Ensemble Assimilation of Ocean Data into the GEOS-5 Coupled GCM

Christian Keppenne\textsuperscript{1,2,*}, Guillaume Vernieres\textsuperscript{1,2}, Michele Rienecker\textsuperscript{1}, Robin Kovach\textsuperscript{1,2}, Jossy Jacob\textsuperscript{1,2} and Atanas Trayanov\textsuperscript{1,2}

\textsuperscript{1}NASA GMAO
\textsuperscript{2}SAIC

Outline:

• NASA GMAO coupled model
• Coupled ensemble assimilation with GMAO ODAS-2
  • Atmospheric analysis “replay” procedure
  • Augmented ocean ensemble Kalman filter
    • Adaptive observation errors
    • Adaptive background-error covariance localization and inflation/deflation
    • Hybrid particle filter
    • Online bias correction
• System validation
  • Assimilation of sea level height
    • Online bias correction
    • Multivariate projection method
  • Assimilation of in situ T and/or S
• Outlook

*contact: christian.keppenne@nasa.gov

June 15, 2010
Atmospheric Observing System

GEOS-5 ADAS  14 May 2008 00UTC
1,557,926 observations - 90% from satellites

The atmospheric observing system today... a 6-hr snapshot (courtesy of Ron Gelaro, GMAO)
Ocean Observing System

ODAS-2 data
- Topex/Jason SSH anomalies
- Argo in situ T and S profiles
- In situ T from TAO, XBT, Pirata and Rama
- Reynolds SST
- Levitus surface salinity while waiting for Aquarius

Historical availability of in situ data

The density and vertical coverage of in situ data has increased tremendously but the ocean is still poorly observed vs. the atmosphere. Hence, assimilating surface measurements from remote sensing is a must.
**ODAS-1**

- Ocean-only runs
- **OGCM**: Poseidon 4
- Analysis algorithms
  - EnKF
  - MvOI (EnKF analysis with steady-state fixed ensemble)
  - UOI (functional univariate background covariances)

**ODAS-2**

- **GEOS-5 Coupled Model**:
  - **OGCM**: MOM-4 (0.5°X 0.167-0.5°X 40L) or any other ESMF-ready model
  - **AGCM**: GEOS-5 AGCM (1.25°X 1°X 72L)
- Analysis algorithms
  - Atmosphere: “replay” of GMAO atmospheric analysis
  - Ocean: “Augmented” hybrid EnKF/lagged EnKF/particle filter approach

- ODAS implemented as ESMF gridded-component -> model independent
GEOS-5 coupled model and coupled ensemble system

GEOSGCM1

AGCM

OGCM

GEOSGCM2

GEOSGCM3

GEOSGCM4

Etc...

GEOSGCM current configuration:
- OGCM: MOM4p1 720x410x40 grid
- AGCM: GEOS5 AGCM 288x181x72 grid

(Each subsystem implemented as ESMF gridded component)
Atmospheric analysis replay

1: AGCM forecast (F)

4: rewind AGCM

5: incremental analysis update (IAU)

2: read atm. analysis (A)

3: calculate atm. increment (A-F)

03z 06z 09z

time

ocean analysis

ocean IAU

ocean analysis
Augmented EnKF

The data assimilation problem

\[ \frac{dx}{dt} = M(x, f) + q \]
\[ E((x - x_t)(x - x_t)^T) = P \]
\[ E(qq^T) = Q \]
\[ y = H(x_t) + r \]
\[ E(rr^T) = R \]

\( x \): model state vector
\( x_t \): unknown true state
\( y \): measurements

Objective: Find the best possible estimate of \( x_t \) given \( x, y \) and their error distributions

The Kalman Filter (Kalman 1960)

\[ \frac{dP}{dt} = \frac{d}{dt} \left[ E((x - x_t)(x - x_t)^T) \right] = \frac{dM}{dx} P \left[ \frac{dM}{dx} \right]^T + Q \]
\[ x^a = x^f + PH^T (HPH^T + R)^{-1} (y - H(x^f)) \]
The ensemble Kalman Filter

Evensen (1994, 1996)
Replace background-covariance evolution with ensemble integration

\[
\frac{dx_i}{dt} = M(x_i, f) + q_i \quad \quad E((x - x_t)(x - x_t)^T) = P
\]

\[
P \approx \frac{1}{n-1} \sum_i (x_i - \bar{x})(x_i - \bar{x})^T
\]

given \( z_i = H(x_i - \bar{x}) \), \( i = 1, \ldots, n \), \( Z = \frac{1}{\sqrt{n-1}}[z_i] \), \( X = \frac{1}{\sqrt{n-1}}[x_i - \bar{x}] \),

the update for ensemble member \( x_i \) is computed as
(from right to left -> only matrix-vector products):

\[
x_i^a = x_i^f + XZ^T(ZZ^T + R)^{-1}(y - Hx_i + \epsilon_i)
\]

\[
P \rightarrow C_p \circ P
\]

\[
R \rightarrow C_r \circ R
\]

\[
XZ^T(ZZ^T + R)^{-1} \rightarrow C_p \circ XZ^T(C_p \circ ZZ^T + C_r \circ R)^{-1}
\]

\( \circ \equiv \text{Hadamard (Schur) product} \)
3 Sources of background-error covariance information

\[ P^f = P_{\text{dyn}}^f + P_{\text{stat}}^f + P_{\text{func}}^f \]

- **\(P_{\text{dyn}}\): State-dependent error-covariance basis vectors from ensemble integration**
  - Current state of each ensemble member minus low pass filter
  - Past states of each ensemble member minus a low pass filter

- **\(P_{\text{stat}}\): Static ensemble of time-independent “error EOFs”**
  Error EOFs calculated from a time series of differences between a coupled model run constrained by replaying the GMAO atmospheric analysis and unconstrained short-term forecasts

- **\(P_{\text{func}}\): Pseudo-Gaussian univariate covariance term**
• Static, not flow adaptive 3D localization along \((x, y, z)\) space dimensions
• Also apply Gaussian filter to deviations from ensemble mean \(x_i - \bar{x}, \ i = 1 \ldots n\)

Marginal Kalman gain: \(T\) obs @\((0n, 156E, 150m)\) on 12/31/01
horizontal section through \(\langle T', T'\rangle\) covariances

- **EnKF-9**
  - Unfiltered, not compactly supported
  - 0.36

- **EnKF-17**
  - Unfiltered, compactly supported
  - 0.51

- **EnKF-33**
  - Filtered, compactly supported
  - 0.63

- **EnKF-65**
  - Filtered, compactly supported
  - 0.77
• Flow adaptive error-covariance localization following neutral density \([(x, y, z, \rho)\) dimensions]  
• Adaptive optimization of error-covariance localization scales \((x, y, z)\) used with each observation  
• Adaptive estimation of representation error associated with each observation  
• Adaptive background-error covariance inflation/deflation  
• Adaptive rescaling of analysis increments  
• Particle pre-filter
1. **Traditional approach (as in ODAS-1)**

   \[ C(\delta x, \delta y, \delta z, \delta t) \text{ is an approximately Gaussian compactly supported correlation function} \]

   \[ P_c = P \circ C \]

2. **Tried hierarchical ensemble filter (Anderson 2007)**

   • Observations must be processed serially (\( \alpha_{kl} P_{kl} \) is not a covariance)

   \[ \alpha = \frac{1}{m-1} \left( \frac{\sum_{i=1}^{m} \beta_i^2}{\sum_{i=1}^{m} \beta_i^2} - 1 \right) \]

3. **Bishop's (2007) flow adaptive moderation of spurious covariances**

   • Some long-range spurious features are amplified.

   • Assimilation performance (OMFA statistics) worse than case 1

   \[ c_{ij}^m = \left( \frac{P_{ij}}{\sqrt{P_{ii} P_{jj}}} \right)^m \]

   \[ G = \text{diag} (C^m), \quad C^{mq} = G^{-1/2} (C^m)^q G^{-1/2} \]

4. **Back to approach 1 with localization in (x, y, z, t, neutral density) space**

   • Respects flow-dependent gradients such as thermocline and fronts

   • Adaptive optimization of localization scales involved in processing each observation

   • Assimilation performance better than case 1
Covariance localization is the most numerically intensive part of the ensemble assimilation system

\[
P \rightarrow C_p \circ P, \quad R \rightarrow C_r \circ R, \quad C = \begin{bmatrix} c_{ij} \end{bmatrix},
\]

\[
c_{ij} = c_0 \left( 2 \frac{|x_i - x_j|}{l_i^x + l_j^x} \right) c_0 \left( 2 \frac{|y_i - y_j|}{l_i^y + l_j^y} \right) c_0 \left( 2 \frac{|z_i - z_j|}{l_i^z + l_j^z} \right) c_0 \left( 2 \frac{|\rho_i - \rho_j|}{l_i^\rho + l_j^\rho} \right) c_0 \left( \frac{|t_i - t_j|}{l_t^t} \right)
\]

\(c_0\) is a compactly supported analytical covariance function (Gaspari and Cohn 1985)

ODAS-1: \(l_x(y)\) and \(l_y(y)\) proportional to Rossby radius of deformation

ODAS-2: \(l_x(x,y,z,t), l_y(x,y,z,t), l_z(x,y,z,t)\) & \(l_p(x,y,z,t)\) optimized iteratively for each datum
Marginal Kalman gain:
unit T innovation at 95m

ODAS-2 flow-dependent error covariances

Marginal Kalman gain:
unit SSH innovation along equator
For each observation $y_0$, process neighboring observations as though they were perfect (R=0) and optimize the localization by iteratively solving for the $l_x$, $l_y$ & $l_z$ that minimize

$$
|y_0 - H_0 C \circ P H_n^T \left(H_n C \circ P H_n^T\right)^{-1} (y_n - H_n x)|
$$

- $y_0$: an observation
- $y_n$: set of neighboring observations of same variable excluding $y_0$
- $H_n$: maps the state vector to $y_n$
- $H_0$: maps the state vector to $y_0$

**Example:**
optimized $l_x$ and $l_y$ localization scales for Reynolds SST data on Jan. 1 2007 expressed as a fraction of the default Rossby-radius dependent localization
For each individual observation, after optimization of the covariance localization parameters $l_x$, $l_y$, & $l_z$, the representation error is estimated as

$$
\sigma_0 = \left| y_0 - H_0 C \circ P H_n^T \left( H_n C \circ P H_n^T \right)^{-1} \left( y_n - H_n x \right) \right|
$$

Estimated representation error for Reynolds SST data Jan. 1 2007

Difference in SST increment:

adaptive (errors + localization + covariance inflation) - standard assimilation (adaptive inflation only)
ODAS-2 adaptive localization and representation-error estimation

Example for one ARGO T profile at (16S, 0W) on Jan. 1, 2007

- Optimal horizontal scales: ~60% of Rossby-radius dependent scales @250m, larger @1000m
- Optimal vertical localization scales: minimum in thermocline. Default (250m) is too short near 1000m
- Representation error estimate ($\sigma_{obs}$): maximum in thermocline, very small below 1000m
Following Desroziers et al. we have:

\[
E[(y - Hx^f)(y - Hx^f)^T] = Tr(HP^f H^T + R)
\]
\[
E[(y - Hx^f)(H(x^a - x^f))^T] = Tr(HPH^T)
\]

Iterate until global convergence is satisfied:

Not prohibitively expensive because does not require calculation of \( C \cdot HPH^T \)

\[
P \rightarrow \alpha P
\]

\[
\alpha = \sum \left[ (y_i - H_i x^f) H_i (x^a - x^f) \right] P
\]

\( H_i(x) \): observation operator (e.g., interpolation) for observation \( i \) (scalar)

**Assimilation increment rescaling**

Parallel algorithm involves each CPU minimizing RMS analysis error variance for a subset of all the observations (all the observations that influence state variables pertaining to that CPU). The increment, \( \Delta \), is then optimized globally by rescaling it \( (\Delta \rightarrow \gamma \Delta) \) such as to globally minimize

\[
f(\gamma) = \sum_i (y_i - H_i x^a)^2 = \sum_i (y_i - H_i (x^f + \gamma \Delta))^2
\]

\[
\frac{d}{d\gamma} f(\gamma) = 0 \quad \Rightarrow \quad \gamma = \frac{\sum (y_i - H_i x^f) H_i \Delta}{\sum (H_i \Delta)^2}
\]
Motivation: ensemble mean is not necessarily a realizable state. Hence we want to improve upon this state by shifting the ensemble mean to the ensemble member that is closest to the observations (a realizable state).

- Find ensemble member $x_p$ that is closest to the data in terms of RMS OMF
- Displace the whole ensemble by an increment $\Delta_p = x_p - x_m$ where $x_m$ is the ensemble mean
- Thereafter, apply the ensemble Kalman filter analysis
ODAS-2 particle pre-filter example: assimilate in situ ARGO T data. Validate against ARGO S data

- **CGCM**
- **Data**
  - Daily assimilation of ARGO T profiles 04/01/06 – 05/31/06 (active data set)
  - ARGO S profiles used for validation (passive data set)

- **Initial condition**
  - 03/01/06 coupled model restart from single coupled model run with atm. Anal. Replay

- **Ensemble initialization (03/01/06 – 04/01/06)**
  - initial perturbation from linear combinations of model signal EOFs
  - daily perturbations with 1% of initial perturbation amplitude

- **Assimilation (04/01/06–05/31/06)**
  - **CE-16**: 16-member control ensemble – no assimilation
  - **EnKF-16×11**: 16 streams (model integrations) and 10 past instances in each stream (lag = 1 day)
  - **HPEnKF-16×11**: reordering particle pre-filter HPF-16 used prior to each EnKF-16×11 analysis
ODAS-2 particle pre-filter example: assimilate in situ ARGO T data. Validate against ARGO S data.

Salinity improvement over control ensemble
Warm (resp. cold) colors denote areas where the analysis is closer to (resp. further away from) the passive S ARGO data than the control ensemble in May 2006 (last month of exp).

Global salt OMFA statistics:
- mean OMF Better than control below 200m
- RMS OMF Better overall
- mean OMA
- RMS OMA
Online bias correction and assimilation of SSH anomalies

- **Challenge 1**: model bias changes as the data are assimilated
- **Challenge 2**: must derive \( T(z) \), \( S(z) \) \( u(z) \) and \( v(z) \) from scalar \( \eta \) measurements

\[
\eta = \int f(\rho(z)) dz
\]

\[
P = P^f + P^b
\]

\[
b^a = b^f - P^b H^T (H(P^b + P^f)H^T + R)^{-1}(y - H(x^f - b^f))
\]

\[
x^a = x^f + P^f H^T (H P^f H^T + R)^{-1}(y - H(x^f - b^a))
\]

\[
b_{k+1}^f = b_k^a
\]

\( y - H(x) \): total innovation
\( y - H(x - b) \): unbiased innovation

\[
P^b \rightarrow C^b \circ P^b
\]

\[
P^f \rightarrow C^f \circ P^f
\]

**SSH bias estimate**

**snapshot 04/01/2006**
Online bias correction and assimilation of SSH anomalies

Experiment duration 01-07 2007
RMS T OMFA statistics at TAO moorings
TAO T data are passive
SLA is active

Note: ensemble initialization during first two months of EnKF run
Validation of surface data assimilation using passive (not assimilated) sub-surface Argo data

T improvement over control: control RMS T OMF - ODAS RMS T OMF

- Assimilation of SST + SSS alone does not improve the subsurface T much (vs. control)
- SLA assimilation with online bias correction improves upon control, but not in Nino-4 area (0-300m)
- Assimilating SST + SSS + SLA mostly corrects the 0-300m Nino-4 area deficiencies
Validation of surface data assimilation using passive (not assimilated) sub-surface Argo data

- SLA assimilation alone very effective is improving $S$ over the control
- Best results for $S$ seen when assimilating SST + SSS + SLA
Online bias correction and assimilation of SSH anomalies

T and S forecast and analysis compared to some un-assimilated in situ profiles near the altimeter track at the time of the first assimilation.
Summary

- Ocean data assimilation into GMAO CGCM with “replay” of the GMAO atmospheric analysis
- Combining static and dynamic ensembles (including lagged ensemble) gives best performance
- Multivariate background covariances effective in improving unobserved model variables
- SLA assimilation improves subsurface T & S, but best results with SST + SSS + SLA assim.
- Ensemble data assimilation system ready for production runs
- Started 1950-present retrospective analysis

Outlook

- Moving towards fully coupled data assimilation system through data assimilation into the skin layer (building upon NCEP GSI work)
- Ready for new data types, starting with Aquarius

GMAO ODAS webpage: http://gmao.gsfc.nasa.gov/research/oceanassim