Use of observations in data assimilation

Gérald Desroziers

Météo-France, Toulouse, France

METEO FRANCE
Toujours un temps d'avance

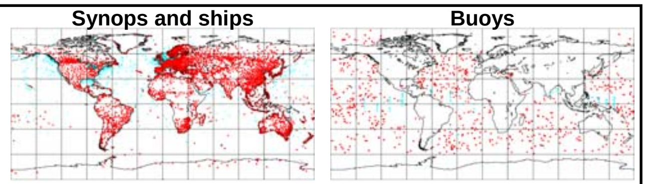
Outline

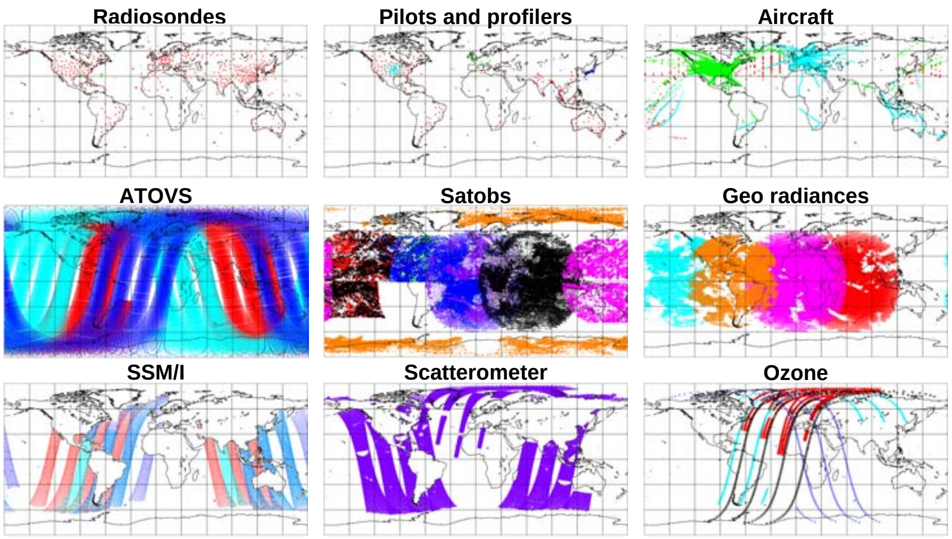
- Introduction
- Optimizing observation error statistics
- Ensembles based on a perturbation of observations
- Impact of observations on analyses and forecasts
- Conclusion and perspectives

Data coverage

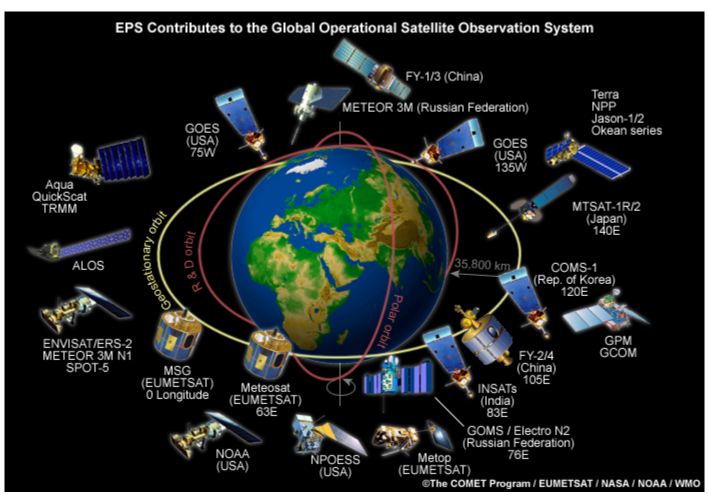
05/09/03 09–15 UTC

(courtesy J.-N. Thépaut)



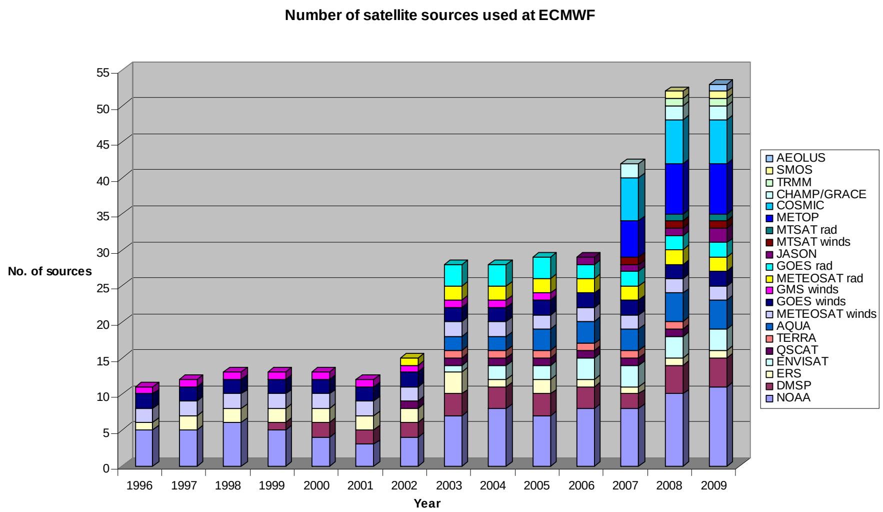


Satellites



(EUMETSAT)

In 2007, ECMWF uses ~ 40 different satellite data sources



(courtesy J-.N.Thépaut, ECMWF)

General formalism

- *Statistical linear estimation* :

$$\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x} = \mathbf{x}^b + \mathbf{K}\ \mathbf{d} = \mathbf{x}^b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})^{-1}\ \mathbf{d},$$

with $\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b)$, *innovation*, **K**, *gain matrix*,
B et **R**, covariances of background and observation errors,

- *H* is called « *observation operator* » (Lorenc, 1986),
- It is most often *explicit*,
- It can be non-linear (satellite observations)
- It can include an error,
- Variational schemes require *linearized* and *adjoint* observation operators,
- 4D-Var generalizes the notion of « *observation operator* » .

Statistical hypotheses

- Observations are supposed un-biased: $E(\varepsilon^o) = \mathbf{0}$.
- If not, they have to be preliminarily de-biased,
- or de-biasing can be made along the minimization (Derber and Wu, 1998; Dee, 2005; Auligné, 2007).
- Observation error variances are supposed to be known (diagonal elements of $R = E(\varepsilon^o \varepsilon^{oT})$).
- Observation errors are supposed to be un-correlated : (non-diagonal elements of $E(\varepsilon^o \varepsilon^{oT}) = 0$),
- but, the representation of observation error correlations is also investigated (Fisher, 2006) .

Outline

- Introduction
- Optimizing observation error statistics
- Ensembles based on a perturbation of observations
- Impact of observations on analyses and forecasts
- Conclusion and perspectives

A posteriori diagnostics

- Is the system consistent?
- We should have

$$E[J(\mathbf{x}^a)] = p,$$

p = total number of observations,

- but also

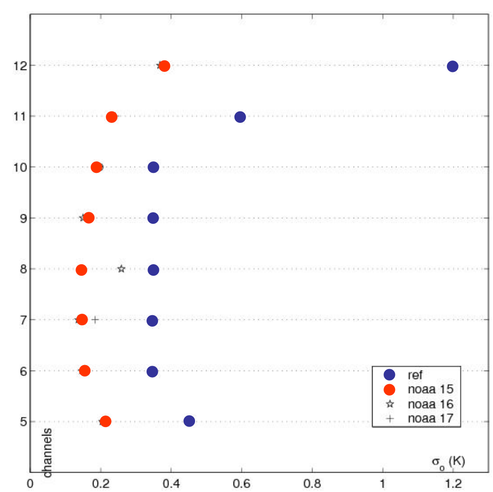
$$E[J^o_i(\mathbf{x}^a)] = p_i - Tr(\mathbf{R}_i^{-1/2}\,\mathbf{H}_i\,\mathbf{A}\,\mathbf{H}_i^T\mathbf{R}_i^{-1/2}),$$

p_i : number of observations associated with J^o_i

(Talagrand, 1999) .

- Computation of optimal $E[J^o_i(\mathbf{x}^a)]$ by a Monte-Carlo procedure is possible.
(Desroziers and Ivanov, 2001) .

Application : optimization of R



One tries to obtain

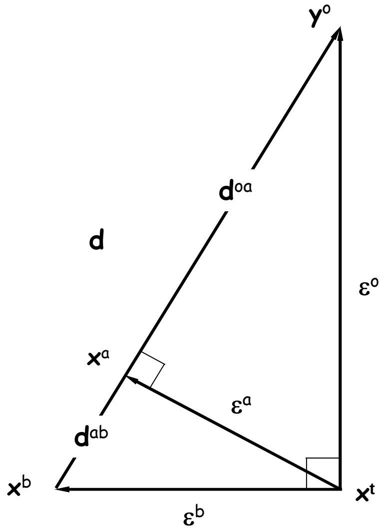
$E[J^o_i(\mathbf{x}^a)] = (E[J^o_i(\mathbf{x}^a)])^{opt}$.

by adjusting the σ^o_i

Optimization of HIRS σ^o

(Chapnik, et al, 2004; Buehner, 2005)

Diagnostics / *observations*



(Desroziers et al, 2005)

$\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b)$

$\mathbf{d}^{oa} = \mathbf{y}^o - H(\mathbf{x}^a) = (\mathbf{I}\text{-}\mathbf{HK})\ \mathbf{d}$

$\mathbf{d}^{ab} = H(\mathbf{x}^a) - H(\mathbf{x}^b) = \mathbf{HK}\ \mathbf{d}$

$E[\mathbf{d}^{oa}\ \mathbf{d}^T] = (\mathbf{I}\text{-}\mathbf{HK})\ E[\mathbf{d}\ \mathbf{d}^T] = \mathbf{R}$

$E[\mathbf{d}^{ab}\ \mathbf{d}^T] = \mathbf{HK}\ E[\mathbf{d}\ \mathbf{d}^T] \quad = \mathbf{HBH}^T$

$\langle \varepsilon, \varepsilon' \rangle = E[\varepsilon\ \varepsilon']$

Implementation in 4D-Var

For any subset i with p_i observations, simply compute

$$(\boldsymbol{\sigma}_i^b)^2 = (\mathbf{d}^{ab})_i^T\ \mathbf{d}_i /\ p_i$$

$$= \sum_{j=1}^{p_i} (y_j^a - y_j^b)(y_j^o - y_j^b)\ /\ p_i$$

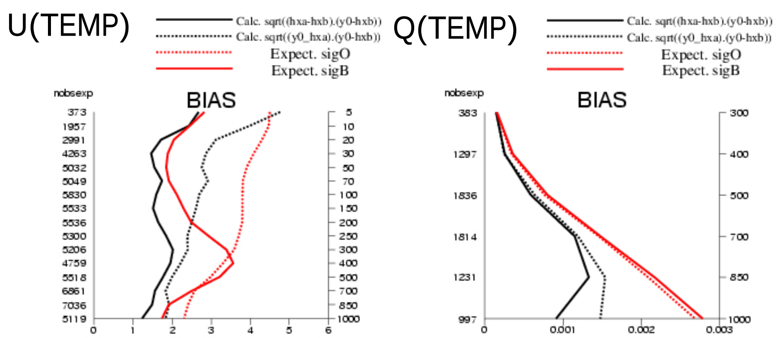
and

$$(\boldsymbol{\sigma}_i^o)^2 = (\mathbf{d}^{oa})_i^T\ \mathbf{d}_i /\ p_i$$

$$= \sum_{j=1}^{p_i} (y_j^o - y_j^a)(y_j^o - y_j^b)\ /\ p_i$$

This is nearly cost-free
and can be computed, a posteriori, over one or several analyses.

Implementation in 4D-Var



Computation over four 6h 4D-Var analyses (one day)

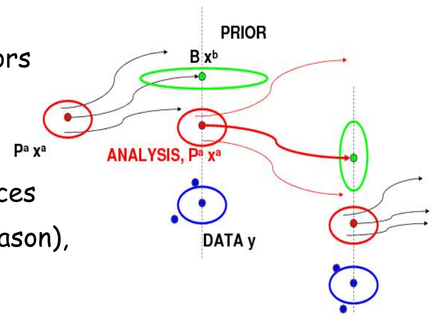
Outline

- Introduction
- Optimizing observation error statistics
- Ensembles based on a perturbation of observations
- Impact of observations on analyses and forecasts
- Conclusion and perspectives

Ensemble of perturbed analyses

- Simulation of the estimation errors along analyses and forecasts.

- Documentation of error covariances
  - over a long period (a month/ a season),
  - for a particular date.



(Ehrendorfer, 2006)

(Evensen, 1997; Fisher, 2004; Berre et al, 2007)

Ensembles based on a perturbation of observations

The same analysis equation and (sub-optimal) operators $\mathbf{K}$ and $\mathbf{H}$ are involved in the equations of $\mathbf{x}^a$ and ε^a:

$$\mathbf{x}^a = (\mathbf{I} - \mathbf{KH})\, \mathbf{x}^b + \mathbf{K}\, \mathbf{x}^o$$

$$\varepsilon^a = (\mathbf{I} - \mathbf{KH})\, \varepsilon^b + \mathbf{K}\, \varepsilon^o$$

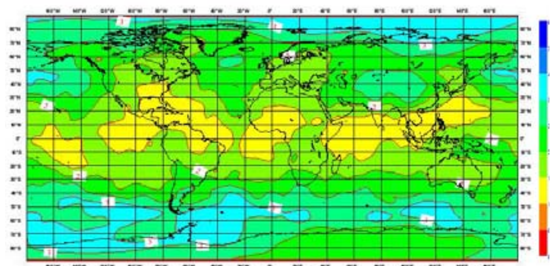
The same equation also holds for the analysis perturbation:

$\mathbf{e}^a = (\mathbf{I} - \mathbf{KH})\, \mathbf{e}^b + \mathbf{K}\, \mathbf{e}^o$, with $\mathbf{e}^o = \mathbf{R}^{1/2}\, \eta$ and η a vector of random numbers

Covariance matrix of analysis error:

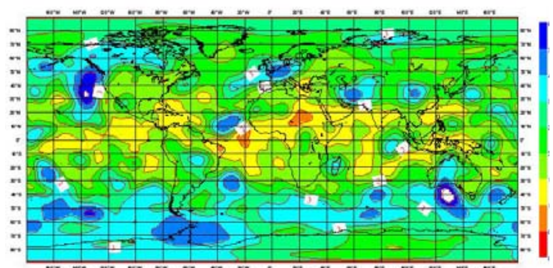
$$\mathbf{P}^a = E(\mathbf{e}^a\, \mathbf{e}^{aT}) = (\mathbf{I} - \mathbf{KH})\, E(\mathbf{e}^b\, \mathbf{e}^{bT})\, (\mathbf{I} - \mathbf{KH})^T + \mathbf{K}\, E(\mathbf{e}^o\, \mathbf{e}^{oT})\, \mathbf{K}^T$$

Background error standard-deviations



Over a month

Vorticity at 500 hPa



For a particular date
08/12/2006 00H
6 member ensemble
Vorticity at 500 hPa

Validation/tuning of ensemble variances

- Background errors $\mathbf{e}^b$ can be projected to observation space by applying the observation operator:
  $\mathbf{H}\,\varepsilon^b$.

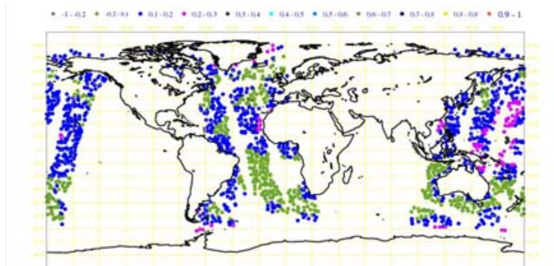
- Using several $\mathbf{e}^b$, ensemble variances in observation space can be computed and compared to

  - variances of innovations (d) minus variances of observation errors

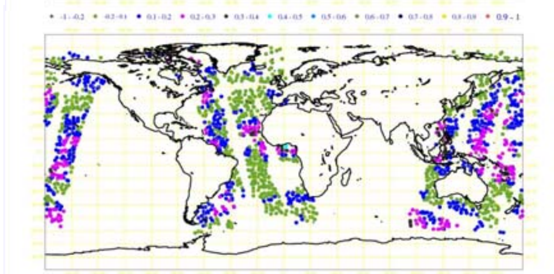
  - or background error variances given by statistics of d^{ab} x d cross products

Ensemble / diagnosed bg error std-dev HIRS 5 (28/8/6 00h)

Ensemble
(6 members)



Diagnosed from
d^{ab} x d
cross-products



Outline

- Introduction
- Optimizing observation error statistics
- Ensembles based on a perturbation of observations
- Impact of observations on analyses and forecasts
- Conclusion and perspectives

Measure of the impact of observations

- Total reduction of estimation error variance:
  $r = Tr(\mathbf{K}\ \mathbf{H}\ \mathbf{B})$
- Reduction due to observation set i :
  $r_i = Tr(\mathbf{K}_i\ \mathbf{H}_i\ \mathbf{B})$
- Variance reduction normalized by $\mathbf{B}$:
  $r_i^{DFS} = Tr(\mathbf{K}_i\ \mathbf{H}_i) = E[J^o_i(\mathbf{x}^a)]$
- Reduction of error projected onto a variable/area:
  $r_i^P = Tr(\mathbf{P}\ \mathbf{K}_i\ \mathbf{H}_i \mathbf{B}\ \mathbf{P}^T)$
- Reduction of error evolved by a forecast model:
  $r_i^{PM} = Tr(\mathbf{P}\ \mathbf{M}\ \mathbf{K}_i\ \mathbf{H}_i \mathbf{B}\ \mathbf{M}^T \mathbf{P}^T) = Tr(\mathbf{L}\ \mathbf{K}_i\ \mathbf{H}_i \mathbf{B}\ \mathbf{L}^T)$

(Cardinali, 2003; Fisher, 2003; Chapnik et al, 2006)

Randomized estimates of error reduction on analyses and forecasts

It can be shown that

$$r_i = tr(\mathbf{L}\mathbf{K}_i\ \mathbf{H}\mathbf{B}\ \mathbf{L}^T) \\ = tr(\mathbf{H}_i\ \mathbf{B}\ \mathbf{L}^T\ \mathbf{L}\mathbf{K}).$$

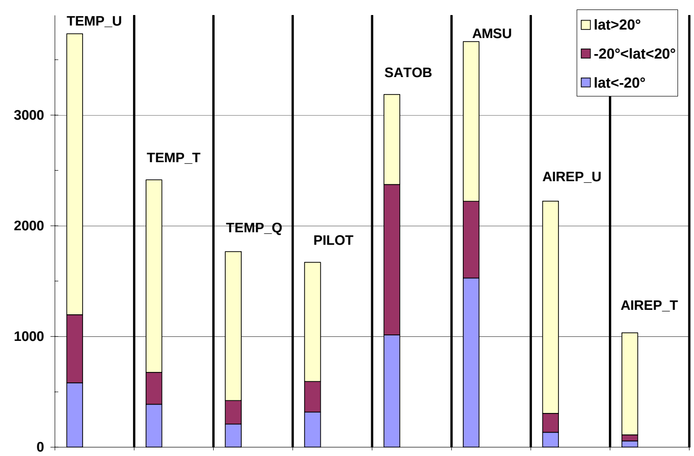
This can be estimated by a randomization procedure:

$$r_i \approx \sum_j (\delta\ \mathbf{y}_i^o)_j^T\ \mathbf{R}_i^{-1}\ \mathbf{H}_i \mathbf{B}\ \mathbf{L}^T\ \mathbf{L}\mathbf{K}(\delta\ \mathbf{y}^o)_j \\ \approx \sum_j (\delta\ \mathbf{y}^o)_j^T\ R_i^{-1} H_i B^{1/2} B^{T/2} L^* L^{'} (\delta\ \mathbf{x}^a)_j$$

where $(\delta\ \mathbf{y}^o)_j$ is a vector of observation perturbations and $(\delta\ \mathbf{x}^a)_j$ the corresponding perturbation on the analysis.

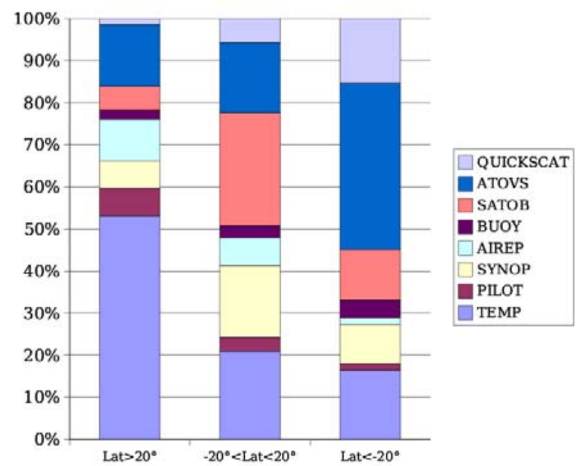
(Fisher, 2003; Desroziers et al, 2005)

Degree of Freedom for Signal (DFS)



(Chapnik et al, 2006)

Error variance reduction



% of error variance reduction for T 850 hPa
by area and observation type

Conclusion and perspectives

- Observation operators allow the use of a wide range of observations

- Statistics on observation errors:
  - variances not perfectly known, but may be tuned by using optimality criteria

- Simulation of analysis and background errors:
  - can be obtained by ensembles based on a perturbation of observations
  - variance of ensembles might be validated in the space of observations

- Measure of the impact of observations:
  - how observations contribute to the reduction of the uncertainty in the analyzed and forecast state?
  - which are the most useful observations?
  - observation impacts are by-products of ensembles based on a perturbation of observations