Application of observation cross-validation method to IASI cloud screening

Deutscher Wetterdienst

Wetter und Klima aus einer Hand

1. Introduction: Cross Validation (CV) diagnostics

The exploitation of remote sensing data for NWP strongly relies on quality control type methods aimed at identifying observations affected by influences (as, e.g., from clouds or land surfaces).
To facilitate the detection of such observations, a cost effective mathematical *cross validation* (**CV**) framework has been developed which computes the conditional probability of observations given the background and other observations.(see present. **11.06**)

This poster demonstrates how the CV diagnostics can be employed for IASI data. Steps towards a cloud screening method based on these diagnostics are presented.

General result:

decompose observations: $\mathbf{y}^o = \{\mathbf{y}^o_{\tau C}, \mathbf{y}^o_{\tau}\}$

Conditional probability of observations $\mathbf{y}^o_\tau$ (given the background and observations $\mathbf{y}^o_{\tau C}$):

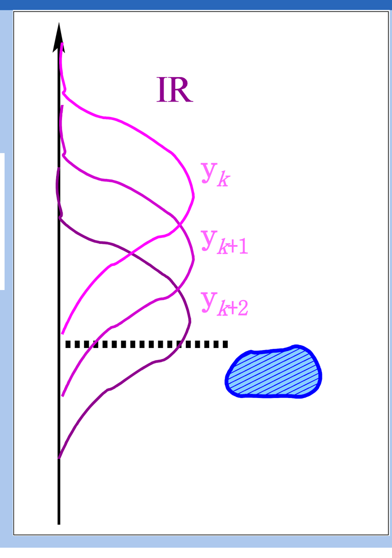
$$P(\mathbf{y}^o_\tau | \mathbf{y}^o_{\tau C}, \mathbf{X}^b) \;\propto\; \exp -\frac{1}{2}\left\{(\mathbf{y}^o_\tau - \overline{\mathbf{y}_\tau})^T \mathbf{D}_\tau (\mathbf{y}^o_\tau - \overline{\mathbf{y}_\tau})\right\}$$

$$\left[\mathbf{H}^T\mathbf{B}\mathbf{H} + \mathbf{R}\right]^{-1} = \begin{pmatrix} \mathbf{D}_{\tau C} & \mathbf{C}_\tau \\ \mathbf{C}^T_\tau & \mathbf{D}_\tau \end{pmatrix}$$

$$\overline{\mathbf{y}_\tau} = -\mathbf{D}_\tau^{-1}\mathbf{C}\,\mathbf{y}^o_{\tau C}$$

$$(\mathbf{y}^o_\tau - \overline{\mathbf{y}_\tau}) = \mathbf{D}_\tau^{-1}\mathbf{z}_\tau$$

$$\mathbf{z} = \left[\mathbf{H}^T\mathbf{B}\mathbf{H} + \mathbf{R}\right]^{-1}\mathbf{y}^o$$



Special case:

observations can be ordered with respect to their vulnerability

$$P(\mathbf{y}_k | \mathbf{y}_{\{l<k\}}, \mathbf{x}^b) = \mathbb{N}^{-1}\exp -\frac{1}{2}\left\{\mathbf{Y}_k^{\,2}\right\}$$

$$\mathbf{Y} = \mathbf{T}_L^{-1}\mathbf{y}$$

Cholesky decomposition:

$$\mathbf{T}_L\mathbf{T}_U = \begin{pmatrix} t_{11} & 0 & & & \\ t_{21} & t_{22} & 0 & & 0 \\ & & \ddots & 0 & \\ t_{k1} & t_{k2} & & t_{kk} & 0 \\ & & & & \ddots \\ & & & & \ddots\; 0 \\ t_{p1} & & & & t_{pp} \end{pmatrix}\begin{pmatrix} t_{11} & t_{21} & & t_{k1} & & t_{p1} \\ 0 & t_{22} & & t_{k2} & & \\ & 0 & \ddots & & & \\ & & 0 & t_{kk} & & \\ & & & 0 & \ddots & \\ & & & & 0 & t_{pp} \end{pmatrix} = \left[\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T\right]$$

$$\mathbf{Y}_k = \frac{\mathbf{y}_k - \mathbf{y}^{a*}_k}{\sqrt{\epsilon^{obs}_k + \epsilon^{a*}_k}}$$

: is a stochastic variable with variance 1

$\mathbf{y}^{a*}_k$: analysis using only observations with index $l < k$

2. Looking at IASI spectra with CV diagnostics

band 1 | **band 2 (humidity channels)**

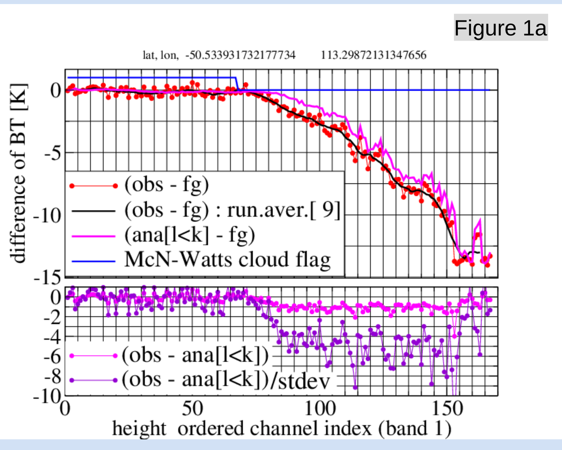


Figure 1a

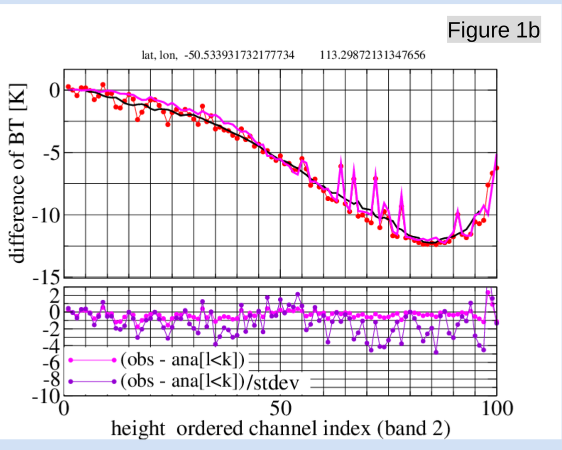


Figure 1b

High clouds:

- Most clouds are cold
- High clouds ⇔ strong signal easy to detect by any method
- Cloud signals are generally weaker in the humidity channels (band 2)

Here: channels are ordered according to their sensitivity with respect to clouds (see Mc Nally & Watts scheme)

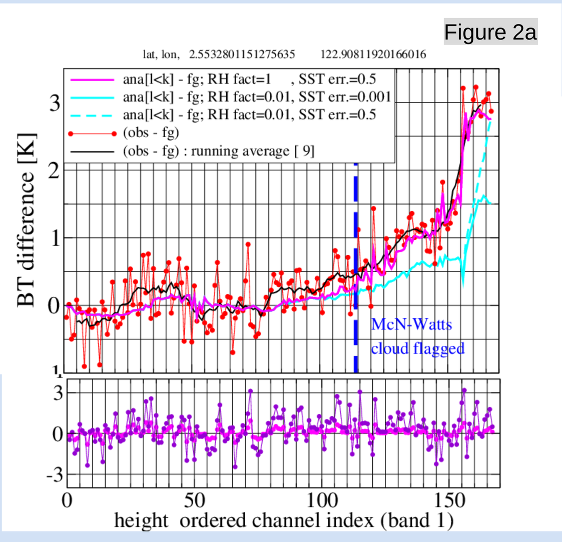


Figure 2a

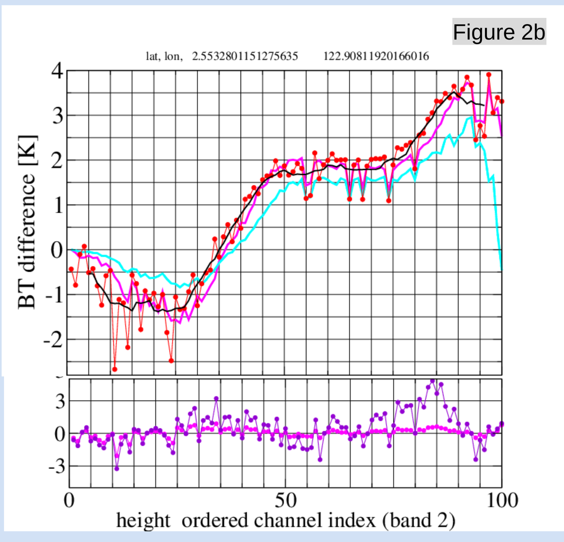


Figure 2b

Low level features

Low level *obs-fg* departures can be caused by departures of:

- Surface temperature
- Atmospheric humidity
- Low level clouds

Which of these explanations is correct can be difficult to determine.
Warm signals (like in Fig.2), however, are mostly not related to clouds (apart from at high latitudes where low clouds over cold surfaces cause warm *obs-fg* departures).

Figure 2 is the same as Fig.1 but for a case with a warm signal in the lower level channels. The cyan curves in the top graphs of Fig.2 also show the analysis values $y_k^{a[l<k]}$ (i.e., the analysis using only obs y_l with $l < k$) which are obtained when background errors for RH are assumed to be extremely small (1% of the normal value). Whether the SST background error is also assumed small (0.001 K, solid line) or at the normal value (0.5 K, dashed line) is seen to affect only the lowest channels of band 1 (window channels).
This shows that the RH background errors are crucial for explaining the large obs-fg departures of the low level channels from band 1 (and obviously also for those of the humidity sensitiv band 2).

Noise reduction through cross validation

For the upper channels (small channel index) of band 1, the (assumed) errors of *obs-fg* departures are dominated by the observation errors (orange lines in Figs 3a&b). For the lower channels the errors are increased through the background errors of RH and, for the lowest channels, also by the SST error. In band 2, *obs-fg* departures are generally dominated by the RH background errors.
Cross validation **strongly reduces** the **correlated errors**. Correspondingly, in Fig.3, the standard deviations of the analysis $y_k^{a[l<k]}$ (i.e., the analysis using only obs y_l with $l < k$) is dominated by the observation errors while the contributions from the background RH errors are strongly reduced.

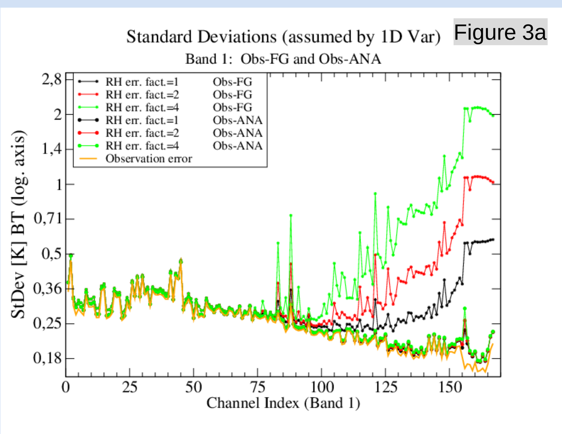


Figure 3a

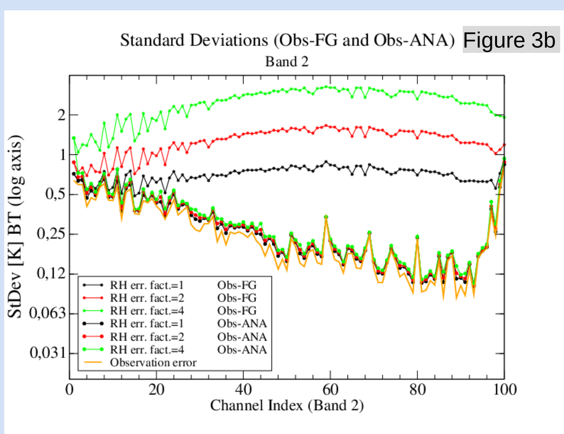


Figure 3b

Obs error estimate band 2

The strong noise reduction by the cross validation method was employed for estimating the observation errors in band 2.

3. Designing cloud screening methods

Aim:	Detecting radiances which are influenced by clouds from analyzing the IASI spectra.
Problem:	Radiances which are only **weakly affected** by clouds are difficult to detect
	• Small cloud fractions (compare Fig.4) • Low level clouds (low altitudes) • Determining the cloud tops
Required:	Diagnostics for **detecting collective structures** involving a larger number of adjacent channels (all affected by the same cloud)

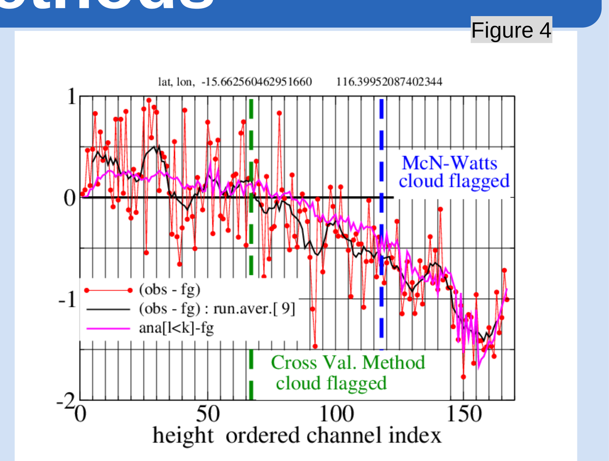


Figure 4

i) First attempt:

Define : $\mathbf{Y}_k \rightarrow \mathbf{Y}_k{}^n \equiv \sum_{j=n-k}^{k} \frac{\mathbf{Y}_j}{\sqrt{n}}$

- $\mathbf{Y}_k{}^n$ also has zero mean and variance 1
- Flag FoVs where $\mathbf{Y}_k{}^n$ exceeds threshold
- However: this diagnostic is far too sensitive to atmospheric perturbations in general (comp. Fig.5)

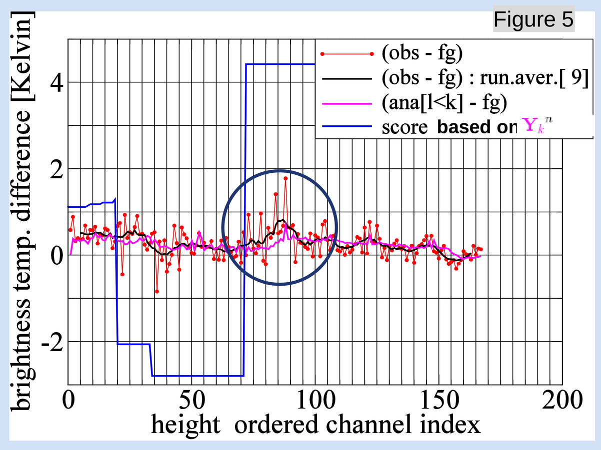


Figure 5

ii) More targeted approach: Project on cloud observation operator

A) Identify cloudy FoVs

Define:

$$\vec{\mathbf{Y}} \rightarrow \tilde{\mathbf{Y}}_l \equiv \frac{\vec{\boldsymbol{h}}_l * \vec{\mathbf{Y}}}{\left\|\vec{\boldsymbol{h}}_l\right\|}$$

$$\vec{\mathbf{Y}} \equiv \mathbf{T}_L^{-1}\boldsymbol{y}$$

$$\vec{\boldsymbol{h}}_l \equiv \mathbf{T}_L^{-1}\mathbf{H}_{cfr[l]}$$

$\mathbf{H}_{cfr[l]}$: column vector of $\mathbf{H}$ related to cloud fraction at level l

- $\tilde{\mathbf{Y}}_l$ is designed to filter (mainly) cloud type structures
- Flag FoVs cloudy where $\tilde{\mathbf{Y}}_l$ exceeds threshold

B) Determining the top of the cloud

Define:

$$\tilde{\mathbf{Y}}_l{}^{(k)} \equiv \frac{\vec{\boldsymbol{h}}_l{}^{(k)} * \vec{\mathbf{Y}}^{(k)}}{\left\|\vec{\boldsymbol{h}}_l\right\|}$$

$\vec{\boldsymbol{h}}_l{}^{(k)}, \vec{\mathbf{Y}}^{(k)}$: considering only obs $\boldsymbol{y}_l$ with $l < k$

Observation $\boldsymbol{y}_k$ is assumed cloud free (i.e., above the cloud) if for all levels $\boldsymbol{l}$:

1) $\tilde{\mathbf{Y}}_l{}^{(k)} <$ Threshold 1

2) $\boldsymbol{grad}_k\left[\tilde{\mathbf{Y}}_l{}^{(k)}\right] <$ Threshold 2

Comparison with McNally-Watts scheme

The new cross-validation scheme

- selects (almost) the same field of views as cloudy if background errors of RH and SST are very small (see Fig.6)
- otherwise, has less low level clouds than McN-Watts (flags them as cloud free, see red curve in Fig.6).
- is (in general) more conservative –> i.e., puts cloud tops to higher levels (see example in Fig.4) (compare Figs.7 a and b)

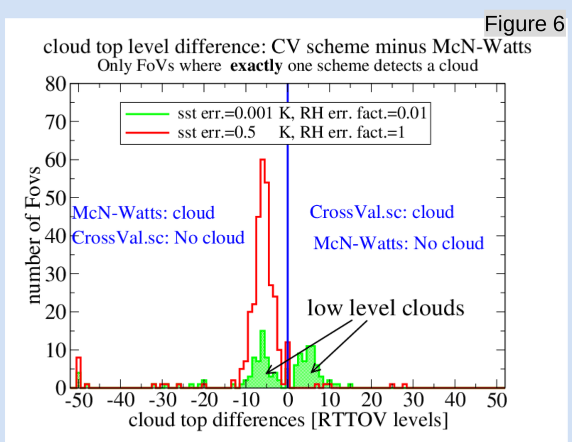


Figure 6

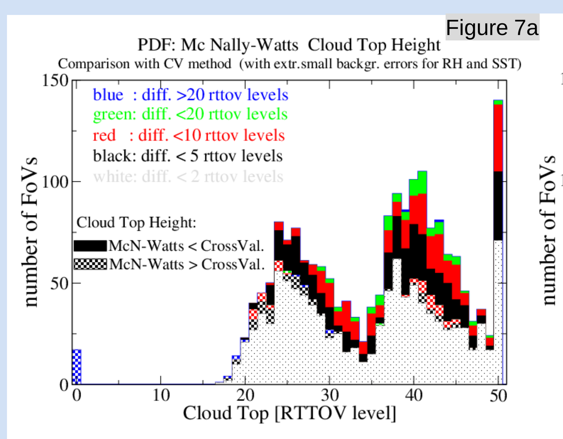


Figure 7a

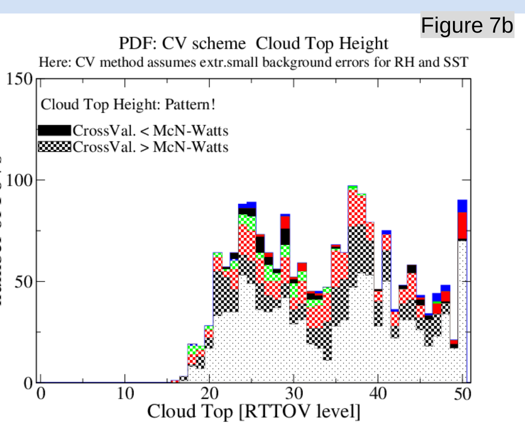


Figure 7b

4. Summary

Cross-validation diagnostics (see presentation **11.06**) have been applied to IASI radiances.
The analysis $y_k^{a[l<k]}$ (i.e., the analysis for y_k using only obs y_l with $l < k$) is seen to be usually quite close to the observations (see Figs. 1, 2, 4 and 5). Only if the background errors for RH are assumed to be extremely small, the values of *obs*-$y_k^{a[l<k]}$ are considerable for the lower channels of band 1 and those of band 2 (see Figs. 2 a and b).
The strong noise reduction by the CV method is consistent with the assumed errors of *obs-fg* and *obs*-$y_k^{a[l<k]}$ shown in Figs.3. For band 2, the *obs - fg* errors are dominated by the background errors (mainly for humidity) while *obs* - $y_k^{a[l<k]}$ errors are always dominated by the observation error.
The strong noise reduction was employed for estimating the observation errors in band 2.

Designing **a cloud screening** method requires diagnostics for detecting collective structures (departures $\mathbf{Y}_k = \frac{\mathbf{y}_k - \mathbf{y}^{a*}_k}{\sqrt{\epsilon^{obs}_k + \epsilon^{a*}_k}}$ of individual observations are generally not sensitive enough).
A general diagnostic (flagging all observations which are not consistent with the assumed error characteristics) is found to be far to restrictive. Instead a more targeted variable which projects *obs* - $y_k^{a[l<k]}$ departures onto a cloud observation operator was found to be more suitable. The resulting cloud screening scheme corresponds well with that of Mc Nally & Watts if the possibility that part of the FG departures may be caused by background humidity or SST errors is discarded. Otherwise the new scheme has considerably less low level clouds.

5. Conclusions/Outlook

The CV method computes from (*obs - fg*) (y_k - $y_k^{a[l<k]}$), These
- have substantially smaller errors (the correlated part of **HBH**^{T}+**R** is subtracted)
- are (mutually) statistically independent

Here : application to IASI cloud screening has been outlined
Hope: method is useful also for screening other impacts like,
- e.g., surface influences (emissivity) not well represented by the employed observation operator

Method requires diagnostic filtering of collective structures which is
- sensitive enough to influences which should be filtered
- selective enough not to filter too many scenes ← determine important directions $\boldsymbol{h}$ in observation space

example above : project (y_k - $y_k^{a[l<k]}$) ☑ obs operator for cloud fraction

Disadvantage: Method is relatively complex, depends on employed error covariance matrixes
Advantage : Method is systematic, will benefit from advances in computing obs error covariances and background error covariances (e.g. Ensemble Kalman Filter)

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