

# The Parallel Data Assimilation Framework (PDAF)

## Upgrade to Version 3

---

Lars Nerger<sup>1</sup>, Yumeng Chen<sup>2</sup>, Armin Corbin<sup>3</sup>, Johannes Keller<sup>4</sup>

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



## DA methodology research – with PDAF

Need functionality for

### With common DA tool:

- Fast usable implementation of new methods
  - Leverage provided functionality to handle ensemble observations, p
- Extensive tests w
  - Use provided t
- Transition to comp
  - Just update ass
- Compare new DA methods with established DA methods
  - Perform experiments with other provided DA methods (e.g. 14 ensemble filters in PDAF) and use provided diagnostics to assess

**Motivation 1:**  
**We need a  
common tool  
for research on data  
assimilation methods**

## DA methodology research – with PDAF

Need functionality for

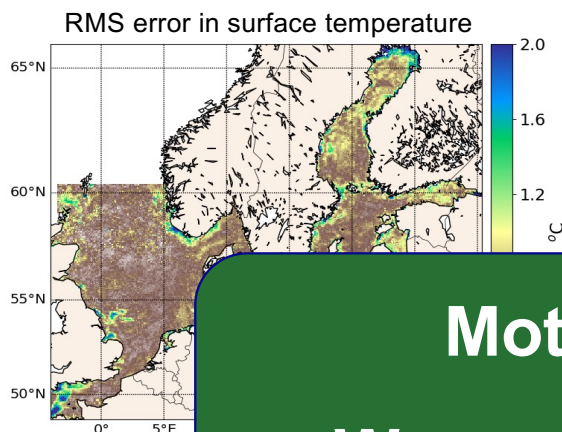
**With common DA tool:**

- Fast usable implementation of new methods  
→ Leverage provided functionality to handle ensemble, observations, parallelization
- Extensive tests with toy models  
→ Use provided toy models fully coded for DA
- Transition to complex models without re-coding  
→ Just update assimilation library
- Compare new DA methods with established DA methods  
→ Perform experiments with other provided DA methods (e.g. 14 ensemble filters in PDAF) and use provided diagnostics to assess

# PDAF Application Examples – at AWI

## Coastal coupled physics/biogeochemistry DA:

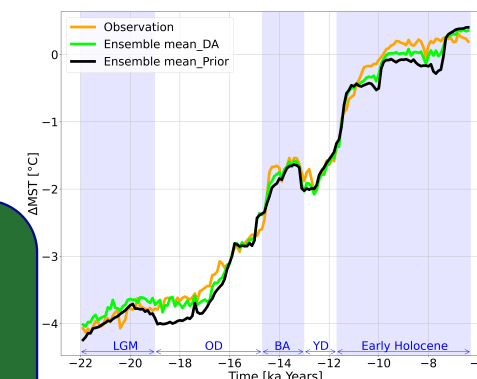
Using operational model systems from CMEMS & BSH



## Paleo-climate DA:

Improve simulation of last deglaciation with

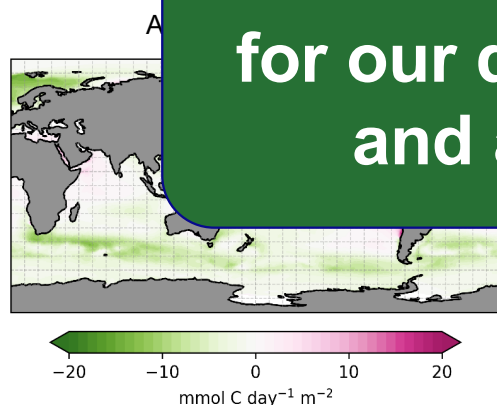
Mean sea surface change over proxy locations



**Motivation 2:**  
**We need a common data assimilation tool for our different models and applications**

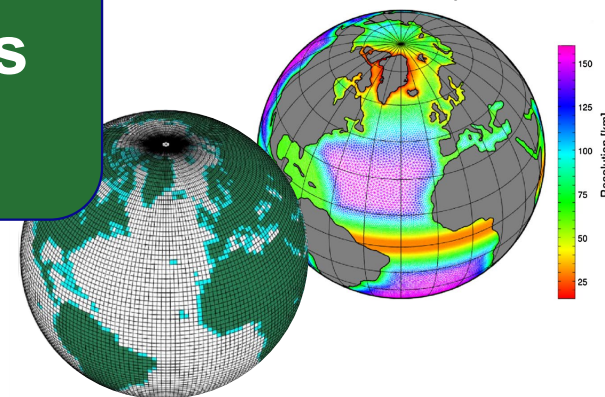
## Global Coupled physics/biogeochemistry DA:

Improving process parameters & ocean carbon flux estimates



ocean obs. into atmosphere with AWI-CM

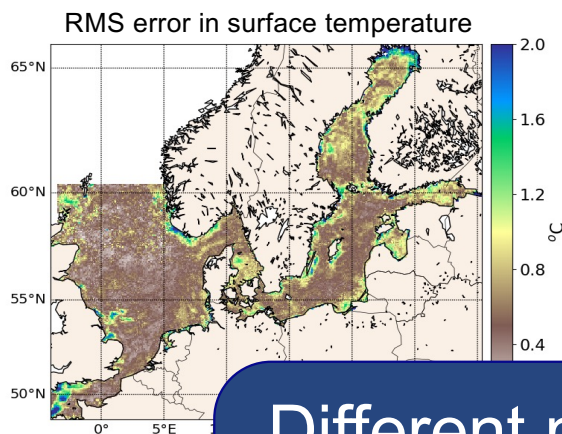
AWI-CM: ECHAM6-FESOM coupled model



# PDAF Application Examples – at AWI

## Coastal coupled physics/biogeochemistry DA:

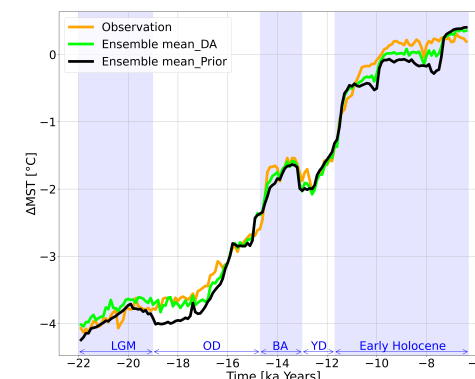
Using operational model systems from CMEMS & BSH



## Paleo-climate DA:

Improve simulation of last deglaciation with CLIMBER-X

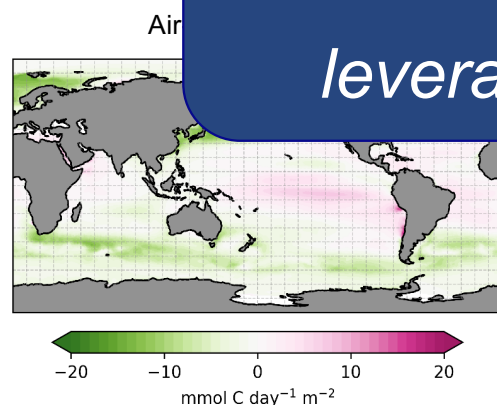
Mean sea surface change over proxy locations



Different models – same assimilation software  
*leverage synergies*

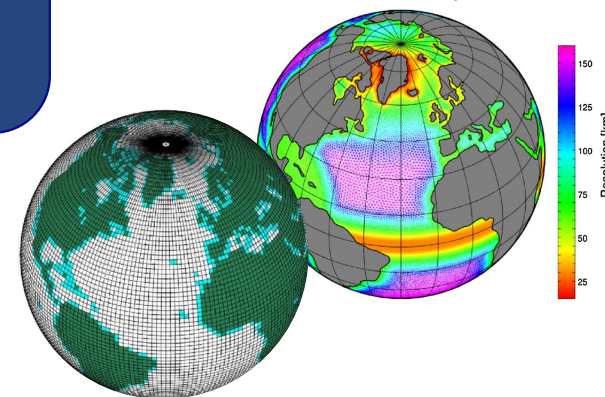
## Global Coupled physics/biogeochemistry DA:

Improving process parameters & ocean carbon flux estimates



strongly + weakly coupled DA of ocean obs. into atmosphere with AWI-CM

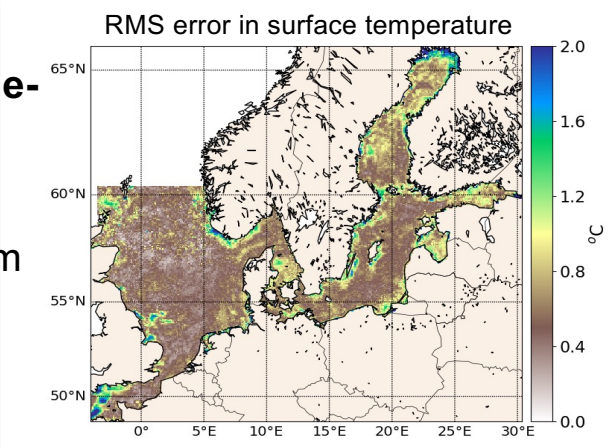
AWI-CM: ECHAM6-FESOM coupled model



# PDAF Application Examples – at AWI

## Coastal coupled physics/biogeochemistry DA:

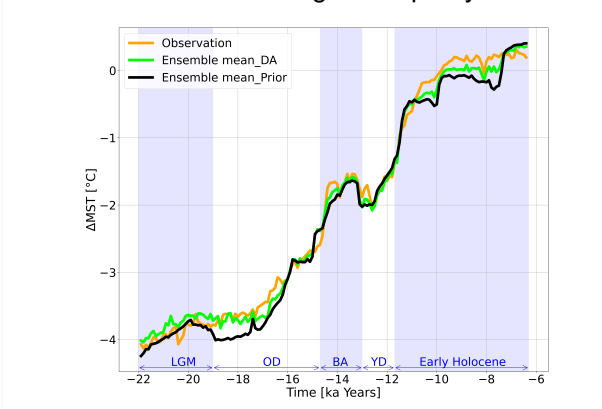
Using operational model systems from CMEMS & BSH



## Paleo-climate DA:

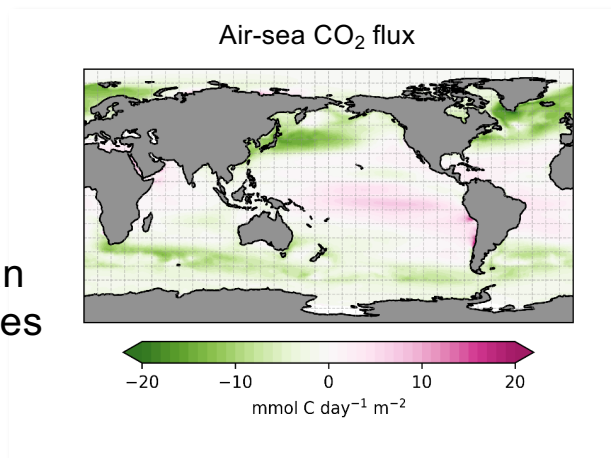
Improve simulation of last deglaciation with CLIMBER-X

Mean sea surface change over proxy locations



## Global Coupled physics/biogeochemistry DA:

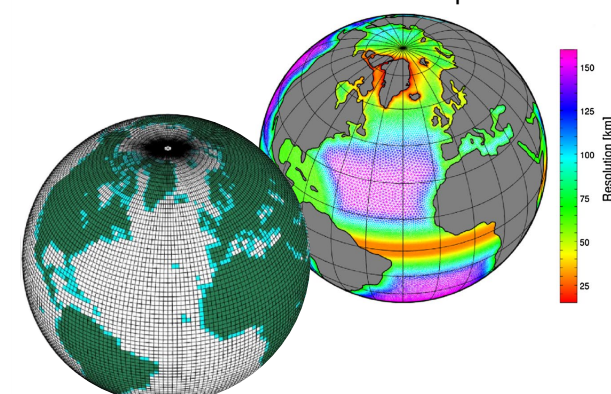
Improving process parameters & ocean carbon flux estimates



## Coupled ocean-atmosphere DA:

strongly + weakly coupled DA of ocean obs. into atmosphere with AWI-CM

AWI-CM: ECHAM6-FESOM coupled model



## PDAF – wide range of applications

### Applications & users, like

#### Operational uses:

- *Germany*: North/Baltic Seas (HBM model), seasonal climate forecasting (MPI-ESM)
- *Europe*: Copernicus marine forecasting center Baltic Sea (NEMO)
- *China*: Arctic ice-ocean pr

#### Beyond ocean

- **WRF** (Weather forecast and reanalysis)
- **HydroGeoSphere** (hydrologic modeling)
- **TSMP** (Terrestrial Systems Modeling)
- **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)
- **Cardiac modeling** (blood flow; pacemakers)
- ... more

**>30 models coupled to PDAF**

Ocean and marine biogeochemistry  
(research applications)

- **CICE** (sea ice) (CFR-ROMS-CICE)
- **ROMB-BOOS** model (biogeochemistry)
- **ROMS-CICE** (biogeochemistry, carbon cycle)

More different models –  
same assimilation software

*... more synergies*

Coupled ocean-atmosphere & climate models

- **AWI-CM**
- **MPI-ESM** (ICON-Ocean)
- **CLIMBER-X** (paleo climate model)
- **GFDL-ESM4**

## Applications & users, like

### Operational uses:

- *Germany*: North/Baltic Seas (HBM model), seasonal climate forecasting (MPI-ESM)
- *Europe*: Copernicus marine forecasting center Baltic Sea (NEMO)
- *China*: Arctic ice-ocean prediction (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)
- **Cardiac modeling** (blood flow; pacemakers)
- ... more

## >30 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)
- **GFDL-ESM4**

## 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

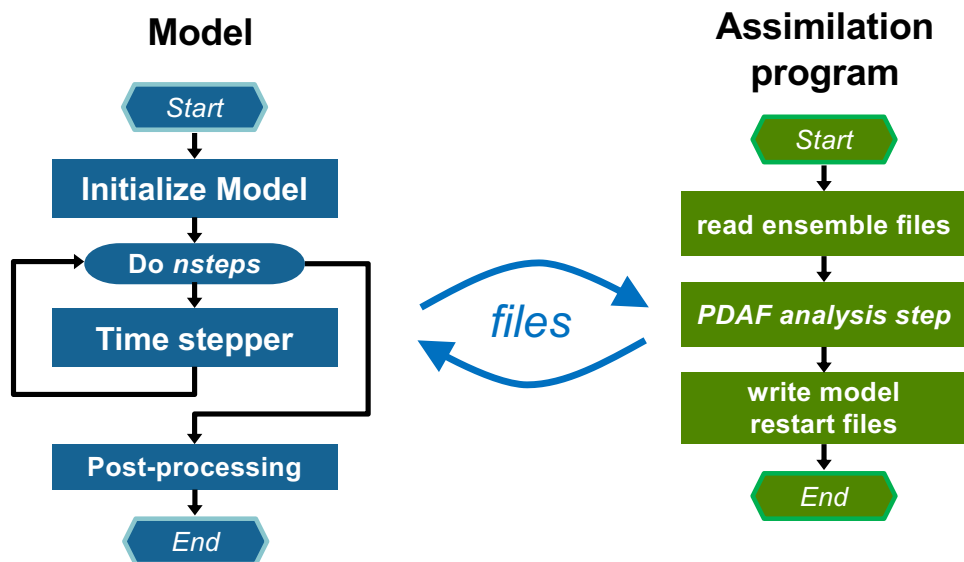
**Open source:**  
Code, documentation, and tutorial available at  
<https://pdaf.awi.de>

*[github.com/PDAF/PDAF](https://github.com/PDAF/PDAF)*



## Coupling Model and Assimilation Code: 2 Variants

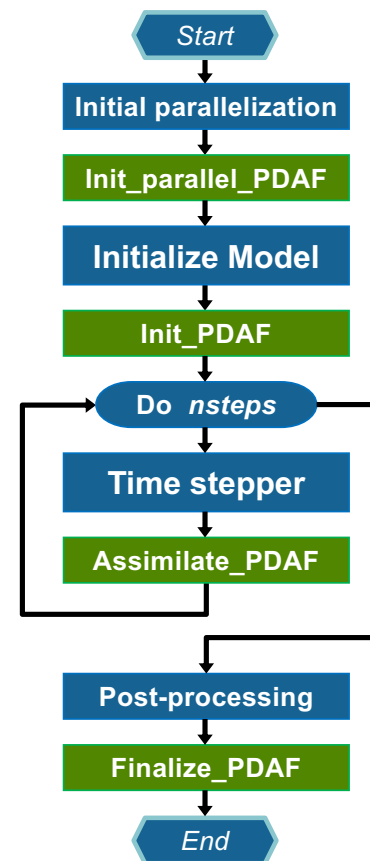
### Offline coupling



- 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

Model code

### single program



DA code

### Online coupling

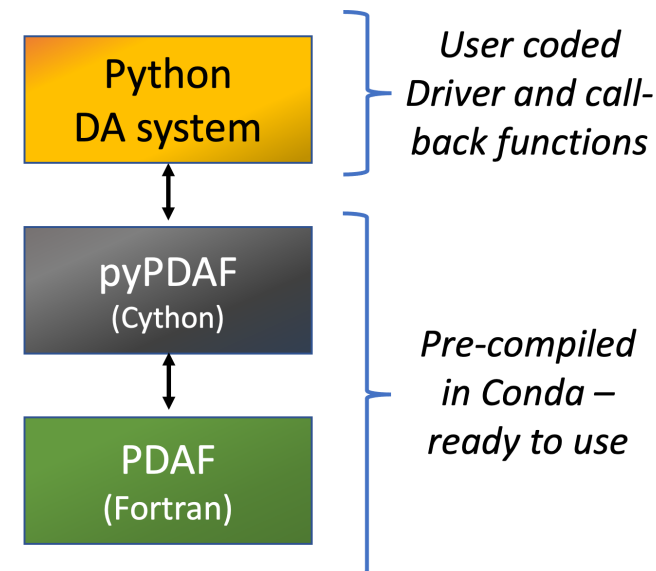
- Create an assimilative model
- Minimal adaptations to model source code: add 4 function calls for PDAF
- easy to implement
- one large compute job
- in-memory data transfers
- computationally very efficient

## 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 library (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](https://github.com/yumench/pyPDAF)



## Release PDAF V3.0 (+ V3.1 released April 16)

---

### Revision and modernization, unifying 21 years development

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

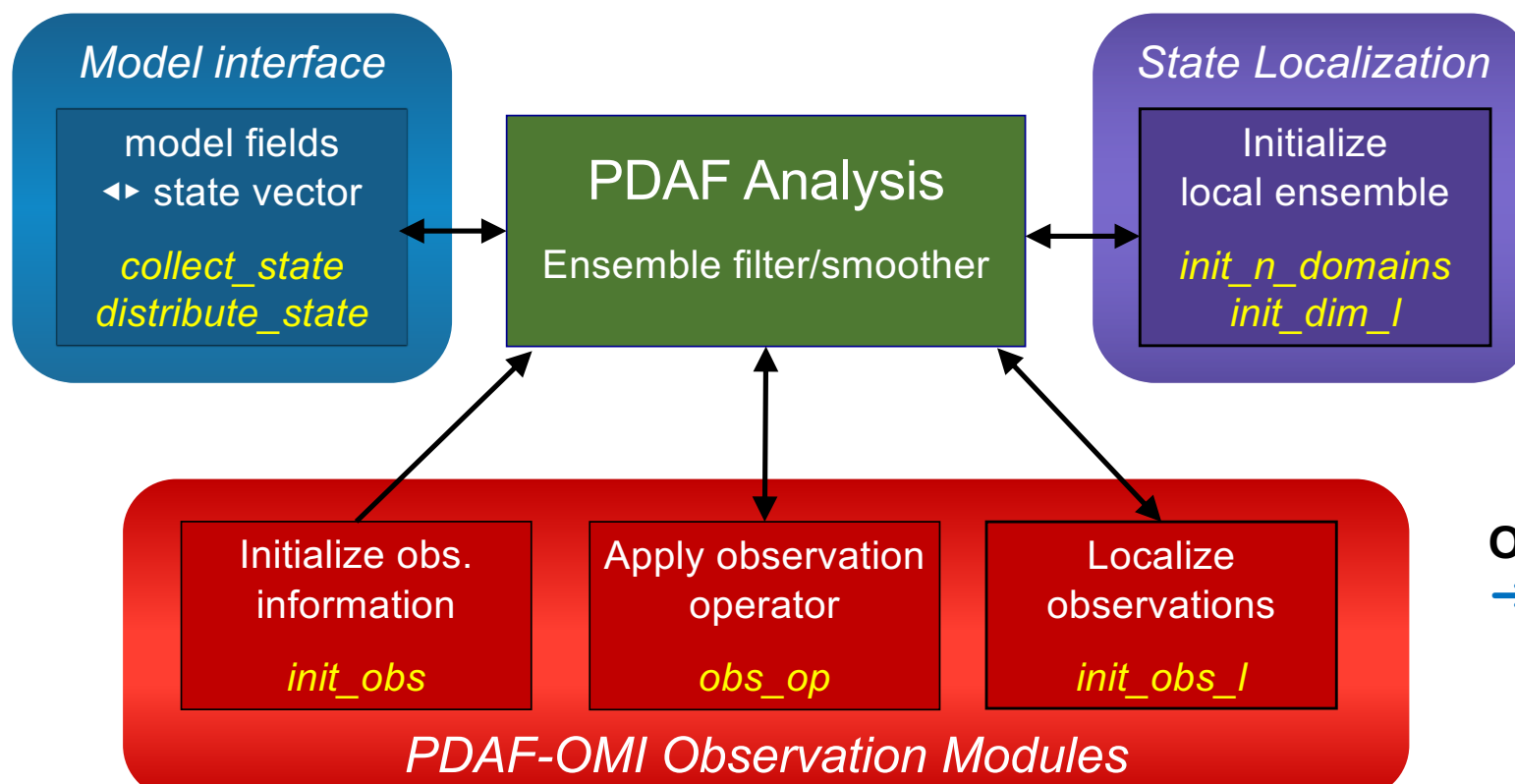
### New features (selection)

- Universal PDAF3 interface
  - 1 function for 14 different ensemble filters/smoother
  - 1 function for 5 different 3D-Var schemes
- EnSRF & EAKF with serial observation processing + related observation handling
  - PDAF probably the first framework to provide LETKF-like filters, serial observation processing filters, nonlinear filters, smoothers and 3D-Vars
- Incremental analysis updates (IAU) – generic for any model
- Observation diagnostics (statistics comparing with observed ensemble)
- Additional ensemble diagnostics (e.g. reliability budget)

## Implementing the Assimilation Analysis/Solver Step

PDAF calls *user-provided functions*

- Intended to provide specific functionality
- Easy to implement using templates and provided functionality



### Observation modules

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

## Summary - PDAF



Thank you!

- 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
- Simplifies development and distribution of new DA methods
- 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

**Post-doc in Ocean Data Assimilation**

at 

European project with 11 partners

*Advance ensemble DA methods*

*Contribute to advance PDAF*

*Apply ensemble DA to real ocean model*

<https://tinyurl.com/Postdoc-AWI>



Application deadline: May 22

<https://pdaf.awi.de>

## 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)
- EAKF (serial obs., Anderson, 2003)
- 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

### Chaotic test models

(full implementations with PDAF)

- Lorenz-96
- Lorenz-63
- Lorenz-2005 models II and III

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

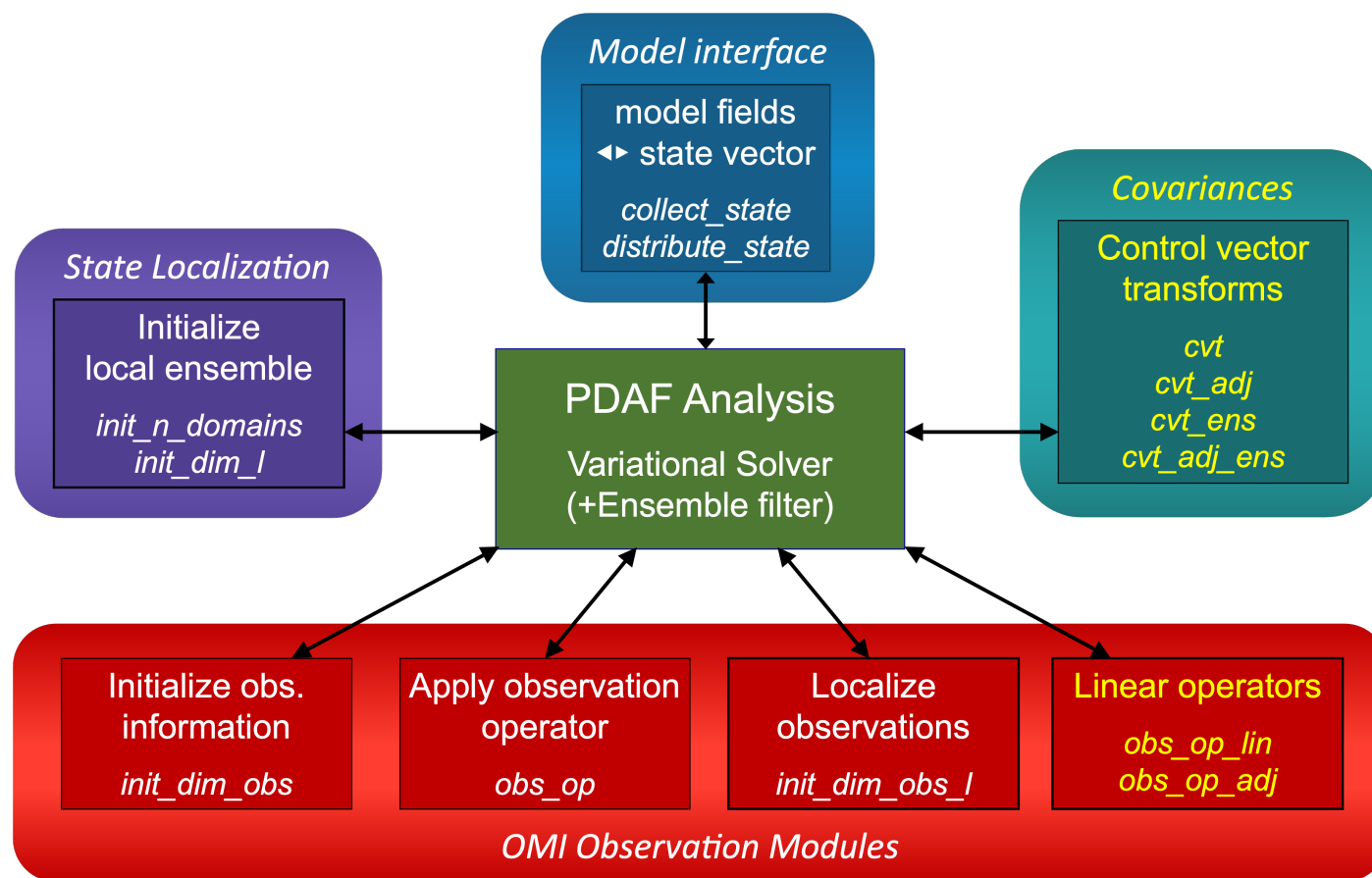
## PDAF features and user base

- 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
  - $\geq 150$  peer-reviewed papers using PDAF ( $\geq 200$  citing PDAF)
  - $\geq 16$  PhD theses completed using PDAF ( $\geq 5$  ongoing)
  - > 20 models coupled to PDAF
  - ~600 users on mail list

## Implementing the Assimilation Analysis/Solver Step

### Abstract interfaces

- Analysis steps needs user-provided routines
- Easy to implement using templates and provided functionality



**Yellow:**  
Additional routines for  
3DVars