TOPAZ

a high-dimensional application of the EnKF to 3D ocean forecasting

L. Bertino, and the Mohn-Sverdrup Center

Nansen Center, Thormoehlensgt. 47, Bergen

NERSC

eVITA winter school, 30th Jan. 2007, Geilo

Mohn-Sverdrup Center

Global Ocean Studies - Operational Oceanography

Motivation

- Objective:
  - Provide short-term forecasts of physical and biogeochemical parameters targeted to users needs (as in weather forecasting)
    - Public users (Met and environmental agencies)
    - Industry (offshore oil & gas, ship routing)
- Strategy
  - Focus on advanced data assimilation techniques
  - Gradual increase of resolution (as affordable…)
  - Nesting on regions of higher interest

NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Outline

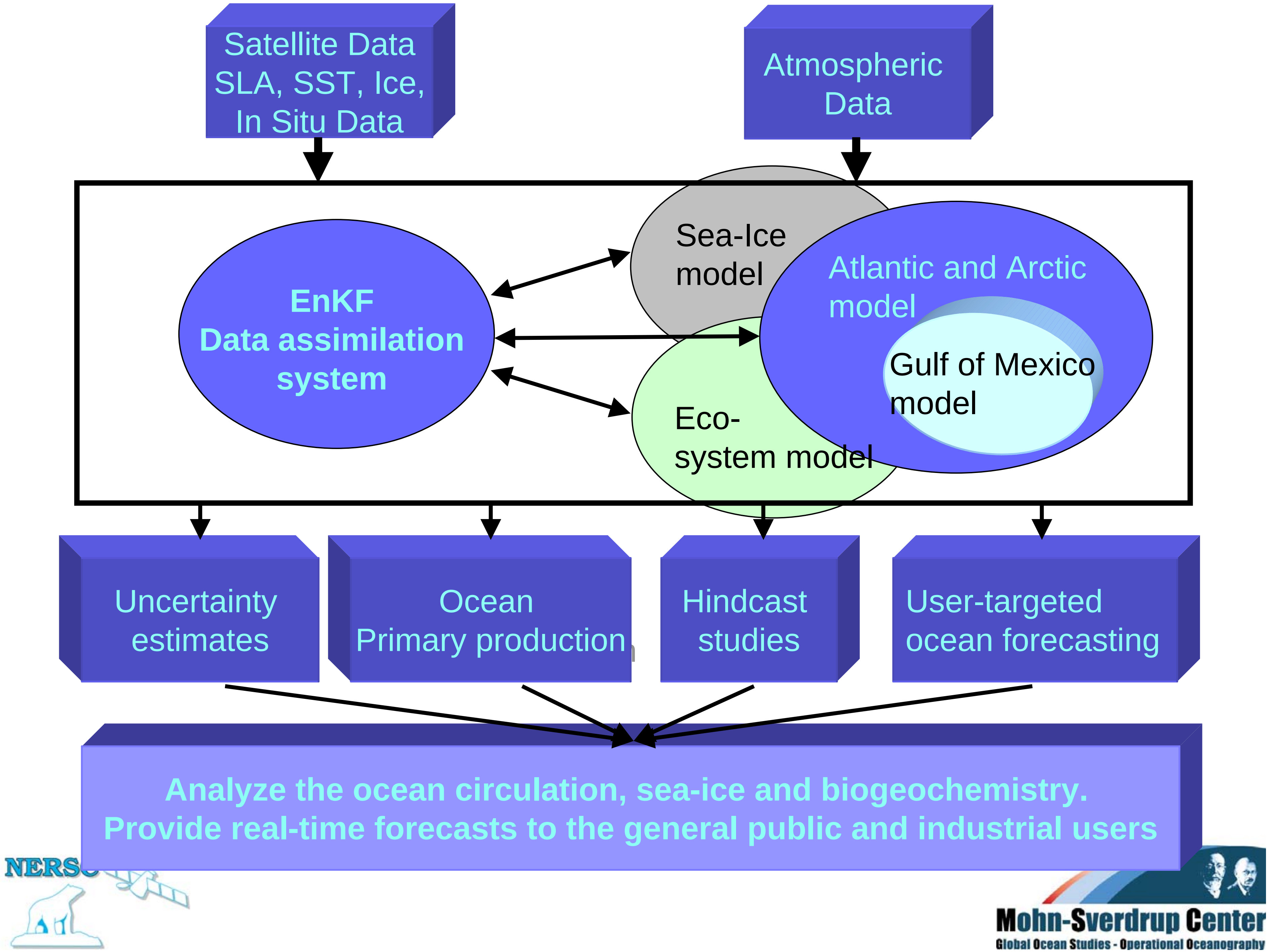
- Set up of the TOPAZ system
- Examples of ensemble statistics
- Problem dimensions
- Results and Applications
- Perspectives

System description

The TOPAZ system

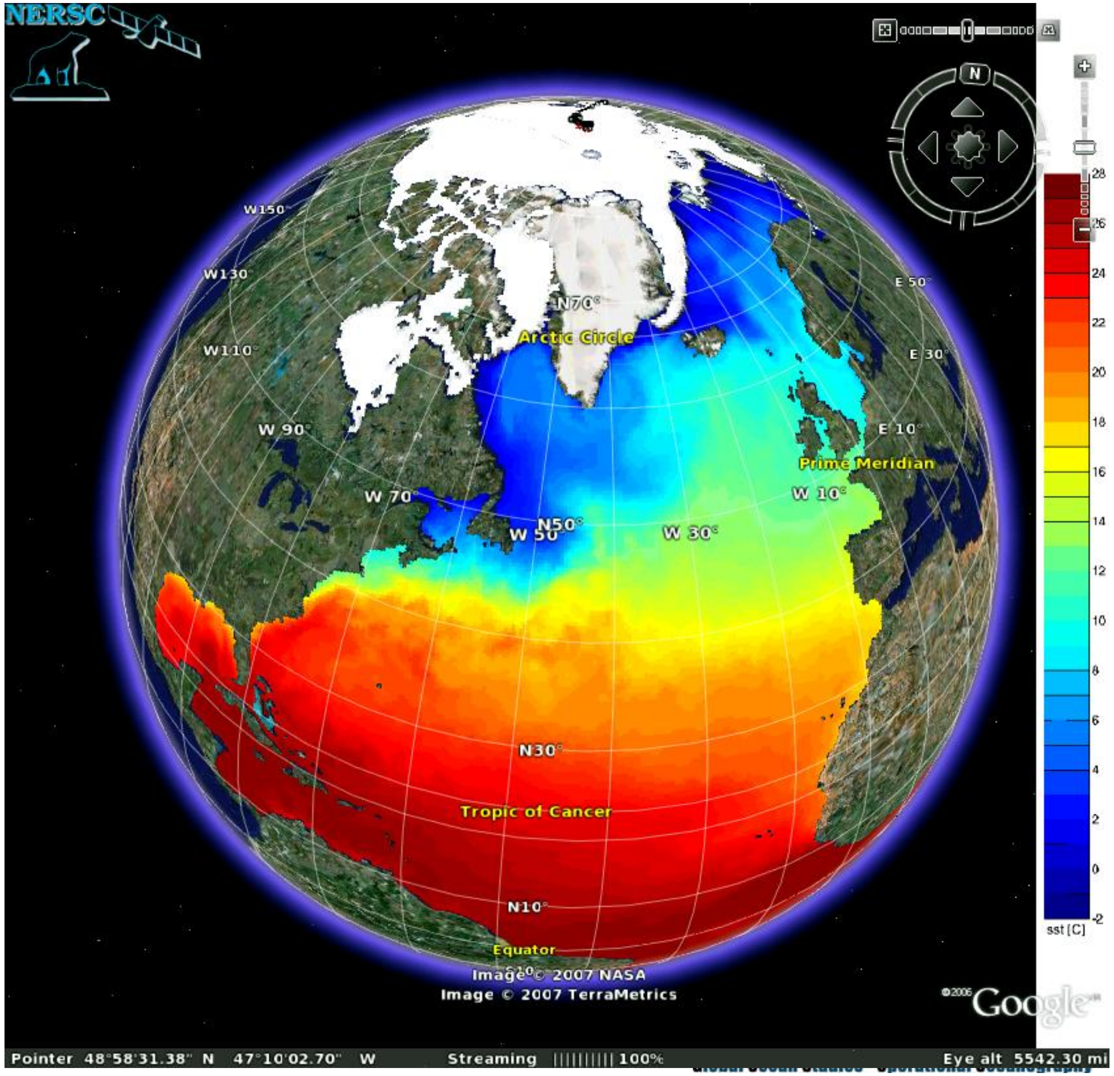
NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography


Satellite Data
SLA, SST, Ice,
In Situ Data
Atmospheric
Data
EnKF
Data assimilation
system
Sea-Ice
model
Atlantic and Arctic
model
Gulf of Mexico
model
Eco-
system model
Uncertainty
estimates
Ocean
Primary production
Hindcast
studies
User-targeted
ocean forecasting
Analyze the ocean circulation, sea-ice and biogeochemistry.
Provide real-time forecasts to the general public and industrial users
NERSC
Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

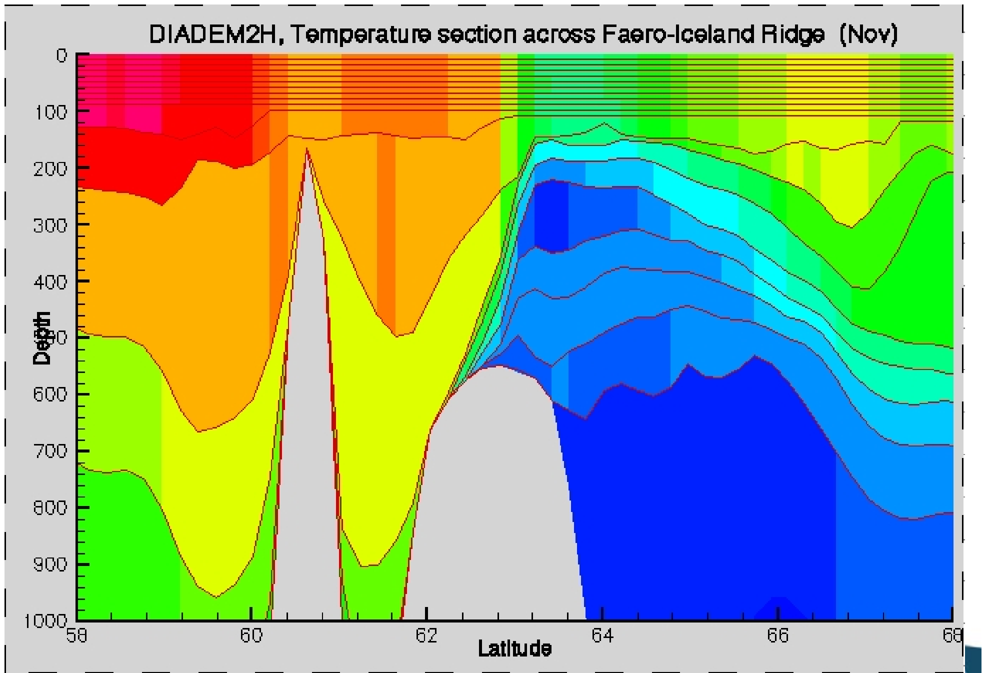
The TOPAZ model system

- Atlantic and Arctic domain
- Dynamic / thermodynamic ice model
- Weekly assimilation cycle
- Surface boundary conditions
  - ECMWF weather forecast



The ingredients

- 3D numerical ocean model
  - Hybrid Coordinate Ocean model, HYCOM (U. Miami)
  - 18-35 km resolution
  - 22 hybrid layers
- Observations
  - Altimetry, SST (CLS, F)
  - Sea Ice (NSIDC, USA)
  - In-situ (CORIOLIS, F)

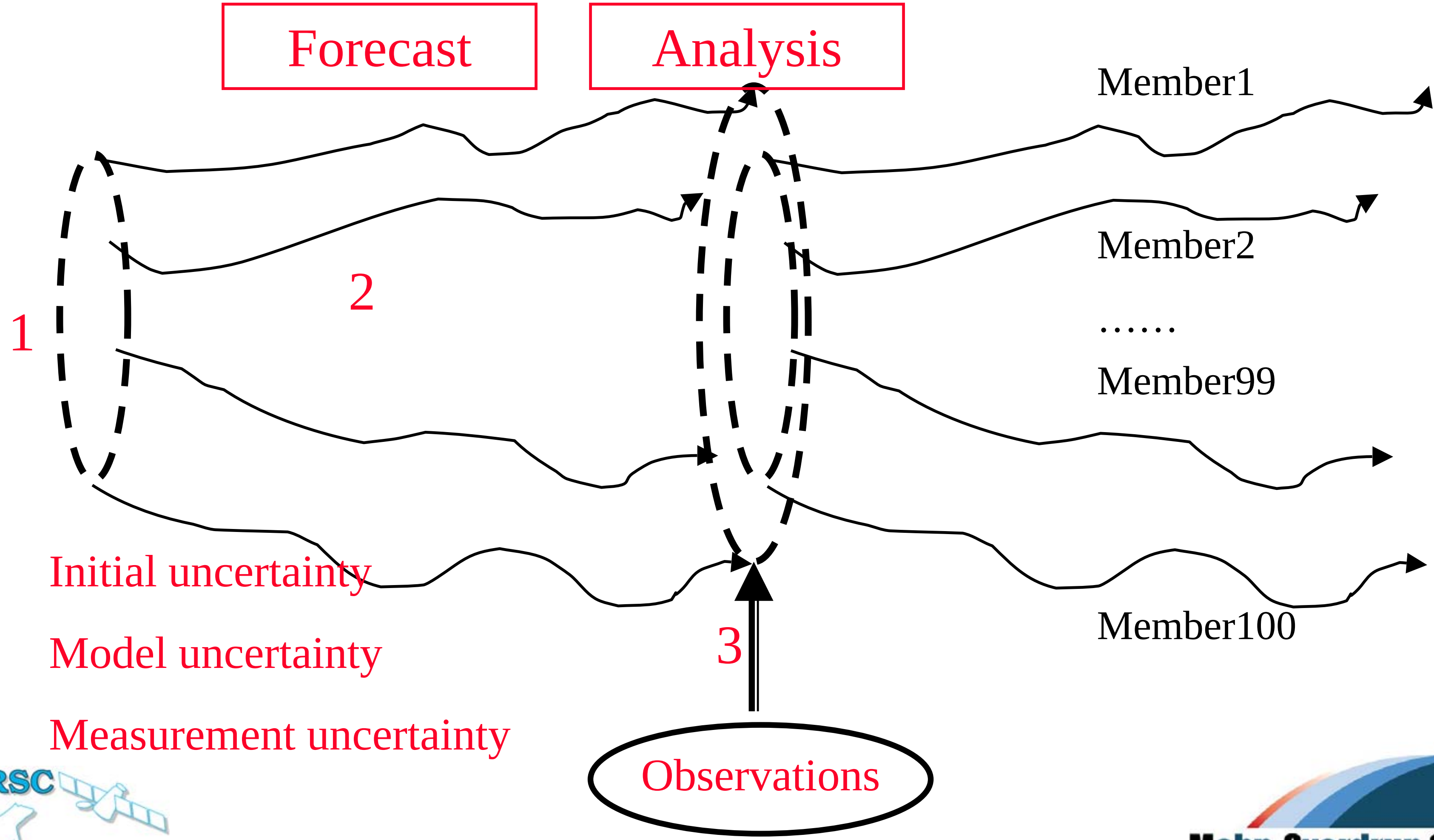


NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Ensemble Kalman filtering

a stochastic process



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

Our Priors

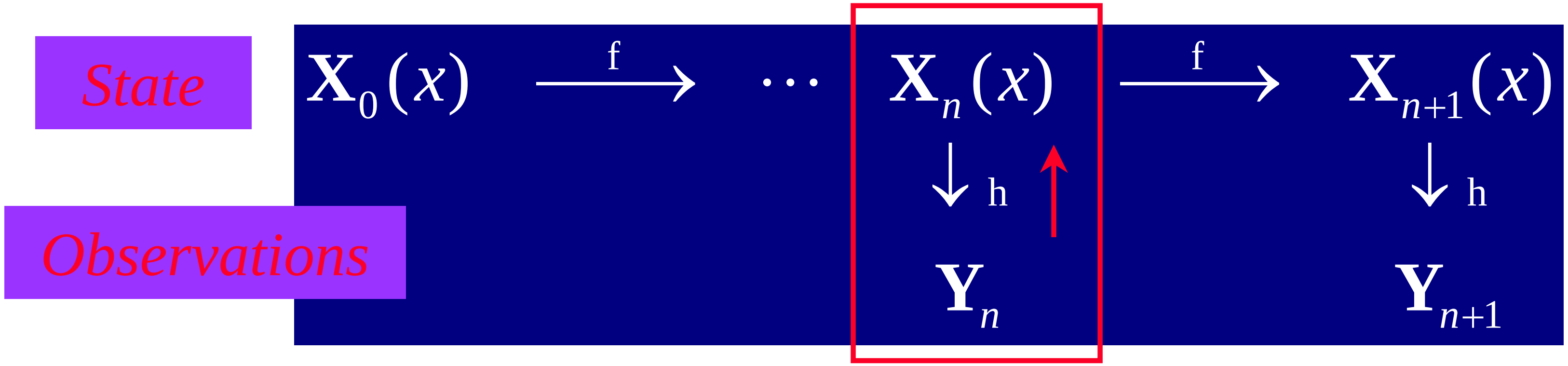
- Initial model error
  - Ocean stratification
  - 10% of the depth of each isopycnal layer
  - Lognormal *pdf*
- Boundary conditions
  - Random errors in
    - Wind speed and stress
    - Radiative heat fluxes
    - Air temperature
  - Gaussian *pdf*
  - Given standard deviation
  - Horizontal radius 250 km

- Measurement errors
  - Gaussian *pdf*
    - Sea surface heights
    - Sea surface temperatures
  - Truncated Gaussian
    - Ice concentrations
  - Given standard deviations
  - Horizontal radius 250 km

NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

The Ensemble Kalman filter



- Assuming
  - Gaussian model state variables
  - Gaussian observation variables
  - Unbiased model and observations
- The EnKF applies the least square estimation

A parenthesis on geostatistics

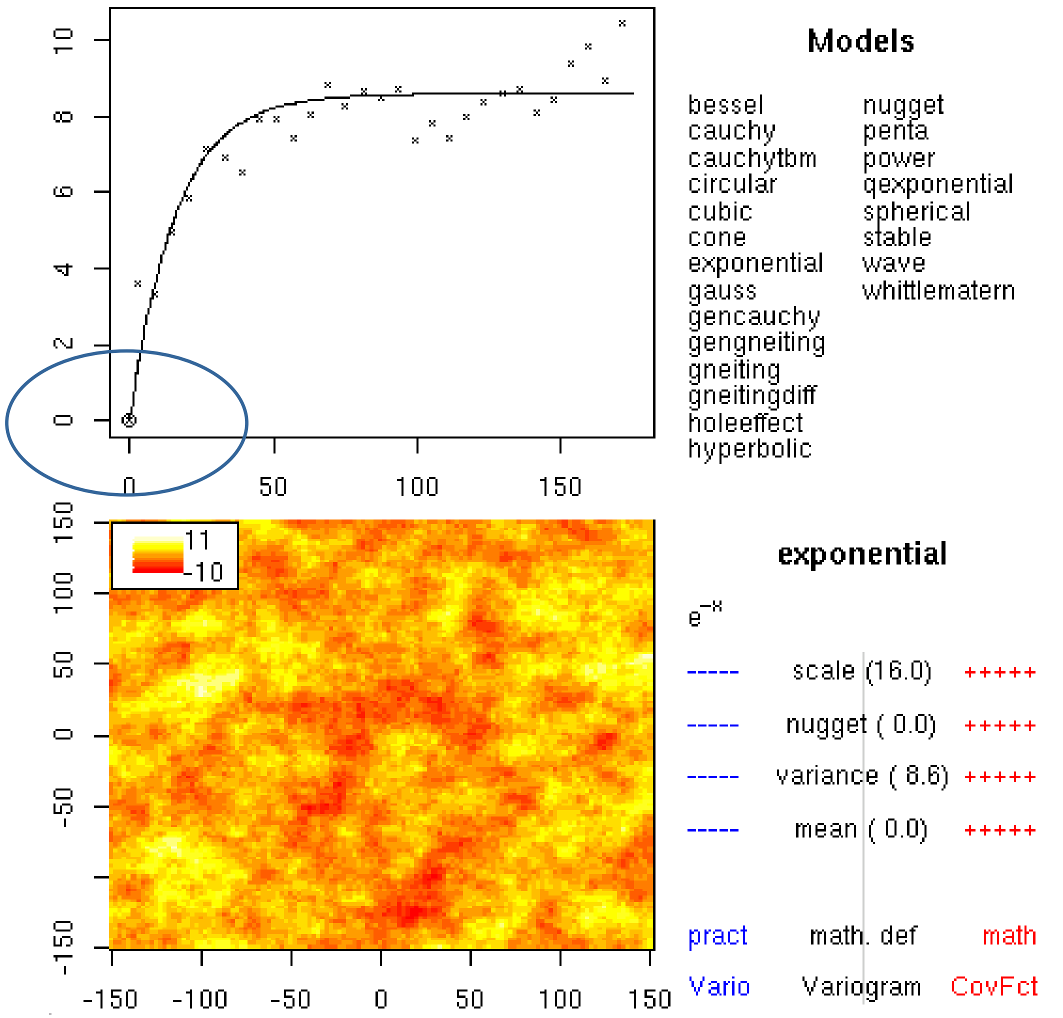
- Kriging (linear least square estimation) depends heavily on the error covariance:
  - Its spatial scale (decorrelation)
  - The Spatial Structure of the covariance
  - In particular its behavior at the origin
- Let us see a simple static example
- (Both under Gaussian *distributions*)

Exponential covariance

- Horizontal scale 50
- Continuous at origin
- Slope at origin
- Field is “rough”.

RandomFields library in R

by M. Schlather, U. Goettingen



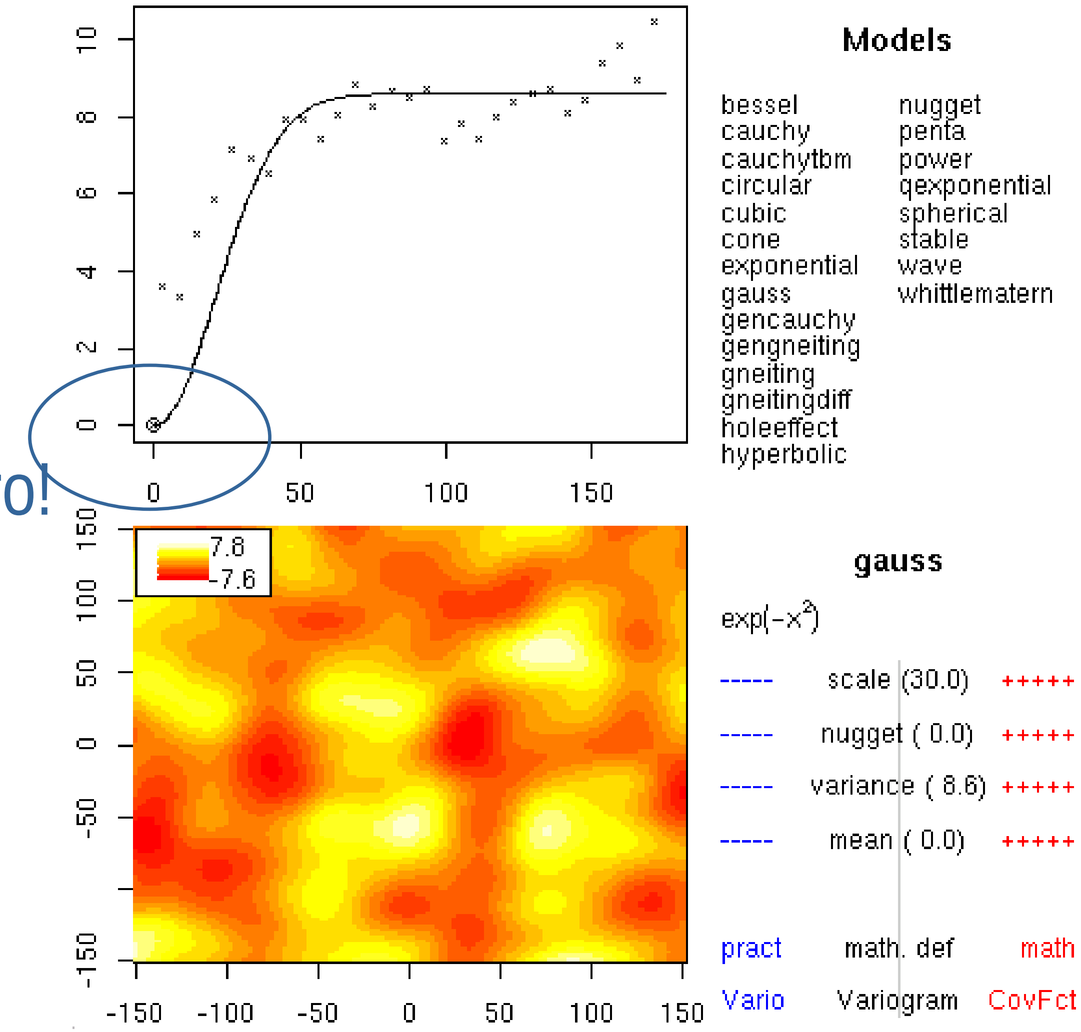
Mohn-Sverdrup Center

Global Ocean Studies - Operational Oceanography

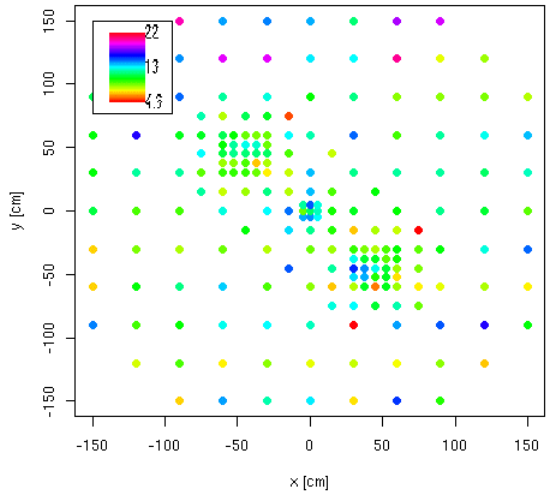
NERSC

Gaussian covariance

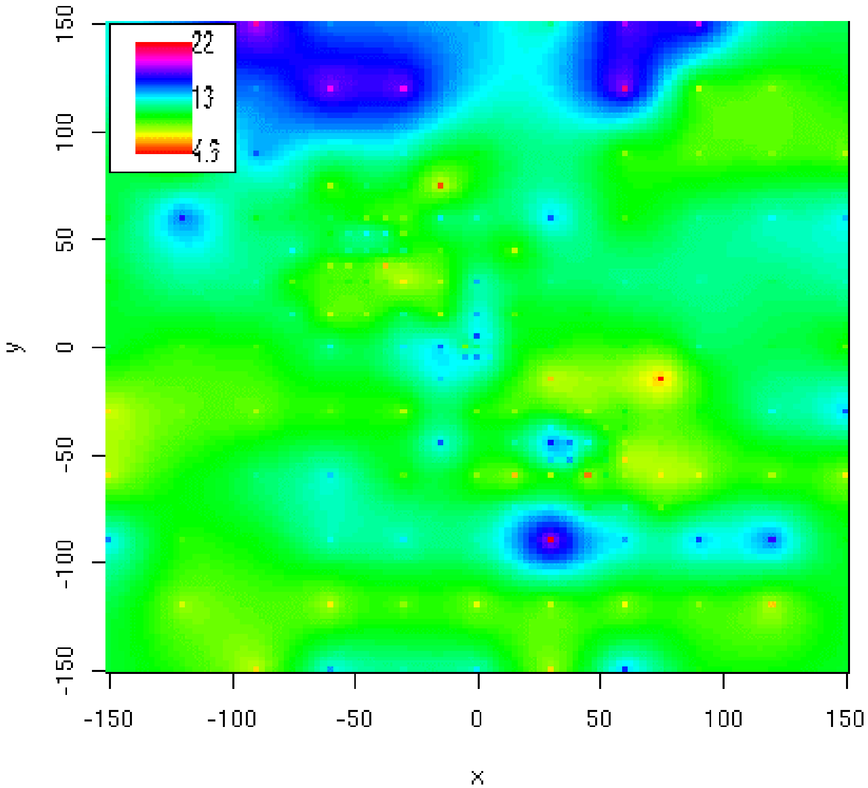
- Horizontal scale 50
- Continuous at origin
- Zero slope at origin
- All derivatives are zero!
- Field is “smooth”



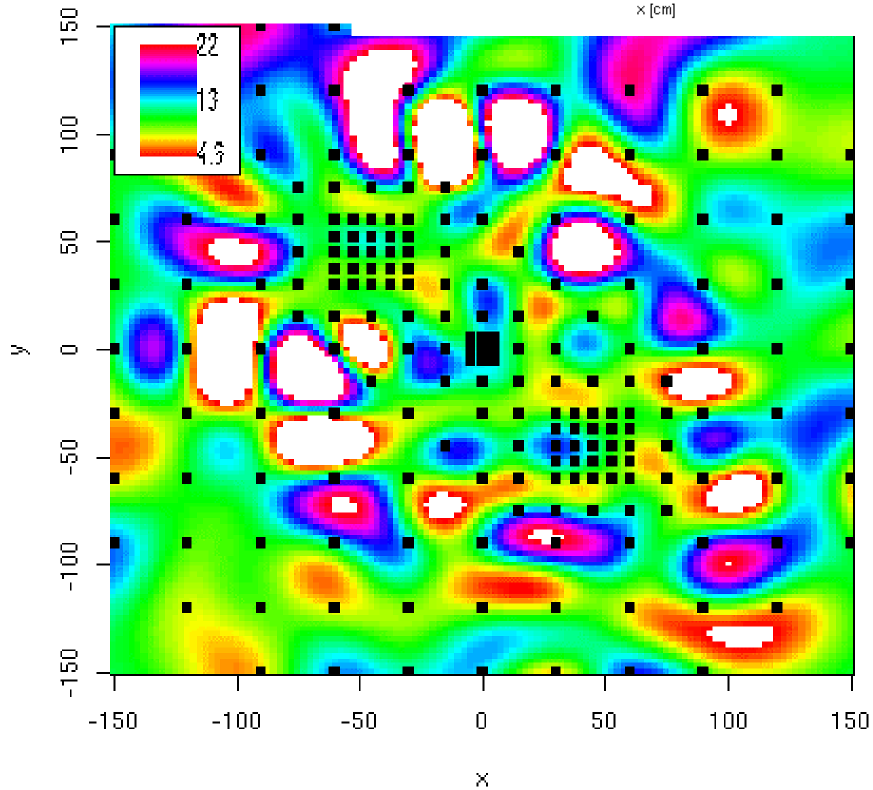
Kriging maps



- Exponential



- Gaussian



The Gaussian covariance makes overshoots!

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Ensemble Statistics

From the TOPAZ system

Ocean dynamics

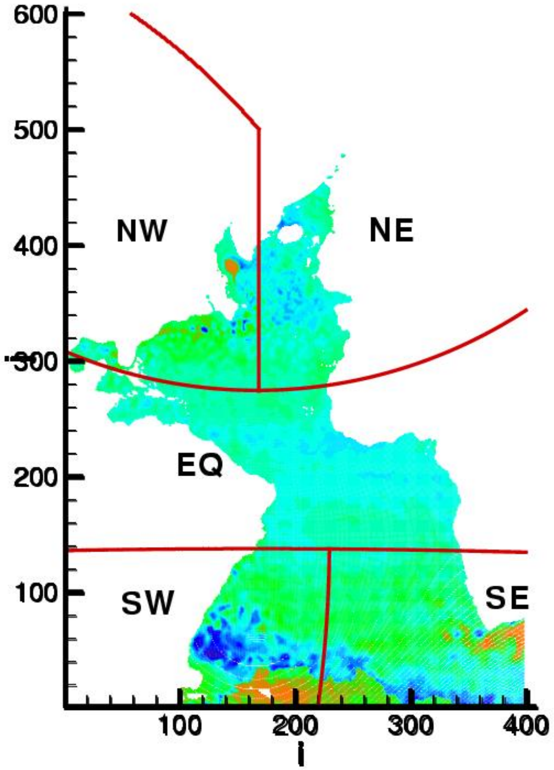
- Statistical properties such as
  - Spatial range
  - Variance
  - Multivariate Cross-covariance
- .. evolve according to the ocean dynamics
  - in space
  - in time
- Monte-Carlo methods provide an “ensemble approximation” to all instantaneous statistics

NERSC

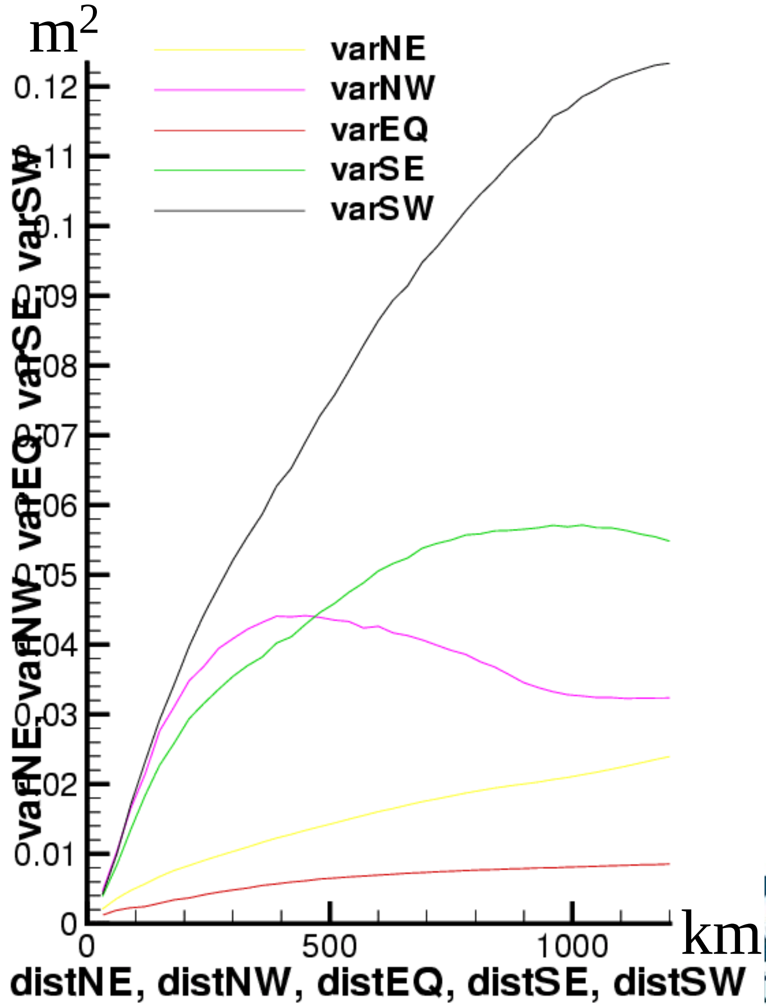
Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Ensemble Covariance

Spatially varying structures

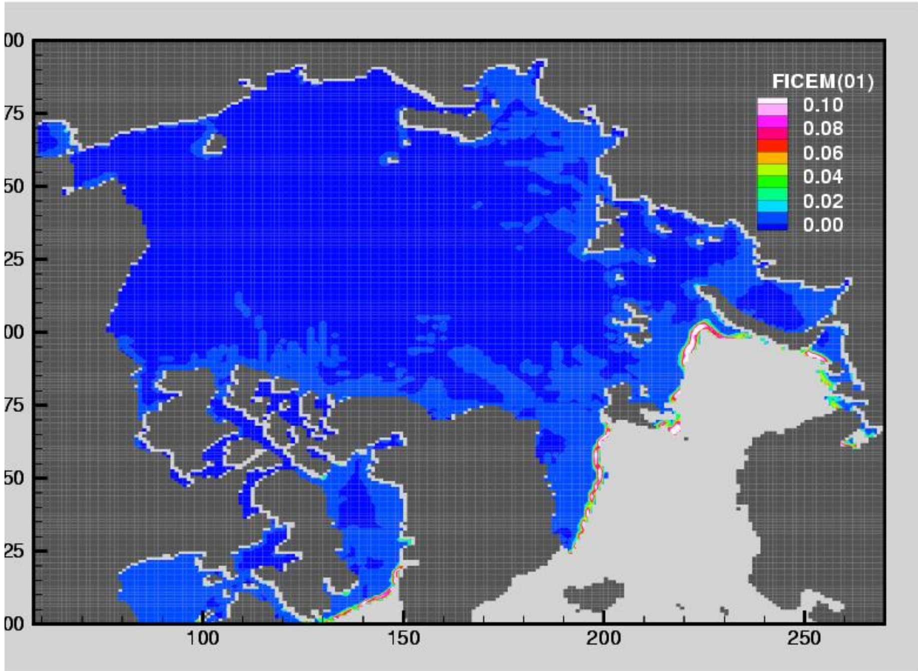


SLA innovations: horizontal variograms

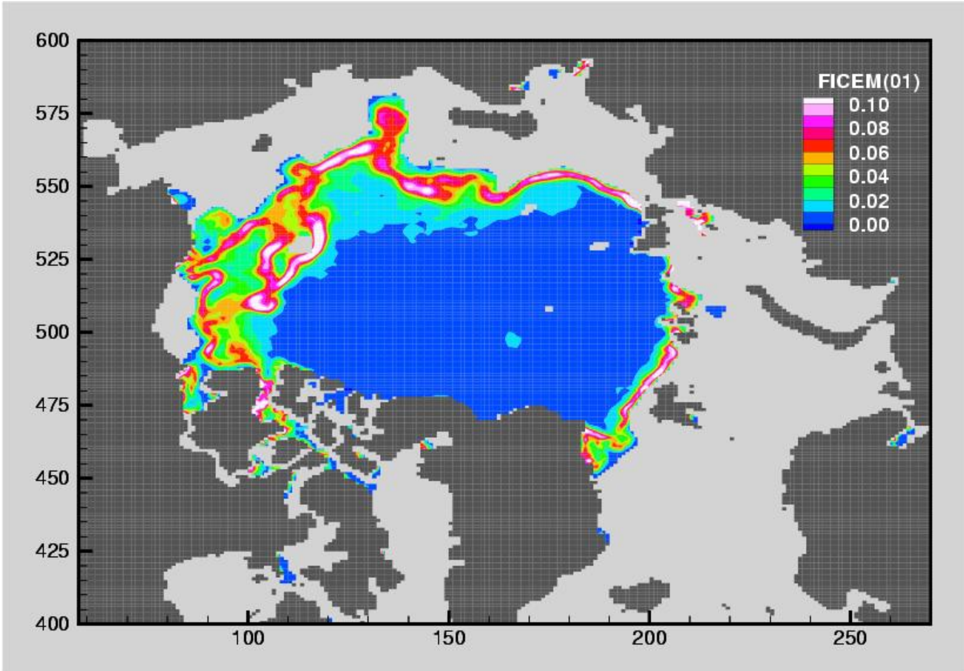


Ensemble Variances
Temporal evolution
(variance of ice concentrations)

- 1st March 2006



- 13th Sept 2006



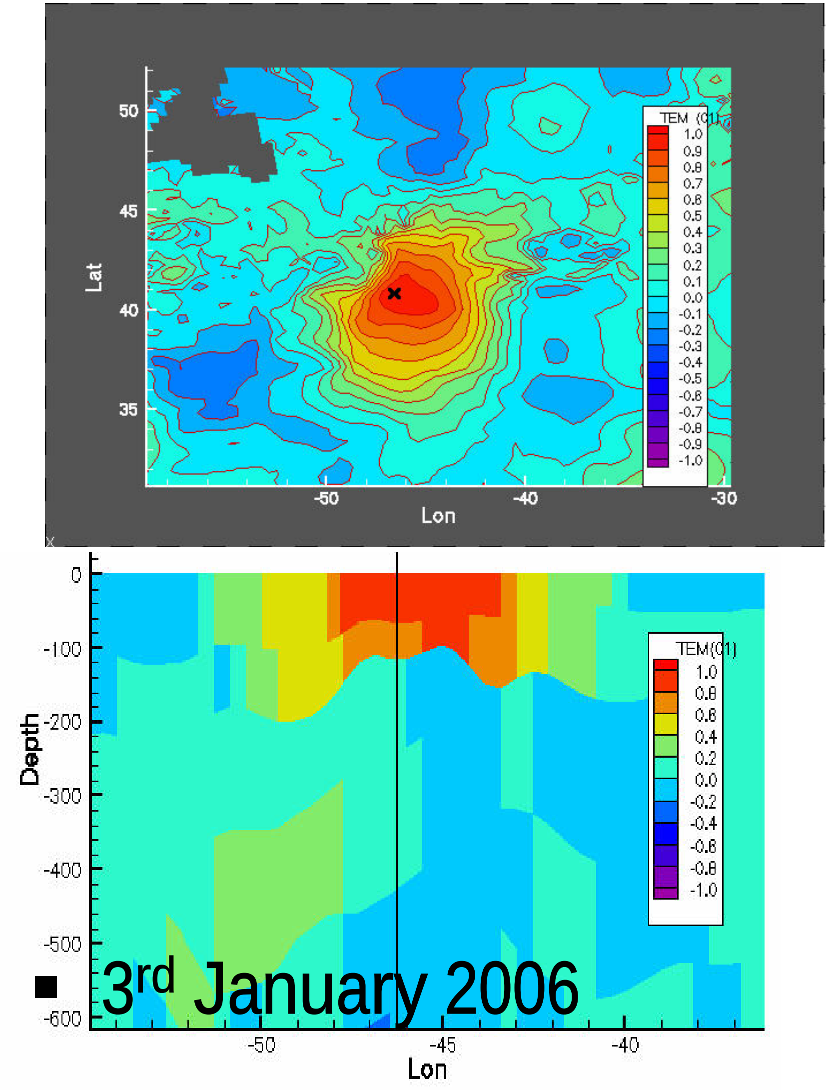
NERSC

http://topaz.nersc.no

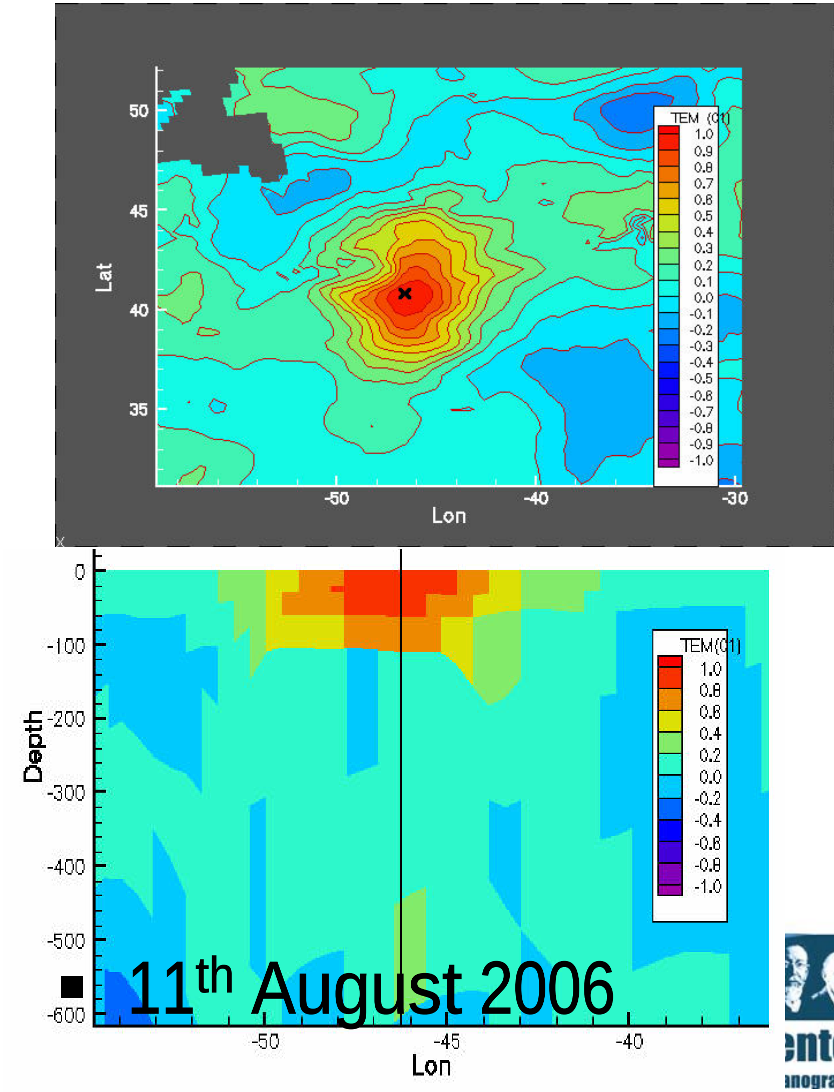
Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Ensemble Covariances

Temporal evolution



■ 3rd January 2006

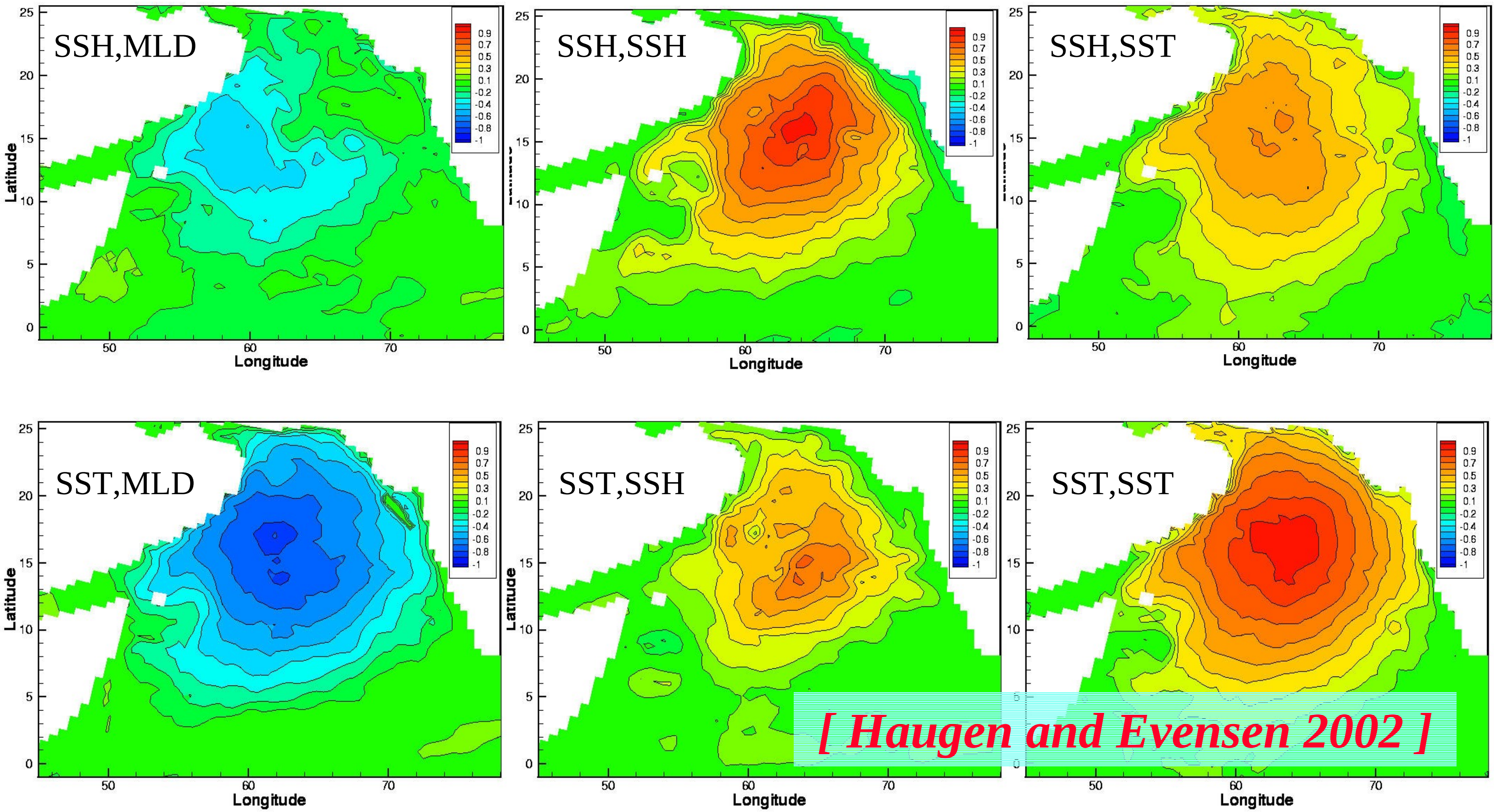


■ 11th August 2006

Ensemble Covariance

Multivariate structures

Sea surface height & temperature, mix layer depth



Problem dimensions

High dimensionality

The State Space

- 2D variables (400 x 600 grid cells)
  - Barotropic pressure, u/v velocity, ice concentration, ice thickness
- 3D variables (400 x 600 x 22 grid cells)
  - Temperature, salinity, u/v current, layer thickness
- TOTAL: n = 27.600.000 state variables
- 100 members in double precision = 21 Gb
- *Next prototype (Due April 2007):*
  - *81 million variables, 60 Gb!*
- *Ecosystem variables: 2 to 3 times more variable depending on ecosystem model formulation*

NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

The observations

- Sea level anomalies – SLA (satellite, radar altimeters)
  - Non linear function of state variables
  - 100.000 observations every week
- Sea-surface temperature – SST (satellite, optical)
  - 8.000 observations every week
- Sea-ice concentrations (satellite, microwave)
  - 40.000 observations every week
- TOTAL: **m=148.000 obs**
- *Coming up:* in-situ *profiles (~500.000 obs.), HR SST (120.000 obs.), HR ice conc. (160.000 obs.) ice drift …*

NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Local analysis

- For each water column (x, y), update with local observations only
  - Local state space n = 115 variables (5x22+5)
  - Local observations m = 49 nearest (within 700km max)
  - Ensemble size N = 100 (as usual)

☺ *N, m, n* are reasonably similar, small matrices

☺ The local analysis loop is *embarrassingly* parallel

☹ The analysis is not necessarily continuous

☹ X_5 is varying with location (x, y).

Computations

- Propagation
- 1000 CPU hours / week
- Embarrassingly parallel
- 100x 4 CPU 3hours jobs
- Each job requires 3 Gb
- Interactive submitting
- Completed within 3 days

- Analysis
- 6 CPU hours / week
- Sequential, 3 datasets
- 3x 4 CPU 40 min jobs
- Each job requires 25 Gb
- MPI parallelization required for clusters

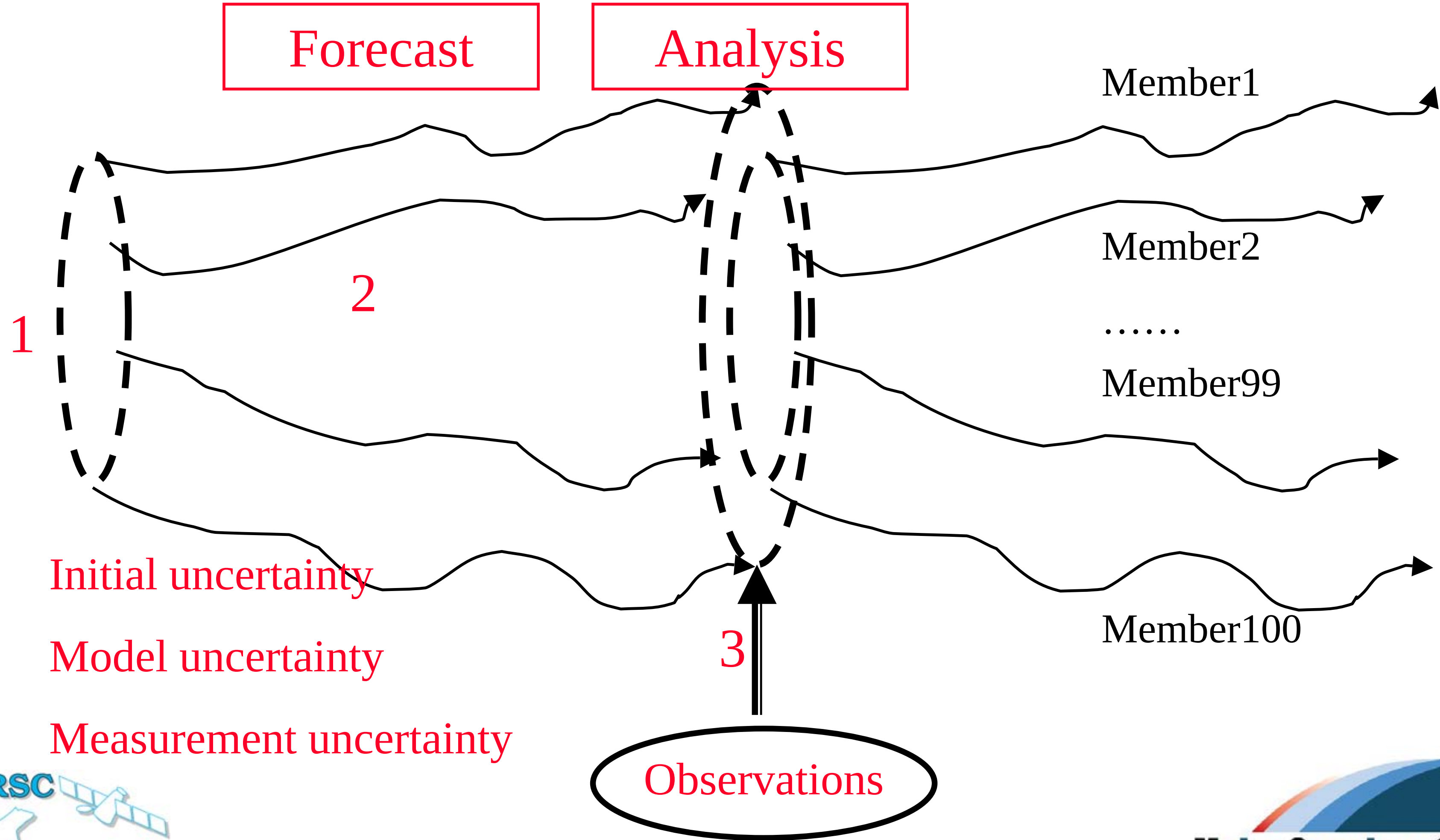
NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Results

Ensemble Kalman filtering

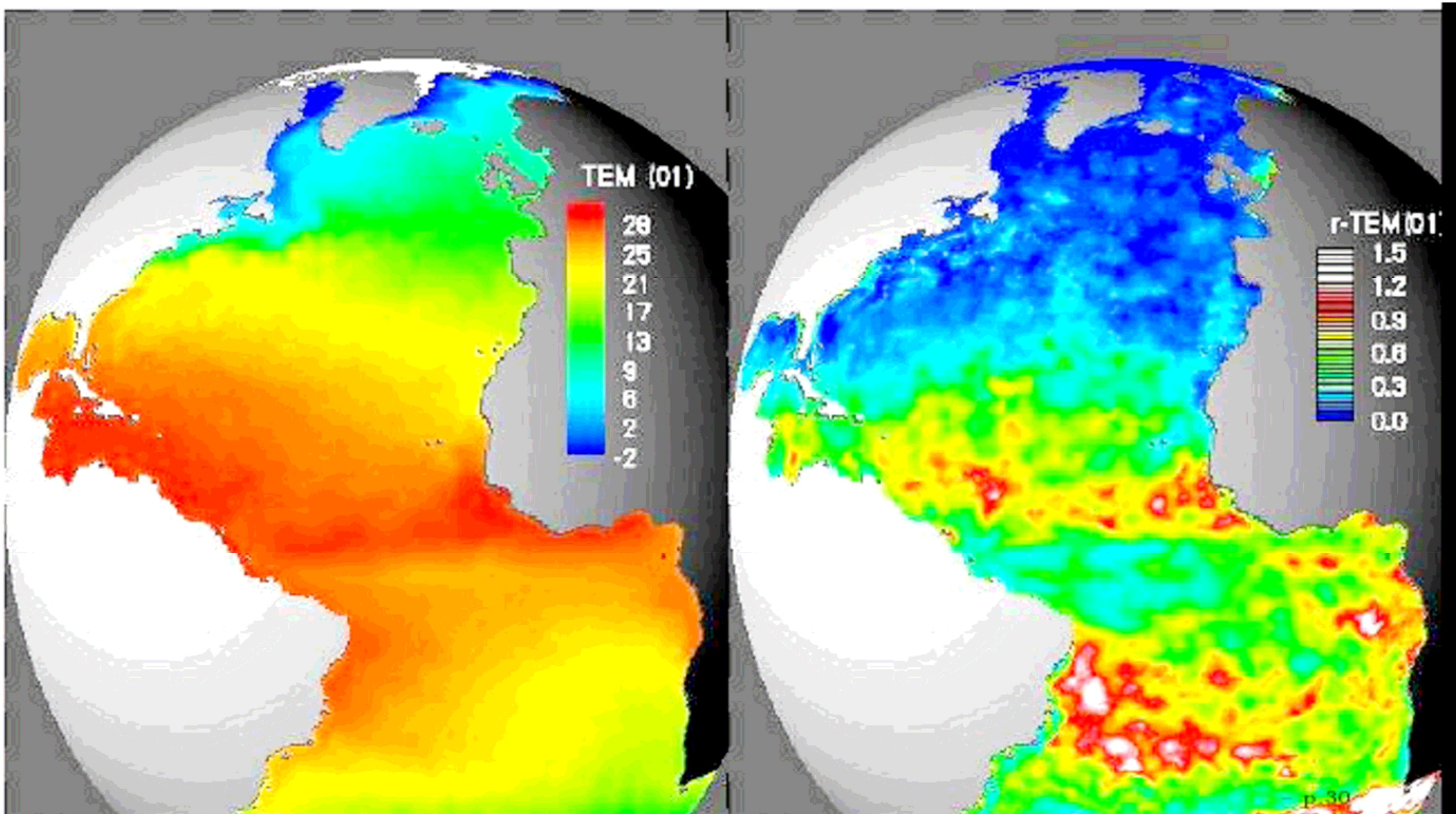
a stochastic process



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

Errors depend on observations density

December 2003 SST before analysis

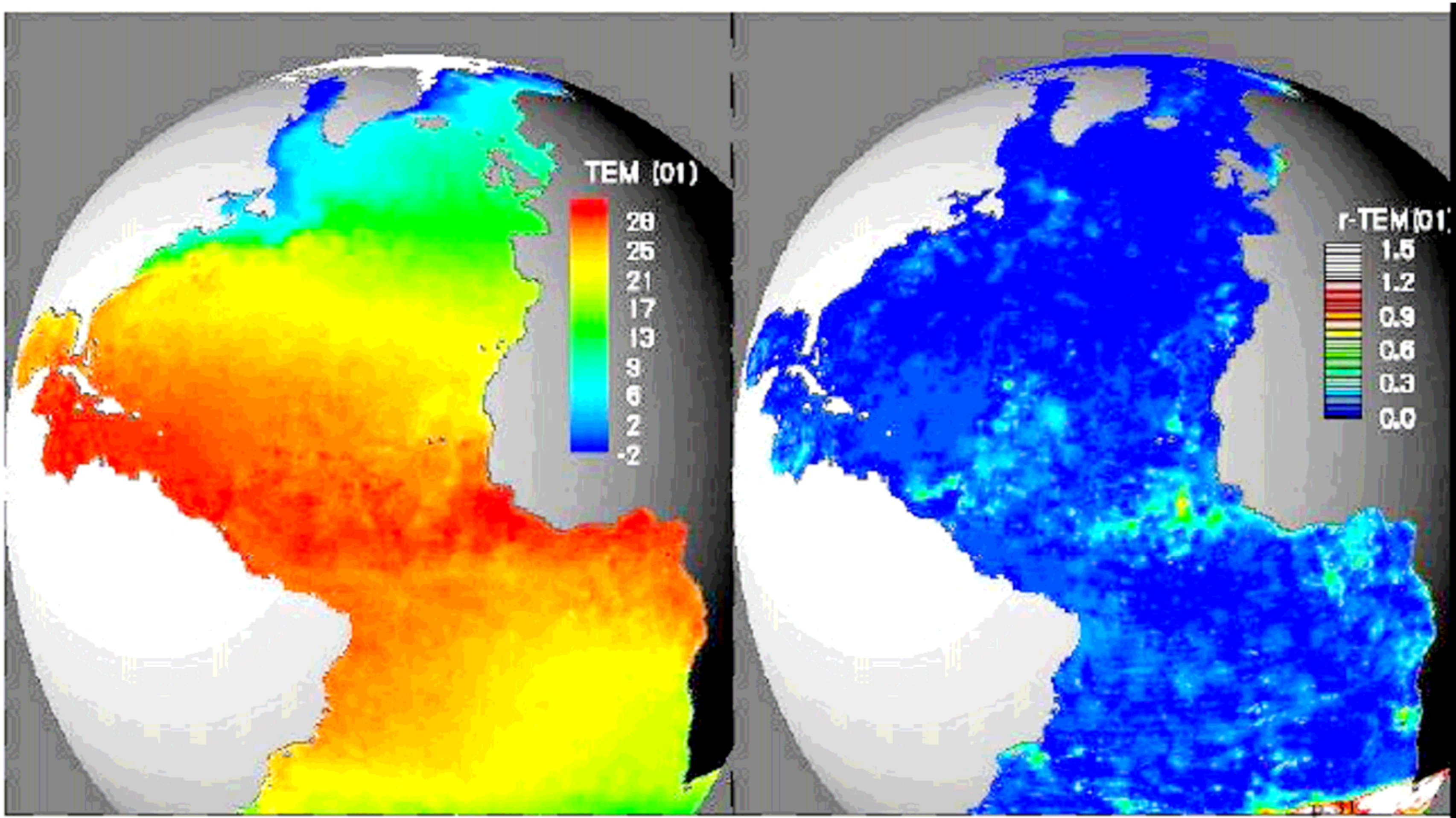


Mean SST

Variance

Errors depend on observations density

Decmber 2003 SST after analysis



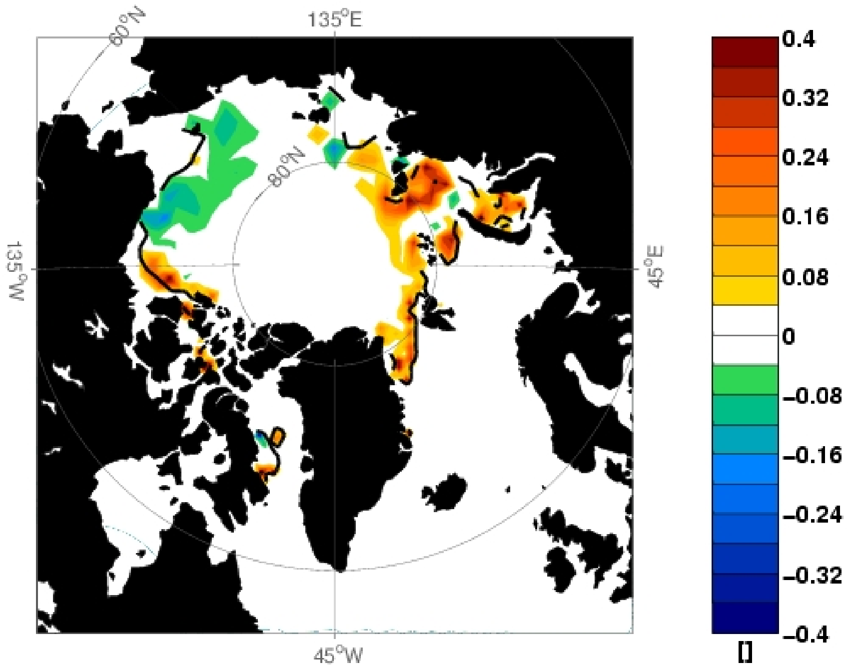
Mean SST

Variance

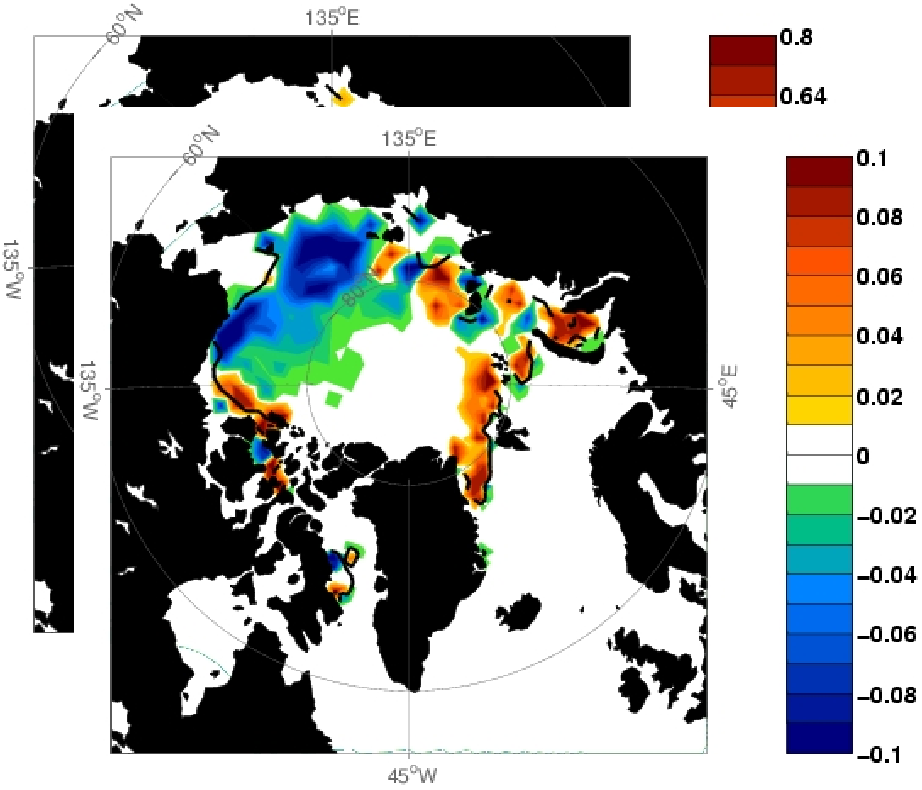
Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Multivariate Assimilation

update - Summer



Ice concentration update



Surface salinity update

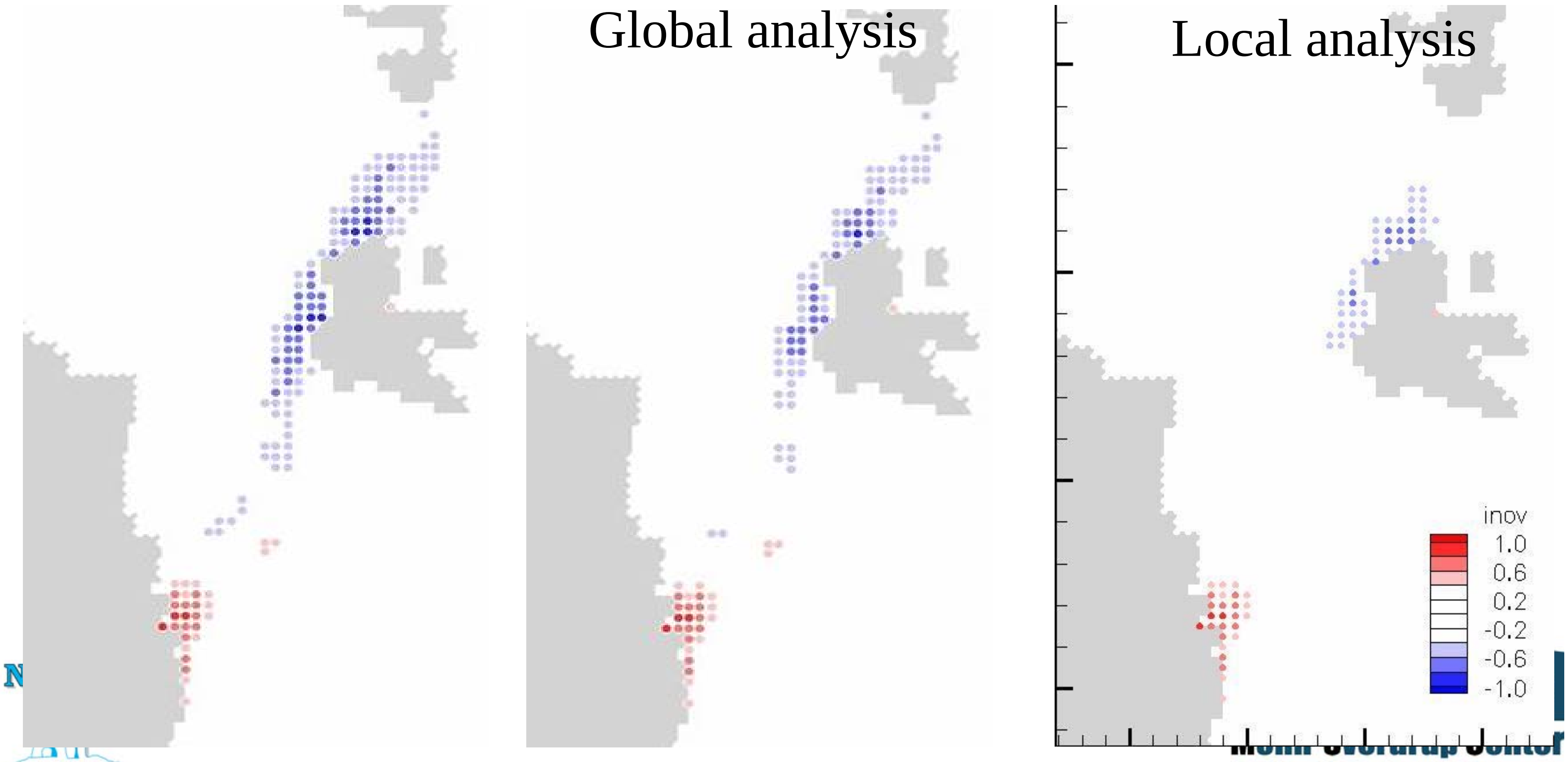
[Lisæter et al. 2003]

NERSC

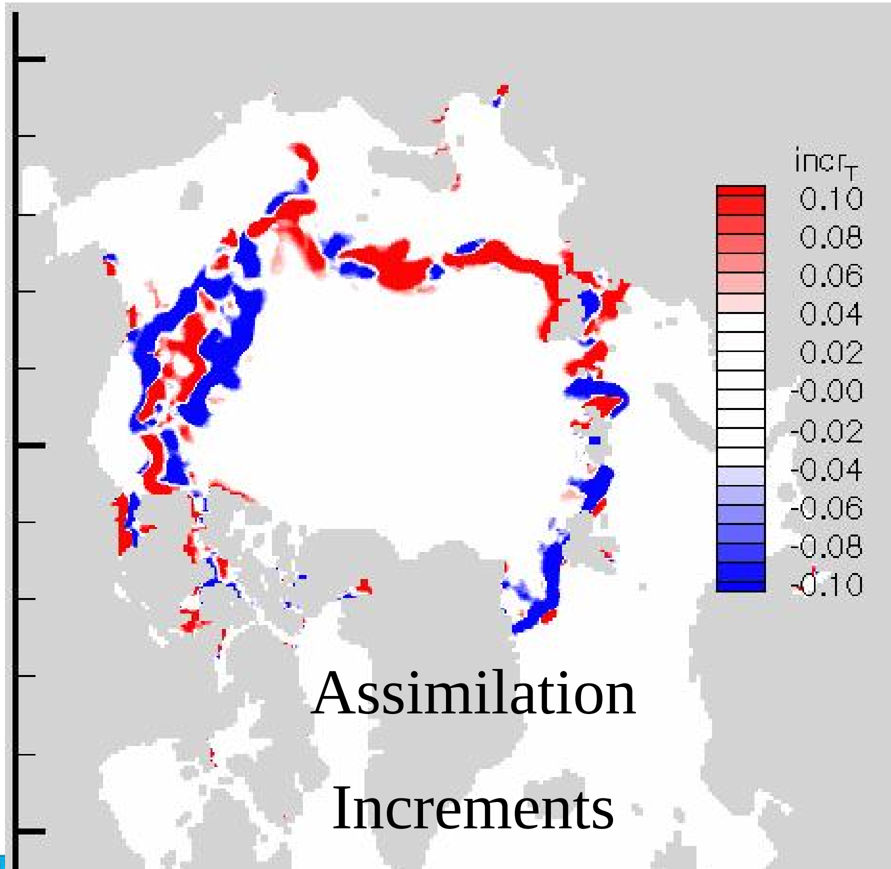
Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

EnKF setup: Effect of localization
Assimilation of ice concentrations

Forecast errors | Analysis errors, Global analysis | Analysis errors, Local analysis

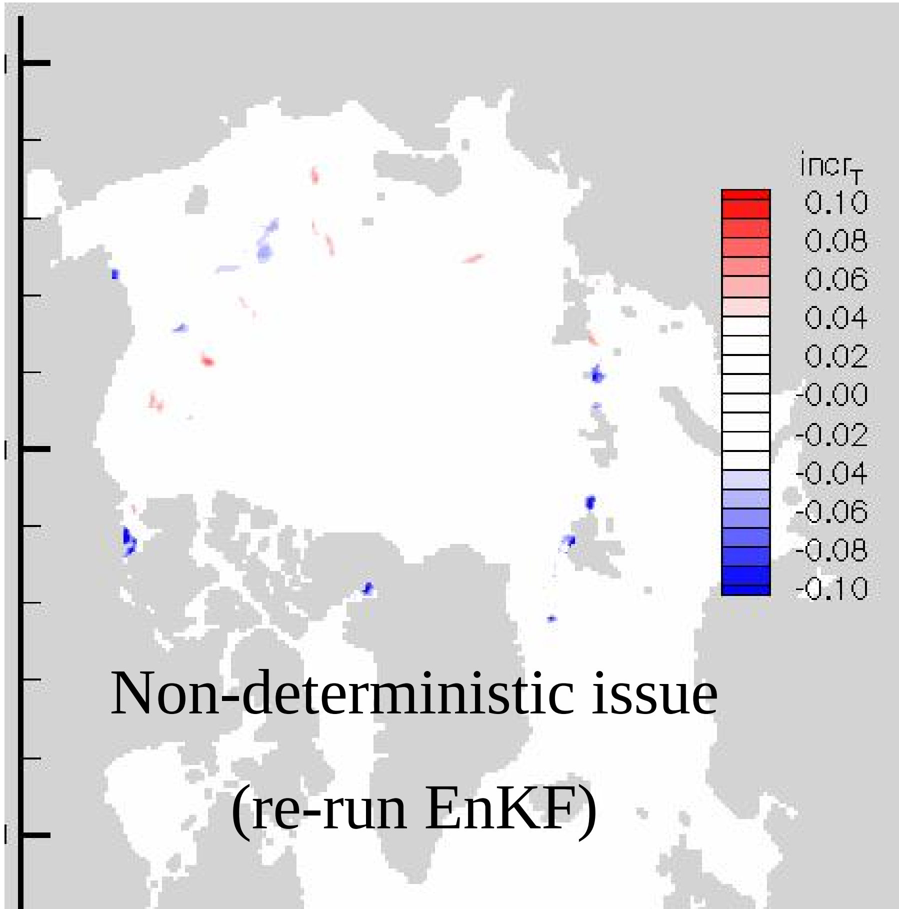


Ice Concentrations assimilation on*19th Sept. 2006*



Assimilation

Increments



Non-deterministic issue

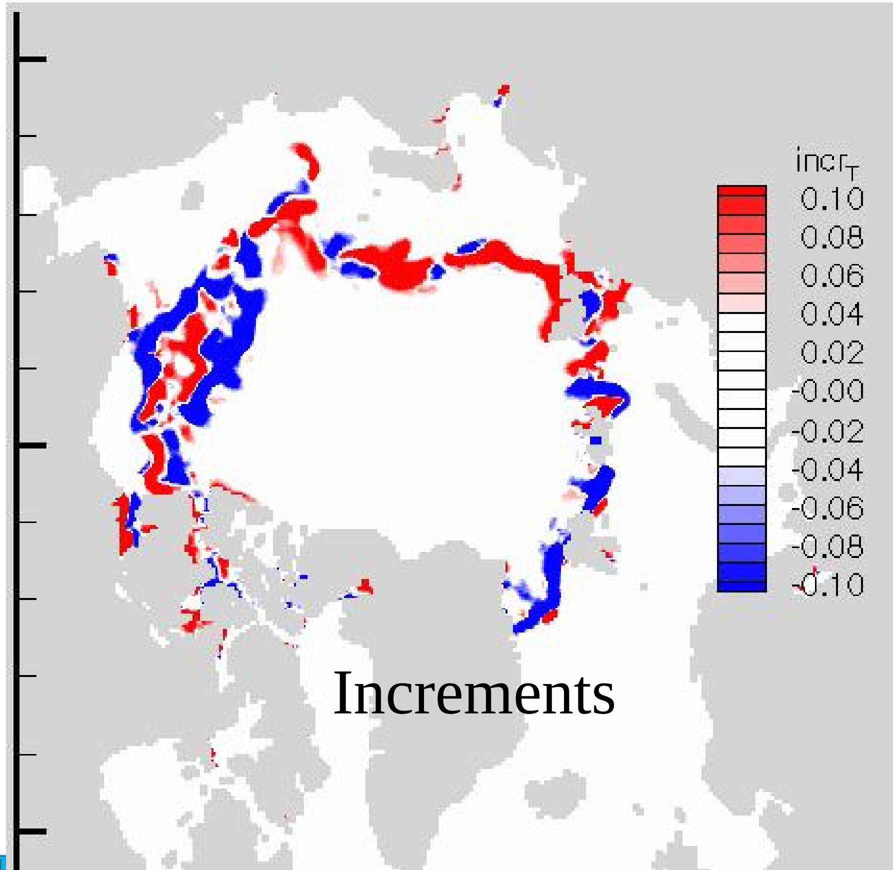
(re-run EnKF)

NERSC

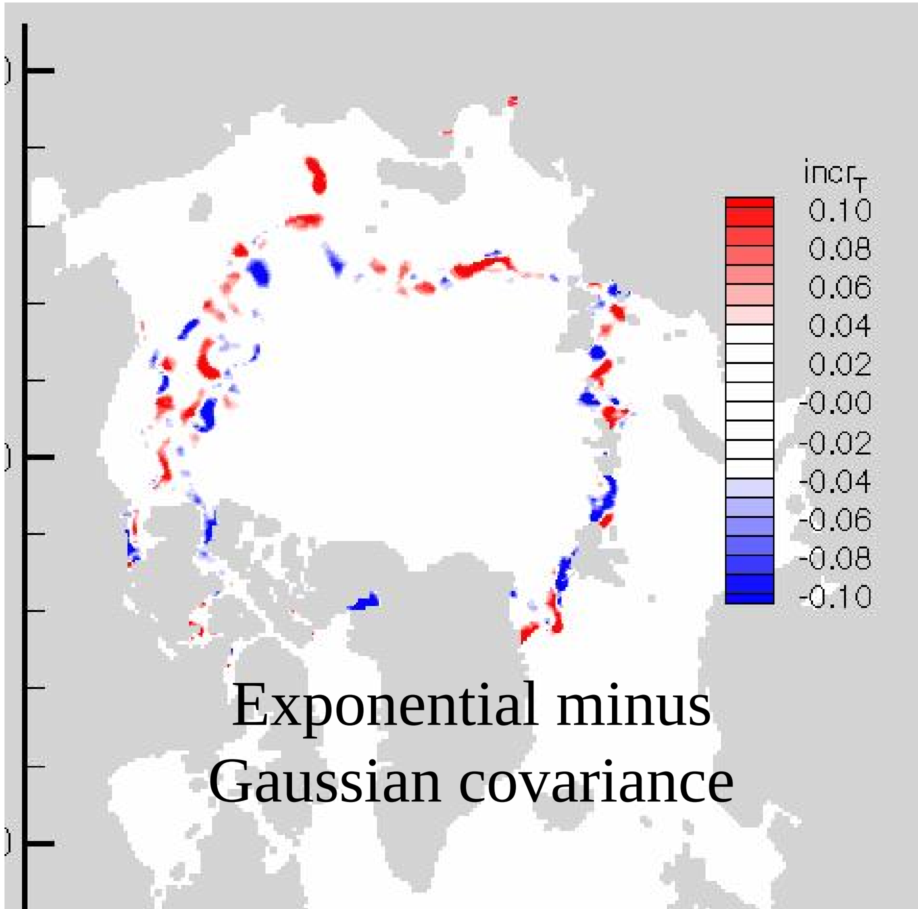
Mohn-Sverdrup Center

Global Ocean Studies - Operational Oceanography

Structure of the measurement errors



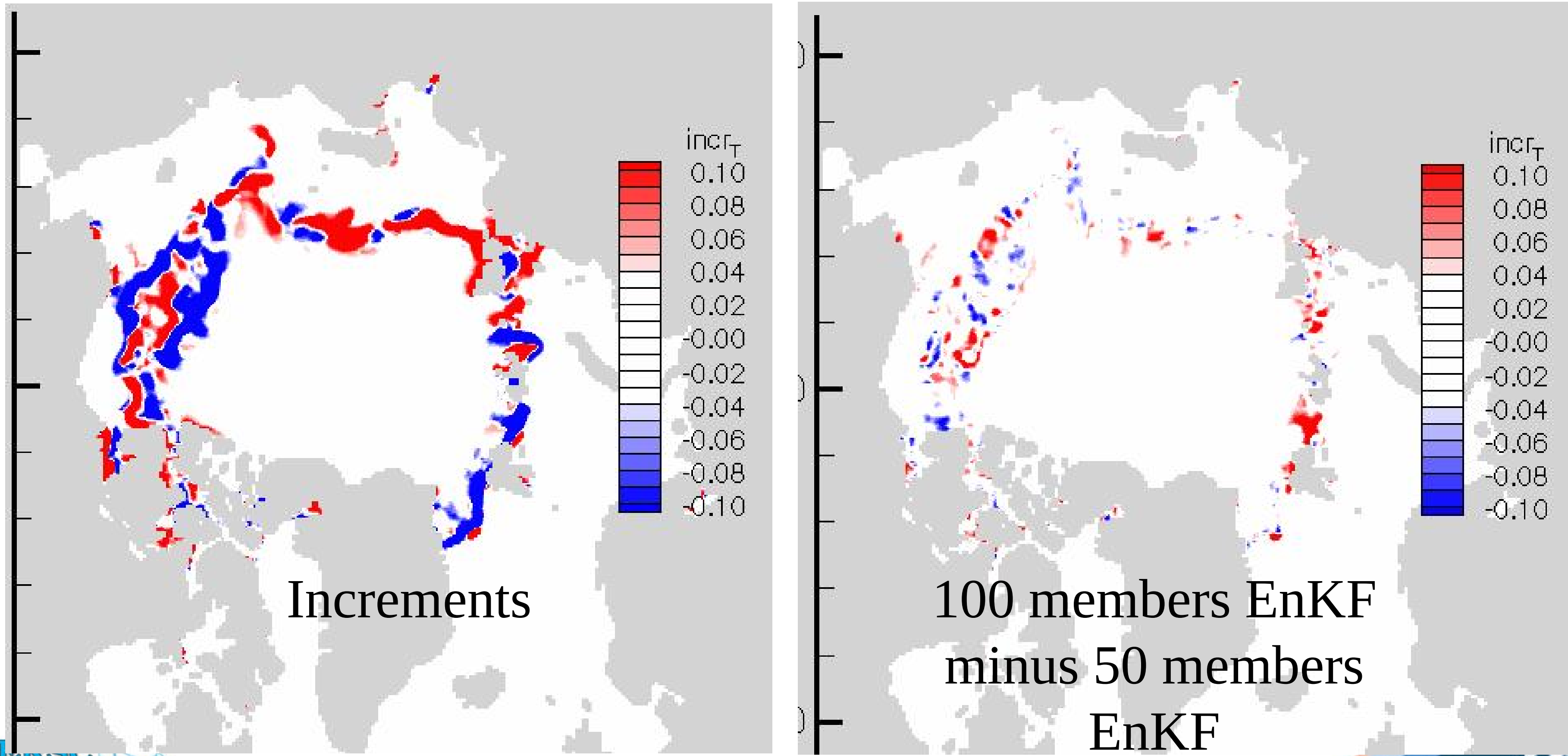
Increments



Exponential minus Gaussian covariance

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Ensemble size



System Applications

Nested systems in

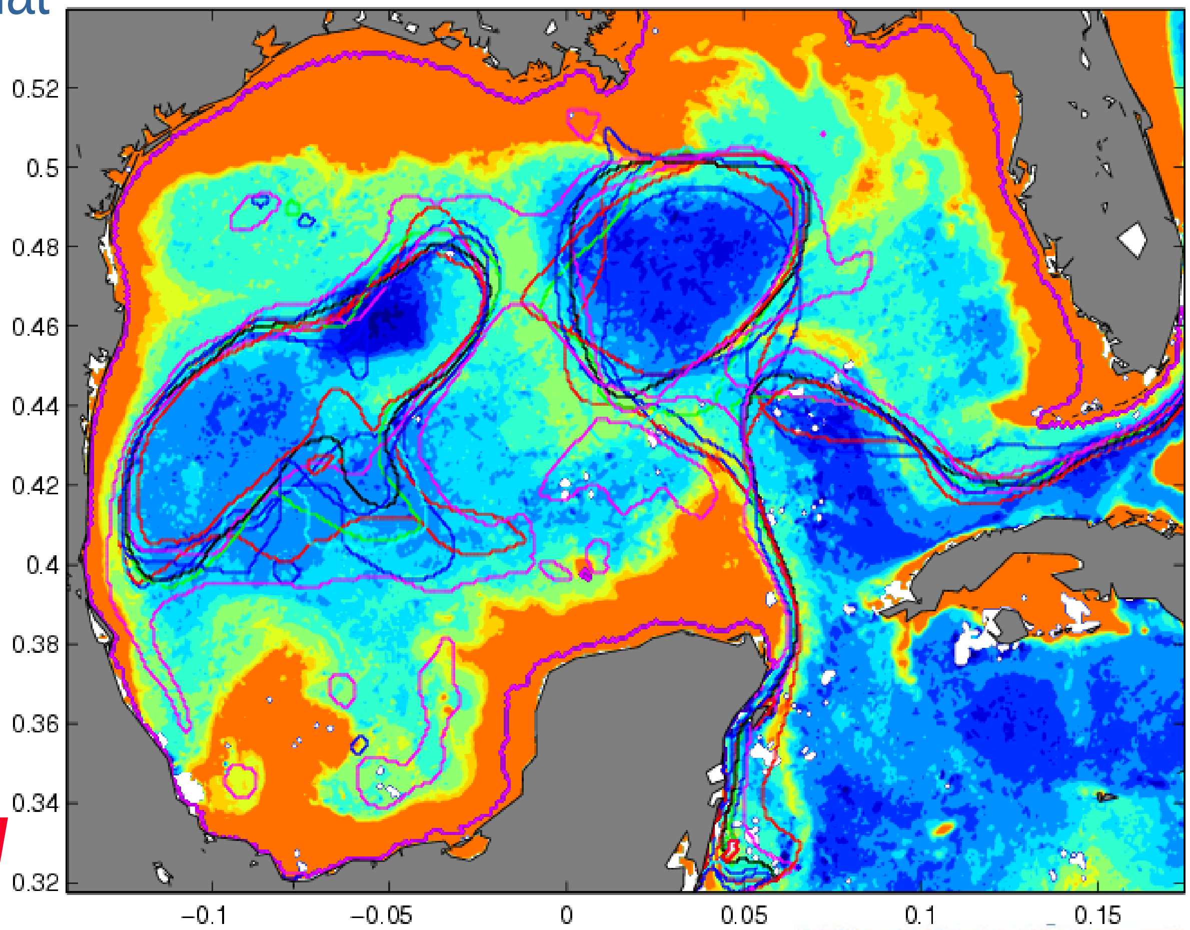
1. North Sea (N. Winther/C. Hansen)
2. Gulf of Mexico (F. Counillon)
3. Barents Sea (I. Keghouche)

Ensemble Forecasting
in the Gulf of Mexico

- What is the probability that an eddy will shed next week?
- Lines ("spaghetti plot")
  - Model fronts
  - 7 days forecast
- Background
  - Satellite data
  - Ocean color (MODIS)
  - Not assimilated

NERSC

[F. Counillon]



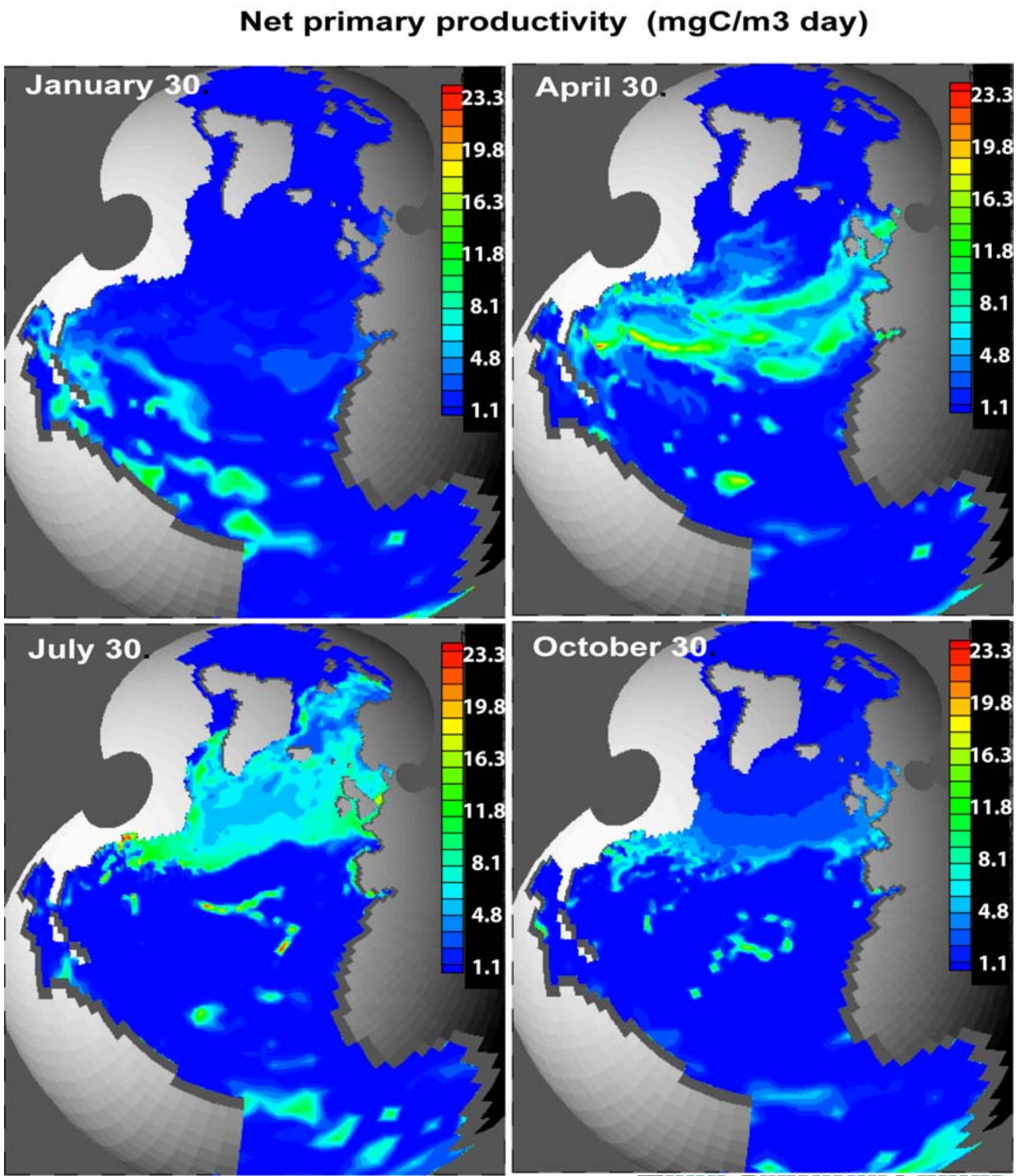
Perspectives

Non-Gaussian estimation

(case of ecosystem variables)

Coupled HYCOM – bio. models

- A physical ocean model can drive an ecosystem model
  - Re-suspension of nutrients from the sea bottom
  - Blooming of phytoplankton
  - Grazing of phytoplankton by zooplankton
- Ecosystem variables are particularly non-Gaussian



[A. Samuelsen

C. Hansen]

NERSC

Global Ocean Studies - Operational Oceanography

Theoretical problems

- Non-linear models
  - No guarantee of Gaussian distributions
- We can apply the Gaussian assumption, but
  - Is a **linear analysis** still optimal?
  - Is a **linear analysis** still unbiased?
- The Gaussian Anamorphosis from geostatistics offers a possible extension

Illustration

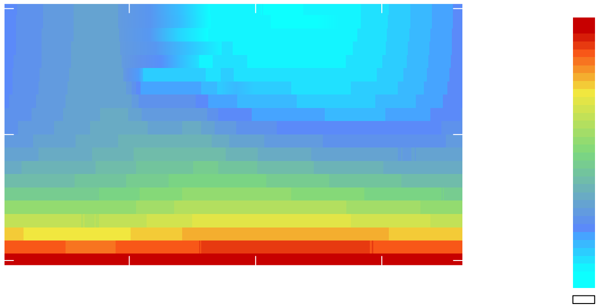
Idealised case: 1-D ecological model

- Spring bloom model, yearly cycles in the ocean
- *Evans & Parslow (1985), Eknes & Evensen (2002)*

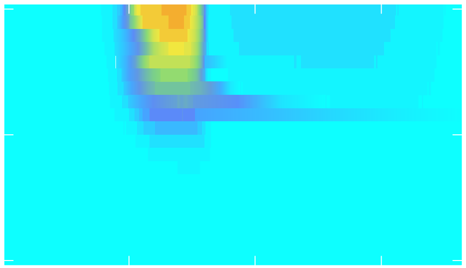
Characteristics

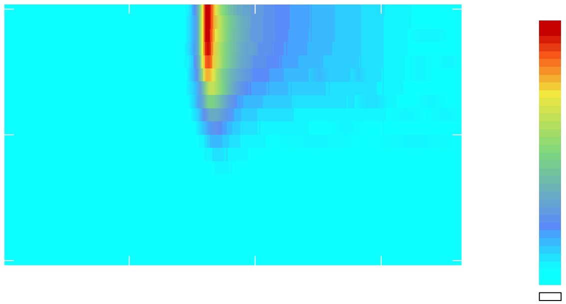
- *Sensitive to initial conditions*
- *Non-linear dynamics*

[Bertino et al. 2003]



Nutrients





Phytoplankton

time-depths plots

Herbivores

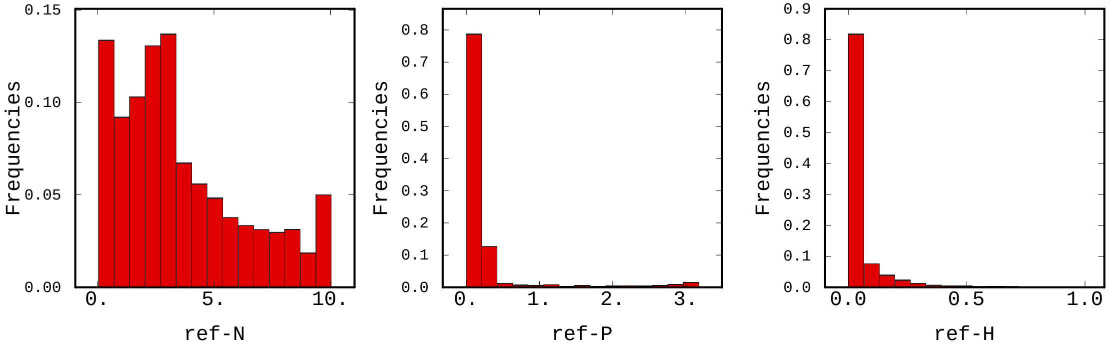
NERSC

Anamorphosis
(logarithmic transform)

__Original__

__histograms__

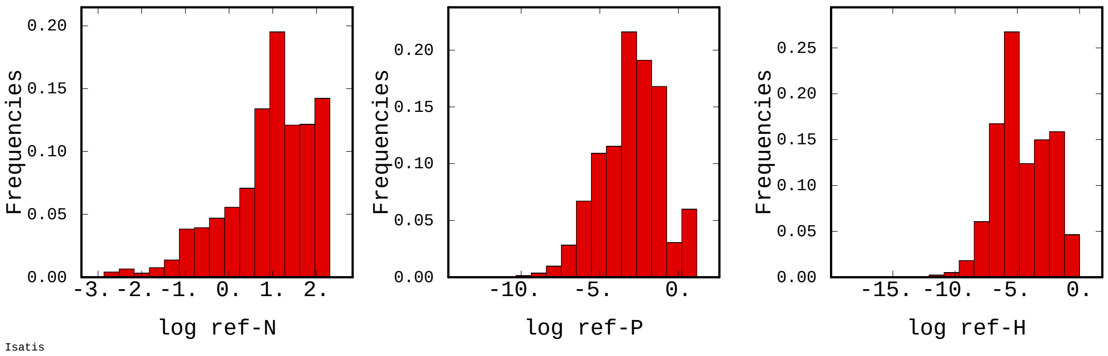
asymmetric



N P H

__Histograms__
__of logarithms__

less
asymmetric



NERSC

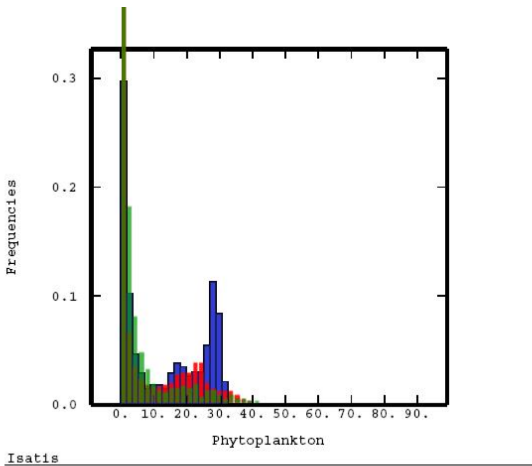
Arbitrary choice, possible refinements (polynomial fit)

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

Anamorphosis

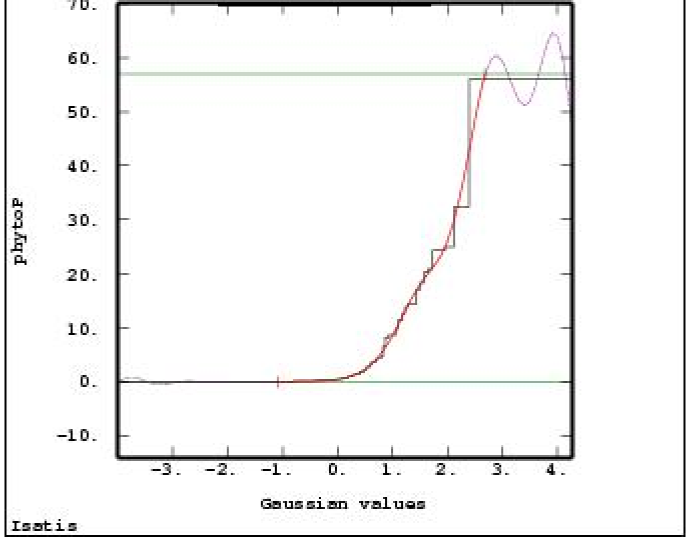
A classical tool from geostatistics

Physical variable

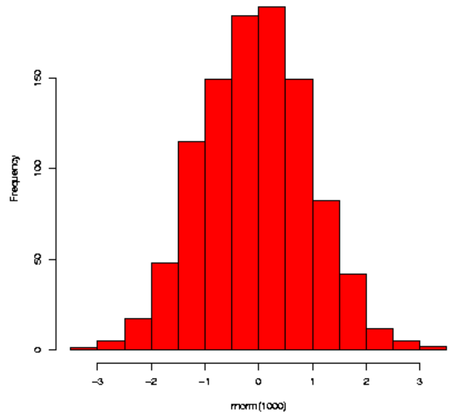


Example: phytoplankton in-situ concentrations

Cumulative density function



Statistical variable



NERSC

Mohn-Sverdrup Center

Global Ocean Studies - Operational Oceanography

Anamorphosis in sequential DA

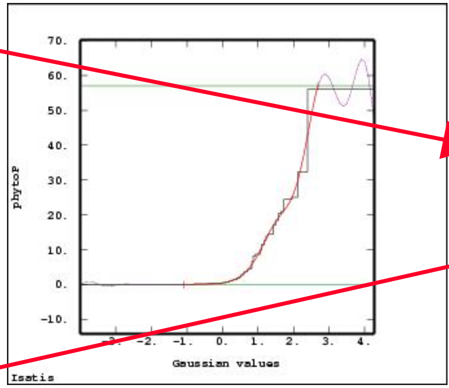
separate the physics from statistics

Physical operations:

Forecast

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

Anamorphosis function



Statistical operation: A and Y transformed

Analysis

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

Forecast

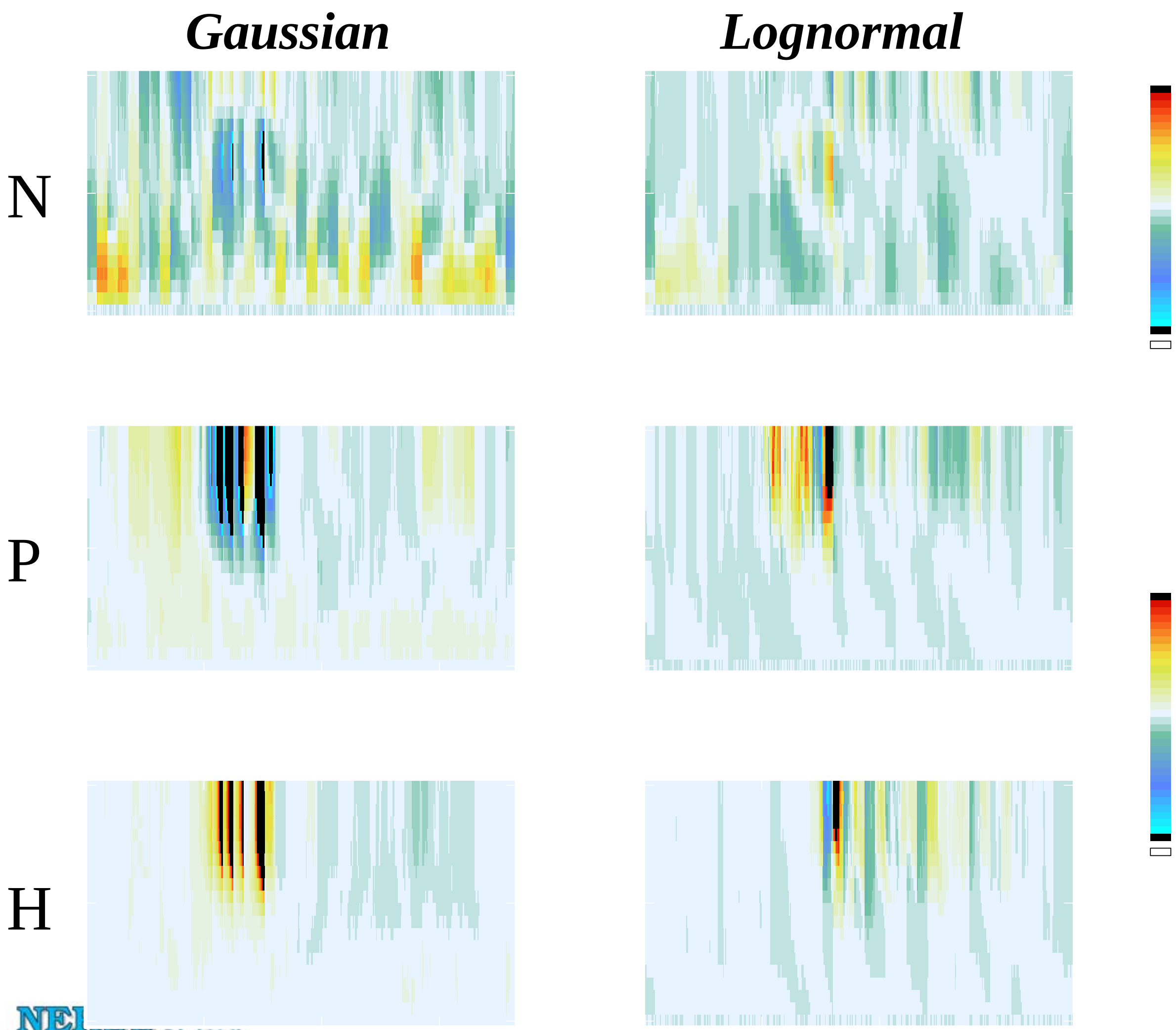
$A^f_{n+1} = f\,(A^a_n)$

- Polynomial fit, distribution tails by hand

NERSC

Mohn-Sverdrup Center
Global Ocean Studies - Operational Oceanography

EnKF assimilation results



RMS errors

- Gaussian assumption
  - Truncated H < 0
  - Low H values overestimated
  - "False starts"
- Lognormal assumption
  - Only positive values
  - Errors dependent on values

Conclusions

- Monte-Carlo methods for operational forecasting
- Large state and observations dimensions
- Non-linear and evolutive system
  - Justifies the use of dynamical data assimilation
- Ensemble statistics make sense
  - Prior Initial/Model errors are critical
- EnKF developments needed
  - Non-Gaussian estimation
  - Bias reduction
  - Improved sampling
  - Parameter estimation …