Data Assimilation of the Global Ocean

Steve Penny

A comparison of data assimilation using 4D-LETKF and SODA.
Advisors

- Eugenia Kalnay
- Jim Carton
- Brian Hunt
- Kayo Ide
- Takemasa Miyoshi
- Gennady Chepurin
Summary

- Introduction of Model, Observation Network, and Assimilation Schemes
- Modifications to LETKF for Ocean domain
- Comparison of results from LETKF and SODA reanalyses
Model

- Modular Ocean Model (MOM2)
- Designed at Geophysical Fluid Dynamics Laboratory (GFDL)
- Finite difference version of primitive equations governing ocean circulation
- Hydrostatic, Boussinesq and rigid lid approximations
- No poles (latitude range from -60° to 60°)
- 360 x 130 x 20, higher resolution near equator.
Model IC's and BC's

- Levitus Climatological Temperature and Salinity as initial conditions
- Sponge layer at northern and southern boundaries
- Interpolated monthly mean surface boundary conditions (NCEP reanalysis)
- Spinup from 1970-1975, historical perturbations taken from 1975-1989
Model Bias

(Obs – Background) for initial ensemble at Jan 1, 1997
Observation Network

- Quality control applied to raw data
- Observations are grouped in 1x1 degree 5-day bins for SODA and 1-day bins for LETKF
- SODA uses 90-day rolling window of obs, LETKF uses a 55-day rolling window.
- Gaussian weighting applied temporally to obs
- World Ocean Atlas XBT and CTD Temperature and Salinity data (no SST)
- Error profiles estimated from SODA analysis
Observation Error Profiles

Temperature error (°C)

Salinity error (°%)

100 m
0.84

500 m

900 m

0.5

0.75

0.7

0.5

1.0
Observation Coverage

Example locations of vertical profiles for a single day
Observation Coverage

Example locations of vertical profiles over 55-day window
LETKF

- “Local Ensemble Transform Kalman Filter”
- Miyoshi's LETKF code adapted for ocean
- Ensemble square root filter method
- The background covariance is calculated from the ensemble spread
- Analysis is performed locally for each grid point
- Assimilates Temp & Salt, modifies U and V velocities
Modifications for Ocean Analysis

- Incremental analysis update
- Perturbations to surface wind forcing fields
- Land/Sea differentiation
- Pre-processing of observations
- Adaptive inflation
- Inflation relaxation to prevent runaway inflation
- Obs error scaling for pre-analysis inflation
- Latitude-dependent localization radius
LETKF Experiment Parameters

- 40-member ensemble
- Wind perturbations ($\alpha_w = 0.1$)
- Initial ensemble perturbation ($\alpha_i = 0.5$)
- Adaptive inflation ($\sigma_b = 0.001$)
- Inflation relaxation where temp spread > 2º C
- Obs error scaling of 0.625
- Quality control on obs data over (10x obs error) distance from mean background value
SODA

- "Simple Ocean Data Assimilation"
- Uses modified Optimal Interpolation
- Background covariance is flow-dependent, but computed a priori and assumed to be steady in time
- Assimilates temperature and salinity
Comparison of LETKF and SODA

- RMS Errors between observations and background, and between observations and analysis
- Temperature and Salinity
- Globally, by region, by vertical level
- 2-year experiment (97-98 el niño) and 3-year experiment (2001-2003 introduction of argo)
SODA Analysis Cycle

Analysis Cycle n
- t=1
  - Background
  - MOM2
  - SODA
  - Incremental Update
  - Background

Observation window (+/- 45 days)

Analysis Cycle n+1
- t=10
  - Background
  - Analysis
4D-LETKF Analysis Cycle

Analysis Cycle n

- t=1
  - Background Ensemble
  - MOM2
  - 4D-LETKF
  - Incremental Update
  - Background Ensemble

Observation window (+/- 25 days)

Analysis Cycle n+1

- t=5 days
  - Background Ensemble
  - MOM2
  - Analysis Ensemble
Daily Observation Counts

Counts are given at LETKF analysis times

1997-1998
Global RMS Error (Temperature)

1997-1998

[Graph showing RMS Error (Degrees C) over dates from 01.01.97 to 01.01.99, with different lines representing SODA Background, SODA Analysis, LETKF Background - rescaled w/ cov limit, LETKF Analysis - rescaled w/ cov limit, and Obs Count / 4000.]
Global RMS Error (Salinity)

1997-1998

- SODA Background
- SODA Analysis
- LETKF Background
- LETKF Analysis

RMS Error (ppt)
Regional RMS Error (Temp)

1997-1998

North Pacific

Equatorial Pacific

Indian Ocean

South Pacific

North Atlantic

Equatorial Atlantic

South Atlantic
Regional RMS Error (Salt)

1997-1998

North Pacific

North Atlantic

Equatorial Pacific

Equatorial Atlantic

Indian Ocean

South Pacific

South Atlantic

No Data
RMS Error by Depth

7.5 meters

97.5 meters

190.61 meters

1099.46 meters

Temperature
Salinity
97-98 El Niño

Sea Surface Temperatures over 25° and 28° C, as analyzed by NCEP and LETKF

December 1997

El Niño

December 1998

La Niña

NCEP graphic source:
http://www.knmi.nl/~oldenbor/publ/bien/
Adaptive Inflation

Dec. 14, 1997
97.5 meters

Shaded from 1 to 2 (0-100% inflation), contoured from 1 and above
Adaptive Inflation

Dec. 14, 1997

443.79 meters

Shaded from 1 to 2 (0-100% inflation), contoured from 1 and above
Daily Observation Count

Counts are given at LETKF analysis times

2001-2003

Nearly 5x increase in daily salinity observation counts from 2001 to 2004

Counts are given at LETKF analysis times
Global RMS Error (Temperature)
Global RMS Error (Salinity)

2001-2003
Regional Observation Counts

2001-2003

North Pacific

Equatorial Atlantic

Indian Ocean

Equatorial Pacific

North Atlantic

South Pacific

South Atlantic
Regional RMS Error (Temp)

2001-2003

North Pacific

North Atlantic

Equatorial Pacific

Equatorial Atlantic

Indian Ocean

South Pacific

South Atlantic
Regional RMS Error (Salt)

2001-2003

North Pacific

North Atlantic

Equatorial Pacific

Equatorial Atlantic

Indian Ocean

South Pacific

South Atlantic
Adaptive Inflation

Shaded from 1 to 2 (0-100% inflation), contoured from 1 and above

Jan. 3, 2002

7.5 meters (SFC)
Adaptive Inflation

Jan. 3, 2002  97.5 meters

Shaded from 1 to 2 (0-100% inflation), contoured from 1 and above
Adaptive Inflation

Jan. 3, 2002  443.79 meters

Shaded from 1 to 2 (0-100% inflation), contoured from 1 and above
Conclusion

- LETKF performs at least as good as SODA in most areas, often better.
- LETKF benefits from larger observation coverage.
- Modifications to the atmospheric-oriented LETKF were necessary to attain useful results in the oceanic domain.
LETKF in one line
(at each grid point)

\[ x^{a(i)} = X^b \left[ \left( (k-1) \left( \frac{1}{p} + Y^b R^{-1} Y^b \right)^{-1} \right)^2 + \left( (k-1) \frac{1}{p} + Y^b R^{-1} Y^b \right)^{-1} Y^b R^{-1} \cdot (\mathbf{y}^o - \mathbf{y}^b) \cdot v \right] + x^b \cdot v \]

- \( X^b \): background perturbations from mean
- \( p \): covariance inflation
- \( Y^b \): background perturbations in obs space
- \( R \): observation covariance
- \( \mathbf{y}^o - \mathbf{y}^b \): observation departure
Initial Ensemble

Initial Ensemble members have mean equal to the base background member used in SODA.

$$z^+ = (1-\alpha)\bar{x} + \alpha y \quad z^- = (1+\alpha)\bar{x} - \alpha y$$

$$= \bar{x} - \alpha x + \alpha y \quad = \bar{x} + \alpha x - \alpha y$$

$$= \bar{x} + \alpha (y - \bar{x}) \quad = \bar{x} + \alpha (\bar{x} - y)$$

$\alpha$ can span from 0 to 1. When $a = 0$, initial ensemble is equal to baseline $\bar{x}$, with zero variance. When $\alpha = 1$, the entire ensemble is made up of historical values or $2\bar{x}$ – historical values, so that the mean of the ensemble is still the baseline $\bar{x}$. 
Surface Wind Forcing

- Random historical years selected
- Perturbation approach identical to initial ensemble generation
- Ensemble forcing =
  \[(1 \pm \alpha) \times \text{reanalysis forcing} \pm \alpha \times \text{historical forcing}\]
Observations Bordering Land
Model Coastline
Analysis at day 80

North Atlantic

Inflation Contour

Obs Increment Improvement (gridpoints)
Adaptive Inflation

- Adaptive Inflation (black contour)

North Pacific

Background spread (white contour)

Analysis Increment (shaded)
Analysis at day 60, obserr 0.5

North Pacific

Background Spread Contour

Observations – Background, (gridded data)
Analysis at day 60, oberr 0.5

North Pacific

Observation Departure Improvement, (gridded data)

Improved (closer to obs)  Got worse (farther from obs)
Analysis at day 70, obserr 0.5

Adaptive Inflation Contour
MAX Inflation = 1.032, or approximately 3%

North Pacific

Observation Departure Improvement, (gridded data)

Improved (closer to obs)
Got worse (farther from obs)
Successive Improvement

Observation Departure Improvement, (gridded data)

Improved (closer to obs)  Got worse (farther from obs)
Latest Experiments

- Use binned observation statistics (mean, variance, count) to develop better estimates of observation error.
- Use letkf 'quality control' to scale down error regime over time.
- Initial ensemble alpha = 1, wind forcing perturbation alpha = 0.3.
LETKF parameters

- 1.0 initial inflation (0%), $\sigma_b = 0.0005$
- Quality control on observation data over $10 \times \sigma_o$ from the background
- Default 0.5º observation variance for bins with 3 or fewer observations
- Possible 0.001-5.0º standard deviation on Temp obs with greater than 2 observations
Analysis Increment at t=1
Analysis Increment at t=26 (9/8)
RMS Errors (vs. retained obs)

Background

Analysis

Obs / 17k

RMS Errors for case with observations within the analysis cycle window only
Next Steps

● Localization
  – Allow to vary across variables
  – Apply Nearest Neighbors reduction
  – Customize with ocean current forecast
  – Apply graph search for land occlusion
  – Extract from LETKF code

● Observations
  – Pre-processing of observations (CGAL)
Heuristic Land Occlusion
Future – Auto Land Occlusion
Future – Auto Land Occlusion
Future – Custom Localization

Zonal Velocity (cm/s)
Balance

- Initial ensemble spread must be large to represent appropriate initial uncertainty.
- Wind forcing spread cannot be too small to force ensemble collapse or too big to create non-linearity.
- Adaptive inflation helps maintain balance between model error and observation error.
- Adaptive inflation cannot be allowed to grow too quickly if obs network is changing.