Parallel
Data
Assimilation
Framework

Ensemble Data Assimilation

for Coupled Models of the Earth System

Lars Nerger, Qi Tang, Mike Goodliff

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

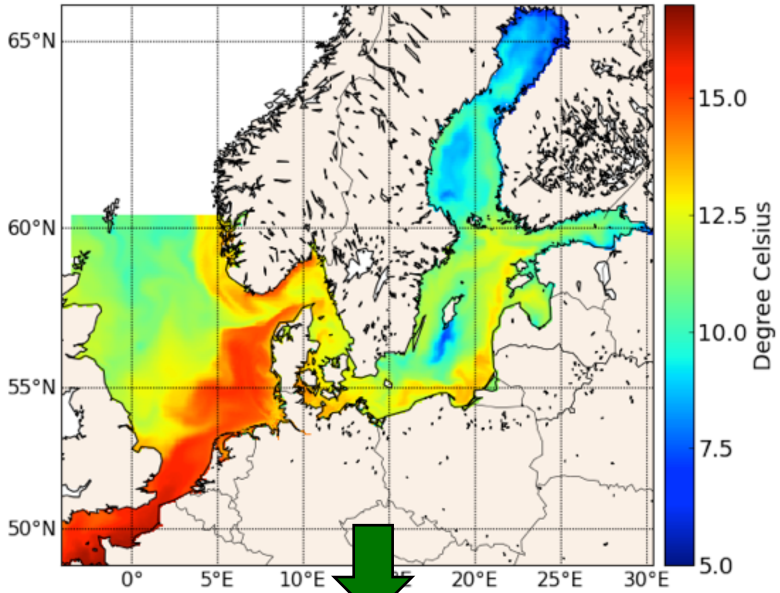
Seminar at SFB 1294, Potsdam, September 13, 2019

Overview

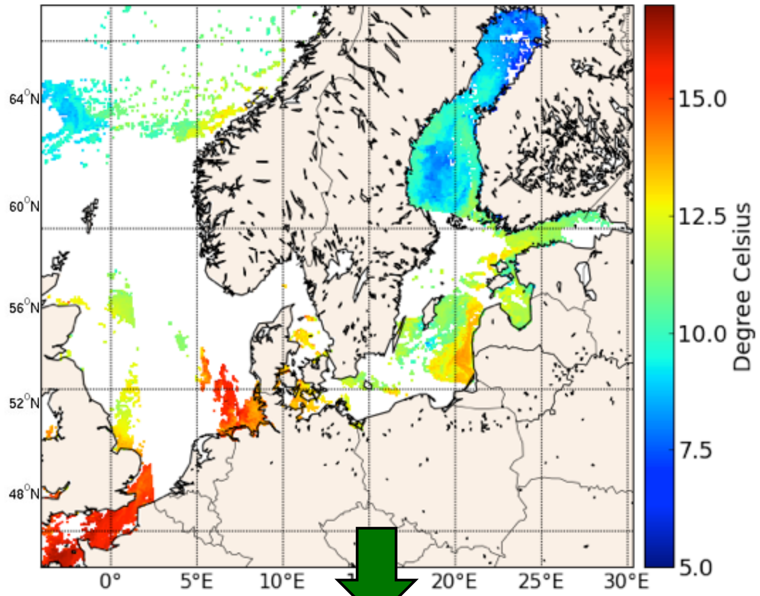
- Ensemble data assimilation
- Importance of software
- Coupled data assimilation
    - Challenges in two application examples
- Nonlinear filter developments

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

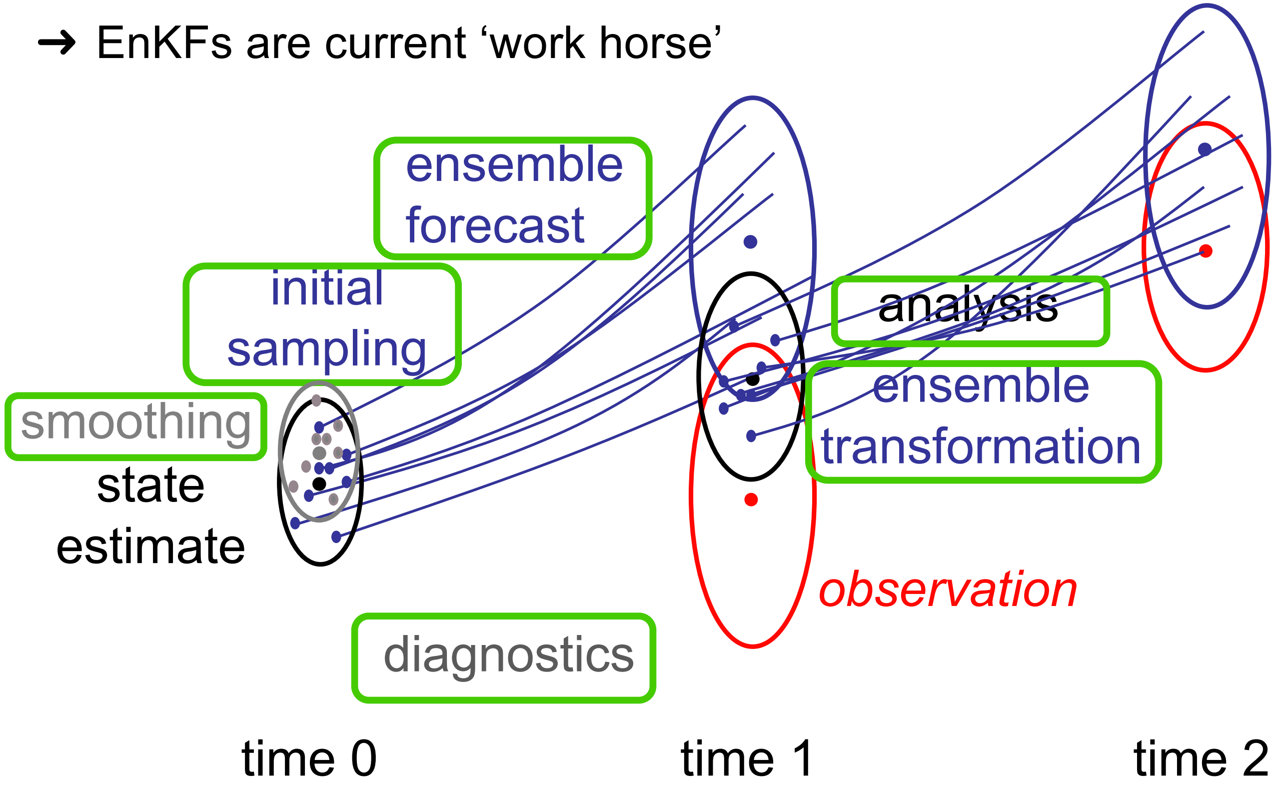
PDAF Parallel Data Assimilation Framework

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

PDAF: A tool for data assimilation

PDAF Parallel Data Assimilation Framework

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

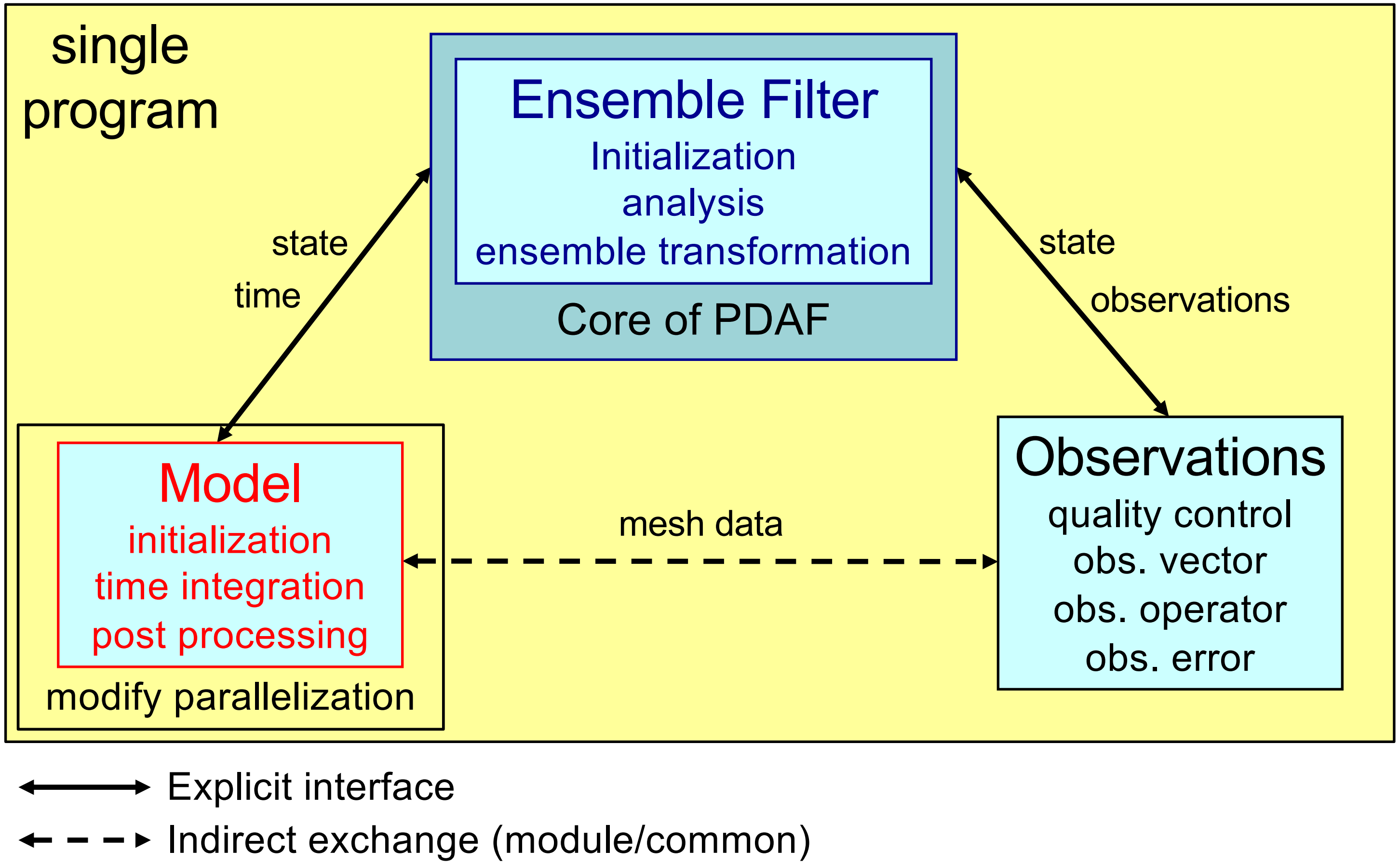
http://pdaf.awi.de

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

AWI

3 Components of Assimilation System

PDAF Parallel Data Assimilation Framework



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

AWI

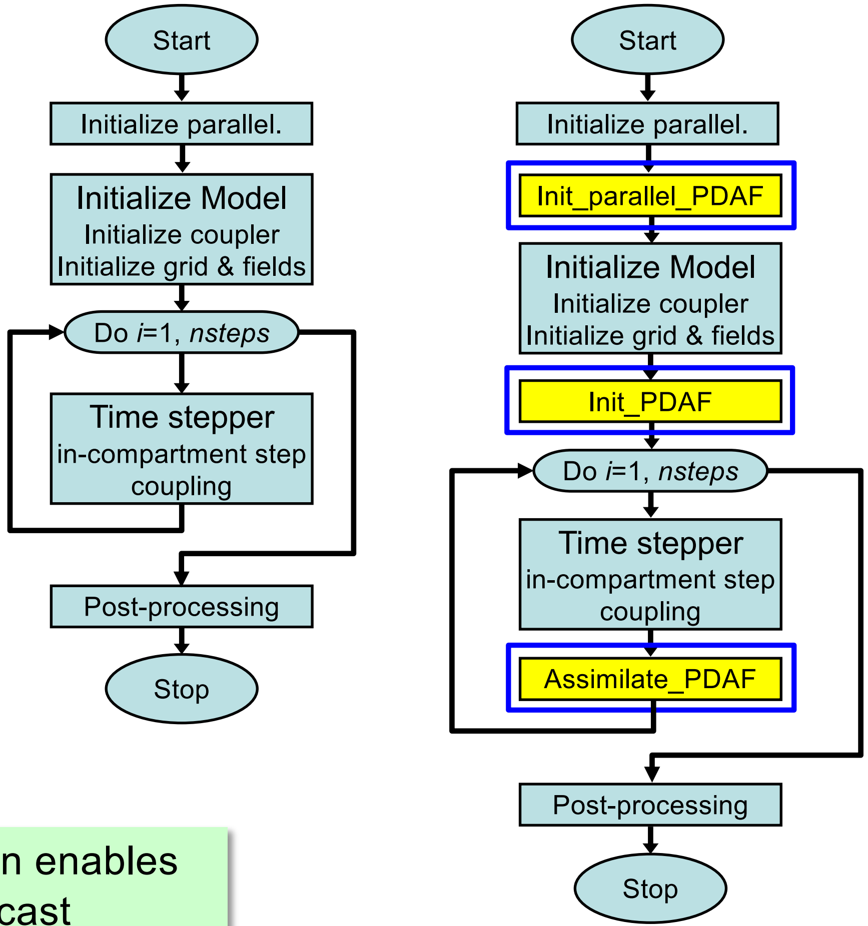
Augmenting a Model for Data Assimilation

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

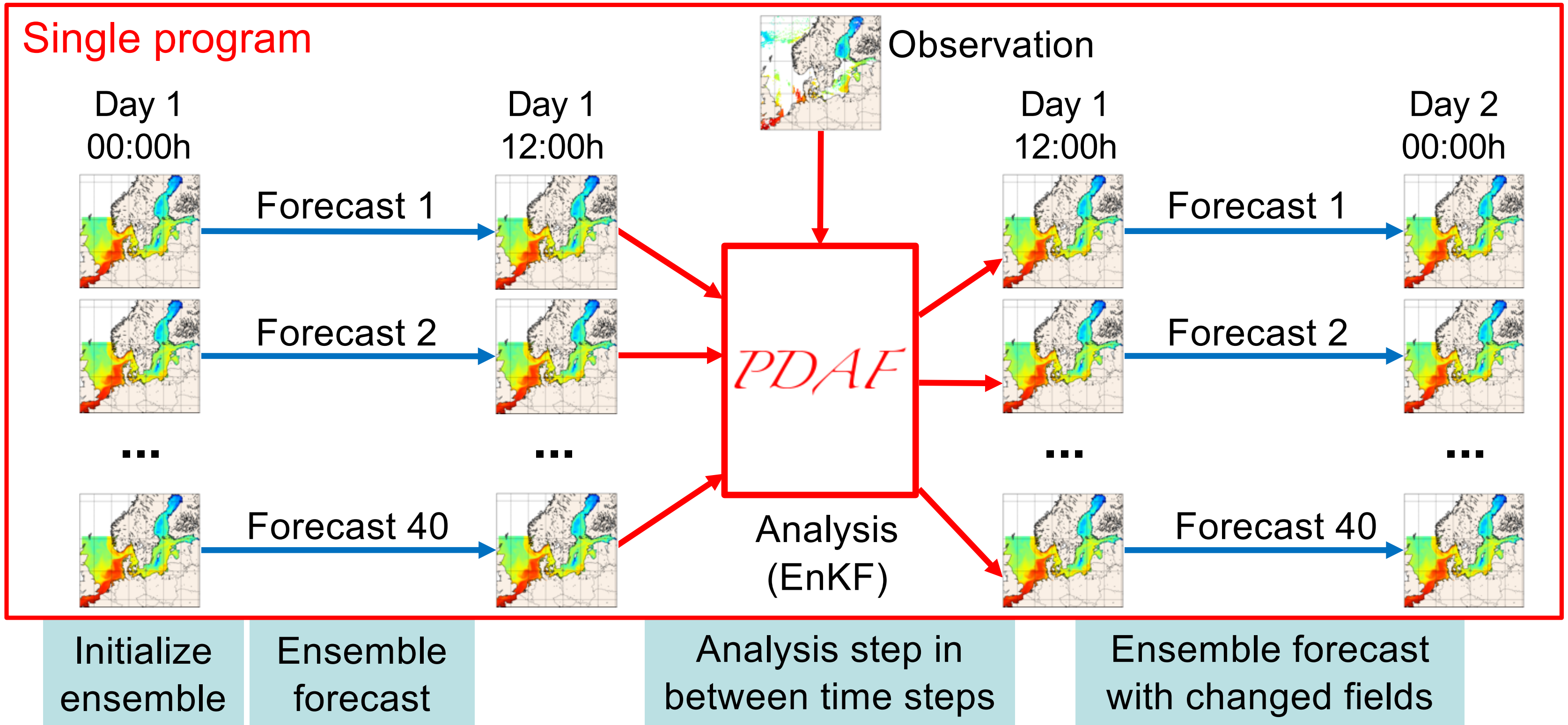
revised parallelization enables ensemble forecast

Augmenting a Model for Data Assimilation

PDAF Parallel Data Assimilation Framework

Couple PDAF with model

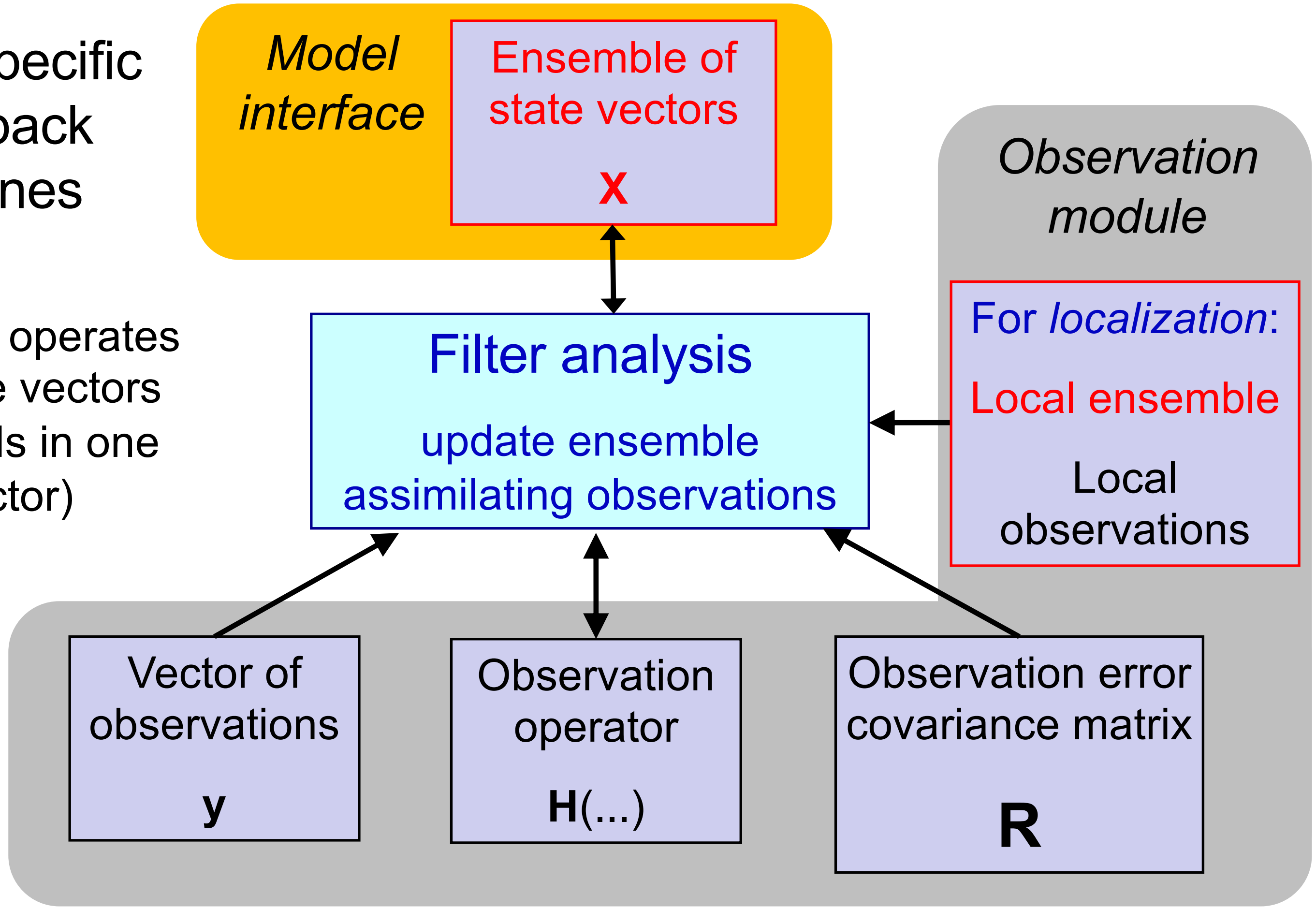
- Modify model to simulate ensemble of model states
- Insert correction step (analysis) to be executed at prescribed interval
- Run model as usual, but with more processors and additional options



Ensemble Filter Analysis Step

case-specific call-back routines

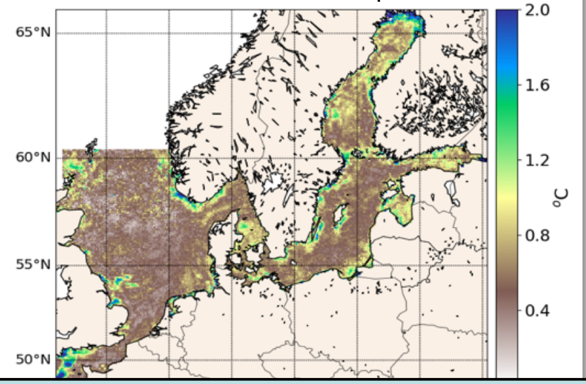
Analysis operates on state vectors (all fields in one vector)



PDAF Application Examples

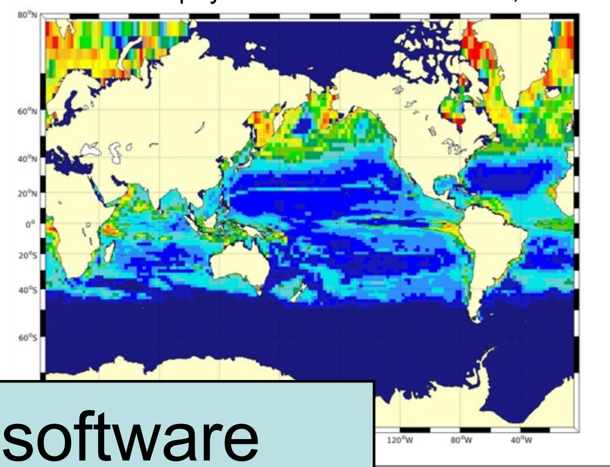
HBM-ERGOM: Coastal assimilation of SST, in situ and ocean color data (Svetlana Losa, Michael Goodliff)

RMS error in surface temperature



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

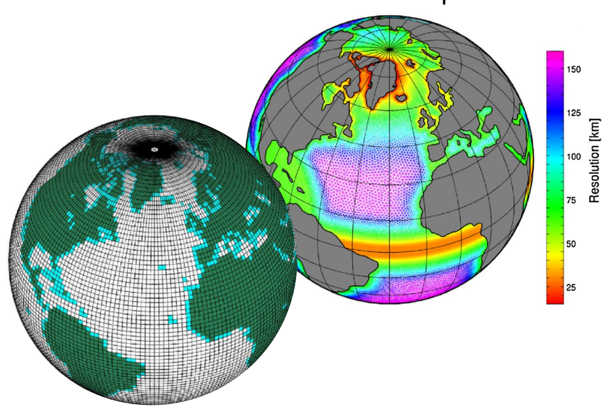
Total chlorophyll concentration June 30, 2012



Different models – same assimilation software

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

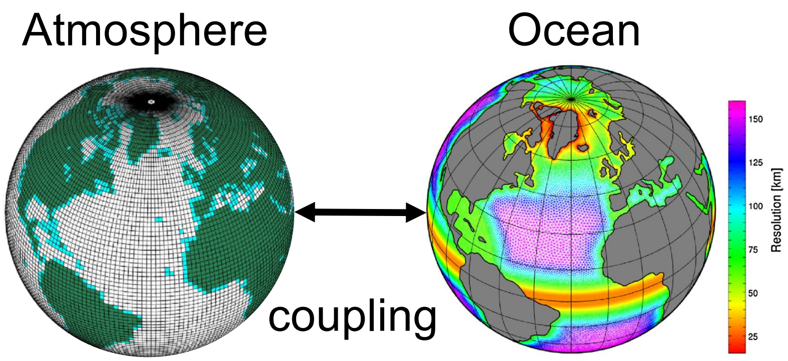
AWI-CM: ECHAM6-FESOM coupled model



+ 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 …

The Ensemble Kalman Filter (EnKF, Evensen 94)

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

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

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

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

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

- K contains
  - observed rows
  - unobserved rows

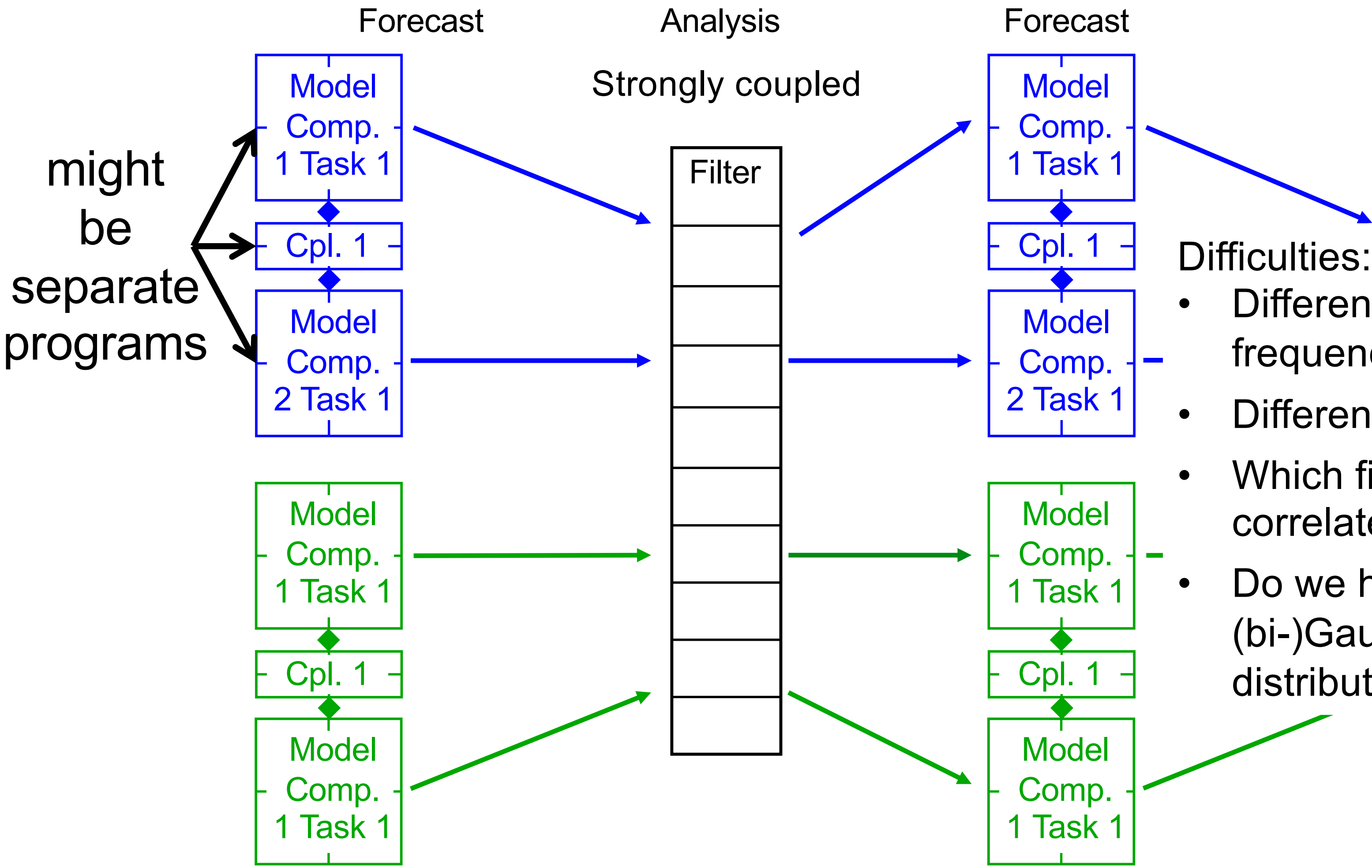
Unobserved variables updated through cross-covariances in $\mathbf{P}$ (linear regression)

Linear and Nonlinear Ensemble Filters

- Represent state and its error by ensemble $\mathbf{X}$ of N states
- **Forecast**:
  - Integrate ensemble with numerical model
- **Analysis step:**
  - update ensemble mean $\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}'^f \tilde{\mathbf{w}}$
  - update ensemble perturbations $\mathbf{X}'^a = \mathbf{X}'^f \mathbf{W}$

  (both can be combined in a single step)
- Ensemble Kalman & nonlinear filters: Different definitions of
  - weight vector $\tilde{\mathbf{w}}$ (dimension N)
  - Transform matrix $\mathbf{W}$ (dimension $N \times N$)

2 compartment system – strongly coupled DA

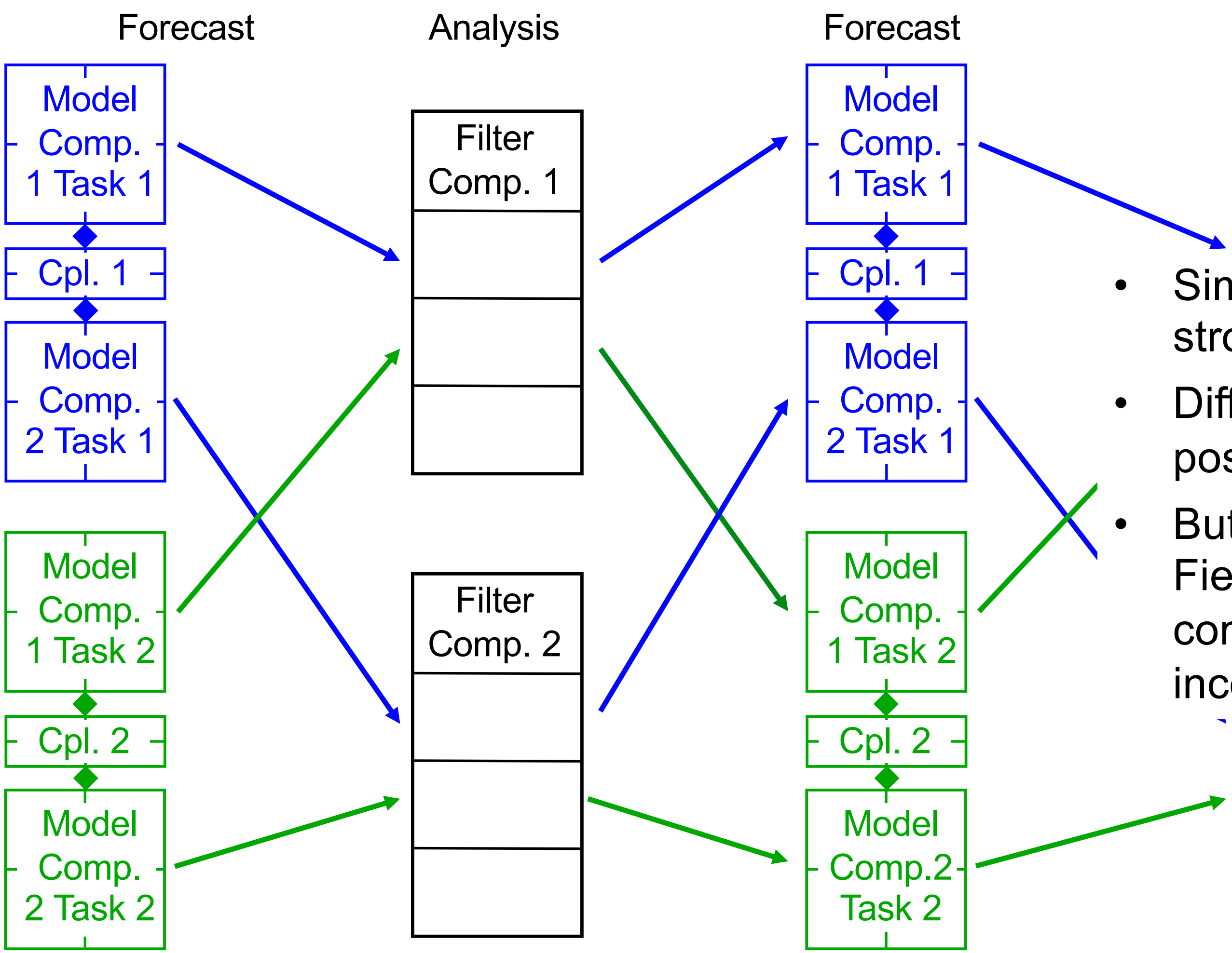


Difficulties:

- 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

AWI

Example 1

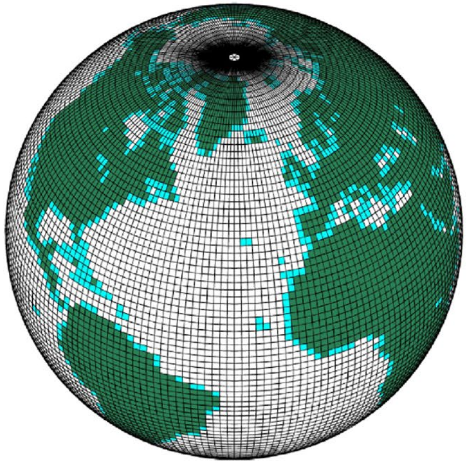
Assimilation into the coupled atmosphere-ocean model AWI-CM

(Qi Tang)

Project: ESM – Advanced Earth System Modeling Capacity

Assimilation into coupled model: AWI-CM

Atmosphere

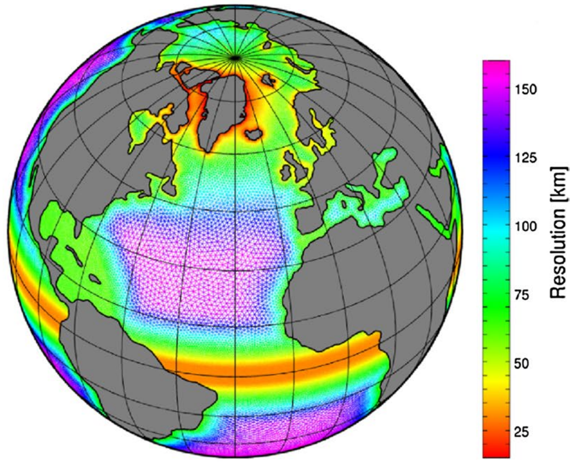


OASIS3-MCT

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

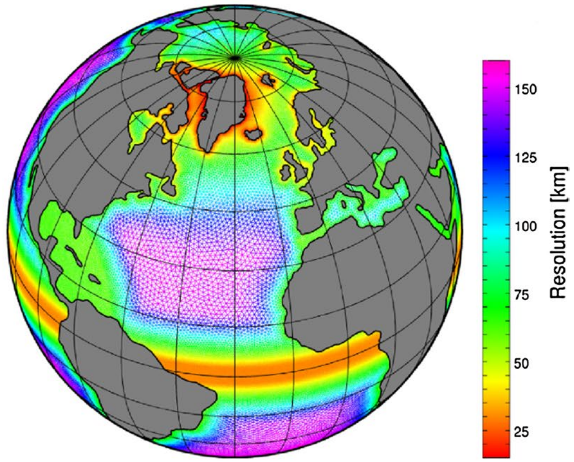
Model setup

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

Data assimilation experiments

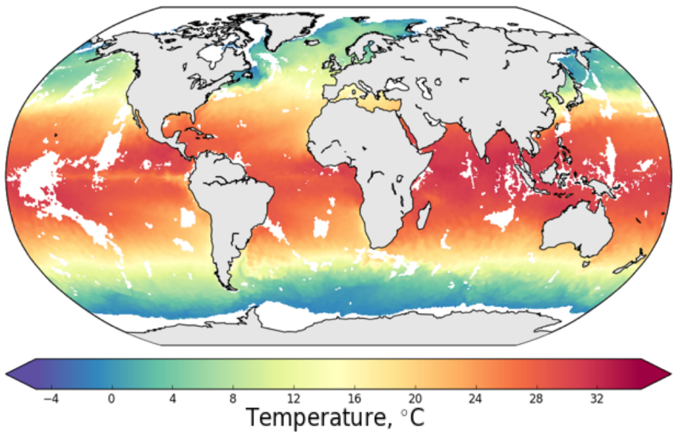
- Observations
  - Satellite SST
  - Profiles temperature & salinity
- Updated: ocean state (SSH, T, S, u, v, w)
- Assimilation method: Ensemble Kalman Filter (LESTKF)
- Ensemble size: 46
- Simulation period: year 2016, daily assimilation update
- Run time: 5.5h, fully parallelized using 12,000 processor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

FESOM mesh resolution

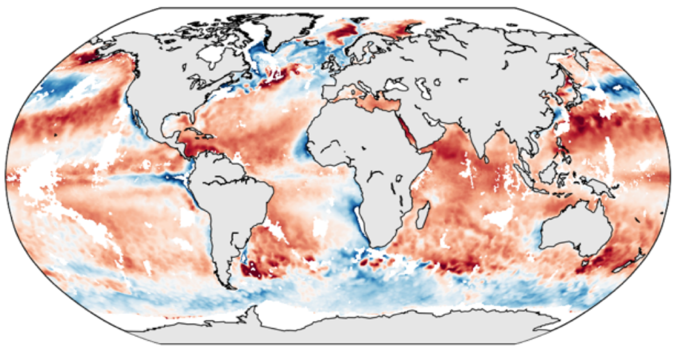


Assimilate sea surface temperature (SST)

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 ^{o}C
- Localization radius: 1000 km

Large initial SST deviation due to using a coupled model: up to 10^{o}C

DA with such a coupled model is unstable!

omit SST observations where

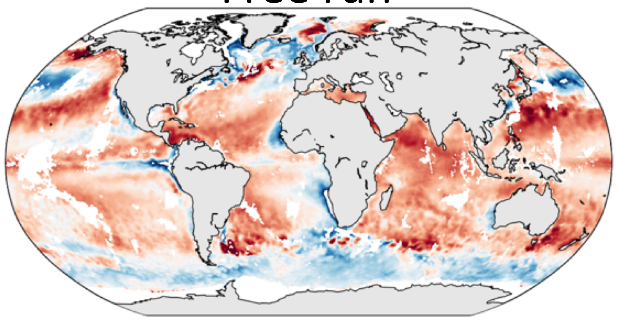
$|SST_{obs} - SST_{ens\_mean}| > 1.6\ ^oC$

(30% initially, <5% later)

SST assimilation: Effect on the ocean

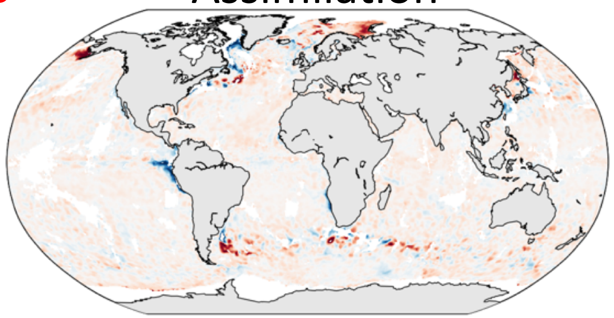
SST difference (obs-model): strong decrease of deviation

Free run



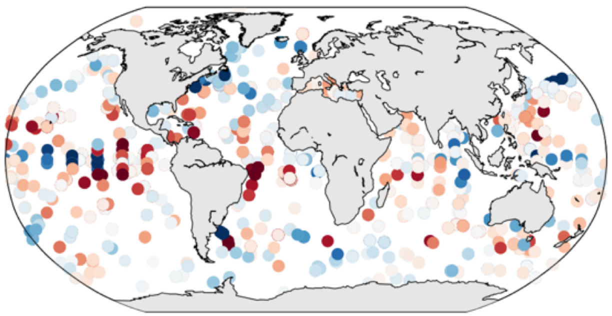
4/30/2016
Day 120

Assimilation



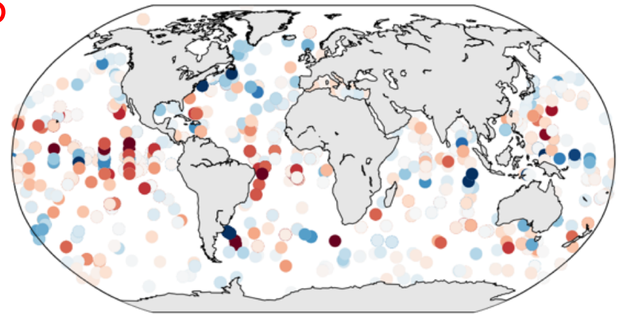
Subsurface temperature difference (obs-model); all the model layers at profile locations

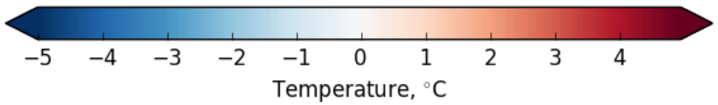
Free run



4/30/2016
Day 120

Assimilation



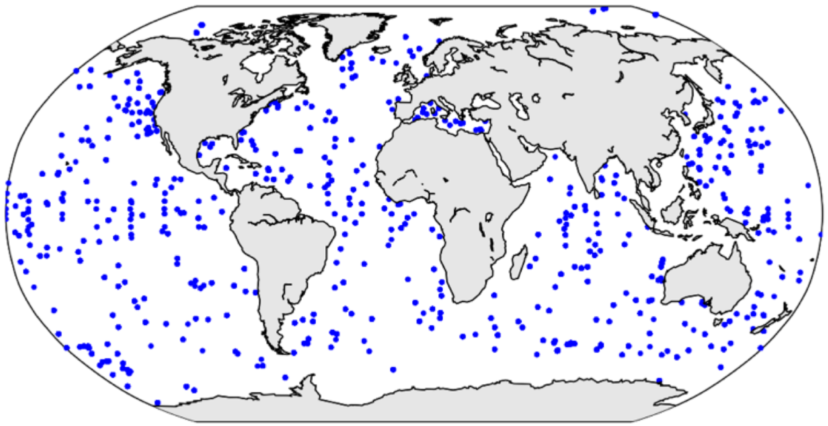


ESM Advanced Earth System Modelling Capacity

AWI

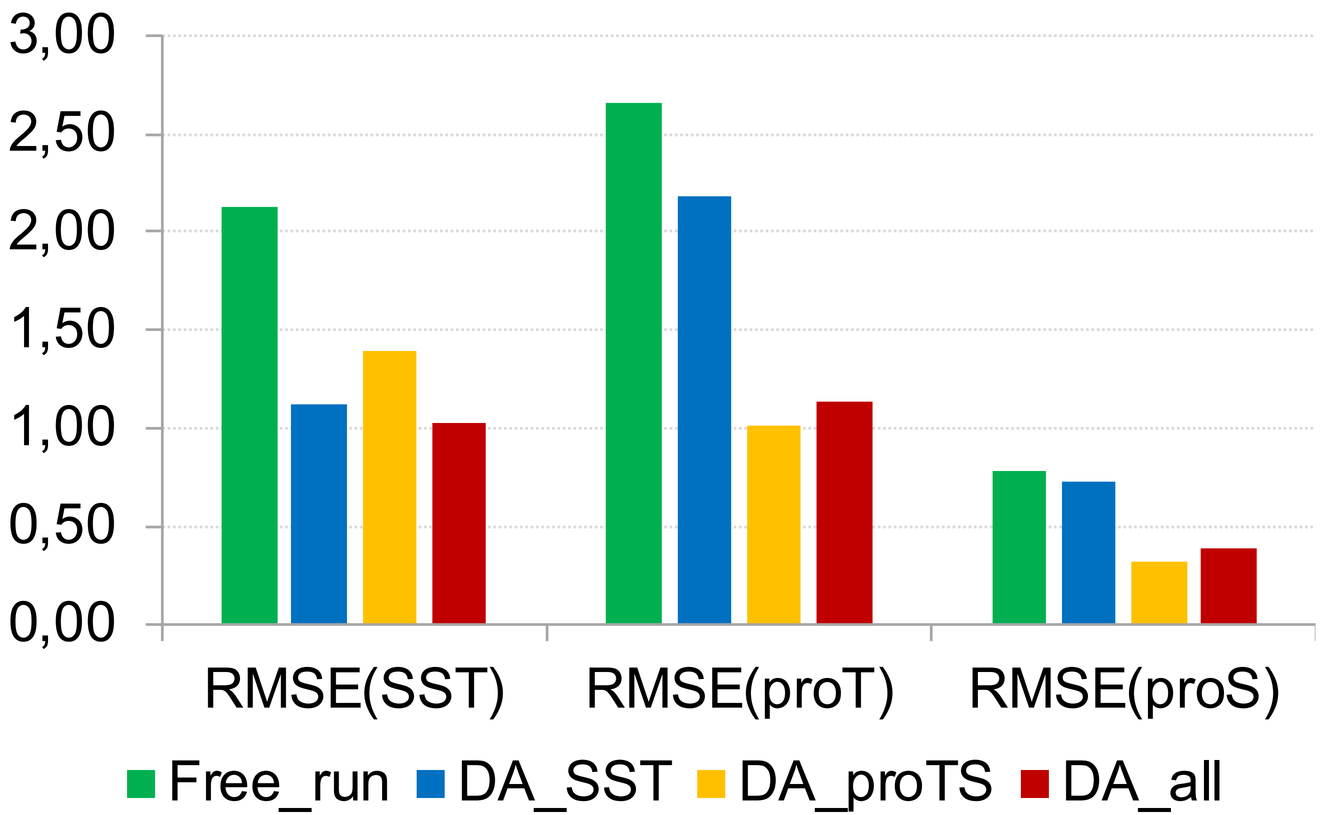
Assimilate subsurface observations: Profiles

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 ^{o}C
  - Salinity profiles: 0.5 psu
- Localization radius: 1000 km

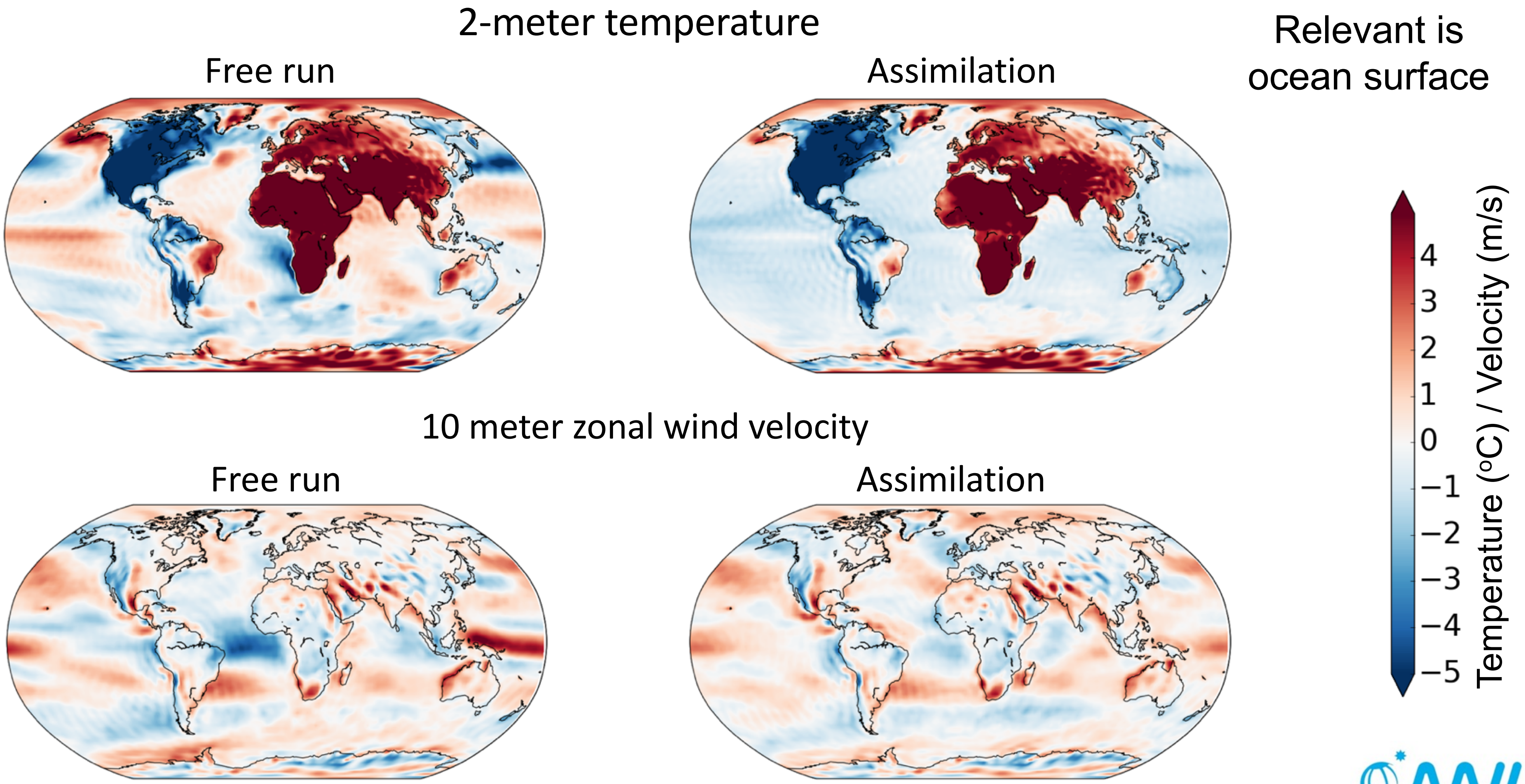
Assimilation effect: RMS errors



Overall lowest errors with combined assimilation

- But partly a compromise

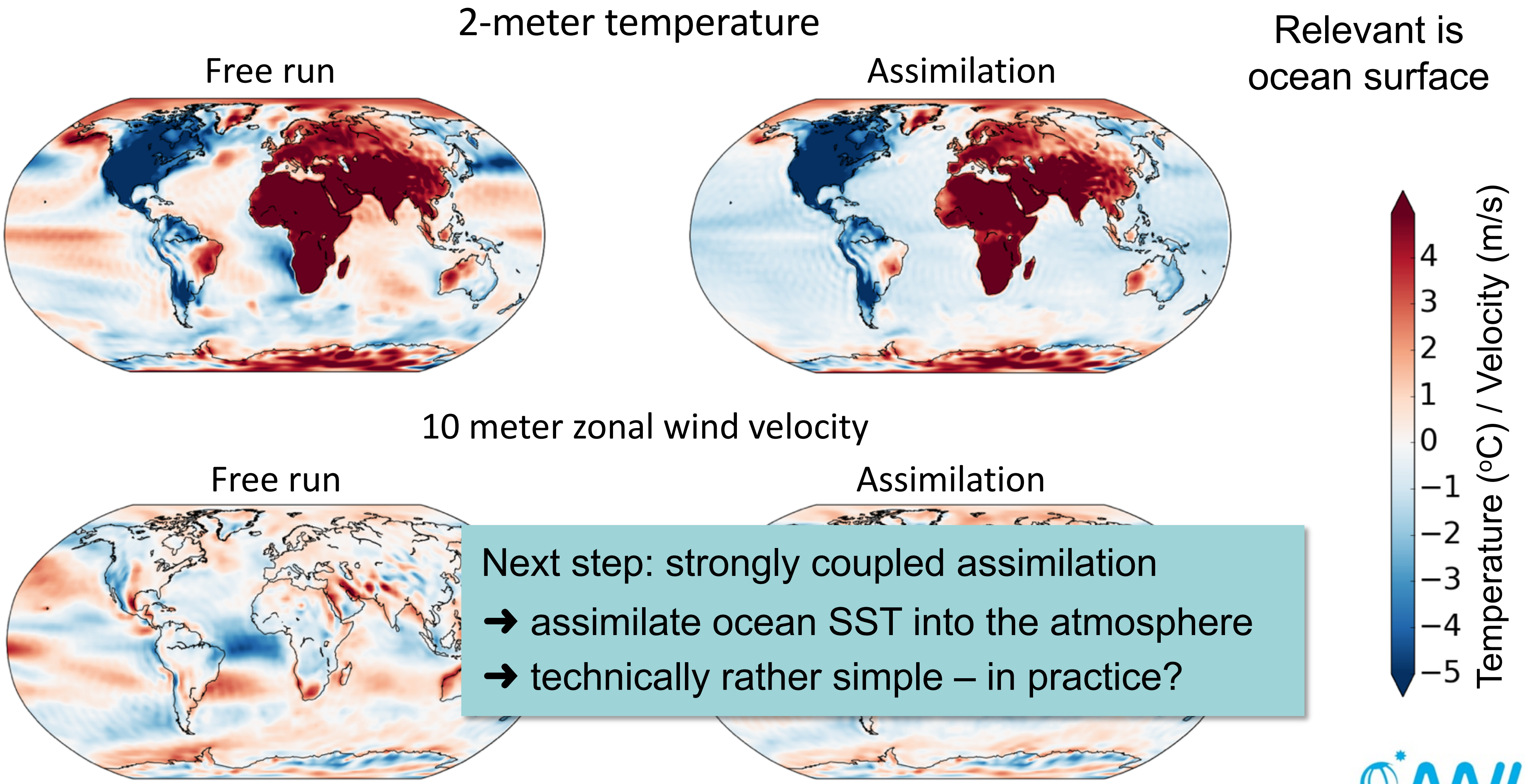
Effect on Atmospheric State (annual mean)



Relevant is ocean surface

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?

Example 2

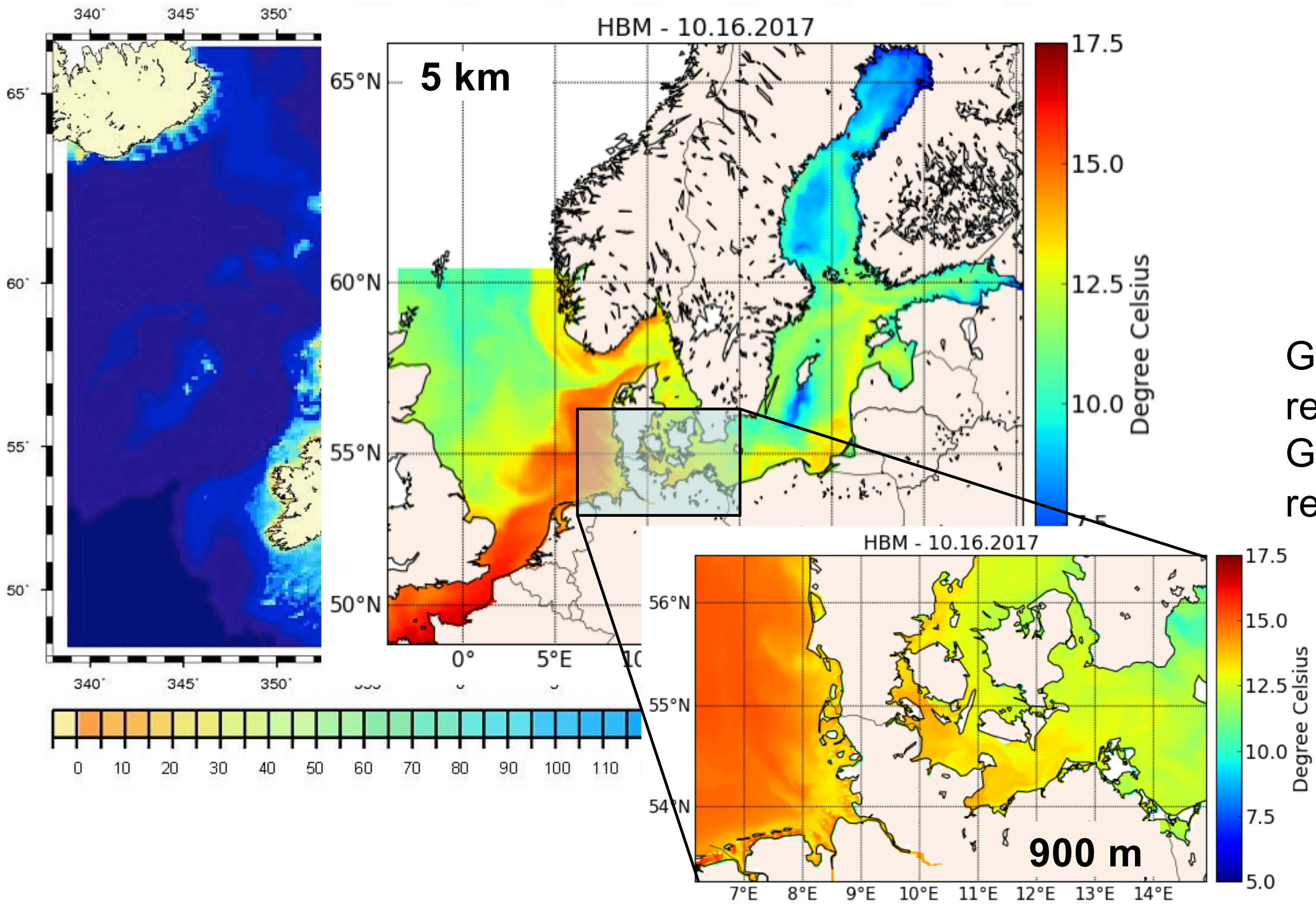
Weakly- and Strongly Coupled Assimilation to Constrain Biogeochemistry with Temperature Data

(MERAMO – Mike Goodliff)

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

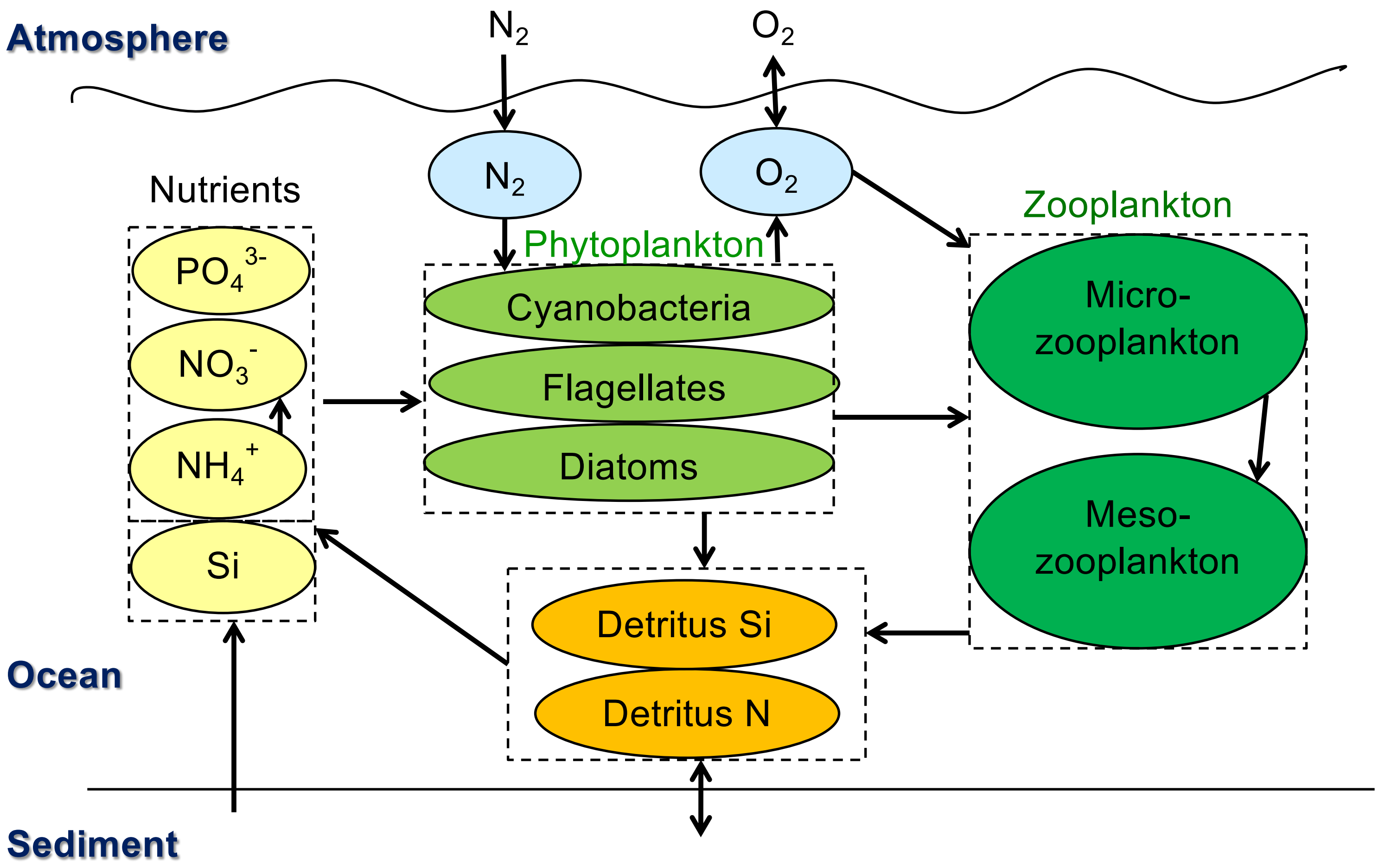
Coastal Model Domain

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



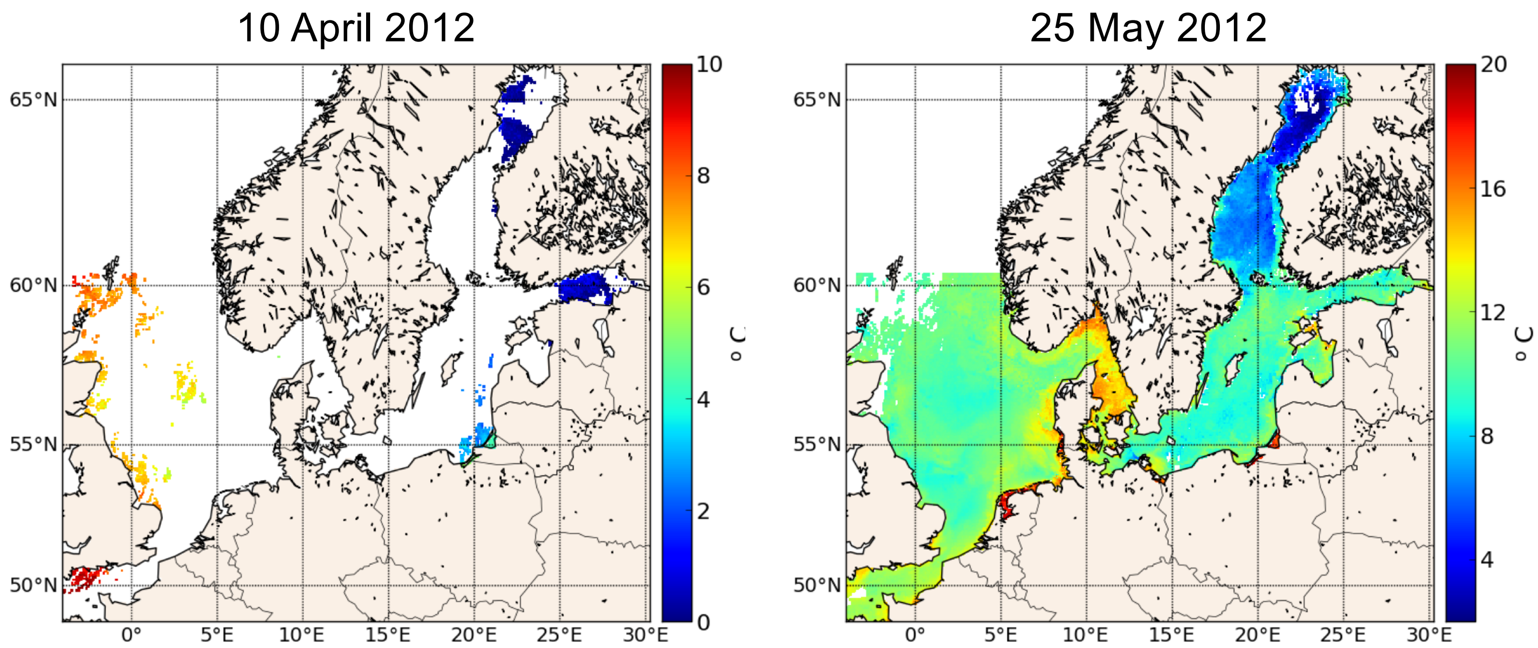
Grid with higher resolution in German coastal region

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

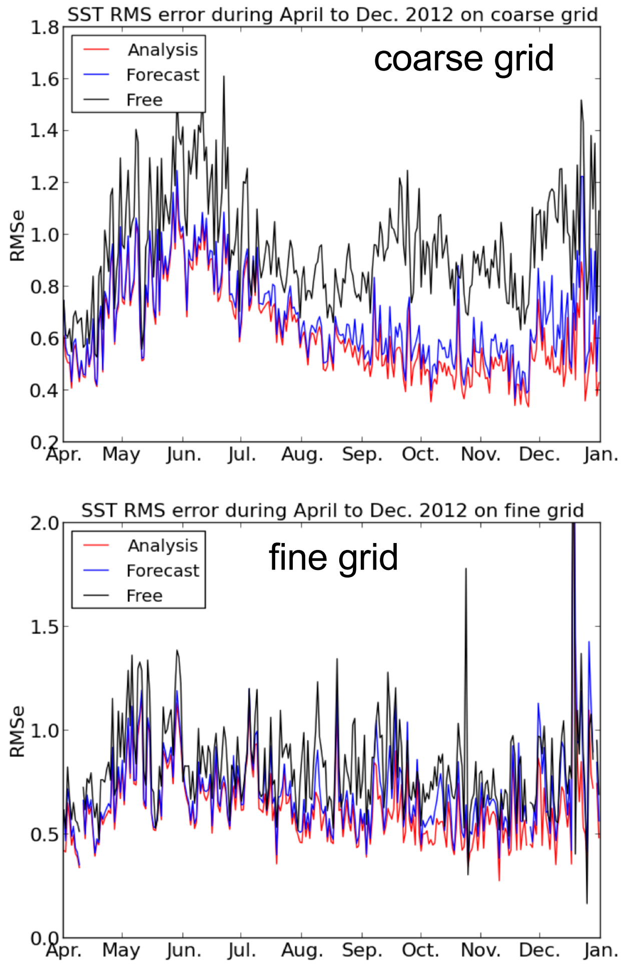
Coarse grid:

- Increasing error-reductions compared to free ensemble run

Fine grid:

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

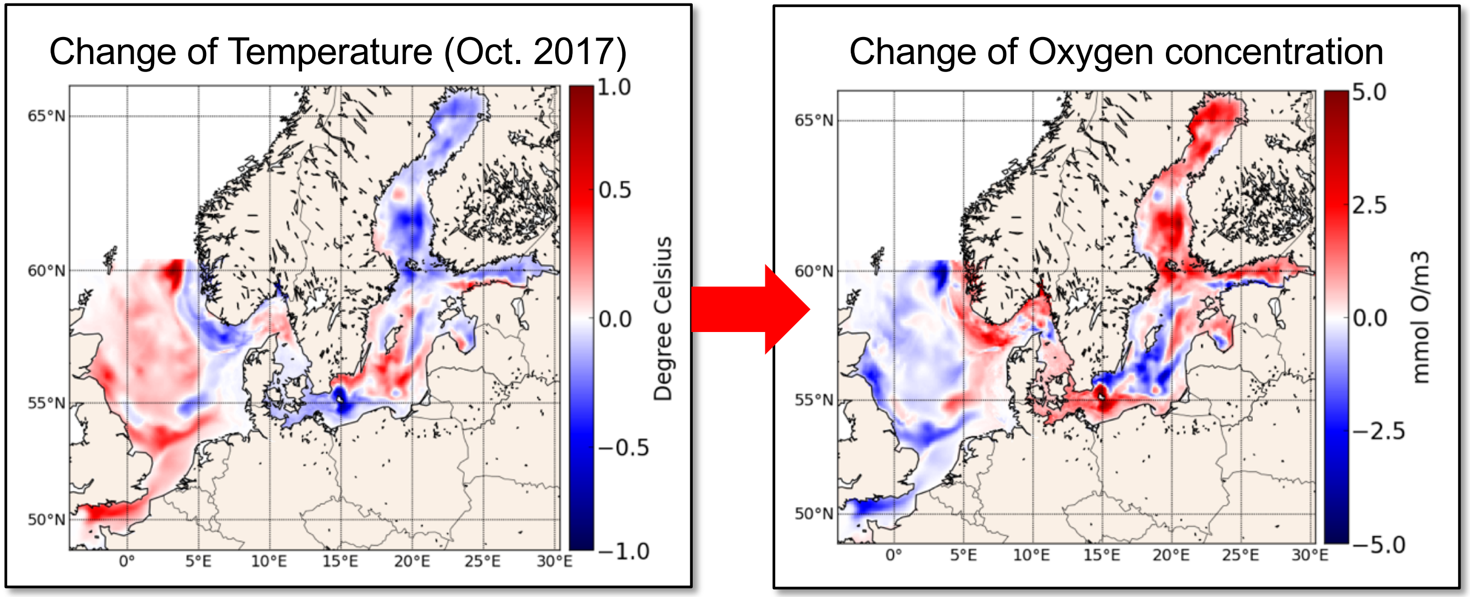
Temperature RMSD



RMS errors (deg. C)

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

Influence of Assimilation on Surface Temperature

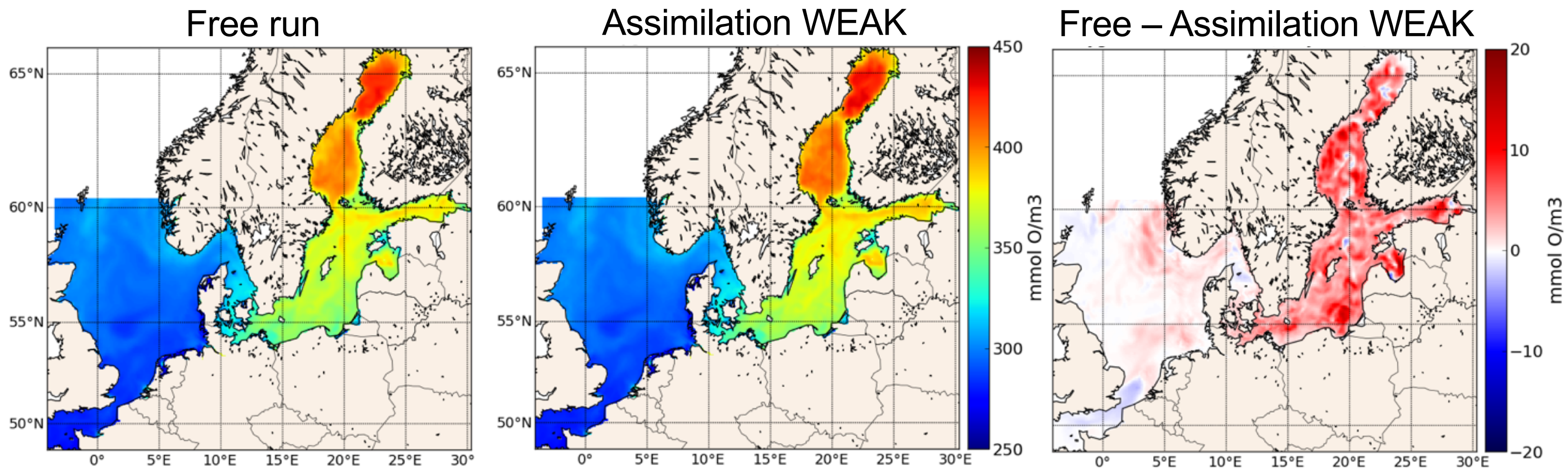


2 ways of influence:

- Indirect - *weakly-coupled assimilation*
  model dynamics react on change in physics
- Direct – *strongly-coupled assimilation*
  use cross-covariances between surface temperature and biogeochemistry

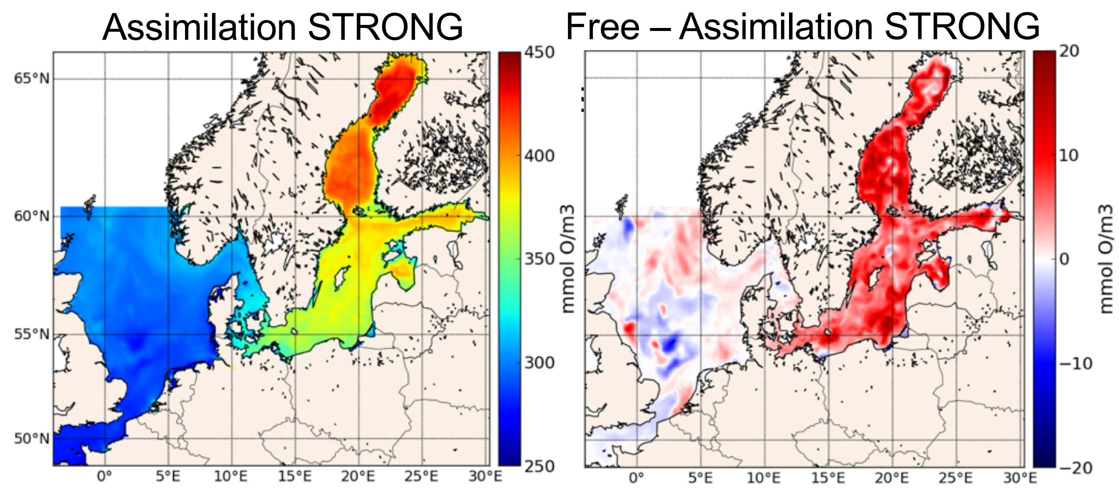
Weakly & strongly coupled effect on biogeochemical model

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

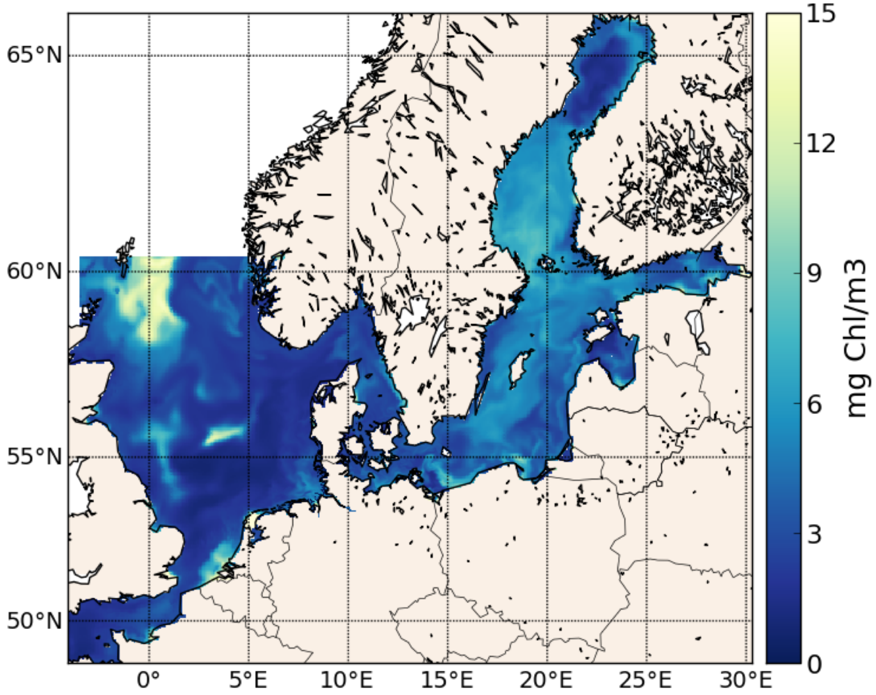
- Chlorophyll is lognormally distributed
- Ensemble Kalman filter
  - Optimality for normal distributions
  - Linear regression between observed and unobserved variables

➜ Apply strongly-coupled DA with logarithm on concentrations?

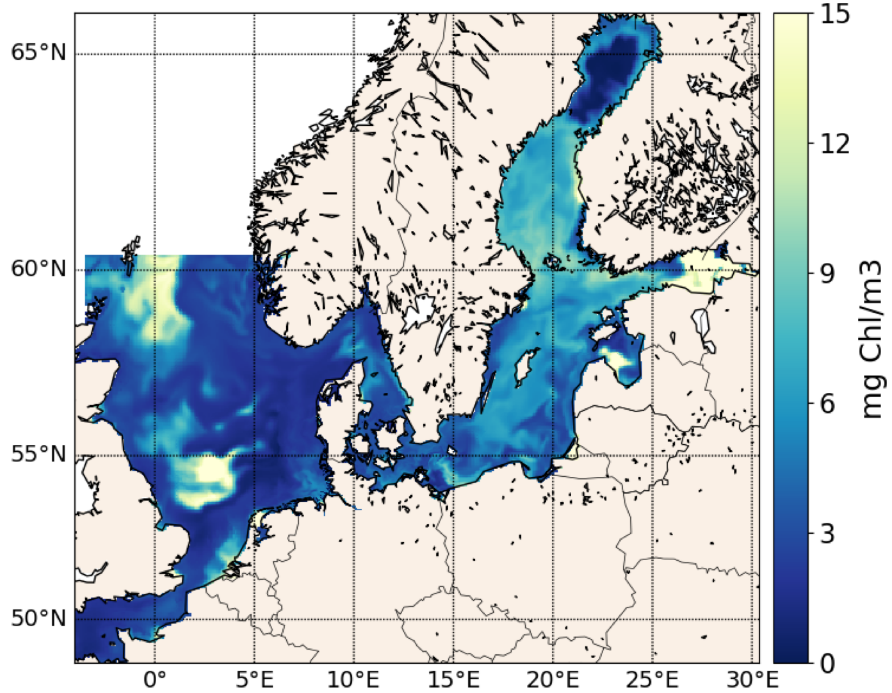
Choice of variable in strongly coupled assimilation

Chlorophyll concentrations
1 May 2012

Weakly coupled



Strongly coupled
linear



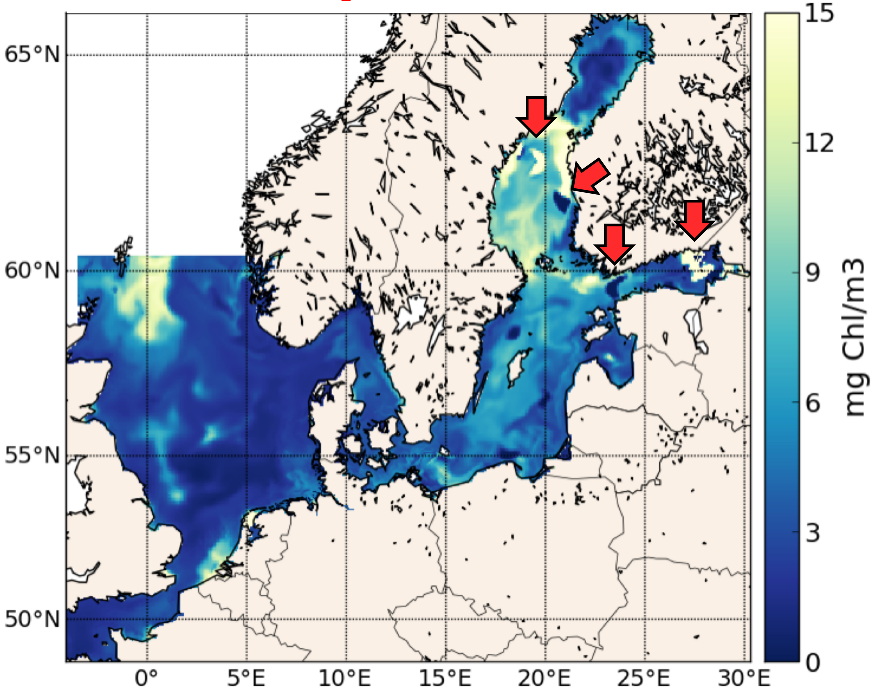
- Larger effect – in particular in North Sea
- Too high in Gulf of Finland

→ Particle filter might help

Choice of variable in strongly coupled assimilation

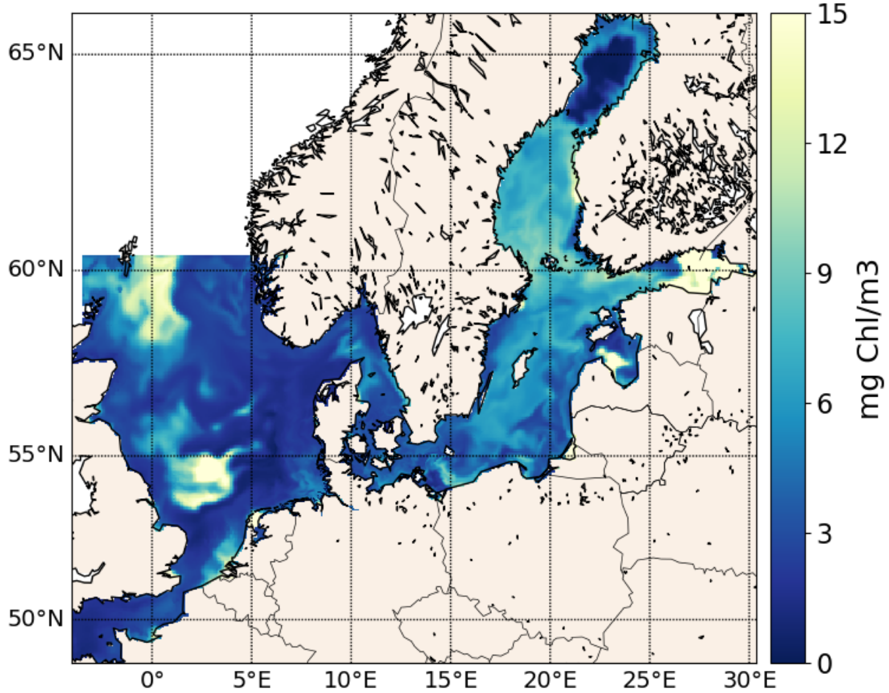
Chlorophyll concentrations
1 May 2012

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

➜ Particle filter might help

Toward usable nonlinear filters:

Hybrid nonlinear-Kalman ensemble filters

Linear and Nonlinear Ensemble Filters

- Represent state and its error by ensemble $\mathbf{X}$ of N states
- **Forecast**:
  - Integrate ensemble with numerical model
- **Analysis step:**
  - update ensemble mean $\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}'^f \tilde{\mathbf{w}}$
  - update ensemble perturbations $\mathbf{X}'^a = \mathbf{X}'^f \mathbf{W}$

  (both can be combined in a single step)

- Ensemble Kalman & nonlinear filters: Different definitions of
  - weight vector $\tilde{\mathbf{w}}$ (dimension N)
  - Transform matrix $\mathbf{W}$ (dimension $N \times N$)

ETKF (Bishop et al., 2001)

- Ensemble Transform Kalman filter
  - Assume Gaussian distributions
  - Transform matrix

$$\mathbf{A}^{-1} = (N-1)\mathbf{I} + (\mathbf{H}\mathbf{X}'^f)^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}'^f$$

  - Mean update weight vector

$$\tilde{\mathbf{w}} = \mathbf{A}(\mathbf{H}\mathbf{X}'^f)^T\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}^f}\right)$$

(depends linearly on $\mathbf{y}$)

  - Transformation of ensemble perturbations

$$\mathbf{W} = \sqrt{(N-1)}\mathbf{A}^{-1/2}\mathbf{\Lambda}$$

$\mathbf{\Lambda}$: mean-preserving random matrix or identity

($\mathbf{W}$ depends only on $\mathbf{R}$, not $\mathbf{y}$)

NETF (Tödter & Ahrens, 2015)

- Nonlinear Ensemble Transform Filter
  - Mean update from Particle Filter weights:
    for Gaussian observation errors for all particles *i*

$$\tilde{w}^i \sim \exp\left(-0.5(\mathbf{y} - \mathbf{H}\mathbf{x}_i^f)^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_i^f)\right)$$

    (nonlinear function of observations $\mathbf{y}$)

  - Ensemble update
    - Transform ensemble to fulfill analysis covariance
      (like ETKF, but not assuming Gaussianity)
    - Derivation gives

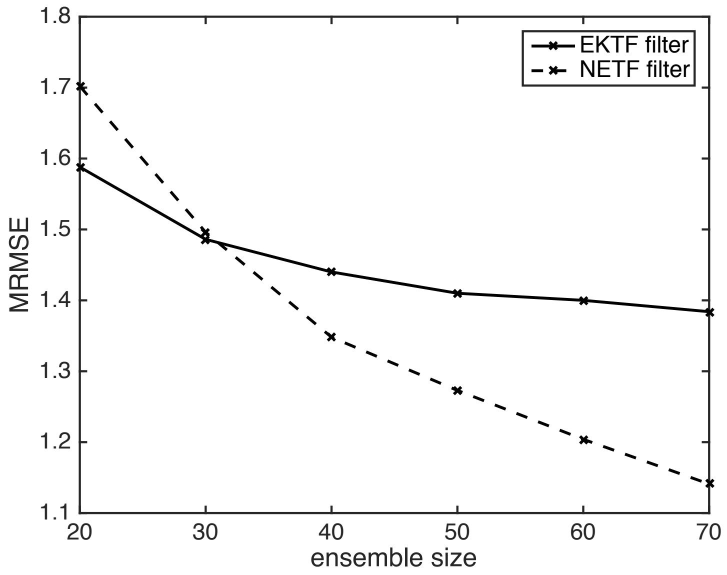
$$\mathbf{W} = \sqrt{N}\left[\mathrm{diag}(\tilde{\mathbf{w}}) - \tilde{\mathbf{w}}\tilde{\mathbf{w}}^T\right]^{1/2}\Lambda$$

      ($\boldsymbol{\Lambda}$: mean-preserving random matrix; useful for stability)

Tödter, J. and Ahrens, B. (2015) *Mon. Wea. Rev.* **143**,1347–1367

Performance of NETF – Lorenz-96

- Double-exponential observation errors
- Run all experiments 10x with different initial ensemble



- NETF beats ETKF for ensemble size N > 30
- Larger ensemble needed for Gaussian errors

Kirchgessner, Tödter, Ahrens, Nerger. (2017) *Tellus A*, **69**, 1327766

ETKF-NETF – Hybrid Filter Variants

1-step update (*HSync*)

$$\mathbf{X}^a_{HSync} = \overline{\mathbf{X}}^f + (1-\gamma)\Delta\mathbf{X}_{NETF} + \gamma\Delta\mathbf{X}_{ETKF}$$

- $\Delta\mathbf{X}$: assimilation increment of a filter
- γ: hybrid weight (between 0 and 1; 1 for fully ETKF)

2-step updates

Variant 1 (*HNK*): NETF followed by ETKF

$$\tilde{\mathbf{X}}^a_{HNK} = \mathbf{X}^a_{NETF}[\mathbf{X}^f, (1-\gamma)\mathbf{R}^{-1}]$$
$$\mathbf{X}^a_{HNK} = \mathbf{X}^a_{ETKF}[\tilde{\mathbf{X}}^a_{HNK}, \gamma\mathbf{R}^{-1}]$$

- Both steps computed with increased $\mathbf{R}$ according to γ

Variant 2 (*HKN*): ETKF followed by NETF

Choosing hybrid weight γ

- Hybrid weight shifts filter behavior

Some possibilities:

- Fixed value
- Adaptive
  - According to which condition?

  - For hybrid particle-EnKF, Frei & Kuensch (2013) suggested using effective sample size $N_{eff} = \sum_i 1/(w^i)^2$
    - Choose γ so that N_{eff} is as small as possible but above minimum limit α

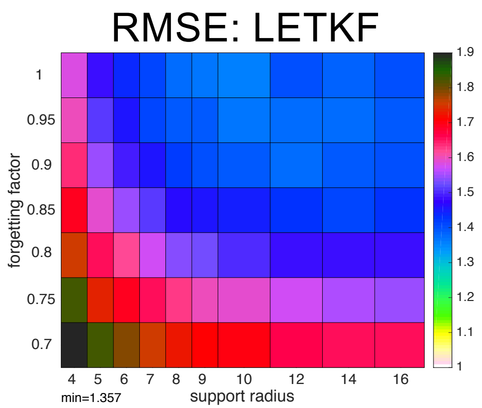
  - Adaptive alternatives

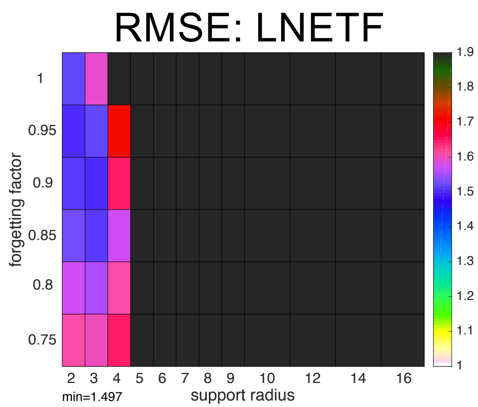
    $$\gamma_{adap} = 1 - N_{eff}/N_e \qquad \gamma_{adap} = \sqrt{1 - N_{eff}/N_e}$$

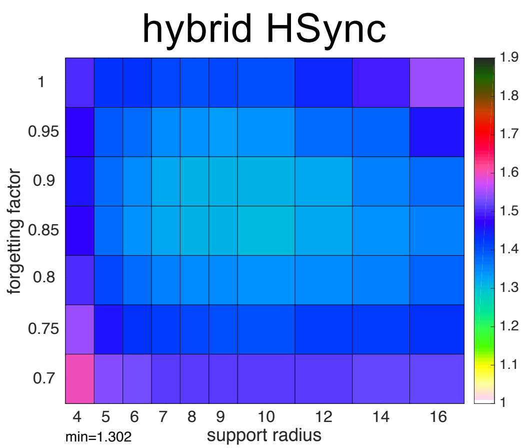
    (close to 1 if N_{eff} small)

Test with Lorenz-96 model (dimension=80)

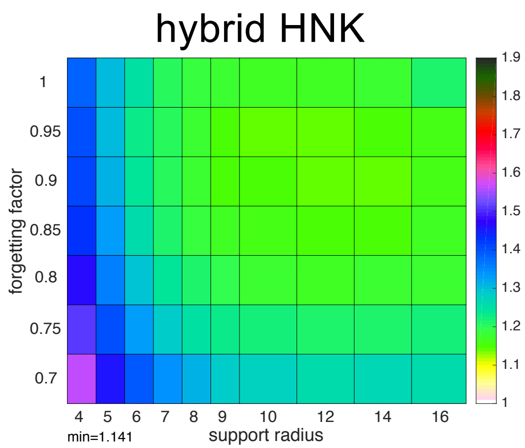
Ensemble size N=50



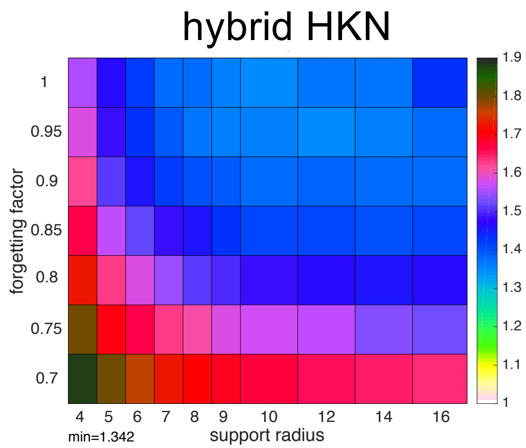




4% improvement



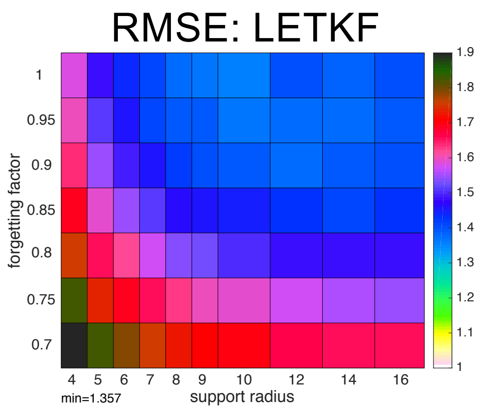
16% improvement



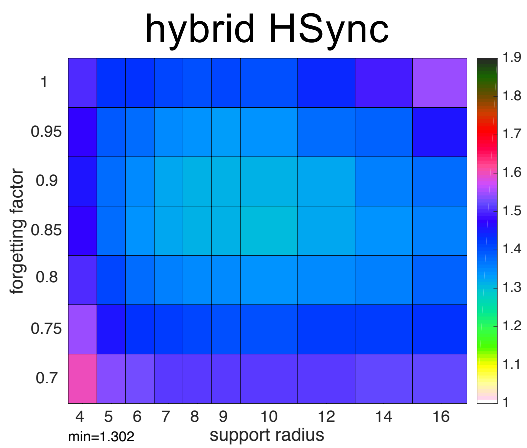
1% improvement

Test with Lorenz-96 model (dimension=80)

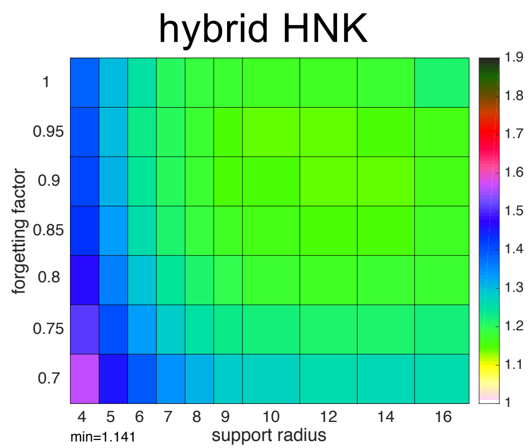
Ensemble size N=50



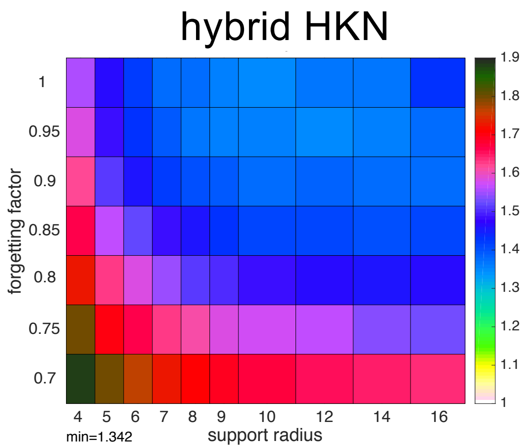
- All hybrid variants improve estimates compared to LETKF & NETF
- Dependence on forgetting factor & localization radius like LETKF
- Similar optimal localization radius
- Largest improvement for variant HNK (NETF before LETKF)



4% improvement



16% improvement



1% improvement

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
- Hybrid nonlinear-linear filters promise to improve estimates while being applicable
- Unified software helps to bring new developments into usage
  - PDAF – Open source available at **http://pdaf.awi.de**

Lars.Nerger@awi.de