Ensemble Data Assimilation in a global coupled sea ice model

Alexander Barth, Mohamed Ouberdous, François Laenen, Yajing Yan, Jean-Marie Beckers

GHER, University of Liege, Belgium. Contact: A.Barth@ulg.ac.be

SANGOMA Annual meeting, November 27, 2012

Context and objectives

- Objective: Understanding and predicting Antarctic sea ice variability at the decadal timescale
- Many gaps in our knowledge of the processes that rule the variability of the sea ice extent in the Southern Ocean are still remaining
- Such as the recent positive trend in sea ice extent in a global warming context
- Our contribution: development of a data assimilative global model system coupled to a sea ice model

Model and state vector

- NEMO-LIM with 2° resolution (global) and 31 z-levels
- Based on NEMO and LIM restart files
- Hydrodynamical variables:
  - u-velocity
  - v-velocity
  - temperature
  - salinity
  - surface elevation
  - rotational of horizontal velocity components
  - divergence of horizontal velocity components
  - turbulent kinetic energy
- Leap-frog time stepping (two time instances: *b and *n) and time averaged surface values (ss*m)

- Sea ice variables:
  - sea ice fraction (transformed variable with Gaussian anamorphosis)
  - Ice thickness
  - Snow thickness
  - Temperature inside the ice/snow layer
  - u-ice velocity
  - v-ice velocity
  - Energy stored in the brine pockets
- in total 32 different variables and 6 million elements.
- to be determined: if the assimilation increment of all variables has a positive impact
- Sea ice Surface Temperature (sist) was removed from state vector

Assimilated Observations

- Global sea surface temperature (OSTIA, reduced to 2° resolution)
  - Error standard deviation is the **average** of the error standard deviation of the original OSTIA SST
  - $\mathrm{var}(\varepsilon_1 + \varepsilon_2) = \mathrm{var}(\varepsilon_1) + \mathrm{var}(\varepsilon_2) + 2\,\mathrm{cov}(\varepsilon_1, \varepsilon_2)$
  - if ε_1 and ε_2 are independent: $\mathrm{var}(\varepsilon_1 + \varepsilon_2) = \mathrm{var}(\varepsilon_1) + \mathrm{var}(\varepsilon_2)$
  - if ε_1 and ε_2 are perfectly correlated: $\mathrm{std}(\varepsilon_1 + \varepsilon_2) = \mathrm{std}(\varepsilon_1) + \mathrm{std}(\varepsilon_2)$
- Global sea ice fraction (OSTIA/OSI-SAF, reduced to 2° resolution), error standard deviation for assimilation is assumed to be 0.1
- Satellite-based sea ice drift (for southern hemisphere only), error standard deviation for assimilation is assumed to be 0.1 m/s
- Error standard deviation needs to be fine-tuned

Data Assimilation algorithm

- The "best" estimator of the model state vector $\mathbf{x}^a$:

$$\begin{aligned} \mathbf{x}^a &= \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \\ \mathbf{K} &= \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \\ \mathbf{P}^a &= \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f \end{aligned}$$

- Decompositions of $\mathbf{P}^f$ in square root matrices $\mathbf{S}^f$ $(n \times r)$:

$$\mathbf{P}^f = \mathbf{S}^f \mathbf{S}^{fT}$$

- Only effective if r is small ($r << n$).
- We assume that $\mathbf{R}$ is diagonal.

n	number of state variables
r	number of ensemble members
$\mathbf{x}^{f/a}$	the model forecast/analysis
$\mathbf{P}^{f/a}$	error covariance of $\mathbf{x}^{f/a}$
$\mathbf{S}^{f/a}$	square root decomposition of $\mathbf{P}^{f/a}$
$\mathbf{y}^o$	observations
$\mathbf{R}$	error covariance of $\mathbf{y}^o$
$\mathbf{H}$	observation operator
$\mathbf{U}$	eigenvectors
$\mathbf{\Lambda}$	eigenvalues

Data Assimilation algorithm

In practice, the following eigenvalue decomposition is made:

$$\left(\mathbf{H}\mathbf{S}^f\right)^T \mathbf{R}^{-1}\mathbf{H}\mathbf{S}^f = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T \tag{1}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and where Λ is diagonal. $\mathbf{U}$ and Λ are both of size $r \times r$. The Kalman gain $\mathbf{K}$ and $\mathbf{S}^a$ can be computed by:

$$\begin{aligned}
\mathbf{K} &= \mathbf{S}^f\mathbf{U}(1+\boldsymbol{\Lambda})^{-1}\mathbf{U}^T(\mathbf{H}\mathbf{S}^f)^T\mathbf{R}^{-1} &(2)\\
\mathbf{S}^a &= \mathbf{S}^f\mathbf{U}(1+\boldsymbol{\Lambda})^{-1/2}\mathbf{U}^T &(3)
\end{aligned}$$

$\mathbf{S}^a$ is the square root of $\mathbf{P}^a$:

$$\mathbf{P}^a = \mathbf{S}^a\mathbf{S}^{aT} \tag{4}$$

Based on $\mathbf{x}^a$ and $\mathbf{S}^a$, an ensemble can be reconstructed:

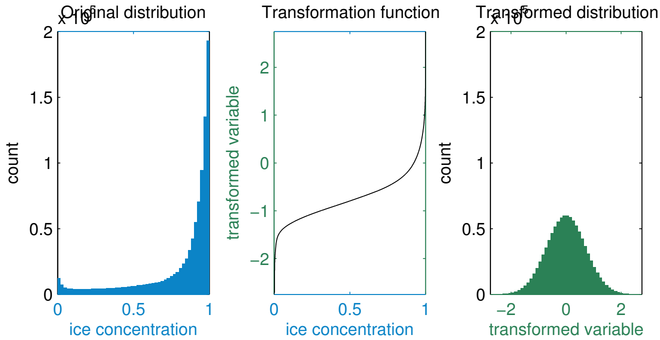
$$\mathbf{x}^{a(k)} = \mathbf{x}^a + \sqrt{r-1}\,\mathbf{S}^{a(k)} \tag{5}$$

Ocean Assimilation Kit (OAK)

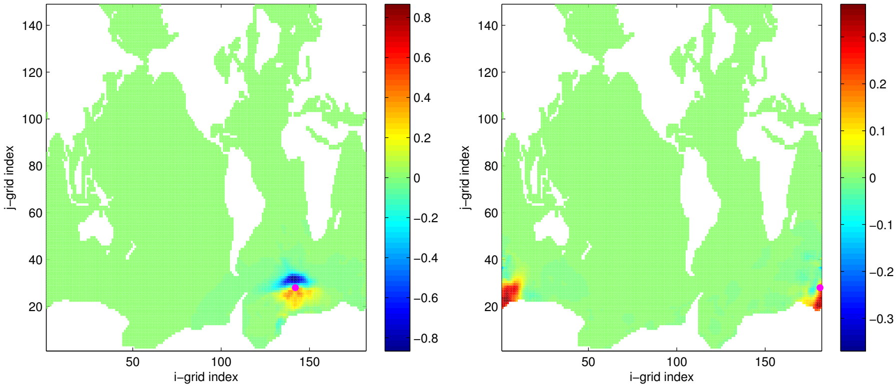
- Reduced rank square root analysis
- Global and local algorithm
- Modular Fortran 90 program
- Flexible definition of state vector
- Supports arbitrary curvilinear grid
- Local algorithm parallelized with OpenMP and MPI
- NetCDF or Fortran binary files as input
- Released as open-source (more info tinyurl.com/assim-ocean or modb.oce.ulg.ac.be/OAK)

Gaussian anamorphosis

- the analysis is the most likely state if errors are Gaussian-distributed
- however some variables are clearly not Gaussian-distributed: e.g. sea ice concentration (between 0 and 1)
- apply non-linear transformation
  - an analytical transformation (e.g. log, for lognormal distributions)
  - empirical transformation (based on cumulative distribution function, cdf)



Covariance localization



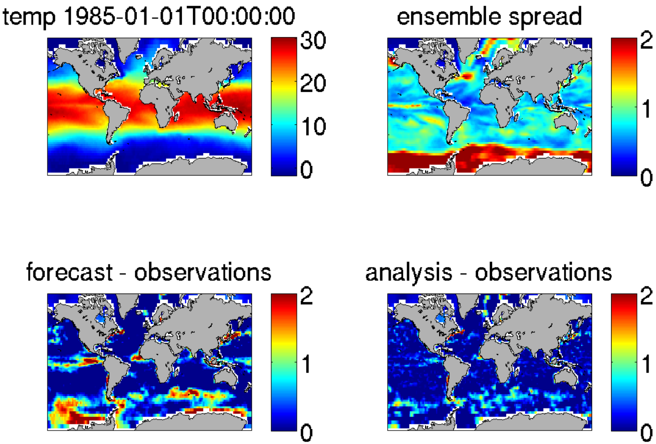
- Assimilation increment for temperature for a point observation (magenta dot). Maximum length-scale is 2000 km (about 20 grid points)
- Model domain extends from -280 E to 80 E.
- Localization needs to take the cyclic boundary condition into account

Model run

- Ensemble spin-up:
  - Start time: 1984-01-01
  - Followed by a one-year ensemble spin-up
  - All members start with the same initial condition
  - Every member is run with perturbed atmospheric forcings (wind and air temperature)
- Assimilative run:
  - First observations assimilated: 1985-01-01
  - Observations are assimilated every 5 days (if available)
  - 50 ensemble members
  - Again perturbed wind and air temperature
  - 1 year $\rightarrow$ 28 hours CPU time on 50 Xeon E5649 CPUs (lemaitre2)

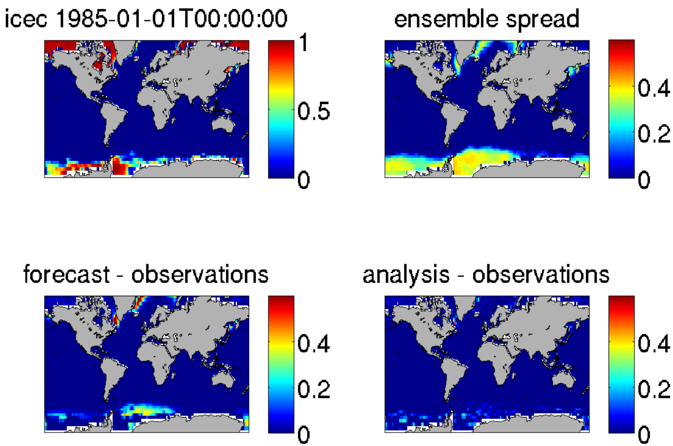
First assimilation cycle

Sea surface temperature



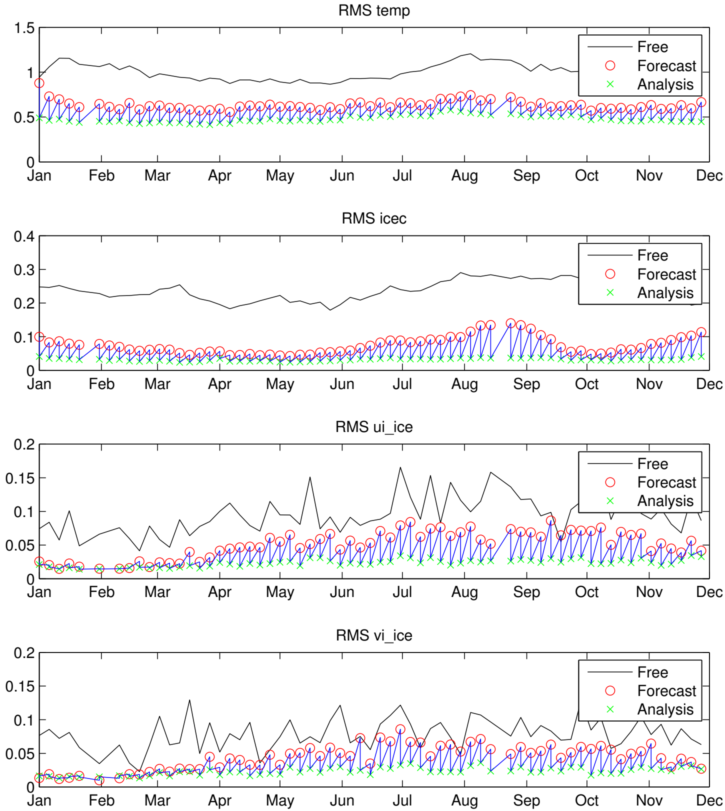
- Relatively large ensemble spread at first assimilation cycle.
- Qualitative agreement between ensemble spread and difference between forecast and observations
- Larger errors in Gulf Stream and Kuroshio region, but no ensemble spread $\rightarrow$ only very small correction.

Ice concentration



- Areas with large ensemble spread (near ice edge) agree well with region with high error
- Uncertainty of sea ice concentration seems to be easier to predict than uncertainty in sea surface temperature

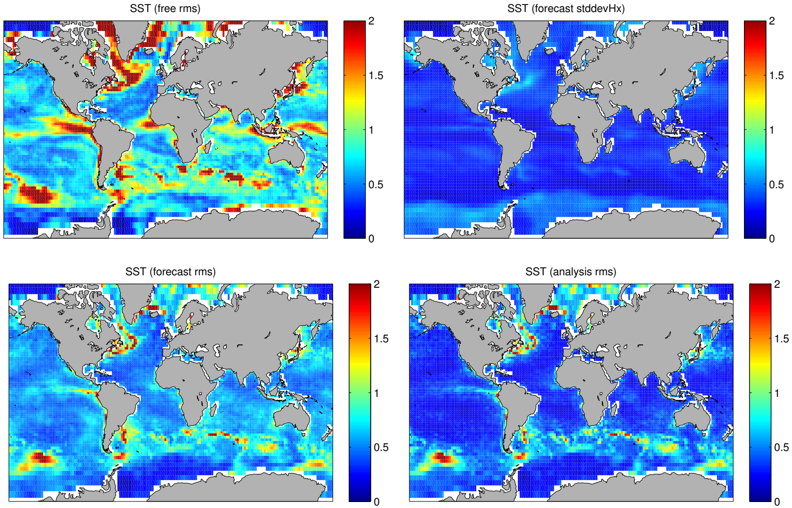
RMS temporal evolution



- RMS difference between model run and observations stabilizes after a few assimilation cycles.
- “Forecast” and “Analysis” refer to the corresponding ensemble mean.
- SST, Ice concentration RMS is averaged over the entire globe.
- Ice drift is averaged over the area where observations are available (i.e. ice-covered areas).

Error statistics averaged over time

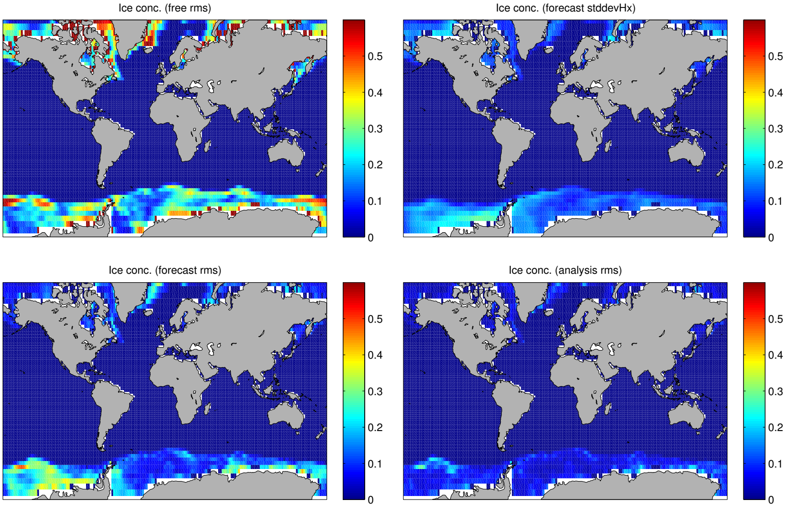
Sea surface temperature



- RMS = averaged over the year
- stddevHx = square root of the averaged ensemble variance of the model forecast
- Free run: large error near boundary currents and equatorial region

- Qualitative agreement of ensemble spread
- Of course, RMS difference between model and observations is reduced during the analysis

Ice concentration

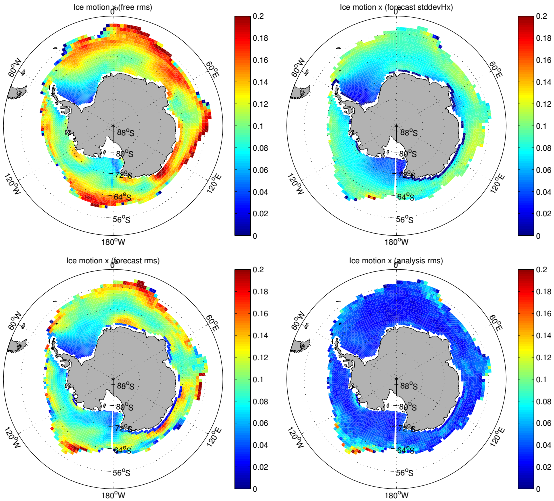


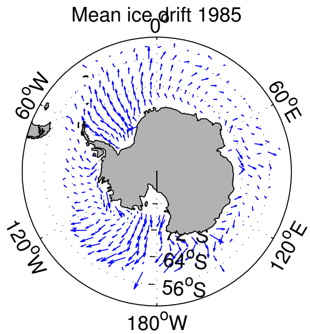
- Relatively good agreement between ensemble spread of forecast and actual RMS difference between forecast and observations

- RMS difference includes also the observation error:

$$E\left[(\mathbf{H}\mathbf{x}^f - \mathbf{y}^o)(\mathbf{H}\mathbf{x}^f - \mathbf{y}^o)^T\right] = \mathbf{H}\mathbf{P}^f\mathbf{H} + \mathbf{R}$$

Zonal ice-drift

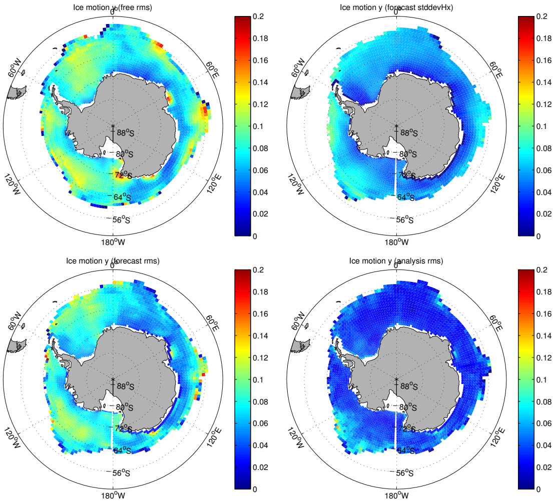


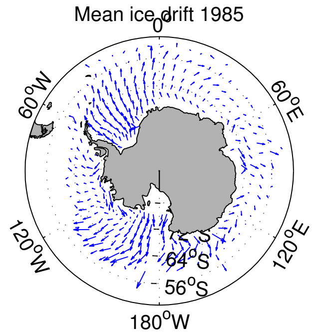


Mean ice drift from the model analysis

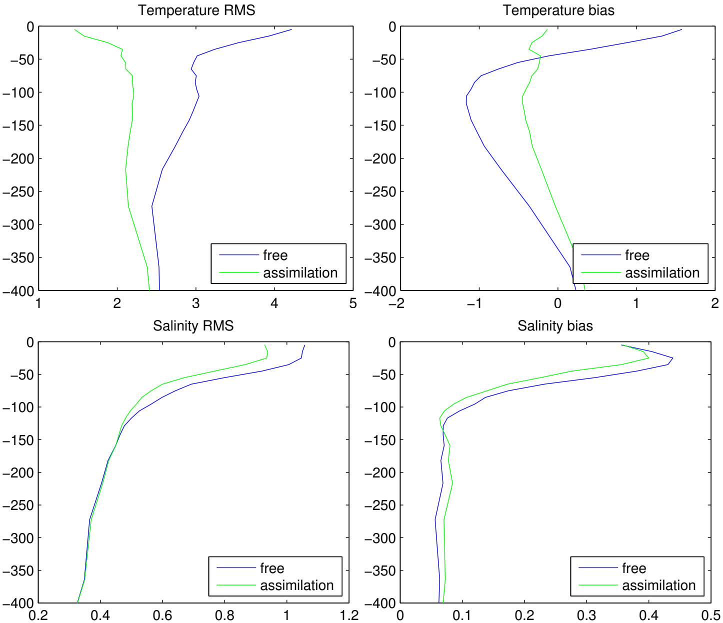
- Low model RMS differences in Weddell and Ross Sea correspond well to the ensemble spread.
- In these areas, ice-drift is directed off-shore.
- Higher ensemble-spread in open ocean, in agreement with RMS difference

Meridional ice-drift



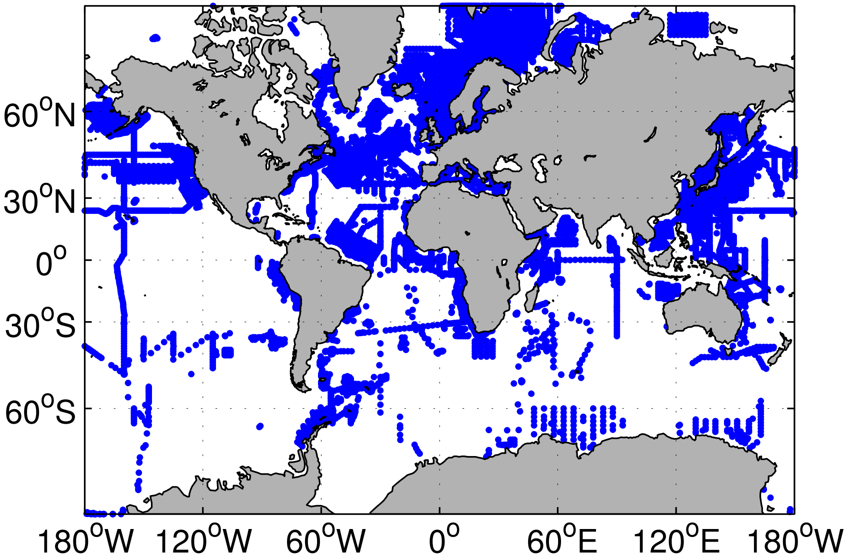


Validation with world ocean data base



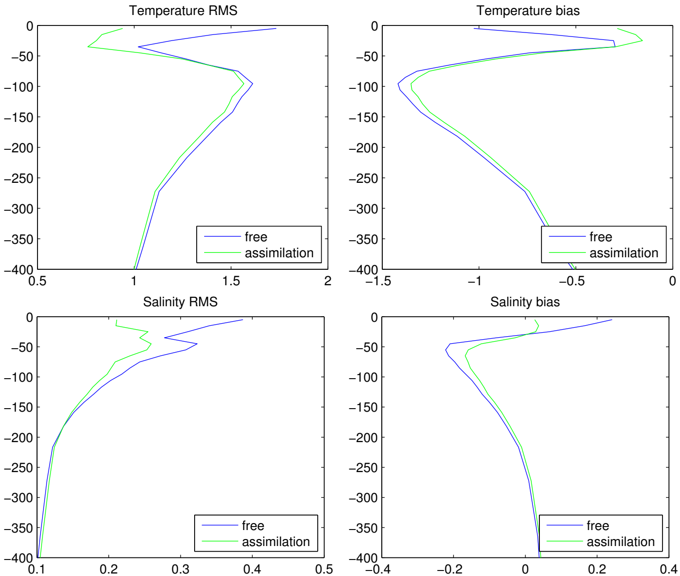
- All profile data in World Ocean Database (observations are interpolated vertically to model grid)
- Temperature presents a quite large error in free model run near the surface.
- Data are often in dynamic regions and average might not be representative.
- RMS and bias reduced compared to this independent data set.

Data distribution



- Data distribution of World Ocean Database for year 1985
- Very inhomogeneous distribution
- Average over complete data-set is biased towards the Northern Hemisphere

Validation with World Ocean Database south of 60°S



- RMS difference between model and observations is dominated by the bias (observation - model)
- In general, model too warm
- RMS and bias reduced also south of 60°S, but less than in Northern Hemisphere.

Conclusions

- First model run of a complete year (~ 30 more to go).
- Start with 1985 since begin of OSTIA time series.
- Ensemble spread is often too small, but structure agrees with the RMS difference of model and observations.
- Improvement of model during analysis persists over the next assimilation cycle.
- Assimilative run reduces error also compared to independent data set (in situ profiles).
  - Significant improvement in temperature
  - Also improvement in salinity

Educational tools

Data Assimilation Demo

This web-page aims to demonstrate the Kalman Filter with some simple linear toy models. First choose model and data assimilation parameters and then click on "Run assimilation"

Model parameters

Model: 1D advection in periodic domain

Equation(s):

$$\mathbf{x}^{(n+1)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \mathbf{x}^{(n)}$$

State vector size: 4

Number of time steps: 40

True initial condition x^i: 1.5, 2, 3, 4

Data assimilation parameters

Method: Kalman Filter

Covariance matrix of initial condition error P^i:

1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Covariance matrix of model error Q:

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Observed every x grid points: 2

Model time steps between observations: 5

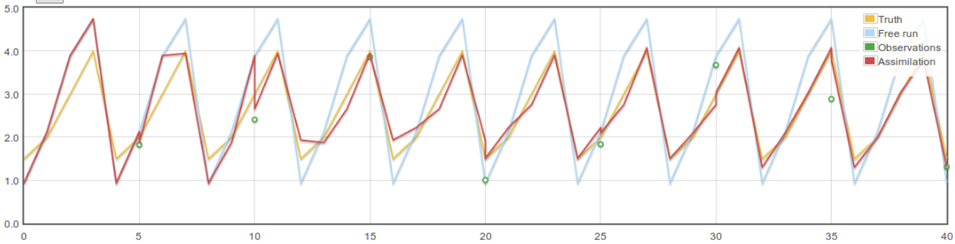
Observation error variance: 0.2

Seed for random numbers: 3

Run assimilation | Reset to defaults

State vector

Index 1 Show: Truth Free run Assimilation Observations

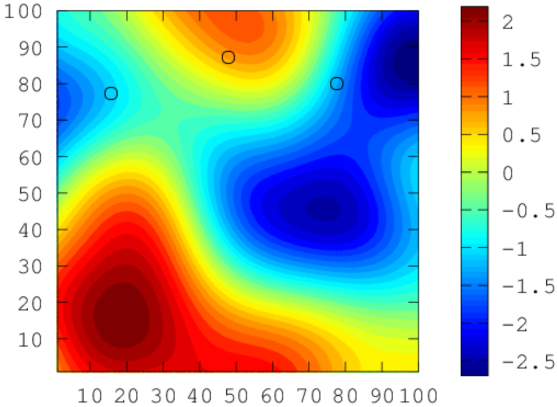


http://www.data-assimilation.net/Tools/

DIVA demo help!

Reference field

Field: Simple random field ▾



Location of observations

Your name:

Correlation length-scale:

Choose the location of your observation (maximum 10)

#	x	y	
1	14.894	77.030	[x]
6	77.508	79.758	[x]
7	47.416	87.030	[x]

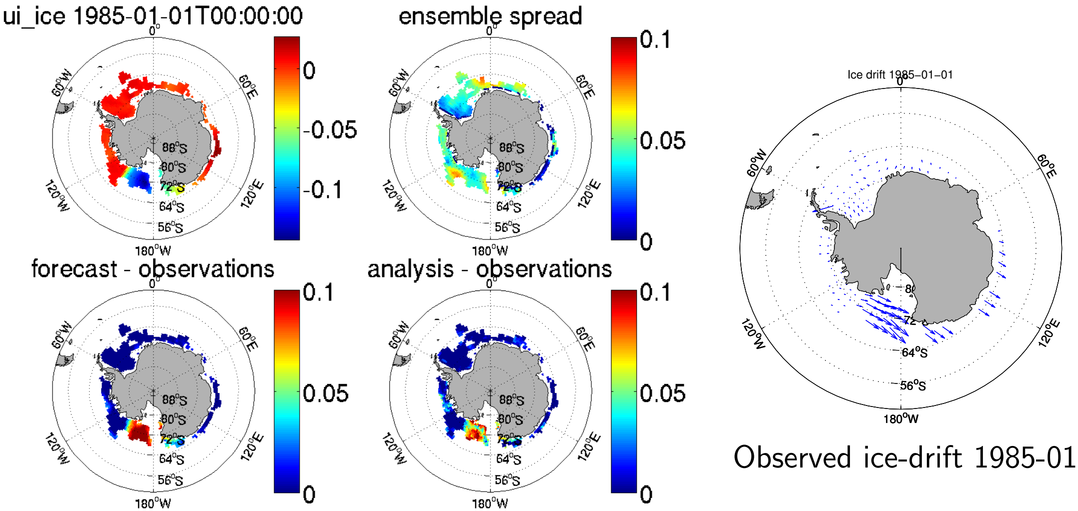
Make analysis | Remove all

Top 5

Name	RMS
JMB	0.2488601797
Ngu	0.2734995952
charles	0.3043944967
it	0.3377809227
Aida	0.3539947015

http://data-assimilation.net/Tools/

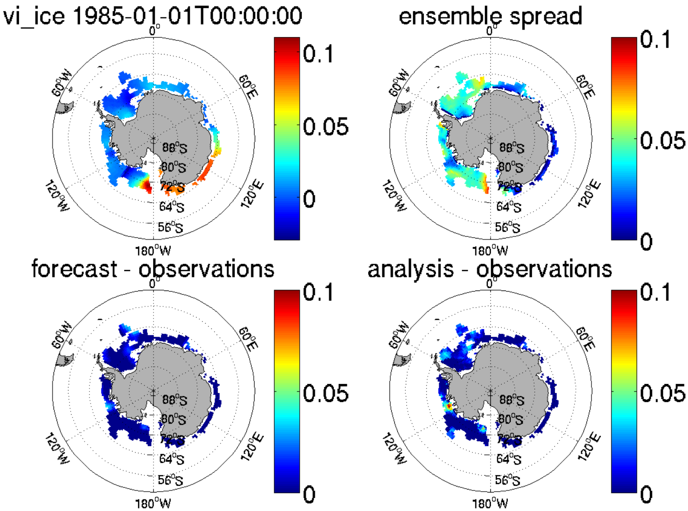
Zonal ice drift

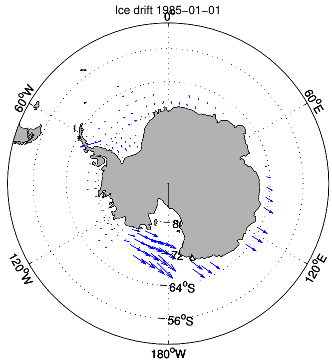


Observed ice-drift 1985-01-01

- Strong zonal ice-drift near Ross Sea, not reproduced by the model
- Not good match for ensemble spread and model forecast-observation difference
- Error reduction but relatively small

Meridional ice drift





Observed ice-drift 1985-01-01

- Surprisingly small error of meridional ice-drift
- Essentially no correction by assimilation