Strongly Coupled Data Assimilation and Initialization with the Parallel Data Assimilation Framework (PDAF)

Lars Nerger, Qi Tang, Longjiang Mu

Alfred-Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany
Contact: Lars.Nerger@awi.de http://www.awi.de

PDAF is open source. The code, and documentation are available at
http://pdaf.awi.de

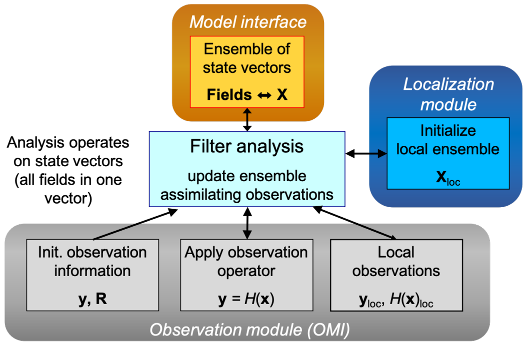
Overview

The Parallel Data Assimilation Framework (PDAF, [1]) is an open-source software framework for highly efficient ensemble data assimilation (DA) with complex models using supercomputers. PDAF was developed to simplify the generation of a DA system from existing models. PDAF provides functionality to perform ensemble integrations, which can be used for ensemble predictions and ensemble DA. Further, PDAF provides fully-implemented ensemble filter and smoother methods for DA. Its internal interface allows to easily add new assimilation methods.

PDAF supports both weakly and strongly coupled DA with easy switching between the approaches. We use a direct connection between the coupled model and PDAF which allows us to set up a DA program with high flexibility and parallel scalability with only small changes to the model. We discuss the structure and features on the example of the coupled atmosphere-ocean model AWI-CM.

Call-back Routines for Analysis Step

Model fields need to be written into the state vectors and back. The filter analysis step needs information on the assimilated observations. PDAF uses call-back routines for this. The programs of the atmosphere and ocean models use distinct user routines for handling observations and model fields. The filter and smoother algorithms are provided by PDAF.



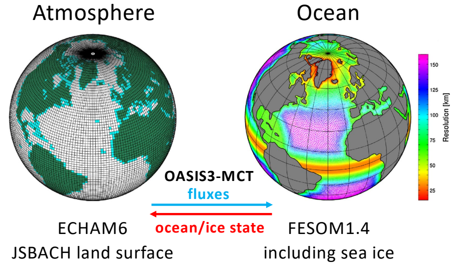
References:

[1] Nerger & Hiller Software for Ensemble-based Data Assimilation Systems - Implementation Strategies and Scalability. Comp. & Geosci. (2013) 55: 110-118

[2] Sidorenko, D. et al. Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate, Clim. Dyn. (2015) 44:757–780

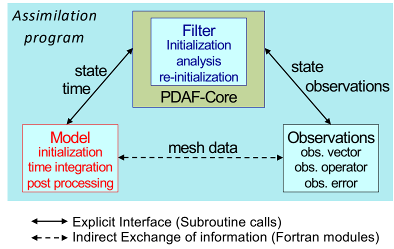
[3] Nerger, Tang, Mu, Efficient ensemble data assimilation for coupled models with the Parallel Data Assimilation Framework: example of AWI-CM (AWI-CM-PDAF 1.0), Geosci. Model Dev. (2020) 13: 4305

Example Coupled Model: AWI-CM



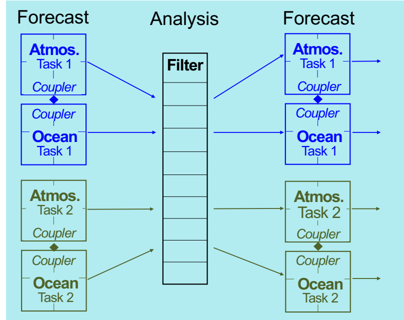
AWI-CM [2] consists of two separate programs: FESOM and ECHAM6. Both are coupled using OASIS3-MCT and run in parallel. Fluxes between the models are computed and exchanged every hour by OASIS3-MCT using parallel communication.

Data Assimilation System



The data assimilation system has three components: Model, filter algorithm, and observations. The filter algorithms are model-agnostic, while the model and subroutines to handle observations are provided by the user. The observation routines are called by PDAF as call-back routines.

Coupled Ensemble Forecasts

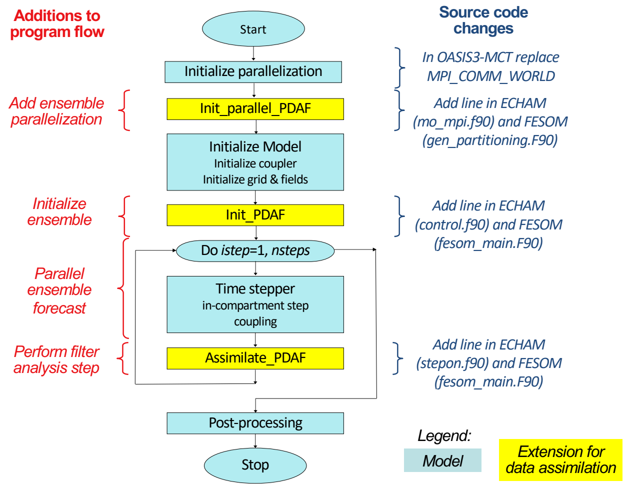


Example of an ensemble integration with two ensemble members. Both models and the filter are parallelized. The ensemble adds one level of parallelization to integrate all members at once. Using a combine atmosphere-ocean state vector permits the application of strongly coupled DA.

PDAF lets you easily build a highly efficient program for coupled ensemble data assimilation

Adapting the Model Codes

We insert three subroutine calls for PDAF into the source codes of both ECHAM6 and FESOM to add data assimilation functionality to the coupled model. Further, we need to adapt a communicator in OASIS3-MCT so that it treats each coupled ensemble task separately.



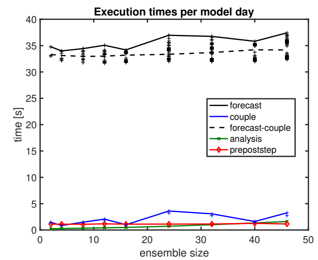
Compute Performance

The experiment

- Daily assimilation of sea surface temperature [3]
- Number of processes: ECHAM 72, FESOM 192
- Assimilation method: Local Error-Subspace Transform Kalman Filter (LESTKF)

Compute Performance

- Run time for ensemble size 46: 3.5 hours (fully parallelized on 12,144 processors)
- Scaling test: increase ensemble size and number of processors
  - Slightly different forecast duration for each ensemble member
  - Run time only increases by 8% for 10-fold ensemble size due to increase time for ensemble communication
  - The Computing time is very similar for strongly and weakly coupled DA



Summary

- Using PDAF we add data assimilation functionality to the coupled model.
- The addition is independent of the actual model coupler.
- PDAF uses in-memory access and parallelization to ensure high efficiency.
- The analysis step is computed in between time steps without stopping the program. There is no need to write the ensemble into files.
- Routines for the model interface and observation handling need to be implemented for each of the two programs for atmosphere and ocean.
- Strongly and weakly coupled Da can be performed based on the configuration of the parallelization for the analysis step (joint or separate state vectors)