Gaussian Processes at the Helm(holtz): A More Fluid Model for Ocean Currents

Renato Berlinghieri [1] **Brian L. Trippe** [2] **David R. Burt** [1] **Ryan Giordano** [1] **Kaushik Srinivasan** [3]
Tamay Özgökmen [4] **Junfei Xia** [4] **Tamara Broderick** [1]

Abstract

Given sparse observations of buoy velocities, oceanographers are interested in reconstructing ocean currents away from the buoys and identifying divergences in a current vector field. As a first and modular step, we focus on the time-stationary case – for instance, by restricting to short time periods. Since we expect current velocity to be a continuous but highly non-linear function of spatial location, Gaussian processes (GPs) offer an attractive model. But we show that applying a GP with a standard stationary kernel directly to buoy data can struggle at both current reconstruction and divergence identification, due to some physically unrealistic prior assumptions. To better reflect known physical properties of currents, we propose to instead put a standard stationary kernel on the divergence and curl-free components of a vector field obtained through a Helmholtz decomposition. 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 the benefits of our method with theory and experiments on synthetic and real ocean data.

1. Introduction

Ocean currents are key to the global distribution of water, heat, and nutrients. To better understand ocean currents, scientists are interested in two tasks: (1) reconstructing ocean currents at different locations and (2) identifying divergences in the current vector field. Reconstructing ocean currents accurately can facilitate weather forecasting, maritime navigation, and forecasting of oil spill dispersion. And current divergences are important to identify since they are responsible for the transport of biomass, carbon, and nutrients – with implications for ecosystem management, climate, and the fishing industry (D'Asaro et al., 2018). With these tasks in mind, researchers release and track GPS-tagged buoys in the ocean (Özgökmen, 2012; D'Asaro et al., 2017).

It remains to choose an appropriate method to reconstruct currents and their divergences from buoy data. Gonçalves et al. (2019) and Lodise et al. (2020) proposed modeling buoy velocities in the latitude and longitude directions according to independent Gaussian processes (GPs) with standard spatiotemporal kernels (e.g., squared exponential kernels). In our work, we focus on the spatial aspects of this task and assume the velocity field is stationary in time. Even under this simplification, an independent spatial GP prior on the velocities is a natural choice due to its ability to handle the sparsity of buoy observations on the ocean surface and its assumption that currents vary continuously but in a nonlinear fashion. We call such a model the *velocity GP*.

However, in what follows, we show that there remains substantial room for improvement. In simulated cases where we have access to ground truth, we observe that the velocity GP approach can fail to complete vortices or fail to connect currents when buoys are observed sparsely. And while we show how to derive divergence estimates for the velocity GP, we also find that these estimates often fail to capture the true divergence when it is known in simulations or real data.

To address these issues, we propose to more directly model known behaviors from fluid dynamics. Scientists know that the motion of a volume element of a continuous fluid medium in two dimensions consists of (i) expansion or contraction in two orthogonal directions, (ii) rotation about an instantaneous axis, and (iii) translation. A *Helmholtz decomposition* (Bhatia et al., 2013; Arfken & Weber, 1999) from fluid dynamics lets us decompose the vector field of ocean currents into a *divergent* component (or *curl-free*, measuring expansion, contraction, and translation) and a *rotational* component (or *divergence-free*, measuring rotation).[1]

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology [2]Department of Statistics, Columbia University [3]Department of Atmospheric and Oceanic Sciences, UCLA [4]Department of Ocean Sciences, Rosenstiel School of Marine and Atmospheric Science, University of Miami. Correspondence to: Renato Berlinghieri <renb@mit.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

[1]The divergent component is called curl-free and the rotational

By contrast to the standard approach, we model the divergent and rotational components (rather than the velocity components) with independent GP priors. The resulting *Helmholtz GP* prior offers several conceptual advantages. For one, oceanographers expect the two components to have substantially different magnitudes and length scales; it is straightforward to encode these differences with a Helmholtz GP. By contrast, we prove that the velocity GP implies an a priori belief that the divergent and rotational components have the same magnitude. Second, we expect correlation between the longitudinal and latitudinal components of a current, which the Helmholtz GP exhibits – and the velocity GP lacks by construction. Finally, the Helmholtz GP is agnostic to the (arbitrary) choice of reference frame, while the velocity GP is not necessarily.

We demonstrate that the Helmholtz GP 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 tractable GP prior on the current itself. Therefore, we can still perform inference given the original data with no extra approximation. And our method suffers no increase in computational complexity relative to the velocity GP.

Finally, we demonstrate the superior performance of the Helmholtz GP at the current-reconstruction task and divergence-estimation task (as well as vorticity estimation) in a variety of experiments on simulated and real data. Code is available at `https://github.com/renatoberlinghieri/Helmholtz-GP`.

Related work. The Helmholtz decomposition has been used for visualization and physical interpretation of an oceanographic field (Rocha et al., 2016; Zhang et al., 2018; 2019; Han & Huang, 2020; Bühler et al., 2014; Caballero et al., 2020). But these methods assume the velocity vector field is known on a grid whereas our goal is to reconstruct the vector field from sparse observations. Prior work in atmospheric statistics used a Helmholtz decomposition to perform regression on the residuals of a physical model (Daley, 1985; Hollingsworth & Lönnberg, 1986); this approach would give the same mean prediction[2] as the Helmholtz GP if the same covariance function were used. However, to estimate the covariance function, the authors rely on a series representation and binning procedure; without hand-tuning that depends on the data and physics of the system (e.g., for determining the number of coefficients), this procedure can result in covariance functions that are not positive definite.

component is called divergence-free because their curl and divergence are zero everywhere, respectively. See Proposition B.5.

[2]Here and throughout, we use the word *prediction* in the machine learning sense; it describes the task of making informed guesses about unseen data points using a trained model and need not imply looking forward in time.

Researchers have developed GP kernels to capture curl- or divergence-free fields (Narcowich & Ward, 1994; Lowitzsch, 2002; Fuselier, 2007; Macêdo & Castro, 2010; Alvarez et al., 2012). Macêdo & Castro (2010) propose using convex combinations of such kernels. However, these works do not propose methods for recovering the weighting of the two components and do not empirically test recovery of the components when the weighting is unknown – the case of interest in the oceans problem. Wahlström et al. (2013); Wahlström (2015); Solin et al. (2018) use curl- and divergence-free kernels for electromagnetic fields. Wahlström (2015) proposes independent GP priors on the terms in a Helmholtz decomposition. But the authors assume direct access to noisy observations of each of the divergence-free and curl-free components separately – whereas we aim to recover the individual components from noisy observations of their sum. Moreover, Wahlström (2015) constrains the two components to have the same prior magnitudes and length scales, but these quantities can be expected to vary substantially between components in ocean currents. Finally, Greydanus & 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, the authors test their method only on low-resolution data available on a dense grid. In our experiments on (sparse) buoy data in Section 5 and Appendix I, we find that their method often produces predictions that are less accurate and less physically plausible.

In sum, then, it is not clear from existing work that divergence and vorticity can be usefully or practically recovered when observations come from a general (noisy) vector field that is neither curl- nor divergence-free. Moreover, there is no existing guidance on how to use a Helmholtz GP in practice for identifying divergences or making predictions from such noisy vector-field observations, and no information is available on how a Helmholtz GP compares to a velocity GP on these tasks – either empirically or theoretically. We discuss related work further in Appendix A.

2. Background

In what follows, we first describe the problem setup. Then we establish necessary notation and concepts from the Helmholtz decomposition and Gaussian processes.

Problem Statement. We consider a dataset D of M observations, $\{(\mathbf{x}_m, \mathbf{y}_m)\}_{m=1}^M$. Here $\mathbf{x}_m = (x_m^{(1)}, x_m^{(2)})^\top \in \mathbb{R}^2$ represents the location of a buoy, typically a longitude and latitude pair. We treat $\mathbf{x}_m$ as a column vector. And $\mathbf{y}_m = (y_m^{(1)}, y_m^{(2)})^\top \in \mathbb{R}^2$ gives the corresponding longitudinal and latitudinal velocities of the buoy (the *drifter trace*). For $m \in \{1, \ldots, M\}$, we consider $\mathbf{y}_m$ as a sparse noisy observation of a 2-dimensional vector field, $F : \mathbb{R}^2 \to \mathbb{R}^2$, mapping spatial locations into longitudinal and latitudinal

velocities, $F(\mathbf{x}_m) = (F^{(1)}(\mathbf{x}_m), F^{(2)}(\mathbf{x}_m))^\top$. We assume that the velocity field is stationary in time, and so F is not a function of time. Our primary goals are (1) prediction of the field F at new locations, not observed in the training data, and (2) estimation of the divergence, itself a function of location and which we define next as part of the Helmholtz decomposition. Secondarily, we are interested in recovering vorticity, another functional of F described below.

The Helmholtz Decomposition. The motion of a volume element of a fluid, such as the ocean, can be decomposed into a divergent velocity and a rotational velocity.

Definition 2.1 (Helmholtz decomposition, Bhatia et al., 2013)**.** A twice continuously differentiable and compactly supported vector field $F : \mathbb{R}^2 \to \mathbb{R}^2$ can be expressed as the sum of the gradient of a scalar potential $\Phi : \mathbb{R}^2 \to \mathbb{R}$, called the *potential function*, and the vorticity operator of another scalar potential $\Psi : \mathbb{R}^2 \to \mathbb{R}$, called the *stream function*:

$$\underbrace{F}_{\text{ocean flow}} = \underbrace{\operatorname{grad}\Phi}_{\text{divergent velocity}} + \underbrace{\operatorname{rot}\Psi}_{\text{rotational velocity}} \tag{1}$$

where

$$\operatorname{grad}\Phi := \begin{bmatrix} \partial\Phi/\partial x^{(1)} \\ \partial\Phi/\partial x^{(2)} \end{bmatrix} \text{ and } \operatorname{rot}\Psi := \begin{bmatrix} \partial\Psi/\partial x^{(2)} \\ -\partial\Psi/\partial x^{(1)} \end{bmatrix}. \tag{2}$$

The *divergence* of F (denoted δ) and the *vorticity* of F (denoted ζ) are

$$\delta := \operatorname{div}(F) := \frac{\partial F^{(1)}}{\partial x^{(1)}} + \frac{\partial F^{(2)}}{\partial x^{(2)}} = \frac{\partial^2\Phi}{\partial^2 x^{(1)}} + \frac{\partial^2\Phi}{\partial^2 x^{(2)}} \tag{3}$$

$$\zeta := \operatorname{curl}(F) := \frac{\partial F^{(1)}}{\partial x^{(2)}} - \frac{\partial F^{(2)}}{\partial x^{(1)}} = \frac{\partial^2\Psi}{\partial^2 x^{(2)}} + \frac{\partial^2\Psi}{\partial^2 x^{(1)}}. \tag{4}$$

In Equation (3), $\operatorname{div}(F)$ depends only on Φ because $\operatorname{div}(\operatorname{rot}\Psi) = 0$. In other words, the rotational velocity is divergence-free. Similarly, in Equation (4), $\operatorname{curl}(F)$ depends only on Ψ because $\operatorname{curl}(\operatorname{grad}\Phi)) = 0$. In other words, the divergent velocity is curl-free. We review the grad, rot, div, and curl operators – and explore the equations above in more detail – in Appendix B. We summarize the various terms in Table 1. In Appendix C, we present a graphical illustration of a Helmholtz decomposition of a selected vector field, and we further discuss the importance of divergence and vorticity within ocean currents.

Bayesian Approach and Gaussian Process Prior. In what follows, we will take a Bayesian approach to inferring F. In particular, we assume a likelihood, or noise model, relating the observed buoy velocities to the field F:

$$\mathbf{y}_m = F(\mathbf{x}_m) + \boldsymbol{\epsilon}_m, \boldsymbol{\epsilon}_m \overset{\text{ind}}{\sim} \mathcal{N}(0, \sigma^2_{\text{obs}}\mathbf{I}_2), 1 \le m \le M, \tag{5}$$

for some $\sigma^2_{\text{obs}} > 0$ and independent ($\overset{\text{ind}}{\sim}$) noise across observations. Here and throughout, we use $\mathbf{I}_p \in R^{p\times p}$ to denote the identity matrix in p dimensions. We use 0 to denote the zero element in any vector space.

Table 1. Terms and notation around the divergence and vorticity.

Φ	potential function
$\operatorname{grad}\Phi$	divergent velocity
$\delta = \operatorname{div}(\operatorname{grad}\Phi)$	divergence
Ψ	stream function
$\operatorname{rot}\Psi$	rotational velocity
$\zeta = \operatorname{curl}(\operatorname{rot}\Psi)$	vorticity

Before defining our prior, we review Gaussian processes (GPs). Let $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$ represent two input vectors. Assume that we want to model a P-dimensional function $G : \mathbb{R}^2 \to \mathbb{R}^P$, $G(\mathbf{x}) = (G^{(1)}(\mathbf{x}), \ldots, G^{(P)}(\mathbf{x}))^\top$. A P-output GP on covariate space $\mathbb{R}^2$ is determined by a mean function $\mu : \mathbb{R}^2 \to \mathbb{R}^P$, $\mu(\mathbf{x}) = (\mu^{(1)}(\mathbf{x}), \ldots, \mu^{(P)}(\mathbf{x}))^\top$, and a positive definite kernel function $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^{P\times P}$. We use $k(\mathbf{x}, \mathbf{x}')_{i,j}$ to denote the (i, j)th output of $k(\mathbf{x}, \mathbf{x}')$. We say that G is GP distributed and write $G \sim \mathcal{GP}(\mu, k)$ if for any $N \in \mathbb{N}$, for any $(\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \mathbb{R}^{2\times N}$, and for any vector of indices $(p_1, \ldots, p_N) \in \{1, \ldots, P\}^N$, $(G^{(p_n)}(\mathbf{x}_n))_{n=1}^N$ is an N-dimensional Gaussian random variable with mean vector $(\mu^{(p_n)}(\mathbf{x}_n))_{n=1}^N$ and covariance matrix with (i, j)th entry $k(\mathbf{x}_i, \mathbf{x}_j)_{p_i,p_j}$. See Alvarez et al. (2012) for a review of multi-output GPs.

Velocity Gaussian Process. In spatial data analysis, commonly μ is chosen to be identically 0. And a conventional choice for k would be an isotropic kernel[3] separately in each output dimension. That is, for any $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$,

$$k_{\text{vel}}(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} k^{(1)}(\mathbf{x}, \mathbf{x}') & 0 \\ 0 & k^{(2)}(\mathbf{x}, \mathbf{x}') \end{bmatrix}. \tag{6}$$

where $k^{(1)}$ and $k^{(2)}$ are isotropic kernels. We call this choice the *velocity GP* to emphasize that the independent priors are directly on the observed velocities. A standard kernel choice for $k^{(i)}$, $i \in \{1, 2\}$, is the squared exponential kernel,

$$k^{(i)}_{\text{SE}}(\mathbf{x}, \mathbf{x}') = \sigma_i^2 \exp\left(-\tfrac{1}{2}\|\mathbf{x} - \mathbf{x}'\|_2^2/\ell_i^2\right). \tag{7}$$

The velocity GP with squared exponential kernels for each component (henceforth, the *SE-velocity GP*) has four hyperparameters: for $i \in \{1, 2\}$, the signal variance $\sigma_i^2 > 0$ determines the variation of function values from their mean in the ith output dimension, and $\ell_i > 0$ controls the length scale on which the function varies.

[3]We say a kernel k is *isotropic* if there exists some $\kappa : \mathbb{R}^+ \to \mathbb{R}$ such that for any $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbb{R}^2$, $k(\mathbf{x}, \mathbf{x}') = \kappa(\|\mathbf{x} - \mathbf{x}'\|)$.

3. Gaussian Processes at the Helm(holtz)

Instead of putting separate GP priors with isotropic kernels on the two components of F as in the velocity GP, we propose to put separate GP priors with isotropic kernels on the Helmholtz scalar potentials Φ and Ψ. In this section, we describe our model and how to retrieve the quantities of interest from it. In the next section, we describe its conceptual strengths over the velocity GP, which we see empirically in Section 5.

Our Helmholtz GP prior. To form our new *Helmholtz GP* prior, we put independent GP priors on the Helmholtz stream and potential functions:

$$\Phi \sim \mathcal{GP}(0, k_\Phi) \quad \text{and} \quad \Psi \sim \mathcal{GP}(0, k_\Psi), \tag{8}$$

where we take k_Φ and k_Ψ to be isotropic kernels. When these kernels are chosen to be squared exponentials (Equation (7)), we call our model the *SE-Helmholtz GP*. The SE-Helmholtz GP has four parameters: ℓ_Φ and σ_Φ^2 for k_Φ, and ℓ_Ψ and σ_Ψ^2 for k_Ψ. We could use any two kernels such that sample paths of the resulting GPs are almost surely continuously differentiable (so that F in Equation (1) is well-defined and continuous). Generally, we will want to be able to consider divergences and vorticities of the implied process, which will require sample paths of the implied process to be at least twice-continuously differentiable. For the latter condition to hold, it is sufficient for $k_\Phi(0, \mathbf{x})$ and $k_\Psi(0, \mathbf{x})$ to have continuous mixed partial derivatives up to order five; see Lindgren (2012, Theorem 2.09 & Section 7.2).

First, we check that our prior yields a GP prior over the vector field F.

Proposition 3.1. *Let F be an ocean current vector field defined by potential and stream functions that are independent and distributed as $\Phi \sim \mathcal{GP}(0, k_\Phi)$ and $\Psi \sim \mathcal{GP}(0, k_\Psi)$, where k_Φ and k_Ψ are such that Φ and Ψ have almost surely continuously differentiable sample paths. Then*

$$F = \operatorname{grad} \Phi + \operatorname{rot} \Psi \sim \mathcal{GP}(0, k_{\text{Helm}}), \tag{9}$$

where, for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$, $i, j \in 1, 2$, $k_{\text{Helm}}(\mathbf{x}, \mathbf{x}')_{i,j}$ is equal to

$$\frac{\partial^2 k_\Phi(\mathbf{x}, \mathbf{x}')}{\partial x^{(i)} \partial (x')^{(j)}} + (-1)^{\mathbf{1}\{i \neq j\}} \frac{\partial^2 k_\Psi(\mathbf{x}, \mathbf{x}')}{\partial x^{(3-i)} \partial (x')^{(3-j)}}. \tag{10}$$

Our proof in Appendix E relies on two observations: (i) the Helmholtz decomposition is based on partial (mixed) derivatives and (ii) the derivative of a GP is a GP; see, e.g., Rasmussen & Williams (2005, Chapter 9.4), and Adler (1981, Theorem 2.2.2).

Making predictions. To make predictions using our Helmholtz GP, we need to choose the hyperparameter values and then evaluate the posterior distribution of the ocean current given those hyperparameters.

We choose the GP hyperparameters by maximizing the log marginal likelihood of the training data. To write that marginal likelihood, we let $\mathbf{X}_{\text{tr}} \in \mathbb{R}^{2 \times M}$ be the matrix with mth column equal to $\mathbf{x}_m$. We define $\mathbf{Y}_{\text{tr}} = (\mathbf{y}_1^{(1)}, \ldots, \mathbf{y}_M^{(1)}, \mathbf{y}_1^{(2)}, \ldots, \mathbf{y}_M^{(2)})^\top \in \mathbb{R}^{2M}$. We extend the definition of the mean and kernel function to allow for arbitrary finite collections of inputs. In particular, for $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \mathbb{R}^{2 \times N}$ and $\mathbf{X}' = (\mathbf{x}'_1, \ldots, \mathbf{x}'_{N'}) \in \mathbb{R}^{2 \times N'}$,

$$\mu(\mathbf{X}) = \begin{pmatrix} \mu^{(1)}(\mathbf{X}) \\ \mu^{(2)}(\mathbf{X}) \end{pmatrix} \text{ and} \tag{11}$$

$$k(\mathbf{X}, \mathbf{X}') = \begin{pmatrix} k(\mathbf{X}, \mathbf{X}')_{1,1} & k(\mathbf{X}, \mathbf{X}')_{1,2} \\ k(\mathbf{X}, \mathbf{X}')_{2,1} & k(\mathbf{X}, \mathbf{X}')_{2,2} \end{pmatrix} \tag{12}$$

where (a) for $i \in \{1, 2\}$, $n \in \{1, \ldots, N\}$, $\mu^{(i)}(\mathbf{X})$ is an N-dimensional column vector with nth entry $\mu^{(i)}(\mathbf{x}_n)$, and (b) for $i, j \in \{1, 2\}$, $n \in \{1, \ldots, N\}$, $n' \in \{1, \ldots, N'\}$, $k(\mathbf{X}, \mathbf{X}')_{i,j}$ is an $N \times N'$ matrix with (n, n')th entry $k(\mathbf{x}_n, \mathbf{x}'_{n'})_{i,j}$. With this notation, we denote the covariance of the training data with itself, under the full model including noise, as $K_{\text{trtr}} = k(\mathbf{X}_{\text{tr}}, \mathbf{X}_{\text{tr}}) + \sigma_{\text{obs}}^2 \mathbf{I}_{2M}$. Then the log marginal likelihood is

$$\begin{aligned} &\log p(\mathbf{Y}_{\text{tr}} \mid \mathbf{X}_{\text{tr}}) = \log \mathcal{N}(\mathbf{Y}_{\text{tr}}; 0, K_{\text{trtr}}) \\ &= -\frac{1}{2} \mathbf{Y}_{\text{tr}}^T K_{\text{trtr}}^{-1} \mathbf{Y}_{\text{tr}} - \frac{1}{2} \log |K_{\text{trtr}}| - \frac{2M}{2} \log 2\pi, \end{aligned} \tag{13}$$

where $|\cdot|$ takes the determinant of its matrix argument. We provide details of our optimization procedure in Section 5.

With hyperparameter values in hand, we form probabilistic predictions using the posterior of the GP. In particular, the posterior mean forms our prediction at a new set of points, and the posterior covariance encapsulates our uncertainty.

Consider N new (test) locations at which we would like to predict the current. We gather them in $\mathbf{X}_{\text{te}} \in \mathbb{R}^{2 \times N}$, with nth column equal to $\mathbf{x}_n^\star$. We denote the covariance of various training and testing combinations as: $K_{\text{tetr}} = k(\mathbf{X}_{\text{te}}, \mathbf{X}_{\text{tr}})$ and $K_{\text{tete}} = k(\mathbf{X}_{\text{te}}, \mathbf{X}_{\text{te}})$. Then a posteriori after observing the training data D, the $2N$-long vector $(F^{(1)}(\mathbf{x}_1^\star), \ldots, F^{(1)}(\mathbf{x}_N^\star), \ldots, F^{(2)}(\mathbf{x}_1^\star) \ldots, F^{(2)}(\mathbf{x}_N^\star))^\top$ describing the current at the test locations has a normal distribution with mean and covariance

$$\mu_{F|D} = K_{\text{tetr}} K_{\text{trtr}}^{-1} \mathbf{Y}_{\text{tr}}, \tag{14}$$

$$K_{F|D} = K_{\text{tete}} - K_{\text{tetr}} K_{\text{trtr}}^{-1} K_{\text{tetr}}^\top. \tag{15}$$

For more details, see Rasmussen & Williams (2005, Section 2.2). Note that these formulas can be used to evaluate posterior moments of the velocity field for either the Helmholtz GP (setting $k = k_{\text{Helm}}$) or the velocity GP (with $k = k_{\text{vel}}$).

Recovering divergence and vorticity. We next show how to recover the posterior distributions on the divergence and

vorticity scalar fields given a posterior on the current field F. We can estimate divergence and vorticity at any location by using the posterior mean at that point, and we can report uncertainty with the posterior variance. Note that our formulas recover divergence and vorticity for either our Helmholtz GP or the velocity GP.

Proposition 3.2. *Let $F \sim \mathcal{GP}(\mu, k)$ be a two-output Gaussian process with almost surely continuously differentiable sample paths. Then, for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$,*

$$\delta = \operatorname{div} F \sim \mathcal{GP}(\operatorname{div} \mu, k^{\delta}) \tag{16}$$

$$\zeta = \operatorname{curl} F \sim \mathcal{GP}(\operatorname{curl} \mu, k^{\zeta}) \tag{17}$$

where

$$k^{\delta}(\mathbf{x}, \mathbf{x}') = \sum_{(i,j)\in\{1,2\}^2} \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')_{i,j}}{\partial x^{(i)} \partial x^{(j)}} \tag{18}$$

$$k^{\zeta}(\mathbf{x}, \mathbf{x}') = \sum_{(i,j)\in\{1,2\}^2} (-1)^{i+j} \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')_{i,j}}{\partial x^{(3-i)} \partial x^{(3-j)}}. \tag{19}$$

We provide the proof for Proposition 3.2 in Appendix F.

Computational Cost. Since the latitude and longitude outputs are correlated under the Helmholtz GP, it generally has a higher computational cost than the velocity GP. We establish that the extra cost is no worse than a small factor.

Proposition 3.3. *Take M training data points. Let $C_{vel}(M)$ and $C_{helm}(M)$ be the computational costs for evaluating the log marginal likelihood (Equation* (13)*) via Cholesky or QR factorization algorithms for the velocity GP and Helmholtz GP, respectively. If we assume worst-case scaling for these algorithms, $\lim_{M\to\infty} C_{helm}(M)/C_{vel}(M) \leq 4$.*

The cost of computing the log marginal likelihood is dominated by the cost of solving the linear system $K_{\mathrm{trtr}}^{-1}\mathbf{Y}_{\mathrm{tr}}$ and computing the log determinant $|K_{\mathrm{trtr}}|$. Both of these costs in turn arise primarily from the cost of factorizing K_{trtr}. Let $\mathrm{CF}(s)$ be the cost of factorizing a square matrix with s rows with Cholesky or QR factorization. Due to the two (correlated) outputs, the cost of the Helmholtz GP is dominated by $\mathrm{CF}(2M)$. In the velocity GP, the two outputs are uncorrelated and can be handled separately, so the cost is dominated by $2\mathrm{CF}(M)$. Therefore, $\lim_{M\to\infty} C_{helm}(M)/C_{vel}(M) \leq \mathrm{CF}(2M)/(2\mathrm{CF}(M))$. When factorizing the matrix costs $\mathrm{CF}(s) \sim cs^p$ for $p \in (0, 3], c > 0$, the result follows by Equation (26). Standard Cholesky and QR factorization algorithms satisfy the condition with $p = 3$ in the worst case (Golub & Van Loan, 2013, p. 164, 249).

In Appendix G we provide similar computational results for the task of prediction and discuss nuances of how any of these results may change in the presence of special structure.

4. Advantages of the Helmholtz prior

We next describe three key advantages of the Helmholtz GP prior over the velocity GP prior: (1) more physically realistic prior assumptions reflecting the relative magnitude and length scales of the divergence and vorticity, (2) more physically realistic correlation of the longitudinal and latitudinal velocities of current at any point, and (3) equivariance to reference frame.

Prior magnitude of the divergence and vorticity. In real ocean flows, except at small-scale frontal features, the divergence is known a priori to have both a substantively different magnitude and different length scale relative to the vorticity (Barkan et al., 2019). In what follows, we argue that the Helmholtz GP is able to capture the relative contributions of divergence and vorticity directly in the prior – whereas the velocity GP does not have this direct control.

On the magnitude side, the divergence is known to contribute much less to the current than the vorticity contributes. If we consider a SE-Helmholtz GP, the signal variance hyperparameters σ_{Φ}^2 and σ_{Ψ}^2 control the magnitude of Φ and Ψ; as a direct consequence of the linearity of the divergence δ and vorticity ζ in Φ and Ψ (Equations (3) and (4)), the marginal variances of δ and ζ scale linearly with σ_{Φ}^2 and σ_{Ψ}^2, respectively. The model can therefore directly and separately control the magnitude of the rotational and divergence components. A similar argument can be applied to more general Helmholtz GPs with parameters controlling the magnitude of Φ and Ψ.

By contrast, the velocity GP provides no such control. In fact, for any isotropic choice of $k^{(1)}$ and $k^{(2)}$ we show that the resulting velocity GP must assume the same variance on the divergence and vorticity in the prior.

Proposition 4.1. *Let $k^{(1)}$ and $k^{(2)}$ be isotropic kernels with inputs $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$. Take $F^{(1)} \sim \mathcal{GP}(0, k^{(1)})$ and $F^{(2)} \sim \mathcal{GP}(0, k^{(2)})$ independent. Suppose $k^{(1)}$ and $k^{(2)}$ are such that $F^{(1)}, F^{(2)}$ have almost surely continuously differentiable sample paths. Let δ and ζ be defined as in Equations* (3) *and* (4)*. Then for any $\mathbf{x}$, $\mathrm{Var}[\delta(\mathbf{x})] = \mathrm{Var}[\zeta(\mathbf{x})]$.*

The proof of Proposition 4.1 appears in Appendix H.1.

Prior length scales of the divergence and vorticity. The divergence and vorticity are also known to operate on very different length scales in real ocean flows. Vorticity operates over long length scales, whereas divergence tends to be more localized. Similarly to the argument above, the Helmholtz GP allows control over the length scale in each of its components, which directly control the length scale of the divergence and vorticity. In particular, if $k_{\Phi}(\mathbf{x}, \mathbf{x}') = \kappa(\|\mathbf{x} - \mathbf{x}'\|/\ell)$, for some $\kappa : \mathbb{R}^+ \to \mathbb{R}$, then $k^{\delta}(\mathbf{x}, \mathbf{x}') = \ell^{-4}\eta(\|\mathbf{x} - \mathbf{x}'\|/\ell)$ for another function $\eta : \mathbb{R}^+ \to \mathbb{R}$ that does not depend on ℓ; see Appendix H.2.

Table 2. Green identifies the lowest RMSE. Dark green indicates the RMSE is at least two times smaller than the next best model.

	VELOCITY F			DIVERGENCE δ			VORTICITY ζ		
	HELM	VEL	D-HNN	HELM	VEL	D-HNN	HELM	VEL	D-HNN
VORTEX	0.24	0.72	0.54	0.0	0.22	0.87	0.77	1.05	1.03
SMALL DIVERGENCE	1.11	1.25	0.67	2.62	1.45	4.14	0.0	1.07	0.31
MEDIUM DIVERGENCE	0.17	0.19	0.55	0.39	0.33	1.32	0.05	0.12	0.38
BIG DIVERGENCE	0.04	0.10	0.19	0.05	0.12	0.27	0.00	0.10	0.11
DUFFING W/ SMALL DIVERGENCE	0.96	2.05	2.14	0.94	0.95	1.89	1.40	2.28	2.64
DUFFING W/ MEDIUM DIVERGENCE	0.19	0.60	1.65	0.14	0.50	1.15	0.24	0.26	2.39
DUFFING W/ BIG DIVERGENCE	0.41	0.22	1.63	0.08	0.17	1.10	0.48	0.16	2.41

By contrast, the velocity GP requires setting the length scales of its priors in tandem, and it is unclear how to control the length scales of the divergence and vorticity.

Correlations between longitudinal and latitudinal current components. Ocean flows have correlation between longitudinal and latitudinal velocities at single locations and across different locations. For instance, within a vortex, the longitudinal velocity at six o'clock (relative to the center of the vortex) coincides with a zero latitudinal velocity at that same location, and also with a non-zero latitudinal velocity at three o'clock. Likewise, the occurrence of divergence at a given point induces a latitudinal velocity at six o'clock (with no longitudinal velocity), as well as a non-zero longitudinal velocity at three o'clock (with no latitudinal velocity). By modeling the divergence and vorticity directly, the Helmholtz prior induces correlation between the longitudinal and latitudinal components, which is absent in the velocity GP prior.

Equivariance to reference frame. We now show the Helmholtz GP is agnostic to the choice of reference frame defined by longitude and latitude, but the velocity GP is not.

Proposition 4.2. *Let $\mu_{F|D}(\mathbf{X}_{\mathrm{te}}, \mathbf{X}_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}})$ denote the Helmholtz GP posterior mean for training data $\mathbf{X}_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}}$ and test coordinates $\mathbf{X}_{\mathrm{te}}$, and let R be an operator rotating coordinates and velocities about $(0, 0)$. Then*

$$\mu_{F|D}(R\mathbf{X}_{\mathrm{te}}, R\mathbf{X}_{\mathrm{tr}}, R\mathbf{Y}_{\mathrm{tr}}) = R\mu_{F|D}(\mathbf{X}_{\mathrm{te}}, \mathbf{X}_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}}). \quad (20)$$

Proposition 4.2 formalizes that it is equivalent to either (1) rotate the data and then predict using the Helmholtz GP or (2) predict using the Helmholtz GP and rotate the prediction. The proof of Proposition 4.2 is given in Appendix H.3.

The equivariance property in Proposition 4.2 need not hold in general for velocity GP priors.

Proposition 4.3. *For isotropic component kernels and zero prior mean, the velocity GP is reference-frame equivariant if and only if the kernels for each component are equal.*

See Appendix H.4 for the proof. Intuitively, if the kernels are equal, both the prior and likelihood (and therefore the entire model) are isotropic, and so there is no special reference frame. For intuition in the other direction, consider the following counterexample. Let $F^{(1)} \sim \mathcal{GP}(0, k^{(1)})$ for some non-identically zero isotropic $k^{(1)}$. And $F^{(2)} = 0$, a trivial isotropic prior. Take any data $\mathbf{X}_{\mathrm{tr}}$, $\mathbf{Y}_{\mathrm{tr}}$, $\mathbf{X}_{\mathrm{te}}$, and a positive (counterclockwise) $90°$ rotation. Due to the trivial prior in the second coordinate, the posterior in the second coordinate has mean $\mu^{(2)}_{F|D}(\mathbf{X}_{\mathrm{te}}, \mathbf{X}_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}}) = 0$. If we rotate the data first, the posterior in the second coordinate is still zero, and generally the posterior in the first coordinate will be nontrivial. But if we first compute the posterior and then rotate the mean, the posterior in the first coordinate will now be zero instead, and the posterior in the second coordinate will be nonzero. Therefore, the equality in Equation (20) will not hold for this velocity GP.

5. Experimental Results

We next empirically compare the SE-Helmholtz GP and SE-velocity GP. We find that the SE-Helmholtz GP yields better current predictions as well as better divergence (and vorticity) identification, across a variety of simulated and real data sets. We also compare to dissipative Hamiltonian neural networks (D-HNNs) (Greydanus & Sosanya, 2022) in Table 2 and Appendix I but find that the GP methods generally perform better.

Data. The real datasets we use consist of drifter traces of GPS-tagged buoys in the ocean. While oceanographers have some knowledge that allows a rough assessment of the real data, only in simulations do we have access to ground truth currents, divergences, and vorticities. Therefore, we run a variety of simulations with current vector fields reflecting known common ocean behaviors. We simulate buoy trajectories by initializing buoys at a starting point and allowing the current field to drive their motion. See Appendix I for more details of our setup in each specific simulation.

Performance. In what follows, we emphasize visual comparisons both because the distinctions between methods are generally clear and because it is illuminating to visually pick out behaviors of interest. We also provide root mean

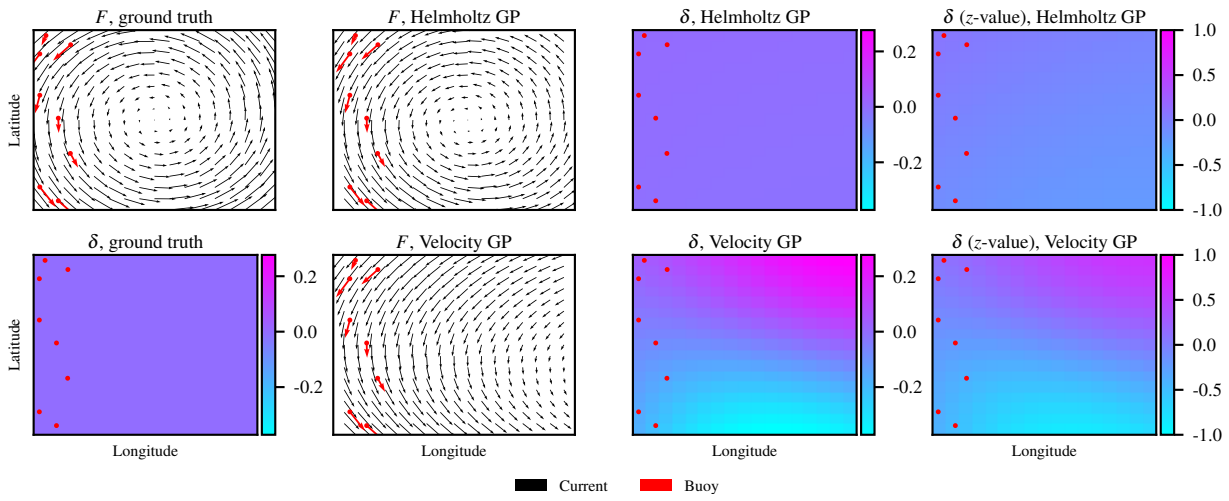

Figure 1. First column: ground truth predictions (upper) and divergence (lower). Second column: current predictions. Third column: divergence estimates. Fourth column: posterior divergence z-values.

squared error (RMSE) comparisons in Table 2. However, we note that the RMSE can be expected to vary as one changes either the ocean area or the densities (or more generally locations) of test points, and both of these choices must always be somewhat arbitrary.

Algorithmic details. In our comparisons, each full model including an SE-Helmholtz GP prior or an SE-velocity GP prior has five hyperparameters: σ_Φ^2, ℓ_Φ, σ_Ψ^2, ℓ_Ψ, σ_{obs}^2 and $\ell_1, \sigma_1^2, \ell_2, \sigma_2^2, \sigma_{\text{obs}}^2$, respectively. In each case, we fit the log of the hyperparameters by maximizing the marginal likelihood using Adam (Kingma & Ba, 2015). We optimize in the log-scale and then exponentiate the optimal values to ensure positivity of the hyperparameters. We run each experiment until the log marginal likelihood changes by less than 10^{-4}, which occurs in fewer than 2000 iterations for all experiments. With the exception of the GLAD data (which presents special difficulties that we describe in Appendix I.3), we found that results were not sensitive to initialization. To train the D-HNN, we ran the code from Greydanus & Sosanya (2022). More algorithmic details are provided individually for each experiment in Appendix I.

5.1. Simulated experiments

We focus on simulations of key ocean behaviors of interest to oceanographers: vortices, concentrated divergences, and combinations thereof. As a summary across simulation experiments in Table 2, we see that – for predictions, divergence estimates, and vorticity estimates – the SE-Helmholtz GP is most accurate by RMSE on a majority of tasks, often by a substantial factor. We next examine individual experiments more closely; all simulated experiments are described in detail in Appendix I.1.

Vortex with zero divergence. First, we consider a vortex with no divergence. The SE-Helmholtz GP is better at predicting the current and identifying the lack of divergence.

In a vortex, water particles rotate around a central point. The black arrows in the upper left plot of Figure 1 show the ground-truth vector field at test points, with longitude on the horizontal axis and latitude on the vertical axis. Red arrows show our simulated buoy trajectories, which we give to all methods as training data. See Appendix I.1.1 for additional details of the setup and results.

The second column in Figure 1 shows predictions from the SE-Helmholtz GP (upper) and SE-velocity GP (lower) at the test points. The red arrows are still the training data. Despite having access to data only from one side of the vortex, the SE-Helmholtz GP is able to reconstruct the full vortex. The SE-velocity GP is not.

The ground truth divergence is identically 0 throughout the domain and depicted in the lower left plot. The third column shows divergence estimates from the SE-Helmholtz GP (upper) and SE-velocity GP (lower) on the same color scale. The fourth column helps us understand if either posterior is reporting a nonzero divergence. In particular, for each point we plot a "z-value": precisely, the posterior mean at that point divided by the posterior standard deviation. One might, for instance, conclude that a method has detected a nonzero divergence if the magnitude of the z-value is greater than 1. From the third column, we conclude that the SE-Helmholtz GP estimate of the divergence is closer to the ground truth of zero than the SE-velocity GP. From the fourth column, we see that neither method concludes nonzero divergence, but the SE-Helmholtz GP posterior is more concentrated near zero.

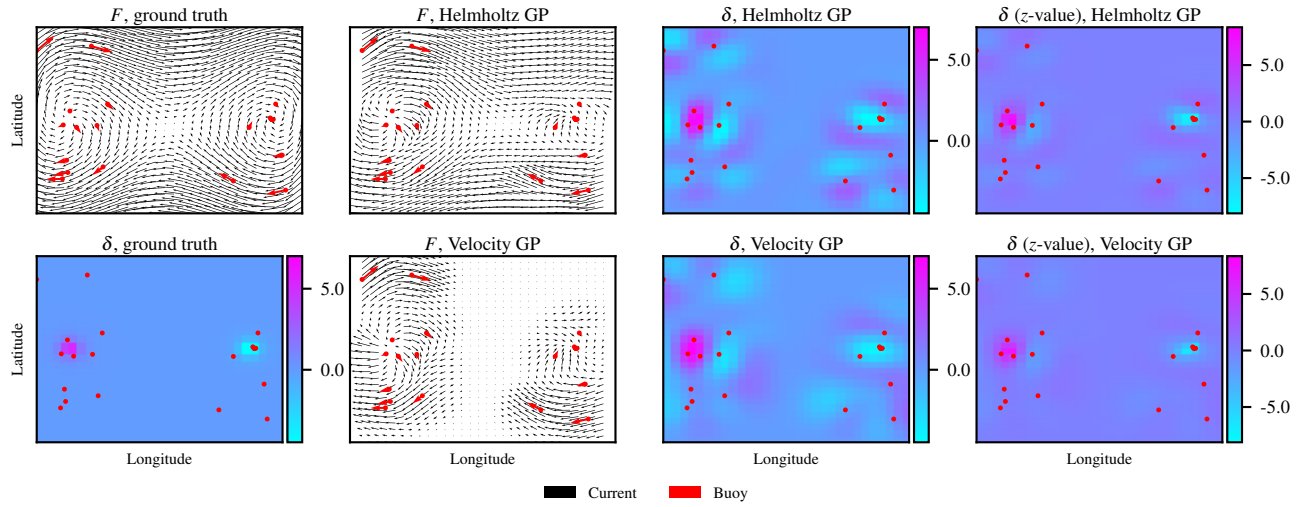

Figure 2. First column: ground truth predictions (upper) and divergence (lower). Second column: current predictions. Third column: divergence estimates. Fourth column: posterior divergence z-values.

Duffing oscillator with areas of concentrated divergence. We next simulate a classic example called a Duffing oscillator, and we add two areas of divergence; the ground truth current appears in the upper left plot of Figure 2, and the ground truth divergence appears in the lower left. The simulated buoy trajectories appear in red. See Appendix I.1.4 for further details on setup and results.

We see in the second column that the SE-Helmholtz GP (upper) is largely able to reconstruct the two vortices in the Duffing oscillator (upper left), though it struggles with the upper right current. By contrast, the SE-velocity GP does not connect the currents continuously across the two sides of the space, in disagreement with conservation of momentum.

Again, the third column depicts divergence estimates from both methods, and the fourth column depicts z-values. In this case, both methods accurately recover the two areas of divergence. In Appendix I.1.4 and Figures 10 to 12, we experiment with smaller and larger areas of divergence with the Duffing oscillator. In Appendix I.1.3, we isolate areas of divergence without the Duffing oscillator. Across the six experiments involving regions of divergence, the SE-Helmholtz GP typically outperforms the SE-velocity GP in detecting these regions – often by a substantial margin, as shown in Table 2. In these same experiments, the SE-Helmholtz GP similarly outperforms the SE-velocity GP at predicting the velocity field.

A note on vorticity. Although we have not focused on vorticity estimation in the main text, generally we find superior performance on this task from the SE-Helmholtz GP relative to the SE-velocity GP (and D-HNNs), similar to divergence estimation. See the righthand side of Table 2 for an RMSE comparison, and see Appendix I for a visual comparison. For example, the SE-Helmholtz GP can predict zero vorticity when there is no vorticity, whereas the SE-velocity GP and D-HNN fail in this task (Figures 7 to 9).

5.2. Real-data experiments

Although ground truth currents and divergences are not available for real data, we can still assess performance against oceanographers' expert knowledge.

LASER data. The LAgrangian Submesoscale ExpeRiment (D'Asaro et al., 2017) was performed in the Gulf of Mexico in 2016. The full dataset spans two winter months and more than 1000 buoys; see Appendix I.2 for details. We focus on a particular spatial area and 2-hour time span in which oceanographers expect to see a particular convergent front. Time information within this range is discarded. We are left with 19 buoys. When recorded spatial coordinates overlap across data points, we observe that both the SE-Helmholtz GP and the SE-velocity GP treat all data as noise. While this issue merits further investigation and model development, for now we downsample across time to form the final data sets here.

The left column of Figure 3 shows the current predictions using the SE-Helmholtz GP (upper) and SE-velocity GP (lower). Red arrows show the observed buoy data, with 55 total observations across all buoys. The black arrows show the current posterior means at test locations. The two sets of predictions are qualitatively very similar.

The second column shows the divergence predictions for the SE-Helmholtz GP (upper) and SE-velocity GP (lower); the third column shows the z-values for the respective posterior distributions. The SE-Helmholtz GP predicts a negative divergence area (light blue diagonal) that agrees with the area

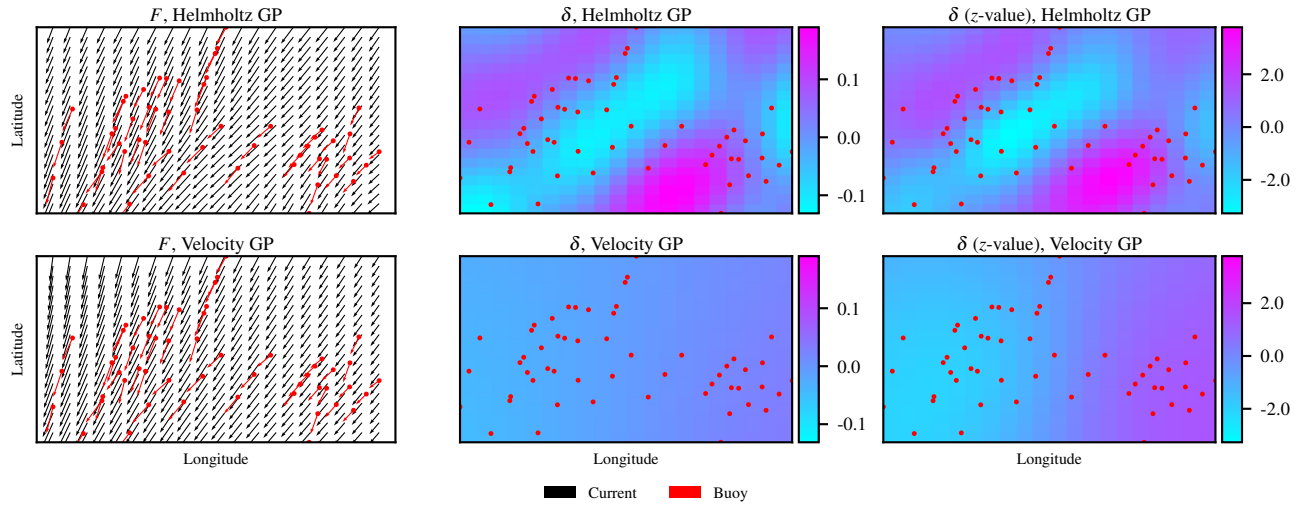

Figure 3. First column: current predictions. Second column: divergence estimates. Third column: posterior divergence z-values.

where oceanographers expect a convergent front. By contrast, the SE-velocity GP does not identify any divergence.

We find that the discrepancy between the SE-Helmholtz GP and the SE-velocity GP observed in Figure 3 depends on the amount of data. When we double the amount of data from that in Figure 3 (by downsampling less), we find that both methods are able to recover the same convergent front; see Appendix I.2 and Figure 14. This finding then also corroborates that the convergent front in fact does exist and looks like the posterior detection by the SE-Helmholtz GP in Figure 3. In that way, the finding further lends support to the superior performance of the SE-Helmholtz GP in Figure 3.

GLAD data. The Grand Lagrangian Deployment (GLAD) experiment (Özgökmen, 2012) was conducted near the Deepwater Horizon site and Louisiana coast in July 2012. The full dataset consists of over 300 buoys. Unlike the winter LASER data, the summer GLAD data faces additional challenges from regular oscillations due to wind (rather than current); see Appendix I.3 for more details. Rather than account for these oscillations, for the moment we ameliorate their effect by downsampling considerably both in time and across buoys. We are left with 85 observations.

The current predictions between the SE-Helmholtz GP and SE-velocity GP are generally quite similar (Figures 15 and 16). Although in this case we do not have any ground truth from oceanographers in advance, some features of the SE-Helmholtz GP prediction seem more realistic from physical intuition: namely, the downturn in the lower left region of the plot and the vortex in the upper left region of the plot. We also note that the SE-Helmholtz GP and SE-velocity GP predict substantially different divergences: respectively, zero and nonzero.

To check that our SE-Helmholtz GP method is computationally reasonable, we run on a larger subset of the GLAD data with 1200 data points. We find that it takes less than 10 minutes to optimize the hyperparameters, form predictions, and estimate the divergences at the test points.

6. Discussion and Future Work

We have demonstrated the conceptual and empirical advantages of our Helmholtz GP relative to the velocity GP. A number of challenges remain. While we have focused on the purely spatial case for modularity, in general we expect currents to change over time, and it remains to extend our method to the spatiotemporal case – which we believe should be straightforward. Moreover, Gonçalves et al. (2019); Lodise et al. (2020) used more complex kernels with two length scales per dimension in their spatiotemporal extension of the velocity GP. From known ocean dynamics, we expect that in fact two length scales would be appropriate in the Helmholtz GP for modeling the vorticity – but unnecessary for the divergence. While our independent Gaussian noise model is standard in spatiotemporal modeling – and shared by Gonçalves et al. (2019); Lodise et al. (2020) – our real-data experimentation suggests that a better noise model might account for short-term wind oscillations and other noise patterns distinctive to oceans.

Acknowledgements

The authors are grateful to the Office of Naval Research for partial support under grant N00014-20-1-2023 (MURI ML-SCOPE). Renato Berlinghieri and Tamara Broderick were also supported in part by an NSF CAREER Award, and Tamay Özgökmen was supported in part by RSMAES, the University of Miami.

References

Adler, R. *The Geometry of Random Fields*. Society for Industrial and Applied Mathematics, 1981.

Alvarez, M. A., Rosasco, L., and Lawrence, N. D. Kernels for vector-valued functions: A review. *Foundations and Trends in Machine Learning*, 4(3):195–266, 2012.

Arfken, G. B. and Weber, H. J. Mathematical methods for physicists, 1999.

Barkan, R., Molemaker, M. J., Srinivasan, K., McWilliams, J. C., and D'Asaro, E. A. The role of horizontal divergence in submesoscale frontogenesis. *Journal of Physical Oceanography*, 49(6):1593–1618, 2019.

Bhatia, H., Norgard, G., Pascucci, V., and Bremer, P.-T. The Helmholtz-Hodge decomposition—a survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(8):1386–1404, 2013.

Bühler, O., Callies, J., and Ferrari, R. Wave–vortex decomposition of one-dimensional ship-track data. *Journal of Fluid Mechanics*, 756:1007–1026, 2014.

Caballero, A., Mulet, S., Ayoub, N., Manso-Narvarte, I., Davila, X., Boone, C., Toublanc, F., and Rubio, A. Integration of HF radar observations for an enhanced coastal mean dynamic topography. *Frontiers in Marine Science*, pp. 1005, 2020.

Daley, R. The analysis of synoptic scale divergence by a statistical interpolation procedure. *Monthly weather review*, 113(6):1066–1080, 1985.

Daley, R. *Atmospheric data analysis*. Number 2. Cambridge university press, 1993.

D'Asaro, E., Guigand, C., Haza, A., Huntley, H., Novelli, G., Özgökmen, T., and Ryan, E. Lagrangian submesoscale experiment (laser) surface drifters, interpolated to 15-minute intervals, 2017. URL `https://data.gulfresearchinitiative.org/data/R4.x265.237:0001`.

D'Asaro, E. A., Shcherbina, A. Y., Klymak, J. M., Molemaker, J., Novelli, G., Guigand, C. M., Haza, A. C., Haus, B. K., Ryan, E. H., Jacobs, G. A., Huntley, H. S., Laxague, N. J. M., Chen, S., Judt, F., McWilliams, J. C., Barkan, R., Kirwan, A. D., Poje, A. C., and Özgökmen, T. M. Ocean convergence and the dispersion of flotsam. *Proceedings of the National Academy of Sciences*, 115 (6):1162–1167, 2018.

Fuselier, Jr., E. J. *Refined error estimates for matrix-valued radial basis functions*. PhD thesis, Texas A&M University, 2007.

Golub, G. H. and Van Loan, C. F. *Matrix computations*. JHU press, 2013.

Gonçalves, R. C., Iskandarani, M., Özgökmen, T., and Thacker, W. C. Reconstruction of submesoscale velocity field from surface drifters. *Journal of Physical Oceanography*, 49(4), 2019.

Greydanus, S. and Sosanya, A. Dissipative Hamiltonian neural networks: Learning dissipative and conservative dynamics separately. *arXiv preprint arXiv:2201.10085*, 2022.

Greydanus, S., Dzamba, M., and Yosinski, J. Hamiltonian neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

Han, L. and Huang, R. X. Using the Helmholtz decomposition to define the Indian Ocean meridional overturning streamfunction. *Journal of Physical Oceanography*, 50 (3), 2020.

Hollingsworth, A. and Lönnberg, P. The statistical structure of short-range forecast errors as determined from radiosonde data. part i: The wind field. *Tellus A: Dynamic Meteorology and Oceanography*, 38(2):111–136, 1986.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Lindgren, G. *Stationary stochastic processes: theory and applications*. CRC Press, 2012.

Lodise, J., Özgökmen, T., Gonçalves, R. C., Iskandarani, M., Lund, B., Horstmann, J., Poulain, P.-M., Klymak, J., Ryan, E. H., and Guigand, C. Investigating the formation of submesoscale structures along mesoscale fronts and estimating kinematic quantities using Lagrangian drifters. *Fluids*, 5(3), 2020.

Lowitzsch, S. *Approximation and interpolation employing divergence-free radial basis functions with applications*. Texas A&M University, 2002.

Macêdo, I. and Castro, R. Learning divergence-free and curl-free vector fields with matrix-valued kernels. Technical report, Instituto Nacional de Matemática Pura e Aplicada, 2010.

Narcowich, F. J. and Ward, J. D. Generalized hermite interpolation via matrix-valued conditionally positive definite functions. *Mathematics of Computation*, 63(208):661–687, 1994.

Novelli, G., Guigand, C. M., Cousin, C., Ryan, E. H., Laxague, N. J., Dai, H., Haus, B. K., and Özgökmen, T. M.

A biodegradable surface drifter for ocean sampling on a massive scale. *Journal of Atmospheric and Oceanic Technology*, 34(11):2509–2532, 2017.

Özgökmen, T. CARTHE: GLAD 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.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian processes for machine learning*. MIT Press, 2005.

Rocha, C. B., Chereskin, T. K., Gille, S. T., and Menemenlis, D. Mesoscale to submesoscale wavenumber spectra in Drake Passage. *Journal of Physical Oceanography*, 46 (2):601–620, 2016.

Solin, A., Kok, M., Wahlström, N., Schön, T. B., and Särkkä, S. Modeling and interpolation of the ambient magnetic field by Gaussian processes. *IEEE Transactions on robotics*, 34(4):1112–1127, 2018.

Srinivasan, K., Barkan, R., and McWilliams, J. C. A forward energy flux at submesoscales driven by frontogenesis. *Journal of Physical Oceanography*, 53(1):287–305, 2023.

Vallis, G. K. *Atmospheric and oceanic fluid dynamics*. Cambridge University Press, 2017.

Wahlström, N. *Modeling of magnetic fields and extended objects for localization applications*. PhD thesis, Linköping University Electronic Press, 2015.

Wahlström, N., Kok, M., Schön, T. B., and Gustafsson, F. Modeling magnetic fields using Gaussian processes. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3522–3526. IEEE, 2013.

Yaremchuk, M. and Coelho, E. F. Filtering drifter trajectories sampled at submesoscale resolution. *IEEE Journal of Oceanic Engineering*, 40(3):497–505, 2014.

Zhang, C., Wei, H., Liu, Z., and Fu, X. Characteristic ocean flow visualization using Helmholtz decomposition. In *2018 Oceans-MTS/IEEE Kobe Techno-Oceans (OTO)*, pp. 1–4. IEEE, 2018.

Zhang, C., Wei, H., Bi, C., and Liu, Z. Helmholtz–Hodge decomposition-based 2D and 3D ocean surface current visualization for mesoscale eddy detection. *Journal of Visualization*, 22, 01 2019.

Appendix

Table of Contents

A Related Work **12**

B Divergence, Gradient, and Curl Operators in 2D **13**

C Helmholtz Decomposition in the Ocean **14**

D Optimal Interpolation vs. Gaussian Processes **14**

E Helmholtz Gaussian Process Prior **16**

F Divergence and Vorticity of A Gaussian Process **17**

G Computational Costs for evaluating Helmholtz GP Posterior **18**

H Benefits of the Helmholtz GP: additional information and supplemental proofs **19**

H.1 Equality of marginal variances of vorticity and divergence with independent velocity priors 19

H.2 Conservation of length scales across k_Φ vs. k^δ (and k_Ψ vs. k^ζ) 20

H.3 Equivariance of Helmoltz GP predictions 21

H.4 Non-Equivariance of Velocity GP predictions 23

I Experimental Results **24**

I.1 Simulated Experiments 24

I.2 Real-world data 1: LASER 43

I.3 Real-world data 2: GLAD 47

A. Related Work

In what follows, we present related work in more detail. We first describe in more detail the differences between our work and that of Gonçalves et al. (2019); Lodise et al. (2020). Then we discuss why we chose to put our priors on the Helmholtz decomposition, rather than an alternative decomposition.

Velocity GP vs. the GP of Gonçalves et al. (2019); Lodise et al. (2020). Gonçalves et al. (2020); Lodise et al. (2020) used many of the components of the velocity GP that we describe in the main text, but their prior was substantially more complex than the velocity GP. Like the velocity GP, they focused on a GP prior with a squared exponential covariance function. Unlike the velocity GP as described in the main text here, their squared exponential prior included not only terms in each of the longitude and latitude directions, but also a term in the time direction. Each term has its own length scale. As a second principle difference, their covariance was in fact a sum of two such squared exponential kernels – introducing a total of 6 length scales (one for longitude, latitude, and time in each of the two kernels), 2 signal variances, and a single noise variance. They mention also trying 3 kernels (instead of 2), but it appears all their results were reported for 2 kernels. We have here tried to take a modular approach to examine the squared exponential prior on its own, so our velocity GP should not be seen as a direct reflection of the performance of the Gonçalves et al. (2019); Lodise et al. (2020) covariance function.

Why we focused on the Helmholtz decomposition. The Helmholtz decomposition is a widely recognized dynamically significant method for dissecting the oceanic velocity field. An alternative – albeit related – decomposition that sees frequent use is the Geostrophic-Ageostrophic (G-Ag hereinafter) decomposition (Vallis, 2017), which relies on the dominance of geostrophic balance at large spatial and time scales in the ocean. Though not directly related to our current work here, we discuss it briefly to provide a holistic oceanographic context to our choice of the Helmholtz decomposition. The Helmholtz decomposition is defined through exact linear operators into velocity components that are orthogonal complements and can be separated easily, allowing priors to be placed on the underlying potentials; the G-Ag decomposition, however, can be derived only by first eliminating the Ageostrophic flow (which represents faster, smaller scales) through an ad hoc time smoothing of drifter velocities using a multi-day filter. The G-Ag components are notably not orthogonal complements and consequently have to be separately estimated through a velocity GP, leading to a more complex modeling pipeline with additional physical hyperparameters (like the smoothing time) that are not easily determined. While there is a measure of correspondence between the geostrophic and rotational components, and the ageostrophic and divergent components, the lack of precision in defining the G-Ag components makes the Helmholtz a natural modeling pathway. Recent oceanographic studies (Barkan et al., 2019; Srinivasan et al., 2023) showing that the Helmholtz decomposition is directly relevant to the dynamics of oceanic components at smaller spatial scales of around O(1 km) offer further justification for our present choice.

B. Divergence, Gradient, and Curl Operators in 2D

In this section we provide some background for the Helmholtz decomposition in 2D. In the first part, we provide definitions for $\mathrm{grad}, \mathrm{div}, \mathrm{curl}$, and rot operators. In Proposition B.5 we then characterize a property of vector fields obtained combining these operators.

Consider a scalar-valued differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$. 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,

$$\mathrm{grad}\, f(\mathbf{x}) := \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x^{(1)}} \\ \frac{\partial f(\mathbf{x})}{\partial x^{(2)}} \end{bmatrix} = \boldsymbol{i}\frac{\partial f(\mathbf{x})}{\partial x^{(1)}} + \boldsymbol{j}\frac{\partial f(\mathbf{x})}{\partial x^{(2)}}$$

where $\boldsymbol{i}$ and $\boldsymbol{j}$ are 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*:

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

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

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

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

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

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

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 *rot* operator - that plays the role of the standard curl operator in the 3D version. In 2D, the rot formally requires the introduction of a third unit vector, $\boldsymbol{k}$ that is orthogonal to the plane containing, $\boldsymbol{i}$ and $\boldsymbol{j}$.

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

$$\mathrm{rot}\, f(\mathbf{x}) \equiv \boldsymbol{k} \times \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x^{(2)}} \\ \frac{-\partial f}{\partial x^{(1)}} \end{bmatrix} = \boldsymbol{i}\frac{\partial f}{\partial x^{(2)}} - \boldsymbol{j}\frac{\partial f}{\partial x^{(1)}}$$

where $\boldsymbol{i}$ and $\boldsymbol{j}$ represents, respectively, the standard unit vectors in the direction of the $x^{(1)}$ and $x^{(2)}$ coordinates; $\boldsymbol{k}$ is the unit vector orthogonal to the plane containing $\boldsymbol{i}$ and $\boldsymbol{j}$ satisfying the identities, $\boldsymbol{k} \times \boldsymbol{j} = -\boldsymbol{i}$ and $\boldsymbol{k} \times \boldsymbol{i} = \boldsymbol{j}$.

Thus the rot operator can be thought of as a $\pi/2$ rotation of the grad operator. The precise reason why we need the introduction of a separate rot operator in 2D is because of a hidden peculiarity that the stream function, Ψ is actually the only non-zero component of a 3D vector potential field, $\boldsymbol{A}(\mathbf{x})$, but that non-zero component is along the $\boldsymbol{k}$ direction, $\boldsymbol{A} \equiv (0, 0, \Psi(\mathbf{x}))$; equivalently $\boldsymbol{A} = \Psi \boldsymbol{k}$. Given this observation, it can be shown that $\nabla_{3D} \times \boldsymbol{A} = \boldsymbol{k} \times \nabla\Psi$, where ∇_{3D} is the direct 3D extension of the 2D ∇ operator defined above. The ideas of gradient, divergence, rot, and curl lead to the following characterization of vector fields.

Definition B.4. A vector field $F : A \to \mathbb{R}^2$ 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 B.5. *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* $\text{curl}(\text{grad } f) = 0$. *Every gradient of a scalar field is rotation free.*
2. *If f is $\mathcal{C}^2$, then* $\text{div}(\text{rot } f) = 0$. *Every* rot *transformation of a scalar field is incompressible.*

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

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

For (2):

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

□

For more material on vector calculus, we refer the reader to Arfken & Weber (1999).

C. Helmholtz Decomposition in the Ocean

In what follows we relate the Helmholtz decomposition to ocean currents. In the first part, we provide intuition of how divergence and vorticity are significant in the context of oceanography. Next, in Figure 4, we present a visual representation of the Helmholtz decomposition and highlight the relevant aspects.

The divergence and vorticity of the ocean flow are relevant for oceanographic studies. Divergence characterizes fronts – small structures with spatial scales on the order of 0.1-10 km and temporal scales on the order of 1-100h. 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, vorticity 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).

In Figure 4 we provide visual intuition on how the Helmholtz theorem decomposes a vector field (ocean flow) into a divergent velocity field and a rotation velocity field. In this 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.

D. Optimal Interpolation vs. Gaussian Processes

Optimal interpolation is a powerful and widely used technique for the analysis of atmospheric data. The technique is only optimal under strong assumptions, and is therefore instead often referred to as *statistical interpolation* (Daley, 1993).

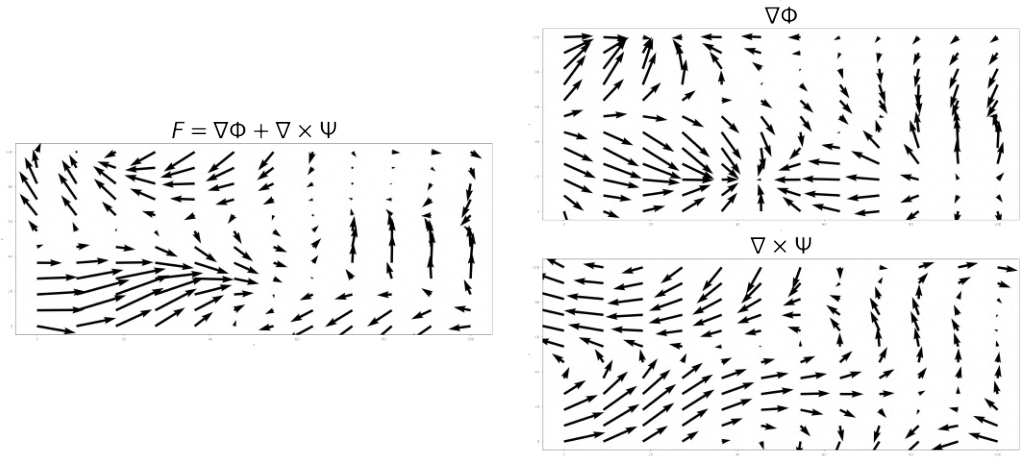

Figure 4. Helmholtz decomposition of vector field F. *Left*: original vector field F. *Top-right*: divergence component of Helmholtz decomposition of F. *Bottom-right*: vorticity component of Helmholtz decomposition of F.

We formulate the statistical interpolation algorithm following Daley (1993, Section 4.2). We are interested in modeling a variable $y = f(\boldsymbol{x})$, where $\mathbf{x}$ represent the 2D spatial coordinates of a point. For simplicity, assume $f : \mathbb{R}^2 \to \mathbb{R}$ is a univariate function. The multivariate statistical interpolation algorithm is very similar. See Chapter 5 in Daley (1993).

We have some locations $\boldsymbol{x}_k, k \in 1, \ldots, K$, where we observe the variable of interest directly, $f_O(\boldsymbol{x}_k)$, and some other locations $\boldsymbol{x}_i, i \in 1, \ldots, I$, where we want to infer (*interpolate*) the value of the function of interest, $f_A(\boldsymbol{x}_i)$. Moreover, there might be some background knowledge about this function (e.g. a climate model providing a first guess), that we denote by $f_B(\boldsymbol{x}_k)$ and $f_B(\boldsymbol{x}_i)$. We assume a linear relationship between the observations and the function we want to model (note that this does not imply the modeled function is linear in the spatial locations). Then

$$f_A(\boldsymbol{x}_i) = f_B(\boldsymbol{x}_i) + \sum_{i=1}^{K} W_{ik}(f_O(\boldsymbol{x}_k) - f_B(\boldsymbol{x}_k)) \tag{21}$$

where K is the number of locations where we observe the variable of interest, and W_{ik} is the weight given to the residual $f_O(\boldsymbol{x}_k) - f_B(\boldsymbol{x}_k)$ when interpolating at $\boldsymbol{x}_i$. Finally, let the true value of the function at points $\boldsymbol{x}_i$ and $\boldsymbol{x}_k$ be $f_T(\boldsymbol{x}_i)$ and $f_T(\boldsymbol{x}_k)$, respectively.

Define $\bar{\boldsymbol{W}}_i$ to be the column vector of length K of weights, $\bar{\boldsymbol{B}}_i$ the column vector of length K whose elements are $\mathbb{E}[(f_B(\boldsymbol{x}_k) - f_T(\boldsymbol{x}_k))(f_B(\boldsymbol{x}_i) - f_T(\boldsymbol{x}_i))]$, and $\boldsymbol{B}$ and $\boldsymbol{O}$ the background and observation error covariance matrices, respectively. That is, for $k, l \in \{1, \ldots, K\}$,

$$\begin{aligned}\boldsymbol{B}_{k,l} &= \mathbb{E}[(f_B(\boldsymbol{x}_k) - f_T(\boldsymbol{x}_k))(f_B(\boldsymbol{x}_l) - f_T(\boldsymbol{x}_l))],\\ \boldsymbol{O}_{k,l} &= \mathbb{E}[(f_O(\boldsymbol{x}_k) - f_T(\boldsymbol{x}_k))(f_O(\boldsymbol{x}_l) - f_T(\boldsymbol{x}_l))].\end{aligned}$$

Then, assuming that (i) the background and the observations are unbiased and (ii) that the correlation between background and observation error is zero, with some algebra it can be shown that the optimal weights satisfy

$$[\boldsymbol{B} + \boldsymbol{O}]\bar{\boldsymbol{W}}_i = \bar{\boldsymbol{B}}_i$$

We refer the reader to Section 4.2 in Daley (1993) for details on this derivation. This is the same as applying a least square approach on the residuals obtained by subtracting the background model estimates from the observations.

Now assume that we do not have any background knowledge ($f_B(\boldsymbol{x}_i) = f_B(\boldsymbol{x}_k) = 0$) and we do not know the ground truth. Instead, we model it with a Gaussian process with zero mean and covariance function $\tilde{K}$, $f_T \sim \mathcal{GP}(0, \tilde{K})$. Assuming that

the model is well-specified and the observation noise is σ^2_{obs}, we have

$$\begin{aligned}
\bar{\boldsymbol{B}}_{i,k} &= \mathbb{E}(f_T(\boldsymbol{x}_i), f_T(\boldsymbol{x}_k)) = \tilde{K}(\boldsymbol{x}_i, \boldsymbol{x}_k) \\
\boldsymbol{B}_{k,l} &= \mathbb{E}(f_T(\boldsymbol{x}_k), f_T(\boldsymbol{x}_l)) = \tilde{K}(\boldsymbol{x}, \boldsymbol{x})_{k,l} \\
\boldsymbol{O}_{k,l} &= \mathbb{E}(f_O(\boldsymbol{x}_k) - f_T(\boldsymbol{x}_k), f_O(\boldsymbol{x}_l) - f_T(\boldsymbol{x}_l)) = \sigma^2_{\text{obs}}
\end{aligned}$$

where by $\tilde{K}(\boldsymbol{x}, \boldsymbol{x})$ we mean the $I \times I$ matrix formed by evaluating the $\tilde{K}$ on all pairs of observed locations. So the optimal weights are

$$\bar{\boldsymbol{W}}_{\ell,k} = \sum_{i=1}^{I} [\tilde{K}(\boldsymbol{x}, \boldsymbol{x}) + \sigma^2_{\text{obs}} \mathbf{I}]^{-1}_{\ell,i} \tilde{K}(\boldsymbol{x}_i, \boldsymbol{x}_k),$$

These weights lead to exactly the same predictions as when doing standard Gaussian process regression. Therefore, we can see Gaussian process regression as a specific case of statistical interpolation, where we do not include any background knowledge (corresponding to a zero mean prior), and we do not have access to the ground truth function. We then choose a covariance function that encodes physical intuition of how the underlying system behaves, and estimate parameters from data (for example, by maximum likelihood). Daley (1993, Section 4.3) provides a detailed discussion of the choice of covariance functions in the oceanographic and atmospheric literature.

E. Helmholtz Gaussian Process Prior

In this section, we state and prove Proposition 3.1 from the main text.

Proposition 3.1. *Let F be an ocean current vector field defined by potential and stream functions that are independent and distributed as $\Phi \sim \mathcal{GP}(0, k_\Phi)$ and $\Psi \sim \mathcal{GP}(0, k_\Psi)$, where k_Φ and k_Ψ are such that Φ and Ψ have almost surely continuously differentiable sample paths. Then*

$$F = \operatorname{grad} \Phi + \operatorname{rot} \Psi \sim \mathcal{GP}(0, k_{\text{Helm}}), \tag{9}$$

where, for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$, $i, j \in 1, 2$, $k_{\text{Helm}}(\mathbf{x}, \mathbf{x}')_{i,j}$ is equal to

$$\frac{\partial^2 k_\Phi(\mathbf{x}, \mathbf{x}')}{\partial x^{(i)} \partial (x')^{(j)}} + (-1)^{\mathbf{1}\{i \neq j\}} \frac{\partial^2 k_\Psi(\mathbf{x}, \mathbf{x}')}{\partial x^{(3-i)} \partial (x')^{(3-j)}}. \tag{10}$$

Proof. We obtain the result in two steps. First, we argue that under the assumptions of the proposition, F is distributed as a Gaussian process and so may be characterized through its mean and covariance function. Second, we show F has mean zero, and the proposed covariance kernel.

To see that F is a Gaussian process, observe that it is the sum of linear transformations of two independent Gaussian processes. This follows from the fact that grad and rot are linear operators on any vector space of differentiable functions, and because k_Φ and k_Ψ are chosen to have almost surely continuously differentiable sample paths. Therefore, $\operatorname{grad}\Phi$ and $\operatorname{rot}\Psi$ are two independent GPs, and so F is a Gaussian process as well.

We next turn to the mean and covariance functions. By linearity of expectation,

$$\begin{aligned}
\mathbb{E}[F] &= \mathbb{E}\left[\operatorname{grad} \Phi\right] + \mathbb{E}\left[\operatorname{curl} \Psi\right] \\
&= \operatorname{grad} \mathbb{E}\Phi + \operatorname{curl} \mathbb{E}\Psi \\
&= 0,
\end{aligned}$$

where the last line follows from the assumption that Φ and Ψ both have mean 0 everywhere. It remains to calculate the covariance function. Since Φ and Ψ are assumed independent we compute the covariance as the sum of covariances for

$\operatorname{grad}\Phi$ and $\operatorname{curl}\Psi$. Consider two points $\mathbf{x}$ and $\mathbf{x}'$.

$$\begin{aligned}
\operatorname{Cov}\left[(\operatorname{grad}\Phi)(\mathbf{x}), (\operatorname{grad}\Phi)(\mathbf{x}')\right] &= \operatorname{Cov}\left[\begin{pmatrix}\frac{\partial\Phi(\mathbf{x})}{\partial x^{(1)}}\\ \frac{\partial\Phi(\mathbf{x})}{\partial x^{(2)}}\end{pmatrix}, \begin{pmatrix}\frac{\partial\Phi(\mathbf{x}')}{\partial (x')^{(1)}}\\ \frac{\partial\Phi(\mathbf{x}')}{\partial (x')^{(2)}}\end{pmatrix}\right] \\
&= \begin{bmatrix}\operatorname{Cov}\left(\frac{\partial\Phi(\mathbf{x})}{\partial x^{(1)}}, \frac{\partial\Phi(\mathbf{x}')}{\partial (x')^{(1)}}\right) & \operatorname{Cov}\left(\frac{\partial\Phi(\mathbf{x})}{\partial x^{(1)}}, \frac{\partial\Phi(\mathbf{x}')}{\partial (x')^{(2)}}\right)\\ \operatorname{Cov}\left(\frac{\partial\Phi(\mathbf{x})}{\partial x^{(2)}}, \frac{\partial\Phi(\mathbf{x}')}{\partial (x')^{(1)}}\right) & \operatorname{Cov}\left(\frac{\partial\Phi(\mathbf{x})}{\partial x^{(2)}}, \frac{\partial\Phi(\mathbf{x}')}{\partial (x')^{(2)}}\right)\end{bmatrix} \\
&= \begin{bmatrix}\dfrac{\partial^2 k_\Phi(\mathbf{x},\mathbf{x}')}{\partial x^{(1)}\partial (x')^{(1)}} & \dfrac{\partial^2 k_\Phi(\mathbf{x},\mathbf{x}')}{\partial x^{(1)}\partial (x')^{(2)}}\\ \dfrac{\partial^2 k_\Phi(\mathbf{x},\mathbf{x}')}{\partial x^{(2)}\partial (x')^{(1)}} & \dfrac{\partial^2 k_\Phi(\mathbf{x},\mathbf{x}')}{\partial x^{(2)}\partial (x')^{(2)}},\end{bmatrix}
\end{aligned}$$

where exchange of integration and differentiation to obtain the final matrix is permissible by the almost surely continuously differentiable sample paths assumption.

Similarly,

$$\begin{aligned}
\operatorname{Cov}\left[(\operatorname{rot}\Psi)(\mathbf{x}), (\operatorname{rot}\Psi))(\mathbf{x}')\right] &= \operatorname{Cov}\left[\begin{pmatrix}\frac{\partial\Psi(\mathbf{x})}{\partial x^{(2)}}\\ -\frac{\partial\Psi(\mathbf{x})}{\partial x^{(1)}}\end{pmatrix}, \begin{pmatrix}\frac{\partial\Psi(\mathbf{x}')}{\partial (x')^{(2)}}\\ -\frac{\partial\Psi(\mathbf{x}')}{\partial (x')^{(1)}}\end{pmatrix}\right] \\
&= \begin{bmatrix}\operatorname{Cov}\left(\frac{\partial\Psi(\mathbf{x})}{\partial x^{(2)}}, \frac{\partial\Psi(\mathbf{x}')}{\partial (x')^{(2)}}\right) & \operatorname{Cov}\left(\frac{\partial\Psi(\mathbf{x})}{\partial x^{(2)}}, -\frac{\partial\Psi(\mathbf{x}')}{\partial (x')^{(1)}}\right)\\ \operatorname{Cov}\left(-\frac{\partial\Psi(\mathbf{x})}{\partial x^{(1)}}, \frac{\partial\Psi(\mathbf{x}')}{\partial (x')^{(2)}}\right) & \operatorname{Cov}\left(\frac{\partial\Psi(\mathbf{x})}{\partial x^{(1)}}, \frac{\partial\Psi(\mathbf{x}')}{\partial (x')^{(1)}}\right)\end{bmatrix} \\
&= \begin{bmatrix}\dfrac{\partial^2 k_\Psi(\mathbf{x},\mathbf{x}')}{\partial x^{(2)}\partial (x')^{(2)}} & -\dfrac{\partial^2 k_\Psi(\mathbf{x},\mathbf{x}')}{\partial x^{(2)}\partial (x')^{(1)}}\\ -\dfrac{\partial^2 k_\Psi(\mathbf{x},\mathbf{x}')}{\partial x^{(1)}\partial (x')^{(2)}} & \dfrac{\partial^2 k_\Psi(\mathbf{x},\mathbf{x}')}{\partial x^{(1)}\partial (x')^{(1)}}\end{bmatrix}.
\end{aligned}$$

The desired expression for k_{Helm} is obtained by taking the sum of these two matrices.

□

F. Divergence and Vorticity of A Gaussian Process

In this section, we state and prove Proposition 3.2 from the main text.

Proposition 3.2. *Let $F \sim \mathcal{GP}(\mu, k)$ be a two-output Gaussian process with almost surely continuously differentiable sample paths. Then, for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$,*

$$\delta = \operatorname{div} F \sim \mathcal{GP}(\operatorname{div}\mu, k^\delta) \tag{16}$$

$$\zeta = \operatorname{curl} F \sim \mathcal{GP}(\operatorname{curl}\mu, k^\zeta) \tag{17}$$

where

$$k^\delta(\mathbf{x},\mathbf{x}') = \sum_{(i,j)\in\{1,2\}^2} \frac{\partial^2 k(\mathbf{x},\mathbf{x}')_{i,j}}{\partial x^{(i)}\partial x^{(j)}} \tag{18}$$

$$k^\zeta(\mathbf{x},\mathbf{x}') = \sum_{(i,j)\in\{1,2\}^2} (-1)^{i+j} \frac{\partial^2 k(\mathbf{x},\mathbf{x}')_{i,j}}{\partial x^{(3-i)}\partial x^{(3-j)}}. \tag{19}$$

Proof. By the assumption that the sample paths are almost surely continuously differentiable, $\operatorname{div} F$ and $\operatorname{curl} F$ are well-defined. Since the image of a Gaussian process under a linear transformation is a Gaussian processes both $\operatorname{div} F$ and $\operatorname{curl} F$ are Gaussian processes. It remains to compute the moments. The expectation can be calculated via linearity,

$$\mathbb{E}(\operatorname{div} F) = \operatorname{div}(\mathbb{E}F) = \operatorname{div}\mu, \tag{22}$$

$$\mathbb{E}(\operatorname{curl} F) = \operatorname{curl}(\mathbb{E}F) = \operatorname{curl}\mu. \tag{23}$$

We next turn to the covariance. Define the centered process $G = F - \mu$. By Equation (22) and Equation (23), $\operatorname{div} G$ and $\operatorname{curl} G$ are centered Gaussian processes with the covariance functions k^δ and k^ζ respectively.

Consider two points $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$. Unpacking the definition of div,

$$\begin{aligned} k^\delta(\mathbf{x}, \mathbf{x}') &= \mathbb{E}\left[\left(\frac{\partial G^{(1)}(\mathbf{x})}{\partial x^{(1)}} + \frac{\partial G^{(2)}(\mathbf{x})}{\partial x^{(2)}}\right)\left(\frac{\partial G^{(1)}(\mathbf{x}')}{\partial (x')^{(1)}} + \frac{\partial G^{(2)}(\mathbf{x}')}{\partial (x')^{(2)}}\right)\right] \\ &= \sum_{(i,j)\in\{1,2\}^2} \mathbb{E}\left[\frac{\partial G^{(i)}(\mathbf{x})}{\partial x^{(i)}} \frac{\partial G^{(j)}(\mathbf{x}')}{\partial (x')^{(j)}}\right] \\ &= \sum_{(i,j)\in\{1,2\}^2} \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')_{i,j}}{\partial x^{(i)} \partial (x')^{(j)}}, \end{aligned}$$

where exchange of integration and differentiation in the final line is permissible given that the sample paths are almost surely continuously differentiable. Similarly,

$$\begin{aligned} k^\zeta(\mathbf{x}, \mathbf{x}') &= \mathbb{E}\left[\left(\frac{\partial G^{(1)}(\mathbf{x})}{\partial x^{(2)}} - \frac{\partial G^{(2)}(\mathbf{x})}{\partial x^{(1)}}\right)\left(\frac{\partial G^{(1)}(\mathbf{x}')}{\partial (x')^{(2)}} - \frac{\partial G^{(2)}(\mathbf{x}')}{\partial (x')^{(1)}}\right)\right] \\ &= \sum_{(i,j)\in\{1,2\}^2} (-1)^{i+j} \mathbb{E}\left[\frac{\partial G^{(i)}(\mathbf{x})}{\partial x^{(3-i)}} \frac{\partial G^{(j)}(\mathbf{x}')}{\partial (x')^{(3-j)}}\right] \\ &= \sum_{(i,j)\in\{1,2\}^2} (-1)^{i+j} \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')_{i,j}}{\partial x^{(3-i)} \partial (x')^{(3-j)}}. \end{aligned}$$

□

G. Computational Costs for evaluating Helmholtz GP Posterior

In this section, we provide a bound for the cost of computing velocity predictions using the Helmholtz GP. We also discuss in more detail the assumption of using a Cholesky factorization or QR decomposition.

Proposition G.1. *Suppose we have observed M training data points and would like to predict the current at N new (test) locations. Assume we use a Cholesky or QR factorization, together with solving the system of equations with back-substitution. Let $C_{vel}(M, N)$ be the total worst-case[4] computational cost for evaluating both the posterior mean (Equation* (14)*) and covariance (Equation* (15)*) for the velocity GP. Let and $C_{helm}(M, N)$ be the analogous total cost for the Helmholtz GP. Then*

$$\lim_{M,N\to\infty} C_{helm}(M, N)/C_{vel}(M, N) \leq 4 \tag{24}$$

where M and N can tend to infinity at arbitrary, independent rates.

Proof. Recall that the posterior mean and covariance can be obtained by solving $K_{\text{tetr}} K_{\text{trtr}}^{-1} \mathbf{Y}_{\text{tr}}$ and computing $K_{\text{tete}} - K_{\text{tetr}} K_{\text{trtr}}^{-1} K_{\text{tetr}}^\top$, respectively Equations (14) and (15). To compute the mean we (A) compute a Cholesky (or QR) factorization of a $2M \times 2M$ matrix, (B) perform a back-solve of a $2M$ dimensional system of equations, and (C) compute N $2M$-dimensional inner products. To compute the covariance we (D) compute a Cholesky (or QR) factorization of a $2M \times 2M$ matrix, (E) perform a back-solve of N distinct $2M$ dimensional systems of equations, (F) compute an $(2N \times 2M) \times (2M \times 2N)$ matrix-multiplication, and (G) subtract $2N \times 2N$ matrices.

We first argue that it suffices to consider steps A and D separately from the remaining steps. In particular, for non-negative numbers $\{a_i\}_{i=1}^I$ and $\{b_i\}_{i=1}^I$, we observe that

$$\frac{\sum_{i=1}^I a_i}{\sum_{i=1}^I b_i} \leq \max \frac{a_i}{b_i}. \tag{25}$$

[4] As in the main text, we assume that the computation incurs the worst-case cost of a Cholesky factorization or QR decomposition. If the matrices involved have special structure, the cost might be much less than the worst-case.

For our purposes, let a_1 be the cost of steps A and D in the Helmholtz GP computation, and let a_2 be the cost of steps B, C, E, F, and G in the Helmholtz GP computation. Analogously, define b_1 and b_2 for the velocity GP computation. The proof of Proposition 3.3 already showed that a_1/b_1 is asymptotically bounded above by 4. By Equation (25), then, it suffices to separately check a_2/b_2.

Next, we will use the following observation to focus on the asymptotically dominant terms of a_2 and b_2, respectively. Consider some index $n \to \infty$. For $r_n \sim s_n$ and $t_n \sim u_n$ with all terms bounded away from zero,

$$\frac{r_n}{s_n} \sim \frac{t_n}{u_n}. \tag{26}$$

Among steps B, C, E, F, and G, the dominant costs are due to performing the triangular back-substitution (step E) and the matrix-matrix multiplication (step F) used to compute the covariance. We can again consider these two steps separately due to Equation (25).

First consider back-substitution. Given a t (square) lower triangular system of equations of dimension s, the cost of solving this system with back-substitution is $\sim ts^2$ floating-point operations (Golub & Van Loan, 2013, Section 3.1.2). When evaluating posteriors with the Helmholtz GP, we have K_{tetr} of shape $2N \times 2M$ and $K_{\text{trtr}} = LL^\top$ (or QR) of shape $2M \times 2M$. So back-substitution in the case of the Helmholtz GP incurs cost (in floating point operations) $\sim 8NM^2$. For the velocity GP, we can exploit the fact that the two outputs are uncorrelated and can be handled separately, so the cost of back-substitution is $\sim 4NM^2$. Therefore, and again using Equation (26), the ratio of the costs of the back-substitution algorithm is asymptotically 2.

Once back-substitution has been performed, matrix-multiplication must be performed. Using textbook matrix-multiplication leads to the same considerations as back-substitution.

In all the steps used to compute the posterior moments, we see that the cost of the Helmholtz GP is not (asymptotically) more than 4 times the cost of the velocity GP, and the result follows. □

Simplifying structure. In Propositions 3.3 and G.1, we have made the assumption that users are solving a linear system or computing a log determinant with a general and standard choice, such as a Cholesky factorization or QR decomposition. We expect essentially the same result to hold for any other general method for computing these quantities. However, if there is special structure that can be used to solve the linear system or compute the log determinant more efficiently, that might change the bounds we have found here. Conversely, we are immediately aware of special structure that we can expect to always apply in the application to modeling current. And any such structure would likely also require special algorithmic development and coding.

H. Benefits of the Helmholtz GP: additional information and supplemental proofs

In what follows, we provide additional information on the benefits of using the Helmholtz GP. In Appendix H.1 we state and prove Proposition 4.1, showing how with independent velocity priors we obtain equal marginal variances for vorticity and divergence. In Appendix H.2 we provide more intuition on the result that length scales are conserved across k_Φ vs. k^δ and across k_Ψ vs. k^ζ. Finally, in Appendix H.3 we state and prove Proposition 4.2 about the equivariance of Helmholtz GP predictions.

H.1. Equality of marginal variances of vorticity and divergence with independent velocity priors

Proposition 4.1. *Let $k^{(1)}$ and $k^{(2)}$ be isotropic kernels with inputs $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$. Take $F^{(1)} \sim \mathcal{GP}(0, k^{(1)})$ and $F^{(2)} \sim \mathcal{GP}(0, k^{(2)})$ independent. Suppose $k^{(1)}$ and $k^{(2)}$ are such that $F^{(1)}, F^{(2)}$ have almost surely continuously differentiable sample paths. Let δ and ζ be defined as in Equations* (3) *and* (4)*. Then for any $\mathbf{x}$, $\mathrm{Var}[\delta(\mathbf{x})] = \mathrm{Var}[\zeta(\mathbf{x})]$.*

Proof. Because $k^{(1)}$ and $k^{(2)}$ are assumed to be isotropic we may write for any $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$

$$k^{(1)}(\mathbf{x}, \mathbf{x}') = \kappa_1(\|\mathbf{x} - \mathbf{x}'\|^2) \text{ and } k^{(2)}(\mathbf{x}, \mathbf{x}') = \kappa_2(\|\mathbf{x} - \mathbf{x}'\|^2)$$

for some $\kappa_1, \kappa_2 : \mathbb{R}^+ \to \mathbb{R}$. Because isotropy implies stationarity, it suffices to consider the variance at any a single point,

and so we consider $\mathbf{x} = \mathbf{x}' = (0, 0)$. By assumption, we have

$$F \sim \mathcal{GP}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}k^{(1)} & 0\\ 0 & k^{(2)}\end{pmatrix}\right)$$

By Proposition 3.2, the induced divergence and vorticity are Gaussian processes with mean 0 and covariances

$$\begin{aligned}
k^{\delta}(\mathbf{x}, \mathbf{x}') &= \frac{\partial^2 k^{(1)}(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(1)}} + \frac{\partial^2 k^{(2)}(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(2)}} \\
k^{\zeta}(\mathbf{x}, \mathbf{x}') &= \frac{\partial^2 k^{(1)}(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(2)}} + \frac{\partial^2 k^{(2)}(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(1)}}
\end{aligned}$$

respectively.

Then, we can compute the variance at $\mathbf{x} = \mathbf{x}' = (0, 0)$ by

$$\begin{aligned}
\mathrm{Var}[\delta(0,0)] &= \frac{\partial^2 k^{(1)}(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(1)}}\Big|_{\mathbf{x}=0, \mathbf{x}'=0} + \frac{\partial^2 k^{(2)}(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(2)}}\Big|_{\mathbf{x}=0, \mathbf{x}'=0} \\
&= \frac{\partial^2 \kappa_1(\|\mathbf{x} - \mathbf{x}'\|^2)}{\partial x^{(1)} \partial (x')^{(1)}}\Big|_{\mathbf{x}=0, \mathbf{x}'=0} + \frac{\partial^2 \kappa_2(\|\mathbf{x} - \mathbf{x}'\|^2)}{\partial x^{(2)} \partial (x')^{(2)}}\Big|_{\mathbf{x}=0, \mathbf{x}'=0} \\
&= \frac{\partial}{\partial x^{(1)}}(-2\kappa_1'(\|\mathbf{x}\|^2)x^{(1)})\Big|_{\mathbf{x}=0} + \frac{\partial}{\partial x^{(2)}}(-2\kappa_2'(\|\mathbf{x}\|^2)x^{(2)})\Big|_{\mathbf{x}=0} \\
&= -2(\kappa_1''(0) + \kappa_2''(0))
\end{aligned}$$

Consequently, we have that for any $\mathbf{x} \in \mathbb{R}^2$, $\mathrm{Var}[\delta(\mathbf{x})] = -2(\kappa_1'(0) + \kappa_2'(0))$.

The computation is similar for the vorticity. We have that

$$\begin{aligned}
\mathrm{Var}[\zeta(0,0)] &= \frac{\partial^2 k^{((1)}(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(2)}}\Big|_{\mathbf{x}=0, \mathbf{x}'=0} + \frac{\partial^2 k^{((2)}(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(1)}}\Big|_{\mathbf{x}=0, \mathbf{x}'=0} \\
&= \frac{\partial^2 \kappa_1(\|\mathbf{x} - \mathbf{x}'\|^2)}{\partial x^{(2)} \partial (x')^{(2)}}\Big|_{\mathbf{x}=0, \mathbf{x}'=0} + \frac{\partial^2 \kappa_2(\|\mathbf{x} - \mathbf{x}'\|^2)}{\partial x^{(1)} \partial (x')^{(1)}}\Big|_{\mathbf{x}=0, \mathbf{x}'=0} \\
&= \frac{\partial}{\partial x^{(2)}}(-2\kappa_1'(\|\mathbf{x}\|^2)x^{(1)})\Big|_{\mathbf{x}=0} + \frac{\partial}{\partial x^{(1)}}(-2\kappa_2'(\|\mathbf{x}\|^2)x^{(2)})\Big|_{\mathbf{x}=0} \\
&= -2(\kappa_1''(0) + \kappa_2''(0))
\end{aligned}$$

Therefore for any $\mathbf{x} \in \mathbb{R}^2$, $\mathrm{Var}[\zeta(\mathbf{x})] = -2(\kappa_1'(0) + \kappa_2'(0))$, and we see $\mathrm{Var}[\zeta(\mathbf{x})] = \mathrm{Var}[\delta(\mathbf{x})]$. This completes the proof. □

H.2. Conservation of length scales across k_Φ vs. k^δ (and k_Ψ vs. k^ζ)

This subsection provides a derivation of the claim that if $k_\Phi(\mathbf{x}, \mathbf{x}'; \ell) = \kappa(\|\mathbf{x} - \mathbf{x}'\|/\ell)$, for some $\kappa : \mathbb{R}_+ \to \mathbb{R}$, then $k^{\delta}_{\mathrm{Helm}}(\mathbf{x}, \mathbf{x}'; \ell) = \ell^{-4}\eta(\|\mathbf{x} - \mathbf{x}'\|/\ell)$ for some $\eta : \mathbb{R}_+ \to \mathbb{R}$ that does not depend on ℓ. The relationship (argument to see it) is identical K_Ψ and $k^{\zeta}_{\mathrm{Helm}}$.

We may see the claim to be true by expanding out the dependencies of $k^{\delta}_{\mathrm{Helm}}$ and $k^{\zeta}_{\mathrm{Helm}}$ on k_Φ and k_Ψ, seeing that they involve fourth order partial derivatives, and applying a change of variables four times; each change of variables contributes one factor of ℓ^{-1}.

In particular,

$$
\begin{aligned}
k^{\delta}_{\mathrm{Helm}} &= \left(\frac{\partial^4}{\partial (x^{(2)})^2 \partial ((x')^{(2)})^2} + \frac{\partial^4}{\partial (x^{(2)})^2 \partial ((x')^{(1)})^2} + \frac{\partial^4}{\partial (x^{(1)})^2 \partial ((x')^{(2)})^2} + \frac{\partial^4}{\partial (x^{(1)})^2 \partial ((x')^{(1)})^2} \right) k_{\Phi} \quad \text{and} \\
k^{\zeta}_{\mathrm{Helm}} &= \left(\frac{\partial^4}{\partial (x^{(2)})^2 \partial ((x')^{(2)})^2} + \frac{\partial^4}{\partial (x^{(2)})^2 \partial ((x')^{(1)})^2} + \frac{\partial^4}{\partial (x^{(1)})^2 \partial ((x')^{(2)})^2} + \frac{\partial^4}{\partial (x^{(1)})^2 \partial ((x')^{(1)})^2} \right) k_{\Psi}.
\end{aligned}
\tag{27}
$$

Consider first the potential function and divergence. If $k_{\Phi}(\mathbf{x}, \mathbf{x}') = \kappa(\|\mathbf{x} - \mathbf{x}'\|)$, then any second order mixed partial derivative may be written through a change of variables (x to x/ℓ) as $\frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}'} k_{\Phi}(\mathbf{x}, \mathbf{x}') = \ell^{-2} \frac{\partial^2}{\partial \ell \mathbf{x} \partial \ell \mathbf{x}'} \kappa(\|\ell \mathbf{x} - \ell \mathbf{x}'\|/\ell) = \ell^{-2} \frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}'} \kappa(\|\mathbf{x} - \mathbf{x}'\|/\ell)$. Analogously, when we differentiate four times rather than twice to obtain $k^{\delta}_{\mathrm{Helm}}$ we have that if $k_{\Phi}(\mathbf{x}, \mathbf{x}') = \kappa(\|\mathbf{x} - \mathbf{x}'\|/\ell)$, for some κ, then $k^{\delta}_{\mathrm{Helm}} = \ell^{-4} \eta(\|\mathbf{x} - \mathbf{x}'\|/\ell)$ for some η that does not depend on ℓ.

H.3. Equivariance of Helmoltz GP predictions

Proposition 4.2. *Let $\mu_{F|D}(\mathbf{X}_{\mathrm{te}}, \mathbf{X}_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}})$ denote the Helmholtz GP posterior mean for training data $\mathbf{X}_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}}$ and test coordinates $\mathbf{X}_{\mathrm{te}}$, and let R be an operator rotating coordinates and velocities about $(0, 0)$. Then*

$$
\mu_{F|D}(R\mathbf{X}_{\mathrm{te}}, R\mathbf{X}_{\mathrm{tr}}, R\mathbf{Y}_{\mathrm{tr}}) = R\mu_{F|D}(\mathbf{X}_{\mathrm{te}}, \mathbf{X}_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}}). \tag{20}
$$

To prove the proposition, it is helpful to distinguish between random variables and the values they take on. We use boldface to denote the random variables, for example $\mathbf{Y}_{\mathrm{tr}}$. When a random variable $\mathbf{Y}_{\mathrm{tr}}$ takes a value Y we write $\mathbf{Y}_{\mathrm{tr}} = Y$. The rotation operator R is characterized by a 2×2 rotation matrix; if $\mathbf{X}_{\mathrm{tr}} = [(x_1^{(1)}, x_1^{(2)})^\top, \ldots, (x_N^{(1)}, x_N^{(2)})^\top]$, then $R\mathbf{X}_{\mathrm{tr}} = [R(x_1^{(1)}, x_1^{(2)})^\top, \ldots, R(x_N^{(1)}, x_N^{(2)})^\top] = [((R\mathbf{x}_1)^{(1)}, (R\mathbf{x}_1)^{(2)})^\top, \ldots, ((R\mathbf{x}_N)^{(1)}, (R\mathbf{x}_N)^{(2)})^\top]$, where we denote by $(R\mathbf{x})^{(1)}$ the rotated first coordinate, and $(R\mathbf{x})^{(2)}$ the rotated second coordinate. When the input is flattened, as in the case of $\mathbf{Y}_{\mathrm{tr}}$ or $\mu_{F|D}$, the R operator is applied as follows: (1) unflatten the vector to get it in the same form as $\mathbf{X}_{\mathrm{tr}}$, then (2) apply the operator R as specified above, and finally (3) flatten the output vector to go back to the original $\mathbf{Y}_{\mathrm{tr}}$ shape. Our proof relies on k_{Φ} and k_{Ψ} being isotropic kernels.

Lemma H.1 (Invariance of the likelihood)**.** *Suppose F is distributed as a Helmholtz GP, and there are M observations $\mathbf{Y}_{\mathrm{tr}} \mid F, \mathbf{X}_{\mathrm{tr}} = X \sim \mathcal{N}\left([F^{(1)}(X), F^{(2)}(X)]^\top, K_{\mathrm{trtr}}(X, X)\right)$, where I_{2M} denotes the identity matrix of size $2M$. Then the marginal likelihood of the observations is* invariant *to rotation. That is, for any 2×2 rotation matrix R,*

$$
p(\mathbf{Y}_{\mathrm{tr}} = Y \mid \mathbf{X}_{\mathrm{tr}} = X) = p(\mathbf{Y}_{\mathrm{tr}} = RY \mid \mathbf{X}_{\mathrm{tr}} = RX).
$$

Proof. By assumption, k_{Φ} is stationary and so, for any two locations $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbb{R}^2$ we may write $k_{\Phi}(\mathbf{x}, \mathbf{x}') = \kappa(\|\mathbf{x} - \mathbf{x}'\|)$ for some function $\kappa : \mathbb{R}_+ \to \mathbb{R}$. Following Appendix C, we may write the induced covariance for $\mathrm{grad}\Phi$ as

$$
\begin{aligned}
\mathrm{Cov}\left[(\mathrm{grad}\,\Phi)(\mathbf{x}), (\mathrm{grad}\,\Phi)(\mathbf{x}')\right] &= \begin{bmatrix} \dfrac{\partial^2 k_{\Phi}(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(1)}} & \dfrac{\partial^2 k_{\Phi}(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(2)}} \\ \dfrac{\partial^2 k_{\Phi}(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(1)}} & \dfrac{\partial^2 k_{\Phi}(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(2)}} \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\partial^2 \kappa(\|\mathbf{x} - \mathbf{x}'\|)}{\partial x^{(1)} \partial (x')^{(1)}} & \dfrac{\partial^2 \kappa(\|\mathbf{x} - \mathbf{x}'\|)}{\partial x^{(1)} \partial (x')^{(2)}} \\ \dfrac{\partial^2 \kappa(\|\mathbf{x} - \mathbf{x}'\|)}{\partial x^{(2)} \partial (x')^{(1)}} & \dfrac{\partial^2 \kappa(\|\mathbf{x} - \mathbf{x}'\|)}{\partial x^{(2)} \partial (x')^{(2)}}, \end{bmatrix}
\end{aligned}
$$

Similarly, we may compute $\mathrm{Cov}\left[(\mathrm{grad}\,\Phi)(R\mathbf{x}), (\mathrm{grad}\,\Phi)(R\mathbf{x}')\right]$ through a change of variables ($\mathbf{x}$ to $R\mathbf{x}$) as

$$\begin{aligned}
\operatorname{Cov}\left[(\operatorname{grad}\Phi)(R\mathbf{x}),(\operatorname{grad}\Phi)(R\mathbf{x}')\right] &= \begin{bmatrix} \frac{\partial^2 k_\Phi(R\mathbf{x},R\mathbf{x}')}{\partial x^{(1)}\partial(x')^{(1)}} & \frac{\partial^2 k_\Phi(R\mathbf{x},R\mathbf{x}')}{\partial x^{(1)}\partial(x')^{(2)}} \\ \frac{\partial^2 k_\Phi(R\mathbf{x},R\mathbf{x}')}{\partial x^{(2)}\partial(x')^{(1)}} & \frac{\partial^2 k_\Phi(R\mathbf{x},R\mathbf{x}')}{\partial x^{(2)}\partial(x')^{(2)}} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial^2 \kappa(\|R\mathbf{x}-R\mathbf{x}'\|)}{\partial x^{(1)}\partial(x')^{(1)}} & \frac{\partial^2 \kappa(\|R\mathbf{x}-R\mathbf{x}'\|)}{\partial x^{(1)}\partial(x')^{(2)}} \\ \frac{\partial^2 \kappa(\|R\mathbf{x}-R\mathbf{x}'\|)}{\partial x^{(2)}\partial(x')^{(1)}} & \frac{\partial^2 \kappa(\|R\mathbf{x}-R\mathbf{x}'\|)}{\partial x^{(2)}\partial(x')^{(2)}}, \end{bmatrix} \\
&= R^\top \operatorname{Cov}\left[(\operatorname{grad}\Phi)(\mathbf{x}),(\operatorname{grad}\Phi)(\mathbf{x}')\right] R
\end{aligned}$$

and see that $\operatorname{Cov}\left[(\operatorname{grad}\Phi)(\mathbf{x}),(\operatorname{grad}\Phi)(\mathbf{x}')\right] = R\operatorname{Cov}\left[(\operatorname{grad}\Phi)(R\mathbf{x}),(\operatorname{grad}\Phi)(R\mathbf{x}')\right])R^\top$.

Similarly, for a collections of M locations X we have that $\operatorname{Cov}\left[(\operatorname{grad}\Phi)(X),(\operatorname{grad}\Phi)(X)\right] = (R \otimes I_M)\operatorname{Cov}\left[(\operatorname{grad}\Phi)(RX),(\operatorname{grad}\Phi)(RX)\right](R^\top \otimes I_M)$, where $\otimes$ denotes the Kronecker product.

An identical argument (up to a change in the sign of off-diagonal terms) can be used to derive the induced covariance for $\operatorname{rot}\Psi$, $\operatorname{Cov}\left[(\operatorname{rot}\Psi)(\mathbf{x}),(\operatorname{rot}\Psi)(\mathbf{x}')\right]$. We obtain

$$\begin{aligned}
\operatorname{Cov}\left[(\operatorname{rot}\Psi)(\mathbf{x}),(\operatorname{rot}\Psi)(\mathbf{x}')\right] &= R\operatorname{Cov}\left[(\operatorname{rot}\Psi)(R\mathbf{x}),(\operatorname{rot}\Psi)(R\mathbf{x}')\right]R^\top \quad \text{and} \\
\operatorname{Cov}\left[(\operatorname{rot}\Psi)(X),(\operatorname{rot}\Psi)(X)\right] &= (R\otimes I_M)\operatorname{Cov}\left[(\operatorname{rot}\Psi)(RX),(\operatorname{rot}\Psi)(RX)\right](R^\top \otimes I_M).
\end{aligned}$$

Together, this implies that if we write the covariance of M vector velocity *tr*aining observations $\mathbf{Y}_{\mathrm{tr}}$ at X as

$$\begin{aligned}
K_{\mathrm{trtr}}(X,X) :&= \operatorname{Var}[\mathbf{Y}_{\mathrm{tr}}|\mathbf{X}_{\mathrm{tr}}=X] \\
&= k_{Helm}(X,X) + \sigma_{obs}^2 I_{2M} \\
&= \operatorname{Cov}\left[(\operatorname{grad}\Phi)(X),(\operatorname{grad}\Phi)(X)\right] + \operatorname{Cov}\left[(\operatorname{rot}\Psi)(X),(\operatorname{rot}\Psi)(X)\right] + \sigma_{obs}^2 I_{2M}
\end{aligned}$$

then

$$K_{\mathrm{trtr}}(X,X) = (R\otimes I_M)K_{\mathrm{trtr}}(RX,RX)(R^\top \otimes I_M)$$

As a result, for any R, Y and X we may compute the log likelihood according to the likelihood model as

$$\begin{aligned}
&\log p(\mathbf{Y}_{\mathrm{tr}} = RY \mid \mathbf{X}_{\mathrm{tr}} = RX) \\
&\quad = \log \mathcal{N}\left(RY; 0, K_{\mathrm{trtr}}(RX,RX)\right) \\
&\quad = -M\log(2\pi) - \frac{1}{2}\log|K_{\mathrm{trtr}}(RX,RX)| \\
&\quad - \frac{1}{2}\log((R\otimes I_M)Y)^\top \left[K_{\mathrm{trtr}}(RX,RX)\right]^{-1}((R\otimes I_M)Y) \\
&\quad = -M\log(2\pi) - \frac{1}{2}\log|K_{\mathrm{trtr}}(X,X)| - \frac{1}{2}\log Y^\top K_{\mathrm{trtr}}(X,X)^{-1}Y \\
&\quad = \log p(\mathbf{Y}_{\mathrm{tr}} = Y \mid \mathbf{X}_{\mathrm{tr}} = X),
\end{aligned}$$

as desired.

□

Lemma H.2 (Invariance of the conditionals)**.** *The conditionals distributions of the Helmoltz GP are invariant to rotation. That is, for any* 2×2 *rotation matrix* R,

$$\begin{aligned}
&p(\mathbf{Y}_{\mathrm{te}} = Y_{\mathrm{te}} \mid \mathbf{X}_{\mathrm{te}} = X_{\mathrm{te}}, \mathbf{X}_{\mathrm{tr}} = X_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}} = Y_{\mathrm{tr}}) \\
=&p(\mathbf{Y}_{\mathrm{te}} = RY_{\mathrm{te}} \mid \mathbf{X}_{\mathrm{te}} = RX_{\mathrm{te}}, \mathbf{X}_{\mathrm{tr}} = RX_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}} = RY_{\mathrm{tr}})
\end{aligned}$$

Proof. The lemma is obtained by applying Bayes' rule and Lemma H.1 as

$$\begin{aligned}
&p(\mathbf{Y}_{\text{te}} = RY_{\text{te}} \mid \mathbf{X}_{\text{te}} = RX_{\text{te}}, \mathbf{X}_{\text{tr}} = RX_{\text{tr}}, \mathbf{Y}_{\text{tr}} = RY_{\text{tr}}) \\
&= \frac{p(\mathbf{Y}_{\text{te}} = RY_{\text{te}}, \mathbf{X}_{\text{te}} = RX_{\text{te}}, \mathbf{X}_{\text{tr}} = RX_{\text{tr}}, \mathbf{Y}_{\text{tr}} = RY_{\text{tr}})}{\int p(\mathbf{Y}_{\text{te}} = RY'_{\text{te}}, \mathbf{X}_{\text{te}} = RX_{\text{te}}, \mathbf{X}_{\text{tr}} = RX_{\text{tr}}, \mathbf{Y}_{\text{tr}} = RY_{\text{tr}}) dRY'_{\text{te}}} \\
&= \frac{p(\mathbf{Y}_{\text{te}} = Y_{\text{te}}, \mathbf{X}_{\text{te}} = X_{\text{te}}, \mathbf{X}_{\text{tr}} = X_{\text{tr}}, \mathbf{Y}_{\text{tr}} = Y_{\text{tr}})}{\int p(\mathbf{Y}_{\text{te}} = Y'_{\text{te}}, \mathbf{X}_{\text{te}} = X_{\text{te}}, \mathbf{X}_{\text{tr}} = X_{\text{tr}}, \mathbf{Y}_{\text{tr}} = Y_{\text{tr}}) dY'_{\text{te}}} \\
&= p(\mathbf{Y}_{\text{te}} = Y_{\text{te}} \mid \mathbf{X}_{\text{te}} = X_{\text{te}}, \mathbf{X}_{\text{tr}} = X_{\text{tr}}, \mathbf{Y}_{\text{tr}} = Y_{\text{tr}}).
\end{aligned}$$

□

Proof of the equivariance proposition:

We now prove the proposition. Recall that

$$\mu_{F|D}(\mathbf{X}_{\text{te}}, \mathbf{X}_{\text{tr}}, \mathbf{Y}_{\text{tr}}) = \mathbb{E}[\mathbf{Y}_{\text{te}} \mid \mathbf{X}_{\text{te}} = X_{\text{te}}, \mathbf{Y}_{\text{tr}} = Y_{\text{tr}}, \mathbf{X}_{\text{tr}} = X_{\text{tr}}].$$

Therefore, for any R, we may compute $\mu_{F|D}(R\mathbf{X}_{\text{te}}, R\mathbf{X}_{\text{tr}}, R\mathbf{Y}_{\text{tr}})$ as

$$\begin{aligned}
\mu_{F|D}(R\mathbf{X}_{\text{te}}, R\mathbf{X}_{\text{tr}}, R\mathbf{Y}_{\text{tr}}) &= \mathbb{E}[\mathbf{Y}_{\text{te}} \mid \mathbf{X}_{\text{te}} = RX_{\text{te}}, \mathbf{Y}_{\text{tr}} = RY_{\text{tr}}, \mathbf{X}_{\text{tr}} = RX_{\text{tr}}] \\
&= \int Y_{\text{te}}\, p(\mathbf{Y}_{\text{te}} = Y_{\text{te}} \mid \mathbf{X}_{\text{te}} = RX_{\text{te}}, \mathbf{Y}_{\text{tr}} = RY_{\text{tr}}, \mathbf{X}_{\text{tr}} = RX_{\text{tr}}) dY_{\text{te}} \\
&= \int RY_{\text{te}}\, p(\mathbf{Y}_{\text{te}} = RY_{\text{te}} \mid \mathbf{X}_{\text{te}} = RX_{\text{te}}, \mathbf{Y}_{\text{tr}} = RY_{\text{tr}}, \mathbf{X}_{\text{tr}} = RX_{\text{tr}}) dY_{\text{te}} \\
&= \int RY_{\text{te}}\, p(\mathbf{Y}_{\text{te}} = Y_{\text{te}} \mid \mathbf{X}_{\text{te}} = X_{\text{te}}, \mathbf{Y}_{\text{tr}} = Y_{\text{tr}}, \mathbf{X}_{\text{tr}} = X_{\text{tr}}) dY_{\text{te}} \\
&= \mathbb{E}[R\mathbf{Y}_{\text{te}} \mid \mathbf{X}_{\text{te}} = X_{\text{te}}, \mathbf{Y}_{\text{tr}} = Y_{\text{tr}}, \mathbf{X}_{\text{tr}} = X_{\text{tr}}] \\
&= R\mu_{F|D}(\mathbf{X}_{\text{te}}, \mathbf{X}_{\text{tr}}, \mathbf{Y}_{\text{tr}})
\end{aligned}$$

Where in the third line we perform a change of variables, noting that $|R| = 1$. The fourth line follows from Lemma H.2. The final line is a result of linearity of expectation and the definition of $\mu_{F|D}$, and provides the desired equality.

H.4. Non-Equivariance of Velocity GP predictions

In this appendix, we show that the velocity GP requires special constraints to exhibit reference-frame equivariance, and these constraints force an undesirable coupling of divergence and vorticity length scales.

Proposition 4.3. *For isotropic component kernels and zero prior mean, the velocity GP is reference-frame equivariant if and only if the kernels for each component are equal.*

Proof. We first show that if the velocity GP is reference-frame equivariant and has isotropic kernels for each component, then the kernels for the two velocity components are equal. In order to show this, it suffices to show that the prior is not equivariant, as this is the posterior given the empty dataset. Let $F \sim GP(0, k)$ be a function from $\mathbb{R}^2 \to \mathbb{R}^2$ by stacking $F(\cdot) = \left[F^{(1)}(\cdot), F^{(2}(\cdot)\right]^T$. Rotational invariance of the prior is equivalent to the condition that for an arbitrary 2-dimensional rotation matrix $R, R^{-1}F(R\cdot) = F(\cdot)$, where equality is in distribution for an entire sample path. Consider a 90 degree rotation of the coordinate axis. In the case of a 90 degree rotation, by isotropy of the kernels, $R^{-1}F(R\cdot) = R^{-1}F(\cdot) = [F_2(\cdot), F_1(\cdot)]^T$. Considering each coordinate of $R^{-1}F(R\cdot)$, equality in distribution to F implies that $F^{(2)}(\cdot) = F^{(1)}(\cdot)$ in distribution, and so the two components must have the same kernel.

We next show that if the two kernels are equal, then the velocity GP is rotationally equivariant. By Lemma H.2, it suffices to show there exist kernels k_Φ, k_Ψ such that the prior is equal in distribution to a Helmholtz GP with these kernels. Let F denote a zero mean velocity GP with kernels $k^{(1)} = k^{(2)}$. Let F' denote a Helmholtz GP with kernels $k_\Phi = k_\Psi$, and $k_\Phi(\mathbf{x}, \mathbf{x}') = \int_{s'=0}^{x^{(1)}} \int_{s=0}^{(x')^{(1)}} k^{(1)}(s, s') ds ds' + \int_{s'=0}^{x^{(2)}} \int_{s=0}^{(x')^{(2)}} k^{(2)}(s, s') ds ds'$. Applying Proposition 3.1 and the fundamental theorem of calculus, we see the covariance functions of F and F' are equal, and since F and F' are zero mean Gaussian processes, they are therefore equal in distribution.

I. Experimental Results

In this section, we provide more details on our experimental results. The section is organized in three parts. Appendix I.1 focuses on experiments with simulated data. Appendix I.2 focuses on experiments with real data from the LASER experiment (D'Asaro et al., 2017). Appendix I.3 focuses on real data from the GLAD experiment (Özgökmen, 2012). In each section, we have one subsection for each experiment. These subsections provide simulation details (e.g., what is the underlying vector field, and how we generated the buoys trajectories), model fitting details (e.g., hyperparameter optimization), and results. At the end of each subsection, we include a figure with these results. All the figures have the same structure. The first column represents ground truths. Second, third and fourth columns contain, respectively, SE-Helmholtz GP, SE-velocity GP, and D-HNN results. The first two rows represent results for the velocity prediction task: row 1 shows reconstructed velocity fields, row 2 differences from ground truth. Rows 3, 4, and 5 are about divergence: first divergence predictions, then standard deviation and z-values for the two GP models. Finally, rows 6, 7, and 8 concern vorticity: vorticity predictions, standard deviation and z-values for the two GP models. See Figure 5 for an example. For real data experiments, where we do not have ground truths, we omit the first column and the second row, i.e., all plots involving comparisons with ground truth quantities. See Figure 13 for an example.

For the simulated experiments, all root mean square errors are evaluated on the grids used to simulate the experiment. Specific grids are discussed in the "simulation details" paragraph of each individual experiments subsection. More explicitly, the root mean square error is calculated as,

$$\text{RMSE} = \sqrt{\frac{1}{|L|}\sum_{\mathbf{x}\in L}\|F(\mathbf{x}) - \hat{F}(\mathbf{x})\|_2^2} \tag{28}$$

where $F(\mathbf{x})$ denotes the simulated vector field, $\hat{F}(\mathbf{x})$ denotes the predictions of a given model and L is the grid used to simulate the vector field.

Initialization In all experiments except the GLAD data, we initialize our parameters such that $\log \ell_\Phi = 0, \log \sigma_\Phi = 0, \log \ell_\Psi = 1, \log \sigma_\Psi = -1, \log \sigma^2_{\text{obs}} = -2$ and likewise $\log \ell_1 = 0, \log \sigma_1 = 0, \log \ell_2 = 1, \log \sigma_2 = -1, \log \sigma^2_{\text{obs}} = -2$. We describe the special difficulties of the GLAD data in Appendix I.3. In all other cases, we found that results were not sensitive to initialization.

I.1. Simulated Experiments

We focus on simulations of key ocean behaviors of interest to oceanographers: vortices, straight currents, concentrated divergences, and combinations thereof.

I.1.1. Simulated experiment 1: single vortex

A single vortex in the ocean is a fluid flow pattern in which water particles rotate around a central point, with the flow pattern resembling a spiral. These vortices can occur due to a variety of factors such as the wind, currents, and tides. Single ocean vortices, also known as ocean eddies, can have a significant impact on ocean circulation and can transport heat, salt, and nutrients across vast distances. They can also affect the distribution of marine life. The vortex constructed has zero divergence and constant vorticity.

Simulation details. To simulate a vortex vector field in a two dimensional space, we first define a grid of points L of size 17 x 17, equally spaced over the interval $[-1, 1] \times [-1, 1]$. For each point $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in L$, we compute the vortex longitudinal and latitudinal velocities by:

$$F^{(1)}(\boldsymbol{x}) = -x^{(2)}$$
$$F^{(2)}(\boldsymbol{x}) = x^{(1)}$$

From these equations we obtain that the divergence of the vortex is 0 for any $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in L$:

$$\delta(\boldsymbol{x}) = \text{div} \cdot F = \frac{\partial F^{(1)}}{\partial x^{(1)}} + \frac{\partial F^{(2)}}{\partial x^{(2)}} = 0 + 0 = 0$$

and the vorticity is -2 for any $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in L$:

$$\zeta(\boldsymbol{x}) = \text{curl} \cdot F = \frac{\partial F^{(1)}}{\partial x^{(2)}} - \frac{\partial F^{(2)}}{\partial x^{(1)}} = -1 - 1 = -2$$

In our simulated experiment, we then use this vector field to simulate buoys trajectories, i.e. the evolution of buoys positions and velocities across time. In doing so, we make an implicit stationarity assumption about the vector field. That is, we assume that across the total time where we want to simulate buoys trajectories, the vector field remains the same. Then we fix starting positions for the desired amount of buoys, in this case 4. We set these to be just on one side of the vortex, to evaluate the ability of the models to reconstruct the full vortex by having access to observations covering only a portion of it. We pick the total time (here 1) for which we observe the trajectories, and the amount of time steps at which we want to observe the buoys trajectories (here 2), to split the total time. To find the trajectories, we solve the velocity-time ordinary differential equation, $d\boldsymbol{x}/dt = F$, where d/dt represents the time-derivative operator. Once we obtain the evolution of buoys' locations, we obtain the corresponding velocities by doing a linear interpolation of the underlying vortex field. By doing this interpolation, we end up with our simulated dataset, consisting in this case of 8 observations.

Model fitting. We are interested in evaluating the models' capabilities of reconstructing the full vortex, and capturing the underlying divergence and vorticity structure. To do so, we consider test locations corresponding to the grid L, so that we can compare our results with the ground truth, for velocities, divergence, and vorticity. To fit the SE-Helmholtz GP, we initialize the hyperparameters as explained in the Initialization paragraph at the start of this appendix: $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.368, \sigma^2_{\text{obs}} = 0.135$. The objective function of our optimization routine is the log marginal likelihood from Equation (5). We optimize the parameter using the gradient-based algorithm Adam (Kingma & Ba, 2015). Note that we optimize the hyperparameters in the log-scale. That is, we consider as parameters in the optimization step $\log \ell_\Phi, \log \sigma_\Phi, \log \ell_\Psi, \log \sigma_\Psi$, and $\log \sigma^2_{\text{obs}}$, and we exponentiate these when evaluating the log marginal likelihood. In doing so, we ensure that the optimal parameters are positive, as needed in this model. We run the optimization routine until the algorithm reaches convergence. In this case, the convergence criterion is the difference of log marginal likelihood in two consecutive optimization steps being less than 10^{-4}. This convergence is achieved in less than 1000 iterations. The optimal hyperparameters are: $\ell_\Phi = 1.1131, \sigma_\Phi = 0.0342, \ell_\Psi = 1.5142, \sigma_\Psi = 0.8884, \sigma^2_{\text{obs}} = 0.1597$. The same optimization routine is performed for the SE-velocity GP. In this case, the initial hyperparameters are $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.368, \sigma^2_{\text{obs}} = 0.135$. The optimal hyperparameters are: $\ell_1 = 1.6191, \sigma_1 = 0.9710, \ell_2 = 2.7183, \sigma_2 = 0.5811, \sigma^2_{\text{obs}} = 0.1759$. For both optimization routines, we tried different initial parametrizations, and the results agree substantially both in terms of RMSEs and visual reconstruction. Finally, to train the D-HNN model, we run the training routine provided in Greydanus & Sosanya (2022) code.

Results. We show the results in Figure 5. For each of the plots, the horizontal and vertical axes represent, respectively, latitude and longitude. The first row represents the ground truth simulated vector field (left), and the reconstruction using the SE-Helmholtz GP (center-left), the SE-velocity GP (center-right) and the D-HNN (right). Red arrows are the observed buoy data, black arrows show the predicted current at test locations. We can see how our method predicts a full vortex covering the spatial domain, whereas the SE-velocity GP predicts a smooth curve with much longer length scale, that does not resemble the ground truth. The D-HNN prediction looks more similar to a vortex, but still not as good as the SE-Helmholtz GP. To support this claim we also show differences from the ground truth in the second row. Finally, note that the RMSE for the SE-Helmholtz GP is 0.24, whereas for the SE-velocity GP it is 0.72 and for the D-HNN is 0.54.

In the third row, we analyze the divergence. The left box shows the constantly zero ground truth. Our model prediction (center-left) correctly captures this behavior, whereas the SE-velocity GP (center-right) predicts an irregular pattern not resembling the truth. The same happens for D-HNN (right box). In the fourth row we show the standard deviation of divergence predictions for the two GP models, and we can see how the SE-Helmholtz GP is very certain that there is no divergence, whereas the uncertainty for the SE-velocity GP predictions is higher. Finally, in the fifth row, we show the z-values for the divergence prediction, defined as the ratio between the mean and the standard deviation. This is a measure of how far from zero the prediction is, measured in terms of standard deviation. Some standard cut-off values for this quantity are -1 and $+1$, and one usually concludes that the prediction is significantly different (in the sense of one standard deviation) from 0 if the corresponding z-value is beyond these thresholds. By using this indicator, we conclude that none of the two predictions are significantly far from zero, so both models are accurate in predicting zero divergence, but our prediction is more precise, in the sense that the mean is closer to the real value and the uncertainty is lower. This is confirmed by looking at RMSEs: 0.0 for the SE-Helmholtz GP, 0.22 for SE-velocity GP, and 0.87 for the D-HNN.

Finally, in the last three rows we analyze results for the vorticity. The left box shows the constant (-2) ground truth. The SE-Helmholtz GP (center-left) predicts that the vorticity is around that value, especially in the center of the vortex, whereas in the corners the behavior is not as good. The SE-velocity GP (center-right) performs much worse also here, by predicting vorticity very close to zero, or positive, on almost all the spatial domain. The D-HNN (right box) predicts negative vorticity in most of the domain, but the pattern is very irregular. In the second-to-last row we show the standard deviation of divergence predictions for the two GP models, and we can see how the range of uncertainties on this task is more similar than before, meaning that there are areas where both models are not very confident. Still, if we look at the z-values in the last row, combined with the prediction plots, we see our model is better at predicting the magnitude and size of the vorticity area. In terms of RMSEs, we have 0.77 for the SE-Helmholtz GP, 1.05 for the SE-velocity GP, and 1.03 for the D-HNN.

In general, in this experiment we have shown that when working with this very simple underlying vector field, our model behaves better than the alternatives. In particular, we have seen how the prediction of the vortex is very accurate for the SE-Helmholtz GP, whereas the two other models are more off (and this is reflected in the respective RMSEs). In terms of divergence, our model predicts with certainty that there is no divergence, whereas the SE-velocity GP approach is less precise (by predicting non-zero divergence with high uncertainty). Finally, we saw how in terms of vorticity our model is the only one able to understand that there is a non-zero vorticity: even if the prediction is not perfect, it is still significantly better than all the other models.

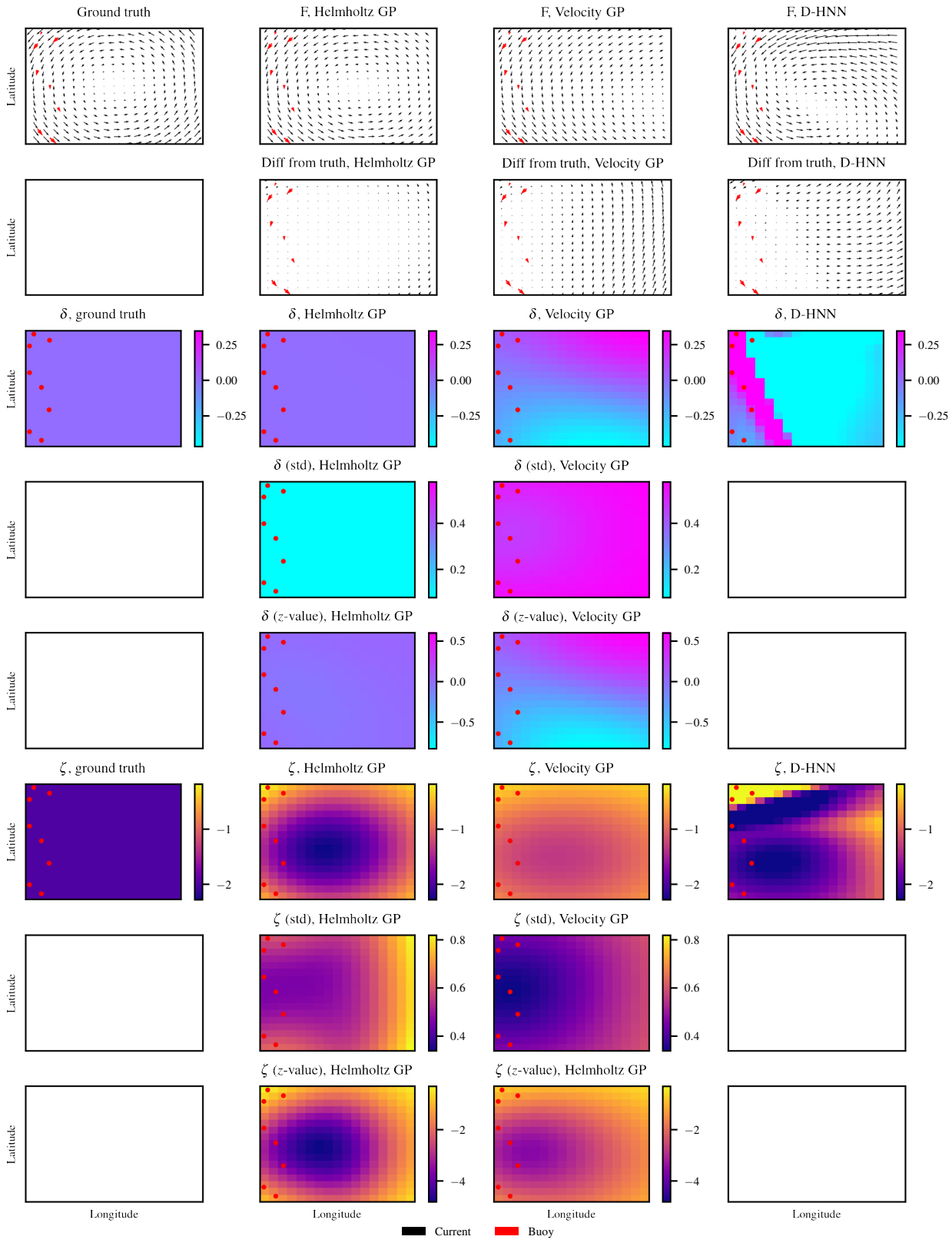

Figure 5. Single vortex. First column: ground truths. Second column: SE-Helmholtz GP results. Third column: SE-velocity GP results. Fourth column: D-HNN results.

I.1.2. SIMULATED EXPERIMENT 2: VORTEX ADJACENT TO STRAIGHT CURRENT

This task elaborates on the previous one by splitting the spatial domain in two regions: the upper region has a single vortex, and the lower region has a constant and straight flow from left to right.

The velocity field is discontinuous at the boundary, and therefore the divergence and vorticity are not defined as fields along this line. In general, though, if we were to consider a smaller sub-region containing the boundary on the lefthand side (where the vortex has downward velocity), we expect such a sub-region to have a negative flux. Analogously, if we were to consider a sub-region containing the boundary on the righthand side (where the vortex has upward velocity), we would expect that sub-region to have a positive flux.

Since the divergence and vorticity fields are not defined everywhere, it is not obvious what the desired behavior is in terms of recovering these fields. Due to this ambiguity, we do not report the results of this experiment in the main text. However, these discontinuities are reminiscent of fronts, which are of substantial interest to oceanographers, so we include this experiment in this appendix for completeness and in case it spurs future advancements.

Simulation details. To simulate such a vortex vector field, we first define a grid of points L of size 25 x 50, equally spaced over the interval $[-1, 1] \times [-1, 2]$. We can see this grid as composed of two subgrids L_1 and L_2, each of dimension 25 x 25, with L_1 representing the top grid and L_2 the lower one. Next, for each point $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in L_1$, we compute the vortex as done in Appendix I.1.1:

$$F^{(1)}(\boldsymbol{x}) = -x^{(2)}$$
$$F^{(2)}(\boldsymbol{x}) = x^{(1)}$$

and we still have $\delta(\boldsymbol{x}) = 0$ and $\zeta(\boldsymbol{x}) = -2$ for any $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in L_1$.

For each point $\boldsymbol{x} \in L_2$, we simulate a constant field with the following equations:

$$F^{(1)}(\boldsymbol{x}) = 0.7$$
$$F^{(2)}(\boldsymbol{x}) = 0.$$

The divergence and vorticity for each $\boldsymbol{x} \in L_2$ are $\delta(\boldsymbol{x}) = 0$ and $\zeta(\boldsymbol{x}) = 0$.

As done for the previous experiment, we then use this vector field to simulate buoys trajectories making the stationarity assumption. Here we consider 7 buoys, covering the full region, observed for a total time of 0.5 and 2 time steps. We reconstruct the buoys trajectories by solving the ODE and interpolating as specified before. By doing this interpolation, the simulated dataset consists of 14 observations.

Model fitting. We fit the three models with the routine specified in Appendix I.1.1. To fit the SE-Helmholtz GP, we initialize the hyperparameters as follows: $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.368, \sigma^2_{\text{obs}} = 0.135$. The optimal hyperparameters are: $\ell_\Phi = 3.8698, \sigma_\Phi = 0.0885, \ell_\Psi = 1.2997, \sigma_\Psi = 0.9773, \sigma^2_{\text{obs}} = 0.0609$. The same optimization routine is performed for the SE-velocity GP. In this case, the initial hyperparameters are $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.368, \sigma^2_{\text{obs}} = 0.135$. The optimal hyperparameters are: $\ell_1 = 0.9397, \sigma_1 = 1.0755, \ell_2 = 2.7183, \sigma_2 = 0.5528, \sigma^2_{\text{obs}} = 0.0087$. For both optimization routines, we tried different initial parametrizations, and the results agree substantially both in terms of RMSEs and visual reconstruction.

Results. We show the results in Figure 6. As before, for each of the plots, the horizontal and vertical axes represent, respectively, latitude and longitude. The first row represents the ground truth simulated vector field (left), and the reconstruction using the SE-Helmholtz GP (center-left), the SE-velocity GP (center-right) and the D-HNN (right). Red arrows are the observed buoy data, black arrows show the predicted current at test locations. We can see how our method predicts accurately the vortex structure, whereas it has some problems in the lower right corner, and smooths out the discontinuity at the boundary. The SE-velocity GP is accurate as well for the vortex part, but has a significant issue in the lower subgrid: the current flows from left to right, then gets interrupted, and then restarts in a different direction. This behavior goes against the idea that currents are continuous (by conservation of momentum). The D-HNN predictions look very similar to the SE-Helmholtz GP. In the second row we include the differences from the ground truth, and these show as well that SE-Helmholtz GP and D-HNN are accurate, whereas the SE-velocity GP has issues in the lower part of the grid. The RMSE for the SE-Helmholtz GP is 0.30, whereas for the SE-velocity GP it is 0.49 and for the D-HNN is 0.28.

In the third row, we analyze the divergence. The left box shows the constantly zero ground truth except for at the boundary line, where the divergence is undefined. Our model captures the zero divergence outside the boundary line, but does not estimate any divergence at or around the boundary line. The SE-velocity GP (center-right) estimates an irregular pattern. The D-HNN divergence estimate has the perhaps desirable property that it is negative on the left side and positive on the right side, which might lead to reasonable predictions of flux into and out of regions containing the boundary between the straight current and vortex. As it is unclear what a reasonable approximation to the divergence is when the divergence is concentrated on a line, we do not report RMSE on this example.

Finally, in the last three rows we analyze results for the vorticity. The left box shows the ground truth, again ignoring the boundary line where the vorticity is undefined. Here both the GP models' predictions look very similar. Both predict that there is a negative vorticity area in the top grid, and a close-to-zero vorticity area in the lower grid. The D-HNN vorticity estimates in the upper and lower regions are similar to those of the GPs. Also the D-HNN estimate suggests a crisper boundary than the GP approaches do. In the standard deviation and z-value plots, we see that both GP models seem quite certain about the existence of a negative vorticity at least in part of the upper region. We again do not report RMSE for vorticity in this example as the vorticity is not well-defined over the entire region we consider.

In summary, in this experiment we showed a situation in which the SE-Helmholtz GP is at least as good as the other two models in predicting the velocity field. It is not entirely clear what desirable reconstruction of the divergence and vorticity field is, as they are not defined on the boundary of the two regions, and the models show substantially different properties. Finally, we note that this field violates the modeling assumptions made by the the two GP models (particularly continuity) and further modeling innovation is likely needed to improve fidelity in examples like this one.

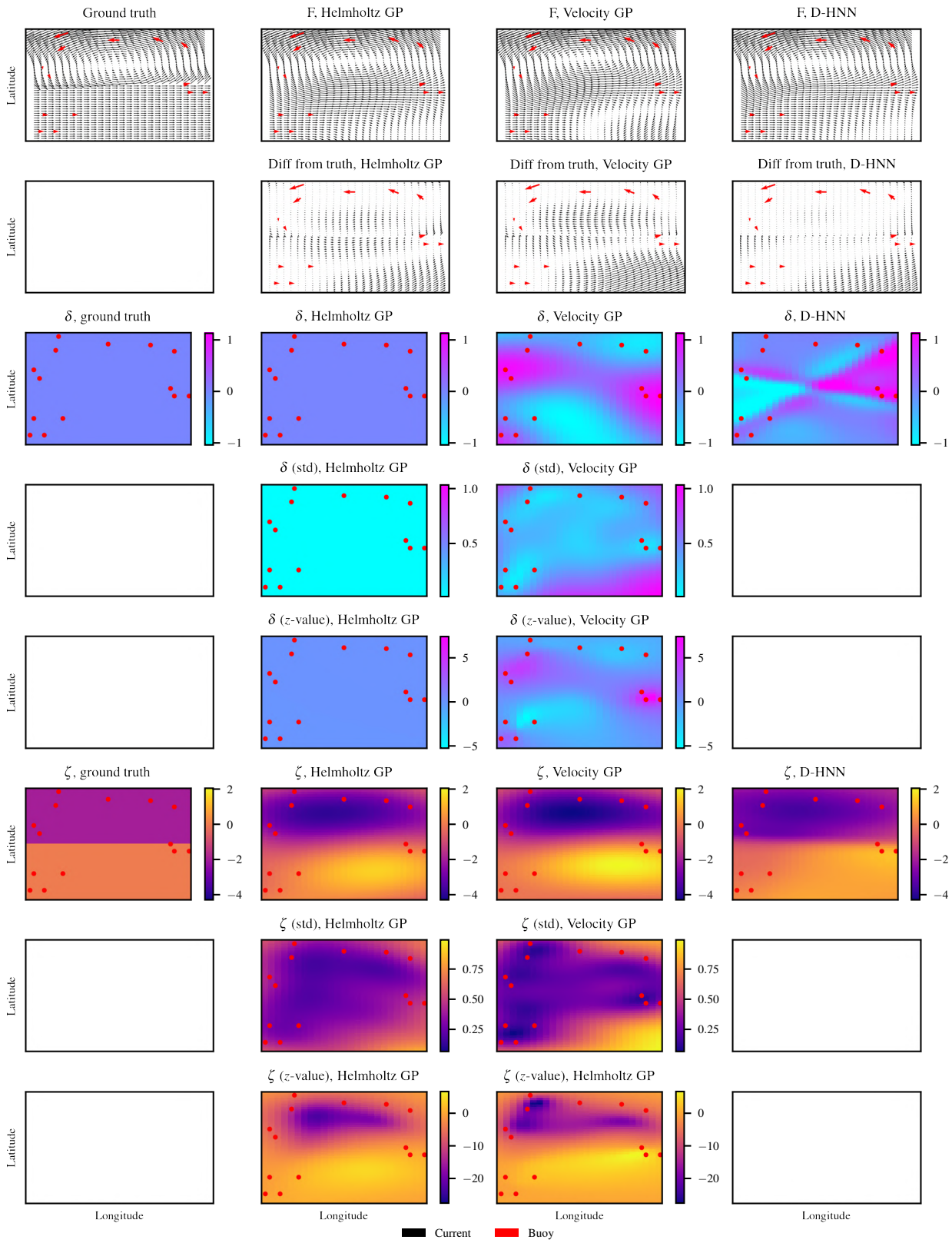

Figure 6. Vortex adjacent to straight current. First column: ground truths. Second column: SE-Helmholtz GP results. Third column: SE-velocity GP results. Fourth column: D-HNN results.

I.1.3. SIMULATED EXPERIMENT 3: AREAS OF CONCENTRATED DIVERGENCE

For this experiment, we consider three different scenarios, differing only in the size of the divergence area. In each of these, we observe 5 buoys, each for 2 time steps. For all the scenarios, the two models perform very well. We show the results in Figures 7 to 9. Both models reconstruct the velocity field well, see Table 2 for more details. Moreover, this field has no vorticity and divergence that peaks at the center of the region and slowly decreases in a circular way. This behavior is captured by both models in an accurate way. We conclude that in this important simulated experiment our model is at least as good as the SE-velocity GP approach.

A vector field with a single diffuse area of divergence simulates the behavior of an ocean fluid flow in which the water particles are spreading out from a particular region. This behavior can be caused by a variety of factors, such as the movement of warm and cold water masses, or the mixing of fresh and salt water, and can lead to increase in nutrient concentration and high primary production and biodiversity. The fluid particles in this area are not rotating in a circular motion as in a vortex, but instead moving away from each other, resulting in a decrease in density and velocity.

Simulation details. To simulate a vector field with a divergence area in a two dimensional space, we first define a grid of points L of size 20 x 20, equally spaced over the interval $[-2, 2] \times [-2, 2]$. The point $(0, 0)$ represents the center of the divergence area. To obtain a vector field with divergence area around this point, for each point $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in L$, we can compute the longitudinal and latitudinal velocities by:

$$F^{(1)}(\boldsymbol{x}) = \frac{x^{(1)}}{b_d + R_d^2(\boldsymbol{x})}$$
$$F^{(2)}(\boldsymbol{x}) = \frac{x^{(2)}}{b_d + R_d^2(\boldsymbol{x})}$$

with $R_d(\boldsymbol{x}) = ((x^{(1)})^2 + (x^{(2)})^2$ being the distance from the center of divergence, and b_d a parameter governing the size of the area of divergence. Larger b_d implies larger area, but also smaller value at the center. Intuitively, this parameter measures how diffuse the divergence around a center point is. This intuition can be confirmed by computing the actual divergence value:

$$\delta(\boldsymbol{x}) = \text{div} \cdot F = \frac{\partial F^{(1)}}{\partial x^{(1)}} + \frac{\partial F^{(2)}}{\partial x^{(2)}} = \frac{b_d + (x^{(1)})^2 - (x^{(2)})^2}{(b_d + R_d^2(\boldsymbol{x}))^2} + \frac{b_d - (x^{(1)})^2 + (x^{(2)})^2}{(b_d + R_d^2(\boldsymbol{x}))^2} = \frac{2b_d}{(b_d + R_d^2(\boldsymbol{x}))^2}.$$

For the vorticity instead we have

$$\zeta(\boldsymbol{x}) = \text{curl} \cdot F = \frac{\partial F^{(1)}}{\partial x^{(2)}} - \frac{\partial F^{(2)}}{\partial x^{(1)}} = \frac{-2(x^{(1)})(x^{(2)})}{(b_d + R_d^2(\boldsymbol{x}))^2} + \frac{2(x^{(1)})(x^{(2)})}{(b_d + R_d^2(\boldsymbol{x}))^2} = 0.$$

The goals for each model then are to (1) reconstruct the velocity field in an accurate way, (2) predict that there is a divergent area and its size, and (3) predict zero vorticity. Finally note that in this experiment, we propose three different scenarios, where the only difference is how diffuse the divergence areas are. Specifically, we run three different experiments with $b_{\text{small}} = 0.4$, $b_{\text{medium}} = 2$, and $b_{\text{big}} = 15$.

As before, our observations are simulated buoy trajectories. For each scenario the simulation part is the same. We simulate 5 buoys, starting in the non-divergent areas, observed for a total time of 3, and we consider 2 time steps. Overall we have 10 observations. As usual, to get these trajectories we solve the velocity-time ODE and interpolate.

Model fitting. For each of the three scenarios, we fit the three models with the routine specified in Appendix I.1.1. The hyperparameter initialization for both GPs is always the same across the three different scenarios: $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.368, \sigma^2_{\text{obs}} = 0.135$ for the SE-Helmholtz GP, $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.368, \sigma^2_{\text{obs}} = 0.135$ for the SE-velocity GP. We provide the optimal hyperparameters for each scenario in the corresponding subsections.

Result: small divergence area, $b_{\text{small}} = 0.5$. The optimal hyperparameters in this scenario are the following:

- $\ell_\Phi = 1.1314, \sigma_\Phi = 1.9422, \ell_\Psi = 5.3132, \sigma_\Psi = 0.1864, \sigma^2_{\text{obs}} = 0.1821$ for the SE-Helmholtz GP

- $\ell_1 = 0.5078, \sigma_1 = 1.6570, \ell_2 = 2.7183, \sigma_2 = 1.8658, \sigma^2_{\text{obs}} = 0.1396$ for the SE-velocity GP.

In Figure 7 we show the results of this scenario. As before, for each of the plots, the horizontal and vertical axes represent, respectively, latitude and longitude. The first row represents the ground truth simulated vector field (left), and the reconstruction using the SE-Helmholtz GP (center-left), the SE-velocity GP (center-right) and the D-HNN (right). Red arrows are the observed buoy data, black arrows show the predicted current at test locations. All three models have some problems in reconstructing the underlying field. The two GPs are particularly problematic, because they predict constant strong current that abruptly stops in regions where there are no buoys. The predictions are particularly bad for the SE-velocity GP, which fails to understand the direction and size of the current in most of the region. The D-HNN prediction is the one that looks better here, but it is still problematic in the sense that far away from the buoys the current starts to rotate. The plots in the second row showing the difference from the ground truth show that all these models provide poor performances on this task. In terms of RMSE, we have 1.11 for the SE-Helmholtz GP, 1.25 for the SE-velocity GP, and 0.67 for the D-HNN, confirming that our model performs much better.

In the third row, we analyze the divergence. The left box shows the divergence structure of this field. As described in the preamble, since b_d is small, we have a small area of divergence with big magnitude. The two GP models identify this area. The SE-velocity GP is more accurate in predicting the size of the divergence area. The SE-Helmholtz GP predicts that there is a divergence area in the middle and gets the correct magnitude, but predicts it to be larger than it actually is. If we consider the z-value plots, we can see that this intuition is confirmed: the SE-velocity GP predicts only a small area to have significant non-zero divergence, whereas our model overestimates the size of this area. The prediction of the D-HNN is less accurate. In terms of RMSEs, we have 2.62 for the SE-Helmholtz GP, 1.45 for the SE-velocity GP, and 4.14 for the D-HNN.

In the last three rows of the plot we have, as usual, the vorticity analysis. The left box shows the ground truth. Here the SE-Helmholtz GP perfectly predicts zero vorticity, and the D-HNN is almost correct too. The SE-velocity GP, on the contrary, predicts very irregular vorticity, with very high uncertainty. If we consider the z-value plots, we see there is one region (in the center) where the vorticity is predicted to be non-zero in a significant manner. This is a problematic behavior that the SE-velocity GP has and our model has not. We have 0.0 RMSE for the SE-Helmholtz GP, 1.07 for the SE-velocity GP, and 0.31 for the D-HNN.

Result: medium divergence area, $b_{\text{small}} = 5$. The optimal hyperparameters in this scenario are the following:

- $\ell_\Phi = 1.9387, \sigma_\Phi = 1.2387, \ell_\Psi = 2.3894, \sigma_\Psi = 0.2192, \sigma^2_{\text{obs}} = 0.0675$ for the SE-Helmholtz GP
- $\ell_1 = 1.6067, \sigma_1 = 0.8181, \ell_2 = 2.7183, \sigma_2 = 0.9859, \sigma^2_{\text{obs}} = 0.0742$ for the SE-velocity GP.

Figure 8 shows the results of this scenario. In the top part we have as always the velocity predictions. Since the divergence area is more diffuse, both the velocity of the current and the lengthscale of variation are smaller, in the sense that there are less sharp deviations. Compared to the previous scenario, this property of the field makes the prediction task easier for all three models. In particular, the SE-Helmholtz GP and SE-velocity GP predict a field that almost resembles identically the ground truth. The D-HNN still has some issues, specifically it predict some rotations far away from the observations. This behavior can be seen by looking at the difference from ground truth in the second row. We have the following RMSEs: 0.17 for the SE-Helmholtz GP, 0.19 for the SE-velocity GP, and 0.55 for the D-HNN.

For the divergence, by looking at the ground truth plot on the left, we see the area of divergence is now more diffuse, and the magnitude is lower. Both the SE-Helmholtz GP and the SE-velocity GP predict this area accurately, both in terms of size and magnitude (they both predict this area to be a bit larger than it actually is). The D-HNN picks up divergence in a very irregular way. In terms of uncertainty, both GP models are more certain about their predictions around the buoys, and the z-values reflect this behavior: the area where the divergence is significantly different from zero (z-value above 1) is almost identical to the actual ground truth. The RMSEs are: 0.39 for the SE-Helmholtz GP, 0.33 for the SE-velocity GP, 1.32 for the D-HNN.

For the vorticity, we observe that the performances of all models are now worse. The SE-Helmholtz GP still predicts vorticity very close to zero almost everywhere, but not exactly zero as before. The predictions for the SE-velocity GP still look less accurate and irregular. The D-HNN performance is very poor. In terms of uncertainty, the SE-Helmholtz GP has low uncertainty about its prediction, and this leads to an area where there is significantly non-zero vorticity (in terms of

z-value). This behavior is somehow problematic, but note that the predicted mean is in absolute value very close to zero in that area too. The z-value for the SE-velocity GP is as in the previous scenario, predicting significantly non-zero divergence in an area where the mean is quite distant from zero. Again, this is a very undesirable behavior. The RMSEs are: 0.05 for the SE-Helmholtz GP, 0.12 for the SE-velocity GP, 0.38 for the D-HNN.

Result: big divergence area, $b_{\text{small}} = 15$. We finally study the last scenario, with the big area of divergence. The optimal hyperparameters in this scenario are the following:

- $\ell_\Phi = 3.3732, \sigma_\Phi = 0.8362, \ell_\Psi = 14.7644, \sigma_\Psi = 0.0659, \sigma^2_{\text{obs}} = 0.0074$ for the SE-Helmholtz GP
- $\ell_1 = 2.3456, \sigma_1 = 0.3376, \ell_2 = 2.7183, \sigma_2 = 0.3355, \sigma^2_{\text{obs}} = 0.0055$ for the SE-velocity GP.

In Figure 9 we show the results of this scenario. Here the divergence areas are even more diffuse, and this seems to help a lot the SE-Helmholtz GP predictions but not so much the other methods.

For the velocity prediction task, the three models produce predictions that are close to the truth. It is clear, however, that the predictions of the SE-Helmholtz GP are more precise, whereas both the SE-velocity GP and D-HNN predict some rotational shapes that should not be there. This result is confirmed by the RMSE: 0.04 for the SE-Helmholtz GP, 0.10 for the SE-velocity GP, and 0.19 for the D-HNN.

In terms of divergence, predictions for the two GP models are similar, but our model is slightly better at predicting the full size of the region, with low uncertainty. The D-HNN prediction is again poor. The z-values show how in the central area, both models significantly predict non-zero divergence, but further away in the corners z-values get closer and closer to zero. This behavior is due to the distribution of the buoys' observations. The RMSEs are: 0.05 for the SE-Helmholtz GP, 0.12 for the SE-velocity GP, and 0.27 for the D-HNN.

Finally, if we consider the vorticity, we can see how here the SE-Helmholtz GP is superior to the other two methods, as in the two previous scenarios. It is able to detect that there is no vorticity, with very low uncertainty. The SE-velocity GP, on the contrary, predicts non-zero positive vorticity in the left side of the plot, and non-zero negative vorticity in the right side. These predictions are with low uncertainty and hence significant, as can be seen by looking at the z-values plot (most of the domain has z-values beyond the thresholds +1 and -1). The prediction with D-HNN is in similar to the SE-velocity GP one. The RMSEs are: 0.0 for the SE-Helmholtz GP, 0.10 for the SE-velocity GP, and 0.11 for the D-HNN.

In general, we saw how in these experiment the SE-Helmholtz GP is at least as good as the other two methods in almost all the velocity prediction tasks, as good as the SE-velocity GP for the divergence tasks, and remarkably better in predicting that there is no vorticity.

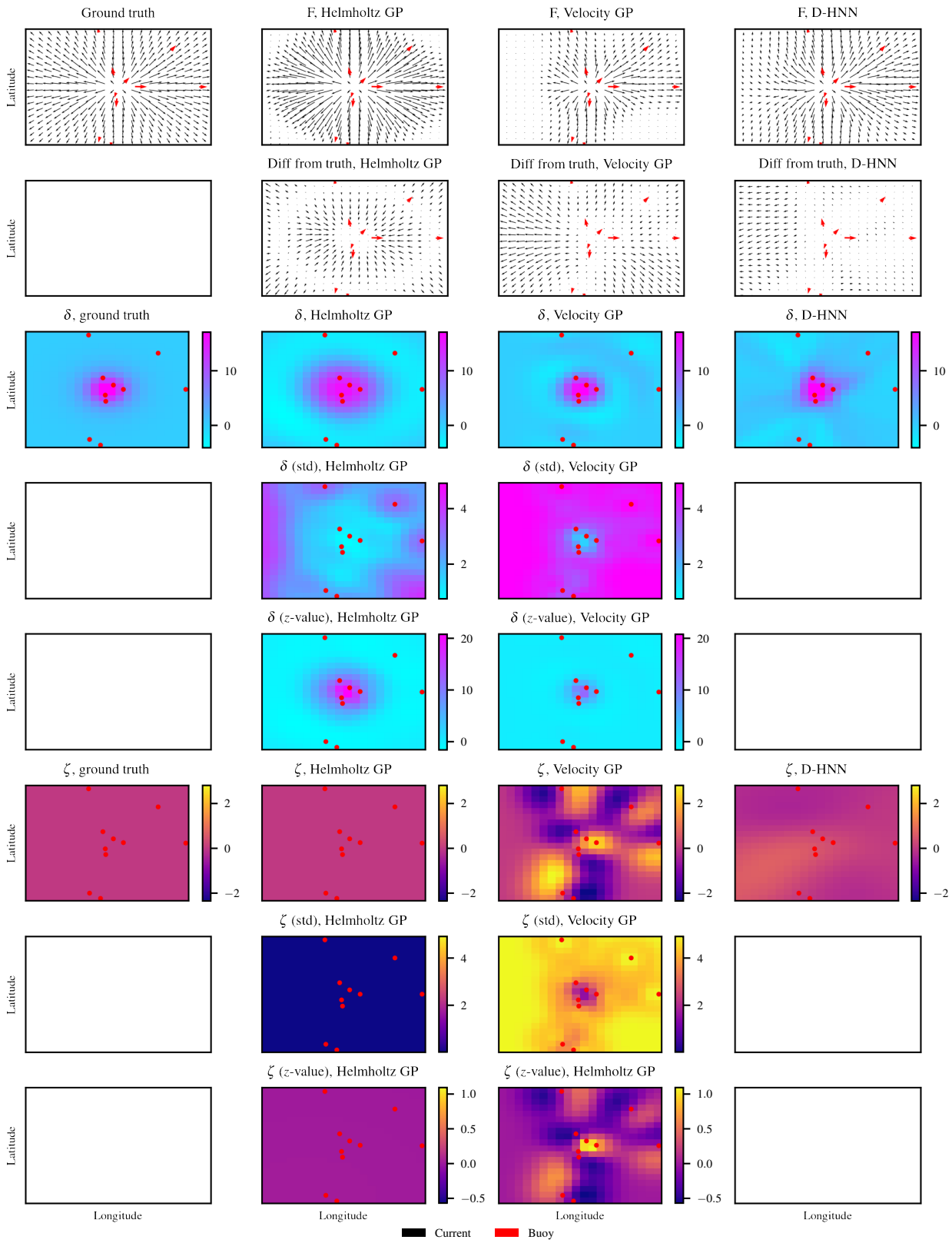

Figure 7. Small area of divergence. First column: ground truths. Second column: SE-Helmholtz GP results. Third column: SE-velocity GP results. Fourth column: D-HNN results.

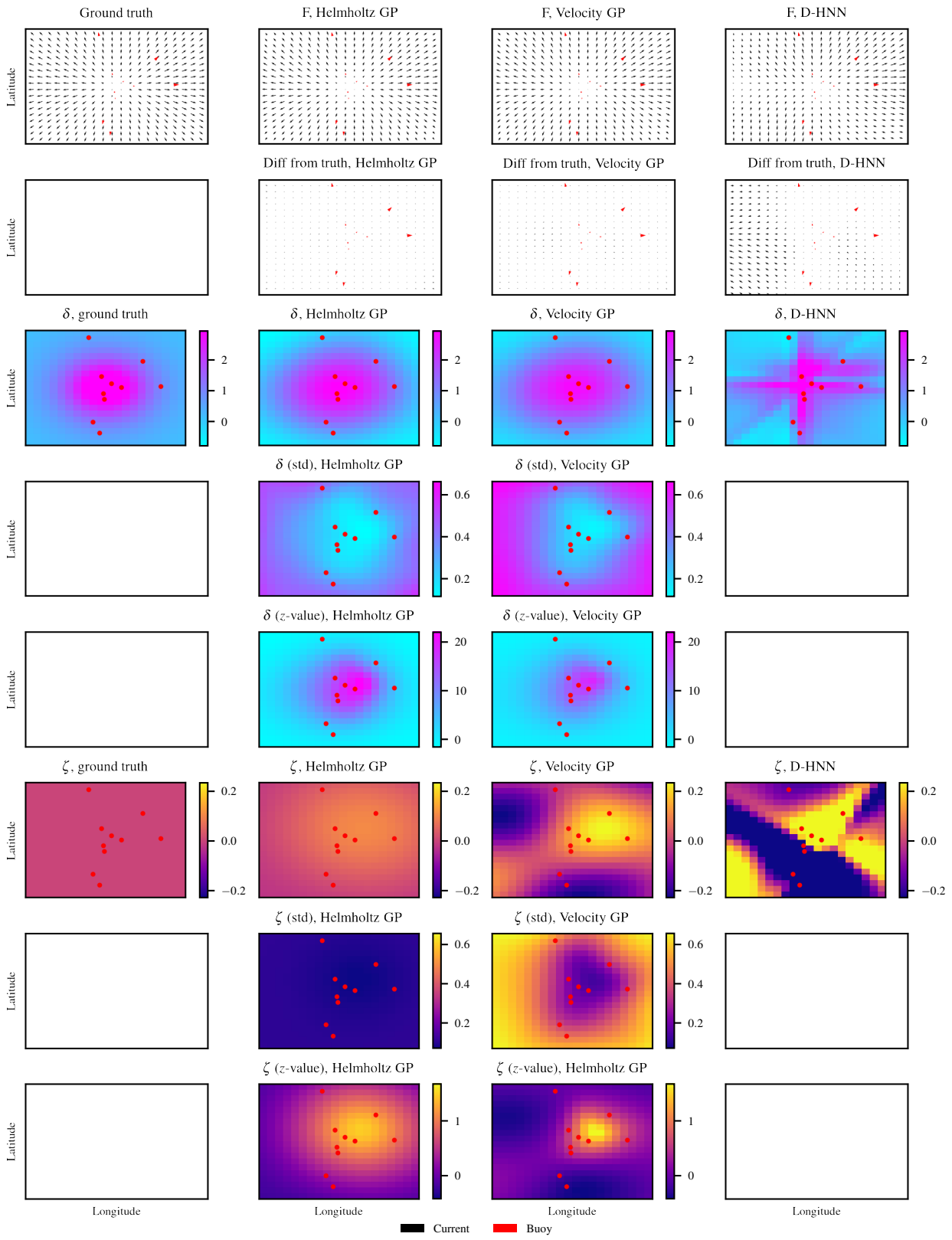

Figure 8. Medium area of divergence. First column: ground truths. Second column: SE-Helmholtz GP results. Third column: SE-velocity GP results. Fourth column: D-HNN results.

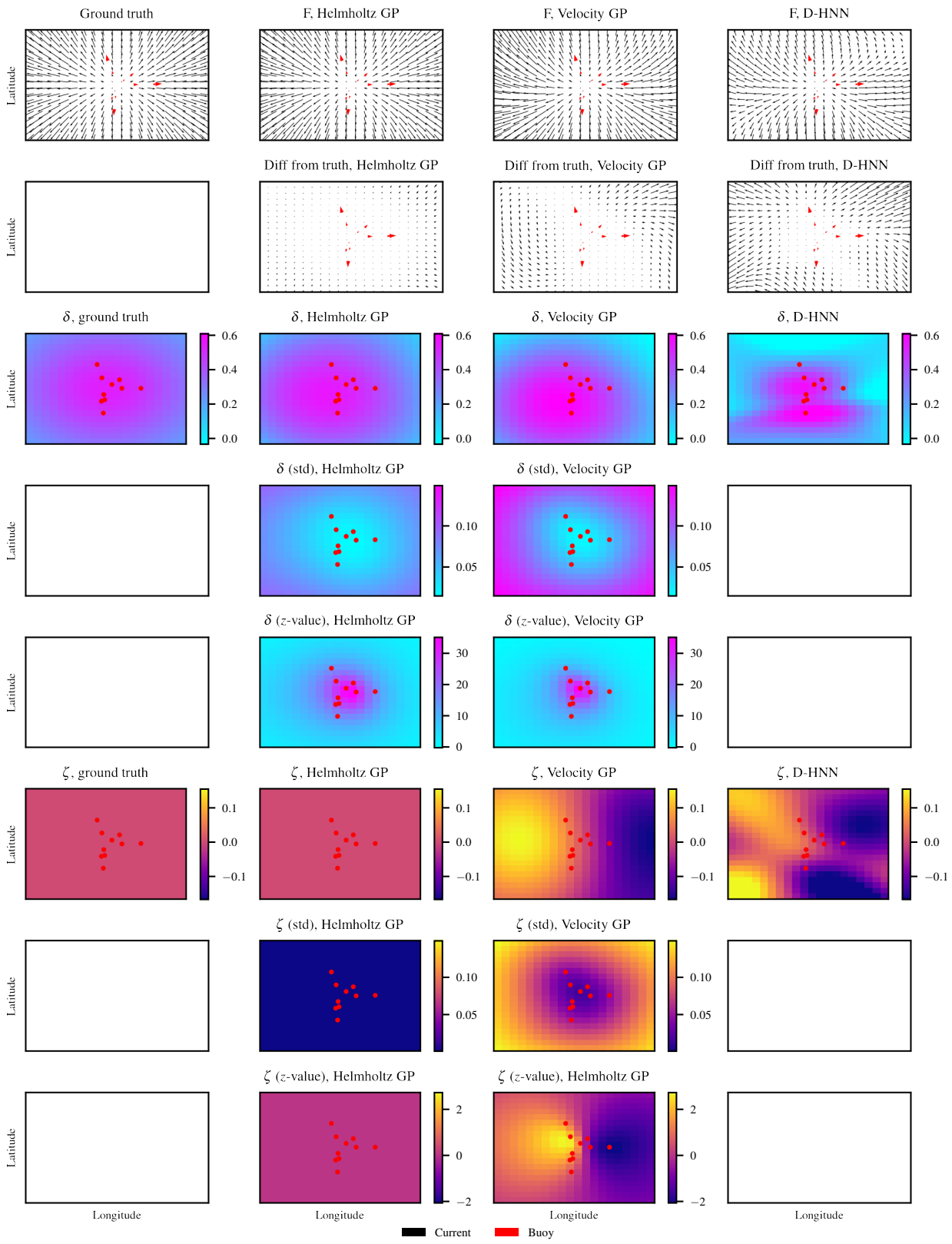

Figure 9. Big area of divergence. First column: ground truths. Second column: SE-Helmholtz GP results. Third column: SE-velocity GP results. Fourth column: D-HNN results.

I.1.4. SIMULATED EXPERIMENT 4: DUFFING OSCILLATOR WITH AREAS OF CONCENTRATED DIVERGENCE

The Duffing oscillator is a nonlinear dynamic system that can be used to study the dynamics of oceanic phenomena such as tides and currents. In this experiment, we add to this system a divergence area on the left region and a convergence area on the right one. See top-left plot in Figure 10. In this way we obtain a field that has both divergence (positive on the right, negative on the left), and vorticity (for the underlying Duffing system).

Simulation details. To simulate a Duffing oscillator in a two dimensional space, we first define a grid of points L of size 30 x 30, equally spaced over the interval $[-4, 4] \times [-4, 4]$. Next, for each point $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in L$, we compute the Duffing longitudinal and latitudinal velocities by:

$$\tilde{F}^{(1)}(\boldsymbol{x}) = x^{(2)}$$
$$\tilde{F}^{(2)}(\boldsymbol{x}) = (x^{(1)} - 0.1 * (x^{(1)})^3) * (1 + 0.1 * \cos(50 * \pi/4)).$$

On top of this field we add a divergent field at location $(-3, 0)$, using equations:

$$D^{(1)}(\boldsymbol{x}) = \frac{(x^{(1)} - (-3))}{b_d + R_d^2(\boldsymbol{x})}$$
$$D^{(2)}(\boldsymbol{x}) = \frac{x^{(2)}}{b_d + R_c^2(\boldsymbol{x})}$$

with $R_d(\boldsymbol{x}) = (x^{(1)} - (-3))^2 + (x^{(2)} - 0)^2$ being the distance from the center of divergence, and b_d a parameter governing the size of the area of divergence. Larger b_d implies larger area, but also smaller value at the center. It can be seen as a parameter measuring how diffuse the divergence around a center point is. We also have a convergent field around $(3, 0)$, determined by the equations:

$$C^{(1)}(\boldsymbol{x}) = -\frac{(x^{(1)} - 3)}{b_c + R_d^2(\boldsymbol{x})}$$
$$C^{(2)}(\boldsymbol{x}) = -\frac{(x^{(2)})}{b_c + R_c^2(\boldsymbol{x})}$$

with $R_c(\boldsymbol{x}) = (x^{(1)} - 3)^2 + (x^{(2)} - 0)^2$, the distance from the center of convergence. To get the full velocity field, we sum up these three quantities:

$$F^{(1)}(\boldsymbol{x}) = \tilde{F}^{(1)}(\boldsymbol{x}) + D^{(1)}(\boldsymbol{x}) + C^{(1)}(\boldsymbol{x})$$
$$F^{(2)}(\boldsymbol{x}) = \tilde{F}^{(2)}(\boldsymbol{x}) + D^{(2)}(\boldsymbol{x}) + C^{(2)}(\boldsymbol{x}).$$

In this system, the divergence and vorticity do not have a simple form, but can be calculated. For the sake of our divergence analysis, it is sufficient to say that there are two areas of interest, around the center of divergence and convergence. In this experiment, we propose three different scenarios, where the only difference is how diffuse the divergence areas are. For simplicity, we assume $b = b_c = b_d$, and we run three different experiments with $b_{\text{small}} = 0.5$, $b_{\text{medium}} = 5$, and $b_{\text{big}} = 15$.

As done before, to predict currents, divergence, and vorticity we simulate buoys. For each scenario the simulation part is the same. We first simulate 3 buoys, starting in the non-divergent areas, observed for a total time of 5, and 2 time steps. We then simulate 4 additional buoys, starting around the divergent areas, for a total time of 5, and 4 time steps. That is, we make observations coarser for buoys in these regions. Overall we have 22 observations. As usual, to get these observations we solve the velocity-time ODE and interpolate.

Model fitting. For each of the three scenarios, we fit the three models with the routine specified in Appendix I.1.1. The hyperparameter initialization for both GPs is the same across the three different scenarios: $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.368, \sigma^2_{\text{obs}} = 0.135$ for the SE-Helmholtz GP, $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.368, \sigma^2_{\text{obs}} = 0.135$ for the SE-velocity GP. We provide the optimal hyperparameters for each scenario in the corresponding subsections.

Result: small divergence area, $b_{\text{small}} = 0.5$. The optimal hyperparameters in this scenario are the following:

- $\ell_\Phi = 0.6335, \sigma_\Phi = 0.3734, \ell_\Psi = 3.9115, \sigma_\Psi = 6.9294, \sigma^2_{\text{obs}} = 0.0083$ for the SE-Helmholtz GP
- $\ell_1 = 0.7212, \sigma_1 = 1.8767, \ell_2 = 2.7183, \sigma_2 = 1.1361, \sigma^2_{\text{obs}} = 0.0084$ for the SE-velocity GP.

In Figure 10 we show the results of this scenario. As before, for each of the plots, the horizontal and vertical axes represent, respectively, latitude and longitude. The first row represents the ground truth simulated vector field (left), and the reconstruction using the SE-Helmholtz GP (center-left), the SE-velocity GP (center-right) and the D-HNN (right). Red arrows are the observed buoy data, black arrows show the predicted current at test locations. First of all, we can see how our method predicts accurately the duffing structure in the left part of the plot, whereas has some issues in the right one, where we have the convergence area. The SE-velocity GP prediction is more problematic: the correct current is predicted around the buoys, but farther away the prediction goes to zero, reverting to the prior mean. This is a problematic behavior, e.g., because it predicts very non-continuous currents. The D-HNN prediction is problematic as well: the current looks more continuous, but the general shape is very different from the ground truth. This behavior can be seen well from the second row, the comparison to the ground truth. In terms of RMSE, we have 0.96 for the SE-Helmholtz GP, 2.05 for the SE-velocity GP, and 2.14 for the D-HNN, confirming that our model performs much better.

In the third row, we analyze the divergence. The left box shows the divergence structure of this field. There is a small area with very positive divergence on the left, and a small area with very negative divergence on the right. The two GP models are good at identifying these areas. At the same time, they both predict some other areas of divergence around the observed buoys. Nonetheless, if we consider the z-value plots (on the fifth row) we can see how the z-values for both models are very high in the two areas of divergence, meaning that there is a strongly significant non-zero mean in those areas, as desired. The D-HNN predicts a quite different divergence structure. The RMSEs are: 0.94 for the SE-Helmholtz GP, 0.95 for the SE-velocity GP, and 1.89 for the D-HNN.

Finally, in the last three rows we analyze results for the vorticity. The left box shows the ground truth. Here the SE-Helmholtz GP prediction looks more accurate than the other two. Nonetheless, even our model is not fully able to capture the full vorticity structure. The predictions for the SE-velocity GP look particularly problematic because it is highly affected by the location of the buoys, and that is reflected in the uncertainty and z-values plots. The D-HNN predicts a very different field on this task as well. The RMSEs are: 1.40 for the SE-Helmholtz GP, 2.28 for the SE-velocity GP, and 2.64 for the D-HNN.

Result: medium divergence area, $b_{\text{small}} = 5$. The optimal hyperparameters in this scenario are the following:

- $\ell_\Phi = 1.2029, \sigma_\Phi = 0.1666, \ell_\Psi = 3.3679, \sigma_\Psi = 9.5514, \sigma^2_{\text{obs}} = 0.0112$ for the SE-Helmholtz GP
- $\ell_1 = 6.7677, \sigma_1 = 4.5316, \ell_2 = 2.7183, \sigma_2 = 23.3219, \sigma^2_{\text{obs}} = 0.0305$ for the SE-velocity GP.

Figure 11 shows the results of this scenario. In the top part we have as always the velocity predictions. In this case, the ground truth field is very similar to before, but the divergence areas are more diffuse, and hence the current has generally longer lengthscale of variation (that is, the deviations are less sharp). This feature helps the predictions for all three methods. We can see indeed how now the three models produce predictions that are closer to the truth than before. Still, by looking at the difference from ground truth plots, we can see that the prediction of our model is slightly better than the SE-velocity GP, and significantly better than the D-HNN. We have the following RMSEs: 0.19 for the SE-Helmholtz GP, 0.60 for the SE-velocity GP, and 1.65 for the D-HNN. These confirm what can see visually in the plots.

In terms of divergence, by looking at the the ground truth plot on the left, one can immediately notice how the areas of divergence are now more diffuse, and the magnitudes are lower. The SE-Helmholtz GP predicts accurately the two areas, with some noise in the central region. The SE-velocity GP is less accurate, but overall understand that there are these two areas. The D-HNN fails in identifying the two regions. It is interesting to observe the z-value plots in this experiment: for the SE-Helmholtz GP, the z-values are very high in the two desired areas, meaning that our model is very certain about divergence being different from zero in those areas. For the SE-velocity GP, the z-values still look good, just less accurate than for our model. The RMSEs are: 0.14 for the SE-Helmholtz GP, 0.50 for the SE-velocity GP, and 1.15 for the D-HNN.

Finally, we consider the vorticity. Here the two GP models agree significantly on the shape of their predictions, and they are both very similar to the ground truth. This result is reflected in the RMSEs: 0.24 for the SE-Helmholtz GP, 0.26 for the SE-velocity GP. The prediction for the D-HNN is far from the truth (RMSE 2.39). The uncertainty is lower close to

the data for both GP models. In general, both GP models seem to work well in recovering divergence and vorticity in this scenario. The SE-Helmholtz GP is superior for the divergence, the SE-velocity GP for the vorticity.

Result: big divergence area, $b_{\text{small}} = 15$. The optimal hyperparameters in this scenario are the following:

- $\ell_\Phi = 2.9194, \sigma_\Phi = 0.4599, \ell_\Psi = 3.2411, \sigma_\Psi = 10.1815, \sigma^2_{\text{obs}} = 0.0137$ for the SE-Helmholtz GP
- $\ell_1 = 7.3457, \sigma_1 = 4.0581, \ell_2 = 2.7183, \sigma_2 = 24.7519, \sigma^2_{\text{obs}} = 0.0202$ for the SE-velocity GP.

In Figure 12 we show the results of this scenario. Here the divergence areas are even more diffuse, and the overall field ends up having longer lengthscale of variation. The results on velocity predictions, divergence, and vorticity are aligned with the medium size scenario.

For the velocity prediction task, the three models produce predictions that are close to the truth. Now the two GP models are similar, as can be seen in the difference from the truth plots, and they are both significantly better than the D-HNN. This result is confirmed by the RMSEs: 0.41 for the SE-Helmholtz GP, 0.22 for the SE-velocity GP, and 1.63 for the D-HNN.

In terms of divergence, the SE-Helmholtz GP accurately predicts the two areas of divergence, still with some noise in the central region. The SE-velocity GP is less accurate, especially in the top right region, but overall understand that there are these two areas. The D-HNN prediction is poor. As in the past experiment, it is interesting to observe the z-value plots: both GP models have very high z-values in the areas of divergence, proving their ability to capture the locations of these. The RMSEs are: 0.08 for the SE-Helmholtz GP, 0.17 for the SE-velocity GP, 1.10 for the D-HNN.

Finally, also if we consider the vorticity, the results are similar to the previous scenario. Predictions are good for the two GPs, with meaningful z-values. Now the SE-velocity GP predictions align almost perfectly with the ground truth, and this is reflected in the lower RMSE (0.16 vs. 0.48 for the SE-Helmholtz GP). The D-HNN still fails to predict structure precisely (2.41 RMSE)

In summary, with this experiment we showed that the SE-Helmholtz GP is generally better than the other models in predicting the underlying velocity field (significantly better in the first scenario). In terms of divergence and vorticity, we do not see a large difference compared to the SE-velocity GP: both models are very good; SE-Helmholtz GP is slightly better for the divergence and SE-velocity GP is slightly better for the vorticity. This behavior is very interesting, showing how both models are able to predict a complex divergence pattern (more complex than the previous experiment).

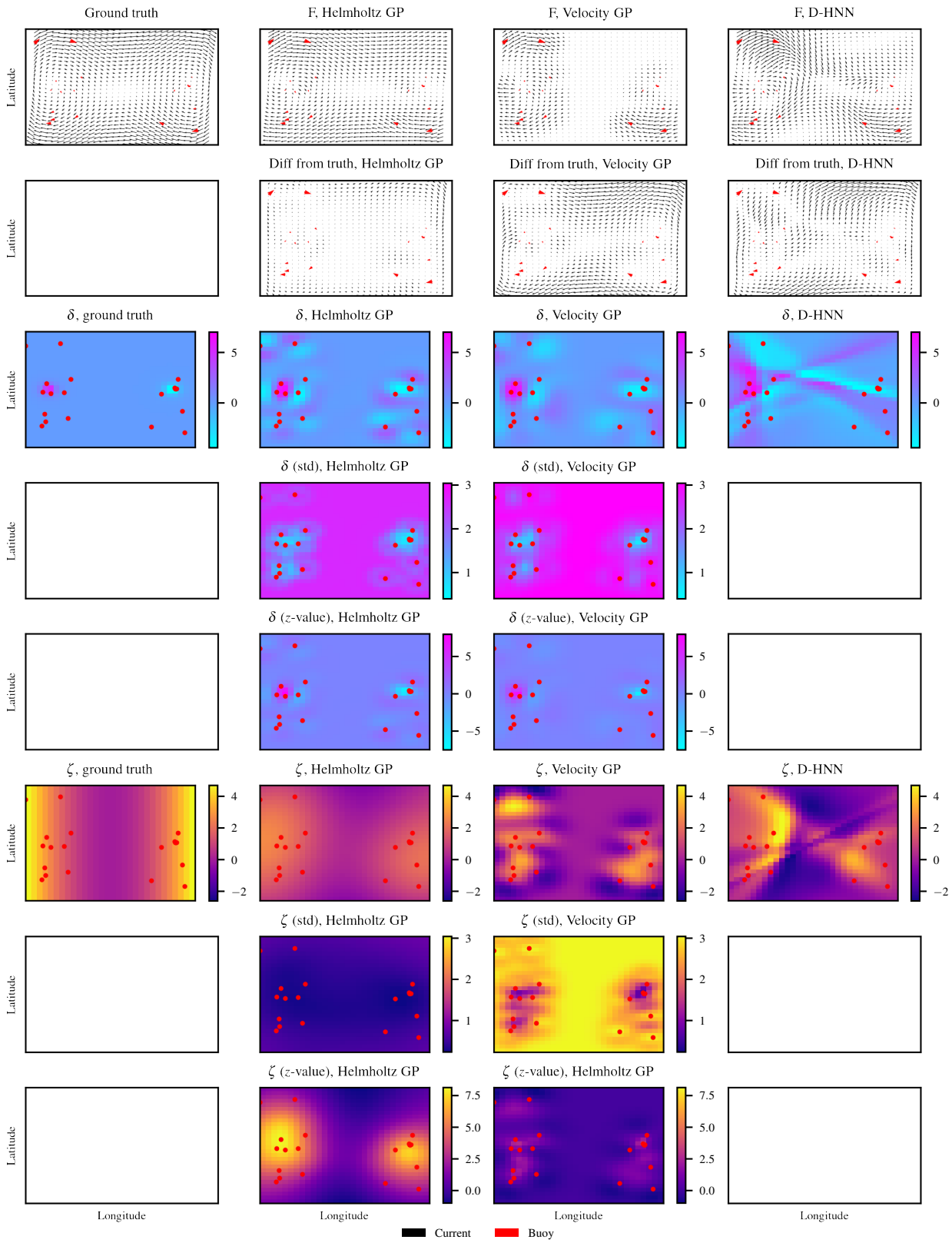

Figure 10. Duffing with small area of divergence. First column: ground truths. Second column: SE-Helmholtz GP results. Third column: SE-velocity GP results. Fourth column: D-HNN results.

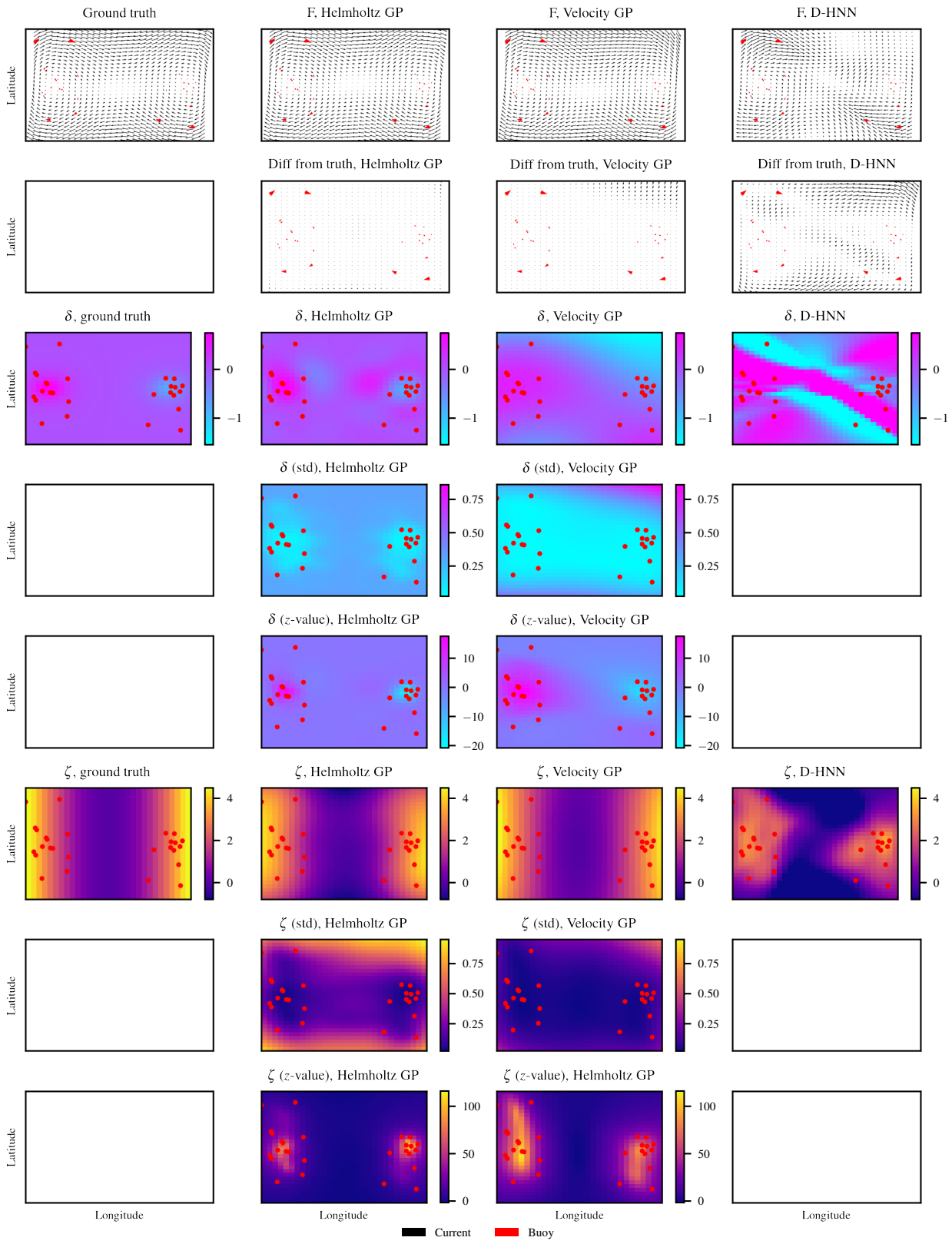

Figure 11. Duffing with medium area of divergence. First column: ground truths. Second column: SE-Helmholtz GP results. Third column: SE-velocity GP results. Fourth column: D-HNN results.

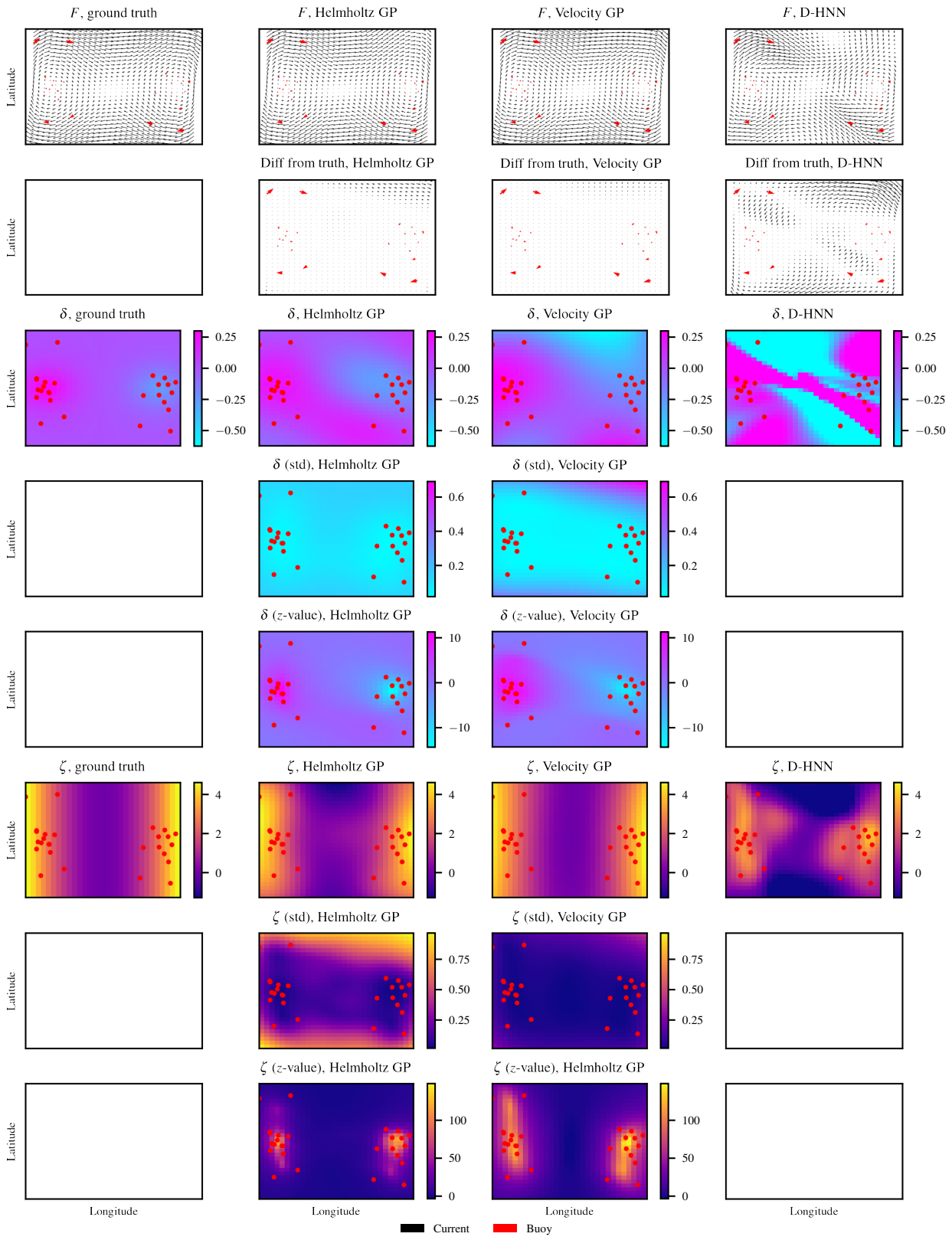

Figure 12. Duffing with big area of divergence. First column: ground truths. Second column: SE-Helmholtz GP results. Third column: SE-velocity GP results. Fourth column: D-HNN results.

I.2. Real-world data 1: LASER

The LAgrangian Submesoscale ExpeRiment, or LASER (Novelli et al., 2017), was performed in the Gulf of Mexico in January-February 2016. Around 10 million data points were retrieved from more than 1000 near-surface biodegradable CODE-type ocean drifters (drogued at a depth of one meter) tracked in real-time using SPOT GPS units. These data were then preprocessed as described in Yaremchuk & Coelho (2014). Finally, since satellite data can have errors and positions of buoys sometimes jump an unrealistic amount, oceanographers removed some bad points that were visible by eye. The preprocessed data are available at `https://data.gulfresearchinitiative.org/data/R4.x265.237:0001` (D'Asaro et al., 2017). In our analysis, we use locations and velocities of buoys as they appear in this dataset.

The main goal of the experiment was to obtain data to understand the submesoscale ocean currents in the open ocean environment near the DeSoto Canyon, as well as how oil or other pollutants might be transported via these currents. In our analysis, we consider a subsample of the LASER data, in an area where the oceanographers expect a convergent front to be (from visual inspection of drifter data). This particular structure in the ocean happens when there are two different masses of water that collide and cause the formation of an area where water sinks. This behavior could happen when two water masses with different temperatures and/or salinities meet, or when water masses from different directions go towards the same area, such as the meeting of warm equatorial water and cold polar water. These fronts are very important for understanding ocean circulation and weather patterns, and can also be a source of nutrients for marine life. To study this structure, we consider two experiments: in the first one, we run our model on a small subset of buoys from this region, collapsing the time dimension and downsampling the observations. To confirm our finds, we then run our models on a dataset that contains more buoys and observations, still from that region.

I.2.1. LASER, CONVERGENT FRONT, SPARSE

In this analysis, we consider 19 buoys, observed every fifteen minutes over a two hour time horizon. By downsampling by a factor of 3 and collapsing the time dimension, we obtain 55 observations. In these data, oceanographers expect to see a clear convergent front in the left region of the spatial domain.

Model fitting. The optimization routine is exactly the same that we do for the simulated experiments: gradient-based Adam algorithm until convergence or a sufficient amount of iterations has elapsed. For the initial hyperparameters, we have tried various alternatives, and found out that the predictions do not change significantly. Hence, for coherence, we stick to the usual initialization done for synthetic data, i.e., $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.368, \sigma^2_{\text{obs}} = 0.135$ for the SE-Helmholtz GP, and $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.368, \sigma^2_{\text{obs}} = 0.135$ for the SE-velocity GP. The optimal hyperparameters obtained are: $\ell_\Phi = 1.6032, \sigma_\Phi = 0.0496, \ell_\Psi = 13.3272, \sigma_\Psi = 1.6392, \sigma^2_{\text{obs}} = 0.0232$ for the SE-Helmholtz GP, and $\ell_1 = 8.3149, \sigma_1 = 0.1384, \ell_2 = 2.7183, \sigma_2 = 0.1318, \sigma^2_{\text{obs}} = 0.0276$

Results. We show the results in Figure 13. The top row shows the predictions for the three models. As before, red arrows are the observed buoy data. The black arrows show the current posterior means at test locations. The test locations are 400 points evenly sparse on a 20 x 20 grid that covers the full range of latitude and longitude of our buoys' observations. The three models produce very similar results: a quasi-constant flow towards the south-west area of the region. There is a slight difference in prediction for the region where buoys seem to converge (SE-velocity GP and D-HNN do not predict different current around there, SE-Helmholtz GP predicts a more converging behavior).

This difference is clear when we look at the posterior divergence plots, in the second row. Our model predicts a negative divergence area (in light-blue) in the area where the oceanographers expect a convergent front. On the contrary, the SE-velocity GP predicts no divergence on the whole spatial domain. This is a very important difference, showing how our model can perform better in recovering this very important property of the ocean. Note that this same intuition is confirmed if we look at the fourth row, where we have z-value plots for both models: the z-values for the SE-Helmholtz GP around the expected convergent front are strongly negative, meaning that the divergence there is significantly non-zero, as desired.

For the vorticity, we just have very small values, almost zero, for both models. Unfortunately, there is no oceanographic knowledge to predict the vorticity far away from the observed drifter traces, and therefore we can not conclude anything related to this point.

I.2.2. LASER, CONVERGENT FRONT, FULL

To further validate the result on the divergence, we consider the same buoys floating over a nine hour time horizon, downsampled by a factor of 3, obtaining 240 observations. We fit our models by performing the usual optimization routine, and we plot the results in Figure 14.

In the top row we show the prediction results. For all the models, the predictions around the buoy agree almost perfectly with predictions from the sparse experiment for the SE-Helmholtz GP; further away models, are more conservative and closer to the prior. The divergence plots in the second row are of the most interest. The prediction according to SE-velocity GP changes remarkably relative to the past experiment. Now it matches closely the Helmholtz result, and both methods detect the convergent front. This result shows the strength of our model in being more data efficient, a very desirable property for a GP model.

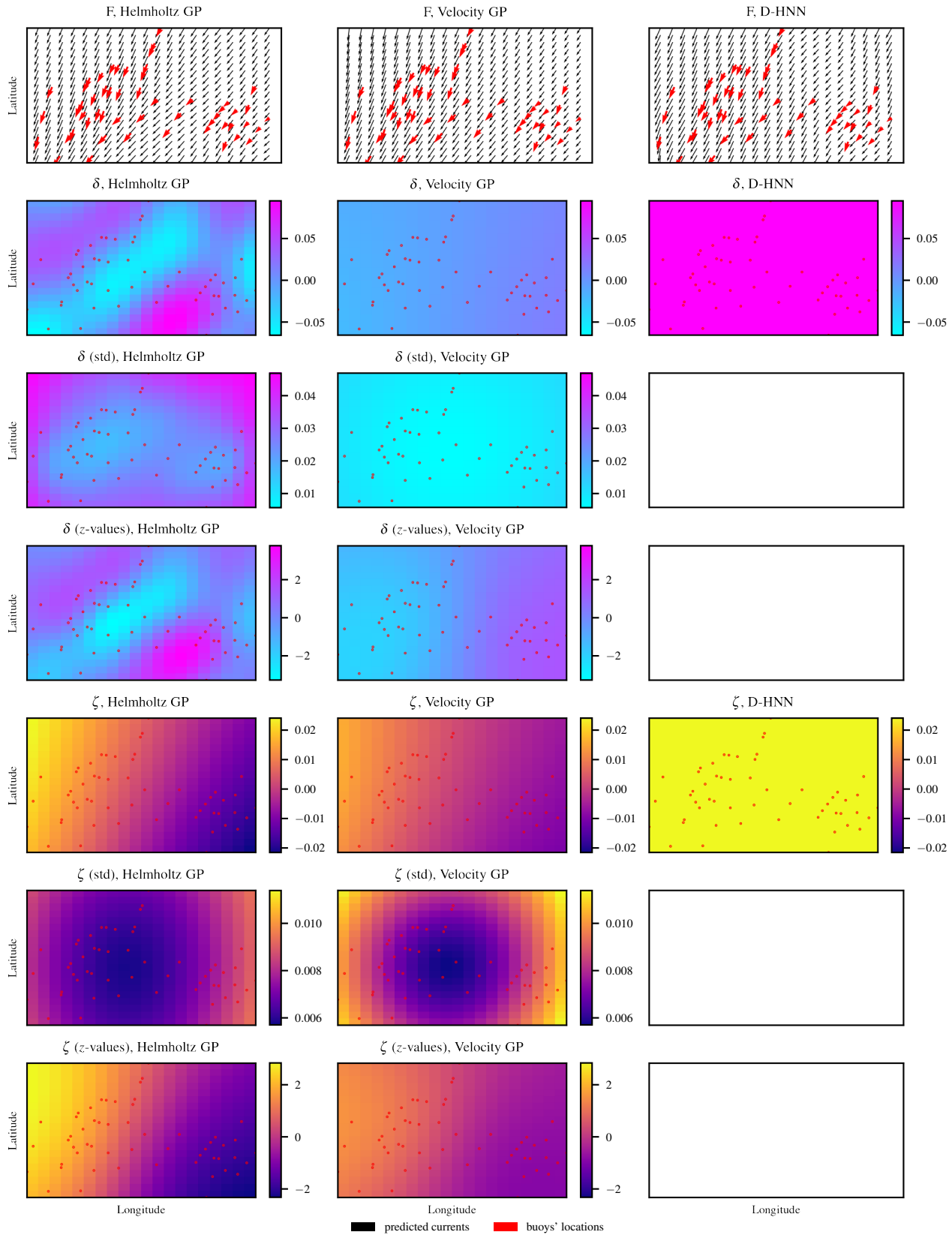

Figure 13. LASER sparse. First column: SE-Helmholtz GP results. Second column: SE-velocity GP results. Third column: D-HNN results.

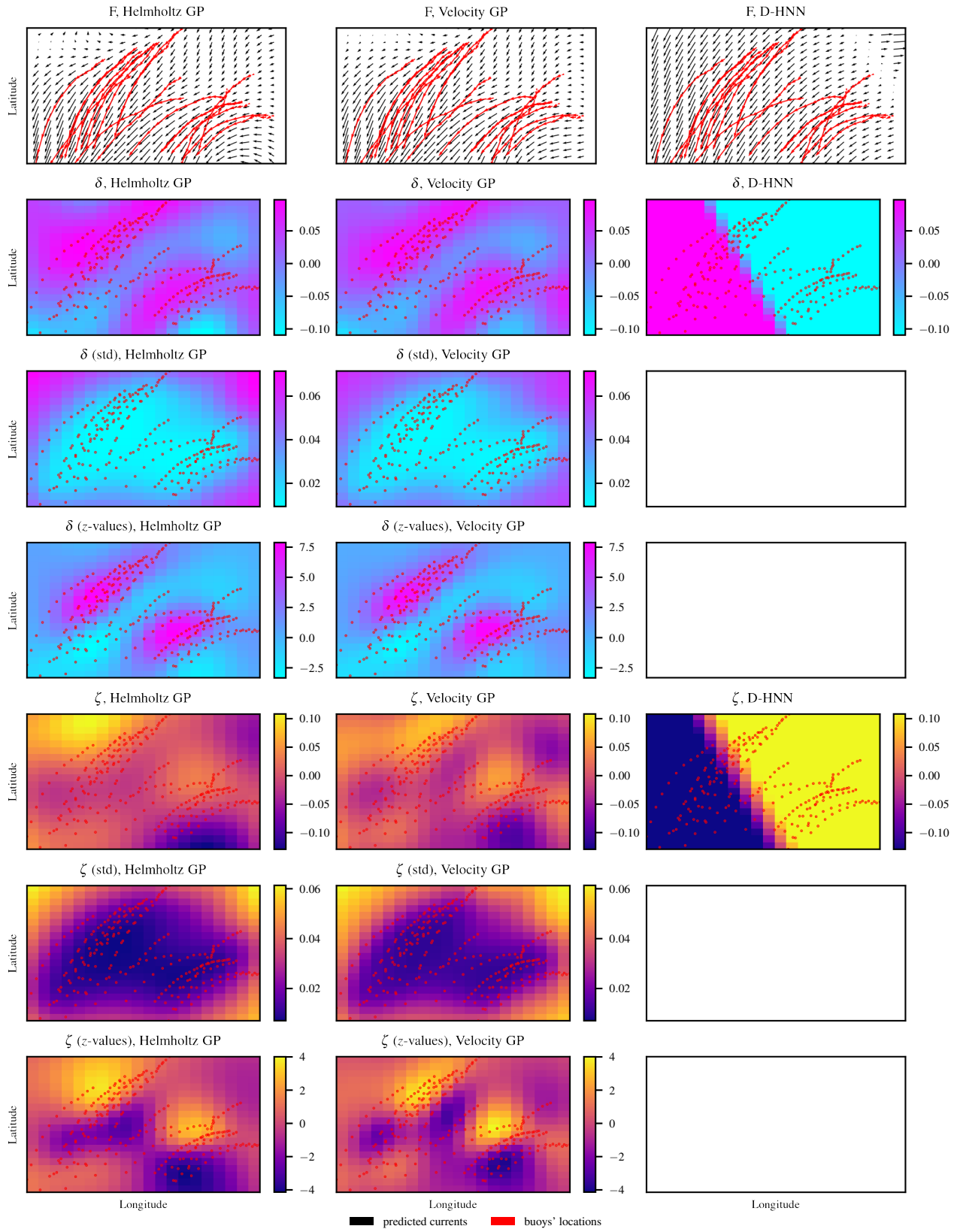

Figure 14. LASER complete. First column: SE-Helmholtz GP results. Second column: SE-velocity GP results. Third column: D-HNN results.

I.3. Real-world data 2: GLAD

The Grand Lagrangian Deployment (GLAD) experiment (Özgökmen, 2012) is another experiment conducted in the northern Gulf of Mexico in July 2012. More than 300 custom-made buoys (of the same type as in the LASER experiment) were deployed near the Deepwater Horizon site and Louisiana coast. This experiment was originally intended to help advance research in understanding the spread and dispersion of oil after the Deepwater Horizon tragedy. Researchers have been using this dataset to study interactions among ocean flows, the levels of influence on transport that large and small flows have, and the size of oil spread at which large flows dominate. Since the GLAD experiment was conducted in the summer time with a shallow 20-meter surface mixed layer for the buoys, the wind has a very strong impact on the trajectories, creating a lot of oscillations. These oscillations are due to a balance of forces due to wind forcing and Earth's rotation, and get amplified during summer time. Filtering these oscillations is a very complicated task, so this wind-induced motions represent a true problem for buoys that are used for measuring oceanographic parameters. Note that we do not see these issues with the LASER data, because that was a winter experiment, where the surface layer is 100-meter deep and devoid of these oscillations.

Model fitting. To deal with this issue, we consider a limited subset of our dataset. We take drifter traces of 12 buoys, observed hourly over a four days time horizon. We collapse the time dimension and downsample these traces by a factor 50, obtaining 85 observations. In terms of optimization routine, we follow very similarly what done in all the other experiments. The only difference is that here different hyperparameter optimization led to different prediction plots for some combinations. In our final results, we decided to stick to the hyperparameter initialization for which both the SE-Helmholtz GP and the SE-velocity GP results were visually more appealing. These are $\ell_\Phi = 12.18, \sigma_\Phi = 0.135, \ell_\Psi = 7.4, \sigma_\Psi = 3, \sigma^2_{\text{obs}} = 0.135$ for the SE-Helmholtz GP, $\ell_1 = 2.7, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 1, \sigma^2_{\text{obs}} = 0.135$ for the SE-velocity GP.

The optimal hyperparameters obtained after the optimization routine are $\ell_\Phi = 45.6840, \sigma_\Phi = 0.0362, \ell_\Psi = 80.1871, \sigma_\Psi = 13.5514, \sigma^2_{\text{obs}} = 0.1715$ for the SE-Helmholtz GP, and $\ell_1 = 72.5835, \sigma_1 = 0.2622, \ell_2 = 2.7183, \sigma_2 = 0.1354, \sigma^2_{\text{obs}} = 0.1739$ for the SE-velocity GP.

Results. In these data, we expect to see a continuous current with no sharp deviations (i.e., lengthscale of variation is long), with few smaller vortices distributed across the region. Unfortunately, here there is no explicit divergence structure that oceanographers expect, so any conclusion from the divergence and vorticity plots is difficult to verify. We show the results of the experiments in Figure 15. We have the predictions in the first row. As before, red arrows are the observed buoy data. The black arrows show the current posterior means at test locations. First of all, the D-HNN model makes physically implausible predictions, likely due to the sparse nature of the data on a large domain. For the GP models, both prediction plots look reasonable, but there are two regions of interest showing important issues with the SE-velocity GP. Consider the bottom right corner. Despite evidence of a strong current making a u-turn, the standard approach shows an abrupt drop in current away from observed data. Our method, on the contrary, predicts a strong current connecting across drifters, in accordance with the continuity of currents (the idea that when a fluid is in motion, it must move in such a way that mass is conserved). This behavior is very problematic. Consider then the top-left corner. Flow behavior around the observations suggests that there might be a vortex in that region. The standard approach shows none. With the SE-Helmholtz GP, instead, we can see the expected vortex between the two lines of current.

To further prove our point, we increase the number of observations to 1200, by decreasing the downsampling factor, and we re-fit the two models with the same optimization routine. The velocity prediction results are included in the first row of Figure 16. Here we can see that our model starts being affected by the oscillations in the data, predicting currents with shorter lengthscale of variation. But also it is still able to reconstruct a continuous current, also far away from the observations, with some vortices with shorter length scale. For the SE-velocity GP, the discontinuity issues increase significantly, and the model is still unable to detect vortices. These are two strong motivations to believe the SE-Helmholtz GP provides a better alternative for this task. The prediction of the D-HNN remains poor.

In terms of divergence and vorticity reconstruction on the sparse dataset, the SE-Helmholtz GP predicts very small divergence almost everywhere, and vorticity coherent with the buoys trajectories. The SE-velocity GP, instead, predicts a reasonable vorticity field, but the divergence shows irregular patterns that look more suspicious. See the second and third blocks in Figure 15 for a visual comparison. By looking at the data, we can see how there are regions on the left where buoys observations seem to be more affected by the oscillations. The SE-velocity GP is more influenced by this noise than our model, and hence predicts divergence areas around the buoys. This claim can be validated by looking at the plots when the dataset size increases. See the second and third blocks in Figure 16. Here, both models seem to be affected more

by the oscillations, but the SE-Helmholtz GP still predicts divergence closer to zero, whereas the SE-velocity GP predicts divergence areas around each conglomerate of buoys in the region. Therefore, we can conclude that our model is at least as good as the SE-velocity GP. Note that we cannot say anything stronger, because there is no expert knowledge suggesting that the SE-Helmholtz GP behavior is the expected one.

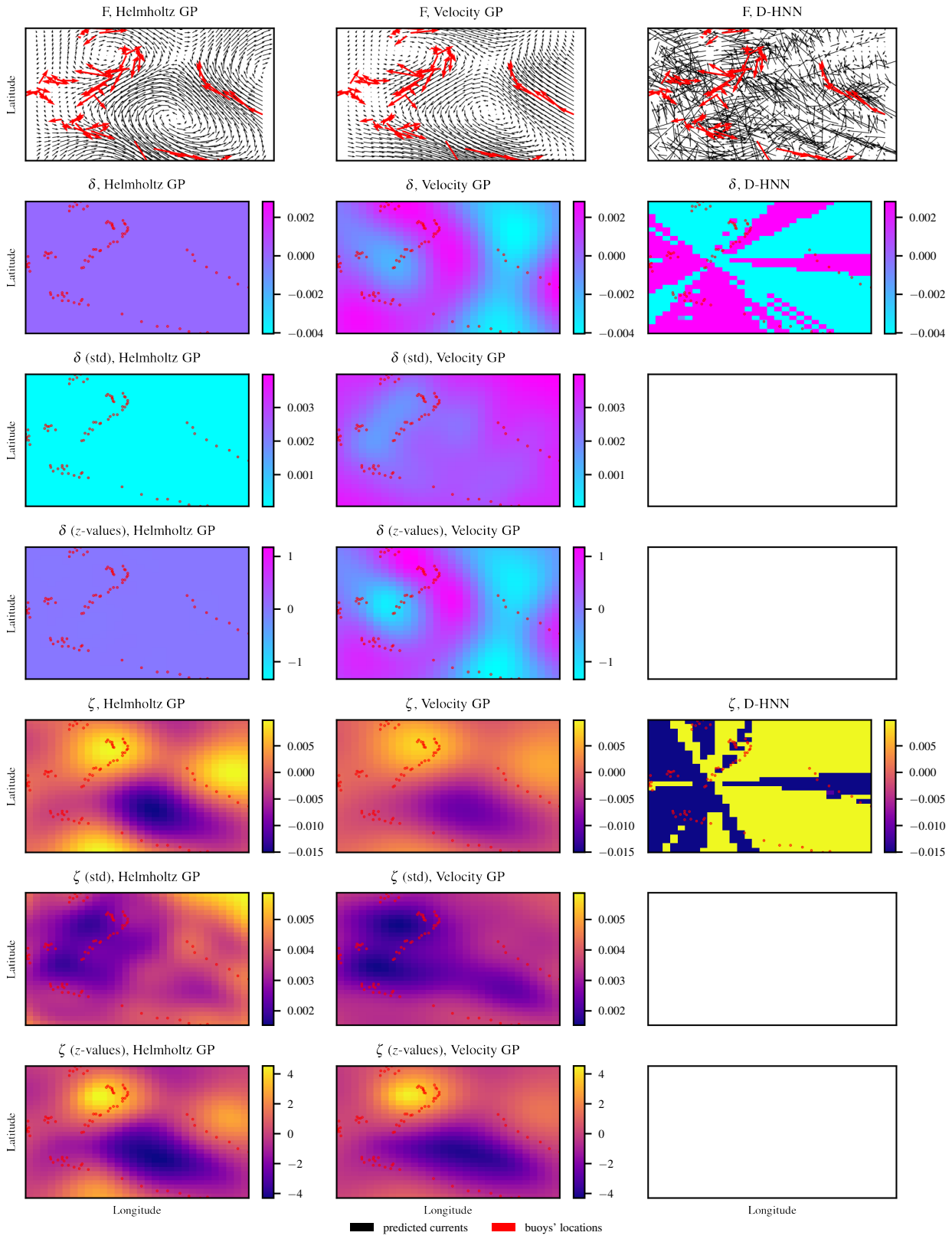

Figure 15. GLAD sparse. First column: SE-Helmholtz GP results. Second column: SE-velocity GP results. Third column: D-HNN results.

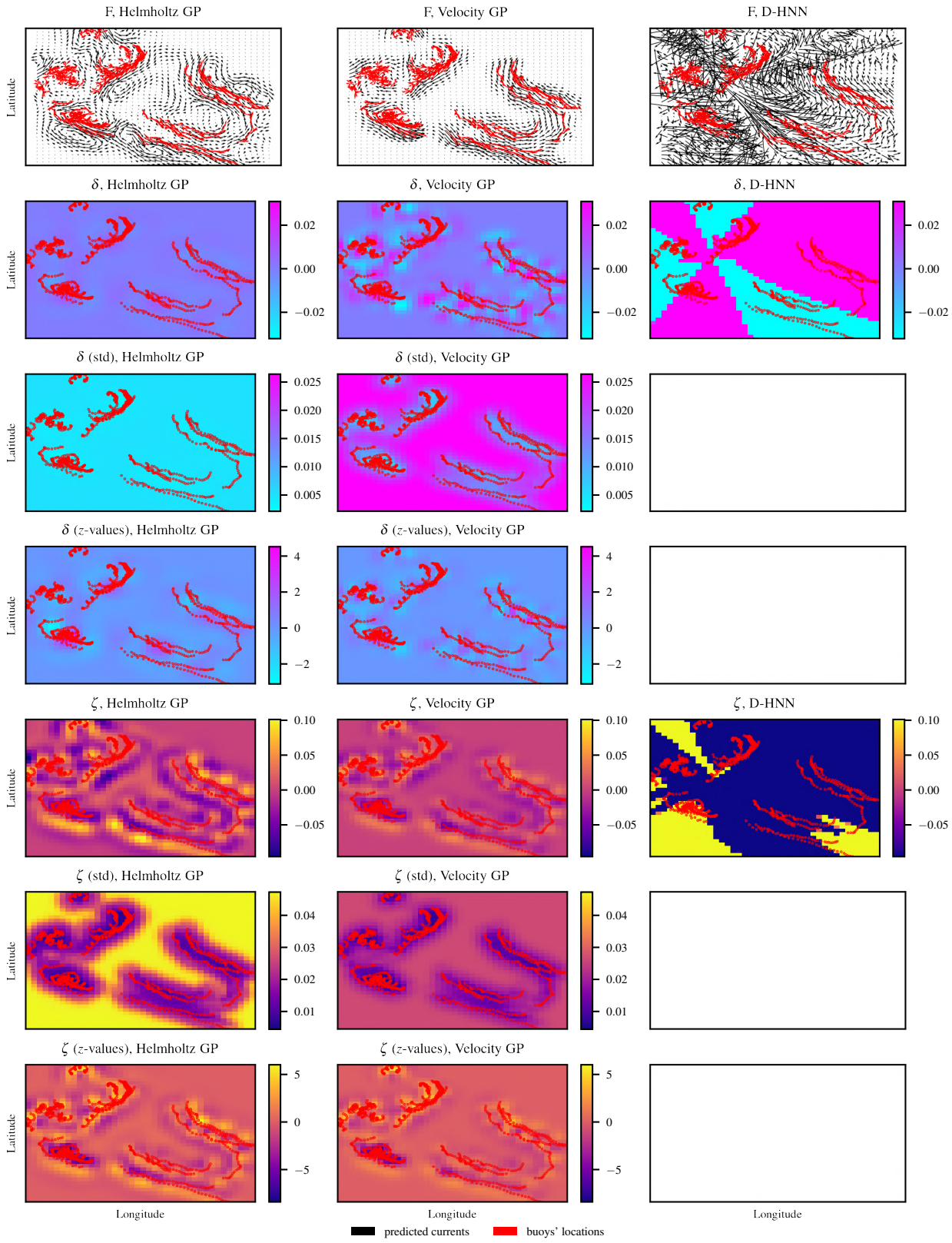

Figure 16. GLAD complete. First column: SE-Helmholtz GP results. Second column: SE-velocity GP results. Third column: D-HNN results.