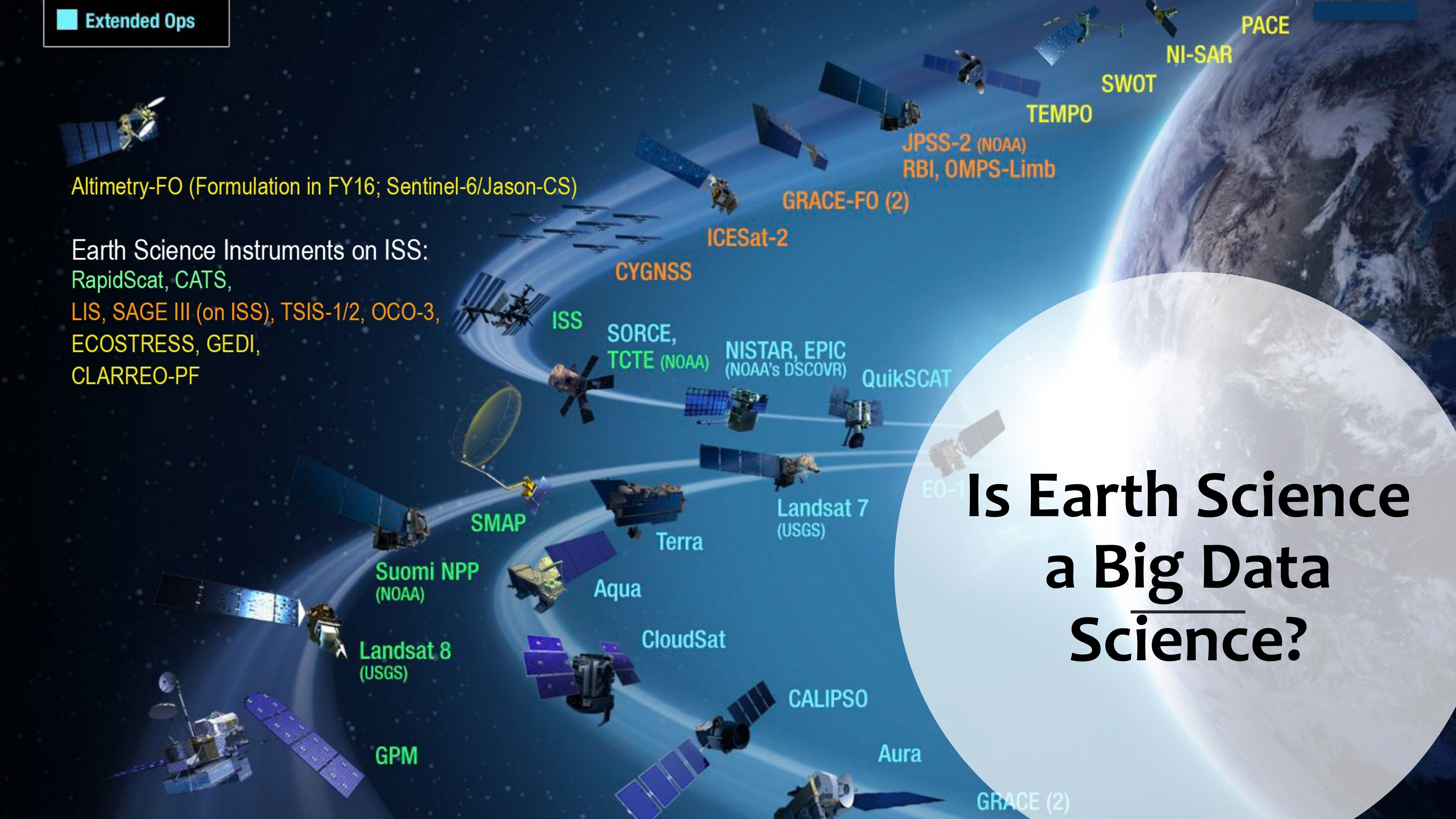
ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

LEARNING FROM (SPARSE) OBSERVATIONS THROUGH THE LENS OF MODELS

Patrick Heimbach, ***the DJ4Earth & ECCO groups***
The University of Texas at Austin, TX

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

Is Earth Science a Big Data Science?

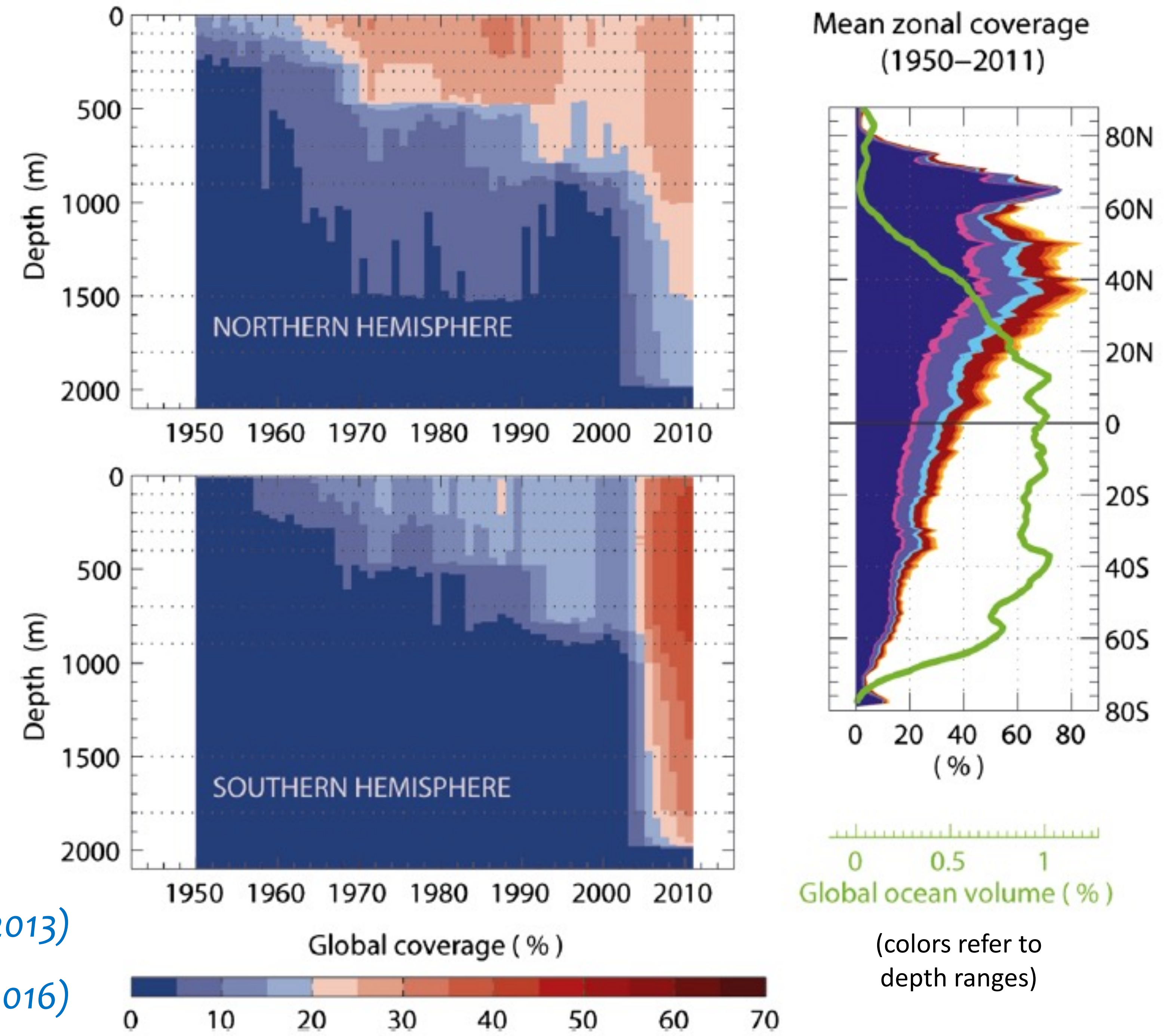


Is Oceanography a “big data” science?

Yes & No …

Oceanography: A *sparse data* problem ...

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



(colors refer to depth ranges)

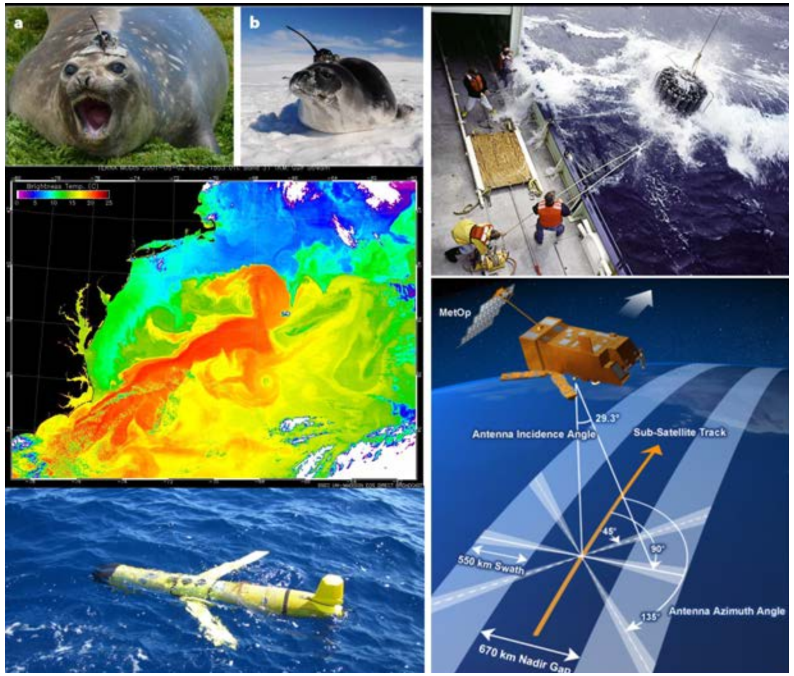
Abraham et al., Rev. Geophys. (2013)

Wunsch (2016)

Two incomplete knowledge reservoirs

an eclectic, patchy, heterogeneous

observing system



numerical models

that require

uncertain

inputs

Session 1: Ocean Dynamics and Circulation

1. How have the recent advances in ocean modeling helped our understanding of the ocean's dynamics and the role of the ocean in the climate system? How does understanding dynamics feed into improving OGCMs?
2. How can we better use observations to evaluate and advance ocean models? Do we have the observations needed, including for evaluating high-resolution models?
3. Is the ocean modeling community tackling the relevant problems, including model development efforts?
4. What are your thoughts on a hierarchical modeling approach from coarse to ultra-high resolution modeling?

What is Data Assimilation / Inverse Modeling?

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."

What is Data Assimilation / Inverse Modeling?

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):

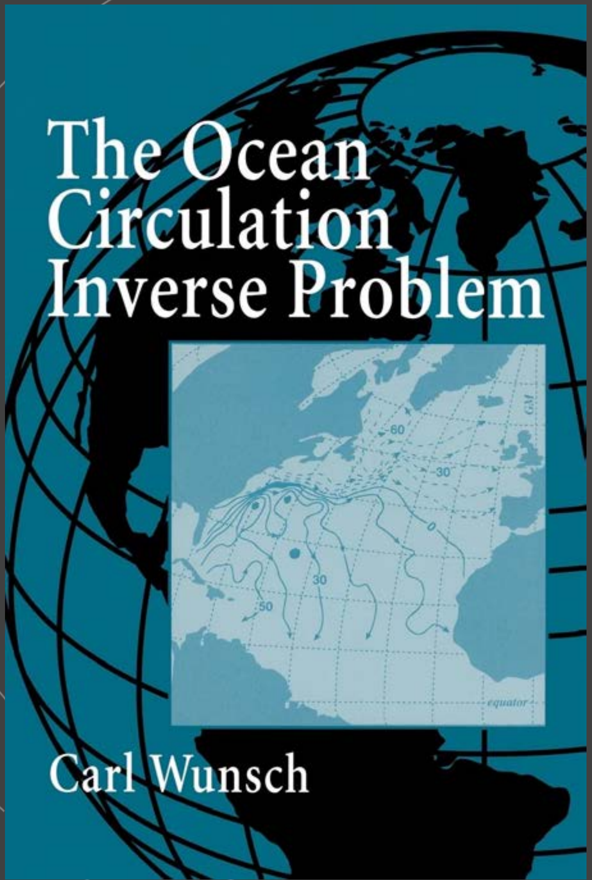
"DA allows information provided from observations to be propagated in time and space to unobserved areas using the dynamical and physical constraints imposed by numerical models."

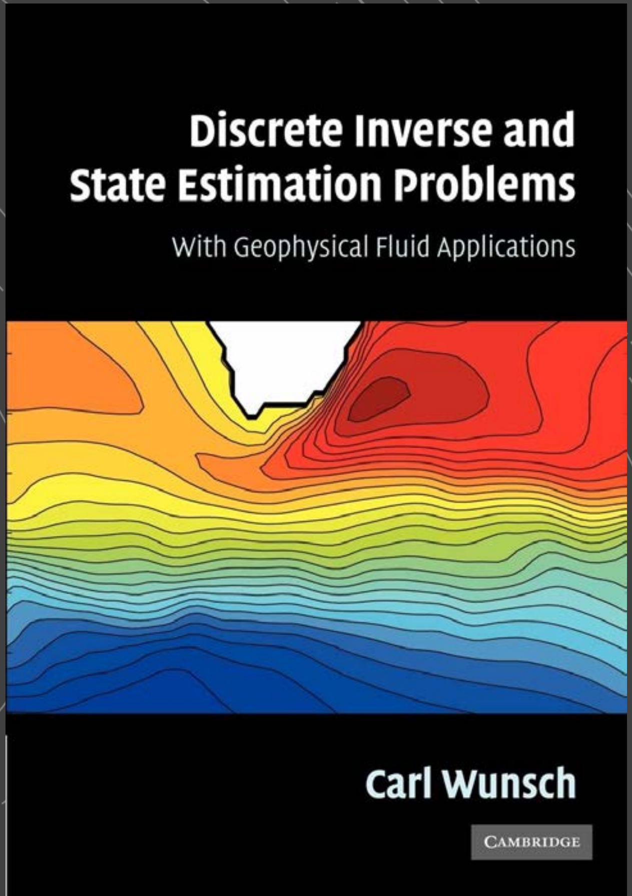
Data Assimilation and Inverse Modeling

The DA / inverse problem is learning from …

- a set of available (usually sparse, heterogeneous) observations
- … *AND* known physics/dynamics,
- … by solving a gigantic least-squares model-data misfit minimization

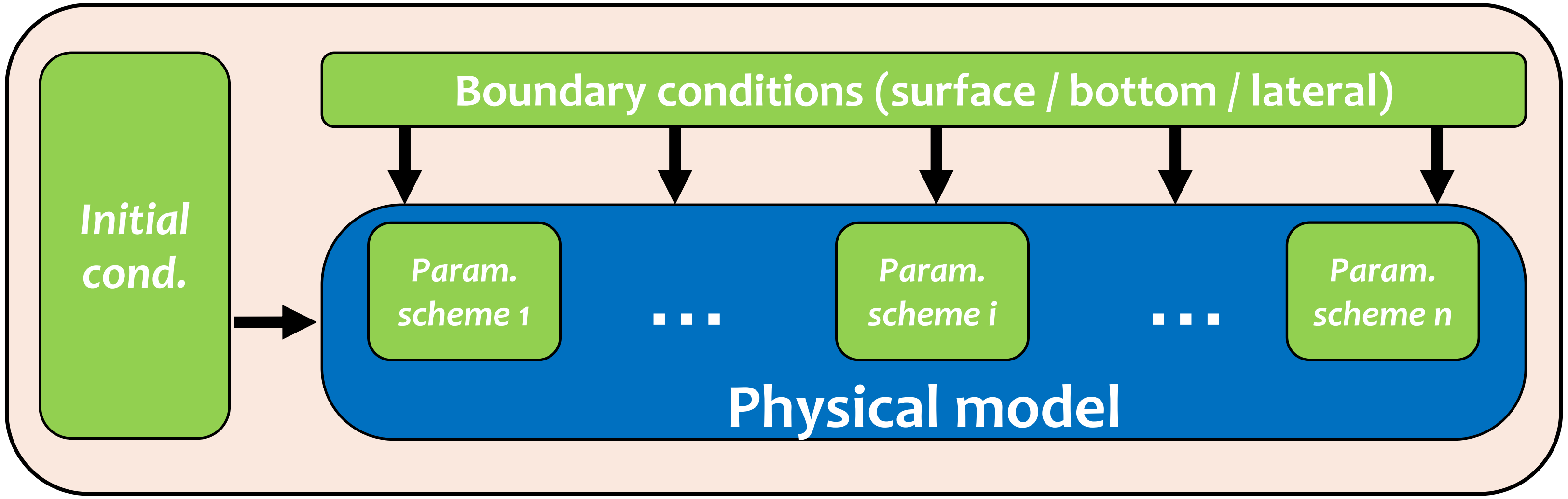
“Data assimilation” is much more than its use in numerical weather prediction





What do we mean by ***"Learning"?***

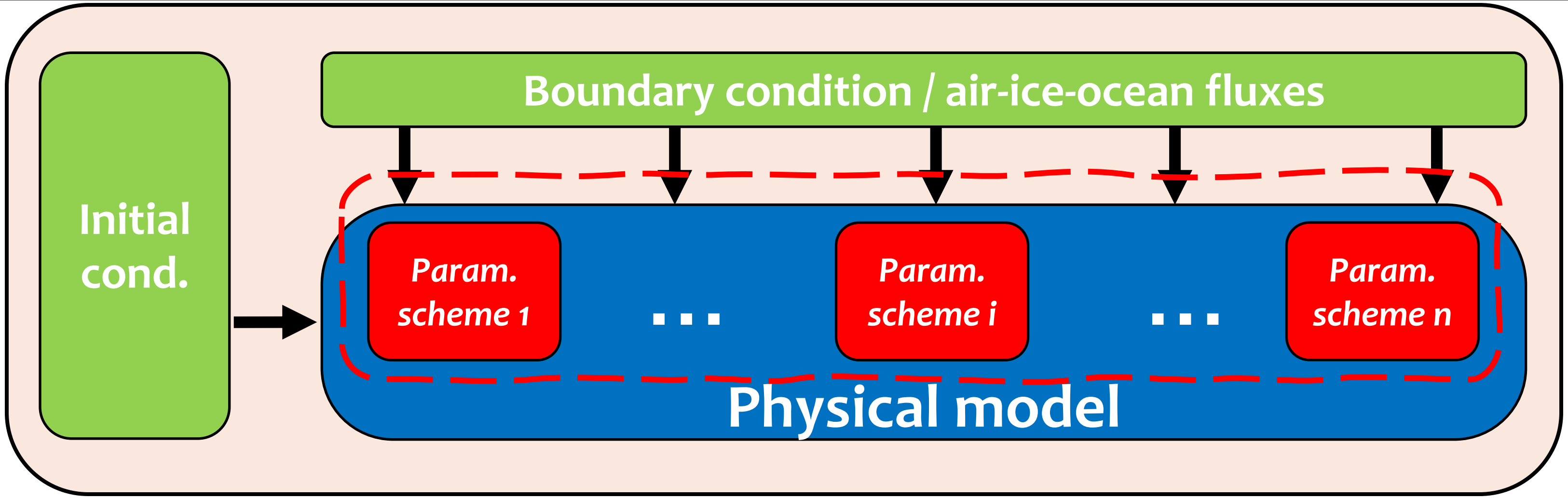
Learn ...


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

Learn model *parameters*

Physical model has many empirical parameters:

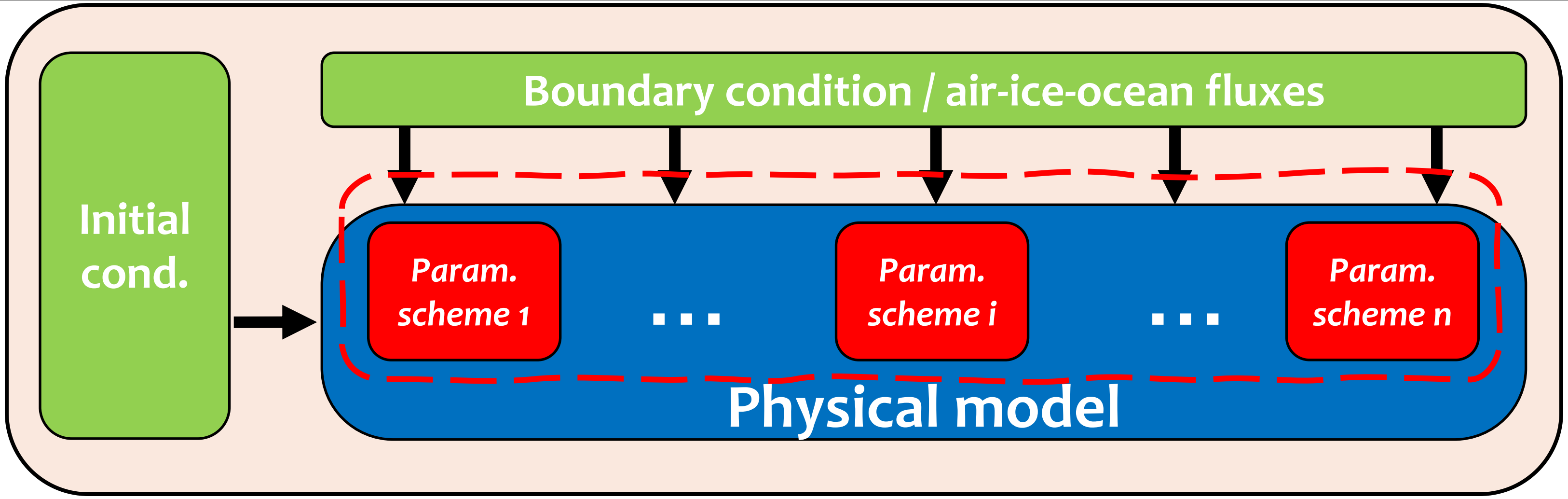
- constitutive laws
- subgrid-scale parameterization schemes



Learn model *parameters*

Physical model has many empirical parameters:

- constitutive laws
- subgrid-scale parameterization schemes



parameter estimation using observations is essential

THE ART AND SCIENCE OF CLIMATE MODEL TUNING

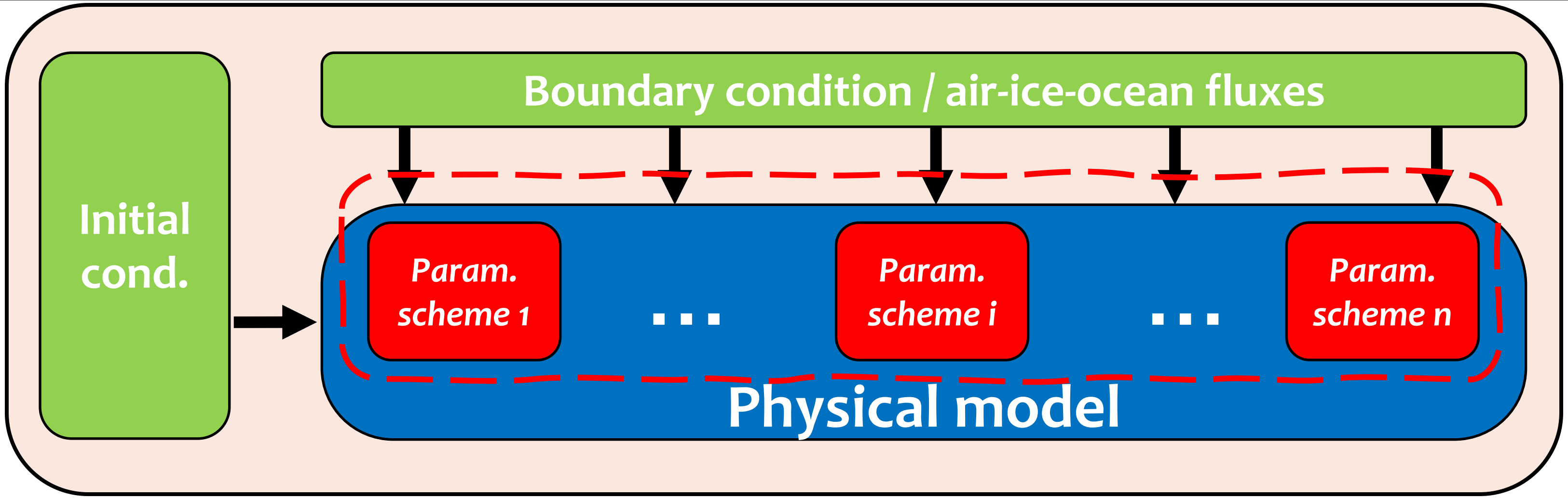
Frédéric Hourdin, Thorsten Mauritsen, Andrew Gettelman, Jean-Christophe Golaz, Venkatramani Balaji, Qingyun Duan, Doris Folini, Duoying Ji, Daniel Klocke, Yun Qian, Florian Rauser, Catherine Rio, Lorenzo Tomassini, Masahiro Watanabe, and Daniel Williamson

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.

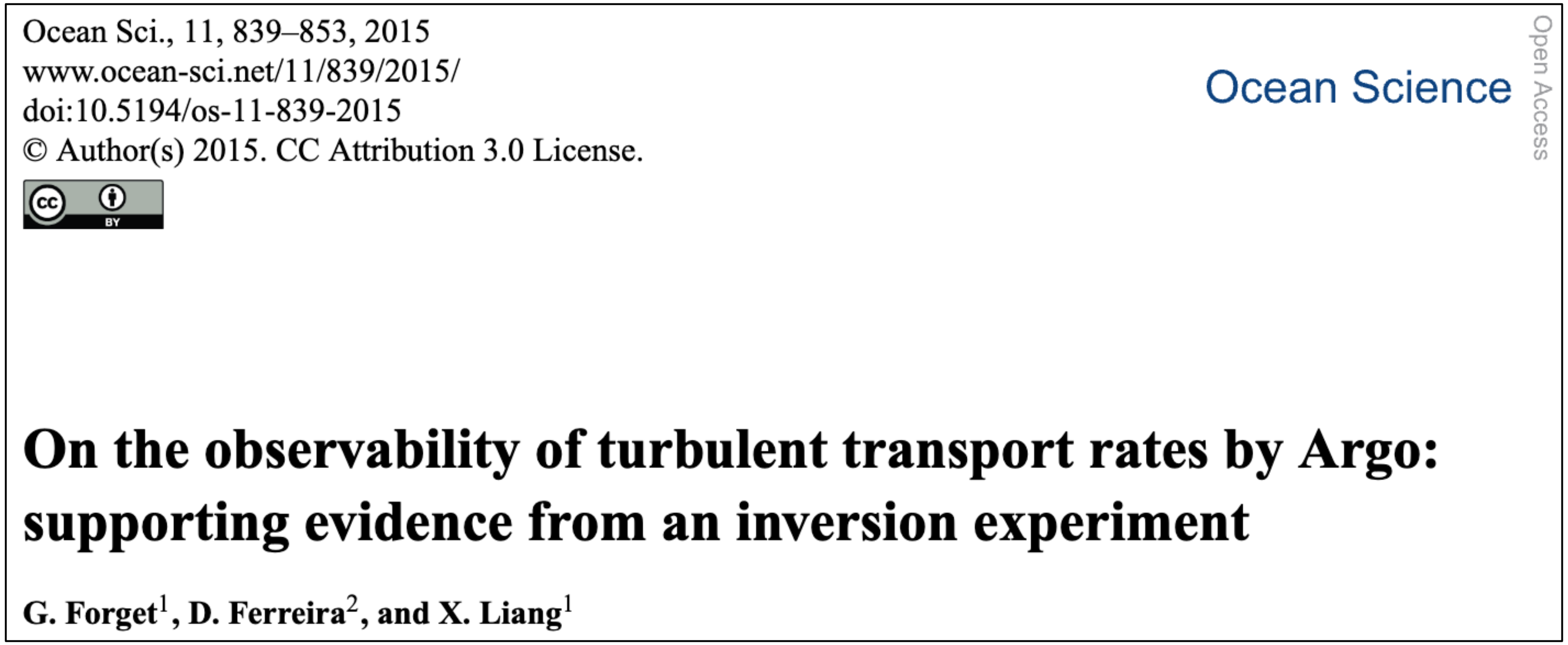
Learn model *parameters*

Physical model has many empirical parameters:

- constitutive laws
- subgrid-scale parameterization schemes



parameter estimation to calibrate model parameters



Ocean Sci., 11, 839–853, 2015
www.ocean-sci.net/11/839/2015/
doi:10.5194/os-11-839-2015
© Author(s) 2015. CC Attribution 3.0 License.

Ocean Science

Open Access

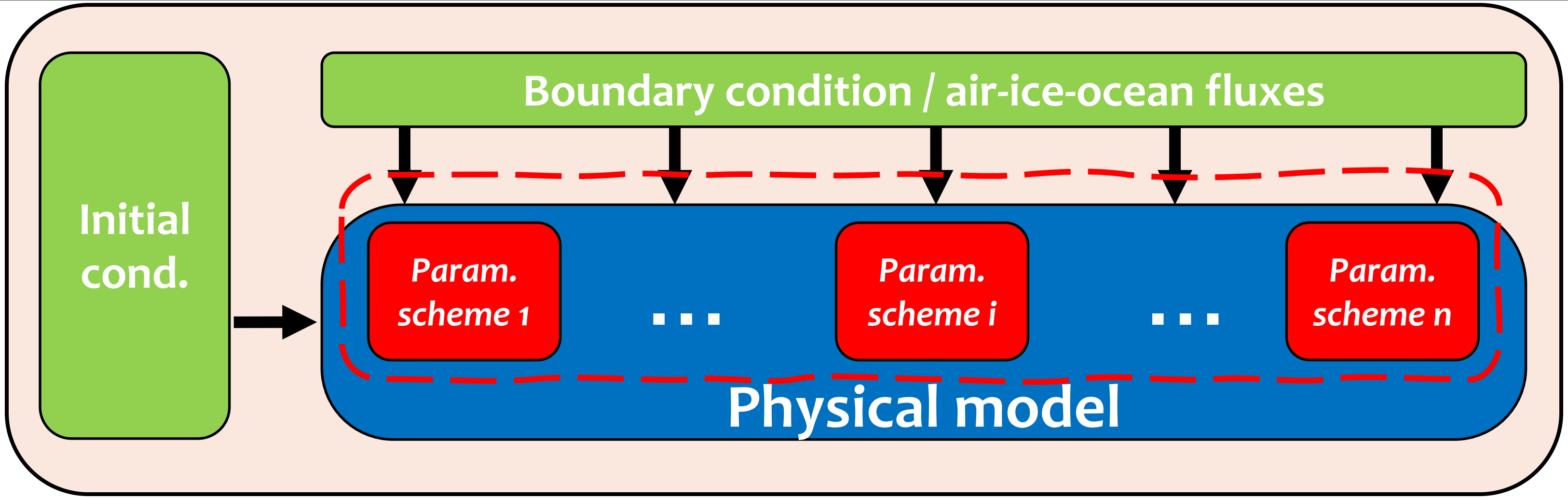
On the observability of turbulent transport rates by Argo: supporting evidence from an inversion experiment

G. Forget[1], D. Ferreira[2], and X. Liang[1]

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

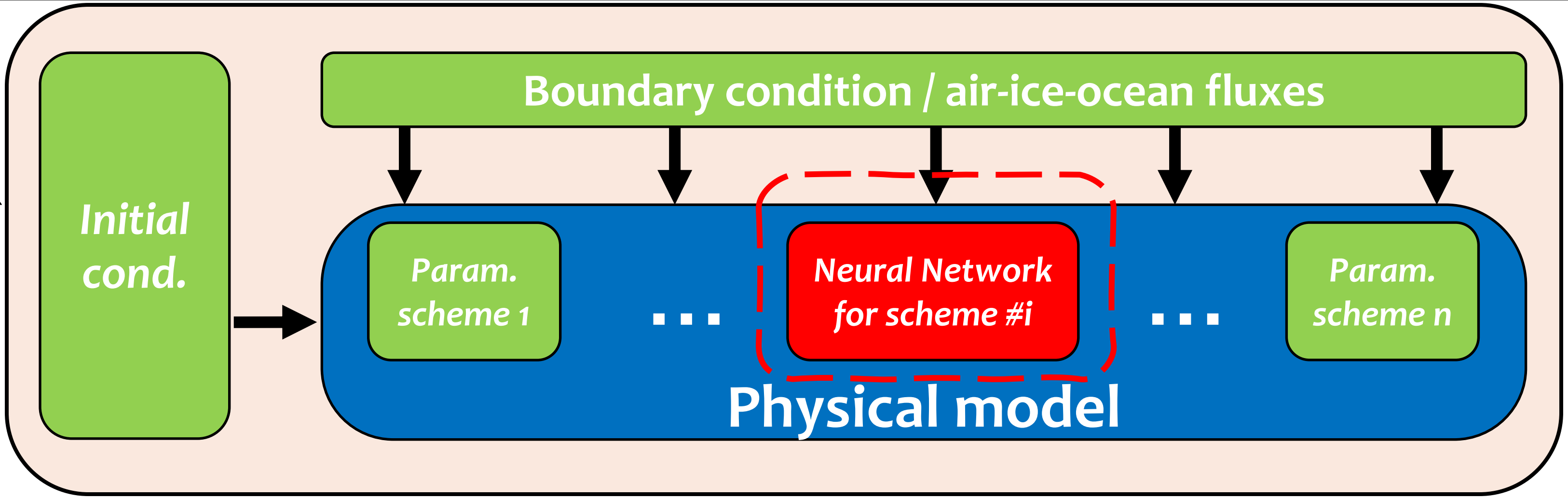
Tracer and observationally derived constraints on diapycnal diffusivities in an ocean state estimate

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

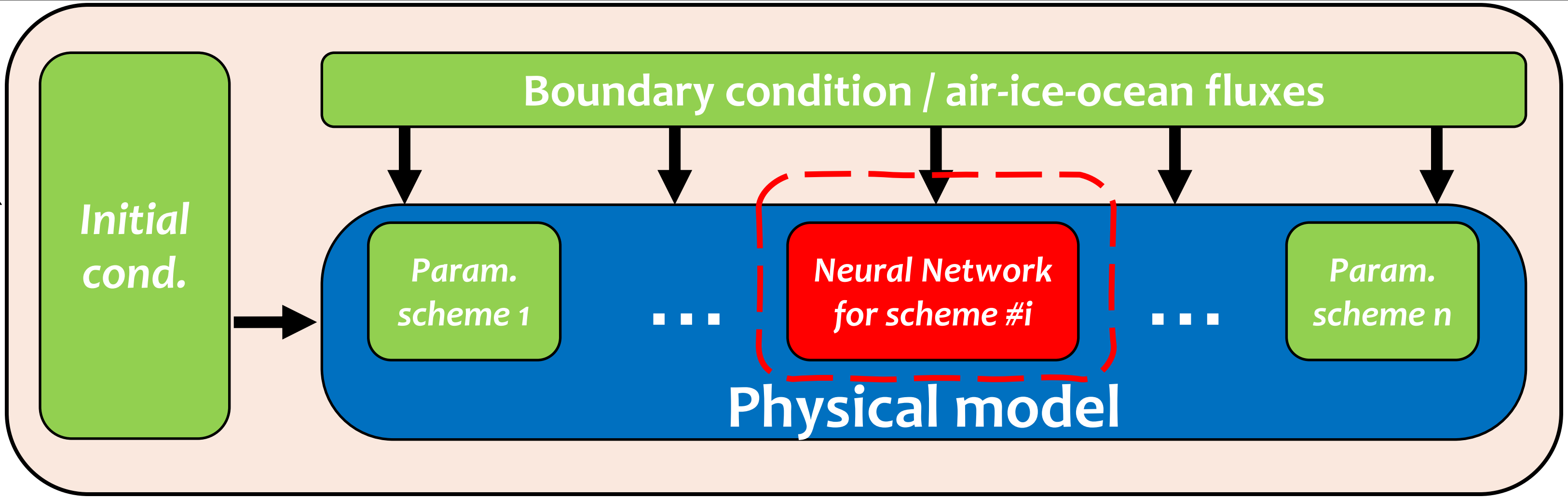
NN is trained on high-fidelity simulation data which resolve scales to be parameterized



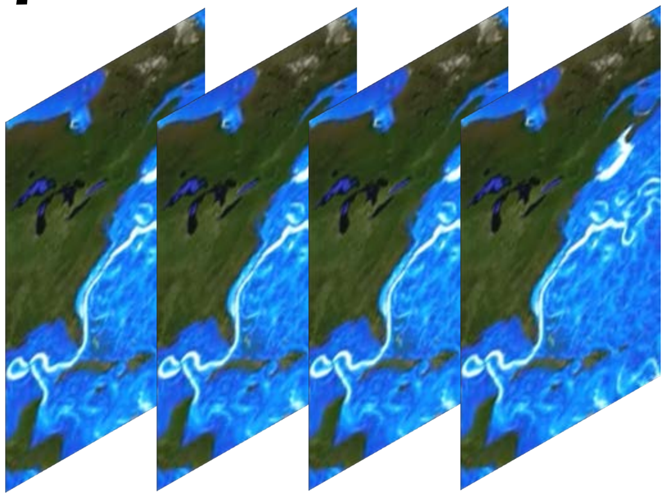
Learn surrogate (e.g., NN) of *model's parameterization scheme*

Parameterization scheme(s) is replaced by neural network

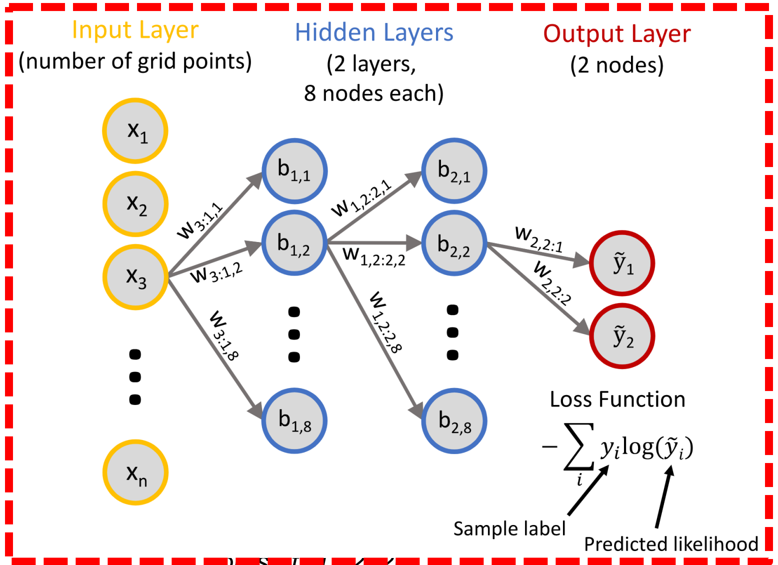
NN is trained on high-fidelity simulation data which resolve scales to be parameterized



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



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



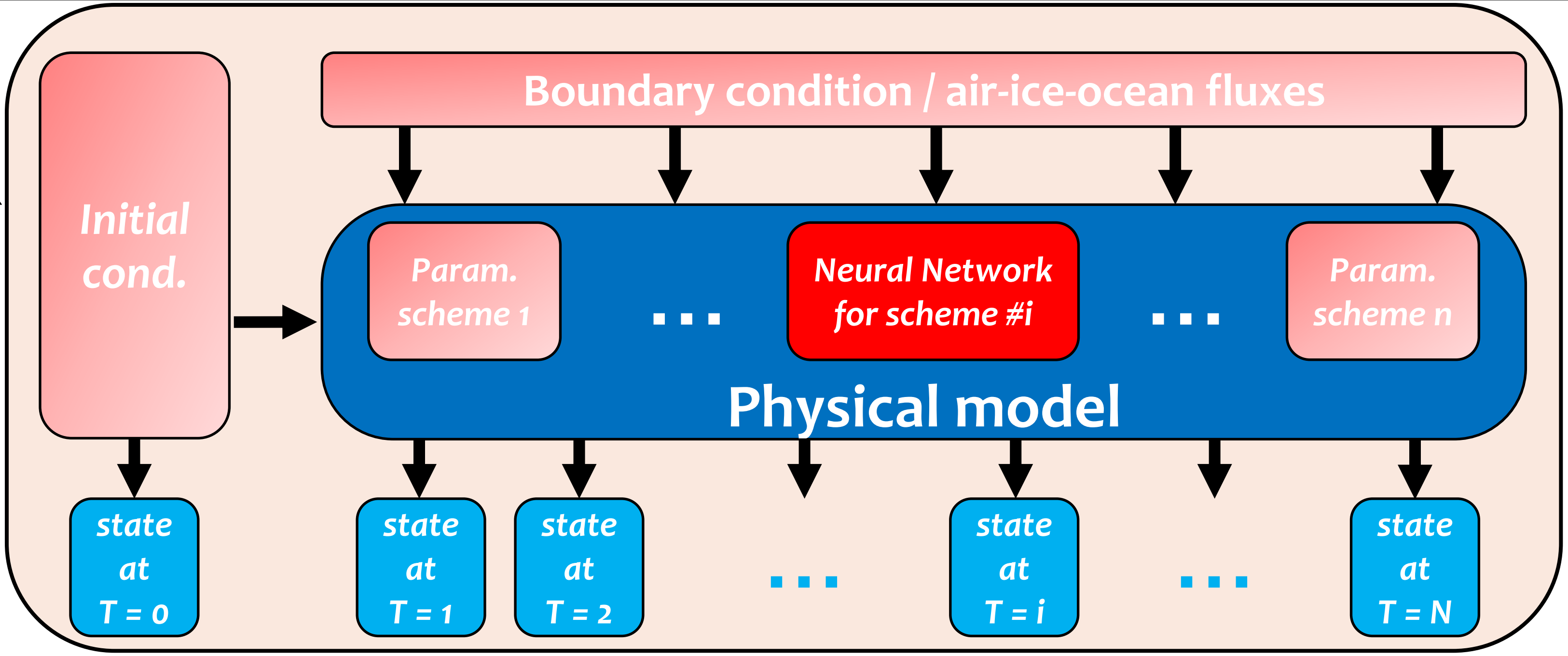
a priori / offline learning

Zanna & Bolton (2021)

Learn *hybrid* physical/surrogate (NN) model

Parameterization scheme(s) is replaced by neural network

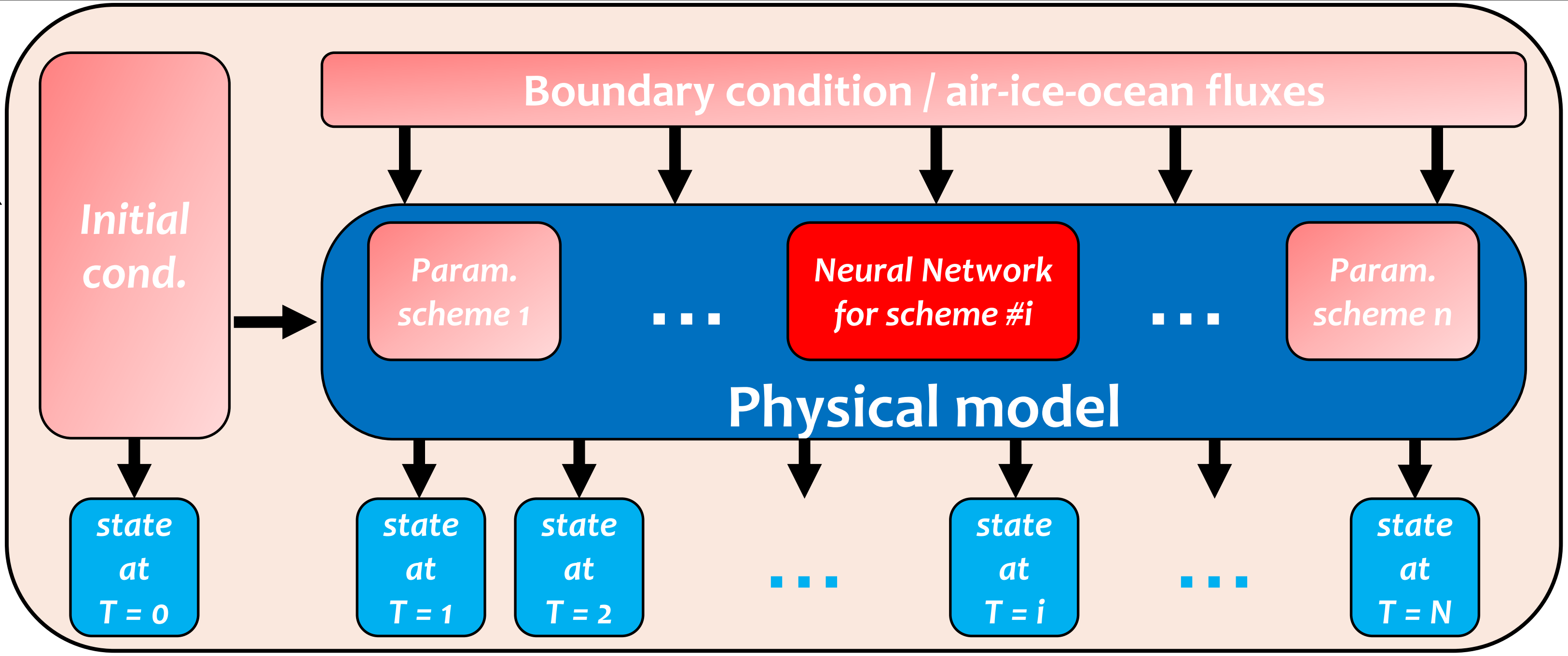
Training of the NN is part of "training" of the physical model on state variables



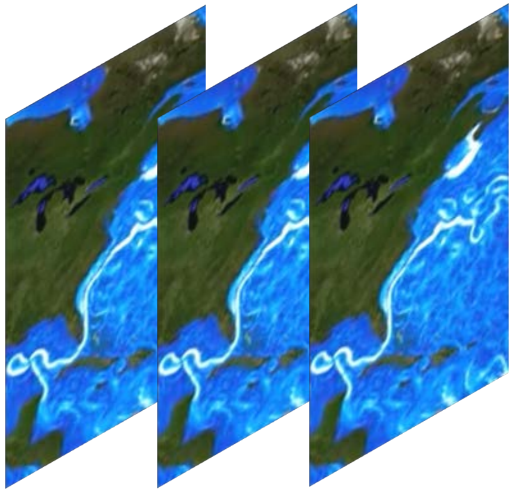
Learn *hybrid* physical/surrogate (NN) model

Parameterization scheme(s) is replaced by neural network

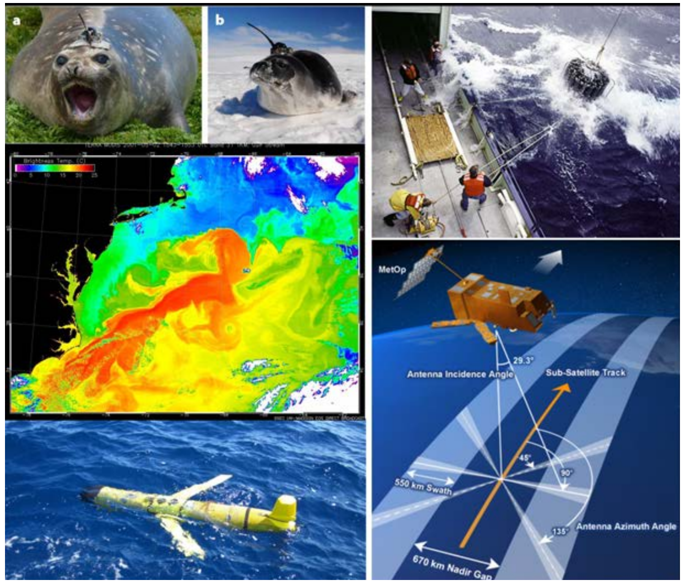
Training of the NN is part of "training" of the physical model on state variables



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



\+



Learn model *initial conditions*

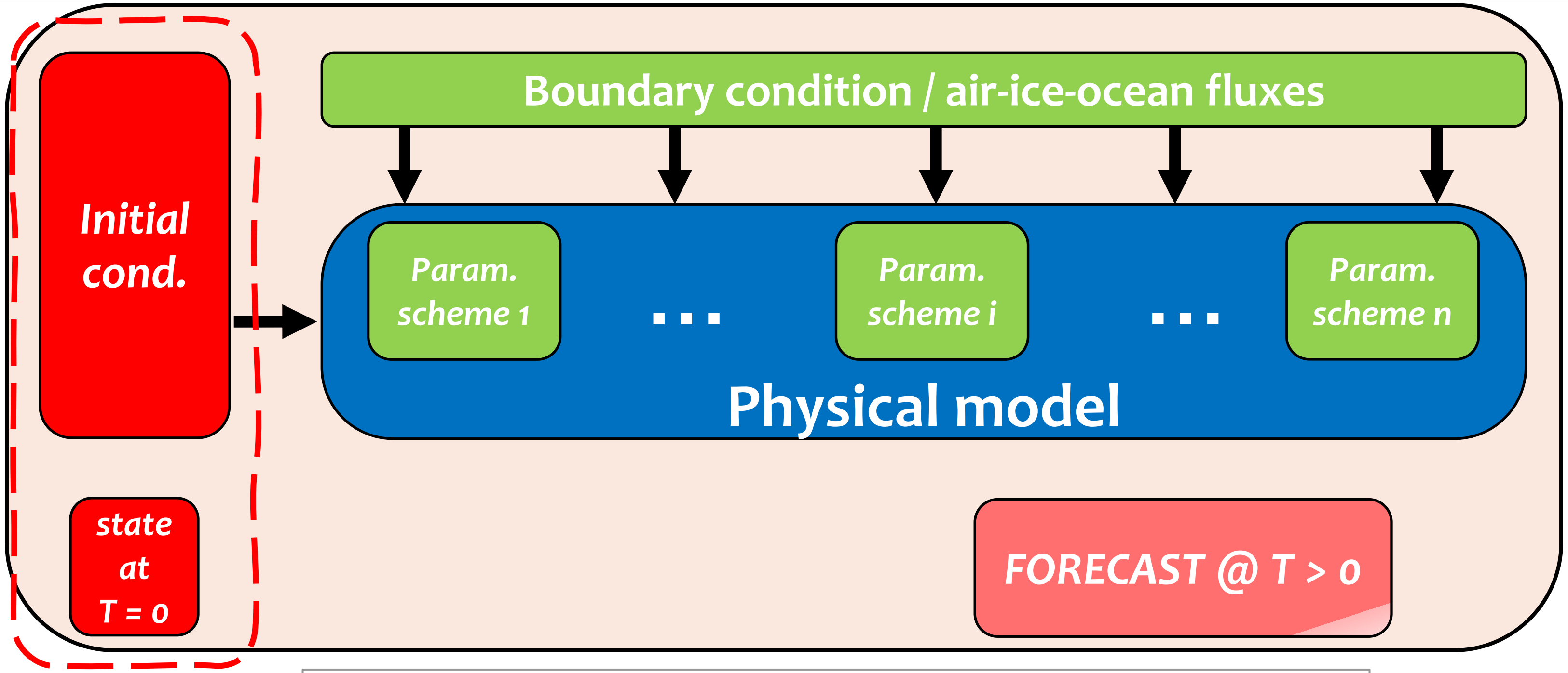
Find best initial conditions that will produce optimal forecast …

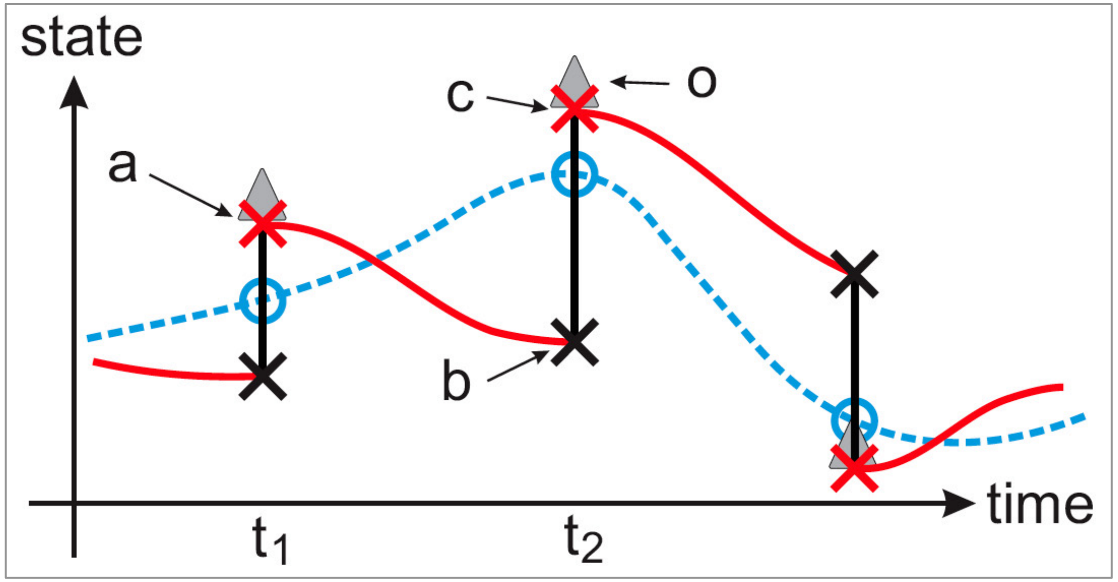
*The **filtering** problem of optimal estimation & control*

Initialization for prediction/extrapolation

as practiced in

numerical weather prediction





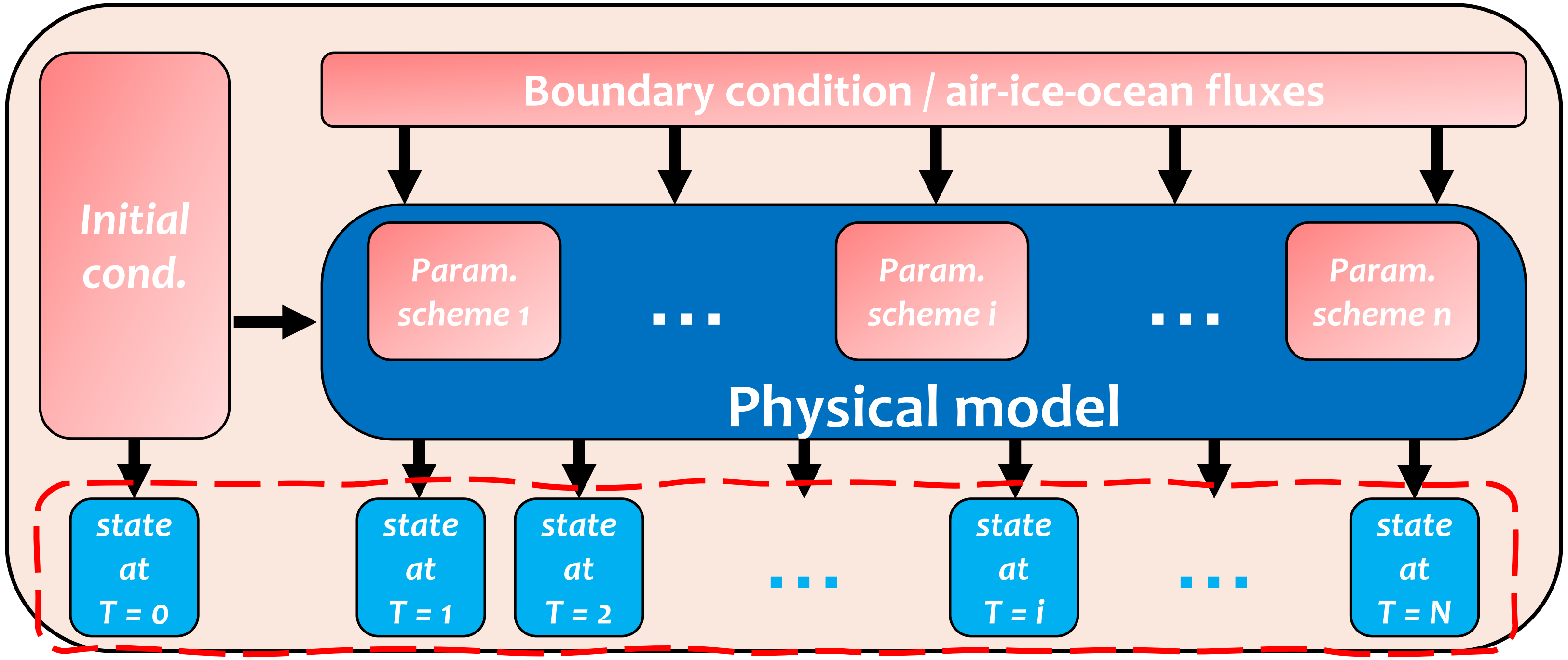
Learn model *time-evolving state*

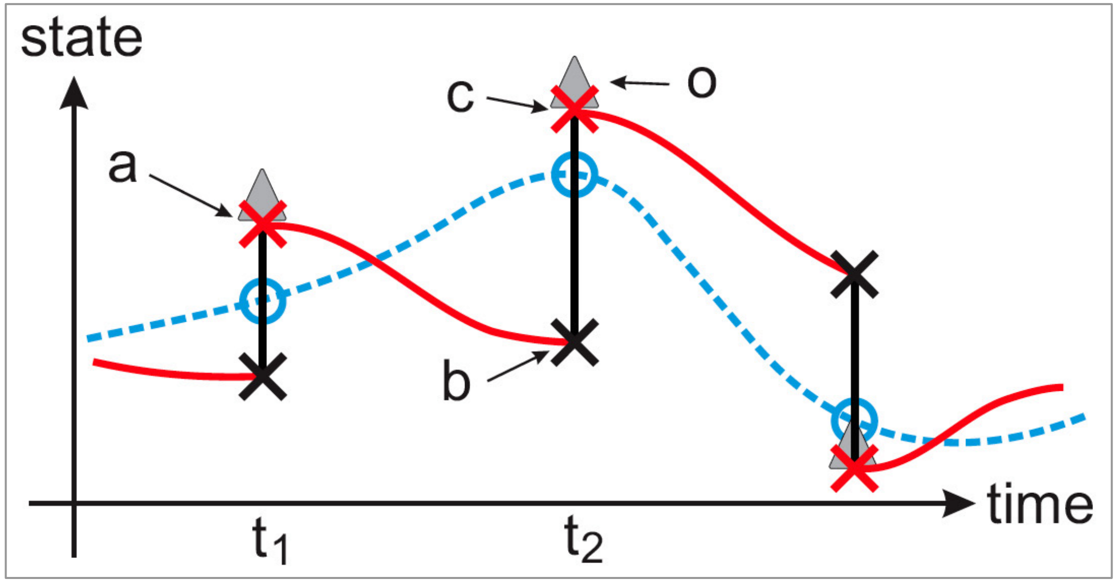
Find model inputs that produce the best dynamically consistent state

The smoothing problem of optimal estimation & control

State & parameter estimation for:

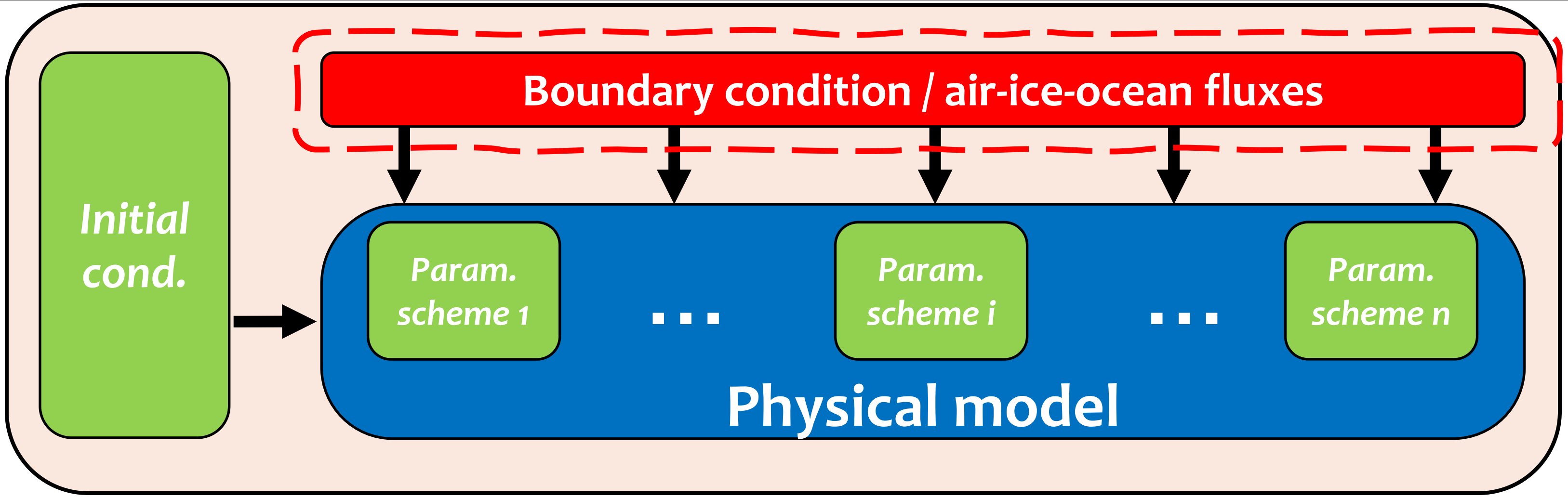
- ***Interpolation/reconstruction***
- ***transient calibration***



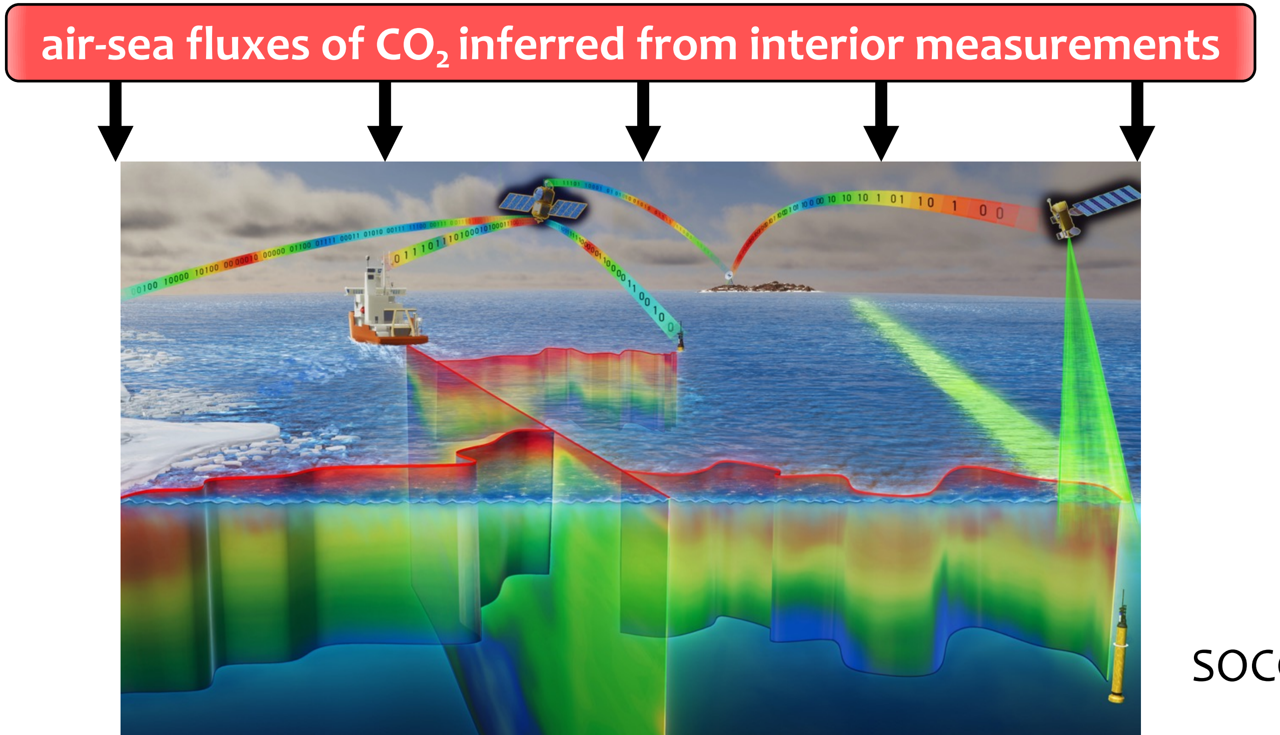


Learn model *boundary conditions*

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)

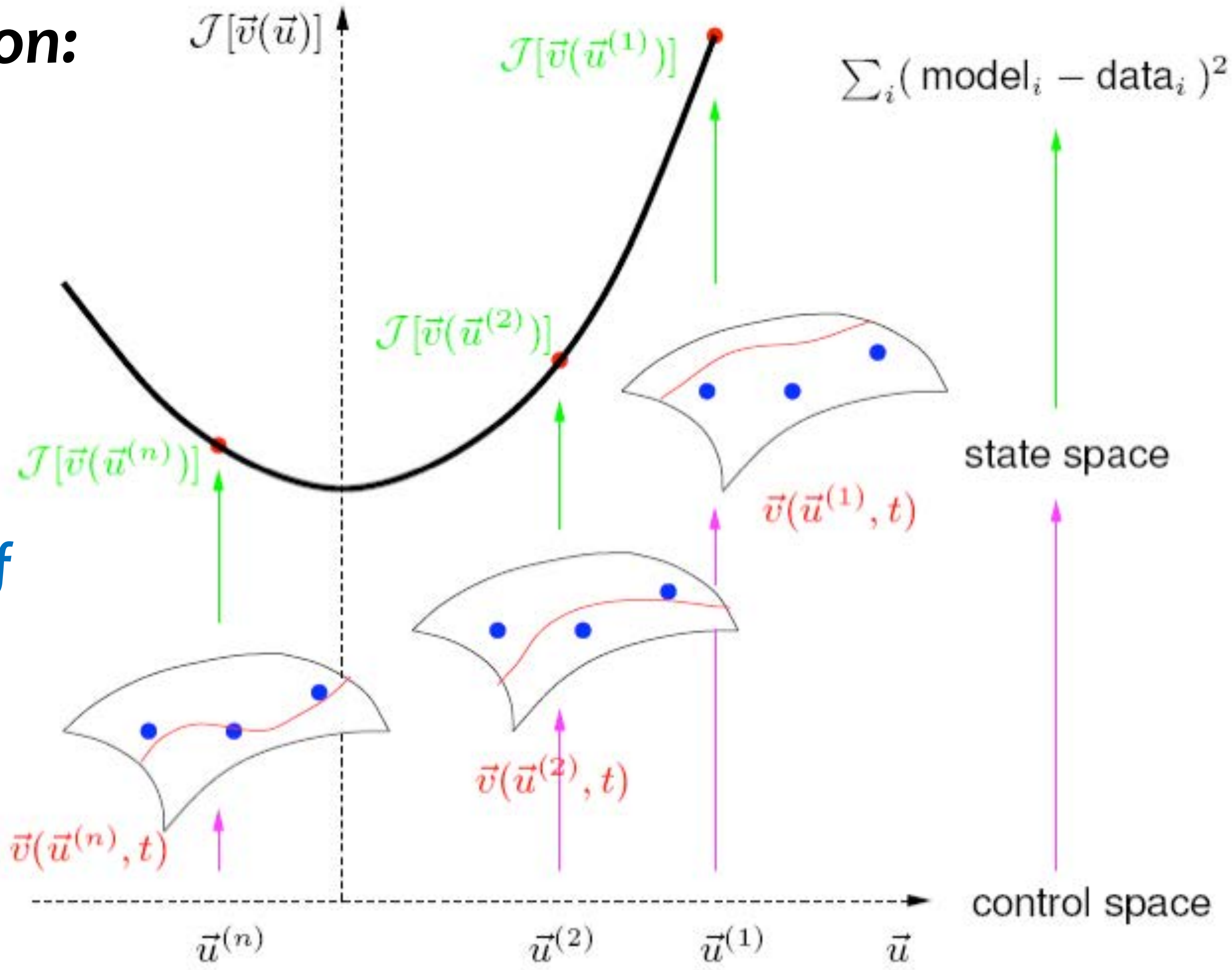
A key unifying computational framework of "*learning from data*"

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***

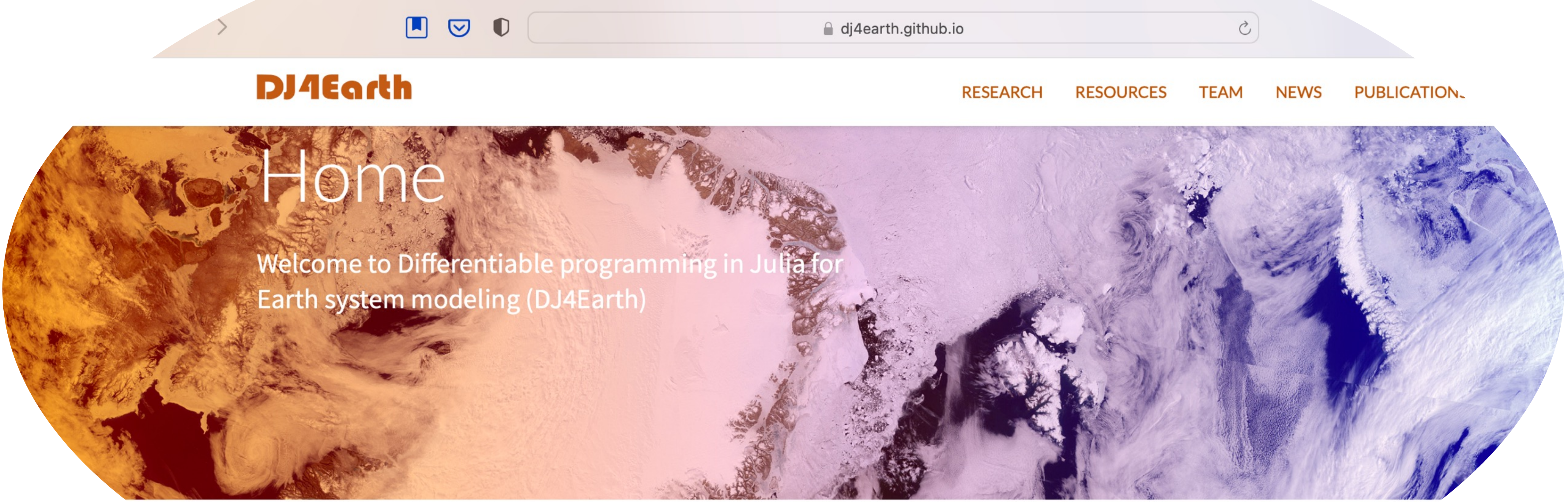
<u>Adjoint / backpropagation</u>

essential tool for computing high-dimensional gradients!



Full-model learning

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



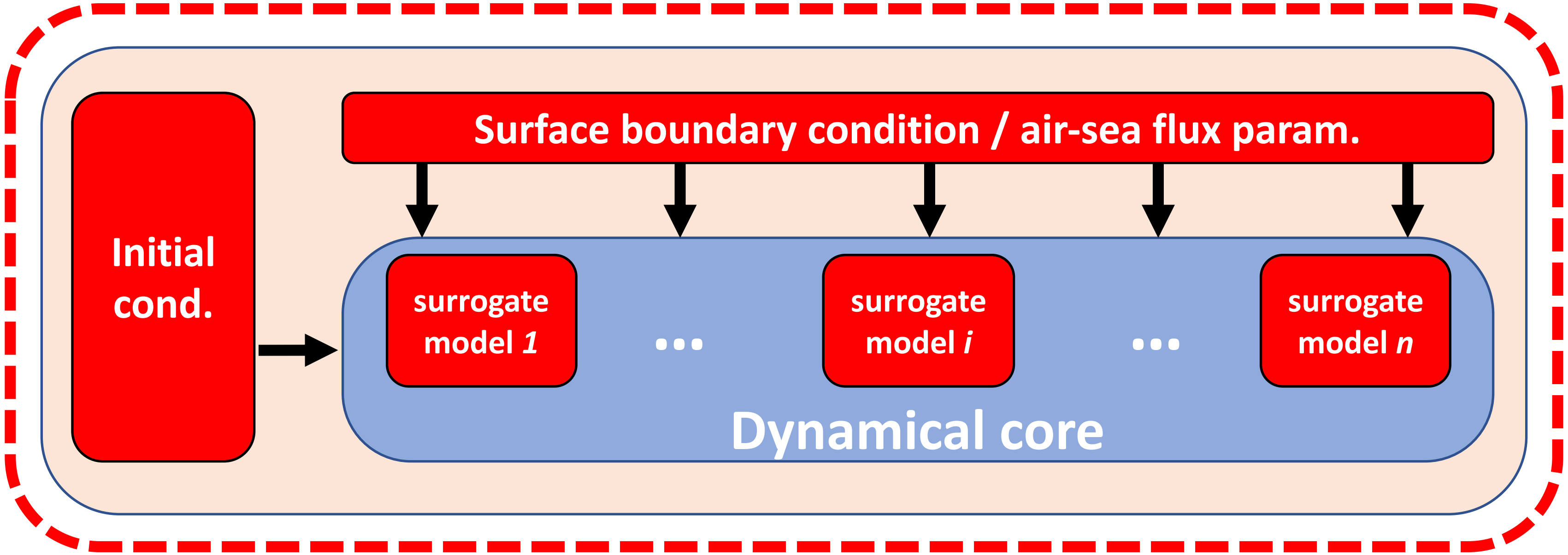
Cyberinfrastructure for Sustained Scientific Innovation (CSSI)

https://DJ4Earth.github.io

NSF CSSI: ***DJ4Earth***

Convergence of Bayesian inverse methods and scientific machine learning through universal differentiable programming

An end-to-end adjoint enables full-model calibration & initialization



Here: use of full-model ***differentiable programming*** to

- replace parts of model by appropriate surrogates
- use all available observations to train/calibrate all uncertain variables
- combines inverse modeling and ML in ***end-to-end learning***

relies on general-purpose automatic differentiation (AD)

Since 2023 the idea of differentiable programming 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 programming has taken off …

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

Geoscientific
Model Development
Open Access
EGU

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,

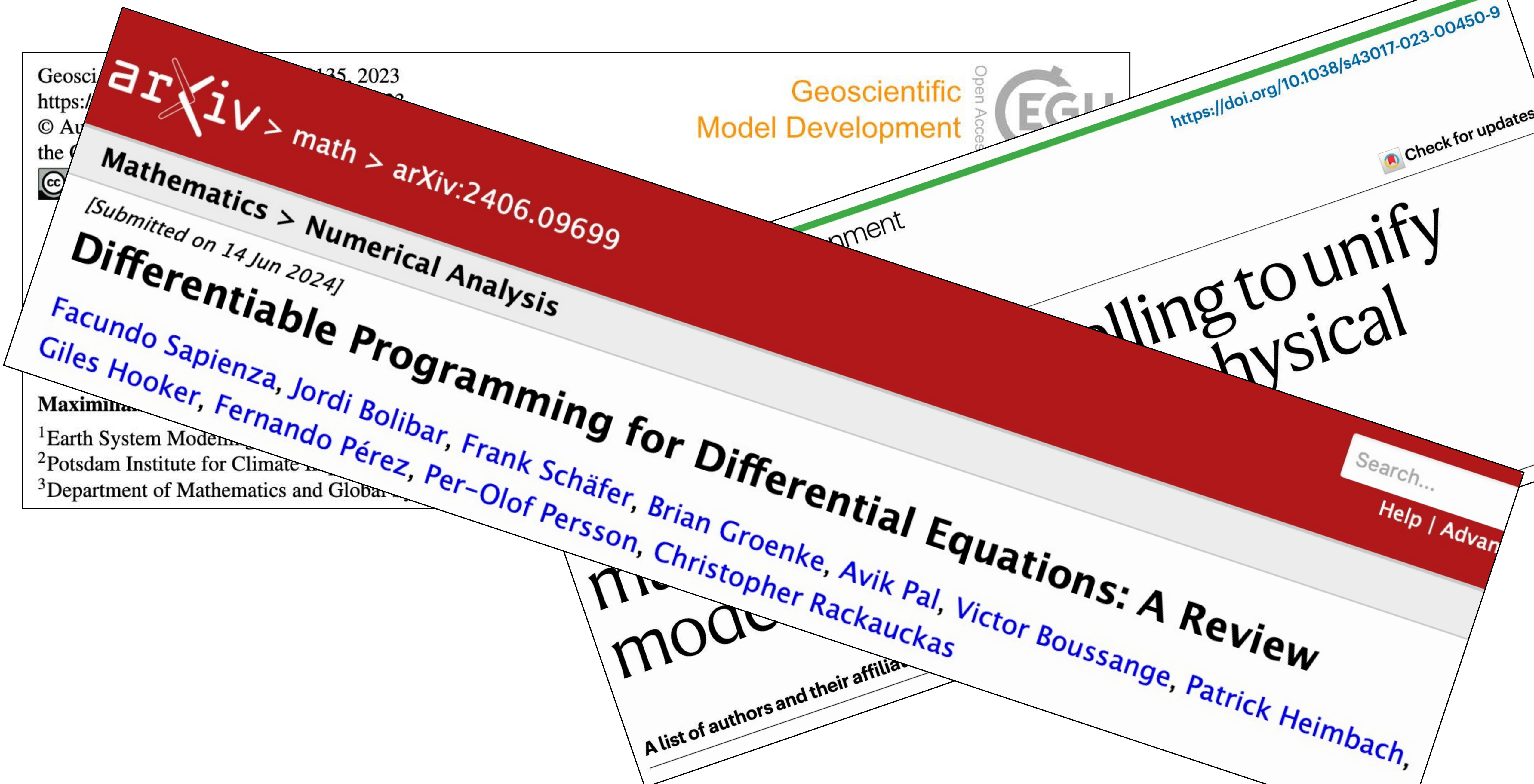
https://doi.org/10.1038/s43017-023-00450-9
Check for updates
nature reviews earth & environment

Perspective

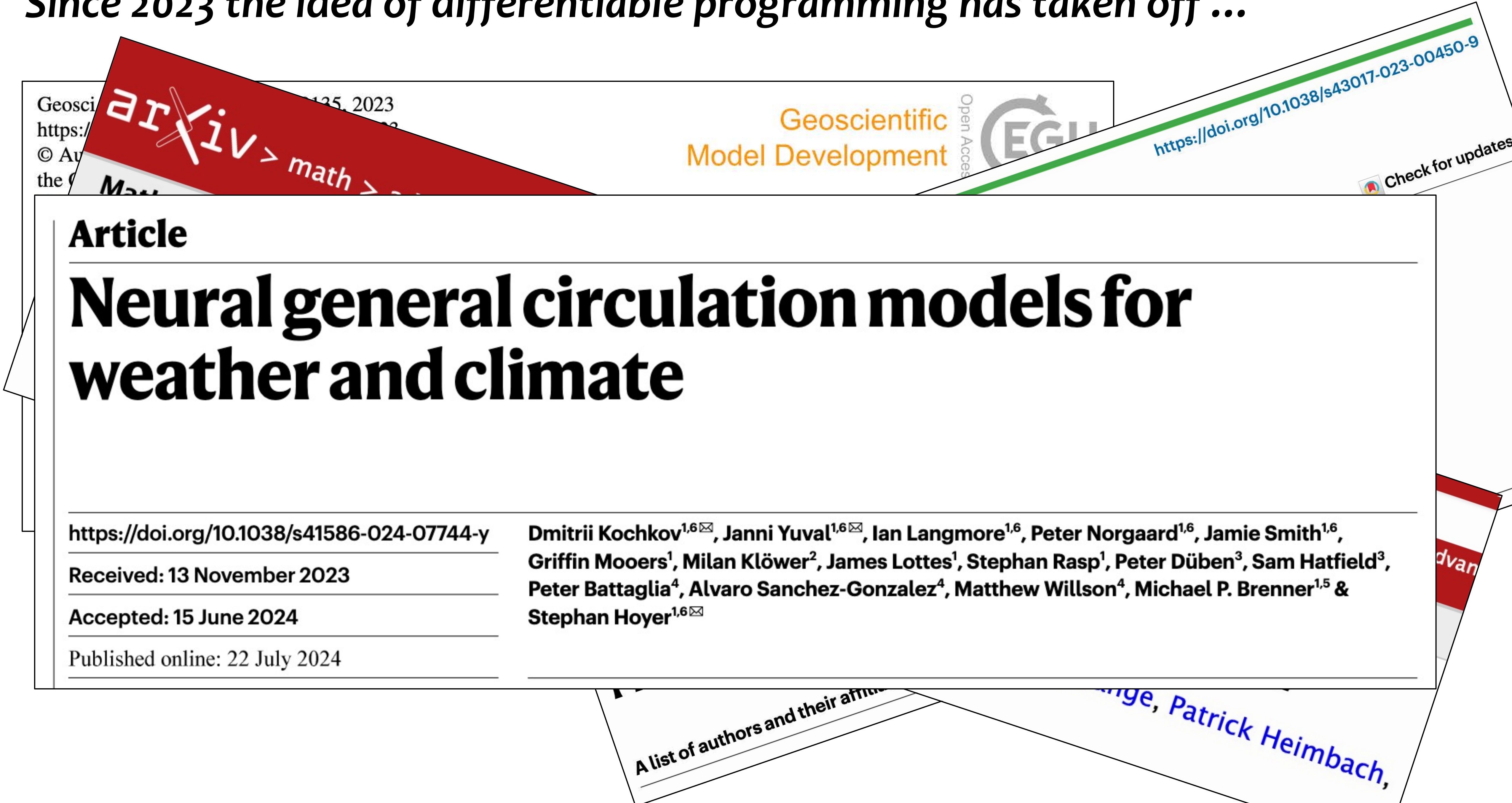
Differentiable modelling to unify machine learning and physical models for geosciences

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

Since 2023 the idea of differentiable programming has taken off …



Since 2023 the idea of differentiable programming has taken off ...

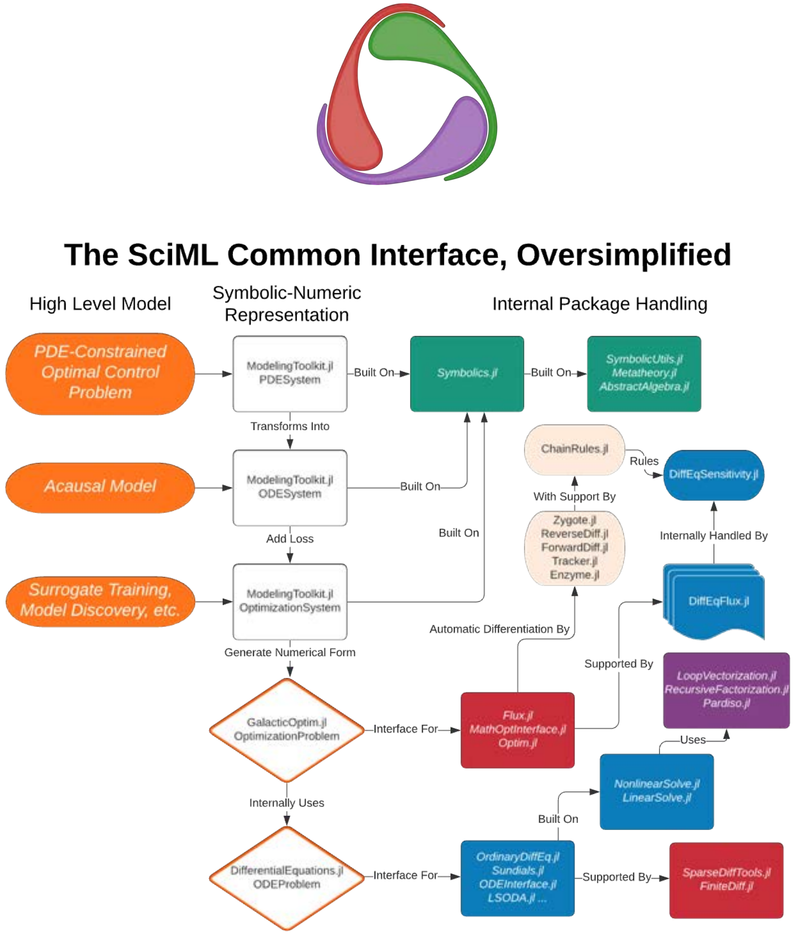


Why Julia?

DJ4Earth

1/ Building on *Climate Modeling Alliance (CliMA)*

2/ Serious efforts in AD, *differentiable programming*



3/ 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*

ClimaOcean.jl:

Ocean model component of the *Climate Model Alliance (CliMA) model*

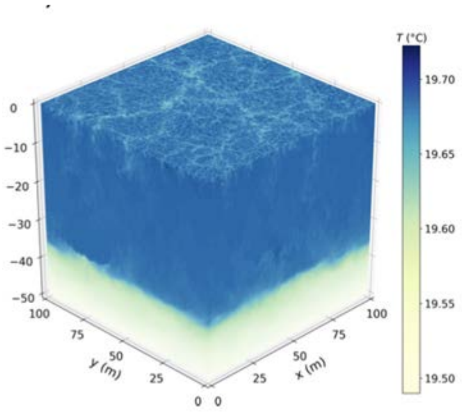
The Journal of Open Source Software

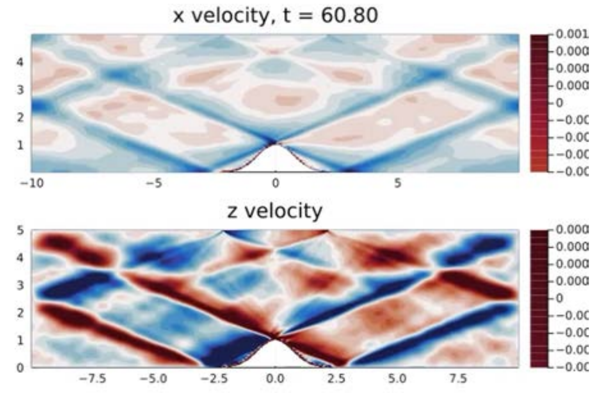
Oceananigans.jl: Fast and friendly geophysical fluid dynamics on GPUs

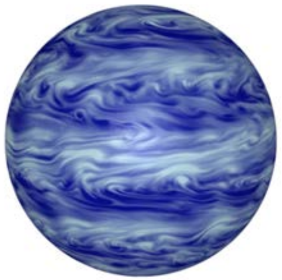
Ali Ramadhan[1], Gregory LeClaire Wagner[1], Chris Hill[1], Jean-Michel Campin[1], Valentin Churavy[1], Tim Besard[2], Andre Souza[1], Alan Edelman[1], Raffaele Ferrari[1], and John Marshall[1]

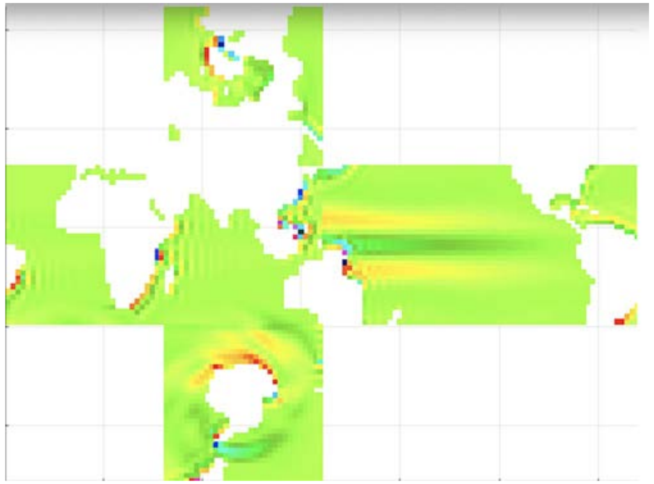
1 Massachusetts Institute of Technology **2** Julia Computing, Inc.

- Finite volume, rotating, stratified fluids model for geophysical fluid dynamics (GFD).
- Written from scratch in Julia
- Multiple simulation options.
- GPU and CPU via kernel abstractions
- Parallelize using MPI.jl and multi-threading





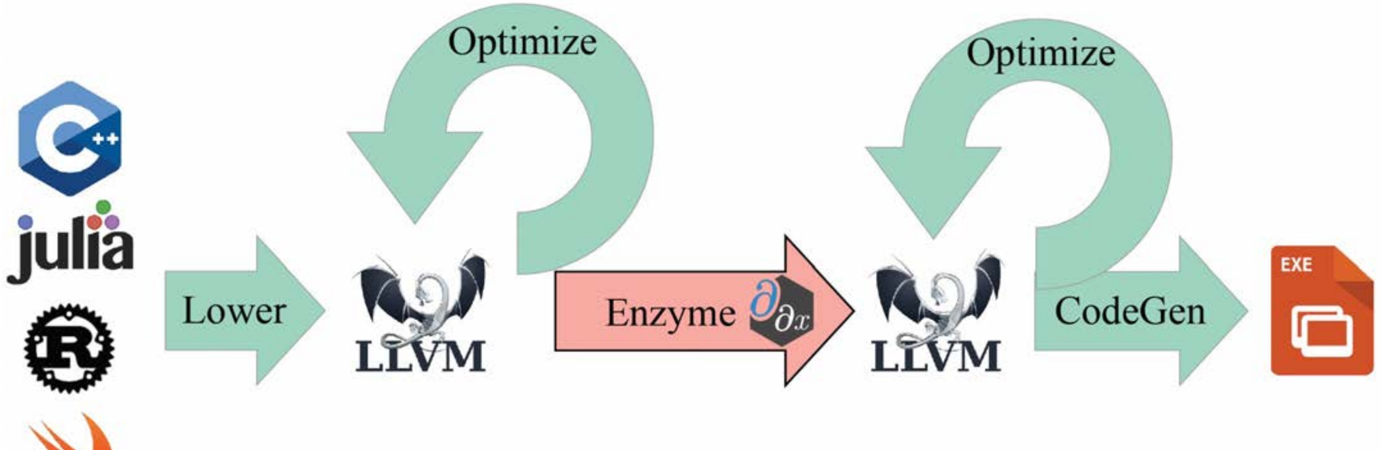




https://github.com/clima/Oceananigans.jl

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

Differentiating GPU-enabled ocean model
in Julia via the AD tool ***Enzyme.jl***

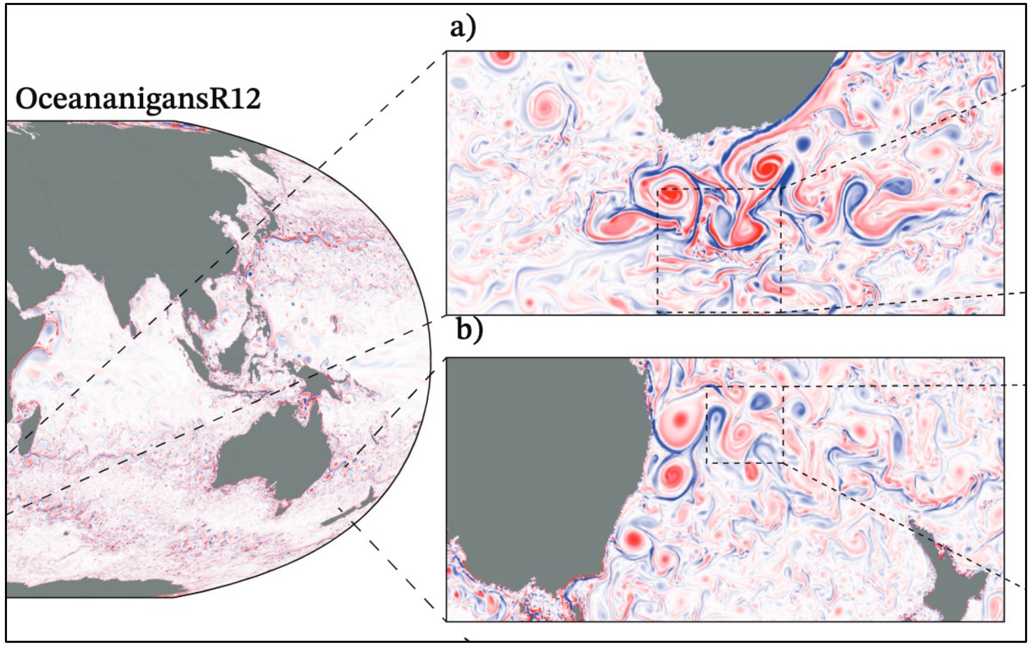


Moses, Churavy, et al., SC'21

CliMA
CLIMATE MODELING ALLIANCE

Oceananigans.jl

(Silvestri et al., arXiv, 2023, 2024)



DJ4Earth

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


W. Moses


V. Churavy


M. Schanen


S H K Narayanan

Three initial Earth system applications

DJ4Earth

Ocean

S. Williamson

J. Kump

N. Loose

Sea ice

S. Silvestri

G. Wagner

C. Hill

Ice sheets

M. Morlighem

C. Gong

- Bringing together concepts from …
  - … big data science & sparse data science
  - … computer science & computational science
  - … scientific machine learning & simulation-based science
- Sensitivity/gradient information is a powerful ingredient; obtained via
  - **differentiable programming / simulators**
  - **general-purpose automatic differentiation (AD)**

Minitutorials during *SIAM Mathematics for Planet Earth* 2024

https://github.com/DJ4Earth/MPE24

SIAM Conference on Mathematics of Planet Earth (MPE24)

Minitutorials

Differentiable Earth System Models in Julia

Joseph L. Kump, University of Texas, U.S.
Sarah M. Williamson, University of Texas, U.S.
Gong Cheng, Dartmouth College, U.S.

Differentiable Programming in Julia with Enzyme

Valentin Churavy, Massachusetts Institute of Technology, U.S.
William Moses, University of Illinois Urbana-Champaign, U.S.
Michel Schanen, Argonne National Laboratory, U.S.