Ocean Sciences Meeting, February 21-26, 2016, New Orleans, LA, USA

Tutorial T001

An Introduction to Ensemble Data Assimilation

Lars Nerger

Alfred Wegener Institute for Polar and Marine Research
Bremerhaven, Germany

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR-
UND MEERESFORSCHUNG

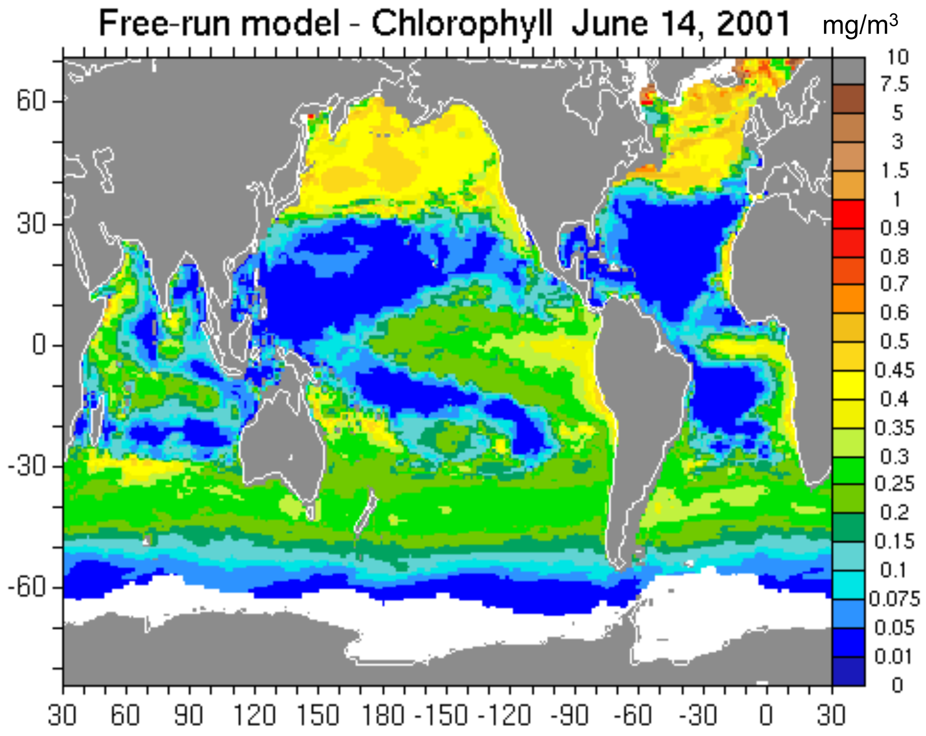
Overview

Introduce to ensemble data assimilation

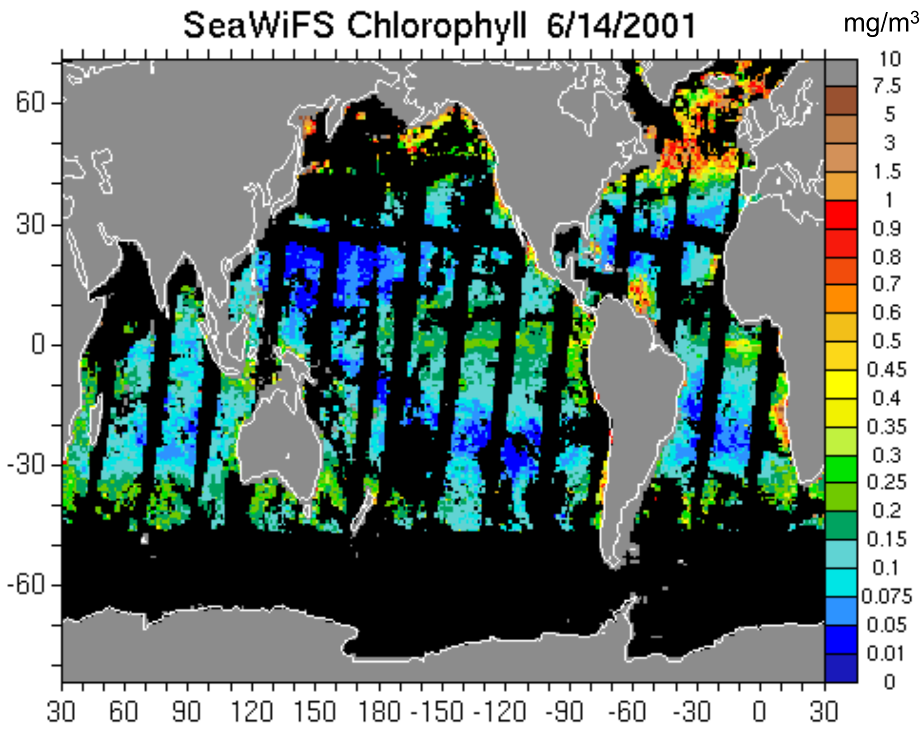
- Application examples
- Data Assimilation: Combine models and observations
- Ensemble Data Assimilation
- Technical aspect: Implement ensemble assimilation

Application examples

Example: Chlorophyll in the ocean



biogeochemical model



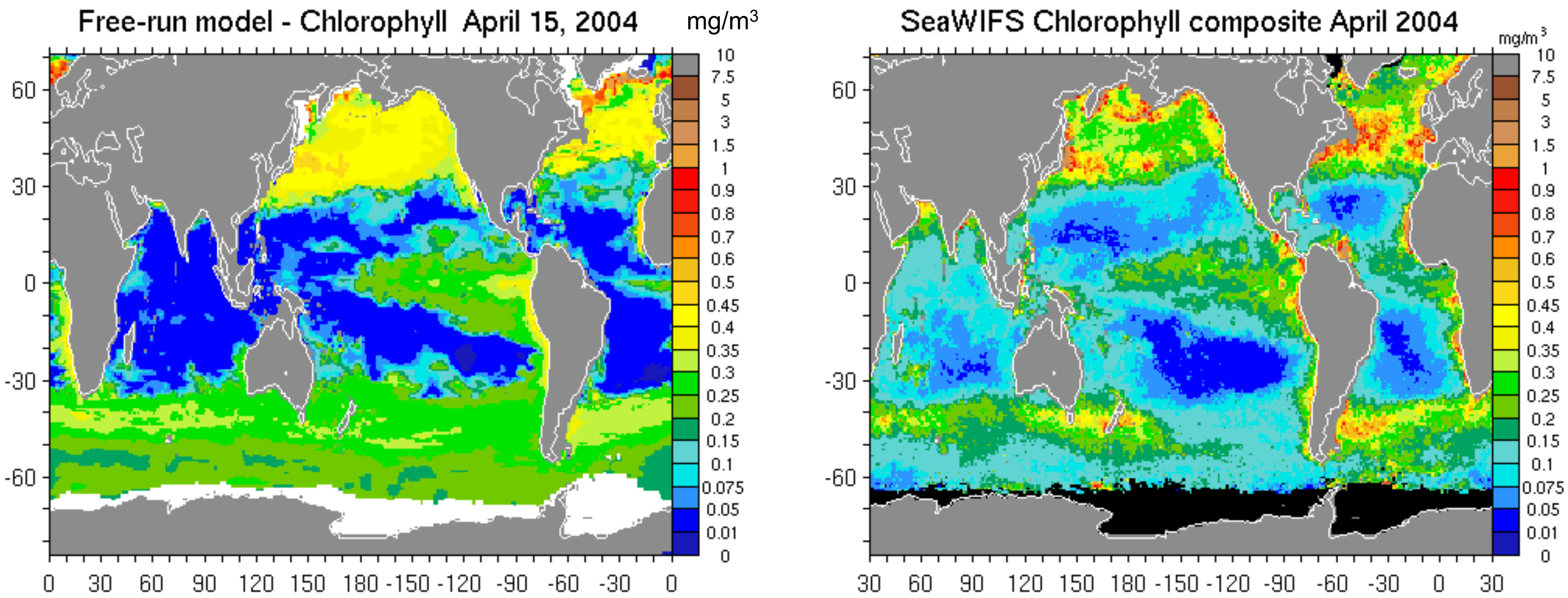
satellite data (1 day)

With data assimilation:

Generate re-analysis fields of chlorophyll concentration with better accuracy than model and filter alone

L. Nerger & W.W. Gregg, *J. Marine Syst.* 68 (2007) 237-254

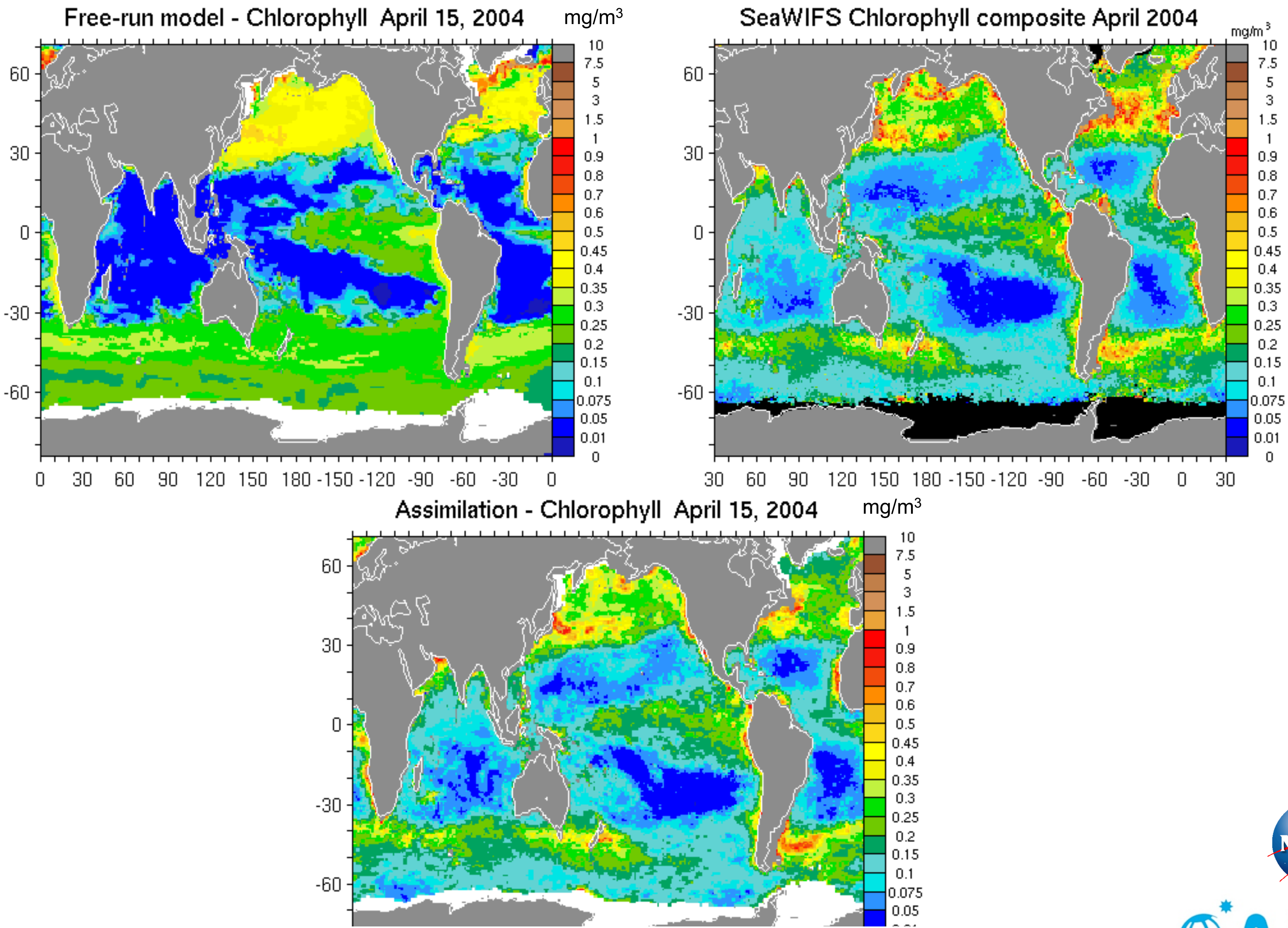
Chlorophyll-a assimilation



From: Nerger & Gregg, *J. Marine Syst.* 68 (2007) 237-254

- Assimilation of satellite chlorophyll-a into global ocean-biogeochemical model
- Generate complete daily chlorophyll fields

Chlorophyll-a assimilation



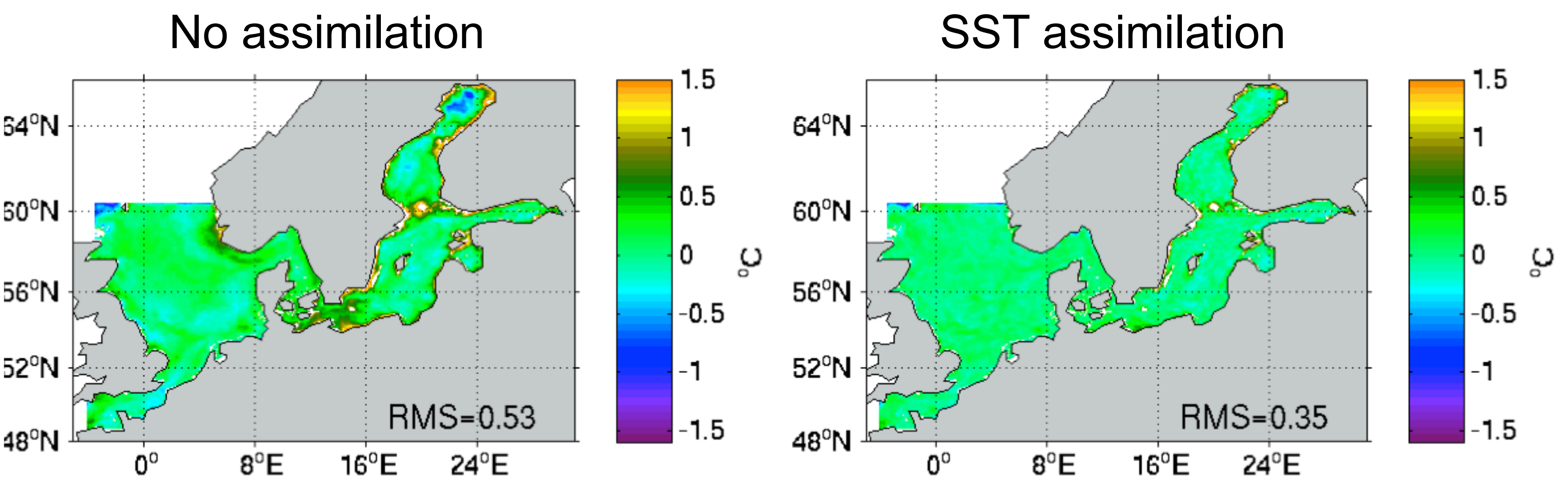
Regional data assimilation application

From: S. Losa et al., J. Mar. Syst. 105–108 (2012) 152–162

Regional application: North Sea & Baltic Sea

Assimilation of satellite sea surface temperature

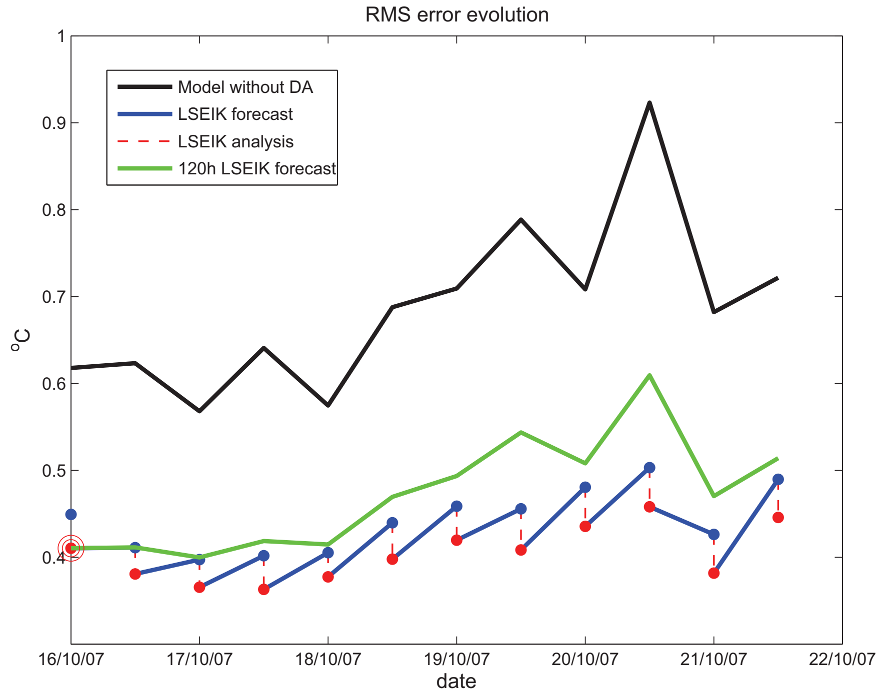
- Mean error for SST in 12-hour forecasts over 1 year



Improving forecasts

From: S. Losa et al., J. Mar. Syst. 105–108 (2012) 152–162

Impact of Assimilation for temperature forecasts

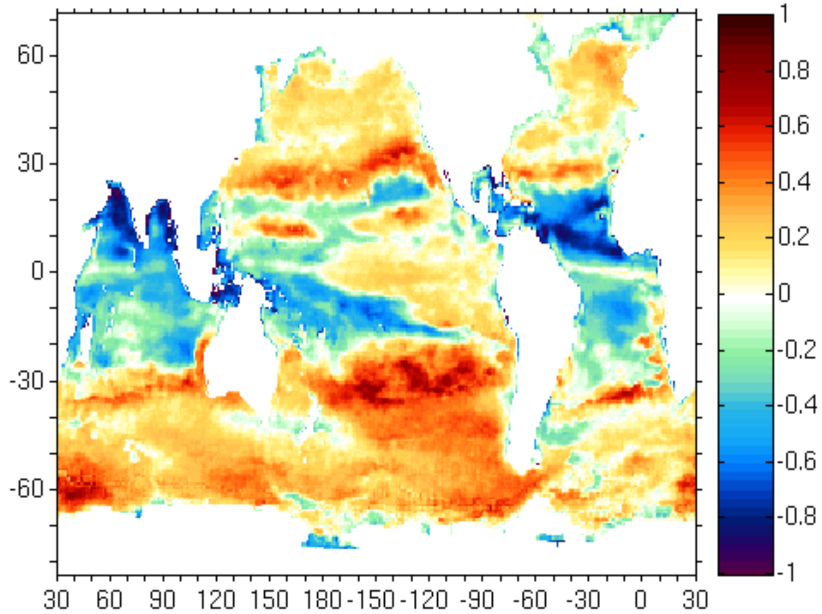


- Very stable 5-days forecasts
- At some point the improvement might break down due to dynamics

Bias Estimation

- *un-biased system*:
  random fluctuation around true state
- *biased system:*
  systematic over- and underestimation
  (common situation with real data)
- *Bias estimation*:
  Separate random from systematic deviations

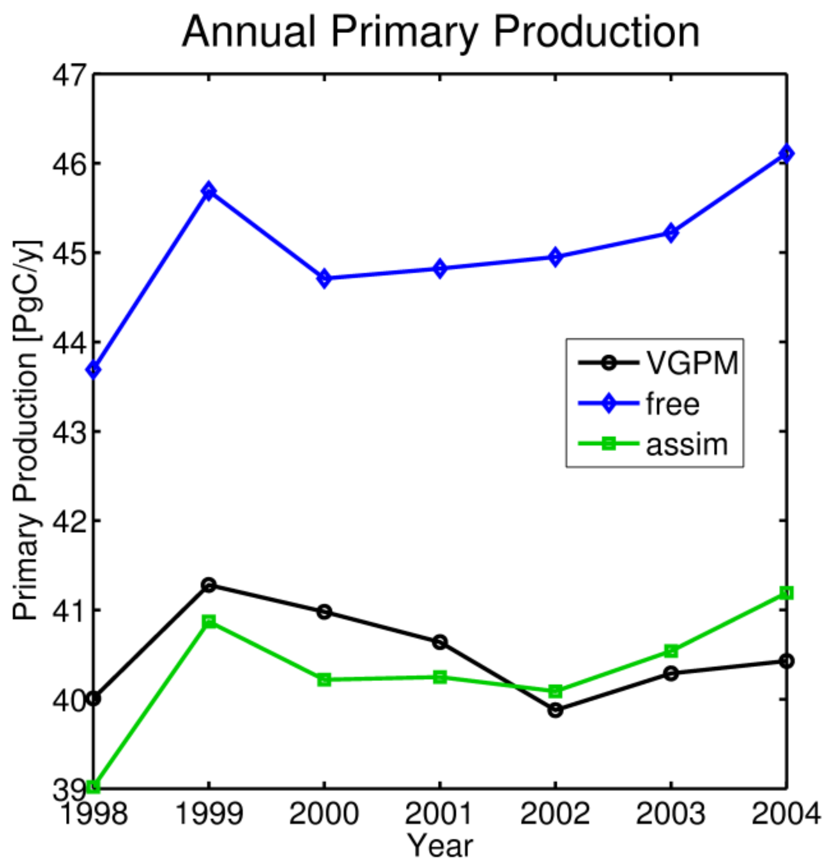
Logarithmic bias estimate
April 15, 2004



Nerger, L., and W.W. Gregg. J. Marine Systems, 73 (2008) 87-102

Estimate Primary Production

- Model: computed as depth-integrated product of growth-rate times Carbon-to-Chlorophyll ratio
- VGPM: Vertical Generalized Production model - satellite data only
- Primary production from assimilation consistent with VGPM-estimate



(VGPM: Behrenfeld, M.J., P.G. Falkowski. Photosynthetic rates derived from satellite-based chlorophyll concentration, Limnol. Oce. 42 (1997) 1-20)

Mean relative difference to VGPM:

Free:	11.2%
Assimilation:	-0.5%

L. Nerger & W.W. Gregg, *J. Marine Syst.* 68 (2007) 237-254

Data Assimilation

Combine Models and Observations

Data Assimilation

Combine model with real data

- Optimal estimation of system state:
  - initial conditions (for weather/ocean forecasts, …)
  - state trajectory (temperature, concentrations, …)
  - parameters (growth of phytoplankton, …)
  - fluxes (heat, primary production, …)
  - boundary conditions and ‘forcing’ (wind stress, …)
- More advanced: Improvement of model formulation
  - Detect systematic errors (bias)
  - Revise parameterizations based on parameter estimates

Needed for Data assimilation

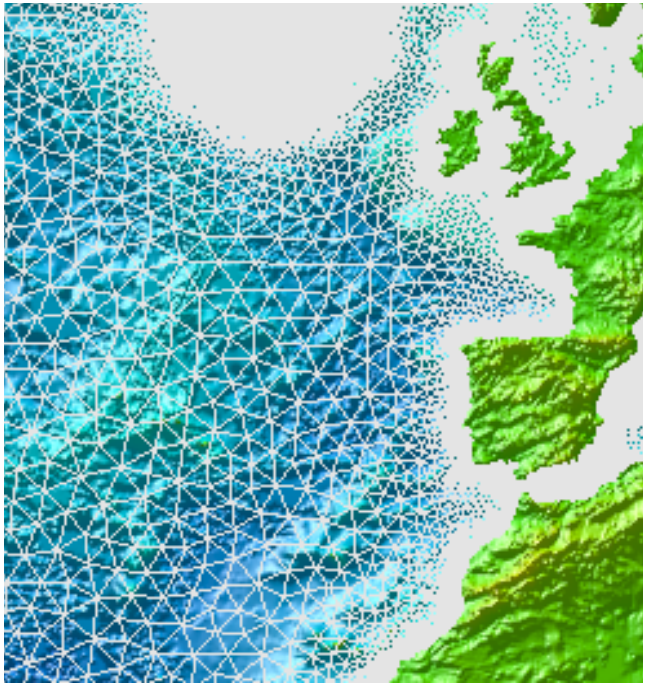
1. Model
   - with some skill
2. Observations
   - with finite errors
   - related to model fields
3. Data assimilation method

Models

Simulate dynamics of ocean

- Numerical formulation of relevant terms
- Discretization with finite resolution in time and space
- “forced” by external sources (atmosphere, river inflows)

- Uncertainties
  - initial model fields
  - external forcing
  - in predictions due to model formulation



Unstructured mesh in North-east Atlantic

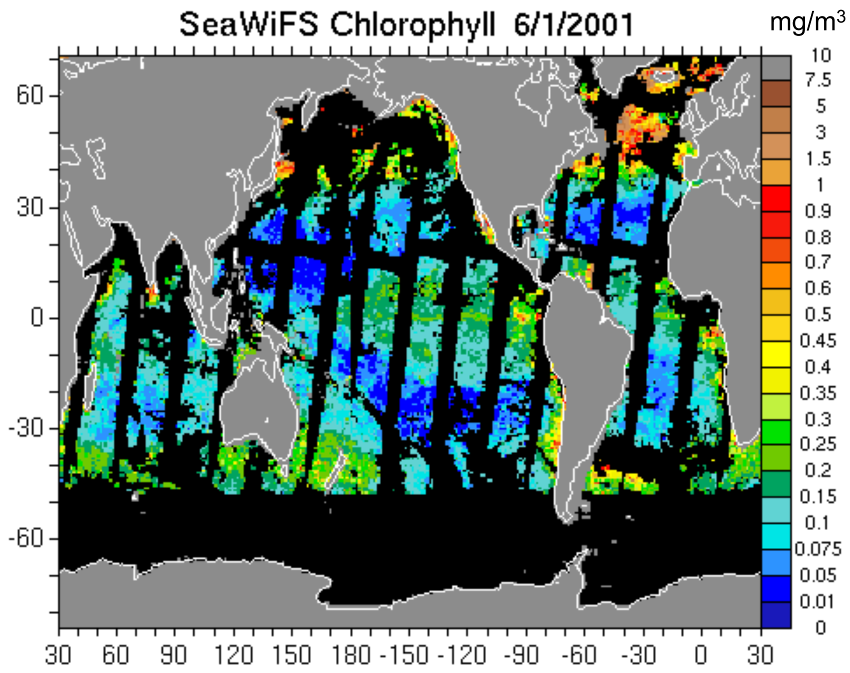
Observations

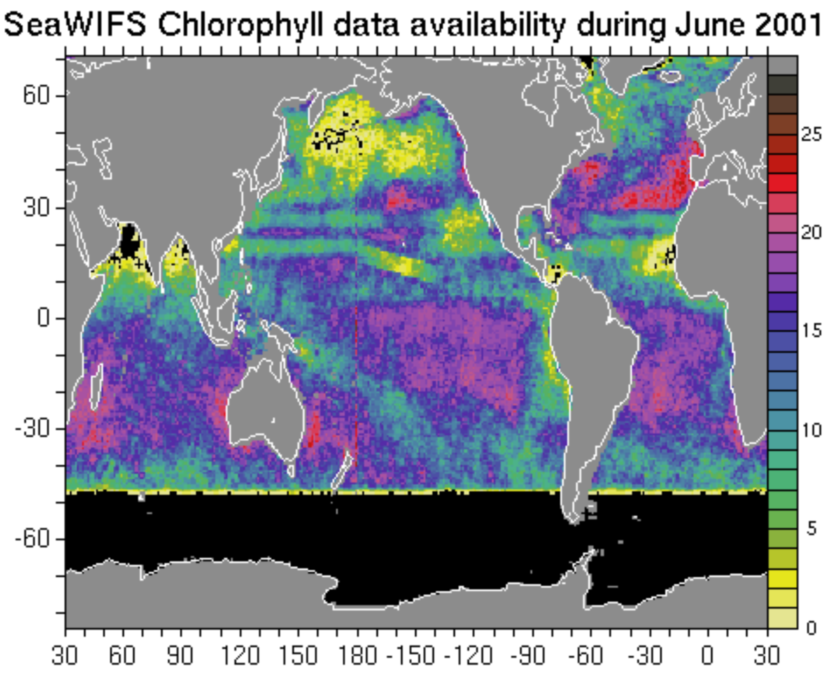
Measure different fields in the Ocean

- Remote sensing
  - E.g. surface temperature, salinity, sea surface height, ocean color, sea ice concentrations & thickness
- In situ
  - Argo, CTD, Gliders, …

- Data is sparse: some fields, data gaps
- Uncertainties
  - Measurement errors
  - Representation errors:
    Model and data do not represent exactly the same
    (e.g. cause by finite model resolution)

Example: Chlorophyll-a (SeaWiFS)





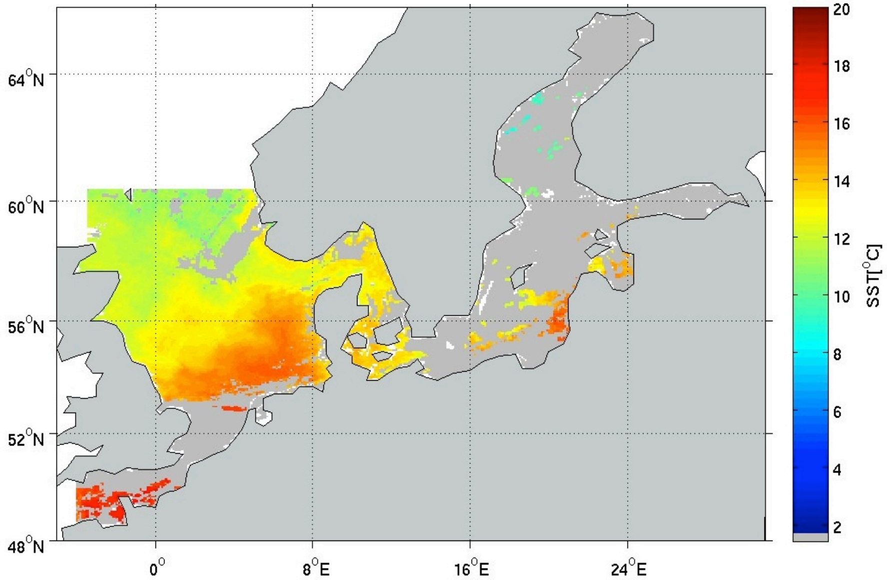
Daily gridded SeaWiFS chlorophyll data

- gaps: satellite track, clouds, polar nights
- On model grid: ~13,000-18,000 data points daily (of 41,000 wet grid points)
- irregular data availability

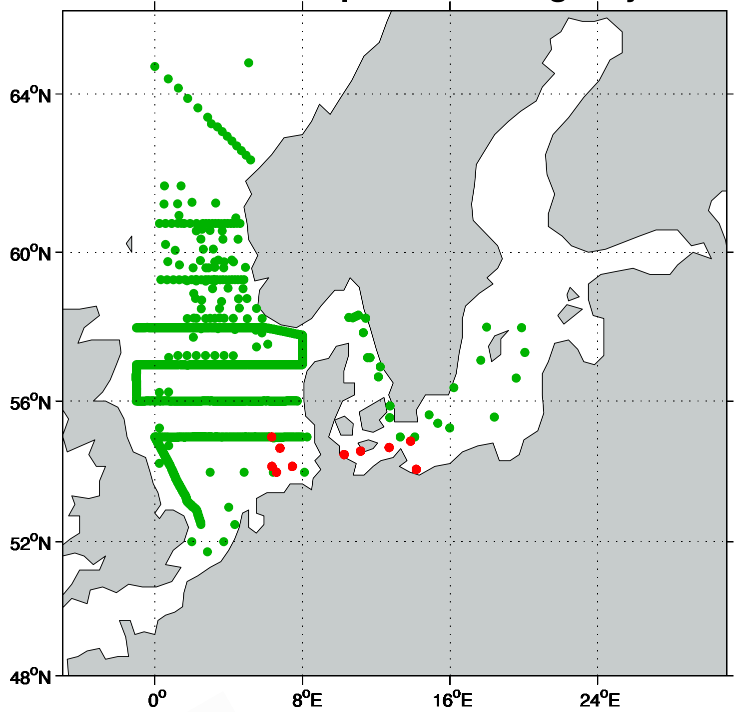
Nerger, L., and W.W. Gregg. J. Marine Systems **68** (2007) 237

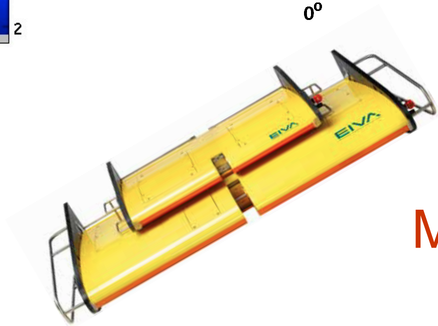
Example: Physical Data in North & Baltic Seas

Satellite surface temperature
(12-hour composite)



Avalable T and S profiles during July 2008





Scanfish and
CTD profiles

MARNET
stations

Ensemble Data Assimilation

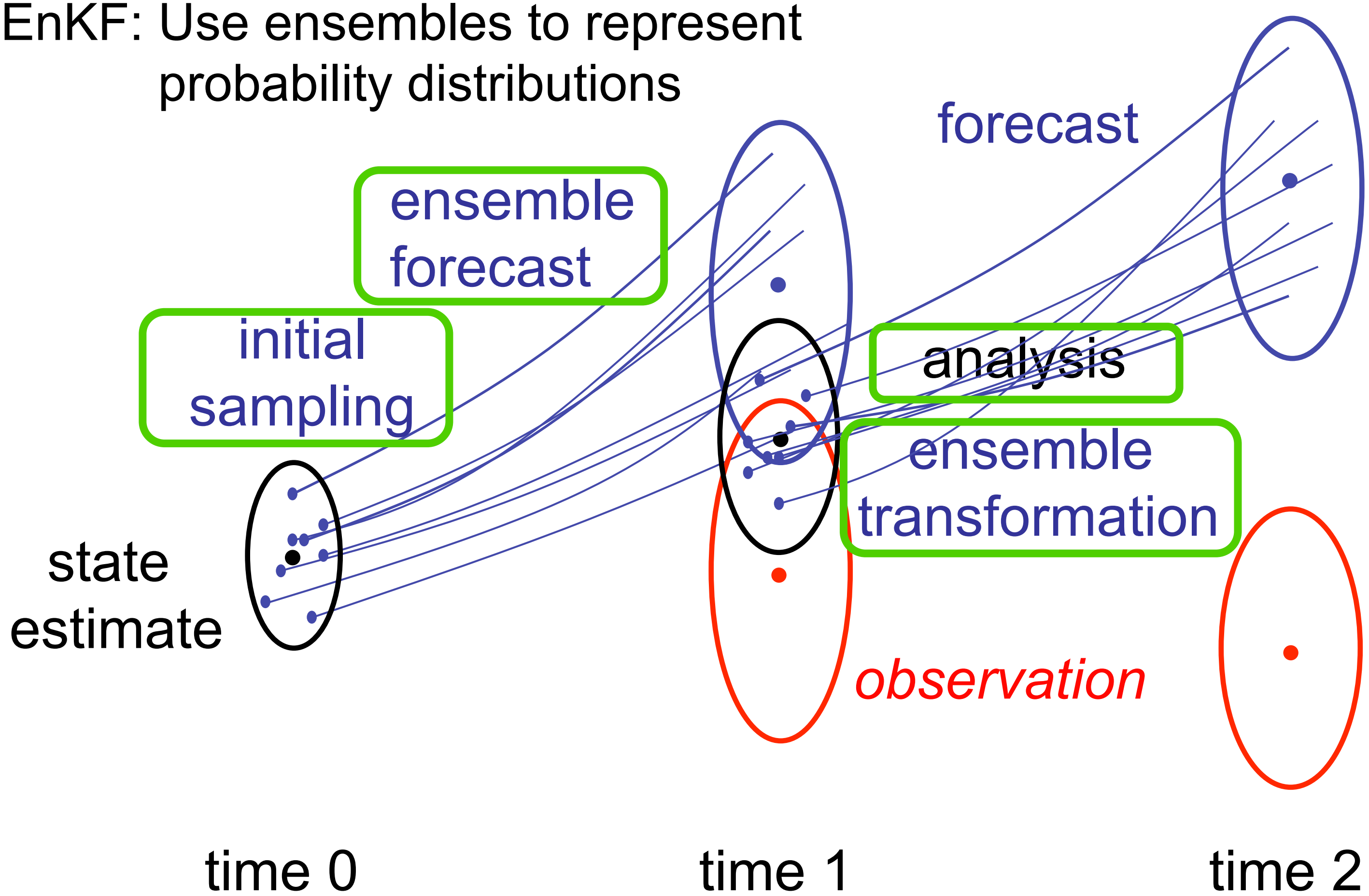
Estimate uncertainty

Ensemble-based Kalman Filter

First formulated by G. Evensen (EnKF, J. Geophys. Res. 1994)

Kalman filter: express probability distributions by mean and covariance matrix

EnKF: Use ensembles to represent probability distributions

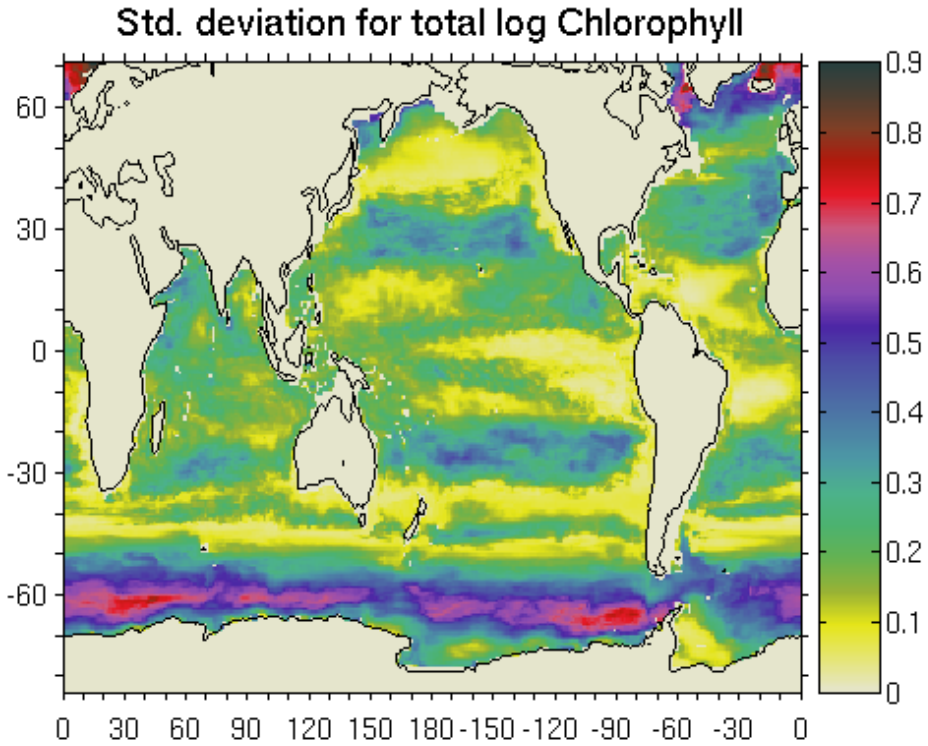


There are many possible choices!

What is optimal is part of our research

Ensemble Covariance Matrix

- Provide uncertainty information (variances + covariances)
- Generated dynamically
  by propagating ensemble of model states



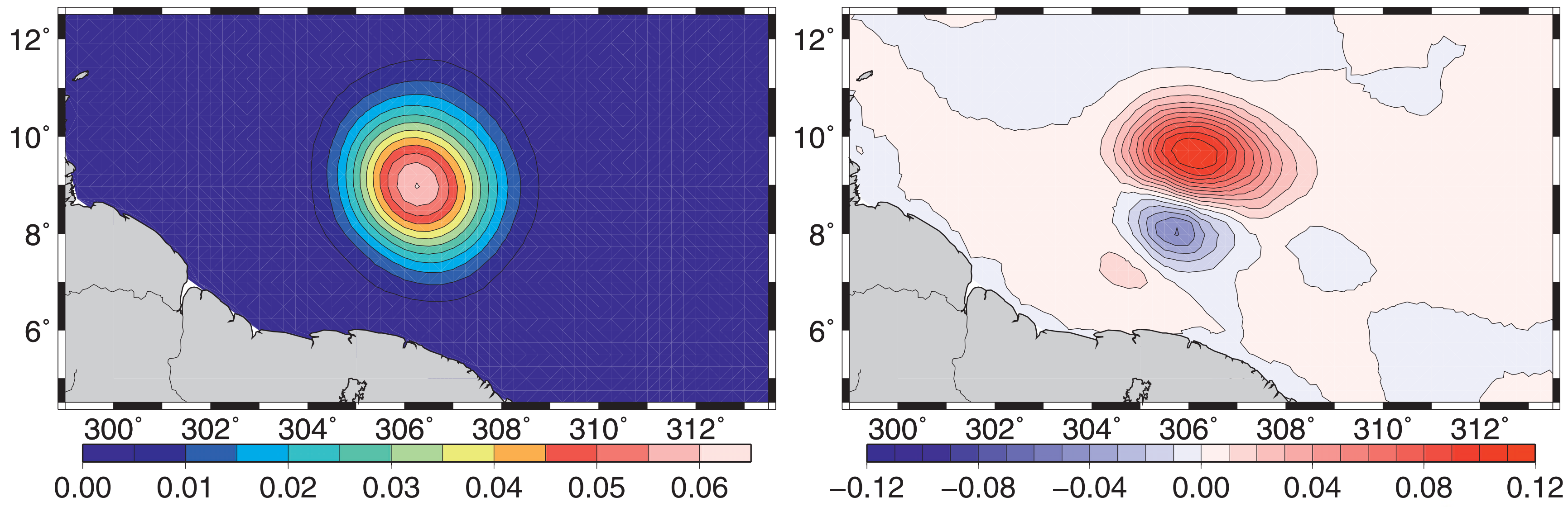
Ensemble Covariance Matrix (II)

- Also:
  Provide information on error correlations
  (between different locations and different fields)

- Example: Assimilation of sea surface height
  (Brankart et al., Mon. Wea. Rev. 137 (2009) 1908-1927)

Assimilation increment in sea surface height

Induced change in zonal velocity



Technical Aspects

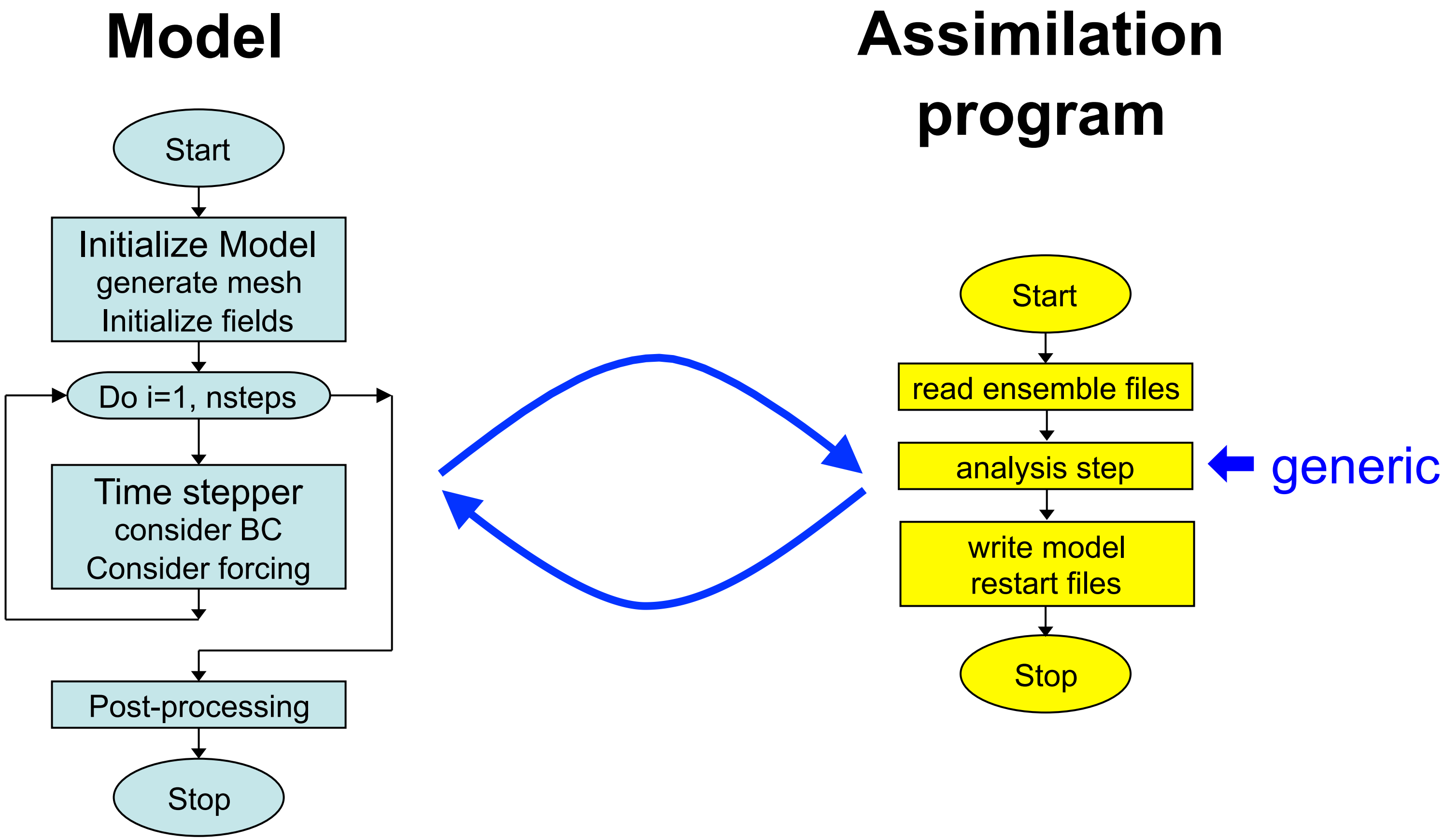
Implement Ensemble Data Assimilation

Computational and Practical Issues

- Running a whole model ensemble is costly
- Ensemble propagation is naturally parallel (all independent)
- Ensemble data assimilation methods need tuning
- No need to go into model numerics (just model forecasts)
- Filter step of assimilation only need to know:
  - Values of model fields an their location
  - Observed values, their location and uncertainty

Ensemble data assimilation can be implemented
in form of a generic code
+ case-specific routines

Offline coupling – separate programs

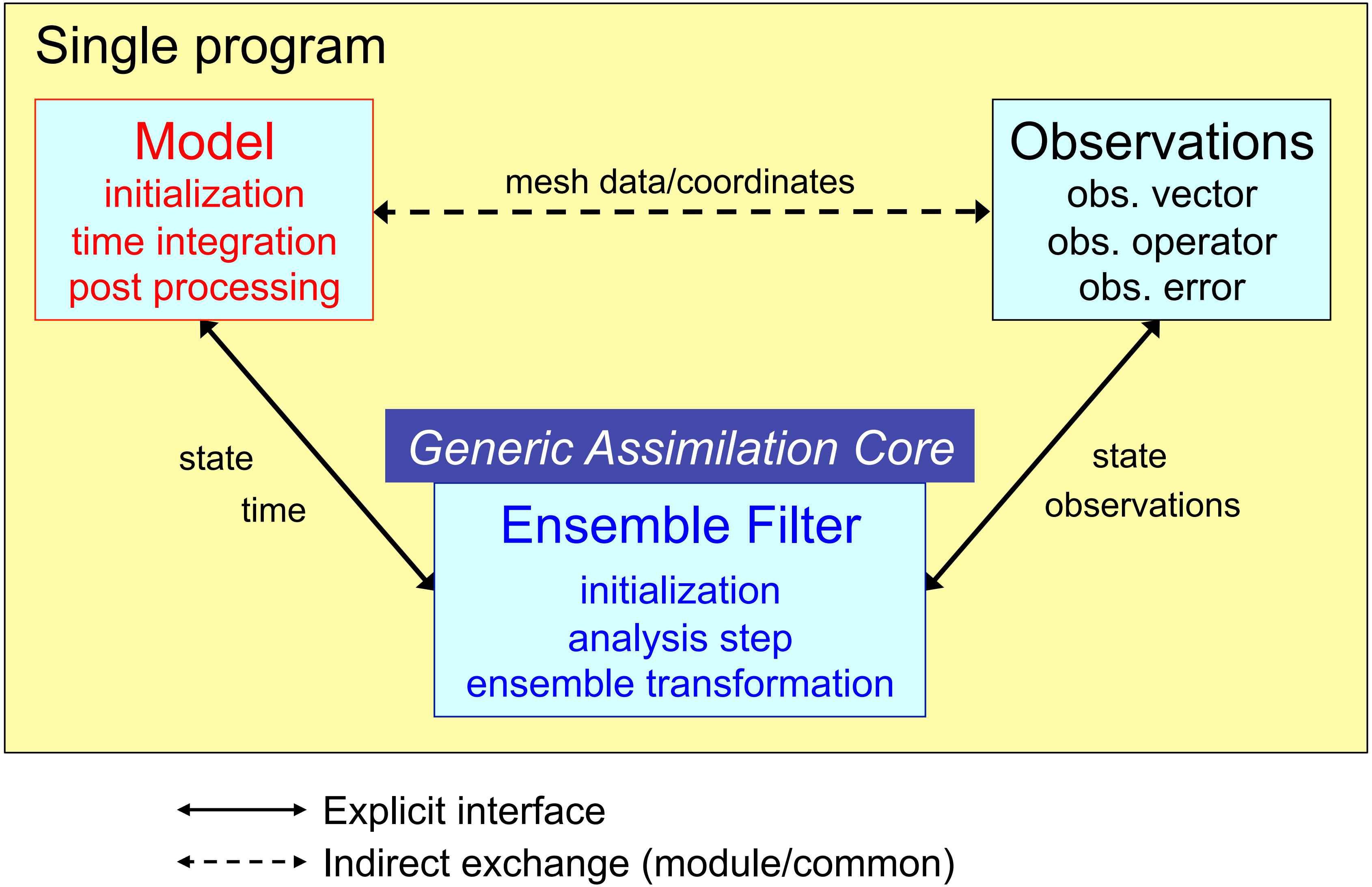


For each ensemble state

- Initialize from restart files
- Integrate
- Write restart files

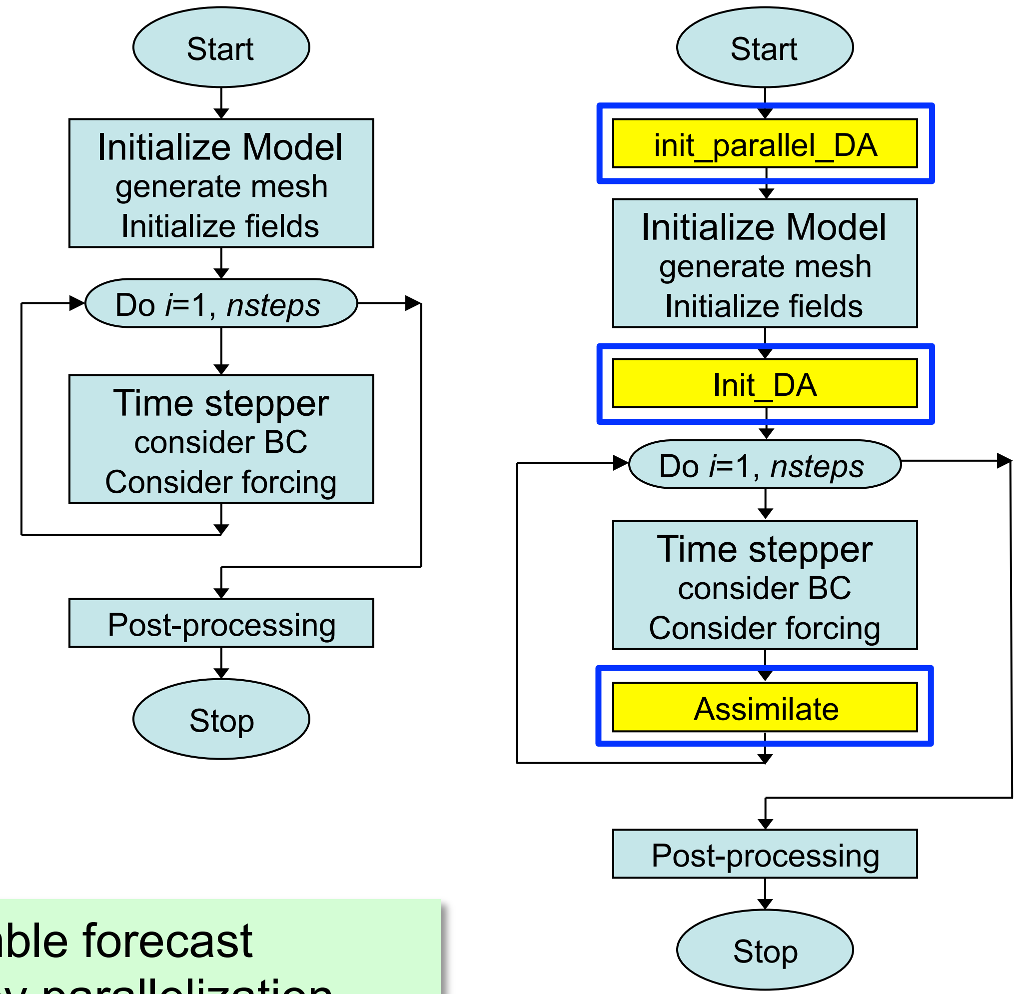
- Read restart files (ensemble)
- Compute analysis step
- Write new restart files

Online Coupling



Extending a Model for Data Assimilation

Model



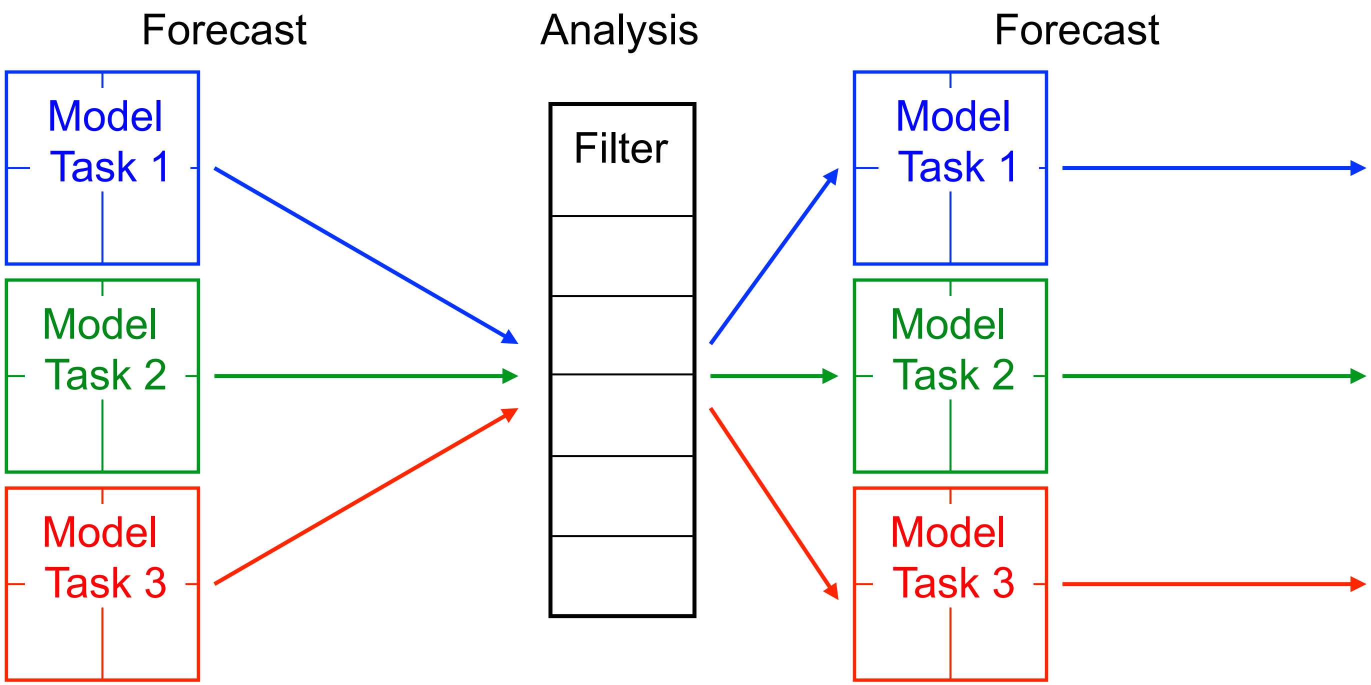
Extension for data assimilation

plus:
Possible model-specific adaption.

E.g. NEMO: Euler time step after assimilation

ensemble forecast enabled by parallelization

2-level Parallelism



1. Multiple concurrent model tasks
2. Each model task can be parallelized

➤ Analysis step is also parallelized

PDAF: A tool for data assimilation

PDAF Parallel Data Assimilation Framework

PDAF - Parallel Data Assimilation Framework

- a program library for data assimilation
- provide support for ensemble forecasts
- provide fully-implemented filter and smoother algorithms
- easily useable with (probably) any numerical model (applied with NEMO, MITgcm, FESOM, MPIOM, HBM)
- makes good use of supercomputers
- first public release in 2004; continued development

Open source:
Code and documentation available at

http://pdaf.awi.de

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

AWI

Summary

Ensemble Data Assimilation

- Use an ensemble of model runs to estimate the model state and its uncertainty
- Combine model ensemble with observations to improve state estimates
- Efficient on today's parallel computers
- Generic program frameworks are available
- Allows to estimate model fields, fluxes, parameters, …

Thank you!

Lars.Nerger@awi.de

http://pdaf.awi.de