ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

October 2024

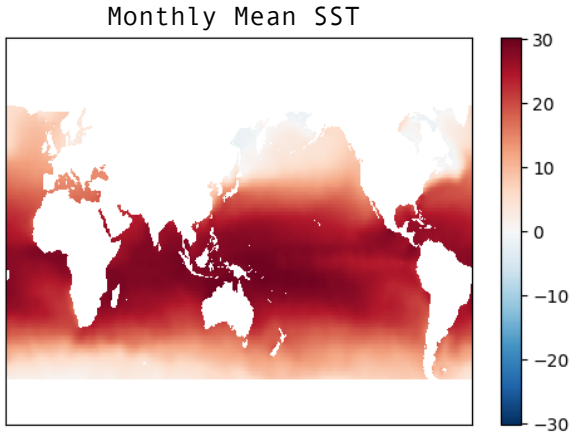
ObsFit

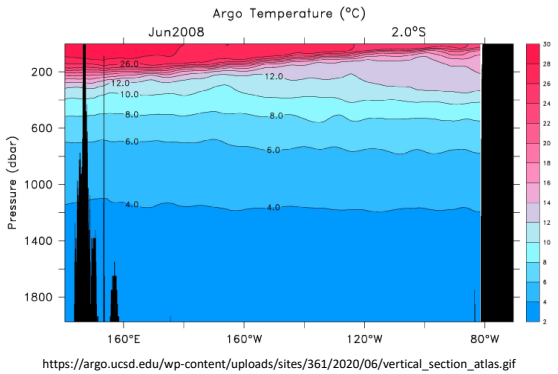
A new package for model-observations comparisons in the MITgcm

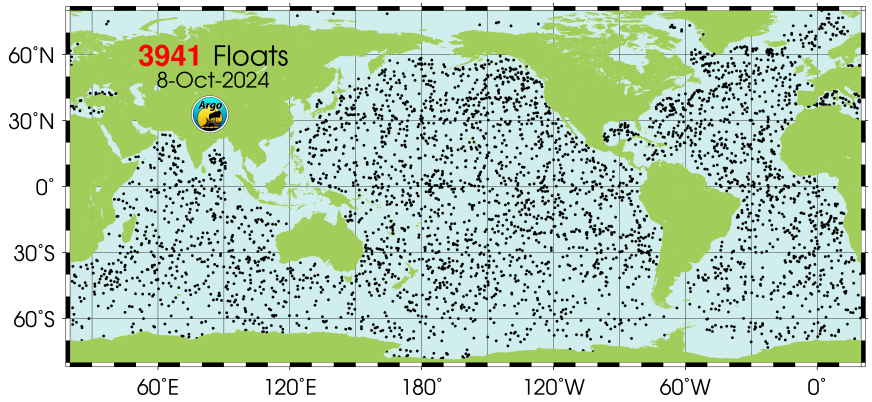
Shoshi Reich
PhD Student, The University of Texas at Austin

MITgcm Data Assimilation Packages

- ECCO: model-data comparisons using gridded datasets
- PROFILES: model-data comparisons at observed locations using vertical columns of data







ObsFit Overview

- A generalization of the profiles package
- Observations do not need to be on a regular grid or a fixed set of depths
  - No “empty” depth levels or horizontal grid points
- Observations can be made of multiple samples that are averaged spatially and/or temporally
- ObsFit samples the model at the time and location of observations, calculates the cost, and produces a model-equivalent output file that is directly comparable to the input file.

[https://github.com/averdy/obsfit_mitgcm]

Example Use Case: SWOT SSH

Assimilating one SWOT swath

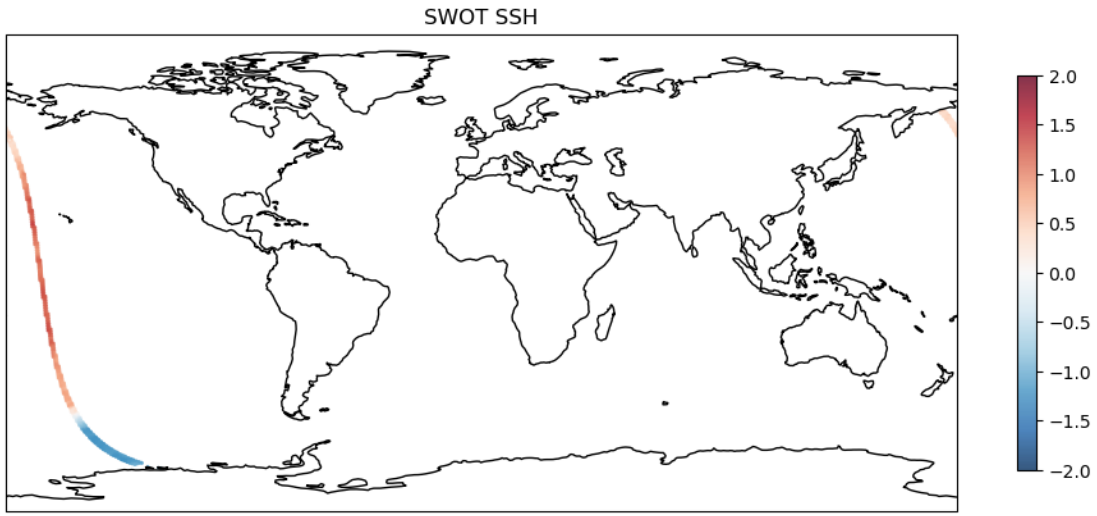
- Spatially sparse
- One hour of data
- LLC270

Using pkg/ecco:

- Need global grid for each hour of the year
- Input file size = 66.4 GB

Using pkg/obsfit

- Average data within each grid cell
- Input file size = 412 MB



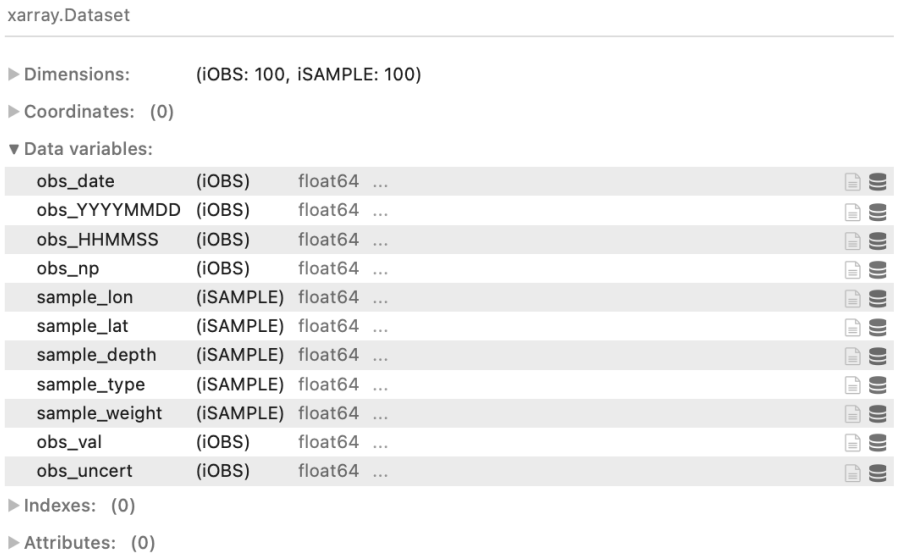
Pre-Processing Data for Spherical-Polar Grid

For each observation, specify:
start time, observed value, uncertainty

An observation is made of *NP* samples; for each sample specify:
spatial coordinates *(lat,lon,depth),* **property type** (*T, S, U, V, SSH*)

Optional fields:
number of samples *(default = 1)*
sample weight *(default = 1)*
observation duration *(default = 0)*

See obsfit_mitgcm/make_obsfit_example.m



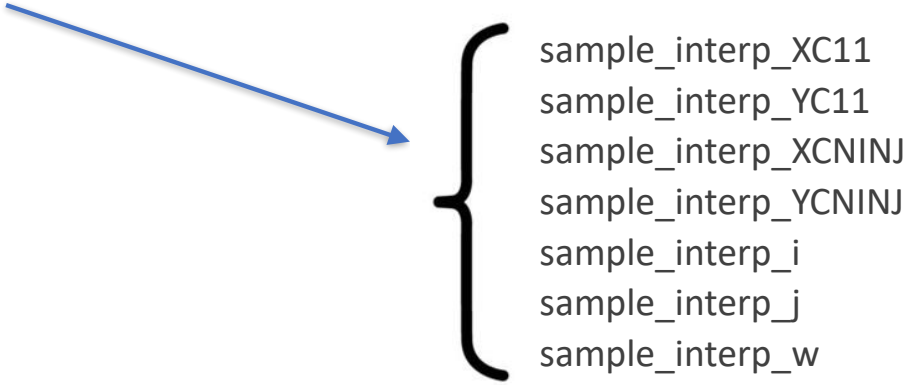
Pre-Processing Data for LLC Grid

For each observation, specify:

start time, observed value, uncertainty

An observation is made of *NP* samples; for each sample specify:

spatial coordinates *(x,y,z),* **property type** (*T, S, U, V, SSH*),
sample grid interpolation values

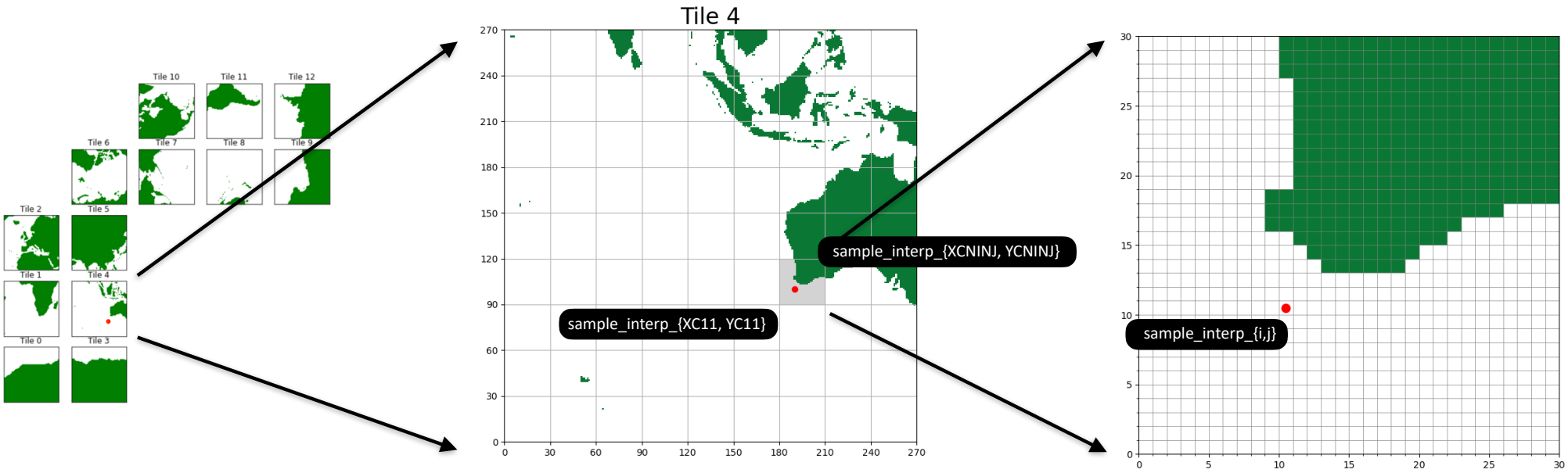


Optional fields:

observation duration *(default = 0)*
number of samples *(default = 1)*
sample weight *(default = 1)*

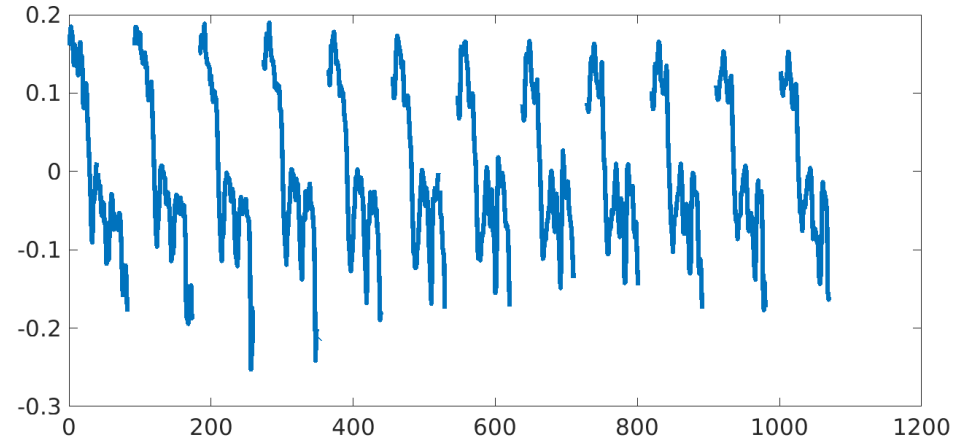
Algorithm based off of https://github.com/gaelforget/MITprof

Pre-Processing Data for LLC Grid



Model - Data Comparison

input:

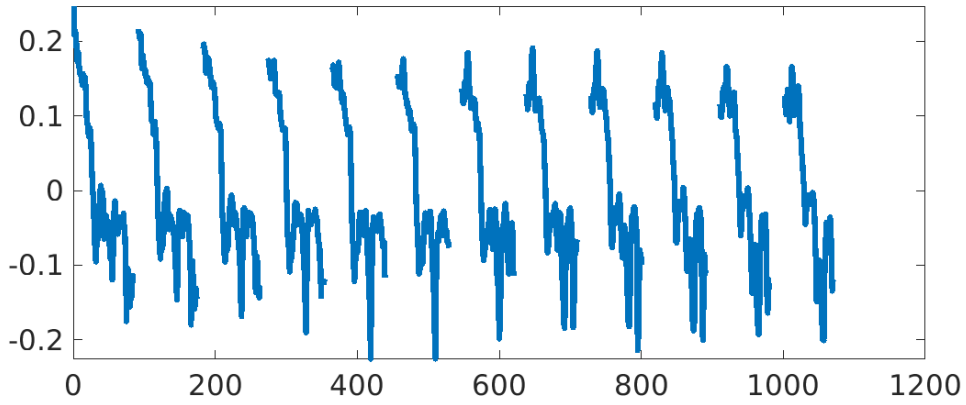


example.nc

observed value
uncertainty
sample type
observation time

latitude
longitude
depth

output:

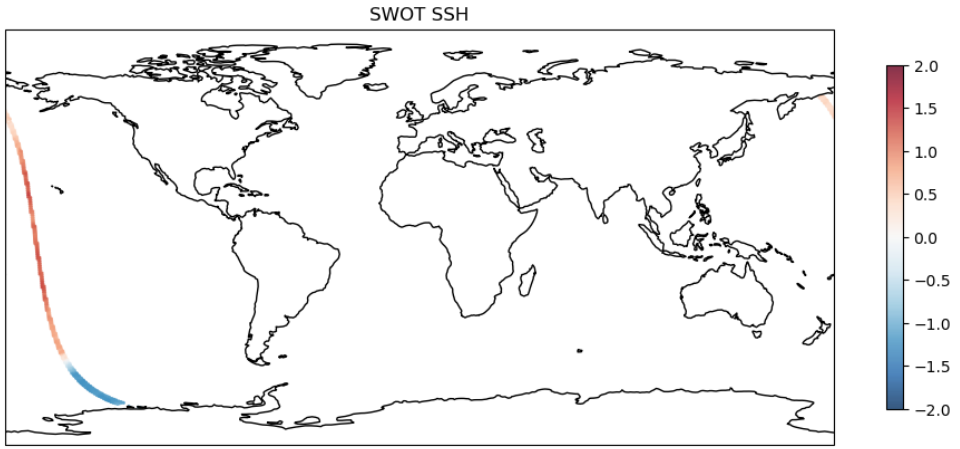


example.equi.nc

model equivalent value
mask

Model - Data Comparison

input:

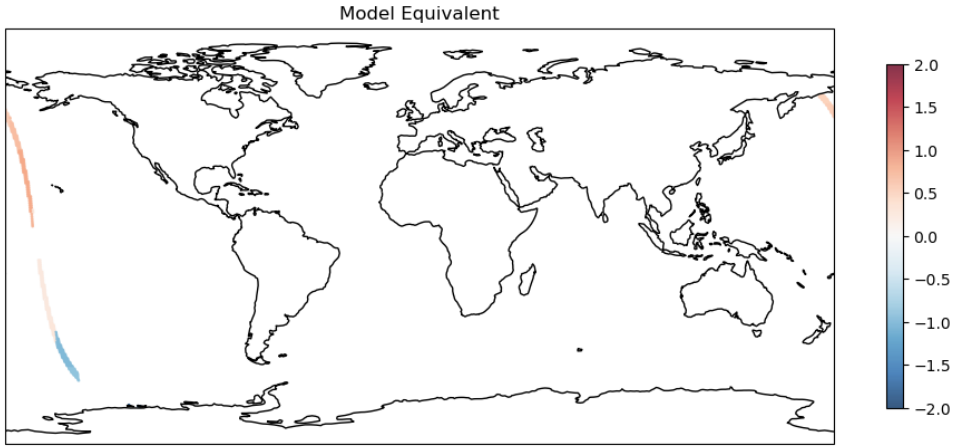


example.nc

observed value
uncertainty
sample type
observation time

longitude
latitude
depth

output:

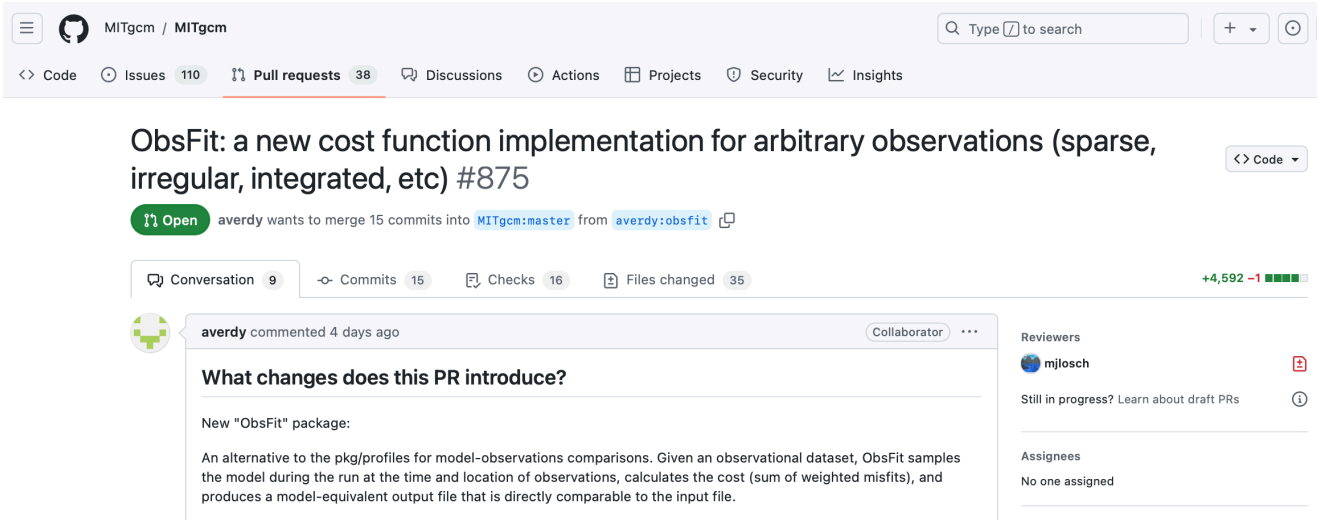


example.equi.nc

model equivalent value
mask

Status Update

- PR submitted last Friday
- LLC grid compatibility still under development (but close!)



[https://github.com/averdy/obsfit_mitgcm]

Extra Slides

Why?

- **Sparse data (memory allocation)**
  Obsfit reads in vector of observations instead of 2D or 3D matrices
- **Respect the data sampling**
  No need to interpolate observations to model grid
- **Flexible averaging operators**
  ObsFit handles spatial and temporal averaging and integration
- **Multi-grid optimization**
  Model-equivalent values are stored in data space, independent of model grid
- **Universal model validation**
  Input and output files are independent of model setup

xarray.Dataset

▶ Dimensions: (iOBS: 270303, iSAMPLE: 270303, iINTERP: 8)

▶ Coordinates: (0)

▼ Data variables:

obs_date	(iOBS)	float64	...
obs_YYYYMMDD	(iOBS)	int64	...
obs_HHMMSS	(iOBS)	int64	...
sample_x	(iSAMPLE)	float64	...
sample_y	(iSAMPLE)	float64	...
sample_z	(iSAMPLE)	float64	...
sample_type	(iSAMPLE)	float64	...
obs_val	(iOBS)	float64	...
obs_uncert	(iOBS)	float64	...
sample_interp_...	(iOBS)	float64	...
sample_interp_...	(iOBS)	float64	...
sample_interp_...	(iOBS)	float64	...
sample_interp_...	(iOBS)	float64	...
sample_interp_i	(iOBS)	float64	...
sample_interp_j	(iOBS)	float64	...
sample_interp_w	(iOBS, iINTERP)	float64	...

▶ Indexes: (0)

▶ Attributes: (0)