Building Ensemble-Based Data Assimilation Systems

Alfred Wegener Institute
Helmholtz Center for Polar and Marine Research

Lars Nerger and Paul Kirchgessner

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

SANGOMA
http://www.data-assimilation.net

Introduction

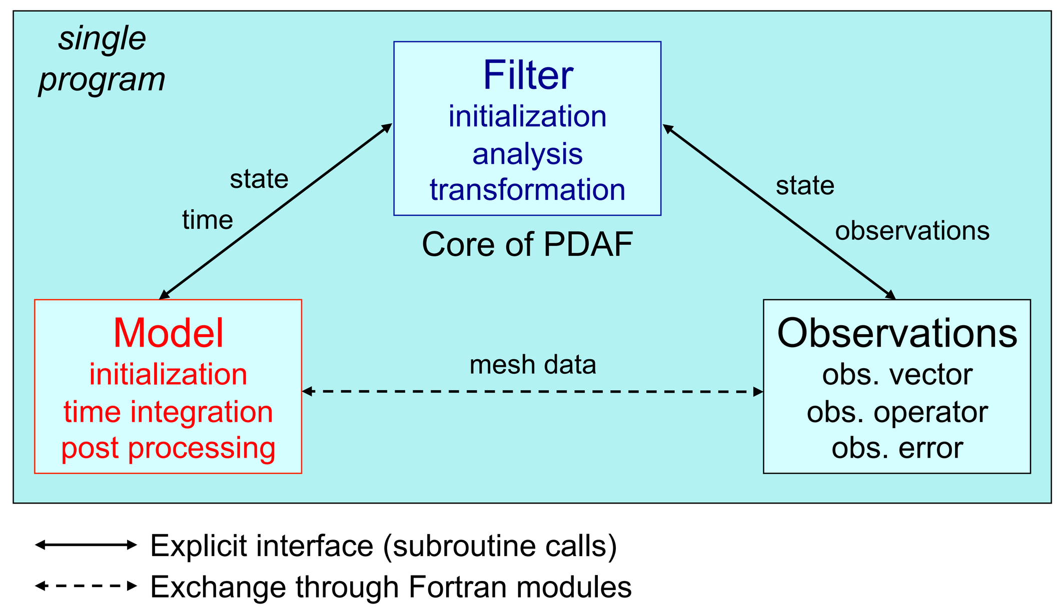
We discuss different strategies for implementing ensemble-based data assimilation systems. Ensemble filters like ensemble Kalman filters and particle filters can be implemented so that they are nearly independent from the model into which they assimilate observations.

Offline coupling through disk files avoids changes to the numerical model, but is computationally not efficient. An online coupling strategy is computational efficient. In this coupling strategy, subroutine calls for the data assimilation are directly inserted into the source code of an existing numerical model and augment the numerical model to become a data assimilative model.

Using the example of the parallel data assimilation framework (PDAF, http://pdaf.awi.de) and the ocean model NEMO, we demonstrate how the online coupling can be achieved with minimal changes to the numerical model.

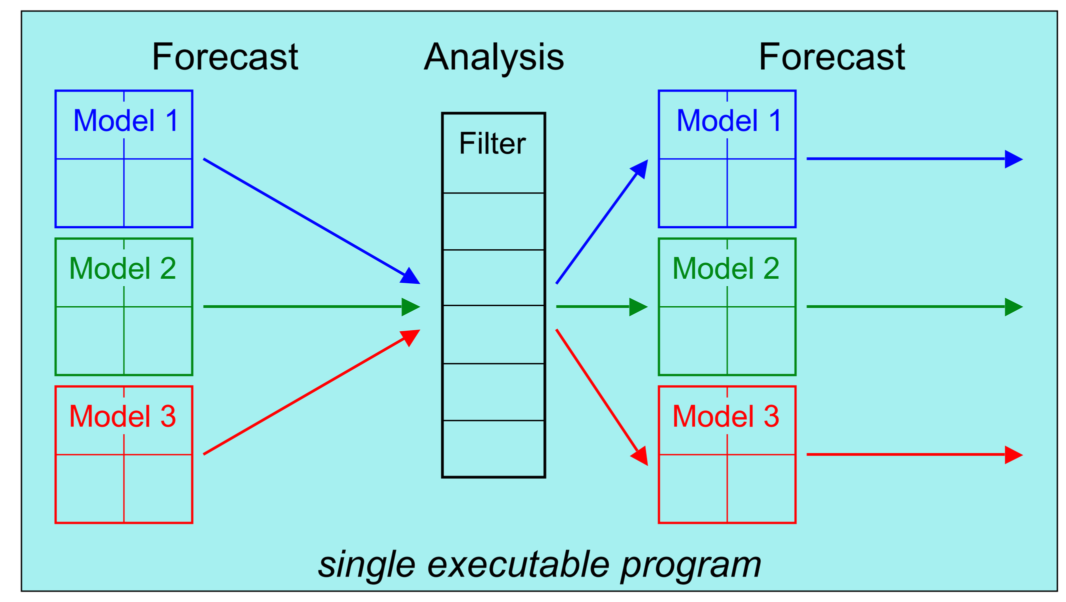
A Parallel Data Assimilation System

Logical separation of the assimilation system



The data assimilation system can be separated into three components: Model, filter algorithm, and observations. The filter algorithms are model-independent, while the model and subroutines to handle observations are provided by the user. The routines are either directly called in the program code or share information, e.g., through Fortran modules.

2-level parallelization of the assimilation system

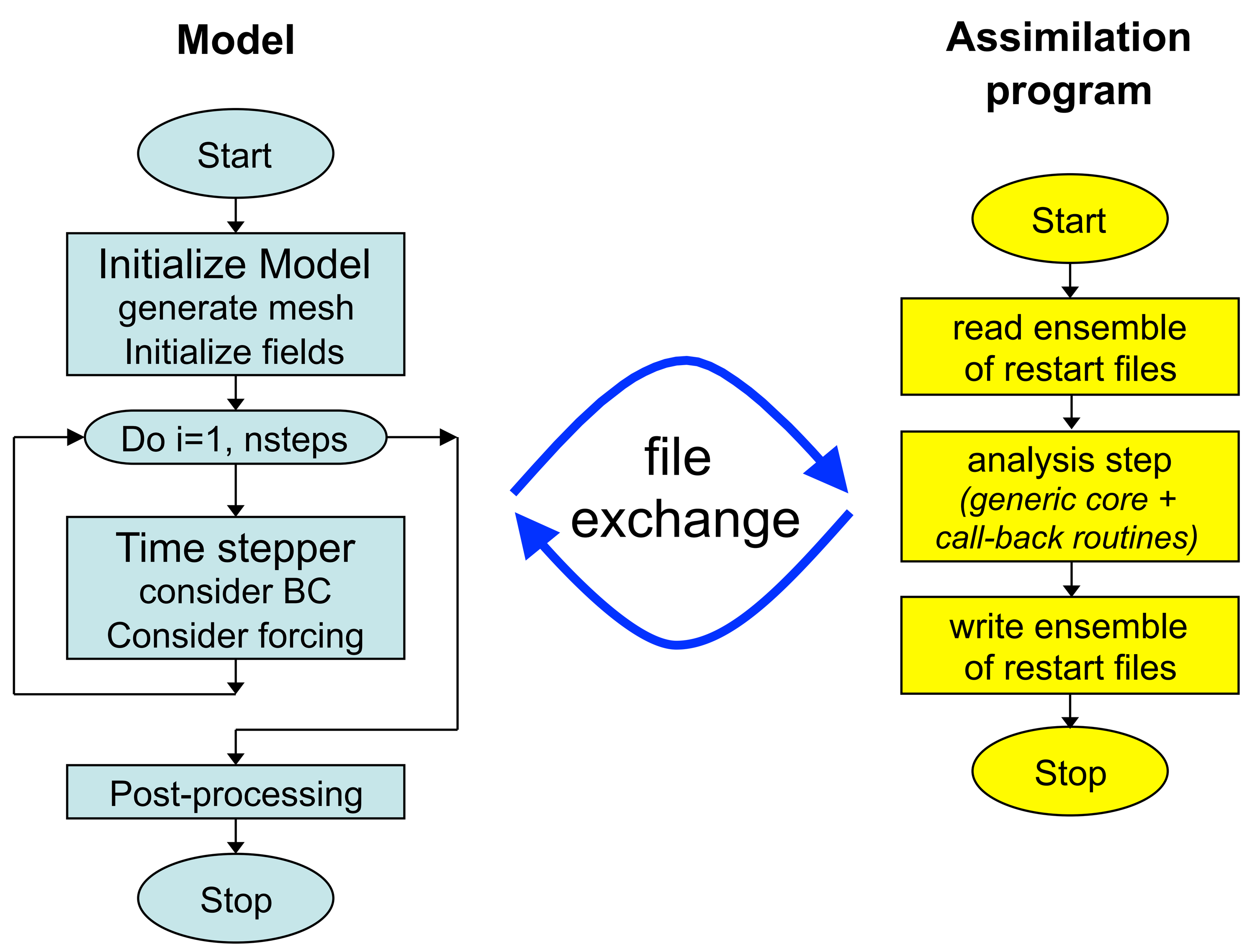


Ensemble-data assimilation can be performed using a 2-level parallelization:

1. Each model integration can be parallelized.
2. All model tasks are executed concurrently.

Thus, ensemble integrations can be performed fully parallel. In addition, the filter analysis step uses parallelization.

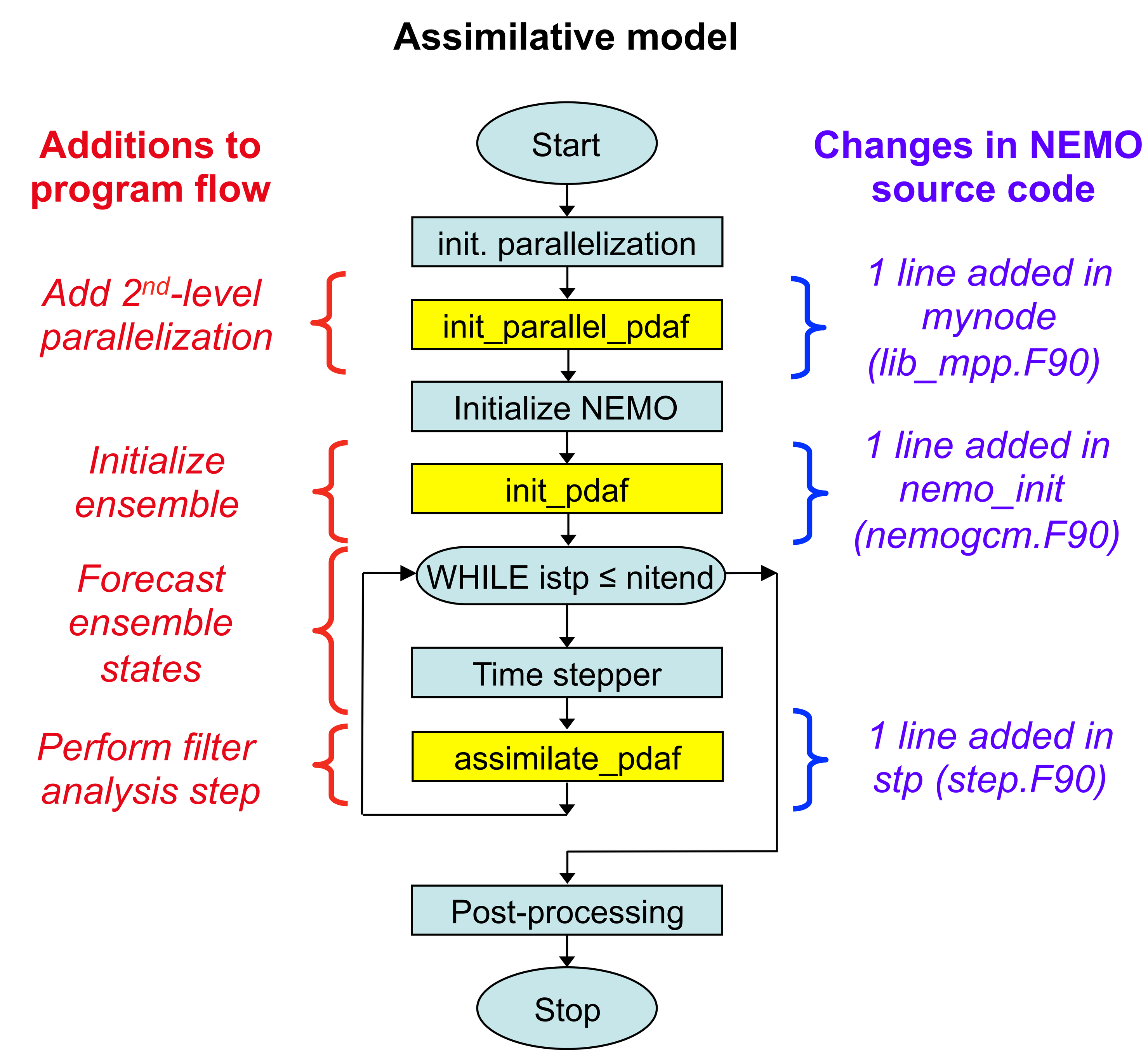
Offline Coupling



For the offline coupling the ensemble forecast is performed by running the model once for each ensemble member. The forecasts are stored in restart files. These files are read in by the assimilation program.

The assimilation program computes the analysis step and writes new restart files. Then the next ensemble forecast is computed by the model. It reads each single restart file and performs the integration.

Online-Coupling of NEMO and PDAF



NEMO is coupled with PDAF [2,3] by adding three subroutine calls the model source code and utilizing parallelization. The model time stepper does not need to exist as a separate subroutine.

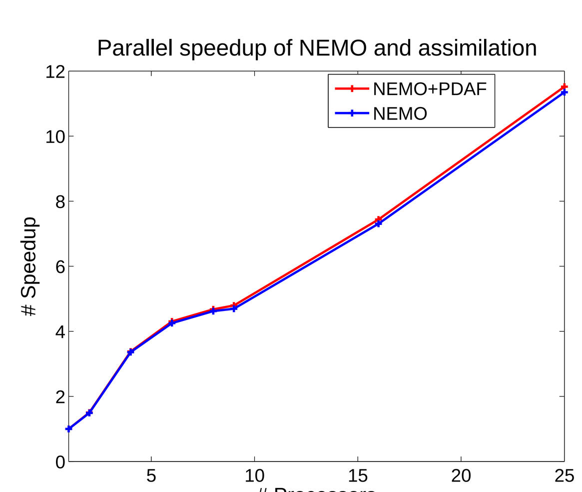
Operations specific to the model and the observations are performed in user-supplied call-back routines that are called through PDAF. The ensemble forecast is also controlled by user-supplied routines.

Parallel Performance of Online Coupling

Assimilation experiments are performed with a box configuration (SEABASS) of NEMO that simulates a double-gyre (see [1]). The configuration is one of the benchmarks of the SANGOMA project. To simulate a high-dimensional model, the resolution is increased to 1/12°. The grid has 361×241 grid points and 11 layers. The state vector has a size of about 3 million.

Synthetic observations of sea surface height at ENVISAT and Jason-1 satellite tracks and temperature profiles on a 3°×3° grid are assimilated each 48 hours over 360 days. Observation errors are respectively set to 5cm and 0.3°C. The assimilation uses the local ESTKF filter [4].

The parallel compute performance of the assimilation system is described by the speedup (ratio of the computing time on n processes to the time on one process). The speedup of the assimilation system is dominated by the speedup of the NEMO model itself. The assimilation slightly increases the speedup due to a better scalability.



Summary

The online coupling shows a good computational scalability on supercomputers and is hence well suited for high-dimensional numerical models, including coupled earth system models.

Further, a clear separation of the model and data assimilation components allows to continue the development of both components separately.

Implementations using online coupling have been performed also for other models like FESOM, BSHcmod, HBM, NOBM, ADCIRC, and MITgcm.

PDAF is coded in Fortran with MPI parallelization. It is available as free software. Further information and the source code of PDAF are available on the web site:

http://pdaf.awi.de

References

[1] Cosme E., Brankart J.-M., Verron J., Brasseur P. and Krysta M. (2010). Implementation of a reduced-rank, square-root smoother for high resolution ocean data assimilation. *Ocean Modelling*, 33: 87–100

[2] Nerger, Hiller, and Schröter (2005). PDAF - The Parallel Data Assimilation Framework: Experiences with Kalman Filtering, in *Use of High Performance Computing in Meteorology - Proceedings of the 11th ECMWF Workshop* / Eds. W. Zwieflhofer, G. Mozdzynski. World Scientific, pp. 63–83

[3] Nerger, L. and W. Hiller (2013). Software for Ensemble-based Data Assimilation Systems – Implementation Strategies and Scalability. *Computers & Geosciences*. 55: 110–118

[4] Nerger, L., T. Janjić, J. Schröter, J., and W. Hiller (2012). A unification of ensemble square root Kalman filters. *Mon. Wea. Rev.* 140: 2335–2345