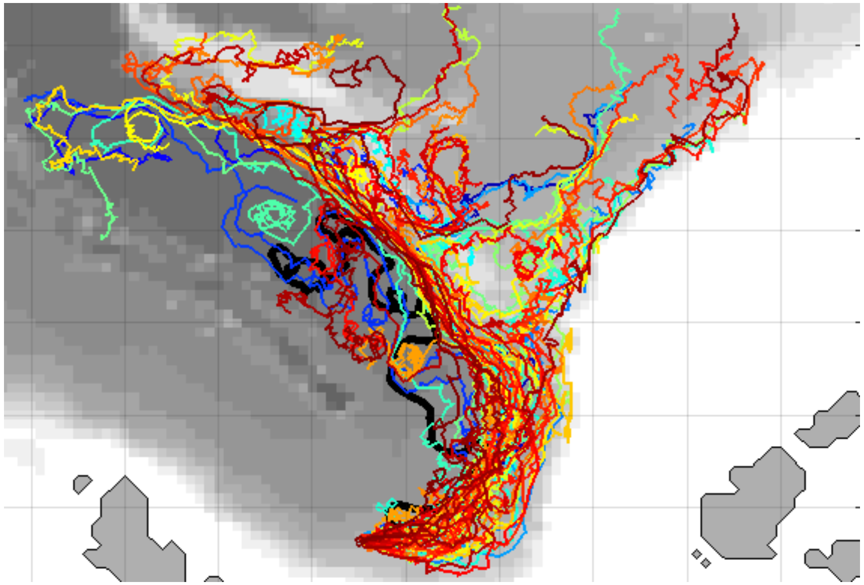
ALPS in state estimation & forecasting frameworks:

A survey of science applications, uncertainty quantifications, and observing network design

An T. Nguyen

University of Texas, Austin



Outline

- Frameworks:
  - Ocean State Estimation
  - Ocean Data Assimilation/Forecasting
  - Examples
- Uncertainty quantification
- Observing network design

What is data assimilation?

It's all about …

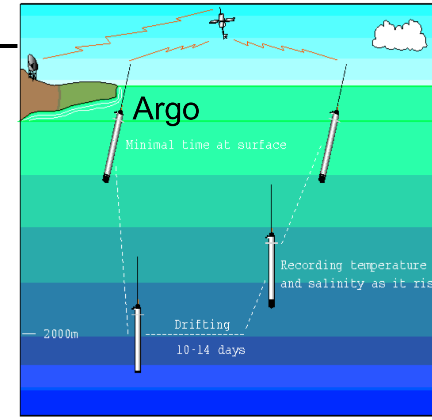
- making optimal use of,
- consistently extracting,
- or combining

information contained in *observations* and physical laws expressed through a *model*, and taking into account all *uncertainties*.

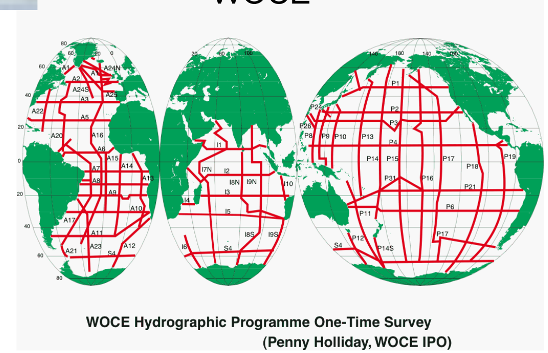
Combine two incomplete information sources

Observations ("data"):

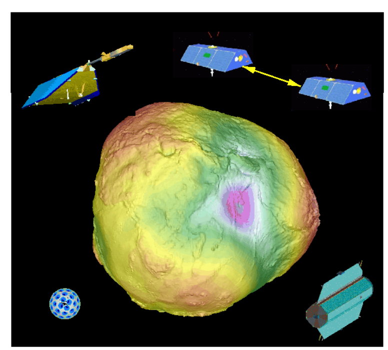
- incomplete/sparse probing of the physical system
  - spatial sampling
  - temporal sampling
  - incomplete state
- different physical variables
- heterogeneous data streams
- measurement errors
- representation errors (later)



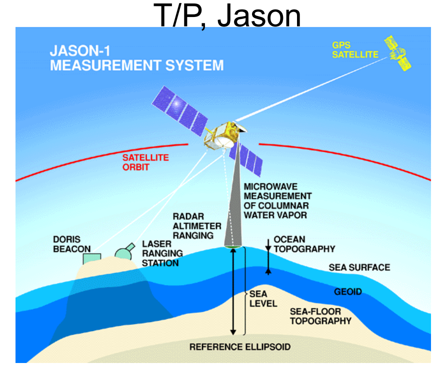
WOCE



GRACE



T/P, Jason



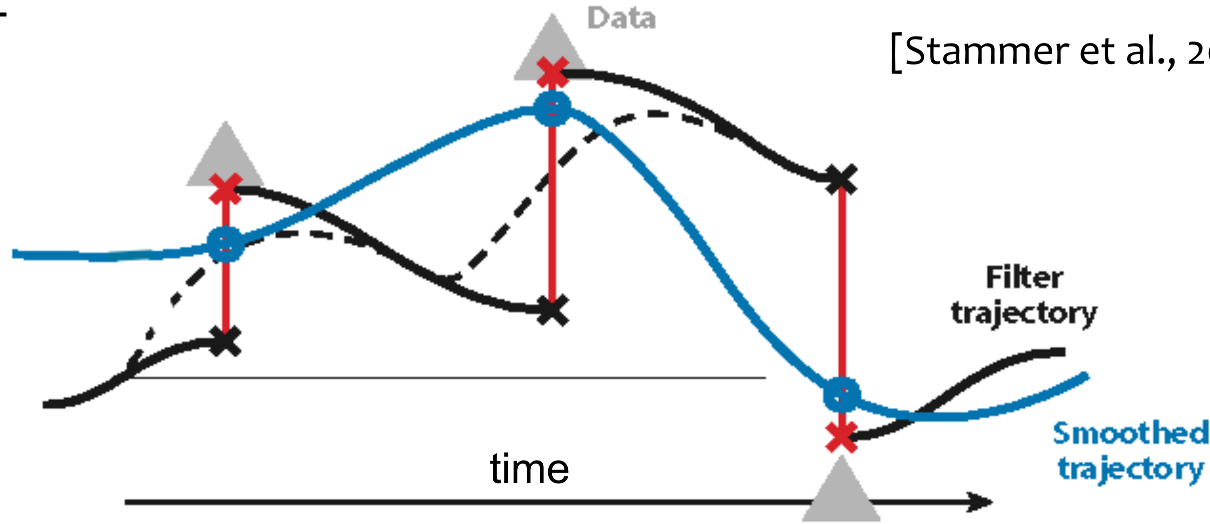
Combine two incomplete information sources

Physical model:

- representation of time-evolving state via equations of motion, conservation laws, theory, …
- An interpolator
- uncertainties/errors:
  - initial conditions
  - boundary conditions (surface, bottom, lateral)
  - model parameters
  - “model errors” (formulation, discretization, …)

$$
\begin{aligned}
\frac{D\vec{\mathbf{v}}_h}{Dt} + f\hat{\mathbf{k}}\times\vec{\mathbf{v}}_h + \frac{1}{\rho_c}\nabla_z p &= \vec{\mathcal{F}} \\
\epsilon_{nh}\frac{Dw}{Dt} + \frac{g\rho}{\rho_c} + \frac{1}{\rho_c}\frac{\partial p}{\partial z} &= \epsilon_{nh}\mathcal{F}_w \\
\nabla_z\cdot\vec{\mathbf{v}}_h + \frac{\partial w}{\partial z} &= 0 \\
\rho &= \rho(\theta, S) \\
\frac{D\theta}{Dt} &= \mathcal{Q}_\theta \\
\frac{DS}{Dt} &= \mathcal{Q}_s
\end{aligned}
$$

Data Assimilation can mean very(!) different things to different people



State Estimation (smoother/adjoint):

- Bring all observations into a dynamically consistent description of the past and recent time-varying ocean circulation.
- Strictly obeys model physics at all time
- non-linear inversion, iterative
- Study ocean dynamics and variability, global-scale and regional energy, heat, and water budgets.
- Decadal to multi-decadal timescale.

Data Assimilation (filter):

- Bring all observations into a model for the purpose of prediction / forecasting
- Model updates can break conservation law.
- Update: weighting between prior knowlkedge and data-model misfits
- Initialization, operational
- days to months timescale

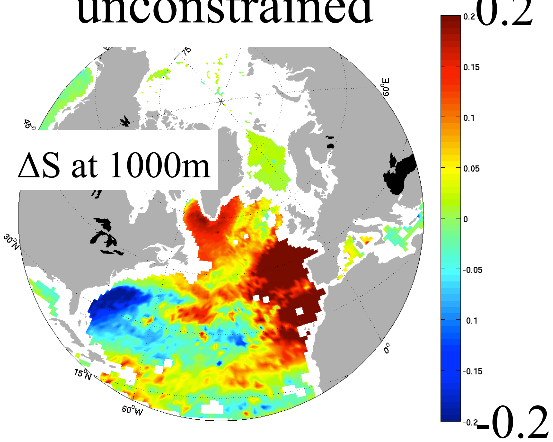
Science goal / application → determines the framework

Frameworks: Science application example

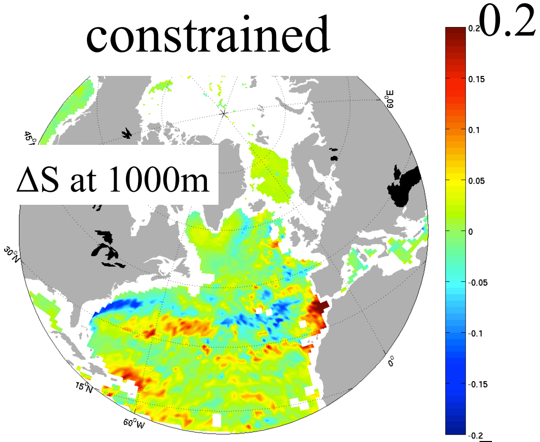
Estimating the Circulation and Climate of the Ocean (ECCO)

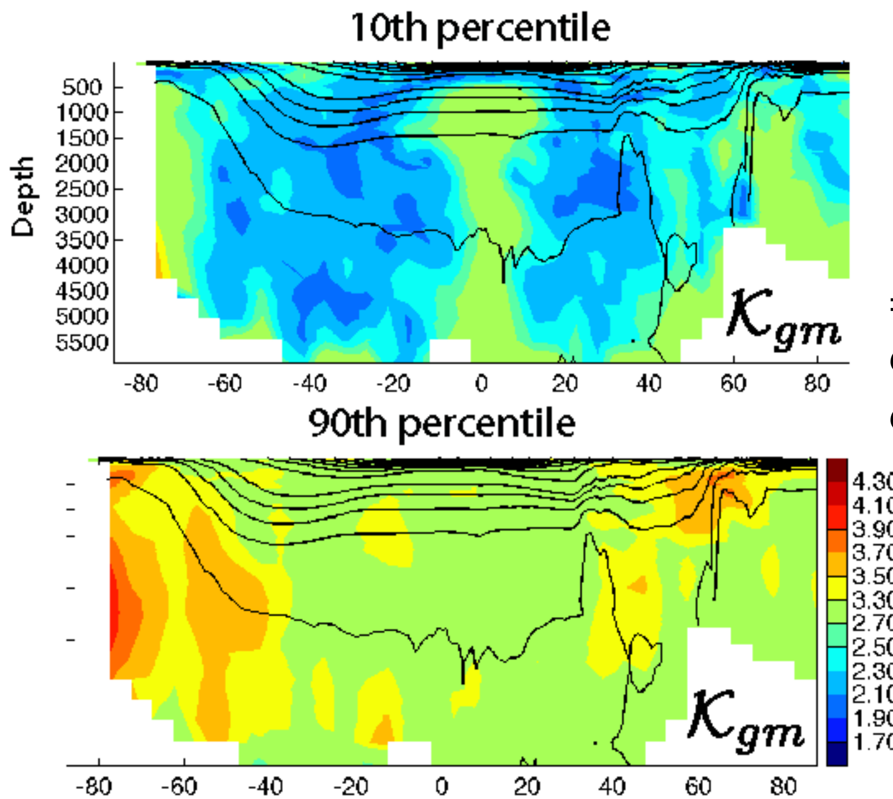
- State estimation
- Study ocean dynamics and variability, budgets.
- Estimate control parameters (e.g., mixing)

unconstrained



constrained





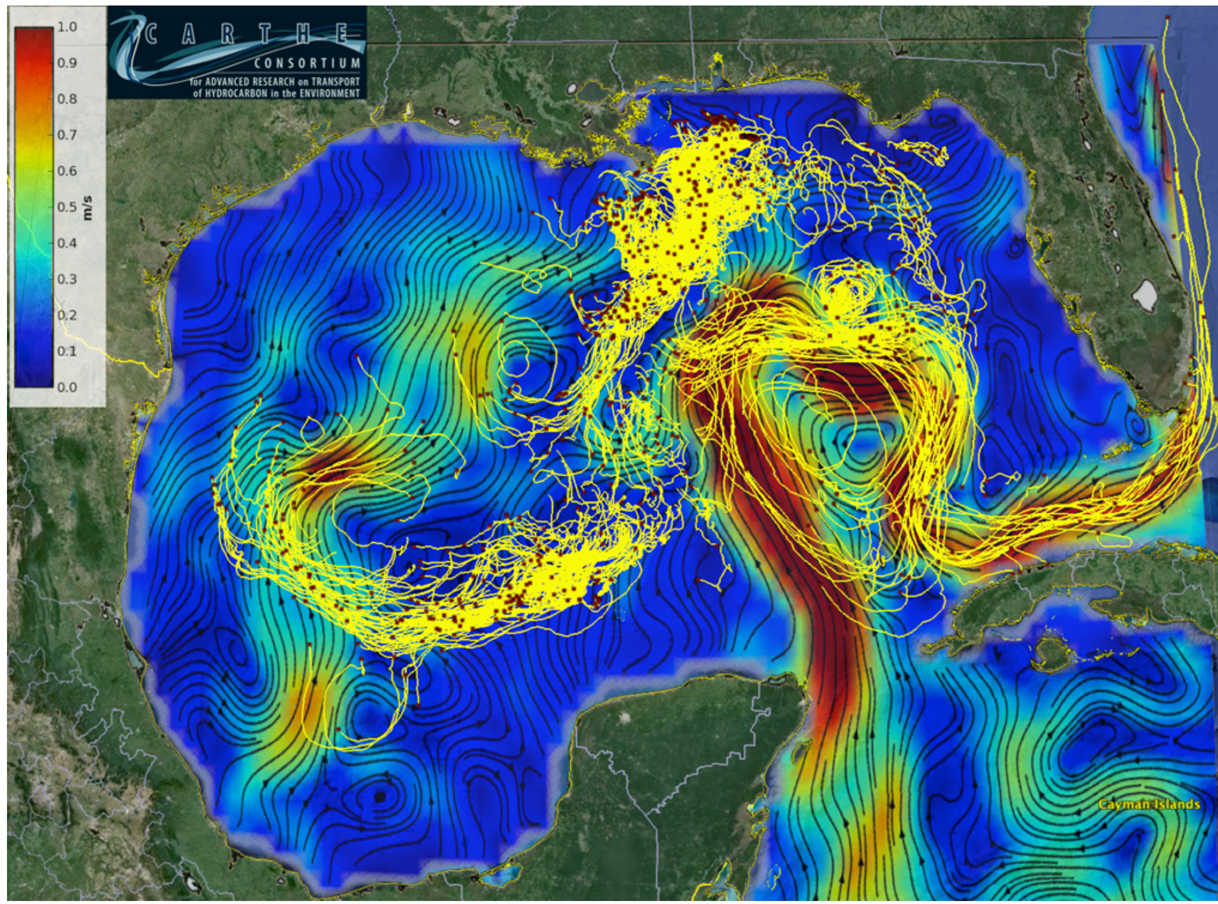
= transport effect of eddies

[Forget et al., 2015a,b, Stammer et al., 2005]

Frameworks: Science application example

Gulf of Mexico Research Initiative

- Data assimilation
- Tracking oil spills with <u>surface drifters</u> and high resolution numerical model
- Timescale ~ days



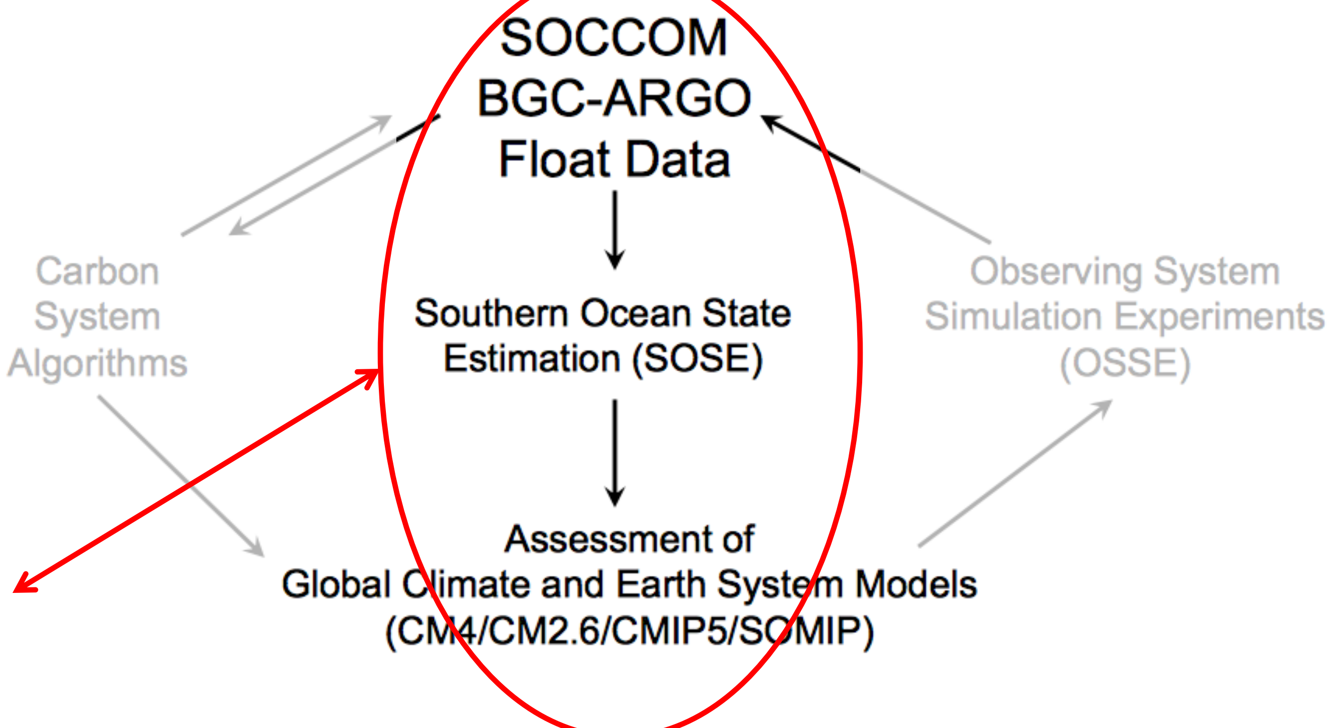
Frameworks: Science application example

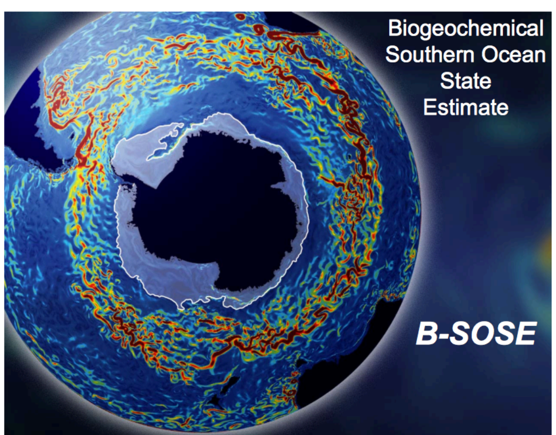
The **Southern Ocean** Carbon and Climate Observations and Modeling project (SOCCOM)

- Improve our understanding of the oceanic uptake of carbon and heat

SOCCOM The SOCCOM Modeling Plan

The SOCCOM float program informs several of the modeling projects and the modeling projects are helping with the planning, design and quality control of the float deployments and data.





Slide from I. Kamenkovich talk

Outline

- ❑ Frameworks:
  - ➢ Ocean State Estimation
  - ➢ Ocean Data Assimilation/Forecasting
  - ➢ Examples
- ❑ Uncertainty quantification
- ❑ Observing network design

How uncertain?

- The estimated state?
- Any climate diagnostic / target output derived from it?
- How affected by …
  - … observation uncertainty?
  - … observation sampling?
  - … prior information on input parameters?
  - … the model itself?

Uncertainty quantification

Turpin et al., [2016] *How essential is Argo for ocean forecasting?*

"The main conclusion is that the performance of the Mercator Ocean 0.25◦ global data assimilation system is **heavily dependent** on the availability of Argo data."

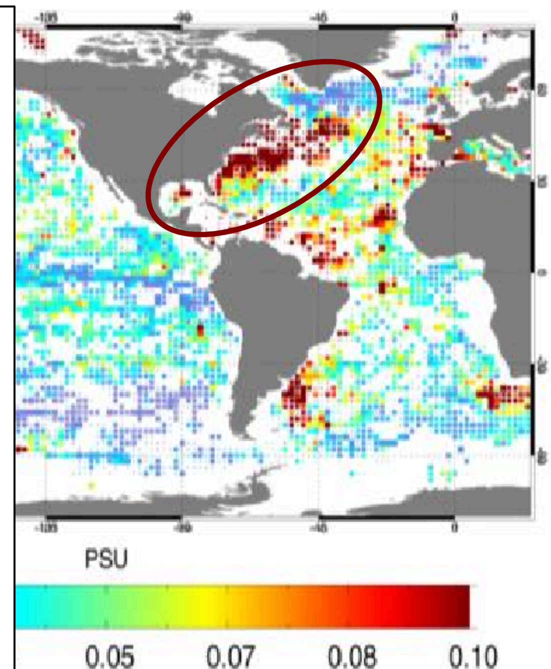
RMS of the difference between fully assimilated run and ARGO salinity

What does large RMS tell us?

a. highly dynamic region

b. variability captured in data?
(e.g. is sampling at 3° x 3° spatial coverage and 10-day rate sufficient?)

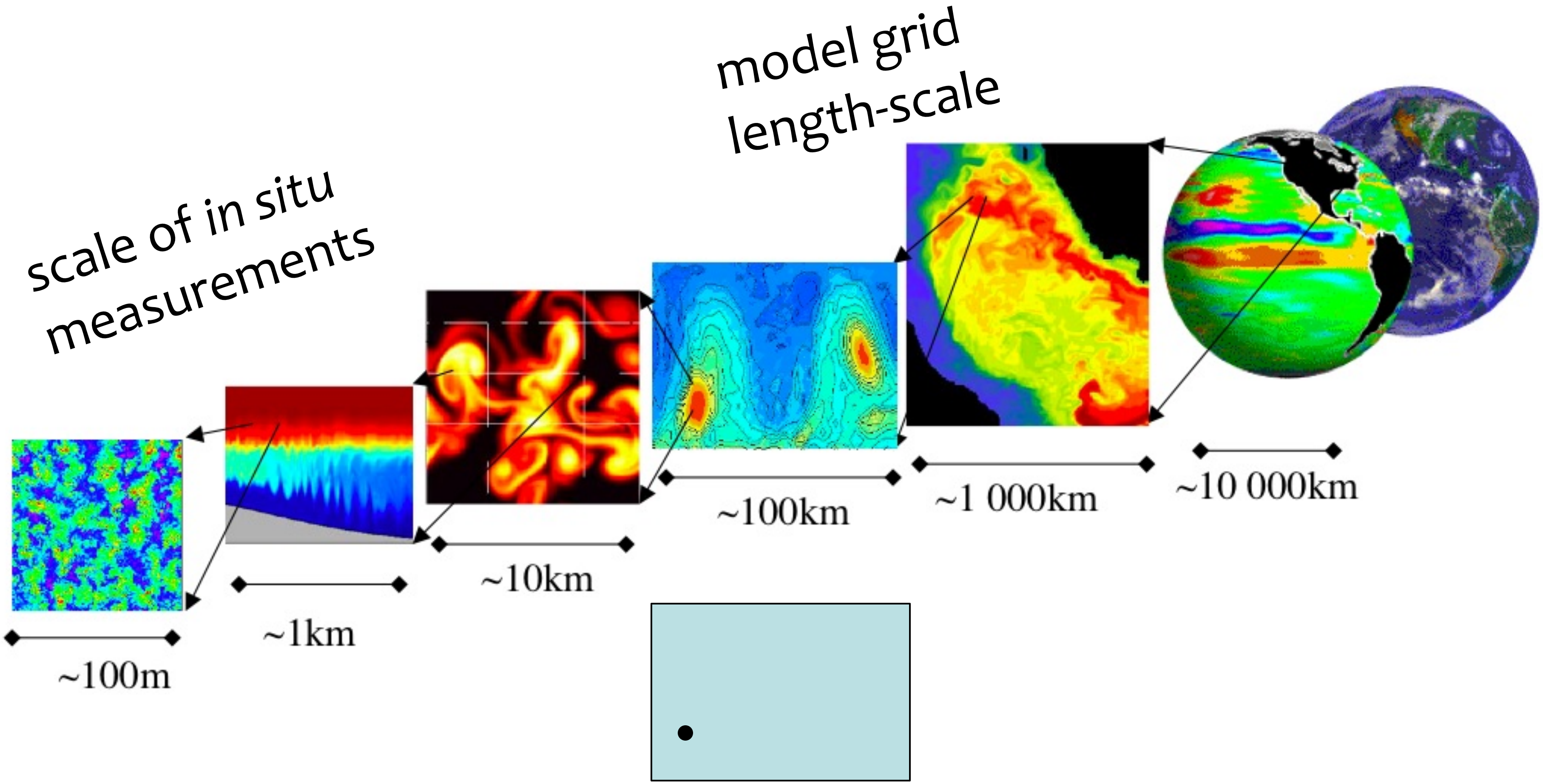
c. model ability to capture observed variability?



Uncertainty quantification

how to represent variability?

→ <u>uncertainty</u>



Uncertainty quantification

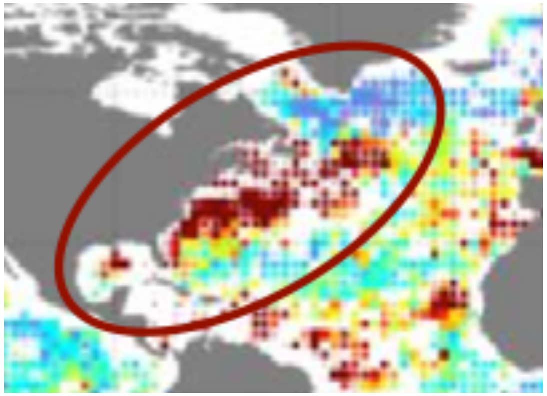
Data → Controls → Target

$\mathbf{y} \rightarrow \mathbf{x} \rightarrow \mathbf{z}$

Data uncertainty → Controls uncertainty → Target uncertainty

$\mathbf{P}_{yy} \rightarrow \mathbf{P}_{xx} \rightarrow \sigma_z^2$

target: misfits
target uncertainty?



Uncertainty quantification:

- Formal quantification : map $\mathbf{P}_{yy}$ and prior knowledge to P**xx,** [*e.g.* Kalmikov & Heimbach, 2014]
  → Computationally expensive
- Prior covariance matrices for controls ($\mathbf{x}$): ad-hoc, difficult to estimate.
- "usefulness" of target: requires knowledge of its uncertainty

Outline

- Frameworks:
  - Ocean State Estimation
  - Ocean Data Assimilation/Forecasting
  - Examples
- Uncertainty quantification
- Observing network design

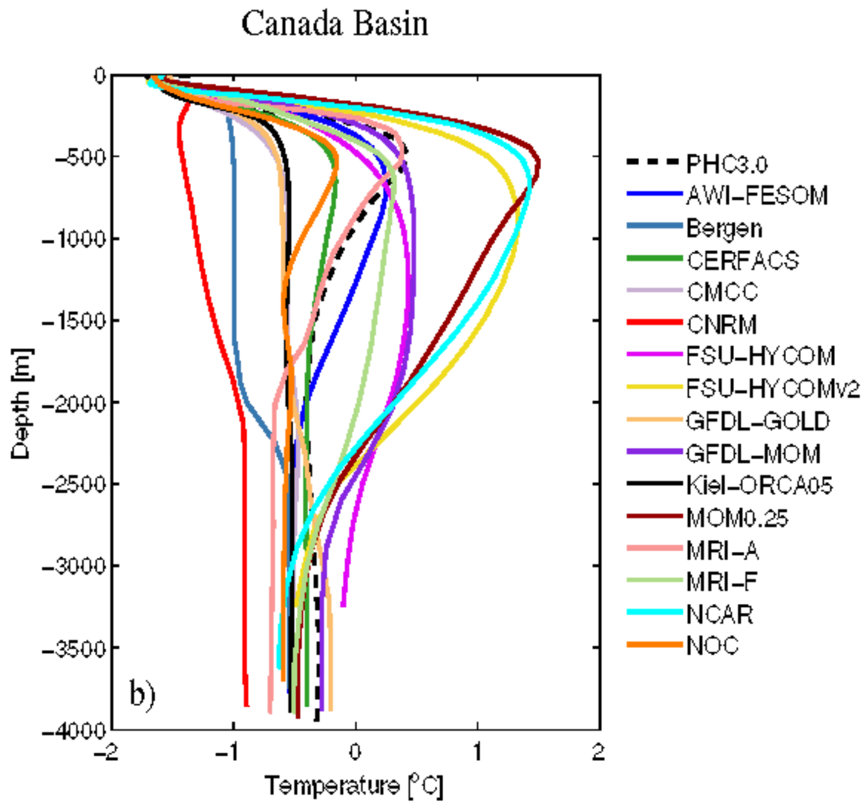
Observing network design:

With current knowledge:
- ➢ Where is uncertainty large?
- ➢ Where are gaps in data coverage?

[Stammer et al., 2016, more ref?]
- High latitude
- Coastal regions
- Gulf Stream path
- Western boundary currents
- Antarctic Circumpolar Current
- The Deep ocean

<u>Parameter uncertainties</u>:
Mixing
Eddy-stirring
Dissipation

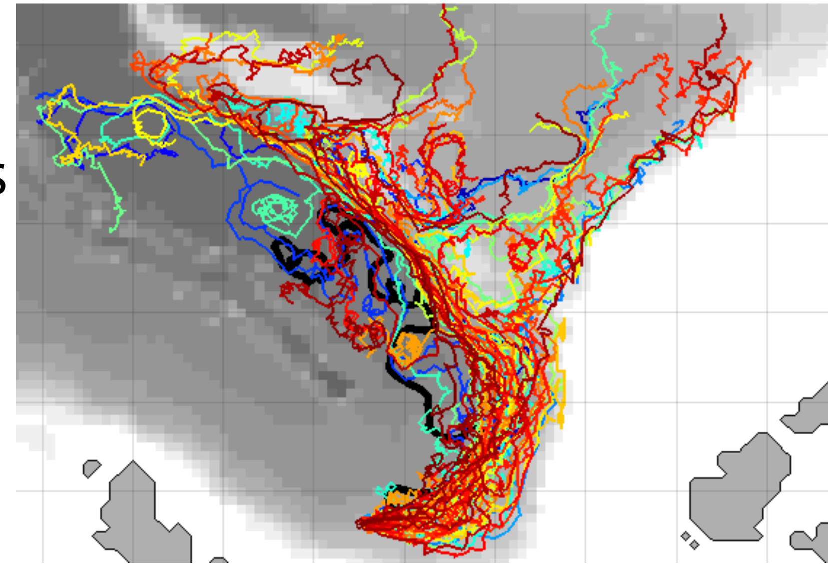
[Ilicak et al., 2016], CMIP5 models



Observing network design: example

Target:
"Usefulness" of measured T/S in the Arctic from potential Argo-type floats in constraining models

Challenge in the Arctic: presence of sea ice → floats cannot surface



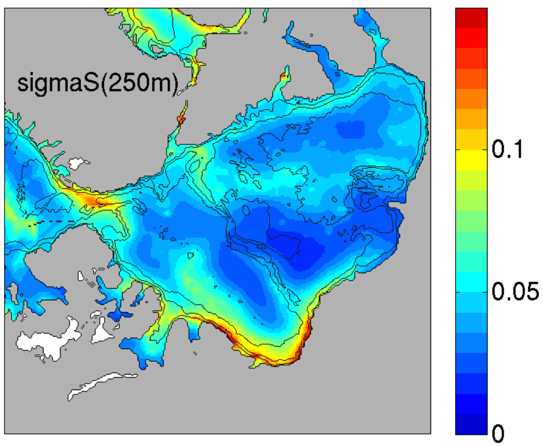
Vikram Garg

→ Errors in positioning accumulates during silent time
→ Map into errors in hydrography

Experiment:
- seeds floats in the Arctic,
- simulate possible trajectories,
- compute accumulated errors during silent time based on model's sea ice cover,
- compute T/S misfits between trajectories to referenced run.

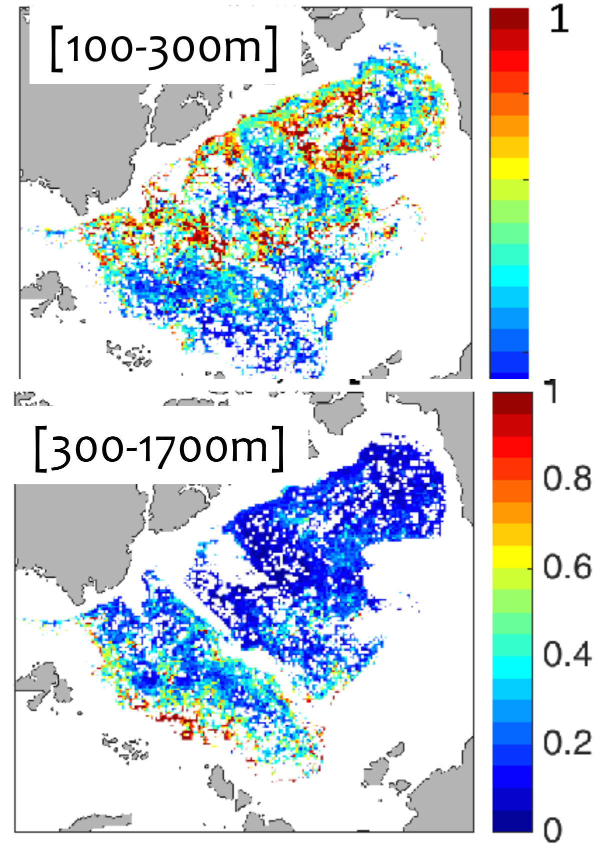
Observing network design: example

Data + model representation errors
[Forget & Wunsch 2007]
updated using ARGO and ITP data



Iterative process:
with additional observations
→ uncertainty should decrease
→ "usefulness" of repeated measurements likely will decrease

"usefulness" = (normalized salinity misfits)



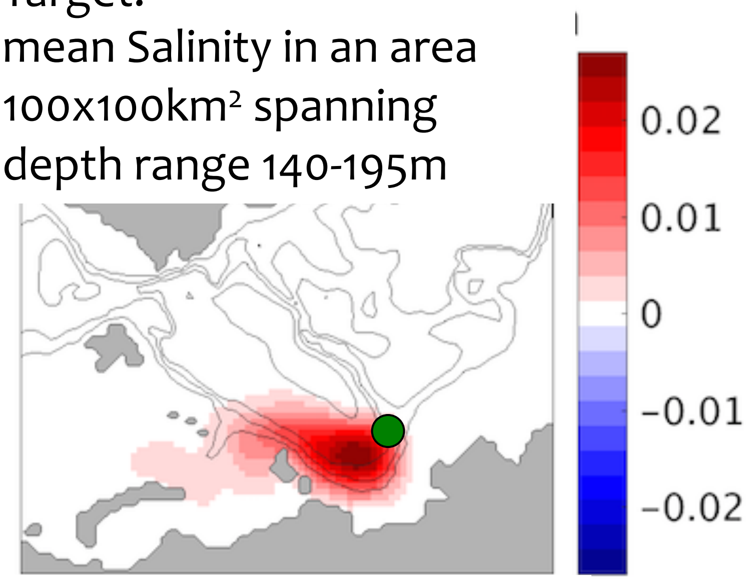
[Nguyen et al., 2017, *in prep*]

Observing network design: example

Experimental designing tool:

- Adjoint sensitivity

Target:
mean Salinity in an area
100x100km² spanning
depth range 140-195m



lag: N months

Define a **target J**.

Adjoint gradients: **∂J/∂ϕ**

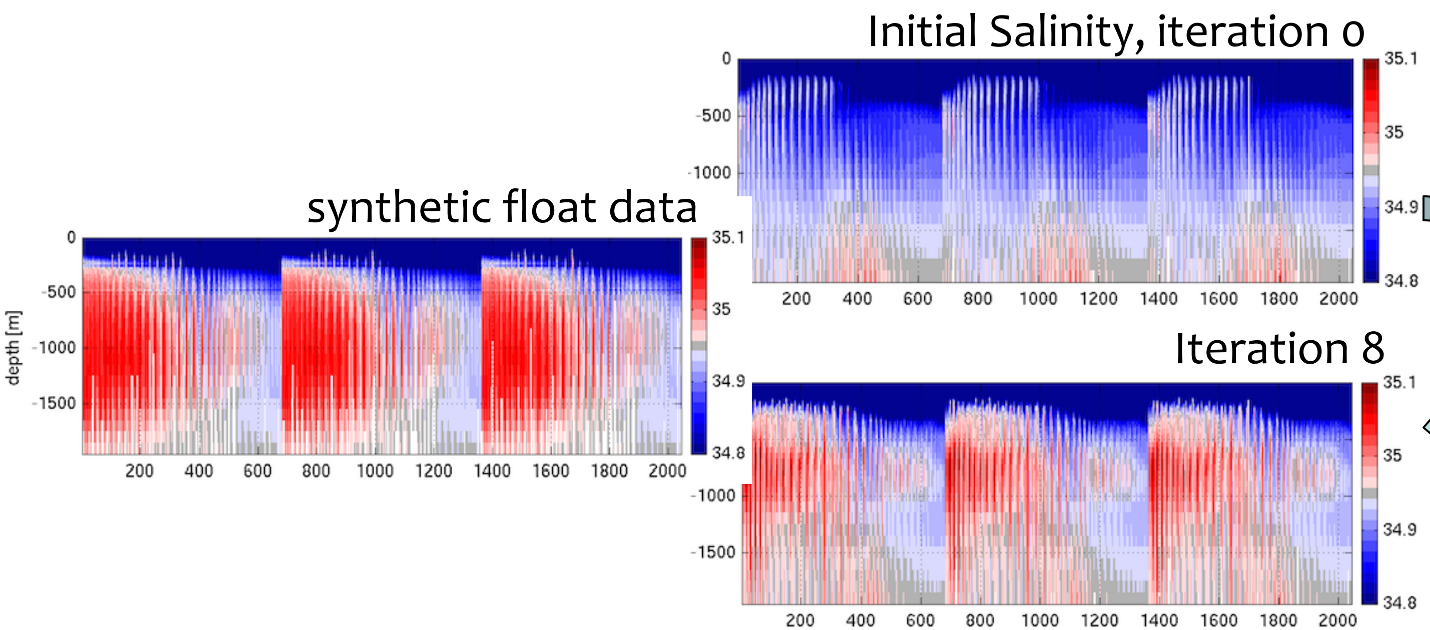
ϕ: ocean state, control variables

→ map sensitivity upstream (and downstream)

→ spatial and temporal knowledge (correlation)

→ Potential upstream sites to drop floats

Back to DA vs state estimation, now with uncertainty: potential impacts



Data assimilation
or
State estimation

<u>State estimation (smoother)</u>:
- Update: control parameters
- Dynamically consistent

→ Use for ocean dynamics and variability studies, budgets analyses,

<u>Data assimilation (filter)</u>:
- Update: state variables T/S

→ Use for initial conditions, forecasting

The machinery is in place and working, needs data to improve estimates of the states.

Summary

- ALPS, in combination with satellite, observations have significantly improved estimates of the ocean states in the past two decades

- ARGO floats: critical to constrain the ocean variability in the upper 2000m

- Challenges: Uncertainty
  - Data: representative of variability? Temporal/spatial aliasing?
  - Model: formal uncertainty quantification vs ad-hoc

- Observing network design
  - address regions of systematic large misfits, data gaps
  - Iterative process