On the Development of an Efficient Ensemble Data Assimilation and Forecasting System for the Red Sea

Sivareddy Sanikommu, Habib Toye, Peng Zhan, Naila Raboudi, and Ibrahim Hoteit

King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

Presentation for the EMMDA-2020 Conference in NCMWRWF, India -February, 2020
Sources of Forecast Errors

- Uncertainties in
  - Initial conditions
  - Atmospheric forcing
  - Model physics
  - Boundary conditions (relevant for regional models)
  - Bathymetry (more relevant near coast)

Information about uncertainty is an important input for ocean data assimilation.
Forecast Error Covariance (B): It’s role

\[ X^a = X^b + B H^T [H B H^T + R]^{-1} [Y - H X^b] \]

\[ X^b = \begin{bmatrix} x_1^b \\ x_2^b \\ x_3^b \\ \vdots \\ x_9^b \end{bmatrix} \]

\[ B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{19} \\ B_{21} & B_{22} & \cdots & B_{29} \\ \vdots & \vdots & \ddots & \vdots \\ B_{91} & B_{92} & \cdots & B_{99} \end{bmatrix} \]

\[ Y = [y_0] \]

\[ R = [\sigma_0^2] \]

\[ H = [0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \]
How to account for these uncertainties in ensemble methods?

• Pre-selecting an appropriate ensemble(s) from an available dataset - EnOIs
  • **Advantages:** Computationally very efficient
  • **Limitations:** Does not well-represent the error-of-the-day, especially when dealing with rapidly varying dynamics

• Dynamically evolve the ensemble using the model - EnKFs
  • **Advantages:** Represents the error-of-the-day
  • **Limitations:** collapses in the ensemble spread, computationally demanding.

• Combine the two methods - Hybrid
Objective
Build an efficient, in terms of cost and performance, high-resolution ocean data assimilation system for the Red Sea (RS), for forecasting and reanalysis.

Configuration of Data assimilation system

Ocean Model: MITgcm
- 30E-50E & 10N-30N covering Red sea and Gulf of Aden, connecting the Arabian sea in the east
- 4km resolution and 50 levels
- Vertical resolution: 50 levels with 5m resolution in the upper 200 m and 300m resolution in the deeper layers
- Forcing: 0.5 x 0.5 ECMWF ensemble
- Boundary Conditions: Daily 25km-resolution GLORYS ocean reanalysis
- GEBCO Bathymetry and KPP vertical mixing scheme

Assimilation package: DART (Data Assimilation Research Package)
- Ensemble Adjustment Kalman Filter (EAKF)
- Localization: 300 km in the horizontal (no-vertical localization)
- Assimilate:
  - Satellite Level-4 Reynolds SST. **0.1 - 0.6 degC** observations error
  - Satellite Level-3 altimeter SLA (merged). **4cm** observations error
  - In situ T & S profiles from EN4 dataset (fully QC’d). **0.5 degC and 0.2 psu** observations error
EnOIs: pre-selection of ensembles

- EnOI - Uses a climatological ensembles
- Adaptive EnOI – Selects a new ensemble at every assimilation cycle
  - AEnOI-L2 –based on the distance from the present forecast
  - AEnOI-OMP –find a smallest possible subset of elements that best represents the forecasted state.

AEnOI-L2

\[ \text{computed distances} \]

\[ \text{dist}_1, \text{dist}_2, \text{dist}_3, \ldots, \text{dist}_L \]

\[ \text{ensemble } X = \{d_{j_1}, d_{j_2}, \ldots, d_{j_N}\} \]

\[ \text{forecast } x' \]

\[ \text{elements sorted based on their distances (from smallest to highest)} \]

AEnOI-OMP

\[ \text{inner products} \]

\[ \text{ip}_1, \text{ip}_2, \ldots, \text{ip}_n \]

\[ \text{dictionary elements} \]

\[ D_{k_0} = [d_{k_0}] \]

\[ \text{kept members at step 1} \]

\[ \text{solved } \min_{j \in \{1, \ldots, N\}} \|x' - y_j d_{k_0}\|_2 \text{ and set } r_1 = x' - y_1 d_{k_0} \]

\[ \text{residual } r_1 \]

\[ \text{inner products} \]

\[ \text{ip}_1, \text{ip}_2, \ldots, \text{ip}_n \]

\[ \text{dictionary elements} \]

\[ D_{k_1} = [d_{k_1}, d_{k_2}] \]

\[ \text{kept members at step 2} \]

\[ \text{solved } \min_{j \in \{1, \ldots, N\}} \|x' - y_j d_{k_2} - y_1 d_{k_1}\|_2 \text{ and set } r_2 = x' - y_1 d_{k_1} - y_2 d_{k_2} \]

\[ \text{residual } r_2 \]

\[ \text{inner products} \]

\[ \text{ip}_1, \text{ip}_2, \ldots, \text{ip}_n \]

\[ \text{dictionary elements} \]

\[ D_{k_N} = [d_{k_N}] \]

\[ \text{kept members at step } N \]

\[ \text{ensemble } X = \{d_{k_1}, d_{k_2}, \ldots, d_{k_N}\} \]
Time-evolution of (daily averaged forecast) SST and SSH RMSDs

Major factors for degradations:

- Repercussions from dynamical imbalances in the subsurface
- Spurious error correlations at the surface
- Data over-fitting
Degradations in the subsurface

Central Red Sea (38E, 22N)

Temperature

Salinity

Majorly factors for degradations:

✓ Spurious-correlations in the background error covariance matrix ???
✓ Dynamical imbalances
EAKF methods: Flow-dependent ensembles

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Initial condition</th>
<th>Atm. Forcing</th>
<th>Physics</th>
<th>Assimilation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fexp</td>
<td>Model state at 1st Jan, 2011</td>
<td>Ensemble mean</td>
<td>STANDARD</td>
<td>No</td>
</tr>
<tr>
<td>Iexp</td>
<td>50-member ensemble based on hindcasts recentered on 1st Jan, 2011</td>
<td>Ensemble mean</td>
<td>STANDARD</td>
<td>Yes</td>
</tr>
<tr>
<td>IAexp</td>
<td>50-member ensemble based on hindcasts recentered on 1st Jan, 2011</td>
<td>50-member ensemble</td>
<td>STANDARD</td>
<td>Yes</td>
</tr>
<tr>
<td>IAPexp</td>
<td>50-member ensemble based on hindcasts recentered on 1st Jan, 2011</td>
<td>50-member ensemble</td>
<td>RANDOM across members</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Iexp* → Accounts uncertainties in the initial conditions

*IAexp* → Accounts uncertainties in the initial conditions and atmospheric forcing

*IAPexp* → Accounts uncertainties in the initial conditions, atmospheric forcing, and model physics
Analysis corrections on 1st October, 2011

Background Error correlations (vertical) on 1st October, 2011

Too Noisy correlations in $iexp$ become more organized in $IAPexp$. 

(a) Temp2m vs Temp 36E 25N
(b) Temp2m vs Temp 36E 21N
(c) Temp2m vs Temp 40E 18N
Comparisons with in-situ SST and SSS observations from WHOI/KAUST cruise
Subsurface Temperature comparisons from WHOI/KAUST cruise

Maximum Vertical Velocity in the ocean column along RS axis

Not accounting for uncertainties from sufficient sources can severely degrade state estimation, particularly in the data sparse regions.

Improved biases in the deep layers with IAPexp
Hybrid-EAKF: Complimenting flow-dependent ensembles with pre-selected ensembles

Traditional Hybrid methods (e.g. Wang et al., 2007)

\[ \Sigma^{p,H} = (1 - \alpha)\Sigma^p + \alpha B, \quad \text{with} \quad 0 \leq \alpha \leq 1 \]

We implemented the Hybrid method in the context of the 2-steps, which act update of EAKF with DART, directly on the ensembles

\[ X^H = [K_d X^{ip}, \quad K_s X^{is}] + \bar{x}^p \]

Where \( K_d = \frac{\sqrt{(1-\alpha)(N_d+N_s-1)}}{N_d-1} \) and \( K_s = \frac{\sqrt{\alpha(N_d+N_s-1)}}{N_s-1} \)

The analysis-ensemble is prepared by re-centering the dynamic analysis-ensemble onto the Hybrid analysis-ensemble -mean

\[ x_{i,H}^u = \bar{x}_{i,H}^u + (x_i^u - \bar{x}^u), \quad i = 1, ..., N_d \]

\( X^H \) Hybrid ensemble
\( X^{ip} \) EAKF prior ensemble anomalies
\( X^{is} \) pre-selected static ensemble anomalies
\( \bar{x}^p \) mean of the prior ensemble
\( \bar{x}_{i,H}^u \) mean of the EAKF updated ensemble
\( \bar{x}_{i,H}^{u,H} \) Hybrid \( i \)th updated ensemble member
\( N_d \) size of the dynamic ensemble
\( N_s \) size of the pre-selected static ensemble

\( \Sigma^p \) flow-dependent covariance
\( B \) Static covariance
\( \Sigma^{p,H} \) Hybrid covariance
\( \alpha \) weighting factor
Forecast-Data comparisons at the surface

The SST and SSH are improved by 20% in Hybrid compared to EAKF.

Non-assimilated SSS is too noisy in EnOI.

Significantly improved by Hybrid, even compared to EAKF.
Forecast-Data comparisons in the sub-surface

Comparisons with independent WHOI/KAUST cruise observations

Max. Vertical Velocities along the Red Sea axis

(a) FREE

(b) EnOI

(c) EAKF

(d) Hybrid
Comparison for Eddy features

Hybrid is best, even better than the interpolated product
Summary

• EnOI-alone methods, which rely on pre-selected ensembles, are prone to spurious corrections.

• Dynamically evolving the ensemble to account for uncertainties in initial conditions, atmospheric forcing and internal physics is greatly beneficial.

• Combining the dynamical ensemble with a pre-selected (climatological) ensemble with the Hybrid method enforces smoothness in the background covariance and greatly reduce computational cost.

• Hybrid method enhances the impact of assimilated observations and provides improved and dynamically balanced state
List of Publications


Thank You