct time scales when it comes to distinguish different dispersion regimes. Indeed, the proposed model for the Lagrangian velocity autocorrelation couples together T , P , and T_L , the last one computed from \mathcal{R}_{uu} and \mathcal{R}_{vv} . Figure 3 panels (j) and (k) show the time distribution of the total absolute dispersion $a^2(t)$ (Enrile et al., 2019; LaCasce, 2008) averaged over each cluster computed using \mathcal{R}_{uu} . From inspecting the time trends of $a^2(t)$, it is clear how Taylor's definition of T_L marks the end of the t^2 scaling. The time scale at which the asymptotic diffusive regime starts is, instead, better represented by T . For the non-looping trajectories, the Taylor integral scale T_L correctly tends to coincide with the estimated T . The signature of the looping character still persists in the $a^2(t)$ curves, see cluster 1 in panel (j). A similar behavior was observed by Enrile et al. (2019) where the absolute dispersion was computed starting from High-Frequency radar measurements. In the above case, the tidal forcing was responsible for the sustainment of looping trajectories and the corresponding absolute dispersion showed oscillations that tend to be dumped around an almost linear trend on a long time-scale, as in the present case. Moreover, anomalous dispersion regime (sub- or super-diffusive) can be observed for the case where the autocorrelation functions showed pronounced negative and

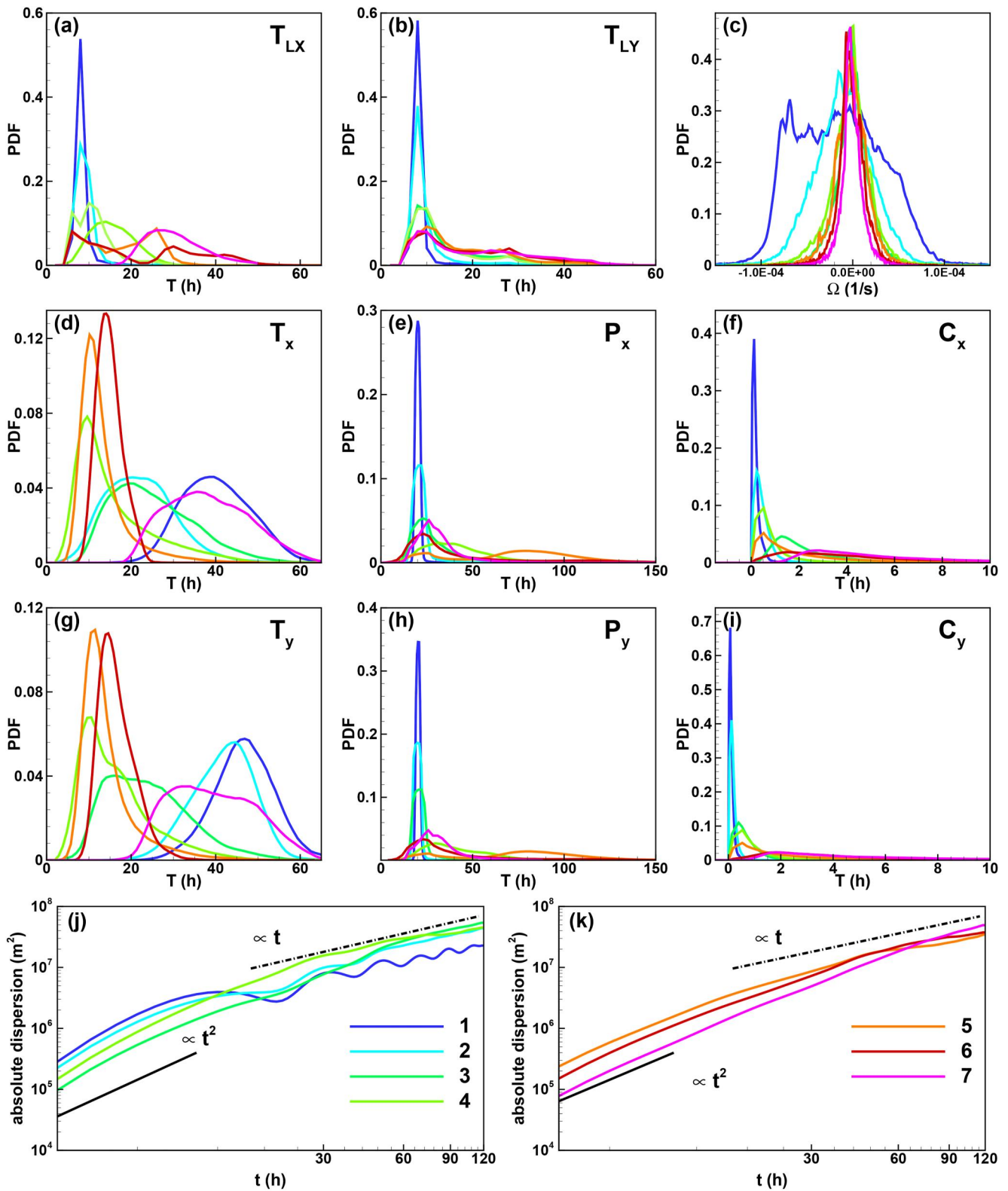


Figure 3. (a, b) Show the pdf of the integral Lagrangian time scale in x and y respectively. (c): pdf of the spin values associated with each cluster. (d–f) Show the pdf distribution for T , P and C of the proposed formula of Equation 7 for \mathcal{R}_{uu} . (g–i) Pdfs of T , P and C for \mathcal{R}_{vv} . (j, k) Distribution of the total absolute dispersion for each \mathcal{R}_{uu} cluster.

positive lobes (Figures 1a4–1a6). Finally, it is worth noting that similar results for the absolute dispersion were obtained averaging over the clusters based on \mathcal{R}_{vv} , see Figure S9 in Supporting Information S1.

4. Conclusions

The purpose of the present analysis was twofold: first, we aimed to apply a technique based on unsupervised clustering to the velocity autocorrelation functions to distinguish regions with similar dispersion properties; second, we proposed a new stochastic model that is able to represent the variability of the autocorrelation functions, extending the work of Veneziani et al. (2004) and Rupolo (2007). We tested the proposed approach using a 1-year long simulation of the coastal circulation of a Gulf in the Western Mediterranean Sea. However, the proposed methodology is generally applicable also in open ocean contexts. Lagrangian velocity autocorrelations were computed using the output of the circulation model. The application of the K-Means algorithm proved to generate clusters that could be clearly interpreted in terms of dispersion processes. The main output is a temporal and spatial mapping of coastal sub domains with similar dispersion properties. The proposed approach showed to be able to efficiently distinguish among looping and non-looping particle trajectories, without any a-priori assumption on the looping character of the trajectories as required by most of the available methodologies. Moreover, the spatial and temporal analysis of the cluster patterns and their joint occurrence matrix showed how the dispersion, although non-homogeneous, seemed to be fairly isotropic. A further important result was to suggest a new analytic expression for the autocorrelation functions, where the exponential decay and the periodic character were included in the same theoretical framework. Three time scales emerged clearly in order to describe the dispersion processes. The Taylor's integral time scale proved to remain valid, however, in the presence of looping trajectories, it must be coupled to the looping period and a longer time scale that describes the long-time exponential decay. These three time scales have a clear signature in the time evolution of the total absolute dispersion.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The relevant data used in the present study are available in a public repository (D. Lagomarsino-Oneto et al., 2023).

Acknowledgments

This research has been partially supported by EU—Next Generation EU Mission 4 “Education and Research” Project IR0000032—ITINERIS—Italian Integrated Environmental Research Infrastructures System—CUP B53C22002150006. A.S. was supported by the Hong Kong RGC-GRF, Grant 15216422.

References

- Azevedo, R., Roja-Solórzano, L. R., & Leal, J. B. (2017). Turbulent structures, integral length scale and turbulent kinetic energy (TKE) dissipation rate in compound channel flow. *Flow Measurement and Instrumentation*, 57, 10–19. <https://doi.org/10.1016/j.flowmeasinst.2017.08.009>
- Berloff, P. S., McWilliams, J. C., & Bracco, A. (2002). Material transport in oceanic gyres. Part I: Phenomenology. *Journal of Physical Oceanography*, 32(3), 764–796. [https://doi.org/10.1175/1520-0485\(2002\)032<0764:mtiogp>2.0.co;2](https://doi.org/10.1175/1520-0485(2002)032<0764:mtiogp>2.0.co;2)
- Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning. *Springer*, 4(4).
- Coppa, S., de Lucia, G. A., Magni, P., Domenici, P., Antognarelli, F., Satta, A., & Cucco, A. (2013). The effect of hydrodynamics on shell orientation and population density of *Pinna nobilis* in the Gulf of Oristano (Sardinia, Italy). *Journal of Sea Research*, 76, 201–210. <https://doi.org/10.1016/j.seares.2012.09.007>
- Coppa, S., Quattrocchi, G., Cucco, A., de Lucia, G. A., Vencato, S., Camedda, A., et al. (2019). Self-organisation in striped seagrass meadows affects the distributional pattern of the sessile bivalve *Pinna nobilis*. *Scientific Reports*, 9(1), 7220. <https://doi.org/10.1038/s41598-019-43214-6>
- Cucco, A., Perilli, A., De Falco, G., Ghezzi, M., & Umgiesser, G. (2006). Water circulation and transport timescales in the gulf of oristano. *Chemistry and Ecology*, 22(1), S307–S331. <https://doi.org/10.1080/02757540600670364>
- Cucco, A., Quattrocchi, G., Satta, A., Antognarelli, F., De Biasio, F., Cadau, E., et al. (2016). Predictability of wind-induced sea surface transport in coastal areas. *Journal of Geophysical Research: Oceans*, 121(8), 5847–5871. <https://doi.org/10.1002/2016jc011643>
- Cucco, A., Sinerchia, M., Lefrançois, C., Magni, P., Ghezzi, M., Umgiesser, G., et al. (2012). A metabolic scope based model of fish response to environmental changes. *Ecological Modelling*, 237, 132–141. <https://doi.org/10.1016/j.ecolmodel.2012.04.019>
- Cucco, A., Sinerchia, M., Ribotti, A., Olita, A., Fazioli, L., Perilli, A., et al. (2012). A high-resolution real-time forecasting system for predicting the fate of oil spills in the strait of bonifacio (western Mediterranean Sea). *Marine Pollution Bulletin*, 64(6), 1186–1200. <https://doi.org/10.1016/j.marpolbul.2012.03.019>
- Davis, R. (1991). Observing the general circulation with floats. *Deep-Sea Research*, 38, S531–S571. [https://doi.org/10.1016/s0198-0149\(12\)80023-9](https://doi.org/10.1016/s0198-0149(12)80023-9)
- De Falco, G., Baroli, M., Cucco, A., & Simeone, S. (2008). Intrabasinal conditions promoting the development of a biogenic carbonate sedimentary facies associated with the seagrass *Posidonia oceanica*. *Continental Shelf Research*, 28(6), 797–812. <https://doi.org/10.1016/j.csr.2007.12.014>
- De Leo, A., Tambroni, N., & Stocchino, A. (2022). Dispersion processes in weakly dissipative tidal channels. *Journal of Geophysical Research: Oceans*, 127(10), e2021JC018315. <https://doi.org/10.1029/2021jc018315>

- Dong, C., Liu, Y., Lumpkin, R., Lankhorst, M., Chen, D., McWilliams, J. C., & Guan, Y. (2011). A scheme to identify loops from trajectories of oceanic surface drifters: An application in the kuroshio extension region. *Journal of Atmospheric and Oceanic Technology*, 28(9), 1167–1176. <https://doi.org/10.1175/jtech-d-10-05028.1>
- Enrile, F., Besio, G., Stocchino, A., & Magaldi, M. G. (2019). Influence of initial conditions on absolute and relative dispersion in semi-enclosed basins. *PLoS One*, 14(7), e0217073. <https://doi.org/10.1371/journal.pone.0217073>
- Farina, S., Quattrocchi, G., Guala, I., & Cucco, A. (2018). Hydrodynamic patterns favouring sea urchin recruitment in coastal areas: A Mediterranean study case. *Marine Environmental Research*, 139, 182–192. <https://doi.org/10.1016/j.marenvres.2018.05.013>
- Ferrarin, C., & Umgiesser, G. (2005). Hydrodynamic modeling of a coastal lagoon: The cabras lagoon in Sardinia, Italy. *Ecological Modelling*, 188(2–4), 340–357. <https://doi.org/10.1016/j.ecolmodel.2005.01.061>
- Griffa, A., Lumpkin, R., & Veneziani, M. (2008). Cyclonic and anticyclonic motion in the upper ocean. *Geophysical Research Letters*, 35(1). <https://doi.org/10.1029/2007gl032100>
- Haigh, M., Sun, L., McWilliams, J. C., & Berloff, P. (2021a). On eddy transport in the ocean. Part II: The advection tensor. *Ocean Modelling*, 165, 101845. <https://doi.org/10.1016/j.ocemod.2021.101845>
- Haigh, M., Sun, L., McWilliams, J. C., & Berloff, P. (2021b). On eddy transport in the ocean. Part I: The diffusion tensor. *Ocean Modelling*, 164, 101831. <https://doi.org/10.1016/j.ocemod.2021.101831>
- Haigh, M., Sun, L., Shevchenko, I., & Berloff, P. (2020). Tracer-based estimates of eddy-induced diffusivities. *Deep Sea Research Part I: Oceanographic Research Papers*, 160, 103264. <https://doi.org/10.1016/j.dsr.2020.103264>
- Holmes, R. M., Zika, J. D., Ferrari, R., Thompson, A. F., Newsom, E. R., & England, M. H. (2019). Atlantic Ocean heat transport enabled by indopacific heat uptake and mixing. *Geophysical Research Letters*, 46(23), 13939–13949. <https://doi.org/10.1029/2019gl085160>
- Jakobsen, P. K., Ribergaard, M. H., Quadfasel, D., Schmith, T., & Hughes, C. W. (2003). Near-surface circulation in the northern North Atlantic as inferred from Lagrangian drifters: Variability from the mesoscale to interannual. *Journal of Geophysical Research*, 108(C8), 3251. <https://doi.org/10.1029/2002jc001554>
- Koszalka, I., LaCasce, J., Andersson, M., Orvik, K., & Mauritzen, C. (2011). Surface circulation in the Nordic seas from clustered drifters. *Deep Sea Research Part I: Oceanographic Research Papers*, 58(4), 468–485. <https://doi.org/10.1016/j.dsr.2011.01.007>
- Koszalka, I. M., & LaCasce, J. H. (2010). Lagrangian analysis by clustering. *Ocean Dynamics*, 60(4), 957–972. <https://doi.org/10.1007/s10236-010-0306-2>
- LaCasce, J. (2008). Statistics from Lagrangian observations. *Progress in Oceanography*, 77(1). <https://doi.org/10.1016/j.pocean.2008.02.002>
- Lagomarsino-Oneto, D., De Leo, A., Stocchino, A., & Cucco, A. (2023). Unraveling the non-homogeneous dispersion processes in ocean and coastal circulations using a clustering approach. [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.10209635>
- Lilly, J. M., & Olhede, S. C. (2009). Bivariate instantaneous frequency and bandwidth. *IEEE Transactions on Signal Processing*, 58(2), 591–603. <https://doi.org/10.1109/tsp.2009.2031729>
- Lilly, J. M., Scott, R. K., & Olhede, S. C. (2011). Extracting waves and vortices from Lagrangian trajectories. *Geophysical Research Letters*, 38(23), L23605. <https://doi.org/10.1029/2011gl049727>
- Lu, Z., & Gan, J. (2023). A modeling study of nutrient transport and dynamics over the northern slope of the South China Sea. *Journal of Geophysical Research: Oceans*, 128(5), e2022JC019225. <https://doi.org/10.1029/2022jc019225>
- Lumpkin, R. (2016). Global characteristics of coherent vortices from surface drifter trajectories. *Journal of Geophysical Research: Oceans*, 121(2), 1306–1321. <https://doi.org/10.1002/2015jc011435>
- Magni, P., Como, S., Cucco, A., De Falco, G., Domenici, P., Ghezzi, M., et al. (2008). A multidisciplinary and ecosystemic approach in the oristano lagoon-gulf system (Sardinia, Italy) as a tool in management plans. *Transitional Waters Bulletin*, 2(2), 41–62.
- Pope, S. B. (2000). *Turbulent flows*. Cambridge University Press.
- Provenzale, A. (1999). Transport by coherent barotropic vortices. *Annual Review of Fluid Mechanics*, 31(1), 55–93. <https://doi.org/10.1146/annurev.fluid.31.1.55>
- Richardson, P. (1993). A census of eddies observed in North Atlantic sofar float data. *Progress in Oceanography*, 31(1), 1–50. [https://doi.org/10.1016/0079-6611\(93\)90022-6](https://doi.org/10.1016/0079-6611(93)90022-6)
- Rupolo, V. (2007). A Lagrangian-based approach for determining trajectories taxonomy and turbulence regimes. *Journal of Physical Oceanography*, 37(6), 1584–1609. <https://doi.org/10.1175/jpo3038.1>
- Taylor, G. (1921). Diffusion by continuous movement. *Proceedings of the London Mathematical Society*, 20, 196–212.
- Umgiesser, G., Canu, D. M., Cucco, A., & Solidoro, C. (2004). A finite element model for the Venice lagoon development, set up, calibration and validation. *Journal of Marine Systems*, 51(1–4), 123–145. <https://doi.org/10.1016/j.jmarsys.2004.05.009>
- Umgiesser, G., Ferrarin, C., Cucco, A., De Pascalis, F., Bellafiore, D., Ghezzi, M., & Bajo, M. (2014). Comparative hydrodynamics of 10 Mediterranean lagoons by means of numerical modeling. *Journal of Geophysical Research: Oceans*, 119(4), 2212–2226. <https://doi.org/10.1002/2013jc009512>
- Veneziani, M., Griffa, A., Reynolds, A. M., Garraffo, Z. D., & Chassignet, E. P. (2005). Parameterizations of Lagrangian spin statistics and particle dispersion in the presence of coherent vortices. *Journal of Marine Research*, 63(6), 1057–1083. <https://doi.org/10.1357/002224005775247571>
- Veneziani, M., Griffa, A., Reynolds, A. M., & Mariano, A. J. (2004). Oceanic turbulence and stochastic models from subsurface Lagrangian data for the northwest Atlantic Ocean. *Journal of Physical Oceanography*, 34(8), 1884–1906. [https://doi.org/10.1175/1520-0485\(2004\)034<1884:otasmf>2.0.co;2](https://doi.org/10.1175/1520-0485(2004)034<1884:otasmf>2.0.co;2)
- Xu, G., Cheng, C., Yang, W., Xie, W., Kong, L., Hang, R., et al. (2019). Oceanic eddy identification using an AI scheme. *Remote Sensing*, 11(11), 1349. <https://doi.org/10.3390/rs11111349>

References From the Supporting Information

- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431–441. <https://doi.org/10.1137/0111030>
- Millot, C. (1999). Circulation in the western Mediterranean sea. *Journal of Marine Systems*, 20(1–4), 423–442. [https://doi.org/10.1016/s0924-7963\(98\)00078-5](https://doi.org/10.1016/s0924-7963(98)00078-5)
- Olita, A., Ribotti, A., Fazioli, L., Perilli, A., & Sorgente, R. (2013). Surface circulation and upwelling in the Sardinia Sea: A numerical study. *Continental Shelf Research*, 71, 95–108. <https://doi.org/10.1016/j.csr.2013.10.011>
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1996). *Numerical recipes in Fortran 90 the art of parallel scientific computing*. Cambridge University Press.

- Quattrocchi, G., Cucco, A., Antognarelli, F., Satta, A., Maicu, F., Ferrarin, C., & Umgiesser, G. (2016). Optimal design of a Lagrangian observing system for hydrodynamic surveys. *Journal of Operational Oceanography*, 9(1), s77–s88. <https://doi.org/10.1080/1755876x.2015.1114805>
- Quattrocchi, G., Simeone, S., Pes, A., Sorgente, R., Ribotti, A., & Cucco, A. (2021). An operational numerical system for oil stranding risk assessment in a high-density vessel traffic area. *Frontiers in Marine Science*, 8, 585396. <https://doi.org/10.3389/fmars.2021.585396>