

École Doctorale des Sciences de l'Environnement d'Île-de-France
Année Universitaire 2011-2012

Modélisation Numérique
de l'Écoulement Atmosphérique
et Assimilation de Données

Olivier Talagrand
Cours 4
15 Mars 2012

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

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

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

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 $N \approx O(10-100)$).

Random vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^T = (x_i)$ (e. g. pressure, temperature, abundance of given chemical compound at n grid-points of a numerical model)

- Expectation $E(\mathbf{x}) \equiv [E(x_i)]$; centred vector $\mathbf{x}' \equiv \mathbf{x} - E(\mathbf{x})$
- Covariance matrix

$$E(\mathbf{x}'\mathbf{x}'^T) = [E(x_i'x_j')]$$

dimension $n \times n$, symmetric non-negative (strictly definite positive except if linear relationship holds between the x_i' 's with probability 1).

- Two random vectors
 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$
 $\mathbf{y} = (y_1, y_2, \dots, y_p)^T$

$$E(\mathbf{x}'\mathbf{y}'^T) = E(x_i'y_j')$$

dimension $n \times p$

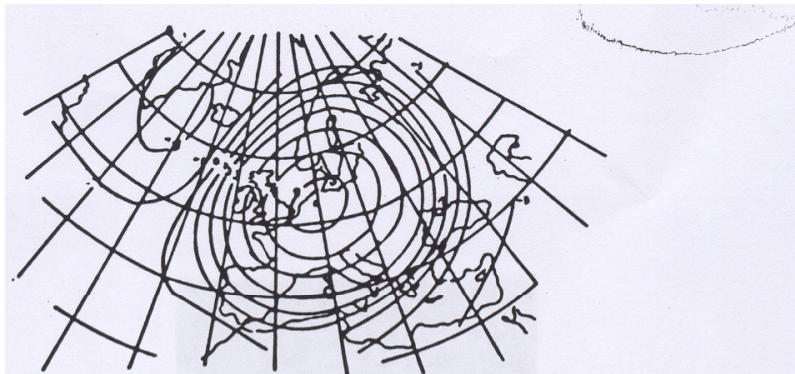
Random function $\Phi(\xi)$ (field of pressure, temperature, abundance of given chemical compound, ... ; ξ is now spatial and/or temporal coordinate)

- Expectation $E[\Phi(\xi)]$; $\Phi'(\xi) \equiv \Phi(\xi) - E[\Phi(\xi)]$
- Variance $Var[\Phi(\xi)] = E\{[\Phi'(\xi)]^2\}$
- Covariance function

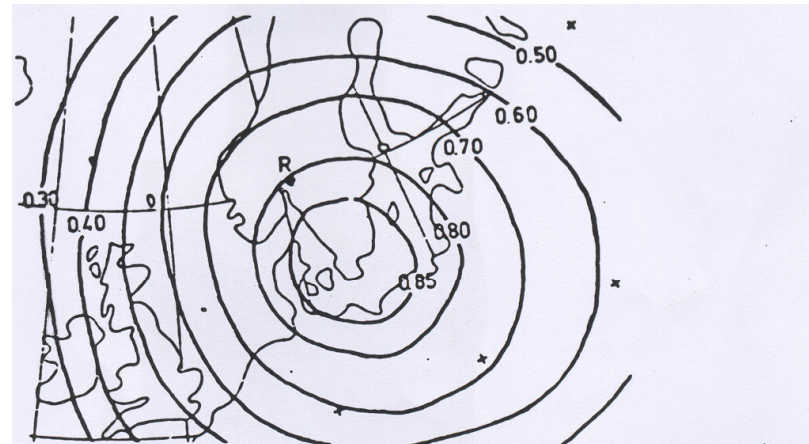
$$(\xi_1, \xi_2) \rightarrow C_\phi(\xi_1, \xi_2) \equiv E[\Phi'(\xi_1) \Phi'(\xi_2)]$$

- Correlation function

$$Cor_\phi(\xi_1, \xi_2) \equiv E[\Phi'(\xi_1) \Phi'(\xi_2)] / \{Var[\Phi(\xi_1)] Var[\Phi(\xi_2)]\}^{1/2}$$



.: Isolines for the auto-correlations of the 500 mb geopotential between the station in Hannover and surrounding stations.
From Bertoni and Lund (1963)



Isolines of the cross-correlation between the 500 mb geopotential in station 01 384 (R) and the surface pressure in surrounding stations.

After N. Gustafsson

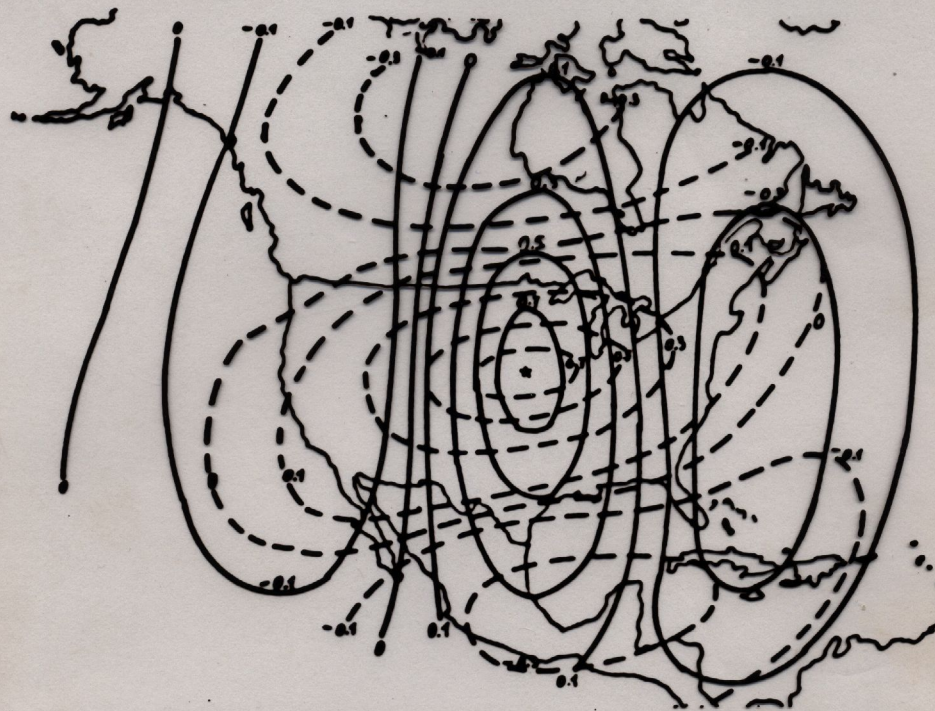
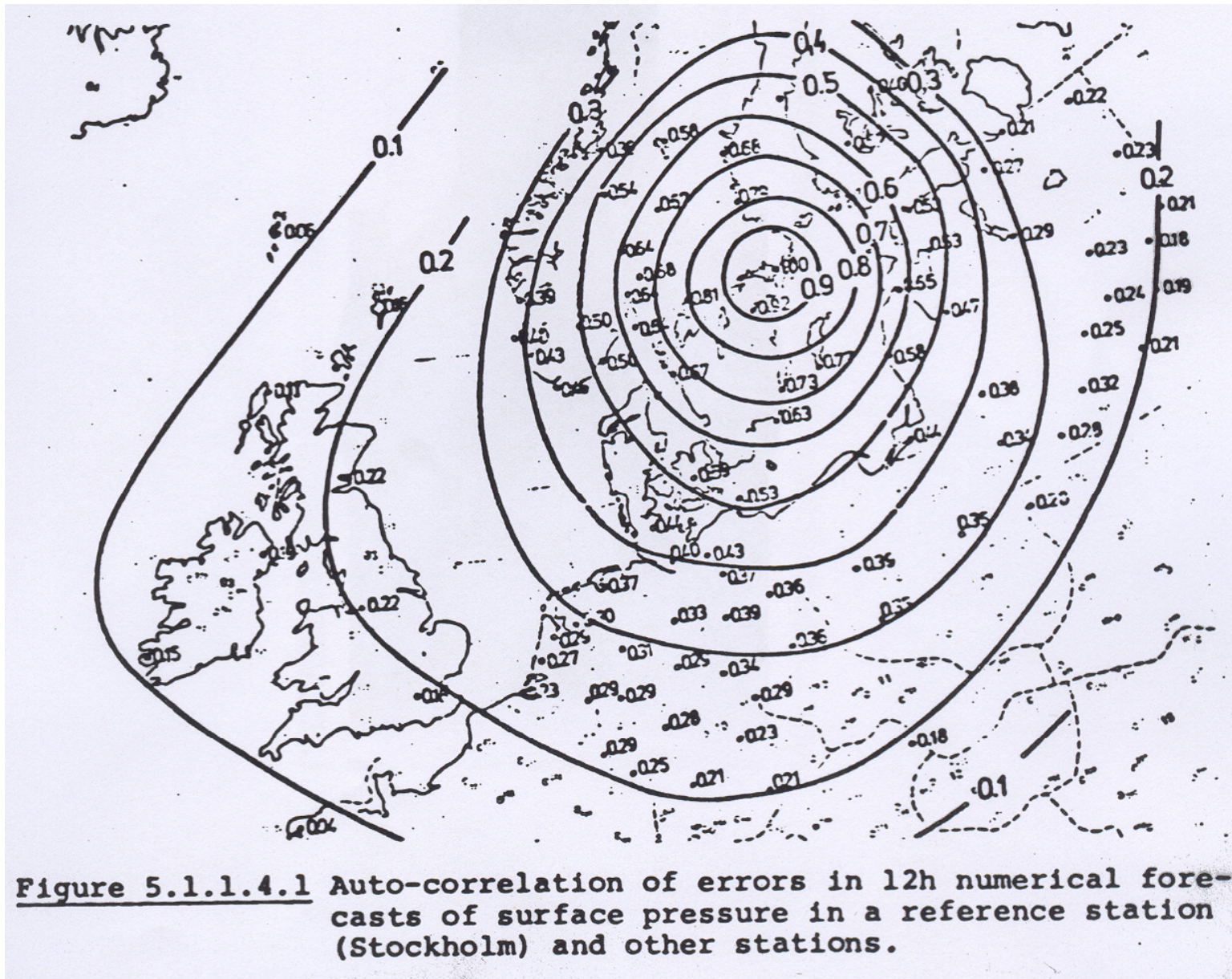


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson



After N. Gustafsson