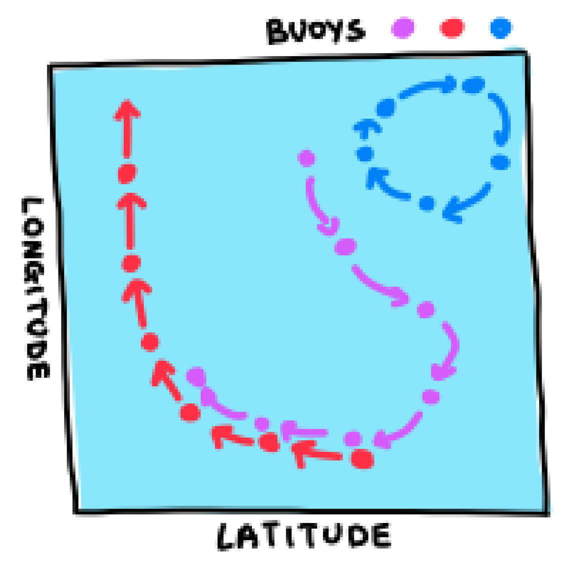
Gaussian Processes at the Helm(holtz)

Renato Berlinghieri[1] Tamara Broderick[1] Ryan Giordano[1] Tamay Özgökmen[2]
Kaushik Srinivasan[3] Brian Trippe[1] Junfei Xia[2]

[1] Laboratory for Information and Decision Systems, Massachusetts Institute of Technology
[2] Rosenstiel School of Marine and Atmospheric Science, University of Miami
[3] Atmospheric and Oceanic Sciences College, University of California Los Angeles

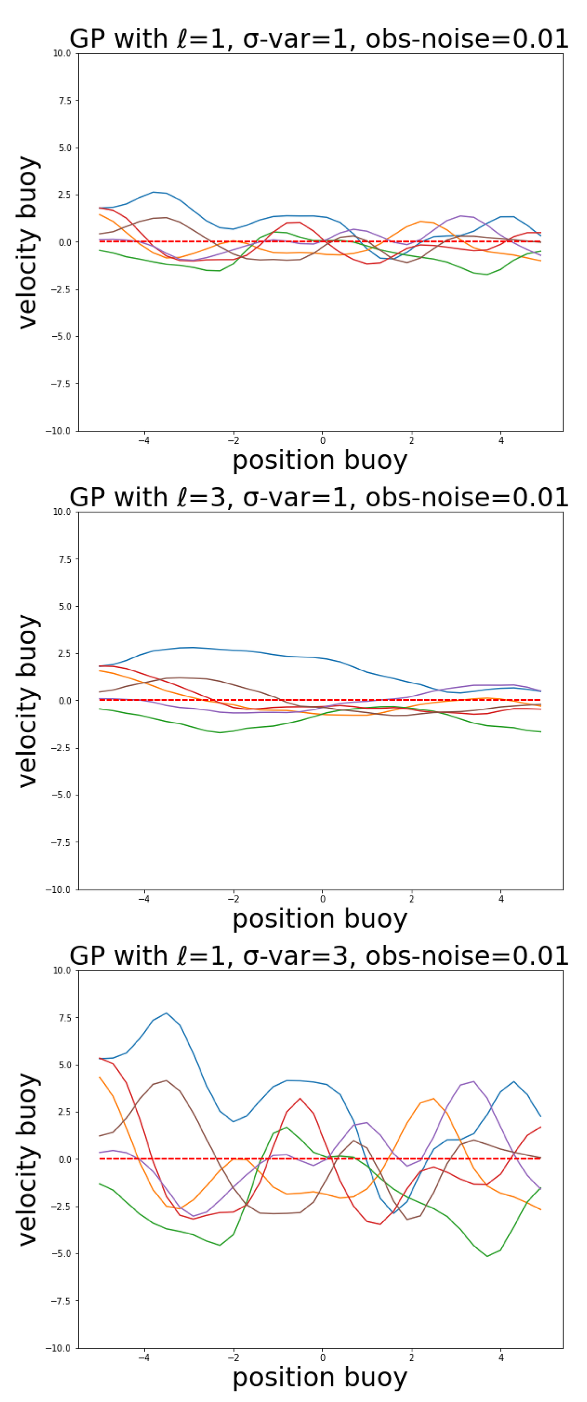
Introduction

- Modeling ocean currents is important for:
  - Forecasting oil spill dispersion
  - Understanding biological productivity
  - Studying energy fluxes
- The statistical framework is complex:
  - Observations are drifter traces from floating buoys
  - Currents vary spatially and temporally
  - The behavior is highly non-linear



- **Goal**: extrapolate a smooth function (*ocean flow*) away from drifters
- Researchers have modeled currents through **Gaussian processes** (GPs)
- Standard approaches in current literature cannot capture some physical properties of interest, e.g., eddies and continuity of currents
- Our approach combines GPs with the **Helmholtz decomposition** and overcomes these issues, with only a small constant multiple of additional computational expense

Background: Gaussian Processes (GP)



- **Motivation**: non-linear regression $\to$ GP is a particular prior over smooth functions
- Formally: a collection of random variables, any finite number of which have a joint Gaussian distribution
- A GP $f(\mathbf{x})$ is completely characterized by its mean and kernel function,

$$m(\mathbf{x}) = \mathbf{E}[f(\mathbf{x})]$$
$$k(\mathbf{x}, \mathbf{x}') = \text{Cov}(f(\mathbf{x}), f(\mathbf{x}'))$$

- Many different kernels $\implies$ many different behaviors. *What is an appropriate kernel choice for our problem?*
- Usual assumption for spatio-temporal problems: we want smooth variations in space/time $\to$ a natural choice is squared-exponential (SE) kernel

$$k(x_i, x_j) = \underbrace{\sigma_f^2}_{\textit{signal var}} \cdot \exp\left(-\frac{(x_i - x_j)^2/2}{\underbrace{\ell^2}_{\textit{lengthscale}}}\right) + \underbrace{\sigma_n^2}_{\textit{obs noise}} \delta_{ij}$$

- Question: *how do we put a prior over ocean currents?*
- Standard solution [1] : independent SE GP priors over velocity components
- **Issue**: this approach does not capture eddies and continuity of currents – see figure in experiments' section
- *Is there room for improvement?* Yes! Use domain knowledge $\to$ Helmholtz decomposition

Background: Ocean flows + Helmholtz

- Ocean flows can be characterized by two quantities:
  - **Divergence**, measures expansion/contraction/translation $\to$ characterizes up/downwelling (particular interest for oceanographers)
  - **Vorticity**, measures rotation $\to$ usually large, long timescale
- Question: *can we represent this mathematically?*

- **Helmholtz decomposition**: express a smooth vector field $\mathsf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ as a sum of the gradient of a scalar potential and the curl of a scalar potential

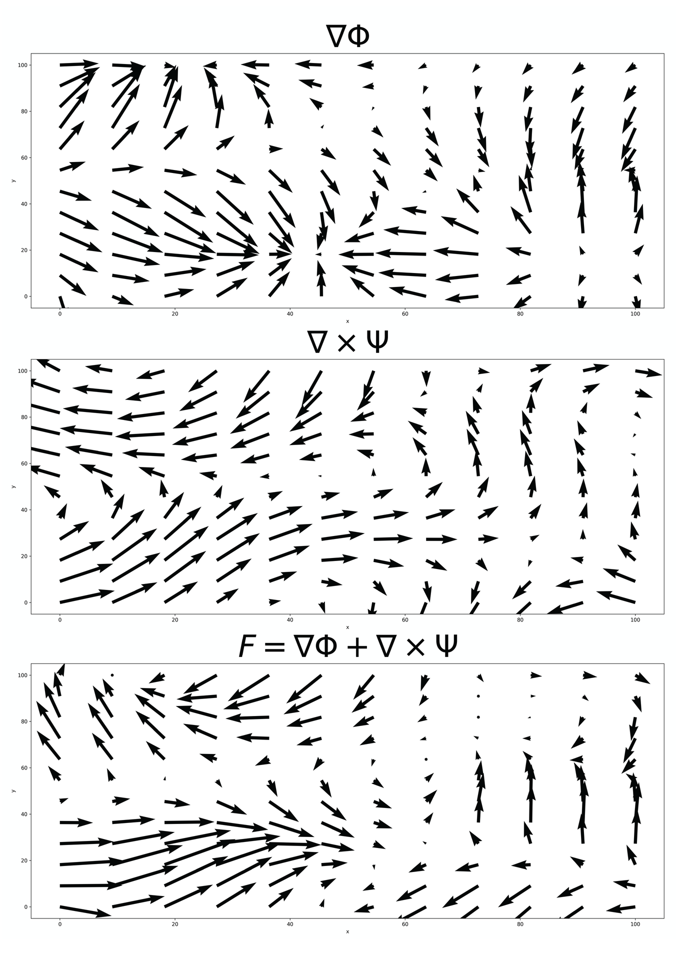
$$\underbrace{\mathsf{F}}_{\textit{ocean flow}} = \underbrace{\nabla\Phi}_{\textit{divergence}} + \underbrace{\nabla \times \Psi}_{\textit{vorticity}}$$

where

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

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

- **Goal**: Estimate Φ and Ψ from data $\to$ extrapolate vector field F.



Our method: GPs @ the Helm(holtz)

- Model drifter traces, $(\mathsf{Y}^{(1)}, \mathsf{Y}^{(2)})$, as sparse noisy observations of a 2D vector field, $\mathsf{F} : \mathbb{R}^2 \to \mathbb{R}^2$, mapping spatial locations, $\mathbf{x} = (x^{(1)}, x^{(2)})$, into horizontal and vertical velocities, $(\mathsf{F}^{(1)}, \mathsf{F}^{(2)})$
- **Likelihood** : independently for each observation,

$$\begin{bmatrix} \mathsf{Y}^{(1)}(\mathbf{x}) \\ \mathsf{Y}^{(2)}(\mathbf{x}) \end{bmatrix} \mid \mathsf{F} \overset{\text{ind}}{\sim} \mathcal{N}\left(\mathsf{F}(\mathbf{x}), \begin{bmatrix} \sigma^2_{\text{obs}} & 0 \\ 0 & \sigma^2_{\text{obs}} \end{bmatrix}\right)$$

with σ^2_{obs} observation noise.
- **Helmholtz prior**: independent SE priors on the two Helmholtz components

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

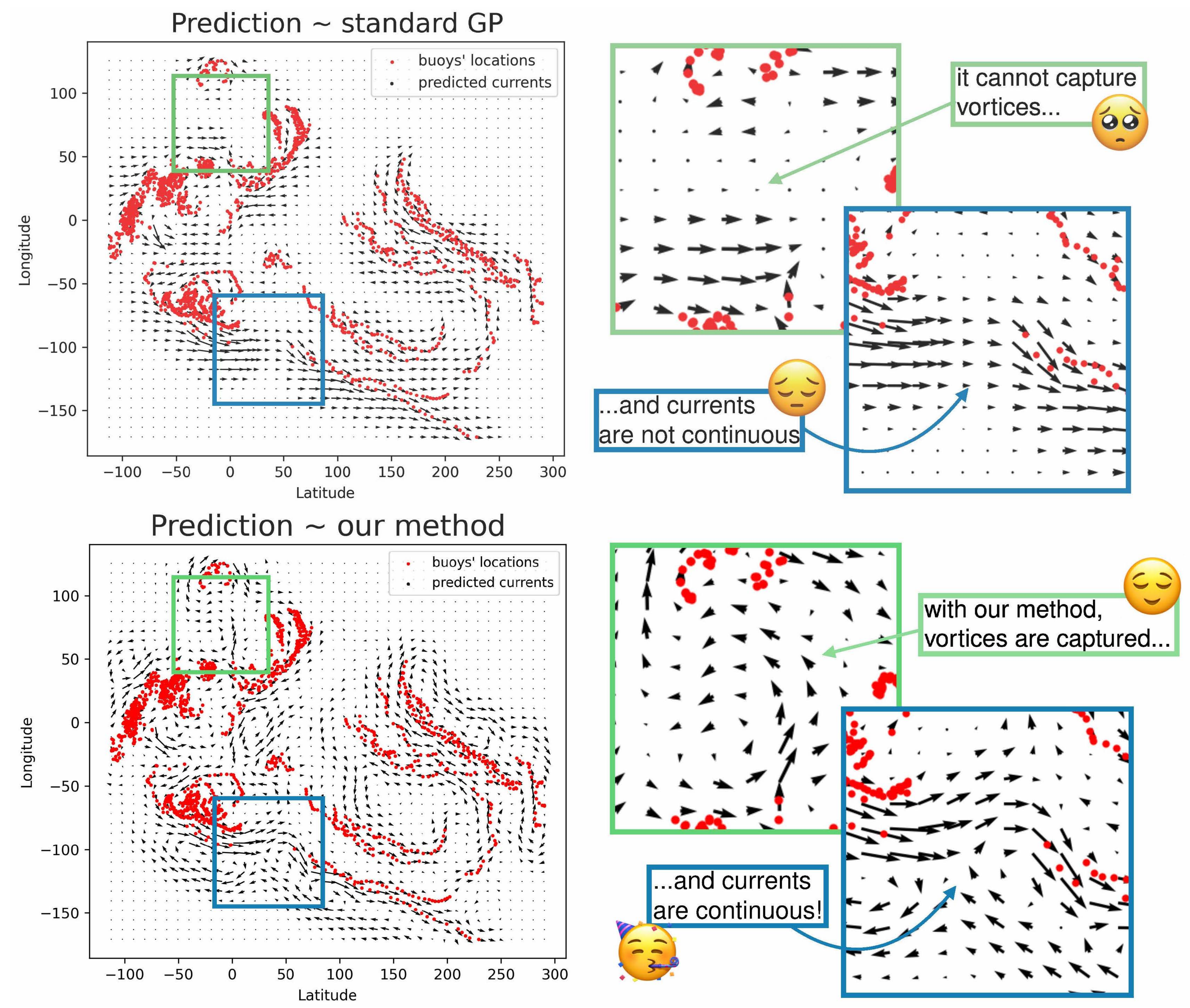
- **Key tool** : the derivative of a GP is a GP [2]. Therefore, e.g., $\nabla\Phi \sim \text{GP}(0, \nabla K_\Phi)$ where, for each pair of datapoints $\mathbf{x}$, $\mathbf{x}'$,

$$\nabla K_\phi(\mathbf{x}, \mathbf{x}') := \text{Cov}\left(\underbrace{\begin{bmatrix} \frac{\partial\Phi(\mathbf{x})}{\partial x^{(1)}} \\ \frac{\partial\Phi(\mathbf{x})}{\partial x^{(2)}} \end{bmatrix}}_{\nabla\Phi(\mathbf{x})}, \underbrace{\begin{bmatrix} \frac{\partial\Phi(\mathbf{x}')}{\partial x'^{(1)}} \\ \frac{\partial\Phi(\mathbf{x}')}{\partial x'^{(2)}} \end{bmatrix}}_{\nabla\Phi(\mathbf{x}')}\right) = \begin{bmatrix} \frac{\partial^2 K(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial x'^{(1)}} & \frac{\partial^2 K(\mathbf{x}, \mathbf{x}')}{\partial x^{(1)} \partial x'^{(2)}} \\ \frac{\partial^2 K(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial x'^{(1)}} & \frac{\partial^2 K(\mathbf{x}, \mathbf{x}')}{\partial x^{(2)} \partial x'^{(2)}} \end{bmatrix}$$

$\implies$ **we show** that with the Helmholtz prior we are still able to make predictions over the original space of interest!

Experiments

- Real data: 60 buoys floating in the Gulf of Mexico for 5 days, measured hourly
- For simplicity, collapse the time dimension and focus on the spatial inference



- Both approaches learn correct magnitude and direction when close to buoys' observations
- While the standard approach fails to capture eddies and continuity of currents, our approach allows us to do so
- *Is there room for improvement?* Yes! Our model tends to create/complete vortices even when not appropriate $\to$ potential solution: combine the two approaches, e.g., by summing up the two kernels

Future work

- Introduce multiple lengthscales in the kernel definition
- Consider time-varying vector fields, i.e., put back time dimension
- Evaluate model on physics simulation (known ground truth)

References

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

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