ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

AUGMENTING A SEA OF DATA WITH DYNAMICS: OCEAN STATE & PARAMETER ESTIMATION

Patrick Heimbach

Nora Loose, An T. Nguyen, Helen Pillar, Tim Smith, … , CRIOS & ECCO groups

The University of Texas at Austin, TX

https://crios-ut.github.io

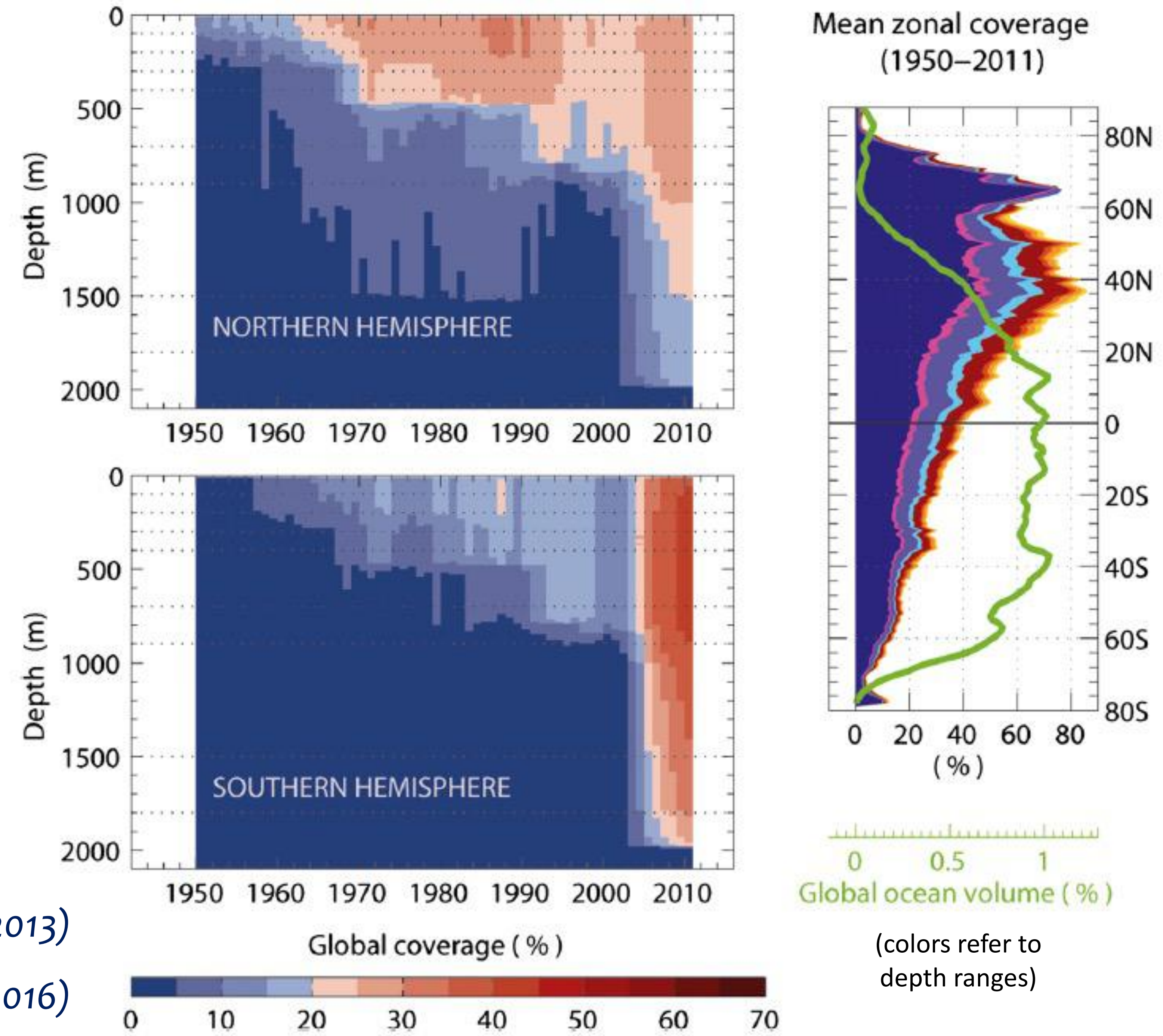
https://ecco-group.org

Overview

1. Oceanography: a sparse-data science
2. ECCO: deterministic inverse problem in a statistical/Bayesian framework
3. DA filters versus smoothers
4. Science applications: time-evolving property budgets
5. Beyond optimization: adjoints for dynamical attribution
6. Hessian uncertainty quantification & observing system design

Oceanography: **<u>A *sparse data* problem</u> …**

Observational sampling coverage for ocean temperature in the upper 2000 m
1950 – 2010
(mean ocean depth: ~ 3900 m)



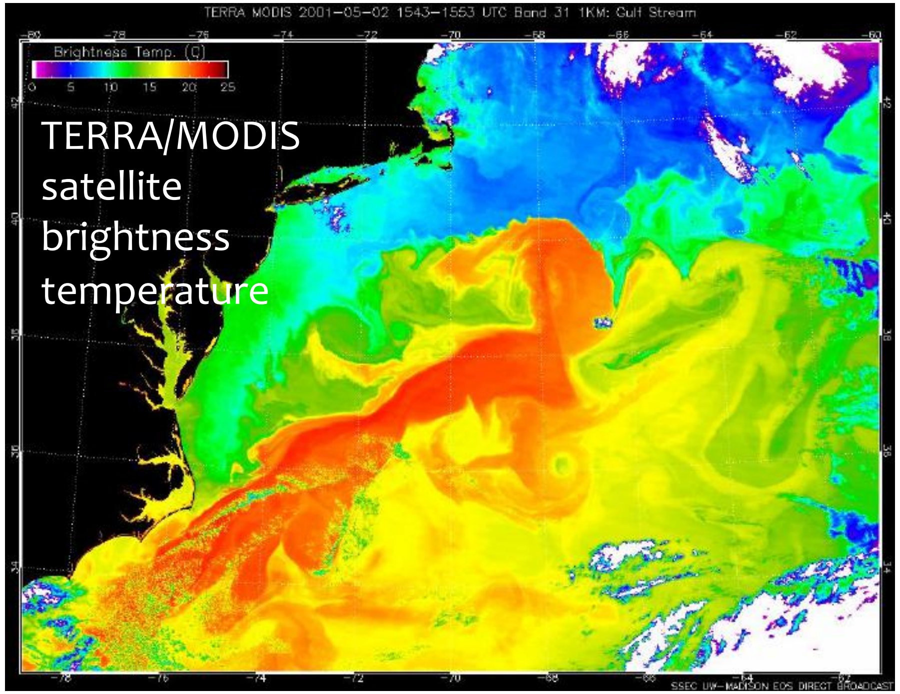
(colors refer to depth ranges)

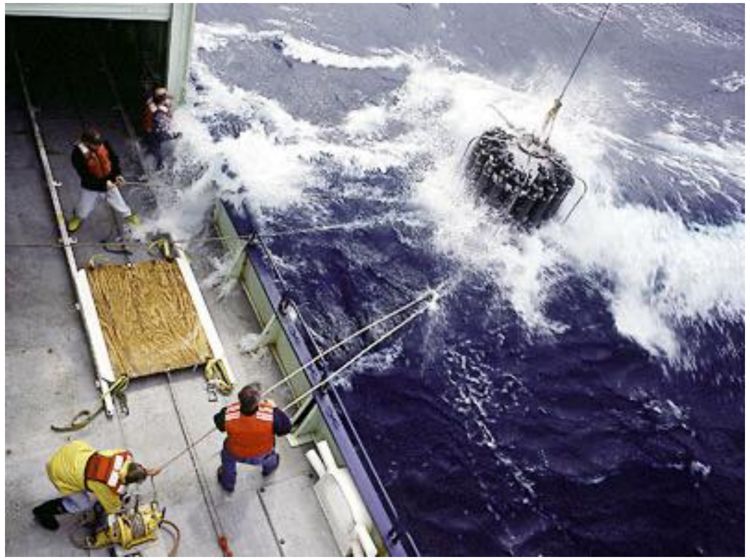
Abraham et al., Rev. Geophys. (2013)

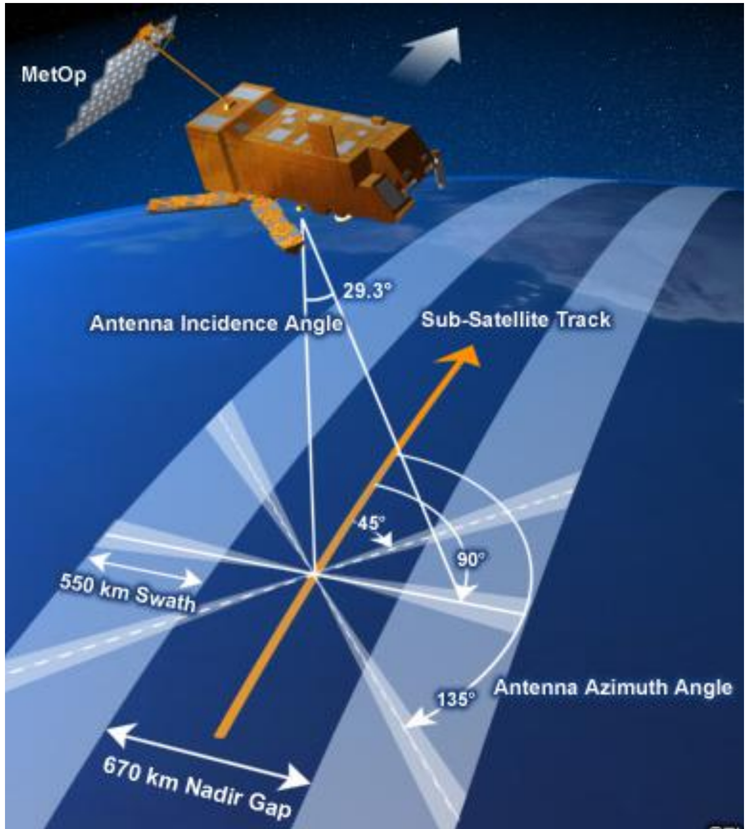
Wunsch (2016)

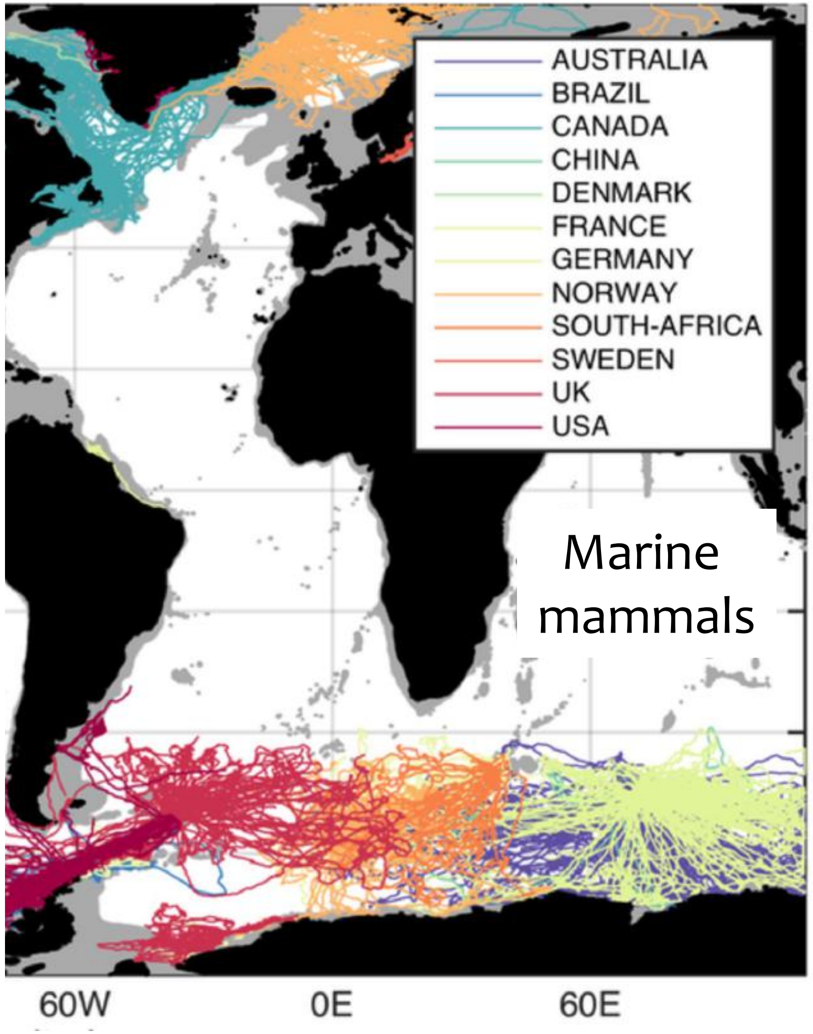
Some of the challenges:

An eclectic observing system with disparate sensors









http://www.meop.net

Two incomplete knowledge reservoirs

- an eclectic, patchy
  global ocean observing system

- **numerical models**
  that require uncertain inputs

Two incomplete knowledge reservoirs

- Can we optimally combine these two knowledge reservoirs?
- Can we do so in a manner that provides useful for climate analysis?
  - causal, dynamical attribution
  - detecting small, residual signals
- Can we provide measures of uncertainties with these?
- Can we use simulation to inform efficient observing strategies?

ECCO: Estimating the Circulation and Climate of the Ocean

A *multi-platform*, *multi-instrument* synthesis effort that integrates ocean and marine ice observations with equations of motion (model)

- *Physical Oceanography (PO)*
- *Cryosphere*
- *Modelling, Analysis, and Prediction (MAP)*
- *Advancing Collaborative Connections for Earth System Science (ACCESS)*

Jet Propulsion Laboratory
California Institute of Technology

TEXAS
The University of Texas at Austin

Massachusetts Institute of Technology

aer Atmospheric and Environmental Research

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

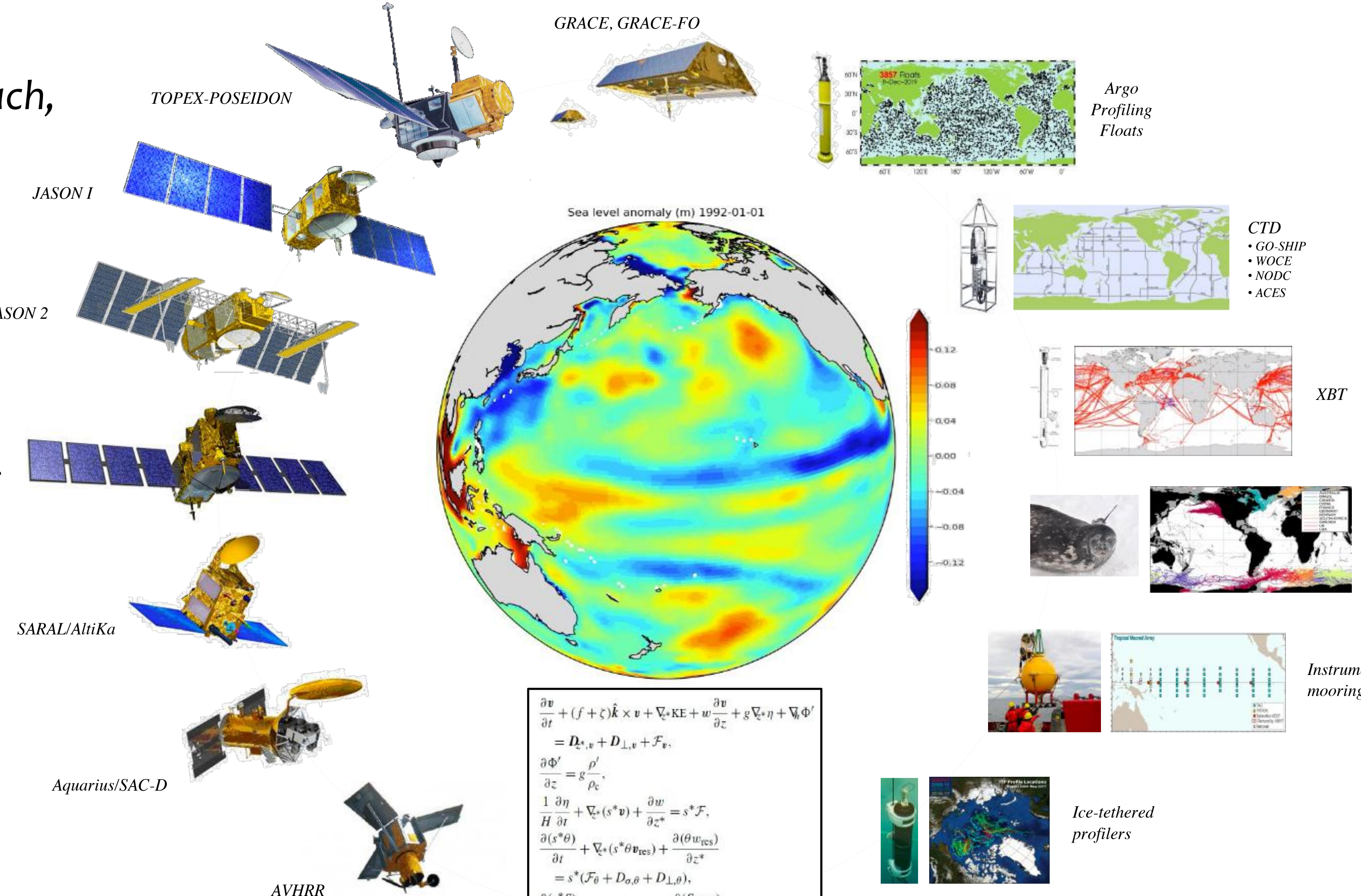
SCRIPPS INSTITUTION OF OCEANOGRAPHY

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

Learning from sparse, heterogeneous observations and models

Wunsch & Heimbach, Physica D (2007)

Heimbach et al., Front. Mar. Sci. (2019)



$$\frac{\partial \boldsymbol{v}}{\partial t} + (f+\zeta)\hat{\boldsymbol{k}} \times \boldsymbol{v} + \nabla_{z^*}\mathrm{KE} + w\frac{\partial \boldsymbol{v}}{\partial z} + g\nabla_{z^*}\eta + \nabla_h\Phi' = \boldsymbol{D}_{z^*,\boldsymbol{v}} + \boldsymbol{D}_{\perp,\boldsymbol{v}} + \mathcal{F}_{\boldsymbol{v}},$$

$$\frac{\partial \Phi'}{\partial z} = g\frac{\rho'}{\rho_c},$$

$$\frac{1}{H}\frac{\partial \eta}{\partial t} + \nabla_{z^*}(s^*\boldsymbol{v}) + \frac{\partial w}{\partial z^*} = s^*\mathcal{F},$$

$$\frac{\partial(s^*\theta)}{\partial t} + \nabla_{z^*}(s^*\theta\boldsymbol{v}_{\mathrm{res}}) + \frac{\partial(\theta w_{\mathrm{res}})}{\partial z^*} = s^*(\mathcal{F}_\theta + D_{\sigma,\theta} + D_{\perp,\theta}),$$

$$\frac{\partial(s^*S)}{\partial t} + \nabla_{z^*}(s^*S\boldsymbol{v}_{\mathrm{res}}) + \frac{\partial(Sw_{\mathrm{res}})}{\partial z^*} = s^*(\mathcal{F}_S + D_{\sigma,S} + D_{\perp,S}),$$

MIT general circulation model

Courtesy: I Fenty (JPL/Caltech)

https://ecco-group.org

What is Data Assimilation?

Kaminski et al., The Cryosphere (2015):

"Ideally, …

- … all observational data streams are interpreted simultaneously,
- … with the process information provided by the model,
- … [which leads to] a consistent picture of the state of the system,
- … that balances all the observational constraints,
- … taking into account all the respective uncertainty ranges."

Penny et al., Front. Mar. Sci. (2019):

"DA allows information provided from observations to be propagated in time and space to unobserved areas using the dynamical and physical constraints imposed by numerical models."

Kaipio & Somersalo (2005)
Statistical and Computational Inverse Problems

- **Formulate** inverse problem as statistical quest for information
- **Given** observable & unobservable quantities
- **Solve** inverse problem:
  - unobservable (or unobserved) quantities are of primary interest
  - extract information and assess uncertainty based on all available knowledge of measurement process, of models, and of prior information of unknowns
- **Solution** of inverse problem is a posterior PDF of unknowns

In the following, we adopt Isaac et al., J. Comp. Phys. (2015)

ECCO as a statistical/Bayesian inverse problem

Incomplete knowledge reservoirs

Model (ocean GCM)

$$\mathbf{x}(t) = \mathcal{L}(\,\mathbf{x}(t-1)\,,\,\mathcal{B}\,\mathbf{u}(t-1)\,)$$

Observations

$$\mathbf{y}(t) = \mathcal{E}\,\mathbf{x}(t) + \mathbf{n}(t)$$

$\mathbf{x}(t)$	:	model state vector at time t
$\mathbf{y}(t)$	:	observation vector at time t
$\mathbf{u}(t)$	:	vector of uncertain (control) variables at time t
$\mathbf{n}(t)$	:	residual noise of model-data misfit at time t
$\mathcal{L}$	:	nonlinear model operator
$\mathcal{E}$	:	state-to-observable map
$\mathcal{B}$	:	parameter-to-state map

Statistical inference over space of uncertain parameters u

$\pi_{post}(\mathbf{u}\,|\,\mathbf{y})$: posterior distribution of uncertain parameters $\mathbf{u}$, given observations $\mathbf{y}$

- combines prior PDF $\pi_{prior}(\mathbf{u})$ with
- likelihood PDF $\pi_{like}(\mathbf{y}\,|\,\mathbf{u})$

Bayes' Theorem

$$\pi_{post}(\mathbf{u}\,|\,\mathbf{y}) \;\propto\; \pi_{prior}(\mathbf{u})\,\pi_{like}(\mathbf{y}\,|\,\mathbf{u})$$

ECCO as a statistical/Bayesian inverse problem

Expensive forward model & high-dimensional space of uncertain parameters necessitates assumptions:

1. $\pi_{prior}(\mathbf{u})$ is Gaussian
2. prior error covariance may impose smoothness, e.g., via elliptic PDE operator (inverse Laplacian)
3. difference between predicted observables, $\mathcal{E}\,\mathbf{x}(t)$, and actual observations, $\mathbf{y}(t)$ captured by noise vector

$$\mathbf{n}(t) = \mathcal{E}\,\mathbf{x}(t) - \mathbf{y}(t)$$

which obeys Gaussian statistics

ECCO as a statistical/Bayesian inverse problem

Generalize to parameter-to-observable map

$$\mathcal{F}(\mathbf{u}, t) = \mathcal{E} \circ \mathcal{L}(\, \mathbf{x}(t-1)\, , \, \mathcal{B}\, \mathbf{u}(t-1)\,)$$

with generally nonlinear $\mathcal{F}$, such that

$$\mathcal{E}\, \mathbf{x}(t) - \mathbf{y}(t) = \mathcal{F}(\mathbf{u}, t) - \mathbf{y}(t)$$

leads to ...

Bayes' Theorem with Gaussian noise & prior

$$\pi_{post}(\mathbf{u} \,|\, \mathbf{y}) \sim \exp \left\{ -\frac{1}{2} || \, \mathcal{F}(\mathbf{u}) - \mathbf{y} \, ||_{\Gamma_{obs}^{-1}} - \frac{1}{2} || \, \mathbf{u} - \mathbf{u}_{prior} \, ||_{\Gamma_{prior}^{-1}} \right\}$$

From statistical to deterministic inversion

Mean $\mathbf{u}_{MAP}$ of posterior distribution $\pi_{post}(\mathbf{u})$ is parameter vector maximizing π_{post}, called **maximum a posteriori** (MAP) point

- found by minimizing $-\log(\pi_{post})$
- solve optimization / deterministic inverse problem

$$\mathbf{u}_{MAP} = \arg\min_{\mathbf{u}} \left\{ -\frac{1}{2} || \mathcal{F}(\mathbf{u}) - \mathbf{y} ||_{\Gamma_{obs}^{-1}} - \frac{1}{2} || \mathbf{u} - \mathbf{u}_{prior} ||_{\Gamma_{prior}^{-1}} \right\}$$

This is the ECCO parameter & state estimation problem

Approximation of the posterior PDF by posterior error covariance Γ_{post}

Posterior error covariance given by inverse of the Hessian at $\mathbf{u}_{MAP}$

$$\Gamma_{post} = \left[H_{misfit}(\mathbf{u}_{MAP}) + \Gamma_{prior}^{-1} \right]^{-1}$$

where

$$H_{misfit}(\mathbf{u}_{MAP}) \approx F^T \, \Gamma_{obs}^{-1} \, F$$

Gauss-Newton Hessian approximation

with:

F : tangent linear operator of $\mathcal{F}$
F^T : adjoint operator of $\mathcal{F}$

The deterministic parameter & state estimation problem

Consider *"perfect"* nonlinear model $\mathcal{L}$, and observations $\mathbf{y}$ with noise $\mathbf{n}$

$$
\begin{aligned}
J(\mathbf{u}) = & \frac{1}{2} \sum_{t=1}^{t_f} \left[\mathcal{E}\mathbf{x}(t) - \mathbf{y}(t) \right]^T \Gamma_{obs}^{-1} \left[\mathcal{E}\mathbf{x}(t) - \mathbf{y}(t) \right] \\
& + \frac{1}{2} \sum_{t=0}^{t_f-1} \left[\mathbf{u}(t) - \mathbf{u}_{prior}(t) \right]^T \Gamma_{prior}^{-1} \left[\mathbf{u}(t) - \mathbf{u}_{prior}(t) \right] \\
= & \; J_{misfit} + J_{prior}
\end{aligned}
$$

Extend to Lagrange function $\mathcal{J}$, introducing Lagrange multipliers $\boldsymbol{\mu}(t)$:

$$
\mathcal{J}(\mathbf{u}, \boldsymbol{\mu}) = J(\mathbf{u}) - \sum_{t=1}^{t_f} \boldsymbol{\mu}^T \left[\mathbf{x}(t) - \mathcal{L}(\mathbf{x}(t-1)) \right]
$$

Lagrange multiplier method:
Stationary point of $\mathcal{J}$ leads to set of normal equations:

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{\mu}(t)} = \mathbf{x}(t) - \mathcal{L}(\mathbf{x}(t-1)) = 0 \qquad 1 \le t \le t_f$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{x}(t)} = \frac{\partial J}{\partial \mathbf{x}(t)} - \boldsymbol{\mu}(t) + \left[\frac{\partial \mathcal{L}(\mathbf{x}(t))}{\partial \mathbf{x}(t)}\right]^T \boldsymbol{\mu}(t+1) = 0 \qquad 0 < t < t_f$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{x}(t_f)} = \frac{\partial J}{\partial \mathbf{x}(t_f)} - \boldsymbol{\mu}(t_f) = 0 \qquad t = t_f$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{x}(0)} = \frac{\partial J}{\partial \mathbf{x}(0)} - \left[\frac{\partial \mathcal{L}(\mathbf{x}(0))}{\partial \mathbf{x}(0)}\right]^T \boldsymbol{\mu}(1) \qquad t_0 = 0$$

The deterministic parameter & state estimation problem

"Variational" hints that we need a gradient:

- gradient of J with respect to *unknown/uncertan* or *control* variables $\mathbf{u}$
- Here: Vary initial conditions, $\mathbf{x}(0)$ such as to minimize J

BUT: J depends not just on $\mathbf{x}(0)$, but on all $\mathbf{x}(t)$.

- consider *nonlinear model* $\mathbf{x}(t+1) = \mathcal{L}(\mathbf{x}(t))$
- linearized operator is *state transition matrix* or *Jacobian* $\mathbf{L}$

$$\delta x(t+1) = \frac{\partial x(t+1)}{\partial x(t)} \delta x(t) = \mathbf{L}\, \delta x(t)$$

Need chain rule of differentiation:

$$\begin{aligned} J &= J\left(x_0, x_1, x_2, \ldots, x_{t_f}\right) \\ &= J\Big(x_0, \mathcal{L}(x_0), \mathcal{L}(\mathcal{L}(x_0)), \ldots, \mathcal{L}^{N_{t_f}}(x_0)\Big) \end{aligned}$$

The deterministic parameter & state estimation problem

$$\mu_0 = \frac{\partial J}{\partial x_0} = \sum_{1 \le t \le t_f} \frac{\partial x_t}{\partial x_0} \left(\frac{\partial J}{\partial x_t} \right)$$

$$= \frac{\partial x_1}{\partial x_0} \left(\frac{\partial J}{\partial x_1} \right) + \frac{\partial x_1}{\partial x_0} \frac{\partial x_2}{\partial x_1} \left(\frac{\partial J}{\partial x_2} \right) + \ldots + \frac{\partial x_1}{\partial x_0} \cdots \frac{\partial x_{t_f}}{\partial x_{t_f - 1}} \left(\frac{\partial J}{\partial x_{t_f}} \right)$$

$$= \mathbf{L}^T \frac{\partial J}{\partial x_1} + \mathbf{L}^T \mathbf{L}^T \frac{\partial J}{\partial x_2} + \ldots + \mathbf{L}^T \cdots \mathbf{L}^T \frac{\partial J}{\partial x_{t_f}}$$

$\mathbf{L}^T$: is the **adjoint model** (and $\mathbf{L}$ is the **tangent linear model**)

$\boldsymbol{\mu}_t = \left(\frac{\partial J}{\partial x_t} \right)$: **Lagrange multiplier** or **dual state** at time t

The deterministic parameter & state estimation problem

For intermediate step of the adjoint model integration one obtains:

$$\begin{aligned}
\boldsymbol{\mu}_t = \frac{\partial J}{\partial x_t} &= \mathbf{L}^T \frac{\partial J}{\partial x_{t+1}} + \mathbf{E}^T \Gamma_{obs}^{-1} \left[\mathcal{E} x_t - y_t \right] \\
&= \mathbf{L}^T \left(\mathbf{L}^T \frac{\partial J}{\partial x_{t+2}} + \mathbf{E}^T \Gamma_{obs}^{-1} \left[\mathcal{E} x_{t+1} - y_{t+1} \right] \right) \\
&\quad + \mathbf{E}^T \Gamma_{obs}^{-1} \left[\mathcal{E} x_t - y_t \right]
\end{aligned}$$

- The adjoint model $\mathbf{L}^T$ propagates $\boldsymbol{\mu}_t$ (the sensitivity of J with respect to all earlier states x_t) backward in time to x_0;
- Each model–data misfit (i.e. innovation vector $\mathcal{E} x_t - y_t$) is a *source* of sensitivity;
- The gradient of J with respect to x_0 takes into account (and weighs) the size of *all* misfit terms, *all* (inverse) error covariances, and *all* (linearized) model dynamics.

The deterministic parameter & state estimation problem

Make explicit time-varying "forcing" in model $\mathcal{L}$,
with known part $\mathbf{B}\, q_k$, uncertain part $\mathbf{\Gamma}\, u_t$

$$x_{t+1} \;=\; \mathcal{L}(x_t,\, \mathbf{B}\, q_t,\, \mathbf{\Gamma}\, u_t)$$

Least-squares estimation problem now is:

$$\begin{aligned} J(x) = \sum_{0 \le t \le t_f} \Big\{ & \left[\mathcal{E} x_t \;-\; y_t\right]^T \mathbf{\Gamma}^{-1}_{obs} \left[\mathcal{E} x_t \;-\; y_t\right] \\ & + \left[u_t \;-\; u_{prior}\right]^T \mathbf{\Gamma}^{-1}_{prior} \left[u_t \;-\; u_{prior}\right] \Big\} \\ & + \left[x_0 \;-\; x^b\right]^T \mathbf{\Gamma}^{-1}_{init} \left[x_0 \;-\; x^b\right] \end{aligned}$$

The deterministic parameter & state estimation problem

Lagrange function $\mathcal{L}$ now takes on the form:

$$\mathcal{J} = J + \sum_{0 \leq t \leq t_f} \boldsymbol{\mu}(t)^T \left[\mathbf{x}(t+1) - \mathcal{L}(\, \mathbf{x}(t), \, \mathbf{B}\, \mathbf{q}(t), \, \boldsymbol{\Gamma}_{prior}\, \mathbf{u}(t) \,) \right]$$

And we have additional normal equation for gradient w.r.t. $\mathbf{u}(t)$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{x}(t)} = \frac{\partial J}{\partial \mathbf{x}(t)} - \boldsymbol{\mu}(t) + \left[\frac{\partial \mathcal{L}(\mathbf{x}, \, \boldsymbol{\Gamma}_{prior} \mathbf{u})}{\partial \mathbf{x}(t)} \right]^T \boldsymbol{\mu}(t+1) = 0 \qquad 0 < t < t_f$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}(t)} = \boldsymbol{\Gamma}_{prior}^{-1} u(t) + \left[\frac{\partial \mathcal{L}(\mathbf{x}, \, \boldsymbol{\Gamma}_{prior} \mathbf{u})}{\partial \mathbf{u}(t)} \right]^T \boldsymbol{\mu}(t+1) = 0 \qquad 0 < t < t_f$$

Machine Learning versus 4D-Variational Data Assimilation

Machine Learning		**(Variational) Data Assimilation**	
labels	y	observations	$y(t)$
features	x	state (at time t)	$x(t)$
neural network (surrogate model)	$y' = W(x)$	nonlinear physical forward model	$x(t+1) = \mathcal{L}(x(t))$
loss function	$J = (y - y')^2$	cost function	$J = \left[y - \mathcal{E}(\mathcal{L}(x))\right]^T \mathbf{R}^{-1} \left[y - \mathcal{E}(\mathcal{L}(x))\right]$
(?)		observation operator	$\mathcal{E}$
regularisation	$\|\|\alpha\|\|$	background	$J^b = \left[x_0 - x^b\right]^T \mathbf{B}^{-1} \left[x_0 - x^b\right]$
stochastic gradient		Newton method	
backpropagation		adjoint model	$\frac{\partial J}{\partial x} = \left[\left(\mathbf{R}^{-1}\right) \left(\frac{\partial J}{\partial y}\right) \mathbf{E} \left(\frac{\partial L}{\partial x}\right) \right]^T$
training		optimization, calibration	
convolutional layers		localization	
differentiable programming for AD		differentiable physics for AD	

Adjoint-based model calibration & state estimation

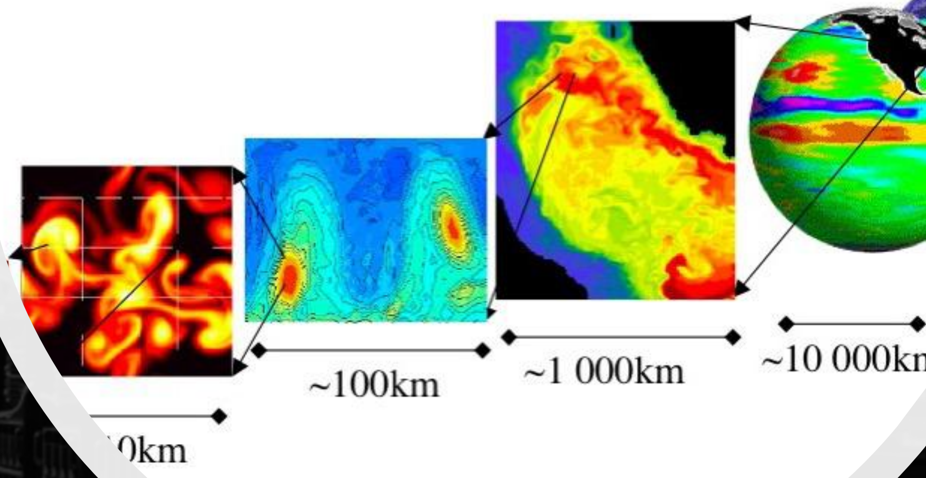
ECCO is learning from …

- (the most complete set of available) ocean observations
- … *AND* known physics/dynamics,
- … by solving a gigantic least-squares model-data misfit minimization
- … using the adjoint / Lagrange Multiplier Method

The MIT general circulation model (MITgcm) and AD-enabled adjoint code generation

Approximated form of Navier-Stokes equations for an incompressible fluid on rotating sphere (hydrostatic or non-hydrostatic), consisting of:

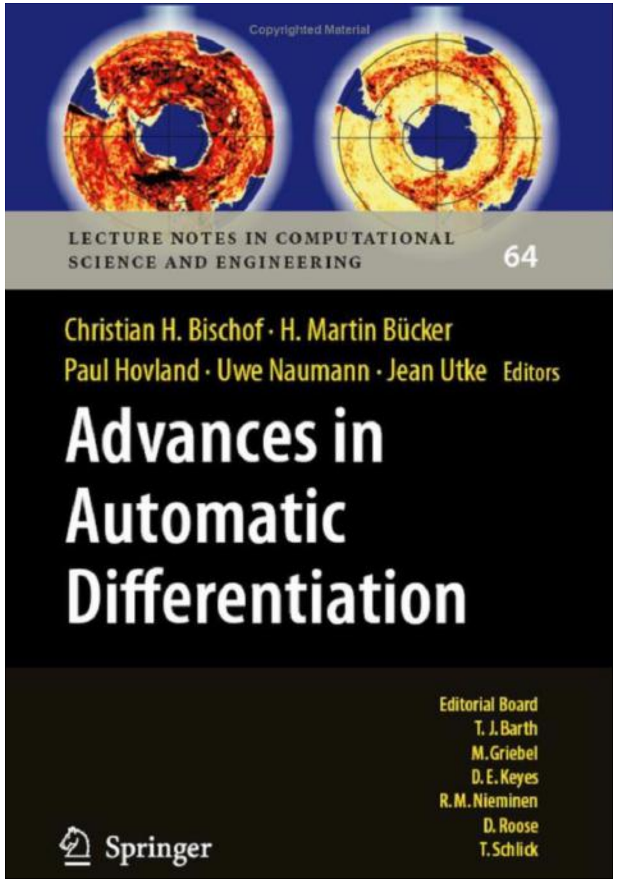
- *momentum equation*
- *conservation of mass*
- *conservation of heat, salt*
- *nonlinear equation of state for seawater*
- *subgrid-scale parameterizations*
- *scalable (domain decomposition)*
- *general curvilinear grid (incl. cubed-sphere)*
- ***adjoint code generation via automatic differentiation (AD)*** *using TAF, OpenAD*



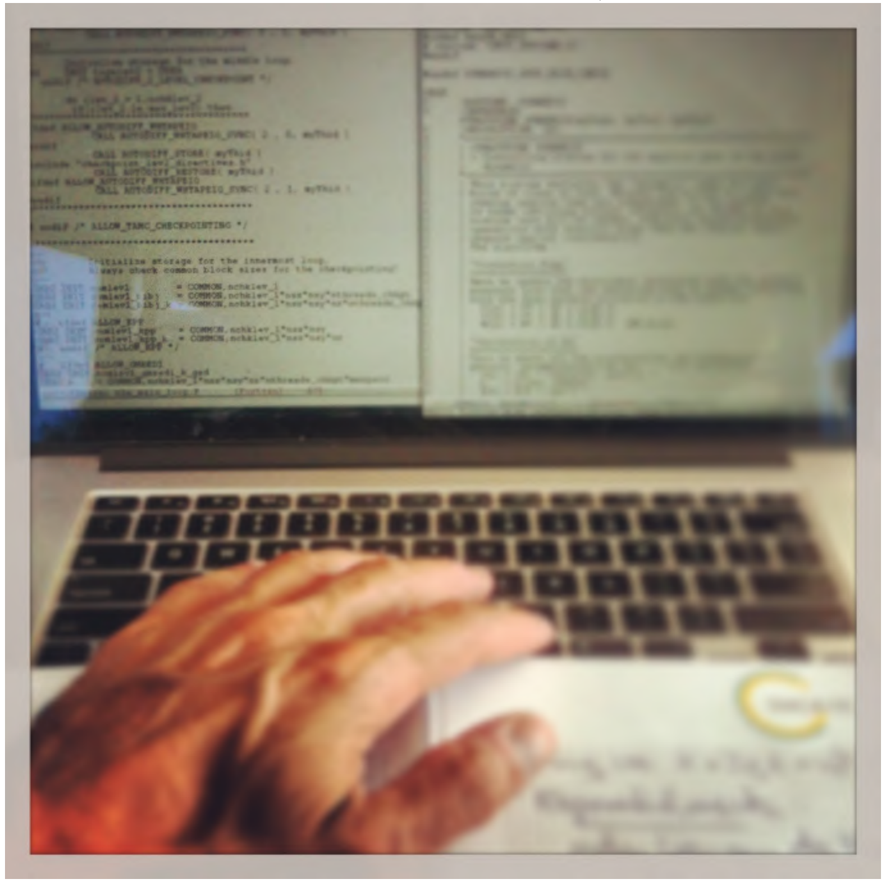
$$\begin{aligned}
\mathbf{k} \times \vec{\mathbf{v}}_h + \frac{1}{\rho_c}\nabla_z p &= \mathcal{F} \\
\epsilon_{nh}\frac{Dw}{Dt} + \frac{g\rho}{\rho_c} + \frac{1}{\rho_c}\frac{\partial p}{\partial z} &= \epsilon_{nh}\mathcal{F}_w \\
\nabla_z \cdot \vec{\mathbf{v}}_h + \frac{\partial w}{\partial z} &=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

Adjoint-based model calibration & state estimation

Generating & maintaining the adjoint of a state-of-the-art ocean circulation model



hand-written adjoint



Automatic Differentiation

Giering & Kaminski (1998); Marotzke et al. (1999); Heimbach et al. (2005); Utke et al. (2007); Griewank & Walther (2008)

Filter vs. Smoother

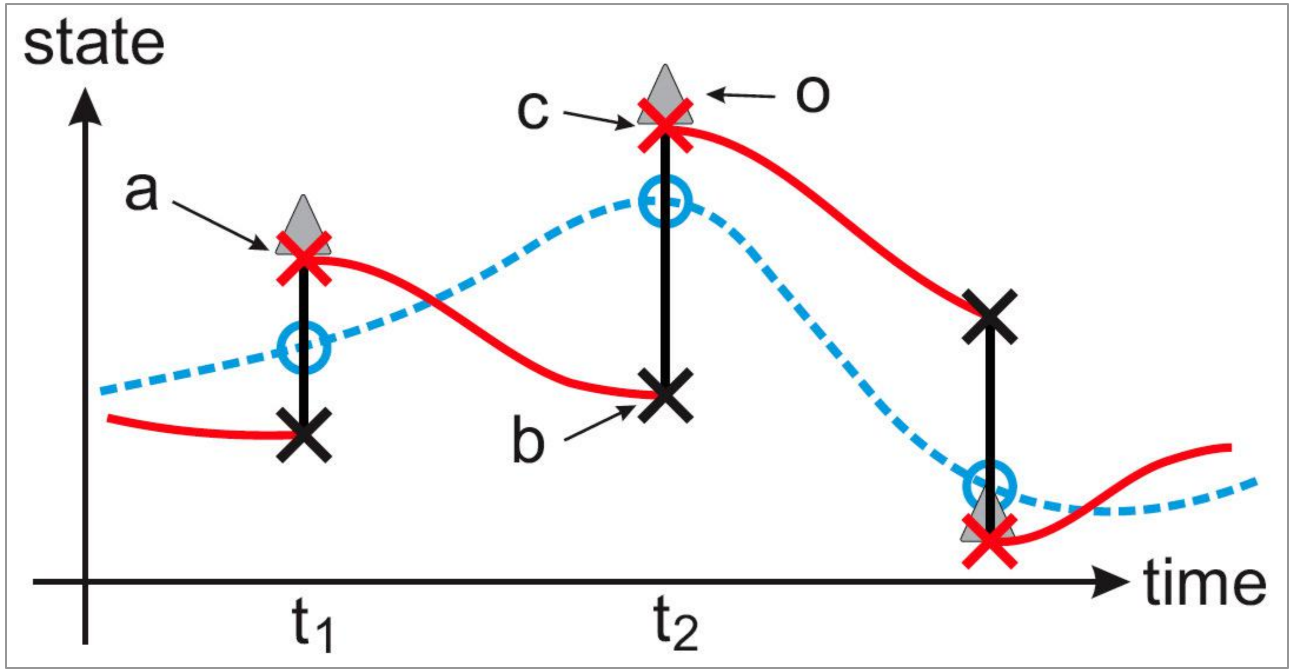
The virtues of property-conserving estimation

Some of the challenges:

Why adjoints: dynamical & kinematical consistency in DA

Numerical Weather Prediction (NWP) – a filtering problem

- Relatively abundant data sampling of the 3-dim. atmosphere
- NWP targets optimal forecasting
  - ➔ find initial conditions which produce best possible forecast;
  - ➔ *dynamical consistency or property conservation NOT required*



Some of the challenges:

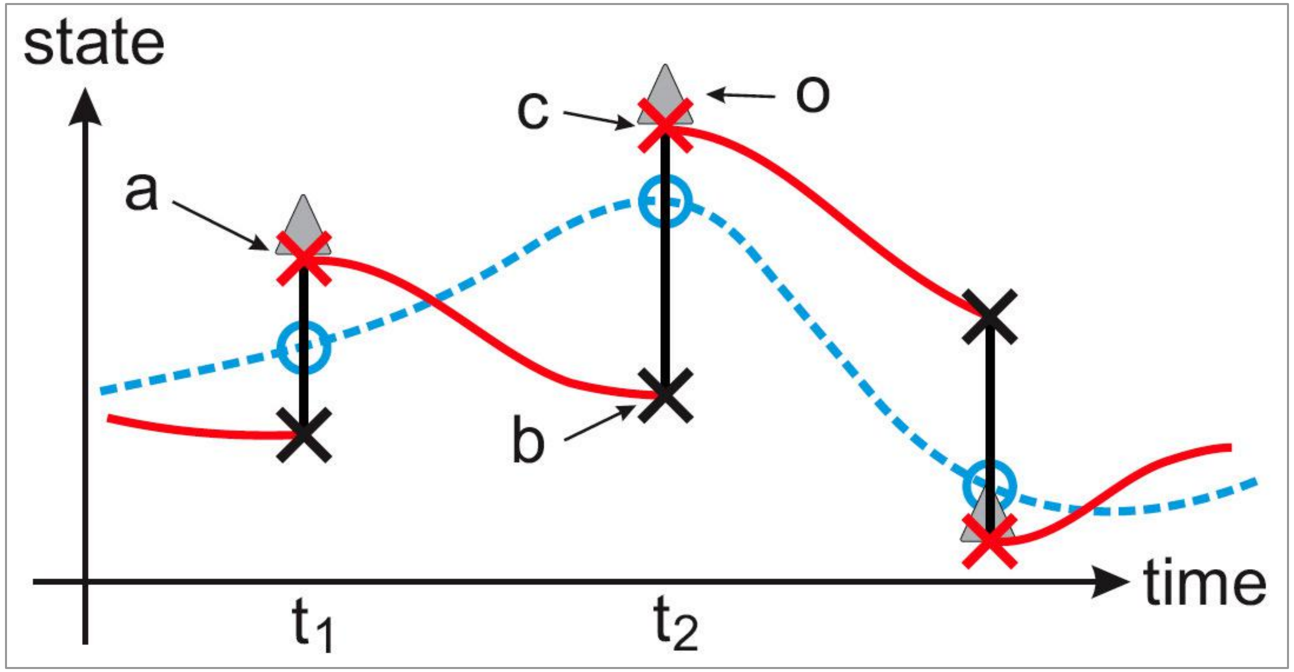
Why adjoints: dynamical & kinematical consistency in DA

Numerical Weather Prediction (NWP) – a filtering problem

- Relatively abundant data sampling of the 3-dim. atmosphere
- NWP targets optimal forecasting
  - ➔ find initial conditions which produce best possible forecast;
  - ➔ *dynamical consistency or property conservation NOT required*

Ocean state estimation/reconstruction – a smoothing problem

- Sparse data sampling of the 3-D. ocean
- Understanding past & present state of the ocean is a major goal all by itself
  - ➔ use observations in an optimal way
  - ➔ *dynamic consistency & property conservation ESSENTIAL* for climate



Some of the challenges:

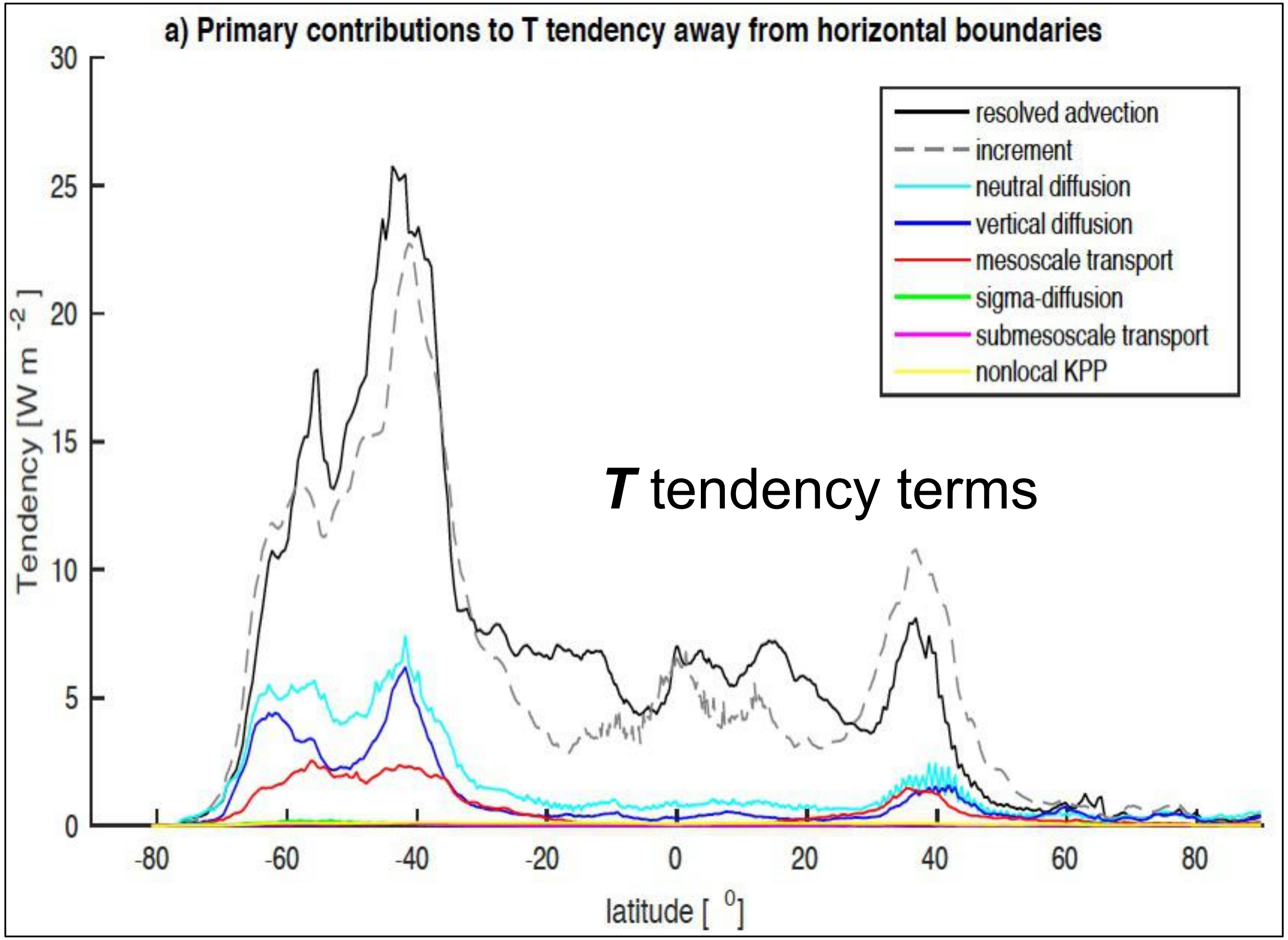
Why adjoints: dynamical & kinematical consistency in DA

Balancing the momentum, freshwater, and heat budgets

Some of the challenges:

Why adjoints: dynamical & kinematical consistency in DA

Tracer budgets in a global ocean reanalysis produced via filtering approach



Components in the tendency equation

dT/dt = r.h.s.

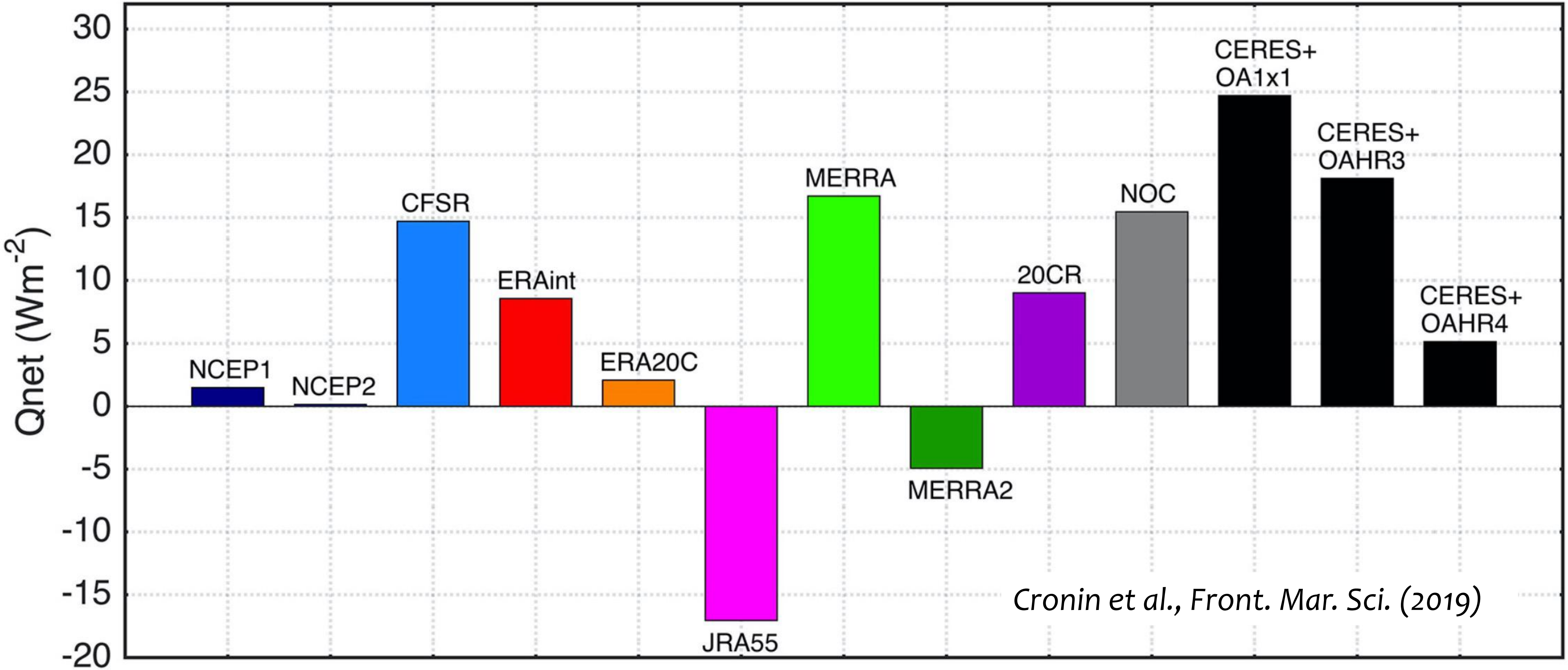
Unphysical analysis increments play leading role in the tracer tendencies

D. Trossman (in perpetual rejection)

Some of the challenges:

Why adjoints: dynamical & kinematical consistency in DA

Global-ocean net heat flux imbalance



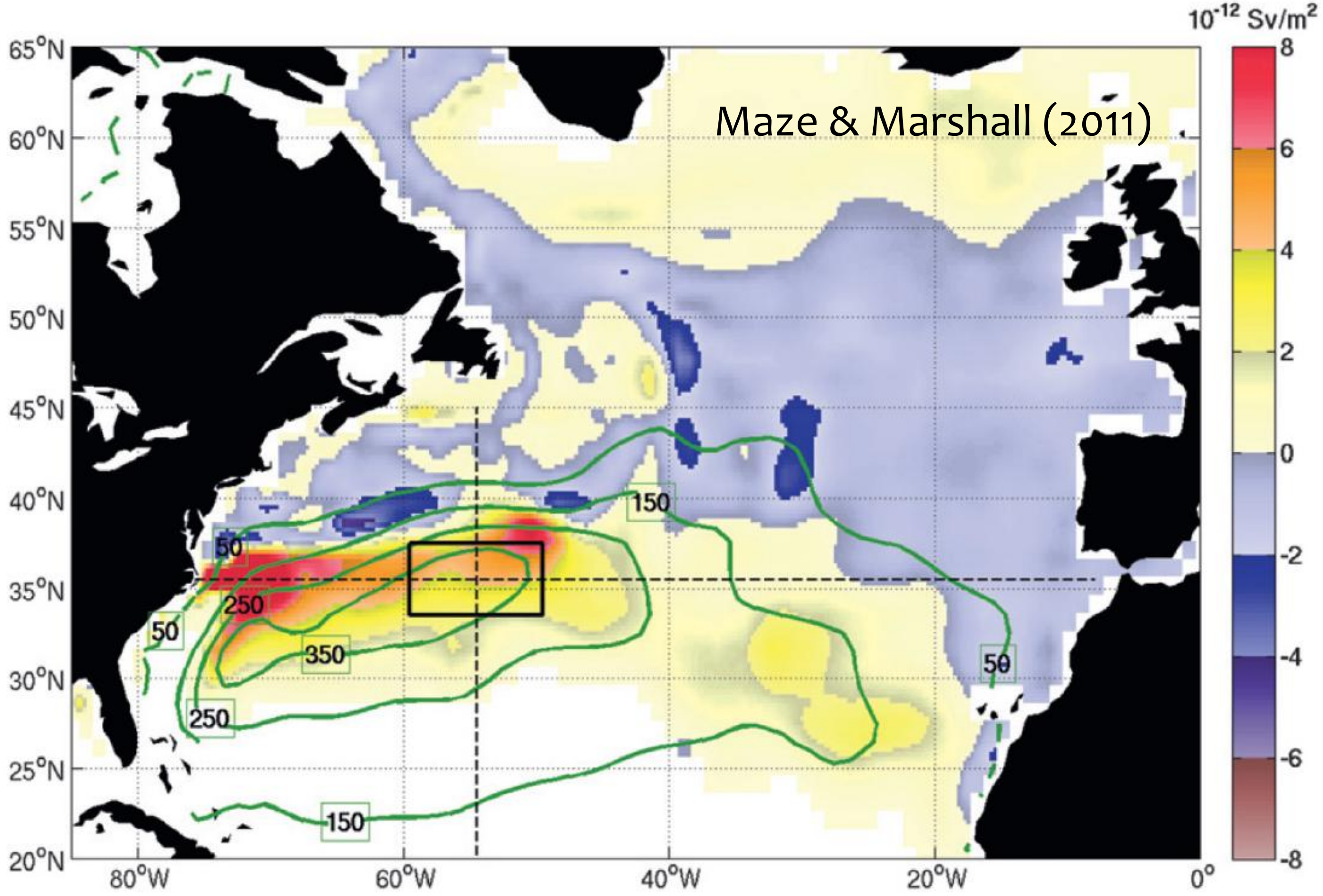
Example Applications

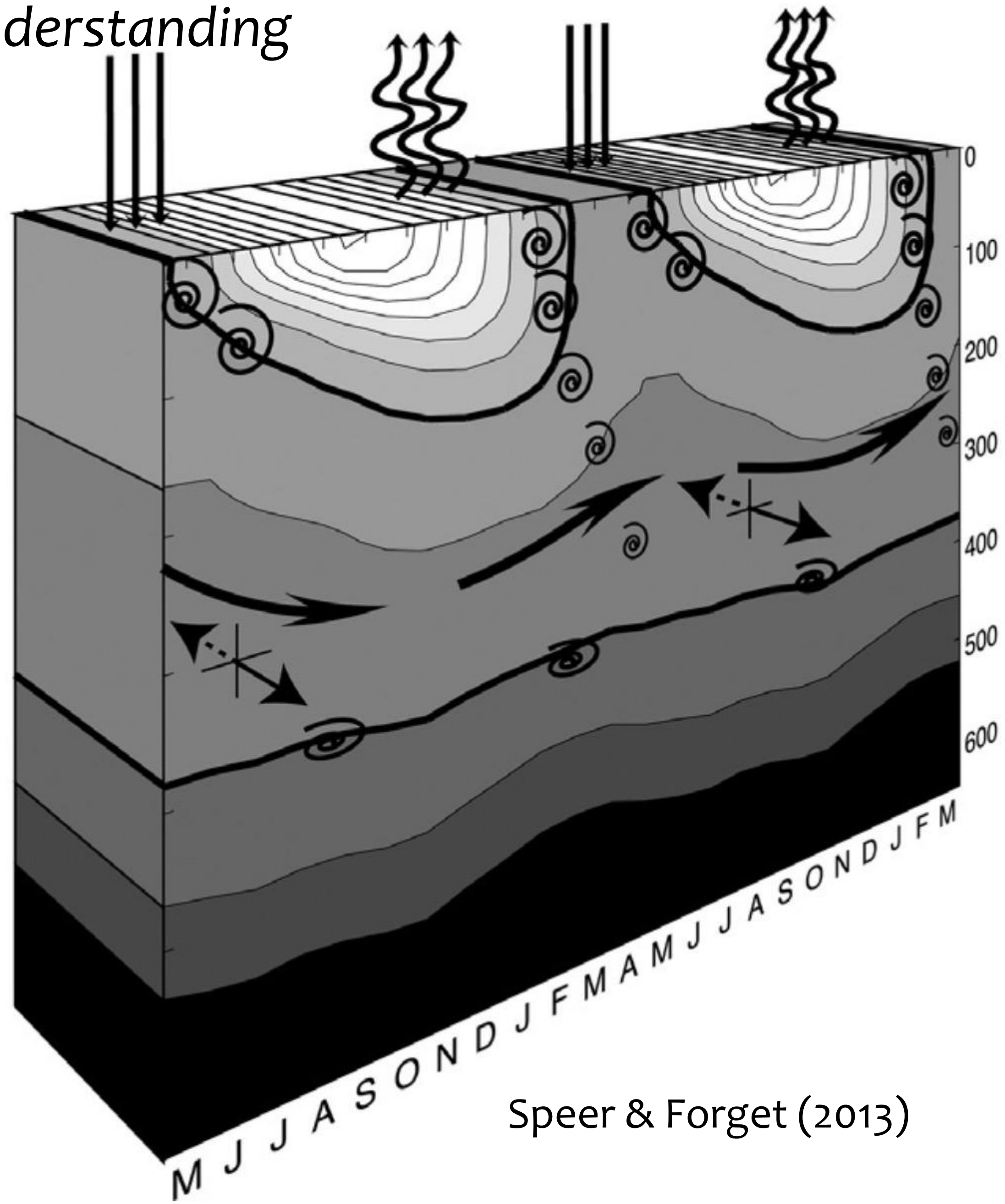
Gaining insight through quantifying time-evolving property budgets

Example use of state estimates:

Use of observations-only vs. state estimates for understanding

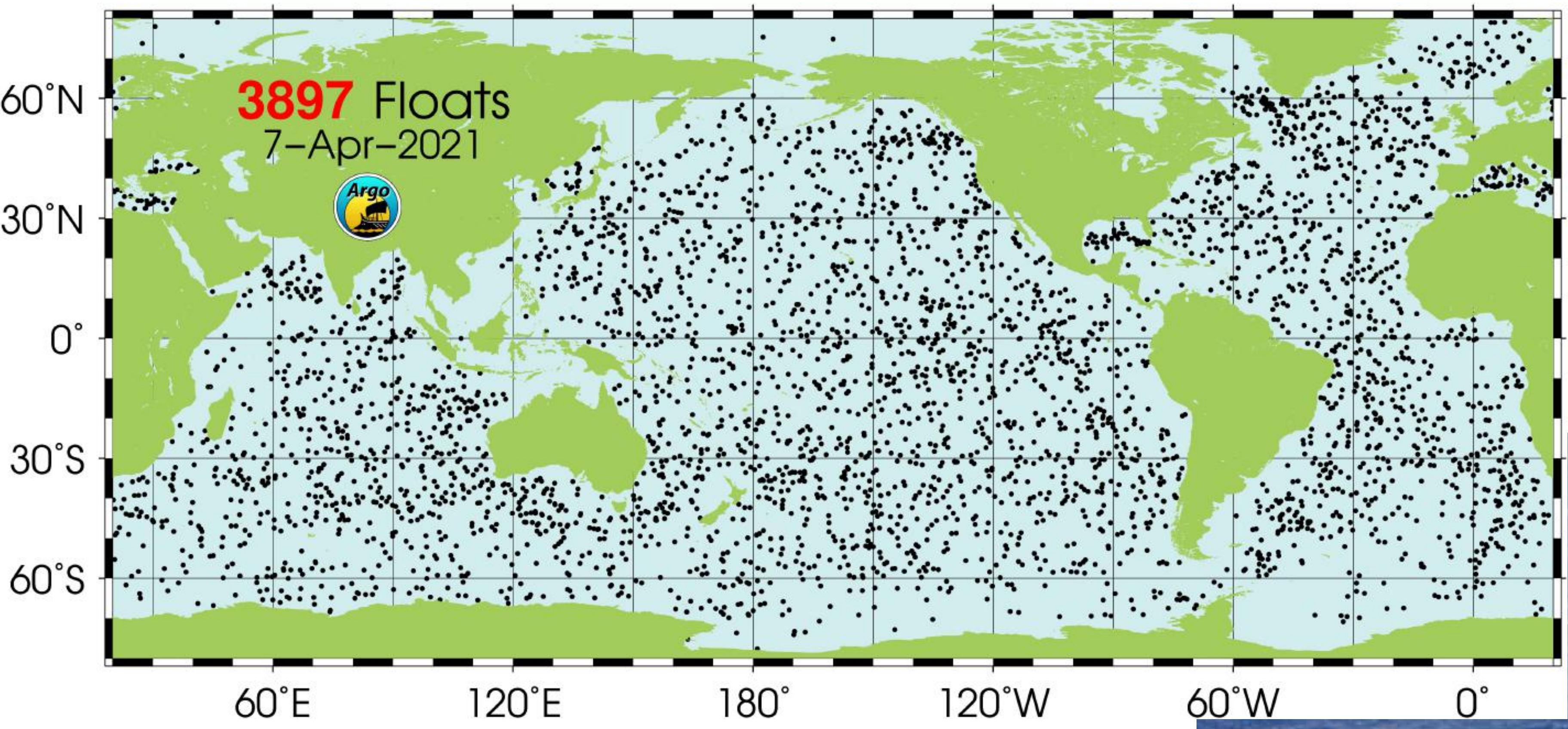
Dynamics & variability of North Atlantic (Eighteen-degree) Mode Water formation

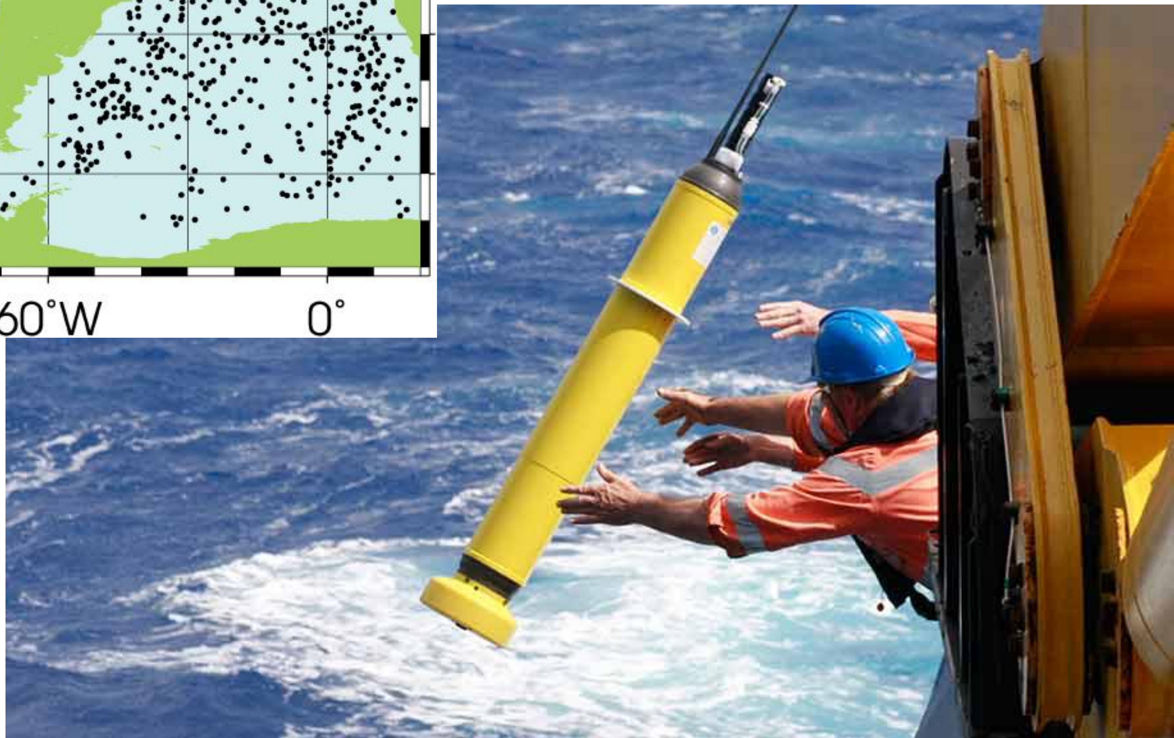




Example use of state estimates:

Use of observations-only vs. state estimates for understanding





The global array of Argo profiling floats

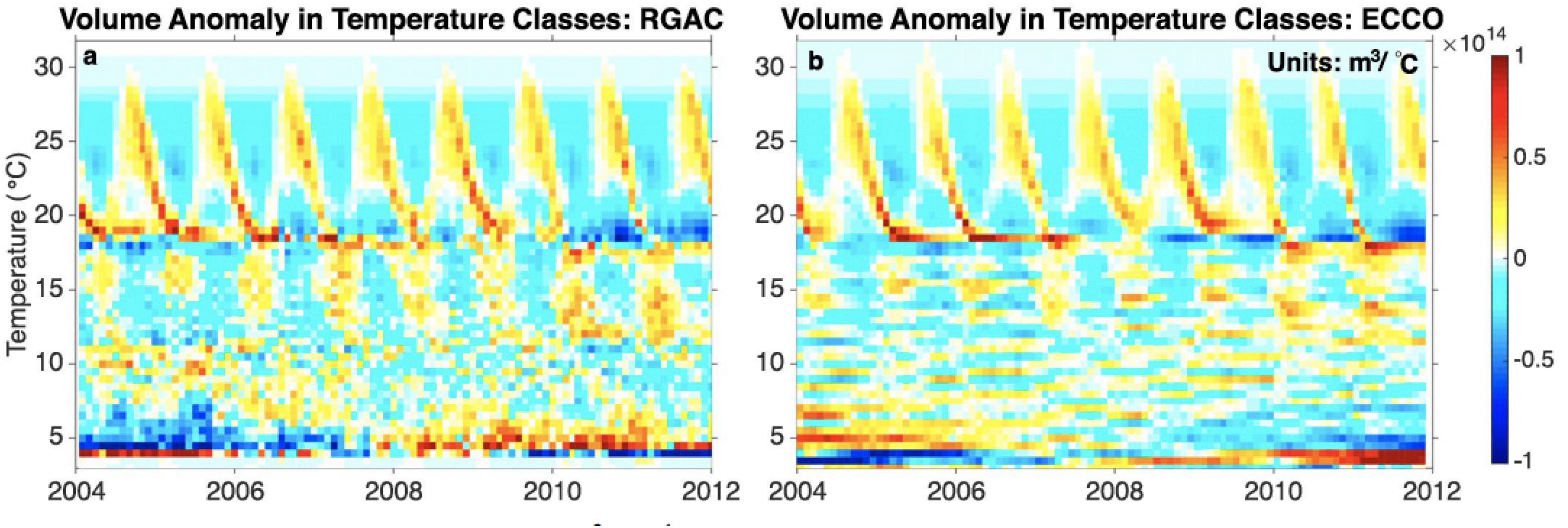
http://www.argo.ucsd.edu

Example use of state estimates:

Use of observations-only vs. state estimates for understanding

Diabatic and adiabatic contributions to water mass volume variability in the North Atlantic subtropical gyre (26°N – 45°N)

Evans et al., JPO (2017)



Roemmich & Gilson, 2009 (Argo) **ECCO version 4**

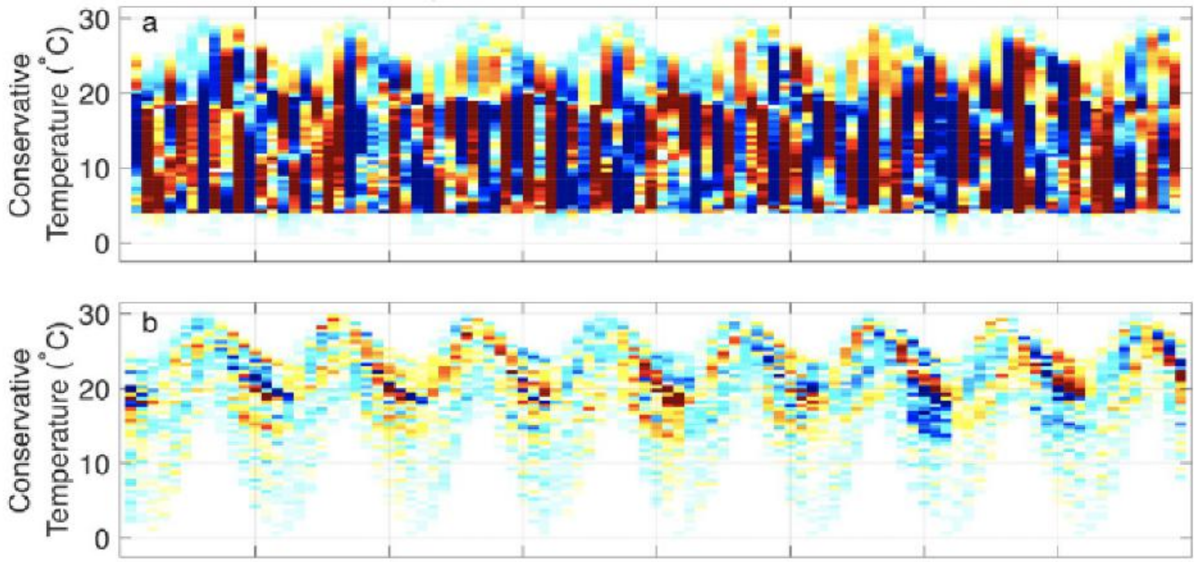
Example use of state estimates:

Use of observations-only vs. state estimates for understanding

Diabatic and adiabatic contributions to water mass volume variability in the North Atlantic subtropical gyre (26°N – 45°N)

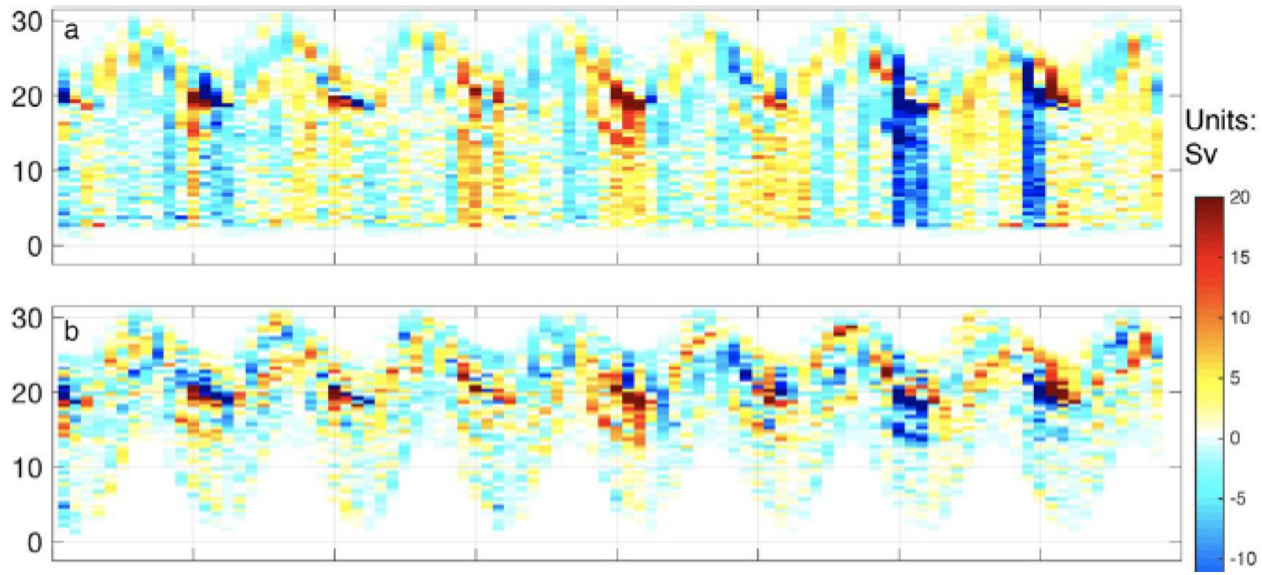
Evans et al., JPO (2017)

Total monthly ***dV/dt*****, Argo**



Monthly total & diathermal transformation due to air–sea heat fluxes NCEP + Reynolds SST

Total monthly ***dV/dt*****, ECCO v4**



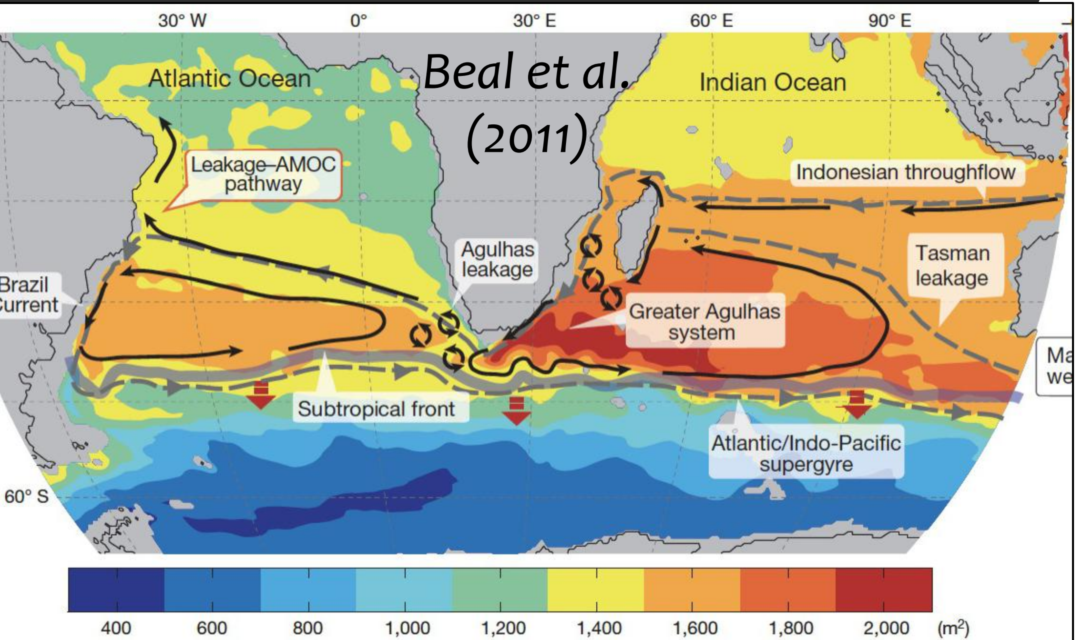
Monthly total & diathermal transformation due to air–sea heat fluxes ECCO v4

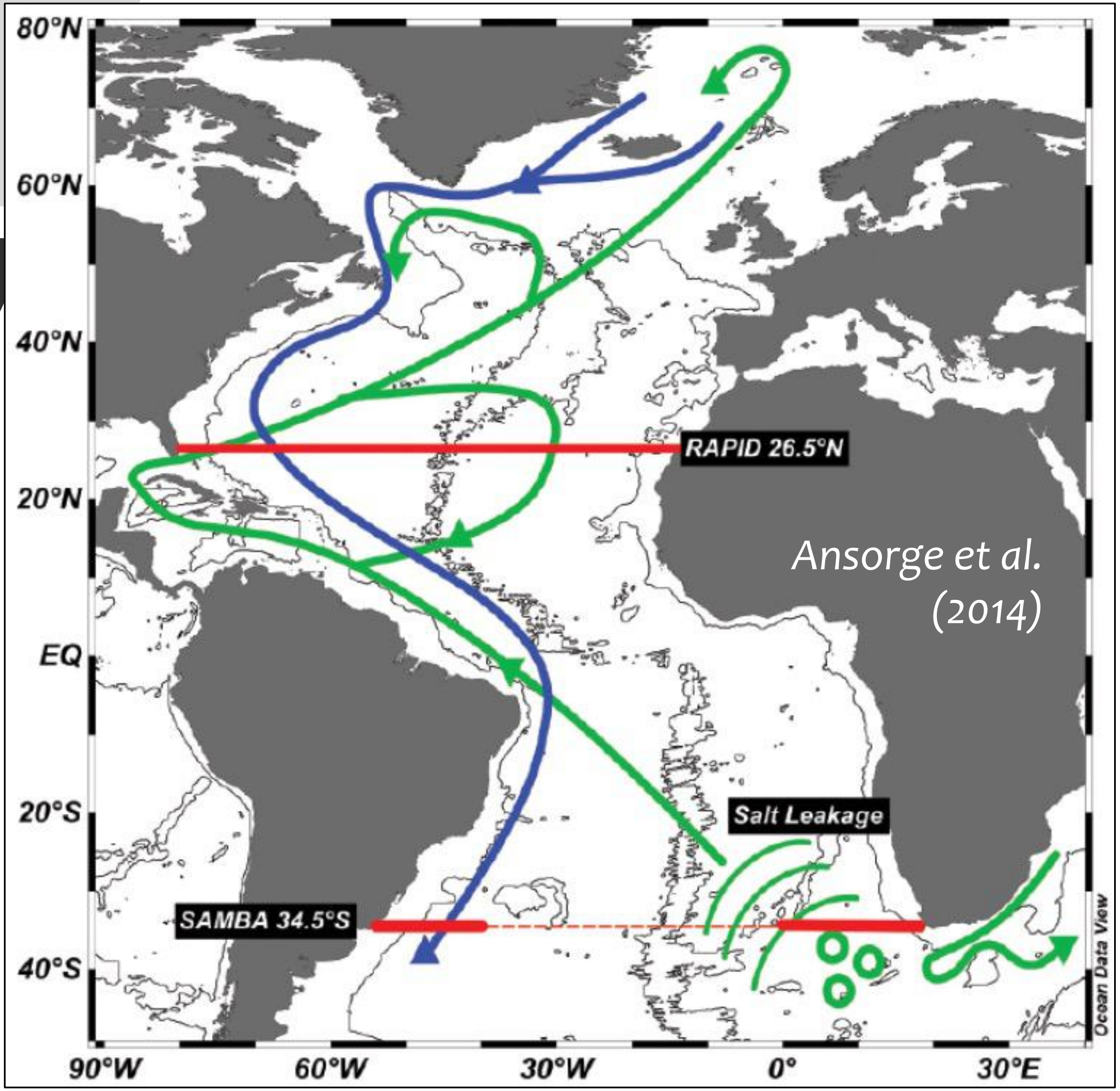
Beyond optimization

Science applications with the adjoint model: sensitivity maps & dynamical attribution

Causal / dynamical attribution: South Atlantic Ocean Circulation

South Atlantic Meridional Overturning Circulation (SAMOC) variability

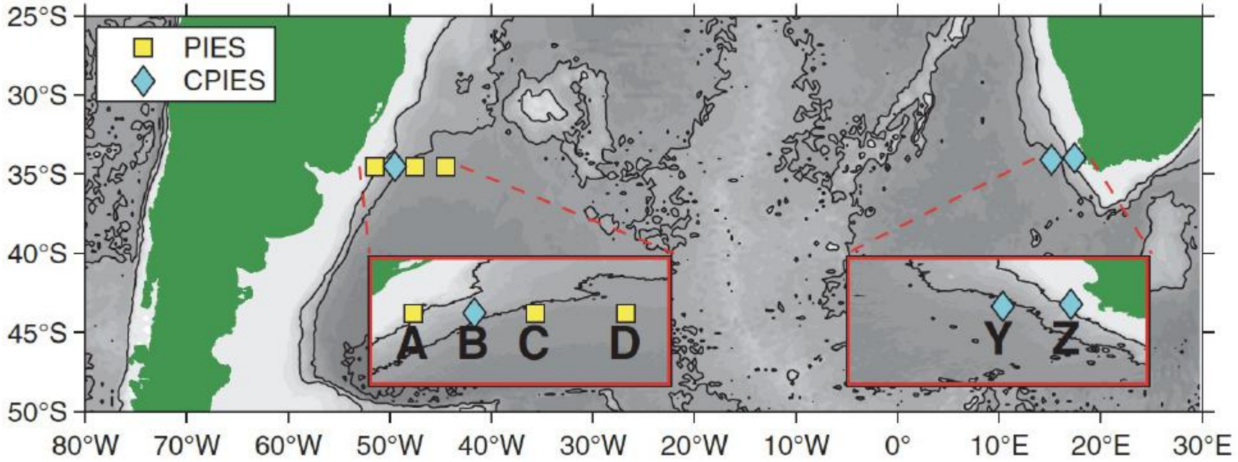




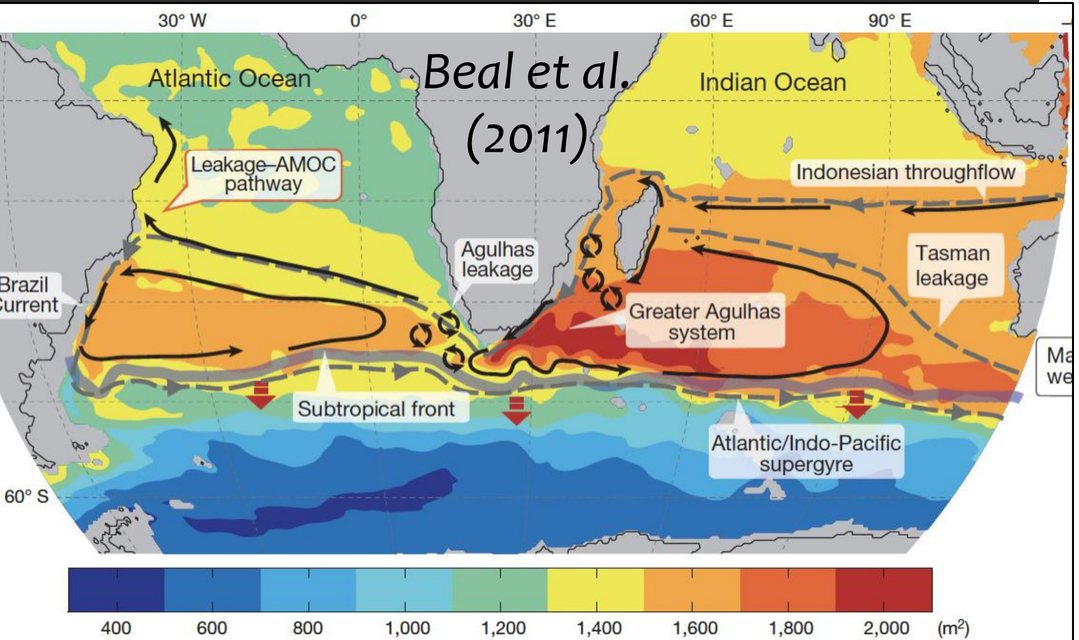
Ansorge et al. (2014)

Causal / dynamical attribution: South Atlantic Ocean Circulation

Smith & Heimbach, J. Clim. (2019)



South Atlantic Meridional Overturning Circulation (SAMOC) variability



Beal et al. (2011)

Quantity of interest:

$$\delta\mathcal{J}(u(x,y,t)) \equiv \text{Monthly AMOC Anomaly @ } 34°\text{S}$$

"controlled" by:

$$\delta u(x,y,t) \equiv \text{Surface Atm. Forcing Perturbations}$$

through (assumed) linear dynamics described by:

$$\frac{\partial \mathcal{J}}{\partial u}(x,y,t) \equiv \text{Sensitivity}$$

Causal / dynamical attribution: South Atlantic Ocean Circulation

Smith & Heimbach, J. Clim. (2019)

Sensitivity of mass transport J with respect to forcings, initial conditions, carried via "backpropagation" by the time-evolving dual/adjoint state

$$\mu_0 = \frac{\partial J}{\partial x_0} = \sum_{1 \le t \le t_f} \frac{\partial x_t}{\partial x_0} \left(\frac{\partial J}{\partial x_t} \right)$$

$$= \frac{\partial x_1}{\partial x_0} \left(\frac{\partial J}{\partial x_1} \right) + \frac{\partial x_1}{\partial x_0} \frac{\partial x_2}{\partial x_1} \left(\frac{\partial J}{\partial x_2} \right) + \ldots + \frac{\partial x_1}{\partial x_0} \cdots \frac{\partial x_{t_f}}{\partial x_{t_f - 1}} \left(\frac{\partial J}{\partial x_{t_f}} \right)$$

$$= \mathbf{L}^T \frac{\partial J}{\partial x_1} + \mathbf{L}^T \mathbf{L}^T \frac{\partial J}{\partial x_2} + \ldots + \mathbf{L}^T \cdots \mathbf{L}^T \frac{\partial J}{\partial x_{t_f}}$$

$\mathbf{L}^T$: is the **adjoint model** (and $\mathbf{L}$ is the **tangent linear model**)

$\mu_k = \left(\frac{\partial J}{\partial x_k} \right)$: **Lagrange multipliers** or **gradients**

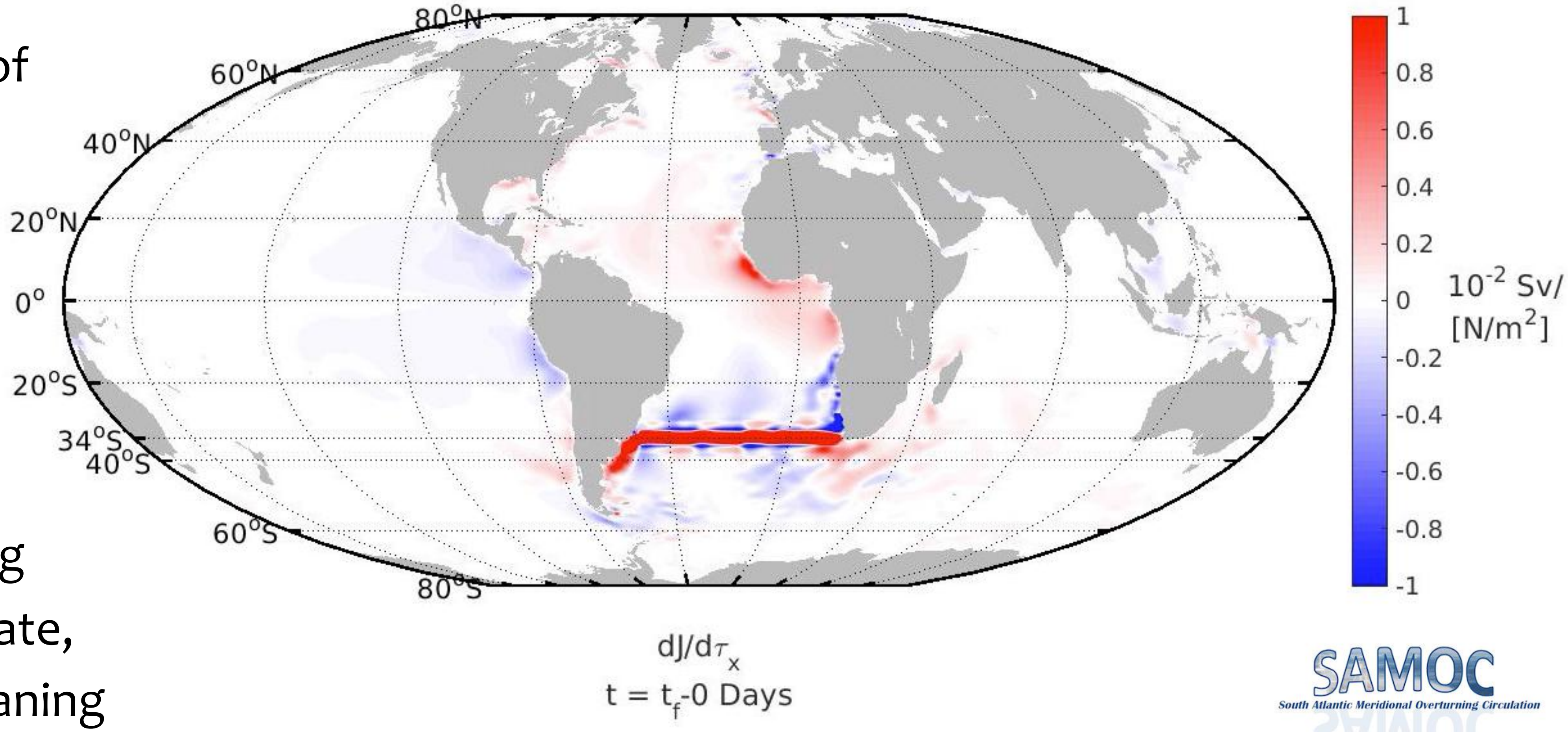
Smith & Heimbach, J. Clim. (2019)

Causal / dynamical attribution:
The South Atlantic Ocean Circulation

- Use of the availability of the *dual* ocean state (time-evolving adjoint state) for scientific analysis.

- Reconstruction of AMOC at 33°S from forcing anomalies via the adjoint;

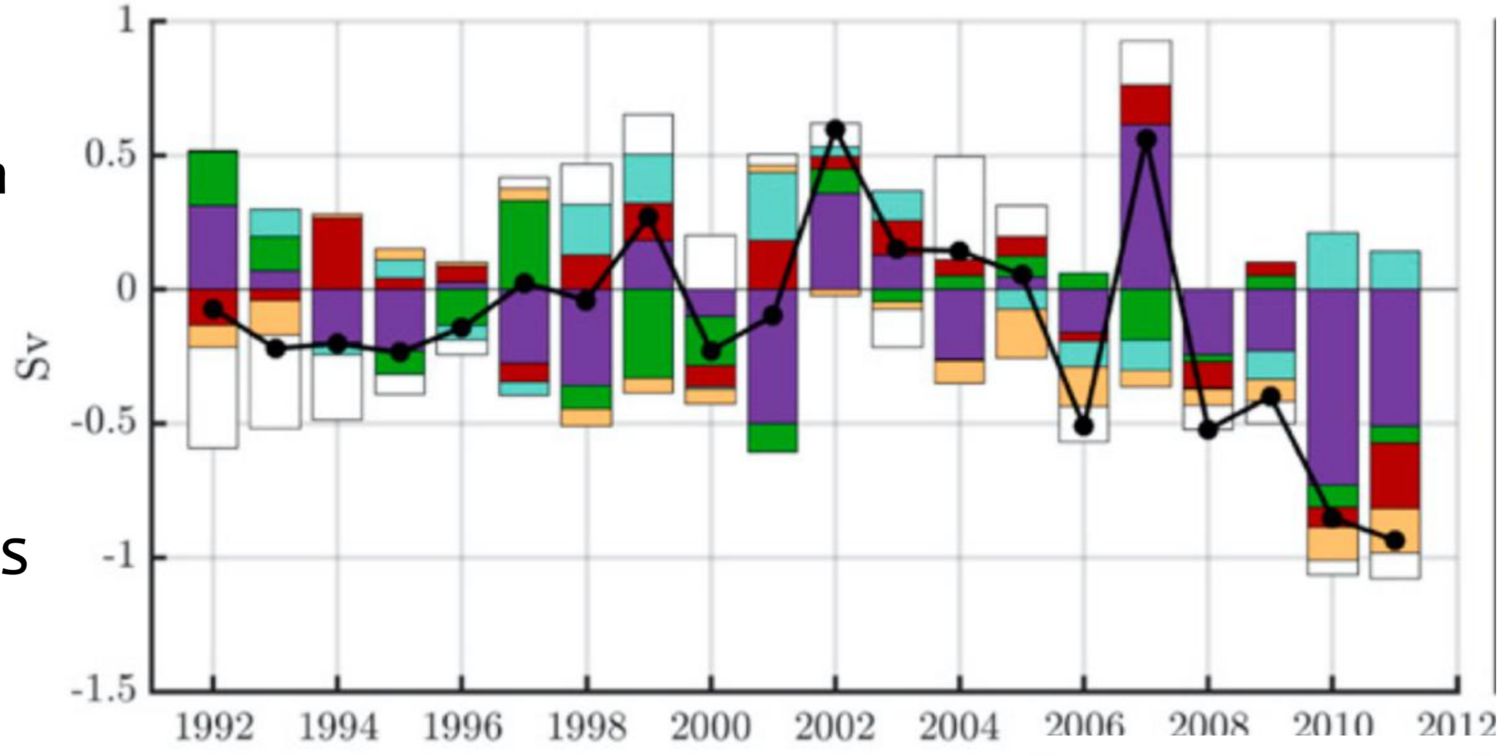
- The time-evolving adjoint = dual state, has physical meaning



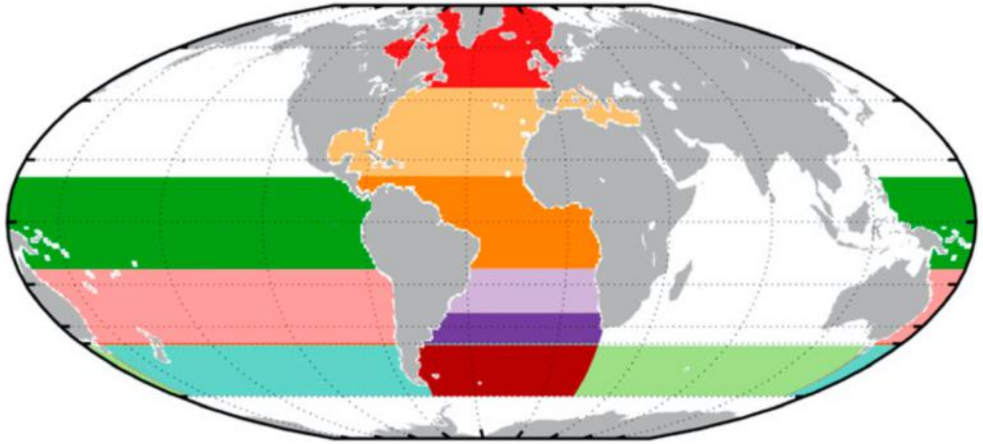
SAMOC
South Atlantic Meridional Overturning Circulation

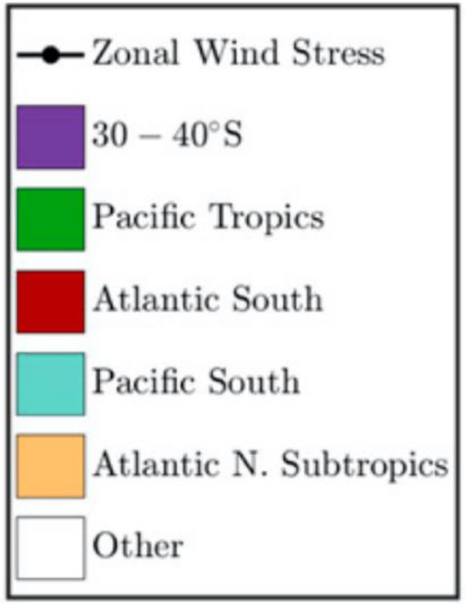
Causal / dynamical attribution:
The South Atlantic Ocean Circulation

Dynamic attribution of interannual SAMOC variability due to wind stress perturbations via Green's functions



Smith & Heimbach, J. Clim. (2019)





$$\delta\mathcal{J}(t) = \sum_{k}^{N_{atm}} \delta\mathcal{J}_k(t)$$

$$= \sum_{k}^{N_{atm}} \int_{t_0}^{t} \int_x \int_y \frac{\partial \mathcal{J}}{\partial u_k}(x, y, \tau - t_f)\, \delta u_k(x, y, \tau)\, dxdyd\tau$$

Beyond optimization

Hessian-based Uncertainty Quantification & formal observing system design

What is Uncertainty Quantification in the context of Data Assimilation?

Different uses:

- *Parameter estimation (<u>calibration</u>)*
  - **uncertainties in the parameters**
- *State estimation / reconstruction / synthesis (<u>interpolation</u>)*
  - **uncertainties in reconstructed state or derived quantities of interest (QoIs)**
- *Forecast initialization (<u>extrapolation</u>)*
  - **uncertainties in the forecast**

Origins of uncertainty in the context of DA

- **Observations:**
  - measurement error & irregular sampling (in space and time)
- **Assimilation scheme:**
  - DA algorithm (and approximations)
  - how observations are ingested in DA
- **Model:**
  - Parametric uncertainties
  - Structural uncertainties (discretization, model inadequacy)
- **All boundary conditions:**
  - external forcing, bathymetry, lateral boundaries
- **Use of prior knowledge:**
  - error covariances, representation error

Why is uncertainty quantification difficult to do?

- **Spatio-temporally irregular sampling of observations**
  - How to account for inhomogeneous sampling uncertainty,
- **Parametric uncertainty**
  - "The curse of dimensionality"
  - N^2 for parameter space dimension N
- **Structural uncertainty / model inadequacy**
  - How to capture? Source terms in the model?
  - Absorb within parameter estimation?
  - Machine learning / neural networks as surrogate models?
  - Stochasticity

A way forward: Bayesian inverse methods

computational frameworks that account *jointly* for model, observations, forcings, prior knowledge, and *all* their uncertainties:

- **MCMC and similar sampling techniques** hard for high-dim. parameter spaces
- **Derivative-based inference**:
  - propagates *all* uncertainties
  - provides dynamical underpinning of uncertainty propagation
  - infers low-order modes

Approximation of the posterior PDF by posterior error covariance Γ_{post}

Posterior error covariance given by inverse of the Hessian at $\mathbf{u}_{MAP}$

$$\Gamma_{post} = \left[H_{misfit}(\mathbf{u}_{MAP}) + \Gamma_{prior}^{-1} \right]^{-1}$$

where

$$H_{misfit}(\mathbf{u}_{MAP}) \approx F^T \, \Gamma_{obs}^{-1} \, F$$

Gauss-Newton Hessian approximation

with:

F : tangent linear operator of $\mathcal{F}$
F^T : adjoint operator of $\mathcal{F}$

How to obtain/extract information from Γ_{post}

- form prior-preconditioned Hessian

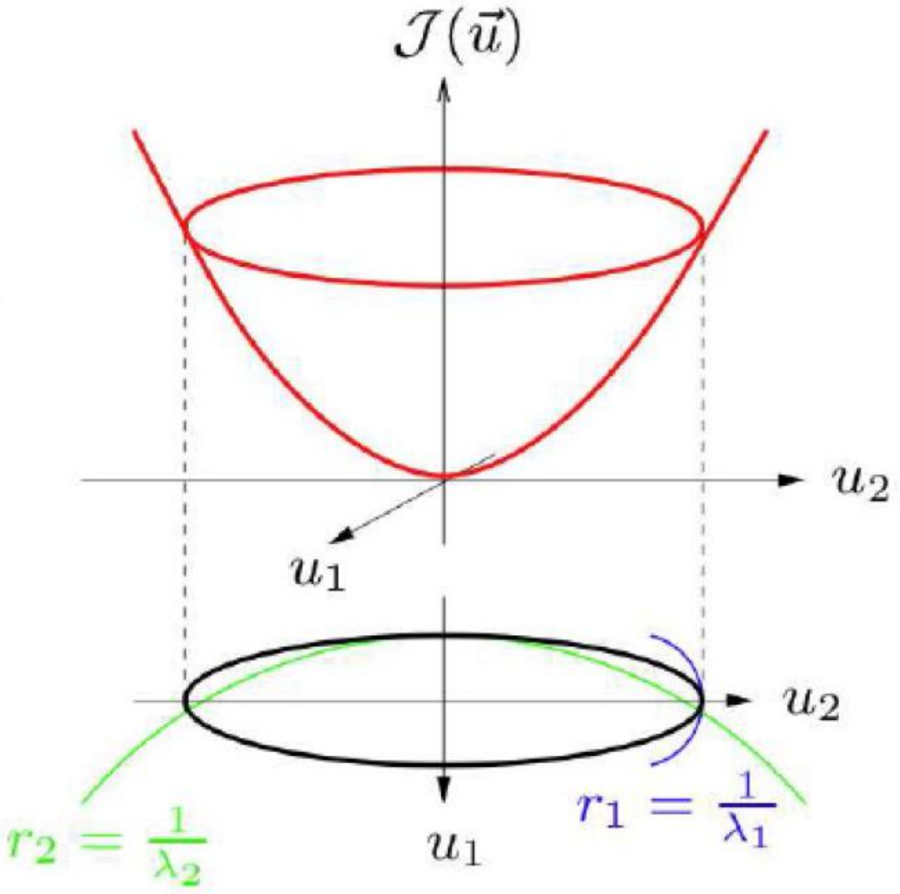
$$\tilde{H}_{misfit} = \Gamma_{prior} H_{misfit} ,$$

- formulate generalized eigenvalue problem

$$\tilde{H}_{misfit} W = W \Lambda$$

where

W	:	eigenvector matrix
$\Lambda = \mathrm{diag}(\lambda_i)$	:	eigenvalue matrix
$V = \Gamma_{prior}^{1/2} W$	:	prior-preconditioned eigenvector matrix



How to obtain/extract information from Γ_{post}

- retain only r largest eigenvalues, yields low-rank approximation:

$$\tilde{H}_{misfit} \approx V_r \Lambda_r V_r^T$$

$$\left[\tilde{H}_{misfit} + I \right]^{-1} \approx I - V_r D_r V_r^T = R_r$$

with low-rank resolution operator R_r and

$$D_r = \mathrm{diag}(\lambda_r / (\lambda_r + 1))$$

Low-rank eigen-decomposition of Γ_{post}

$$\Gamma_{post} = \Gamma_{prior} - \sum_{i=1}^{r} \frac{\lambda_i}{\lambda_i + 1} \left(\Gamma_{prior}^{1/2} w_i \right) \left(\Gamma_{prior}^{1/2} w_i \right)^T$$

Hessian uncertainty quantification …

… applied to observing system design in the North Atlantic

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

10.1029/2020MS002386

Leveraging Uncertainty Quantification to Design Ocean Climate Observing Systems

Nora Loose[1] and **Patrick Heimbach**[1,2,3]

[1]Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, USA, [2]Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, USA, [3]Institute for Geophysics, The University of Texas at Austin, Austin, TX, USA

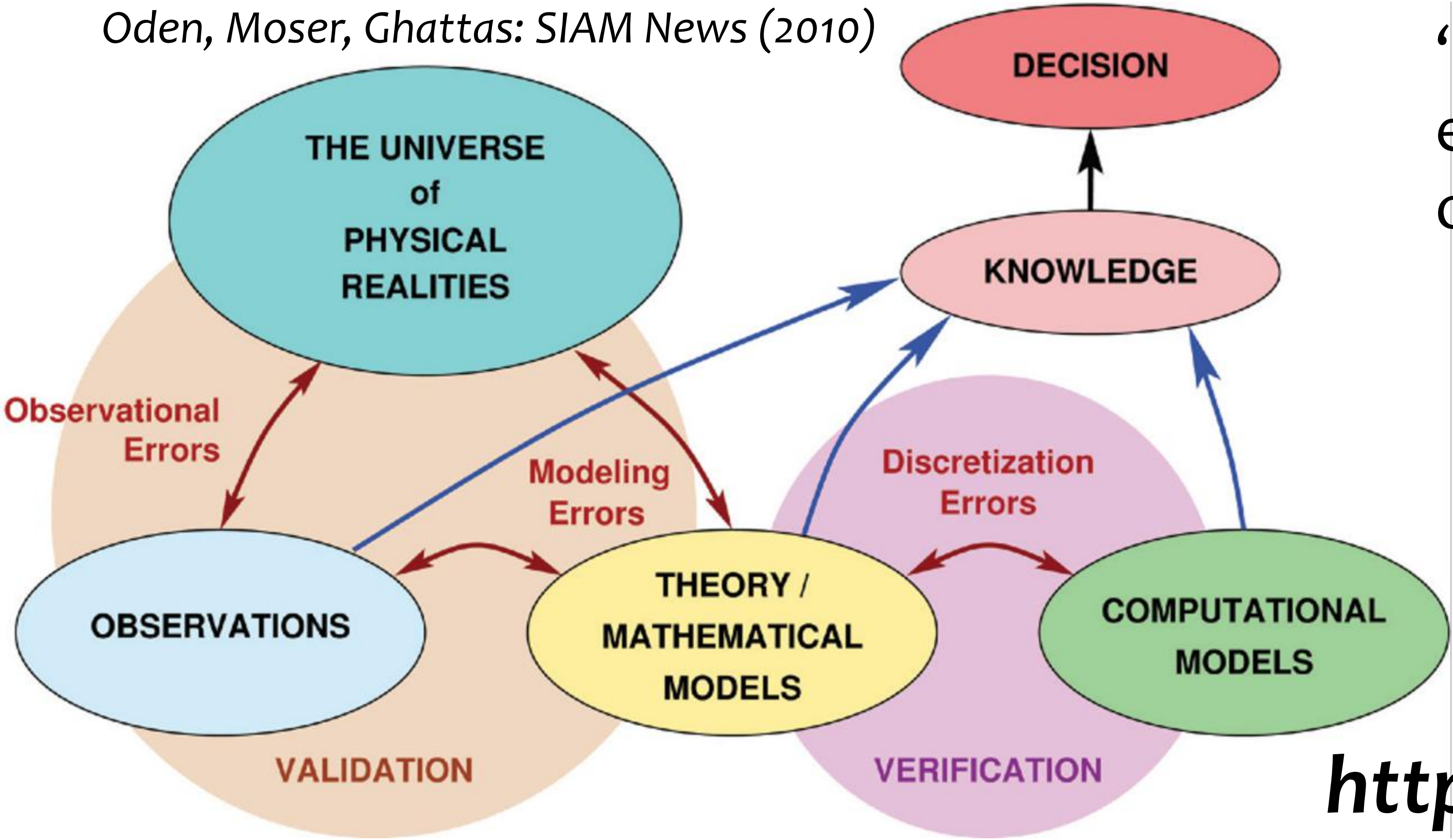
Key Points:

- We apply Hessian uncertainty quantification (UQ) to the global ocean state estimate ECCO, and explore its use for observing system design
- Hessian UQ elucidates oceanic teleconnections that communicate observational constraints over basin-scale distances
- Going beyond previous adjoint ocean modeling techniques, Hessian UQ rigorously assesses redundancy and optimality of observing systems

Abstract Ocean observations are expensive and difficult to collect. Designing effective ocean observing systems therefore warrants deliberate, quantitative strategies. We leverage adjoint modeling and Hessian uncertainty quantification (UQ) within the ECCO (Estimating the Circulation and Climate of the Ocean) framework to explore a new design strategy for ocean climate observing systems. Within this context, an observing system is optimal if it minimizes uncertainty in a set of investigator-defined quantities of interest (QoIs), such as oceanic transports or other key climate indices. We show that

Predictive data science:
from observation & simulation to decision

Oden, Moser, Ghattas: SIAM News (2010)



"Data assimilation is essentially an automation of the scientific method"

S. Penny et al.,

Front. Mar. Sci. (2019)

https://crios-ut.github.io

DILBERT BY SCOTT ADAMS