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

Année Universitaire 2013-2014

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 4

12 Mai 2014

$z_1 = x + \zeta_1$ 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

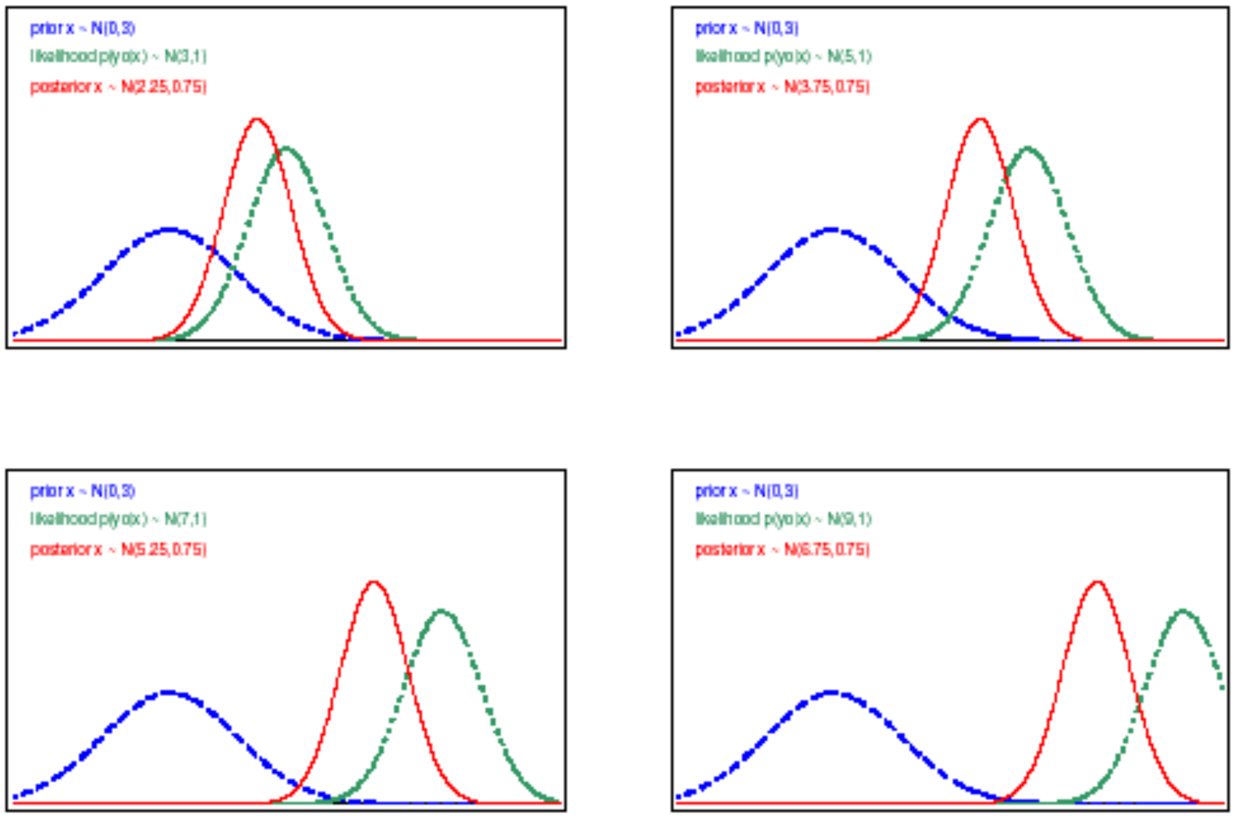


Fig. 1.1: Prior pdf $p(x)$ (dashed line), posterior pdf $p(x|y^o)$ (solid line), and Gaussian likelihood of observation $p(y^o|x)$ (dotted line), plotted against x for various values of y^o. (Adapted from Lorenc and Hammon 1988.)

Conditional expectation x^a minimizes following scalar *objective function*, defined on ξ-space

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

In addition

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

Conditional probability distribution in Gaussian case

$$P(x = \xi \mid z_1, z_2) \propto \exp[\ -\ \underbrace{(\xi - x^a)^2/2p^a}_{\mathcal{J}(\xi)}]$$

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 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*)

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

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)~~

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

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

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

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 $Var[\Phi(\xi)] = E\{[\Phi'(\xi)]^2\}$

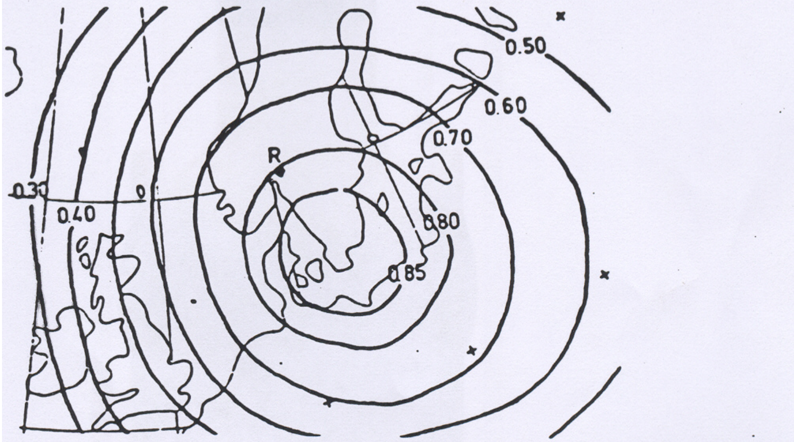
- Covariance function

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

- 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

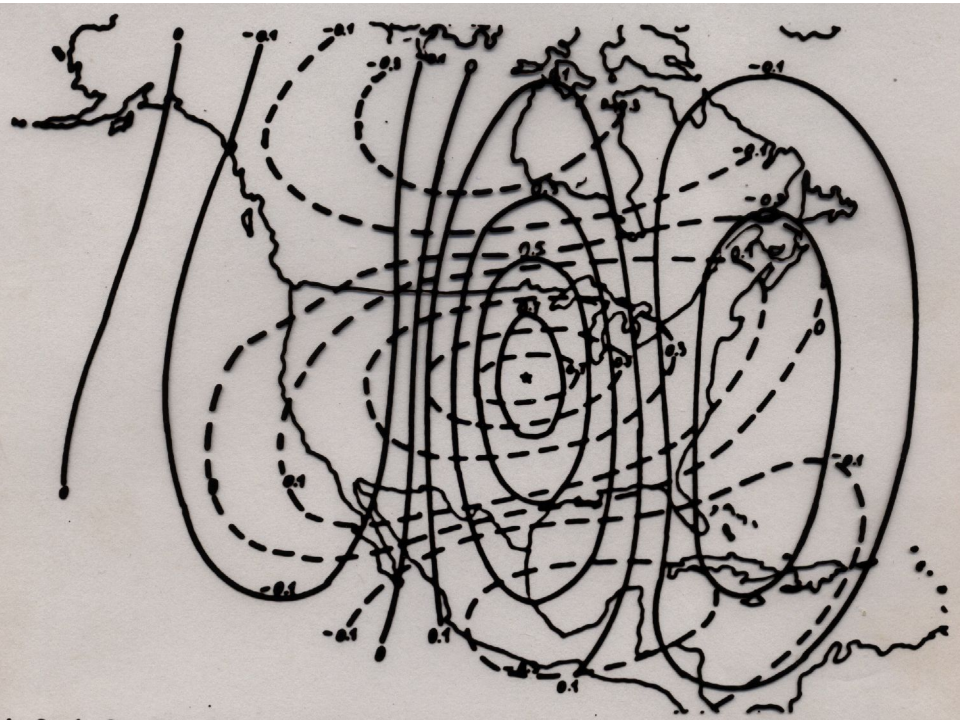


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson

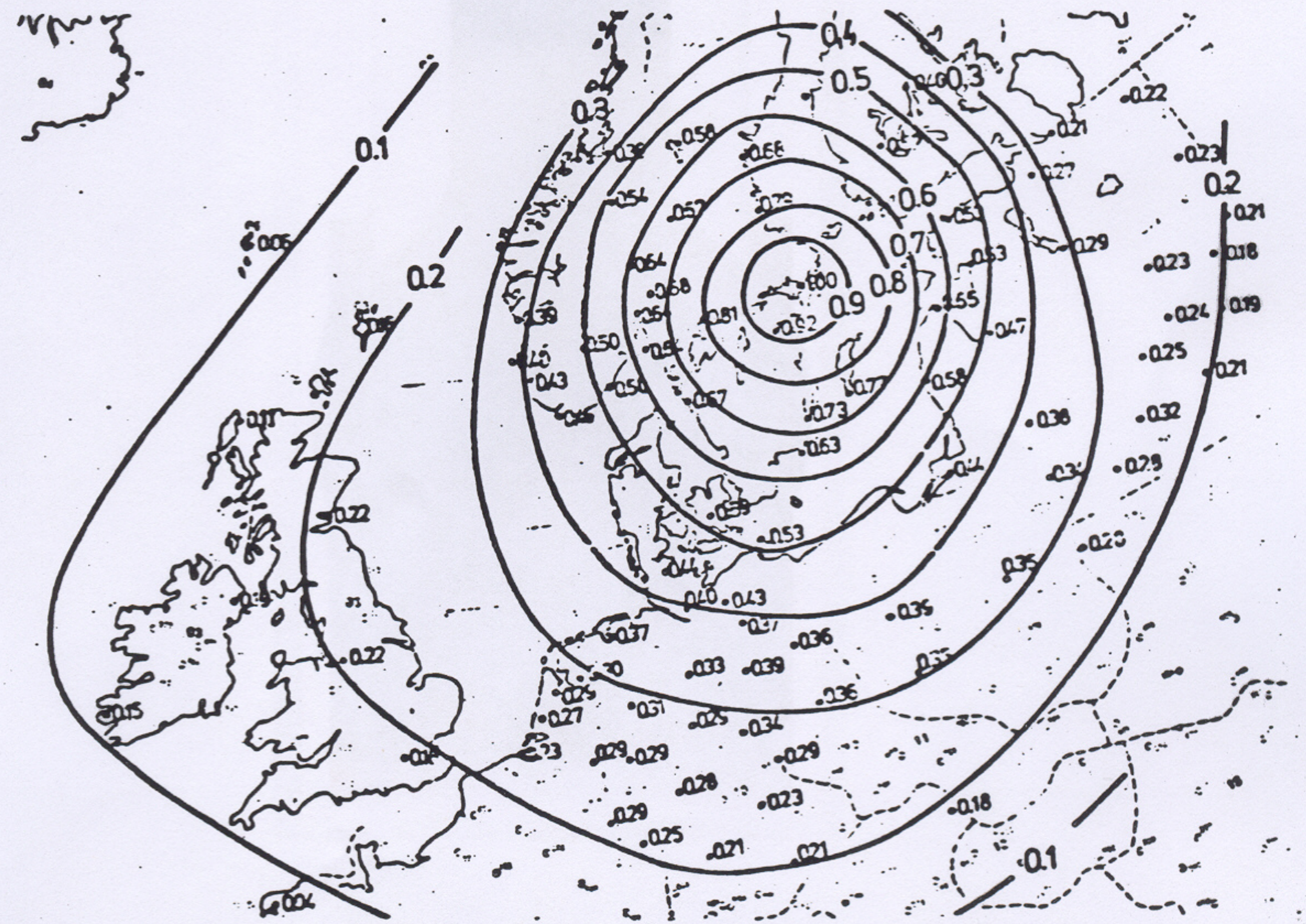


Figure 5.1.1.4.1 Auto-correlation of errors in 12h numerical forecasts of surface pressure in a reference station (Stockholm) and other stations.

After N. Gustafsson

Optimal Interpolation

Random field $\Phi(\xi)$

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

For one particular realization of the field, observations

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

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

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

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

Optimal Interpolation (continued 1)

Solution

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

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

$$E[(x-x^a)^2] = E(x'^2) - E[(x'^a)^2])$$
$$= C_{xx} - C_{x\mathbf{y}}\,[C_{\mathbf{yy}}]^{-1}\,C_{\mathbf{y}x}$$

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

Optimal Interpolation (continued 2)

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

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

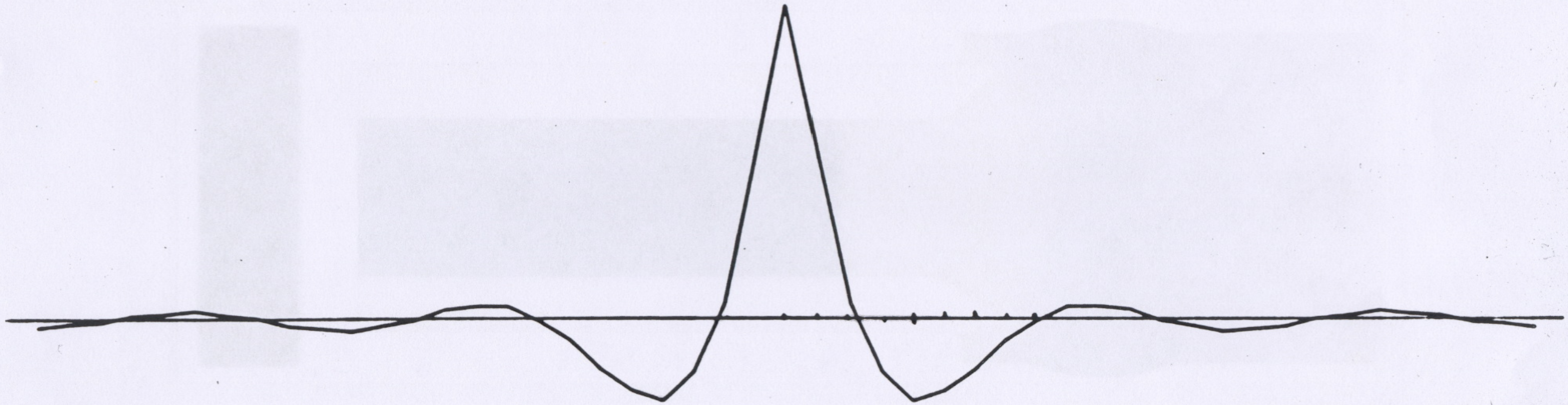
$$E(y_j'y_k') = E[\boldsymbol{\Phi}'(\xi_j) + \varepsilon_j'][\boldsymbol{\Phi}'(\xi_k) + \varepsilon_k']$$

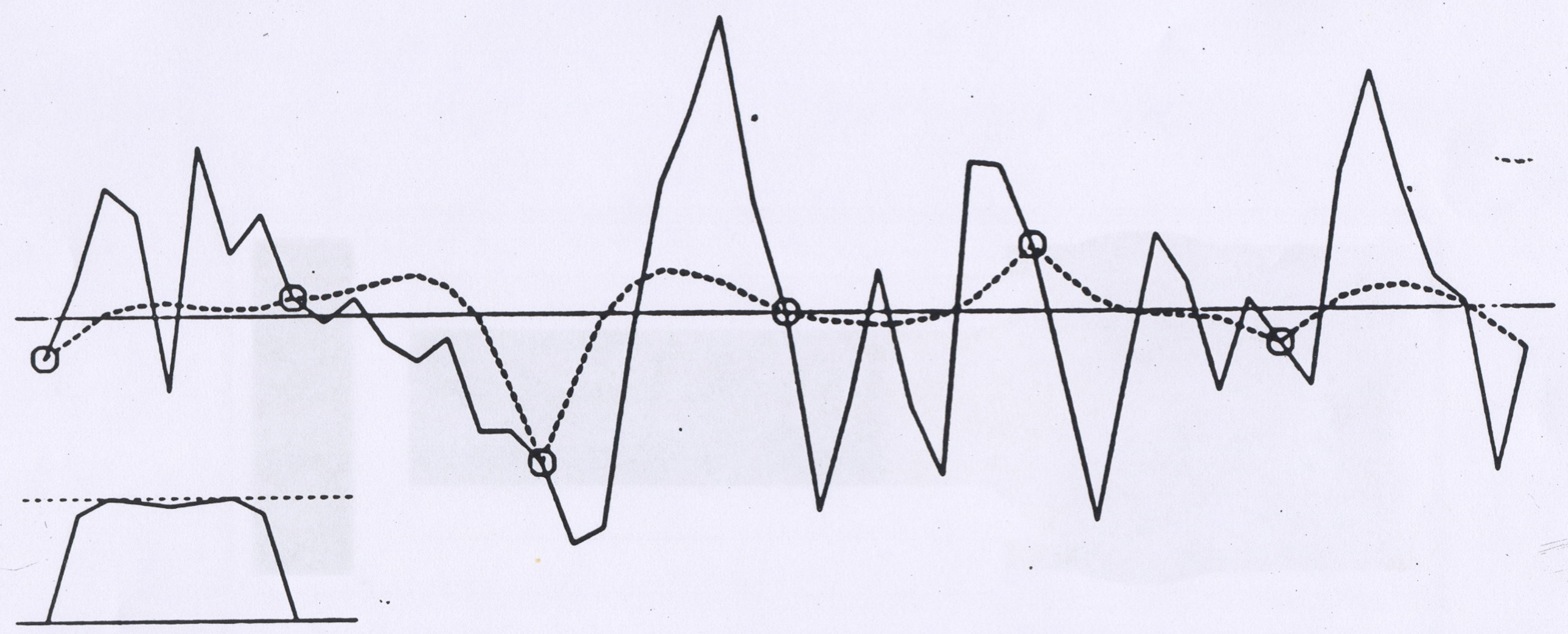
If observation errors ε_j are mutually uncorrelated, have common variance r, and are uncorrelated with field $\boldsymbol{\Phi}$, then

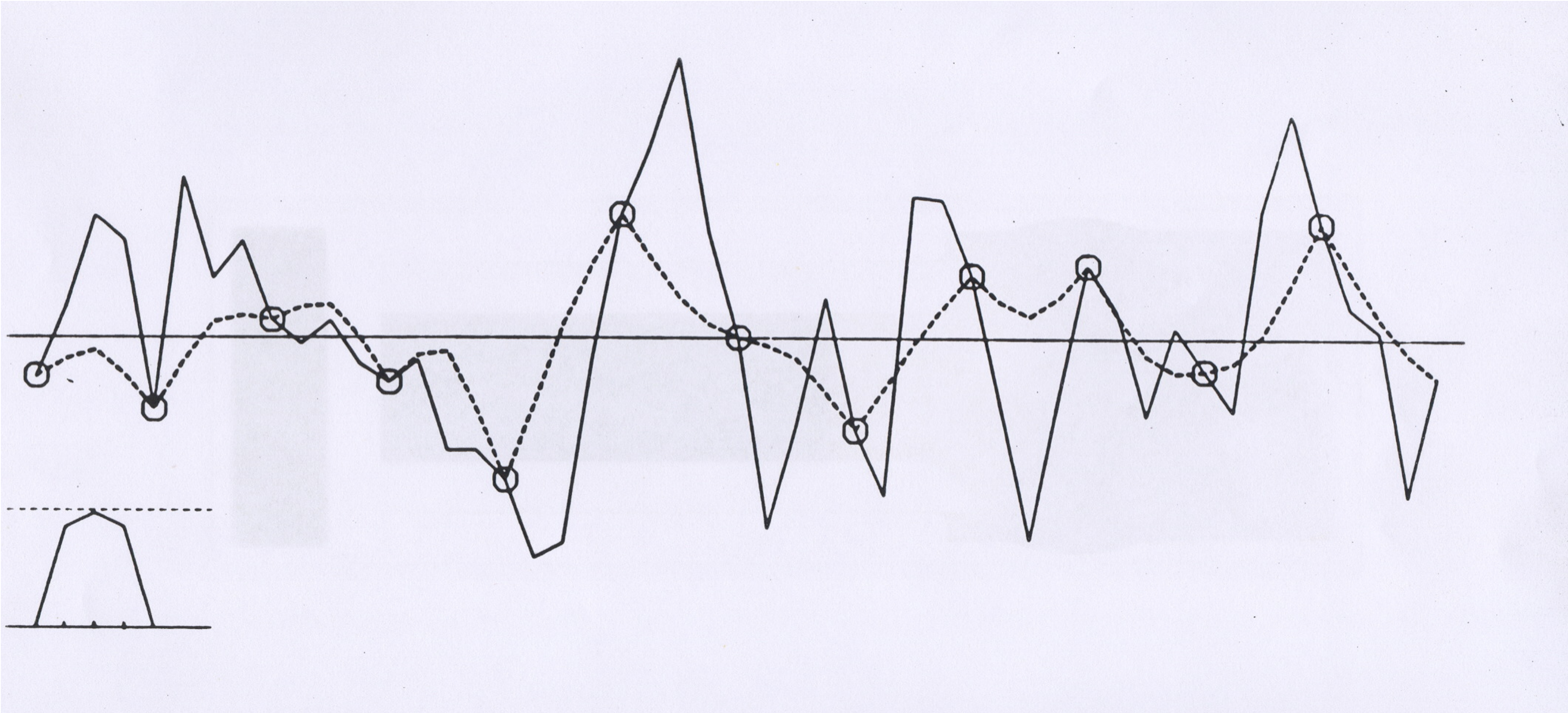
$$E(y_j'y_k') = C_{\Phi}(\xi_j, \xi_k) + r\delta_{jk}$$

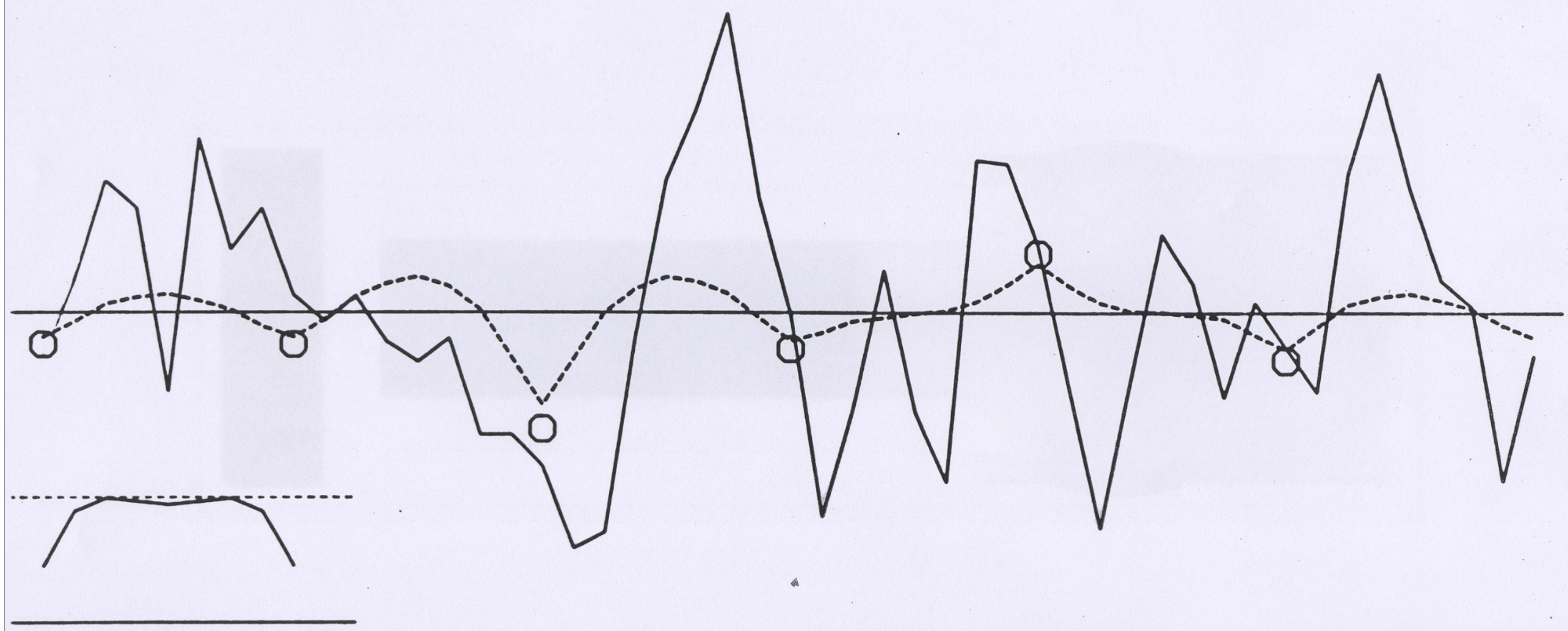
and

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









Optimal Interpolation (continued 3)

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

Vector

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

is independent of variable to be estimated

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

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

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.

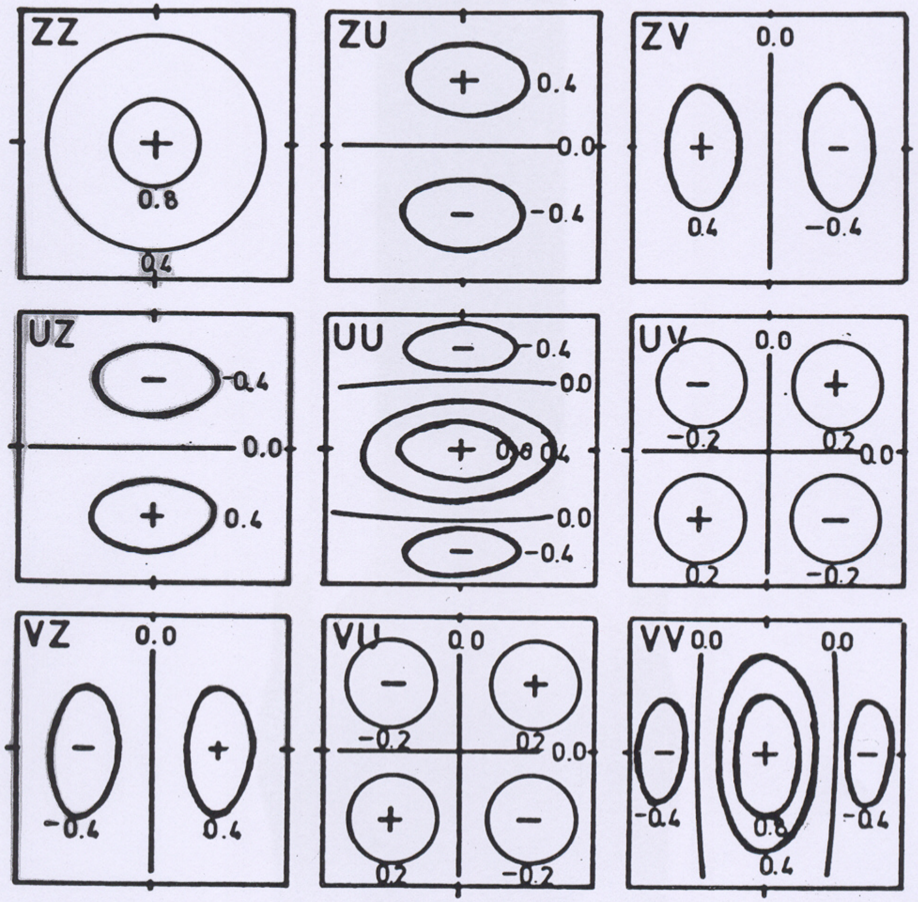
Optimal Interpolation (continued 4)

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

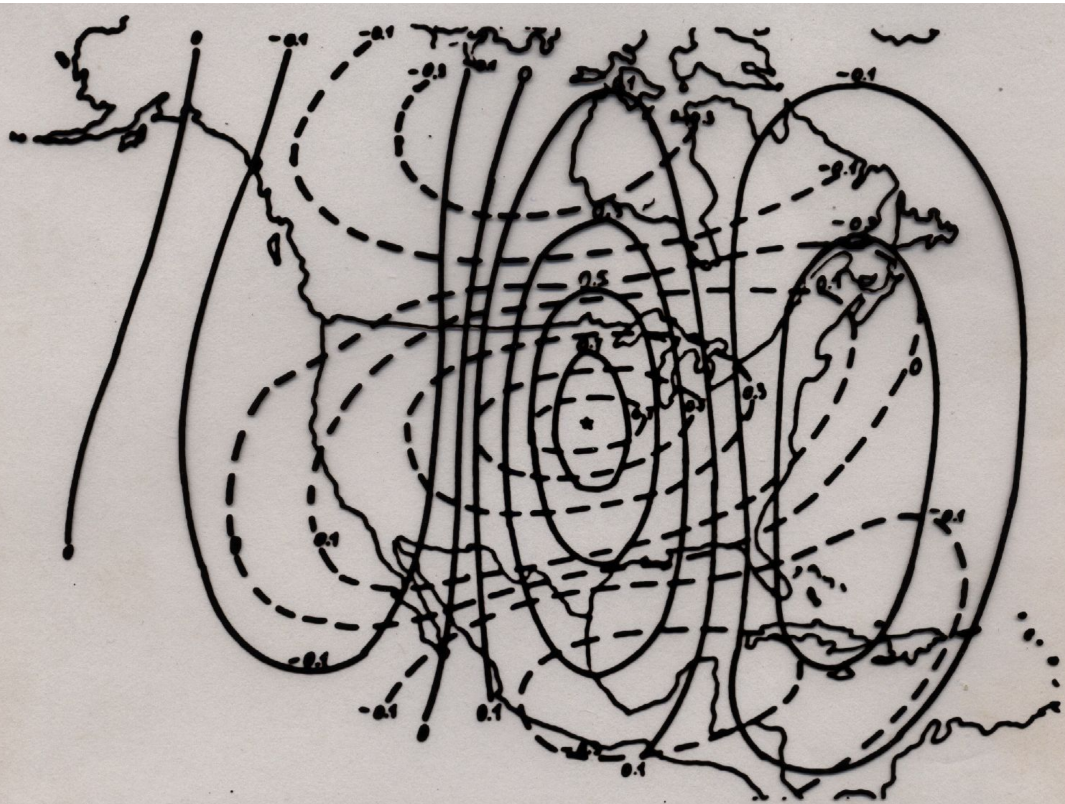


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson

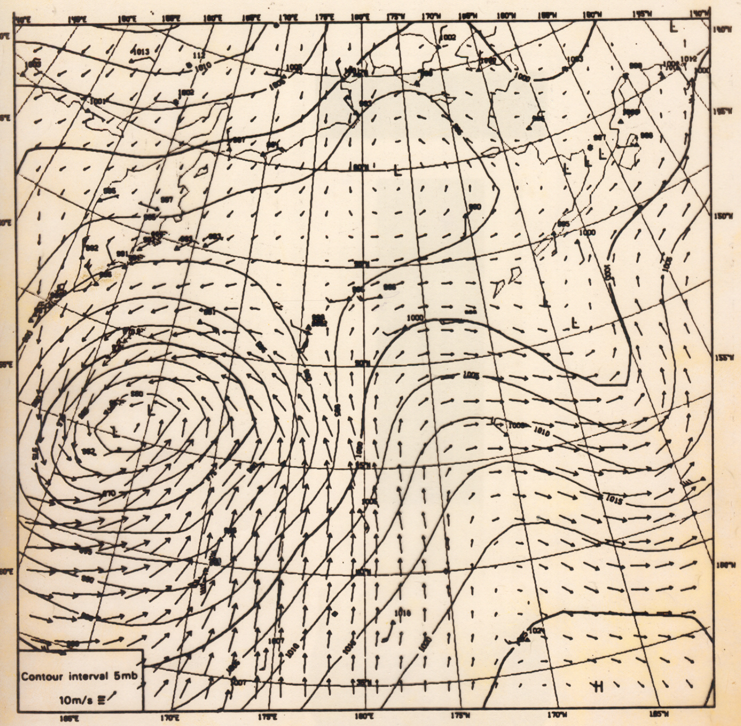


Fig. 4 Sea level pressure and wind forecast corresponding to the central area of Fig. 1, with plotted surface observations of pressure and wind (each fleche = 5 m/s).

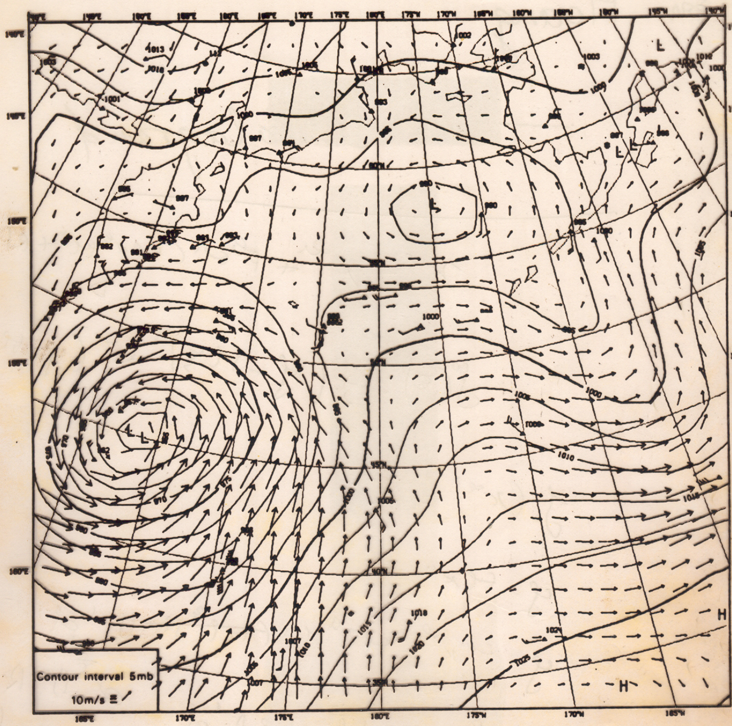


Fig. 5 As Fig. 4 for the analysis in the data assimilation cycle

After A. Lorenc, MWR, 1981

1200 GMT 19 January 1979

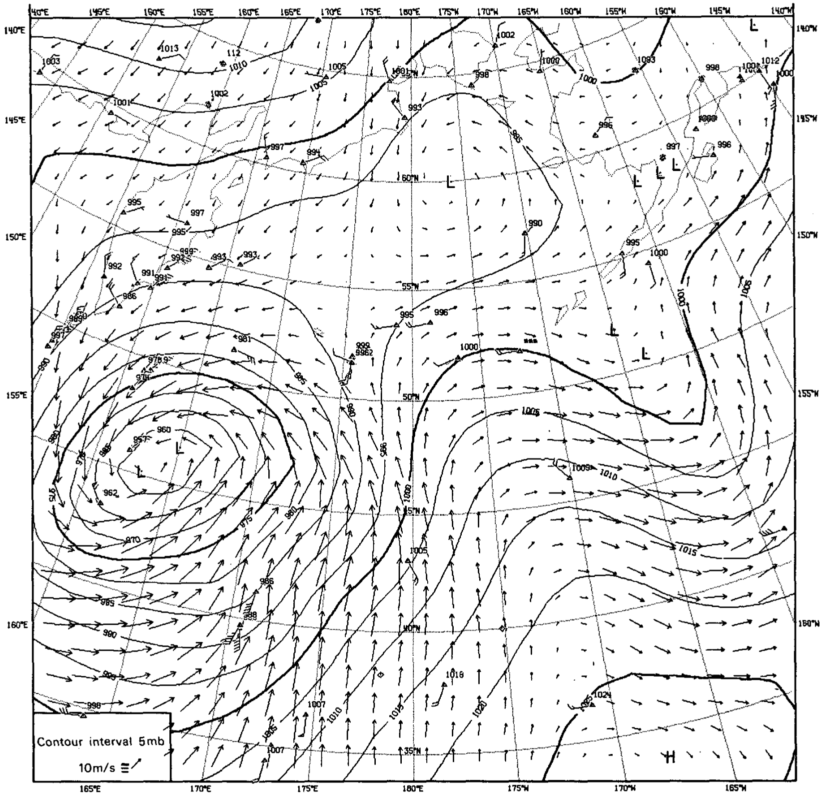


FIG. 14. Sea level pressure and wind forecast corresponding to the central area of Fig. 11, with plotted surface observations of sea level pressure and wind (each barb = 5 m s^{-1}).

1200 GMT 19 January 1979

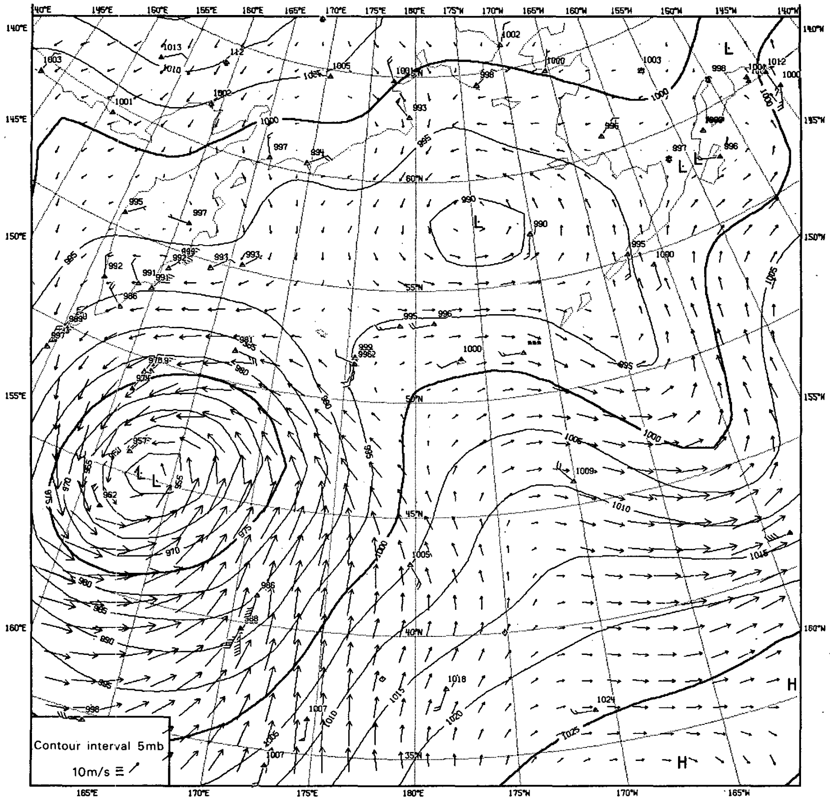


FIG. 15. As in Fig. 14 for the analysis in the data-assimilation cycle.

After A. Lorenc, MWR, 1981

Optimal Interpolation (continued 5)

Observation vector $\boldsymbol{y}$

Estimation of a scalar x

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

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

Estimation of a vector $\boldsymbol{x}$

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

$$E[(\boldsymbol{x}-\boldsymbol{x}^a) (\boldsymbol{x}-\boldsymbol{x}^a)^T] = E(\boldsymbol{x}'\boldsymbol{x}'^T) - E(\boldsymbol{x}'^a \boldsymbol{x}'^{aT})$$
$$= C_{\boldsymbol{xx}} - C_{\boldsymbol{xy}} [C_{\boldsymbol{yy}}]^{-1} C_{\boldsymbol{yx}}$$