# Ensemble Data Assimilation for Coupled Models of the Earth System

Lars Nerger, Qi Tang, Longjiang Mu, Mike Goodliff

Alfred Wegener Institute
Helmholtz Center for Polar and Marine Research
Bremerhaven, Germany

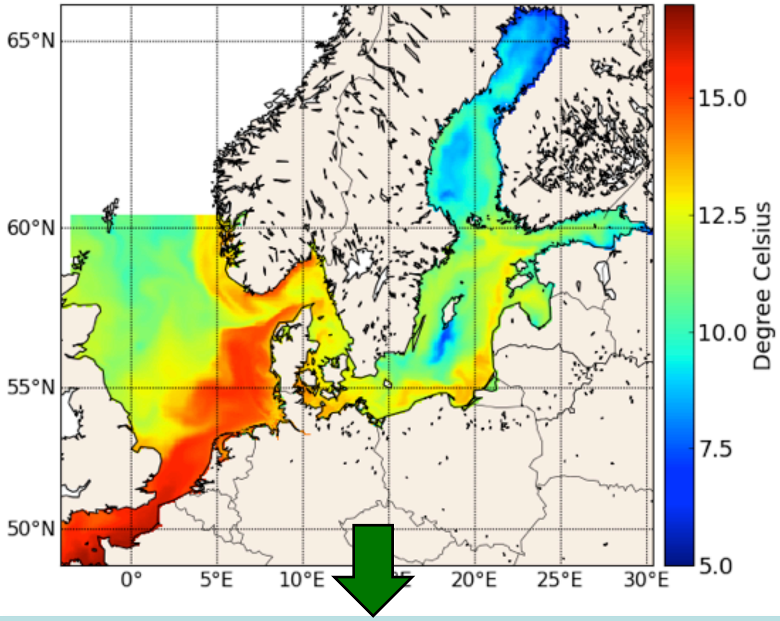
Sun Yat-sen University, Zhuhai, China, November 5, 2019

# Overview

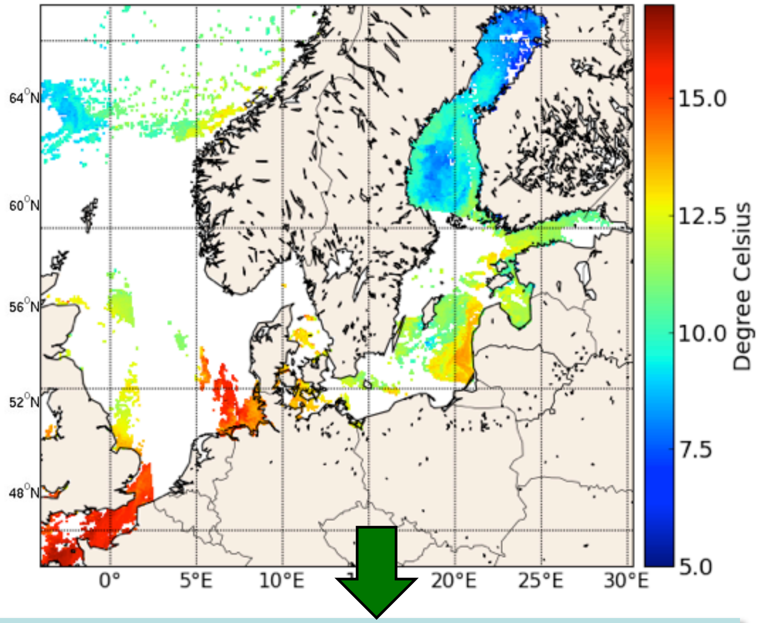
- Ensemble data assimilation
- Importance of software
- Coupled data assimilation
    - Challenges in two application examples

# Data assimilation

*Model* surface temperature




*Satellite* surface temperature




Combine both sources of information

quantitatively by computer algorithm

→ Data Assimilation

# Data Assimilation

Methodology to combine model with real data

- Optimal estimation of system state:
  - 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, …)
- More advanced: Improvement of model formulation
  - Detect systematic errors (bias)
  - Revise parameterizations based on parameter estimates

# Ensemble Data Assimilation

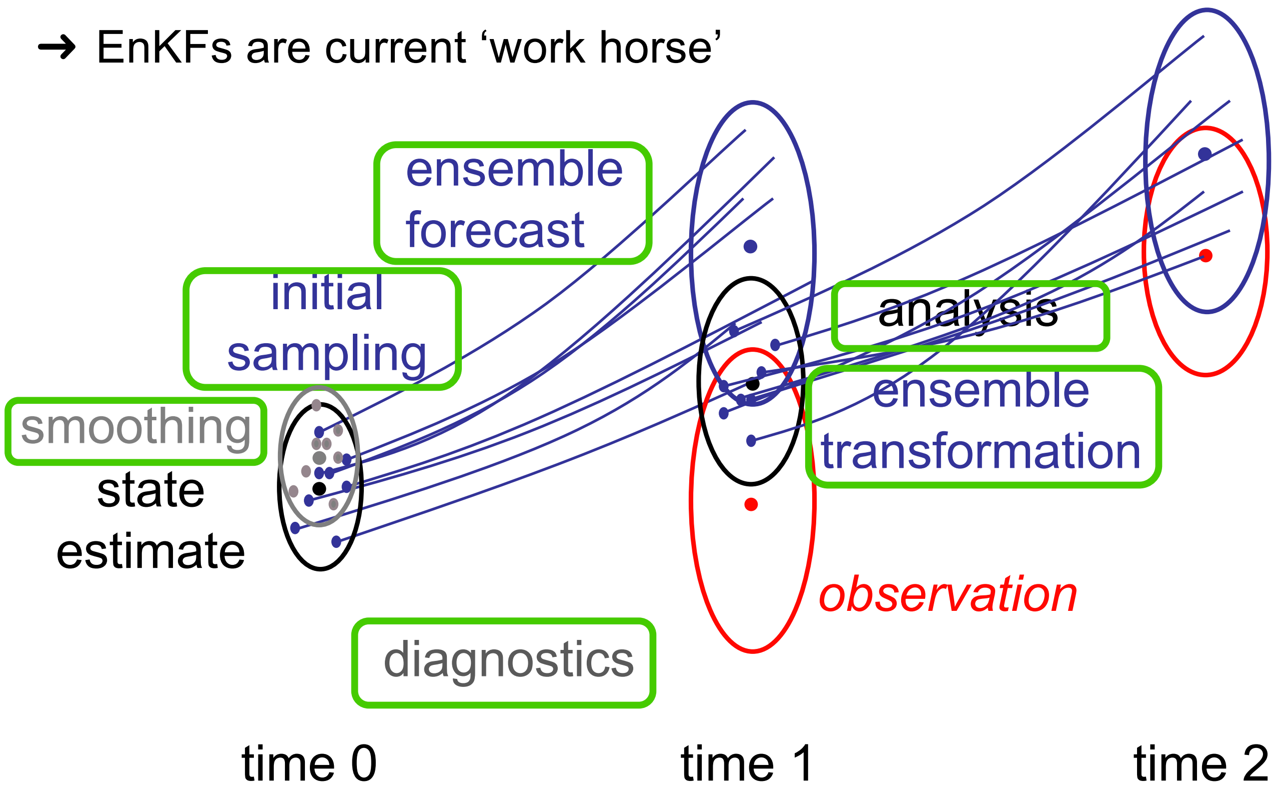


Ensemble Kalman Filters (EnKFs) & Particle Filters

➜ Use ensembles to represent probability distributions (uncertainty)

➜ Use observations to update ensemble

➜ EnKFs are current 'work horse'




There are many possible choices!

What is optimal is part of our research

Different choices in PDAF

## Data Assimilation Group @ AWI: Research Interests

- Ensemble-based data assimilation algorithms
  - Understanding, improvement and development of algorithms
  - In particular for high-dimensional and nonlinear systems
  - Ensemble Kalman filters, particle filters, ensemble variational schemes
- Applicability of ensemble assimilation methods to complex models

  ➜ Software PDAF
- Applications of data assimilation
  - Ocean physics, sea ice, biogeochemistry
  - Coupled Earth system models

  ➜ Applications provide insight into skill of assimilation method
  (cannot assessed purely mathematically)

# PDAF: A tool for data assimilation

PDAF - Parallel Data Assimilation Framework

- a program library for ensemble data assimilation
- provides support for parallel ensemble forecasts
- provides filters and smoothers - fully-implemented & parallelized
  (EnKF, LETKF, LESTKF, NETF, PF … easy to add more)
- easily useable with (probably) any numerical model
- run from laptops to supercomputers (Fortran, MPI & OpenMP)
- Usable for real assimilation applications and to study assimilation methods
- first public release in 2004; continued development
- ~400 registered users; community contributions

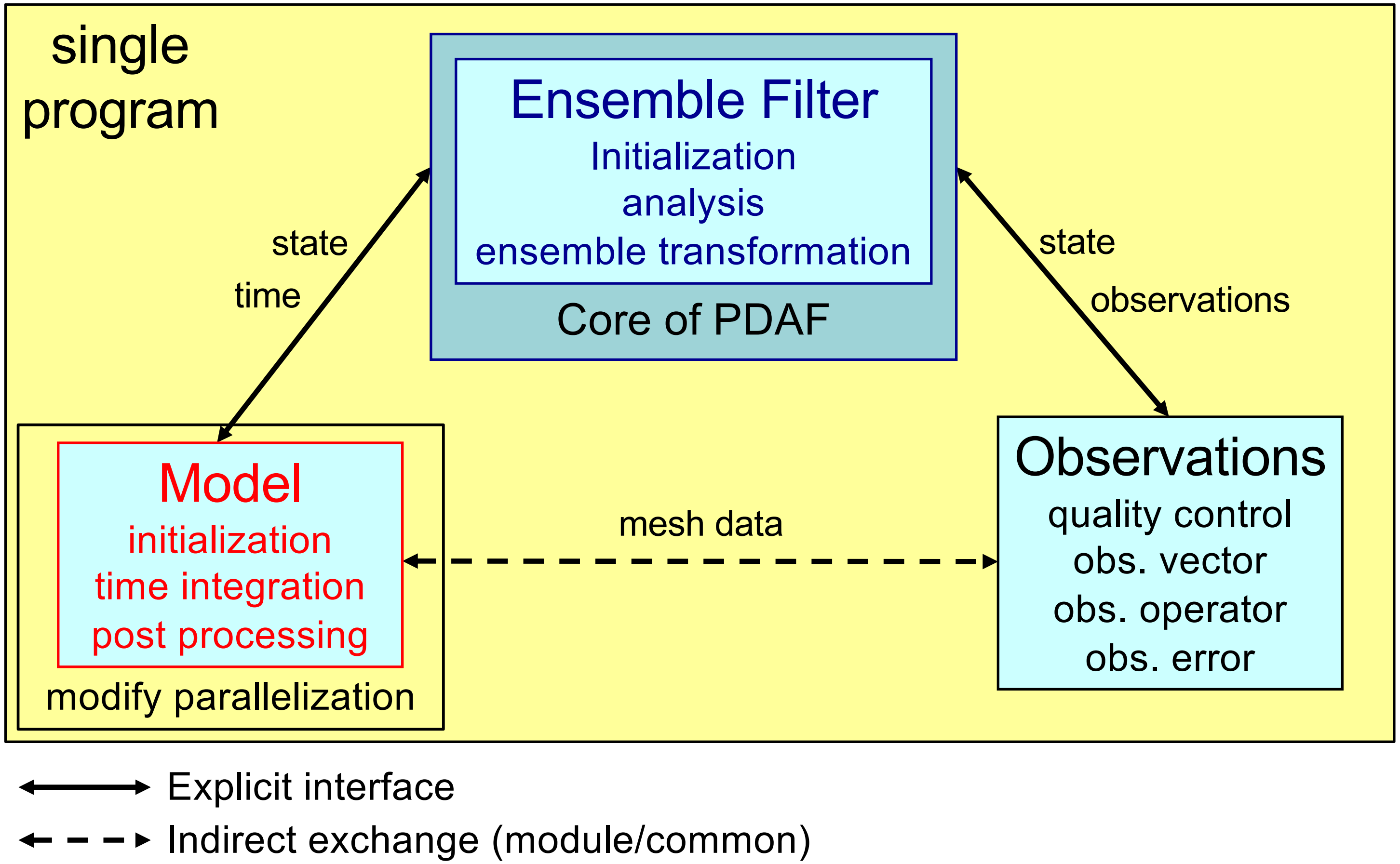
Open source:
Code, documentation, and tutorial available at

http://pdaf.awi.de

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

# 3 Components of Assimilation System

PDAF Parallel Data Assimilation Framework




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

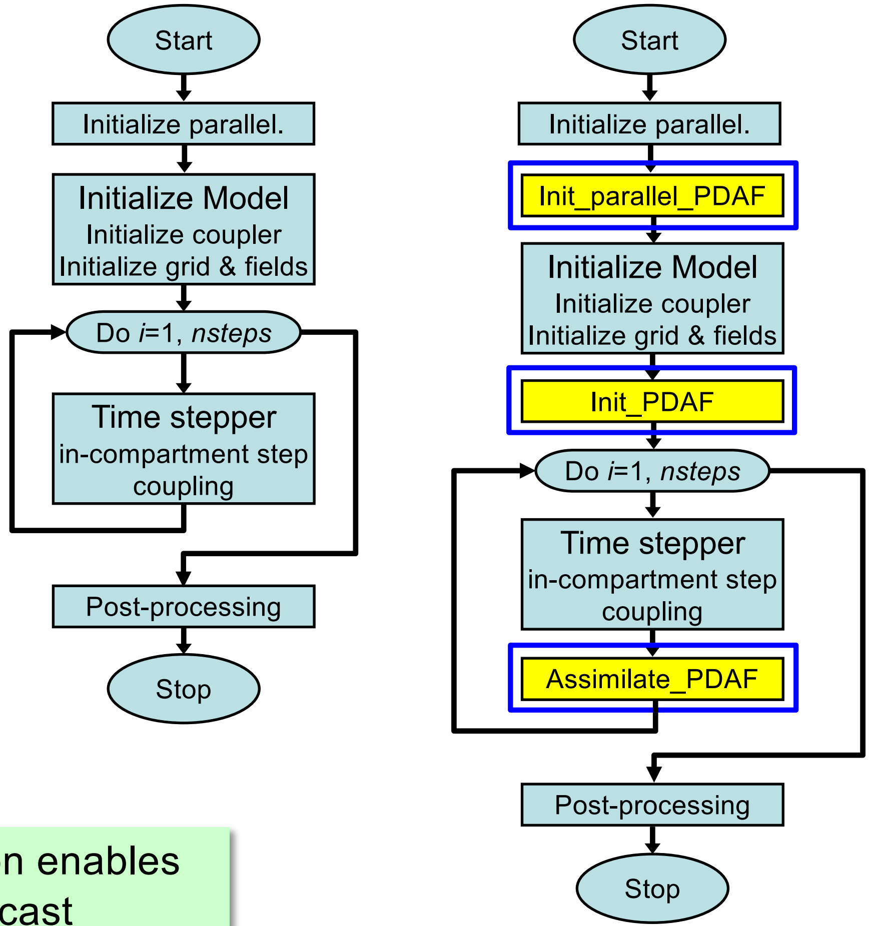
# Augmenting a Model for Data Assimilation

PDAF Parallel Data Assimilation Framework

Model

*single or multiple executables*

*coupler might be separate program*




Extension for data assimilation

*plus:*
Possible model-specific adaption

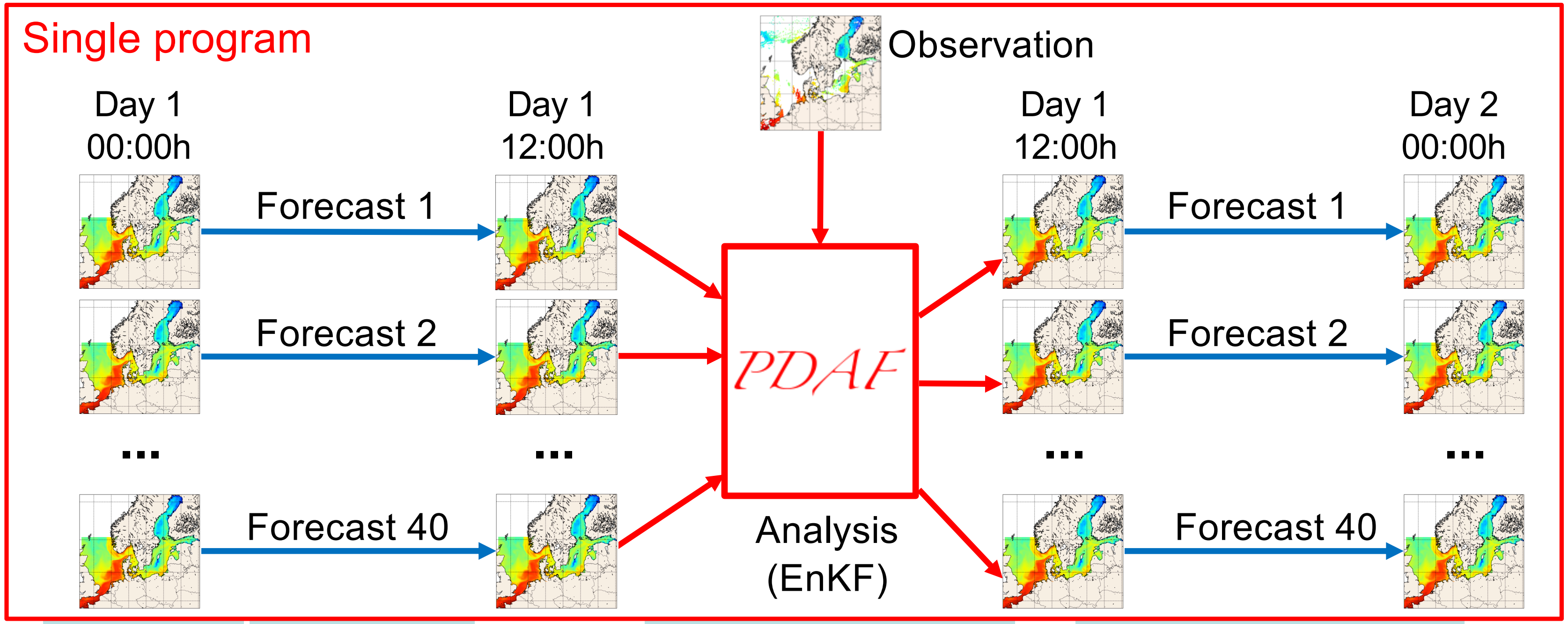
e.g. in NEMO: treat leap-frog time stepping

revised parallelization enables ensemble forecast

# Augmenting a Model for Data Assimilation

Couple PDAF with model

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



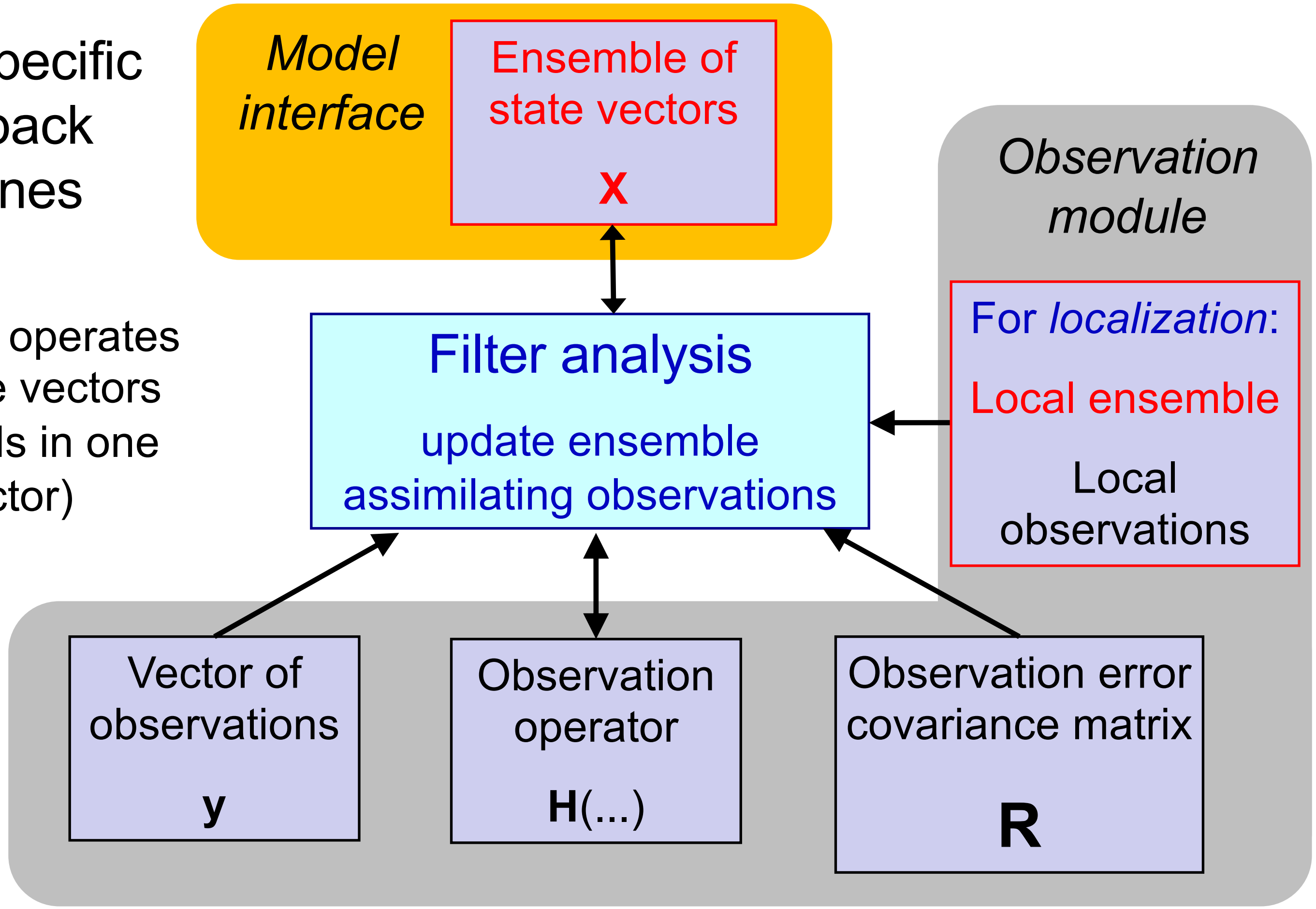

Initialize ensemble | Ensemble forecast | Analysis step in between time steps | Ensemble forecast with changed fields

# Ensemble Filter Analysis Step

PDAF Parallel Data Assimilation Framework

case-specific call-back routines

Analysis operates on state vectors (all fields in one vector)



## The Ensemble Kalman Filter (EnKF, Evensen 94)

Ensemble $\{\mathbf{x}_0^{a(l)}, l = 1, \dots, N\}$

Ensemble covariance matrix
$$\mathbf{P}_k^f := \frac{1}{N-1}\sum_{l=1}^{N}\left(\mathbf{x}_k^{f(l)} - \overline{\mathbf{x}_k^f}\right)\left(\mathbf{x}_k^{f(l)} - \overline{\mathbf{x}_k^f}\right)^T$$

Ensemble mean (state estimate)
$$\mathbf{x}_k^a := \frac{1}{N}\sum_{l=1}^{N}\mathbf{x}_k^{a(l)}$$

**Analysis step:**

Update each ensemble member

**Kalman filter**

$$\boxed{\begin{aligned}\mathbf{x}_k^{a(l)} &= \mathbf{x}_k^{f(l)} + \mathbf{K}_k\left(\mathbf{y}_k^{(l)} - \mathbf{H}_k\mathbf{x}_k^{f(l)}\right)\\ \mathbf{K}_k &= \mathbf{P}_k^f\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}\end{aligned}}$$

**Expensive to compute**
(in practice we use a more efficient formulation)

If elements of $\mathbf{x}$ are observed:

- K contains
  - observed rows
  - unobserved rows

Unobserved variables updated through cross-covariances in $\mathbf{P}$ (linear regression)

# Current algorithms in PDAF

PDAF Parallel Data Assimilation Framework

PDAF originated from comparison studies of different filters

**Filters and smoothers**

- EnKF (Evensen, 1994 + perturbed obs.)
- (L)ETKF (Bishop et al., 2001)
- SEIK filter (Pham et al., 1998)
- ESTKF (Nerger et al., 2012)
- NETF (Toedter & Ahrens, 2015)

- Particle filter (PF)
- Generate synthetic observations

Not yet released:
- serial EnSRF
- EWPF

**All methods include (except PF)**

- global and localized versions
- smoothers

**Model binding**

- MITgcm

**Toy models**

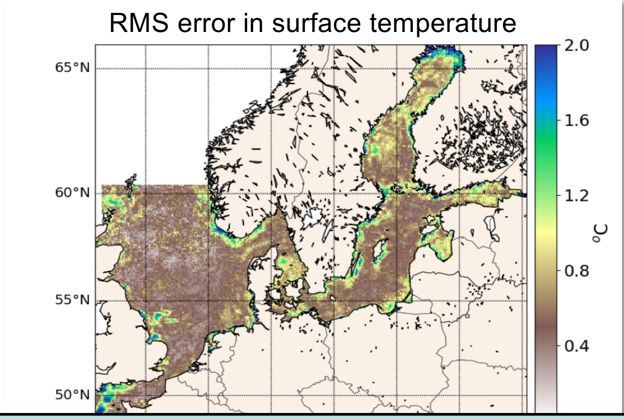
- Lorenz-96, Lorenz63

Not yet released:
- AWI-CM model binding
- NEMO model binding

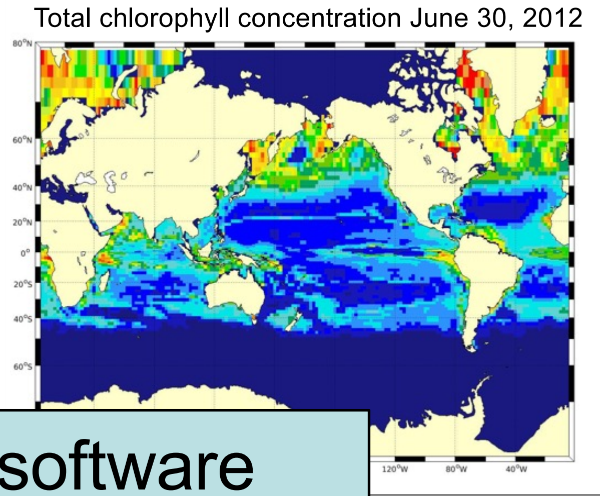
# PDAF Application Examples

PDAF Parallel Data Assimilation Framework

**HBM-ERGOM**: Coastal assimilation of SST, in situ and ocean color data (Svetlana Losa, Michael Goodliff)

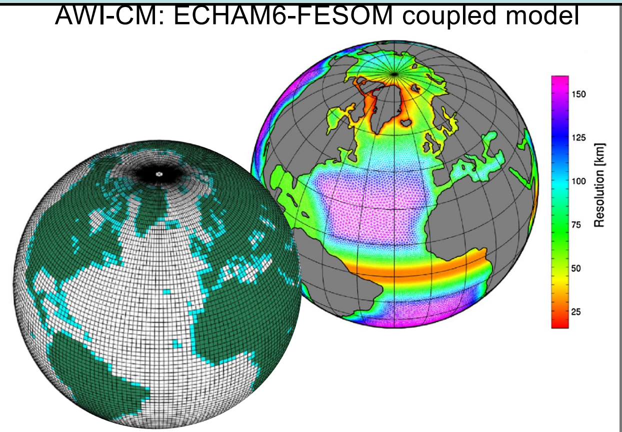



**MITgcm-REcoM**: global ocean color assimilation (Himansu Pradhan)




## Different models – same assimilation software

**AWI-CM**: coupled atmos.-ocean assimilation (Qi Tang, Longjiang Mu)

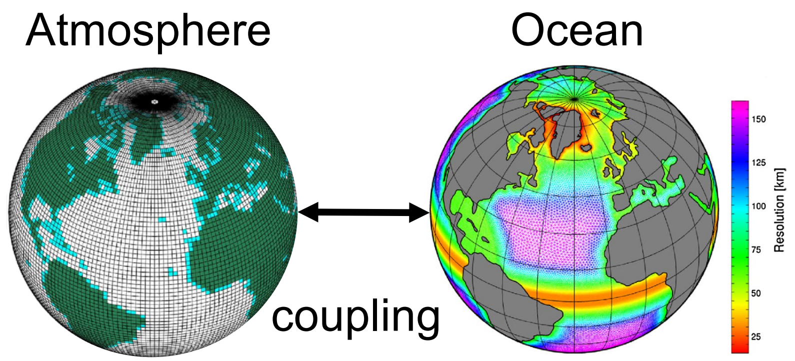



+ external applications & users, like

- MITgcm sea-ice assim (NMEFC Beijing)
- Geodynamo (IPGP Paris, A. Fournier)
- TerrSysMP-PDAF (hydrology, FZ Juelich)
- CMEMS Baltic-MFC (operational, DMI/BSH/SMHI)
- CFSv2 (J. Liu, IAP-CAS Beijing)
- NEMO (U. Reading , P. J. van Leeuwen)

# Coupled Models and Coupled Data Assimilation






Coupled models

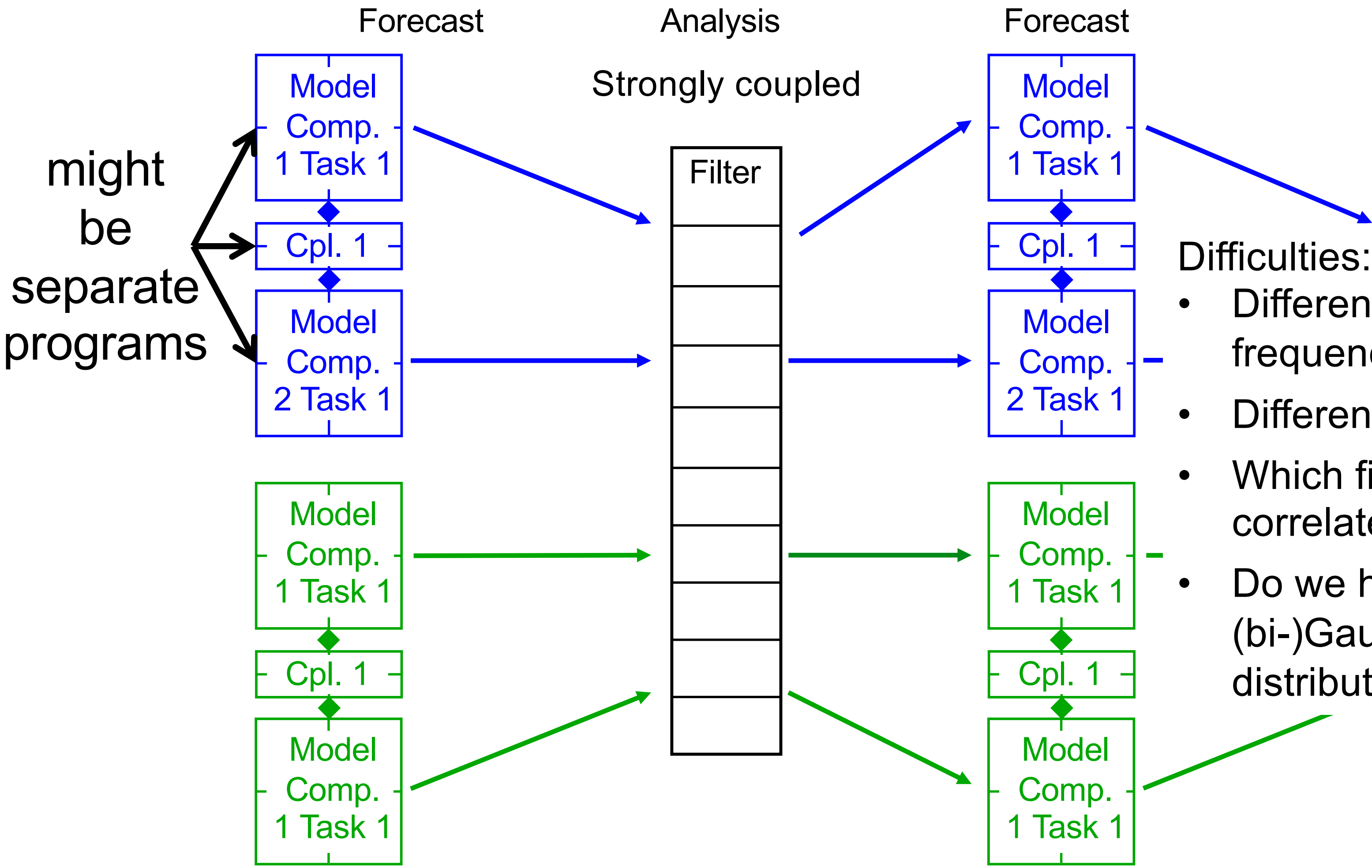
- Several interconnected compartments, like
  - Atmosphere and ocean
  - Ocean physics and biogeochemistry (carbon, plankton, etc.)

Coupled data assimilation

- Assimilation into coupled models
  - Weakly coupled: separate assimilation in the compartments
  - Strongly coupled: joint assimilation of the compartments
    - ➜ Use cross-covariances between fields in compartments
  - Plus various “in between” possibilities …

# 2 compartment system – strongly coupled DA



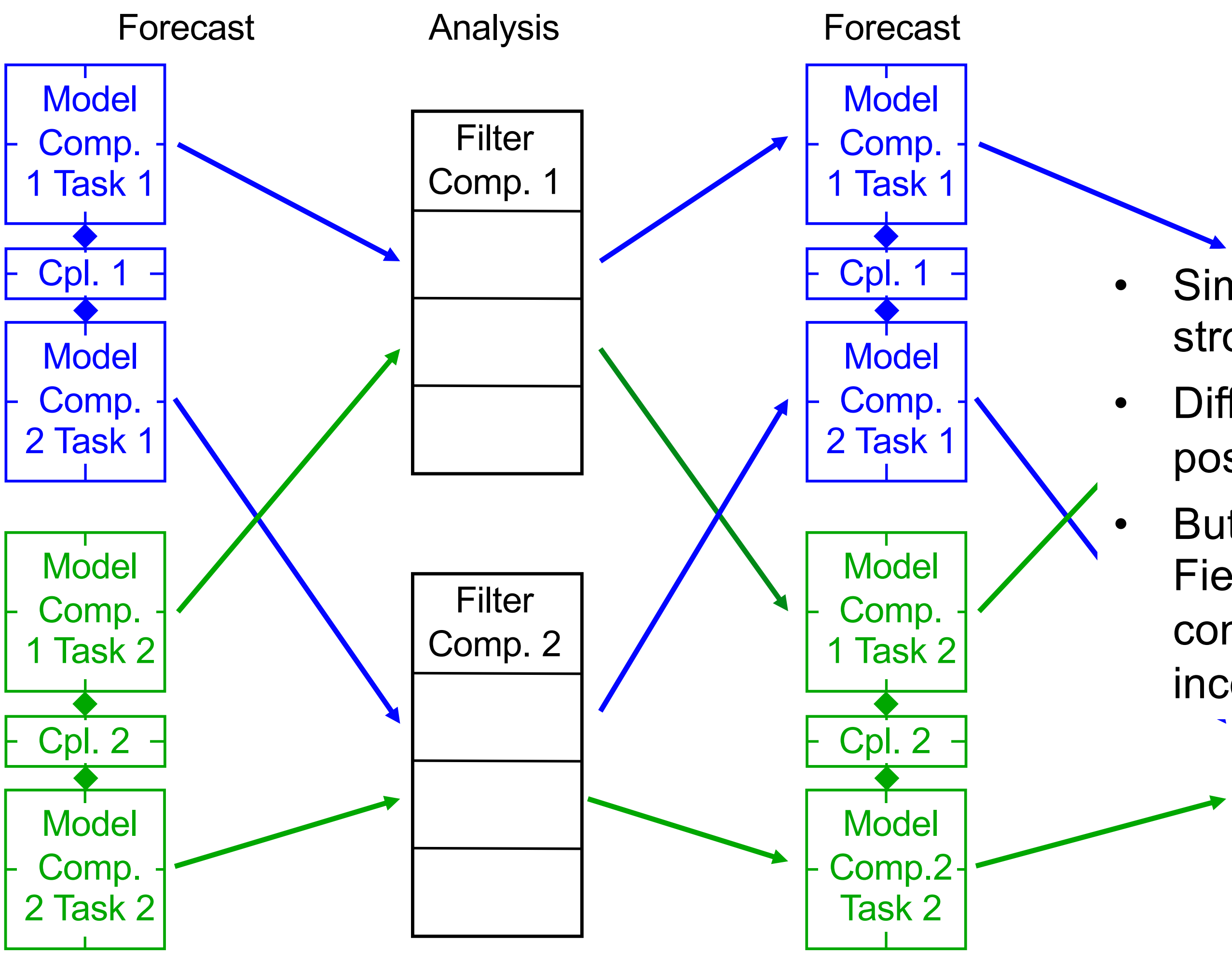



Difficulties:

- Different assimilation frequency
- Different time scales
- Which fields are correlated?
- Do we have (bi-)Gaussian distributions?

# 2 compartment system – weakly coupled DA

PDAF Parallel Data Assimilation Framework




- Simpler setup than strongly coupled
- Different DA methods possible
- But:
  Fields in different compartments can be inconsistent

Example 1

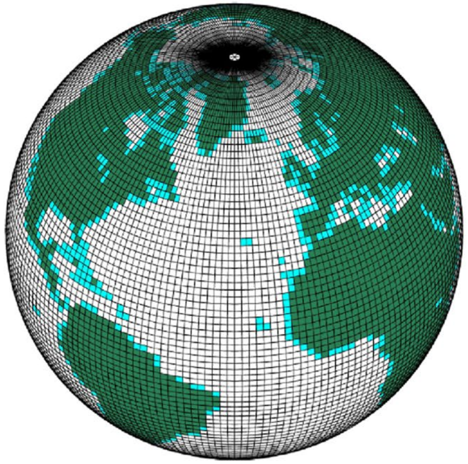
# Assimilation into the coupled atmosphere-ocean model AWI-CM

(Qi Tang)

Project: ESM – Advanced Earth System Modeling Capacity

# Assimilation into coupled model: AWI-CM

Atmosphere

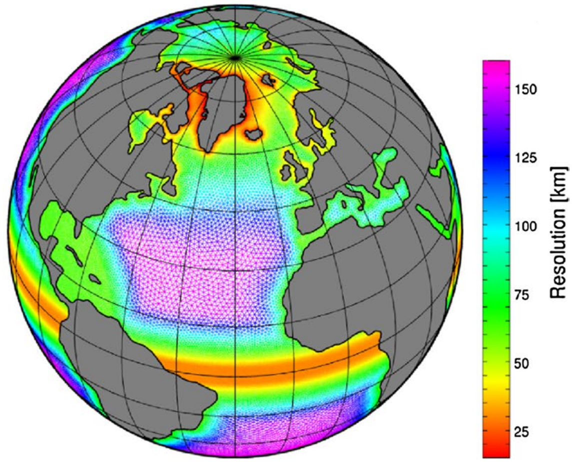


OASIS3-MCT

fluxes

ocean/ice state

Ocean




**Atmosphere**
- ECHAM6
- JSBACH land

**Coupler library**
- OASIS3-MCT

**Ocean**
- FESOM
- includes sea ice

Two separate executables for atmosphere and ocean

**Goal: Develop data assimilation methodology for cross-domain assimilation (“strongly-coupled”)**

ESM Advanced Earth System Modelling Capacity

AWI-CM: Sidorenko et al., Clim Dyn 44 (2015) 757

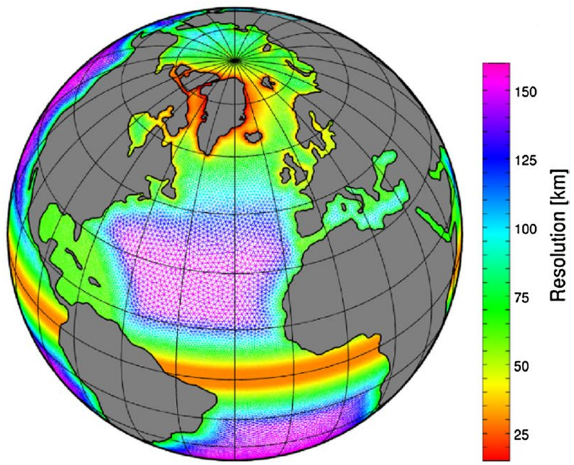
AWI

# Data Assimilation Experiments

## Model setup

- Global model
- ECHAM6: T63L47
- FESOM: resolution 30-160km

FESOM mesh resolution




## Data assimilation experiments

- Observations
  - Satellite SST
  - Profiles temperature & salinity
- Updated: ocean state (SSH, T, S, u, v, w)
- Assimilation method: Ensemble Kalman Filter (LESTKF)
- Ensemble size: 46
- Simulation period:  year 2016, daily assimilation update
- Run time: 5.5h, fully parallelized using 12,000 processor cores

# Offline coupling - Efficiency

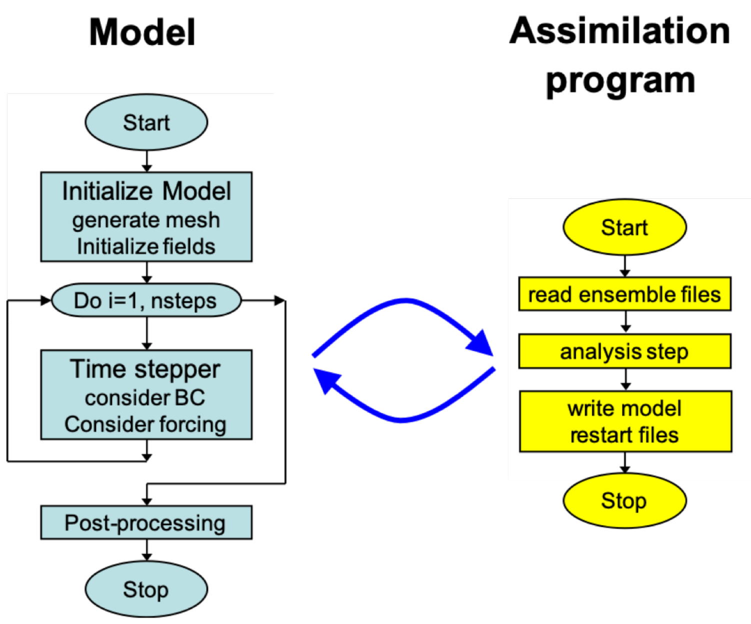
Offline-coupling is simple to implement
but can be very ineffcient

**Example:**
Timing from atmosphere-ocean
coupled model (AWI-CM)
with daily analysis step:

| | | |
|---|---|---|
| Model startup: | 95 s | overhead |
| Integrate 1 day: | 28 s | |
| Model postprocessing: | 14 s | overhead |
| Analysis step: | 1 s | |

Restarting this model is ~3.5 times
more expensive than integrating 1 day

➜ avoid this for data assimilation

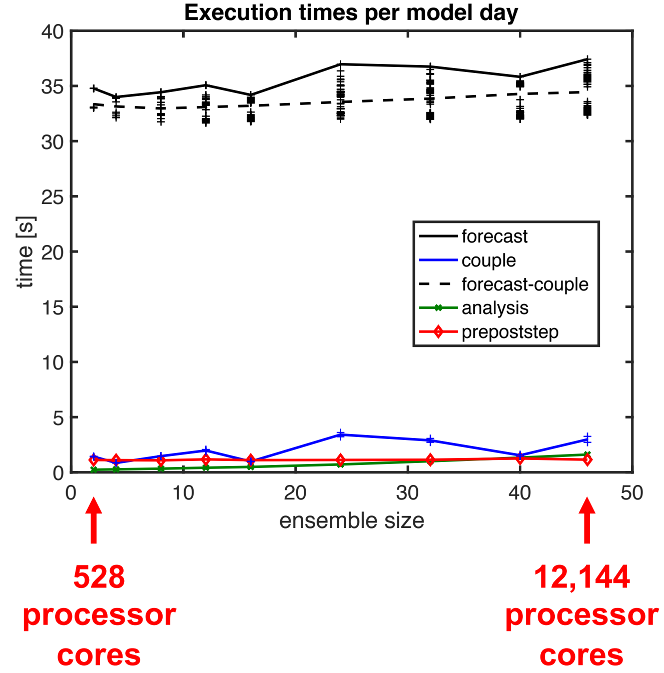


# Execution times (weakly-coupled, DA only into ocean)

MPI-tasks

- ECHAM: 72
- FESOM: 192

- Increasing integration time with growing ensemble size (11%; more parallel communication; worse placement)
- some variability in integration time over ensemble tasks

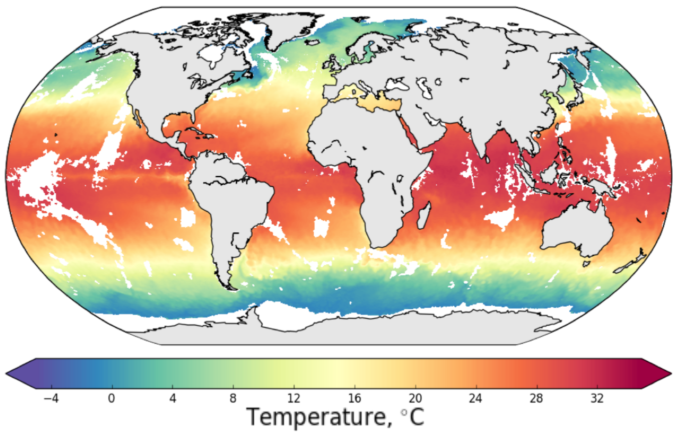



Important factors for good performance

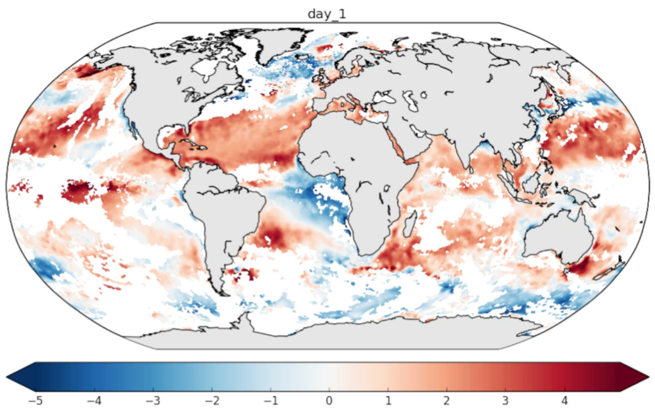
- Need optimal distribution of programs over compute nodes/racks (here set up as ocean/atmosphere pairs)
- Avoid conflicts in IO (Best performance when each AWI-CM task runs in separate directory)

# Assimilate sea surface temperature (SST)

SST on Jan 1st, 2016




SST difference: observations-model




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

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

DA with such a coupled model is unstable!

omit SST observations where

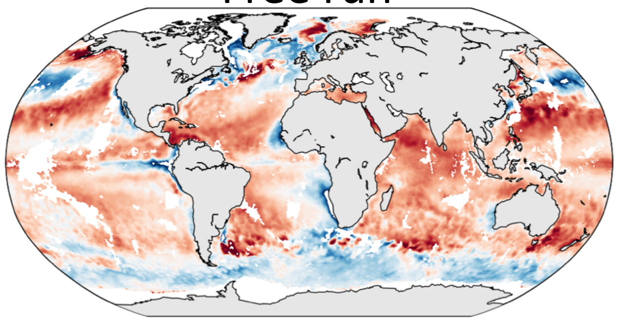
$|SST_{obs} - SST_{ens\_mean}| > 1.6\ ^{o}C$

(30% initially, <5% later)

# SST assimilation: Effect on the ocean

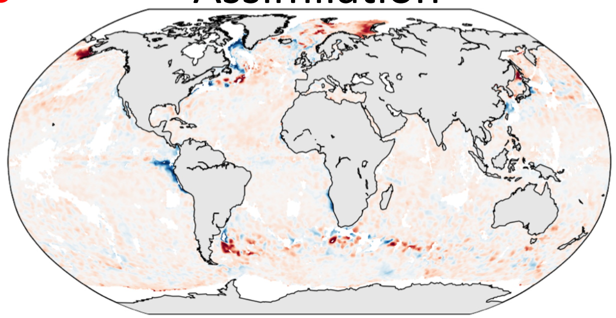
SST difference (obs-model): strong decrease of deviation

Free run



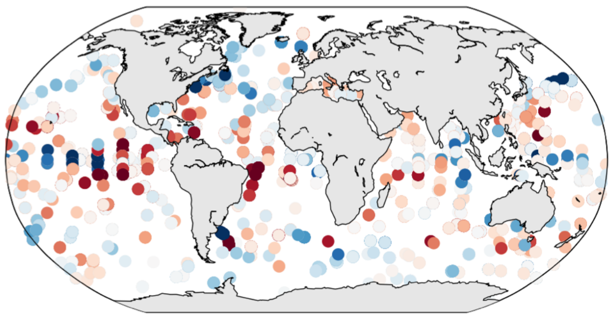
4/30/2016
Day 120

Assimilation



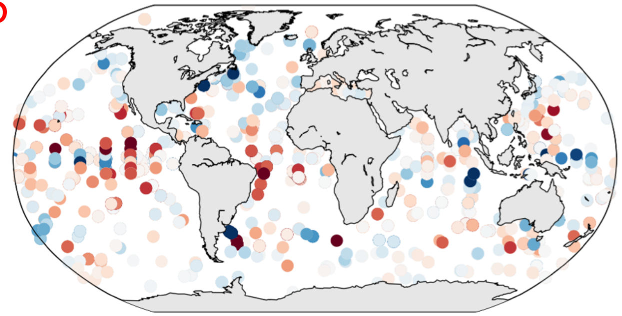
Subsurface temperature difference (obs-model); all the model layers at profile locations

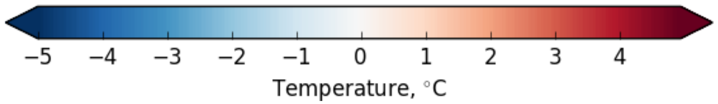
Free run



4/30/2016
Day 120

Assimilation



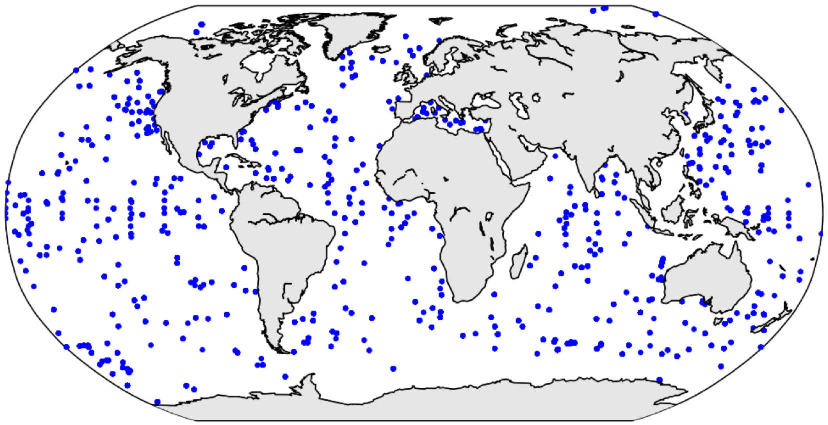



ESM Advanced Earth System Modelling Capacity

AWI

# Assimilate subsurface observations: Profiles

Profile locations on Jan 1st, 2016

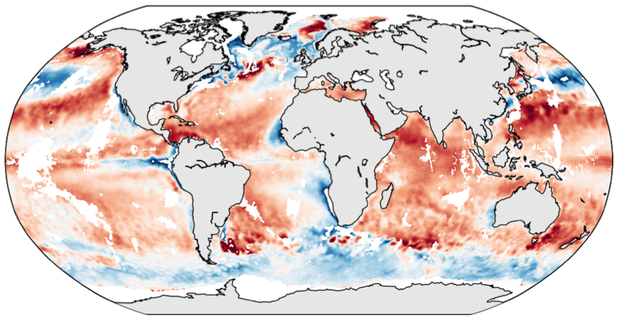


- Temperature and Salinity
- EN4 data from UK MetOffice
- Daily data
- Subsurface down to 5000m
- About 1000 profiles per day
- Observation errors
  - Temperature profiles: 0.8 °C
  - Salinity profiles: 0.5 psu
- Localization radius: 1000 km

# SST assimilation: Effect on the ocean

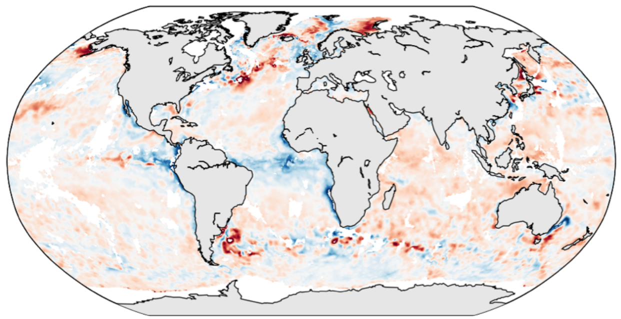
SST difference (obs-model)

Free run



4/30/2016
Day 120

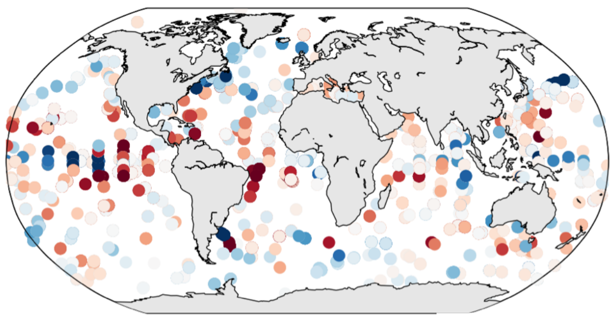
Assimilation



larger deviations than for SST assimilation

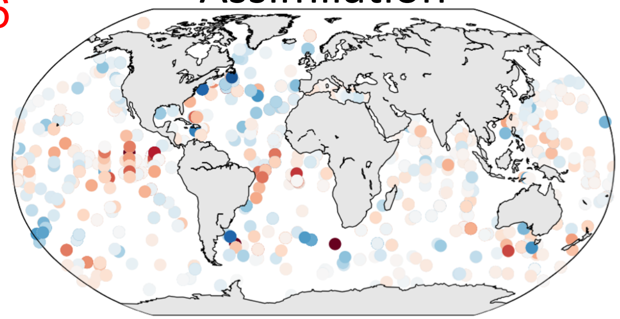
Subsurface temperature difference (obs-model); all the model layers at profile locations

Free run

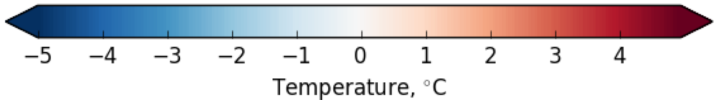


4/30/2016
Day 120

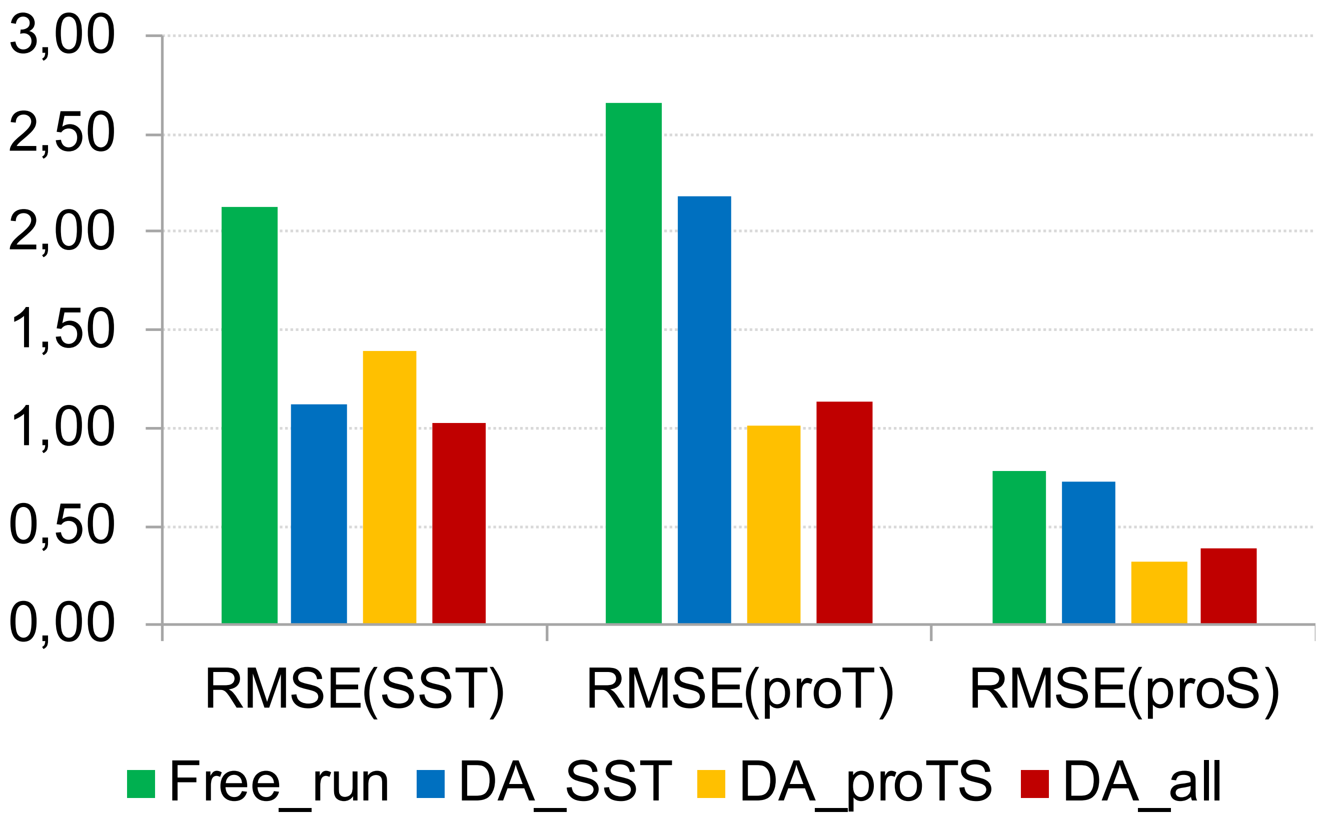
Assimilation



smaller deviations than for SST assimilation



# Assimilation effect: RMS errors




Overall lowest errors with combined assimilation

- But partly a compromise

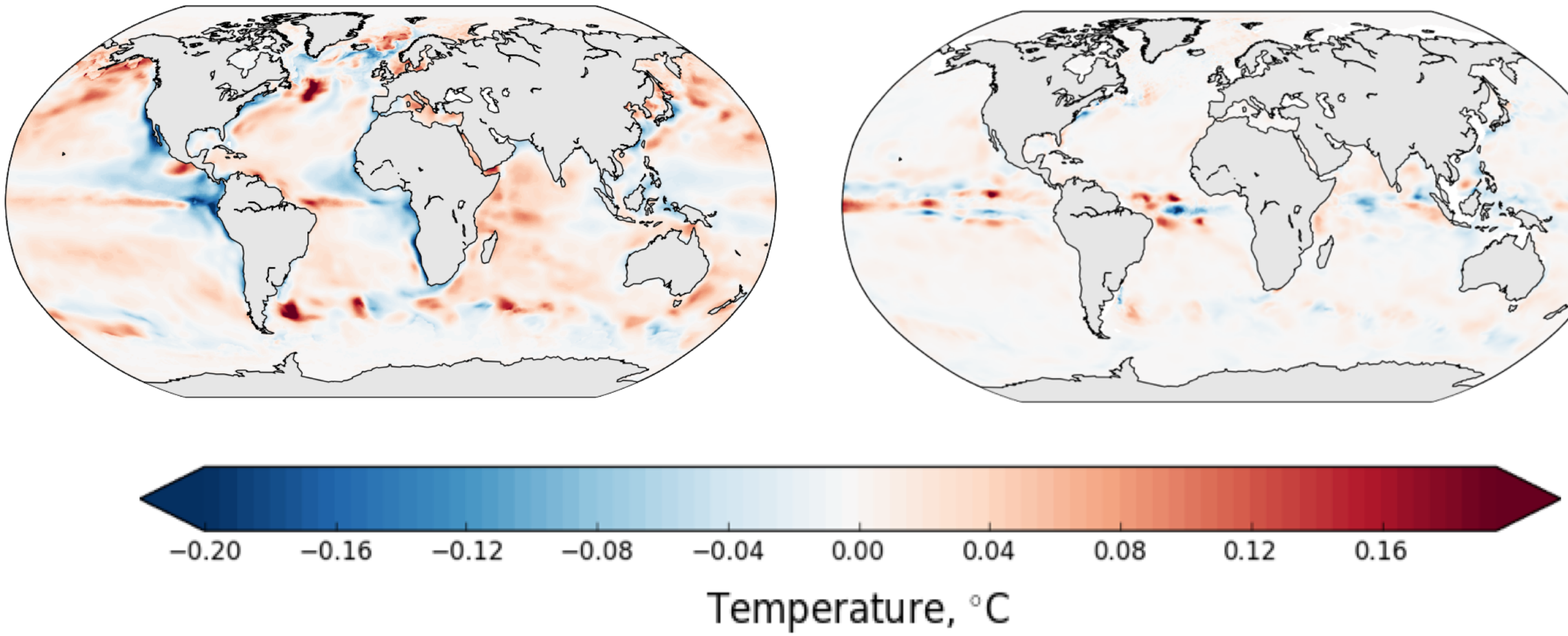
# Mean increments

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

→ non-zero values indicate regions with possible biases

Surface temperature

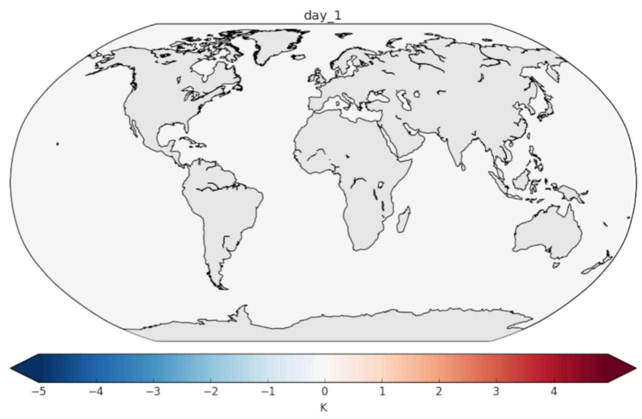
Temperature at 100m depth



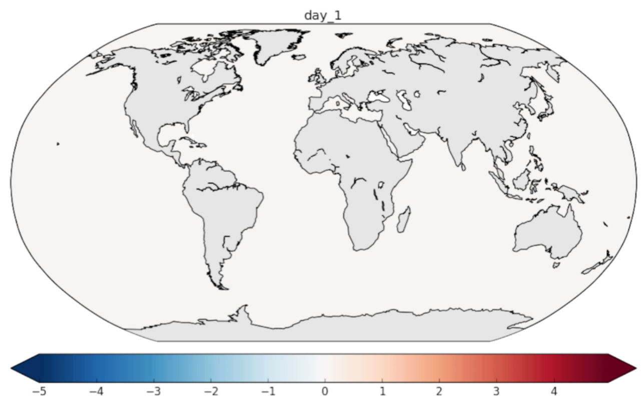
# Assimilation Effect on the Atmosphere

Difference between assimilation runs and the free run

Temperature at 2m



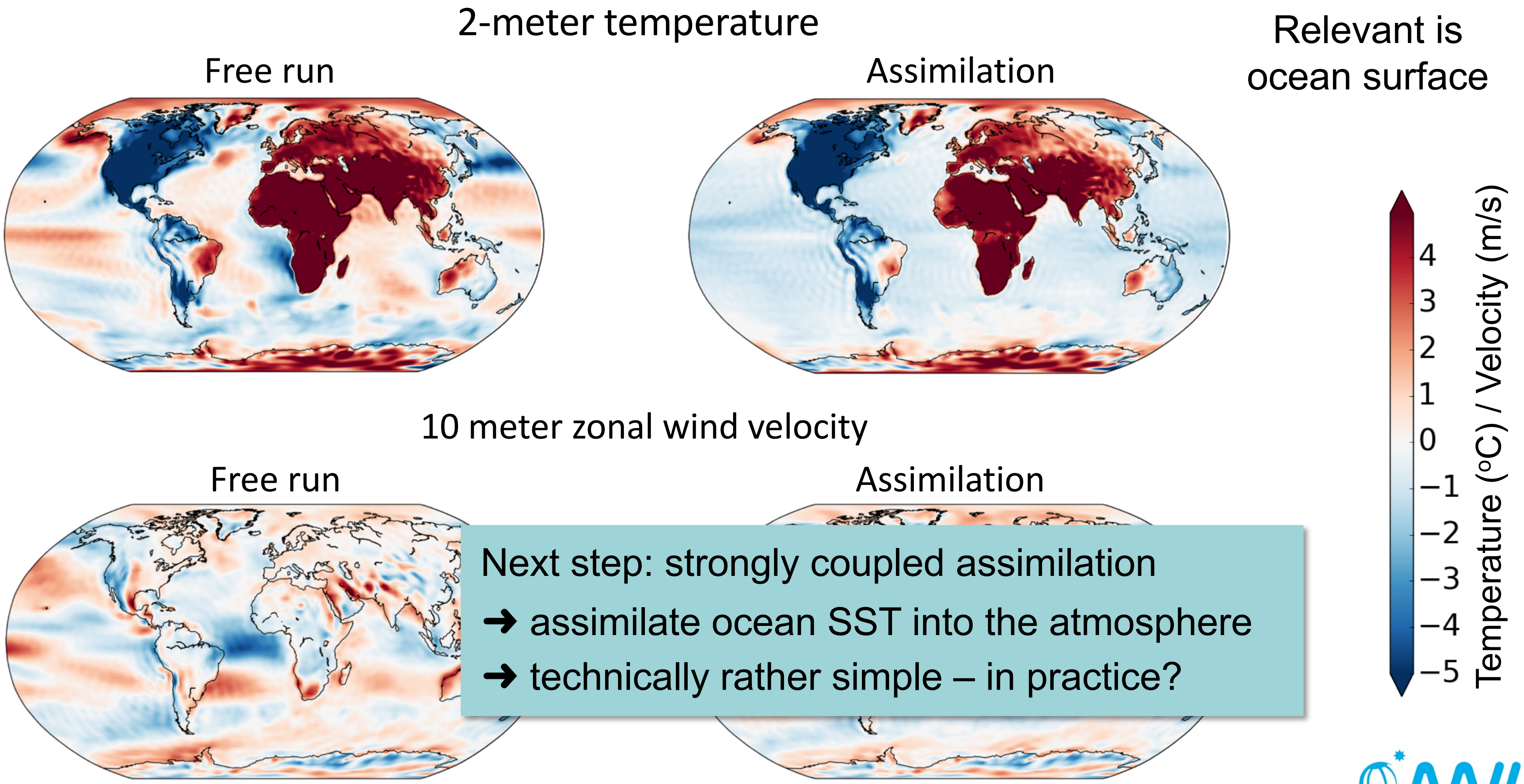

Sea surface temperature




Atmosphere reacts quickly on the changed ocean state

Does it make the atmosphere more realistic?

# Effect on Atmospheric State (annual mean)



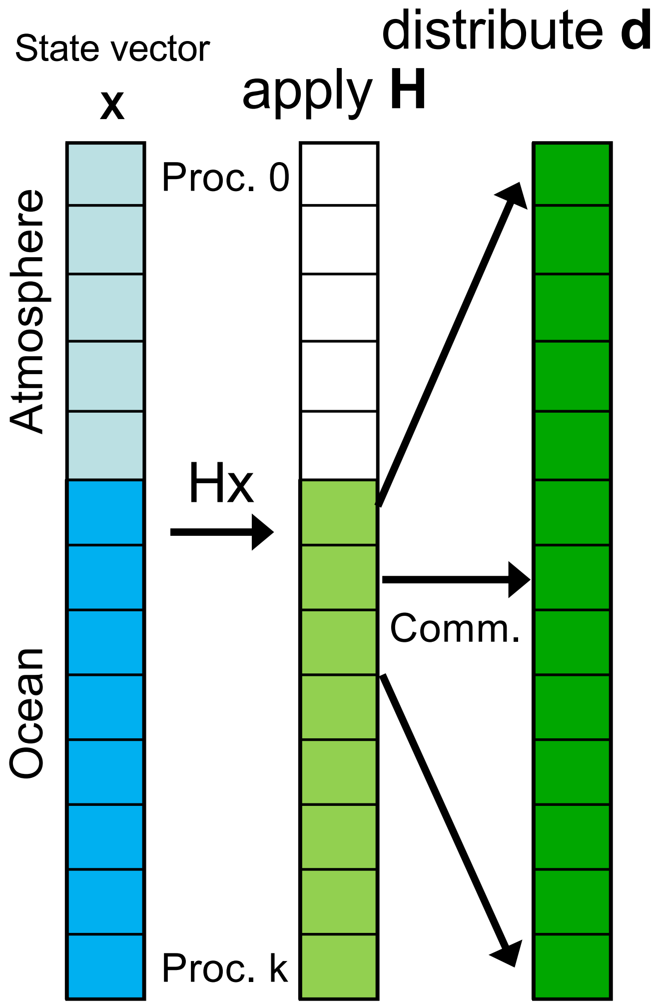

Relevant is ocean surface

Next step: strongly coupled assimilation

➜ assimilate ocean SST into the atmosphere

➜ technically rather simple – in practice?

# Strongly coupled: Parallelization of analysis step




We need innovation: $\mathbf{d} = \mathbf{Hx} - \mathbf{y}$

Observation operator links different compartments

1. Compute part of **d** on process 'owning' the observation
2. Communicate **d** to processes for which observation is within localization radius

Example 2

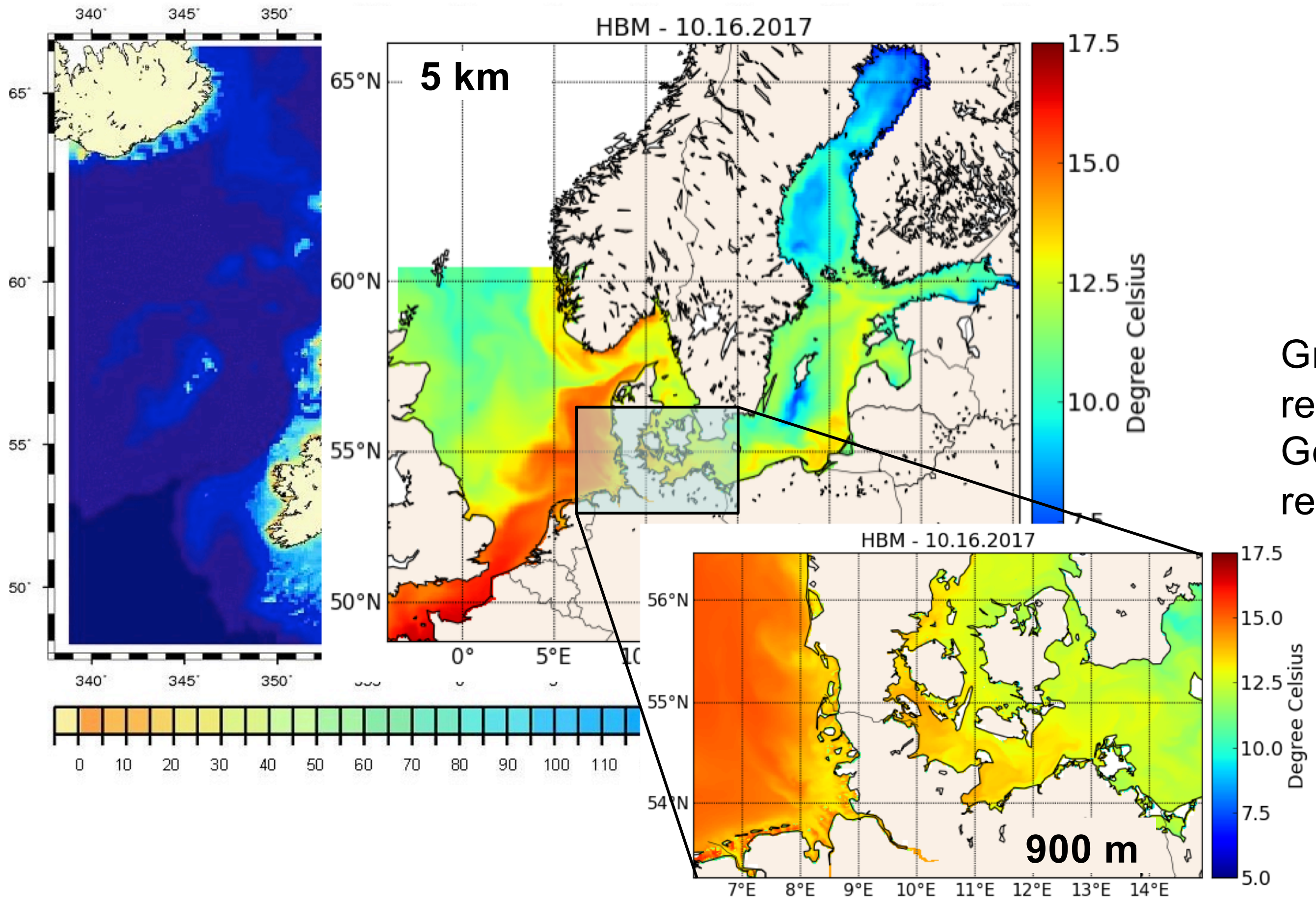
# Weakly- and Strongly Coupled Assimilation to Constrain Biogeochemistry with Temperature Data

(MERAMO – Mike Goodliff)

Cooperation with German Hydrographic Agency (BSH)
(Ina Lorkowski, Xin Li, Anja Lindenthal, Thoger Brüning)

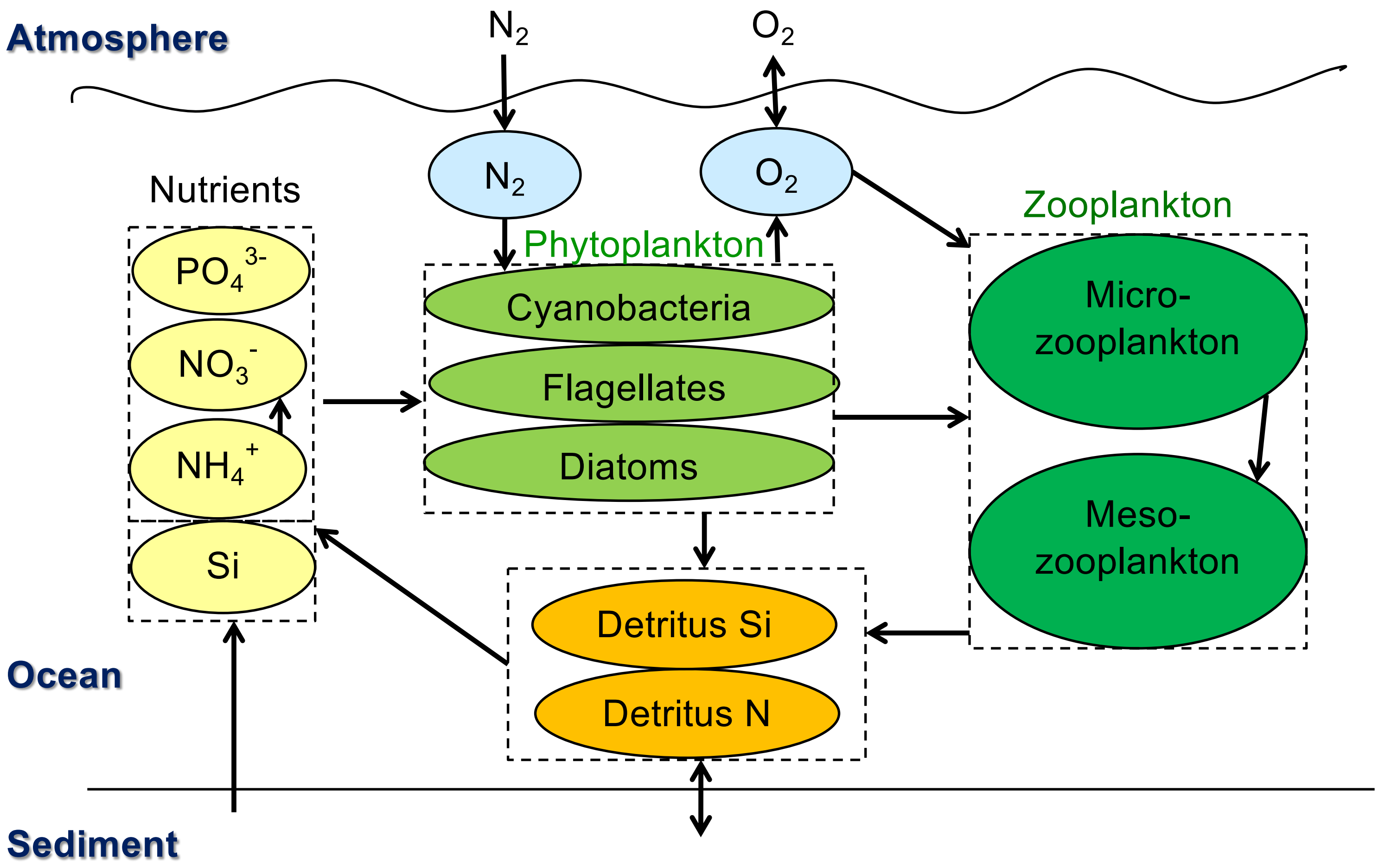
## Coastal Model Domain

**HBM (Hiromb-BOOS Model) – operationally used at German Federal Maritime and Hydrographic Agency (BSH)**



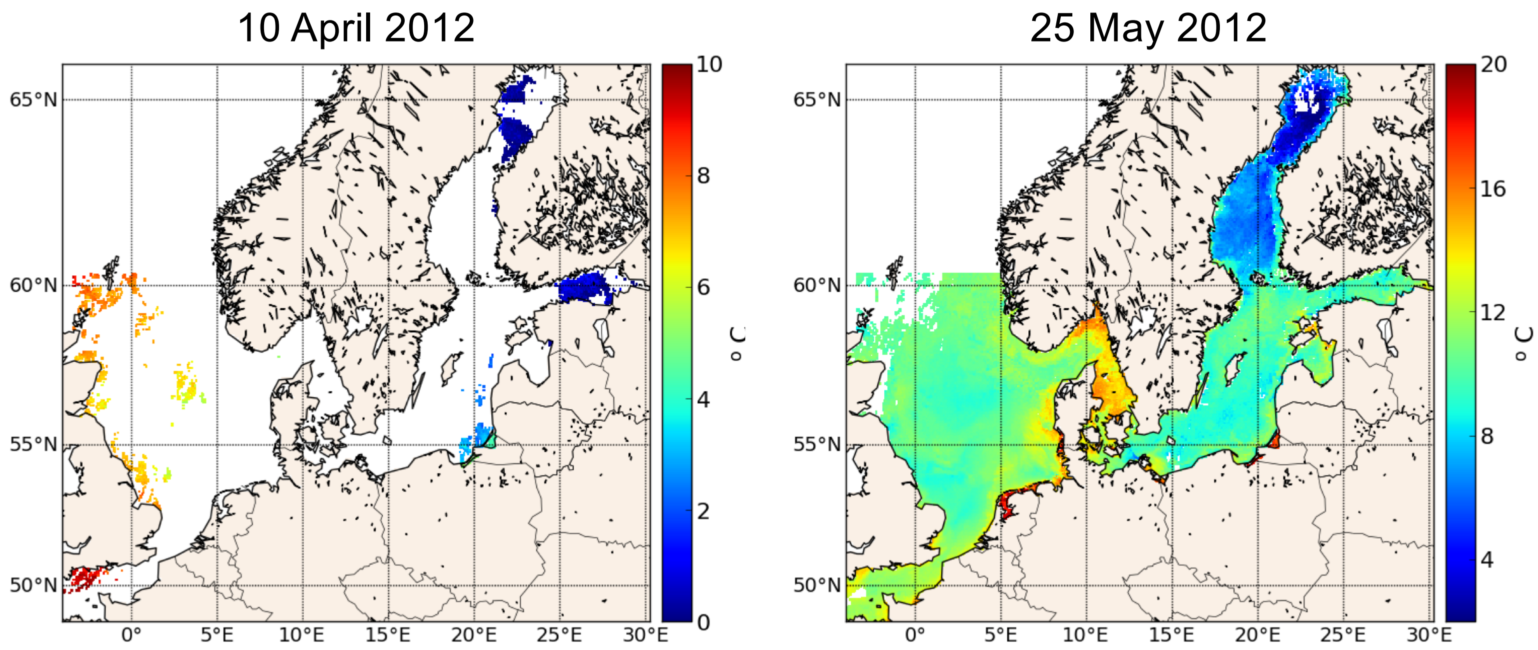

Grid with higher resolution in German coastal region

# Biogeochemical model: ERGOM



# Observations – Sea  Surface Temperature (SST)

NOAA/AVHRR Satellite data




- 12-hour composites on both model grids
- Vastly varying data coverage (due to clouds)
- Effect on biogeochemistry?

# Comparison with assimilated SST data (4-12/2012)

- RMS deviation from SST observations up to ~0.4 °C

**Coarse grid**:

- Increasing error-reductions compared to free ensemble run

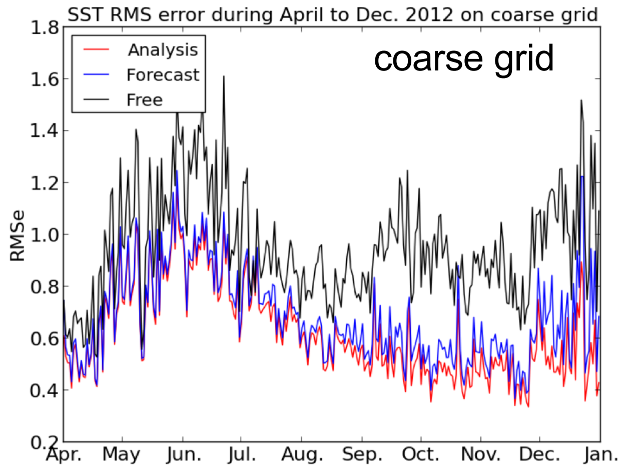
**Fine grid**:

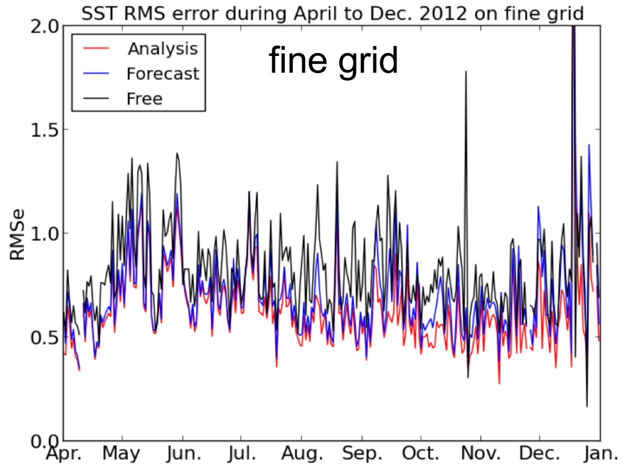
- much stronger variability
- Forecast errors sometimes reach errors of free ensemble run

**RMS errors (deg. C)**

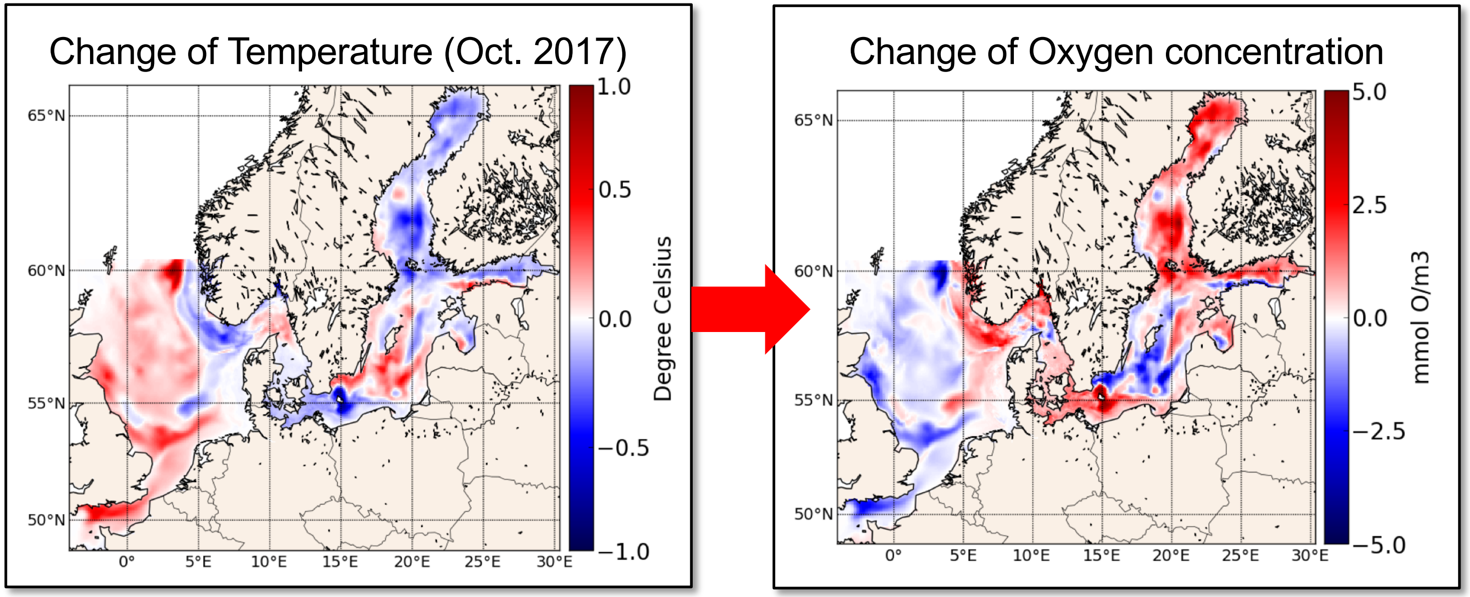
| | Free | Forec. | Ana. |
|---|---|---|---|
| Coarse | 0.95 | 0.68 | 0.63 |
| Fine | 0.83 | 0.70 | 0.63 |

Temperature RMSD



# Influence of Assimilation on Surface Temperature



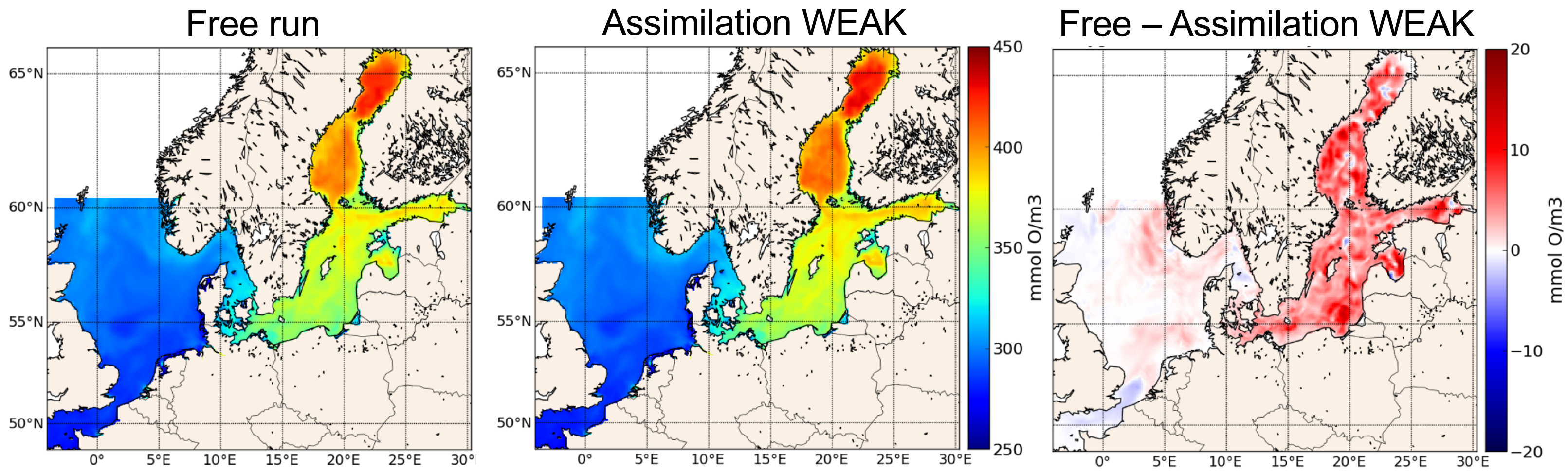

2 ways of influence:

- Indirect - *weakly-coupled assimilation*
  model dynamics react on change in physics

- Direct – *strongly-coupled assimilation*
  use cross-covariances between surface temperature and biogeochemistry

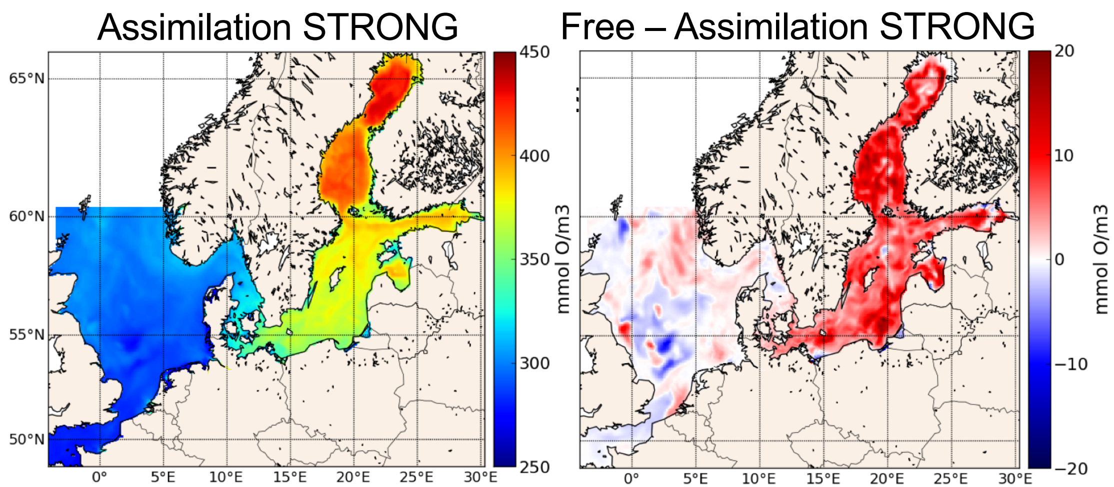
# Weakly & strongly coupled effect on biogeochemical model

Oxygen mean for May 2012 (as mmol O / $m^3$)




Strongly coupled

- slightly larger changes
- Strongly coupled DA further improves oxygen




Goodliff et al., Ocean Dynamics, 2019, doi:10.1007/s10236-019-01299-7

## Choice of variable in strongly coupled assimilation

- Chlorophyll is lognormally distributed
- Ensemble Kalman filter
  - Optimality for normal distributions
  - Linear regression between observed and unobserved variables

➜ Apply strongly-coupled DA with logarithm on concentrations?

**Kalman filter**

$$\mathbf{x}_k^{a(l)} = \mathbf{x}_k^{f(l)} + \mathbf{K}_k\left(\mathbf{y}_k^{(l)} - \mathbf{H}_k\mathbf{x}_k^{f(l)}\right)$$

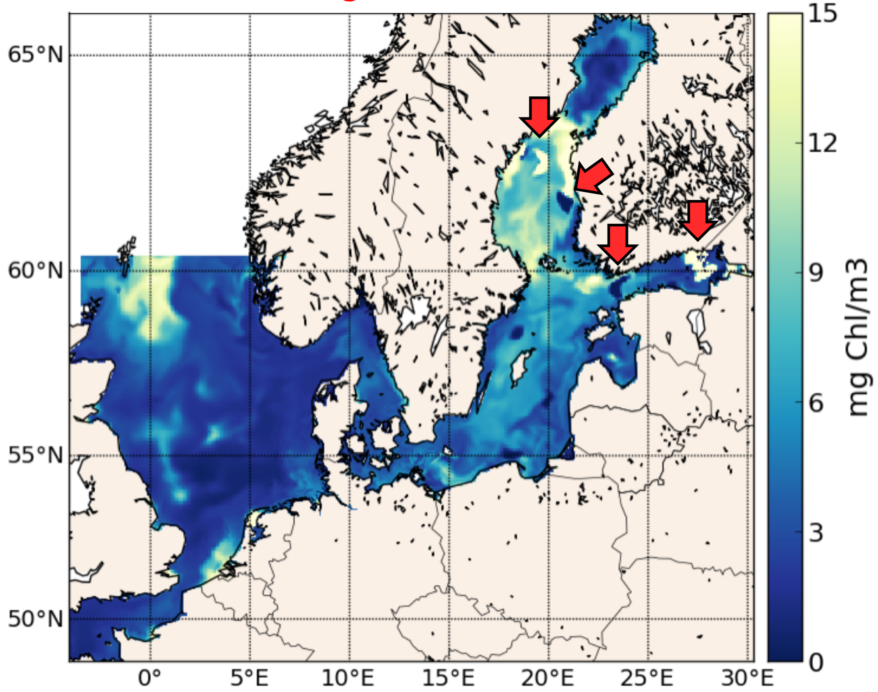
$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}$$

$$\mathbf{K}_k = \underbrace{\mathbf{X}'_k}_{\text{model}}\ \underbrace{(\mathbf{H}_k\mathbf{X}'_k)^T\left(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}}_{\text{observations}}$$

# Choice of variable in strongly coupled assimilation

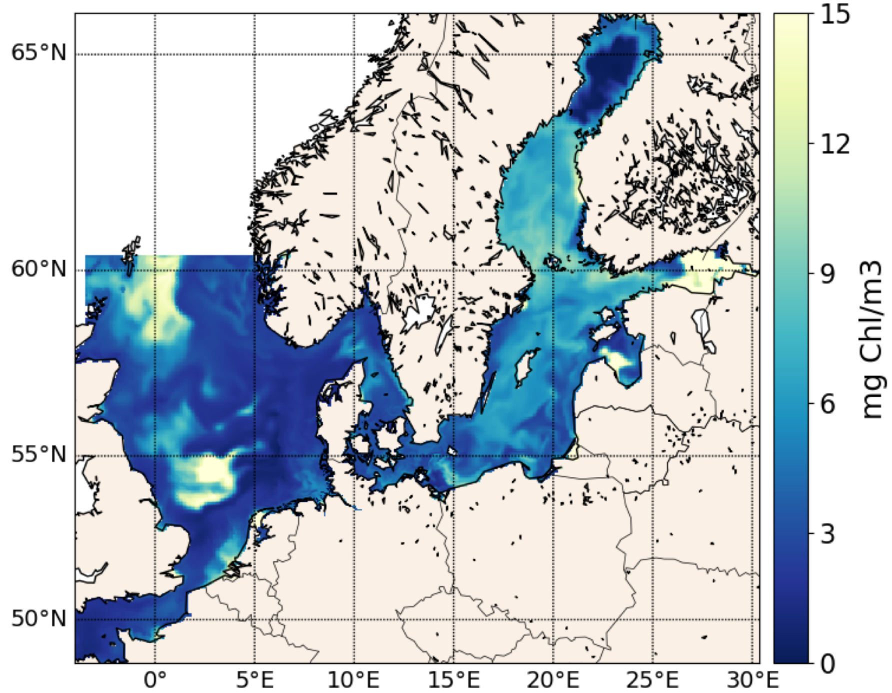
*Chlorophyll concentrations*
*1 May 2012*

Strongly coupled
logarithmic




- locally unrealistically high and low concentrations
  ➜ Linear regression with lognormal concentration not general solution

Strongly coupled
linear




- Larger effect – in particular in North Sea
- Too high in Gulf of Finland

## Summary

- Coupled data assimilation:
  - Weakly-coupled easy to apply
    - But changing one part can disturb the other
  - Strongly-coupled depends on cross-covariances
    - EnKF uses linear regression – variables not well defined
- Unified software helps to bring new developments into usage
  - PDAF – Open source available at **http://pdaf.awi.de**

Lars.Nerger@awi.de

# References

PDAF Parallel Data Assimilation Framework


- http://pdaf.awi.de
- Nerger, L., Hiller, W. *Software for Ensemble-based DA Systems – Implementation and Scalability*. Computers and Geosciences 55 (2013) 110-118
- Nerger, L., Hiller, W., Schröter, J.(2005). *PDAF - The Parallel Data Assimilation Framework: Experiences with Kalman Filtering*, Proceedings of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology, Reading, UK, 25 - 29 October 2004, pp. 63-83.