47th International Liège Colloquium, Liège, Belgium, 4 – 8 May 2015

Building Ensemble-Based Data Assimilation Systems for High-Dimensional Models

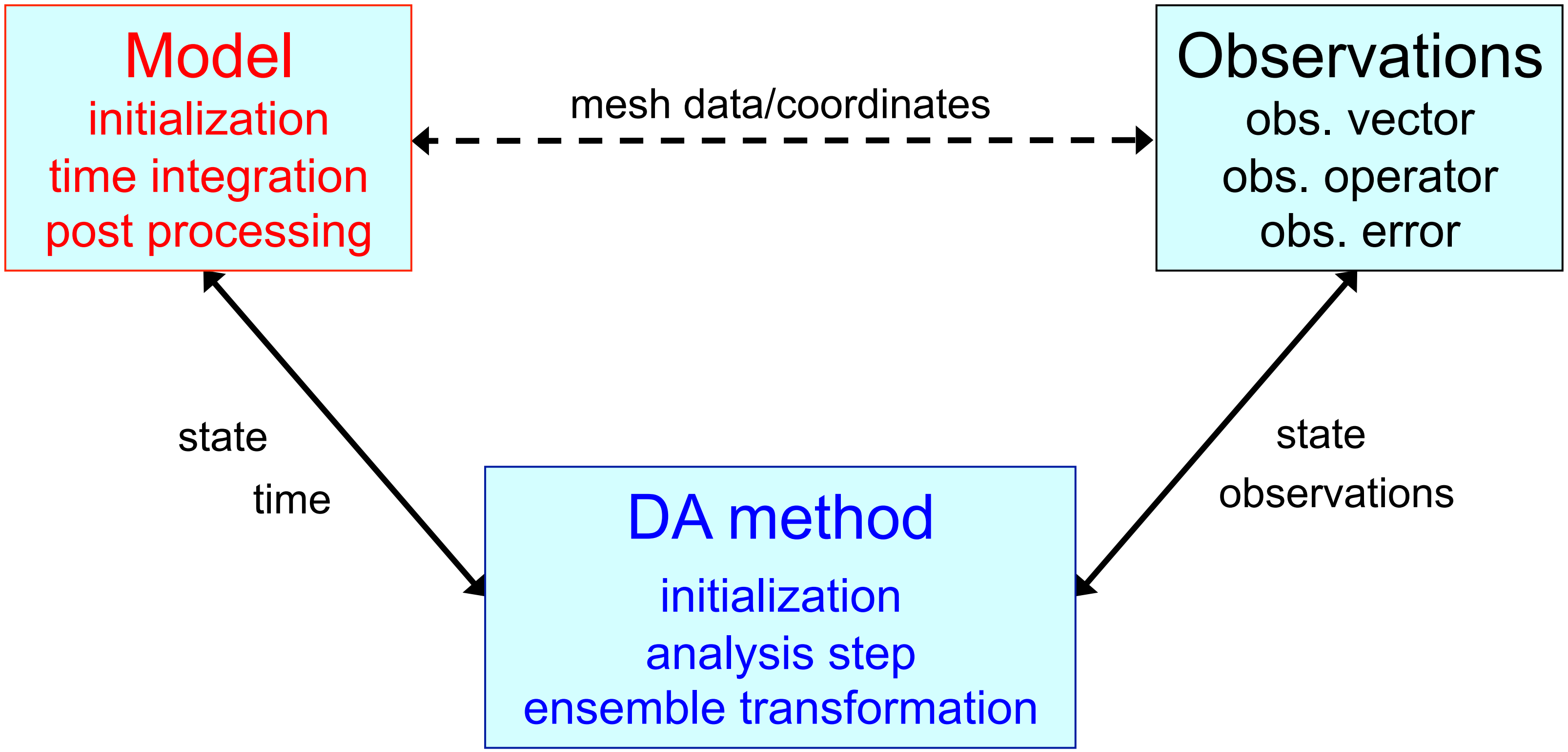
Lars Nerger, Paul Kirchgessner

Alfred Wegener Institute for Polar and Marine Research
Bremerhaven, Germany

SANGOMA

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

3 Components of an Assimilation System

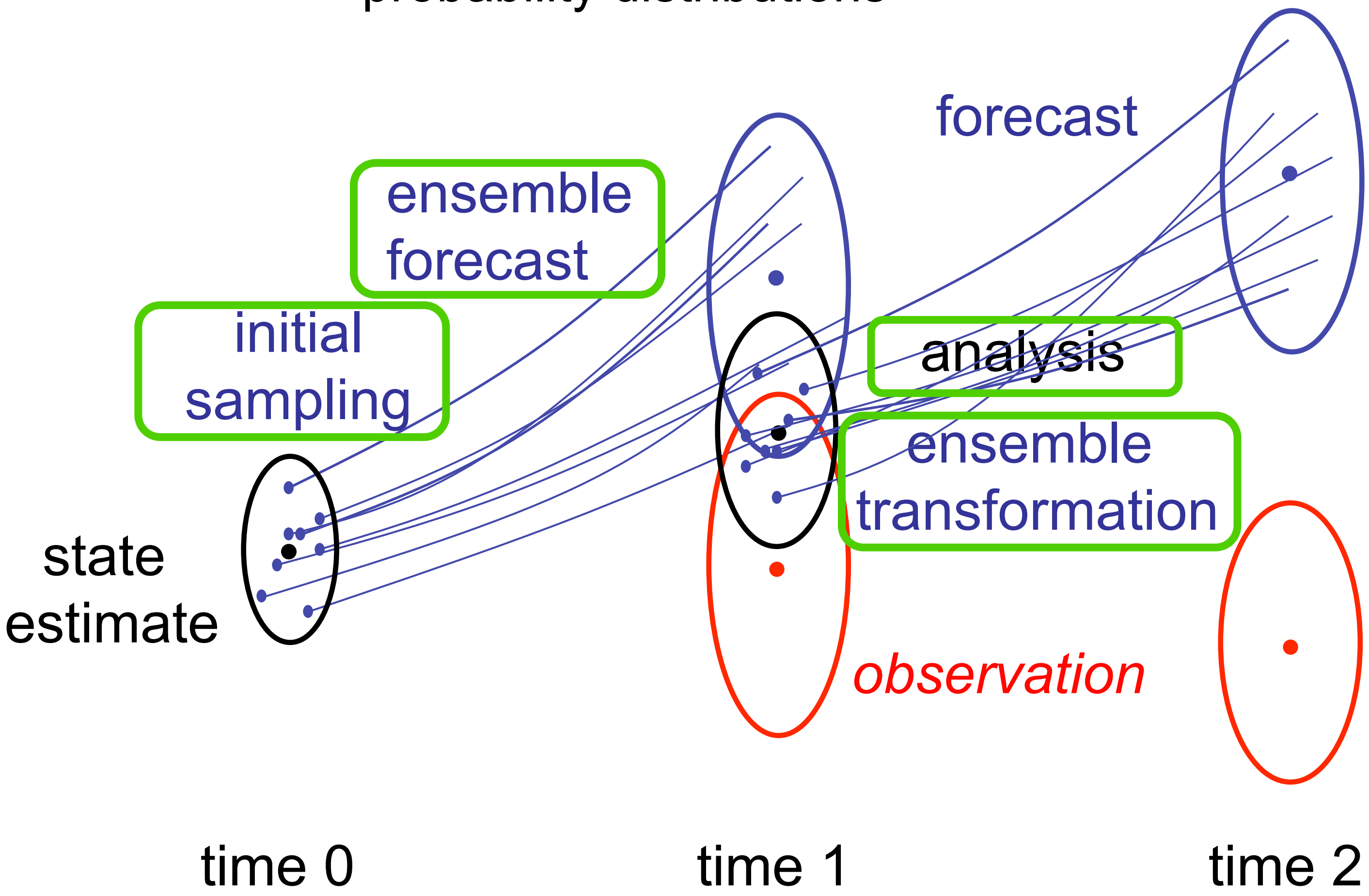


Nerger, L., Hiller, W. Software for Ensemble-based DA Systems – Implementation and Scalability. Computers and Geosciences 55 (2013) 110-118

Ensemble-based Kalman Filter

Kalman filter: express probability distributions by mean and covariance matrix

EnKF (Evensen, 1994): Use ensembles to represent probability distributions



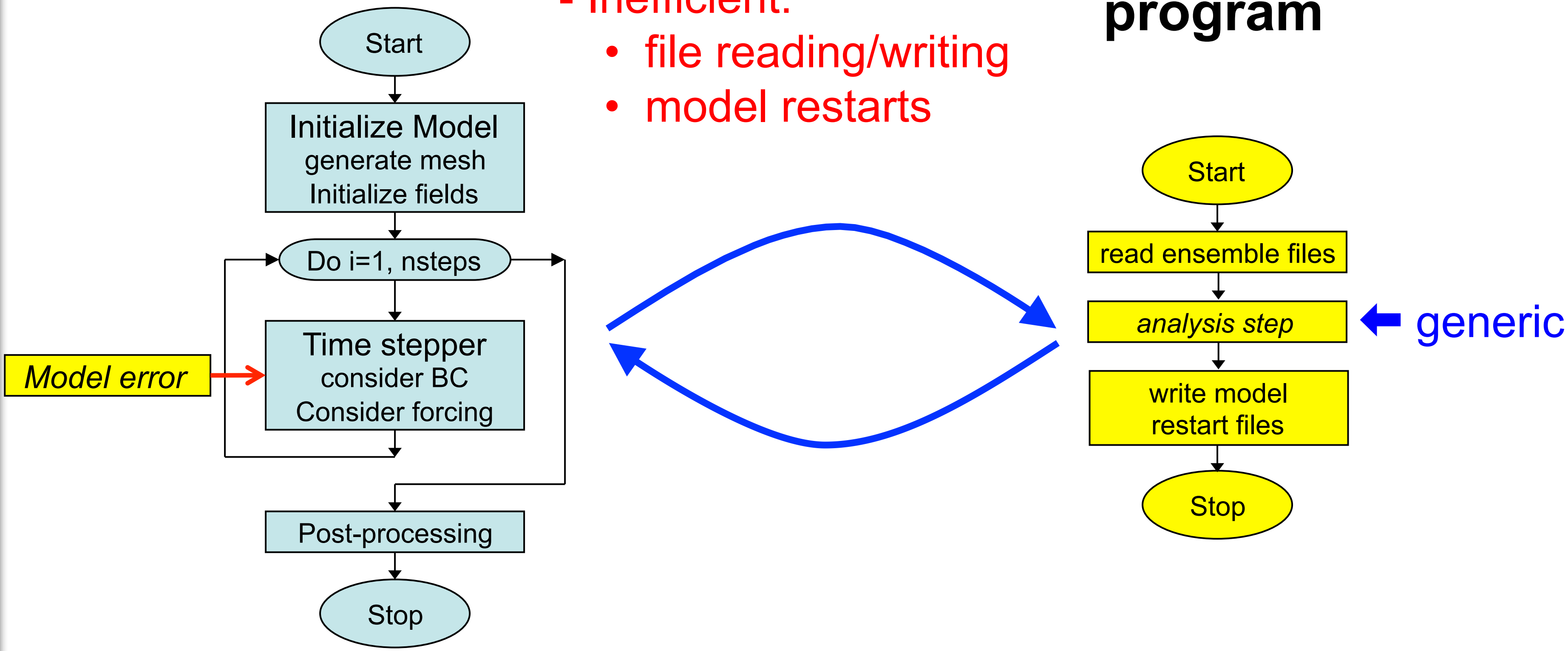
Offline Coupling – Separate Programs

Model

+ Simple to implement

- Inefficient:
  - file reading/writing
  - model restarts

Assimilation program

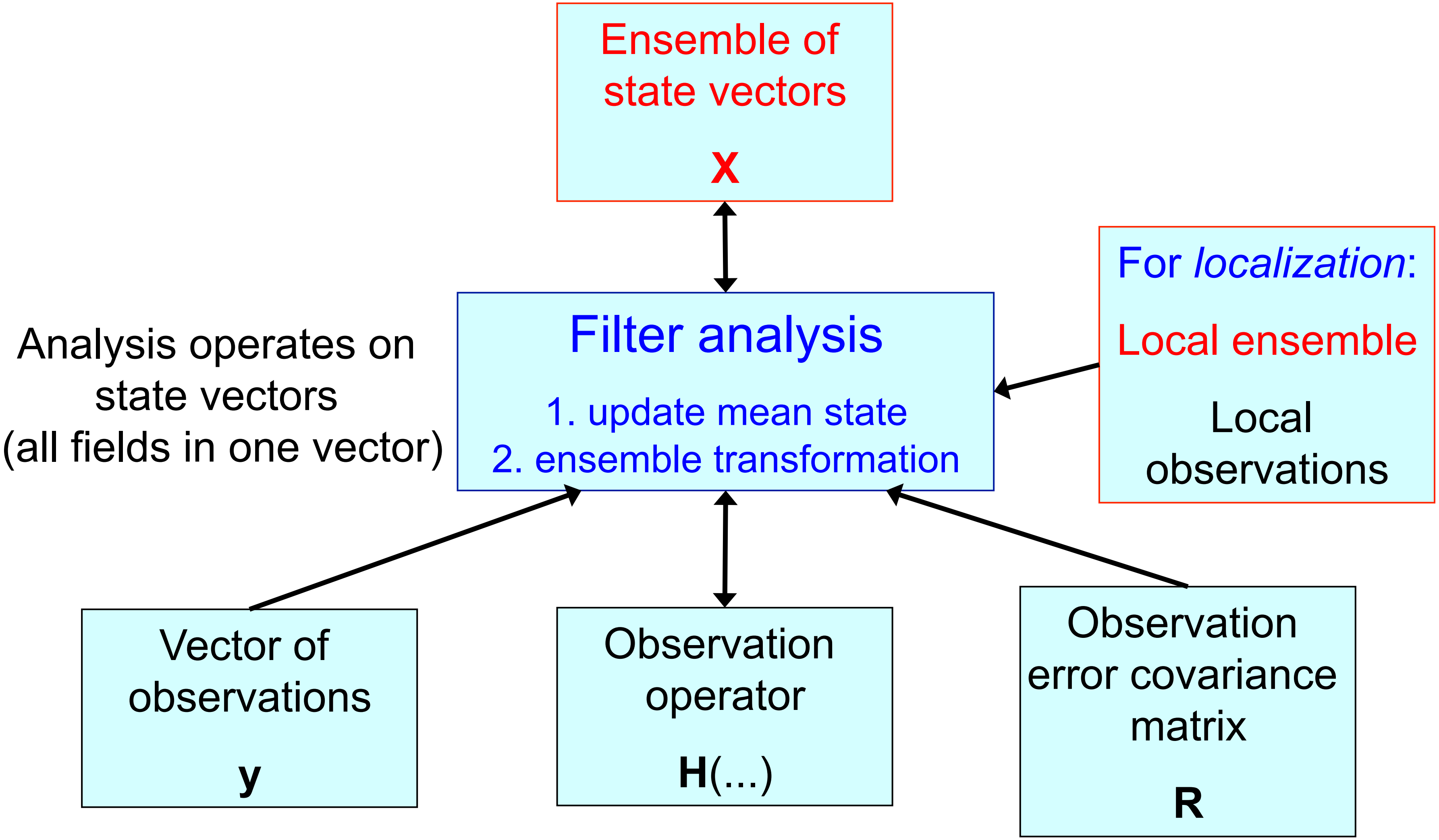


For each ensemble state

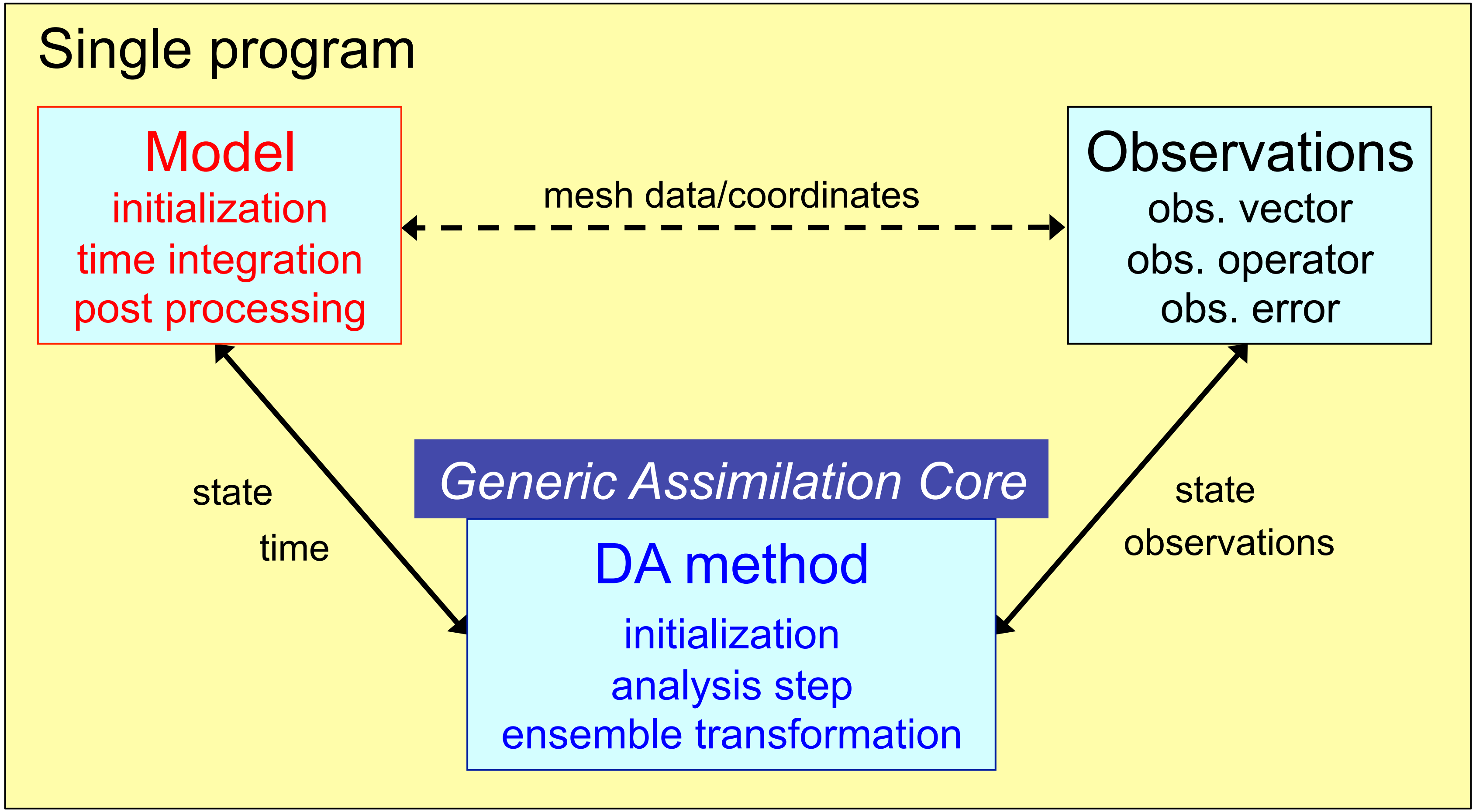
- Initialize from restart files
- Integrate
- Write restart files

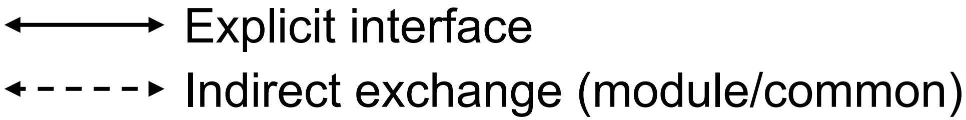
- Read restart files (ensemble)
- Compute analysis step
- Write new restart files

Ensemble Filter Analysis Step



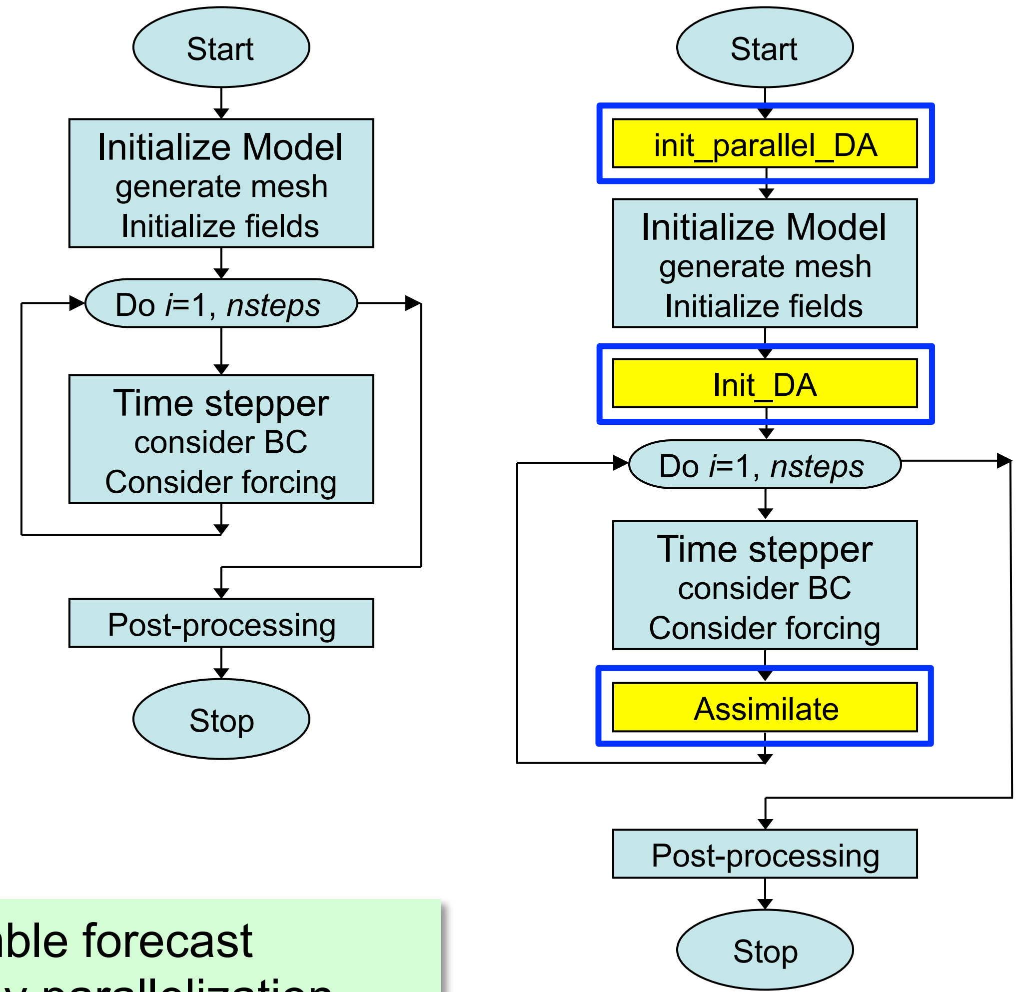
Online Coupling





Extending a Model for Data Assimilation

Model



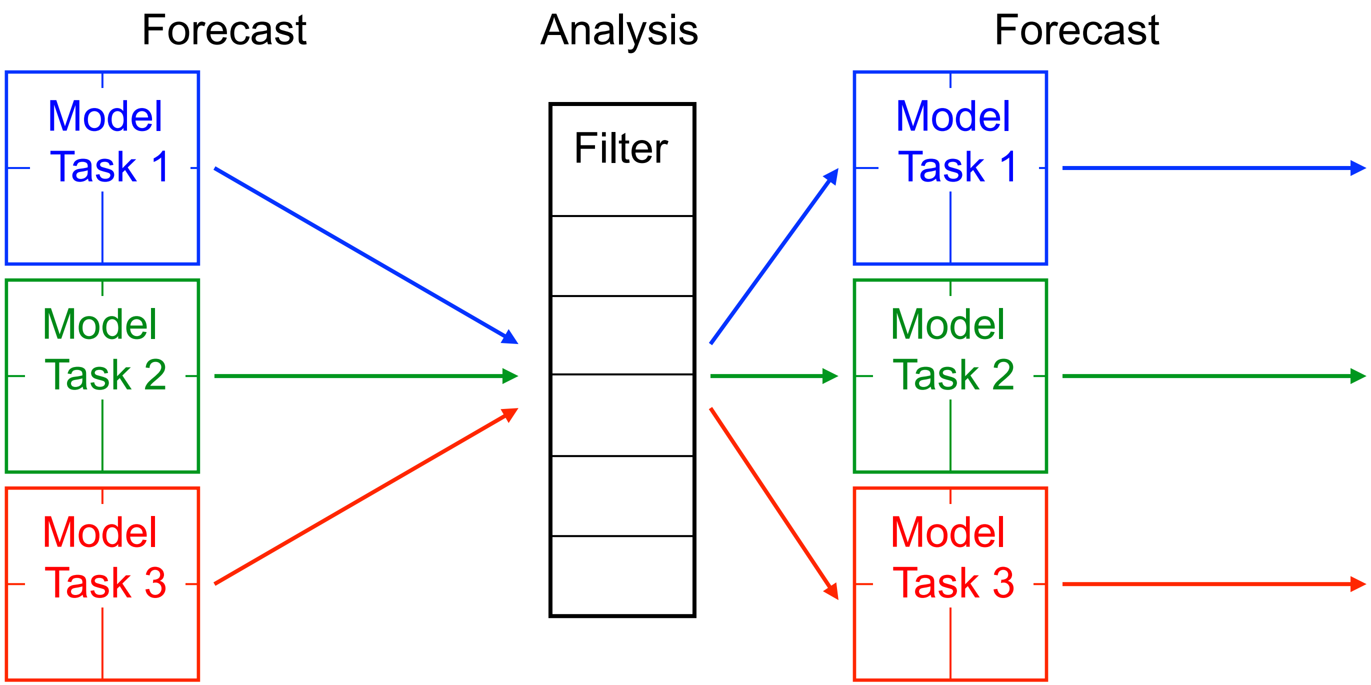
Extension for data assimilation

plus:
Possible model-specific adaption.

NEMO: Euler time step after assimilation

ensemble forecast
enabled by parallelization

2-level Parallelism

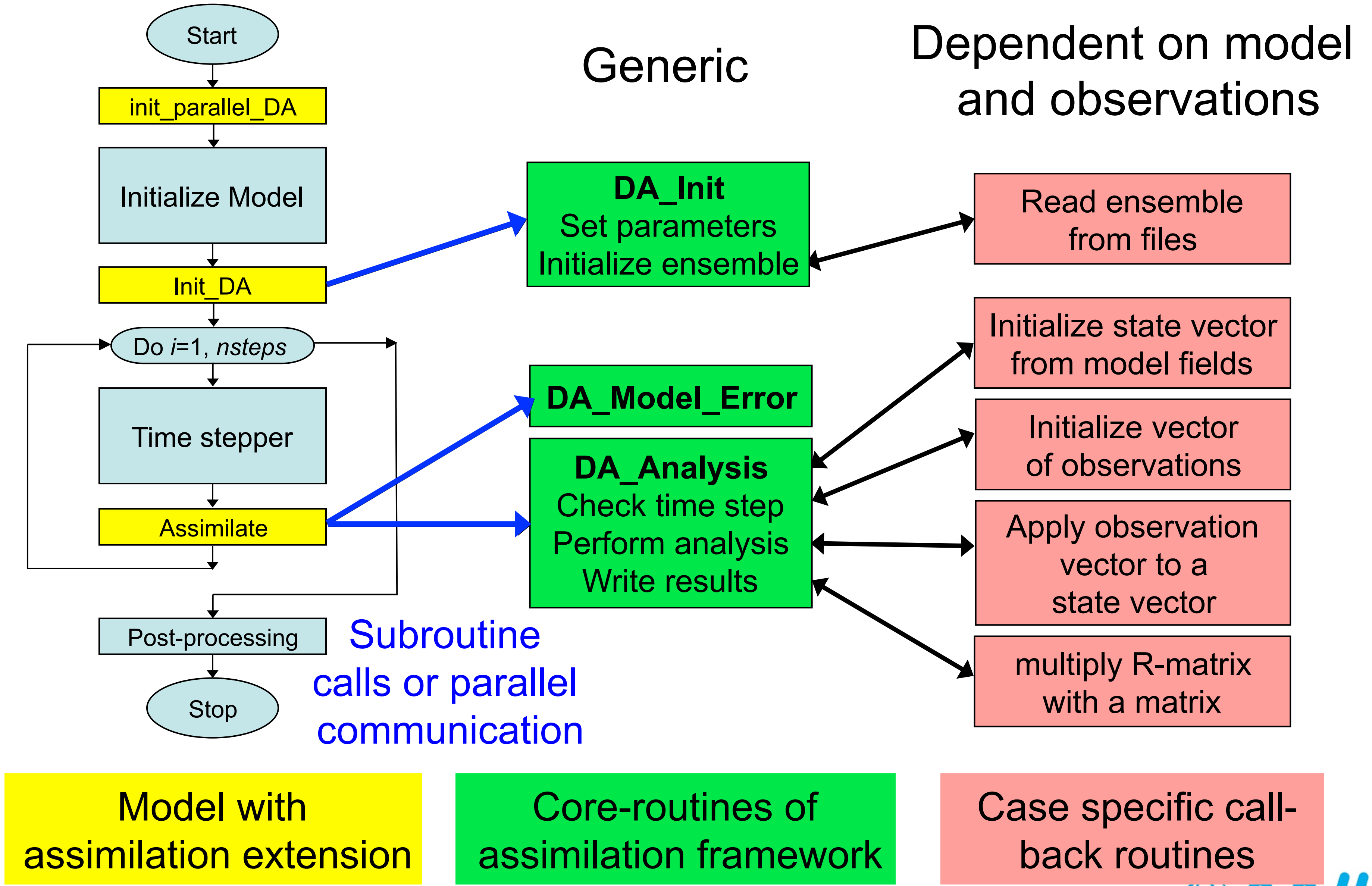


1. Multiple concurrent model tasks
2. Each model task can be parallelized

➢ Analysis step is also parallelized

➢ Configured by "*MPI Communicators*"

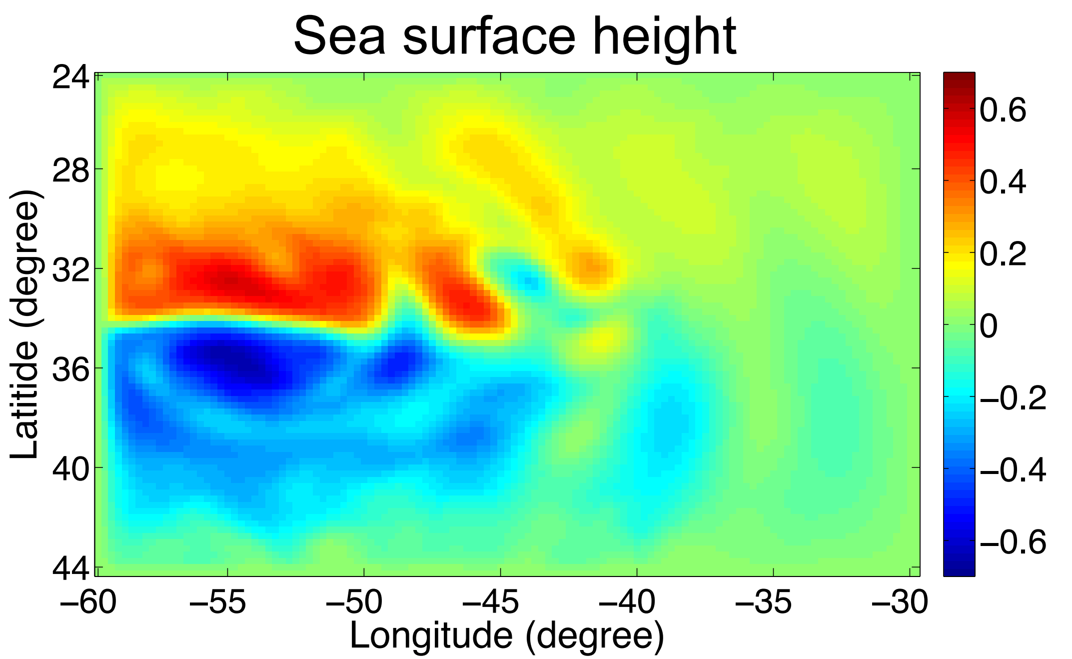
Framework Solution with Generic Filter Implementation



Assimilation Example with NEMO

Model configuration

- medium size SANGOMA benchmark
- box-configuration SQB (SEABASS)
- wind-driven double gyre
- 1/12^{o} resolution
- 361x241 grid points, 11 layers



PDAF: A tool for data assimilation

PDAF Parallel Data Assimilation Framework

PDAF - Parallel Data Assimilation Framework

- provide support for parallel ensemble forecasts
- provide fully-implemented filter and smoother algorithms
- makes good use of supercomputers (Fortran with MPI & OpenMP parallelization)
- easily useable with (probably) any numerical model (coupled e.g. to NEMO, MITgcm, HBM, ADCIRC, FESOM)
- allows for separate development of model and assimilation algorithms

Open source:
Code and documentation available at

http://pdaf.awi.de

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

AWI

Minimal changes to NEMO

Add to *mynode* (lin_mpp.F90) just before init of myrank

```
#ifdef key_USE_PDAF
   CALL init_parallel_pdaf(0, 1, mpi_comm_opa)
#endif
```

Add to *nemo_init* (nemogcm.F90) at end of routine

```
   CALL init_pdaf()
```

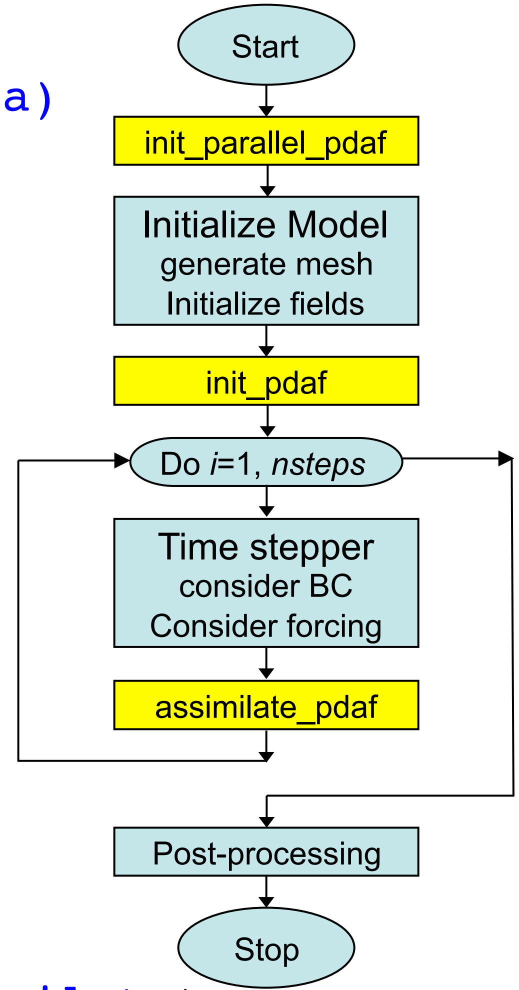
Add to *stp* (step.F90) at end of routine

```
   CALL assimilate_pdaf()
```

For Euler time step after analysis step:

Modify *dyn_nxt* (dynnxt.F90)

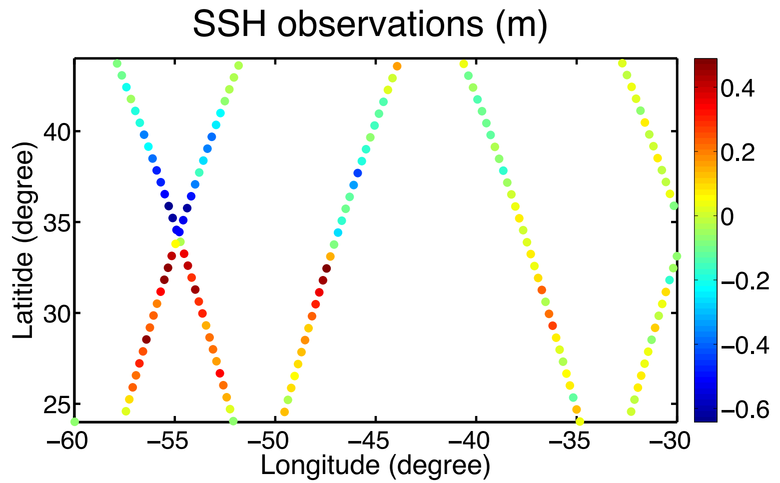
```
#ifdef key_USE_PDAF
   IF((neuler==0 .AND. kt==nit000) .or. assimilate)
#else
```



Assimilation Example with NEMO - Observations

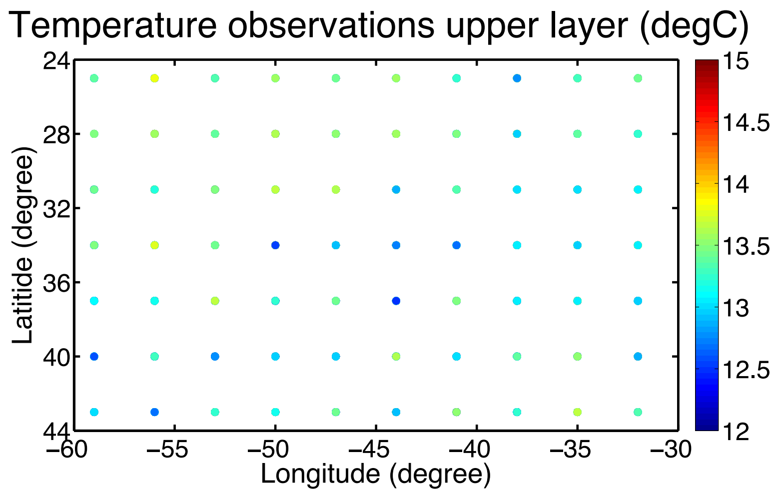
Observations – twin experiment

- Simulated satellite SSH (Envisat & Jason-1 tracks), 5cm error
- Temperature profiles on 3°x3° grid, 0.3°C error



Ensemble data assimilation

- Local ESTKF
- Assimilate each 48h

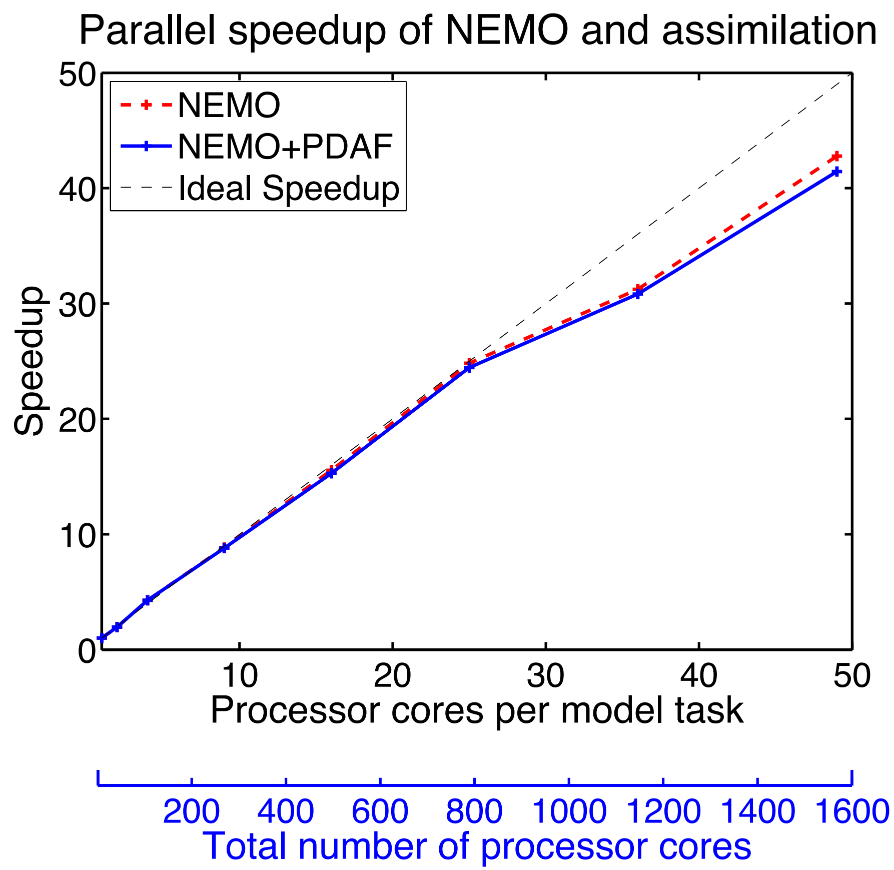


Case-specific routines utilize mesh information from Fortran modules of NEMO

Parallel Performance

- Speedup of NEMO-PDAF SEABASS 1/12° assimilation
- Ensemble size 32
- State dimension ~$3 \cdot 10^6$

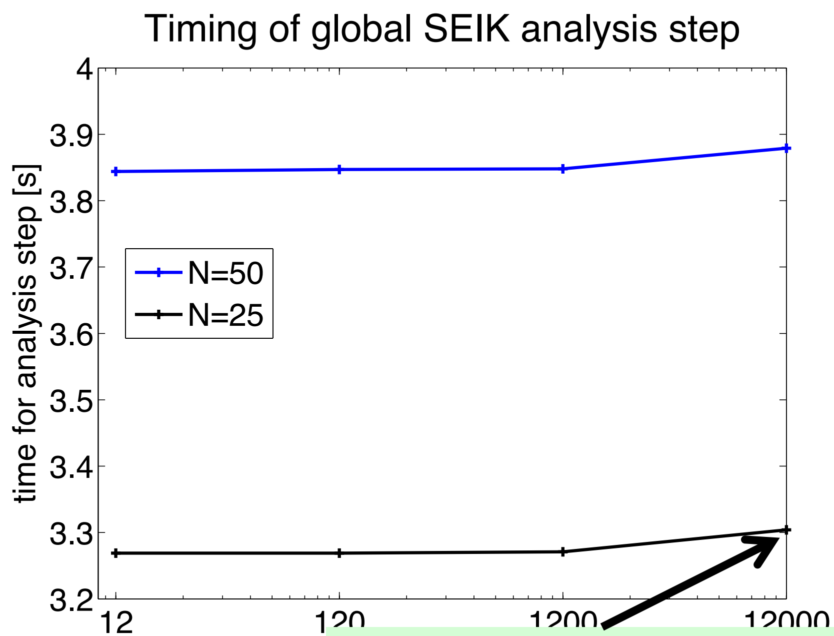
- Speedup determined by speedup of NEMO
- Almost same speedup with assimilation
- Analysis step takes < 8% of total time (0.9s for largest case)



Very big test case

PDAF Parallel Data Assimilation Framework

- Simulate a “model”
- Choose an ensemble
  - state vector per processor: 10^7
  - observations per processor: $2 \cdot 10^5$
  - Ensemble size: 25
  - 2GB memory per processor
- Apply analysis step for different processor numbers
  - 12 – 120 – 1200 – 12000
  - Increase total state and obs. accordingly



State dimension: 1.2e11

Observation dimension: 2.4e9

- Very small increase in analysis time (~1%)
- Didn’t try to run a real ensemble of largest state size (no model yet)

Summary

- Online coupling more efficient than offline coupling
- Generic model interface for online ensemble data assimilation
- Minimal changes to model code
- Parallelization allows for ensemble forecasts
- Data assimilation framework PDAF (http://pdaf.awi.de) supports high-dimensional models
- Coding you own Ensemble Kalman filter usually not necessary

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
- Nerger, L., Hiller, W., Schröter, J. (2005). PDAF - The Parallel Data Assimilation Framework: Experiences with Kalman Filtering, Use of high performance computing in meteorology : proceedings of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology, Reading, UK, 25 - 29 October 2004 / Eds.: Walter Zwieflhofer; George Mozdzynski, Singapore: World Scientific, 63-83