Université Pierre et Marie Curie
Master de Sciences et Technologies
Spécialité Océan, Atmosphère, Climat et Observations Spatiales

Année 2010-2011
Cours *Introduction à l'assimilation de données et modélisation inverse en géophysique*

De la modélisation à l'assimilation de données

Olivier Talagrand
8 Novembre 2010

Fig. 1: Members of day 7 forecast of 500 hPa geopotential height for the ensemble originated from 25 January 1993.

Figure 6 Hurricane Katrina mean-sea-level-pressure (MSLP) analysis for 12 UTC of 29 August 2005 and t+84h high-resolution and EPS forecasts started at 00 UTC of 26 August:

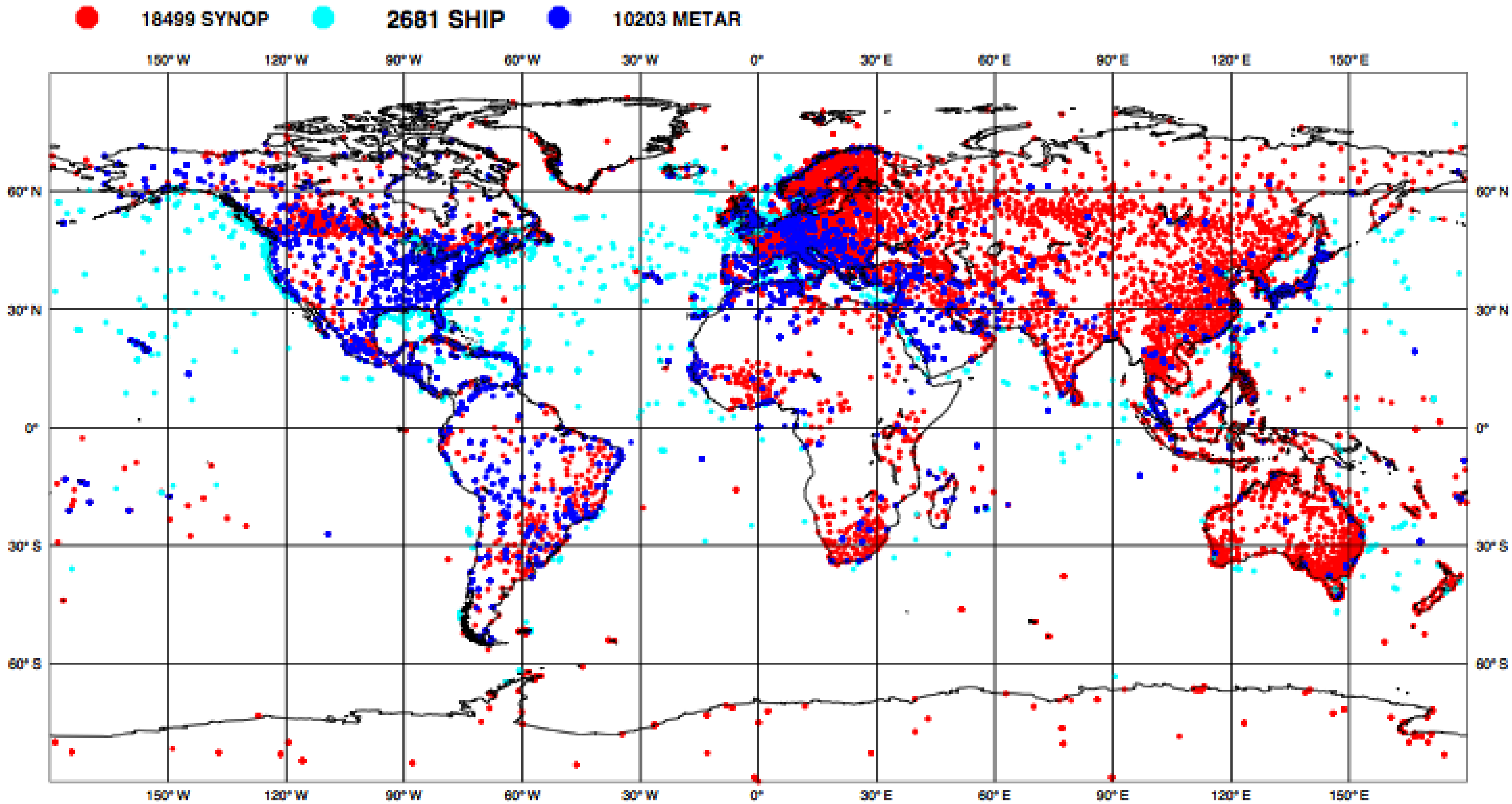
1st row: 1st panel: MSLP analysis for 12 UTC of 29 Aug
2nd panel: MSLP t+84h T_L511L60 forecast started at 00 UTC of 26 Aug
3rd panel: MSLP t+84h EPS-control T_L255L40 forecast started at 00 UTC of 26 Aug
Other rows: 50 EPS-perturbed T_L255L40 forecast started at 00 UTC of 26 Aug.

The contour interval is 5 hPa, with shading patters for MSLP values lower than 990 hPa.

Pourquoi les météorologistes ont-ils tant de peine à prédire le temps avec quelque certitude ? Pourquoi les chutes de pluie, les tempêtes elles-mêmes nous semblent-elles arriver au hasard, de sorte que bien des gens trouvent tout naturel de prier pour avoir la pluie ou le beau temps, alors qu'ils jugeraient ridicule de demander une éclipse par une prière ?[...] un dixième de degré en plus ou en moins en un point quelconque, le cyclone éclate ici et non pas là, et il étend ses ravages sur des contrées qu'il aurait épargnées. Si on avait connu ce dixième de degré, on aurait pu le savoir d'avance, mais les observations n'étaient ni assez serrées, ni assez précises, et c'est pour cela que tout semble dû à l'intervention du hasard.

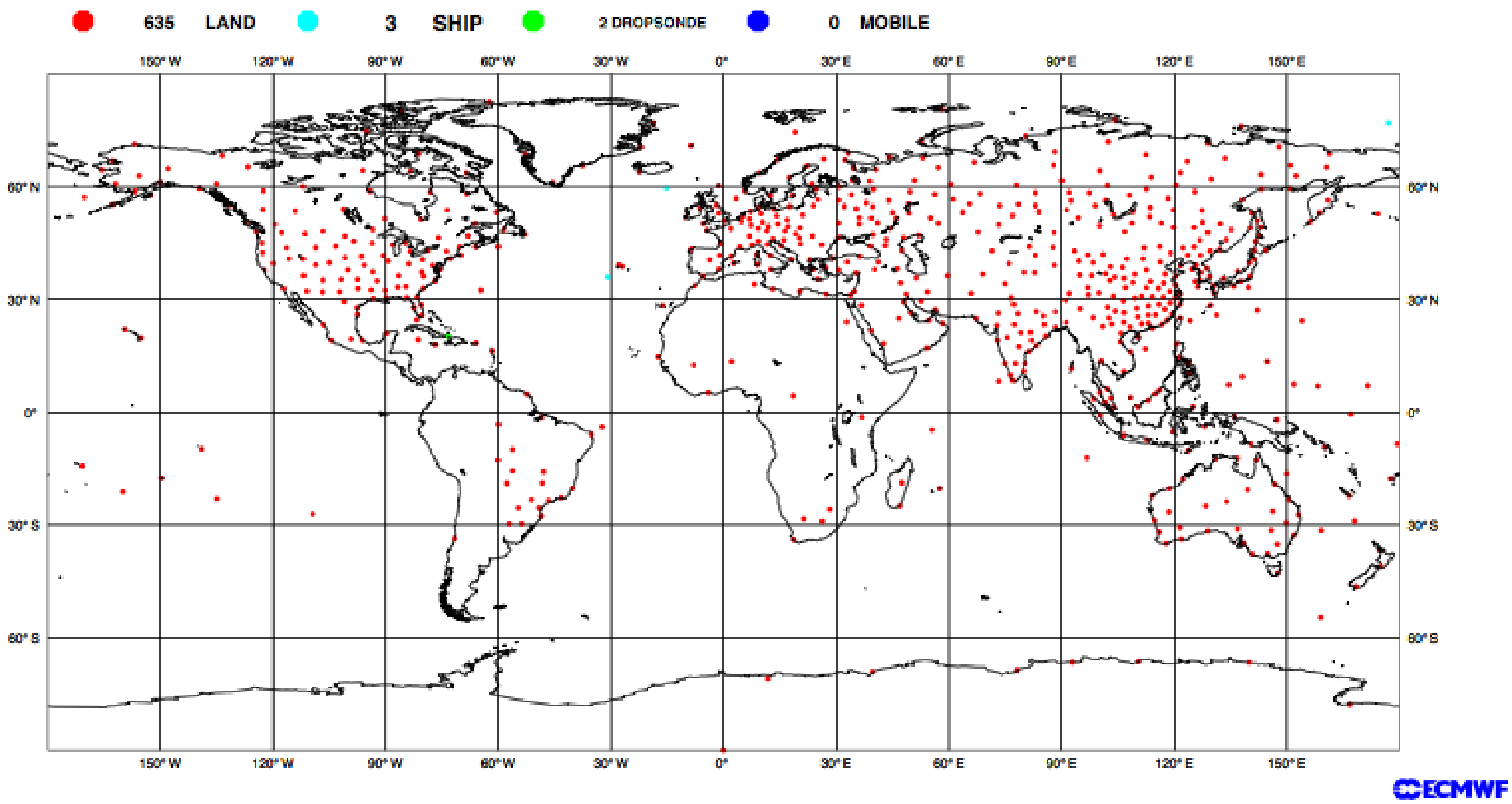
H. Poincaré, *Science et Méthode*, Paris, 1908

ECMWF Data Coverage (All obs DA) - SYNOP/SHIP
06/NOV/2010; 00 UTC
Total number of obs = 31383

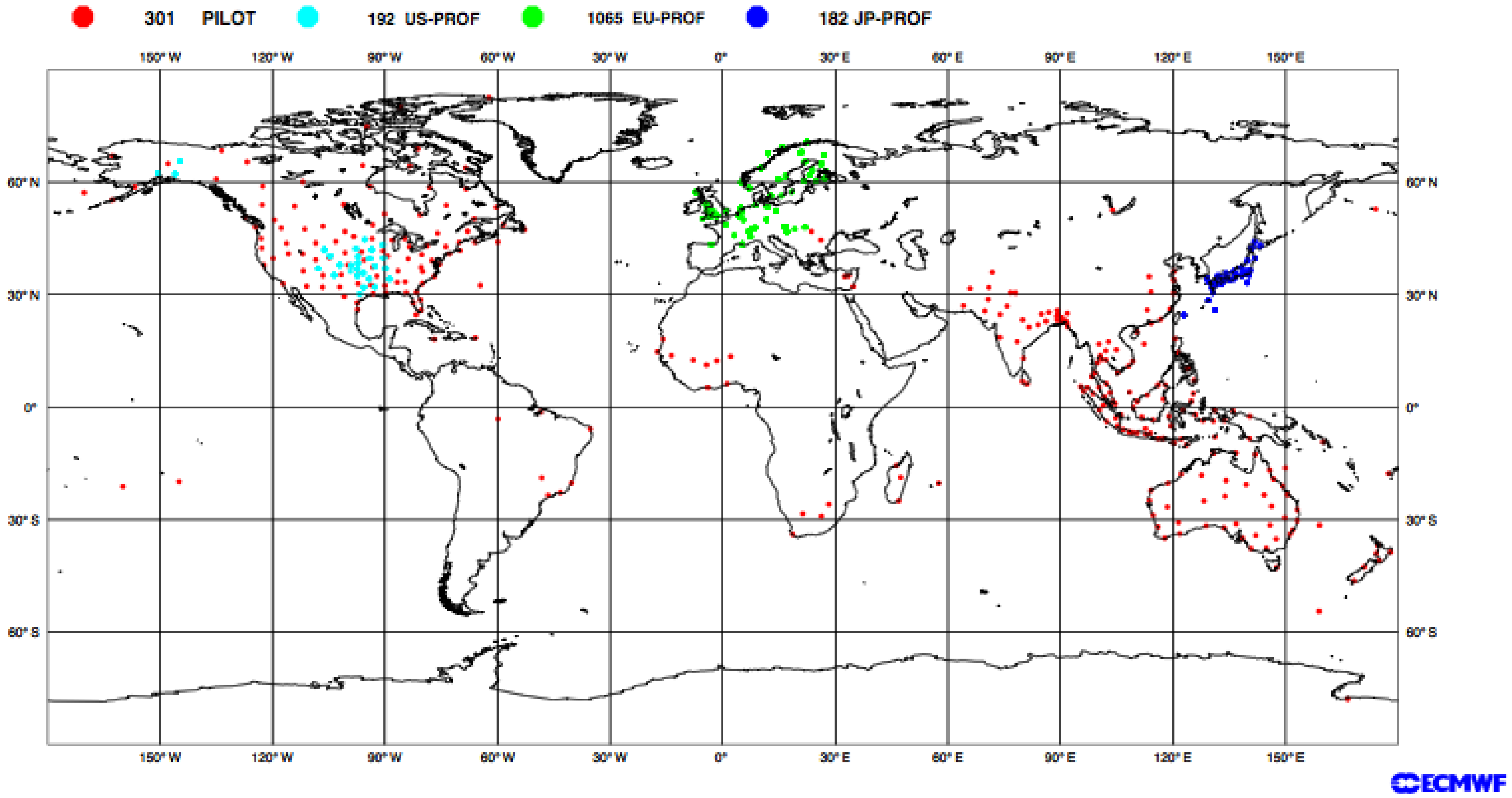


ECMWF

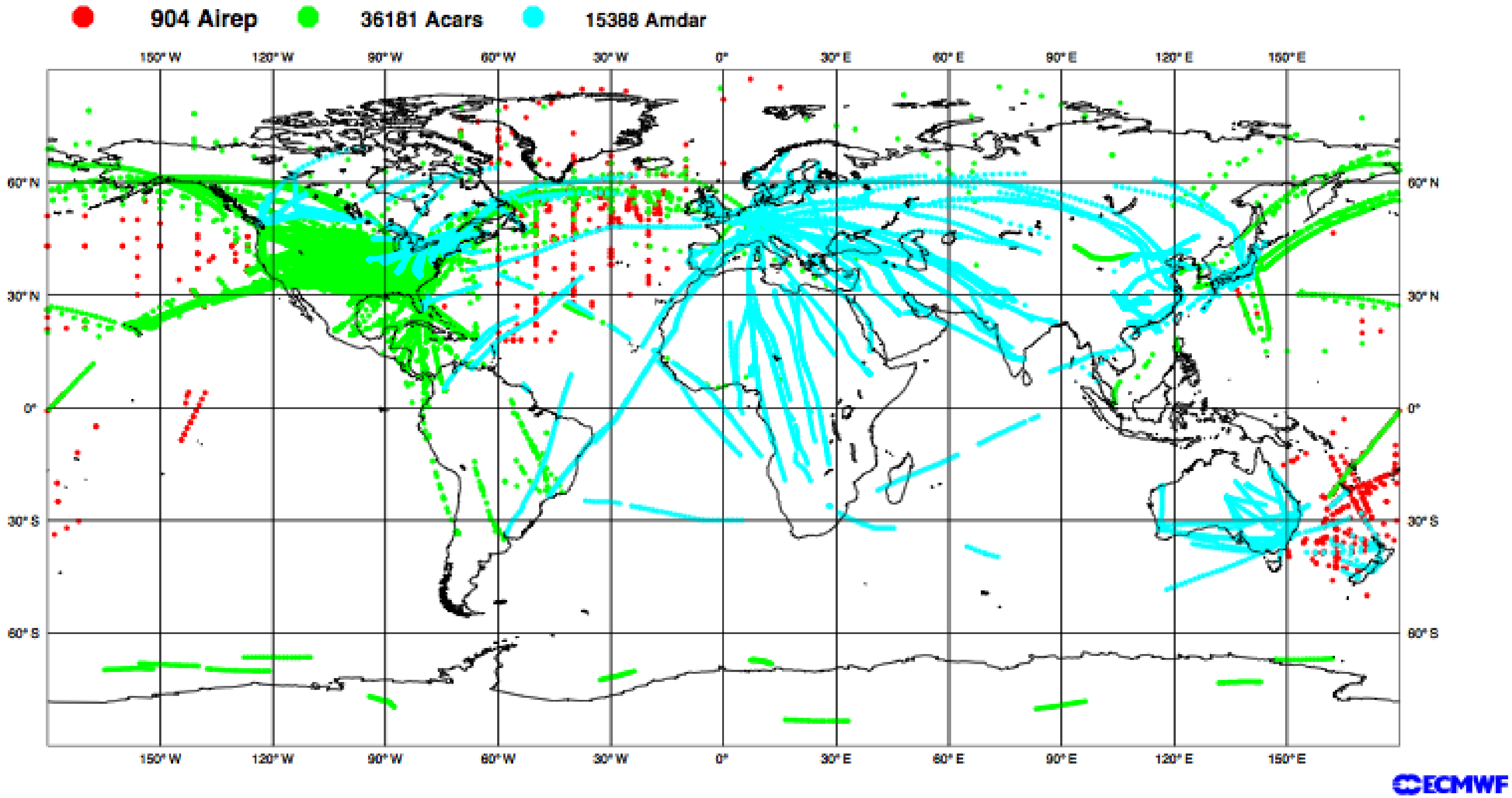
ECMWF Data Coverage (All obs DA) - TEMP
06/NOV/2010; 00 UTC
Total number of obs = 640



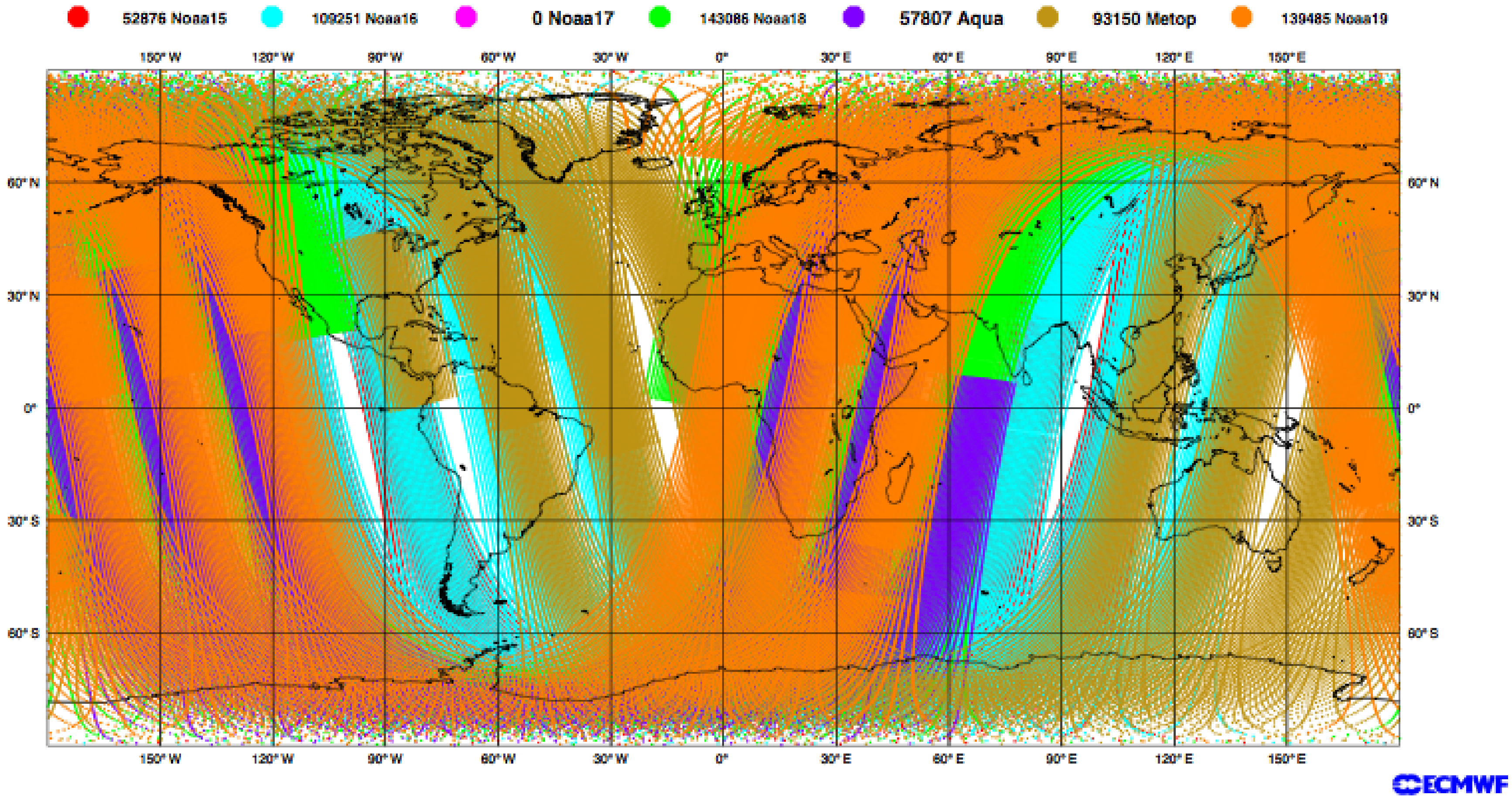
ECMWF Data Coverage (All obs DA) - PILOT/PROFILER
06/NOV/2010; 00 UTC
Total number of obs = 1740



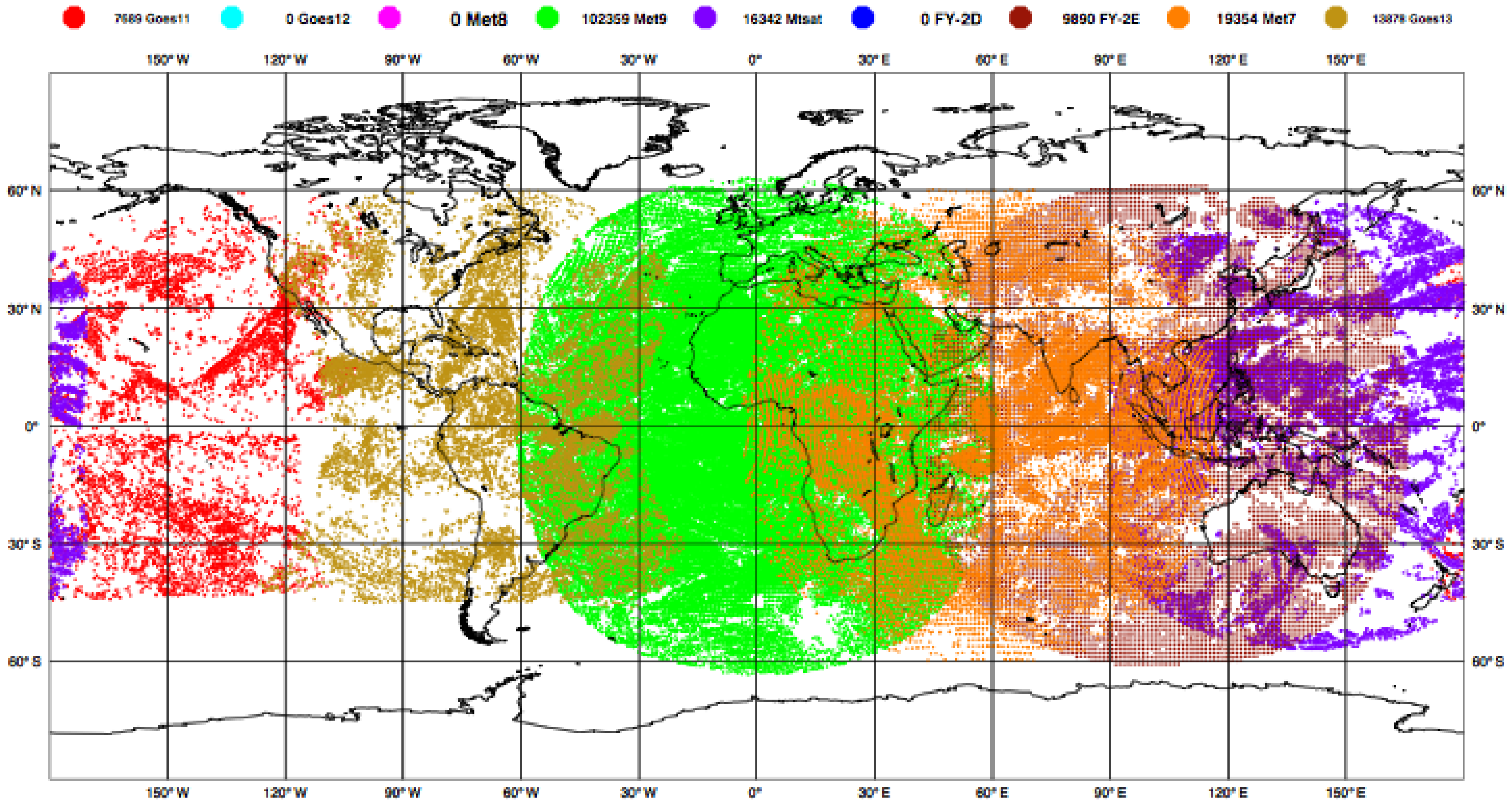
ECMWF Data Coverage (All obs DA) - AIRCRAFT
06/NOV/2010; 00 UTC
Total number of obs = 52473



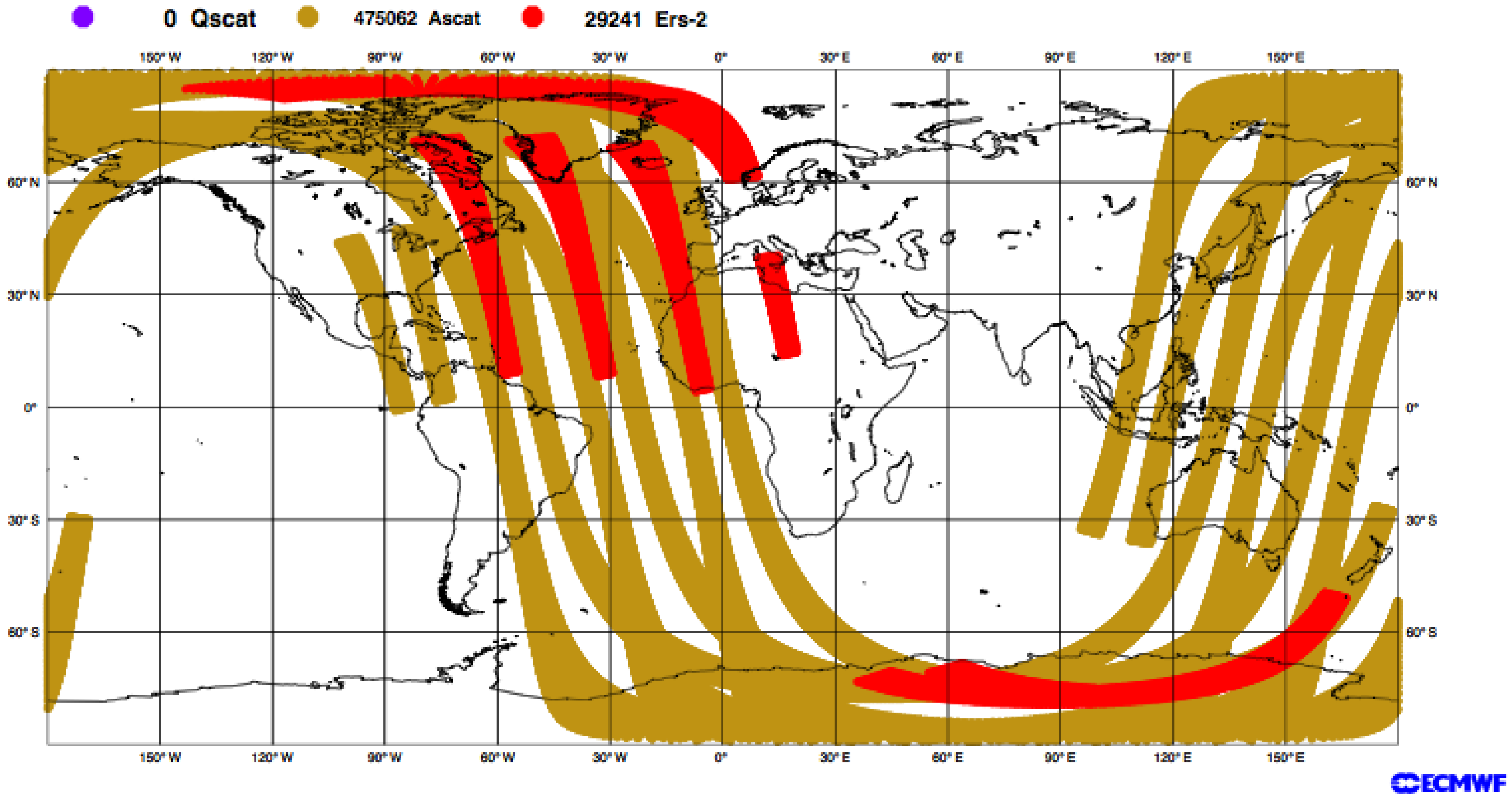
ECMWF Data Coverage (All obs DA) - AMSU-A
06/NOV/2010; 00 UTC
Total number of obs = 595655



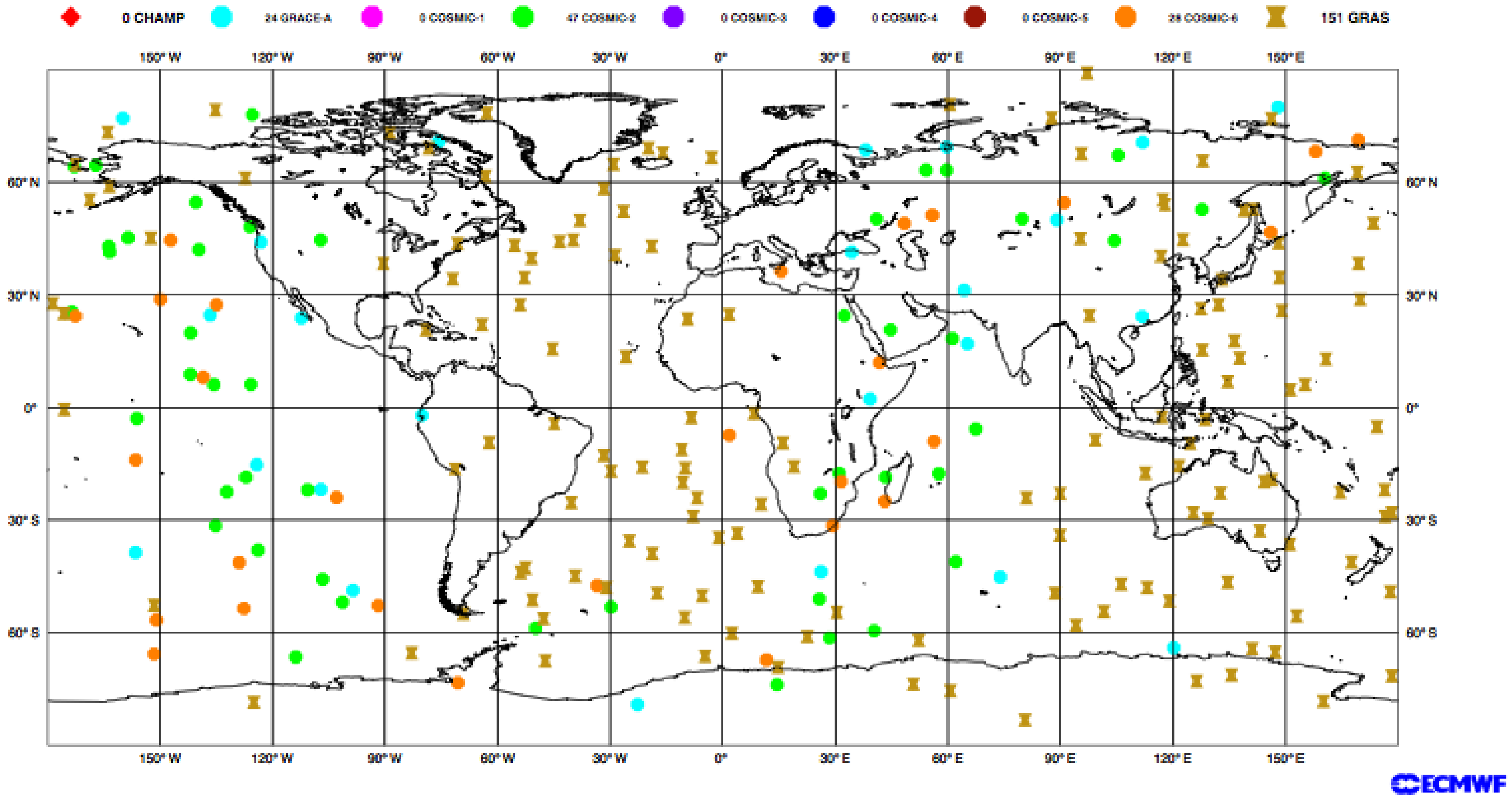
ECMWF Data Coverage (All obs DA) - AMV WV
06/NOV/2010; 00 UTC
Total number of obs = 169512



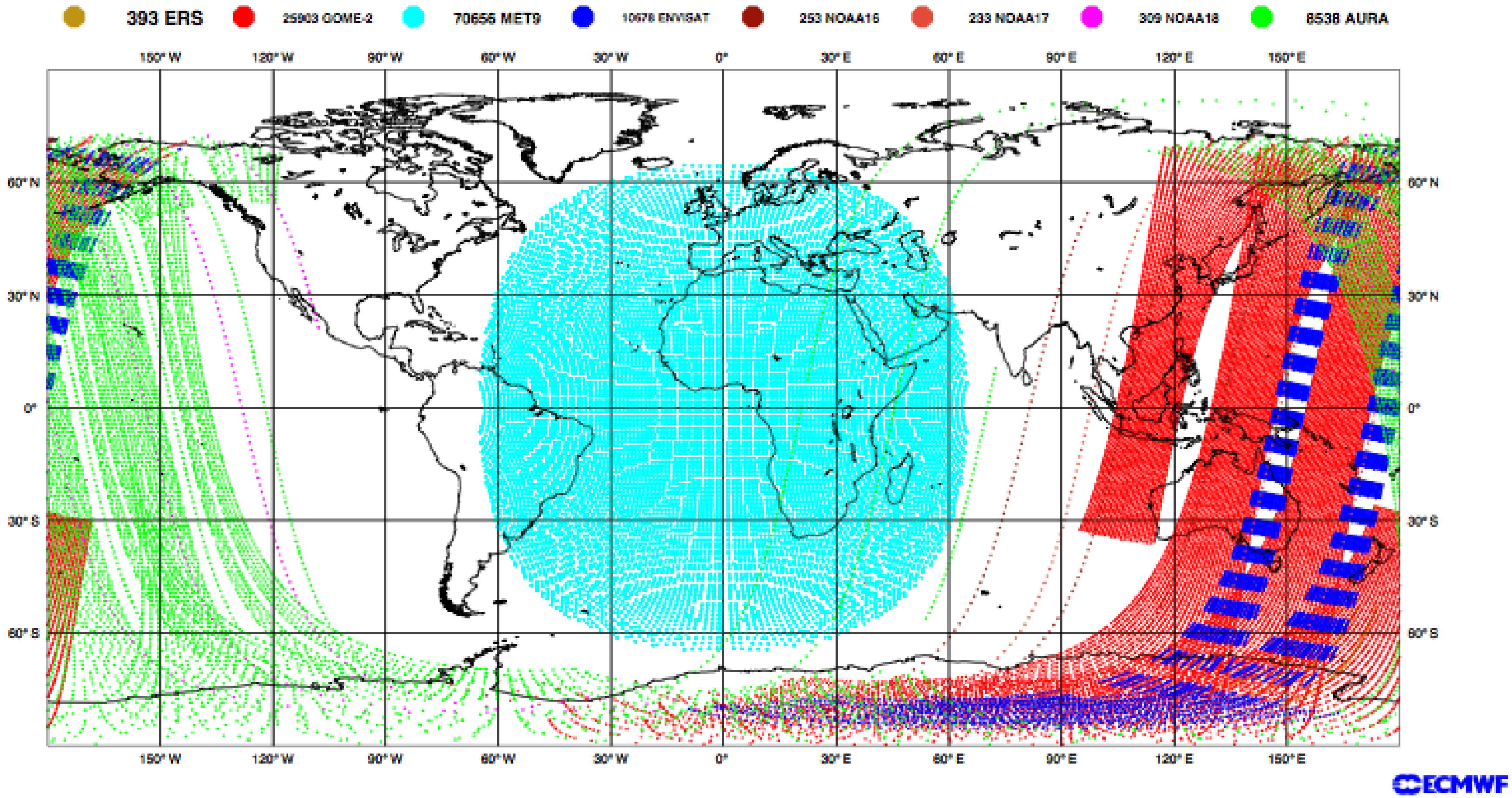
ECMWF Data Coverage (All obs DA) - SCAT
06/NOV/2010; 00 UTC
Total number of obs = 504303



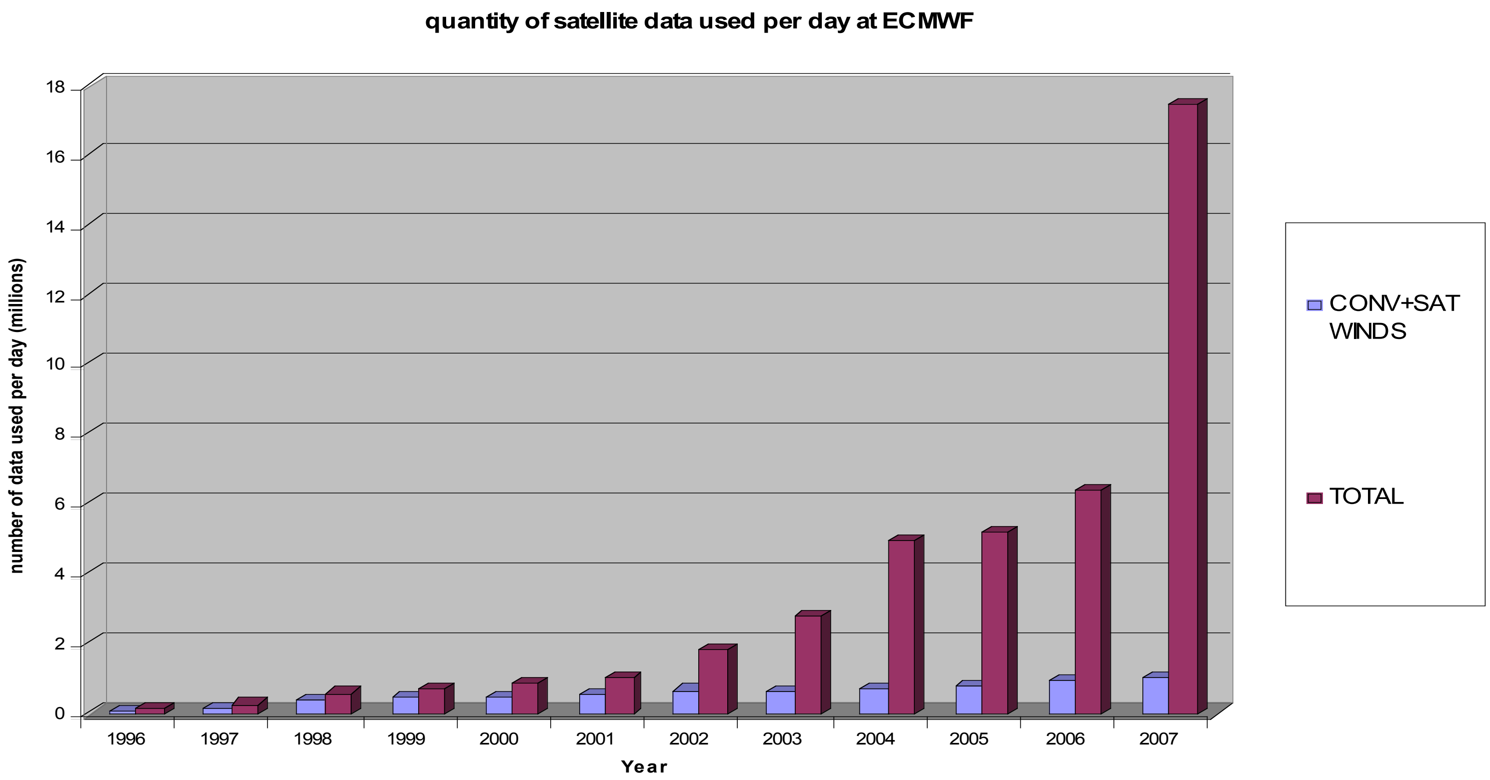
ECMWF Data Coverage (All obs DA) - GPSRO
06/NOV/2010; 00 UTC
Total number of obs = 250



ECMWF Data Coverage (All obs DA) - OZONE
06/NOV/2010; 00 UTC
Total number of obs = 116963



December 2007: Satellite data volumes used: around 18 millions per day



Value as of March 2010 : 25 millions per day

- Observations *synoptiques* (observations au sol, radiosondages), effectuées simultanément, par convention internationale, dans toutes les stations météorologiques du globe (00:00, 06:00, 12:00, 18:00 TU)

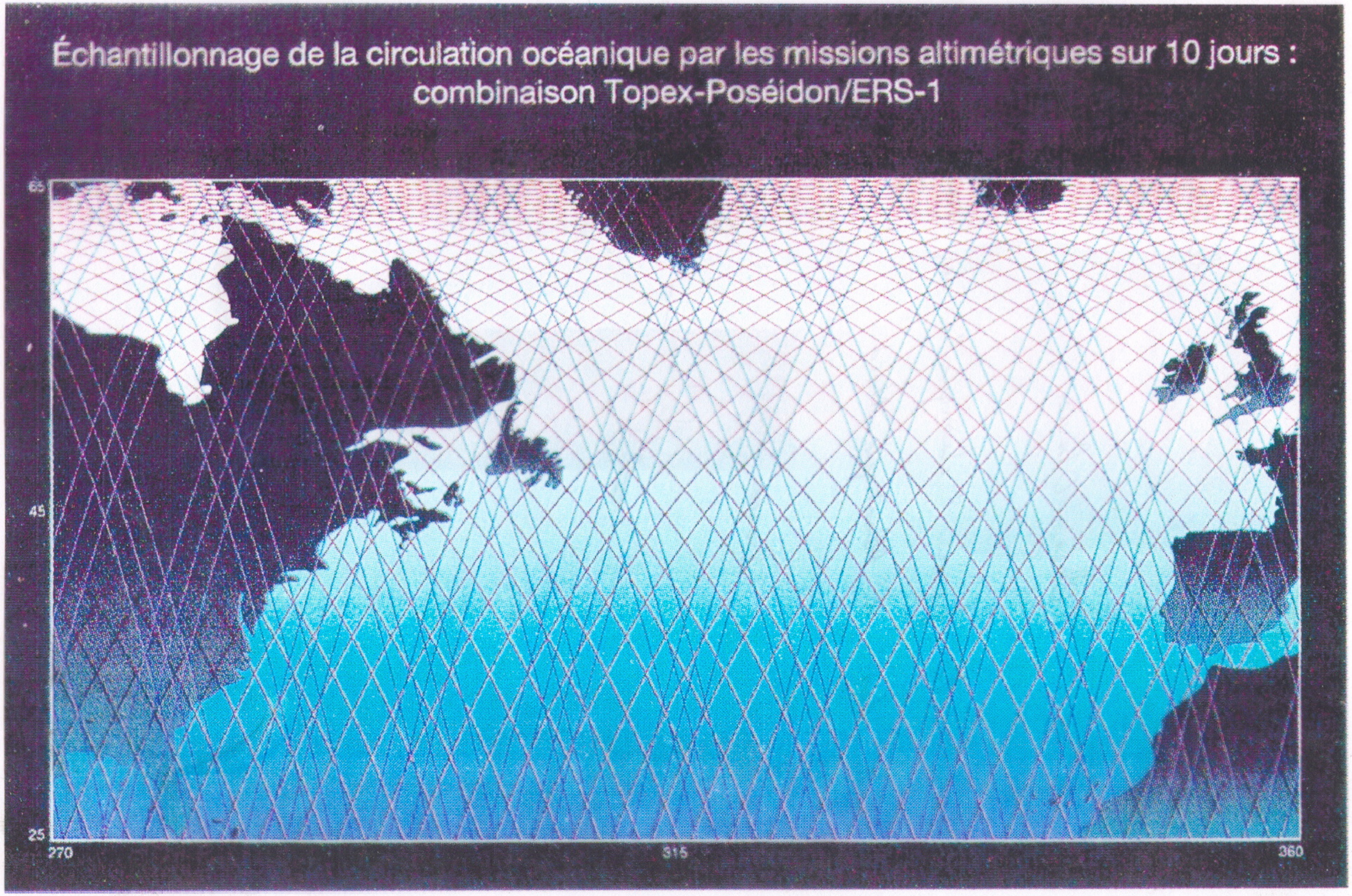
- Observations *asynoptiques* (satellites, avions), effectuées plus ou moins continûment dans le temps.

- Observations *directes* (température, pression, composantes du vent, humidité), portant sur les variables utilisées pour décrire l'état de l'écoulement dans les modèles numériques

- Observations *indirectes* (observations radiométriques, …), portant sur une combinaison plus ou moins complexe (le plus souvent, une intégrale d'espace unidimensionnelle) des variables utilisées pour décrire l'état de l'écoulement

$$y = H(\boldsymbol{x})$$

H : *opérateur d'observation* (par exemple, équation de transfert radiatif)



S. Louvel, Doctoral Dissertation, 1999

Modèles numériques de prévision météorologique

Construits sur les lois physiques qui gouvernent l'évolution de l'écoulement atmosphérique (conservation de la masse, de l'énergie et de la quantité de mouvement), discrétisées de façon appropriée dans l'espace et le temps.

Lois physiques régissant l'écoulement

- Conservation de la masse

  $D\rho/Dt + \rho \, \mathrm{div}\underline{U} = 0$

- Bilan d'énergie interne

  $De/Dt - (p/\rho^2) \, D\rho/Dt = Q$

- Bilan de quantité de mouvement

  $D\underline{U}/Dt + (1/\rho) \, \underline{\mathrm{grad}}p - \underline{g} + 2 \, \underline{\Omega} \wedge \underline{U} = \underline{F}$

- Equation d'état thermodynamique

  $f(p, \rho, e) = 0$ ($p/\rho = rT$, $e = C_v T$ pour un gaz parfait)

- Bilan de masse pour les composants secondaires (eau pour l'atmosphère, sel pour l'océan, …)

  $Dq/Dt + q \, \mathrm{div}\underline{U} = S$

Vocabulaire du métier :

- Processus adiabatiques et inviscides, et donc thermodynamiquement réversibles (tout sauf Q, $\underline{F}$ et S) :: *'dynamique'*

- Processus décrits par les termes Q, $\underline{F}$ et S : *'physique'*

Plusieurs hypothèses simplificatrices sont faites pour les besoins de la modélisation du climat

- Dans la direction verticale, approximation *hydrostatique* :

$\partial p/\partial z + \rho g = 0$

Élimine l'équation du mouvement pour la direction verticale; en outre, l'écoulement est incompressible dans les coordonnées (x, y, p) $\Rightarrow$ nombre d'équations diminué de deux unités.

Approximation hydrostatique valide dans l'atmosphère pour échelles horizontales > 20-30 km

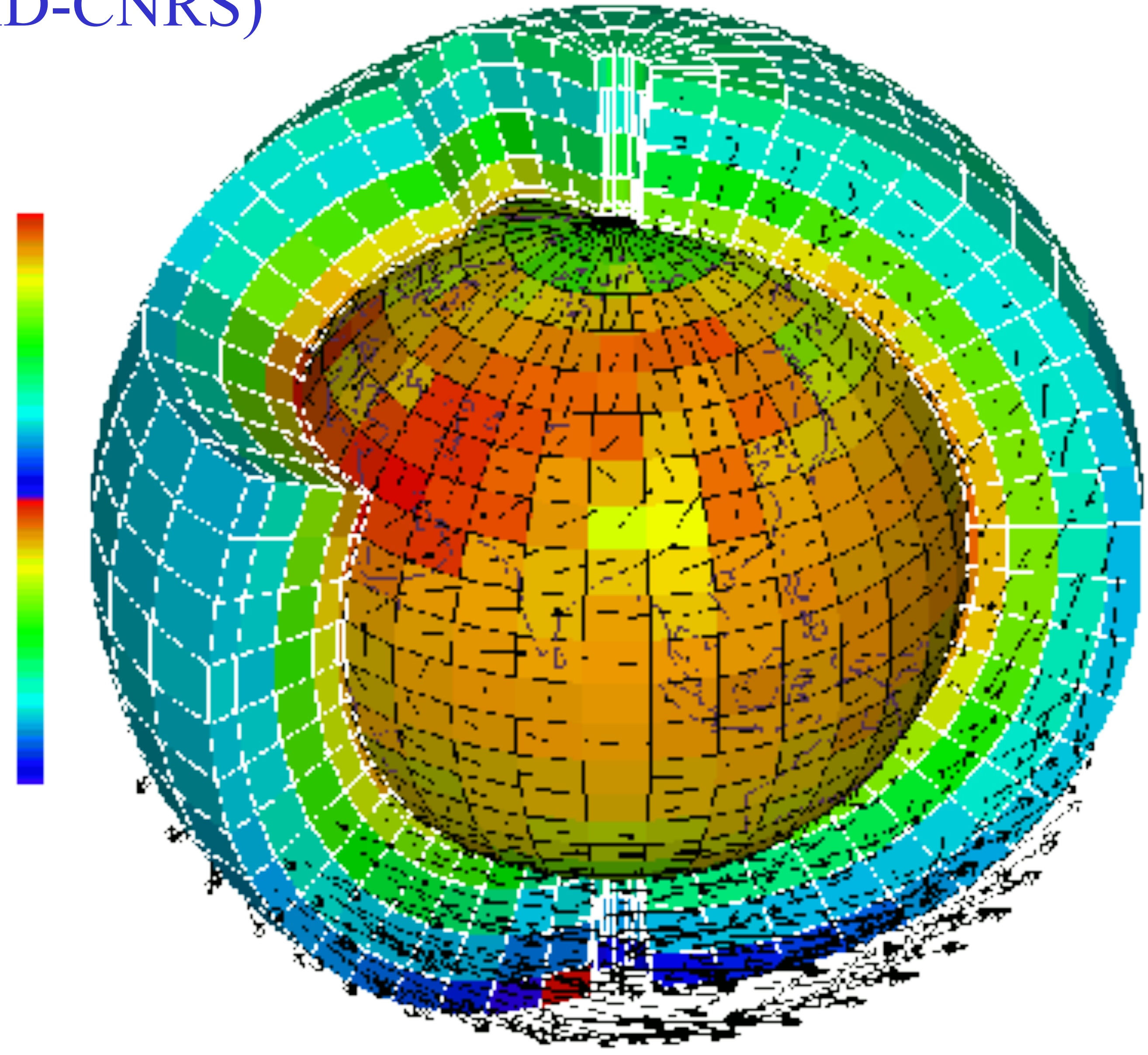
- Atmosphère et océan sont contenus dans une couche sphérique d'épaissseur négligeable devant le rayon de la Terre

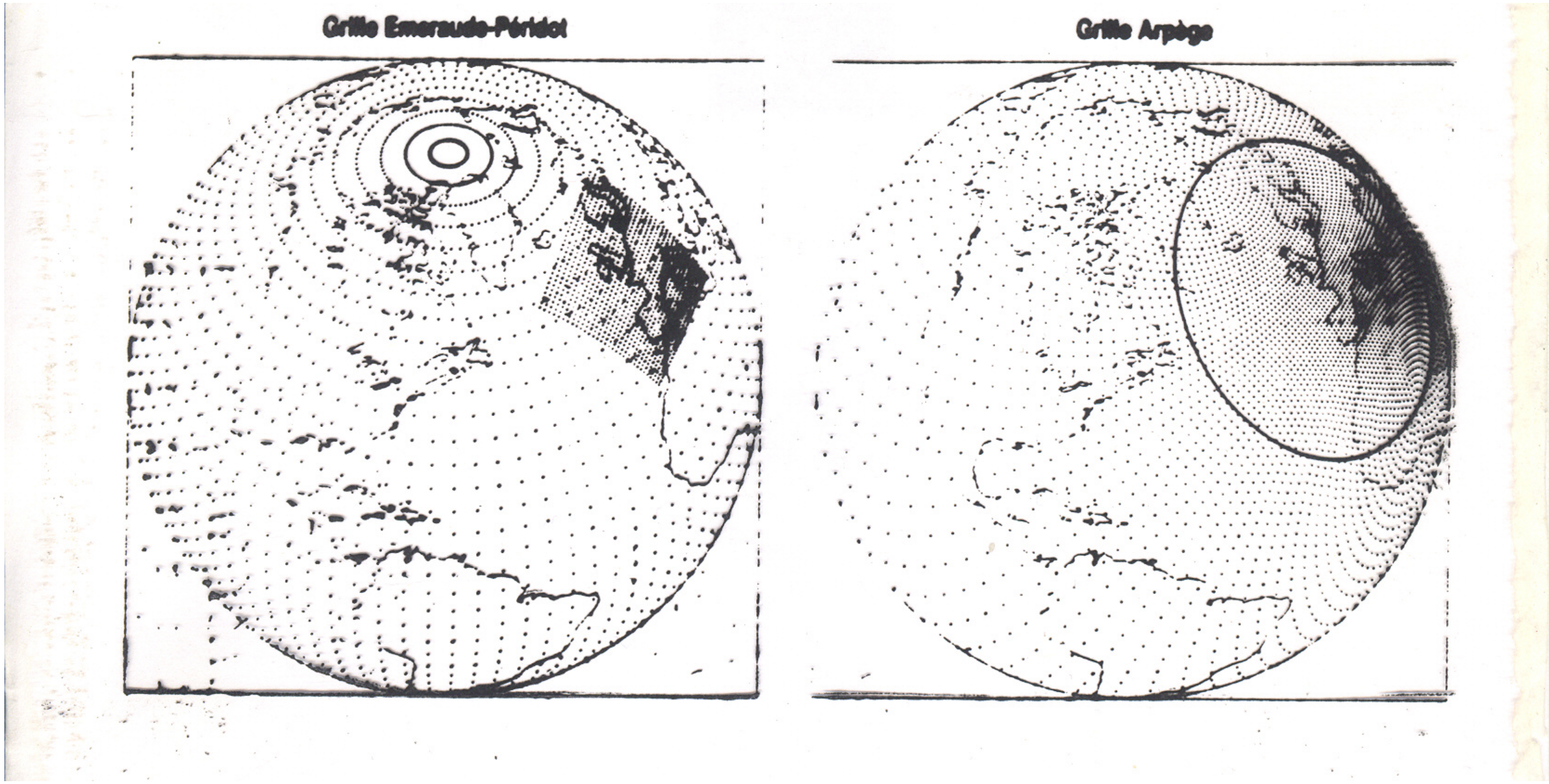
- …

$\Rightarrow$ Equations dites *primitives*

Modèles non-hydrostatiques, plus coûteux, sont utilisés pour la météorologie de petite échelle.

Schéma de principe d'un modèle atmosphérique
(L. Fairhead /LMD-CNRS)





Grilles de modèles de Météo-France *(La Météorologie)*

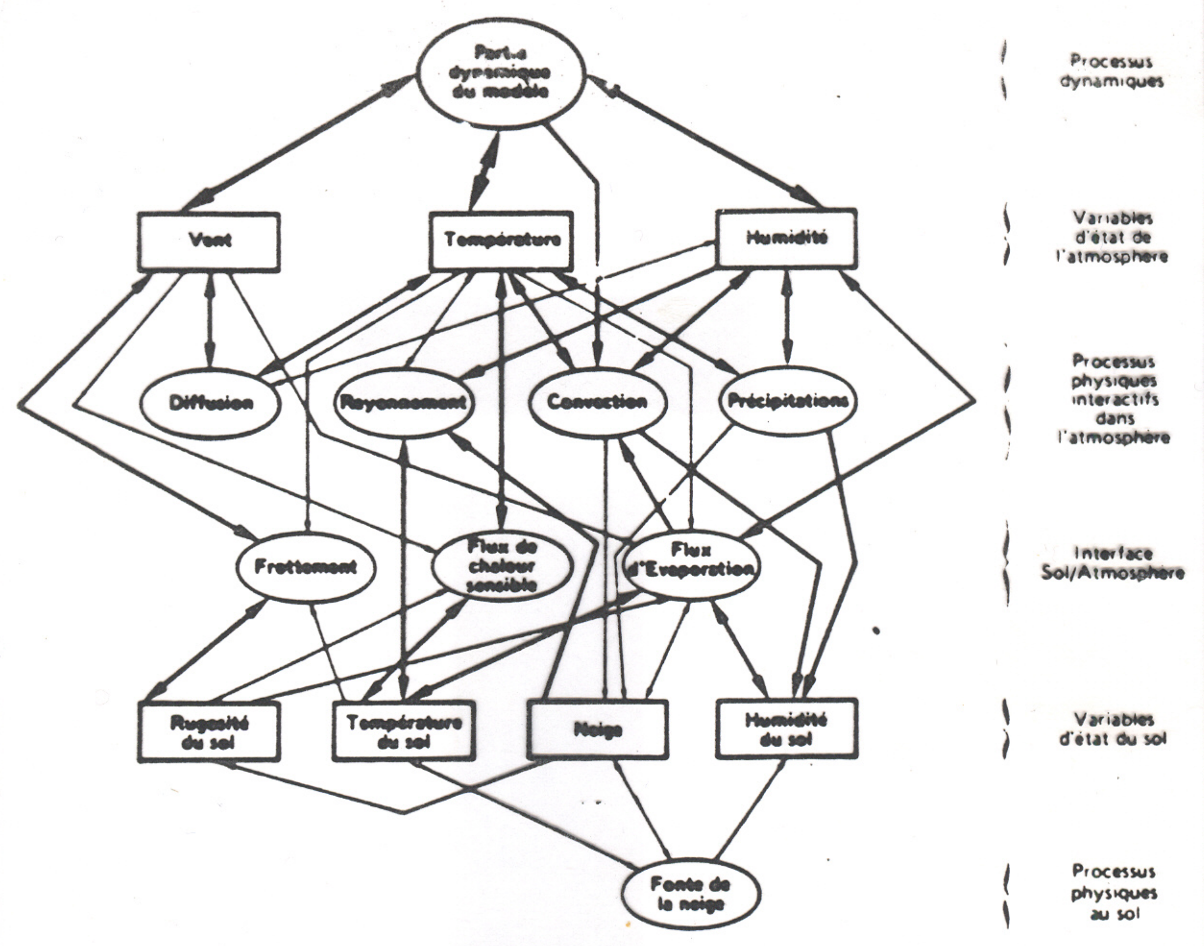
Discrétisation spatiale

Deux grandes classes de discrétisation

- discrétisation en points de grille (peu de volumes finis, en particulier pour l'atmosphère)

- discrétisation (semi-) spectrale, suivant les harmoniques sphériques. Seules les opérations linéaires relatives à la 'dynamique' sont effectuées dans l'espace spectral, les opérations non-linéaires et les opérations relatives à la 'physique' sont effectuées dans l'espace physique. Nécessité de passer en permanence d'un espace à l'autre. Possible grâce à l'utilisation des Transformées de Fourier Rapides (FFT)

5 · SCHEMA DES INTERACTIONS PHYSIQUES DANS LE MODELE



Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT, Reading, GB)

(European Centre for Medium-range Weather Forecasts, ECMWF)

Depuis le 26 Janvier 2010

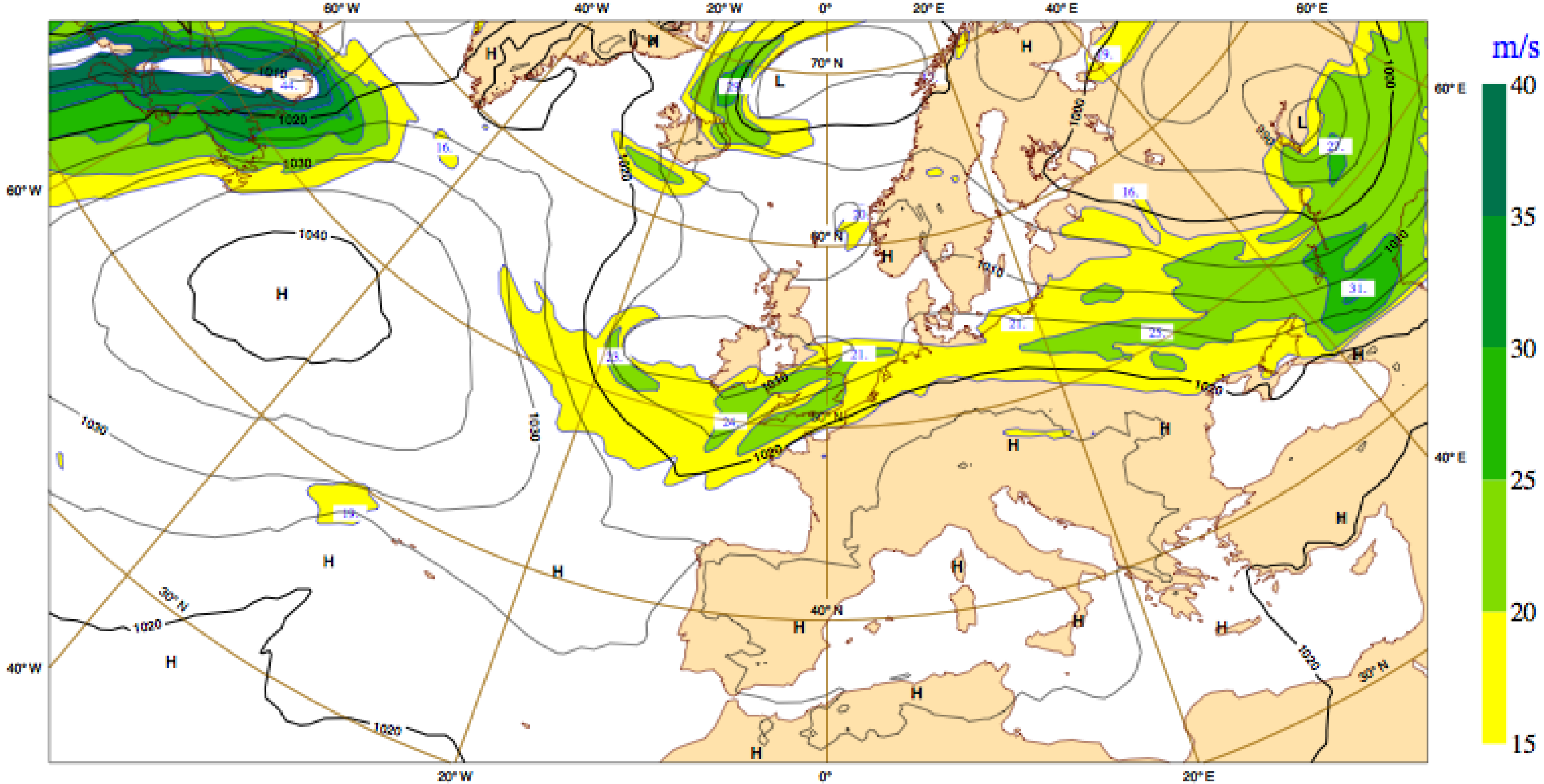
Troncature 'triangulaire' T1279 (résolution horizontale ≈ 16 kilomètres)

91 niveaux dans la direction verticale (0 - 80 km)

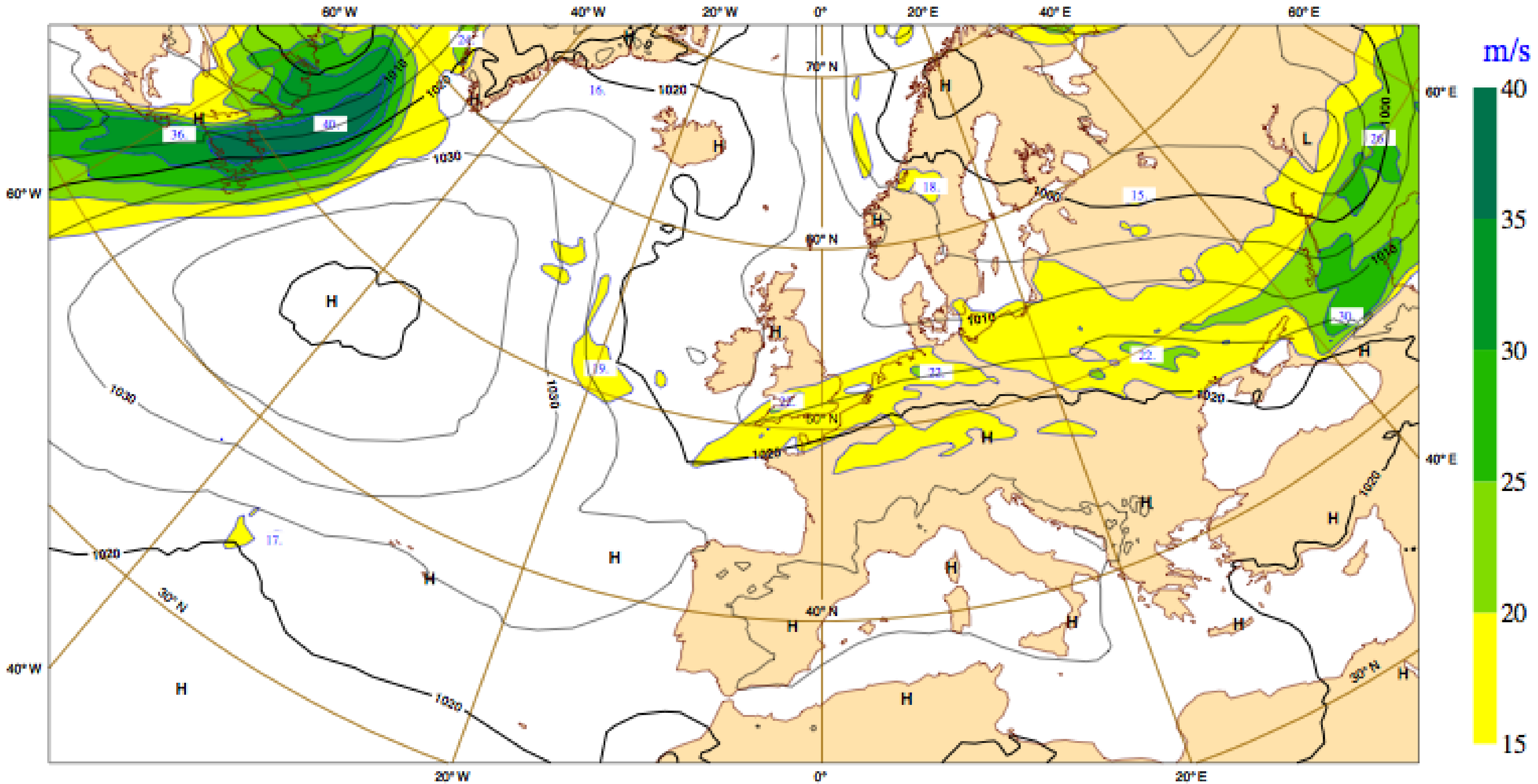
Dimension du vecteur d'état correspondant $\approx 1{,}5\ 10^9$

Pas de discrétisation temporelle : 10 minutes

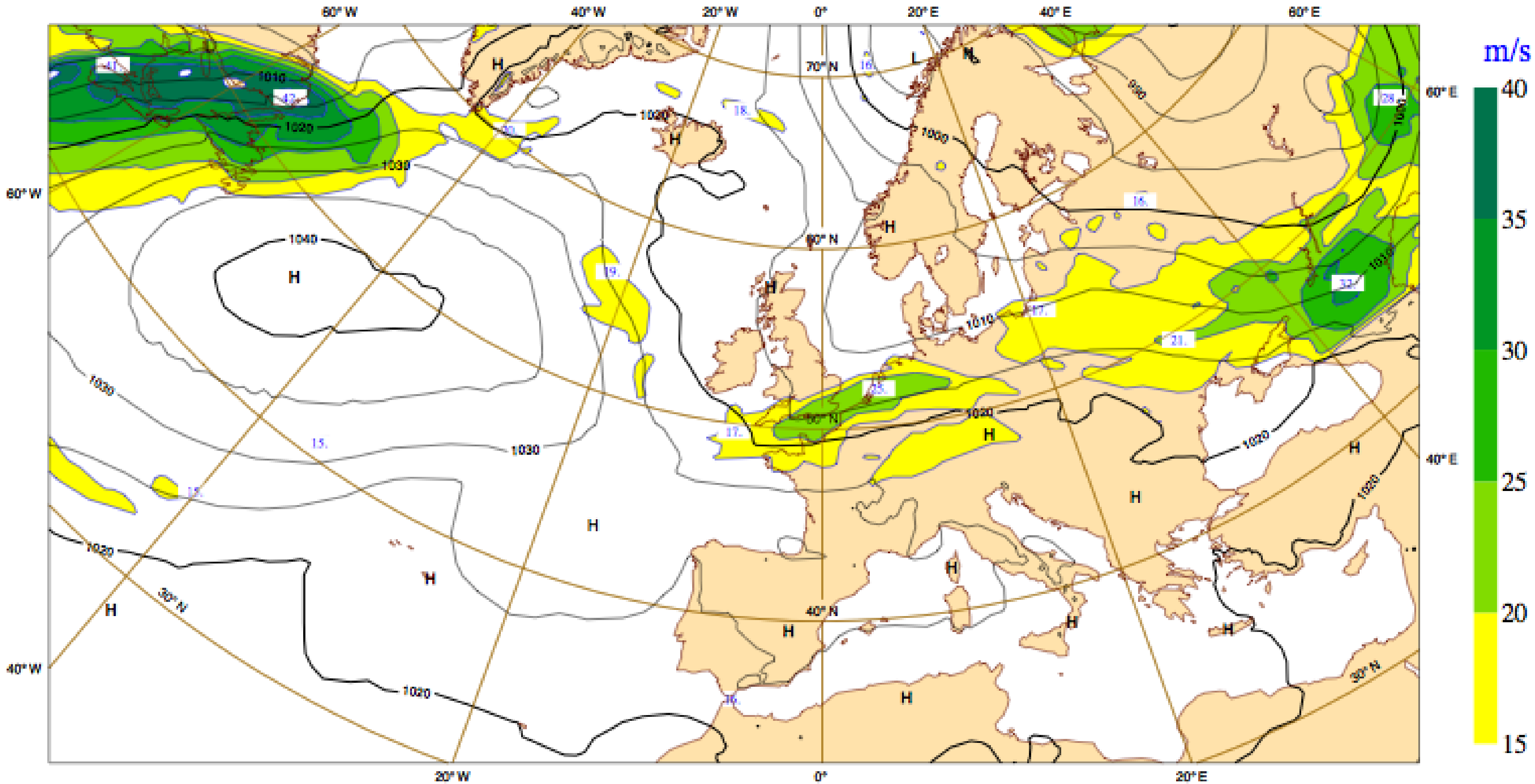
Sunday 31 October 2010 00UTC ©ECMWF Forecast t+144 VT: Saturday 6 November 2010 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



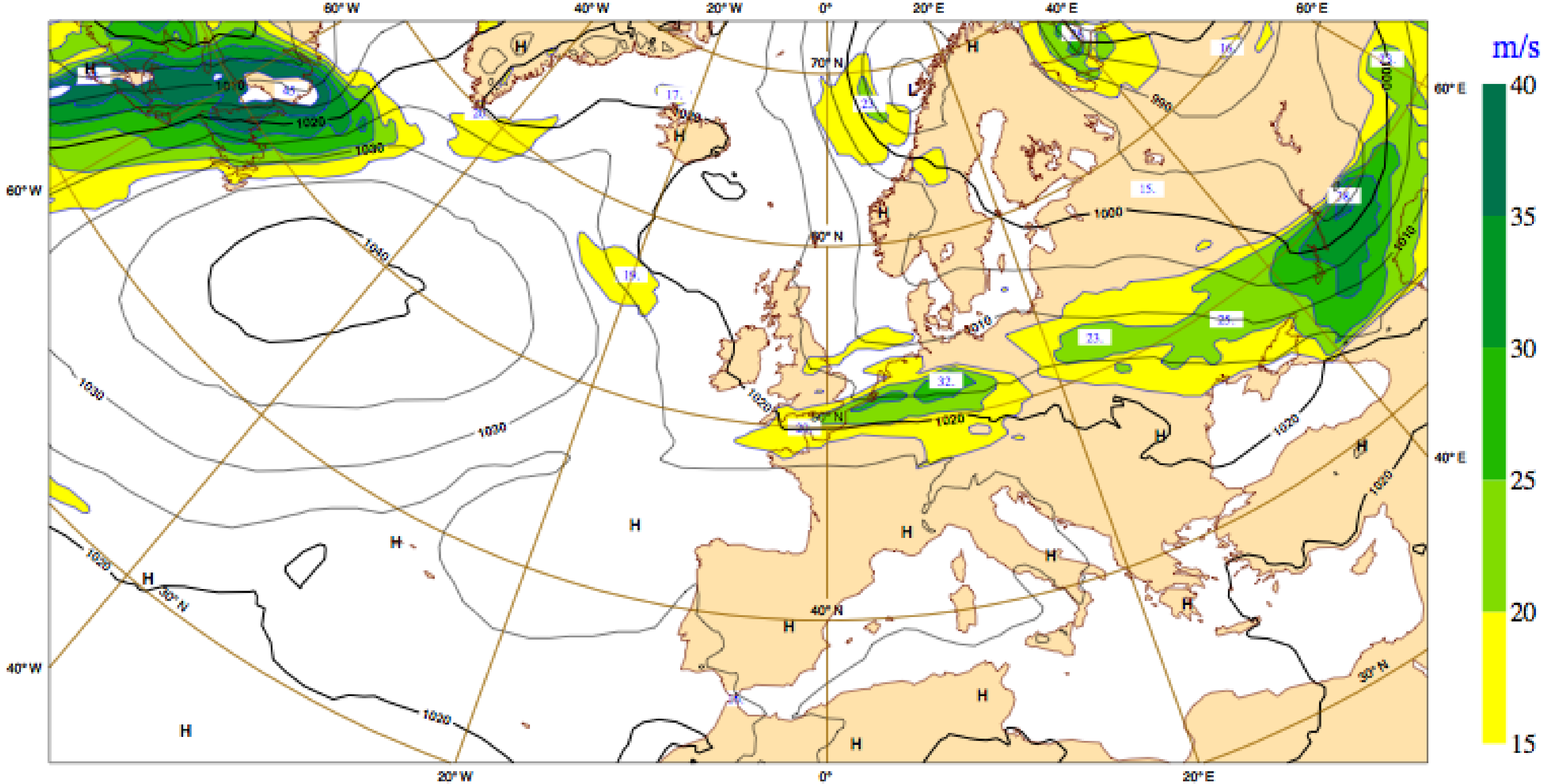
Tuesday 2 November 2010 00UTC ©ECMWF Forecast t+096 VT: Saturday 6 November 2010 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



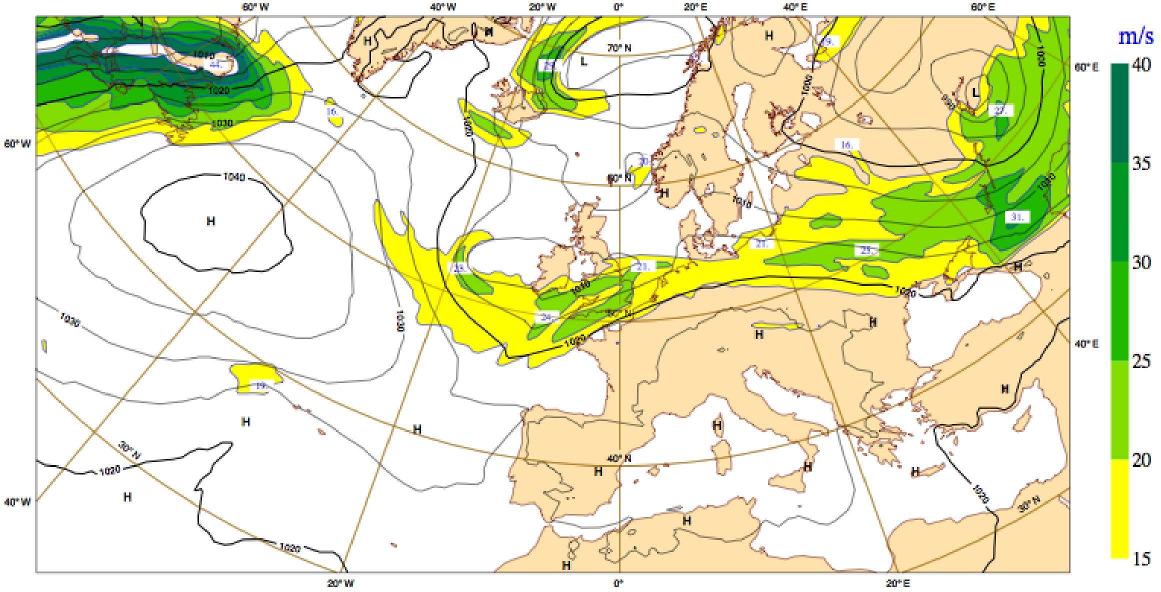
Thursday 4 November 2010 00UTC ©ECMWF Forecast t+048 VT: Saturday 6 November 2010 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



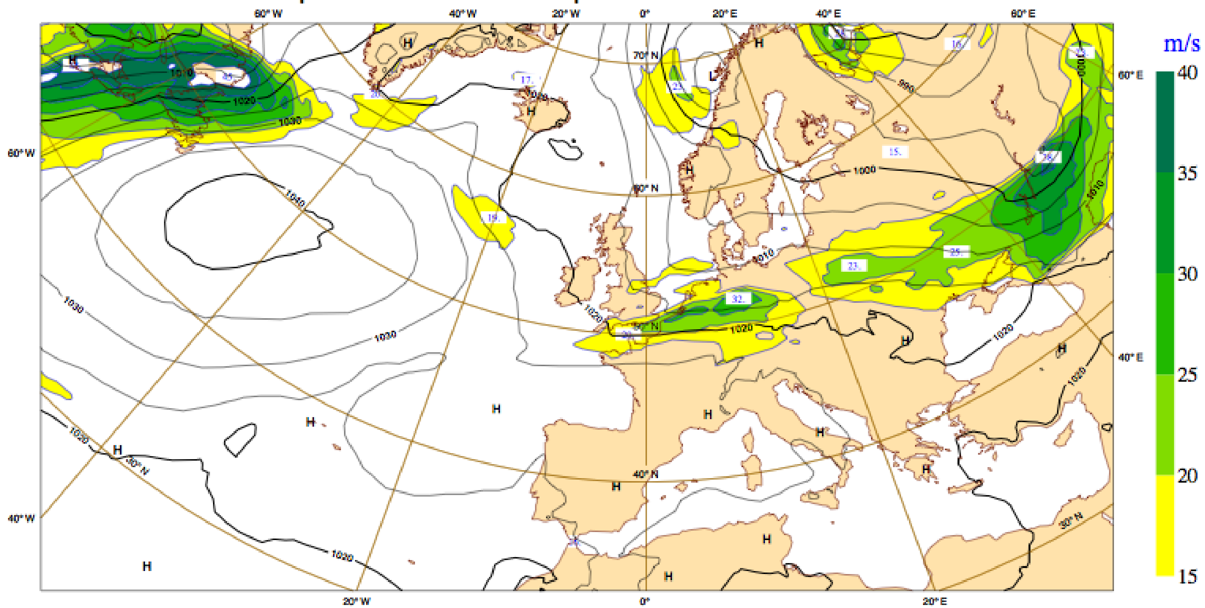
Saturday 6 November 2010 00UTC ©ECMWF Analysis t+000 VT: Saturday 6 November 2010 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



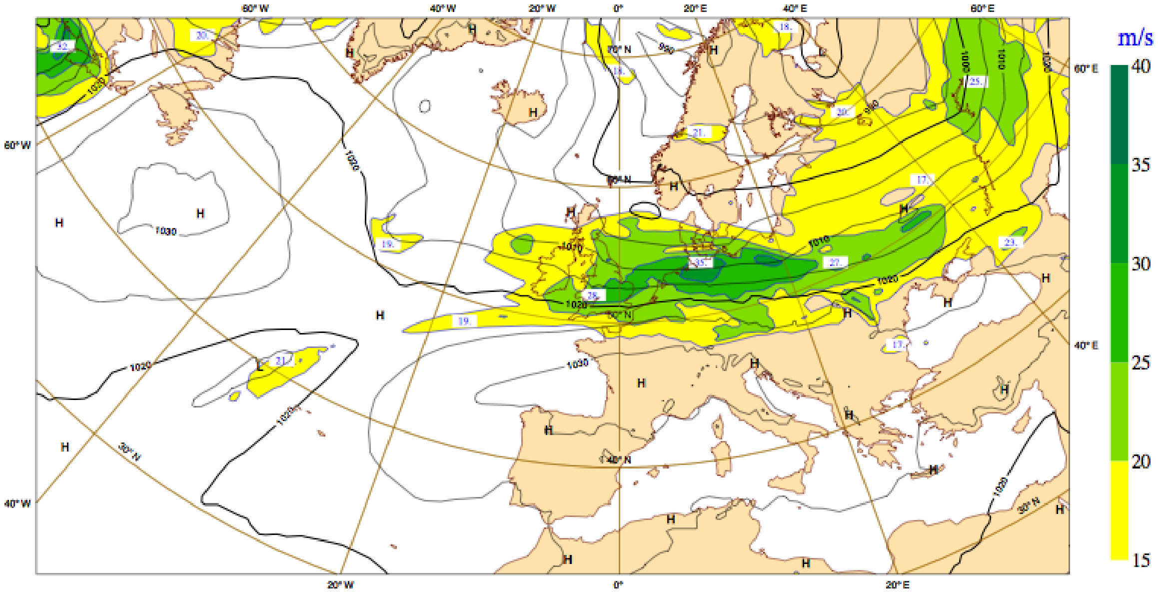
Sunday 31 October 2010 00UTC ©ECMWF Forecast t+144 VT: Saturday 6 November 2010 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



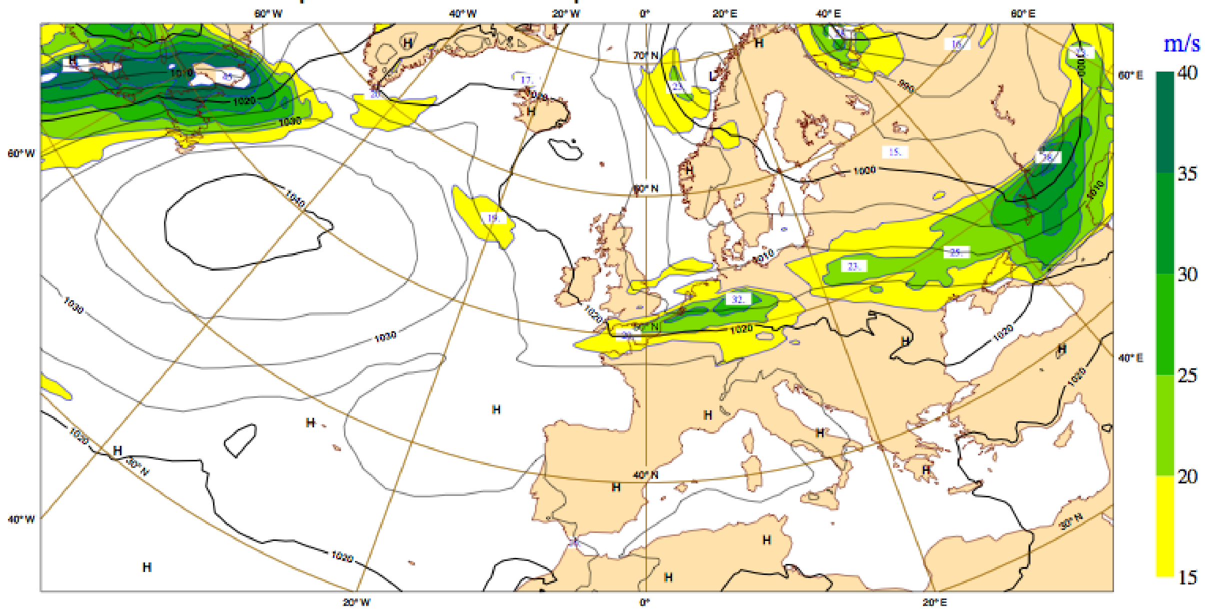
Saturday 6 November 2010 00UTC ©ECMWF Analysis t+000 VT: Saturday 6 November 2010 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



Friday 5 November 2010 00UTC ©ECMWF Analysis t+000 VT: Friday 5 November 2010 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



Saturday 6 November 2010 00UTC ©ECMWF Analysis t+000 VT: Saturday 6 November 2010 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



Résultats extraits de

Richardson *et al.*, 2009, *Verification statistics and evaluations of ECMWF forecasts in 2008-2009*, Memorandum Technique 606 CEPMMT, Reading, GB.

Disponible à l'adresse

http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/601-700/tm606.pdf

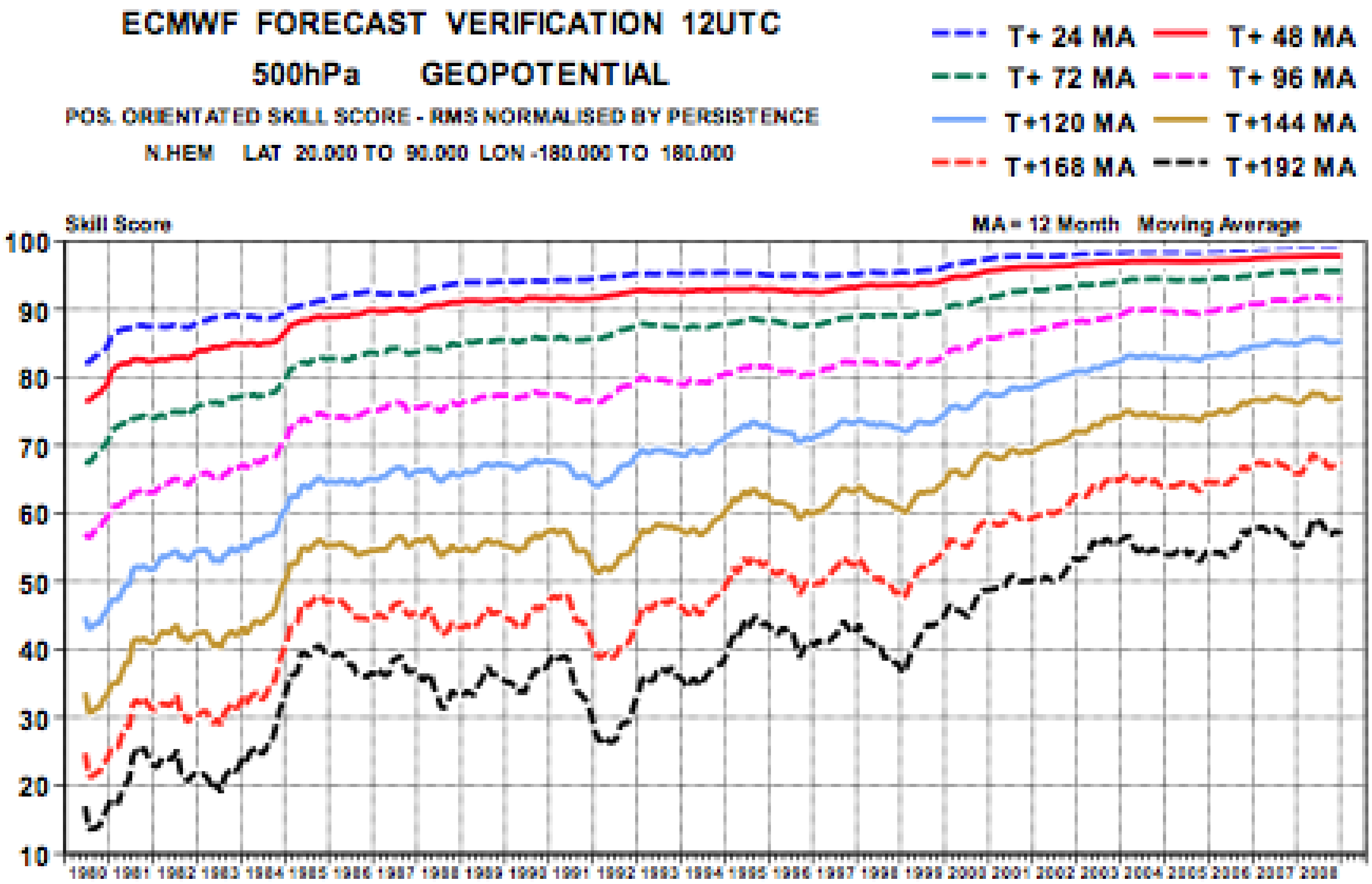


Figure 1: 500 hPa height skill score for Europe (top) and the northern hemisphere extra-tropics (bottom), 12-month moving averages, forecast ranges from 24 to 192 hours. The last point on each curve is for the 12-month period August 2008 - July 2009.

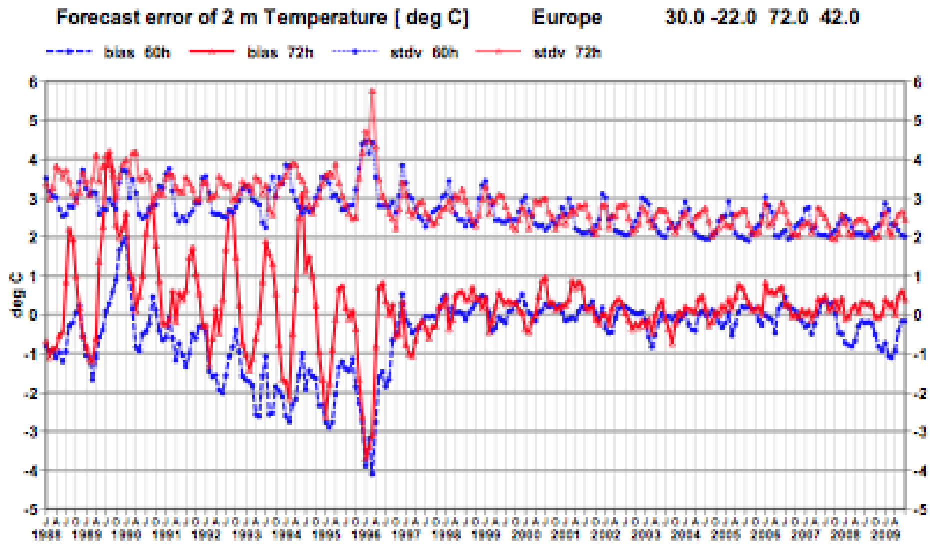


Figure 15: Verification of 2 metre temperature forecasts against European SYNOP data on the GTS for 60-hour (night-time) and 72-hour (daytime) forecasts. Lower pair of curves are bias, upper curves are standard deviation of error.

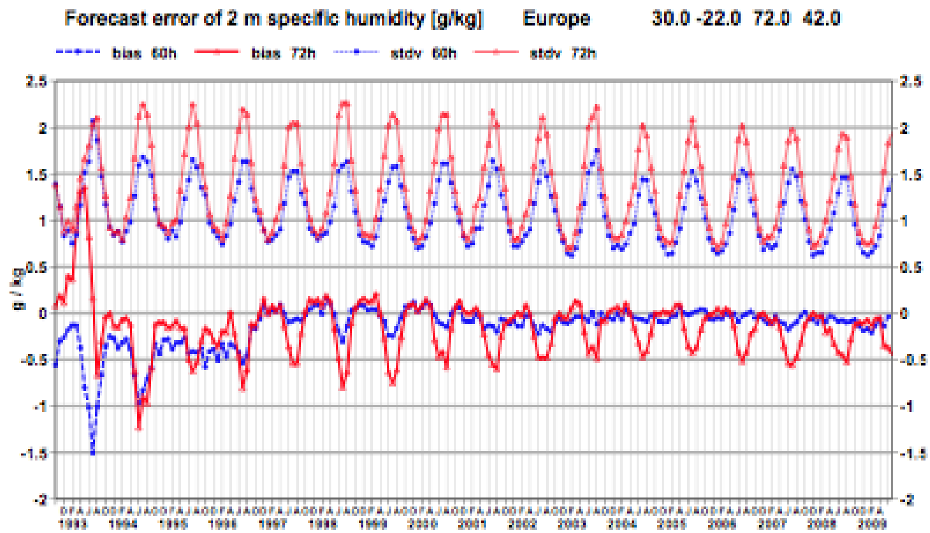
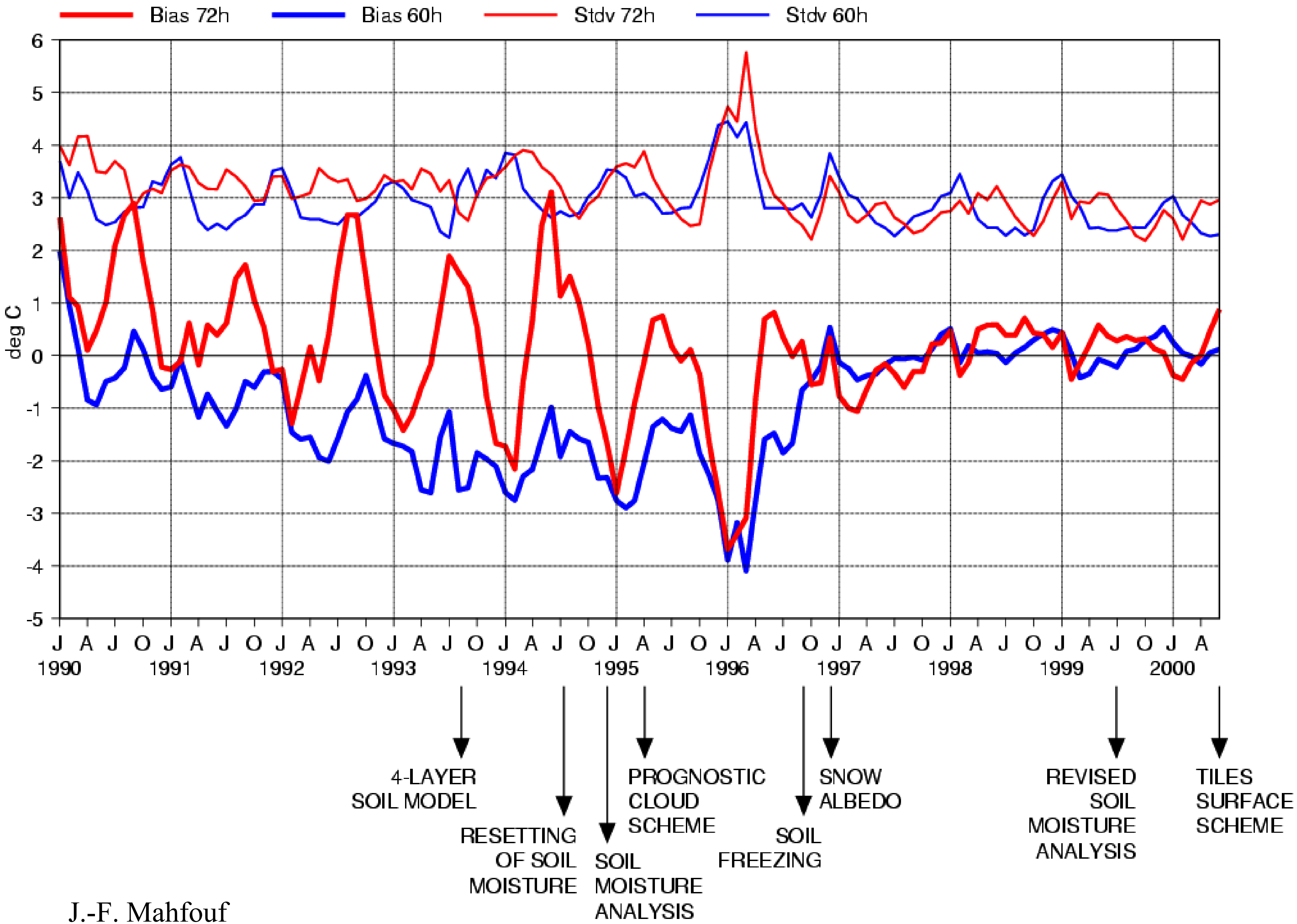


Figure 16: Verification of 2 metre specific humidity forecasts against European SYNOP data on the GTS for 60-hour (night-time) and 72-hour (daytime) forecasts. Lower pair of curves is bias, upper curves are standard deviation of error.



Bias 72h
Bias 60h
Stdv 72h
Stdv 60h
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
deg C
J A J O J A J O J A J O J A J O J A J O J A J O J A J O J A J O J A J O J A J O J A
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
4-LAYER SOIL MODEL
RESETTING OF SOIL MOISTURE
SOIL MOISTURE ANALYSIS
PROGNOSTIC CLOUD SCHEME
SOIL FREEZING
SNOW ALBEDO
REVISED SOIL MOISTURE ANALYSIS
TILES SURFACE SCHEME

J.-F. Mahfouf

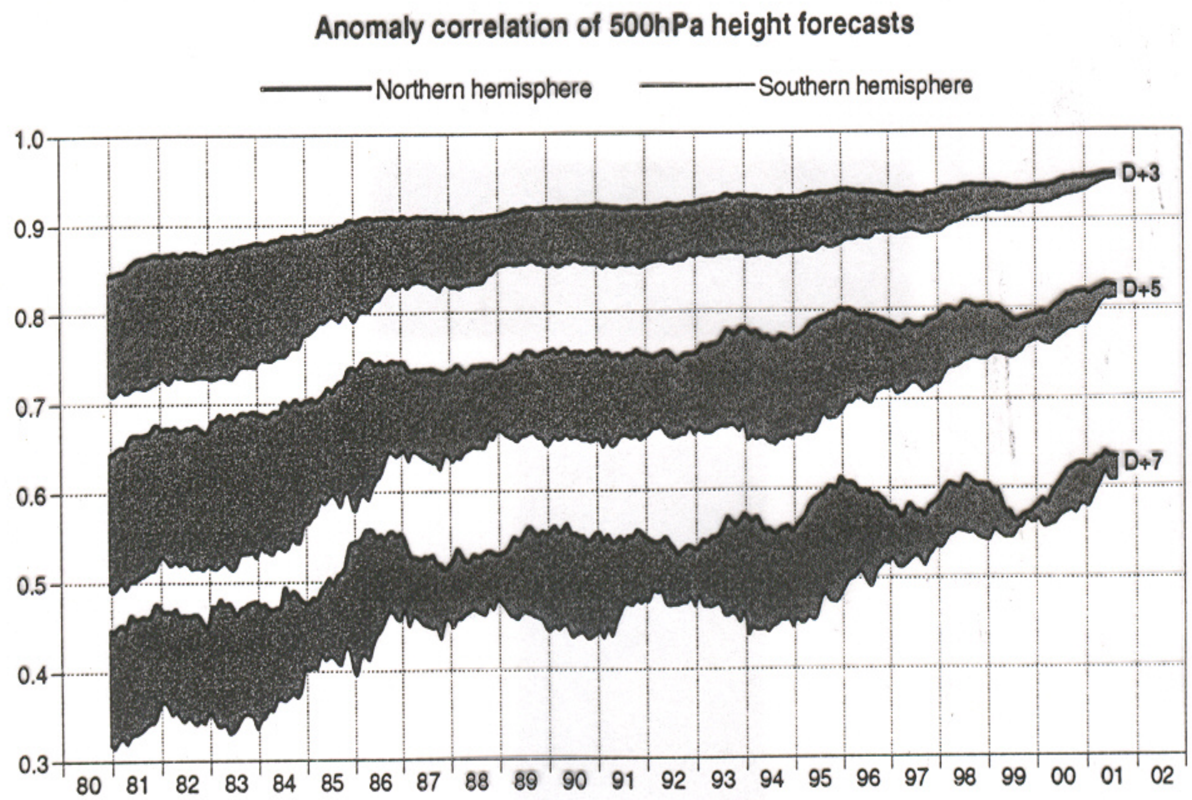


Fig 4. Anomaly correlation coefficients of 3-, 5- and 7-day ECMWF 500hPa height forecasts for the extratropical northern and southern hemispheres, plotted in the form of annual running means of archived monthly-mean scores for the period from January 1980 to August 2001. Values plotted for a particular month are averages over that month and the 11 preceding months. The shading shows the differences in scores between the two hemispheres at the forecast ranges indicated.

Simmons et Hollingsworth, 2002, *Q. J. R. Meteorol. Soc.,* **128**, 647-677

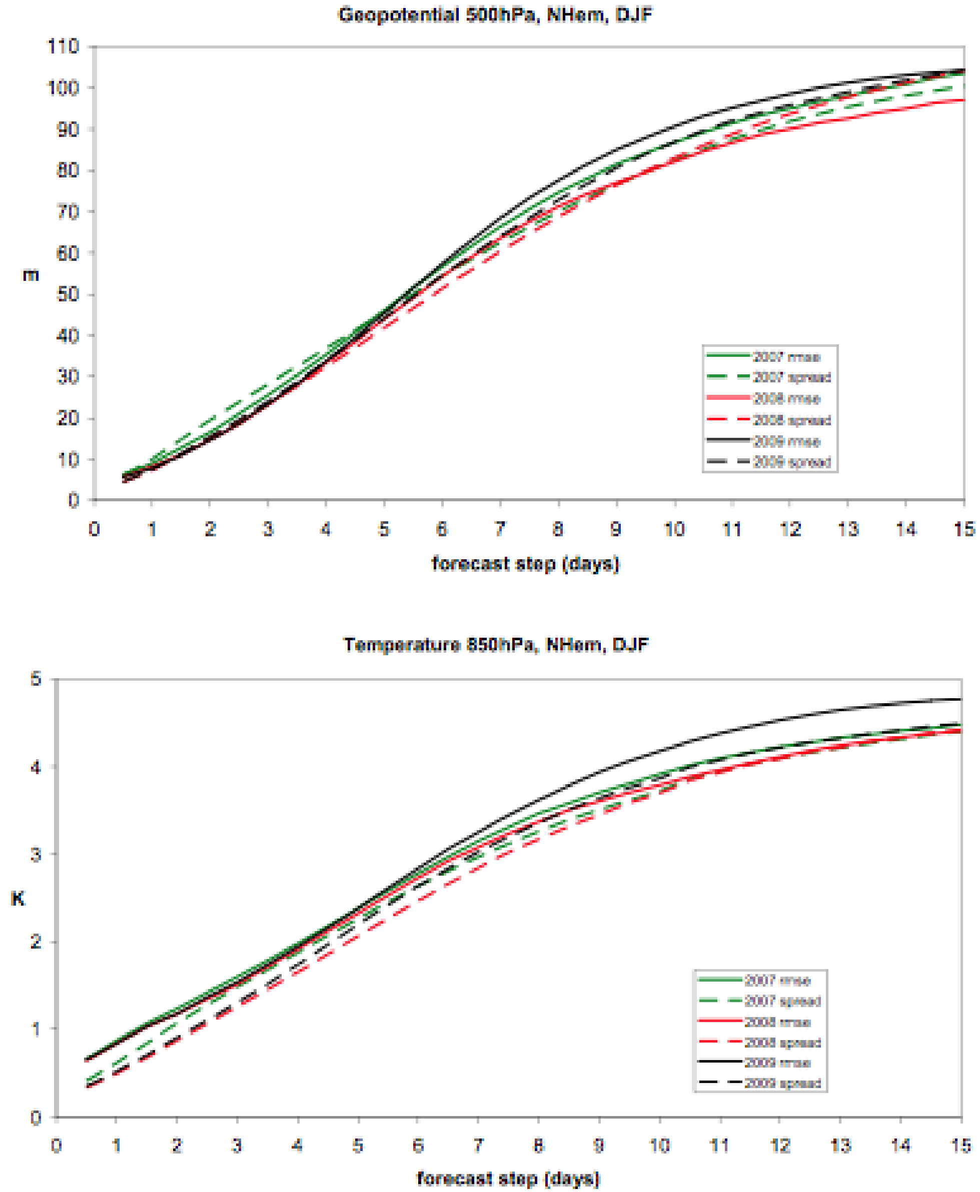


Figure 8: Ensemble spread (standard deviation, dashed lines) and root mean square error of ensemble-mean (solid lines) for 500 hPa geopotential (top) and 850 hPa temperature (bottom) for winter 2008-09 (black), 2007-08 (red) and 2006-07 (green) over the extra-tropical northern hemisphere.

Problèmes restants

- Cycle de l’eau (évaporation, condensation, influence sur le rayonnement absorbé ou émis par l’atmosphère)

- Échanges avec l’océan ou la surface continentale (chaleur, eau, quantité de mouvement, …)

- …

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.

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).

Assimilation is a problem in bayesian estimation.

Determine the conditional probability distribution for the state of the system, knowing everything we know (unambiguously defined if a prior probability distribution is defined; see Tarantola, 2005).

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, $p \approx$ 1-3.10^7 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

Coût des différentes composantes de la chaîne de prévision opérationnelle du CEPMMT (mars 2010, J.-N. Thépaut) :

Analysis	Daily	Weekly
DA	44394	311325
DCDA	78306	536834
Total	122700	**848159**

Forcast	Daily	Weekly
DA	49673	347546
DCDA	4517	30788
Total	54190	**378334**

EPS	Daily	Weekly
	193028	**1351576**

Monthly	Daily	Weekly
	N/A	46129

Hindcast	Daily	Weekly
	N/A	256724

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

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

$x = \xi \iff \zeta_1 = z_1 - \xi$ 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

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

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)

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.

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

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.

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

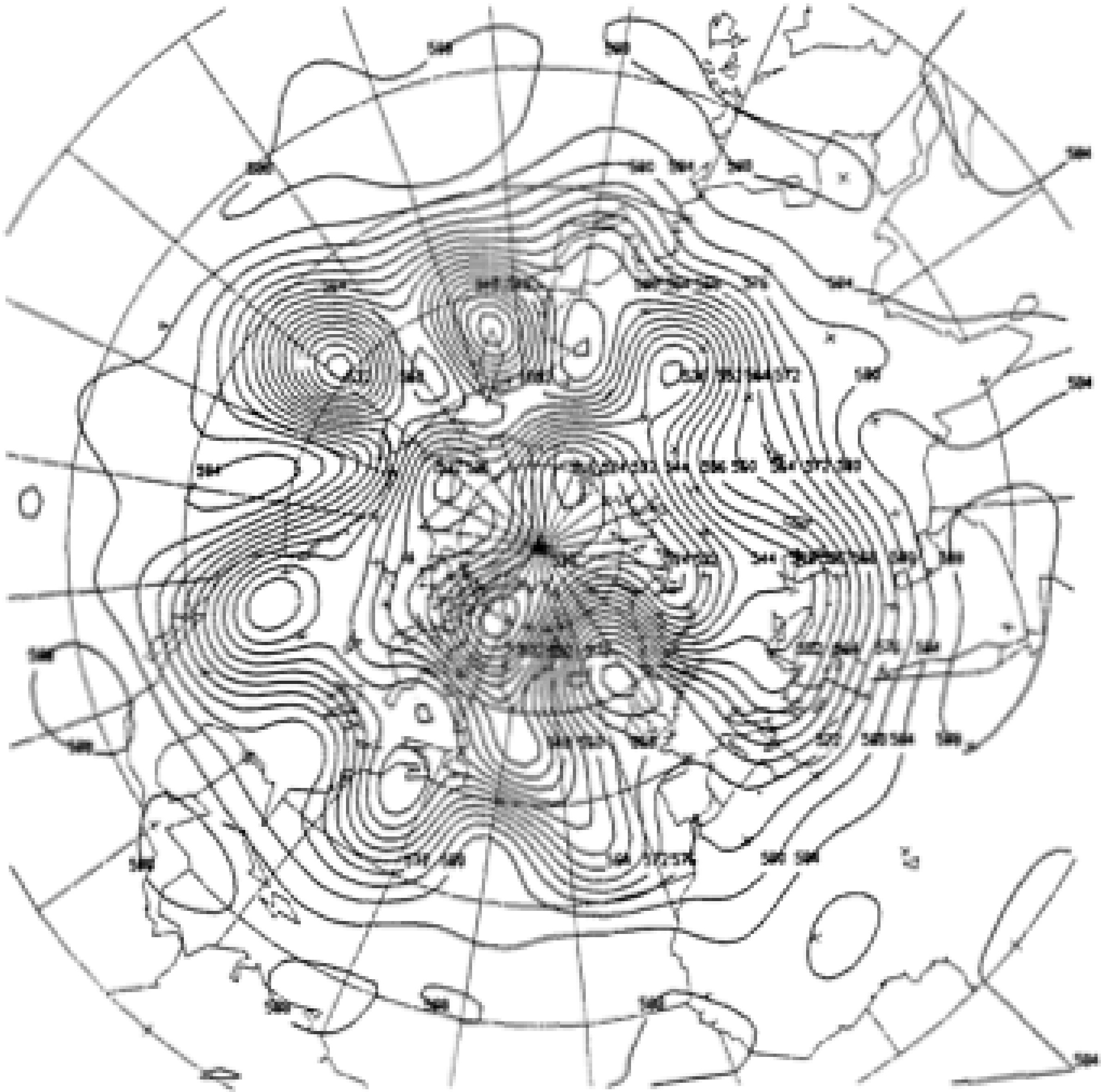


Figure 2. 500 mb height field produced by the operational analysis procedure of Direction de la Météorologie for 00 GMT, 26 April 1984. Units: dam, contour interval: 4 dam. The field has been truncated to the truncation of the model used for the experiments described in the article.

Courtier and Talagrand, *QJRMS*, 1987

(a)

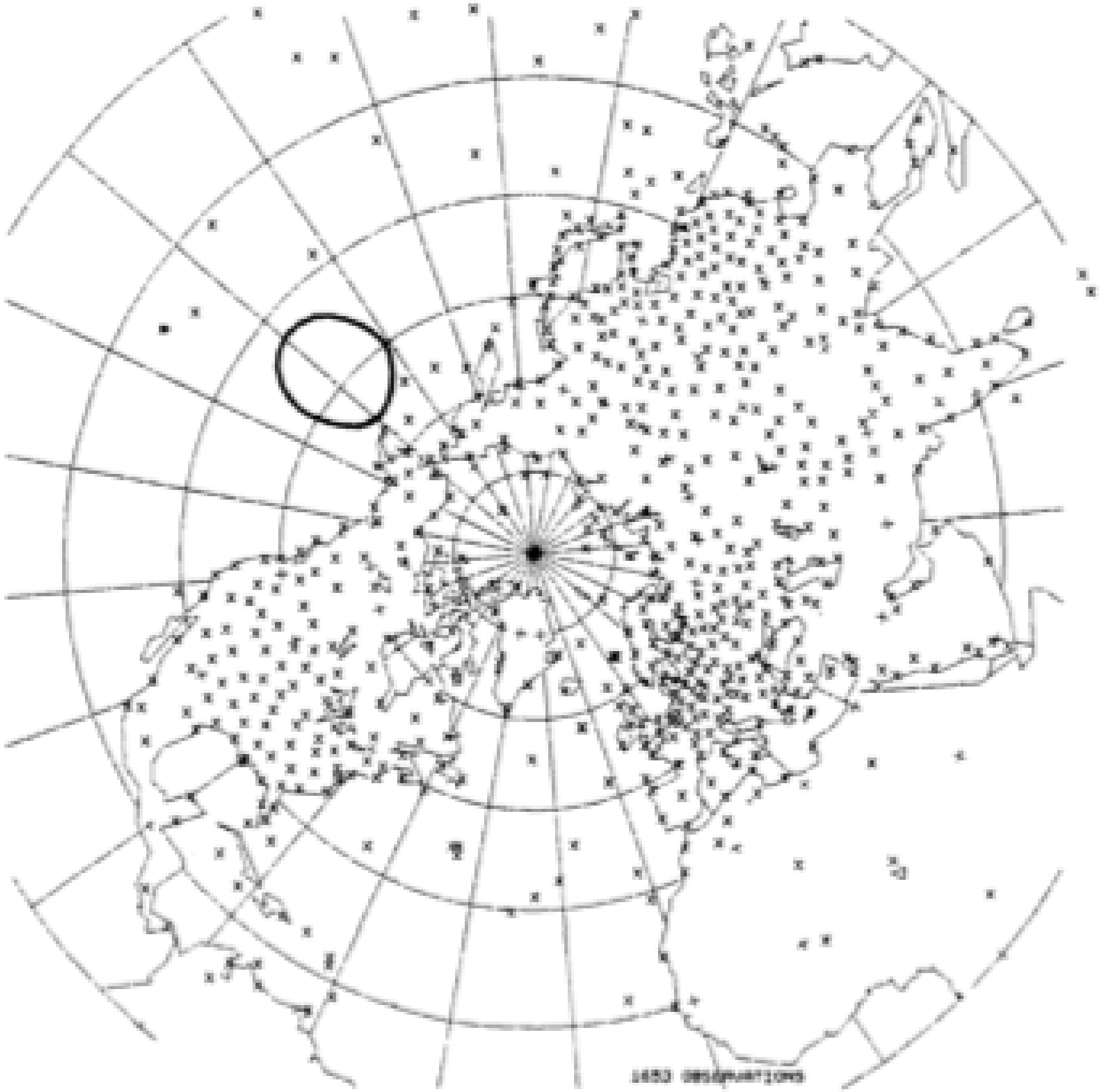


Figure 1. Geographical distribution of the observations used for the assimilation experiments. (a): geopotential observations; (b): wind observations. At most of the points plotted, several observations were made at successive synoptic hours. On each of the two charts, the heavy line delineates the Aleutian depression (see Figure 2).

Courtier and Talagrand, *QJRMS*, 1987

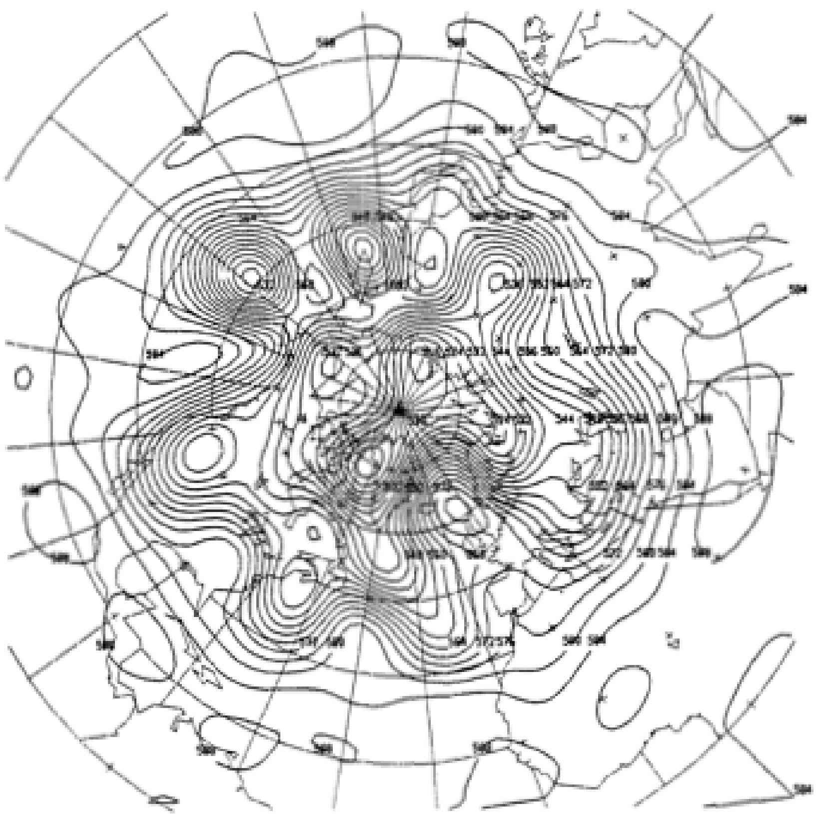


Figure 2. 500 mb height field produced by the operational analysis procedure of Direction de la Météorologie for 00 GMT, 26 April 1984. Units: dam, contour interval: 4 dam. The field has been truncated to the truncation of the model used for the experiments described in the article.

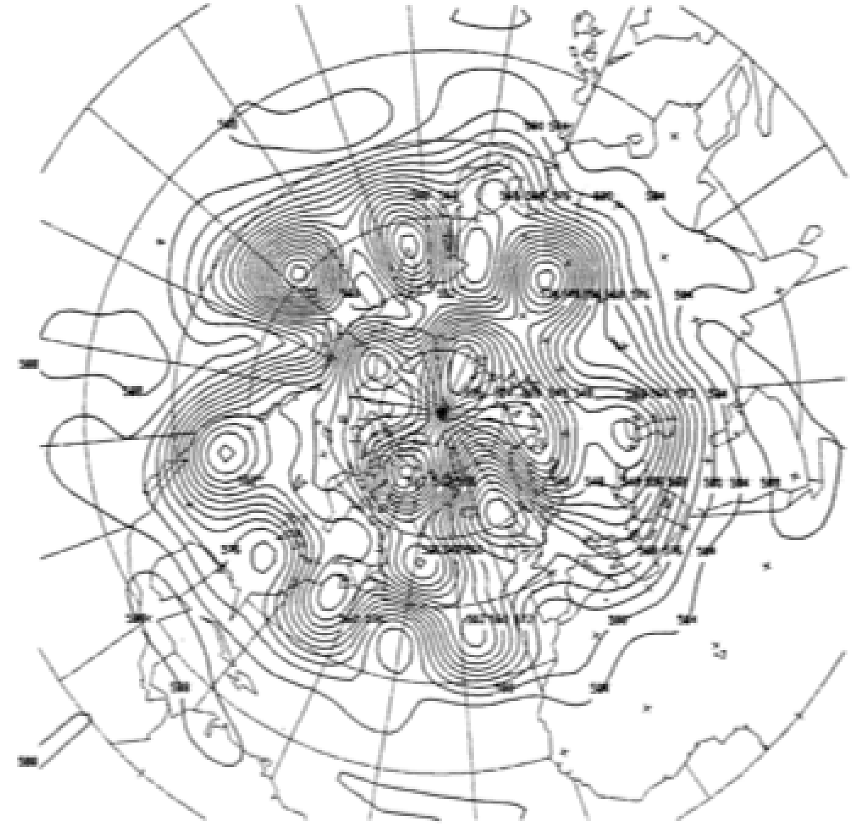


Figure 3. 500 mb height field produced for 00 GMT, 26 April 1984, by the variational analysis minimizing the distance function defined by Eqs. (1)–(2) over a 24-hour period. Units: dam; contour interval: 4 dam.

500-hPa geopotential field as determined by : (left) operational assimilation system of French Weather Service (3D, primitive equation) and (right) experimental variational system (2D, vorticity equation)

Courtier and Talagrand, *QJRMS*, 1987

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

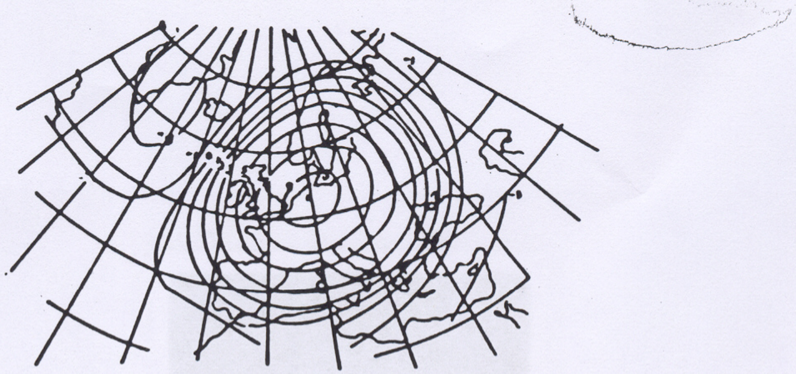
Random function $\Phi(\xi)$ (field of pressure, temperature, abundance of given chemical compound, … ; ξ is now spatial and/or temporal coordinate)

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

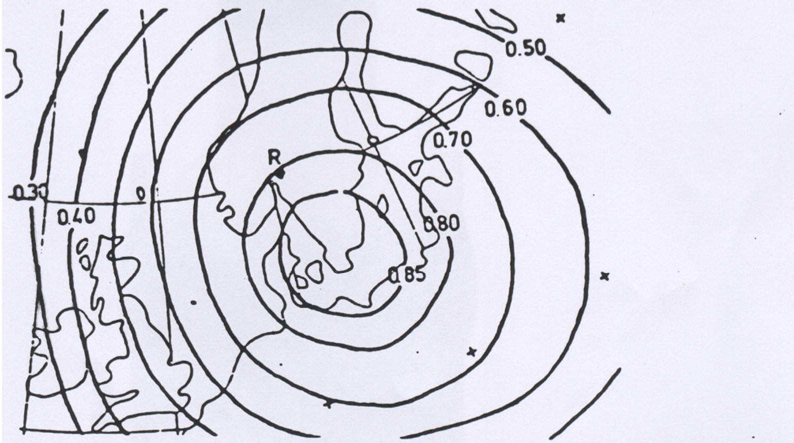
$$(\xi_{1,}\ \xi_2) \rightarrow\ C_{\varphi}(\xi_{1,}\ \xi_2) \equiv E[\Phi'(\xi_1)\ \Phi'(\xi_2)]$$

- Correlation function

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



.: 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

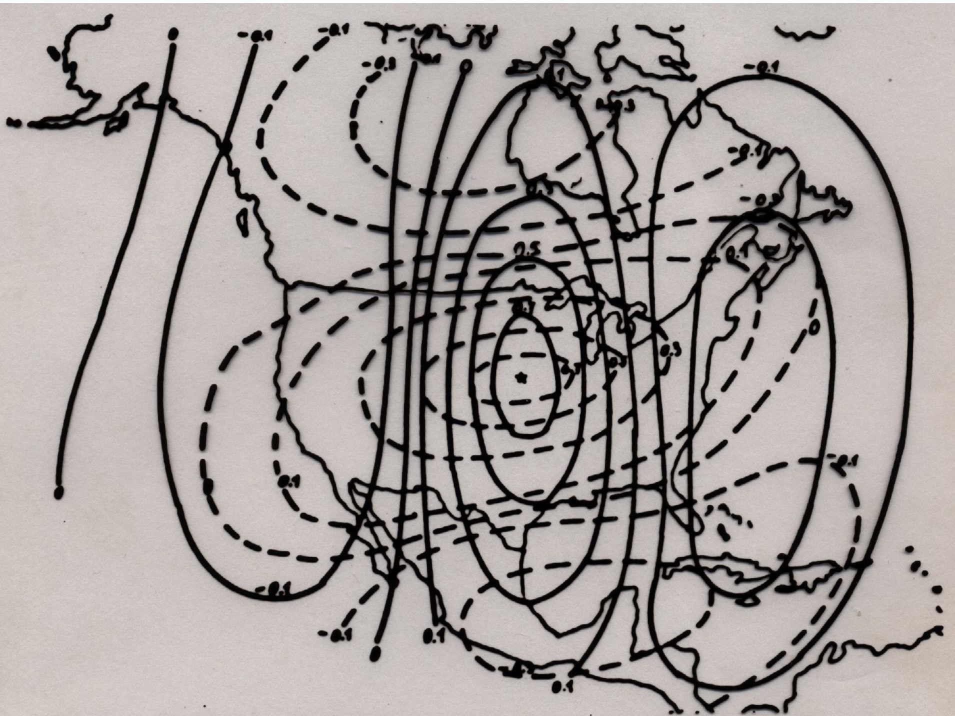
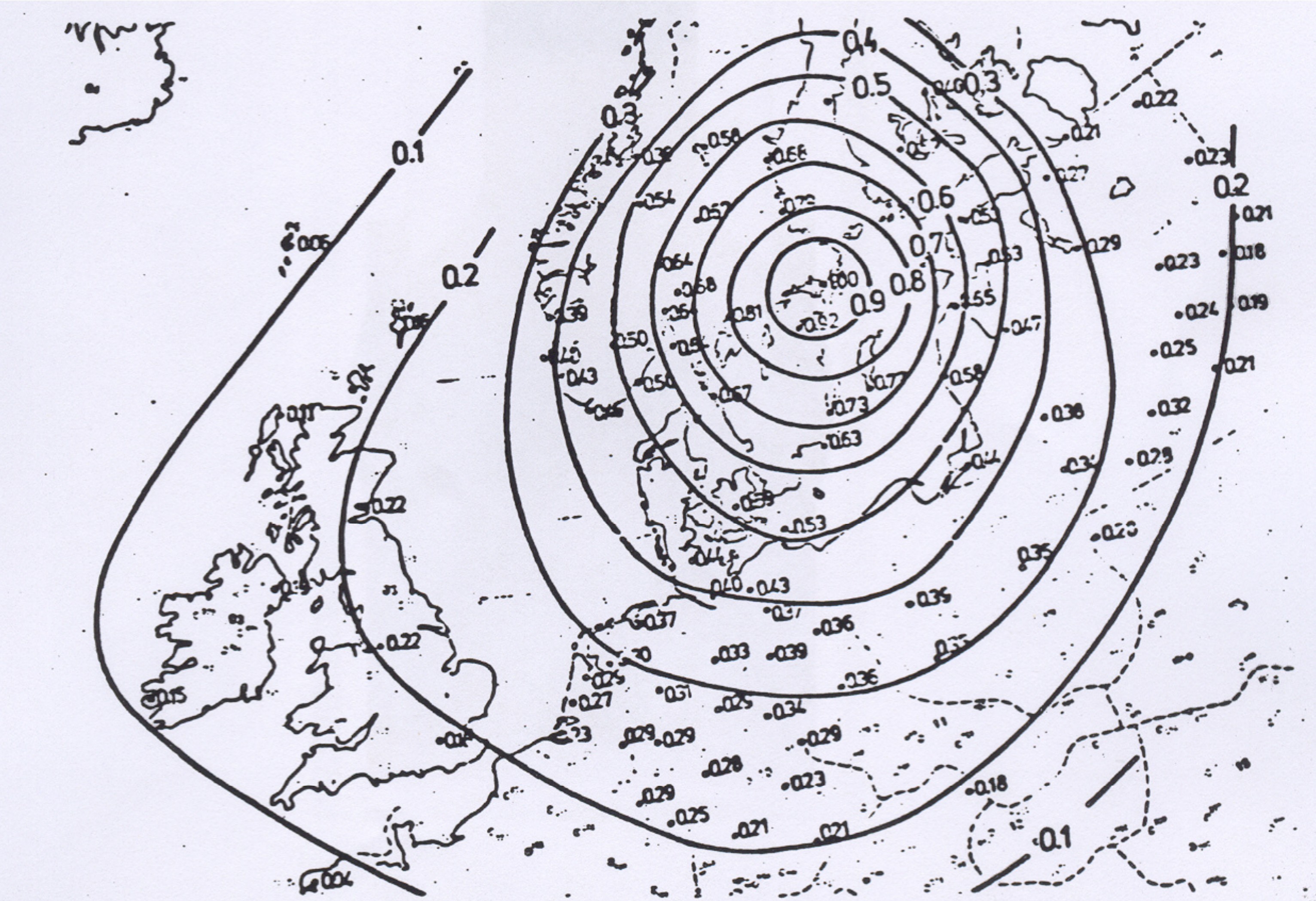


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson



<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

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

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

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

i.e.,

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

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

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

Minimized quadratic estimation error

$$E[(x-x^a)^2] = E(x'^2) - E(x'\mathbf{y}'^{\mathrm{T}})\,[E(\mathbf{y}'\mathbf{y}'^{\mathrm{T}})]^{-1}\,E(\mathbf{y}'x')$$

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

Optimal Interpolation (continued 2)

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

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

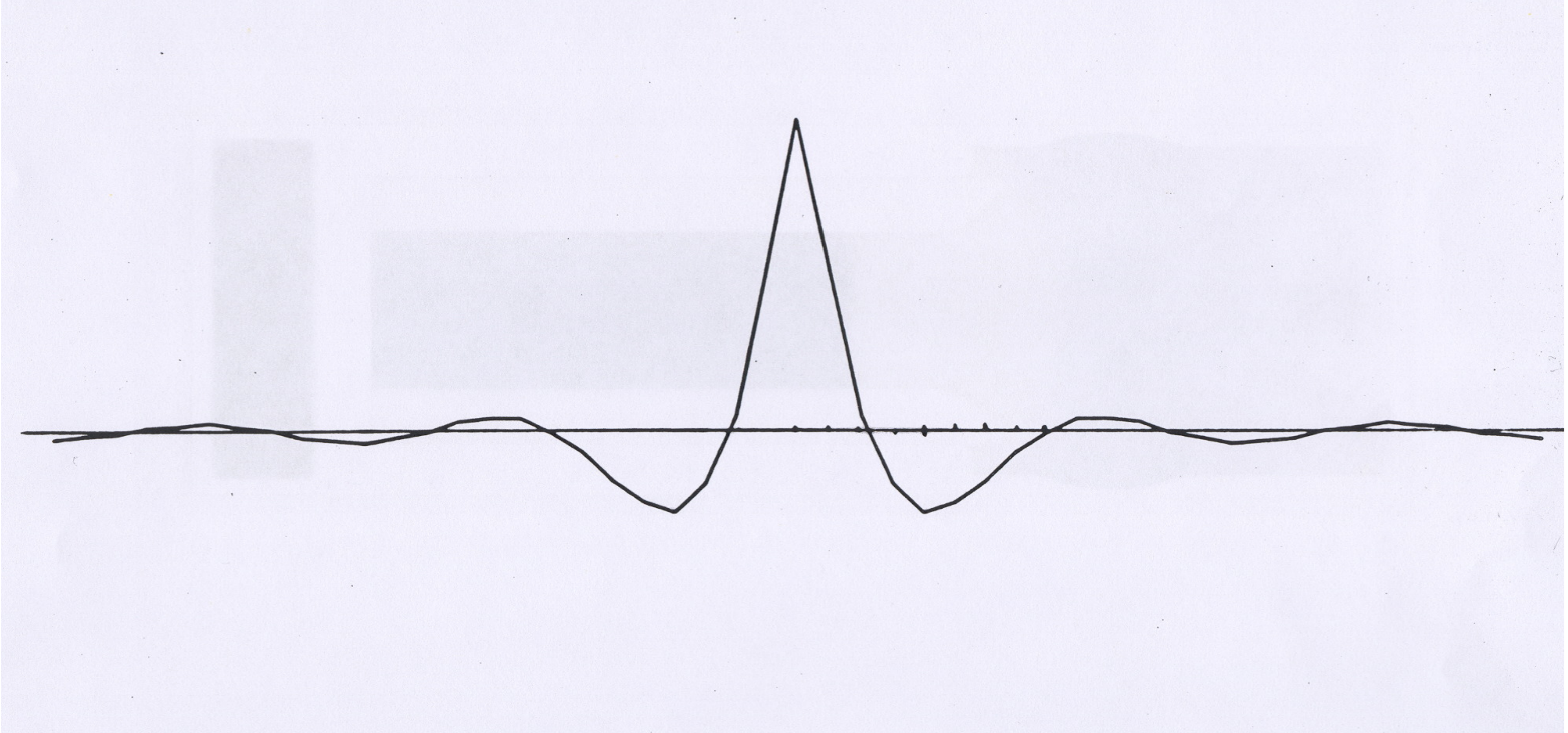
$$E(y_j'y_k') = E[\Phi'(\xi_j) + \varepsilon_j'][\Phi'(\xi_k) + \varepsilon_k']$$

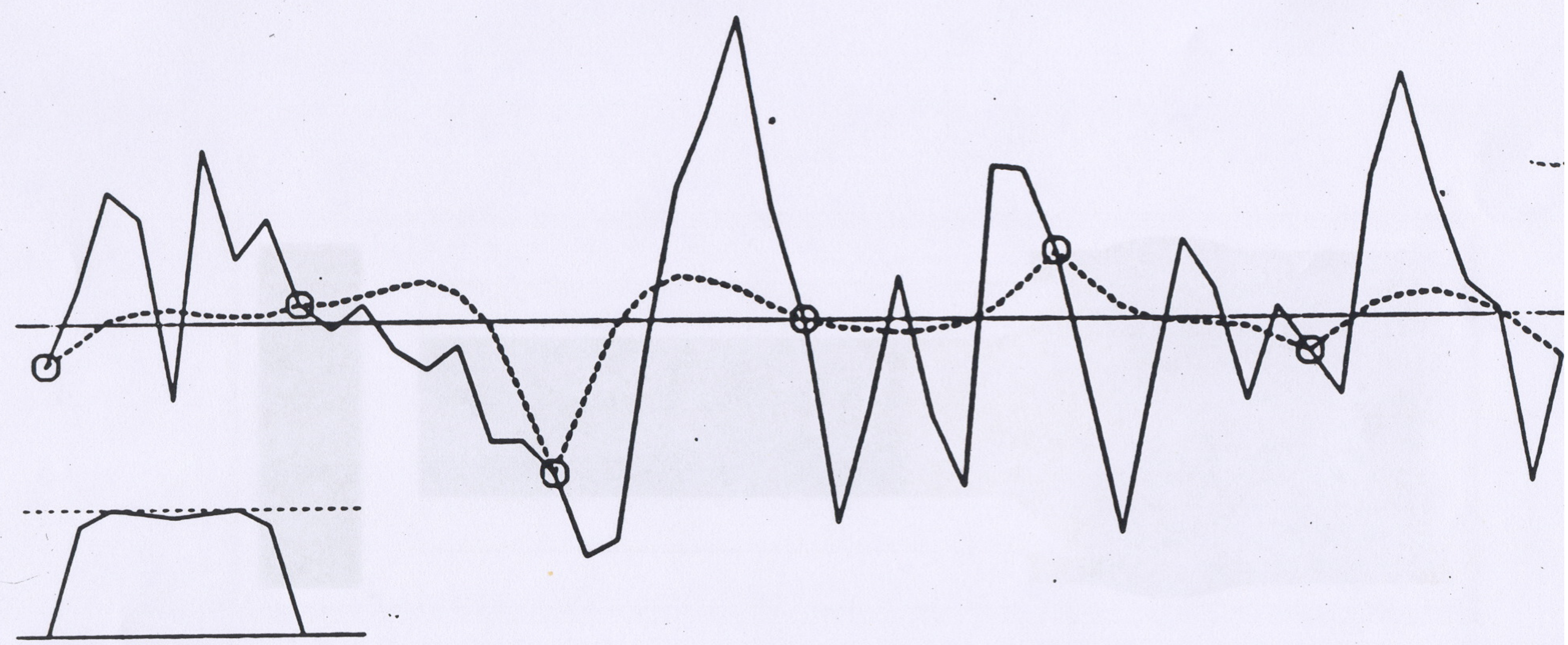
If observation errors ε_j are mutually uncorrelated, have common variance s, and are uncorrelated with field Φ, then

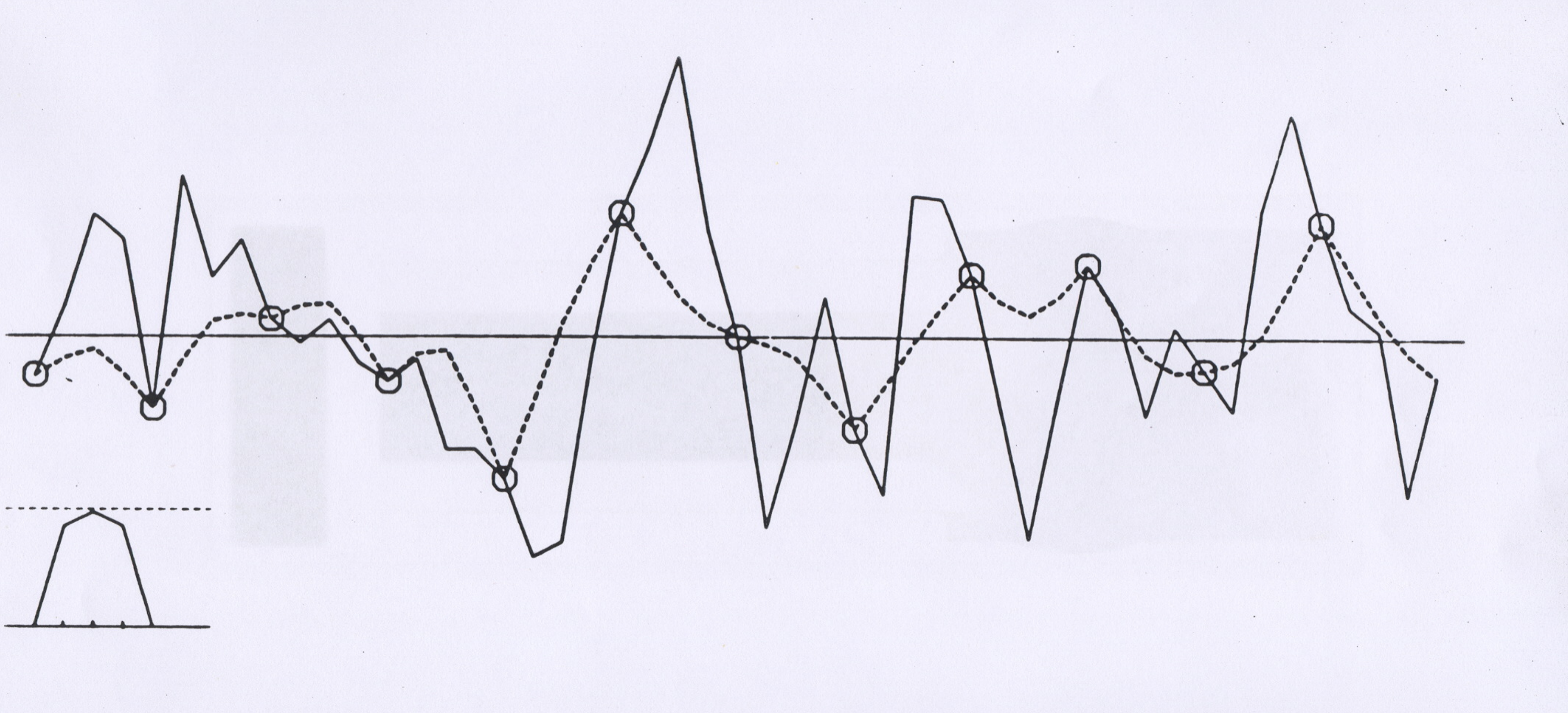
$$E(y_j'y_k') = C_\Phi(\xi_{j,}\ \xi_k) + s\delta_{jk}$$

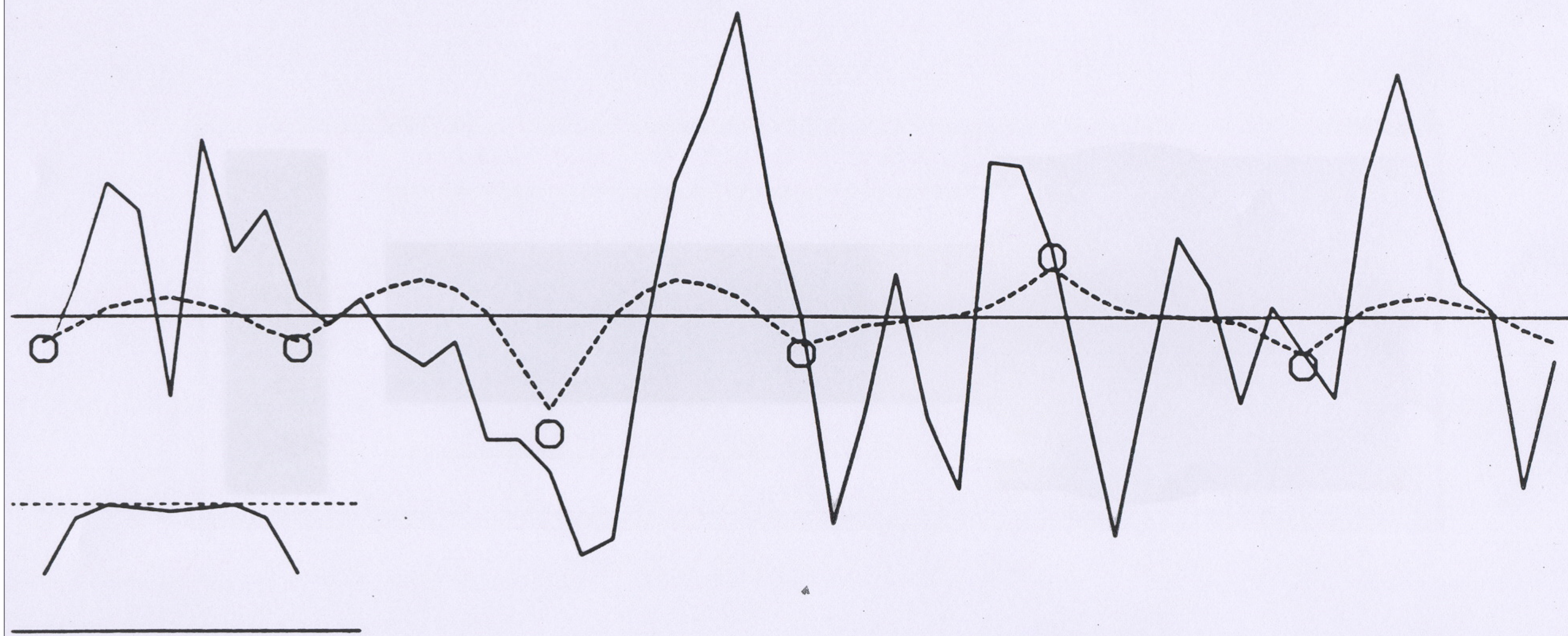
and

$$E(x'y_j') = C_\Phi(\xi_,\ \xi_j)$$









Optimal Interpolation (continued 3)

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

Vector

$$\boldsymbol{\mu} = (\mu_j) \equiv [E(\boldsymbol{y}'\boldsymbol{y}'^{\mathrm{T}})]^{-1}\,[\boldsymbol{y} - E(\boldsymbol{y})]$$

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

$$E[\Phi'(\xi)\, y_j']. \; E[\Phi'(\xi)\, y_j']\; [\, = C_{\Phi}(\xi, \xi_j)\,]$$

considered as a function of estimation position ξ, is the *representer* associated with observation y_j.

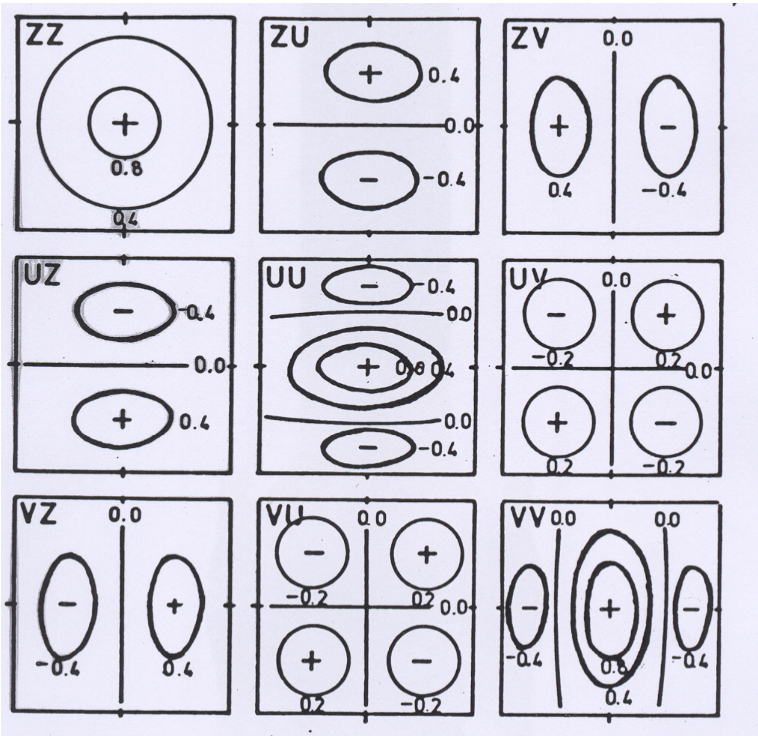
Optimal Interpolation (continued 4)

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