Ensemble Forecast Sensitivity to Observations (EFSO) and Proactive Quality Control

Daisuke Hotta\textsuperscript{(2,1)}, Eugenia Kalnay\textsuperscript{(1)}, Yoichiro Ota\textsuperscript{(2,3)}, Takemasa Miyoshi\textsuperscript{(4,1)}

\textsuperscript{(1)} University of Maryland
\textsuperscript{(2)} Japan Meteorological Agency
\textsuperscript{(3)} National Centers for Environmental Prediction
\textsuperscript{(4)} RIKEN, Kobe, Japan

Acknowledgements: NESDIS JPSS for support, and JCSDA for access to S4 supercomputer and Dr. Jung (NCEP) for advice
Introduction:
The NCEP “5-day skill dropout” problem

From Kumar et.al (2009)
Introduction:
The NCEP “5-day skill dropout” problem

From Kumar et.al (2009)

- “Culprit” is not the model but “bad observations” (or inability of DA system to properly assimilate them)

- How can we detect those “flawed” observations without resorting to an external system?
EFSO: Ensemble Forecast Sensitivity to Observations

- Quantifies how much each observation improved/degraded the forecast
- Economical alternative to OSE
- First invented for a variational DA-system using the adjoint method by Langland and Baker (2004)
- Liu and Kalnay (2008) adapted it to LETKF
- Kalnay, Ota, Miyoshi and Liu (Tellus, 2012) has given an improved, simpler formulation
- The new formulation is
  - more accurate
  - simpler and easier to implement
  - applicable to any formulation of enKF
Ota et.al (2013): Applied EFSO to NCEP GFS/EnSRF using all operational observations

They identified regional 24hr “forecast failures” by the following procedures:

1. Divide the globe into 30°x30° lat-lon boxes
2. Find all cases which satisfy the following criteria
   1. the 24hr regional forecast error is at least 20% larger than the 36hr forecast error verifying at the same time, and
   2. the 24hr forecast has errors at least twice the time average.

3. Perform EFSO to the above-identified cases and attribute that failure to the top observation type that had most negative impact on the forecast

- → Found 7 cases of regional forecast failure (next slide). In every case, forecast was improved by not assimilating “bad observations”.
24-hr forecast error correction (Ota et al. 2013)
- identified 7 cases of large $30^\circ \times 30^\circ$ regional errors,
- rerun the forecasts denying bad obs.
- the forecast errors were substantially reduced

<table>
<thead>
<tr>
<th>Initial UTC</th>
<th>Area</th>
<th>Size</th>
<th>Rate</th>
<th>N</th>
<th>Denied observation</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 UTC JAN 10</td>
<td>90S<del>60S 100E</del>130E</td>
<td>2.04</td>
<td>1.20</td>
<td>1</td>
<td>GPSRO (80S<del>60S, 90E</del>120E) ASCAT (60S<del>50S, 100E</del>120E)</td>
<td>-6.6%</td>
</tr>
<tr>
<td>06 UTC JAN 12</td>
<td>50N~80N 150E ~ 180</td>
<td>2.18</td>
<td>1.40</td>
<td>1</td>
<td>AMSUA (ch4: 45N<del>75N, 160E</del>170W, ch5:40N<del>55N, 155E</del>180, NOAA15 ch6: 50N<del>75N, 140E</del>170W, ch7: 70N<del>80N, 130E</del>170E)</td>
<td>-11.4%</td>
</tr>
<tr>
<td>00 UTC JAN 16</td>
<td>30N<del>60N 30W</del>0</td>
<td>2.13</td>
<td>1.31</td>
<td>2</td>
<td>Radiosonde wind (Valentia, Ireland), ASCAT (40N<del>47N, 20W</del>10W, 50N<del>55N, 35W</del>30W)</td>
<td>-1.0%</td>
</tr>
<tr>
<td>12 UTC JAN 22</td>
<td>90S<del>60S 130E</del>160E</td>
<td>2.34</td>
<td>1.22</td>
<td>2</td>
<td>AMSUA (ch5: 65S<del>50S, 90E</del>110E, 60S<del>50S, 120E</del>127E, ch6: 60S<del>45S, 110E</del>125E)</td>
<td>-2.2%</td>
</tr>
<tr>
<td>06 UTC FEB 2</td>
<td>50N<del>80N 150W</del>120W</td>
<td>3.10</td>
<td>1.32</td>
<td>4</td>
<td>IASI (35N<del>45N, 155W</del>150W) NEXRAD (55N<del>60N, 160W</del>135W)</td>
<td>-5.5%</td>
</tr>
<tr>
<td>18 UTC FEB 6</td>
<td>60N<del>90N 50E</del>80E</td>
<td>2.06</td>
<td>1.71</td>
<td>2</td>
<td>MODIS_Wind (60N<del>90N, 30E</del>90E)</td>
<td>-39.0%</td>
</tr>
<tr>
<td>18 UTC FEB 6</td>
<td>90S<del>60S 20W</del>10E</td>
<td>3.56</td>
<td>1.22</td>
<td>1</td>
<td>MODIS_Wind (80S<del>50S, 30W</del>0)</td>
<td>-22.5%</td>
</tr>
</tbody>
</table>
Proactive QC: Identify bad observations by EFSO and then repeat analysis/forecast withdrawing them.

- After identifying MODIS polar winds producing bad 24hr forecasts, the withdrawal of these winds reduced the regional forecast errors by 39%, as projected by EFSO.
1. Investigate if the same methodology also works for:
   - enKF/GSI hybrid DA
   - shorter lead-time (6 or 12 hours)

2. Improve “forecast failure” detection criteria
   - Divide the globe by zeros of isotropic spherical harmonics (e.g. m=6, n=3) instead of regular lat-lon rectangle (as in Ota et.al)
   - Use of forecast spread to capture flow-dependent dynamic instability
Implementation of EFSO to the hybrid GSI

New Experiment

• Analysis: **Low-resolution hybrid GSI**
  – **T254L64** for 3D-Var & deterministic forecast
  – **T126L64** for enKF & ensfcst
  – Ensemble size: 80 members
  – **LETKF** instead of the default serial enSRF
  – Observations: same as operational system

• EFSO settings:
  – Verifying Truth: both **GSI analysis** and **enSRF ensemble-mean analysis**
  – Evaluation Lead-time: **6, 12 and 24hrs**

• Period:
  – 31 days from 2012-Jan-08-00Z to 2012-Feb-07-18Z
  – 4 cycles per day (124 cases in total)

Ota et.al (2013)

• Analysis: **pure enKF**
  – **T254L64** for enKF & ensfcst
  – Ensemble size: 80 members
  – **serial enSRF** (Whitaker and Hamill, 2002)
  – Observations: same as operational system

• EFSO settings:
  – Verifying Truth: **enSRF ensemble-mean analysis**
  – Evaluation Lead-time: **24hrs**

• Period:
  – 31 days from 2012-Jan-08-00Z to 2012-Feb-07-18Z
  – 4 cycles per day (124 cases in total)
Average total observation impact classified by observation types: Sensitivity to Verifying Truth

Lead-Time: 6 hours
Verified against GSI analysis

Lead-Time: 6 hours
Verified against LETKF analysis

Time Averaged EFSO’s are rather insensitive to the choice of verifying truth!
Average total observation impact classified by observation types: Comparison with Ota et.al (2013)

Averaged total Obs. Impact by obs. type Moist Energy norm, EFT=6hr

Lead-Time: 6 hours
Verified against GSI analysis

Lead-Time: 6 hours
Verified against LETKF analysis

Lead-Time: 24 hours
Verified against enSRF analysis from Ota et.al (2013)

Fairly good agreement among different verifications!
Average total observation impact classified by observation types:
Dependence on Evaluation Lead-Time

Fairly good agreement among different verifications and lead times!
Example of the detected “dropout” and its EFSO result
Feb. 06 18UTC, near the North Pole

Flawed MODIS-Wind obs. detected even with 6-hour lead time!
Summary

• **EFSO (Ensemble Forecast Sensitivity to Observations)** has been shown to be capable of detecting “flawed observations” that degrade the forecasts.

• This will allow us to perform *flow-dependent “Proactive QC”* in which
  – observations that cause regional forecast failure are identified with short lead-time EFSO, and
  – the analysis & forecasts are repeated rejecting the identified “flawed observations”.

• We have found that we can use EFSO for the hybrid GSI with 6-hour forecast and that it is not very sensitive to the choice of verification or the lead-time.

• Our next step is to actually perform Proactive QC and confirm that this method improves the performance of NWP system.
Future work

(Immediate future)
• Improve criteria for regional “forecast failure” detection
• Perform Proactive QC using regional EFSO with 6-hour lead-time to identify “bad observations”
• Try use of forecast error metrics in the observation space (cf. backup slide)

(Ultimate goal)
• Repeat all processes using the full-resolution operational system, improve NCEP’s operational forecast!
• Accumulate cases of flawed observations along with detailed information about the atmospheric characteristics and how the observations were used
• → Provide that information to algorithm developers in hope of assisting them to improve algorithms
Backup slides
Ensemble Sensitivity in Model space and in Observation space
(after Todling’s adjoint sensitivity formulation, 2009, 2012)

\[
\Delta e^2 = \left[ MK(y - H(x_{0|t-6}^b)) \right]^T C (e_{t0} + e_{t-6})
\]
\[
= \left[ (y - H(x_{0|t-6}^b)) \right]^T R^{-1} Y_0 X_{t0}^T C (e_{t0} + e_{t-6}) / (K - 1)
\]

\[
\Delta d^2 = \left[ d_t^T C y d_t^f - d_t^g C y d_t^g \right] \approx
\]
\[
- \delta y_o^T R^{-1} \left( H_0 X_0^a \right) \left( H_t X_t^f \right)^T C_y \left( d_t^f + d_t^g \right) / (K - 1)
\]

We can use \( C = C_y = R^{-1} \), and  \( \Delta e^2 = \left[ e^T R^{-1} e^f - e^g T R^{-1} e^g \right] \)

where  \( e^f = H(x^f) - H(x^a) \);  \( e^g = H(x^g) - H(x^a) \)

so the results can be directly compared.
Ensemble Forecast Sensitivity to Observations

\[ \Delta e^2 = (e_{t|0} - e_{t|-6})^T \left( e_{t|0} + e_{t|-6} \right) \]
\[ = (\bar{x}_{t|0}^f - \bar{x}_{t|-6}^f)^T \left( e_{t|0} + e_{t|-6} \right) \]
\[ = \left[ M(\bar{x}_0^a - \bar{x}_0^b) \right]^T \left( e_{t|0} + e_{t|-6} \right), \text{ so that} \]

\[ \Delta e^2 = \left[ MK(y - H(x_0^b)) \right]^T \left( e_{t|0} + e_{t|-6} \right) \]

Langland and Baker (2004), Gelaro and Zhu, solve this with the adjoint:

\[ \Delta e^2 = \left[ (y - H(x_0^b)) \right]^T K^T M^T \left( e_{t|0} + e_{t|-6} \right) \]

This requires the adjoint of the model \( M^T \) and of the data assimilation system \( K^T \) (Langland and Baker, 2004)
Ensemble Forecast Sensitivity to Observations

With EnKF we can use the original equation without “adjointing”:

Recall that

\[ K = P^a H^T R^{-1} = X^a X^{aT} H^T R^{-1} / (K - 1) \]

so that

\[ MK = MX^a (X^{aT} H^T) R^{-1} / (K - 1) = X^f_{t|0} Y^a T R^{-1} / (K - 1) \]

Thus,

\[ \Delta e^2 = \left[ MK(y - H(x^b_{0|6})) \right]^T (e_{t|0} + e_{t|-6}) \]

\[ = \left[ (y - H(x^b_{0|6})) \right]^T R^{-1} Y^a_{0} X^f_{t|0} (e_{t|0} + e_{t|-6}) / (K - 1) \]

This uses the available **nonlinear** forecast ensemble products.
Tested ability to detect *a poor quality ob impact* on the forecast in the Lorenz 40 variable model

Observation impact from LB(+) and from ensemble sensitivity (●)

- The adjoint and the ensemble sensitivity give *similar observation impact* on the 24 hr forecast.
- The ensemble sensitivity is nonlinear and is able to *detect bad obs* for longer forecasts.
- This was done ignoring EnKF localization.

The localization center point for observation impact estimate is now moved with the horizontal wind: an approximation
Time series of 6hr EFSO impact (UV, GPS, O3) verifying against LETKF and hybrid GSI
Time series of 6hr EFSO impact ($T$, $Q$, $P_s$) verifying against LETKF and hybrid GSI

- **EFSO obs=$T$ area=NH**
- **EFSO obs=$Q$ area=NH**
- **EFSO obs=$P_S$ area=NH**
- **EFSO obs=$T$ area=TR**
- **EFSO obs=$Q$ area=TR**
- **EFSO obs=$P_S$ area=TR**
- **EFSO obs=$T$ area=SH**
- **EFSO obs=$Q$ area=SH**
- **EFSO obs=$P_S$ area=SH**
Time series of 6hr EFSO impact (radiances) verifying against LETKF and hybrid GSI

The day-to-day verifications against LETKF and GSI Hybrid are generally consistent but LETKF estimates a slightly larger impact.