EGU General Assembly 2024

19 Apr 2024

PDAF Parallel Data Assimilation Framework

SC5.12 Getting Started with Data Assimilation: Theory and Application

Qi Tang[1], Lars Nerger[2], Armin Corbin[3], Nabir Mamnun[4], Yumeng Chen[5]

[1]University of Neuchâtel, Centre for Hydrogeology and Geothermics (CHYN), Switzerland

[2]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany

[3]University of Bonn, Institute for Geodesy and Geoinformation, Astronomical, Physical and Mathematical Geodesy Group, Germany

[4]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany

[5]University of Reading, National Centre for Earth Observation, Department of Meteorology, United Kingdom of Great Britain

unine Université de Neuchâtel

AWI

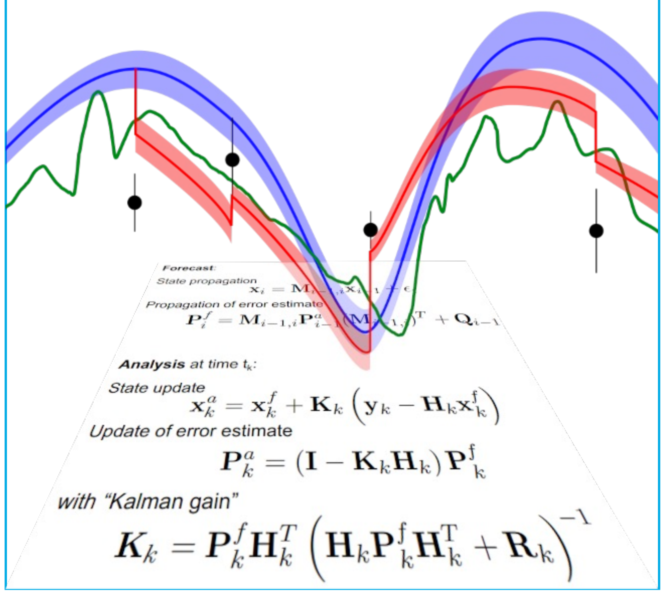
UNIVERSITÄT BONN

University of Reading

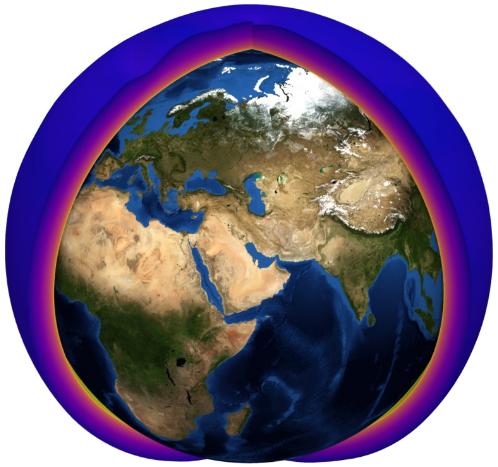
Schedule

PDAF Parallel
Data Assimilation
Framework

I Theory (25 min)



II Applications (20 min)



III Hands-on (45 min)

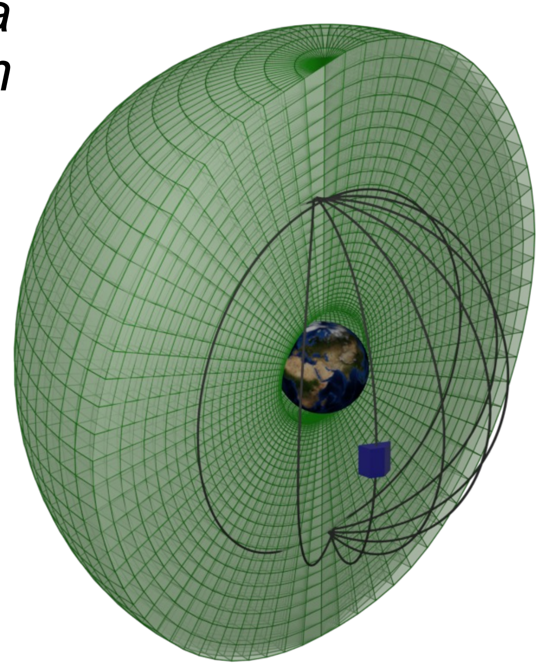
```
     ___        _____         ___         ___
    /  /\      /  /::\       /  /\       /  /\
   /  /::\    /  /:/\:\     /  /::\     /  /:/_
  /  /:/\:\  /  /:/  \:\   /  /:/\:\   /  /:/ /\
 /  /:/~/:/ /__/:/ \__\:| /  /:/~/::\ /  /:/ /:/
/__/:/ /:/  \  \:\ /  /:/ /__/:/ /:/\:\/__/:/ /:/
\  \:\/:/    \  \:\  /:/  \  \:\/:/__\/\  \:\/:/
 \  \::/      \  \:\/:/    \  \::/      \  \::/
  \  \:\       \  \::/      \  \:\       \  \:\
   \  \:\       \__\/        \  \:\       \  \:\
    \__\/                     \__\/        \__\/
```

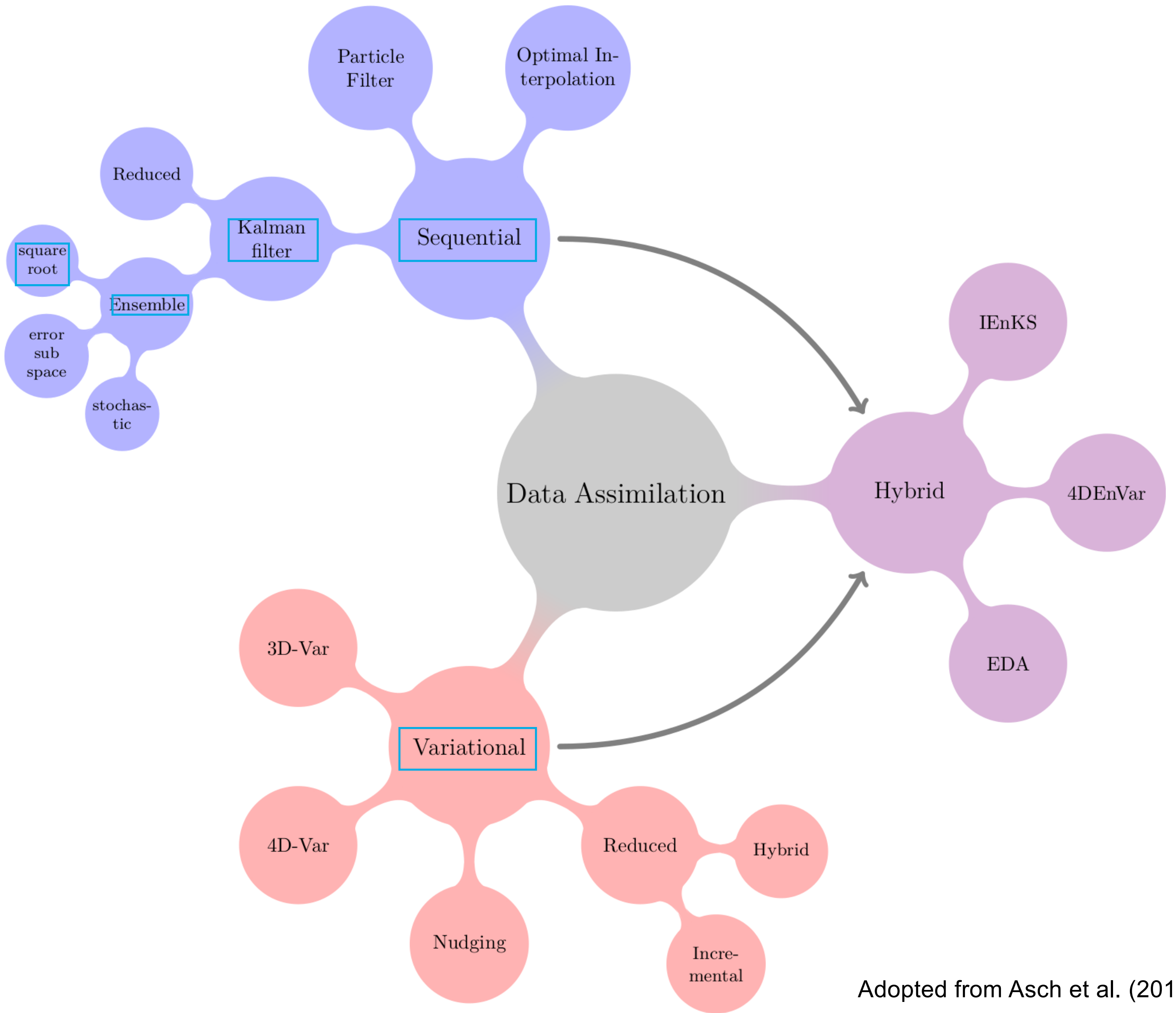
I Theory

Data Assimilation (DA)

Data assimilation (DA) is the science of ***combining observations*** *of a system,* ***including their uncertainty****, with estimates of that system from a dynamical* ***model****, including its* ***uncertainty****, to obtain a new and more accurate description of the system including an uncertainty estimate of that description.*

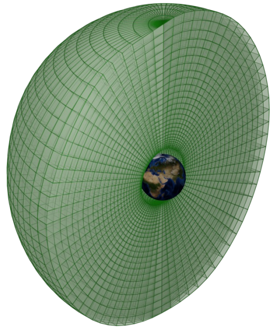
Vetra-Carvalho et al. (2018)





Adopted from Asch et al. (2016)

Requirements for DA



1. Model
   - With some skill

2. Observations
   - With finite errors
   - Related to model fields

3. Data assimilation method

PDAF Parallel Data Assimilation Framework

Observation Operator

PDAF Parallel Data Assimilation Framework

observations — state — observation errors

$$y(t) = \mathcal{H}\left(x(t)\right) + \varepsilon(t)$$

observation operator: maps state to observation

Linearized Operator: $$H = \left.\frac{\partial \mathcal{H}(x)}{\partial x}\right|_{x=x(t)}$$

Model Operator

state

model errors

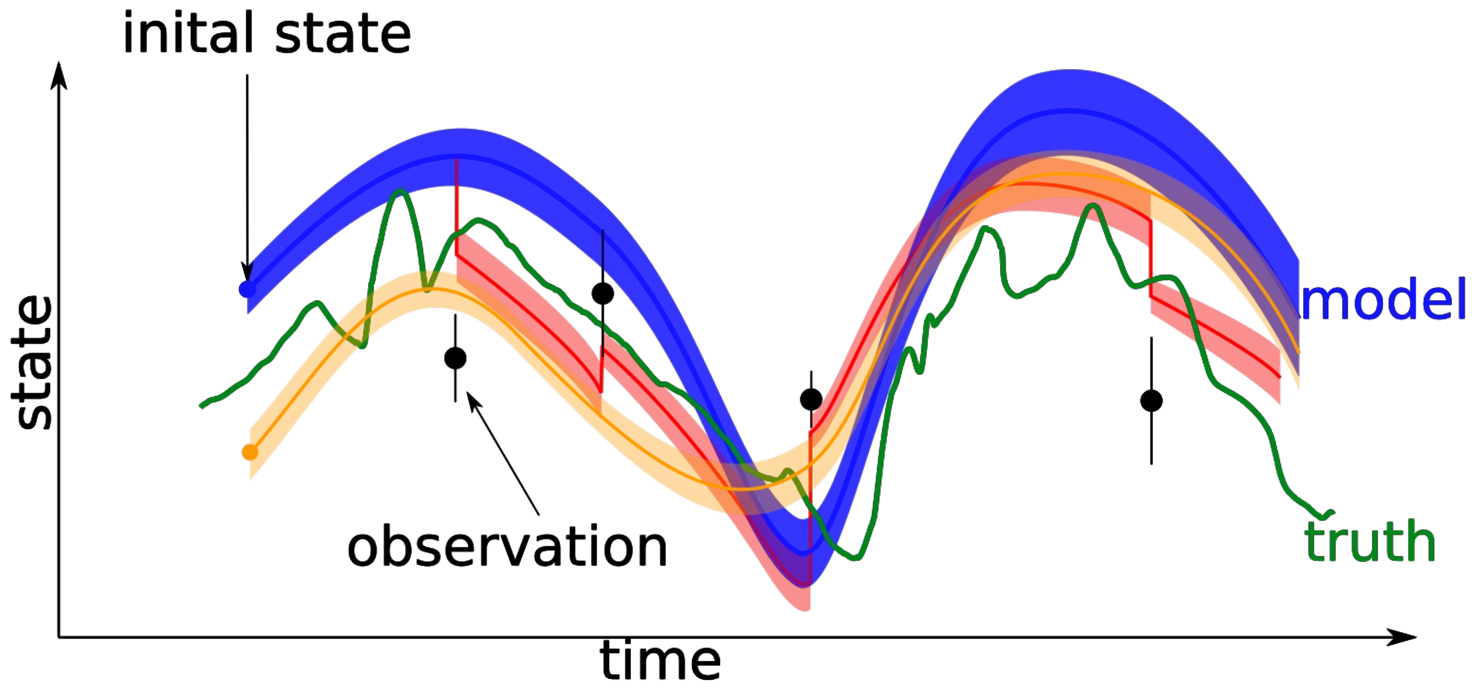
$$x(t) = \mathcal{M}_{s,t}(x(s)) + \eta(t)$$

model/forward operator: propagates state from time s to t

Linearized Operator: $$M_{s,t} = \left.\frac{\partial \mathcal{M}_{s,t}(x)}{\partial x}\right|_{x=x(s)}$$

Sequential and Variational DA

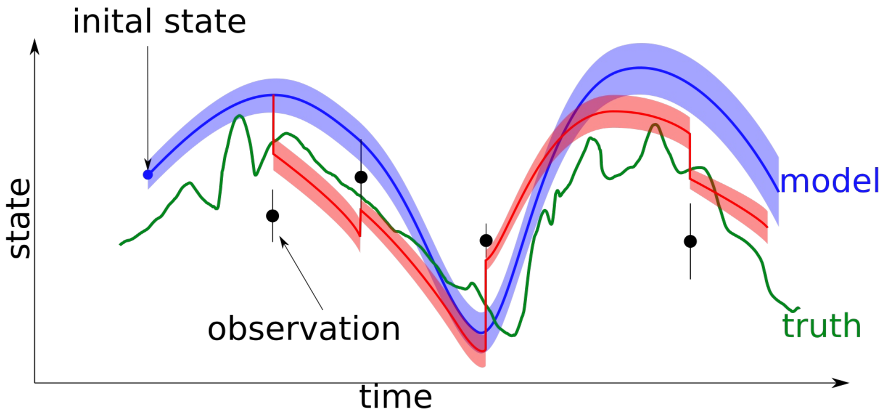
PDAF Parallel Data Assimilation Framework



Optimization

PDAF Parallel Data Assimilation Framework

Sequential



forecast

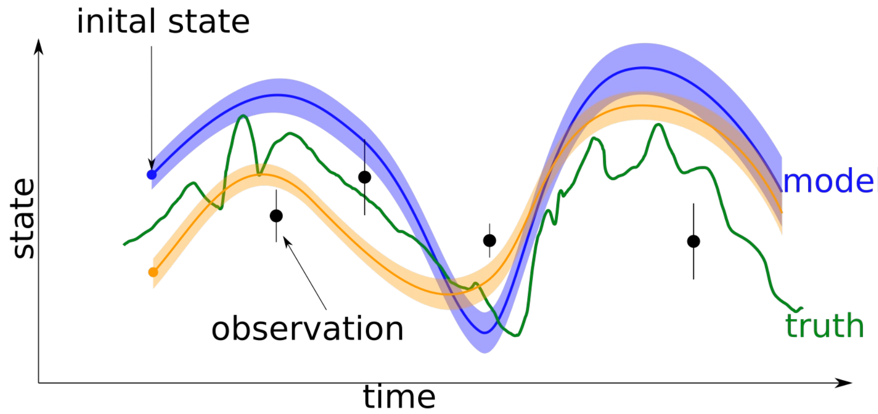
observations

$$J(x) = \left|x - x_i^f\right|^2_{\left(P_i^f\right)^{-1}} + \left|\mathcal{H}(x) - y_i^o\right|^2_{(R_i)^{-1}}$$

state

observation operator

Variational (4D Var)



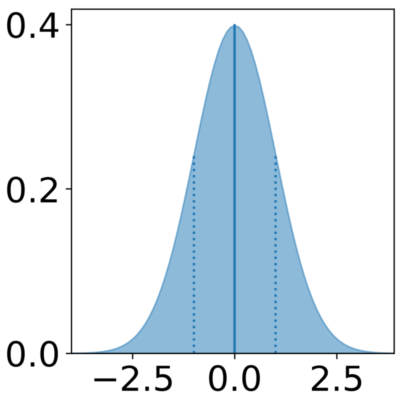
$$J(x_0) = \left|x_0 - x_0^b\right|^2_{(P_0)^{-1}} + \sum_{i=0}^{n} \left|\mathcal{H}(x_i) - y_i^o\right|^2_{(R_i)^{-1}}$$

Kalman filter is optimal

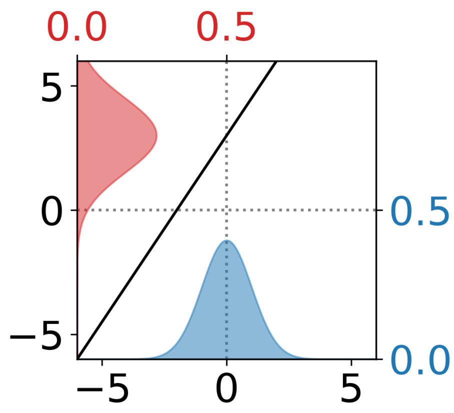
Optimal: state is **unbiased** and has **minimal variance**

Assumptions:

1. everything is Gaussian



2. model and observation operator are linear



3. model errors are not correlated with state or observation errors

$$\begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix}$$

Kalman Filter

PDAF Parallel Data Assimilation Framework

1. Forecast/Prediction

State propagation

$$x_i = M_{i-1,i} x_{i-1} + \varepsilon_i$$

Propagation of error estimate

$$P_i^f = \boxed{M_{i-1,i} P_{i-1}^a M_{i-1,i}^T} + Q_{i-1}$$

1. M explicitly required
2. Scales poorly with the size of the problem

2. Analysis/Update at time t_k

State update

$$x_k^a = x_k^f + K_k (y_k^o - H_k x_k^f)$$

Propagation of error estimate

$$P_k^a = (I - K_k H_k) P_k^f$$

with Kalman gain

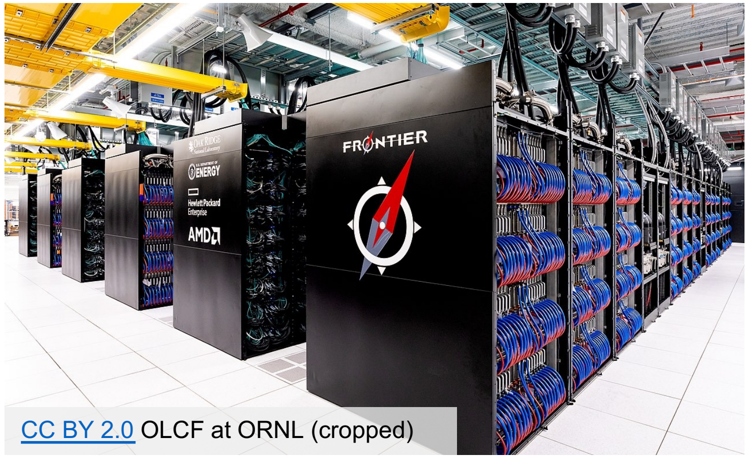
$$K_k = P_k^f H_k^T \left(H_k P_k^f H_k^T + R_k \right)^{-1}$$

Large Scale Models

PDAF Parallel Data Assimilation Framework

Dimension of state vector	Required memory (double)	
	x	P
1 000 000	8 MB	8 TB
10 000 000	80 MB	800 TB
100 000 000	800 MB	80 000 TB

Combined memory of best super computer[1] accumulates to **9 200 TB**

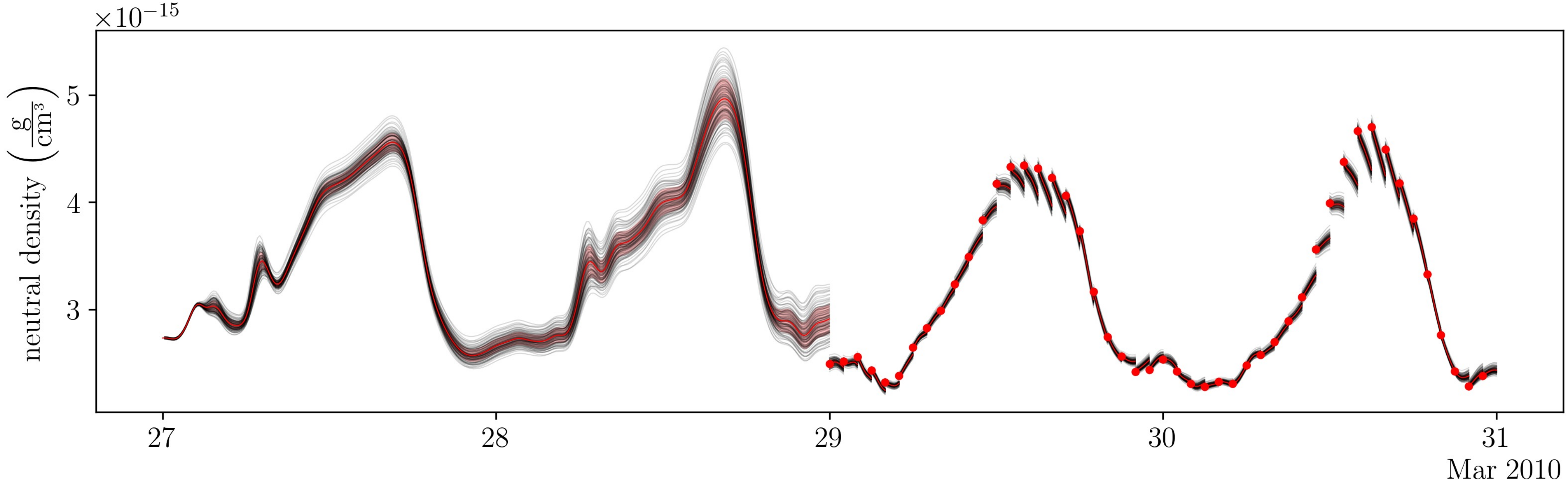


Storing and multiplication of the covariance matrix of the state becomes too expensive

[1]According to TOP500 List of Nov 2023

Ensemble Kalman Filters

PDAF Parallel Data Assimilation Framework



ensemble matrix

$$X = [x_1\, x_2\, \cdots\, x_n]$$

ensemble mean

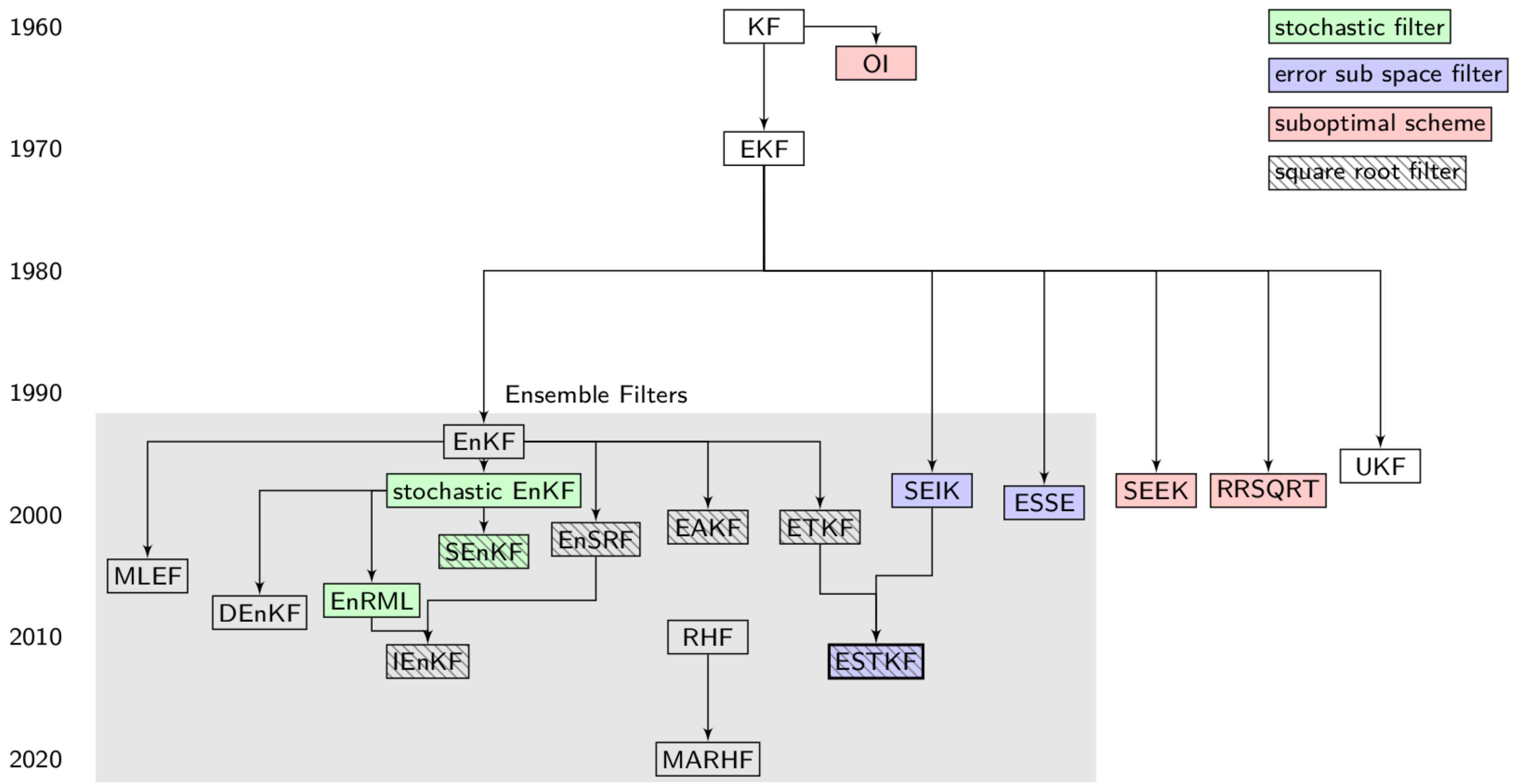
$$\bar{x} = \frac{1}{n} X I$$

ensemble variance

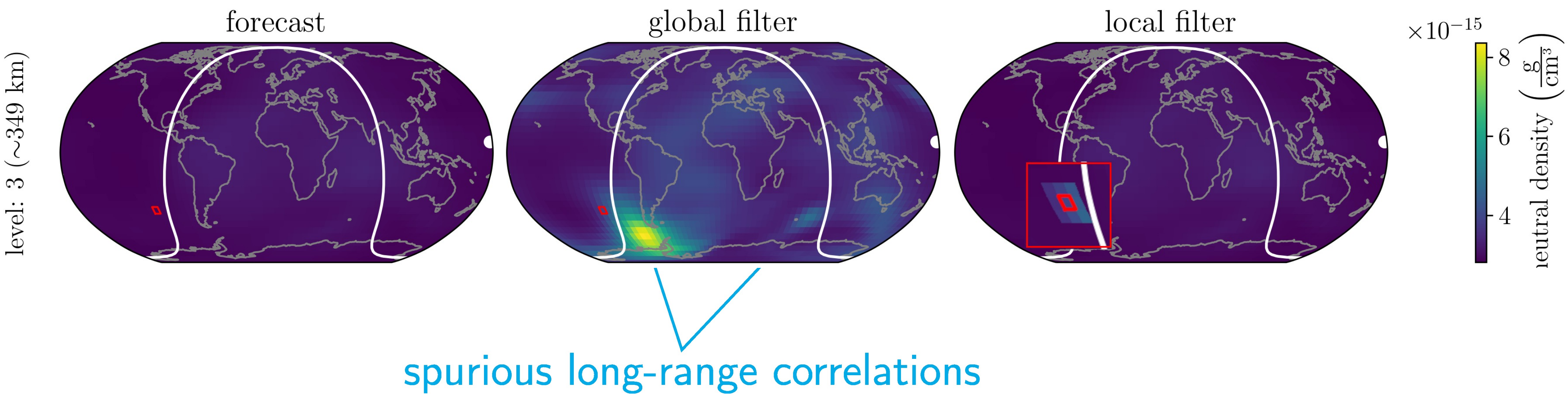
$$P^f \approx \frac{1}{n-1}(X - \overline{X})(X - \overline{X})^T$$

Filters

PDAF Parallel Data Assimilation Framework

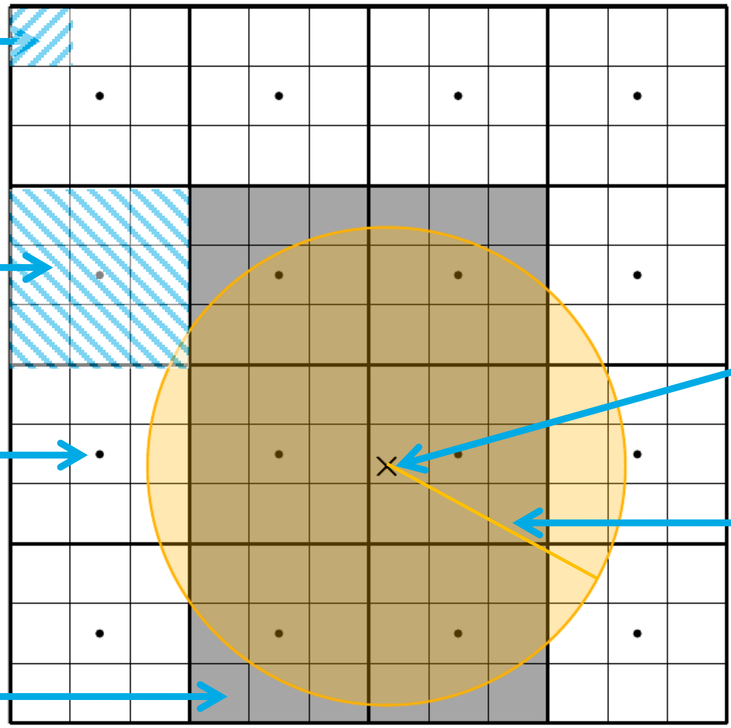


Localization



Domain Localization

- subdivide model into disjoint sub-domains
- update each sub-domain individually taking only observations within specific distance into account

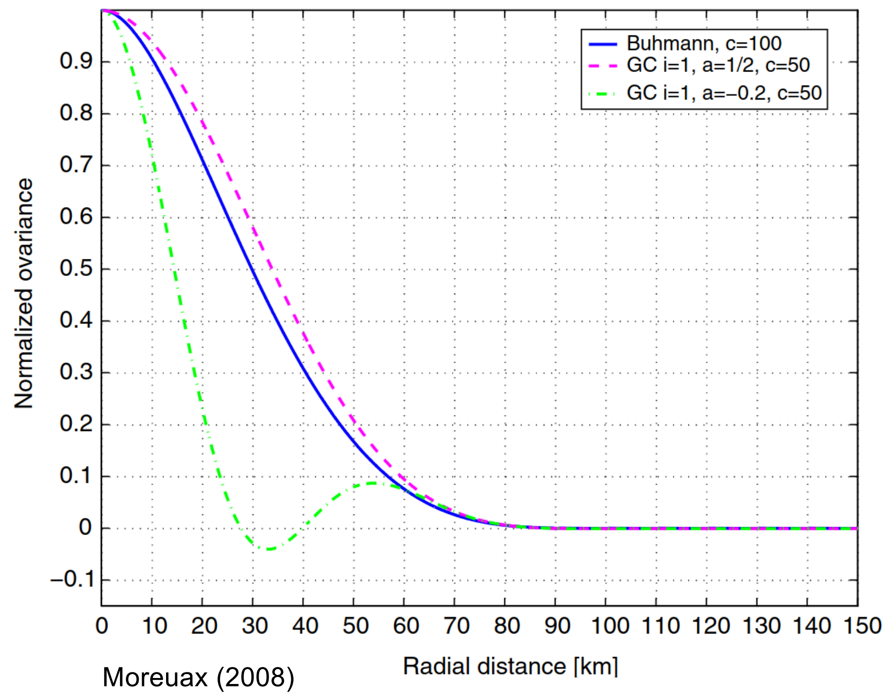


Covariance Localization

PDAF Parallel Data Assimilation Framework

- Multiply covariance matrix of forecasted state with finite covariance function

properties of auto covariance functions
- positive semi-definite
- $f(0) \geq 0$
- $|f(x)| \leq f(0)$
- $f(-x) = f(x)$



Moreuax (2008)

Inflation

PDAF Parallel Data Assimilation Framework

- True variance is always underestimated
  - small ensemble size
  - sampling errors (unknown structure of P)
  - model errors

→ can lead to filter divergence

- Simple remedy

→ Increase error estimate before analysis

- Inflation
  - Increase ensemble spread by constant factor
  - Some filters allow multiplication of a small matrix
  - Needs to be experimentally tuned

Co-Estimation of Model Dynamics (Model Calibration)

PDAF Parallel Data Assimilation Framework

augment state vector with model parameters

state

$$z = \begin{bmatrix} x \\ c \end{bmatrix}$$

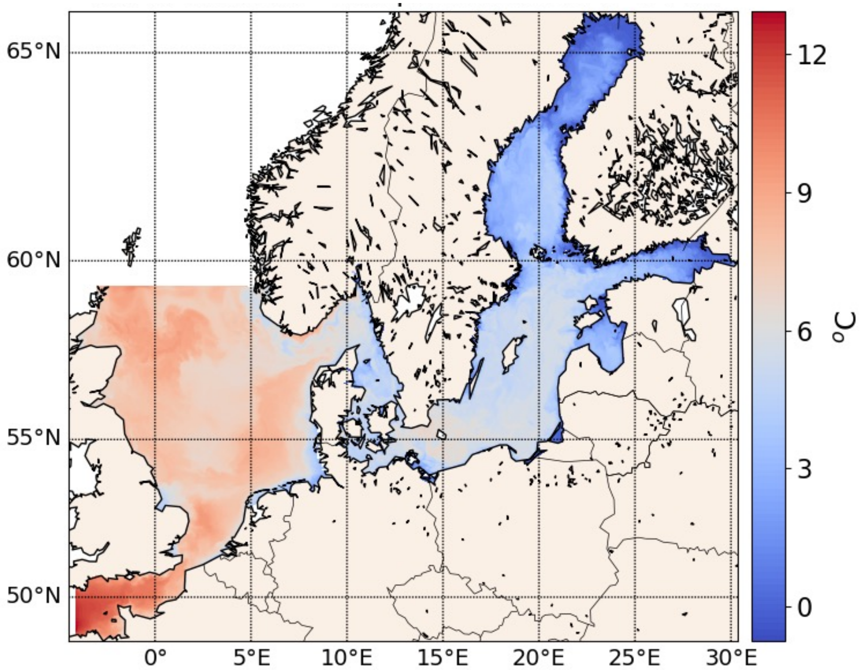
model parameters

II Applications

Coastal Ocean DA

Improving forecasts of **ocean physics** and **biogeochemistry**

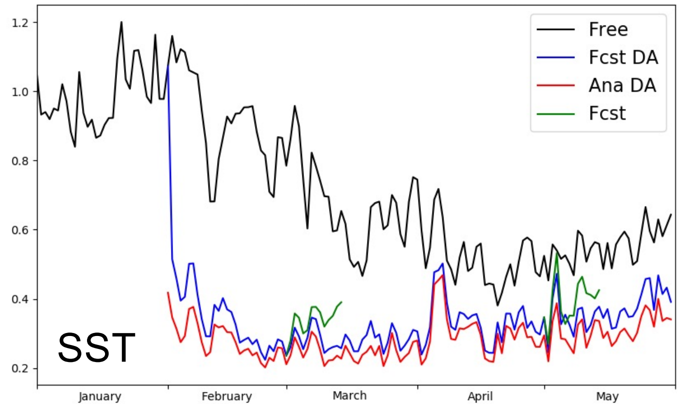
Model domain: North Sea and Baltic Sea
1.8km resolution, 56 layers

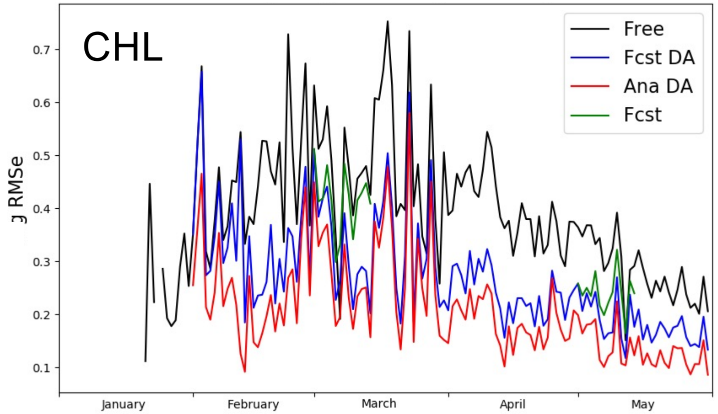


Surface temperature:
RMSe in Baltic Sea

Black: no DA
Blue: 1-day forecasts
Red: analysis
Green: 14-day forecasts

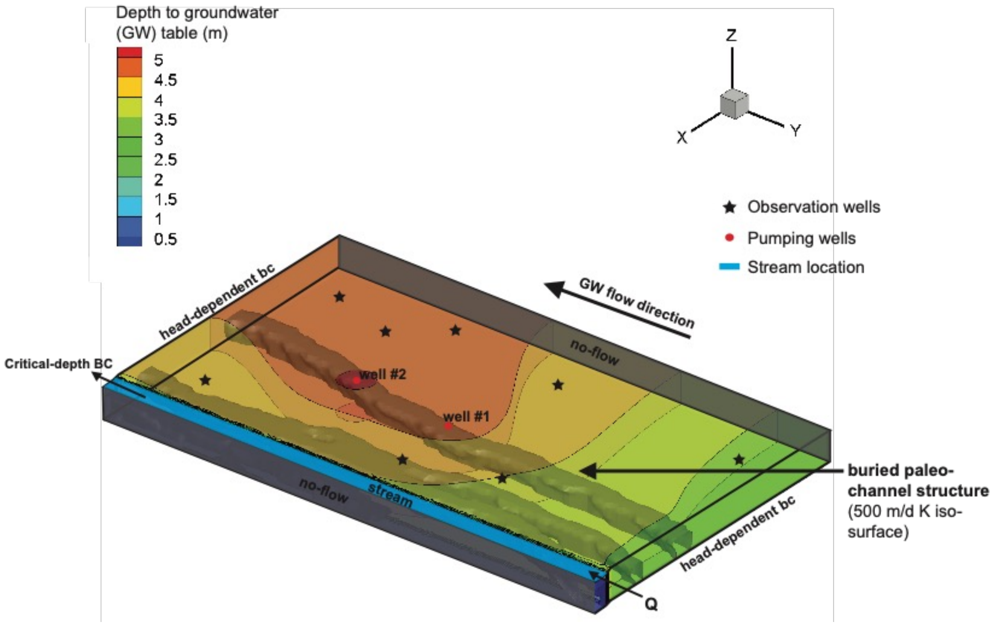
Chlorophyll:
Log10-RMSe
in Baltic Sea

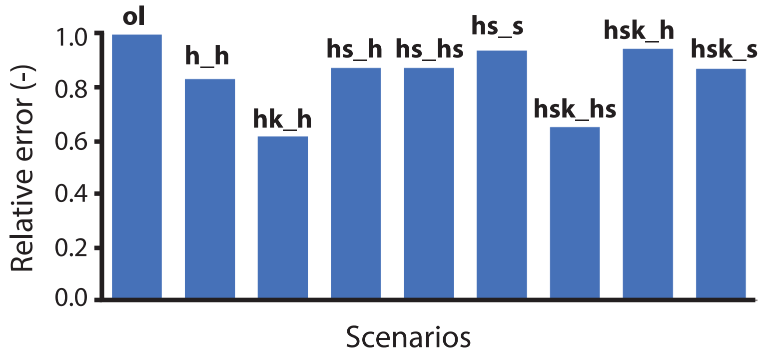


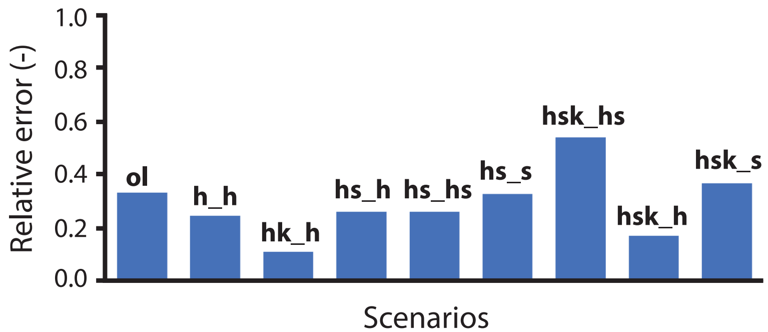


HGS-PDAF

A modular data assimilation framework for an integrated surface and subsurface **hydrological** model





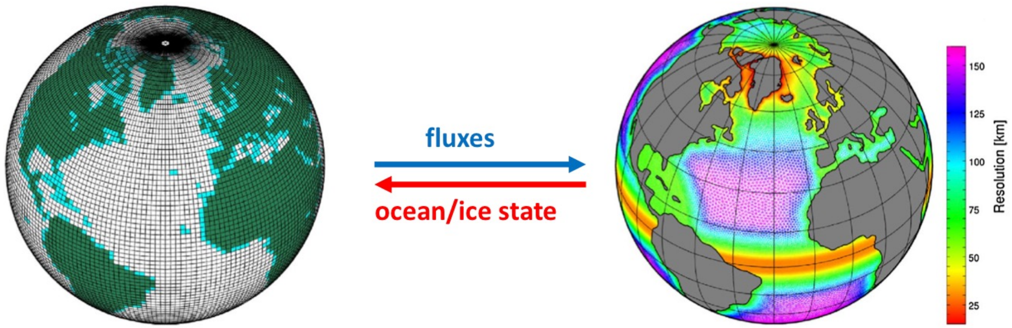


Tang et al. (2023): HGS-PDAF (version 1.0): A modular data assimilation framework for an integrated surface and subsurface hydrological model, Geosci. Model Dev. Discuss.

PDAF Parallel Data Assimilation Framework

AWI-CM-PDAF

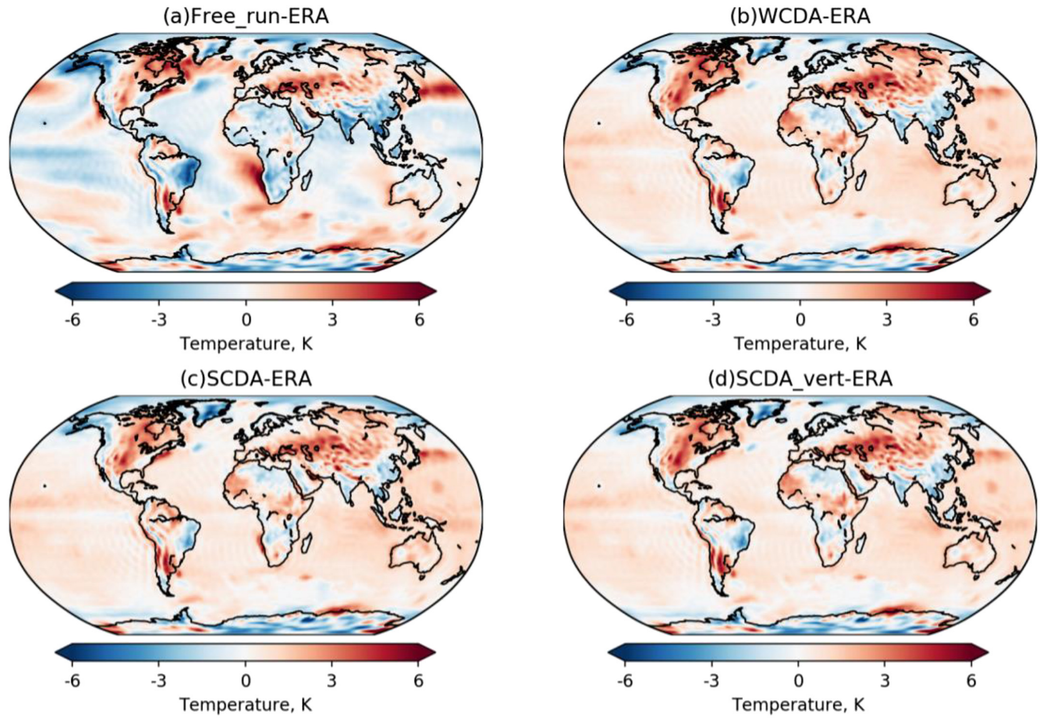
A data assimilation framework for coupled **ocean-atmosphere** models



Atmosphere
ECHAM6
JSBACH land

Coupler library
OASIS3-MCT

Ocean
FESOM
includes sea ice

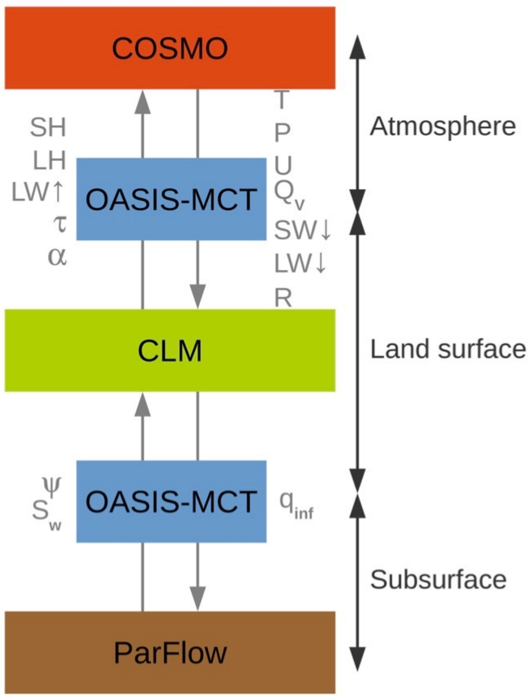


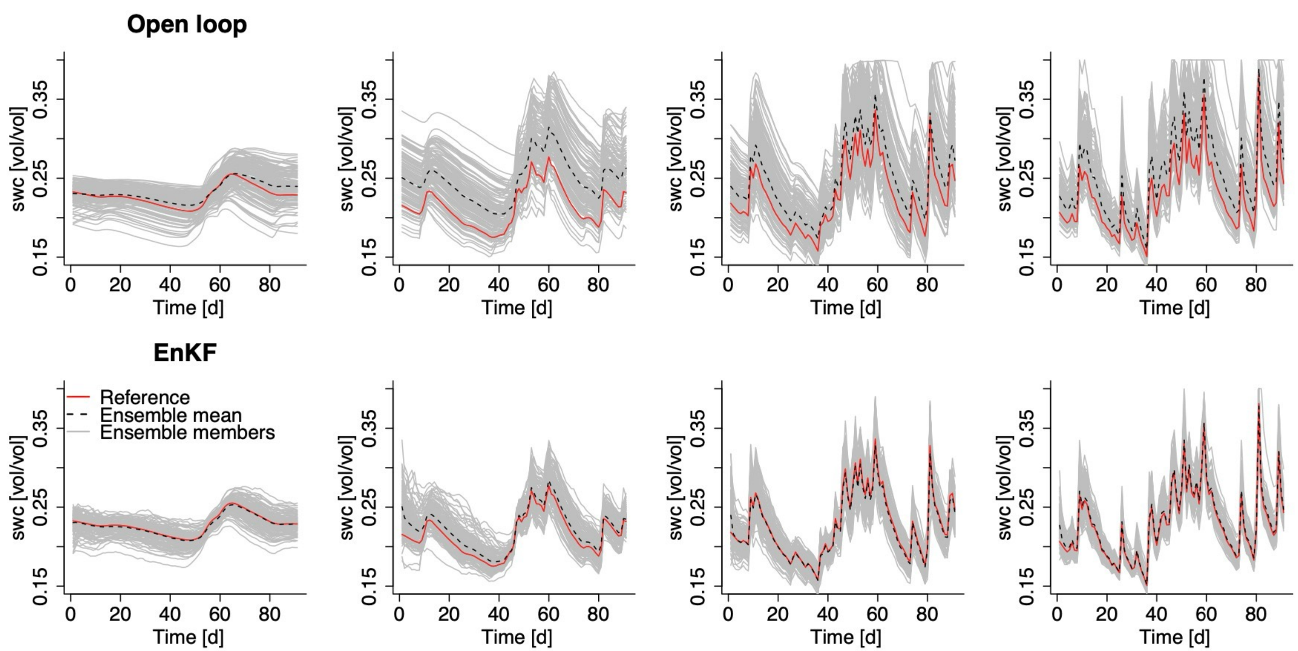
Nerger et al. (2020): Efficient ensemble data assimilation for coupled models with the Parallel Data Assimilation Framework: example of AWI-CM (AWI-CM-PDAF 1.0), Geosci. Model Dev.
Tang et al. (2021): Strongly coupled data assimilation of ocean observations into an ocean-atmosphere model. Geophysical Research Letters

TerrSysMP–PDAF

PDAF Parallel Data Assimilation Framework

a modular high-performance data assimilation framework for an integrated **land surface–subsurface** model

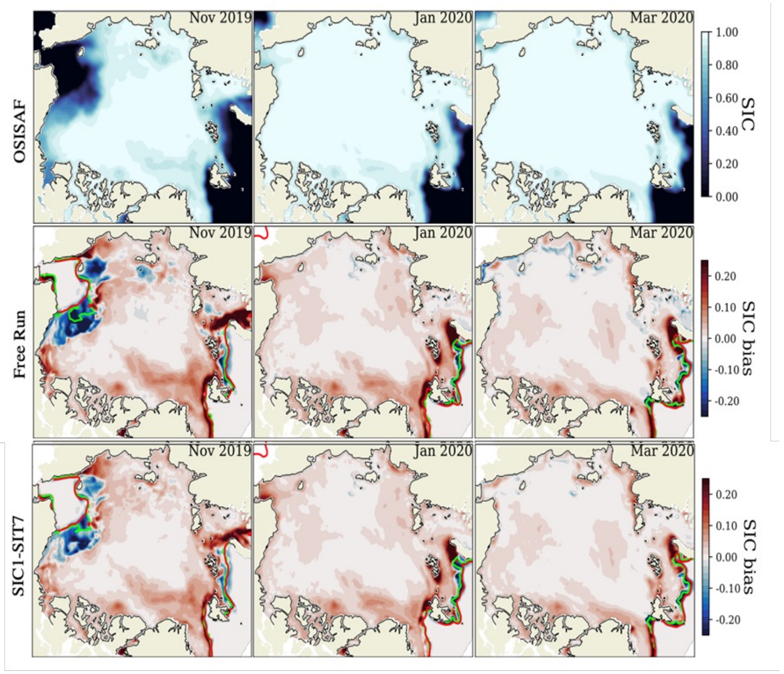




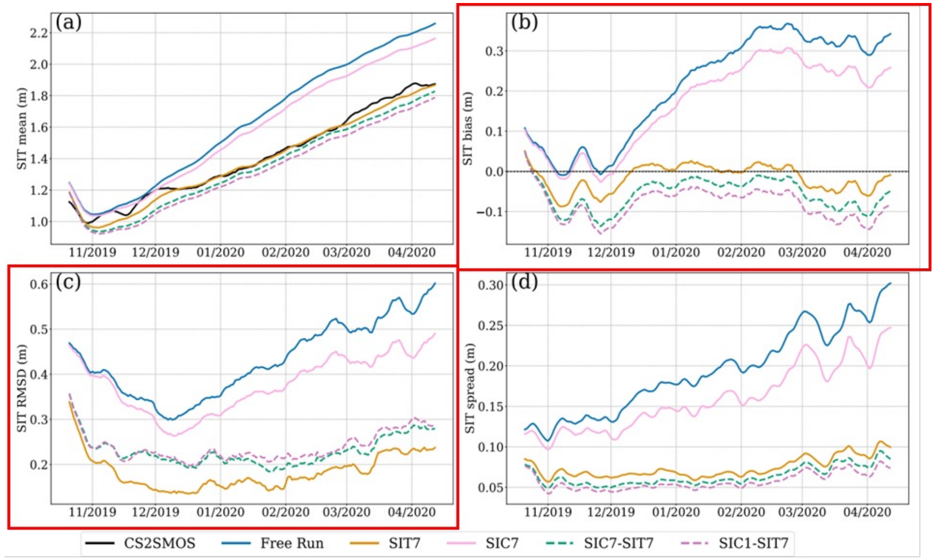
Kurtz et al. (2016): TerrSysMP–PDAF (version 1.0): a modular high-performance data assimilation framework for an integrated land surface–subsurface model, Geosci. Model Dev.

DEnKF on a Lagrangian **sea ice** model

PDAF Parallel Data Assimilation Framework



Reduction in sea ice extent bias (green: obs., red:forecast)



Improved sea ice thickness estimates

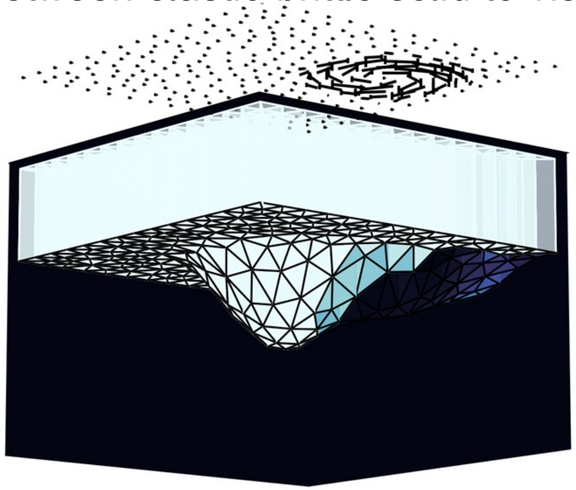
Cheng, S., Chen, Y., Aydoğdu, A., Bertino, L., Carrassi, A., Rampal, P., and Jones, C. K. R. T.: Arctic sea ice data assimilation combining an ensemble Kalman filter with a novel Lagrangiansea ice model for the winter 2019–2020, The Cryosphere, 17, 1735–1754, https://doi.org/10.5194/tc-17-1735-2023, 2023.

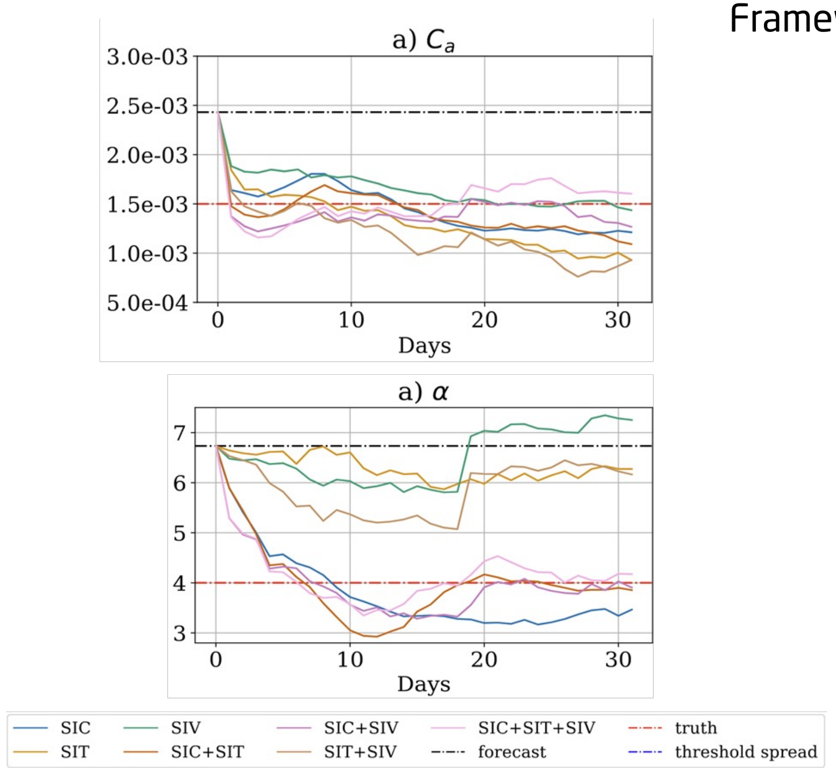
University of Reading SCHMIDT FUTURES National Centre for Earth Observation NATURAL ENVIRONMENT RESEARCH COUNCIL

Joint state and parameter estimation for sea ice model

- Idealised experiment on parameter estimation for a dynamics-only Arctic sea ice model

- Two parameters are estimated:
  - Air drag coefficient () - determines the influence of wind on the sea ice motion
  - Damage parameter (α) – determines the transition between elastic-brittle solid to viscous fluid behaviour





- **Estimation can get close to the truth**
- **Some issues exist**

Chen, Y., Smith, P., Carrassi, A., Pasmans, I., Bertino, L., Bocquet, M., Finn, T. S., Rampal, P., and Dansereau, V.: Multivariate state and parameter estimation with data assimilation on sea-ice models using a Maxwell-Elasto-Brittle rheology, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-1809, 2023.

NCAR TIE-GCM PDAF

PDAF Parallel Data Assimilation Framework

Improving neutral mass density estimation in **upper atmosphere**

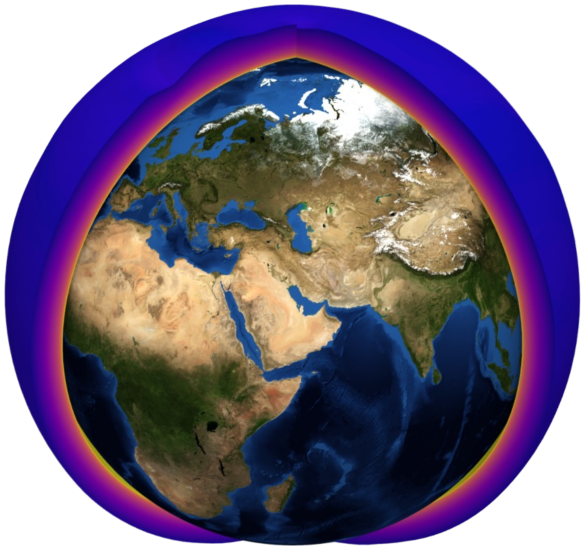
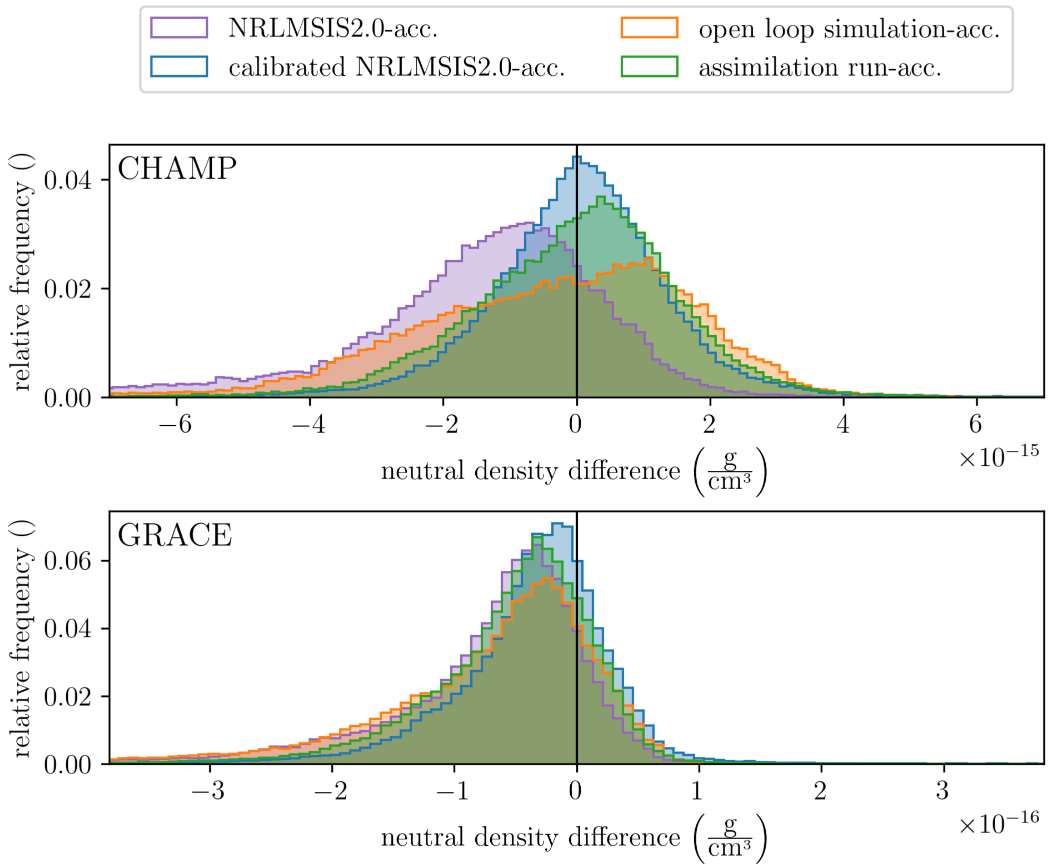


Image of Earth: Reto Stöckli, NASA Earth Observatory



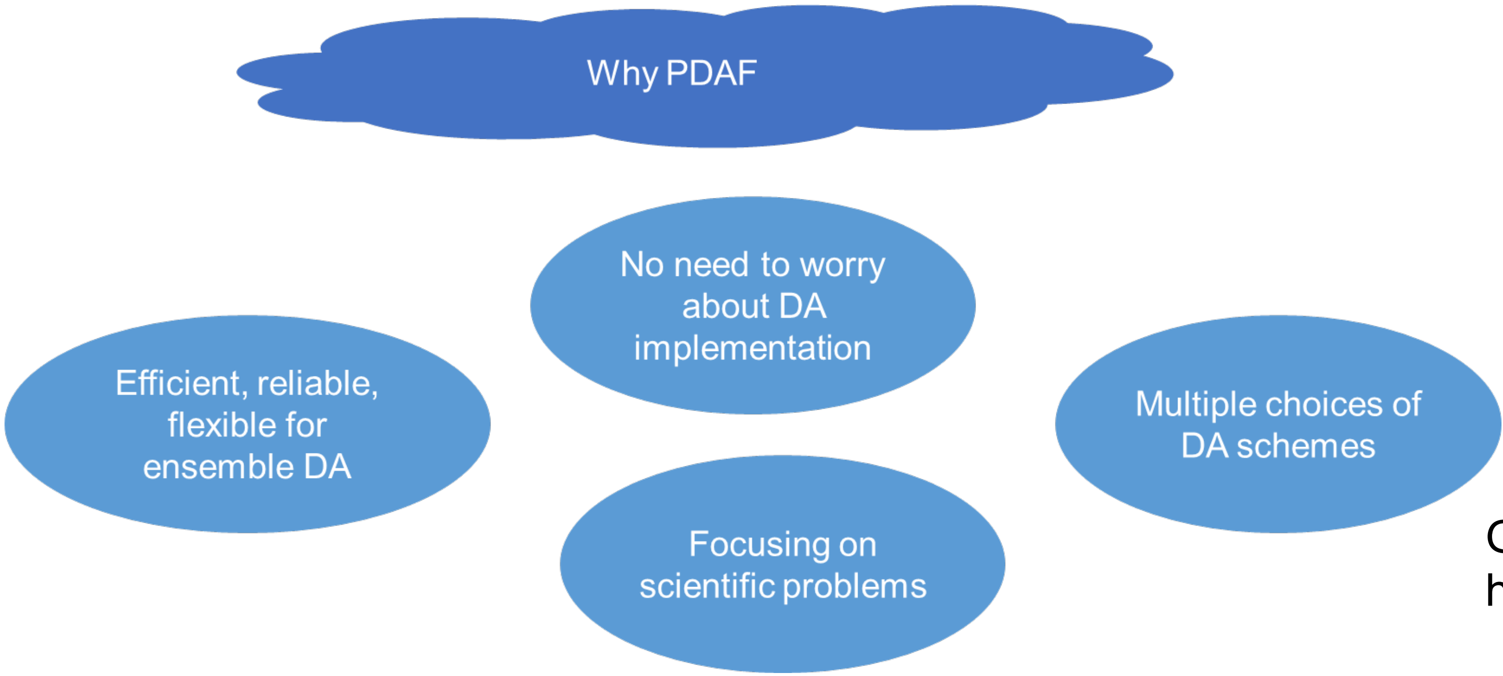
Corbin, A. & Kusche, J. Improving the estimation of thermospheric neutral density via two-step assimilation of in situ neutral density into a numerical model. Earth, Planets and Space 74, 183 (2022).

III Hands On

Building a DA system

PDAF Parallel **Data Assimilation** Framework

- We will use PDAF to build a simple data assimilation system



Website: https://pdaf.awi.de/trac/wiki

Github repo: https://github.com/PDAF/PDAF

pyPDAF – A Python interface to PDAF

PDAF Parallel Data Assimilation Framework

Efficient

Fortran models/
Offline DA

git repository:
https://github.com/yumengch/pypdaf

Python models

PDAF
(Fortran)

pyPDAF
(Python)

Intensive code
development

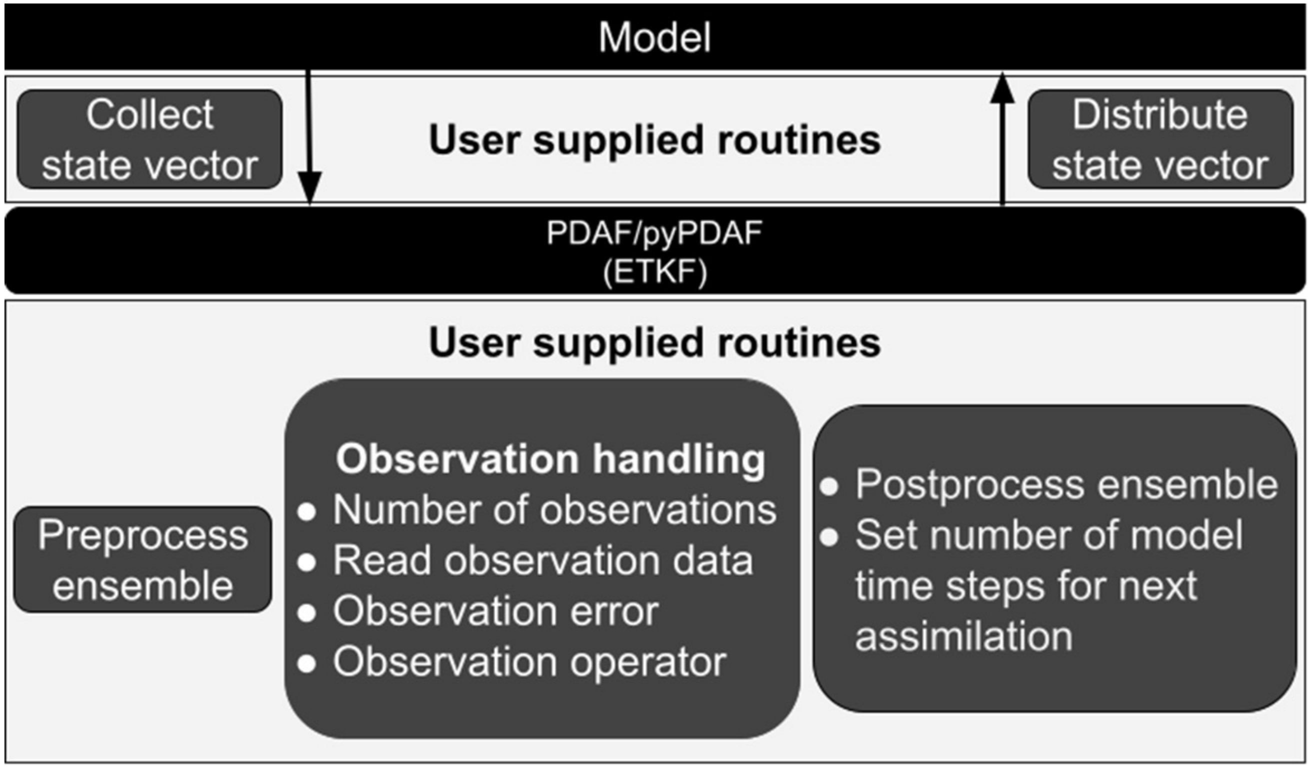
High-dimensional
systems

Less
efficient

Easier code
development

Overview of DA system

PDAF Parallel Data Assimilation Framework



Forecasts and their uncertainties

$$x_k^a = x_k^f + K_k\left(\boxed{y_k^o} - \boxed{H_k} x_k^f\right)$$

$$K_k = P_k^f \boxed{H_k^T}\left(\boxed{H_k} P_k^f \boxed{H_k^T} + \boxed{R_k}\right)^{-1}$$

Observations and their uncertainties

Hands-on example

PDAF Parallel Data Assimilation Framework

https://tinyurl.com/2p938fne

- The jupyter notebook can be run directly in Google colab

- If you download the jupyter notebook on your local computer, you can also install pyPDAF and jupyter notebook with conda using

```
conda install -c yumengch -c conda-forge pypdaf jupyter
```

and running jupyter notebook from the terminal with

```
jupyter notebook
```

References

Asch, M, M. Bocquet, M. Nodet, *Data Assimilation: Methods, Algorithms, and Applications*, SIAM, 2017

Moreaux, G., *Compactly Supported Radial Covariance Functions*, Journal of Geodesy 82.7, 2008

Vetra-Carvalho, S., Van Leeuwen, P.J., Nerger, L., Barth, A., Altaf, M.U., Brasseur, P., Kirchgessner, P. and Beckers, J.-M., *State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems,* Tellus A: Dynamic Meteorology and Oceanography, 70(1), 2018

Evensen, G., F. Vossepoel, P. J. van Leeuwen, *Data Assimilation Fundamentals*, Springer, 2022, doi:10.1007/978-3-030-96709-3 (open access)