Topological Feature Tracking for Submesoscale Eddies

Sam Voisin[1], Jay Hineman[1], James B. Polly[1], Gary Koplik[1], Ken Ball[1], Paul Bendich[1,4], Joseph D'Addezio[2], Gregg A. Jacobs[2], Tamay Özgökmen[3]

[1]Geometric Data Analytics, Inc., Durham, NC, USA
[2]Ocean Dynamics and Prediction, Naval Research Laboratory, Stennis Space Center, MS, USA
[3]Rosenstiel School of Marine and Atmospheric Science, University of Miami, FL, USA
[4]Department of Mathematics, Duke University, Durham, NC, USA

Key Points:

- Topological Feature Tracking (TFT) is introduced as a way to identify features in scalar fields and associate those features through time.
- We identify and track submesoscale eddies over 1-year of ocean surface velocity data computed via the Navy Coastal Ocean Model.
- Eddy statistics provide insight on lifetime, speed, and distance traversed for understanding eddy motions and scale interactions.

Corresponding author: James B. Polly, `james.polly@geomdata.com`

Abstract

Current state-of-the art procedures for studying modeled submesoscale oceanographic features have made a strong assumption of independence between features identified at different times. Therefore, all submesoscale eddies identified in a time series were studied in aggregate. Statistics from these methods are illuminating but oversample identified features and cannot determine the lifetime evolution of the transient submesoscale processes. To this end, the authors apply the Topological Feature Tracking (TFT) algorithm to the problem of identifying and tracking submesoscale eddies over time. TFT identifies critical points on a set of time-ordered scalar fields and associates those points between consecutive timesteps. The procedure yields tracklets which represent spatio-temporal displacement of eddies. In this way we study the time-dependent behavior of submesoscale eddies, which are generated by a 1-km resolution submesoscale-permitting model. We summarize the submesoscale eddy dataset produced by TFT, which yields unique, time-varying statistics.

Plain Language Summary

Current state-of-the art procedures for studying small-scale features in the ocean do not take the effects of time into account. Instead, features like small vortices are studied as a single population across many points in time. This method has provided oceanographers with many valuable insights. New insights can be added by identifying vortices and then tracking them over time to study their behavior through an algorithm designed to identify and track features on a grid.

1 Introduction

Submesoscale eddies occupy length scales between large-scale forcings and micro-scale dissipation. Their larger, mesoscale counterparts are well-studied, yet submesoscale currents have, until recently, received less attention despite their importance. In addition to influencing the transport of nutrients (Lévy et al., 2018) and pollutants (Poje et al., 2014), submesoscale currents form an important link in the turbulent energy cascade and the global oceanic circulation (see McWilliams, 2016, for a summary of submesoscale eddy dynamical theory, observational findings, and modeling approaches).

Studies considering the temporal evolution of mesoscale eddies have been performed (e.g., Chelton et al., 2007; Kurian et al., 2011; Faghmous et al., 2015), but similar investigations have yet to be done for the submesoscale. While dissipation-scale phenomena are typically unresolved and parameterized with subgrid-scale closure models, the "intermediate" length scales occupied by submesoscale eddies are able to be resolved in models such as the Navy Coastal Ocean Model (NCOM; Barron et al., 2006), among others. In these models, time tracking and statistical reporting of submesoscale eddies is not currently done but would be useful for model evaluation, e.g., inspecting performance of eddy viscosity and parameterized closure schemes. The method we describe herein permits comparison of eddy statistics between model-generated and observational data. Our algorithm provides a tool to (among other things) evaluate eddy dissipation by providing lifetime metrics of these features in a similar (but more automated) way that Liu et al. (2021) found that horizontal model resolution was correlated with overestimation of vertical velocities. Furthermore, statistical summaries of transient submesoscale eddy behavior is needed for satellite altimetry data assimilation efforts (D'Addezio et al., 2019) and has motivated the statistical investigations in D'Addezio et al. (2020).

In this study we apply the algorithm (henceforth referred to as Topological Feature Tracking, or TFT) introduced in Soler et al. (2018) to the problem of submesoscale eddy identification and temporal association. In this way, we extend the study of D'Addezio

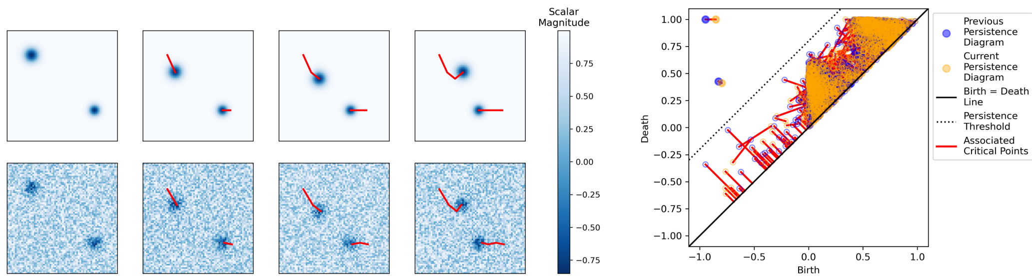

Figure 1: Illustration of TFT algorithm on a notional example: Left: Tracking two Gaussian features on a time-ordered series of scalar fields. Right: Matching between persistence diagrams (blue dots and orange dots) associated to scalar fields (bottom row) at $t = 2, 3$, respectively.

et al. (2020) by computing statistics of eddy lifetimes and trajectories to supplement the time-independent statistical analysis presented therein. Using one year of NCOM simulation data, we provide statistical summaries of eddy speed, lifespan, and displacement in aggregate over the Gulf of Mexico. We also provide analysis of these characteristics conditioned on season and regions selected for the presence of mesoscale features. While extending the technique used in D'Addezio et al. (2020) with the TFT-based method, we are introducing the community to the TFT approach in the context of surface-based submesoscale eddies.

2 Method

In this section we give a brief description of the TFT algorithm (Section 2.2), along with the elementary topological data analysis (TDA) concepts needed to understand it (Section 2.1). For more details on TFT and TDA in general, see Soler et al. (2018) and Edelsbrunner and Harer (2010), respectively. Finally, we describe the Okubo–Weiss parameter used to generate the scalar fields to which we apply TFT (Section 2.3).

2.1 Persistence Diagrams

Suppose that f is a *scalar field*, that is, a real-valued function on some domain U. The domain can be of arbitrary dimension and shape and we need no assumptions about the smoothness of f. For a working example, suppose U is any of the two-dimensional squares shown on the left side of Figure 1, with the values of f indicated by the color bar. The *persistence diagrams* of f provide a compact summary of the location and importance of topological features as observed by f. More precisely, consider $U_\alpha = \{x \in U \mid f(x) \leq \alpha\}$. As the threshold value α increases, these create a nested *filtration* of *sublevel sets* that start with the empty set and finish with U itself. Along the way, topological features (*connected components* and *holes*) are created and then subsequently destroyed, each of which corresponds (Milnor, 1963) to a *critical point* of f that occurs at a *critical value*. The birth and death critical values of each feature are plotted as dots in the plane, and the multi-set of such dots, along with the major diagonal $y = x$, forms the persistence diagram $D(f)$ of the scalar field. Two such diagrams can be seen on the right side of Figure 1, where blue (orange) dots correspond to features in the scalar fields in the second (third) columns, bottom row. The *persistence* of a dot is the difference between its death and birth values (i.e, the vertical distance to the major diagonal). Higher-persistence dots tend to be less likely to be noise. For example, all of the example scalar

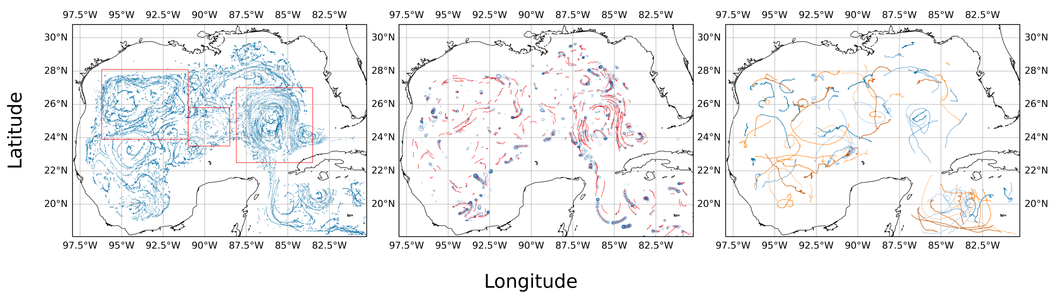

Figure 2: Examples of TFT applied to the masked O.–W. dataset. Left to right: (1) Submesoscale eddies identified in the period 2016 January 1–18, depicted as blue points in the Gulf of Mexico. Zones 1, 2 and 3 (west to east) are defined here and referenced in the text. (2) Submesoscale eddies tracked via TFT, where blue solid line contours are eddies identified at January 5, 2016 03:00, dotted blue line contours depict eddy locations over the previous five days, and the corresponding eddy tracks are shown in red. This subset depicts only tracks of 25km or longer. (3) Selection of tracks of eddies lasting for 15 days or more. These relatively long lived tracks demonstrate both the meandering nature of the eddy, and the persistent tracking capability of TFT.

fields have two prominent connected components indicated by the two dots far from the major diagonal.

Persistence diagrams have two important properties that we exploit. First, they are *stable to noise* in a precise sense. The *Wasserstein distance* between two diagrams can be defined as the cost of an optimal matching between the dots in the diagrams, where dots can be matched to the major diagonal if needed; the right side of Figure 1 shows an optimal matching. Precise theorems (Cohen-Steiner et al., 2007) bound the Wasserstein distance between two diagrams $D(f), D(g)$ in terms of the ℓ_∞ distance between the scalar fields f, g. In particular, this guarantees that the diagrams associated to a smoothly time-varying sequence of scalar fields will themselves form a time-varying sequence, which facilitates the TFT algorithm. Second, various theorems (Edelsbrunner et al., 2006; Laudenbach, 2013) guarantee the following: given a two-dimensional scalar field f and a threshold value ϵ, there exists a simplified scalar field g with exactly the same critical point structure of f except that all critical points of persistence less than ϵ have been removed. For example, with ϵ being the distance between the major diagonal and the dotted line on the right side of Figure 1, the scalar fields in the top row on the left are the topological simplifications of the scalar fields in the bottom row.

2.2 Topological Feature Tracking

Now suppose that we have a time-ordered sequence $f_1, \ldots f_T$ of scalar fields, such as the four fields across either row on the left of Figure 1, all defined on the same domain U. Computing persistence leads to a time-ordered sequence $D(f_1), \ldots, D(f_T)$ of persistence diagrams. The user may choose a persistence threshold to topologically simplify the scalar fields as desired. Then the TFT algorithm connects critical points to produce a series of *tracks*, as follows.

Consider a time-adjacent pair of (possibly simplified) scalar fields f_i and f_{i+1}. Each dot in the two diagrams corresponds to a topological feature, and has an associated pair of critical points in U, one which created the feature and one which destroyed it. The *lifted Wasserstein* distance of Soler et al. (2018) defines the cost of associating two dots in $D(f_i)$ and $D(f_{i+1})$ as a (user-specified) weighted combination of the distance between the pair of dots in the persistence diagram and the geometric distance between the associated critical points in the domain U. An optimal matching between the two diagrams is then computed via this cost function. If this optimal matching connects two dots, a track segment is drawn between their associated critical points. If it connects a dot at time i with the diagonal at time $i+1$, then a track segment ends. If it connects a dot at time $i+1$ with the diagonal at time i, a new track segment is started. The end result, over all time steps in the sequence, is a set of tracks which move in time through the domain U.

Figure 1 shows the tracks for our notional example, indicated as thick red lines on the left side of the figure. Figure 2 shows tracks for submesoscale eddies, identified by the same procedure and further described in the following sections.

The matching procedure described above must be applied to each consecutive pair of persistence diagrams in the time series. Computationally, this may be done in parallel so long as the time order is maintained. Once matching is completed for all consecutive time steps, the matchings of associated critical pairs may be applied to coordinates in the domain to combine the track segments and form full tracks of the identified features.

2.3 Okubo–Weiss Parameter

The Okubo–Weiss (O.–W.) parameter is one of many dynamical quantities used to define eddies and has been utilized in numerous studies (see Isern-Fontanet et al., 2003; Kurian et al., 2011; D'Addezio et al., 2020 and references therein). Aside from the well-established use of O.–W., we use this quantity to identify eddies because TFT utilizes information at critical points to calculate persistent homology and simplify noisy scalar fields, making the O.–W. parameter more suitable than those where a gradient (rather than a critical point) is associated with the feature of interest. Aditionally, D'Addezio et al. (2020) utilized O.–W. for eddy identification, and the extension of that work presented herein maintains this approach for consistency.

The O.–W. parameter is defined as

$$W = S_n^2 + S_s^2 - \zeta^2 \tag{1}$$

where S_n, S_s, and ζ are respectively the normal strain ($S_n = \partial_x u - \partial_y v$), shear strain ($S_s = \partial_x v + \partial_y u$), and relative vorticity ($\zeta = \partial_x v - \partial_y u$), with u and v being velocity components. When $W < 0$ the relative vorticity term overwhelms the shear terms, $|\zeta| > S_n^2 + S_s^2$, and indicates a flow dominated by rotation. By finding critical points ($\partial_i W = 0$) in the negative portion of this field, TFT can rapidly identify rotationally dominant regimes without the need to mask fields based on additional criteria, i.e., Rossby number or eddy shape, as discussed in the following section.

3 Data & Procedure

The dataset used as an input to TFT is a year-long simulation of the Gulf of Mexico generated by the Navy Coastal Ocean Model (NCOM), with three-hourly temporal resolution, spanning the 2016 calendar year. NCOM solutions have a spatial resolution of one kilometer, which *permits* submesoscale eddy generation, but does not fully resolve submesoscale dynamical features (see e.g., Capet et al., 2008 for additional discussion). Results presented herein summarize the behavior of "submesoscale" eddies permitted by a 1-km resolution model, but note that results are dependent on the resolution of the input solution data.

Two derivative datasets were generated from the NCOM simulation. The first is a replication of the dataset generated in D'Addezio et al. (2020), in which eddies are identified using closed contours of a filtered, normalized O.–W. field computed from small-scale velocities. Those identified eddies must meet criteria for O.–W. value, Rossby number, and circularity of the closed contour. All non-eddy regions are masked, and thus we call this the "masked" dataset (and see D'Addezio et al., 2020 for details). By applying TFT to the "masked" dataset, the critical point (eddy) identification step is deemphasized as the features of interest are the only suriving data in the masked fields, thus the novel TFT contribution is primarily the temporal association between timesteps.

The second dataset is a less restrictive version of the first in which the same procedure is followed until the normalized O.–W. field is generated. We refrain from applying the second smoothing filter, any circularity tests, or masking of this dataset; we therefore refer to it as "unmasked" and task the TFT algorithm to perform eddy identification as described in Section 2.2. By limiting TFT to the negative portions of the unmasked scalar fields, the algorithm identifies critical points corresponding to rotationally dominant flow structures, per Section 2.3. It is known however that submesoscale eddies are ageostrophic, i.e., $Ro \approx \zeta/f >> 1$ (where Ro is the local Rossby number and f is the Coriolis frequency; see Capet et al., 2008; Zhong & Bracco, 2013; Gula, Molemaker, & McWilliams, 2014). Unlike the masked dataset, the unmasked dataset does not impose the ageostrophic requirement.

Limiting the O.–W. field to only negative values focuses on eddies, and results in improved track quality, which is subjectively determined, e.g., by limiting the number of ephemeral tracks lasting only one or two timesteps, or eliminating temporal associativity between eddies that are spatially far apart. Note that the persistence threshold (ϵ) controls the number of critical points identified at a given timestep, and some experimentation was performed to remove noise from the Okubo–Weiss fields without removing eddies of interest.

The output of the TFT algorithm is a set of tracks representing the historical behavior of individual submesoscale eddies in the Gulf of Mexico. Two mild postprocessing routines were applied to this set of tracks. We first removed tracks which began or ended on the boundary of the Gulf of Mexico. These erroneous tracks are caused by the abrupt end of the scalar field at its edges. We also applied a filter which removed any tracks whose average speed was greater than the maximum surface speed at any point in the NCOM simulation. These tracks which have been filtered out due to excessive speeds are nonphysical, and are a numerical artifact of the temporal matching process. A subset of the resulting tracks can be seen in the middle and right images of Figure 2.

4 Results

In this section we provide insights gleaned from tracking submesoscale eddies identified in the Okubo–Weiss field. All figures correspond to results obtained from the masked O.–W. dataset. In Section 4.1 we present seasonality studies, and in Section 4.2 we provide descriptive statistics of submesoscale eddy behavior observed through tracks identified using TFT on both masked and unmasked datasets.

4.1 Identifying Seasonal Mesoscale Patterns via Submesoscale Tracks

Mesoscale features are responsible for transporting submesoscale eddies throughout the Gulf of Mexico (Zhong & Bracco, 2013; Gula et al., 2014; McWilliams, 2016). By tracking those submesoscale eddies, we also gain insight into the evolving mesoscale phenomena, as seen in Figure 3. Each frame of Figure 3 represents three months of submesoscale eddies with track length greater than or equal to 25 km in length. In winter, the greatest track density appears in the Loop Current, which by spring has split into a southern current exiting the gulf to the east, and a mesoscale eddy further north off the western coast of Florida. In summer this large eddy has moved west with less track

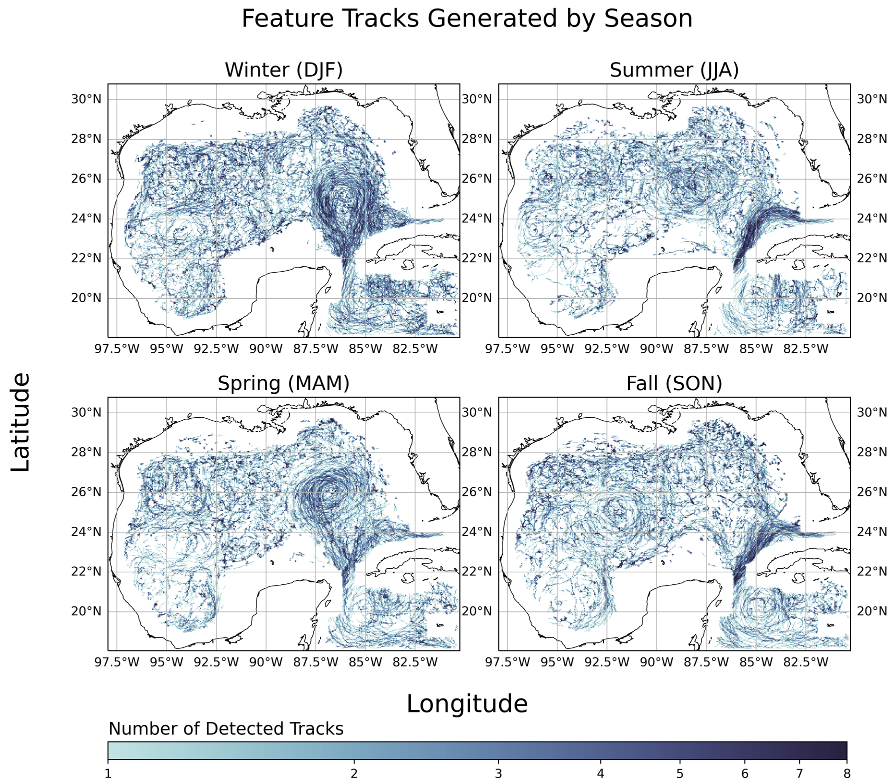

Figure 3: Illustration of submesoscale eddy behavior in aggregate over four seasons of the masked dataset. Tracks shown have been filtered to include those $\geq$ 25 km.

density, compensated by greater track density in the current to the northwest of Cuba. In the fall this large mesoscale eddy moves further west, deeper into the Gulf, while the current near Cuba carries a high density of eddies toward the Gulf Stream.

Submesoscale tracks do not follow any consistent directional pattern. Their trajectories appear to be governed by large-scale background flow, dictated primarily by both the synoptic jet and the interior mesoscale eddies. This is in contrast with the mesoscale eddy field which is known to propagate westward outside the influence of boundary currents (Chelton et al., 2007). Our results demonstrate the utility of submesoscale eddy tracks for characterizing mesoscale dynamics, such as the seasonality of the Loop Current.

To highlight seasonal differences we sum the track densities for winter and spring, and then difference that sum by the combined densities from summer and fall. This difference in track density is shown in Figure 4. Most notable is the Loop Current fluctuation, but the lack of a clear signal in the western gulf is also apparent.

4.2 Statistical Summary of Tracks

Statistics of tracks generated by TFT are shown in Table 1. We calculate track statistics in aggregate, but also on subsets of the tracks. We subset temporally (by season) and spatially (in three "zones" associated with large scale features). These zones are labeled Zone 1, Zone 2, and Zone 3 from west to east, and are shown in the left image of Figure 2. Zone 1 captures an irregularly shaped, mesoscale flow pattern. Zone 2 is an intermediate region, and Zone 3 attempts to capture the Loop Current structure.

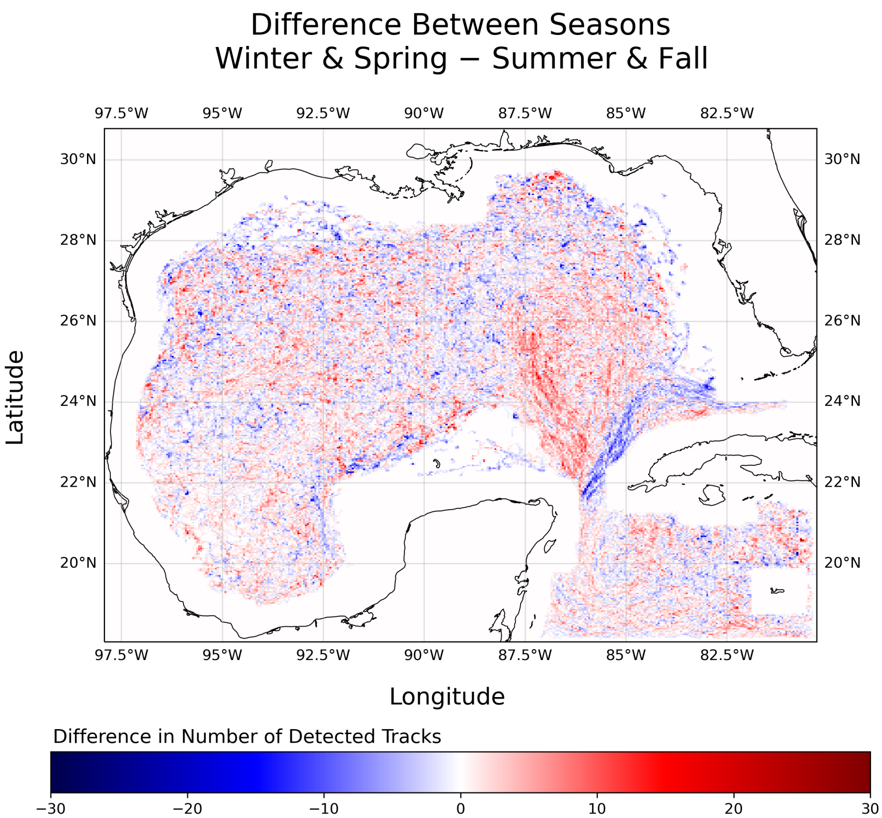

Figure 4: Seasonal difference in masked dataset track density, computed as (winter + spring) − (summer + fall). Tracks shown have been filtered to include those ≥ 25 km.

Broadly, eddies in the Gulf of Mexico tend to move fastest in the spring and summer. However, the seasonal variance is low. Overall, submesoscale eddy velocity is $O(0.5$ m/s), furthering previous results which showed mesoscale and submesoscale horizontal velocities to be similar (Capet et al., 2008). If, as we have documented, the submesoscale eddy motion is largely a function of the jet and mesoscale eddies (Figure 2), then these horizontal velocities are of similar orders of magnitude.

Lifespans tend to be longer in the winter and fall. This is likely due to the known relationship between submesoscale generation and maintenance, and the depth of the mixed layer (McWilliams, 2016). Using this relationship, one can calculate a mixed-layer deformation radius that dictates the maximum size of submesoscale eddies as a function of mixed-layer depth. In the summer, the mixed layer shoals in the presence of strong surface heating, dramatically reducing the mixed-layer deformation radius. With a 1-km horizontal resolution, this NCOM simulation cannot support the generation and maintenance of such small features, leading to a decline in the number of identified submesoscale eddies during this season (D'Addezio et al., 2020). This is evident in the seasonality of the submesoscale eddy sample size shown in Table 1 (last column).

While there is some numerical component to the seasonality we observe herein, decreased eddy activity in summer has also been found in observational measurements of submesoscale turbulent kinetic energy spectra (Callies et al., 2015), and is theoretically expected. In contrast, winter features much deeper mixed layers, and can therefore support the creation of more, relatively larger submesoscale eddies and allow them to propagate longer in the more favorable mixed layer environment. It is expected that these seasonal differences in sample size will be more pronounced with increased temporal output frequency.

Some notable differences exist between tracking results for the masked and unmasked datasets. Compared with the masked fields, distances, lifetimes, and speeds are greater for the unmasked fields. Near the Loop Current (Zone 3) differences in speed between masked and unmasked datasets are relatively attenuated compared to regions away from persistent mesoscale structures (Zones 1 and 2). Lifespan and displacement remain greater for the unmasked dataset in Zone 3, making the similarity in speed between these two datasets somewhat unique.

The unmasked dataset also contains more samples and greater variance in nearly all cases. This is likely due to the limiting nature of traditional eddy identification methods (e.g., D'Addezio et al., 2020). In these traditional methods, identification criteria (e.g., "circularity") may change over the lifetime such that identification criteria are not satisfied throughout the lifespan of the eddy. The statistical effect of this is that a single, long-lived eddy is broken up into multiple, short-lived parts. While the unmasked dataset is less restrictive and contains more samples, the masking procedure can shorten the lifespan and displacement of long-lived eddies, as observed in the statistical summaries shown herein. Further work is required to quantify the influence of this difference in eddy identification.

5 Conclusions

We introduce Topological Feature Tracking to the oceanographic community by applying it to NCOM solutions of the Gulf of Mexico. TFT minimizes preprocessing of data by simplifying noisy scalar fields and tracking critical points between timesteps. Using TFT, we compute eddy statistics of lifetime, displacement, and speed for 1 year of NCOM solutions. Insights on submesoscale eddy propagation speeds of 0.5 m/s, lifetime of 18 hours, and displacement of 30 km are novel results. Seasonal differences are summarized and compared with models and observations.

	Speed (m/s)		Lifespan (h)		Displacement (km)		Sample Size	
	Unmasked Mean (St. Dev.)	Masked Mean (St. Dev.)	Unmasked Mean (St. Dev.)	Masked Mean (St. Dev.)	Unmasked Mean (St. Dev.)	Masked Mean (St. Dev.)	Unmasked	Masked
GoM Aggregate	**0.4436 (0.2343)**	**0.3808 (0.2124)**	**17.8 (28.8)**	**12.3 (26.0)**	**30.9 (60.4)**	**16.2 (31.4)**	**655,727**	**119,775**
GoM Winter (DJF)	0.4184 (0.2333)	0.3760 (0.2171)	19.0 (30.5)	13.5 (27.6)	31.3 (62.0)	17.4 (33.7)	182,522	31,319
GoM Spring (MAM)	0.4726 (0.2367)	0.3949 (0.2167)	16.7 (25.7)	11.1 (20.7)	31.4 (58.5)	15.4 (27.3)	171,134	31,292
GoM Summer (JJA)	0.4703 (0.2354)	0.3928 (0.2156)	15.8 (25.1)	11.6 (27.6)	29.1 (54.9)	15.8 (32.8)	154,453	28,545
GoM Fall (SON)	0.4133 (0.2241)	0.3586 (0.1966)	19.4 (33.3)	13.0 (27.5)	31.6 (66.0)	16.1 (31.7)	147,618	28,619
Zone 1 Aggregate	**0.4316 (0.2154)**	**0.3457 (0.1689)**	**18.4 (30.0)**	**12.8 (26.4)**	**30.9 (58.9)**	**15.2 (28.6)**	**141,626**	**27,081**
Zone 1 Winter	0.3862 (0.2034)	0.3205 (0.1598)	20.6 (33.7)	14.6 (29.3)	31.1 (61.2)	16.1 (30.1)	38,219	6,859
Zone 1 Spring	0.4556 (0.2189)	0.3526 (0.1697)	17.8 (26.8)	11.3 (20.2)	32.0 (58.2)	13.8 (22.9)	39,754	7,848
Zone 1 Summer	0.4622 (0.2187)	0.3596 (0.171)	16.3 (25.7)	11.6 (26.5)	29.5 (53.7)	14.4 (29.1)	35,998	6,737
Zone 1 Fall	0.4200 (0.2106)	0.3501 (0.1727)	19.1 (33.4)	14.2 (29.8)	30.8 (63.0)	17.2 (32.8)	27,655	5,637
Zone 2 Aggregate	**0.4315 (0.2172)**	**0.3725 (0.1887)**	**16.8 (29.0)**	**12.8 (27.1)**	**26.6 (47.8)**	**16 (30.9)**	**24,571**	**5,773**
Zone 2 Winter	0.4137 (0.2101)	0.3528 (0.1745)	18.2 (29.8)	13.8 (26.7)	27.5 (47.4)	16.1 (27.3)	6,601	1,506
Zone 2 Spring	0.4304 (0.2187)	0.3523 (0.1837)	15.6 (25.9)	12.2 (20.9)	25.2 (44.6)	14.4 (22.4)	5,576	1,443
Zone 2 Summer	0.4887 (0.2317)	0.4427 (0.2218)	13.6 (21.6)	11.3 (28.8)	25.2 (46.5)	17.0 (35.2)	5,481	1,245
Zone 2 Fall	0.4040 (0.2019)	0.3542 (0.1631)	18.8 (34.9)	13.6 (30.7)	27.8 (51.6)	16.6 (36.5)	6,913	1,579
Zone 3 Aggregate	**0.5167 (0.2513)**	**0.4917 (0.2566)**	**14.8 (24.3)**	**12.0 (22.3)**	**29.3 (52.9)**	**21.0 (38.0)**	**93,578**	**19,608**
Zone 3 Winter	0.5196 (0.2581)	0.5177 (0.2642)	16.4 (25.6)	12.5 (23.6)	33.1 (59.6)	23.1 (42.2)	29,903	5,849
Zone 3 Spring	0.5621 (0.2464)	0.5459 (0.2562)	13.8 (21.2)	10.9 (17.2)	30.5 (54.0)	21.9 (37.6)	23,629	4,813
Zone 3 Summer	0.5275 (0.2458)	0.4833 (0.2536)	13.5 (21.5)	11.3 (22.2)	27.1 (47.8)	19.3 (35.5)	22,881	4,773
Zone 3 Fall	0.4348 (0.2333)	0.4023 (0.2238)	15.3 (28.8)	13.2 (25.4)	23.7 (43.8)	18.8 (34.5)	17,165	4,173

Table 1: Statistics (calculated seasonally and in aggregate) of submesoscale eddy tracks across the Gulf of Mexico (three zones as depicted in Figure 2.)

Further investigation should focus on the differences in eddy identification methods, and how TFT can be improved based on these efforts. Also, modifications to the Lifted Wasserstein distance function (to penalize incorrect matchings in a nonlinear manner) should improve the method broadly. Additionally, an automated method of suggesting or selecting weight parameters and the persistence threshold may be explored.

Open Research

Data Availability Statement: Ocean surface velocity data, used to identify and track features in this study, were obtained via the Navy Coastal Ocean Model (NCOM). The solution data used herein was generated using the same NCOM modeling framework (i.e., domain, boundary and initial conditions, numerical and physical parameterizations, etc.) as described in D'Addezio et al. (2020) (`https://doi.org/10.1175/JPO-D-19-0100.1`).

Acknowledgments

The authors thank the anonymous reviewers for their helpful comments. Research by the first six authors was partially supported by the DARPA Ocean of Things project, under contract N6600121C4006. Joseph M. DAddezio was funded by the Office of Naval Research (ONR) under grant N0001421WX02086. Gregg Jacobs and Tamay Özgökmen were funded by the DARPA Ocean of Things project under contract HR0011150953. This document has been cleared for public release under Distribution Statement "A" (Approved for Public Release, Distribution Unlimited). The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

References

Barron, C. N., Kara, A. B., Martin, P. J., Rhodes, R. C., & Smedstad, L. F. (2006). Formulation, implementation and examination of vertical coordinate choices in the global navy coastal ocean model (ncom). *Ocean Modelling*, *11*(3-4), 347–375.

Callies, J., Ferrari, R., Klymak, J. M., & Gula, J. (2015). Seasonality in submesoscale turbulence. *Nature communications*, *6*(1), 1–8.

Capet, X., McWilliams, J. C., Molemaker, M. J., & Shchepetkin, A. F. (2008). Mesoscale to submesoscale transition in the california current system part i:

Flow structure, eddy flux, and observational tests. *J. Phys. Oceanogr.*, *38*, 29-43.

Chelton, D. B., Schlax, M. G., Samelson, R. M., & de Szoeke, R. A. (2007). Global observations of large oceanic eddies. *Geophysical Research Letters*, *34*(15).

Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007, January). Stability of persistence diagrams. *Discrete Comput. Geom.*, *37*(1), 103–120. Retrieved from `http://dx.doi.org/10.1007/s00454-006-1276-5` doi: 10.1007/s00454-006-1276-5

D'Addezio, J. M., Jacobs, G. A., Yaremchuk, M., & Souopgui, I. (2020). Submesoscale eddy vertical covariances and dynamical constraints from high-resolution numerical simulations. *Journal of Physical Oceanography*, *50*(4), 1087–1115.

D'Addezio, J. M., Smith, S., Jacobs, G. A., Helber, R. W., Rowley, C., Souopgui, I., & Carrier, M. J. (2019). Quantifying wavelengths constrained by simulated swot observations in a submesoscale resolving ocean analysis/forecasting system. *Ocean Modelling*, *135*, 40–55.

Edelsbrunner, H., & Harer, J. (2010). *Computational topology: an introduction.* AMS.

Edelsbrunner, H., Morozov, D., & Pascucci, V. (2006). Persistence-sensitive simplification functions on 2-manifolds. In *Proceedings of the twenty-second annual symposium on computational geometry* (pp. 127–134).

Faghmous, J. H., Frenger, I., Yao, Y., Warmka, R., Lindell, A., & Kumar, V. (2015). A daily global mesoscale ocean eddy dataset from satellite altimetry. *Scientific data*, *2*(1), 1–16.

Gula, J., Molemaker, M. J., & McWilliams, J. C. (2014). Submesoscale cold filaments in the gulf stream. *Journal of Physical Oceanography*, *44*(10), 2617–2643.

Isern-Fontanet, J., García-Ladona, E., & Font, J. (2003). Identification of marine eddies from altimetric maps. *Journal of Atmospheric and Oceanic Technology*, *20*(5), 772–778.

Kurian, J., Colas, F., Capet, X., McWilliams, J. C., & Chelton, D. B. (2011). Eddy properties in the california current system. *Journal of Geophysical Research: Oceans*, *116*(C8).

Laudenbach, F. (2013). A proof of morse's theorem about the cancellation of critical points. *Comptes Rendus de l'Academie des Sciences*, 483–488.

Lévy, M., Franks, P. J. S., & Smith, K. S. (2018). The role of submesoscale currents in structuring marine ecosystems. *Nature communications*, *9*(1), 1–16.

Liu, G., Bracco, A., & Sitar, A. (2021). Submesoscale mixing across the mixed layer in the gulf of mexico. *Frontiers in Marine Science*, *8*, 615066.

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

Milnor, J. W. (1963). *Morse theory.* Princeton University Press.

Poje, A. C., Özgökmen, T. M., Lipphardt, B. L., Haus, B. K., Ryan, E. H., Haza, A. C., ... others (2014). Submesoscale dispersion in the vicinity of the deepwater horizon spill. *Proceedings of the National Academy of Sciences*, *111*(35), 12693–12698.

Soler, M., Plainchault, M., Conche, B., & Tierny, J. (2018). Lifted wasserstein matcher for fast and robust topology tracking. *2018 IEEE 8th Symposium on Large Data Analysis and Visualization (LDAV)*, 23–33. doi: 10.1109/ldav.2018.8739196

Zhong, Y., & Bracco, A. (2013). Submesoscale impacts on horizontal and vertical transport in the gulf of mexico. *Journal of Geophysical Research: Oceans*, *118*(10), 5651–5668.