Ensemble-based approximation of observation impact using an observation-based verification metric

Matthias Sommer and Martin Weissmann

Hans–Ertel–Centre for Weather Research
Data Assimilation Branch
Ludwig–Maximilians–Universität München

1 June 2015
Research question

In a complex systems of observations, data assimilation and forecasts...

- How much do the individual observations contribute to the forecast quality?
- Are the observations used in an optimal way?

Motivation

The assessment of observation impact can help...

- to improve the interaction of observations, data assimilation and model
- to exclude data that systematically degrades the forecast.

Methods to determine observation impact

- Data-denial-experiments: Big computational cost
- Adjoint-based methods: Not available for all models, e.g. COSMO
- Ensemble-based methods Kalnay et al. [2012], Liu and Kalnay [2008], Sommer and Weissmann [2014]
Observation impact: Definition

Data denial impact of observations $d'$ relative to all observations $d$

$$J(d') = |e_f^d|^2 - |e_f^{d-d'}|^2$$

- $d$: All available observations
- $d'$: Small subset of observations whose impact one is interested in

Algorithm

```
ANALYSIS
forecasts
ekf
Y*W

FORECAST
forecasts
fof/ekf
ObsImpact
J
```

Example

- (a) $|e_f^d|^2$
- (b) $|e_f^{d-d'}|^2$
- (c) $J = \frac{1}{2} \left( |e_f^d|^2 - |e_f^{d-d'}|^2 \right)$
**LETKF update equation**

\[
\bar{x}_{aj} = \bar{x}_{bj} \tilde{P}_a(j) Y_b^\top R^{-1}(j) (y_o - \bar{y}_b) + \bar{x}_{bj}
\]

**Variables**

- \( j \): Grid point
- \( \bar{x}_a \): Analysis mean
- \( X_b \): Background ensemble
- \( \tilde{P}_a \): Ensemble analysis error covariance matrix
- \( W^a(j) = \left( (K - 1) \tilde{P}_a(j) \right)^{\frac{1}{2}} \): Weight matrix
- \( Y_b \): Background ensemble in observation space
- \( R \): Observation error covariance matrix
- \( d = y_o - \bar{y}_b \): Observational increment
- \( \bar{x}_b \): Background mean
Approximation with ensemble perturbations

**LETKF update equation**

\[ \bar{x}_{aj} = x_{bj} \tilde{P}_a(j) Y_b^T R^{-1}(j) (y_o - \bar{y}_b) + \bar{x}_{bj} \]

**Data denial observation impact**

\[ J(d') = |e^d_f|^2 - |e^d_f - e^0_f|^2 = (e^d_f + e^{d-d'}_f) \cdot (e^d_f - e^{d-d'}_f) \]

**Direct derivation** [Kalnay et al., 2012]

\[ e^d_f - e^0_f = e^d_f - e^0_f \approx \frac{1}{K-1} x^d_f (Y_b W^d)^T R^{-1} d \]

\[ \Rightarrow J(d') = (e^d_f + e^{d-d'}_f) \cdot (e^d_f - e^0_f - (e^{d-d'}_f - e^0_f)) \]

\[ \approx (e^d_f + e^{d-d'}_f) \cdot \left( \frac{1}{K-1} x^d_f (Y_b W^d)^T R^{-1} d' \right) \]

\[ \approx (e^d_f + e^0_f) \cdot \left( \frac{1}{K-1} x^d_f (Y_b W^d)^T R^{-1} d' \right) \]
Approximation with ensemble perturbations

**LETKF update equation**

\[ x_{aj} = x_{bj} \tilde{P}_a(j) Y_b \mathbf{R}^{-1}(j) (y_o - y_b) + x_{bj} \]

**Data denial observation impact**

\[ J(d') = |e^{d'}_f|^2 - |e^{d-d'}_f|^2 = (e^{d}_f + e^{d-d'}_f) \cdot (e^{d}_f - e^{d-d'}_f) \]

**Direct derivation** [Kalnay et al., 2012]

\[
\begin{align*}
    e^{d}_f - e^{0}_f &= \bar{x}^{d}_f - \bar{x}^{0}_f \\
    &= \left( e^{d}_f + e^{d-d'}_f \right) \cdot \left( e^{d}_f - e^{0}_f - (e^{d-d'}_f - e^{0}_f) \right) \\
    &\approx \left( e^{d}_f + e^{d-d'}_f \right) \cdot \left( \frac{1}{K-1} X^{d}_f (Y_b W^{d})^\top \mathbf{R}^{-1} d' \right) \\
    &\approx \left( e^{d}_f + e^{0}_f \right) \cdot \left( \frac{1}{K-1} X^{d}_f (Y_b W^{d})^\top \mathbf{R}^{-1} d' \right)
\end{align*}
\]

**Taylor expansion** [Sommer and Weissmann, 2015]

\[
\begin{align*}
    J(d') &= J(0) + \frac{d}{dd'} \bigg|_{d'=0} J(d') d' + \mathcal{O} \left( |d'|^2 \right) \\
    &= 2e^{d}_f \cdot \left( -\frac{d}{dd'} \bigg|_{d'=0} e^{d-d'}_f \right) d' + \mathcal{O} \left( |d'|^2 \right) \\
    &\approx 2e^{d}_f \cdot \left( \frac{1}{K-1} X^{d}_f (Y_b W^{d})^\top \mathbf{R}^{-1} d' \right)
\end{align*}
\]
Verification with...

\[ e_f = \overline{x_f} - x_a \]
\[ |e_f|^2 = \sum_{\text{gridpoints}} \frac{1}{2} (\overline{u_f} - \overline{u_a})^2 + \frac{1}{2} (\overline{v_f} - \overline{v_a})^2 \]
\[ \Rightarrow J(d') \approx 2e_f^d \cdot \left( \frac{1}{K-1} X_f^d (Y_b W^d)^T R^{-1} d' \right) \]

- Homogeneous in space and time
- Strongly correlated to forecast

\[ e_f = H(\overline{x_f}) - y_o \]
\[ |e_f|^2 = \sum_{\text{observations}} \left( \frac{H(\overline{x_f}) - y_o}{\sigma} \right)^2 \]
\[ \Rightarrow J(d') \approx 2e_f^d \cdot \left( \frac{1}{K-1} Y_f^d (Y_b W^d)^T R^{-1} d' \right) \]

- Independent of forecast
- Computationally easy
- Unobserved regions/variables may be ignored
**DWD Convective-scale assimilation and forecasting systems**

### Kilometer-scale Ensemble Data Assimilation (KENDA)
- Localized Ensemble Transform Kalman Filter for use with COSMO-DE (in development)

### Consortium for Small-scale Modelling (COSMO)
- Operational limited-area model of Deutscher Wetterdienst
- Grid point model of non-hydrostatic equations
- Horizontal resolution: 2.8 km; 50 vertical levels

![Figure: COSMO-DE domain (~ 1300 km × 1200 km)](image)

### Experimental settings
- Test period: 10 June 2012 12:00 UTC – 13 June 2012 15:00 UTC
- Initialization every 3 h
- Forecast length 6 h
- 40-members ensemble
- Observations used:
  - AIREP (Aircrafts):  $$U, V, T$$
  - PROF (Wind profiler):  $$U, V$$
  - SYNOP (Ground stations):  $$U, V, T, RH$$
  - TEMP (Weather Balloons):  $$U, V, T, RH$$
10 June 2012 12:00 UTC – 13 June 2012 15:00 UTC

Number of assimilated observations per station

Number of AIREP observations

Number of PROF observations

Number of SYNOP observations

Number of TEMP observations

Matthias Sommer and Martin Weissmann
Ensemble-based approximation of observation impact
**Impact per observation type**

### Total impact

![Impact bar chart](chart)

- **AIREP**: -0.06
- **PROF**: -0.04
- **SYNOP**: -0.02
- **TEMP**: 0

### Number of observations

- **AIREP**: 0
- **PROF**: 1
- **SYNOP**: 2
- **TEMP**: $10^5$

### Number of stations

- **AIREP**: 0
- **PROF**: 500
- **SYNOP**: 1000
- **TEMP**: 5

### Impact per observation

- **AIREP**: $10^{-4}$
- **PROF**: $10^{-2}$
- **SYNOP**: 0
- **TEMP**: $10^{-3}$

### One wind profiler equivalents...

- **AIREP**: 134
- **PROF**: 73
- **SYNOP**: 16

Matthias Sommer and Martin Weissmann

Ensemble-based approximation of observation impact
Non-Gaussian distribution
- Ratio of negative to positive values ca. 52:48
- Width of distribution $\gg$ Mean
Distribution of impact values

Histogram of individual observations impact values

AIREP. Ratio neg/pos: 0.516 : 0.484

SYNOP. Ratio neg/pos: 0.521 : 0.479

PROF. Ratio neg/pos: 0.51 : 0.49

TEMP. Ratio neg/pos: 0.518 : 0.482

Transformation of x-axis

\[ J(d') = |e^d|^2 - |e^{d-d'}|^2 \quad \rightarrow \quad \hat{J}(d') = \text{sign}(J(d')) \sqrt{|J(d')|} \]
Distribution of impact values

Histogram of individual observations impact values

Transformation of x-axis

\[ J(d') = |e^d|^2 - |e^{d-d'}|^2 \quad \rightarrow \quad \hat{J}(d') = \text{sign}(J(d')) \sqrt{|J(d')|} \]
Different slopes of negative and positive impact values

Mismatch with PROF observations
Distribution of impact values

Histogram of individual observations impact values

Probability distribution

\[ p(J) \sim e^{-\alpha \sqrt{J} + \beta} \Rightarrow \langle J \rangle = \int dJ \ J p(J) = -\frac{2}{\alpha^4} e^{-\alpha \sqrt{J} + \beta} \left( 6 + 6\alpha \sqrt{J} + 3\alpha^2 J + \alpha^3 J^{3/2} \right) \]
Impact per observation type

**Total impact**

![Impact per observation type chart]

- **AIREP, SYNOP, TEMP**
  - Qualitative match between approximation and data denial impact

- **PROF**
  - Bad match between approximation and data denial impact
  - Discrepancy between estimated and smoothed impact hints at insufficient sampling

**Reliability indicator**

|                | AIREP   | PROF    | SYNOP   | TEMP    |
|----------------|---------|---------|---------|---------|
| Unfitted impact| -0.0094 | -0.0216 | -0.0421 | -0.0061 |
| Fitted impact  | -0.0101 | -0.0090 | -0.0433 | -0.0055 |
| Ratio          | 0.93    | 2.39    | 0.972   | 1.11    |

Matthias Sommer and Martin Weissmann

Ensemble-based approximation of observation impact
Cumulative distribution function of observation impact from experiment (green) and fit (blue)

- Extreme values contribute only little to total impact (except for PROF)
Temporal impact distribution

Observation time vs. impact

- Temporally homogeneous distributions (low dependency on forecast time)
- Extreme PROF values during precipitation event
- Low specificity of regions with positive and negative impact
Generally large temperature impact
Small SYNOP wind impact
Anisotropy of wind components impact
Dependency on verification

Verification with conventional observation types

- Each observation group has the largest impact by verification with itself
- Definition of suitable metric including radar and satellite observations

Weighted metric

\[ \tilde{J}_B^A = \alpha_{\text{AIREP}} \frac{J_{\text{TOTAL}}^A}{J_{\text{TOTAL}}^{\text{AIREP}}} J_{\text{AIREP}}^A + \alpha_{\text{PROF}} \frac{J_{\text{TOTAL}}^A}{J_{\text{TOTAL}}^{\text{PROF}}} J_{\text{PROF}}^A + \alpha_{\text{SYNOP}} \frac{J_{\text{TOTAL}}^A}{J_{\text{TOTAL}}^{\text{SYNOP}}} J_{\text{SYNOP}}^A + \alpha_{\text{TEMP}} \frac{J_{\text{TOTAL}}^A}{J_{\text{TOTAL}}^{\text{TEMP}}} J_{\text{TEMP}}^A \]

| Verification norm | AIREP impact | PROF impact | SYNOP impact | TEMP impact |
|------------------|--------------|-------------|--------------|-------------|
| \(J_{25/25/25/25}\) | 23%          | 31%         | 32%          | 13%         |
| \(J_{30/30/30/10}\) | 25%          | 35%         | 31%          | 9%          |
| \(J_{PS}\)        | 37%          | -1%         | 49%          | 16%         |
Signal propagation of AIREP observations

Data denial

\[ t = 0h \]

\[ t = 6h \]
Signal propagation of AIREP observations

Approximation

$t = 0h$

$t = 6h$
Summary

Tool for an approximated assessment of observation impact in an LETKF

- Fast a posteriori estimation of observation impact in a combined analysis and forecasting system
- Modification for the use of observations as verification
- Reliability indication (→ long averaging needed for stable results)
- Limit the approximation to short forecast times because of
  - Linearisation
  - (Static) localization
- Results depend on verification metric

Outlook

- Assessment of impact of more complex observations (Satellites, radar)
- Longer experiment period and operational implementation (DWD)

Literature

Eugenia Kalnay, Yoichiro Ota, Takemasa Miyoshi, and Junjie Liu. A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus A*, 64, 2012. ISSN 1600-0870. URL http://www.tellusa.net/index.php/tellusa/article/view/18462.

Junjie Liu and Eugenia Kalnay. Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 134(634):1327–1335, 2008. ISSN 1477-870X. doi: 10.1002/qj.280. URL http://dx.doi.org/10.1002/qj.280.

Matthias Sommer and Martin Weissmann. Observation impact in a convective-scale localized ensemble transform Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2672–2679, 2014. ISSN 1477-870X. doi: 10.1002/qj.2343. URL http://dx.doi.org/10.1002/qj.2343.

Matthias Sommer and Martin Weissmann. Estimating observation impact using an observation-based verification metric. *Tellus A (submitted)*, 2015.