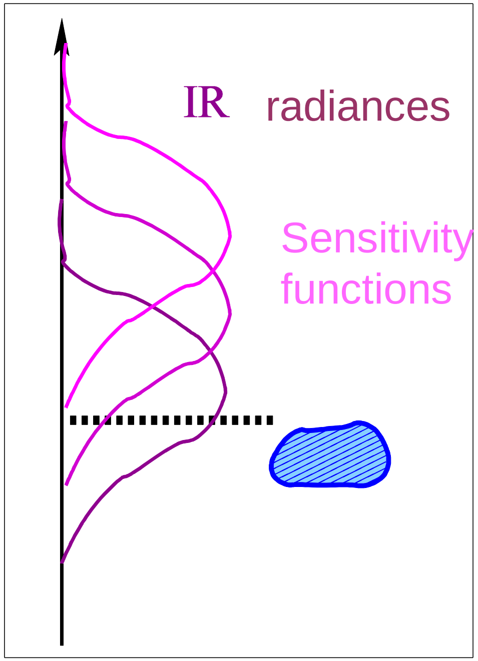
A probabilistic tool for the cross validation of observations

Olaf Stiller, Deutscher Wetterdienst

Are the observations consistent

- *with the background given*
  - *assumed errors*
  - *observation operator ?*
- *with the other observations ?*

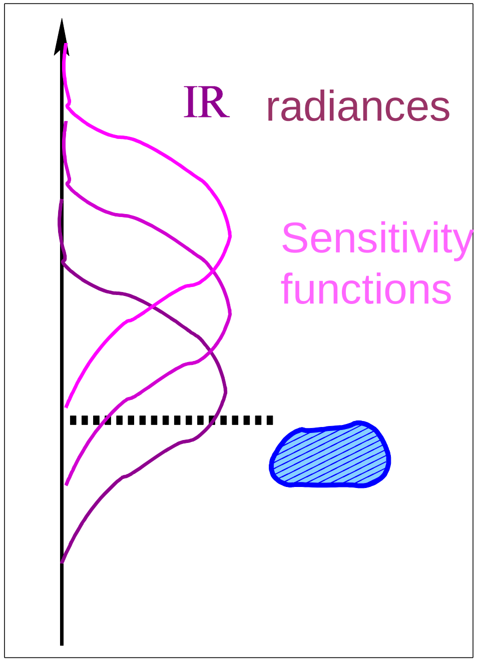


Does the observation operator (properly) deal with the cloud???

Aim: identify observations which are not consistent with the assumptions made in the DA system

First Guess checks

- Check consistency of observation with first guess
- Only works if inconsistencies are **very strong**



Are the observations consistent

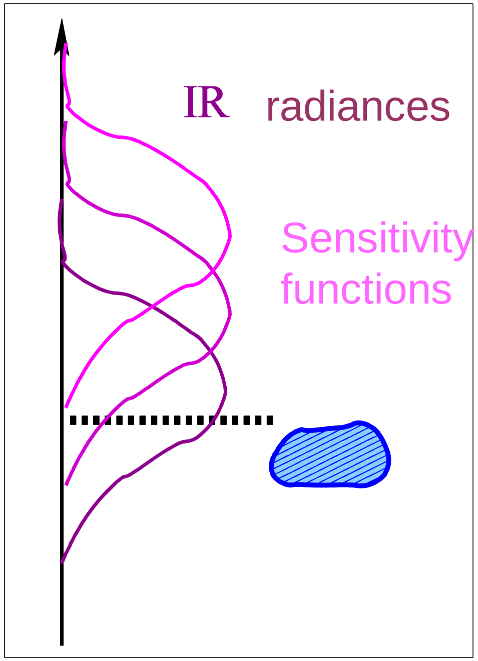
- *with the background given*
  - *assumed errors*
  - *observation operator ?*
- *with the other observations ?*

Does the observation operator (properly) deal with the cloud???

Olaf.Stiller@dwd.de

Aim: identify observations which are not consistent with the assumptions made in the DA system

Q. : Can we check the consistency of observations
- with other observations
- using the probabilistic framework of the DA system?



Are the observations consistent

- *with the background given*
  - *assumed errors*
  - *observation operator ?*
- *with the other observations ?*

Does the observation operator (properly) deal with the cloud???

Assumed uncertainties in data assimilation

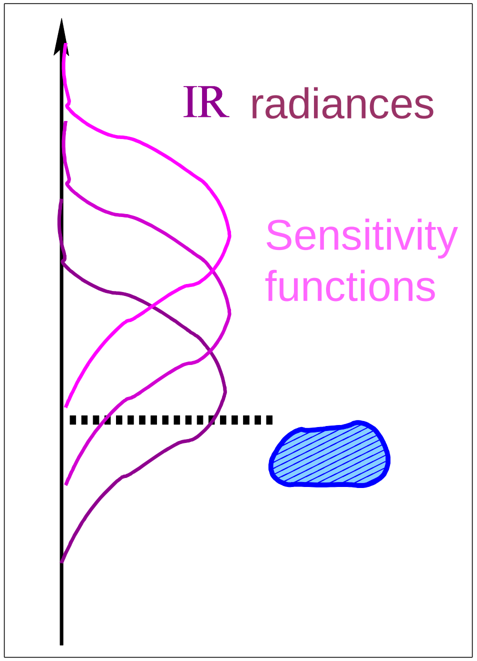
Most DA schemes minimise a quadratic costfunction

$$J(\mathbf{x}) = \frac{1}{2}\left[\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} + (\mathbf{y}^o - \mathbf{H}\mathbf{x})^T\,\mathbf{R}^{-1}\,(\mathbf{y}^o - \mathbf{H}\mathbf{x})\right]$$

$$\mathbf{x} = \mathbf{X} - \mathbf{X}^b$$

$$\mathbf{y}^o = \mathbf{Y}^o - \mathbf{Y}^b$$

$\mathbf{H}$: liearised observation operator



Assumed Gaussian error statistic:

Covariance matrix of

- Observation error $\quad \mathbf{R} = \left\langle (\mathbf{Y} - \mathbf{Y}^o)^T (\mathbf{Y} - \mathbf{Y}^o) \right\rangle$
- background error $\quad \mathbf{B} = \left\langle \left(\mathbf{X} - \mathbf{X}^b\right)^T \left(\mathbf{X} - \mathbf{X}^b\right) \right\rangle$
- background error in observation space $\quad \mathbf{H}^T\mathbf{B}\mathbf{H} = \left\langle (\mathbf{H}\mathbf{x})^T (\mathbf{H}\mathbf{x}) \right\rangle$
- $\mathbf{y}^o = \mathbf{Y}^o - \mathbf{Y}^b$
  obs - first guess $\quad \left\langle (\mathbf{y}^o)^T\,\mathbf{y}^o \right\rangle = \mathbf{H}^T\mathbf{B}\mathbf{H} + \mathbf{R}$

Validation against Background

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Assumed Gaussian error statistic:

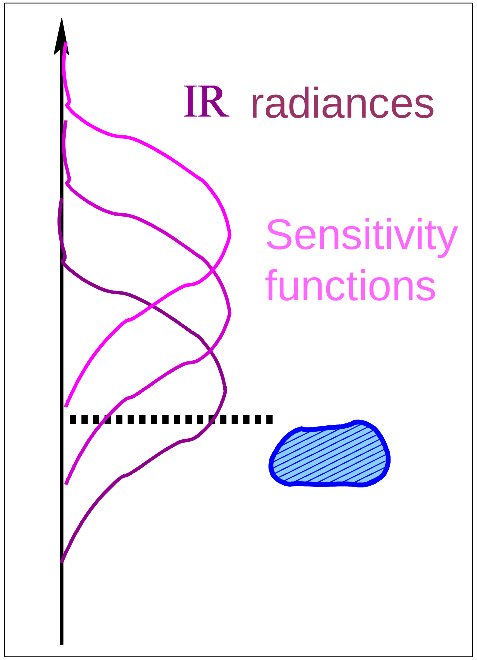
- $\mathbf{y}^o = \mathbf{Y}^o - \mathbf{Y}^b$
  obs - first guess

$$\left\langle (\mathbf{y}^o)^T \mathbf{y}^o \right\rangle = \mathbf{H}^T\mathbf{B}\mathbf{H} + \mathbf{R}$$

$$\left\langle \left(\mathbf{y}_k^o\right)^2 \right\rangle = \left[\mathbf{H}^T\mathbf{B}\mathbf{H} + \mathbf{R}\right]_{kk} = \sigma_k^2$$

Conditional probability of the observations $\mathbf{y}_k^o$ (given the background):

$$P(\mathbf{y}_k^o|\mathbf{X}^b) \;\propto\; \exp -\frac{1}{2}\left(\frac{\mathbf{y}_k^o}{\sigma_k}\right)^2$$



Cross-Validation with background (standard Quality Control check):

$$n \text{ sigma check:} \qquad \left|\frac{\mathbf{y}_k^o}{\sigma_k}\right| < n$$

Olaf.Stiller@dwd.de

Cross Validation

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Assumed Gaussian error statistic:

- $\mathbf{y}^o = \mathbf{Y}^o - \mathbf{Y}^b$
  obs - first guess

$$\left\langle (\mathbf{y}^o)^T \mathbf{y}^o \right\rangle = \mathbf{H}^T\mathbf{B}\mathbf{H} + \mathbf{R}$$

Cross-Validation with background and observations $\mathbf{y}^o_{\tau^C}$:

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 $\mathbf{y}^o_{\tau^C}$):

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

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

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

If R matrix is diagonal:

$$\overline{\mathbf{y}_\tau} = \mathbf{y}_\tau^{a\{\tau^C\}} = \mathbf{H}_\tau \left[\mathbf{x}^{a\{\tau^C\}}\right]$$

(Analysis using only $\mathbf{y}^o_{\tau^C}$)

Olaf.Stiller@dwd.de

Application: Identifying the Impact of Localized Features

Testing: *Can the observation operator properly deal with clouds?*

Radiance measurements can be **ordered** (at least roughly) according to their **likeliness** to be **affected by clouds**

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

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

$$[\mathbf{R} + \mathbf{HBH}^T] = \mathbf{T}_L \mathbf{T}_U$$

$$[\mathbf{R} + \mathbf{HBH}^T]^{-1} = \mathbf{T}_U^{-1}\mathbf{T}_L^{-1}$$

Cholesky decomposition

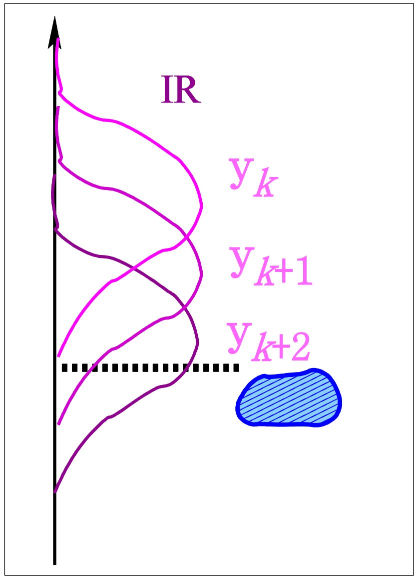
$$\mathbf{T}_U = \mathbf{T}_L^T$$

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

$$\mathbf{T}_L\mathbf{Y} = \begin{pmatrix} t_{11} & 0 & & & & \\ t_{21} & t_{22} & & & 0 & \\ & & \ddots & 0 & & \\ t_{k1} & t_{k2} & & t_{kk} & 0 & \\ & & & & \ddots & 0 \\ t_{p1} & & & & & t_{pp} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \\ \mathbf{Y}_k \\ \\ \mathbf{Y}_p \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \\ \mathbf{y}_k \\ \\ \mathbf{y}_p \end{pmatrix}$$

Stratospheric channels are never affected by clouds

Vulnerability increases the deeper weighting functions enter into the troposphere



Application: 1D Var for dealing with Localized Features

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

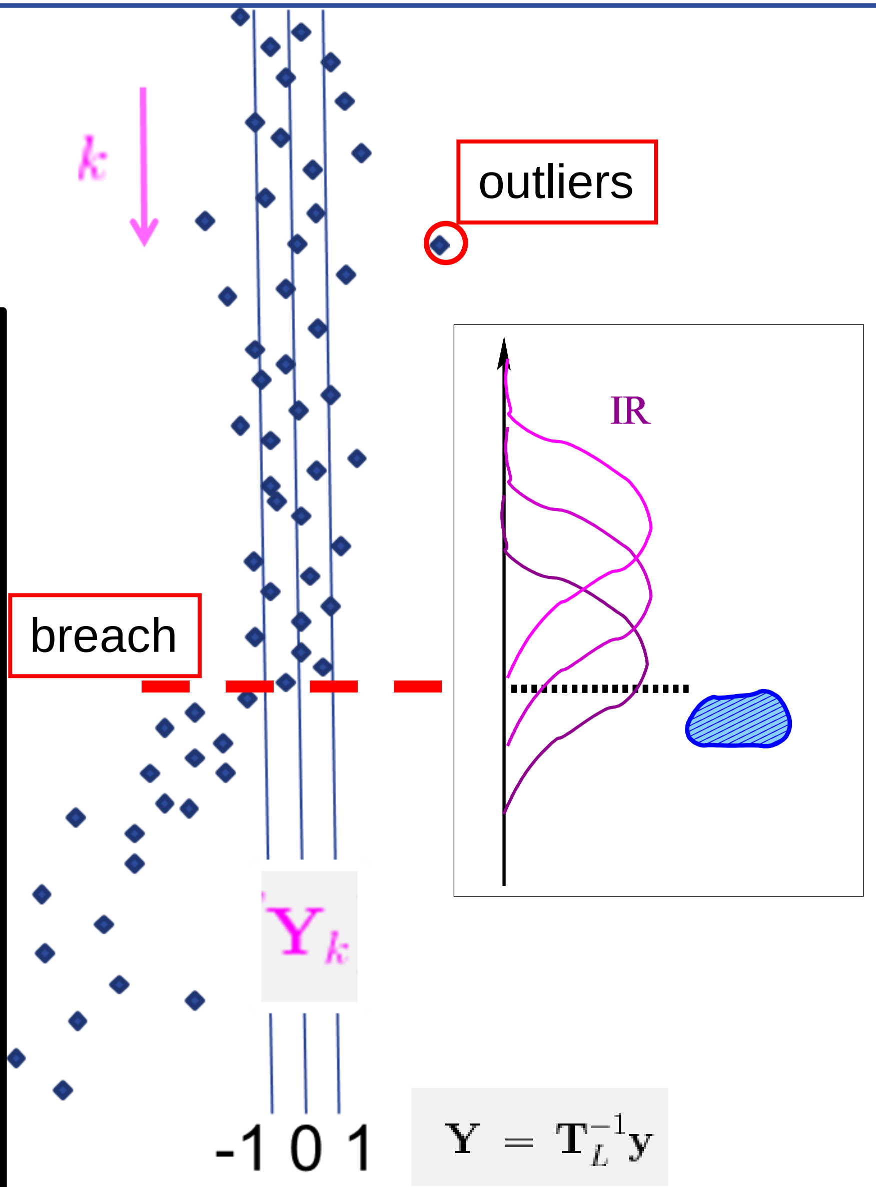
$$\begin{aligned} \mathbf{x}^a &= \mathbf{B}\mathbf{H}^T \left[\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T\right]^{-1} \mathbf{y} \\ &= \mathbf{B}\mathbf{H}^T \mathbf{T}_U^{-1}\mathbf{T}_L^{-1} \mathbf{y} \\ &= \mathbf{B}\mathbf{H}^T \mathbf{T}_U^{-1} \mathbf{Y} \end{aligned}$$

- Order observations according to their vulnarability
- Do outer_loop=1, Nmax

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

  - Do k =1, numb.of observations
    - Compute $\mathbf{Y}_k$
    - Flag (reject) outliers
    - Test for breaches
      - ✓ Reject all observations after breach
  - End Do

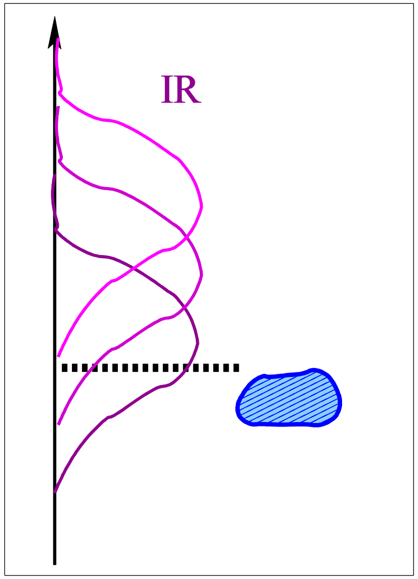
  Compute $\mathbf{x}^a = \mathbf{B}\mathbf{H}^T \mathbf{T}_U^{-1} \mathbf{Y}$ only from *good* observations
- End Do ! outer loops



Some applications of the 1D Var system

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- **Stand alone:**
  - Testing of observation operators dealing with strong nonlinearities related to
    - clouds
    - landsurfaces
    - any localised feature or perturbation

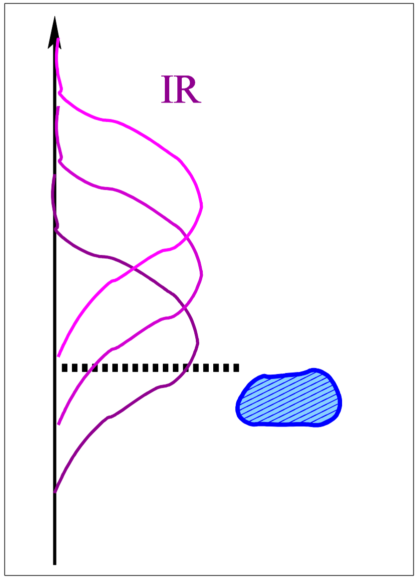


- **Inline with full DA system:**
  - Preprocessing (Quality Control) of satellite radiances: Flagging channels inconsistent with
    - the channels above
    - the linearised form of the observation operator **used at a given outer loop**

Summary and Conclusions

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- A cross validation method for observations has been developed which
  - works within the probabilistic framework of the DA system
  - puts a powerful constraint on observations if
    - observation operators sufficiently overlap
    - observation errors are sufficiently small



- The design of a 1D Var system has been outlined which
  - allows
    - observation cross validation at **almost no additional cost**
    - the detection of outliers
    - the detection of breaches
    - the flagging of bad observation **before** they enter into the analysis
  - can be used for
    - Stand alone: E.g., testing of observation operators
    - Inline with the full DA system: Preprocessing of observations
    -

Thank you for listening

Olaf.Stiller@dwd.de

Probability given background and observations $\mathbf{y}^o_{\tau^C}$:

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

$$\propto \exp -\frac{1}{2}\left\{\mathbf{z}_\tau^T \mathbf{D}_\tau^{-1} \mathbf{z}_\tau\right\}$$

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

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

Expectation value $\overline{\mathbf{y}_\tau}$ given background and observations $\mathbf{y}^o_{\tau^C}$:

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

$$\overline{\mathbf{y}_\tau} = \mathbf{y}_\tau^{a\{\tau^C\}} + \mathbf{R}_{\tau,\tau^C} \mathbf{R}_{\tau^C}^{-1} \left(\mathbf{y}^o_{\tau^C} - \mathbf{y}^{a\{\tau^C\}}_{\tau^C}\right)$$

$$\mathbf{y}_\tau^{a\{\tau^C\}} = \mathbf{H}_\tau \left[\mathbf{x}^{a\{\tau^C\}}\right]$$

(Analysis using only $\mathbf{y}^o_{\tau^C}$*)*

$$\text{if} \quad \mathbf{R}_{\tau,\tau^C} = \mathbf{0}$$

$$\mathbf{R}_{\tau,\tau^C} = \left\langle \left(\mathbf{y}^o_\tau - \mathbf{y}^t_\tau\right)\left(\mathbf{y}^o_{\tau^C} - \mathbf{y}^t_{\tau^C}\right)\right\rangle$$

error cross correlations:

$$(\mathbf{y}^o_\tau - \overline{\mathbf{y}_\tau}) = \left\{\left[\mathbf{R}^{-1}(\mathbf{1} - \mathbf{S})\right]_\tau\right\}^{-1} \mathbf{R}_\tau^{-1} (\mathbf{y}^o_\tau - \mathbf{y}^a_\tau)$$

$$\underbrace{\text{if} \quad \mathbf{R}_{\tau,\tau^C} = 0}_{=} \quad (\mathbf{1}_\tau - \mathbf{S}_\tau)^{-1} (\mathbf{y}^o_\tau - \mathbf{y}^a_\tau)$$

self sensitivity matrix:

$$\mathbf{S} = \mathbf{H}\mathbf{K}$$

Kalman gain matrix

compare*:* Liu, Kalnay, Miyoshi and Cardinali, QJ 2009 *: Analysis sensitivity calculation in an ensemble Kalman filter*

Olaf.Stiller@dwd.de