Observability of flow dependent structure functions and their use in data assimilation

Pierre Gauthier

Based on work done in collaboration with

Cristina Lupu (MSc 2006, PhD thesis 2010, UQAM)

and Stéphane Laroche (Env. Canada)

Presentation at the

2010 ESA Earth Observation Summer School on

Earth system monitoring and modeling

Frascati, Italy, 2-13 August 2010

UQÀM Department of Earth and Atmospheric Sciences
Université du Québec à Montréal

Outline

- **Measuring the impact of observations in data assimilation systems**
  - * Impact on the analysis (information content)
  - * Impact on short-term forecasts based on adjoint methods
- **Impact of flow-dependent structures in data assimilation and link with precursors to dynamic instability**
  - * Evaluation of the observability of structure functions (Lupu, 2010)
- **Implications for a hybrid 4D-Var**
  - * The earlier experiments of Fisher and Andersson with a reduced rank Kalman filter
  - * The hybrid 4D-Var/EnKF (Buehner *et al.*, 2009)
- **Conclusions**

Statistical nature of the assimilation

* To correct a short-term forecast ($\mathbf{x}_b$, the background state) with error covariance **B** based on information contained in observation **y** with observation error covariance **R**
* The resulting analysis $\mathbf{x}_a$ has an accuracy measured by its error covariance $\mathbf{P}_a$ which is "less" than that of the background

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$
$$\mathbf{P}_a = \mathbf{B} - \mathbf{KHB}$$

* The weight is given by the gain matrix **K** set to minimize the total analysis error variance

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$$

* Observation operator **H** has been *linearized* around the current background state

Approaches to measuring the impact of assimilated observations

Information content

* based on the relative accuracy of observations and the background state

Observing System Experiments

* Data denials
* Global view of the impact of observations on the quality of the forecasts

Observation impact on the quality of the forecasts

* Sensitivities with respect to observations based on adjoint methods (Baker and Daley, 2000; Langland and Baker, 2003)
* Ensemble Kalman filter methods

Information content

- **Ratio of the analysis error covariance to B**

$$tr\left(\mathbf{P}_a\mathbf{B}^{-1}\right) = tr(\mathbf{I}) - tr(\mathbf{KH}) = N - tr(\mathbf{KH})$$

The information gained from assimilating a given set of observations is represented by the second term, where N is the dimension of the *model space*

DFS = Degrees of Freedom per signal

- **… and in observation space**

$$\mathbf{P}_a \rightarrow \mathbf{HP}_a\mathbf{H}^T \quad \mathbf{B} \rightarrow \mathbf{HBH}^T$$

$$tr\left(\left(\mathbf{HP}_a\mathbf{H}^T\right)\left(\mathbf{HBH}^T\right)^{-1}\right) = M - tr(\mathbf{HK})$$

with M being the number of observations

Diagnosing the statistical information from the results of analysis

- **Desroziers (2005)**
  - use the results of the assimilation to estimate the observation, background and analysis error covariances in observation space

$$\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}_b \quad \mathbf{a} = \mathbf{y} - \mathbf{H}\mathbf{x}_a \quad \mathbf{d}_a^b = \mathbf{H}(\mathbf{x}_a - \mathbf{x}_b) = \mathbf{HKd}$$

  - and then,

$$\langle \mathbf{d}\mathbf{d}^T \rangle \equiv \widetilde{\mathbf{D}} = (\mathbf{R} + \mathbf{HBH}^T) \equiv \mathbf{D}$$

$$\langle \mathbf{a}\mathbf{d}^T \rangle \equiv \widetilde{\mathbf{R}} = \mathbf{R}(\mathbf{D}^{-1}\widetilde{\mathbf{D}})$$

$$\langle \mathbf{d}_a^b \mathbf{d}^T \rangle \equiv \mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T = \mathbf{HBH}^T(\mathbf{D}^{-1}\widetilde{\mathbf{D}})$$

$$\langle \mathbf{d}_a^b \mathbf{a}^T \rangle = \mathbf{H}\widetilde{\mathbf{P}}_a\mathbf{H}^T = \mathbf{HBH}^T\mathbf{D}^{-1}\widetilde{\mathbf{D}}\mathbf{D}^{-1}\mathbf{R}$$

Estimating the information content (or Degrees of Freedom per signal)

- **Noticing that** $\tilde{\mathbf{R}} = \mathbf{R}\left(\mathbf{D}^{-1}\tilde{\mathbf{D}}\right)$ $\quad \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T = \mathbf{HBH}^T\left(\mathbf{D}^{-1}\tilde{\mathbf{D}}\right)$

- **If the *a priori* and *a posteriori* error statistics are consistent, then** $\tilde{\mathbf{D}} = \mathbf{D}$ **and therefore,**

$$\tilde{\mathbf{R}} = \mathbf{R} \qquad \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T = \mathbf{HBH}^T$$

- **Estimation of the DFS**

$$DFS = tr(\mathbf{HK}) = tr\left(\mathbf{HBH}^T\left(\mathbf{R} + \mathbf{HBH}^T\right)^{-1}\right) = tr\left(\mathbf{HBH}^T\mathbf{D}^{-1}\right)$$

$$tr\left(\mathbf{H}\tilde{\mathbf{K}}\right) = tr\left(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T\tilde{\mathbf{D}}^{-1}\right) = tr\left(\mathbf{HBH}^T\mathbf{D}^{-1}\tilde{\mathbf{D}}\tilde{\mathbf{D}}^{-1}\right) = DFS$$

This gives the same information content as obtained from the *a priori* error statistics

Estimating the information content

- **Estimate of the information content is based solely on diagnostics from the assimilation process**

$$\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}_b \quad \mathbf{a} = \mathbf{y} - \mathbf{H}\mathbf{x}_a \quad \mathbf{d}_a^b = \mathbf{H}(\mathbf{x}_a - \mathbf{x}_b) = \mathbf{HKd}$$

$$DFS = tr\left(\left\langle \mathbf{d}_b^a \mathbf{d}^T \right\rangle \left\langle \mathbf{d}\mathbf{d}^T \right\rangle^{-1}\right) = \left\langle tr\left(\mathbf{d}_b^a \mathbf{d}^T \left\langle \mathbf{d}\mathbf{d}^T \right\rangle^{-1}\right)\right\rangle = \left\langle \mathbf{d}^T \left\langle \mathbf{d}\mathbf{d}^T \right\rangle^{-1} \mathbf{d}_b^a \right\rangle$$

- **Need to estimate and *invert* $\left\langle \mathbf{d}\mathbf{d}^T \right\rangle$ which is a full matrix because it contains the background error**
- **Alternate form**

$$DFS = tr\left(\widetilde{\mathbf{R}}^{-1}\mathbf{H}\widetilde{\mathbf{P}}_a\mathbf{H}^T\right) = tr\left(\left\langle \mathbf{a}\mathbf{d}^T \right\rangle^{-1} \left\langle \mathbf{d}_a^b \mathbf{a}^T \right\rangle\right)$$

$$= \left\langle \mathbf{a}^T \left\langle \mathbf{a}\mathbf{d}^T \right\rangle^{-1} \mathbf{d}_b^a \right\rangle = \left\langle \mathbf{a}^T \widetilde{\mathbf{R}}^{-1} \mathbf{d}_b^a \right\rangle$$

Additional assumption: $\widetilde{\mathbf{R}}$ is diagonal

Robustness of the estimate: experiments with a simple 1D-Var

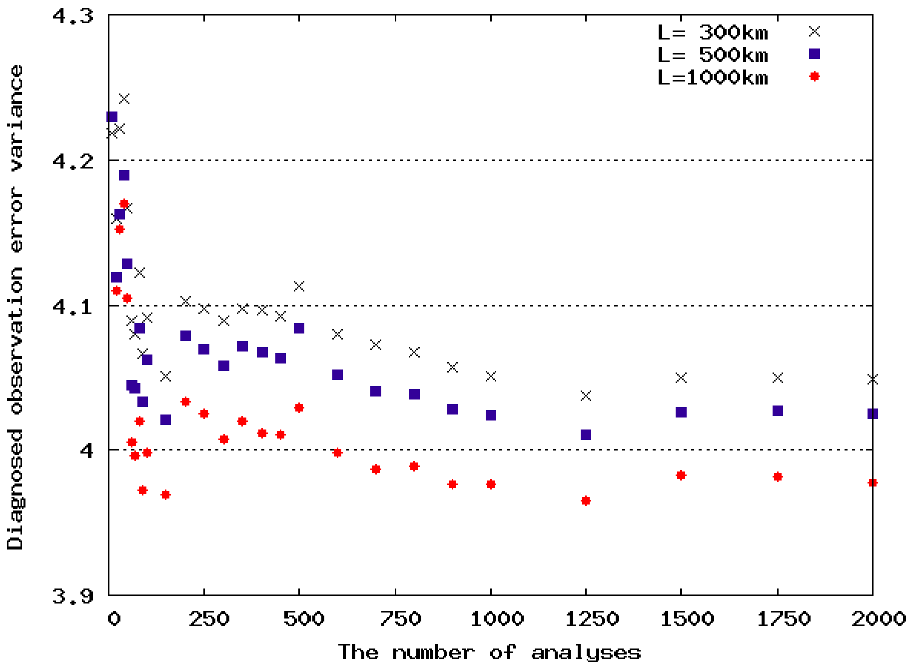
- 1Dvar assimilation of 60 observations with a covariance model with homogeneous and isotropic correlations
- Statistical average over 2000 analyses

L(km)	$\sigma_o^2 = \sigma_{o(t)}^2$	$\sigma_b^2 = \sigma_{b(t)}^2$	$(\tilde{\sigma}_o^2)$	$(\tilde{\sigma}_b^2)$
300	4.	1.	4.04	0.98
500	4.	1.	4.02	0.96
1000	4.	1.	3.98	0.94

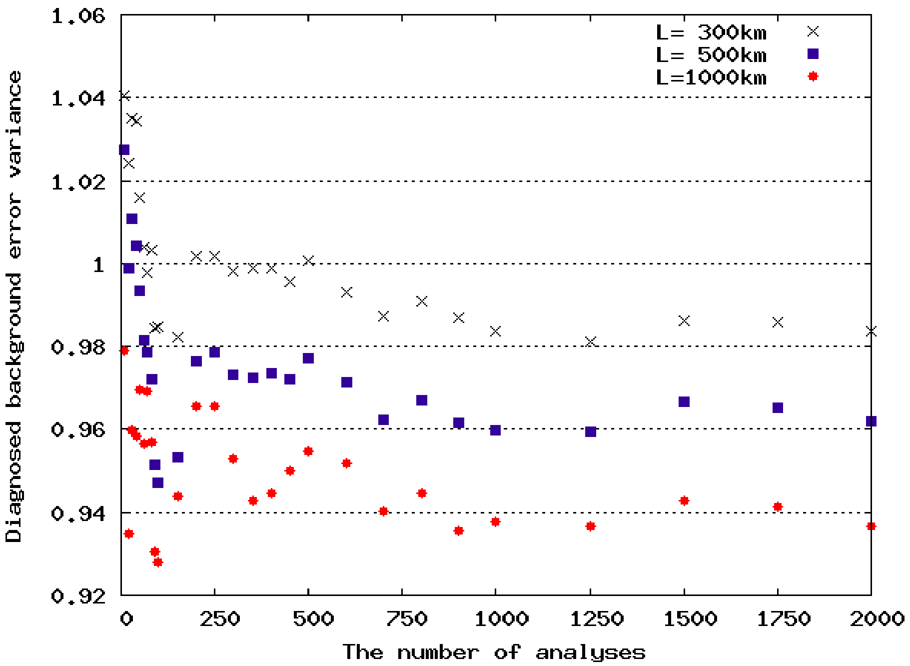
,

Robustness of the results with the size of the sample

Observation error

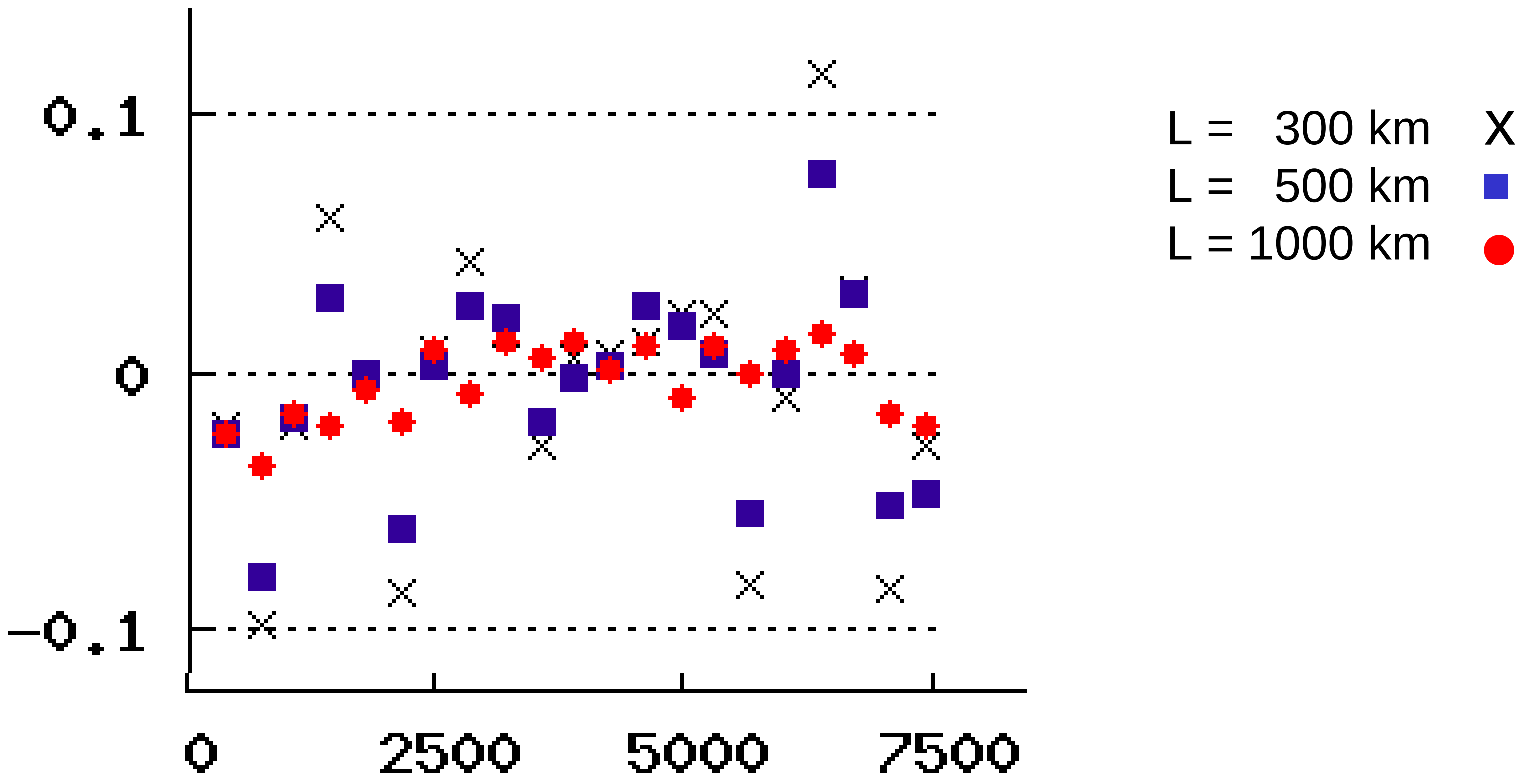


Background error



Estimating the observation error covariance $\tilde{R}$

- **Estimate of the off-diagonal terms of** $\tilde{\mathbf{R}}_{i,j} = \overline{\left(\mathbf{a}_i \mathbf{d}_j\right)}$ **as a function of distance** $\mathbf{r_{i,j}}$



Estimation of the information content

$$D\widetilde{F}S^{(1)}_{APOST} = tr\left(\mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T\widetilde{\mathbf{D}}^{-1}\right)$$

$$D\widetilde{F}S^{(2)}_{APOST} = tr\left(\widetilde{\mathbf{R}}^{-1}\left(\mathbf{H}\widetilde{\mathbf{P}}_a\mathbf{H}^T\right)\right)$$

$D\widetilde{F}S_{DIAG}$: only the diagonal terms of the second method are used

L (km)	DFS_{THEOR}	DFS_{GIRARD}	$D\widetilde{F}S^{(1)}_{APOST}$	$D\widetilde{F}S^{(2)}_{APOST}$	$D\widetilde{F}S_{DIAG}$
300	11.03	10.88	10.81	10.80	10.70
500	9.50	9.37	9.21	9.20	9.07
1000	7.34	7.08	6.79	6.79	6.75

Easiest to compute

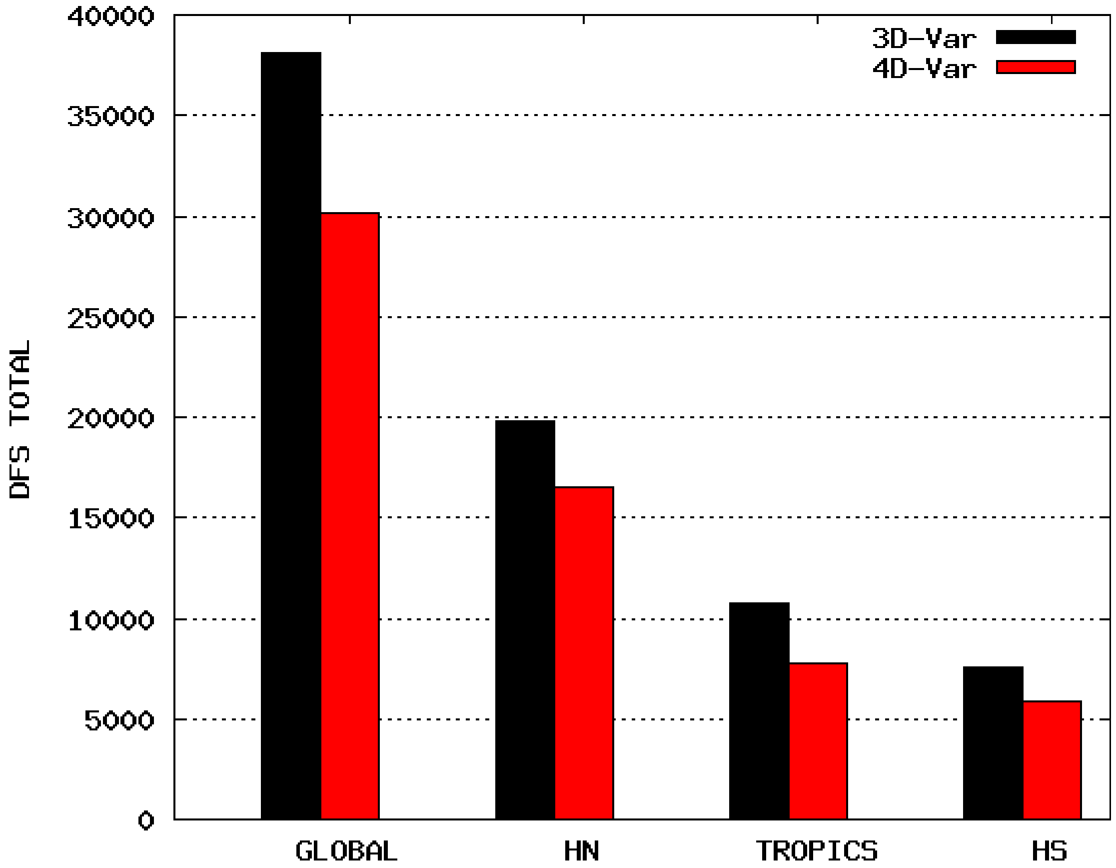
DFS_{GIRARD} : estimation obtained from perturbed analysis

DFS_{THEOR} : estimation obtained from the true values

Information content in 3D-Var and 4D-Var analyses from Environment Canada's system

- **Results from the assimilation experiments of Laroche and Sarrazin (2010 a,b) over the period December 21, 2006 to February 28, 2007**
  - Exclude the first 11 days (spin-up of the assimilation cycle)
- **Observations include**
  - Radiosondes, aircraft, surface and ship data, wind profilers
  - Atmospheric motion vectors from geostationary satellites
  - Radiances from polar-orbiting satellites (AMSU-a,b) and geostationary satellites (GOES-East and West)
- **Diagnostic of statistical consistency: $\chi^2 / M \sim 1$**
  - Both in 3D-Var and 4D-Var it was found that $\chi^2 / M = 0.56$
  - Error statistics used in the system are overestimated
  - Desroziers and Ivanov (2001) and Chapnik et al. (2004) use this information to recalibrate the statistics
  - This was not the object of this work

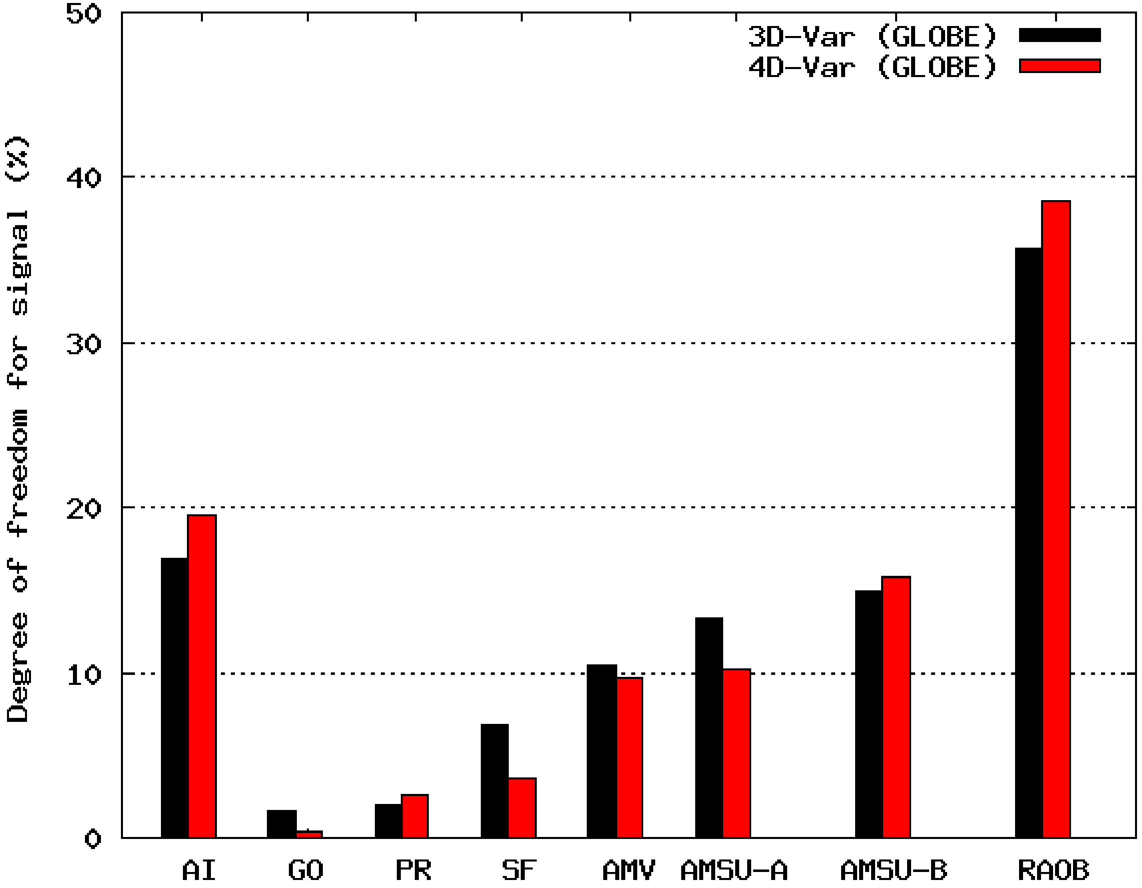
Total DFS estimated over different regions for 3D-Var and 4D-Var (January-February 2007)



Computation of DFS for each type of observations *in MSC's* 3D-Var *and* 4D-Var *systems*

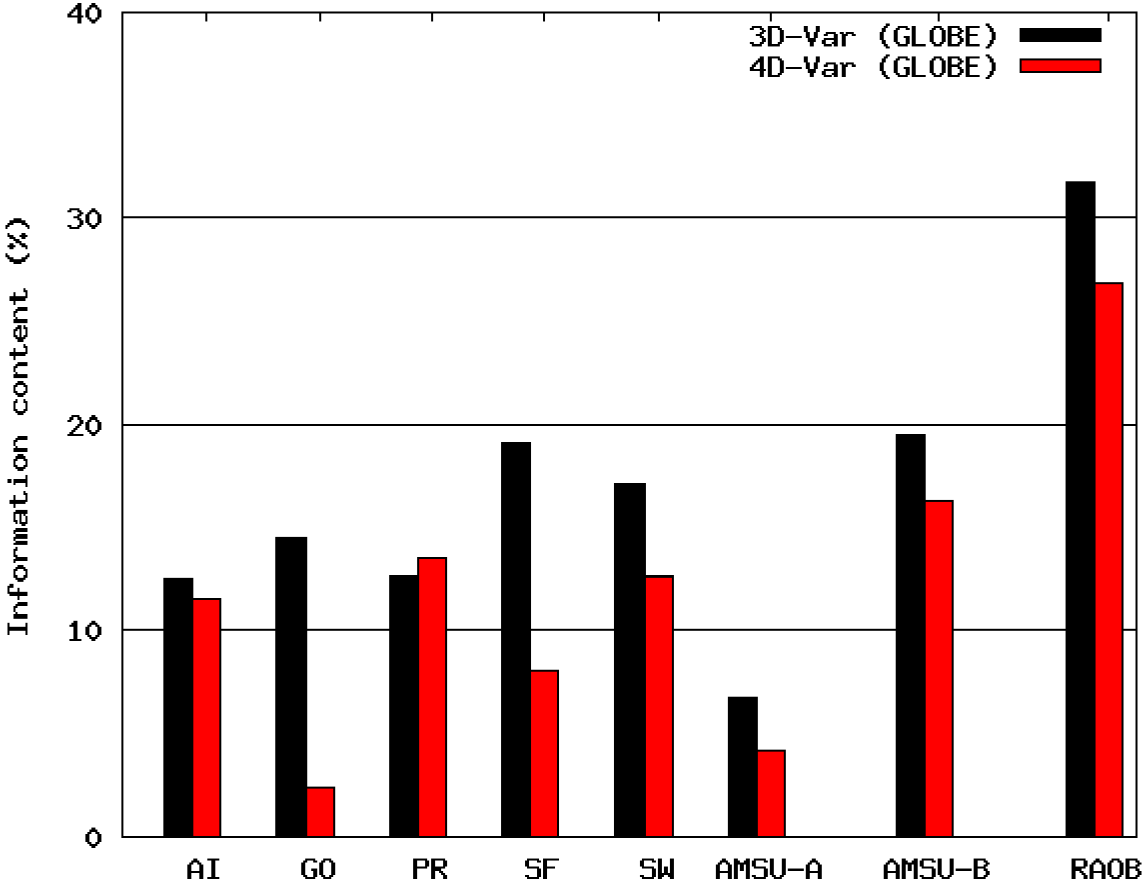
$$\mathrm{DFS}^{\mathrm{Region}}_{\mathrm{Obs\_type}}(\%) = 100 \cdot \frac{\mathrm{DFS}^{\mathrm{Region}}_{\mathrm{Obs\_type}}}{\mathrm{DFS}^{\mathrm{Globe}}_{\mathrm{Total\_obs}}}$$

Region : Globe
Obs_types : AI, GO, PR, SF, SW, AMSU-A, AMSU-B, RAOB



Lupu *et al.* (2010)

Observation impact per observation in each region



$$IC(\%) = 100 \cdot \frac{DFS_k^{Region}}{p_k}$$

Lupu et al. (2010)

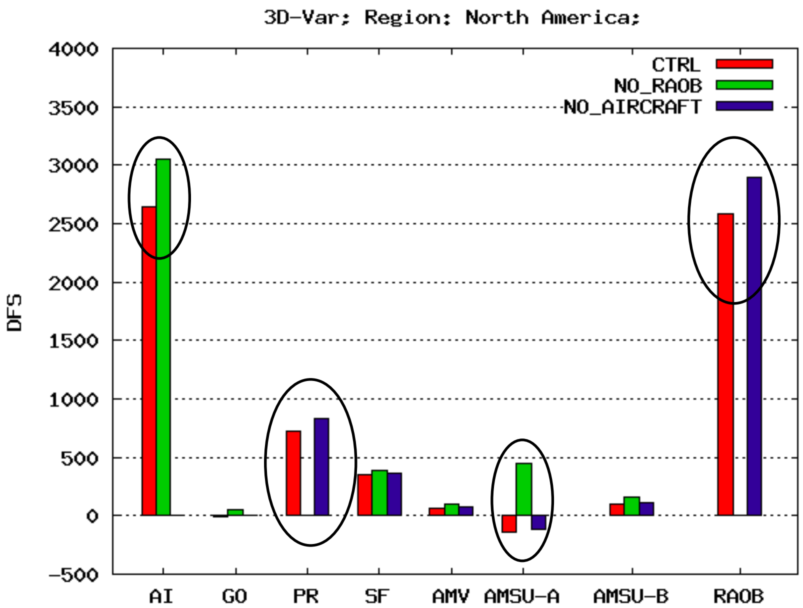
Observing System Experiments (OSEs)

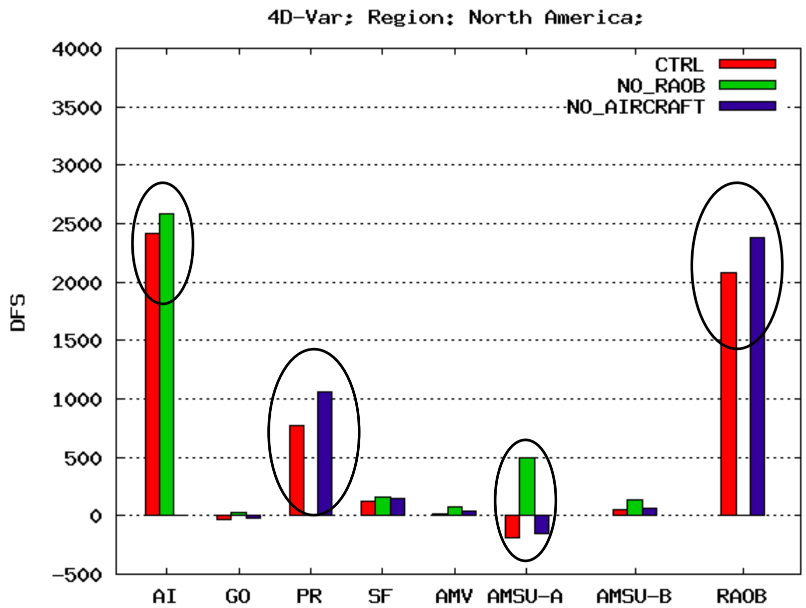
- **Experiments reported in Laroche and Sarrazin (2010-a,b)**
- **Evaluation of the impact of observations through data denials**
    * Take an analysis using all observations as a reference and then remove one observation type and measure the degradation
    * Modification of the observation environment alters the relative importance of the observations
- **Comparison of the information content for these experiments gives a detailed view of the interactions between observations**

OSEs experiments: ***3D-Var and 4D-Var, North America***

$$\frac{DFS_k^{NA}}{p_k}$$

DFS values per obstype normalized by the number of observations.





NO_RAOB: DFS per single observation notably increases, especially for AMSU-A and GO;

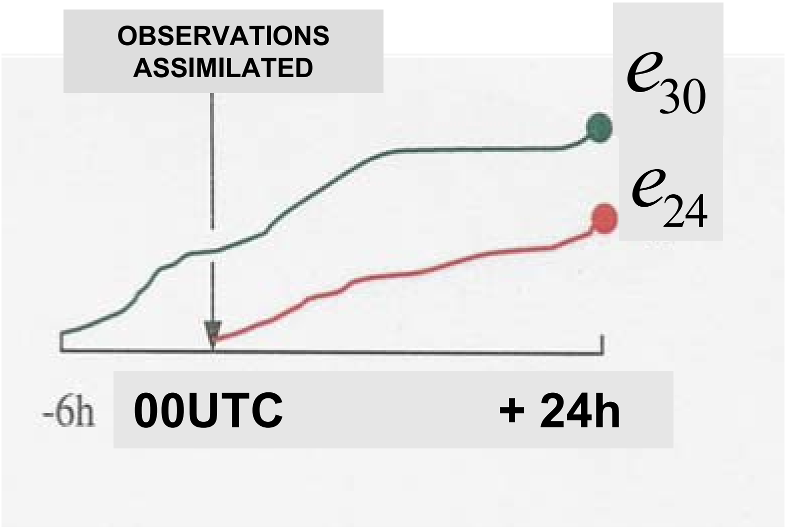
NO_AIRCRAFT: DFS per single observation notably increases, especially for RAOB and PR; For other observations (GO, SW and AMSU-B) DFS per obs also increases slightly.

Summary

- Information content can be evaluated by diagnosing the results of an assimilation
- Provides a detailed view of the impact of the observations within the original observation environment
- Application to the results from OSEs show how the impact of observations on analyses depend on the observation environment
- OSEs on the other hand measure the impact of observations on the subsequent forecasts

Observation Impact Methodology

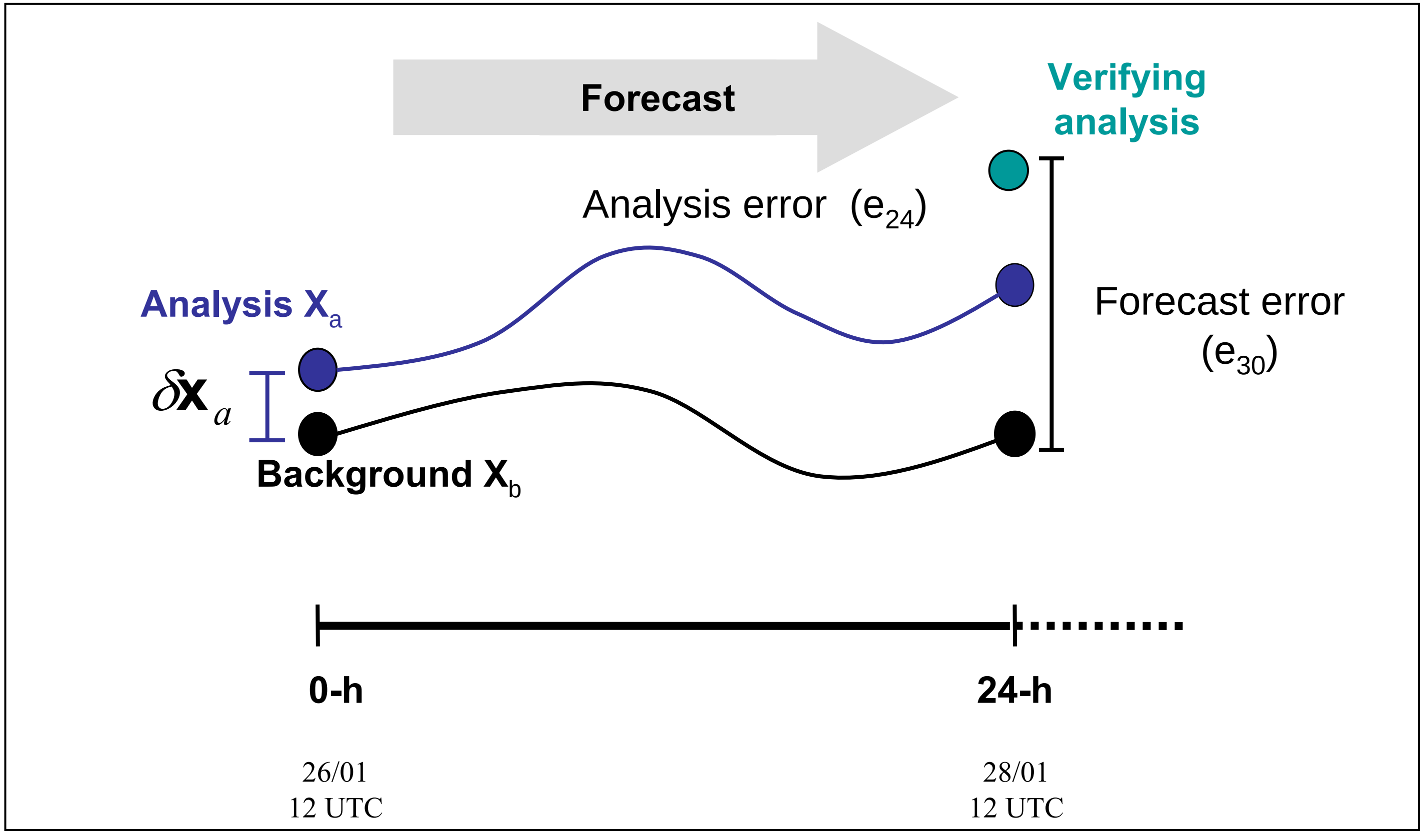
(Langland and Baker, 2004)



Observations move the model state from the "**background**" trajectory to the new "**analysis**" trajectory

The difference in forecast error norms, $e_{24} - e_{30}$, is due to the combined impact of all observations assimilated at 00UTC

Observability of flow dependent structure functions



$$\Delta e_a^b = e_{30} - e_{24} = \delta \mathbf{x}_a^T \left(\mathbf{L}_a^T \frac{\partial J_a}{\partial \mathbf{x}_a} + \mathbf{L}_b^T \frac{\partial J_b}{\partial \mathbf{x}_b} \right) = \left(\mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}_b \right) \right)^T \left(\mathbf{L}_a^T \frac{\partial J_a}{\partial \mathbf{x}_a} + \mathbf{L}_b^T \frac{\partial J_b}{\partial \mathbf{x}_b} \right)$$

Evaluation of the impact of observations

At initial time

$$\Delta e_a^b = [\mathbf{y} - \mathbf{H}(\mathbf{x}_b)]^T \left[\mathbf{K}^T \left(\frac{\partial J_a}{\partial \mathbf{x}_a} + \frac{\partial J_b}{\partial \mathbf{x}_b} \right)_{t=t_0} \right]$$

where $\mathbf{y} - \mathbf{H}(\mathbf{x}_b)$: observation departure from the background state

Evaluation of the impact of observations

At initial time

$$\Delta e_a^b = [\mathbf{y} - \mathbf{H}(\mathbf{x}_b)]^T \left[\mathbf{K}^T \left(\mathbf{L}_a^T \frac{\partial J_a}{\partial \mathbf{x}_a} + \mathbf{L}_b^T \frac{\partial J_b}{\partial \mathbf{x}_b} \right) \right] \leftarrow \left(\frac{\partial J}{\partial \mathbf{x}_a} + \frac{\partial J}{\partial \mathbf{x}_b} \right)_{t=t_0}$$

where $\mathbf{y} - \mathbf{H}(\mathbf{x}_b)$: observation departure from the background state

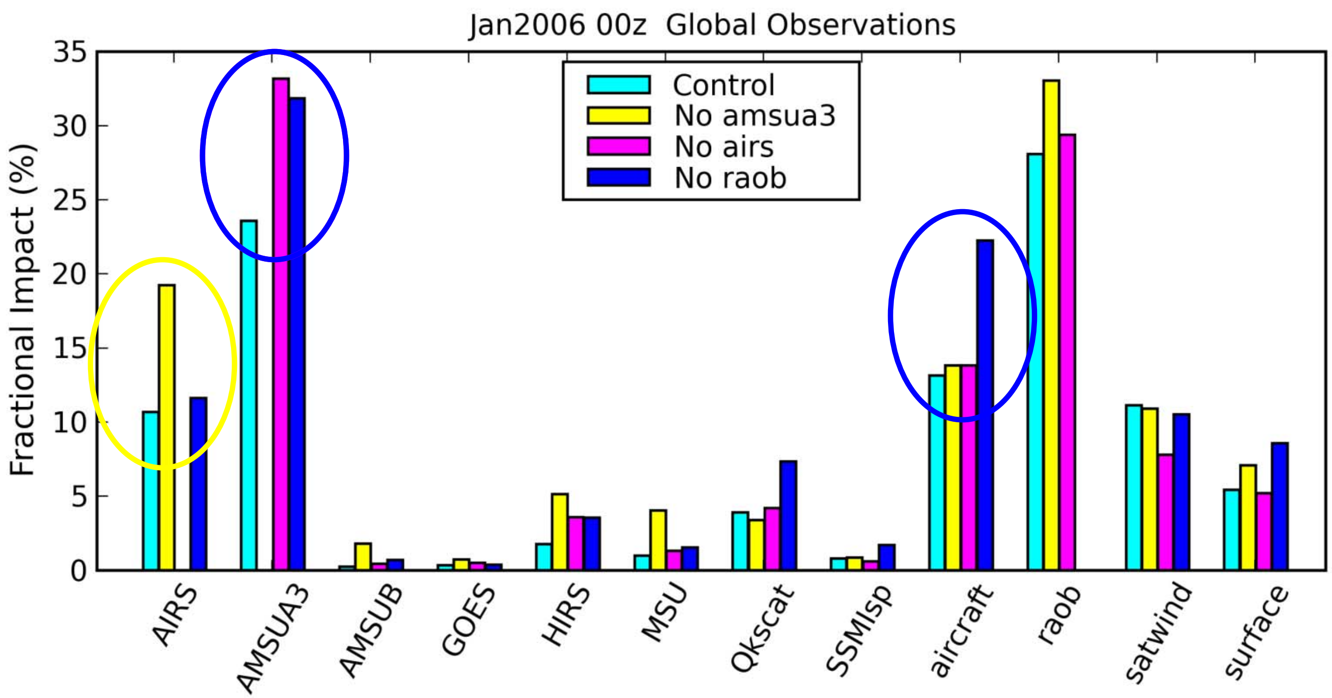
Computation of the Observation Impact: $\mathbf{K}^T \left(\frac{\partial J_a}{\partial \mathbf{x}_a} + \frac{\partial J_b}{\partial \mathbf{x}_b} \right)_{t=t_0} = \mathbf{R}^{-1}\mathbf{H}\mathbf{w}$

One can obtain **w** by slightly adapting the assimilation to solve

$$F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{B}^{-1}\mathbf{w} + \frac{1}{2}(\mathbf{H}\mathbf{w})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{w}) - \mathbf{w}^T\left(\frac{\partial J}{\partial \mathbf{x}_a} + \frac{\partial J}{\partial \mathbf{x}_b} \right)_{t=t_0}$$

Combined Use of ADJ and OSEs (Gelaro *et al.*, 2008)

…ADJ applied to *various* OSE members to examine how the mix of observations influences their impacts

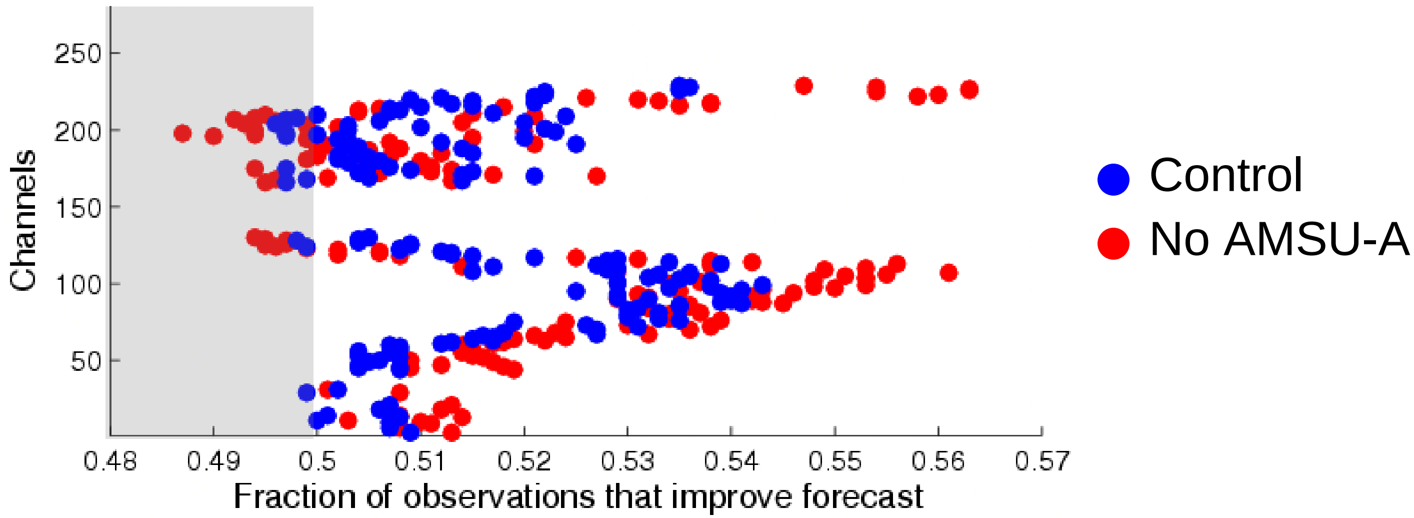


- Removal of AMSUA results in large increase in AIRS (and other) impacts
- Removal of AIRS results in significant increase in AMSUA impact
- Removal of raobs results in significant increase in AMSUA, aircraft and other impacts (but not AIRS)

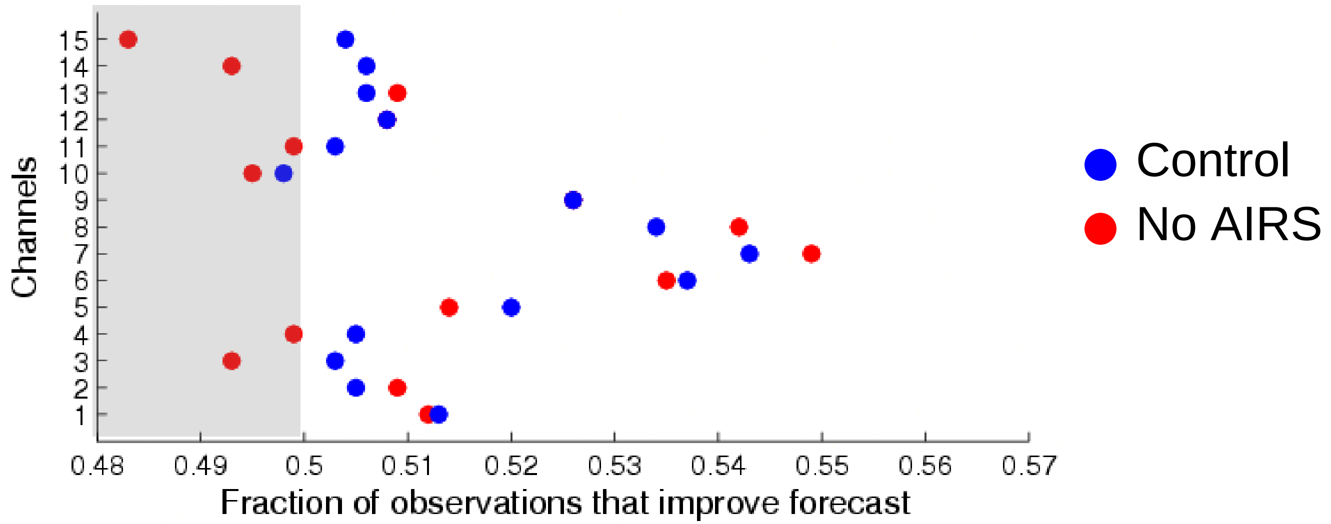
Fraction of Observations that Improve the Forecast

GEOS-5 July 2005 00z (Gelaro, 2008)

AIRS



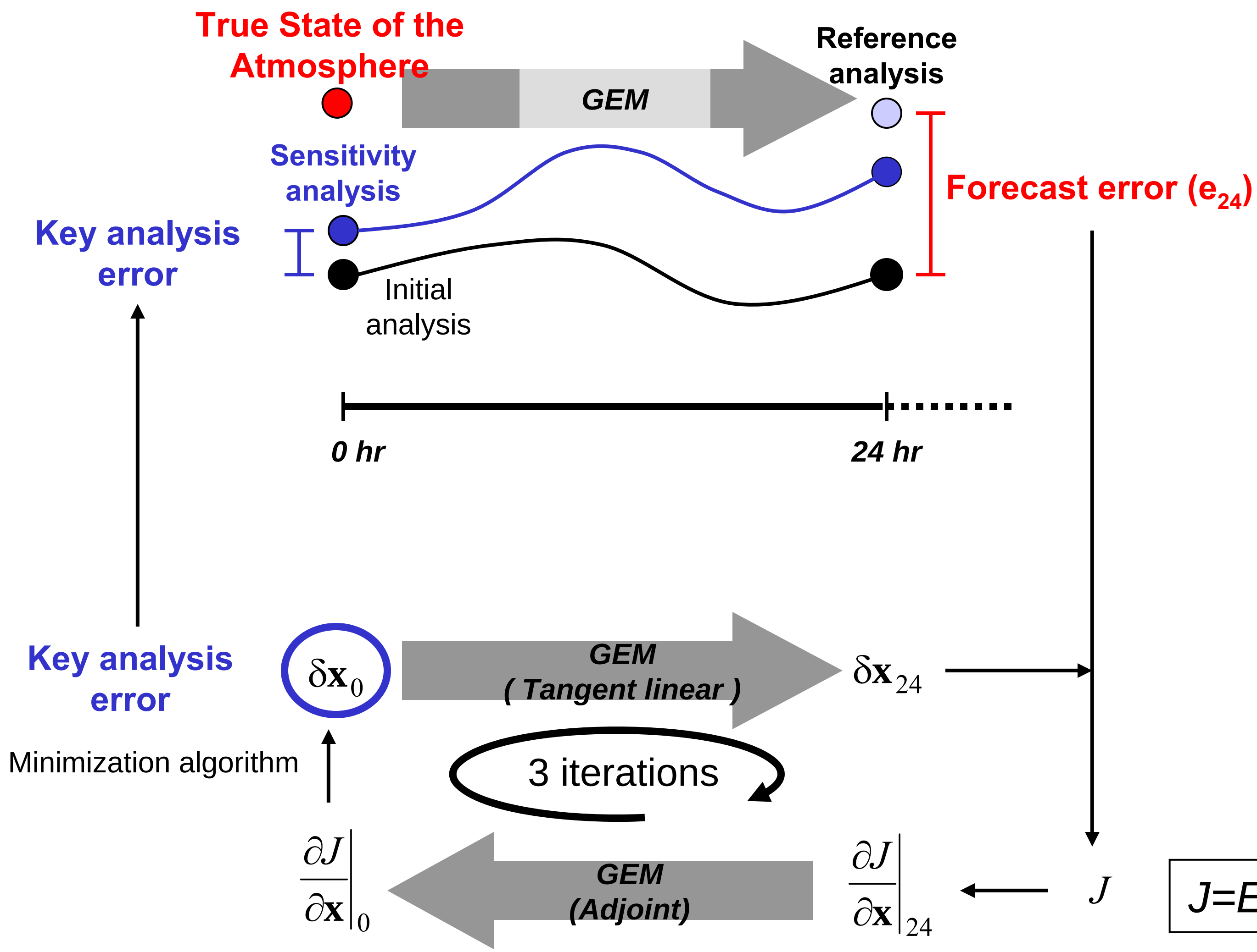
AMSU-A



…only a small majority of the observations improve the forecast

Key analysis errors algorithm – configuration

(Laroche *et al.*, 2002)



J=Energy of $(\delta \mathbf{x}_{24} - e_{24})$

Modelling background-error covariances using sensitivities

▪The adapted 3D-Var

- Structure functions defined with respect to a posteriori sensitivities;
- Flow dependent structure functions were introduced in the 3D-Var;

$$\widetilde{\mathbf{B}}_{\boldsymbol{\xi}} = \mathbf{I} + \widetilde{\mathbf{B}} = \mathbf{I} + \lambda_{\mathbf{1}}^{\mathbf{2}} \tilde{\mathrm{f}} \tilde{\mathrm{f}}^{\mathrm{T}}$$

- Error variance along f: $$1 + \lambda_1^2 = \sigma_1^2$$

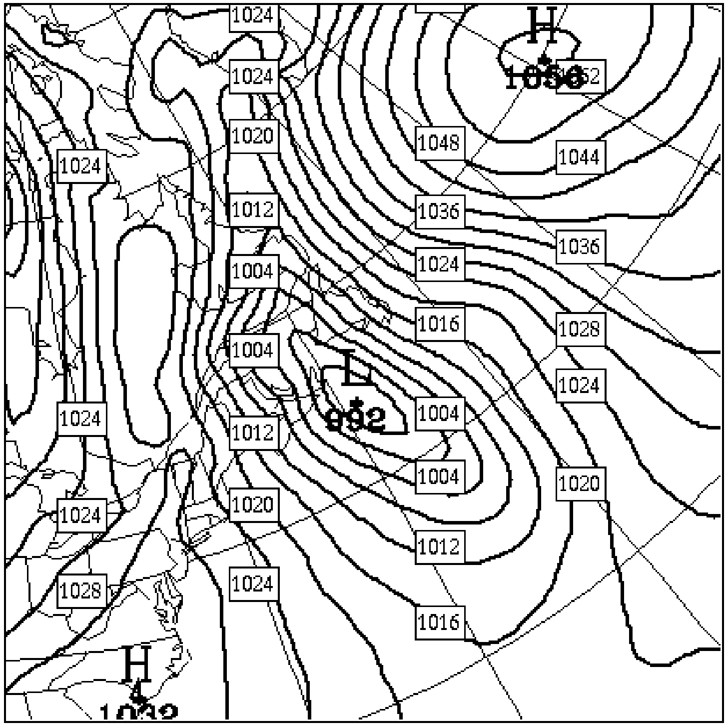
Does a flow-dependent background error formulation improve the analysis and subsequent forecast?

(Lupu 2006)

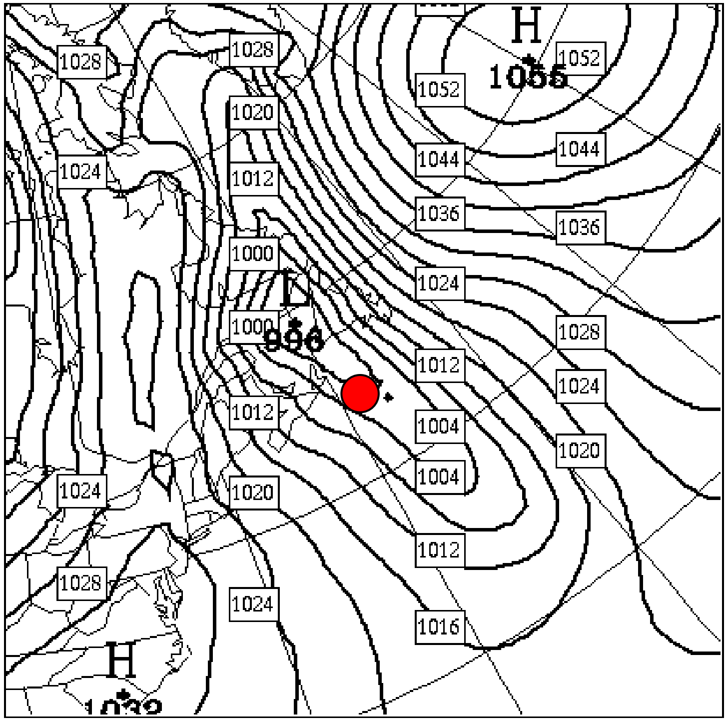
Case study of January 27, 2003

Forecast verification, 12 UTC January 28, 2003

CMC analysis



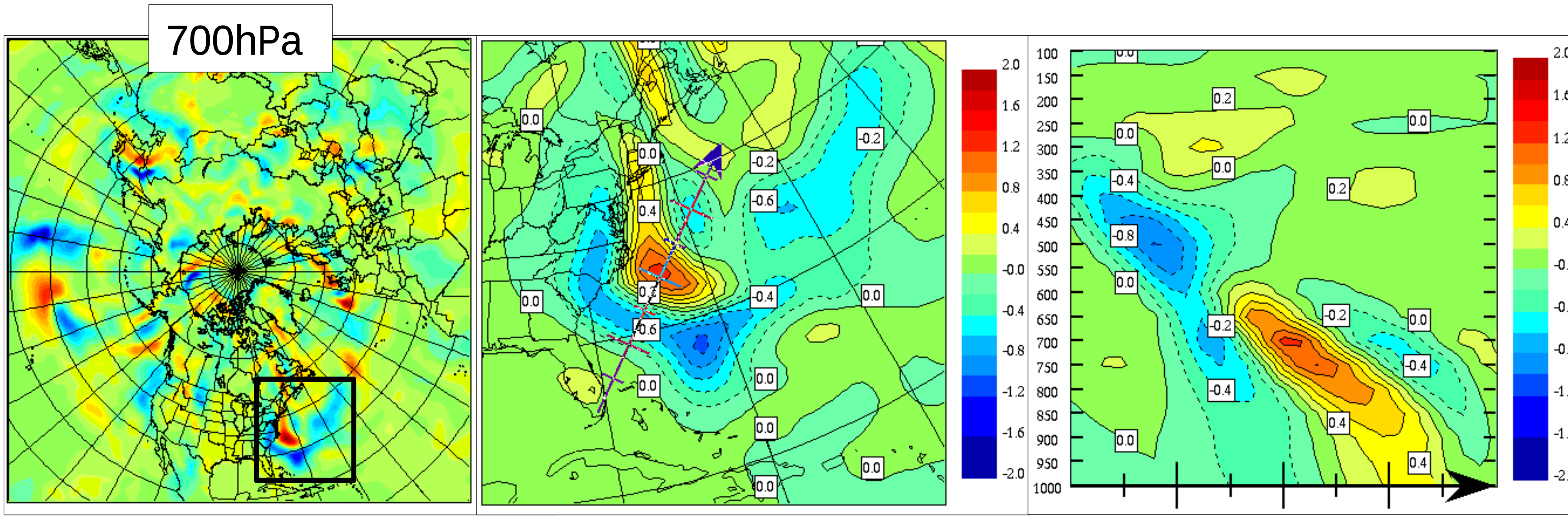
Global-GEM 24hr operational forecast



Sea Level Pressure (4 hPa)

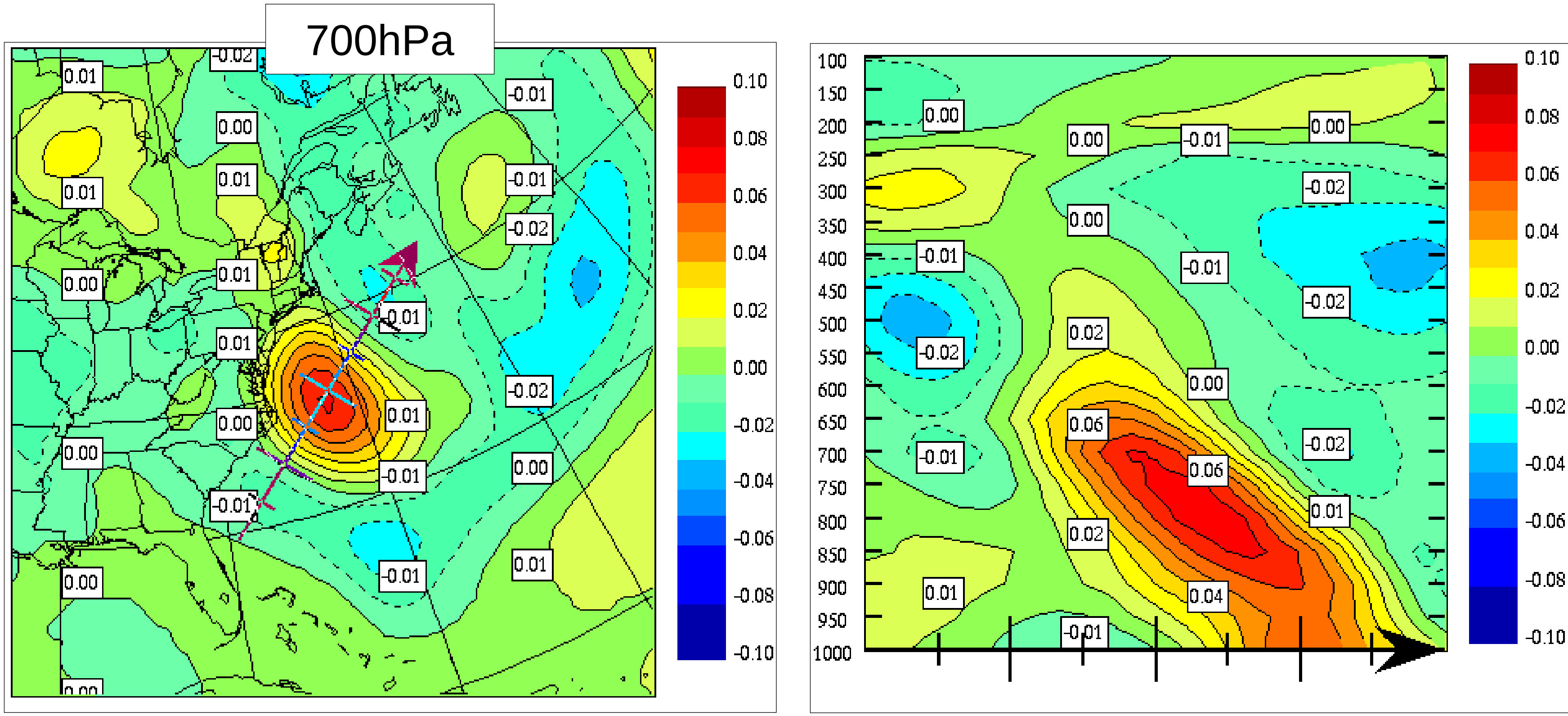
Case study – Global sensitivity function

Initial temperature corrections for the 12 UTC January 27, 2003 analysis



Corrections responsible for the forecast improvement of the Canadian Maritimes system and cross section of initial temperature correction made along the arrow.

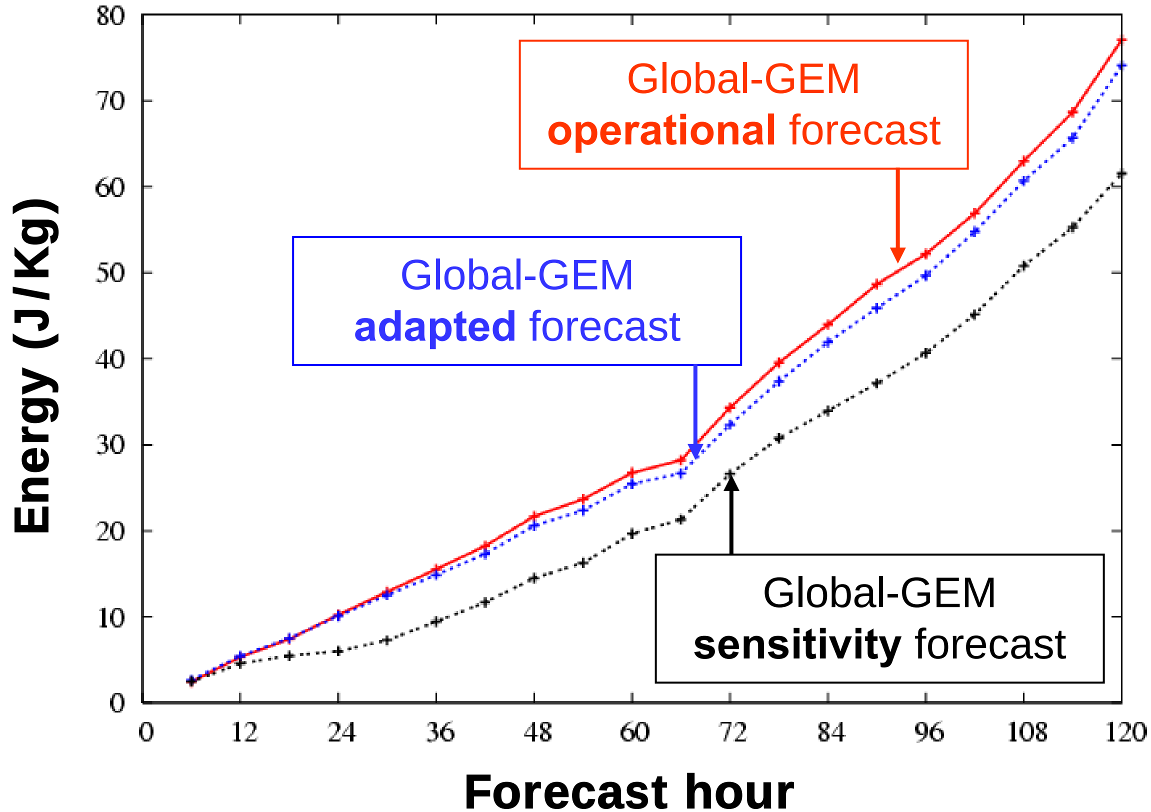
Impact of the adapted 3D-Var in the analysis



Difference between the temperature analysis increments for 12 UTC January 27, 2003 analysis 3D adapted -3D standard and cross section.

Case study –Forecast improvement

Energy (total) of the forecast error average over Northern Hemisphere Extra-tropics (25N - 90N)

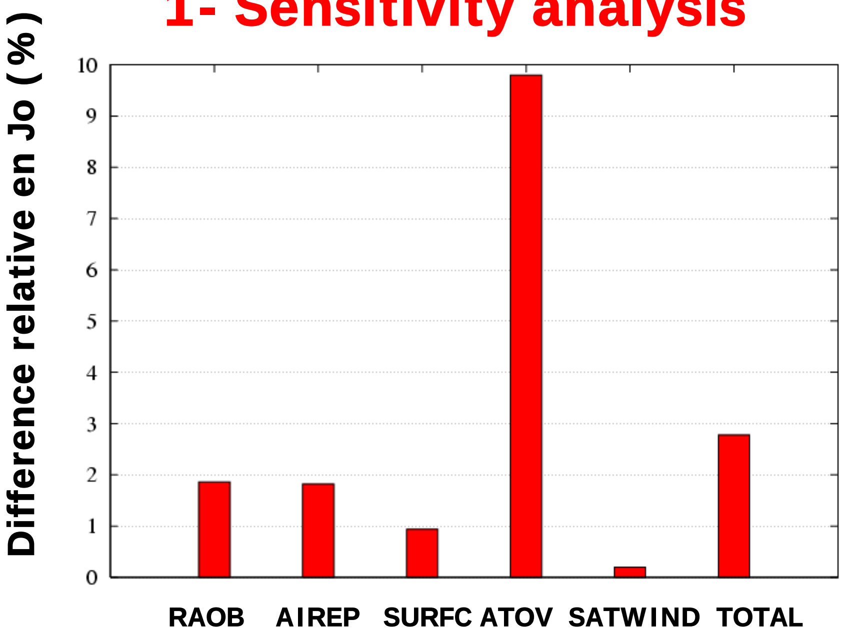


Fit to the observational Data

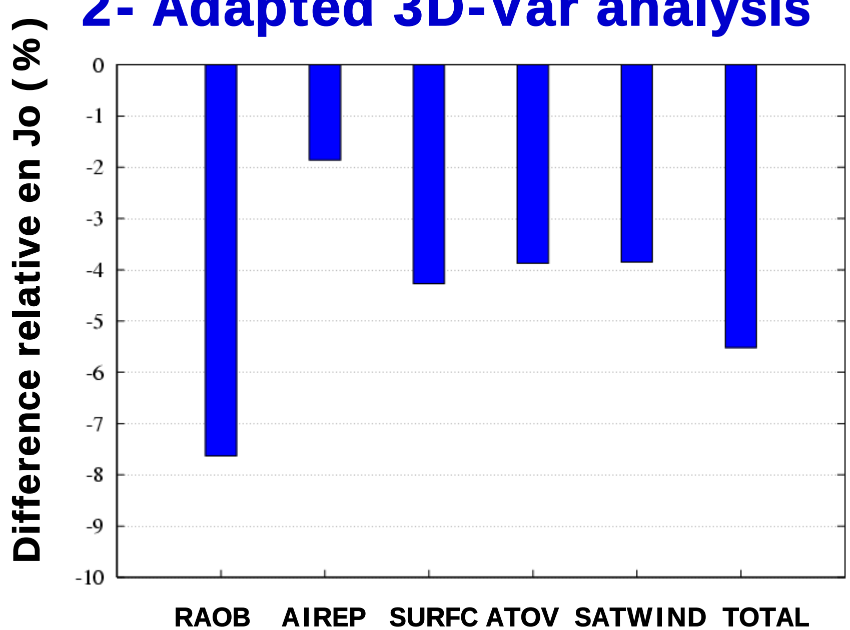
Do the corrections **decrease** or **increase** the departure between **the analysis** and **the observations** ?

$$\Delta J_o = \frac{J_o(\mathbf{x}^{1,2}) - J_o(\mathbf{x}^{3D-Var})}{J_o(\mathbf{x}^{3D-Var})} \begin{cases} > 0 = \textbf{increase} \\ < 0 = \textbf{decrease} \end{cases}$$

1- Sensitivity analysis



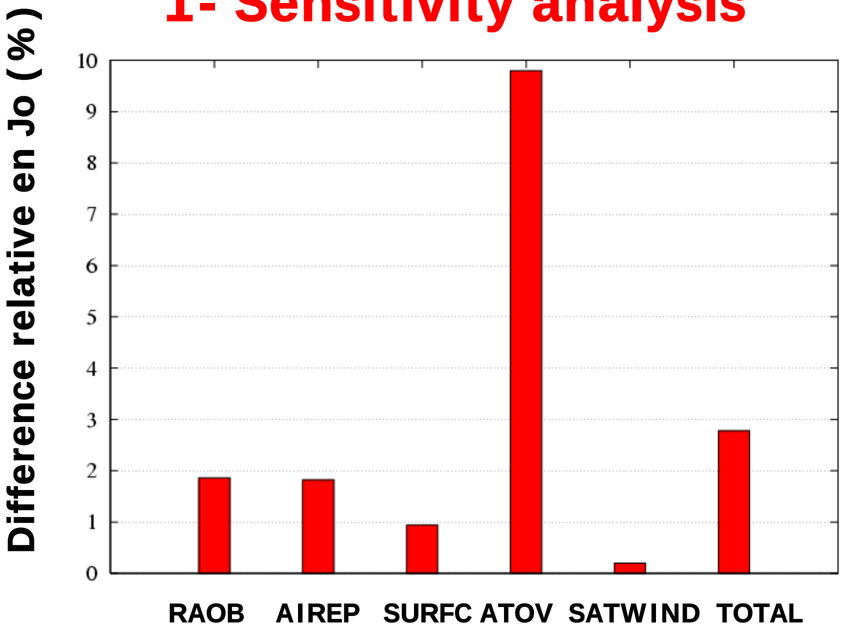
2- Adapted 3D-Var analysis



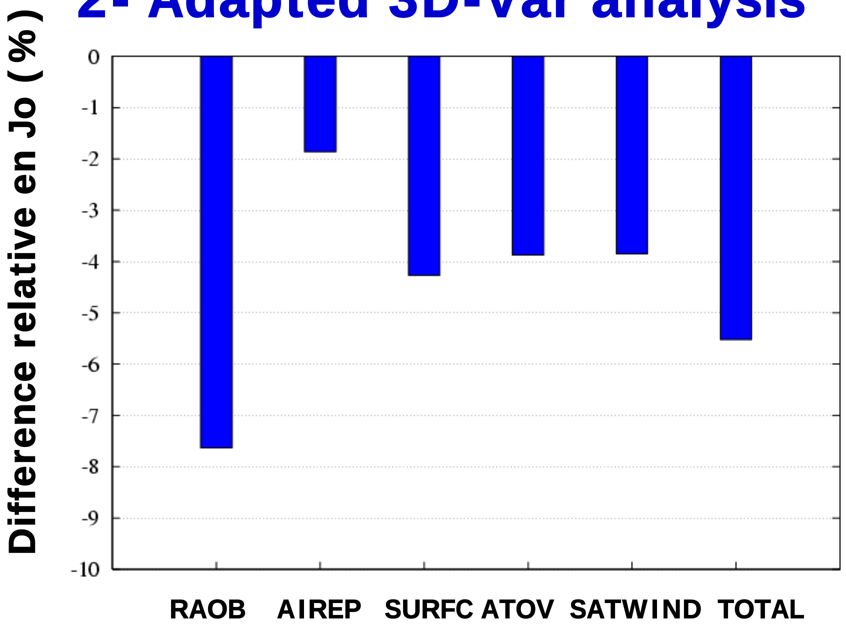
Fit to the observational Data

- Positive values mean that the sensitivity analysis is further away from the obs. than the initial analysis (same conclusions from ECMWF, Isaksen et al., 2004);
- Negative values mean that the adapted 3D-Var analysis is closer to the obs. (due to the increase background-error variance);

1- Sensitivity analysis



2- Adapted 3D-Var analysis



Observability of flow-dependent structures

- **Adapted 3D-Var for which the structure functions where defined by normalizing the *a posteriori* sensitivity function**
- **Consider the case where $\mathbf{B} = \sigma^2 \mathbf{v}\mathbf{v}^T$ and the analysis increment is then**

$$\delta \mathbf{x}_a = \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) = \mathbf{K}\mathbf{d} = \alpha \mathbf{v}$$

with

$$\alpha = \frac{(\mathbf{H}\mathbf{v})^T \mathbf{R}^{-1} \mathbf{d}}{\sigma^{-2} + (\mathbf{H}\mathbf{v})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{v})} = \frac{\sigma^2 C_1}{1 + \sigma^2 C_2}$$

and

$$C_1 = (\mathbf{H}\mathbf{v})^T \mathbf{R}^{-1} \mathbf{d} \qquad C_2 = (\mathbf{H}\mathbf{v})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{v})$$

Associated information content and observability

- **Evaluation of the DFS in this case**

$$DFS = tr(\mathbf{HK}) = \frac{\sigma^2 C_2}{1+\sigma^2 C_2}$$

$$\lim_{\sigma \to \infty} DFS = 1$$

- **Correlation between the innovations and a structure function**

$$\rho = \frac{(\mathbf{Hv})^T \mathbf{R}^{-1} \mathbf{d}}{\left[(\mathbf{Hv})^T \mathbf{R}^{-1} (\mathbf{Hv})\right]^{1/2} \left[\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}\right]^{1/2}} = \frac{C_1}{(2C_2 J_o(0))^{1/2}}$$

- **This defines the observability of a structure functions**

  * Can the observations detect a given structure function

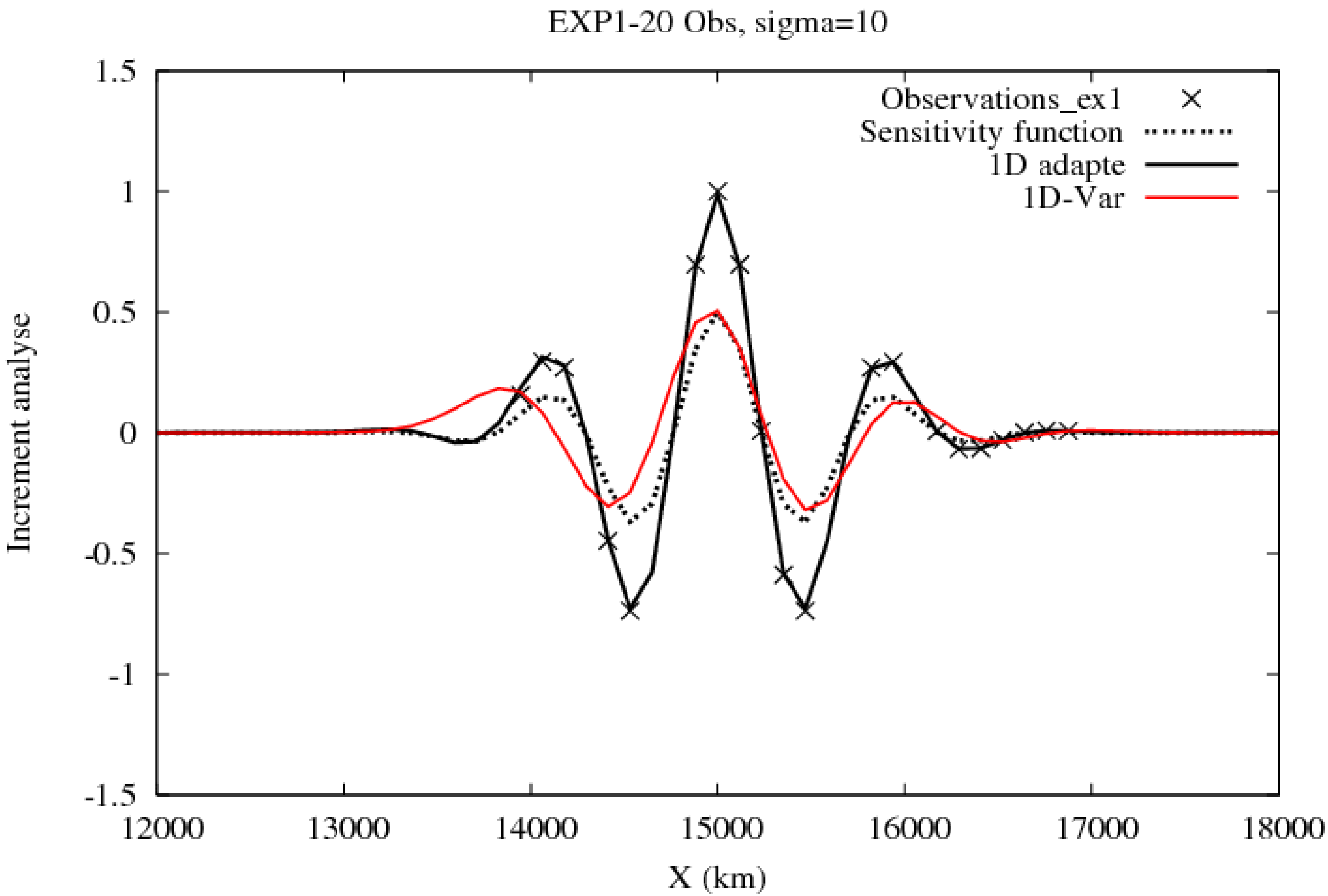
Example from 1D-Var experiments

- **Consider the following cases**
    - * Observations are generated from the same structure function as that used in the assimilation
    - * Observations are generated from a different structure function (phase shift)
    - * Signal has an amplitude lower than the level of observation error

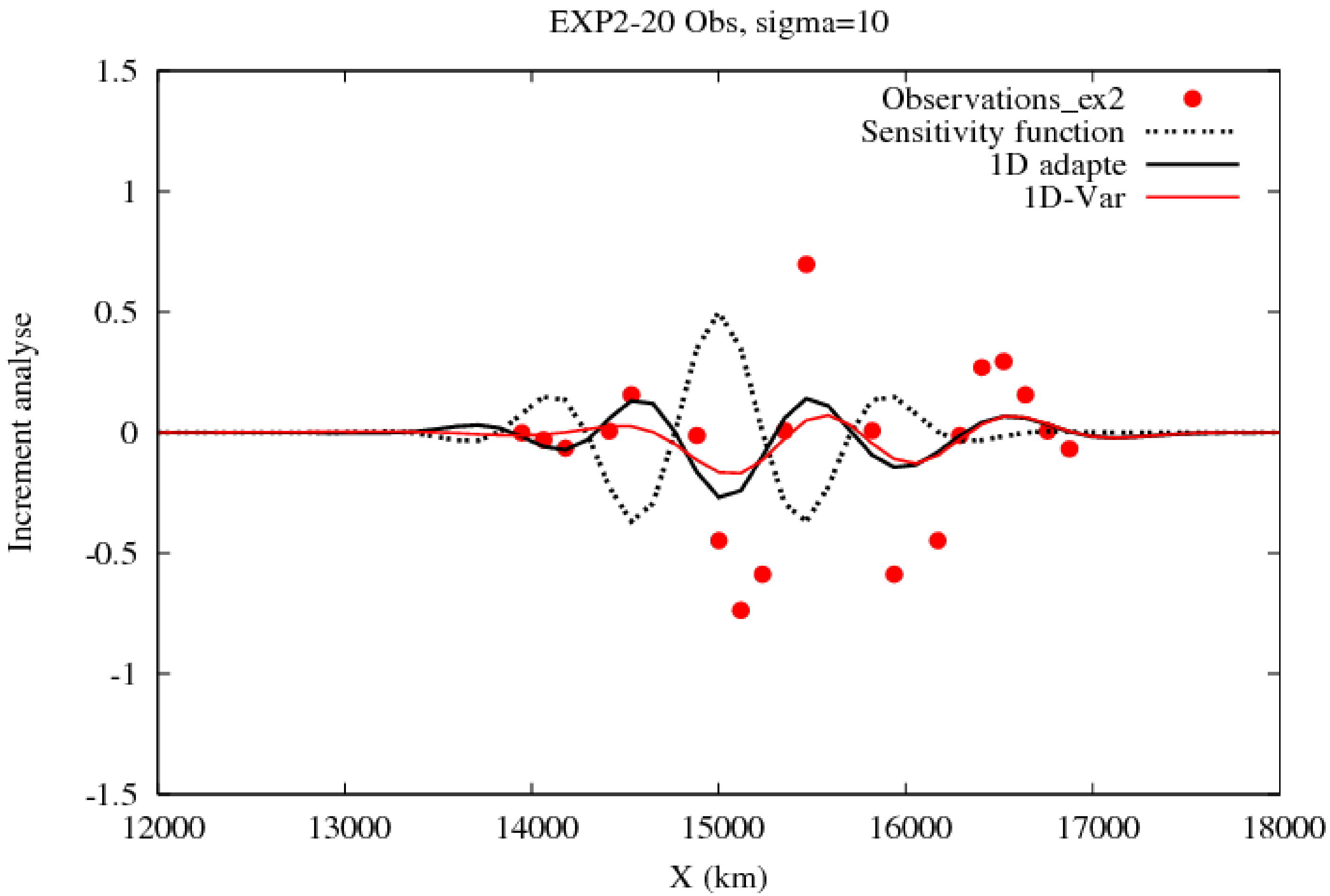
Observability as a function of observation error

	Nb obs.	$\mathbf{C_1}$	$\mathbf{C_2}$	ρ
$\mathbf{y}' = 2(\mathbf{Hv})$	10 obs.	1.29	0.64	0.99
	20 obs.	1.96	0.97	0.99
	40 obs.	2.26	1.13	1.
$\mathbf{y}' = 2(\mathbf{Hv}) + \varepsilon_o$	10 obs.	0.95	0.64	0.38
	20 obs.	1.15	0.97	0.22
σ_o^2 =1	40 obs.	1.48	1.13	0.20
$\mathbf{y}' = 2(\mathbf{Hv}) + \varepsilon_o$	10 obs.	0.89	0.64	0.17
	20 obs.	0.89	0.97	0.11
σ_o^2 =4	40 obs.	0.87	1.13	0.08

Experiment with the same function



Experiment with a shifted function



Experiments with an adapted 3D-Var

- ***A posteriori* sensitivities depend on**
  * Target area
  * Norm used to measure the forecast error
  * Initial norm
  * Definition of the tangent-linear and adjoint model
- **Experiments with an adapted 3D-Var based on EC's 3D-Var assimilation**
  * Dry energy norm
  * Four cases documented in Caron *et al. (2007): January 19, 2002, 00UTC, Feburary 6, 2002, 00UTC January 6, 2003 12UTC; January 27, 2003 12UTC*
  * Target area: global, hemispheric (25-90N) and local (area on the East Coast of North America)
  * Imposition of a nonlinear balance constraint (Caron *et al*., 2007)

Preliminary test: does it work?

- **Normalized analysis increment of a 3D-Var as a structure function**
  - * Limiting case where $\mathbf{B} = \sigma^2 \mathbf{v}\mathbf{v}^T$
  - * Does the adapted 3D-Var recover the right amplitude
  - * This particular choice insures that we have a structure that can fit the observations.

Observability for the test case

Obs. type	Correlation coefficient ρ			
	January 27, 2003	January 06, 2003	February 06, 2002	January 19, 2002
RAOB	0.73	0.76	0.77	0.76
AIREP	0.73	0.73	0.73	0.72
AMV	0.68	0.72	0.72	0.73
SURFC	0.69	0.74	0.75	0.76
ATOVS	0.59	0.58	0.71	0.65
TOTAL	0.71	0.73	0.75	0.74

Observability of different structure functions based on key analyses

Structure functions	Obs. type	ρ, correlation coefficient			
		January 27, 2003	January 06, 2003	February 06, 2002	January 19, 2002
GLOBAL	RAOB	0.01	0.02	0.03	-0.01
	AIREP	0.00	0.02	-0.01	-0.01
	ATOVS	0.13	0.11	0.07	0.12
	TOTAL	**0.05**	**0.05**	**0.05**	**0.03**
LOCAL	RAOB	-0.01	0	-0.01	-0.02
	AIREP	-0.03	-0.01	-0.03	-0.03
	ATOVS	0.05	0.01	0.06	0.02
	TOTAL	0	0	0	-0.01
HEMISPHERIC	RAOB	0.00	0.02	0.01	0.01
	AIREP	-0.05	0.02	-0.02	-0.03
	ATOVS	0.08	0.07	0.07	0.04
	TOTAL	**0.03**	**0.04**	**0.04**	**0.02**
PV-BAL	RAOB	0.01	0	0.01	0
	AIREP	-0.03	0.01	-0.03	0
	ATOVS	0.09	0.08	0.08	0.05
	TOTAL	**0.03**	**-0.01**	**0.06**	**0.02**

Observability of a pseudo-inverse obtained from a finite number of singular vectors (Mahidjiba et al., 2007)

- **Leading singular vectors are the structures that will grow the most rapidly over a finite period of time**
  - Leading 60 SVs were computed based on a total dry energy norm at a lead time of 48-h
  - The forecast error is projected onto those SVs at the final time which allows to express the error at initial time that explains that forecast error (pseudo-inverse)
- **Experiments**
  - 18 cases were considered in December 2007
  - Are those structures observable from available observations?
  - Observability of SV_1, the leading singular vectors
  - Observability of the pseudo-inverse

Observability of the leading singular vector and pseudo-inverse

Date	Obs. type	Correlation coefficient ρ		
		SV no. 1 Initial time	SV no. 1 Final time	Pseudo-inverse
2007120100	TOTAL	0.0098	0.0067	0.0169
2007120212	TOTAL	0.0140	-0.0179	-0.0011
2007120400	TOTAL	-0.0187	-0.0211	-0.0034
2007120512	TOTAL	0.0022	-0.0020	0.0124
2007120700	TOTAL	0.0159	0.0020	-0.0033
2007120812	TOTAL	0.0019	0.0212	0.0062
2007121000	TOTAL	-0.0029	-0.0151	0.0040
2007121112	TOTAL	0.0054	0.0148	0.0096
2007121300	TOTAL	0.0125	-0.0241	-0.0028
2007121412	TOTAL	0.0224	-0.056	0.0209
2007121600	TOTAL	0.0125	0.0235	0.0234
2007121712	TOTAL	0.0041	0.0465	-0.0064
2007121900	TOTAL	0.0119	-0.0097	-0.0010
2007122012	TOTAL	0.0067	0.0217	0.0047
2007122200	TOTAL	0.0103	-0.0084	-0.0053
2007122312	TOTAL	0.0099	-0.0068	0.0110
2007122500	TOTAL	-0.0020	-0.0065	-0.0059
2007122612	TOTAL	-0.0086	0.0056	-0.0117

Summary and conclusions

- **Evaluation of the information content of observations can be obtained from simple diagnostics using information generated by any assimilation system**
  - Impact of observations depends on the observing environment
  - Offer a measure of the consistency between the statistics used in the assimilation and those diagnosed through comparison to observations
- **Impact of observations on forecasts can be quantified as well, based on a method proposed by Langland and Baker (2004)**
  - Measurement based on a backward integration of the adjoint model
  - Same "ingredients" that are used to compute key analyses to pinpoint the source of the forecast error
  - Observation impact is defined with respect to their correlation with respect to that particular structure
  - Our results may explain in part why only half the observations have a positive impacts

Conclusion (cont'd)

- **Observability of structure functions has been defined in observation space as a correlation between innovations and the structure function**
- **Even though those structures do correspond to structure that will grow the most or grow to correct the forecast error at a given lead time**
    - *A posteriori* sensitivities are not well correlated with observations
        - → *This has been tested for different ways to compute the sensitivities*
    - Singular vectors were not found to be observable either
- **Reduced rank Kalman filters do not seem to be appropriate to represent the background error covariances in an assimilation system**
- **Evolved covariances as estimated with an Ensemble Kalman filter would be more appropriate for an hybrid 4D-Var assimilation**