

École Doctorale des Sciences de l'Environnement d'Île-de-France
Année 2009-2010

Modélisation Numérique
de l'Écoulement Atmosphérique
et Assimilation d'Observations

Olivier Talagrand
Cours 4
30 Avril 2010

- Observations *synoptiques* (observations au sol, radiosondages), effectuées simultanément, par convention internationale, dans toutes les stations météorologiques du globe (00:00, 06:00, 12:00, 18:00 TU)
- Observations *asynoptiques* (satellites, avions), effectuées plus ou moins continûment dans le temps.
- Observations *directes* (température, pression, composantes du vent, humidité), portant sur les variables utilisées pour décrire l'état de l'écoulement dans les modèles numériques
- Observations *indirectes* (observations radiométriques, ...), portant sur une combinaison plus ou moins complexe (le plus souvent, une intégrale d'espace unidimensionnelle) des variables utilisées pour décrire l'état de l'écoulement

$$y = H(x)$$

H : opérateur d'observation (par exemple, équation de transfert radiatif)

Échantillonnage de la circulation océanique par les missions altimétriques sur 10 jours :
combinaison Topex-Poséidon/ERS-1



S. Louvel, Doctoral Dissertation, 1999

Purpose of assimilation : reconstruct as accurately as possible the state of the atmospheric or oceanic flow, using all available appropriate information. The latter essentially consists of

- The observations proper, which vary in nature, resolution and accuracy, and are distributed more or less regularly in space and time.
- The physical laws governing the evolution of the flow, available in practice in the form of a discretized, and necessarily approximate, numerical model.
- ‘Asymptotic’ properties of the flow, such as, *e. g.*, geostrophic balance of middle latitudes. Although they basically are necessary consequences of the physical laws which govern the flow, these properties can usefully be explicitly introduced in the assimilation process.

Assimilation is one of many '*inverse problems*' encountered in many fields of science and technology

- solid Earth geophysics
- plasma physics
- 'nondestructive' probing
- navigation (spacecraft, aircraft,)
- ...

Solution most often (if not always) based on Bayesian, or probabilistic, estimation. 'Equations' are fundamentally the same.

Difficulties specific to assimilation of meteorological observations :

- Very large numerical dimensions ($n \approx 10^6$ - 10^9 parameters to be estimated, $p \approx 1$ - $3 \cdot 10^7$ observations per 24-hour period). Difficulty aggravated in Numerical Weather Prediction by the need for the forecast to be ready in time.
- Non-trivial, actually chaotic, underlying dynamics

Coût des différentes composantes de la chaîne de prévision opérationnelle du CEPMMT (mars 2010, J.-N. Thépaut) :

Analysis	Daily	Weekly
DA	44394	311325
DCDA	78306	536834
Total	122700	848159

Forecast	Daily	Weekly
DA	49673	347546
DCDA	4517	30788
Total	54190	378334

EPS	Daily	Weekly
	193028	1351576

Monthly	Daily	Weekly
	N/A	46129

Hindcast	Daily	Weekly
	N/A	256724

$$\begin{array}{ll}
 z_1 = x + \zeta_1 & \text{density function } p_1(\zeta) \propto \exp[-(\zeta^2)/2s_1] \\
 z_2 = x + \zeta_2 & \text{density function } p_2(\zeta) \propto \exp[-(\zeta^2)/2s_2]
 \end{array}$$

$$x = \xi \Leftrightarrow \zeta_1 = z_1 - \xi \text{ and } \zeta_2 = z_2 - \xi$$

$$\begin{aligned}
 P(x = \xi | z_1, z_2) &\propto p_1(z_1 - \xi) p_2(z_2 - \xi) \\
 &\propto \exp[-(\xi - x^a)^2/2p^a]
 \end{aligned}$$

where $1/p^a = 1/s_1 + 1/s_2$, $x^a = p^a (z_1/s_1 + z_2/s_2)$

Conditional probability distribution of x , given z_1 and z_2 : $\mathcal{N}[x^a, p^a]$
 $p^a < (s_1, s_2)$ independent of z_1 and z_2

$$z_1 = x + \xi_1$$

$$z_2 = x + \xi_2$$

Same as before, but ξ_1 and ξ_2 are now distributed according to exponential law with parameter a , *i. e.*

$$p(\xi) \propto \exp[-|\xi|/a] \quad ; \quad \text{Var}(\xi) = 2a^2$$

Conditional probability density function is now uniform over interval $[z_1, z_2]$, exponential with parameter $a/2$ outside that interval

$$E(x | z_1, z_2) = (z_1 + z_2)/2$$

$$\text{Var}(x | z_1, z_2) = a^2 (2\delta^3/3 + \delta^2 + \delta + 1/2) / (1 + 2\delta), \text{ with } \delta = |z_1 - z_2| / (2a)$$

Increases from $a^2/2$ to ∞ as δ increases from 0 to ∞ . Can be larger than variance $2a^2$ of original errors (probability 0.08)

(Entropy $-f \ln p$ always decreases in bayesian estimation)

Bayesian estimation

State vector x , belonging to *state space* \mathcal{S} ($\dim \mathcal{S} = n$), to be estimated.

Data vector z , belonging to *data space* \mathcal{D} ($\dim \mathcal{D} = m$), available.

$$z = F(x, \xi) \quad (1)$$

where ξ is a random element representing the uncertainty on the data (or, more precisely, on the link between the data and the unknown state vector).

For example

$$z = \Gamma x + \xi$$

Bayesian estimation (continued)

Probability that $x = \xi$ for given ξ ?

$$x = \xi \Rightarrow z = F(\xi, \zeta)$$

$$P(x = \xi | z) = P[z = F(\xi, \zeta)] / \int_{\xi'} P[z = F(\xi', \zeta)]$$

Unambiguously defined iff, for any ζ , there is at most one x such that (1) is verified.

\Leftrightarrow data contain information, either directly or indirectly, on any component of x .
Determinacy condition.

Bayesian estimation is however impossible in its general theoretical form in meteorological or oceanographical practice because

- It is impossible to explicitly describe a probability distribution in a space with dimension even as low as $n \approx 10^3$, not to speak of the dimension $n \approx 10^{6-9}$ of present Numerical Weather Prediction models.
- Probability distribution of errors on data very poorly known (model errors in particular).

One has to restrict oneself to a much more modest goal. Two approaches exist at present

- Obtain some ‘central’ estimate of the conditional probability distribution (expectation, mode, ...), plus some estimate of the corresponding spread (standard deviations and a number of correlations).
- Produce an ensemble of estimates which are meant to sample the conditional probability distribution (dimension $N \approx O(10-100)$).