# Gaussian Processes for Spatio-Temporal Mapping of Oceanographic Observations

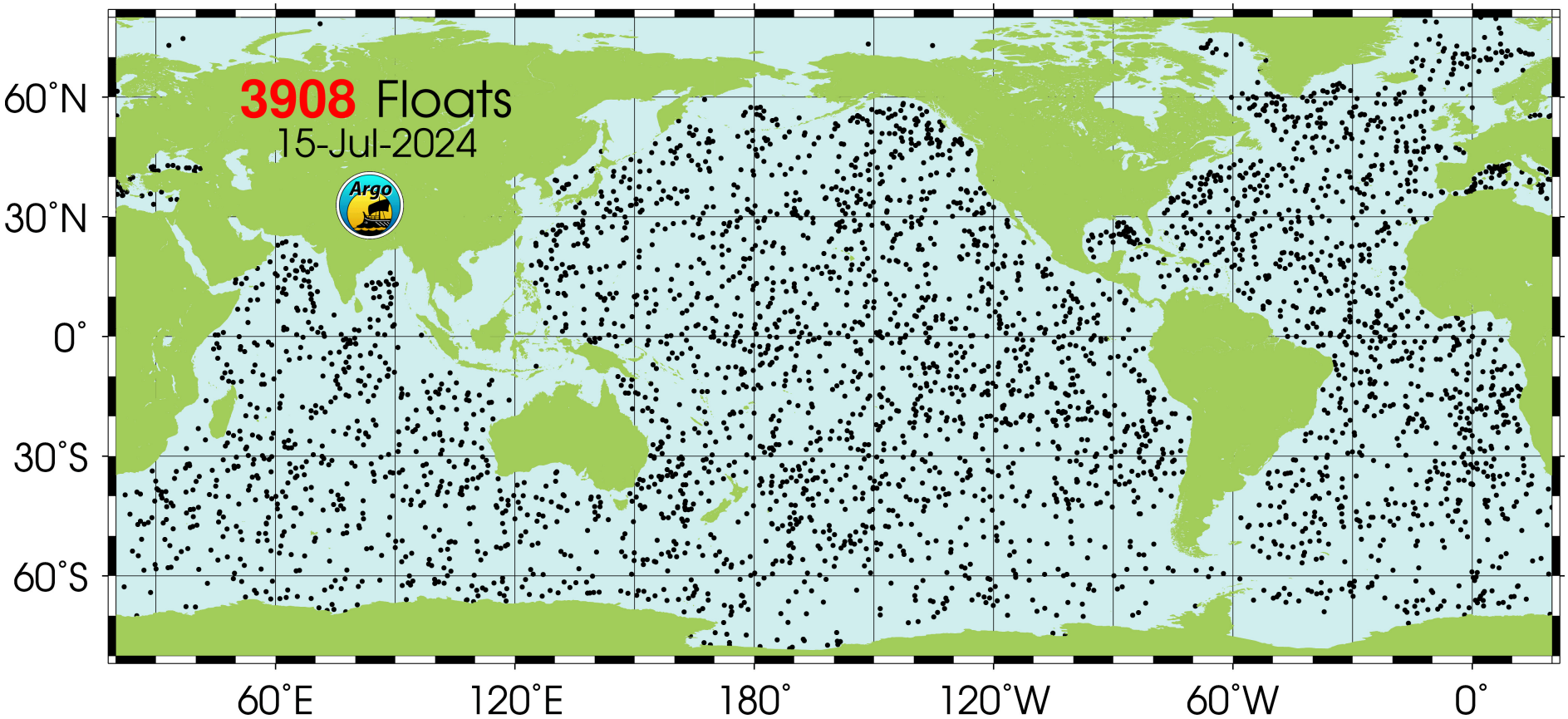
Mikael Kuusela

Department of Statistics and Data Science,
Carnegie Mellon University

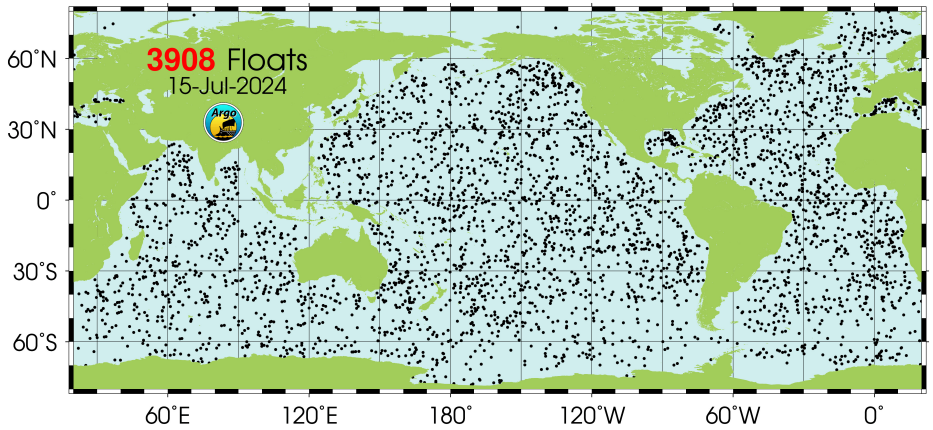
OceanUQ Summer School
Miami, FL

July 17, 2024

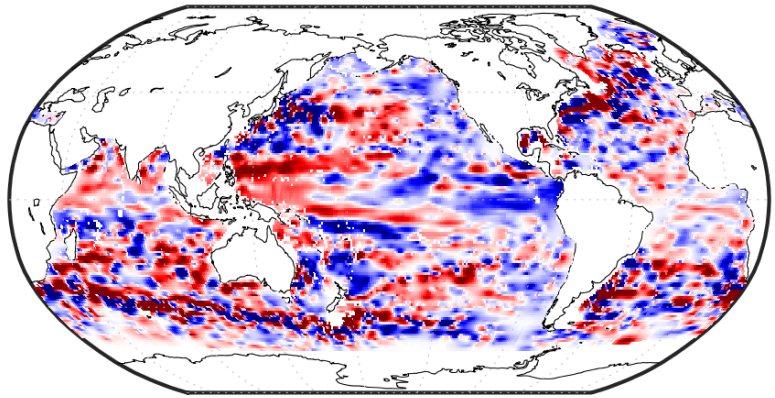
# Motivation: Argo floats



# Need for spatio-temporal interpolation




$\Longrightarrow$ ??



Many components of the ocean observing system produce in situ point observations (Argo floats, moorings, drifters, gliders, ships,...)

In order to do science with these data, it is often necessary to solve the *spatio-temporal interpolation* problem of mapping the point observations onto a regular grid over space and time

# Outline

In this lecture, we are going to develop the theory and practice of spatio-temporal interpolation through *Gaussian process (GP) regression*

This yields the standard spatio-temporal interpolants widely used in oceanography

- Note, however, that GPs are not the only way of arriving at these same interpolants and other interpolants are also possible

Outline:

1. A primer on Gaussian processes
2. Mean functions, covariance functions and parameter estimation
3. Gaussian process regression for interpolating oceanographic data

# Stochastic processes for spatio-temporal data

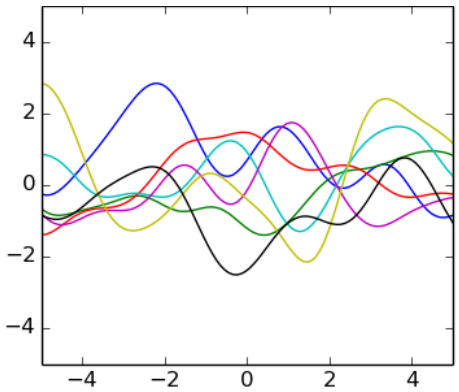
We take here the perspective that spatial fields are random realizations of an underlying stochastic process

A (continuously indexed) *stochastic process* is a collection of random variables $\{f(\boldsymbol{x})\}$ indexed by $\boldsymbol{x} \in D \subset \mathbb{R}^d$, where $f(\boldsymbol{x})$ is random for each $\boldsymbol{x}$

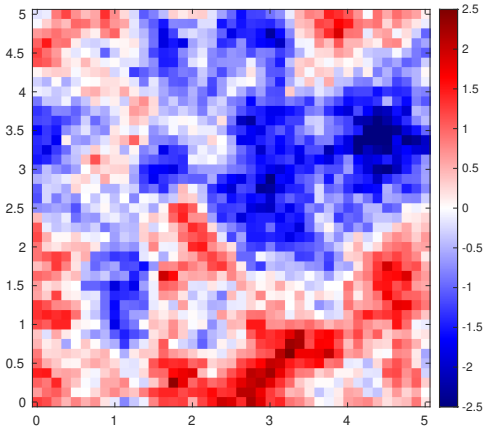
Equivalently, we can understand $f(\boldsymbol{x})$ as a *random function* of $\boldsymbol{x}$

Cases of specific interest in oceanography:

- Temporal processes: $D \subset \mathbb{R}$, $x$ represents time
- Spatial processes: $D \subset \mathbb{R}^2$, $\boldsymbol{x}$ represents space $[\boldsymbol{x} = (\text{lon}, \text{lat})]$
- Spatio-temporal processes: $D \subset \mathbb{R}^3$, $\boldsymbol{x}$ represents space and time $[\boldsymbol{x} = (\text{lon}, \text{lat}, t)]$




(Figure: Wikipedia)



# A Primer on Gaussian Processes

# Multivariate Gaussian distribution

A random vector $\boldsymbol{y} \in \mathbb{R}^n$ has an $n$-variate Gaussian distribution, denoted by $\boldsymbol{y} \sim N(\boldsymbol{m}, \boldsymbol{\Sigma})$, if its pdf is given by

$$p(\boldsymbol{y}|\boldsymbol{m}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{m})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{m})\right)$$

This is parameterized by the *mean vector* $\boldsymbol{m} \in \mathbb{R}^n$ and the symmetric and positive definite *covariance matrix* $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ so that

$$\begin{aligned} \mathbb{E}[y_i] &= m_i, \qquad \text{for all } i = 1, \ldots, n \\ \mathrm{Cov}[y_i, y_j] &= \Sigma_{ij}, \qquad \text{for all } i, j = 1, \ldots, n \end{aligned}$$

# Multivariate Gaussian distribution

Multivariate Gaussian random vectors have a number of nice properties

For example, consider the decomposition

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}, \quad \boldsymbol{m} = \begin{bmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Then the marginal distribution of $\boldsymbol{y}_1$ is

$$\boldsymbol{y}_1 \sim N(\boldsymbol{m}_1, \boldsymbol{\Sigma}_{11})$$

and the conditional distribution of $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ is

$$(\boldsymbol{y_1}|\boldsymbol{y}_2) \sim N(\boldsymbol{m}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{m}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

By rearranging the elements of $\boldsymbol{y}$, we can have any subset of elements in the component $\boldsymbol{y}_1$ and the remaining elements in the component $\boldsymbol{y}_2$. In other words:

- Any subset of elements of $\boldsymbol{y}$ has a multivariate Gaussian distribution
- Any subset of elements of $\boldsymbol{y}$ conditioned on the rest has a multivariate Gaussian distribution

# Gaussian process: Definition

Now, imagine that $n$ is very large. We then have a large collection of random variables

$$\{y_1, y_2, \ldots, y_{n-1}, y_n\} = \{y_i\}_{i=1}^n,$$

whose joint behavior is described by the multivariate Gaussian distribution. This collection is indexed by the discrete index $i \in [n]$.

A Gaussian process is an infinite-dimensional generalization of this to a collection of random variables indexed on a continuum.

In order words, a Gaussian process is a stochastic process satisfying the following:

**Definition**

A *Gaussian process* is a random function $f(\boldsymbol{x})$ whose values $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$ at any finite set of inputs $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ follow a multivariate Gaussian distribution.

# Gaussian process: Definition

## Definition

A *Gaussian process* is a random function $f(\boldsymbol{x})$ whose values $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$ at any finite set of inputs $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ follow a multivariate Gaussian distribution.

A Gaussian process is parameterized by a *mean function* $m(\boldsymbol{x})$ and a *covariance function* $k(\boldsymbol{x}_1, \boldsymbol{x}_2)$ so that

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})], \quad \text{for all } \boldsymbol{x}$$
$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathrm{Cov}[f(\boldsymbol{x}_1), f(\boldsymbol{x}_2)], \quad \text{for all } \boldsymbol{x}_1, \boldsymbol{x}_2.$$

We then denote $f \sim GP(m(\boldsymbol{x}), k(\boldsymbol{x}_1, \boldsymbol{x}_2))$.

The covariance function $k(\boldsymbol{x}_1, \boldsymbol{x}_2)$ has to be such that the covariance matrix of $[f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)]^{\mathsf{T}}$ for any inputs $\boldsymbol{x}_i$, $i = 1, \ldots, n$, is positive definite.

Functions $k(\boldsymbol{x}_1, \boldsymbol{x}_2)$ with this property are called *positive definite*. There are various well-known families of positive definite functions, but it's good to keep in mind that not all bivariate functions are valid covariance functions.

# Gaussian process: Inference

Let $f \sim GP(m(\boldsymbol{x}), k(\boldsymbol{x}_1, \boldsymbol{x}_2))$ and assume that we get to observe

$$y_1 = f(\boldsymbol{x}_1), y_2 = f(\boldsymbol{x}_2), \ldots, y_n = f(\boldsymbol{x}_n).$$

What can we then say about $y_* = f(\boldsymbol{x}_*)$ at some unobserved location $\boldsymbol{x}_*$?

Since $y_*$ is a random quantity, our task is to *predict* $y_*$.

Denote $\boldsymbol{y}_n = [y_1, \ldots, y_n]^\mathsf{T}$. Then, by definition:

$$\begin{bmatrix} y_* \\ \boldsymbol{y}_n \end{bmatrix} = \begin{bmatrix} y_* \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \sim N(\boldsymbol{m}, \boldsymbol{\Sigma}), \quad \text{where} \quad \boldsymbol{m} = \begin{bmatrix} m(\boldsymbol{x}_*) \\ m(\boldsymbol{x}_1) \\ \vdots \\ m(\boldsymbol{x}_n) \end{bmatrix} = \begin{bmatrix} m(\boldsymbol{x}_*) \\ \boldsymbol{m}_n \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} k(\boldsymbol{x}_*, \boldsymbol{x}_*) & k(\boldsymbol{x}_*, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_*, \boldsymbol{x}_n) \\ k(\boldsymbol{x}_1, \boldsymbol{x}_*) & k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_n, \boldsymbol{x}_*) & k(\boldsymbol{x}_n, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_n, \boldsymbol{x}_n) \end{bmatrix} = \begin{bmatrix} k(\boldsymbol{x}_*, \boldsymbol{x}_*) & \boldsymbol{k}_*^\mathsf{T} \\ \boldsymbol{k}_* & \boldsymbol{K}_n \end{bmatrix}$$

# Gaussian process: Inference

Then, by the properties of the multivariate Gaussian distribution, the conditional distribution of $y_*$ given $\boldsymbol{y}_n$ is

$$(y_*|\boldsymbol{y}_n) \sim N(m(\boldsymbol{x}_*) + \boldsymbol{k}_*^\mathsf{T}\boldsymbol{K}_n^{-1}(\boldsymbol{y}_n - \boldsymbol{m}_n), k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^\mathsf{T}\boldsymbol{K}_n^{-1}\boldsymbol{k}_*)$$

Since we are trying to predict $y_*$ given $\boldsymbol{y}_n$, this is also known as the *predictive distribution* of $y_*$. We can directly extract from this the predictive mean

$$\hat{y}_* = \mathbb{E}[y_*|\boldsymbol{y}_n] = m(\boldsymbol{x}_*) + \boldsymbol{k}_*^\mathsf{T}\boldsymbol{K}_n^{-1}(\boldsymbol{y}_n - \boldsymbol{m}_n)$$

and the predictive variance

$$\hat{\sigma}_*^2 = \mathrm{Var}[y_*|\boldsymbol{y}_n] = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^\mathsf{T}\boldsymbol{K}_n^{-1}\boldsymbol{k}_*.$$

We can then predict $y_*$ using $\hat{y}_*$. We know from Monday that this is the mean squared error optimal predictor of $y_*$.

The predictive uncertainty can be quantified using the $1-\alpha$ prediction interval which is given by $[\hat{y}_* - z_{1-\alpha/2}\hat{\sigma}_*, \hat{y}_* + z_{1-\alpha/2}\hat{\sigma}_*]$.

# Gaussian process: Inference

As a result, we conclude that $y_*$ should be predicted using

$$\hat{y}_* = m(\boldsymbol{x}_*) + \boldsymbol{k}_*^{\mathsf{T}} \boldsymbol{K}_n^{-1} (\boldsymbol{y}_n - \boldsymbol{m}_n)$$

and the uncertainty of the prediction at level $1 - \alpha$ is given by

$$[\hat{y}_* - z_{1-\alpha/2}\hat{\sigma}_*, \hat{y}_* + z_{1-\alpha/2}\hat{\sigma}_*]$$

This has various names depending on the context, including *kriging* (spatial statistics / geostatistics), *objective mapping* (oceanography) or *optimal interpolation* (atmospheric science)

# Gaussian process: Inference

Notice also that we can repeat this same calculation for other $\boldsymbol{x}_*$'s to obtain pointwise predictions of $f(\boldsymbol{x})$ on a fine grid, for example.

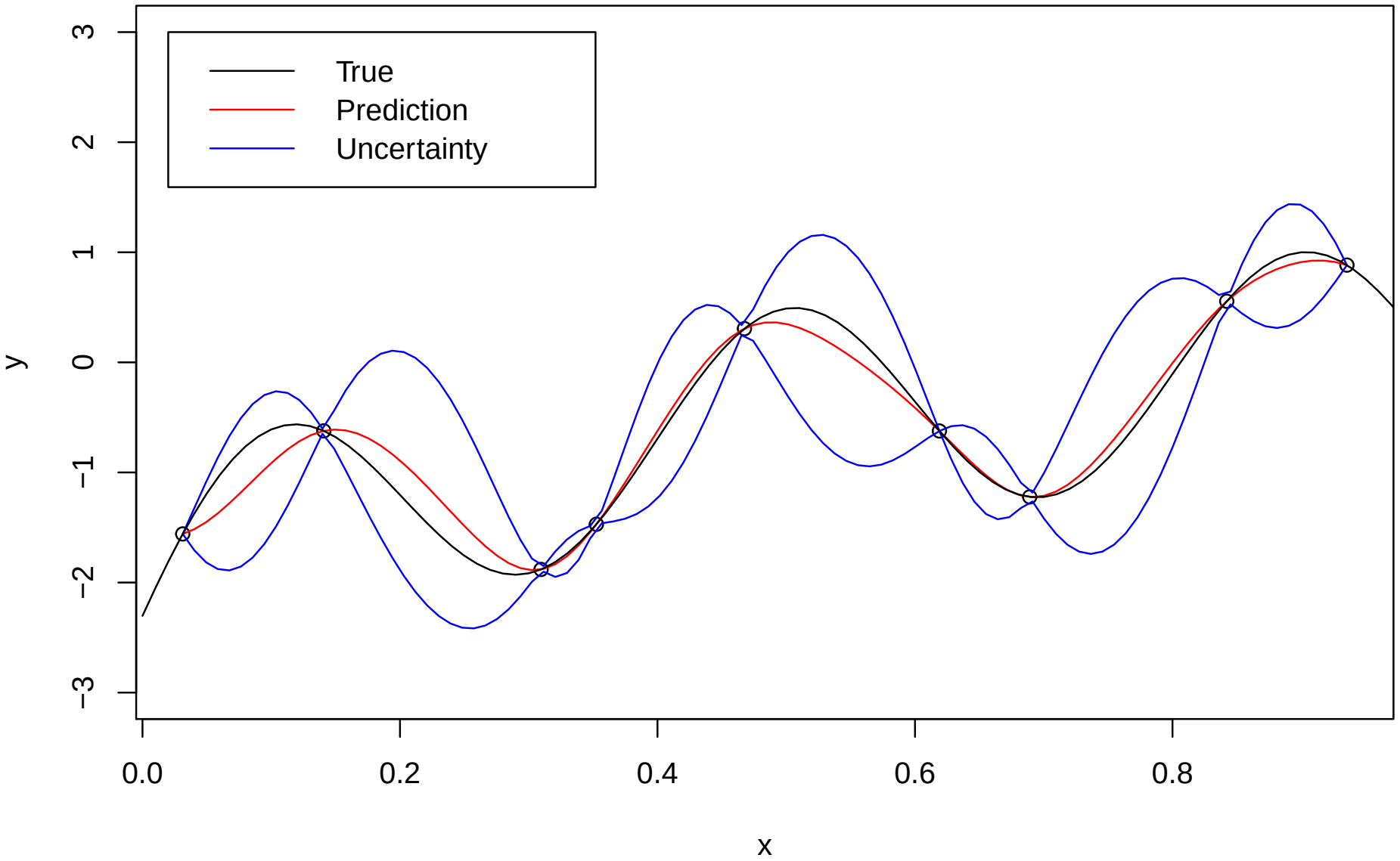
We can also repeat the same steps for the vector

$$[y_{1*}, \ldots, y_{p*}, y_1, \ldots, y_n]^\mathsf{T} = [f(\boldsymbol{x}_{1*}), \ldots, f(\boldsymbol{x}_{p*}), f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)]^\mathsf{T}$$

to obtain the predictive distribution of $[y_{1*}, \ldots, y_{p*}]^\mathsf{T}$ given $[y_1, \ldots, y_n]^\mathsf{T}$, which also provides the predictive covariance between different locations $\boldsymbol{x}_{i*}$.

**Key observation:** Because finite evaluations of a Gaussian process follow a multivariate Gaussian distribution, we immediately know how to make a finite number of predictions given a finite number of observations.

# Illustration



# Gaussian process regression

In practice, we do not necessarily want to force the prediction to go through the observations $y_1, \ldots, y_n$.

Therefore, the following *Gaussian process regression* model is commonly employed:

$$y_i = f(\boldsymbol{x}_i) + \varepsilon_i,$$

where $f \sim GP(m(\boldsymbol{x}), k(\boldsymbol{x}_1, \boldsymbol{x}_2))$, $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$ and $f$ is independent of the $\varepsilon_i$'s.

The extra term $\varepsilon_i$ is called the *nugget effect* and corresponds to measurement error, unexplained variation or microscale variation, depending on the context.

One might then be interested in predicting either $f_* = f(\boldsymbol{x}_*)$ or $y_* = f(\boldsymbol{x}_*) + \varepsilon_*$

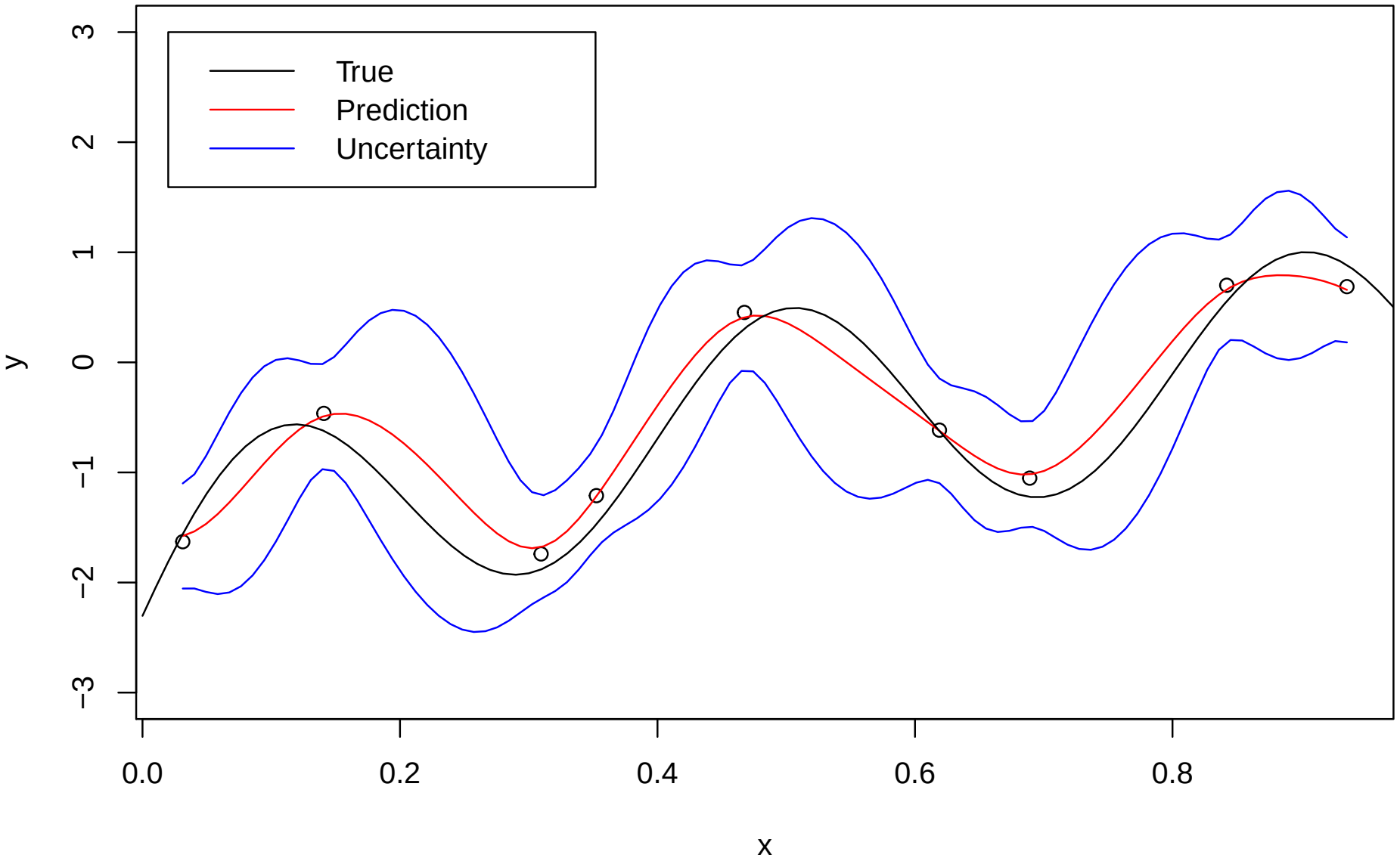
The predictive distribution in the first case is

$$(f_*|\boldsymbol{y}_n) \sim N(m(\boldsymbol{x}_*)+\boldsymbol{k}_*^{\mathsf{T}}(\boldsymbol{K}_n+\sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y}_n-\boldsymbol{m}_n), k(\boldsymbol{x}_*, \boldsymbol{x}_*)-\boldsymbol{k}_*^{\mathsf{T}}(\boldsymbol{K}_n+\sigma^2\boldsymbol{I})^{-1}\boldsymbol{k}_*)$$

The latter case is otherwise the same but the predictive variance is

$$\text{Var}[y_*|\boldsymbol{y}_n] = \text{Var}[f_*|\boldsymbol{y}_n] + \sigma^2 = k(\boldsymbol{x}_*, \boldsymbol{x}_*)+\sigma^2-\boldsymbol{k}_*^{\mathsf{T}}(\boldsymbol{K}_n+\sigma^2\boldsymbol{I})^{-1}\boldsymbol{k}_*$$

# Illustration



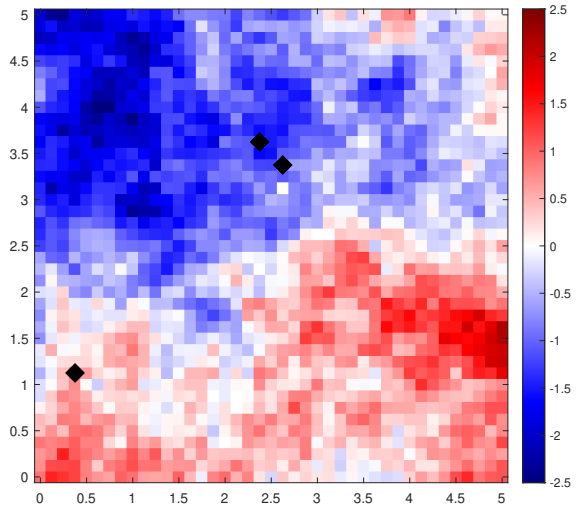


Figure: GP realization

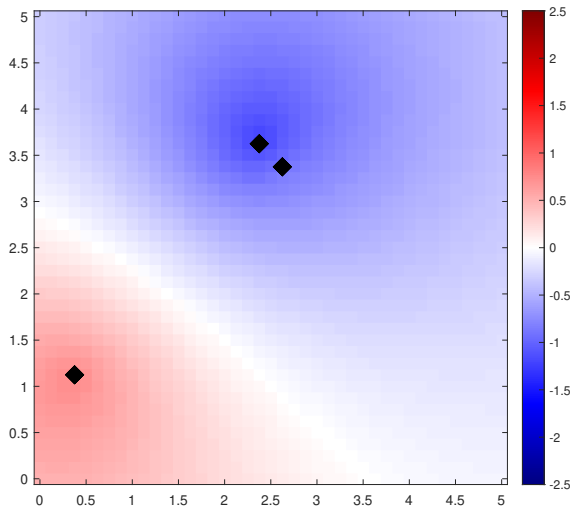


Figure: Conditional mean

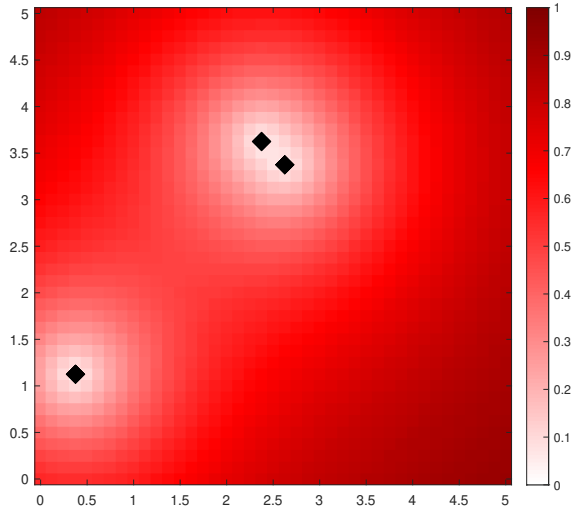


Figure: Conditional variance

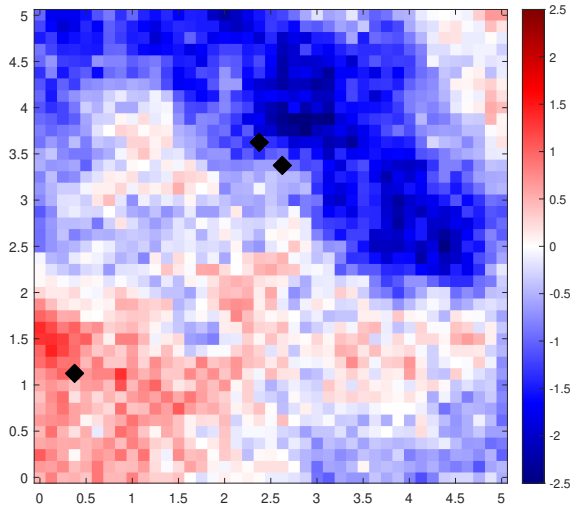


Figure: Conditional simulation

# Mean functions, covariance functions and parameter estimation

# Gaussian process modeling

A Gaussian process $f \sim GP(m(\boldsymbol{x}), k(\boldsymbol{x}_1, \boldsymbol{x}_2))$ is parameterized by the *mean function* $m(\boldsymbol{x})$ and the *covariance function* $k(\boldsymbol{x}_1, \boldsymbol{x}_2)$

In order to model data using a GP, one needs to decide how to choose these functions.

A significant portion of GP literature revolves around this question.

Sometimes there is ambiguity with regards to what portion of the data should be explained using $m(\boldsymbol{x})$ and what portion using $k(\boldsymbol{x}_1, \boldsymbol{x}_2)$, especially if there is only a single realization of $f$

- "One person's mean structure is another person's covariance structure"

Some authors claim that one can simply set $m(\boldsymbol{x}) = 0$ without loss of generality, but it's not that simple

In practice, we tend to use certain parametric classes of functions for both:

$$m(\boldsymbol{x}) = m(\boldsymbol{x}; \boldsymbol{\beta}), \quad k(\boldsymbol{x}_1, \boldsymbol{x}_2) = k(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{\theta})$$

# Choice of the mean function

The mean function $m(\boldsymbol{x})$ should be flexible enough to model the average shape of the random function $f(\boldsymbol{x})$, but also rigid enough to not fit the stochastic fluctuations in the data

It can be difficult to strike a balance here, but luckily the final predictions are usually quite robust against modest misspecification of the mean

Common choices for $m(\boldsymbol{x}; \boldsymbol{\beta})$:

- Linear in $\boldsymbol{x}$ and $\boldsymbol{\beta}$: $m(\boldsymbol{x}; \boldsymbol{\beta}) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i$
- Splines (especially in 1D): $m(x; \boldsymbol{\beta}) = \sum_{i=1}^{p} \beta_i B_i(x)$, where $B_i(\cdot)$ are B-spline basis functions
- Nonlinear (in both $\boldsymbol{x}$ and $\boldsymbol{\beta}$) regression functions (e.g., neural nets)

# Choice of the covariance function

Recall that $k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \text{Cov}[f(\boldsymbol{x}_1), f(\boldsymbol{x}_2)]$.

Which bivariate function $k(\cdot, \cdot)$ to use? (Remember that $k(\cdot, \cdot)$ needs to be positive definite.)

A common assumption is to say that $k(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is *stationary* (i.e., translation invariant): $k(\boldsymbol{x}_1, \boldsymbol{x}_2) = k(\boldsymbol{x}_1 - \boldsymbol{x}_2)$

Furthermore, it is common to assume *isotropy*

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = k(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)$$

or *geometric anisotropy*

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = k(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\boldsymbol{A}}),$$

where $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\boldsymbol{A}} = \sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\mathsf{T}} \boldsymbol{A} (\boldsymbol{x}_1 - \boldsymbol{x}_2)}$ for a positive definite matrix $\boldsymbol{A}$

# Choice of the covariance function

Let's focus on the case with geometric anisotropy. Denote $s = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\boldsymbol{A}}$.

At this point, we need to choose the matrix $\boldsymbol{A}$ and the function $k(s)$.

Here $\boldsymbol{A}$ controls the length scales (decorrelation scales) and orientation of the dependence in $f(\boldsymbol{x})$ over $\boldsymbol{x}$.

The function $k(s)$ controls the remaining properties of the random function $f(\boldsymbol{x})$, such as smoothness, periodicity, etc.

# Choice of the covariance function

Popular models for $k(s)$ include:

- Exponential: $k(s) = \phi \exp(-s),\ \phi > 0$
  - $f(\boldsymbol{x})$ continuous but not differentiable
- Squared exponential: $k(s) = \phi \exp(-s^2),\ \phi > 0$
  - $f(\boldsymbol{x})$ infinitely differentiable
- Matérn: $k(s) = \phi \frac{2^{1-\nu}}{\Gamma(\nu)} s^\nu K_\nu(s),\ \phi > 0$, where $\nu > 0$ is a smoothness parameter and $K_\nu$ is a modified Bessel function
  - $f(\boldsymbol{x})$ $k$ times differentiable if and only if $\nu > k$
  - Gives exponential for $\nu = \frac{1}{2}$ and squared exponential for $\nu \to \infty$
  - Has simplified form when $\nu$ is half integer, i.e., $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$

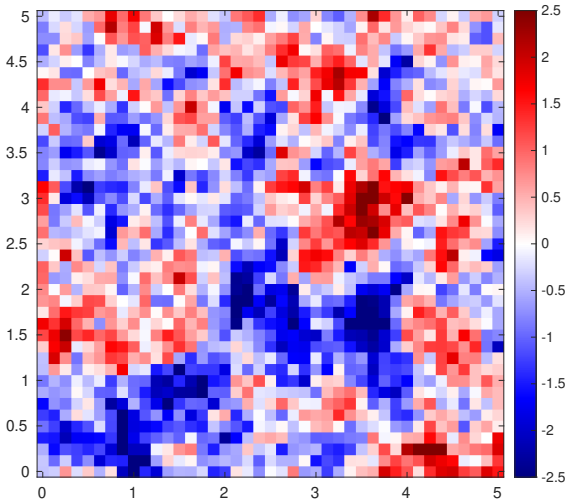
For example, if we pick $\boldsymbol{A} = \text{diag}(1/\theta_1^2, \ldots, 1/\theta_d^2)$ and let $k(s)$ be exponential, then we have the following covariance model

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2; \phi, \theta_1, \ldots, \theta_d)$$

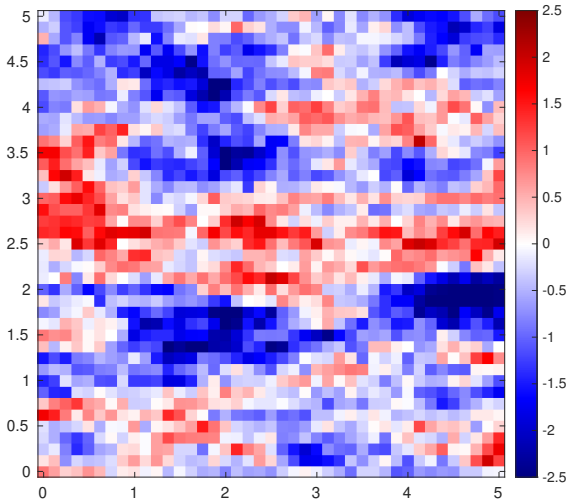
$$= \phi \exp\left( - \sqrt{\left(\frac{x_{11} - x_{21}}{\theta_1}\right)^2 + \left(\frac{x_{12} - x_{22}}{\theta_2}\right)^2 + \cdots + \left(\frac{x_{1d} - x_{2d}}{\theta_d}\right)^2} \right)$$

parameterized by $\phi, \theta_1, \ldots, \theta_d > 0$

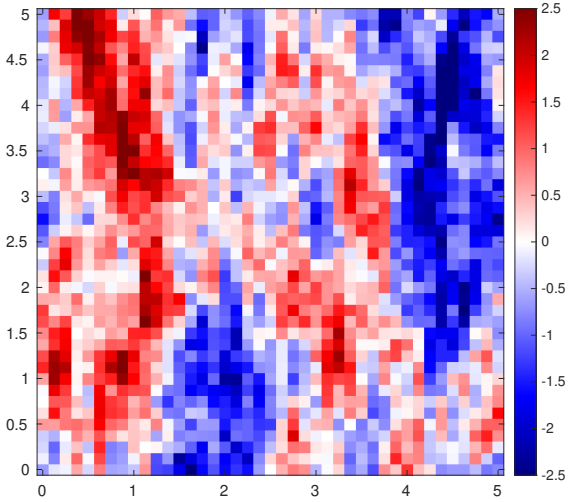
# Illustration: Effect of covariance length scales



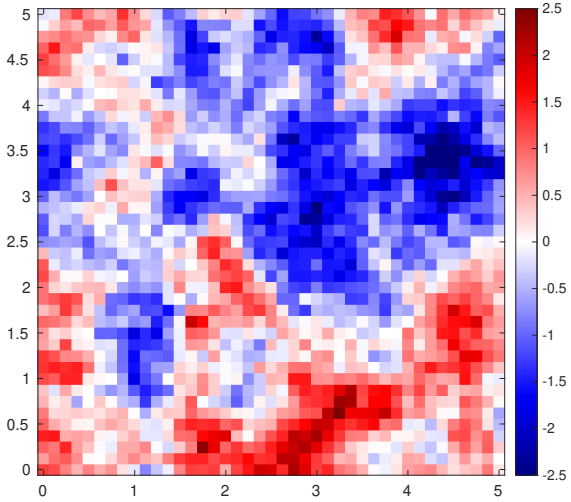
(a) $\theta_1 = 0.3,\ \theta_2 = 0.3$



(b) $\theta_1 = 1,\ \theta_2 = 0.3$



(c) $\theta_1 = 0.3,\ \theta_2 = 1$



(d) $\theta_1 = 1,\ \theta_2 = 1$

# Parameter estimation

Let $\boldsymbol{\theta}$ denote the vector of parameters affecting to covariance function $k(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{\theta})$

Then the unknown parameters of the GP model are $(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$ and we wish to learn these parameters using the observed data $\boldsymbol{y}_n$

Various techniques for estimating these parameters exist, but the most common approach is to use maximum likelihood.

Since $\boldsymbol{y}_n$ follows a multivariate Gaussian, the log-likelihood of $(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$ is

$$\begin{aligned}\ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) &= \log p(\boldsymbol{y}_n | \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) \\ &= -\frac{1}{2}\Big[n \log(2\pi) + \log\det\left(\boldsymbol{K}_n(\boldsymbol{\theta}) + \sigma^2 \boldsymbol{I}\right) \\ &\qquad + (\boldsymbol{y}_n - \boldsymbol{m}_n(\boldsymbol{\beta}))^{\mathsf{T}} (\boldsymbol{K}_n(\boldsymbol{\theta}) + \sigma^2 \boldsymbol{I})^{-1} (\boldsymbol{y}_n - \boldsymbol{m}_n(\boldsymbol{\beta}))\Big]\end{aligned}$$

The estimates $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2)$ are those values that maximize $\ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$

For linear mean functions, $\boldsymbol{\beta}$ can be solved in closed-form (for given $(\boldsymbol{\theta}, \sigma^2)$), but to solve $(\boldsymbol{\theta}, \sigma^2)$ one needs to typically use numerical optimization

# Gaussian Process Regression for Interpolating Oceanographic Data

# Gaussian process model for ocean observations

The following Gaussian process model is often used (either explicitly or implicitly) for interpolating oceanographic observations:

$$y_{i,j} = f_i(x_{\text{lon},i,j}, x_{\text{lat},i,j}, t_{i,j}) + \varepsilon_{i,j},$$
$$f_i \overset{\text{iid}}{\sim} \text{GP}(m, k), \quad \varepsilon_{i,j} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

where:

- $y_{i,j}$ is some in situ observation (temperature, salinity, oxygen,...)
- $i = 1, \ldots, N$ refers to years and $j = 1, \ldots, n_i$ to observations in the $i$th year
- $x_{\text{lon},i,j}$, $x_{\text{lat},i,j}$ and $t_{i,j}$ are the longitude, latitude and time of $y_{i,j}$
- $\varepsilon_{i,j}$ captures the sensor noise and microscale variation

This model says that year-to-year variations of the oceanographic field can be regarded as i.i.d. realizations from a Gaussian process

In this model, the mean function (or mean field) $m(\cdot)$ is the *climatology* (long-term average of the oceanographic field)

The mean-centered process $f_i(\cdot) - m(\cdot)$ is called the *anomaly* and the covariance function $k(\cdot, \cdot)$ characterizes spatio-temporal dependence in these anomalies

# Mean field for ocean observations

A potential choice for the mean field is:

$$
\begin{aligned}
m(x, y, t) = \beta_0 &+ \beta_1(x - x_0) + \beta_2(y - y_0) \\
&+ \beta_3(x - x_0)(y - y_0) + \beta_4(x - x_0)^2 + \beta_5(y - y_0)^2 \\
&+ \sum_{k=1}^{6} \gamma_k \sin\left(2\pi k \frac{\tau(t)}{365}\right) + \sum_{k=1}^{6} \delta_k \cos\left(2\pi k \frac{\tau(t)}{365}\right) \\
&+ \nu_1(t - t_0) + \nu_2(t - t_0)^2,
\end{aligned}
$$

where $x$ is longitude, $y$ latitude, $t$ is Julian day, $\tau(t)$ is the yearday corresponding to $t$, $(x_0, y_0)$ is the mid-point of the spatial domain and $t_0$ is the Julian day at the mid-point of the analyzed time period

This is inspired by the Roemmich and Gilson (2009) mean fit, except that we add linear and quadratic climatological time trend terms

# Covariance function for ocean observations

The Matérn covariance is often a good choice for physical fields

So a reasonable choice of a covariance function is

$$k(x_1, y_1, t_1, x_2, y_2, t_2) = M_\nu\Big(-d\big((x_1, y_1, t_1), (x_2, y_2, t_2)\big)\Big),$$

where $M_\nu(\cdot)$ is the Matérn kernel with smoothness $\nu$ and

$$d\big((x_1, y_1, t_1), (x_2, y_2, t_2)\big) = \sqrt{\left(\frac{x_1 - x_2}{\theta_x}\right)^2 + \left(\frac{y_1 - y_2}{\theta_y}\right)^2 + \left(\frac{t_1 - t_2}{\theta_t}\right)^2}$$

is an anisotropic space-time distance metric with positive decorrelation scales $\theta_x$, $\theta_y$ and $\theta_t$

For example, $\nu = 1/2$ is often a good choice and gives:

$$k(x_1, y_1, t_1, x_2, y_2, t_2) = \phi \exp\Big(-d\big((x_1, y_1, t_1), (x_2, y_2, t_2)\big)\Big)$$

If the field needs to be differentiable, can use $\nu = 3/2$

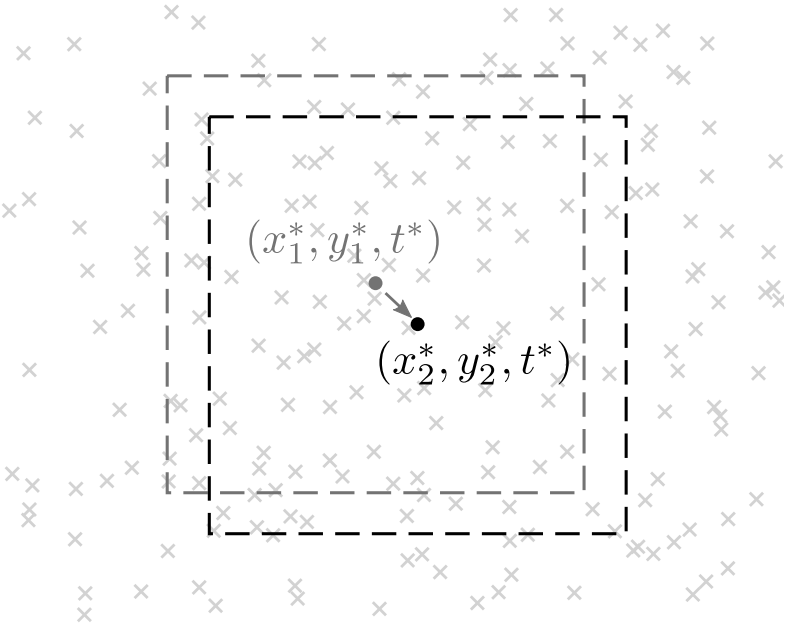
# Locally stationary GP regression

The previous model can work well over a small spatial domain

However, this would be a terrible model for basin-scale or global interpolation

There are two problems:

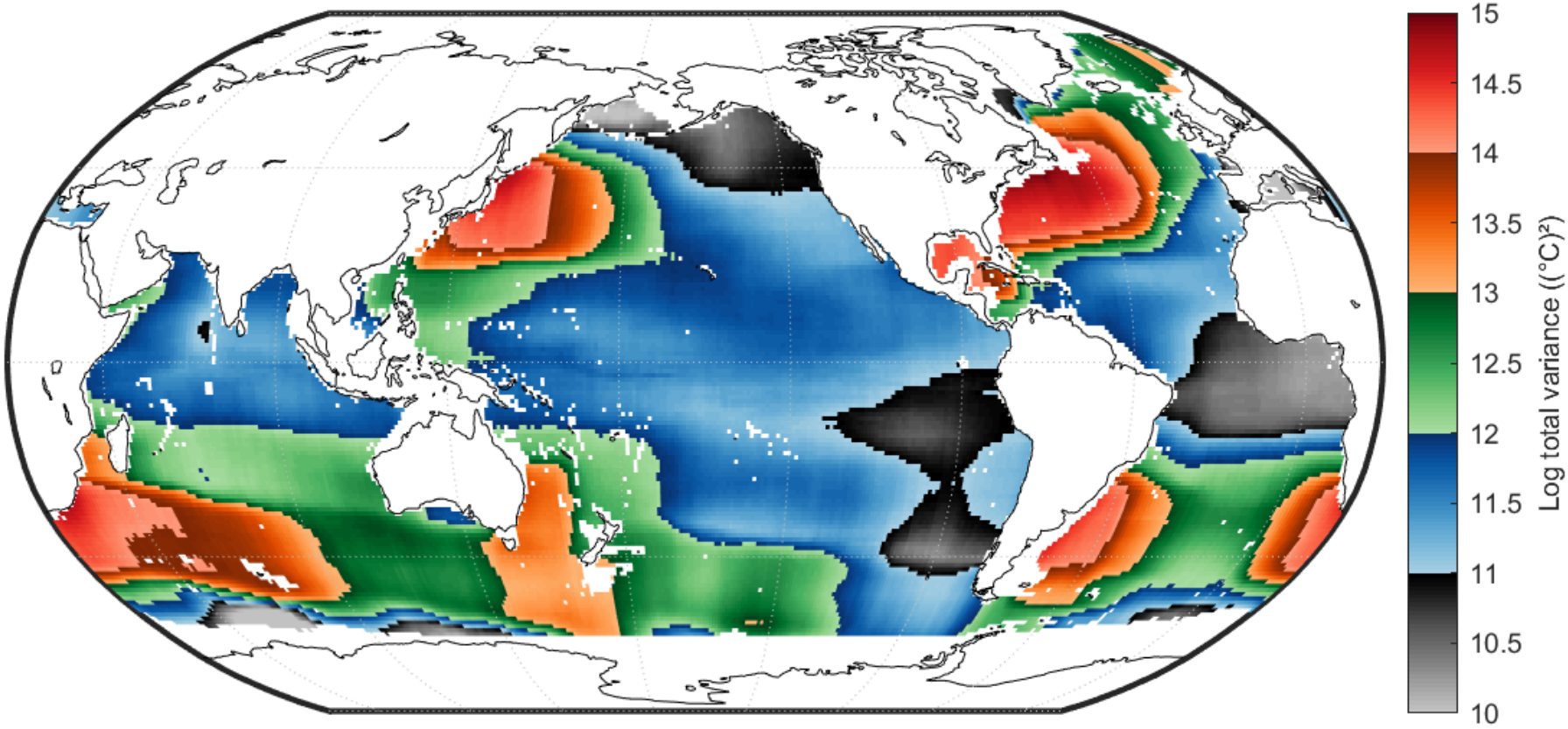
1. The assumed mean model is too simple over large domains
2. Any realistic oceanographic field does not satisfy the stationarity assumption in the covariance function




However, there is an easy fix that we have found to work well for many oceanographic fields: use the previous model only locally within overlapping moving windows! (Kuusela and Stein, 2018)

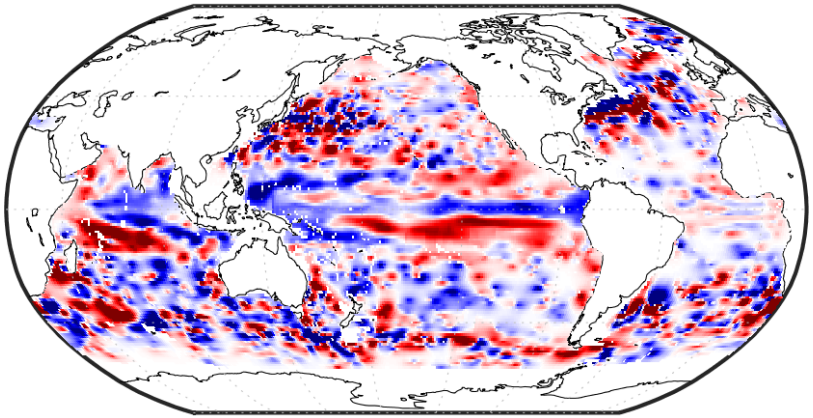
This approach also has major computational benefits since fits are done using only subsets of the data and the computations can be parallelized

# Locally stationary GP regression

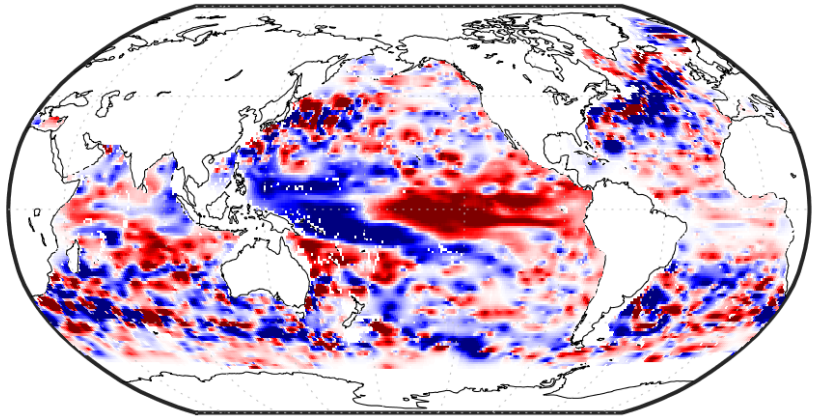



Estimated total variance $\hat{\phi} + \hat{\sigma}^2$ for ocean heat content (15–975 m) from Argo floats in a locally stationary GP model

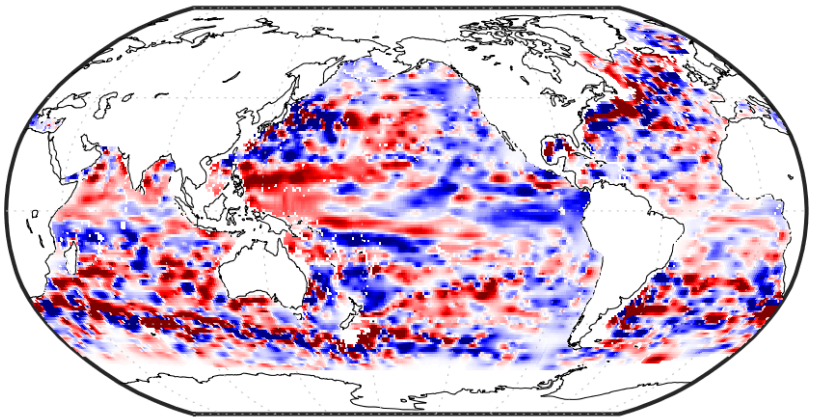
# Ocean heat content anomalies



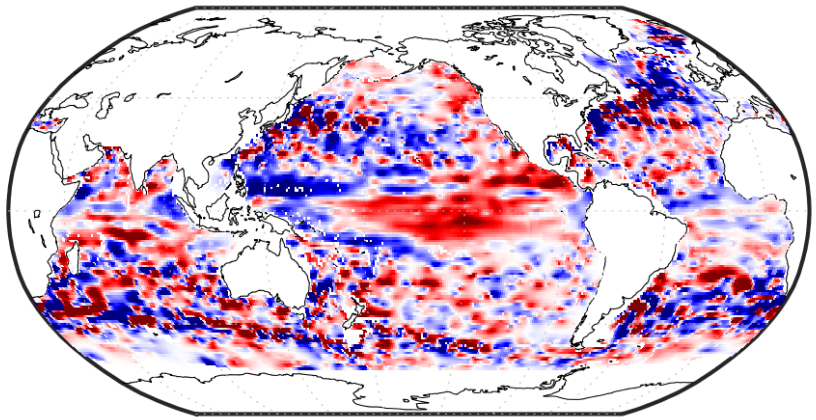
(a) 02/2007



(b) 02/2010



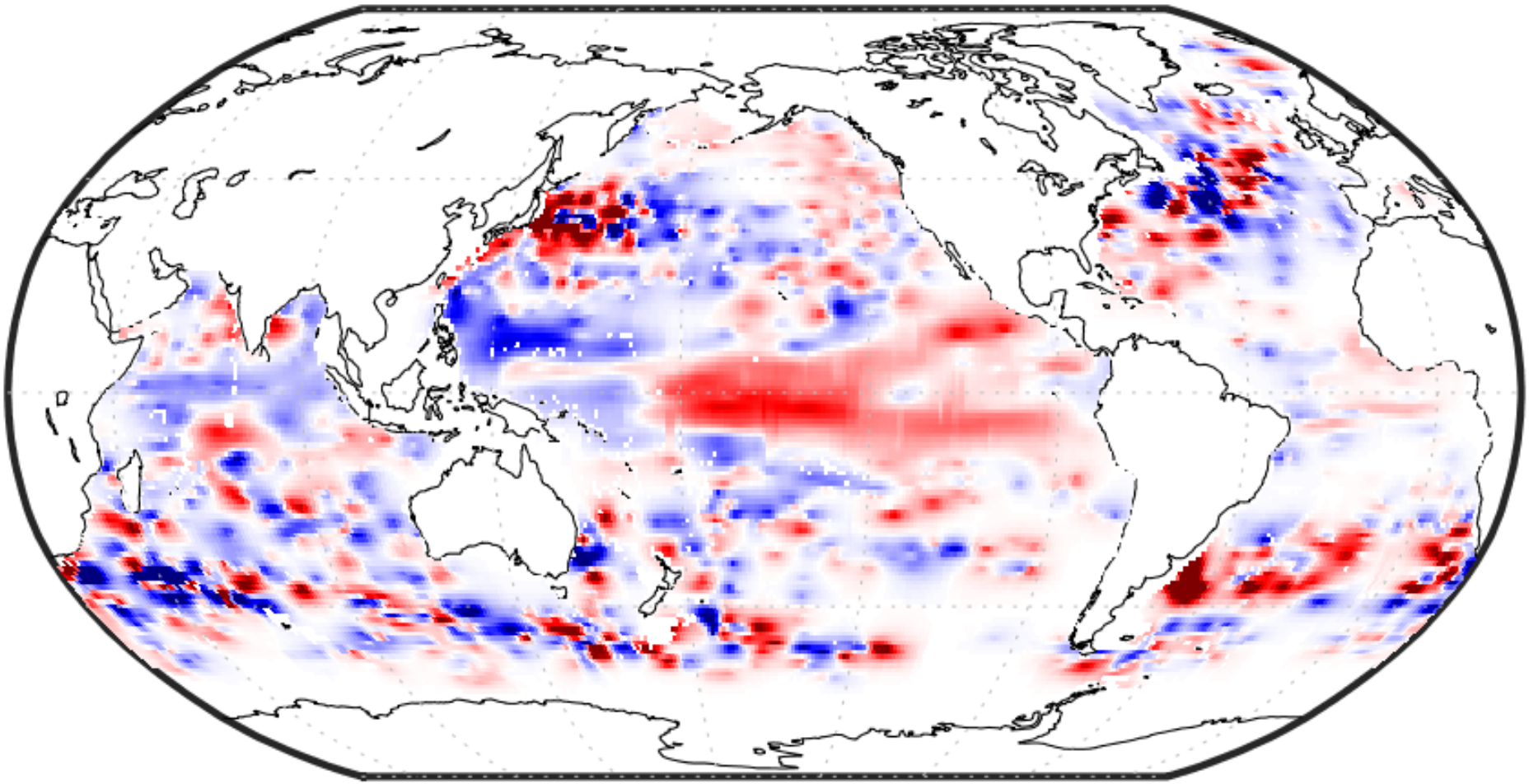
(c) 02/2013



(d) 02/2015

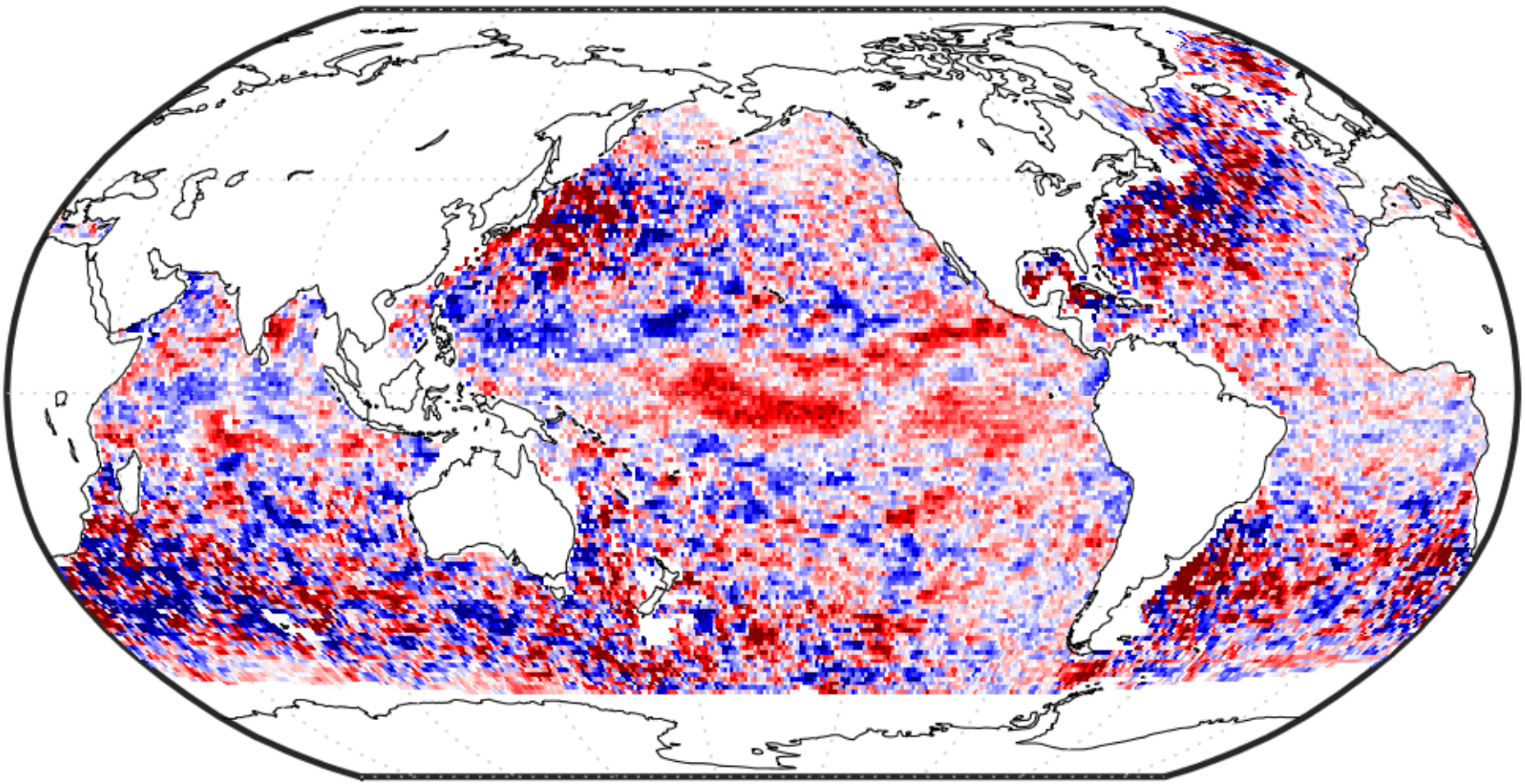
Monthly ocean heat content anomalies (15–975 m) interpolated from Argo float data using a locally stationary Gaussian process

# UQ with local conditional simulations



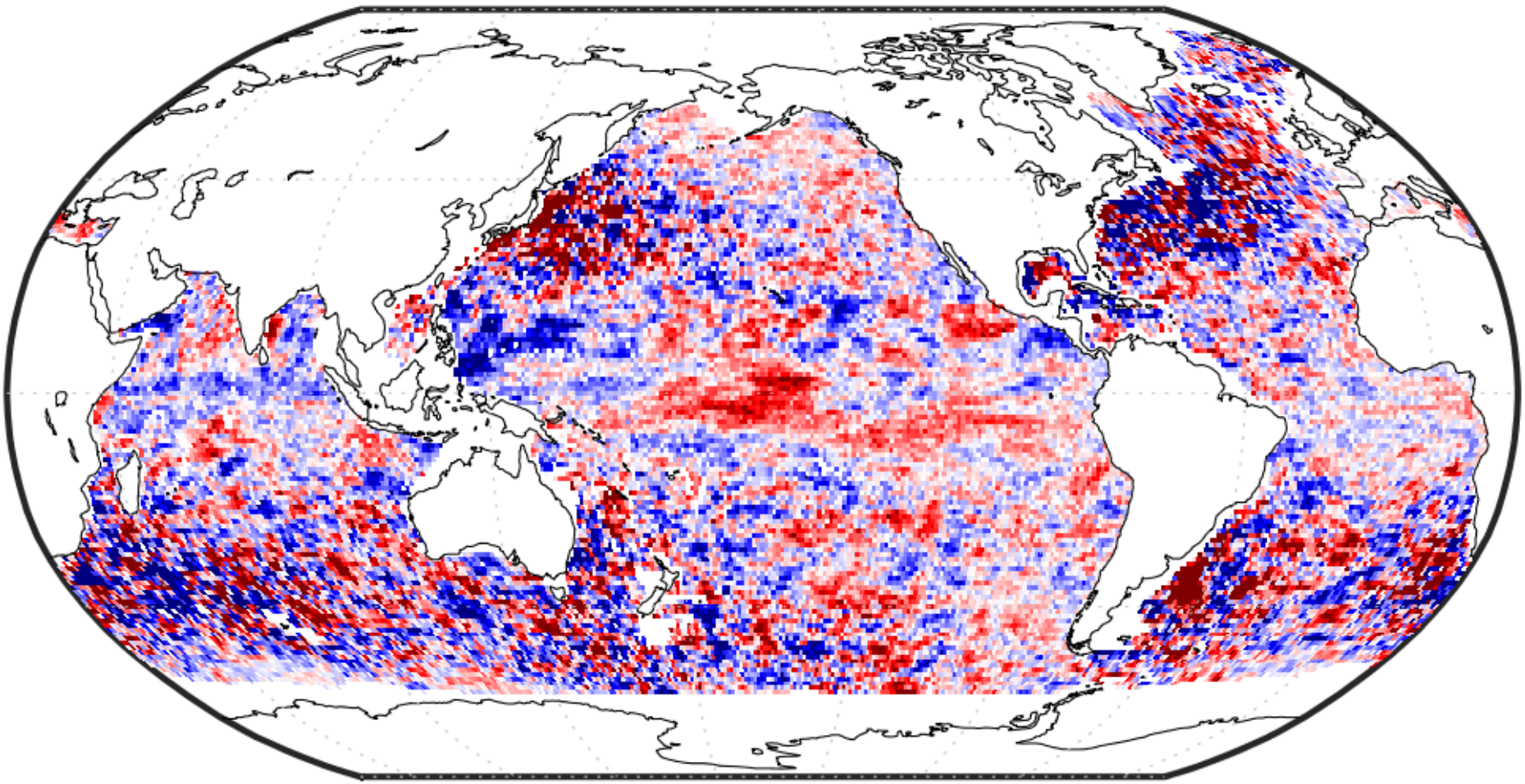
Conditional mean

# UQ with local conditional simulations



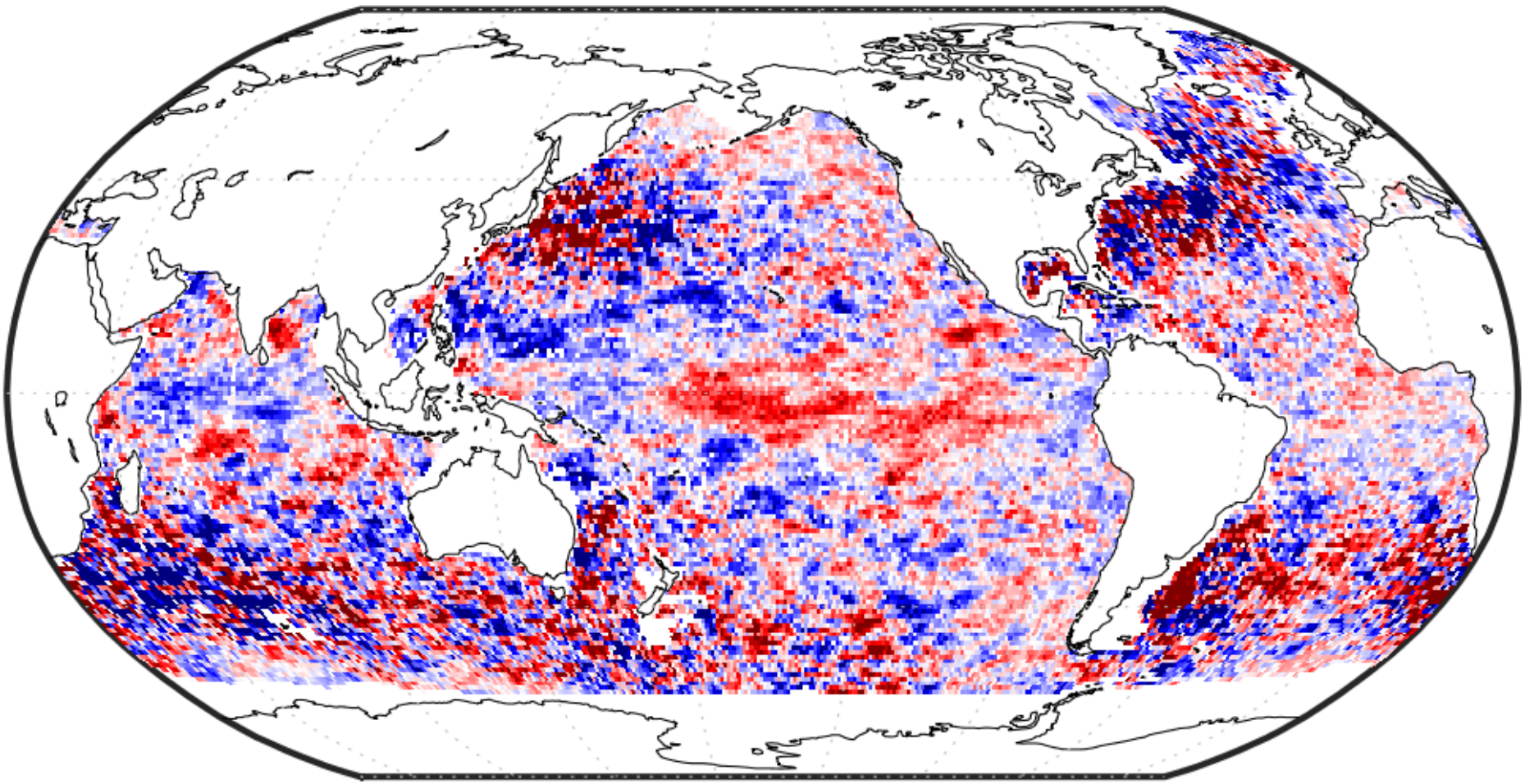
Conditional simulation realization 1

# UQ with local conditional simulations



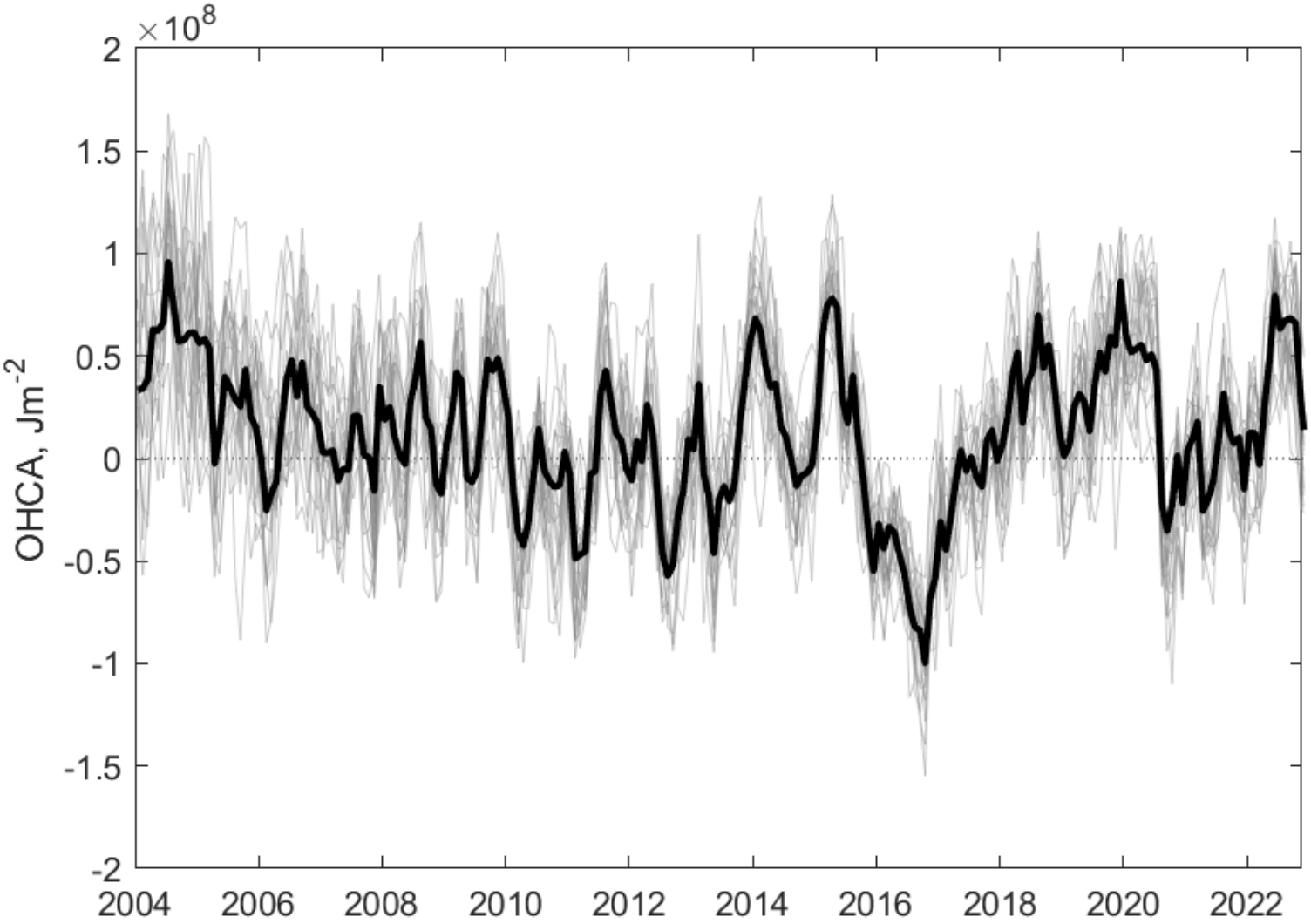
Conditional simulation realization 2

# UQ with local conditional simulations



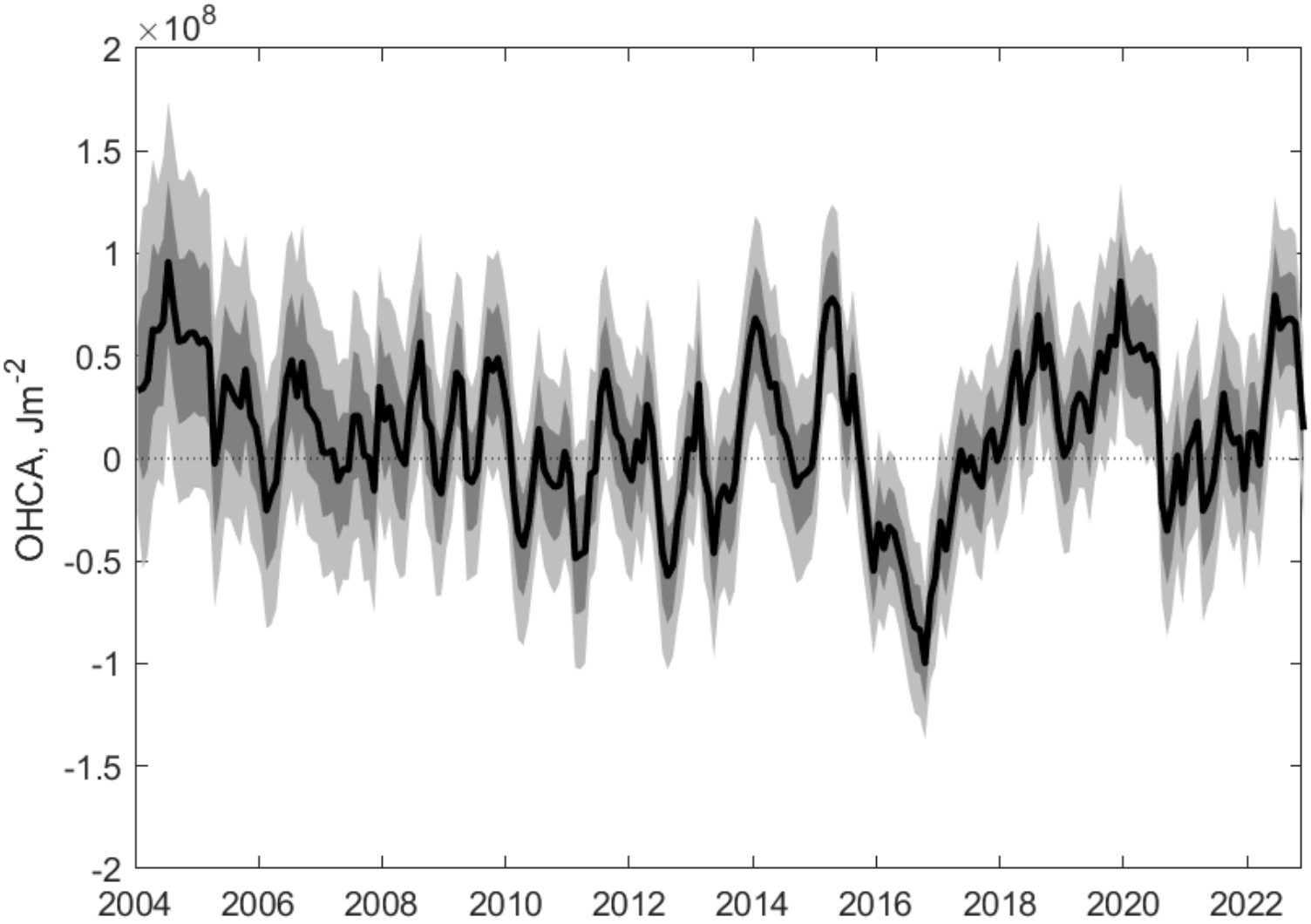
Conditional simulation realization 3

# UQ with local conditional simulations




Conditional mean (black) and 20 conditional simulations (gray) for upper ocean OHC anomalies

# UQ with local conditional simulations




Upper ocean OHC anomalies with 68% (dark gray) and 95% (light gray) uncertainties

# Additional reading

The following textbooks are good starting points for learning more:

- C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006
- M.L. Stein, Interpolation of spatial data: Some theory for kriging, Springer, 1999
- N.A.C. Cressie, Statistics for spatial data, Revised edition, John Wiley & Sons, 1993
- N. Cressie and C. K. Wikle, Statistics for spatio-temporal data, Wiley, 2011
- A.E. Gelfand, P.J. Diggle, M. Fuentes, and P. Guttorp (editors), Handbook of Spatial Statistics, CRC Press, 2010

# References I

M. Kuusela and M. L. Stein. Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proceedings of the Royal Society A*, 474:20180400, 2018.

D. Roemmich and J. Gilson. The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. *Progress in Oceanography*, 82:81–100, 2009.

# Backup