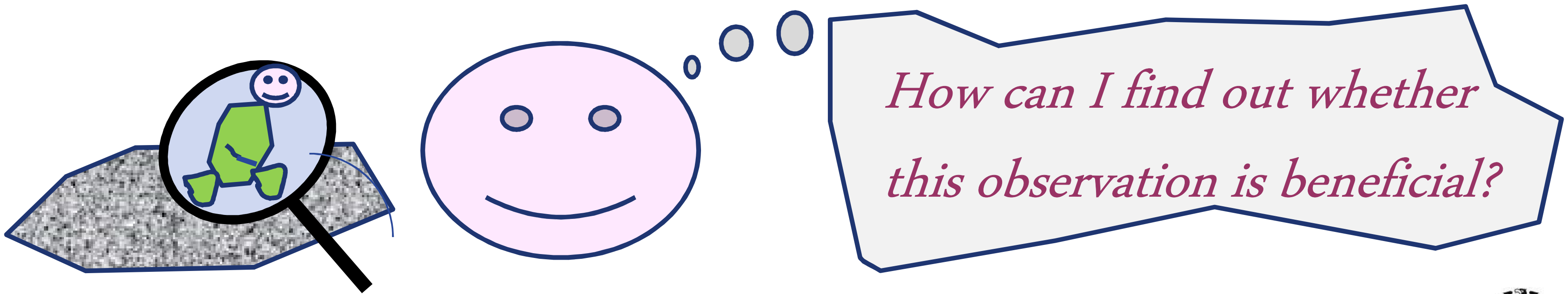
Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Observation impact diagnostics in an Ensemble Data Assimilation System

Olaf Stiller, Deutscher Wetterdienst

denial experiments/
observation system experiments (OSE) → **extremely expensive**



Observation impact diagnostics in an Ensemble Data Assimilation System

Motivation

denial experiments/
observation system experiments (OSE) → **extremely expensive**

Aim: Find diagnostic tools to

- indicate impact of ***observation subsets*** on analysis/forecast
- identify where observation impact is sub-optimal/negative

Define cost function: “model fit to the *truth*”

(*Langland and Barker, 2004*)

using obs - not using obs

Here I follow

Sommer and Weissmann (2014,2016)

and use observations for verification

$$\boldsymbol{y}^{\mathrm{v};t} = \boldsymbol{y}_{\mathrm{v}}^{o;t} \quad \text{observation } \boldsymbol{y}_{\mathrm{v}}^{o} \text{ at time } t$$

$$\boldsymbol{H}_{\mathrm{v}} \;:\; \text{observation operator for } \boldsymbol{y}_{\mathrm{v}}^{o}$$

$$C = \boldsymbol{R}_{\mathrm{vv}}^{-1} = \frac{1}{r_{\mathrm{v}}}$$

$$J \stackrel{\text{def!}}{=} \frac{1}{2}\left(\left\|\boldsymbol{e}^{t|a}\right\|^2 - \left\|\boldsymbol{e}^{t|b}\right\|^2\right) \qquad \begin{aligned} \boldsymbol{e}^{t|a} &= \boldsymbol{y}^{t|a} - \boldsymbol{y}^{\mathrm{v};t} \\ \boldsymbol{e}^{t|b} &= \boldsymbol{y}^{t|b} - \boldsymbol{y}^{\mathrm{v};t} \end{aligned}$$

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

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

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

Olaf.Stiller@dwd.de

$$J = \frac{1}{2}\left(\boldsymbol{e}^{t|a} + \boldsymbol{e}^{t|b}\right)^T \boldsymbol{C}\left(\boldsymbol{e}^{t|a} - \boldsymbol{e}^{t|b}\right)$$

$$\begin{aligned} J &= \sum_{l\in\{obs\}} \frac{1}{2}\left(\boldsymbol{e}^{t|a} + \boldsymbol{e}^{t|b}\right)^T \boldsymbol{C}\boldsymbol{H}_{\mathrm{v}} M_{0\to t}\mathbf{K}_l\left(\boldsymbol{y}_l^o - \boldsymbol{y}_l^b\right) \\ &= \sum_{l\in\{obs\}} \frac{1}{2}\left(\boldsymbol{J}_l^{t|a} + \boldsymbol{J}_l^{t|b}\right) \end{aligned}$$

$\mathbf{K}_l$: column of Kalman gain matrix acting on $\boldsymbol{y}_l^o$

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

$\boldsymbol{H}_{\mathrm{v}}$: operator computing `modelequivalent` to $\boldsymbol{y}^{\mathrm{v};t}$

$$J \stackrel{\text{def!}}{=} \frac{1}{2}\left(\left\|\boldsymbol{e}^{t|a}\right\|^2 - \left\|\boldsymbol{e}^{t|b}\right\|^2\right) \qquad \begin{aligned} \boldsymbol{e}^{t|a} &= \boldsymbol{y}^{t|a} - \boldsymbol{y}^{\mathrm{v};t} \\ \boldsymbol{e}^{t|\mathrm{b}} &= \boldsymbol{y}^{t|\mathrm{b}} - \boldsymbol{y}^{\mathrm{v};t} \end{aligned}$$

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

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

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

Time evolution

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

$$
\begin{aligned}
J &= \sum_{l \in \{obs\}} \frac{1}{2} \left(\boldsymbol{e}^{t|a} + \boldsymbol{e}^{t|b} \right)^T \boldsymbol{C} \boldsymbol{H}_{\mathrm{v}} M_{0 \to t} \mathrm{K}_l \left(\boldsymbol{y}_l^o - \boldsymbol{y}_l^b \right) \\
&= \sum_{l \in \{obs\}} \frac{1}{2} \left(\boldsymbol{J}_l^{t|a} + \boldsymbol{J}_l^{t|b} \right)
\end{aligned}
$$

K_l : column of Kalman gain matrix acting on $\boldsymbol{y}_l^o$
$M_{0 \to t}$: time evolution operator
$\boldsymbol{H}_{\mathrm{v}}$: operator computing `modelequivalent` to $\boldsymbol{y}^{\mathrm{v};t}$

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

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

Ensemble Kalman Filter (EKF)
→ use ensemble

Olaf.Stiller@dwd.de

Optimality Condition

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

$$J = \sum_{l \in \{obs\}} \frac{1}{2} \left(\boldsymbol{e}^{t|a} + \boldsymbol{e}^{t|b} \right)^T \boldsymbol{C}\boldsymbol{H}_{\mathrm{v}} M_{0\to t} \mathrm{K}_l \left(\boldsymbol{y}_l^o - \boldsymbol{y}_l^b \right)$$

$$= \sum_{l \in \{obs\}} \frac{1}{2} \left(\boldsymbol{J}_l^{t|a} + \boldsymbol{J}_l^{t|b} \right)$$

Choose initial condition:

$$\boldsymbol{x}^{\mathrm{init}}\left(\{\lambda_l\}\right) = \boldsymbol{x}^{\mathrm{b}} + \sum_{l \in \{obs\}} \mathrm{K}_l \lambda_l \left(\boldsymbol{y}_l^o - \boldsymbol{y}_l^b \right)$$

$$\boldsymbol{x}^a = \boldsymbol{x}^{\mathrm{init}}\left(\{\lambda_l = 1\}\right)$$

If J has a minimum for $\boldsymbol{x}^{\mathrm{init}} = \boldsymbol{x}^a$
one finds

$$\boxed{\begin{aligned} \left\langle \boldsymbol{J}_l^{t|a} \right\rangle &\equiv \left\langle \left(\boldsymbol{e}^{t|a} \right)^T \boldsymbol{C}\boldsymbol{H}_{\mathrm{v}} M_{0\to t} \mathrm{K}_l \left(\boldsymbol{y}_l^o - \boldsymbol{y}_l^b \right) \right\rangle \\ &= 0 \qquad \text{for all } l \in \{obs\} \end{aligned}}$$

$\langle .. \rangle$: statistical mean

Our System

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

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

- LETKF (Hunt et al. 2007)
  40 ensemble members
- verification vs obs.

  (Sommer and Weissmann 2014)
- time evolution via analysis ensemble

  (Kalnay et al. 2012)

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

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

Our System

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

$$J = \sum_{l\in\{obs\}} \frac{1}{2}\left(e^{t|a} + e^{t|b}\right)^T C H_{\mathrm{v}} M_{0\to t} \mathrm{K}_l \left(y_l^o - y_l^b\right)$$
$$= \sum_{l\in\{obs\}} \frac{1}{2}\left(J_l^{t|a} + J_l^{t|b}\right)$$

- LETKF (Hunt et al. 2007)
  40 ensemble members
- verification vs obs.

  (Sommer and Weissmann 2014)
- time evolution via analysis ensemble

  (Kalnay et al. 2012)

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

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

$$\mathrm{w}_k^a = \sum_{\alpha,\beta\in\{obs\}} Y_\alpha^{a(k)} R_{\alpha\beta}^{-1}\left(y_\beta^o - y_\beta^b\right)$$
$$Y_\alpha^{a(k)} = H_\alpha X^{a(k)}$$

: $X^{a(k)}$ in obs space

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

analysis covariance matrix in obs space
(estimated from ensemble)

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

Olaf.Stiller@dwd.de

Interpretation for *t=0*

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

$$\left\langle -J_l^{0|a} \right\rangle = P_{\mathrm{v}l}^a \frac{\langle (y_\mathrm{v}^o - y_\mathrm{v}^a)(y_l^o - y_l^b) \rangle}{R_{ll} R_{\mathrm{vv}}} = \frac{P_{\mathrm{v}l}^a}{R_{ll} R_{\mathrm{vv}}} R_{\mathrm{v}l} \quad = 0 \text{ if } y_\mathrm{v}^o \text{ , } y_l^o \text{ statistically independent}$$

$$\left\langle -J_l^{0|b} \right\rangle = P_{\mathrm{v}l}^a \frac{\langle (y_\mathrm{v}^o - y_\mathrm{v}^b)(y_l^o - y_l^b) \rangle}{R_{ll} R_{\mathrm{vv}}} = \frac{P_{\mathrm{v}l}^a}{R_{ll} R_{\mathrm{vv}}} P_{\mathrm{v}l}^b \quad > 0 \text{ if } P_{\mathrm{v}l}^a \text{ and } P_{\mathrm{v}l}^b \text{ have the same sign}$$

Results: What is shown?

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

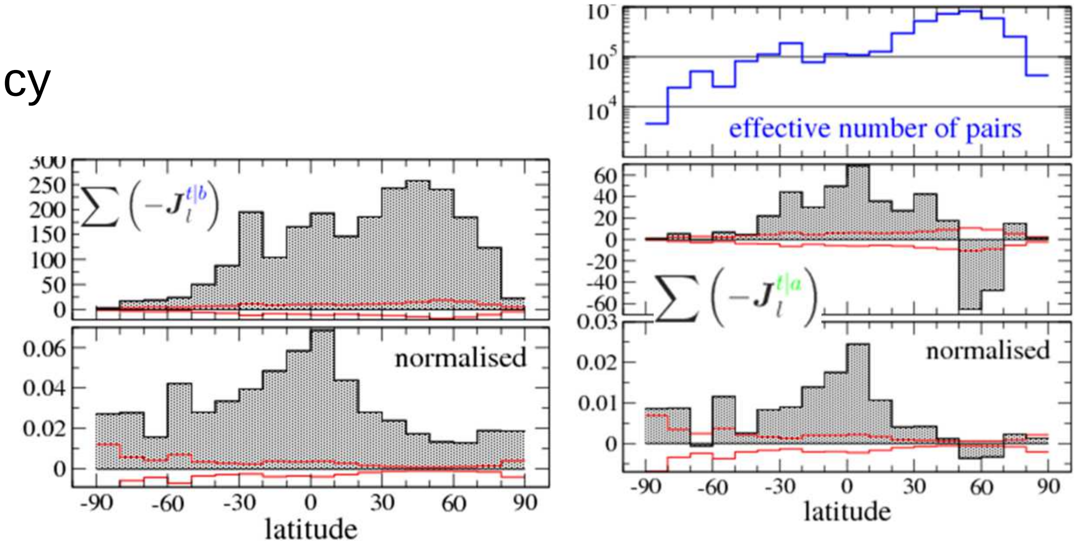
$$\sum_{stats} \left(-J_l^{0|a} \right) = \sum_{stats} P_{\mathrm{v}l}^a \frac{(y_\mathrm{v}^o - y_\mathrm{v}^a)\,(y_l^o - y_l^b)}{R_{ll} R_{\mathrm{vv}}}$$
← should be small --- "optimality condition"

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

.

Results:
Temps verified by GPSRO

Excellent correspondence/consistency between Temps and GPSRO



– impact variable : $\sum\left(-J^{0|a}\right)$ or $\sum\left(-J^{0|b}\right)$

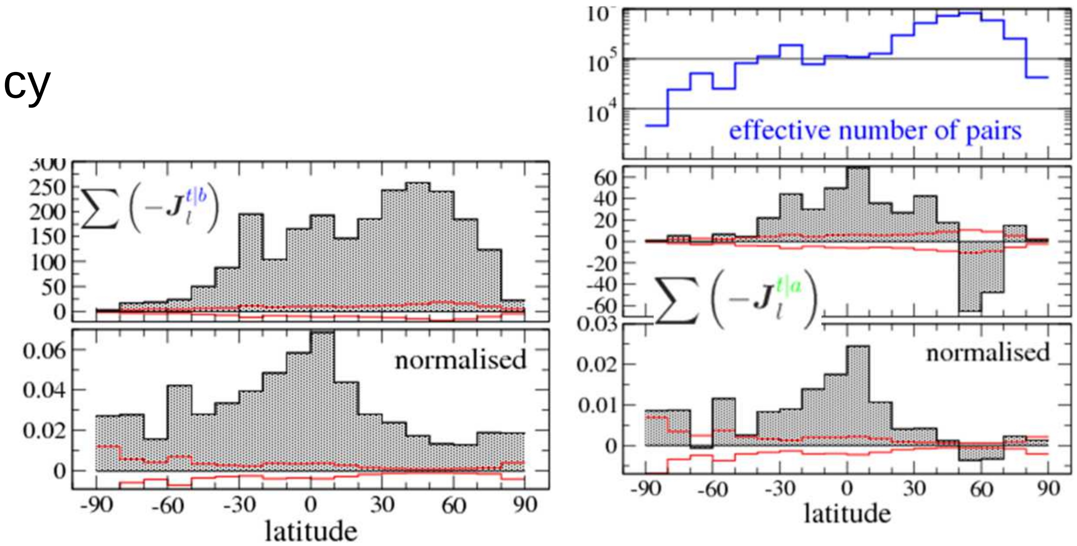
– "normalised" : $\sum\left(-J^{0|a}\right) / \left(\sum \frac{|P^a_{\mathbf{v}l}|}{\sqrt{R_{\mathbf{vv}} R_{ll}}}\right)$ $\sum\left(-J^{0|b}\right) / \left(\sum \frac{|P^a_{\mathbf{v}l}|}{\sqrt{R_{\mathbf{vv}} R_{ll}}}\right)$

– ⟨uncorr. noise⟩ : $\sqrt{\sum\left(-J^{0|a}\right)^2}$ $\sqrt{\sum\left(-J^{0|b}\right)^2}$

– effective number : $\left(\sum |P^a_{\mathbf{v}l}|\right)^2 / \sum |P^a_{\mathbf{v}l}|^2$

Results:
Temps verified by GPSRO

Excellent correspondence/consistency between Temps and GPSRO



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

$\sum\left(-J_l^{t|a}\right)$: is

1.) much smaller than $\sum\left(-J_l^{t|b}\right)$

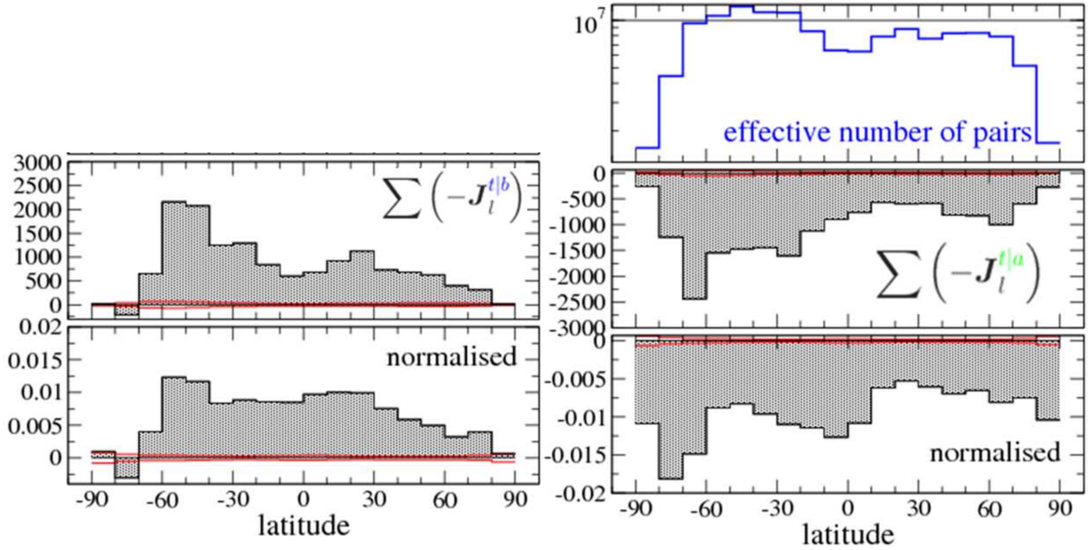
→ optimality condition largely fulfilled

2.) mostly positive

→ weight on TEMPS in assimilation could be slightly increased in tropical regions

Results:
AMSU-A verified by GPSRO

Deutscher Wetterdienst
Wetter und Klima aus einer Hand



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

$\sum\left(-J_l^{t|a}\right)$: clearly negative
→ obs have too strong weight

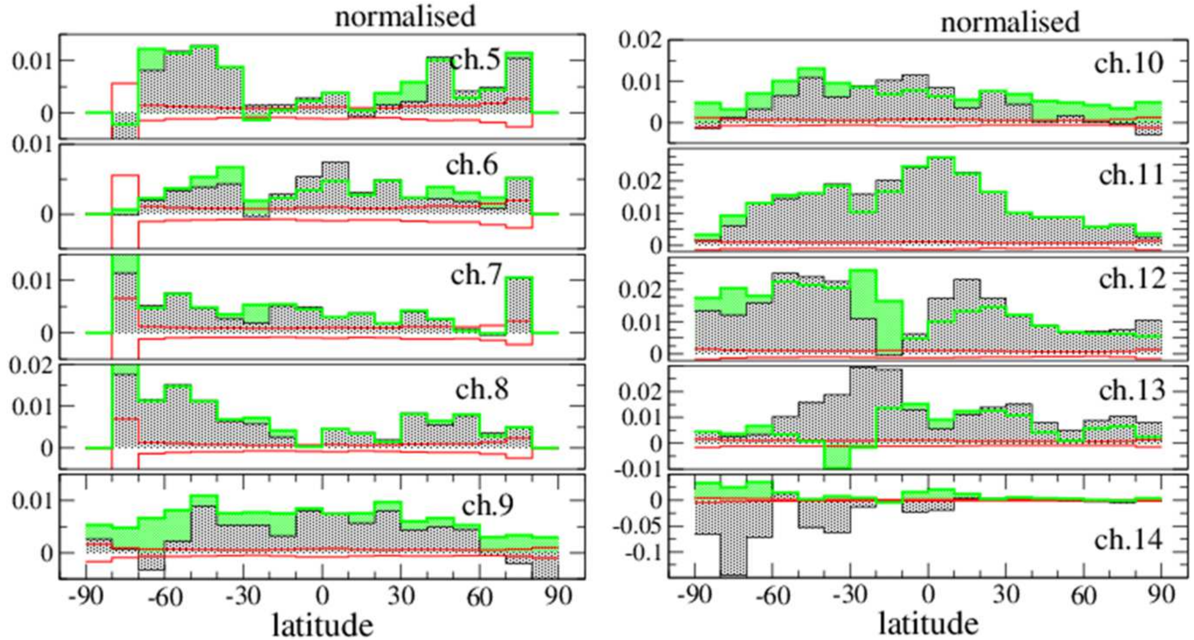
AMSU-A observation errors are too small
(also according to Desroziers diagnostics)

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

Results:
AMSU-A verified by GPSRO

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- individual channels   only showing $\sum\left(-J_l^{t|b}\right)$ - normalised



Results:
AMSU-A verified by GPSRO

- individual channels   only showing $\sum\left(-J_l^{t|b}\right)$ - normalised

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

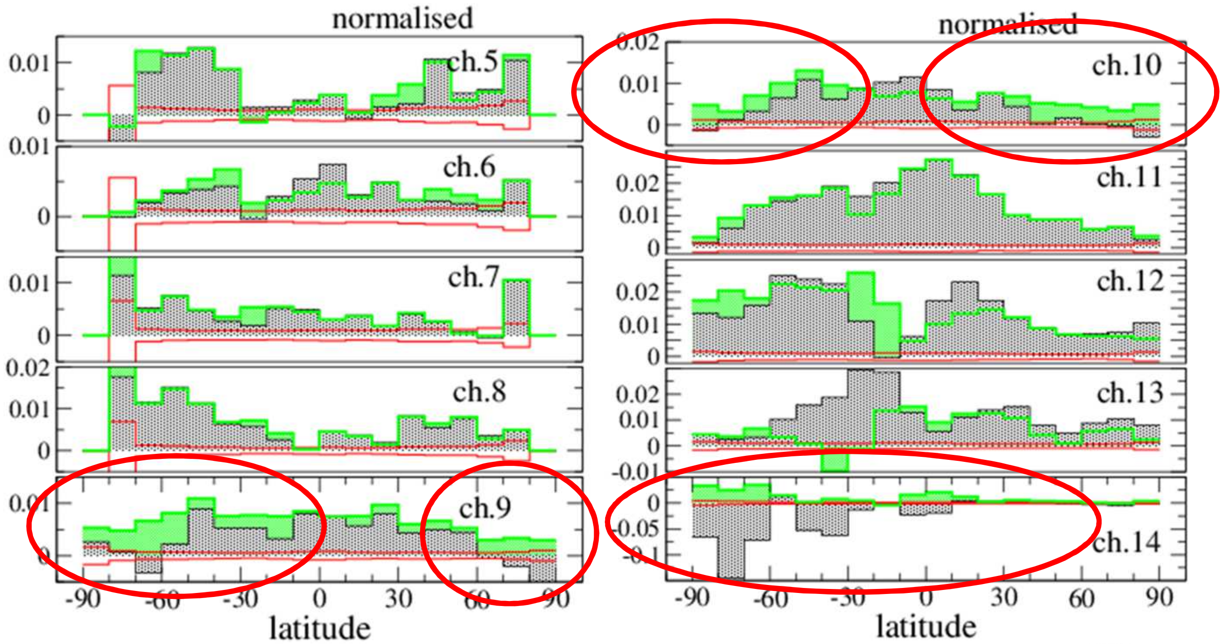
In some regions
- dark shaded areas are negative or neutral
- green (bias removed a posteriori) clearly positive

→ **AMSU - GPSRO opposite bias**

Significant bias problems occur for :

(i) ch.14 (everywhere)

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



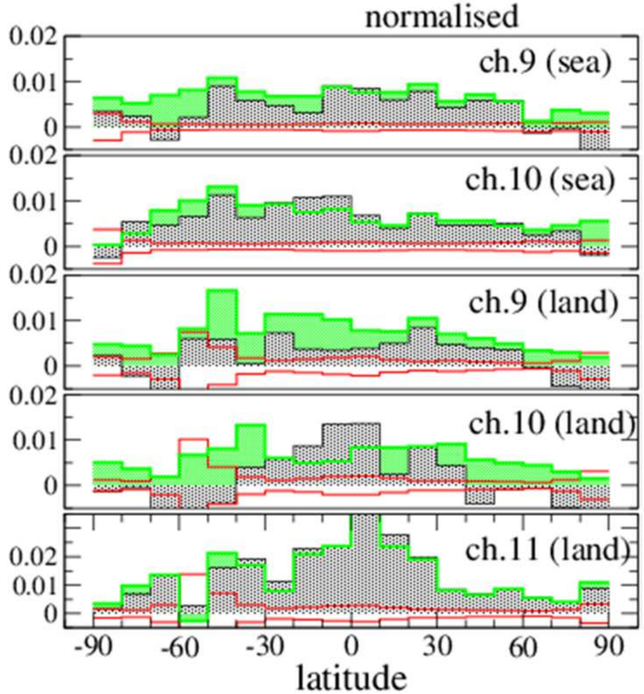
Results:
AMSU-A verified by GPSRO

- individual channels only showing $\sum\left(-J_i^{t|b}\right)$ - normalised

Bias problems for channels 9 and 10 are stronger over land than over sea

(Channel 11 no significant bias found – neither over land)

data over sea/land only



Outlook: Assessing impact of individual observations

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Work so far:
Cost-function J gives **impact of all observations** assimilated at time $t = 0$.

Interpretation of different components J_l
corresponding to individual observations $\boldsymbol{y}_l$
is *suggestive but not rigorous*.

$$J = \sum_{l \in \{obs\}} \frac{1}{2}\left(\boldsymbol{J}_l^{t|a} + \boldsymbol{J}_l^{t|b}\right)$$

More rigorous approach:
Cost-function for **data denial exp**eriment. Replace: $\boldsymbol{x}^{b} \rightarrow \boldsymbol{x}^{a/k}$ **(analysis not using $\boldsymbol{y}_k^o$)**

$$J = \frac{1}{2}\left(\left\|\boldsymbol{e}^{t|a}\right\|^2 - \left\|\boldsymbol{e}^{t|b}\right\|^2\right) \quad \rightarrow \quad \frac{1}{2}\left(\left\|\boldsymbol{e}^{t|a}\right\|^2 - \left\|\boldsymbol{e}^{t|a/k}\right\|^2\right)$$

$$\boxed{\begin{aligned}\boldsymbol{J}_k^{t|a} &= -P_{\mathrm{v}k}^a(t)\frac{\left(\boldsymbol{y}_k^o - \boldsymbol{y}_k^{a/k}\right)\left(\boldsymbol{y}_{\mathrm{v}}^{o;t} - \boldsymbol{y}_{\mathrm{v}}^{t|a}\right)}{\boldsymbol{R}_{kk}\boldsymbol{R}_{\mathrm{vv}}} \\ \boldsymbol{J} &= \boldsymbol{J}_k^{t|a} - \frac{1}{2}\left[P_{\mathrm{v}k}^a(t)\right]^2\frac{\left(\boldsymbol{y}_k^o - \boldsymbol{y}_k^{a/k}\right)^2}{\boldsymbol{R}_{kk}^2\boldsymbol{R}_{\mathrm{vv}}}\end{aligned}}$$

with

$$\begin{aligned}\left(\boldsymbol{y}_l^o - \boldsymbol{y}_l^{a/k}\right) &= \left(\mathbf{1} - \boldsymbol{H}_{\mathrm{v}}\mathbf{K}_k\right)^{-1}\left(\boldsymbol{y}_k^o - \boldsymbol{y}_k^{a}\right) \\ &= \frac{\left(\boldsymbol{y}_k^o - \boldsymbol{y}_k^{a}\right)}{1 - P_{kk}^a(t=0)/\boldsymbol{R}_{kk}}\end{aligned}$$

Summary and Discussion

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- Different parts of the cost function $J = \sum_{l \in \{obs\}} \frac{1}{2}\left(\boldsymbol{J}_l^{t|a} + \boldsymbol{J}_l^{t|b}\right)$ should be considered (interpreted) separately.

- Examples were given for how the diagnostics could be linked to:
  - 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 interpretation of observation impact diagnostics is often not trivial.
  - Statistical significance/spurious correlations is a big issue (particularly for large forecast lead times).

- A method is under development to show the “denial impact” *for individual observations* (e.g., a single AMSU channel).

Olaf.Stiller@dwd.de

Thank you – any questions?

Olaf.Stiller@dwd.de