my Ocean

The Research Council of Norway

Data assimilation at NERSC

L. BERTINO, P. SAKOV, F. COUNILLON,
E. SIMON

NERSC, Bergen

SANGOMA KO meeting, Ulg, 25th Nov. 2011

NERSC

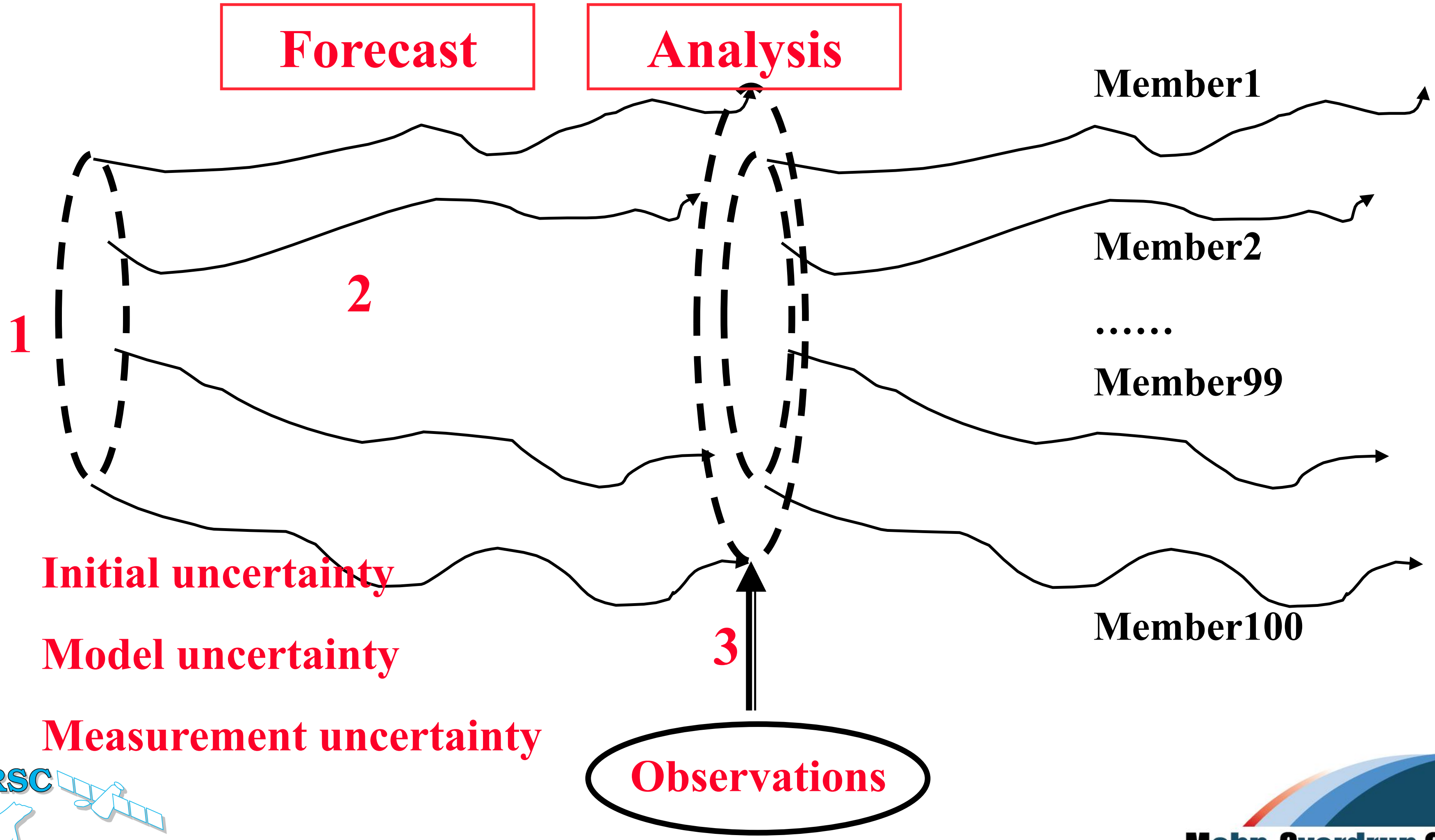
With the support from the eVITA-EnKF project from the RCN and the MyOcean project from the EU

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Outline

- Results from the TOPAZ pilot reanalysis (2003-2008)
- EnKF with Gaussian anamorphosis
- Iterated EnKF for highly non-linear models
- Relevant “Tier 1” R&D activities in MyOcean

Ensemble Kalman filtering



Forecast

Analysis

Member1

Member2

……

Member99

Member100

1

2

3

Observations

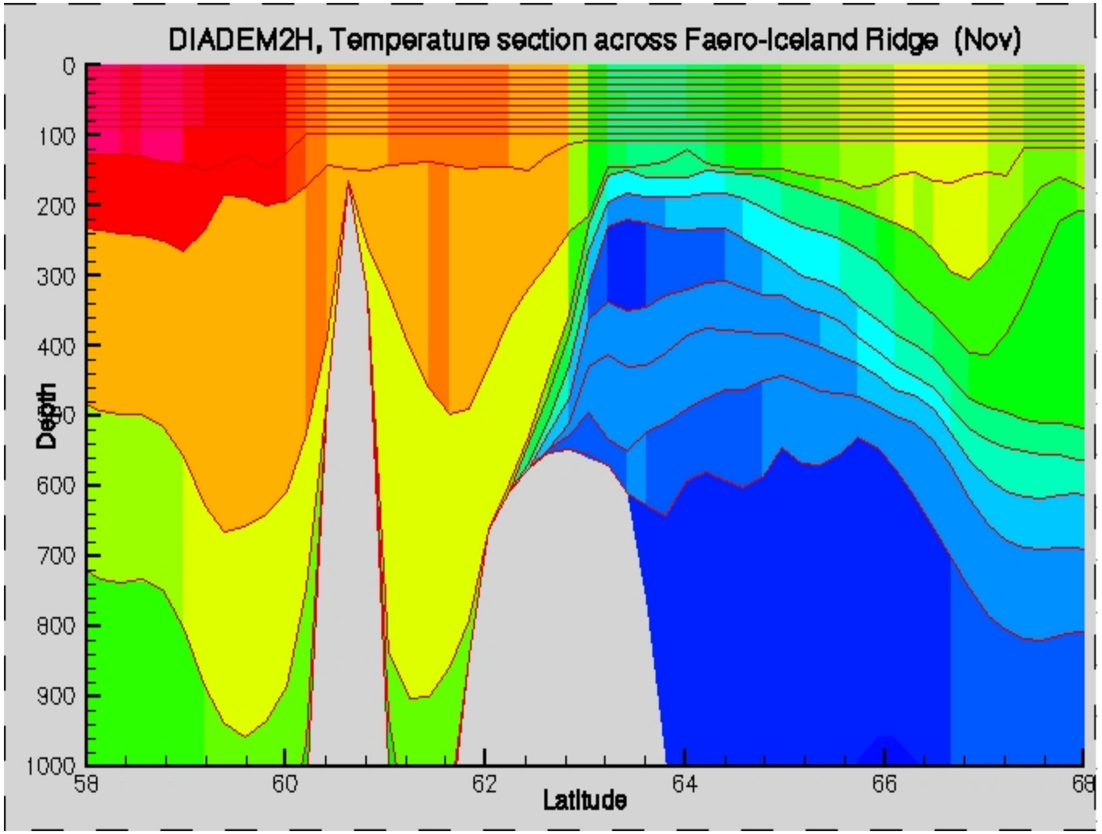
1. Initial uncertainty
2. Model uncertainty
3. Measurement uncertainty

NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

The HYCOM model at NERSC

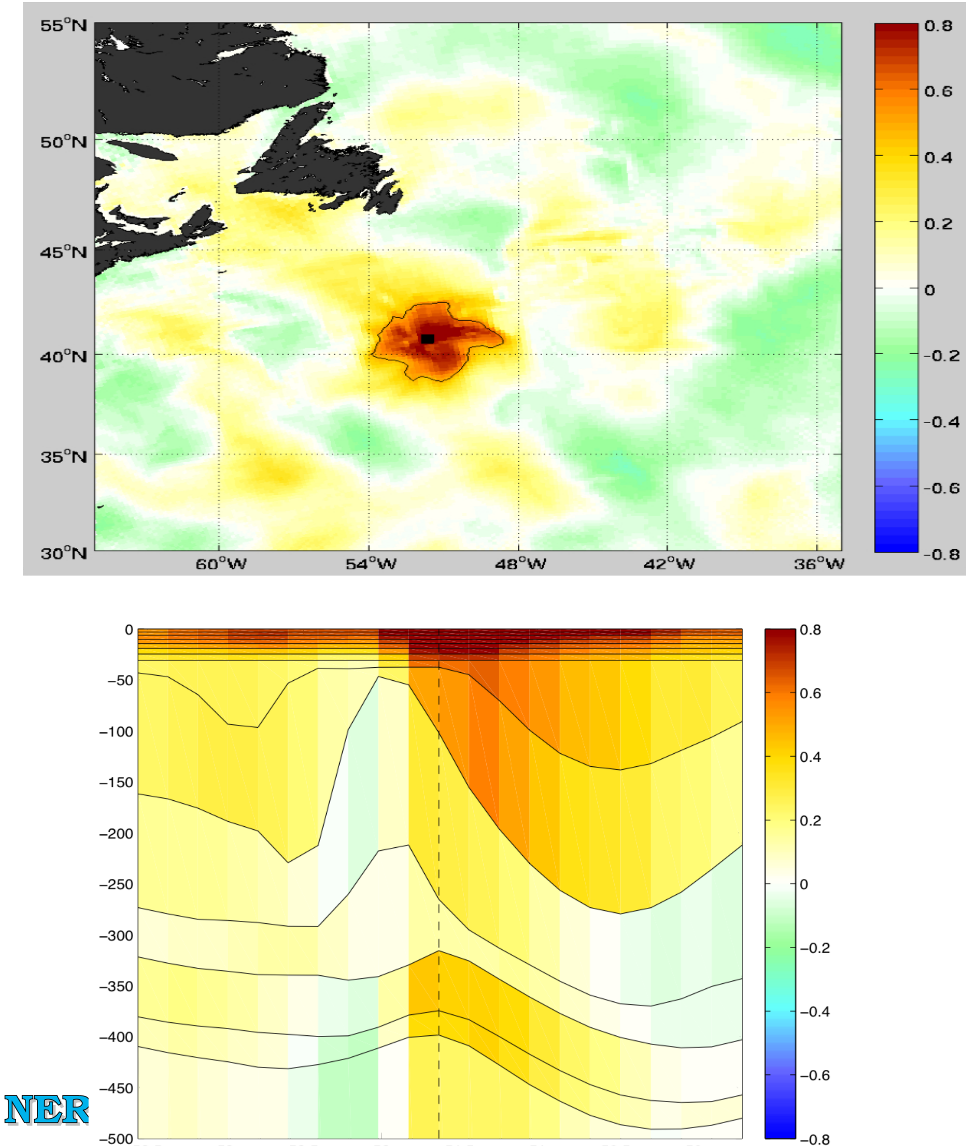
- 3D numerical ocean model
    - Hybrid Coordinate Ocean model, HYCOM (U. Miami)
- Hybrid vertical coordinate
    - Isopycnal in the interior
    - Z-coordinate at the surface
    - TOPAZ4 uses 28 layers
- Coupling to sea ice model
    - EVP dynamics
    - Semtner Thermodynamics
- Data assimilation: EnKF
    - 3D State variables (u,v,T,S,dp)
    - 2D State variables (ub,vb,pb,ice...)
    - Why the EnKf?



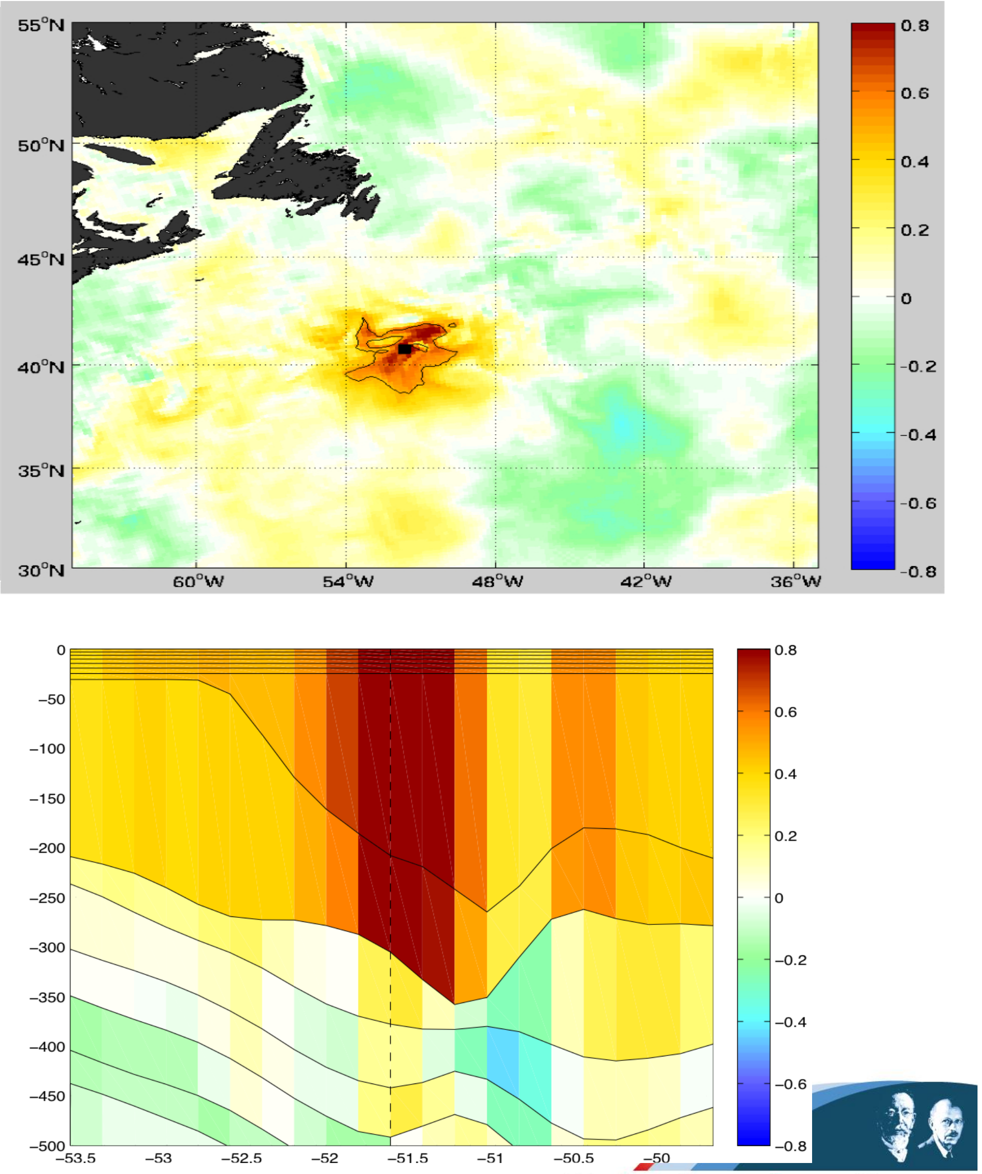
NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

EnKF Correlations, SST



15th June 2008

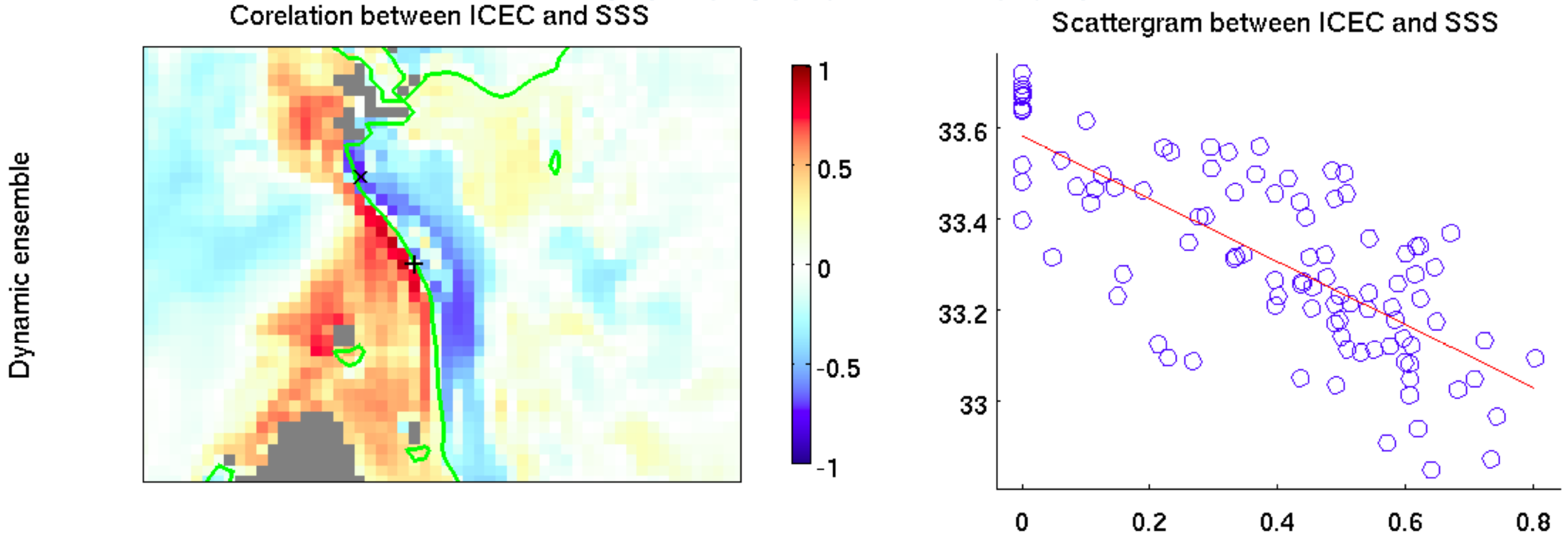


24 Dec. 2008

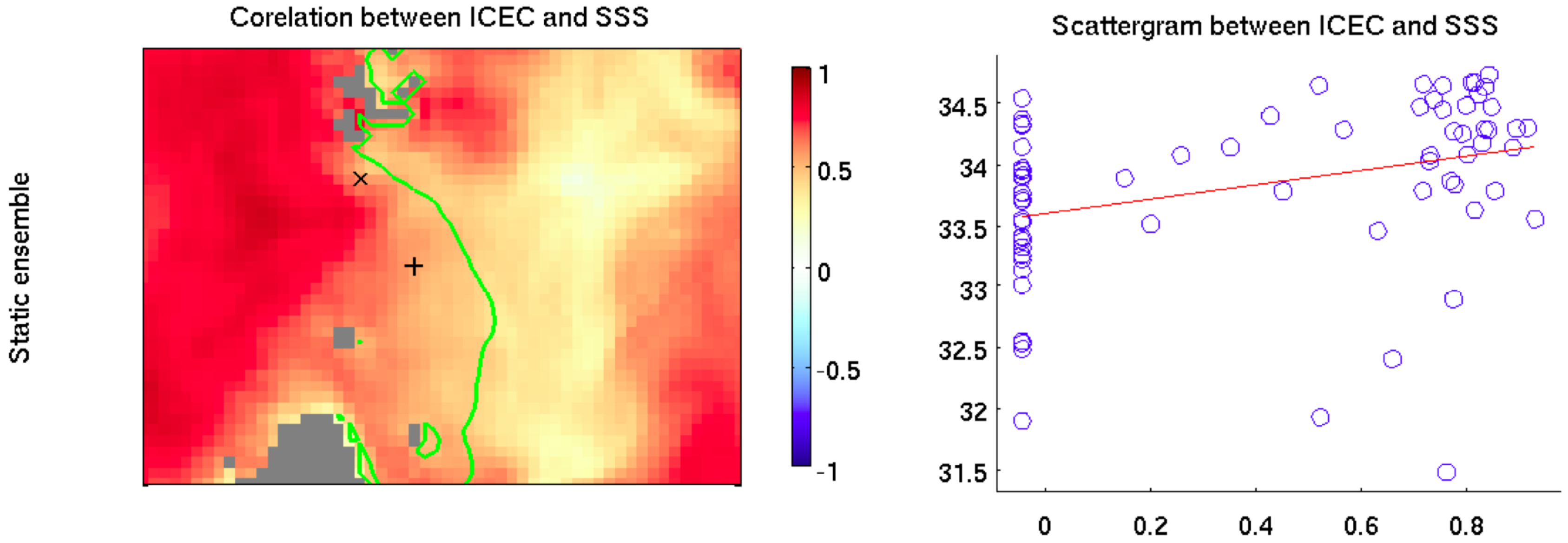
NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Why dynamic Data Assimilation in coupled ice-ocean model?



Also see *Lisæter et al.* Oc. Dyn. 2003



Conservation of properties

Evensen (2003)

- Update equation

  $X^a_n = X^f_n + K_n (Y_n - H(X^f_n))$

- Factorize by X^f_n

  $X^a_n = X^f_n \cdot T$

  T: Transform matrix (size 100 x 100), also sometimes called X_5

Kalman gain:

$$\mathbf{K_n = X_n^{\,f} X'^f_n{}^T H^T \,.}$$

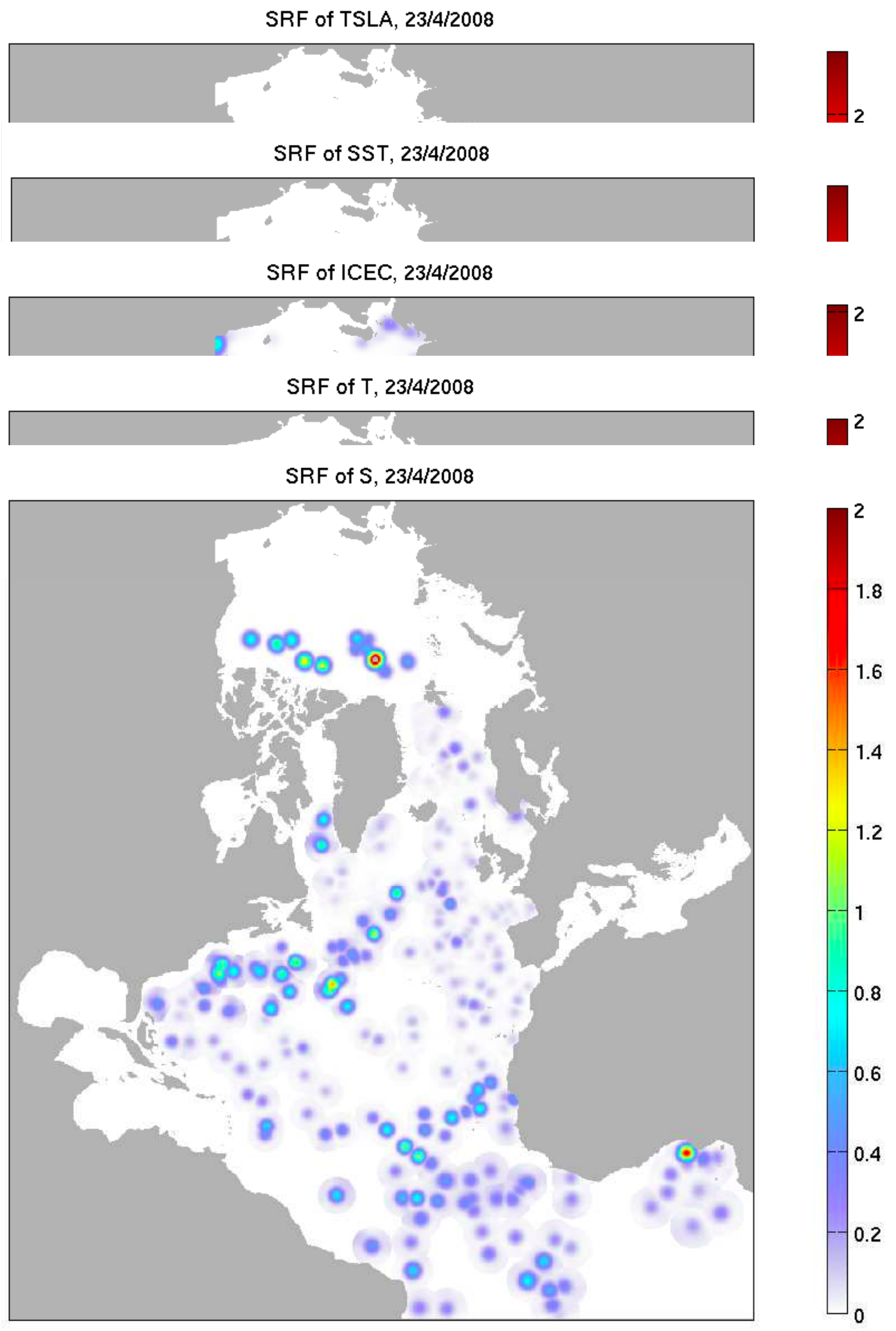
$$\mathbf{(H X'^f_n X'^f_n{}^T H^T + R)^{-1}}$$

The transform T ensures **conservation of linear properties** (geostrophic balance), but not the others.

Ensemble **X**, anomalies **X'** = **X** - <**X**>

The TOPAZ system

- DEnKF, asynchronous
  - 100 members
  - Local analysis (~90 km radius)
- Model state:
  - 3D variables (u,v, T, S, d)
  - 2D variables (ice, …)
  - 800x880x148 = **104 million** variables
- Observations:
  - Sea Level Anomalies (CLS)
  - SST (NOAA, then OSTIA)
  - Sea Ice Concentr. (AMSR, NSIDC)
  - Sea ice drift (CERSAT)
  - T/S profiles (Argo, ITPs, field exps.)
  - **400.000 observations** per week



Computations

DEnKF 100 members

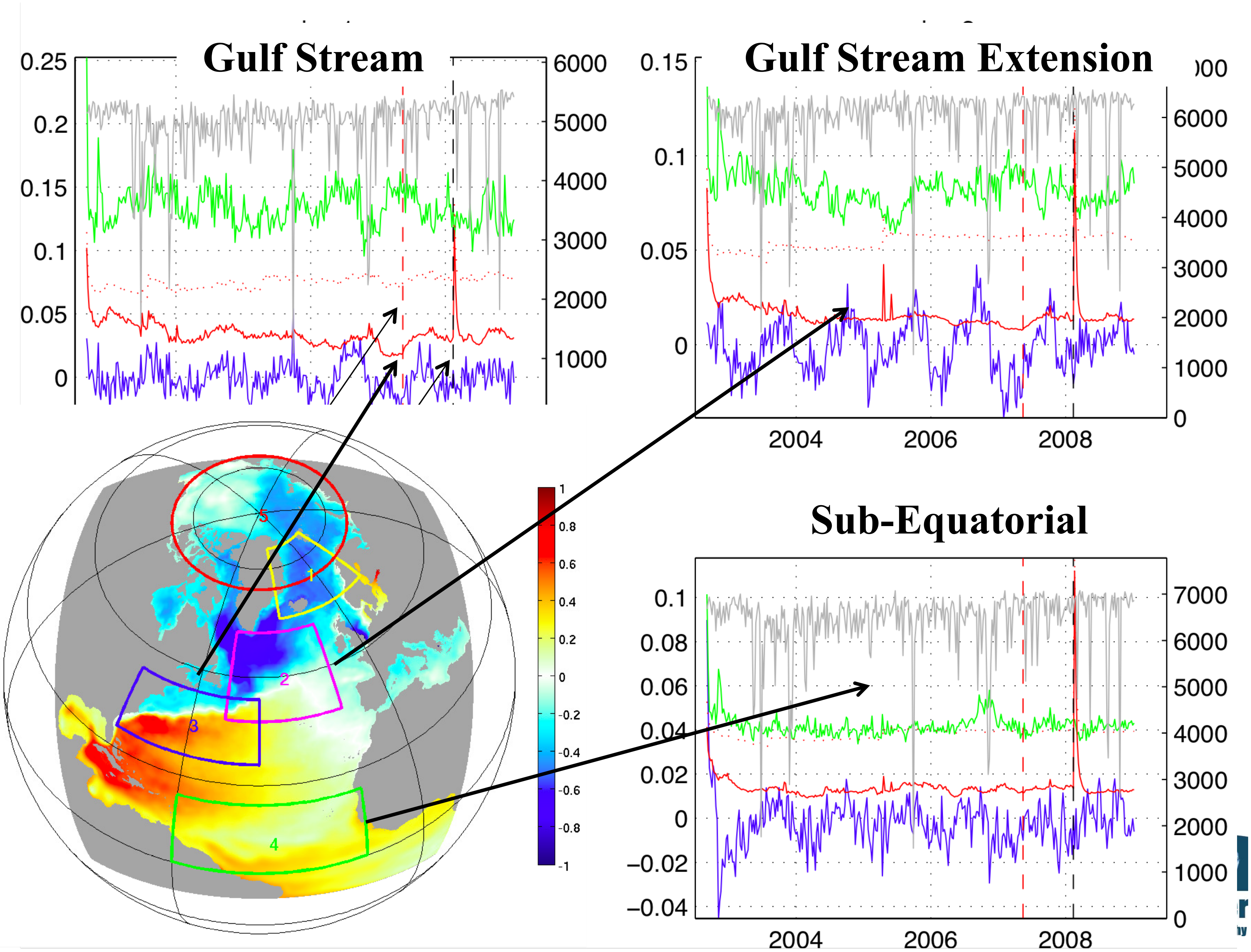
- <u>**Ensemble Forecast**</u>
- 2500 CPU hours / cycle
- Embarrassingly parallel
- 100x **133 CPU 11 min** jobs
- Each job requires **400 Mb**
    - MPI parallelization

- <u>**Analysis**</u>
- 20 CPU hours / update
- 6 datasets simultaneously
- One **20 CPU 1h job**
- Memory required **1 Gb**
    - MPI parallelization

- **HPC Machine:**
- Cray XT4, Installed 2008
    - 5500 CPUs, 55 Tflop/s
    - 1375 nodes (quad-core)
    - 1-4 Gb per node

Avoiding ensemble collapse?

- Initial error
  - Interannual variability + ensemble run with model errors
- Model errors:
  - Winds, air temperature,
  - e-minus-p, cloud cover,
  - Static parameters:
    - mean SSH, mean SST, sea ice rheology parameters (stress tensor)
- Remediation of EnKF shortcomings:
  - Inflation: 1%
  - *Moderation* of observations
    - *Adaptive pre-screening of observations (if pdfs do not match: stretch!)*
    - *Exaggerate obs errors x2 for anomalies update only*

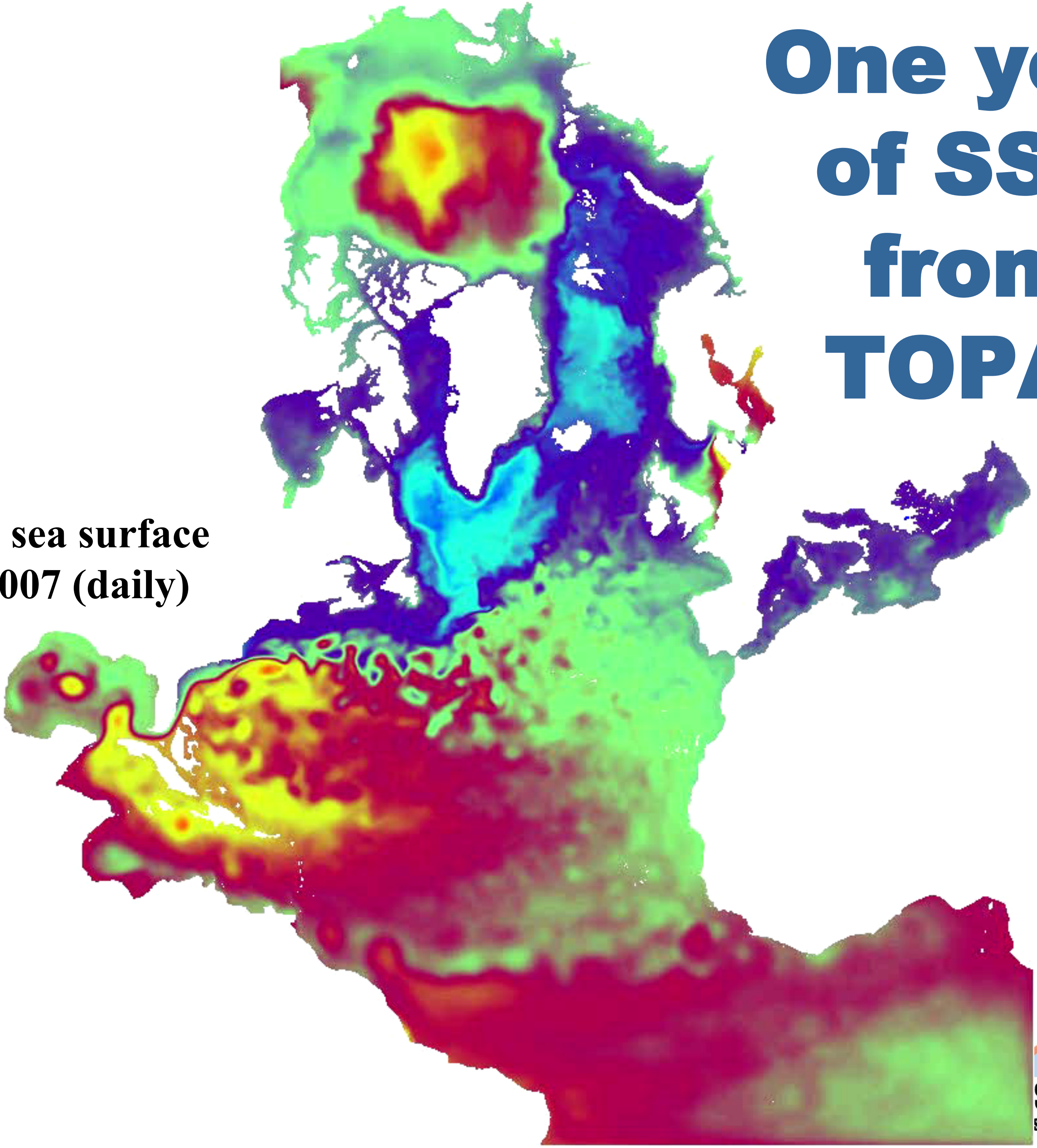


Gulf Stream

Gulf Stream Extension

Sub-Equatorial

One year of SSH from TOPAZ

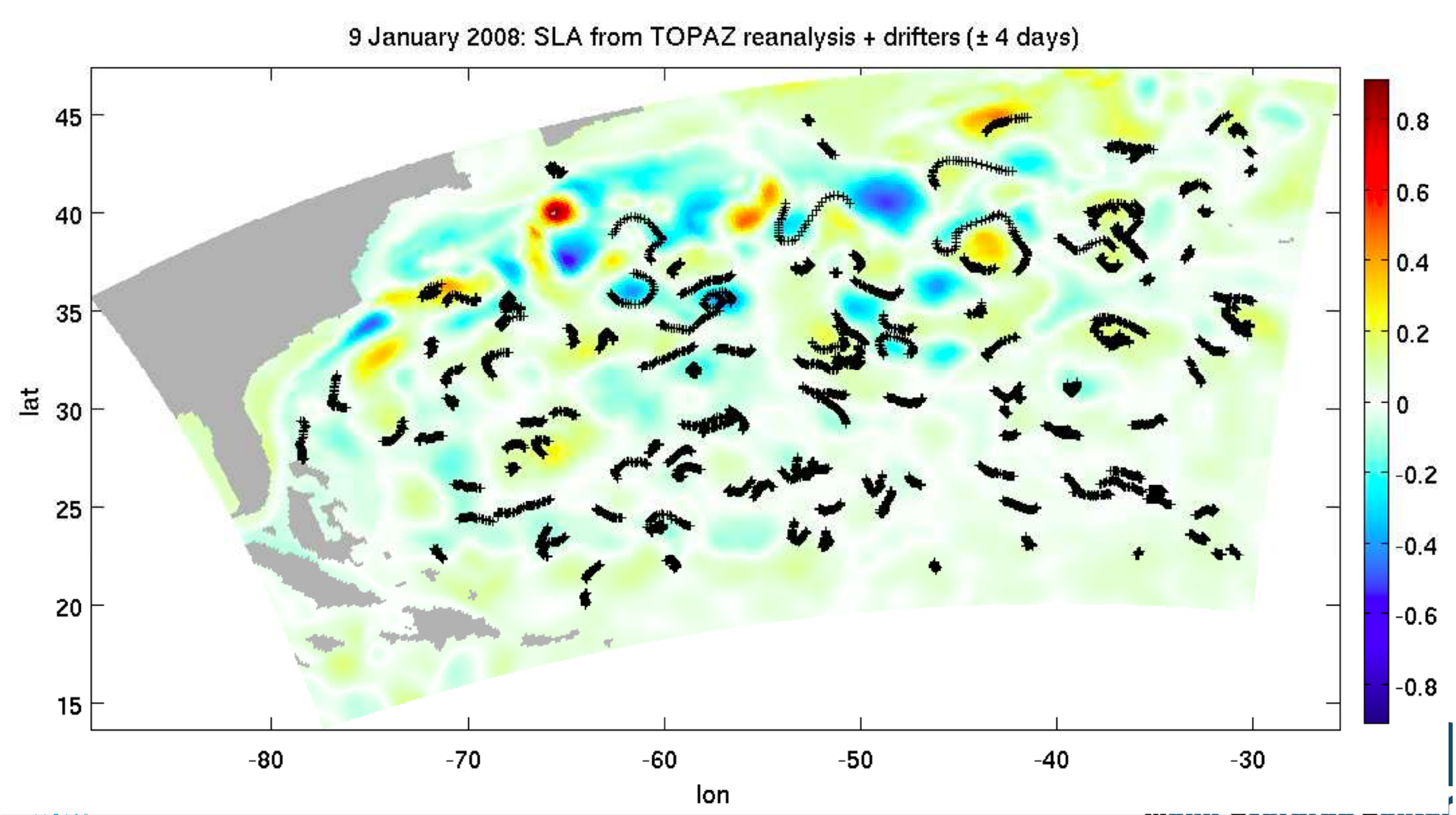


Reanalyzed sea surface heights in 2007 (daily)

NERSC

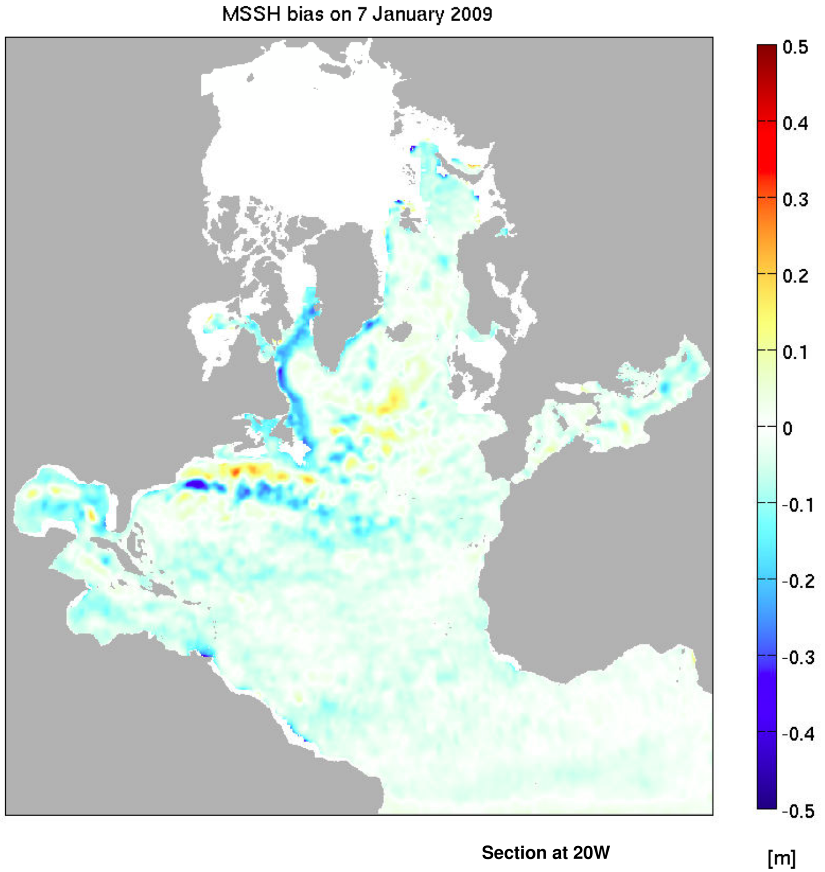
Sverdrup Center
Studies - Operational Oceanography

Independent data: surface drifters

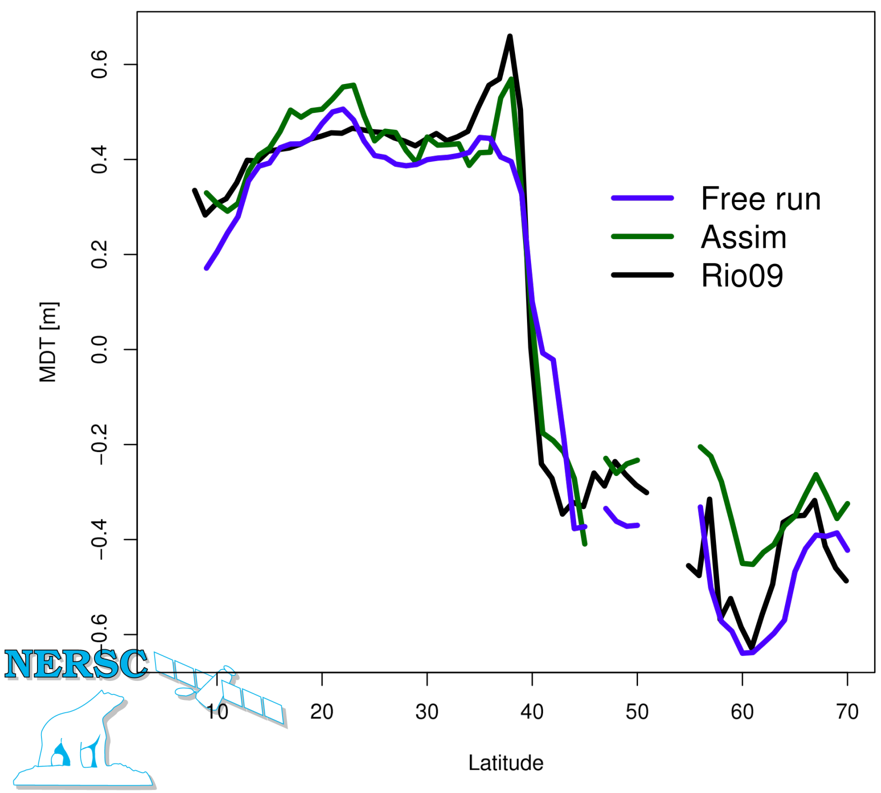


9 January 2008: SLA from TOPAZ reanalysis + drifters (± 4 days)

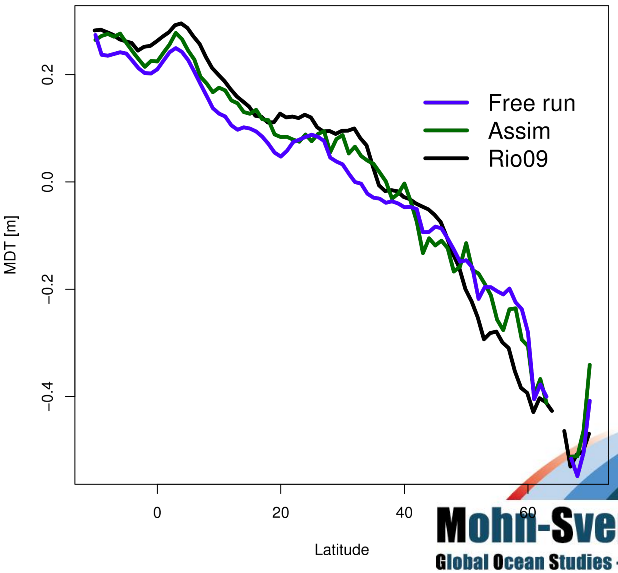
Bias estimation


MSSH bias on 7 January 2009

Section at 60W



Section at 20W

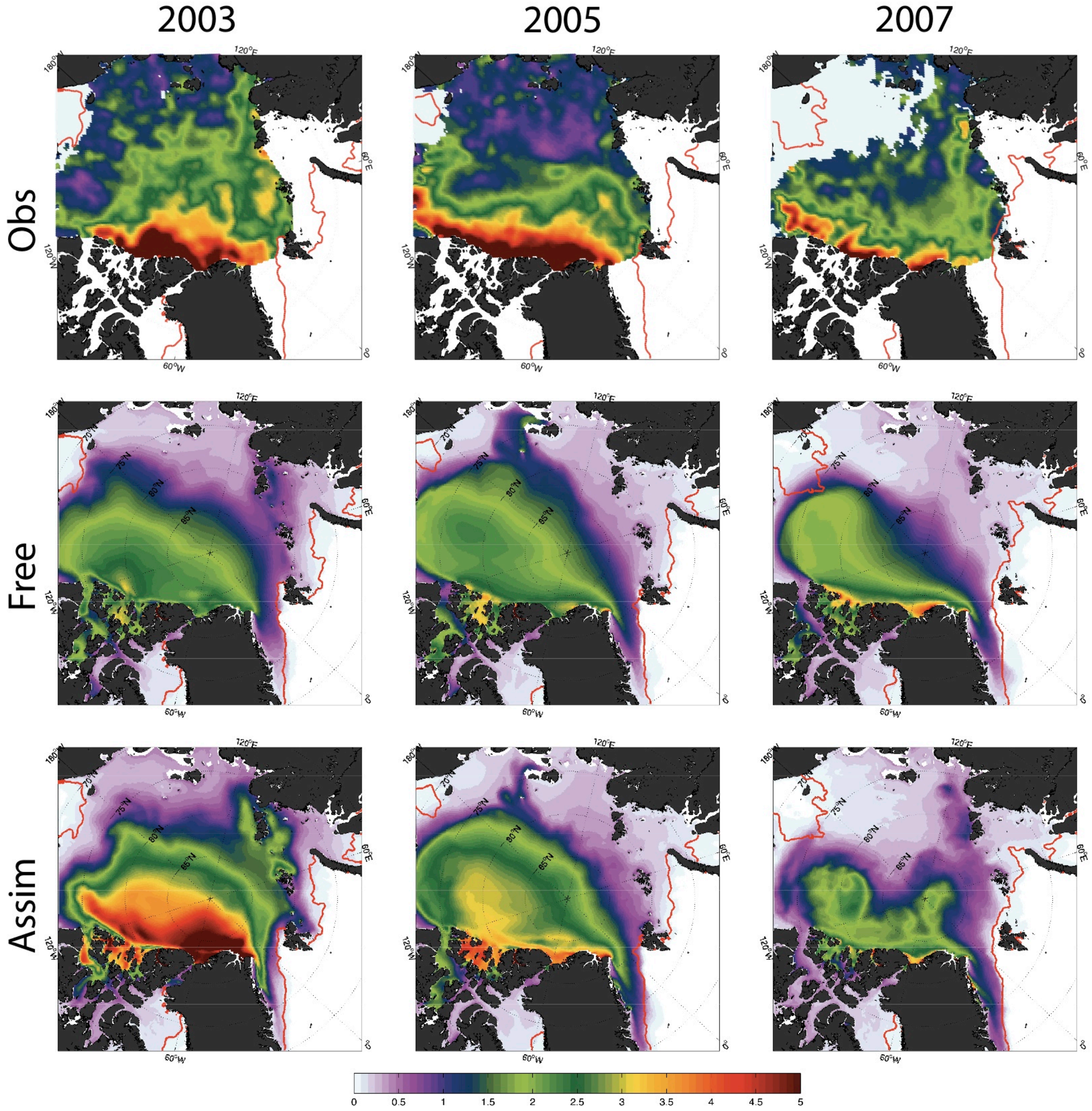


NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Not observed: Ice thickness

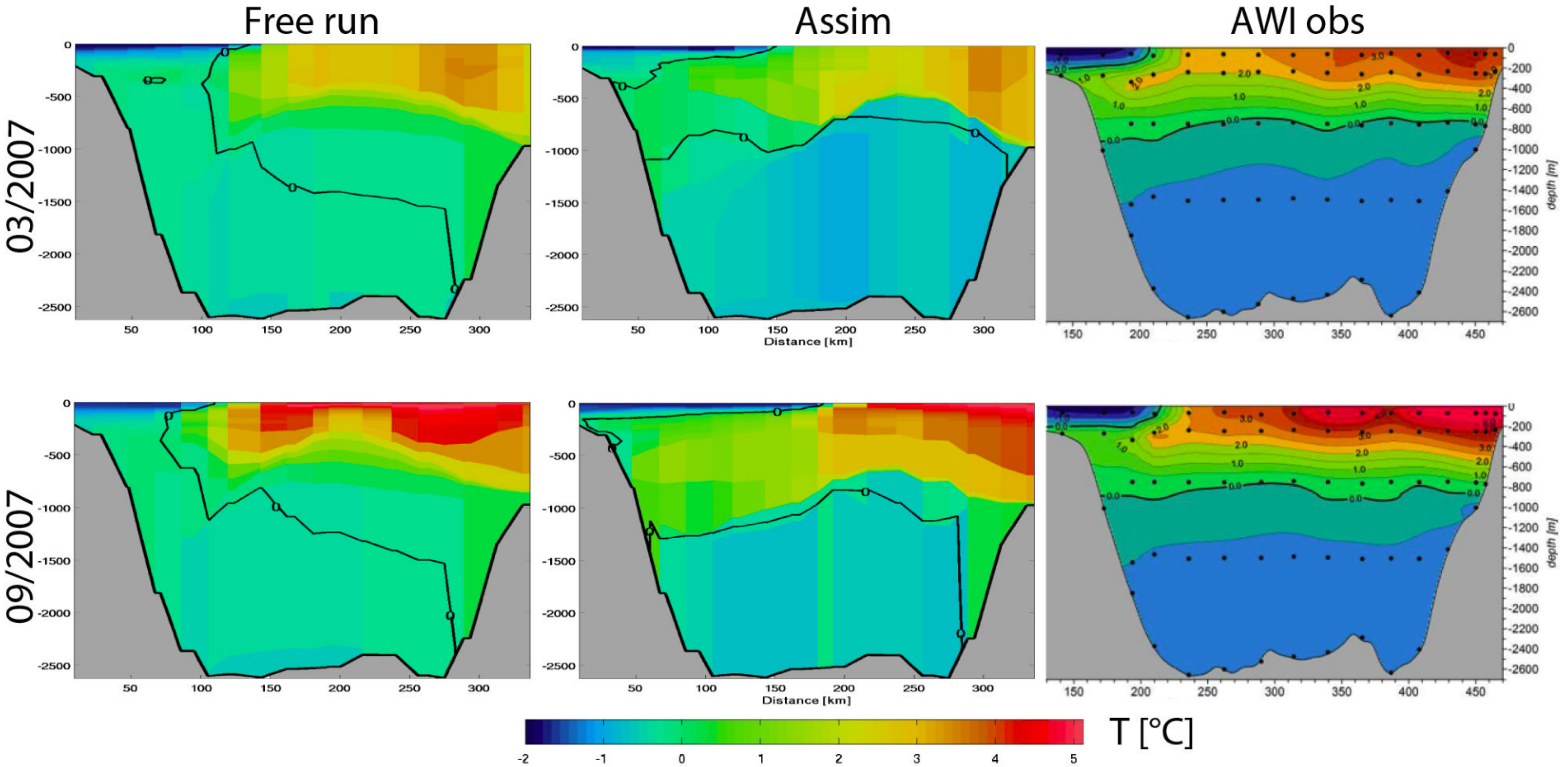
Independent satellite data



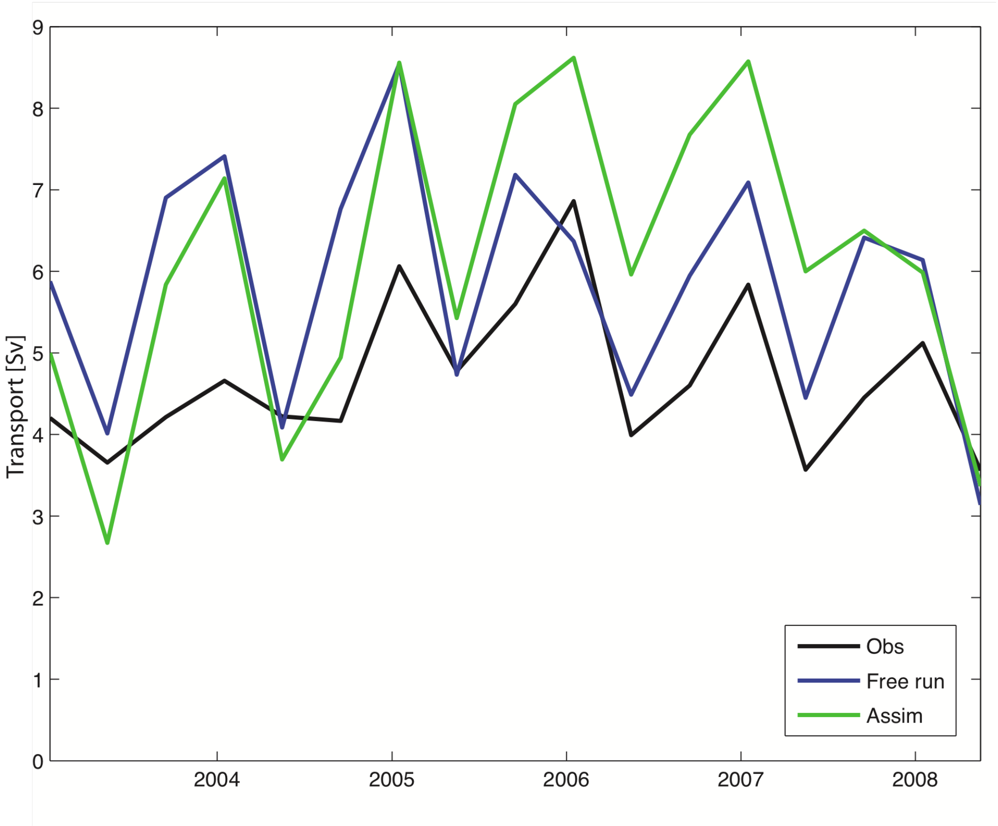
NERSC

ın-Sverdrup Center
cean Studies - Operational Oceanography

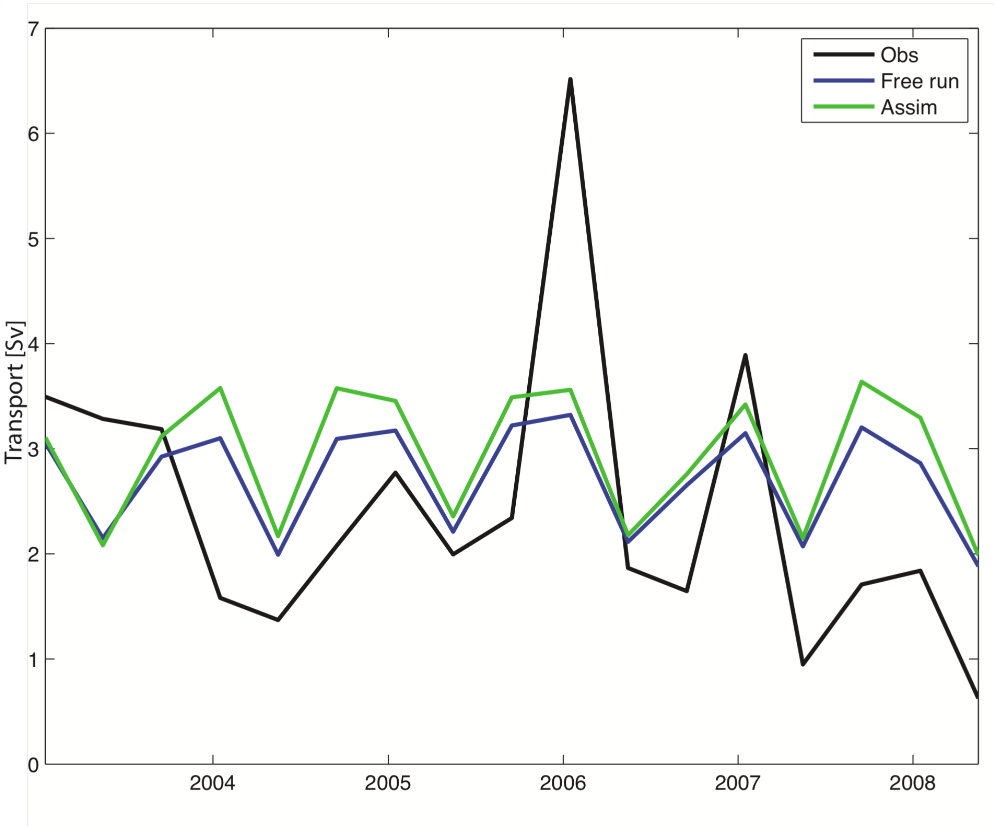
Oceanographic Validation



Water Transport across sections



Svinøy Section
Net flux



Barents Sea Opening
Net flux

NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

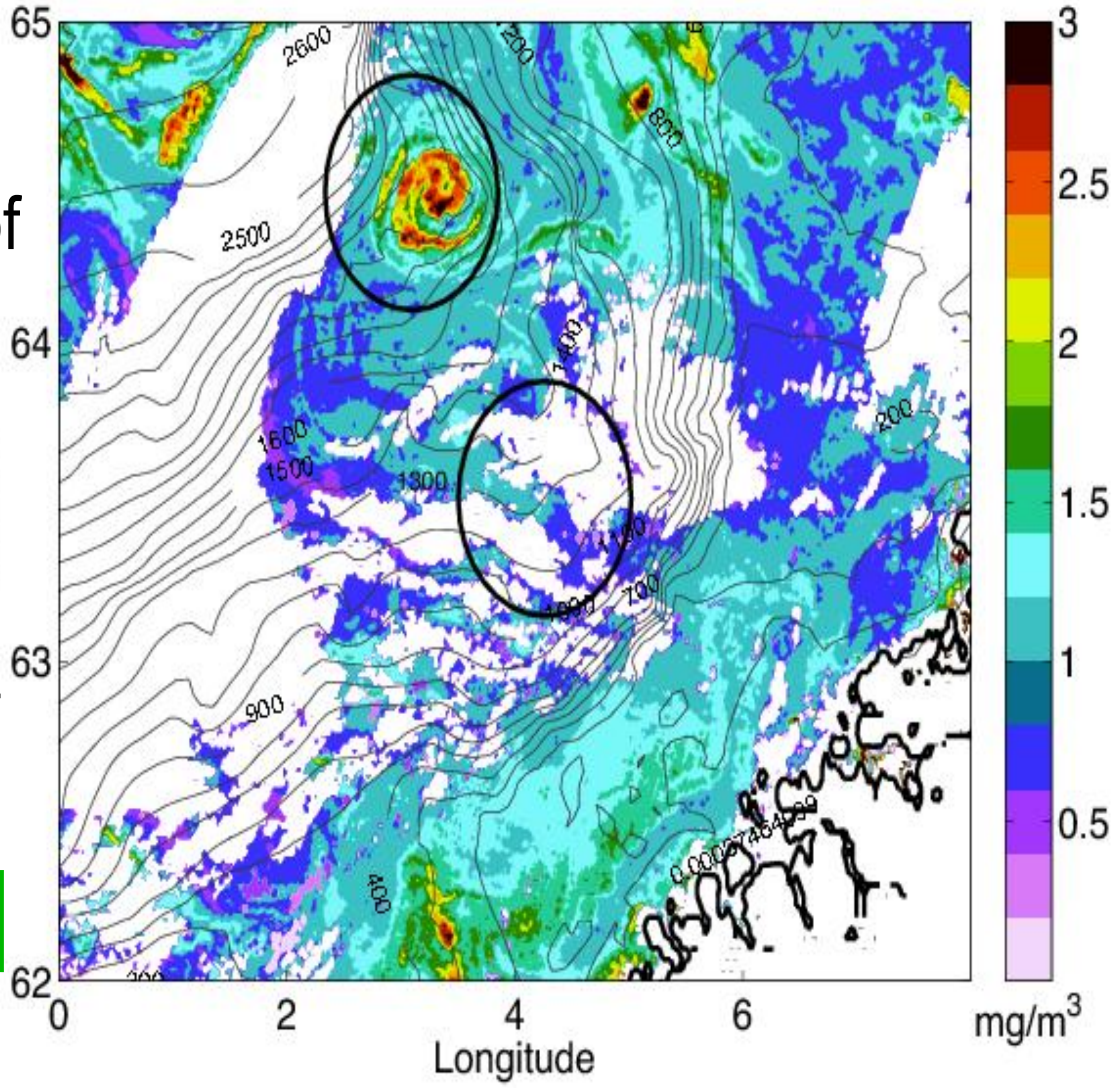
Conclusion

- Reanalysis publicly available on http://topaz.nersc.no and http://myocean.eu.org
- Code on https://svn.nersc.no/enkf
- The TOPAZ4 system is running a 20-years coupled ice-ocean reanalysis (ongoing)
    - No ensemble collapse or innovation drift
    - No assimilation “shocks” in transport time series
- Allows assimilation of various data types
    - Based on Monte-Carlo framework
    - Identified sources of error (Bayesian philosophy)
    - Fully multivariate method

Weather IN the oceans:
Mesoscale physical-ecosystem interactions

- Eddies can make local deserts or local oases of marine life
- **Hansen, C., A. Samuelsen**: Influence of horizontal model grid resolution on the simulated primary production in an embedded primary production model in the Norwegian Sea: Journal of Marine Systems. Vol. 75, Issue: 1-2, pp 236-244, 2009.

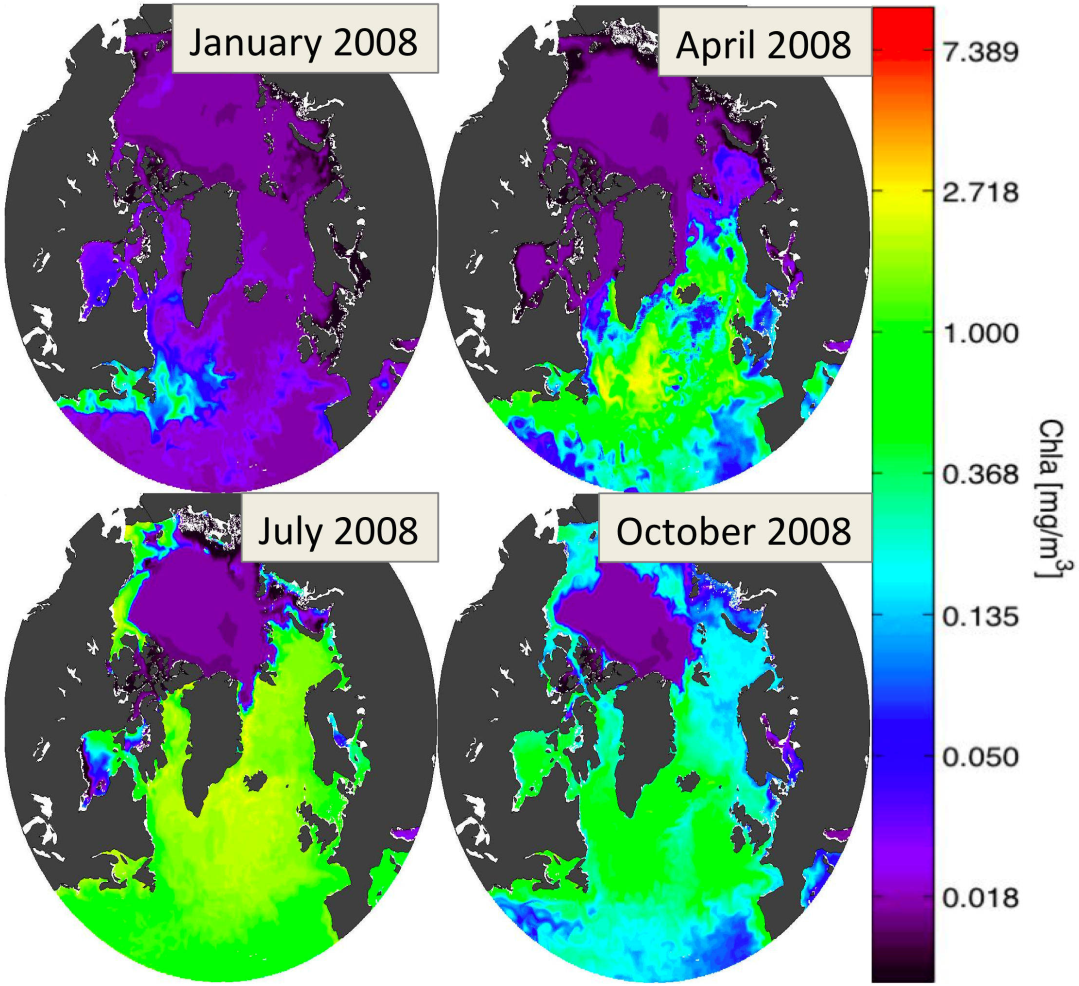
MERIS ocean colour data



C. Hansen

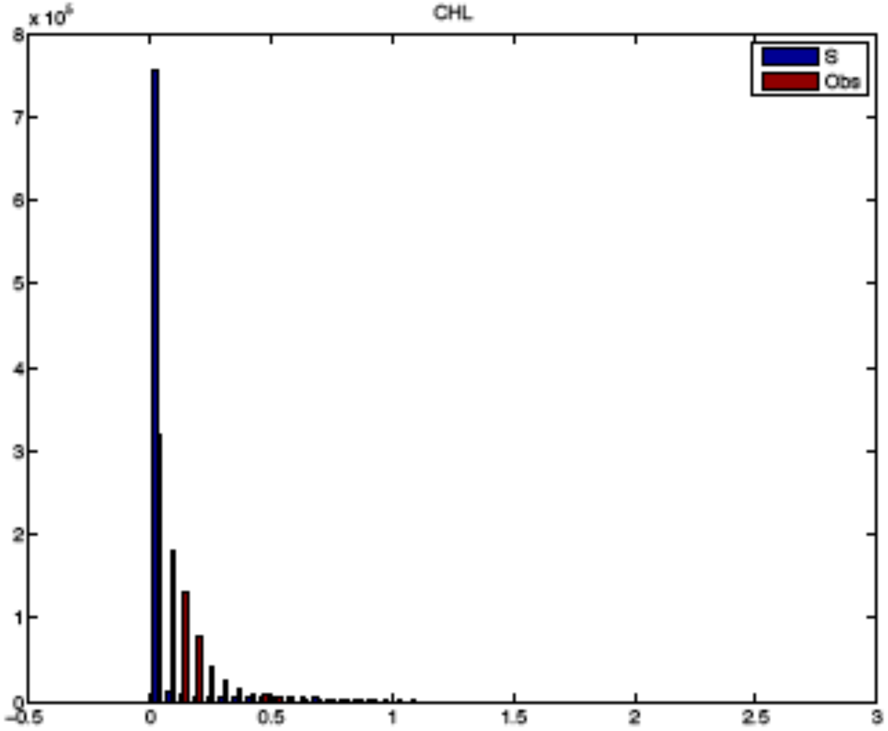
Ocean ecosystem anamorphosis – *EnKF (E. Simon)*

- Data:
  - Satellite, ocean colour
  - SeaWIFS
- Problem
  - Coupled 3-dimensional physical-biological model
  - High-dimensional
  - Non-Gaussian variables

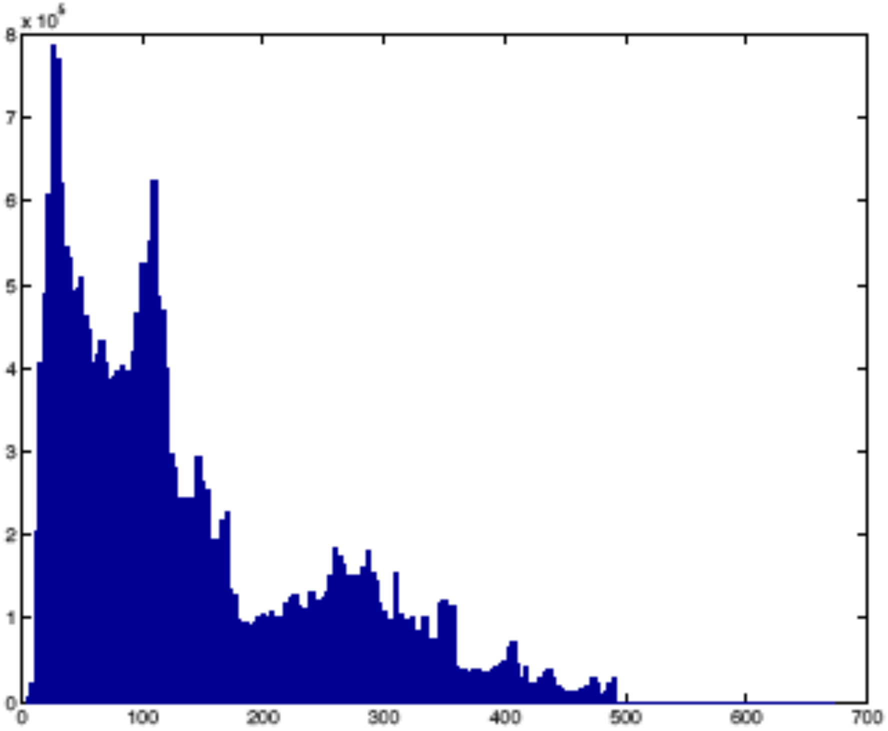


Non-Gaussian variables

Chlorophyll



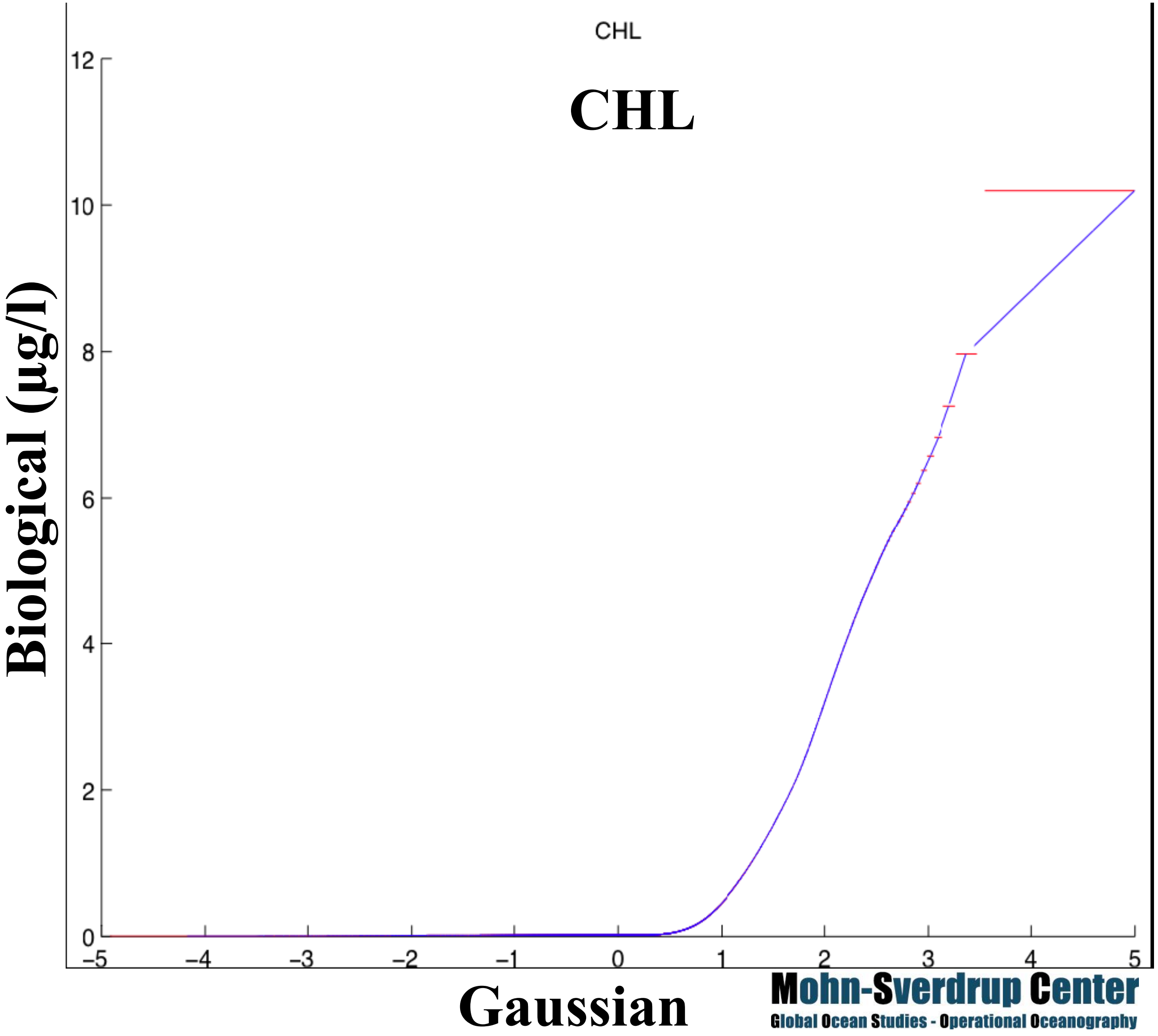
Nitrate nutrients



- Asymmetric distributions (right skewed)
- Positive valued

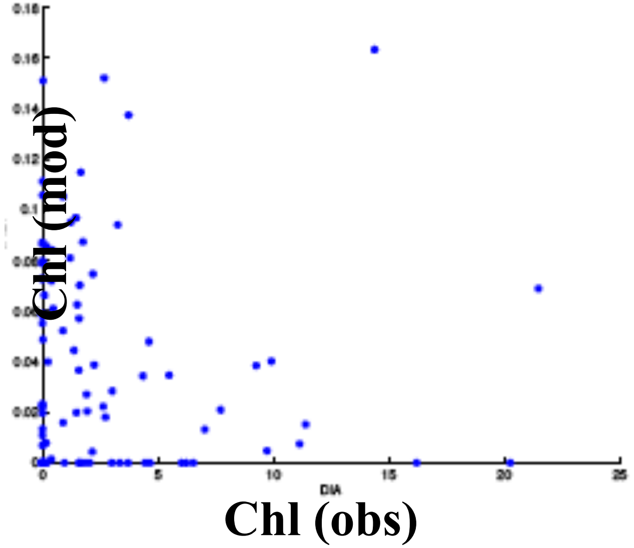
Gaussian anamorphosis

- Univariate transform
  - to Gaussian.
- More adequate for linear analysis
- Idea taken from Geostatistics

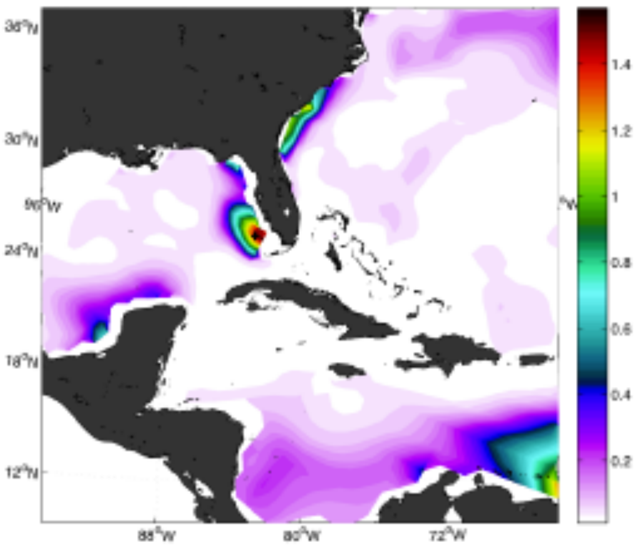


Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

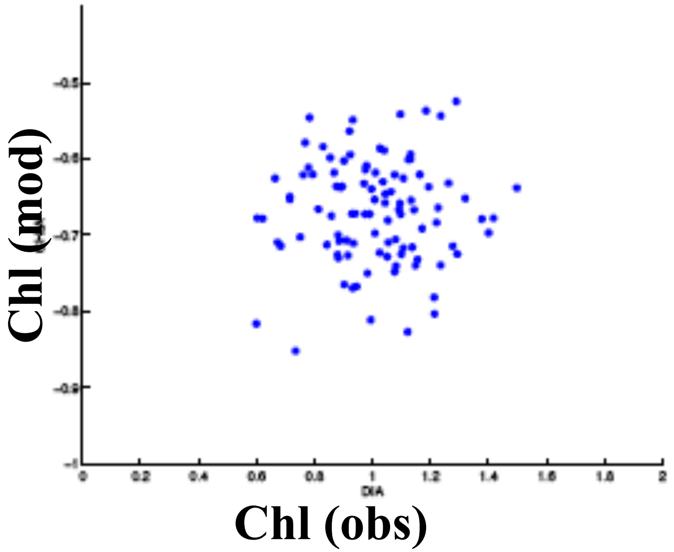
Why is Gaussianity important?



Standard



ECO : Mean of the forecast



ANA

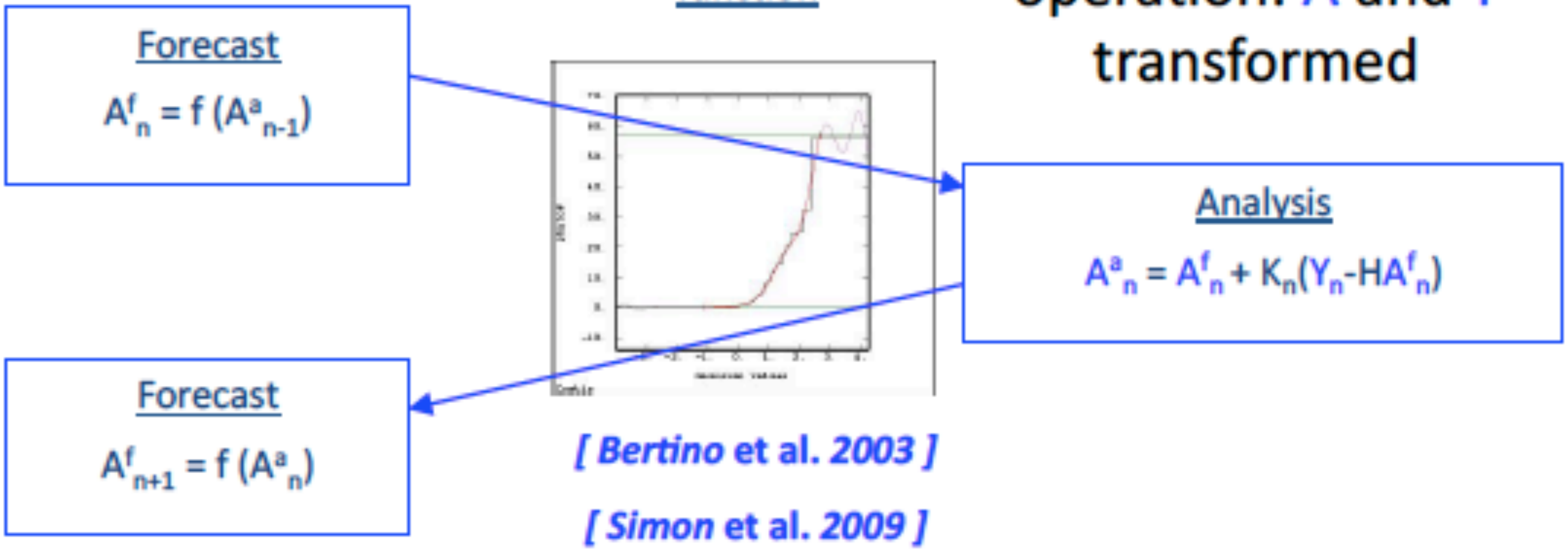
- Makes scatterplots more linear
- Reduces the impact of outliers

Gaussian Anamorphosis

Physical operations:

Anamorphosis function

Statistical operation: A and Y transformed



Forecast

$A^f_n = f\,(A^a_{n-1})$

Analysis

$A^a_n = A^f_n + K_n(Y_n - HA^f_n)$

Forecast

$A^f_{n+1} = f\,(A^a_n)$

[Bertino et al. 2003]

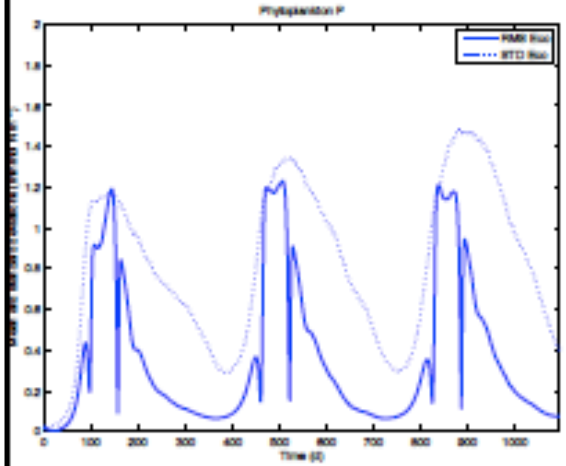
[Simon et al. 2009]

Estimating *loss to carnivores*

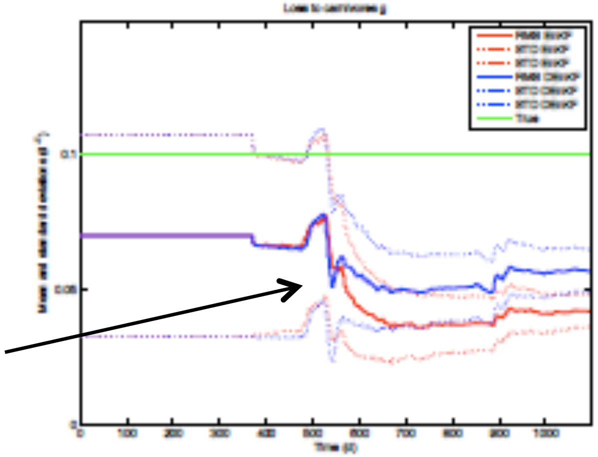
Biased g: parameters

Simon & Bertino, JMS 2011

Errors on phytoplankton



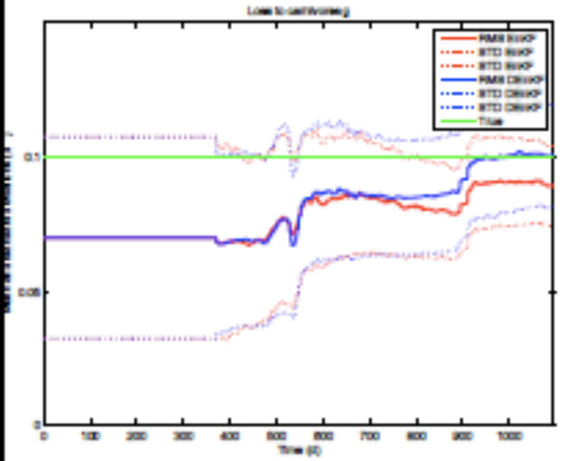
Without anamorphosis



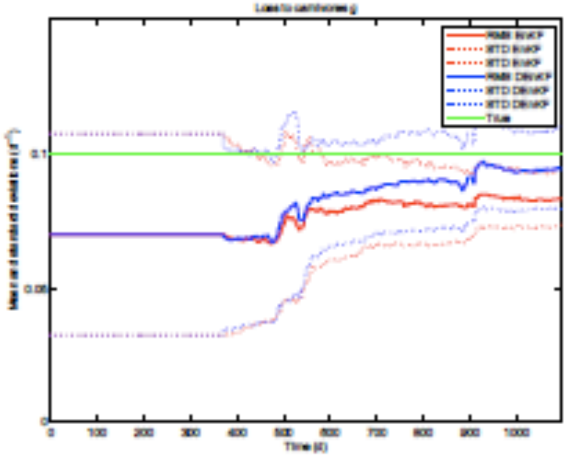
Biased loss to carnivores (g): mean and spread.

Wrong direction!

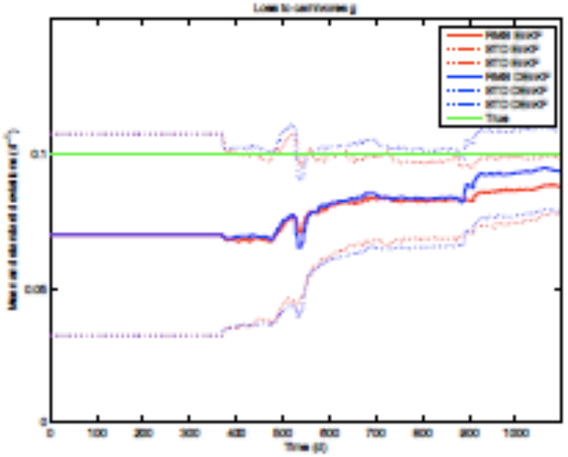
Static anamorphosis



Dynamical anamorphosis



Hybrid anamorphosis



- End of the 1st bloom: correction in the opposite direction without anam.
- Anamorphosis: slow convergence towards the true value (2 blooms).

The iterated EnKF for strongly non-linear systems

P. Sakov, D. Oliver, L. Bertino

Under final (?) review in MWR

NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

The concept

We consider:

- A nonlinear system with infrequent but rather precise observations
- So that propagation of the ensemble anomalies about the ensemble mean is substantially nonlinear ("strongly nonlinear system")
- The system would be less nonlinear after assimilating future observations
- But the assimilation can not be done properly because the sensitivities are imprecise due to the nonlinearity (too big ensemble spread)

We seek solution through iterations of the linear solution

global (can it work?)

Jazwinski (1970): iterations

local

Iterative solution

The objective function: (perfect model framework)

$$\mathbf{x}_1^a = \arg\min_{\{x_1\}} \left\{ (\mathbf{x}_1 - \mathbf{x}_1^f)^{\mathrm{T}} (\mathbf{P}_1^f)^{-1} (\mathbf{x}_1 - \mathbf{x}_1^f) + \left[\mathbf{y}_2 - \mathcal{H}_2(\mathbf{x}_2)\right]^{\mathrm{T}} (\mathbf{R}_2)^{-1} \left[\mathbf{y}_2 - \mathcal{H}_2(\mathbf{x}_2)\right] \right\}, \qquad \mathbf{x}_2 = \mathcal{M}_{12}(\mathbf{x}_1)$$

Solution:

$$\begin{aligned} \mathbf{x}_1^{i+1} =& \mathbf{x}_1^i + \mathbf{K}_{12}^i \left\{ \mathbf{y}_2 - \mathcal{H}_2\left[\mathcal{M}_{12}(\mathbf{x}_1^i)\right] \right\} \\ & + \mathbf{P}_1^i (\mathbf{P}_1^f)^{-1} (\mathbf{x}_1^f - \mathbf{x}_1^i) \qquad \text{(Tarantola, 2005, eq. 3.51)} \end{aligned} \tag{1}$$

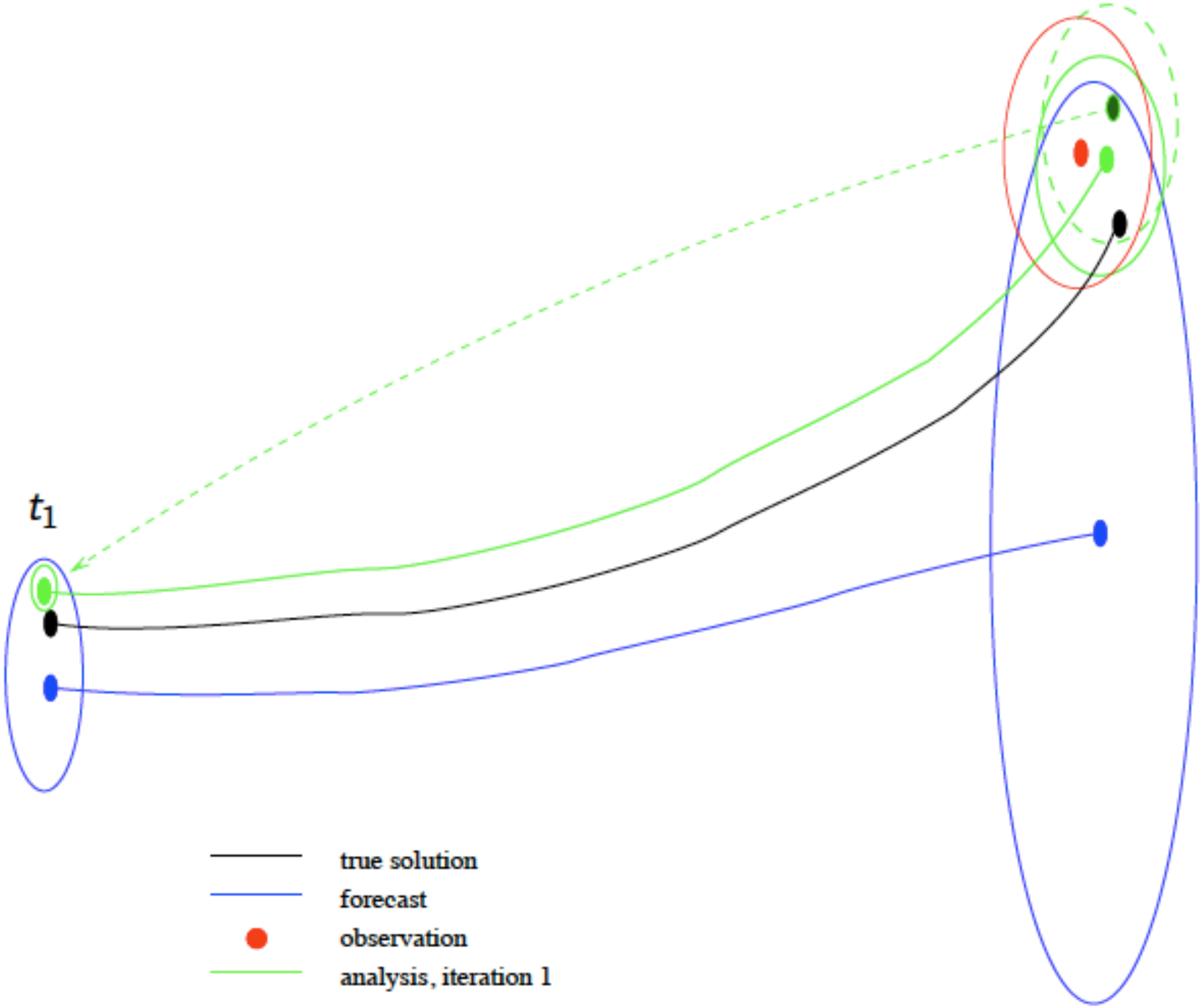
or

$$\begin{aligned} \mathbf{x}_1^{i+1} =& \mathbf{x}_1^f + \mathbf{K}_{12}^i \left\{ \mathbf{y}_2 - \mathcal{H}_2\left[\mathcal{M}_{12}(\mathbf{x}_1^i)\right] \right. \\ & \left. + \mathbf{H}_2^i \mathbf{M}_{12}^i (\mathbf{x}_1^i - \mathbf{x}_1^f) \right\} \qquad \text{(Gu and Oliver, 2007, eqs. 12,13)} \end{aligned} \tag{2}$$

where

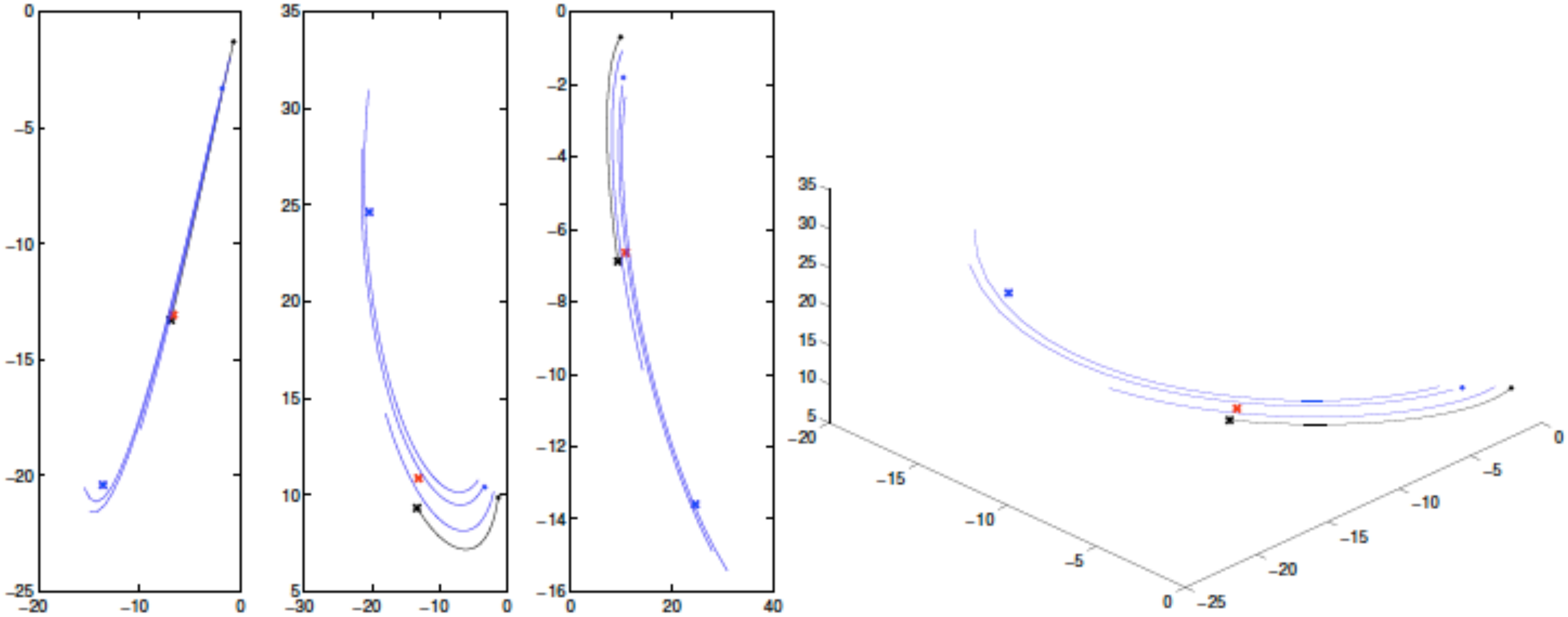
$$\begin{aligned} \mathbf{K}_{12}^i &= \mathbf{P}_1^f (\mathbf{H}_2^i \mathbf{M}_{12}^i)^{\mathrm{T}} \left[\mathbf{H}_2^i \mathbf{M}_{12}^i \mathbf{P}_1^f (\mathbf{H}_2^i \mathbf{M}_{12}^i)^{\mathrm{T}} + \mathbf{R}_2 \right]^{-1}, \\ \mathbf{P}_1^i &= \left[\mathbf{I} - \mathbf{K}_{12}^i \mathbf{H}_2^i \mathbf{M}_{12}^i \right] \mathbf{P}_1^f \end{aligned}$$

The Iterative EnKF



Example

Cycle 10 (1):

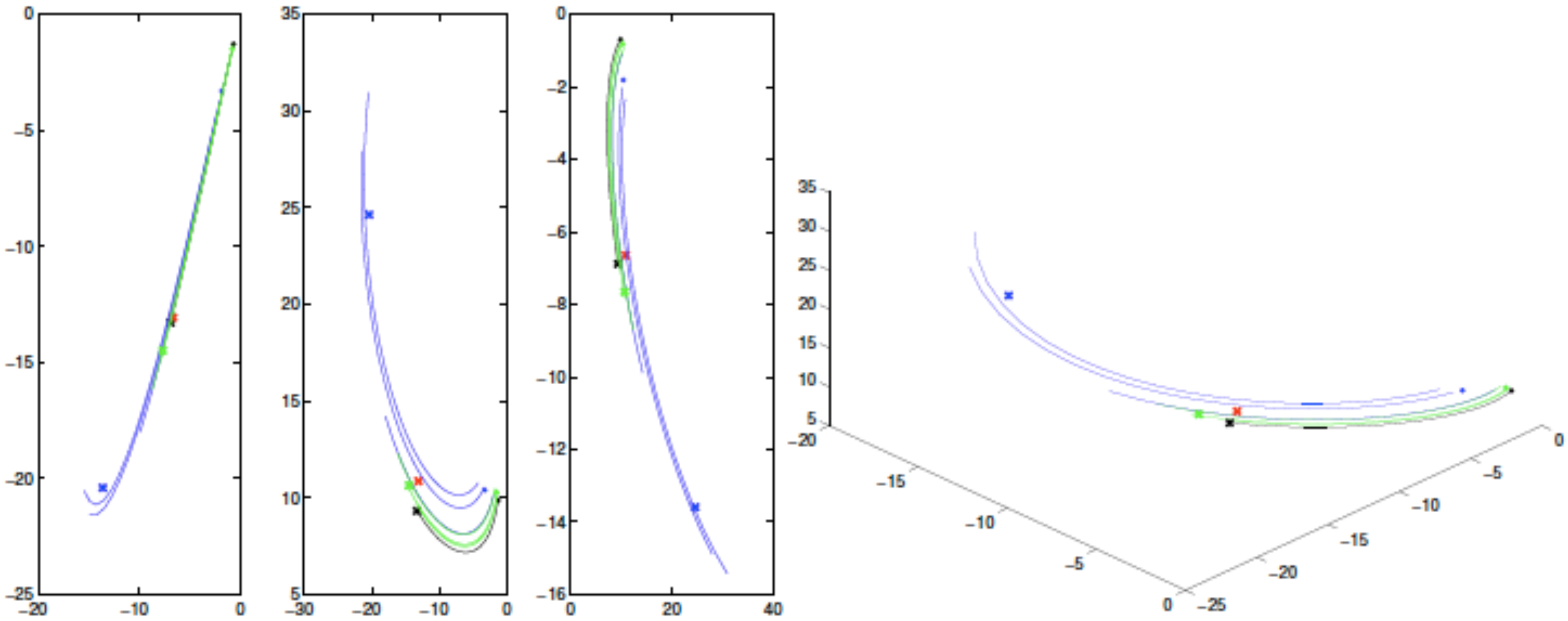


NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

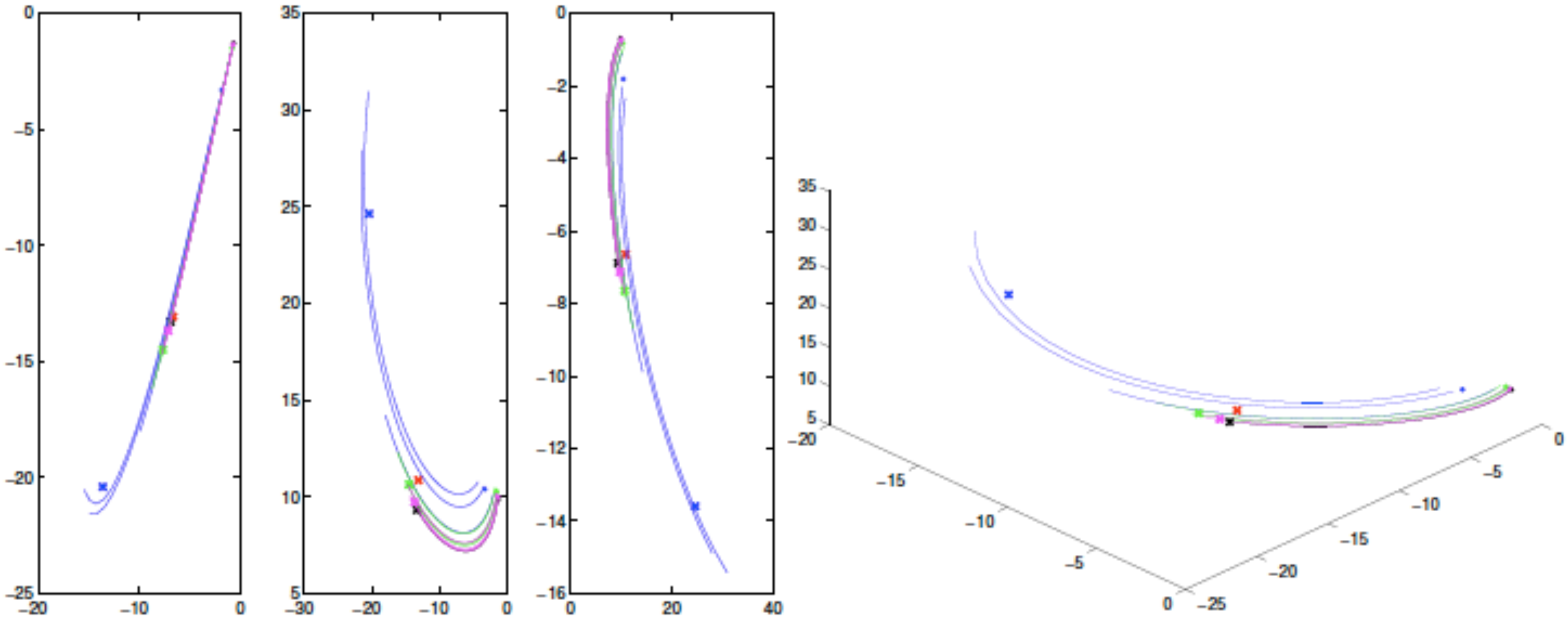
Example

Cycle 10 (2):



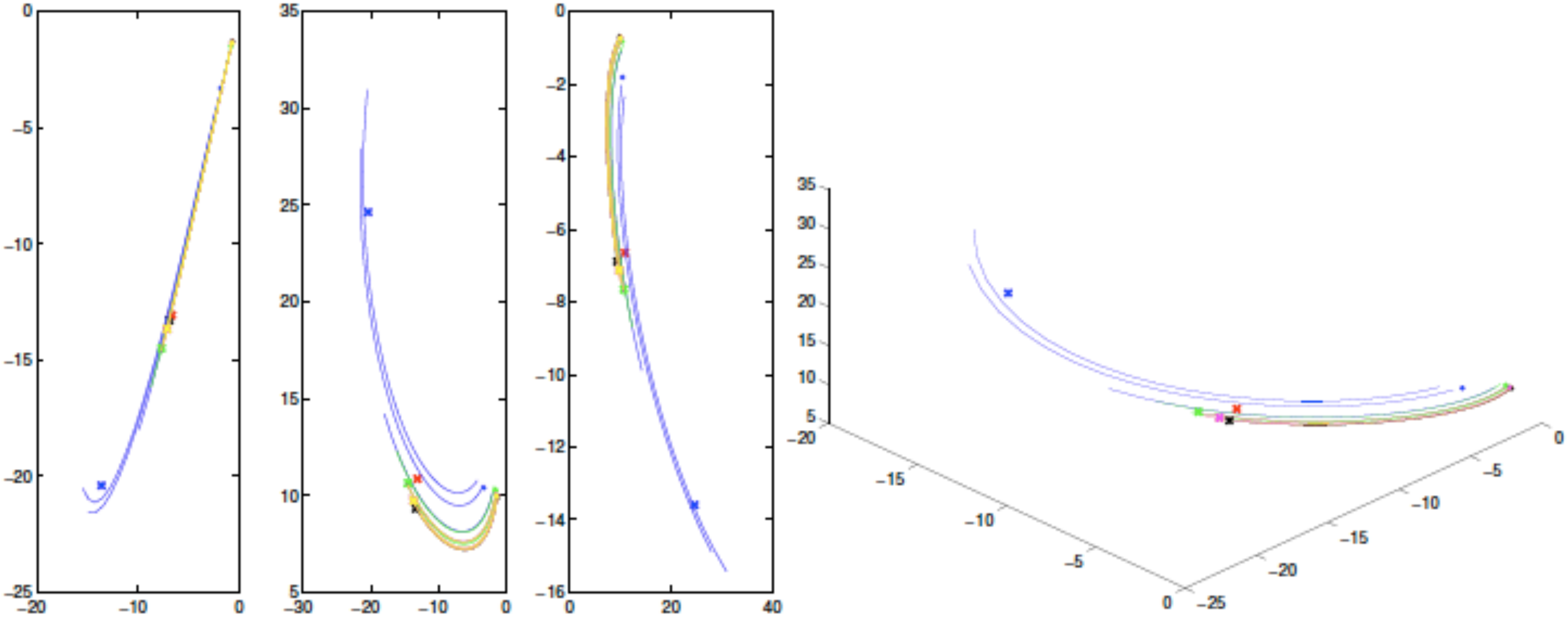
Example

Cycle 10 (3):



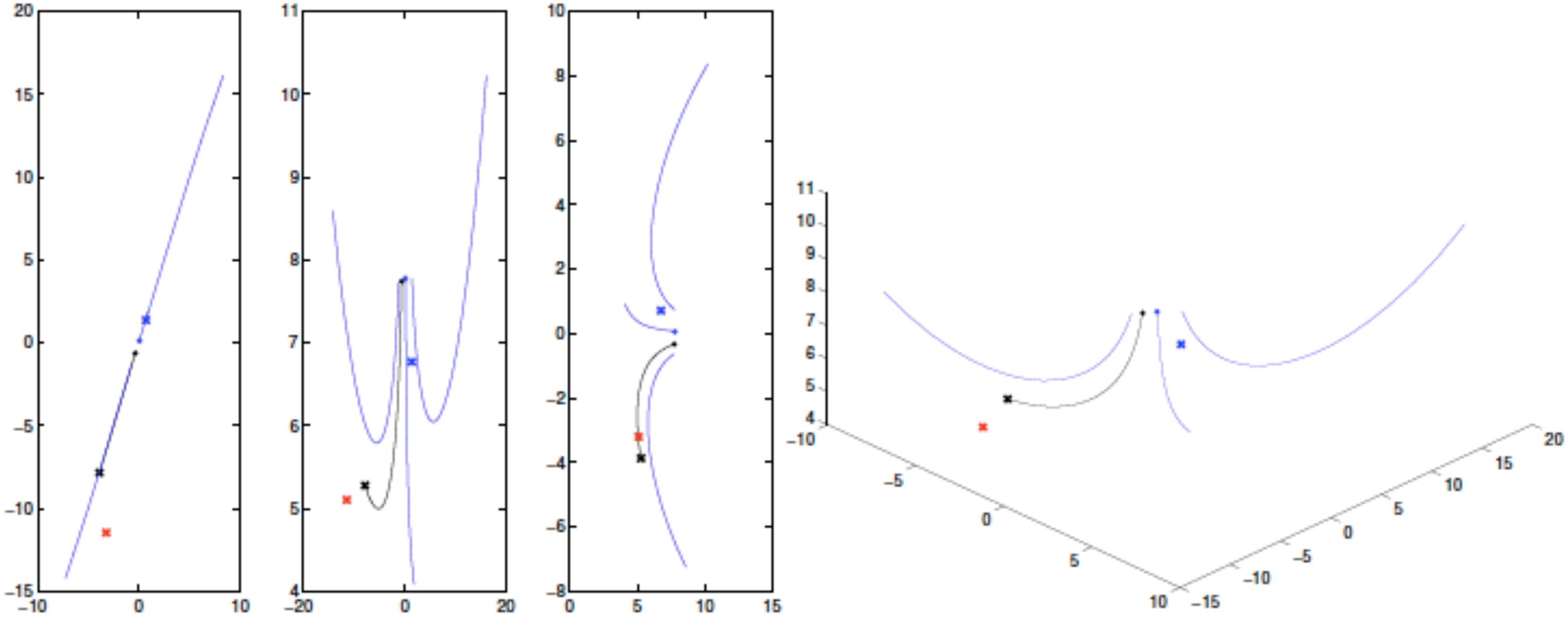
Example

Cycle 10 (4):



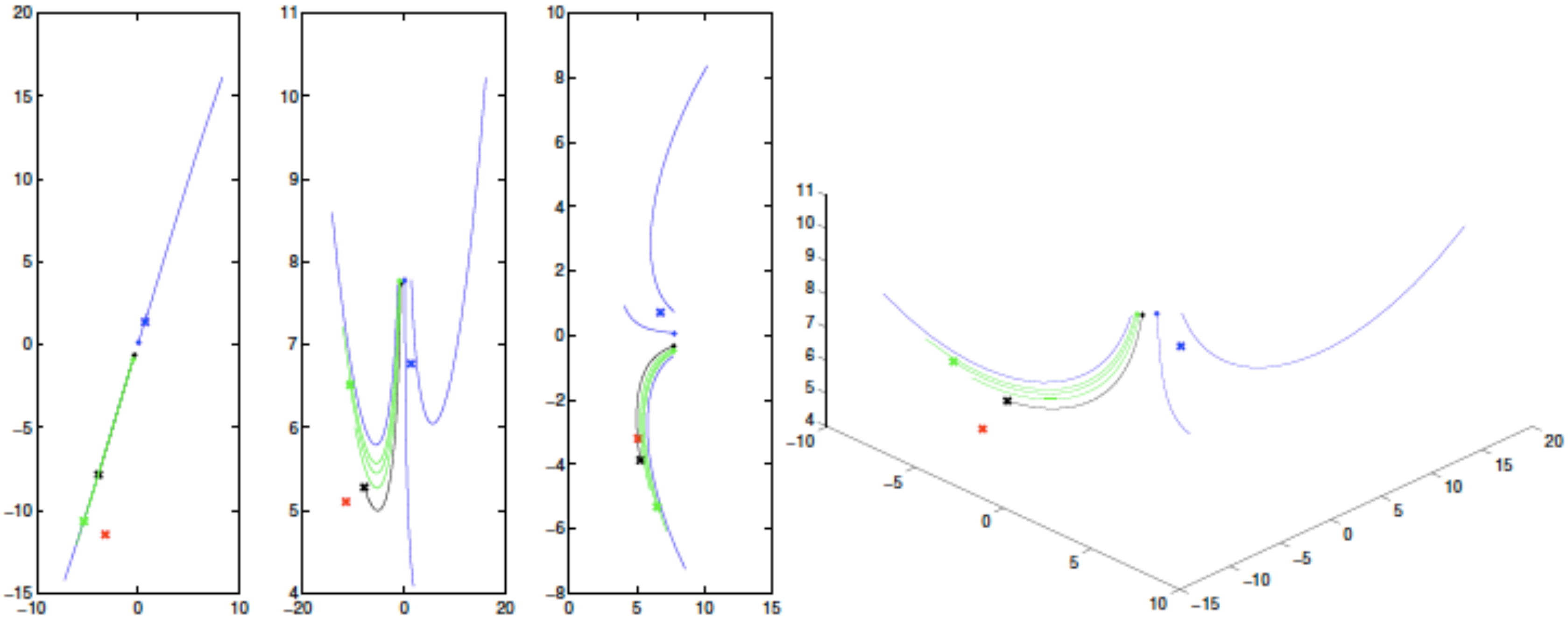
Example

Cycle 484 (1):



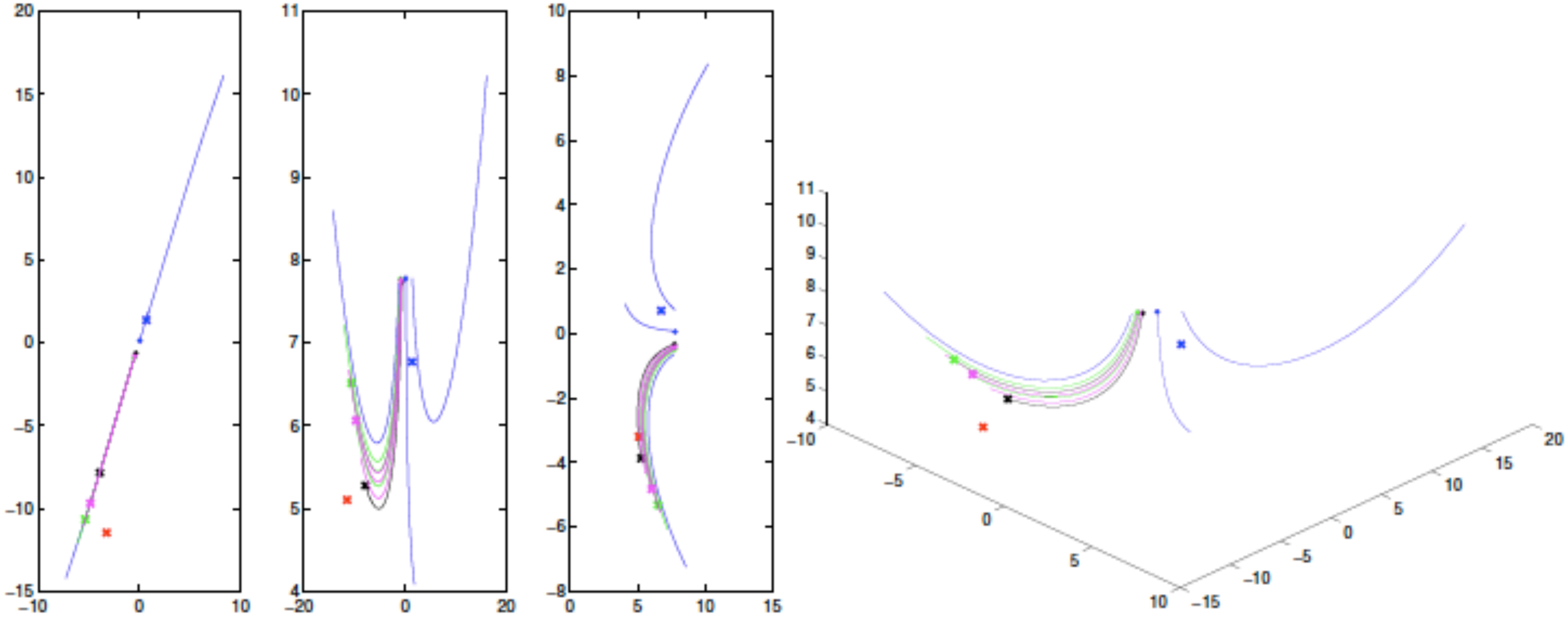
Example

Cycle 484 (2):



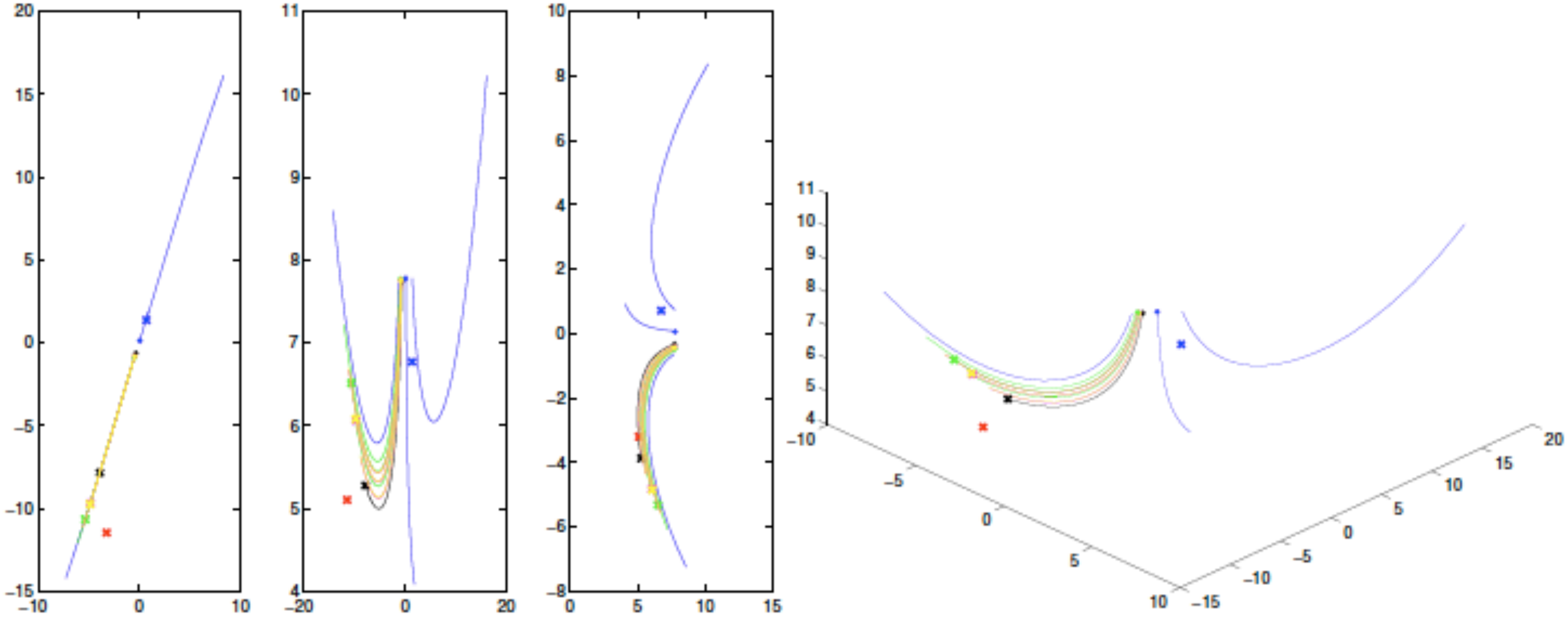
Example

Cycle 484 (3):



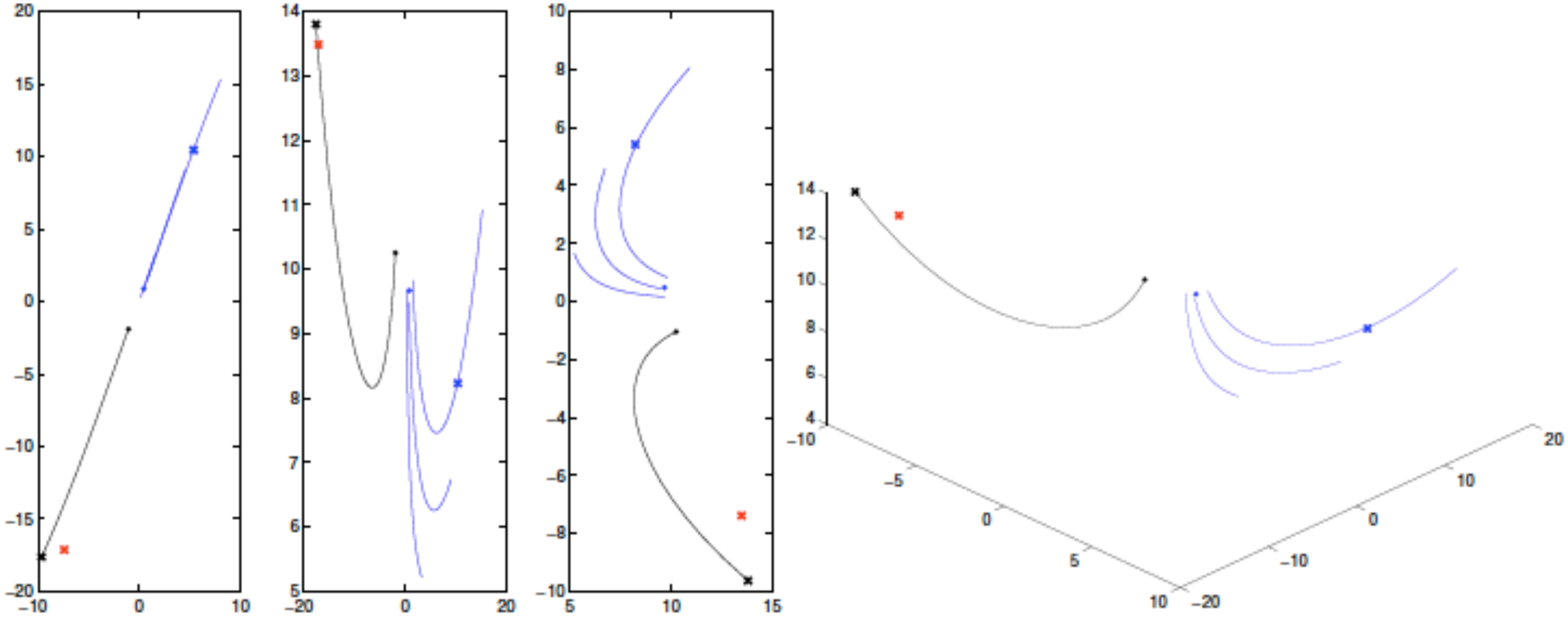
Example

Cycle 484 (4):



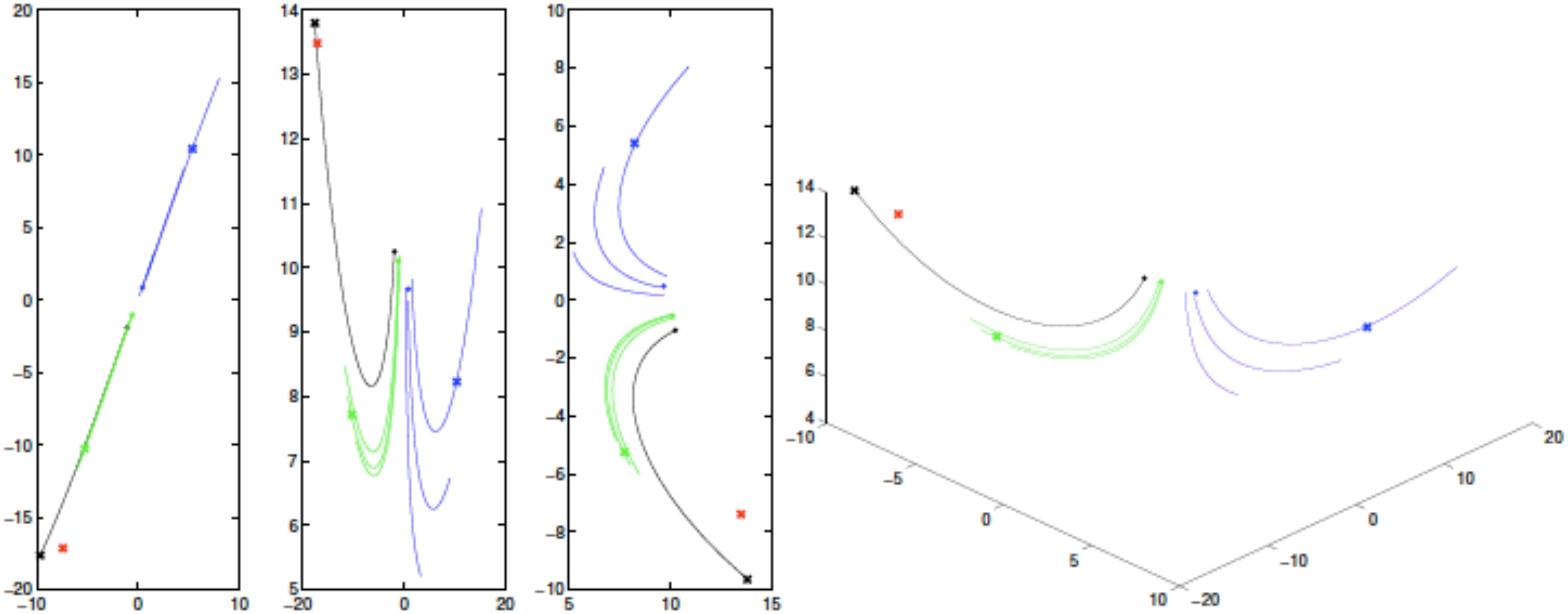
Example

Cycle 1392 (1):



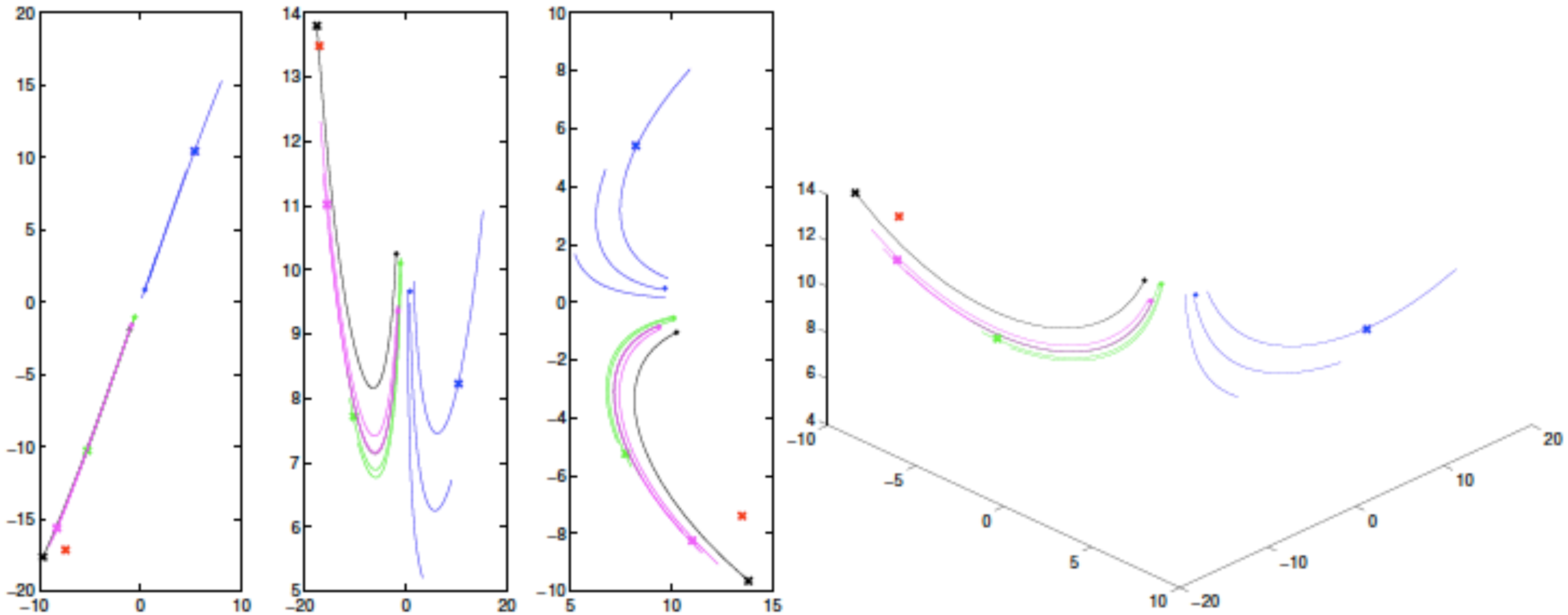
Example

Cycle 1392 (2):



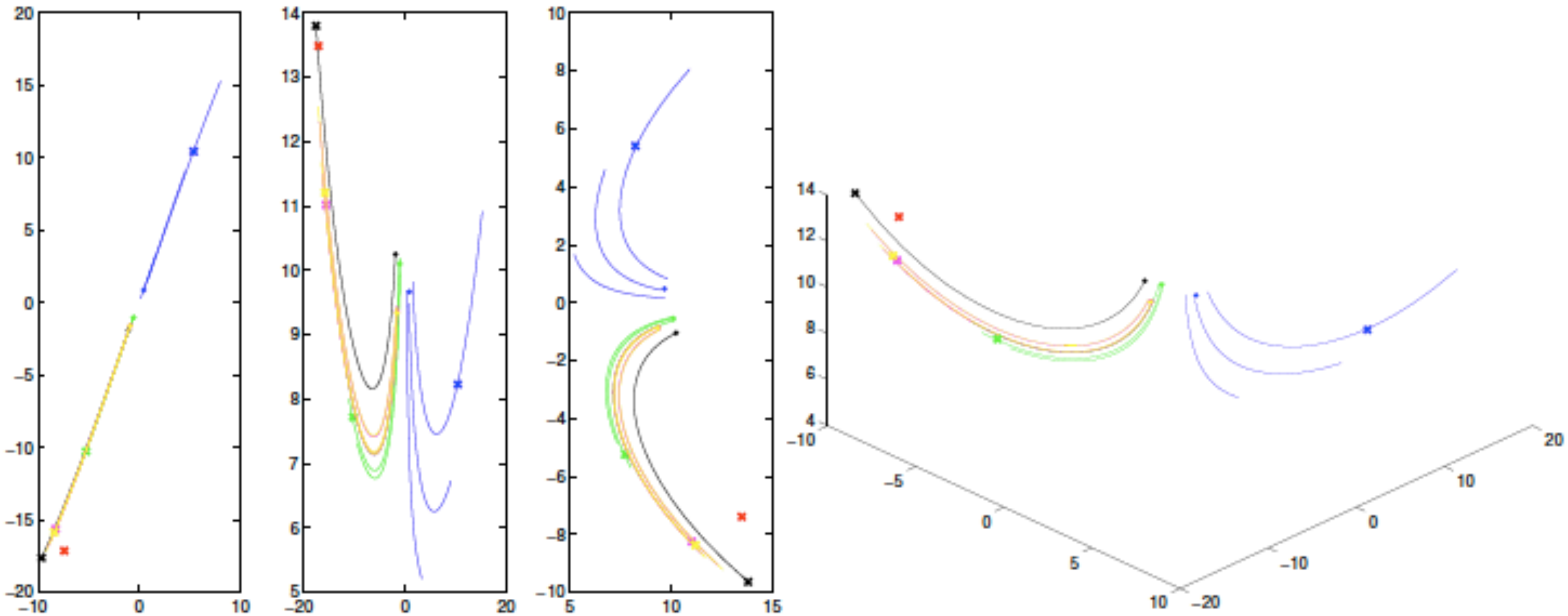
Example

Cycle 1392 (3):



Example

Cycle 1392 (4):



Experiment 1: L3, long window

(as in Miller et al. 1994)

scheme estimate	EnKF	IEnKF	IEKF

$m = 3$

analysis rmse	0.82	0.33	0.32
# iterations	1	2.8	2.7
inflation used	1.35	1.08	1.06

$m = 10$

analysis rmse	0.65	0.30	0.32
# iterations	1	2.6	2.7
inflation used	1.15	1.02	1.06

Kalnay et al. (2007):

- EnKF, $m = 3$: 0.71
- EnKF, $m = 6$: 0.59
- 4D-Var (assim. window = 75): 0.53

Yang et al. (2006):

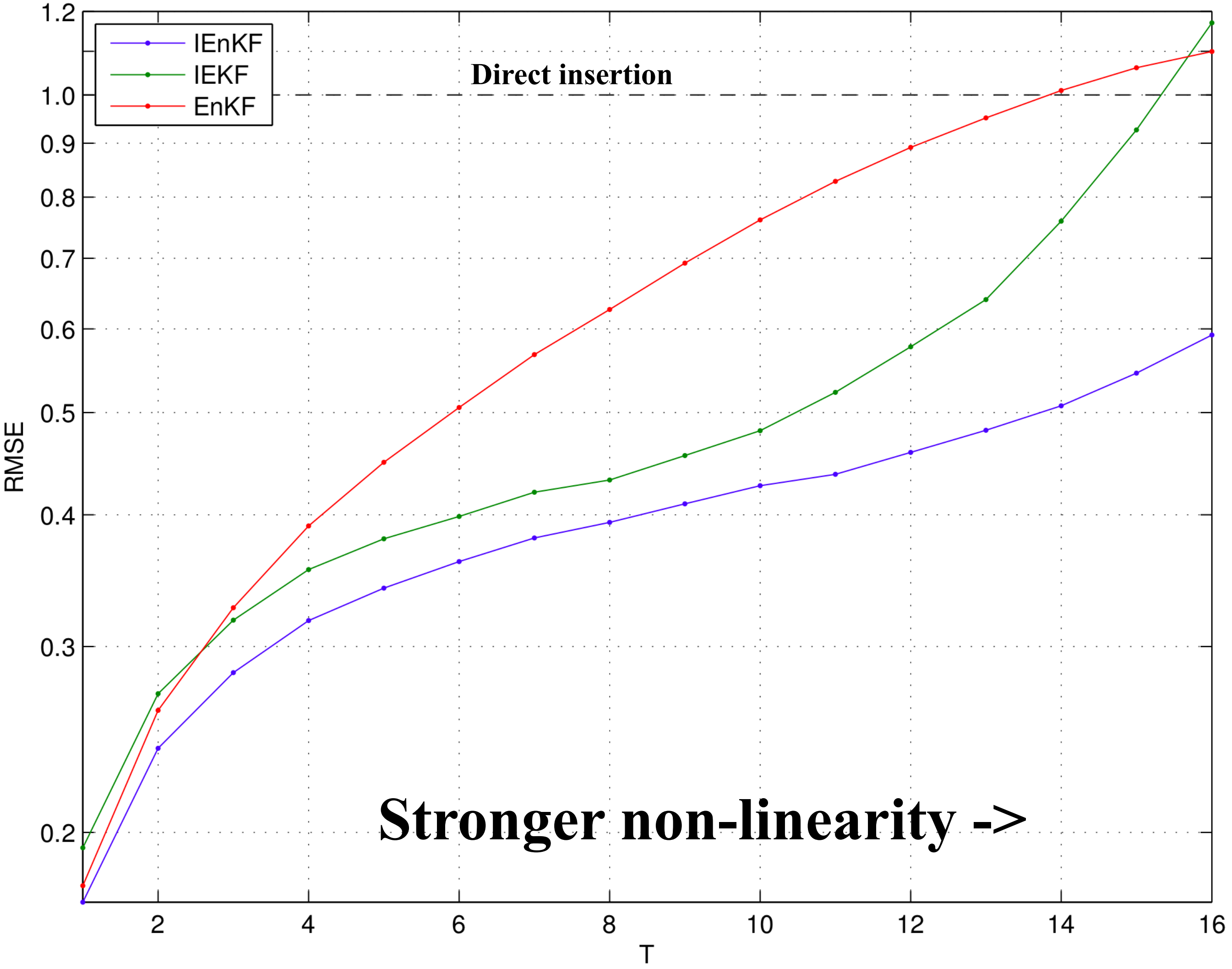
- EKF: 0.63

Kalnay and Yang,
http://www.mpipks-dresden.mpg.de/~ecodyc10/Contributions/Kalnay.pdf:

- RIP, $m =?$: 0.39

Benefits of Monte Carlo methods

RMS errors



- Sakov, Oliver & Bertino, subm. MWR

Conclusions

- The iterative EnKF is a natural generalisation of the EnKF for strongly nonlinear systems
- In the linear case the IEnKF is equivalent to the EnKF (the second and following iterations return zero increments)
- The IEnKF outperforms other schemes in the experiments
- The IEnKF should be considered in the case of a well constrained system with nonlinear growth during the assimilation cycle (the propagation is initially nonlinear but becomes linear through iterations)
- The above implies a high degree of optimality (including a perfect or nearly perfect model)
- Similarly to the EnKF, the IEnKF can be easily transformed into the ensemble formulation of the IEKF
- Future work: include model error; localisation

Relevant tasks in MyOcean2

- Task 19.3.4 **Assimilation of new types satellite observations**
- (a) Assimilation of sea-ice properties (NERSC, Mercator, METNO, DMI). A strategy will be derived for combined assimilation of ice concentrations and thickness as from CryoSAT. Different anamorphic transformations will be tested (static, variable in time or space), the observation operator for sea-ice concentrations will be tested as a choice for the observations. Diagnostics to be checked: scatterplots, representers. The multivariate impact on thickness, surface temperature and salinity will be monitored.

Google image "Sangoma"

We have nicer bugs!

Painless assimilation!