Covariance modeling and data assimilation with the Navy Ocean Models

Dr. HANS NGODOCK

Drs. MATTHEW CARRIER, SCOTT SMITH, MAX YAREMCHUK, PHIL MUSCARELLA

Naval Research Laboratory
Stennis Space Center, MS
Goal: replace OI-based assimilation system with 3d and 4d-Var

3DVAR assimilation existing operational data streams

4DVAR or 3DVAR with high resolution

4DVAR high resolution with In-situ data

Existing operational data streams + other available in situ data

Gliders, HF Radar, Moorings
OUTLINE

• **3DVAR & CORRELATION MODELING**
  – Generalized diffusion equation
  – Anisotropic correlations
  – Explicit in Data space, Implicit in Data space and State space
  – Hybrid 3DVAR
  – Multiscale 3DVAR

• **4DVAR**
  – Cycling representers
  – Application with NCOM

• **OTHER PROJECTS INVOLVING DA**
  – Hycom
  – Wave
  – Ice
  – Tides
Errors, errors and more errors

- The model and data contain errors: *external forcing errors, initial and boundary condition errors, bad parameterization, unresolved processes, instrument error and error of representativeness ... etc.*

- These errors are unknown and not measurable. We can only make assumptions about them.

- Those assumptions are contained in the covariances that appear in the cost function.

- The covariances will determine how the assimilation fits the data and propagate/extend the information from the measurement sites to all model variables through the domain.

- It is critical that the covariances be prescribed carefully and assessed through a significance test (usually ignored in most DA applications).
COVARIANCE MODELLING: the science of DA

Prescribing covariances is the toughest problem in data assimilation

“The most important element in the statistical interpolation algorithm is the background error covariance. To a large extent, the form of this matrix governs the resulting objective analysis.” Roger Daley (1992)

“Construction of these error estimates is the most challenging and scientifically important task.” Olivier Talagrand, LMD Paris, France (1999)

“It is difficult to develop covariances. It follows that the resulting inverse estimate or analysis of the circulation also lack credibility.” Andrew Bennett, OSU (2002)

“... the choice of the weights dominates the current effort. At this stage, we have a problem primarily of oceanography and meteorology rather than one of mathematics.” Carl Wunsch, MIT (2006).
Background Error Covariance

Assume the background error covariance is of the form

$$ B(x, x') = \exp\left(\frac{-|x - x'|^2}{2L^2}\right) $$

The multiplication of $B$ with a state variable $u(x)$

$$ \lambda(x) = \int B(x, x')u(x')dx' $$

may be obtained by solving the diffusion equation

$$ \frac{\partial v}{\partial t} = \Delta v \quad 0 \leq t \leq L^2/2 $$

$$ v(x, 0) = u(x) $$

$$ \lambda(x) = v(x, L^2/2) $$

$u(x)$ is a Dirac delta
Time integration

- The diffusion equation can be integrated with either an explicit or an implicit time stepping.

\[
\nu^{n+1} = \nu^n + \delta t D \nu^n \\
B = \left( I + \frac{tD}{n} \right)^n
\]

\[
\nu^{m+1} = \nu^m + \delta t D \nu^{m+1} \\
B = \left( I - \frac{tD}{m} \right)^{-m}
\]

- The explicit solution suffers from harsh CFL criterion.
- Implicit approach more attractive, being unconditionally stable.
Anisotropy: examples

To account for spatial variations in the correlation length, Weaver and Ricci (2003) introduce a function of space as diffusion coefficients instead of a constant, replacing $\kappa$ with $\kappa_\text{SSH}$ in the diffusion equation.

Choices for $\kappa$ : Bathymetry, SSH, Velocity, Density
The structure of B is poorly known: only a small number of eigenvectors can be captured with confidence. Formulations are equivalent only if the increment $\delta x$ is restricted to the subspace spanned by the eigenvectors of B. Result of assimilation crucially depends on the model of B (or $B^{-1}$). Thus, given the same model background, covariance and data, both data space and state space approaches should yield the same analysis.
Computational cost

\[ D = \nabla \nu \nabla \]

**Obs space formulation:**  
\[ B = \exp(Dt) \]

\[ \delta x = B H^T \left[ H B H^T + R \right]^{-1} \delta d \]

explicit “time integration”  
\[ [I + Dt/n]^n = B \]

implicit “time integration”  
\[ [I - Dt/m]^{-m} = [I - Dt/m]^m = (B^{-1})^{-1} \quad (m \sim 2-3) \]

\[ n_{expl} \sim (\rho/dx)^2 \quad n_{impl} \sim m(\rho/dx) \quad \text{max}\{\rho/dx\} \sim 10-20 \]

\[ n_{expl} \sim 100-400 \quad n_{impl} \sim 20 - 60 \]

\[ \text{CPU(obs)} \sim [C(HB^HT+R)]^{1/2} m(\rho/dx) \sim 100m(\rho/dx) \sim 2000 - 6000 \]

**Model space formulation:**  
\[ B^{-1} = \exp(-Dt) \]

\[ \delta x = \left[ B^{-1} + H^T R^{-1} H \right]^{-1} H^T R^{-1} \delta d \]

\[ \text{CPU(mod)} \sim C\left[ (I-Dt/m)^m + H^T R^{-1} H \right]^{1/2} < C\left[ (I-Dt/m)^m \right] \sim (\rho/dx)^m \sim 100-8000 \]
control misfits (solid red), implicit-3DVAR misfits (dash green), and explicit-3DVAR misfits (dash blue) at the initial time (a) and at days 5 (b), 10 (c), 15 (d), 20 (e), and 25 (f).
Relative forecast error in temperature (red), salinity (blue), SSH (black), and total velocity (green) for implicit-3DVAR (solid) and explicit-3DVAR (dashed). Relative error is normalized by the error in each field at day 1.
RELO NCOM/NCODA configured for the Hawaiian Island region during RIMPAC 2008 (6/18/08 – 7/10/08)
- 6 km grid using NRL DBDB2 2' bathymetry
- Global NCOM lateral boundary conditions
- OSU OTIS tidal heights and transports
- Surface forcing using bulk formulae with NOGAPS fields
- Data assimilation: GFO/Jason/Envisat GAC/LAC/ship XBT/ARGO/gliders
- 24h update cycle

Nested NCOM data coverage for 00Z 16 June 2008
RIMPAC 08 Experiment

Validation with independent temperature profiles on 6/16/08
RIMPAC 08 Experiment

Validation with independent salinity profiles on 6/16/08
Given a model forecast $x^f$ and observations $y$, minimize the cost function

$$J(\delta x) = 0.5\delta x^T B^{-1} \delta x + 0.5 (H\delta x - d)^T R^{-1} (H\delta x - d) \quad \delta x = x - x^f, \quad d = y - Hx^f$$

**Decomposition into large and small scale components**

$$x^f = x_L^f + x_S^f, \quad y = y_L + y_S$$

$$H = H_L + H_S, \quad B = B_L + B_S, \quad R = R_L + R_S$$

**Assimilate incrementally for large and small scales**

$$J_L(\delta x_L) = 0.5\delta x_L^T B_L^{-1} \delta x_L + 0.5 (H\delta x_L - d)^T \left( R + H B_S H^T \right)^{-1} (H\delta x_L - d)$$

$$J_S(\delta x_S) = 0.5\delta x_S^T B_S^{-1} \delta x_S + 0.5 (H\delta x_S - d)^T \left( R + H B_L H^T \right)^{-1} (H\delta x_S - d)$$
Multiscale 3DVAR

If dense observations are available, in addition to coarse observations:

\[
J_L (\delta x_L) = 0.5 \delta x_L^T B_L^{-1} \delta x_L + 0.5 \left( H^c \delta x_L - d^c \right)^T \left( R^c + H^c B_S H^{cT} \right)^{-1} \left( H^c \delta x_L - d^c \right)
+ 0.5 \left( H^d \delta x_L - d^d \right)^T \left( R^d + H^d B_S H^{dT} \right)^{-1} \left( H^d \delta x_L - d^d \right)
\]

\[
J_S (\delta x_S) = 0.5 \delta x_S^T B_S^{-1} \delta x_S + 0.5 \left( H^d \delta x_S - d^d \right)^T \left( R^d + H^d B_L H^{dT} \right)^{-1} \left( H^d \delta x_S - d^d \right)
+ 0.5 \left( H^c \delta x_S - d^c \right)^T \left( R^c + H^c B_L H^{cT} \right)^{-1} \left( H^c \delta x_S - d^c \right)
\]
Multiscale 3DVAR

- Navy Coastal Ocean Model (NCOM)
- 3km horizontal resolution
- 46 vertical levels
- Kuroshio Extension western boundary current
- Multi-Scale 3DVAR system (Li et al. 2011)
- Assimilating simulated observations of temperature and salinity
- Large Scale – 100km, small Scale – 10km
- Simulate a control solution from which observations are sample: 17-day delay
- Simulate erroneous solution to be corrected by the assimilation
High resolution flow in the Kuroshio extension

RELO-NCOM Sea surface temperature

Data Type: temp
Date = 09/01/2010:00
Level = 1
Error metric evaluated only in the upper 500m to isolate the effects of small scale processes

Multiscale assimilation fits observations better than traditional method at observation locations

Main difference with the control solution is clearer in the small scales region

**Multiscale 3DVAR**

\[
\text{Analysis Error metric} = \frac{1}{N} \sum_{n=1}^{N} \frac{|O_n - M_n|}{\sigma_n}
\]
Forecast Error metric

\[ \frac{1}{N} \sum_{n=1}^{N} \frac{|O_n - M_n|}{\sigma_n} \]

- Forecast error growth on the 19\textsuperscript{th} to 22\textsuperscript{nd} due to local disturbance in the small scale region
- Large scale assimilation failed to capture the disturbance for lack of coarse data
- Dense observation coverage captured the disturbance, therefore the multiscale assimilation was able to correct the forecast error
Multiscale 3DVAR: real observations

1. Assimilation of routinely collected observations (operational data streams processed through NCODA).

2. Multiscale algorithm shows similar accuracy as with the simulated observations

3. Muscarella, Carrier and Ngodock: Multiscale 3DVAR assimilation in the Kuroshio extension (Manuscript in preparation)
Hybrid model: formulation

\[ J(\delta x) = \frac{1}{2} [\delta x^\top B^{-1} \delta x + (H\delta x - \delta y)^\top R^{-1} (H\delta x - \delta y)] \rightarrow \min_{\delta x} \]
\[ \delta y = Hx_b - d \]

\[ B = B^{-1} = \frac{1}{\alpha} B_m^{-1} + \frac{1}{\beta} B_0^{-1} \]

\[ B^{-1} = \alpha P \Lambda_m^{-1} P^\top + \beta P_\perp B_0^{-1} P_\perp^\top \]
\[ P_\perp = I - PP^\top \]

\[ B = \frac{1}{\alpha} P \Lambda_m P^\top + \frac{1}{\beta} [P_\perp \exp(-Dt)P_\perp^\top]^{-1} \]
\[ D = \nabla^\top \nu \nabla \]

- heuristic (Gaussian)
- dynamical (derived from model statistics)
Hybrid model: Twin data experiments

Gliders:
- duration: 27 days
- $K \approx 1.5 \times 10^5$

Model:
- NCOM 1.5km 40 levels
- $M = 515,450$

Validation:
- a) 12hr forecast errors
- b) Two moorings with
  - 18 levels U/V
  - 11 levels T/S
Hybrid model: Twin data experiments

“True”

First guess
**Hybrid model:** Skill assessment

**Metric**

\[ G = \text{diag}\{g_T, g_S, g_u, g_v\} \quad g_\xi(x) = \left[ \xi(x) - \bar{\xi}(x) \right]^{2^{1/2}} \]

**Distances**

\[ r_\xi^s(x_1, x_2) = \left\langle (\xi_1 - \xi_2)^2 g_\xi^{-2} \right\rangle^s_{1/2} \quad r_\xi^g(x_1, x_2) = \left\langle (\xi_1 - \xi_2)^2 R^{-1} \right\rangle^g_{1/2} \]

state space  \quad  \text{obs space}

**Skill**

\[ q_{\xi}^{g,s}(t) = \frac{r_\xi^g(x^t, x^f)|_t}{r_\xi^g(x^t, x^f)|_0} \]
Hybrid model: Twin data experiments results
4DVAR

\[
\begin{aligned}
\frac{\partial X}{\partial t} &= F(X) + f, \quad 0 \leq t \leq T \\
X(t = 0) &= I(x) + i(x) \\
\bar{f} &= 0, \quad \bar{f}f^T = C_f, \quad \bar{i} = 0, \quad \bar{i}i^T = C_i \\
y_m &= H_m X + \varepsilon_m, \quad 1 \leq m \leq M, \\
\bar{\varepsilon} &= 0 \quad \bar{\varepsilon}\varepsilon^T = C_\varepsilon
\end{aligned}
\]

\[
J = \iiint_{\Omega} \iiint_{\Omega} f(x, t) W_f(x, t, x', t') f(x', t') dx' dt' dx dt
\]

\[
+ \iint_{\Omega} \iint_{\Omega} i(x) W_i(x, x') i(x') dx' dx
\]

\[
+ \varepsilon^T W_\varepsilon \varepsilon
\]

\[
C_i(x, x') = v(x)^{1/2} v(x')^{1/2} \exp\left(-\frac{|x-x'|^2}{2L^2}\right)
\]

\[
C_f(x, t, x', t') = C_i(x, x') \exp\left(-\frac{|t-t'|}{\tau}\right)
\]

\[
W_i(x, x') = C_i^{-1}(x, x')
\]

\[
W_f(x, t, x', t') = C_f^{-1}(x, t, x', t')
\]

\[
W_\varepsilon = C_\varepsilon^{-1}
\]
Euler-Lagrange (EL) equations for minimizing the cost function

\[
\begin{aligned}
&\frac{\partial X}{\partial t} = F(X) + C_f \cdot \lambda, \quad 0 \leq t \leq T, \\
&X(t = 0) = I(x) + C_i \circ \lambda(x, 0) \\
&- \frac{\partial \lambda}{\partial t} = \left[ \frac{\partial F}{\partial X}(X) \right]^T \lambda + \sum_{m=1}^{M} \sum_{n=1}^{M} W_{\varepsilon, mn} \left( y_m - H_m X \right) \delta(x - x_m) \delta(t - t_m), \quad 0 \leq t \leq T \\
&\lambda(T) = 0
\end{aligned}
\]

Classic 4DVAR (strong constraints):

Assume no model and solve iteratively using a Newton-type technique

Same technique is not practical with model errors and high dimensions problems
4DVAR: The representer method (*Bennett* 1992)

Only applicable to linear EL equations: linearize the EL

\[
\begin{align*}
\frac{\partial X^k}{\partial t} &= F(X^{k-1}) + \left[ \frac{\partial F}{\partial X}(X^{k-1}) \right] (X^k - X^{k-1}) + C_f \cdot \lambda, \quad 0 \leq t \leq T, \\
X^k(t = 0) &= I(x) + C_i \cdot \lambda(x, 0) \\
- \frac{\partial \lambda}{\partial t} &= \left[ \frac{\partial F}{\partial X}(X^{k-1}) \right]^T \lambda + \sum_{m=1}^{M} \sum_{n=1}^{M} W_{\epsilon, mn} \left( y_m - H_m X^k \right) H^T \delta(x - x_m) \delta(t - t_m), \quad 0 \leq t \leq T \\
\lambda(T) &= 0
\end{align*}
\]

Expand the solution as a first guess plus a finite linear combination of representer functions

\[
X^k(x, t) = X^k_F(x, t) + \sum_{m=1}^{M} \beta^k_m r^k_m(x, t)
\]
4DVAR: The representer method (Bennett 1992)

Representer functions

\[
\begin{aligned}
- \frac{\partial \alpha_m^k}{\partial t} &= \left[ \frac{\partial F}{\partial X} \left( X^{k-1} \right) \right]^T \alpha_m^k + H^T \delta \left(x - x_m\right) \delta \left(t - t_m\right), \quad 0 \leq t \leq T \\
\alpha_m^k \left(T\right) &= 0 \\
\frac{\partial r_m^k}{\partial t} &= F \left( X^{k-1} \right) + \left[ \frac{\partial F}{\partial X} \left( X^{k-1} \right) \right] r_m^k + C_f \cdot \alpha_m^k, \quad 0 \leq t \leq T, \\
r_m^k \left(t = 0\right) &= C_i \circ \alpha_m^k \left(x, 0\right)
\end{aligned}
\]

Representer coefficients

\[
\left( R^k + C_\varepsilon \right) \beta = Y - H X^k_F
\]

\[
R_m^{kk} = H_n r_m \left(x, t\right) = r_m^k \left(x_n, t_n\right)
\]
1.5 nonlinear reduced gravity model

\[ \begin{align*}
\frac{\partial hu}{\partial t} + \frac{\partial uhu}{\partial x} + \frac{\partial uhv}{\partial y} - fhv + g'h \frac{\partial h}{\partial x} &= A_M \left( \frac{\partial^2 hu}{\partial x^2} + \frac{\partial^2 hu}{\partial y^2} \right) + \tau^x - \text{drag}_x, \\
\frac{\partial hv}{\partial t} + \frac{\partial vhu}{\partial x} + \frac{\partial vhv}{\partial y} + fhu + g'h \frac{\partial h}{\partial y} &= A_M \left( \frac{\partial^2 hv}{\partial x^2} + \frac{\partial^2 hv}{\partial y^2} \right) + \tau^y - \text{drag}_y, \\
\frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} &= 0,
\end{align*} \]

\[ \Delta x=20\text{Km}, \Delta x=18.75\text{Km}, \Delta t=20\text{min} \]

Active upper 300m

35Sv inflow/outflow

~4-month periodic eddy shedding

<table>
<thead>
<tr>
<th>Network</th>
<th>1 300km 10d</th>
<th>2 300km 5d</th>
<th>3 200km 10d</th>
<th>4 200km 5d</th>
<th>5 100km 10d</th>
<th>6 100km 5d</th>
<th>7 60km 10d</th>
<th>8 60km 5d</th>
</tr>
</thead>
<tbody>
<tr>
<td>#data</td>
<td>540</td>
<td>1080</td>
<td>1152</td>
<td>2304</td>
<td>5184</td>
<td>10368</td>
<td>14976</td>
<td>29952</td>
</tr>
</tbody>
</table>
1.5 nonlinear reduced gravity model: experiments setup

Initial condition for the representer = initial ensemble mean

Assimilation challenge: nonlinearity and purposefully misplaced eddy.

The EnKF, EnKS and the representer method will be compared based on the accuracy in the assimilation and the ensuing forecast.

\[ Q(x,t,x',t') = V(x,x') \exp\left(-\frac{(x-x')^2}{2L^2}\right) \exp\left(-\frac{|t-t'|}{\tau}\right), \quad L = 100km, \quad \tau = 10days, \quad V(x,x)=0.1Nm^{-2} \]
1.5 nonlinear reduced gravity model assimilation accuracy comparison

1.5 nonlinear reduced gravity model

forecast accuracy comparison (from network 3)
Cycling the representer algorithm

Using the representer method in a cycling assimilation-forecast setting:

- Xu and Daley 2000: 1D linear advection problem with update of the initial error covariance
- Xu and Daley 2002: 2D linear unstable barotropic problem, no initial error covariance update

In both cases, the cycling solution was more accurate (lower RMS misfit to the data/truth) than the non-cycling solution

Apply the concept to nonlinear models:
  Can we expect the same RMS behavior?
  How does cycling interact with potential TLM issues?
Cycling the representer algorithm: benefits

Background from previous cycle’s **nonlinear** forecast: ‘better’ than the corresponding portion of ‘global’ background

Shorter cycles: stable + accurate TLM

More accurate analysis

Cost reduction:
1) smaller minimization problem:
   conjugate gradient

   non-cycling \( N \log(N) \)
   cycling \( N \log(N_{cy}) \)
   \( N = \#\text{data in entire time interval} \)
   \( N_{cy} = \#\text{data within a cycle} \)

2) One outer loop !?!

⚠️ May require more outer loops in early cycles
Cycling the representer algorithm: Lorenz attractor

\[
\begin{align*}
\frac{dx}{dt} &= \sigma(y - x) + q^x \\
\frac{dy}{dt} &= \rho x - y - xz + q^y \\
\frac{dz}{dt} &= xy - \beta z + q^z
\end{align*}
\]

\[
\begin{align*}
x(0) &= 1.50887 + i^x \\
y(0) &= -1.531271 + i^y \\
z(0) &= 25.46091 + i^z
\end{align*}
\]

\[
\sigma = 28, \quad \rho = 10, \quad \beta = \frac{8}{3}
\]
Cycling the representor algorithm: Lorenz attractor

Ngodock, Smith, Jacobs,
Cycling the representer algorithm: Lorenz & Emanuel 1998

\[
\frac{\partial X_i}{\partial t} = (X_{i+1} - X_{i-2})X_{i-1} - X_i + 8, \quad i = 1, \ldots, 40 \\
\Delta t = 0.05 \sim 6\text{hr}
\]
1.5 nonlinear reduced gravity: truth – solution after 3 months

Network 3
U, V, SSH data

Network 2
U, V, SSH data

Network 1
U, V, SSH data

Network 1
SSH data
4DVAR: Application to NCOM in the Monterey Bay

Model domain

TLM stability

TLM Stability Test for All Fields

- T
- S
- U
- V
- SSH
4DVAR: twin data experiment

Strong constraints: wrong IC

Weak constraints: wrong surface forcing

Weak constraints: wrong IC & forcing

Obs. Locations, 6 cycles

Obs. Locations, 1 cycle

Obs. Locations, 1 cycle

All gridpoints

All gridpoints

Restored correct forcing after the first cycle

\[
\frac{1}{N} \sum_{n=1}^{N} \left| \frac{O_n - M_n}{\sigma_n} \right|
\]
Representers

Temperature

Salinity

Temperature cross-section
Representers

Temperature

Salinity

Temperature cross-section
4DVAR: Glider data

Temperature differences

Salinity differences
4DVAR: Glider data

Slocum 11

Slocum 6
4DVAR: Qualitative fit to the data

\[
\frac{1}{N} \sum_{n=1}^{N} \left| \frac{O_n - M_n}{\sigma_n} \right|
\]

10km correlation

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assimilation</td>
<td>86</td>
<td>95</td>
</tr>
<tr>
<td>First guess</td>
<td>68</td>
<td>80</td>
</tr>
<tr>
<td>Free run</td>
<td>64</td>
<td>76</td>
</tr>
</tbody>
</table>
4DVAR: Glider data

Withheld glider data

M1

M2
Conclusion

Successful implementation of diffusion-based correlation model in 3DVAR assimilation experiments with simulated and real data

Significant computational savings when the diffusion model is solved implicitly with the state space formulation

Promising results with hybrid and multiscale 3DVAR

Cycling the representer algorithm overcomes TLM limitations.

The representer method can be applied to operational NWP and ocean 4DVAR assimilation in place of the gradient descent, eventually at the same cost, albeit for model error covariance multiplications

No need for the ‘perfect model’ assumption

Promising results with the newly developed NCOM 4DVAR
QUESTIONS??
Time integration and anisotropy

Explicit \((I + Dt/n)^n\)

Implicit \((I - Dt/m)^{-m}\)
Cycling the representer algorithm: Open issues

Optimal choice of cycle length

Initial error covariance not being updated

Model error should be redefined in subsequent cycles

Forecasted open boundary conditions for nested models may be incompatible with corrected BCs

Idem for external forcing
Fitting the data quantitatively and qualitatively

Assimilated solution fits 80%, 90% and 95% of the observations to within one, two and three standard deviations respectively, while the corresponding numbers for the first guess are 60%, 75% and 82%, and 45%, 63% and 73% for the free running model.
Outer Loops

Analysis error

Forecast error
3DVAR/4DVAR assimilation in Monterey Bay

NCOM Fixed 2km resolution
Nested into Global 1/8 degree NCOM

NOGAPS atmospheric forcing

SSH
SST
MODAS T/S profiles

New implementation of NCOM with TL and adjoint models.

Weak constraint 4DVAR

Sample SST increment with MODAS synthetic T/S profile locations
Temperature profiles
OTHER PROJECTS INVOLVING DA

Observation Impact
Multiscale DA
Large scale HYCOM
TIDES
WAVE
Reduced order modeling
Biospace
... and more
Observation Impact Concept -

Observations move the forecast from the **background trajectory** to the **trajectory starting from the new analysis**

“Observation impact” is the combined effect of all of the observations on the difference in forecast error ($e_f - e_g$)

assumes a 24 hr assimilation window

**Langland and Baker (2004)**
Forecast error differences

\[ \delta e^g_f = \left< \left( y - Hx_b \right), K^T \left\{ \frac{\partial e_f}{\partial x_a} + \frac{\partial e_g}{\partial x_b} \right\} \right> \]

Adjoint sensitivity gradients in model grid-point space

Observations assimilated

Adjoint of assimilation

\[ K^T = \left[ H P_b H^T + R \right]^{-1} H P_b \]

• Forecasts are made with HYCOM/NCOM and NCODA 3D-Var/4D-Var

• Adjoint versions of HYCOM/NCOM and NCODA 3D-Var/4D-Var are used to calculate the observation impact

• The impact of observation subsets (observing systems, platforms, analysis variables, geographic regions) are easily quantified
Observation impact interpretation -

For any observation / innovation assimilated, if ...

$$\delta e^g_f < 0.0 \quad \text{the observation is BENEFICIAL}$$

forecast errors decrease

- the effect of the observation is to make the error of the forecast started from $x_a$ less than the error of the forecast started from $x_b$

$$\delta e^g_f > 0.0 \quad \text{the observation is NON-BENEFICIAL}$$

forecast errors increase
Identify observation impacts -

• Non-beneficial impacts:
  - not expected, all observations should decrease forecast error
  - look for problems in data QC, instrument accuracy, model biases, specification of assimilation error statistics

• Beneficial impacts:
  - associated with observations more accurate than the background in regions where adjoint sensitivity gradients are large
  - extreme beneficial impacts from isolated observations indicate the need for greater observation density

• Best Outcome:
  - many observations that produce small to moderate impacts, not few observations that produce large impacts
Variational twin data experiments with the linearized and adjoint models of SWAN

Computational Domain (Santa Rosa Island, FL)

**TA1: Tri-Axis Buoy #1**  
(input data location)

**TA2: Tri-Axis Buoy #2**

**SAB: Sentry ADCP**

**SIB: Iridium ADCP**
Ice Modeling Assimilation from Satellites

Arctic Cap Nowcast/Forecast System (ACNFS)

- Couples HYbrid Coordinate Ocean Model (HYCOM) and Los Alamos Community Ice CodE (CICE)
- $1/12^\circ$ (3.5 km at pole) resolution
- Ocean/ice data assimilated via Navy Coupled Ocean Data Assimilation (NCODA)

ARCc0.08-03.0 Ice Concentration: 20100210

Black line is the independent ice edge location (NIC)
This system will:
• Provide a tidal prediction system that automatically computes high-resolution tides along the open boundaries of NCOM.
• Operate as a stand-alone tidal prediction system.
• Operate within COAMPS (ESPC).
Small Scale Prediction: OTIS Ver.1

Validation of smaller domains

### M2 Tidal RMS Errors for Incheon, Korea

<table>
<thead>
<tr>
<th>Tidal Solution</th>
<th>Resolution</th>
<th>Relative to TIL Data (19 points)</th>
<th>Relative to IHO Data (26 points)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Amp (m) Ph (deg) Amp (m) Ph (deg)</td>
<td></td>
</tr>
<tr>
<td>Local OTIS Sol</td>
<td>1 km</td>
<td>0.018 0.6</td>
<td>0.244 9.5</td>
</tr>
<tr>
<td>Local PCTides Sol</td>
<td>1 km</td>
<td>0.181 4.3</td>
<td>0.492 8.1</td>
</tr>
<tr>
<td>OSU Yellow Sea DB</td>
<td>1/12 deg</td>
<td>0.050 2.4</td>
<td>1.624 81.9</td>
</tr>
<tr>
<td>OSU Global DB</td>
<td>1/4 deg</td>
<td>0.128 6.9</td>
<td>1.391 20.3</td>
</tr>
<tr>
<td>FES 2004 Global DB</td>
<td>1/8 deg</td>
<td>0.091 6.0</td>
<td>0.409 8.0</td>
</tr>
<tr>
<td>NCOM with BCs from Local OTIS</td>
<td>1 km</td>
<td>0.119 4.8</td>
<td>0.557 25.8</td>
</tr>
<tr>
<td>NCOM with BCs from Local PCTides</td>
<td>1 km</td>
<td>0.305 10.7</td>
<td>0.726 33.5</td>
</tr>
<tr>
<td>NCOM with BCs from OSU Yellow Sea DB</td>
<td>1 km</td>
<td>0.204 8.5</td>
<td>0.635 34.1</td>
</tr>
<tr>
<td>NCOM with BCs from OSU Global DB</td>
<td>1 km</td>
<td>0.291 9.1</td>
<td>0.716 33.4</td>
</tr>
</tbody>
</table>