Observation processing in a LAM operational data assimilation system

Roger Randriamampianina
Hungarian Meteorological Service

Summer school on nonhydrostatic dynamics and fine scale data assimilation,
St. Petersburg, 10-17 June 2006

NWP ⟹ initial/boundary value problem

Given

➔ an estimate of the present state of the atmosphere (initial conditions)

➔ appropriate surface and lateral boundary conditions the model simulates or forecasts the evolution of the atmosphere.

➔ The more accurate the estimate of the initial conditions, the better the quality of the forecasts.

In case of a LAM, initial conditions can be provided by an operational global model

OR,

we can produce initial conditions through a statistical combination of observations and short-range forecasts.

➔ This approach is called data assimilation

The advantage of such a system is that it takes into account as many local observations as possible

- If the model state is overdetermined by the observations, the analysis is reduced to an interpolation problem

- But observations are inhomogeneous in space and time, and some are indirectly related to the model variables in many geographical areas
  $\Rightarrow$ problem is well-posed, only if some background information is available

- In some areas, observations are too dense $\Rightarrow$ one must thin them

- Observations are not perfect $\Rightarrow$ we have to deal with observation errors
  $\Rightarrow$ screening, and bias correction for several types of observations
  (e.g. satellite observations)

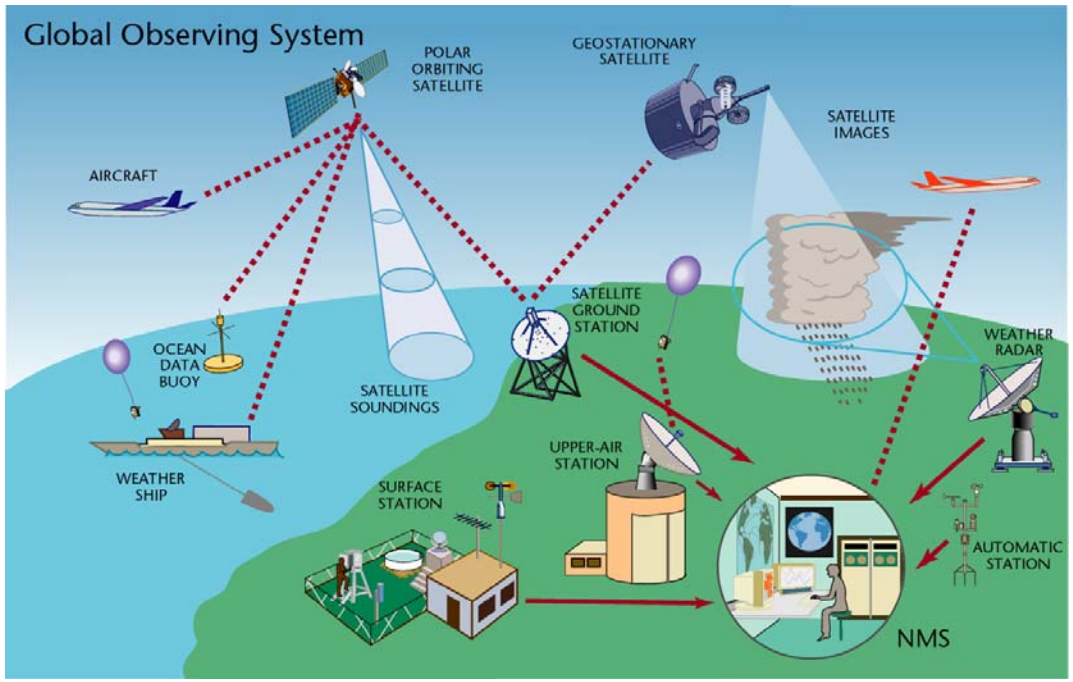
Outline

Need of control and monitoring of the observations

Observation pre-processing

LAM DA system scheduling

Need of DA system monitoring



Source: www.wmo.int

Real time available data:

Networks

Observations arrive through

- **GTS (Global Telecommunication System) in ASCII or BUFR format**
- **Direct Sat. Reception – (HRPT Station)**
- **Special TCP/IP lines**
- **Internet FTP in some cases**

Providers

- **All WMO countries for most of the observation types**
- **Space agencies for Satellite observations (Eumetsat, ESA, NESDIS, NASA, DMSP)**
- **NMS for the local national data**

Types of observations

Conventional Observations

→ Surface data:
- Synop, Ship
- Bathy, Tesac
- Buoy

→ Upperair data:
- Airep, Amdar, Acar
- Temp, Temp-ship, Temp-mobil, Temp-drop
- Pilot, Pilot-ship, Europrofil, Profiler
- Satob, Satgeo, geowind

Satellite Observations

→Meteosat:
- SEVIRI
- Clear sky radiances

→ NOAA Atovs:
- Hirs, Amsua, Amsub, Mhs

→ DMSP:
- Ssm/I

→ AQUA:
-Airs

- The quality of the forecasts depends on the quality of the observations (data) used in the data assimilation system
- The data are not perfect : they contain errors of different kinds and sources
- The control and the monitoring of the data is important to eliminate the wrong ones and determine the list of stations not to be used : blacklist (or we can have a whitelist)
- Importance of the feedback to the producers
- Identification of local problems during the test of new observations …
- Exchange of results with other centers : the WMO monitoring of the observations is done by several national centers (for example, ECMWF is the lead center for monitoring the quality of sonde data) $\Rightarrow$ monthly control reports, SAF NWP etc …

General Principle of the data monitoring

- Use of daily and monthly statistics of the observations against a reference : a 6-hour forecast (called guess, or background)

- This reference is interpolated at the obs point. We can define in the observation space :

  O-G = (O – Otrue) – (G – Gtrue) – (Gtrue – Otrue)

  obs bias model bias scale bias

  the observation errors and the model errors are supposed to be not correlated

- An observation is said to be suspicious when its statistics (compared to the model) is too far away from that of the other observations

Sources of model and observations errors

In the model (guess or background) :

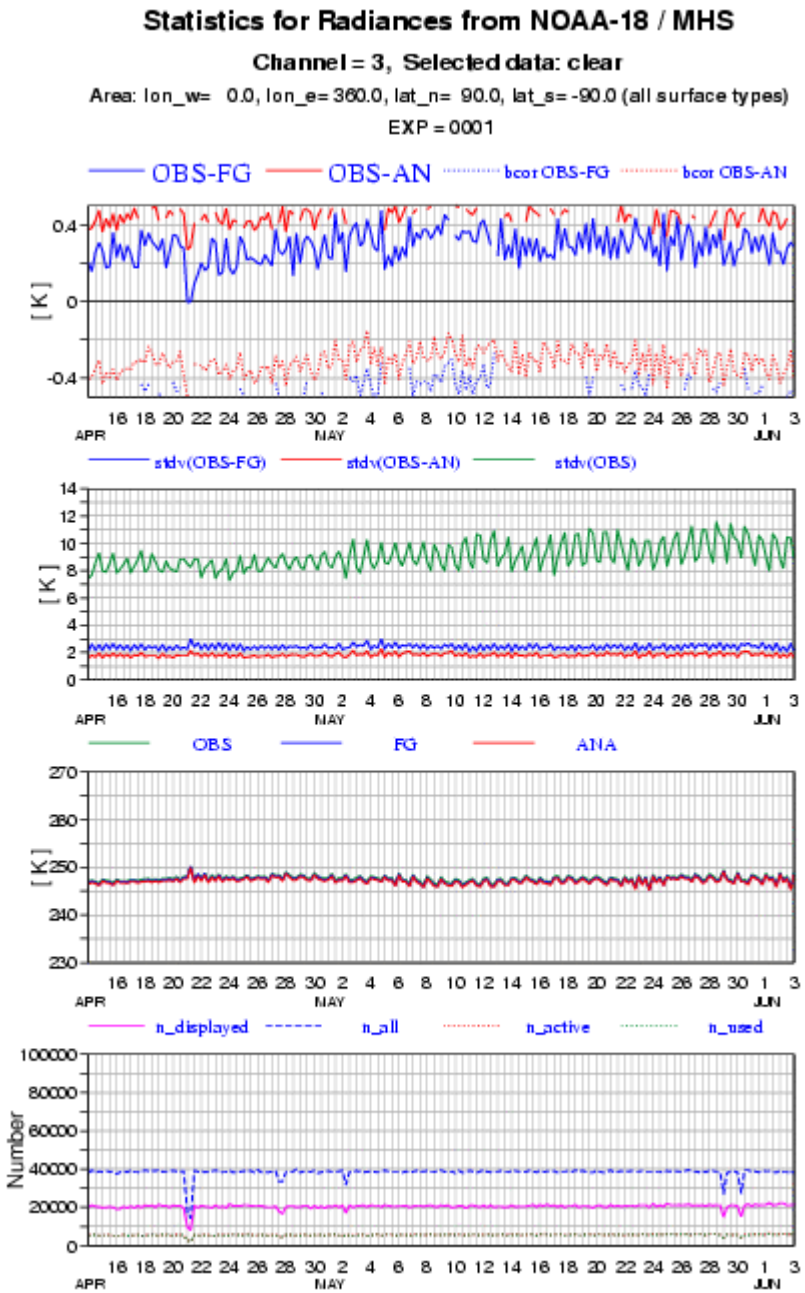
- Systematic errors of prediction
- Lack of resolution
- Weakness in the description of the physical processes
- Errors in the starting point (analysis with not enough obs)
- Amplification of errors (hydrodynamical instabilities)

In the observations :

- Instrumental errors (little in most cases). These informations are exchanged between the centers involved in the monitoring of the observations.
- Errors in the coding / decoding pocedures
- Representativity / resolution of the model
- Observation operator error (when calculating the values of the model at obs points)

Blacklisting the observations

- Knowledge of the mean behaviour (in term of statistics of obs-guess) of the observations
- Defining thresholds of large errors (WMO thesholds for example)
- Assessing each platform/station taking into account model and representativity problems
- Looking at the stability in time of the errors
- Looking at the other observations available in the same region
- Putting the observations in the blacklist file and removing them from there when their quality becomes better



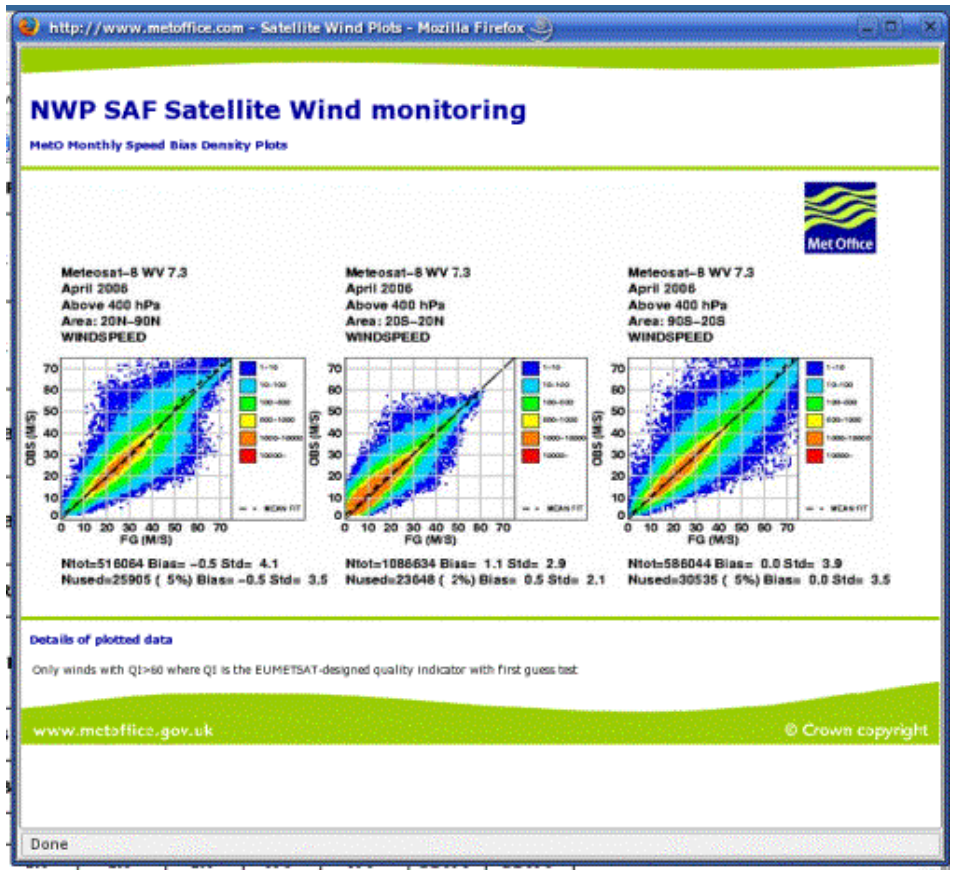
Obs. Monitoring on the SAF NWP page

← ECMWF

and

MetOffice

↓

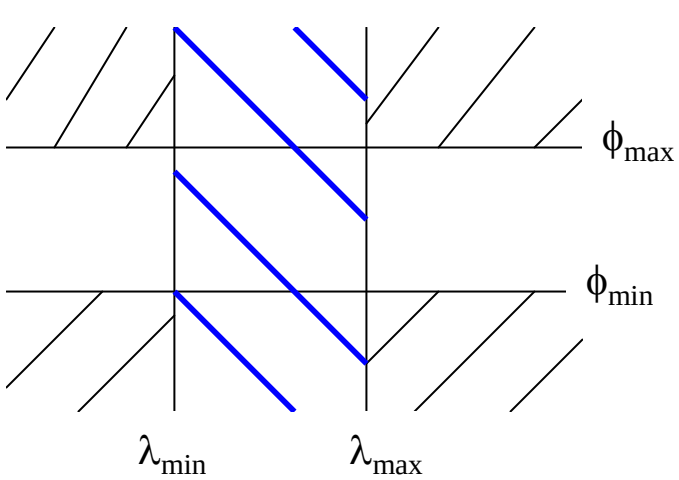


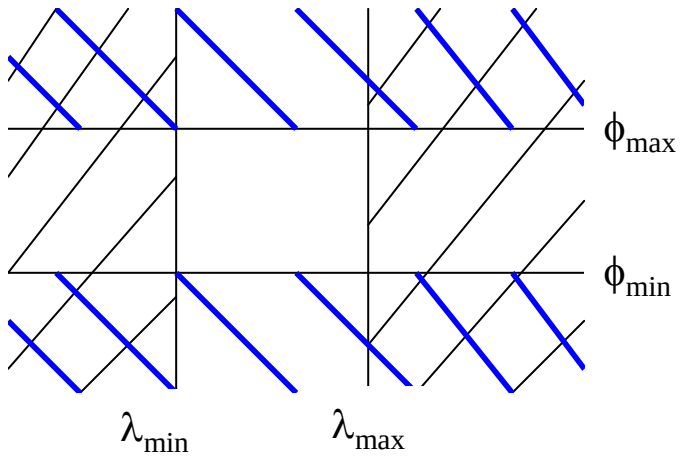
The blacklisting

Blacklist can be limited:

→ to a concrete parameter or group of parameters from a certain platform/station

→ to different geographical zones (Geo. sat. data)

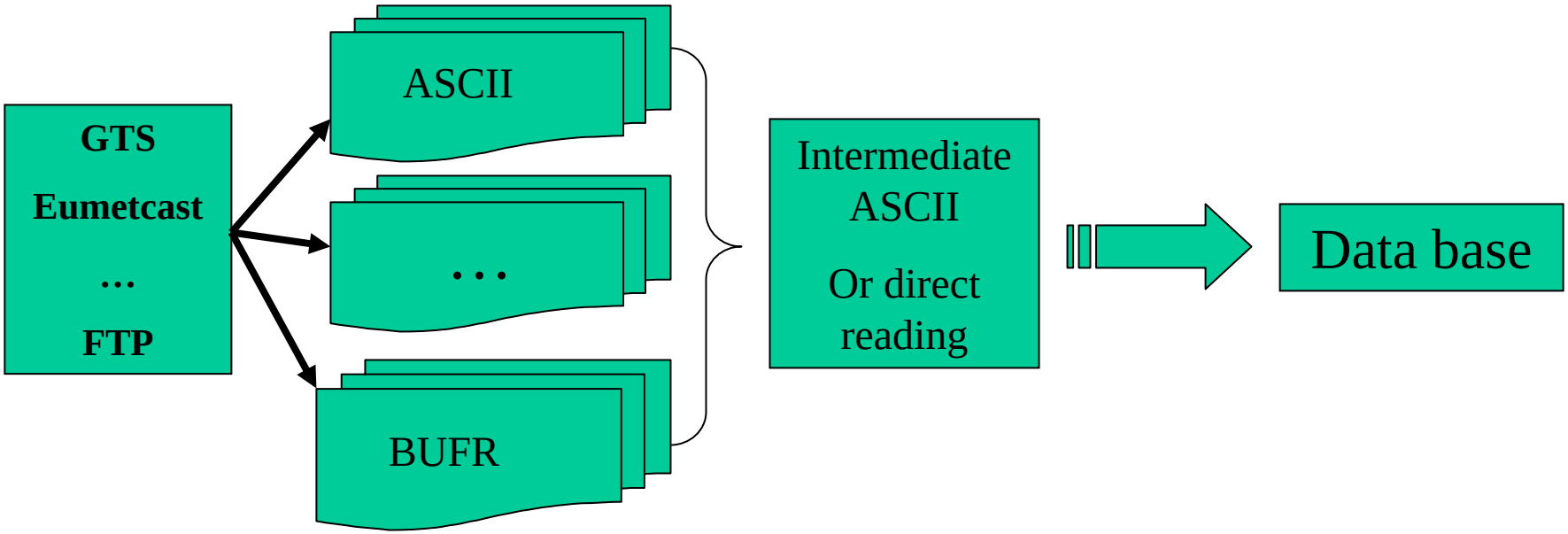




→ to different channels of satellite instruments

→ to different pressure thicknesses (ex. applied for AMV and wind profiler data)

Ex. Pre-processing of data at HMS



Pre-processing chain:

Reading of data and coding of additional information

Blacklisting and initialization of the observation errors

Complex quality check and data selection

Observation usage (in surface analysis)

Message type	Observation type	Extracted parameters	Used parameters	Parameter blacklisted or not used
SYNOP SYNOP-SHIP (TEMP TEMP-SHIP TEMP-MOBIL TEMP-DROP)	Surface observation (over land and sea)	• P_{mer} or P_{st} • u_{10m} & v_{10m} computed from dd_{10m} & ff_{10m} • T_{2m} • *SST* • Hu_{2m} computed from T_{2m} and Td_{2m} • q_{2m} computed from T_{2m} and Hu_{2m} • *RR* • *SF* • *TTC*	• P_{mer} or P_{st} following the principle: - if z_{st} <= 500 m, P_{mer} converted into geopotential, - if z_{st} > 500 m, P_{st} converted into geopotential. • u_{10m} and v_{10m} over sea • T_{2m} • *SST* • Hu_{2m} • q_{2m}	• files of blacklist identifiers • u_{10m} and v_{10m} over land • *RR* • *SF* • *TTC*
BATHY TESAC	Oceanic observations (bathymetric sounding)	• u_{10m} & v_{10m} computed from dd_{10m} & ff_{10m} • T_{2m} • SST	• u_{10m} & v_{10m} • T_{2m} • *SST*	• files of blacklist identifiers
BUOY	Oceanic observations	• P_{mer} • u_{10m} & v_{10m} computed from dd_{10m} & ff_{10m} • T_{2m} • *SST*	• P_{mer} converted into geopotential • u_{10m} & v_{10m} • T_{2m} • *SST*	• files of blacklist identifiers
PILOT PILOT-SHIP EUROPROFIL PROFILER	Surface and altitude observations (land and sea radiosonding)	• u_{10m} & v_{10m} computed from dd_{10m} & ff_{10m}	• u_{10m} & v_{10m}	• files of blacklist identifiers • u_{10m} & v_{10m} over land

Observation usage (in variational analysis)

Message type	Observation type	Extracted parameters	Used parameters	Parameter blacklisted or not used
TEMP **TEMP-SHIP** **TEMP-MOBIL** **TEMP-DROP**	**Altitude and surface observations (land, sea and airplane radiosonding)**	• ***Z*** u_{10m} & v_{10m} computed from dd_{10m} & ff_{10m} • ***u*** & ***v*** computed from ***dd*** & ***ff*** • ***T*** • ***Hu*** computed from ***T*** and ***Td*** • ***q*** computed from ***T*** and ***Hu***	u_{10m} & v_{10m} over sea • ***u*** & ***v*** on standard levels and on significant levels • ***T*** on significant and standard levels • ***q*** on the significant and standard levels	• files of blacklist identifiers • levels under ground • ***Hu*** • ***P*** < 300 hPa and on all standard levels • u_{10m} & v_{10m} over land • ***Z***
PILOT **PILOT-SHIP** **EUROPROFIL** **PROFILER**	**Surface and altitude observations (land and sea radiosonding)**	• u_{10m} & v_{10m} computed from dd_{10m} & ff_{10m} • ***u*** & ***v*** computed from ***dd*** & ***ff***	• u_{10m} & v_{10m} • ***u*** & ***v*** on the standard and significant levels	• files of blacklist identifiers • u_{10m} & v_{10m} over land
SATOB **SATGEO** **GEOWIND**	**Altitude observations (geostationary satellite)**	• ***u*** & ***v*** computed from ***dd*** & ***ff*** in channels VIS, IR and WV • Quality indicator	• ***u*** & ***v*** • ***QI***	• ***u*** and ***v*** originated from SATOB INSAT
AIREP **AMDAR** **ACAR**	**Multilevel observations (aircraft messages)**	• ***u*** & ***v*** computed from ***dd*** & ***ff*** • ***T*** • ***Hu*** computed from ***T*** and ***Td***. In practice, ***Td*** is always missing	• ***u*** & ***v*** • ***T***	• files of blacklist identifiers • ***Hu***

Observation usage (Satellite radiances)

Message type	**Observation type**	**Extracted parameters**	**Used parameters**	**Parameters blacklisted or not used**
1C level Radiances	Observation: multichannels Polar Satellites	•*Tb* •*Sat identifier* •*Scanning line* •*Scan angle (number of FOV)* •*Zenith angle* •*Azimuth angle* •*Satellite altitude*	•*Tb in some channels, depending on the nature of the surface (land, sea or ice), cloud fraction and model orography*	•*Tb in some channels, depending on the nature of the surface (land, sea or ice), cloud fraction and model orography*

Example: Use of RAD1C AMSU-A

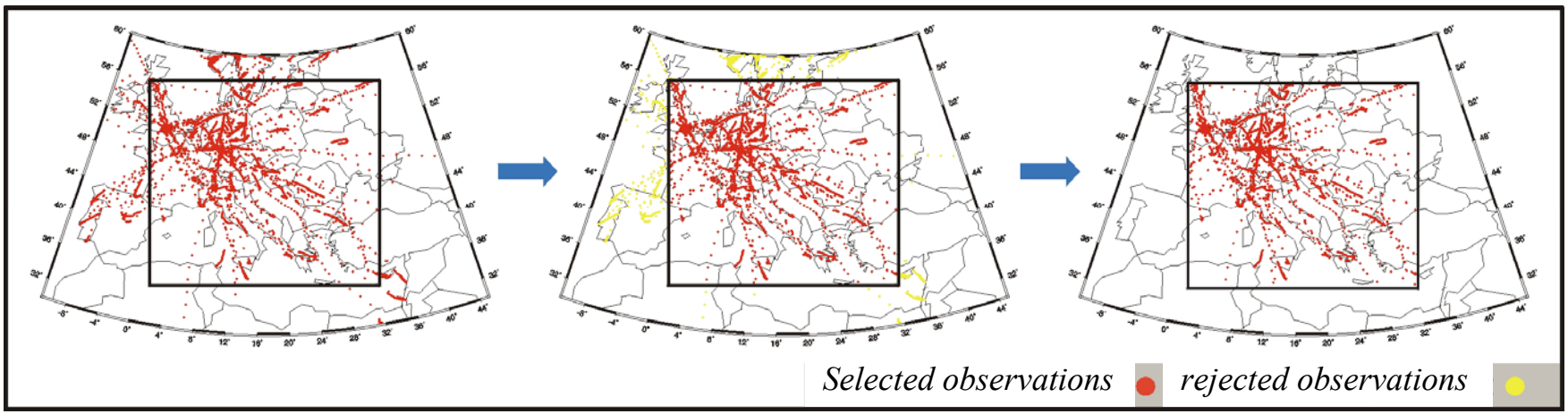
Channel number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Over Land					x	x	x	x	x	x	x	x			
Over Sea					x	x	x	x	x	x	x	x			
Over Sea ice							x	x	x	x	x	x			
Cloudy pixel								x	x	x	x	x			

Over land channels 5 and 6 are used when the model orography is less than 500m and 1500m, respectively

Selection of the observations

- Selection of observations

→ Select observations inside the LAM domain



The observation screening

→ The screening is the last step of the pre-processing in data assimilation

→ We want to remove the wrong data and make the final choice inside the set of data which are said to be potentially acceptable by the control and the monitor

→ We need to control them against the background, to verify their vertical consistency and to thin them when their spatial resolution is too high

The observation screening

1. ➔ Preliminary check of observations

→ Check of completeness of the reports

➔ Check of the reporting practice for the surface & radio-sonde mass observations [Ps & Φ]

→ Scanning for blacklisting :

➔A check for the blacklisted parameters is applied to discard the excessively noisy or biased observations.

➔A selection of variables useful for the analysis system is done using the data selection apart of the blacklist file and the information hard-coded in the model (orographic rejection limit, land-sea rejection…).

The observation screening

This is when we assume that both the observation and the background errors are unbiased and the errors covariance, associated with the probability distribution of the innovations to be Gaussian

2. ➔ Backg

→ It is pe ssimilation

! The va by :

$$\langle y - H(x_b), y - H(x_b) \rangle \sim \sigma_o^2 + \sigma_b^2$$

H- observation operator, *y*- observation array, σ_0 and σ_b – observation and background error standard deviations, < >- variance

! The estimate of variance for the normalized departure is given by :

$$(1 + \frac{\sigma_o^2}{\sigma_b^2}) = \lambda$$

→It is very unlikely that $\left(\frac{y - H(x_b)}{\sigma_b} \right)^2 > n * \lambda$

In the literature it refers to the optimal interpolation quality control

→ OI QC

The observation screening

3. ➔ Vertical consistency of multilevel reports

→ The duplicated levels, in multi-level reports, are removed from the reports

→ If 4 consecutive layers are found to be of suspicious quality then these layers are rejected

4. ➔ Removal of duplicated reports

→ In case of co-located reports of the same observation types, some or all of the content of one of the reports is rejected

The observation screening

5. ➔ Redundancy check

→ It is performed for the active reports that are co-located and originate from the same station/platform

→ For land SYNOP, the report closest to the centre of the screening time window with most active data is retained

→ The ship SYNOP are redundant if the moving platforms are within a circle of 1 degree radius

→ TEMP and PILOT from the same station are considered at the same time in the redundancy check

→ A SYNOP mass observation is redundant if there are any TEMP geopotential height observations (made in the same time and the same station) that are no more than 50hPa above the SYNOP mass observation

The observation screening

6. ➔ Thinning

→ The horizontal and vertical resolutions of several types of observations are too high to allow to use all of them in the analysis (knowing that no horizontal correlation of observation errors is supported)

→ A horizontal thinning is performed for the Aircraft (**AIREP**, **AMDAR**, ...), the AMV (**SATOP** or **GEOWIND**) and the sat. observations (**ATOVS**, **SSM/I**, **AIRS**, ...etc) reports

→ A vertical thinning is also performed for the Aircraft and the AMV reports

→ A predefined minimum horizontal distance between the nearby reports from the same platform is enforced

The observation screening

6/a. ➔ Thinning of the Aircraft data

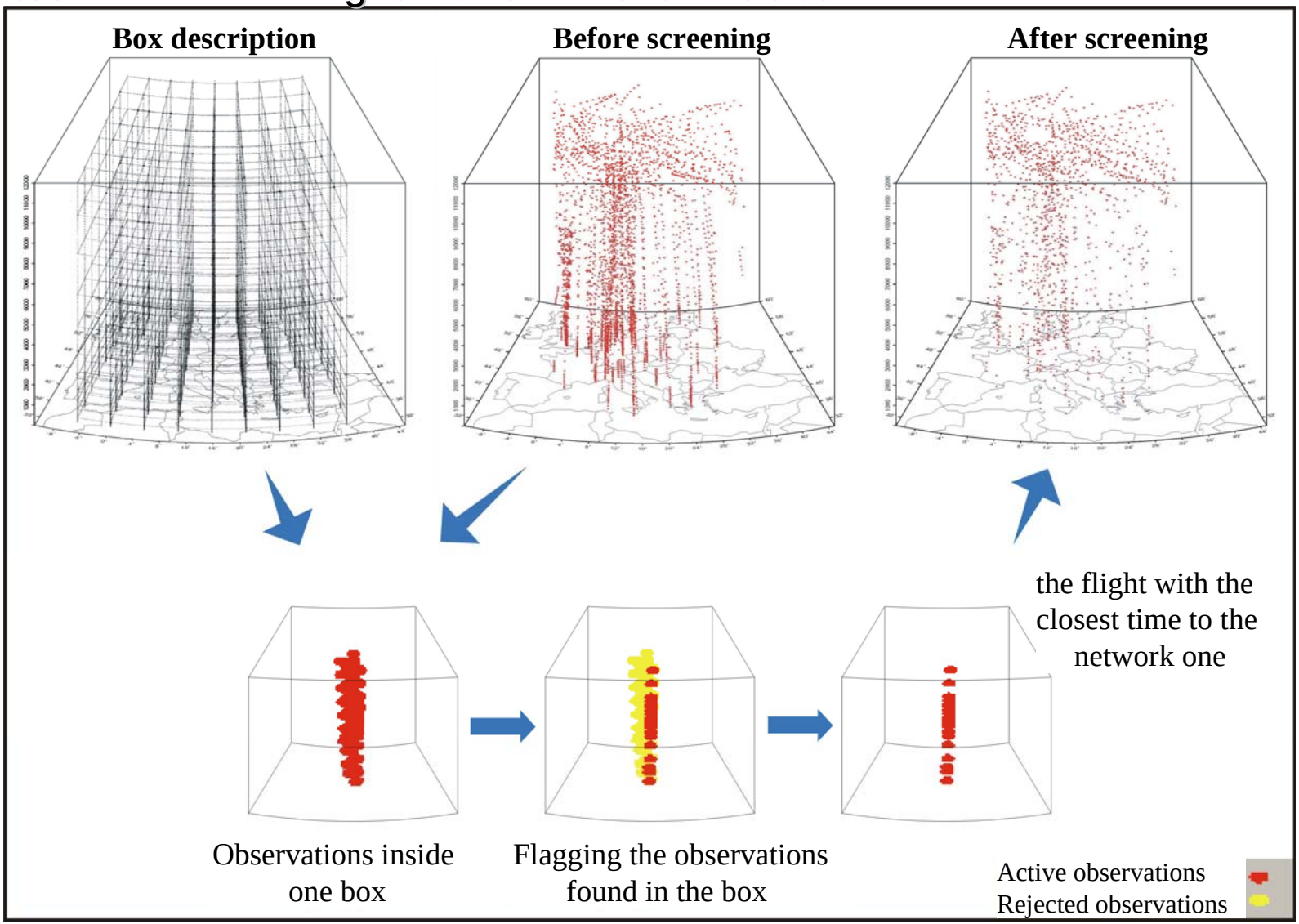
→ A minimum distance is imposed.

→We want to keep only one obs per box defined horizontally by the minimum distance and vertically around the standard or model levels

(in the ARPEGE/ALADIN : the screening is applied separately to different flights
➔ problem when in LAM DA we would like to use these data in thinner resolution)

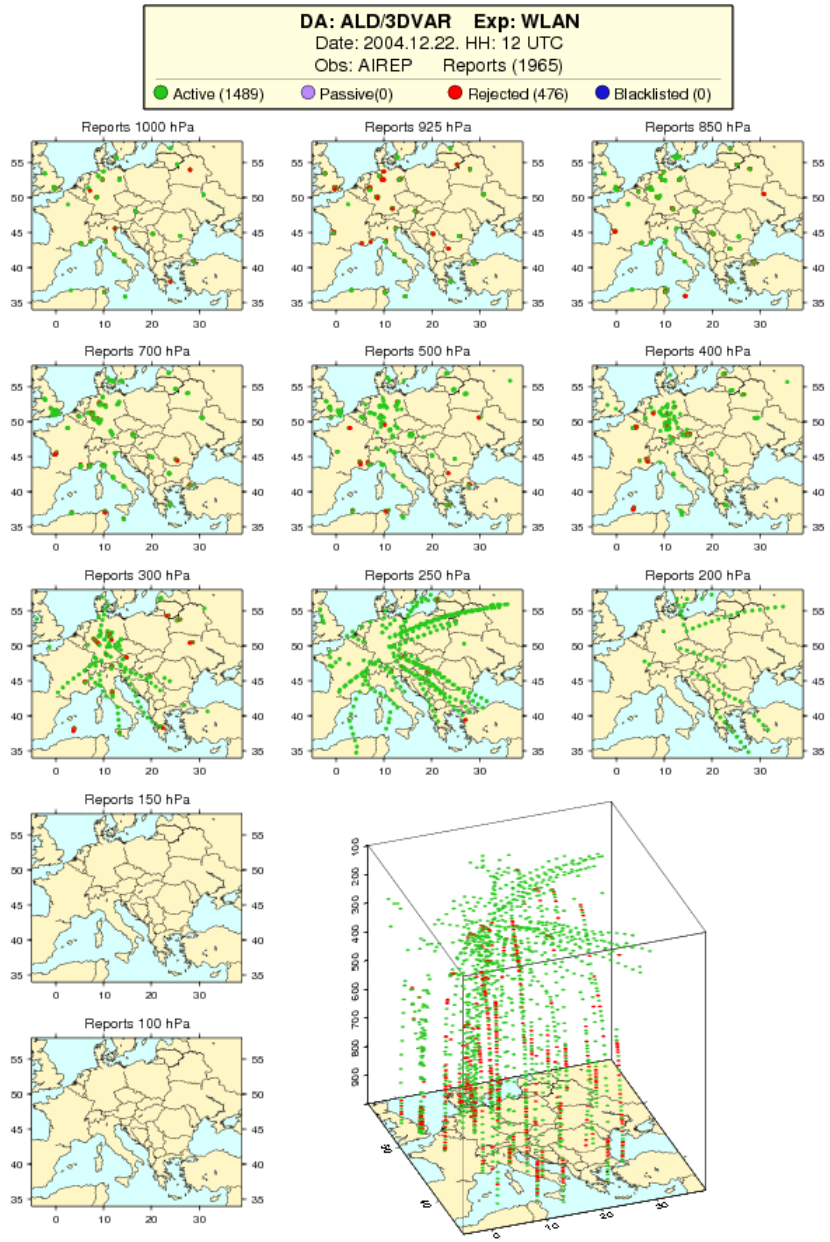
The observation screening

6/a. ➔ Thinning of the Aircraft data



The observation screening

6/a. ➔ Thinning of the Aircraft data



The observation screening

6/b. → Thinning of the AMV data

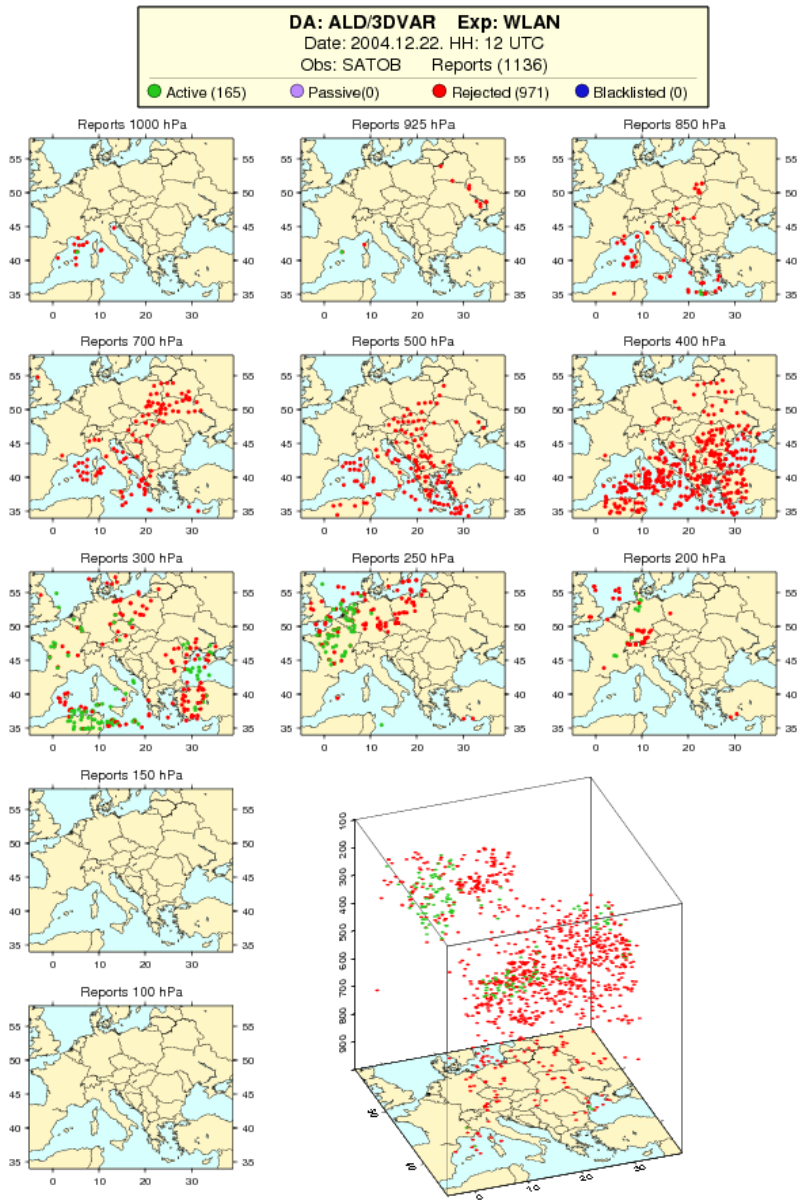
→ First a minimum distance is enforced

→ Then a repeated scan is performed to achieve the final separation

→ We want to keep only one obs per box defined horizontally by the minimum distance and vertically around the standard levels

The observation screening

6/a. ➔ Thinning of the Atmospheric Motion Vector (AMV) data

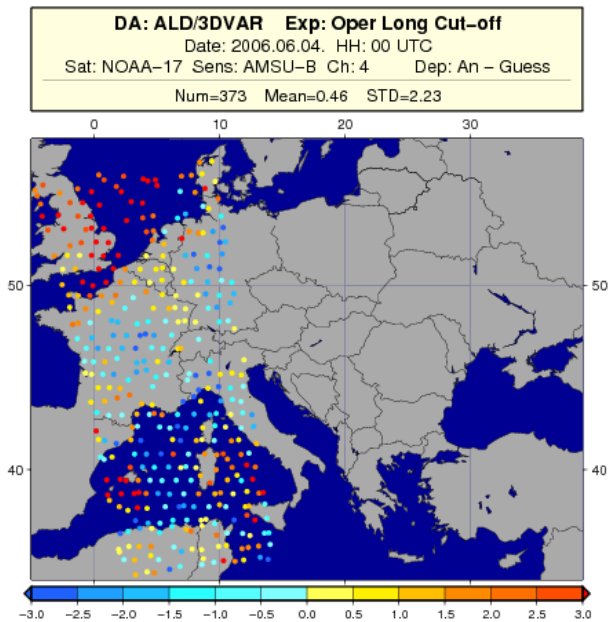


The observation screening

6/c. ➔ Thinning of the Satellite data

→ First a minimum distance is enforced.

→ Then a repeated scan is performed to achieve the final separation.



The observation screening

1- Surface observations

2- Aircraft observations

5- Radiosonde observations

7- Satellite observations

*** FOR WHOLE OBSERVATION ARRAY

STATUS SUMMARY OF REPORTS:

OB.TYP	REPORTS	ACTIVE	PASSIVE	REJECTED	BLACKLISTED
1	759	758	0	1	0
2	1851	1825	0	26	0
3	0	0	0	0	0
4	0	0	0	0	0
5	6	6	0	0	0
6	0	0	0	0	0
7	6304	274	0	6030	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
TOT	8920	2863	0	6057	0

STATUS SUMMARY OF DATA:

OB.TYP	DATA	ACTIVE	PASSIVE	REJECTED	BLACKLISTED
1	4849	2829	2016	2020	0
2	5523	5417	2	106	31
3	0	0	0	0	0
4	0	0	0	0	0
5	1057	899	153	158	0
6	0	0	0	0	0
7	28629	576	0	28053	20078
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
TOT	40058	9721	2171	30337	20109

Variational analysis and quality control

Good data
First guess

→ Analysis

↓ Minimization of the cost function

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(H(x) - y)^T R^{-1}(H(x) - y)$$

$J_o(x)$- *obs. cost function*

With the assumption that errors are Gaussian in nature

If the observations errors are uncorrelated, then $J_o(x)$ simplifies to:

$$J_o(x) = \sum_{i=1}^{N} \frac{1}{2}\left[\frac{y - H(x)}{\sigma_o}\right]_i^2$$

The general expression for the observation cost function is based on the probability density function (pdf) of the observation error distribution (Lorenc 1986)

$$J_o = -\ln p + const$$

p *is the* pdf *of obs. error*

Arbitrary constant chosen Jo=0 when y=H(x)

The variational QC

➔ The normal (Gaussian) distribution (N) $$N = \frac{1}{\sigma_o \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{y - H(x)}{\sigma_o} \right)^2 \right]$$

➔ In an attempt to better describe the tails of the observed distributions, the pdf is written as a sum of two distinct distributions by Ingleby and Lorenc (1993), see also Andersson and Järvinen (1999):

$$p^{QC} = (1 - A)N + Ap^G$$

*Normal distribution (*pdf*) for "good" data*

pdf *for data affected by gross errors.* A is the prior probability of gross error

➔ In VarQC a non-Gaussian pdf will be used, resulting a non-quadratic expression for *Jo*.

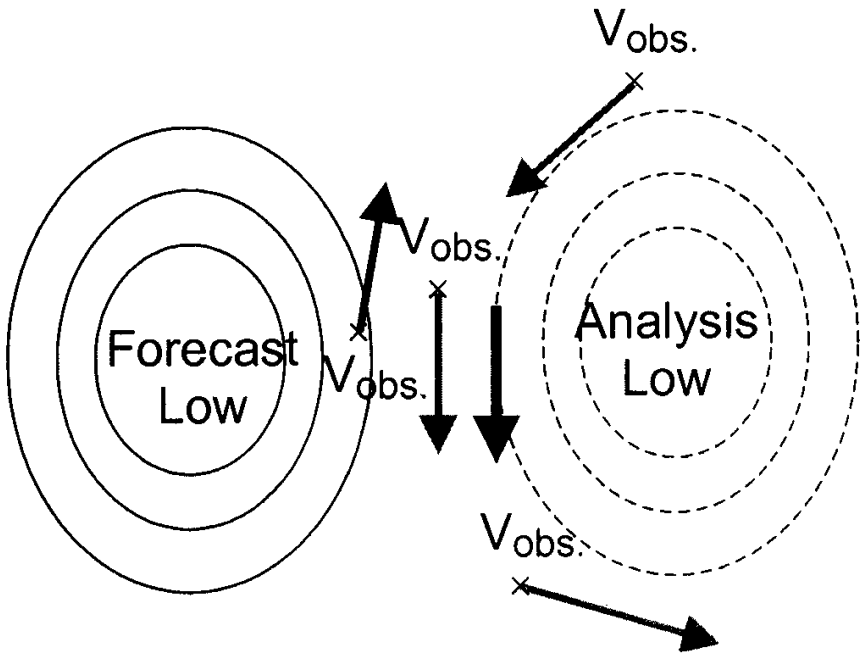
➔ Data which are found likely to be incorrect are given reduced weight in the analysis.

➔ VarQC requires a good "preliminary analysis". Otherwise incorrect QC-decision will occur.

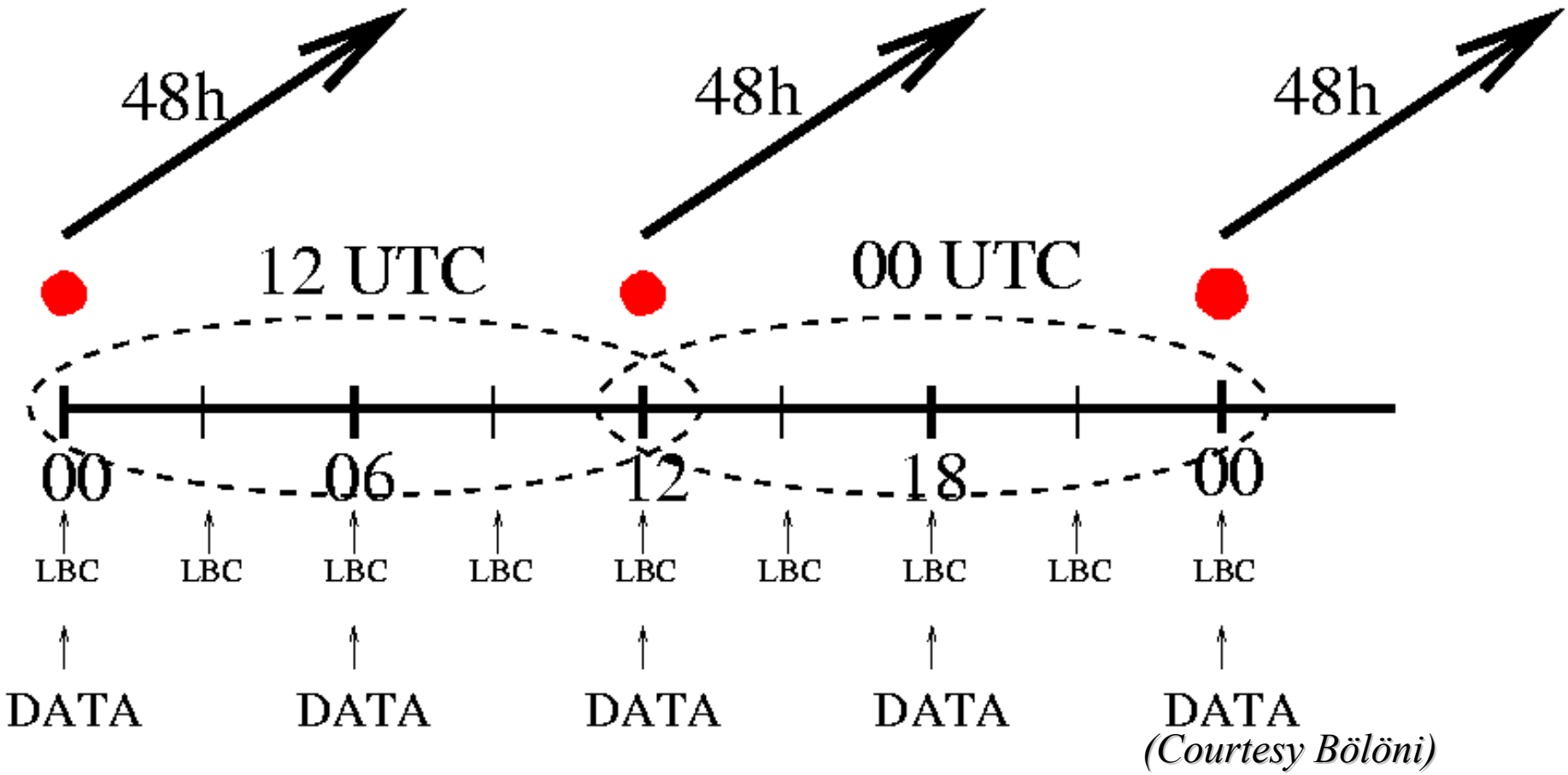
➔ Therefore, in operational VarQC is switched on after certain iterations.

The variational QC

Gauthier et al., 2003: Describing the advantage of the VarQC in addition to the OI one.

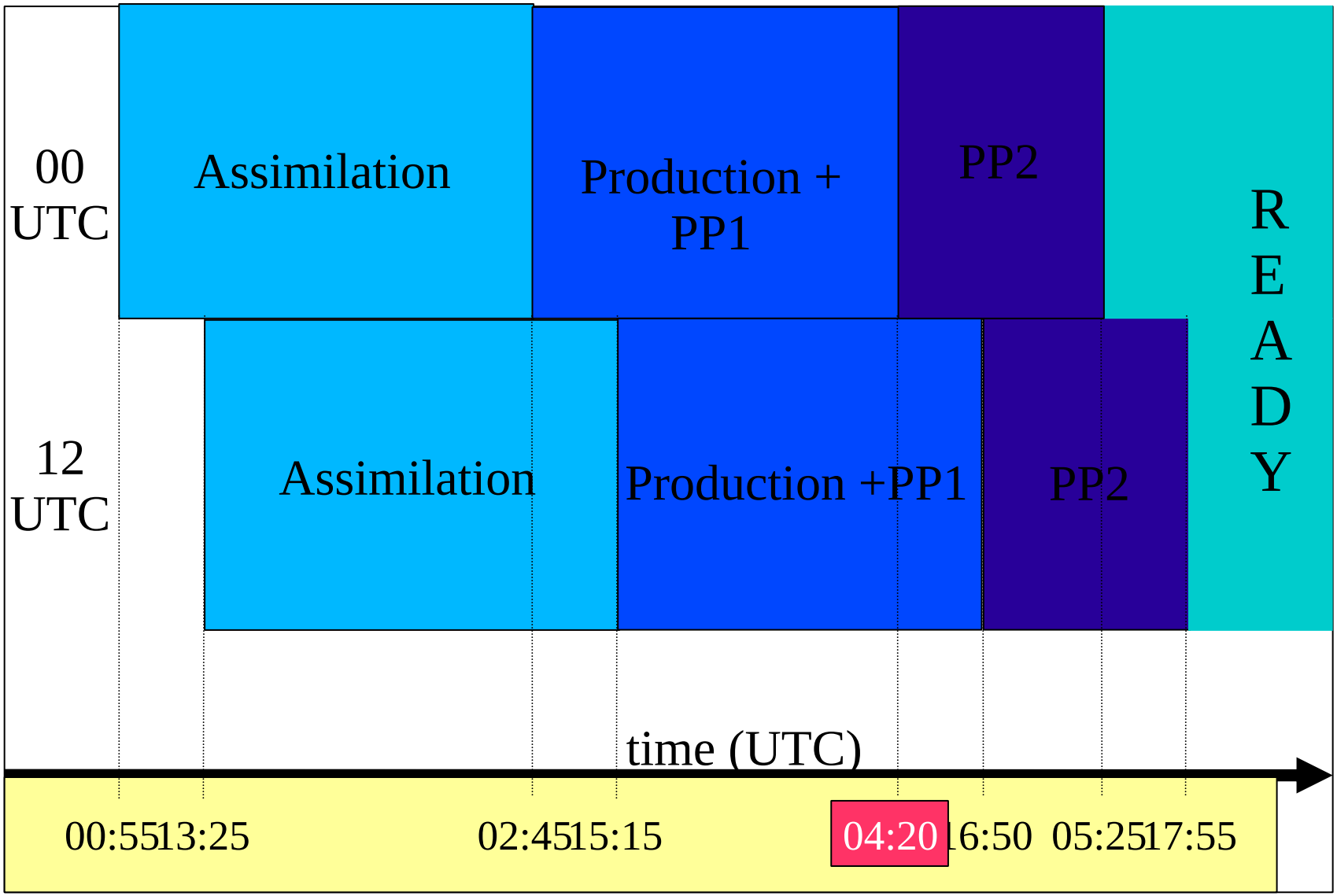


Assimilation cycle



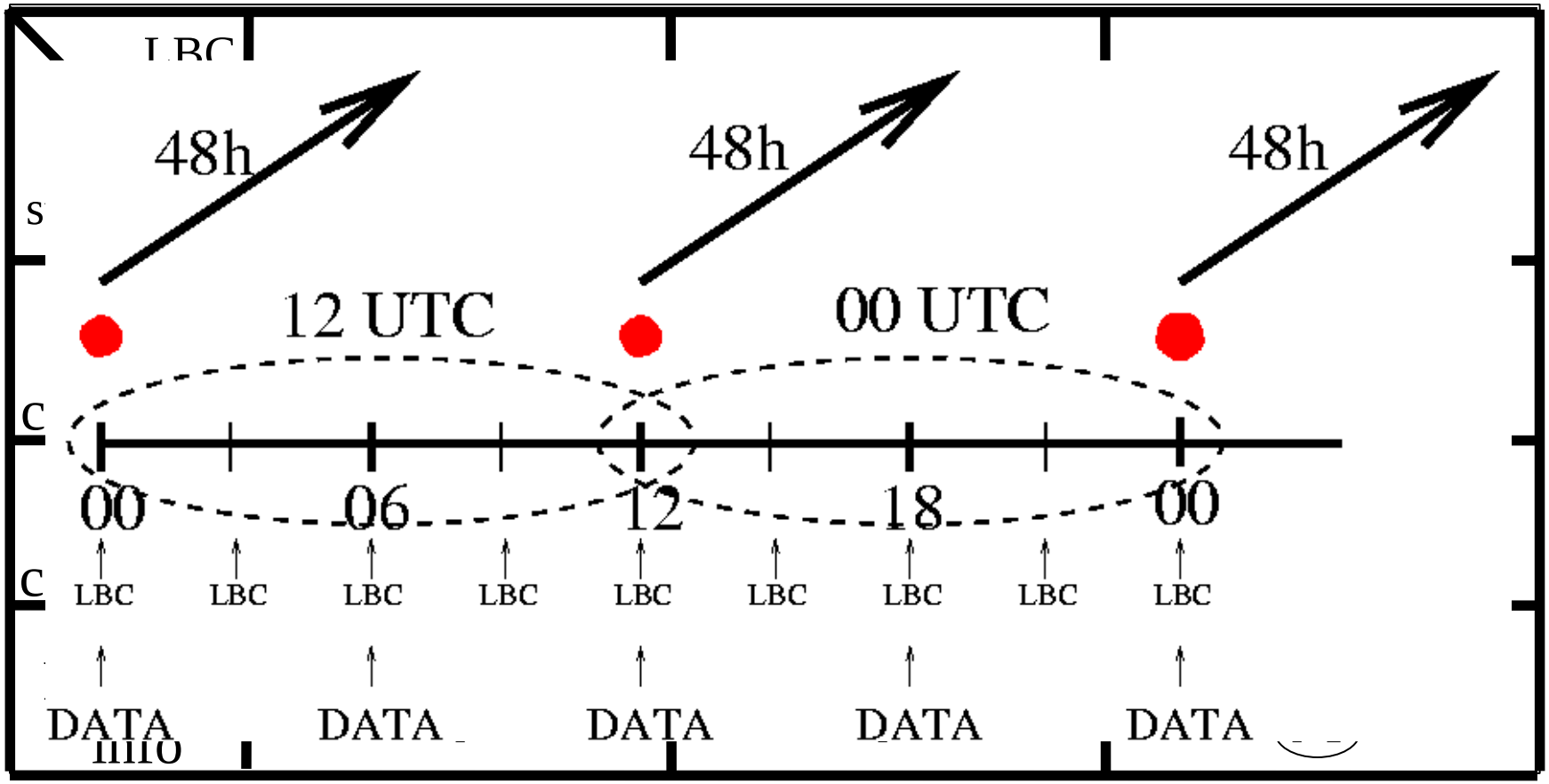
(Courtesy Bölöni)

Assimilation cycle - schedule



(Courtesy Bölöni)

Assimilation cycle – Choice of coupling files
Lateral boundary conditions



Strategies in ALADIN

(Courtesy Bölöni)

Assimilation cycle – Choice of coupling files
Lateral boundary conditions

LBC range / BG forecast	+000 hours	+003 hours	+006 hours
00 ⟶ 06 UTC	ARP $_{long}$ +000	ARP $_{long}$ +003	ARP $_{long}$ +000
06 ⟶ 12 UTC	ARP $_{long}$ +000	ARP $_{long}$ +003	ARP $_{short}$ +000
12 ⟶ 18 UTC	ARP $_{long}$ +000	ARP $_{long}$ +003	ARP $_{long}$ +000
18 ⟶ 00 UTC	ARP $_{short}$ +000	ARP $_{short}$ +003	ARP $_{short}$ +006

(Courtesy Bölöni)

The surface analysis

1. Take the ARPEGE analysis.

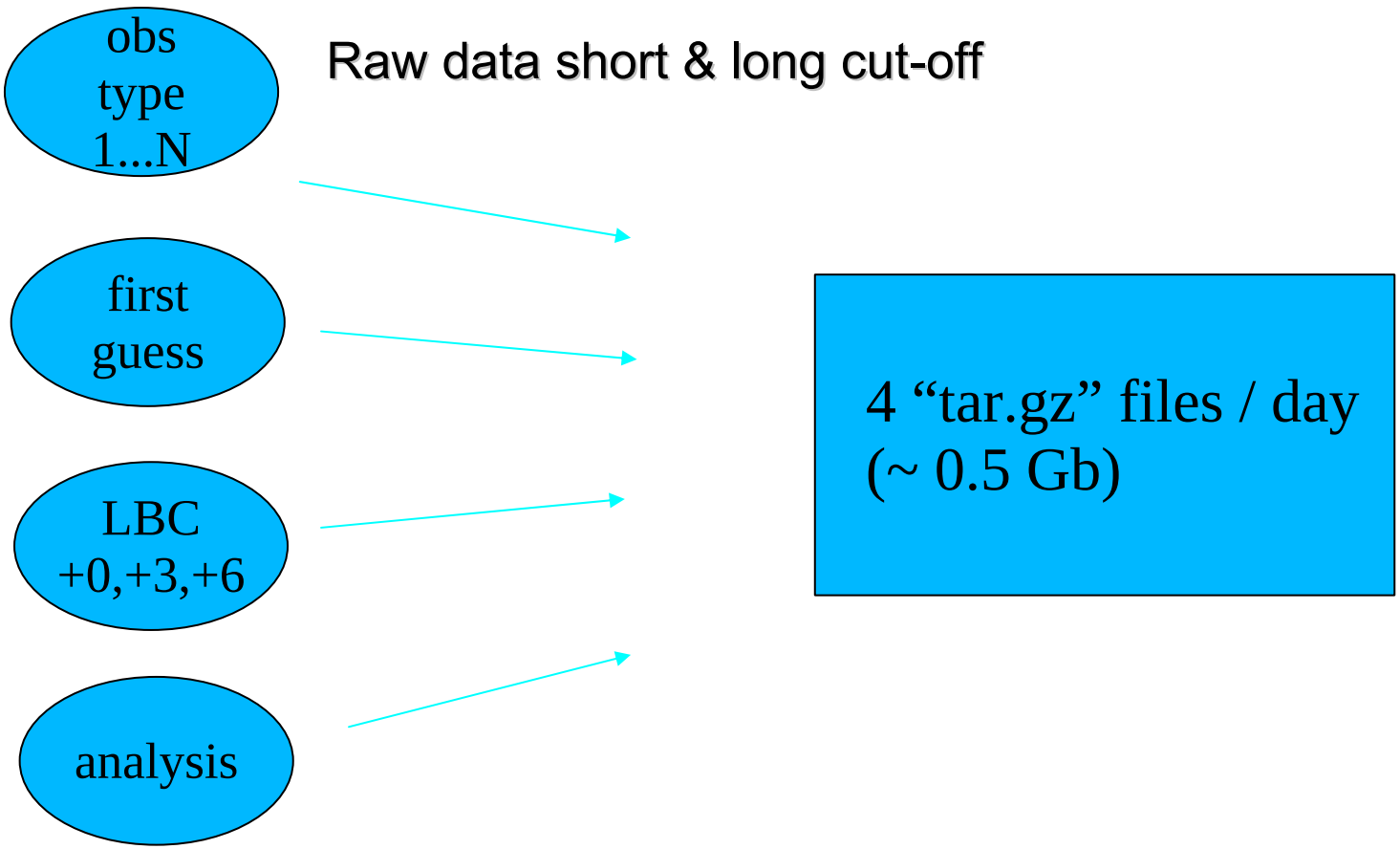
2. Local OI analysis

network time	surface field
06 UTC	ARP_{long} +000
12 UTC	ARP_{short} +000
18 UTC	ARP_{long} +000
00 UTC	ARP_{short} +000

(Courtesy Bölöni)

Archive

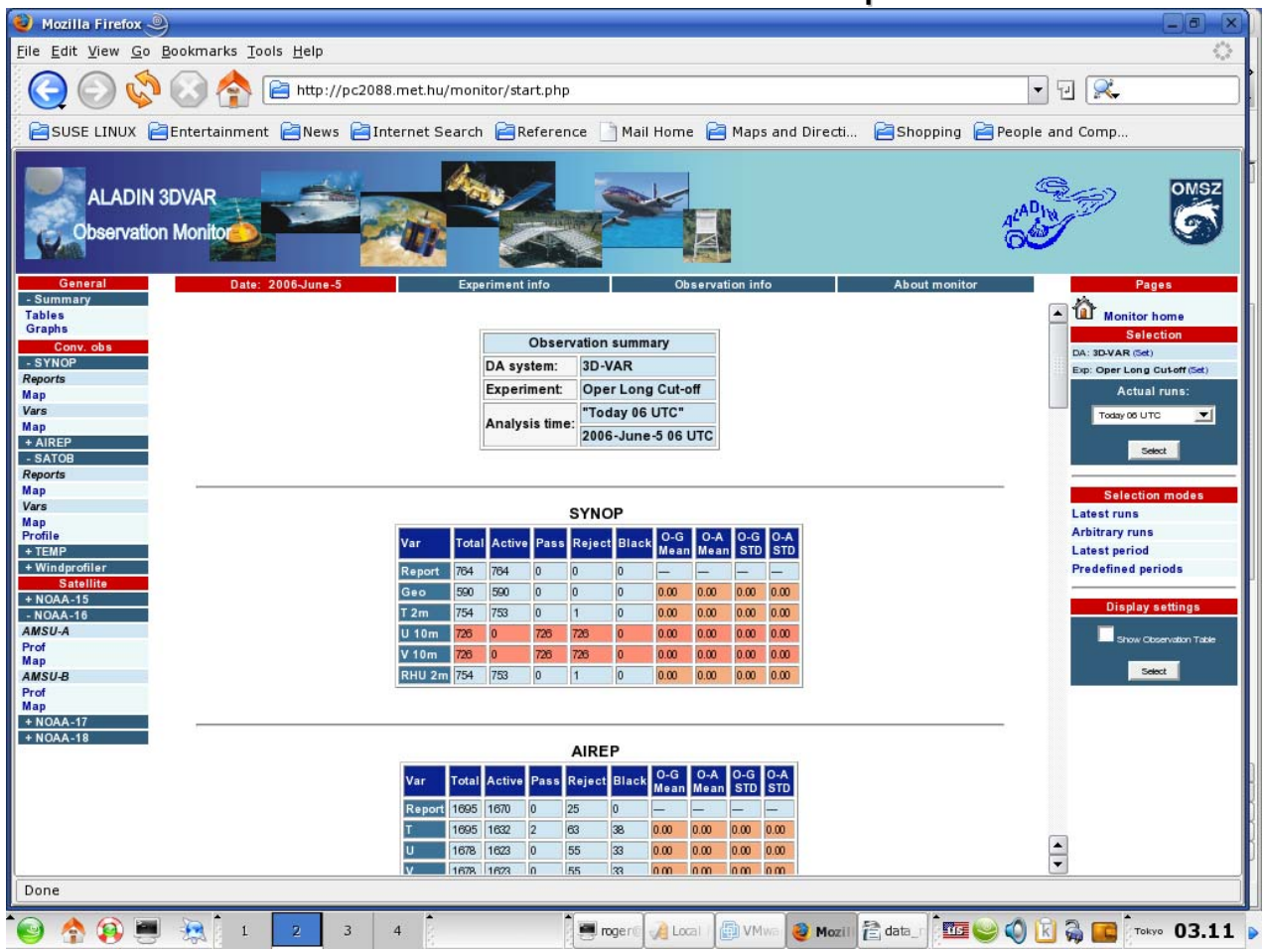
Purpose: reproducible assimilation cycle



Need of the DA system monitoring

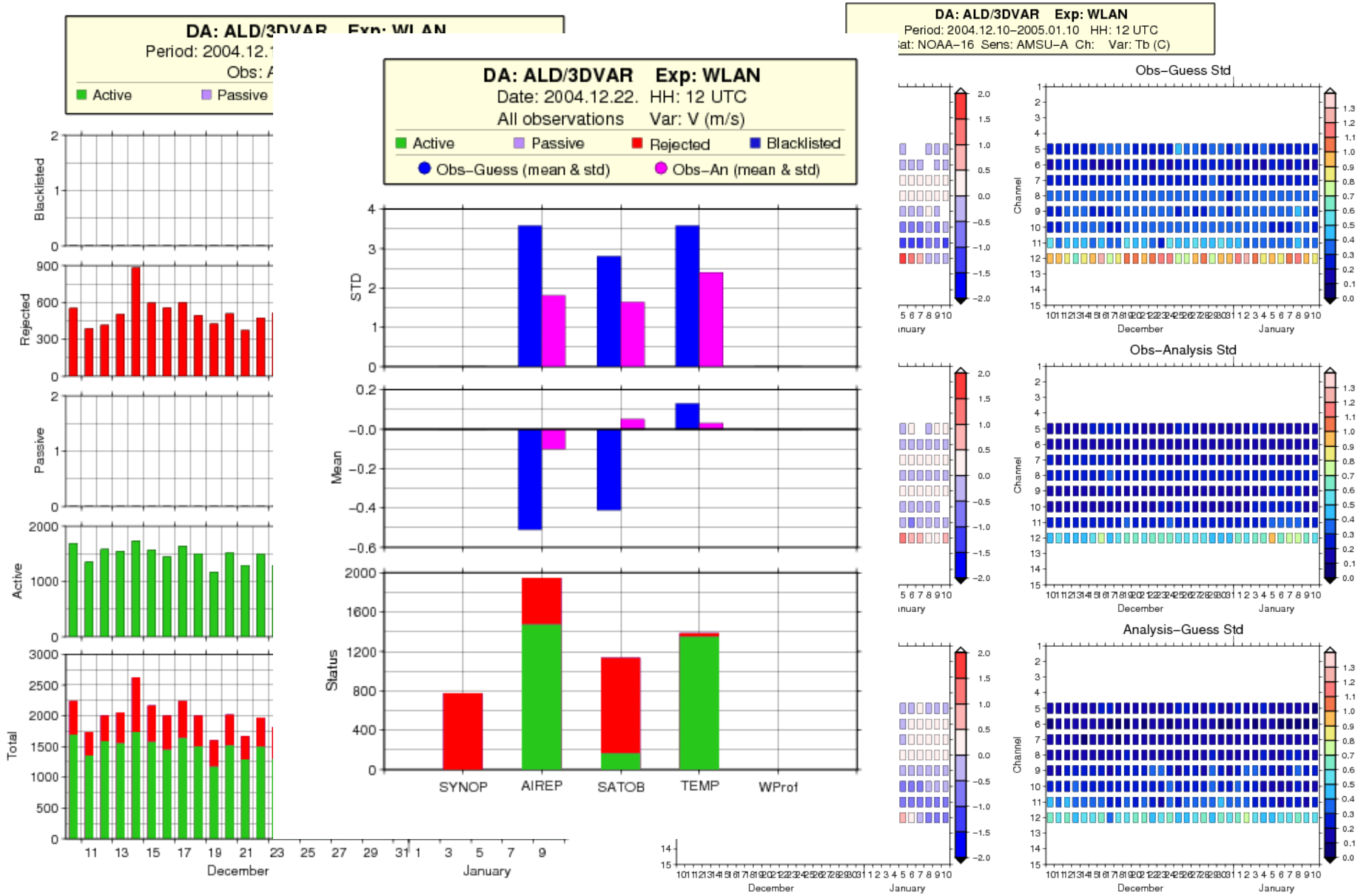
➔Once the system is ready for operational suite, we need to build a post-processing able to give complex description about the use and the quality of different observations involved in the assimilation process

➔ The post-processing is needed to meet technical and scientific aspects of the analysis system

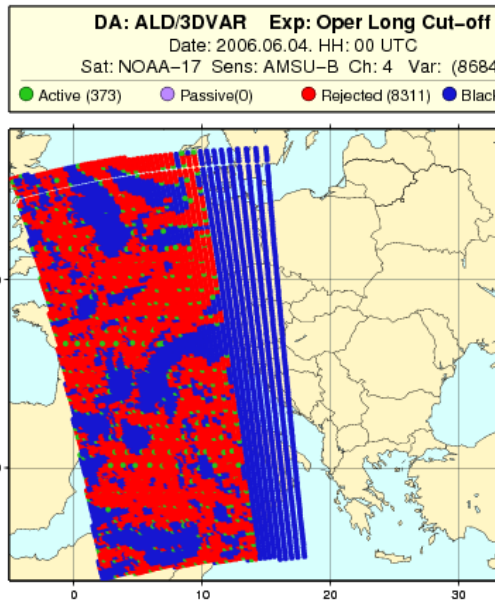


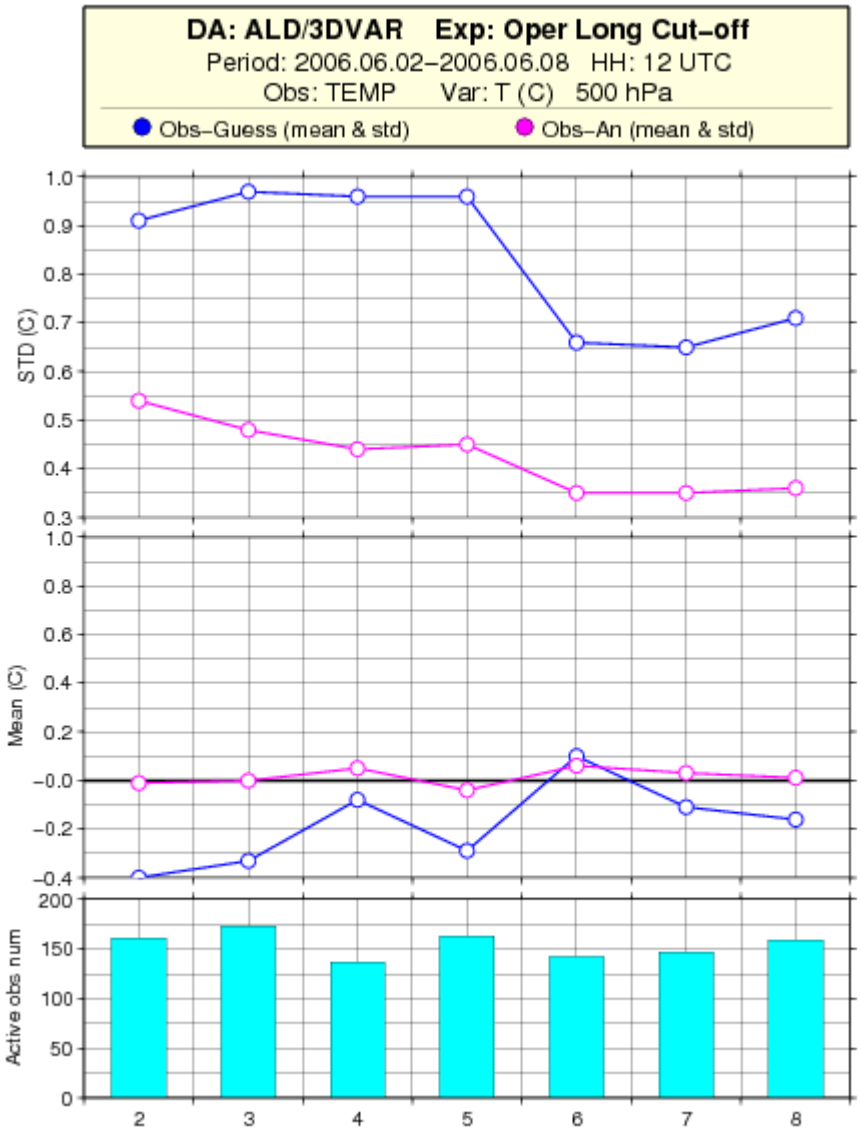
➔ We need system

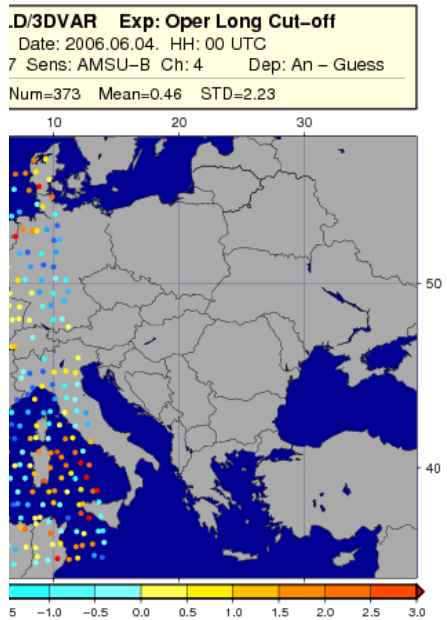
ready to diagnose an experimental run



➔ Ready to diagnose the operational run







Summary

➔We saw the importance of the continuous observation monitoring to guarantee the use of “good observations” through the blacklisting technique

➔We discussed the process of data quality control in LAM data assimilation system

➔We showed how data assimilation scheduling is built up

➔A complex monitoring system is needed to ensure an effective operational data assimilation

Thank you for your attention !

References

Andersson, E and Järvinen, H, 1999: *Variational quality control. Q. J. R. Meteorol. Soc.,* 125, 697-722.

Andersson, E, 2006: Variational quality control, *ECMWF training course on data assimilation and the use of satellite data*, 22-31 March 2006, Reading, UK, available on the ECMWF web page.

Gauthier, P, Chouinard, C. and Brasnett, B, 2003: *Quality control: Methodology and application, IV Earth and Environmental Science,* vol. 26, 177-187.

Ingleby, N. B. and Lorenc, A. C. 1993: *Bayesian quality control using multivariate normal distributions, Q. J. R. Meteorol. Soc.,* 119, 1195-1225.

Moll, P and Gérard, É, 2001: *ALATNET Seminar on Data Assimilation,* Gourdon, 11-22 June 2001.

Peter Lynch, Lecture on Numerical Weather Prediction, MAPH (313), http://maths.ucd.ie/met/msc/NWP/MAPH-P313.html