Preliminary results comparing adjoint- and ensemble-based approaches to observation impact using the GMAO data assimilation system

Fábio L. R. Diniz*
and Ricardo Todling

Global Modeling and Assimilation Office
NASA

UMD Weather-Chaos Group Meeting
December 11th, 2017

*Centro de Previsão de Tempo e Estudos Climáticos, Cachoeira Paulista, SP, Brazil

Thanks to D. Groff (NCEP) for pointing to Ota’s EFSOI upgrade to Whitaker’s EnKF.
Outline

Introduction

Recap of FSOI Basics

Preliminary Results

Closing Remarks
Evolution of Forecast Sensitivity and Observation Impact (FSOI) at GMAO:

- GMAO has been calculating FSOI in its Forward Processing (FP) system for several years.
- FP has evolved from 3dVar to Hybrid-3dVar to what is presently Hybrid-4dEnVar.
- Along the years the GMAO forward model has gone from FV to FV$^3$; accordingly, the adjoint model has gone from AD-FV to AD-FV$^3$.
- Linearized physics, and corresponding adjoint, has evolved from simple diffusion and vertical drag to more elaborate accountability of convection (Holdaway, Errico, Gelaro & Kim).

Ensemble DA opens the door to bypass the Adjoint Model:

- In a dual-analysis system (Var & Ens) the possibility exists to base FSOI fully on the ensemble - this has its caveats (see what follows).
- Alternatively:
  - Method I The AD-Var-analysis can be adapted to make use of an ensemble forecast to implicitly estimate forecast sensitivities;
  - Method II Or, similarly, but not identically, the ensemble might be used to explicitly estimate forecast sensitivities required by the AD-Var-analysis.

This presentation provides insights from preliminary evaluation of these possibilities.
Hybrid GEOS DAS

Forward/Backward procedures

**Forward Data Assimilation-Forecast Procedure:**

**Input:** Observations and Central Background

Central Analysis System (GSI) → Central Initial State → Forecast Model → **Output:** Central Forecast

**Input:** Observations and Ensemble of Backgrounds

Ensemble Analysis System (EnSRF) → Ensemble of Initial States → Forecast Model → **Output:** Ensemble of Forecasts

**Adjoint Data Assimilation-Forecast Procedure:**

**Output:** Forecast sensitivity to observations and background

Adj. GSI or Adj. EnSRF → Forecast Sensitivity to Initial State → Adj. Fcst. Model or Ens. Pert. → **Input:** Forecast (measure)

Adj. GSI or Adj. EnSRF
Error reduction measure and FSOI

Forecast error:

\[ e^s(t_v | t_0) = \langle [x^f(t_v | t_0) - x^v(t_v)]^T T [x^f(t_v | t_0) - x^v(t_v)] \rangle \]

The impact of observations is typically evaluated by studying how the error measure above changes as a consequence of assimilating observations. Whether based on adjoint or ensemble techniques, these methods require evaluation of expressions of the form:

\[ \delta e \approx \langle d^T K^T g_0 \rangle \]

with \( d \) and \( K \) being the background residual vector and the analysis gain matrix, and \( g \) amounting to a forecast sensitivity vector whose approximation leads to all kinds of formula.

<table>
<thead>
<tr>
<th>AD-Solver (( K^T ))</th>
<th>Forecast Sensitivity (( g_0 ))</th>
<th>This Talk</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA-FSOI Var ADM done</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE-FSOI En En done</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VE-FSOI Var En done</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EA-FSOI En ADM —</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
VA-FSOI vs EE-FSOI in a Dual-Analysis Hybrid System
## Adjoint- and Ensemble-based FSOI

### Variational-Adjoint-FSOI (VA-FSOI)

Second-order Approximation (Trapezoidal rule; Langland & Baker 2004; Tellus):

\[ g_0^A \equiv \frac{1}{2} (M_a^T T_a + M_b^T T_b) \cdot \]

And in a system such as GSI, the calculation of \( \delta e \) can be done as in:

\[ \delta e \approx \langle d^T R^{-1} H \tilde{g} \rangle, \]

where \( \tilde{g} \) is derived from the GSI-hybrid solver as in (double-CG or Bi-CG):

\[
(B + BH^T R^{-1} HB)z = Bg_0^A \\
\tilde{g} = Bz
\]

for \( B = \beta_c B_c + \beta_e B_e \).

### Ensemble-Ensemble-FSOI (EE-FSOI)

In Ensemble systems, the gradient is defined with respect to the ensemble mean:

\[ g_0^E \equiv \frac{1}{2} X^f_a T(\bar{e}^a + \bar{e}^b), \]

where \( X^f_a \equiv X^f(t_v | t_a) \) is a matrix created from the ensemble perturbation of forecasts issued from \( t_a \) and valid at \( t_v \), and the over-bar represents ensemble average. And in a system such as the EnSRF, calculation of \( \delta e \) amount to:

\[ \delta e \approx \frac{1}{2} \langle d^T R^{-1} H (L \cdot X^a X^f T) T(\bar{e}^a + \bar{e}^b) \rangle, \]

where \( X^a_a \equiv X^a(t_a) \) is a matrix formed from ensemble analysis perturbations (Kalnay et al., 2012, Tellus; Ota et. al, 2013, Tellus). An argument has been made to have \( L \) above as an advected form of the \( L \) used in the forward ensemble analysis.
Error reductions are similar between central and ensemble forecasts, thought latter is slightly smaller in absolute value for 12-hour forecasts.

Left: compares FSOI when backward Var changed from Hyb-4dEnVar to 4dEnVar.

Right: presents EE-FSOI.
VA-FSOI vs EE-FSOI
Observation Impacts per Observation

- In this case, to render fair comparison, the VA-FSOI employs a 4DEnVar solver.
- EE-FSOI presents considerable more impacts per observation for the entire observing system.

Overall this comparison reflects that:
- Ensemble mean forecasts are unrelated to the central forecast.
- But more importantly, the ensemble analysis handles observations largely differently to how the hybrid analysis does.
The difference in treatment of observations between ensemble and central analyses is evidenced in the observation count.

The GSI and EnSRF solvers have considerably different convergence criteria.

Even with the ideal DFS-based criterion (chosen here), the EnSRF ignores a very large percentage of the observations.

All-in-all we don’t think observation impacts derived from the EnSRF solver represent well how the deterministic (central) analysis system uses observations.
VA-FSOI vs VE-FSOI in a Dual-Analysis Hybrid System
Variational-Ensemble-FSOI (VE-FSOI) Method I

In an En-Var, such that $B = B_e$, the ensemble background covariance allows for the following to be written

$$B_e g_0^A = L \cdot X_b (X_b)^T g_0^A$$
$$\approx L \cdot X_b X_a^T M^T(t_a, t_b)g_0^A$$

We can replace $g_0^A$ with $g_0^E$ (using central forecast errors) in the RHS to get

$$B_e g_0^A \approx B_e g_0^E$$
$$\approx \frac{1}{2} L \cdot X_b X_a^T M^T(t_a, t_b)X_a^{rT} T(e^a + e^b)$$
$$= \frac{1}{2} L \cdot X_b X_b^{rT} T(e^a + e^b)$$

which amounts to a simple change to the RHS of the minimization problem solved for calculation of observation impacts in the Var system.

Note: $e^a(e^b)$ replaces $\bar{e}^a(\bar{e}^b)$

Variational-Ensemble-FSOI (VE-FSOI) Method II

Alternatively, we can try to use the approach of Ancell & Hakim (2007; MWR) to estimate forecast sensitivities using an ensemble of forecasts.

In this case, the forecast sensitivity is estimated as in:

$$\frac{\partial f}{\partial x} = D^{-1} \begin{bmatrix} \delta x_1 \delta e_1^T \\ \delta x_2 \delta e_2^T \\ \vdots \\ \delta x_n \delta e_n^T \end{bmatrix}$$

where $D = diag(\|\delta x_1\|^2, \|\delta x_2\|^2, \ldots, \|\delta x_n\|^2)$, $dim(x_i) = dim(e_i) = M \times 1$, and $n$ is the state-space dimension.
Lack of advection of localization scales in the RHS of the Var impact expression motivates following Pellerin et al. (2016; WMO) and evaluating 12-hr instead of 24-hr FSOI.

Left: compares FSOI when backward Var changed from Hyb-4dEnVar to 4dEnVar.

Right: compares VA-FSOI with VE-FSOI Method I.

Remark: 32-member ensemble perturbations seem rather reasonable replacement for ADM for 12-hr sensitivity calculation in 4dEnVar context.
To have consistency in the test, we calculate Adjoint-based impacts for a changed Adjoint analysis integration where the climatological term is shut off, thus converting the backward run into a 4dEnVar instead of its default (FP-like) Hybrid-4dEnVar.

The two approaches treat the observations in exactly the same way, and fully consistent with how the forward (Hyb-4dEnVar) solver treats them.

Replacing the Adjoint Model with Ensemble perturbations to estimate forecast sensitivities leaves the analysis solver untouched wrt each other.

The figure on the right shows observation counts between the VA-FSOI and VE-FSOI Method I techniques, for backward integrations of 4dEnVar, covering a 10-day period.
Overall, impacts per observation don’t seem to change much and are largely comparable when ADM is replaced with Ensemble perturbations.

Closer look reveals pilot ballons (Pibal), dropsondes (Dropsnd) and near surface observations (DriftBuoy, LandSfc & MarineSfc) to have larger impact per observation when Ens-Perts are used compared to when ADM is used to estimate forecast sensitivity.

The above seems to be consistent with the fact that the simply parameterized adjoint physics is expected to mis-represent water and near surface fields as compared to the full GCM.
More importantly, though the observation operators in the hybrid GSI and EnSRF are shared, GSI and EnSRF treat observations rather differently.

In a dual hybrid DA system, when a (low resolution) ensemble analysis filter is used to provide flow dependence to a (high resolution) hybrid analysis, certain configurations of the ensemble filter might discourage assessing observation impact using the EE-FSOI based on the EnDA part of system.

The comment above applies particularly to GSI-EnSRF-based systems.

As in Pellerin et al., we have shown that it is possible to enable the Var system to derive observation impacts with forecast sensitivities calculated from the ensemble thus avoiding the adjoint model.
Future Works

- Expand implementation of VE-FSOI Method-I to accommodate the advection of localization.
- Extend all results for 24-h.
- Extend the 10-day case to full month.