Unsupervised clustering of oceanic Lagrangian particles: identification of the main pathways of the Labrador Current

M. Jutras[1], N. Planat[2], C. O. Dufour[2], L. C. Talbot[2]

[1]Department of Earth Sciences, McGill University
[2]Department of Atmospheric and Oceanic Sciences, McGill University

Key Points:

- Unsupervised clustering can identify the main pathways in geospatial Lagrangian trajectories.
- The clusters provide information on the properties and origin of the pathways.
- The Labrador Current breaks in an east-west see-saw at the tip of the Grand Banks.

Corresponding author: Mathilde Jutras, `mathilde.jutras@mail.mcgill.ca`

Abstract

Modelled geospatial Lagrangian trajectories are widely used in Earth Science, including in oceanography, atmospheric science and marine biology. The typically large size of these dataset makes them arduous to analyze, and their underlying pathways challenging to identify. Here, we show that a Machine Learning unsupervised k-means++ clustering method can successfully identify the pathways of the Labrador Current from a large set of modelled Lagrangian trajectories. The presented method requires simple pre-processing of the data, including a Cartesian correction on longitudes and a PCA reduction. The clustering is performed in a kernalized space and uses a larger number of clusters than the number of expected pathways. During post-processing, similar clusters are grouped into pathway categories by experts in the circulation of the region of interest. We find that the Labrador Current mainly follows a westward-flowing and an eastward retroflecting pathway (20% and 50% of the flow, respectively) that compensate each other through time in a see-saw behaviour. These pathways experience a strong variability of up to 96%. We find that two thirds of the retroflection occurs at the tip of the Grand Banks, and one quarter at Flemish Cap. The westward pathway is mostly fed by the on-shelf branch of the Labrador Current, and the eastward pathway by the shelf-break branch. Pathways of secondary importance feed the Labrador Sea, the Gulf of St. Lawrence through the Belle Isle Strait, and the subtropics across the Gulf Stream.

Plain language summary

Lagrangian trajectories, in which we follow a parcel of water or air parcel as it is moved around by currents, are widely used in Earth Science, including in oceanography, atmospheric science and marine biology. They typically come in very large and chaotic dataset, from which it is difficult to identify the main pathways of a flow. Here, we use a Machine Learning based algorithm, more specifically an unsupervised clustering algorithm, to identify the main pathways of the Labrador Current based on a large set of Lagrangian trajectories obtained from an ocean model. This study shows the power of such a method to help analyze this type of data, and provides a detailed description of the applied recipe so it can be used by people in the field. We find that, when it reached the Grand Banks of Newfoundland, most of the Labrador Current flows either westward towards the Slope Sea or eastward towards the North Atlantic Ocean, in a see-saw behaviour. We also identify a previously unknown minor pathway that brings Labrador Current waters south of the Gulf Stream front.

1 Introduction

Lagrangian trajectories are diagnostics that are widely used across climate sciences. Such trajectories are obtained from the positioning of observational platforms such as drifting floats in the ocean (e.g. Argo floats, surface drifters, RAFOS floats) and balloons in the atmosphere, as well as from the advection of virtual particles derived from velocity fields reconstructed from satellite altimetry or output from numerical simulations (among others, A. F. Thompson & Sallée, 2012; van Sebille et al., 2018). Lagrangian trajectories are used to study ocean and atmospheric circulations (e.g., Schulze Chretien & Frajka-Williams, 2018; Gillard et al., 2016; Bower et al., 2011; Fischer & Schott, 2002) and sea ice drift (e.g., Williams et al., 2016; Brunette et al., 2019), to identify the origin and fate of water masses (e.g., Kawasaki et al., 2022; Kelly et al., 2019), to assess connectivity timescales (e.g., Jönsson & Watson, 2016), and to study the fate of atmospheric and oceanic pollutants (e.g., Hertwig et al., 2015; Viikmäe et al., 2013), plastic (e.g., Lebreton et al., 2012), larvae (e.g., Ayata et al., 2010; Cetina-Heredia et al., 2015; Phelps et al., 2015; Simons et al., 2013), icebergs (e.g., Marson et al., 2018; Merino et al., 2016), and debris or people during search and rescue (e.g., Hart-Davis & Backeberg, 2021). Yet, sets of Lagrangian trajectories are challenging to analyze. It is often not possible to clearly distinguish pathways given the chaotic nature of geophysical flows, which generally prevents the use of simple and objective criteria to produce classification. In oceanography, traditional classification methods of Lagrangian trajectories are based on counting particles crossing sections based on hydrography (Jutras et al., 2023; Daher et al., 2020; Merino et al., 2016; Bower et al., 2011), topography, or dynamic water properties (e.g. on fronts, Roach & Speer, 2019; Schulze Chretien & Frajka-Williams, 2018). Apart from passages, straits or other clearly defined topographic features that provide non-ambiguous physical boundaries for the flow, criteria used for classification of trajectories often appear adhoc or subjective (Fig. 1b). Besides, in modelling studies, dataset typically reach up to millions of trajectories, making visual inspection overwhelming and non-efficient.

Machine Learning (ML) offers several algorithms that can help analyze (extremely) large and complex Lagrangian datasets. Here, we consider clustering algorithms, which automatically classify objects into "clusters", or groups of elements with similar properties. *Supervised* clustering is trained on a pre-classified dataset, which can be obtained, e.g., based on visual inspection. These types of methods are useful when the classification is already known or obvious to the human eye. On the other hand, *unsupervised* clustering lets the algorithm identify the clusters itself, removing potential biases in the choice of classes. Unsupervised clustering has already successfully been applied to vertical profiles from Conductivity-Temperature-Depth sensors (CTD; Boehme & Rosso, 2021) and Argo floats (Houghton & Wilson, 2020; Rosso et al., 2020; Jones et al., 2019), to radar data (Tiira & Moisseev, 2020), to cyclones tracks (Kremer et al., 2020) and to air pollutants (Brankov et al., 1998), as well as to identify mean flows (Koszalka & Lacasce, 2010), ocean fronts (Thomas et al., 2021), and finite-time (couple of days) coherent structures in a flow (Filippi, Hadjighasem, et al., 2021; Wichmann et al., 2021; Schneide et al., 2018), to name a few. The above-mentioned studies use various ML clustering methods, including Gaussian-Mixture Models (Boehme & Rosso, 2021; Rosso et al., 2020; Jones et al., 2019; Thomas et al., 2021), k-means (Houghton & Wilson, 2020; Kremer et al., 2020; Schneide et al., 2018; Koszalka & Lacasce, 2010), optimized-parameter spectral methods based on k-means (Filippi, Hadjighasem, et al., 2021; Filippi, Rypina, et al., 2021), hierarchical clustering, and density-oriented clustering like DBSCAN (Wichmann et al., 2021). Yet, to our knowledge, no study has applied unsupervised clustering to large-scale (more than a couple of days) geophysical Lagrangian trajectories, nor used such a method to identify the main pathways of a geophysical flow. This technique appears especially suitable to identify and characterize the pathways of an ocean current, removing the subjectivity inherent to more traditional methods mentioned earlier.

In this study, we use such a method to study the Labrador Current (Fig. 1a). The Labrador Current is a western boundary current. It forms the western limb of the subpolar gyre, and as such is a critical component of the North Atlantic circulation. The Labrador Current is composed of an inshore and a shelf-break branch that flow south on and along the Labrador shelf, respectively (Florindo-López et al., 2020; Loder et al., 1998; Lazier & Wright, 1993), until the tip of the Grand Banks. Eventually, most of the Labrador Current Water is entrained into the subpolar North Atlantic by the North Atlantic Current (NAC) and the remainder follows the continental shelf southwestward (Fig. 1a; Townsend et al., 2015; Fratantoni & McCartney, 2010; Pérez-Brunius et al., 2004). By doing so, the Labrador Current carries cold, relatively fresh and well-oxygenated waters from the subarctic to both the subpolar North Atlantic and to the Slope Sea and eastern American continental shelf. Variability in the strength and exact path of the Labrador Current therefore affects the water properties in both regions (e.g., Jutras et al., 2023; Gonçalves Neto et al., 2021; Chen et al., 2020; Holliday et al., 2020; Claret et al., 2018; B. D. Petrie & Drinkwater, 1993) and in connected bodies of water such as the Gulf of St. Lawrence Estuary (Jutras et al., 2020; Gilbert et al., 2005; Han et al., 1999) and the Gulf of Maine (Whitney et al., 2022; Pershing et al., 2016), with direct consequences on marine ecosystems (Poitevin et al., 2019; Chabot & Dutil, 1999) and fisheries (Pershing et al., 2016; Mills et al., 2013). Some Labrador Current Waters also leak into the Labrador

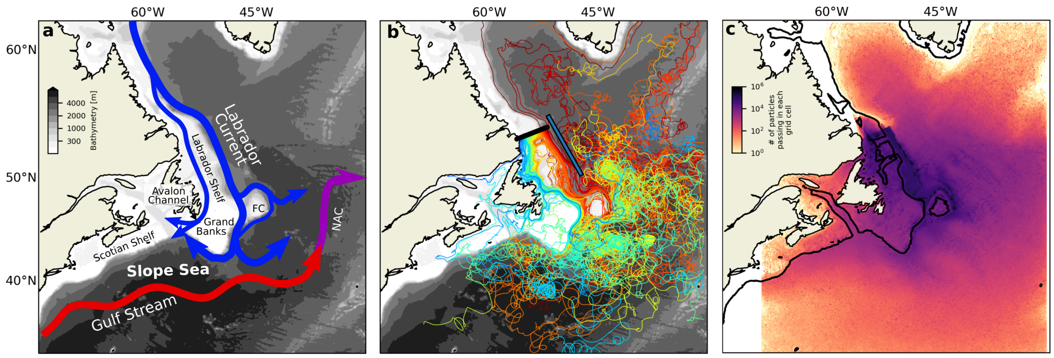


Figure 1. **(a)** Region of the Labrador Current. The arrows show the approximate location of the main currents of the region. The main topographic and oceanographic features are labelled. FC indicates Flemish Cap. NAC indicates the North Atlantic Current. **(b)** Example subset of Lagrangian trajectories launched in the Labrador Current. The colour indicates the distance from the shore at initialization. The thick black line indicates the launch section. The blue line indicates the entry point to the Labrador Sea, used in section 3.2.1. **(c)** Probability density plot of the complete dataset of Lagrangian trajectories. We stop tracking particles east of 50E.

Sea (Schulze Chretien & Frajka-Williams, 2018; Howatt et al., 2018; Palter et al., 2008; Myers, 2005), possibly affecting stratification and modulating deep water formation and the Atlantic Meridional Overturning Circulation (AMOC; New et al., 2021). Still, little is known about the spatio-temporal characteristics, magnitude and drivers of these pathways, or about other possible pathways (Jutras et al., 2023; Fratantoni & McCartney, 2010).

This paper uses the Labrador Current as a case study to demonstrate that unsupervised clustering can be used to identify pathways in geophysical Lagrangian tracks. To do so, we implement an unsupervised kmeans++ clustering method on a large set of Lagrangian trajectories within the Labrador Current (Fig. 1b), to identify and characterize the main and secondary pathways of this current. The trajectories are almost impossible to distinguish through traditional methods, being continuously distributed in the western North Atlantic (Fig. 1b,c). Section 2 presents a step-by-step description of our method intended for non-experts, hoping that this case study can inspire applications in other Earth system contexts. Section 3.1 presents the results of the clustering, including the identification of the pathways, a characterization of their properties, and quantitative comparisons with the literature. Section 3.2 uses the classification of the complete dataset to look at the variability of each pathway of the Labrador Current on seasonal and decadal time scales. Section 4 offers some concluding remarks on the method and the results.

2 Methods

2.1 Lagrangian trajectories

The Lagrangian trajectories are generated from virtual particles advected offline by 3D velocity from the GLORYS12V1 ocean reanalysis (Lellouche et al., 2018). GLORYS12V1 is based on the NEMO3.1 modelling platform (Madec et al., 2019). It has a spatial resolution of 1/12° on an ORCA grid and 50 levels in the vertical, with thicknesses ranging from 0.5 m at the surface to 160 m at a 1 km depth and with 18 levels in the top 50 m. The simulation covers the 1993 to 2018 period and is forced with the ERA-Interim atmospheric reanalysis (ECMWF Re-Analysis, Dee et al., 2011).

The virtual particles are tracked with the OceanParcels tool for Python (Probably A Really Computationally Efficient Lagrangian Simulator; Delandmeter & Van Sebille, 2019). We use the daily horizontal velocity outputs on a longitude-latitude grid provided on the Copernicus Marine Service (CMS) website. The vertical velocities are reconstructed from sea surface height. Particles are seeded every 1/12° along the (53°N, 56.7°W) – (54.3°N, 52.0°W) line (Fig. 1b) and every 10 m in the vertical, in waters with a salinity lower than 34.8, for a total of 966 particles per seeding event. The salinity cut-off is used to delineate the Labrador Current from the Labrador Sea (Myers, P., *personal communication*; Loder et al., 1998). Particles are released every week from January 1^{st} 1993 to January 1^{st} 2015 and are tracked with a 10-minute time step. The complete data set contains 1.2 millions trajectories. The seeding temporal and spatial frequencies are chosen so that increasing the number of particles does not change their general distribution downstream, optimizing the use of computational resources (van Sebille et al., 2018). We stop tracking the particles when they hit topography or the boundaries of the domain (Fig. 1c) or after 550 days, time after which they have left the Labrador Shelf and reached their final export zone. These experiments are also described in Jutras et al. (2023), who look at the variability in the retroflection of the Labrador Current. In addition to the position and depth of the particles, we track their temperature, salinity and age since release. We use a purely advective scheme. Tamsitt et al. (2017) showed that the addition of turbulent diffusion did non affect Lagrangian trajectories significantly in eddy-resolving models. In addition, there is no consensus on a realistic value for diffusive coefficients, especially when covering both coastal and open ocean areas (van Sebille et al., 2018).

2.2 Observational dataset

We compare the trajectories of the virtual particles with those from actual observational platforms, namely surface drifters, Argo floats and RAFOS/SOFAR floats. A direct comparison is not possible because virtual particles can move vertically, while floats and drifters flow at a fixed depth. In addition, most of the Argo and RAFOS/SOFAR floats drift deeper than the virtual particles, more specifically into the Deep Western Boundary Current. We therefore expect the trajectories to differ, in particular where the Labrador Current waters dive as they interact with the Gulf Stream – NAC front. Still, we use the observations to validate qualitatively the simulated pathways, as well as to offer a rough comparison of the magnitude of each pathway.

We use surface drifters deployed as part of the Global Drifter Program. These satellite-tracked buoys drift at the surface of the ocean and are equipped with 15 m or 1 m drogues. We consider the floats that are carried by the Labrador Current by selecting the ones that cross the virtual particles seeding line and that enter the Grand Banks area, as defined by the (55°W; 41°W) – (45°N; 50°N) box (Fig. 6). Based on this criterion, we identify 79 drifters from 2000 to 2018.

Argo floats are autonomous profilers that drift passively with ocean currents at a parking depth (typically 1 km) and profile temperature, salinity and pressure down to approximately 2 km every 10 days. RAFOS/SOFAR floats are autonomous platforms that drift at a fixed depth between 500 m and 1 km. We select the floats based on the same criteria as for the surface drifters, except that we extend the seeding line and the box offshore by two degrees (Fig. 6) to account for the fact that floats drift deeper over the continental slope. We identify 64 Argo floats fitting these criteria between 2001 and 2019 and 50 RAFOS/SOFAR floats between 2003 and 2007.

A visual inspection suggests that the pathways of observational platforms and of virtual particles generally agree (Fig. 1b and 6). The small number of drifters and floats rules out applying a clustering algorithm to their trajectories. Hence, we manually classify the platforms into pathways using the following hydrographic sections (Fig. 6):

- Westward-flowing: crosses the 54^{th} meridian south of the Grand Banks;
- Westward then retroflected: crosses the 54^{th} meridian south of the Grand Banks and eventually drifts eastward;
- Retroflecting: enters the zone from 0°W to 60°W and from 47°N to 65°N;
- Southward-flowing: enters the zone from 54°W to 35°W and from 35°N to 47°N.

2.3 Clustering algorithm, step by step

2.3.1 Overview

Machine Learning unsupervised clustering algorithms build a classification model that attributes each object (here, trajectories) to a cluster. The model is characterized by parameters called *hyperparameters* that can include, for instance, the number of transformations applied to the data, the number of clusters, or criteria on the within-cluster maximal distance. Three independent data subsets are used to feed the model, namely the training, validation and test sets. These sets must be large (at least hundreds of objects) and of high quality (e.g. evenly sampled or without missing values). The training set is used to train the model, which is validated with the validation set for a range of hyperparameter values. By comparing the results with performance metrics, the most performant hyperparameters values are determined. Once the model is ready, its performance is validated with the test set. To avoid overfitting the model to the subsets, the test set must be used only once, to validate the final results. Overfitting would lead to a model that offers a good classification of the training subset, but not of new data. Finally, once the model is ready, it can be applied to the complete dataset or to new dataset. An overview of the method is presented in Figure 2.

2.3.2 Pre-processing

Before building this model, we need to prepare the data. Since the goal of the study is to identify the various pathways of the Labrador Current as it flows over and along the Labrador Shelf, we are interested in the shape of the Lagrangian trajectories. We therefore base our classification on latitude and longitude coordinates. Additional variables (temperature, salinity and depth) were also considered to be used in the clustering algorithm, but showed no significant improvement on the classification results. We build the clustering model with a subset of 100 000 trajectories out of a total 1.2 millions (Fig. 2). These trajectories are selected randomly every four years, as preliminary analyses showed no periodicity in the preferred pathways over that timescale. This subset is further separated into an 80 000 particle training set, a 10 000 particle validation set, and a 10 000 particle test set. While there is no universal rule on the number of objects required in each set, an 0.8-0.1-0.1 ratio is commonly used.

We apply the following pre-processing to each set (Fig. 2):

- To avoid a bias by which the particles would be clustered based on their initialization location, we translate all the particles to the same starting point. This *translation* step increases the efficiency of the clustering (not shown).
- Trajectories shorter than 550 days – e.g. due to the particles reaching the bottom of the ocean, the shore, or the boundaries of the domain – are filled with zeros.
- To account for the sphericity of the Earth, we apply a longitudinal correction. The particles flow approximately from 54°N, where one degree of latitude represents 65 km, to 30°N, where one degree of latitude represents 96 km. Because we are interested in the shape of the trajectory in a Cartesian space (km) but operate the classification in latitude-longitude space, we apply a "$\cos\lambda$" weight to the longitudes, where λ represents the latitude. The resulting Euclidian distance in modified latitude-longitude space offers a good approximation of the real (physical) Cartesian distance at the surface of the ocean.

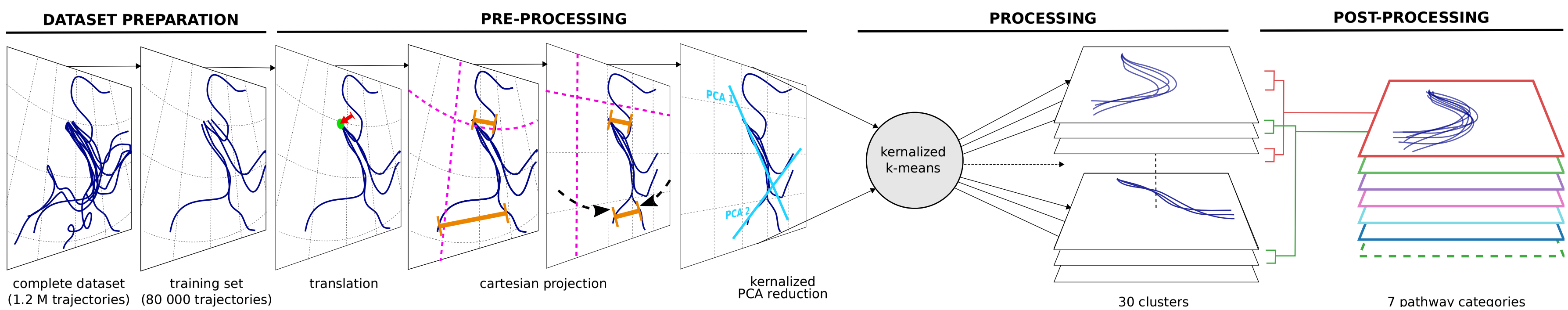


Figure 2. Schematic diagram of the unsupervised clustering method. A detailed description of the method is provided in Section 2.3. The 30 clusters are presented in Fig. 4 and the seven pathway categories are presented in Fig. 5.

- To lower the computational cost of the analysis (the training set has a size of 80 000 × 550 locations × 2 coordinate variables), it is common practice to reduce the number of features by implementing a Principal Component Analysis (PCA). By keeping the features responsible for 99.99% of the total variance, we reduce the dataset to 12% of its original size. We implement the PCA and the k-mean clustering (see Section 2.3.3) in a kernalized space, i.e. a transformed variable space. A kernel helps convergence when using linear classifiers on non-linear data, by implicitly adding non-linearities to the algorithm (Hofmann et al., 2008). Among the tested kernels (sigmoid, polynomial, cosinus), the cosinus transformation, defined as follows, led to the most efficient clustering:

$$k(x,y) = \frac{xy^T}{||x|| \cdot ||y||} \tag{1}$$

  where x and y are the vectors containing the variables (here, the coordinates of the trajectories).

Computing the kernalized k-means and kernalized PCA requires high RAM, as large matrices need to be temporarily loaded. In our case, the computation takes about one day on a HPC system with 186 GB of RAM. In a non-kernalized space, the clustering algorithm could run on a regular work station.

2.3.3 Processing

We here apply a k-means++ clustering algorithm (Fig. 2), which is common, easy to implement, and requires only one hyperparameter: the number of clusters. The k-means method classifies the data by minimizing the within-cluster variance of the Euclidean distance between each object. More specifically, each cluster is characterized by a centroid, or mean vector, to which the distance with each object belonging to that cluster is minimized. In the k-means++, the spread between the initial centroids is maximized by testing multiple initializations and keeping the one offering the best classification, significantly improving the convergence and speed compared to the traditional k-means method. We here implement 20 random initializations. To accelerate the convergence of the classification itself, we then fold it 15 times: we randomly split the dataset in 15 pieces, iteratively apply the classification to 14 pieces and evaluate the results on the 15th. The results are not sensitive to a higher number of folds or initializations. These steps are implemented using the k-means++ functions of the Python scikit-learn package (`scikit-learn.org/`).

For the value of the hyperparameter, namely the number of clusters, prior knowledge of the circulation of the Labrador Current suggests two major pathways plus some minor ones (see Section 1). The k-means method has difficulty converging in the presence of clusters of unequal sizes (i.e. containing unequal number of objects). Using a large number of clusters and grouping them afterwards has been shown to improve the performance of the classification (Echols et al., 2020) and helps reveal secondary pathways. To find the optimal number of clusters, we use two performance metrics: the silhouette score (Rousseeuw, 1987), and a physics-based metrics that is adapted to our scientific question. The silhouette score measures the overall performance of the clustering algorithm based on the intra and inter cluster distances. This metric is expected to monotonously decrease with the number of clusters, since a higher number of clusters necessarily improves the performance on average (i.e. the intra cluster spread decreases as the number of objects per clusters decreases). The number of clusters can therefore be chosen based on a stabilization of this score (Fig. 3a). We then define a physics-based metric that evaluates the spatial and temporal coherence of the particles. We first define five regions that the particles are likely to visit (Fig. 3b). For each cluster, we identify the most popular region at each time step, and compute the fraction of particles found in that region. This provides a score for each cluster, between 0 and 1. We then average

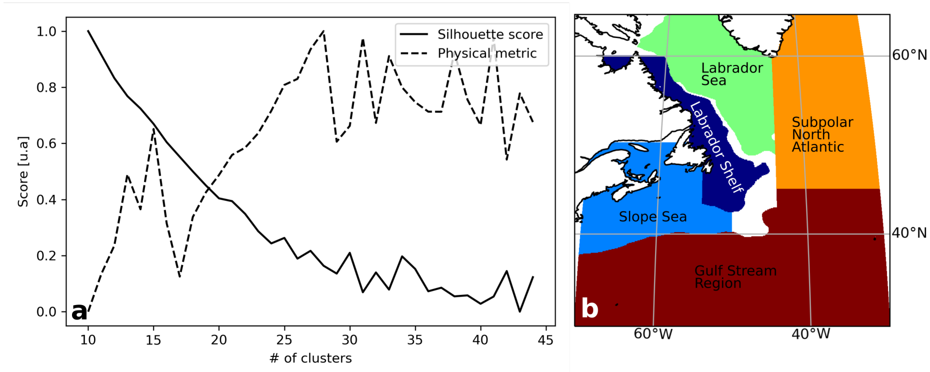


Figure 3. **(a)** Algorithm's performances for a varying number of clusters, based on silhouette and a physics-based scores. **(b)** Regions used for the physics-based performance metric.

the scores of all clusters to obtain a global score. The model's performance is highest when this metric is maximized, indicating that a high number of particles simultaneously visit the same region. Both the numerical and physics base metrics show a plateau around 30 clusters (Fig. 3a).

2.3.4 Post-processing

As expected, a visual inspection of the obtained clusters reveals some redundancies in the pathways represented in some clusters (Fig. 4). While, as we will show in Section 3.2.3, most of the clusters which look alike actually present differences that are not visible in the particle trajectories, it is useful to group the clusters identifying similar pathways. Based on the shape of the trajectories and on their export location, we visually identified six pathway categories which are described in detail in Section 3.1.1. The identification of these categories was nourished by prior knowledge of the circulation discussed in the literature (Section 3.1.3). To avoid biases in the categorization, we invited eight experts of the northwestern Atlantic circulation to sort the 30 clusters into the six identified pathway categories or to new ones they would discern. The experts overall agree on the classification (see Appendix B for details).

3 Results

3.1 Pathways

The unsupervised clustering method successfully classifies the trajectories into 30 clusters showing similar trajectories (Fig. 4), which were combined into six pathway categories (Table 1 and Fig. 5). Note that throughout this section, we display results from the test set. We first describe how we identified the pathway categories (section 3.1.1) before assessing the qualitative agreement with the scarce observations (section 3.1.2) and with the literature (section 3.1.3 and 3.1.4).

3.1.1 Definition of the pathway categories

From the 30 clusters, we identify two main pathway categories: particles retroflected towards the subpolar North Atlantic (henceforth referred to as *retroflected*), and particles following the shelf westward into the Slope Sea and along the eastern American continental shelf (henceforth referred to as *westward-flowing*, Fig. 4 and 5). These pathways account in total for respectively 48% and 21% of all the trajectories (Table 1), meaning that close to 70% of the water from the Labrador Current feeds either the subpolar North Atlantic or the Slope Sea and eastern American continental shelf. Less than 1% of the particles first enter the Slope Sea before retroflecting towards the subpolar North Atlantic (cluster #22 on Fig. 4). We include these particles in the westward-flowing category, as they first affect the water properties of the Slope Sea and have lost most of their Labrador Current water signature once they retroflect. Another significant pathway category comprises the particles that are killed on the Labrador Shelf as they hit the seafloor (22% of the trajectories; referred to as *Labrador Shelf*). This category does not represent a real pathway, and is rather an artifact of the virtual Lagrangian tracking. We also identify three secondary pathway categories: the particles that travel southward from the tip of the Grand Banks (8%; referred to as *southward-flowing*), the ones that enter the Gulf of St. Lawrence through the Belle Isle Strait (1%; referred to as *Belle Isle*), and the ones that feed the Labrador Sea (<1%; referred to as *Labrador Sea*, Fig. 4 and 5).

The uncertainty on this classification comes from two sources: the clustering algorithm itself (algorithm uncertainty), and the categorization of clusters (human-induced uncertainty). The two are not independent, since a large algorithm error will lead to disagreement in the experts' classification. First, the algorithm error manifests as trajectories that are classified into a cluster even if, from a visual inspection, they would have fitted better in another. For instance, cluster #17 belongs to the Labrador Shelf pathway category, but a few particles still reach the Scotian Shelf, and should have been classified in a cluster belonging to the westward-flowing pathway category. There currently exists no widely accepted method to evaluate the error from unsupervised clustering algorithms (e.g., Abdar et al., 2021; Kläs & Vollmer, 2018). We cannot use the within-cluster spread to assess the algorithm's error, because particles can end up quite far from each other but still belong to the same cluster (e.g., particles retroflecting eastward can reach from 30N to 55N). Hence, we simply report the algorithm's performance based on the physical metric presented in Section 2.3. We find that the score is high for all (> 0.7) but some Labrador Shelf clusters, in which a few particles enter the Belle Isle Strait, and for the Belle Isle cluster, in which a few particles flow along the Scotian Shelf (Fig. A3).

Second, for the errors in the categorization of clusters, we find that the experts are almost unanimous in classifying the clusters in the Belle Isle, southward-flowing and Labrador Sea pathway categories. For other categories, the error ranges between 7 and 10% (Table 1, see also Appendix B and Table B1). Overall, the errors appear sufficiently small to go forward with the analysis of the results.

3.1.2 Comparison against trajectories of observational platforms

We perform a visual comparison between the obtained pathways and that of Argo floats, RAFOS/SOFAR floats, and surface drifters (see Section 2.2). We find that the retroflected, westward-flowing and southward-flowing pathways clearly appear in the trajectories of autonomous platforms (Fig. 6). There is also a significant amount of platforms going westward and then retroflecting, more than in the virtual particles. We do not expect any observational platforms to follow the Labrador Sea pathway because our selection criteria filter out these platforms (section 2.2). The same holds for the Labrador Shelf pathway, which is an artifact of the virtual Lagrangian tracking, and for the Belle Isle pathway, since no autonomous platforms have been launched within the coastal current that feeds this strait. The agreement in the pathways provides confidence in the clustering. The bulk sizes of the observed and modelled pathway categories generally agree (Table 1), keeping in mind that (i) there are too few observational platforms to allow a statistically robust comparison and (ii) observational platforms drift at a fixed depth while Lagrangian particles can move vertically (see Section 2.2), and (iii) the on-shelf category, composing >20% of the virtual particles, is an artifact of the Lagrangian tracking and is absent from the observations. Note that most of the surface drifters retroflect eastward (Fig. 6).

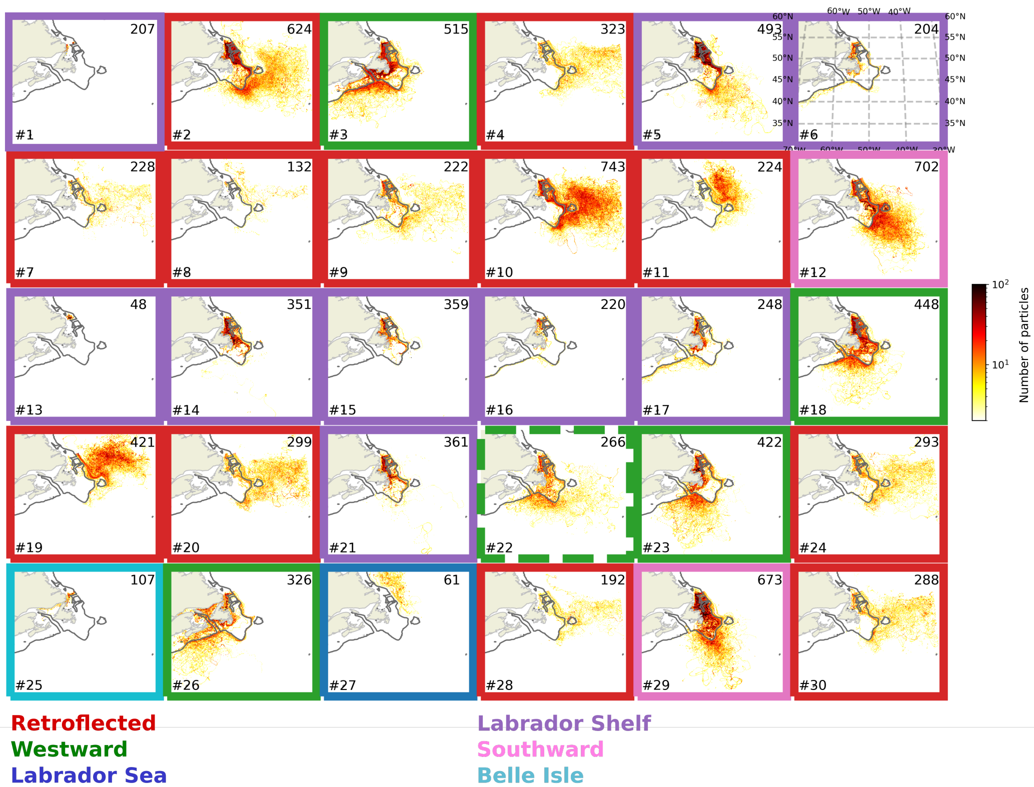


Figure 4. Density map of the trajectories for each of the 30 clusters identified by the k-means++ model for the test set (see Section 2.3.3). The numbers on the top right indicate the number of particles (or trajectories) in each cluster, while the numbers on the bottom left correspond to the cluster identification number. The dark grey line shows the 350 m isobath. The coloured frames indicate in which pathway category the cluster is classified by the experts: retroflected (red); westward-flowing (green); Labrador Shelf (purple); Labrador Sea (blue); southward-flowing (pink); Belle Isle (cyan). The westward-flowing cluster with a dashed contour contains particles that go westward first and are then retroflected. See Section 3.1.1 for a description of each pathway category.

Pathway category	**Cluster ID**	**Percentage: mean** (min-max)	**Expert's error**	**% of observations**
Retroflected	2, 4, 7*, 8*, 9, 10, 11*, 19*, 20, 24, 28*, 30	47.6 % (24.9-73.7)	10 %	74 %
Westward-flowing	3, 18, 22, 23, 26	21.0 % (5.4-42.2)	6 %	10 %
Westward then retroflected	22	0.6 % (0.0-2.2)	7 %	4 %
Labrador Sea	27	0.4 % (0-3.2)	0 %	-
Labrador Shelf	1, 5, 6, 13, 14, 15, 16, 17, 21	21.8 % (12.7-31.3)	9 %	-
Southward-flowing	12, 29	7.8 % (4.2-13.0)	0 %	16 %
Belle Isle	25	1.4 % (0.0-6.5)	0 %	0 %

Table 1. Classification of the 30 clusters into the six pathway categories (see Section 3.1.1 for a description of each pathway category). The first column indicates the name of the pathway category; the second column indicates the identity number (ID) of the clusters classified within that category (see Fig. 4 for the IDs); the third column indicates the mean percentage of trajectories classified into a given category, computed from the complete dataset, as well as the lowest and highest percentage over the 1993 to 2018 period; the fourth column indicates the error coming from the disagreement in the experts' categorization; the last column indicates the percentage of observational platforms corresponding to each category (see Section 2.2). In the retroflected category, the clusters marked with an asterisk retroflect at Flemish Cap while the others retroflect at the tip of the Grand Banks.

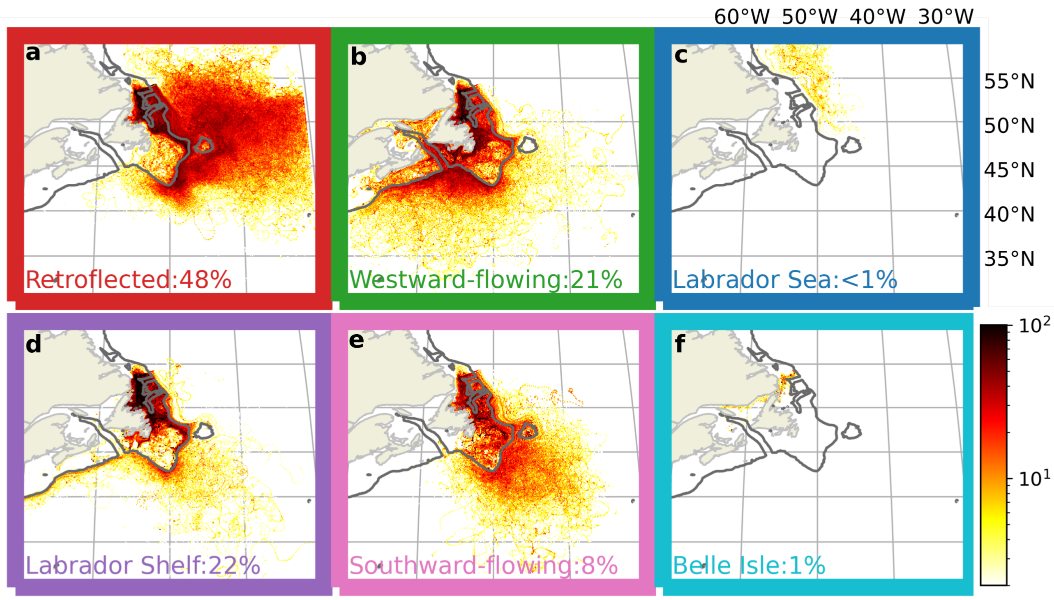


Figure 5. Density maps for each of the six pathway categories of the Labrador Current (see Section 2.3.4). See Figure 4 for a detailed description of the plot. The percentage provides the average magnitude of each pathway category.

3.1.3 Validation of pathway categories against the literature

We compare the relative importance of each pathway category with results from previous studies. First, many studies also report that the retroflected and westward-flowing pathways are, respectively, the main and secondary pathways for the Labrador Current (e.g., Gonçalves Neto et al., 2023; Holliday et al., 2020; Han et al., 2014; Fratantoni & McCartney, 2010). Fox et al. (2022) also observed that modelled Lagrangian trajectories can retroflect after having flowed westward. Our estimate of the Labrador Current export towards the Labrador Sea (0 - 3%, Table 1) is in good agreement with observation-based studies (0 - 3%; Howatt et al., 2018; Schmidt & Send, 2007) and with model-based studies, (6 - 8%; Myers, 2005). We expect an underestimation, because the above-mentioned studies focus on the shelf-break branch of the Labrador Current, while we also consider the inshore, on-shelf branch of the current (∼15% of the volume transport). The inflow of water through the Belle Isle strait has been estimated to range from 0.1 Sv in the spring to 0.4 Sv in the winter, based on observations (Shaw & Galbraith, 2023; B. Petrie et al., 1988), and from 0.15 Sv to 1 Sv during winter storms, based on a model (Saucier et al., 2003). Relative to the mean 8.1 Sv Labrador Current volume transport found in GLORYS12V1, this represents 1 - 12% of the current, in broad agreement with the results of the clustering (0 - 7%). The southward-flowing pathway has not been explicitly described as a Labrador Current pathway in the literature. However, such a pathway has been described for the Deep Western Boundary Current (DWBC), which exports Labrador Sea Waters equatorward below ∼1500 m (Bower et al., 2009). The virtual particles that follow the southward-flowing pathway first sink to a depth of ∼1000 m (see Fig. 11), thus reaching the upper limb of the DWBC (Handmann et al., 2018). This pathway could therefore emerge from interactions between the two currents. Overall, the relative importance of each pathway obtained from the clustering agrees well with previously model-based and observation-based estimates, further supporting the method.

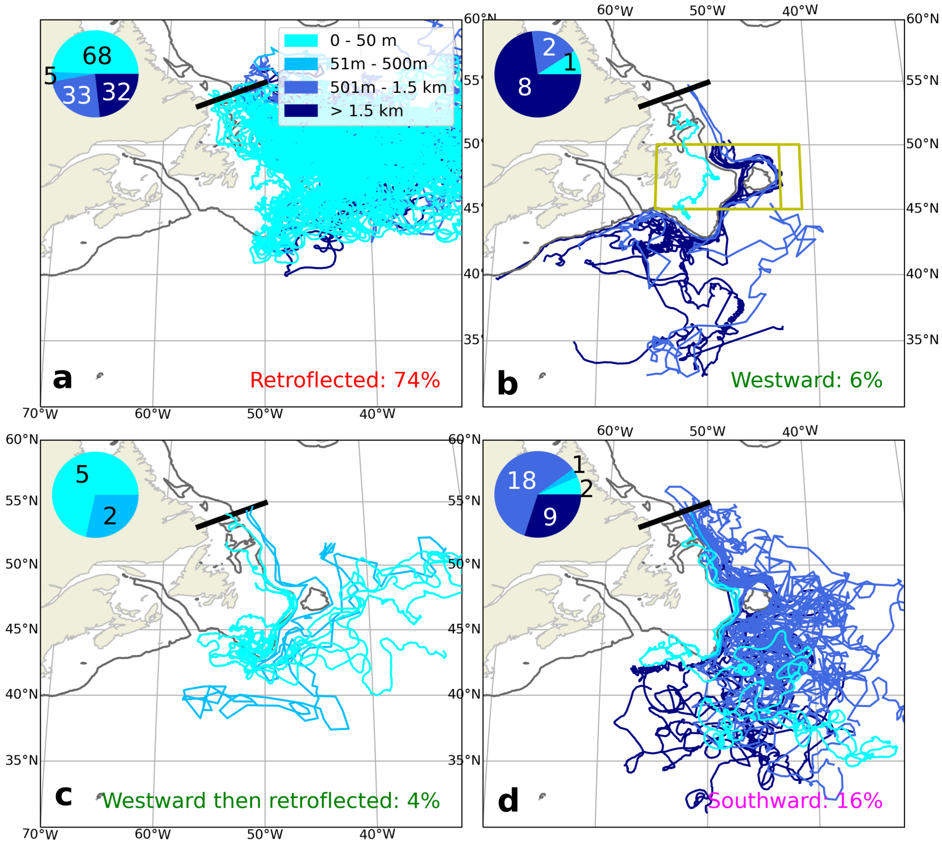


Figure 6. Observational platforms drifting within the Labrador Current between year 2000 and 2018 – a time period that overlaps with the GLORYS12V1 reanalysis period – sorted into four of the pathway categories identified from the clustering algorithm (see Section 2.3.4). The name of the pathway category is indicated at the bottom right, along with the percentage of platforms classified in this category. The colours of the trajectories refers to the drifting depth of the platform. For each panel, a pie chart represents the fraction of platforms in depth classes, with the corresponding number of platforms indicated within each class. The grey contour delineates the 350 m isobath. The black straight line and the yellow boxes indicate the criteria used to select the platforms of interest (see section 2.2 for further details).

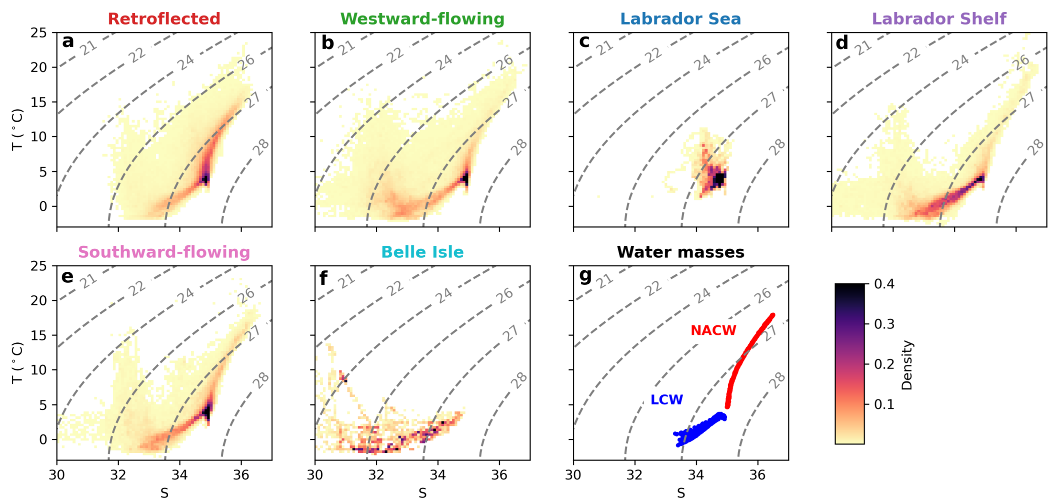


Figure 7. (a-f) Potential temperature – salinity diagrams of all the virtual particles classified according to pathway categories. The color shading represents the density of particles for a given T-S combination.(g) Potential temperature – salinity diagram of the two main water masses interacting in the region of interest: the North Atlantic Central Waters (NACW) originating from the Gulf Stream, and the Labrador Current Waters (LCW) formed by the Labrador Current flowing southward on the Labrador Shelf. Data for these signatures come from the World Ocean Circulation Experiment (WOCE) climatology (Gouretski, 2018). The LCW is defined as the waters lying between 150 m and the seafloor on the Labrador Shelf and slope, and the NACW as the waters lying between 250 and 1300 m within the Gulf Stream jet (see Jutras et al., 2020). The dashed grey lines show isopycnals (+1000 kg m^{-3}).

3.1.4 Thermohaline signature across pathways

We further verify the ability of the clustering algorithm to properly classify the trajectories by comparing the thermohaline properties of each category with what is expected for these pathways. The Labrador Sea category clearly and almost exclusively shows the signature of Labrador Current Waters (Fig. 7), which makes sense since these waters leave the continental shelf before any contamination can occur (see section 3.2.2). All the other categories show the signature of the LCW getting fresher as they receive river outflow along the Labrador coast. The Belle Isle category contains only the LCW. In addition to the signature of the LCW, the retroflected, westward-flowing and southward-flowing categories show that of the warm and salty North Atlantic Central Waters (NACW), which progressively mix with the LCW along the Labrador Current – NAC front. In the westward-flowing category, we only find the signature of the coldest, freshest NACW, as the contact time with the NAC is shorter than for the retroflected category. Most of the pathways show the additional weak signature (few trajectories) of warm and fresh waters formed on the Labrador Shelf during the summer. The fact that each pathway category has a thermohaline signature that fits with what can be expected from the circulation supports the algorithm and our choice of categories.

3.2 Spatio-temporal characteristics of the pathway categories

Now that the different pathways of the Labrador Current are identified and validated, we document their temporal evolution (section 3.2.1) and spatial characteristics (sections 3.2.2, 3.2.3 and 3.2.4). To do so, we use the classification of the complete dataset (1.2 millions trajectories) from 1993 to 2018.

3.2.1 Temporal variability in the pathways

The classification provides time series of the relative importance of each identified pathway (Fig. 8). The relative importance of the westward-flowing pathway from 1996 to 1998 fits with a salinification of the subpolar North Atlantic (Holliday et al., 2020), while that of the retroflected pathway since 2011 coincides with a strong freshening of the subpolar North Atlantic (Holliday et al., 2020), a warming of the eastern American continental shelf (Chen et al., 2020), and a deoxygenation of the western North Atlantic and adjacent basins (Jutras et al., 2020; Claret et al., 2018).

We find that the two main pathways display a strong interannual and seasonal variability (Fig. 8). Their magnitude varies largely: between 24% and 73% of the particles are retroflected and between 4% and 42% flow westward, respectively a 77% and 96% variability (variance/mean $\times$ 100). The retroflected pathway is always dominant, while the westward-flowing pathway can be almost shut down when the retroflection is strong. The southward-flowing pathway is the most stable pathway, with a variability of 13%.

The retroflected and westward-flowing pathway categories are strongly anti-correlated (correlation coefficient (c.c.) = -0.97, p-value < 0.001, top panel of Fig. 8a). This agrees well with a see-saw behaviour of the Labrador Current at the tip of the Grand Banks (Jutras et al., 2023; Han et al., 2019): when the retroflection towards the subpolar North Atlantic is strong, little water reaches the Slope Sea or the continental shelf break, and vice-versa. The Labrador Sea category is the opposite, being anti-correlated with the westward-flowing one (c.c. = -0.64, p < 0.001) and hence correlated with the retroflected one (c.c. = 0.55, p < 0.001). These correlations support the idea that the westward-flowing pathway is mainly associated with similar branches of the Labrador Current, namely the inshore branch, while the Labrador Sea and retroflected pathways are associated with the offshore branch of the current (see Section 3.2.2). The Belle Isle and southward-flowing pathway categories show no correlation with other pathways, suggesting that they are forced by different mechanisms.

There is a significant seasonal cycle in the time at which the particles veer westward or are retroflected at the tip of the Grand Banks (Fig. 8b). The retroflection is strongest in late summer (Aug.-Sep.) and generally weakest in the winter (Jan.), although it is strong in some winters, while the westward-flowing pathway is greatest in the winter (Feb.) and weakest in the summer (Jun.-Aug.). The opposite seasonal cycles between the two main pathways suggest that the see-saw behaviour also occurs at a seasonal scale. Seasonal variations in the circulation patterns near the Grand Banks are discussed in the literature, and are suggested to be driven by seasonal variations in the water temperature and salinity (advection of meltwater; Fratantoni & McCartney, 2010; Lazier & Wright, 1993) affecting stratification (Fratantoni & McCartney, 2010), in the density gradients across the shelf-break (Schneider et al., 2015), in the winds (Holliday et al., 2020; Han, 2005; K. R. Thompson et al., 1986), and to a southern drift of the Gulf Stream in the summer (Seidov et al., 2021).

In addition to the two main pathway categories, there is a marked seasonality in the trajectory of the particles leaving the Labrador Shelf towards the Labrador Sea (crossing the blue line on Fig. 1b). More particles do so in the early summer (Jun.-Jul.) compared to other seasons (not shown). This behavior agrees with the observations of Howatt et al. (2018), who suggest that northward winds, which are only present in the summer, drive an offshore Ekman transport that supports the export of freshwater to the Labrador Sea.

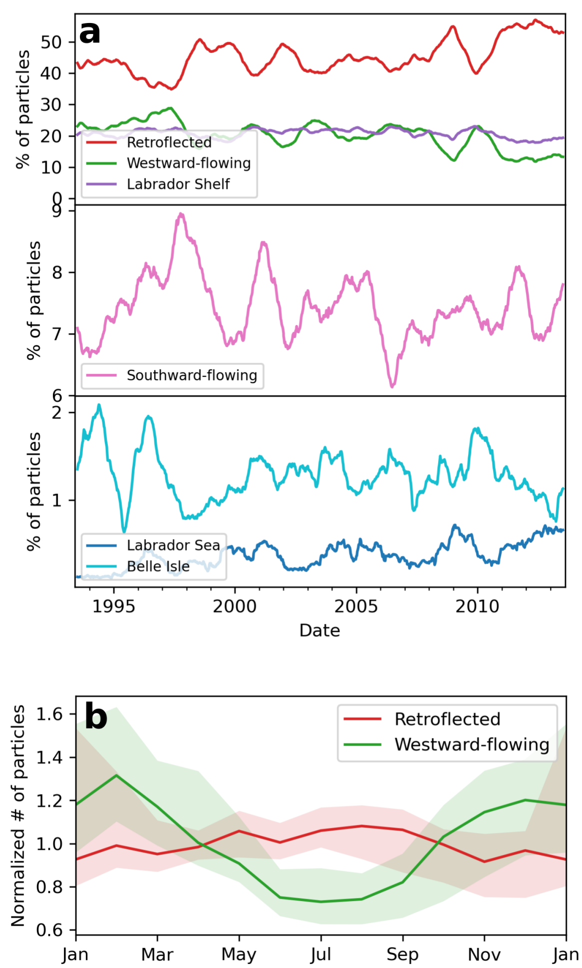


Figure 8. Temporal variability of the pathway categories. **(a)** Time series of the percentage of the total number of particles for each of the six pathway categories, with the time recorded at the seeding time. The time series are smoothed over one year and are presented in three different panels for readability. Note the differences between y-axes. **(b)** Seasonal cycle of the retroflected and westward-flowing pathways, for the time at which the particles reach the tip of the Grand Banks (i.e. when they cross the 49°N line, see Fig. 1a). This way of recording the time gives a better estimate of the local seasonal export variability, given the wide range of propagation times for the particles (Fig. 10; Fox et al., 2022). The amplitude is normalized. The shaded area shows the interannual spread in the seasonal cycle, computed from the squared sum of each year difference.

3.2.2 Spatial characteristics of the pathways

We can use the characteristics of the different pathway categories to deduce information about their origin. In the retroflected pathway category, the virtual particles have slightly higher velocities on the Labrador Shelf compared to the other pathways (Fig. 9iii), and most (though not all) particles originate from and flow within the offshore portion of the shelf (Fig. 9i and ii). The retroflected pathway thus seems to be mostly fed by the offshore, or shelf-break, faster branch of the Labrador Current. In contrast, the virtual particles in the westward-flowing and in the Labrador Shelf pathway categories originate equally from across the seeding line (Fig. 9iv), and then most converge towards the inshore section of the shelf (Fig. 9v). Still, more than 5% of particles drifting within the shelf-break branch of the Labrador Current join the westward-flowing pathway (Fig. 9v). The westward-flowing and Labrador Shelf pathways are also associated with generally slower velocities (Fig. 9vi and xii). The Labrador Shelf category overall seems to be fed predominantly by the offshore branch of the Labrador Current (Fig. 9xi). The particles that end up in the Labrador Sea originate from close to the shelf-break, and are carried by the offshore-most and fastest portion of the Labrador Current (Fig. 9viii and ix). In contrast, the particles entering the Belle Isle Strait are very slow (Fig. 9xviii). They can originate from any location across the shelf, although not from the offshore-most portion of the current (Fig. 9xvi), and more specifically from shallower depths ($<$ 50 m) than the other particles (Supplementary fig. A1). The depth distribution of the particles at initialization does not play a role for the other pathways. Finally, similarly to the westward-flowing category, the particles associated with the southward-flowing pathway do not appear to have a preferred origin and travel across the whole shelf (Fig. 9xiii and xiv). They also show a wider range of velocities than the other pathway categories (Fig. 9xv). This suggests that the westward-flowing and southward-flowing categories are not fed by a particular branch of the Labrador Current. We find that the particles in the southward-flowing category show turbulent motion soon after they leave the Grand Banks (Fig. 1b and 6). Since that region is located in the transition zone between the Gulf Stream and the more stable NAC, we suggest that the southward-flowing pathway emerges as particles get caught in small-scale features such as eddies, common in that region (Rossby, 1999; Brooks, 1987), explaining why these particles do not follow the average circulation of the Gulf Stream/NAC, directed northeastward (Bower et al., 2011).

3.2.3 Specific circulation patterns

Grouping the different clusters into pathway categories is useful to concentrate on the general properties of the pathways of the Labrador Current. Yet, within a pathway category, individual clusters often show distinct characteristics. These characteristics reveal important details of the circulation that can refine our view of the Labrador Current pathways. For instance, the particles in different retroflecting clusters veer at different locations. About one third of the retroflecting particles do so near Flemish Cap, and the remainder at the tip of the Grand Banks (Table 1 and Fig. 4). These proportions do not vary significantly with time. We notice that depending on where the particles retroflect, they reach slightly different regions of the North Atlantic. The particles that retroflect at Flemish Cap feed the north of the subpolar gyre (~52–57°N) and the particles that retroflect further downstream, at the tip of the Grand Banks, feed the center of the subpolar gyre (~45–52°N).

A more detailed look at the westward-flowing category provides information on the specific pathways of Labrador Current Waters. Some of the waters reaching the Slope Sea do so through the Avalon Channel (cluster #23), while others flow over the Grand Banks (cluster #18). We find that the waters entering the Laurentian Channel and reaching further south along the Scotian Shelf mostly go through the Avalon Channel (cluster #3 and 26). We also mentioned in Section 3.1.1 how cluster #22 contains particles that first go westward and are then retroflected (Fig. 4).

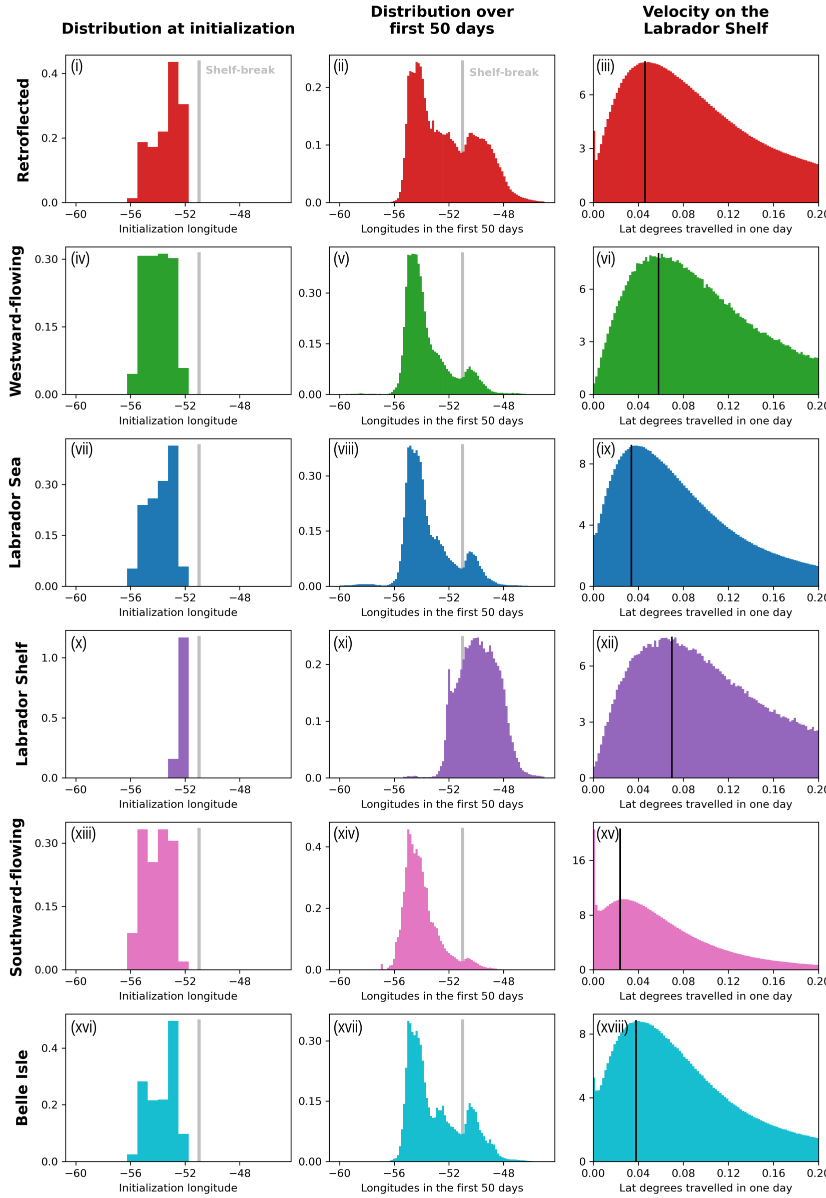


Figure 9. Left: Histograms of the longitudes of origin (along the seeding line), for each pathway. Middle: Histograms of the longitudes covered during the first 50 days, for each pathway. The first 50 days represent the portion of the trajectory north of Flemish Cap, where the current spreads longitudinally because of the presence of the Grand Banks. Right: For each pathway category, histograms of the average velocity of the particles (degrees of latitude travelled per day; zero indicates zonal displacement) over their first 100 days, the average time that particles spend north of the tip of the Grand Banks. The black vertical line indicates the peak of the histogram. The vertical grey lines in the left and middle panels show the location of the shelf-break. A cut-off at 0.20 was chosen for clarity, as the focus is on the bulk of the particles.

Finally, we find that the two clusters that belong to the southward-flowing pathway category are actually associated with two slightly different pathways (#12 and 29; Fig. 4). A first pathway flows *along* the Grand Banks shelf-break and, once it detaches from the shelf, veers slightly east and reaches deeper depths (maximal depths of 1500 m, cluster #12). Another pathway flows *over* the Grand Banks and, once it detaches from the shelf, continues straight to the south and reaches maximal depths of 700 m (cluster #29).

When looking at individual clusters, we also notice that the algorithm classified particles based on their propagation time (Fig. 10). This is not surprising, since it classified them based on their location at every time step, which implicitly contains information on the velocity. For instance, the particles classified in the westward-flowing pathway category take about 8 months to reach the Slope Sea from their seeding position, except for the particles classified in cluster #18, which take about one year (Fig. 10). Similarly, the particles classified in the retroflected pathway category take 2-3 months to reach the tip of the Grand Banks, except for the particle classified in clusters #2, 9 and 24, which take 4-9 months. While, here, our interest is on the various pathways of the Labrador Current, information on the propagation time is useful to evaluate how long it takes for anomalies carried by the Labrador Current to reach different export zones.

3.2.4 Characteristic depths of the pathways

Each identified pathway category has a distinct signature in depth (Fig. 11). The particles moving eastward stay at shallow depths, while the particles moving southward and westward reach deeper. The southward-flowing particles reach the deepest depths, diving on average to maximal depths of ∼1200 m or below 2000 m for 10% of them. These particles cross the Gulf Stream front, which acts as a barrier to cross-front flow down to 700 m, and as a stirrer below (Palter et al., 2013; Bower et al., 1985). Hence, it is expected that the southward-flowing particles travel at such great depths, since only the particles that subduct can cross the front and follow this pathway (Fig. 12). The westward-flowing particles dive on average to maximal depths of ∼900 m, or even down to 1800 m for 10% of them. Cluster #18 however remains above 700 m, gathering the particles that are not entrained below the front. Within the retroflected pathway category, particles descend once they quit the shelf, near the 50th meridian (Fig. 11b), reaching on average depths of ∼500 m, or 1000 m for 10% of them. The particles retroflecting at the tip of the Grand Banks (clusters #2 and 10) reach greater depths (∼1500 m) compared to particles retroflecting at Flemish Cap (∼700 m). Hence, the latter feed the core of the North Atlantic Current (down to ∼800 m, Gouretski, 2018), while the former feed the deep ocean. Finally, the particles of the Labrador Sea pathway category remain relatively close to the surface (above ∼150 m). This is probably due to the low salinity (33-34.5) of these waters compared to the open ocean (34.5-36.5), and suggests that the weak freshwater export from the Labrador Current contributes to increasing the stratification in the Labrador Sea (Howatt et al., 2018).

4 Discussion and conclusion

In this study, we present a method to classify geophysical Lagrangian trajectories using unsupervised clustering. Our results demonstrate that this method is useful and efficient to (i) classify Lagrangian tracks that are challenging to classify with more traditional methods (e.g. counting particles crossing hydrographic sections), (ii) assist in the treatment of huge Lagrangian tracks datasets, (iii) identify the main pathways of an ocean current, and (iv) analyze the variability in the magnitude of these pathways. The method was applied to 1.2 millions modelled trajectories along the Labrador Current and was successful in identifying the different pathways of the Labrador Current, including a previously unknown pathway directed southward from the tip of the Grand Banks. The

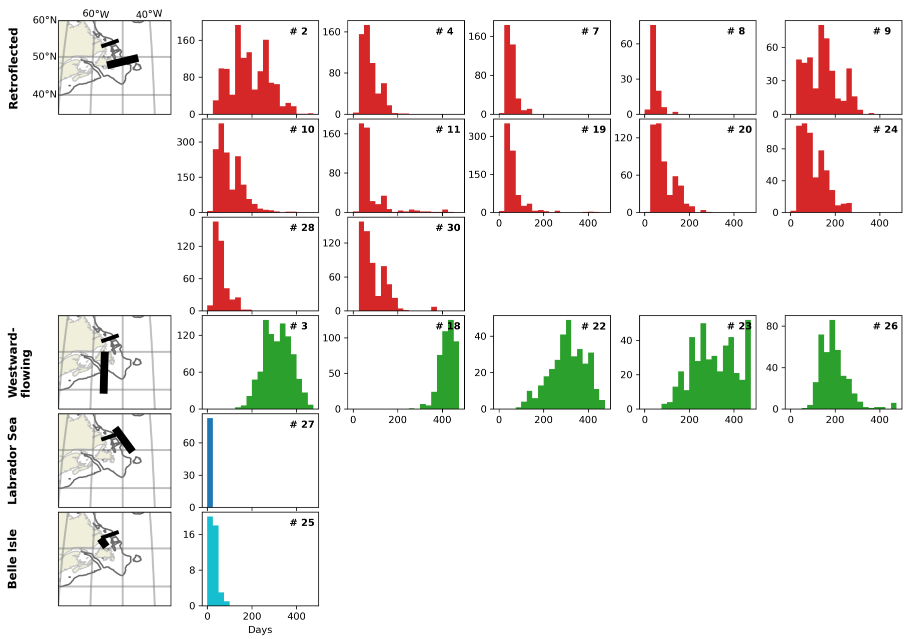


Figure 10. (**Histograms, right**) For each cluster in four pathway categories, histogram of the propagation time (in days) of the virtual particles, from the initialization line (thickest black on the maps on the left) to the entry point to the export zones associated with each category (medium-thickness black line, maps on the left). The cluster ID is indicated in the top right of the panel. We do not show the propagation time for the Labrador Shelf pathway category because they are not exported, and for the Southward-flowing pathway category because the export zone is not clearly definable. The thin black contour on the maps corresponds to the 350 m isobath.

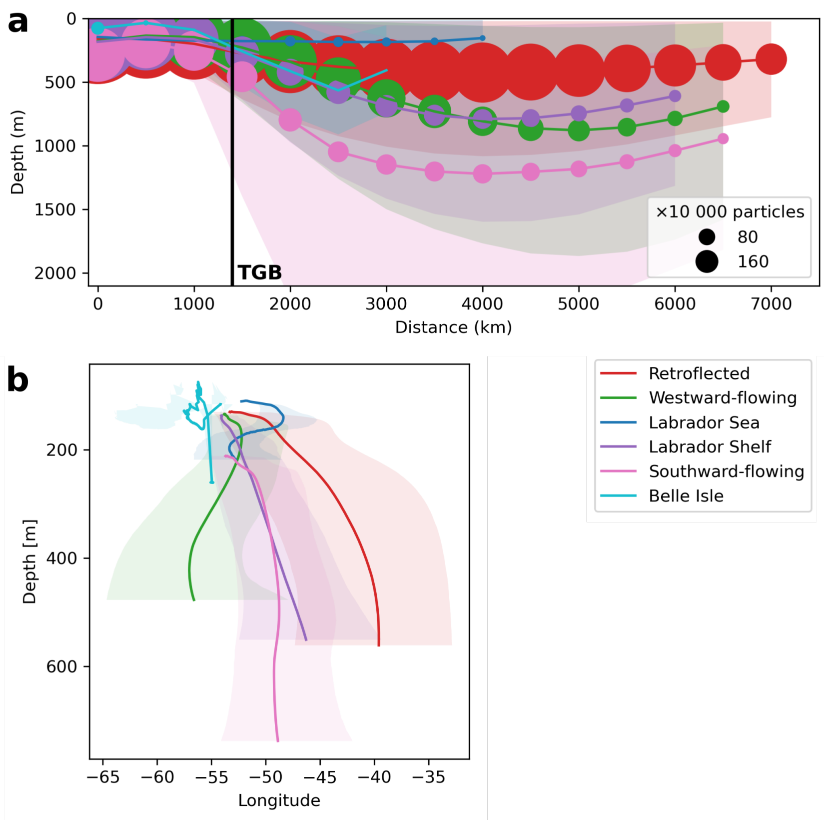


Figure 11. **(a)** Average depth of the particles in each pathway category along the distance travelled by the particles, using 100 km bins. The size of the dots represent the number of particles used in the average for each distance bin. This number tends to decrease with the travelled distance because particles can exit the domain or hit bathymetry. We stop displaying the data when less than 25% of the particles remain. The vertical black line indicates the approximate location of the Tip of the Grand Banks. **(b)** For each pathway category, averaged longitude of the particles in a given depth bin. In both plots, the shaded areas show the zone encompassing the 10% and 90% longitude percentile.

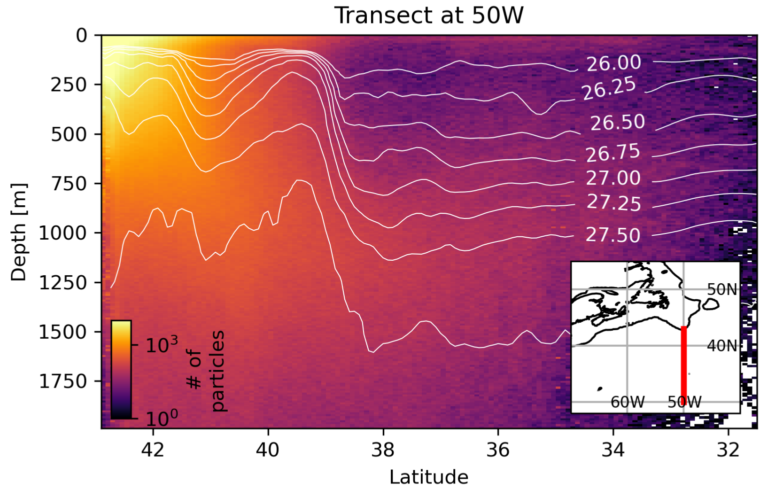


Figure 12. Particle distribution along a transect at 50°W that runs across the Gulf Stream front, from the tip of the Grand Banks (43°N, left) to 32.5°N (right). The inset shows the location of the transect (red line). Colours provide the number of particles passing by. White lines show isopycnals of potential density (+1000 kg m^{-3}). The strong tilt in the isopycnals around 39°N is due to the Gulf Stream front.

pre-processing applied here is relatively straightforward to implement, and the k-means++ algorithm is simple to use and converges well (Section 2.3). The use of a larger number of clusters than the expected number of pathways proves adequate, as it leads to a good performance of the clustering and to the identification of details of the circulation that we were initially not hoping to resolve. The choice of the number of clusters relies on two metrics: the silhouette score and an ad hoc "physics-based" metrics tailored to our scientific question. Overall, the algorithm is relatively cheap to run, except for the kernalized PCA step (see Section 2.3), which requires a lot of computational resources (here, 186 GB of RAM for one day, on a HPC system).

The results of the clustering confirm that the Labrador Current splits into two main branches: a branch retroflecting east towards the subpolar North Atlantic, representing ∼50% of the Labrador Current water, and a branch flowing west along the eastern American continental shelf-break and into the Slope Sea, representing ∼20% of it (Fig. 13). Two-thirds of the eastward retroflection occurs at the tip of the Grand Banks, and a quarter at Flemish Cap. The waters retroflecting at the tip of the Grand Banks reach deeper and get close to the Northeast corner, while the water retroflecting at Flemish Cap stay shallower and reach higher north. Secondary pathways of the Labrador Current include one exporting water from the Labrador Shelf to the Labrador Sea, one entering the Gulf of St. Lawrence through the Strait of Belle Isle, and one retroflecting after vising the Slope Sea (Fig. 13). In addition to these pathways, which were already documented in the literature, the clustering reveals a pathway bringing waters southwards from the tip of the Grand Banks and representing on average 8% of the Lagrangian trajectories. This pathway has been described for the underlying Deep Western Boundary Current (Bower et al., 2011, 2009) and suggests a connection between the two currents.

The variability of the two main branches is strong, reaching up to 96% of the mean state. These two branches strongly compensate each other through time, which is characteristic of a see-saw system. The time series of the magnitude of each pathway can be very useful to study what drives their variability. For instance, Jutras et al. (2023) introduce an index for the retroflection of the Labrador Current and use it to study the drivers of the retroflection and its impact on the physical and biogeochemical properties of the northwestern Atlantic. That index is based on counting the number of particles reaching a hydrographic line south of the Grand Banks, therefore not discriminating the particles going southward from the tip of the Grand Banks, or reaching the Labrador Sea, from those actually retroflected towards the subpolar North Atlantic. The time series of the magnitude of the retroflected pathway obtained in the present study through unsupervised clustering represent a more precise estimate of the magnitude of the retroflection of the Labrador Current.

Finally, by analyzing the origin of each pathway and the associated water masses (section 3.1.1), we can discuss what influences water properties in the export regions of each pathway. The two main pathways, retroflected and westward-flowing, are fed by both the inshore and shelf-break branches of the Labrador Current, but the shelf-break branch contributes slightly more to the retroflected pathway (about 30%) and the inshore branch slightly more to the westward-flowing pathway (about 90%, Fig. 13). The inshore branch is fed by Arctic outflow through Davis Strait and by river outflow along the Labrador Shelf, while the shelf-break branch is mostly fed by the Greenland Current (section 1, Florindo-López et al., 2020). Hence, variations in the magnitude of rivers outflow along the Labrador Shelf likely affect salinity in the Slope Sea slightly more than in the subpolar North Atlantic. In contrast, variations in salinity in the Greenland Current, due to changes in the Arctic freshwater outflow (de Steur et al., 2018) or to Greenland ice sheet melt (Marson et al., 2021), will affect salinity in the subpolar North Atlantic slightly more than in the Slope Sea. These variations would also likely affect salinity in the Labrador Sea through the Labrador Sea pathway, fed exclusively by this pathway. The southward-flowing pathway has a weak variability and is not associated with a particular branch of the Labrador Current. Hence, its variability does not contribute to that in western North Atlantic Ocean water properties, but changes in its water properties could. Finally, we also find that the pathways exporting water to the Labrador Sea and to the subpolar North Atlantic supply the surface ocean (Fig. 11 and 13). Since the Labrador Current carries freshwater, variations in these exports would likely affect the stratification in these regions, including the occurrence and intensity of deep convection, with potential effects on the Atlantic Meridional Overturning Circulation (AMOC; Lozier, 2012), carbon uptake (Fontela et al., 2016), and on oxygen repletion of the deep North Atlantic waters (Koelling et al., 2022; Atamanchuk et al., 2021).

To finish, this paper offers first and foremost methodological advancements for the geophysical community. The method, extensively described in this paper, could be applied to other oceanic currents or other types of geophysical Lagrangian trajectories.

5 Open Research

The Lagrangian tracking experiments can be reproduced by downloading the publicly available GLORYS12V1 outputs from the Copernicus Marine Environment Monitoring Service (CMS) website: `resources.marine.copernicus.eu/product-detail/GLOBAL_MULTIYEAR_PHY_001_030/INFORMATION`. Information about the OceanParcels tool for Python is available at `oceanparcels.org`. The scripts used to run the Lagrangian tracking experiments can be found as a supplementary material to Jutras et al. (2023).

The ML tools are available through the Python scikit-learn package (`scikit-learn.org/`). The scripts of the unsupervised clustering method are available at `https://github.com/noemieplanat/Clustering_Lagragian_particles`.

The data from the Global drifter program was obtained from the Atlantic Oceanographic and Meteorological Laboratory of the National Oceanic and Atmospheric Administration (AOML/NOAA, `ftp.aoml.noaa.gov/phod/pub/buoydata`). The RAFOS/SOFAR subsurface float trajectories are compiled from 52 experiments by the WOCE Subsurface Float Data Assembly Center (WFDAC, `www.aoml.noaa.gov/phod/float_traj/`).

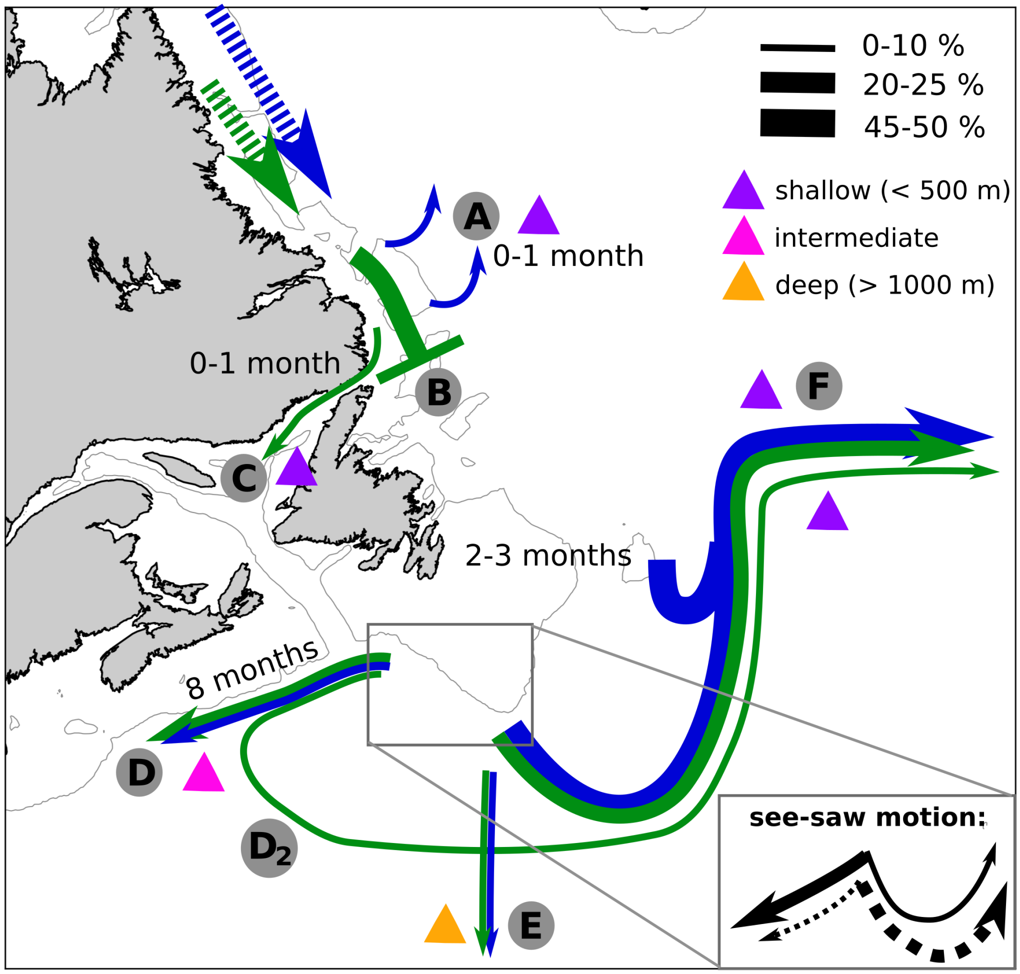


Figure 13. Summary of the pathways of the Labrador Current as identified by the unsupervised clustering of Lagrangian trajectories. The dashed arrows indicate the shelf-branch (green) and the shelf-break (blue) branches of the Labrador Current. The full arrows indicate the different pathways, and are identified with a letter: (A) Labrador Sea, (B) Labrador Shelf, (C) Belle Isle, (D) westward-flowing, (E) southward-flowing, (F) retroflected eastward, (D_2) westward-flowing and then retroflected. Pathway (B) ends with a bar because it contains particles that die on the shelf after they hit bathymetry. For each pathway, the width of the arrow indicates its averaged magnitude (corresponding to the fraction of the particles classified in that pathway category), and the color of the arrows indicates which of the Labrador Current branches mainly feeds the pathway. Colored triangles indicate the depth reached by the particles. The month labels indicate the average transit time from initialization to the export zone. The insert in the bottom-right illustrates the see-saw behaviour of the two main pathways (westward-flowing and retroflected): one weakens as the other strengthens, and vice-versa. The thin gray line indicates the 250 m isobath.

The Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it (`argo.ucsd.edu`, `https://www.ocean-ops.org`). The Argo Program is part of the Global Ocean Observing System.

655 # Appendix A Supplementary material

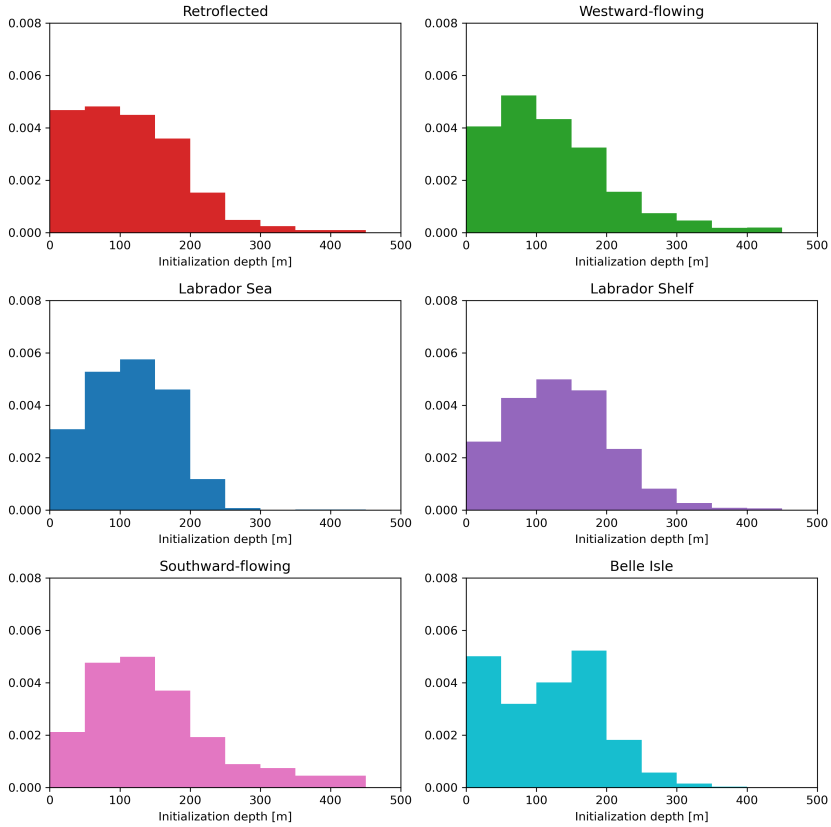


Figure A1. Histograms of the initialization depth of the particles associated with the different pathways.

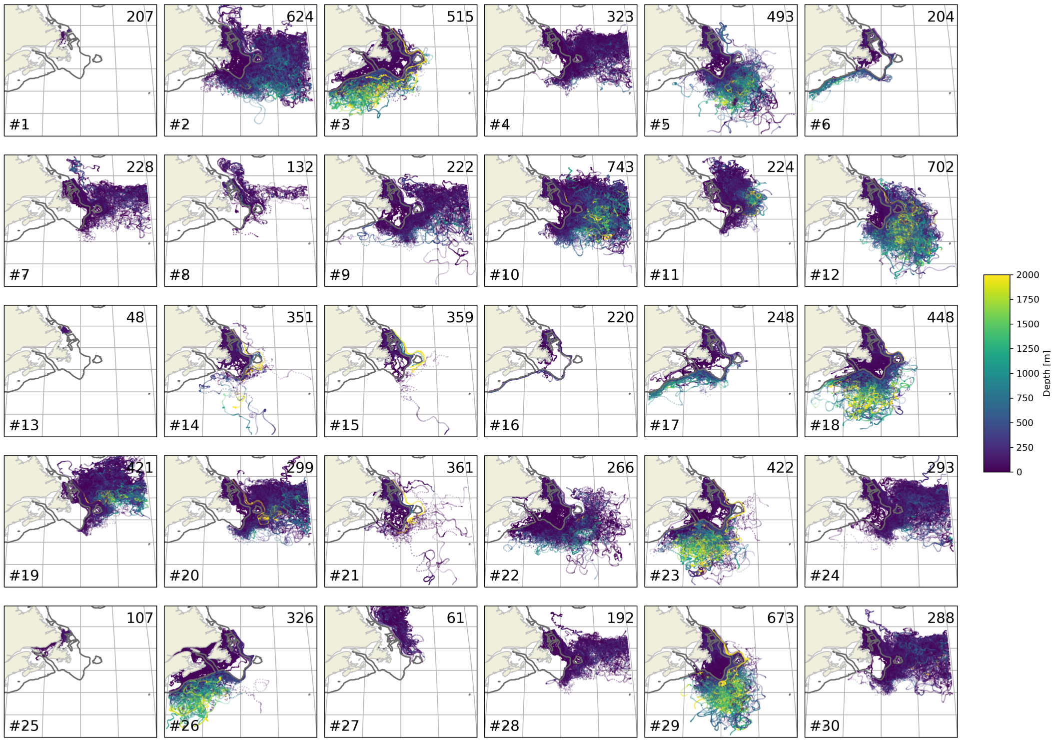


Figure A2. Trajectories in each cluster, for the test set. The color indicates the depth of the particles.

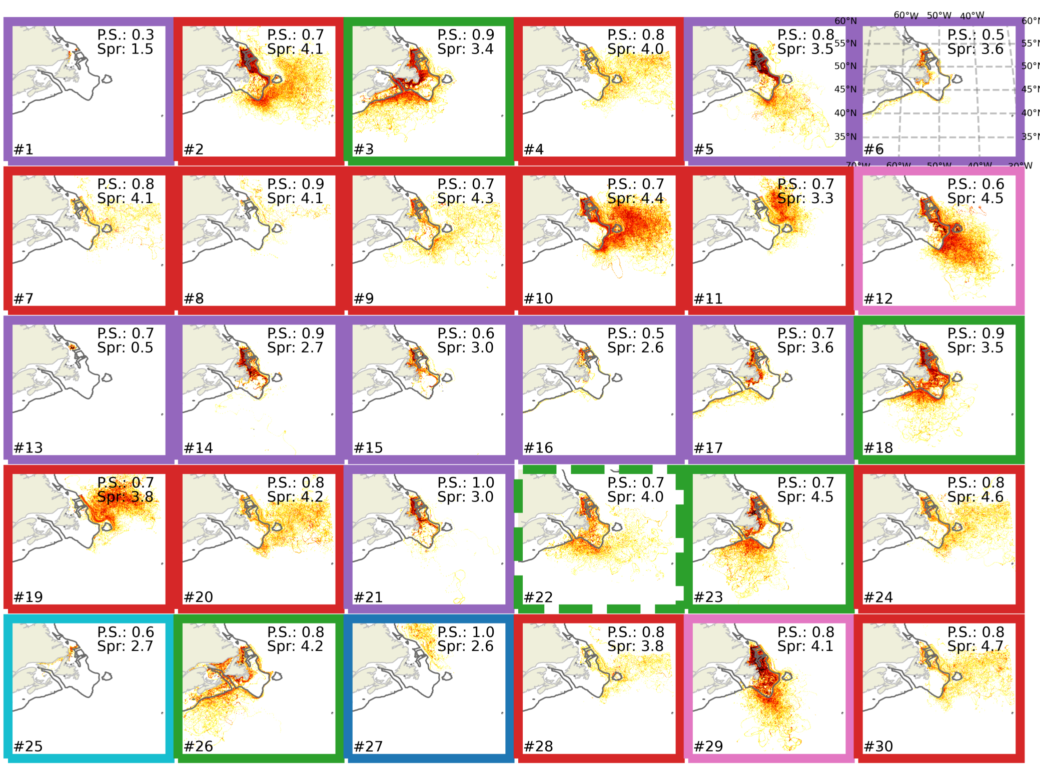


Figure A3. Same as Figure 4, but indicating in the top right the score of the physical metrics (P.S.) defined in section 2.3 and the internal spread (Spr).

1	67%	11	90%	21	100%
2	70%	12	100%	22	66%
3	90%	13	90%	23	90%
4	90%	14	100%	24	60%
5	90%	15	100%	25	100%
6	90%	16	100%	26	50%
7	100%	17	55%	27	90%
8	100%	18	90%	28	100%
9	90%	19	55%	29	100%
10	90%	20	100%	30	100%

Table B1. Agreement rate during the expert's classification into the different pathway categories, for each cluster (section 2.3.4).

Appendix B Clusters grouping exercise

The agreement rate between the consulted experts is of 100% in 12 clusters, all agree but one or two in 10 clusters, and is above 60% in 5 clusters (Table B1). The agreement rate is of 55% for the westward-flowing cluster #17, that 45% of the experts assign to the Labrador Shelf category. We assign this cluster to the Labrador Shelf category, because the strong majority of the particles remain on the Labrador Shelf. Disagreement regarding the Labrador Sea pathway is probably due to a lack of clear distinction between the Labrador Sea and the north of the subpolar North Atlantic. For the retroflected cluster #19, 45% of the experts assign to the Labrador Sea category. We assign cluster it to the retroflected category because, even if some of the particles in this cluster cross the southern portion of the Labrador Sea, they eventually feed the subpolar North Atlantic, where they will affect the water properties. For similar reasons, we go against the expert agreement on cluster #11 (which was categorized as belonging to the Labrador Sea category), because the particles are retroflected at Flemish Cap before aiming North. Finally, there is equality in the vote for cluster #26, between the Belle Isle and westward-flowing categories. An analysis of the individual trajectories, as opposed to a density view, reveals that while a fair amount of particles enter Belle Isle Strait, most go around Newfoundland and some around the Grand Banks (not shown). We therefore assign this cluster to the westward-flowing category.

Acknowledgments

The authors would like to thank Pr. Daniel Bourgault, Frédéric Cyr, Pr. Juliana Marson, Pr. Alfonso Mucci, Pr. Paul Myers, Pr. Jaime Palter, Nicolai von Oppeln-Bronikowski, Mariona Claret and Peter S. Galbraith for acting as experts in classifying the clusters into pathway categories, as well as Jan Klaus Rieck, David Straub and Bruno Tremblay for testing the classification protocol. We also want to thank Pr. David Rolnick for an informal conversation on Machine Learning applications to geosciences.

This research was enabled in part by support provided by Calcul Québec (`calculquebec.ca`) and Compute Canada (`computecanada.ca`). MJ was supported by the The Natural Sciences and Engineering Research Council of Canada (NSERC), the Fonds de recherche du Québec - Nature et technologie (FRQNT), Ouranos; NP by the FRQNT; and LCT by McGill University Science Undergraduate Research Award. This research was also funded by NSERC Accelerator Supplements (grant no. RGPAS/2018-522502) and the Canada Research Chair program awarded to COD, as well as by a NSERC Discover Grant awarded to Pr. Bruno Tremblay (grant no. RGPIN/2018-04838). Finally, we acknowledge the support of Québec-Océan.

MJ and NP designed and conducted the study with input from COD. MJ did the preprocessing of the data, and NP developed the ML model, with the assistance of MJ. LCT processed and analyzed the drifter and float dataset. MJ and NP analyzed the results and COD contributed to the interpretation. MJ wrote the manuscript, with contribution from NP for the method section, from LCT for the observation section and from COD for editing.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., … Nahavandi, S. (2021, December). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, *76*, 243–297. Retrieved 2023-03-02, from https://www.sciencedirect.com/science/article/pii/S1566253521001081 doi: 10.1016/j.inffus.2021.05.008

Atamanchuk, D., Palter, J., Palevsky, H., Le Bras, I., Koelling, J., & Nicholson, D. (2021, December). Linking Oxygen and Carbon Uptake with the Meridional Overturning Circulation Using a Transport Mooring Array. In *Frontiers in Ocean Observing: Documenting Ecosystems, Understanding Environmental Changes, Forecasting Hazards.* (E.S. Kappel, S.K. Juniper, S. Seeyave, E. Smith, and M. Visbeck ed., p. 9). Retrieved 2023-03-17, from https://doi.org/10.5670/oceanog.2021.supplement.02-03

Ayata, S. D., Lazure, P., & Thiébaut, (2010). How does the connectivity between populations mediate range limits of marine invertebrates? A case study of larval dispersal between the Bay of Biscay and the English Channel (North-East Atlantic). *Progress in Oceanography*, *87*(1-4), 18–36. doi: 10.1016/j.pocean.2010.09.022

Boehme, L., & Rosso, I. (2021). Classifying Oceanographic Structures in the Amundsen Sea, Antarctica. *Geophysical Research Letters*, *48*(5), 1–8. doi: 10.1029/2020GL089412

Bower, A. S., Lozier, M. S., Gary, S. F., & Böning, C. W. (2009). Interior pathways of the North Atlantic meridional overturning circulation. *Nature*, *459*(7244), 243–247. doi: 10.1038/nature07979

Bower, A. S., Lozier, S., & Gary, S. (2011). Export of Labrador Sea Water from the subpolar North Atlantic: A Lagrangian perspective. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *58*(17-18), 1798–1818. Retrieved from http://dx.doi.org/10.1016/j.dsr2.2010.10.060 (Publisher: Elsevier) doi: 10.1016/j.dsr2.2010.10.060

Bower, A. S., Rossby, H. T., & Lillibridge, J. L. (1985). The Gulf Stream—Barrier or Blender? *Journal of Physical Oceanography*, *15*(1), 24–32. doi: 10.1175/1520-0485(1985)015⟨0024:tgsob⟩2.0.co;2

Brankov, E., Trivikrama Rao, S., & Steven Porter, P. (1998). A Trajectory-Clustering-Correlation Methodology for Examining the Long-Range Transport of Ari Pollutants. *Atmospheric Environment*, *32*(9), 1525–1534.

Brooks, D. A. (1987). The Influence of Warm-Core Rings on Slope Water Entering the Gulf of Maine. *Journal of Geophysical Research*, *92*(C8), 8183–8196.

Brunette, C., Tremblay, B., & Newton, R. (2019, February). Winter Coastal Divergence as a Predictor for the Minimum Sea Ice Extent in the Laptev Sea. *Journal of Climate*, *32*(4), 1063–1080. Retrieved 2022-11-25, from https://journals.ametsoc.org/doi/10.1175/JCLI-D-18-0169.1 doi: 10.1175/JCLI-D-18-0169.1

Cetina-Heredia, P., Roughan, M., van Sebille, E., Feng, M., & Coleman, M. A. (2015). Strengthened currents override the effect of warming on lobster larval dispersal and survival. *Global Change Biology*, *21*(12), 4377–4386. doi: 10.1111/gcb.13063

Chabot, D., & Dutil, J. D. (1999). Reduced growth of Atlantic cod in non-lethal hypoxic conditions. *Journal of Fish Biology*, *55*(3), 472–491. doi: 10.1006/jfbi.1999.1005

Chen, Z., Kwon, Y. O., Chen, K., Fratantoni, P. S., Gawarkiewicz, G., & Joyce, T. M. (2020). Long-Term SST Variability on the Northwest Atlantic Continental Shelf and Slope. *Geophysical Research Letters*, *47*(1), 1–11. doi: 10.1029/2019GL085455

Claret, M., Galbraith, E. D., Palter, J. B., Bianchi, D., Fennel, K., Gilbert, D., & Dunne, J. P. (2018). Rapid coastal deoxygenation due to ocean circulation shift in the northwest Atlantic. *Nature Climate Change*, *8*. Retrieved from http://www.nature.com/articles/s41558-018-0263-1 doi: 10.1038/s41558-018-0263-1

Daher, H., Beal, L. M., & Schwarzkopf, F. U. (2020). A New Improved Estimation of Agulhas Leakage Using Observations and Simulations of Lagrangian Floats and Drifters. *Journal of Geophysical Research: Oceans*, *125*(4), 1–17. doi: 10.1029/2019JC015753

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. doi: 10.1002/qj.828

Delandmeter, P., & Van Sebille, E. (2019). The Parcels v2.0 Lagrangian framework: New field interpolation schemes. *Geoscientific Model Development*, *12*(8), 3571–3584. doi: 10.5194/gmd-12-3571-2019

de Steur, L., Peralta-Ferriz, C., & Pavlova, O. (2018). Freshwater Export in the East Greenland Current Freshens the North Atlantic. *Geophysical Research Letters*, *45*(24), 13,359–13,366. Retrieved 2023-06-22, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL080207 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018GL080207) doi: 10.1029/2018GL080207

Echols, R., Rocap, G., & Riser, S. (2020). Exploring biogeographical provinces with machine learning and Argo float data. In *AGU 2020 fall meeting* (pp. OS022–06).

Filippi, M., Hadjighasem, A., Rayson, M., Rypina, I. I., Ivey, G., Lowe, R., ... Peacock, T. (2021). Investigating transport in a tidally driven coral atoll flow using Lagrangian coherent structures. *Limnology and Oceanography*, *66*(11), 4017–4027. doi: 10.1002/lno.11939

Filippi, M., Rypina, I. I., Hadjighasem, A., & Peacock, T. (2021). An optimized-parameter spectral clustering approach to coherent structure detection in geophysical flows. *Fluids*, *6*(1), 1–34. doi: 10.3390/fluids6010039

Fischer, J., & Schott, F. A. (2002). Labrador Sea Water tracked by profiling floats - From the boundary current into the Open North Atlantic. *Journal of Physical Oceanography*, *32*(2), 573–584.

Florindo-López, C., Bacon, S., Aksenov, Y., Chafik, L., Colbourne, E., Penny Holliday, N., & Holliday, N. P. (2020). Arctic Ocean and Hudson Bay Freshwater Exports : New Estimates from 7 Decades of Hydrographic Surveys on the Labrador Shelf. *Journal of Climate*, *33*(20), 1–62. doi: 10.1175/jcli-d-19-0083.1

Fontela, M., García-Ibáñez, M. I., Hansell, D. A., Mercier, H., & Pérez, F. F. (2016). Dissolved Organic Carbon in the North Atlantic Meridional Overturning Circulation. *Scientific Reports*, *6*, 1–9. Retrieved from http://dx.doi.org/10.1038/srep26931 (Publisher: Nature Publishing Group ISBN: 2045-2322 (Electronic) 2045-2322 (Linking)) doi: 10.1038/srep26931

Fox, A. D., Biastoch, A., Cunningham, S. A., Gary, S. F., Handmann, P., Penny, N., ... Schmidt, C. (2022). How reduced Labrador Sea surface heat loss caused exceptional freshening and cooling in the eastern subpolar North Atlantic.

Ocean Science, *in prep.*(April), 1–35.

Fratantoni, P. S., & McCartney, M. S. (2010). Freshwater export from the Labrador Current to the North Atlantic Current at the Tail of the Grand Banks of Newfoundland. *Deep-Sea Research Part I: Oceanographic Research Papers*, *57*(2), 258–283. Retrieved from `http://dx.doi.org/10.1016/j.dsr.2009.11.006` doi: 10.1016/j.dsr.2009.11.006

Gilbert, D., Sundby, B. B., Gobeil, C., Mucci, A., & Tremblay, G.-H. (2005). A seventy-two-year record of diminishing deep-water oxygen in the St. Lawrence estuary: The northwest Atlantic connection. *Limnology and Oceanography*, *50*(5), 1654–1666.

Gillard, L. C., Hu, X., Myers, P. G., & Bamber, J. L. (2016). Meltwater pathways from marine terminating glaciers of the Greenland ice sheet. *Geophysical Research Letters*, *43*(20), 10,873–10,882. doi: 10.1002/2016GL070969

Gonçalves Neto, A., Langan, J. A., & Palter, J. B. (2021). Changes in the Gulf Stream precede rapid warming of the Northwest Atlantic Shelf. *Nature Communications Earth & Environment*, *2*(74). Retrieved from `http://dx.doi.org/10.1038/s43247-021-00143-5` doi: 10.1038/s43247-021-00143-5

Gonçalves Neto, A., Palter, J. B., Xu, X., & Fratantoni, P. (2023). Temporal Variability of the Labrador Current Pathways Around the Tail of the Grand Banks at Intermediate Depths in a High-Resolution Ocean Circulation Model. *Journal of Geophysical Research: Oceans*, *128*(3), e2022JC018756. Retrieved 2023-03-13, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2022JC018756` doi: 10.1029/2022JC018756

Gouretski, V. (2018). *WOCE-Argo Global Hydrographic Climatology (WAGHC Version 1.0).* (Publication Title: World Data Center for Climate (WDCC) at DKRZ)

Han, G. (2005). Wind-driven barotropic circulation off Newfoundland and Labrador. *Continental Shelf Research*, *25*(17), 2084–2106. doi: 10.1016/j.csr.2005.04.015

Han, G., Chen, N., & Ma, Z. (2014). Is there a north-south phase shift in the surface Labrador Current transport on the interannual-to-decadal scale? *Journal of Geophysical Research: Oceans*, *119*(1), 276–287. (ISBN: 2169-9275) doi: 10.1002/2013JC009102

Han, G., Loder, J. W., & Smith, P. C. (1999). Seasonal-Mean Hydrography and Circulation in the Gulf of St. Lawrence and on the Eastern Scotian and Southern Newfoundland Shelves. *Journal of Physical Oceanography*, *29*(6), 1279–1301. Retrieved from `http://journals.ametsoc.org/doi/abs/10.1175/1520-0485%281999%29029%3C1279%3ASMHACI%3E2.0.CO%3B2` (ISBN: 0022-3670) doi: 10.1175/1520-0485(1999)029⟨1279:SMHACI⟩2.0.CO;2

Han, G., Ma, Z., & Chen, N. (2019). Ocean climate variability off Newfoundland and Labrador over 1979–2010: A modelling approach. *Ocean Modelling*, *144*(October), 101505. Retrieved from `https://doi.org/10.1016/j.ocemod.2019.101505` doi: 10.1016/j.ocemod.2019.101505

Handmann, P., Fischer, J., Visbeck, M., Karstensen, J., Biastoch, A., Böning, C. W., & Patara, L. (2018). The Deep Western Boundary Current in the Labrador Sea From Observations and a High-Resolution Model. *Journal of Geophysical Research: Oceans*, *123*(4), 2829–2850. doi: 10.1002/2017JC013702

Hart-Davis, M., & Backeberg, B. (2021, April). Towards a particle trajectory modelling approach in support of South African search and rescue operations at sea. *Journal of Operational Oceanography*, 1–9. Retrieved 2022-11-25, from `https://www.tandfonline.com/doi/full/10.1080/1755876X.2021.1911485` doi: 10.1080/1755876X.2021.1911485

Hertwig, D., Burgin, L., Gan, C., Hort, M., Jones, A., Shaw, F., … Zhang, K. (2015). Development and demonstration of a Lagrangian dispersion modeling system for real-time prediction of smoke haze pollution frombiomass burning in

Southeast Asia. *Journal of Geophysical Research*, *120*(24), 12,605–12,630. doi: 10.1002/2015JD023422

Hofmann, T., Schölkopf, B., & Smola, A. J. (2008, June). Kernel methods in machine learning. *The Annals of Statistics*, *36*(3), 1171–1220. Retrieved 2022-11-28, from https://projecteuclid.org/journals/annals-of-statistics/volume-36/issue-3/Kernel-methods-in-machine-learning/10.1214/009053607000000677.full (Publisher: Institute of Mathematical Statistics) doi: 10.1214/009053607000000677

Holliday, N. P., Bersch, M., Berx, B., Chafik, L., Cunningham, S., Florindo-López, C., ... Yashayaev, I. (2020). Ocean circulation causes the largest freshening event for 120 years in eastern subpolar North Atlantic. *Nature Communications*, *11*(1). doi: 10.1038/s41467-020-14474-y

Houghton, I. A., & Wilson, J. D. (2020). El Niño Detection Via Unsupervised Clustering of Argo Temperature Profiles. *Journal of Geophysical Research: Oceans*, *125*(9), 1–12. doi: 10.1029/2019JC015947

Howatt, T. M., Palter, J. B., Matthews, J. B. R., deYoung, B., Bachmayer, R., & Claus, B. (2018). Ekman and eddy exchange of freshwater and oxygen across the Labrador Shelf break. *Journal of Physical Oceanography*, *48*(5), 1015–1031. doi: 10.1175/JPO-D-17-0148.1

Jones, D. C., Holt, H. J., Meijers, A. J., & Shuckburgh, E. (2019). Unsupervised Clustering of Southern Ocean Argo Float Temperature Profiles. *Journal of Geophysical Research: Oceans*, *124*(1), 390–402. doi: 10.1029/2018JC014629

Jutras, M., Dufour, C. O., Mucci, A., Cyr, F., & Gilbert, D. (2020). Temporal Changes in the Causes of the Observed Oxygen Decline in the St. Lawrence Estuary. *Journal of Geophysical Research: Oceans*, *125*(12), 1–20. doi: 10.1029/2020JC016577

Jutras, M., Dufour, Carolina O., Mucci, Alfonso, & Talbot, Lauryn. (2023). Remote control of the retroflection of the Labrador Current.

Jönsson, B. F., & Watson, J. R. (2016, April). The timescales of global surface-ocean connectivity. *Nature Communications*, *7*(1), 11239. Retrieved 2022-11-25, from https://www.nature.com/articles/ncomms11239 (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/ncomms11239

Kawasaki, T., Matsumura, Y., & Hasumi, H. (2022). Deep water pathways in the North Pacific Ocean revealed by Lagrangian particle tracking. *Scientific Reports*, *12*(1), 1–10. Retrieved from https://doi.org/10.1038/s41598-022-10080-8 (Publisher: Nature Publishing Group UK ISBN: 4159802210080) doi: 10.1038/s41598-022-10080-8

Kelly, S. J., Proshutinsky, A., Popova, E. K., Aksenov, Y. K., & Yool, A. (2019). On the Origin of Water Masses in the Beaufort Gyre. *Journal of Geophysical Research: Oceans*, *124*(7), 4696–4709. doi: 10.1029/2019JC015022

Kläs, M., & Vollmer, A. M. (2018). Uncertainty in Machine Learning Applications: A Practice-Driven Classification of Uncertainty. In B. Gallina, A. Skavhaug, E. Schoitsch, & F. Bitsch (Eds.), *Computer Safety, Reliability, and Security* (pp. 431–438). Cham: Springer International Publishing. doi: 10.1007/978-3-319-99229-7_36

Koelling, J., Atamanchuk, D., Karstensen, J., Handmann, P., & Wallace, D. W. R. (2022, January). Oxygen export to the deep ocean following Labrador Sea Water formation. *Biogeosciences*, *19*(2), 437–454. Retrieved 2023-06-22, from https://bg.copernicus.org/articles/19/437/2022/ (Publisher: Copernicus GmbH) doi: 10.5194/bg-19-437-2022

Koszalka, I. M., & Lacasce, J. H. (2010, August). Lagrangian analysis by clustering. *Ocean Dynamics*, *60*(4), 957–972. doi: 10.1007/s10236-010-0306-2

Kremer, T., Schömer, E., Euler, C., & Riemer, M. (2020). Cluster analysis tailored to structure change of tropical cyclones using a very large number of trajectories. *Monthly Weather Review*, *148*(10), 4209–4229. doi:

10.1175/MWR-D-19-0408.1
Lazier, J. R. N., & Wright, D. G. (1993). Annual velocity variations in the Labrador Current. *Journal of Physical Oceanography*, *23*(4), 659–678. doi: 10.1175/1520-0485(1993)023⟨0659:AVVITL⟩2.0.CO;2
Lebreton, L. C., Greer, S. D., & Borrero, J. C. (2012). Numerical modelling of floating debris in the world's oceans. *Marine Pollution Bulletin*, *64*(3), 653–661. Retrieved from `http://dx.doi.org/10.1016/j.marpolbul.2011.10.027` (Publisher: Elsevier Ltd) doi: 10.1016/j.marpolbul.2011.10.027
Lellouche, J.-M., Greiner, E., Le Galloudec, O., Garric, G., Regnier, C., Drevillon, M., ... Le Traon, P. Y. (2018). Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12° high-resolution system. *Ocean Science*, *14*(5), 1093–1126. doi: 10.5194/os-14-1093-2018
Loder, J. W., Petrie, B. D., & Gawarkiewicz, G. (1998). The coastal ocean off north-eastern North America: a large-scale view. In *The Sea* (Vol. 11, pp. 105–133).
Lozier, M. S. (2012). Overturning in the North Atlantic. *Annual Review of Marine Science*, *4*(1), 291–315. Retrieved 2023-02-07, from `https://doi.org/10.1146/annurev-marine-120710-100740` (_eprint: https://doi.org/10.1146/annurev-marine-120710-100740) doi: 10.1146/annurev-marine-120710-100740
Madec, G., Bourdallé-Badie, R., Chanut, J., Clementi, E., Coward, A., Ethé, C., ... (2019). *NEMO ocean engine (Version v4. 0).* Notes Du Pôle De Modélisation De L'institut Pierre-simon Laplace (IPSL), Zenodo [Dataset].
Marson, J. M., Gillard, L. C., & Myers, P. G. (2021). Distinct Ocean Responses to Greenland's Liquid Runoff and Iceberg Melt. *Journal of Geophysical Research: Oceans*, *126*(12), e2021JC017542. Retrieved 2023-06-22, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2021JC017542` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021JC017542) doi: 10.1029/2021JC017542
Marson, J. M., Myers, P. G., Hu, X., & Le Sommer, J. (2018). Using Vertically Integrated Ocean Fields to Characterize Greenland Icebergs' Distribution and Lifetime. *Geophysical Research Letters*, *45*(9), 4208–4217. doi: 10.1029/2018GL077676
Merino, N., Le Sommer, J., Durand, G., Jourdain, N. C., Madec, G., Mathiot, P., & Tournadre, J. (2016). Antarctic icebergs melt over the Southern Ocean : Climatology and impact on sea ice. *Ocean Modelling*, *104*, 99–110. Retrieved from `http://dx.doi.org/10.1016/j.ocemod.2016.05.001` (Publisher: Elsevier Ltd) doi: 10.1016/j.ocemod.2016.05.001
Mills, K. E., Pershing, A. J., Brown, C. J., Chen, Y., Chiang, F. S., Holland, D. S., ... Wahle, R. A. (2013). Fisheries management in a changing climate: Lessons from the 2012 ocean heat wave in the Northwest Atlantic. *Oceanography*, *26*(2). doi: 10.5670/oceanog.2013.27
Myers, P. G. (2005). Impact of freshwater from the Canadian Arctic Archipelago on Labrador Sea Water formation. *Geophysical Research Letters*, *32*(6), 1–4. doi: 10.1029/2004GL022082
New, A., Smeed, D. A., Czaja, A., Blaker, A. T., Mecking, J. V., Mathews, J. P., ... Sanchez-Franks, A. (2021). Labrador Slope Water Connects the Subarctic with the Gulf Stream. *Environmental Research Letters*, *16*(8), 084019. (Publisher: IOP Publishing) doi: 10.1088/1748-9326/ac1293
Palter, J. B., Lozier, M. S., & Lavender, K. L. (2008). How does Labrador sea water enter the deep western boundary current? *Journal of Physical Oceanography*, *38*(5), 968–983. doi: 10.1175/2007JPO3807.1
Palter, J. B., Marinov, I., Sarmiento, J. L., & Gruber, N. (2013). Large-Scale, Persistent Nutrient Fronts of the World Ocean: Impacts on Biogeochemistry. In (pp. 1–38). Berlin, Heidelberg: Springer. Retrieved 2023-03-03, from `https://doi.org/10.1007/698_2013_241` doi: 10.1007/698_2013_241

Pershing, A. J., Kerr, L. A., Bris, A. L., Mills, K. E., Sherwood, G. D., Alexander, M. A., ... Thomas, A. C. (2016). Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science*, *352*(6284), 423. doi: 10.1126/science.aae0463

Petrie, B., Toulany, B., & Garrett, C. (1988). The transport of water, heat and salt through the strait of Belle Isle. *Atmosphere - Ocean*, *26*(2).

Petrie, B. D., & Drinkwater, K. F. (1993). Temperature and Salinity Variability on the Scotian Shelf and in the Gulf of Maine 1945-1990. *Journal of Geophysical Research*, *98*(C11), 20079–20089.

Phelps, J. J., Polton, J. A., Souza, A. J., & Robinson, L. A. (2015). Behaviour influences larval dispersal in shelf sea gyres: Nephrops norvegicus in the Irish Sea. *Marine Ecology Progress Series*, *518*, 177–191. doi: 10.3354/meps11040

Poitevin, P., Thebault, J., Siebert, V., Donnet, S., Archambault, P., Doré, J., ... Lazure, P. (2019). Growth response of Arctica islandica to North Atlantic oceanographic conditions since 1850. *Frontiers in Marine Science*, *6*, 1–14. doi: 10.3389/fmars.2019.00483

Pérez-Brunius, P., Rossby, T., & Watts, D. R. (2004). Absolute transports of mass and temperature for the North Atlantic Current-subpolar front system. *Journal of Physical Oceanography*, *34*(8), 1870–1883.

Roach, C. J., & Speer, K. (2019). Exchange of Water Between the Ross Gyre and ACC Assessed by Lagrangian Particle Tracking. *Journal of Geophysical Research: Oceans*, *124*(7), 4631–4643. doi: 10.1029/2018JC014845

Rossby, T. (1999). On gyre interactions. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *46*(1-2), 139–164. doi: 10.1016/S0967-0645(98)00095-2

Rosso, I., Mazloff, M. R., Talley, L. D., Purkey, S. G., Freeman, N. M., & Maze, G. (2020). Water Mass and Biogeochemical Variability in the Kerguelen Sector of the Southern Ocean: A Machine Learning Approach for a Mixing Hot Spot. *Journal of Geophysical Research: Oceans*, *125*(3), 1–23. doi: 10.1029/2019JC015877

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(C), 53–65. doi: 10.1016/0377-0427(87)90125-7

Saucier, F. J., Roy, F., Gilbert, D., Pellerin, P., & Ritchie, H. (2003). Modeling the formation and circulation processes of water masses and sea ice in the Gulf of St. Lawrence, Canada. *Journal of Geophysical Research C: Oceans*, *108*(8), 25–1. Retrieved from `http://doi.wiley.com/10.1029/2000JC000686` (ISBN: 2156-2202) doi: 10.1029/2000jc000686

Schmidt, S., & Send, U. (2007). Origin and composition of seasonal Labrador Sea freshwater. *Journal of Physical Oceanography*, *37*(6), 1445–1454. doi: 10.1175/JPO3065.1

Schneide, C., Pandey, A., Padberg-Gehle, K., & Schumacher, J. (2018). Probing turbulent superstructures in Rayleigh-Bénard convection by Lagrangian trajectory clusters. *Physical Review Fluids*, *3*(11), 1–15. (arXiv: 1810.07396 Publisher: American Physical Society) doi: 10.1103/PhysRevFluids.3.113501

Schneider, L., Kieke, D., Jochumsen, K., Colbourne, E., Yashayaev, I., Steinfeldt, R., ... Rhein, M. (2015). Variability of Labrador Sea Water transported through Flemish Pass during 1993–2013. *Journal of Geophysical Research: Oceans*, *120*, 5514–5533. doi: 10.1002/2015JC010939

Schulze Chretien, L. M., & Frajka-Williams, E. (2018). Wind-driven transport of fresh shelf water into the upper 30m of the Labrador Sea. *Ocean Science*, *14*(5), 1247–1264. doi: 10.5194/os-14-1247-2018

Seidov, D., Mishonov, A., & Parsons, R. (2021). Recent warming and decadal variability of Gulf of Maine and Slope Water. *Limnology and Oceanography*, *66*(9), 3472–3488. Retrieved 2022-11-29, from `https://onlinelibrary.wiley.com/doi/abs/10.1002/lno.11892` (_eprint:

https://onlinelibrary.wiley.com/doi/pdf/10.1002/lno.11892) doi: 10.1002/lno.11892

Shaw, J.-L., & Galbraith, P. S. (2023). Climatology of Transport in the Strait of Belle Isle. *Journal of Geophysical Research: Oceans*, *128*(2), e2022JC019084. Retrieved 2023-03-15, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2022JC019084` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022JC019084) doi: 10.1029/2022JC019084

Simons, R. D., Siegel, D. A., & Brown, K. S. (2013). Model sensitivity and robustness in the estimation of larval transport: A study of particle tracking parameters. *Journal of Marine Systems*, *119-120*, 19–29. Retrieved from `http://dx.doi.org/10.1016/j.jmarsys.2013.03.004` doi: 10.1016/j.jmarsys.2013.03.004

Tamsitt, V., Drake, H. F., Morrison, A. K., Talley, L. D., Dufour, C. O., Gray, A. R., ... Weijer, W. (2017). Spiraling pathways of global deep waters to the surface of the Southern Ocean. *Nature Communications*, *8*(1), 1–10. doi: 10.1038/s41467-017-00197-0

Thomas, S. D. A., Jones, D. C., Faul, A., Mackie, E., & Pauthenet, E. (2021, November). Defining Southern Ocean fronts using unsupervised classification. *Ocean Science*, *17*(6), 1545–1562. Retrieved 2023-01-16, from `https://os.copernicus.org/articles/17/1545/2021/` doi: 10.5194/os-17-1545-2021

Thompson, A. F., & Sallée, J.-B. (2012, June). Jets and Topography: Jet Transitions and the Impact on Transport in the Antarctic Circumpolar Current. *Journal of Physical Oceanography*, *42*(6), 956–972. Retrieved 2023-06-27, from `http://journals.ametsoc.org/doi/10.1175/JPO-D-11-0135.1` doi: 10.1175/JPO-D-11-0135.1

Thompson, K. R., Lazier, J. R. N., & Taylor, B. (1986). Wind-forced changes in Labrador Current transport. *Journal of Geophysical Research*, *91*(C12), 14261. doi: 10.1029/jc091ic12p14261

Tiira, J., & Moisseev, D. (2020). Unsupervised classification of vertical profiles of dual polarization radar variables. *Atmospheric Measurement Techniques*, *13*(3), 1227–1241. doi: 10.5194/amt-13-1227-2020

Townsend, D. W., Pettigrew, N. R., Thomas, M. A., Neary, M. G., Mcgillicuddy, D. J., & Donnell, J. O. (2015). Water masses and nutrient fluxes to the Gulf of Maine. *Journal of Marine Research*, *73*, 93–122. doi: 10.1038/141548c0

van Sebille, E., Griffies, S. M., Abernathey, R., Adams, T. P., Berloff, P., Biastoch, A., ... Zika, J. D. (2018). Lagrangian ocean analysis: Fundamentals and practices. *Ocean Modelling*, *121*(November 2017), 49–75. doi: 10.1016/j.ocemod.2017.11.008

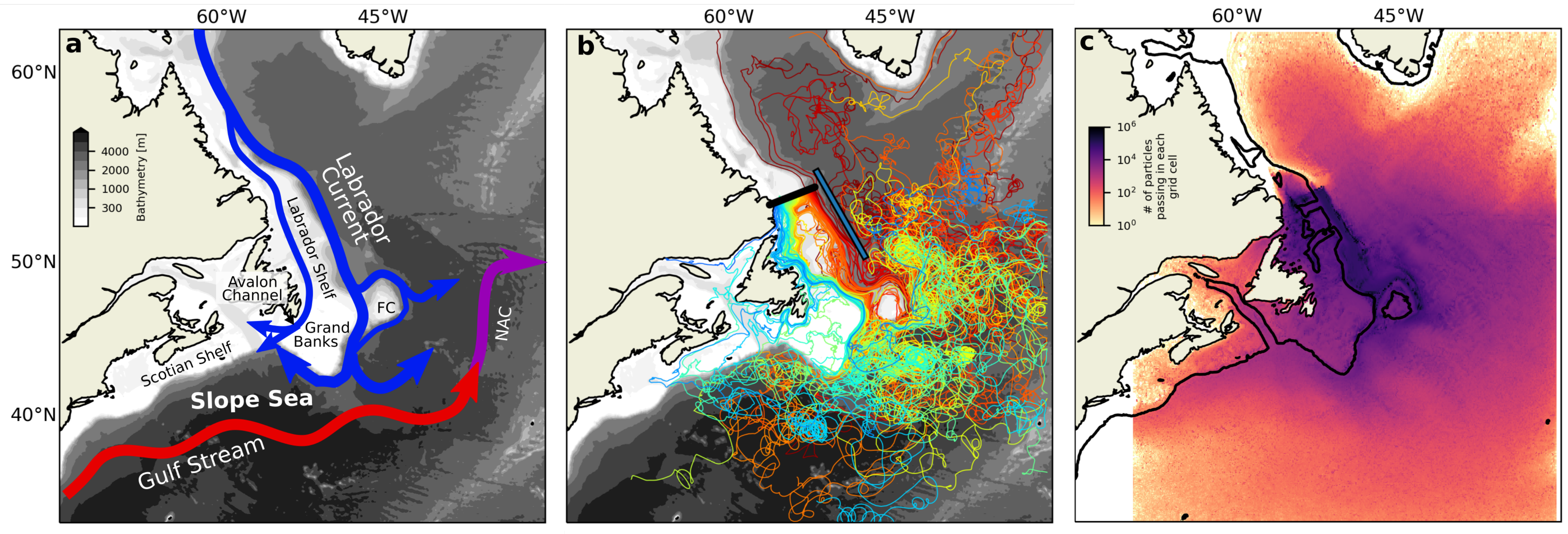
Viikmäe, B., Torsvik, T., & Soomere, T. (2013). Impact of horizontal eddy diffusivity on Lagrangian statistics for coastal pollution from a major marine fairway Topical Collection on the 16th biennial workshop of the Joint Numerical Sea Modelling Group (JONSMOD) in Brest, France 21-23 May 2012. *Ocean Dynamics*, *63*(5), 589–597. doi: 10.1007/s10236-013-0615-3

Whitney, N. M., Wanamaker, A. D., Ummenhofer, C. C., Cresswell-clay, N., Kreutz, K. J., & Johnson, B. J. (2022). Rapid 20th century warming reverses 900-year cooling in the Gulf of Maine. *Nature Communications Earth & Environment*, *3*(179), 1–15. doi: 10.1038/s43247-022-00504-8

Wichmann, D., Kehl, C., Dijkstra, H. A., & Van Sebille, E. (2021). Ordering of trajectories reveals hierarchical finite-time coherent sets in Lagrangian particle data: Detecting Agulhas rings in the South Atlantic Ocean. *Nonlinear Processes in Geophysics*, *28*(1), 43–59. doi: 10.5194/npg-28-43-2021

Williams, J., Tremblay, B., Newton, R., & Allard, R. (2016). Dynamic preconditioning of the minimum september sea-ice extent. *Journal of Climate*, *29*(16), 5879–5891. doi: 10.1175/JCLI-D-15-0515.1

Figure 1.

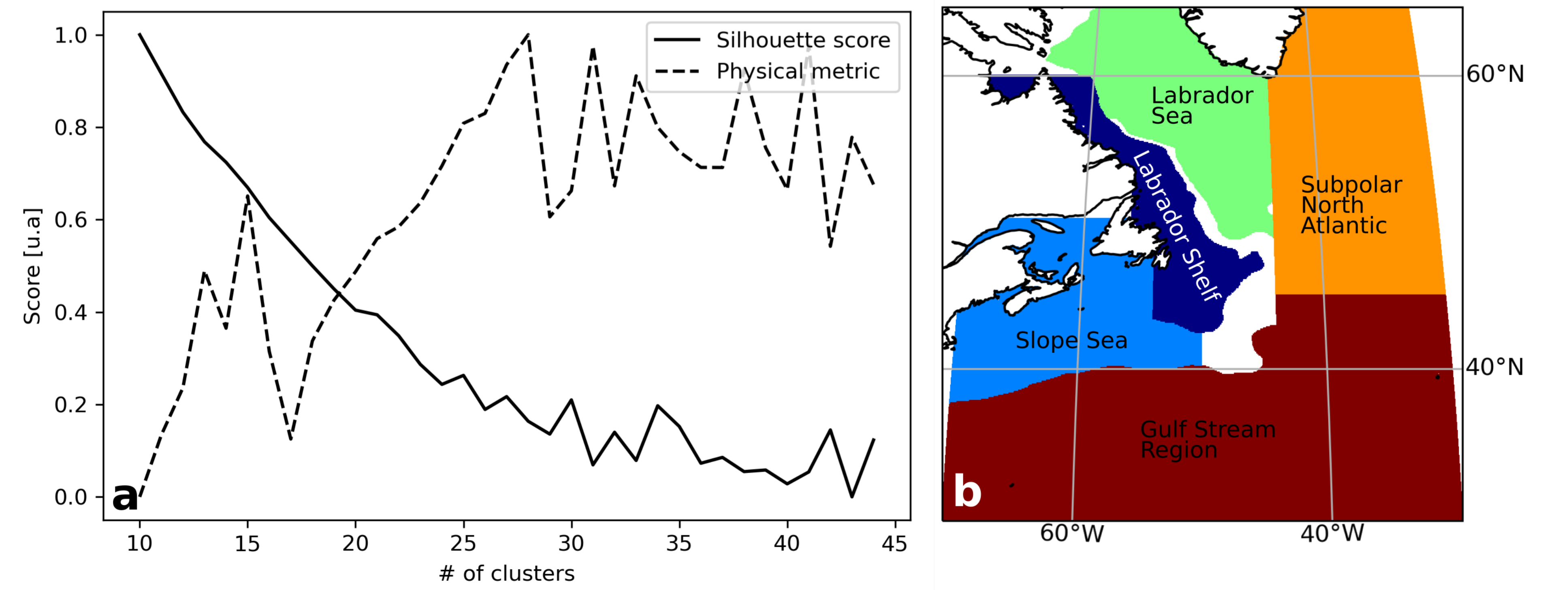


60°W
45°W
60°N
50°N
40°N
a
Bathymetry [m]
4000
2000
1000
300
Labrador Current
Labrador Shelf
Avalon Channel
Grand Banks
FC
NAC
Scotian Shelf
Slope Sea
Gulf Stream
b
c
# of particles passing in each grid cell
10^6
10^4
10^2
10^0

Figure 2.

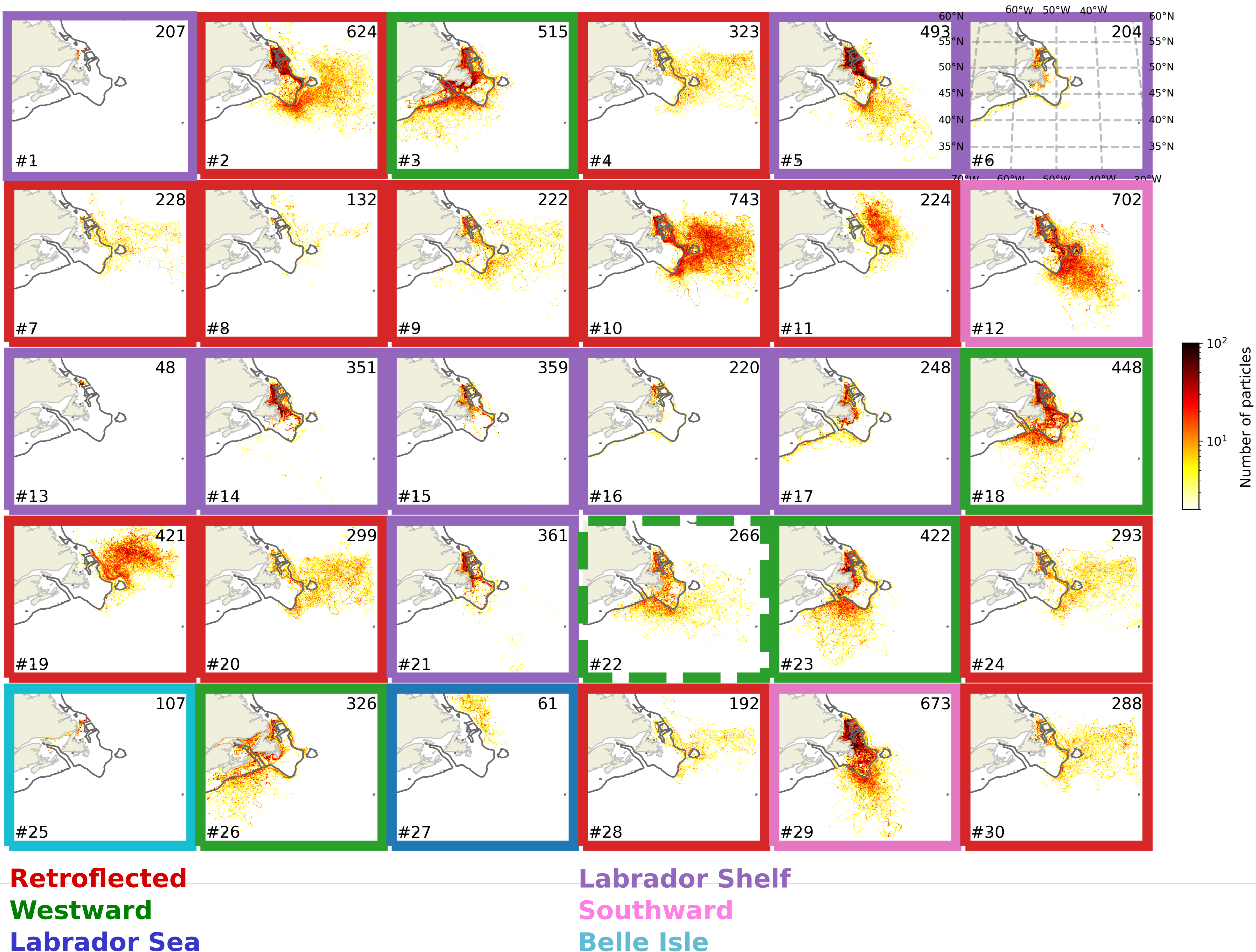
DATASET PREPARATION
PRE-PROCESSING
PROCESSING
POST-PROCESSING
PCA 1
PCA 2
kernalized
k-means
complete dataset
(1.2 M trajectories)
training set
(80 000 trajectories)
translation
cartesian projection
kernalized
PCA reduction
30 clusters
7 pathway categories

Figure 3.



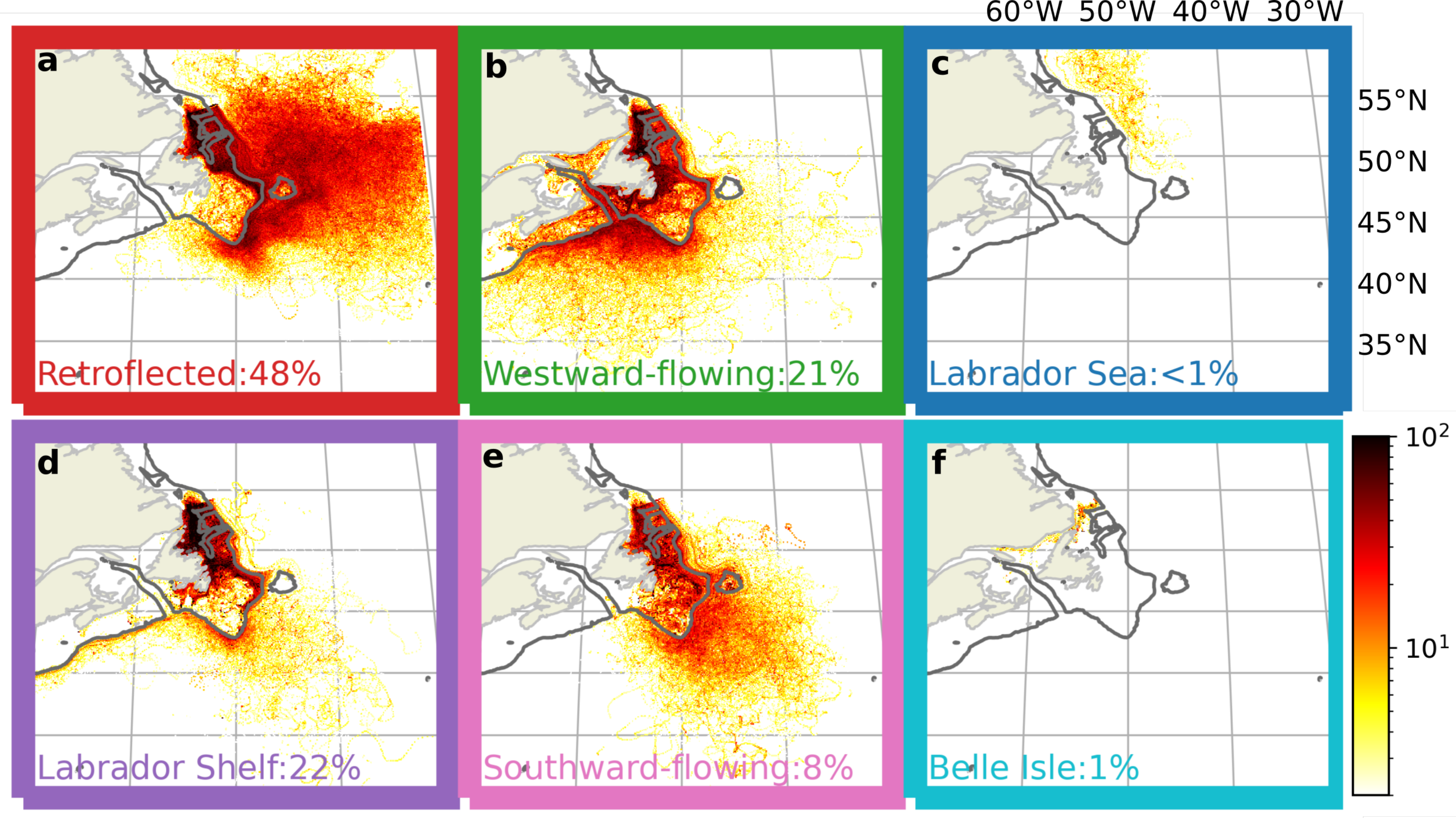
Silhouette score
Physical metric
Score [u.a]
# of clusters
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
35
40
45
a
Labrador Sea
Labrador Shelf
Subpolar North Atlantic
Slope Sea
Gulf Stream Region
60°N
40°N
60°W
40°W
b

Figure 4.



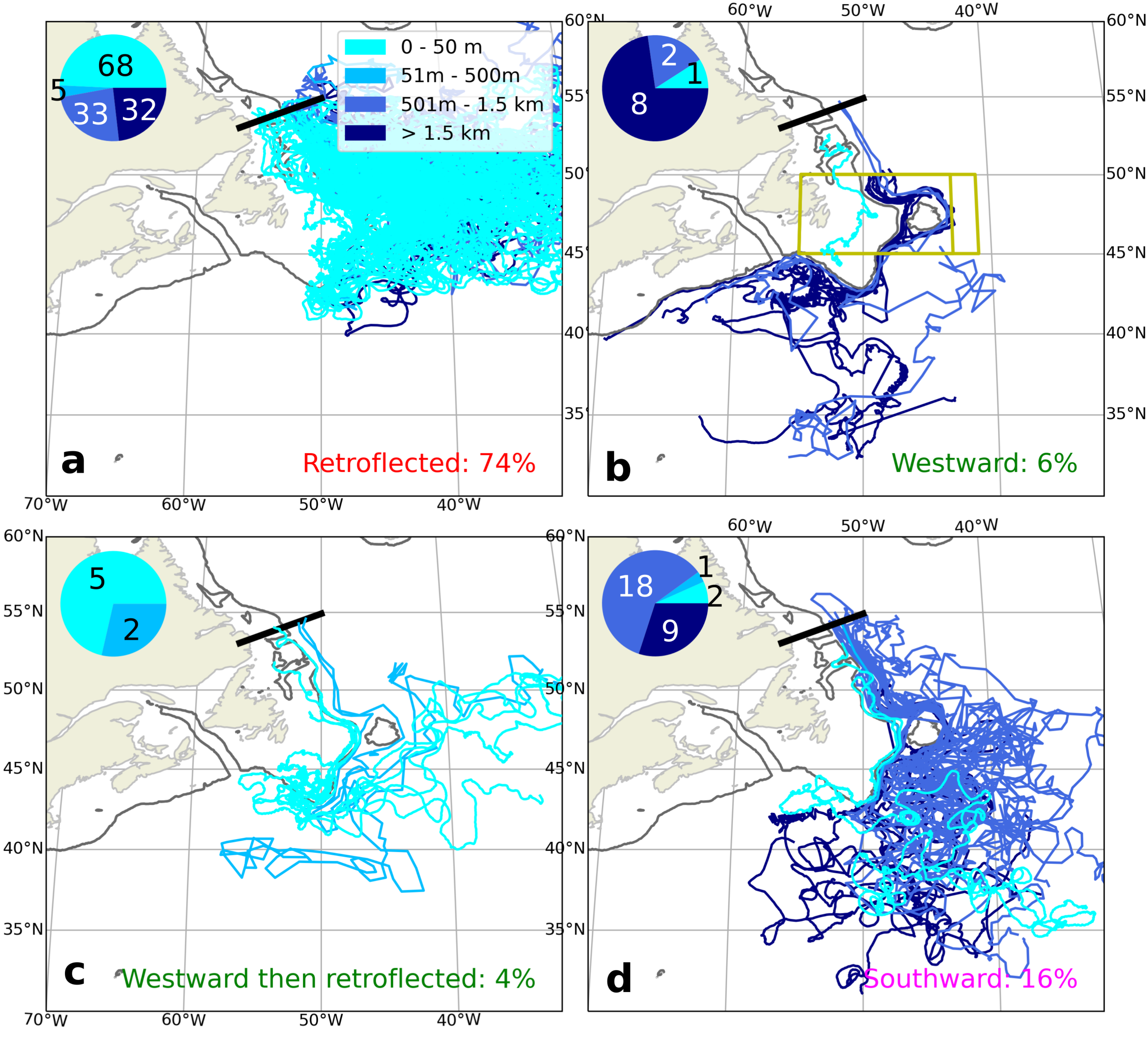
207
#1
624
#2
515
#3
323
#4
493
#5
204
#6
60°W 50°W 40°W
60°N
55°N
50°N
45°N
40°N
35°N
228
#7
132
#8
222
#9
743
#10
224
#11
702
#12
48
#13
351
#14
359
#15
220
#16
248
#17
448
#18
421
#19
299
#20
361
#21
266
#22
422
#23
293
#24
107
#25
326
#26
61
#27
192
#28
673
#29
288
#30
Number of particles
10^2
10^1
Retroflected
Westward
Labrador Sea
Labrador Shelf
Southward
Belle Isle

Figure 5.



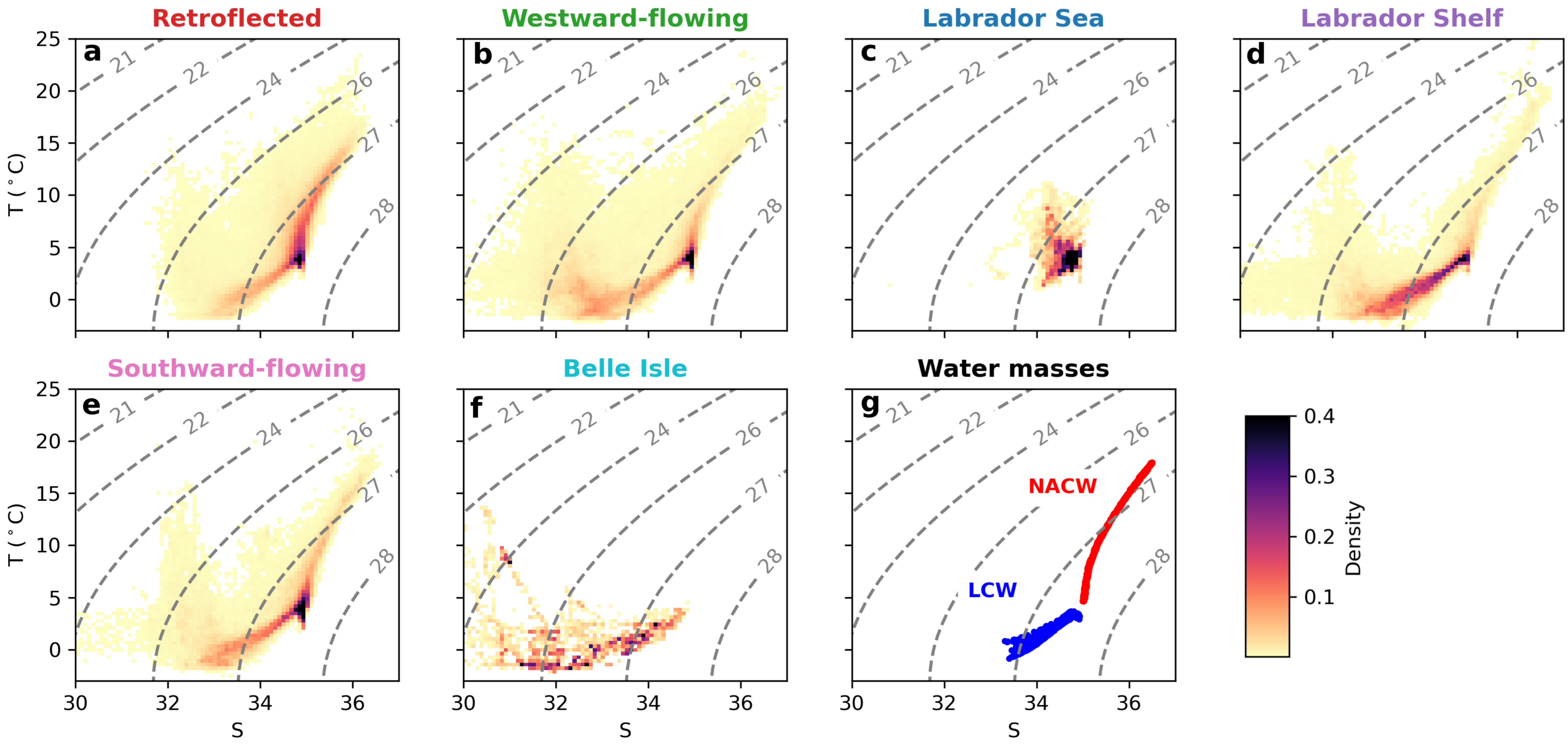
60°W 50°W 40°W 30°W
55°N
50°N
45°N
40°N
35°N
a
Retroflected:48%
b
Westward-flowing:21%
c
Labrador Sea:<1%
d
Labrador Shelf:22%
e
Southward-flowing:8%
f
Belle Isle:1%
10^2
10^1

Figure 6.



68
5
33
32
0 - 50 m
51m - 500m
501m - 1.5 km
> 1.5 km
a
Retroflected: 74%
70°W
60°W
50°W
40°W
2
1
8
b
Westward: 6%
60°N
55°N
50°N
45°N
40°N
35°N
5
2
c
Westward then retroflected: 4%
18
1
2
9
d
Southward: 16%

Figure 7.



Retroflected
Westward-flowing
Labrador Sea
Labrador Shelf
Southward-flowing
Belle Isle
Water masses
a
b
c
d
e
f
g
21
22
24
26
27
28
T ($^\circ$C)
S
NACW
LCW
Density

Figure 8.

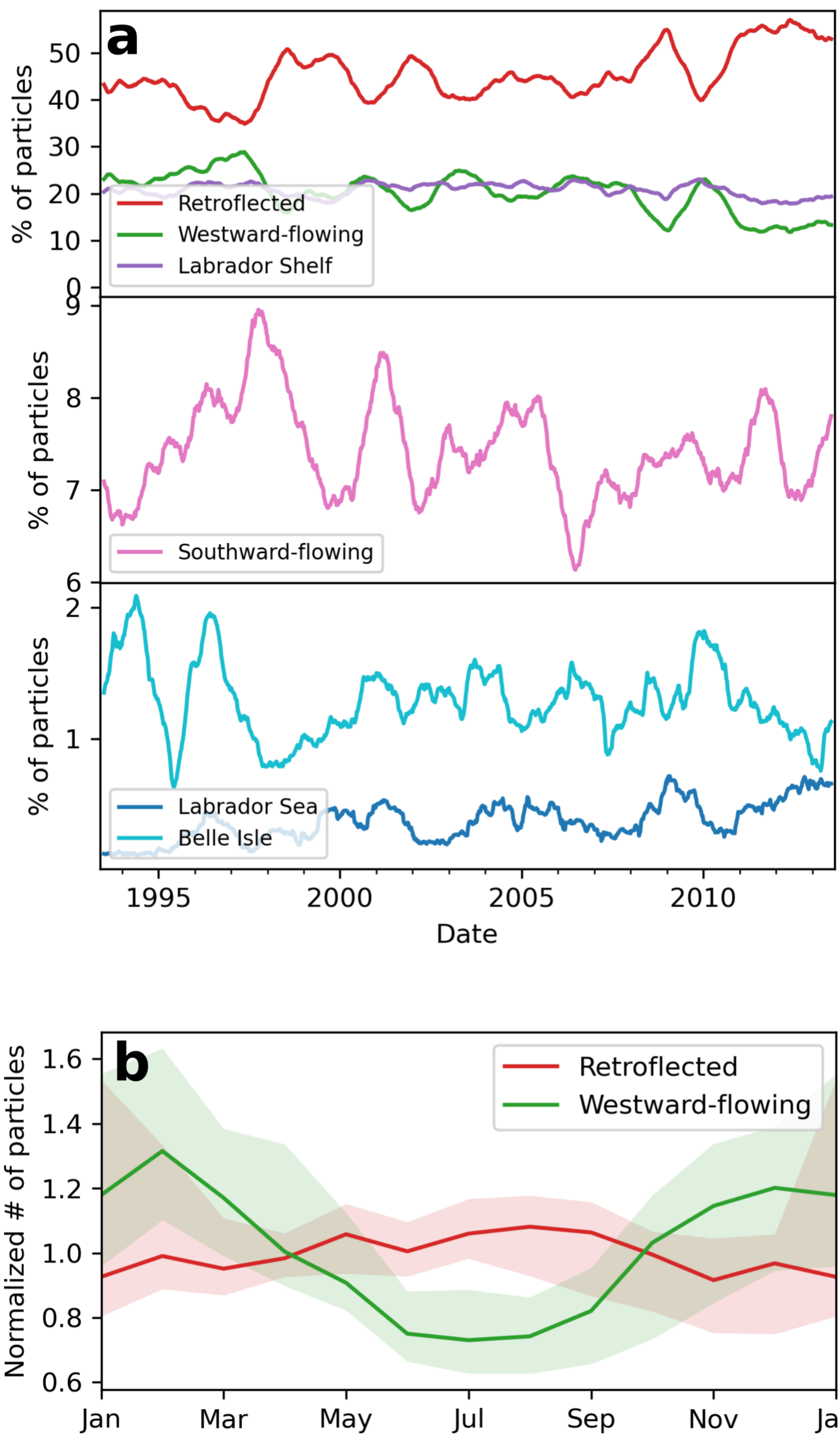
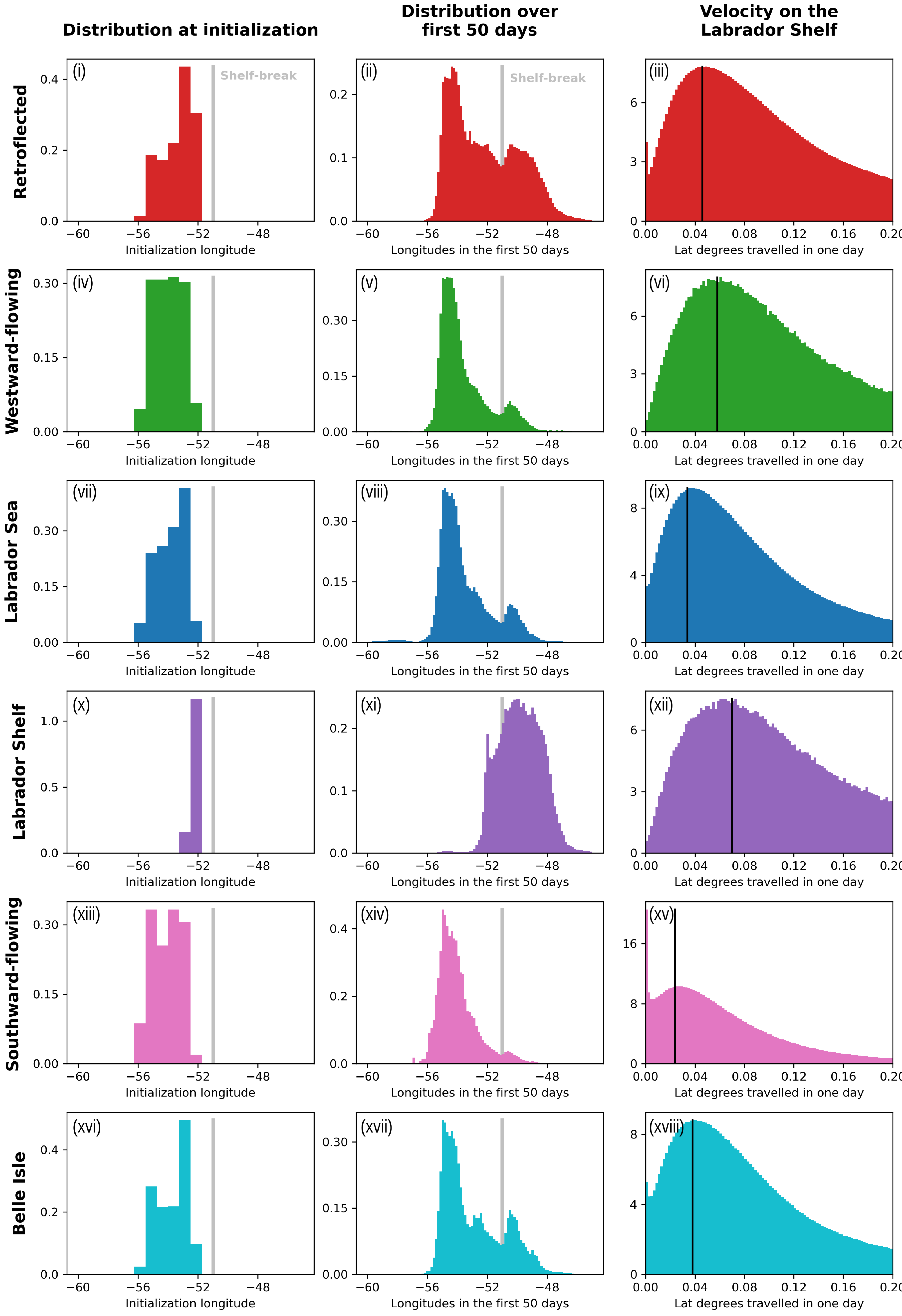
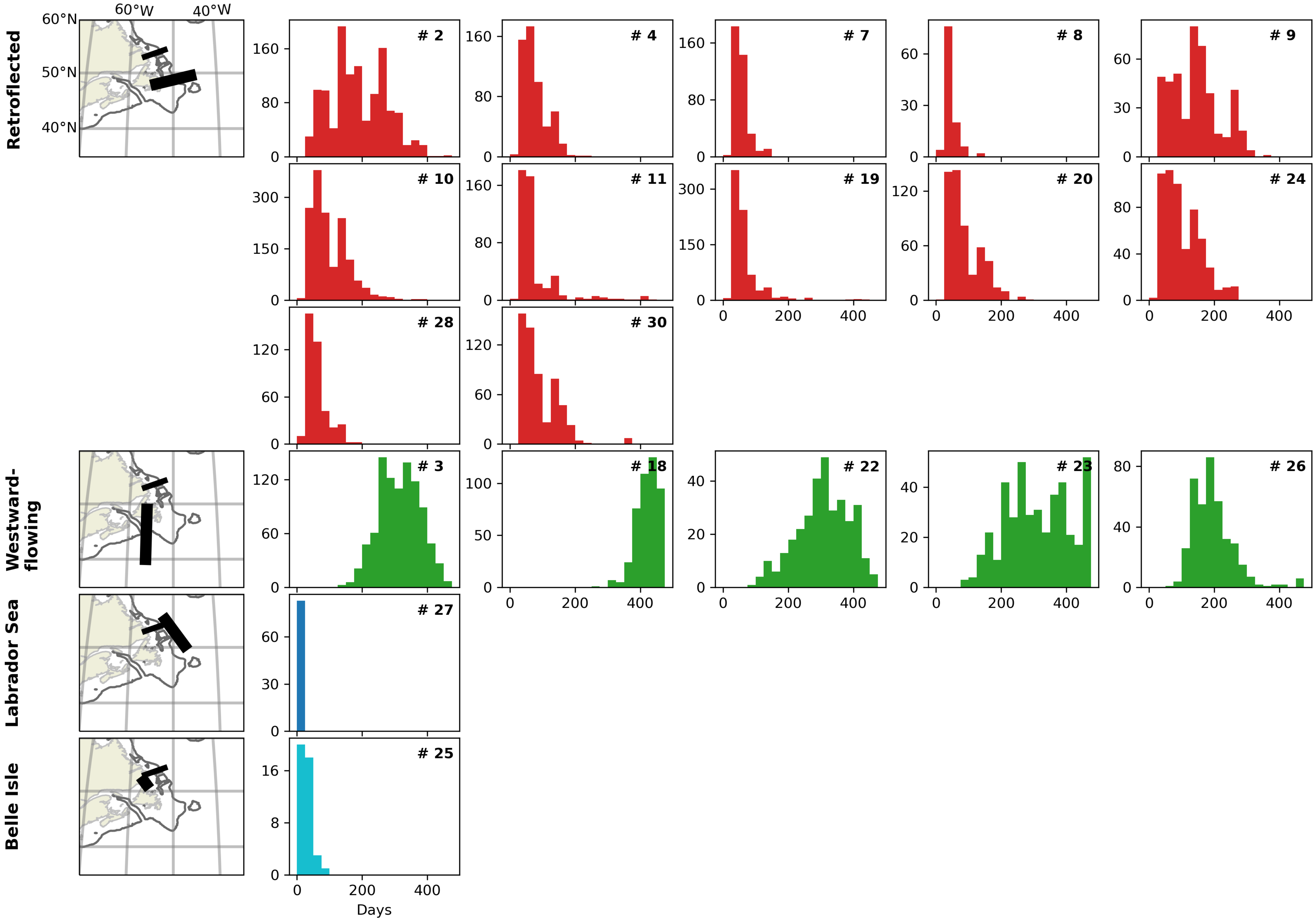

a
% of particles
50
40
30
20
10
0
Retroflected
Westward-flowing
Labrador Shelf
9
8
7
6
Southward-flowing
2
1
Labrador Sea
Belle Isle
1995
2000
2005
2010
Date
b
Normalized # of particles
1.6
1.4
1.2
1.0
0.8
0.6
Retroflected
Westward-flowing
Jan
Mar
May
Jul
Sep
Nov
Jan

Figure 9.



Distribution at initialization
Distribution over first 50 days
Velocity on the Labrador Shelf
Retroflected
Westward-flowing
Labrador Sea
Labrador Shelf
Southward-flowing
Belle Isle
Shelf-break
Initialization longitude
Longitudes in the first 50 days
Lat degrees travelled in one day
(i)
(ii)
(iii)
(iv)
(v)
(vi)
(vii)
(viii)
(ix)
(x)
(xi)
(xii)
(xiii)
(xiv)
(xv)
(xvi)
(xvii)
(xviii)

Figure 10.



Retroflected
Westward-flowing
Labrador Sea
Belle Isle
60°N
50°N
40°N
60°W
40°W
2
4
7
8
9
10
11
19
20
24
28
30
3
18
22
23
26
27
25
Days

Figure 11.

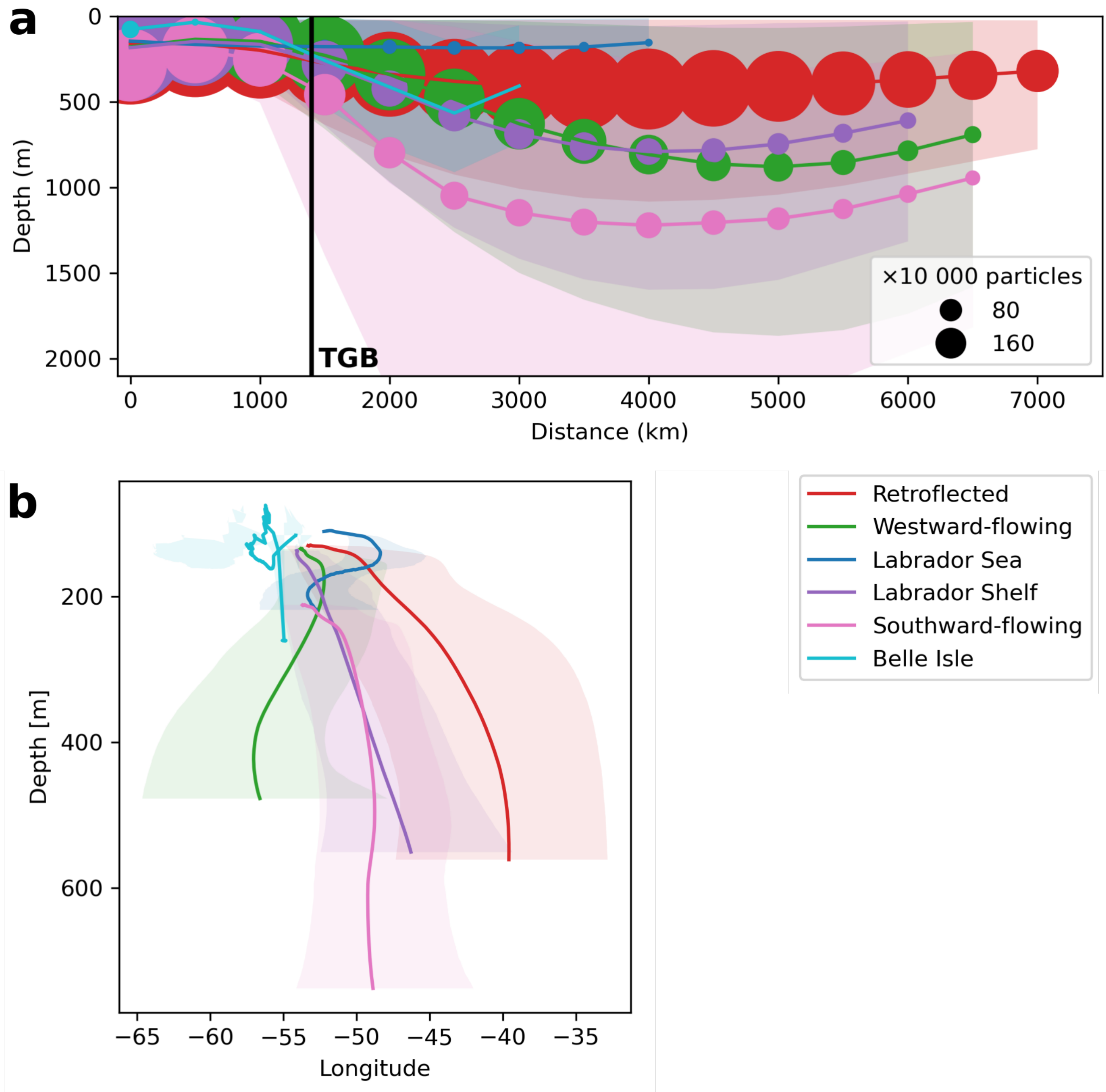
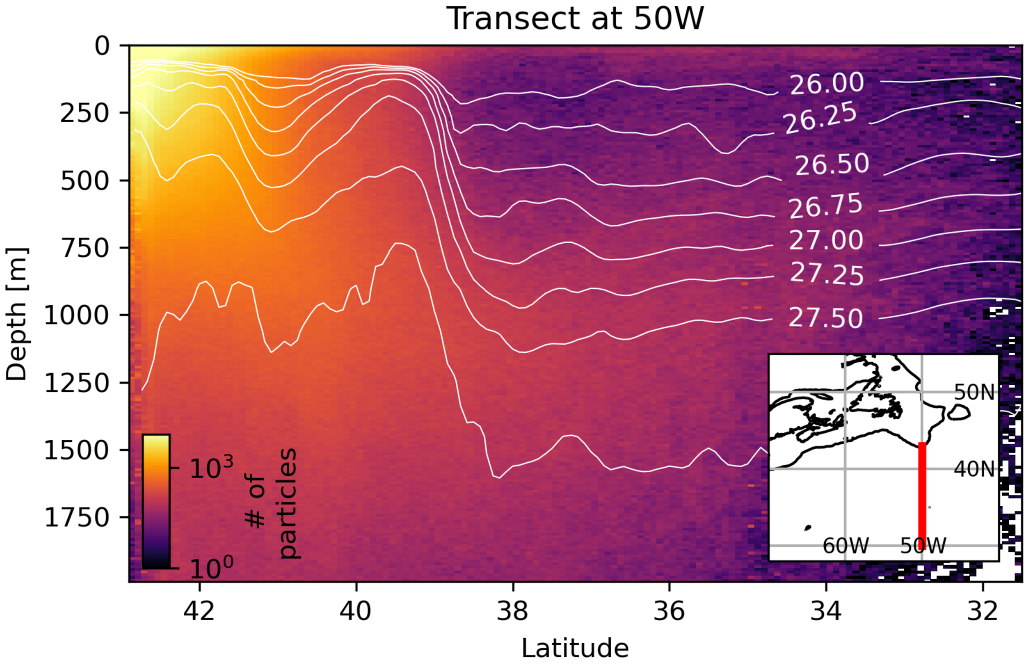

a
b
Depth (m)
Distance (km)
TGB
×10 000 particles
80
160
Depth [m]
Longitude
Retroflected
Westward-flowing
Labrador Sea
Labrador Shelf
Southward-flowing
Belle Isle

Figure 12.



Transect at 50W
Depth [m]
0
250
500
750
1000
1250
1500
1750
Latitude
42
40
38
36
34
32
26.00
26.25
26.50
26.75
27.00
27.25
27.50
10^3
10^0
# of particles
50N
40N
60W
50W

Figure 13.

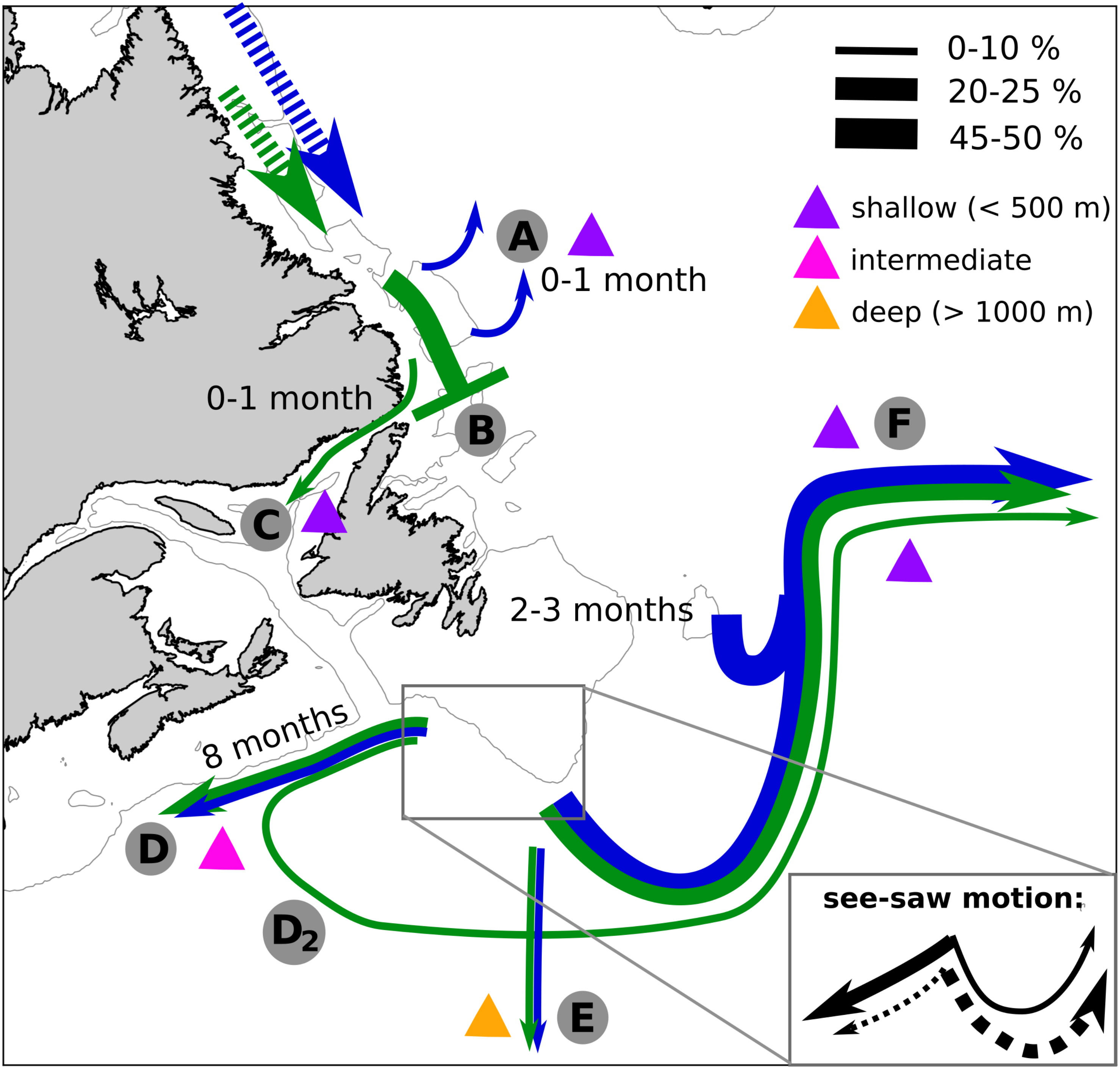

0-10 %
20-25 %
45-50 %
shallow (< 500 m)
intermediate
deep (> 1000 m)
A
0-1 month
0-1 month
B
C
F
2-3 months
8 months
D
D2
E
see-saw motion:

Figure A1.

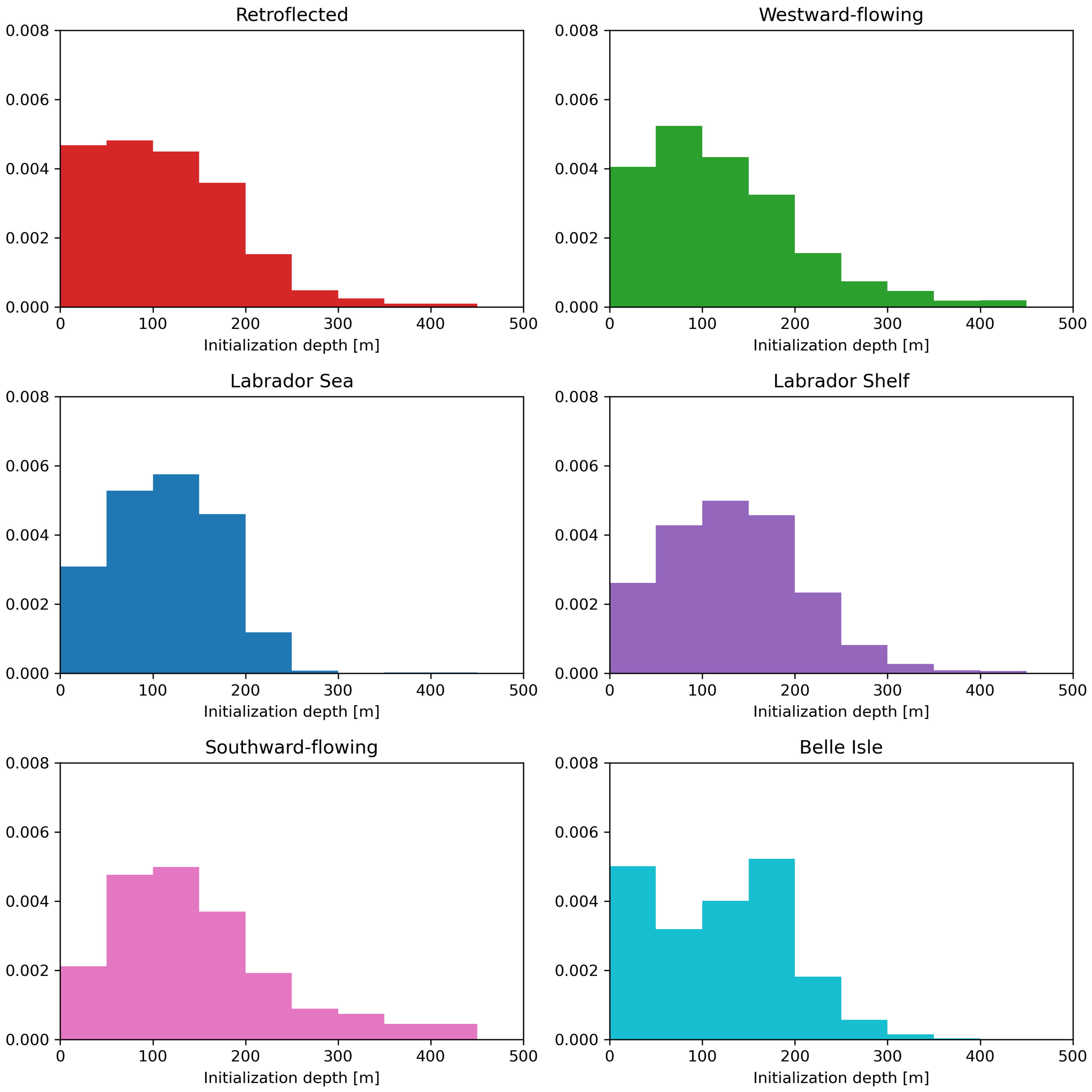

Retroflected
Westward-flowing
Labrador Sea
Labrador Shelf
Southward-flowing
Belle Isle
0.008
0.006
0.004
0.002
0.000
0
100
200
300
400
500
Initialization depth [m]

Figure A2.

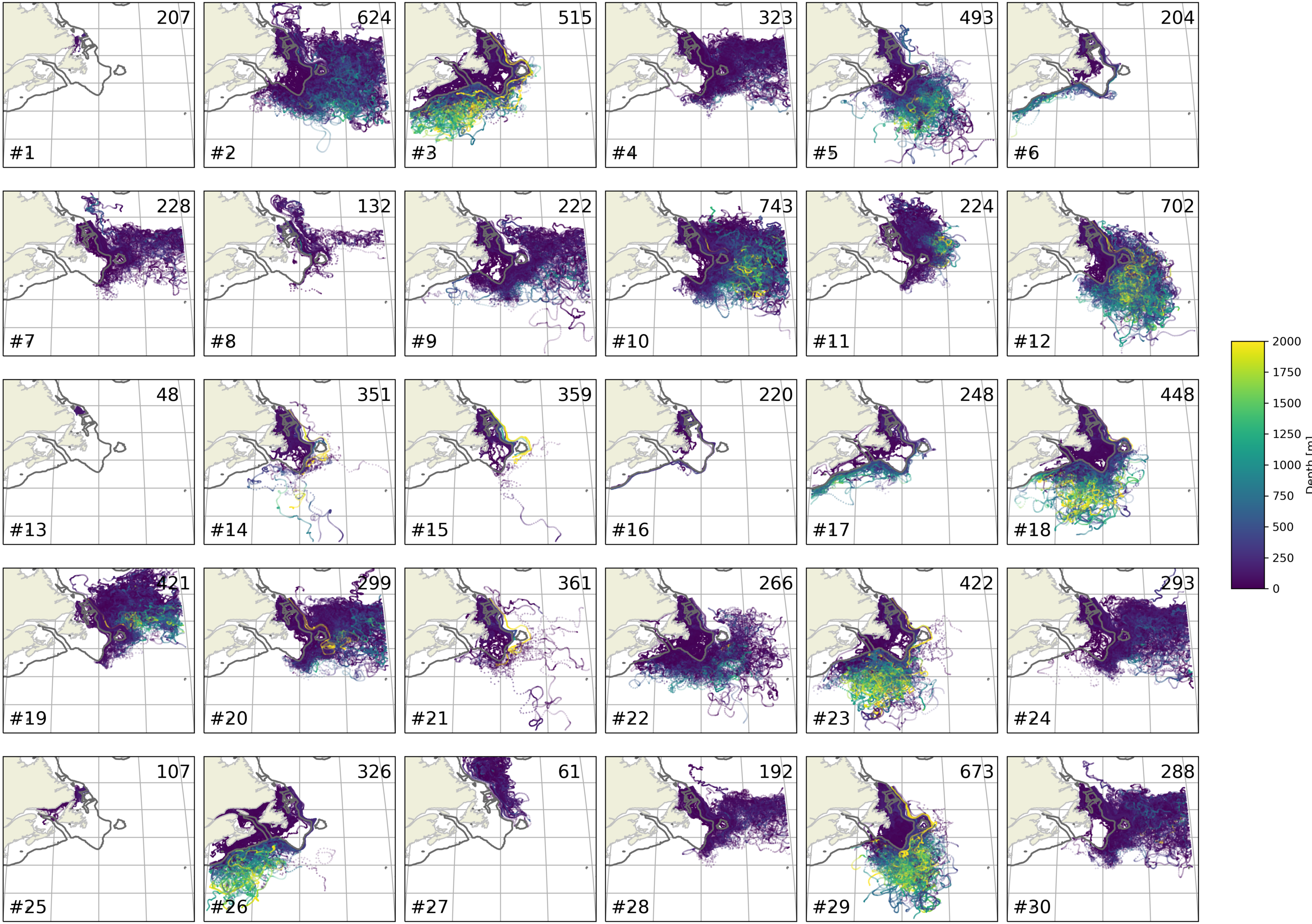
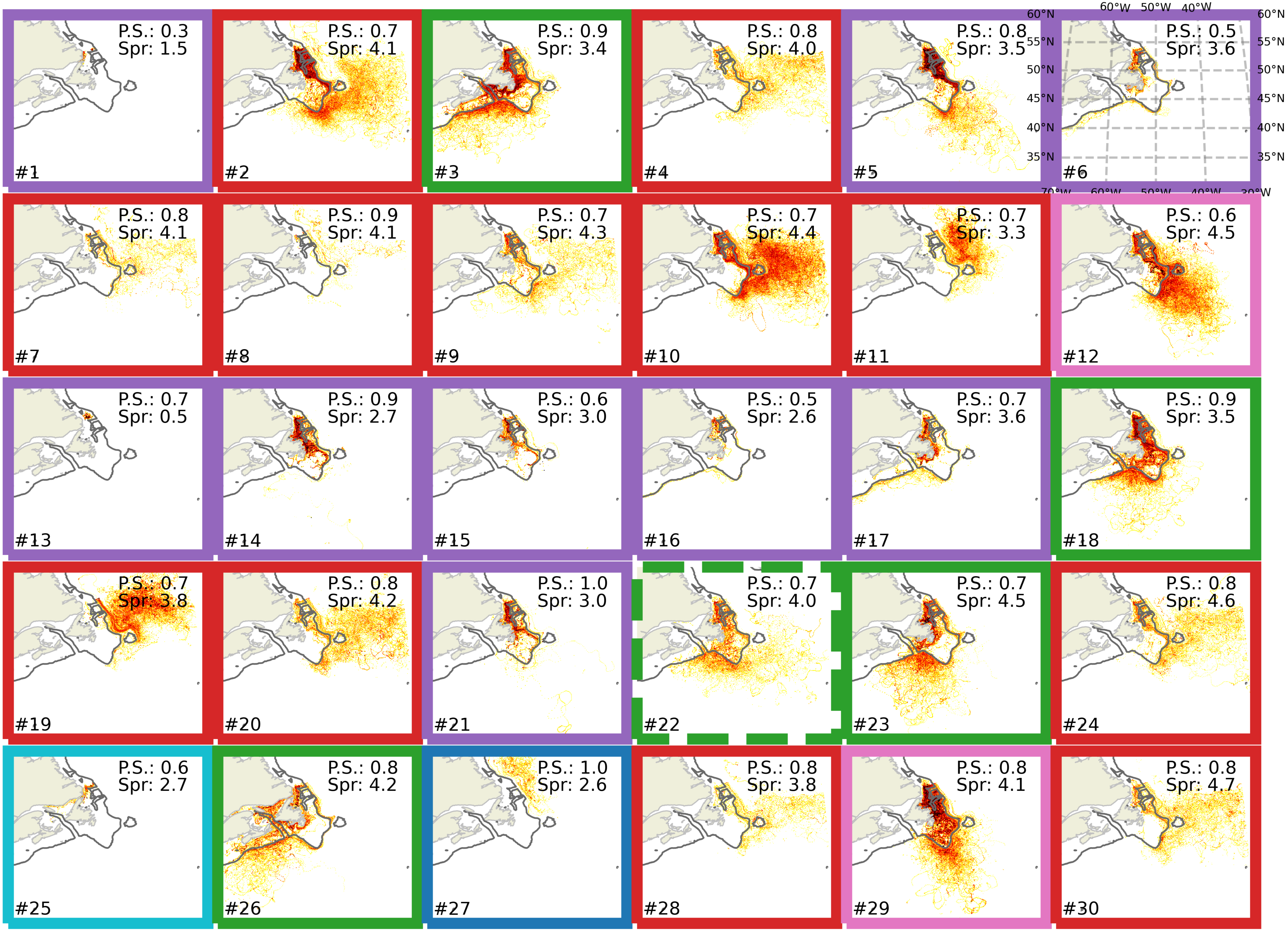

207
#1
624
#2
515
#3
323
#4
493
#5
204
#6
228
#7
132
#8
222
#9
743
#10
224
#11
702
#12
48
#13
351
#14
359
#15
220
#16
248
#17
448
#18
421
#19
299
#20
361
#21
266
#22
422
#23
293
#24
107
#25
326
#26
61
#27
192
#28
673
#29
288
#30
2000
1750
1500
1250
1000
750
500
250
0
Depth [m]

Figure A3.



P.S.: 0.3
Spr: 1.5
#1
P.S.: 0.7
Spr: 4.1
#2
P.S.: 0.9
Spr: 3.4
#3
P.S.: 0.8
Spr: 4.0
#4
P.S.: 0.8
Spr: 3.5
#5
P.S.: 0.5
Spr: 3.6
#6
60°W 50°W 40°W
60°N
55°N
50°N
45°N
40°N
35°N
70°W 60°W 50°W 40°W 30°W
P.S.: 0.8
Spr: 4.1
#7
P.S.: 0.9
Spr: 4.1
#8
P.S.: 0.7
Spr: 4.3
#9
P.S.: 0.7
Spr: 4.4
#10
P.S.: 0.7
Spr: 3.3
#11
P.S.: 0.6
Spr: 4.5
#12
P.S.: 0.7
Spr: 0.5
#13
P.S.: 0.9
Spr: 2.7
#14
P.S.: 0.6
Spr: 3.0
#15
P.S.: 0.5
Spr: 2.6
#16
P.S.: 0.7
Spr: 3.6
#17
P.S.: 0.9
Spr: 3.5
#18
P.S.: 0.7
Spr: 3.8
#19
P.S.: 0.8
Spr: 4.2
#20
P.S.: 1.0
Spr: 3.0
#21
P.S.: 0.7
Spr: 4.0
#22
P.S.: 0.7
Spr: 4.5
#23
P.S.: 0.8
Spr: 4.6
#24
P.S.: 0.6
Spr: 2.7
#25
P.S.: 0.8
Spr: 4.2
#26
P.S.: 1.0
Spr: 2.6
#27
P.S.: 0.8
Spr: 3.8
#28
P.S.: 0.8
Spr: 4.1
#29
P.S.: 0.8
Spr: 4.7
#30