École Doctorale des Sciences de l'Environnement d'Île-de-France
Année Universitaire 2024-2025

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand
Cours 3

18 Avril 2025

Bayesian estimation

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

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

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

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

$$\boldsymbol{z} = \boldsymbol{\Gamma x} + \zeta$$

Bayesian estimation (continued)

Probability that $\boldsymbol{x} = \xi$ for given ξ ?

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

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

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

$\Leftrightarrow$ data contain information, either directly or indirectly, on any component of $\boldsymbol{x}$. *Determinacy* condition. Implies $m \geq n$.

Bayesian estimation is actually 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 $N \approx O(10\text{-}100)$).

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

- *Optimal Interpolation*. Principle, simple examples, basic properties.

- *Best Linear Unbiased Estimate* (*BLUE*)

Scalar random variable x

Observed outcome of ‘realizations’ of a process that is repeated a large number of times. And also, *a priori* uncertainty on that result.

For any interval $[a, b]$, the probability $P(a < x < b)$ is known (whether inequalities are strict or not may matter).

Probability density function (pdf). Function $p(\xi)$ such that, for any interval $[a, b]$

$$P[a < x < b] = \int_a^b p(\xi)\, d\xi \qquad \int_{-\infty}^{\infty} p(\xi)\, d\xi = 1$$

($p(\xi)$ may contain diracs)

Expectation. Mean of a large number of realizations of x

$$E(x) \equiv \int_{-\infty}^{\infty} \xi\, p(\xi) d\xi$$

(may not exist)

Scalar random variable x (continued)

Variance

$$Var(x) \equiv E\{[x - E(x)]^2\} = E(x^2) - [E(x)]^2$$

Standard deviation

$$\sigma(x) \equiv \sqrt{Var(x)}$$

Centred variable $x' \equiv x - E(x)$

Couple of random variables $\boldsymbol{x} = (x_1, x_2)^T$

For any intervals $[a_1, b_1]$, $[a_2, b_2]$, probability $P(a_1 < x_1 < b_1 \text{ and } a_2 < x_2 < b_2)$ is known

Extends to any measurable domain $\mathcal{D} \subset R^2$

$$P[(x_1, x_2) \in \mathcal{D}] = \iint_{\mathcal{D}} p(\xi_1, \xi_2)\, d\xi_1 d\xi_2$$

where $p(\xi_1, \xi_2)$ is probability density function

Expectation

$$E(x_1 + x_2) = E(x_1) + E(x_2)$$

Couple of random variables $\boldsymbol{x} = (x_1, x_2)^T$

Covariance

$$Cov(x_1, x_2) \equiv E(x_1' \, x_2')$$

$$Corr(x_1, x_2) \equiv Cov(x_1, x_2) \,/\, (\sigma(x_1)\ \sigma(x_2)) \ = \ \cos\varphi$$

Covariance is a scalar product, and defines Euclidean geometry (on space of finite-variance random variables on a given trial space)

Modulus = standard deviation σ, angle = $\cos^{-1}$ $(Corr)$, orthogonality = decorrelation

If x_1 and x_2 uncorrelated,

$$Var(x_1 + x_2) = Var(x_1) + Var(x_2) \qquad \text{(Pythagorean theorem)}$$

$$E(x_1\, x_2) = E(x_1)\, E(x_2)$$

Couple of random variables $\boldsymbol{x} = (x_1, x_2)^T$ (continued)

Independence

x_1 and x_2 independent : knowledge about either one of the variables brings no knowledge about the other one.

For any intervals $[a_1, b_1]$, $[a_2, b_2]$

$$P(a_1 < x_1 < b_1 \text{ and } a_2 < x_2 < b_2) = P(a_1 < x_1 < b_1)\, P(a_2 < x_2 < b_2)$$

Equivalently, pdf's verify

$$p(\xi_1, \xi_2) = p_1(\xi_1)\, p_2(\xi_2)$$

Independence implies decorrelation. Converse is not true
(consider $S = \sin \alpha$, $C = \cos \alpha$, where α is uniformly distributed over $[0, 2\pi]$)

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

Non-random vector $\boldsymbol{\lambda} = (\lambda_i)_{i = 1, .., n}$

$$G \equiv \Sigma_i \lambda_i x_i' \qquad\qquad G^2 = \Sigma_{i,j} \lambda_i \lambda_j x_i' x_j'$$

$$E(G^2) = \Sigma_{i,j} \lambda_i \lambda_j E(x_i'x_j') = \boldsymbol{\lambda}^T E(\boldsymbol{x}'\boldsymbol{x}'^T) \boldsymbol{\lambda} \geq 0$$

Covariance matrix $E(\boldsymbol{x}'\boldsymbol{x}'^T)$ is symmetric non negative (strictly definite positive except if linear relationship holds between the x_i's with probability 1).

Change

$$\boldsymbol{x} \rightarrow \boldsymbol{y} \equiv P\boldsymbol{x}$$

$$\boldsymbol{y}'\boldsymbol{y}'^{\mathrm{T}} = P\boldsymbol{x}'(P\boldsymbol{x}')^{\mathrm{T}} = P\,\boldsymbol{x}\,\boldsymbol{x}'^{\,\mathrm{T}}\,P^{\mathrm{T}}$$

$$E(\boldsymbol{y}'\boldsymbol{y}'^{\mathrm{T}}) = P\,E(\boldsymbol{x}'\boldsymbol{x}'^{\mathrm{T}})\,P^{\mathrm{T}}$$

In change $\boldsymbol{x} \rightarrow \boldsymbol{y}$, eigenvalues of covariance matrix remain > 0, but can be modified (conserved if $P^{\mathrm{T}} = P^{-1}$, orthogonal matrix).

Eigenvalues can actually take any positive values.

In particular, covariance matrix can be made equal to the unit matrix, for instance in the basis of *principal components*.

- Two random vectors

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$$
$$\boldsymbol{z} = (z_1, z_2, \ldots, z_p)^{\mathrm{T}}$$

$$E(\boldsymbol{x}'\boldsymbol{z}'^{\mathrm{T}}) = E(x_i' z_j')$$

dimension nxp

Change

$$\boldsymbol{x} \rightarrow \boldsymbol{u} \equiv A\boldsymbol{x} \qquad \boldsymbol{z} \rightarrow \boldsymbol{v} \equiv B\boldsymbol{z}$$

$$E(\boldsymbol{u}'\boldsymbol{v}'^{\mathrm{T}}) = A\, E(\boldsymbol{x}'\boldsymbol{z}'^{\mathrm{T}})\, B^{\mathrm{T}}$$

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) (aka *stochastic process* if function of time)

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

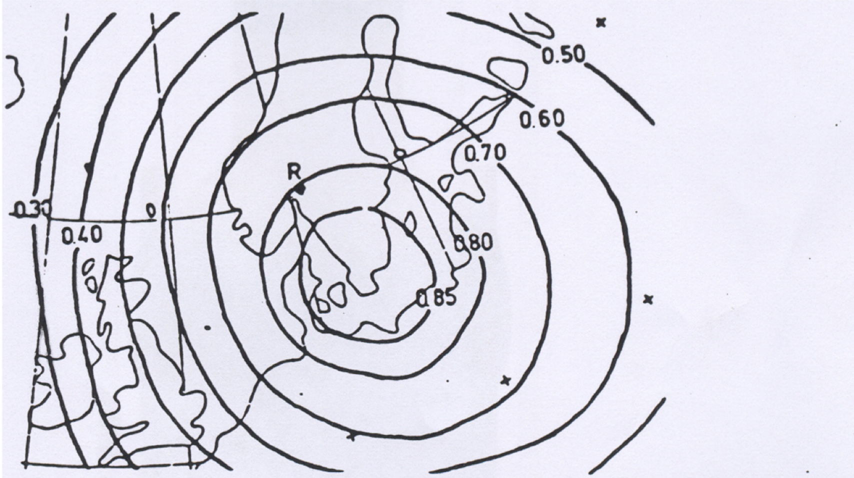
- Covariance function

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

- 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

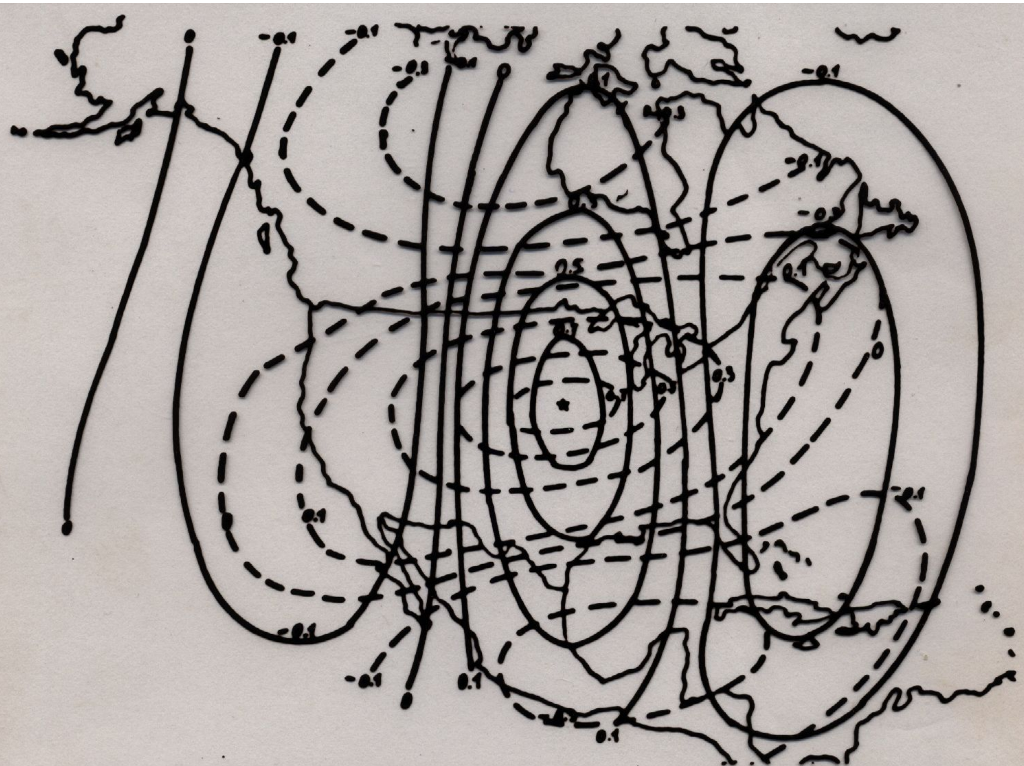


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson

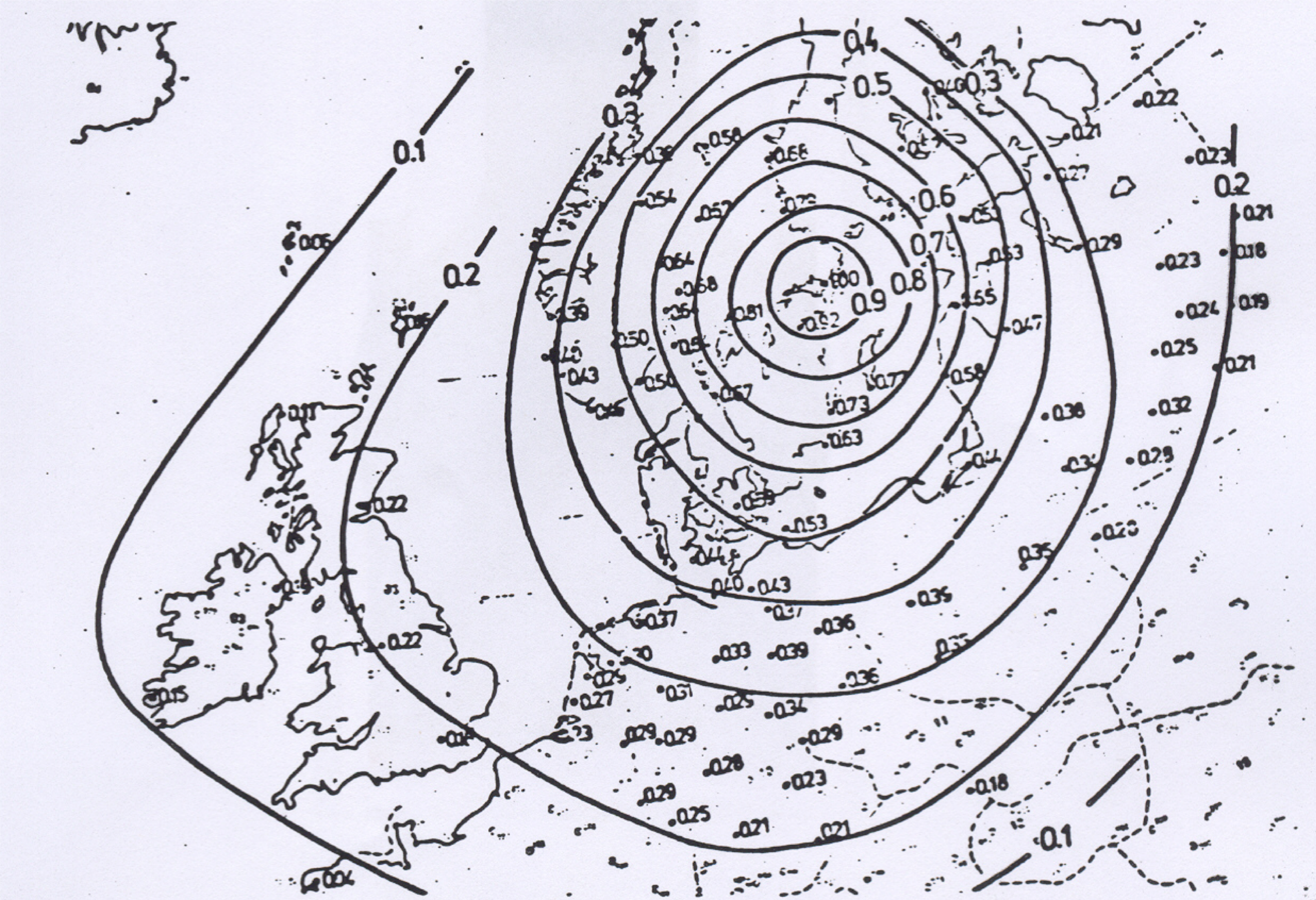


Figure 5.1.1.4.1 Auto-correlation of errors in 12h numerical forecasts of surface pressure in a reference station (Stockholm) and other stations.

After N. Gustafsson

Covariance function can be

homogeneous $C_{\Phi}(\xi_1, \xi_2) = H(\xi_1 - \xi_2)$

or *isotropic* $C_{\Phi}(\xi_1, \xi_2) = K(|\xi_1 - \xi_2|)$

(on the sphere, no difference)

N points $\xi_1, \xi_2, \ldots, \xi_N$ in state space

N non-random coefficients $\lambda_1, \lambda_2, \ldots, \lambda_N$

$$G \equiv \Sigma_i \lambda_i \Phi'(\xi_i)$$

$$E(G^2) = \Sigma_{i,j} \lambda_i \lambda_j C_{\Phi}(\xi_i, \xi_j) \geq 0$$

$$E(G^2) = \Sigma_{i,j}\, \lambda_i\, \lambda_j\, C_\Phi(\xi_i,\, \xi_j) \geq 0$$

covariance functions are of *positive type* (or *definite positive*). Conversely, a function of positive type can be shown to be the covariance function of a random function.

Example

On a circle, function $C(\xi_1, \xi_2) = \cos(\xi_1 - \xi_2)$ is covariance function of random function $\Phi(\xi) = 2\cos(\xi + \alpha)$, where α is uniformly distributed over $[0, 2\pi]$.

More generally, random function on 2π-circle of the form

$$\Phi(\xi) = \Sigma_{k=-K, +K} \; \phi_k \exp(ik\xi)$$

with $\phi_k = \rho_k \exp(i\theta_k)$, ρ_k real, $k \geq 0$, $\phi_{-k} = \rho_k \exp(-i\theta_k)$

All ρ_k and θ_k random, the θ_k's being uniformly distributed over $[0, 2\pi]$, mutually independent, and independent of the ρ_k ‘s.

$\Phi(\xi)$ is the superposition of a spatially uniform random ρ_0 (we assume $E(\rho_0)=0$) and of K sine waves with random and mutually independent (uniformy distributed) phases θ_k and amplitudes ρ_k.

$$\Phi'(\xi_1)\ \Phi'(\xi_2) = [\Sigma_k\, \rho_k \exp(i\theta_k) \exp(ik\xi_1)]$$

$$\times\ [\Sigma_{k'}\, \rho_{k'} \exp(-i\theta_{k'}) \exp(-ik'\xi_2)]$$

$$= \Sigma_{kk'}\, \rho_k\, \rho_{k'} \exp[i(\theta_k - \theta_{k'})] \exp[i\,(k\xi_1 - k'\xi_2)]$$

On taking expectation, $E[\exp[i(\theta_k - \theta_{k'})] = 0$ if $k \neq k'$ and there remains

$$E[\Phi'(\xi_1)\ \Phi'(\xi_2)] = C_\Phi(\xi_1, \xi_2) = \Sigma_k\, E(\rho_k^{\,2}) \exp[ik(\xi_1 - \xi_2)]$$

$$C_\Phi(\xi_1, \xi_2) = E(\rho_0^{\,2}) + 2\, \Sigma_{k>0}\, E(\rho_k^{\,2}) \cos\, [k(\xi_1 - \xi_2)]$$

Bochner-Khintchin theorem. Homogeneous function $C(\xi_1, \xi_2) = H(\xi_1 - \xi_2)$ over R^n of positive type ⇔ Fourier Transform of H is real ≥ 0.

In R^n, squared exponential

$$C(\xi_1, \xi_2) = \exp[-(\xi_1 - \xi_2)^T B^{-1} (\xi_1 - \xi_2)] \quad B > 0$$

is of positive type

Gaussian variables

Unidimensional (expectation m, variance a)

$$\mathcal{N}[m, a] \sim (2\pi a)^{-1/2} \exp [- (1/2a) (\xi - m)^2]$$

Dimension n (expectation $\boldsymbol{m}$, covariance matrix $\boldsymbol{A}$)

$$\mathcal{N}[\boldsymbol{m}, \boldsymbol{A}] \sim$$
$$[(2\pi)^n \det \boldsymbol{A}]^{-1/2} \exp [- (1/2) (\boldsymbol{\xi} - \boldsymbol{m})^T \boldsymbol{A}^{-1} (\boldsymbol{\xi} - \boldsymbol{m})]$$

Gaussian variables

Gaussian couple $\boldsymbol{z} = (\boldsymbol{x}^{\mathrm{T}}, \boldsymbol{y}^{\mathrm{T}})^{\mathrm{T}}$ with distribution $\mathcal{N}[0, \boldsymbol{C}]$

pdf $\sim$ $\exp\left[-(1/2)\, \boldsymbol{z}^{\mathrm{T}} \boldsymbol{C}^{-1} \boldsymbol{z}\right]$

$$\boldsymbol{C} = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix}$$

$\boldsymbol{x}$ and $\boldsymbol{y}$ uncorrelated $\boldsymbol{C}_{xy} = 0$, $\boldsymbol{C}_{yx} = 0$

$$\boldsymbol{C}^{-1} = \begin{pmatrix} \boldsymbol{C}_{xx}{}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{yy}{}^{-1} \end{pmatrix}$$

$$\boldsymbol{z}^{\mathrm{T}} \boldsymbol{C}^{-1} \boldsymbol{z} = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{C}_{xx}{}^{-1} \boldsymbol{x} + \boldsymbol{y}^{\mathrm{T}} \boldsymbol{C}_{yy}{}^{-1} \boldsymbol{y}$$

Gaussian variables

$$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{z} = \boldsymbol{x}^{\mathrm{T}}\,\boldsymbol{C}_{xx}^{-1}\,\boldsymbol{x} + \boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{C}_{yy}^{-1}\,\boldsymbol{y}$$

$$\exp\,[-\,(1/2)\,\boldsymbol{z}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{z}] = \exp\,[-\,(1/2)\,\boldsymbol{x}^{\mathrm{T}}\,\boldsymbol{C}_{xx}^{-1}\,\boldsymbol{x}\,]\,\exp\,[-\,(1/2)\,\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{C}_{yy}^{-1}\,\boldsymbol{y}]$$

$$p(\boldsymbol{z}) = p(\boldsymbol{x})\,p(\boldsymbol{y})$$

For globally Gaussian variables, decorrelation implies independence

- ‘Optimal Interpolation’. Basic theory and basic properties. A simple example.

Optimal Interpolation

x ξ_1 x ξ_3

X ξ

x ξ_2 x ξ_5

x ξ_4

Random field $\Phi(\xi)$, with known probability distribution

Observations y_j at points ξ_j , $j = 1, \ldots, p$

Value $x = \Phi(\xi)$ at point ξ ?

Optimal Interpolation (continued 1)

Random field $\Phi(\xi)$

Observation network $\xi_1, \xi_2, \ldots, \xi_p$

For one particular realization of the field, observations

$y_j = \Phi(\xi_j) + \varepsilon_j \quad , \ j = 1, \ldots, p \quad ,$ making up vector $\boldsymbol{y} = (y_j)$

Estimate $x = \Phi(\xi)$ at given point ξ, in the form

$$x^a = \alpha + \Sigma_j \beta_j y_j = \alpha + \boldsymbol{\beta}^T \boldsymbol{y}, \text{ where } \boldsymbol{\beta} = (\beta_j)$$

α and the β_j's being determined so as to minimize the expected quadratic estimation error $E[(x-x^a)^2]$

Optimal Interpolation (continued 2)

$E[(x\text{-}x^a)^2]$ minimum $\Rightarrow$ $E(x\text{-}x^a) = 0$ Estimate x^a is unbiased.

$$x^a = \alpha + \Sigma_j \beta_j y_j$$

$$E(x^a) = \alpha + \Sigma_j \beta_j E(y_j)$$

$$x^a - E(x) = \Sigma_j \beta_j [y_j - E(y_j)]$$

Computations are to be made on centred variables

$x'^a \equiv x^a - E(x)$ is the linear combination of the $y_j' = y_j - E(y_j)$ that minimizes the distance to $x' = x - E(x)$. It is the *orthogonal projection, in the sense of covariance*, of x' onto the space spanned by the y_j''s.

Optimal Interpolation (continued 3)

$x' - x'^a$ uncorrelated with y_j'

$$E[(x' - x'^a)\, y_j'] = 0$$
$$x'^a = \Sigma_k\, \beta_k y_k'$$

$$\Rightarrow \quad \Sigma_k\, \beta_k E(y_k'\, y_j') = E(x'\, y_j')$$

in matrix form $\quad C_{yy}\, \boldsymbol{\beta} = C_{yx}$

Optimal Interpolation (continued 4)

Solution

$$x^a = E(x) + E(x'\boldsymbol{y}'^{\mathrm{T}})\,[E(\boldsymbol{y}'\boldsymbol{y}'^{\mathrm{T}})]^{-1}\,[\boldsymbol{y} - E(\boldsymbol{y})]$$
$$= E(x) + \boldsymbol{C}_{xy}\,[\boldsymbol{C}_{yy}]^{-1}\,[\boldsymbol{y} - E(\boldsymbol{y})]$$

i. e.,

$$\boldsymbol{\beta}^{\mathrm{T}} = C_{xy}\,[C_{yy}]^{-1}$$
$$\alpha = E(x) - \boldsymbol{\beta}^{\mathrm{T}} E(\boldsymbol{y})$$

Estimate is unbiased $E(x\text{-}x^a) = 0$

Minimized quadratic estimation error

$$E[(x\text{-}x^a)^2] = E(x'^2) - E[(x'^a)^2]$$
$$= \boldsymbol{C}_{xx} - \boldsymbol{C}_{xy}\,[\boldsymbol{C}_{yy}]^{-1}\,\boldsymbol{C}_{yx}$$

Estimation made in terms of deviations x' and $\boldsymbol{y}'$ from expectations $E(x)$ and $E(\boldsymbol{y})$.

Optimal Interpolation (continued 5)

$$x^a = E(x) + E(x'\mathbf{y}'^{\mathrm{T}})\,[E(\mathbf{y}'\mathbf{y}'^{\mathrm{T}})]^{-1}\,[\mathbf{y} - E(\mathbf{y})]$$

$$y_j = \Phi(\boldsymbol{\xi}_j) + \varepsilon_j$$

$$E(y_j'y_k') = E\,\{[\Phi'(\boldsymbol{\xi}_j) + \varepsilon_j'][\Phi'(\boldsymbol{\xi}_k) + \varepsilon_k']\}$$

If observation errors ε_j are mutually uncorrelated, have common variance r, and are uncorrelated with field Φ, then

$$E(y_j'y_k') = C_\Phi(\boldsymbol{\xi}_j, \boldsymbol{\xi}_k) + r\delta_{jk}$$

and

$$E(x'y_j') = C_\Phi(\boldsymbol{\xi}, \boldsymbol{\xi}_j)$$

Optimal Interpolation (continued 6)

Unique observation (p=1) $y_1 = \Phi(\xi_1) + \varepsilon_1$

Value $x = \Phi(\xi)$ at some point ξ to be estimated
(all values assumed to be centred)

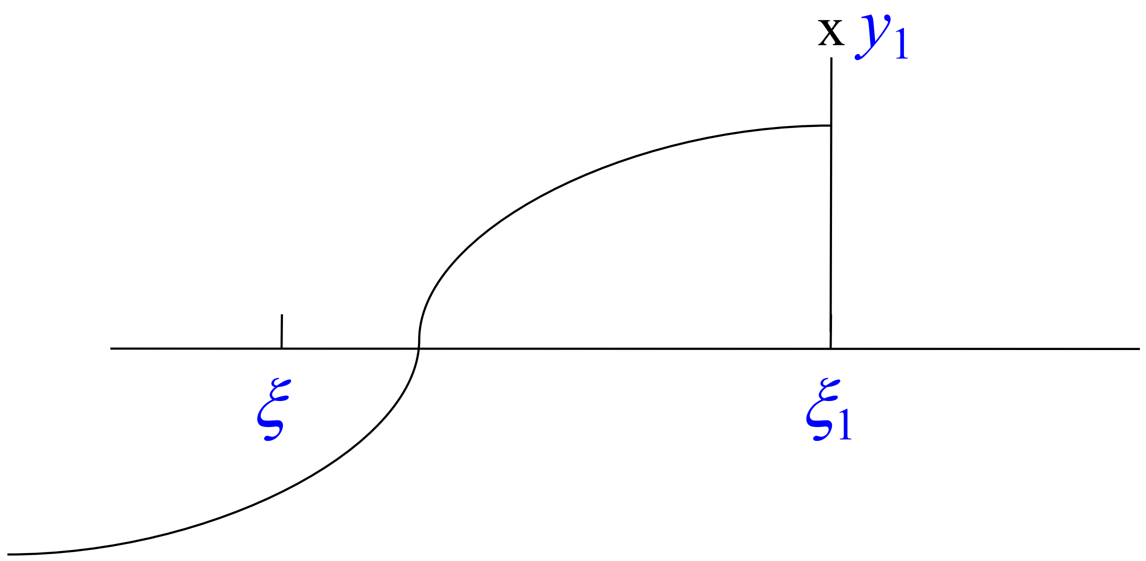
$$C_{yy}\,\beta = C_{yx}$$

$C_{yy} = E(y_1{}^2) = C_\Phi(\xi_1, \xi_1) + r$ $C_{yx} = C_\Phi(\xi, \xi_1)$

$$x^a = \Phi^a(\xi) = \frac{C_\Phi(\xi, \xi_1)}{C_\Phi(\xi_1, \xi_1) + r}\, y_1$$

Optimal Interpolation (continued 7)

$$x^a = \Phi^a(\xi) = \frac{C_\Phi(\xi, \xi_1)}{C_\Phi(\xi_1, \xi_1) + r} y_1$$



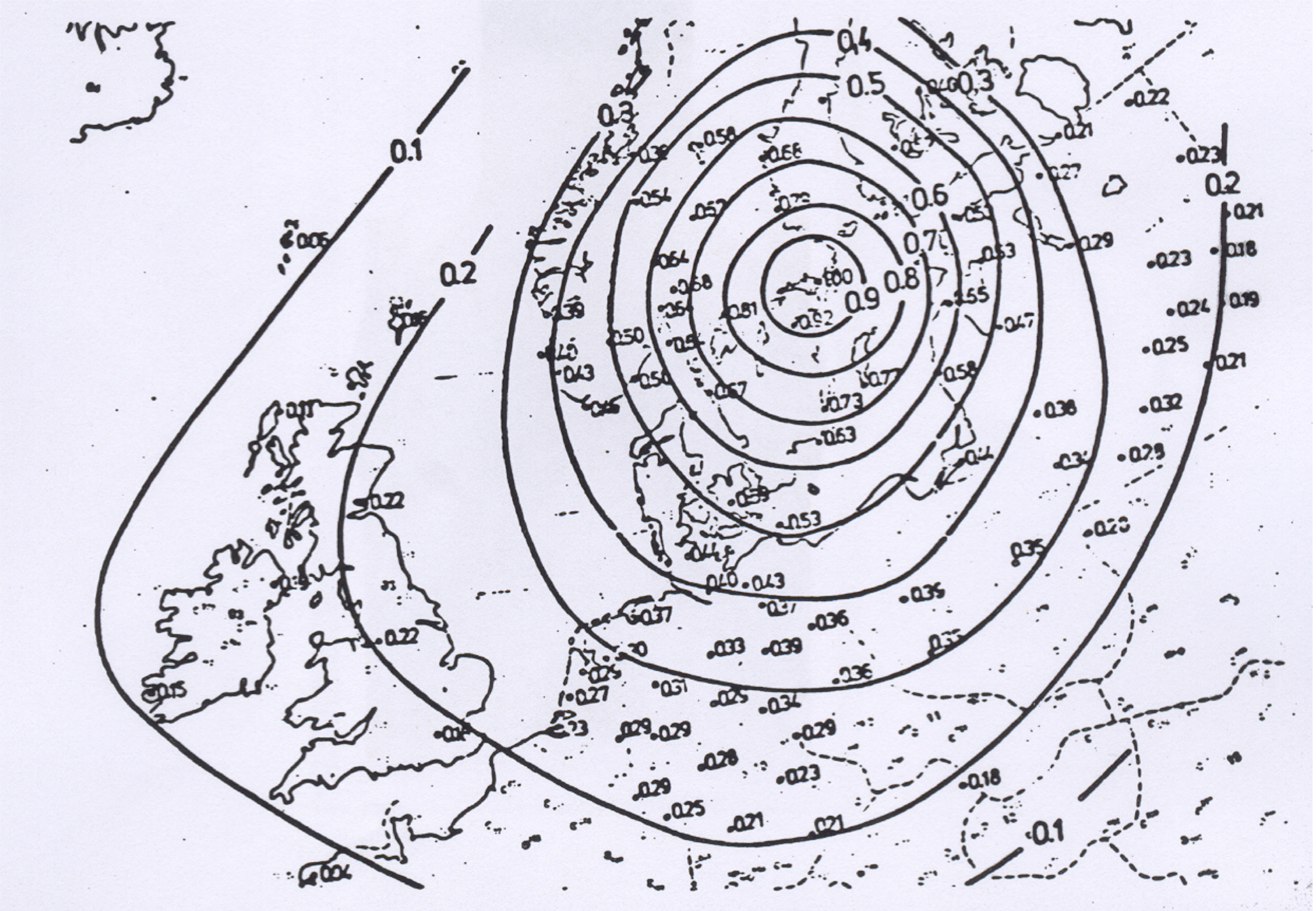
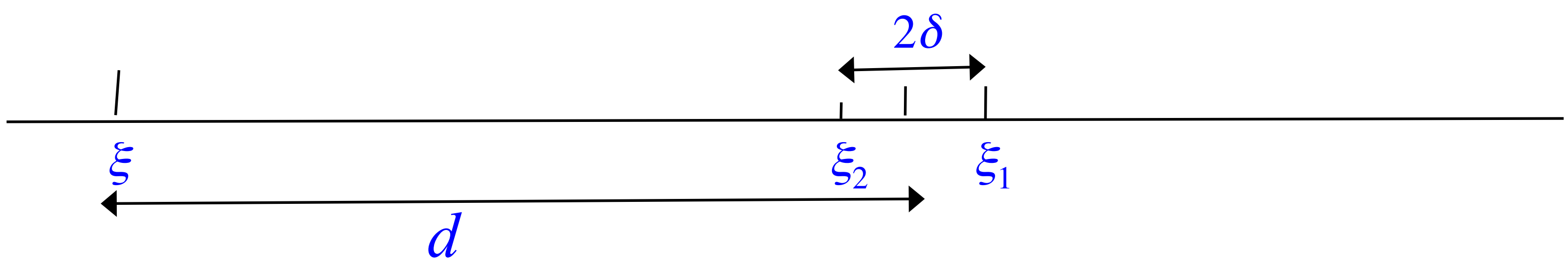


Figure 5.1.1.4.1 Auto-correlation of errors in 12h numerical forecasts of surface pressure in a reference station (Stockholm) and other stations.

After N. Gustafsson

Optimal Interpolation (continued 8)

Two mutually close observations (p=2) $y_j = \Phi(\xi_j) + \varepsilon_j \quad , \; j = 1,2$



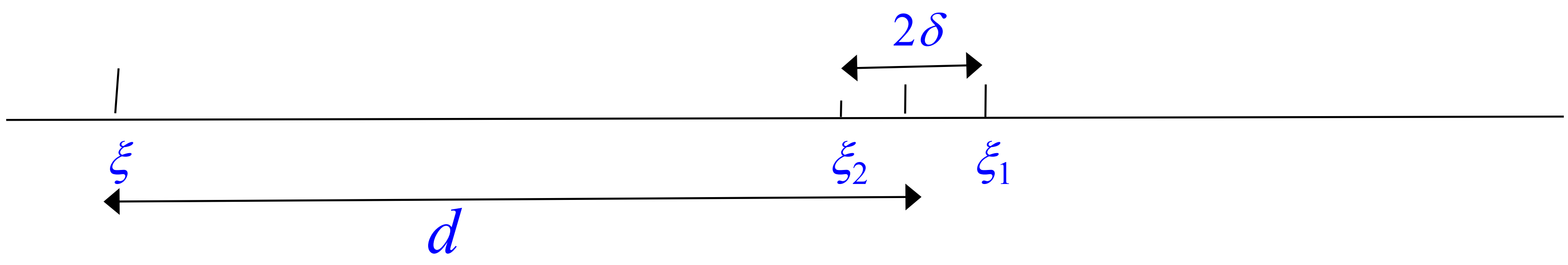
Homogeneous covariance function $C_\Phi(\chi_1, \chi_2) = \Gamma(\chi_1 - \chi_2)$

Linear system for weights β_j's

$$\begin{pmatrix} \Gamma(0) + r & \Gamma(2\delta) \\ \Gamma(2\delta) & \Gamma(0) + r \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \Gamma(d + \delta) \\ \Gamma(d - \delta) \end{pmatrix}$$

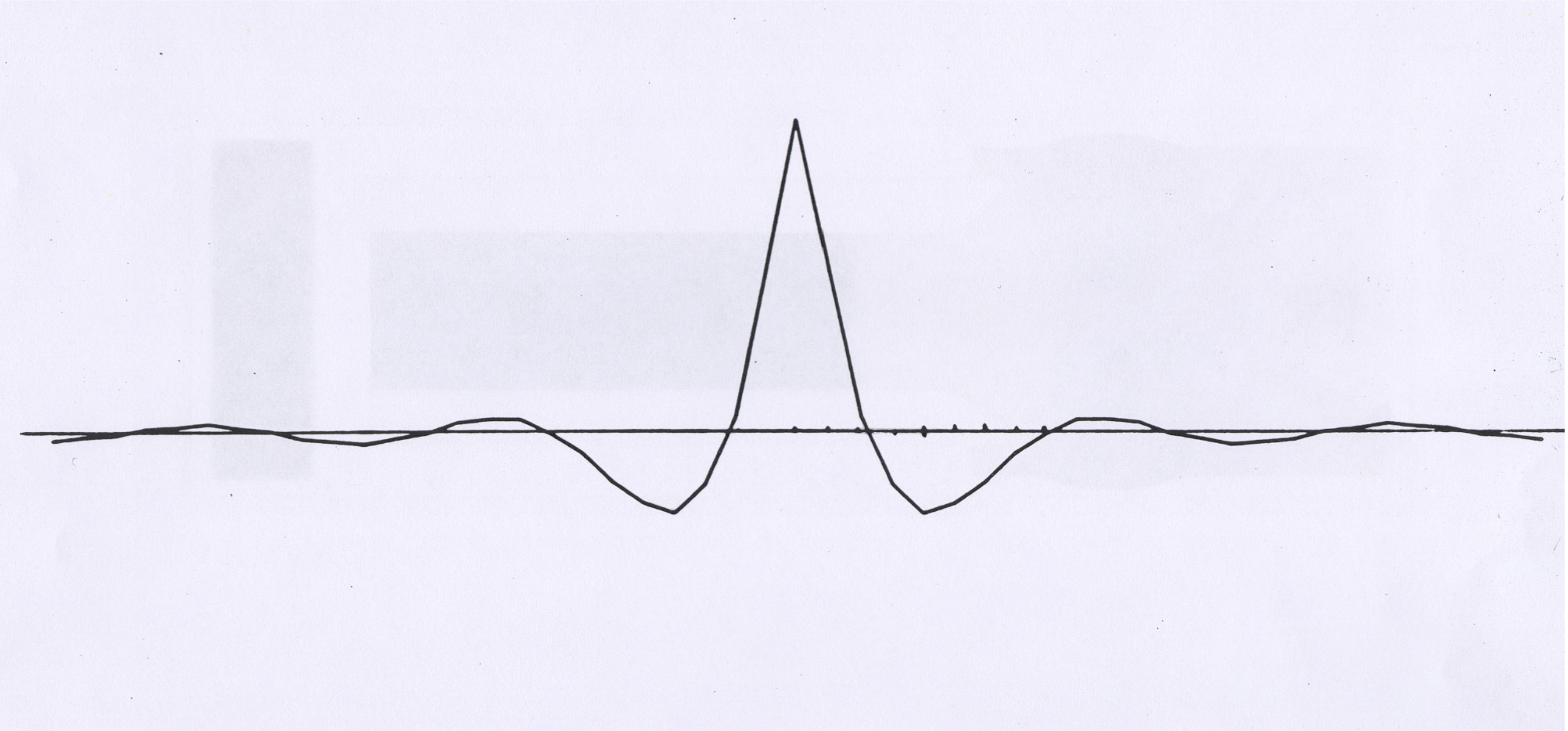
Optimal Interpolation (continued 9)

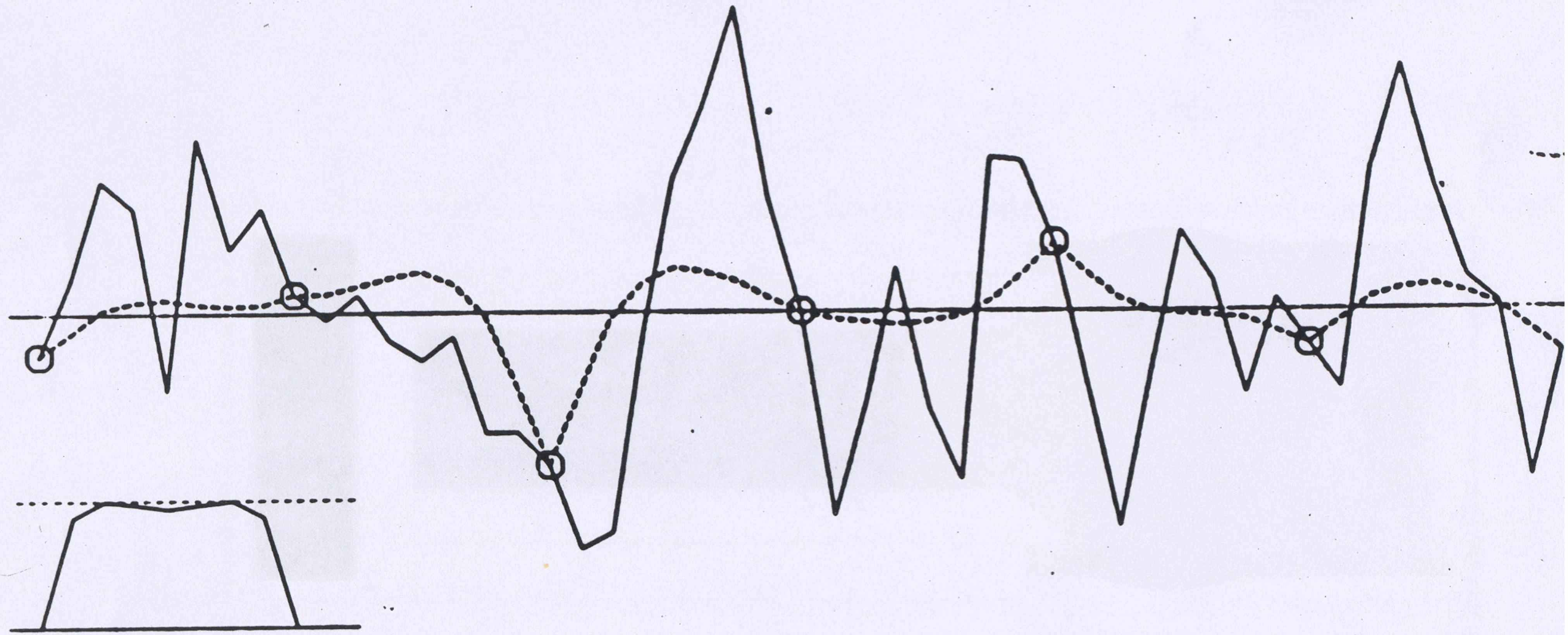
Two mutually close observations (p=2) $y_j = \Phi(\xi_j) + \varepsilon_j \ , \ j = 1,2$

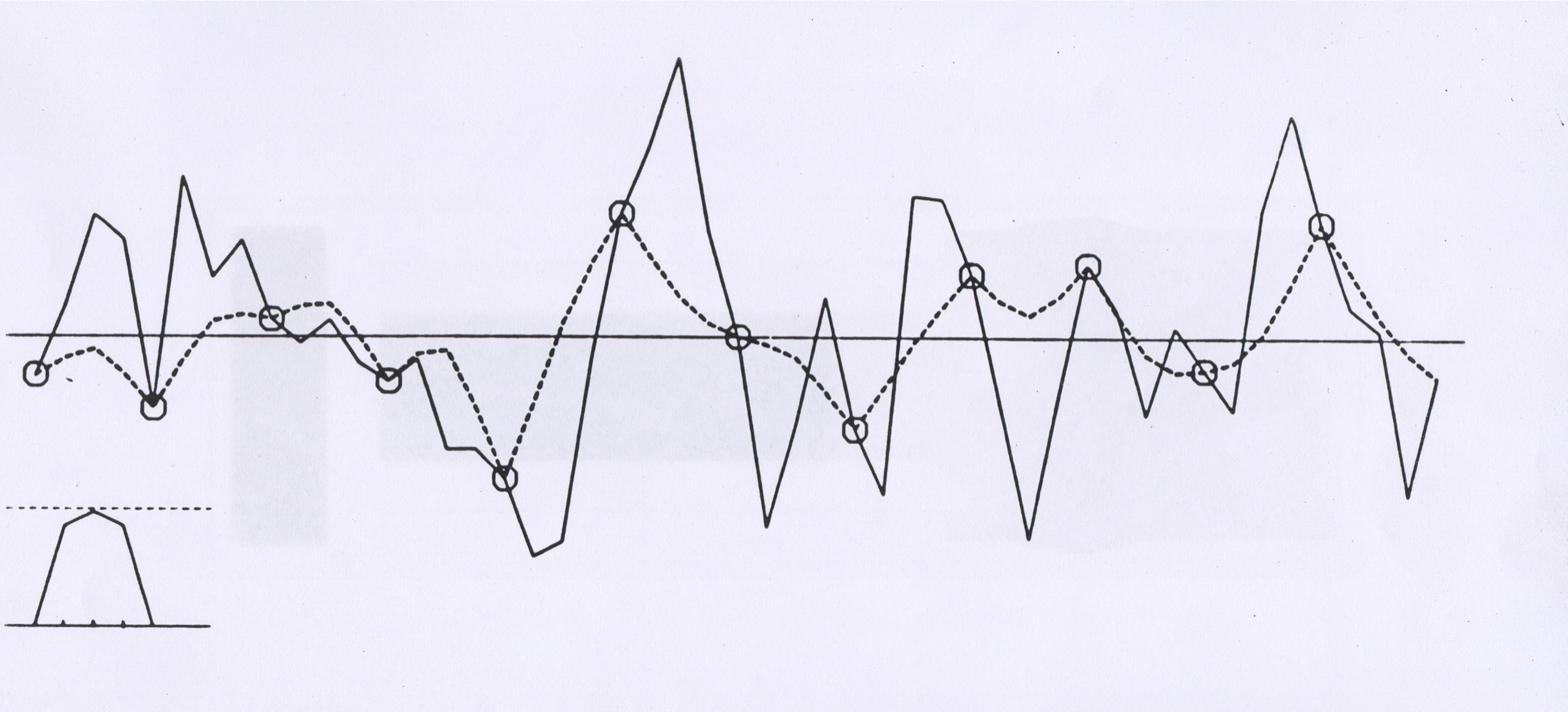


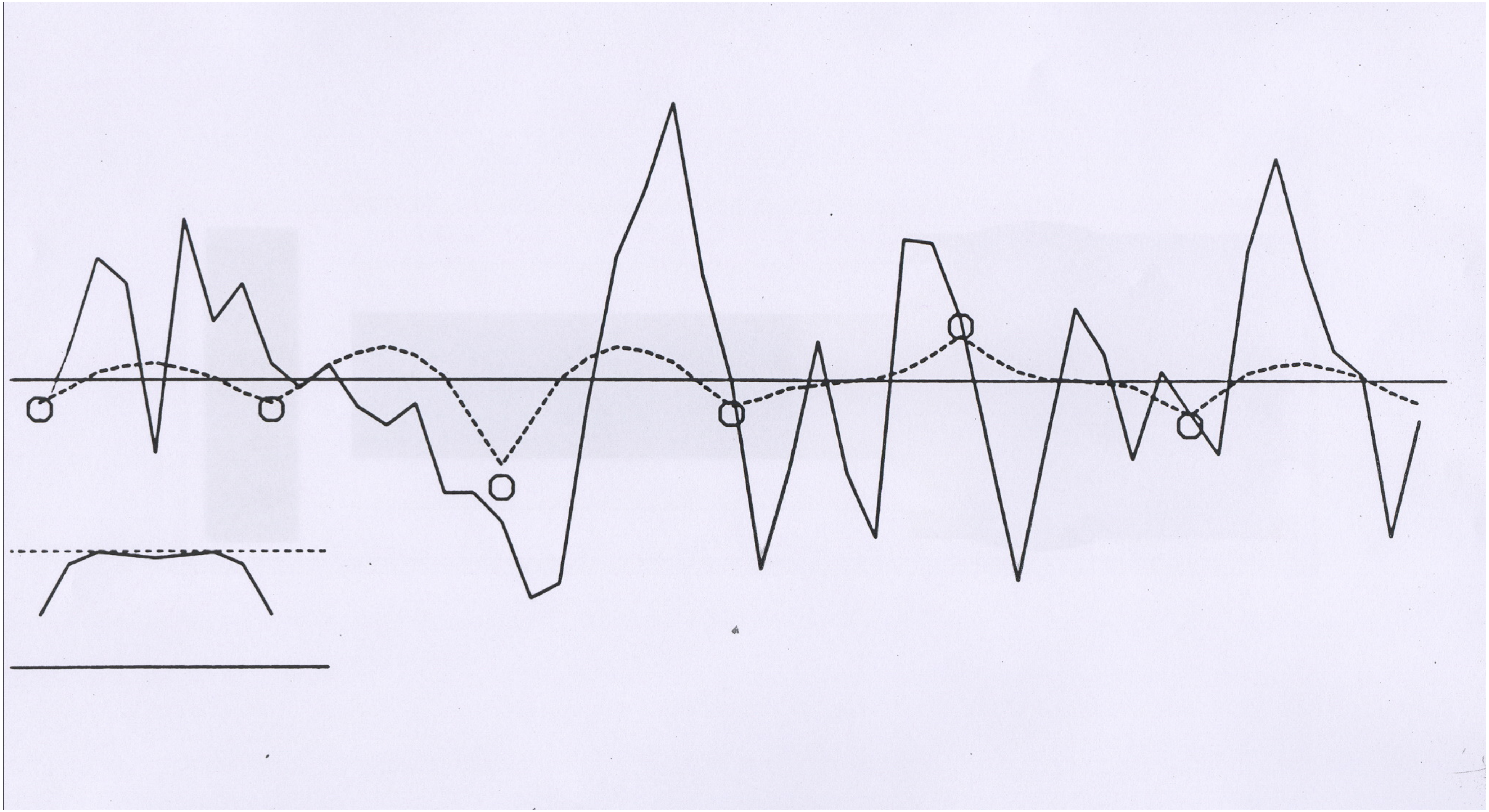
$$\beta_1 + \beta_2 = \frac{\Gamma(d+\delta)+\Gamma(d-\delta)}{\Gamma(0)+\Gamma(2\delta)+r}$$

$$\beta_1 + \beta_2 = \frac{\Gamma(d)}{\Gamma(0)+r/2}$$









Optimal Interpolation (continued 10)

$$x^a = E(x) + \boldsymbol{C}_{x\boldsymbol{y}} [\boldsymbol{C}_{\boldsymbol{yy}}]^{-1} [\boldsymbol{y} - E(\boldsymbol{y})]$$

Vector

$$\boldsymbol{\mu} = (\mu_j) \equiv [\boldsymbol{C}_{\boldsymbol{yy}}]^{-1} [\boldsymbol{y} - E(\boldsymbol{y})]$$

is independent of variable to be estimated

$$x^a = E(x) + \Sigma_j \, \mu_j \, E(x'y_j')$$

Optimal Interpolation (continued 11)

$$x^a = E(x) + \Sigma_j\, \mu_j\, E(x'y_j')$$

$$\Phi^a(\xi) = E[\Phi(\xi)] + \Sigma_j\, \mu_j\, E[\Phi'(\xi)\, y_j']$$

Under hypotheses made above, $E[\Phi'(\xi)\, y_j'] = C_\Phi(\xi, \xi_j)$

$$\Phi^a(\xi) = E[\Phi(\xi)] + \Sigma_j\, \mu_j\, C_\Phi(\xi, \xi_j)$$

Correction made on background expectation is a linear combination of the p functions $C_\Phi(\xi, \xi_j)$

$C_\Phi(\xi, \xi_j)$, considered as a function of estimation position ξ, is the *representer* associated with observation y_j.

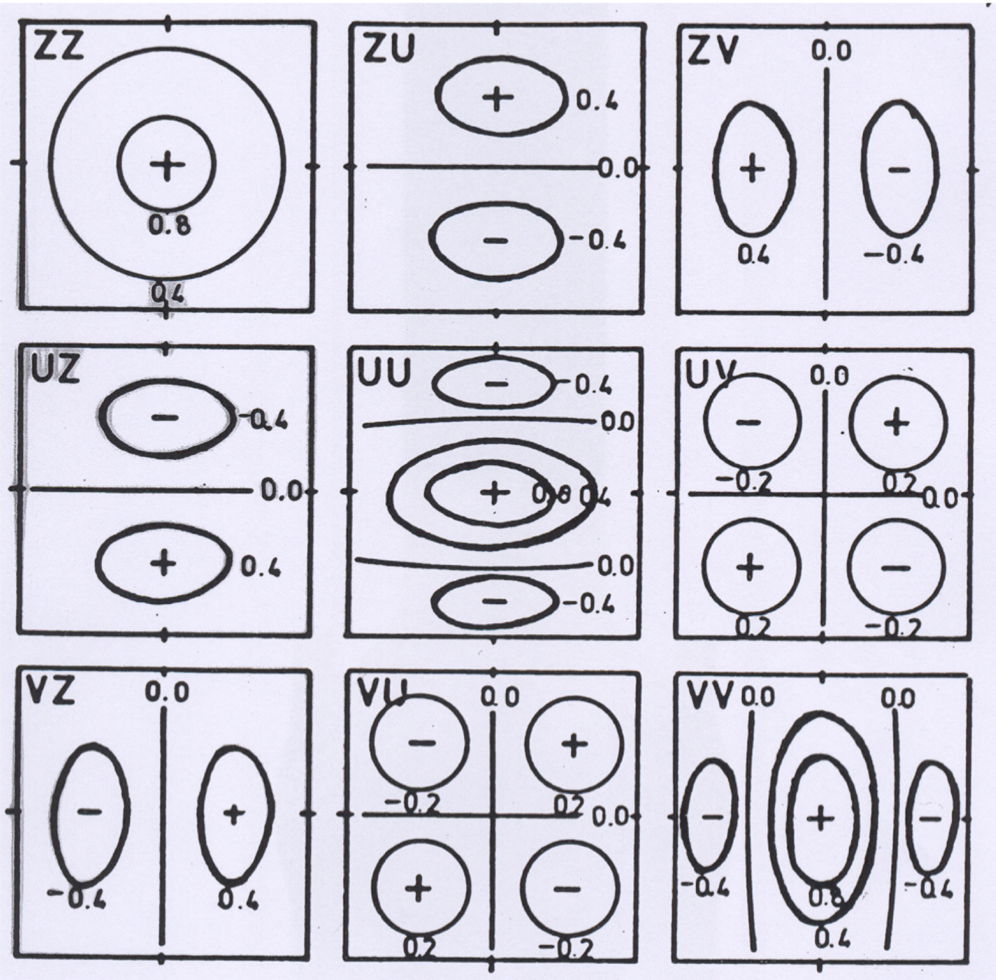
Optimal Interpolation (continued 12)

Univariate interpolation. Each physical field (*e*. *g*. temperature) determined from observations of that field only.

Multivariate interpolation. Observations of different physical fields are used simultaneously. Requires specification of cross-covariances between various fields.

Cross-covariances between mass and velocity fields can simply be modelled on the basis of geostrophic balance.

Cross-covariances between humidity and temperature (and other) fields still a problem.



4.: Schematic illustration of correlation functions and cross-correlation functions for multi-variate analysis derived by the geostrophic assumption.

After N. Gustafsson

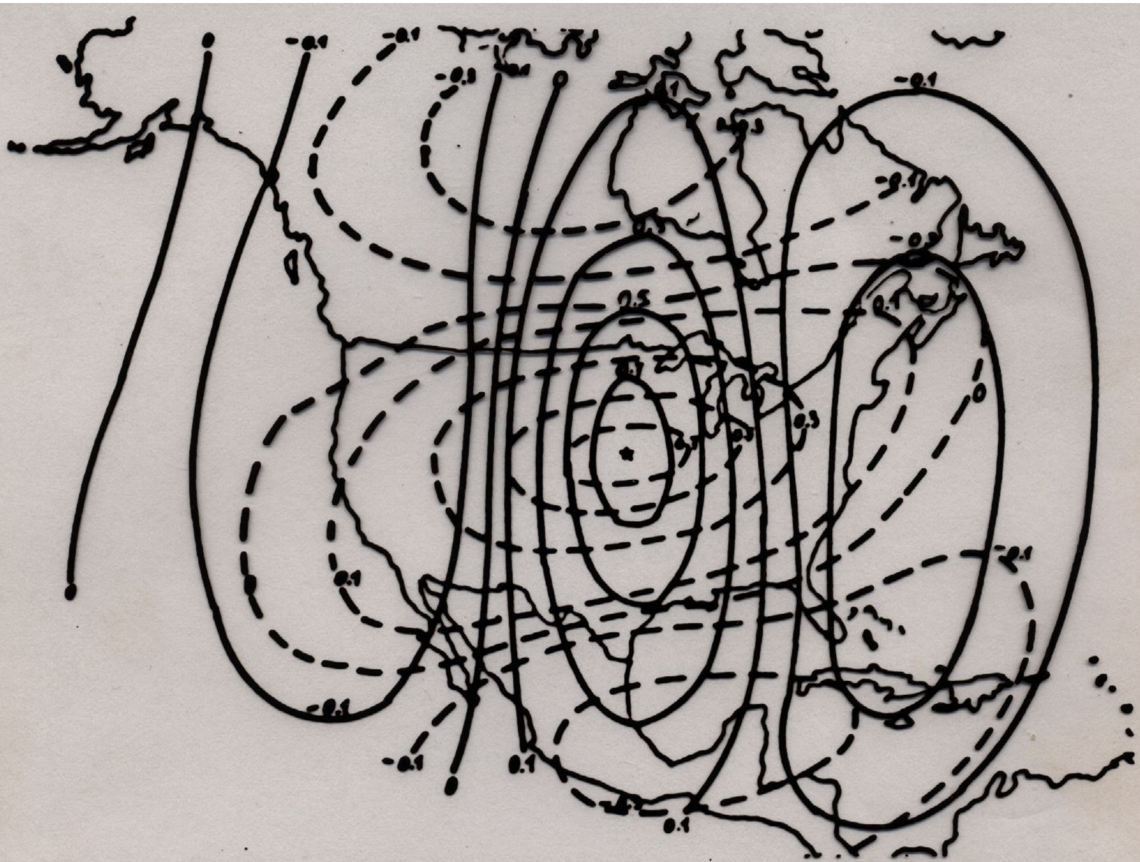


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson

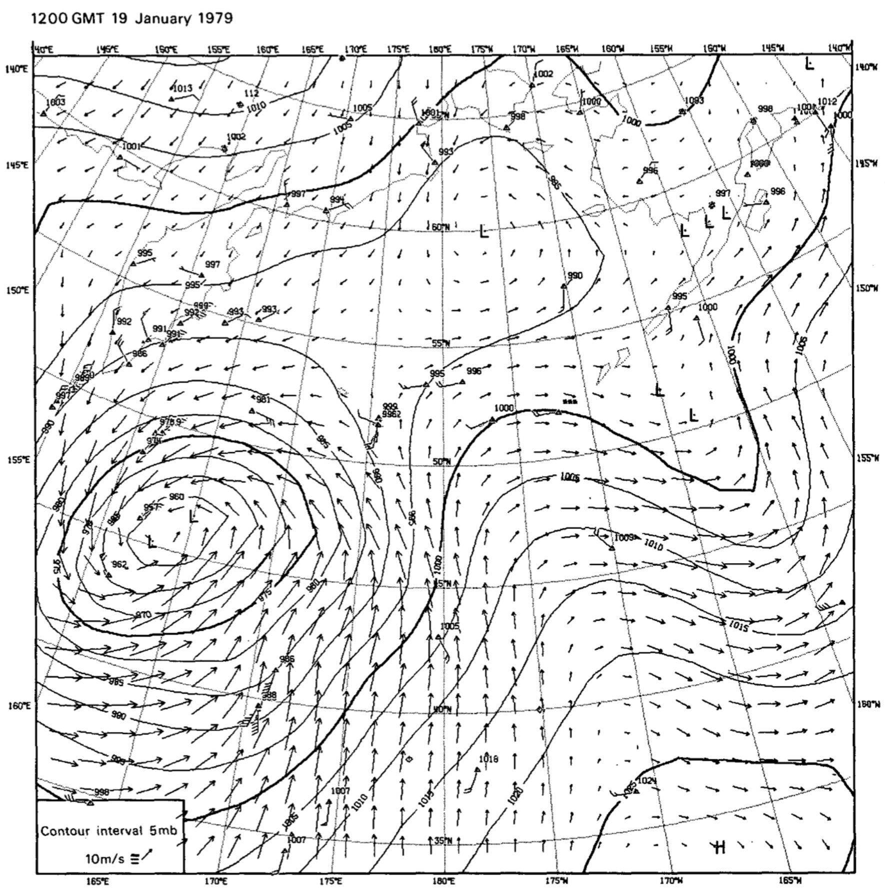


FIG. 14. Sea level pressure and wind forecast corresponding to the central area of Fig. 11, with plotted surface observations of sea level pressure and wind (each barb = 5 m s⁻¹).

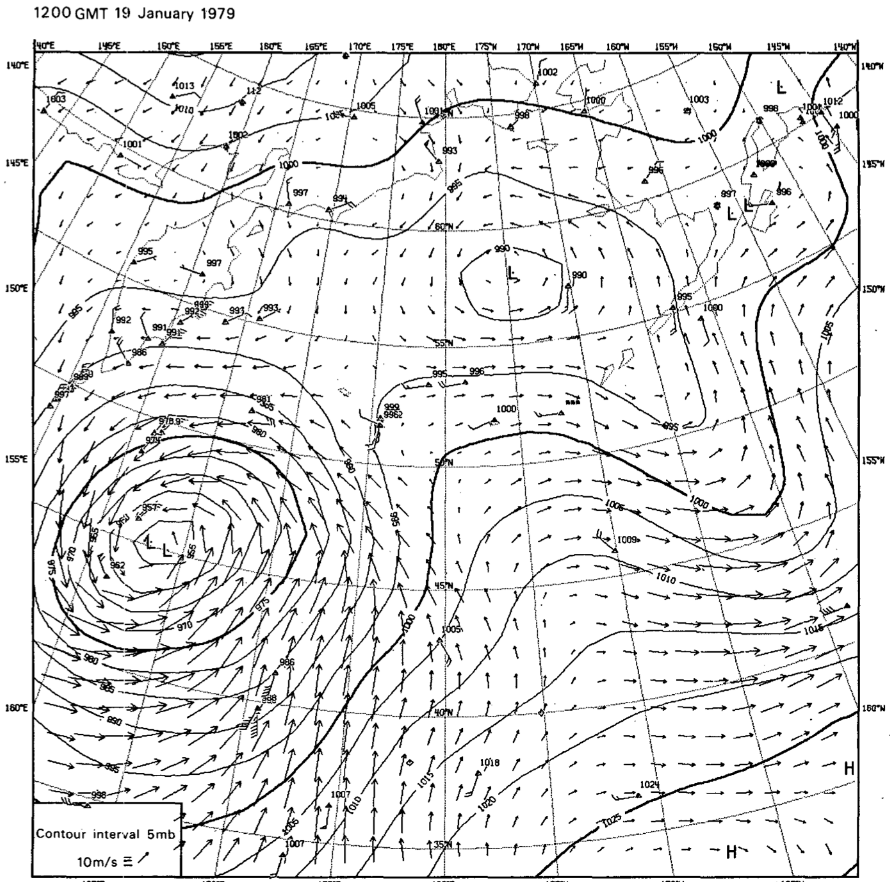


FIG. 15. As in Fig. 14 for the analysis in the data-assimilation cycle.

After A. Lorenc, MWR, 1981

Optimal Interpolation (continued 13)

Observation vector $\boldsymbol{y}$

Estimation of a scalar x

$$x^a = E(x) + \boldsymbol{C}_{xy} [\boldsymbol{C}_{yy}]^{-1} [\boldsymbol{y} - E(\boldsymbol{y})]$$

$$p^a \equiv E[(x-x^a)^2] = E(x'^2) - E[(x'^a)^2])$$
$$= C_{xx} - \boldsymbol{C}_{xy} [\boldsymbol{C}_{yy}]^{-1} \boldsymbol{C}_{yx}$$

Estimation of a vector $\boldsymbol{x}$

$$\boldsymbol{x}^a = E(\boldsymbol{x}) + \boldsymbol{C}_{xy} [\boldsymbol{C}_{yy}]^{-1} [\boldsymbol{y} - E(\boldsymbol{y})]$$

$$\boldsymbol{P}^a \equiv E[(\boldsymbol{x}-\boldsymbol{x}^a) (\boldsymbol{x}-\boldsymbol{x}^a)^{\mathrm{T}}] = E(\boldsymbol{x}' \boldsymbol{x}'^{\mathrm{T}}) - E(\boldsymbol{x}'^a \boldsymbol{x}'^{a\mathrm{T}})$$
$$= \boldsymbol{C}_{xx} - \boldsymbol{C}_{xy} [\boldsymbol{C}_{yy}]^{-1} \boldsymbol{C}_{yx}$$

Optimal Interpolation (continued 14)

$$\boldsymbol{x}^a = E(\boldsymbol{x}) + \boldsymbol{C}_{\boldsymbol{xy}} [\boldsymbol{C}_{\boldsymbol{yy}}]^{-1} [\boldsymbol{y} - E(\boldsymbol{y})]$$

$$\boldsymbol{P}^a = \boldsymbol{C}_{\boldsymbol{xx}} - \boldsymbol{C}_{\boldsymbol{xy}} [\boldsymbol{C}_{\boldsymbol{yy}}]^{-1} \boldsymbol{C}_{\boldsymbol{yx}}$$

If probability distribution for couple $(\boldsymbol{x}, \boldsymbol{y})$ is Gaussian (with, in particular, covariance matrix

$$C = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix}$$

then Optimal Interpolation achieves Bayesian estimation, in the sense that

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \mathcal{N}[\boldsymbol{x}^a, \boldsymbol{P}^a]$$

Optimal Interpolation (continued 15)

Optimal Interpolation is a particular (and relatively simple) case of a more general approach called *kriging*, originally developed for the estimation of the content of an ore field.

Best Linear Unbiased Estimate

State vector $\boldsymbol{x}$, belonging to *state space* $\mathcal{S}$ ($\dim \mathcal{S} = n$), to be estimated.
Available data in the form of

- A '*background*' estimate (*e*. *g*. forecast from the past), belonging to *state space*, with dimension n

  $$\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b$$

- An additional set of data (*e*. *g*. observations), belonging to *observation space*, with dimension p

  $$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

  $\boldsymbol{H}$ is known linear *observation operator*.

Assume probability distribution is known for the couple $(\boldsymbol{\zeta}^b, \boldsymbol{\varepsilon})$.
Assume $E(\boldsymbol{\zeta}^b) = 0$, $E(\boldsymbol{\varepsilon}) = 0$, $E(\boldsymbol{\zeta}^b \boldsymbol{\varepsilon}^{\mathrm{T}}) = 0$ (not restrictive)
Set $E(\boldsymbol{\zeta}^b \boldsymbol{\zeta}^{b\mathrm{T}}) \equiv \boldsymbol{P}^b$ (also often denoted $\boldsymbol{B}$), $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) \equiv \boldsymbol{R}$

Best Linear Unbiased Estimate (continuation 1)

$$\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b \qquad (1)$$

$$\boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{\varepsilon} \qquad (2)$$

A probability distribution being known for the couple $(\boldsymbol{\zeta}^b, \boldsymbol{\varepsilon})$, eqs (1-2) define probability distribution for the couple $(\boldsymbol{x}, \boldsymbol{y})$, with

$$E(\boldsymbol{x}) = \boldsymbol{x}^b, \quad \boldsymbol{x}' = \boldsymbol{x} - E(\boldsymbol{x}) = -\boldsymbol{\zeta}^b$$

$$E(\boldsymbol{y}) = \boldsymbol{Hx}^b, \quad \boldsymbol{y}' = \boldsymbol{y} - E(\boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{Hx}^b = \boldsymbol{\varepsilon} - \boldsymbol{H\zeta}^b \qquad (\boldsymbol{H} \text{ is linear})$$

$\boldsymbol{d} \equiv \boldsymbol{y} - \boldsymbol{Hx}^b$ is called the *innovation vector*.

Best Linear Unbiased Estimate (continuation 2)

Apply formulæ for Optimal Interpolation for estimating $\boldsymbol{x}$

$$\boldsymbol{x}^a = E(\boldsymbol{x}) + \boldsymbol{C}_{xy}\,[\boldsymbol{C}_{yy}]^{-1}\,[\boldsymbol{y} - E(\boldsymbol{y})]$$
$$\boldsymbol{P}^a = \boldsymbol{C}_{xx} - \boldsymbol{C}_{xy}\,[\boldsymbol{C}_{yy}]^{-1}\,\boldsymbol{C}_{yx}$$

$$E(\boldsymbol{x}) = \boldsymbol{x}^b\,,\quad \boldsymbol{x}' = \boldsymbol{x} - E(\boldsymbol{x}) = -\,\boldsymbol{\zeta}^b$$
$$E(\boldsymbol{y}) = \boldsymbol{H}\boldsymbol{x}^b\,,\quad \boldsymbol{y}' = \boldsymbol{y} - E(\boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b = \boldsymbol{\varepsilon} - \boldsymbol{H}\boldsymbol{\zeta}^b$$

$$\boldsymbol{C}_{xy} = E(\boldsymbol{x}'\boldsymbol{y}'^{\mathrm{T}}) = E[-\boldsymbol{\zeta}^b(\boldsymbol{\varepsilon} - \boldsymbol{H}\boldsymbol{\zeta}^b)^{\mathrm{T}}] = -\,\underbrace{E(\boldsymbol{\zeta}^b\boldsymbol{\varepsilon}^{\mathrm{T}})}_{0} + \underbrace{E(\boldsymbol{\zeta}^b\boldsymbol{\zeta}^{b\mathrm{T}})}_{\boldsymbol{P}^b}\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}$$

$$\boldsymbol{C}_{yy} = E(\boldsymbol{y}'\boldsymbol{y}'^{\mathrm{T}}) = E[(\boldsymbol{\varepsilon} - \boldsymbol{H}\boldsymbol{\zeta}^b)\,(\boldsymbol{\varepsilon} - \boldsymbol{H}\boldsymbol{\zeta}^b)^{\mathrm{T}}] = \underbrace{E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})}_{\boldsymbol{R}} + \boldsymbol{H}\underbrace{E(\boldsymbol{\zeta}^b\boldsymbol{\zeta}^{b\mathrm{T}})}_{\boldsymbol{P}^b}\boldsymbol{H}^{\mathrm{T}}$$

$$\boldsymbol{C}_{yy} = \boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}$$

Best Linear Unbiased Estimate (continuation 3)

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b)$$
$$\boldsymbol{P}^a = \boldsymbol{P}^b - \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} \boldsymbol{H}\boldsymbol{P}^b$$

$\boldsymbol{x}^a$ is the *Best Linear Unbiased Estimate* (*BLUE*) of $\boldsymbol{x}$ from $\boldsymbol{x}^b$ and $\boldsymbol{y}$.

Equivalent set of formulæ

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^a \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b)$$
$$[\boldsymbol{P}^a]^{-1} = [\boldsymbol{P}^b]^{-1} + \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1}\boldsymbol{H}$$

Vector $\boldsymbol{d} \equiv \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b$ is *innovation vector*
Matrix $\boldsymbol{K} \equiv \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} = \boldsymbol{P}^a \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1}$ is *gain matrix*.

If couple $(\boldsymbol{\zeta}^b, \boldsymbol{\varepsilon})$ is Gaussian, *BLUE* achieves bayesian estimation, in the sense that
$P(\boldsymbol{x} \mid \boldsymbol{x}^b, \boldsymbol{y}) = \mathcal{N}[\boldsymbol{x}^a, \boldsymbol{P}^a]$.

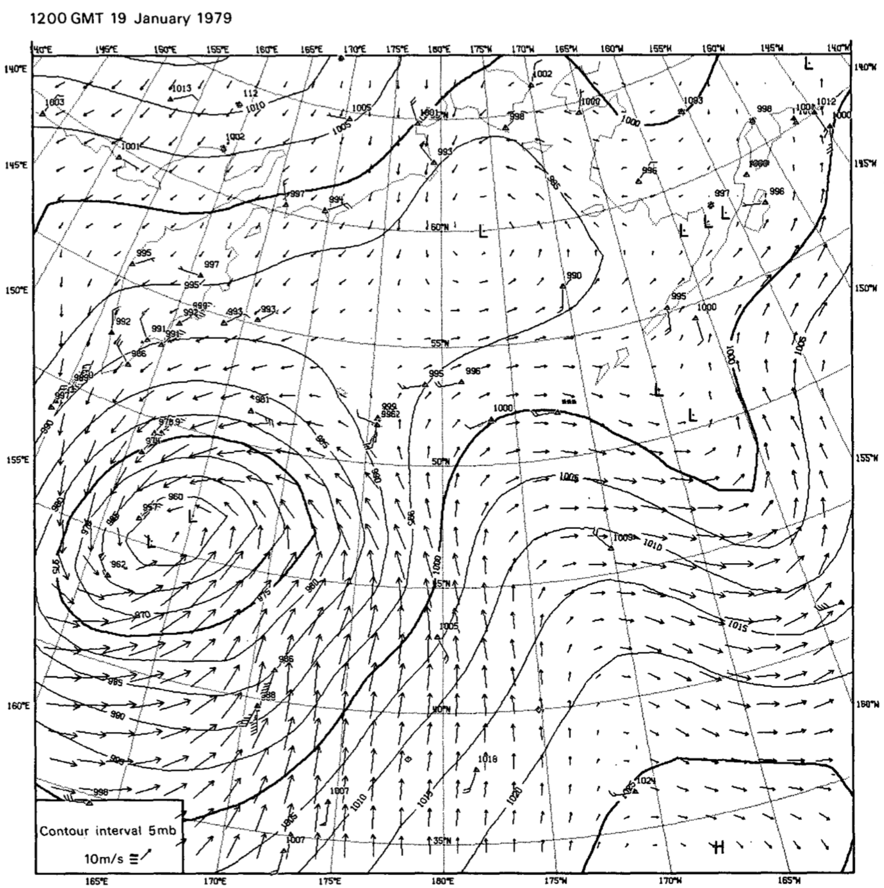


FIG. 14. Sea level pressure and wind forecast corresponding to the central area of Fig. 11, with plotted surface observations of sea level pressure and wind (each barb = 5 m s⁻¹).

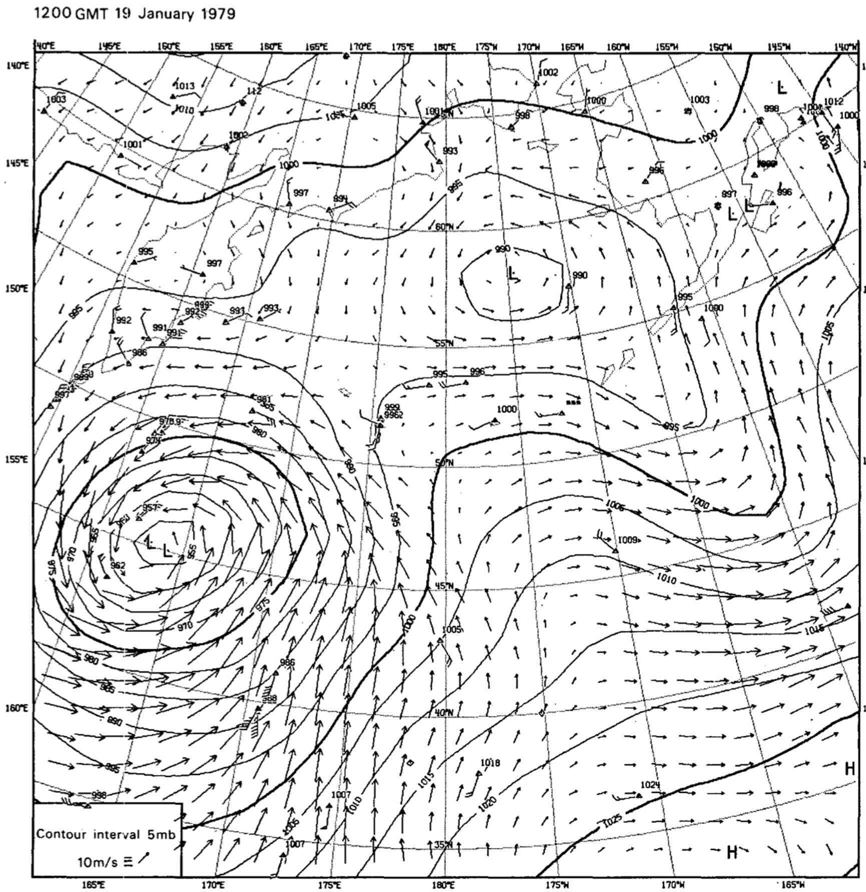


FIG. 15. As in Fig. 14 for the analysis in the data-assimilation cycle.

After A. Lorenc, MWR, 1981

Best Linear Unbiased Estimate (continuation 4)

$\boldsymbol{H}$ can be any linear operator

Example : (scalar) satellite observation

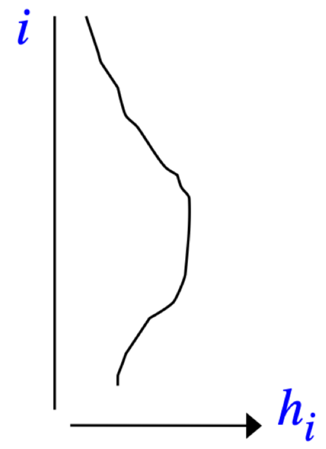
$$\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}} \quad \text{temperature profile}$$

Observation $\quad y = \Sigma_i h_i x_i + \varepsilon = \boldsymbol{Hx} + \varepsilon$, $\boldsymbol{H} = (h_1, \ldots, h_n)$, $E(\varepsilon^2) = r$

Background $\quad \boldsymbol{x}^b = (x_1^{\,b}, \ldots, x_n^{\,b})^{\mathrm{T}}$, error covariance matrix $\boldsymbol{P}^b = (p_{ik}^{\,b})$

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{HP}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{Hx}^b)$$

$$[\boldsymbol{HP}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{Hx}^b) = (y - \Sigma_\iota h_\iota x_\iota^{\,b}) / (\Sigma_{ik} h_i h_k p_{ik}^{\,b} + r) \equiv \mu$$ scalar !



- $\boldsymbol{P}^b = p^b \boldsymbol{I}_n \qquad x_i^a = x_i^b + p^b h_i \mu$
- $\boldsymbol{P}^b = \mathrm{diag}(p_{ii}^{\,b}) \qquad x_i^a = x_i^b + p_{ii}^{\,b} h_i \mu$
- General case $\qquad x_i^a = x_i^b + \Sigma_k p_{ik}^{\,b} h_k \mu$

Each level i is corrected, not only because of its own contribution to the observation of the contribution of the other levels with which its background error is correlated.

Best Linear Unbiased Estimate (continuation 6)

BLUE is invariant in any invertible linear change of variables, in either state or observation space.

Equivalently, *BLUE* is independent of the possible choice of a scalar product in either one of the two spaces.

Again, if the couple (ζ^b, ε) is Gaussian, the *BLUE* is Bayesian in the sense that $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{x}^b, \boldsymbol{y}) = \mathcal{N}[\boldsymbol{x}^a, \boldsymbol{P}^a]$

Next step

How to introduce temporal dynamics in assimilation ?

Kalman Filter. Variational Assimilation

Conference
on **"IA and mathematics for meteorology and climatology »**

May 5th 2025 at the Collège de France, organised by Pierre-Louis Lions and Stéphane Mallat

….

- Marc Bocquet, École Nationale des Ponts et Chaussées, "Artificial Intelligence for gephysical data assimilation"

Program and information available at: https://www.college-de-france.fr/fr/agenda/grand-evenement/ia-et-les-mathematiques-pour-la-meteorologie-et-la-climatologie

Cours à venir

~~Mercredi 2 avril~~

~~Vendredi 11 avril~~

~~Vendredi 18 avril~~

Mercredi 23 avril

Lundi 12 mai

Mercredi 28 mai

Mercredi 11 juin

Mercredi 18 juin