# Machine-Learning Mesoscale and Submesoscale Surface Dynamics from Lagrangian Ocean Drifter Trajectories

NIKOLAS O. AKSAMIT

*Institute for Mechanical Systems, Swiss Federal Institute of Technology, Zurich, Switzerland*

THEMISTOKLIS SAPSIS

*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts*

GEORGE HALLER

*Institute for Mechanical Systems, Swiss Federal Institute of Technology, Zurich, Switzerland*

## ABSTRACT

Lagrangian ocean drifters provide highly accurate approximations of ocean surface currents but are sparsely located across the globe. As drifters passively follow ocean currents, there is minimal control on where they will be making measurements, providing limited temporal coverage for a given region. Complementary Eulerian velocity data are available with global coverage but are themselveslimited by the spatial and temporal resolution possible with satellite altimetry measurements. In addition, altimetry measurements approximate geostrophic components of ocean currents but neglect smaller submesoscale motions and require smoothing and interpolation from raw satellite track measurements. In an effort to harness the rich dynamics available in ocean drifter datasets, we have trained a recurrent neural network on the time history of drifter motion to minimize the error in a reduced-order Maxey–Riley drifter model. This approach relies on a slow-manifold approximation to determine the most mathematically relevant variables with which to train, subsequently improving the temporal and spatial resolution of the underlying velocity field. By adding this neural-network component, we also correct drifter trajectories near submesoscale features missed by deterministic models using only satellite and wind reanalysis data. The effect of varying similarity between training and testing trajectory datasets for the blended model was evaluated, as was the effect of seasonality in the Gulf of Mexico.

## 1. Introduction

Ocean currents play a critical role in the climate at time scales from weeks to decades. Ocean mixing is responsible for the scalar transport of temperature and salinity, impacting physical processes ranging from weather system formation, the melting rate of sea ice, the abundance and location of ocean ecosystems and the dispersion of microplastics (Trenberth and Hurrell 1994; Martin et al. 2002; Maximenko et al. 2012; Lévy et al. 2018). With the public availability of satellite altimetry measurements, researchers actively study global ocean currents and mesoscale features in the ocean, ranging in size from 10 to 200 km (Stewart 2008), in near–real time from an Eulerian perspective. These data have allowed a

better understanding of the role of prominent circulation features in the deep ocean, like the Gulf Stream and the Pacific Gyre, as well as smaller coherent structures like the Agulhas rings (Wang et al. 2015). In shallower waters, or at the interface of two mesoscale features, the influence of submesoscale dynamics on ocean transport increases. The interaction of meso- and submesoscale motions, however, is still poorly understood (Beron-Vera et al. 2019a). Satellite observations typically represent the mixing and energy transport in submesoscale features (1–10 km) inaccurately, with coastal features and tidal influences largely overlooked as they do not have significant sea surface height signatures (Thomas et al. 2008; Ferrari and Wunsch 2009).

Complementary Lagrangian measurements of ocean currents are also available, with their origins dating back to drifter deployments during the 1872 voyage of the

*Corresponding author*: N. O. Aksamit, naksamit@ethz.ch

HMS *Challenger* (Niiler 2001). Since then, the quality and coverage of Lagrangian drifter data have improved through the use of satellite positioning systems and improved drifter designs. Lagrangian drifters come in varying geometries but are universally composed of a drogue that is fully submerged in the water, behaving as a sail, a float on the ocean surface that communicates location and velocity data, and a tether connecting the two. The size of the entire apparatus is typically on the order of 1–5 m, with recent designs at the lower end of that range. The finite-size (inertial) effects and the buoyancy of every drifter prevent perfect adherence to fluid-particle path lines. Considerable research has been conducted quantifying the ocean-drifter slip velocity; thorough reviews of such processes have been presented over the years in great detail by Niiler et al. (1987), Niiler and Paduan (1995), D'Asaro (2003), and others. Even with the disadvantages of nonzero slip velocities, Lagrangian drifters have been pivotal in our understanding of the oceans by providing velocity and position data for a range of motions spanning meters to thousands of kilometers. Such drifters continue to play a crucial role in understanding physical oceanography in changing climates with improving technology (Lumpkin et al. 2017).

Typically, efforts to model Lagrangian drifter motions suffer in accuracy away from large structures, as noted in a recent evaluation of remotely sensed products for drifter modeling near the Gulf Loop Current by Liu et al. (2014). This increased error is, in large part, due to the relatively low temporal and spatial resolution available for ocean and wind velocity fields that drive drifter models. The focus of the present research is to harness the rich dynamics available in Lagrangian drifter data to enable better predictions of drifter velocities and trajectories from low-resolution Eulerian measurements of geostrophic ocean currents and wind reanalysis.

Our research capitalizes on theoretical work by Haller and Sapsis (2008), who developed reduced-order approximations of the computationally expensive Maxey–Riley (MR) equation for asymptotic inertial particle dynamics in unsteady flows. They identified an attracting slow manifold in the equation and gave sufficient conditions for the manifold to attract other trajectories. This theoretical advance then allowed Beron-Vera and LaCasce (2015) and Beron-Vera et al. (2016, 2019b) to subsequently develop deterministic reduced-order models of ocean drifters that capture inertial aspects witnessed in cyclonic and anticyclonic eddies. In a numerical setting, Wan et al. (2018) and Wan and Sapsis (2018) also utilized this slow-manifold reduction to develop a leading-order particle motion model by introducing a blended model approach with a recurrent neural network (RNN) to learn the higher-order terms in the slow dynamics. This blended machine-learning and reduced-order model proved very successful in representing bubble motion for both laminar and turbulent flows while only using a limited laminar dataset for neural-network training (Wan and Sapsis 2018).

Here we use these recent modeling developments by first constructing a reduced-order model of Lagrangian ocean drifter motion similar to that of Beron-Vera et al. (2016), but also including a priori unknown functions for wind drag, wave rectification, and near-surface shear forces. We then train a long short-term memory (LSTM) RNN (Hochreiter and Schmidhuber 1997) to learn the unknown functions and missing higher-order motion terms. Our slow-manifold approach provides us with an improved Lagrangian drifter model, and also determines the most relevant parameters and variables to use for learning submesoscale dynamics. This development provides a new avenue for representing and understanding the role of submesoscale motions in Lagrangian ocean dynamics and is fundamentally trainable for any basin worldwide. It is worth mentioning that the machine-learning corrections developed may depend on the particular drifter geometry. Different drifters may not necessarily be characterized by the same data-driven model corrections just as different shapes will follow different trajectories in turbulent flow. As is shown below, the model is robust to different flow features, seasons, and weather systems, and, as with other machine-learning applications, the similarity of trained and modeled features should be considered prior to widespread use.

## 2. Methods

Our main objective is to improve upon publicly available geostrophic ocean current estimates for the Gulf of Mexico. Our approach fundamentally differs from data-assimilation methods by the inclusion of drifter history and special consideration of blended machine-learning and deterministic models. We start by adding an Ekman current component induced by wind shear at the sea surface to geostrophic velocities obtained from satellite altimetry measurements. We then use this flow field in a reduced-order MR equation for baseline Lagrangian drifter-velocity predictions. An LSTM RNN (Hochreiter and Schmidhuber 1997) is then trained to minimize the modeled and real velocity residual along the real drifter trajectories. We evaluated the ability of our blended MR–LSTM model to predict drifter velocities given low-resolution input, as well as

the ability to trace major flow features in a chaotic flow regime. Details of each of these steps can be found in the following subsections.

### a. Data

#### 1) FLUID FLOW FIELDS

The geostrophic velocity component $\mathbf{u}_g$ used in our ocean current analysis is the freely available Archiving, Validation and Interpretation of Satellite Ocean Data (AVISO) L4 gridded velocity field hosted by the Copernicus Marine Environment Monitoring Service. The geostrophic current is the result of balancing pressure, as measured through sea surface height by satellite altimetry, and the Coriolis effect of a spinning Earth. The equations of motion of the geostrophic flow are

$$\frac{1}{\rho}\frac{\partial p}{\partial x} = fu_{g,2}, \quad \frac{1}{\rho}\frac{\partial p}{\partial y} = -fu_{g,1}, \quad \text{and} \quad \frac{1}{\rho}\frac{\partial p}{\partial z} = -g, \quad (1)$$

where $\rho$ is the density of water, $p$ is the pressure, $g$ is the constant of gravity, $f$ is the Coriolis parameter, and $\mathbf{u}_g = (u_{g,1}, u_{g,2})$.

We modify this flow field with an Ekman wind correction derived from ERA-5 reanalysis of 10-m wind fields (Copernicus Climate Change Service 2017). Several recent ocean drifter modeling studies have shown that the addition of windage influence in ocean current estimation can greatly improve modeled drifter velocities (Liu et al. 2014; Beron-Vera et al. 2016, 2019b). The ERA-5 data are freely available from the European Centre for Medium-Range Weather Forecasting (ECMWF) and are currently hosted by the Copernicus Programme Climate Change Service. We estimate the wind shear contribution to ocean current using the Ekman model of wind influence, such that our resultant fluid velocity is $\mathbf{u} = \mathbf{u}_g + \mathbf{u}_e$, where

$$\mathbf{u}_e = \left[ U_{10}\frac{0.0127}{\sqrt{\sin(\phi)}}\cos\left(\theta - \frac{\pi}{4}\right), \right.$$
$$\left. U_{10}\frac{0.0127}{\sqrt{\sin(\phi)}}\sin\left(\theta - \frac{\pi}{4}\right) \right], \quad (2)$$

$U_{10}$ is the magnitude of the 10-m wind, $\theta$ is the wind direction in radians, and $\phi$ is the latitude (Stewart 2008). More detailed estimations of wind drag are possible at this stage in the model development. For example, we know the angle of Ekman influence actually varies between 40° and 70° at this latitude, but the seasonal and spatial variability of this angle is also most extreme in regions like the Gulf of Mexico (Rio and Hernandez 2003; Zavala-Hidalgo et al. 2014). Instead, we rely on

machine learning to account for the variability around the classic 45° Ekman theory estimate (Ekman 1905). This approach to including wind stress influences is comparable to the Geostrophic and Ekman Current Observatory (GEKCO) velocity products (Sudre et al. 2013). For our blended model, we used a spline interpolant to upscale the geostrophic velocity from the original 24-h time step and ¼° resolution to the 15-min resolution of the drifter experiments. In addition, the ERA-5 reanalysis was originally collected at a 6-h time step and ¼° resolution, and interpolated to the drifter resolution.

#### 2) OCEAN DRIFTER DATA

The Lagrangian ocean drifter data used for training and validation in the current study were obtained from deployments during the Lagrangian Submesoscale Experiment (LASER; Novelli et al. 2017; D'Asaro et al. 2018; Haza et al. 2018) in January–February of 2016 and the Grand Lagrangian Deployment (GLAD) in July of 2012 (Olascoaga et al. 2013; Poje et al. 2014; Jacobs et al. 2014; Coelho et al. 2015). Position and velocity data during these extensive field campaigns in the Gulf of Mexico were measured on Consortium for Advanced Research on Transport of Hydrocarbon in the Environment (CARTHE; Novelli et al. 2017) and Coastal Dynamics Experiment (CODE; Davis 1985) drifters and were interpolated to 15-min intervals. The GLAD experiment was conducted with CODE drifters near the Deepwater Horizon spill site in the Gulf of Mexico during summer, with the follow-up LASER experiment, designed to identify submesoscale motions in the same and different regions of the Gulf during winter with biodegradable CARTHE drifters. These experiments provide an unprecedented density of Lagrangian drifter trajectories near very complex flow features that are not clearly resolved in satellite altimetry or wind reanalysis data. The LASER and GLAD experiment data are a perfect test for assessing the feasibility of machine learning to adapt a reduced-order, low-resolution drifter model to complex flow features that are only resolved at high spatial and temporal resolution.

During LASER and GLAD, more than 1100 drifters were deployed in the Gulf of Mexico, with the GLAD deployments being replicated by the first LASER deployment, and subsequent LASER deployments located near various structures in the Gulf. Many drifters were released on the continental shelf, close to the outlet of the Mississippi river in a region of strong salinity and temperature gradients at the freshwater–saltwater interface. Other drifters were deployed in deeper waters, thus providing drifter trajectories that measure submesoscale motions of varying physical origin. As noted

previously, the spatial and seasonal variability of wind influence in the Gulf of Mexico is dramatic (Zavala-Hidalgo et al. 2014), and, as part of our study, we sought to identify the shortcomings of a single data-driven model applied to both GLAD and LASER weather conditions. This helps to identify the current limitations of data-driven approaches given different physical inputs.

Throughout the LASER experiment, several large El Niño winter storms occurred with significant surface wave generation, resulting in considerable drifter motion and sporadic drogue loss (Haza et al. 2018). As this loss is noted in the published dataset, we focused our training and modeling on the drogued drifters. The summertime GLAD deployment was affected by much stronger onshore–offshore winds than LASER, and the velocity data contained much stronger evidence of windage and inertial oscillations (Beron-Vera and LaCasce 2015).

### b. Data-driven model error reduction

We utilize the drifter data to improve the performance of an existing model (to be described in the next subsection). Specifically, we complement the imperfect model, a slow-manifold reduction of the MR equation, with a machine-learning component (Wan and Sapsis 2018). In Wan and Sapsis (2018) the authors represented the true dynamics as a superposition of an imperfect dynamical system and an unknown function of both the current system state and its history (expressed through discrete time delays). A motivation for the inclusion of time delays is the observation (through Takens's embedding theorem; Takens 1981) that the inclusion of memory into the model essentially increases the degrees of freedom allowing for a more complex family of models than those obtained with the Markov property.

For the unknown function, we adopt a data-driven RNN representation with an LSTM structure (Hochreiter and Schmidhuber 1997). RNN is a deep neural-network architecture that can be viewed as a nonlinear dynamical system mapping from sequences to sequences. The most distinct feature of an RNN is the presence of hidden states whose values are dependent on those of the previous temporal steps and the current input. These units represent the internal memory of the model and are designed to address the strong temporal dependence of the state variables. LSTM takes advantage of the sequential nature of the time-delayed reduced space coordinates by processing the input in chronological order and keeping memory of the useful state information that complements the imperfect dynamics at each time step.

### c. Blended reduced-order drifter model

Starting from the MR equation of motion for a small, perfectly spherical particle in an unsteady flow (Maxey and Riley 1983), we present a modified version for the motion of our Lagrangian drifter in an ocean current in the form:

$$
\begin{aligned}
\rho_p \dot{\mathbf{v}} = &\ \underbrace{\rho_f \frac{D\mathbf{u}}{Dt}}_{(3a)} \\
&+ \underbrace{(\rho_p - \rho_f)\mathbf{g}}_{(3b)} \\
&- \underbrace{\frac{9\nu_f \rho_f}{2a^2}\left(\mathbf{v} - \mathbf{u} - \frac{a^2}{6}\Delta\mathbf{u}\right)}_{(3c)} \\
&- \underbrace{\frac{\rho_f}{2}\left[\dot{\mathbf{v}} - \frac{D}{Dt}\left(\mathbf{u} + \frac{a^2}{10}\Delta\mathbf{u}\right)\right]}_{(3d)} \\
&- \underbrace{\frac{9\nu_f \rho_f}{2a}\sqrt{\frac{\nu_f}{\pi}}\int_0^t \frac{1}{\sqrt{t-s}}\left[\dot{\mathbf{v}}(s) - \frac{d}{ds}\left(\mathbf{u} + \frac{a^2}{6}\delta\mathbf{u}\right)\right]\,ds}_{(3e)} \\
&- \underbrace{\nu_a \rho_a \alpha(\mathbf{v} - \mathbf{u}_{\text{wind}})}_{(3f)} - \underbrace{f(\rho_p \mathbf{v} - \rho_f \mathbf{u})^\perp}_{(3g)} + \underbrace{F(\mathbf{u}_e, \mathbf{u})}_{(3i)},
\end{aligned}
\tag{3}
$$

where $\mathbf{u}$ and $\mathbf{u}_e$ are the ocean and Ekman velocities defined by Eq. (2); $\rho_f$, $\rho_p$, and $\rho_a$ are representative densities of the surface ocean water, drifter, and air, respectively; $a$ is the radius of the drifter; $\nu_f$ and $\nu_a$ are the kinematic viscosities of the ocean and air; and $t$ and $s$ are time variables. The first five terms on the right-hand side of Eq. (3) are standard in MR frameworks and represent the force of the current on the drifter (3a), the buoyancy effect (3b), the Stokes drag on the drogue (3c), the force from the fluid moving with the particle (3d), and the Basset–Boussinesq memory term (3e). Specific to the motion of ocean drifters, we identified three additional forces that contribute significantly to a slip velocity between the drifter and ocean current. These last forces include the Stokes drag of the surface wind on the exposed drifter float (3f), the Coriolis force (3g), the drag on the tether and drogue induced by shear currents, and motions induced by surface waves, (Geyer 1989; Niiler and Paduan 1995; Beron-Vera and LaCasce 2015). Because of the complexity of the processes, we temporarily represent the combined shear current and surface wave effects in one unknown function of current velocity and wind speed (3i). Following the slow-manifold approach described in the appendix, we reduce this set of equations to the leading-order approximation

$$\dot{\mathbf{x}} = \mathbf{u}(\mathbf{x}, t) + \varepsilon \left\{ \left( 1 - \frac{3R}{2} \right) \left[ \frac{D\mathbf{u}(\mathbf{x}, t)}{Dt} - \mathbf{g} \right] \right.$$
$$\left. - Rf(\delta_p - 1)\mathbf{u}^{\perp} + H[\mathbf{u}_e(\mathbf{x}, t), \mathbf{u}] \right\}, \qquad (4)$$

where $H$ is assumed to be a bounded function of the Ekman velocity and the current velocity, representing wind-wave and shear drag contributions; there is no known a priori differential operation on $\mathbf{u}$, or $\mathbf{u}_e$ in $H$. For example, similar slip velocity estimates have been found to scale as linear functions of the ocean surface velocity (Geyer 1989; Edwards et al. 2006). We avoid the assumptions necessary for such a linear slip-velocity estimate and relegate the learning of potentially more complicated wind-influence and shear drag terms to the data-driven component of our model. As detailed in the appendix, the missing higher-order terms are shown to be functions of $\mathbf{u}$, $\mathbf{u}_e$, and $D\mathbf{u}/Dt$.

From this reduced-order model, we follow an approach similar to that of Wan and Sapsis (2018) by employing an LSTM to learn the contribution from the missing higher-order terms and the wind-wave function $H$. Through the model order reduction detailed in the appendix, we can identify which variables the higher-order terms are functions of. These variables inform with which time series to meaningfully train our RNN. We write out the 2D blended drifter model in the following form, neglecting vertical motion:

$$\dot{\mathbf{x}} = \mathbf{u}(\mathbf{x}, t) + \varepsilon \left[ \left( \frac{3R}{2} - 1 \right) \frac{D\mathbf{u}}{Dt} - Rf(\delta_p - 1)\mathbf{u}^{\perp} \right]$$
$$+ \mathbf{G}[\boldsymbol{\xi}(t), \boldsymbol{\xi}(t - \tau), \boldsymbol{\xi}(t - 2\tau), \dots]$$
$$\boldsymbol{\xi}(t) = \left[ \mathbf{u}(\mathbf{x}, t), \mathbf{u}_e(\mathbf{x}, t), \frac{D\mathbf{u}}{Dt} \right]. \qquad (5)$$

Here $\mathbf{G}$ is an LSTM neural network and $\boldsymbol{\xi}$ is a vector of time-delayed values of $\mathbf{u}$, $\mathbf{u}_e$, and the material derivative $D\mathbf{u}/Dt$. Using our training set of real drifter trajectories, $\mathbf{G}$ is designed to learn both the error in our drifter approximation in Eq. (4) and the unknown function $H$, as a function of $\boldsymbol{\xi}$ over the full drifter history. We use a half-mean-squared-error loss function for our training.

The results presented here come from using a simple LSTM architecture similar to that used by Wan and Sapsis (2018). This architecture consists of one LSTM layer with 200 hidden units, and one fully connected layer trained for 100 epochs, and a time delay of $\tau = 15$ min. This choice of training parameters has not been fully optimized, although we found it to result in sufficiently accurate predictions without overfitting when compared with training between 50 and 400 epochs and using LSTM layers of 50–400 units.

Note that a sensitivity analysis was conducted by removing individual variables from $\boldsymbol{\xi}$, and removing $D\mathbf{u}/Dt$ and the Coriolis term from the baseline model in Eq. (5). While these results have not been presented for brevity, the blended model performed best in the form of Eq. (5). The largest modeling errors were caused by removing $\mathbf{u}$ from $\boldsymbol{\xi}$, followed by removing $\mathbf{u}_e$. The blended model was actually able to appropriately correct for modifications in the deterministic model, but this provided a poorer performing baseline comparison for the Lagrangian metrics discussed below.

## 3. Results

We tested the blended model approach with a variety of training datasets from the LASER and GLAD experiments. The two experiments were similar, differing mainly in season (winter and summer), the number of drifters and a subset of the seeding locations. Though differing in geometry, the CARTHE and CODE drifter designs used in the LASER and GLAD experiments, respectively, have been found to have nearly identical trajectories in both coastal and deep ocean currents (Novelli et al. 2017). We trained and tested blended models on each experiment separately, as well as training a blended model with a combined set of drifters spanning both seasons and experiments. As mentioned previously, the combined drifter set includes trajectories influenced by varying weather systems with different magnitudes of inertial oscillation, thus providing much more complex physics for the neural network to interpret.

Many of the drifters in both experiments were deployed in small clusters. By randomly selecting drifters, we created a rich training set that included many trajectories similar to those found in the test set. Because machine-learning approaches are most accurate when interpolating between features that the networks have already seen, using randomized training test sets provided a relatively simple task when testing the blended model. A cartoon example of random training-test drifter designation is shown in Fig. 1a. However, we also wanted to test the extrapolative ability of the blended model for a scenario with limited similarity between training and test sets. We achieved this by handpicking the drifters, making sure drifters from the same tight deployment clusters were not separated into both the training and test sets. This approach is demonstrated in contrast in Fig. 1b. Hereinafter we will refer to these two methods of separating training and test drifters as random and clustered.

As with all machine learning, potentially overfitting the model to the training data is an important consideration. With this in mind, we required our training set
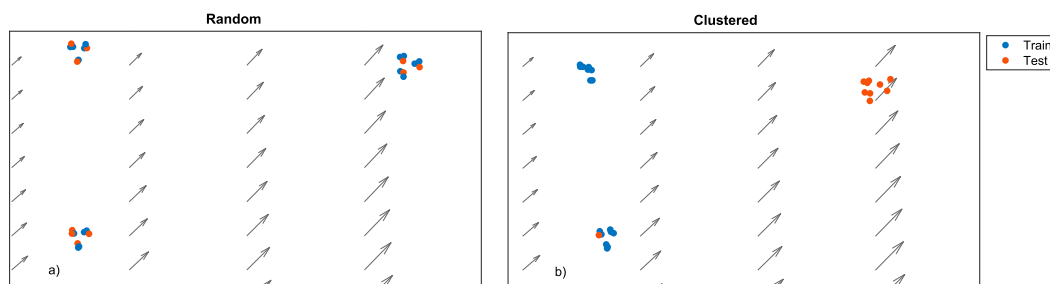
FIG. 1. Two types of training and test drifter selection used in this study: (a) Random selection maximizes training and test drifter trajectory similarity, whereas (b) clustered training provides a more difficult test for blended-model performance by limiting training and test set similarities. Drifter initial positions are overlain on ocean current vectors.

to consist of 80% of our drifter data for both the random and clustered training. This forced an imperfect separation of drifters for the clustered method, but, as is discussed below, still generated a more complex testing scenario. A comparison of the modeling capabilities of each method is detailed in the context of statistical measures of similarity in sections 3b and 4.

The LASER campaign involved a much larger deployment of drifters than GLAD and hence serves as the primary source of discussion for the specific abilities of the blended model. We also briefly present the results from the summer GLAD model to complement the conclusions from the LASER data. By training and testing on a combined LASER and GLAD dataset, we show it may also be possible to train blended models for a progressively wider range of meso- and submesoscale features, but additional modeling considerations are likely necessary.

### a. Interpolative LASER testing

To test the ability of our blended model to recognize and predict motion around similar submesoscale features, we initially trained the data-driven component **G** in Eq. (5) with 668 randomly selected Lagrangian drifter trajectories from the winter (LASER) dataset with time series length ranging from 16 to 60 days. The remaining 167 drifters were reserved as a test dataset in order to evaluate the model on new data. For model comparison, we used our Ekman-modified flow field and the MR drifter model of Beron-Vera and LaCasce (2015) as a baseline. The Beron-Vera et al. model is precisely Eq. (4) without the wind-wave function $H$.

After training, we evaluated the blended model's single-step prediction ability of drifter velocity, that is, the model's ability to predict velocity at each time step along the actual drifter trajectory. As shown in Fig. 2, the randomly trained blended model showed considerable improvement in accuracy over the baseline model. Figure 2a displays histograms of the root-mean-square

error (RMSE) of the zonal velocity for the MR model and the blended model along all trajectories in the randomized LASER test set. A significant shift toward smaller errors with the blended model indicates an improvement in the model's ability to accurately predict the 15-min drifter velocity from daily $\frac{1}{4}°$ geostrophic velocity and 6-h wind reanalysis data. Example time series of the zonal component of modeled and actual drifter velocities are shown in Figs. 2b–d, with the MR model in blue, ocean current $\mathbf{u}(\mathbf{x}, t)$ in green, output of the blended model in black and the real drifter velocity in orange. While potentially important for Lagrangian prediction, the $\varepsilon$-term contribution in Eq. (5) is very small, making the blue and green plots hard to differentiate. An inset has been included to highlight the small difference. The distance to the nearest drifter in the training dataset, the time to the soonest deployment of a trained drifter, and the RMSE of the blended model are all noted on the respective subplots.

We also tested the model's multistep predictive capabilities to evaluate how accurately we can forecast the trajectory of a real drifter in an environment with rich submesoscale features. For multistep prediction, we used blended model velocity output to update drifter positions over time with a fourth-order Runge–Kutta scheme, allowing modeling errors to accumulate in both the neural-network memory and in the drifter position. For a chaotic flow such as in the ocean, it is nearly impossible for a perfect one-to-one model-drifter trajectory match. To account for uncertainty in the underlying chaotic dynamics, we introduced uncertainty in the initial conditions by seeding a grid of synthetic drifters in a small neighborhood ($<10$ km) of a real drifter position. Each drifter in the grid had the same initial velocities, but small perturbations in their positions. This ensemble approach provided an insightful evaluation as dominant flow features began to appear in the cluster trajectories.

Two test examples of this ensemble trajectory model are shown in Fig. 3. The real LASER drifter trajectory is mapped as a bold red line in each row. The left panels

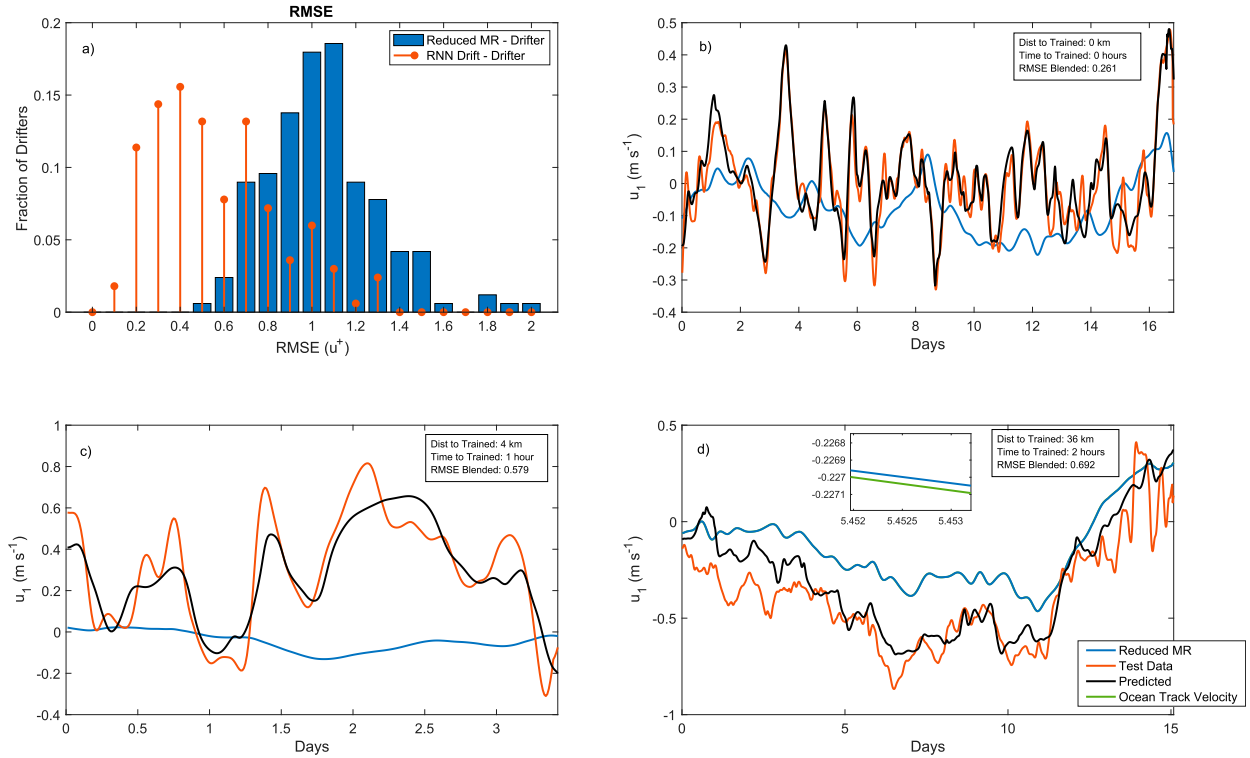FIG. 2. Single-step prediction velocity test comparisons: (a) Histograms of RMSE for nondimensionalized zonal components $[u^+ = (u_1/\overline{\mathbf{u}})]$ for the reduced MR equations and the blended model. (b)–(d) Examples of reduced-MR velocities, real drifter velocities, blended model velocities, and the underlying low-resolution ocean current velocity. A significant reduction in error can be found with the blended approach. The minor difference between the reduced-MR velocity and underlying ocean current is shown in the inset in (d). The magnitude of the data-driven component $G$ can be seen as the difference between the blue and black curves.

show the trajectories of all blended model drifters initialized in a small neighborhood in black, with their median trajectory displayed in yellow. The right panels show analogous trajectories with blue MR model output and the green median path. Over nearly 50 days of transport, the upper-left blended model drifters were able to trace out the large spiraling flow feature that was largely missed by the purely deterministic MR equations in the upper right. In the bottom plots, the real drifter followed a flow path more closely mimicked by the MR model. Upon further investigation, we found that the several major flow pathways drawn by the blended model correspond closely with trajectories of different drifter clusters (not pictured). This means that while small perturbations in the blended model initial conditions can reveal multiple flow features, only a subset of these features were actually revealed by the reduced-MR equations.

We quantified the ability of the ensemble models to approximate real trajectories with the normalized cumulative separation distance, or skill score, $s$ from Liu and Weisberg (2011). For a drifter time series of length $n$, this metric normalizes the cumulative separation distance between the modeled drifter position $\mathbf{y}(t_j)$ and the actual drifter position $\mathbf{x}_D(t_j)$ by the cumulative length of the real drifter trajectory, $l(t_j)$, in the following form:

$$c = \frac{\sum_{j=1}^{n} \|\mathbf{y}(t_j) - \mathbf{x}_D(t_j)\|}{\sum_{j=1}^{n} l(t_j)}, \quad s = 1 - c.$$

A skill score $s$ of less than 0 indicates that the modeled and real drifter paths are deviating faster than the real drifter is actually moving. As shown in Fig. 4a, the mean skill score for an ensemble of blended drifters was better than the reduced-order model 67% of the time. This improved trajectory modeling is also apparent in the skill score of the median ensemble trajectory (yellow and green paths in Fig. 3) in Fig. 4b, where the blended drifters outperformed the reduced-order model 68% of the time. Of particular note are the different ranges of skill score values for the MR and blended model outputs. Specifically, there are large negative values present
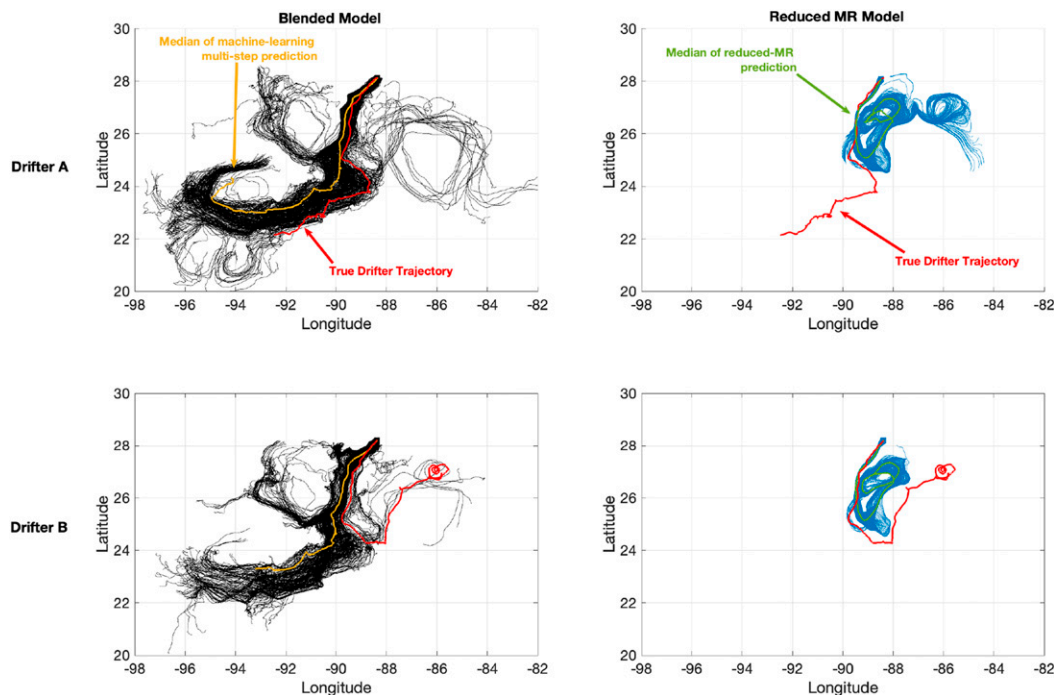
FIG. 3. Multistep ensemble trajectory prediction comparisons: (left) Ensembles of two blended model drifters in black, and their median trajectories in yellow. (right) Reduced-MR drifter ensembles and their median trajectory in green. On all plots, the true drifter trajectory is indicated by the thick red line. Note that, for both trajectories, some fraction of the blended model ensemble was able to make the correct prediction, whereas the reduced-MR trajectories were largely confined to one flow feature.

in the MR model evaluation that are nonexistent for the blended model.

We investigated these negative skill scores and found that 75% of them correspond to drifters deployed on 7 February 2016. In Figs. 5a and 5b, we show the 7 February location of drifters near this deployment, with drifters colored by their skill scores for the blended and MR models, respectively. These colored coordinates are overlaid on the daily sea surface temperature from *Aqua* MODIS thermal infrared measurements on 7 February. The MR drifters with the worst MR skill scores are clustered in the strong mixing regions between the cold coastal and warm offshore waters in Fig. 5b. From this position, the warm waters extend south and east throughout the rest of the Gulf of Mexico and into the Loop Current. For the blended model, there was no discernible difference in model performance across this interface, or elsewhere in the domain.

In Figs. 5c and 5d, we show the median path drawn by 37 blended and MR drifters residing on the freshwater–saltwater interface in Figs. 5a and 5b. The MR ensemble exits the cool waters and quickly becomes entrained by the mesoscale geostrophic motions away from the coast. In contrast, the real drifters and blended model drifters stay near the mixing interface, in the cooler waters in the northern part of the domain. This submesoscale model improvement was precisely the result of training the blended model on drifter velocity and acceleration data from similar mixing regions. The LSTM component was able to identify a signature of this mixing in $\xi$ and prevent entrainment into the larger currents, forcing drifters to remain in the high temperature gradient mixing region on the shelf.

RMSE histograms of the random training of GLAD and LASER+GLAD datasets are shown in the following section. Randomly selecting training drifters always resulted in model improvements over the MR equations. However, we analyze the robustness of these data-driven improvements in the following section by looking at minimally repetitive training and test sets.

### b. Extrapolative drifter testing

We compared the blended model's ability to correct velocities and trajectories for clustered datasets to the randomized training method for LASER, GLAD, and combined GLAD+LASER datasets. In hopes to understand the different model performance with each training set, we generated several metrics comparing
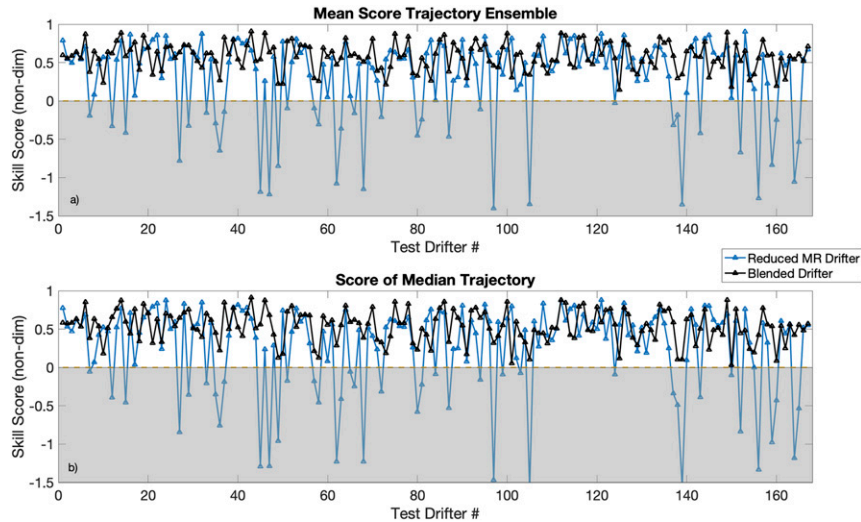
FIG. 4. Skill scores of the randomly trained blended model and MR model on the LASER dataset: (a) The mean skill score of all drifters in an ensemble prediction for each LASER test drifter. (b) Skill scores for the median trajectory from an ensemble prediction for LASER drifters. The gray-shaded area indicates the region where model and real drifter trajectory separation outpaces the actual drifter motion.

trajectories in test and training sets. To quantify how similar the training and test datasets were in the time domain, we calculated the maximum normalized correlation for zonal velocity of each test drifter against all drifters in the training set. To complement these values in the frequency domain, we also calculated the maximum mean magnitude spectral coherence between each test drifter and the training datasets. Last, we calculated distance of the initial position for each drifter to the nearest training drifter, and the time to the closest training drifter. Mean values of all these metrics from the randomized data separation are displayed in Table 1 for the LASER and GLAD data, with values inside parentheses indicating the clustered separation scheme.

As can be seen in Table 1, maximum correlation and magnitude squared coherence change very little for the LASER data between the random and clustered training schemes, but there is a significant increase in both the mean distance and time to nearest trained drifter. For the GLAD dataset, there was a decrease in correlation and MS coherence as well as an increase in distance and time to trained drifters when using the clustered separation. To highlight the details of these influences, a drifter by drifter comparison of the single-step zonal velocity model RMSE values for the random and clustered LASER sets are compared to the metrics from Table 1 (Fig. 6). There appears to be little relationship between similar test and training power spectra and low RMSE values, though a slight decrease in RMSE can be

noted with increasing correlation, especially for the random training set. For both the random and clustered models, increasing distance between test and training drifters resulted in higher RMSE values, but there is notable scatter in RMSE for short distances and time lapses. The red dot indicates a drifter present in both the random and clustered test sets that is further analyzed in section 4.

The overall influence of the decreased similarity between trained and tested trajectories can be found in the histograms of Fig. 7. For GLAD and LASER separately, the blended model was able to obtain better single-step velocity prediction when there was more similarity in trained and tested flow features and the training and test drifters were geographically and temporally proximal. Even with the handicap of selectively limiting the kinds of features previously seen by the data-driven component, the blended model was still able to outperform the purely deterministic approach. Analysis of one such trajectory is further discussed in section 4.

The last dataset that the blended model was tested against was a combination of both LASER and GLAD drifters, referred to as LASER+GLAD in Figs. 7e and 7f. This increasingly complex collection of drifters spanned both winter and summer conditions in the Gulf with different flow features and disparate evidence of inertial oscillations. Again, training with a random selection of LASER and GLAD drifters allowed the model to outperform the single-step predictive
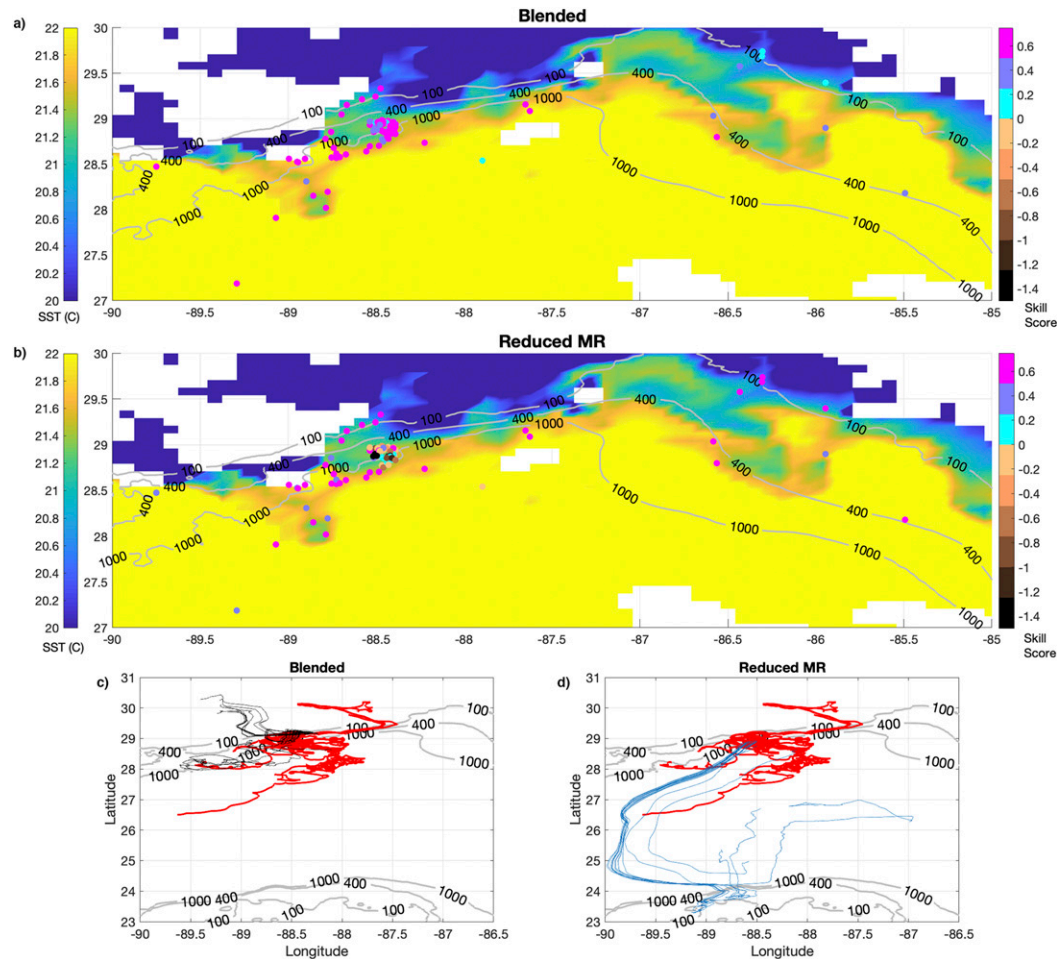
FIG. 5. Positions and skills scores for the (a) blended and (b) reduced-MR models on 7 Feb 2016 overlaid on daily *Aqua* MODIS Thermal IR sea surface temperature measurements and bathymetry contours. White areas in the top left correspond to the outlet of the Mississippi river. Also shown for the (c) blended and (d) reduced-MR models are the median trajectories of the drifters with negative skill scores. Reduced-MR drifters quickly leave the coastal zone, whereas the drifters in the blended model wander for their entire lifetime. The time elapsed for trajectories in (c) and (d) is approximately 23 days.

capability of clustered test method, and both blended model approaches outperformed the purely deterministic model. There still remain large RMSE values for the combined blended models, indicating an inability to always make the proper corrections, or recognize submesoscale features from a wider library of options.

For all three datasets, training with a randomly selected training set resulted in the lowest RMSE for single-step predictions. As the LSTM was trained with single-step predictions, and because of the greater similarity in training and test drifter trajectories for random training, this improvement supports the notion that a blended model can increase model accuracy around flow features similar to those seen previously. Because of the currently limited amount of high-resolution training data

available globally, a more realistic forecasting situation was simulated by multistep predictions for clustered training on the LASER training set. Further supporting the blended model approach, we found comparable skill

TABLE 1. Mean values of velocity statistics (correlation and mean MS coherence) that compare test and training datasets for LASER and GLAD drifter trajectories. The first values are for randomly selected training sets, and the values in parentheses are for ordered, clustered datasets.

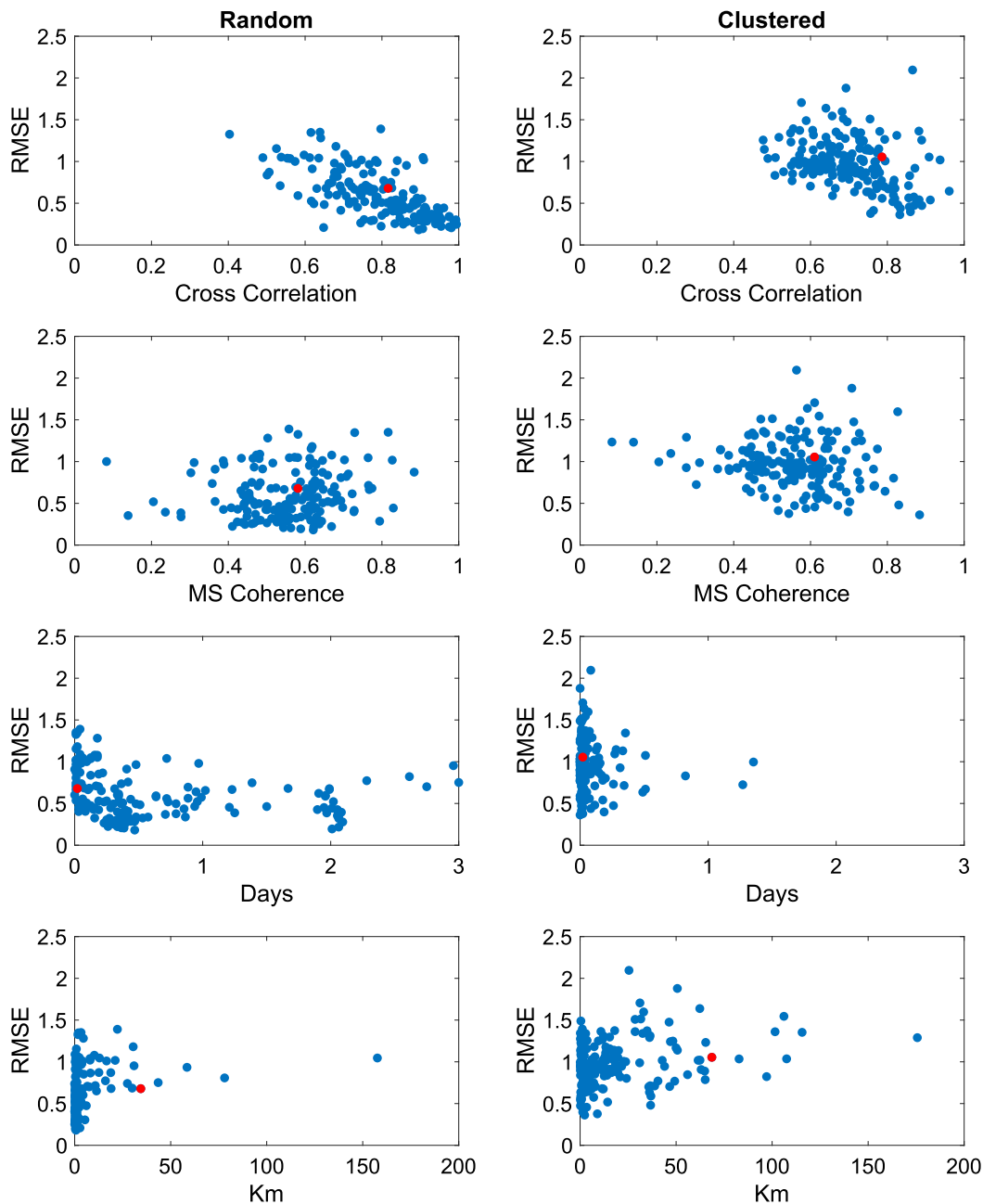| | LASER (winter) | GLAD (summer) |
|---|---|---|
| Max correlation | 0.8 (0.7) | 0.8 (0.74) |
| Max mean MS coherence | 0.56 (0.56) | 0.8 (0.62) |
| Distance to trained drifter (km) | 5.2 (19.1) | 4.6 (13.7) |
| Time to trained drifter (h) | 0.95 (18.72) | 0.25 (79) |

FIG. 6. Cross correlation, mean MS coherence, time, and distance between test drifters' deployment and soonest or closest train drifter deployment vs RMSE values for zonal velocity of blended drifter modeling from (left) random and (right) clustered training methods. There is the clearest decrease in RMSE with increased correlation for the random scheme and a general increase in RMSE with increasing distance between training and test drifters. All other relationships remain less defined. The same test drifter marked by a red dot in all plots is further explored in section 4.

score improvements for both clustered and random training, when test and training drifters were often from the same deployment cluster. Figure 8 shows analogous skill score data as Fig. 4 for the clustered training. Again, many of the lowest skill scores from the reduced-MR

predictions were avoided with the blended model, and the blended model outperformed for 63.4% of the test set. This is a slight decrease from 67% to 68% from the random training and the range of skill scores was larger for the clustered case, actually extending into negative
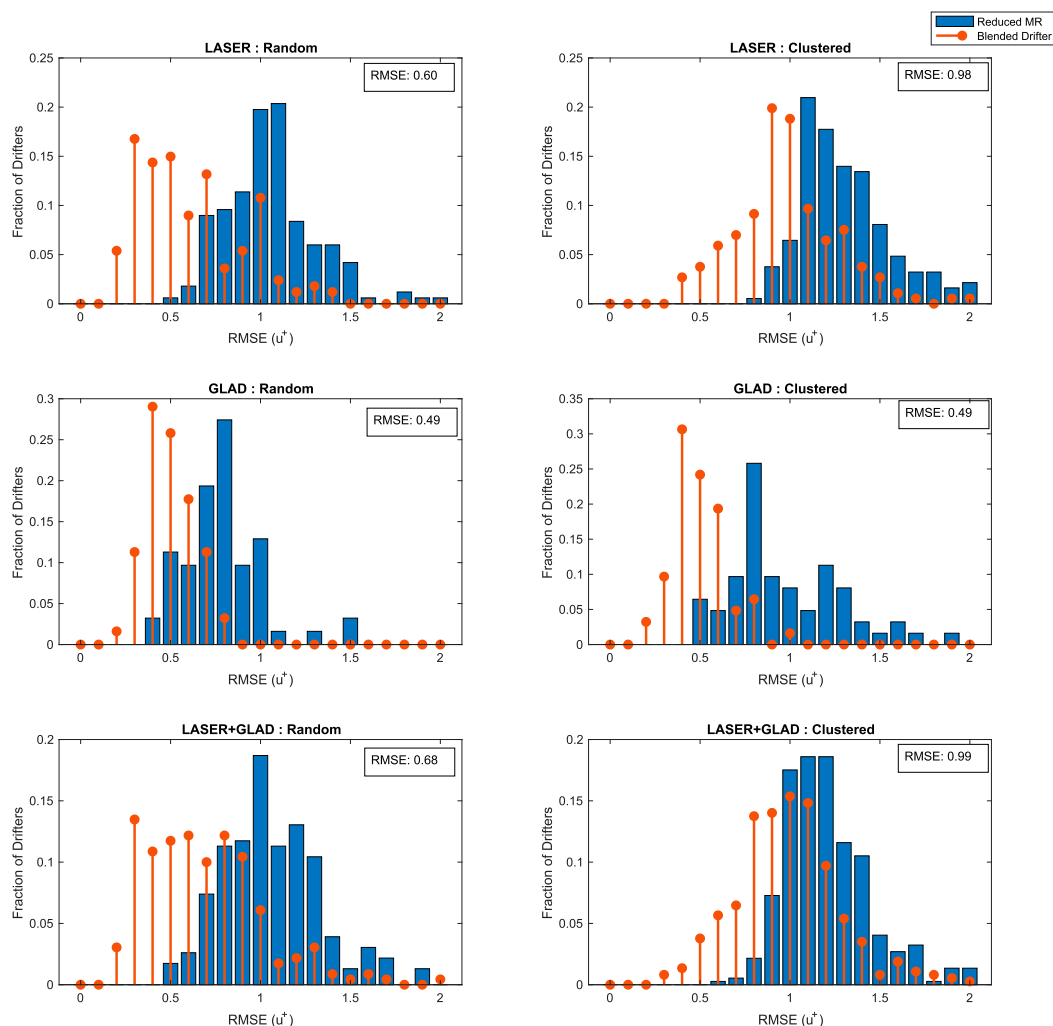
FIG. 7. RMSE histograms of nondimensionalized zonal velocity for surface drifters during the (top) LASER (winter), (middle) GLAD (summer), and (bottom) LASER+GLAD datasets for (left) random and clustered trajectory training. Blue bars indicate the performance of the deterministic reduced MR equations, and orange stem plots represent the performance of the blended model.

values. Because we are testing with a different drifter set, we also do not have the strong negative skill scores from the reduced-MR drifters as in Fig. 4.

## 4. Discussion

A blended deterministic and neural-network model for predicting Lagrangian ocean drifter trajectories substantially outperformed a deterministic reduced-order MR model for two experimental datasets in the Gulf of Mexico. We approximated drifter velocities with significantly less error (Fig. 7) along drifter paths, indicating a better representation of the underlying flow field is available if given the appropriate initial conditions. The blended model also had increased multistep predictive capabilities, outperforming the

ability of the reduced-MR equations to correctly predict drifter locations at time scales up to two months. Skill scores were typically higher for both repetitive and clustered training, with improved behavior linked to correct identification of dominant flow features in several cases.

Specifically, the blended model was also able to accurately trace several large vortical features in Fig. 3 that were missed by the MR equation. Though these features were mesoscale in size, inaccuracies in the baseline model combined with low-resolution flow data prevented the MR model from resolving many features for several drifters, even with perturbations in initial conditions. For oil spill, search and rescue, and ocean ecosystem research purposes, it is precisely these transport features that are of utmost importance to model and resolve.
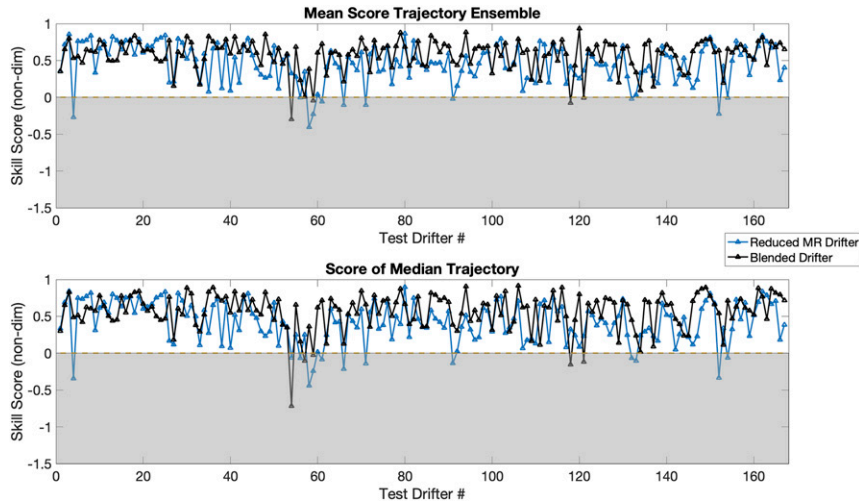
FIG. 8. Skill scores calculated for the clustered winter (LASER) test and training dataset, similar to Fig. 4.

Model improvement was also evident in the shallow waters near the outlet of the Mississippi River where a large number of the LASER drifters remained for several weeks, but currents in the geostrophic data brought the MR drifters off the continental shelf. In this region, there is complex fresh–saltwater mixing as previously highlighted by Gonçalves et al. (2019), and strongly influential submesoscale signatures not immediately evident in the geostrophic currents or 10-m wind fields. The collocation of poorly performing drifters and the mixing interface was verified in satellite sea surface temperature measurements during the LASER campaign. Near this transport barrier, the MR model obtained negative skill scores whereas the randomly trained blended model performed just as well as in other areas. As the two models only differed by the inclusion of the LSTM component, this improvement can be attributed to the ability of the neural network to mimic a real drifter velocity from a similar signature in $\boldsymbol{\xi}(t)$ that was seen during training. It is also worth reiterating that $\boldsymbol{\xi}(t)$ contains no position information, and thus the model was able to make changes based solely on the time history of the physical motion of the drifter and the surrounding fluid flow.

Testing the blended model with more difficult tasks, such as using distinctly dissimilar training and test trajectories or testing over a broader range of dynamics (combined summer and winter) started to highlight the current limitations of a blended-model approach. While the blended-model always outperformed the baseline deterministic model upon which it was improving (Fig. 7), the improvements were less significant than with random training. An incremental model improvement

from baseline MR equations, to clustered training, to random training is examined in Fig. 9. Single-step predictions for the drifter represented by the red dot in Fig. 6 are examined for the three models in Fig. 9 (top). Clearly, the modeled velocity approaches the drifter velocity as the data-driven component is introduced and improves with more similar training data.

This single-step improvement is complemented by trajectories prediction improved in Fig. 9 (bottom row). Because the MR model underpredicts the drifter speed, the ensemble trajectories quickly diverge from the real drifter path, resulting in a poor skill score (inset). The clustered model performs with enhanced physics, resulting in an improved skill score, even though the nearest training drifter was deployed nearly 70 km and 4 h away. The random training provided a drifter only 34 km away from the test, and the arcing features of the real drifter were most closely mimicked. An increased proximity to training drifters likely aided in this model performance, but there are other strong influences as can be seen in the lack of clear trends in Fig. 6.

There is currently no metric for determining the domain of applicability for this kind of machine learning with a given training dataset. As seen in Figs. 6 and 9, neither correlation nor spectral coherence between a tested drifter and the training set can fully account for the differing performance between the random and clustered training methods. Higher correlation and spectral coherence values typically result in lower RMSE values for single-step prediction, but low RMSEs are also possible with poor correlation or coherence. While using random training to obtain similar deployment coordinates (space and time) for test and training drifter results
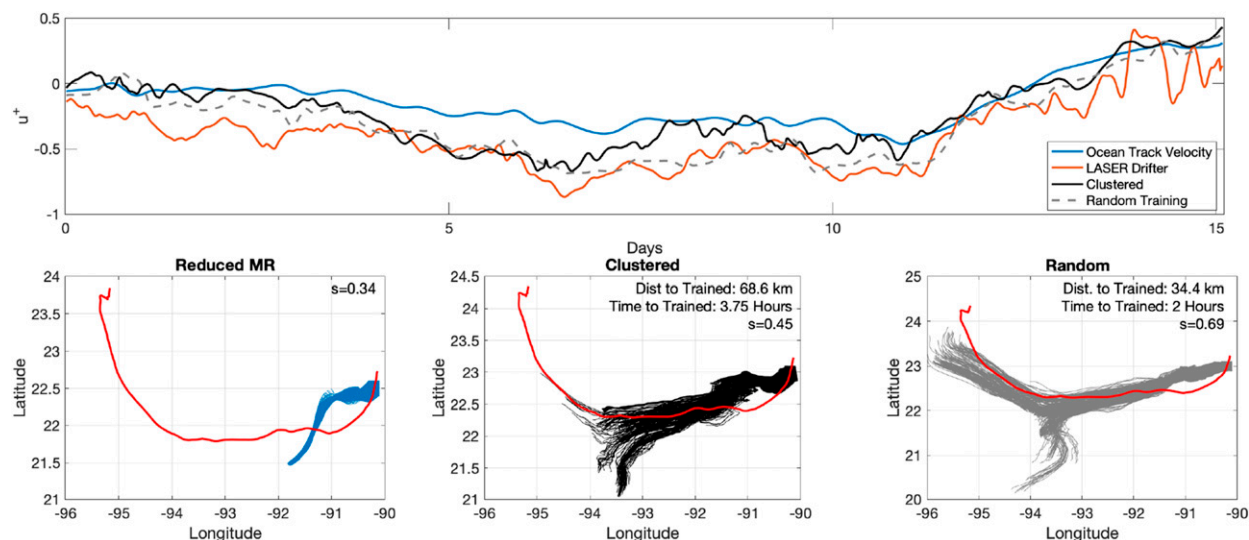
FIG. 9. Trajectory and velocity time series predictions for the clustered and random training sets, and the reduced-MR model, as well as the true drifter trajectory and zonal velocity.

in better blended model performance, this does little to help us understand why drifters seeded away from the training set can also perform well, or to predict if a blended Lagrangian model will perform well in a region without drifters nearby. Further investigation of a definition of Lagrangian similarity for this kind of forecasting and modeling would be beneficial for constraining models to appropriate domains of application, but is currently beyond the scope of this initial investigation as there is no guarantee of stationarity with this data.

## 5. Conclusions

Because of the lack of availability of high spatial- and temporal-resolution flow data that can resolve submesoscale motions, the use of machine learning for improving Lagrangian drifter models shows great promise. The blended model approach pursued here prevents user-defined parameterization and tuning with respect to one specific drifter geometry, drag, buoyancy corrections, and other factors that can complicate physical models of drifter trajectories. For drifters of the same type in similar environments and similar seasons, signatures in drifter velocity and acceleration time series can be identified with recurrent neural networks and used to autocorrect velocity fields, even in the presence of strong coastal temperature and salinity gradients, and inertial oscillations. With the current availability of high-resolution drifter data, it is difficult to evaluate the utility of a one-size-fits-all blended model for all seasons. Seasonally restricting the domain of training and testing,

however, provides significant improvements over deterministic drifter modeling.

Similar to other machine-learning approaches, this blended model may be further improved by training on additional datasets that include more submesoscale features. This added physical complexity will likely need to be met with increased LSTM complexity. The flexible blended approach is also adaptable to include additional physical parameters in its training (e.g., temperature, salinity) should a physical basis for their inclusion be justified for a different baseline model. As well, the blended approach can be adapted for further advances in deterministic drifter modeling, such as the recent slow-manifold developments of Beron-Vera et al. (2019b). The broad implications of machine-learning improvements could be particularly advantageous to the oceanography community as data from future field campaigns can be harnessed to better understand the motions present in a particular region, and compared to others. Because the blended approach uses highly accurate Lagrangian data to approximate the underlying physics of drifter transport, we believe there is a theoretical finite limit on additional training datasets that are necessary for accurate drifter modeling in all regions of the globe. Applications of this method to other regions are currently under way.

hourly 10-m wind fields are available from The Copernicus Programme Climate Change Service (https://climate.copernicus.eu). The AVISO geostrophic current velocity product used in this study, "Global Ocean Gridded L4 Sea Surface Heights and Derived Variables Reprocessed," is freely available and is hosted by the Copernicus Marine Environment Monitoring Service (http://marine.copernicus.eu).

# APPENDIX

## Blended Drifter Model Derivation

The MR equation defines the path $\mathbf{x}(t)$ of a small spherical particle in a time-dependent flow as follows:

$$\dot{\mathbf{x}} = \mathbf{v}(\mathbf{x}, t)$$

$$\rho_p \dot{\mathbf{v}} = \rho_f \frac{D\mathbf{u}}{Dt} + (\rho_p - \rho_f)\mathbf{g}$$

$$- \frac{9\nu\rho_f}{2a^2}\left(\mathbf{v} - \mathbf{u} - \frac{a^2}{6}\Delta\mathbf{u}\right)$$

$$- \frac{\rho_f}{2}\left[\dot{\mathbf{v}} - \frac{D}{Dt}\left(\mathbf{u} + \frac{a^2}{10}\Delta\mathbf{u}\right)\right]$$

$$- \frac{9\nu\rho_f}{2a}\sqrt{\frac{\nu}{\pi}}\int_0^t \frac{1}{\sqrt{t-s}}\left[\dot{\mathbf{v}}(s) - \frac{d}{ds}\left(\mathbf{u} + \frac{a^2}{6}\delta\mathbf{u}\right)\right]ds,$$

$$(A1)$$

where the individual force terms on the right-hand side of Eq. (A1) are the force of the fluid on the particle by the undisturbed flow, the buoyancy force, the Stokes drag, an added mass term from the fluid moving with the particle, and the Basset–Boussinesq memory term. We consider four additional influences on an ocean drifter trajectory following the suggestions of Niiler et al. (1987), Geyer (1989), Edwards et al. (2006), and Beron-Vera et al. (2016). These forces include the wind drag on floats above the sea surface, drag on the tether and drogue induced by shear currents, motions induced by surface waves, and the Coriolis force.

The wind drag on the exposed float and Coriolis effect can be considered directly, but explicit forms of the shear current drag and surface wave influences are beyond the scope of this research and remain open problems in oceanography. Near-surface shear current effects are significantly complicated by stratification in the ocean caused by temperature or salinity gradients, and the degree of turbulence in a specific ocean region (Niiler and Paduan 1995). It is assumed that the Ekman effect decreases exponentially over the Ekman layer depth $D_E$

(estimated at 10–40 m for this study) in a well-mixed upper-ocean layer. The current changes from $\mathbf{u} = \mathbf{u}_e + \mathbf{u}_g$ at the surface to $\mathbf{u} = \mathbf{u}_g$ at a depth of $D_E$ (Stewart 2008). Thus the resulting shear acting on the drogue and tether is a function of $du/dz$ which is a function of $u_e/(zD_E)$, depending on stratification and the shape of the exponential. The effect of this motion is thus included as an unknown function of Ekman current, a linear function of wind speed, that will be learned by the machine-learning corrections. For a summary of wind-induced wave effects and specific details of wave effects on drifters from this study, refer to Niiler et al. (1987) and Haza et al. (2018), respectively. The effect of surface waves will be included as an unknown function of the Ekman contribution (which itself is a linear function of wind speed) and the surface ocean velocity. By training to predict the discrepancy between our deterministic model that does not account for surface waves and true drifter we motion, the neural network will then learn to account for surface wave contributions, among other motions, as discussed below.

We then write an adapted form of Eq. (A1) for ocean drifter motion as follows [this is identical to Eq. (3) in the main text]:

$$\rho_p \dot{\mathbf{v}} = \rho_f \frac{D\mathbf{u}}{Dt}$$

$$+ (\rho_p - \rho_f)\mathbf{g}$$

$$- \frac{9\nu_f\rho_f}{2a^2}\left(\mathbf{v} - \mathbf{u} - \frac{a^2}{6}\Delta\mathbf{u}\right)$$

$$- \frac{\rho_f}{2}\left[\dot{\mathbf{v}} - \frac{D}{Dt}\left(\mathbf{u} + \frac{a^2}{10}\Delta\mathbf{u}\right)\right]$$

$$- \frac{9\nu_f\rho_f}{2a}\sqrt{\frac{\nu_f}{\pi}}\int_0^t \frac{1}{\sqrt{t-s}}\left[\dot{\mathbf{v}}(s) - \frac{d}{ds}\left(\mathbf{u} + \frac{a^2}{6}\delta\mathbf{u}\right)\right]ds$$

$$- \nu_a\rho_a\alpha(\mathbf{v} - \mathbf{u}_{\text{wind}}) - f(\rho_p\mathbf{v} - \rho_f\mathbf{u})^\perp + F(\mathbf{u}_e, \mathbf{u}),$$

where $\alpha$ is a drag coefficient depending on the unique geometry and surface area of the float not submerged in the ocean at any point in time, $f$ is the Coriolis parameter, and $F$ is an unknown function combining surface wave effects and shear influences. We can remove the Fauxen terms because $a/L \ll 1$, where $L$ is a dominant length scale of our ocean flow (Wan and Sapsis 2018). If we can assume that $a/(\nu_f)^{1/2}$ is very small, then we can remove the Basset–Boussinesq term. After rescaling space, time, and velocity by a characteristic length scale $L$, characteristic time scale $T = L/U$, and characteristic velocity $U$, we have

$$\dot{\mathbf{v}} - \frac{3R}{2}\frac{D\mathbf{u}}{Dt} = -\mu(\mathbf{v}-\mathbf{u}) + \left(1 - \frac{3R}{2}\right)\mathbf{g} - R[f(\delta_p\mathbf{v}-\mathbf{u})^{\perp}$$
$$+ \delta_a\nu_a\alpha(\mathbf{v}-\mathbf{u}_{\text{wind}}) - F(\mathbf{u}_e,\mathbf{u})],$$

(A2)

with $R = 2\rho_f/(\rho_f + 2\rho_p)$, $\mu = R/\text{St}$, $\text{St} = (2/9)(a/L)^2\text{Re}$, $\delta_a = \rho_a/\rho_f \ll 1$, $\delta_p = \rho_p/\rho_f \approx 1$ by design, and $R$ is close to $2/3$. Introducing the small parameter $\varepsilon = 1/\mu \ll 1$, we can rewrite our system as

$$\dot{\mathbf{x}} = \mathbf{v}$$

$$\varepsilon\dot{\mathbf{v}}(\mathbf{x},t) = \mathbf{u}(\mathbf{x},t) - \mathbf{v} + \varepsilon\left\{\frac{3R}{2}\frac{D\mathbf{u}(\mathbf{x},t)}{Dt} + \left(1 - \frac{3R}{2}\right)\mathbf{g}\right.$$
$$- R[f(\delta_p\mathbf{v}-\mathbf{u})^{\perp} + \delta_a\nu_a\alpha(\mathbf{v}-\mathbf{u}_{\text{wind}})$$
$$\left. - F(\mathbf{u}_e,\mathbf{u})]\right\}.$$

(A3)

Here, assuming the Ekman current velocity to be

$$\mathbf{u}_e = (0.0127/\sqrt{\sin|\lambda|})\mathbf{A}\mathbf{u}_{\text{wind}},$$

where $\mathbf{A}$ is a 45° rotation to the right in the Northern Hemisphere and $\lambda$ is latitude, we combine the additional slow wind functions into one function of Ekman velocity and ocean current, and drifter velocity $H[\mathbf{u}_e(x,t),\mathbf{u},\mathbf{v}]$:

$$\varepsilon\dot{\mathbf{v}}(\mathbf{x},t) = \mathbf{u}(\mathbf{x},t) - \mathbf{v} + \varepsilon\left\{\frac{3R}{2}\frac{D\mathbf{u}(\mathbf{x},t)}{Dt} + \left(1 - \frac{3R}{2}\right)\mathbf{g}\right.$$
$$\left. - Rf(\delta_p\mathbf{v}-\mathbf{u})^{\perp} + H[\mathbf{u}_e(\mathbf{x},t),\mathbf{u},\mathbf{v}]\right\}.$$

(A4)

Upon introducing the fast time by letting $\varepsilon\tau = t - t_0$, $\phi = t_0 + \varepsilon\tau$ be a dummy variable, and denoting differentiation with respect to $\tau$ by prime, we can rewrite our system as the autonomous dynamical system

$$x' = \varepsilon\mathbf{v}$$

$$\phi' = \varepsilon$$

$$\mathbf{v}'(\mathbf{x},\varphi) = \mathbf{u}(\mathbf{x},\phi) - \mathbf{v} + \varepsilon\left[\frac{3R}{2}\frac{D\mathbf{u}(\mathbf{x},\phi)}{Dt} + \left(1 - \frac{3R}{2}\right)\mathbf{g}\right.$$
$$\left. - Rf(\delta_p\mathbf{v}-\mathbf{u})^{\perp} + H(\mathbf{u}_e(\mathbf{x},\phi),\mathbf{u},\mathbf{v})\right].$$

(A5)

The $\varepsilon = 0$ limit of the system

$$x' = 0$$
$$\phi' = 0$$
$$\mathbf{v}'(\mathbf{x},\varphi) = \mathbf{u}(\mathbf{x},\phi) - \mathbf{v}$$

(A6)

has a set of fixed points. Taylor expanding these solutions in $\varepsilon$ gives

$$M_\varepsilon = [(\mathbf{x},\phi,\mathbf{v}): \mathbf{v} = \mathbf{u}(\mathbf{x},\phi) + \varepsilon\mathbf{u}^1(\mathbf{x},\phi) + \cdots$$
$$+ \varepsilon^r\mathbf{u}^r(\mathbf{x},\phi) + O(\varepsilon^{r+1}), (\mathbf{x},\phi) \in D_0].$$

(A7)

Equation (A6) restricted to $M_\varepsilon$ is a slowly varying system of the form

$$\mathbf{x}' = \varepsilon\mathbf{v}\big|_{M_\varepsilon}$$
$$= \varepsilon[\mathbf{u}(\mathbf{x},\phi) + \varepsilon\mathbf{u}^1(\mathbf{x},\phi) + \cdots + \varepsilon^r\mathbf{u}^r(\mathbf{x},\phi) + O(\varepsilon^{r+1})].$$

(A8)

Now, differentiating

$$\mathbf{v} = \mathbf{u}(\mathbf{x},\phi) + \sum_{k=1}^{r}\varepsilon^k\mathbf{u}^k(\mathbf{x},\phi) + O(\varepsilon^{r+1})$$

(A9)

with respect to $\tau$ gives

$$\mathbf{v}' = \mathbf{u}_x x' + \mathbf{u}_\phi \phi' + \sum_{k=1}^{r}\varepsilon^k(\mathbf{u}_x^k x' + \mathbf{u}_\phi^k \phi') + O(\varepsilon^{r+1}).$$

(A10)

From Eqs. (A6) and (A10) we also have

$$\mathbf{v}' = -\sum_{k=1}^{r}\varepsilon^k\mathbf{u}^k(\mathbf{x},\phi) + \varepsilon\left\{\frac{3R}{2}\frac{D\mathbf{u}(\mathbf{x},\phi)}{Dt} + \left(1 - \frac{3R}{2}\right)\mathbf{g}\right.$$
$$- Rf\left(\delta_p\sum_{k=0}^{r}\varepsilon^k\mathbf{u}^k(\mathbf{x},\phi) - \mathbf{u}\right)^{\perp}$$
$$\left. + H\left[\mathbf{u}_e(\mathbf{x},\phi),\mathbf{u},\sum_{k=0}^{r}\varepsilon^k\mathbf{u}^k(\mathbf{x},\phi)\right]\right\} + O(\varepsilon^{r+1}).$$

(A11)

Equating terms of equal power of epsilon in Eqs. (A10) and (A11) gives a leading-order approximation in our original time $t$: [this is identical to Eq. (4) in the main text]:

$$\dot{\mathbf{x}} = \mathbf{u}(\mathbf{x},t) + \varepsilon\left\{\left(1 - \frac{3R}{2}\right)\left[\frac{D\mathbf{u}(\mathbf{x},t)}{Dt} - \mathbf{g}\right]\right.$$
$$\left. - Rf(\delta_p - 1)\mathbf{u}^{\perp} + H[\mathbf{u}_e(\mathbf{x},t),\mathbf{u}]\right\}$$

The sign of the Coriolis term is the same as in Beron-Vera and LaCasce (2015) but appears different because of multiplication by $R$ and differing definitions $\delta$ and $\delta_p$.

Without better models of the wind–wave coupling, upper-ocean stratification, or more information on drifter specifics (size, geometry, density, compressibility, float–drifter tension, etc.), we have limited our modeling approach to a leading-order approximation of drifter behavior, and let a machine-learning algorithm learn the remaining terms as introduced by Wan et al. (2018) and Wan and Sapsis (2018). We remove gravity and are left with the following form of our blended model [this is identical to Eq. (5) in the main text]:

$$\dot{\mathbf{x}} = \mathbf{u}(\mathbf{x},t) + \varepsilon \left[ \left( \frac{3R}{2} - 1 \right) \frac{D\mathbf{u}}{Dt} - Rf(\delta_p - 1)\mathbf{u}^{\perp} \right]$$
$$+ \, \mathbf{G}[\boldsymbol{\xi}(t), \boldsymbol{\xi}(t-\tau), \boldsymbol{\xi}(t-2\tau), \ldots]$$
$$\boldsymbol{\xi}(t) = \left[ \mathbf{u}(\mathbf{x},t), \mathbf{u}_e(\mathbf{x},t), \frac{D\mathbf{u}}{Dt} \right],$$

where $\tau$ is a time delay, optimized for the frequency at which we want the LSTM to include prior trajectory information. We are thus forcing the data-driven model $\mathbf{G}$ to learn our as yet unknown float–drifter–wind coupling function $H$, as well as the higher-order terms from the Taylor expansion. The actual function we then model with our LSTM is of the form

$$\mathbf{G}[\boldsymbol{\xi}(t), \boldsymbol{\xi}(t-\tau), \boldsymbol{\xi}(t-2\tau), \ldots] = \underbrace{\dot{\mathbf{x}}}_{\text{Real drifter data}}$$
$$- \underbrace{\mathbf{u}(\mathbf{x},t) - \varepsilon \left[ \left( \frac{3R}{2} - 1 \right) \frac{D\mathbf{u}}{Dt} - Rf(\delta_p - 1)\mathbf{u}^{\perp} \right]}_{\text{Remote sensing and reanalysis data}},$$

$$(A12)$$

where all forces influencing $\dot{\mathbf{x}}$ that are not explicitly calculated from the remote sensing and reanalysis data on the right-hand side of Eq. (A12) are implicitly learned by $\mathbf{G}$.

## REFERENCES

Beron-Vera, F. J., and J. H. LaCasce, 2015: Statistics of simulated and observed pair separations in the Gulf of Mexico. *J. Phys. Oceanogr.*, **46**, 2183–2199, https://doi.org/10.1175/JPO-D-15-0127.1.

——, M. J. Olascoaga, and R. Lumpkin, 2016: Inertia-induced accumulation of flotsam in the subtropical gyres. *Geophys. Res. Lett.*, **43**, 12 228–12 233, https://doi.org/10.1002/2016GL071443.

——, A. Hadjighasem, Q. Xia, M. J. Olascoaga, and G. Haller, 2019a: Coherent Lagrangian swirls among submesoscale motions. *Proc. Natl. Acad. Sci. USA*, **116**, 18 251–18 256, https://doi.org/10.1073/pnas.1701392115.

——, M. J. Olascoaga, and P. Miron, 2019b: Building a Maxey–Riley framework for surface ocean inertial particle dynamics. *Phys. Fluids*, **31**, 096602, https://doi.org/10.1063/1.5110731.

Coelho, E. F., and Coauthors, 2015: Ocean current estimation using a multi-model ensemble Kalman filter during the Grand Lagrangian Deployment experiment (GLAD). *Ocean Modell.*, **87**, 86–106, https://doi.org/10.1016/j.ocemod.2014.11.001.

Copernicus Climate Change Service, 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store, digital media, https://cds.climate.copernicus.eu/.

D'Asaro, E. A., 2003: Performance of autonomous Lagrangian floats. *J. Atmos. Oceanic Technol.*, **20**, 896–911, https://doi.org/10.1175/1520-0426(2003)020<0896:POALF>2.0.CO;2.

——, and Coauthors, 2018: Correction for D'Asaro et al., Ocean convergence and the dispersion of flotsam. *Proc. Natl. Acad. Sci. USA*, **115**, E2664, https://doi.org/10.1073/pnas.1802701115.

Davis, R. E., 1985: Drifter observations of coastal surface currents during CODE: The method and descriptive view. *J. Geophys. Res.*, **90**, 4741–4755, https://doi.org/10.1029/JC090iC03p04741.

Edwards, K. P., F. E. Werner, and B. O. Blanton, 2006: Comparison of observed and modeled drifter trajectories in coastal regions: An improvement through adjustments for observed drifter slip and errors in wind fields. *J. Atmos. Oceanic Technol.*, **23**, 1614–1620, https://doi.org/10.1175/JTECH1933.1.

Ekman, V., 1905: On the influence of the Earth's rotation on ocean currents. *Ark. Mat. Aston. Fys.*, **2**, 1–53.

Ferrari, R., and C. Wunsch, 2009: Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annu. Rev. Fluid. Mech.*, **41**, 253–282, https://doi.org/10.1146/annurev.fluid.40.111406.102139.

Geyer, W. R., 1989: Field calibration of mixed-layer drifters. *J. Atmos. Oceanic Technol.*, **6**, 333–342, https://doi.org/10.1175/1520-0426(1989)006<0333:FCOMLD>2.0.CO;2.

Gonçalves, R. C., M. Iskandarani, T. Özgökmen, and W. C. Thacker, 2019: Reconstruction of submesoscale velocity field from surface drifters. *J. Phys. Oceanogr.*, **49**, 941–958, https://doi.org/10.1175/JPO-D-18-0025.1.

Haller, G., and T. Sapsis, 2008: Where do inertial particles go in fluid flows? *Physica D*, **237**, 573–583, https://doi.org/10.1016/j.physd.2007.09.027.

Haza, A. C., and Coauthors, 2018: Drogue-loss detection for surface drifters during the Lagrangian Submesoscale Experiment (LASER). *J. Atmos. Oceanic Technol.*, **35**, 705–725, https://doi.org/10.1175/JTECH-D-17-0143.1.

Hochreiter, S., and J. Schmidhuber, 1997: Flat minima. *Neural Comput.*, **9**, 1–42, https://doi.org/10.1162/neco.1997.9.1.1.

Jacobs, G. A., and Coauthors, 2014: Data assimilation considerations for improved ocean predictability during the Gulf of Mexico Grand Lagrangian Deployment (GLAD). *Ocean Modell.*, **83**, 98–117, https://doi.org/10.1016/j.ocemod.2014.09.003.

Lévy, M., P. J. Franks, and K. S. Smith, 2018: The role of submesoscale currents in structuring marine ecosystems. *Nat. Commun.*, **9**, 4758, https://doi.org/10.1038/s41467-018-07059-3.

Liu, Y., and R. H. Weisberg, 2011: Evaluation of trajectory modeling in different dynamic regions using normalized cumulative Lagrangian separation. *J. Geophys. Res.*, **116**, C09013, https://doi.org/10.1029/2010JC006837.

——, ——, S. Vignudelli, and G. T. Mitchum, 2014: Evaluation of altimetry-derived surface current products using Lagrangian drifter trajectories in the eastern Gulf of Mexico. *J. Geophys. Res. Oceans.*, **119**, 2827–2842, https://doi.org/10.1002/2013JC009710.

Lumpkin, R., T. Özgökmen, and L. Centurioni, 2017: Advances in the application of surface drifters. *Annu. Rev. Mar. Sci.*, **9**, 59–81, https://doi.org/10.1146/annurev-marine-010816-060641.

Martin, A. P., K. J. Richards, A. Bracco, and A. Provenzale, 2002: Patchy productivity in the open ocean. *Global Biogeochem. Cycles*, **16**, 1025, https://doi.org/10.1029/2001GB001449.

Maxey, M., and J. Riley, 1983: Equations of motion for a small rigid sphere in a nonuniform flow. *Phys. Fluids*, **26**, 883–889, https://doi.org/10.1063/1.864230.

Maximenko, N., J. Hafner, and P. Niiler, 2012: Pathways of marine debris derived from trajectories of Lagrangian drifters. *Mar. Pollut. Bull.*, **65**, 51–62, https://doi.org/10.1016/j.marpolbul.2011.04.016.

Niiler, P., 2001: The World Ocean surface circulation. *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*, 1st ed. International Geophysics Series, Vol. 77, Academic Press, 193–204, https://doi.org/10.1016/S0074-6142(01)80119-4.

Niiler, P. P., and J. D. Paduan, 1995: Wind-driven motions in the northeast pacific as measured by Lagrangian drifters. *J. Phys. Oceanogr.*, **25**, 2819–2830, https://doi.org/10.1175/1520-0485(1995)025<2819:WDMITN>2.0.CO;2.

——, R. E. Davis, and H. J. White, 1987: Water-following characteristics of a mixed layer drifter. *Deep-Sea Res.*, **34A**, 1867–1881, https://doi.org/10.1016/0198-0149(87)90060-4.

Novelli, G., C. M. Guigand, C. Cousin, E. H. Ryan, N. J. Laxague, H. Dai, B. K. Haus, and T. M. Özgökmen, 2017: A biodegradable surface drifter for ocean sampling on a massive scale. *J. Atmos. Oceanic Technol.*, **34**, 2509–2532, https://doi.org/10.1175/JTECH-D-17-0055.1.

Olascoaga, M. J., and Coauthors, 2013: Drifter motion in the Gulf of Mexico constrained by altimetric Lagrangian coherent structures. *Geophys. Res. Lett.*, **40**, 6171–6175, https://doi.org/10.1002/2013GL058624.

Poje, A. C., and Coauthors, 2014: Submesoscale dispersion in the vicinity of the Deepwater Horizon spill. *Proc. Natl. Acad. Sci. USA*, **111**, 12 693–12 698, https://doi.org/10.1073/pnas.1402452111.

Rio, M.-H., and F. Hernandez, 2003: High-frequency response of wind-driven currents measured by drifting buoys and altimetry over the world ocean. *J. Geophys. Res.*, **108**, 3283, https://doi.org/10.1029/2002JC001655.

Stewart, R. H., 2008: *Introduction to Physical Oceanography*. University Press of Florida, 345 pp, https://doi.org/10.1163/221160083x00124.

Sudre, J., C. Maes, and V. Garçon, 2013: On the global estimates of geostrophic and Ekman surface currents. *Limnol. Oceanogr. Fluids Environ.*, **3**, 1–20, https://doi.org/10.1215/21573689-2071927.

Takens, F., 1981: Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, D. Rand and L. Young, Eds., Springer-Verlag, 366–381.

Thomas, L. N., A. Tandon, and A. Mahadevan, 2008: Submesoscale processes and dynamics. *Ocean Modeling in an Eddying Regime, Geophys. Monogr. Ser.*, Vol. 177, Amer. Geophys. Union, 17–38.

Trenberth, K. E., and J. W. Hurrell, 1994: Decadal atmosphere-ocean variations in the Pacific. *Climate Dyn.*, **9**, 303–319, https://doi.org/10.1007/BF00204745.

Wan, Z. Y., and T. P. Sapsis, 2018: Machine learning the kinematics of spherical particles in fluid flows. *J. Fluid Mech.*, **857**, R2, https://doi.org/10.1017/jfm.2018.797.

——, P. Vlachas, P. Koumoutsakos, and T. Sapsis, 2018: Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PLOS ONE*, **13**, e0197704, https://doi.org/10.1371/journal.pone.0197704.

Wang, Y., M. J. Olascoaga, and F. J. Beron-Vera, 2015: Coherent water transport across the South Atlantic. *Geophys. Res. Lett.*, **42**, 4072–4079, https://doi.org/10.1002/2015GL064089.

Zavala-Hidalgo, J., R. Romero-Centeno, A. Mateos-Jasso, S. L. Morey, and B. Martínez-López, 2014: The response of the Gulf of Mexico to wind and heat flux forcing: What has been learned in recent years? *Atmosfera*, **27**, 317–334, https://doi.org/10.1016/S0187-6236(14)71119-1.