Efficient Ensemble Data Assimilation For Earth System Models with the Parallel Data Assimilation Framework (PDAF)

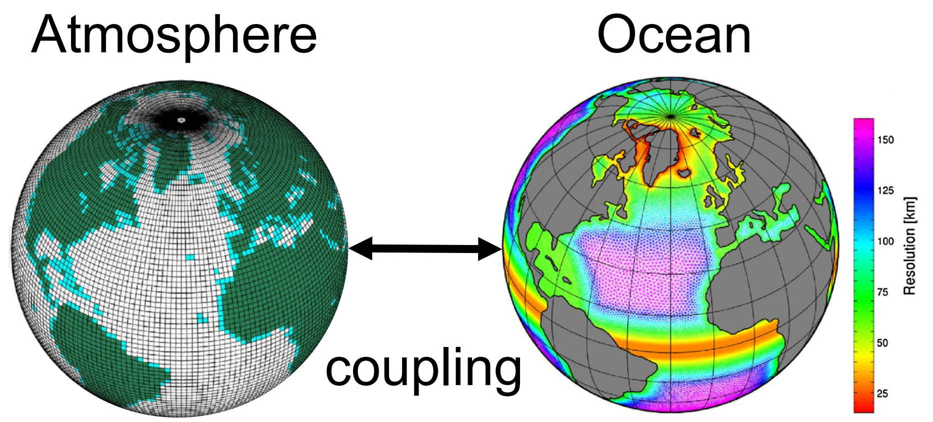
Lars Nerger, Qi Tang, Longjiang Mu, Dmitry Sidorenko

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

Overview

- Coupled Data Assimilation
- PDAF – Parallel Data Assimilation Framework
- Combining coupled model and PDAF
- Example: AWI Climate Model (ECHAM6 & FESOM)

Coupled Models and Coupled Data Assimilation



Coupled models

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

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 …

Ensemble Data Assimilation

Ensemble Kalman Filters & Particle Filters

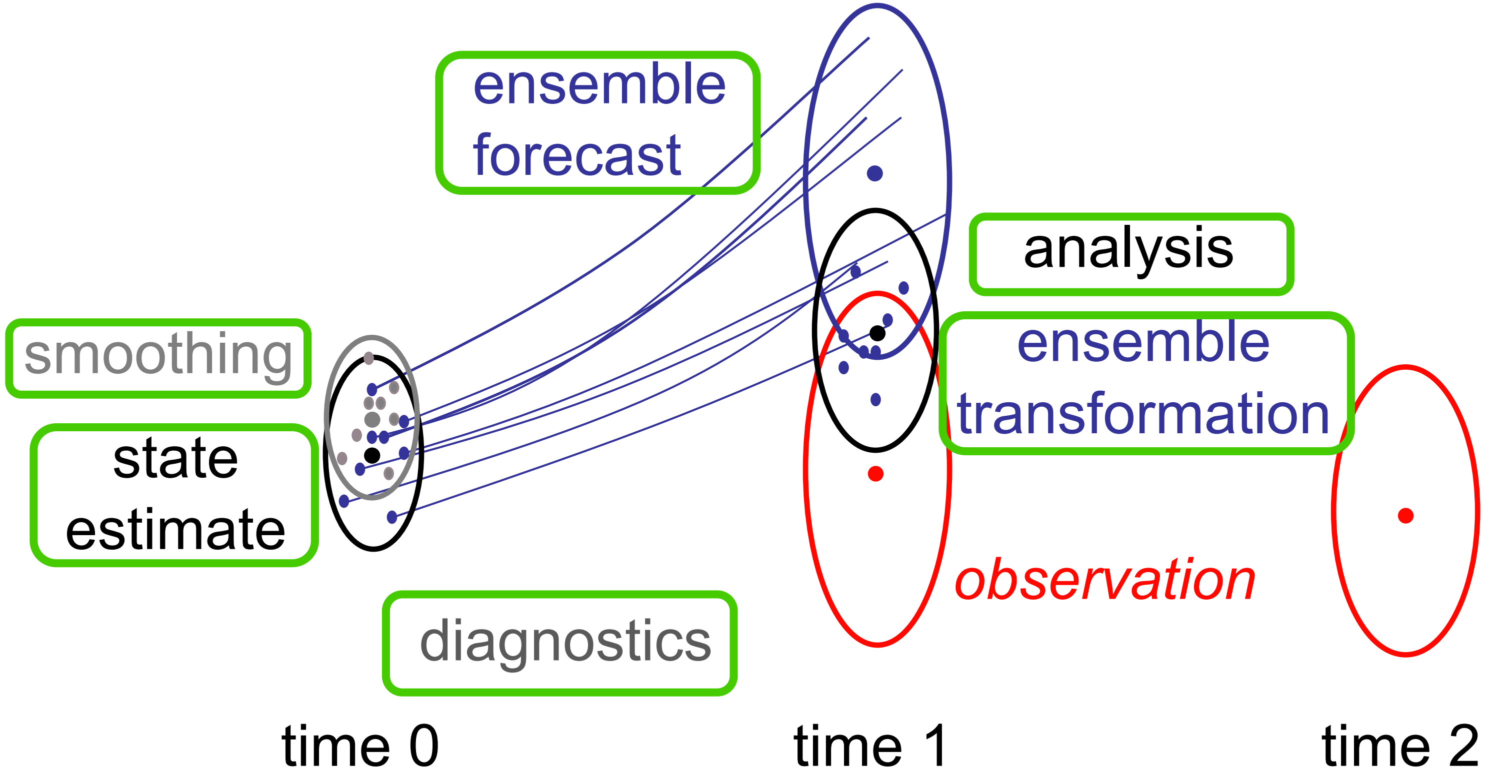
→ Use ensembles to represent state and uncertainty

→ Propagate ensemble using numerical model

→ Use observations to update ensemble

→ EnKFs are current 'work horse'

PDAF provides methods for each of the steps



Nerger and Hiller, Comp & Geosci., 2013

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
  (coupled to e.g. NEMO, MITgcm, FESOM, HBM, MPI-ESM, SCHISM/ESMF)
- run from laptops to supercomputers (Fortran, MPI & OpenMP)
- Usable for real assimilation applications and to study assimilation methods
- ~470 registered users; community contributions

Open source:
Code, documentation, and tutorial available at

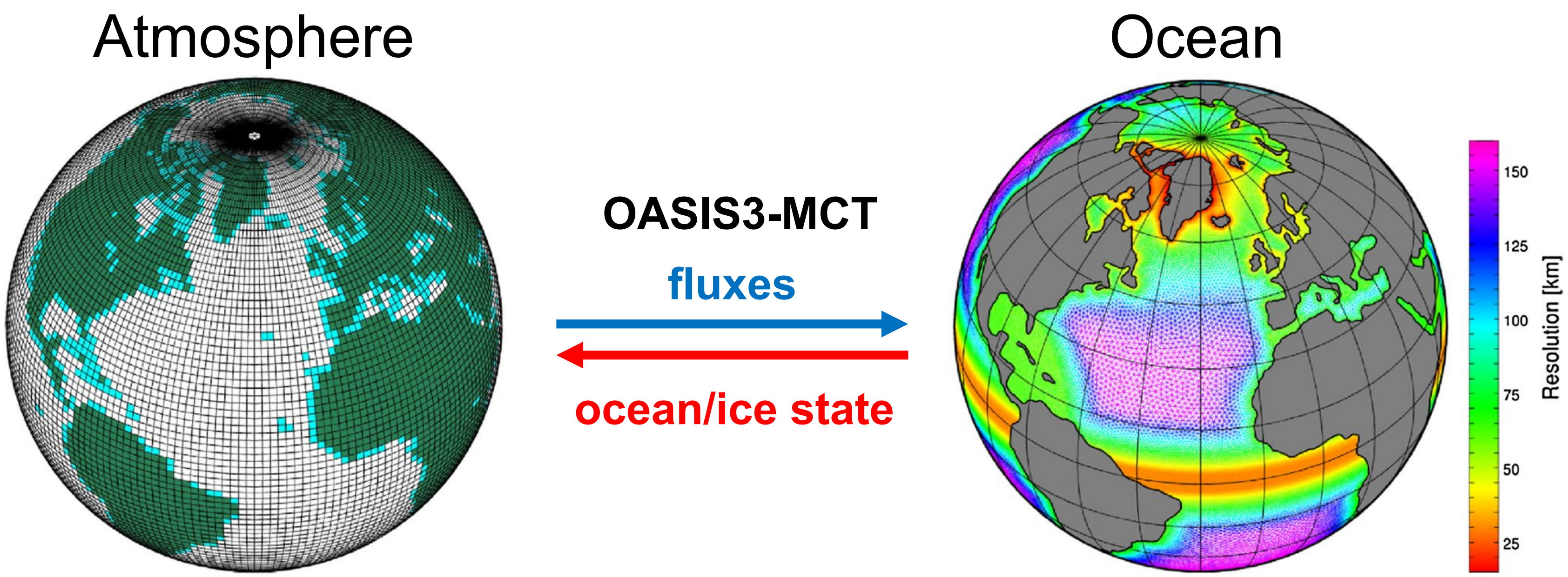
http://pdaf.awi.de

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

AWI

Combining coupled model and PDAF

Example for assimilation into coupled model: AWI-CM



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

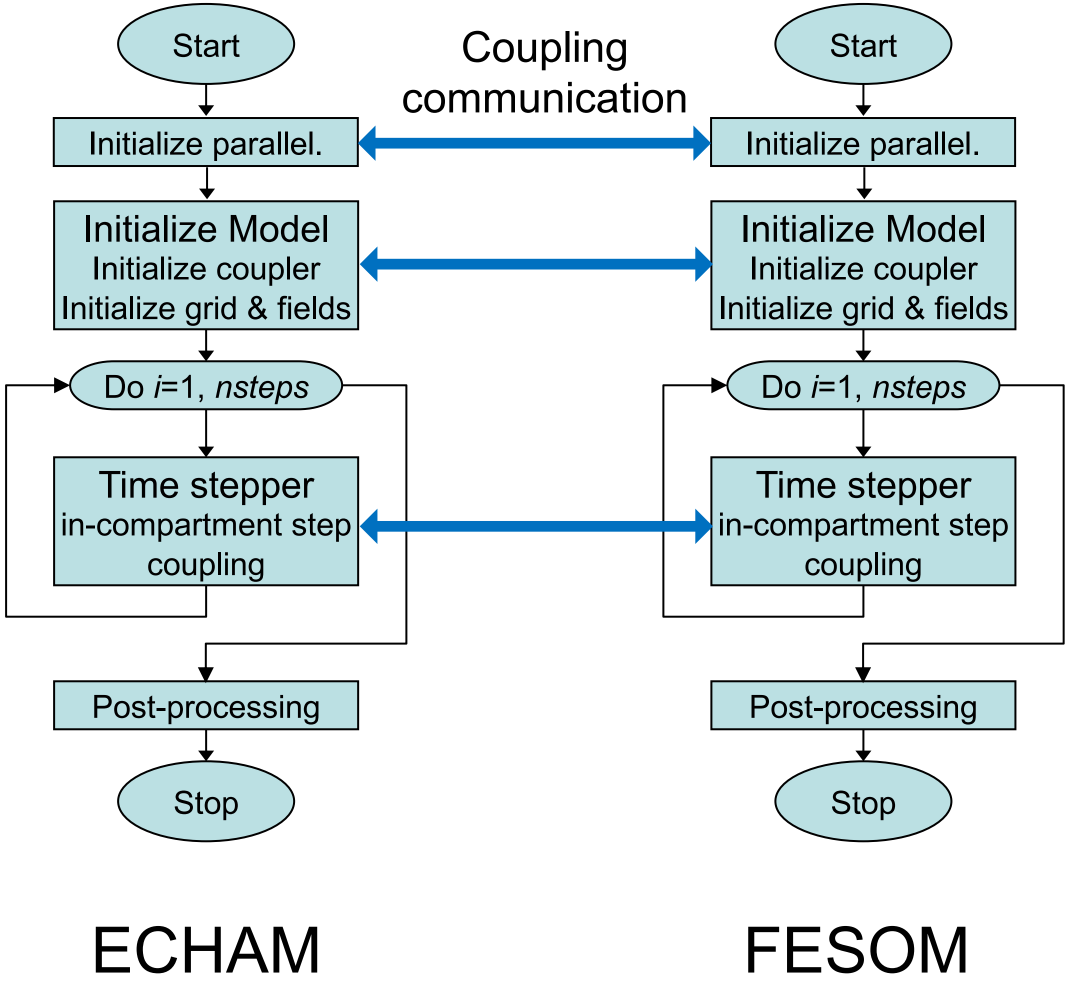
AWI

Augmenting a Model for Data Assimilation

Model

single or multiple executables

coupler might be separate program



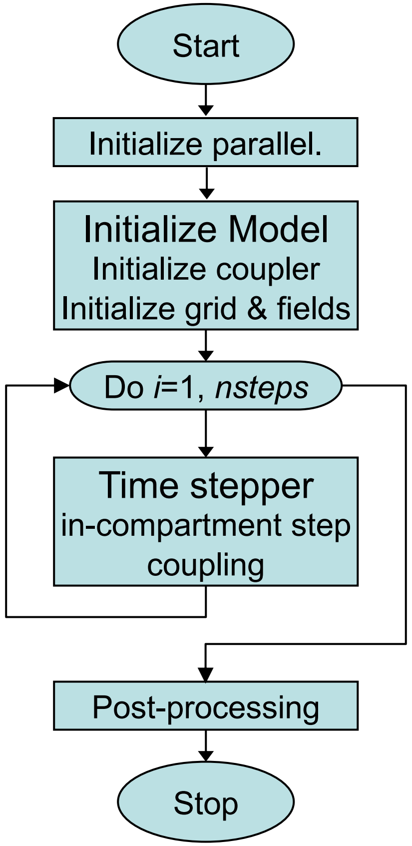
Augmenting a Model for Data Assimilation

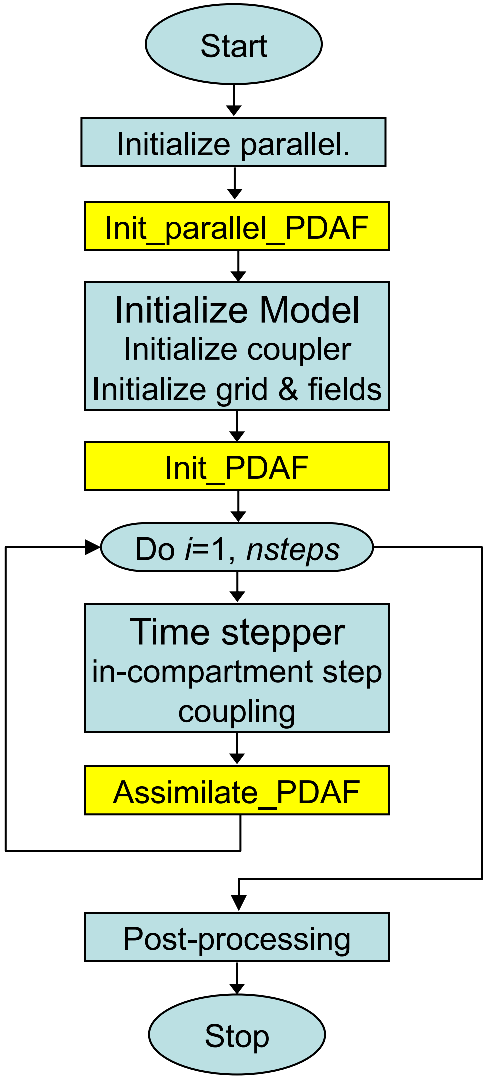
PDAF Parallel Data Assimilation Framework

Model

single or multiple executables

coupler might be separate program





Extension for data assimilation

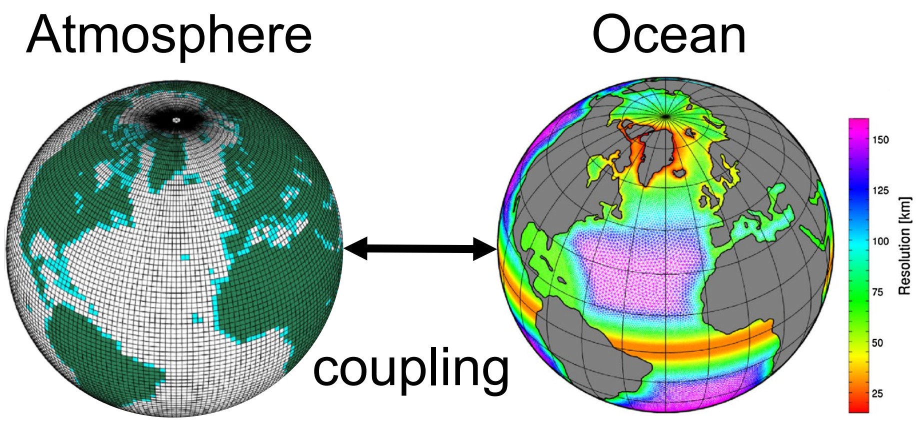
Augment both ECHAM & FESOM

plus:
Possible model-specific adaption

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

revised parallelization enables ensemble forecast

Requirements on the Coupler



- Coupling to PDAF bypasses model coupler
  - Provides direct access to model fields and mesh information
  - Should be compatible with any coupler

- Coupler has to support ensemble integrations
  - Run several model instances concurrently
  - Example OASIS3-MCT (version in AWI-CM)
    - uses MPI_COMM_WORLD ➜ need to be replaced
    - Current version allows to specify '*commworld*'

MPI Process setup

Communicators for AWI-CM (single model instance)

0	1	2	3	4	5
0	1	2	3	0	1

← Set by OASIS3-MCT

Color legend:

MPI_COMM_WORLD
COMM_FESOM
COMM_ECHAM

MPI Processes – setup for ensemble run

PDAF Parallel Data Assimilation Framework

Communicators for AWI-CM (single model instance)

0	1	2	3	4	5
0	1	2	3	0	1

Color legend:

MPI_COMM_WORLD	COMM_CPLMOD
COMM_FESOM	COMM_COUPLE
COMM_ECHAM	COMM_FILTER

Communicators for ensemble run (ensemble size 3)

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	3	4	5	← Set by PDAF
0	1	2	3	0	1	0	1	2	3	0	1	0	1	2	3	0	1	← Set by OASIS3

Realization 1 | Realization 2 | Realization 3

Nerger et al., GMD (2020), doi:10.5194/gmd-13-4305-2020

AWI

MPI Processes – typical setup for assimilation

Communicators for AWI-CM (single model instance)

0	1	2	3	4	5
0	1	2	3	0	1

Color legend:

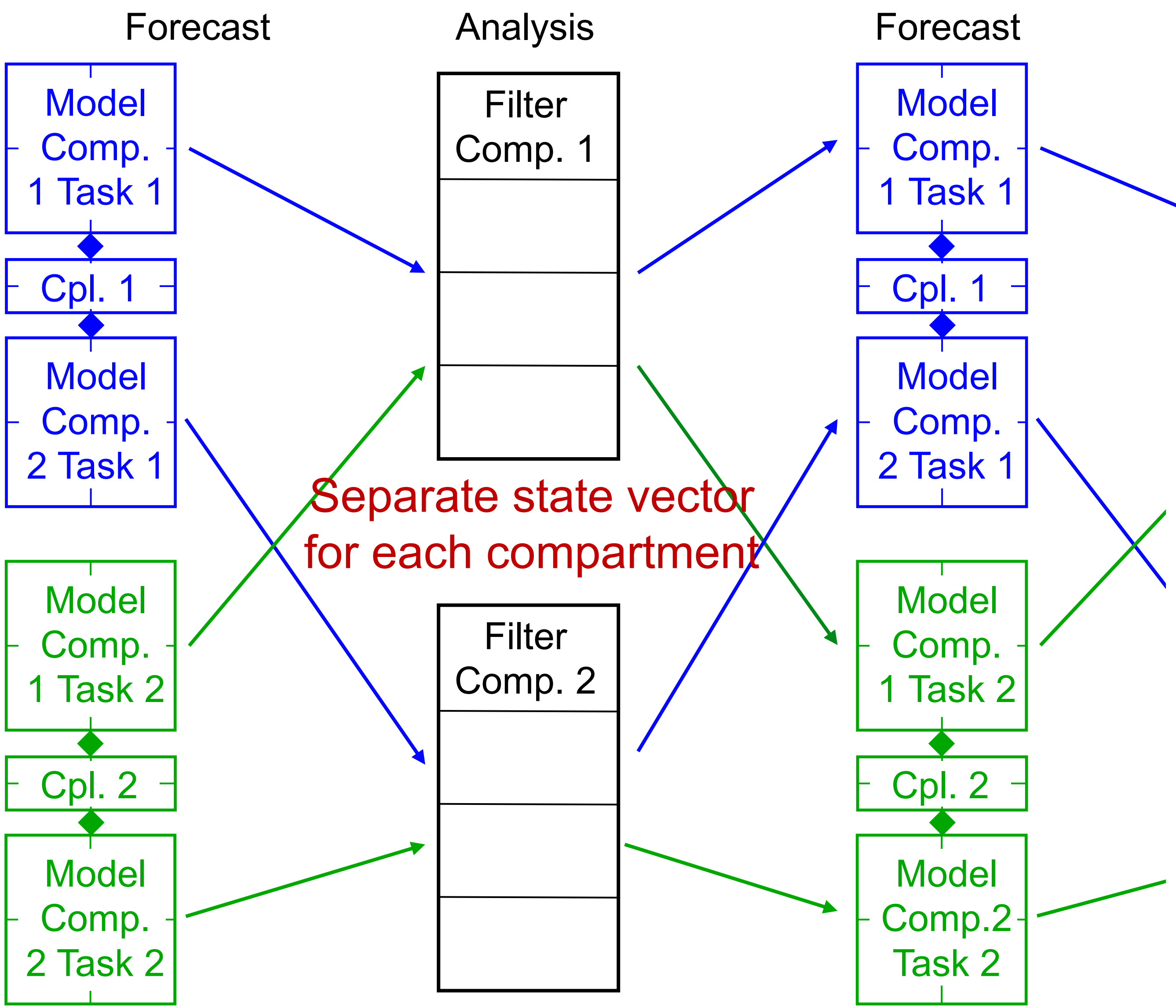
MPI_COMM_WORLD	COMM_CPLMOD
COMM_FESOM	COMM_COUPLE
COMM_ECHAM	COMM_FILTER

Communicators for ensemble run (ensemble size 3)

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	3	4	5	← Set by PDAF
0	1	2	3	0	1	0	1	2	3	0	1	0	1	2	3	0	1	← Set by OASIS3
0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	← Set by PDAF
0	1	2	3	4	5	← For **strongly coupled** assimilation (ECHAM and FESOM combined)												

0	1	2	3	0	1	← For **weakly coupled** assimilation (separate ECHAM and FESOM)

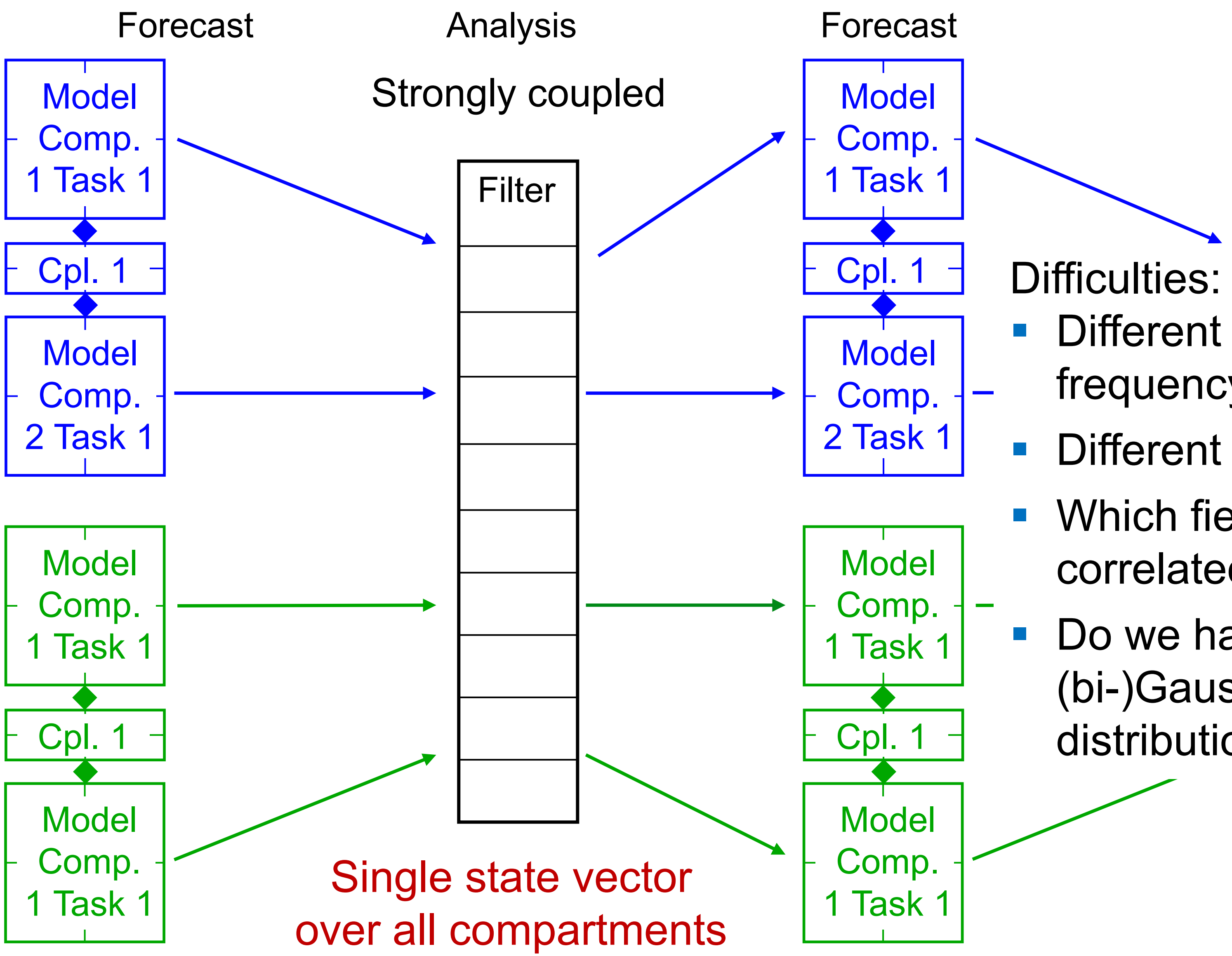
2 compartment system – weakly coupled DA



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

2 compartment system – strongly coupled DA

Switch between weakly coupled and strongly coupled assimilation by change in MPI configuration



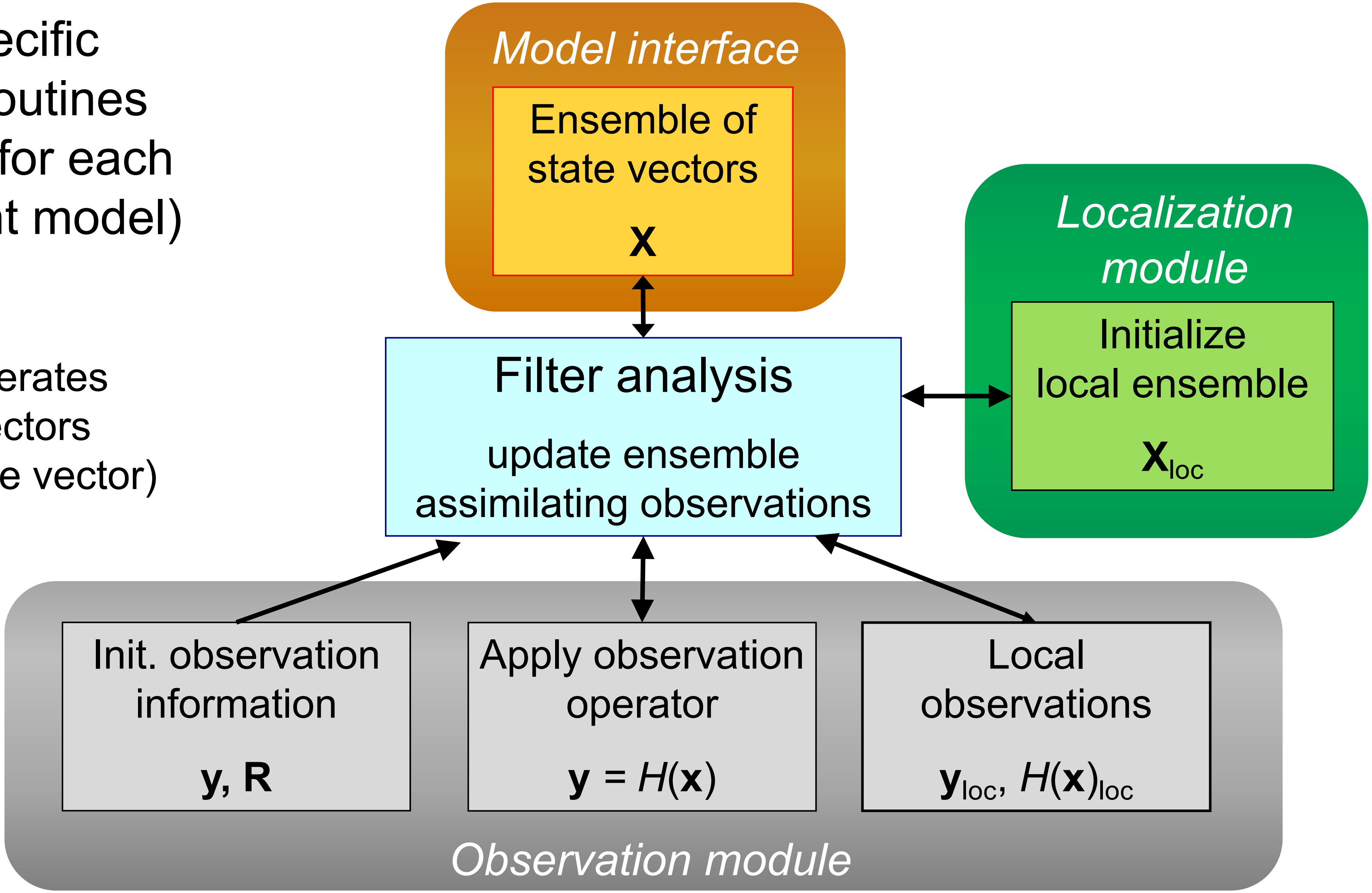
Difficulties:

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

Implementing the Ensemble Filter Analysis Step

case-specific
call-back routines
(implement for each
compartment model)

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



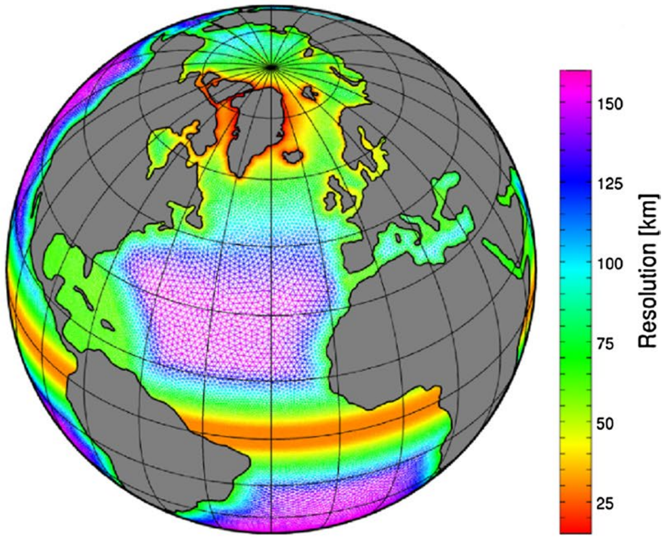
Numerical results

Data Assimilation Experiments

Model setup

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

FESOM mesh resolution



Data assimilation experiments

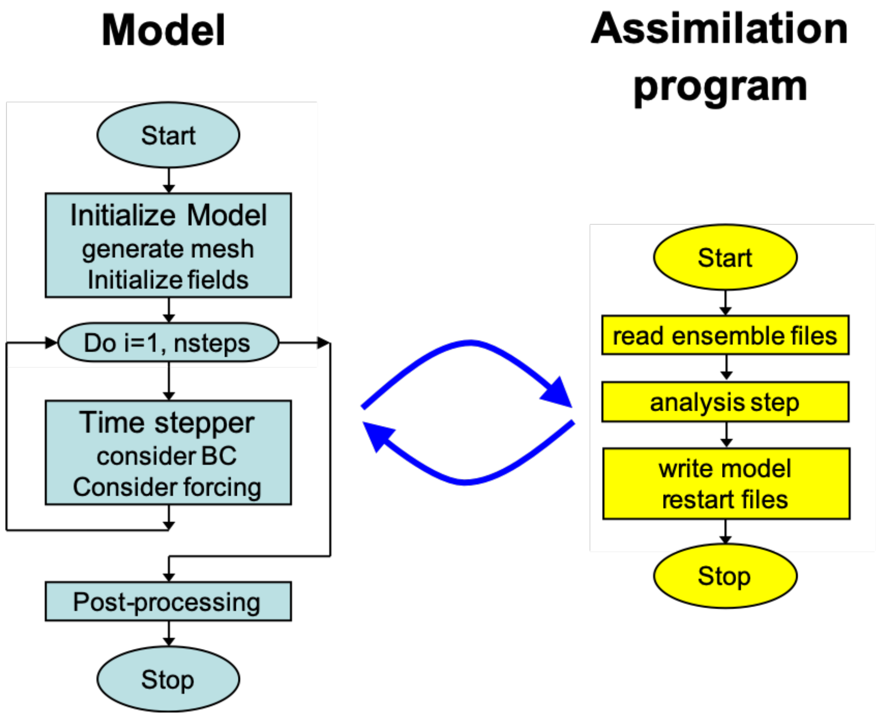
- Observations
  - Satellite Sea surface temperature
  - Temperature and salinity profiles (EN4)
- Updated: ocean (SSH, T, S, u, v, w)
  atmosphere (T, surf. P, vorticity, divergence, humidity, wind velocity)
- Assimilation method: Ensemble Kalman Filter (LESTKF)
- Ensemble size: 46
- Simulation period: year 2016, daily assimilation update
- Run time: ~4h, fully parallelized using 12,000 processor cores

Online and 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
Integrate 1 day:	33 s
Model postprocessing:	14 s
Analysis step:	1 s



Online and Offline Coupling - Efficiency

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

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

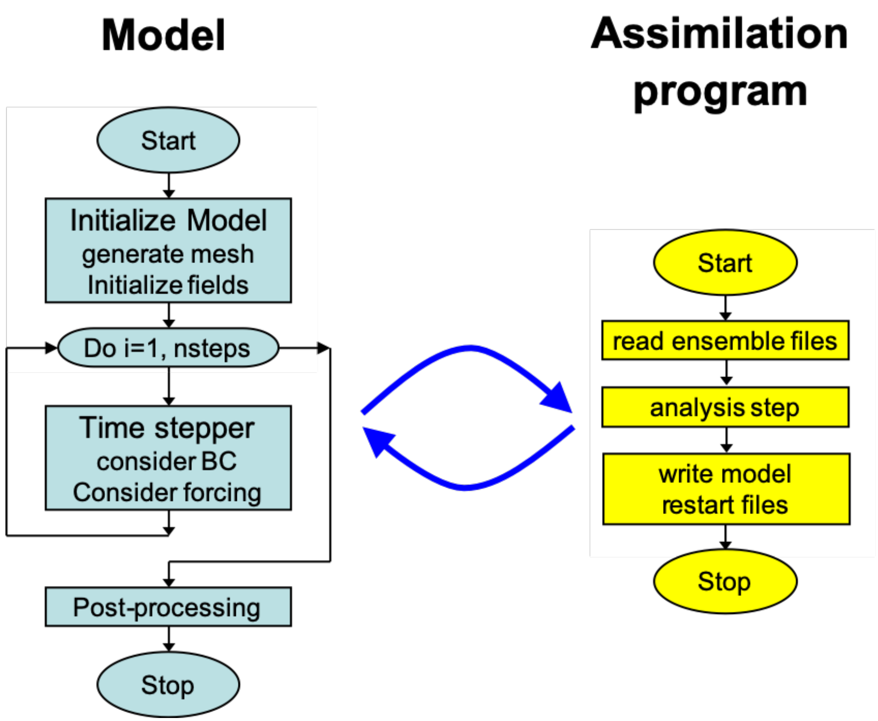
Model startup:	95 s
Integrate 1 day:	33 s
Model postprocessing:	14 s

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)

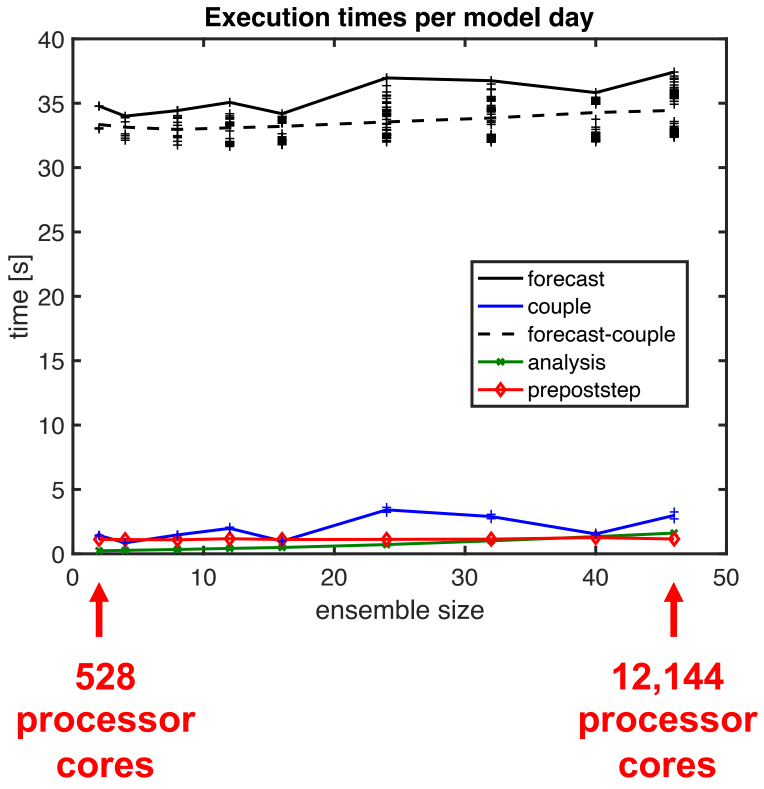
MPI-tasks (each model instance)

- ECHAM: 72
- FESOM: 192
- Vary ensemble size

- 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

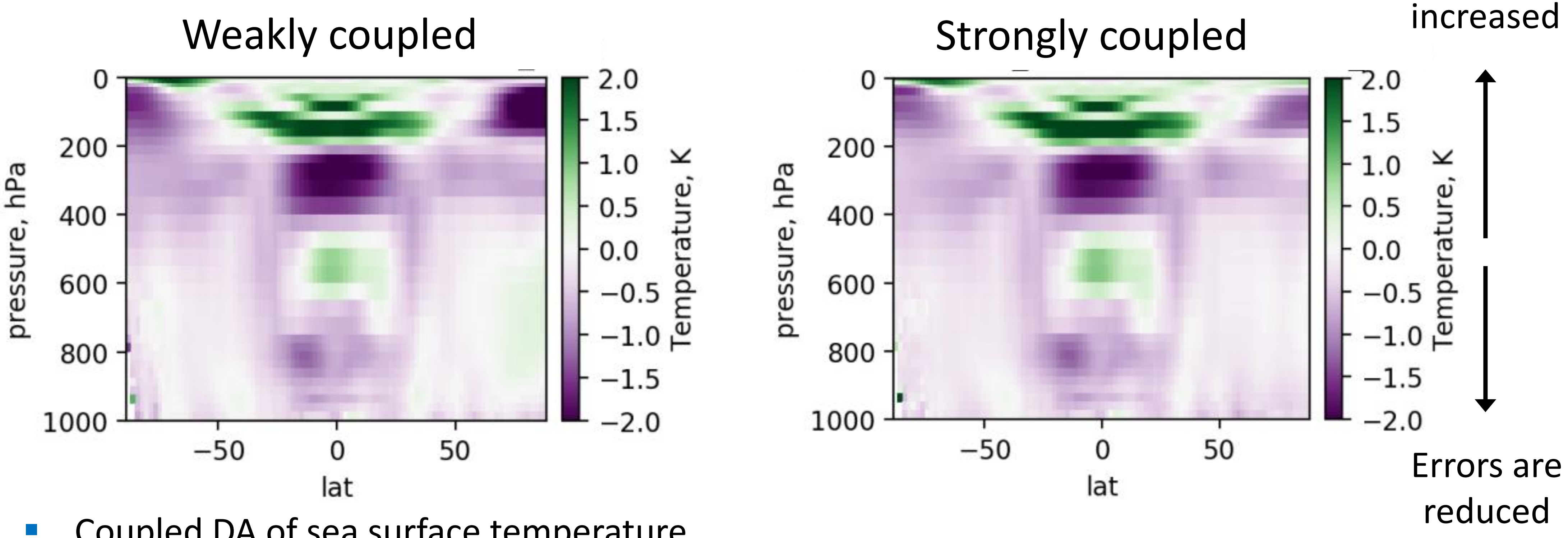
- Need optimal distribution of programs over compute nodes/racks (here set up as ocean/atmosphere pairs)
- Avoid conflicts in IO (Best performance when each AWI-CM task runs in separate directory)



Nerger et al., GMD (2020), doi:10.5194/gmd-13-4305-2020

Strongly and weakly coupled DA

Difference of RMS errors: Assimilation – Free run (zonal averages)



- Coupled DA of sea surface temperature
  - Effect throughout the atmosphere
  - Strongly coupled: reduced errors in Arctic troposphere compared to weaky
  - (currently analyzing results in detail)

Summary

- Efficient assimilative coupled model
  - by combining of coupled model with PDAF (“online-coupling”)
  - bypass the model coupler
  - avoid excessive file IO
  - avoid model restarts
- Resulting model is run like original model
  - with more processes and additional options
- Strongly coupled DA can be easily implemented

  → Making it efficient is the real issue
- PDAF is open source (http://pdaf.awi.de)

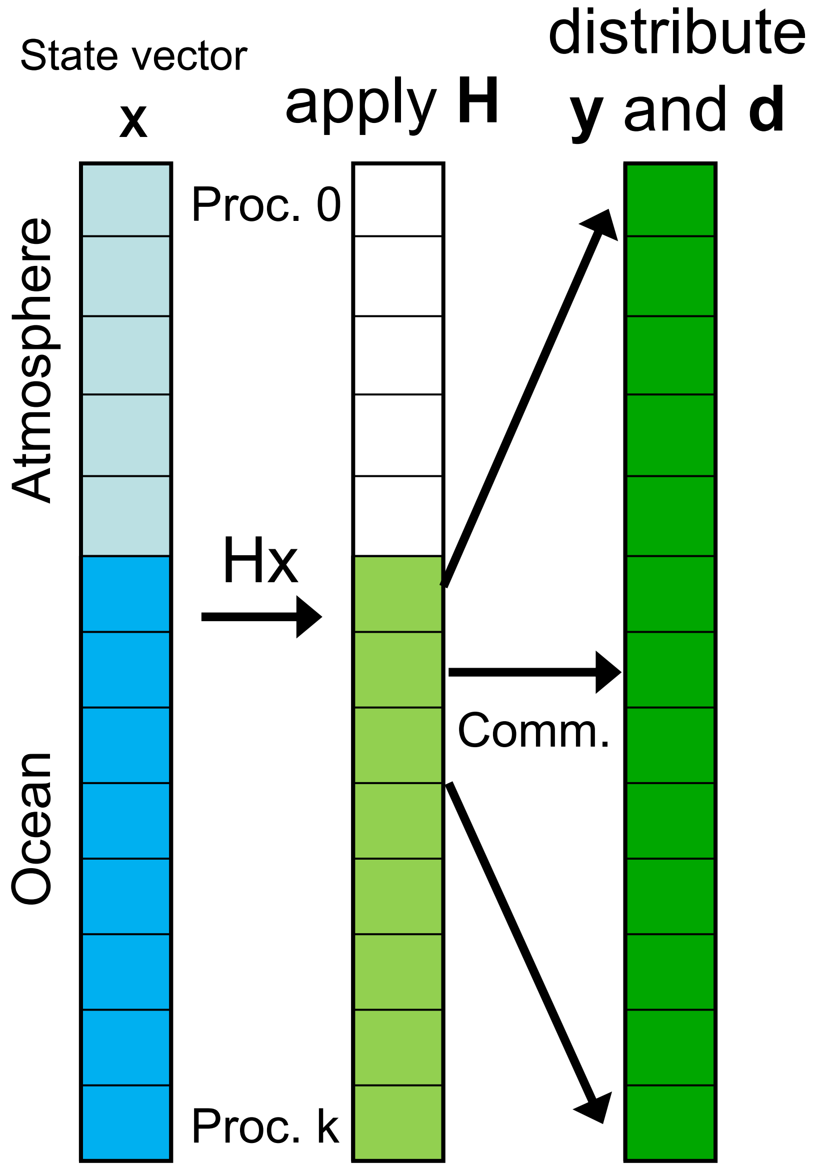
Lars.Nerger@awi.de

References

- **http://pdaf.awi.de**
- Nerger, L., Hiller, W. (2013). Software for Ensemble-based Data Assimilation Systems - Implementation Strategies and Scalability. Computers and Geosciences, 55, 110-118. doi:10.1016/j.cageo.2012.03.026
- Nerger, L., Tang, Q., Mu, L. (2020). Efficient ensemble data assimilation for coupled models with the Parallel Data Assimilation Framework: Example of AWI-CM. Geoscientific Model Development, 13, 4305–4321, doi:10.5194/gmd-13-4305-2020
- Tang, Q., Mu, L., Sidorenko, D., Goessling, H., Semmler, T., Nerger, L. (2020) Improving the ocean and atmosphere in a coupled ocean-atmosphere model by assimilating satellite sea surface temperature and subsurface profile data. Q. J. Royal Metorol. Soc., in press doi:10.1002/qj.3885
- Mu, L., Nerger, L., Tang, Q., Losa, S. N., Sidorenko, D., Wang, Q., Semmler, T., Zampieri, L., Losch, M., Goessling, H. F. (2020) Towards a data assimilation system for seamless sea ice prediction based o the AWI climate model. Journal of Advances in Modeling Earth Systems, 12, e2019MS001937 doi:10.1029/2019MS001937

Lars.Nerger@awi.de

Strongly coupled: Parallelization of analysis step



We need innovation: **d** = **Hx** - **y**

Observation operator H 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

In PDAF:

achieved by changing the communicator for the filter processes (i.e. getting a joint state vector decomposed over the processes)