

# A new open-source gridded altimeter product

**Jonathan M. Lilly**

The Planetary Science Institute, Tucson, Arizona

*October 31, 2022*

Created with `{Liminal}` using `{Remark.js}` + `{Markdown}` + `{KaTeX}`

If you are reading this as a pdf, kindly [{click here}](#) for the html version in order to view the embedded animations.

# Overview

Gridded sea level anomalies and associated velocities are among the most important oceanographic data products.

# Overview

Gridded sea level anomalies and associated velocities are among the most important oceanographic data products.

The standard (essentially, only) product is the CMEMS/AVISO product, based on Optimal Interpolation, which is not open source.



# Overview

Gridded sea level anomalies and associated velocities are among the most important oceanographic data products.

The standard (essentially, only) product is the CMEMS/AVISO product, based on Optimal Interpolation, which is not open source.

An extensive review of mapping methods suggests that a different method, local polynomial fitting (a.k.a. loess), may be preferable.

# Overview

Gridded sea level anomalies and associated velocities are among the most important oceanographic data products.

The standard (essentially, only) product is the CMEMS/AVISO product, based on Optimal Interpolation, which is not open source.

An extensive review of mapping methods suggests that a different method, local polynomial fitting (a.k.a. loess), may be preferable.

In order to isolate the mapping problem from multi-satellite merging, here we employ only the 30-year Jason-class record.

# Overview

Gridded sea level anomalies and associated velocities are among the most important oceanographic data products.

The standard (essentially, only) product is the CMEMS/AVISO product, based on Optimal Interpolation, which is not open source.

An extensive review of mapping methods suggests that a different method, local polynomial fitting (a.k.a. loess), may be preferable.

In order to isolate the mapping problem from multi-satellite merging, here we employ only the 30-year Jason-class record.

A sweep through the parameter space of local polynomial fitting with an observing system simulation experiment leads to a best fit.

# Overview

Gridded sea level anomalies and associated velocities are among the most important oceanographic data products.

The standard (essentially, only) product is the CMEMS/AVISO product, based on Optimal Interpolation, which is not open source.

An extensive review of mapping methods suggests that a different method, local polynomial fitting (a.k.a. loess), may be preferable.

In order to isolate the mapping problem from multi-satellite merging, here we employ only the 30-year Jason-class record.

A sweep through the parameter space of local polynomial fitting with an observing system simulation experiment leads to a best fit.

The best fit parameters, when applied to the Jason-class data, yields maps that rival CMEMS despite using a fraction of the data.

# Overview

Gridded sea level anomalies and associated velocities are among the most important oceanographic data products.

The standard (essentially, only) product is the CMEMS/AVISO product, based on Optimal Interpolation, which is not open source.

An extensive review of mapping methods suggests that a different method, local polynomial fitting (a.k.a. loess), may be preferable.

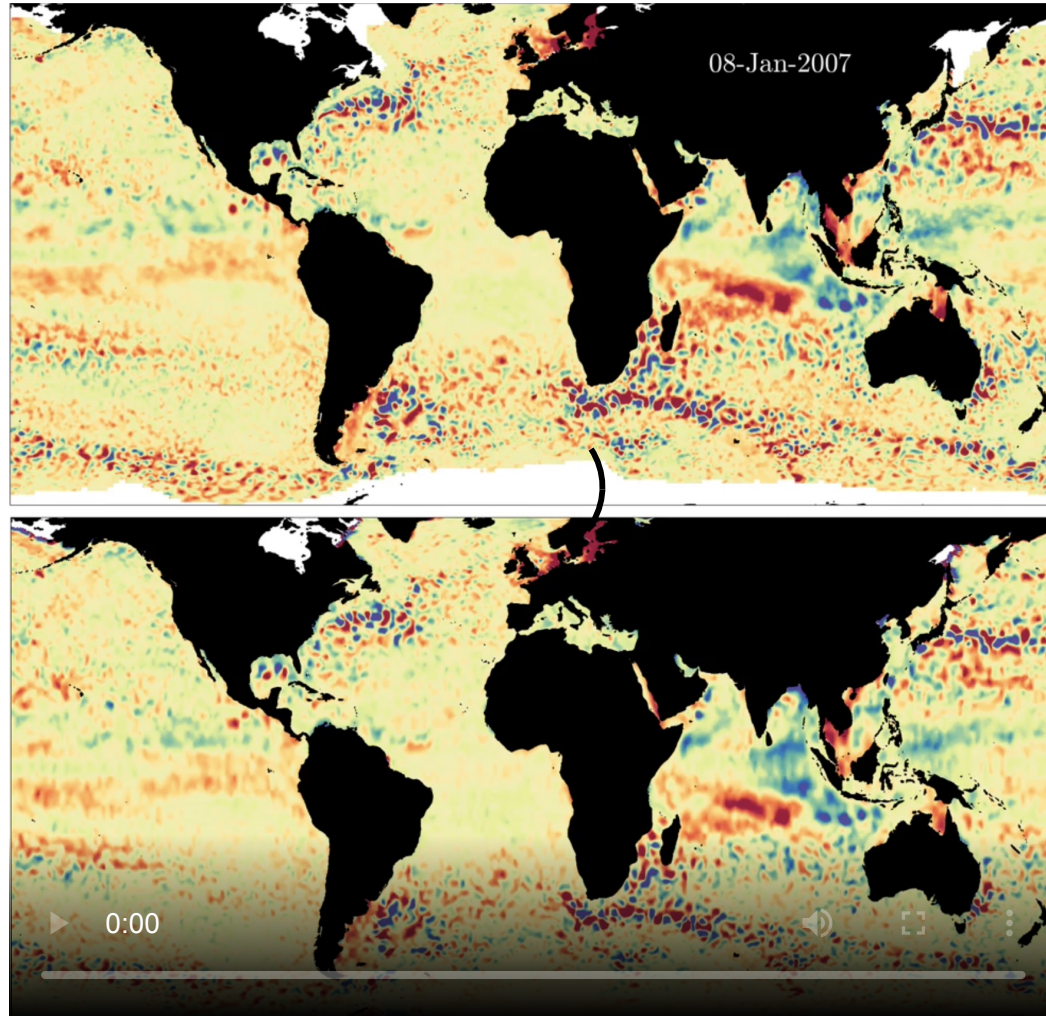
In order to isolate the mapping problem from multi-satellite merging, here we employ only the 30-year Jason-class record.

A sweep through the parameter space of local polynomial fitting with an observing system simulation experiment leads to a best fit.

The best fit parameters, when applied to the Jason-class data, yields maps that rival CMEMS despite using a fraction of the data.

This suggests that the use improved mapping methods could greatly improve altimetric gridded products.

# CMEMS vs. New Jason-Only Product



SLA for 2007. CMEMS (top) vs. new product (bottom).

# Comparison of the Two Products

The two animations in the previous slide appear quite comparable.

Yet whereas the upper panel incorporates Jason-class, ERS-class, Geosat-class, and other missions such as Saral and Cryosat, the lower panel uses Jason-class only.

In other words, a refinement in the mapping algorithm has led to a gridded product comparable to that of CMEMS despite using a fraction of the data.

This suggests that the current Optimal Interpolation algorithm is suboptimal and is substantially oversmoothing the data.

It also suggests that incorporating all available altimeter data into a refined mapping method would lead to greatly improved products.

This however is a large task that would require a dedicated effort.

(Note: close inspection shows occasional “stripes” in the lower panel due to long-wavelength error; this should be seen as an upstream data processing issue rather than a mapping issue.)

# Local Polynomial Fitting

The idea of local polynomial fitting is to minimize, within the vicinity of some mapping point  $(x_o, y_o)$ , the weighted squared error  $\Delta_p^2(x_o, y_o)$  between the observations and a  $p$ th-order polynomial.

The  $N$  observations are arranged into  $N$ -vectors, with  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{y}}$  being the  $x$ - and  $y$ -locations, and  $\tilde{\mathbf{z}}$  being the observed values.

Then the weighted squared error in the vicinity of point  $(x_o, y_o)$  is

$$\Delta_p^2(x_o, y_o) \equiv \left\| \mathbf{W}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, x_o, y_o) \left[ \tilde{\mathbf{z}} - \mathbf{X}_p(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, x_o, y_o) \widehat{\boldsymbol{\beta}}_p(x_o, y_o) \right] \right\|^2$$

where  $\mathbf{W}$  is a matrix of weights,  $\mathbf{X}_p$  is a matrix of coordinate deviations containing powers up to  $p$ , and  $\widehat{\boldsymbol{\beta}}_p$  is the estimated field and its first  $p$  derivatives.



# The Estimate Vector $\widehat{\beta}_p$

The vector  $\beta_p$  contains the mean field  $z$  and its first  $p$  derivatives.

$$\beta_0(x_o, y_o) \equiv [z(x_o, y_o)]$$

$$\beta_1(x_o, y_o) \equiv \begin{bmatrix} z(x_o, y_o) \\ \frac{\partial z}{\partial x}(x_o, y_o) \\ \frac{\partial z}{\partial y}(x_o, y_o) \end{bmatrix}$$

$$\beta_2(x_o, y_o) \equiv \begin{bmatrix} z(x_o, y_o) \\ \frac{\partial z}{\partial x}(x_o, y_o) \\ \frac{\partial z}{\partial y}(x_o, y_o) \\ \frac{\partial^2 z}{\partial x^2}(x_o, y_o) \\ \frac{\partial^2 z}{\partial x \partial y}(x_o, y_o) \\ \frac{\partial^2 z}{\partial y^2}(x_o, y_o) \end{bmatrix}$$

which is of length  $Q$ , with  $Q = 1, 3$ , and  $6$  for  $p = 0, 1$ , and  $2$ . The estimate vector  $\widehat{\beta}_p$  is then our error-minimizing estimate of  $\beta_p$ .

# The Deviation Matrix $\mathbf{X}_p$

The matrix  $\mathbf{X}_p$  contains the coordinate deviations necessary for a Taylor expansion up to  $p$ th order.

$$\mathbf{X}_0(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, x_o, y_o) \equiv \left\{ \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\} N \text{ terms}$$

$$\mathbf{X}_1(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, x_o, y_o) \equiv \begin{bmatrix} 1 & \tilde{x}_1 - x_o & \tilde{y}_1 - y_o \\ 1 & \tilde{x}_2 - x_o & \tilde{y}_2 - y_o \\ \vdots & \vdots & \vdots \\ 1 & \tilde{x}_N - x_o & \tilde{y}_N - y_o \end{bmatrix}$$

$$\mathbf{X}_2(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, x_o, y_o) \equiv \begin{bmatrix} 1 & \tilde{x}'_1 & \tilde{y}'_1 & \frac{1}{2} \tilde{x}'_1{}^2 & \tilde{x}'_1 \tilde{y}'_1 & \frac{1}{2} \tilde{y}'_1{}^2 \\ 1 & \tilde{x}'_2 & \tilde{y}'_2 & \frac{1}{2} \tilde{x}'_2{}^2 & \tilde{x}'_2 \tilde{y}'_2 & \frac{1}{2} \tilde{y}'_2{}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \tilde{x}'_N & \tilde{y}'_N & \frac{1}{2} \tilde{x}'_N{}^2 & \tilde{x}'_N \tilde{y}'_N & \frac{1}{2} \tilde{y}'_N{}^2 \end{bmatrix}$$

In the last expression, we define  $\tilde{x}'_n \equiv \tilde{x}_n - x_o$  and  $\tilde{y}'_n \equiv \tilde{y}_n - y_o$ .

# The Weighting Matrix $\mathbf{W}$

Finally, the weighting matrix  $\mathbf{W}$  is defined as

$$\mathbf{W}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, x_o, y_o) \equiv \begin{bmatrix} K_h(\tilde{r}_1) & 0 & \cdots & 0 \\ 0 & K_h(\tilde{r}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K_h(\tilde{r}_N) \end{bmatrix}$$

where  $\tilde{r}_n \equiv \sqrt{(\tilde{x}_n - x_o)^2 + (\tilde{y}_n - y_o)^2}$  is the distance between the mapping point  $(x_o, y_o)$  and the  $n$ th observation point  $(x_n, y_n)$ .

Here  $K_h(r) \equiv K(r/h)$  where  $K(r)$  is a symmetric, decaying function called the kernel. Generally  $K(r)$  is set to vanish outside of a prescribed radius, chosen here to be  $r = 1$ .

The rescaled kernel  $K_h(r)$  then vanishes outside of radius  $r = h$ , where  $h$  is known as the *bandwidth*.

# The Error-Minimizing Solution

The problem of finding the local polynomial that minimizes the weighted error

$$\Delta_p^2(x_o, y_o) \equiv \left\| \mathbf{W}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, x_o, y_o) \left[ \tilde{\mathbf{z}} - \mathbf{X}_p(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, x_o, y_o) \widehat{\boldsymbol{\beta}}_p(x_o, y_o) \right] \right\|^2$$

can be shown to have the solution

$$\widehat{\boldsymbol{\beta}}_p(x_o, y_o) = (\mathbf{X}_p^T \mathbf{W} \mathbf{X}_p)^{-1} \mathbf{X}_p^T \mathbf{W} \tilde{\mathbf{z}}$$

the heart of which involves inverting the  $Q \times Q$  matrix  $\mathbf{X}_p^T \mathbf{W} \mathbf{X}_p$ .

In the case that  $p = 0$ , a linear fit, it can be shown that this reduces to a simple kernel smoothing.

The main free parameters of local polynomial fitting are thus the fit order  $p$ , the bandwidth  $h$ , and the choice of kernel  $K(r)$ .

Anisotropic smoothing, involving a matrix-valued bandwidth parameter, can also be employed, but will not be considered here.

# Choice of Kernel

A variety of different choices of weighting kernel  $K(r)$  have been employed in the literature.

Here we propose the general form

$$K(r; \alpha, \beta) \equiv \kappa_{\alpha, \beta} (1 - r^\alpha)^\beta$$

where  $\kappa_{\alpha, \beta}$  is a normalizing constant. This broad family subsumes the most commonly used kernels as special cases.

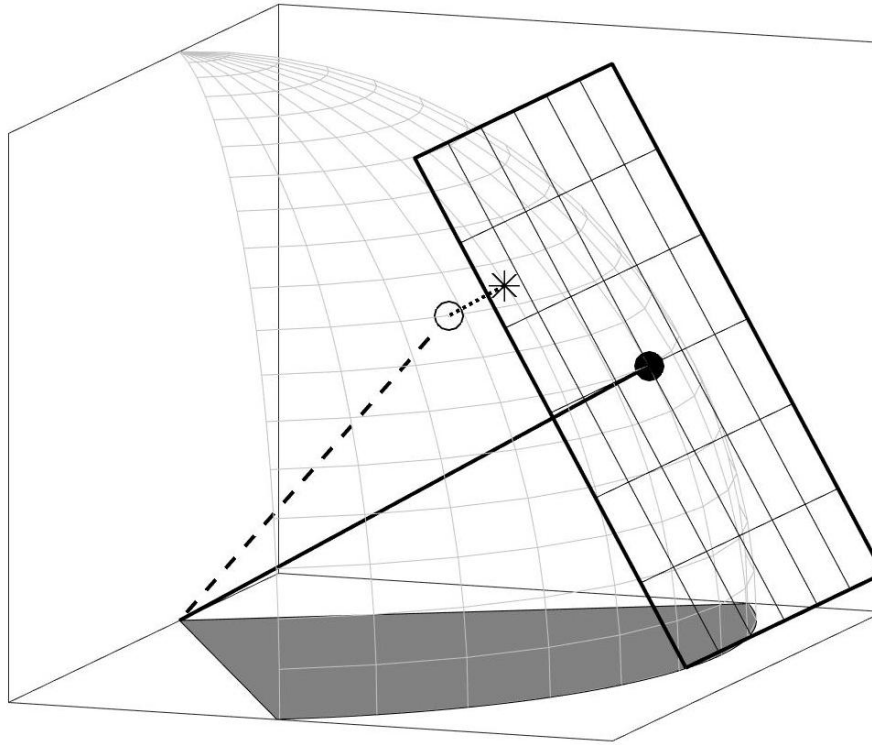
The use of a general form allows us to readily determine the optimal choice of parameters in an observing system simulation experiment through numerical optimization.

This kernel is usefully reparameterized in terms of the *half-power radius*  $R$  and *shape parameter*  $\alpha$ , with  $\beta$  then given by

$$\beta(R, \alpha) \equiv \frac{\ln(1/2)}{\ln(1 - R^\alpha)}$$

where we note that  $K(R; \alpha, \beta)/K(0; \alpha, \beta) = 1/2$ .

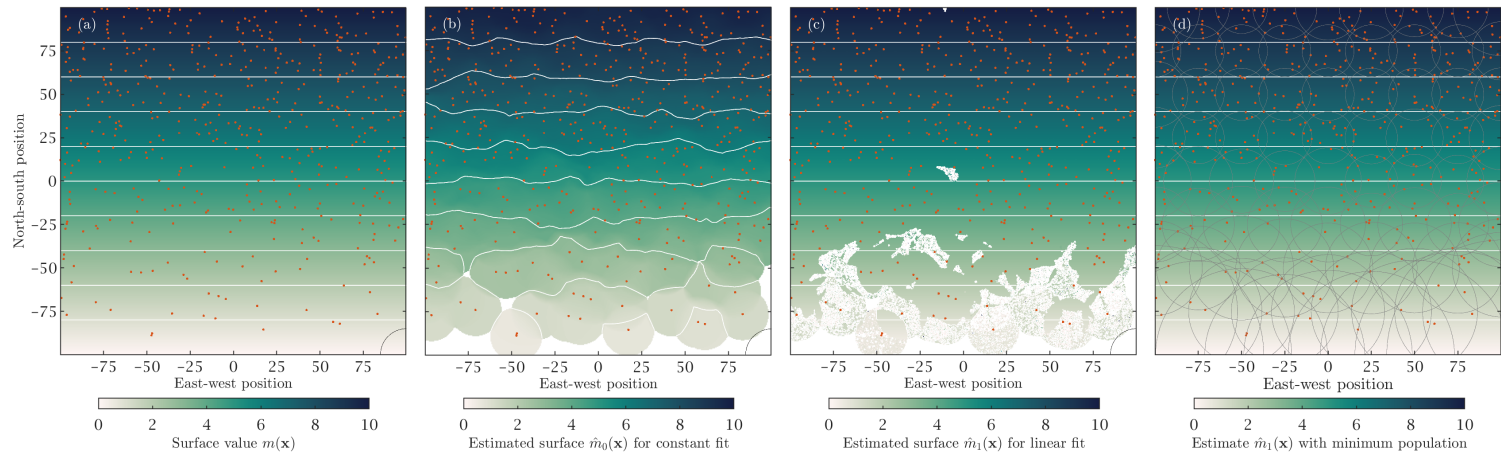
# Local Polynomial Fitting on the Sphere



Local polynomial fitting on the sphere can be accomplished by projecting data into a plane tangent to the mapping point  $(x_o, y_o)$ .

Then the previously presented equations all hold, with  $\tilde{\mathbf{x}}(x_o, y_o)$  and  $\tilde{\mathbf{y}}(x_o, y_o)$  now giving the location within the tangent plane.

# Design Adaptivity and Fixed Population

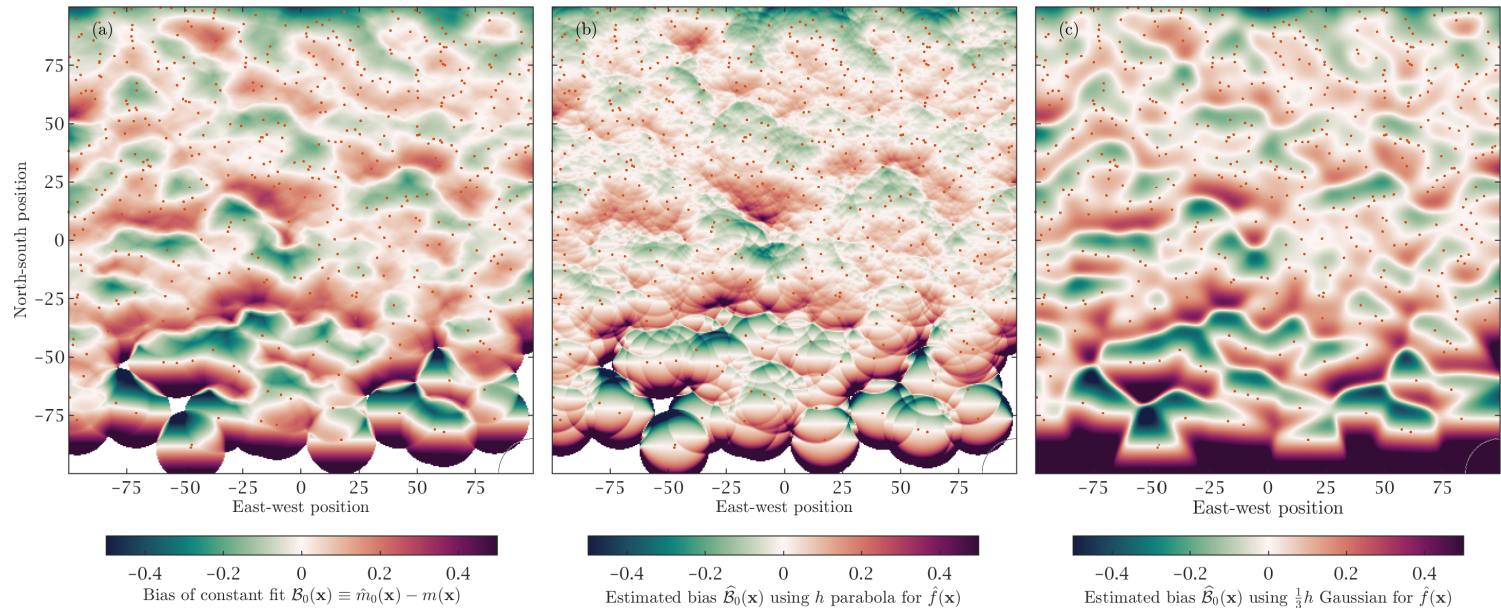


Fits with  $p = 1$  or larger have a crucial property in that bias does not arise due to spatial variations in the distribution of data points.

This property, called *design adaptivity* by Fan (1992), corrects a severe defect in the zeroth-order fit. Compare (a) the true field, (b) the zeroth-order or  $p = 0$  fit, and (c) the first-order or  $p = 1$  fit.

The problem of gaps arising from low data density with a fixed bandwidth, as in (c), is solved by letting the bandwidth locally expand to encompass a specified number of observation points, as in panel (d). This is termed the *fixed population* algorithm herein.

# Bias Estimates



The existence of extensive theory surrounding local polynomial fitting allows errors to be estimated with reasonable accuracy.

Here, the true mapping bias (a) is compared with two different estimates that do not rely on the true (generally unknown) field.



# Comparison with Optimal Interpolation

In the oceanographic community, Optimal Interpolation (OI) is by far the most common mapping method.

Local polynomial fitting (LPF) and OI differ in both theory and practice. Factors in favor of LPF are numerous, including:

- LPF does not require prior knowledge of the covariance structure of the field being mapped.
- Optimal LPF parameters can be chosen or estimated empirically using theoretical results.
- LPF has several degrees of adjustability, whereas OI, strictly speaking, has zero.
- LPF is much faster because the matrix being inverted is vastly smaller— $Q \times Q$ , i.e.  $6 \times 6$  for a  $p = 2$  fit, versus  $N \times N$  for OI.
- Although OI is theoretically optimal in some sense, key caveats are that (i) real-world performance may depart from optimality on paper and (ii) the sense of optimality is crucial.

For example, OI applied to an eddying ocean is optimal if one wishes to smooth out the eddies to estimate the mean, yet paradoxically is commonly applied to estimate eddy properties.

# Creation of a New Gridded Product

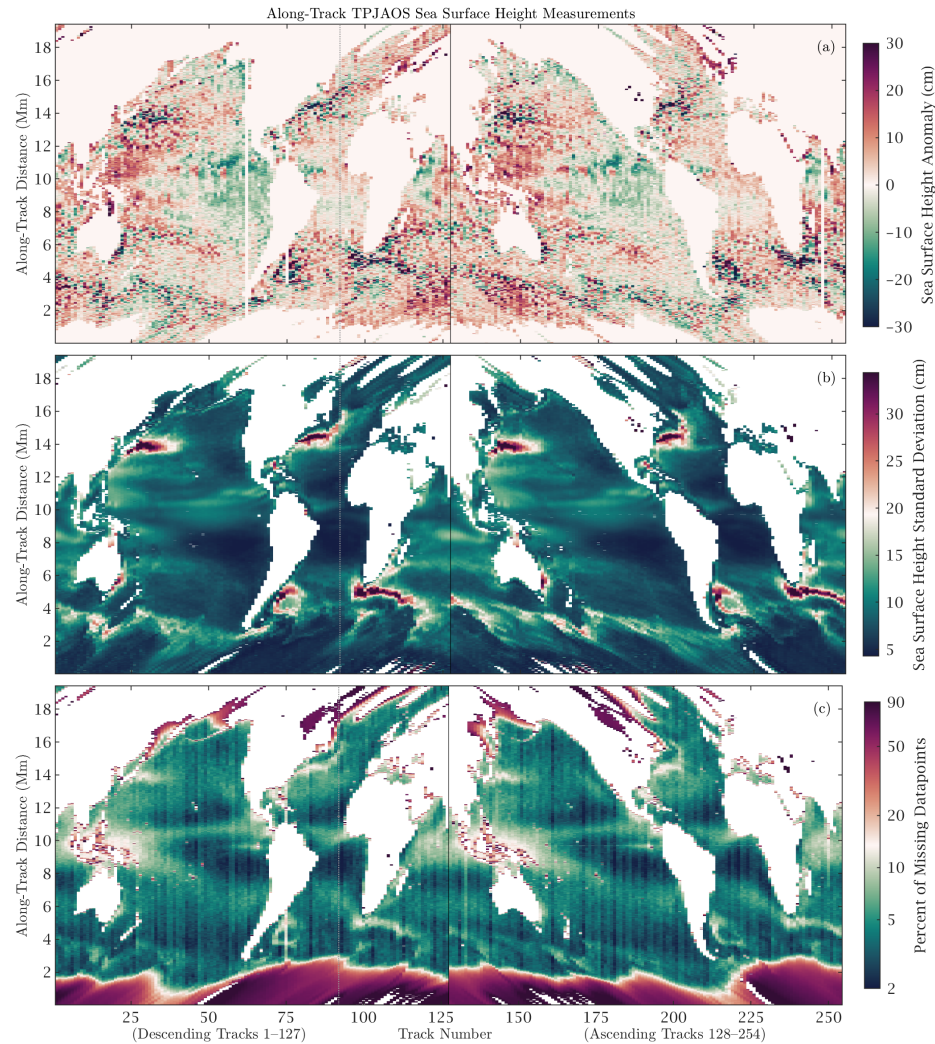
We now turn to the creation of an new open-source gridded product using local polynomial fitting.

For expediency, in order to isolate the mapping aspects of the problem and avoid the great challenge of merging different classes of altimeter data, here we focus entirely on the Jason-class data.

The Jason-class missions have sampled the same tracks continuously since 1992.

To determine the optimal mapping parameters, an observing system simulation experiment will be performed using a synthetic alongtrack dataset constructed from a global numerical model.

# Alongtrack Jason-Class Data



(a) SLA snapshot (b) standard deviation (c) % missing cycles

# About the Data

For alongtrack data, we use the {Integrated Multi-Mission Ocean Altimeter Data for Climate Research Version 5.0} of Beckley et al.

This dataset has been re-organized such that descending and ascending tracks are separated, as shown on the previous slide, and subjected to some additional despiking.

An advantage of this re-organization is that one can simply plot the alongtrack data, with track number along the x-axis and alongtrack distance along the y-axis, and still make sense of it.

We will also form an empirical estimate of the small-scale noise, to be used later in an observing system simulation experiment.

This noise level estimate is formed by a high-wavenumber alongtrack bandpass using a wavelet transform.

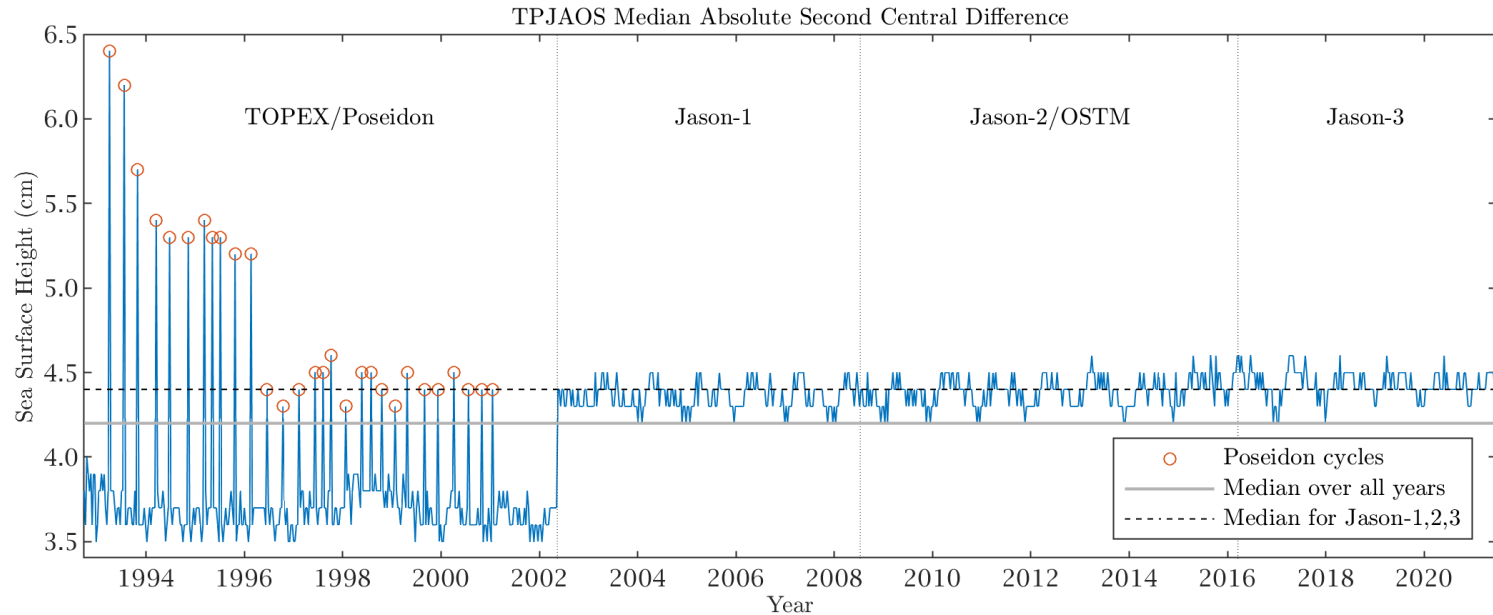
A very high wavenumber band shows atmospheric structure, and can consequently be taken as a noise estimate, while an adjacent lower band shows ocean structure, as shown shortly.

# Despiking Illustration



Selected tracks from the Beckley product before (a) and after (b) median-based despiking.

# Change in Noise Level Between Missions



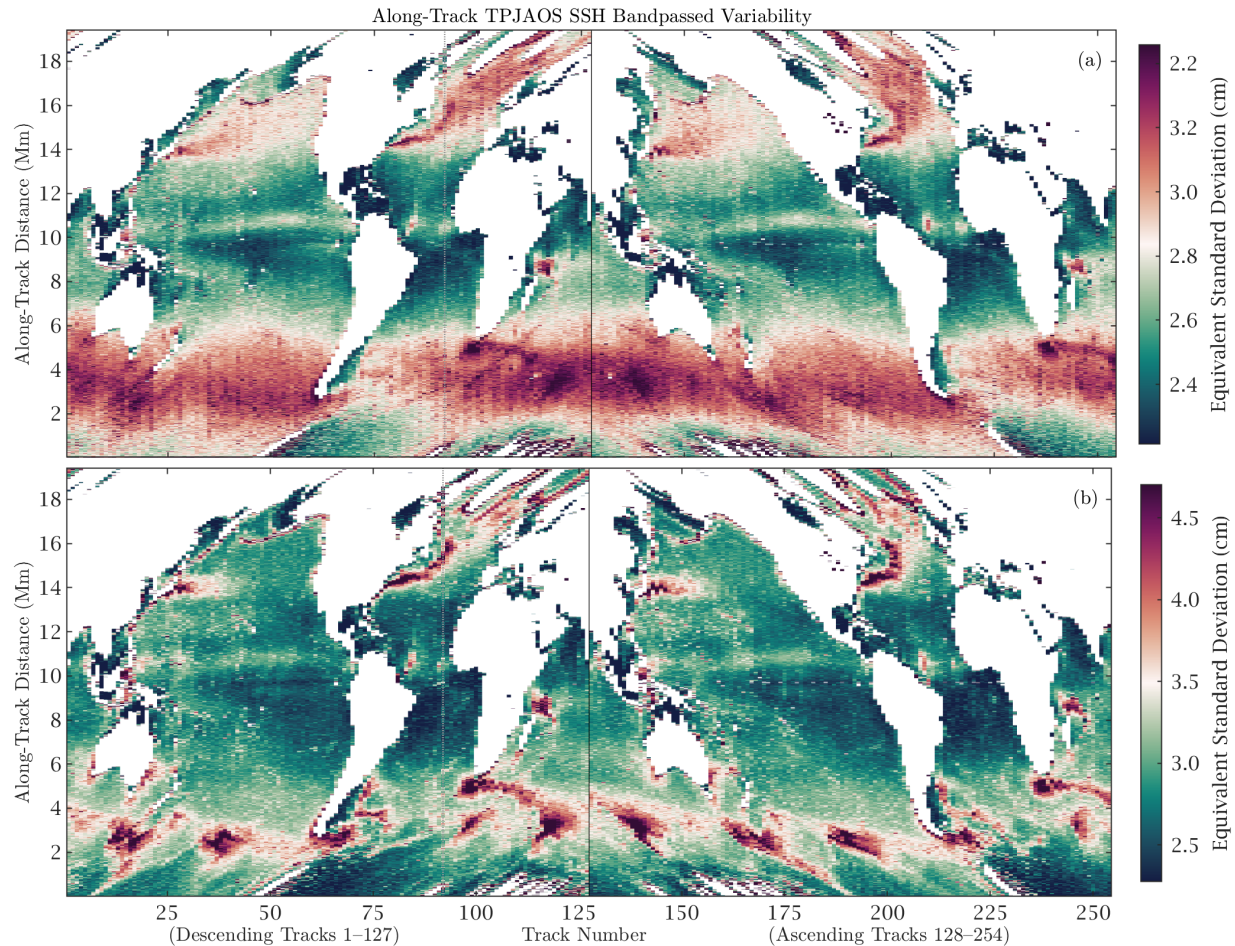
The median value of the absolute second central difference represents an estimate of the noise level, as it identifies the very smallest-scale variability in the absence of other structure.

Divisions between the TOPEX and Poseidon instruments are clear, as is the transition from TOPEX/Poseidon to the Jason satellites.

Nevertheless the time variability of the noise level will be ignored in a first approximation.



# An Empirical Noise Estimate



Alongtrack variability in (a) a very high band or (b) a somewhat lower band. We can use (a) as an empirical noise estimate.

# Creating a Synthetic Alongtrack Dataset

To determine the best mapping parameters, we perform an observing system simulation experiment.

A synthetic alongtrack dataset is created by looking up sea surface height within an ocean model at the exact sample times and locations occurring in the data. Spatially-varying noise is then added to match the empirically estimated noise in the real data.

For a model, we use a year-long,  $1/8^{\text{th}}$  degree simulation with the GOLD model performed by H. Simmons (Simmons and Alford, 2012) and generously made available for this study.

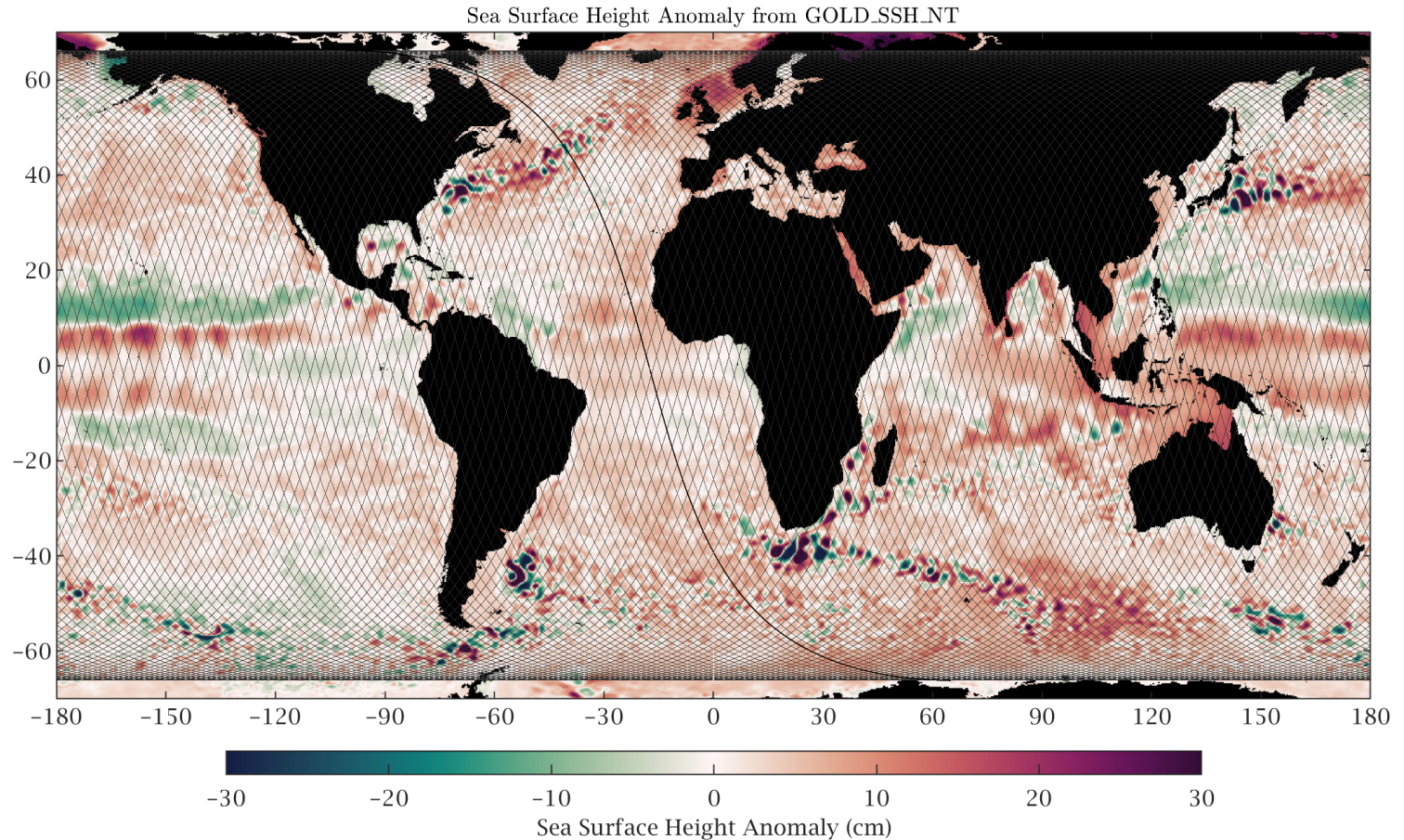
This simulation has the advantage of its surface fields having been saved at the full model resolution every hour.

This means that alongtrack spatial and temporal structure are both well resolved, leading to a high-quality simulated product.

Because tides are removed from the real alongtrack altimeter data in upstream processing, we use a version of the GOLD simulation that does not include tidal forcing.



# A Snapshot from the GOLD Model



A snapshot of sea surface height anomaly from the GOLD model, with Jason-class altimeter tracks overlaid and with the longest continuous ocean-only track shown as a heavy line.

# Determining the Best Mapping Parameters

Next we sweep through parameter space in order find the best set of parameters.

Parameters are allowed to vary as a function of latitude.

Error is quantified by the averaged squared deviation between the resulting map and the colocated true model value.

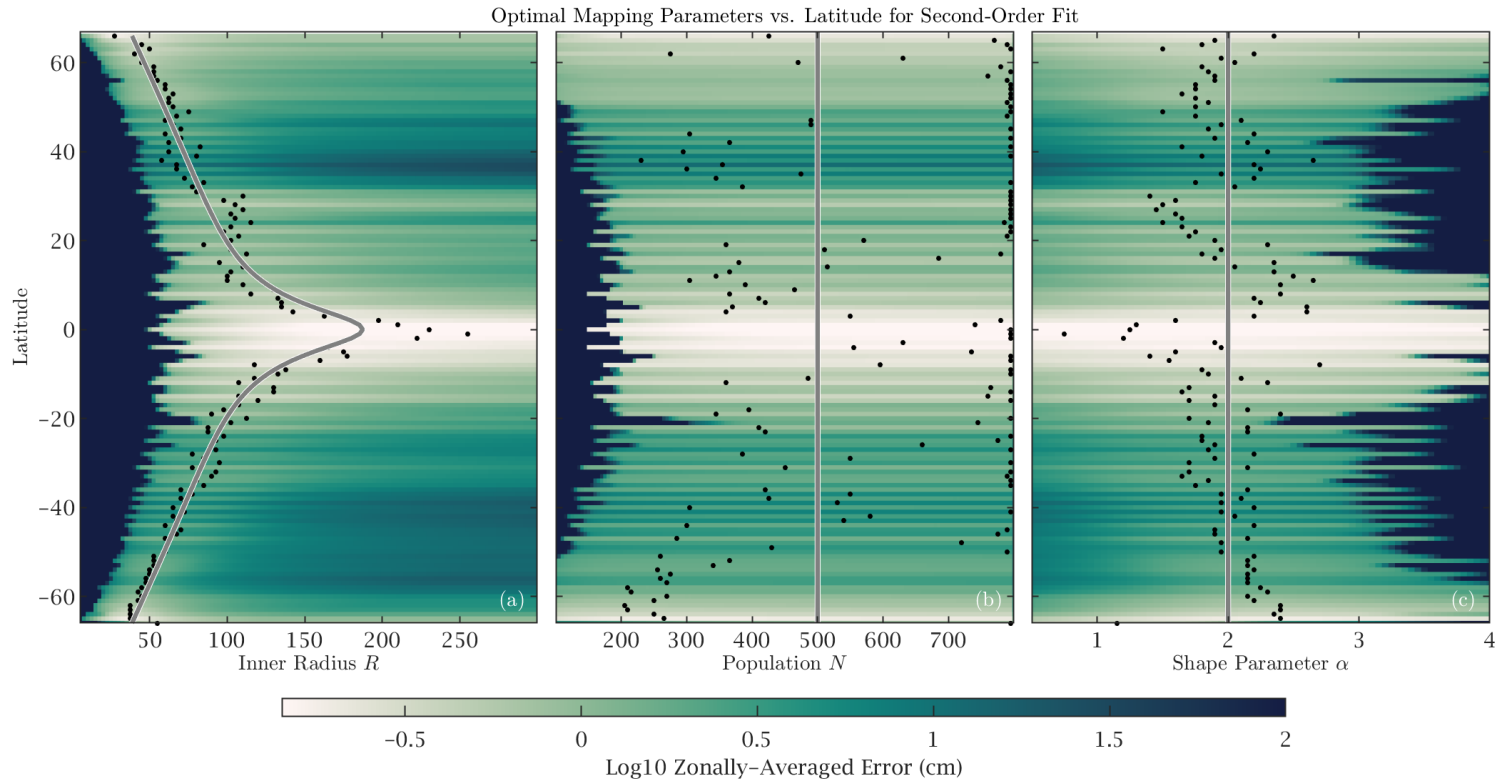
If any gaps emerge in the mapped product between  $\pm 66$  degrees for a certain parameter choice, the error is set to infinity.

After this sweep, a simple parametric form is chosen if the error-minimizing choice is observed to vary with latitude, and its parameters are estimate through numerical optimization.

Otherwise, the best constant value across latitudes is determined.

This is done within the context of fixed population fits, as these naturally adapt to data gaps.

# Optimal Latitude-Dependent Parameters



An exhaustive sweep through (a) half-power radius  $R$ , (b) population  $N$ , and (c) shape parameter  $\alpha$ . Importantly, the former exhibits clear latitude dependence, but the other variables do not.

The heavy gray curves are the best fits as a function of latitude.

# A Beta Version of a New Gridded Product

Using these best-fit curves, maps were created from the real-world alongtrack data for every altimeter cycle during the single year 2007, which is the same year for which the model was run.

This leads to the animation shown in the bottom panel earlier.

The favorable comparison with CMEMS, despite using only a fraction of the data, suggests considerable improvements in the gridded products could be realized through refining our mapping algorithms.

In other words, applying the local polynomial fitting method to all available altimeter data should result in greatly improved maps.

## Next Up

The work presented in this talk is formulated in a series of drafts, which are expected to be submitted for publication in early 2023.

The new open-source gridded product for the entire 30-year Jason record will be made publicly available at that time.

The mapping code upon which this is based is already available as a part of {jLab}, the author's Matlab toolbox.

This software is being substantially revised and refactored in conjunction with the forthcoming publications.

# Thank you!

For all submitted papers as well as my software toolbox, kindly  
visit my {website}.