PDAF Parallel Data Assimilation Framework

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

Ensemble Data Assimilation for Coupled Modeling in the Earth System

Lars Nerger

Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research
Computing Center & Section Marine Biogeosciences

Acknowledgements: Qi Tang, Nabir Mamnun,
Christoph Völker, Helge Gössling, Tido Semmler,
Yumeng Chen, Changliang Shao

University of Hohenheim, May 16, 2025

HELMHOLTZ
SPITZENFORSCHUNG FÜR
GROSSE HERAUSFORDERUNGEN

Overview

Data assimilation

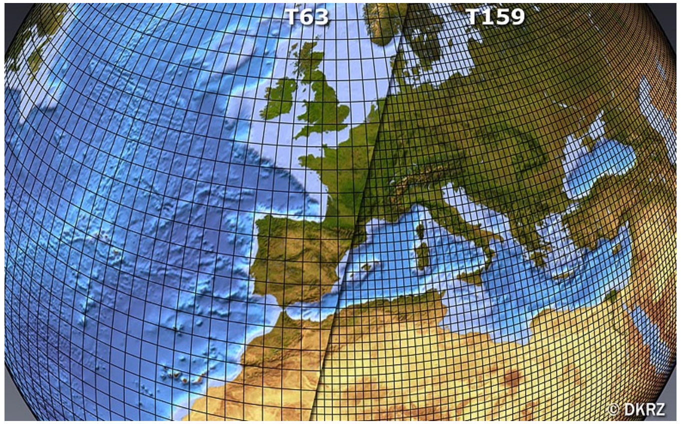
- Assimilation software

Application cases:

- Coupled ocean-atmosphere assimilation
- Estimation of biogeochemical process parameters

Data Assimilation – Combining Models and Observations

Models



+

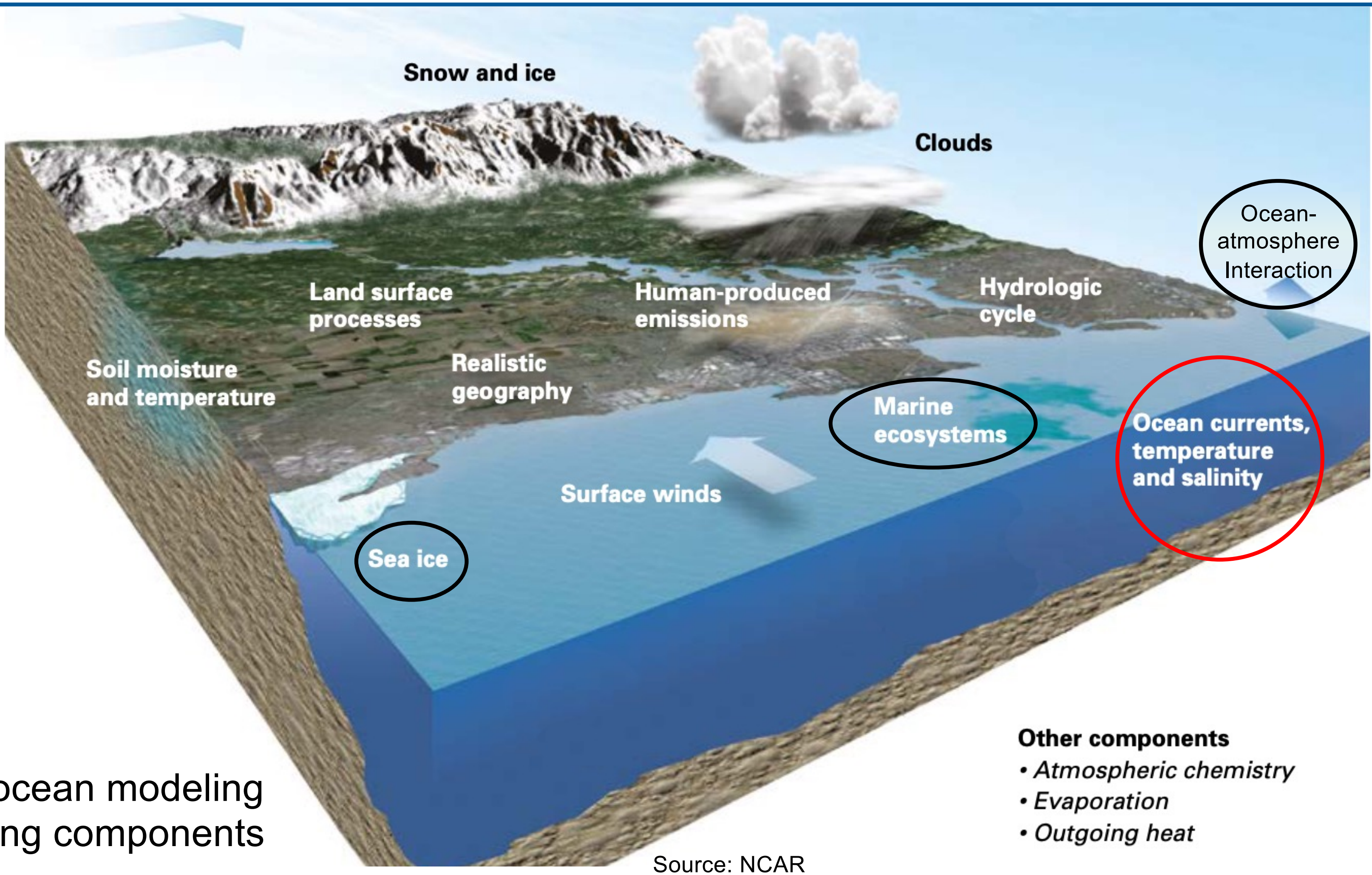
Observations

Combine both sources of information

quantitatively and optimally by computer algorithm

→ Data Assimilation

Coupled Assimilation - Ocean-Centric Perspective



Source: NCAR

Focus on ocean modeling
+ interacting components

Data Assimilation - Possibilities

"Observation-constrained (ensemble) modeling"

Aims

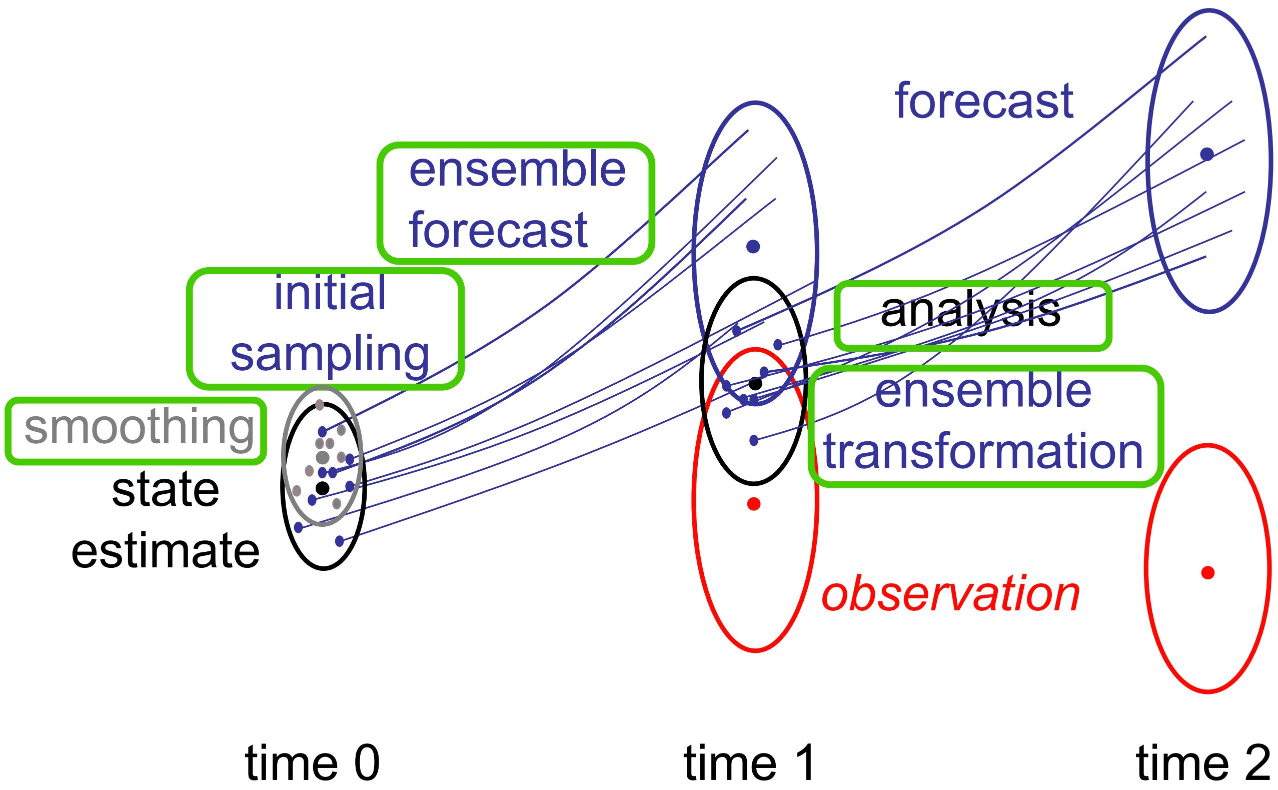
1. Optimal estimation of some modelled system:
   - initial conditions (for weather/ocean forecasts, …)
   - state trajectory (temperature, concentrations, …)
   - parameters (ice strength, plankton growth, …)
   - fluxes (heat, primary production, …)
   - boundary conditions and 'forcing' (wind stress, …)

2. More advanced: Improvement of model formulation and observations
   - detect systematic errors (bias)
   - revise parameterizations based on parameter estimates
   - detect relevant observations for observation system design

Ensemble Data Assimilation

Ensemble Kalman filters (Evensen (1994), see Vetra-Carvalho et al. (2018)…

Particle filters (see van Leeuwen et al. (2019), …)



Much research into how to perform these operations

Most can be implemented in generic form

Available in our DA software PDAF

PDAF – Parallel Data Assimilation Framework

PDAF Parallel Data Assimilation Framework

A unified tool for interdisciplinary data assimilation …

- provide support for parallel ensemble forecasts
- provide DA methods (EnKFs, smoothers, PFs, 3D-Var) - fully-implemented & parallelized
- provide tools for observation handling and for diagnostics
- easy implementation with (probably) any numerical model (<1 month)
- a program library (PDAF-core) plus additional functions & templates
- run from notebooks to supercomputers (Fortran, MPI & OpenMP)
- ensure separation of concerns (model – DA method – observations – covariances)
- first release in year 2004; continuous further development

Focus on

- Easy implementation
- Performance for complex models
- Flexibility to extend system

Open source:
Code, documentation, and tutorial available at
https://pdaf.awi.de

github.com/PDAF/PDAF

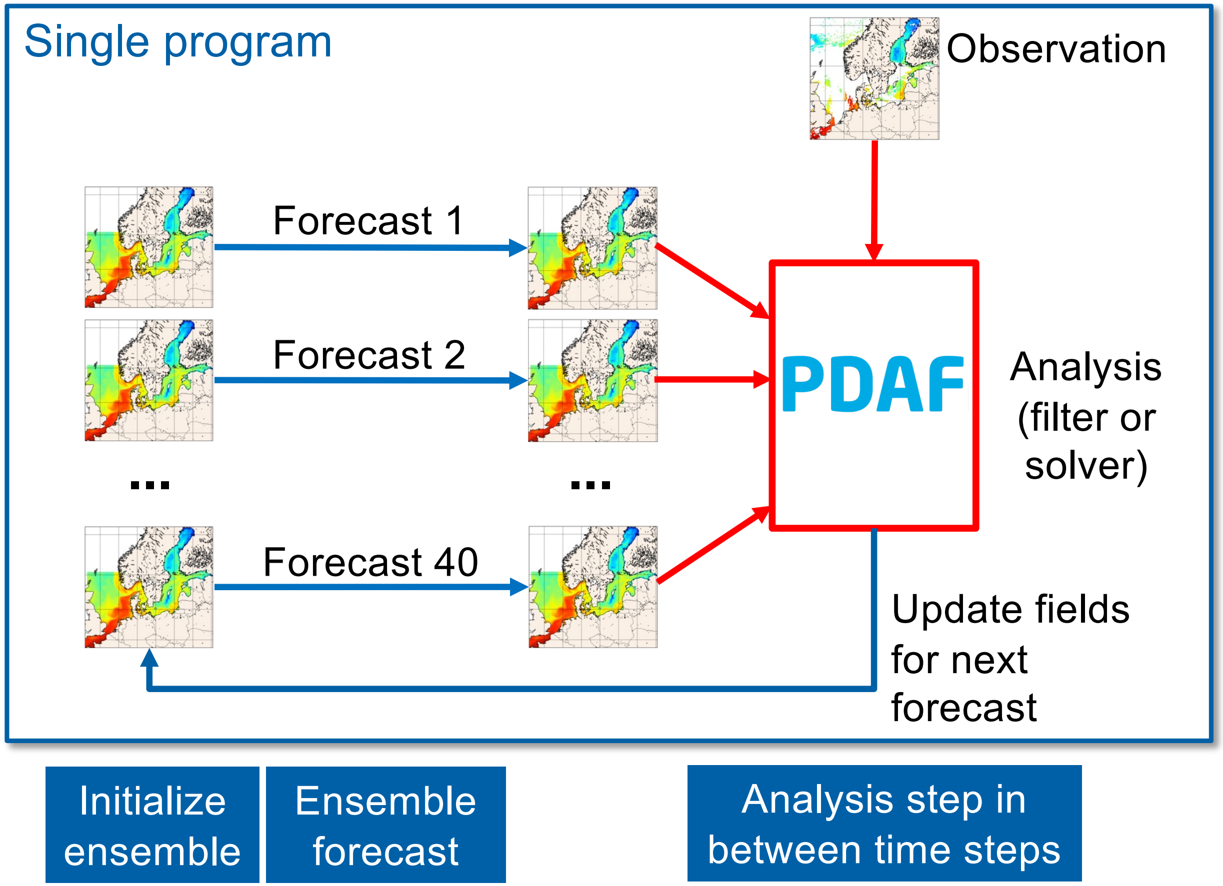
L. Nerger, W. Hiller, Computers & Geosciences 55 (2013) 110-118

Online-Coupling – Assimilation-enabled Model

Couple a model with PDAF

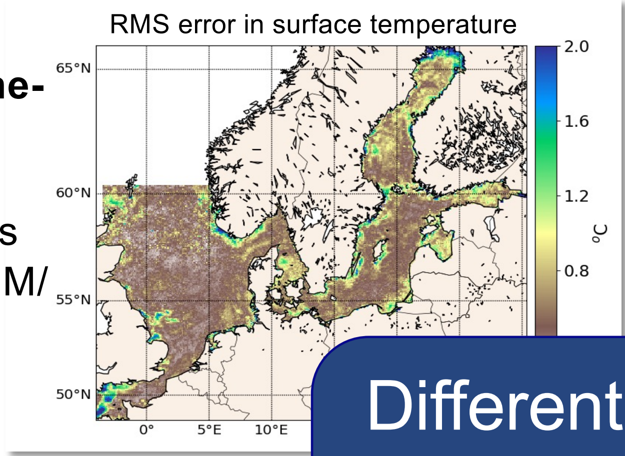
- Modify model to simulate ensemble of model states
- Insert analysis step/solver to be executed at prescribed interval
- Run model as usual, but with more processors and additional options

- EnOI and 3D-Var also possible:
  - Evolve single model state
  - Prescribe ensemble perturbations or covariance

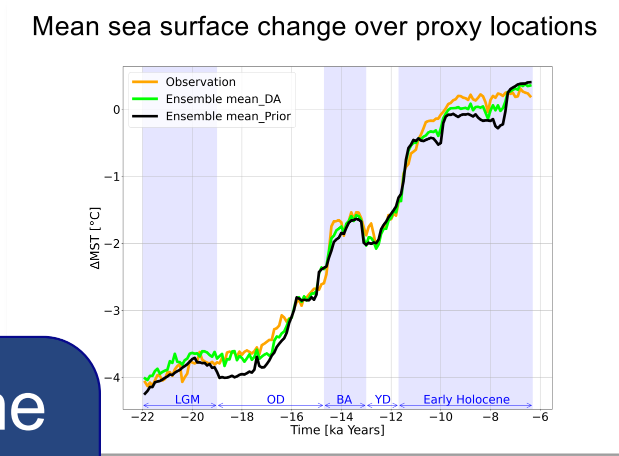


PDAF Application Examples – at AWI

Coastal coupled physics/biogeochemistry DA: CMEMS/BSH - Improving forecasts with NEMO-ERGOM/ HBM-ERGOM: *(S. Vliegen, A. Sathanarayanan)*



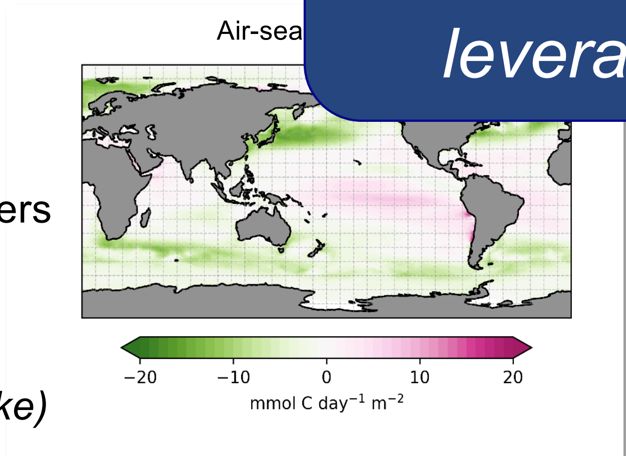
Paleo-climate DA: improve simulation of last deglaciation with CLIMBER-X *(A. Masoum)*



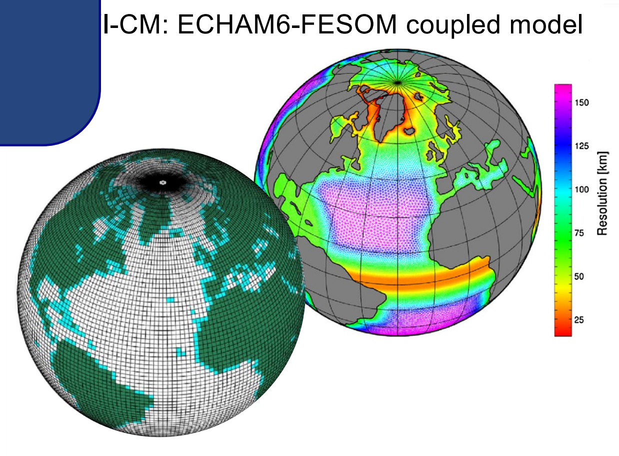
Different models – same assimilation software

leverage synergies

Coupled physics/biogeochemistry DA: Improving parameters & carbon flux in REcoM *(N. Mamnun, F. Bunsen, A. Broschke)*



atmosphere DA: Assimilate ocean observations into atmosphere with AWI-CM *(Q. Tang)*



PDAF Application Examples - External

PDAF Parallel Data Assimilation Framework

External applications & users, like

Operational uses:

- *Germany:* North/Baltic Seas (HBM model)
- *Europe:* Copernicus marine forecasting center Baltic Sea (NEMO)
- *China:* Arctic ice-ocean prediction system (MITgcm)

Ocean, biogeochemistry & climate models (research applications)

- **SCHISM/ESMF**
- **CICE** (sea ice)
- **MEDUSA** (biogeochemistry)
- **MPI-ESM** (ICON-O)
- **COAWST** (WRF-ROMS-CICE)
- **NEMO**
- **MITgcm**

Beyond ocean

- **TSMP** (Terrestrial Systems Modeling Platform)
- **WRF** (Weather forecast and research model)
- **TIE-GCM** (Thermosphere Ionosphere Electrodynamics GCM)
- **WRF** (Wether Research and Forecast Model)
- **VILMA** (Viscoelastic Lithosphere and Mantle Model)
- **Parody** (Geodynamo model)
- **HYSPLIT** (Volcanic Ash Transport and Dispersion model)
- **HydroGeosphere** (hydrology)
- **Cardiatic modeling** (blood flow)
- … more

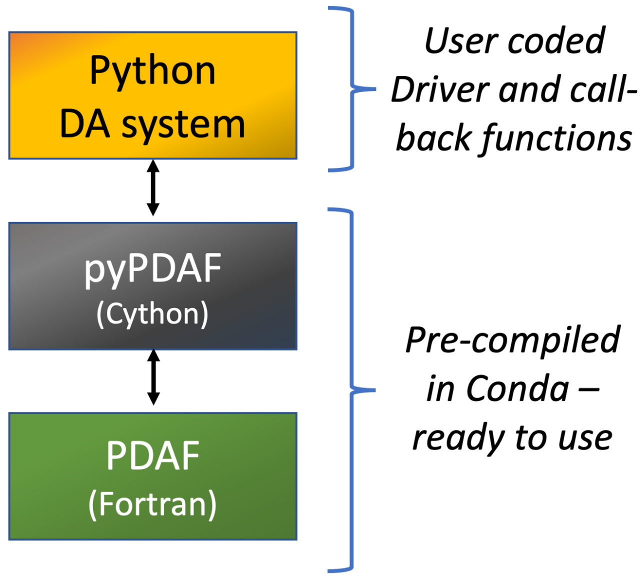
Different models – same assimilation software

... more synergies

pyPDAF

Python interface to PDAF

- allows to code all application-specific functionality in Python (not touching Fortran!)
- assimilation analysis computed inside PDAF (excellent performance due to compiled Fortran code)
- supports all functionality of PDAF including MPI-parallelization
    - online coupling (e.g. for Python-coded models)
    - offline coupling (using files from model runs)
- demonstrated low overhead for localized ensemble filters
- installation using Conda



github.com/yumench/pyPDAF

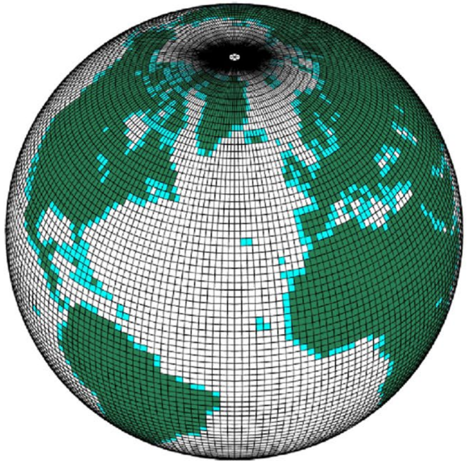
Y. Chen et al. (2024). Geoscientific Model Development, submitted

Application I

Coupled atmosphere-ocean data assimilation

Assimilation into coupled model: AWI-CM1

Atmosphere

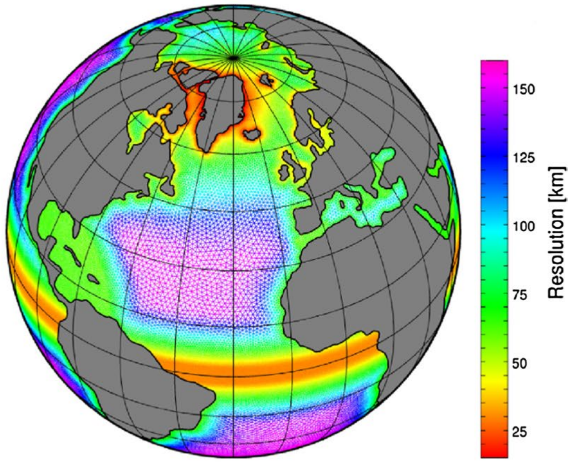


OASIS3-MCT

fluxes

ocean/ice state

Ocean



Alternative:
OpenIFS
(AWI-CM3)

Atmosphere
- ECHAM6
- JSBACH land

Coupler library
OASIS3-MCT

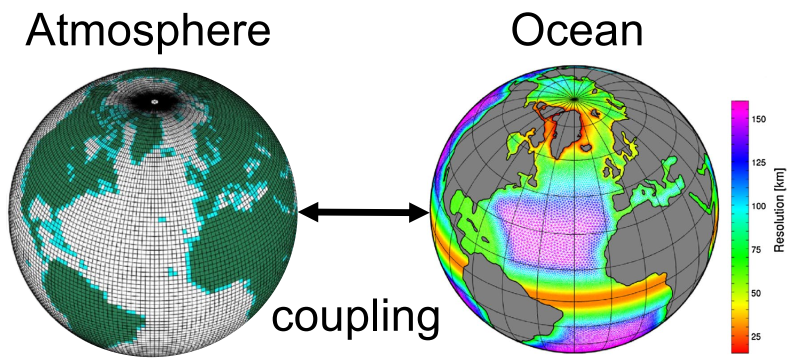
Ocean
- FESOM
- includes sea ice

Two separate programs for atmosphere and ocean

Goal: Develop data assimilation methodology for cross-component assimilation ("strongly-coupled")
"assimilate ocean observations into the atmosphere"

Nerger et al. (2020). Geoscientific Model Development, 13, 4305–4321, doi:10.5194/gmd-13-4305-2020

Coupled Models and Coupled Data Assimilation



Coupled models

- Several interconnected components, like
  - Atmosphere and ocean
  - Ocean physics and biogeochemistry (carbon, plankton, etc.)
  - Atmosphere and land surface

Coupled data assimilation

- Assimilation into coupled models
  - **Weakly coupled**: separate assimilation in the components
  - **Strongly coupled**: joint assimilation of the components
    - Use cross-covariances between fields in components
    - Expected to ensure consistent state across components

Challenges:

- Distinct spatial and temporal scales
- nonlinearity

Data Assimilation Experiments

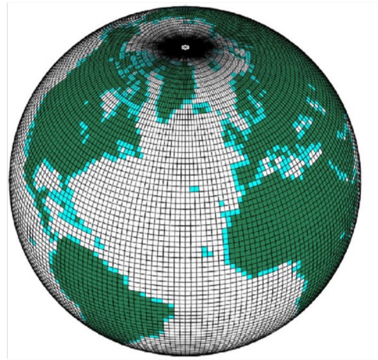
Model setup

- Global model
- ECHAM6: T63L47 , ~1.8^{o} horizontal resolution
- FESOM: resolution 30-160km

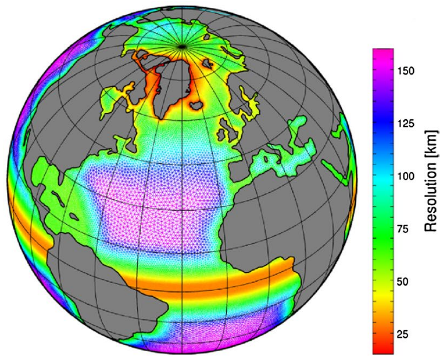
Data assimilation experiments

- Observations: satellite sea surface temperature
- Simulation period: year 2016, daily assimilation update
- Updated variables:
  - weakly-coupled: ocean
  - strongly-coupled: ocean + atmosphere
- ensemble filter LESTKF (Nerger et al., 2012) in PDAF
- Ensemble size: 46
- Computing time: 3.5h
  - fully parallelized using ~13,000 processor cores

ECHAM mesh

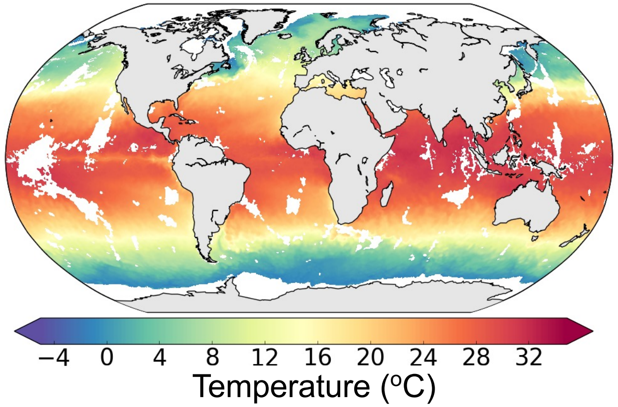


FESOM mesh resolution

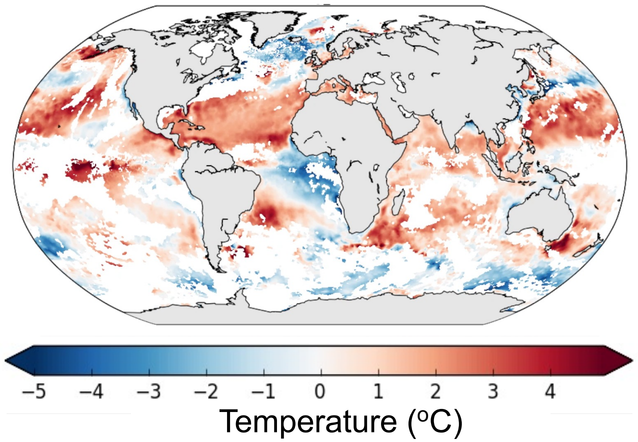


Assimilate sea surface temperature (SST)

Satellite SST on January 1, 2016



SST difference: observations-model



- Satellite sea surface temperature (multi-satellite, level 3, EU Copernicus)
- Daily data, data gaps due to clouds
- Observation error: 0.8 °C
- Localization radius: 1000 km

→ no atmospheric data assimilated (by intention)

Large initial SST deviation due to using a coupled model: up to 10°C

Assimilation with such a coupled model is unstable!

omit SST observations where

$|SST_{obs} - SST_{ens\_mean}| > 1.6\ ^{o}C$

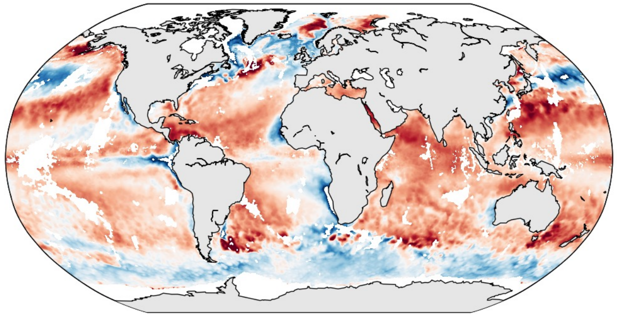
(30% initially, <5% later)

Tang et al. (2020) Q. J. Royal Metorol. Soc., 146, 4014-4029, doi:10.1002/qj.3885
Tang, et al., (2021) Geophys. Res. Lett, 48, e2021GL094941,

Assimilation of satellite SST: Effect on the ocean

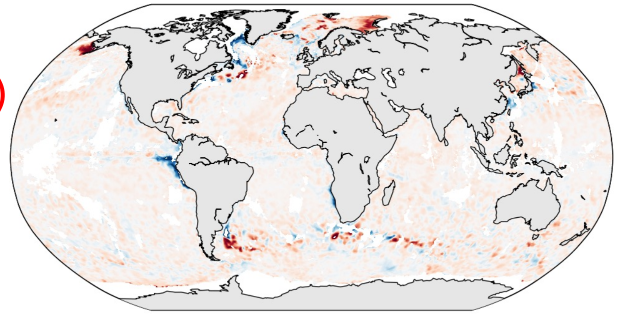
SST difference (obs-model): strong decrease of deviation

No Assimilation



4/30/2016
Day 120
(April 30)

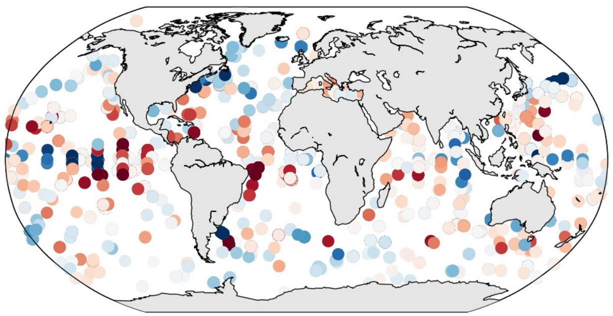
Assimilation



Necessary effect: dependent data

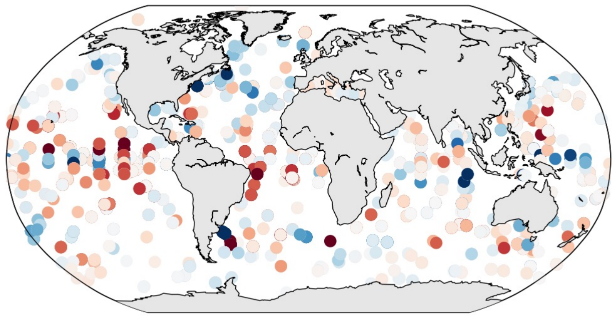
Subsurface temperature difference (obs-model); all model layers at profile locations

No Assimilation

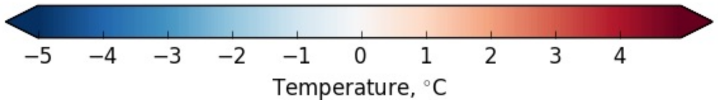


4/30/2016
Day 120
(April 30)

Assimilation



independent data

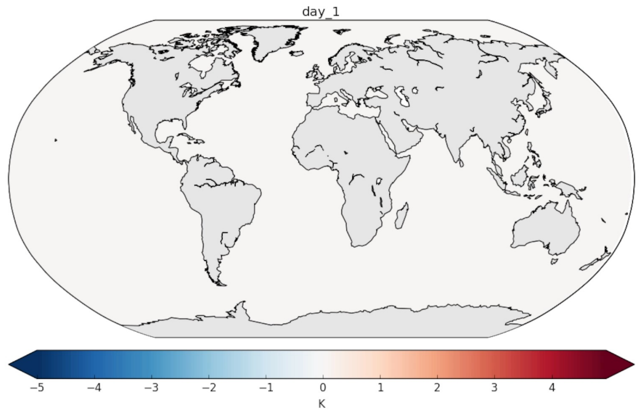


Weakly coupled DA: Effect on the Atmosphere

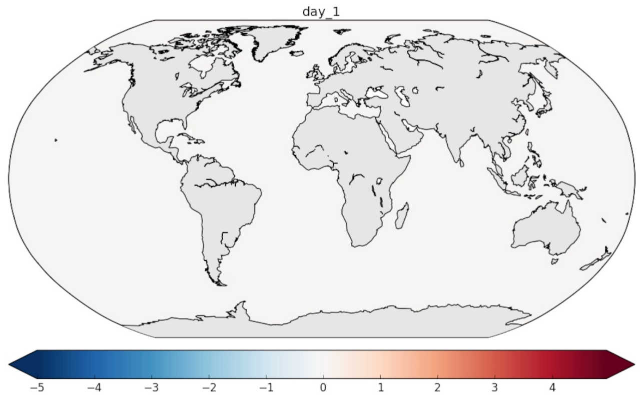
Difference between assimilation run and the free run

Weakly coupled: DA only changes ocean variables

Temperature at 2m



Sea surface temperature



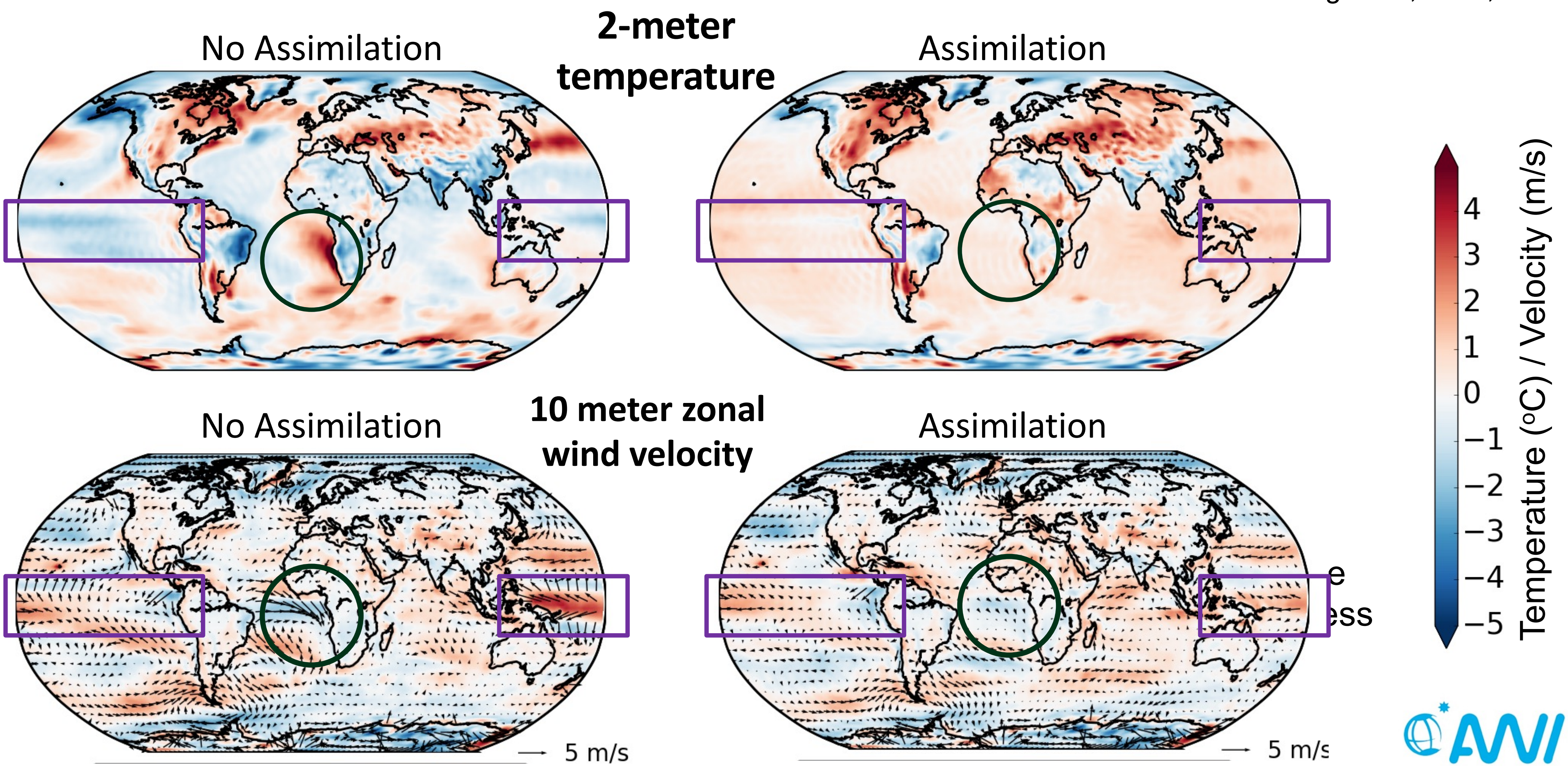
Atmosphere reacts quickly on the changed ocean state

Does it make the atmosphere more realistic?

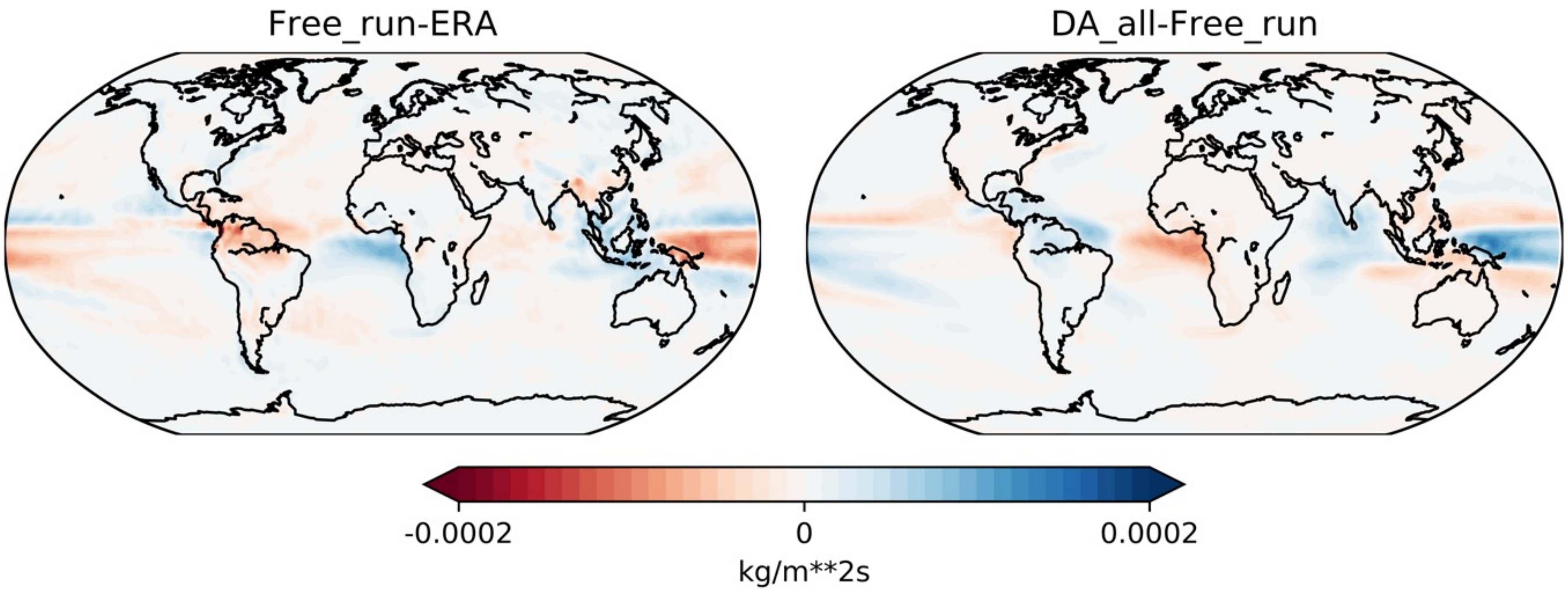
Effect on Atmospheric State

Difference Model – ERA-Interim Reanalysis (mean over 2016)

Tang et al., 2020, 2021



Weakly-coupled DA – Change in precipitation



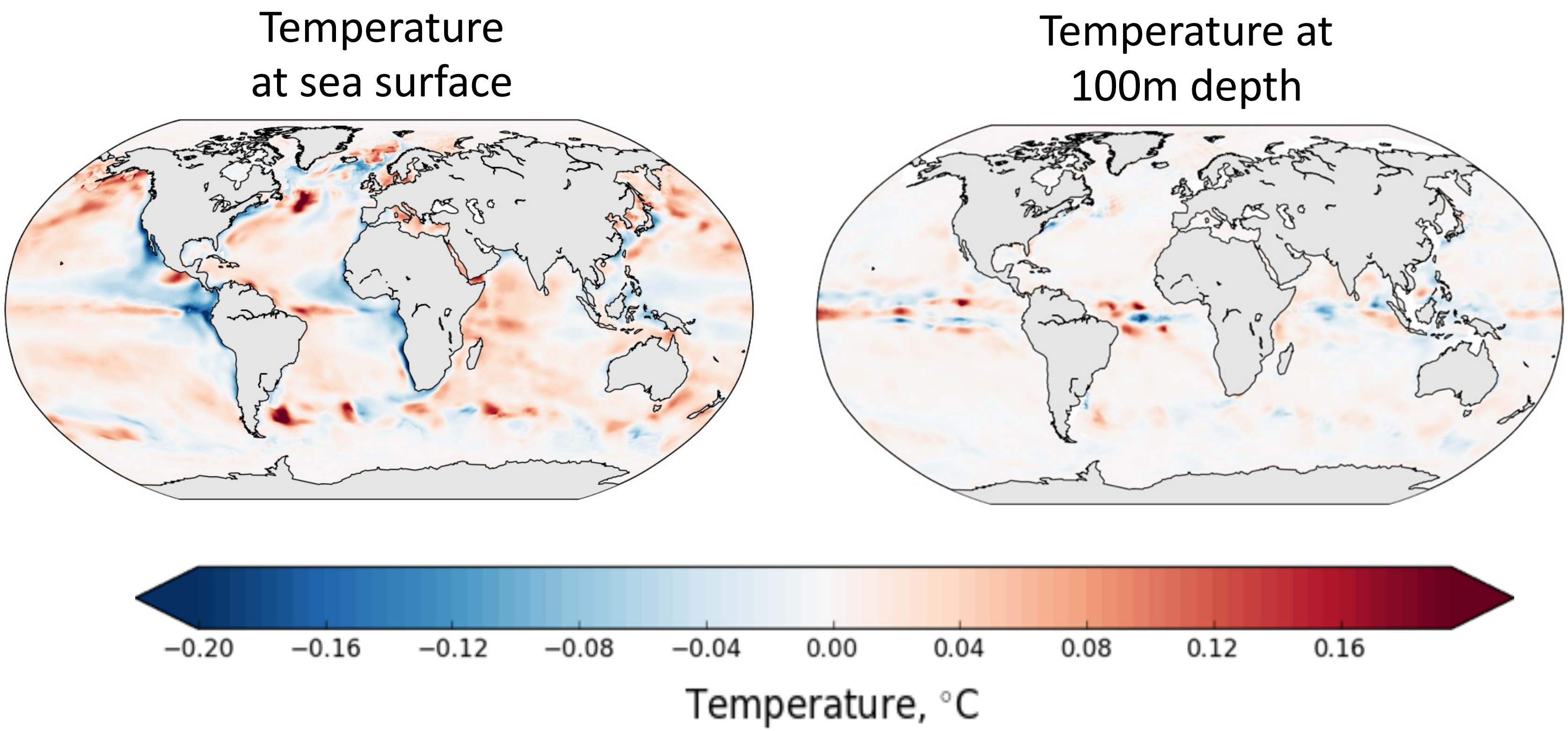
- Data assimilation removes most of the precipitation bias

Tang et al. (2020) Q. J. Royal Metorol. Soc., 146, 4014-4029, doi:10.1002/qj.3885
Tang, et al., (2021) Geophys. Res. Lett, 48, e2021GL094941,

Average assimilation increments

Average increments (analysis – forecast) for days 61-366 (after spinup)

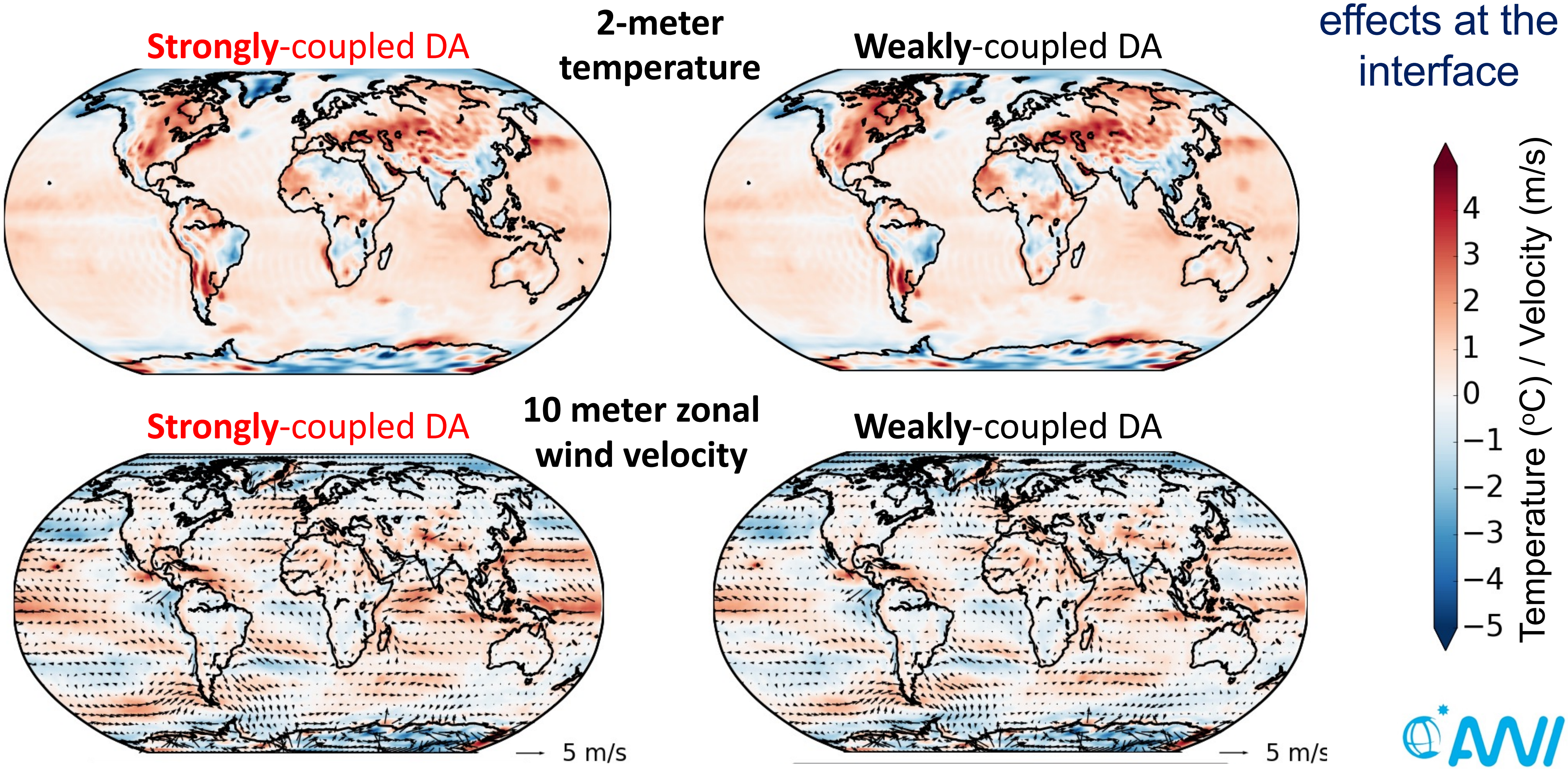
- non-zero values indicate regions with possible biases
- Equatorial region likely due to resolution



Strongly vs. weakly coupled DA

Difference Model – ERA-Interim Reanalysis (mean over 2016)

Very similar effects at the interface

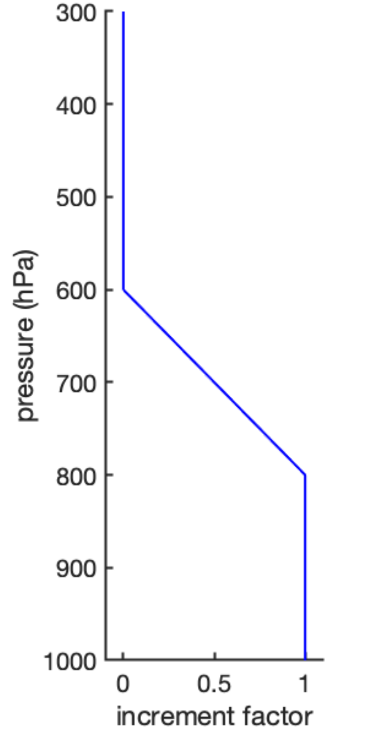


Strongly-coupled assimilation of SST observations

- Assimilate sea surface temperature into the atmosphere

- 2 cases:
  - **SCDA**: update all atmospheric layers
    → cross-covariances with SST used throughout the atmosphere

  - **SCDA_vloc**: apply vertical localization to assimilation increments
    - Full increments in lower atmosphere (sea surface – 800 hPa)
    - Linear decrease to zero from 800 – 600 hPa
    - No update higher than 600 hPa
      → No direct effect on assimilation in boundary layer
      → Dynamics in free atmophere influence lower atmosphere

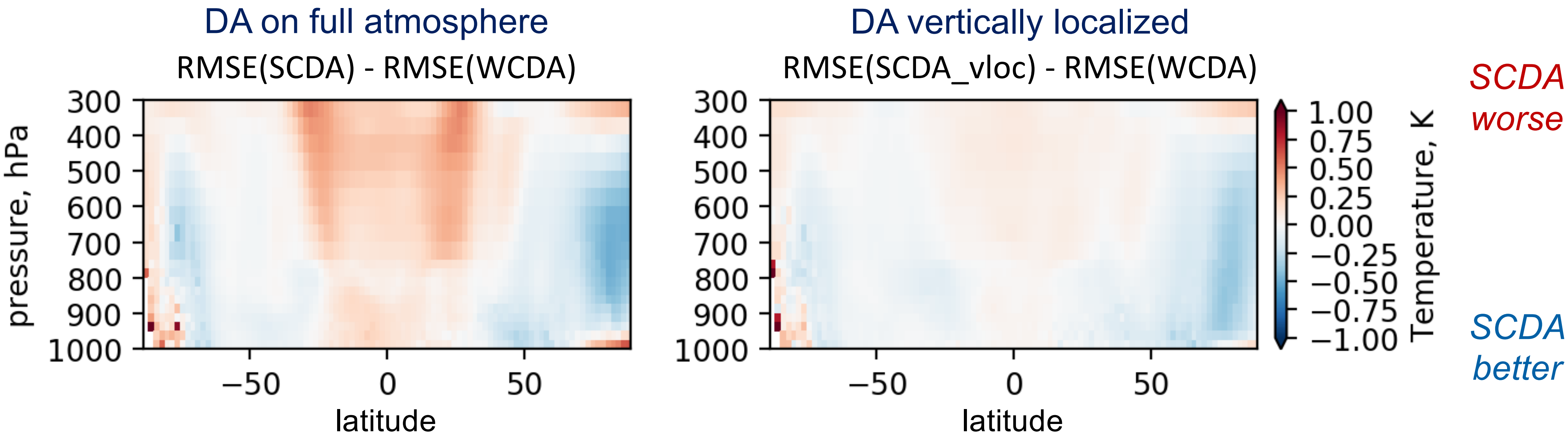


Factor for vertical localization

Tang et al. (2021) GRL, 48, e2021GL094941

Strongly vs. weakly coupled DA

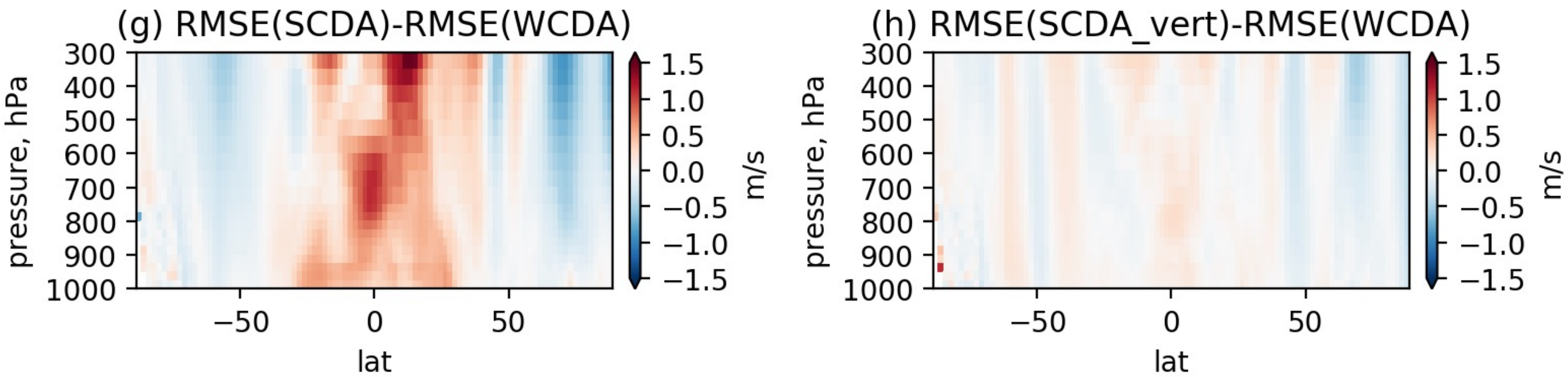
Atmospheric temperature – zonally averaged root-mean square errors compared to ERA-Interim



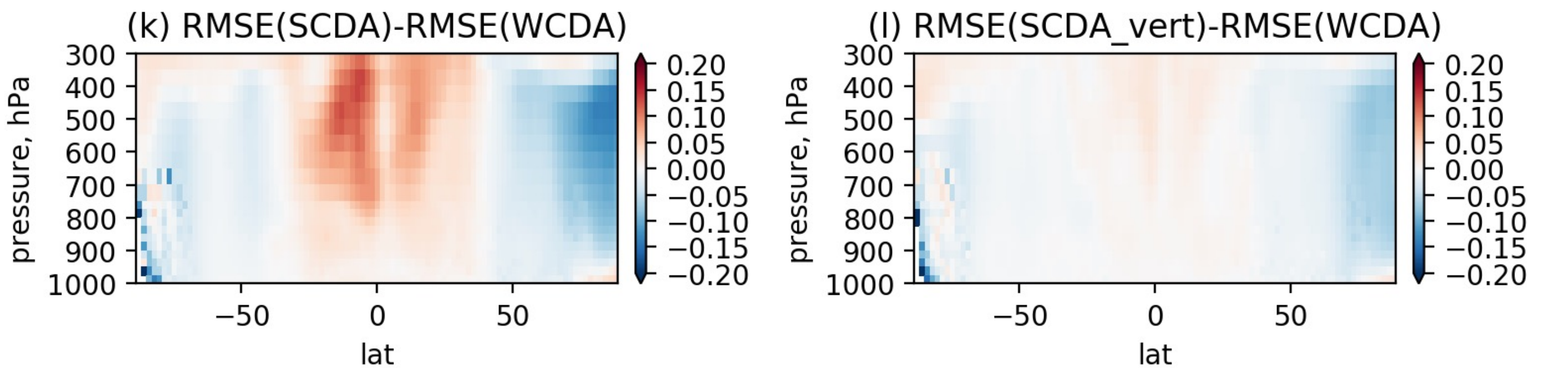
- SCDA in full atmosphere
  - deteriorates temperature in tropical region (due to too strong vertical correlations)
  - Improves high latitude regions (also in wind speed and specific humidity)
- Vertical localization (SCDA_vloc)
  - removes deterioration in tropical region
  - keeps (some) positive effect in high latitude regions

Weakly vs. Strongly coupled DA

- **Wind speed** – zonally averaged rms error



- **Specific humidity** (normalized by ERA-Interim – zonally averaged rms error



Weakly and strongly-coupled assimilation of SST observations

- Improvement of atmospheric state
- Fast reaction of atmosphere
- Strongly-coupled assimilation compared to weakly coupled
  - Slightly increased errors in tropical region
  - Improvements in higher latitudes
  - Utilizing vertical localization
    - Overall smaller effects, but still positive in high latitudes

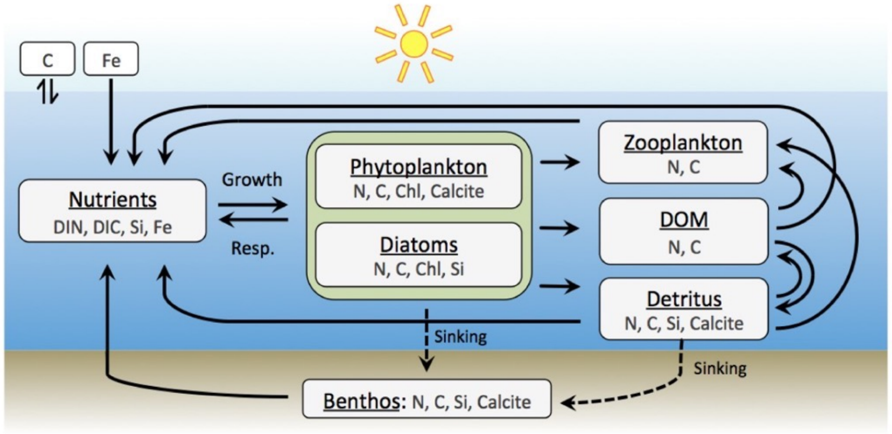
Tang et al. (2021) GRL, 48, e2021GL094941

Application II

Estimating biogeochemical process parameters

REcoM2 coupled to MITgcm

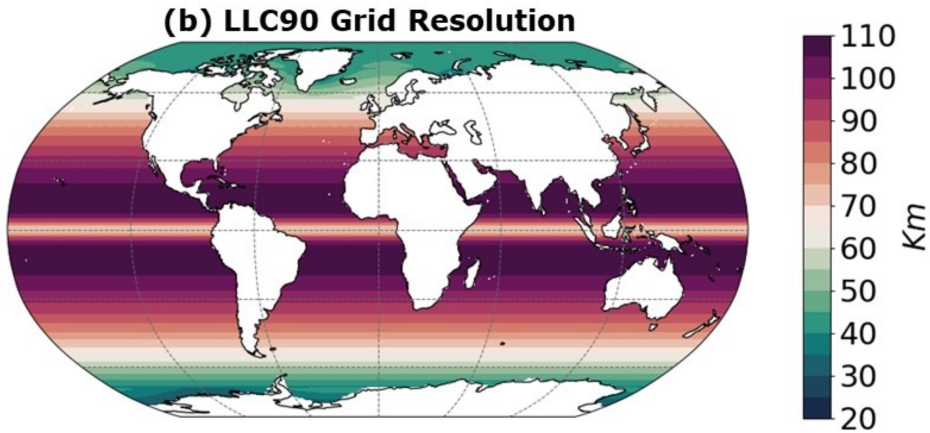
Ecosystem: REcoM



coupling

velocities
temperature

Physics: MITgcm



Forget et al. (2015)

- Biogeochemical process parameters:
  - 65 process parameters
  - constant over the globe
  - most values have significant uncertainties

Coupled Data Assimilation

Ensemble states

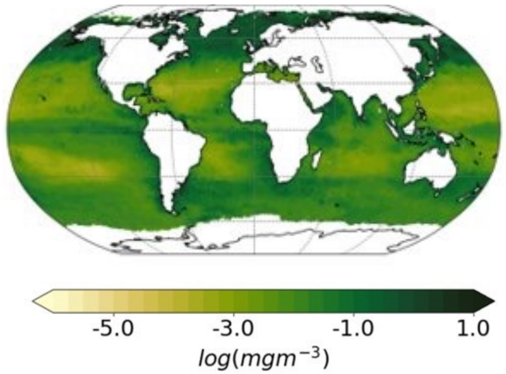
- 8 REcoM state variables describing phytoplankton
- 9 sensitive process parameters
  - Defined as 2 dimensional fields

Assimilation setup

- satellite observations:
  - 5-day chlorophyll concentration from ESA OCCCI
- simulation period: year 2020
- ensemble Kalman filter LESTKF in PDAF
- ensemble size: 40

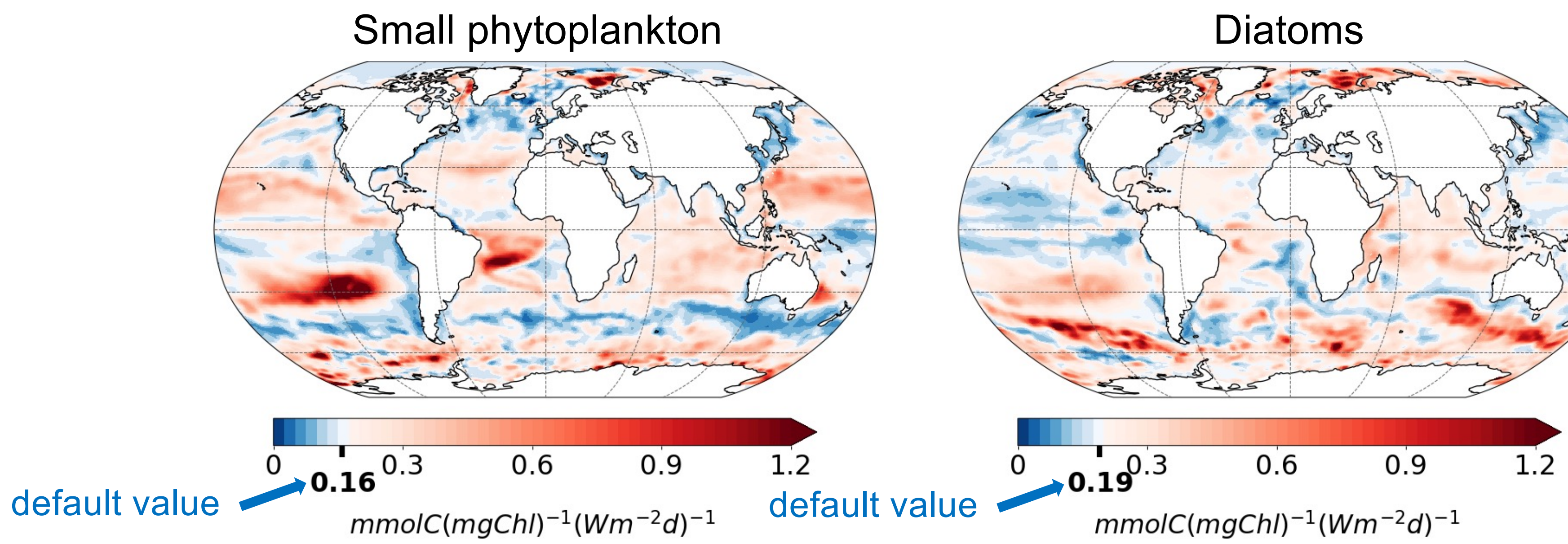
- **estimate 2-dimensional parameter fields**
- assimilation applied to log-transformed variables and parameters
- at initial time:
  - stochastic perturbation of process parameters for ensemble spread

Chlorophyll observations average for September



Estimated parameter: initial slope of photosynthesis curve (α)

α: Efficiency at which phytoplankton converts light energy into chemical energy



Estimation of spatially varying process parameter

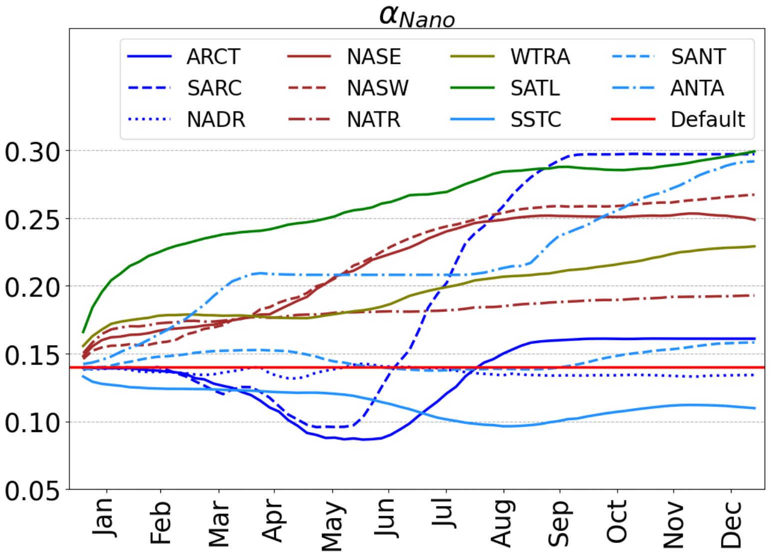
- Different distributions for both plankton groups
- Significant deviations from constant default values
- Generally consistent with observations (where they exist)

Mamnun et al. (2025) J. Geophys. Res., submitted, preprint: doi:10.22541/essoar.174594990.07411710/v1

Estimated parameter: initial slope of photosynthesis curve (α)

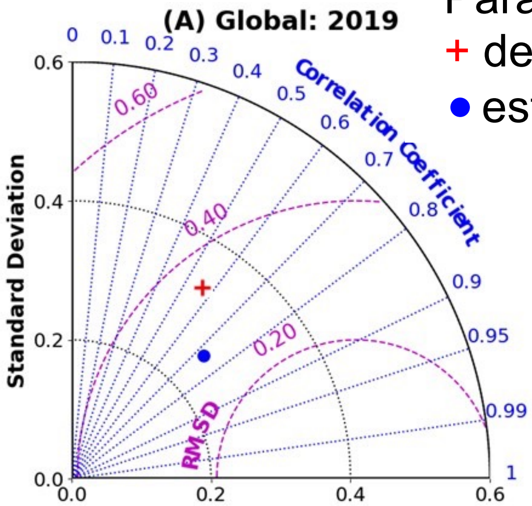
α: Efficiency at which phytoplankton converts light energy into chemical energy

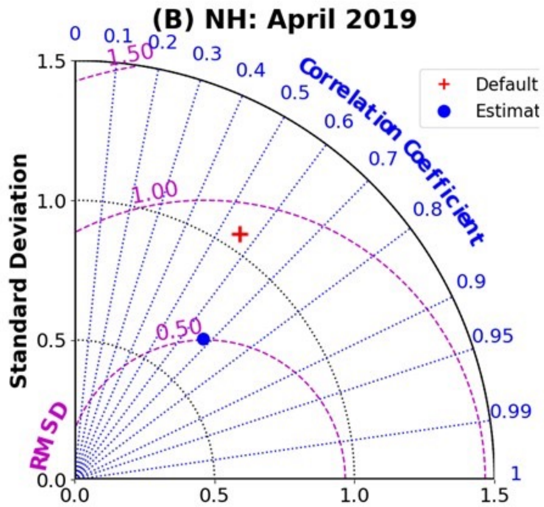
Small phytoplankton
Regional parameter value over time



- Estimates develop over time
- Values typically converge
  - But not always: variation over time possible relevant for some parameters

Model run with estimated parameters

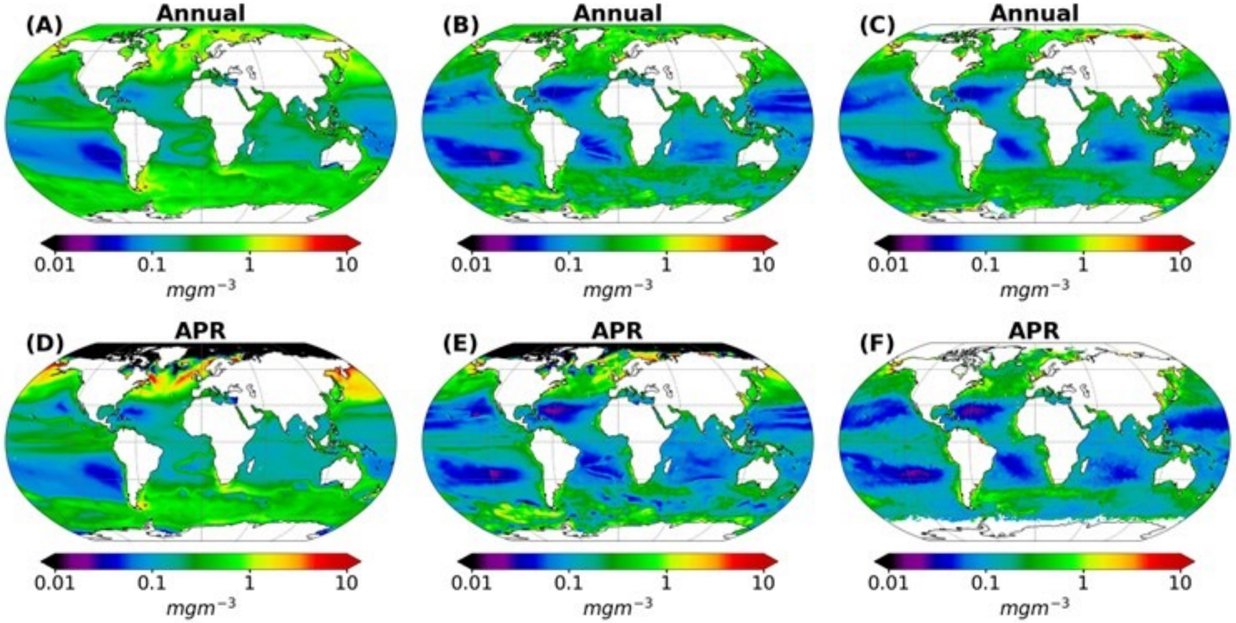




Parameters:
+ default
● estimated

- Comparison with satellite chlorophyll observations
- Over 5 years (2017 – 2021)
- 26% reduction of RMSE for annual average

Default parameters | optimized parameters | Observations



Summary

Ensemble Data Assimilation

- allows for ‘observation-constrained’ modeling and to let the model learn from observations
- rather easy to implement and to apply

Coupled data assimilation

- Joint update beyond single Earth system components
- Can improve overall consistency, but depends on time scales

Data assimilation is not just for forecasting!

- Useful applications in understanding model deficits and improving the models

Thank you!