ISDA 2025, Melbourne, Australia, Sept. 29 – Oct. 3, 2025

PDAF Parallel Data Assimilation Framework

The Parallel Data Assimilation Framework (PDAF)

Upgrade to Version 3

Lars Nerger[1], Yumeng Chen[2], Armin Corbin[3], Johannes Keller[4]

1) Alfred Wegener Institute, Bremerhaven, Germany
2) University of Reading and NCEO, UK
3) University of Bonn, Germany
4) Research Center Jülich, Germany

Overview

PDAF

- What it is and how it works
- Latest enhancements in PDAF V3.0
- pyPDAF – PDAF with Python

PDAF – Parallel Data Assimilation Framework

PDAF Parallel Data Assimilation Framework

A unified tool for interdisciplinary data assimilation …

- provide support for parallel ensemble forecasts
- provide DA methods (EnKFs, smoothers, PFs, 3D-Var) - fully-implemented & parallelized
- provide tools for observation handling
- provide diagnostics for state, ensemble and observation space
- easy implementation with (probably) any numerical model (<1 month)
- technically a program library (PDAF-core) plus additional functions & templates
- run from notebooks to supercomputers (Fortran, MPI & OpenMP)
- ensure separation of concerns (model – DA method – observations – covariances)
- first release in year 2004; continuous further development & community involvement

Focus on

- Easy implementation
- Performance for complex models
- Flexibility to extend system

Open source:
Code, documentation, and tutorial available at

https://pdaf.awi.de

github.com/PDAF/PDAF

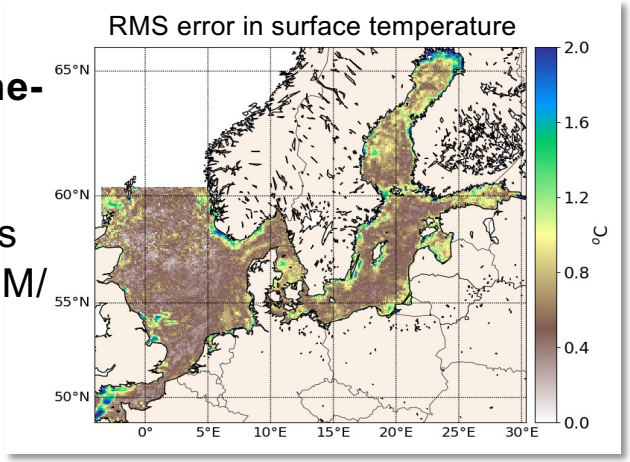
L. Nerger, W. Hiller, Computers & Geosciences 55 (2013) 110-118

AWI

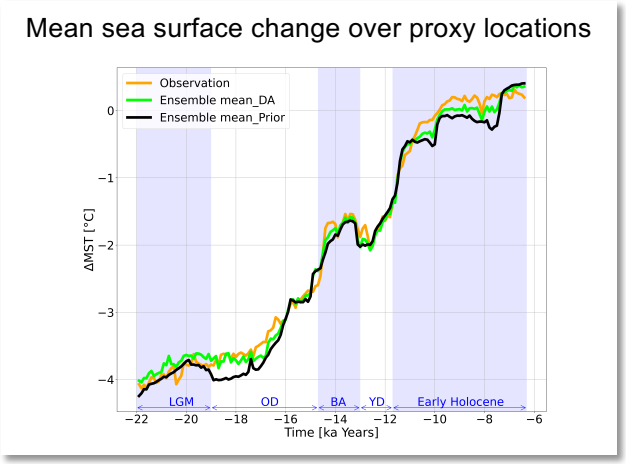
PDAF Application Examples – at AWI

PDAF Parallel Data Assimilation Framework

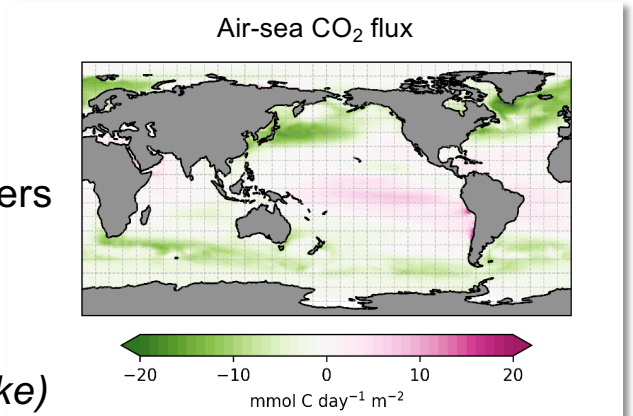
Coastal coupled physics/biogeochemistry DA: CMEMS/BSH - Improving forecasts with NEMO-ERGOM/ HBM-ERGOM: *(S. Vliegen, A. Sathanarayanan)*



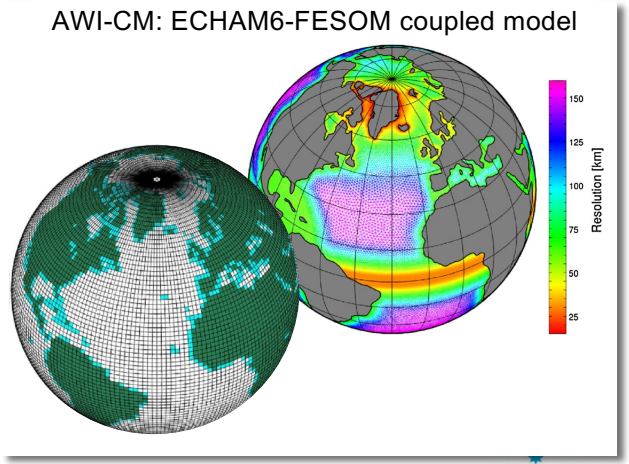
Paleo-climate DA: improve simulation of last deglaciation with CLIMBER-X *(A. Masoum)*



Coupled physics/biogeochemistry DA: Improving parameters & carbon flux in REcoM *(N. Mamnun, F. Bunsen, A. Broschke)*



Coupled ocean-atmosphere DA: strongly + weakly coupled DA of ocean obs. into atmosphere with AWI-CM *(Q. Tang)*



→ Talk by Nabir Mamnun on Thursday

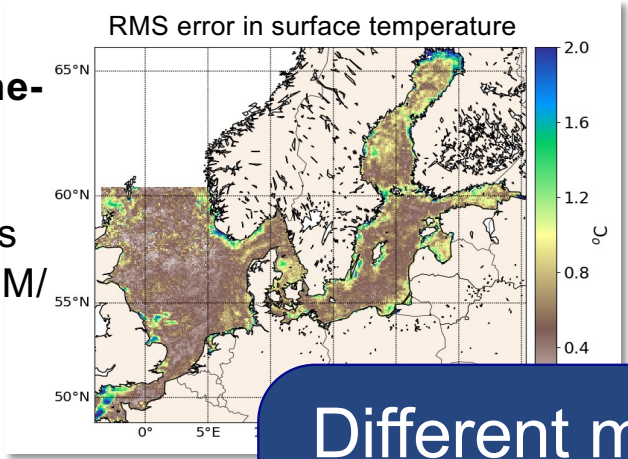
PDAF Release V3.0

AWI

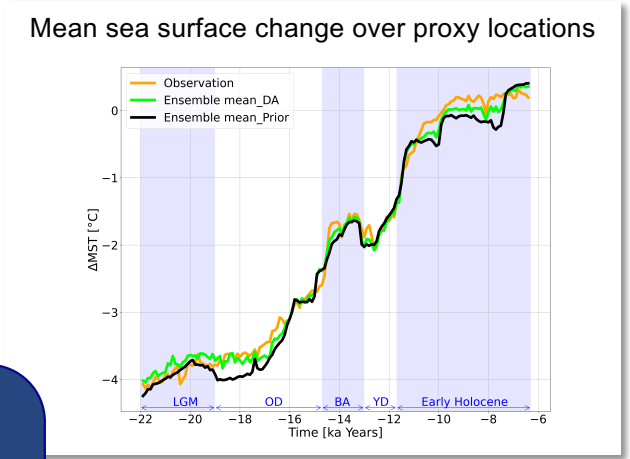
PDAF Application Examples – at AWI

PDAF Parallel Data Assimilation Framework

Coastal coupled physics/biogeochemistry DA: CMEMS/BSH - Improving forecasts with NEMO-ERGOM/ HBM-ERGOM: *(S. Vliegen, A. Sathanarayanan)*



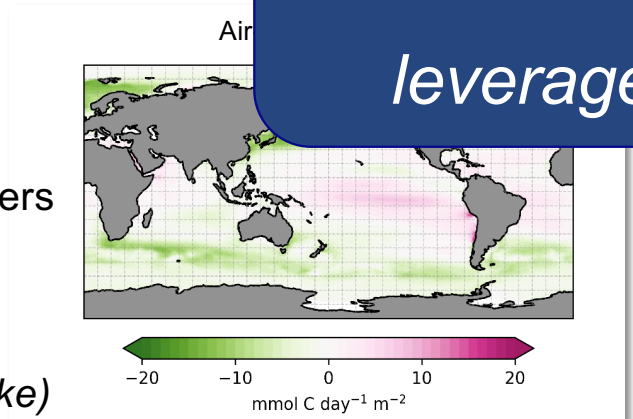
Paleo-climate DA: improve simulation of last deglaciation with CLIMBER-X *(A. Masoum)*



Different models – same assimilation software

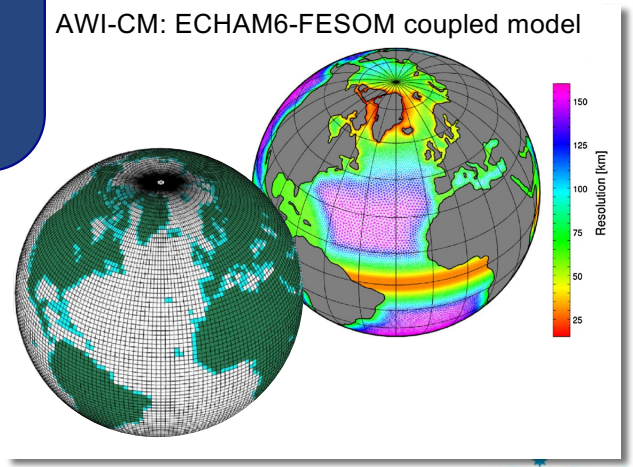
leverage synergies

Coupled physics/biogeochemistry DA: Improving parameters & carbon flux in REcoM *(N. Mamnun, F. Bunsen, A. Broschke)*



→ Talk by Nabir Mamnun on Thursday

strongly + weakly coupled DA of ocean obs. into atmosphere with AWI-CM *(Q. Tang)*



PDAF Release V3.0

AWI

PDAF – wide range of applications

PDAF Parallel Data Assimilation Framework

Applications & users, like

Operational uses:

- *Germany:* North/Baltic Seas (HBM model)
- *Europe:* Copernicus marine forecasting center Baltic Sea (NEMO)
- *China:* Arctic ice-ocean prediction system (MITgcm)

Beyond ocean

- **WRF** (Weather forecast and research model)
- **HydroGeoSphere** (hydrology)
- **TSMP** (Terrestrial Systems Modeling Platform)
- **CLM5** (Community land model)
- **TIE-GCM** (Thermosphere Ionosphere Electrodynamics)
- **VILMA** (Viscoelastic Lithosphere and Mantle Model)
- **Parody** (Geodynamo model)
- **HYSPLIT** (Volcanic Ash Transport and Dispersion model)
- **Cardiatic modeling** (blood flow; pacemakers)
- … more

>20 models coupled to PDAF

Ocean and marine biogeochemistry (research applications)

- **CICE** (sea ice)
- **COAWST** (WRF-ROMS-CICE)
- **FESOM**
- **HBM** (regional HiROMB-BOOS model)
- **MEDUSA** (biogeochemistry)
- **MITgcm**
- **NEMO**
- **REcoM** (biogeochemistry, carbon cycle)
- **SCHISM/ESMF**

Coupled ocean-atmosphere & climate models

- **AWI-CM**
- **MPI-ESM** (ICON-Ocean)
- **CLIMBER-X** (paleo climate model)

PDAF – wide range of applications

PDAF Parallel Data Assimilation Framework

Applications & users, like

Operational uses:

- *Germany:* North/Baltic Seas (HBM model)
- *Europe:* Copernicus marine forecasting center Baltic Sea (NEMO)
- *China:* Arctic ice-ocean predi… (MITgcm)

Beyond ocean

- **WRF** (Weather forecast and…
- **HydroGeoSphere** (hydrolog…
- **TSMP** (Terrestrial Systems M…
- **CLM5** (Community land model)
- **TIE-GCM** (Thermosphere Ionosphere Electrodynamics)
- **VILMA** (Viscoelastic Lithosphere and Mantle Model)
- **Parody** (Geodynamo model)
- **HYSPLIT** (Volcanic Ash Transport and Dispersion model)
- **Cardiatic modeling** (blood flow; pacemakers)
- … more

Different models – same assimilation software

… more synergies

>20 models coupled to PDAF

Ocean and marine biogeochemistry (research applications)

- **CICE** (sea ice)
- **COAWST** (WRF-ROMS-CICE)
- …al HiROMB-BOOS model)
- …ogeochemistry)
- …geochemistry, carbon cycle)
- **SCHISM/ESMF**

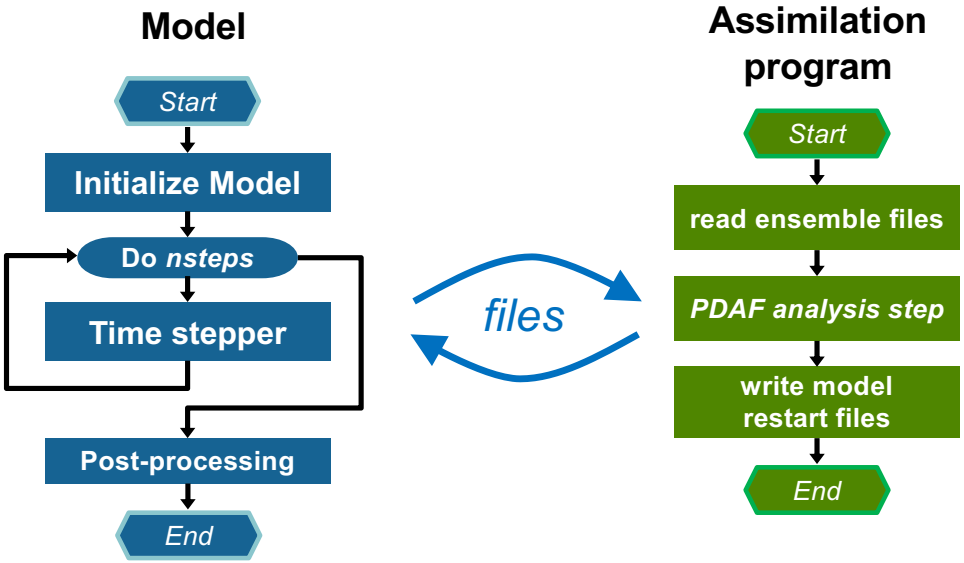
Coupled ocean-atmosphere & climate models

- **AWI-CM**
- **MPI-ESM** (ICON-Ocean)
- **CLIMBER-X** (paleo climate model)

Coupling Model and Assimilation Code: 2 Variants

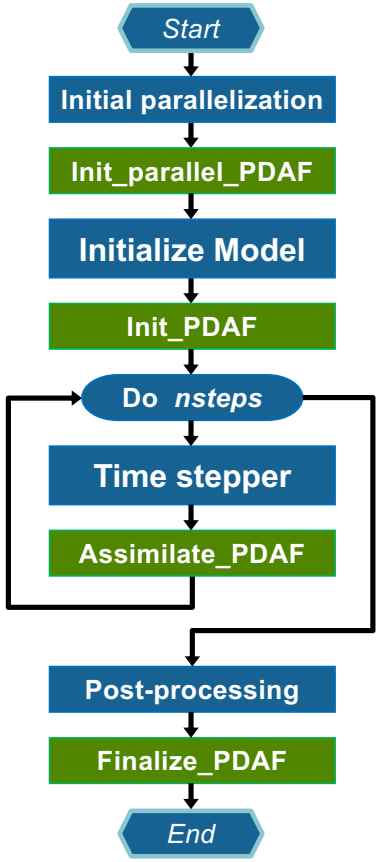
PDAF Parallel Data Assimilation Framework

Offline coupling



- separate programs for model and assimilation
- standard setup at many operational centers
- can be computed by group of compute jobs
- limited computational efficiency
  - model restarts
  - amount of disk IO

Online coupling



- augment model with assimilation functionality (insert PDAF function calls)
- adapt model source code adding 4 routine calls
- easy to implement
- one large compute job
- in-memory data transfers
- computationally very efficient
- For coupled Earth system models: independent of model coupler

Model code | DA code

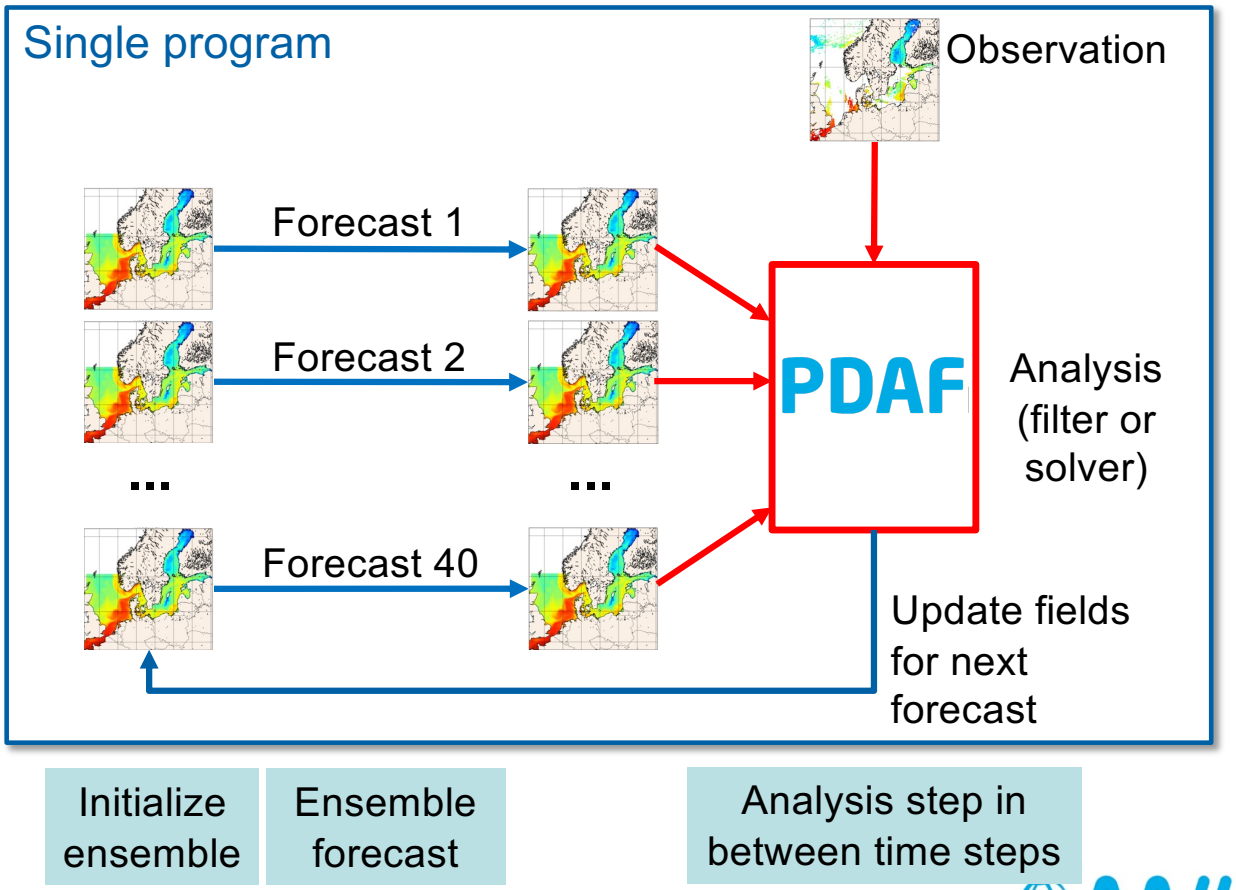
Online-Coupling – Assimilation-enabled Model

PDAF Parallel Data Assimilation Framework

Couple a model with PDAF

- Modify model to simulate ensemble of model states
- Insert analysis step/solver to be executed at prescribed interval
- Run model as usual, but with more processors and additional options (easy to use…)
- Implementation usually below 1 month (often only days)

- EnOI and 3D-Var also possible:
  - Evolve single model state
  - Prescribe ensemble perturbations or covariance

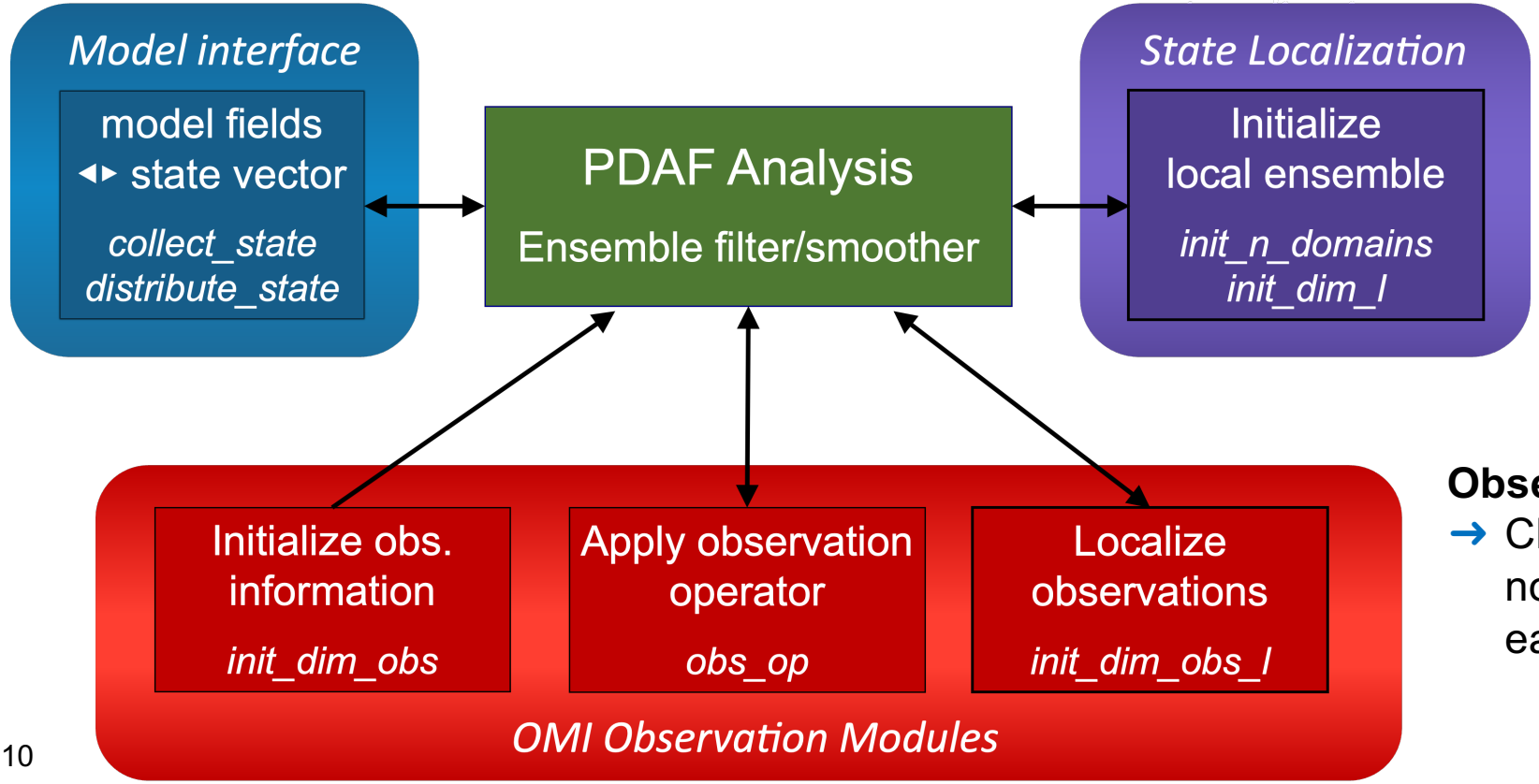


Implementing the Assimilation Analysis/Solver Step

PDAF Parallel Data Assimilation Framework

Abstract interfaces for *user-provided functions*

- Intended to provide specific functionality
- Easy to implement using templates and provided functionality
- 3 sets of of functions for ensemble filters



Observation modules

→ Class-like structure, but not object-oriented for easier implementation

PDAF-OMI (Observation Module Infrastructure)

Provides generic interfaces & functionality for

- **Observation initialization** (set values, errors/R, coordinates, link to model grid)
- **Observation operators** (nonlinear, adjoint and linearized)
  - Functional form: gets state vector, returns observed state vector
- **Localization**
  - Cartesian possible with 1 to n>3 directions
  - 2 and 3 dimensions support geographic coordinates
  - non-isotropic localization supported (fully 3D or factorized vertical-horizontal)
- **Diagnostics** in observation space (e.g. comparing ensemble and observations)

- Use Fortran modules to handle each observation like a class
- Ensures correct parallelization in case of MPI

Release PDAF 3.0

Revision and modernization, unifying 21 years development

- Generally: Fortran 2003+, MPI & OpenMP
- All parts are modularized, but no object-oriented programming

New features (selection)

- Universal PDAF3 interface
  – 1 call for 14 different ensemble filters/smoothers
- Incremental analysis updates (IAU) – generic for any model
- EnSRF & EAKF with serial observation processing
  – PDAF probably the first framework to provide LETKF-like filters, serial observation processing filters, nonlinear filters and 3DVars
- Observation diagnostics for PDAF-OMI (statistics comparing with observed ensemble)
- Additional ensemble diagnostics (e.g. reliability budget)

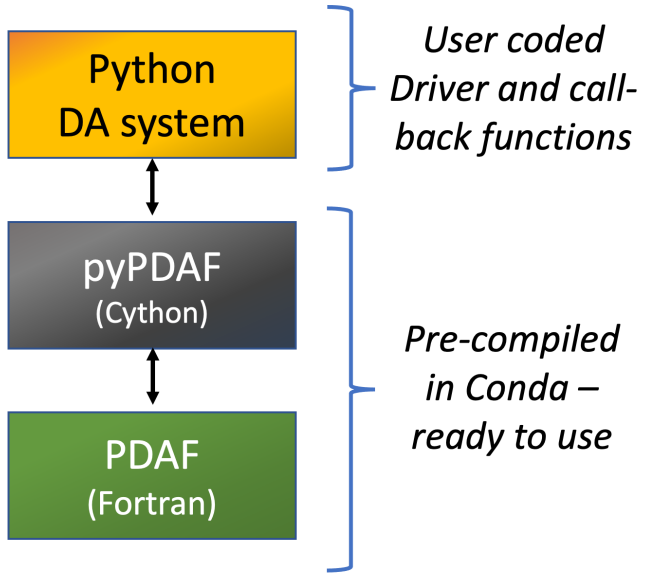
pyPDAF

PDAF Parallel Data Assimilation Framework

Python interface to PDAF - using PDAF as library

- Implement calls to PDAF and user-supplied functions in Python (no Fortran programming)
- analysis step computed inside PDAF (excellent performance due to compiled Fortran code)
- supports
  - Online coupling (e.g. for Python-coded models)
  - Offline coupling (using files from model runs)
  - all PDAF functions can be called
- Allows linkage to Python ML-tools
- installation using Conda

  conda create -n pypdaf -c conda-forge yumengch::pypdaf



github.com/yumench/pyPDAF

Chen, Y., Nerger, L., and Lawless, A. S, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-1078, 2024

AWI

PDAF features and user base

PDAF Parallel Data Assimilation Framework

- Designed to be used with ‘any’ model
- Works with any model grid
- Model agnostic – easy to couple a model to PDAF (usually <1 month)
- Computationally highly efficient and scalable
  (tested on 57600 processor cores; state dimension $860 \cdot 10^9$; $17 \cdot 10^9$ observations)
- Easily handles large number of observation types & complex obs. operators
- Simple code structure for user-sided code
- Internal interface to DA methods - easy to add new methods
- Open-source community software – developments reach a large user group
- Large user base
  - ≧ 150 peer-reviewed papers using PDAF (≧ 200 citing PDAF)
  - ≧ 16 PhD theses completed using PDAF (≧ 5 ongoing)
  - > 20 models coupled to PDAF
  - ~600 users on mail list

DA Methods in PDAF

PDAF originated from comparison studies of different filters

Ensemble Filters and smoothers - *global and localized*

- EnKF, LEnKF (Evensen, 1994, Burgers et al., 1998)
- LETKF, ETKF (Bishop et al., 2001/Hunt et al. 2007)
- LESTKF, ESTKF (Nerger et al., 2012)
- LSEIK, SEIK (Pham, 2001)
- EnsRF (serial obs., Whitaker/Hamill 2002) *new in PDAF V3.0*
- EAKF (serial obs., Anderson, 2003) *new in PDAF V3.0*
- LNETF, NETF (Toedter & Ahrens, 2015)
- Particle filter
- LKNETF (Hybrid Nonlinear Kalman Filter Nerger, 2022)
- *EnOI mode*

3D-Var schemes

- 3D-Var with parameterized covar.
- 3D Ensemble Var
- Hybrid 3D-Var

For details see https://pdaf.awi.de

Summary - PDAF

PDAF Parallel Data Assimilation Framework

- Unified tool for interdisciplinary data assimilation
- Provide range of assimilation methods
  - *Ensemble*: domain localized, serial obs, nonlinear
  - *3DVar*: parameterized, ensemble, hybrid
- Simplifies application of data assimilation
  - Easy coupling to models & implementation of observations
- Computationally very efficient
  - applicable from small models on notebook computers to complex models running on supercomputers
- PDAF 3 brings major code modernization and extension of functionality
- Inviting users to add their models, observations, DA methods
  - Reach a large number of users

https://pdaf.awi.de

Thank you!

Lars.Nerger@awi.de – Data Assimilation and PDAF

AWI