GAUSSIAN PROCESSES AT THE HELM(HOLTZ): A MORE FLUID MODEL FOR OCEAN CURRENTS

Renato Berlinghieri*, **Brian L. Trippe**†, **David R. Burt***, **Ryan Giordano***,
Kaushik Srinivasan‡, **Tamay Özgökmen**§, **Junfei Xia**§, **Tamara Broderick***
* MIT, † Columbia, ‡ UCLA, § University of Miami

ABSTRACT

Oceanographers are interested in predicting ocean currents and identifying divergences in a current vector field based on sparse observations of buoy velocities. Since we expect current dynamics to be smooth but highly non-linear, 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 prediction 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 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) ocean current prediction and (2) identification of divergences, also known as upwellings and downwellings, in the current vector field. Predicting 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; Haza et al., 2018).

It remains to choose an appropriate method to reconstruct currents and their divergences from the buoy data. Gonçalves et al. (2019) and Lodise et al. (2020) proposed modeling buoy velocities in the latitude and longitude directions with independent Gaussian processes (GPs) with standard spatiotemporal kernels (e.g., squared exponential kernels). See Appendix B for a short review on GPs. 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 data on the ocean surface and its assumption that currents vary smoothly but in a highly nonlinear fashion. We call this 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 shortcomings, we propose to instead model more directly known behaviors from fluid dynamics. In particular, a *Helmholtz decomposition* (Bhatia et al., 2013) from fluid dynamics lets us decompose the vector field of ocean currents into: (1) a *divergent* component (measuring expansion, contraction, and translation) that can be represented as the gradient of a scalar potential function, $\mathrm{grad}\Phi$, and (2) a *rotational* component (measuring rotation) that can be represented as the rotational operator of another scalar potential function, $\mathrm{rot}\Psi$. See Appendix C for more details.

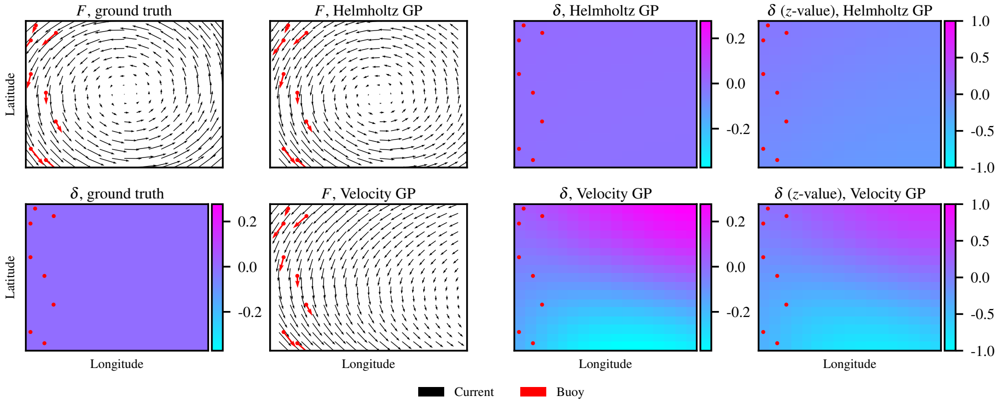

Figure 1: First column: ground truth predictions (upper) and divergence (lower). Second column: current predictions. Third column: divergence estimates. Fourth column: posterior divergence z-values.

By contrast to the standard approach, we model Φ and Ψ with independent GP priors, rather than the velocity components. We demonstrate that our Helmholtz GP is amenable to practical inference. We demonstrate the superior performance of our Helmholtz GP at the current prediction task and divergence estimation task (as well as vorticity estimation) in simulated and real experiments.

Related work. The Helmholtz decomposition has been used extensively for interpretation and visualization of an observed field on a grid, without solving a prediction task (Rocha et al., 2016; Zhang et al., 2018; 2019; Han & Huang, 2020; Bühler et al., 2014; Caballero et al., 2020). Researchers have developed Gaussian process kernels to capture curl- or divergence-free fields (Narcowich & Ward, 1994; Lowitzsch, 2002; Fuselier Jr, 2007; Macêdo & Castro, 2010; Alvarez et al., 2012), and 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) effectively constrains the two components to have the same magnitudes and length scales – whereas 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 to ours, the authors test their method only on low-resolution data available on a dense grid. In our experiments on (sparse) buoy data in Appendix I, we find that it often produces physically implausible predictions.

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 OUR METHOD: GAUSSIAN PROCESSES AT THE HELM(HOLTZ)

We consider a dataset of M observations, $\{\mathbf{x}_m, \mathbf{y}_m\}_{m=1}^{M}$, where $\mathbf{x}_m = (x_m^{(1)}, x_m^{(2)})$ represents the spatial location of a buoy, and $\mathbf{y}_m = (\mathrm{y}_m^{(1)}, \mathrm{y}_m^{(2)})$ the corresponding horizontal and vertical velocities (the "drifter trace"). We assume a likelihood, or noise model, relating the observed buoy velocities to a field $F : \mathbb{R}^2 \to \mathbb{R}^2$, mapping spatial locations into horizontal and vertical velocities, $(F^{(1)}, F^{(2)})$:

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

for some $\sigma^2_{\text{obs}} > 0$ and independent ($\overset{\text{ind}}{\sim}$) noise across observations.

By the Helmholtz decomposition theorem, we can rewrite $F(\mathbf{x}_m) = \text{grad}\Phi(\mathbf{x}_m) + \text{rot}\Psi(\mathbf{x}_m)$, for each $\mathbf{x}_m$. To exploit this decomposition, we put independent GP priors on the Helmholtz stream and potential functions:

$$\Phi \sim \mathcal{GP}(0, k_\Phi) \quad \text{and} \quad \Psi \sim \mathcal{GP}(0, k_\Psi), \tag{2}$$

where we take k_Φ and k_Ψ to be squared exponential (Equation (4) in Appendix B) with parameters ℓ_Φ, σ^2_Φ, and ℓ_Ψ, σ^2_Ψ, respectively. In principle we could use any two kernels that are twice differentiable with respect to each of their arguments (Adler 1981, Theorem 2.2.2).

Our result below shows that, with this prior, we can still make predictions in the buoy space. Intuitively, the result holds since (i) the Helmholtz decomposition is based on partial (mixed) derivatives and (ii) the derivative of a GP is a GP. In particular, in deriving the covariance of the induced GP on F, we use Equation (9.1) in Rasmussen & Williams (2005).

Proposition 1. *Let F be the ocean current vector field and Φ and Ψ the divergence and vorticity components of an Helmholtz decomposition of the field, respectively. The prior choice in Equation* (2) *implies that the prior on the original vector field F is a GP as well, with mean 0 and kernel* $\text{grad}\, k_\Phi + \text{rot}\, k_\Psi$*, as in Appendix D.*

We show in Appendix E that, with this prior choice, we can recover the posteriors on, respectively, (1) the current field F, (2) the divergence δ, and (3) the vorticity ζ. We estimate divergence or vorticity at a location with the posterior mean, and we report uncertainty via the posterior variance. Moreover, we show in Appendix F that using the Helmholtz GP induces only a small constant multiple of additional computational expense relative to the velocity GP – because the gradient operators "double" the size of the covariance matrices involved.

Finally, in Appendix G, we establish key advantages of the Helmholtz GP prior over the velocity GP prior: (1) more physically realistic prior assumptions that the magnitude and length scales of the divergence and vorticity can be substantially different, (2) more physically realistic correlation of the longitudinal and latitudinal velocities of current at any point, and (3) equivariance to reference frame.

3 EXPERIMENTAL RESULTS

We now show that the Helmholtz GP yields better current predictions as well as better divergence (and vorticity) identification across a variety of simulated and real data sets, compared to the velocity GP. In Appendix I we show that both methods improve on the Greydanus & Sosanya (2022) approach.

In our experiments, we use only the spatial locations of buoys and discard time; that is, for now we make the simplifying assumption that currents are stationary across time. The real datasets consist of drifter traces of GPS-tagged buoys in the ocean. We first simulate common ocean behaviors of interest, though, because we have access to full ground truth only in simulations. See Appendix I for more details of our simulation setups.

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, we find that – for predictions, divergence estimates, and vorticity estimates – the Helmholtz GP is most accurate by RMSE on a majority of tasks, often by a substantial factor; see Table 2 in Appendix H.

As one example, we consider a single vortex with no divergence. In a vortex, water particles rotate around a 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 the simulated buoy trajectories, our training data. See Appendix I.1.1 for additional details.

The second column shows predictions from the Helmholtz GP (upper) and velocity GP (lower) at the test points. The red arrows are still the training data. Despite having access only to data from one side of the vortex, the Helmholtz GP is able to reconstruct the full vortex. The 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 Helmholtz GP (upper) and velocity GP (lower) on the same color scale. The fourth column helps us understand if either posterior is reporting a

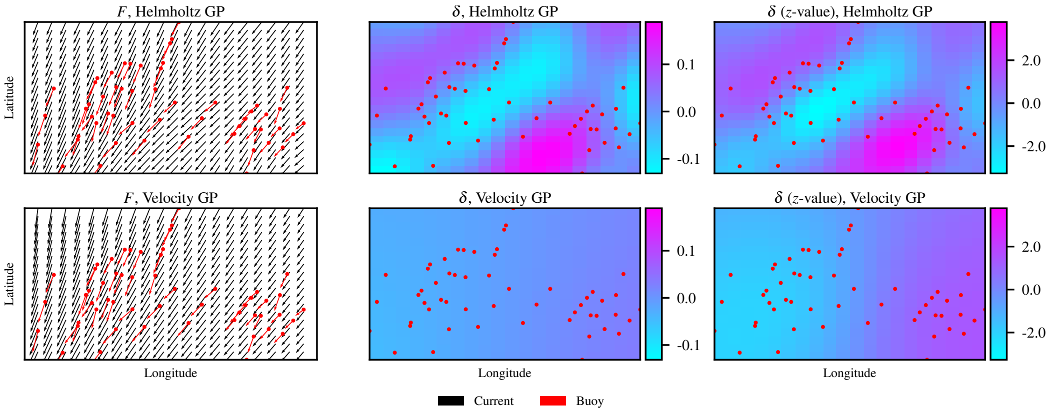

Figure 2: First column: current predictions. Second column: divergence estimates. Third column: posterior divergence z-values.

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 Helmholtz GP estimate of the divergence is closer to the ground truth of zero than the velocity GP. From the fourth column, we see that neither method concludes nonzero divergence, but the Helmholtz GP posterior is more concentrated near zero. In Appendix I.1.2 we provide an example with zero ground truth divergence, where the Helmholtz GP correctly estimates no divergence, and the velocity GP erroneously and very confidently reports nonzero divergence.

Real-data experiments. Although ground truth currents and divergences are not available for real data, we can still assess performance against oceanographers' expert knowledge. We consider the LAgrangian Submesoscale ExpeRiment (Novelli et al., 2017), performed in the Gulf of Mexico in 2016. 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 convergent front. The left column of Figure 2 shows in black the current predictions using the Helmholtz GP (upper) and velocity GP (lower). In this case, the two sets of predictions are qualitatively very similar.

The second column shows the divergence predictions for the Helmholtz GP (upper) and velocity GP (lower); the third column shows the z-values for the posterior distributions. The Helmholtz GP predicts a negative divergence area (light blue diagonal) that agrees with the region where oceanographers expect a convergent front. By contrast, the velocity GP does not identify any divergence.

We find that the discrepancy between the Helmholtz GP and the velocity GP observed in Figure 2 depends on the amount of data. When we increase the amount of data by considering the same buoys over a 9-hour time span, we find that both methods are able to recover the same convergent front; see Appendix I.2 and Figure 13. This finding then also corroborates that the convergent front in fact does exist and looks like the posterior detection by the Helmholtz GP in Figure 2. In that way, the finding further lends support to the superior performance of the Helmholtz GP in Figure 2.

In Appendix I.3 we provide results for the Grand Lagrangian Deployment (GLAD) experiment (Özgökmen, 2012) near the Deepwater Horizon site. In this case, various features of the Helmholtz GP prediction appear 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.

Discussion and future work. 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) – 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

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

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

George B Arfken and Hans J Weber. Mathematical methods for physicists, 1999.

Maristella Berta, Annalisa Griffa, Marcello G Magaldi, Tamay M Özgökmen, Andrew C Poje, Angelique C Haza, and M Josefina Olascoaga. Improved surface velocity and trajectory estimates in the Gulf of Mexico from blended satellite altimetry and drifter data. *Journal of Atmospheric and Oceanic Technology*, 32(10):1880–1901, 2015.

Harsh Bhatia, Gregory Norgard, Valerio Pascucci, and Peer-Timo Bremer. The Helmholtz-Hodge decomposition—a survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(8): 1386–1404, 2013.

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

Ainhoa Caballero, Sandrine Mulet, Nadia Ayoub, Ivan Manso-Narvarte, Xabier Davila, Christine Boone, Florence Toublanc, and Anna Rubio. Integration of HF radar observations for an enhanced coastal mean dynamic topography. *Frontiers in Marine Science*, pp. 1005, 2020.

Cédric P Chavanne and Patrice Klein. Can oceanic submesoscale processes be observed with satellite altimetry? *Geophysical Research Letters*, 37(22), 2010.

Eric D'Asaro, Cedric Guigand, Angelique Haza, Helga Huntley, Guillaume Novelli, Tamay Özgökmen, and Ed Ryan. Lagrangian submesoscale experiment (laser) surface drifters, interpolated to 15-minute intervals, 2017. URL `https://data.gulfresearchinitiative.org/data/R4.x265.237:0001`.

Eric A. D'Asaro, Andrey Y. Shcherbina, Jody M. Klymak, Jeroen Molemaker, Guillaume Novelli, Cédric M. Guigand, Angelique C. Haza, Brian K. Haus, Edward H. Ryan, Gregg A. Jacobs, Helga S. Huntley, Nathan J. M. Laxague, Shuyi Chen, Falko Judt, James C. McWilliams, Roy Barkan, A. D. Kirwan, Andrew C. Poje, and Tamay M. Özgökmen. Ocean convergence and the dispersion of flotsam. *Proceedings of the National Academy of Sciences*, 115(6):1162–1167, 2018.

Lee-Lueng Fu and Raffaele Ferrari. Observing oceanic submesoscale processes from space. *Eos, Transactions American Geophysical Union*, 89(48):488–488, 2008.

Lee-Lueng Fu, Dudley B Chelton, Pierre-Yves Le Traon, and Rosemary Morrow. Eddy dynamics from satellite altimetry. *Oceanography*, 23(4):14–25, 2010.

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

Rafael C. Gonçalves, Mohamed Iskandarani, Tamay Özgökmen, and W. Carlisle Thacker. Reconstruction of submesoscale velocity field from surface drifters. *Journal of Physical Oceanography*, 49(4), 2019.

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

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

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

A. C. Haza, E. D'Asaro, H. Chang, S. Chen, M. Curcic, C. Guigand, H. S. Huntley, G. Jacobs, G. Novelli, T. M. Özgökmen, A. C. Poje, E. Ryan, and A. Shcherbina. Drogue-loss detection for surface drifters during the lagrangian submesoscale experiment (LASER). *Journal of Atmospheric and Oceanic Technology*, 35(4), 2018.

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

John Lodise, Tamay Özgökmen, Rafael C. Gonçalves, Mohamed Iskandarani, Björn Lund, Jochen Horstmann, Pierre-Marie Poulain, Jody Klymak, Edward H. Ryan, and Cedric Guigand. Investigating the formation of submesoscale structures along mesoscale fronts and estimating kinematic quantities using Lagrangian drifters. *Fluids*, 5(3), 2020.

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

Ives Macêdo and Rener Castro. *Learning divergence-free and curl-free vector fields with matrix-valued kernels*. IMPA, 2010.

A. J. Mariano, E. H. Ryan, H. S. Huntley, L.C. Laurindo, E. Coelho, A. Griffa, T. M. Özgökmen, M. Berta, D. Bogucki, S. S. Chen, M. Curcic, K.L. Drouin, M. Gough, B. K. Haus, A. C. Haza, P. Hogan, M. Iskandarani, G. Jacobs, A. D. Kirwan Jr., N. Laxague, B. Lipphardt Jr., M. G. Magaldi, G. Novelli, A. Reniers, J. M. Restrepo, C. Smith, A. Valle-Levinson, and M. Wei. Statistical properties of the surface velocity field in the northern Gulf of Mexico sampled by GLAD drifters. *Journal of Geophysical Research: Oceans*, 121(7):5193–5216, 2016.

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

Guillaume Novelli, Cédric M Guigand, Charles Cousin, Edward H Ryan, Nathan JM Laxague, Hanjing Dai, Brian K Haus, and Tamay M Özgökmen. A biodegradable surface drifter for ocean sampling on a massive scale. *Journal of Atmospheric and Oceanic Technology*, 34(11):2509–2532, 2017.

TM Özgökmen. 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.

Andrew C. Poje, Tamay M. Özgökmen, Bruce L. Lipphardt, Brian K. Haus, Edward H. Ryan, Angelique C. Haza, Gregg A. Jacobs, A. J. H. M. Reniers, Maria Josefina Olascoaga, Guillaume Novelli, Annalisa Griffa, Francisco J. Beron-Vera, Shuyi S. Chen, Emanuel Coelho, Patrick J. Hogan, Albert D. Kirwan, Helga S. Huntley, and Arthur J. Mariano. Submesoscale dispersion in the vicinity of the Deepwater Horizon spill. *Proceedings of the National Academy of Sciences*, 111 (35):12693–12698, 2014.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2005.

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

Arno Solin, Manon Kok, Niklas Wahlström, Thomas B Schön, and Simo Särkkä. Modeling and interpolation of the ambient magnetic field by Gaussian processes. *IEEE Transactions on robotics*, 34(4):1112–1127, 2018.

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

Niklas Wahlström, Manon Kok, Thomas B Schön, and Fredrik Gustafsson. Modeling magnetic fields using Gaussian processes. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3522–3526. IEEE, 2013.

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

Cuicui Zhang, Hao Wei, Zhilei Liu, and Xiaomei Fu. Characteristic ocean flow visualization using Helmholtz decomposition. In *2018 Oceans-MTS/IEEE Kobe Techno-Oceans (OTO)*, pp. 1–4. IEEE, 2018.

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

A RELATED WORK

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

The Helmholtz decomposition has been used extensively in the oceanographic literature as a way to interpret and visualize currents' features when the field is observed on a grid (Rocha et al., 2016; Zhang et al., 2018; 2019; Han & Huang, 2020; Bühler et al., 2014; Caballero et al., 2020). These authors decompose ocean currents into vortical (divergence-free) and divergent (curl-free) components of the flow, and different characteristic ocean features are recognized within different components: eddies, which are rotational, are only present in the vortical component; fronts, being irrotational, are only present in the divergent one. These works, however, rely mostly on measurements from satellites, which unfortunately have spatial and temporal resolution that is too coarse to precisely capture most of the features of interest (Fu & Ferrari, 2008; Chavanne & Klein, 2010; Fu et al., 2010). Moreover, they assume data are available on a dense grid, which is not the case for real ocean current data. As a consequence, to more accurately capture behaviors about ocean features at this resolution, researchers release GPS-tagged buoys in the ocean – and track the corresponding surface drifter traces, varying across time and space (Poje et al., 2014; Berta et al., 2015; Mariano et al., 2016; Novelli et al., 2017; Haza et al., 2018). Since many patches of ocean end up having no observations, there is inherent modeling uncertainty that must be confronted, and the problem gets challenging. In particular, it is not obvious how to simultaneously interpolate and extrapolate away from drifter data, with an appropriate measure of uncertainty, while still leveraging the Helmholtz decomposition to learn structures of ocean currents.

In the GP literature, it is well known that specific covariance functions can be used to capture curl- and divergence-free vector fields (Narcowich & Ward, 1994; Lowitzsch, 2002; Fuselier Jr, 2007; Macêdo & Castro, 2010; Alvarez et al., 2012). In particular, Macêdo & Castro (2010) is the most related to our work. The authors investigate kernel ridge regression using divergence and curl-free kernels. They consider convex combinations of divergence and curl-free kernels for modelling of general fluid fields, and emphasize the advantage of being able to decompose a field into its divergence and curl-free components. However, they do not investigate methods for selecting the weighting of the curl- and divergence-free kernels (or other parameters in the model). In our modelling task, it is critical that we have a data-driven method to select the relative weighting (variances) of the kernels as well as the lengthscales, as the curl and divergence-free components of the ocean current generally have different magnitudes and act across different scales. We achieve this with maximum marginal likelihood, and discuss this benefit of our approach in greater detail in Appendix G. Wahlström et al. (2013); Wahlström (2015); Solin et al. (2018) uses curl- and divergence-free kernels in the context of electromagnetic fields, similarly to what we are doing for ocean currents. In particular, Wahlström (2015) already suggests using the Helmholtz expansion and placing independent GP priors on each term for vector field modeling. However, the authors assume to have direct access to noisy observations of the divergence-free component alone (rather than the sum of divergence- and curl-free components), and they constrain the two components to be exactly equal (with opposite direction) at observed data points, so that they can say to have noisy observations of the curl-free component directly as well. By doing so, their work does not allow different magnitudes or length scales of the divergence- and curl-free components. So they cannot assess recovery of these quantities from a noisy sum of components. But, even more, their work cannot assess recovery of two unconstrained divergence- and curl-free components from a noisy sum of these components, which is what we care about in our ocean setting. In general, 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.

Finally, a recent work from the deep learning community (Greydanus & Sosanya, 2022) extended Hamiltonian Neural Networks (Greydanus et al., 2019) in such a way that, the authors suggest, allows one to model both curl- and divergence-free dynamics simultaneously, for example for reconstructing surface flows from a noisy ocean current dataset. Although the prediction problem is similar, the authors test this method only on low-resolution data available on a dense grid. We show results of this method for sparse buoy data in Appendix I, and find that it often produces physically implausible predictions.

B GAUSSIAN PROCESSES REVIEW

In this section, we review Gaussian processes. Let $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$ represent two input vectors. Assume that we want to model a two dimensional function $G : \mathbb{R}^2 \to \mathbb{R}^2$, $G(\mathbf{x}) = (G^{(1)}(\mathbf{x}), G^{(2)}(\mathbf{x}))^\top$. A 2-output Gaussian process on covariate space $\mathbb{R}^2$ is determined by a mean function $\mu : \mathbb{R}^2 \to \mathbb{R}^2$, $\mu(\mathbf{x}) = (\mu^{(1)}(\mathbf{x}), \mu^{(2)}(\mathbf{x}))^\top$, and a positive definite kernel function $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$. We use $k(\mathbf{x}, \mathbf{x}')_{i,j}$ to denote the (i, j)th output of $k(\mathbf{x}, \mathbf{x}')$. We say $F \sim \mathcal{GP}(\mu, k)$ 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 $(p_1, \ldots p_N) \in \{1, 2\}^N$, $(G^{(p_n)}(\mathbf{x}_i))_{n=1}^N$ is an N-dimensional Gaussian random variable with mean vector $(\mu^{(p_n)}(\mathbf{x}_i))_{n=1}^N$ and covariance matrix with (i, j)th entry $k(\mathbf{x}_i, \mathbf{x}_j)_{p_i,p_j}$.

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

$$k_{\text{vel}}(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} k_{\text{SE}}^{(1)}(\mathbf{x}, \mathbf{x}') & 0 \\ 0 & k_{\text{SE}}^{(2)}(\mathbf{x}, \mathbf{x}') \end{bmatrix}. \tag{3}$$

where, for $i \in \{1, 2\}$,

$$k_{\text{SE}}^{(i)}(\mathbf{x}, \mathbf{x}') = \sigma_i^2 \exp\left(-\tfrac{1}{2}\|\mathbf{x} - \mathbf{x}'\|_2^2/\ell_i^2\right) \tag{4}$$

is the squared exponential kernel in one dimension. Here there are four hyperparameters for the GP prior: 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. In our work, we refer to this GP prior as the *velocity GP* to emphasize that the squared exponential priors are directly on the observed velocities. See Alvarez et al. (2012) for a review of multi-output GPs.

C DIVERGENCE, GRADIENT, AND CURL OPERATORS IN 2D

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 **x** is the vector whose components are the partial derivatives of f at **x**. Formally,

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

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

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 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 $\operatorname{div} F : A \to \mathbb{R}$, defined by

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

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

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

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, $\boldsymbol{k}$ which is orthogonal to the plane containing, $\boldsymbol{i}$ and $\boldsymbol{j}$.

Definition 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

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

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 grad and rot operators are the building blocks for the Helmholtz decomposition.

Definition 4 (Helmholtz decomposition, Bhatia et al., 2013)**.** *A smooth 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 $\Psi : \mathbb{R}^2 \to \mathbb{R}$, the stream function:*

$$\underbrace{F(\mathbf{x})}_{\textit{ocean flow}} = \underbrace{\operatorname{grad}\Phi(\mathbf{x})}_{\textit{divergent velocity}} + \underbrace{\operatorname{rot}\Psi(\mathbf{x})}_{\textit{rotational velocity}} \tag{5}$$

where $\mathbf{x}$ is a 2-dimensional input, $\mathbf{x} = (x^{(1)}, x^{(2)})^\top$, and

$$\operatorname{grad}\Phi(\mathbf{x}) := \begin{bmatrix} \partial\Phi(\mathbf{x})/\partial x^{(1)} \\ \partial\Phi(\mathbf{x})/\partial x^{(2)} \end{bmatrix} \textit{ and } \operatorname{rot}\Psi(\mathbf{x}) := \begin{bmatrix} \partial\Psi(\mathbf{x})/\partial x^{(2)} \\ -\partial\Psi(\mathbf{x})/\partial x^{(1)} \end{bmatrix}. \tag{6}$$

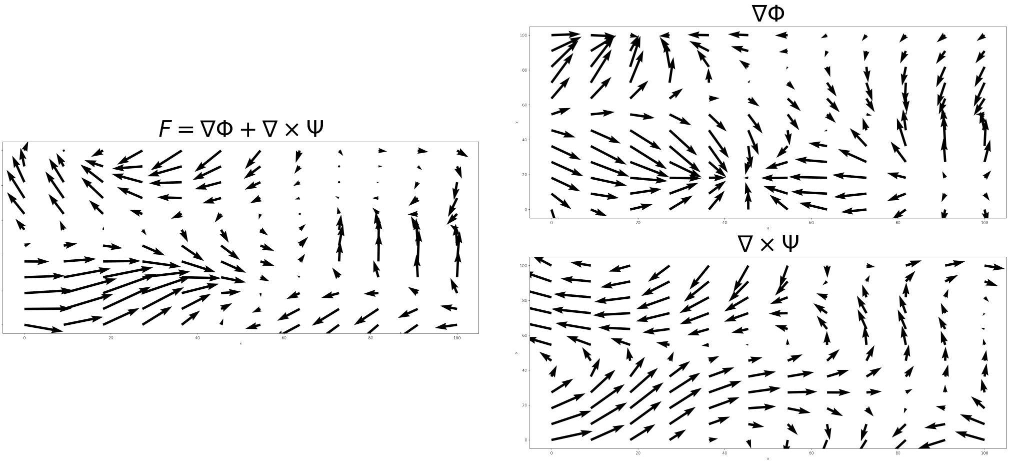

Figure 3: Helmholtz decomposition of vector field F. *Left*: original vector field F. *Top-right*: divergence component of Helmholtz decomposition of F. *Bottom-right*: vorticity component of Helmholtz decomposition of F.

The *divergence* of F, δ, and the *vorticity* of F, ζ, at a point $\mathbf{x}$, are

$$\delta(\mathbf{x}) := \mathrm{div}(F(\mathbf{x})) := \frac{\partial F^{(1)}(\mathbf{x})}{\partial x^{(1)}} + \frac{\partial F^{(2)}(\mathbf{x})}{\partial x^{(2)}} = \frac{\partial^2 \Phi(\mathbf{x})}{\partial^2 x^{(1)}} + \frac{\partial^2 \Phi(\mathbf{x})}{\partial^2 x^{(2)}} \tag{7}$$

$$\zeta(\mathbf{x}) := \mathrm{curl}(F(\mathbf{x})) := \frac{\partial F^{(1)}(\mathbf{x})}{\partial x^{(2)}} - \frac{\partial F^{(2)}(\mathbf{x})}{\partial x^{(1)}} = \frac{\partial^2 \Psi(\mathbf{x})}{\partial^2 x^{(2)}} + \frac{\partial^2 \Psi(\mathbf{x})}{\partial^2 x^{(1)}}. \tag{8}$$

We summarize the terms and notation introduced in this appendix in Table 1, and we show an example of a Helmholtz decomposition of a vector field in Figure 3.

Finally, note that the ideas of gradient, divergence, rot, and curl lead to the following characterization of vector fields.

Definition 5. A vector field $F : A \to \mathbb{R}^2$ is called *rotation-free* (or curl-free) if the curl is zero, $\mathrm{curl}F = 0$, and it is called *incompressible* (or divergence-free) if the divergence is zero, $\mathrm{div}F = 0$.

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

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

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

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

For (2):

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

$\square$

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

D HELMHOLTZ PRIOR IN DETAIL

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

Proposition 3. *Let F be an ocean current vector field defined by potential and stream functions that are a priori 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 squared exponential kernels. Then*

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

where, for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$*,*

$$k_{\text{Helm}}(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} \dfrac{\partial^2 k_\Phi(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(1)}} + \dfrac{\partial^2 k_\Psi(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(2)}} & \dfrac{\partial^2 k_\Phi(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(2)}} - \dfrac{\partial^2 k_\Psi(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(1)}} \\[2ex] \dfrac{\partial^2 k_\Phi(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(1)}} - \dfrac{\partial^2 k_\Psi(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(2)}} & \dfrac{\partial^2 k_\Phi(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial (x')^{(2)}} + \dfrac{\partial^2 k_\Psi(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial (x')^{(1)}} \end{bmatrix}. \tag{10}$$

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 be square exponential kernels, Ψ and Φ are almost surely infinitely differentiable. 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 so long as $\partial^2 k_\Phi(\mathbf{x}_1, \mathbf{x}_2)/\partial x^{(i)} \partial x^{(j)}$ exists (Adler, 1981, Theorem 2.2.2); this condition holds for k_Φ (and k_Ψ) because these are assumed to be square exponential kernels.

Similarly,

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

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

E POSTERIORS USING HELMHOLTZ PRIOR

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 that end, we let $\mathbf{X}_{\mathrm{tr}} \in \mathbb{R}^{2\times M}$ be the matrix with mth column equal to $\mathbf{x}_m$. We define $\mathbf{Y}_{\mathrm{tr}} = (\mathbf{y}_1^{(1)}, \ldots, \mathbf{y}_M^{(1)}, \mathbf{y}_1^{(2)}, \ldots, \mathbf{y}_M^{(2)})^\top \in \mathbb{R}^{2M}$. 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 $\mathbf{X}' = (\mathbf{x}'_1, \ldots, \mathbf{x}'_{N'}) \in \mathbb{R}^{2\times N'}$,

$$
\mu(\mathbf{X}) = \begin{pmatrix} \mu^{(1)}(\mathbf{X}) \\ \mu^{(2)}(\mathbf{X}) \end{pmatrix} \quad \text{and} \quad k(\mathbf{X}, \mathbf{X}') = \begin{pmatrix} k(\mathbf{X}, \mathbf{X}')_{1,1} & k(\mathbf{X}, \mathbf{X}')_{1,2} \\ k(\mathbf{X}, \mathbf{X}')_{2,1} & k(\mathbf{X}, \mathbf{X}')_{2,2} \end{pmatrix} \tag{11}
$$

where (a) for $i \in \{1, 2\}$, $n \in \{1, \ldots, N\}$, $\mu^{(i)}(\mathbf{X})$ is an N-dimensional 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 *tr*aining data with itself, under the full model including noise, as $K_{\mathrm{trtr}} = k_{\mathrm{Helm}}(\mathbf{X}_{\mathrm{tr}}, \mathbf{X}_{\mathrm{tr}}) + \sigma^2_{\mathrm{obs}} \mathbf{I}_{2M}$. Then the log marginal likelihood is

$$
\begin{aligned}
\log p(\mathbf{Y}_{\mathrm{tr}} \mid \mathbf{X}_{\mathrm{tr}}) &= \log \mathcal{N}(\mathbf{Y}_{\mathrm{tr}}; 0, K_{\mathrm{trtr}}) \\
&= -\frac{1}{2}\mathbf{Y}_{\mathrm{tr}}^T K_{\mathrm{trtr}}^{-1} \mathbf{Y}_{\mathrm{tr}} - \frac{1}{2}\log|K_{\mathrm{trtr}}| - \frac{2M}{2}\log 2\pi,
\end{aligned} \tag{12}
$$

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

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

Proposition 4 (Helmholtz posterior)**.** *Consider N new locations at which we would like to predict the current. We gather them in $\mathbf{X}_{\mathrm{te}} \in \mathbb{R}^{2\times N}$, with nth column equal to $\mathbf{x}_n^\star$. We denote the covariance of various* tr*aining and* te*sting combinations as: $K_{\mathrm{tetr}} = k_{\mathrm{Helm}}(\mathbf{X}_{\mathrm{te}}, \mathbf{X}_{\mathrm{tr}})$ and $K_{\mathrm{tete}} = k_{\mathrm{Helm}}(\mathbf{X}_{\mathrm{te}}, \mathbf{X}_{\mathrm{te}})$. Then a posteriori the $2N$-long vector $(F^{(1)}(\mathbf{x}_1^\star), \ldots, F^{(1)}(\mathbf{x}_N^\star), \ldots, F^{(2)}(\mathbf{x}_1^\star) \ldots, F^{(2)}(\mathbf{x}_N^\star)|D)^\top$ describing the current at these locations has a normal distribution with mean*

$$
\mu_{F|D} = K_{\mathrm{tetr}} K_{\mathrm{trtr}}^{-1} \mathbf{Y}_{\mathrm{tr}} \tag{13}
$$

and covariance

$$
K_{F|D} = K_{\mathrm{tete}} - K_{\mathrm{tetr}} K_{\mathrm{trtr}}^{-1} K_{\mathrm{tetr}}^\top. \tag{14}
$$

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 5. *Let $F \sim \mathcal{GP}(\mu, k)$ be a two-output Gaussian process, indexed by $\mathbb{R}^2$, that is continuously differentiable. Then, for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$,*

$$\delta = \operatorname{div} F \sim \mathcal{GP}(\operatorname{div} \mu, k^\delta) \tag{15}$$

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

where

$$k^\delta(\mathbf{x}, \mathbf{x}') = \sum_{(i,j)\in\{1,2\}^2} \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')_{i,j}}{\partial x^{(i)} \partial x^{(j)}} \tag{17}$$

$$k^\zeta(\mathbf{x}, \mathbf{x}') = \sum_{(i,j)\in\{1,2\}^2} (-1)^{i+j} \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')_{i,j}}{\partial x^{(3-i)} \partial x^{(3-j)}}. \tag{18}$$

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,

$$\mathbb{E}(\operatorname{div} F) = \operatorname{div}(\mathbb{E}F) = \operatorname{div} \mu, \tag{19}$$

$$\mathbb{E}(\operatorname{curl} F) = \operatorname{curl}(\mathbb{E}F) = \operatorname{curl} \mu. \tag{20}$$

We next turn to the covariance. Define the centered process $G = F - \mu$. By Equation (19) and Equation (20), $\operatorname{div} G$ 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,

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

□

F COMPUTATIONAL COST OF HELMHOLTZ GP

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

Proposition 6. *Take M training data points. Let $C_{vel}(M)$ and $C_{helm}(M)$ be the computational costs for evaluating the log marginal likelihood (Equation* (12)*) for the velocity GP and Helmholtz GP, respectively. Then*

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

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 $|K_{\mathrm{trtr}}|$. The linear system can be solved in not more than the cost of inverting the matrix. Let $\mathrm{MI}(s)$ be the cost of inverting a square matrix with s rows. Due to the two (correlated) current outputs, the cost of the Helmholtz GP is dominated by $\mathrm{MI}(2M)$. In the velocity GP, the two outputs are uncorrelated and can be handled separately, so the cost is dominated by $2\mathrm{MI}(M)$. So $\lim_{M\to\infty} C_{helm}(M)/C_{vel}(M) \leq \mathrm{MI}(2M)/(2\mathrm{MI}(M))$. Since the cost of matrix inversion is a polynomial with degree at most 3, the result follows. The same considerations apply for the log determinant.

G ADVANTAGES OF THE HELMHOLTZ PRIOR

We 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, the divergence is known a priori to have both a substantively different magnitude and different length scale relative to the vorticity. 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. In the 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 the Φ and Ψ, the marginal variances of δ and ζ scale linearly with σ_Φ^2 and σ_Ψ^2, respectively. The model can therefore easily capture the larger magnitude of the vorticity expected in ocean flows.

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

Proposition 7. *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. Let δ and ζ be defined as in Equations* (7) *and* (8)*. 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 5, the induced divergence and vorticity are Gaussian processes with mean 0 and covariances

---

[1] 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 $\mathbb{R}^2$, $k(\mathbf{x}, \mathbf{x}') = \kappa(\|\mathbf{x} - \mathbf{x}'\|)$. For example, squared exponential kernels are isotropic.

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

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. ☐

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 $k_\Phi(\mathbf{x},\mathbf{x}') = \kappa(\|\mathbf{x}-\mathbf{x}'\|/\ell)$ and $k_\Psi(\mathbf{x},\mathbf{x}') = \kappa(\|\mathbf{x}-\mathbf{x}'\|/\ell)$, for some $\kappa : \mathbb{R}^+ \to \mathbb{R}$, when we expand out the dependence of $k^{\delta}_{\mathrm{Helm}}$ and $k^{\zeta}_{\mathrm{Helm}}$ with respect to k_Φ and k_Ψ we see that they involve fourth order partial derivatives:

$$
\begin{aligned}
k^{\delta}_{\mathrm{Helm}} &= \left(\frac{\partial^4}{\partial (x^{(2)})^2 \partial ((x')^{(2)})^2} + \frac{\partial^4}{\partial (x^{(2)})^2 \partial ((x')^{(1)})^2} + \frac{\partial^4}{\partial (x^{(1)})^2 \partial ((x')^{(2)})^2} + \frac{\partial^4}{\partial (x^{(1)})^2 \partial ((x')^{(1)})^2}\right) k_\Phi \\
k^{\zeta}_{\mathrm{Helm}} &= \left(\frac{\partial^4}{\partial (x^{(2)})^2 \partial ((x')^{(2)})^2} + \frac{\partial^4}{\partial (x^{(2)})^2 \partial ((x')^{(1)})^2} + \frac{\partial^4}{\partial (x^{(1)})^2 \partial ((x')^{(2)})^2} + \frac{\partial^4}{\partial (x^{(1)})^2 \partial ((x')^{(1)})^2}\right) k_\Psi .
\end{aligned}
\tag{21}
$$

Considering first the potential function and divergence. If $k_\Phi(\mathbf{x},\mathbf{x}') = \ell^{-1}\kappa(\|\mathbf{x}-\mathbf{x}'\|)$, then any second order mixed partial derivative may be written 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 which does not depend on ℓ. The same holds true for the stream function and vorticity. 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 horizontal and vertical velocity components. For example, in a vortex, a longitudinal velocity at six o'clock occurs together with a latitudinal velocity at three o'clock, relative to the center of a vortex. Similarly, divergence at a point induces a latitudinal velocity at six o'clock together with a longitudinal velocity at three o'clock. 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 horizontal and vertical coordinates, but the velocity GP is not.

Proposition 8. *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*

$$\mu_{F|D}(R\mathbf{X}_{\mathrm{te}}, R\mathbf{X}_{\mathrm{tr}}, R\mathbf{Y}_{\mathrm{tr}}) = R\mu_{F|D}(\mathbf{X}_{\mathrm{te}}, \mathbf{X}_{\mathrm{tr}}, \mathbf{Y}_{\mathrm{tr}}). \tag{22}$$

Proposition 8 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.

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 $\mathbf{X}_{\mathrm{tr}} = [(x_1^{(1)}, x_1^{(2)})^\top, \dots, (x_N^{(1)}, x_N^{(2)})^\top]$, then $R\mathbf{X}_{\mathrm{tr}} = [R(x_1^{(1)}, x_1^{(2)})^\top, \dots, R(x_N^{(1)}, x_N^{(2)})^\top] = [((R\mathbf{x}_1)^{(1)}, (R\mathbf{x}_1)^{(2)})^\top, \dots, ((R\mathbf{x}_N)^{(1)}, (R\mathbf{x}_N)^{(2)})^\top]$, 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 the forms of k_Φ and k_Ψ. While the Helmholtz GP assumes k_Φ and k_Ψ are squared exponential kernels, we require only that they are isotropic. We first prove two lemmas that will be used in Proposition 8 proof.

Lemma 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

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

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

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

and see that $\mathrm{grad}k_\Phi(\mathbf{x}, \mathbf{x}') = R\mathrm{grad}k_\Phi(R\mathbf{x}, R\mathbf{x}')R^\top$.

Similarly, for a collections of M locations X we have that $\mathrm{grad}k_\Phi(X, X) = (R \otimes I_M)\mathrm{grad}k_\Phi(RX, RX)(R^\top \otimes I_M)$, 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 $\mathrm{rot}\Psi$, $\mathrm{rot}k_\Psi(\mathbf{x}, \mathbf{x}') := \mathrm{Cov}\left[(\mathrm{rot}\,\Psi)(\mathbf{x}), (\mathrm{rot}\,\Psi)(\mathbf{x}')\right]$. We obtain $\mathrm{rot}k_\Psi(\mathbf{x}, \mathbf{x}') = R\mathrm{rot}k_\Psi(R\mathbf{x}, R\mathbf{x}')R^\top$ and $\mathrm{rot}k_\Psi(X, X) = (R\otimes I_M)\mathrm{rot}k_\Psi(RX, RX)(R^\top \otimes I_M)$.

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

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

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

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

as desired.

□

Lemma 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 1 as

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

□

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 2. The final line is a result of linearity of expectation and the definition of $\mu_{F|D}$, and provides the desired equality. □

Finally, we observe that the equivariance property in Proposition 8 need not hold for velocity GP priors. To see this, consider the following counterexample. Let $F^{(1)} \sim \mathcal{GP}(0, k^{(1)})$ for some nontrivial isotropic $k^{(1)}$. And $F^{(2)} = 0$, a trivial isotropic prior. Take any data $\mathbf{X}_{\text{tr}}$, $\mathbf{Y}_{\text{tr}}$, $\mathbf{X}_{\text{te}}$, and a positive 90° rotation. Due to the trivial prior in the second coordinate, the posterior in the second coordinate has mean $\mu^{(2)}_{F|D}(\mathbf{X}_{\text{te}}, \mathbf{X}_{\text{tr}}, \mathbf{Y}_{\text{tr}}) = 0$. 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. Conversely, 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 (22) does not hold for this velocity GP.

H ROOT MEAN SQUARED ERROR (RMSE) TABLE FOR SIMULATED EXPERIMENTS

In this section, we provide RMSEs for the simulated experiments described in detail in Appendix I. Each row represent an experiment. The first column block shows RMSEs for the velocity field F predictions, for the Helmholtz GP (first column), velocity GP (second column), and D-HNN (third column). The second block shows RMSEs for the divergence δ predictions, and the last block RMSEs for the vorticity ζ.

Table 2: Green identifies the lowest RMSE. Dark green indicates the MSE 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
VORTEX & CONSTANT	0.30	0.49	0.28	0.0	0.57	0.51	0.88	1.08	1.91
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 & SMALL DIV	0.96	2.05	2.14	0.94	0.95	1.89	1.40	2.28	2.64
DUFFING & MEDIUM DIV	0.19	0.60	1.65	0.14	0.50	1.15	0.24	0.26	2.39
DUFFING & BIG DIV	0.41	0.22	1.63	0.08	0.17	1.10	0.48	0.16	2.41

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 (Haza et al., 2018). 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 represent ground truths. Second, third and fourth columns contain, respectively, Helmholtz GP, 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, row 6, 7, and 8 concern vorticity: vorticity predictions, standard deviation and z-values for the two GP models. See Figure 4 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 12 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,

$$\text{RMSE} = \sqrt{\frac{1}{|G|} \sum_{\mathbf{x} \in G} \| F(\mathbf{x}) - \hat{F}(\mathbf{x}) \|_2^2} \tag{23}$$

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

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 G of size 17 x 17, equally spaced over the interval $[-1, 1] \times [-1, 1]$. For each point $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in G$, we compute the vortex horizontal and vertical 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 $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in G$:

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

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

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

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, 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 G, so that we can compare our results with the ground truth, for velocities, divergence, and vorticity. To fit the Helmholtz GP, we initialize the hyperparameters as follows: $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.369, \sigma^2_{\text{obs}} = 0.135$. The objective function of our optimization routine is the log marginal likelihood from Equation (1). 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 consider T optimization steps, where $T = \min(1000, T_{\text{conv}})$, where T_{conv} is the time at which 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}. 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 velocity GP. In this case, the initial hyperparameters are $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.369, \sigma^2_{\text{obs}} = 0.135$. 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 4. 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 Helmholtz GP (center-left), the 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 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 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 Helmholtz GP is 0.24, whereas for the 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 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 Helmholtz GP is very certain that there is no divergence, whereas the uncertainty for the 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 Helmholtz GP, 0.22 for 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 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 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 Helmholtz GP, 1.05 for the 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 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 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.

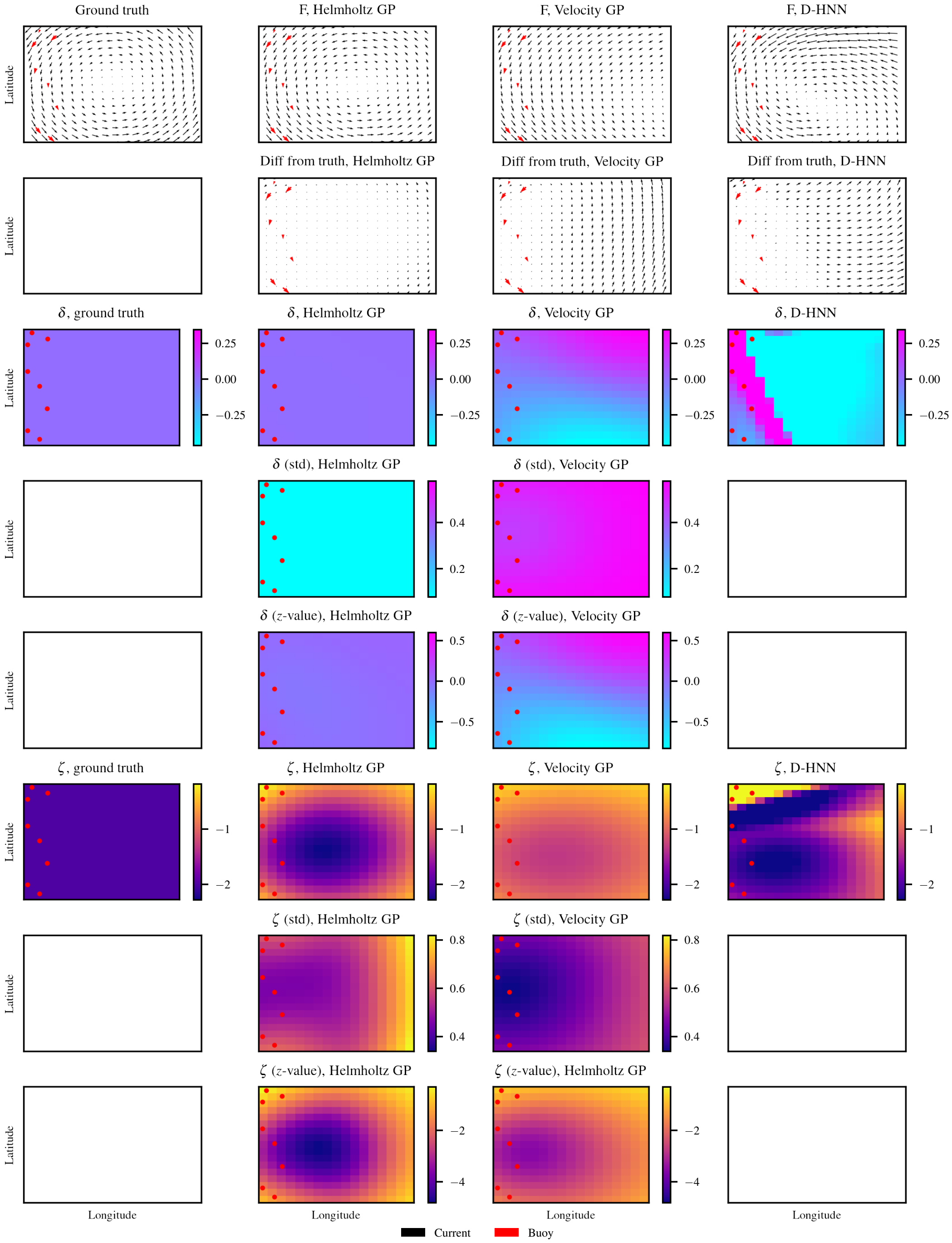

Figure 4: Single vortex. First column: ground truths. Second column: Helmholtz GP results. Third column: Velocity GP results. Fourth column: D-HNN results.

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: a region with a single vortex on top and a constant flow from left to right at the bottom. In this region, fluid particles move in a straight line with a constant velocity. This could be caused by a steady wind or current that is blowing or flowing in one direction.

Simulation details. To simulate such a vortex vector field, we first define a grid of points G 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 $G1$ and $G2$, each of dimension 25 x 25, with $G1$ representing the top grid and $G2$ the lower one. Next, for each point $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in G1$, 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 $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in G1$.

For each point $\boldsymbol{x} \in G2$, 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 $\boldsymbol{x} \in G2$ 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, 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 Helmholtz GP, we initialize the hyperparameters as follows: $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.369, \sigma^2_{\text{obs}} = 0.135$. 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 velocity GP. In this case, the initial hyperparameters are $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.369, \sigma^2_{\text{obs}} = 0.135$. 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 5. 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 Helmholtz GP (center-left), the 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. The 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 goes against the idea that currents are continuous (by conservation of momentum), and is very undesirable. The D-HNN predictions looks very similar to the Helmholtz GP. In the second row we include the differences from the ground truth, and these show as well that Helmholtz GP and D-HNN are accurate, whereas the velocity GP has issues in the lower part of the grid. The RMSE for the Helmholtz GP is 0.30, whereas for the 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. As in the previous experiment, our model prediction (center-left) correctly captures this behavior , whereas the velocity GP (center-right) predicts an irregular pattern not resembling the truth. The same happens for D-HNN as well (right box). The main difference from before is the size of the predicted divergence in the velocity GP, which is now much bigger. This is reflected in the z-values: now there are various regions in the domain where z-values for the velocity GP divergence are beyond the thresholds +1 and -1. This means that the velocity GP predicts significantly non-zero divergence, and this is a big difference compared to our model. We have 0.0 RMSE for the Helmholtz GP, 0.57 for the velocity GP, and 0.51 for the D-HNN.

Finally, in the last three rows we analyze results for the vorticity. The left box shows the ground truth. Here both the GP models' predictions look very similar. Neither is able to model the vorticity well at the boundary of $G1$ and $G2$, where the vorticity is discontinuous. Nonetheless, 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 predicts accurately the boundary, but the magnitudes are slightly off. In terms of standard deviation, both models are quite uncertainty about their predictions, and this is reflected in the z-values plot. We have the following RMSEs: 0.88 for the Helmholtz GP, 1.08 for the velocity GP, and 1.91 for the D-HNN.

In summary, in this experiment we showed a situation in which the Helmholtz GP is (i) at least as good as the other two models in predicting the velocity field, (ii) significantly better than the other models in understanding there is no divergence, and (iii) has the same problems as the velocity GP for predicting a discontinuous vorticity field.

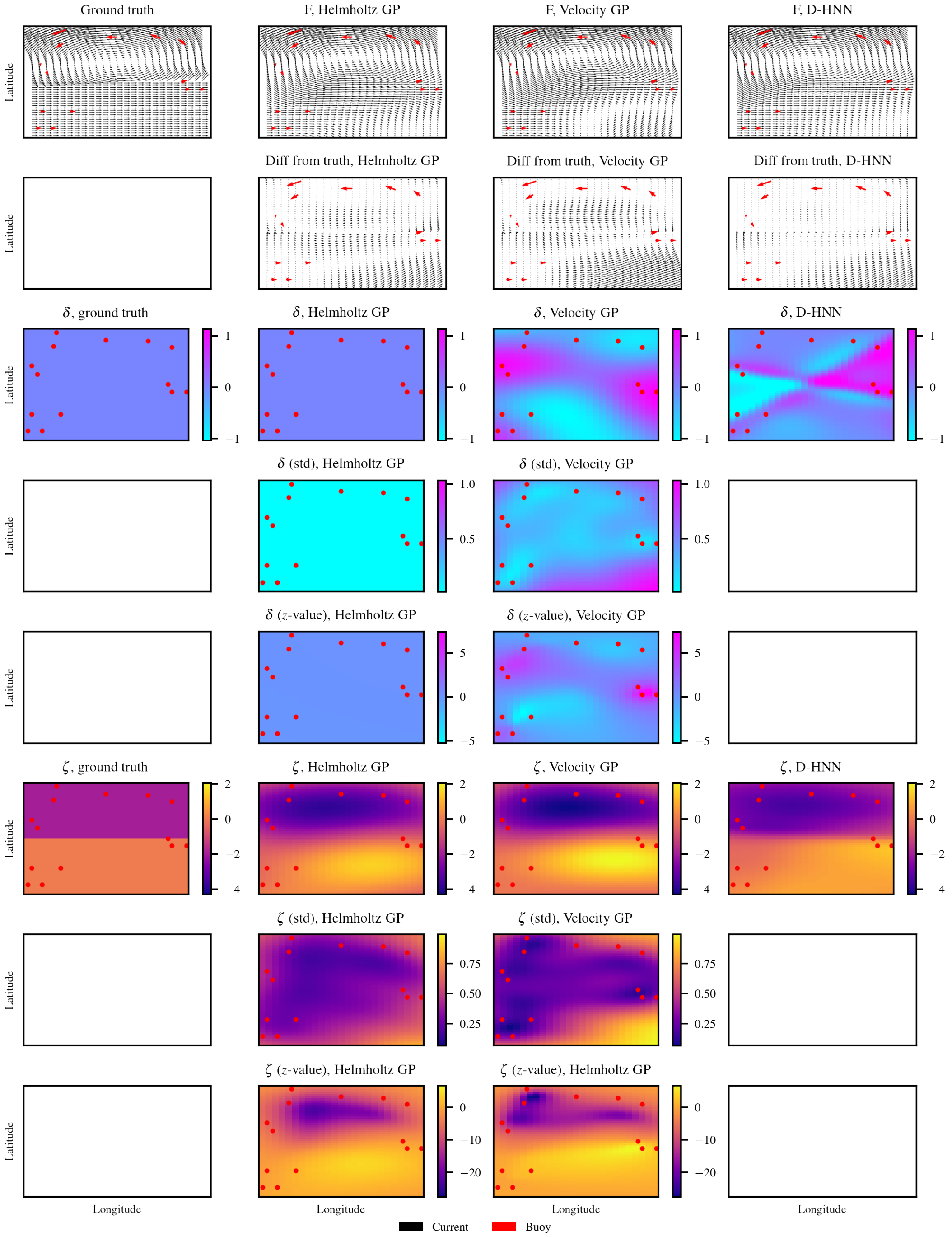

Figure 5: Vortex adjacent to straight current. First column: ground truths. Second column: Helmholtz GP results. Third column: Velocity GP results. Fourth column: D-HNN results.

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 6 to 8. Both models are 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 on this important simulated experiment our model is at least as good as the 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 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 G 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 $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in G$, we can compute the horizontal and vertical 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. It can be seen as a parameter measuring how diffuse the divergence around a center point is. This can be seen 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: $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.369, \sigma^2_{\text{obs}} = 0.135$ for the Helmholtz GP, $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.369, \sigma^2_{\text{obs}} = 0.135$ for the 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 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 velocity GP.

In Figure 6 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 Helmholtz GP (center-left), the 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 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 Helmholtz GP, 1.25 for the 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 desribed 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 velocity GP is more accurate in predicting the size of the divergence area. The 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 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 Helmholtz GP, 1.45 for the 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 Helmholtz GP perfectly predicts zero vorticity, and the D-HNN is almost correct too. The 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 velocity GP has and our model has not. We have 0.0 RMSE for the Helmholtz GP, 1.07 for the 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 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 velocity GP.

Figure 7 shows the results of this scenario. In the top part we have as always the velocity predictions. Since the divergence area is more diffuse, the velocity of the current is generally smaller and evolves more smoothly. Compared to the previous scenario, this makes the prediction task easier for all three models. In particular, the Helmholtz GP and velocity GP predict a field that almost resembles identically the ground truth. The D-HNN still has some issues, specifically it predicts some rotations far away from the observations. This can be seen by looking at the difference from ground truth in the second row. We have the following RMSEs: 0.17 for the Helmholtz GP, 0.19 for the velocity GP, and 0.55 for the D-HNN.

For the divergence, by looking at the the ground truth plot on the left, we see the area of divergence is now more diffuse, and the magnitude is lower. Both the Helmholtz GP and the 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: 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 Helmholtz GP, 0.33 for the velocity GP, 1.32 for the D-HNN.

For the vorticity, we observe that the performances of all models are now worse. The Helmholtz GP still predicts vorticity very close to zero almost everywhere, but not exactly zero as before. The predictions for the velocity GP still looks less accurate and irregular. The D-HNN performance is very poor. In terms of uncertainty, the Helmhotz 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 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 velocity GP are 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 Helmholtz GP, 0.12 for the velocity GP, 0.38 for the D-HNN.

Result: big divergence area, $b_{\textbf{small}} = 15$. 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 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 velocity GP.

In Figure 8 we show the results of this scenario. Here the divergence areas are even more diffuse, and this seems to help a lot the 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 Helmholtz GP are more precise, whereas both the velocity GP and D-HNN predicts some rotational shapes that should not be there. This is confirmed by the RMSE: 0.04 for the Helmholtz GP, 0.10 for the 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 in 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 is due to the distribution of the buoys' observations. The RMSEs are: 0.05 for the Helmholtz GP, 0.12 for the velocity GP, and 0.27 for the D-HNN.

Finally, if we consider the vorticity, we can see how here the 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 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 velocity GP one. The RMSEs are: 0.0 for the Helmholtz GP, 0.10 for the velocity GP, and 0.11 for the D-HNN.

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

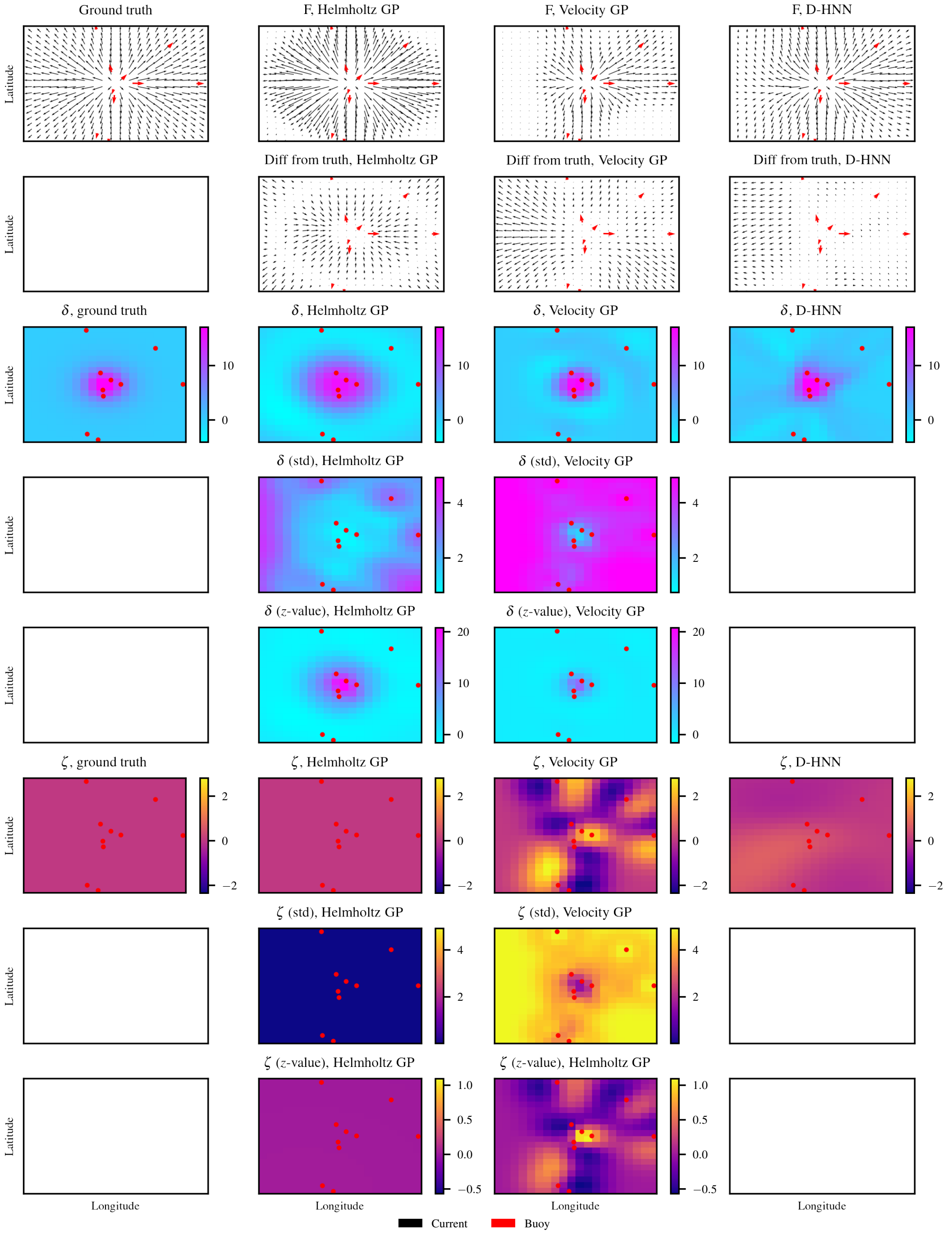

Figure 6: Small area of divergence. First column: ground truths. Second column: Helmholtz GP results. Third column: Velocity GP results. Fourth column: D-HNN results.

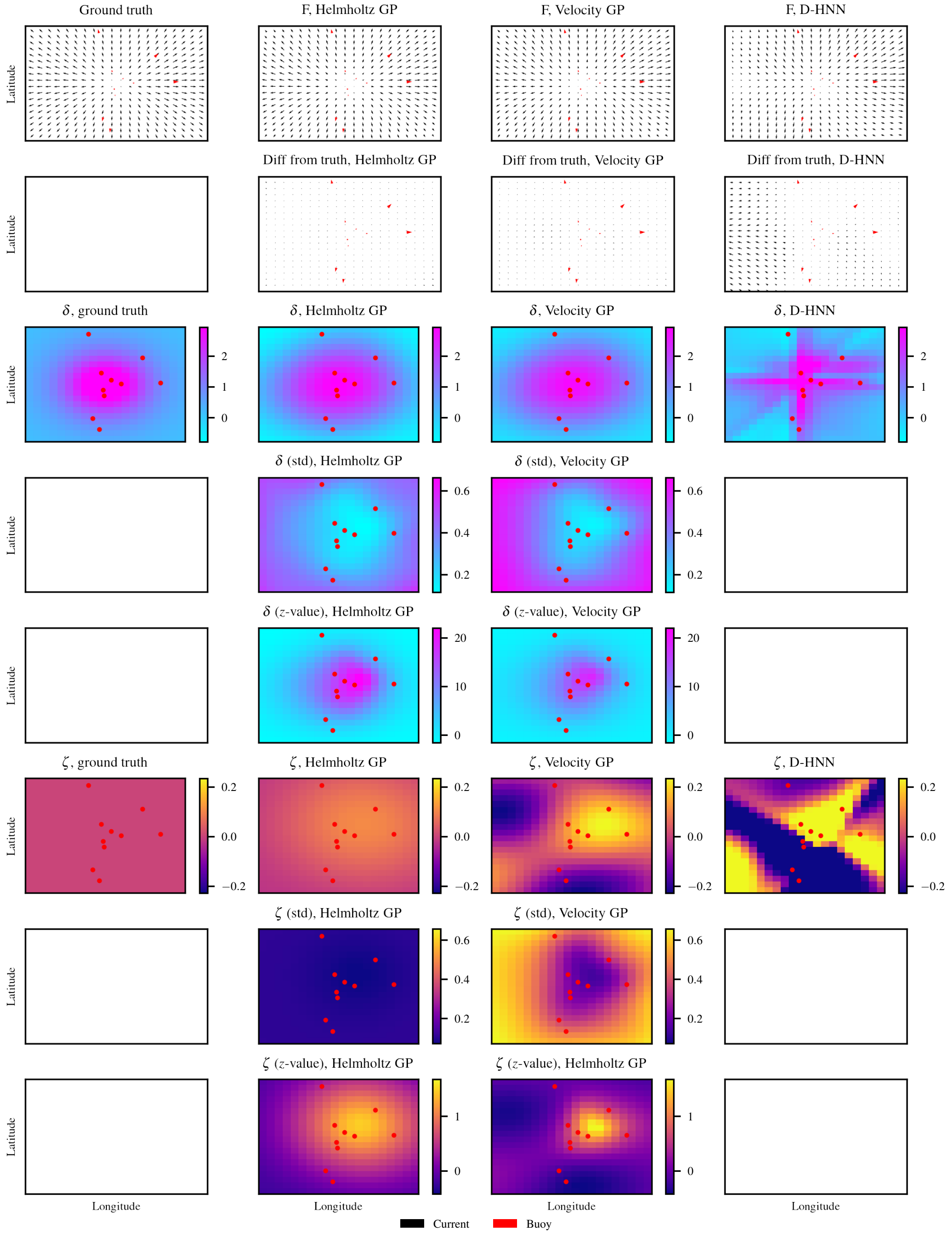

Figure 7: Medium area of divergence. First column: ground truths. Second column: Helmholtz GP results. Third column: Velocity GP results. Fourth column: D-HNN results.

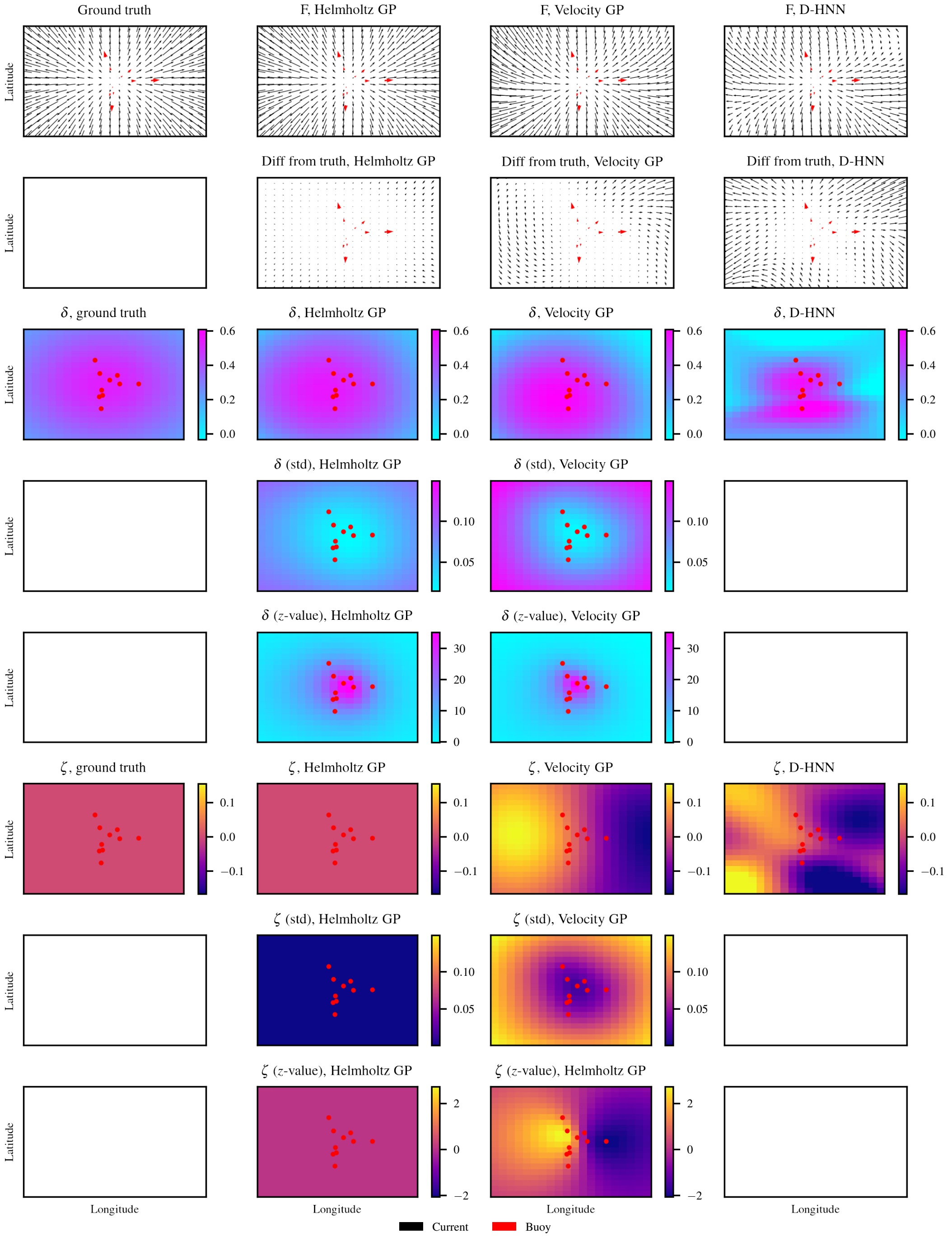

Figure 8: Big area of divergence. First column: ground truths. Second column: Helmholtz GP results. Third column: Velocity GP results. Fourth column: D-HNN results.

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 9. 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 G of size 30 x 30, equally spaced over the interval $[-4, 4] \times [-4, 4]$. Next, for each point $\boldsymbol{x} = (x^{(1)}, x^{(2)}) \in G$, we compute the Duffing horizontal and vertical velocities by:

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

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

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

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:

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

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:

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

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: $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.369, \sigma^2_{\text{obs}} = 0.135$ for the Helmholtz GP, $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.369, \sigma^2_{\text{obs}} = 0.135$ for the 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 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 velocity GP.

In Figure 9 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 Helmholtz GP (center-left), the 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 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 can be seen well from the second row, the comparison to the ground truth. In terms of RMSE, we have 0.96 for the Helmholtz GP, 2.05 for the 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 in 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 Helmholtz GP, 0.95 for the 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 Helmholtz GP prediction look more accurate than the other two. Nonetheless, even our model is not fully able to capture the full vorticity structure. The predictions for the velocity GP lools 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 Helmholtz GP, 2.28 for the 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 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 velocity GP.

Figure 10 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 is generally smoother. This 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 velocity GP, and significantly better than the D-HNN. We have the following RMSEs: 0.19 for the Helmholtz GP, 0.60 for the 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 Helmholtz GP predicts accurately the two areas, with some noise in the central region. The 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 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 velocity GP, the z-values still look good, just less accurate than for our model. The RMSEs are: 0.14 for the Helmholtz GP, 0.50 for the 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 is reflected in the RMSEs: 0.24 for the Helmholtz GP, 0.26 for the 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 Helmholtz GP is superior for the divergence, the 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 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 velocity GP.

In Figure 11 we show the results of this scenario. Here the divergence areas are even more diffuse, and the overall field ends up being smoother. 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 is confirmed by the RMSEs: 0.41 for the Helmholtz GP, 0.22 for the velocity GP, and 1.63 for the D-HNN.

In terms of divergence, the Helmholtz GP accurately predicts the two areas of divergence, still with some noise in the central region. The 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 Helmholtz GP, 0.17 for the 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 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 Helmholtz GP). The D-HNN still fails to predict structure precisely (2.41 RMSE)

In summary, with this experiment we showed that the 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 velocity GP: both models are very good; Helmholtz GP is slightly better for the divergence and velocity GP is slightly better for the vorticity. This is very interesting, showing how both models are able to predict a complex divergence pattern (more complex than the previous experiment).

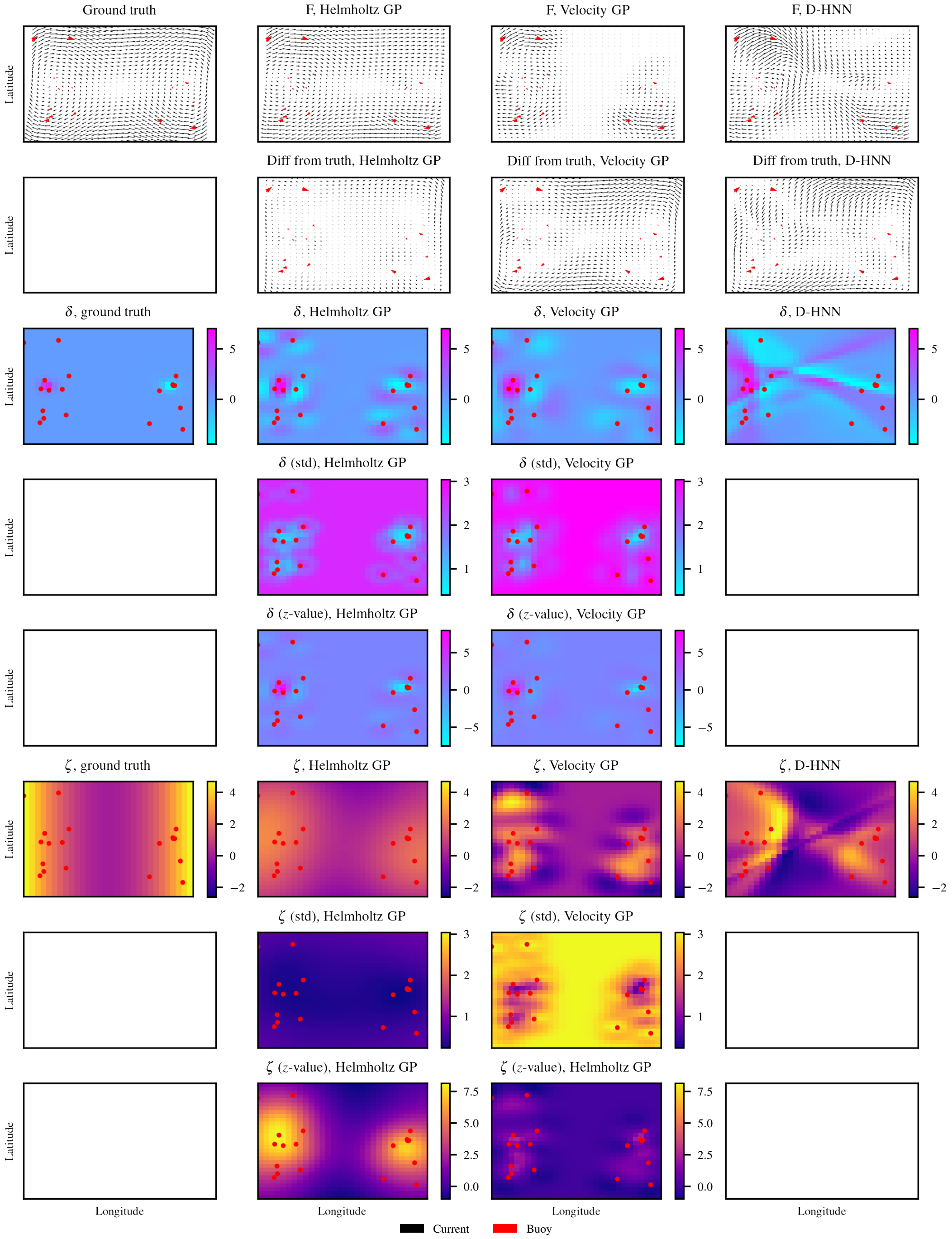

Figure 9: Duffing with small area of divergence. First column: ground truths. Second column: Helmholtz GP results. Third column: Velocity GP results. Fourth column: D-HNN results.

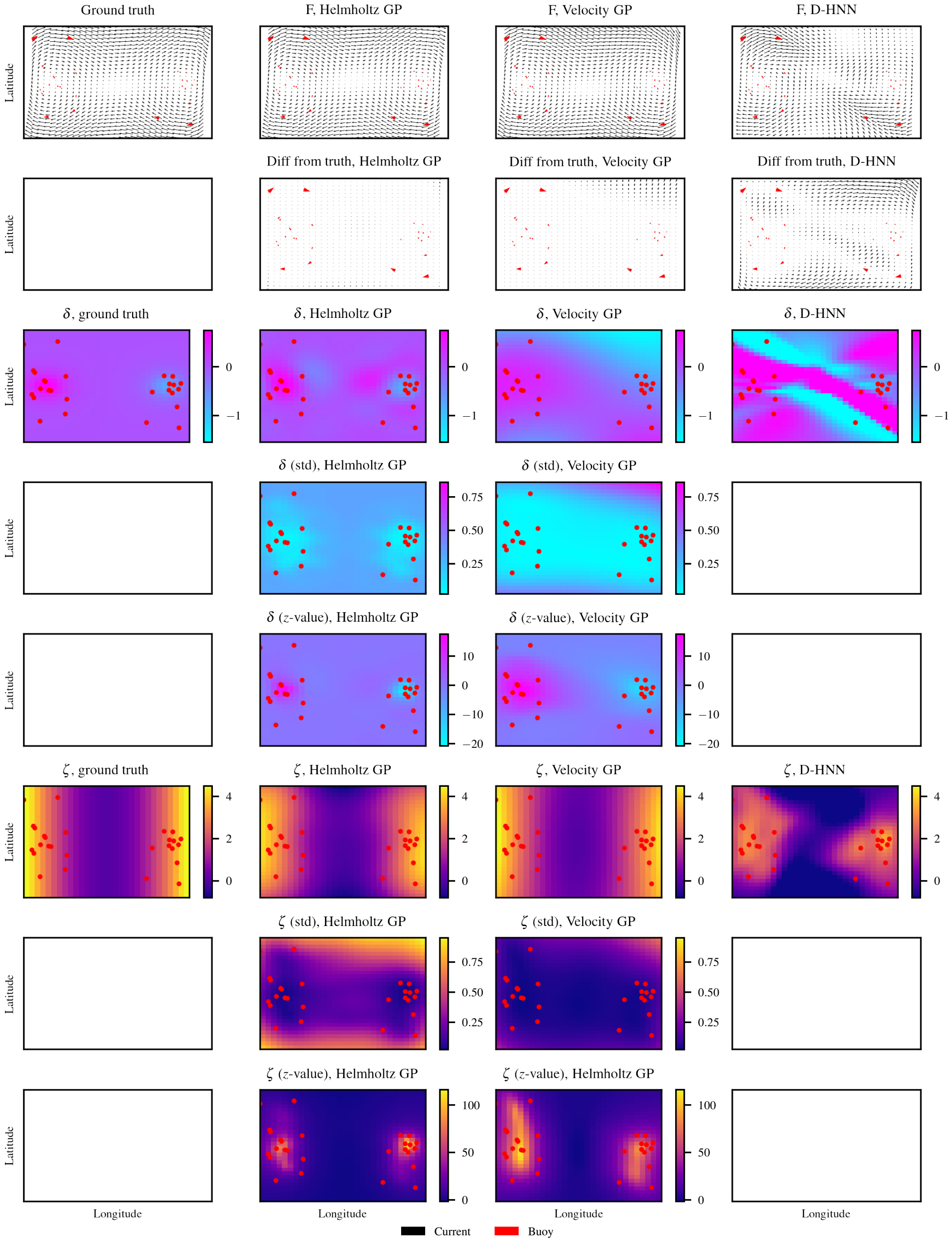

Figure 10: Duffing with medium area of divergence. First column: ground truths. Second column: Helmholtz GP results. Third column: Velocity GP results. Fourth column: D-HNN results.

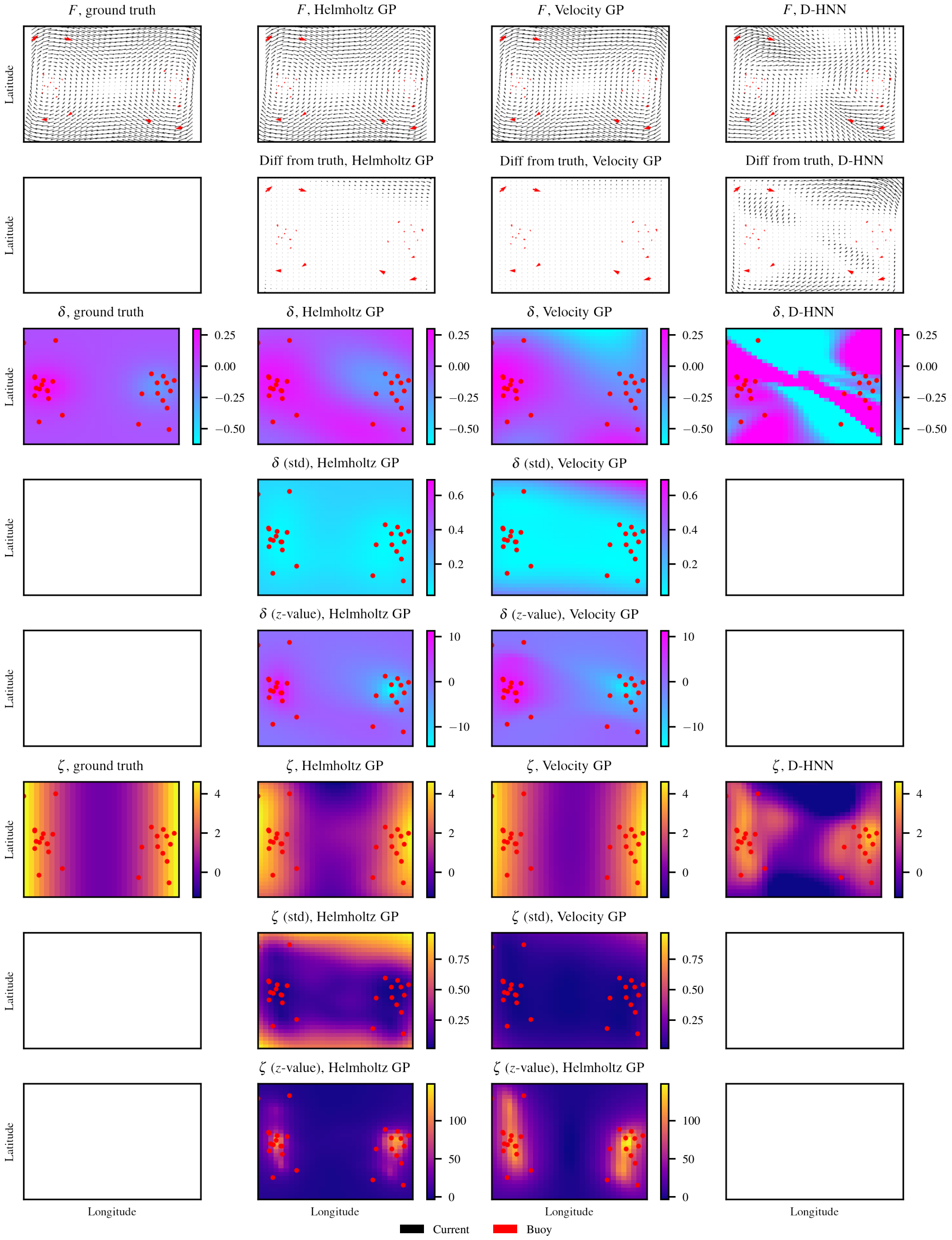

Figure 11: Duffing with big area of divergence. First column: ground truths. Second column: Helmholtz GP results. Third column: Velocity GP results. Fourth column: D-HNN results.

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 could happen when two water masses with different temperatures and/or salinities meet, or when water masses from different directions goes 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 fron 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., $\ell_\Phi = 1, \sigma_\Phi = 1, \ell_\Psi = 2.7, \sigma_\Psi = 0.369, \sigma^2_{\text{obs}} = 0.135$ for the Helmholtz GP, and $\ell_1 = 1, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 0.369, \sigma^2_{\text{obs}} = 0.135$ for the 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 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 12. 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 sparsed 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 (velocity GP and D-HNN do not predict different current around there, 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 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 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 13.

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 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 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.

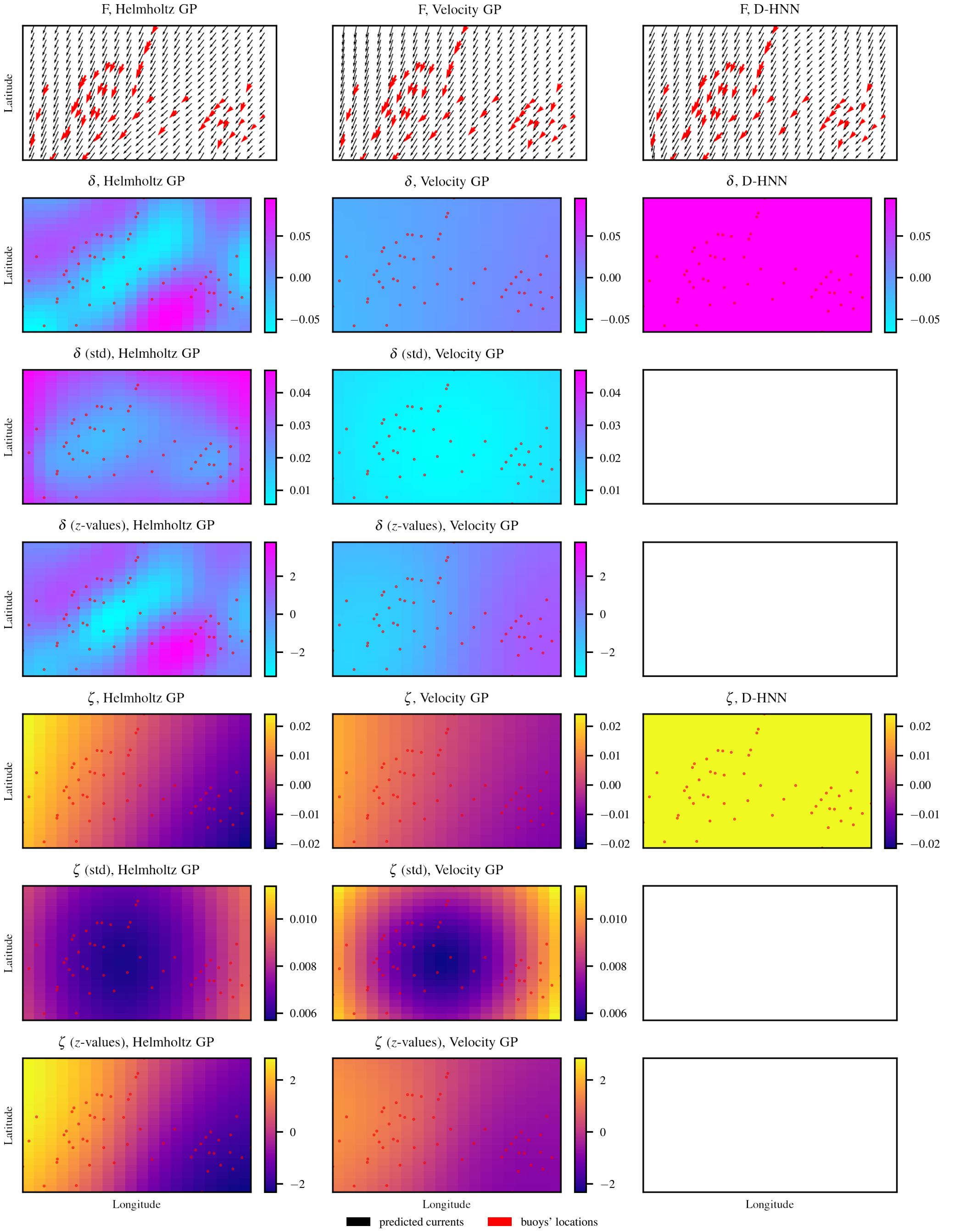

Figure 12: LASER sparse. First column: Helmholtz GP results. Second column: Velocity GP results. Third column: D-HNN results.

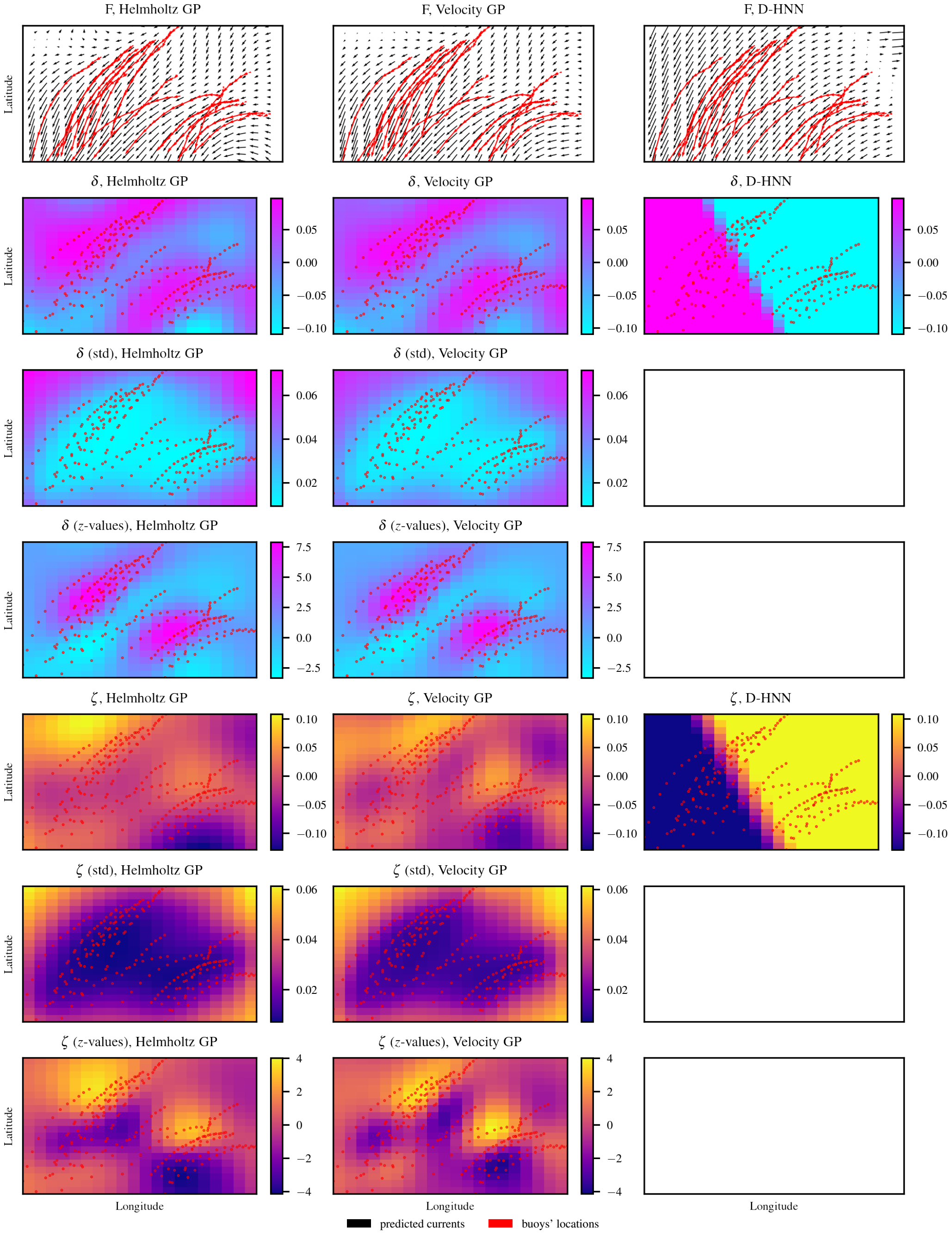

Figure 13: LASER complete. First column: Helmholtz GP results. Second column: Velocity GP results. Third column: D-HNN results.

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 conduced in the summer time with a shallow 20-meter surface mixed layer for the buoys, the wind have 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 of 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 Helmholtz GP and the 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 Helmholtz GP, $\ell_1 = 2.7, \sigma_1 = 1, \ell_2 = 2.7, \sigma_2 = 1, \sigma^2_{\text{obs}} = 0.135$ for the 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 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 velocity GP.

Results. In these data, we expect to see a smooth current, 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 are difficult to verify. We show the results of the experiments in Figure 14. 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 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 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 15. Here we can see that our model starts being affected by the oscillations in the data, predicting less smooth varying currents. 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 velocity GP, the discontinuity issues increase significantly, and the model is still unable to detect vortices. These are two strong motivations to believe the Helmholtz GP provide 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 Helmholtz GP predicts very small divergence almost everywhere, and vorticity coherent with the buoys trajectories. The velocity GP, instead, predicts a reasonable vorticity field, but the divergence shows irregular patterns that look more suspicious. See the second and third block in Figure 14 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 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 block in Figure 15. Here, both models seem to be affected more by the oscillations, but the Helmholtz GP still predicts divergence closer to zero, whereas the 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 velocity GP. Note that we cannot say anything stronger, because there is no expert knowledge suggesting that the Helmholtz GP behavior is the expected one.

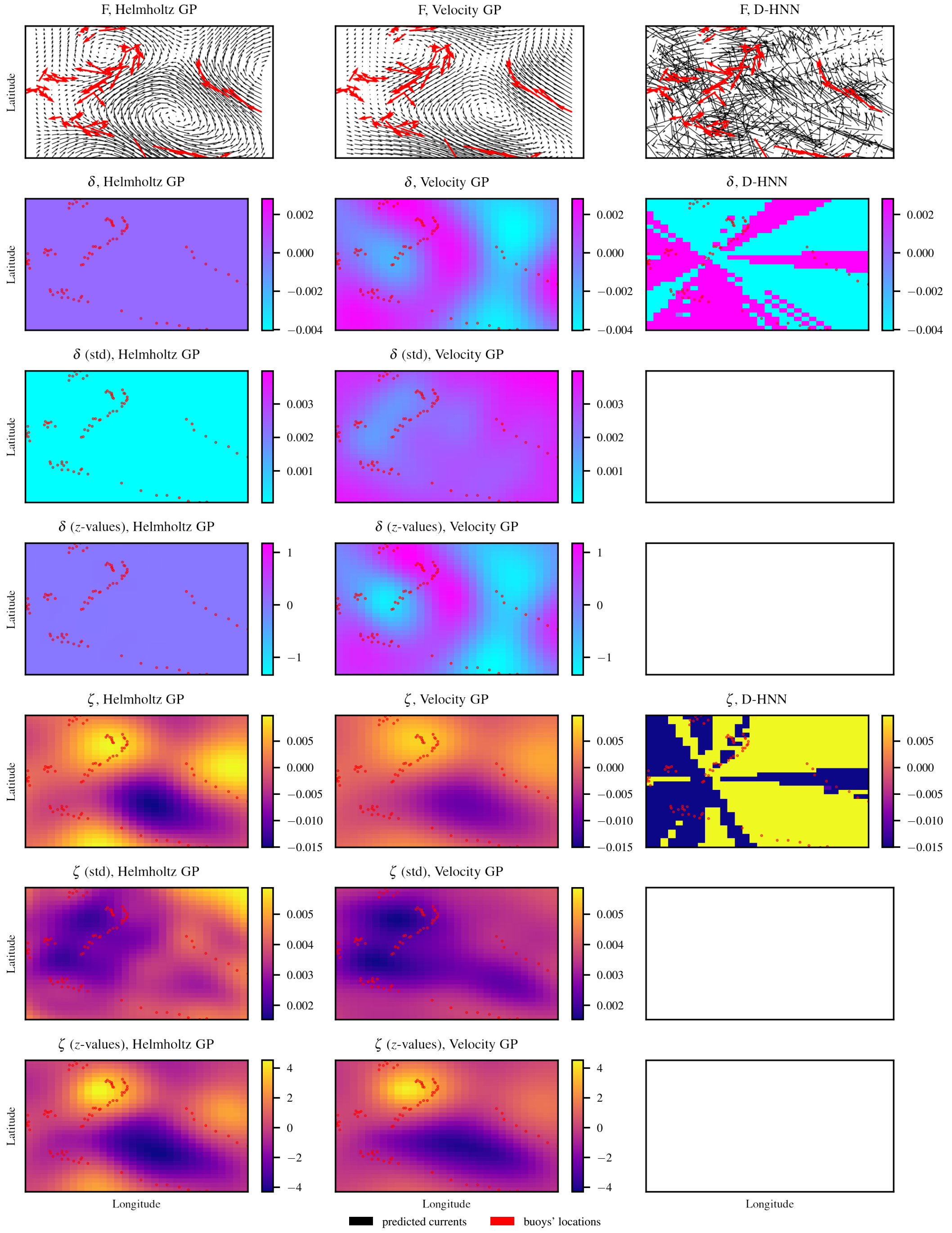

Figure 14: GLAD sparse. First column: Helmholtz GP results. Second column: Velocity GP results. Third column: D-HNN results.

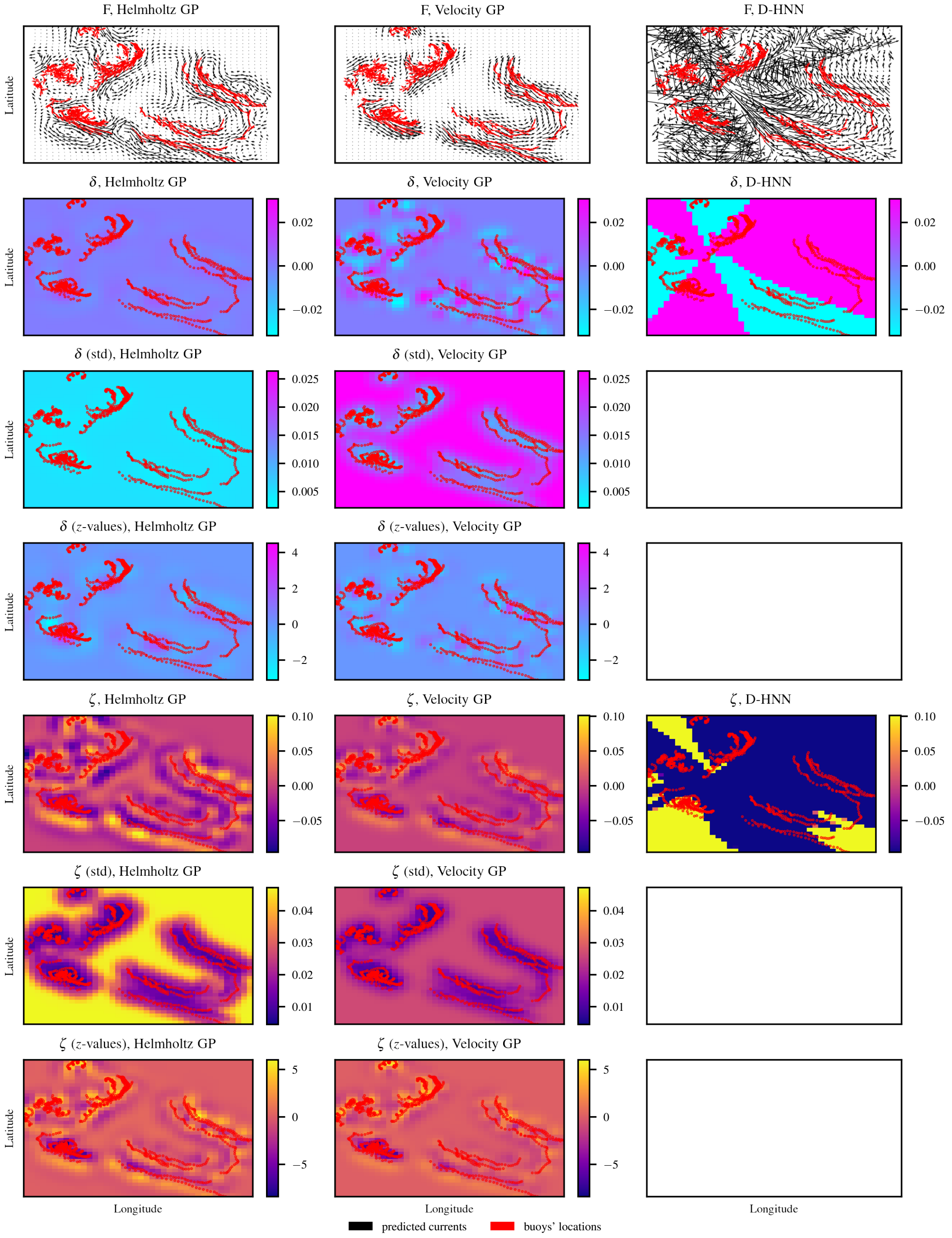

Figure 15: GLAD complete. First column: Helmholtz GP results. Second column: Velocity GP results. Third column: D-HNN results.