ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

DIFFERENTIABLE PROGRAMMING FOR HYBRID DATA ASSIMILATION & MACHINE LEARNING

Patrick Heimbach, William Moses, Joseph Kump, Valentin Churavy, Cheng Gong, Sarah Williamson, Dhruv Apte, Paul Berg, Mose Giordano, Christopher Hill, Nora Loose, Alexis Montoison, Mathieu Morlighem, Sri Hari Krishna Narayanan, Avik Pal, Michel Schanen, Greg Wagner

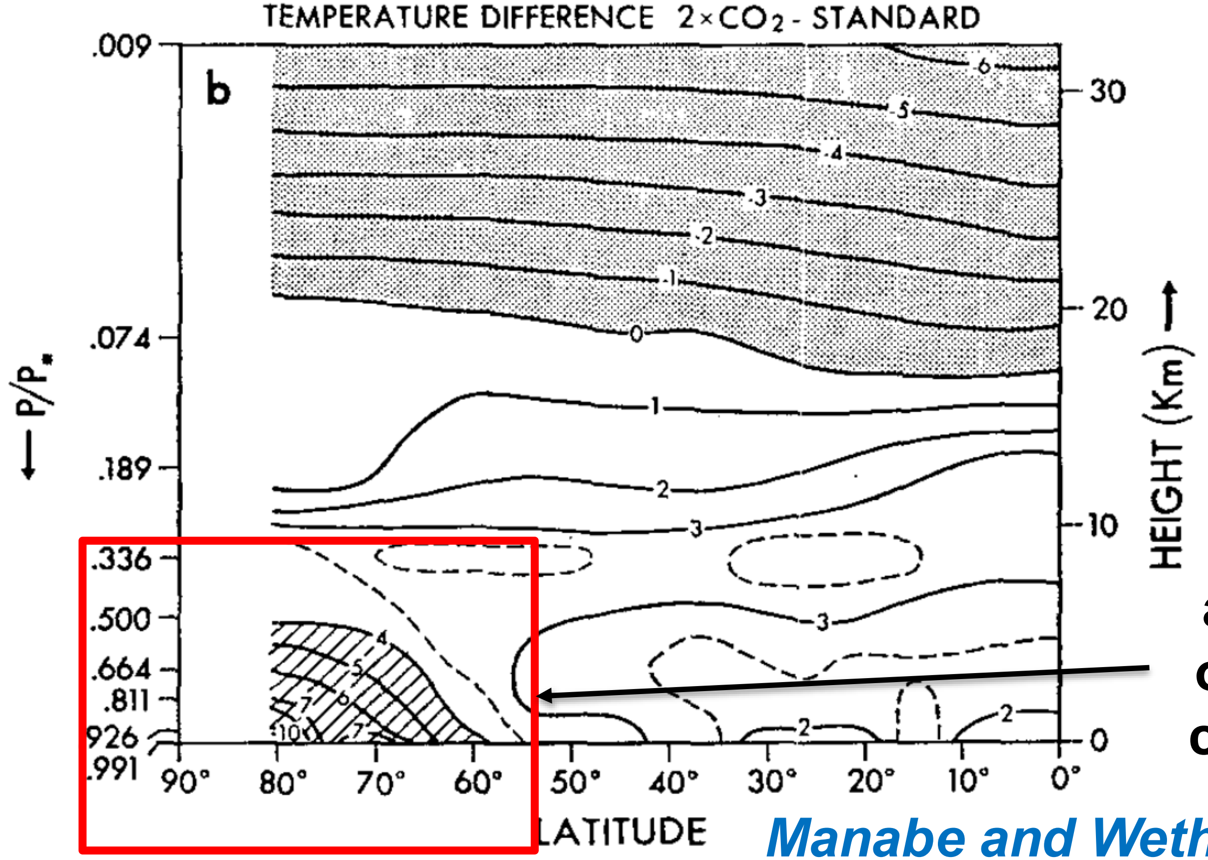
University of Texas at Austin
University of Illinois, Urbana Champaign
Dartmouth College
Massachusetts Institute of Technology
Argonne National Lab
[C]Worthy

https://ecco-group.org
https://crios-ut.github.io
https://dj4earth.github.io

Prelude: An early triumph of climate modeling

Sukyoro Manabe (NOAA/GFDL)
Nobel Prize in Physics 2021
(together with K. Hasselmann and G. Parisi):
“for the physical modelling of Earth’s climate, quantifying variability and reliably predicting global warming”

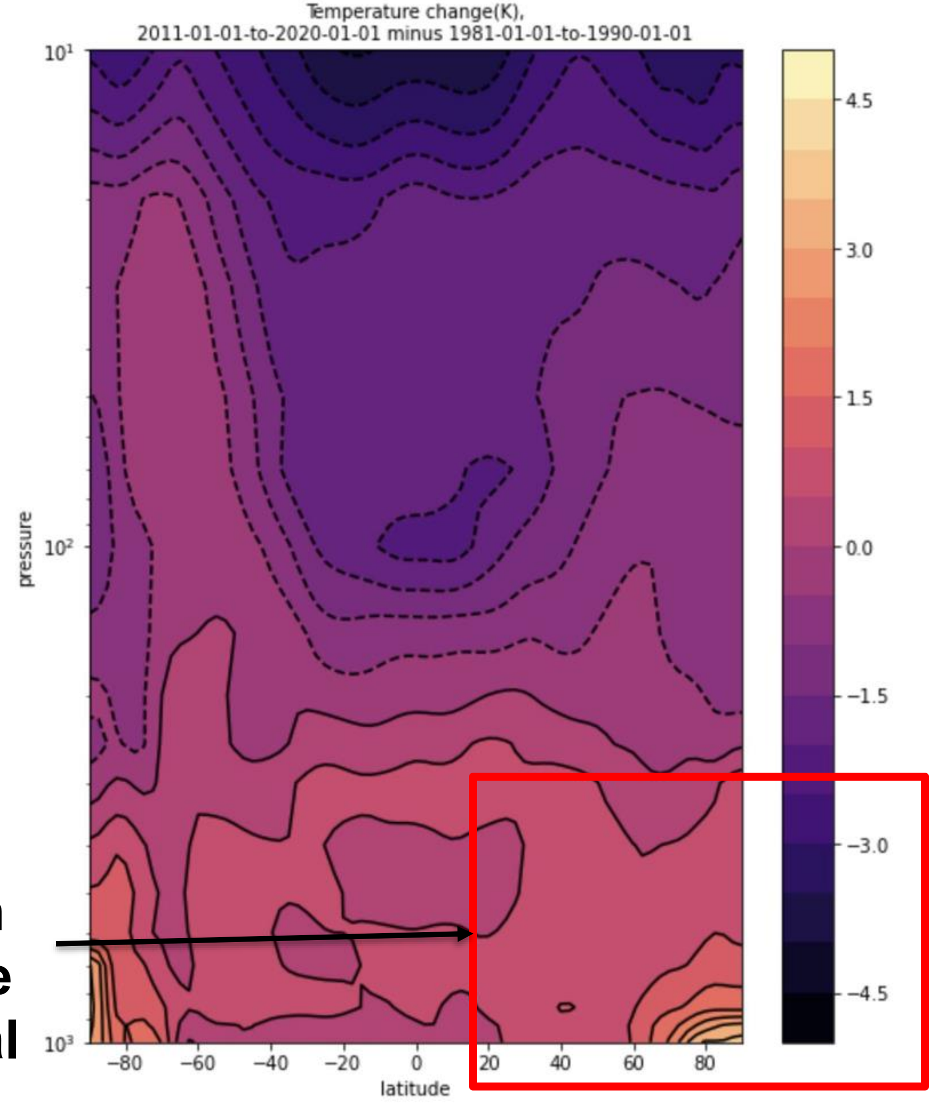
Simulated response to CO_2 doubling in the first climate model



Arctic amplification of the climate change signal

Manabe and Wetherald, JAS, 1975

Observed recent temperature change 2011-2020 minus 1981-1990



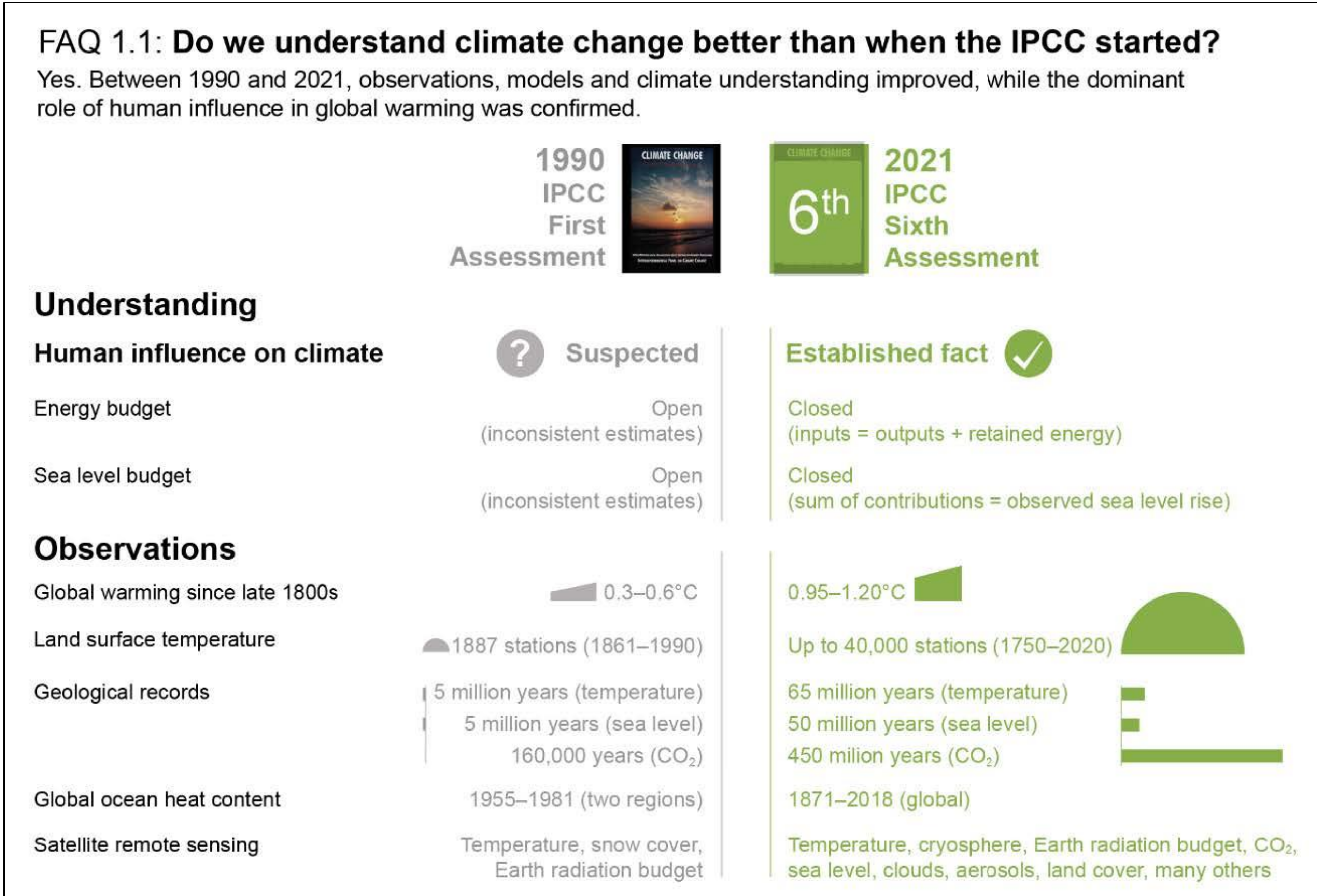
Fast-forward to 2021: IPCC AR6

Intergovernmental Panel on Climate Change (IPCC)

6th Assessment Report (2021)

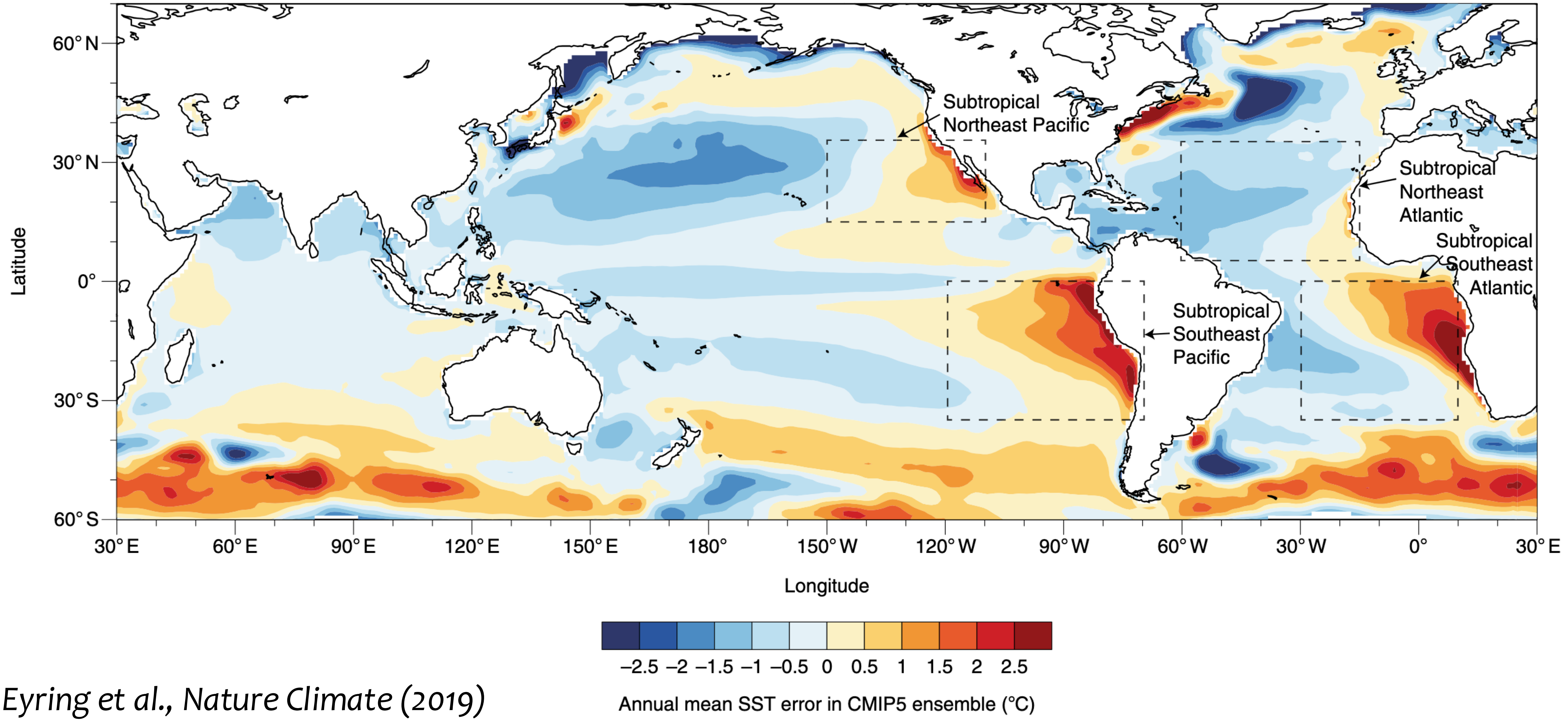
https://ipcc.ch

(historical notes in: Randall et al., Meteorol. Monogr., 2019)



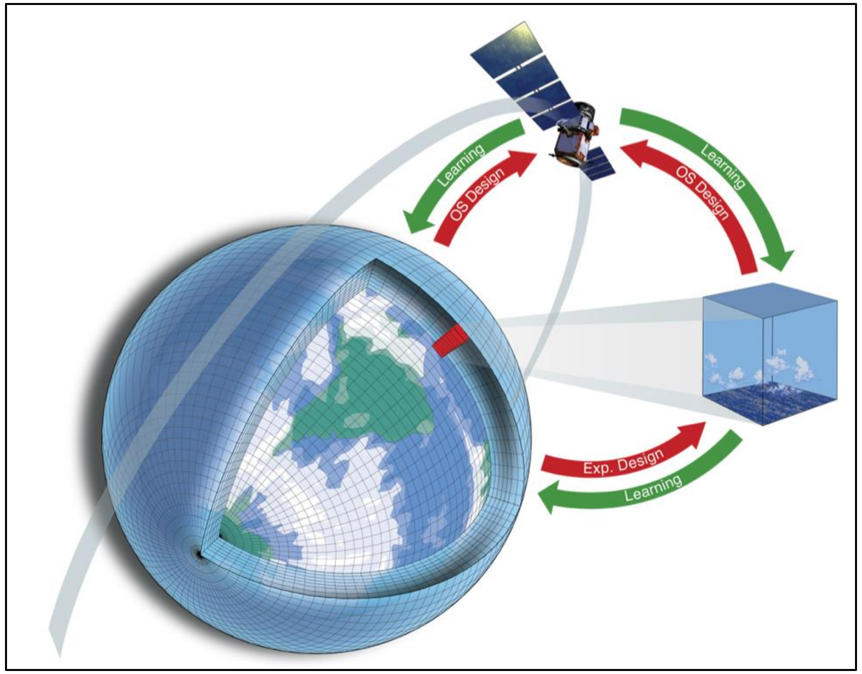
But challenges remain *in the details*: Example of CMIP model biases

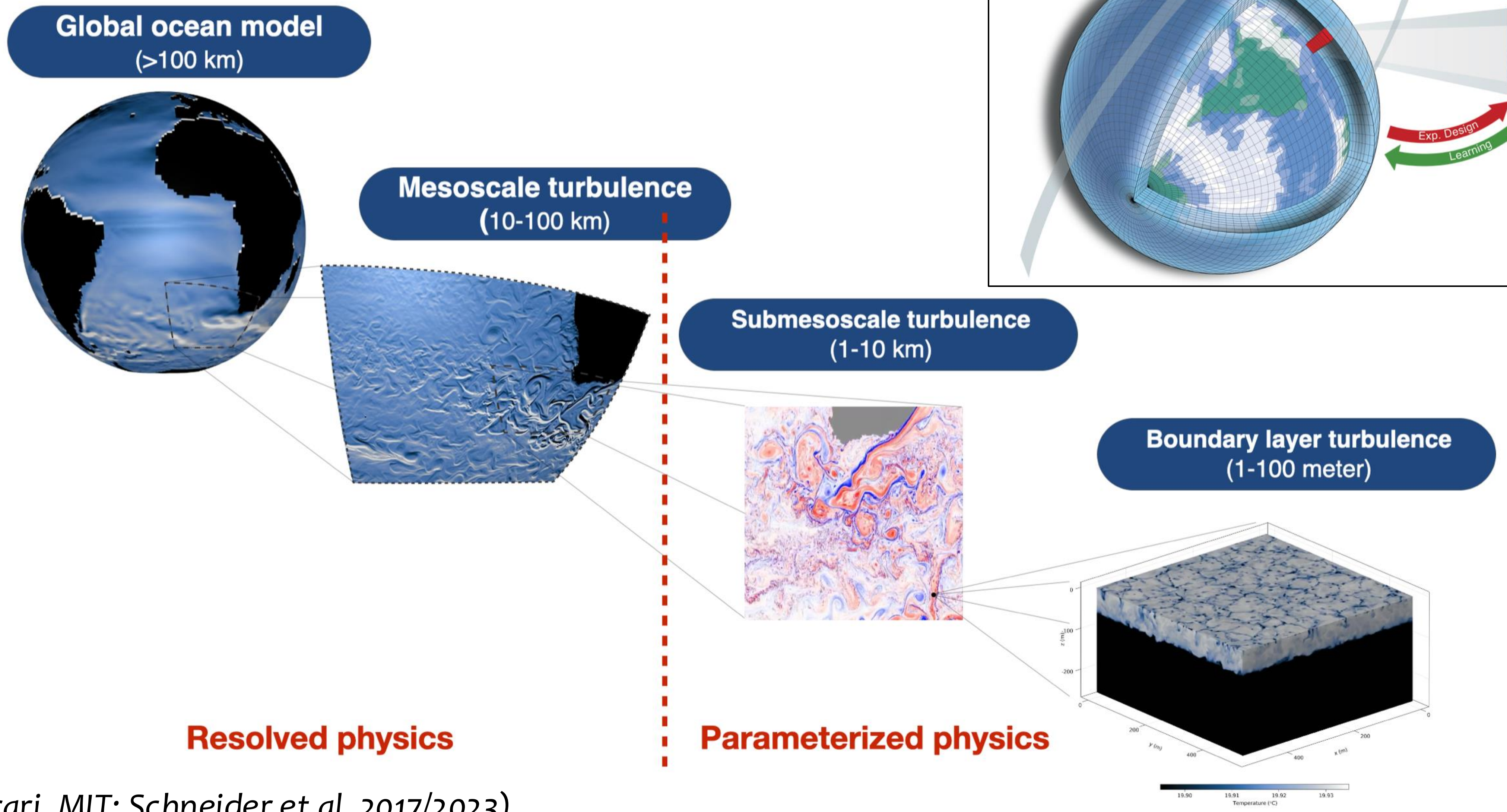
Annual mean SST error from the CMIP5 multi-model ensemble



Eyring et al., Nature Climate (2019)

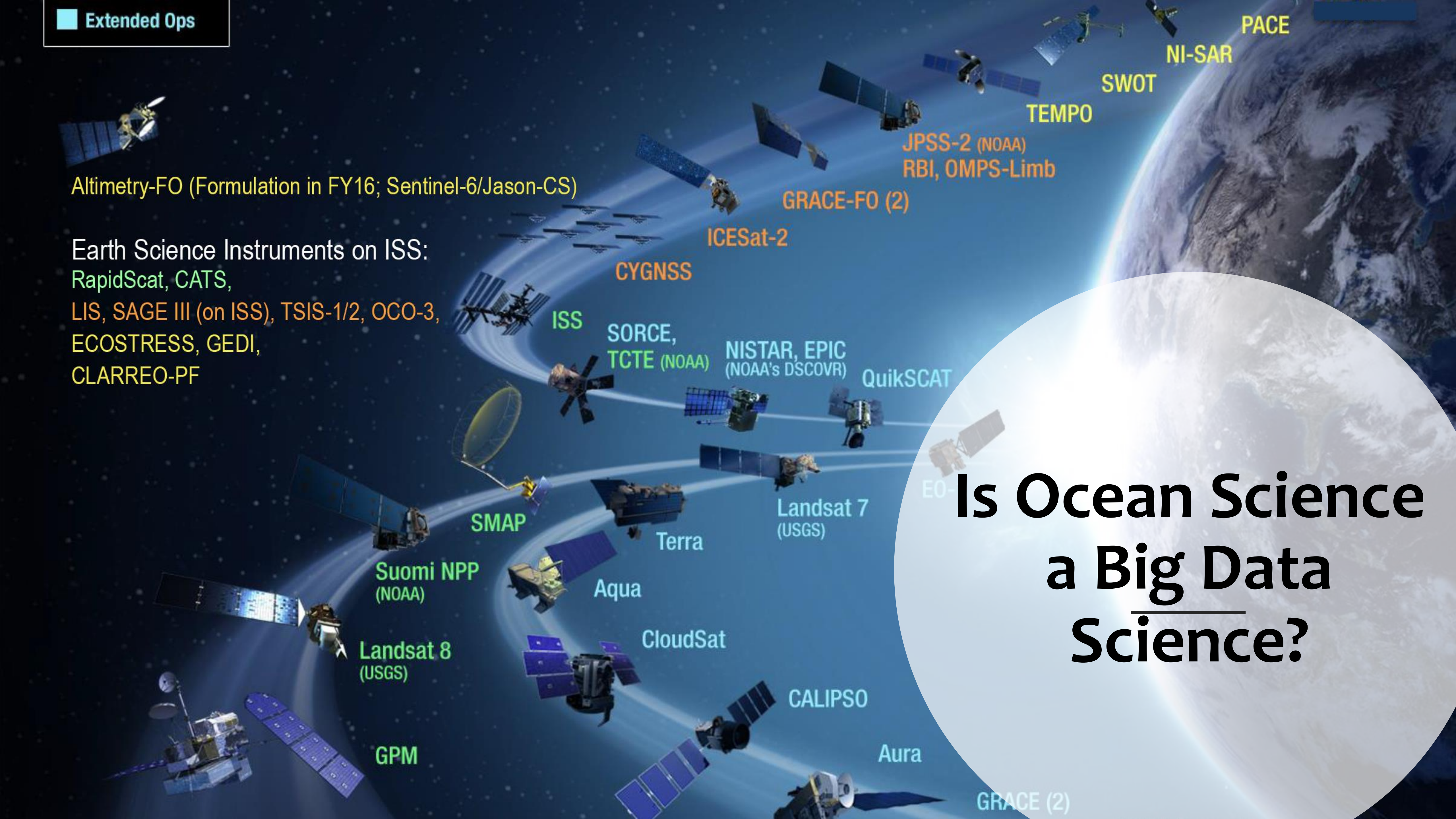
Need for parameterization & data-informed calibration in ocean (and climate) models





(R. Ferrari, MIT; Schneider et al. 2017/2023)

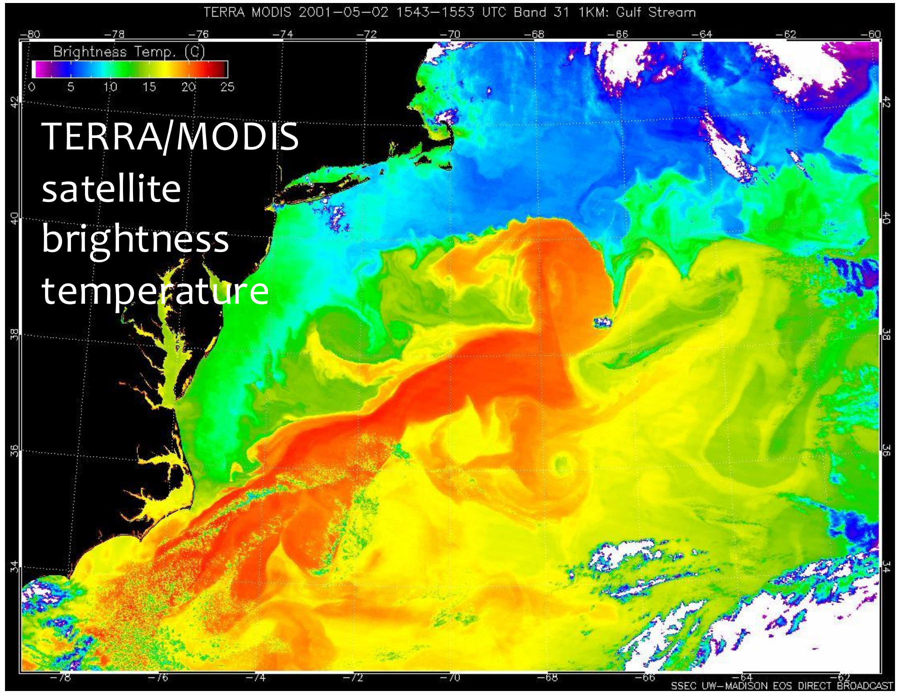
Is Ocean Science a Big Data Science?

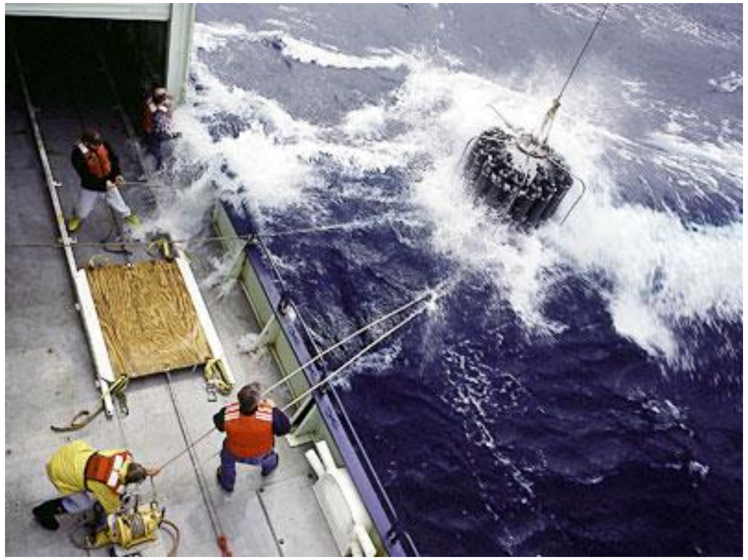


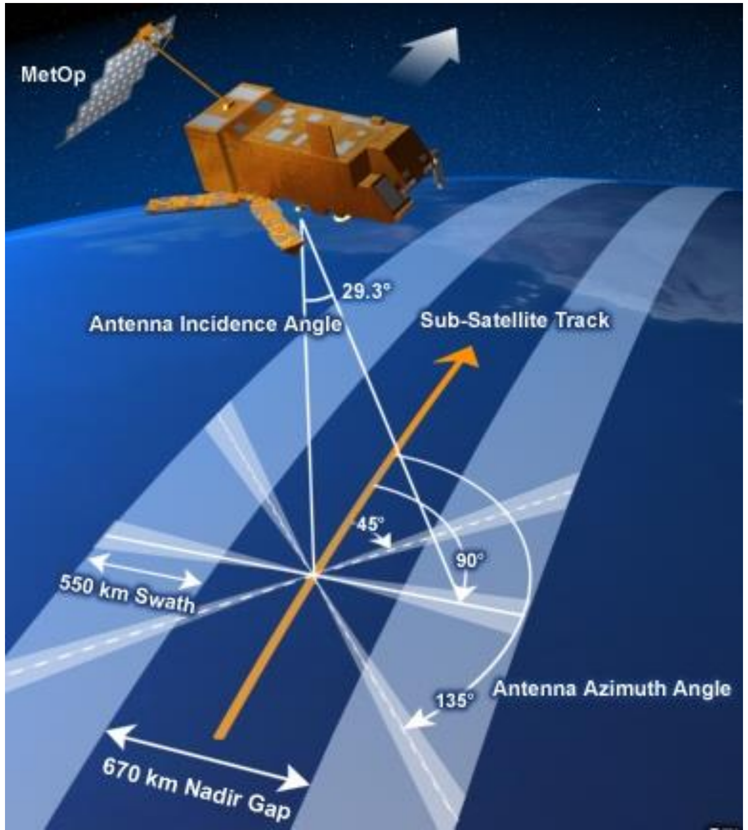
Observing challenge:

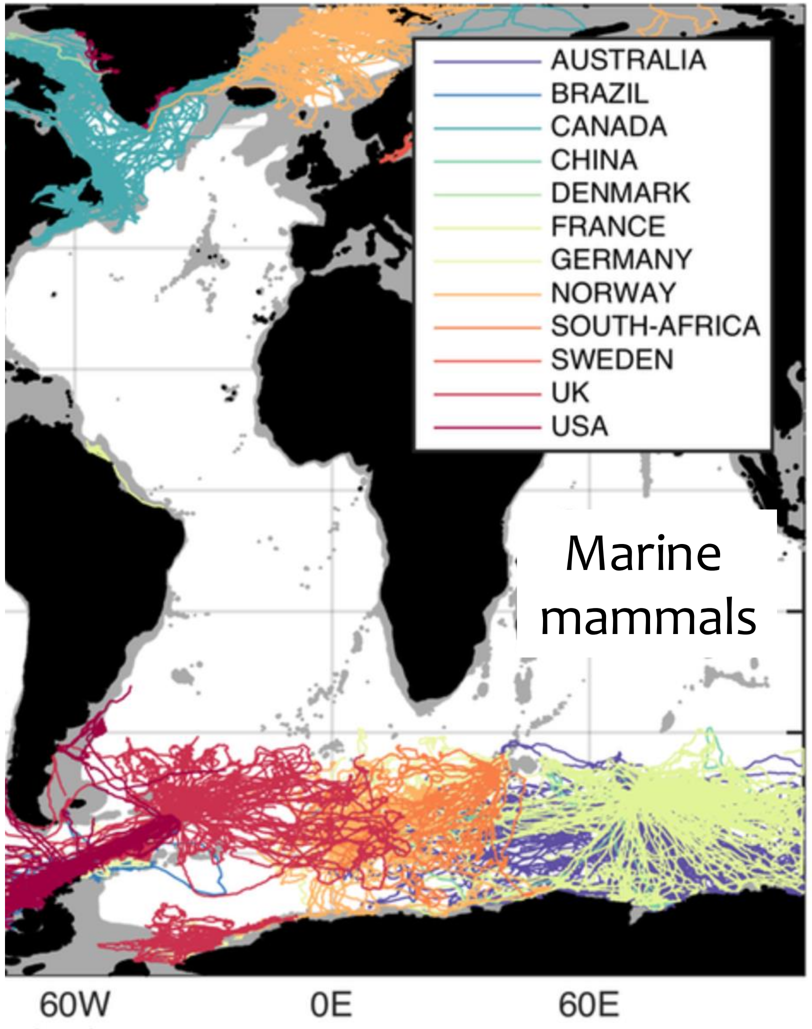
An eclectic observing system

Sparse & disparate sensors

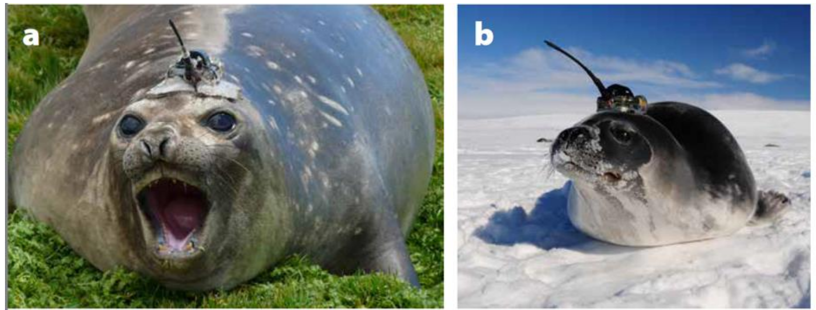








http://www.meop.net



Data Assimilation and Inverse Modeling

The DA / inverse problem is
learning from …

- … available, usually sparse, heterogeneous observations
- … *AND* known physics/dynamics

ECCO: Estimating the Circulation and Climate of the Ocean

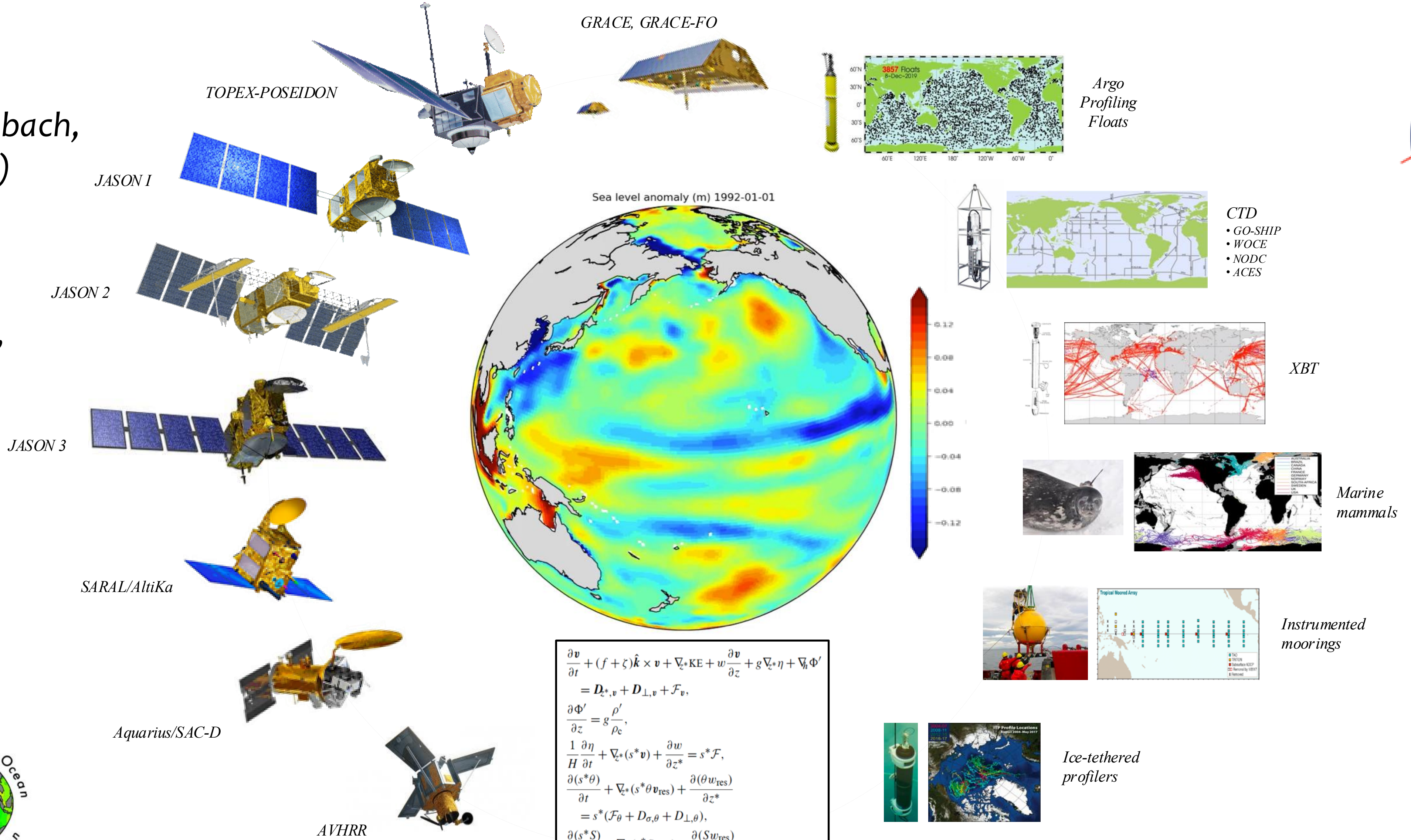
Learning from sparse, heterogeneous observations and models

Stammer et al., JGR (2002)

Wunsch & Heimbach, Physica D (2007)

Forget et al., GMD (2015)

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



$$\frac{\partial \boldsymbol{v}}{\partial t} + (f+\zeta)\hat{\boldsymbol{k}} \times \boldsymbol{v} + \nabla_{z^*}\mathrm{KE} + w\frac{\partial \boldsymbol{v}}{\partial z} + g\nabla_{z^*}\eta + \nabla_h \Phi'$$
$$= \boldsymbol{D}_{z^*,\boldsymbol{v}} + \boldsymbol{D}_{\perp,\boldsymbol{v}} + \mathcal{F}_{\boldsymbol{v}},$$
$$\frac{\partial \Phi'}{\partial z} = g\frac{\rho'}{\rho_c},$$
$$\frac{1}{H}\frac{\partial \eta}{\partial t} + \nabla_{z^*}(s^*\boldsymbol{v}) + \frac{\partial w}{\partial z^*} = s^*\mathcal{F},$$
$$\frac{\partial (s^*\theta)}{\partial t} + \nabla_{z^*}(s^*\theta \boldsymbol{v}_{\mathrm{res}}) + \frac{\partial(\theta w_{\mathrm{res}})}{\partial z^*}$$
$$= s^*(\mathcal{F}_\theta + D_{\sigma,\theta} + D_{\perp,\theta}),$$
$$\frac{\partial (s^*S)}{\partial t} + \nabla_{z^*}(s^*S \boldsymbol{v}_{\mathrm{res}}) + \frac{\partial(S w_{\mathrm{res}})}{\partial z^*}$$
$$= s^*(\mathcal{F}_S + D_{\sigma,S} + D_{\perp,S}),$$

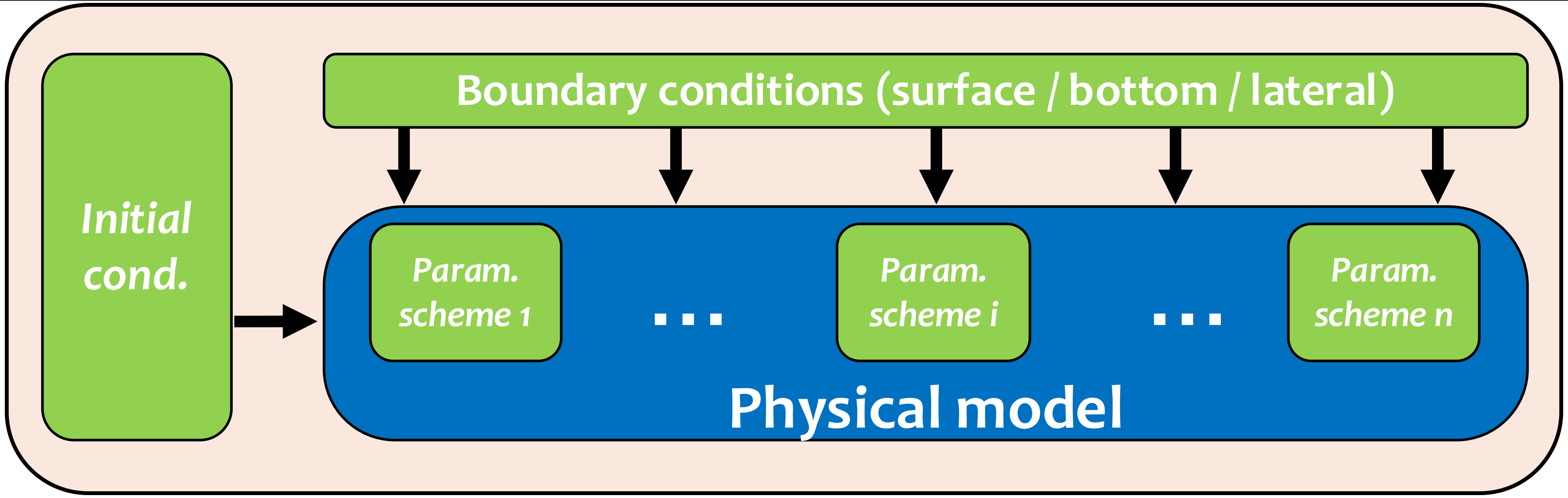
MIT general circulation model

Courtesy: I Fenty (JPL/Caltech)

https://ecco-group.org

What do we mean by
"Learning"?

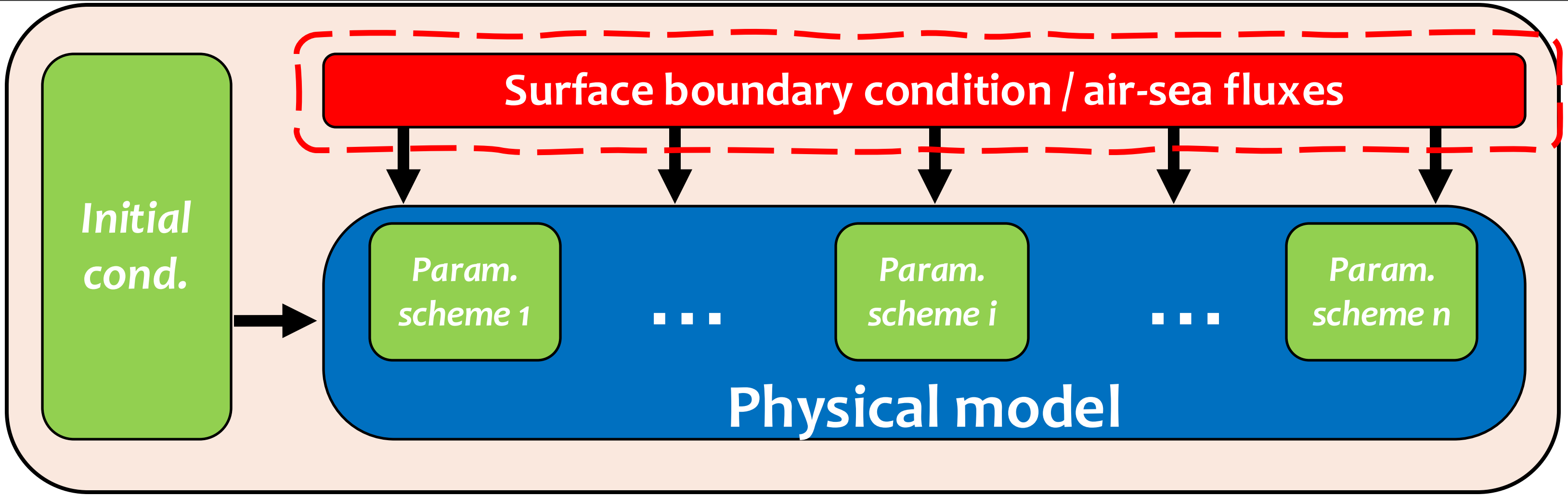
Learn …



Boundary conditions (surface / bottom / lateral)
Initial cond.
Param. scheme 1
…
Param. scheme i
…
Param. scheme n
Physical model

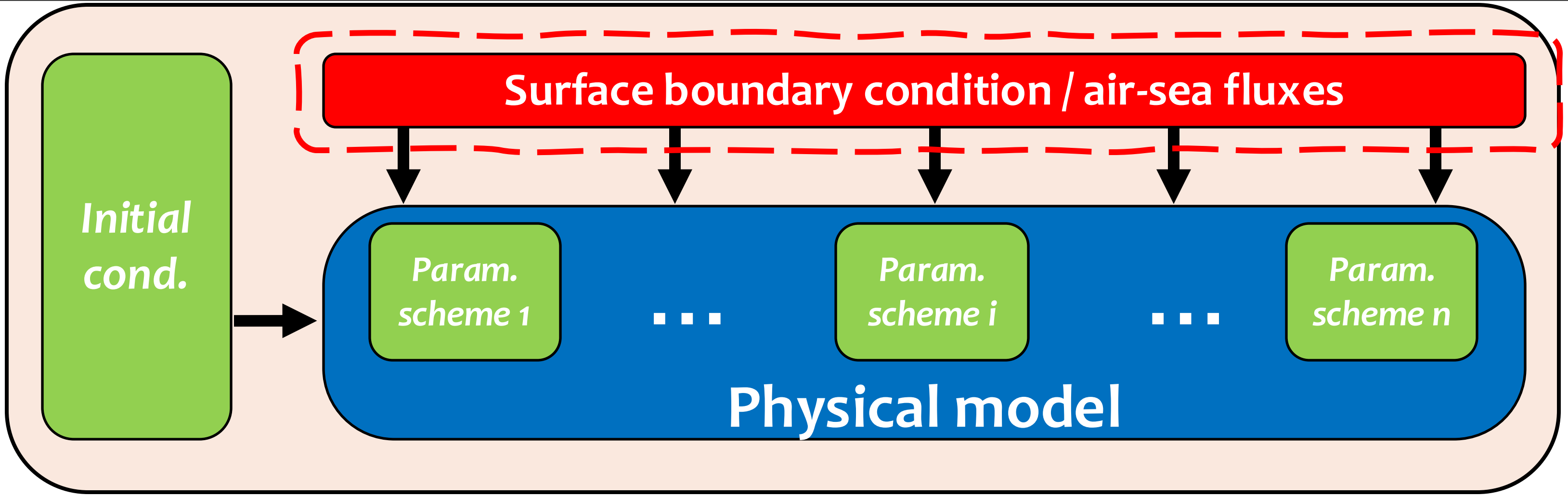
Learn model *boundary conditions*

Observations of **ocean interior**, combined with global & local mass/tracer conservation enables …

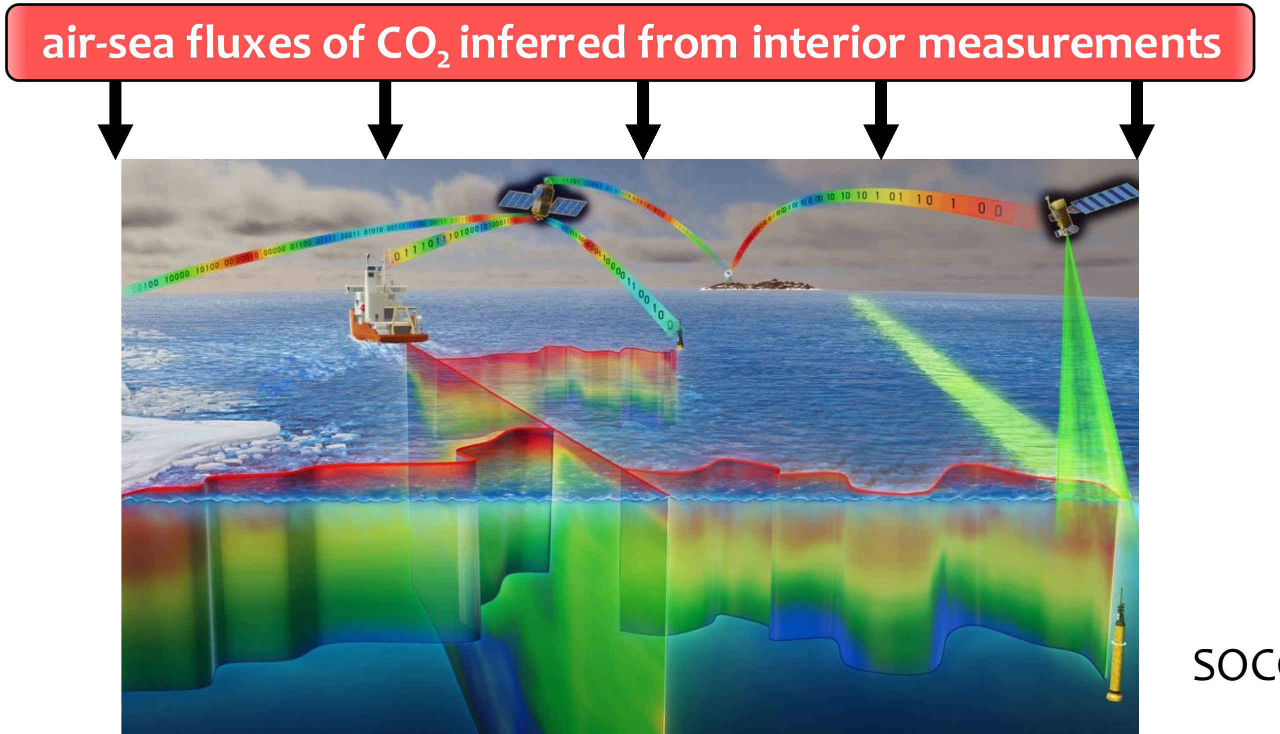


Learn model *boundary conditions*

Observations of **ocean** interior, combined with global & local mass/tracer conservation enables …



… inversion for surface fluxes that are required to match interior observations



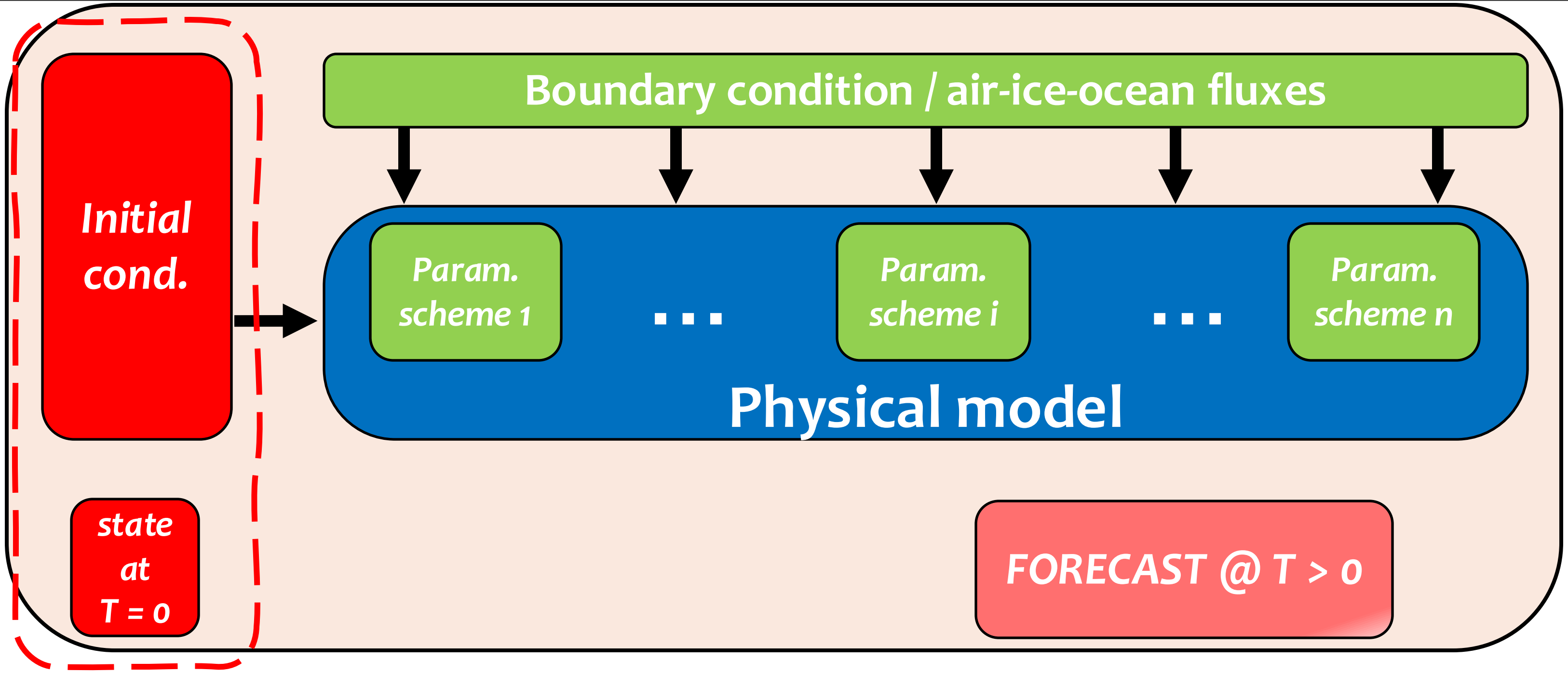
SOCCOM

Example for CO_2 air-sea fluxes (similar for heat fluxes)

Learn model *initial conditions*

Find best initial conditions that will produce optimal forecast …

*The **filtering** problem of optimal estimation & control*



Learn model *initial conditions*

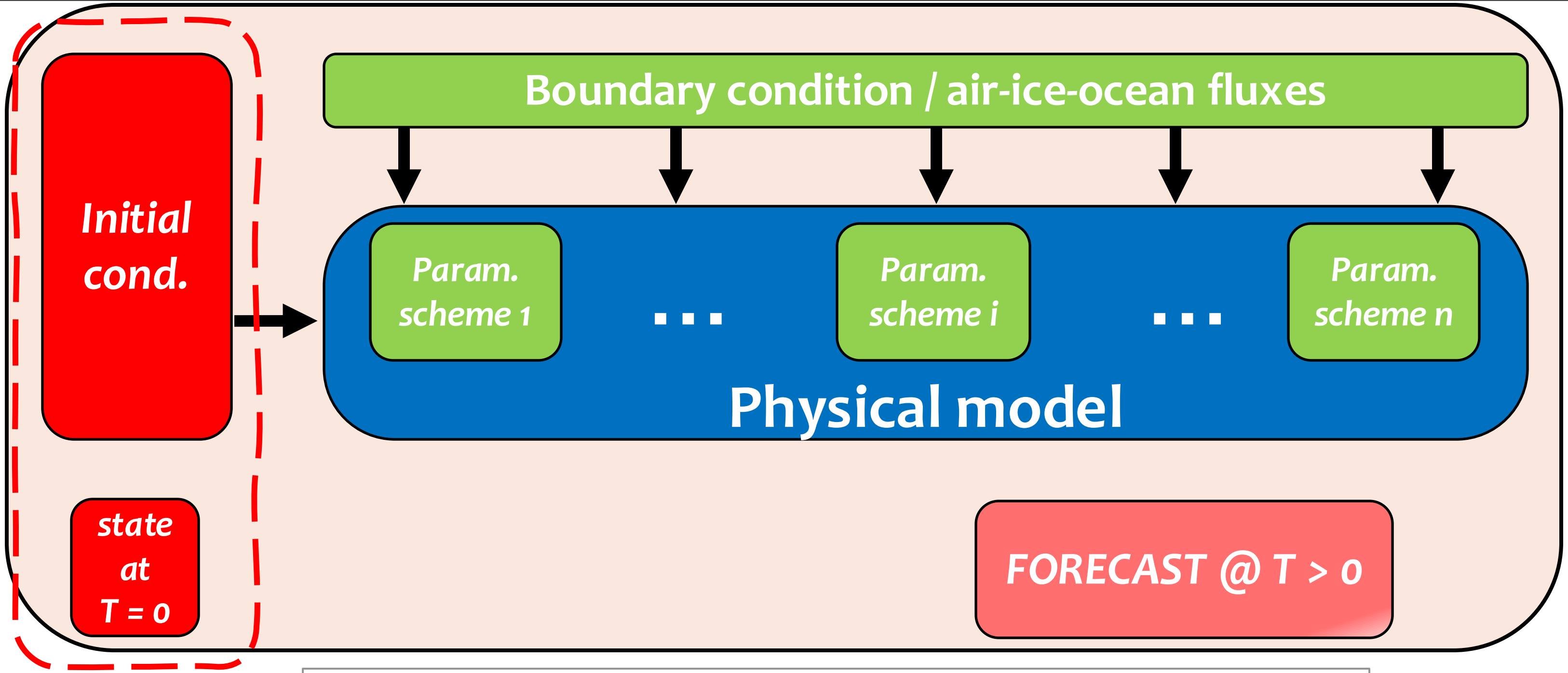
Find best initial conditions that will produce optimal forecast …

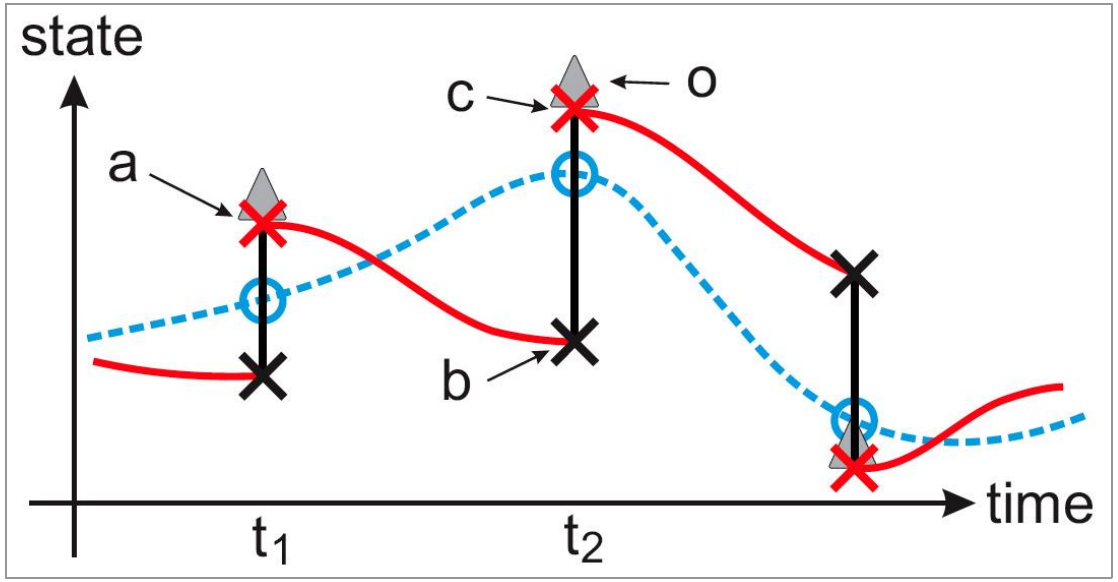
*The **filtering** problem of optimal estimation & control*

Initialization for prediction/extrapolation

as practiced in

numerical weather prediction





Learn model *initial conditions*

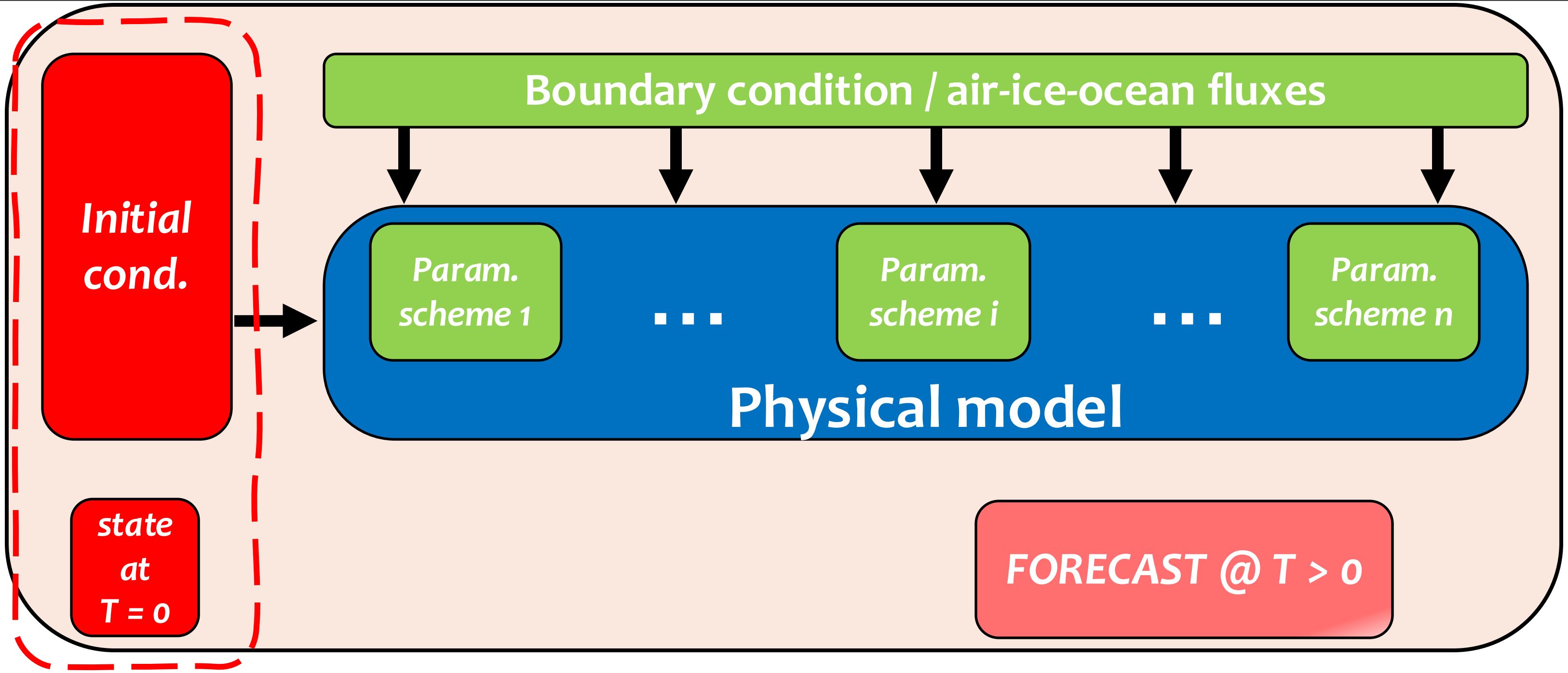
Find best initial conditions that will produce optimal forecast …

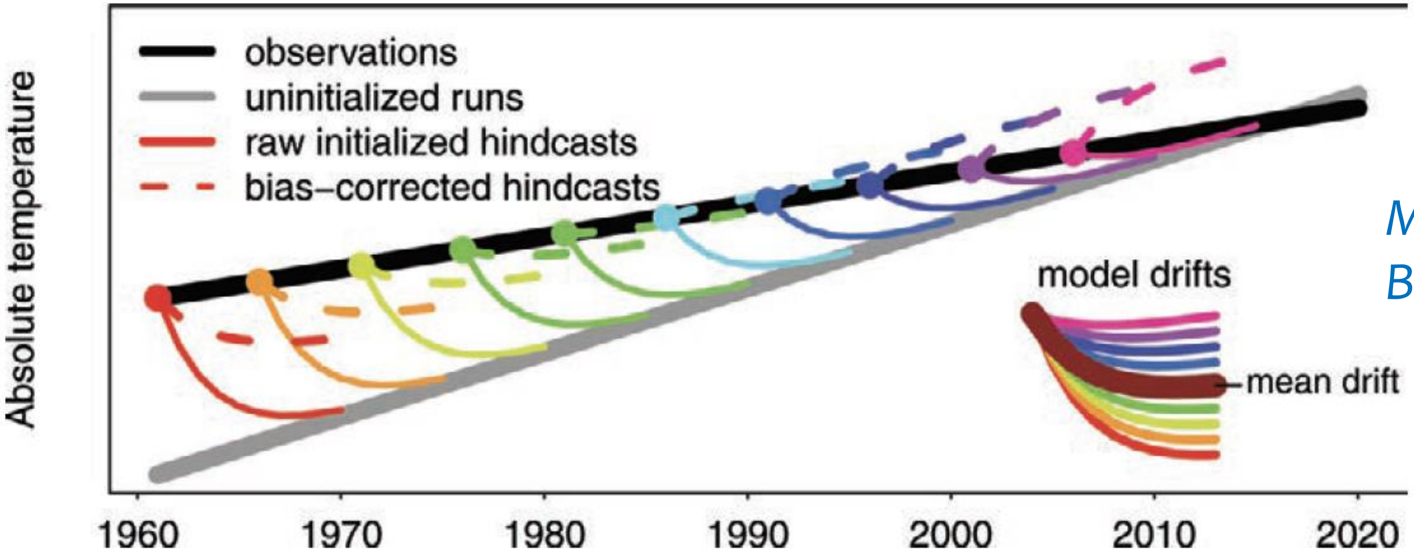
*The **filtering** problem of optimal estimation & control*

Initialization for prediction/extrapolation

as practiced in

decadal prediction





Meehl et al. BAMS (2014)

Learn model *initial conditions*

Find best initial conditions that will produce optimal forecast …

The

pr

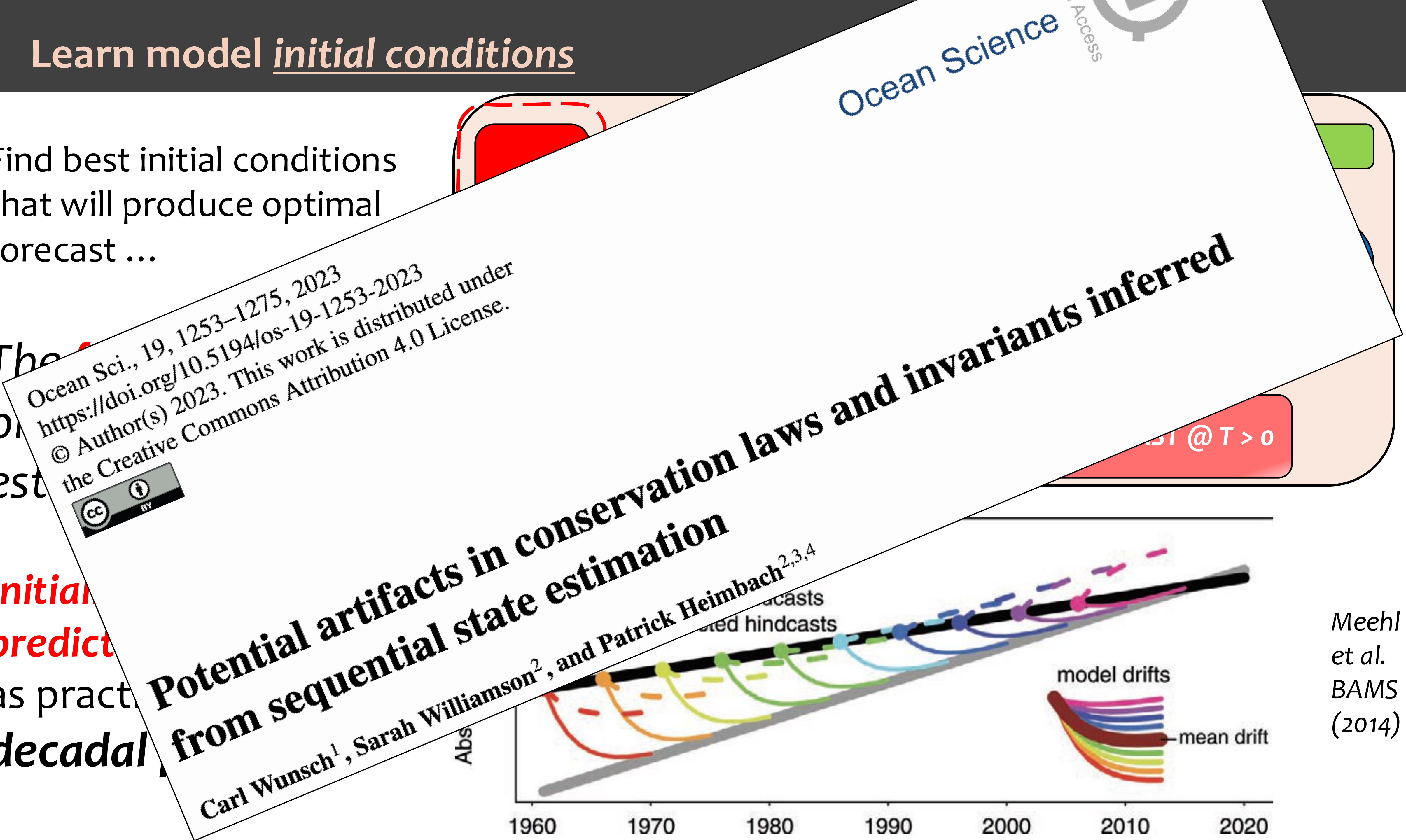
est

Initial

predict

as pract

decadal


Ocean Science
Ocean Sci., 19, 1253–1275, 2023
https://doi.org/10.5194/os-19-1253-2023

Potential artifacts in conservation laws and invariants inferred from sequential state estimation

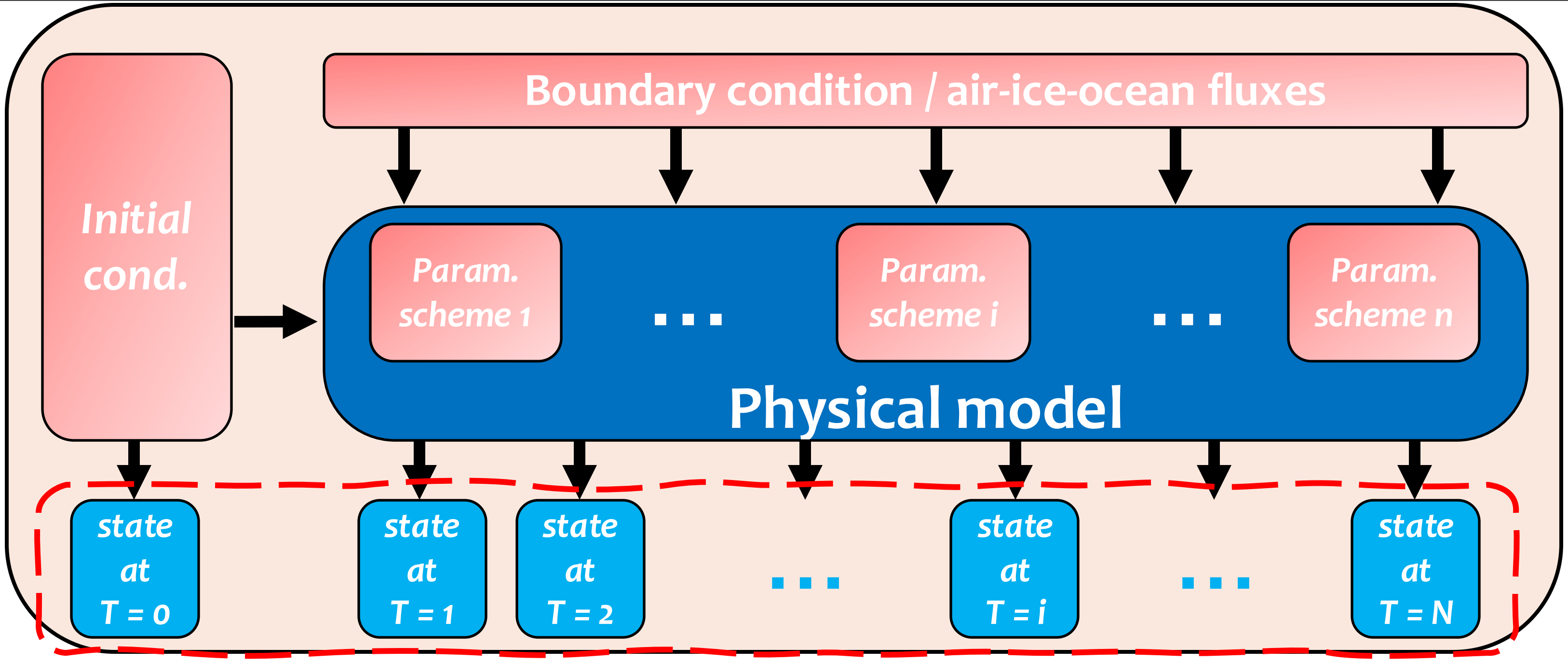
Carl Wunsch[1], Sarah Williamson[2], and Patrick Heimbach[2,3,4]

Meehl et al. BAMS (2014)

Learn model *time-evolving state*

Find model inputs that produce the best dynamically consistent state

*The **smoothing** problem of optimal estimation & control*



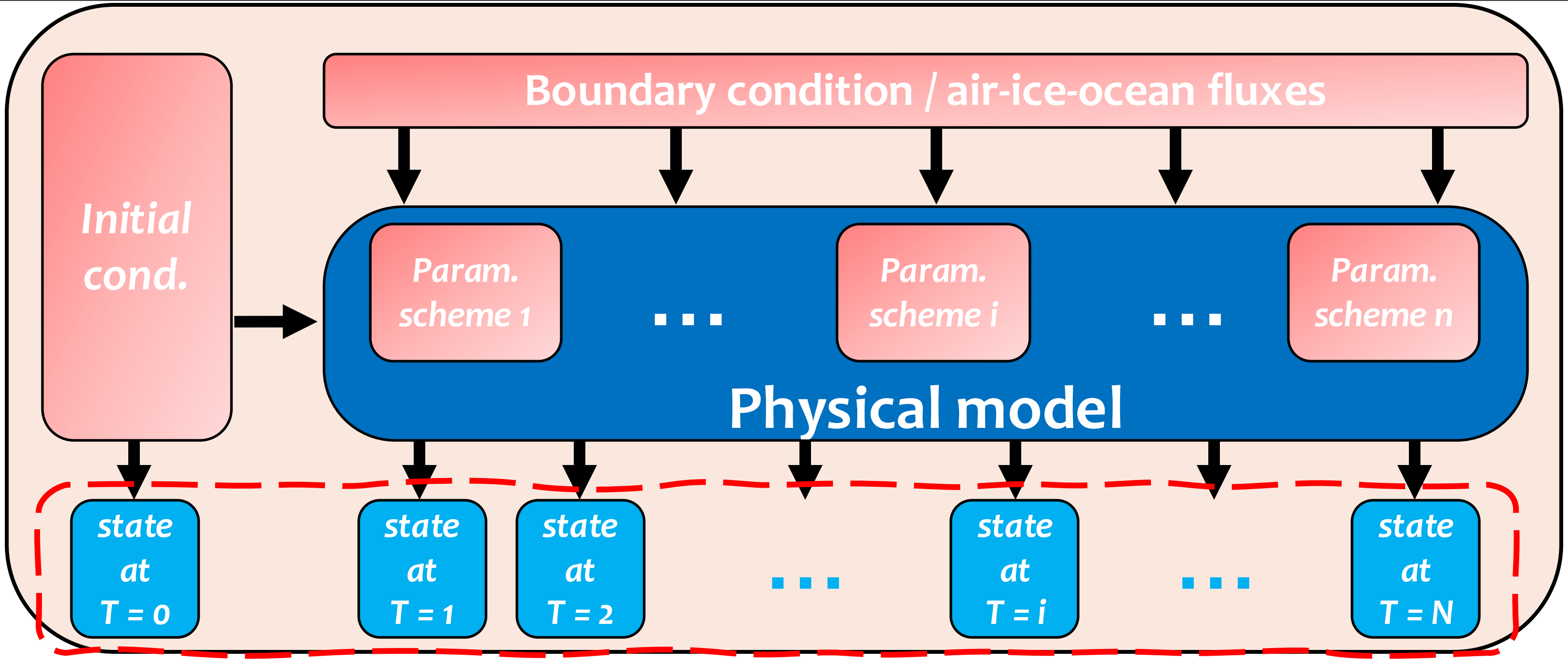
Learn model *time-evolving state*

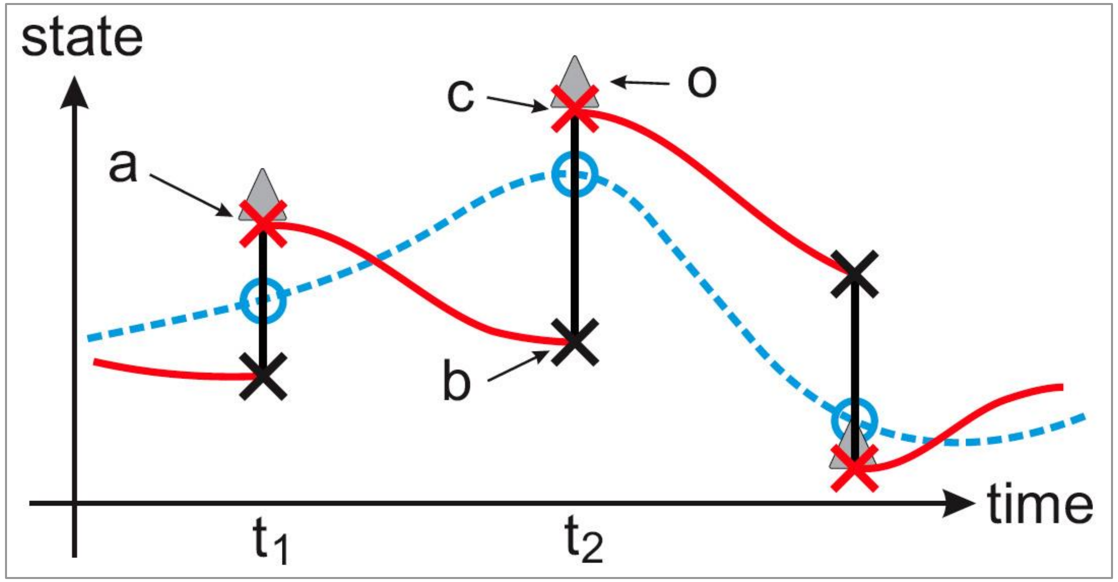
Find model inputs that produce the best dynamically consistent state

*The **smoothing** problem of optimal estimation & control*

State & parameter estimation for:

- ***interpolation/reconstruction***
- ***(transient) calibration***





Bryson & Frazier (1963)

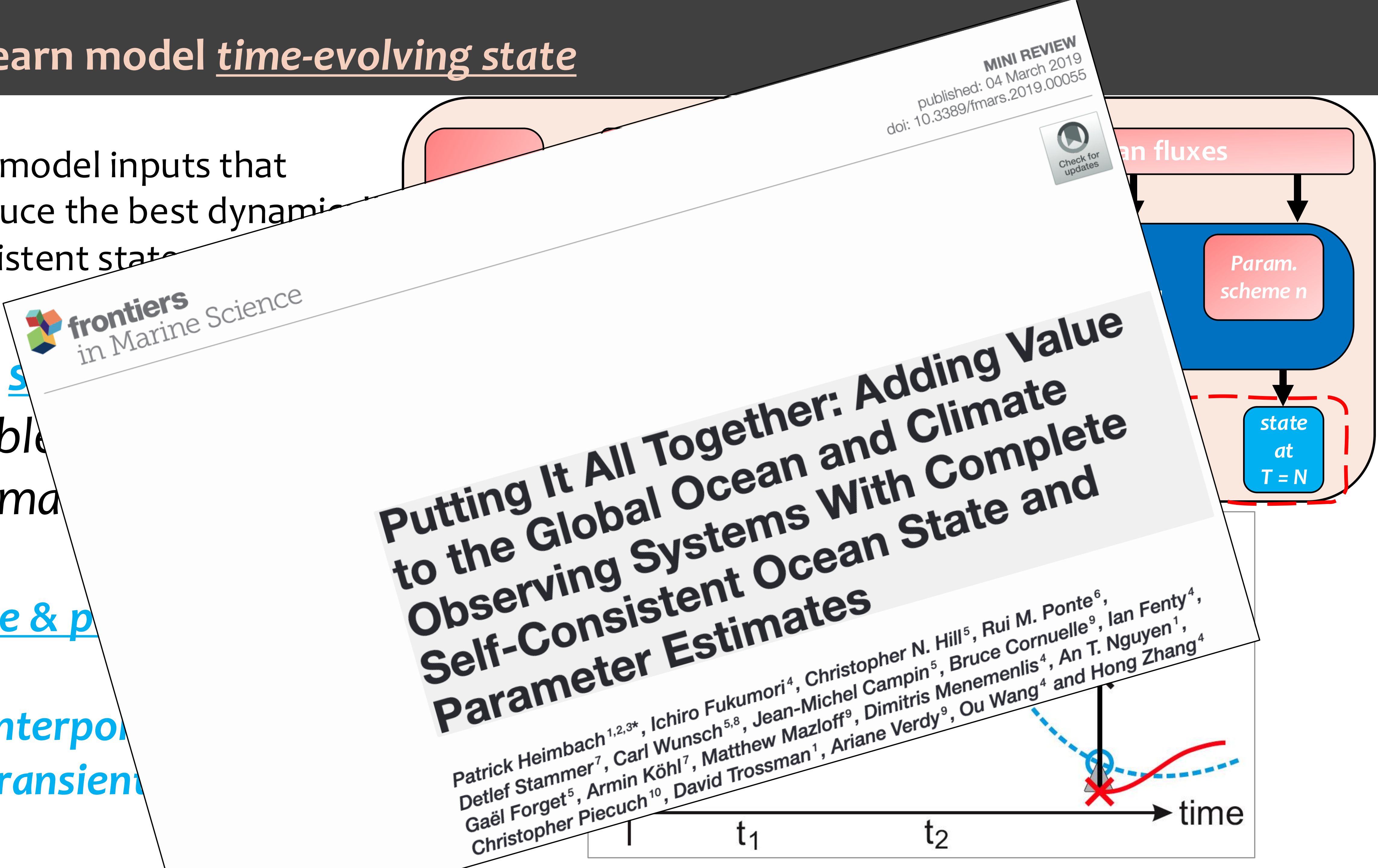
Learn model *time-evolving state*

Find model inputs that
produce the best dynami
consistent stat

The s
probl
estima

State & p
for:

- ***Interpo***
- ***transient***



MINI REVIEW
published: 04 March 2019
doi: 10.3389/fmars.2019.00055

Check for updates

frontiers in Marine Science

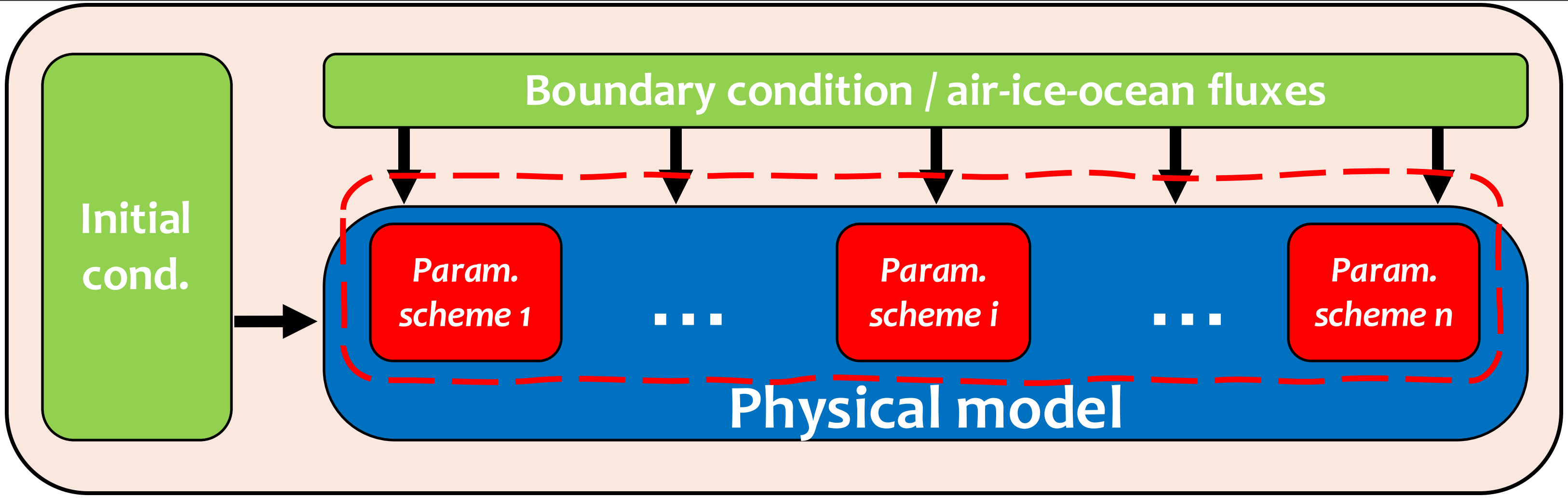
Putting It All Together: Adding Value to the Global Ocean and Climate Observing Systems With Complete Self-Consistent Ocean State and Parameter Estimates

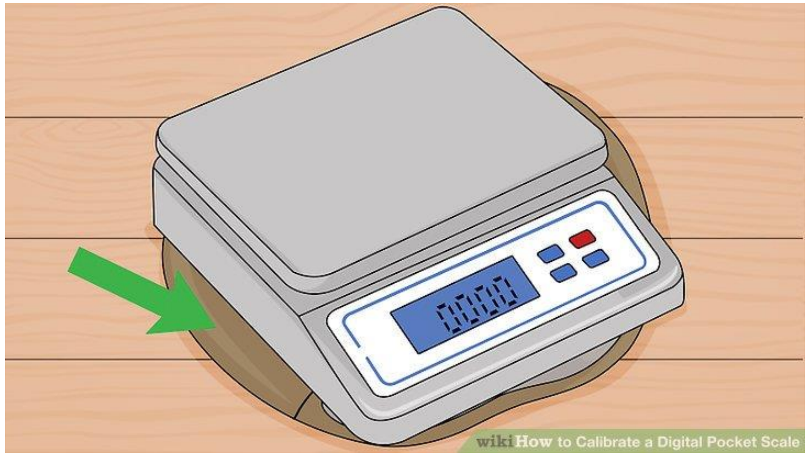
Patrick Heimbach[1,2,3*], Ichiro Fukumori[4], Christopher N. Hill[5], Rui M. Ponte[6], Detlef Stammer[7], Carl Wunsch[5,8], Jean-Michel Campin[5], Bruce Cornuelle[9], Ian Fenty[4], Gaël Forget[5], Armin Köhl[7], Matthew Mazloff[9], Dimitris Menemenlis[4], An T. Nguyen[1], Christopher Piecuch[10], David Trossman[1], Ariane Verdy[9], Ou Wang[4] and Hong Zhang[4]

Learn model *parameters*

Physical model has many empirical parameters:

- constitutive laws
- subgrid-scale parameterization schemes

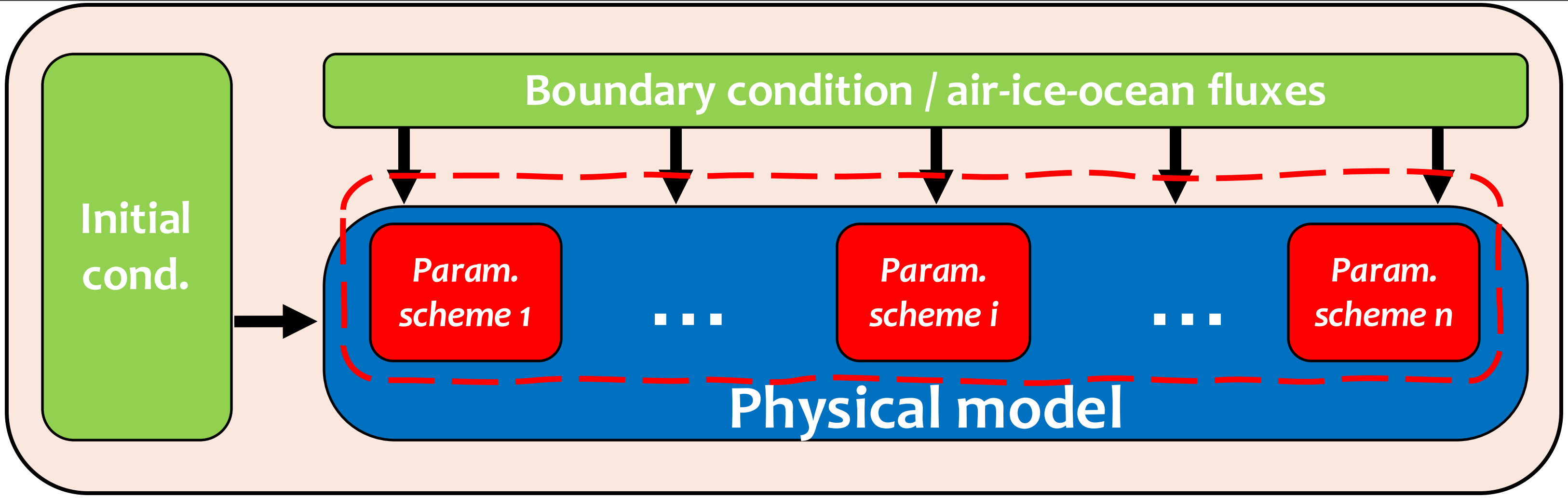




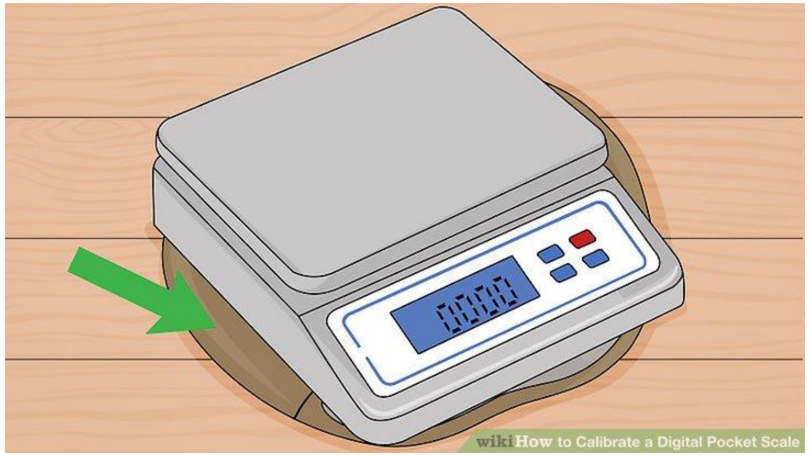
Learn model *parameters*

Physical model has many empirical parameters:

- constitutive laws
- subgrid-scale parameterization schemes



parameter estimation using observations is important



THE ART AND SCIENCE OF CLIMATE MODEL TUNING

BAMS

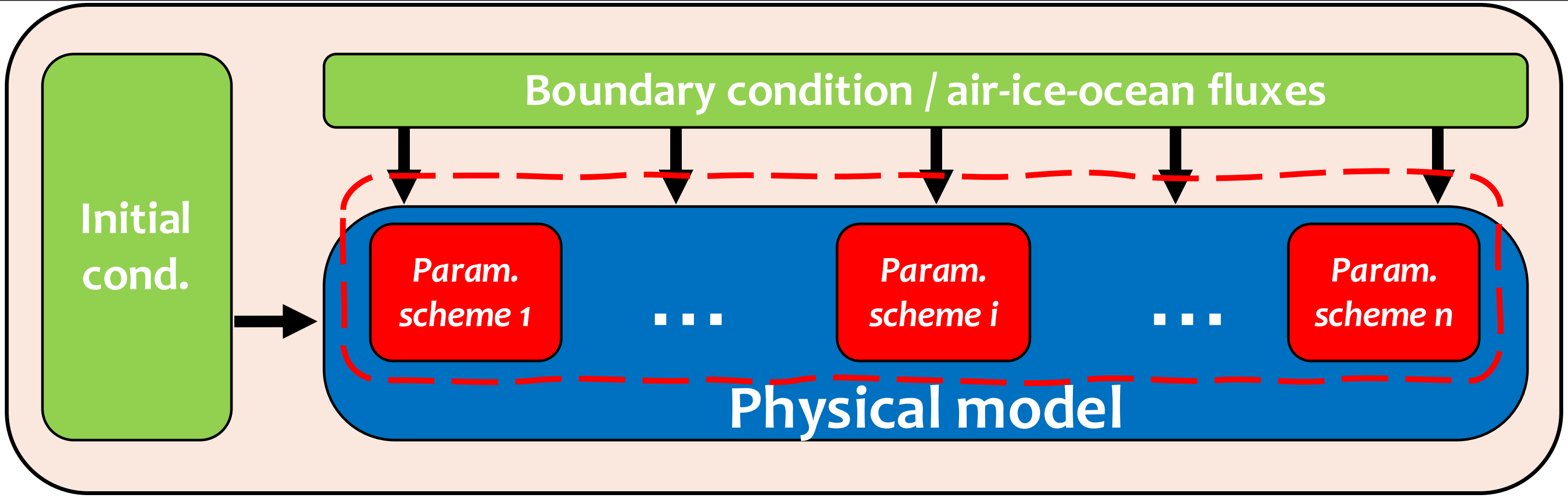
Frédéric Hourdin, Thorsten Mauritsen, Andrew Gettelman, Jean-Christophe Golaz, Venkatramani Balaji, Qingyun Duan, Doris Folini, Duoying Ji, Daniel Klocke, Yun Qian, Florian Rauser, Catherine Rio, Lorenzo Tomassini, Masahiro Watanabe, and Daniel Williamson

We survey the rationale and diversity of approaches for tuning, a fundamental aspect of climate modeling, which should be more systematically documented and taken into account in multimodel analysis.

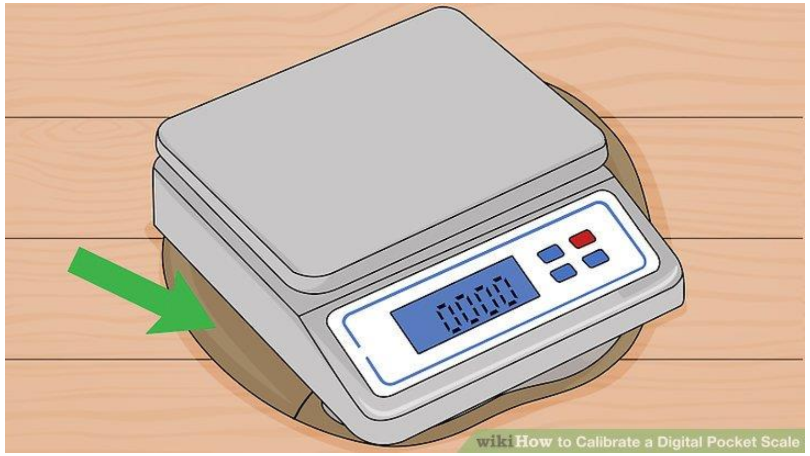
Learn model *parameters*

Physical model has many empirical parameters:

- constitutive laws
- subgrid-scale parameterization schemes



parameter estimation to calibrate model parameters




Ocean Sci., 18, 729–759, 2022
https://doi.org/10.5194/os-18-729-2022

Ocean Science
Open Access

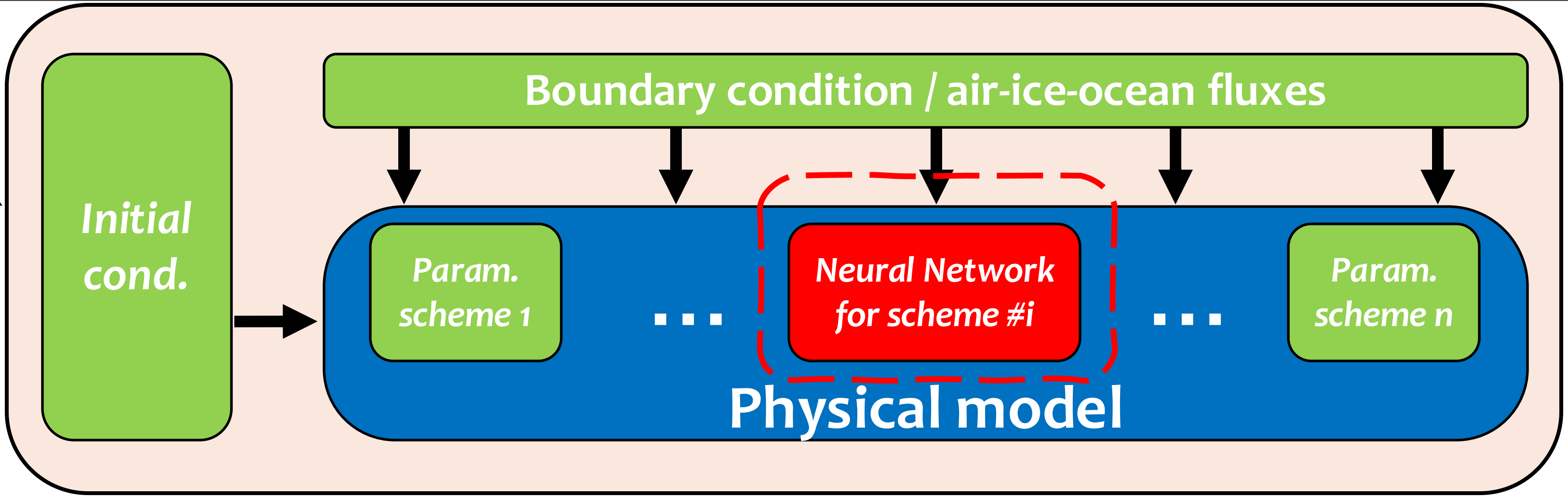
Tracer and observationally derived constraints on diapycnal diffusivities in an ocean state estimate

David S. Trossman[1,2], Caitlin B. Whalen[3], Thomas W. N. Haine[4], Amy F. Waterhouse[5], An T. Nguyen[6], Arash Bigdeli[7], Matthew Mazloff[5], and Patrick Heimbach[6,8]

Learn surrogate (e.g., NN) of *model's parameterization scheme*

Parameterization scheme(s) is replaced by neural network

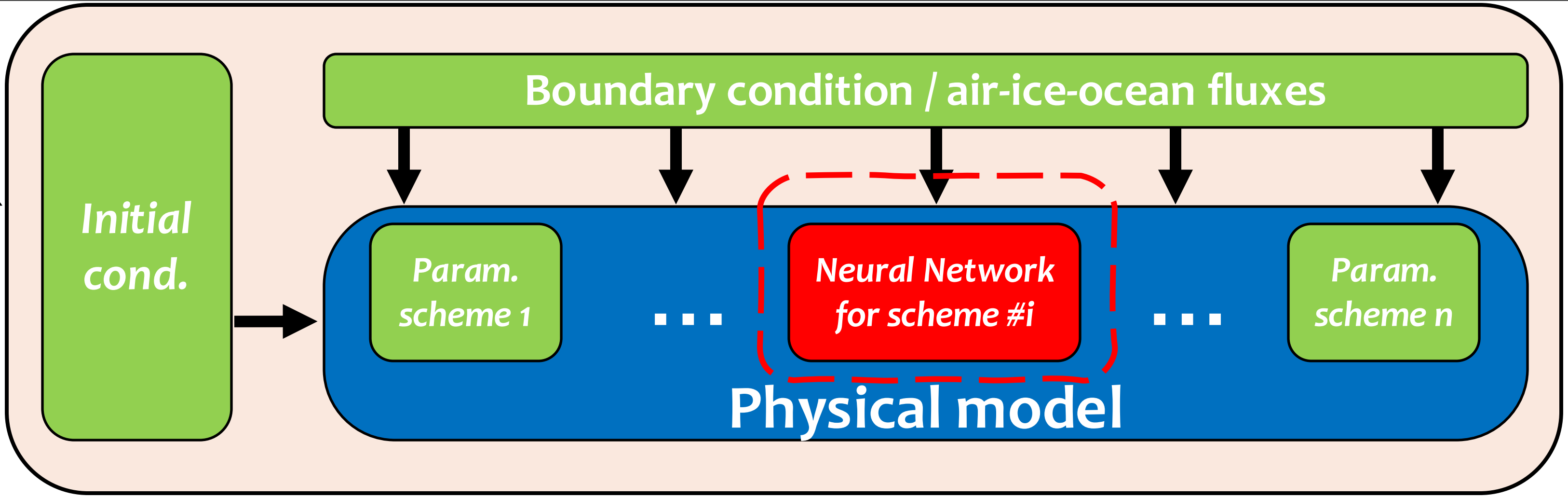
NN is trained on high-fidelity simulation data (e.g. LES) which resolve scales to be parameterized



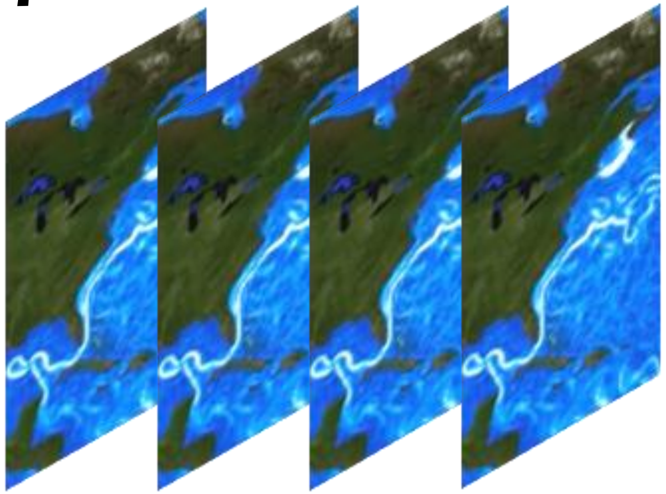
Learn surrogate (e.g., NN) of *model's parameterization scheme*

Parameterization scheme(s) is replaced by neural network

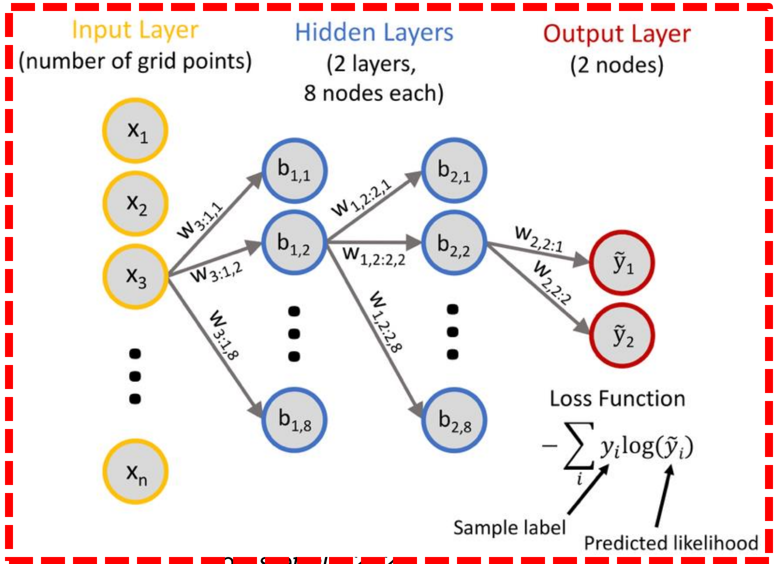
NN is trained on high-fidelity simulation data (e.g. LES) which resolve scales to be parameterized



$$\frac{\partial\overline{\Phi}}{\partial t} + (\overline{\mathbf{u}}\cdot\nabla)\,\overline{\Phi} = \nabla\cdot\left(\kappa\nabla\overline{\Phi}\right) + \overline{F}_{\Phi} + \boxed{\nabla\cdot\mathbf{S}}$$



$$\boxed{\nabla\cdot\mathbf{S}} = \nabla\cdot\left(\overline{\mathbf{u}\,\Phi} - \overline{\mathbf{u}}\,\overline{\Phi}\right) =$$



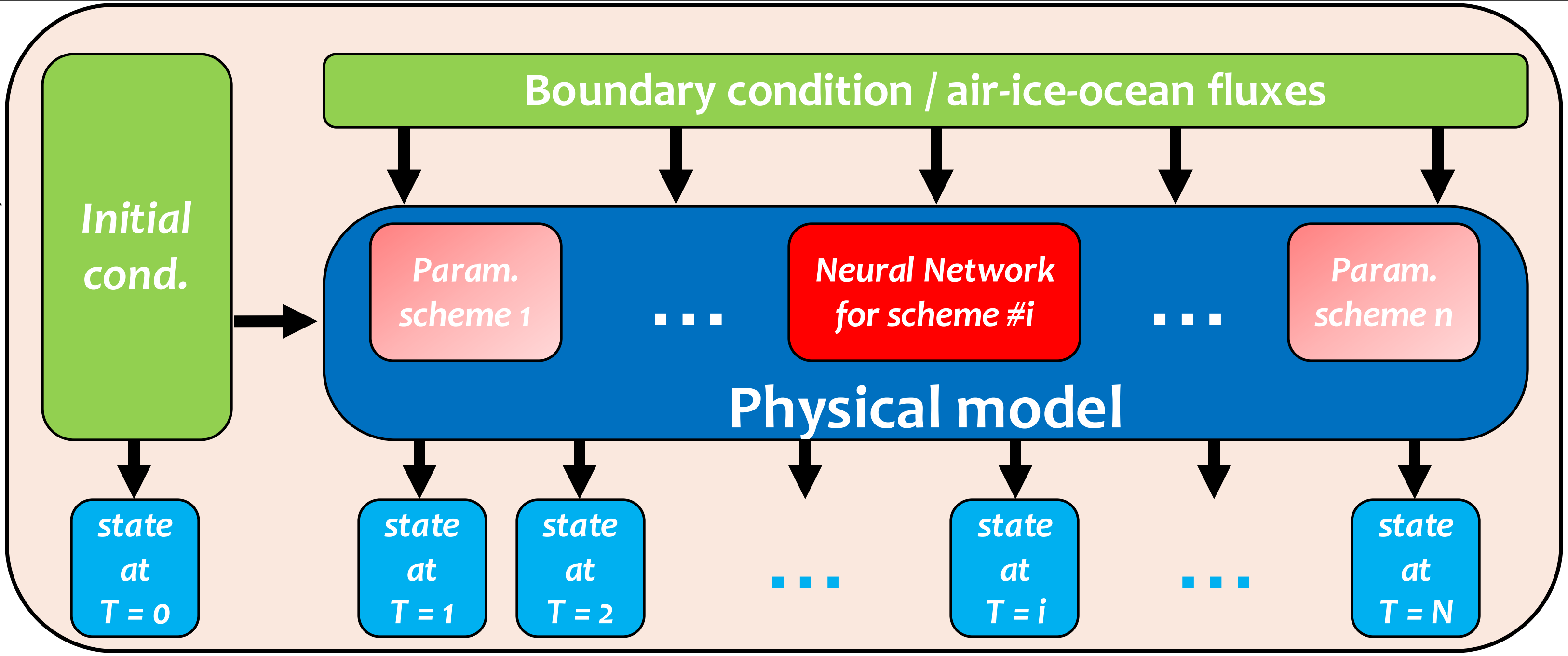
a priori / offline learning

Zanna & Bolton, GRL (2020)

Learn *hybrid* physical/surrogate (NN) model

Parameterization scheme(s) is replaced by neural network

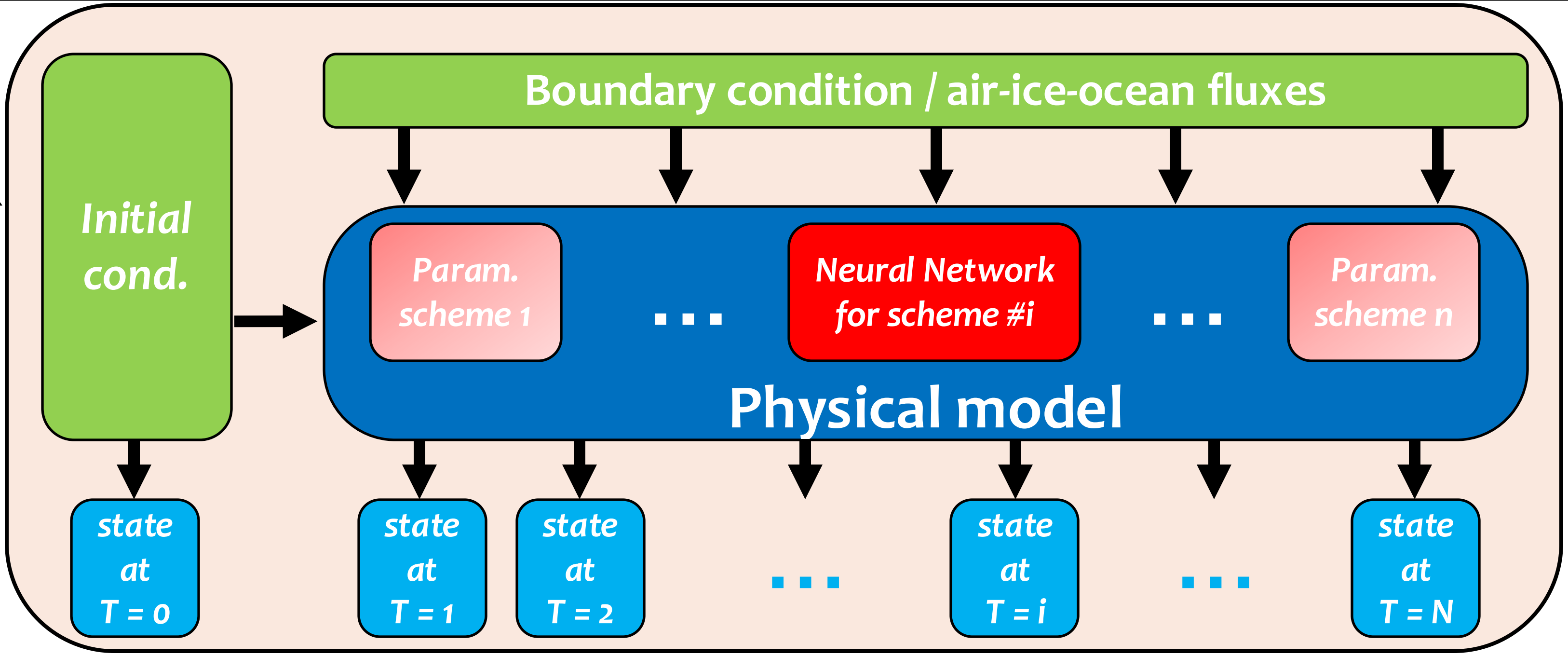
Training of the NN is part of "training" of the physical model on state variables

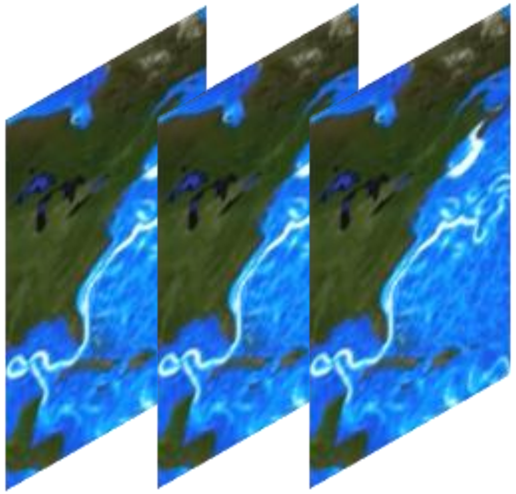


Learn *hybrid* physical/surrogate (NN) model

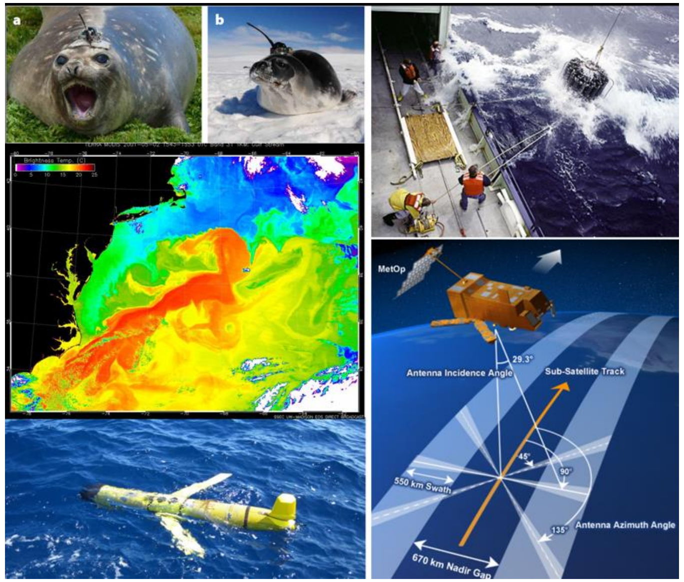
Parameterization scheme(s) is replaced by neural network

Training of the NN is part of "training" of the physical model on state variables





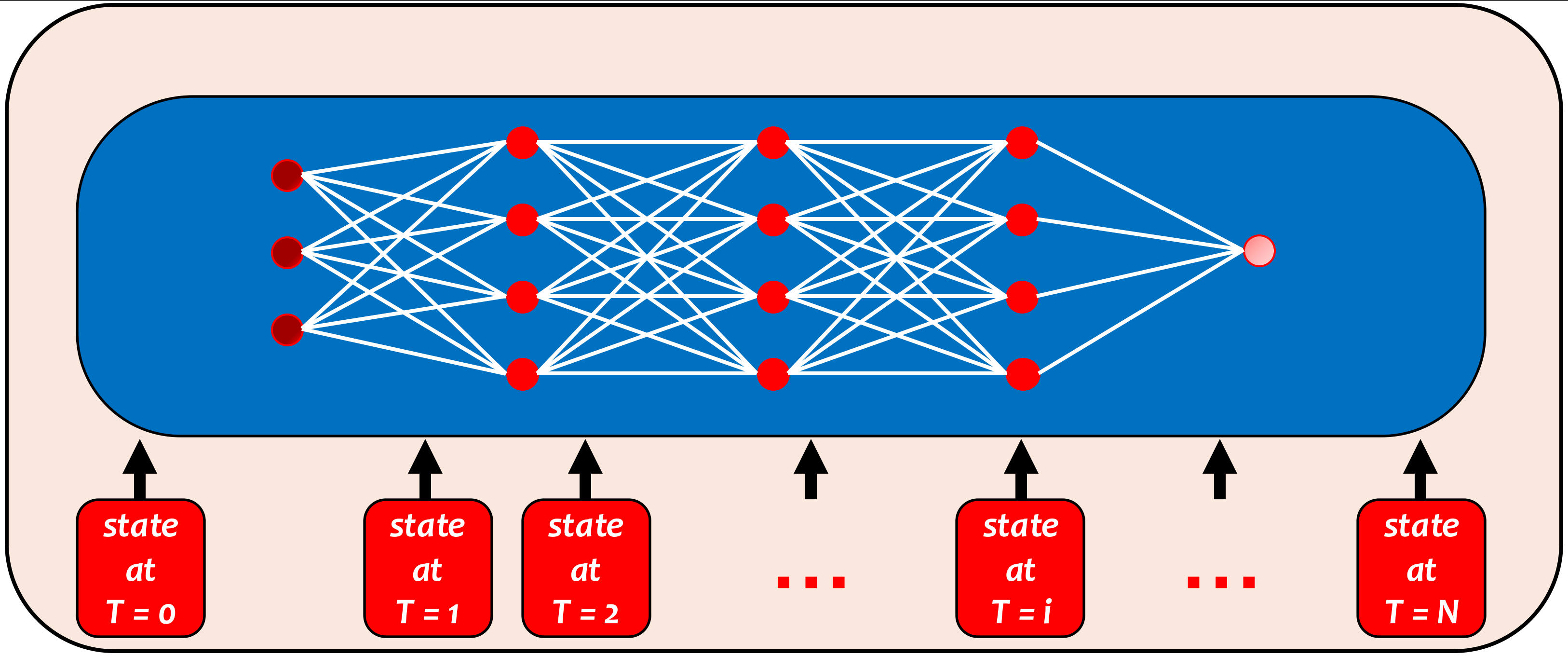
+



a posteriori / full-model / online / end-to-end learning

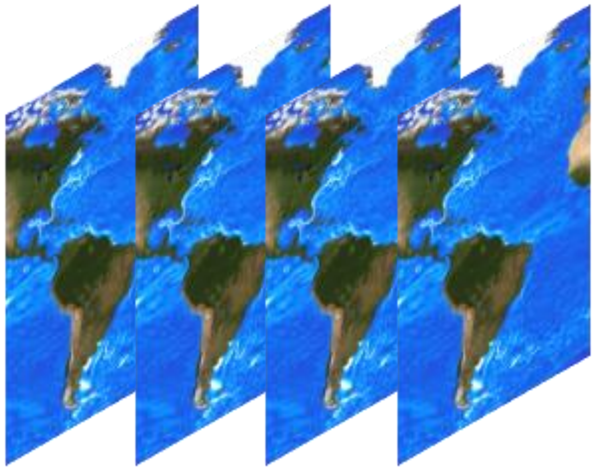
Learn surrogate (e.g., NN) of the *entire* physical model

Physical model is replaced entirely by surrogate model, e.g., neural network (NN) (or reduced-order model)



Weights of neural network trained on simulated model states from:

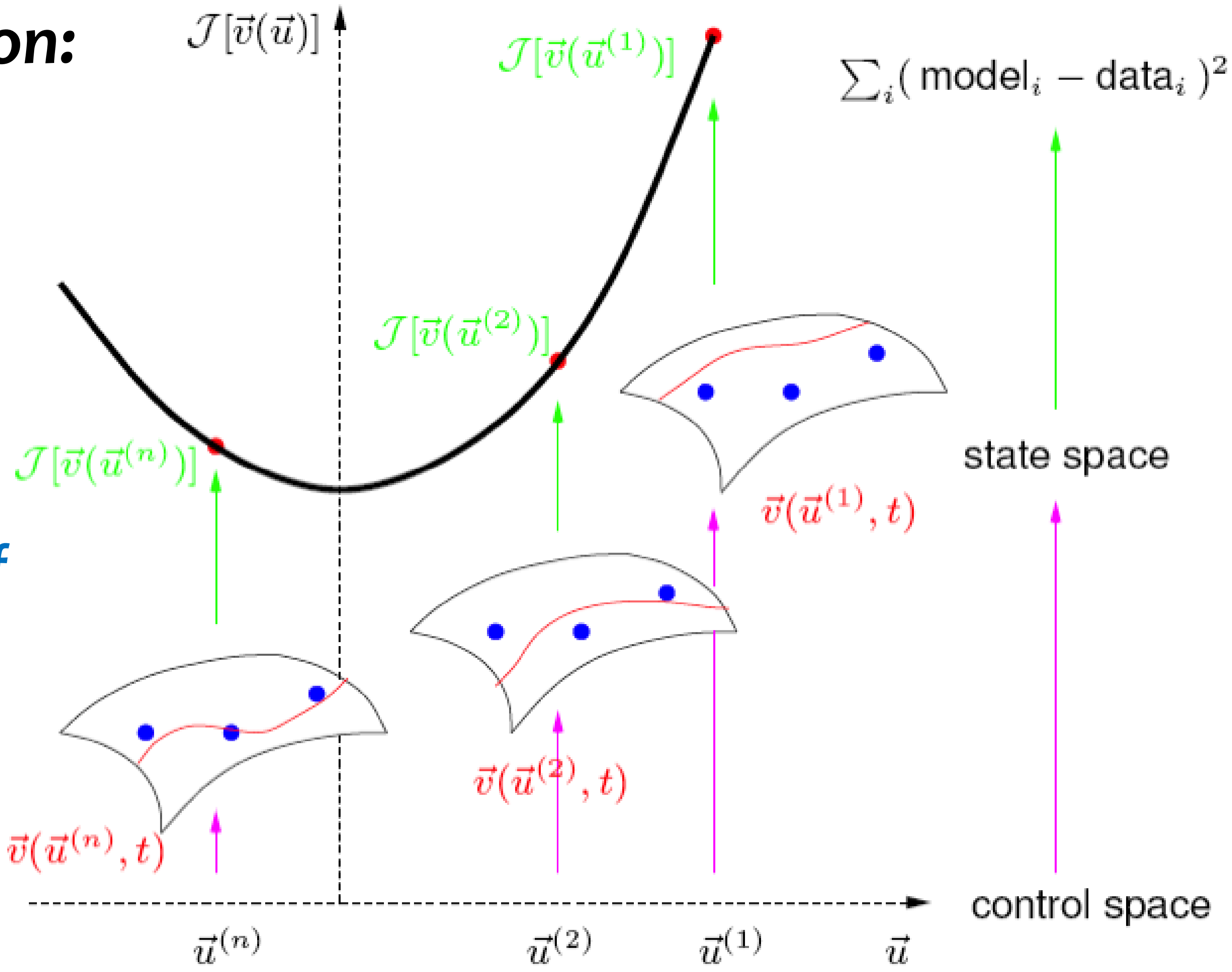
- *high-fidelity models (e.g., LES)*
- *"reanalyses" (those are NOT observations!)*



A key unifying computational framework of “*learning from data*”

Gradient-based optimization:

- <u>*inversion*</u> *(physical models)*
  - ***seek uncertain input / control variables / parameters***
- <u>*training*</u> *(neural networks)*
  - ***seek uncertain weights of NN representation***



<u>Adjoint / backpropagation</u>

powerful tool for computing high-dimensional gradients!

1/ “Data Assimilation” is more than its use in numerical weather prediction (initial conditions)

-> optimal estimation/control (or inverse) methods

2/ Conceptually strong similarities between scientific machine learning and optimal estimation/control theory

1/ "Data Assimilation" is more than its use in numerical weather prediction (initial conditions)

-> *optimal estimation/control (or inverse) methods*

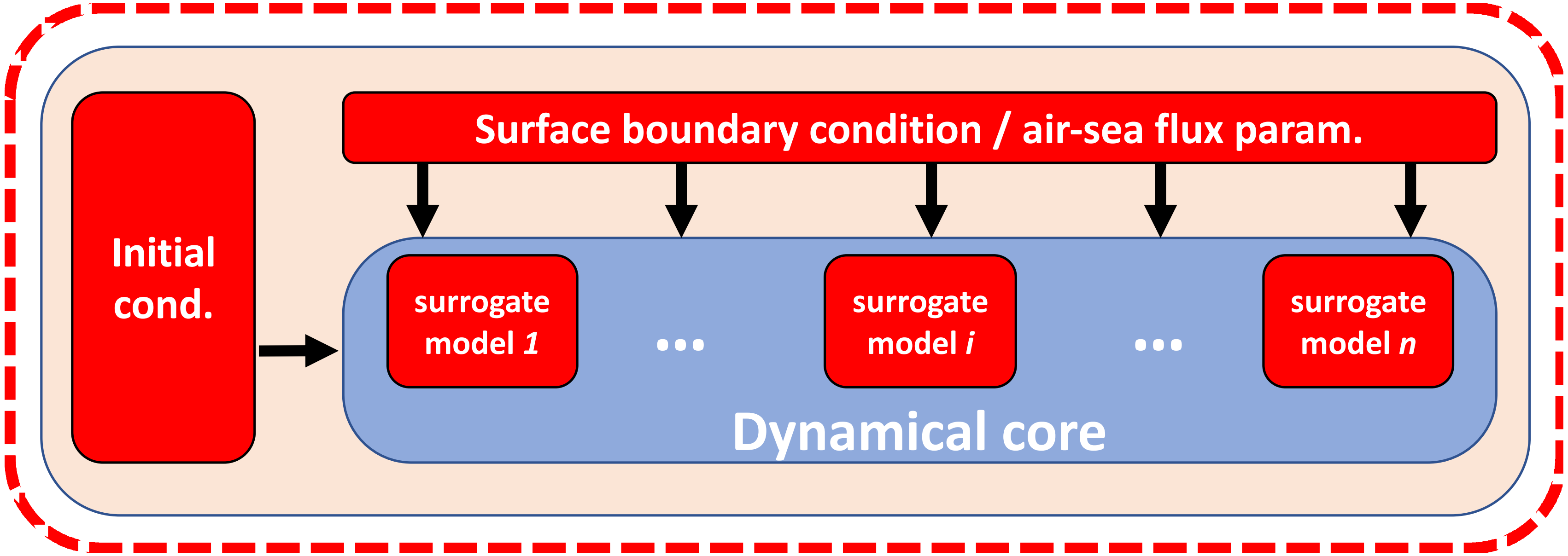
2/ Conceptually strong similarities between scientific machine learning and optimal estimation/control theory

3/ Physics-based models remain key for producing "high-fidelity" neural network training data

Full-model learning

Can we integrate the surrogate model training within full-model calibration?

Calibrating (learning) model parameterizations – ***online learning***



Use of full-model differentiable programming to

- replace parts of model by appropriate surrogates
- use available observations to train/calibrate uncertain variables
- combines inverse modeling and machine learning in ***full-model learning***
  - *e.g., NeuralGCM (Kochkov et al., Nature, 2024)*

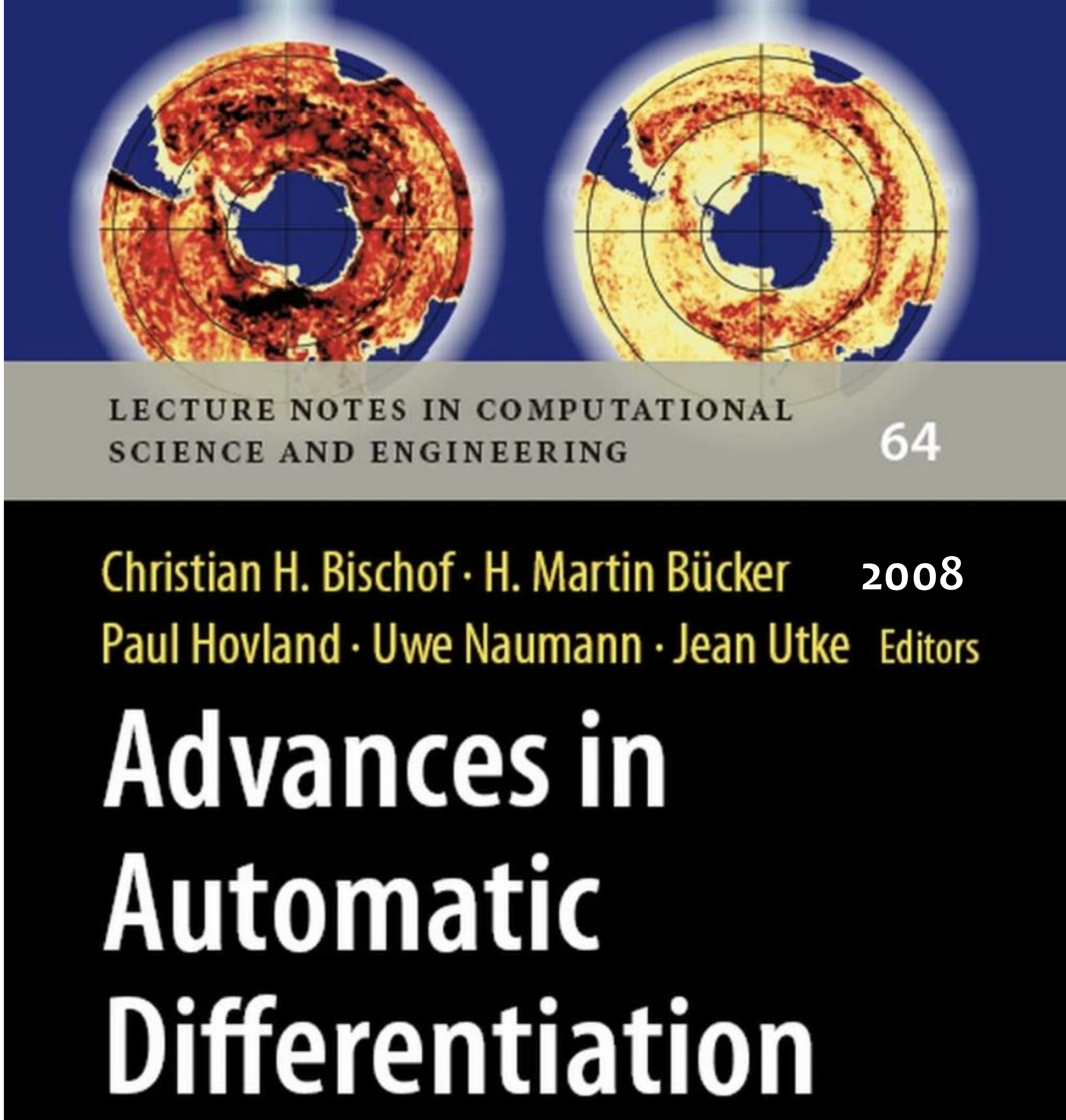
relies on general-purpose automatic differentiation (AD)

Automatic Differentiation

- ***One of the main algorithmic workhorses of ML***

- (***Aside:*** We have been using AD for 2.5 decades in the context of an ocean general circulation model written in Fortran)

Marotzke et al. (1999)
Stammer et al. (2002)
Heimbach et al. (2005)
Wunsch & Heimbach (2006)
Utke et al. (2007) – open source
Gaikwad et al. (2024) – open source



Since ~2023 the idea of differentiable Earth System Models has taken off …

Geosci. Model Dev., 16, 3123–3135, 2023
https://doi.org/10.5194/gmd-16-3123-2023

Geoscientific
Model Development
Open Access

Differentiable programming for Earth system modeling

Maximilian Gelbrecht[1,2], Alistair White[1,2], Sebastian Bathiany[1,2], and Niklas Boers[1,2,3]

[1]Earth System Modelling, School of Engineering and Design, Technical University of Munich, Munich, Germany
[2]Potsdam Institute for Climate Impact Research, Potsdam, Germany
[3]Department of Mathematics and Global Systems Institute, University of Exeter, Exeter, UK

Since ~2023 the idea of differentiable Earth System Models has taken off …

Geosci. Model Dev., 16, 3123–3135, 2023
https://doi.org/10.5194/gmd-16-3123-2023

Geoscientific
Model Development
Open Access

Differentiable programming

Maximilian Gelbrecht[1,2], Alistair White[1,2], Sebastian
[1]Earth System Modelling, School of Engineering and De
[2]Potsdam Institute for Climate Impact Research, Potsdam,
[3]Department of Mathematics and Global Systems Institute,

nature reviews earth & environment
https://doi.org/10.1038/s43017-023-00450-9

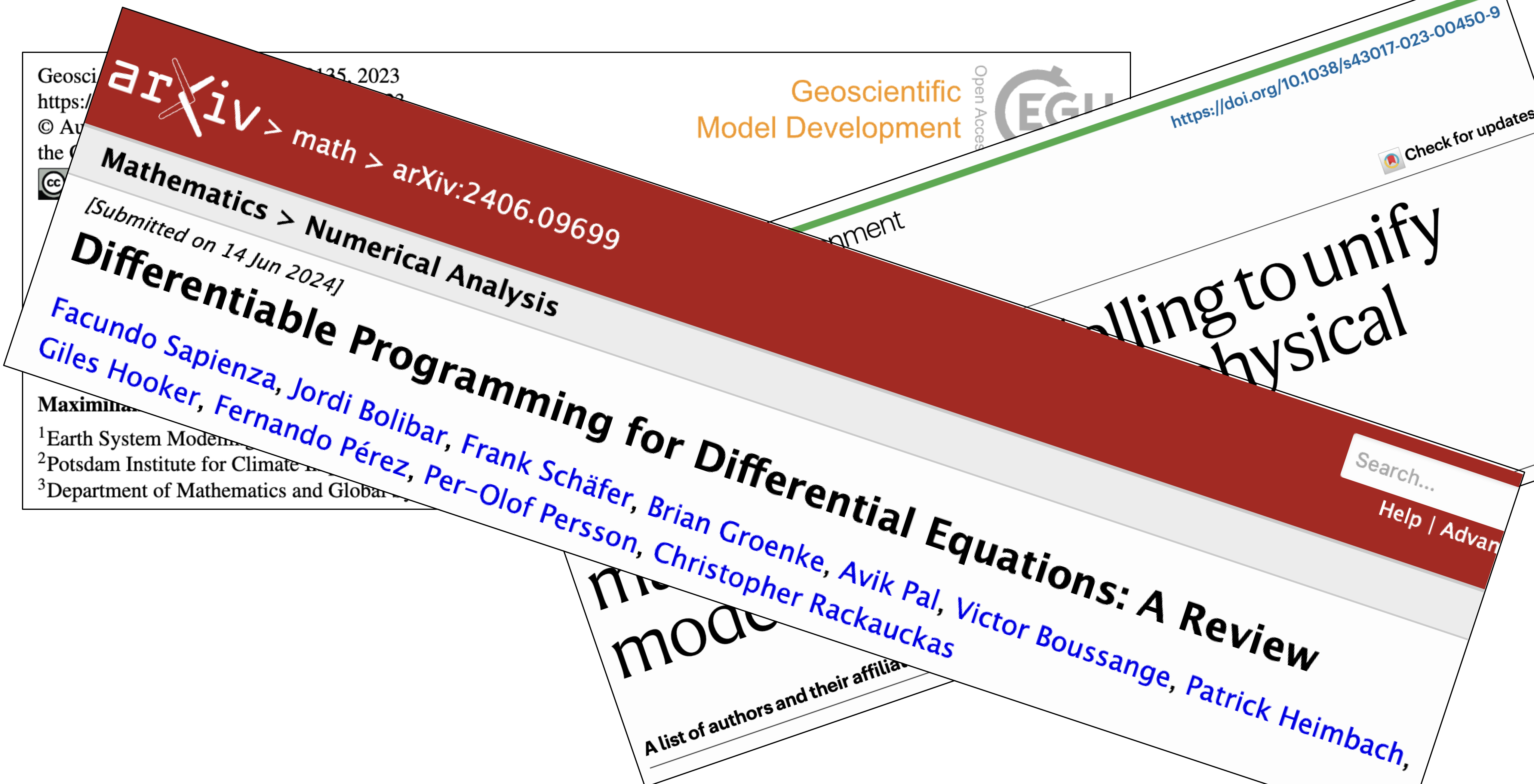
Check for updates

Perspective

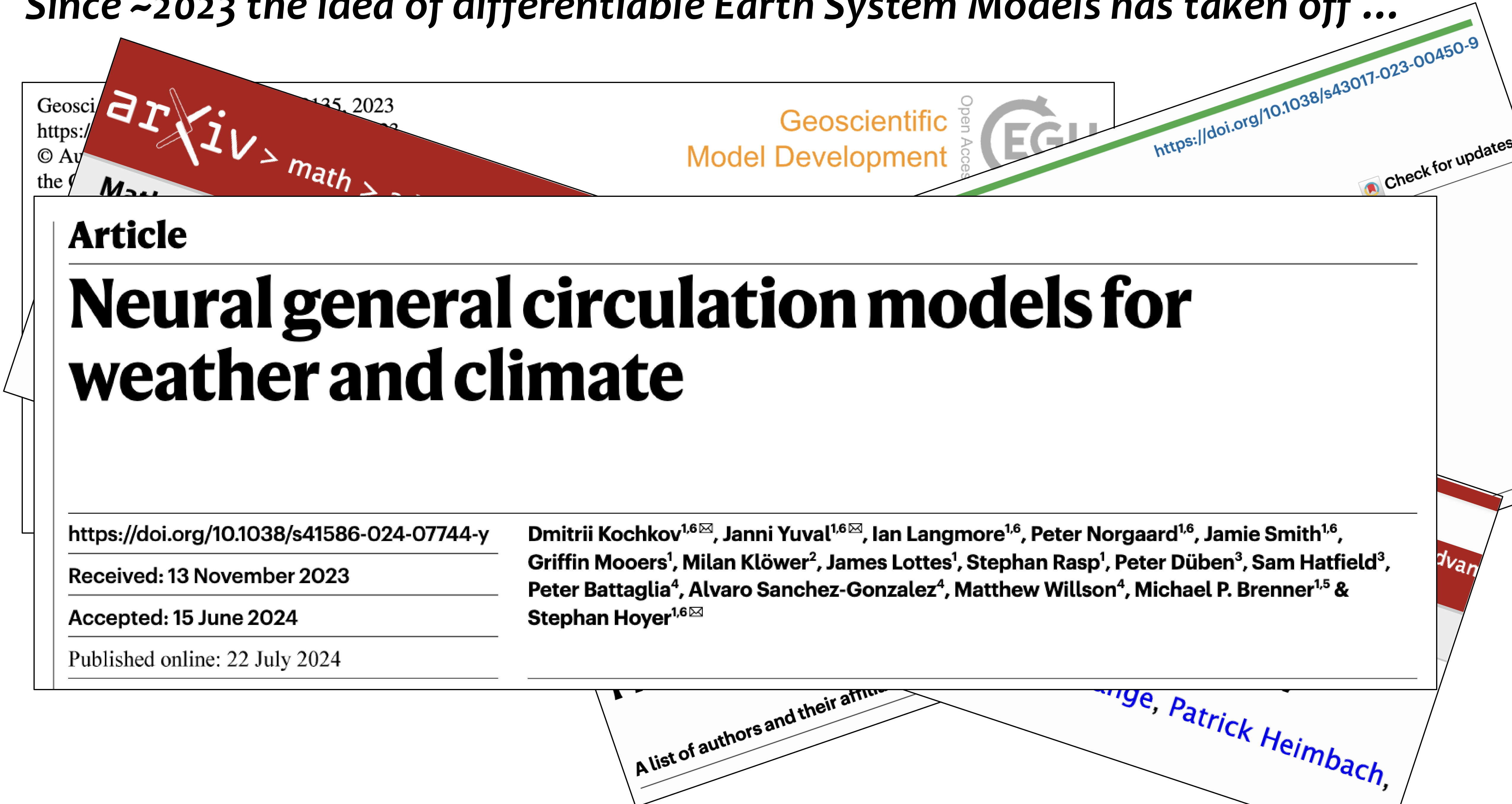
Differentiable modelling to unify machine learning and physical models for geosciences

A list of authors and their affiliations appears at the end of the paper

Since ~2023 the idea of differentiable Earth System Models has taken off …



Since ~2023 the idea of differentiable Earth System Models has taken off …



Article

Neural general circulation models for weather and climate

https://doi.org/10.1038/s41586-024-07744-y

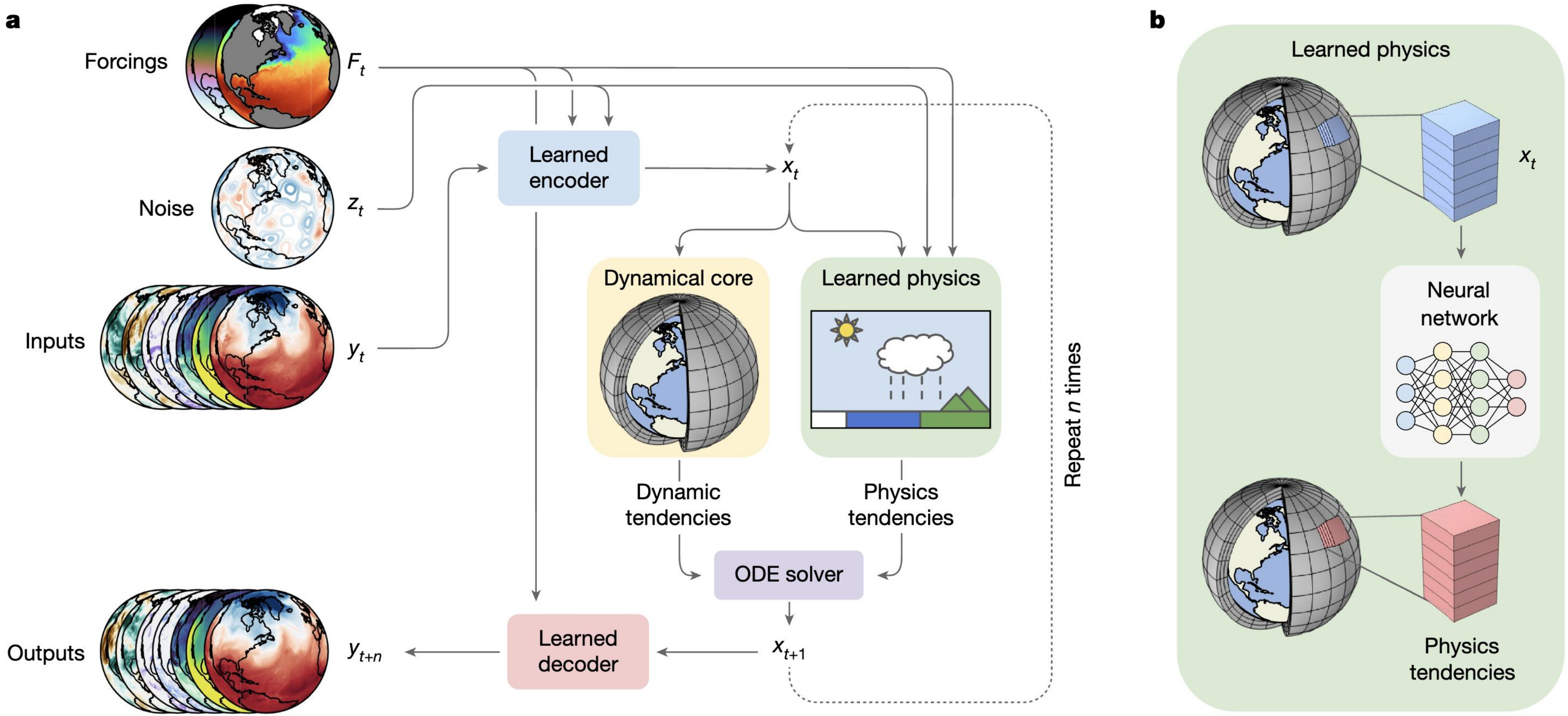
Received: 13 November 2023

Accepted: 15 June 2024

Published online: 22 July 2024

Dmitrii Kochkov[1,6] ✉, Janni Yuval[1,6] ✉, Ian Langmore[1,6], Peter Norgaard[1,6], Jamie Smith[1,6], Griffin Mooers[1], Milan Klöwer[2], James Lottes[1], Stephan Rasp[1], Peter Düben[3], Sam Hatfield[3], Peter Battaglia[4], Alvaro Sanchez-Gonzalez[4], Matthew Willson[4], Michael P. Brenner[1,5] & Stephan Hoyer[1,6] ✉

NeuralGCM (Kochkov et al., Nature, 2024)

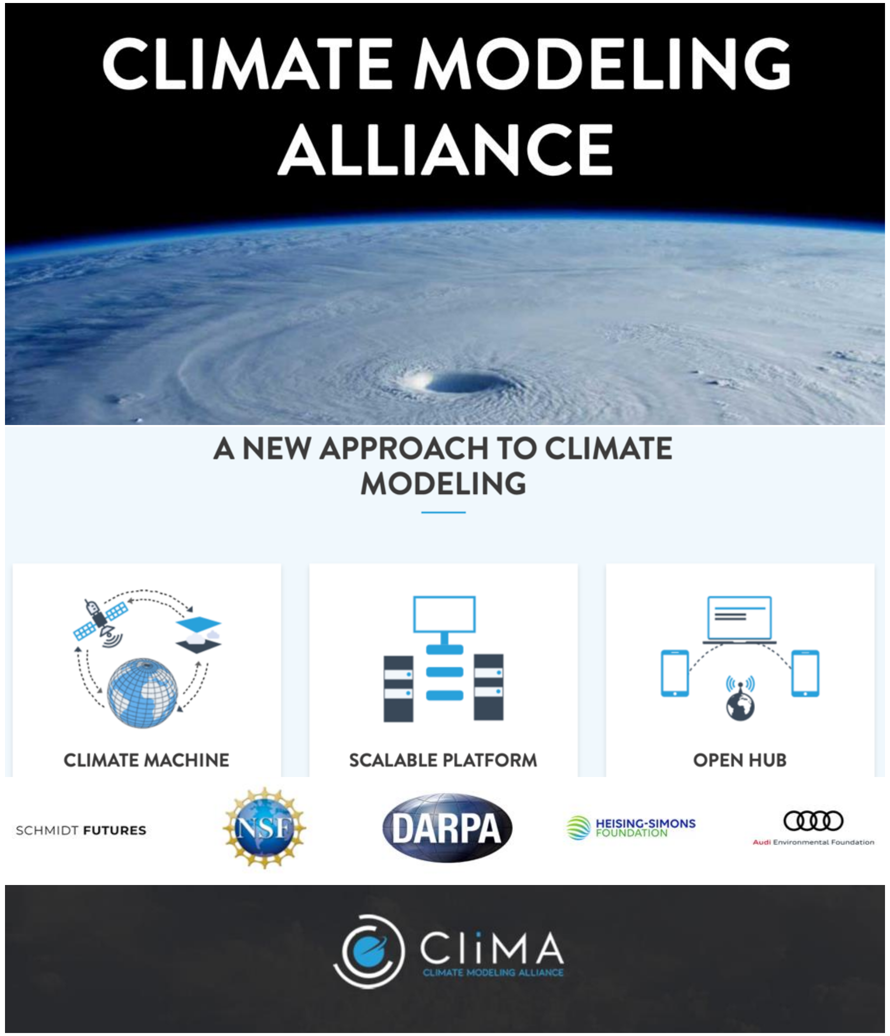


Dynamical core FV3 of atmospheric GCM rewritten in JAX for differentiability via AD

Why Julia?

DJ4Earth

Building on CliMA



Harness next-gen. compute architecture

TOOLBOX

JULIA: COME FOR THE SYNTAX, STAY FOR THE SPEED

Researchers often find themselves coding algorithms in one programming language, only to have to rewrite them in a faster one. An up-and-coming language could be the answer.

1 AUGUST 2019 | VOL 572 | NATURE

SIAM REVIEW
Vol. 59, No. 1, pp. 65–98

Julia: A Fresh Approach to Numerical Computing*

Oceananigans.jl & ClimaOcean.jl: The ocean model component of the *Climate Model Alliance (CliMA) model*

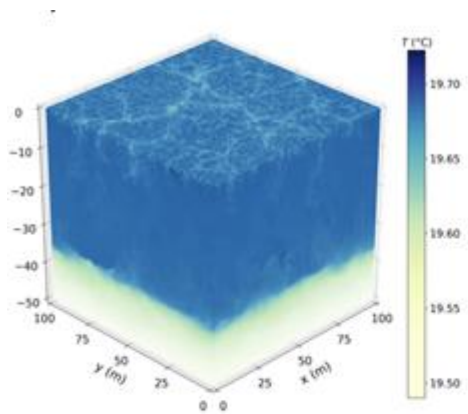
The Journal of Open Source Software

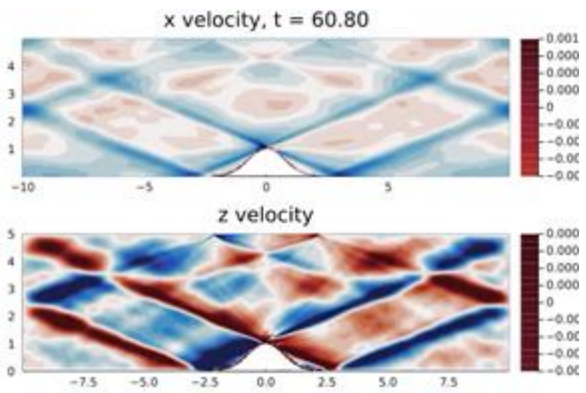
Oceananigans.jl: Fast and friendly geophysical fluid dynamics on GPUs

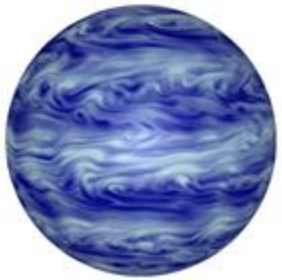
Ali Ramadhan[1], Gregory LeClaire Wagner[1], Chris Hill[1], Jean-Michel Campin[1], Valentin Churavy[1], Tim Besard[2], Andre Souza[1], Alan Edelman[1], Raffaele Ferrari[1], and John Marshall[1]

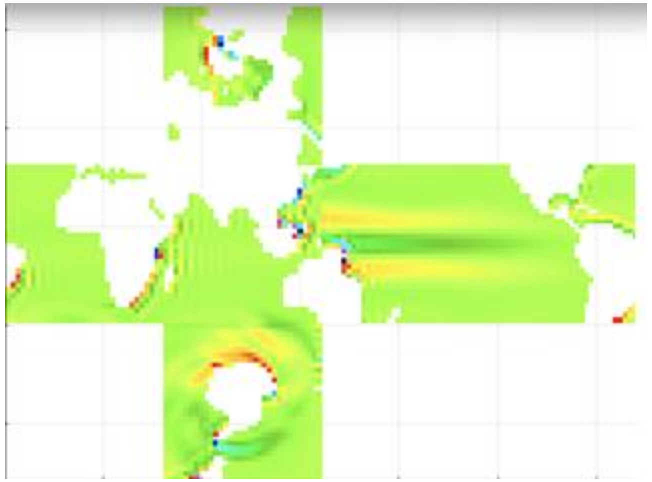
1 Massachusetts Institute of Technology 2 Julia Computing, Inc.

- Finite volume, rotating, stratified fluids model for geophysical fluid dynamics (GFD).
- Written from scratch in Julia, rooted in finite volume algorithms of M.I.T General Circulation Model (MITgcm).
- Multiple simulation options.
- GPU and CPU via kernel abstractions
- Parallelize using MPI.jl and multi-threading









https://github.com/clima/Oceananigans.jl

Oceananigans.jl & ClimaOcean.jl: The ocean model component of the *Climate Model Alliance (CliMA) model*

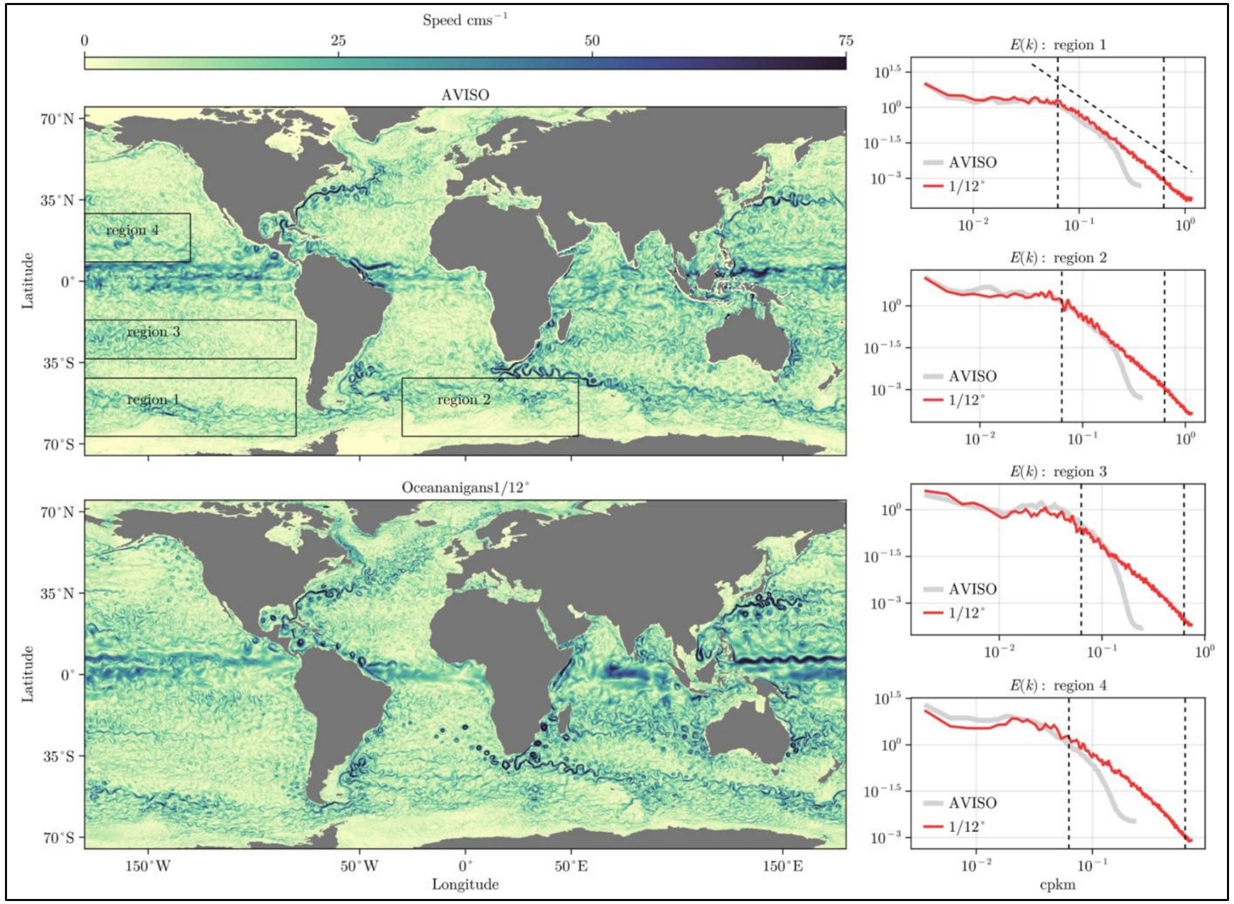
JAMES | Journal of Advances in Modeling Earth Systems®

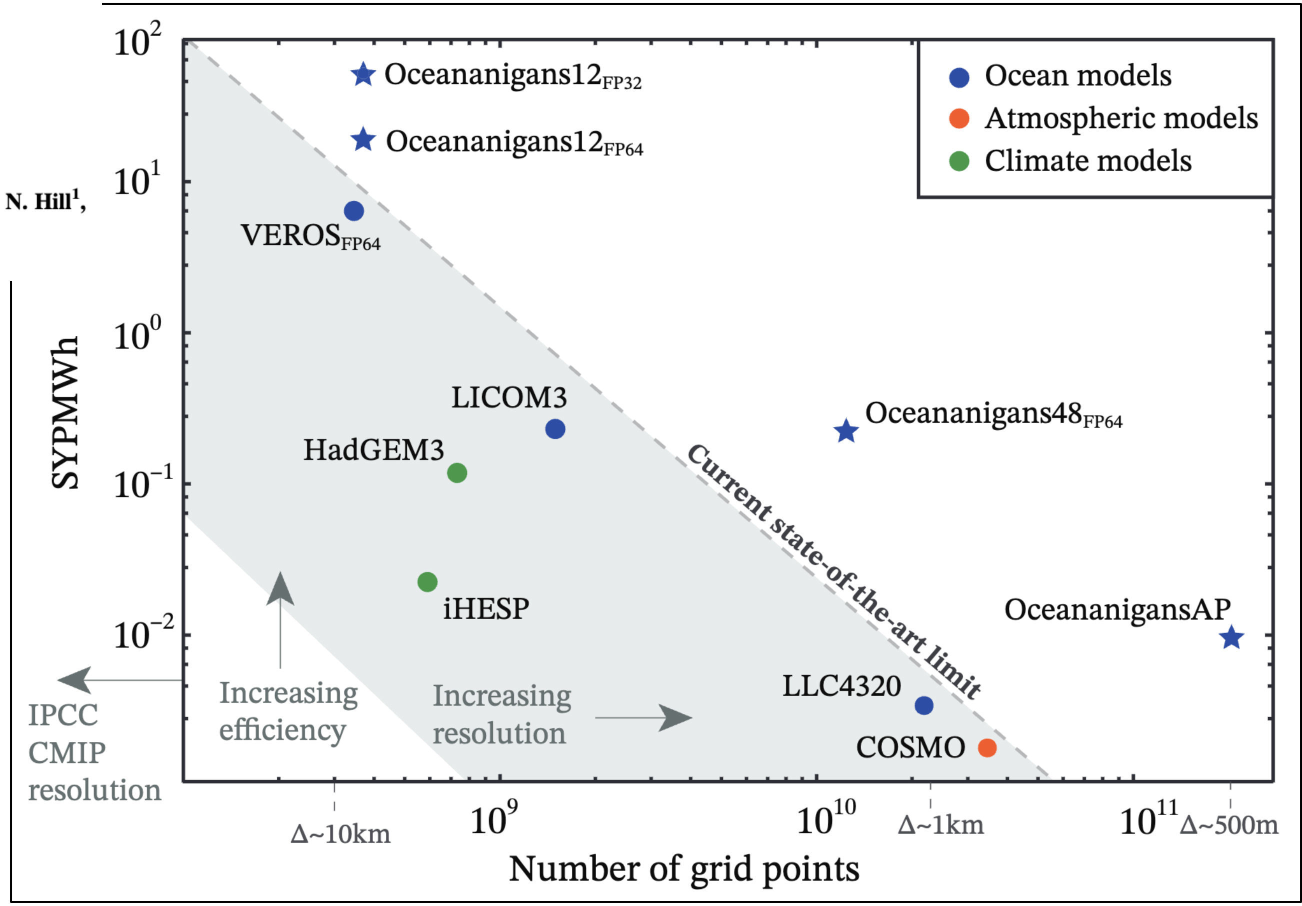
RESEARCH ARTICLE
10.1029/2024MS004465

Special Collection:
The CliMA Earth System Model

A GPU-Based Ocean Dynamical Core for Routine Mesoscale-Resolving Climate Simulations

Simone Silvestri[1], Gregory L. Wagner[1], Navid C. Constantinou[2,3], Christopher N. Hill[1], Jean-Michel Campin[1], Andre N. Souza[1], Siddhartha Bishnu[1], Valentin Churavy[1], John Marshall[1], and Raffaele Ferrari[1]





https://github.com/clima/Oceananigans.jl
https://github.com/clima/ClimaOcean.jl

Differentiable programming for full-model / end-to-end learning

Differentiating a GPU-enabled ocean model

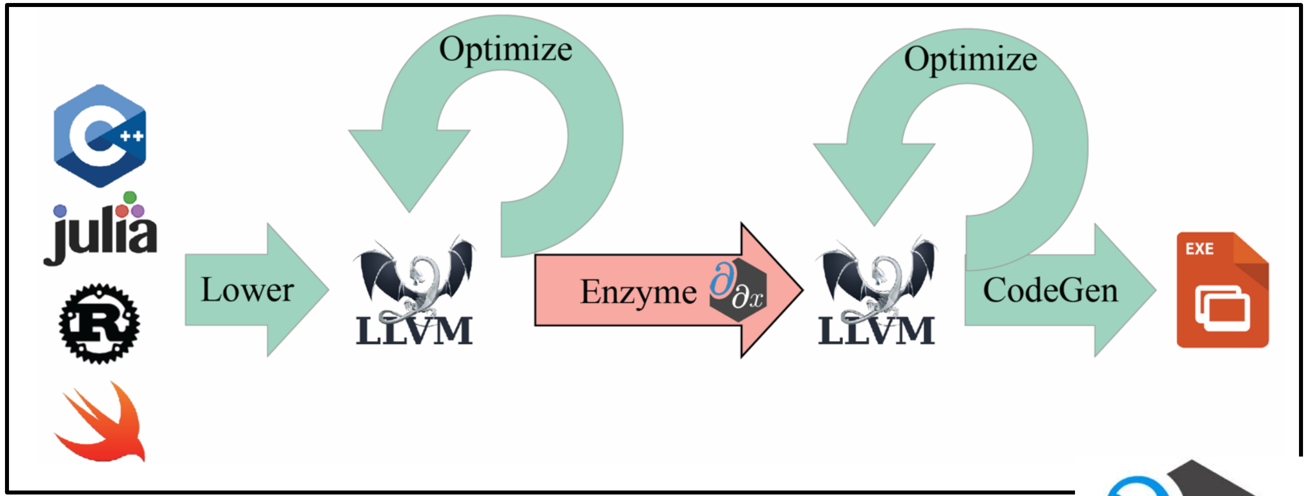
in Julia via AD & transpilation tools

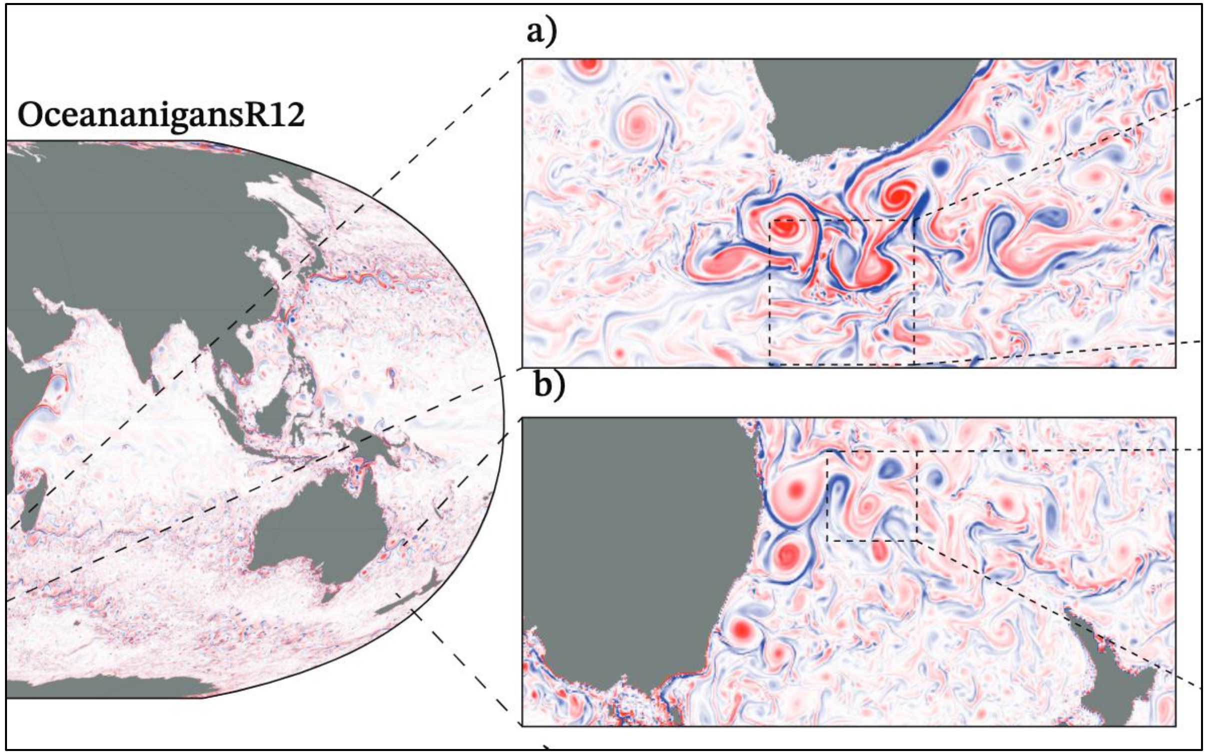
Enzyme.jl & Reactant.jl

DJ4Earth

Oceananigans.jl & ClimaOcean.jl

Silvestri et al., JAMES, (2025);Wagner et al., arXiv (2025)





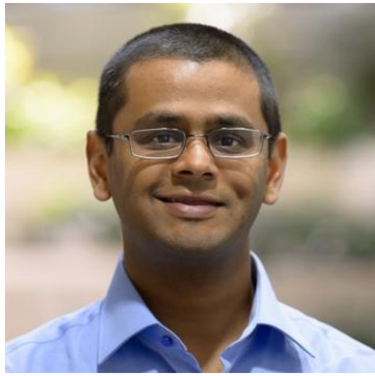
***Enzyme.jl**: Moses et al., SC'22, ICCS'23*
***Reactant.jl**: PPoPP '23; Moses et al. (2025) submitted.*

Enzyme: Fast, Parallel, and Rewrite-Free Derivatives

- Derivatives are ubiquitous in machine learning (training neural networks, Bayesian inference), scientific computing (uncertainty quantification, simulation)

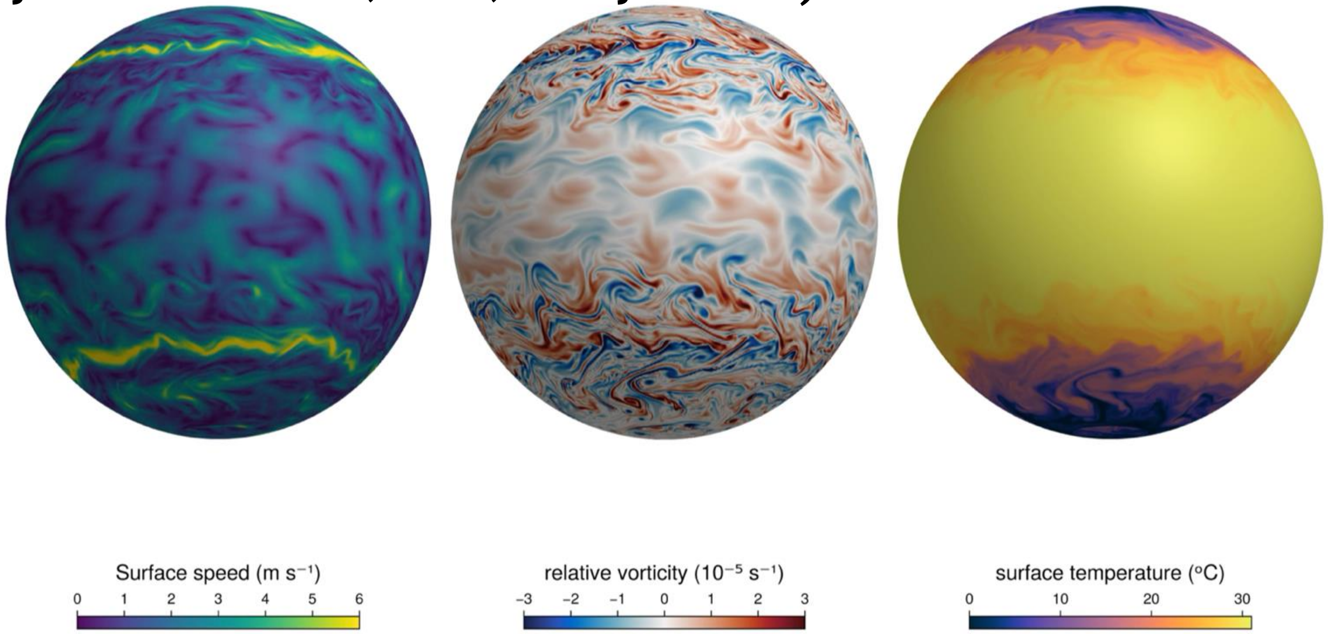
- Enzyme synthesizes derivatives of arbitrary code within the compiler
  - Differentiate code in any LLVM-based language (C/C++, Julia, Rust, Swift, Fortran, Python, etc) *without rewriting it*!
  - Operating after and alongside program optimization generates asymptotically and empirically faster derivatives
  - First automatic differentiation tool to handle arbitrary GPU kernels



Reactant.jl - https://enzymead.github.io/Reactant.jl

- **Optimize Julia Functions With Multi-Level Intermediate Representation (MLIR) and XLA for High-Performance Execution on CPU, GPU, TPU and more**
- Preserves high-level structures from Julia code into the compiler
  - Effectively optimize programs
  - Effectively retarget programs (allowing them to run on distributed clusters of your favorite accelerator, including CPU/GPU/TPU)
- Can leverage XLA (SOTA runtime, used by TensorFlow, JaX, & PyTorch)
- Compatible with mutation, control flow
- Compiled single-node CUDA version of Oceananigans.jl to execute on thousands of TPUs and GPUs



Moses et al., 2025, submitted

Conclusion

- Oceanography (like much of geoscience) is a ***sparse-data science (obs. data)***
- Important to use information encapsulated in equations of motion
  - ***inverse methods ("data assimilation")*** provide a framework for doing so
- Correspondence between ***adjoint and backpropagation*** in ML should be exploited in coupling physics models & surrogate models
  - ***full-model learning/calibration/estimation/initialization***
- *Julia language* provides promising opportunity through its focus on ***differentiable programming*** & *general-purpose AD*
- Powerful **uses of gradients beyond optimization/"learning"**
  - Reduced-order modeling & dynamical attribution using sensitivity kernels
  - Hessian-based uncertainty quantification (UQ)
  - Optimal Experimental Design (OED)