Data assimilation in meteorology

Loïk Berre
Météo-France/CNRM

Plan of the talk

- Numerical Weather Prediction (NWP),
  Data Assimilation (DA)

- Observations (in-situ and remote sensing)

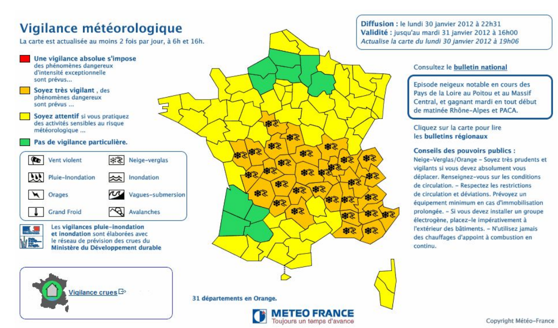
- Error covariances : estimation and modelling

1. Numerical Weather Prediction and Data Assimilation

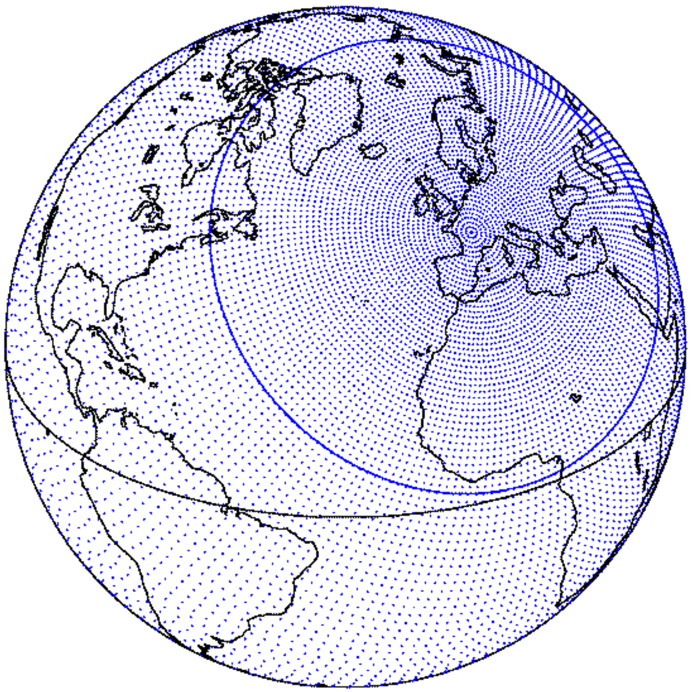
Numerical Weather Prediction

Numerical resolution of fluid mechanics equations (computer code), to **forecast the atmospheric evolution** from an **estimated initial state** (which is called the « analysis »).

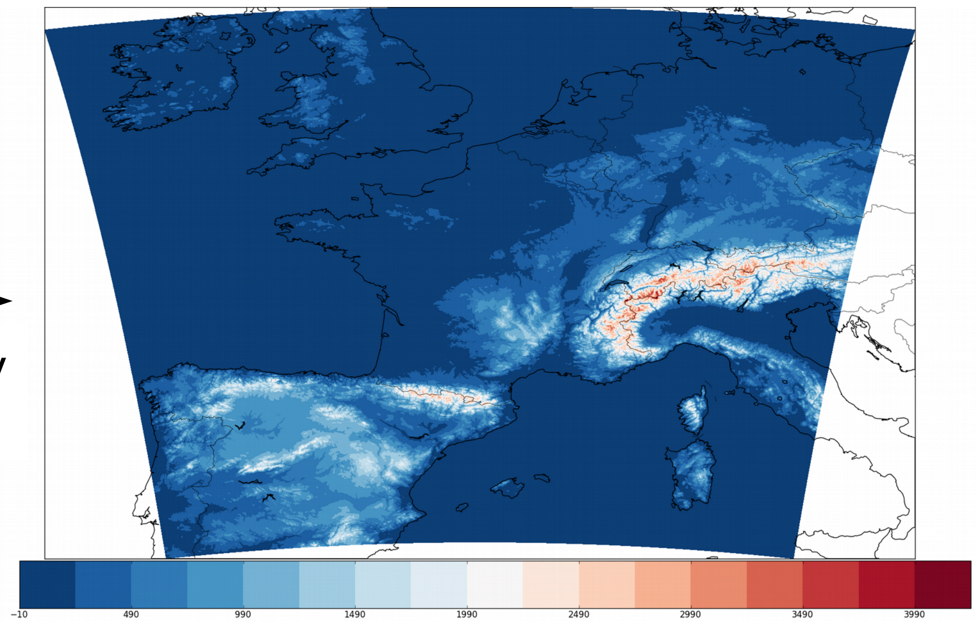
Initial state → Numerical model → Forecast state



NWP models at Météo-France (in collaboration with e.g. ECMWF)



Lateral Boundary Conditions

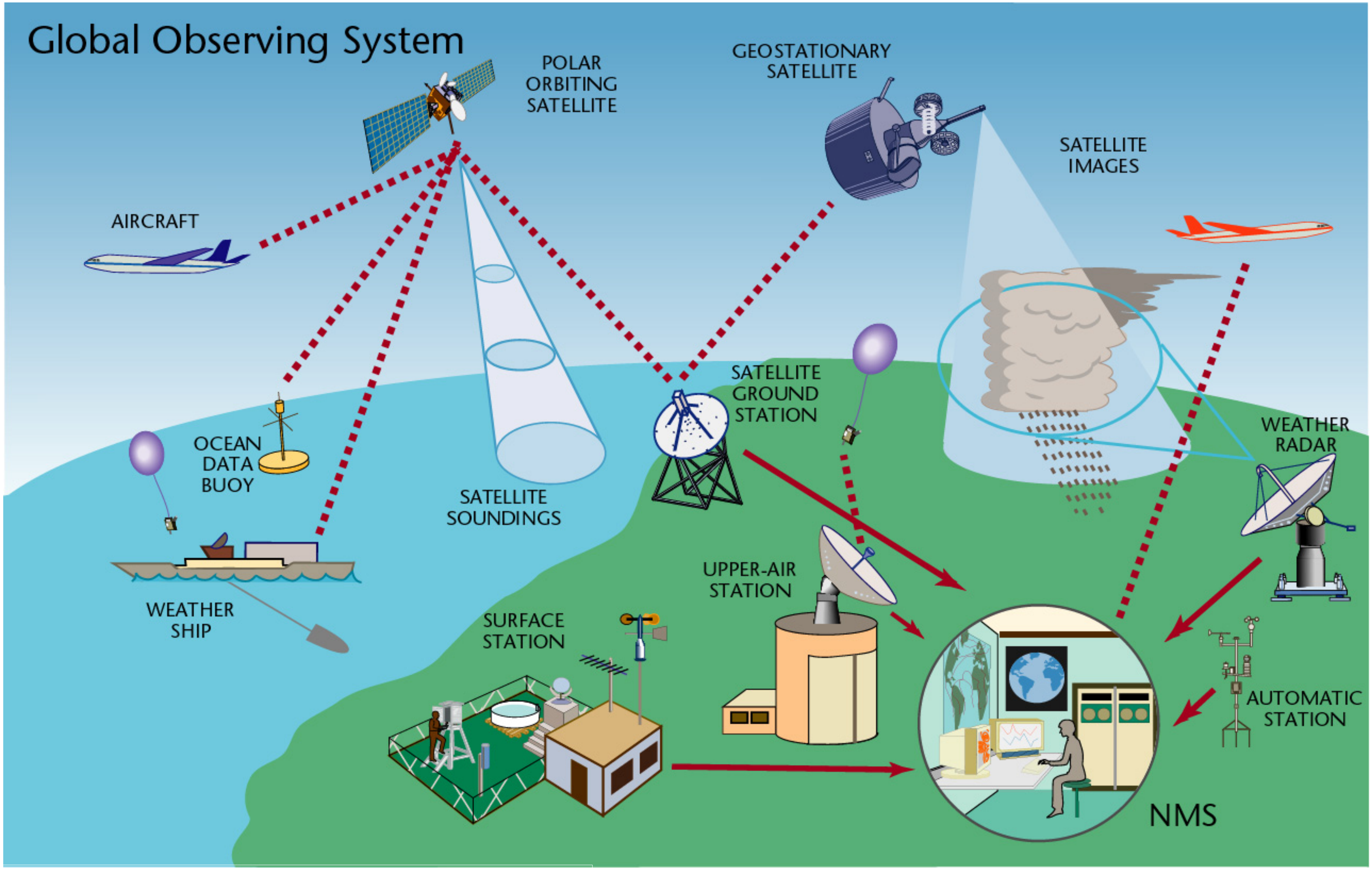


ARPEGE (7 km - 40 km)
10^9 model variables

AROME (1.3 km)
$1{,}4 \times 10^9$ model variables

Equations of dynamics and physical parametrizations (radiation, convection, …) to predict the evolution of temperature, wind, humidity, etc.

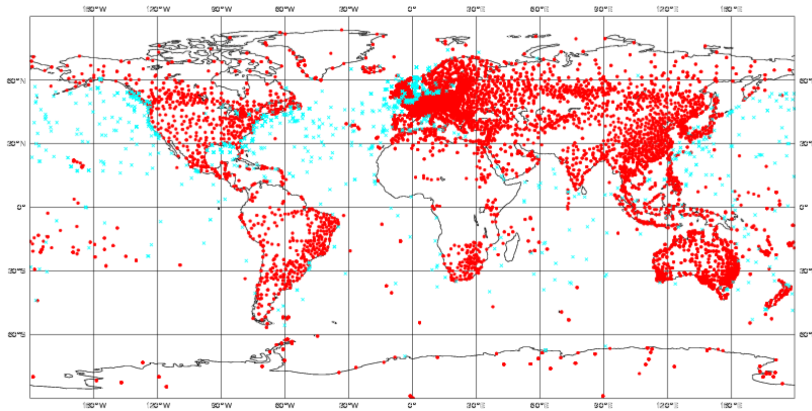
Data which are assimilated in NWP models



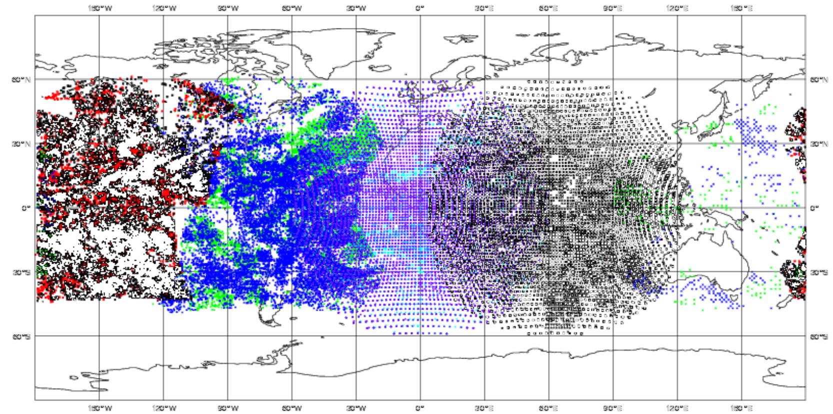
ARPEGE	AROME
10^9 model variables	1,4 x 10^9 model variables
5×10^6 observations / 6h	2×10^5 observations / 6h
90 % satellite	Up to 70 % radar, 10 % satellite
4D assimilation : 40 min	3D assimilation : 7 min

Spatial coverage & density of observations

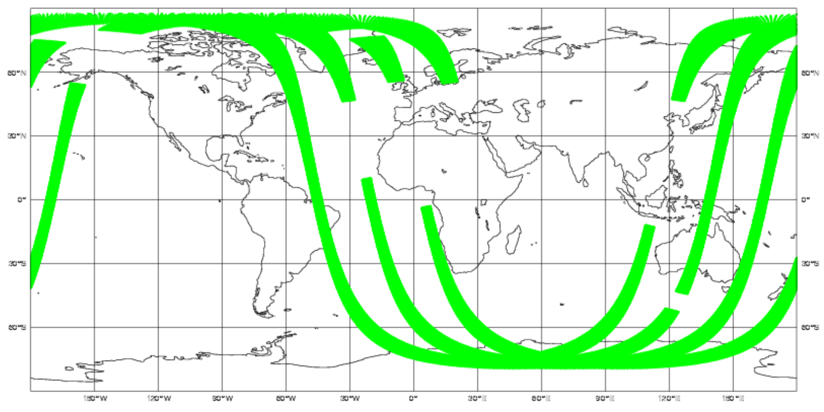
SURFACE DATA



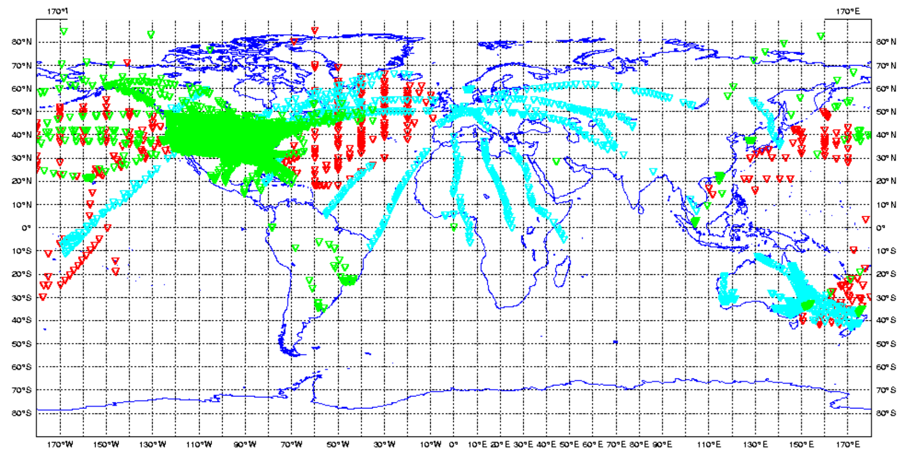
GEOSAT. WINDS



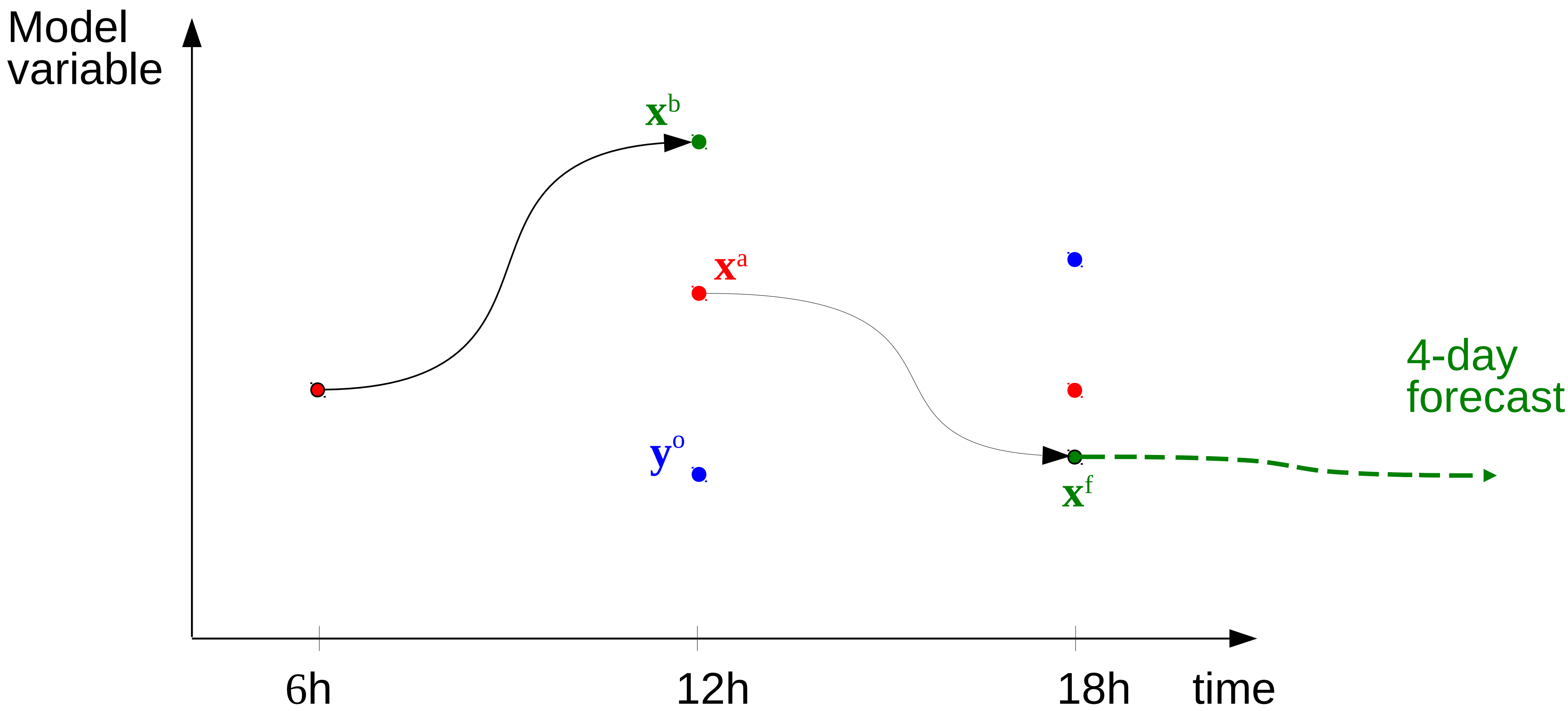
SCATTEROMETER



AIRCRAFT DATA

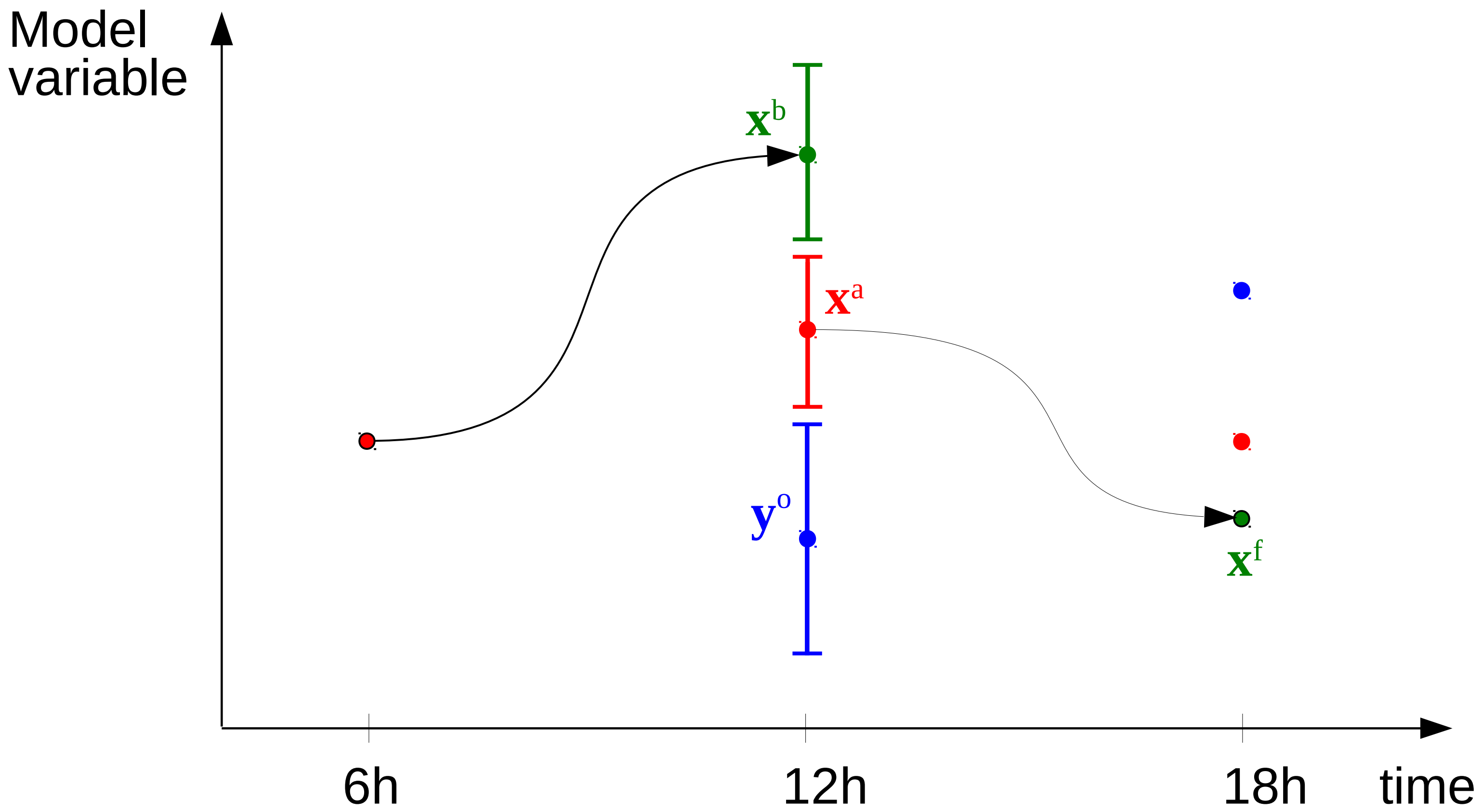


Data assimilation cycling :
temporal succession of analysis and forecast steps



=> The model state is propagated and then updated, e.g. every 6h :
the memory of DA system is updated ~ continuously in time, through cycling.

Uncertainties in observations, background, analysis



Uncertainties are often measured by error variances
(ex : accurate observation ↔ small observation error variance).

=> Use linear estimation theory to account for errors.

Analysis equation

- BLUE formalism : $\mathbf{x}^a = \mathbf{x}^b + \mathbf{K} \, (\mathbf{y}^o - H[\mathbf{x}^b] \,)$

- *H* = non linear observation operator

  = projection from model space to observation space : $\mathbf{y} = H[\mathbf{x}]$

H ~ spatial interpolation : from model grid to obs locations (e.g. for radiosondes) ;

H ~ radiative transfer : from model temperature to satellite radiances ;

H ~ NWP model: for observations available at different times within DA window.

- **K** ~ observation weights : $\mathbf{K} = \mathbf{B} \, \mathbf{H}^T \, (\mathbf{H} \, \mathbf{B} \, \mathbf{H}^T + \mathbf{R} \,)^{-1}$

with **H** = tangent linear version of ***H***,

B = background error covariance matrix,

R = observation error covariance matrix.

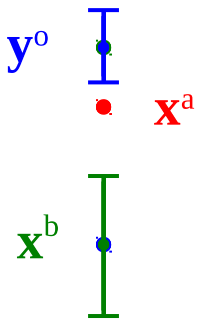
=> **K** accounts for relative accuracy of observations,

and for spatial structures of background errors.

Background error covariances : filtering and propagation of $\mathbf{y}^o - H(\mathbf{x}^b)$

- Variances

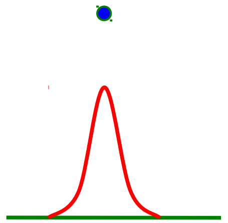
  - Weighting/filtering of observations.



- 3D spatial correlations

  - Spatial propagation of observations.

  - Spatial coherence of analysis.

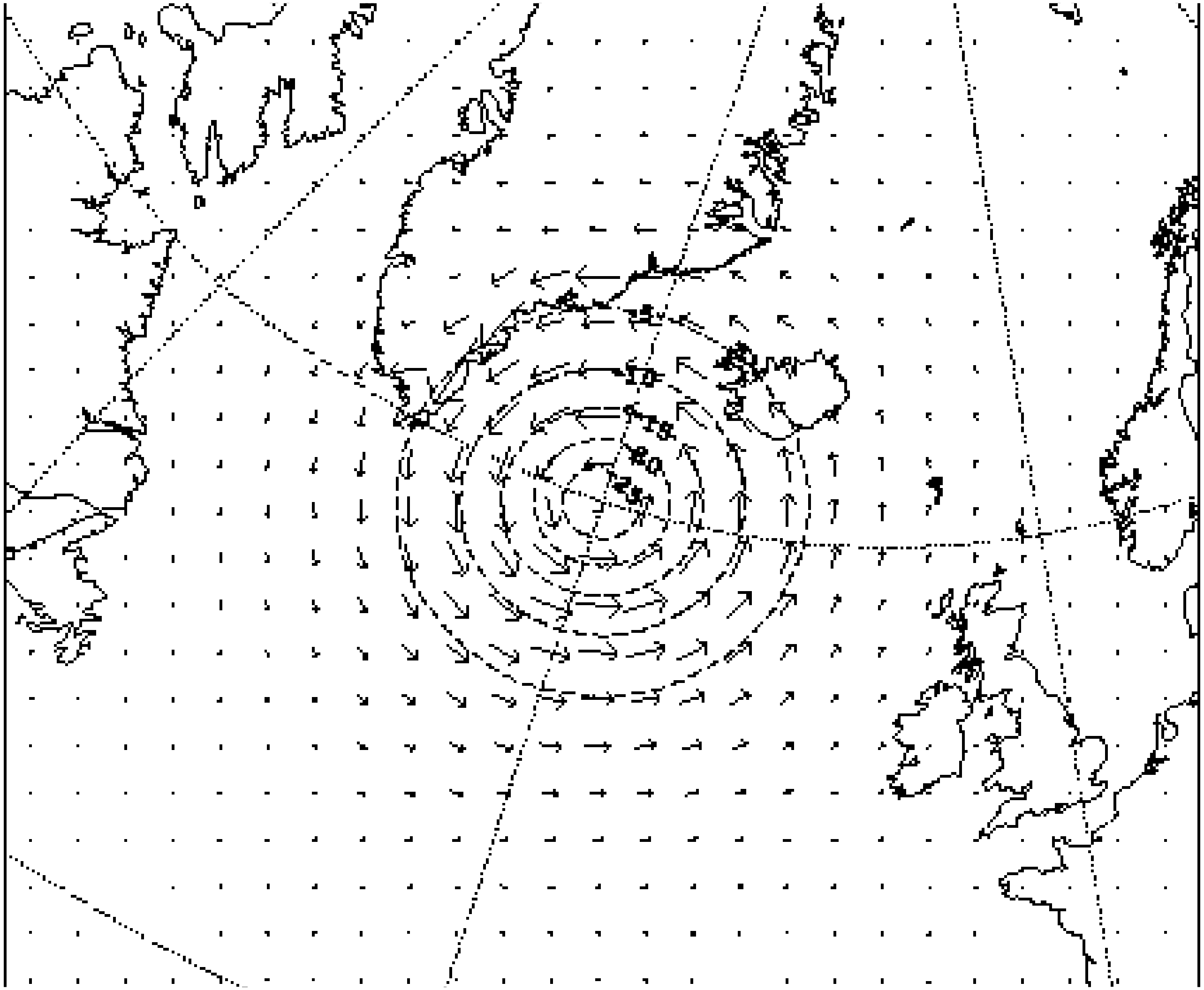


- 4D spatio-temporal correlations

  - Spatial and temporal propagation of observations.

  - Spatial and temporal coherence of trajectory.

Impact of one surface pressure observation on the pressure and wind analysis (2D)



B contains & provides information about typical scales in the atmosphere and mass/wind couplings (such as geostrophy)

Variational analysis

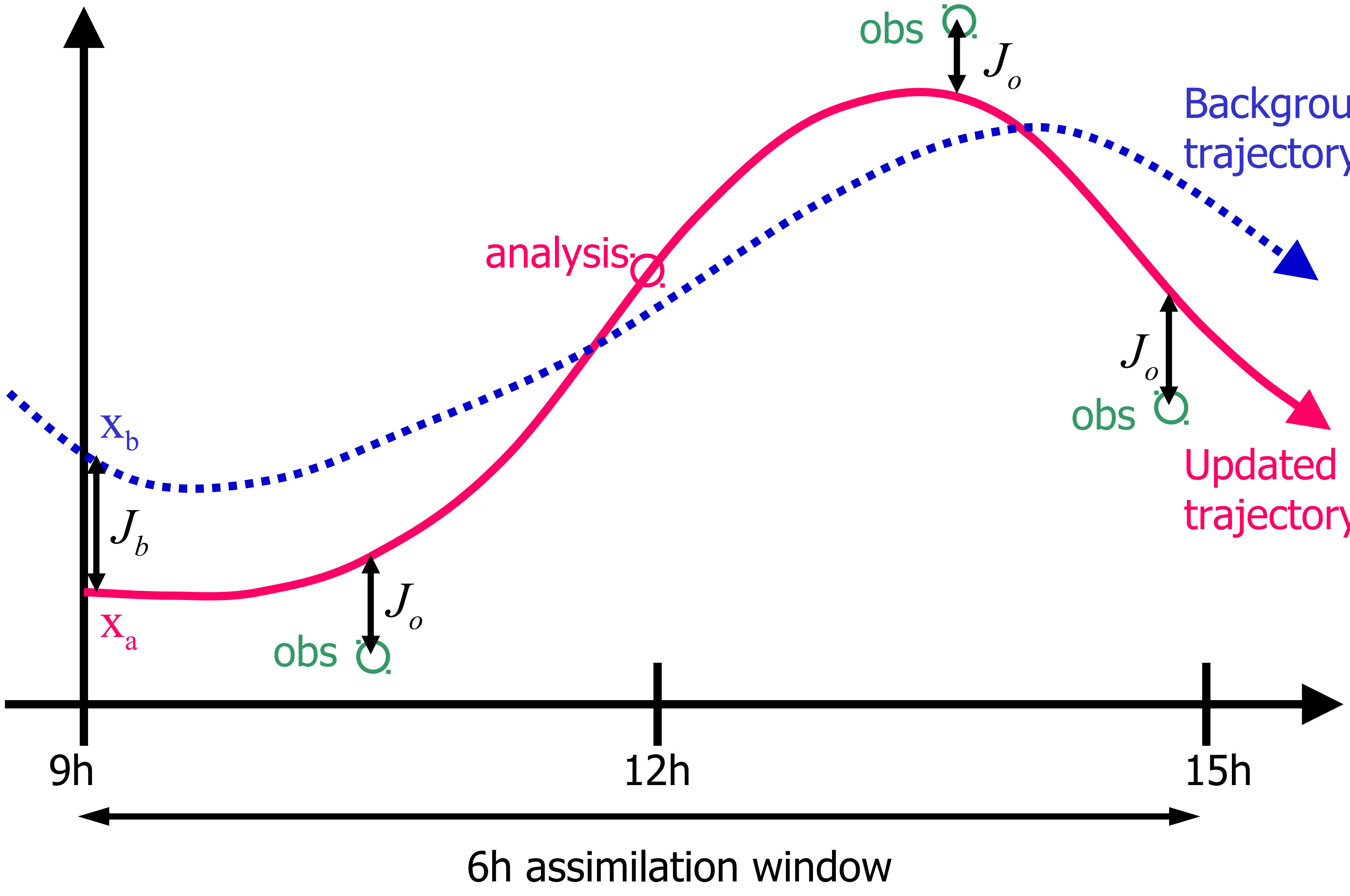
- Size of **B** is huge : square of model size ~ $(10^9)^2 = 10^{18}$
  => error covariances need to be estimated, simplified and modelled.

- The matrix ($\mathbf{H}\,\mathbf{B}\,\mathbf{H}^T + \mathbf{R}$) in $\mathbf{K} = \mathbf{B}\,\mathbf{H}^T\,(\mathbf{H}\,\mathbf{B}\,\mathbf{H}^T + \mathbf{R})^{-1}$
  is too large to be explicitly inverted.
  => minimize distance $J(\mathbf{x}^a)$ to $\mathbf{x}^b$ and $\mathbf{y}^o$ (4D-Var),
  without explicit matrix inversions (e.g. Talagrand and Courtier 1987).

- Some (weakly) non linear features are accounted for
  in calculation of departures $\mathbf{y}^o - H(\mathbf{x}^b)$ (e.g. non linear radiative transfer),
  and by updating the non linear trajectory in 4D-Var.

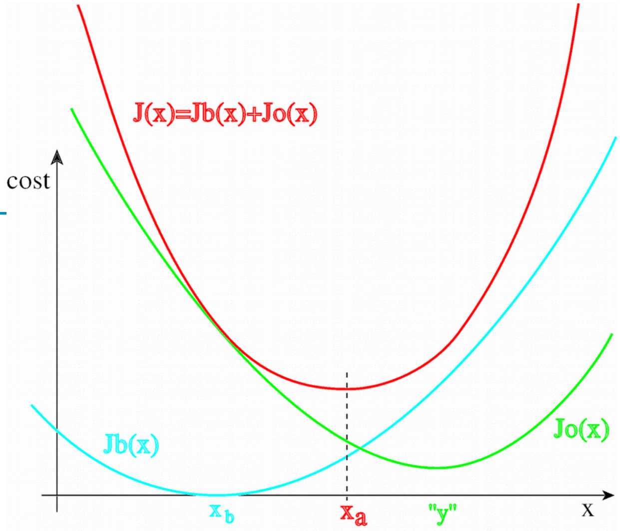
METEO FRANCE

Principle of 4D-Var assimilation



Minimisation of J=Jb+Jo allows an updated trajectory to be computed, consistent with observations at different times.

Implementation of 4D-Var



- Variational formulation :

  cost function $J(\mathbf{x}^a) = ||\mathbf{x}^a - \mathbf{x}^b||^2_{\mathbf{B}^{-1}} + ||H(\mathbf{x}^a) - \mathbf{y}^o||^2_{\mathbf{R}^{-1}}$

  minimised when gradient $J'(\mathbf{x}^a)=0$ (equivalent to BLUE).

- Computation of gradient J' : development and use of adjoint operators (i.e. transpose of tangent-linear operators).

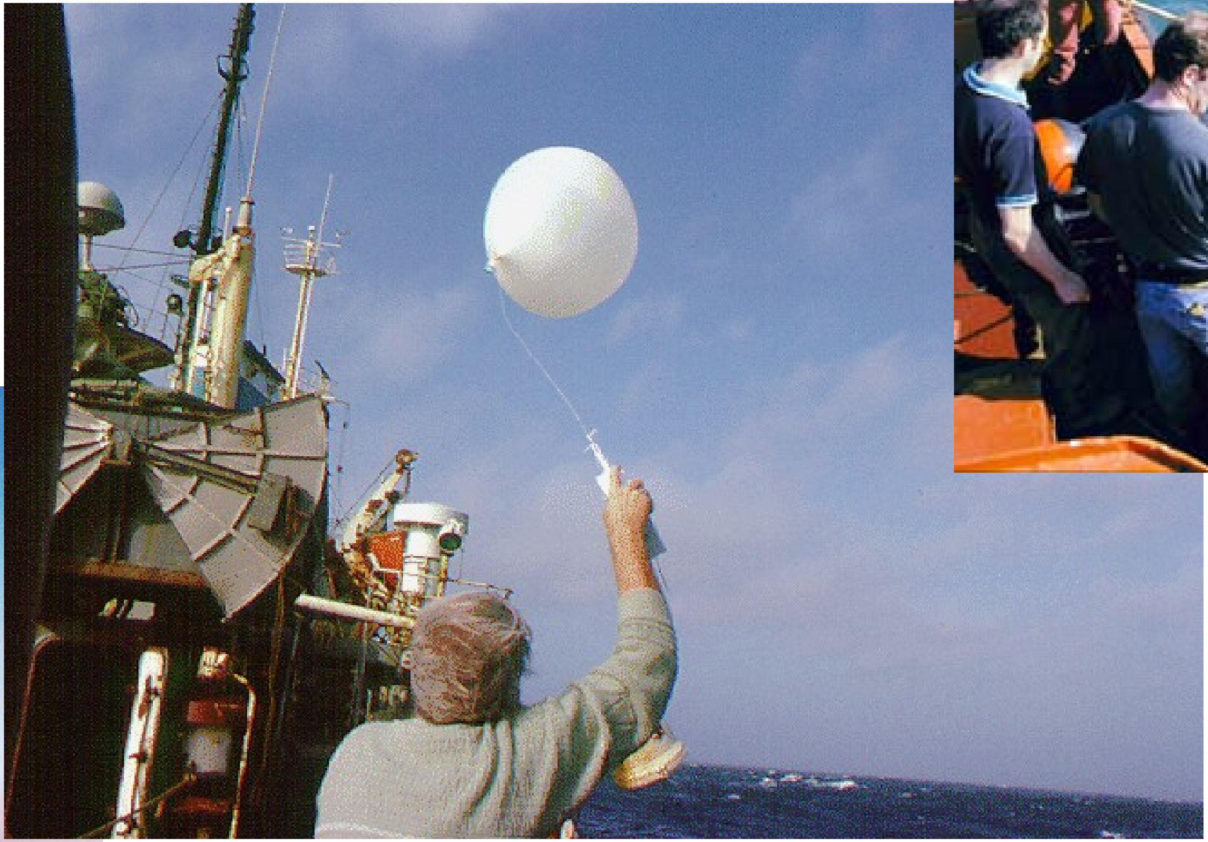
- Generalized observation operator *H* : includes NWP model *M*, in order to assimilate observations distributed in time over a 6h window.

- Reduction of computation cost : analysis increment $\delta\mathbf{x} = \mathbf{x}^a - \mathbf{x}^b$ can be computed at low resolution (Courtier et al 1994).

METEO FRANCE

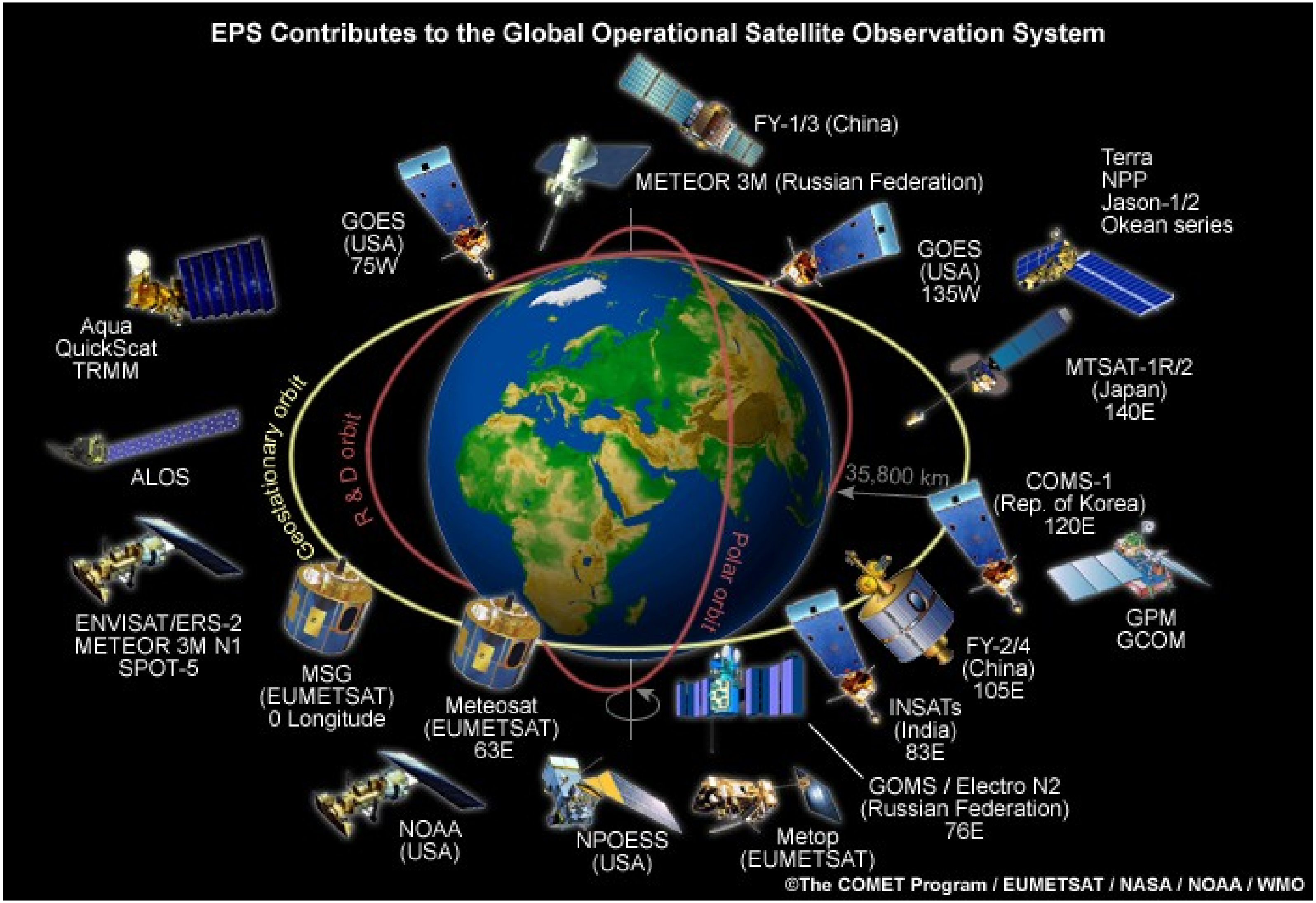
2. In-situ observations and remote sensing data

Observation networks in meteorology : in situ measurements



* Direct measurements of temperature, wind, humidity.
* Relatively easy to compare with the model, and to assimilate.
* High quality data, with relatively small biases.
* Poor horizontal coverage over the globe.

Observation networks in meteorology : satellite data

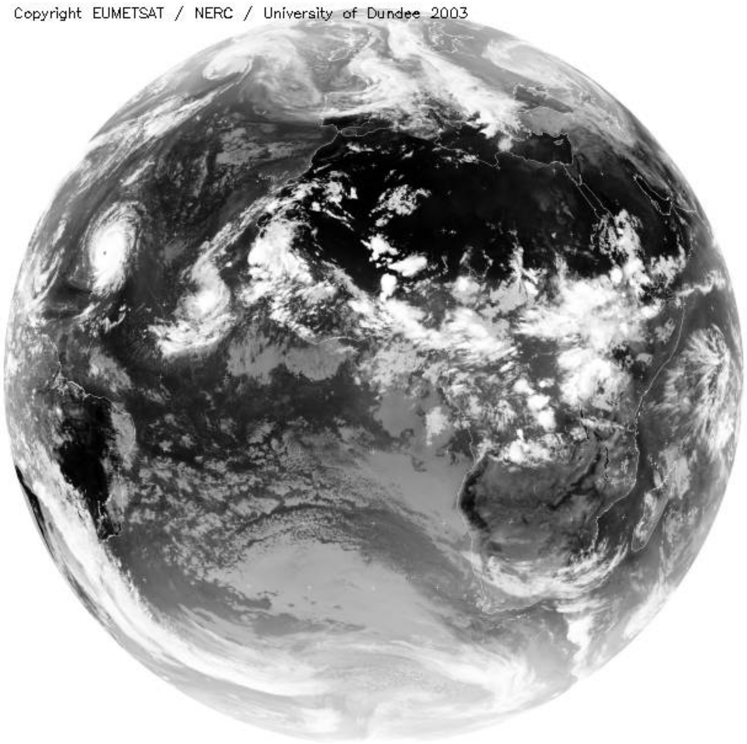


Constellation of polar orbiting or geostationary satellites

Geostationary satellites

Fix position / earth, at 36 000 km height, above equator.

Same area of the globe (disk) is always observed.



- Advantages

    Very high temporal resolution (~ 15 min).

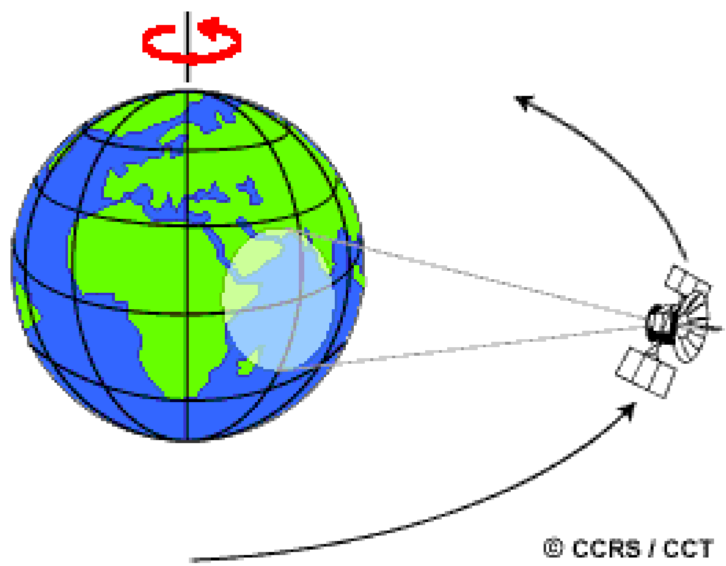
    Useful for nowcasting
    (= very short range forecasts, e.g. within the next 2 hours).

    Dynamics of meteorological structures
    (e.g. fronts, tropical cyclones).

- Drawbacks

    Insufficient spatial coverage of 1 satellite / whole globe.

    Not adapted to polar regions, due to position.



Polar orbiting satellites

Low orbit satellites (800 km height) :

- Advantages

  High spatial resolution (~10 km).

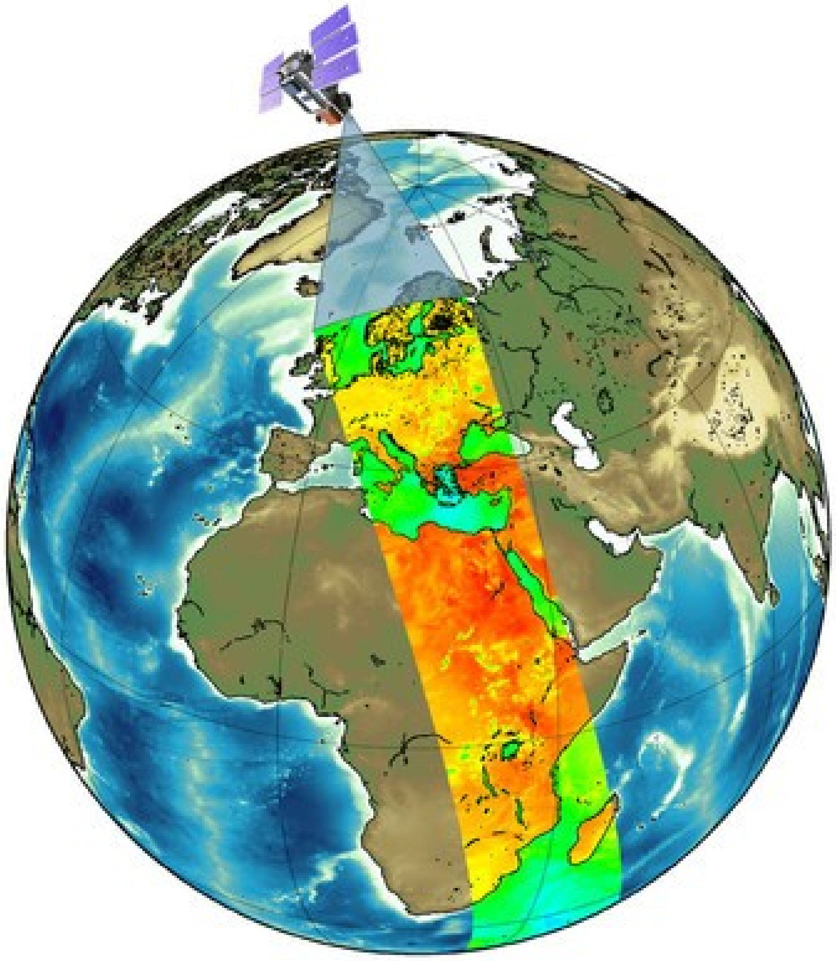
  Global spatial coverage (twice a day)

  Sounding instruments
  (over several vertical layers)

- Drawbacks

  Insufficient temporal resolution :
  a given location is only observed every 12h

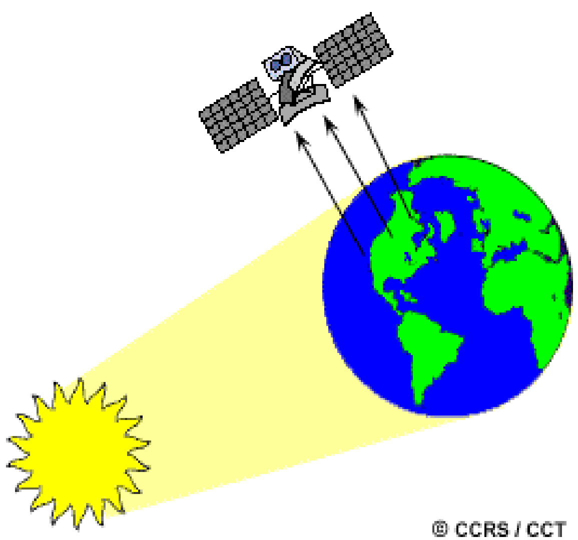
  (several satellites are needed, to have
  frequent observations over the same area)



Two types of satellite measurements

Passive measures

(no energy is emitted from instrument)

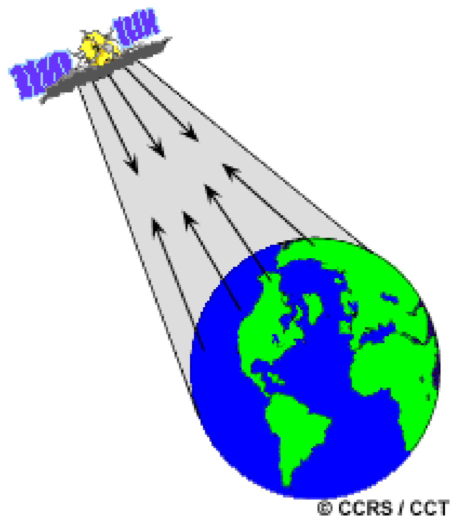


Measures natural radiation emitted

by Earth or Atmosphere (with Sun origin)

Active measures

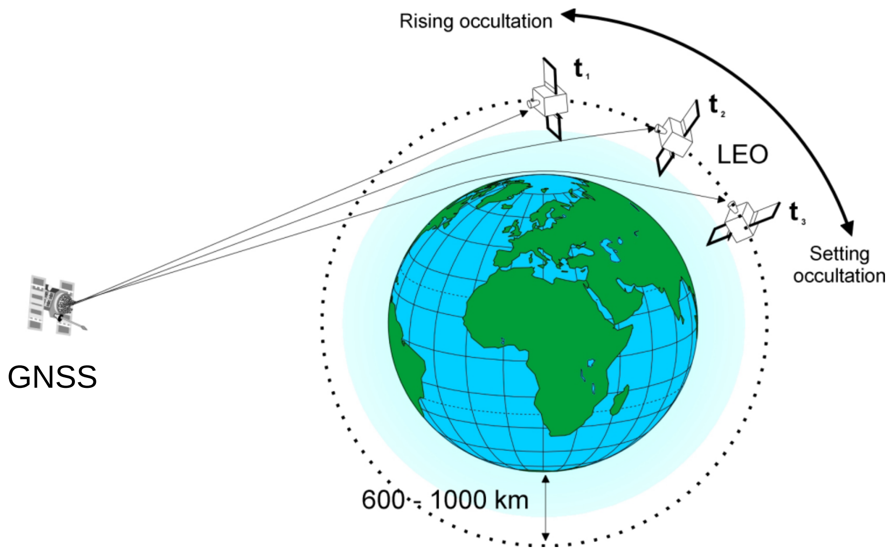
(energy is emitted from instrument)

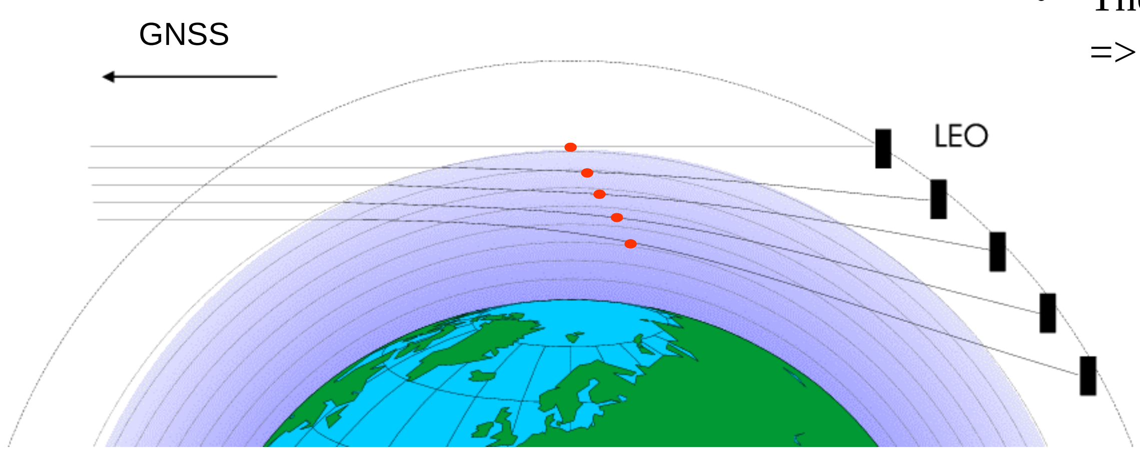


Measures radiation emitted by satellite and then

reflected or diffused by Earth or Atmosphere

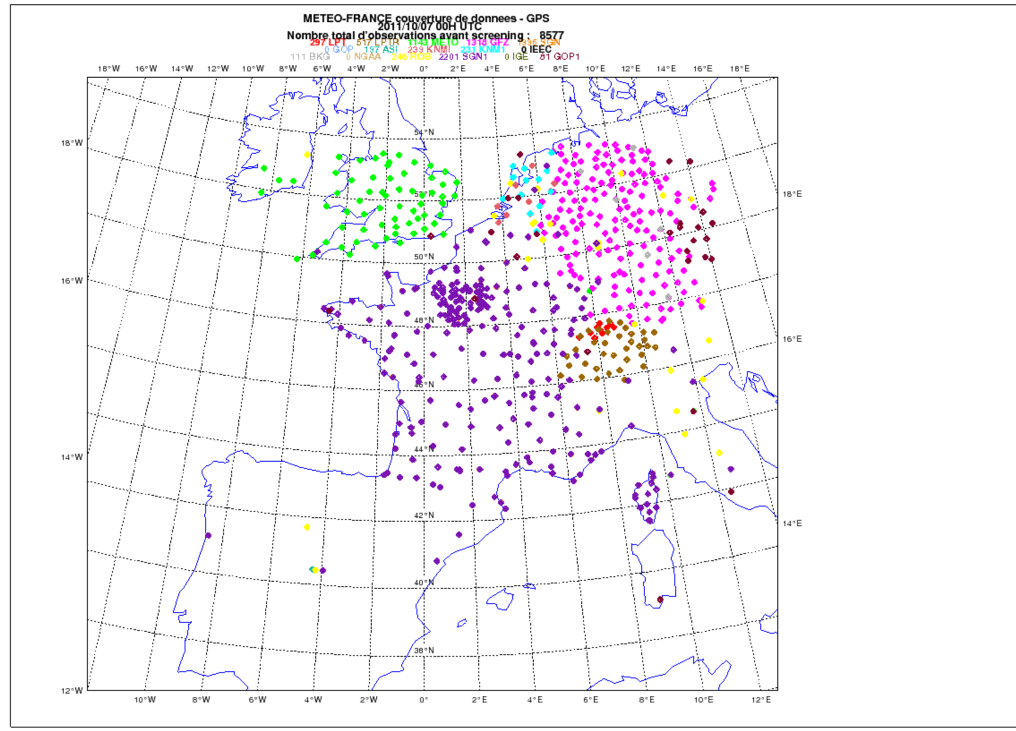
GNSS radio-occultation data (1st example of active remote sensing)





- GNSS is the Global Navigation Satellite System = GPS (USA) or Galileo (Europe).
- Low-Earth Orbit (LEO) satellites receive a signal from a GNSS satellite.
- The GNSS signal passes through the atmosphere and it gets refracted along the way.
- The magnitude of the refraction depends on temperature, moisture and pressure.
- The relative position of GNSS and LEO changes over tim
  => vertical scanning of the atmosphere,
  with information on temperature and humidity.

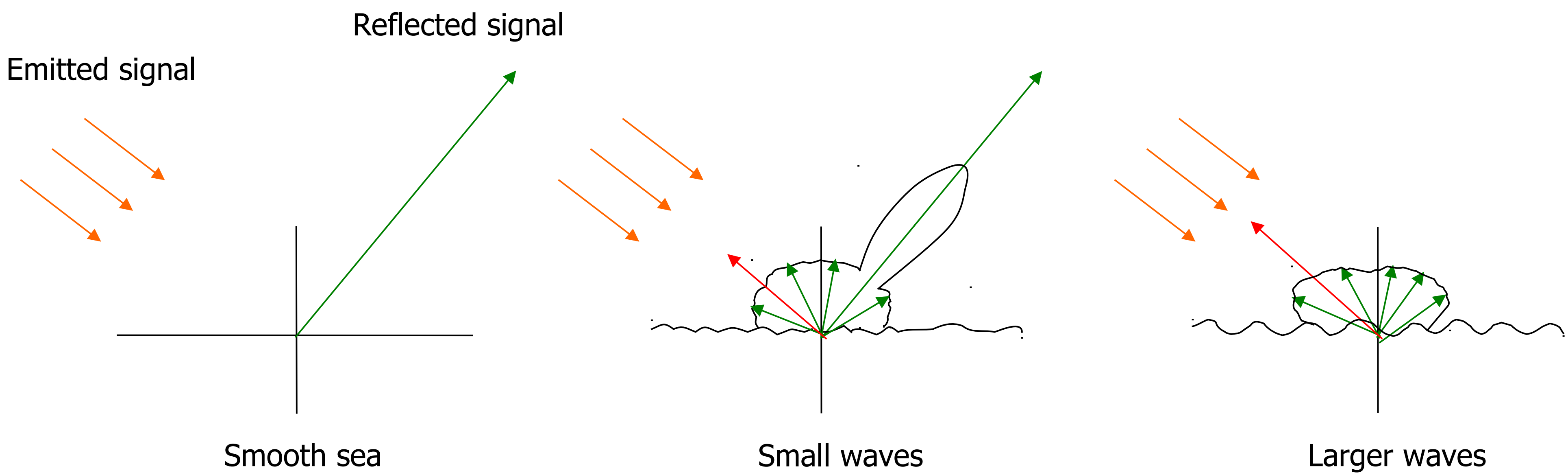
Ground-based data from GNSS (2nd example of active remote sensing)



- Propagation of GNSS signal is slowed by atmosphere (dry air and water vapour) : the propagation delay provides information about humidity in particular.
- More than 900 GNSS stations over Europe provide an estimation of Zenith Total Delay (ZTD) in real time to weather centres.
  - All weather instrument
  - High temporal resolution

Scatterometers

They send out a microwave signal towards a sea target.
The fraction of energy returned to the satellite depends on wind speed and direction.



=> Measurements of near surface wind over the ocean,
through backscattering of microwave signal reflected by waves.

Passive remote sensing :
what is measured by satellite sensors ?

- Sensors do not measure directly atmospheric temperature and humidity, but electromagnetic radiation : brightness temperature or radiance.

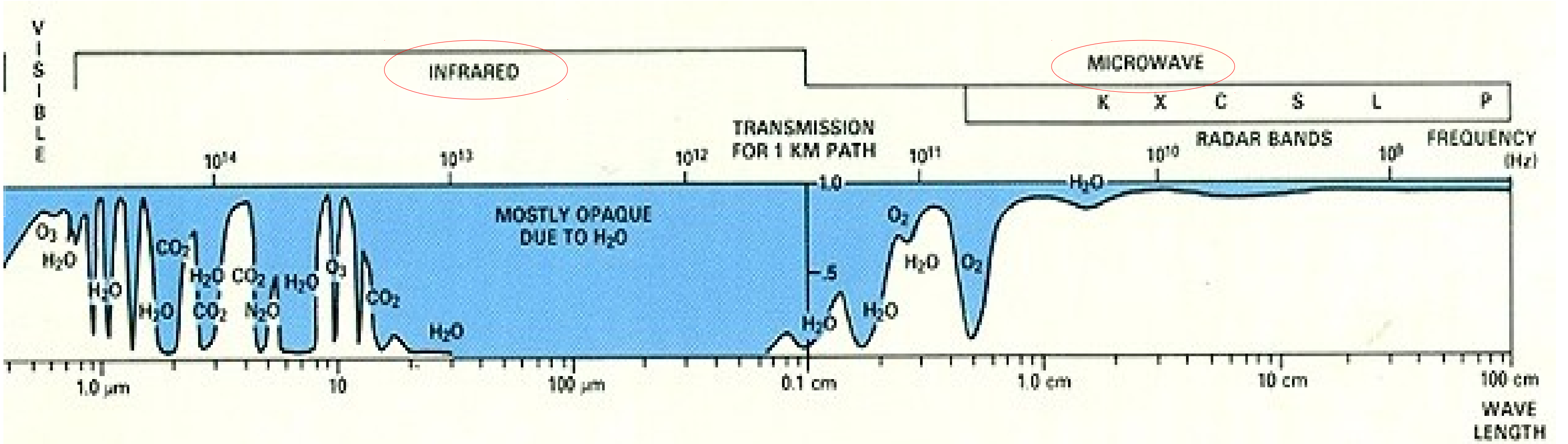
- Depending on wave length, indirect information on gas concentration (e.g. humidity) or on physical properties of atmosphere (temperature or pressure).

- Observations are made in « atmospheric windows » (in white, below) : frequencies with « low atmospheric transmittance » (= « low opacity »), e.g. in microwave and some infrared.

100 %

Atmospheric
Transmittance

0 %



Passive remote sensing : radiative transfer equation

- What is observed is a **radiance** = quantity of energy per time unit, going through a surface, in a solid angle, and for a wave number interval of the radiation.
  Unit [$W/m^2 Sr.cm^{-1}$]

- Planck function:
  $B_\upsilon(T)$= radiance emitted by a black body at temperature T, for wave number υ.

- Intensity of the radiation, emitted by the atmosphere at wave number υ:

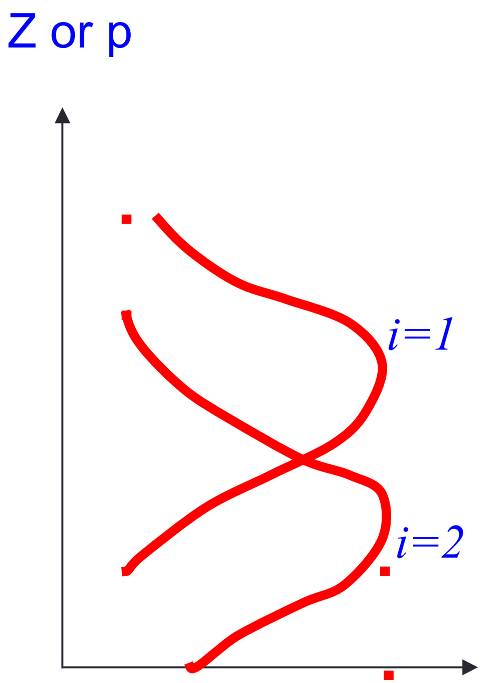
$$R_\upsilon = (I_0)_\upsilon \tau_\upsilon(z_0) + \int_{z0} B_\upsilon(T(z)) \, (d\tau_\upsilon(z)/dz) \, dz$$

$(I_0)_\upsilon$ is the surface emission at altitude z_0.

$\tau_\upsilon(z)$ is the *transmittance* from z to the top of the atmosphere ;
it accounts for atmospheric absorption of radiation.

$K_\upsilon(z) = d\tau_\upsilon(z)/dz$ is called *weighting function* :
it weights the Planck function in the radiance equation, and it determines the vertical layer of the atmosphere that is sounded at this frequency.

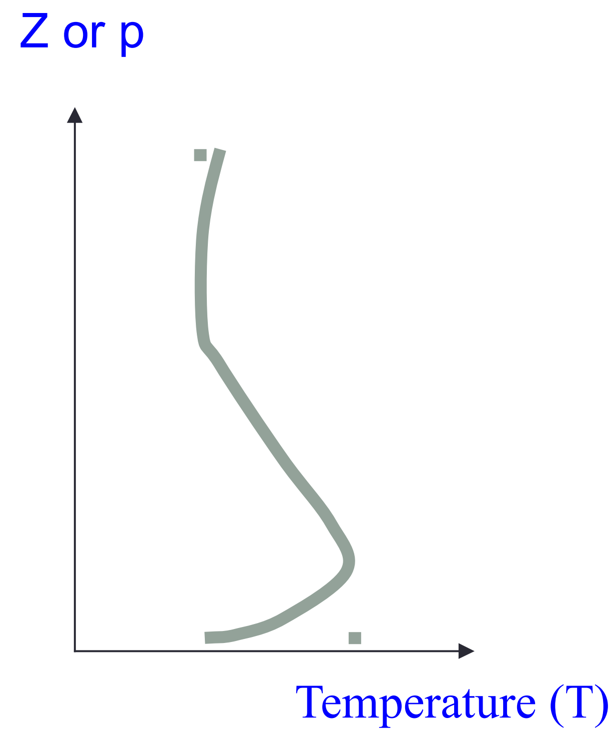
Retrieval of temperature vertical profiles



Contributions to radiances R_i
(for a given channel *i*, with specific weighting function)

Radiative transfer
(observation operator)

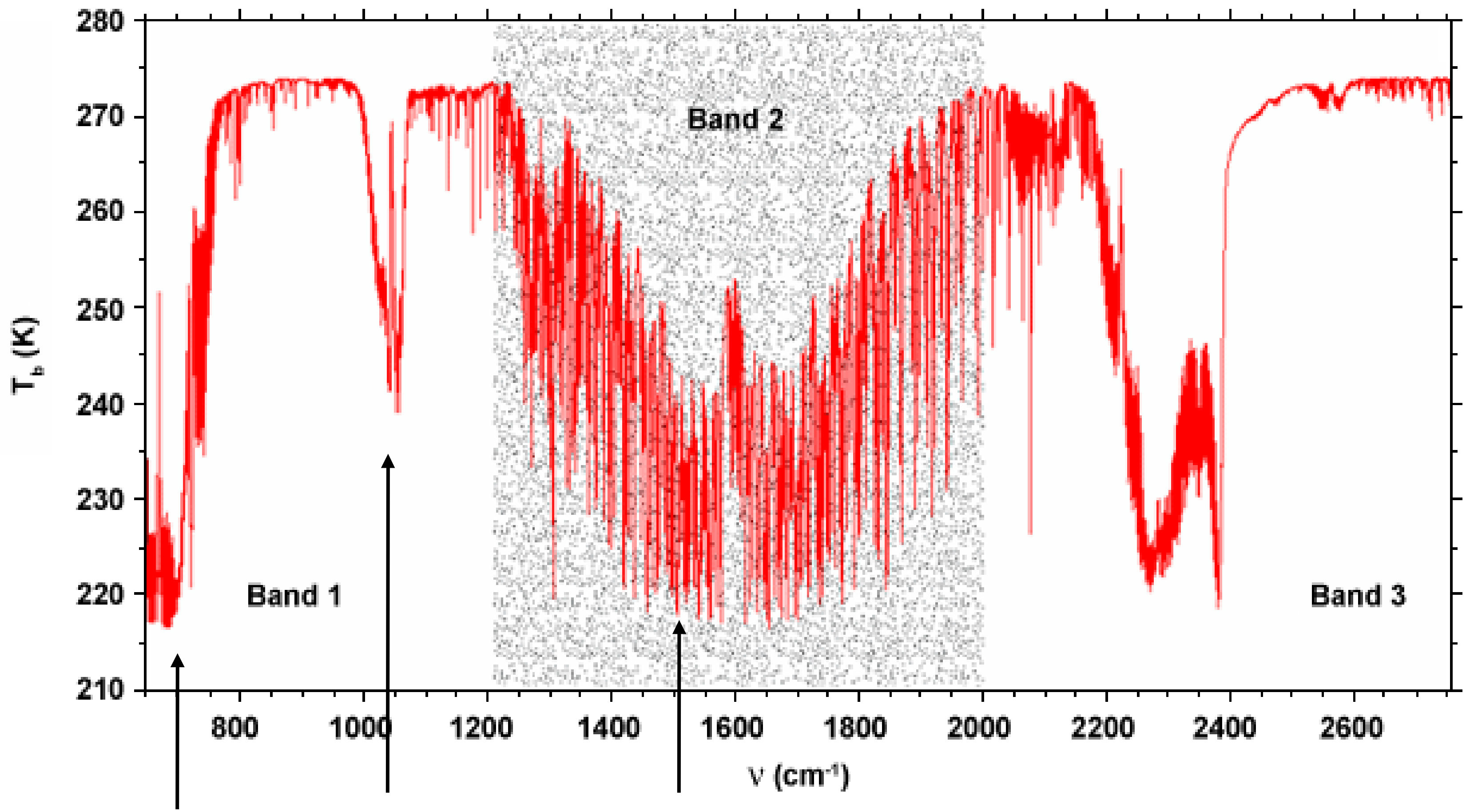
Inversion
(e.g. through DA)



$$R_i \;=\; B_i[T(p_0)].\tau_i(p_0) + \int_p B_i[T(p)].[d\tau_i(p)/dz].dz$$

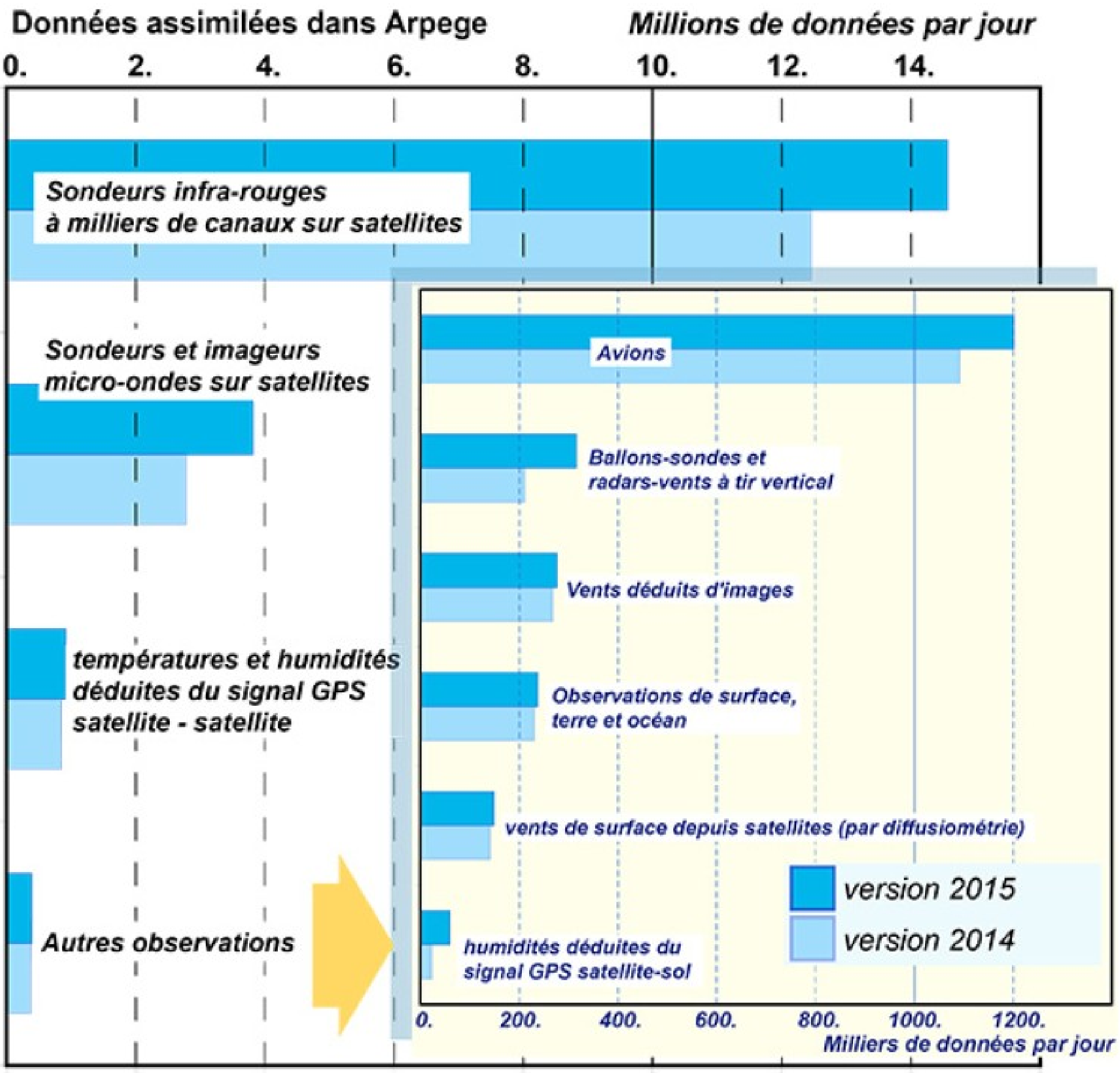
IASI : infra-red interforemeter developed by CNES and EUMETSAT

IASI offers a very high spectral resolution (~ 8000 channels)



Temperature ozone Water vapor

Number of observations used in ARPEGE (global DA at Météo-France)



Total ~ 20 million obs per day

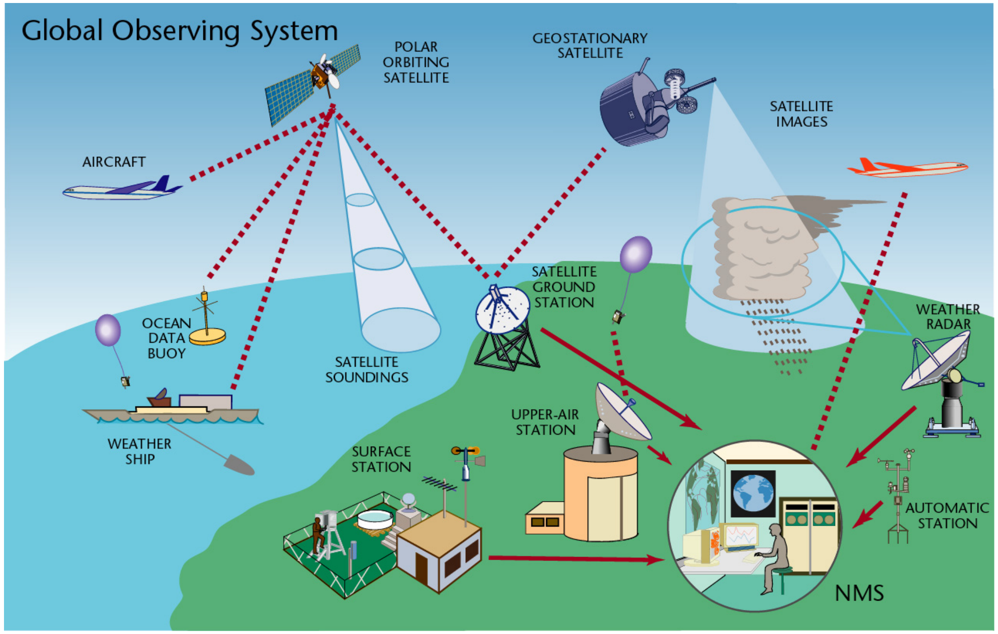
How do observations meet global NWP requirements ?

- **Surface observations**
  good coverage over land, sparse coverage over sea;
  observations not suited to describe upper levels.

- **Aircraft observations**
  good accuracy,
  but do not describe the 3D state of the atmosphere
  (except near airports).

- **Radiosonde data**
  good accuracy, good vertical resolution,
  but poor horizontal coverage over the globe.

- **Satellite data**
  good horizontal coverage over the globe,
  but poor vertical resolution (reduced to 1 level for satellite winds or imagers).



METEO FRANCE

Radar network in France

- 30 radars (19 C-band, 5S, 6X)

  every 15 minutes, at 1 km resolution.

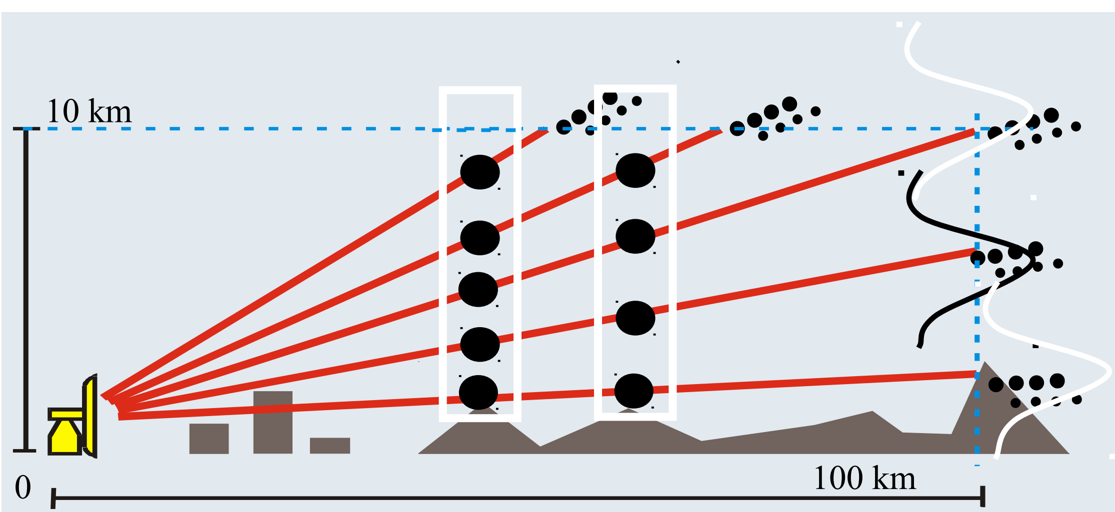
- Observations :

reflectivities Z (related to precipitation) ;

radial winds Vr (doppler effect) :

the emitted microwave signal returns to the radar with a modified frequency, when the target is moving.

=> invert Doppler equation to obtain a wind observation.



Observations assimilated
as vertical profiles,
after estimating the pixel altitude

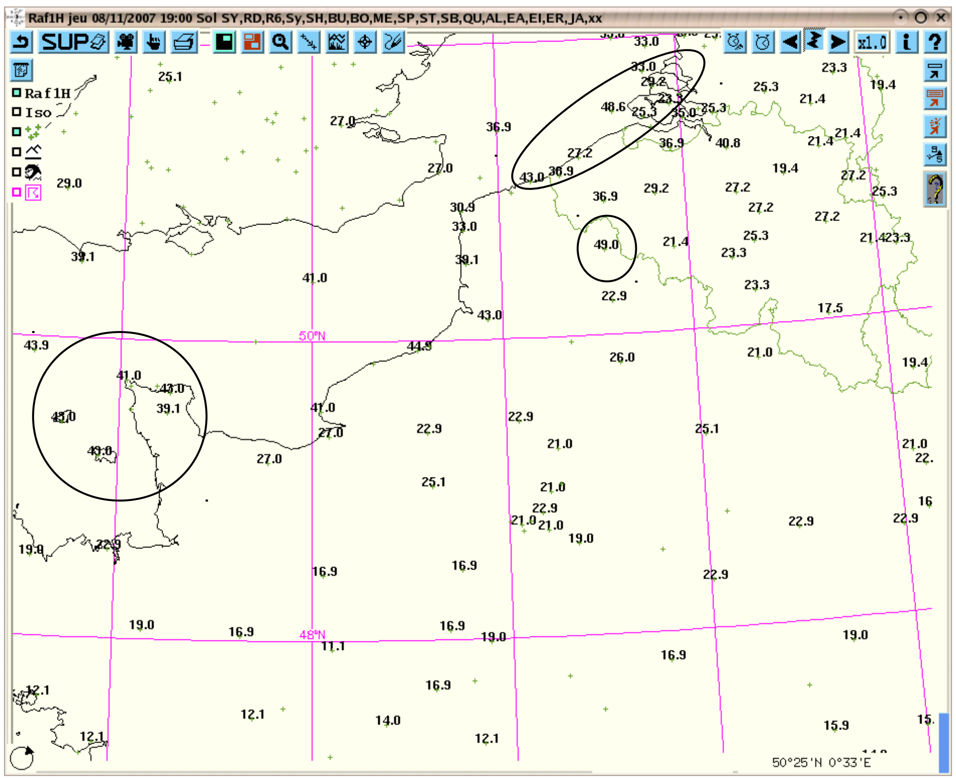
(Pixel altitude is computed using
a constant refractivity index along the path)
(= effective radius approximation)

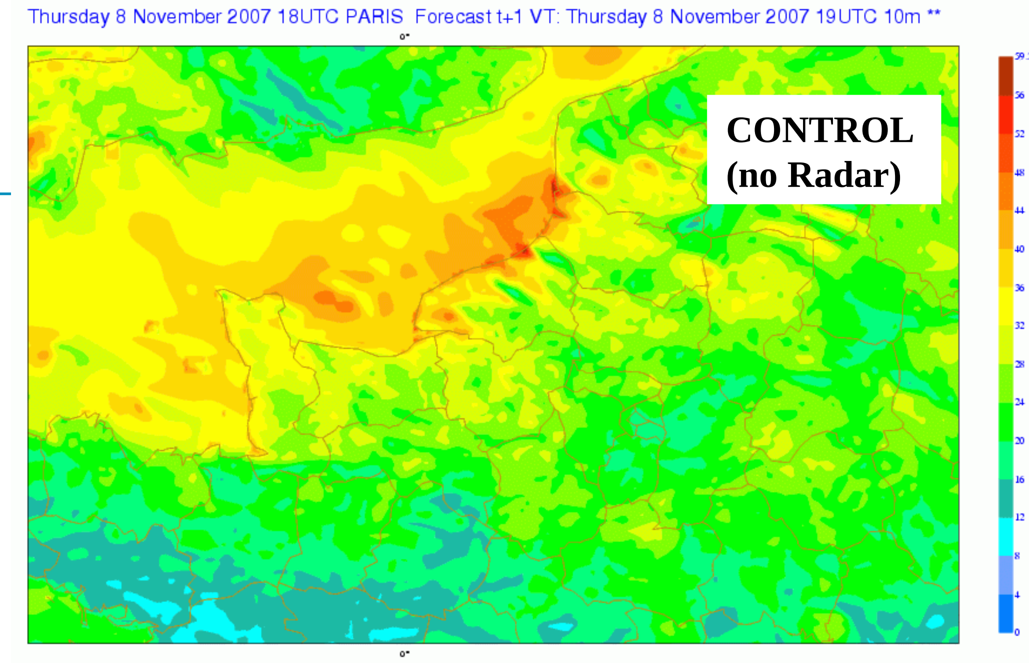
Assimilation of radar radial winds

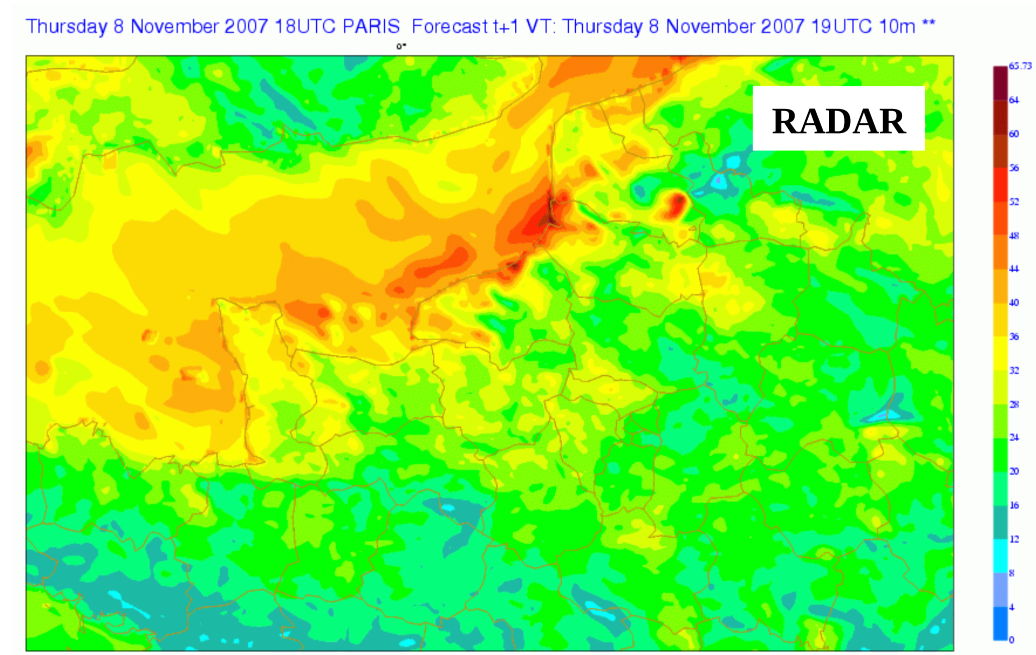
Wind gust at 10 m (kt)

Forecast +1h (19 UTC)

OBS







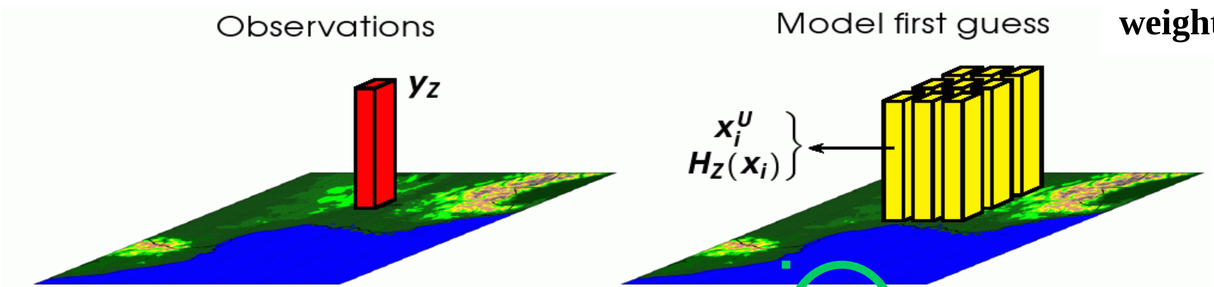
Bayesian inversion of reflectivity profiles

1. **Compute model reflectivities** from model relative humidity (RH) profiles (using observation operator for relectivities).

2. **Estimate likelihoods of model reflectiv.**, by comparing them with obs. reflectivities and by computing associated exponantials.

3. Compute (pseudo-)observed RH profile, to be assimilated : **average of model RH profiles, weighted by their likelihoods**.

Caumont, 2006: use model profiles in the neighborhood of each observation (in 3 steps)



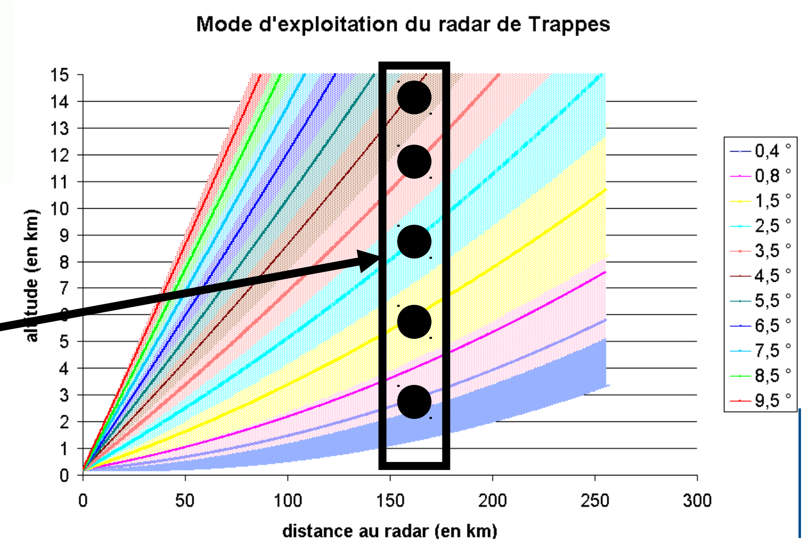
$$y_{po}^U = \sum_{\substack{i \in \\ \text{neighbours}}} x_i^U \frac{\exp\left(-\frac{1}{2}\|y_Z - H_Z(x_i)\|^2\right)}{\sum_{\substack{j \in \\ \text{neighbours}}} \exp\left(-\frac{1}{2}\|y_Z - H_Z(x_j)\|^2\right)}$$

Observation operator for reflectivities

y_{po}^U: column of pseudo-observed relative humidity,
y_Z: column of observed reflectivities,
x_i^U: column of relative humidity,
$H_Z(x_i)$: column of simulated reflectivities.

The pseudo-observed RH profile can then be assimilated as a conventional profile (like a radiosonde profile).

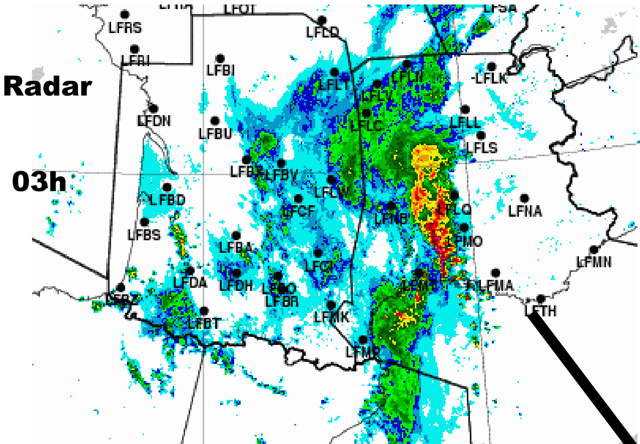
y_{po}^U

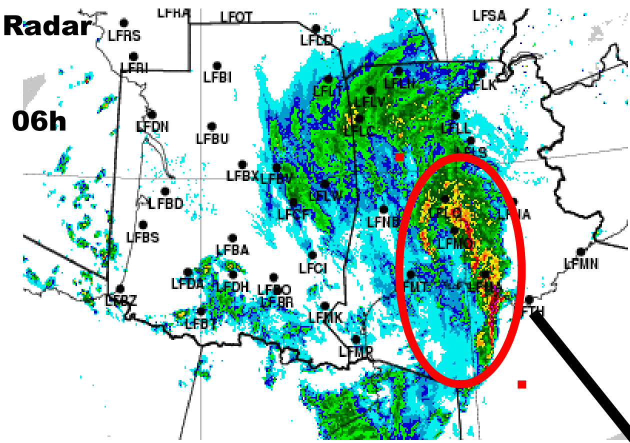


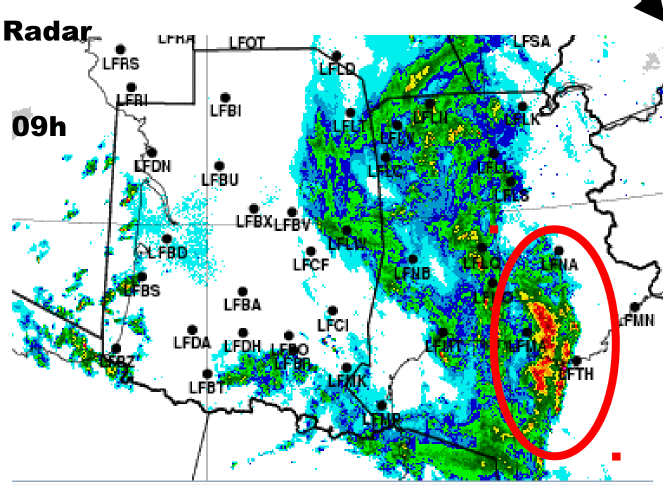
Impact of radar reflectivities

Example of High Precipitation Event (South-East of France) :

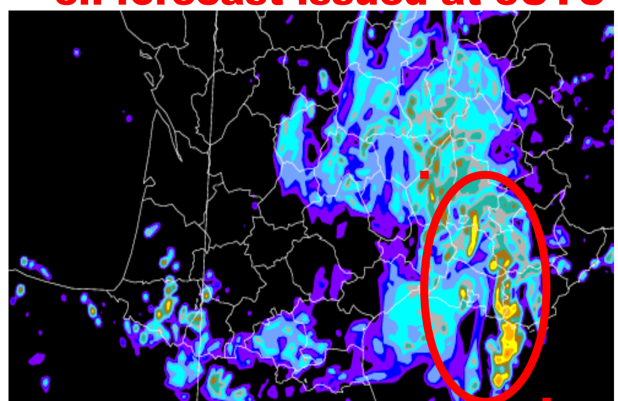
REFL (assimilation of reflectivities) vs **CONTROL** (no reflectivities in DA)



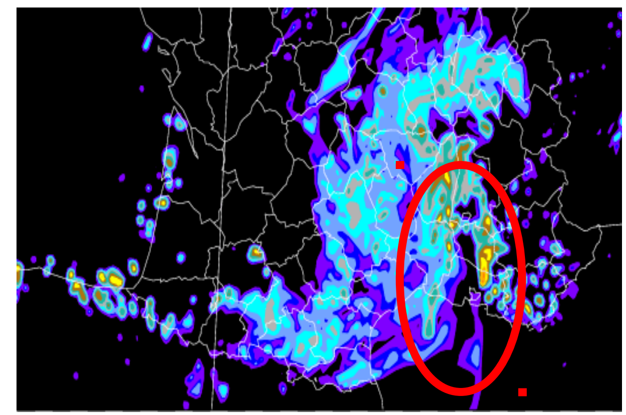




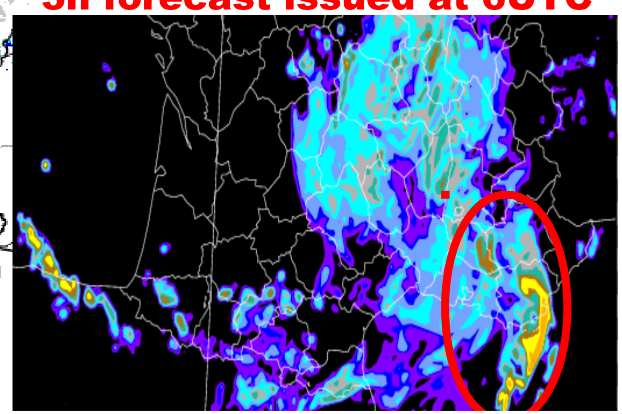
3h forecast issued at 3UTC



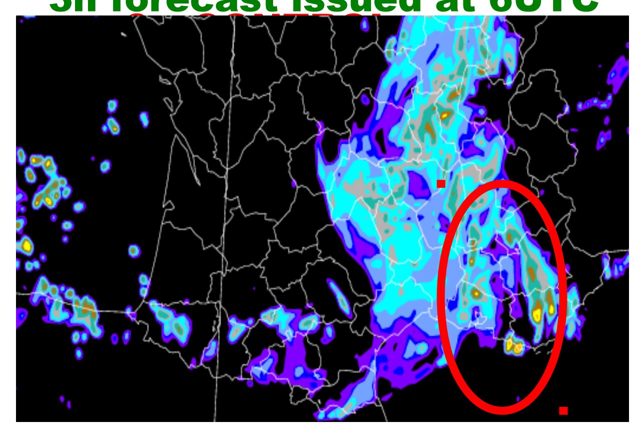
3h forecast issued at 3UTC



3h forecast issued at 6UTC



3h forecast issued at 6UTC



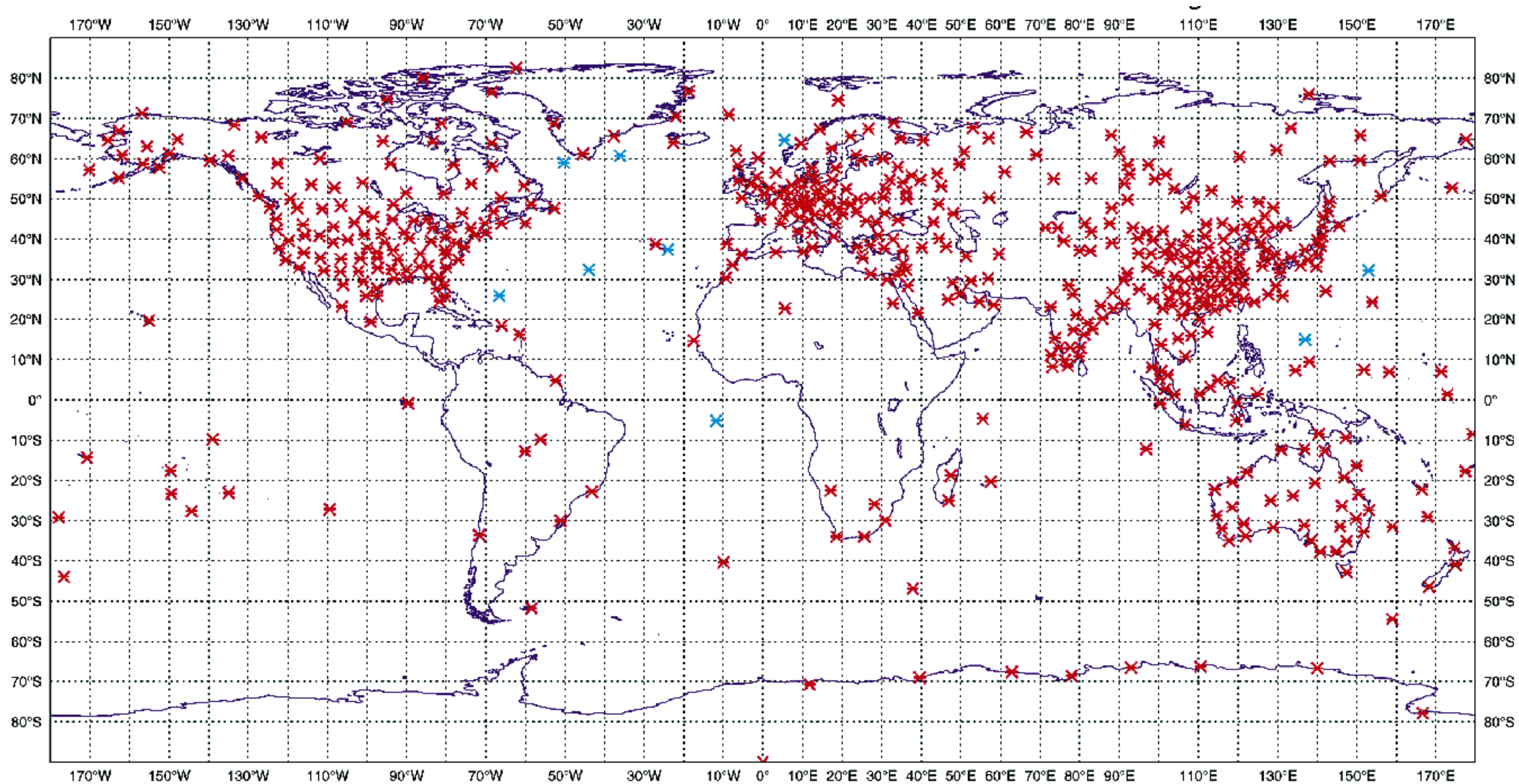
Line of heavy precipitation is well analysed in REFL run.

Error covariances : estimation and modelling

How can we estimate error covariances ?

- The **true atmospheric state** is never (exactly) known.
- Use **observation-minus-background** departures
  to estimate some average variances and correlations of **R** and **B**, using assumptions on spatial structures of errors.
- Use an **ensemble** to simulate the error evolution and to estimate space- and time-dependent background error structures.
- Use **covariance modelling** to filter out sampling noise and other uncertainties in the ensemble.

Radiosonde observation network



Covariances of innovations

- Innovations = observation-background departures :

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

$$\sim \mathbf{e}^o - \mathbf{H}\mathbf{e}^b$$

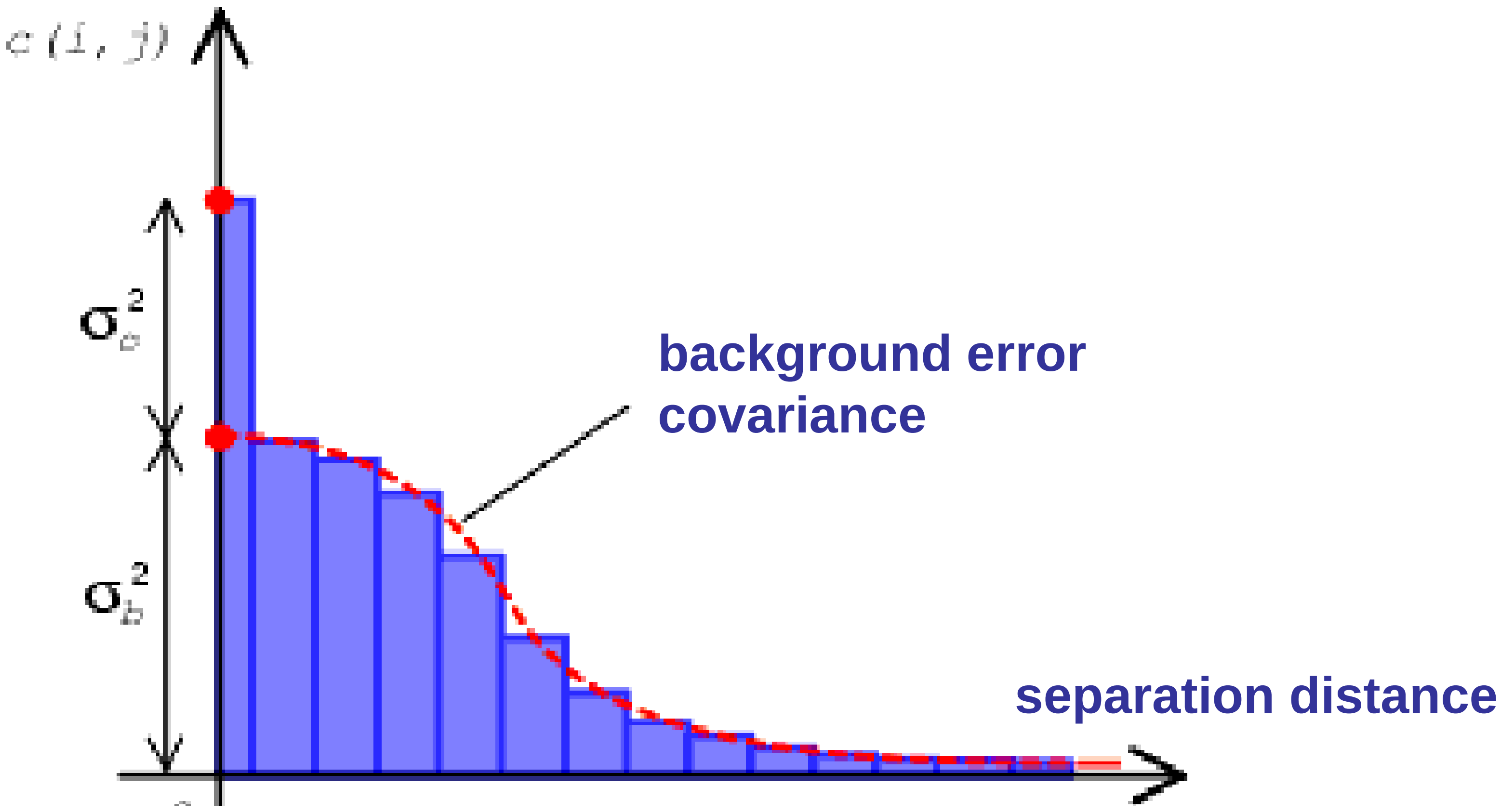
- Innovation covariances :

$$\mathrm{E}[(\mathbf{y}^o - H(\mathbf{x}^b))(\mathbf{y}^o - H(\mathbf{x}^b))^{\mathrm{T}}] = \mathbf{R} + \mathbf{H}\,\mathbf{B}\,\mathbf{H}^{\mathbf{T}}$$

assuming that $\mathrm{E}[\,\mathbf{e}^o\,(\mathbf{H}\mathbf{e}^b)^{\mathrm{T}}\,] = \mathbf{0}$.

(e.g. Hollingsworth and Lönnberg 1986).

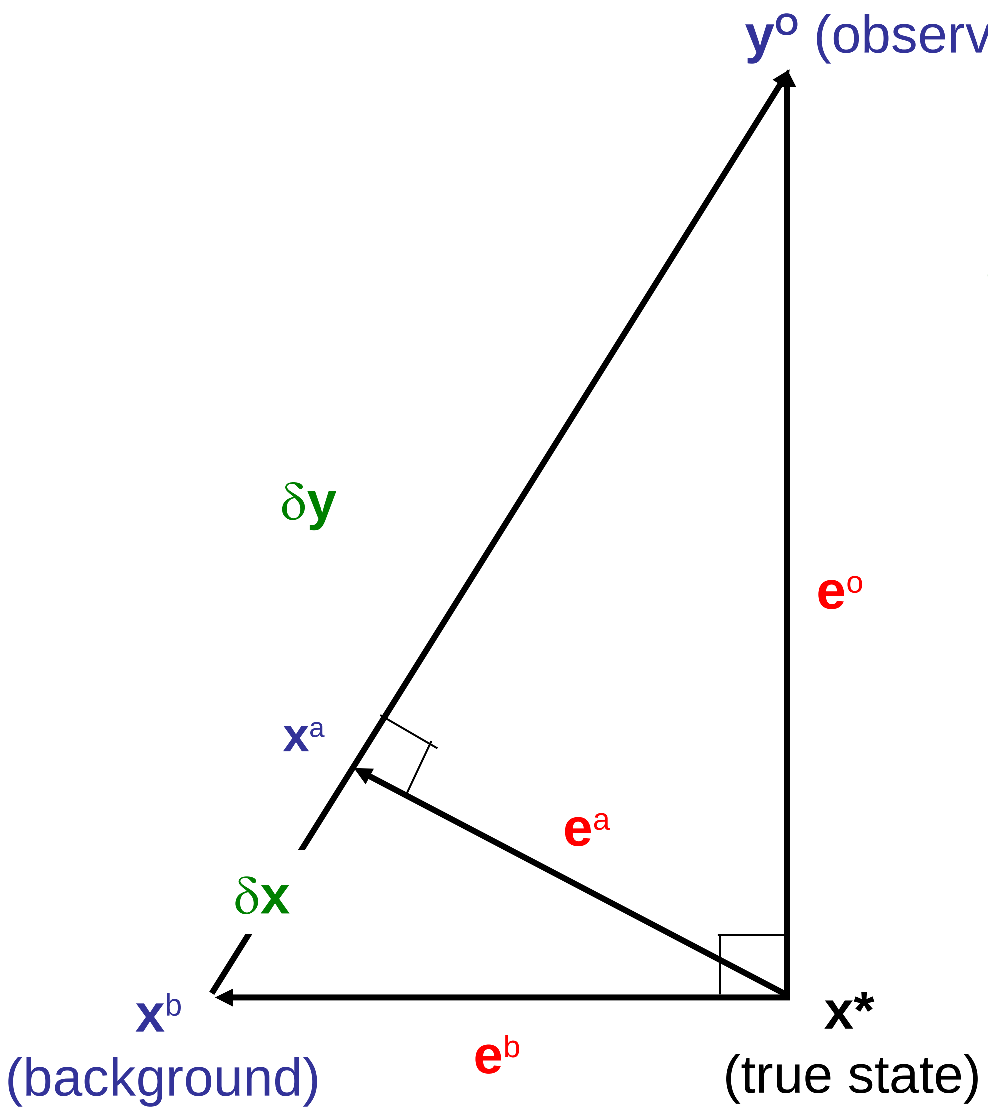
Covariances of innovations (with extrapolation to zero separation distance)



$$E[(\mathbf{y}^o - H(\mathbf{x}^b))(\mathbf{y}^o - H(\mathbf{x}^b))^T] = \mathbf{R} + \mathbf{H}\ \mathbf{B}\ \mathbf{H}^\mathbf{T}$$

Extrapolation to zero separation distance allows different contributions to be estimated.

Covariances of analysis residuals



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

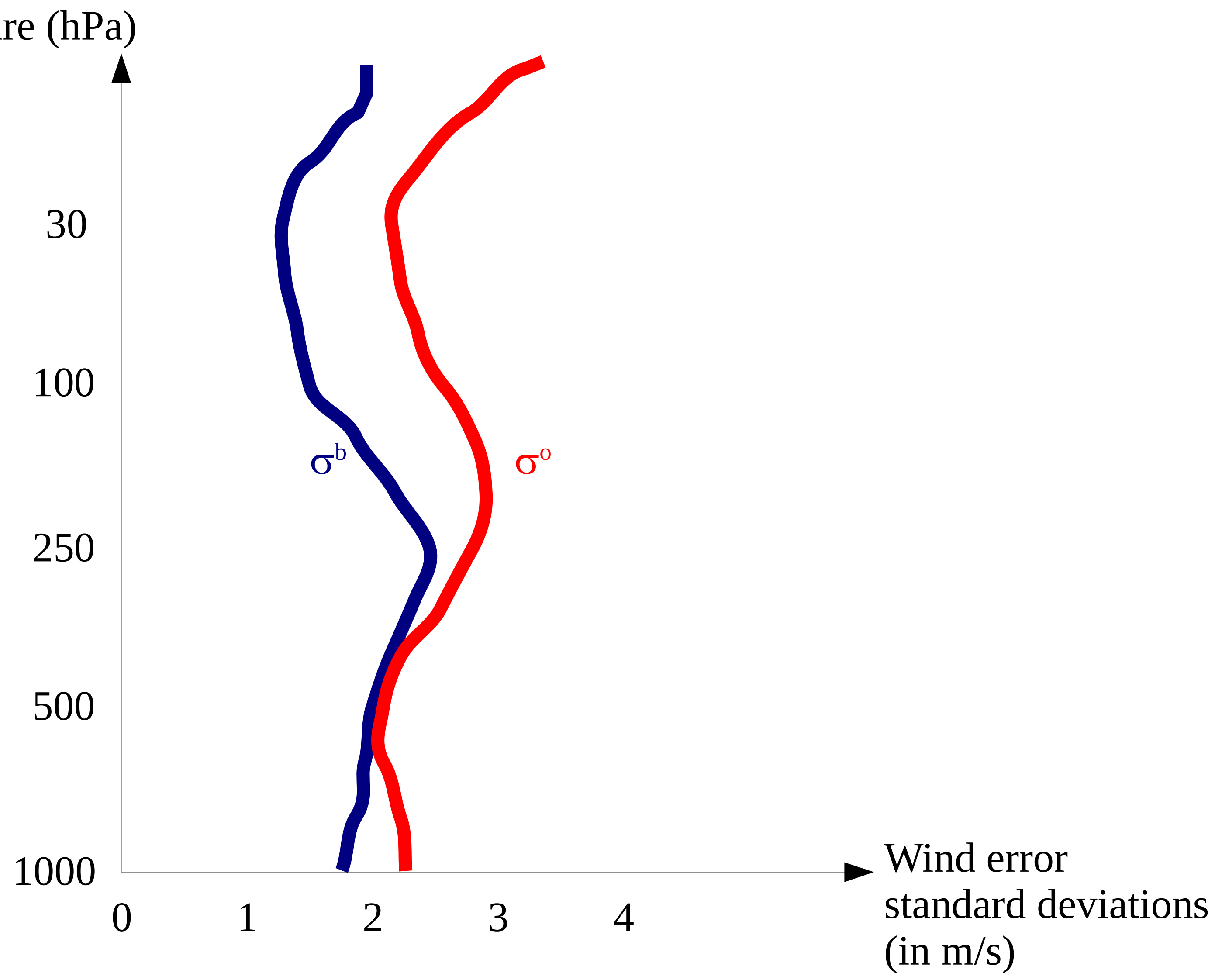
$\mathbf{H}\,\delta\mathbf{x} = H(\mathbf{x}^a) - H(\mathbf{x}^b)$ (increment)

$E[\mathbf{H}\,\delta\mathbf{x}\,\delta\mathbf{y}^T] = \mathbf{HBH}^T$

$E[(\mathbf{y}^o - H(\mathbf{x}^a))\,\delta\mathbf{y}^T] = \mathbf{R}$

(Desroziers et al 2005)

Vertical profiles of standard deviations of background errors and observation errors



Space & time averages of innovation-based covariances

$H(\mathbf{x}^b)$

$\delta \mathbf{y} \sim \mathbf{e}^o - \mathbf{H}\mathbf{e}^b$

$\mathbf{y}^o$

At a given location and time,

there is only 1 innovation value $\delta \mathbf{y}$:

a single error realization is available locally

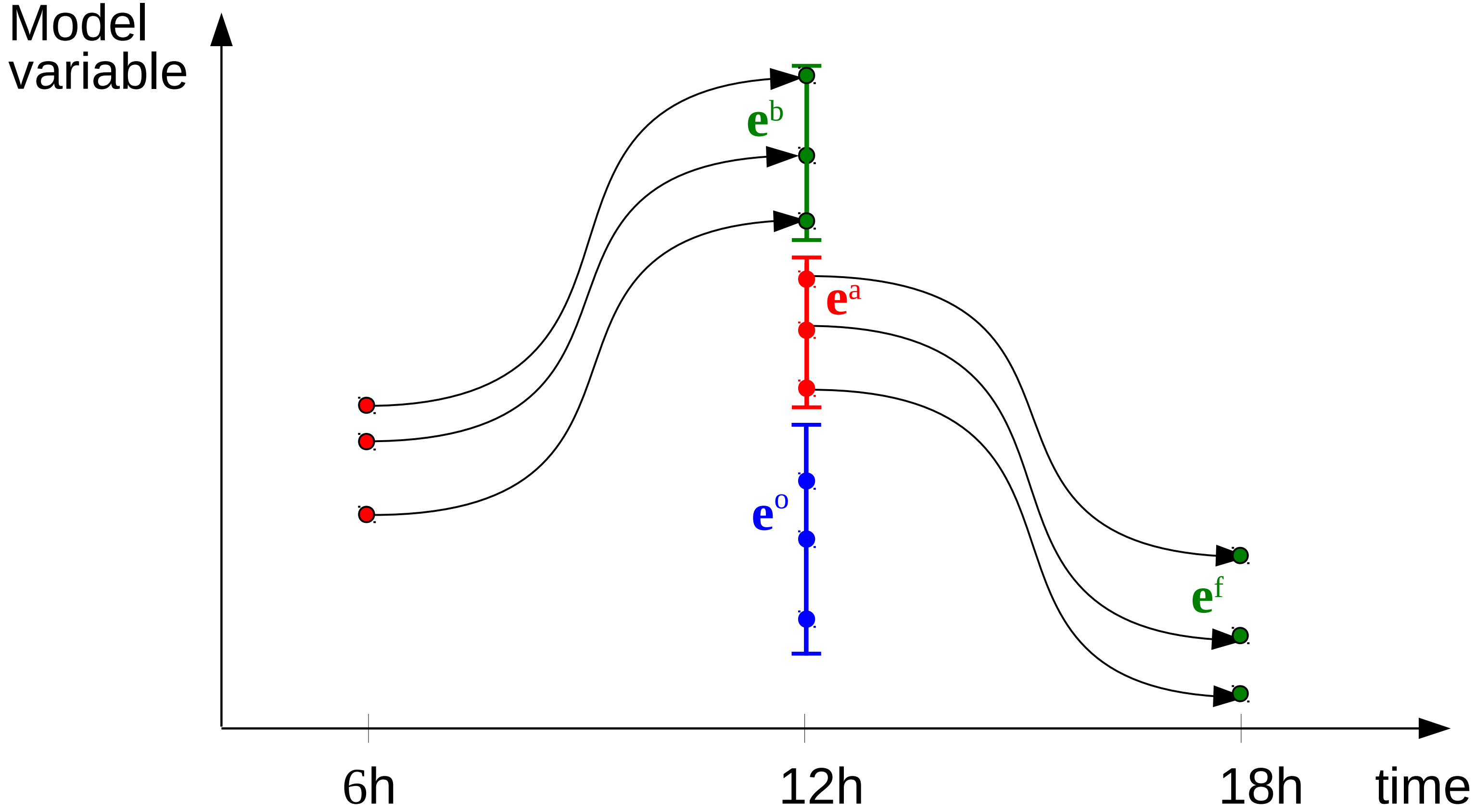
(e.g. for estimating background errors).

=> Statistical averages (mathematical expectations)
need to be replaced by
space and time averages (ergodic assumption).

=> only space or time averages of **B** and **R**
can be estimated from innovation data.

=> consider other approaches, such as ensemble methods.

Ensemble Data Assimilation (EDA) : simulation of error cycling



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

$\mathbf{e}^f = \mathbf{M}\mathbf{e}^a + \mathbf{e}^m$

with $\mathbf{e}^b = \mathbf{e}^{f-}$

and $\mathbf{e}^o = \mathbf{R}^{1/2}\, \eta$ (random draws of **R**)

(e.g. Houtekamer et al 1996, Fisher 2003, Berre et al 2006 ;
ARPEGE : 50 members to estimate flow-dependent **B**)

Simulation and propagation of observation errors and model errors

- Observation errors can be simulated by

  adding random draws of **R** : $\mathbf{e^o = R}^{1/2}\, \eta^o$.

- Model errors can be simulated by

  adding random draws of **Q** : $\mathbf{e^m = Q}^{1/2}\, \eta^m$ (additive or mult. inflation);

  using a multi-model approach (or multi-physics) ;

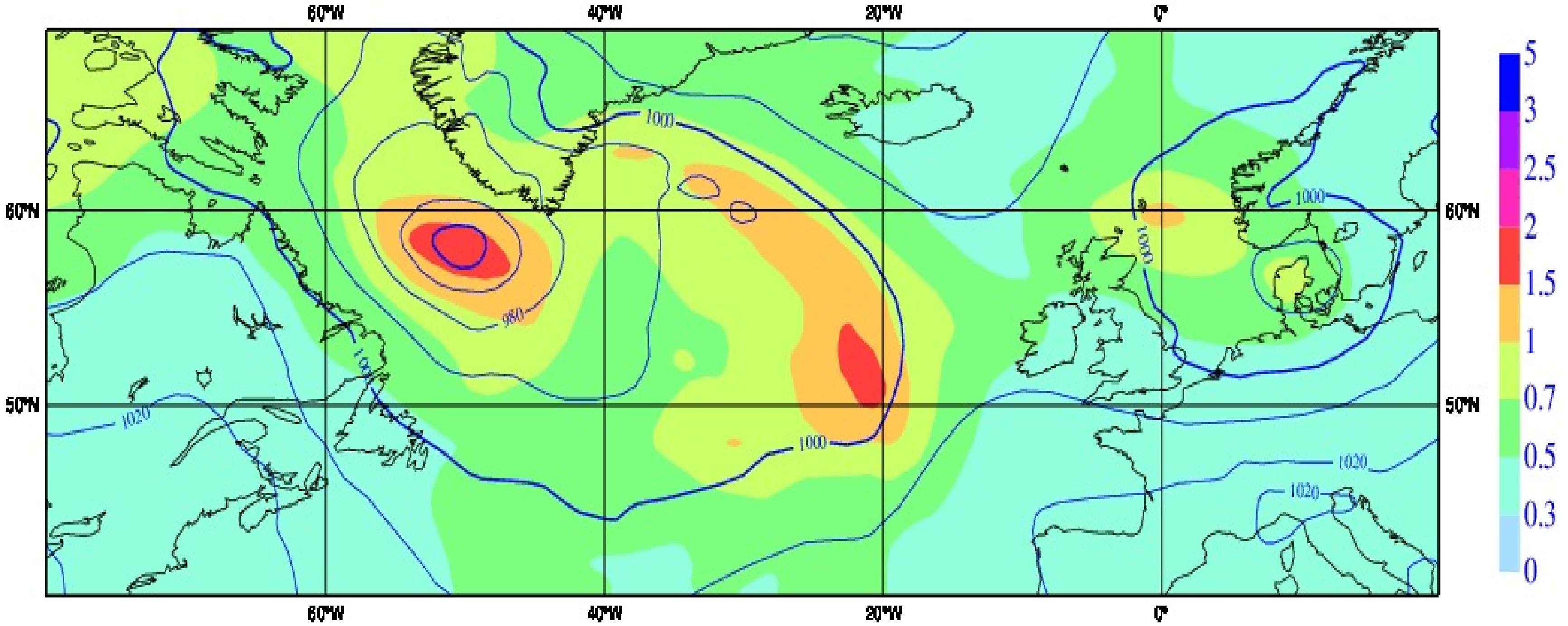
  perturbing physical tendencies of the model ;

  perturbing model parameters. (...)

- Observation and model perturbations are propagated during the successive analysis/forecast steps of DA cycling.

- Flow-dependent background error covariances can be estimated from the ensemble (with 50 members at Météo-France).

Dynamics of background error variances



Standard deviations of surface pressure errors (hPa)
(superimposed with MSLP analysis (hPa)).

Modelling and filtering covariances

- Huge size of **B** : model it with operators which are sparse and/or of small size.
- Sampling noise, and other uncertainties. => Spatio-temporal filtering.
- Factorisation : $\mathbf{B} = \mathbf{B}^{1/2}\,\mathbf{B}^{T/2}$

$$\mathbf{B}^{1/2} = \mathbf{L}\,\mathbf{S}\,\mathbf{C}^{1/2}$$

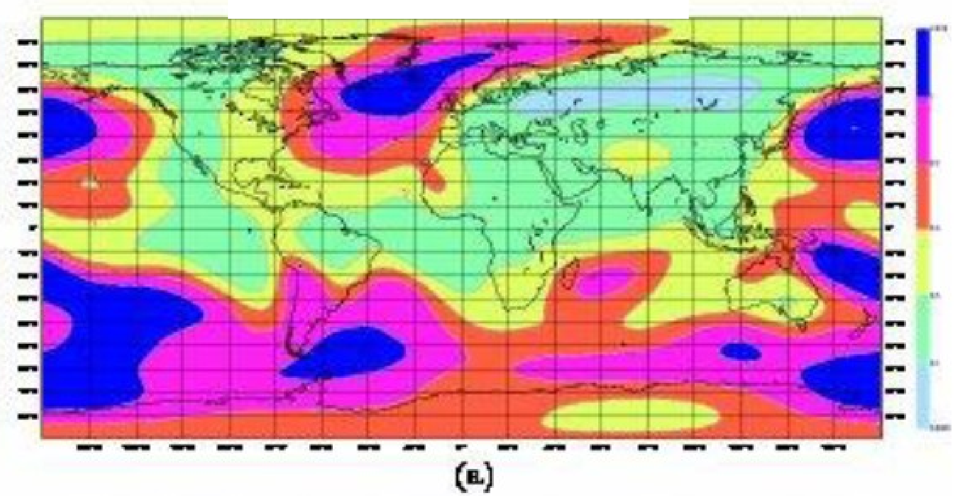
L ~ mass/wind cross-covariances (related to geostrophy),
including flow-dependence (non linear balances).

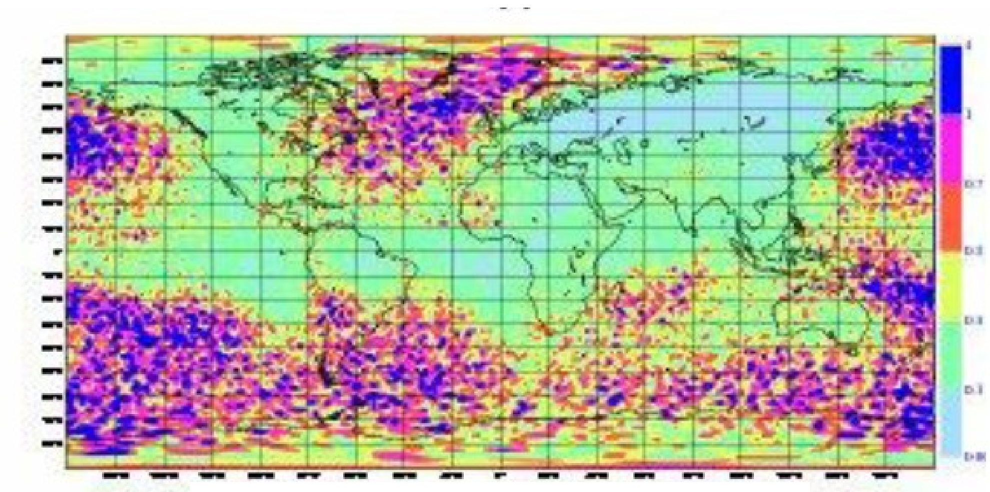
S flow-dependent standard deviations (~ expected error amplitudes),
filtered spatially.

C matrix of 3D spatial correlations (~ spatial structures of errors),
filtered in wavelet space (block-diagonal model).

Spatial filtering of variance field

EXACT
VARIANCES

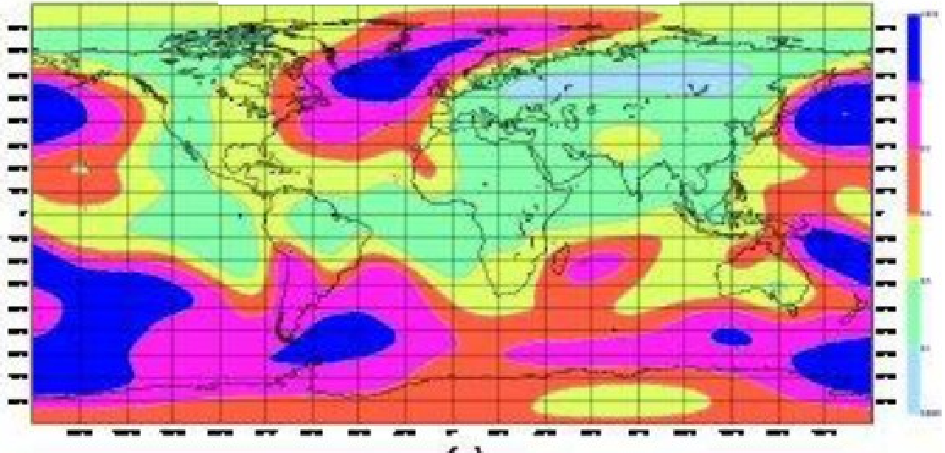




RAW ENSEMBLE
VARIANCES (N=6)

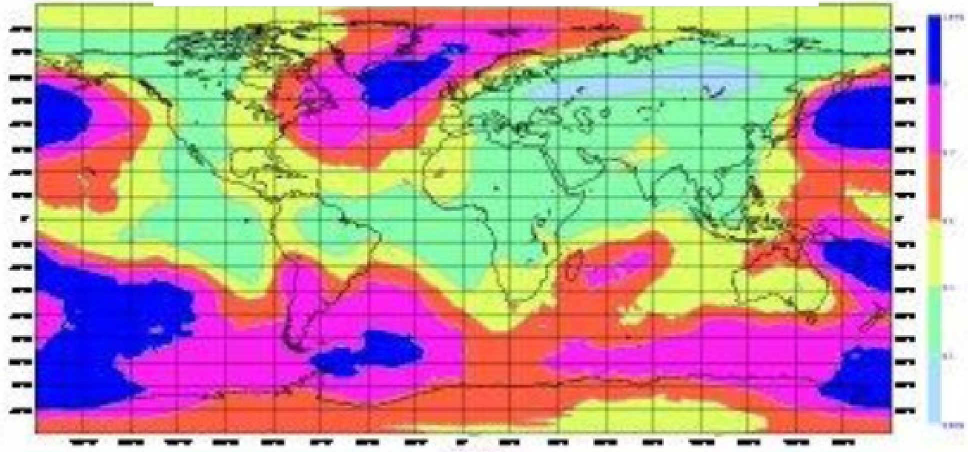
Spatial filtering of variance field

EXACT VARIANCES

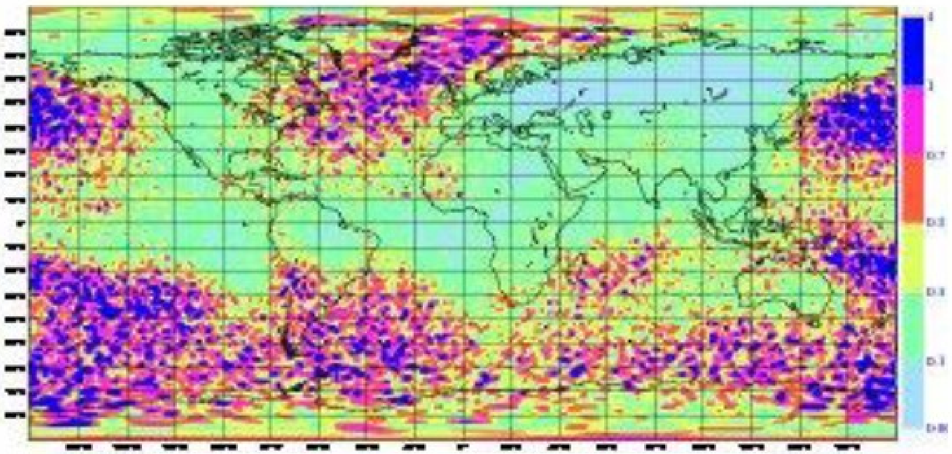


(a)

FILTERED ENSEMBLE VARIANCES (N=6)



(b)



RAW ENSEMBLE VARIANCES (N=6)

Low-pass filter applied to raw variances :

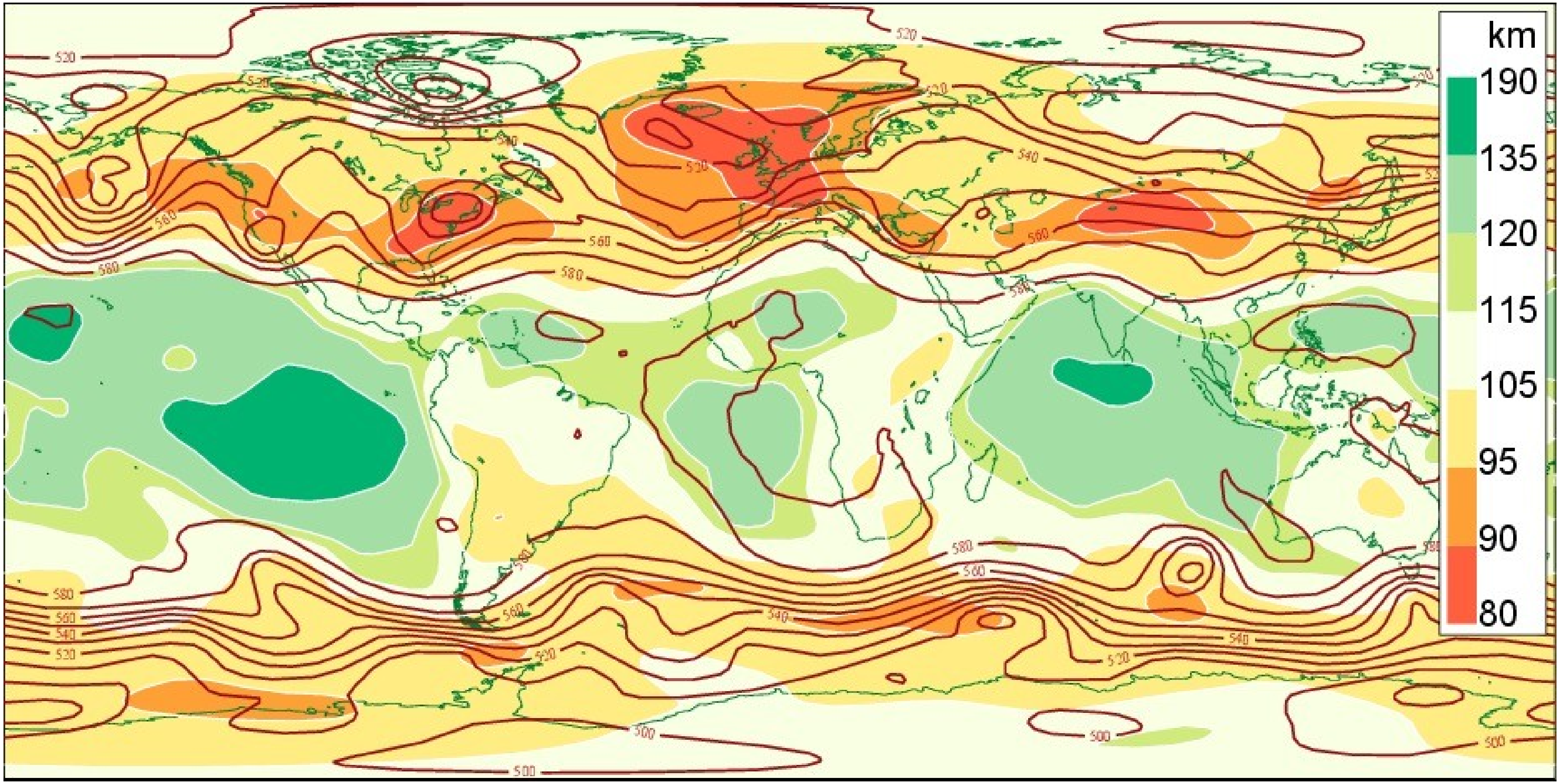
$$\mathbf{v'}_b = \mathbf{F}\,\mathbf{v}_b$$

with **F** optimized in spectral space / spatial structures of signal & noise :

$$F = 1 / (1 + E[noise^2]/signal^2)$$

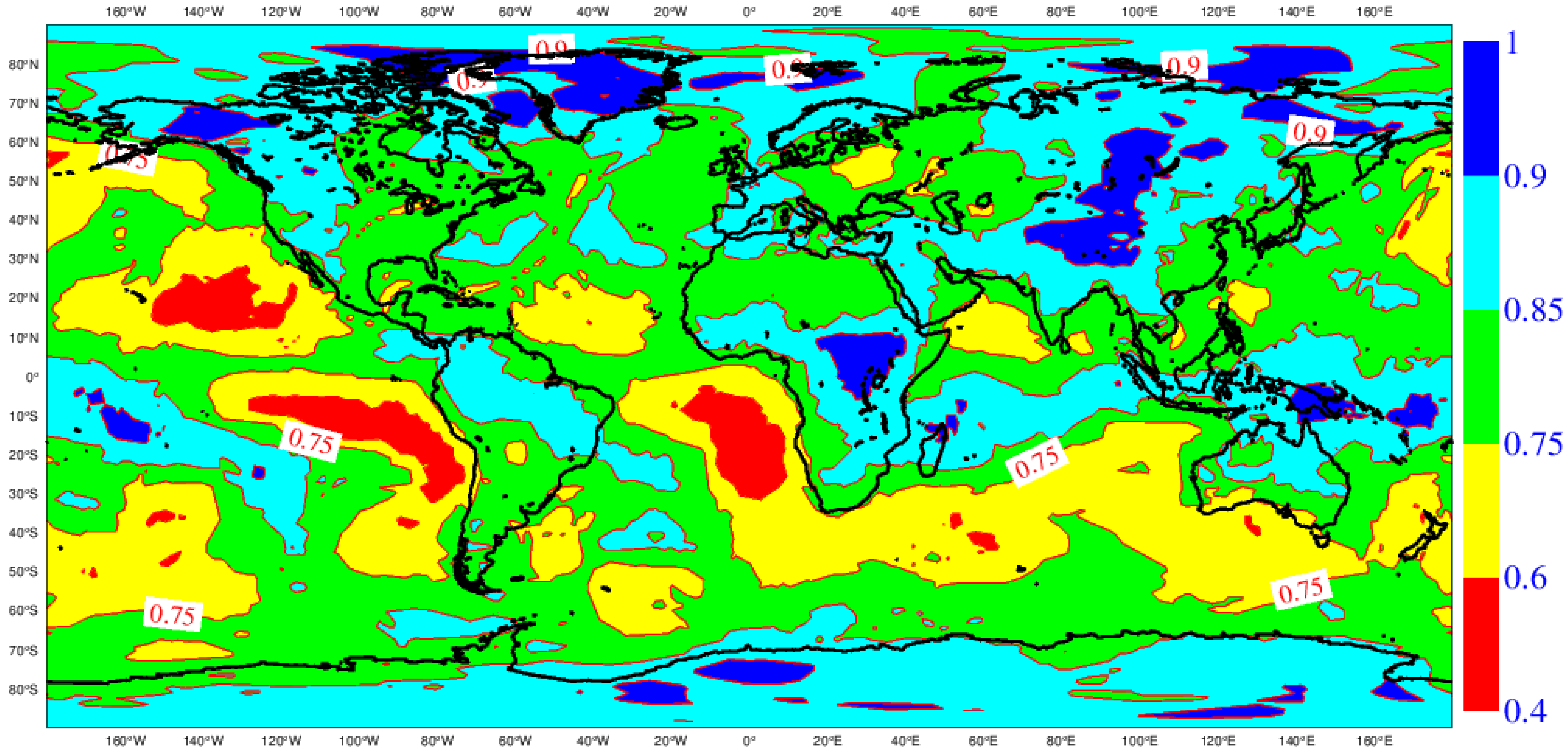
and $E[v_b^2] = signal^2 + E[noise^2]$

Dynamics of horizontal correlations



Horizontal length-scales (in km) of wind errors near 500 hPa, superimposed with geopotential

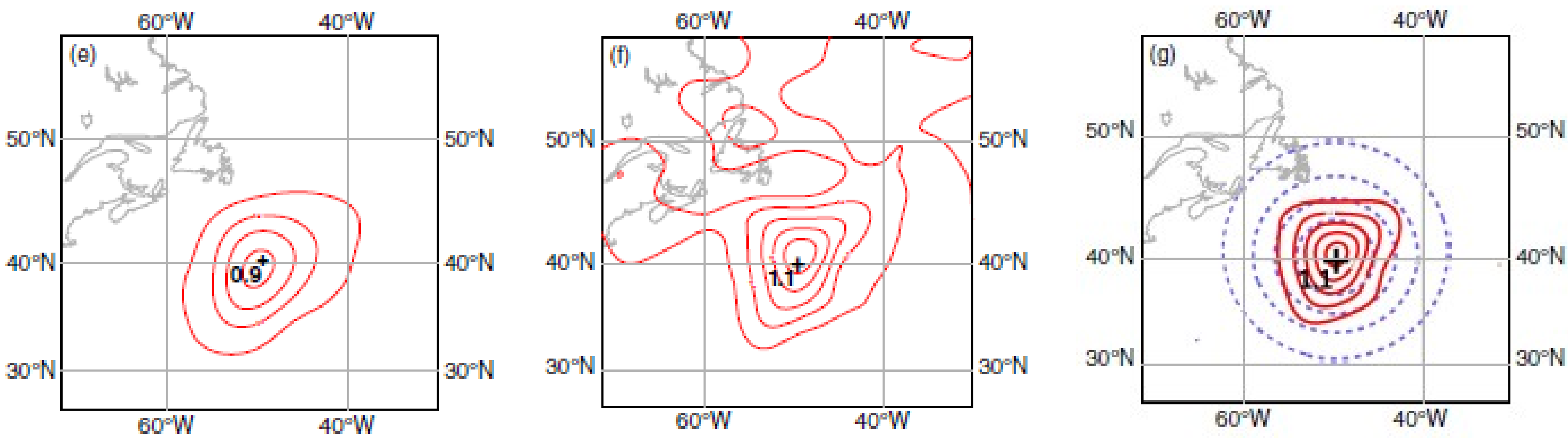
Dynamics of vertical correlations



Vertical correlations of temperature errors
between 850 & 870 hPa

Covariance anisotropy and localisation

Ensemble to get information on anisotropy,
but requires filtering = localisation.



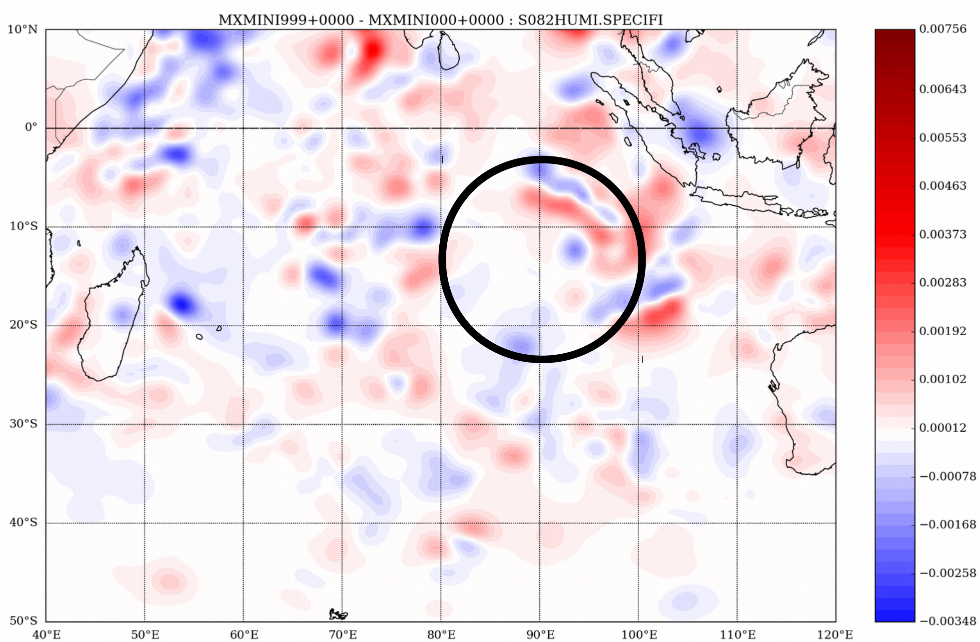
« Exact » covariances

Raw covariances (200 members)

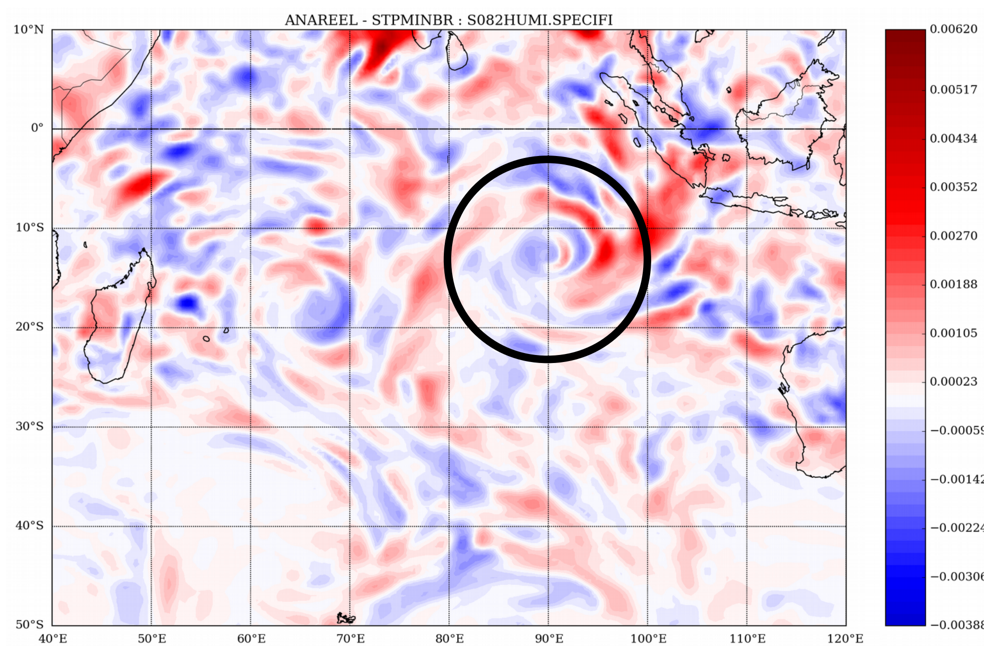
Localised covariances (200 members)

Flow-dependent anisotropic increments

Humidity analysis increments (near 850 hPa)



With isotropic correlations



With anisotropic correlations, filtered by localisation

Conclusions

Conclusions

- Data assimilation is vital for weather forecasting.
- Observations are very diverse in type, density and quality.
- 4D-Var for temporal and non linear aspects.

- Observation-background departures for estimation of average variances and correlations in **R** and **B**.
- Ensemble DA for error simulation and for covariance dynamics.

- Sampling noise issues and filtering methods.
- Towards 4DEnVar (variational assimilation based on a 4D ensemble).

Thanks for your attention

Liens principaux entre thématiques ensemblistes

Spécification des incertitudes
(observations, modèle) du système d'analyse/prévision

⇩

Assimilation d'ensemble : simulation de la propagation
des erreurs au cours du "cyclage" de l'assimilation

⇩

Spécification des covariances d'erreur d'ébauche,
formulations EnVar de l'assimilation

⇩

Prévision d'ensemble : simulation de la propagation /
amplification des erreurs au cours de la prévision

⇩

Prévision probabiliste : traitement statistique
des prévisions de l'ensemble

Properties of innovation methods

- Provides estimates in observation space.
- A good quality data dense network is needed.
- Assumption that observation errors are spatially uncorrelated.
- An objective source of information on **B** and **R**.
- At a given location and time, only 1 innovation value :
  only a single error realization is available.

  => Statistical averages (expectations) are replaced by
  space and time averages (ergodic assumption).

4DEnVar
Variational analysis based on a 4D Ensemble

Minimisation of $J(\underline{\delta \mathbf{x}})$ where $\underline{\delta \mathbf{x}}$ is a 4D analysis increment :

$$J(\underline{\delta \mathbf{x}}) = \underline{\delta \mathbf{x}}^T \underline{\mathbf{B}}^{-1} \underline{\delta \mathbf{x}} + (\underline{\mathbf{d}}\text{-}\underline{\mathbf{H}}\ \underline{\delta \mathbf{x}})^T \underline{\mathbf{R}}^{-1} (\underline{\mathbf{d}}\text{-}\underline{\mathbf{H}}\ \underline{\delta \mathbf{x}})$$

with $\underline{\mathbf{B}} = \underline{\mathbf{X}}^{b'} \underline{\mathbf{X}}^{b'T} \circ \underline{\mathbf{L}}$, where $\mathbf{L}$ is a localization matrix,

$$\underline{\mathbf{X}}^{b'} = (\underline{\mathbf{x}}^{b'}{}_1, \ldots, \underline{\mathbf{x}}^{b'}{}_{Ne}),$$

$$\underline{\mathbf{x}}^{b'}{}_{ne} = \underline{\mathbf{x}}^{b}{}_{ne} - \langle \underline{\mathbf{x}}^{b} \rangle / (N^e-1)^{1/2}, \ ne = 1, N^e.$$

$\underline{\mathbf{x}}^{b'}$ of dimension K+1 (time) x M (3D variables) x N (dim 3D).

(Liu et al, 2008, 2009 ; Buehner et al, 2010 ; Lorenc, 2012 ;

Desroziers et al 2014).

4DEnVar
Variational analysis based on a **4D En**semble

- 4D covariances from an ensemble of trajectories.

- Improved realism of 4D background error covariances (anisotropies, non linear evolution including all physical processes).

- No need to develop and maintain an adjoint model in this case.
  - Especially important for AROME.

- Pursue within the variational framework
  - Global assimilation of all available observations, distributed in space and in time.

- Introduces additional levels of parallelism (space, time, ensemble).