# Data Assimilation in GOAPP North Atlantic model

*Entcho Demirov, Keith Thompson and Dan Wright*

# OUTLINE

- Overview of SEEK filter
- Ongoing research in Memorial University
- Future plans

# The conceptual model

- Consider a physical system described by

$$x^t_t = \mathcal{M}\, x^t_{t-1} + \eta(t)$$

where the transition operator $\mathcal{M}$ is often described in terms of system of PDEs

- and observations $y_t$

$$y_t = \mathcal{H}\, x^t_t + \varepsilon(t)$$

# Bayesian Data Assimilation

- A major task of ocean data assimilation is to estimate accurately the probability density function (pdf) for the current ocean state, given all current and past observations $\mathbf{y_t}$.

- Adopting a probabilistic approach, the state of the ocean $x_t$ at some tine step $t$ has a conditional probability density $\mathbf{p(x_t \mid y_t)}$

## Two steps to the general data assimilation

Assume that a pdf of the ocean state is available (in the lack of knowledge this may be climatological pdf)

- The first step is to assimilate recent observations, thereby sharpening the pdf
- The second step is to propagate pdf forward in time until new observations become available. If the pdf is initial sharp, the chaotic dynamics and model uncertainty will broaden the probability distribution.

# Bayesian Updating

- **The update problem is to accurately estimate $P(x^t_t \mid y_t)$ , the probability of the current ocean state, given present and past observations, which is given by the Bayes' rule with accuracy to a normalization constant:**

$$P(x^t_t \mid y_t, y_{t-1}, \ldots, y_1) \approx P(y_t, y_{t-1}, \ldots, y_1 \mid x^t_t)\, P(x^t_t)$$

# Observational Error

- Assume that observation errors are independent from one time to the next

$$P(y_t, y_{t-1}, \ldots, y_1 | x^t_t) \approx P(y_t | x^t_t)\, P\,(y_{t-1}, \ldots, y_1 | x^t_t)$$

- This may be not always true especially in satellite data, where the bias error is not always easy to remove. Errors of observation may be flow dependent (Daley, 1993)

# Bayesian Updating

- Under this assumption and using the Bayes' rule again we may obtain:

$$P(x^t_t \mid y_t, y_{t-1}, y_{t-2}, \ldots, y_1) \approx P(y_t \mid x^t_t)\, P(x^t_t \mid y_{t-1}, y_{t-2}, \ldots y_1)$$

- This is a recursive relations which allows to estimate the 'posterior' pdf by using pdf of current observations and 'prior' (background) pdf.

# Forecasting of probability density

- With an updated pdf a method of forecasting the evolution of this pdf in time is needed
- Assume
  $\eta(t) = G(x^t_t)\, d\, q$
  where $dq$ is a Brownian-motion process with covariance $Q_t dt$.
- Conceptually the time evolution of the pdf can be modeled with Fokker-Planck equation: pdf diffuses in time according to

  (a) chaotic dynamics of the forecast model and

  (b) model error, including the increasing diffusion of pdf due to model uncertainty and drift.

# Limitations of Bayesian data assimilation

- Neither the update nor the forecast steps of Bayesian data assimilation can be applied directly to real-world ocean predictions
- For the update step one problem is the high dimension of the ocean state.
- Fokker-Planck equation can not be integrated in a high-dimensional system

# The Kalman filter

- The Kalman filter is an approximation to Bayesian approach which assumes:

  (i) linearity of error growth
  (ii) normality of the error distribution

- There are two steps of the Kalman filter:

  (i) update state, where the state estimate and model uncertainty are adjusted to new observations
  (ii) forecast step – uncertainty estimate is propagated forward

# The extended Kalman Filter

- Kalman filter was designed for linear systems only. For non-linear system $\mathcal{M}$, and $\mathcal{H}$ are linearized around $x^t_t$ :

$$H = \partial\mathcal{H}/\partial x \quad \text{and} \quad M = \partial\mathcal{M}/\partial x$$

- It assumes that the background and observational error distributions are Gaussian

# EKF equations

- EKF is recursive: assuming the estimate of the state vector $x^a_i$ and analysis error covariance $P^a_i$ are known at time $t_i$.
- Forecast:

$$x^f_{i+1} = M(x^a_i)$$
$$P^f_{i+1} = M\, P^a_i\, M^T + Q$$

- Analysis (update):

$$K_{i+1} = P^f_{i+1} H^T ( H P^f_{i+1} H^T + R)^{-1}$$
$$x^a_{i+1} = x^f_{i+1} + K_{i+1} (y_{i+1} - H\, x^f_{i+1})$$
$$P^a_{i+1} = ( I - K_{i+1} H )\, P^f_{i+1}$$

# Limitations to extended Kalman filter

- Normality of the pdf can be not appropriate for some ocean parameters.
- Application of KF to nonlinear systems results in instability in covariance equation.

# Limitations to extended Kalman filter

- Error covariances must be carefully estimated and monitored.
- Estimating Q and R may be particularly difficult.
- Computationally heavy

# Sub-optimal Kalman filter

- The Kalman filter is optimal when all **Q, R $P^a_0$** are well known.

- The dimension of these matrices is high for real models and the estimation of all their elements is impossible with high accuracy

- If KF is stable, then the choice of **$P^a_0$** may be less restrictive.

# The SEEK filter

- Covariance matrix $P = L\,U\,L^T$
- The equation of analysis error covariance is projected onto singular modes.

  $L_i = M_{i-1,i}\,L_{i-1}$

  $$U_i^{-1} = [U_{i-1} + (L^T{}_i L_i)^{-1}\,L^T{}_i\,Q_i\,L_i\,(L^T{}_i L_i)^{-1}]^{-1} +$$

  $$+\,L^T{}_i\,H^T{}_i\,R_i^{-1}\,H_i\,L_i$$
- Error subspace $S_o \approx L\,(U)^{1/2}$, $P = S_o S_o^T$

# Initial build of sub-space

- EOFs from climatology
- Prescribed covariance structure (using analytical models)
- Breeding modes
- Monte Carlo method

# Model set up

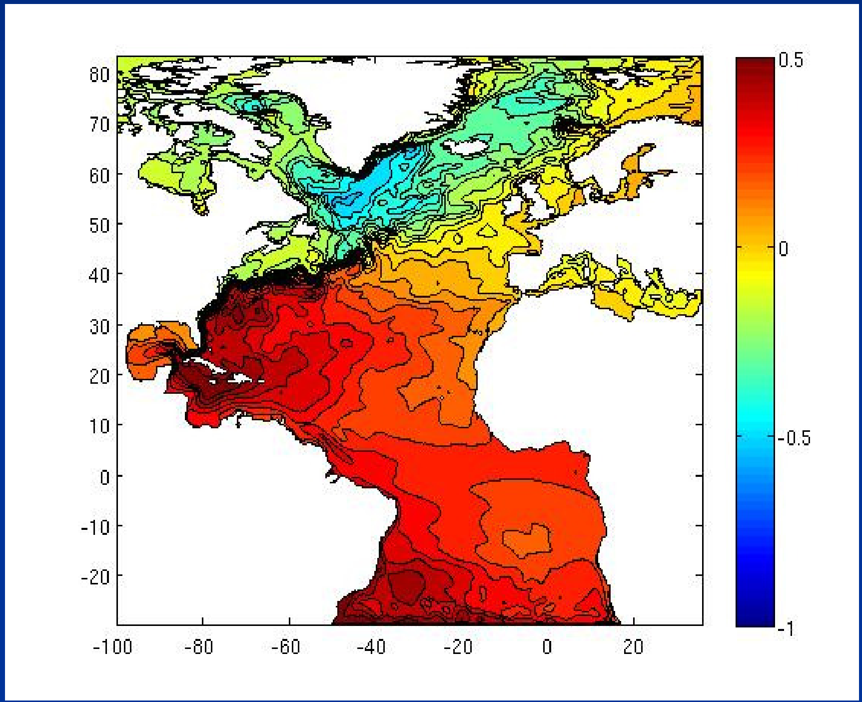
- North Atlantic model in two versions:

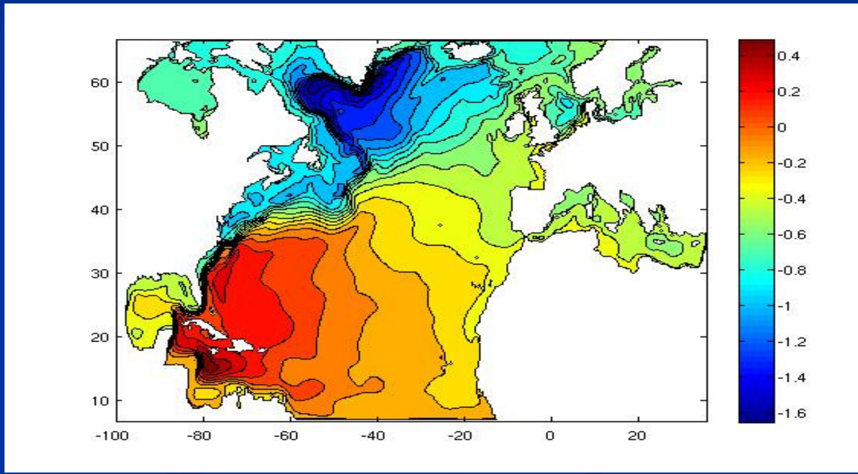
  (a) CMEP domain from 7 to 67 N
  (b) North Atlantic between 30 S an 81N

- Model uses standard BIO implementation

- Atmospheric forcing is from 6 hrs NCEP reanalysis

# The two model domains



# Model set up

- Spectral nudging
- OBCS
- AGRIF

# Plan of data assimilation development

- Two data assimilation schemes 3-D VAR and SEEK
- Comparison of data assimilation schemes
- Merged data assimilation scheme

# Dissemination

- Model experiment for evaluation of the data assimilation skills. Focused on predictability.
- Near – real time pilot forecasting
- Preoperational GOAPP North Atlantic Forecasting System (Yimin Liu)
- Operational GOAPP North Atlantic Forecasting System