
Gaussian processes at the Helm(holtz): A better way to model ocean currents

Renato Berlinghieri¹ Brian Trippe¹ Ryan Giordano¹ Kaushik Srinivasan³
Tamay Özgökmen² Junfei Xia² Tamara Broderick¹
¹Massachusetts Institute of Technology ²University of Miami ³UCLA

Abstract

Understanding the behavior of ocean currents has important practical applications. Since we expect current dynamics to be smooth but highly non-linear, Gaussian processes (GPs) offer an attractive model. In particular, one existing approach is to consider the velocities of the buoys as sparse observations of a vector field in two spatial dimensions and one time dimension. But we show that applying a GP, e.g. with a standard square exponential kernel, directly to this data fails to capture real-life current structure, such as continuity of currents and the shape of vortices. By contrast, these physical properties are captured by divergence and curl-free components of a vector field obtained through a Helmholtz decomposition. So we propose instead to model these components with a GP directly. We show that, because this decomposition relates to the original vector field just via mixed partial derivatives, we can still perform inference given the original data with only a small constant multiple of additional computational expense. We illustrate our method on real oceans data.

1 Introduction

Accurately capturing the behavior of ocean currents has the potential to improve ecosystem management, forecasting of oil spill dispersion, and general understanding of how matter traverses the ocean. To this end, ocean researchers release GPS-tagged buoys in the ocean – and track the corresponding drifter traces, varying across time and space. We wish to predict currents, i.e., the ocean’s velocity vector field, away from the observed drifters. A priori, we expect currents to vary smoothly but in a highly nonlinear fashion. Thus, multi-output Gaussian processes (GPs) with standard spatiotemporal kernels form a natural model class. Indeed, Gonçalves et al. [2019] and Lodise et al. [2020] take exactly such an approach; they consider the velocities of the buoys as sparse observations of a vector field describing ocean dynamics in two spatial dimensions and one time dimension.

However, we show that naively applying a multi-output GP with a standard square exponential kernel directly to drifter trace data fails to capture real-life current structure. See Figure 1a. For simplicity, we start by assuming that the currents are stationary across time and focus on prediction across space. Red dots represent observed locations from 60 buoys in the Gulf of Mexico over 4 days (measured hourly). The horizontal and vertical axes represent, respectively, latitude and longitude. Black arrows represent posterior means of the current. The lower plots zoom in on parts of the upper plot. Despite evidence of a strong current moving left to right in the blue box, the naive approach shows an abrupt drop in current away from observed data. Moreover, current behavior around observations suggests a vortex should be present in the pink box, but the naive approach shows none. The naive approach captures current behavior only very near the observed drifters.

The aim of this work is to extrapolate the behavior of currents more realistically. To that end, we propose to incorporate fluid dynamics information directly into our model via a *Helmholtz decomposition* [Bhatia et al., 2013, Arfken and Weber, 1999]. In particular, we know that the motion of

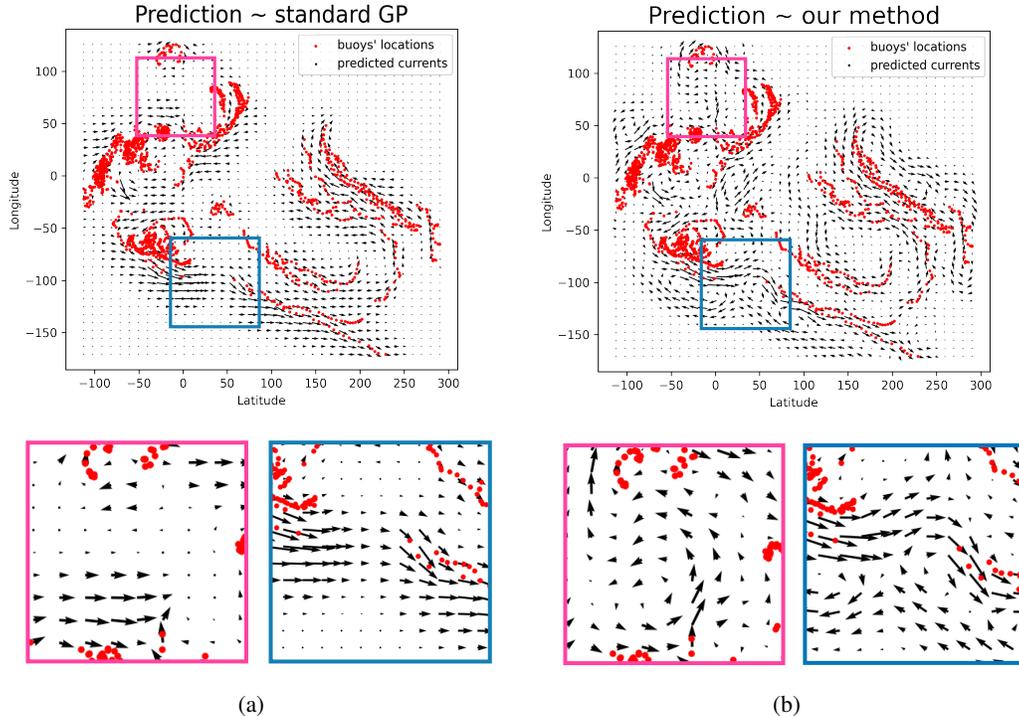


Figure 1: (a) Empirical results using the naive approach. *Top*: predictive means (black arrows) for a grid of test points, given training observations (red dots). *Bottom*: zoom on areas with predicted abrupt current change (blue) and missing a potential vortex (pink). (b) Empirical results using our new Helmholtz approach predict more realistic current behavior.

a volume element of a continuous fluid media consists of (i) expansion or contraction in three orthogonal directions, (ii) rotation about an instantaneous axis, and (iii) translation. Oceanographers characterize these movements using two quantities: *divergence*, measuring expansion/contraction and translation, and *vorticity*, measuring rotation. The key new idea in our work relative to the naive approach is that we model the *Helmholtz components* as smoothly varying according to standard spatial GPs, rather than applying standard spatial GPs directly to the current vector-field components.

We show that our resulting model is amenable to practical inference. Since (i) the Helmholtz decomposition is based on partial (mixed) derivatives, and (ii) the derivative of a GP is a GP, we show that our prior choice implies a GP prior on the current itself. Therefore, we can still perform inference given the original data with no extra approximation. Moreover, our method increases computational expense relative to the naive approach only by a (small) constant multiple.

Related work. The Helmholtz decomposition has been used extensively in the oceanographic literature as a way to interpret and visualize currents' features when the field is observed on a grid [Rocha et al., 2016, Zhang et al., 2018, 2019, Han and Huang, 2020, Bühler et al., 2014, Caballero et al., 2020]. These works decompose the observed vector field in the two Helmholtz components (curl- and divergence-free, respectively) but do not solve any prediction task. In the GP literature, it is well known that specific covariance functions can be used to capture curl- and divergence-free vector fields [Alvarez et al., 2012, Fuselier Jr, 2007, Lowitzsch, 2002]; moreover, a line of research related to electromagnetism [Wahlström et al., 2013, Wahlström, 2015, Solin et al., 2018] uses curl- and divergence-free kernels to try to reconstruct the overall magnetization through Maxwell's equations, similarly to what we are doing for ocean currents. However, none of the aforementioned work can be directly applied to our setting, either because the observation model is different, or because curl- and divergence-free fields are not considered together, as we are doing instead with the Helmholtz decomposition. Finally, recent work from the deep learning community [Greydanus and Sosanya, 2022] extended Hamiltonian Neural Networks [Greydanus et al., 2019] to model both

curl- and divergence-free dynamics simultaneously. Although the prediction problem is similar, this method relies on having low-resolution data available on a dense grid, which is not available for ocean drifters. We refer the reader to Appendix A for a more detailed discussion of related literature.

2 Background

Gaussian Processes. We briefly introduce our GP notation before proceeding. We consider a GP prior on a function f defined on the covariate space \mathbb{R}^2 ; the prior is characterized by a mean function $m : \mathbb{R}^2 \rightarrow \mathbb{R}$ and a positive definite covariance function $k : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$. We write $f \sim \text{GP}(m, k)$.

In spatial problems, analysts often assume smooth variation across space and may choose a squared-exponential (SE) kernel to capture these assumptions:

$$k(\mathbf{x}_m, \mathbf{x}_{m'}) = \sigma_f^2 \cdot \exp\left(-\frac{\|\mathbf{x}_m - \mathbf{x}_{m'}\|_2^2/2}{\ell^2}\right) + \sigma_n^2 \delta_{mm'}$$

where $\mathbf{x}_m, \mathbf{x}_{m'}$ represent two spatial locations in \mathbb{R}^2 , the signal variance σ_f^2 determines the variation of function values from their mean, ℓ controls the lengthscale on which the function varies, σ_n^2 captures noise around the trend, and $\|\cdot\|$ denotes the Euclidean norm.

Gonçalves et al. [2019], Lodise et al. [2020] model the horizontal and vertical velocity components of current with two independent SE GP priors, but this model fails to capture physical properties of interest (Figure 1a). We propose that, when modeling ocean currents, it is more natural to put SE GP priors on meaningful physical components of the ocean flow, rather than the two spatial directions directly. We describe these components next.

Ocean flows and the Helmholtz decomposition. The motion of a volume element of a continuous fluid media, e.g., the ocean flow, is characterized by two quantities: the *divergence*, measuring expansions, contractions, and translations, and the *vorticity*, measuring rotations around an instantaneous axis. Both quantities are relevant for oceanographic studies. Divergence, for example, characterizes fronts – small structures with spatial scales of $\mathcal{O}(0.1 - 10)$ km and temporal scales of $\mathcal{O}(1 - 100)$ h. These are associated with strong vertical motions comprised of a narrow and intense downwelling (flow into the ocean from the surface) and broad, diffuse upwelling (flow from depths to the surface). The strong downwelling regions play a crucial role in air-sea fluxes (including uptake of gases into the ocean) and for biological productivity, since floating particles in the ocean (that include plankton and algae) are concentrated at these fronts. On the other hand, rotation characterizes eddies, larger structures that usually evolve over a long timescale. These account for kinetic energy in the ocean, which makes them a crucial part of global balances of energy, momentum, heat, salt, and chemical constituents (such as carbon dioxide).

From a mathematical perspective, one can observe that since the divergence is an irrotational field, its curl is zero (curl-free), and can thus be represented as the gradient of a scalar potential function, $\nabla\Phi$. On the other hand, since the rotation is an incompressible field, its divergence is zero (divergence-free), and can thus be represented as the vorticity operator of another scalar potential function, $\nabla \times \Psi$. See Appendix B for more details. The Helmholtz decomposition puts together these two observations in a formal mathematical theorem.

Theorem 1 (Helmholtz decomposition). *A smooth vector field $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ can be expressed as the sum of the gradient of a scalar potential and the vorticity operator of another scalar potential,*

$$\underbrace{F}_{\text{ocean flow}} = \underbrace{\nabla\Phi}_{\text{divergence}} + \underbrace{\nabla \times \Psi}_{\text{vorticity}}$$

where

$$\nabla\Phi(\mathbf{x}) := \begin{bmatrix} \partial\Phi(\mathbf{x})/\partial x^{(1)} \\ \partial\Phi(\mathbf{x})/\partial x^{(2)} \end{bmatrix} \quad \nabla \times \Psi(\mathbf{x}) := \begin{bmatrix} \partial\Psi(\mathbf{x})/\partial x^{(2)} \\ -\partial\Psi(\mathbf{x})/\partial x^{(1)} \end{bmatrix}.$$

We provide some visual intuition about Theorem 1 in Appendix E.

In this work, we aim to estimate Φ and Ψ from observed data – and use the estimated components to reconstruct the vector field F . Our approach has an additional advantage: since Φ and Ψ are scalar potentials, they are invariant under changes of coordinate system. As a consequence, when we put a prior on them, the prior is also invariant under these changes. On the contrary, since the velocity field is a vector field, this invariance does not hold for the standard approach.

3 Our method: Gaussian processes at the Helm(holtz)

We consider a dataset of M observations, $\{\mathbf{x}_m, \mathbf{y}_m\}_{m=1}^M$, where $\mathbf{x}_m = (x_m^{(1)}, x_m^{(2)})$ represents the spatial location of a buoy, and $\mathbf{y}_m = (y_m^{(1)}, y_m^{(2)})$ the corresponding horizontal and vertical velocities (the so-called drifter trace). We treat these \mathbf{y} 's as sparse noisy observations of a 2D vector field, $F: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, mapping spatial locations into horizontal and vertical velocities, $(F^{(1)}, F^{(2)})$. We model the drifter traces as follows, with observation noise σ_{obs}^2 :

$$\begin{bmatrix} y_m^{(1)} \\ y_m^{(2)} \end{bmatrix} | F(\mathbf{x}_m) \stackrel{\text{ind}}{\sim} \mathcal{N}\left(F(\mathbf{x}_m), \begin{bmatrix} \sigma_{\text{obs}}^2 & 0 \\ 0 & \sigma_{\text{obs}}^2 \end{bmatrix}\right) \quad \text{independently across } m.$$

By the Helmholtz decomposition theorem, we can rewrite $F(\mathbf{x}_m) = \nabla\Phi(\mathbf{x}_m) + \nabla \times \Psi(\mathbf{x}_m)$, for each \mathbf{x}_m . To exploit this decomposition, we put independent GP priors with squared-exponential kernel on each of the two Helmholtz components:

$$\Phi \sim \text{GP}(0, K_\Phi) \quad \text{and} \quad \Psi \sim \text{GP}(0, K_\Psi). \quad (1)$$

A principal contribution of our work is to show that, with this prior choice in the Helmholtz-transformed space, we are still able to make predictions over the original space of interest. We further show that our method incurs only a small constant multiple of additional computation.

Proposition 1. *Let F be the ocean current vector field and Φ and Ψ the divergence and vorticity components of an Helmholtz decomposition of the field, respectively. The prior choice in Equation (1) implies that the prior on the original vector field F is a GP as well, with mean 0 and kernel $\nabla K_\Phi + \nabla K_\Psi$, as in Appendix C.*

Intuitively, the result holds since (i) the Helmholtz decomposition is based on partial (mixed) derivatives and (ii) the derivative of a GP is a GP. In particular, in deriving the covariance of the induced GP on F , we use Equation (9.1) in Rasmussen and Williams [2005]. This transformation induces a small constant multiple of additional computational expense compared to the direct approach to drifter trace data because the gradient operators “double” the size of the covariance matrices involved. Roughly speaking, if the standard approach computational cost is N , our Helmholtz approach cost would be $8 \cdot N \simeq \mathcal{O}(N)$.

4 Experimental results

We now check the empirical performance of putting GP priors directly on the Helmholtz components. We use the GLAD dataset [Özgökmen, 2012]: 297 trajectories from near-surface CODE-type ocean drifters (drogued at a depth of one meter) tracked in real-time using SPOT GPS units, launched in the northern Gulf of Mexico near DeSoto Canyon in July 2012 as part of the CARTHE Grand Lagrangian Deployment (GLAD) experiment.

For ease of analysis, we take drifter traces of 60 of these buoys, observed hourly over a 4-day time horizon, and we downsample these traces by a factor 12, obtaining 1906 observations. As before, we collapse the time dimension and focus on the spatial inference. With our method, we have five hyperparameters to choose: σ_Φ and ℓ_Φ for the divergence kernel, σ_Ψ and ℓ_Ψ for the vorticity kernel, and the observation noise σ^2 . To optimize these, we perform a two-stage grid search with marginal log-likelihood as objective. This is inspired by oceanographic prior knowledge: usually the divergence component is substantially less strong (one or two order of magnitude) than the vorticity, whereas the respective lengthscales should be more similar. See Appendix D for additional details.

Our results appear in Figure 1b. The black arrows show the current posterior means at test locations with this alternative kernel choice. The test locations are 1600 points evenly sparsed on a 40 x 40 grid that covers the full range of latitude and longitude of our buoys’ observations. Strengths of our approach are illustrated in the two colored boxes. In the blue box, our method predicts a strong current connecting across drifters, in accordance with the known continuity of currents (the idea that when a fluid is in motion, it must move in such a way that mass is conserved). Furthermore, in the pink box we now see the expected vortex between the two lines of current. Finally, note that fitting our new model takes approximately 3.10 seconds (0.75 s for building the kernel matrix); fitting the standard model takes about 2.66 s (0.4 s for the kernel).

Discussion and future work. Challenges remain. For instance, Figure 1b demonstrates that our model has an (unrealistic) tendency to close vortices within a short space horizon. We expect that further model development can address this issue – e.g., by incorporating multiple lengthscales (an idea supported by oceanographic prior knowledge) or even summing over both types of kernels. Another extension of interest is allowing currents to vary in time.

References

- M. A. Alvarez, L. Rosasco, N. D. Lawrence, et al. Kernels for vector-valued functions: A review. *Foundations and Trends® in Machine Learning*, 4(3):195–266, 2012.
- G. B. Arfken and H. J. Weber. *Mathematical methods for physicists*, 1999.
- M. Berta, A. Griffa, M. G. Magaldi, T. M. Özgökmen, A. C. Poje, A. C. Haza, and M. J. Olascoaga. Improved surface velocity and trajectory estimates in the gulf of mexico from blended satellite altimetry and drifter data. *Journal of Atmospheric and Oceanic Technology*, 32(10):1880–1901, 2015.
- H. Bhatia, G. Norgard, V. Pascucci, and P.-T. Bremer. The helmholtz-hodge decomposition—a survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(8):1386–1404, 2013.
- O. Bühler, J. Callies, and R. Ferrari. Wave–vortex decomposition of one-dimensional ship-track data. *Journal of Fluid Mechanics*, 756:1007–1026, 2014.
- A. Caballero, S. Mulet, N. Ayoub, I. Manso-Narvarde, X. Davila, C. Boone, F. Toublanc, and A. Rubio. Integration of hf radar observations for an enhanced coastal mean dynamic topography. *Frontiers in Marine Science*, page 1005, 2020.
- C. P. Chavanne and P. Klein. Can oceanic submesoscale processes be observed with satellite altimetry? *Geophysical Research Letters*, 37(22), 2010.
- G. Esposito, M. Berta, L. Centurioni, T. S. Johnston, J. Lodise, T. Özgökmen, P.-M. Poulain, and A. Griffa. Submesoscale vorticity and divergence in the alboran sea: Scale and depth dependence. *Frontiers in Marine Science*, 8:678304, 2021.
- L.-L. Fu and R. Ferrari. Observing oceanic submesoscale processes from space. *Eos, Transactions American Geophysical Union*, 89(48):488–488, 2008.
- L.-L. Fu, D. B. Chelton, P.-Y. Le Traon, and R. Morrow. Eddy dynamics from satellite altimetry. *Oceanography*, 23(4):14–25, 2010.
- E. J. Fuselier Jr. *Refined error estimates for matrix-valued radial basis functions*. PhD thesis, Texas A&M University, 2007.
- R. C. Gonçalves, M. Iskandarani, T. Özgökmen, and W. C. Thacker. Reconstruction of submesoscale velocity field from surface drifters. *Journal of Physical Oceanography*, 49(4), 2019.
- S. Greydanus and A. Sosanya. Dissipative hamiltonian neural networks: Learning dissipative and conservative dynamics separately. *arXiv preprint arXiv:2201.10085*, 2022.
- S. Greydanus, M. Dzamba, and J. Yosinski. Hamiltonian neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.
- L. Han and R. X. Huang. Using the helmholtz decomposition to define the indian ocean meridional overturning streamfunction. *Journal of Physical Oceanography*, 50(3), 2020.
- A. C. Haza, E. D’Asaro, H. Chang, S. Chen, M. Curcic, C. Guigand, H. S. Huntley, G. Jacobs, G. Novelli, T. M. Özgökmen, A. C. Poje, E. Ryan, and A. Shcherbina. Drogue-loss detection for surface drifters during the lagrangian submesoscale experiment (laser). *Journal of Atmospheric and Oceanic Technology*, 35(4), 2018.
- J. Lodise, T. Özgökmen, R. C. Gonçalves, M. Iskandarani, B. Lund, J. Horstmann, P.-M. Poulain, J. Klymak, E. H. Ryan, and C. Guigand. Investigating the formation of submesoscale structures along mesoscale fronts and estimating kinematic quantities using lagrangian drifters. *Fluids*, 5(3), 2020.
- S. Lowitzsch. *Approximation and interpolation employing divergence-free radial basis functions with applications*. Texas A&M University, 2002.
- I. Macêdo and R. Castro. *Learning divergence-free and curl-free vector fields with matrix-valued kernels*. IMPA, 2010.

- A. Mariano, E. Ryan, H. Huntley, L. Laurindo, E. Coelho, A. Griffa, T. Özgökmen, M. Berta, D. Bogucki, S. Chen, et al. Statistical properties of the surface velocity field in the northern gulf of mexico sampled by glid drifters. *Journal of Geophysical Research: Oceans*, 121(7):5193–5216, 2016.
- G. Novelli, C. M. Guigand, C. Cousin, E. H. Ryan, N. J. Laxague, H. Dai, B. K. Haus, and T. M. Özgökmen. A biodegradable surface drifter for ocean sampling on a massive scale. *Journal of Atmospheric and Oceanic Technology*, 34(11):2509–2532, 2017.
- T. Özgökmen. Carthe: Glid experiment code-style drifter trajectories (lowpass filtered, 15 minute interval records), northern gulf of mexico near desoto canyon, july-october 2012. *Gulf of Mexico Research Initiative*, 10:N7VD6WC8, 2012.
- A. C. Poje, T. M. Özgökmen, B. L. Lipphardt, B. K. Haus, E. H. Ryan, A. C. Haza, G. A. Jacobs, A. Reniers, M. J. Olascoaga, G. Novelli, et al. Submesoscale dispersion in the vicinity of the deepwater horizon spill. *Proceedings of the National Academy of Sciences*, 111(35):12693–12698, 2014.
- C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2005.
- C. B. Rocha, T. K. Chereskin, S. T. Gille, and D. Menemenlis. Mesoscale to submesoscale wavenumber spectra in drake passage. *Journal of Physical Oceanography*, 46(2):601–620, 2016.
- A. Solin, M. Kok, N. Wahlström, T. B. Schön, and S. Särkkä. Modeling and interpolation of the ambient magnetic field by gaussian processes. *IEEE Transactions on robotics*, 34(4):1112–1127, 2018.
- N. Wahlström. *Modeling of magnetic fields and extended objects for localization applications*. PhD thesis, Linköping University Electronic Press, 2015.
- N. Wahlström, M. Kok, T. B. Schön, and F. Gustafsson. Modeling magnetic fields using gaussian processes. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3522–3526. IEEE, 2013.
- C. Zhang, H. Wei, Z. Liu, and X. Fu. Characteristic ocean flow visualization using helmholtz decomposition. In *2018 Oceans-MTS/IEEE Kobe Techno-Oceans (OTO)*, pages 1–4. IEEE, 2018.
- C. Zhang, H. Wei, C. Bi, and Z. Liu. Helmholtz–hodge decomposition-based 2d and 3d ocean surface current visualization for mesoscale eddy detection. *Journal of Visualization*, 22, 01 2019.

A Related work

Recent works have implemented GP regression to learn characteristic features of the oceans from drifter data. In Gonçalves et al. [2019], the authors model drifter traces with a GP with squared-exponential covariance function characterizing how the correlation in the direction and magnitude of the currents decay with latitude, longitude, and time, at two different scales. Lodise et al. [2020] further develop, apply, and validate this approach, by exploring three different variations of the core GP regression method. In both of these works, the components of the 2D velocity field are separately modeled as independent GPs, and we show that this leads to a poor reconstruction of some important ocean’s features, e.g., eddies. In this work, we overcome this problem by performing the GP regression in the latent space obtained through an Helmholtz decomposition.

The Helmholtz decomposition has been used extensively in the oceanographic literature as a way to interpret and visualize currents’ features when the field is observed on a grid [Rocha et al., 2016, Zhang et al., 2018, 2019, Han and Huang, 2020, Bühler et al., 2014, Caballero et al., 2020]. These authors decompose ocean currents into vortical (divergence-free) and divergent (curl-free) components of the flow, and different characteristic ocean features are recognized with different components: eddies, which are rotational, are only present in the vortical component; fronts, being irrotational, are only present in the divergent one. These works, however, rely mostly on measurements from satellites, which unfortunately have spatial and temporal resolution that is too coarse to precisely capture most of the features of interest [Fu and Ferrari, 2008, Chavanne and Klein, 2010, Fu et al., 2010]. Moreover, they assume data are available on a dense grid, which is not the case for real ocean current data. As a consequence, to more accurately capture behaviors about ocean features at this resolution, researchers release GPS-tagged buoys in the ocean – and track the corresponding surface drifter traces, varying across time and space [Poje et al., 2014, Berta et al., 2015, Mariano

et al., 2016, Novelli et al., 2017, Haza et al., 2018]. Since many patches of ocean end up having no observations, there is inherent modeling uncertainty that must be confronted, and the problem gets challenging. In particular, it is not obvious how to simultaneously interpolate and extrapolate away from drifter data, with an appropriate measure of uncertainty, while still leveraging the Helmholtz decomposition to learn structures of ocean currents.

In the GP literature, it is well known that specific covariance functions can be used to capture curl- and divergence-free vector fields [Alvarez et al., 2012, Fuselier Jr, 2007, Lowitzsch, 2002]. But existing work does not directly apply in the setting at hand. In Macêdo and Castro [2010], the authors model fluids with matrix valued radial-basis kernels, built for curl- and divergence-free vector fields. However, no guidance is provided on what to do when there are prior beliefs, or making inference about the underlying potential functions. Moreover, their predictions do not comprise a measure of uncertainty. The same kernels are used in modeling magnetic fields, where Maxwell’s equations give intuitive explanation of the two components [Wahlström et al., 2013, Wahlström, 2015]. These works, though, use a different observation model (the vector field of interest is not observed), focus on 3D and not 2D, and model the curl- and divergence-free components directly, rather than the scalar fields that are then passed through the gradient operators. This is done in Solin et al. [2018], where the magnetic field components are modeled jointly by imposing a GP prior to the latent scalar potential of the magnetic field. Their results confirm the intuition that modeling directly the latent potentials has some advantages, but cannot be applied to our setting, where we decompose the ocean vector field into the sum of a curl- and a divergence-free component.

Finally, a recent work from the deep learning community [Greydanus and Sosanya, 2022] extended Hamiltonian Neural Networks [Greydanus et al., 2019] in such a way that, the authors suggest, allows to model both curl- and divergence-free dynamics simultaneously, for example for reconstructing surface flows from a noisy ocean current dataset. Although the prediction problem is similar, this method does not allow to incorporate prior beliefs, which instead is a key feature for the hyperparameter tuning of our model. Furthermore these authors only attempt their approach on low-resolution data which is already available on a low-spatial resolution grid, and it is unclear how this method applies to ocean drifter data at the resolution we are interested in. In particular, this method seems to strongly rely on having a dense grid of data, and more work would need to be done to adapt this to our drifter data.

B Divergence, gradient, and curl operators in 2D

Consider a scalar-valued differentiable function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. The *gradient* of f is the vector-valued function ∇f whose value at point \mathbf{x} is the vector whose components are the partial derivatives of f at \mathbf{x} . Formally,

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x^{(1)}} \\ \frac{\partial f(\mathbf{x})}{\partial x^{(2)}} \end{bmatrix} = \hat{i} \frac{\partial f(\mathbf{x})}{\partial x^{(1)}} + \hat{j} \frac{\partial f(\mathbf{x})}{\partial x^{(2)}}$$

where \hat{i} and \hat{j} represents, respectively, the standard unit vectors in the direction of the $x^{(1)}$ and $x^{(2)}$ coordinates. From this rewriting, one can note that taking the gradient of a function is equivalent to taking a vector operator ∇ , called *del*:

$$\nabla = \hat{i} \frac{\partial}{\partial x^{(1)}} + \hat{j} \frac{\partial}{\partial x^{(2)}} = \left(\frac{\partial}{\partial x^{(1)}}, \frac{\partial}{\partial x^{(2)}} \right)$$

Using this operator, two operations on vector fields can be defined.

Definition 1. Let $A \subset \mathbb{R}^2$ be an open subset and let $F : A \rightarrow \mathbb{R}^2$ be a vector field. The *divergence* of F is the scalar function $\text{div}F : A \rightarrow \mathbb{R}$, defined by

$$\text{div}F(\mathbf{x}) := (\nabla \cdot F)(\mathbf{x}) = \frac{\partial F^{(1)}}{\partial x^{(1)}} + \frac{\partial F^{(2)}}{\partial x^{(2)}}$$

Definition 2. Let $A \subset \mathbb{R}^2$ be an open subset and let $F : A \rightarrow \mathbb{R}^2$ be a vector field. The *curl* of F is the scalar function $\text{curl}F : A \rightarrow \mathbb{R}$, defined by

$$\text{curl}F(\mathbf{x}) := \frac{\partial F^{(1)}}{\partial x^{(2)}} - \frac{\partial F^{(2)}}{\partial x^{(1)}}$$

Note that this curl definition follows directly from the definition of curl in three dimensions, where this quantity describes infinitesimal circulation.

In the 3D world, curl and divergence are enough to characterize the Helmholtz decomposition. For the 2D version, however, we need to characterize an additional operator - which we call *vorticity* operator - that plays the role of the standard curl operator in the 3D version.

Definition 3. Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a scalar field. The *vorticity* of f is the vector field $\nabla \times f$, defined by

$$\nabla \times f = \begin{bmatrix} \frac{\partial f}{\partial x^{(2)}} \\ -\frac{\partial f}{\partial x^{(1)}} \end{bmatrix} = \hat{i} \frac{\partial f}{\partial x^{(2)}} - \hat{j} \frac{\partial f}{\partial x^{(1)}}$$

where \hat{i} and \hat{j} represents, respectively, the standard unit vectors in the direction of the $x^{(1)}$ and $x^{(2)}$ coordinates.

Note that we use the notation $\nabla \times$ to draw the parallel with the curl operator in three dimension. The ideas of gradient, divergence, vorticity, and curl lead to the following characterization of vector fields.

Definition 4. A vector field $F : A \rightarrow \mathbb{R}^2$ is called *rotation-free* (or curl-free) if the curl is zero, $\text{curl} F = 0$, and it is called *incompressible* (or divergence-free) if the divergence is zero, $\text{div} F = 0$.

Proposition 2. Let f be a scalar field and \mathcal{C}^2 the class of functions whose second derivatives exist and are continuous.

1. If f is \mathcal{C}^2 , then $\text{curl}(\nabla f) = 0$. Every gradient of a scalar field is rotation free.
2. If f is \mathcal{C}^2 , then $\text{div}(\nabla \times f) = 0$. Every vorticity of a scalar field is incompressible.

Proof. For (1), we have the following:

$$\text{curl}(\nabla f) = \text{curl} \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x^{(1)}} \\ \frac{\partial f(\mathbf{x})}{\partial x^{(2)}} \end{bmatrix} = \frac{\partial f(\mathbf{x})/\partial x^{(1)}}{\partial x^{(2)}} - \frac{\partial f(\mathbf{x})/\partial x^{(2)}}{\partial x^{(1)}} = 0.$$

For (2):

$$\text{div}(\nabla \times f) = \text{div} \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x^{(2)}} \\ -\frac{\partial f(\mathbf{x})}{\partial x^{(1)}} \end{bmatrix} = \frac{\partial f(\mathbf{x})/\partial x^{(2)}}{\partial x^{(1)}} + \frac{-\partial f(\mathbf{x})/\partial x^{(1)}}{\partial x^{(2)}} = 0.$$

□

For more material on vector calculus, we refer the reader to Arfken and Weber [1999].

C Proof Proposition 1

First, we state a more complete version of Proposition 1, that we are going to prove.

Proposition. Let F be the ocean current vector field and Φ and Ψ the divergence and vorticity components of an Helmholtz decomposition of the field, respectively. If

$$\Phi \sim GP(0, K_\Phi) \quad \Psi \sim GP(0, K_\Psi).$$

then the prior on the original vector field F is a GP as well,

$$F \sim GP(0, \nabla K_\Phi + \nabla K_\Psi)$$

where, for each pair of datapoints $\mathbf{x}_m, \mathbf{x}_{m'}$,

$$\nabla K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'}) := \begin{bmatrix} \frac{\partial^2 K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(1)} \partial x_{m'}^{(1)}} & \frac{\partial^2 K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(1)} \partial x_{m'}^{(2)}} \\ \frac{\partial^2 K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(2)} \partial x_{m'}^{(1)}} & \frac{\partial^2 K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(2)} \partial x_{m'}^{(2)}} \end{bmatrix},$$

$$\nabla K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'}) := \begin{bmatrix} \frac{\partial^2 K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(2)} \partial x_{m'}^{(2)}} & \frac{-\partial^2 K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(2)} \partial x_{m'}^{(1)}} \\ -\frac{\partial^2 K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(1)} \partial x_{m'}^{(2)}} & \frac{\partial^2 K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(1)} \partial x_{m'}^{(1)}} \end{bmatrix}.$$

Proof. Recall first of all that since differentiation is a linear operator, the derivative of a Gaussian process is another Gaussian process. This implies that we can make inference based on the joint Gaussian distribution of function values and partial derivatives. In particular, a covariance function $k(\cdot, \cdot)$ on function f on covariate space \mathbb{R}^2 implies the following covariance between (mixed) partial derivatives

$$\text{Cov}\left(\frac{\partial f(\mathbf{x}_m)}{\partial x_m^{(i)}}, \frac{\partial f(\mathbf{x}_{m'})}{\partial x_{m'}^{(i)}}\right) = \frac{\partial^2 k(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(i)} \partial x_{m'}^{(i)}} \quad (2)$$

for $i, j \in 1, 2$. See e.g., paragraph 9.4 in Rasmussen and Williams [2005].

By the Helmholtz decomposition, for each spatial input \mathbf{x}_m ,

$$F(\mathbf{x}_m) = \nabla \Phi(\mathbf{x}_m) + \nabla \times \Psi(\mathbf{x}_m) = \begin{bmatrix} \partial \Phi(\mathbf{x}_m) / \partial x_m^{(1)} + \partial \Psi(\mathbf{x}_m) / \partial x_m^{(2)} \\ \partial \Phi(\mathbf{x}_m) / \partial x_m^{(2)} - \partial \Psi(\mathbf{x}_m) / \partial x_m^{(1)} \end{bmatrix}.$$

Since F is related to Φ and Ψ just by (mixed) partial derivatives, and Φ and Ψ are two independent GPs, by the result on derivatives of GPs we have that F is a GP as well. In particular, since both Φ and Ψ have mean zero, the induced GP on F also has mean zero. For the covariance, for each pair of datapoints $\mathbf{x}_m, \mathbf{x}_{m'}$, we have the following:

$$\begin{aligned} \text{Cov}(F(\mathbf{x}_m), F(\mathbf{x}_{m'})) &= \text{Cov}\left(\begin{bmatrix} \frac{\partial \Phi(\mathbf{x}_m)}{\partial x_m^{(1)}} + \frac{\partial \Psi(\mathbf{x}_m)}{\partial x_m^{(2)}} \\ \frac{\partial \Phi(\mathbf{x}_m)}{\partial x_m^{(2)}} - \frac{\partial \Psi(\mathbf{x}_m)}{\partial x_m^{(1)}} \end{bmatrix}, \begin{bmatrix} \frac{\partial \Phi(\mathbf{x}_{m'})}{\partial x_{m'}^{(1)}} + \frac{\partial \Psi(\mathbf{x}_{m'})}{\partial x_{m'}^{(2)}} \\ \frac{\partial \Phi(\mathbf{x}_{m'})}{\partial x_{m'}^{(2)}} - \frac{\partial \Psi(\mathbf{x}_{m'})}{\partial x_{m'}^{(1)}} \end{bmatrix}\right) \\ &\stackrel{(i)}{=} \begin{bmatrix} \text{Cov}\left(\frac{\partial \Phi(\mathbf{x}_m)}{\partial x_m^{(1)}}, \frac{\partial \Phi(\mathbf{x}_{m'})}{\partial x_{m'}^{(1)}}\right) & \text{Cov}\left(\frac{\partial \Phi(\mathbf{x}_m)}{\partial x_m^{(1)}}, \frac{\partial \Phi(\mathbf{x}_{m'})}{\partial x_{m'}^{(2)}}\right) \\ \text{Cov}\left(\frac{\partial \Phi(\mathbf{x}_m)}{\partial x_m^{(2)}}, \frac{\partial \Phi(\mathbf{x}_{m'})}{\partial x_{m'}^{(1)}}\right) & \text{Cov}\left(\frac{\partial \Phi(\mathbf{x}_m)}{\partial x_m^{(2)}}, \frac{\partial \Phi(\mathbf{x}_{m'})}{\partial x_{m'}^{(2)}}\right) \end{bmatrix} + \\ &\quad \begin{bmatrix} \text{Cov}\left(\frac{\partial \Psi(\mathbf{x}_m)}{\partial x_m^{(2)}}, \frac{\partial \Psi(\mathbf{x}_{m'})}{\partial x_{m'}^{(2)}}\right) & \text{Cov}\left(\frac{\partial \Psi(\mathbf{x}_m)}{\partial x_m^{(2)}}, -\frac{\partial \Psi(\mathbf{x}_{m'})}{\partial x_{m'}^{(1)}}\right) \\ \text{Cov}\left(-\frac{\partial \Psi(\mathbf{x}_m)}{\partial x_m^{(1)}}, \frac{\partial \Psi(\mathbf{x}_{m'})}{\partial x_{m'}^{(2)}}\right) & \text{Cov}\left(-\frac{\partial \Psi(\mathbf{x}_m)}{\partial x_m^{(1)}}, -\frac{\partial \Psi(\mathbf{x}_{m'})}{\partial x_{m'}^{(1)}}\right) \end{bmatrix} \\ &\stackrel{(ii)}{=} \begin{bmatrix} \frac{\partial^2 K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(1)} \partial x_{m'}^{(1)}} & \frac{\partial^2 K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(1)} \partial x_{m'}^{(2)}} \\ \frac{\partial^2 K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(2)} \partial x_{m'}^{(1)}} & \frac{\partial^2 K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(2)} \partial x_{m'}^{(2)}} \end{bmatrix} + \\ &\quad \begin{bmatrix} \frac{\partial^2 K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(2)} \partial x_{m'}^{(2)}} & -\frac{\partial^2 K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(2)} \partial x_{m'}^{(1)}} \\ -\frac{\partial^2 K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(1)} \partial x_{m'}^{(2)}} & \frac{\partial^2 K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'})}{\partial x_m^{(1)} \partial x_{m'}^{(1)}} \end{bmatrix} \\ &:= \nabla K_\Phi(\mathbf{x}_m, \mathbf{x}_{m'}) + \nabla K_\Psi(\mathbf{x}_m, \mathbf{x}_{m'}) \end{aligned}$$

where equality (i) follows from the definition of cross-covariance, independence of Φ and Ψ , and separating the two components' terms, and equality (ii) follows from Equation (2). \square

D Hyperparameters' choice description

In Section 4, we have seen that we have five hyperparameters to choose for fitting our model: σ_Φ and ℓ_Φ for the divergence kernel, σ_Ψ and ℓ_Ψ for the vorticity kernel, and the observation noise σ^2 . The optimization procedure is done performing *two* grid searches inspired by oceanographic prior knowledge [Esposito et al., 2021]: we know that usually the divergence component is substantially less strong than the vorticity, whereas the respective lengthscales should be more similar. Exploiting this information, first we explore a wide range of values for each parameters to understand the general magnitude of the hyperparameters. Then, we do a finer search around the optimal values from

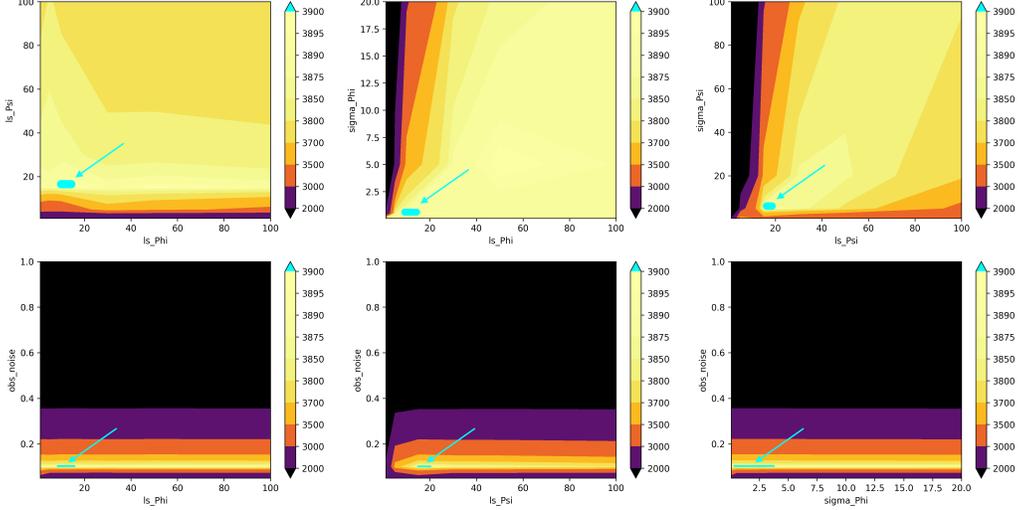


Figure 2: Profile likelihoods plot for the first (coarser) grid search. In each subplot, we highlight in cyan the area where corresponding hyperparameters lead to maximum log marginal likelihood. The six subplots are as follows. *Top-left*: ℓ_Φ vs. ℓ_Ψ ; *Top-center*: ℓ_Φ vs. σ_Φ ; *Top-right*: ℓ_Ψ vs. σ_Ψ ; *Bottom-left*: ℓ_Φ vs. σ^2 ; *Bottom-center*: ℓ_Ψ vs. σ^2 ; *Bottom-right*: σ_Φ vs. σ^2 .

the first search, to actually optimize the hyperparameters' choice. For both the two grid searches, we pick the hyperparameters' combination that maximize the log marginal likelihood.

For the first grid search we have:

- $\ell_\Phi \in [1, 5, 10, 30, 50, 100]$
- $\sigma_\Phi \in [0.05, 0.5, 5, 20]$
- $\ell_\Psi \in [1, 5, 15, 30, 50, 100]$
- $\sigma_\Psi \in [0.5, 5, 20, 100]$
- $\sigma^2 \in [0.05, 0.1, 0.5, 1]$

We plot the results of this search in Figure 2. For each of the six subplots, we pick two parameters (that name the axis, e.g., ℓ_Ψ vs. ℓ_Φ for the top-left), and we contour plot the log marginal likelihood as a function of these two hyperparameters, where the other three are chosen to be the optimal according to our search. In other words, for $\ell_\Phi = 1, \ell_\Psi = 1$, we plot the value $\max_{\sigma_\Phi, \sigma_\Psi, \sigma^2} L(1, \sigma_\Phi, 1, \sigma_\Psi, \sigma^2)$, where L is the log marginal likelihood function, taking as input the five parameters $(\ell_\Phi, \sigma_\Phi, \ell_\Psi, \sigma_\Psi, \sigma^2)$. We call this plot a *profile likelihood* plot, in analogy to profile likelihoods in linear-mixed models, where to optimize likelihood with variable random effects parameters, I need to take into account the values of fixed hyperparameters. The optimal combination, highlighted in cyan in the plot, is $(\ell_\Phi, \sigma_\Phi, \ell_\Psi, \sigma_\Psi, \sigma^2) = (10, 0.5, 15, 5, 0.1)$. This confirms our intuition that the divergence component is less strong than the vorticity, and the respective lengthscales are more similar.

Next, we run the second grid search in a neighborhood of this optimal combination, to refine the hyperparameters' choice. In particular, we have:

- $\ell_\Phi \in [3, 6, 10, 13, 16, 20]$
- $\sigma_\Phi \in [0.2, 0.5, 0.7, 1]$
- $\ell_\Psi \in [8, 12, 15, 20, 24]$
- $\sigma_\Psi \in [4, 5, 6, 7]$
- $\sigma^2 \in [0.09, 0.1, 0.12, 0.15]$

The results of this search are presented in the corresponding profile likelihood plot, Figure 3. In this case, the optimal combination is $(\ell_\Phi, \sigma_\Phi, \ell_\Psi, \sigma_\Psi, \sigma^2) = (6, 0.5, 24, 7, 0.1)$, and this is the hyperparameters' choice that leads to the posterior plot in Figure 1b.

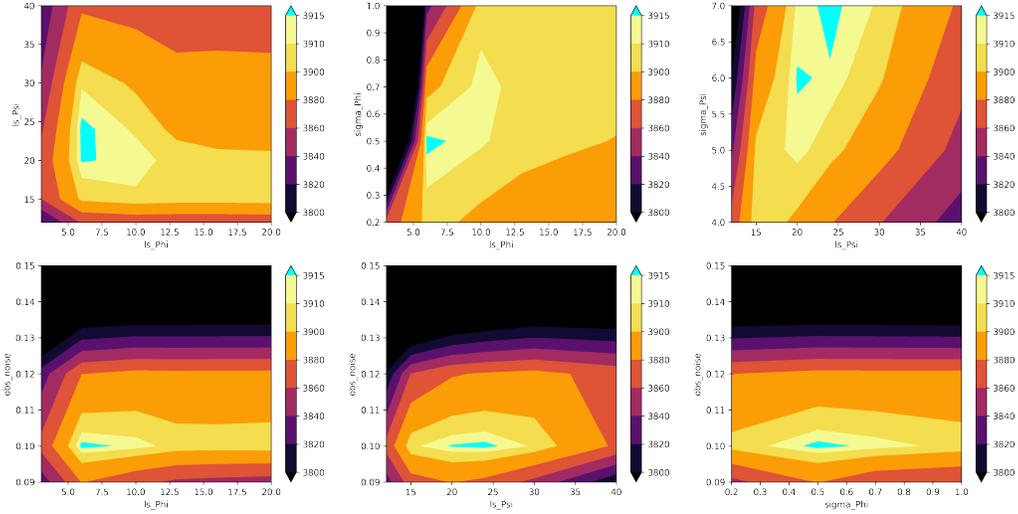


Figure 3: Profile likelihoods for the second (finer) grid search. The six subplots are as in Figure 2. As before, we highlight in cyan the areas where the hyperparameters' choice leads to higher marginal log-likelihood.

E Visualize the Helmholtz decomposition

To provide further intuition on the Helmholtz theorem, we plot a possible decomposition for a vector field F into the divergence $\nabla\Phi$ and vorticity components $\nabla \times \Psi$. In the plot, one can see that from the divergence we can read areas of downwelling/sink (arrows pointing inwards to a single point) and upwelling/source (arrows pointing outwards from a single point). The vorticity, instead, characterizes rotational elements of the vector field, e.g., vortices/eddies in our ocean setting.

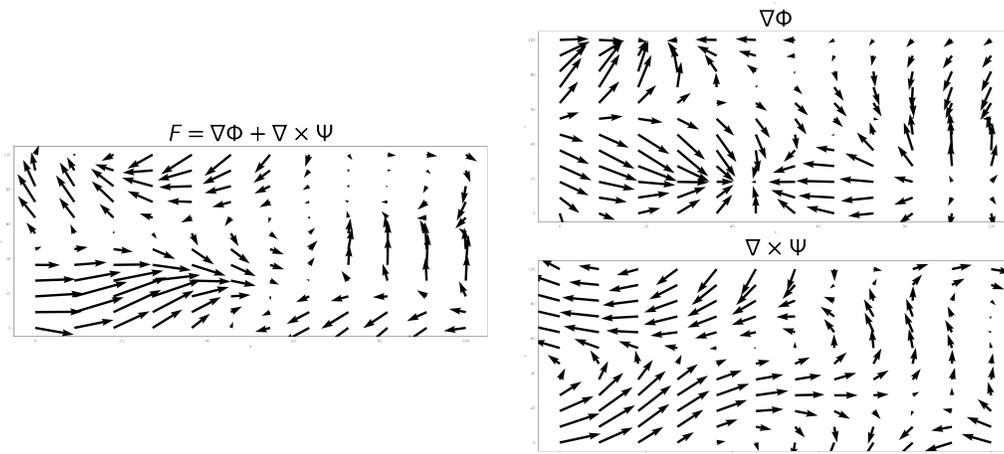


Figure 4: Helmholtz decomposition of vector field F . *Left*: original vector field F . *Top-right*: divergence component of Helmholtz decomposition of F . *Bottom-right*: vorticity component of Helmholtz decomposition of F .