

## RESEARCH LETTER

10.1002/2016GL071269

## Key Points:

- The phase angle between the components of the geostrophic velocity anomalies is proposed to identify eddy structures
- Machine learning models are trained to detect the phase patterns of eddies; we demonstrate that this method is largely region independent
- The proposed method eliminates the previously reported issue of eddy misclassification and is also capable of estimating the eddy lifetime

## Supporting Information:

- Supporting Information S1

## Correspondence to:

M. D. Ashkezari,  
demo@mit.edu

## Citation:

Ashkezari, M. D., C. N. Hill, C. N. Follett, G. Forget, and M. J. Follows (2016), Oceanic eddy detection and lifetime forecast using machine learning methods, *Geophys. Res. Lett.*, 43, 12,234–12,241, doi:10.1002/2016GL071269.

Received 20 SEP 2016

Accepted 15 NOV 2016

Accepted article online 21 NOV 2016

Published online 15 DEC 2016

## Oceanic eddy detection and lifetime forecast using machine learning methods

Mohammad D. Ashkezari<sup>1</sup>, Christopher N. Hill<sup>1</sup>, Christopher N. Follett<sup>1</sup>, Gaël Forget<sup>1</sup>, and Michael J. Follows<sup>1</sup>
<sup>1</sup>Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

**Abstract** We report a novel altimetry-based machine learning approach for eddy identification and characterization. The machine learning models use daily maps of geostrophic velocity anomalies and are trained according to the phase angle between the zonal and meridional components at each grid point. The trained models are then used to identify the corresponding eddy phase patterns and to predict the lifetime of a detected eddy structure. The performance of the proposed method is examined at two dynamically different regions to demonstrate its robust behavior and region independency.

## 1. Introduction

Ocean eddies play a key role in the dynamics of the world's oceans, mediating both the horizontal and vertical transport of physical, chemical, and biological tracers [e.g., see *Forget et al.*, 2015b]. These fluxes, in turn, become critical for problems ranging from ocean primary production to climate change [*Chelton et al.*, 2011a; *Beal et al.*, 2011]. However, the transport is nonlinear, so calculating eddy fluxes requires measuring and predicting the distributions and evolution of eddy characteristics. Considering the complex nature of eddy dynamics, combining machine learning algorithms with decades of satellite data could greatly enhance our ability to first identify eddy features and then further forecast their evolutions. Machine learning (statistical learning) approaches particularly stand out in analyzing complex systems that have extensive observations and yet are not fully understood.

During the last few decades various models have been suggested and employed for eddy detection using altimetry data. Arguably, one of the most established metrics to distinguish hyperbolic (high deformation) from elliptic (high vorticity) subdomains within a flow field is the Okubo-Weiss criterion [*Okubo*, 1970; *Weiss*, 1991]:

$$W(\mathbf{x}, t) = s^2(\mathbf{x}, t) + \zeta^2(\mathbf{x}, t), \quad (1)$$

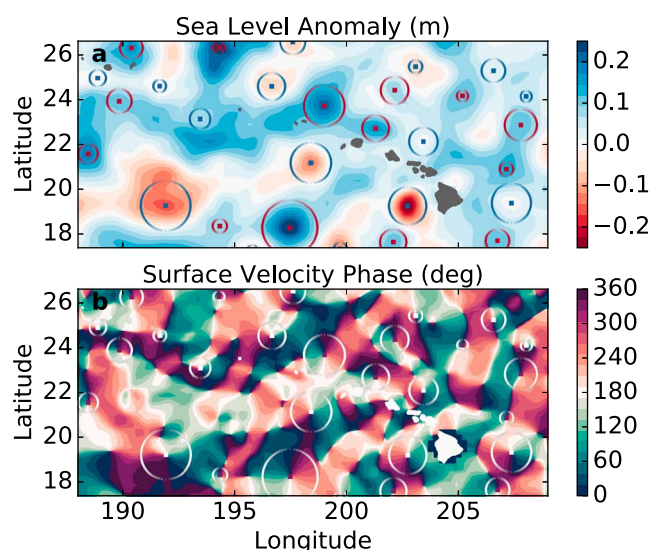
where  $s$  and  $\zeta$  represent strain and vorticity, respectively such that

$$s^2(\mathbf{x}, t) = \left( \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right)^2 + \left( \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right)^2 \quad (2)$$

$$\zeta^2(\mathbf{x}, t) = \left( \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right)^2, \quad (3)$$

where  $u$  and  $v$  are the zonal and meridional components of the flow and  $\mathbf{x} = (x, y)$  represents the coordinates. Choosing a predefined threshold value  $W_o$ , one can partition the domain into smaller subdomains for eddy identification. This method, however, strongly relies on the predefined threshold  $W_o$  which varies across different regions. Therefore, although it can be applied globally (with a constant  $W_o$  value) [*Chelton et al.*, 2007], the best results are achieved on a specific region with a tuned  $W_o$  parameter [*d'Ovidio et al.*, 2009; *Bracco et al.*, 2000].

As an alternative to the Okubo-Weiss method, one may directly use Sea Level Anomaly (SLA) data to identify local extrema and define the eddy boundary as the outermost closed contour of the SLA. The challenge,



**Figure 1.** (top) One daily map of SLA around the Hawaiian Islands. (bottom) The corresponding map of velocity phase (equation (4)) ranging from  $0^\circ$  to  $360^\circ$ . Eddy cores identified in section 3 are denoted by squares, and estimates of eddy radius are shown by circles. In Figure 1 (top) blue squares represent cyclonic eddies while red squares represent anticyclonic eddies.

however, is that methods like these still require expert-tuned parameters [Chelton *et al.*, 2011b; Mason *et al.*, 2014; Chaigneau *et al.*, 2009] or exhibit notable rates of eddy misclassification (inconsistency between eddy polarity and the corresponding sign of relative vorticity) [Faghmous *et al.*, 2015].

Aiming to improve on the previous methods, we combine a simple eddy model with machine learning techniques to detect eddy structures and forecast their lifetimes. The proposed method does not require region-specific tuned parameters, significantly alleviates the issue of eddy misclassification, and offers a foundation for a machine-learning-based platform to build predictive analytics solutions for mesoscale eddies. The quality of the method's outcomes is assessed by comparing them with other methods and with in situ measurements.

## 2. Methodology and Data Source

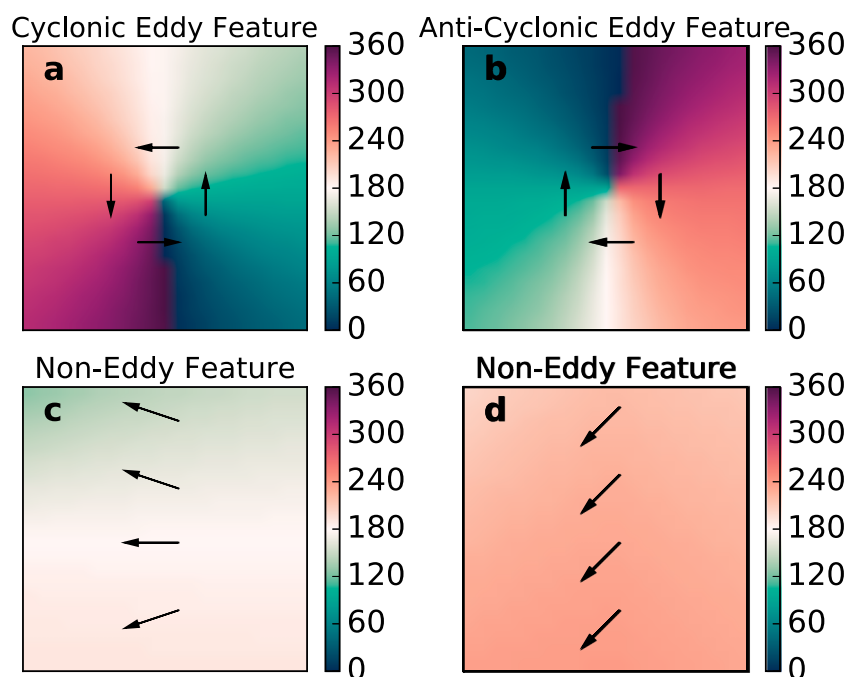
We employ geostrophic velocity anomalies inferred from satellite altimetry data as inputs to our model and compute the phase angle between the meridional  $V$  and zonal  $U$  components of the flow field:

$$\Phi = \tan^{-1} \left( \frac{V}{U} \right). \quad (4)$$

Therefore, for every set of flow fields  $U, V$  (at a given day) we calculate the corresponding phase field  $\Phi$ , and then investigate the phase field to identify eddy patterns. An example along with the corresponding SLA map is shown in Figure 1. An eddy core emerges as circular sequences of phase values (direction of the velocity vectors) from  $0^\circ$  to  $360^\circ$  and produces a distinct signature that can be detected by the machine learning algorithms (Figure 2). On the other hand, a mean flow, for instance, would translate in rather uniform phase contours aligned with the flow (Figure 2).

The altimetry data used in this work are produced by Salto/Duacs and are provided by the Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO; <http://aviso.altimetry.fr/en/home.html>) [Capet *et al.*, 2014]. The bulk of our analysis relies on the “Delayed-Time-Allsat” daily global geostrophic velocity anomalies data set. The altimetry data are on a grid with  $0.25^\circ$  resolution and covers a nearly 23 year period, from January 1993 to September 2015.

This study is motivated by the Simons Collaboration on Ocean Processes and Ecology (SCOPE) field experiment. SCOPE is a collaborative effort to conduct highly resolved experiments and analyses over multiple levels of biological organization at a representative ocean benchmark, Station ALOHA, located in the North Pacific Subtropical Gyre. Therefore, we focus our analysis on the region delimited by  $17^\circ\text{N}$ – $29^\circ\text{N}$  and  $151^\circ\text{W}$ – $172^\circ\text{W}$  (that includes the Hawaiian Islands). We interpolate the geostrophic velocity anomalies onto a finer grid with a spatial resolution of  $0.01^\circ$  before computing the phase values using equation (4) so that each AVISO grid cell is associated with a smooth phase pattern (Figure 2). Interpolating to fine resolutions will make it possible to consistently apply this method on the model-generated fine-resolution flow fields, as well. We discard detected eddies with radii smaller than 25 km since they are smaller than the AVISO grid spacing (section 3.2) or the size of the training samples (section 3). Depending on the study, one may apply a different radius filter and only accept larger eddy features [Chelton *et al.*, 2011b]. The source code developed for this study to identify and characterize eddies is publicly accessible at <https://github.com/eddML/core>.



**Figure 2.** A selection of the features extracted for training-testing purposes. Color shading represents phase value (equation (4)). (a, b) Cyclonic and anticyclonic eddy cases. Eddies are characterized by circular patterns where phase values show nearly concentric patterns of  $0^\circ$  to  $360^\circ$ . (c, d) Two noneddy patterns. Figure 2c represents a weakly divergent region, while Figure 2d exhibits a mean flow.

### 3. Eddy Core Detection

To prepare the training set, 40 phase fields were computed using equation (4). The first 40 days of the year 2014 were used to calculate these phase fields. A balanced set of eddy and noneddy structures were manually identified for training and validation purposes. To extract the eddy and noneddy structures, we used a fixed-size selection mask (a matrix with dimensions  $21 \times 21$ ), approximately  $23 \text{ km} \times 23 \text{ km}$  in size. In the case of eddy structures, the selection mask only picks up the core region of the eddy.

The training-test data set consists of a total 356 instances out of which half (178 instances) are eddies while the others represent noneddy structures. Also, 50% of the eddy instances ( $178/2 = 89$ ) are cyclonic eddies, and the rest are anticyclonic eddies. The number of training-test samples was increased until the desired classification performances were achieved (see section 3.1) and the effect of adding more training samples was marginal. Each instance is numerically labeled as eddy (label = 1) or noneddy (label = -1). Resolving all possible noneddy configurations is not necessary since we do not aim to subclassify noneddies. When an eddy structure is identified, the polarity of the eddy is also recorded in the label vector (cyclonic = 1, anticyclonic = -1). Figure 2 shows samples of these eddy/noneddy features.

Each instance then is turned into a  $21 \times 21 = 441$  dimensional numerical vector storing the phase values. These vectors construct the inputs in order to train and test the classifier.

#### 3.1. Eddy Classification: Training and Testing

Given the extracted feature data set, we take 75% of the data set for training and the rest is used for testing purposes. The splitting procedure is carried out so that the balance between different classes (eddy versus noneddy, and cyclonic versus anticyclonic) is maintained.

In order to reach higher identification accuracy, we designed a double-tier classification process; first, eddy cores are identified and then the polarity of the detected eddies is determined. Therefore, we trained two separate classifiers, one for eddy identification and the other for eddy polarity. We examined multiple classification models for each tier of the identification process. Two of the models stood out: Support Vector Machines (SVMs) classifier with radial basis function kernel [Guyon *et al.*, 1992; Cortes and Vapnik, 1995] and random forest classifier [Breiman, 2001] (see supporting information Text S1). In comparison with more

complex models such as artificial neural network [Lippmann, 1987], SVM and Random Forest involve simpler algorithms and require less training samples.

We examined the quality of predictions made by the above mentioned models on the test data set by evaluating various classification metrics including “accuracy” and “precision.” The SVM scored an accuracy of  $0.92 \pm 0.03$  and a precision of  $0.99 \pm 0.01$ , while random forest classifier resulted in an accuracy of  $0.97 \pm 0.02$  and a precision of  $0.96 \pm 0.03$  (see supporting information for more details, Table S1). These scores are associated with identification of eddy structures from those of noneddies. The uncertainties indicate one standard deviation on multiple (tenfold) realizations. As a quick reminder the presented classification metrics are defined as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \end{aligned} \quad (5)$$

where TP indicates the number of true positive recognitions, TN is the number of true negative recognitions, FP is the number of false positive recognitions (a noneddy recognized as eddy), and FN corresponds to the number of false negative recognitions (an eddy recognized as noneddy). Although the random forest classifier showed a higher accuracy metric, SVM was chosen to identify eddy structures due to its superior precision and speed. In the same vein, SVM classifier was also used for eddy polarity identification.

### 3.2. Eddy Classification: Implementation

After training and validation processes, we apply the trained classifiers to the entire regional data set—the nearly 23 years of satellite altimetry. For each day we compute the phase field, similar to Figure 1b, and then scan the field using a typical “sliding window” algorithm to search for eddy cores. At any given location  $(i, j)$ , the  $21 \times 21$  selection mask is applied and the extracted phase data are then fed into the classifier to determine whether it corresponds to an eddy or a noneddy structure. If the region is recognized as an eddy core, the phase data are passed to the second classifier to identify the eddy polarity. The region corresponding to the detected eddy core is marked to avoid multievaluation of a single eddy.

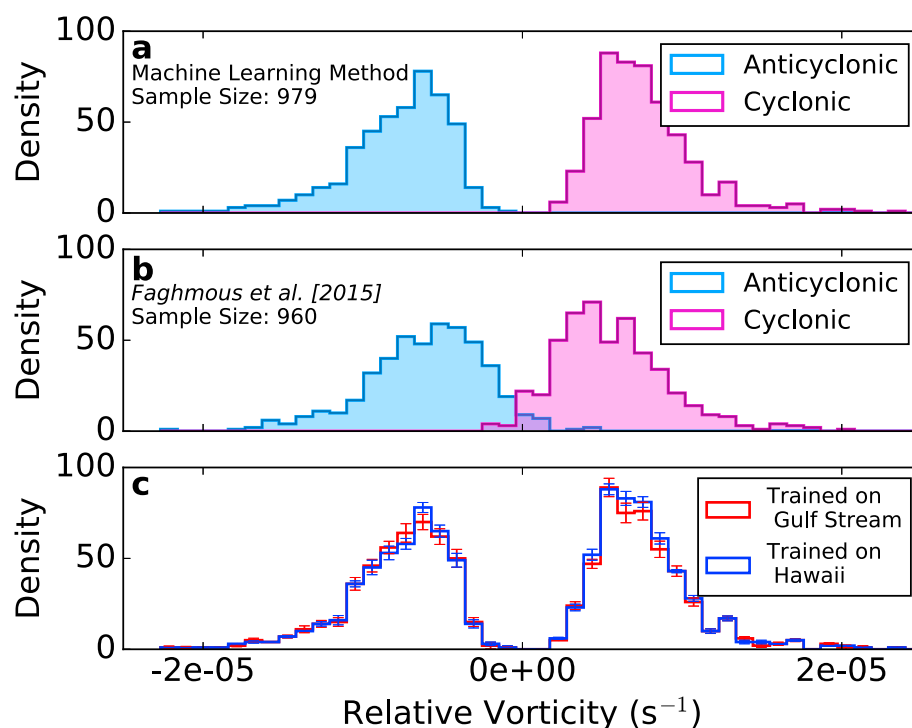
Once the eddy core is identified, we further estimate the corresponding eddy radius. From the surface velocity phase perspective, an ideal eddy is described with concentric circular phase values starting from  $0^\circ$  and linearly increasing to  $360^\circ$ . Therefore, to estimate the eddy radius, we extract the phase values on the border of a circle located at eddy core with a radius  $r$ . We then fit a line on the extracted phase values; if the goodness of fit ( $R^2$ : coefficient of determination) is above a predefined value ( $R^2 \geq 0.80$ ), the extracted region is considered inside the eddy domain. This process is repeated with larger  $r$  values until the fit quality falls below the predefined threshold. In order to reach lower rates of false positive detections, we impose a minimum radius constraint of 25 km. The results are illustrated in Figure 1 for one snapshot.

### 3.3. Eddy Classification: Assessment

To assess the proposed method, we first compare its results with those of the recent SLA-based eddy identification algorithm proposed by Faghmous *et al.* [2015] who provide a global database of identified eddies (<https://github.com/jfaghm/OceanEddies> and <http://climatechange.cs.umn.edu/eddies/>). For this comparison, we select 30 random days and examine the overlap between the two sets of detected eddy cores. Both methods find about 1000 eddy cores and 75% of the detected eddy cores match to within 50 km.

We further compute relative vorticity maps for the 30 selected days and evaluate relative vorticity at the location of each eddy core. Figures 3a and 3b compare the relative vorticity distributions obtained by both methods for cyclonic and anticyclonic eddies. Our method (Figure 3a) eliminates the overlap between vorticity distributions of the identified cyclonic and anticyclonic eddies (Figure 3b) and alleviates the issue of misclassification (up to 16%) reported by Faghmous *et al.* [2015]. Moreover, on average the relative vorticity associated with eddy cores identified using our method is approximately 30% higher than those reported by Faghmous *et al.* [2015], which may indicate a more precise localization of eddy cores.

In order to investigate the regional sensitivity of the method, the machine learning classifiers used to identify eddies are trained at geographically different domains and then are applied to the Hawaiian region. Figure 3c is identical to Figure 3a except that it compares the results of our method using two different training sets. In one case we train our model at the Hawaiian region, while in the other case we take samples from the Gulf



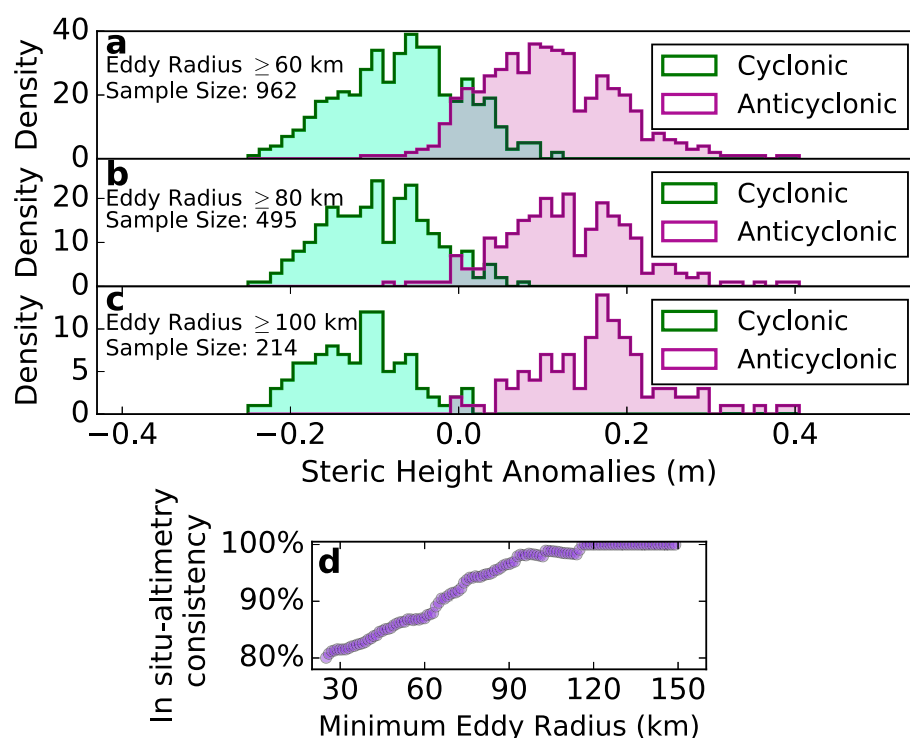
**Figure 3.** (a, b) The distribution of relative vorticity at the location of detected eddy cores using the method presented in this study and by Faghmous et al. [2015]. Relative vorticity is computed to within 30 km of the cores. (c) Comparison of the distribution of relative vorticity of the detected eddies at the Hawaiian region using two different training sets. The error bars indicate one standard deviation on multiple realizations.

Stream region, a dynamically distinct context, for training purposes. Under the null hypothesis, namely, both distributions shown in Figure 3c are drawn from the same parent distribution, we employ a two-sample  $K$ -S test statistic [Stephens, 1970] which results in a corresponding  $p$  value  $\approx 0.99$ , confirming a strong consistency between the two distributions. Therefore, the observed consistency between the results associated with the two distinctly trained models demonstrates that the method does not rely on local or region-specific training samples. The same procedure was applied to the Gulf Stream region. Again, close similarity is observed between the identified eddies using different trained models confirming that the proposed method is largely region independent (see supporting information Figure S4).

Besides errors that may derive from imperfect algorithms, eddy detection results could be contaminated by errors inherited from the AVISO product (which involves elaborate data processing, noise filtering, and interpolation procedures). To further assess our results, we therefore rely on independent and pointwise in situ profile data: the MITprof data collections [Forget et al., 2015a; <http://dx.doi.org/10.7910/DVN/EE3C40>]. It should be stressed that such data are relatively sparse (particularly until the advent of Argo) and carry uncertainties of their own, so they can only provide an imperfect “ground truth.”

The comparison procedure is as follows: for each in situ measurement we search for an eddy core detected by our algorithm at the same date and to within a 50 km radius. If a match is found, the detected eddy polarity and steric height anomaly inferred from the in situ measurement are recorded. Other in situ characteristics of the eddy field could have been chosen for this comparison, but steric height is relatively simple to evaluate from profile data, and it is comparable with altimetry [e.g., Forget and Wunsch, 2007; Forget and Ponte, 2015].

Figures 4a–4c show distributions of steric height anomalies at detected eddy cores of both polarities. It appears that the sign of steric height anomalies is consistent with the polarity of detected eddies in 80% of cases, and the consistency increases when excluding the smaller detected eddies (see Figure 4d). The polarity of cyclones and anticyclones with radius larger than 80 km identified using AVISO maps and our method is thus found consistent with the steric height anomalies sign inferred from in situ measurements to within 95% confidence level.



**Figure 4.** (a–c) The distribution of steric height anomalies computed from in situ profiles located in close proximity of eddy cores (cyclones or anticyclones) detected by our method. When progressively excluding the smaller eddies, the overlap of the two distributions is reduced. (d) This behavior is illustrated and shows the degree of consistency between steric height anomaly sign and polarity of the matched eddies as a function of minimum eddy radius.

#### 4. Eddy Tracking and Statistics

Eddy cores identified in consecutive phase fields are then aggregated to construct eddy trajectories and time series (in analogy with *Chelton et al.* [2011b] and *Faghmous et al.* [2015]). At each time step (a day), we record the current time, eddy polarities, and eddy core locations. We then attempt to associate eddies in the current frame with eddies in the successive frame. For each eddy core in the current frame, we search within a predefined radius (approximately 40 km) in the successive frame for an eddy core that matches the polarity of the original eddy.

Since AVISO maps can be contaminated by errors and the recognition accuracy of the machine learning model is not perfect, it is possible that an eddy core is missed at some point during the lifetime of an eddy. Therefore, we allow for temporary disappearance of eddies, but if the disappearance exceeds 2 days, then the trajectory is terminated and the eddy is considered dead.

The tracking algorithm was applied to the entire regional Hawaii data set, and approximately 17,000 trajectories were reconstructed. Analysis of such estimates provides various eddy characteristics, such as lifetime, displacement, propagation velocity, and acceleration. In particular, it is encouraging that the lifetime and displacement distributions are qualitatively similar to those reported by *Chelton et al.* [2011b] and *Mason et al.* [2014] who use very different numerical methods (see supporting information Figure S2).

#### 5. Eddy Lifetime Prediction

The finite lifetime of ocean eddies implies that their energy gets reduced through dissipative interactions with external forces imposed by winds, collisions with neighboring eddies, etc. Although the dissipative mechanisms may not be fully known, the machine learning algorithm allows us to use their effects on measured variables to predict the lifetime of an eddy. One potential application of an eddy lifetime predictive model is in planning field campaigns where a clear prediction of the longevity of the targeted feature can be helpful.



Under small perturbations, an eddy is expected to follow a “smooth evolution” curve. For instance, its corresponding SLA value may start from a background level, pass through an extremum (maximum for anticyclonic and minimum for cyclonic eddies), and then return to the background level [Samelson *et al.*, 2014; Escudier *et al.*, 2016]. In this section we focus on such smoothly evolved eddies. We searched through the data set of eddy trajectories (section 4) and selected 900 smoothly evolved trajectories with a nearly balanced population of cyclonic and anticyclonic eddies. These eddies are then used for training and validation of the predictive model. Distributions and statistical summaries regarding the smoothly evolved eddies may be found in the supporting information (Table S3 and Figure S5).

The goal is to develop a machine learning method to estimate the lifetime of an eddy based on the instantaneous features of the eddy that can be inferred from altimetry. Therefore, for each eddy and each day during the course of its evolution, we construct a numerical vector containing a list of 10 potentially relevant features to the lifetime of the eddy. They involve eddy properties such as the associated phase values, SLA, and relative vorticity (see supporting information Table S4).

Each of these vectors represents an eddy sample for which we know the remaining eddy lifetime. More than 22,000 numerical vectors are generated and fed to an “Extra Trees Regressor” model [Geurts *et al.*, 2006]. The model is trained with a random selection of 90% of the samples and is tested on the remaining samples. This process is then repeated 10 times to obtain a more accurate estimate of the prediction quality metrics associated with the trained model. The results show that the predictive model forecasts the lifetime of eddy samples with a root-mean-square error of  $5.4 \pm 4.7$  days. Alternatively, the quality of the predictions may be expressed in terms of the coefficient of determination  $R^2 = 0.92 \pm 0.01$ . Predictions made by the trained model are illustrated in the supporting information (Figure S6).

## 6. Conclusion and Prospect

In this study we propose that the collective phase patterns associated with the angle between the zonal and meridional components of the flow field are valuable indicators of the eddy cores and are correlated with eddy characteristics such as the eddy size.

To explore this idea, we take advantage of the 23 years of the geostrophic velocity anomalies made available by AVISO and train machine learning models to identify eddy cores and their polarity. The detection method is solely based on phase patterns and does not require any a priori knowledge of eddy sizes. As a result of the very distinct phase patterns associated with eddies of different polarities, our experiments show a reduction in misclassification as compared with previous SLA-based studies [Faghmous *et al.*, 2015]. Moreover, we demonstrate that the proposed method is significantly insensitive to the choice of local training samples and does not require a region-specific tuned parameter.

The scope of our study further includes the modeling and prediction of eddy lifetime. For that application we construct a numerical vector compiling 10 instantaneous features of an eddy inferred from altimetry data. An Extra Trees Regressor model was used for training, and our tests indicate that the model can determine the lifetime of an eddy with a root-mean-square error of  $\sim 5$  days.

The proposed method could, of course, be refined further. For example, it only provides a crude estimate of the eddy size as compared with methods that are based on closed SLA contours. Also, the suggested predictive lifetime model relies on smoothly evolved eddies while this assumption may be violated under strong external perturbations. We suspect that introducing additional features in the lifetime predictive model to account for the influence of neighboring eddies would enhance the forecast accuracy.

This study focuses on a region of North Pacific Subtropical Gyre centered around the Hawaiian Islands. In the future, the method presented here will be applied at the global scale. The computational performance of the proposed eddy identification method is expected to scale linearly with the area. The aim of the global study will be toward mapping out both eddy properties and their influence on transport.

### Acknowledgments

This work was supported by a grant from the Simons Foundation (SCOPE Award 329108, Michael J. Follows). The altimetry data were produced by Salto/Duacs and distributed by AVISO [Capet *et al.*, 2014; <http://aviso.altimetry.fr/en/home.html>].

### References

- Beal, L. M., W. P. De Ruijter, A. Biastoch, R. Zahn, and SCOR/WCRP/IAPSO Working Group 136 (2011), On the role of the Agulhas System in ocean circulation and climate, *Nature*, 472, 429–436.
- Bracco, A., J. LaCasce, C. Pasquero, and A. Provenzale (2000), The velocity distribution of barotropic turbulence, *Phys. Fluids*, 12, 2478–2488.
- Breiman, L. (2001), Random Forests, *Machine Learning*, 45(1), 5–32.

- Capet, A., E. Mason, V. Rossi, C. Troupin, Y. Faugère, I. Pujol, and A. Pascual (2014), Implications of refined altimetry on estimates of mesoscale activity and eddy-driven offshore transport in the eastern boundary upwelling systems, *Geophys. Res. Lett.*, **41**, 7602–7610, doi:10.1002/2014GL061770.
- Chaigneau, A., E. Gérard, and B. Dewitte (2009), Eddy activity in the four major upwelling systems from satellite altimetry (1992–2007), *Prog. Oceanogr.*, **83**, 117–123.
- Chelton, D., P. Gaube, M. Schlax, M. G. Early, and R. M. Samelson (2011a), The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll, *Science*, **334**, 328–332.
- Chelton, D., M. Schlax, and R. Samelson (2011b), Global observations of nonlinear mesoscale eddies, *Prog. Oceanogr.*, **91**, 167–216.
- Chelton, D., M. Schlax, R. Samelson, and R. de Szoeke (2007), Global observations of large oceanic eddies, *Geophys. Res. Lett.*, **34**, L15606, doi:10.1029/2007GL030812.
- Cortes, C., and V. Vapnik (1995), Support-vector networks, *Machine Learning*, **20**, 273–297.
- d'Ovidio, F., J. Isern-Fontanet, C. López, E. Hernández-García, and E. García-Ladona (2009), Comparison between Eulerian diagnostics and finite-size Lyapunov exponents computed from altimetry in the Algerian basin, *Deep Sea Res., Part I*, **56**, 15–31.
- Escudier, R., L. Renault, A. Pascual, P. Brasseur, D. Chelton, and J. Beuvier (2016), Eddy properties in the western Mediterranean Sea from satellite altimetry and a numerical simulation, *J. Geophys. Res. Oceans*, **121**, 3990–4006, doi:10.1002/2015JC011371.
- Faghmous, J. H., I. Frenger, Y. Yao, R. Warmka, A. Lindell, and V. Kumar (2015), A daily global mesoscale ocean eddy dataset from satellite altimetry, *Sci. Data*, **2**, 150028, doi:10.1038/sdata.2015.28.
- Forget, G., J.-M. Campin, P. Heimbach, C. N. Hill, R. M. Ponte, and C. Wunsch (2015a), ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation, *Geosci. Model Develop.*, **8**, 3071–3104.
- Forget, G., D. Ferreira, and X. Liang (2015b), On the observability of turbulent transport rates by Argo: Supporting evidence from an inversion experiment, *Ocean Sci.*, **11**, 839–853.
- Forget, G., and R. M. Ponte (2015), The partition of regional sea level variability, *J. Phys. Oceanogr.*, **137**, 173–195.
- Forget, G., and C. Wunsch (2007), Estimated global hydrographic variability, *J. Phys. Oceanogr.*, **37**, 1997–2008.
- Geurts, P., D. Ernst, and L. Wehenkel (2006), Extremely randomized trees, *Machine Learning*, **63**(1), 3–34.
- Guyon, I., B. Boser, and V. Vapnik (1992), Automatic capacity tuning of very large VC-dimension classifiers: Advances in neural information processing systems 5, in *Proceeding Advances in Neural Information Processing Systems*, vol. 5, pp. 147–155, Morgan Kaufmann, San Francisco, Calif.
- Lippmann, R. P. (1987), An introduction to computing with neural nets, *IEEE Acoustics, Speech, and Signal Processing Newsletter*, **4**, 4–22.
- Mason, E., A. Pascual, and J. C. McWilliams (2014), A new sea surface height-based code for oceanic mesoscale eddy tracking, *J. Atmos. Oceanic Technol.*, **31**, 1181–1188.
- Okubo, A. (1970), Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences, *Deep Sea Res.*, **17**, 445–454.
- Samelson, R. M., M. A. Schlax, and D. B. Chelton (2014), Randomness, symmetry, and scaling of mesoscale eddy life cycles, *Am. Meteorol. Soc.*, **44**, 1012–1029.
- Stephens, M. A. (1970), Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables, *J. R. Stat. Soc.*, **32**, 115–122.
- Weiss, J. (1991), The dynamics of enstrophy transfer in two-dimensional hydrodynamics, *Physica D*, **48**, 273–294.