

**Disentangling resolution, precision, and inherent stochasticity in nonlinear systems**Lei Fang,<sup>1</sup> Sanjeeva Balasuriya<sup>2</sup>, and Nicholas T. Ouellette<sup>1,\*</sup><sup>1</sup>*Department of Civil and Environmental Engineering, Stanford University, Stanford, California 94305, USA*<sup>2</sup>*School of Mathematical Sciences, University of Adelaide, Adelaide, SA 5005, Australia*

(Received 23 July 2019; accepted 1 June 2020; published 16 June 2020)

Reliable measurement, simulation, and analysis of dynamical systems rely on appropriately bounded uncertainty. Errors that lead to uncertainty naturally arise from finite precision or resolution, but an additional unappreciated source of uncertainty is the effective stochasticity associated with nonlinear dynamics. Here we describe and quantify the interplay between these three sources of uncertainty using a recently developed framework known as stochastic sensitivity theory. Using fluid mixing as a test case and considering data from an analytical flow, a laboratory experiment, and geophysical observations, we show how to delimit regimes that are limited by finite resolution or by inherent stochasticity. We arrive at the surprising conclusion that in some cases, refining the resolution of a measurement or simulation can actually be counterproductive and lead to an outcome that is less faithful to the true dynamics. Our results have significant implications for the measurement and analysis of nonlinear systems.

DOI: [10.1103/PhysRevResearch.2.023343](https://doi.org/10.1103/PhysRevResearch.2.023343)**I. INTRODUCTION**

Nonlinear systems are inherently more challenging to handle than their linear counterparts. Nonlinearities tend to amplify tiny differences in initial and boundary conditions as well as the effects of noise, making them prone to instability and chaos. These issues can be even more severe in spatially extended nonlinear systems, where the effects of the nonlinearities are often highly spatially nonuniform, or when the time evolution of the system is desired, where the effects can compound multiplicatively. These difficulties are also endemic in every mode of scientific inquiry. Nonlinear systems often defy analytical solution; they require specialized, computationally expensive algorithms to handle numerically; and they place stringent constraints on experimental or observational resolution and measurement accuracy.

Given all of these difficulties, it is often not clear how best to design empirical studies of nonlinear systems, whether numerical, experimental, or observational, to achieve optimal fidelity given inevitable constraints. We can distinguish three logically distinct sources of potential error that must all be accounted for in this design process. First, error can come from limited *resolution*. Processes that occur, for example, on length scales smaller than the grid size in a computational simulation do not exist in the numerical results; and even though experiments contain all of the proper physics, the impacts of processes on scales smaller than the experimental resolution may be improperly accounted for. Second, errors

can arise due to limited *precision*. Such errors can come from, for example, the finite accuracy of numerical methods or measurement errors in experiments. And finally, effective errors can arise from the *inherent dynamics* of the system—that is, from dynamical effects from the equations of motion that lead to effective stochasticity that will be present for any finite resolution or precision. The well-known extreme sensitivity to tiny changes in initial conditions in chaotic dynamical systems is an example of this [1]. A well-designed empirical study must confront all of these sources of error and weigh them appropriately. However, this can be difficult to do *a priori*; in particular, the interplay of the inherent nonlinearity due to the dynamics and the potentially controllable effects of resolution and precision has received little attention.

Here, we present a framework for disentangling and analyzing these interwoven effects in computations using spatiotemporal data, using the recently developed stochastic sensitivity theory [2]. We illustrate our results using examples drawn from fluid mixing, a canonical nonlinear problem that displays all of the challenges discussed above and that has a wide range of important applications [3]. Studies of mixing start from the Eulerian fluid velocity field, which is governed by the Navier–Stokes equations; when the Reynolds number is large (as is the case in most flows of industrial or environmental interest), these equations are highly nonlinear. But since mixing is fundamentally a question of how material is transported by the flow, it is natural to consider it from a Lagrangian perspective instead by studying the trajectories followed by material parcels [4]. These trajectories are computed by integrating the velocity field in time, and so even small errors in these fields can compound multiplicatively and produce large deviations [5]. It is thus natural to try to increase both resolution and precision as much as possible to enable the best possible computation of the trajectories. However, such an approach ignores the inherent unpredictability in the problem, which arises from both the nonlinear complexity

\*nto@stanford.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

resulting from the dynamics generated from the data, and fact that the data itself possesses uncertainties. Here, we demonstrate that because of this inherent unpredictability, in some limits increased resolution and precision not only does not help but can in fact counterintuitively hinder our ability to construct accurate trajectories.

We illustrate the use of our theoretical framework with three examples. First, we describe a canonical model flow, the double gyre, that has become a standard testbed for studying coherent structures [4,6,7]. In this case, we have full knowledge of the flow field, and so can independently study the effects of resolution, precision, and dynamics. Next, we consider a laboratory realization of two-dimensional turbulence, where the nonlinearities are stronger and the flow is not *a priori* known. In this case, the resolution is very high, but measurement error is unavoidable and is typically tacitly considered to be the main limitation. Finally, we analyze observational flow data from the Gulf of Mexico, where finite resolution is usually assumed to be the limiting factor.

## II. STOCHASTIC SENSITIVITY

For our investigation of nonlinearity, precision, and resolution, we apply tools from a recent theoretical development in stochastic differential equations that we term stochastic sensitivity theory. The details of this theory, including proofs of its main conclusions and guarantees of computability, are set out in Ref. [2]. Here, we only lay out the essential ideas and notation necessary for our context; further details are given in Appendix A.

The central question addressed by the theory in Ref. [2] is how to account for the effects of error on the solution trajectories of nonlinear systems given by the ordinary differential equation(s),

$$\dot{\mathbf{x}} = \mathbf{u}(\mathbf{x}, t). \quad (1)$$

In the fluid flow examples we will focus on here,  $\mathbf{u}$  should be interpreted as the Eulerian velocity field that is a solution of the Navier–Stokes equations. However, nothing in the theory rests on the details of fluid mechanics, and so the results are generalizable. Formally, the governing ODE admits the solution

$$\mathbf{F}_0^t(\mathbf{x}) = \mathbf{x} + \int_0^t \mathbf{u}(\mathbf{F}_0^\tau(\mathbf{x}), \tau) d\tau, \quad (2)$$

where  $\mathbf{F}_0^t(\mathbf{x})$  is the flow map that transforms an initial condition  $\mathbf{x}$  at time 0 to its location at time  $t$  under the action of the flow  $\mathbf{u}$ . This integral should be interpreted as a path integral that is evaluated along the solution trajectory. This formal solution, however, makes the implicit assumption that knowledge of  $\mathbf{u}$  is perfect and error-free, which will never be the case in any practical situation. An actually observed solution  $\mathbf{y}_t$  would therefore be different from  $\mathbf{F}_0^t(\mathbf{x})$ . Treating the error in the velocity as random, the idea in Ref. [2] is to write  $\mathbf{y}_t$  as the solution of the stochastic differential equation

$$d\mathbf{y}_t = \mathbf{u}(\mathbf{y}_t, t)dt + \varepsilon d\mathbf{W}_t, \quad (3)$$

where  $\varepsilon$  is a positive nondimensional parameter that expresses the expected level of error and  $d\mathbf{W}_t$  is the (vector) Wiener process. The impact of the Eulerian velocity errors on the

Lagrangian locations is then encapsulated in the statistics of the random variable

$$\mathbf{z}_\varepsilon(\mathbf{x}, t) = \frac{\mathbf{y}_t - \mathbf{F}_0^t(\mathbf{x})}{\varepsilon}. \quad (4)$$

Of particular interest is the limit  $\varepsilon \rightarrow 0$ , which captures the inherent amplification of infinitesimal uncertainty due to the nonlinear dynamics of the system. Such effects will be present for *any* nonzero error, no matter how small.

Since  $\mathbf{z}_\varepsilon$  can be oriented in any direction relative to the true location  $\mathbf{F}_0^t(\mathbf{x})$ , Ref. [2] considers the projection of  $\mathbf{z}_\varepsilon$  onto a general unit vector  $\hat{\mathbf{n}}$ , labeled here as  $P_\varepsilon(\mathbf{x}, t, \hat{\mathbf{n}})$ . Stochastic sensitivity is defined as the maximum variance of the set of all  $P_\varepsilon$  in the limit  $\varepsilon \rightarrow 0$ ; that is,

$$S^2(\mathbf{x}, t) \equiv \lim_{\varepsilon \rightarrow 0} \sup_{\hat{\mathbf{n}}} \text{var} [P_\varepsilon(\mathbf{x}, t, \hat{\mathbf{n}})]. \quad (5)$$

$S^2$  has units of area, so that  $\sqrt{S^2}$  is a length. Defined in this way,  $S^2$  is a scalar field that takes a value at each position  $\mathbf{x}$  in the domain of interest. Importantly, this seemingly complicated expression for  $S^2$  is nonetheless explicitly computable for any given velocity data without having to perform stochastic simulations; Appendix A details this computation, derived in Ref. [2]. The  $S^2$  field encapsulates the interaction between the nonlinearity in the data—which can generate spatially varying unpredictability *à la* chaotic systems—and the inherent uncertainty in the data itself in quantifying the uncertainty in the eventual Lagrangian location.

We now adapt this theory to investigate the role of the resolution length scale. Going back to the definition of  $\mathbf{z}_\varepsilon$ , we can argue that, to leading order, the size of the deviation between our estimate of the solution to the governing ODE and the true solution will be  $L_S(\mathbf{x}) \equiv \varepsilon\sqrt{S^2}$ ; note that  $L_S$  is still a length scale, since  $\varepsilon$  is nondimensional. Additionally, since  $S^2$  depends on  $\mathbf{x}$ ,  $L_S(\mathbf{x})$  is in fact a field on the particle locations at time 0, and can vary strongly between different regions of a dynamical system.

Next, we observe that regions of the flow field with small  $S^2$  will be highly robust to errors, since the product  $\varepsilon\sqrt{S^2}$  can remain small in those cases even when  $\varepsilon$  is on its own relatively large. But more usefully, it allows us to draw inferences about the required resolution with which we must know the velocity field as well as how much we can trust different features we observe. In many cases, the error  $\varepsilon$  is more or less spatially uniform and is set by the measurement technique (for an experiment) or the accuracy of a numerical scheme (for a computational model). In that case, we can directly compare the available spatial resolution  $L_R$  of the field—either the typical distance between measurement locations in an experiment or the grid spacing in a simulation—with  $L_S(\mathbf{x})$ . In regions where  $L_S < L_R$ , the uncertainty engendered from the dynamics—incorporating the uncertainty in the Eulerian velocity field—is smaller than the spatial resolution. Thus, finite spatial resolution is the dominant factor that controls our ability to reconstruct proper solution trajectories, and we say that we are *resolution limited*. In resolution-limited regions of the flow, increasing the spatial resolution will improve our estimate of the dynamics. However, when  $L_S > L_R$ , the dominant factor that determines the quality of our estimate is the inherent uncertainty in the dynamics, and we say that

we are *stochasticity limited*. In this case, enhancing the spatial resolution does not improve our ability to estimate solutions, because the uncertainty from the dynamics,  $L_S$ , dominates.

Optimally, we argue that one wants to choose  $L_R \approx L_S$ . In this case, the resolution is sufficiently high that it is not a limiting factor. But, as we show below, increasing the spatial resolution by making  $L_R \ll L_S$  is not only not helpful but actually counterproductive, since in that case one finely resolves the inherent unpredictability in the system and allows it to influence the solutions multiplicatively. A better choice is to keep  $L_R \approx L_S$  and simply interpolate to smaller scales, bypassing most of the inherent unpredictability resulting from the dynamics driven by uncertain velocities.

### III. RESULTS

#### A. Analytical flow

We now illustrate the use of this stochastic sensitivity framework in three examples. We begin with the double gyre [6], an analytical flow that has been well studied in the coherent-structures literature [4,8]. The essential transport dynamics of the double gyre is understood; and, because it is analytically specified and thus does not require the solution of any differential equations, we can use it to test the effects of arbitrary levels of resolution and precision.

The double gyre was developed as a two-dimensional, simplified model of flow in an ocean basin, and consists of two vortices whose relative size oscillates back and forth. The components of the velocity field are given by

$$u_x = -\pi A \sin(\pi f) \cos(\pi y) \tag{6}$$

and

$$u_y = \pi A \cos(\pi f) \sin(\pi y) \frac{\partial f}{\partial x}, \tag{7}$$

where  $f = ax^2 + bx$ ,  $a = \alpha \sin \omega t$ , and  $b = 1 - 2\alpha \sin(\omega t)$ , and it is defined on the domain  $x \in [0, 2]$  and  $y \in [0, 1]$ . Here, we used  $A = 1$ ,  $\alpha = 0.02$ , and  $\omega = 2\pi$ . Since  $f$  is a function of time, this flow is unsteady. The velocity field for these parameters is shown at  $t = 0$  in Fig. 1(a).

The two vortices in the double gyre are separated by a transport barrier, a co-dimension one structure (that is, a curve, in this case) that fluid elements cannot cross. Although this barrier is defined in the infinite-time limit in this periodic flow, it is typically revealed by computing the finite-time Lyapunov exponents (FTLEs). As has become standard practice [7], we compute FTLEs by first constructing the Cauchy–Green strain tensor, defined as the inner product of the gradient of  $F_0^T(\mathbf{x})$  with itself. The logarithm of the square root of the largest eigenvalue of this tensor, when scaled by the advection time  $T$ , is the FTLE. The FTLE gives an estimate of the local stretching rate of fluid elements, and ridges of the FTLE field tend to correspond to transport barriers [7]. The FTLE field for the double gyre is shown in Fig. 1(b). The expected transport barrier is revealed by the sharp ridge of the FTLE field between the two vortices.

We now compute  $L_S = \varepsilon \sqrt{S^2}$  as a function of space, as shown in Fig. 1(c). The required intermediate calculation of  $S^2$  is performed using the method described in Appendix A. Here we have used  $\varepsilon = 0.1$ , but our results are fairly insen-

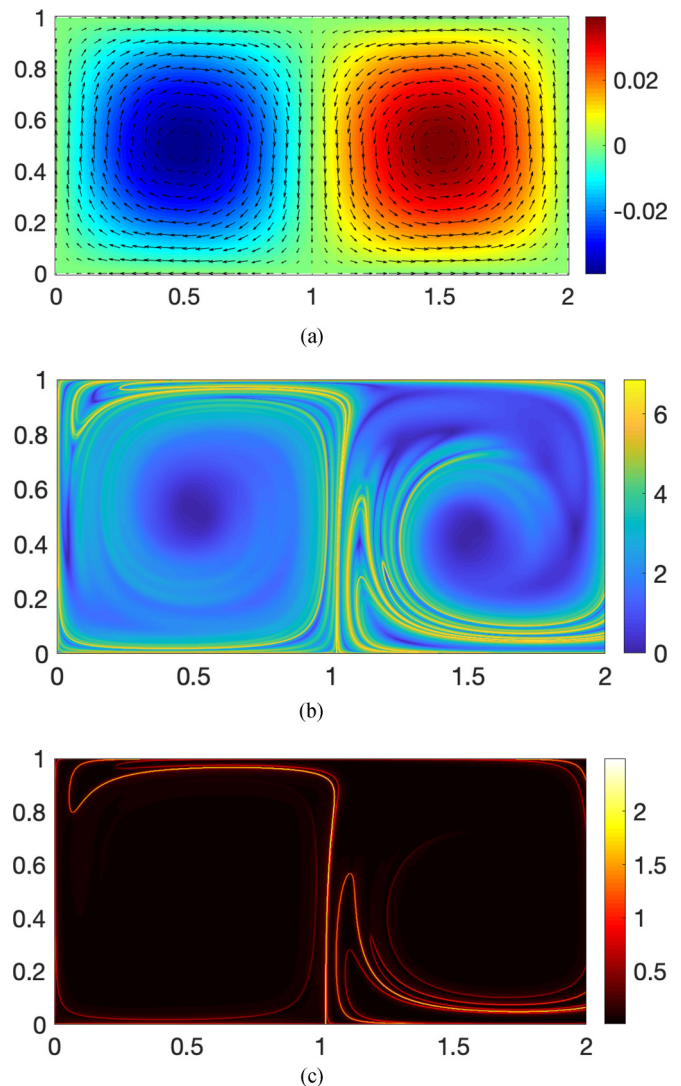


FIG. 1. The double gyre. (a) Velocity and vorticity fields. (b) FTLE field, computed for an integration time of two flow cycles. (c)  $L_S$  field, computed for a time of two flow cycles.

sitive to a reasonably chosen value of  $\varepsilon$ . Recall that the local value of  $L_S(\mathbf{x})$  gives us information about the strength of the inherent stochasticity due to the dynamics of the system; thus, it should not be surprising that the  $L_S$  field takes on high values exactly on the FTLE ridge. The high local stretching due to the flow that is picked out by the FTLE also makes the dynamics highly sensitive to noise in this region of the flow, since fluid elements on either side of the barrier will experience very different flows. However, the FTLE and  $L_S$  measure slightly different susceptibilities associated with the flow: the former measures amplifications to deviations in the initial condition, whereas the latter quantifies the impact of *continuing uncertainty* in the velocity (or, more generally, the nonlinear evolution model). Additionally, the  $L_S$  field gives us much more information than just the location of the transport barrier, because the values it takes tell us something about the resolution we need to locate this barrier. In this case, we find that  $L_S \approx 2$  at maximum in the double gyre, suggesting that we only require a resolution of about  $L_R \approx 2$ . Recall, though,

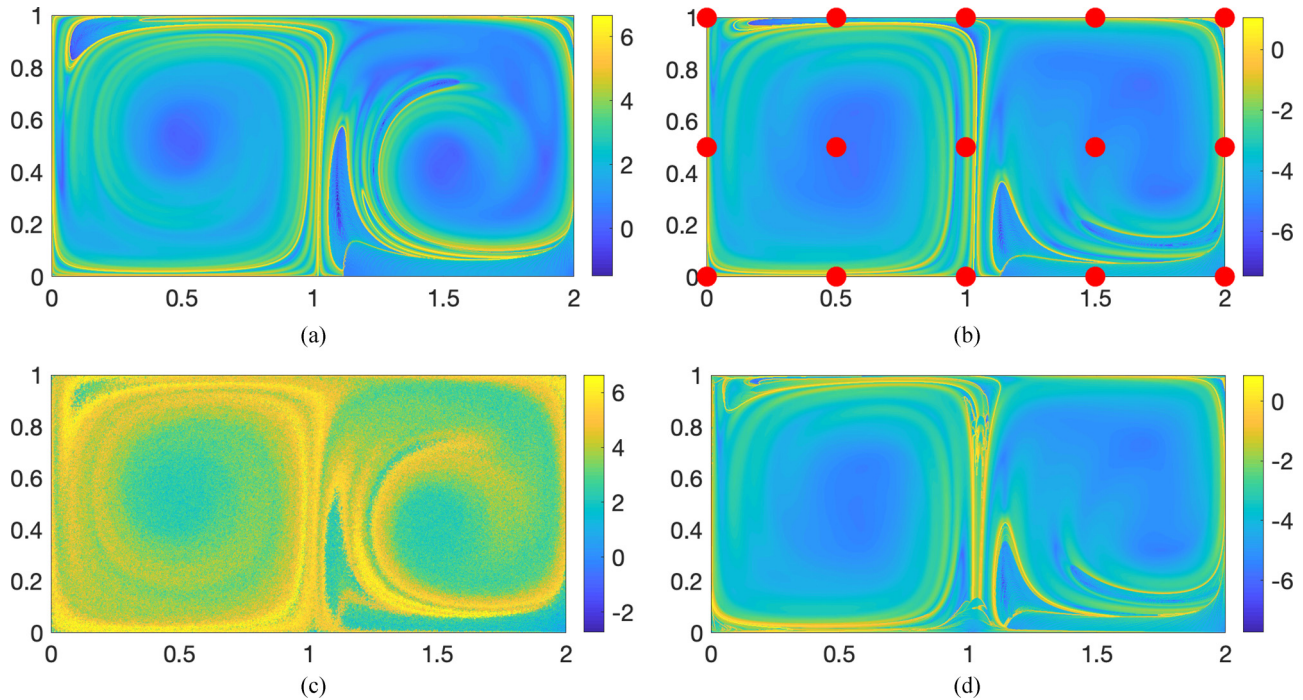


FIG. 2. FTLE fields computed for the double gyre with reduced resolution and precision. (a) FTLE field computed by interpolating in a discretized but highly resolved ( $L_R = 0.002$ ) velocity field. (b) FTLE field computed by interpolating a poorly resolved ( $L_R = 0.5$ ) velocity field. Symbols show the locations of the velocity vectors from which the field is interpolated. (c) FTLE field computed for  $L_R = 0.002$  but with added noise with a magnitude of 10% of the root-mean-square velocity at each point (see text for details). (d) FTLE field computed for  $L_R = 0.5$  with 10% added noise.

that the double gyre is only defined in the domain  $x \in [0, 2]$  and  $y \in [0, 1]$ —suggesting the surprising conclusion that our resolution can be as coarse as the entire domain and still give us a good estimate of the essential dynamics of the flow.

To test this prediction, we computed the FTLE field not using the analytical values of the flow field everywhere but rather only on a discrete grid of locations and then interpolating to estimate the velocity of a fluid element at other spatial locations. In this way, we are mimicking the kind of data one would obtain from a numerical simulation or an experiment. In such cases—and particularly for experiments—continuous data are simply not available, and refining the resolution is often not possible due to constraints such as computational resources or measurement technology. Interpolation is then the only possible way to recover data at locations away from measurements. In this example, we use simple linear interpolation; our results are qualitatively insensitive to the interpolation scheme chosen. In Fig. 2(a), we show the FTLE field computed in this way using a high resolution of  $L_R = 0.002$ , much smaller than  $L_S$  and therefore deep in the stochasticity limited regime. As one would expect, the FTLE fields shown in Figs. 1(b) and 2(a) are very similar. What is surprising, however, is how good the estimate of the FTLE field is in Fig. 2(b), where we set  $L_R = 0.5$  and specify only 15 velocity vectors over the entire domain [indicated by the symbols in Fig. 2(b)] as opposed to  $5 \times 10^5$  vectors in Fig. 2(a). Some of the fine detail is lost, but the essential structure of the transport barrier remains. This result provides strong support for the prediction from stochastic sensitivity theory that the structure of the dynamics is preserved even for  $L_R \approx L_S$ , as well as

the surprising conclusion that the resolution-limited regime essentially does not exist for the double gyre.

Our theory also suggests that this low-resolution field ought to be more robust to low precision, since small errors will not be amplified by the inherent dynamics. To test this prediction, we added uniformly distributed random noise to the velocity vectors specified on our discrete grids by adding an additional vector to each velocity with a magnitude of 10% of the root-mean-square velocity and a randomly chosen direction. The FTLE fields for these noisy velocity fields are shown in Fig. 2(c) for  $L_R = 0.005$  and Fig. 2(d) for  $L_R = 0.5$ . Although both cases still show the transport barrier, the effects of the lowered precision are indeed more evident in the high resolution case (deeper in the stochasticity-limited regime).

### B. Experimental quasi-two-dimensional turbulence

In an analytical flow field like the double gyre, variation of resolution or precision is inherently artificial. Additionally, since the double gyre is periodic, the transport barrier that separates the two vortices is not a transient feature of the dynamics but rather is present for all time, and thus very robust. We therefore turn now to a highly resolved laboratory experiment to understand what our theory can tell us about a more realistic situation where less is known *a priori*.

We consider a laboratory quasi-two-dimensional turbulent flow. Our apparatus and measurement techniques are described in Appendix B. We have access to highly resolved (in both space and time) velocity fields, sampled finely with respect to the dominant length scale  $L_m$  and associated time

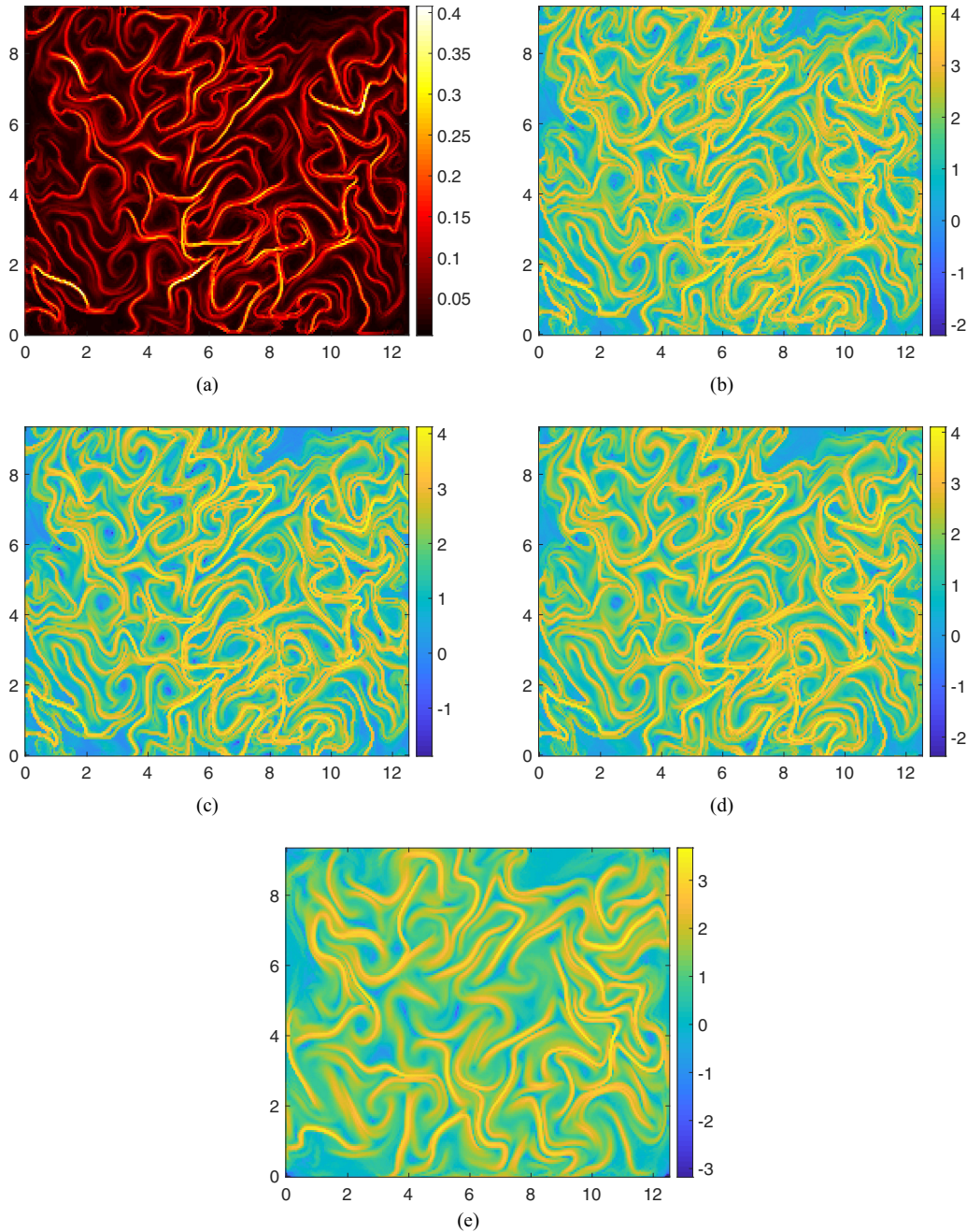


FIG. 3. Dynamical structure of the experimental quasi-two-dimensional turbulent flow. (a)  $L_S$  field, given in units of  $L_m$ , the forcing scale of the experiment (see Methods in Appendix B). The horizontal and vertical extent of the measurement domain are also given in units of  $L_m$ . (b) FTLE field computed using the full experimental resolution of  $L_{R,max} \approx 0.06L_m$ . FTLEs are given in units of  $T_L$ , the typical time scale of the flow. (c) FTLE field computed with a reduced resolution of  $L_R = 0.1L_m$ , in the stochasticity limited regime. (d) FTLE field computed with a resolution of  $L_R = 0.18L_m = 2\langle L_S^2 \rangle^{1/2}$ . (e) FTLE computed with a resolution of  $L_R = 0.5L_m$ , well into the resolution limited regime.

scale  $T_L$  of the turbulence. However, we do not know the relevant length scale of the inherent unpredictability, which may not map onto these scaling estimates exactly. Additionally, as in any experiment, our measurements of the velocity field have some associated finite precision that is also not known exactly. Thus, in this experimental case, we do not have ground-truth information in the same way that we do for an analytical flow, so we must analyze the data in a different way.

In Fig. 3(a), we show the  $L_S$  field for a representative snapshot of the experimental data, again choosing  $\varepsilon = 0.1$ . It

shows a much more complex pattern of high values, indicating the additional complexity added to the dynamics by the turbulence. As expected, this pattern again corresponds well to the FTLE field computed for the same data using the full resolution of the data set,  $L_{R,max} \approx 0.06L_m$  [Fig. 3(b)]. As we artificially coarsen the resolution by reconstructing gridded velocity fields using only subsamples of the full data set (and recomputing Lagrangian trajectories based on these coarser fields), the FTLE patterns change somewhat, though many of the qualitative features remain as long as we stay in the

stochasticity limited regime [Figs. 3(c) and 3(d)]. We note that we keep the same number of grid points for all values of  $L_R$ ; what changes is the number of measurements used to create these gridded fields. In this case, since the  $L_S$  field is highly variable in space, we use a cutoff of  $2\langle L_S^2 \rangle = 0.18L_m$  to separate the stochasticity limited and resolution limited regimes. This cutoff separates out the ridges of the  $L_S$  field from the background, delineating the two; note, however, that our results are relatively insensitive to the choice of this cutoff as long as it identifies the ridges adequately. Once we make the resolution coarse enough that we fall into the resolution limited regime, the pattern begins to change qualitatively as well as quantitatively, as smaller-scale features are lost and the FTLE ridges shift position [Fig. 3(e)].

To quantify the effects of reduced resolution, we computed the standard cross-correlation of the FTLE fields for different  $L_R$  with the field computed for  $L_{R,\max}$ . We show the correlation coefficient (that is, the normalized correlation function) in Fig. 4(a). This correlation is close to one for small  $L_R$ , but begins to fall off more rapidly for larger  $L_R$  values, as one would expect: with a coarser resolution, fine-scale features cannot be resolved. The change in slope occurs near the transition from the stochasticity-limited regime to the resolution-limited regime.

We can also evaluate the effect of adding artificial noise to the experimental data, which we implemented just as we did above for the double gyre. For each value of  $L_R$ , we created a set of artificially noisy velocity fields with different noise amplitudes  $\xi$ , which we used to compute FTLE fields. We then computed the cross-correlation of the FTLE fields with artificial noise with the fields computed with no added noise. For each value of  $L_R$ , we found that these correlations decayed roughly exponentially as the noise amplitude  $\xi$  increased. We thus fit them (using a least-squares approach) with a function of the form  $C \exp(-b\xi)$ , where  $C$  is an overall constant and  $b$  is a decay rate. The decay rate  $b$  gives a measure of how rapidly the structure of the FTLE field degrades as the noise is increased; for smaller  $b$ , the added noise is less significant. In Fig. 4(b), we show how  $b$  varies with  $L_R$ . In accordance with the counterintuitive prediction we made above, the decay of the correlation is *slower* for *lower-resolution* data. The impact of limited precision is indeed more severe in the stochasticity limited regime where it is amplified by the inherent complex dynamics of the system. However, the overall structure of the field does become less faithful to the real dynamics when  $L_R$  increases [Fig. 4(a)], and so a balance must be struck. Hence, as argued above, choosing a resolution comparable to the typical  $L_S$  value is a good choice.

### C. Remotely sensed ocean flow

As a final case study, we turn to satellite altimetry data for the Gulf of Mexico. Details of the data are given in Appendix C. FTLEs (and related tools for detecting Lagrangian coherent structures) have been used previously in the Gulf of Mexico to elucidate the dynamical structures that guide transport [9]. They have been used, for example, to draw conclusions about the presence of a persistent transport barrier on the West Florida Shelf [10,11], to understand how large-scale flow features evolve [12], and to characterize the spread

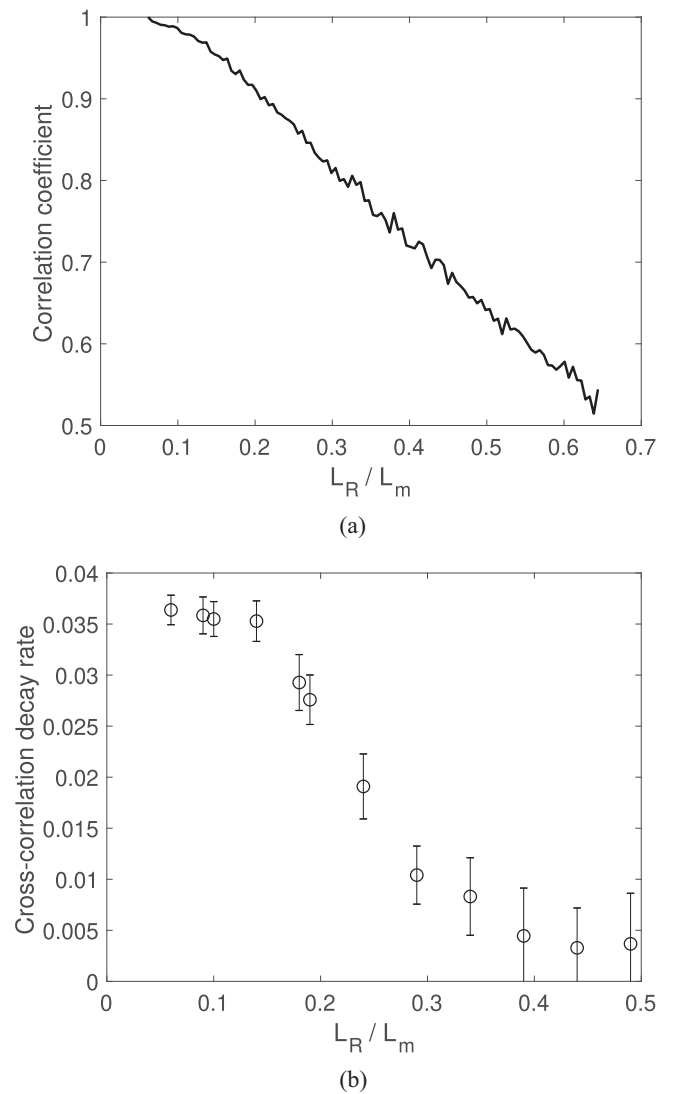


FIG. 4. Effects of reduced resolution and precision on the experimental data. (a) Cross-correlation of the full-resolution FTLE field and the reduced resolution FTLE fields as a function of the resolution  $L_R$ . (b) Decay rate  $b$  (see text) of the cross-correlation of the raw FTLE fields and FTLE fields computed with artificially added noise as a function of the resolution  $L_R$ . Error bars show the 95% confidence interval.

of oil after the Deepwater Horizon spill in 2010 [13]. The raw data for such studies is typically acquired from satellite remote sensing. Such an approach has intrinsic resolution limitations relative to the actual scales of ocean flow, which in principle range from planetary scales down to the finest scales of motion that can be of order meters or smaller. As we have argued here, resolution plays a key role in determining which features of the FTLE field are faithful to the underlying dynamics; however, the implications of this observation for examining transport in the Gulf of Mexico (or other geophysical flows) have not been examined.

The dataset we use has a spatial resolution of  $L_R = 0.25^\circ$  and a temporal resolution of one day. Using these data, we can compute an FTLE field, as shown in Fig. 5(a). These FTLEs reproduce many of the features commonly reported in the

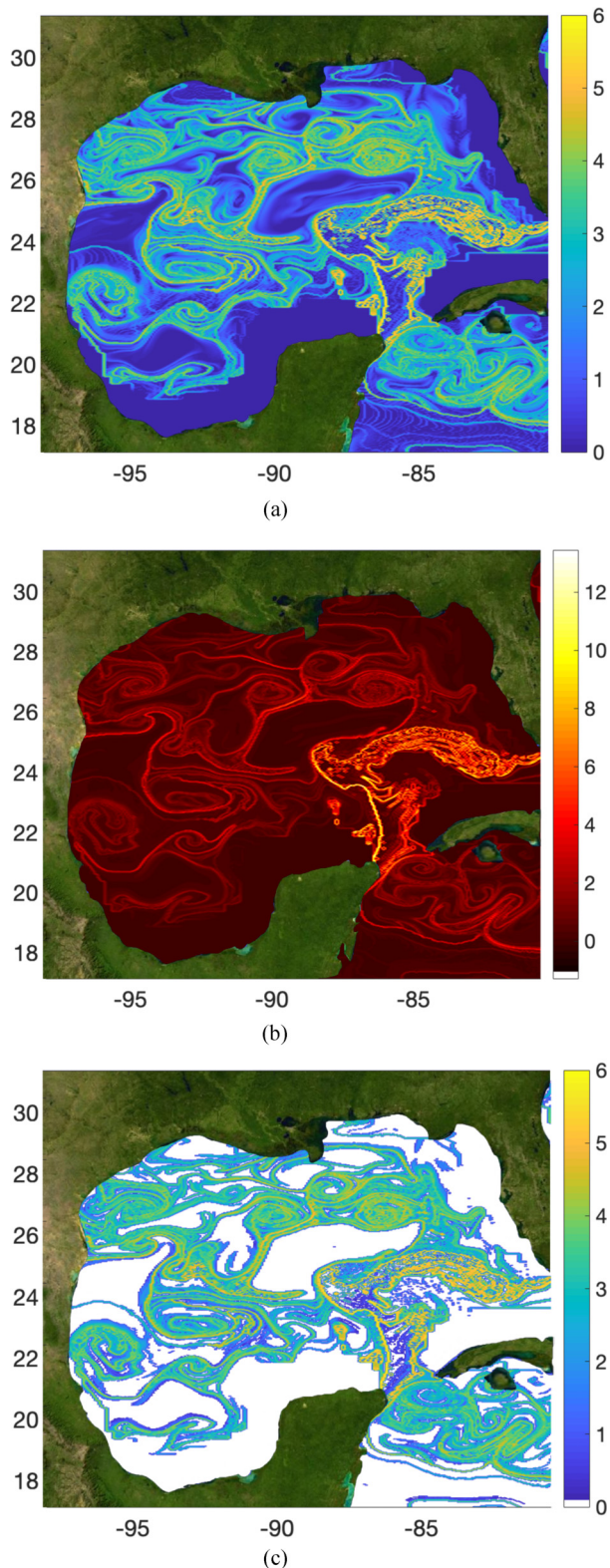


FIG. 5. Our theory applied to flow in the Gulf of Mexico. (a) FTLE field computed for a 60-day period beginning March 2, 2017, in units of inverse days. (b)  $L_S$  field for the same time range, in units of degrees. (c) The same data as in panel (a), but with regions that are resolution-limited (i.e., those with  $L_R > L_S$ ) masked out in white. Any structures identified in these regions are highly questionable, in that they may appear in the wrong places due to the unresolved physical processes.

Gulf, including the weakly mixing region off the west coast of Florida. We can also now use our theory to compute the  $L_S$  field, as shown in Fig. 5(b). In this case, we choose a smaller  $\varepsilon$  ( $\varepsilon = 0.05$ ), as this will allow us to make a conservative estimate of trustworthy regions below. The  $L_S$  field again has many of the same qualitative features as the FTLE field. But since  $L_S$  is a length scale, we use it in this case to assess where we can trust the FTLE values and where they might be questionable. As we showed above for the experimental data, FTLE ridges for the resolution-limited case are often not faithful to the underlying dynamics, and may appear in the wrong places. In this case, we can identify these resolution-limited regions using the  $L_S$  field: wherever  $L_S < L_R$ , the data will be resolution limited. In Fig. 5(c), we mask out these regions of the domain, and show only the FTLEs in regions that are stochasticity limited. Although these may contain noise and will be affected by the inherent unpredictability in the flow, they will in general appear in the proper places, and can thus be interpreted in terms of transport.

#### IV. DISCUSSION

Measurement of complex dynamical systems is challenging for a number of reasons, including imprecise techniques, limited resolution, and inherent complexity in the dynamics. Our central message here is that by disentangling these logically distinct sources of uncertainty, one can more appropriately design measurement and simulation schemes and better interpret empirical data. Our adaptation of stochastic sensitivity theory allows us to precisely compute the length scale at which the inherent dynamics will affect measurements of the system. Over-resolving this scale does not necessarily help to characterize the system, since measurements in this stochasticity limited regime will always reflect this unavoidable source of uncertainty, and any measurement error will be amplified. At the same time, the resolution cannot be too coarse, since in that resolution limited regime small-scale features are simply not present. Thus, the optimal choice is to set the resolution to be comparable to the stochastic sensitivity length scale.

This argument also allows one to estimate how much precision is required in an experiment, simulation, or observation given a fixed resolution to reliably uncover features given the dynamics. Since  $L_S = \varepsilon\sqrt{S^2}$  and we argue that we want  $L_R \approx L_S$ , a tolerable level of precision is given by  $\varepsilon \approx L_R/\sqrt{S^2}$ . Thus, both the resolution and the inherent dynamics come into play when making this estimate. If  $\varepsilon$  is larger than this bound, then  $L_R > L_S$  and the results will be untrustworthy. Since  $\varepsilon$  can be thought of as a relative error estimate in the velocity data, we are thus able to link resolution  $L_R$  and precision  $\varepsilon$  in relation to certainties in predicting Lagrangian trajectories.

Finally, even though we have demonstrated these ideas with examples drawn from fluid mixing, there is nothing in our analysis that is unique to fluid mechanics. Thus, all of the conclusions we have drawn here are straightforwardly extendable to other nonlinear complex systems, and we encourage researchers to consider them when designing empirical studies.

## ACKNOWLEDGMENTS

L.F. and N.T.O. acknowledge support from the US National Science Foundation under Grant No. CMMI-1563489. S.B. acknowledges support from the Australian Research Council under Grants No. FT130100484 and No. DP200101764.

## APPENDIX A: COMPUTING STOCHASTIC SENSITIVITY

To compute the stochastic sensitivity  $S^2(\mathbf{x}, t)$ , we use the formulas derived in Ref. [2]. We specialize to the situation of two-dimensional, incompressible flow subject to an isotropic constant diffusivity tensor (more generalized expressions which relax these conditions are available in Ref. [2]). Under these conditions,  $S^2(\mathbf{x}, t)$  in Eq. (5) can be expressed as

$$S^2(\mathbf{x}, t) = \frac{1}{2} \int_0^t \sum_{i,j=1}^2 \Lambda_{ij}(\mathbf{w}, \tau) \Lambda_{ij}(\mathbf{w}, \tau) d\tau + N(\mathbf{w}, t), \quad (\text{A1})$$

where  $\mathbf{w} = \mathbf{F}_0^t(\mathbf{x})$  is the final location to which  $\mathbf{x}$  is advected exactly under  $\mathbf{F}$ ,  $\Lambda_{ij}$  are the components of the matrix  $\Lambda$  given by

$$\Lambda(\mathbf{w}, t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \nabla \mathbf{F}_t^t(\mathbf{w}), \quad (\text{A2})$$

and

$$N(\mathbf{w}, t) = \sqrt{P^2(\mathbf{w}, t) + M^2(\mathbf{w}, t)}, \quad (\text{A3})$$

$$P(\mathbf{w}, t) = \frac{1}{2} \int_0^t \left[ \sum_{i=1}^2 \Lambda_{i2}^2(\mathbf{w}, \tau) - \sum_{i=1}^2 \Lambda_{i1}^2(\mathbf{w}, \tau) \right] d\tau, \quad (\text{A4})$$

$$M(\mathbf{w}, t) = \int_0^t \sum_{i=1}^2 [\Lambda_{i1}(\mathbf{w}, \tau) \Lambda_{i2}(\mathbf{w}, \tau)] d\tau. \quad (\text{A5})$$

For a detailed theoretical derivation of Eq. (A1), see Ref. [2]. Given velocity data  $\mathbf{u}$  on a spatiotemporal grid and  $\mathbf{x}$  on a spatial grid at time 0, we can find the locations  $\mathbf{w}$  to which  $\mathbf{x}$  is taken by the flow at time  $t$ . By tracking the locations along this trajectory and others nearby and using numerical differentiation to estimate gradients, we can compute  $\Lambda(\mathbf{w}, \tau)$  for all relevant times. Numerical integration can then be used to find  $S^2$  in Eq. (A1). By performing this calculation for each  $\mathbf{x}$  on our initial grid, we therefore can compute  $S^2$  as a field.

## APPENDIX B: EXPERIMENTAL DATA ACQUISITION

The experimental quasi-two-dimensional turbulence data are obtained in a laboratory apparatus that we have described in detail previously [14–17]. Briefly, in our apparatus we drive flow in a thin layer of salt water (NaCl in deionized water, 16% by mass), with lateral dimensions of  $86 \times 86 \text{ cm}^2$  and a depth of roughly 5 mm. A second layer of pure water, also about 5 mm in depth, floats on top of the salt water. A square lattice of dipole magnets with their poles oriented vertically lies below the fluid. The magnets are spaced by  $L_m = 2.54 \text{ cm}$ , and are arranged in stripes of alternating polarity. When we drive a d.c. electric current horizontally through the NaCl solution, a Lorentz body force is generated that drives a flow that lies almost entirely in the plane. To measure the flow, we seed the fluid with  $51\text{-}\mu\text{m}$  fluorescent polystyrene particles and track their motion with a 4 megapixel digital camera at a rate of 60 frames per second. The camera only images the central  $31 \times 23 \text{ cm}^2$  of the flow so that any effects of the lateral boundary of the apparatus are negligible in the measurement region. We track about 35 000 particles per frame, with a typical spacing of 0.14 cm, setting the spatial resolution of the measurement. The fields are then projected onto a streamfunction eigenbasis to ensure two-dimensionality [14]. We define a Reynolds number as  $\text{Re} = UL_m/\nu$ , where  $U$  is the in-plane root mean square velocity and  $\nu = 0.0124 \text{ cm}^2 \text{ s}^{-1}$  is the kinematic viscosity. For the data used here, the Reynolds number is roughly 200. Lagrangian trajectories are computed from these gridded velocity fields using second-order Runge-Kutta integration in time [5] and cubic spatial interpolation of the discrete velocity fields.

## APPENDIX C: OCEAN DATA SET

Surface velocity data for the Gulf of Mexico was taken from the Global Ocean Gridded L4 Sea Surface Heights and Derived Variables NRT product provided by the Copernicus Marine Environment Monitoring Service [18]. Using data from satellite-based sea surface altimetry, this data set contains gridded surface velocity estimates with a spatial resolution of  $0.25^\circ \times 0.25^\circ$  and a temporal resolution of one day. Lagrangian trajectories are again computed by Runge-Kutta integration, and cubic interpolation is again used spatially.

- 
- [1] E. Lorenz, Deterministic nonperiodic flow, *J. Atmos. Sci.* **20**, 130 (1963).
- [2] S. Balasuriya, SIAM Review (to be published).
- [3] H. Aref, J. R. Blake, M. Budisic, S. S. S. Cardoso, J. H. E. Cartwright, H. J. H. Clercx, K. El Omari, U. Feudel, R. Golestanian, E. Guillard *et al.*, *Frontiers of chaotic advection, Rev. Mod. Phys.* **89**, 025007 (2017).
- [4] S. Balasuriya, N. T. Ouellette, and I. I. Rypina, Generalized Lagrangian coherent structures, *Physica D* **372**, 31 (2018).
- [5] N. T. Ouellette, P. J. J. O'Malley, and J. P. Gollub, Transport of Finite-Sized Particles in Chaotic Flow, *Phys. Rev. Lett.* **101**, 174504 (2008).
- [6] S. C. Shadden, F. Lekien, and J. E. Marsden, Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows, *Phys. D* **212**, 271 (2005).
- [7] G. Haller, Lagrangian coherent structures, *Annu. Rev. Fluid Mech.* **47**, 137 (2015).
- [8] A. Hadjighasem, M. Farazmand, D. Blazeviski, G. Froyland, and G. Haller, A critical comparison of Lagrangian methods for coherent structure detection, *Chaos* **27**, 053104 (2017).
- [9] M. J. Olascoaga *et al.*, Drifter motion in the Gulf of Mexico constrained by altimetric Lagrangian coherent structures, *Geophys. Res. Lett.* **40**, 6171 (2013).



- [10] M. J. Olascoaga, I. I. Rypina, M. G. Brown, F. J. Beron-Vera, H. Koçak, L. E. Brand, G. R. Halliwell, and L. K. Shay, Persistent transport barrier on the West Florida Shelf, *Geophys. Res. Lett.* **33**, L22603 (2006).
- [11] M. J. Olascoaga, Isolation on the West Florida Shelf with implications for red tides and pollutant dispersal in the Gulf of Mexico, *Nonlin. Proc. Geophys.* **17**, 685 (2010).
- [12] F. Andrade-Canto, J. Sheinbaum, and L. Zavala Sansón, A Lagrangian approach to the loop current eddy separation, *Nonlin. Proc. Geophys.* **20**, 85 (2013).
- [13] M. J. Olascoaga and G. Haller, Forecasting sudden changes in environmental pollution patterns, *Proc. Natl. Acad. Sci. USA* **109**, 4738 (2012).
- [14] D. H. Kelley and N. T. Ouellette, Onset of three-dimensionality in electromagnetically forced thin-layer flows, *Phys. Fluids* **23**, 045103 (2011).
- [15] Y. Liao and N. T. Ouellette, Spatial structure of spectral transport in two-dimensional flow, *J. Fluid Mech.* **725**, 281 (2013).
- [16] L. Fang and N. T. Ouellette, Multiple stages of decay in two-dimensional turbulence, *Phys. Fluids* **29**, 111105 (2017).
- [17] L. Fang and N. T. Ouellette, Influence of lateral boundaries on transport in quasi-two-dimensional flow, *Chaos* **28**, 023113 (2018).
- [18] [http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com\\_csw&view=details&product\\_id=SEALEVEL\\_GLO\\_PHY\\_L4\\_NRT\\_OBSERVATIONS\\_008\\_046](http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=SEALEVEL_GLO_PHY_L4_NRT_OBSERVATIONS_008_046).