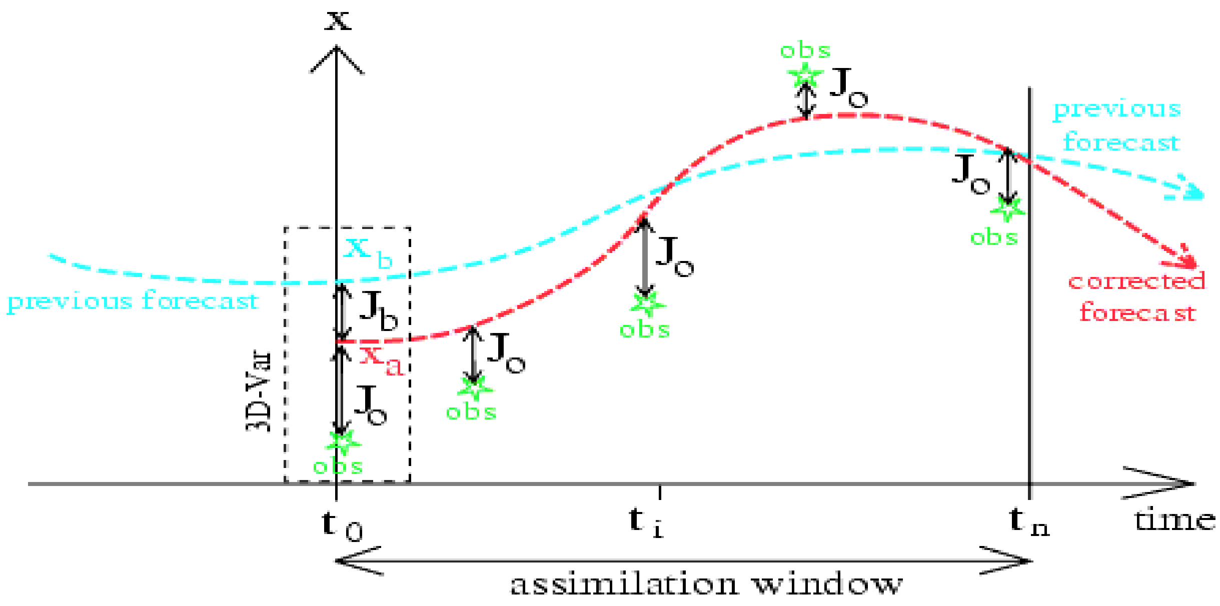
Background Error Covariance Modelling

- Variational data assimilation
- Importance of background covariances
- Diagnosing background error statistics
- Estimating background errors statistics

Variational data assimilation



$J_b(\delta x)=1/2\ \delta x^T$ B$^{-1}\delta x$ ✉ measure of discrepancy

between state vector and “first guess”.

Importance of Background Covariances

The formulation of B matrix (Jb term of the cost function) is crucial to the performance of current analysis systems.

Exemple:

- We suppose we have a single observation of the value of a model field at one gridpoint.

→For this simple case, the observation operator is:

$$H = (0,\dots,0,1,0,\dots,0) .$$

→The gradient of the 3dVar cost function is:

$$\nabla J = B^{-1}(x-x_b) + H^T R^{-1}(Hx-y) = 0$$

$$x - x_b = B\, H^T R^{-1}(y-Hx)$$

Importance of Background Covariances

- $x - x_b = B\,H^T R^{-1}(y-Hx)$
- For this simple case, $R^{-1}(y-Hx)$ is a scalar => $x-x_b \propto BH^T$

  But, $H = (0,\ldots,0,1,0,\ldots,0)$

=> The analysis increment is proportional to a column of B.

- **The role of B is:**
  1. To spread out the information from the observations.
  2. To provide statistically consistent increments at the neighbouring gridpoints and levels of the model.
  3. To ensure that observations of one model variable (e.g. temperature) produce dynamically consistent increments in the other model variables (e.g. vorticity and divergence).

Main Issues in Covariance Modelling

- There are 2 problems to be addressed in specifying B:

1. **We want to describe the statistics of the errors in the background**.
   - However, we don't know what the errors in the background are, since we don't know the true state of the atmosphere.

2. **The B matrix is enormous (~10^7x10^7).**
   - We are forced to simplify it just to fit it into the computer.
   - Even if we could fit it into the computer, we don't have enough statistical information to determine all its elements.

Diagnosing Background Error Statistics

- Problem:
  - We cannot produce samples of background error. (We don't know the true state.)
- Instead, we must either:
  - Dissociate background errors from the information we do have: innovation (observation-minus-background) statistics.
- Or:
  - Use a surrogate quantity whose error statistics are similar to those of background error. Two possibilities are:
    - Differences between forecasts that verify at the same time.
    - differences between background fields from an ensemble of analyses.

Diagnosing Background Error Statistics

- Three approaches to estimating Jb statistics:

1. **The Hollingsworth and Lönnberg (1986) method**
   - Differences between observations and the background are a combination of background and observation error.
   - The method tries to partition this error into background errors and observation errors by assuming that the observation errors are spatially uncorrelated.
2. **The NMC method (Parrish and Derber, 1992)**
   - This method assumes that the spatial correlations of backgound error are similar to the correlations of differences between 48h and 24h forecasts verifying at the same time.
3. **The Analysis-Ensemble method (Fisher, 2003)**
   - This method runs the analysis system several times for the same period with randomly-perturbed observations. Differences between background fields for different runs provide a surrogate for a sample of background error.

The Hollingsworth and Lönnberg method

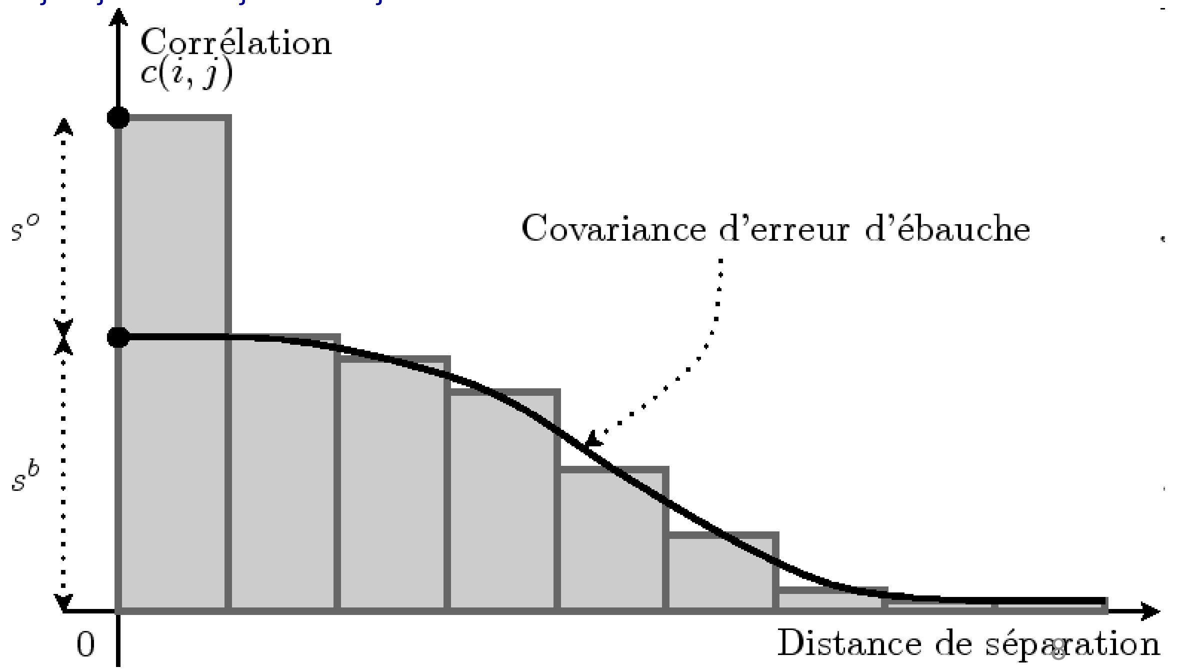
- Hypotheses :

1) Background errors are independent of observation errors.
2) Observations have spatially uncorrelated errors (for some observation types).

The innovation (obs-guess) for the i^{th} observation is $d_i = y_i - H_i(x_b)$

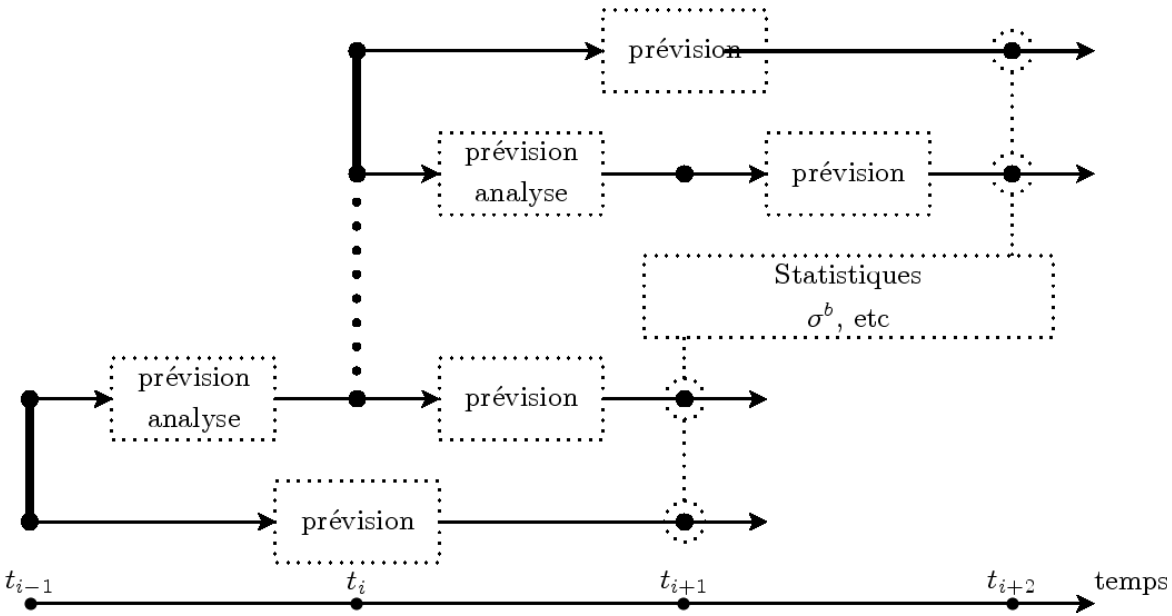
$$C(d_i, d_j) = cov[(y_i - H_i x^b), (y_j - H_j x^b)] = R_{ij} + H_i B H_j^T$$

We can extract a lot of useful information by plotting $Cov(d_i , d_j)$ as a function of the distance between pairs of observations.



The NMC method

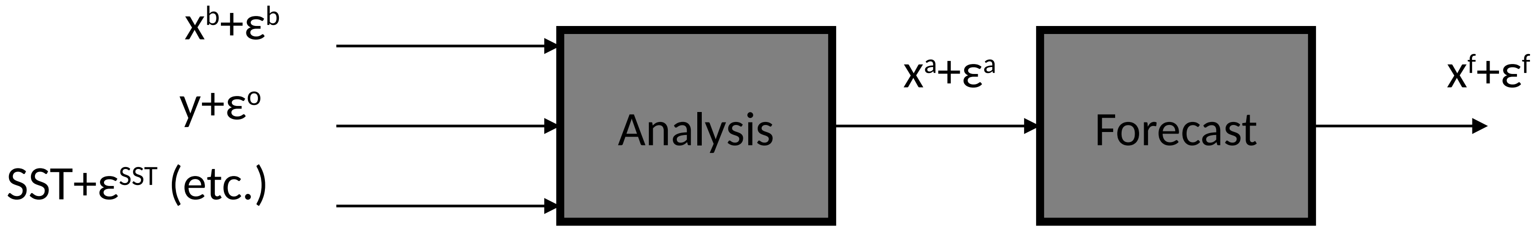
- The background error statistics are extracted from statistics on differences between two forecasts valid at the same time.



- A forecast starts at t_{i-1} and ends at t_{i+1}.
- At the same time t_{i-1}, an assimilation cycle is performed and makes it possible to obtain an initial state at the instant t_i, from which a forecast of ti to t_{i+1} is made.

The Analysis-Ensemble method

- Suppose we perturb all the inputs to the analysis/forecast system with random perturbations, drawn from the relevant distributions:



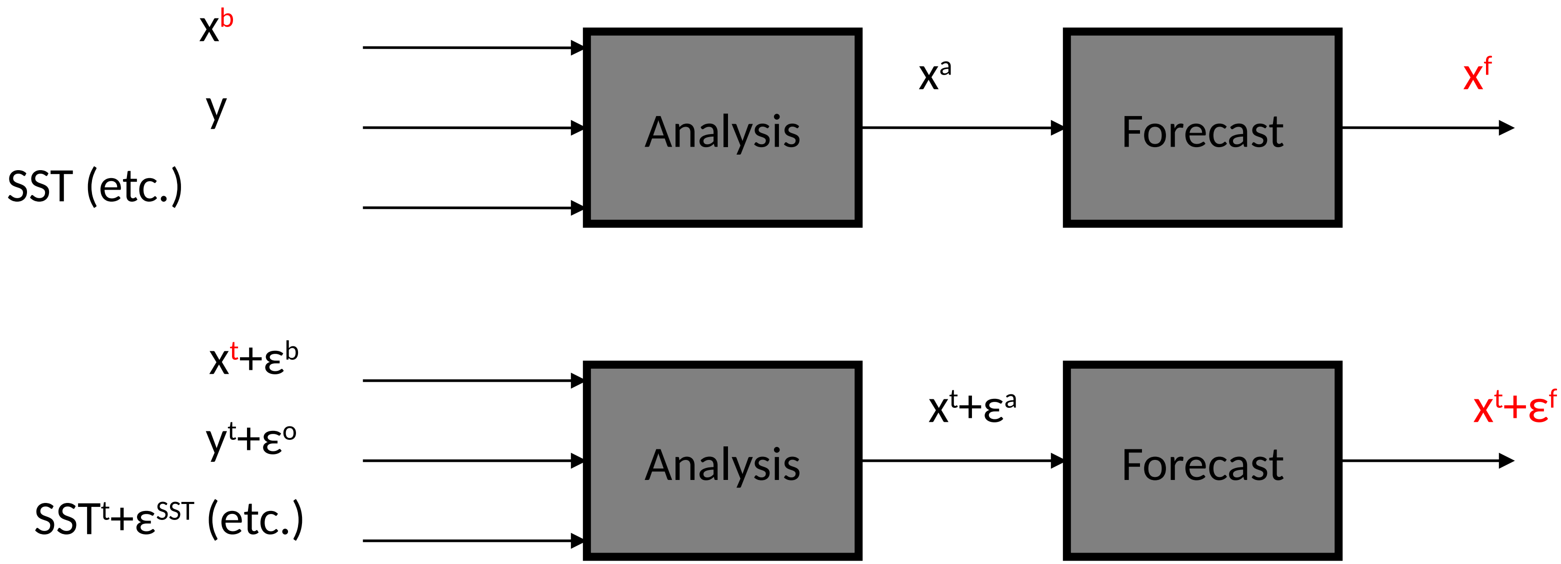
- The result will be a perturbed analysis and forecast, with perturbations characteristic of analysis and forecast error.

- The perturbed forecast may be used as the background for the next (perturbed) cycle.

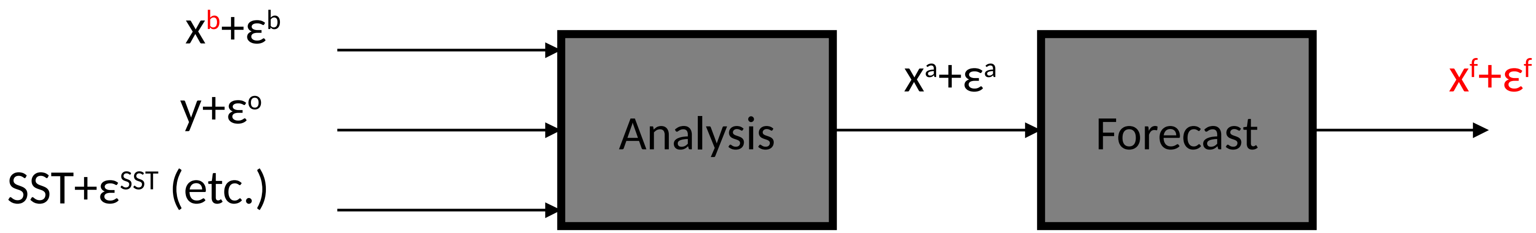
- After a few cycles, the system will have forgotten the original initial background perturbations.

The Analysis-Ensemble method

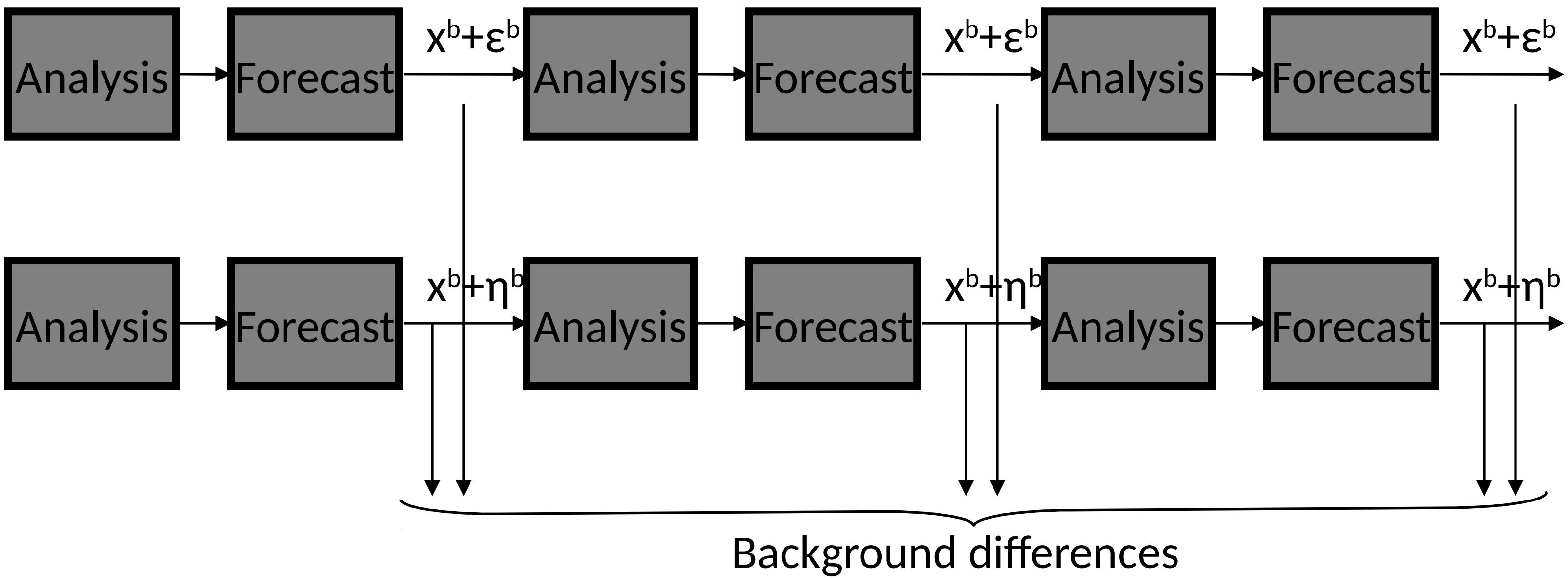
Normal Analysis



Perturbed Analysis



The Analysis-Ensemble method



- Run the analysis system several times with different perturbations, and form differences between pairs of background fields.
- These differences will have the statistical characteristics of background error.

Comparison between methods

- **Innovation statistics:**
  - ☺The only direct method for diagnosing background error statistics.
  - ☹Provides statistics of background error in observation space.
  - ☹Statistics are not global, and do not cover all model levels.
  - ☹Requires a good uniform observing network.
  - ☹Statistics are biased towards data-dense areas.
- **Forecast Differences (NMC):**
  - ☺Inexpensive
  - ☹Generates global statistics of model variables at all levels.
  - ☹Statistics are a mixture of analysis and background error.
  - ☹Not good in data-sparse regions.
- **Ensembles of Analyses:**
  - ☺Diagnoses the statistics of the actual analysis system.
  - ☹Assumes statistics of observation error (and SST, etc.) are well known.
  - ☹Danger of feedback. (Noisy analysis system => noisy stats => noisier system.)

References

- Mike Fisher, Massimo Bonavita, Elias Holm, 2016: Background Error Covariance Modelling ECMWF Data Assimilation Training Course