Using ensemble based diagnostics to identify sub-optimally used observations

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Olaf Stiller, DWD, German Meteorological Service, Data assimilation section

1. Introduction: *How we use EFSOI type statistics*

Aim

Identify subgroups of observations whose impact is negative/sub-optimal

Method

i. Use verification function ***J*** as a measure for the impact of the observations
ii. write ***J*** as a sum over contributions from the individual observations
iii. take statistics over predefined observation subgroups (to test their impact)

In this work

A) We consider 2 cases (2 types of functions ***J***)
1. „usual case“ (comp., e.g., Langland and Baker 2004) impact of obs in full analysis (proxy for denial exp.)
2. single obs case impact if *one obs is assimilated alone*

B) The function ***J*** is written in terms of two parts indicating primarily:
I. the consistency of obs and covariance matrixes
II. problems with the analysis(weigths, localisation, etc)

using obs -- not using obs

$$J \overset{\text{def!}}{=} \frac{1}{2}\left(\left\|\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|a}_{\mathrm{v}}\right\|^2_C - \left\|\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right\|^2_C\right)$$
$$= -\frac{1}{2}\left(\left(\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|a}_{\mathrm{v}}\right) + \left(\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)\right) C \left(\boldsymbol{y}^{t|a}_{\mathrm{v}} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)$$

$$\left(\boldsymbol{y}^{t|a}_{\mathrm{v}} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right) = \boldsymbol{H}_{\mathrm{v}} M_{0\to t}\mathbf{K}\left(\boldsymbol{y}^o - \boldsymbol{y}^b\right)$$

$M_{0\to t}$: time evolution operator
$\boldsymbol{H}_{\mathrm{v}}$: operator computing **model equivalent** to $\boldsymbol{y}^{o;t}_{\mathrm{v}}$

$\boldsymbol{y}^{o;t}_{\mathrm{v}}$: verifying observation at time t
$\boldsymbol{y}^{t|a}_{\mathrm{v}}$: forecast starting from analysis
$\boldsymbol{y}^{t|b}_{\mathrm{v}}$: forecast starting from background

2 ways to write Kalman gain matrix K :

$$\begin{array}{llll} \text{I)} & \mathbf{K} = & P^a\boldsymbol{H}^T & \boldsymbol{R}^{-1} \\ \text{II)} & \phantom{\mathbf{K}} = & P^b\boldsymbol{H}^T & \left[\boldsymbol{H}P^b\boldsymbol{H}^T + \boldsymbol{R}\right]^{-1} \end{array}$$

$$\boldsymbol{H}_{\mathrm{v}} M_{0\to t}\mathbf{K} = \boldsymbol{H}_{\mathrm{v}} M_{0\to t} P\boldsymbol{H}^T \; \mathfrak{G}^{-1} = P[\mathbf{v}, t;\alpha]\; \mathfrak{G}^{-1}$$

Impact function *J* can be written with two components:

$$J = -\frac{1}{2}\sum_{\alpha\in obs}\left(J^{t|b}_\alpha + \left\{J^{t|b}_\alpha - J^{t|ab}_\alpha\right\}\right)$$
$$\boldsymbol{J}^{t|b}_\alpha = \left(\boldsymbol{y}^{o;t}_{\mathrm{v}} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right) C \left(\sum_\beta P[\mathbf{v}, t;\beta]\, \mathfrak{G}^{-1}_{\beta\alpha}\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)\right)$$
$$\boldsymbol{J}^{t|ab}_\alpha = \left(\boldsymbol{y}^{t|a}_{\mathrm{v}} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right) C \left(\sum_\beta P[\mathbf{v}, t;\beta]\, \mathfrak{G}^{-1}_{\beta\alpha}\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)\right)$$

I) $P \to P^a$, $\mathfrak{G} \to \boldsymbol{R}$

II) $P \to P^b$, $\mathfrak{G} \to \left[\boldsymbol{H}P^b\boldsymbol{H}^T + \boldsymbol{R}\right]$

Using the Ensemble
Estimating the covariance matrix:

$$P^a[\mathbf{v}, t;\alpha] = (N_{ens}-1)^{-1}\sum_{l=1}^{N_{ens}}\left(\{\boldsymbol{Y}^a_l\}_{\mathrm{v}}(t)\,\{\boldsymbol{Y}^a_l\}_\alpha\right) * \eta_{loc}$$
$$P^b[\mathbf{v}, t;\alpha] = (N_{ens}-1)^{-1}\sum_{l=1}^{N_{ens}}\left(\{\boldsymbol{Y}^b_l\}_{\mathrm{v}}(t)\,\{\boldsymbol{Y}^b_l\}_\alpha\right) * \eta_{loc}$$
$$\{\boldsymbol{Y}^a_l\}_{\mathrm{v}}(t) = \boldsymbol{H}_{\mathrm{v}}\boldsymbol{X}^a_l(t)$$
$$\{\boldsymbol{Y}^b_l\}_{\mathrm{v}}(t) = \boldsymbol{H}_{\mathrm{v}}\boldsymbol{X}^b_l(t)$$

X_l (t) : incr. ensemble member „***l***“

What we compute:

- verification against observations
- metric ***C*** → ***R***$^{-1}$ (diagonal!!)

I) “usual case” impact on full analysis

$$\boldsymbol{J}^{t|b}_\alpha = P^a[\mathbf{v}, t;\alpha]\,\frac{\left(\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)}{\boldsymbol{R}_{\mathrm{vv}}\boldsymbol{R}_{\alpha\alpha}}$$
$$\boldsymbol{J}^{t|ab}_\alpha = P^a[\mathbf{v}, t;\alpha]\,\frac{\left(\boldsymbol{y}^{t|a}_{\mathrm{v}} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)}{\boldsymbol{R}_{\mathrm{vv}}\boldsymbol{R}_{\alpha\alpha}}$$

II) single obs case assimilation of 1 obs only

$$\boldsymbol{J}^{t|b}_\alpha = P^b[\mathbf{v}, t;\alpha]\,\frac{\left(\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)}{\boldsymbol{R}_{\mathrm{vv}}\left(\boldsymbol{P}^b_{\alpha\alpha} + \boldsymbol{R}_{\alpha\alpha}\right)}$$
$$\boldsymbol{J}^{t|ab}_\alpha = P^b[\mathbf{v}, t;\alpha]\,\frac{\left(\boldsymbol{y}^{t|a}_{\mathrm{v}} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)}{\boldsymbol{R}_{\mathrm{vv}}\left(\boldsymbol{P}^b_{\alpha\alpha} + \boldsymbol{R}_{\alpha\alpha}\right)}$$

2. Optimality condition

When writing the verification function ***J*** as a function of the initial model state $\boldsymbol{x}^{init}$,

it can be shown that the **optimality condition** $\left\langle J^{t|ab}_\alpha\right\rangle \simeq \left\langle J^{t|b}_\alpha\right\rangle$ holds if either:

A. $\boldsymbol{J}(\boldsymbol{x}^{init})$ has a minimum when $\boldsymbol{x}^{init}$ is the analysis
Or:
B. i) All first guess departures are bias free and

ii) $$\sum_{\gamma,\beta\in obs} P^b[\mathbf{v}, t;\gamma]\;\underbrace{\left(\boldsymbol{P}^b_{\gamma\beta} + \boldsymbol{R}_{\gamma\beta}\right)^{-1}\left\langle\left(\boldsymbol{y}^o_\beta - \boldsymbol{y}^b_\beta\right)\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)\right\rangle}_{\delta_{\gamma\alpha}} = \left\langle\underbrace{\left(\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)}_{P^b[\mathbf{v}, t;\alpha]}\right\rangle$$

(***P***^{b}+***R***) = Cov. matrix(obs –fg)

Ensemble Covariance verifying obs ←→ analysis obs

3. Can we check consistency of model covariances and observations more directly?

Single obs case:

$$\left\langle\left[\frac{\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)^2}{\left(\boldsymbol{P}^b_{\alpha\alpha} + \boldsymbol{R}_{\alpha\alpha}\right)}\right]\right\rangle P^b[\mathbf{v}, t;\alpha] = \left\langle\left(\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)\right\rangle$$

Comparing estim. covariances

model vs obs:

Agreement is

- generally good to excellent for local obs.
- slightly more problematic for GPSRO and particularly for satellite radiances

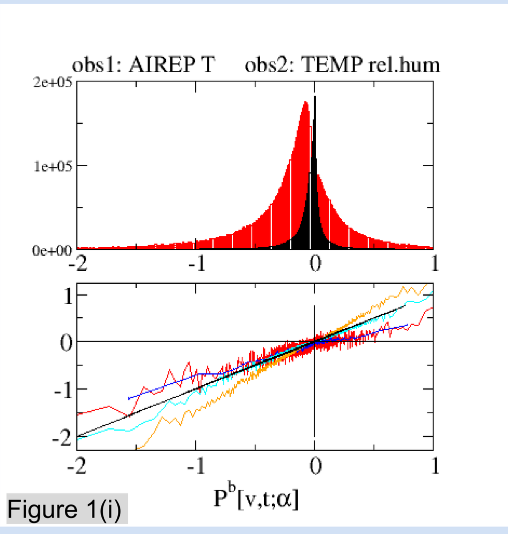

Figure 1(i)

All quantities scaled by obs errors

Histograms:
black : nb (number of points in equidistant bins)
red : nb * P^b (arbitrary units)

Taking averages:
1) order all points $\left(\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)$
2) take running mean

aver.length 500
aver.length 5000
$\left\langle\left(\boldsymbol{y}^{o;t}_{\mathrm{v};} - \boldsymbol{y}^{t|b}_{\mathrm{v}}\right)\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)\right\rangle$

obs 1 → α
obs 2 → α
$\left\langle\left[\frac{\left(\boldsymbol{y}^o_\alpha - \boldsymbol{y}^b_\alpha\right)^2}{\left(\boldsymbol{P}^b_{\alpha\alpha} + \boldsymbol{R}_{\alpha\alpha}\right)}\right]\right\rangle P^b[\mathbf{v}, t;\alpha]$

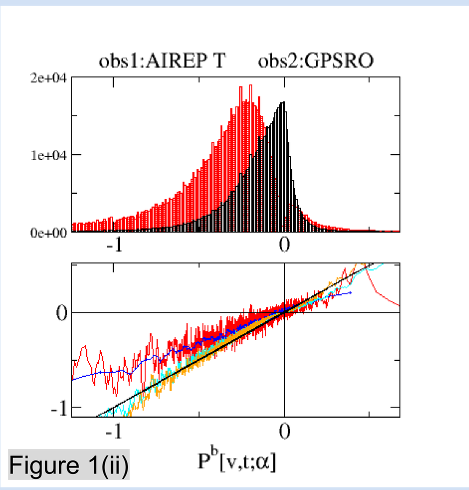

Figure 1(ii)

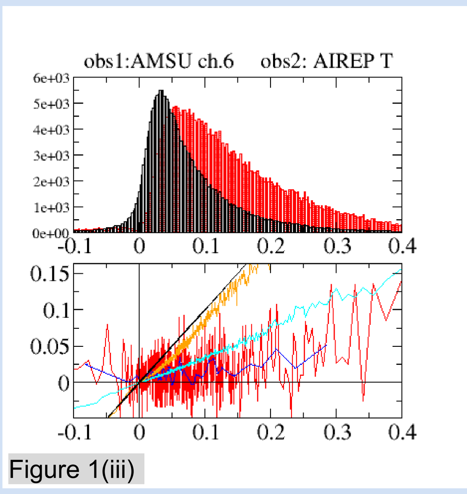

Figure 1(iii)

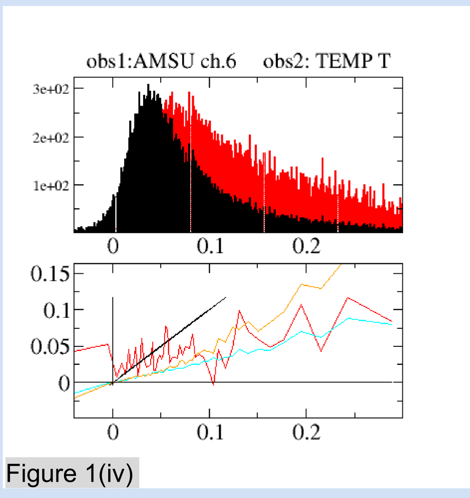

Figure 1(iv)

4. Results: Impact on analysis (t=0)

Our statistical tool

- Columns: cases I) „full analysis“ and II) „single obs“
- Bottom graphs are normalised by $\sum\left|\frac{P^a[\mathbf{v},t;\alpha]}{\sqrt{\boldsymbol{R}_{vv}\boldsymbol{R}_{\alpha\alpha}}}\right|$
- Our noise estimate for a stochastic sum $\sum_j A_j$ is $\sqrt{\sum_j A_j^2}$

- usually very similar qualitative behaviour for cases I) and II).
- generally execellent agreememt for localised observations. (note functions need to be positive for good impact – compare definitions)

Below, (due to the limited space) only data for case I) are shown.

— $\sum_{\alpha\in obs}\left(J^{t|b}_\alpha + \left\{J^{t|b}_\alpha - J^{t|ab}_\alpha\right\}\right)$
— $\sum J^{t|b}_\alpha$
noise estimate
— $\sum J^{t|ab}_\alpha$
noise estimate

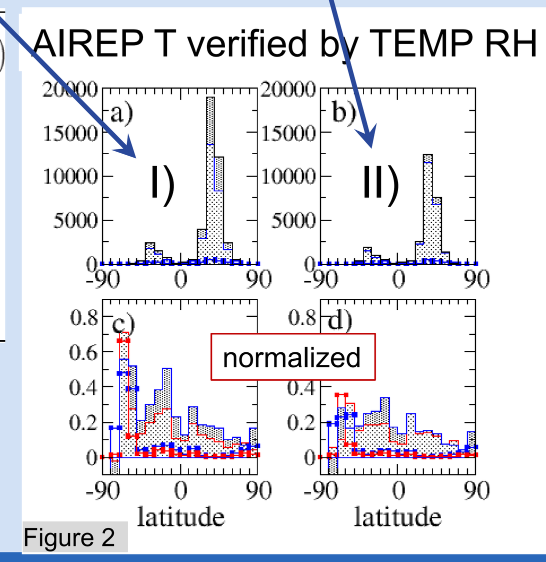

Figure 2

AMSU-A verified by GPSRO

channels
8 & 11 : mostly significant positive impact
11 : weak performance near the equator

9 & 10 : weak performance particularly near poles.
this can be explained by known model biases for the height of these channels

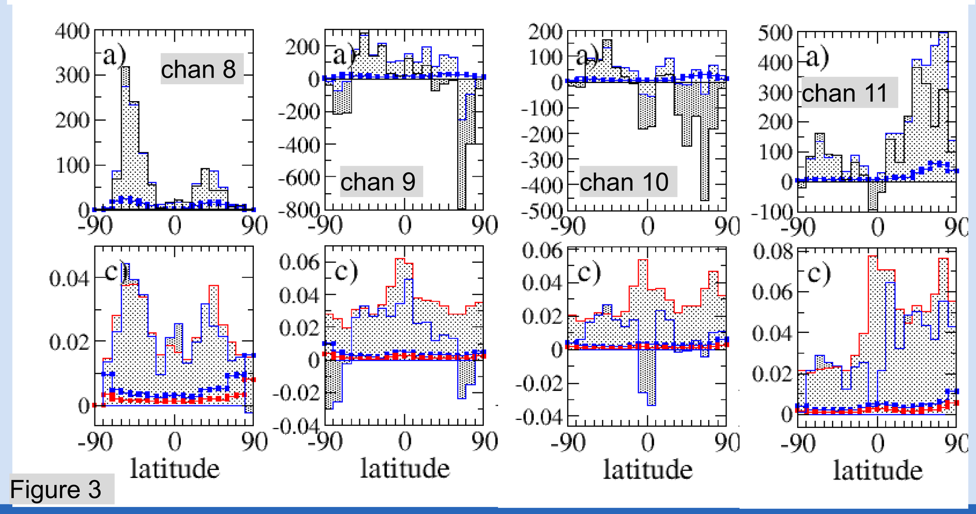

Figure 3

Dependance on the vertical distance between observations

- For localized obs. EFSOI statistics are dominated by very small distances (log(Δp) ∊ [-0.1 , 0.1], Fig.4).
- For GPSRO those peaks in log(Δp)-space are substantially broader (Fig.5).
- For AMSU A (see Fig.6) peaks are even broader and the normalized data exhibit no or strongly shifted peaks. normalized data → largely linked to vert.structure of the GPSRO FG departure biases (weak performance of chan. 9 is mostly due to latitude depend. biases discussed above)

Performance of the different data types is consistent with the agreement with the ensemble covariances (comp.Sec.3 above).

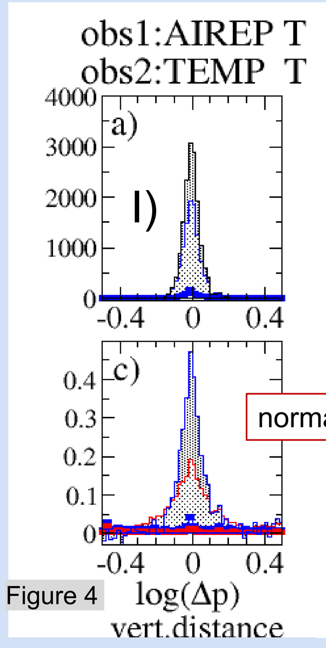

Figure 4

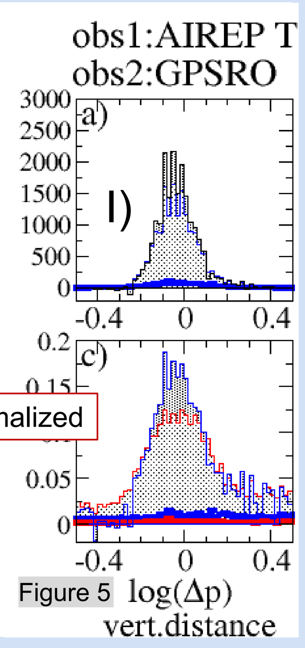

Figure 5

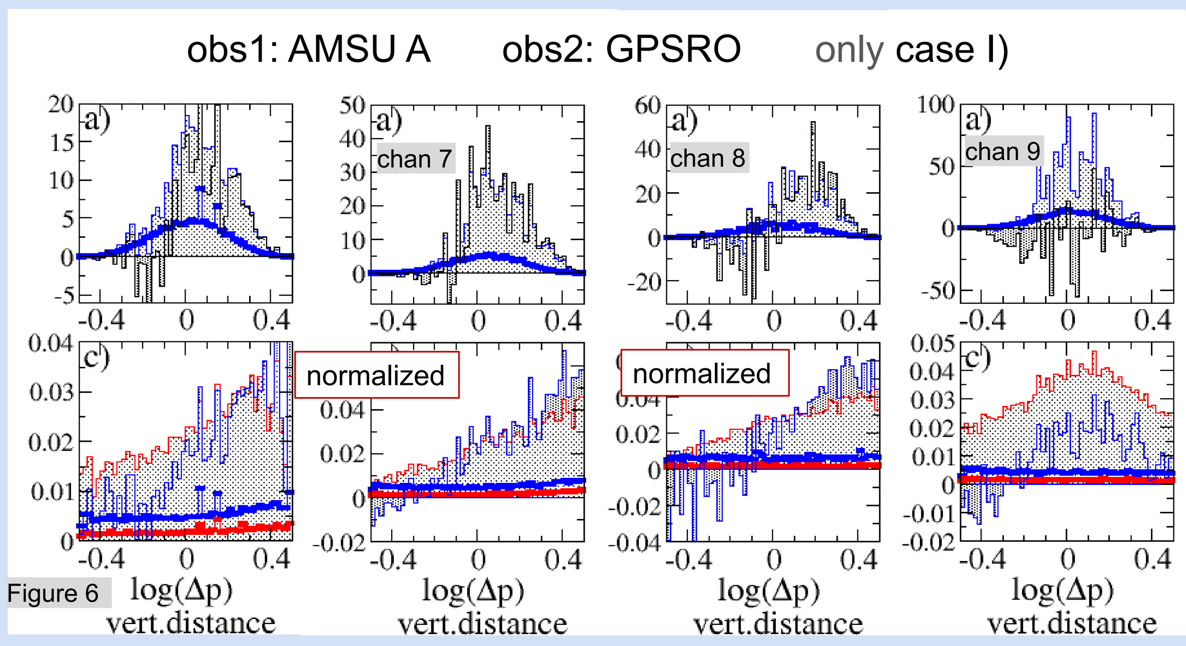

Figure 6

5. Summary and Discussion

- We use observations for verification (Sommer and Weissmann, 2014) but consider EFSOI type statistics as consistency relations

(obs – fg) obs1 | (obs – fg) obs2 | ensemble covariances

- Generalising the method from Kalnay et al. (2012) a verification function formalism corresponding to single obs cases has been derived. So far results look very similar to the „usual“ EFSOI (*proxy for denial experiments*).
- Also a method to directly compare covariances from the ensemble and observations is explored.
- A statistical tool (software) has been developed which is flexible to generate conditional statistics
  -- dependent on, eg., latitude, height, distance between observations, sun angle, ...
  -- comparing different observation types (e.g., conventional, satellite)
  -- comparing different parameters (e.g., temperature, humidity, velocities)
- The impact (or agreement) of localized observations seems to be good to excellent.
- Also for GPSRO generally good agreement is observed

The impact of Satellite radiances (AMSU A):

While the impact is overall positive, strong variations and negative impact is observed in some conditions. Important reasons are:

- The quality of ensemble covariances reduces with distance between the observations. → less impact for strongly non-local observations.
- **AMSU A biases opposite to model bias**. Most regions with negative impact can be linked to (***obs – fg***)-biases of radiances being opposite to those of the verifying observations (the latter being dominated by model biases and couple to biases of the radiances).

References

- R. H. Langland and N. L. Baker. Estimation of observation impact using the nrl atmosphericvariational 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 observations: application to ensemble kalman filters. Tellus A: Dynamic Meteorology and Oceanography, 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).

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