Coupled Ensemble Data Assimilation
with the Climate Model AWI-CM

Lars Nerger, Qi Tang, Longjiang Mu

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

4th DA-TT Meeting, Toulouse, 20-22.1.2020
• Assimilation system: AWI-CM and PDAF
• Weakly-coupled assimilation into the ocean component
• Toward strongly coupled assimilation
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 …
Assimilation System

AWI-CM-PDAF
Assimilation into coupled model: AWI-CM

Atmosphere
- ECHAM6
- JSBACH land

Ocean
- FESOM
- includes sea ice

Coupler library
- OASIS3-MCT

Goal: Develop data assimilation methodology for cross-domain assimilation ("strongly-coupled")

AWI-CM: Sidorenko et al., Clim Dyn 44 (2015) 757
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 (coupled to e.g. NEMO, MITgcm, FESOM, HBM, MPI-ESM, SCHISM)
- run from laptops to supercomputers (Fortran, MPI & OpenMP)
- Usable for real assimilation applications and to study assimilation methods
- ~400 registered users; community contributions

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

Augmenting a Model for Data Assimilation

Model
- single or multiple executables
- coupler might be separate program

Extension for data assimilation

plus:
Possible model-specific adaption
- e.g. in NEMO or ECHAM: treat leap-frog time stepping

revised parallelization enables ensemble forecast
Assimilation-enabled Model

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

Single program

Day 1
00:00h
Forecast 1
Forecast 2
... Forecast 40

Day 1
12:00h

Day 1
12:00h
Analysis step in between time steps

Day 2
00:00h
Forecast 1
Forecast 2
... Forecast 40

Day 2
12:00h

Observation

Initialize ensemble

Ensemble forecast

Ensemble forecast with changed fields

Analysis (EnKF)
**Ensemble Filter Analysis Step**

- **Model interface**: Ensemble of state vectors $x$
- **Localization module**: Local ensemble $x_{loc}$
- **Filter analysis**: update ensemble assimilating observations

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

- **Observation module**:
  - Vector of observations $y$
  - Observation operator $H(...)$
  - Observation error covariance matrix $R^{-1} \cdot A$
  - Local observations $y_{loc}$, $H(x)_{loc}$

**case-specific call-back routines**
Weakly-coupled Assimilation in Ocean
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
Offline Coupling - Efficiency

Offline-coupling is simple to implement but can be very inefficient.

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

- Model startup: 95 s
- Integrate 1 day: 33 s (overhead)
- Model postprocessing: 14 s
- 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)

- Satellite sea surface temperature (level 3, EU Copernicus)
- Daily data
- Data gaps due to clouds
- Observation error: 0.8 °C
- Localization radius: 500 km
SST DA: Achieving stable assimilation

Coupled model only represents climate, not weather:
Large initial SST deviation up to 10°C

DA in this case is unstable!

For stabilization:
omit SST observations where
\(|SST_{\text{obs}} - SST_{\text{ens\_mean}}| > 1.6 \degree C|

(30% initially, <5% after 2 months)

Further omit SST observations at grid points
where model has ice
(mismatch between ice and no-ice conditions)

Lars Nerger et al. – Coupled Ensemble Data Assimilation
Assimilation of satellite SST: 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 model layers at profile locations

Free run 4/30/2016 Day 120 Assimilation

Necessary effect: dependent data

independent data
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
Assimilation of Profiles: Effect on the ocean

SST difference (obs-model)

Free run 4/30/2016 Day 120

Assimilation

larger deviations than for SST assimilation (independent data)

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 (dependent data)
Assimilation effect: RMS errors

Overall lowest errors with combined assimilation
- But partly a compromise

* Independent data
Mean increments

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

- non-zero values indicate regions with possible biases
Effect on Atmospheric State

- Compare to ERA-Interim
- Mean over 2016

2-meter temperature
(ERA-Interim – Model)

- Strong improvements over oceans – model SST slightly too cold
- Smaller improvements over land
Effect on Atmospheric State

ERA-Interim – Model (mean over 2016)

Assimilation generally positive

10 meter zonal wind velocity

Free run

Assimilation

10 meter meridional wind velocity

Free run

Assimilation

Lars Nerger et al. – Coupled Ensemble Data Assimilation
Toward Strongly-coupled Assimilation
Strongly Coupled Data Assimilation

- joint assimilation of the compartments
- First step: assimilation ocean observations into atmosphere
- Unfortunately, no results yet

Technical Challenges:

- ECHAM is spectral model
  - Need fields in grid point space for localization (just identified the right place in the code; thank to ECHAM developers)
  - Need coordinate information in ECHAM (hidden in the code, but found it)
2 compartment system – weakly coupled DA

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

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

might be separate programs

Lars Nerger et al. – Coupled Ensemble Data Assimilation
Strongly coupled: Parallelization of analysis step

We need innovation: $d = Hx - y$

**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

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

- Assimilation system of AWI-CM with PDAF for coupled DA
- Weakly coupled assimilation
  - Good effects of assimilation for ocean
  - Improvements in atmosphere
- Strongly coupled
  - Getting there
  - Technically not difficult for analysis step
  - ECHAM6 is tricky
- Further current work
  - Upgrade to FESOM2 (finite-volume) coupled to IFS

Lars.Nerger@awi.de