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

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

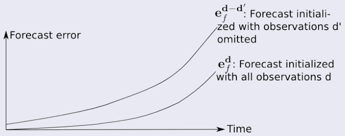
The assessment of observation impact can help. . .

- 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]

Data denial impact of observations $\mathbf{d}'$ relative to all observations $\mathbf{d}$

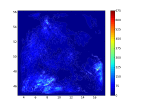


$$J(\mathbf{d}') = |\mathbf{e}_f^{\mathbf{d}}|^2 - |\mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}|^2$$

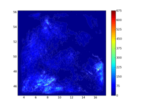
$\mathbf{d}$: *All* available observations

$\mathbf{d}'$: Small subset of observations whose impact one is interested in

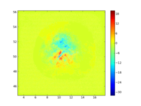
Example



(a) $|\mathbf{e}_f^{\mathbf{d}}|^2$

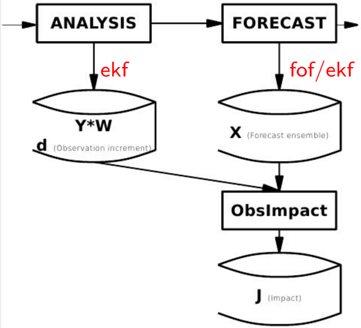


(b) $|\mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}|^2$



(c) $J = \frac{1}{2}\left(|\mathbf{e}_f^{\mathbf{d}}|^2 - |\mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}|^2\right)$

Algorithm



Approximation with ensemble perturbations

LETKF update equation

$$\bar{\mathbf{x}}_{aj} = \mathrm{X}_{bj}\widetilde{\mathrm{P}}_a(j)\mathrm{Y}_b{}^{\intercal}\mathrm{R}^{-1}(j)(\mathbf{y}_o - \bar{\mathbf{y}}_b) + \bar{\mathbf{x}}_{bj}$$

Variables

j : Grid point

$\bar{\mathbf{x}}_a$: Analysis mean

X_b : Background ensemble

$\widetilde{\mathrm{P}}_a$: Ensemble analysis error covariance matrix

$\mathrm{W}^a(j) = \left((K-1)\widetilde{\mathrm{P}}^a(j)\right)^{\frac{1}{2}}$: Weight matrix

Y_b : Background ensemble in observation space

R : Observation error covariance matrix

$\mathbf{d} = \mathbf{y}_o - \bar{\mathbf{y}}_b$: Observational increment

$\bar{\mathbf{x}}_b$: Background mean

Approximation with ensemble perturbations

LETKF update equation

$$\overline{\mathbf{x}}_{aj} = \mathrm{X}_{bj}\widetilde{\mathrm{P}}_a(j)\mathrm{Y}_b{}^\intercal \mathrm{R}^{-1}(j)(\mathbf{y}_o - \overline{\mathbf{y}}_b) + \overline{\mathbf{x}}_{bj}$$

Data denial observation impact

$$J(\mathbf{d}') = |\mathbf{e}_f^{\mathbf{d}}|^2 - |\mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}|^2 = \left(\mathbf{e}_f^{\mathbf{d}} + \mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}\right) \cdot \left(\mathbf{e}_f^{\mathbf{d}} - \mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}\right)$$

Direct derivation [Kalnay et al., 2012]

$$\mathbf{e}_f^{\mathbf{d}} - \mathbf{e}_f^{\mathbf{0}} = \overline{\mathbf{x}}_f^{\mathbf{d}} - \overline{\mathbf{x}}_f^{\mathbf{0}} \approx \frac{1}{K-1}\mathrm{X}_f^{\mathbf{d}}(\mathrm{Y}_b\mathrm{W}^{\mathbf{d}})^\intercal \mathrm{R}^{-1}\mathbf{d}$$

$$\Rightarrow J(\mathbf{d}')$$

$$= \left(\mathbf{e}_f^{\mathbf{d}} + \mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}\right) \cdot \left(\mathbf{e}_f^{\mathbf{d}} - \mathbf{e}_f^{\mathbf{0}} - \left(\mathbf{e}_f^{\mathbf{d}-\mathbf{d}'} - \mathbf{e}_f^{\mathbf{0}}\right)\right)$$

$$\approx \left(\mathbf{e}_f^{\mathbf{d}} + \mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}\right) \cdot \left(\frac{1}{K-1}\mathrm{X}_f^{\mathbf{d}}(\mathrm{Y}_b\mathrm{W}^{\mathbf{d}})^\intercal \mathrm{R}^{-1}\mathbf{d}'\right)$$

$$\approx \left(\mathbf{e}_f^{\mathbf{d}} + \mathbf{e}_f^{\mathbf{0}}\right) \cdot \left(\frac{1}{K-1}\mathrm{X}_f^{\mathbf{d}}(\mathrm{Y}_b\mathrm{W}^{\mathbf{d}})^\intercal \mathrm{R}^{-1}\mathbf{d}'\right)$$

Approximation with ensemble perturbations

LETKF update equation

$$\bar{\mathbf{x}}_{aj} = \mathrm{X}_{bj}\widetilde{\mathrm{P}}_a(j)\mathrm{Y}_b{}^\intercal \mathrm{R}^{-1}(j)(\mathbf{y}_o - \bar{\mathbf{y}}_b) + \bar{\mathbf{x}}_{bj}$$

Data denial observation impact

$$J(\mathbf{d}') = |\mathbf{e}_f^{\mathbf{d}}|^2 - |\mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}|^2 = \left(\mathbf{e}_f^{\mathbf{d}} + \mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}\right) \cdot \left(\mathbf{e}_f^{\mathbf{d}} - \mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}\right)$$

Direct derivation [Kalnay et al., 2012]

$$\mathbf{e}_f^{\mathbf{d}} - \mathbf{e}_f^{\mathbf{0}} = \bar{\mathbf{x}}_f^{\mathbf{d}} - \bar{\mathbf{x}}_f^{\mathbf{0}} \approx \frac{1}{K-1}\mathrm{X}_f^{\mathbf{d}}(\mathrm{Y}_b\mathrm{W}^{\mathbf{d}})^\intercal \mathrm{R}^{-1}\mathbf{d}$$

$$\Rightarrow J(\mathbf{d}')$$

$$= \left(\mathbf{e}_f^{\mathbf{d}} + \mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}\right) \cdot \left(\mathbf{e}_f^{\mathbf{d}} - \mathbf{e}_f^{\mathbf{0}} - \left(\mathbf{e}_f^{\mathbf{d}-\mathbf{d}'} - \mathbf{e}_f^{\mathbf{0}}\right)\right)$$

$$\approx \left(\mathbf{e}_f^{\mathbf{d}} + \mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}\right) \cdot \left(\frac{1}{K-1}\mathrm{X}_f^{\mathbf{d}}(\mathrm{Y}_b\mathrm{W}^{\mathbf{d}})^\intercal \mathrm{R}^{-1}\mathbf{d}'\right)$$

$$\approx \left(\mathbf{e}_f^{\mathbf{d}} + \mathbf{e}_f^{\mathbf{0}}\right) \cdot \left(\frac{1}{K-1}\mathrm{X}_f^{\mathbf{d}}(\mathrm{Y}_b\mathrm{W}^{\mathbf{d}})^\intercal \mathrm{R}^{-1}\mathbf{d}'\right)$$

Taylor expansion [Sommer and Weissmann, 2015]

$$J(\mathbf{d}') = J(\mathbf{0}) + \left.\frac{d}{d\mathbf{d}'}\right|_{\mathbf{d}'=\mathbf{0}} J(\mathbf{d}')\mathbf{d}' + \mathcal{O}\left(|\mathbf{d}'|^2\right)$$

$$= 2\mathbf{e}_f^{\mathbf{d}} \cdot \left(-\left.\frac{d}{d\mathbf{d}'}\right|_{\mathbf{d}'=\mathbf{0}} \mathbf{e}_f^{\mathbf{d}-\mathbf{d}'}\right)\mathbf{d}' + \mathcal{O}\left(|\mathbf{d}'|^2\right)$$

$$= 2\mathbf{e}_f^{\mathbf{d}} \cdot \left(\left.\frac{d}{d\mathbf{d}'}\right|_{\mathbf{d}'=\mathbf{d}} \bar{\mathbf{x}}_f^{\mathbf{d}'}\right)\mathbf{d}' + \mathcal{O}\left(|\mathbf{d}'|^2\right)$$

$$\approx 2\mathbf{e}_f^{\mathbf{d}} \cdot \left(\frac{1}{K-1}\mathrm{X}_f^{\mathbf{d}}(\mathrm{Y}_b\mathrm{W}^{\mathbf{d}})^\intercal \mathrm{R}^{-1}\mathbf{d}'\right)$$

...analysis [Kalnay et al., 2012]

$$\mathbf{e}_f = \overline{\mathbf{x}_f} - \mathbf{x}_a$$

$$|\mathbf{e}_f|^2 = \sum_{gridpoints} \frac{1}{2}(\overline{\mathbf{u}}_f - \overline{\mathbf{u}}_a)^2 + \frac{1}{2}(\overline{\mathbf{v}}_f - \overline{\mathbf{v}}_a)^2$$

$$\Rightarrow J(\mathbf{d}') \approx 2\mathbf{e}_f^{\mathbf{d}} \cdot \left(\frac{1}{K-1} \mathrm{X}_f^{\mathbf{d}} (\mathrm{Y}_b \mathrm{W}^{\mathbf{d}})^{\mathsf{T}} \mathrm{R}^{-1} \mathbf{d}'\right)$$

\+ Homogeneous in space and time

\- Strongly correlated to forecast

...observations [Sommer and Weissmann, 2015]

$$\mathbf{e}_f = H(\overline{\mathbf{x}_f}) - \mathbf{y}_o$$

$$|\mathbf{e}_f|^2 = \sum_{observations} \left(\frac{H(\overline{\mathbf{x}_f}) - \mathbf{y}_o}{\sigma}\right)^2$$

$$\Rightarrow J(\mathbf{d}') \approx 2\mathbf{e}_f^{\mathbf{d}} \cdot \left(\frac{1}{K-1} \mathrm{Y}_f^{\mathbf{d}} (\mathrm{Y}_b \mathrm{W}^{\mathbf{d}})^{\mathsf{T}} \mathrm{R}^{-1} \mathbf{d}'\right)$$

\+ Independent of forecast

\+ Computationally easy

\- Unobserved regions/variables may be ignored

DWD Convective-scale assimilation and forecasting systems

Kilometer-scale **E**nsemble **D**ata **A**ssimilation (KENDA)

- **L**ocalized **E**nsemble **T**ransform **K**alman **F**ilter for use with COSMO-DE (in development)

Consortium for **S**mall-scale **M**odelling (COSMO)

- Operational limited-area model of Deutscher Wetterdienst
- Grid point model of non-hydrostatic equations
- Horizontal resolution: 2.8 km; 50 vertical levels

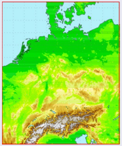


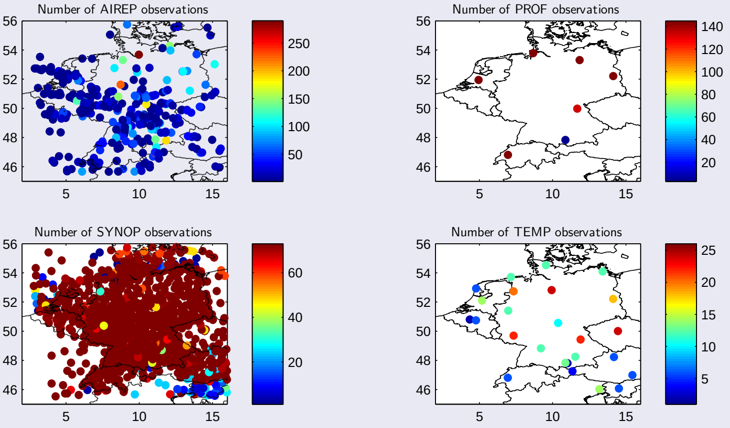
Figure : COSMO-DE domain ($\approx$ 1300 km $\times$ 1200 km)

Experimental settings

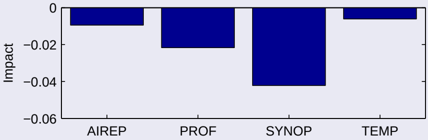
- Test period: 10 June 2012 12:00 UTC – 13 June 2012 15:00 UTC
- Initialization every 3 h
- Forecast length 6 h
- 40-members ensemble
- Observations used:
  - AIREP (Aircrafts): U, V, T
  - PROF (Wind profiler): U, V
  - SYNOP (Ground stations): U, V, T, RH
  - TEMP (Weather Balloons): U, V, T, RH

Number of assimilated observations per station

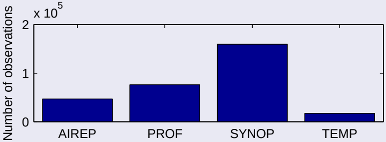
10 June 2012 12:00 UTC – 13 June 2012 15:00 UTC



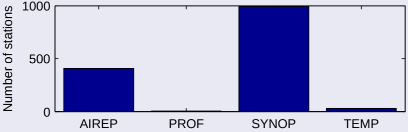
Total impact



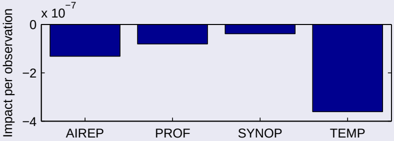
Number of observations



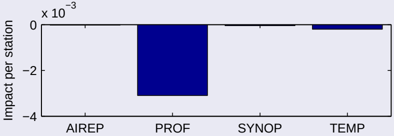
Number of stations



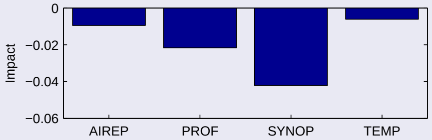
Impact per observation



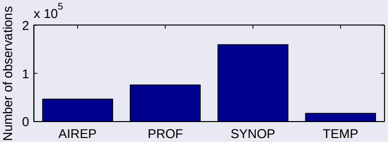
Impact per station



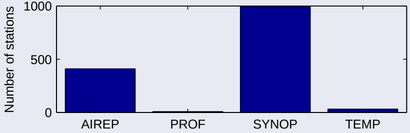
Total impact



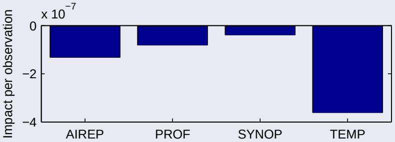
Number of observations



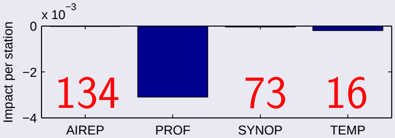
Number of stations



Impact per observation

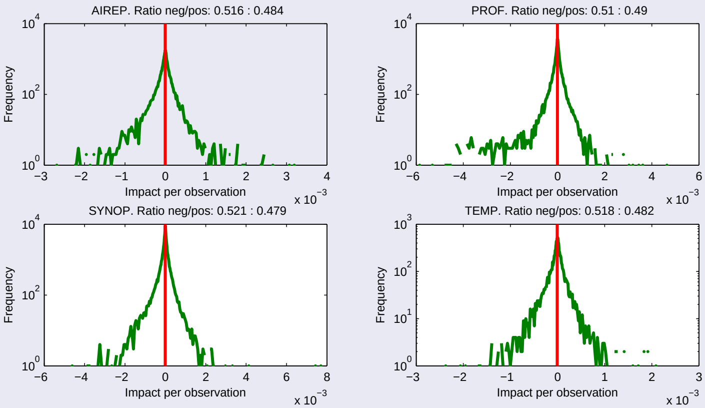


One wind profiler equivals...



Distribution of impact values

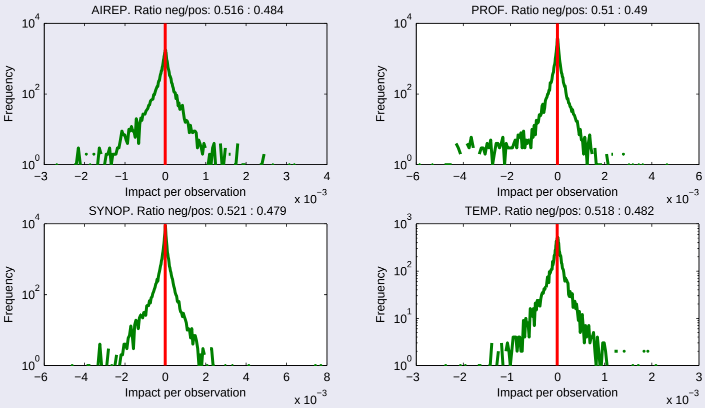
Histogram of individual observations impact values



- Non-Gaussian distribution
- Ratio of negative to positive values ca. 52:48
- Width of distribution >> Mean

Distribution of impact values

Histogram of individual observations impact values

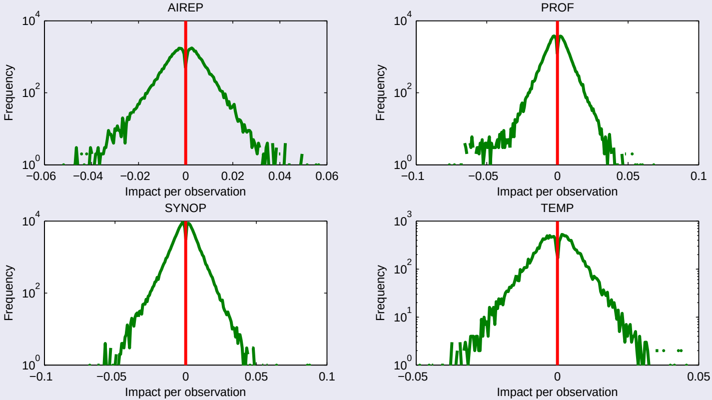


Transformation of x-axis

$$J(\mathbf{d}') = |\mathbf{e}^{\mathbf{d}}|^2 - |\mathbf{e}^{\mathbf{d}-\mathbf{d}'}|^2 \quad \rightarrow \quad \widehat{J}(\mathbf{d}') = \text{sign}(J(\mathbf{d}'))\sqrt{|J(\mathbf{d}')|}$$

Distribution of impact values

Histogram of individual observations impact values

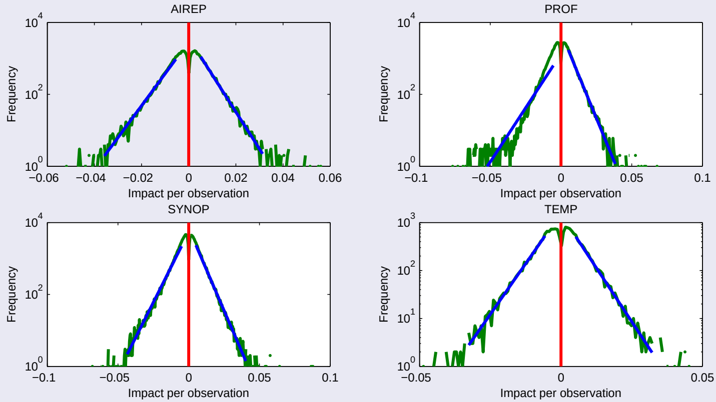


Transformation of x-axis

$$J(\mathbf{d}') = |\mathbf{e}^{\mathbf{d}}|^2 - |\mathbf{e}^{\mathbf{d}-\mathbf{d}'}|^2 \quad \rightarrow \quad \widehat{J}(\mathbf{d}') = \mathrm{sign}(J(\mathbf{d}'))\sqrt{|J(\mathbf{d}')|}$$

Distribution of impact values

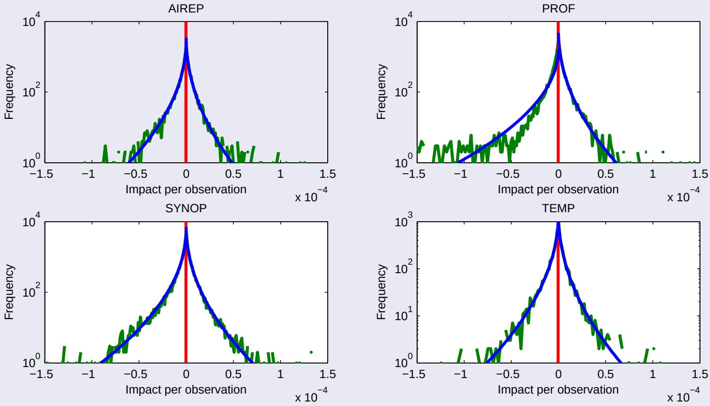
Histogram of individual observations impact values



- Different slopes of negative and positive impact values
- Mismatch with PROF observations

Distribution of impact values

Histogram of individual observations impact values

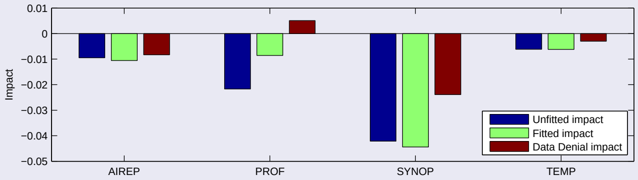


Probability distribution

$$p(J) \sim e^{-\alpha\sqrt{J}+\beta} \Rightarrow \langle J \rangle = \int \mathrm{d}J\, Jp(J) = -\frac{2}{\alpha^4} e^{-\alpha\sqrt{J}+\beta} \left(6 + 6\alpha\sqrt{J} + 3\alpha^2 J + \alpha^3 J^{\frac{3}{2}}\right)$$

Impact per observation type

Total impact



AIREP, SYNOP, TEMP

- Qualitative match between approximation and data denial impact

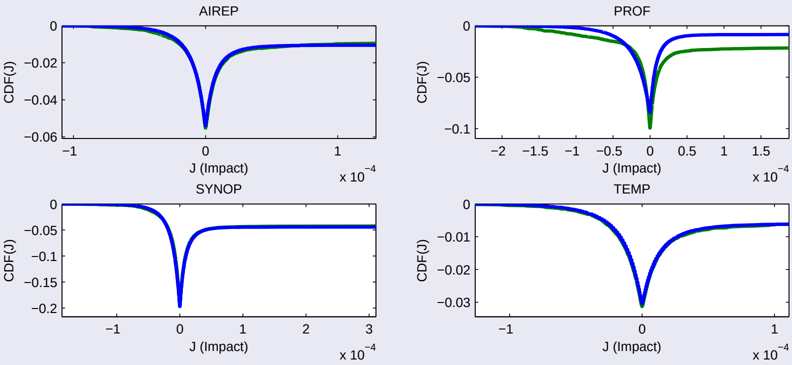
PROF

- 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

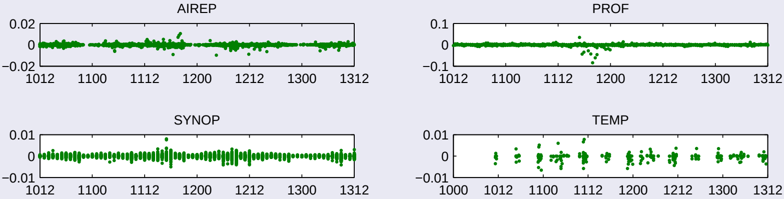
Cumulative distribution function of observation impact from experiment (green) and fit (blue)



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

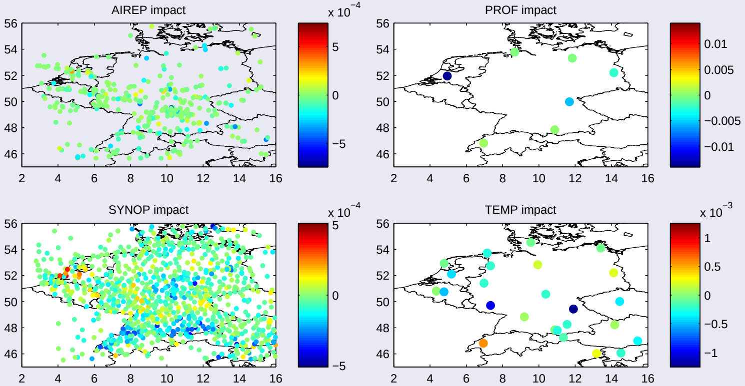
Temporal impact distribution

Observation time vs. impact



- Temporally homogeneous distributions (low dependency on forecast time)
- Extreme PROF values during precipitation event

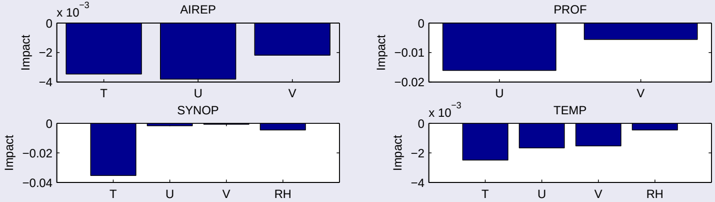
Impact per ident



- Low specificity of regions with positive and negative impact

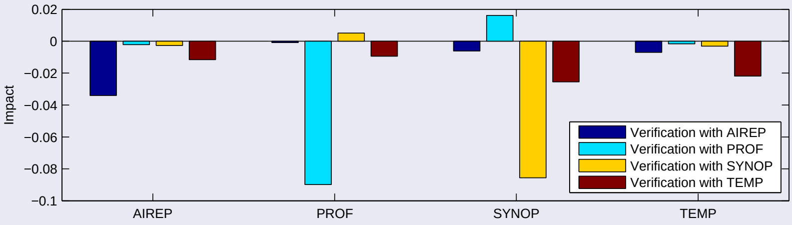
Impact per observed variable

Normalized with number of observations



- Generally large temperature impact
- Small SYNOP wind impact
- Anisotropy of wind components impact

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

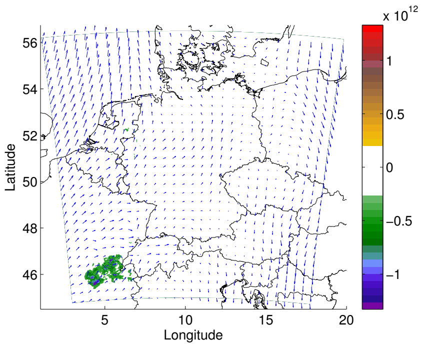
J_B^A : Impact of A when verified with B

$$\tilde{J}_\alpha^A = \frac{\alpha_{\text{AIREP}}}{J_{\text{AIREP}}^{\text{TOTAL}}} J_{\text{AIREP}}^A + \frac{\alpha_{\text{PROF}}}{J_{\text{PROF}}^{\text{TOTAL}}} J_{\text{PROF}}^A + \frac{\alpha_{\text{SYNOP}}}{J_{\text{SYNOP}}^{\text{TOTAL}}} J_{\text{SYNOP}}^A + \frac{\alpha_{\text{TEMP}}}{J_{\text{TEMP}}^{\text{TOTAL}}} J_{\text{TEMP}}^A$$

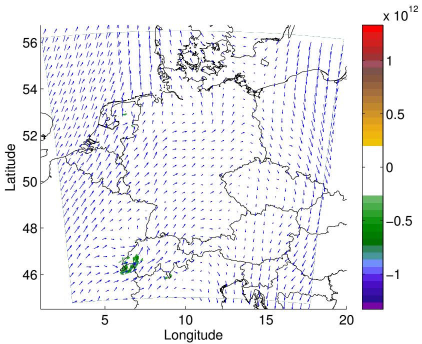
Verification norm	AIREP impact	PROF impact	SYNOP impact	TEMP impact
$J_{25/25/25/25}$	23%	31%	32%	13%
$J_{30/30/30/10}$	25%	35%	31%	9%
J_{PS}	37%	-1%	49%	16%

Data denial

$t = 0\text{h}$

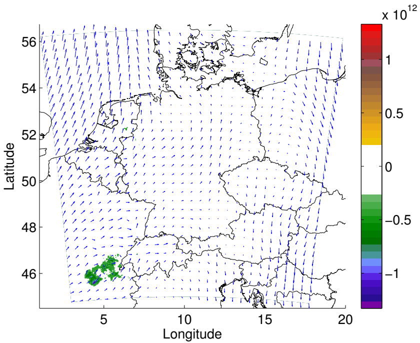


$t = 6\text{h}$

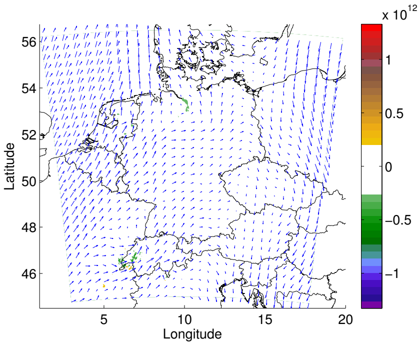


Approximation

$t = 0h$



$t = 6h$



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 obeservation impact using an observation-based verification metric. *Tellus A (submitted)*, 2015.