École Doctorale des Sciences de l'Environnement d’Île-de-France

Année Universitaire 2018-2019

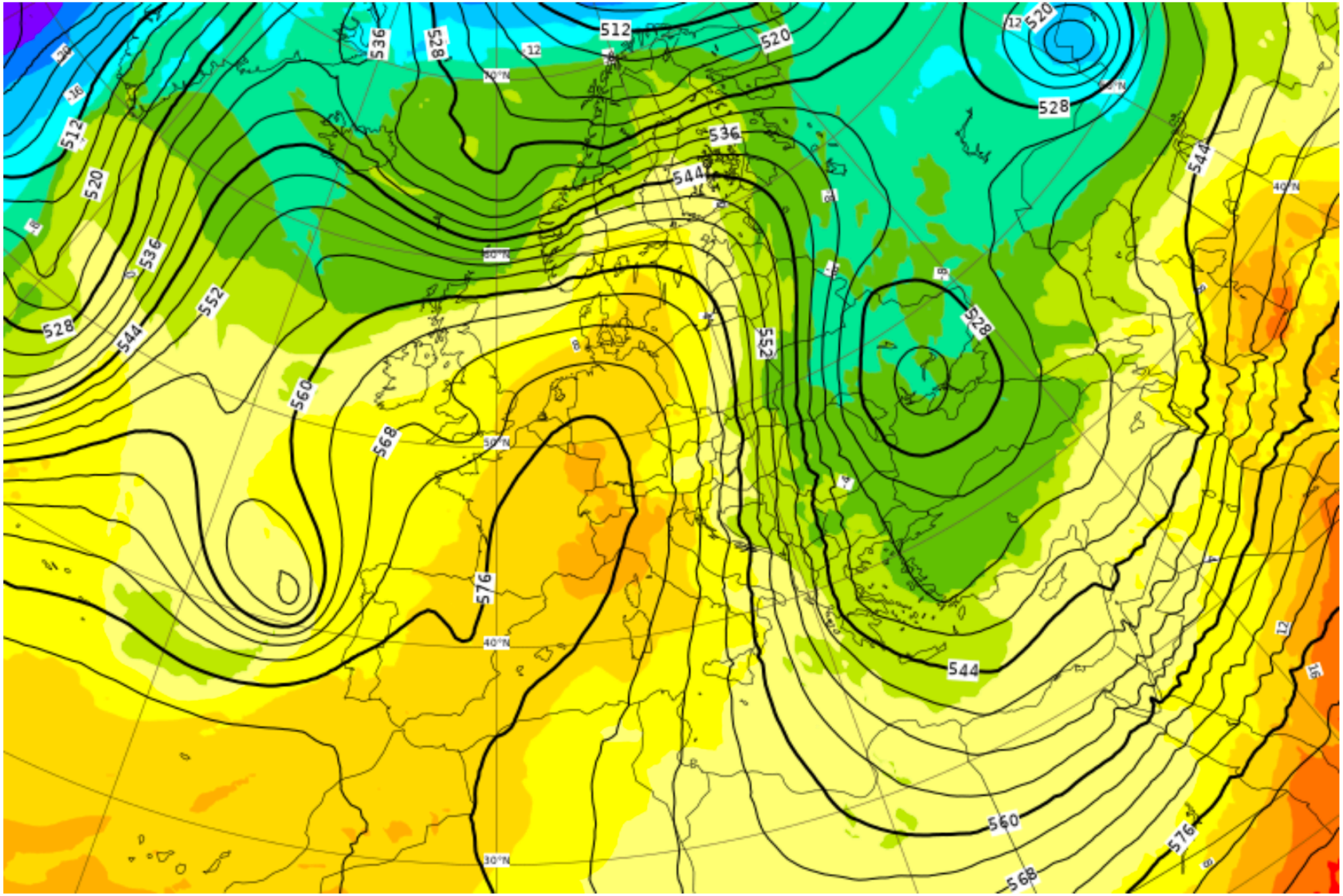
Modélisation Numérique de l’Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand
Cours 3

28 Février 2019

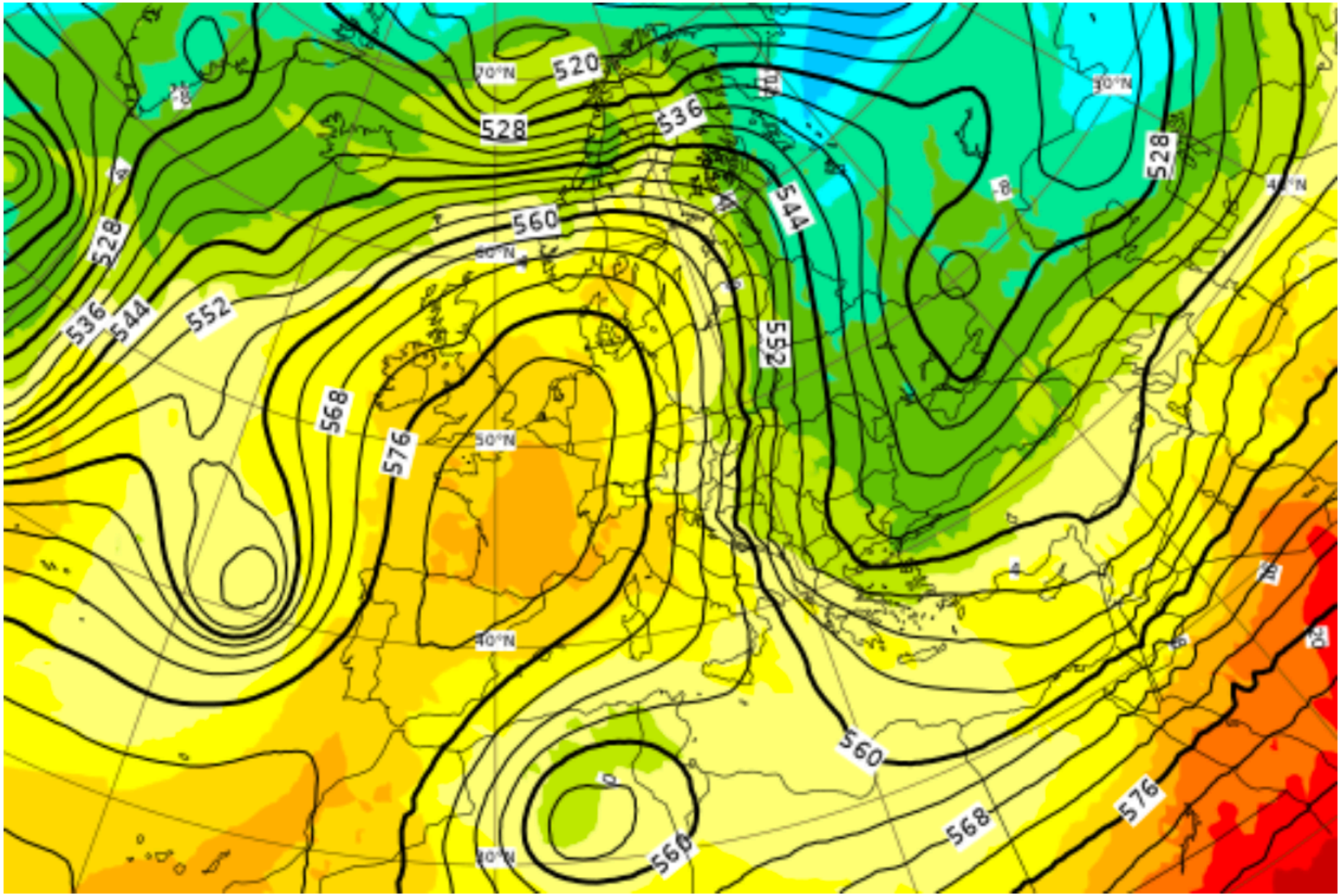
850 hPa temperature / 500 hPa geopotential

Tuesday 19 Feb, 00 UTC T+192 Valid: Wednesday 27 Feb, 00 UTC

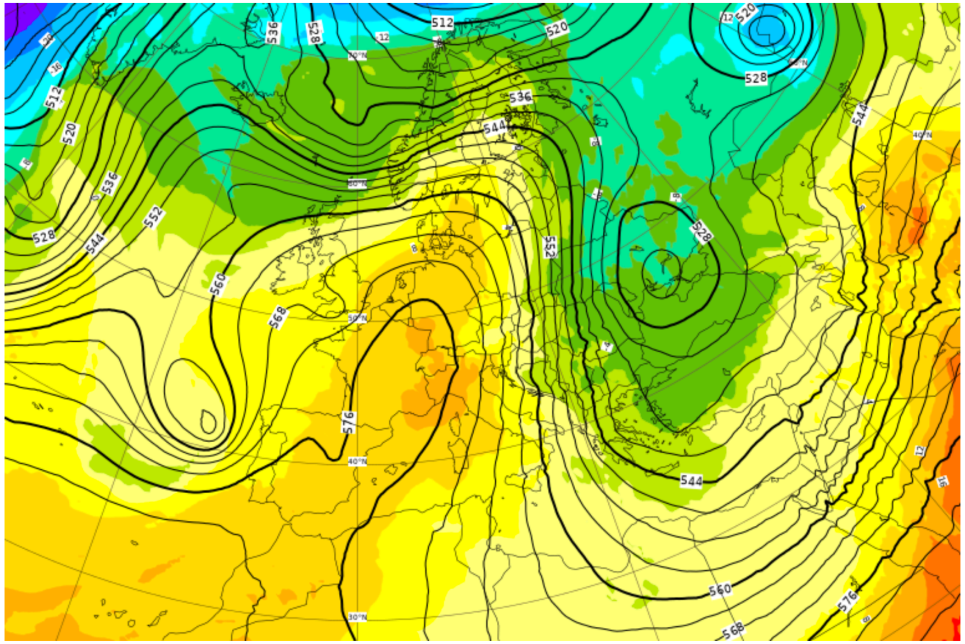


850 hPa temperature / 500 hPa geopotential

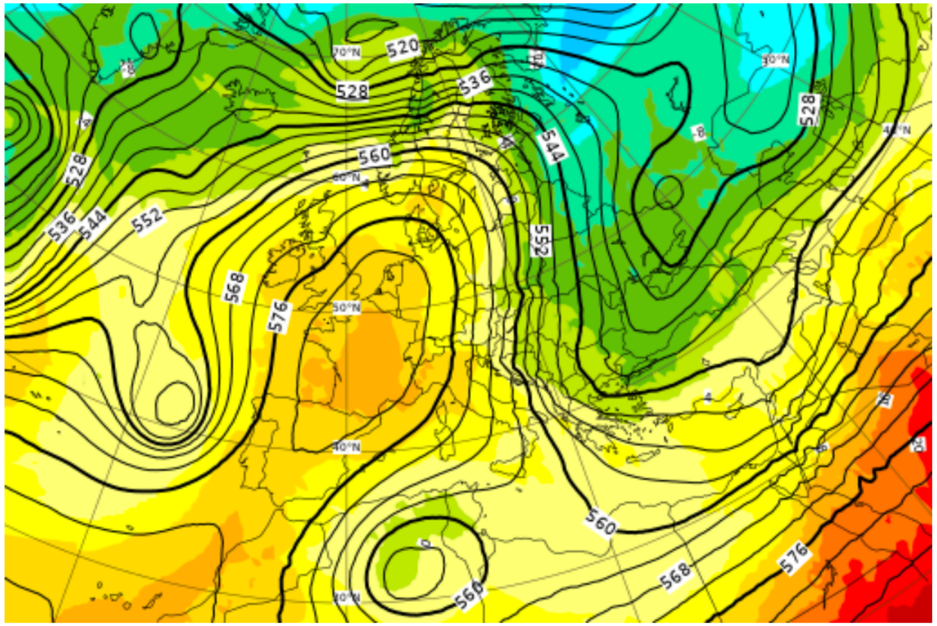
Wednesday 27 Feb, 00 UTC T+0 Valid: Wednesday 27 Feb, 00 UTC



850 hPa temperature / 500 hPa geopotential
Tuesday 19 Feb, 00 UTC T+192 Valid: Wednesday 27 Feb, 00 UTC

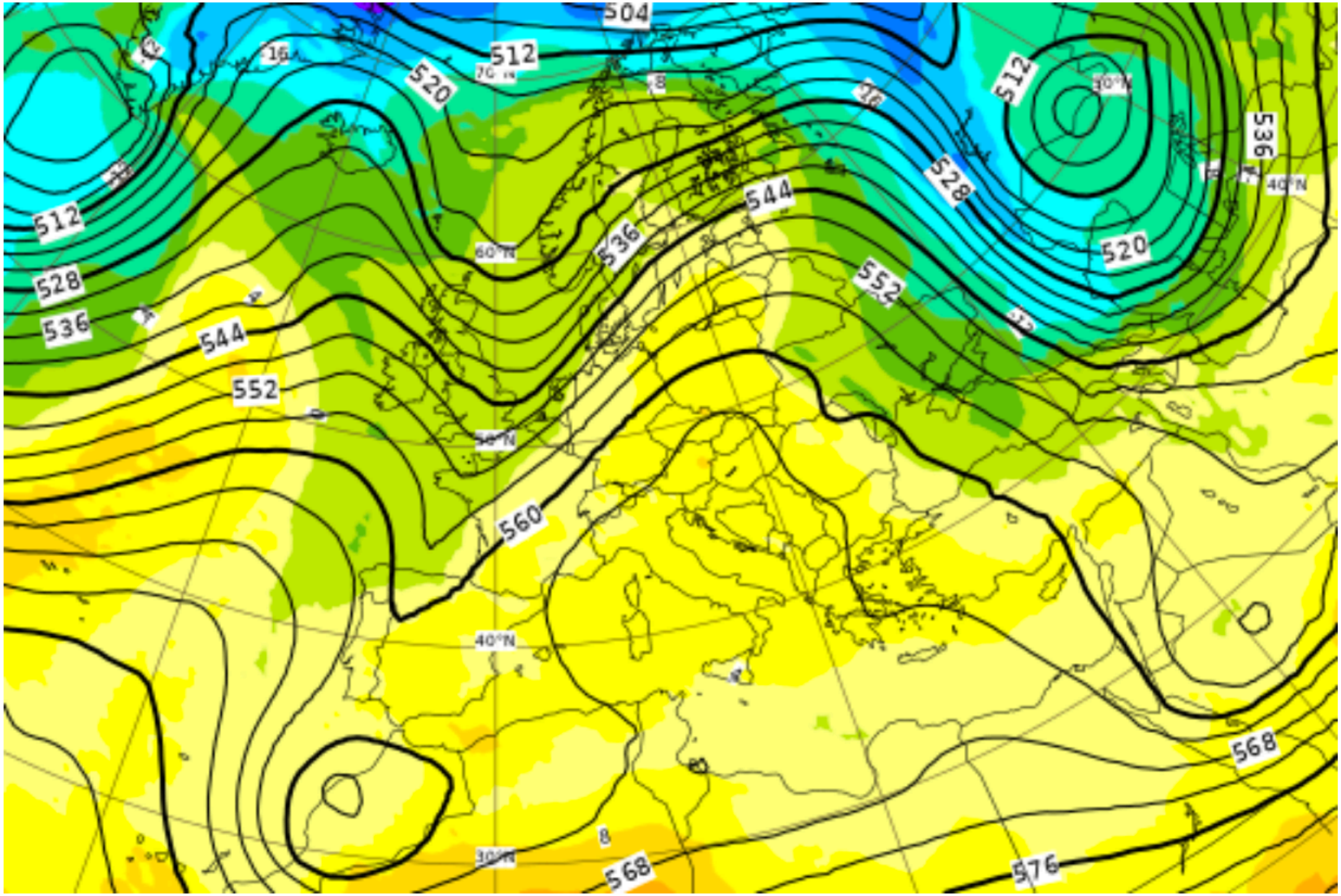


850 hPa temperature / 500 hPa geopotential
Wednesday 27 Feb, 00 UTC T+0 Valid: Wednesday 27 Feb, 00 UTC



850 hPa temperature / 500 hPa geopotential

Tuesday 19 Feb, 00 UTC T+0 Valid: Tuesday 19 Feb, 00 UTC



- Bayesian estimation. Continuation. A simple example.

- Reminder on elementary probability theory. Random vectors and covariance matrices, random functions and covariance functions

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, $p \approx$ 4-5.10^7 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

Both observations and 'model' are affected with some uncertainty ⇒ uncertainty on the estimate.

For some reason, uncertainty is conveniently described by probability distributions (don't know too well why, but it works; see, *e.g*. Jaynes, 2007, *Probability Theory: The Logic of Science,* Cambridge University Press).

Assimilation is a problem in bayesian estimation.

Determine the conditional probability distribution for the state of the system, knowing everything we know (see Tarantola, A., 2005*, Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM).

Coût des différentes composantes de la chaîne de prévision opérationnelle du CEPMMT (septembre 2015, J.-N. Thépaut) :

4DVAR: 9.5%

HRES FC: 4.5%

EDA: 30%

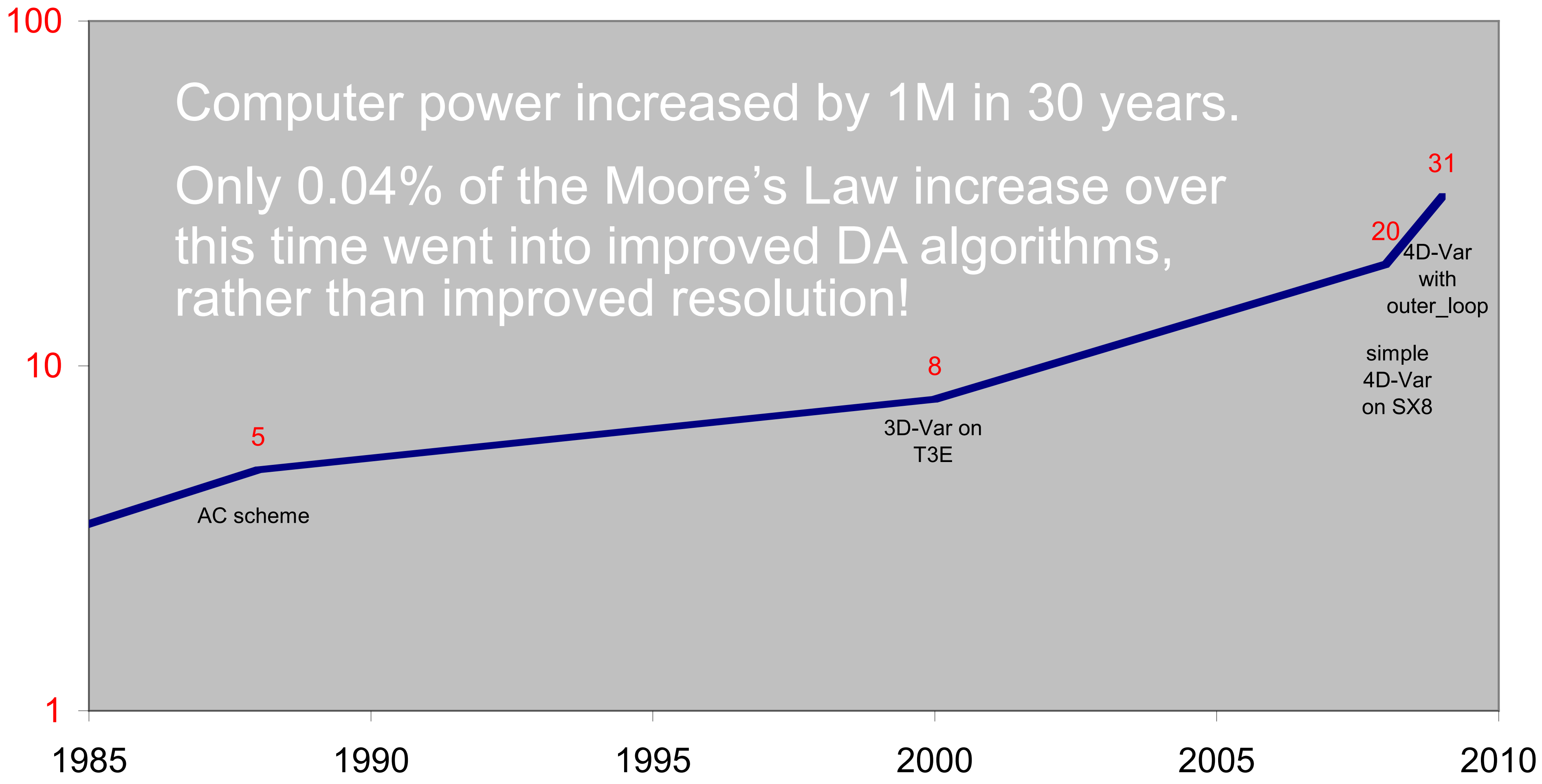
ENS: 22%

ENS: hindcasts 14%

Other: 20% of which BC AN: 3.5% BC FC: 4% BC ENS: 9.5%

L'EDA fournit à la fois les variances d'erreur d'ébauche du 4D-Var, et les perturbations initiales (en complément des vecteurs singuliers) de l'EPS.

ratio of supercomputer costs:
1 day's assimilation / 1 day forecast



Computer power increased by 1M in 30 years.

Only 0.04% of the Moore’s Law increase over this time went into improved DA algorithms, rather than improved resolution!

Courtesy A. Lorenc

Bayesian Estimation

Determine conditional probability distribution of the state of the system, given the probability distribution of the uncertainty on the data

$z_1 = x + \zeta_1$ $\zeta_1 = \mathcal{N}[0, s_1]$

density function $p_1(\zeta) \propto \exp[\, - (\zeta^2)/2s_1]$

$z_2 = x + \zeta_2$ $\zeta_2 = \mathcal{N}[0, s_2]$

density function $p_2(\zeta) \propto \exp[\, - (\zeta^2)/2s_2]$

- ζ_1 and ζ_2 mutually independent

What is the conditional probability $P(x = \xi \mid z_1, z_2)$ that x be equal to some value ξ ?

$z_1 = x + \zeta_1$ density function $p_1(\zeta) \propto \exp[- (\zeta^2)/2s_1]$

$z_2 = x + \zeta_2$ density function $p_2(\zeta) \propto \exp[- (\zeta^2)/2s_2]$

ζ_1 and ζ_2 mutually independent

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

- $P(x = \xi \mid z_1, z_2) \propto p_1(z_1 - \xi)\, p_2(z_2 - \xi)$

  $\propto \exp[- (\xi - x^a)^2/2p^a]$

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

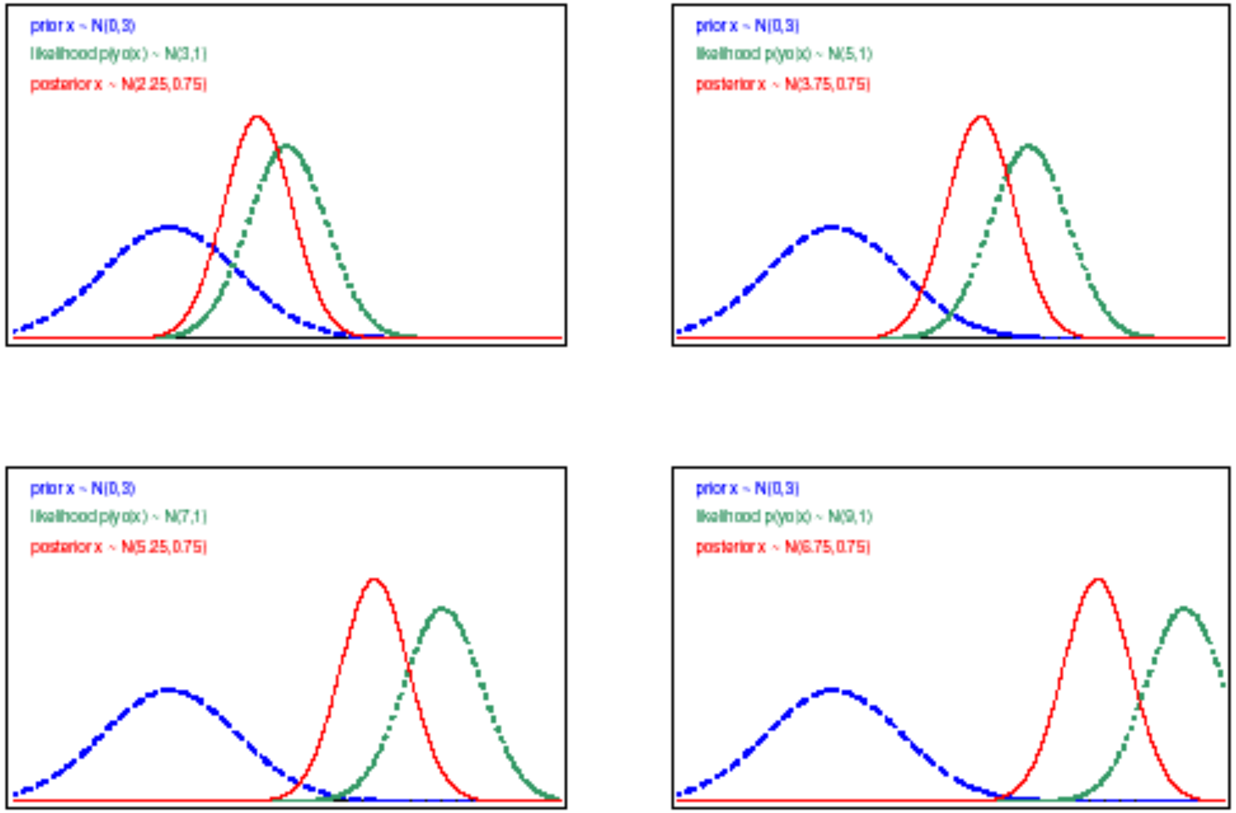


Fig. 1.1: Prior pdf $p(x)$ (dashed line), posterior pdf $p(x|y^o)$ (solid line), and Gaussian likelihood of observation $p(y^o|x)$ (dotted line), plotted against x for various values of y^o. (Adapted from Lorenc and Hammon 1988.)

$z_1 = x + \zeta_1$
$z_2 = x + \zeta_2$

Same as before, but ζ_1 and ζ_2 are now distributed according to exponential law with parameter a, *i. e.*

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

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

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

$\mathrm{Var}(x \mid z_1, z_2) = a^2\,(2\delta^3/3 + \delta^2 + \delta + 1/2) / (1 + 2\delta)$, 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 $\int p \ln p$ always decreases in bayesian estimation)~~

We started from

$$\xi \rightarrow \mathcal{J}(\xi) \equiv (1/2)\,[(z_1 - \xi)^2 / s_1 + (z_2 - \xi)^2 / s_2\,]$$

$$= (1/2)\,(\xi - x^a)^2 / p^a + \ldots$$

$$P(x = \xi \mid z_1, z_2) \propto \exp[-\mathcal{J}(\xi)]$$

Conditional expectation x^a minimizes *objective function* $\mathcal{J}(\xi)$ defined on ξ-space ⇒ *variational assimilation*

In addition

$$p^a = 1/\,\mathcal{J}''(x^a)$$

Estimate

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

with error p^a such that

$$1/p^a = 1/s_1 + 1/s_2$$

can also be obtained, independently of any Gaussian hypothesis, as simply corresponding to the linear combination of z_1 and z_2 that minimizes the error $E\,[(x^a - x)^{\,2}]$

Best Linear Unbiased Estimator (BLUE)

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.

$$z = F(x, \zeta) \qquad (1)$$

where ζ 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

$$z = \Gamma x + \zeta$$

Bayesian estimation (continued)

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

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

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

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

⇔ 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 $n \approx 10^{6\text{-}9}$ of present Numerical Weather Prediction models (the *curse of dimensionality)*.

- 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 $N \approx O(10\text{-}100)$).

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

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

- Covariance matrix

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

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

- Two random vectors

$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$

$\boldsymbol{y} = (y_1, y_2, \ldots, y_p)^T$

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

dimension nxp

Covariance matrices will be denoted

$$C_{\boldsymbol{xx}} \equiv E(\boldsymbol{x}'\boldsymbol{x}'^{\mathrm{T}})$$

$$C_{\boldsymbol{xy}} \equiv E(\boldsymbol{x}'\boldsymbol{y}'^{\mathrm{T}})$$

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

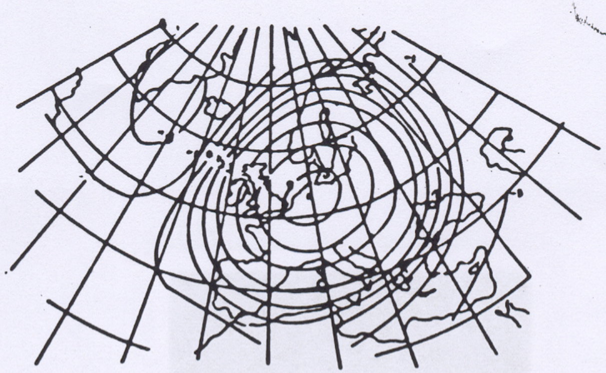
- Expectation $E[\Phi(\xi)]$; $\Phi'(\xi) \equiv \Phi(\xi) - E[\Phi(\xi)]$
- Variance $Var[\Phi(\xi)] = E\{[\Phi'(\xi)]^2\}$

- Covariance function

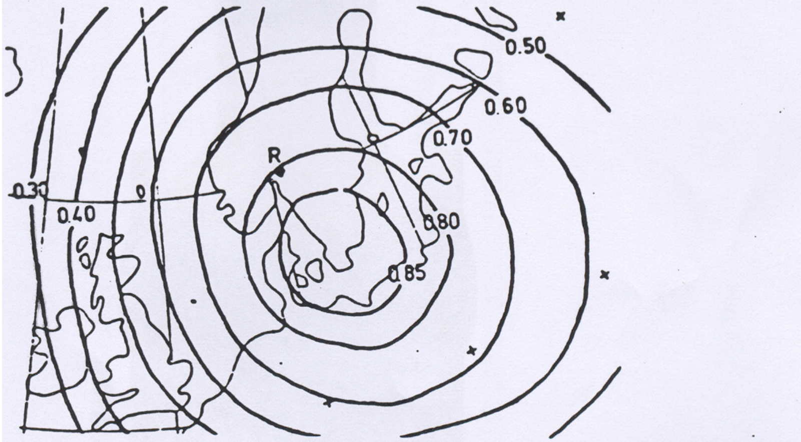
$$(\xi_{1,}\ \xi_2) \rightarrow\ C_\Phi(\xi_{1,}\ \xi_2) \equiv\ E[\Phi'(\xi_1)\ \Phi'(\xi_2)]$$

- 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

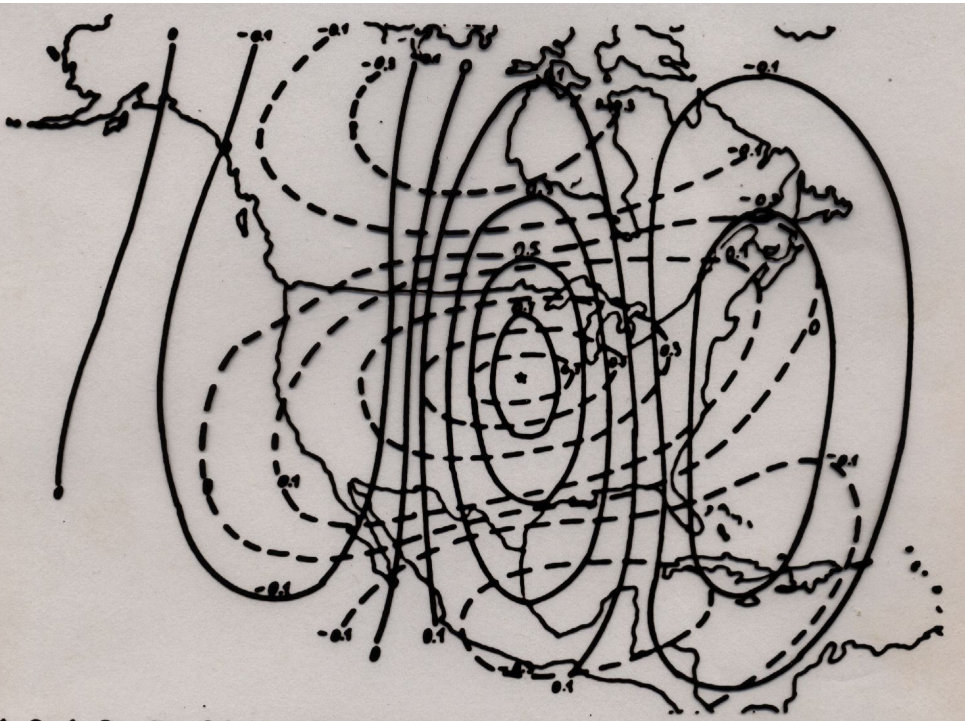


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson

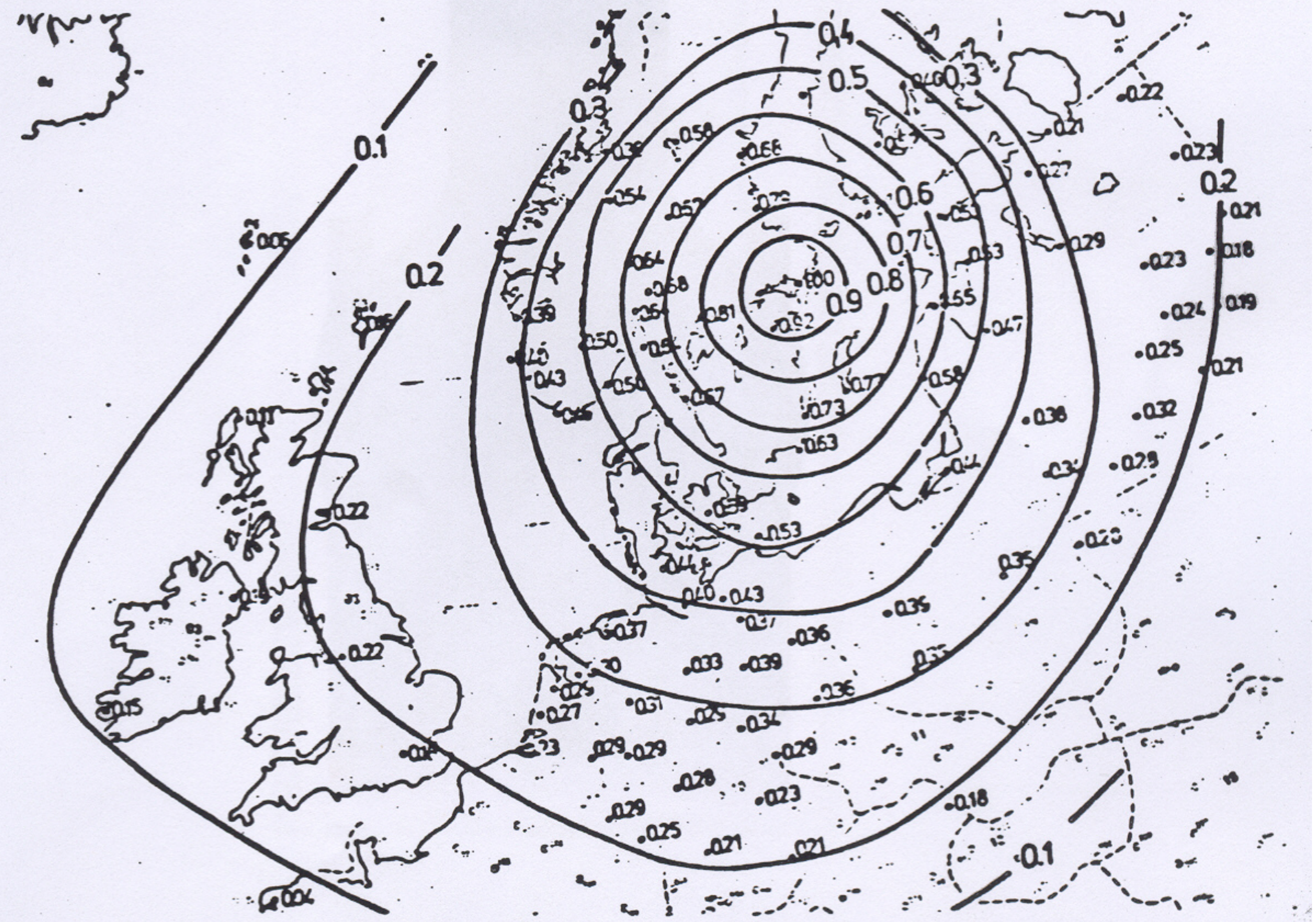


Figure 5.1.1.4.1 Auto-correlation of errors in 12h numerical forecasts of surface pressure in a reference station (Stockholm) and other stations.

After N. Gustafsson

Cours à venir

~~Jeudi 14 Février~~

~~Jeudi 21 Février (**)~~

~~Jeudi 28 Février~~

Jeudi 7 Mars

Vendredi 15 Mars

Jeudi 21 Mars (**)

Jeudi 28 Mars (*)

De 10h00 à 12h30, Département de Géosciences, École Normale Supérieure, 24, rue Lhomond, Paris 5, Salle de la Serre, 5ième étage,

(*) Salle E314, 3ième étage

(**) Salle E350, 3ième étage