Scalable spatio-temporal inference for large-scale ocean data: From classical methods to simulation-based inference

Thea Sukianto[1], Donata Giglio[2] and Mikael Kuusela[1]

Carnegie Mellon University[1], University of Colorado Boulder[2]

Introduction

- Reliable ocean heat content (OHC) estimates and uncertainties are crucial for understanding the evolution of Earth's climate ($\sim$ 91% of excess heat in the climate system is stored in the ocean)
- The Argo float array is uniquely capable of observing changes in OHC at various spatial and temporal scales
- Estimating OHC from Argo temperature profile data requires careful use of spatio-temporal mapping and new statistical methodology for uncertainty quantification

Challenges in estimating OHC from Argo data:

1. Data availability in vertical dimension
2. Nonstationarity
3. Large dataset size (1.7 million+ temperature profiles)

Here we introduce our locally stationary Gaussian process-based mapping framework and leverage neural inference to improve the computational efficiency.

Mapping Methodology

- **Quantity of interest**

We wish to estimate OHC at time t which is the triple integral of potential temperature T over space S and pressure z:

$$\mathrm{OHC}_{z_u}^{z_d}(t) = \rho_0\, c_{p,0} \iint_{(x,y)\in D} \int_{z=z_u}^{z=z_d} T(x,y,z,t)\,\mathrm{d}z\,\mathrm{d}S(x,y)$$

However, OHC is only observed at profile locations, so we need statistical interpolation to map onto a regular grid to estimate regional/global OHC:

$$\approx \sum_{i,j} \widetilde{\mathrm{OHC}}_{z_u}^{z_d}(x_{i,j}, y_{i,j}, t)\cdot S_{i,j}$$

- **Locally stationary Gaussian process (GP) regression**

Our OHC mapping framework is based on the following statistical model

$$\widetilde{\mathrm{OHC}}_{z_u}^{z_d}(x,y,t) = \mu(x,y,t) + a(x,y,t) + \varepsilon(x,y,t),$$

where μ is a climatological mean field, a is a zero-mean anomaly field, and ε is a zero-mean nugget effect.

Due to sampling inhomogeneities in the vertical dimension, we model the upper ocean (15-975 dbar) and midocean (975-1850 dbar) anomalies separately.

We use the locally stationary Gaussian process regression model from Kuusela and Stein (2018)

$$a_i \overset{\text{i.i.d.}}{\sim} \mathrm{GP}(0, k(s_1,t_1,s_2,t_2;\theta_W)),\ \varepsilon \overset{\text{i.i.d.}}{\sim} N(0,\sigma_W^2)$$

where $k = \phi_W \exp(-d(s_1,t_1,s_2,t_2))$ is an anisotropic exponential space-time covariance function with distance function

$$d(s_1,t_1,s_2,t_2) = \sqrt{\left(\frac{x_1-x_2}{\theta_{W,\mathrm{lon}}}\right)^2 + \left(\frac{y_1-y_2}{\theta_{W,\mathrm{lat}}}\right)^2 + \left(\frac{t_1-t_2}{\theta_{W,t}}\right)^2}$$

- **Parameter estimation and kriging**

We focus on a window W around a spatio-temporal grid point $s^* = (x^*, y^*, t^*)$ and regard each year i as an i.i.d. replicate from the above GP.

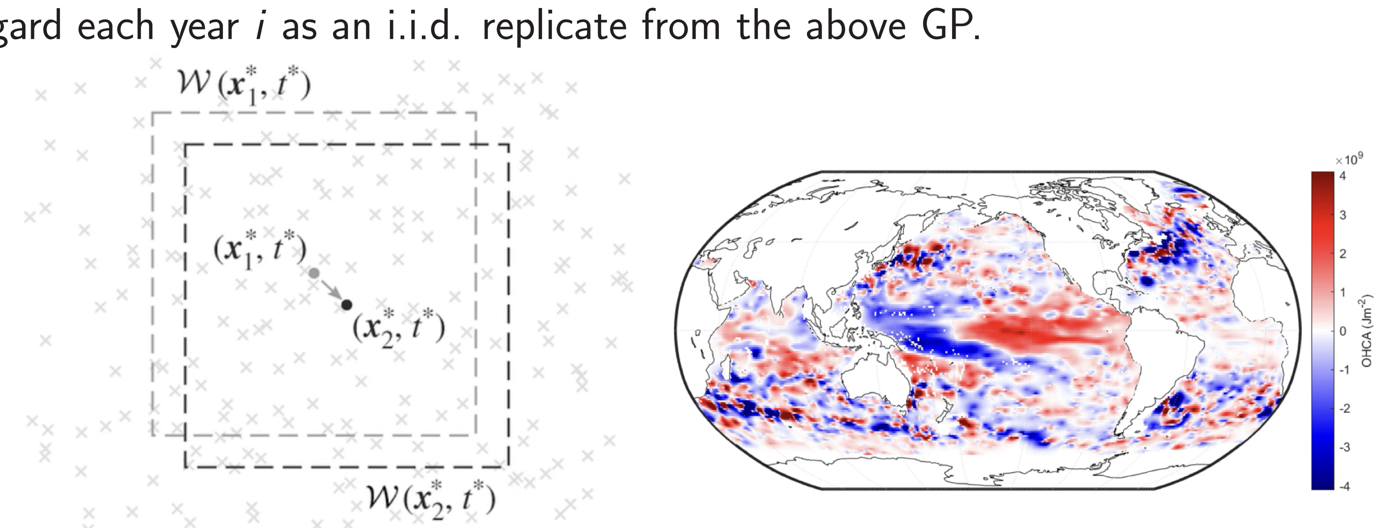


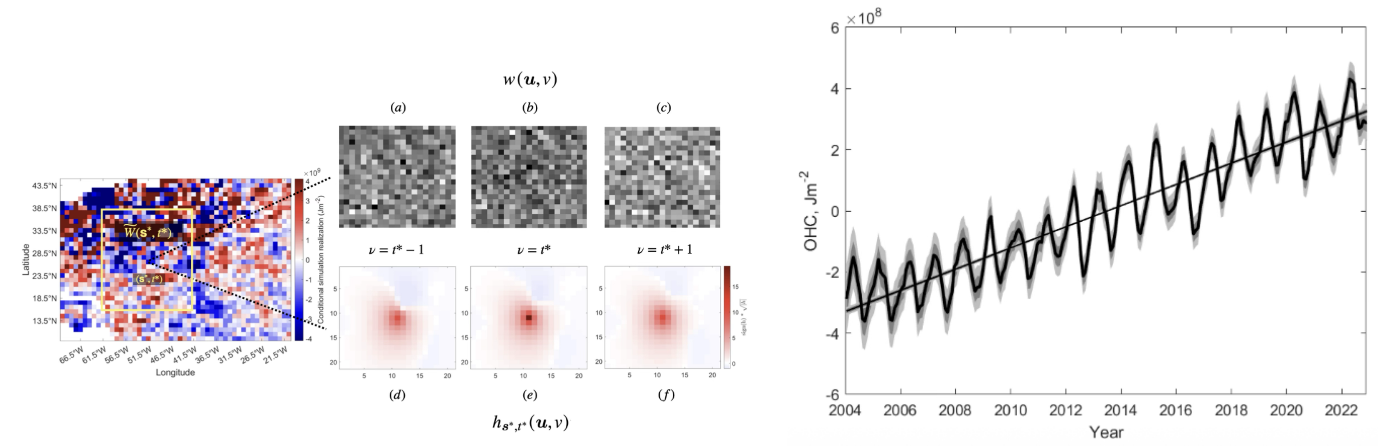
Figure: Window (left) and upper ocean mapped OHC anomaly (February 2010; right)

This produces spatially-varying parameter estimates and point predictions!

Uncertainty Quantification

- **Local conditional simulation**

Since computing the predictive covariance between every grid point across the global ocean is infeasible, we developed a novel local conditional simulation algorithm to simulate from $p(\mathrm{OHC} \mid \text{data})$ based on convolving Gaussian white noise with a kernel.



This algorithm produces an ensemble of conditional simulation realizations (left), which we have used to contribute OHC uncertainty estimates (right) to climatological intercomparison efforts led by JPL and NOAA (e.g., Hakuba et al. 2024)

Likelihood Estimation with Neural Inference

To estimate the local GP model parameters, we need to maximize the log-likelihood $\log(\mathcal{L}_W(\theta_f, \theta_\epsilon | y_t)) =$

$$-\frac{1}{2}\sum_{t\in D_t}[\log\det(K_t(\theta_f) + \Sigma_\epsilon(\theta_\epsilon)) + y_t^\top (K_t(\theta_f) + \Sigma_\epsilon(\theta_\epsilon))^{-1} y_t + n_t \log(2\pi)]$$

for $\approx$ 30,000 grid points. Note that the determinant/inverse of the total covariance matrix are $\mathcal{O}(n_t^3)$, which means directly maximizing the likelihood can quickly become costly.

- **Neural likelihood**

Neural likelihood (e.g., Walchessen et al. 2024) relies on the "classifier trick", i.e. training a neural network classifier h so

$$\mathcal{L}(\theta|y_t) \propto \frac{h(y_t,\theta)}{1-h(y_t,\theta)}$$

Evaluating the classifier creates neural likelihood surfaces that can be used to produce parameter estimates $\Rightarrow$ Neural likelihood is amortized, so we just need to train the network once and we can produce vectorized MLEs!

- **Extension to irregular data with GNNs**

We extend neural likelihood to the irregular setting using a graph neural network (GNN) for the first time to anisotropic spatio-temporal processes, inspired by the neural Bayes estimator architecture from Sainsbury Dale et al. (2025).

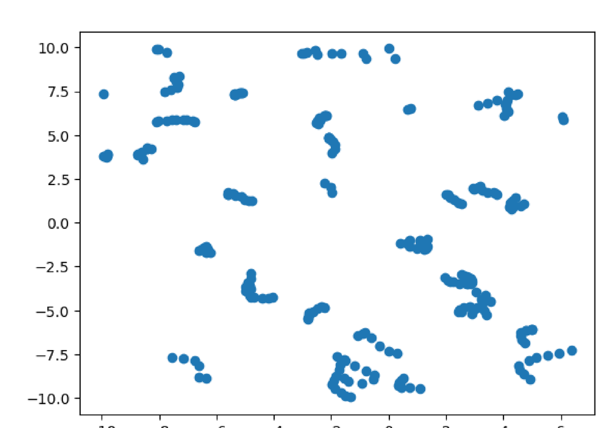


Figure: Argo sampling pattern in 20° $\times$ 20° spatial window (time flattened)

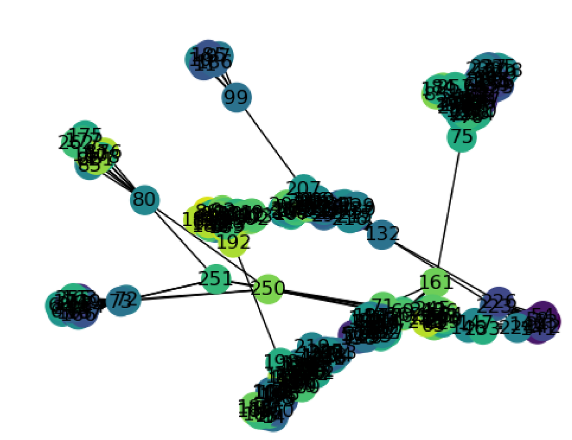
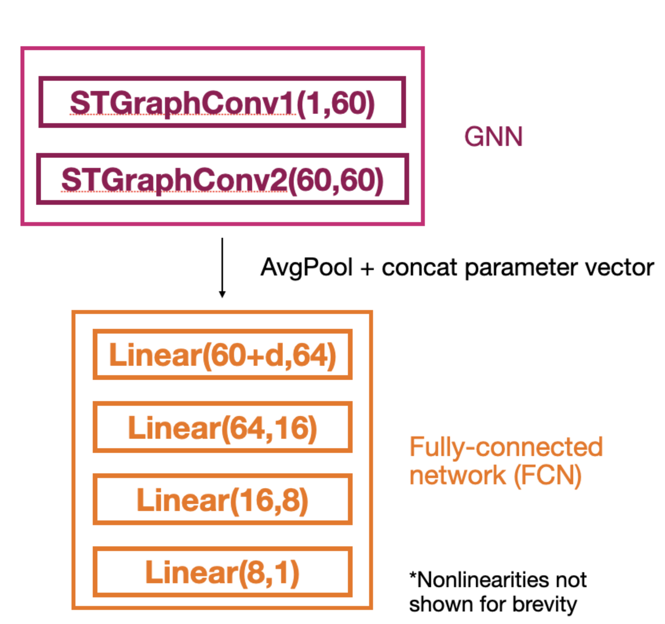


Figure: Graph representation (nodes = GP field value, edges = lat, long, time distances)



- The graph first passes through two graph convolution layers defined by the following node update:

$$x^{(l)} = W_1 x_j^{(l-1)} + W_2^{(l)} \sum_{j'\in\mathcal{N}(j)} \omega^{(l)}(e_{j,j'}) f^{(l)}(x_j^{(l-1)}, x_{j'}^{(l-1)})$$

- The output is then averaged for each channel and concatenated with the parameter vector, which is passed into the FCN part of the classifier that produces the likelihood
- Note that we can vectorize (batch evaluate) over many windows at once, which makes parameter estimation fast

Likelihood Surfaces and Validation

Training data: 600,000 upper ocean Argo sampling patterns from 2004-2022 (60,000 parameter vectors x 10 patterns)

Target: Local GP with anisotropic exponential covariance function parameterized by nugget-total variance ratio $\frac{\sigma_W^2}{\sigma_W^2+\phi_W}$ and length scales $\theta_{W,\mathrm{lon}}, \theta_{W,\mathrm{lat}}, \theta_{W,t}$

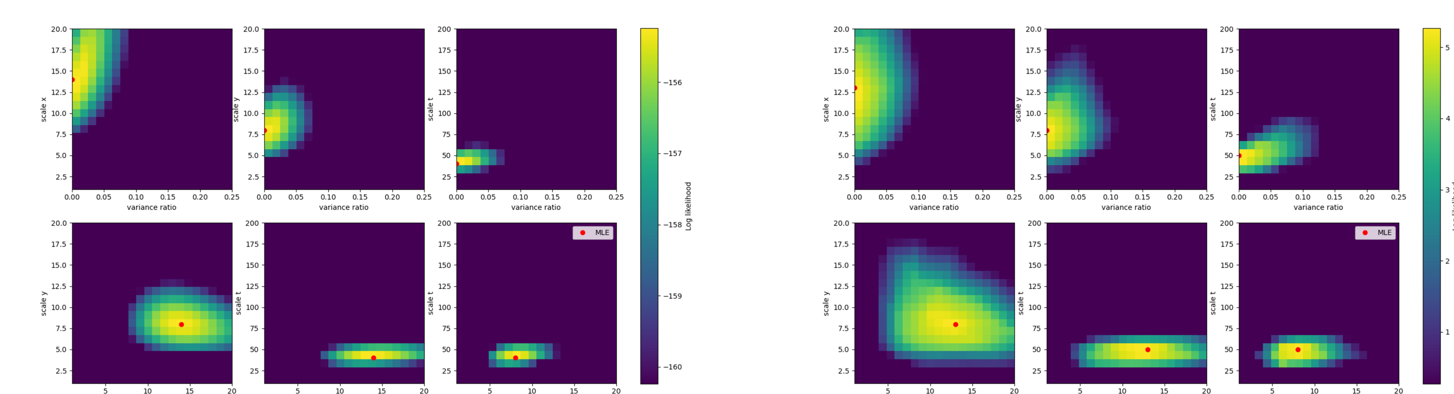


Figure: Exact (left) and neural (right) likelihood surfaces for a 20° $\times$ 20° $\times$ 3 month window in the Niño 3.4 region in February 2010. Note that the neural surface is $\approx$ 350x faster!

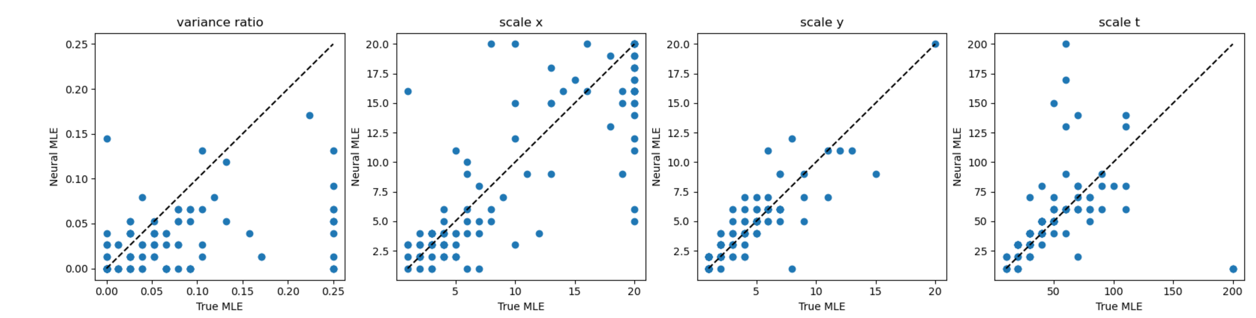


Figure: Exact (x-axis) vs. neural (y-axis) MLEs for windows centered on 100 grid points in the equatorial region (10° N-10° S) in February 2010.

Preliminary Global MLE Results

We apply the above GNN classifier to find the MLE across the full Argo sampling period from 2004-2022 (i.e., produce yearly neural likelihood surfaces as above and sum across years before finding the maximum):

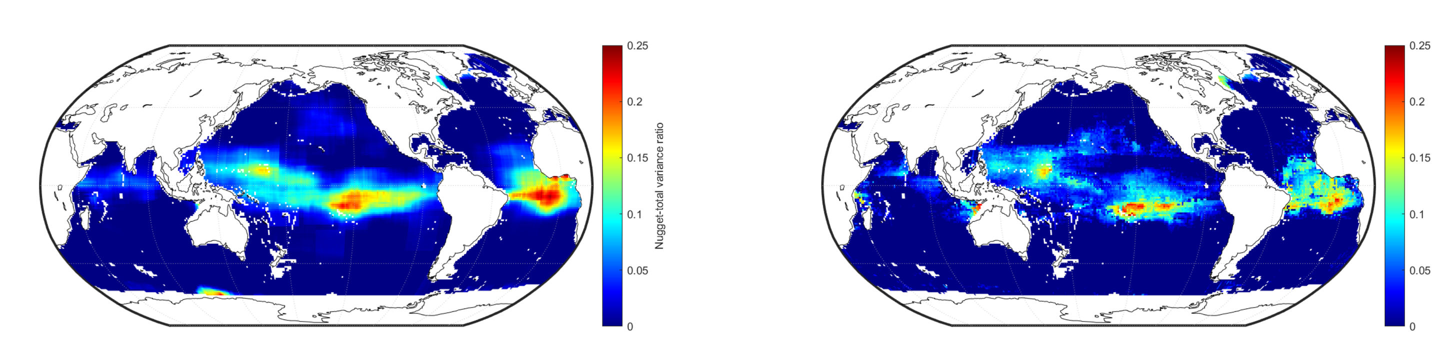


Figure: Exact (left) vs. neural (right) nugget-total variance ratio MLE (February)

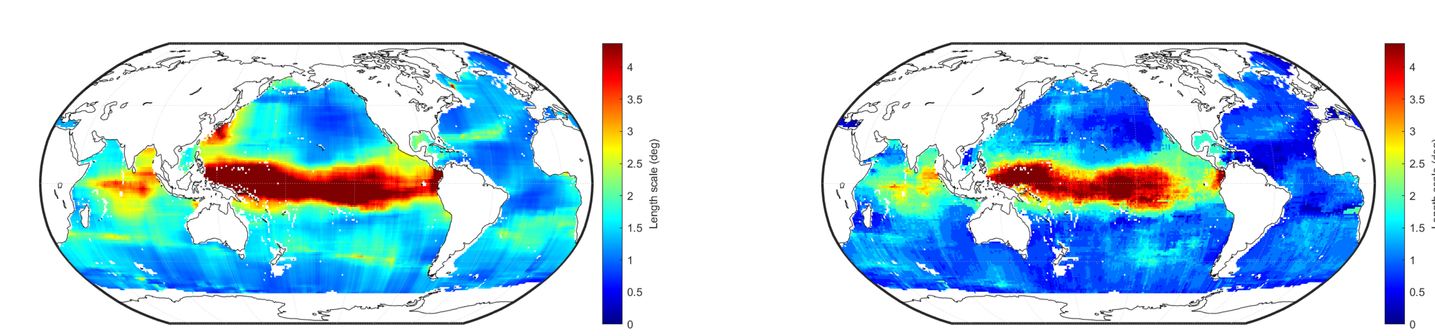


Figure: Exact (left) vs. neural (right) latitude length scale MLE (February)

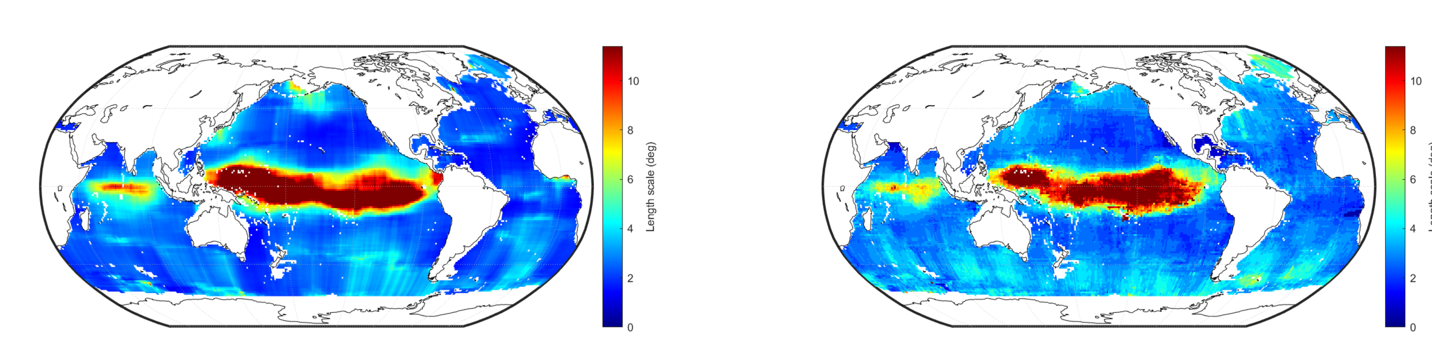


Figure: Exact (left) vs. neural (right) longitude length scale MLE (February)

*Note about the temporal length scale: The neural MLE currently has a tendency to choose values close to the boundary, especially in higher latitudes. We are actively exploring ways to improve the estimates, such as changing the nonlinearities in the GNN architecture or increasing the evaluation grid resolution.

Contact: thea@stat.cmu.edu