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

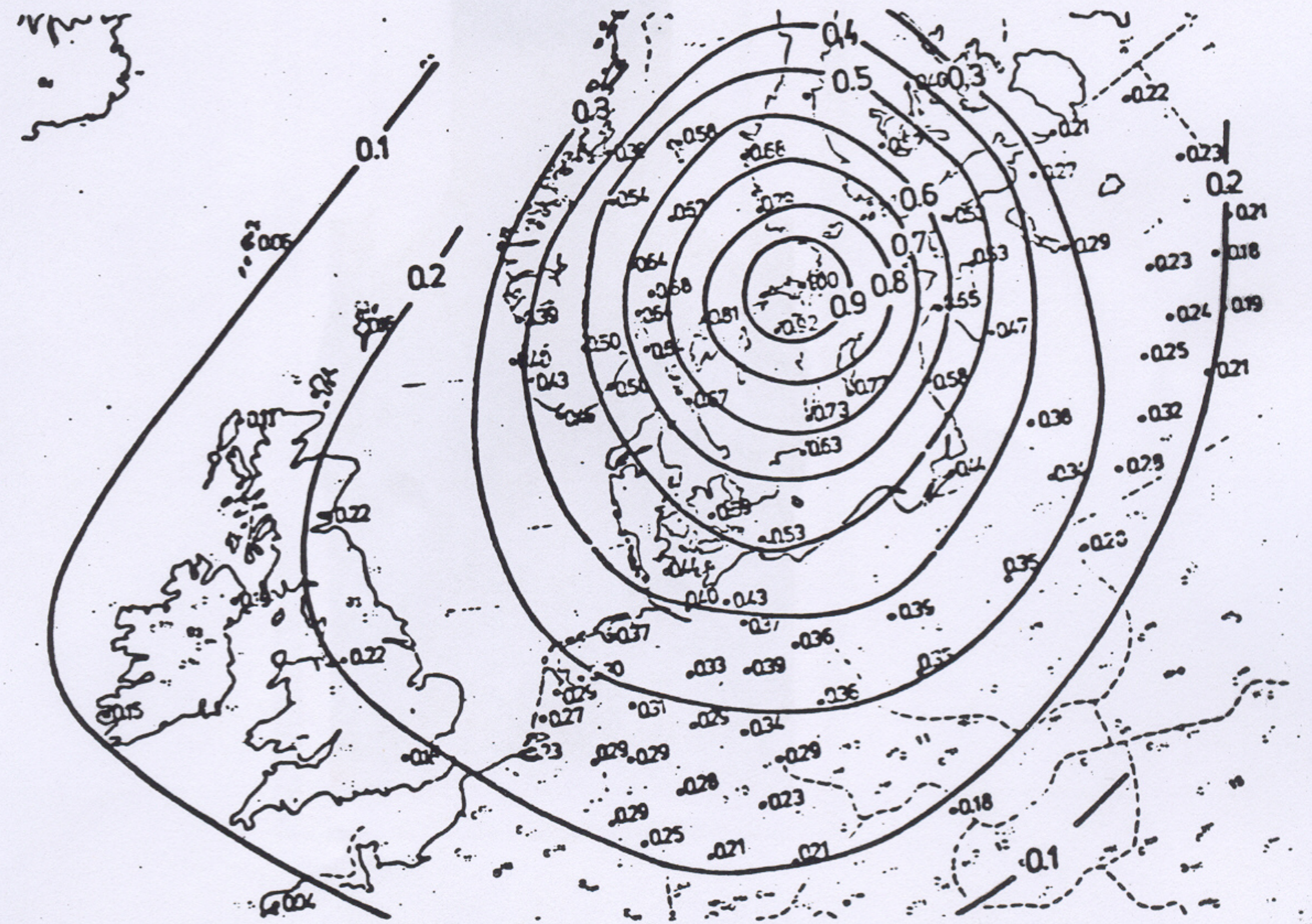
Année Universitaire 2015-2016

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 4

4 Mai 2016



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

After N. Gustafsson

Optimal Interpolation

Random field $\Phi(\xi)$

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

For one particular realization of the field, observations

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

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

$$x^a = \alpha + \Sigma_j \beta_j y_j = \alpha + \boldsymbol{\beta}^T \boldsymbol{y}$$

, where $\boldsymbol{\beta} = (\beta_j)$

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

Optimal Interpolation (continued 1)

Solution

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

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

i.e.,

$$\boldsymbol{\beta}^{\mathrm{T}} = C_{x\mathbf{y}}\,[C_{\mathbf{yy}}]^{-1}$$

$$\alpha = E(x) - \boldsymbol{\beta}^{\mathrm{T}} E(\mathbf{y})$$

Estimate is unbiased $E(x-x^a) = 0$

Minimized quadratic estimation error

$$E[(x-x^a)^2] = E(x'^2) - E[(x'^a)^2])$$

$$= \mathbf{C}_{xx} - \mathbf{C}_{x\mathbf{y}}\,[\mathbf{C}_{\mathbf{yy}}]^{-1}\,\mathbf{C}_{\mathbf{y}x}$$

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

Optimal Interpolation (continued 2)

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

$$y_j = \Phi(\xi_j) + \varepsilon_j$$

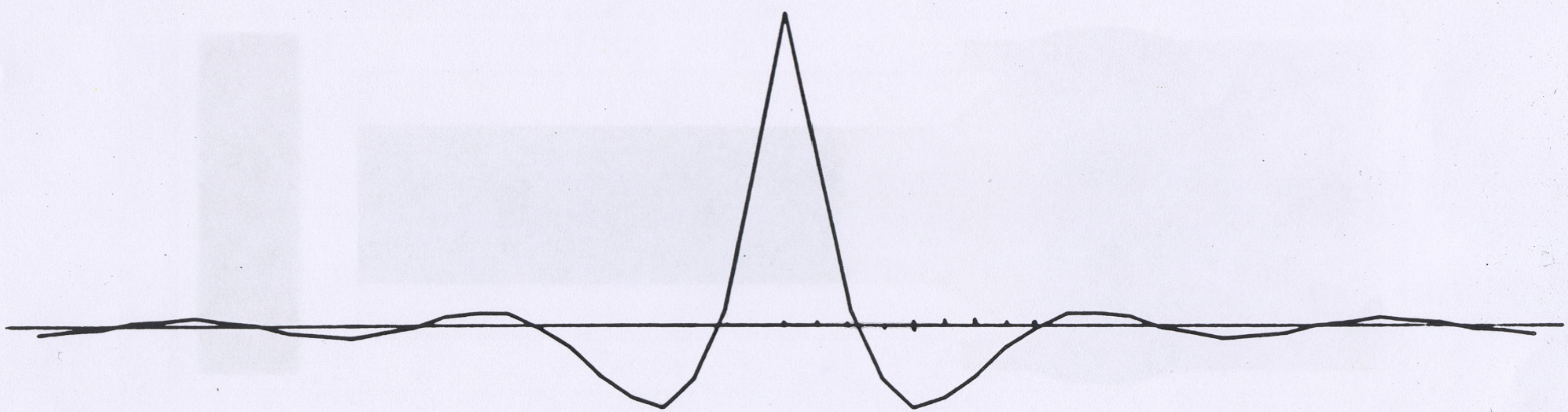
$$E(y_j' y_k') = E[\Phi'(\xi_j) + \varepsilon_j'][\Phi'(\xi_k) + \varepsilon_k']$$

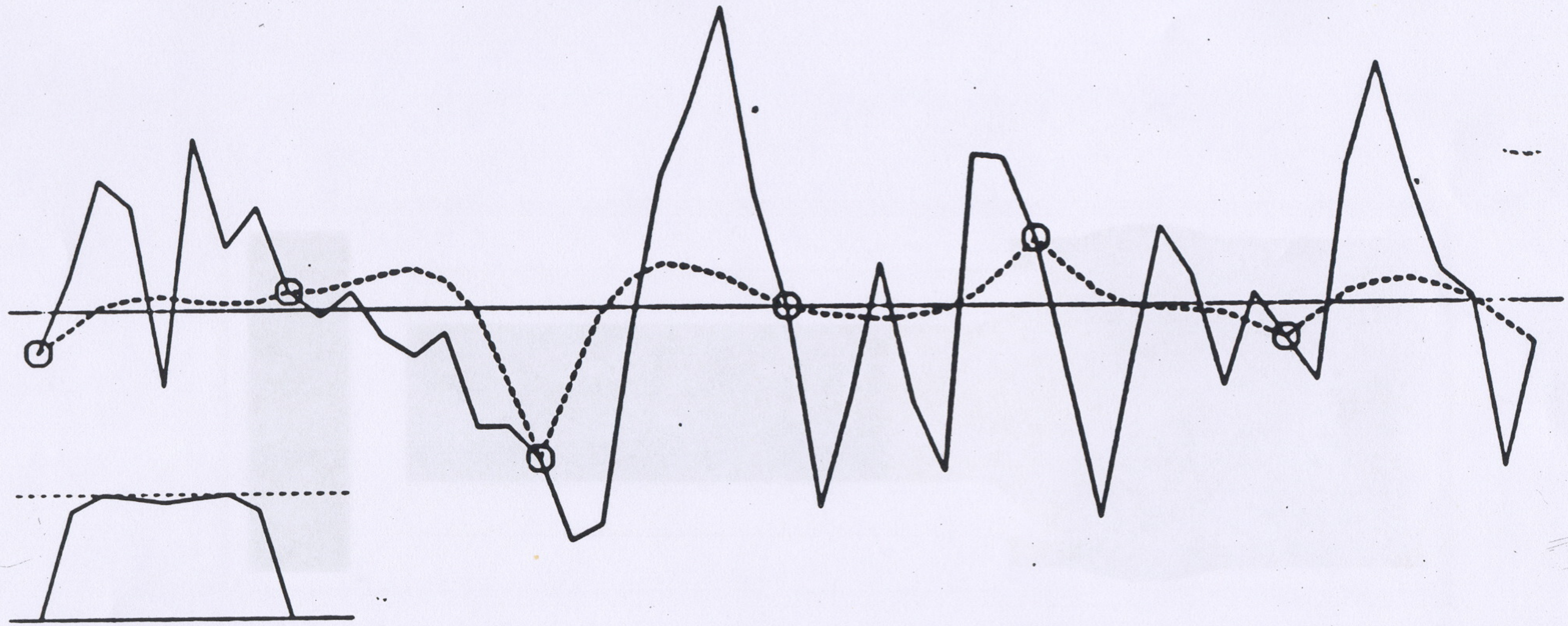
If observation errors ε_j are mutually uncorrelated, have common variance r, and are uncorrelated with field Φ, then

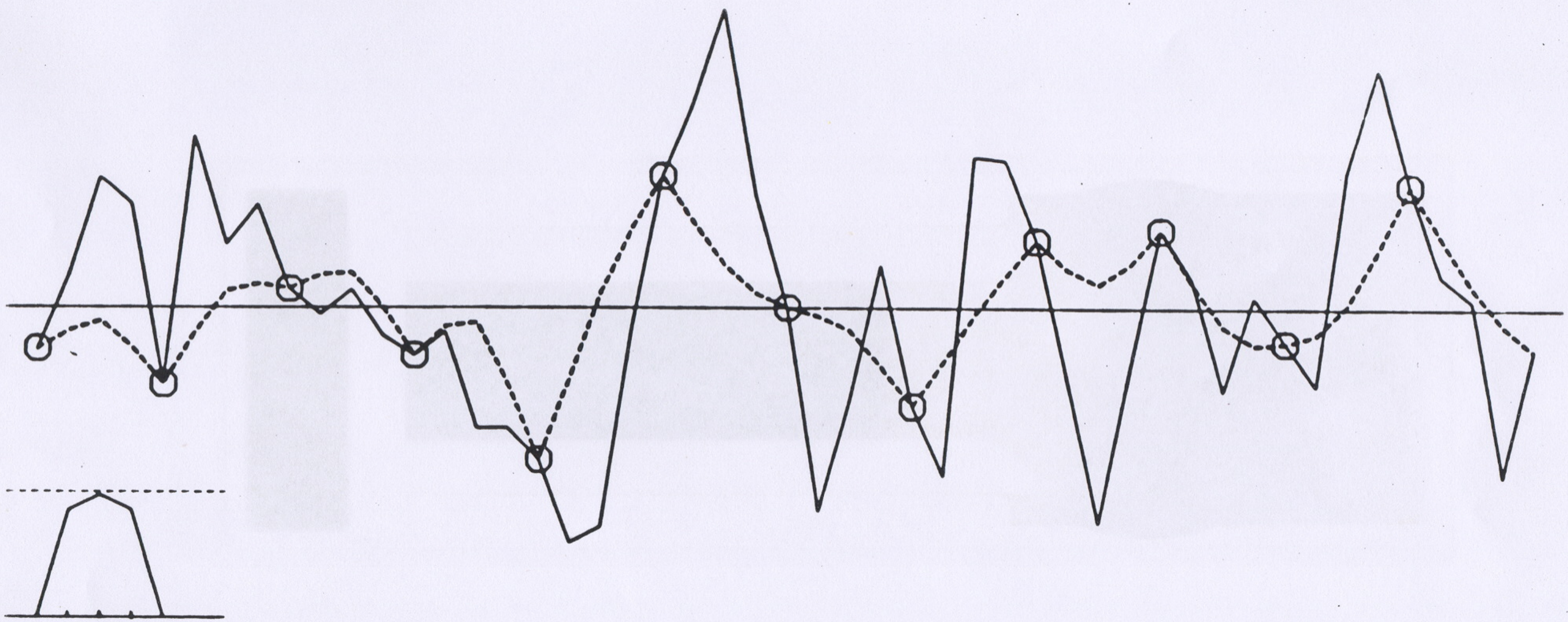
$$E(y_j' y_k') = C_{\Phi}(\xi_j, \xi_k) + r\delta_{jk}$$

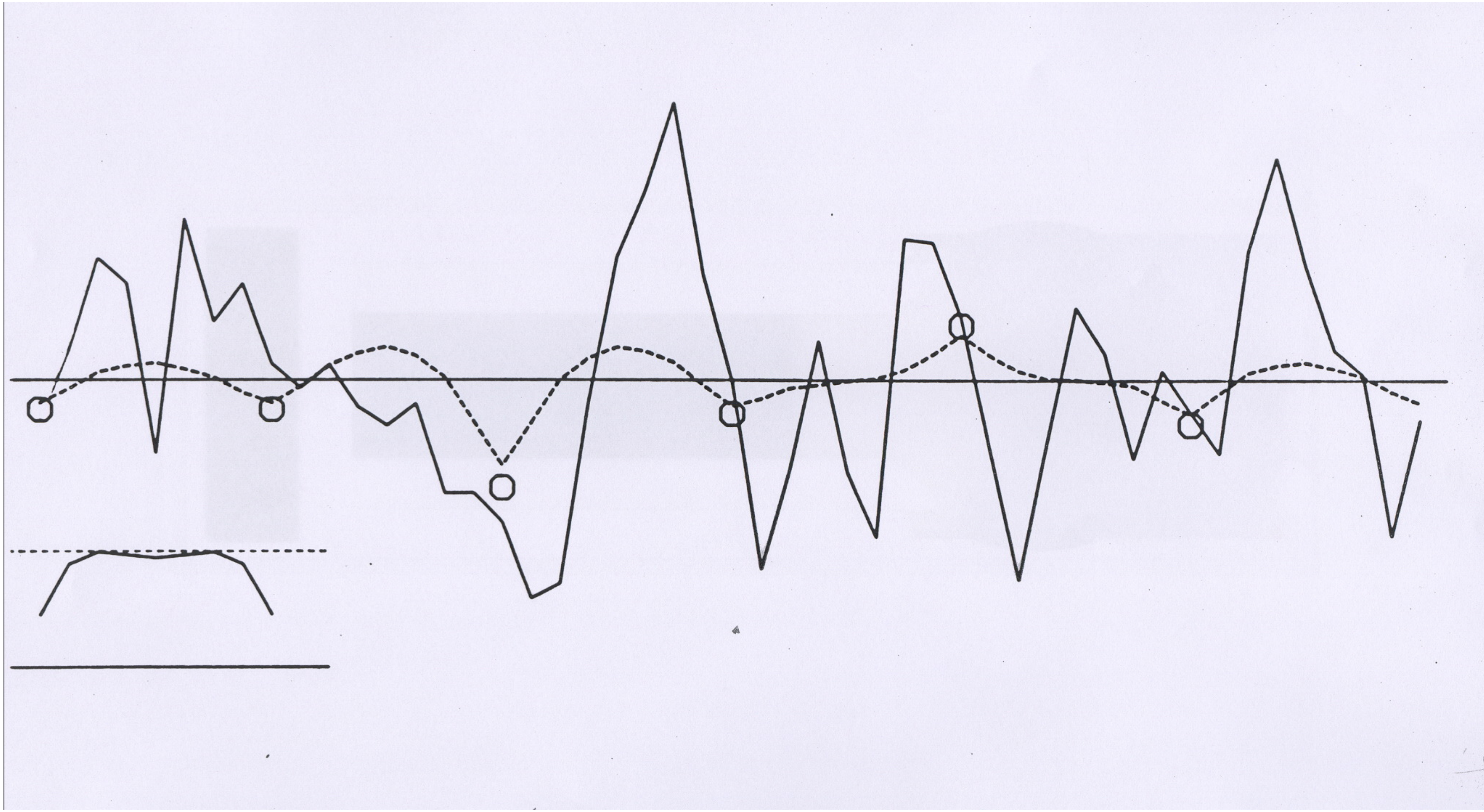
and

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









Optimal Interpolation (continued 3)

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

Vector

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

is independent of variable to be estimated

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

$$\begin{aligned} \Phi^a(\xi) &= E[\Phi(\xi)] + \Sigma_j \mu_j E[\Phi'(\xi) y_j'] \\ &= E[\Phi(\xi)] + \Sigma_j \mu_j C_\Phi(\xi, \xi_j) \end{aligned}$$

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.

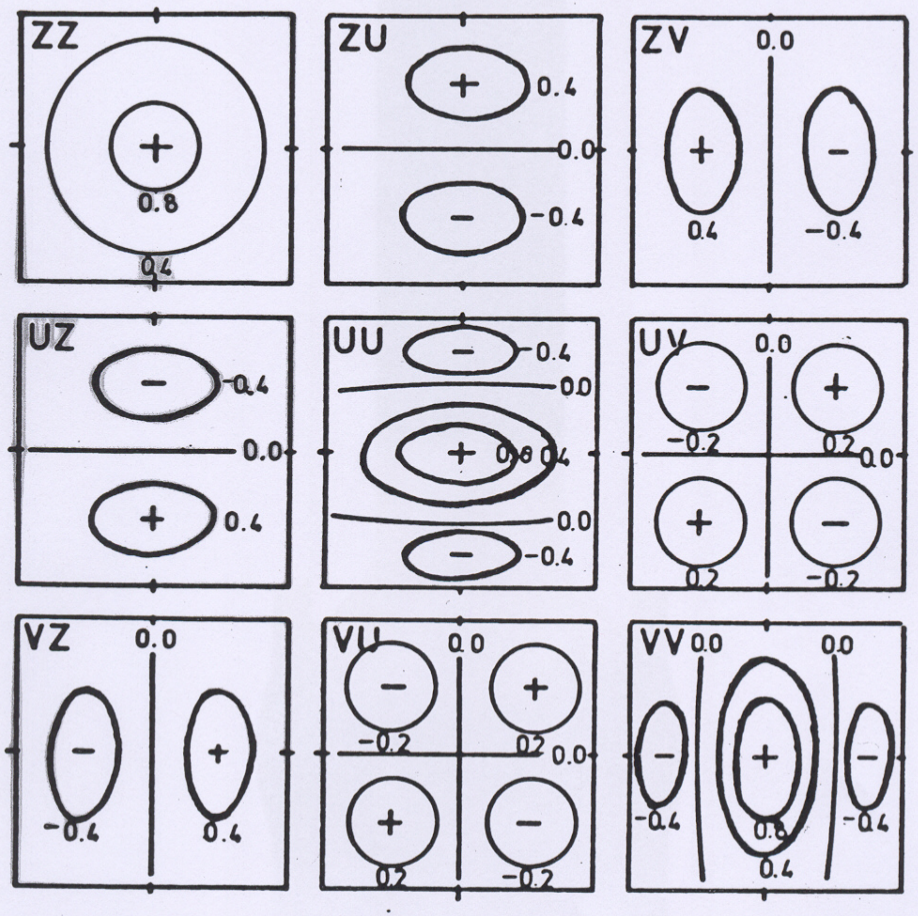
Optimal Interpolation (continued 4)

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

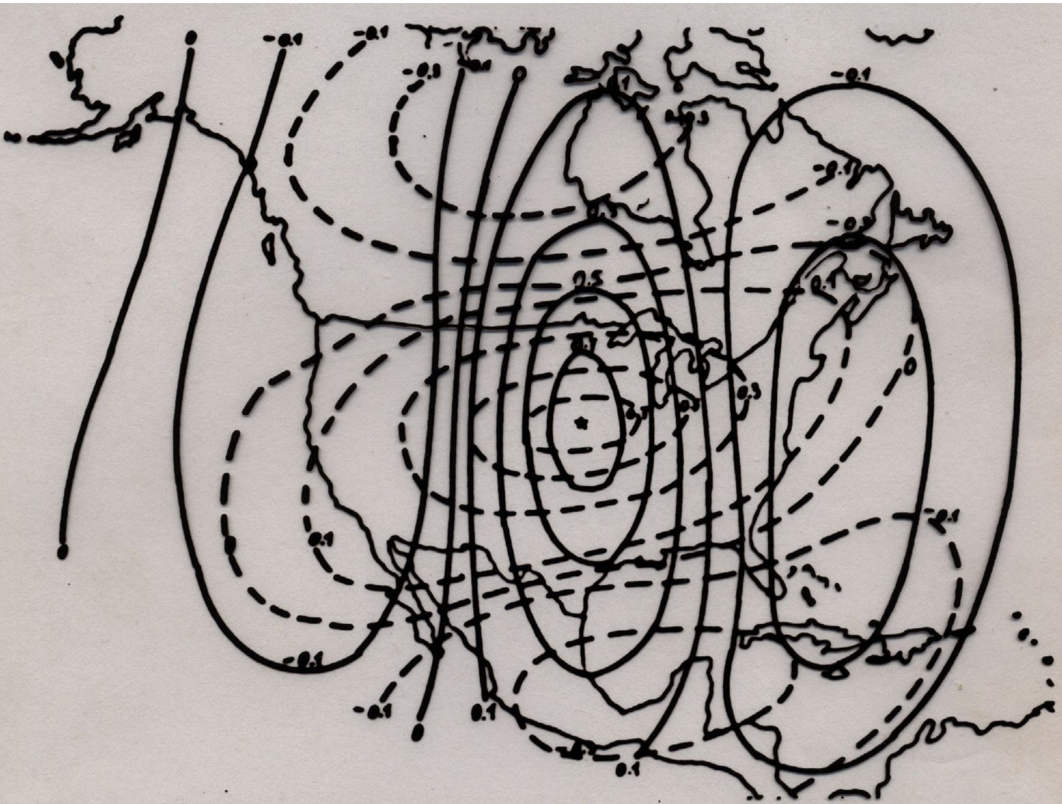


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson

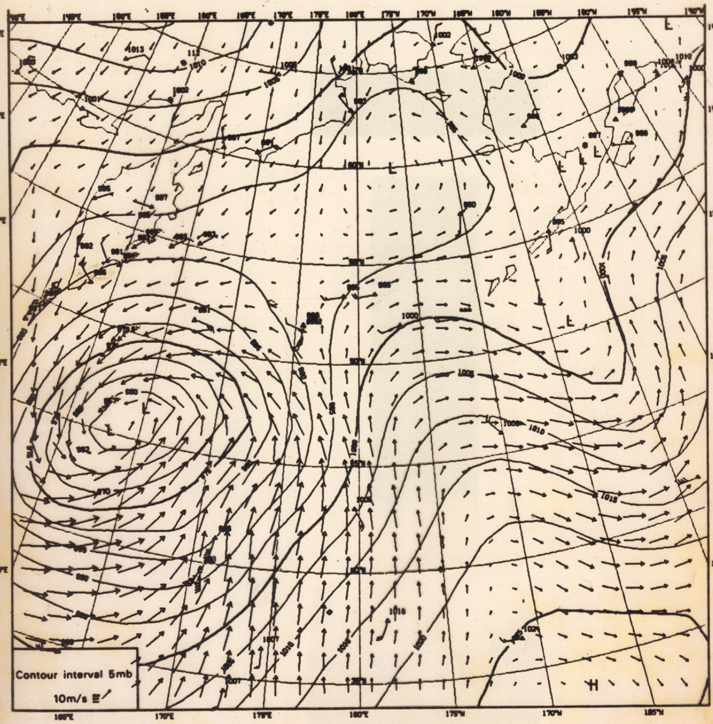


Fig. 4 Sea level pressure and wind forecast corresponding to the central area of Fig. 1, with plotted surface observations of pressure and wind (each fleche = 5 m/s).

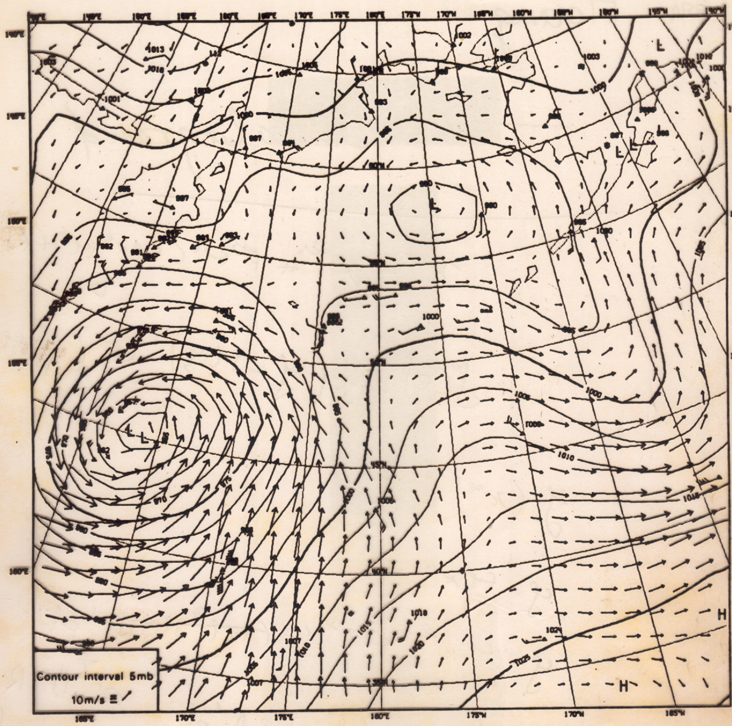


Fig. 5 As Fig. 4 for the analysis in the data assimilation cycle

After A. Lorenc, MWR, 1981

1200 GMT 19 January 1979

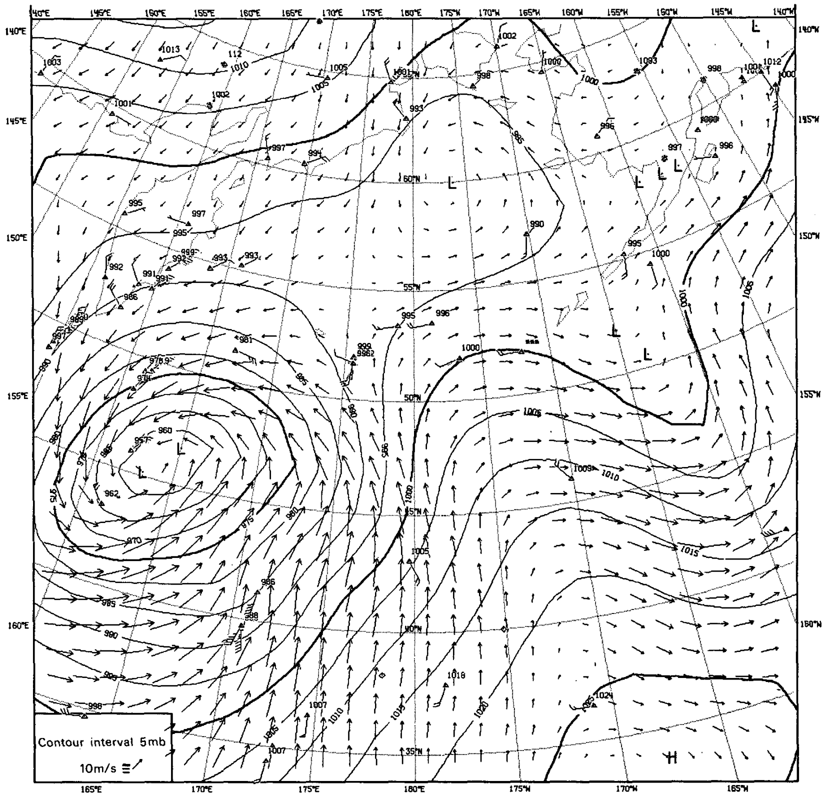


FIG. 14. Sea level pressure and wind forecast corresponding to the central area of Fig. 11, with plotted surface observations of sea level pressure and wind (each barb = 5 m s⁻¹).

1200 GMT 19 January 1979

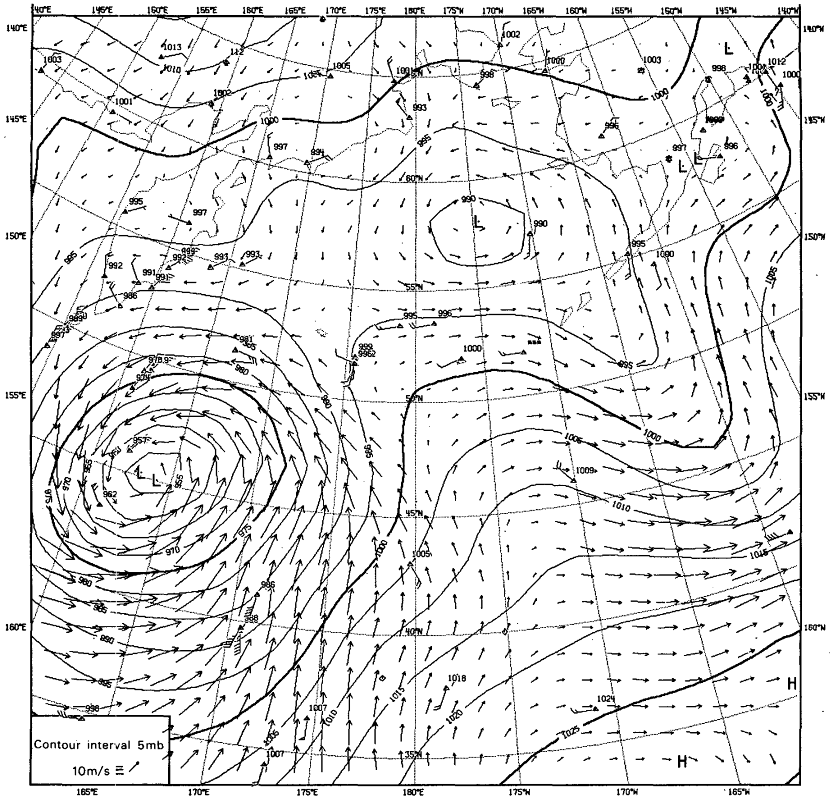


FIG. 15. As in Fig. 14 for the analysis in the data-assimilation cycle.

After A. Lorenc, MWR, 1981

Optimal Interpolation (continued 5)

Observation vector $\boldsymbol{y}$

Estimation of a scalar x

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

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

Estimation of a vector $\boldsymbol{x}$

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

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

Optimal Interpolation (continued 6)

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

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

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

then Optimal Interpolation achieves Bayesian estimation, in the sense that

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

Best Linear Unbiased Estimate

State vector 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

  $x^b = x + \zeta^b$

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

  $y = Hx + \varepsilon$

  H is known linear *observation operator.*

Assume probability distribution is known for the couple (ζ^b, ε).

Assume $E(\zeta^b) = 0$, $E(\varepsilon) = 0$, $E(\zeta^b \varepsilon^T) = 0$ (not restrictive)

Set $E(\zeta^b \zeta^{bT}) = P^b$ (also often denoted B), $E(\varepsilon\varepsilon^T) = R$

Best Linear Unbiased Estimate (continuation 1)

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

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

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

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

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

Best Linear Unbiased Estimate (continuation 2)

Apply formulæ for Optimal Interpolation

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

x^a is the *Best Linear Unbiased Estimate* (*BLUE*) of x from x^b and y.

Equivalent set of formulæ

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

Matrix $K = P^b H^T [HP^bH^T + R]^{-1} = P^a H^T R^{-1}$ is *gain matrix*.

If probability distributions are *globally* gaussian, *BLUE* achieves bayesian estimation, in the sense that $P(\boldsymbol{x} \mid \boldsymbol{x}^b, \boldsymbol{y}) = \mathcal{N}[\boldsymbol{x}^a, \boldsymbol{P}^a]$.