A reproducible UQ framework for ocean heat content via local conditional simulations

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

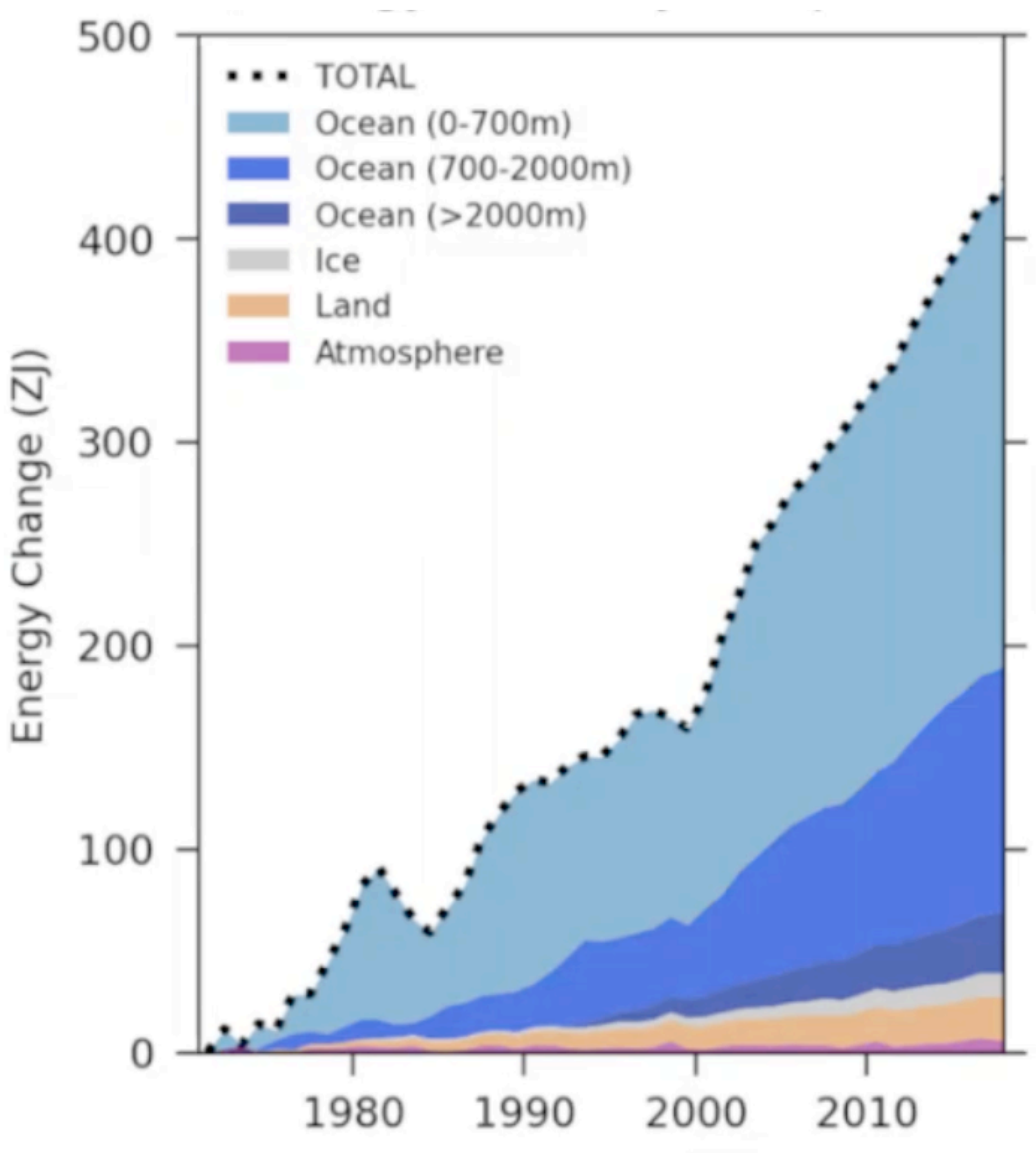
[1]Department of Statistics and Data Science, Carnegie Mellon University
[2]Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder

JSM 2024
4 August 2024

Carnegie Mellon University
Statistics & Data Science

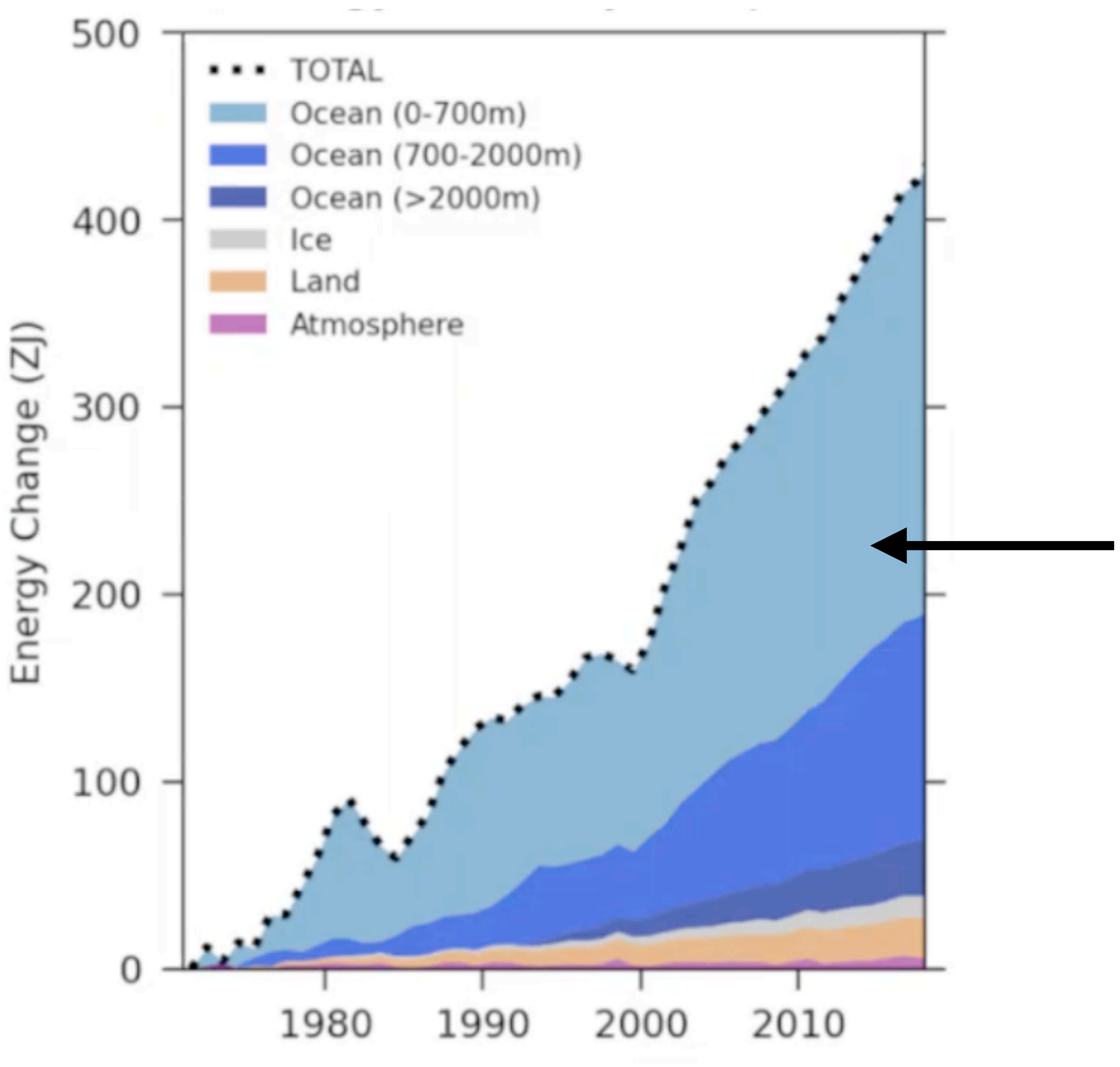
University of Colorado
Boulder

Changes in OHC contribute to extreme climate events



(IPCC AR6)

Changes in OHC contribute to extreme climate events



(IPCC AR6)

~91% of the excess heat in the climate system is stored **in the ocean**.

Changes in OHC contribute to extreme climate events

Is sea level rising?

Yes, sea level is **rising at an increasing rate**

With continued ocean and atmospheric warming, sea levels will likely rise for many centuries at rates highe
In the United States, almost 40 percent of the population lives in relatively high-population-density coastal
role in flooding, shoreline erosion, and hazards from storms. Globally, eight of the world's 10 largest cities c
U.N. Atlas of the Oceans.

(NOAA NOS, 2024)

Climate change is probably increasing the intensity of tropical cyclones

BY , MAYA V. CHUNG, GABE VECCHI, JINGRU SUN, , AND
PUBLISHED MARCH 31, 2021

HIGHLIGHTS

- Warming of the surface ocean from human-induced climate change is likely fueling more powerful tropical cyclones (TCs).
- The destructive power of individual TCs through flooding is amplified by rising sea level, which very likely has a substantial contribution at the global scale from anthropogenic climate change.
- TC precipitation rates are projected to increase due to enhanced atmospheric moisture associated with anthropogenic global warming.
- The proportion of Category 4 & 5 TCs has increased, possibly due to anthropogenic climate change, and is

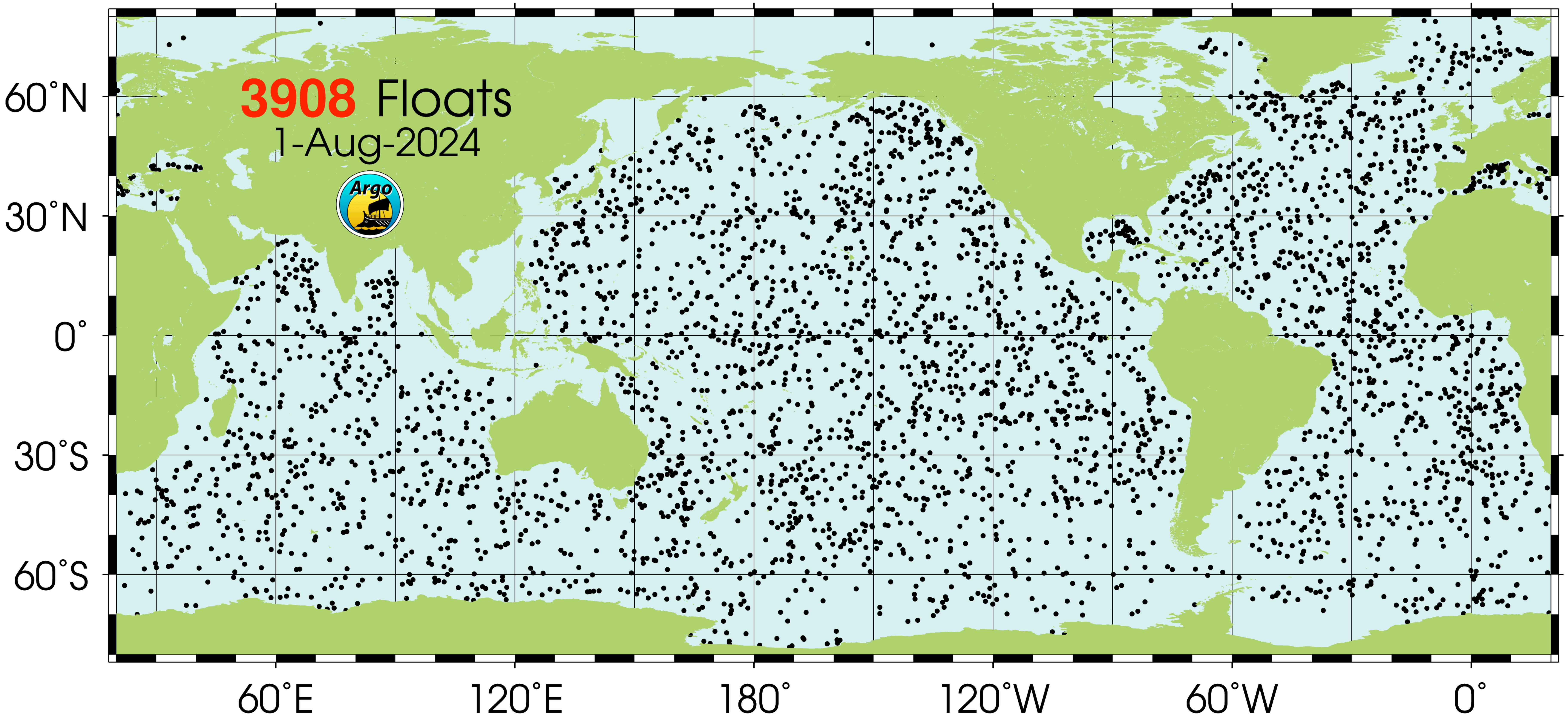
This post is adapted from a ScienceBrief by the same authors. The findings are consistent with the Intergovernmental Panel on Climate Change's Special Report on the Oceans and Cryosphere in a Changing Climate and with the reports (Part 1, Part 2) from the World Meteorological Organization's Task Team on Tropical Cyclones and Climate Change. The images that appear in the post were not part of the original brief.

This ScienceBrief presents a summary of the state of the science on tropical cyclones (tropical storms, hurricanes, and typhoons) and climate change. The authors assessed more than 90 peer-reviewed scientific articles, with a focus on articles describing observations of, or projected future changes to, the frequency and intensity of tropical cyclones (TCs) globally or in key regions, as well as changes in tropical cyclone-related rainfall and storm surge.

Warming of the surface ocean from anthropogenic (human-induced) climate change is likely fueling more powerful TCs. The destructive power of individual TCs through flooding is amplified by rising sea level, which very likely has a substantial contribution at the global scale from anthropogenic climate change. In addition, TC precipitation rates are projected to increase due to enhanced atmospheric moisture associated with anthropogenic global warming.

(NOAA Climate.gov, 2021)

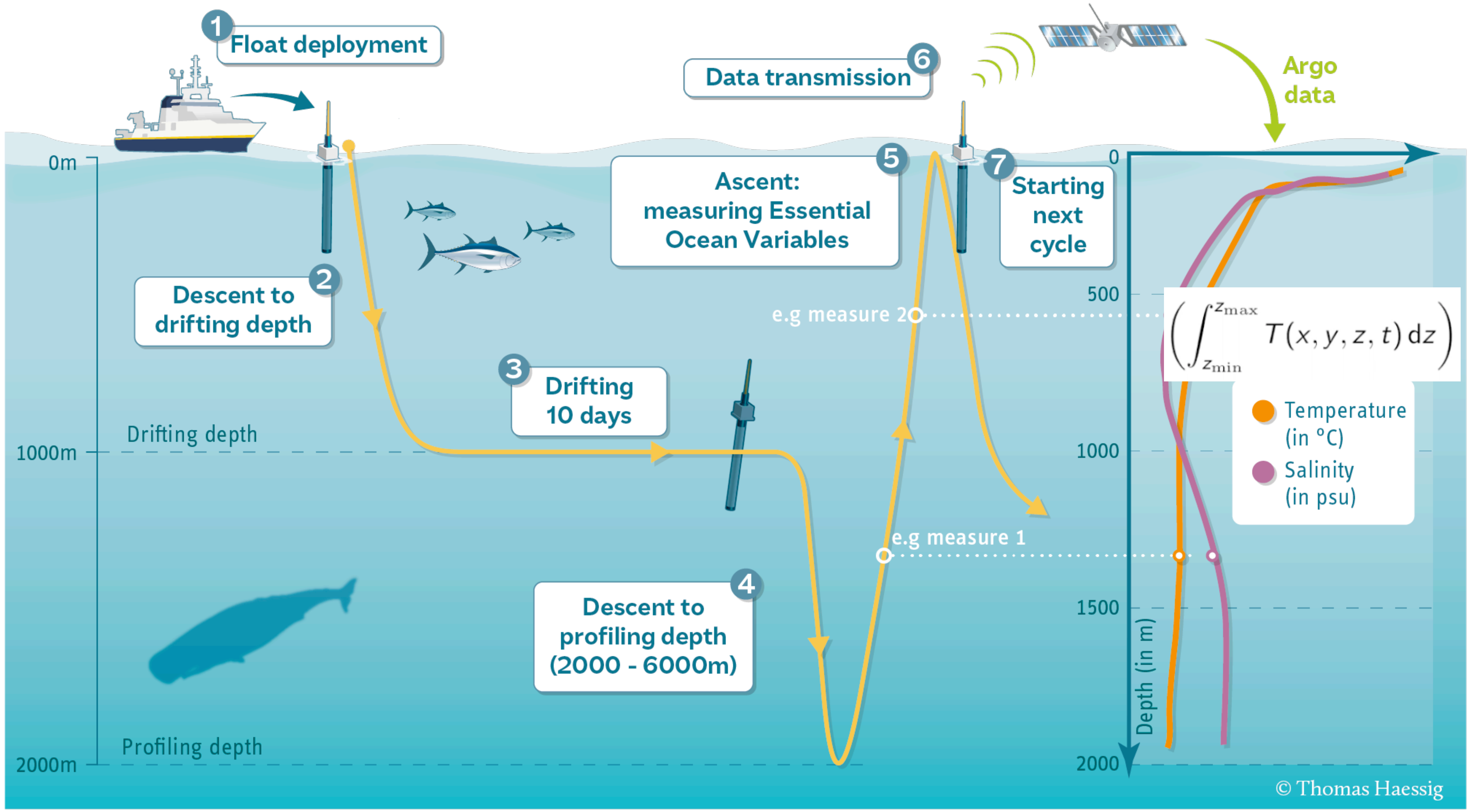
We use Argo data (2004-2022) to estimate 15-1850 dbar OHC



(Argo Program)

We use Argo data (2004-2022) to estimate 15-1850 dbar OHC

$$\mathrm{OHC}(t) = \rho_0 c_{p,0} \int\!\!\int \left(\int_{z_{\min}}^{z_{\max}} T(x,y,z,t)\,\mathrm{d}z \right) \mathrm{d}x\,\mathrm{d}y$$

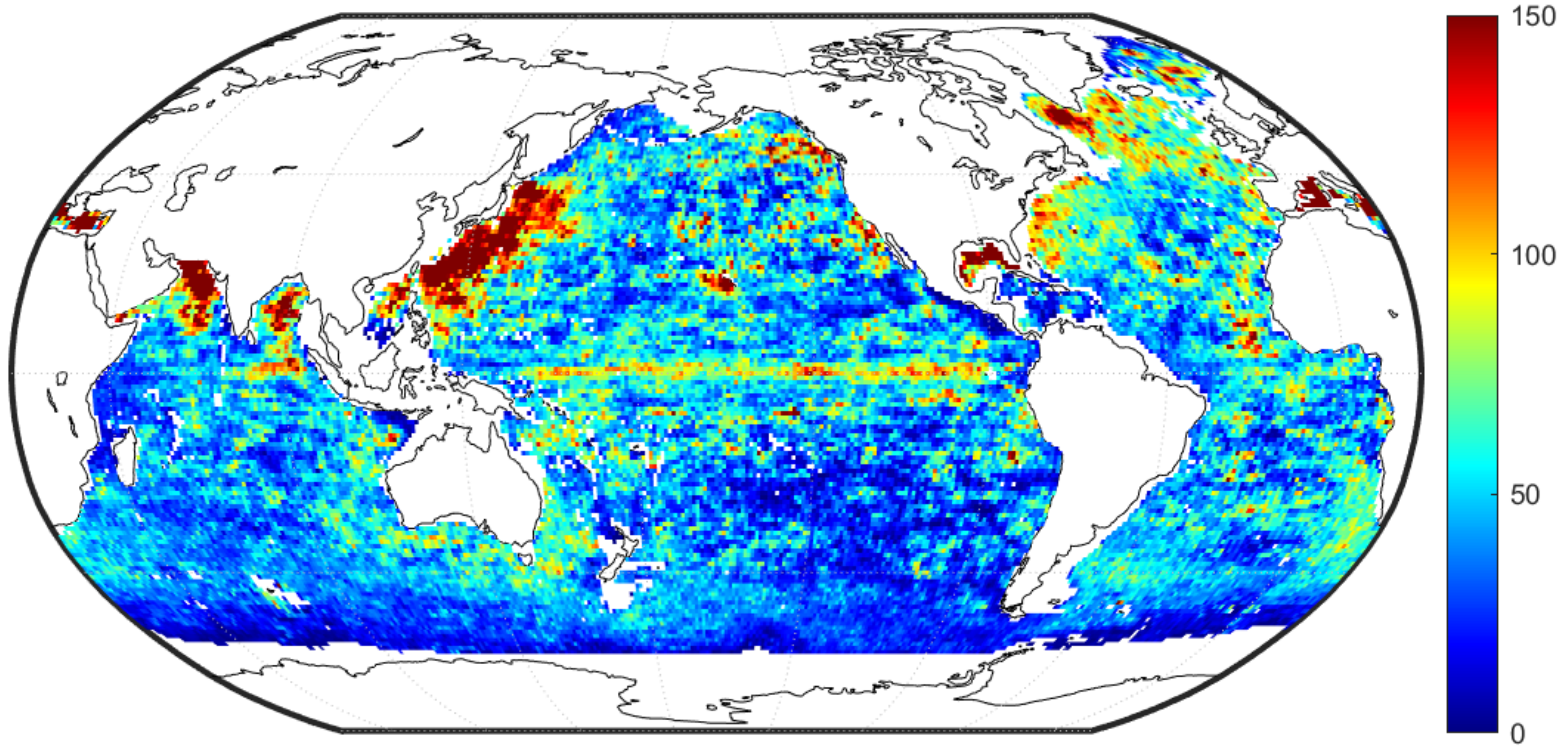


(Argo Program)

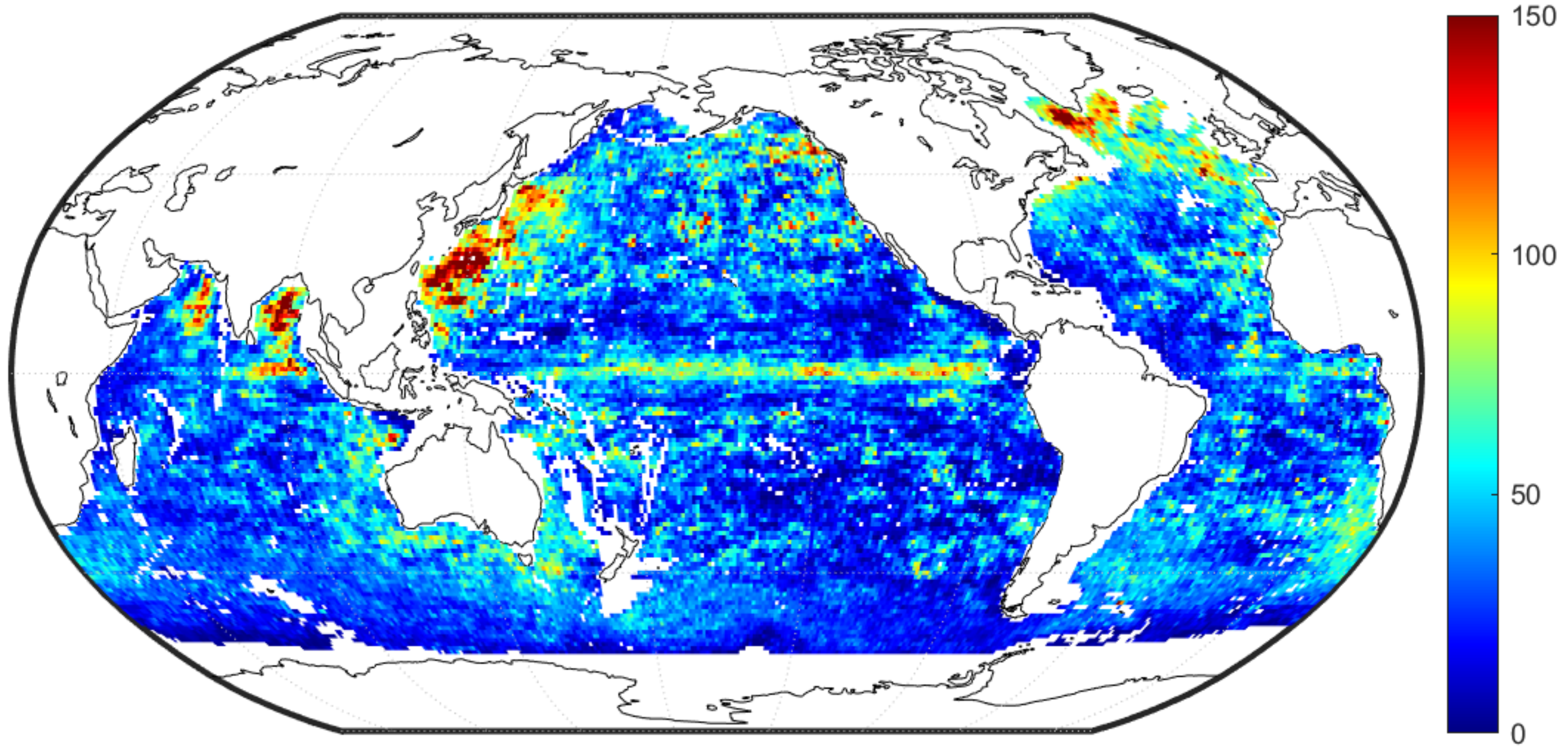
Modeling OHC from sparse observations: a 4-step framework

1. Integrate the **vertical dimension**
2. Model **horizontal and temporal dimension**
3. Estimate **uncertainties** based on (1) and (2)
4. Integrate the **horizontal dimension**

We use Argo data (2004-2022) to estimate 15-1850 dbar OHC



→ **Top layer profiles (15-975 dbar)**



→ **Bottom layer profiles (975-1850 dbar)**

Modeling OHC from sparse observations: a 4-step framework

1. Integrate the **vertical dimension**

2. Model **horizontal and temporal dimension**

3. Estimate **uncertainties** based on (1) and (2)

4. Integrate the **horizontal dimension**

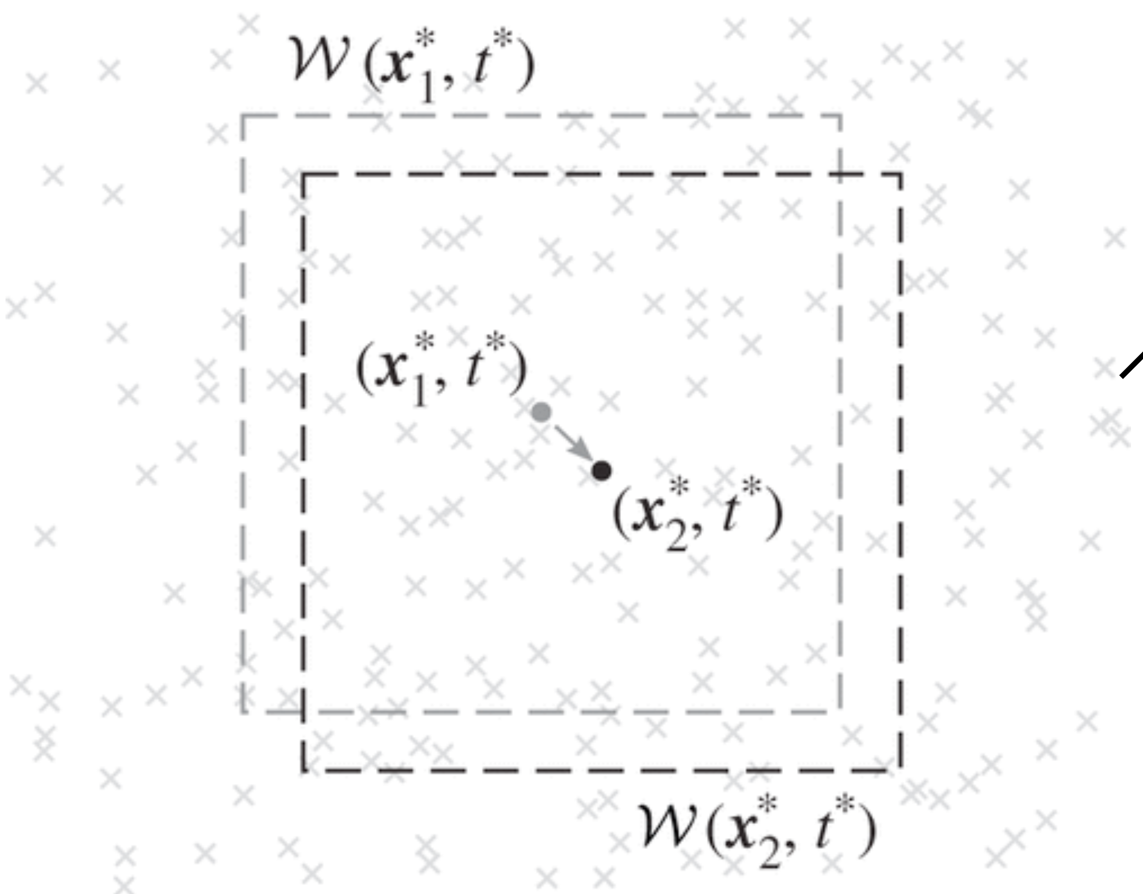
Modeling challenges arise from the data size and nonstationarity

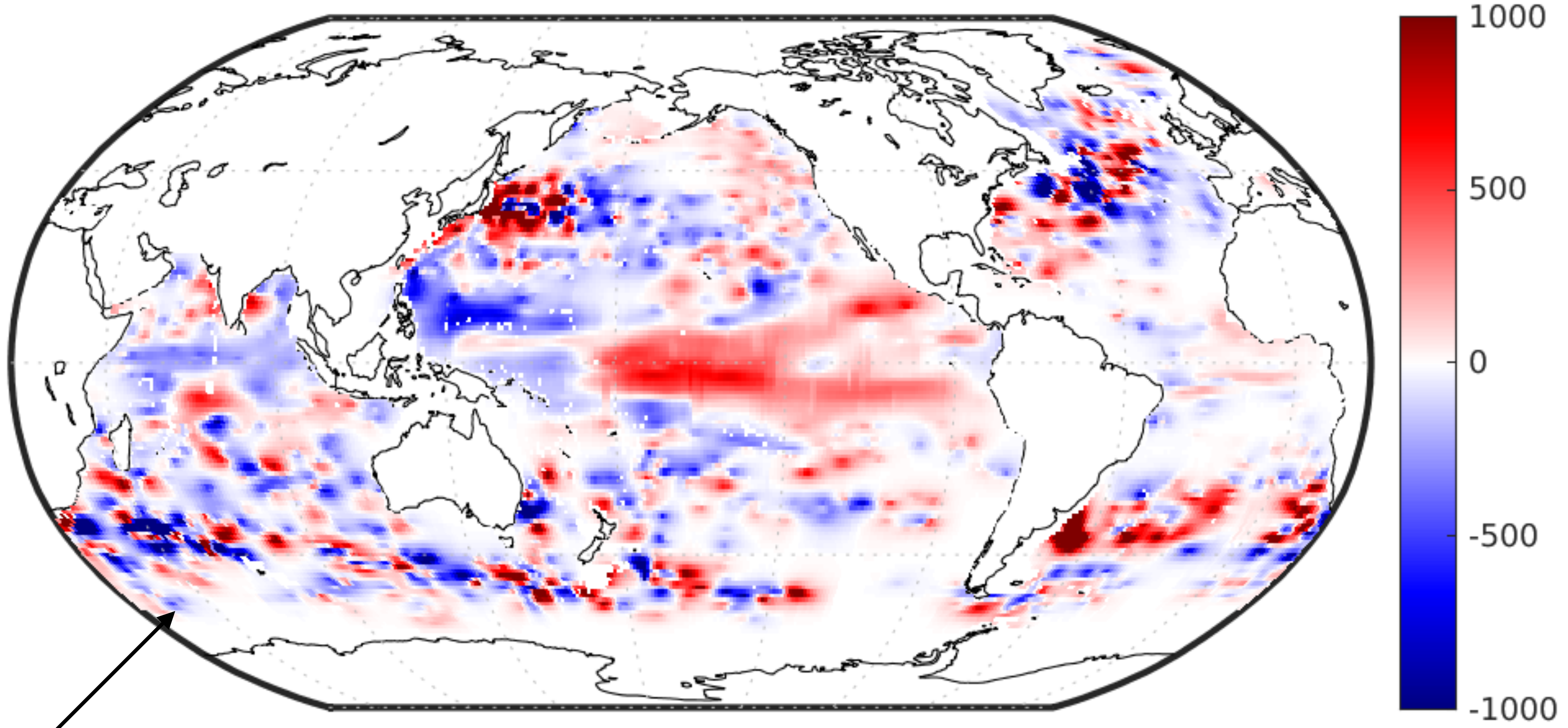
- **Size:** $>$ 2.5 million Argo profiles (matrix inversion for covariance parameter estimation and kriging is infeasible)
- **Nonstationarity:** Challenging to define a nonstationary covariance function flexible enough to explain variability across entire global ocean

We can estimate a gridded temperature field with GP regression

1. Model temperature mean field: local least squares regression (Roemmich and Gilson 2009) + linear time trend

2. Model residuals: **locally stationary** Gaussian process (GP) regression (Kuusela and Stein 2018)





Modeled top layer residuals (02/2005)

We can estimate a gridded temperature field with GP regression

$$\begin{bmatrix} y_{\text{top}} \\ y_{\text{bot}} \end{bmatrix}_{i,j} = f_i(\begin{bmatrix} x_{\text{top}} \\ x_{\text{bot}} \end{bmatrix}_{i,j}, \begin{bmatrix} t_{\text{top}} \\ t_{\text{bot}} \end{bmatrix}_{i,j}) + \begin{bmatrix} \epsilon_{\text{top}} \\ \epsilon_{\text{bot}} \end{bmatrix}_{i,j}$$

Temperature residuals	**Latitude Longitude**	**Date**	**Nugget effect**

We can estimate a gridded temperature field with GP regression

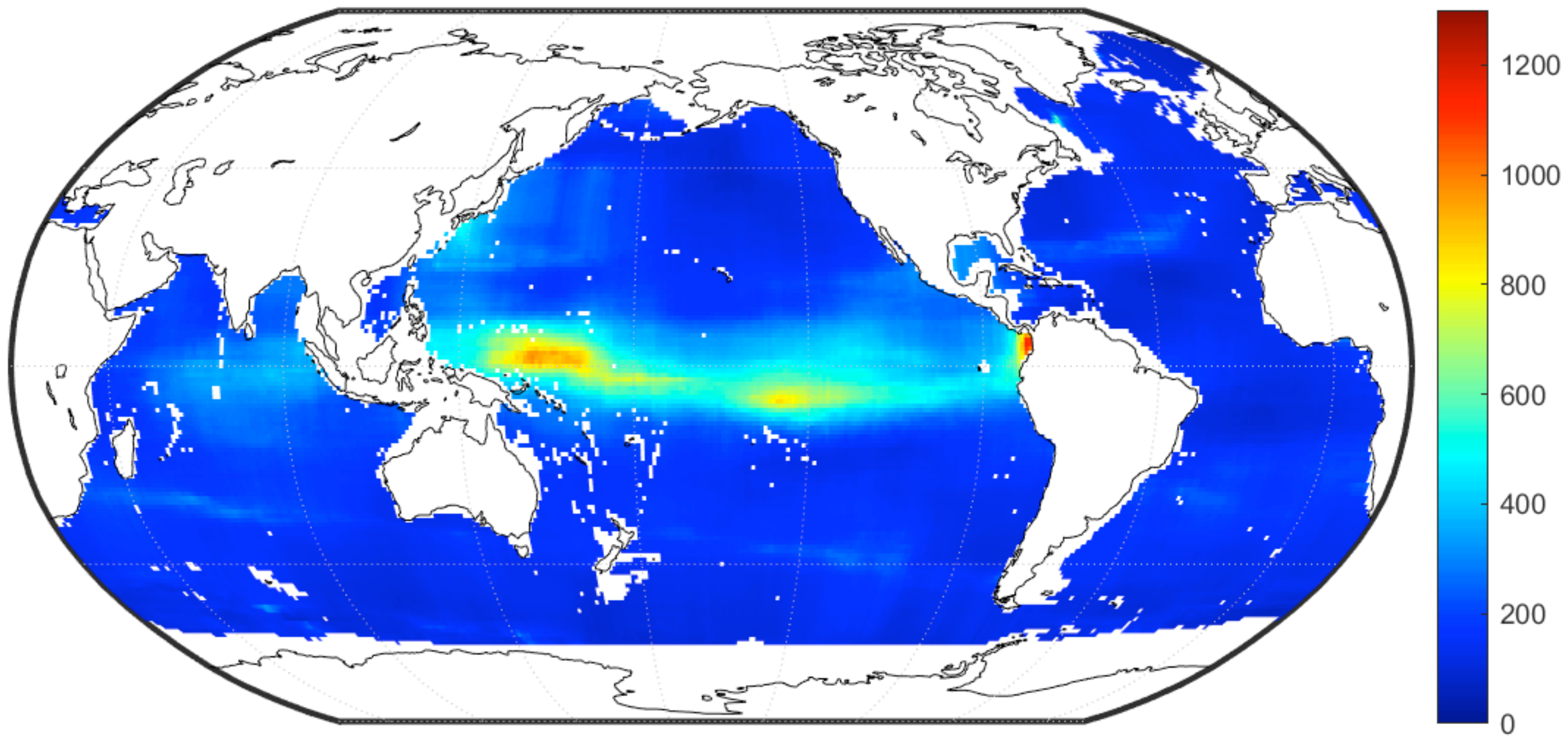
$$\begin{bmatrix} y_{\text{top}} \\ y_{\text{bot}} \end{bmatrix}_{i,j} = f_i\left(\begin{bmatrix} x_{\text{top}} \\ x_{\text{bot}} \end{bmatrix}_{i,j}, \begin{bmatrix} t_{\text{top}} \\ t_{\text{bot}} \end{bmatrix}_{i,j}\right) + \begin{bmatrix} \epsilon_{\text{top}} \\ \epsilon_{\text{bot}} \end{bmatrix}_{i,j}$$

Temperature residuals | **Latitude Longitude** | **Date** | **Nugget effect**

$$f_i \overset{\text{iid}}{\sim} \mathrm{GP}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{K}(x_1, t_1, x_2, t_2; \boldsymbol{\theta})\right)$$

$$\begin{bmatrix} \epsilon_{\text{top}} \\ \epsilon_{\text{bot}} \end{bmatrix} \overset{\text{iid}}{\sim} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_{\boldsymbol{\epsilon}})\right)$$

$$\boldsymbol{K}_{\text{ii}}(\boldsymbol{z_1}, \boldsymbol{z_2}; \boldsymbol{\theta}) = \frac{\delta_i^2}{\sqrt{|\boldsymbol{\Theta}_i|}} \exp\left(-\sqrt{(\boldsymbol{z_1} - \boldsymbol{z_2})^{\mathrm{T}} \boldsymbol{\Theta}_i^{-1} (\boldsymbol{z_1} - \boldsymbol{z_2})}\right)$$



Length scale MLE (latitude; top layer)

Modeling OHC from sparse observations: a 4-step framework

1. Integrate the vertical dimension

2. Model **horizontal and temporal dimension**

3. Estimate **uncertainties** based on (1) and (2)

4. Integrate the **horizontal dimension**

Obtaining uncertainties is facilitated by local conditional simulations

full field | data ?

Obtaining uncertainties is facilitated by local conditional simulations

full field | data - **multivariate normal** with conditional covariance Σ_i

(parameterized by estimated GP variance, length scales, cross-correlation)

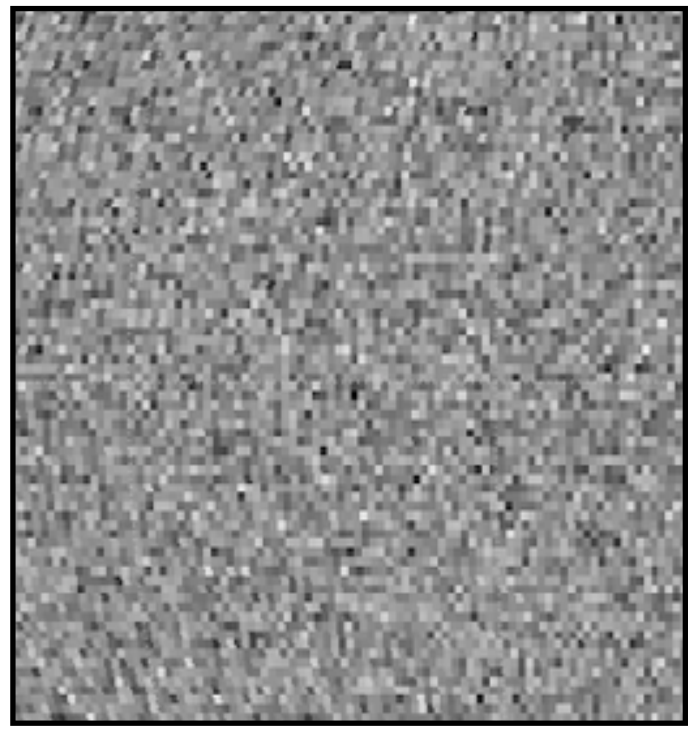
Obtaining uncertainties is facilitated by local conditional simulations

full field | data - **multivariate normal** with conditional covariance Σ_i

(parameterized by estimated GP variance, length scales, cross-correlation)

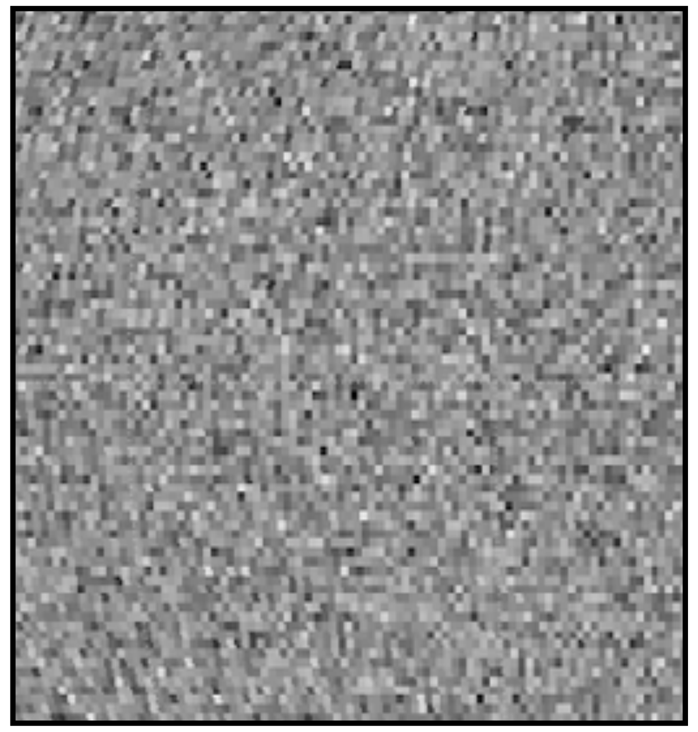
Local conditional simulations! (extension of Nychka et.al. 2018)

Obtaining uncertainties is facilitated by local conditional simulations



Simulate Gaussian
white noise over grid
(keep fixed)

Obtaining uncertainties is facilitated by local conditional simulations

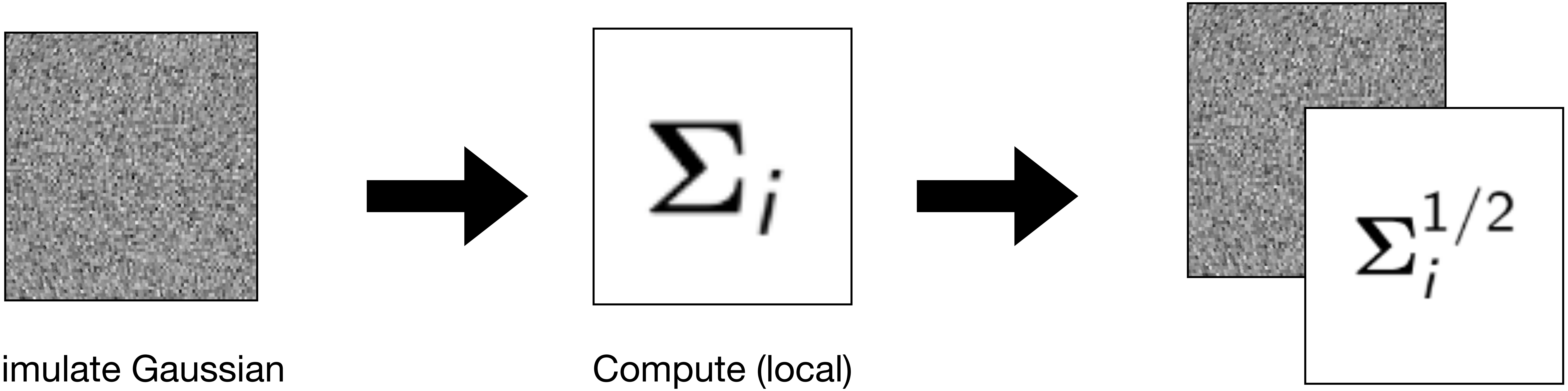


Simulate Gaussian
white noise over grid
(keep fixed)

$$\Sigma_i$$

Compute (local)
conditional covariance
matrix

Obtaining uncertainties is facilitated by local conditional simulations

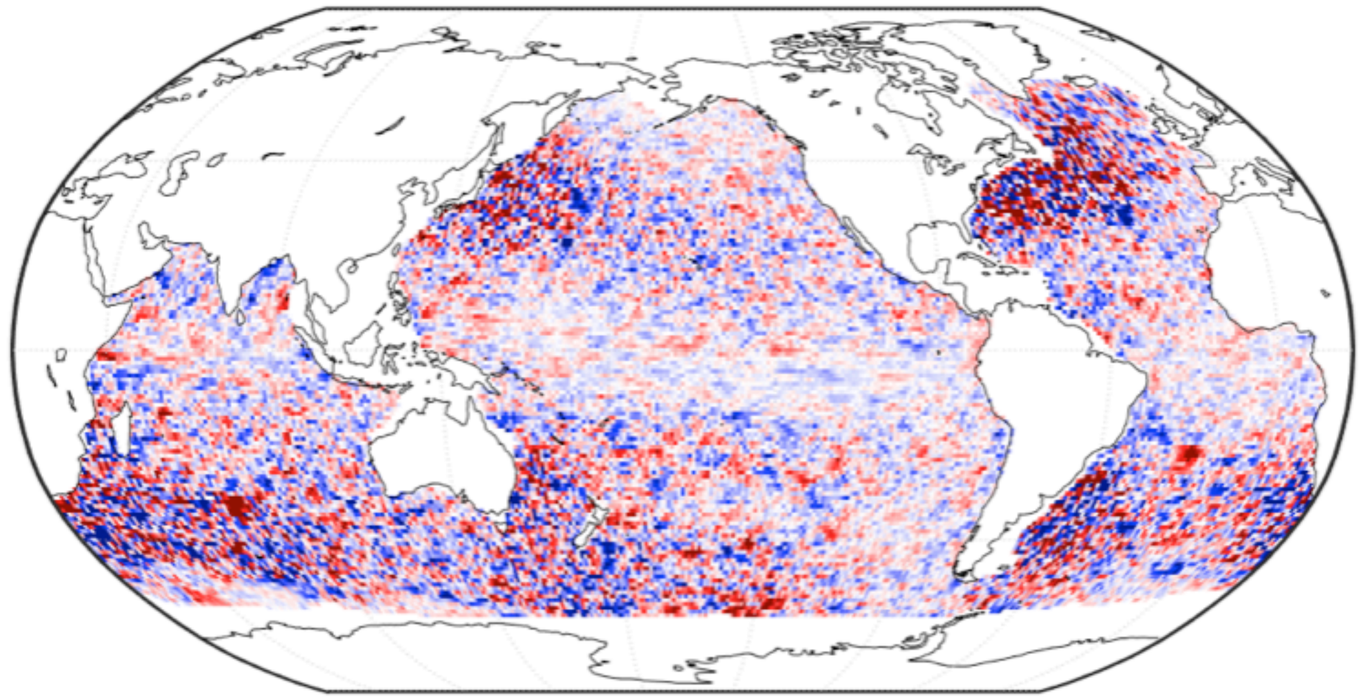


Simulate Gaussian white noise over grid (keep fixed)

Compute (local) conditional covariance matrix

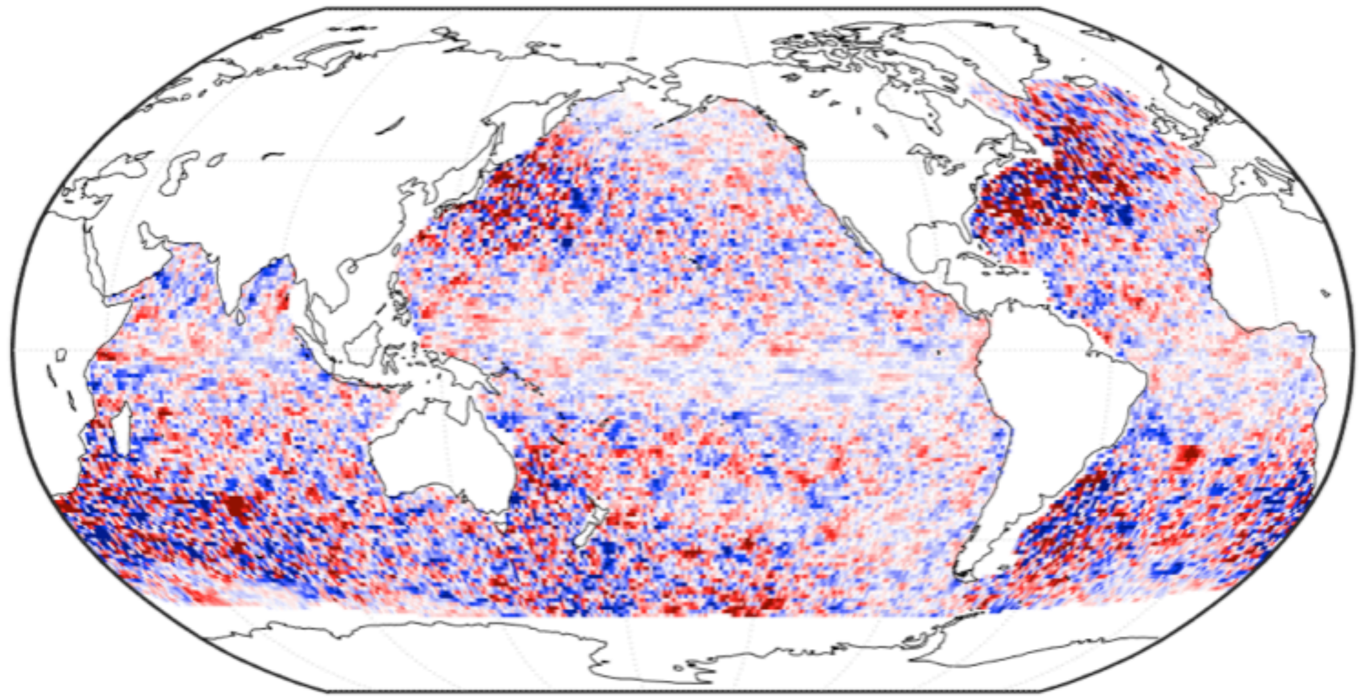
Take symmetric square root and multiply by white noise

Obtaining uncertainties is facilitated by local conditional simulations

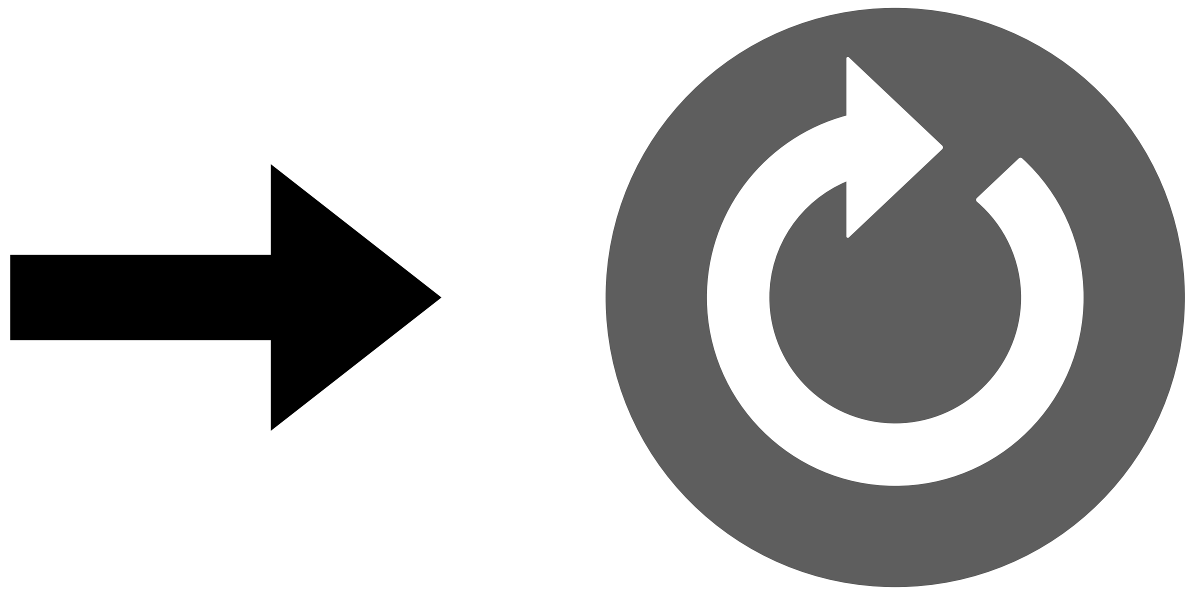


Keep the center point and repeat for all grid points

Obtaining uncertainties is facilitated by local conditional simulations

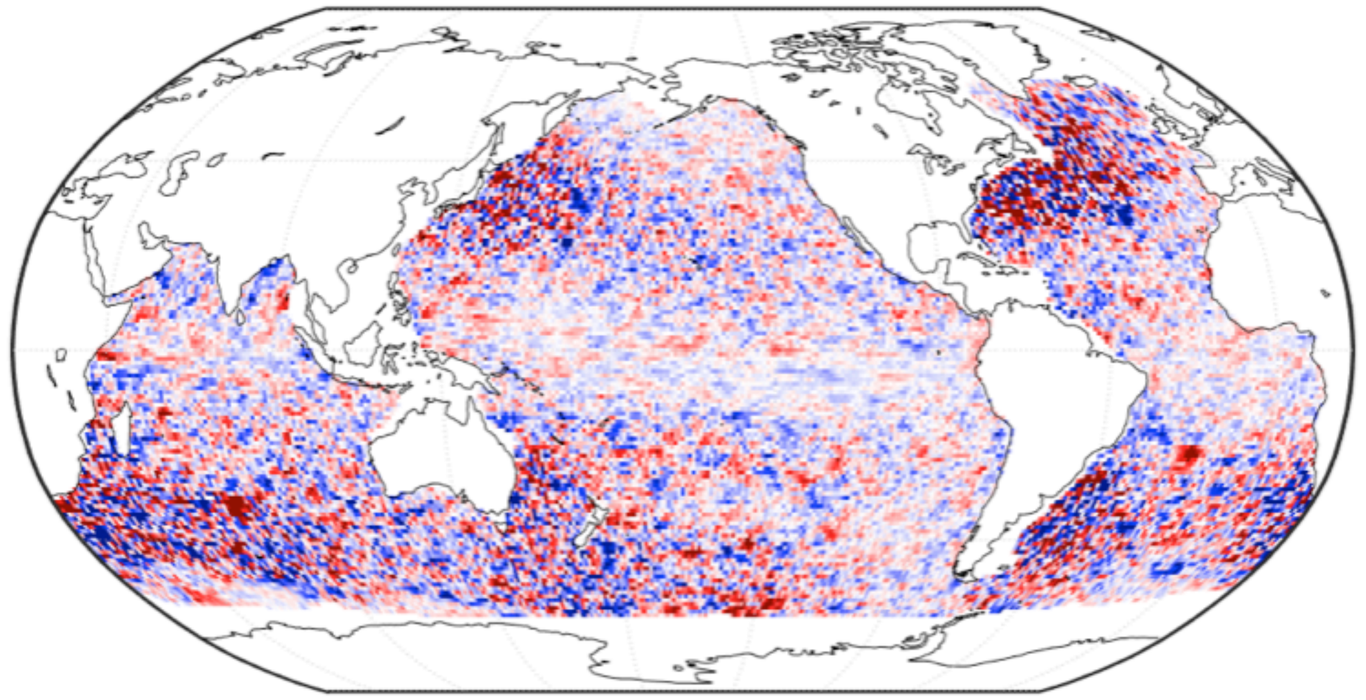


Keep the center point and repeat for all grid points



Repeat for desired number of ensemble members and other layer

Obtaining uncertainties is facilitated by local conditional simulations



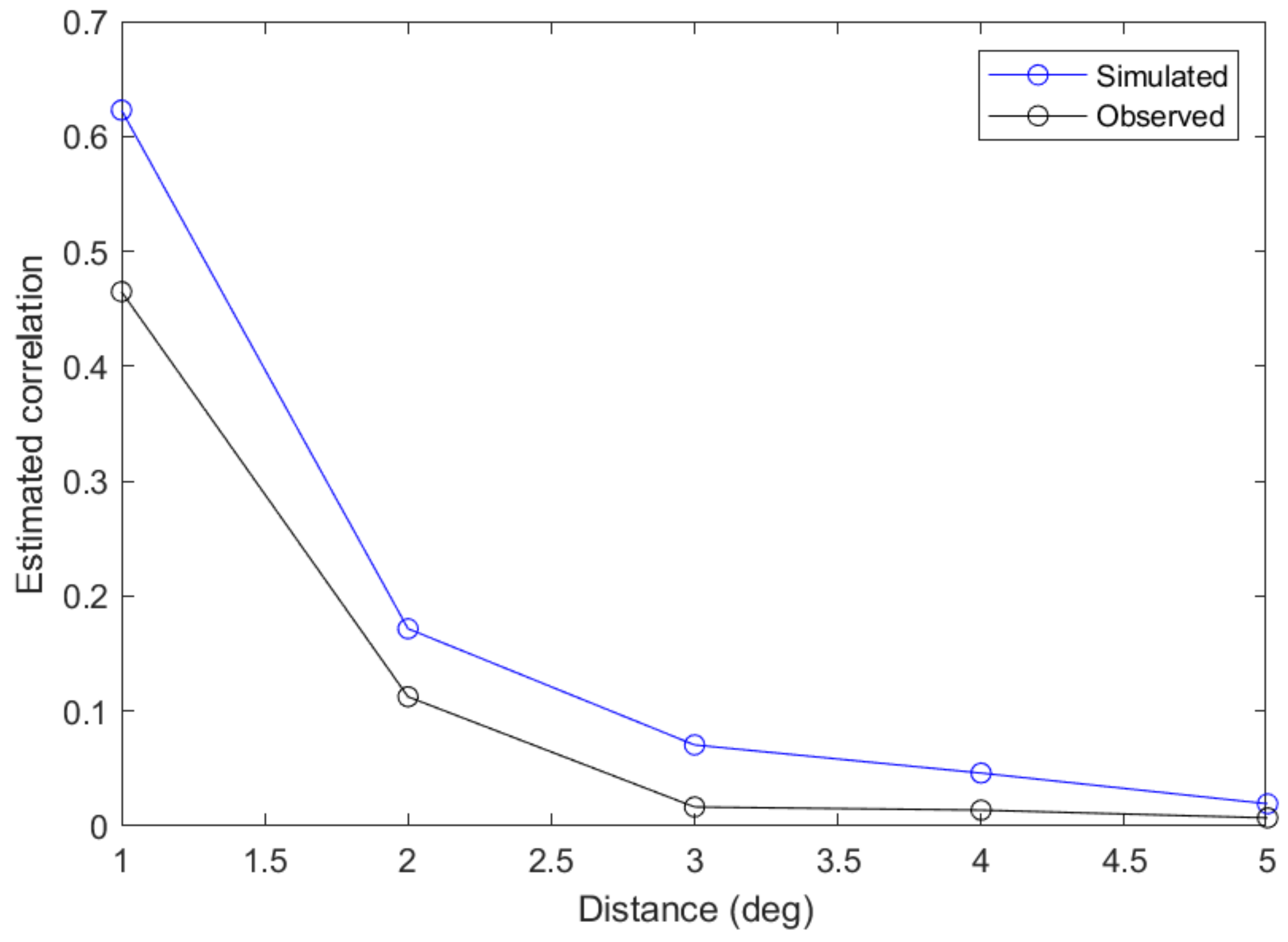
Keep the center point and repeat for all grid points

Repeat for desired number of ensemble members and other layer

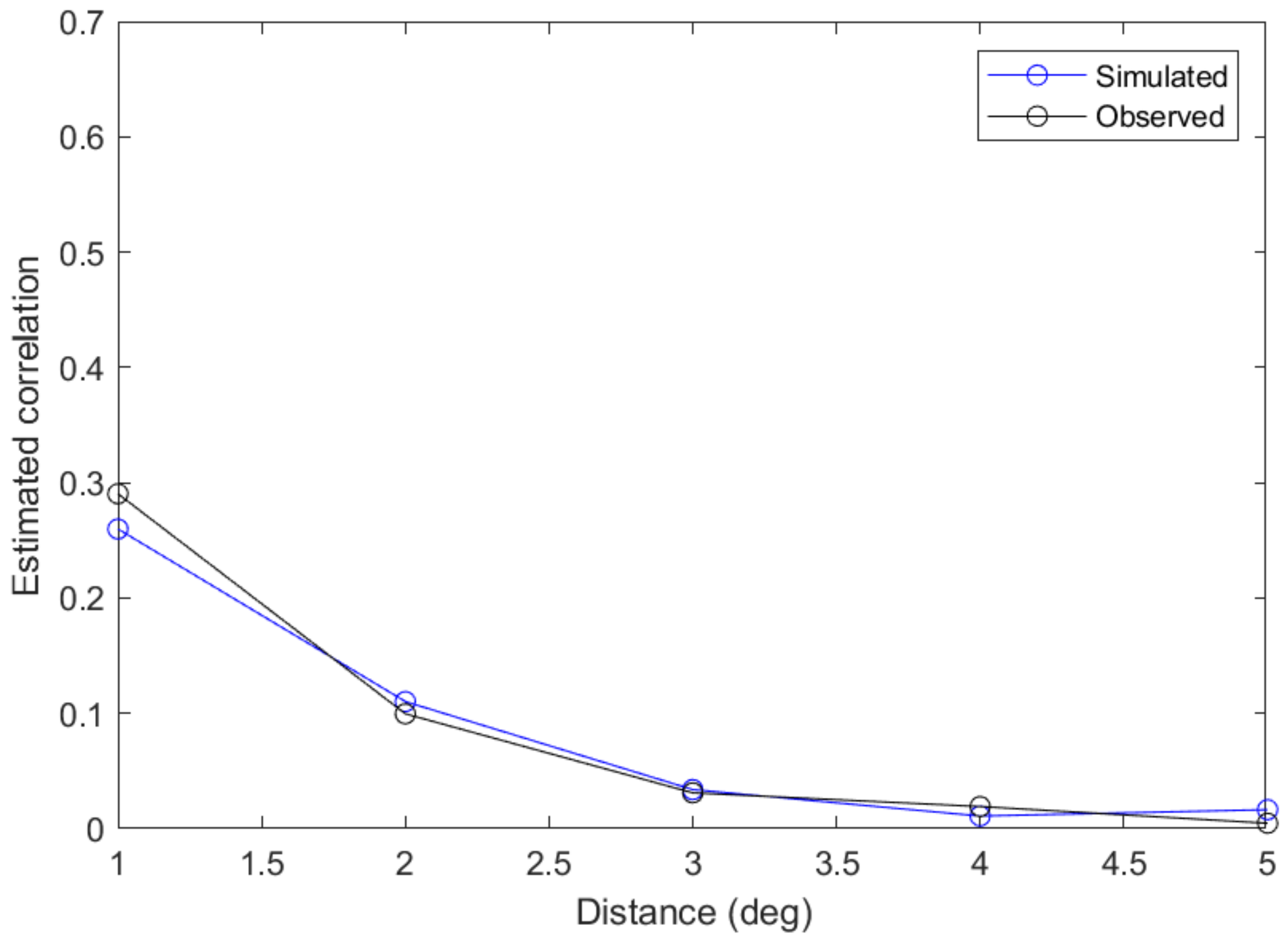
Compute conservative estimate of OHC total conditional variance using sample variances

$$\left(\sqrt{\mathrm{Var}(\mathrm{OHC}_{\mathrm{top}}|\mathrm{data})} + \sqrt{\mathrm{Var}(\mathrm{OHC}_{\mathrm{bot}}|\mathrm{data})}\right)^2$$

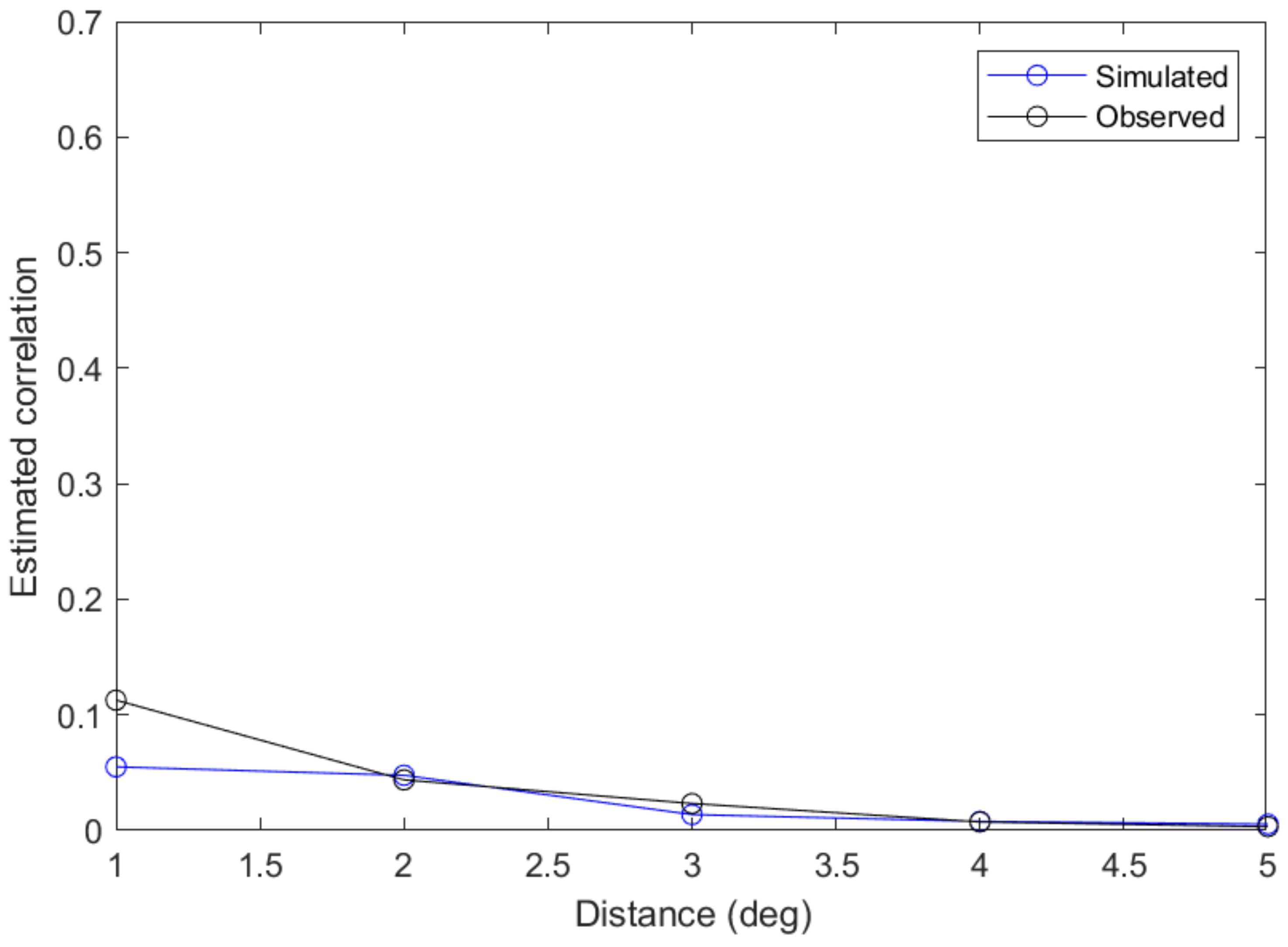
CV shows simulations capture conditional spatial dependence reasonably well



0~30 days



30~60 days



60~90 days

Modeling OHC from sparse observations: a 4-step framework

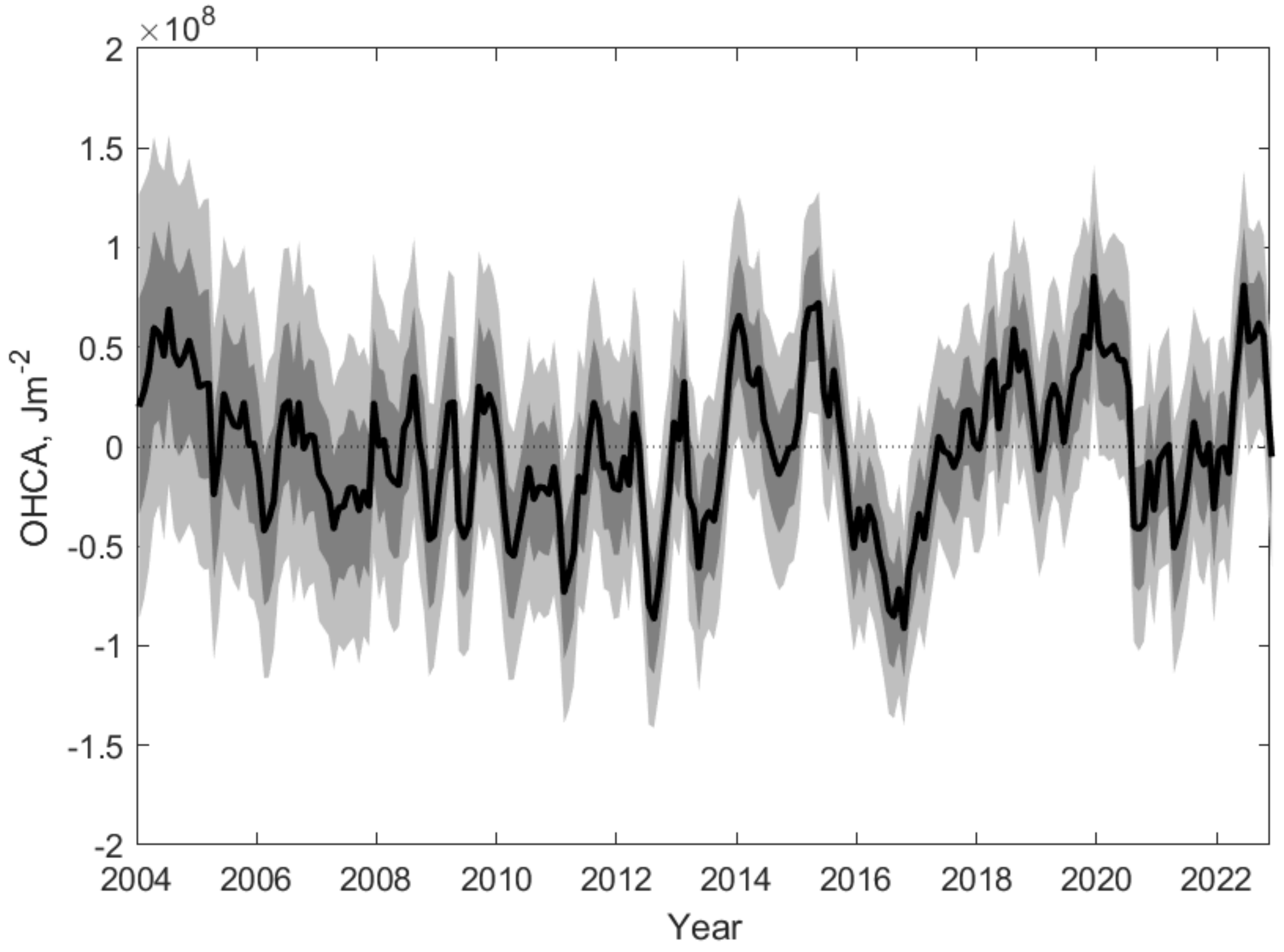
1. Integrate the vertical dimension

2. Model **horizontal and temporal dimension**

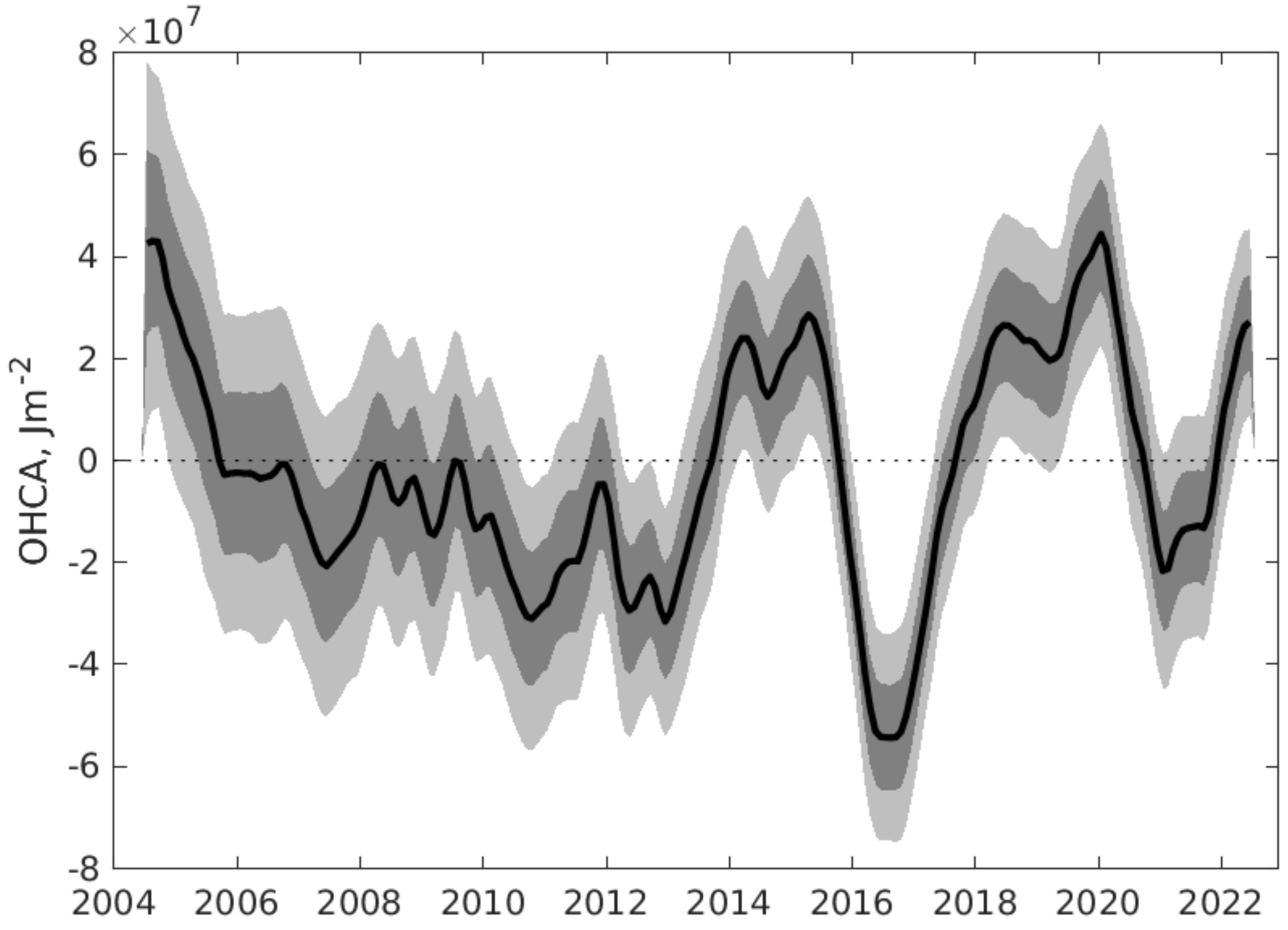
3. Estimate **uncertainties** based on (1) and (2)

4. Integrate the **horizontal dimension**

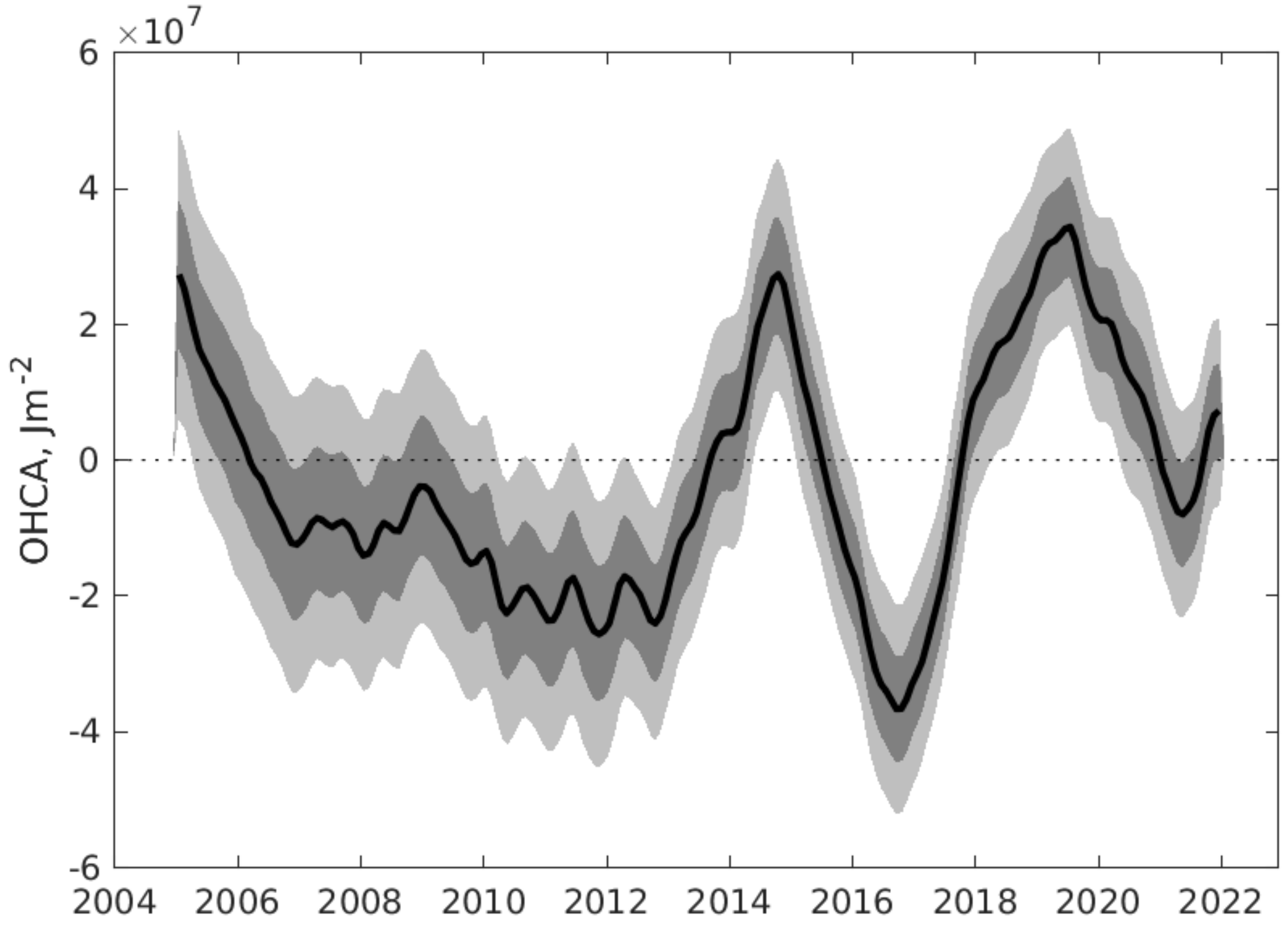
Uncertainties: Global OHC anomalies



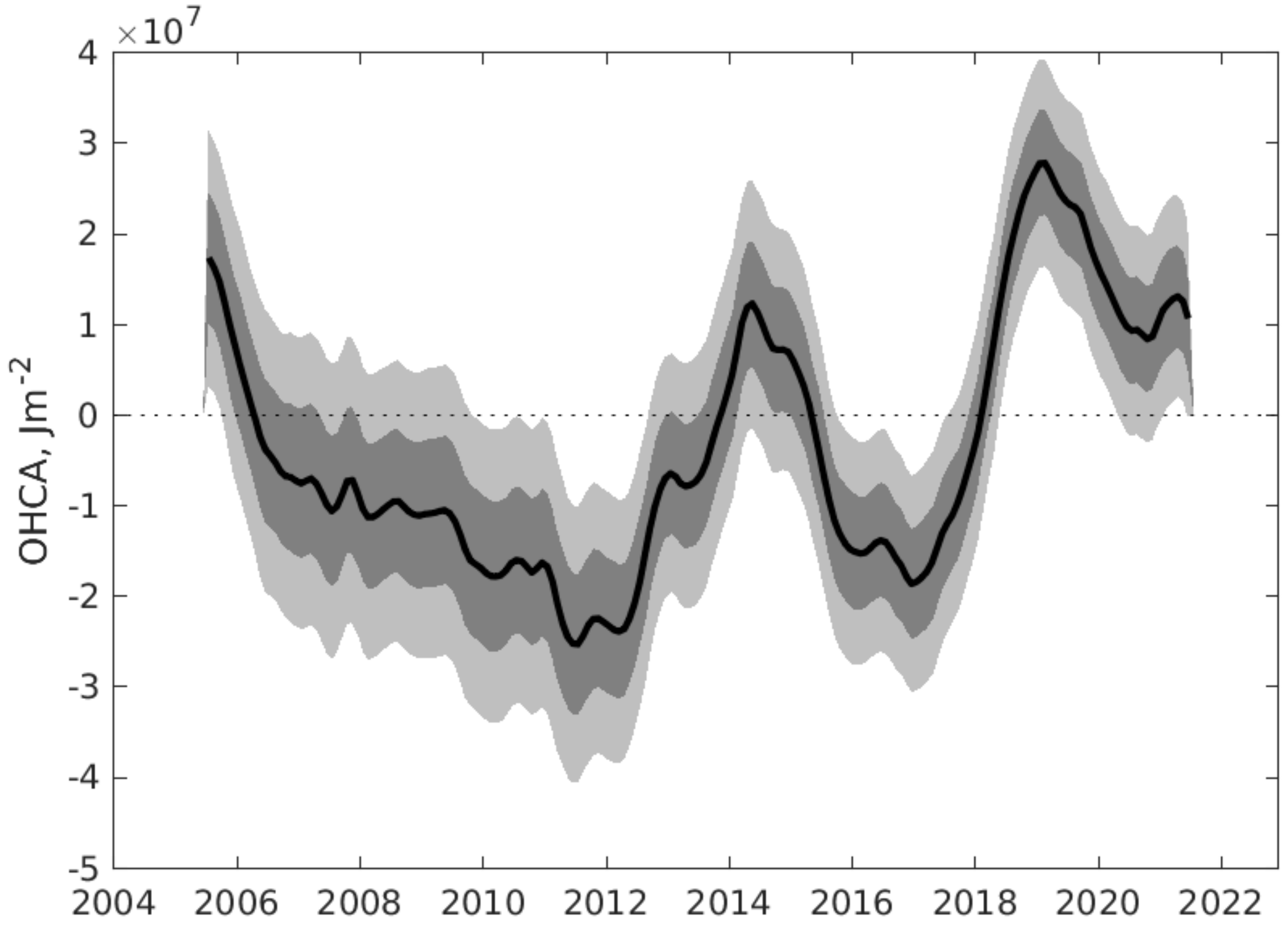
Monthly



12-month moving average

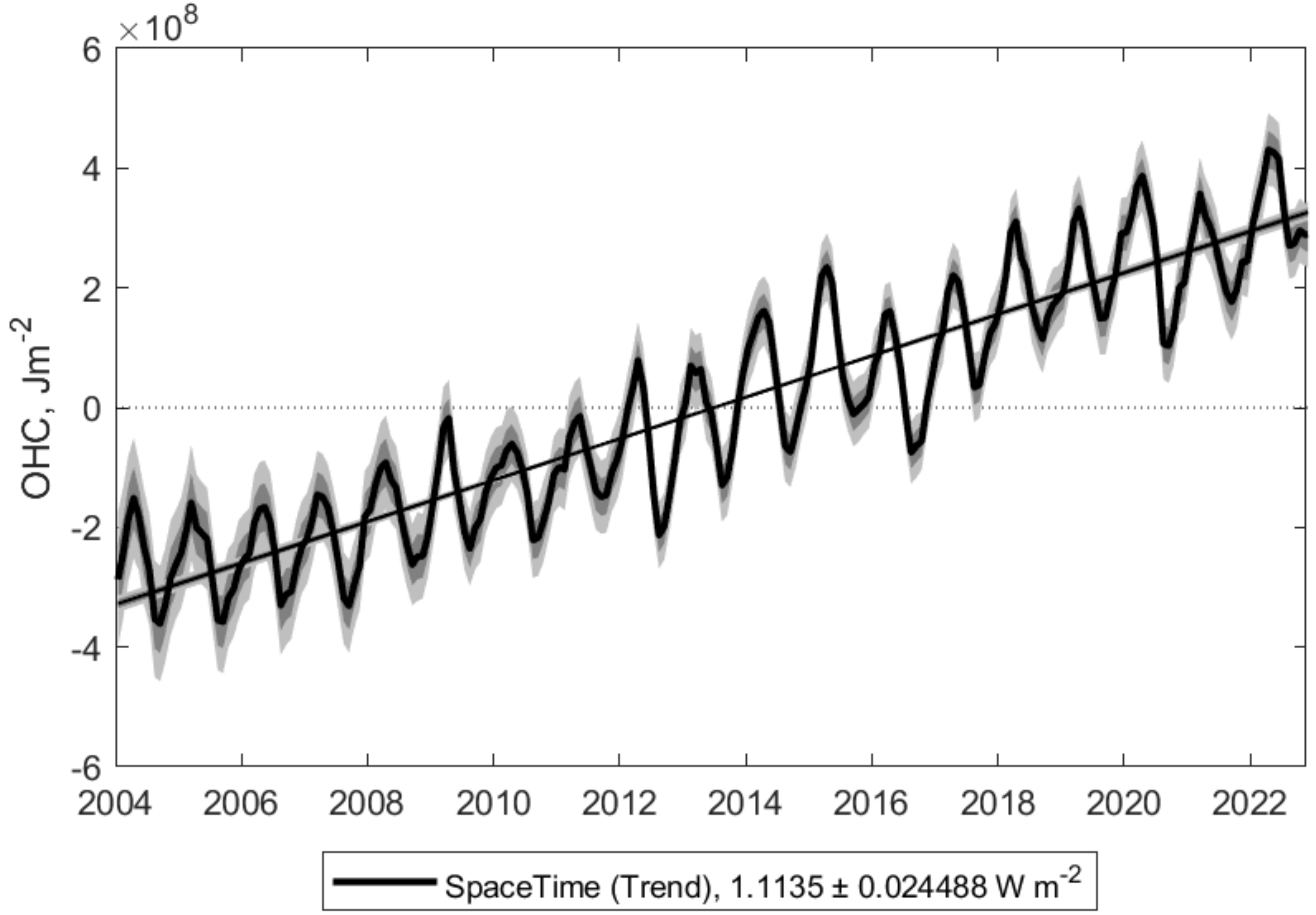


24-month moving average

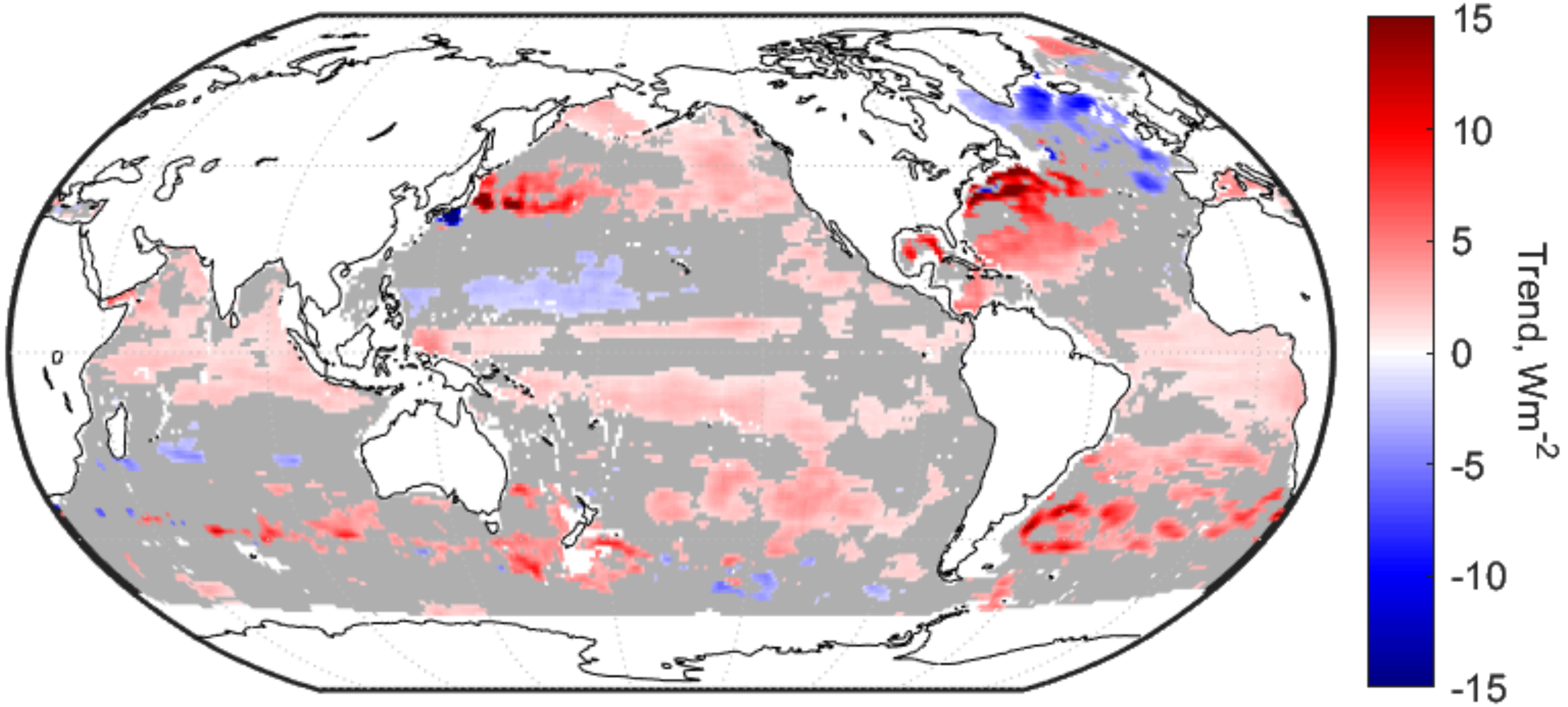


36-month moving average

Uncertainties: Global OHC trend

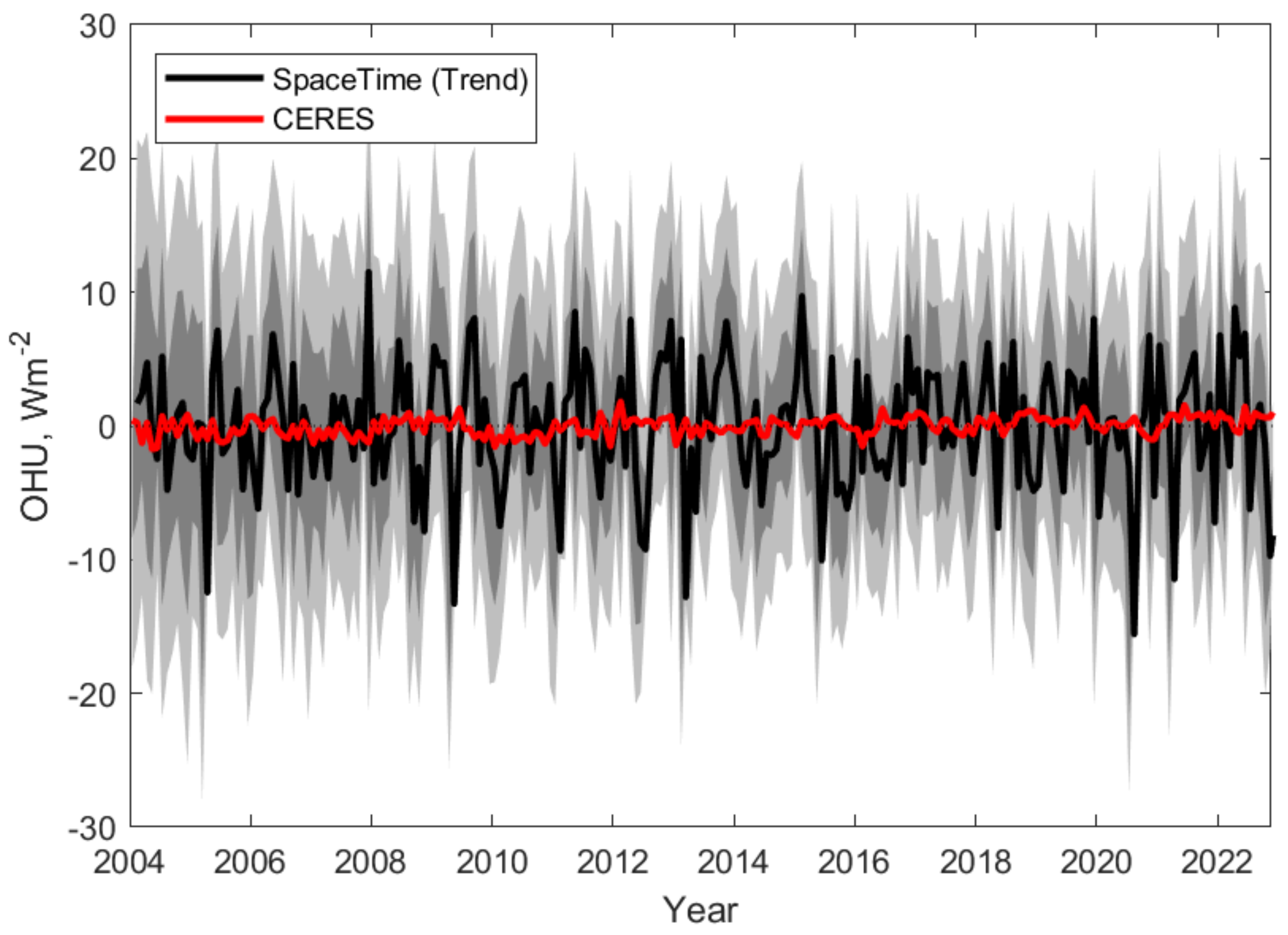


Uncertainties: Regional OHC trends

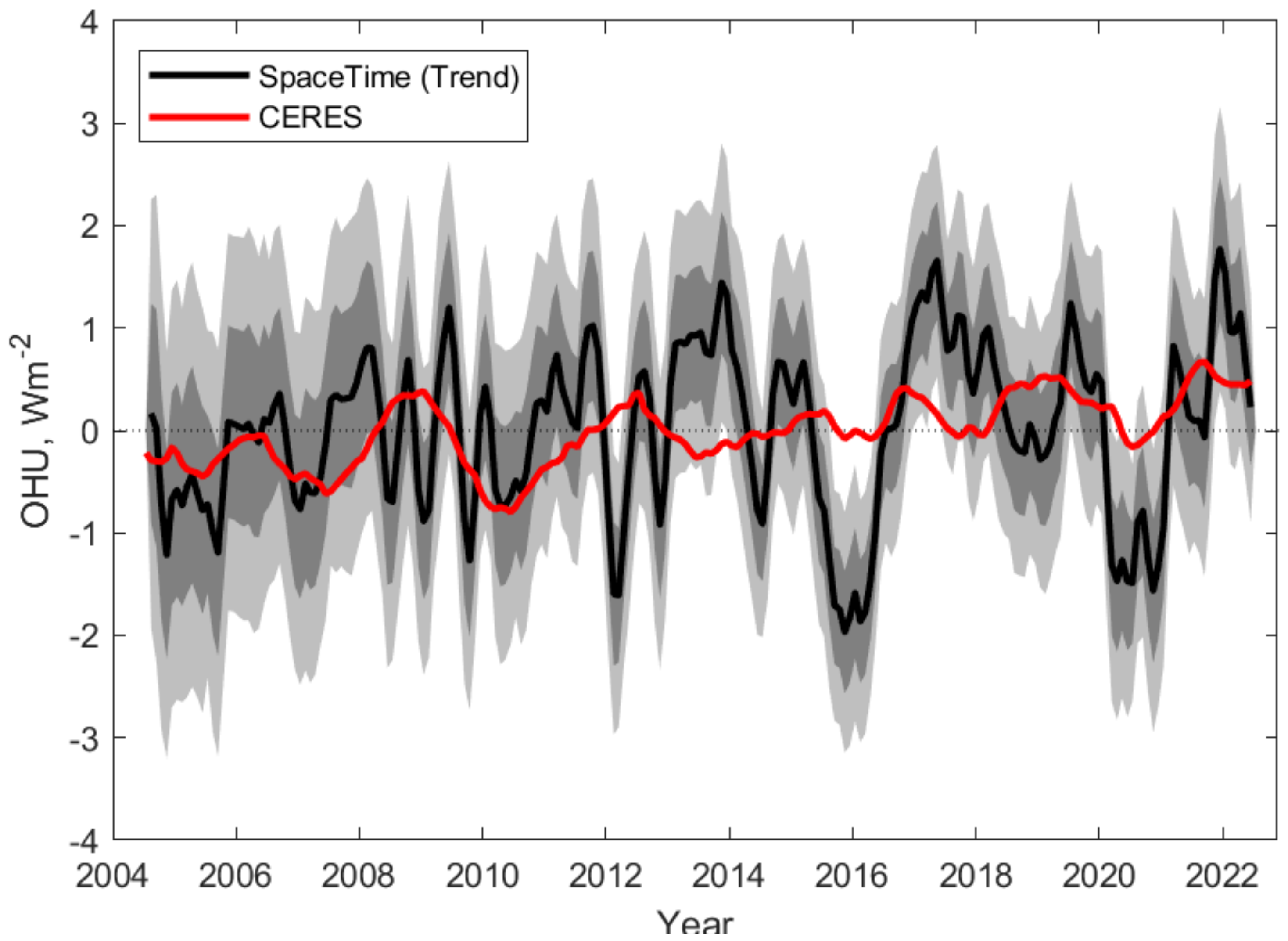


Significant OHC trends at 5% level (2004-2022)

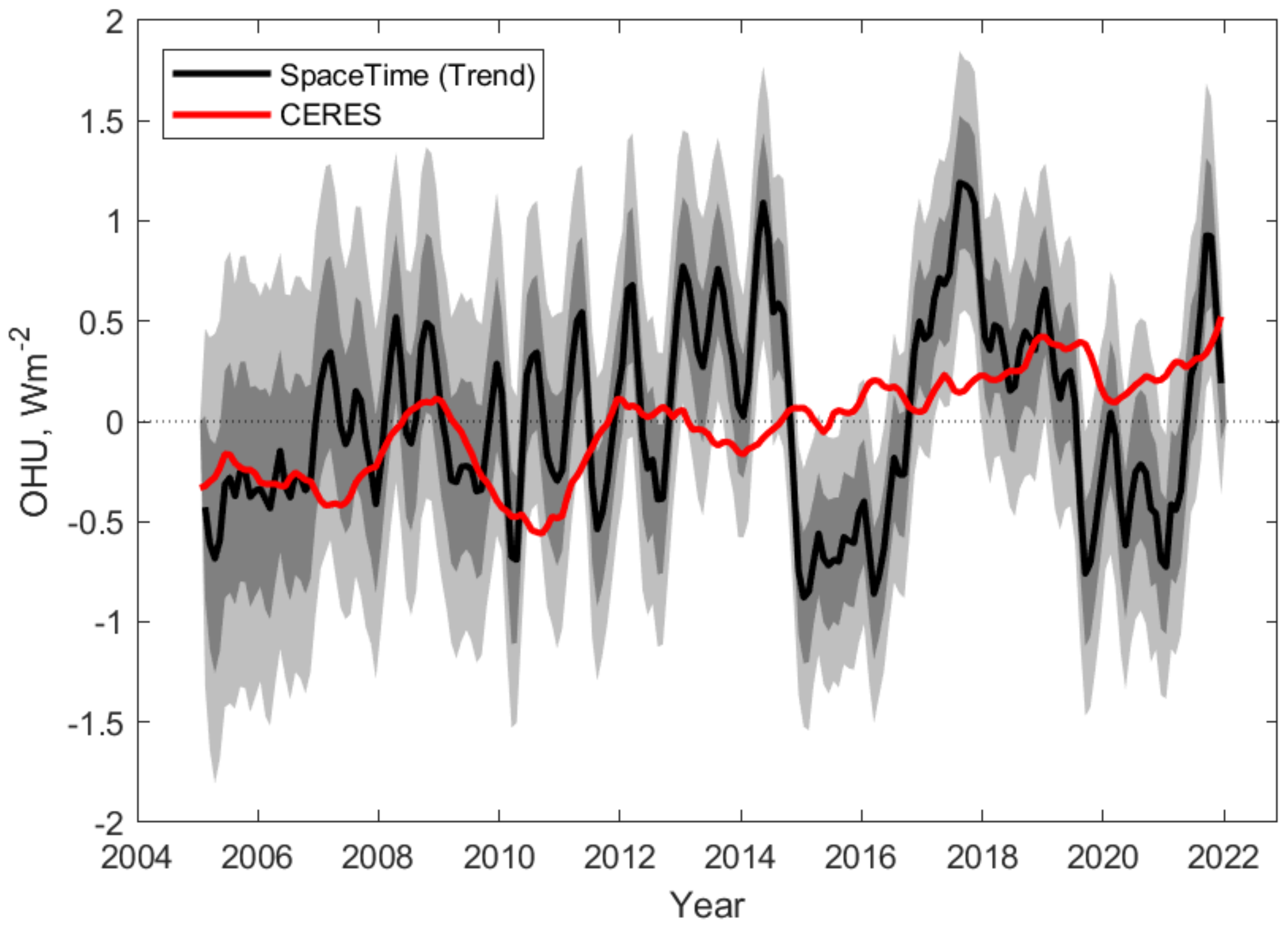
Uncertainties: Global ocean heat uptake (OHU) anomalies



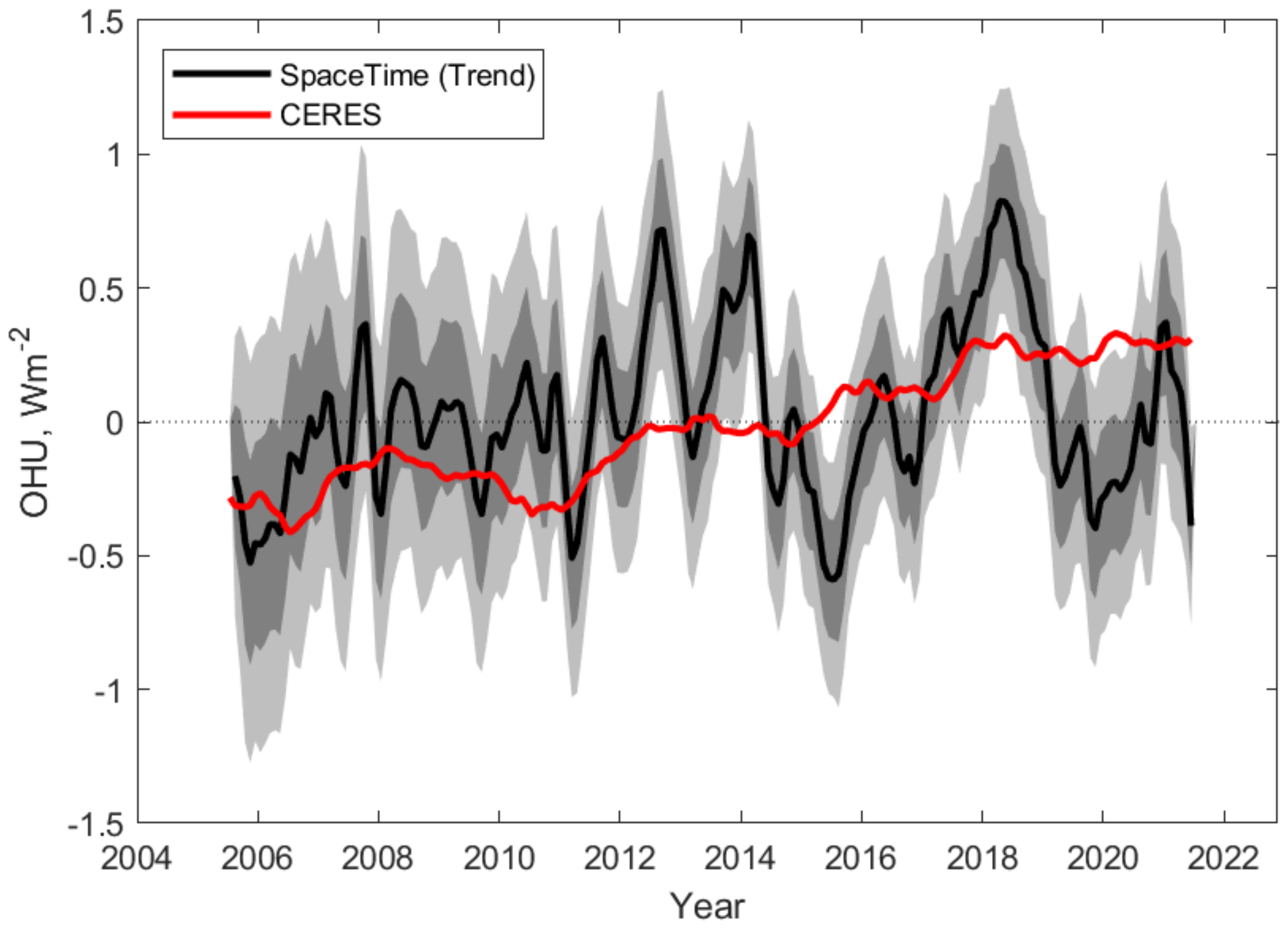
Monthly



12-month
moving average

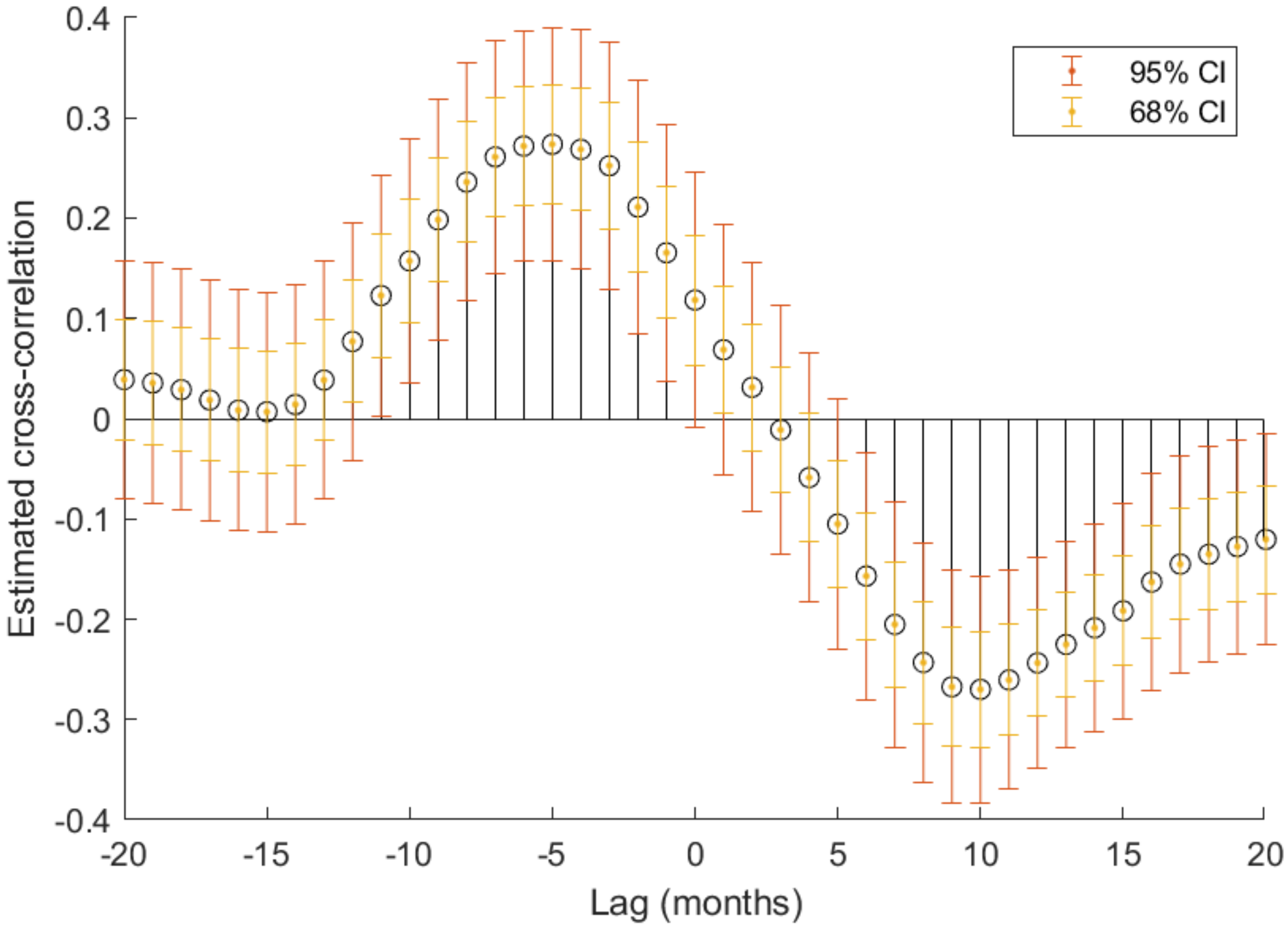


24-month
moving average

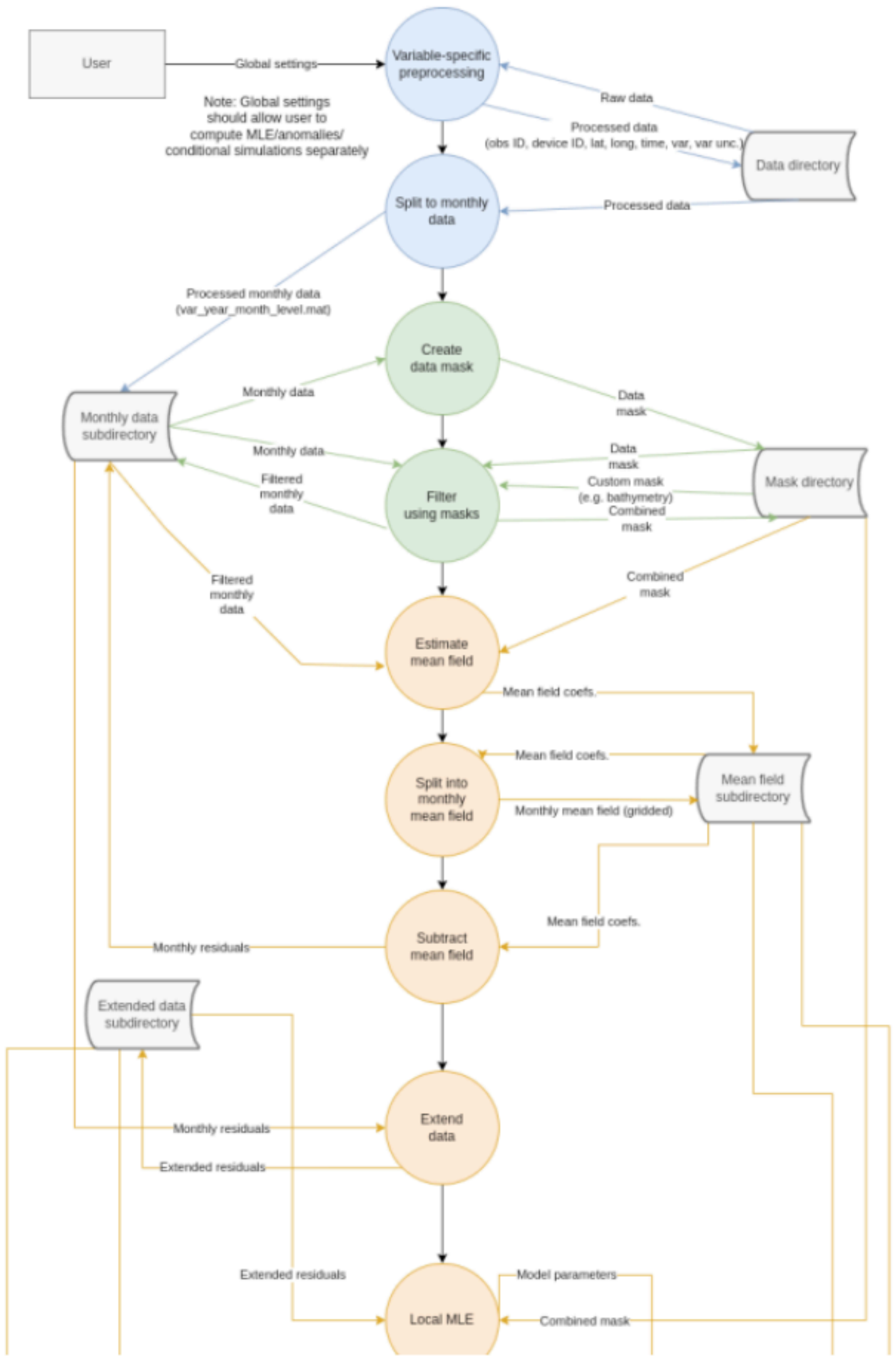


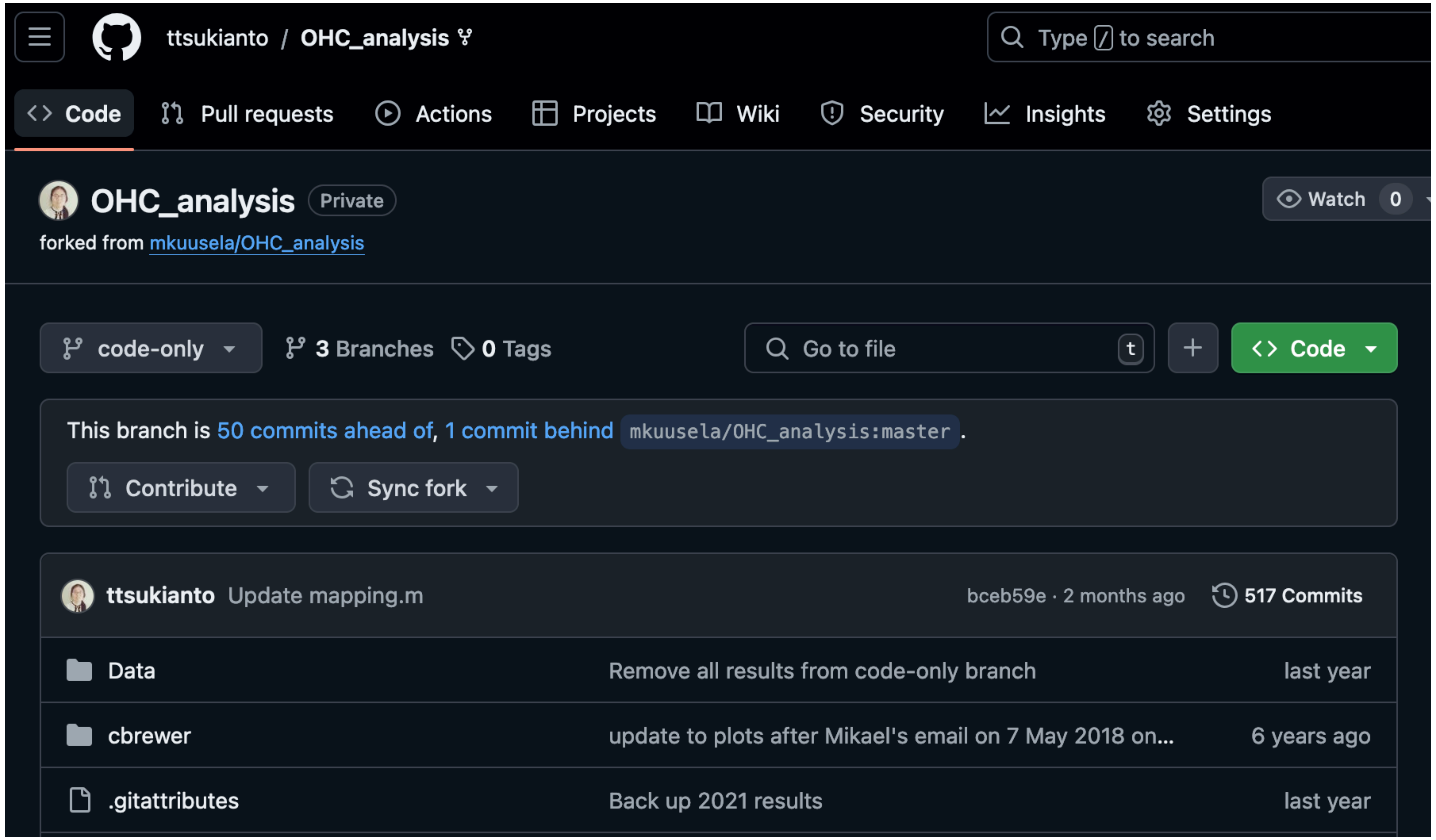
36-month
moving average

Uncertainties: Correlation between OHC anomalies and ONI



A reproducible, open-source pipeline (work-in-progress)





The framework in action: Data product

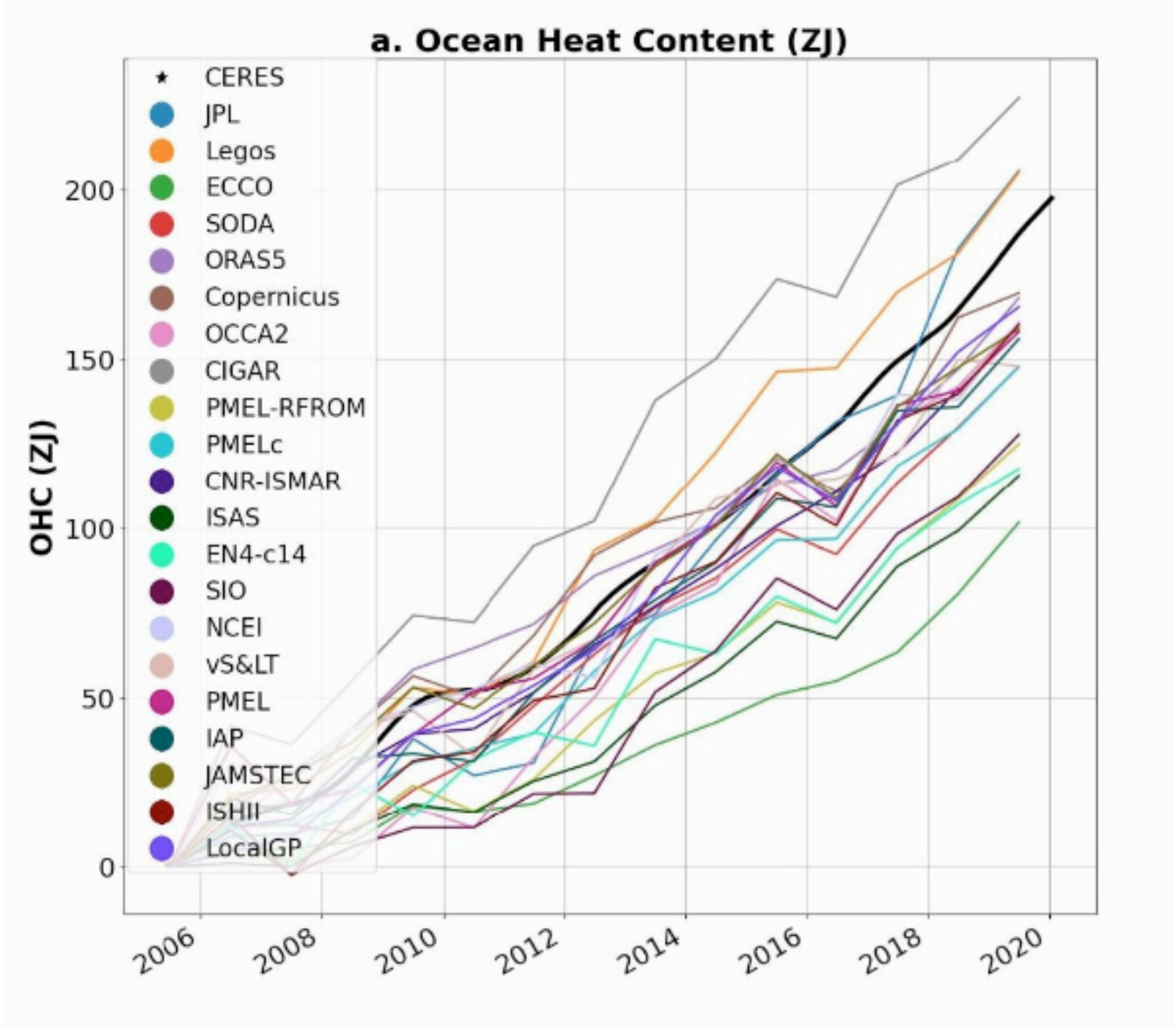
Global Ocean Heat Content Anomalies and Ocean Heat Uptake based on mapping Argo data using local Gaussian processes

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

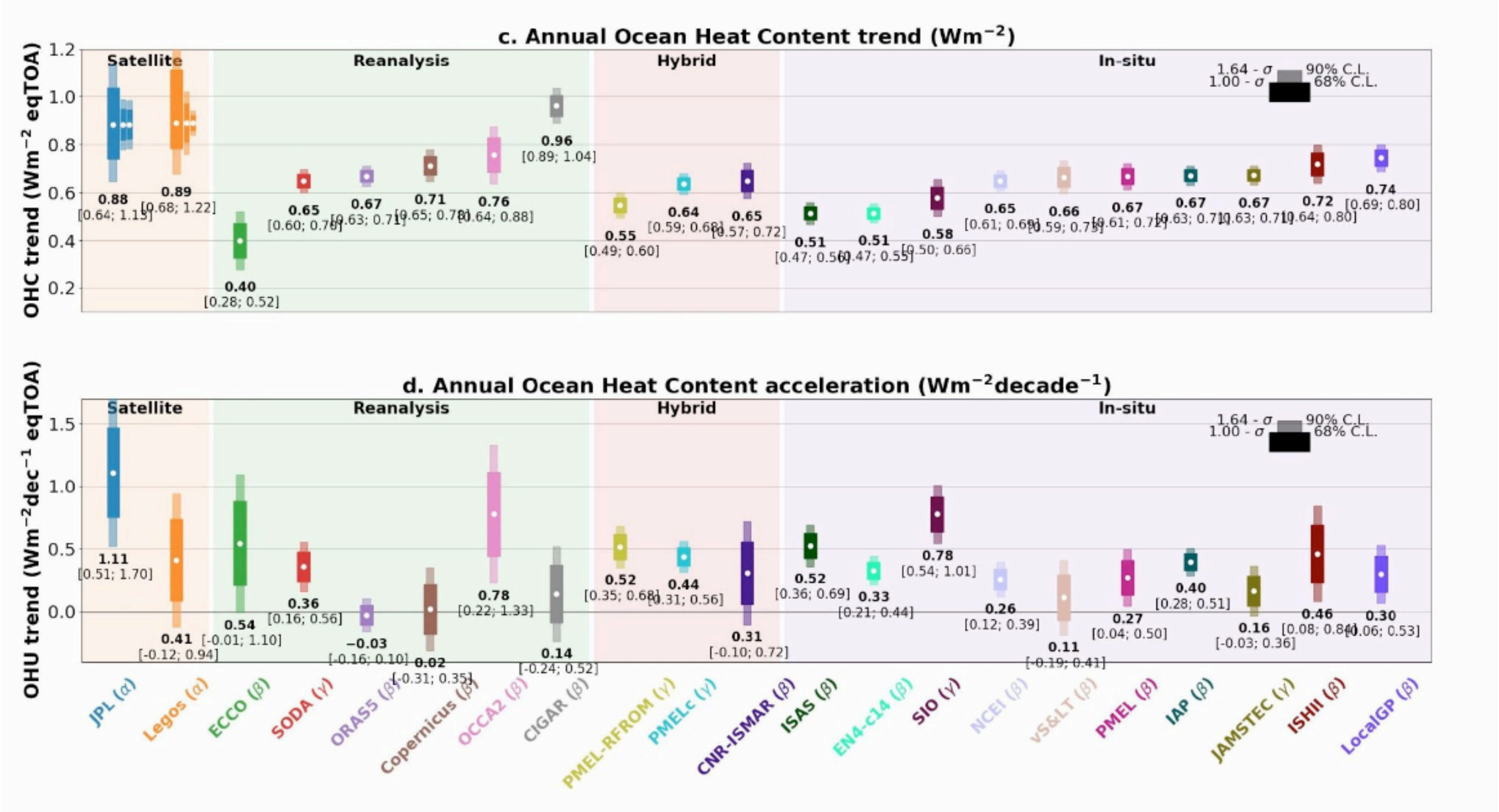
Show affiliations

NOTE for users: please use the latest version of the product. Monthly Ocean Heat Content Anomalies (OHCA) in the top 2000 dbar of the ocean are calculated (during 2004-2022, equatorward of 65 degree latitude) subtracting the mean over the period 2004-2022 from the monthly time series of OHC. Yearly OHCA time series are then calculated that include two points per year, i.e., from averaging Jan to Dec and Jul to Jun, respectively. OHC fields are mapped using locally stationary Gaussian processes (defined over space and time) with data-driven decorrelation scales (Kuusela and Stein, 2018). A linear time trend was included in the estimate of the mean field (along with spatial terms and harmonics for the annual cycle). Mapping is done separately for different vertical sections: 15-20 dbar, 15-300 dbar, 300-700 dbar, 700-1850 dbar, 1800-1850 dbar. The 15-20 dbar (1800-1850 dbar) section is used to estimate OHCA for 0-15 dbar (1850-2000 dbar), where observations are sparser. Different vertical sections are combined to estimate global OHCA time series for 0-2000 dbar. The attribute "area" is included in the netcdf files and it tells the corresponding surface area for the estimates. Regions of the ocean that are shallower than 300 m or are not sufficiently well sampled by the Argo array are not included. Ocean Heat Uptake is also calculated from the monthly OHCA and then averaged as described above to produce a yearly time series and included in the file.

The framework in action: Intercomparison efforts



(Hakuba et. al 2024)



A previous version of this data product was contributed to the 2022 WMO State of the Global Climate Report.

The implementation is still computationally challenging

- For every grid point, use data in a 10 degree/3 month window
- Estimate GP model parameters numerically with MLE + BFGS algorithm
- How many grid points? **360 long x 180 lat = 64,800 grid points (!)**
- Embarrassingly parallel, but still computationally challenging
  - Fit parameters (optimize all): ~**48h @ PSC**
  - Obtain conditional simulations for Feb of every year: **~48h**

Conclusions

- Estimating ocean heat content (with reliable uncertainties) is crucial for **tracking climate change and mitigation**
- To address nonstationarity and the large Argo dataset size, we model the temperature residuals with **locally stationary Gaussian process regression**
- To estimate the uncertainties of the total OHC in the water column (top + bottom), we generate ensembles of **local conditional simulations**
- Using this reproducible framework, we are able to obtain reliable uncertainties for **OHCA, regional/global OHC trends, and OHU**

Future work:

Bivariate extension, SSH/OHC joint mapping, neural inference (Walchessen et al. 2024)…

Data product

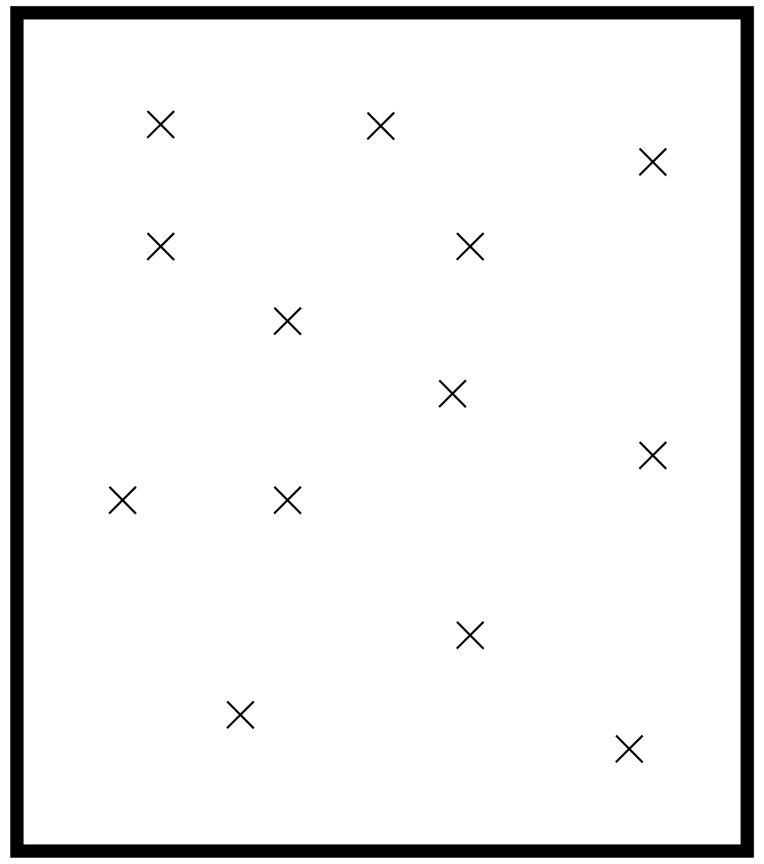
Questions? Comments? Concerns?

Contact: thea@stat.cmu.edu

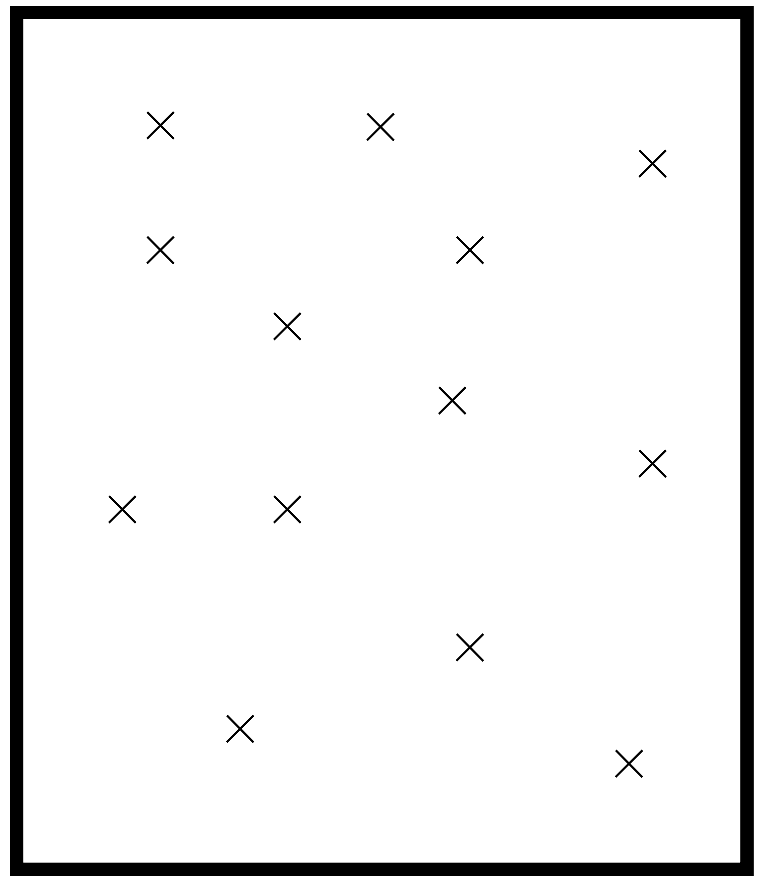
Intercomparison paper

Backup

How do we estimate uncertainties for the total (top + bottom layers) OHC?



$$\mathrm{Var}(\mathrm{OHC}_{\mathrm{top}}|\mathrm{data})$$



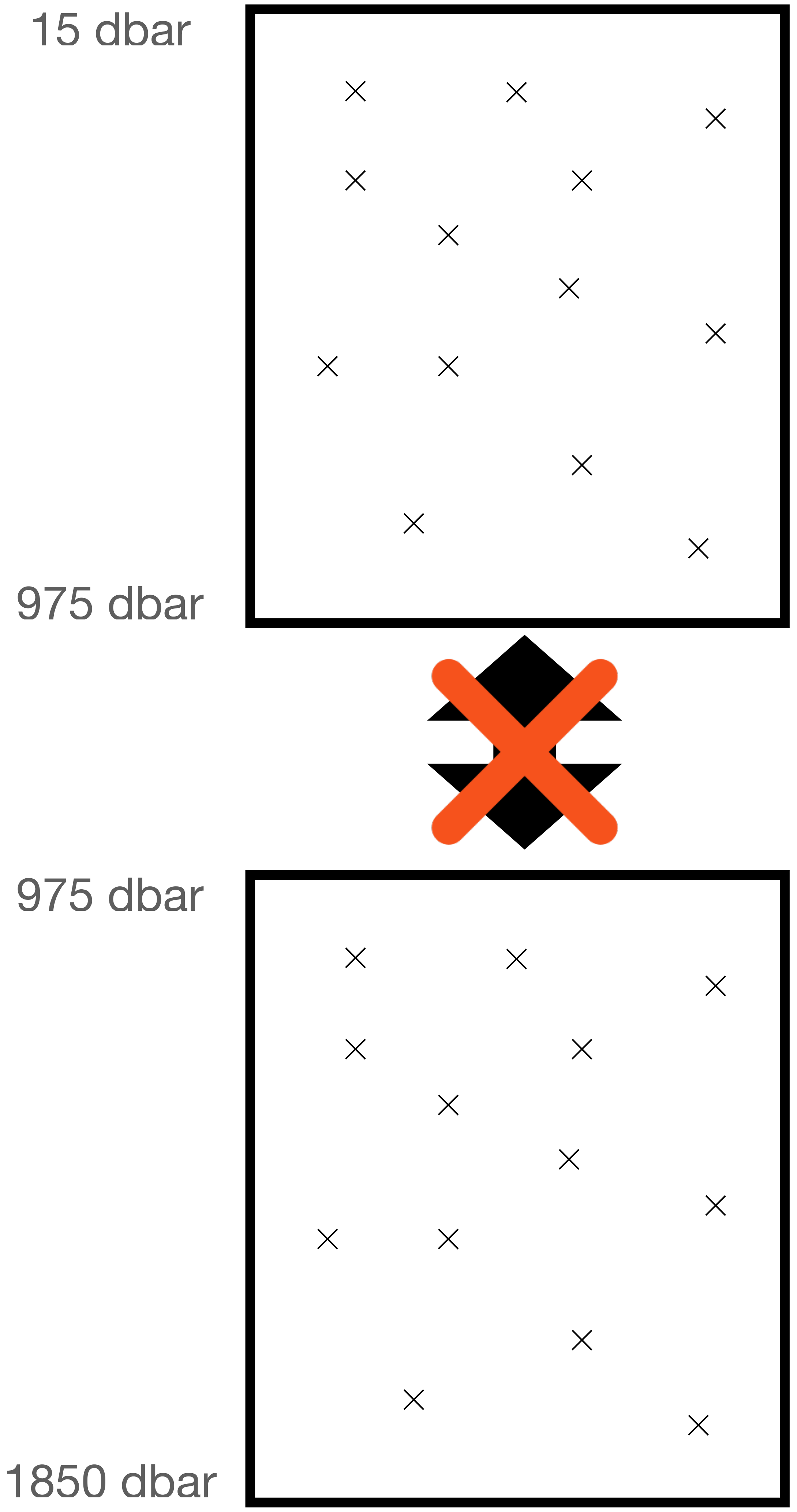
$$\mathrm{Var}(\mathrm{OHC}_{\mathrm{bot}}|\mathrm{data})$$

$$\mathrm{Var}(\cancel{\mathrm{OHC}_{\mathrm{total}}}|\mathrm{data})$$

$$\mathrm{Var}(\mathrm{OHC}_{\mathrm{top}}|\mathrm{data}) + \mathrm{Var}(\mathrm{OHC}_{\mathrm{bot}}|\mathrm{data})$$

Summing the variances for each layer **underestimates** the uncertainties of the total OHC.

How do we estimate uncertainties for the total (top + bottom layers) OHC?



$$\mathrm{Var}(\mathrm{OHC}_{\mathrm{top}}|\mathrm{data})$$

$$2\mathrm{Cov}(\mathrm{OHC}_{\mathrm{top}}, \mathrm{OHC}_{\mathrm{bot}}|\mathrm{data})$$

$$\mathrm{Var}(\mathrm{OHC}_{\mathrm{bot}}|\mathrm{data})$$

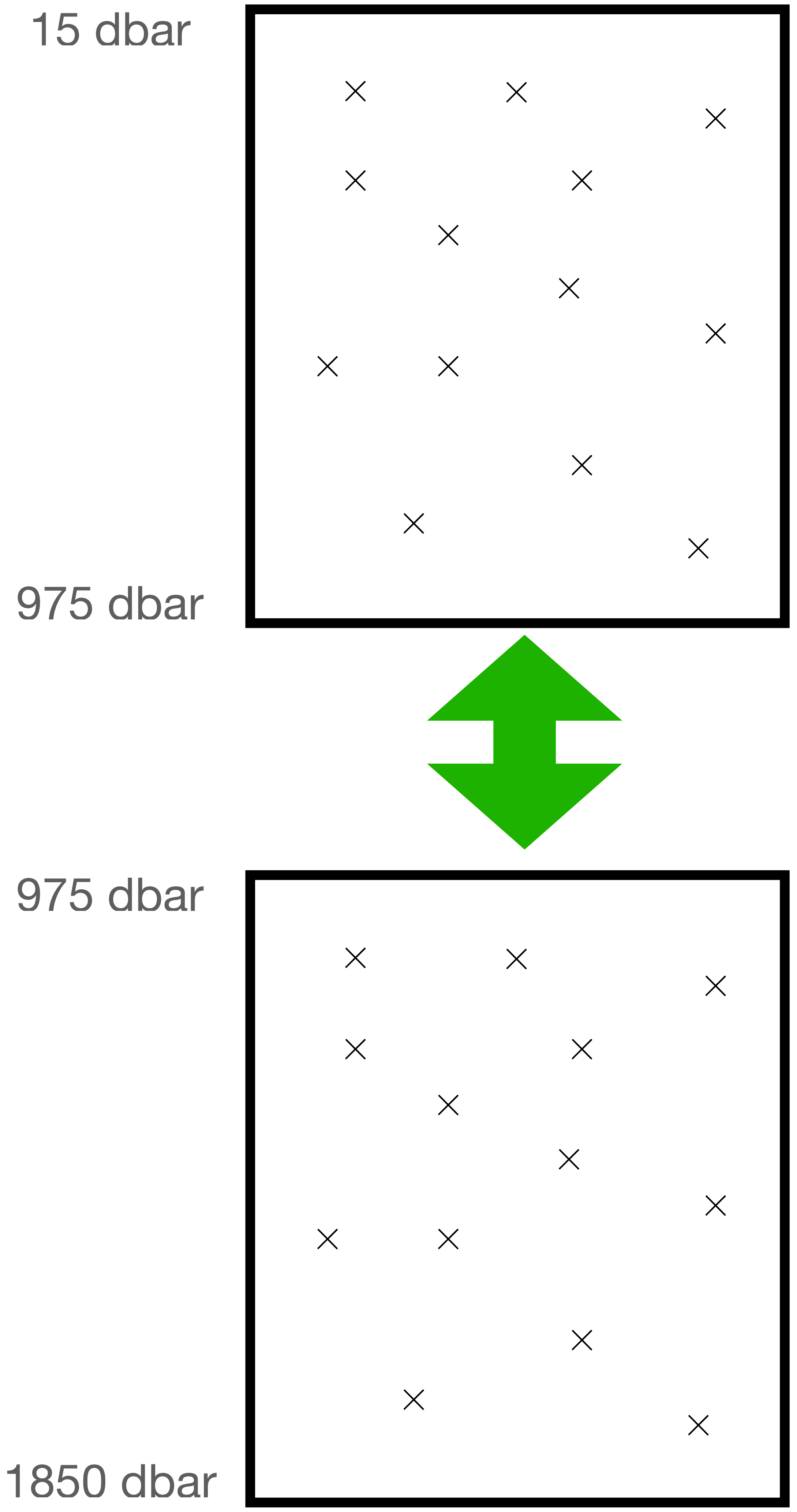
$$\mathrm{Var}(\mathrm{OHC}_{\mathrm{total}}|\mathrm{data})$$

(conservative upper bound)

$$\left(\sqrt{\mathrm{Var}(\mathrm{OHC}_{\mathrm{top}}|\mathrm{data})} + \sqrt{\mathrm{Var}(\mathrm{OHC}_{\mathrm{bot}}|\mathrm{data})}\right)^2$$

Squaring the sum of the standard deviations for each layer **overestimates** the uncertainties of the total OHC.

How do we estimate uncertainties for the total (top + bottom layers) OHC?



$$\mathrm{Var}(\mathrm{OHC}_{\mathrm{top}}|\mathrm{data})$$

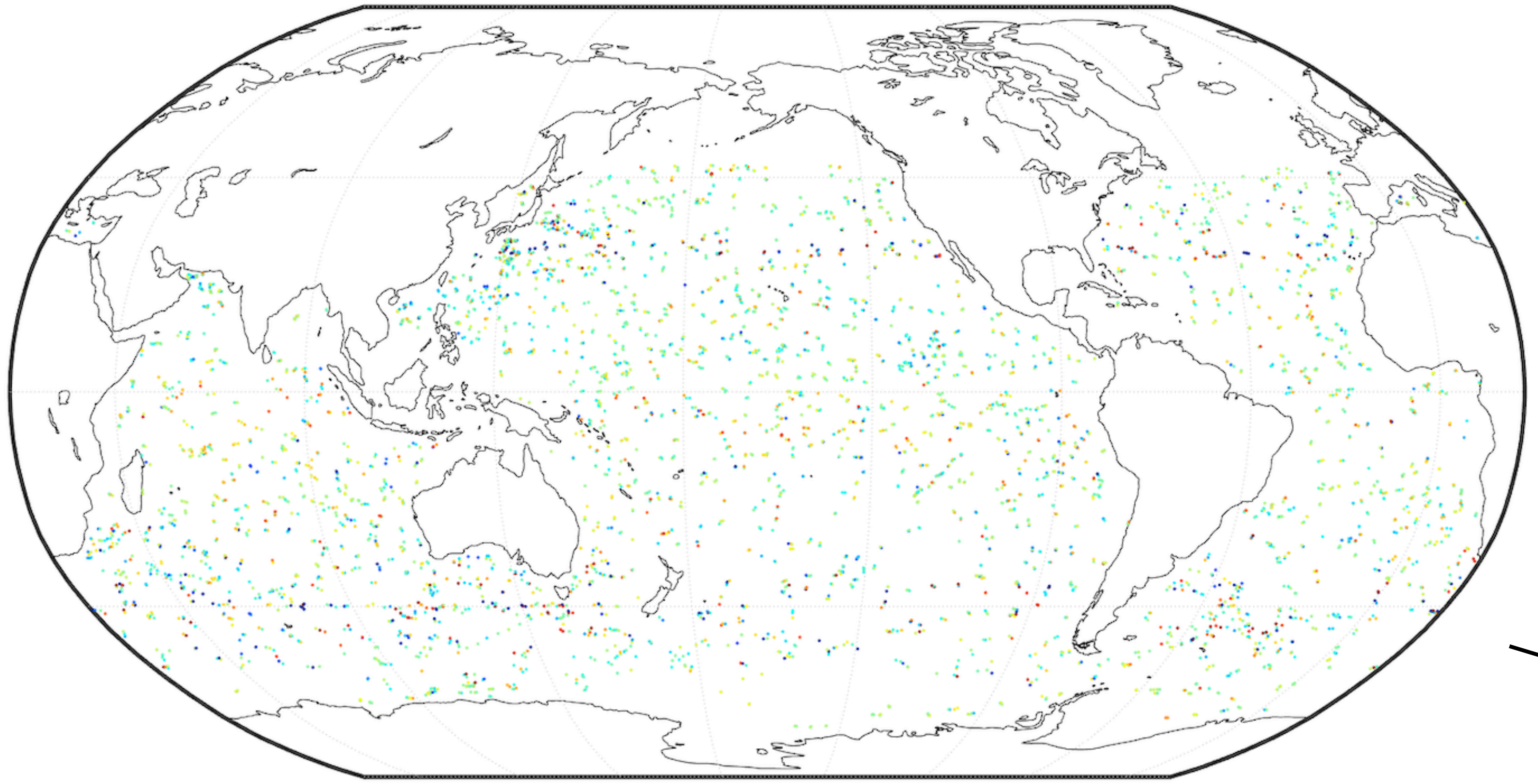
$$2\mathrm{Cov}(\mathrm{OHC}_{\mathrm{top}}, \mathrm{OHC}_{\mathrm{bot}}|\mathrm{data})$$

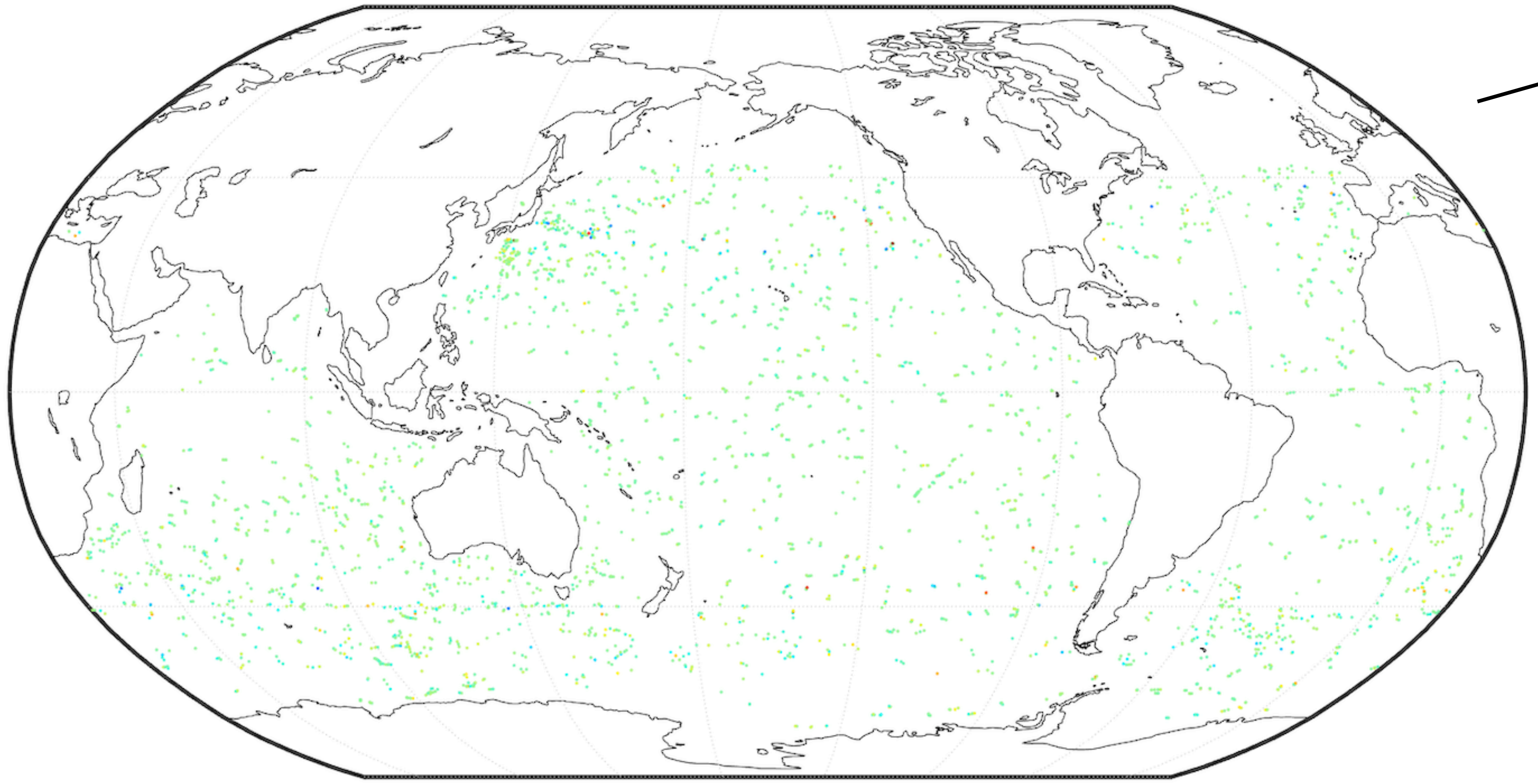
$$\mathrm{Var}(\mathrm{OHC}_{\mathrm{bot}}|\mathrm{data})$$

$$\mathrm{Var}(\mathrm{OHC}_{\mathrm{total}}|\mathrm{data})$$

We can improve the uncertainty estimates by also modeling the **correlation**.

A bivariate GP model accounts for cross-layer correlation





Temperature residuals	Latitude Longitude	Date	Nugget effect

$$\begin{bmatrix} y_{\text{top}} \\ y_{\text{bot}} \end{bmatrix}_{i,j} = f_i\left(\begin{bmatrix} x_{\text{top}} \\ x_{\text{bot}} \end{bmatrix}_{i,j}, \begin{bmatrix} t_{\text{top}} \\ t_{\text{bot}} \end{bmatrix}_{i,j}\right) + \begin{bmatrix} \epsilon_{\text{top}} \\ \epsilon_{\text{bot}} \end{bmatrix}_{i,j}$$

$$f_i \overset{\text{iid}}{\sim} \mathrm{GP}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{K}(x_1, t_1, x_2, t_2; \boldsymbol{\theta})\right)$$

(Covariance function)

$$\begin{bmatrix} \epsilon_{\text{top}} \\ \epsilon_{\text{bot}} \end{bmatrix} \overset{\text{iid}}{\sim} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{\epsilon}(\boldsymbol{\theta}_{\epsilon})\right)$$

A bivariate GP model accounts for cross-layer correlation

Marginal covariance
(Kuusela and Stein 2018)

$$\boldsymbol{K}_{\text{ii}}(\boldsymbol{z_1}, \boldsymbol{z_2}; \boldsymbol{\theta}) = \frac{\delta_i^2}{\sqrt{|\boldsymbol{\Theta}_i|}} \exp(-\sqrt{(\boldsymbol{z_1} - \boldsymbol{z_2})^{\mathrm{T}} \boldsymbol{\Theta}_i^{-1} (\boldsymbol{z_1} - \boldsymbol{z_2})})$$

GP variance

Space-time distance w/ length scale parameters

Cross-covariance
(Kleiber and Nychka 2012)

$$\boldsymbol{K}_{\text{top,bot}}(\boldsymbol{z_1}, \boldsymbol{z_2}; \boldsymbol{\theta}) = \beta \frac{\delta_{\text{top}} \delta_{\text{bot}}}{\sqrt{|\boldsymbol{\Theta}_{\text{top,bot}}|}} \exp(-\sqrt{(\boldsymbol{z_1} - \boldsymbol{z_2})^{\mathrm{T}} \boldsymbol{\Theta}_{\text{top,bot}}^{-1} (\boldsymbol{z_1} - \boldsymbol{z_2})})$$

(Cross) correlation