Predicting Observation Impact on Forecast Skill with Machine Learning

Francois Vandenberghe & Dick Dee

Joint Center for Satellite Data Assimilation (JCSDA)
Boulder, CO
vandenb@ucar.edu
Forecast Sensitivity to Observations Impact (FSOI)

Assimilation cost function:

\[
J = (x - x^b)^T B^{-1} (x - x^b) + (y^o - Hx)^T R^{-1} (y^o - Hx)
\]

Analysis

\[
x^a = x^b + BH^T (HBH^T + R)^{-1} (y^o - Hx^b)
\]

Analysis uncertainty

\[
A = (I - KH)B
\]
Forecast Sensitivity to Observations Impact (FSOI)

\[ \delta e \approx d^T K^T [M^T_b e(x^f_b) + M^T_a e(x^f_a)] \]

Adjoint-derived (single outer-loop) observation impact

\[ \delta e \approx d^T R^{-1} L(HX^0_a)X^f_a [e(x^f_b) + e(x^f_a)] \]

Ensemble-derived observation impact

Langland and Baker (2004)
Forecast Sensitivity to Observations Impact (FSOI)

Total Impact
EMC 24h Observation Impact Summary
Global, 00Z 07Apr-06May, 2018
Total Impact (J/kg)

Observations count
EMC 24h Observation Impact Summary
Global, 00Z April 2018
Observation Count per Analysis
Fractional Impact

![Graph showing fractional impact with ts for travel ban March 11 and COSMIC-2 March 25, and AIRCRAFT and GNSSRO indicators.](image)
Seek for the radiance **bias correction** that maximizes FSOI:

- **Use Machine Learning** to compute bias correction coefficients (MLBC)
- Do not limit the number of **predictors**. (big data)

\[ I = \left( y^o - Hx^b \right)^T \frac{\partial F}{\partial y} \]

impact  innovation  sensitivity

Can Forecast Sensitivity to Observations be predicted?
Machine Learning

Software: TensorFlow on Amazon
Data: Dec 2014, Jan 2015 & Feb 2015
Focus initially on AMSU
Training: Dec 1 – Feb 14
Prediction: Feb 14 - 28
Predictors: all FV3 first guess variables:

# Data Size

<table>
<thead>
<tr>
<th></th>
<th>Number of files</th>
<th>Avg file size</th>
<th>Total files size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>360</td>
<td>48.9 MB</td>
<td>17.2 GB</td>
</tr>
<tr>
<td>Model background</td>
<td>360</td>
<td>464.2 MB</td>
<td>163.2 GB</td>
</tr>
</tbody>
</table>
Data Analysis

GMAO Dec-Jan-Feb 2015 scaled sensitivity
AMSU N18 channel 7

Scaled Sensitivity

Scaled Sensitivity between [-10, 10]

p-value = 0.0, sample does not look Gaussian
GMAO Dec-Jan-Feb 2015 scaled sensitivity
AMSU N18 channel 7

Kurtosis measure: 54.68
Machine Learning

GMAO Forecast Sensitivity to AMSUA Channel 14: Test data (x-axis) and predictions (y-axis) 00z analyses for December 2014 (~3.5x10^5 obs)

- Linear regression
- Gradient Boosting
Test data (x-axis) and predictions (y-axis) for GMAO Dec-Jan-Feb 2015 sensitivity AMSU N18 channel 7
Test data (x-axis) and predictions (y-axis) for GMAO Dec-Jan-Feb 2015 sensitivity AMSU N18 channel 7
Results

Mean Squared Errors AMSUA N18 CH7

- Only zeros: 0.9515
- Linear Regression: 0.9507
- Multi-Input NN: 0.9503
- Convolutional NN: 0.949
- Dense NN: 0.9477
- Gradient Boosting: 0.946
Summary

- Forecast to Observation Sensitivity is not gaussian.
- Sophisticated ML techniques lead to better results than simple regression
- Very limited predictability.