GMAO Ocean Data Assimilation

Michele Rienecker
Christian Keppenne, Robin Kovach, Shu-Chih Yang, Jossy Jacob
David Adamec
Watson Gregg, Lars Nerger

Global Modeling and Assimilation Office (GMAO)
NASA/Goddard Space Flight Center

JCSDA Workshop
Application of Remotely Sensed Observations in Data Assimilation
August 1, 2007
Overall Goals:
• Improve ocean analyses for initialization of coupled forecasts.
• Ocean climate variability
• Ocean color Climate Data Records (CDRs)

Methods:
• Optimal Interpolation (univariate)
• Ensemble Kalman Filter (multi-variate, state-dependent; satellite altimetry)
• Bred-vectors to capture dominant growing modes of error
• SEIK filter (Ocean Color)

Model:
• Poseidon V4 and V5 - quasi-isopycnal model (Paul Schopf)
• MOM4 (GFDL)
Ocean in situ observations
TAO/TRITON/PIRATA moorings
+ XBTs
+ ARGO
Salinity Profiles per annum

Year: 1999  Obs: 1148
Year: 2000  Obs: 1922
Year: 2001  Obs: 5867
Year: 2002  Obs: 15798
Year: 2003  Obs: 27983
Year: 2004  Obs: 41938
Year: 2005  Obs: 20795

ARGO Float
PALACE Float
Fixed Buoy
CTD Profile at TAO Locations
Surface Height
TOPEX: August 1992-2005
JASON: December 2001---
JASON-2: June 2008

Surface Winds
SSM/I: July 1987 ---
QuikSCAT: June 1999 ---

Ocean Color
SeaWiFS: August 1997 -- MODIS: 2000---

Sea Surface Temperature
AVHRR: 1982 ---
MODIS: 2000 ---
TMI: 1997 ---
Aqua/AMSR-E: 2002 ---
Optimal Interpolation (univariate)
• Fixed Gaussian covariances: $x_s=20^\circ$, $y_s=5^\circ$, $z_s=100\text{m}$; more isotropic with increasing latitude
• Temperature ($T$) and Salinity ($S$) assimilated separately

Ensemble Kalman Filter (multi-variate, state-dependent; satellite altimetry)
• Surface height is a diagnostic - calculate $<\delta\text{SSH}, \delta T(z)>$ and $<\delta\text{SSH}, \delta S(z)>$ to “project” corrections to surface height anomalies through the water column
• Temperature data used to update salinity and currents
• Salinity data used to update temperature and currents

Observations
• Instrument error and Representation error
• Synthetic salinity used to constrain water masses

Surface Forcing
• One of the major source of errors!
• Heat, fresh water and momentum fluxes
We assimilate:
In-situ temperature profiles
In-situ salinity profiles from Argo floats
Synthetic salinity profiles from observed T(z) and climatological T-S relations
T/P and Jason-1 SSH anomalies ⇒ Bias must be accounted for when assimilating SSH

Side by side estimation of:
• Unbiased error
• Climatological error (bias)
EnKF

Propagate $t_{k-1}$ to $t_k$:

$x_k^{i-} = f(x_{k-1}^{i+}) + w_k^i$

$w = \text{model error}$

Update at $t_k$:

$x_k^{i+} = x_k^{i-} + K_k(y_k^{i} - x_k^{i-})$

for each ensemble member $i=1\ldots N$

$K_k = P_k (P_k + R_k)^{-1}$

with $P_k$ computed from ensemble spread

$x_k^i$ state vector (T, S, u,v, SSH)

$P_k$ state error covariance

$R_k$ observation error covariance

Integrate ensemble of states and compute sample covariance $P$
**Compactly supported EnKF (bias estimation omitted)**

\[ x_{i,k}^f = M(x_{i,k-1}^a, f_{k-1}) + N_{i,k-1}, \quad E(N_{i,k-1}N_{i,k-1}^T) \approx Q_{k-1}, \quad i = 1, \ldots, n, \quad (1a) \]

\[ S = \{s_1, s_2, \ldots, s_n\} = \left\{H(\Phi(x_1^f - \bar{x}^f)), H(\Phi(x_2^f - \bar{x}^f)), \ldots, H(\Phi(x_n^f - \bar{x}^f))\right\} \quad (1b) \]

\[ HP^f H^T = \frac{1}{n-1} SS^T, \quad (1c) \]

\[ a_i = [C \cdot (HP^f H^T + R)]^{-1} (y + e_i - H(x_i^f)), \quad E(e_i e_i^T) = R, \quad i = 1, \ldots, n, \quad (1d) \]

\[ x_{i,i}^a = x_{i,i}^f + \frac{1}{n-1} \sum_{j=1}^{n} (\Phi(x_{j,i}^f - \bar{x}_i^f))s_j^T (c_i \cdot a_i), \quad i = 1, \ldots, n. \quad (1e) \]

Compensating for the effects of small ensemble size:

**Φ:** smoothing operator for small scales

**C:** Compact support operator (Schur product) from Gaspari and Cohn (1985)

Variance inflation to avoid filter collapse
Ocean state-dependent covariances with the EnKF

Temporal evolution of Kalman gain for T obs.

EnKF-33: filter
Schur(C,P) @(0N, 156E, 150m)

Christian Keppenne
Ocean climate for June 2007 along the equatorial Pacific

OI - XBT

ENKF

NCEP’s GODAS
Ocean climate for June 2007 along 155°W

OI - XBTT

ENKF

NCEP’s GODAS
Independent Validation
RMSD of analysis c.f.
TAO servicing cruise CTDs
1994-1998
Independent Validation
RMSD of analysis c.f.
TAO ADCP zonal currents
1993-2006
Temperature-Salinity Diagrams (Density Contours in kg/m$^3$)

Central Indian Ocean: 70E-75E, 10S-15S (Black Dots are Argo T-S Values)
Temperature-Salinity Diagrams (Density Contours in kg/m$^3$)
Equatorial Indian Ocean: 70E-75E, 2.5S-2.5N (Black Dots are Argo T-S Values)
Forecast skill (ACC) from CGCMv1
Heat content anomaly in upper 300m
1993-2006

EnKF

1-month lead

3-month lead

6-month lead

OI-TS

Heat content anomaly in upper 300m
1993-2006
Forecast skill (ACC) from CGCMv1
SST anomaly
1993-2006

EnKF

OI-TS

1-month lead

3-month lead

6-month lead
The impact of Argo - preparing for Aquarius

Christian Keppenne and Robin Kovach

February 2006 Surface Fields: Salinity

EnKF E011 T, S, SSH

EnKF E015 No Argo S

Argo Obs z < 5m

Levitus Climatology
Augmenting covariance estimates with information from bred vectors

Shu-Chih Yang
Christian Keppenne, Eugenia Kalnay
Coupled breeding technique is designed to capture the growing errors related to slow-varying coupled instabilities, like ENSO-related growing errors.

Breeding is a nonlinear approach and tightly related to the Ensemble Kalman Filter.

Coupled breeding is implemented in the NASA/GMAO coupled general circulation model (CGCM). The applications of bred vectors (BVs) are explored for the purpose of improving coupled forecasting: 

- use BVs as the initial ensemble perturbations of the ensemble forecast system for ENSO prediction
- Augment the background error covariances in ocean data assimilation system with the structure of BVs.
Breeding in the GMAO coupled GCM

- **NASA/GMAO coupled GCM** *(Poseidon+ NSIPP-1 AGCM)*
- **Bred vectors**: Differences between the control forecast and perturbed run
- Coupled breeding cycle needs to choose physically meaningful breeding parameters in order to choose the type of instability
4 different rescaling norms are chosen to measure the coupled atmosphere-ocean instability (10% of Climate variability, rescale every month)

1. \( |\text{SST}_{BV}| = 0.1^\circ C \) (in 150°W~90°W, 5°S~5°N)
2. \( |\text{D20}_{BV}| = 1.5 \text{ m} \) (in 160°E~140°W, 2.5°S~2.5°N)
3. \( |[u'_{BV}, h'_{BV}]| = 6.5 \times 10^{-3} \) (in 130°E-80°W, 5°S~5°N)
   >> the first 4 long wave modes (Kelvin+3 Rossby waves)
4. \( |[u_{BV}\tau_{xc} + u_c\tau_{x_BV}]| = 0.1 \) (in 130°E-80°W, 5°S~5°N)
   >> work done on the ocean by the atmosphere (Goddard and Philander, 1999)

Initial conditions for CGCM:
- Ocean analysis (\( T, S \) assimilated with optimal interpolation scheme)
  + AMIP restart
- 4 pairs of ± coupled BVs are centered at this initial condition
Dominant growing modes from BVs in Pacific
Dominant growing modes from EnKF in Pacific
Ensemble forecasts initialized from 4 ±BV

pattern correlation: SSTA vs. Reynolds SSTA at 9-month lead time (1993~2002)

4 BV ensemble mean has higher skill than control

---

**BVs ensemble mean**

- Feb start
- May start
- Aug start
- Nov start

**Control**

- Feb start
- May start
- Aug start
- Nov start

---
Ensemble-based covariance in hybrid-OI scheme

\[ P_f = (1- \alpha) P_{OI} + \alpha P_f^0 \]

\( P_f \): the background error covariance
\( P_f^0 \): Ensemble-based background error covariance
\( P_{control} \): Gaussian covariance \((x_s=20^\circ, y_s=5^\circ, z_s=100\text{m})\)
\( \alpha \): the hybrid coefficient

\[ x^a - x^f = K[y - H(x^f)] = Kd \text{ (analysis increment)} \]
\[ Kd = P_f H^T [HP_f H^T + R]^{-1} d \]
\[ = (P_f^0 + P_{contr}) H^T [H(P_f^0 + P_{contr}) H^T + R]^{-1} d \]
\[ = P_f H^T [H(P_f^0 + P_{contr}) H^T + R]^{-1} d + P_{contr} H^T [H(P_f^0 + P_{contr}) H^T + R]^{-1} d \]

\( d \): the difference between forecast and observations (innovation vector)
\( \alpha = 0 \): Fully \( P_{control} \), approximate to Univariate OI
Assimilation experiment setup

<table>
<thead>
<tr>
<th>Observations</th>
<th>TAO, XBT, ARGO, Pirata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assimilation interval</td>
<td>4-day (Jan2002 ~ Dec2002)</td>
</tr>
<tr>
<td>Covariance localization for $P_f^0$</td>
<td>$x_s=8^\circ, y_s=4^\circ, z_s=100m$</td>
</tr>
<tr>
<td>Gaussian horizontal filter for $P_f^0$</td>
<td>$x_f=4^\circ, y_f=2^\circ$</td>
</tr>
<tr>
<td>Background error</td>
<td>$\sigma_T=0.7^\circ C, \sigma_S=0.1psu$</td>
</tr>
</tbody>
</table>

Experiments:

1. only the Gaussian function (control)
   - used as the benchmark
2. $P_f$ is based on 4 EOF modes
   - EOFs are constructed from long and large ensemble runs
3. $P_f$ is based on 4 BVs (updated every month)
4. $P_f$ is based on 4 BVs (updated every 4 days by linear interpolation)
Temperature correlation of the location at (156°E, EQ, 150m)

- **control (UOI)**
- **4EOF**
- **4BVs**
- **4BVs**

Fixed in time

Warm event

Cold event

Normalized Error covariance structure
Annual Mean 2002 structure in the equatorial Pacific

Temperature

Salinity

Zonal current
Comparisons with independent observations

Temp. observations from Global Temperature Salinity Profile Program

[GTSPPP T profiles – monthly analysis] in Niño3 region

control  4BVs_4day  4EOF

- Both the 4BV_4day and 4EOFs runs show improvement over the Control.
- The 4BV_4day run has positive impact on (i) summer season and (ii) the upper ocean of Nov&Dec.
RMS of Temp./Salin OMF in Pacific

E01: 4-day BVs
E03: monthly tendency BVs
Ensemble forecasts initialized from 4 coupled ±BVs have increased skill when starting from cold phase of the annual cycle.

Augmenting the Gaussian background error covariance by 4BVs (a hybrid system) has positive impact when assimilating real T and S observations.

The optimal hybrid weighting is 30-40% of the total background error covariance.

Overall, between the two hybrid experiments, the one with the BVs applicable at the analysis time (BVs_4day) generates the better T and S analyses.

- For T, the improvement over the control is seen in the tropical Pacific.
- For S, the improvement is mainly located in the western Pacific during late spring to summer season.
- BVs_4day carries the error structures most dynamically relevant to the slowly growing mode.
Ocean Color Assimilation
Watson Gregg and Lars Nerger
Ocean Color Data Assimilation complete, products available GES-DISC Giovanni (http://reason.gsfc.nasa.gov)

Goal: Consistent (climate) products from CZCS - MODIS

http://gmao.gsfc.nasa.gov/research/oceanbiology/
Constraining a Global Three-Dimensional Ocean Biogeochemical Model by SeaWiFS Ocean Chlorophyll Data Using a Local SEIK Filter

Lars Nerger, Watson W. Gregg

Smaller error than SeaWiFS

Nerger and Gregg, 2007
J. Mar. Syst. (submitted)
Sampling biases in MODIS ocean chlorophyll were determined by “flying” the MODIS daily sampling over the complete daily coverage provided by data assimilation. The results showed that MODIS annual mean chlorophyll estimates are about 8% too high. Considering that the maximum interannual variability in the 10-year SeaWiFS record is about 3%, this sampling bias should be considered when making estimates of global chlorophyll.
Some References:


