Ocean data assimilation

L. Bertino, NERSC
With inputs from Johnny Johannessen, François Counillon

ECMWF seminar 8th Sept 2014
Ensemble Kalman filtering

1. Initial uncertainty
2. Model uncertainty
3. Measurement uncertainty

Observations

Member1
Member2
......
Member99
Member100
Vocabulary and Analogies with 4D-VAR

**4D-VAR**
- Cost function (quadratic)
- Adjoint sensitivities
- 4D assimilation
- Optimal solution in linear cases

**Not analogous**
- Powerful iterative gradient descent
- Strong constraint

**EnKF**
- Posterior variance (min)
- Cross-covariances
- Asynchronous EnKF
- Optimal solution in linear cases
  - but sampling errors

**Not analogous**
- Monte-Carlo framework
- Explicit model errors
MyOcean
GMES Marine Service
1. Global Modeling and Forecasting Center
   - Lead Mercator
     • NEMO + fixed based SEEK filter

2. Arctic Modeling and Forecasting Center
   - Lead developments NERSC
   - Exploited operationally at MET Norway
   - Based on the TOPAZ system
     • HYCOM + EnKF

3. until 7, see [http://myocean.eu](http://myocean.eu)
The HYCOM model at NERSC

• 3D numerical ocean model
  – Hybrid Coordinate Ocean model, HYCOM (U. Miami)

• Hybrid vertical coordinate
  – Isopycnal in the interior
  – Z-coordinate at the surface
  – TOPAZ4 uses 28 layers

• Coupling to sea ice model
  – EVP dynamics ...
  – Semtner Thermodynamics

• Data assimilation:
  – EnKF (probabilistic) ...
The state vector $X$

- **3D variables**
  - Temperature
  - Salinity
  - Layer thickness (can be zero)
  - X-current
  - Y-current

- **2D variables**
  - Sea ice area
  - Sea ice thickness
  - Snow depths
  - Barotropic currents + pressure

- **Typical grid size**
  - Horizontal: 800x880
  - Vertical: 28

- **Total unknowns**: \(~10^8\)
  - Need to perform *local* analyses

Evensen 2002
Computations

**DEnKF 100 members**

- **Ensemble Forecast**
  - 2500 CPU hours / cycle
  - Embarrassingly parallel
  - 100x **133 CPU 11 min** jobs
  - Each job requires **400 Mb**
    - MPI parallelization

- **Analysis**
  - 20 CPU hours / update
  - 6 datasets simultaneously
  - One **20 CPU 1h job**
  - Memory required **1 Gb**
    - MPI parallelization

- **HPC Machine:**
  - Cray XE6m, updated 2012
    - 22272 cores, 205 Tflop/s
    - 676 nodes (32-cores)
    - 1-4 Gb per node
The TOPAZ system

- Exploited operationally at met.no
  - Since 2008
  - Ecosystem added in Jan. 2012
- 20 years reanalysis at NERSC
  - Took 2 years to produce
  - 3-years ecosystem reanalysis
- MyOcean (Arctic MFC)
  - Free distribution of data
  - Dynamical viewing (Godiva2)
- Data used by ECMWF wave model (J. Bidlot)
  - Sea ice edge forecast
  - Surface currents

Ice thickness forecast for 14th Aug. 2012
TOPAZ Assimilation

- DEnKF, asynchronous
  - 100 members
  - Local analysis (~90 km radius)
  - Ensemble inflation by 1%
- Observations:
  - Sea Level Anomalies (CLS)
  - SST (NOAA, then UK Met)
  - Sea Ice Concentr. (OSI-SAF)
  - Sea ice drift (CERSAT)
  - T/S profiles (Coriolis)
  - **400,000 observations** per week
  - ~100 in each local radius

SRF: local spread reduction factor

\[
SRF = \sqrt{\frac{\text{tr}(H^p H^T R^{-1})}{\text{tr}(H^a H^T R^{-1})}} - 1
\]
EnKF Correlations, SST
Why dynamic Data Assimilation in the Arctic?

Example of ice-salinity correlations in the Barents Sea

*Sakov et al.*, the TOPAZ4 system, OS 2012

Also see *Lisæter et al.* Oc. Dyn. 2003
Comparison to static / climatological covariances
Data assimilation statistics SLA

Stable / decreasing errors

Stable ensemble spread
Independent data: surface drifters

9 January 2008: SLA from TOPAZ reanalysis + drifters (± 4 days)
Data assimilation statistics SST
SST forecasts in real-time

Bias, Sea surface temperature vs. drifting buoy data, forecast day: 6

RMS, Sea surface temperature vs. drifting buoy data, forecast day: 6
In situ profiles assimilated

• A “Good period” 2003-2008
  – Argo floats
  – Sections
  – Ice-Tethered Profilers from Damocles IPY
  – All reprocessed quality controlled data
  – Not all profiles contain salinity

• Still very poor coverage compared to atmosphere
TEM biases at depths

Temperature Long-term Mean Difference
Period: 1991–2010, Depth: 100m

TOPAZ Reanalysis – WOA2013
RMSE: 0.93 °C

TOPAZ FREE – WOA2013
RMSE: 1.10 °C
TEM biases at depths

Temperature Long-term Mean Difference
Period: 1991–2010, Depth: 300m

TOPAZ Reanalysis – WOA2013
RMSE: 0.95 °C

TOPAZ FREE – WOA2013
RMSE: 1.29 °C
Data assimilation stats S100-300

Salinity Innovation Statistics 100–300m

- Innovation Bias
- Ensemble Standard Deviation
- Innovation Standard Deviation
- Number of Observations

IPY

Time:

Bias / Standard Deviation [PSU]
-0.25 -0.20 -0.15 -0.10 -0.05 0.00 0.05 0.10 0.15 0.20 0.25

Number of Observations
0 1000 2000

Reynolds SST
OSTIA Reanalysis
OSTIA NRT
OSI SAF (no repro)
AMSR–E
OSI SAF (E)
Along–track SLA
Argo floats

www.myocean.eu
Salinity bias at 100m depths

Salinity Long-term Mean Difference
Period: 1991–2010, Depth: 100m

TOPAZ Reanalysis – WOA2013
RMSE: 0.34 PSU

TOPAZ FREE – WOA2013
RMSE: 0.63 PSU

Salinity difference [PSU]
Salinity Long-term Mean Difference
Period: 1991–2010, Depth: 300m

TOPAZ Reanalysis – WOA2013
RMSE: 0.11 PSU

TOPAZ FREE – WOA2013
RMSE: 0.20 PSU
Validation of 1993-2009 reanalyses
Solfrid Hjøllo, Vidar Lien, Morten, Henning, Einar, Gilles, Francois

TASK

- Validation of 1993-2009 reanalyses, focus on vol & heat fluxes, hydrography
- Global / Arctic MFC / (ROMS)
- Monthly means, both free and assimilated runs
- Mean, std, seasonal cycle and trends
Atlantic water $T>5^\circ C$, $S>35.0$
Berx et al 2013
Nemo free:
Slightly higher salinity, temperature and speed than in assimilated run

TOPAZ free:
More saline AW core than in assimilated run, but AW depth similar
Nemo assim:
Realistic hydrography: AW core at Shetland shelf slope; sloping T and S surfaces; AW above ~500 m. Too weak currents

TOPAZ assim:
Realistic hydrography: AW core at Shetland shelf slope; sloping T and S surfaces; AW above ~500 m.
All model simulations show too low AW volume and heat transports.
Assimilation improves correlation slightly.
Seasonal variation in the Faroe–Shetland Channel

- **Obs**
- **TOPAZ FREE**
- **TOPAZ ASSIM**
- **NEMO FREE**
- **NEMO ASSIM**
Problem of AW representation

Laptev Sea mooring, 2004–2005

(a) Observations
(b) TOPAZ
(c) free run 2003–2008
(d) PHC climatology

cold temperature [°C]

depth [m]
salinity

Canadian mooring, 2006–2007

cold temperature [°C]

depth [m]
salinity
Ice concentrations climatology 1991-2010

OSI SAF

MARCH

REANALYSIS

FREE RUN

SEPTEMBER
Icea area anomalies

Arctic ice area anomalies

Free run shows a slower trend. Corrected by assimilation (as expected)
Validation of operational forecasts

Ice edge (A. Melsom met.no)

- http://myocean.met.no/ARC-MFC/V2Validation/index.html
- Weekly monitoring of forecast skills
- Error on ice edge: 50 km on average in European areas
- Larger errors in Summer
  - Expected from reanalysis

RMS, ice edge

![Graphs showing RMS distance over time](image-url)
Ice thickness validation

Independent satellite IceSAT (Kwok, JPL)

TOPAZ free run

TOPAZ pilot reanalysis

Underestimates thick ice

Overestimates thin ice

Common feature of AOMIP models (Johnson et al. JGR 2012)
Ice thickness validation
Assimilation in regional models

- We do not afford an EnKF for nested high-res models
  - Resort to a cheaper, locally-tuned EnOI
    - “Static ensemble” instead of dynamic ensemble
    - Relies on a model climatology
  - Most operational ocean data assimilation methods today are similar to an EnOI
    - Srinivasan et al. OM, 2011
- Our experience: Gulf of Mexico, South China Sea, Agulhas currents
  - Able to constrain identifiable mesoscale features
  - Also able to handle tides while assimilating Altimeter data
Agulhas current

MEAN SURFACE VELOCITY (m/s): 2008-2009

(a) ASAR
(b) FREE
(c) ASSIM_{SLA}
(d) ASSIM_{SST}
Agulhas current

MEAN SEA SURFACE TEMPERATURE (°C): 2008-2009

(a) Drifters

(b) FREE

(c) ASSIM_{SLA}

(d) ASSIM_{SST}
Seasonal-to-decadal prediction with the Norwegian Climate Prediction Model

Counillon F., Bethke I., Keenlyside N., Wang Y., Bentsen M., Bertino L., Zheng F.
Norwegian Climate prediction system

- Model: NorESM
  - Ocean: UniRe Klima
  - Carbon: UiB/UniRe
- Assimilation: EnKF
  - NERSC

Global skill assessment: Upper ocean temperature

RMSE calculated over the full model domain (averaged over the 10 prediction cycles)

For all model variables at 1-year lead average; 2-5 lead year average
- Analyze reduction of RMSE in EnKF-SST relative to Free
- Compare the improvements relative to Perfect
Conclusions

• Ocean data assimilation is worth the hassle
  – EnKF framework makes probabilistic forecasts seamless.
  – Also useful in a coupled climate model (…)
• Possible to correct both (poorly) observed and non-observed variables in the ocean
  – Still some regressions but not catastrophic
  – Ice edge accuracy within 50km, SST less than 1 deg C
• Other features are difficult to reproduce
  – Acceleration of ice drift
    • Model drift is too fast
    • Even the drift seasonality is not respected
  – Thinning of the sea ice
• R’n D to do
  – Ocean models Arctic water mass properties: better numerics or resolution
  – Sea ice validation argues for a change of the model EVP rheology