A parameter-free spatio-temporal pattern mining model to catalog global ocean dynamics

James H. Faghmous, Matthew Le, Muhammed Uluyol, and Vipin Kumar
Department of Computer Science and Engineering
The University of Minnesota
Minneapolis, USA
jfagh@cs.umn.edu

Snigdhansu Chatterjee
School of Statistics
The University of Minnesota
Minneapolis, USA
chatterjee@stat.umn.edu

***Abstract*—As spatio-temporal data have become ubiquitous, an increasing challenge facing computer scientists is that of identifying discrete patterns in continuous spatio-temporal fields. In this paper, we introduce a parameter-free pattern mining application that is able to identify dynamic anomalies in ocean data, known as ocean eddies. Despite ocean eddy monitoring being an active field of research, we provide one of the first quantitative analyses of the performance of the most used monitoring algorithms. We present an incomplete information validation technique, that uses the performance of two methods to construct an imperfect ground truth to test the significance of patterns discovered as well as the relative performance of pattern mining algorithms. These methods, in addition to the validation schemes discussed provide researchers direction in analyzing large unlabeled climate datasets.**

I. Introduction

The World Ocean covers more than 70% of the globes's surface and is the site of intense physical, chemical, and biological activity that impact virtually every other aspect of our planet. A plethora of phenomena occur globally and there are significant scientific questions to be answered by effectively monitoring such phenomena. Given that most climate phenomena are dynamic, a typical workflow is to identify such phenomena, track their evolution, and report global statistics. The focus of this paper is on enabling the aforementioned workflow for monitoring mesoscale ocean eddies in large climate data.

Mesoscale ocean eddies (hereafter eddies) are coherent rotating structures of ocean spanning tens to hundreds of kilometers and lasting a few days to several months (see Figure 1). Eddies are critical phenomena as they dominate the ocean's kinetic energy and are responsible for the transport of heat, salt, nutrients, and energy across the ocean [12]. Eddies also have had significant impacts on marine and terrestrial ecosystems. For instance, one study found that 7000-year-old coral reefs were asphyxiated due to massive phytoplankton blooms, which were enhanced by a large eddy [24]. Similarly, some of the recent devastating landfalling hurricanes, including Hurricane Katrina, gained intensity in the Gulf of Mexico when passing over a warm-core eddy [18]. As a result, hurricane intensity forecasts now account for eddy activity when making projections [25]. Subsequently, in order to understand how these mesoscale features impact other phenomena it is imperative that we understand their properties on a global scale.

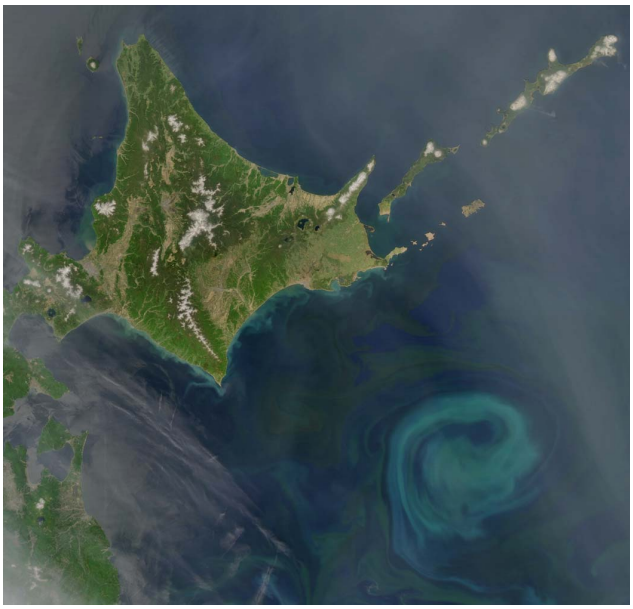

Fig. 1. A mesoscale ocean eddy off the coast of Japan (near bottom right corner). These large whirlpools are a source of intense physical and biological activity. We are able to see the eddy, which is submerged *under* the surface because of the enhanced phytoplankton activity (reflected in the bright blue color). Image courtesy of the NASA Earth Observatory. Best seen in color.

A. Ocean eddies: A brief overview

Ocean eddies are three dimensional features that extend up to tens of meters deep under the ocean's surface (think of a submerged cyclone). Therefore, eddies would be easy to identify given global three dimensional measurements of key ocean variables such as salinity and turbulence. Unfortunately, most of the global data available do no provide subsurface information and thus we must resort to surface data to monitor eddies globally. The ocean's surface is influenced by a variety of oceanographic and atmospheric phenomena and since eddies cannot be observed directly we must rely on the impact eddies have on the sea surface as a proxy for inferring the presence of an eddy.

Traditionally, the automatic detection and tracking of ocean eddies were achieved using sea surface temperature or ocean color satellite data [23, 10, 6]. Although eddies do impact

1550-4786/13 $31.00 © 2013 IEEE
DOI 10.1109/ICDM.2013.162

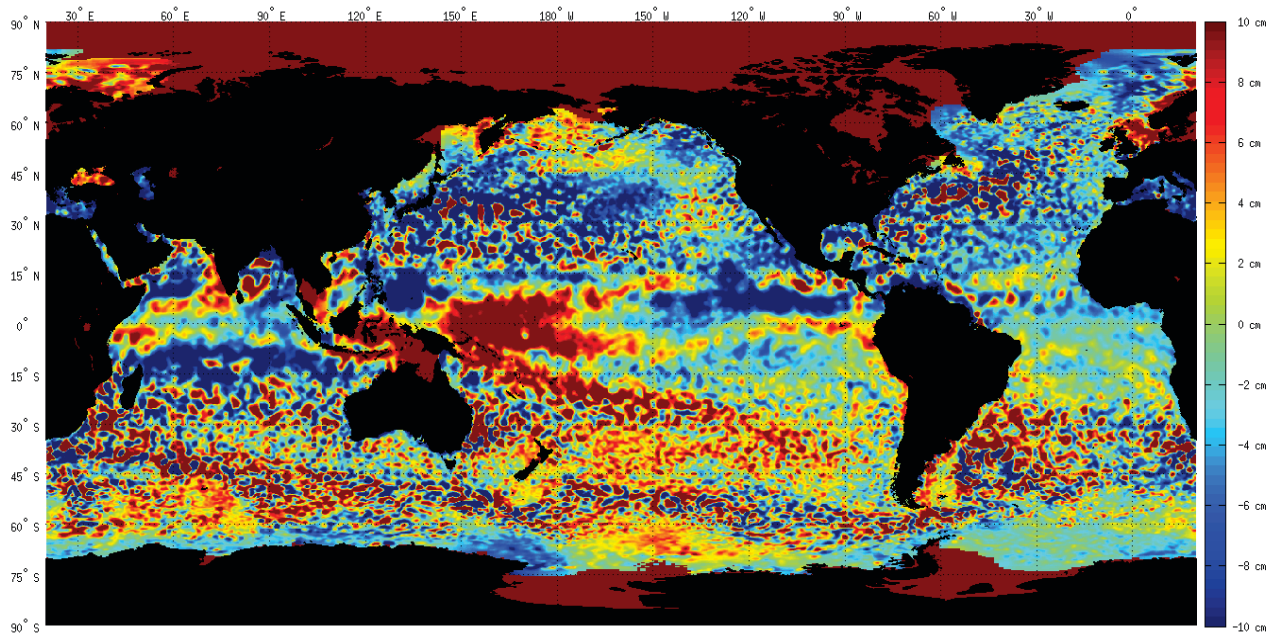

Fig. 2. Global unfiltered SSH anomalies for one week in 2005. Large-scale variability makes global pattern mining challenging.

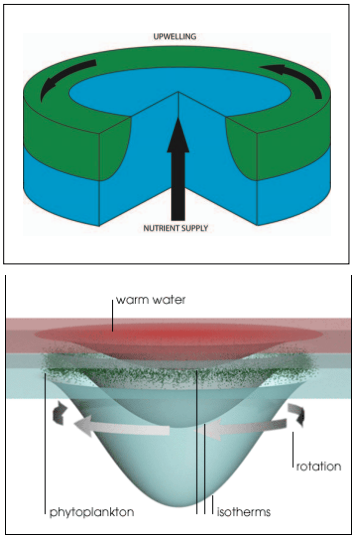

Fig. 3. **Top:** A schematic cross section of an anti-cyclonic eddy (in the Northern Hemisphere) density surfaces are depressed within the eddy causing an increase in SSH. The elevation of subsurface density surfaces replenishes the upper part of the ocean with nutrients needed for primary production. **Bottom**: A cyclonic eddy causes an decrease in SSH. Bottom image by Robert Simmons of NASA. Best seen in color.

surface temperature and near-surface color, other phenomena do so as well. This made identifying eddies in such datasets prone to false positives. The advent of sea surface height (SSH) observations from satellite radar altimeters became a better alternative to these datasets. This is because eddy dynamics are intimately linked to SSH. Eddies are classified by their rotational direction. Cyclonic eddies rotate counter-clockwise (in the Northern Hemisphere), while anti-cyclonic eddies rotate clockwise. Cyclonic eddies, like the one in Figure 3 (bottom panel), cause a decrease in SSH and elevations in subsurface density surfaces. Anti-cyclonic eddies, such as the one depicted in Figure 3 (top panel), cause an increase in SSH and depressions in subsurface density surfaces. These characteristics allow us to identify ocean eddies in SSH satellite data as close contours of positive (anti-cyclonic) and negative (cyclonic) anomalies.

Eddies are commonly identified in the SSH field by assigning binary values to the SSH data based on whether or not a varying threshold was exceeded, and subsequently saving the eddy-like connected component features that remain after thresholding. The identified features are further pruned based on expert-defined criteria that characterize eddies [5, 7]. Such threshold-based algorithms have two distinguishing characteristics: first, they are designed by experts who have extensive, yet incomplete, knowledge of the application at hand. Second, such expert knowledge is generally encapsulated in many necessary, yet arbitrary, parameters. While parameter-laden algorithms are sometimes needed to control for the chaos and noise in the system, they quickly become a double-edge sword: at what point do we stop finding novel features due to constraining parameters that are based on incomplete current knowledge?

In addition to potentially jeopardizing knowledge discovery, expert parameterization has other notable drawbacks as noted by [19]: first, it makes it hard to compare across studies since it is difficult to control for the effect of different parameterizations. Second, if the parameters were estimated using the full dataset, the method becomes subject to overfitting and might not generalize to unseen data. Finally, strong paramterization may lead to overestimating the significance of

spurious patterns.

The above observations are especially true when it comes to climate data as they tend to also be highly variable. Sources of variability include: (i) natural variability, where wide-range fluctuations within a single field exist between different locations on the globe, as well as at the same location across time (see Figure 2); (ii) variability from measurement errors; (iii) variability from model parameterization, initial conditions, and post-processing; and (iv) variability from our limited understanding of how the world functions (*i.e.* model representation). Even if one accounts for such variability, it is not clear if these biases are additive and there are limited approaches to de-convolute such biases a posteriori. Additionally, ocean eddies and their related properties (size, propagation speed, *etc.*) vary by latitude [12].

To address these issues, we introduce a parameter-free ocean eddy monitoring application and novel evaluation methods that assess the quality of unsupervised learning algorithms using the spatio-temporal consistency of features as a measure of accuracy. The majority of eddy monitoring applications take the information-rich four-dimensional ocean data and reduce it to 2 or 3 dimensions and introduce unnecessary uncertainty in the process. We propose that by monitoring the spatio-temporal consistency of features we are better able to identify features compared to using space or time information alone. Furthermore, we introduce several experiments to evaluate the performance of our unsupervised learning method without any "ground truth" data readily available.

The methods introduced in this work illustrate new methods for identifying closed contour features in a continuous spatio-temporal field. Other examples of computer science research include the works of Mesrobian et al. [21] and Stolorz et al. [27] who tracked cyclones as local minima within a closed contour sea level pressure (SLP) field, as well as Bain et al. [1] and Henke et al. [13] who identified and tracked the InterTropical Convergence Zone (ITCZ), a prominent climate phenomena over the east Pacific.

II. Previous work

Traditionally, the automatic detection and tracking of ocean eddies were achieved using sea surface temperature or ocean color satellite data [23, 11, 10, 6]. The advent of SSH observations from satellite radar altimeters provided researchers with an unprecedented opportunity to study eddy dynamics on a global scale. The earliest automated eddy detection methods in SSH data relied on a measure of rotation and deformation in fluid flow known as the Okubo-Weiss (W) parameter [15]. In such studies, eddies were defined as features where the W-parameter was below an expert-specified negative threshold. The majority of these studies applied region-specific parameters to study eddy activity in the Mediterranean Sea [15, 16, 17] as well as major currents [22]. Another regional study identifying eddies as closed streamlines with a total 360° angle between adjacent streamlines [3]. Chelton et al. [4] performed the first W-parameter-based global eddy monitoring study.

Threshold-based methods have since gained popularity with works from Fang and Morrow [9] and Chaigneau and Pizarro [2] analyzing regional eddy activity with a single threshold value of $\pm 10cm$ and $\pm 6cm$ respectively. Chelton et al. [5] used an iterative thresholding method to monitor eddies globally. Faghmous et al. [7] extended the iterative thresholding method by enforcing a minimal convexity ratio on the features to ensure the most compact features were preserved.

While there has been a wealth of research on the subject, identifying eddies still remains a challenge. As noted by [5], the W-parameter algorithms are highly susceptible to the noise in the SSH field. Furthermore, most of the studies above use arbitrary parameter values to separate eddies from noise. In addition to the biases introduced by parameterization the iterative thresholding methods are unable to separate features that are in close physical proximity as Chelton et al. [5] points out: "The algorithm described above can yield eddies with more than one local extremum of SSH (*i.e.* merges). This could occur because of multiple eddies in close proximity that are contained within a single outermost closed contour of SSH, or because of irregularity of the SSH structure within a single eddy from noise in the SSH fields. We attempted to separate, or split, multiple eddies ... after much experimentation, the eddy splitting procedure was abandoned." As we will show in the results, merged eddies have a notable impact on reported ocean dynamics.

III. Methods

The most widely used eddy finding algorithms employ a top-down thresholding approach (TD) [5]. At a high level, the algorithm extracts candidate connected components from SSH data by iteratively thresholding the data and assigning binary values to the SSH field based on whether or not a varying threshold was exceeded, and subsequently identifying mesoscale connected component features. We refer to this approach as top-down because the algorithm attempts to find features at their largest possible close contour. This is achieved by repeatedly thresholding the data at regular $1cm$ intervals from $-100cm$ to $+100cm$. At each threshold tr_i, all connected components that have an SSH anomaly of at least tr_i are identified. Each connected component is then analyzed based on 5 expert-specific criteria to determine whether a connected component may be deemed an eddy-like feature. These are: a pixel count ranging strictly between 9 and 1000 pixels, at least $1cm$ amplitude, each feature must have at least one extrema, every pixel within the feature's contour must be within a predefined maximum distance from any other pixel within the eddy, and every feature must meet a strict latitude-dependent convexity criterion. If the feature meets these criteria, the algorithm then removes from consideration all pixels belonging to the identified "eddy-like" connected component and tr_i is incremented. For identification of anticyclonic eddies, tr_i is initialized at $-100cm$ and increased in $1cm$ steps to $+100cm$. Conversely, detection of cyclonic eddies is accomplished by decreasing tr_i from $+100cm$ to $-100cm$.

One of the main reasons TD suffers from the limitations reported in section II is because it consistently over-estimates the features' contour, which in addition to the noise in the SSH field and the aggressive thresholding steps results in noise being be part of a feature, and in some cases if the noise is less then the thresholding step, the noise between features causes them to be merged as a single large feature.

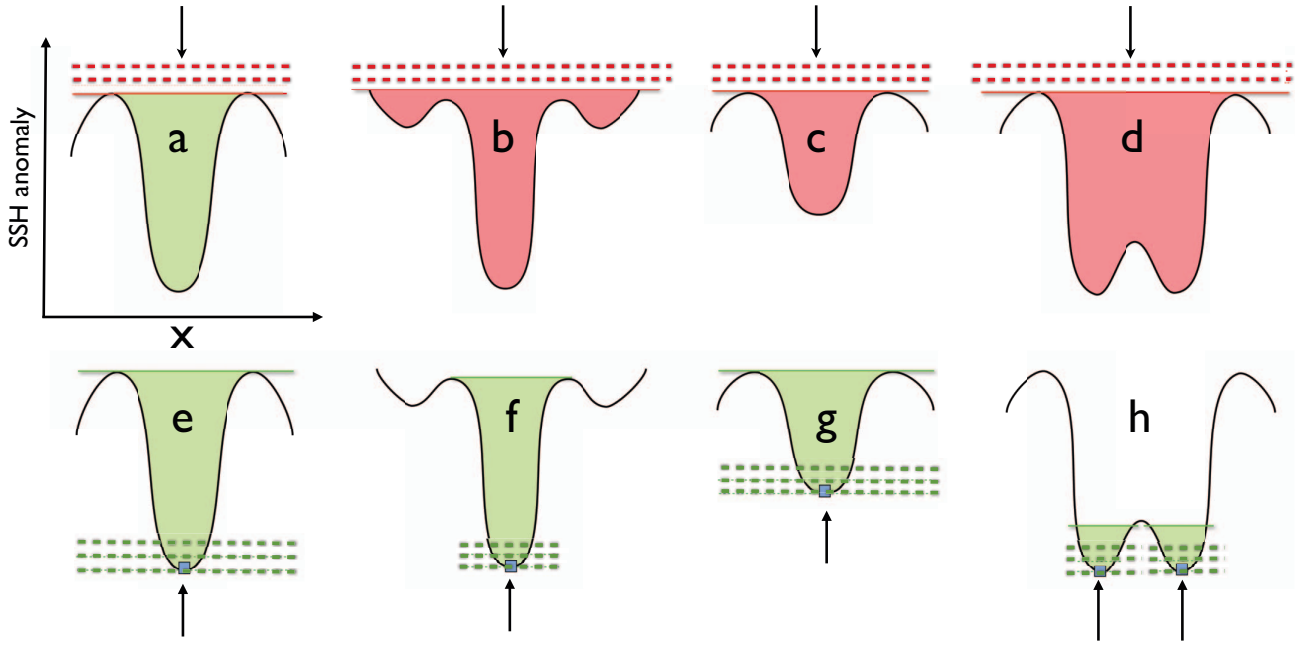

Fig. 4. A two-dimensional cross section of SSH anomalies. The arrows and dashed lines represent the direction of the iterative thresholding method. The color of the feature (red or green) represents whether each method was able to accurately recover each features boundaries. TD starts from a very high threshold and gradually decreases and stops as soon as it finds a close contour that meets its expert-criteria. Alternatively, BU (bottom row) thresholds locally starting from each local minima and gradually grows to reconstruct the features body and stops once a feature contains two extrema. Unlike TD, BU is able to avoid adding noise to the features contour (panel f), does not discard features due to arbitrary parameters (panel g), and is able to separate features in close proximity (panel h) when TD effectively merges them (panel d)

A. *Bottom-up (BU) thresholding*

A more intuitive approach to monitoring global ocean eddy activity starts from the simple notion that every eddy should have a single extrema. Therefore, we can construct a superset of all extrema in the SSH field for any given SSH satellite snapshot. While not every extrema may represent an eddy's core, starting from the extrema allows us to search for anomalies locally as opposed to globally as in TD's case.

We define extrema as pixels that are strictly greater/less than their 5×5 neighbors and seed our algorithm with the extrema of the SSH field. For each extrema, we set a threshold to its current value and we incrementally increase (decrease) the threshold to construct a concave down (up) feature from that extrema. Since every feature can have only a single extrema, instead of having numerous stoping conditions, we stop thresholding once the connected component contains more than a single extrema. At this point, it is very likely that we have overestimated the feature's contour. When this occurs, we simply set the feature's contour as that of the step prior to merging the two extrema. This intuitive observation – that a feature cannot contain two local extrema – allows us to abandon all expert parameters proposed in [5, 7], except for the 9 pixel minimal feature size, because as Chelton et al. [5] points out, 8 pixels is the minimal feature size that can be resolved given the post-processing applied to the satellite product used in this study. This method is fundamentally different from the traditional TD approach in that it starts with the most certain part of the feature – its extrema – and then builds the body up. Furthermore, if we were to constrain BU using similar expert-conditions it would cause BU to severely underestimate a feature's contour and, in many cases, discard the feature. Once the feature's contour is identified, we track the features in space and time by attaching a feature in one time-frame to the nearest feature in following time-step within a predefined search space. Although this tracking method has limitations [8], novel eddy tracking methods are beyond the scope of this paper. The full source code, an interactive eddy and track viewer, as well as all results from this study are available as an open-source project for download from: www.ucc.umn.edu/eddies

Figure 4 illustrates how each method performs under a variety of scenarios. The top row (panels a-d), shows the resulting features from a TD analysis, while the bottom row (panels e-h) shows the features recovered by the BU approach. Each feature is represented as a two-dimensional cross section in the SSH anomaly field. TD methods select the largest close contour possible that meet all expert-criteria. This works well if there is no noise along the feature's contour such as in panel a. In practice, however, TD is susceptible to: (1) including noise in the feature's perimeter (panel b); (2) missing features that slightly fail to meet any of the expert-criteria (panel c); and (3) merge features that are in close proximity (panel d). The BU approach, however, is able to recover all features, including those that were discarded through parameterization since it is parameter-free (panel g).

Using this algorithm, we monitored global eddy activity using the Version 3 dataset of the Archiving, Validation, and Interpretation of Satellite Oceanographic (AVISO) which

contains 7-day averages of SSH on a 0.25° grid from October 1992 through January 2011 [1].

IV. Evaluation

The evaluation of pattern mining algorithms is extremely challenging in unlabeled climate data. Given the large data size and common parameterizations, the validity and significance of identified features must be questioned. Once mesoscale features have been identified, the quality of both the features and propagation paths must be evaluated. This can be done in three ways: by analyzing field studies and in-situ data, applying methods to simulations and idealized models, or by analyzing the types of biases inherent in the data and quantify each method's robustness to such biases.

Large unlabeled datasets are not foreign to computer science. A similar problem faced other data mining applications such as optical character recognition and autonomous image labeling. The Internet's hyper-scale, along with creative crowdsourcing initiatives (also known as Human Computation) such reCAPTCHA [29], PHETCH (later renamed Google Image Labeler) [28], and Amazon's Mechanical Turk [26] allowed us to make significant gains in labeling large complex datasets that cannot be autonomously labeled. A crucial distinction between labeling pictures and physical features is scale. While millions of people can easily distinguish between the picture of a cat and a dog, only a minute fraction could identify mesoscale features in SSH data. In fact, until recently, even experts misidentified non-linear eddies for linear Rossby waves in satellite data [20]! In most studies, results of automated eddy identification methods are tested anecdotally. For example, Chaigneau et al. [3] randomly selected 10 altimeter snapshots off the coast of Peru (out of 700+ possible snapshots) and asked 5 expert oceanographers to manually draw the contour of every eddy they could identify in the sample snapshots. The features identified by the authors' algorithm were then compared to "the expert eddies" and false positive and negative rates were computed based on how many features were introduced or missed by the algorithms compared to the experts.

Ideally, one would use "ground truth" data where the eddy tracks are known in advance and test how well each method recovers such tracks under varying conditions. One way to generate such data would be through a numerical simulation (*i.e.* ocean model) with idealized eddies as ground truth and then gradually add noise. However, such simulations are computationally expensive and require sophisticated physics-based models to simulate eddies and their trajectories. Furthermore, developing unbiased methods to introduce noise into the data is challenging. An alternative would be to use field studies data, where floats are dropped in the ocean and subsequently tracked. Eddies are identified when the float rotates while translating, which occurs when the float is trapped within the eddy interior and moving along its translating path. However, such data make up a small sample size (in both space and time) and are not sufficient to significantly differentiate between the two methods.

Given that no ground truth is available, we instead leverage the fact that both TD and BU methods perform reasonably well and we can leverage the performance of both algorithms to identify a set of features that are more certain than any feature discovered by either methods alone. To do so, we construct three datasets: first, features that are identified by both TD and BU. These are the features in TB and BU that overlap by at least a single pixel. Second, we identify the features that are identified by TD but not BU. Finally, we find the features that are identified by BU but not TD. This allows us to frame our problem from an unsupervised learning problem into a classification task, where the training set may have mislabeled observations (*i.e.* imperfect ground truth). Assume that the j^{th} algorithm has probabilities of false positives $\alpha_j, j = 1, 2$ and those of false negatives $\beta_j, j = 1, 2$. Suppose there are m observations in the dataset, for which the labels are obtained by running the first algorithm on a given dataset. If m_1 of these m observations are classified as positives, we may expect that $m_1\alpha_1$ of these are mislabeled, while $(m - m_1)\beta_1$ of those that are labeled negative and misclassified. Using the second algorithm on the same dataset partitions it into eight sets, whose probabilities are obtained by using standard multinomial probability-based algebra.

For each set of features, we record characteristics that are important to understanding ocean dynamics, these are: (1) eddy amplitude, which is the difference in SSH between the eddy's extrema and its mean perimeter [5]. Eddy amplitude informs of the strength of the feature's rotation; (2) Rotational speed: the rotation of the feature is the single most distinguishing characteristic between an eddy and other ocean phenomena. The rotational speed is approximated by calculating the mean gradient in both x and y directions for each pixel in the feature; (3) Pixel count: we measure the area of each feature by counting the number of pixels within its contour, these can be further transformed to m^2 by measuring pixel area based on the latitude of the pixel; (4) Feature lifetime: once a feature is identified we are interested in tracking its trajectory over time. Due to the noise in the SSH field, Chelton et al. [5] considered significant only features that persist 4 weeks or more. We follow the same convention in our analysis; (5) Position in lifetime: In addition the total length of a feature's lifetime, we are interested in knowing which stage it is in. We can quantify a feature's relative lifetime by dividing its current position from its total lifetime (*e.g.* if a feature survives for 10 weeks and it is in week 5, it has reached 50% of its lifetime).

V. Results

Although we analyzed global eddy activity between 1992-2011, for simplicity, we will focus our results on a single year of data 2005. A full year gives us the full range of seasonal variability while remaining manageable for analysis and discussion.

A. Difference in features and tracks between TD and BU

For 2005, both methods found 93,603 overlapping features. BU identified 36,120 unique features that TD did not find. Finally, TD found 3,572 unique features that BU did not identify. We then performed a multivariate distribution analysis on the quantities measured (rotational speed, amplitude, *etc.*) to determine whether the TD-only and BU-only features were more similar to the more certain overlapping features.

[1] Available at http://www.aviso.oceanobs.com/es/data/products/sea-surface-height-products/

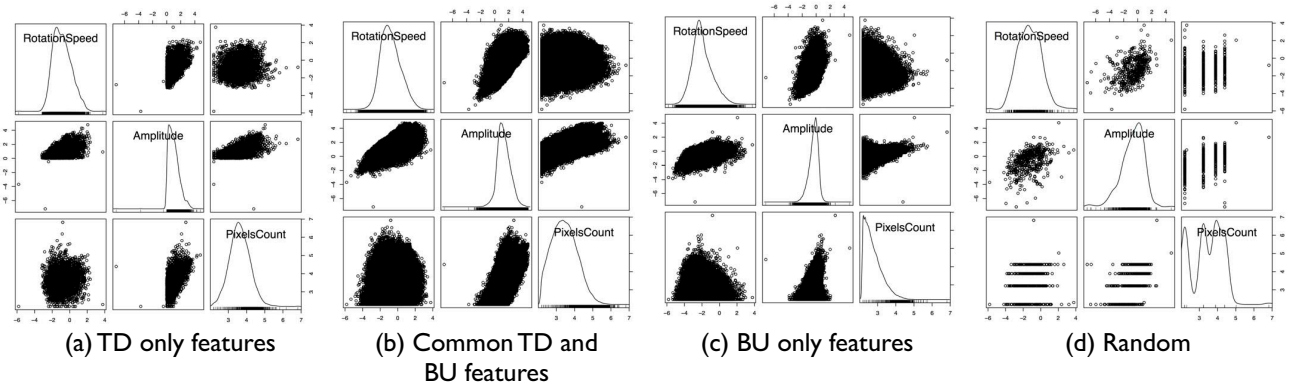

Fig. 5. Scatter plots and density estimates for rotational speed (top row), amplitude (middle row), and pixel count (bottom row) in TD only features (panel a), shared features between TD and BU (panel b), BU only features (panel c), and random SSH regions (panel d)

In addition to considering data from the cases where features were detected by both TD and BU methods, or by one but not both, we also selected a random sample of 1000 points where neither method detected an eddy. For each randomly selected non-eddy pixel, we further randomly select a $k \times k$ neighborhood of pixels to form the body of this random feature. k ranged between 3 and 9 pixels for a random feature size of 9 to 81 pixels. In this random case, feature lifetime and current position are not defined since these are not real features. Our first analysis consisted of determining whether both TD and BU did reasonably well versus random noise to support our assumption that we could use the features they both identify as imperfect ground truth.

The data for rotational speed, amplitude, and pixel count are all positive-valued for any identified eddy. However, the amplitude for random formations may be negative. In our random selection of 1000 non-eddy pixels, 527 had negative amplitudes. We removed these cases from the comparison, so that the feature space for the random formations and potential eddies are the same. A preliminary analysis then suggested that rotational speed and amplitudes may be modeled as log-normal variables, hence a logarithmic transformation was implemented on these features to achieve approximate Gaussianity.

Figure 5 shows the scatter plots and density estimates for these variables in the three datasets we used. The diagonal shows the probability density estimates for rotational speed (top row), amplitude (middle row), and feature pixel count (bottom row). The remaining panels show the scatter plots of the log-transformed dimensions between each row and column. For instance the scatter plot in the second box in the top row of panel (a) shows the TD-only features' rotational speeds and associated amplitudes.

By observing the TD only features, one may notice that that estimated amplitude density is truncated at 1, due to the 1cm constraint imposed by TD. Furthermore, many of its scatter plots are unable to recover the full range of the sample population captured by the TD and BU common features. For example, the TD-only density severely underestimated the number of features with low pixel count and high amplitude (the rightmost middle scatter plot in panel (a)). Small feature size along with high amplitude are signs of high energy compact features, which are important to ocean dynamics. Such features would be missed by TD. Such a lack of representation is a clear example of how highly parametrized methods fail to generalize to unseen data.

Hotelling's T-squared [14] is a method of comparing multivariate observations. We use it to verify whether TD is closer to the random data compared to the BU method. The T-squared distances between random data and TD/BU are 32571.48 and 39436.91 respectively, although both of these methods are significantly distant from the random data. Note that these numbers reflect the units in which the features are measured, and should be used only for comparison (they carry no meaning on an absolute scale). This result gives us confidence that both methods significantly outperform random chance and we can use the features detected by both as an imperfect ground truth.

We used several linear discriminant discriminant analyses on the logarithmically-transformed data to estimate the *sensitivity* and *specificity* of the TD and BU methods. In order to study sensitivity, we consider the percentage of cases where these methods detect, or fail to detect, a true eddy from a random formation in the ocean. When the transformations were used on rotational speed, amplitude, and area in pixels, both methods classified every eddy and random formation correctly, even when the training sample size was as low as 10% of the data size. Without the transformations, each methods misclassified about 5% of the cases, which is not unexpected since the standard assumptions of linear discriminant analysis fail to hold.

While both methods may be good at discriminating random noise and eddy features, the next question of interest is the relative performance of these methods. That is, are both methods equally able to detecting eddies, or does one of them tend to miss a few more than the other? This is a question of specificity. Note that for a potential eddy identified by either method, we have additional information in feature lifetime and the logit-transformed relative lifetime position. Using these variables, TD has a misclassification error of 12%, while BU has a misclassification error of 3.5%. This figure is obtained by accounting for the fact that the BU method misclassifies about 4% of the potential eddies identified by TD, which in turn correctly labels eddies about 88% of the time. Some amount of algebra reveals that the true misclassification rate for BU is

thus around 3.5%.

Note that in the above analysis, we assumed that the data has standard independence and exchangeability properties that are required for the statistical analysis. We ensured that random sampling assumptions are valid for our study by not only comparing *detected* eddies with a sample of random features, but also by using randomized training samples for our discriminant analyses. In the absence of ground truth, we had to use the three sets of data on eddies detected by both methods, or by only one of the two methods. Our results indicate that eddies and random formations are very well separated in a log-transformed space.

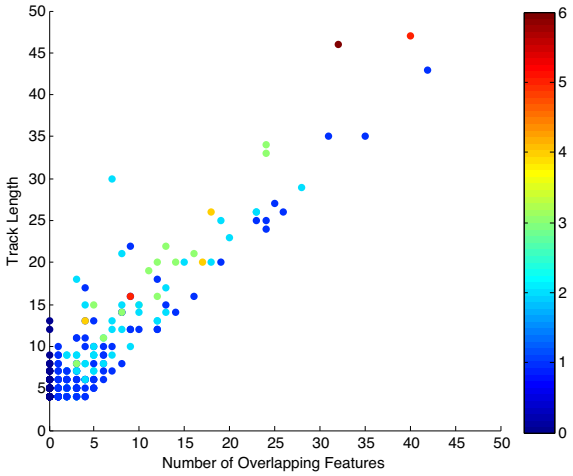

Fig. 6. Number of overlapping features versus track lengths, color-coded based on the number shorter distinct TD tracks associated with a single BU track.

B. Impact of parameters on reported track lengths

The majority of parameters used in previous studies can be strict and arbitrary. To investigate the gains of introducing a parameter-free method, we identified all tracks that have a feature that violates one of the criteria used in the pervious methods. Due to space limitations we will only discuss the impact of removing the 1cm amplitude criterion. In this analysis, we find all tracks that have at least one BU feature with an amplitude less than 1cm. We then look at every BU feature in the track and see if it overlaps with a TD feature. Figure 6 shows a scatter plot of overlapping features versus track lengths, where each point in the plot is colored based on how many distinct tracks the TD approach broke the track up into due to its harsh 1cm amplitude restriction. That is, for every track that contains at least a single feature with an amplitude less than 1cm, how many TD tracks are associated with it? This plot reveals the different effects a harsh criterion can have on resulting tracks. Take the bright red point in the upper right corner for example, it represents a single 45 week BU track which has 40 corresponding TD features, yet TD identified five distinct tracks along the single 45-week BU track. These disjoint tracks are due to the five feature discrepancy $(45 - 40)$ between TD and BU. At least one of the missed TD features were due to the amplitude criterion.

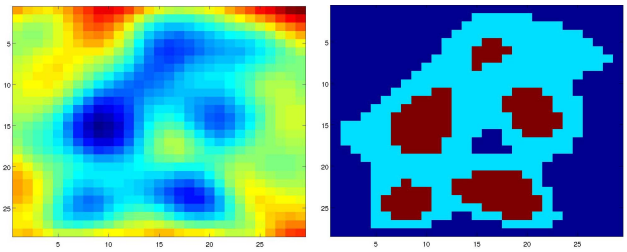

Fig. 7. An image of five distinct eddies that were merged together by the TD approach, but broken up by the BU approach. Left panel shows the SSH anomaly. It can be easily seen that there are at least five well-defined features in the frame. Right panel: the features as identified by TD (light blue) and BU (maroon). Although this is an extreme case, artificial merges have significant implications on a variety of eddy dynamics.

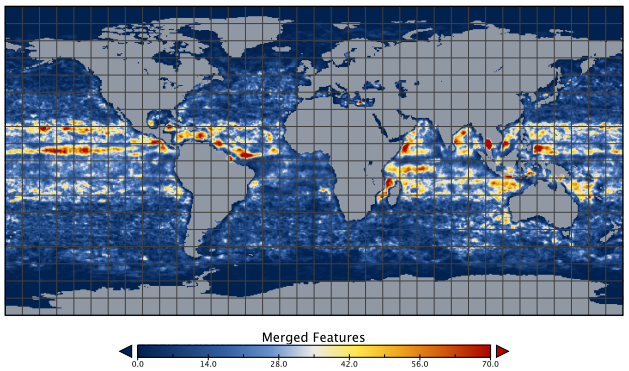

Fig. 8. A density map showing the regions with high concentration of features with more than a single extrema (merges).

It is important to also note the points that are along the diagonal. These are tracks that have the same exact number of TD and BU features (*i.e.* matches), yet we know that at least a single BU feature in every track violates the 1cm amplitude condition. A closer examination of these tracks reveals that even though BU identified a feature with a less than 1cm amplitude, TD either merged or had a larger contour which resulted in the same feature passing the 1cm amplitude criterion.

Imposing a 1cm amplitude is an example of how one can overestimate the significance of patterns identified using a parameter-laden approach. This is because if one was to report a high number of short lived eddies in a certain region, these "patterns" would be incorrect because these tracks would not be short-lived, rather tracks would be prematurely terminated because a single feature along its path temporarily fell below the 1cm amplitude threshold.

C. Impact of merged features

As previously mentioned, merging eddies is a significant issue with threshold-based methods. Figure 7 shows the SSH anomalies of five cyclonic features. As it can be seen, the TD method merged the five features into a single large feature (light blue feature, right panel) while BU (maroon feature) did not. To identify merges within TD features, we analyzed all TD features that contained more than a single BU feature. Of

the 91,177 features identified, 5,875 were identified as merges. The large prevalence of merged features (see Figure 8) alters the results reported in terms of feature properties, such as surface area, amplitude, and position of centroid, as well as their tracks.

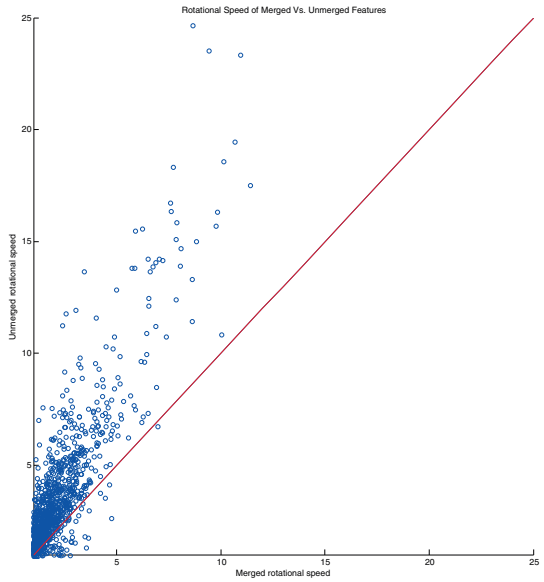

Fig. 9. Rotational speeds of the merged TD features, and their corresponding un-merged BU counterparts. BU is able to un-merge a large number of large TD features while maintaining high rotational speeds, which are characteristics of non-linear eddies.

1) Impact on rotational speed: Figure 9 shows the mean rotational speed of the large TD merged feature and the maximum of the n un-merged or corrected features from BU. We select the maximum speed to show a one-to-one comparison. The red diagonal line denotes the features where both the merged and un-merged features have the same rotational speed. The points that are above the red line represent the features that have a faster rotational speed after un-merging. In most cases, the corrected features have a larger rotational speed than the merged feature. A closer inspection of the area near the $(0, 0)$ origin, shows that many large features that have weak rotational speed tend to also have equally low un-merged rotational speeds. These are the cases where a true merge occurred and two weakening eddies might merge or split and dissipate (terminate).

2) Impact on displaced centroids: Another effect of merging multiple features is that the centroids for merged features will be in the middle of the merged features instead of at the center of each individual feature. To quantify the centroid accuracy of each method, we used Chelton et al. [5]'s definition of the tightest possible contour of the feature by selecting the contour with the highest average rotational speed. Given that such a feature would be the most compact, its centroid is most likely to be the feature's "true" centroid. For each feature identified by TD and BU, we compute its tightest contour and

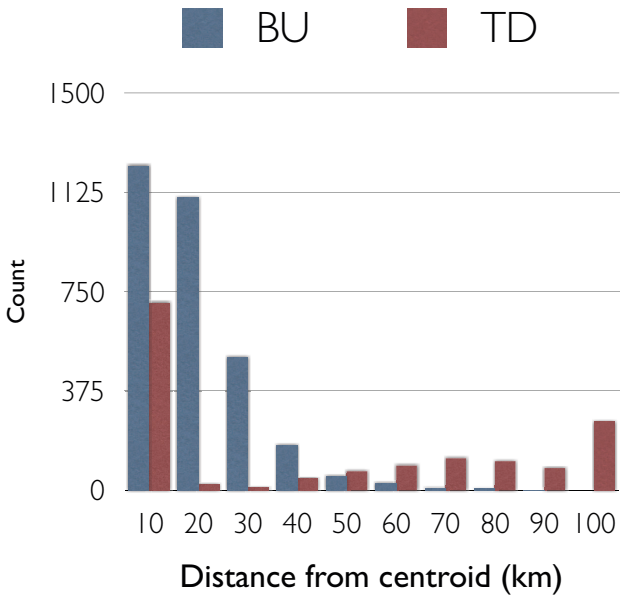

Fig. 10. The distance between the centroid of the contour with maximum rotational speed and the TD (maroon) and BU (blue) centroids. The centroid of the contour with maximum rotational speed is the most certain part of the eddy. BU centroids tend to be closer than TD centroids due to more compact bodies and by avoiding merges where centroids are severely displaced.

measure the distance between the optimal contour's centroid and that of the original feature. We find that in most instances BU centroids are closest to the "true" centroids.

3) Impact on track Lengths: A final significant impact of artificial merges is altering reported track lengths. When a feature is propagating and is attached to a merged feature it causes the features associated with the merge to have one of their tracks terminated. Additionally, artificially merged features tend to extend eddy lifetimes since the distorted centroid would be closer to unrelated features and effectively extend certain tracks.

To investigate both of these side effects, we analyzed the tracks associated with features that were labeled as potential merges. Figure 11 shows how features that were deemed insignificant due to artificial merges may have a more significant lifetimes than reported. While artificially merged features can have a multitude of cascading effects on neighboring tracks, we will focus on the impact of merges on non-persisting features. In panel (a) of Figure 11, we notice that a large number of tracks associated with merged features do not persist for more than four weeks. However, it is unclear if these features are truly spurious or whether the short lifetimes are due to the artificial merging. We looked at all the tracks associated with BU features that resulted from un-merging the TD features that did not persist for more than a week (leftmost column in panel (a)). The lifetimes of such un-merged BU features are shown in panel (b). Although 500+ of the un-merged features do not persist over a single week, nearly 200 persist for more than 4 weeks compared to the TD reports of such features terminating after a single week.

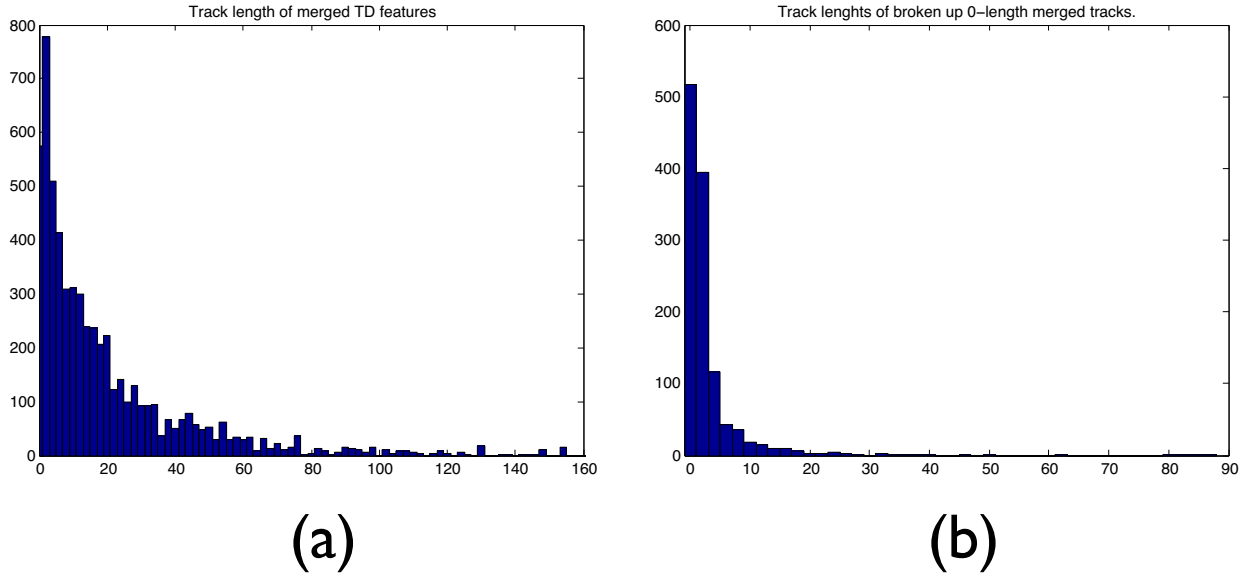

Fig. 11. Impact of merges on reported track lifetimes. Panel (a) shows the track lifetimes of features associated with TD merges. Of the features that persisted for 1 week (far left bar in panel a), we analyzed the lifetimes of the un-merged features found by BU (panel b). We find a significant number of features that we label as insignificant (less than 4 weeks) by TD as having persisted more than 4 weeks by BU after un-merging

VI. Conclusion and future work

In this paper we presented a parameter-free method to identify patterns in continuous spatio-temporal data. We are able to reproduce more significant features than the most widely used eddy identification scheme which employes numerous expert-defined parameters. We also presented numerous analyses that give an in-depth look into the various challenges researchers face when mining large unlabeled climate datasets. As the size of climate datasets continue to grow and the need for rapid exploratory research tools become crucial, we must pay special attention to three aspect of the pattern mining process:

1) Object definition and identification: The first challenge is to be able to define a signal that characterizes the feature of interest. This has usually been done using domain expertise to define a feature's signal on the continuous field. Such an approach is not always desirable since we have significant knowledge gaps in many domains where large dataset exist. Therefore, an exploratory feature identification process might be preferable, especially in large datasets.
2) Performance of spatio-temporal learning algorithms: Most of the problems at hand have no reliable "ground truth" data and therefore rely on unsupervised learning techniques. Hence, it is crucial to develop objective performance measures and experiments that allow to compare the performance of different spatio-temporal data mining algorithms.
3) Significance testing of features: A major challenge is the ability to distinguish a meaningful signal from noise, that is once a signal has been discovered how likely does a feature match a signal at random? This is especially true in exploratory research in large datasets where a very large number of relationships are tested and, effectively increase the likelihood of observing a strong statistic by random chance.

Acknowledgment

This work was funded by a University of Minnesota Doctoral Dissertation Fellowship and an U.S. National Science Foundation Expeditions in Computing Grant IIS-1029711. Access to computing facilities was provided by the University of Minnesota Supercomputing Institute.

References

[1] C. L. Bain, J. De Paz, J. Kramer, G. Magnusdottir, P. Smyth, H. Stern, and C.-c. Wang. Detecting the itcz in instantaneous satellite data using spatiotemporal statistical modeling: Itcz climatology in the east pacific. *Journal of Climate*, 24(1):216–230, 2011.
[2] A. Chaigneau and O. Pizarro. Mean surface circulation and mesoscale turbulent flow characteristics in the eastern south pacific from satellite tracked drifters. *J. Geophys. Res*, 110:C05014, 2005.
[3] A. Chaigneau, A. Gizolme, and C. Grados. Mesoscale eddies off peru in altimeter records: Identification algorithms and eddy spatio-temporal patterns. *Progress in Oceanography*, 79(2-4):106–119, 2008.
[4] D. Chelton, M. Schlax, R. Samelson, and R. de Szoeke. Global observations of large oceanic eddies. *Geophysical Research Letters*, 34:L15606, 2007.
[5] D. Chelton, M. Schlax, and R. Samelson. Global observations of nonlinear mesoscale eddies. *Progress in Oceanography*, 2011.
[6] C. Dong, F. Nencioli, Y. Liu, and J. McWilliams. An automated approach to detect oceanic eddies from satellite

remotely sensed sea surface temperature data. *Geoscience and Remote Sensing Letters, IEEE*, (99):1–5, 2011.

[7] J. H. Faghmous, L. Styles, V. Mithal, S. Boriah, S. Liess, F. Vikebo, M. d. S. Mesquita, and V. Kumar. Eddyscan: A physically consistent ocean eddy monitoring application. In *Intelligent Data Understanding (CIDU), 2012 Conference on*, pages 96 –103, oct. 2012.

[8] J. H. Faghmous, M. Uluyol, L. Styles, M. Le, V. Mithal, S. Boriah, and V. Kumar. Multiple hypothesis object tracking for unsupervised self-learning: An ocean eddy tracking application. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[9] F. Fang and R. Morrow. Evolution, movement and decay of warm-core leeuwin current eddies. *Deep Sea Research Part II: Topical Studies in Oceanography*, 50(12-13): 2245–2261, 2003.

[10] A. Fernandes. Identification of oceanic eddies in satellite images. *Advances in Visual Computing*, pages 65–74, 2008.

[11] A. Fernandes and S. Nascimento. Automatic water eddy detection in sst maps using random ellipse fitting and vectorial fields for image segmentation. In *Discovery Science*, pages 77–88. Springer, 2006.

[12] L. Fu, D. Chelton, P. Le Traon, and R. Morrow. Eddy dynamics from satellite altimetry. *Oceanography*, 23(4): 14–25, 2010.

[13] D. Henke, P. Smyth, C. Haffke, and G. Magnusdottir. Automated analysis of the temporal behavior of the double intertropical convergence zone over the east pacific. *Remote Sensing of Environment*, 123:418–433, 2012.

[14] H. Hotelling. The generalization of student's ratio. *The Annals of Mathematical Statistics*, 2(3):pp. 360–378, 1931. ISSN 00034851. URL http://www.jstor.org/stable/2957535.

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

[16] J. Isern-Fontanet, J. Font, E. García-Ladona, M. Emelianov, C. Millot, and I. Taupier-Letage. Spatial structure of anticyclonic eddies in the algerian basin (mediterranean sea) analyzed using the okubo–weiss parameter. *Deep Sea Research Part II: Topical Studies in Oceanography*, 51(25):3009–3028, 2004.

[17] J. Isern-Fontanet, E. García-Ladona, and J. Font. Vortices of the mediterranean sea: An altimetric perspective. *Journal of physical oceanography*, 36(1):87–103, 2006.

[18] B. Jaimes and L. K. Shay. Mixed layer cooling in mesoscale oceanic eddies during hurricanes katrina and rita. *Monthly Weather Review*, 137(12):4188–4207, 2009.

[19] E. Keogh, S. Lonardi, and C. A. Ratanamahatana. Towards parameter-free data mining. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 206–215. ACM, 2004.

[20] D. McGillicuddy Jr. Eddies masquerade as planetary waves. *Science*, 334(6054):318–319, 2011.

[21] E. Mesrobian, R. Muntz, E. Shek, J. Santos, J. Yi, K. Ng, S.-Y. Chien, C. Mechoso, J. Farrara, P. Stolorz, et al. Exploratory data mining and analysis using conquest. In *Communications, Computers, and Signal Processing, 1995. Proceedings., IEEE Pacific Rim Conference on*, pages 281–286. IEEE, 1995.

[22] R. Morrow, F. Birol, D. Griffin, and J. Sudre. Divergent pathways of cyclonic and anti-cyclonic ocean eddies. *Geophysical Research Letters*, 31(24), 2004.

[23] W. Pegau, E. Boss, and A. Martínez. Ocean color observations of eddies during the summer in the gulf of california. *Geophysical Research Letters*, 29(9):1295, 2002.

[24] P. Rahul, P. Salvekar, B. Sahu, S. Nayak, and T. S. Kumar. Role of a cyclonic eddy in the 7000-year-old mentawai coral reef death during the 1997 indian ocean dipole event. *Geoscience and Remote Sensing Letters, IEEE*, 7(2):296–300, 2010.

[25] E. N. Rappaport, J. L. Franklin, A. B. Schumacher, M. DeMaria, L. K. Shay, and E. J. Gibney. Tropical cyclone intensity change before us gulf coast landfall. *Weather and Forecasting*, 25(5):1380–1396, 2010.

[26] A. Sorokin and D. Forsyth. Utility data annotation with amazon mechanical turk. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2008.

[27] P. Stolorz, E. Mesrobian, R. Muntz, J. Santos, E. Shek, J. Yi, C. Mechoso, and J. Farrara. Fast spatio-temporal data mining from large geophysical datasets. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, pages 300–305, 1995.

[28] L. von Ahn, S. Ginosar, M. Kedia, and M. Blum. Improving image search with PHETCH. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV–1209. IEEE, 2007.

[29] L. Von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum. reCAPTCHA: Human-based character recognition via web security measures. *Science*, 321 (5895):1465–1468, 2008.