Deutscher Wetterdienst
Wetter und Klima aus einer Hand

# Observation impact diagnostics in an Ensemble Data Assimilation System

## 1. Introduction: Forecast impact diagnostics developed so far

### Motivation

Assessing the impact of observations in data assimilation systems is generally **extremely expensive**
- denial experiments
- observation system experiments (OSE).

**Aim**: Find diagnostic tools to
- indicate impact of observation subsets on analysis/forecast
- identify where observation impact is sub-optimal/negative

$$J \stackrel{\text{def!}}{=} \frac{1}{2}\left(\left\|e^{t|a}\right\|^2 - \left\|e^{t|b}\right\|^2\right) \qquad e^{t|a} = y^{t|a} - y^{v;t}, \quad e^{t|b} = y^{t|b} - y^{v;t}$$

$y^{v;t}$ : data used for verification (at time $t$)

$y^{t|a}, y^{t|b}$ : modelequivalent to $y^{v;t}$ based on forecast from analysis $x^a \rightarrow y^{t|a}$, background $x^b \rightarrow y^{t|b}$

$$\left\|e^{t|a}\right\| = \left(e^{t|a}\right)^T C \left(e^{t|a}\right) \quad \text{"scalar product with metric } C\text{"}$$

Cost-function *J*
- gives impact of obs assimilated at time *t* = 0
- can be written as **sum over the individual observations** $y_l^o$

$$J = \sum_{l\in\{obs\}} \frac{1}{2}\left\langle\left(e^{t|a} + e^{t|b}\right)^T C H_v M_{0\to t} K_l \left(y_l^o - y_l^b\right)\right\rangle = \sum_{l\in\{obs\}} \frac{1}{2}\left(J_l^{t|a} + J_l^{t|b}\right)$$

$K_l$ : column of Kalman gain matrix acting on $y_l^o$

$M_{0\to t}$ : time evolution operator

$H_v$ : operator computing modelequivalent to $y^{v;t}$

Time evolution $M_{0\to t}$ :
Different DA systems use different approaches for computing the time evolution

4D Var (Langland and Barker 2004)
→ use linear (adjoint) model

Ensemble Kalman Filter (EKF)
→ use ensemble

Here : Ensemble Kalman Filter
Verification with observations

## 2. Optimality condition

Consider the cost function *J* for different initial conditions

$$y^{t|init}(\{\lambda_l\}) = y^t\left[x^{init}(\{\lambda_l\})\right]$$
$$x^{init}(\{\lambda_l\}) = x^b + \sum_{l\in\{obs\}} K_l \lambda_l \left(y_l^o - y_l^b\right)$$
$$x^a = x^{init}(\{\lambda_l = 1\})$$

If *J* has a minimum for $x^{init} = x^a$ one finds

$$\left\langle J_l^{t|a}\right\rangle \equiv \left\langle\left(e^{t|a}\right)^T C H_v M_{0\to t} K_l \left(y_l^o - y_l^b\right)\right\rangle = 0 \quad \text{for all } l \in \{obs\}$$

$\langle .. \rangle$ : statistical mean

## 3. Our System

- LETKF (Hunt et al. 2007) 40 ensemble members
- verification vs obs. (Sommer and Weissmann 2014)
- time evolution via analysis ensemble (Kalnay et al. 2012)

$y^{v;t} = y_v^{o;t}$ observation $y_v^o$ at time $t$

$H_v$ : observation operator for $y_v^o$

$C = R_{vv}^{-1} = \frac{1}{r_v}$

$$\left(x^a - x^b\right) = \sum_k w_k^a X^{a(k)}$$
$$M_{0\to t}\left(x^a - x^b\right) = \sum_k w_k^a X^{a(k)}(t)$$

$X^{a(k)}$ : $k^{th}$ incr. analysis ensemble member

weights are computed as:

$$w_k^a = \sum_{\alpha,\beta\in\{obs\}} Y_\alpha^{a(k)} R_{\alpha\beta}^{-1}\left(y_\beta^o - y_\beta^b\right)$$
$$Y_\alpha^{a(k)} = H_\alpha X^{a(k)} \quad : X^{a(k)} \text{ in obs space}$$

$$P_{vl}^a(t) = \sum_{k=1}^{n_{ens}}\left(Y_v^{a(k)}(t) Y_l^{a(k)}\right)\rho_{loc}[v,l]$$

analysis covariance matrix in obs space (estimated from ensemble)

⇒ (For simplicity *R* diagonal)

$$J_l^{t|a} = -P_{vl}^a(t)\frac{\left(y_v^{o;t} - y_v^{t|a}\right)\left(y_l^o - y_l^b\right)}{R_{vv}R_{ll}}$$
$$J_l^{t|b} = -P_{vl}^a(t)\frac{\left(y_v^{o;t} - y_v^{t|b}\right)\left(y_l^o - y_l^b\right)}{R_{vv}R_{ll}}$$

## 4. Results - What is shown?

So far only results for t=0 ("impact on analysis").
Statistics have been computed for different cost-function components separately:

$$J = \frac{1}{2}\left(J^{t|a} + J^{t|b}\right)$$

$$\sum_{stats}\left(-J_l^{0|a}\right) = \sum_{stats} P_{vl}^a \frac{\left(y_v^o - y_v^a\right)\left(y_l^o - y_l^b\right)}{R_{ll}R_{vv}}$$ ← should be small --- "optimality condition"

$$\sum_{stats}\left(-J_l^{0|b}\right) = \sum_{stats} P_{vl}^a \frac{\left(y_v^o - y_v^b\right)\left(y_l^o - y_l^b\right)}{R_{ll}R_{vv}}$$ ← should be positive (and large) --- "potential benefit"

.

- impact variable : $\sum\left(-J^{0|a}\right)$ or $\sum\left(-J^{0|b}\right)$
- "normalised" : $\sum\left(-J^{0|a}\right) / \left(\sum \frac{|P_{vl}^a|}{\sqrt{R_{vv}R_{ll}}}\right)$ $\quad$ $\sum\left(-J^{0|b}\right) / \left(\sum \frac{|P_{vl}^a|}{\sqrt{R_{vv}R_{ll}}}\right)$
- impact variable : bias removed from bins
- ⟨uncorr. noise⟩ : $\sqrt{\sum\left(-J^{0|a}\right)^2}$ $\quad$ $\sqrt{\sum\left(-J^{0|b}\right)^2}$
- effective number : $\left(\sum|P_{vl}^a|\right)^2 / \sum|P_{vl}^a|^2$

## 6. Outlook:
## Assessing impact of individual observations

**Work so far:**
Cost-function *J* gives **impact of all observations** assimilated at time *t* = 0.
Interpretation of different components (corresponding to individual observations) is *suggestive but not rigorous* ($[x^a - x^b]$ depends on all observations used in assimilation).

**More rigorous approach:**
Cost-function for **data denial** experiment. Replace: $x^b \rightarrow x^{a/l}$ **(analysis not using** $y_l^o$**)**

⇒

$$J_l^{t|a} = -P_{vl}^a(t)\frac{\left(y_l^o - y_l^{a/k}\right)\left(y_v^{o;t} - y_v^{t|a}\right)}{R_{ll}R_{vv}}$$
$$J = J_l^{t|a} - \frac{1}{2}\left[P_{vl}^a(t)\right]^2 \frac{\left(y_l^o - y_l^{a/k}\right)^2}{R_{ll}^2 R_{vv}}$$

with

$$\left(y_l^o - y_l^{a/k}\right) = \left(1 - H_v K_l\right)^{-1}\left(y_l^o - y_l^a\right) = \frac{\left(y_l^o - y_l^a\right)}{1 - P_{ll}^a(t=0)/R_{ll}}$$

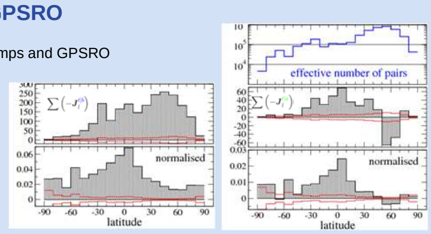
## 5. Results: Impact on analysis (t=0)

### Temps verified by GPSRO

Excellent correspondence/consistency between Temps and GPSRO

$\sum\left(-J_l^{0|b}\right)$ : clearly positive everywhere

$\sum\left(-J_l^{0|a}\right)$ : is 1.) much smaller than $\sum\left(-J_l^{0|b}\right)$
→optimality condition largely fulfilled
2.) mostly positive
→ weight on TEMPS in assimilation could be slightly increased in tropical regions



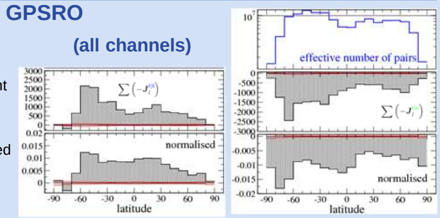

### AMSU-A verified by GPSRO

**(all channels)**

$\sum\left(-J_l^{0|b}\right)$: generally positive → strong potential

$\sum\left(-J_l^{0|a}\right)$: clearly negative → obs have too strong weight
AMSU-A observation errors are too small
(also according to Desroziers diagnostics)

Increased observation errors have been tested for operational implementation.
But: Positive impact on forecast only after reduced thinning of AMSU-A




### Data from individual AMSU-A channels ($\sum\left(-J_l^{0|a}\right)$ - normalised)

The correspondence between AMSU-A channels and GPSRO is positive or neutral for most channels and most latitudes ($\sum\left(-J_l^{0|b}\right)$ is mostly positive).

In some regions the dark shaded areas are negative or neutral but the green curve (bias removed a posteriori) is clearly positive. This indicates that **AMSU and GPSRO** have **opposite bias**.

Significant bias problems occur for :

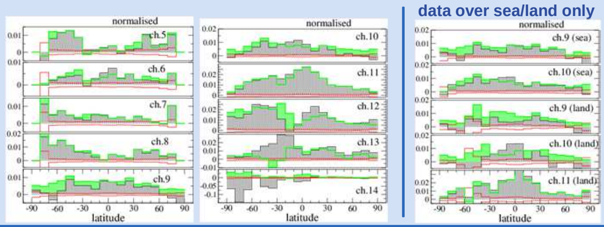
(i) ch.14 (everywhere)

(ii) chs. 9+10 (esp.towards poles).

In both cases (i) and (ii), GPSRO has *strong bias vs model*. Most likely the GPSRO is correct → **model bias**.

Ch. 9+10 are the lowest channels used over land.
As seen (on the right below) bias problems are much stronger over land than over sea.
(Model biases are different over land and sea).

**data over sea/land only**




## 7. Discussion

The interpretation of observation impact diagnostics is often not trivial. Statistical significance is a big issue (particularly for large forecast lead times). To facilitate the interpretation and to differentiate between model and DA issues the work presented here has been (so far) restricted to t=0 (impact on the analysis).

It is explained that different parts of the cost function $J = \frac{1}{2}\left(J^{t|a} + J^{t|b}\right)$ should be considered (interpreted) separately.
Examples are given for how the diagnostics could be linked to the following observational problems:
- The use of too small observation errors in the DA system for AMSU-a
- Biases of AMSU-A channels ←→ Bias of GPSRO

Bias problems only show up if the bias is opposite to the bias which the verifying obs (here GPSRO) have with respect to the model.
The diagnostics reveal inconsistencies. The separation into contributions from different observations is, however, not rigorous. Particularly for strongly overlapping observations the interpretation in terms of impact (on analysis or forecast) is problematic. A method is under development to show the "denial impact" for individual observations (e.g., a single AMSU channel). See sec. **6.**.


**References**
- R. H. Langland and N. L. Baker. Estimation of observation impact using the nrl atmospheric variational data assimilation adjoint system. Tellus A, 56(3):189–201 (2004).
- E. Kalnay, Y. Ota, T. Miyoshi, and J. Liu. A simpler formulation of forecast sensitivity to obser-vations: application to ensemble kalman filters. Tellus A: Dynamic Meteorology and Oceanog-raphy, 64(1):18462 (2012).
- M. Sommer and M. Weissmann. Observation impact in a convective-scale localized ensemble transform kalman filter. Quarterly Journal of the Royal Meteorological Society, 140(685):2672–2679 (2014).
- B. R. Hunt, E. J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatio temporal chaos: A local ensemble transform kalman filter. Physica D: Nonlinear Phenomena, 230(1):112–126 (2007).



**DWD, German Meteorological Service, Data assimilation section**
**Author: Olaf Stiller, Olaf.Stiller@dwd.de**