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

Année 2013-2014

Course *Introduction to data assimilation*

From numerical modelling to data assimilation

Olivier Talagrand

7 January 2014

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.

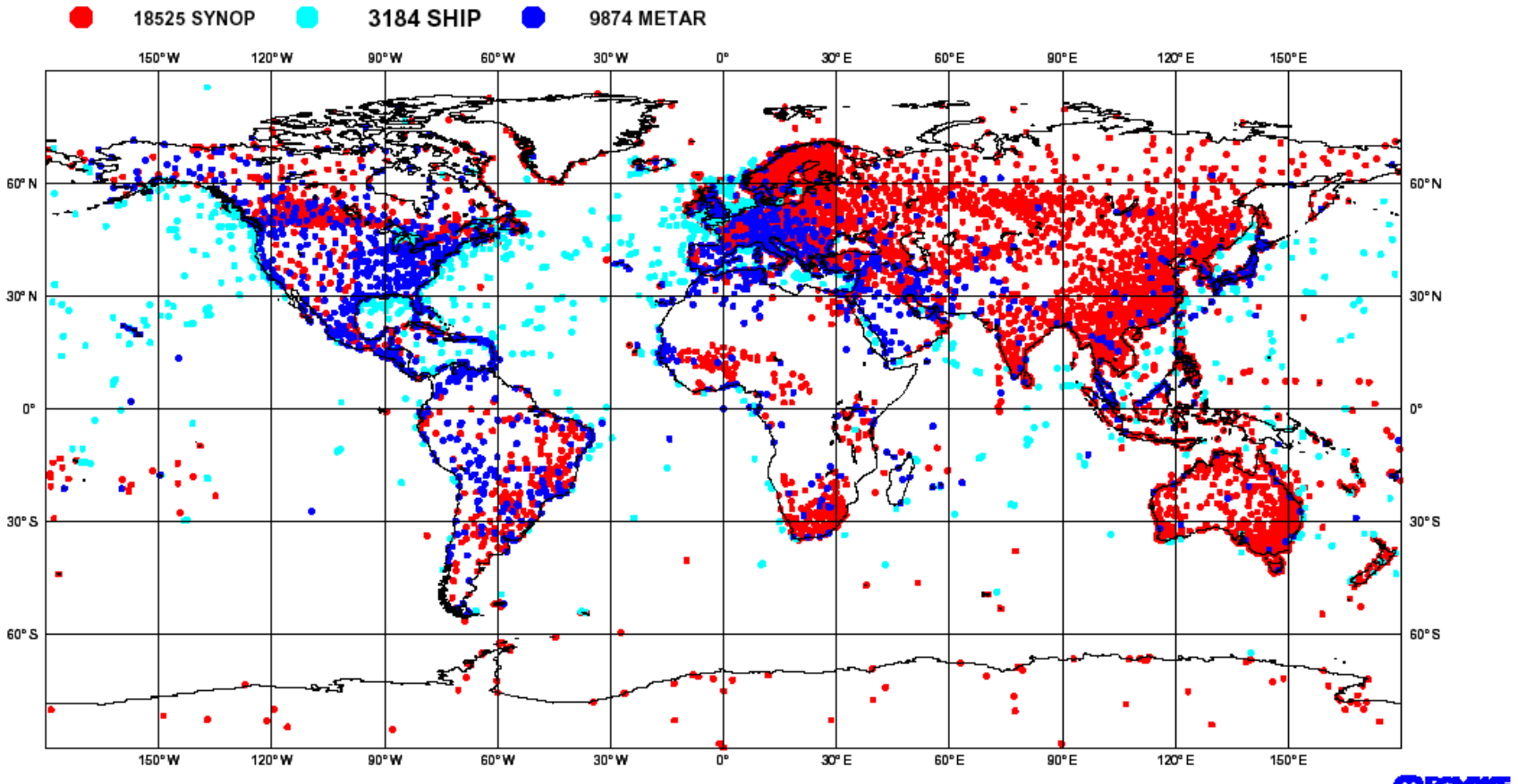
The contour interval is 5 hPa, with shading patters for MSLP values lower than 990 hPa.

Why have meteorologists such difficulties in predicting the weather with any certainty ? Why is it that showers and even storms seem to come by chance, so that many people think it is quite natural to pray for them, though they would consider it ridiculous to ask for an eclipse by prayer ? […] *a tenth of a degree more or less at any given point, and the cyclone will burst here and not there, and extend its ravages over districts that it would otherwise have spared. If they had been aware of this tenth of a degree, they could have known it beforehand, but the observations were neither sufficiently comprehensive nor sufficiently precise, and that is the reason why it all seems due to the intervention of chance.*

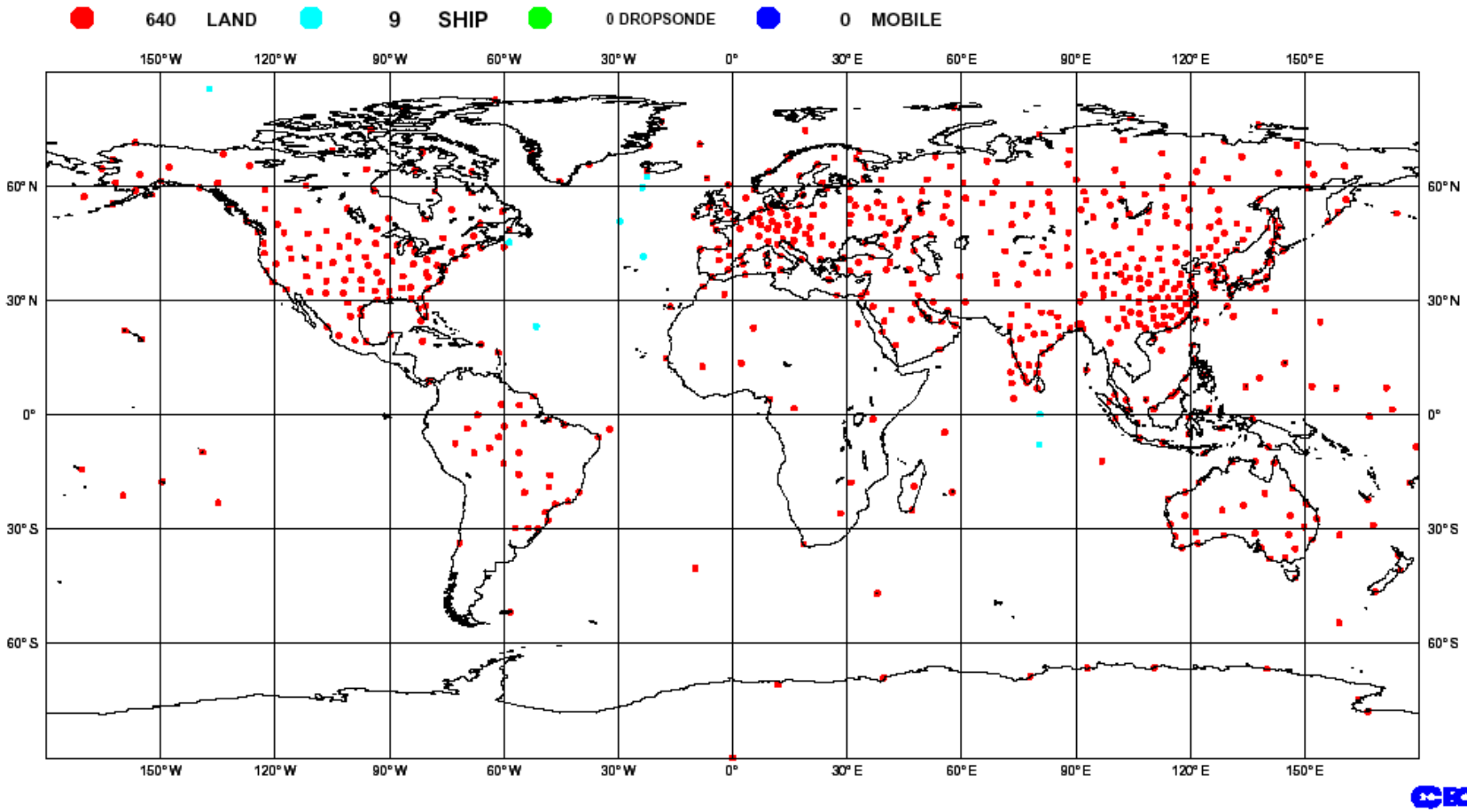
H. Poincaré, *Science et Méthode*, Paris, 1908

(translated Dover Publ., 1952)

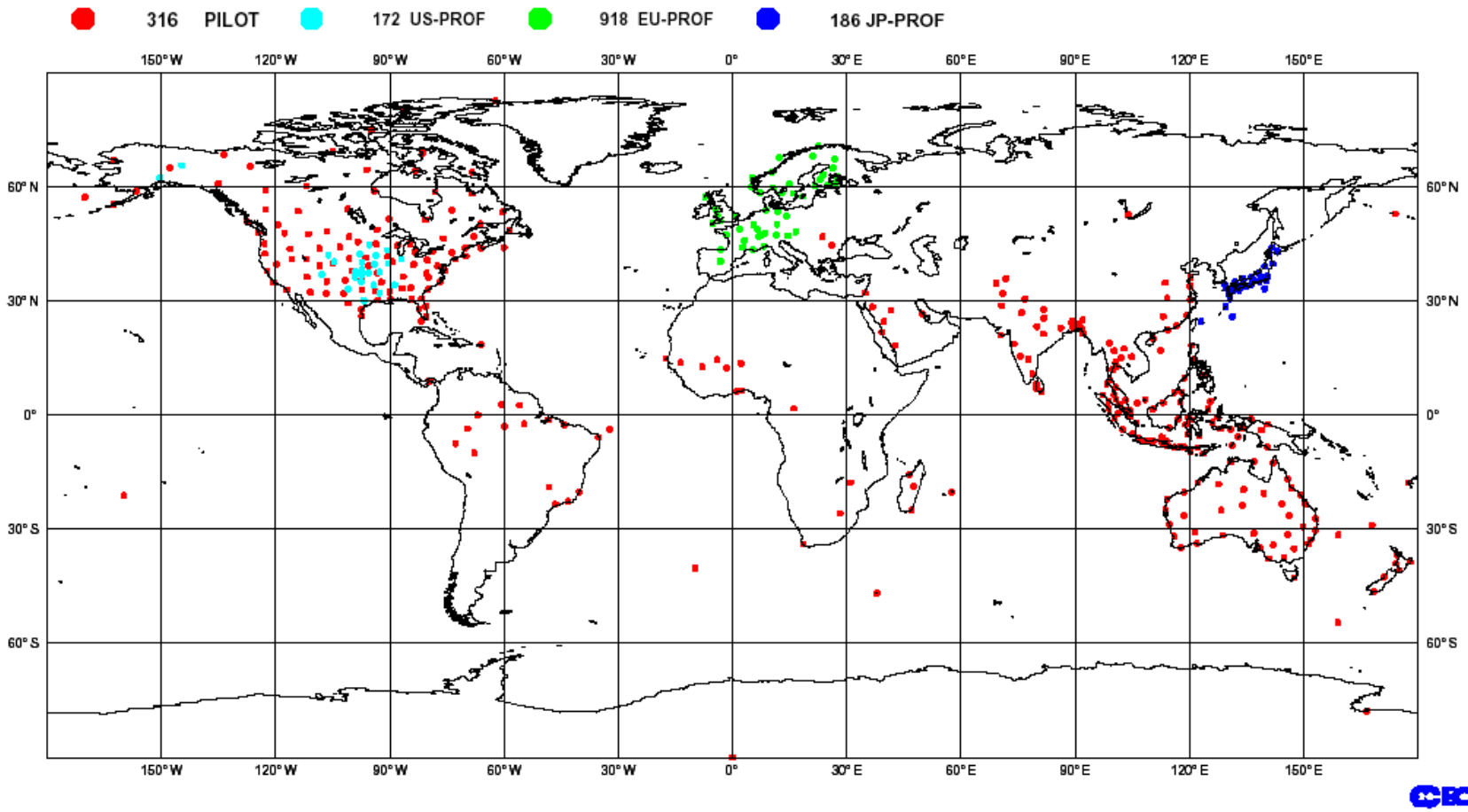
ECMWF Data Coverage (All obs DA) - Synop-Ship-Metar
13/Nov/2011; 00 UTC
Total number of obs = 31583



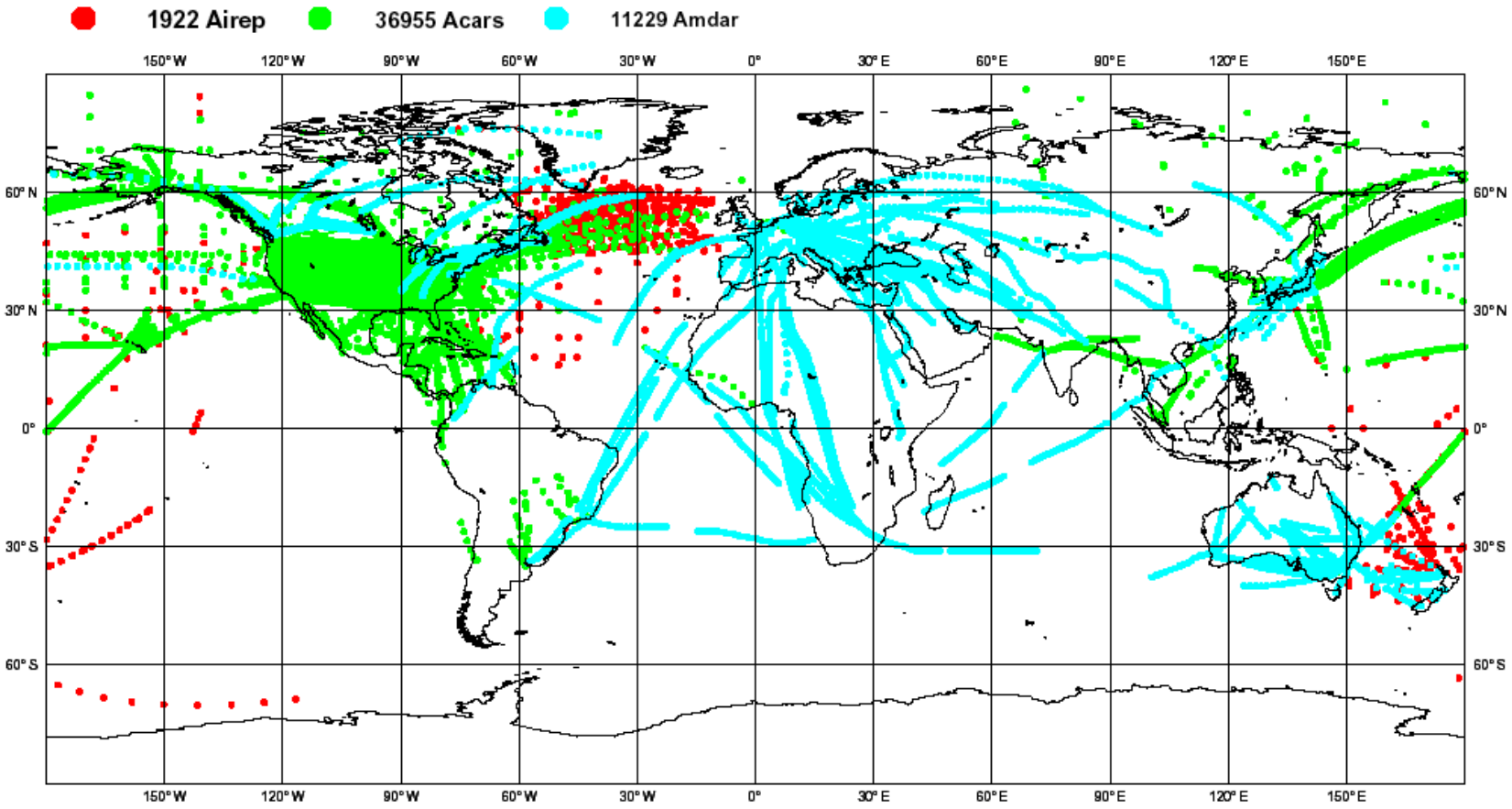
ECMWF Data Coverage (All obs DA) - Temp
13/Nov/2011; 00 UTC
Total number of obs = 649



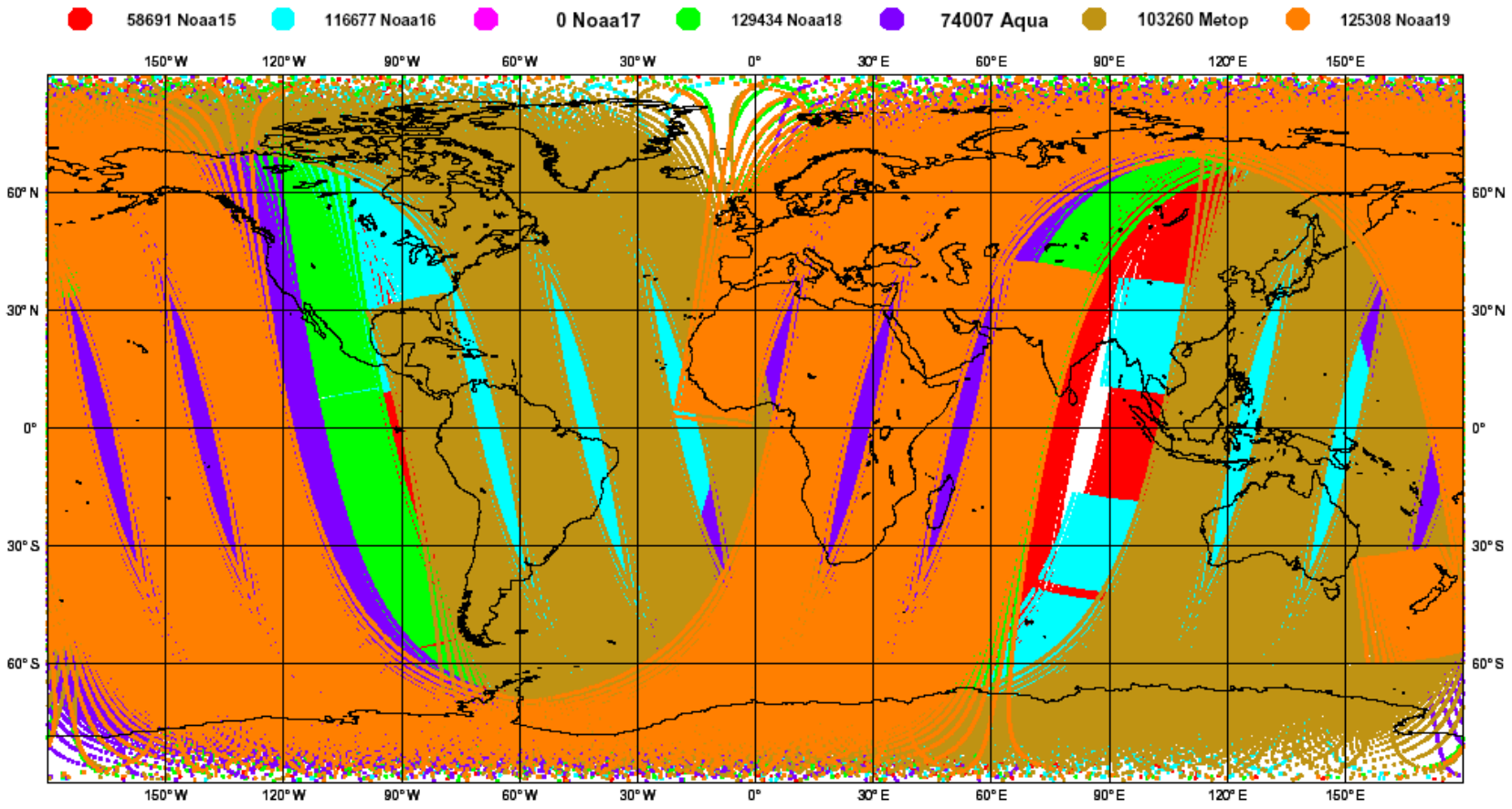
ECMWF Data Coverage (All obs DA) - Pilot-Profiler
13/Nov/2011; 00 UTC
Total number of obs = 1592



ECMWF Data Coverage (All obs DA) - Aircraft
13/Nov/2011; 00 UTC
Total number of obs = 50106

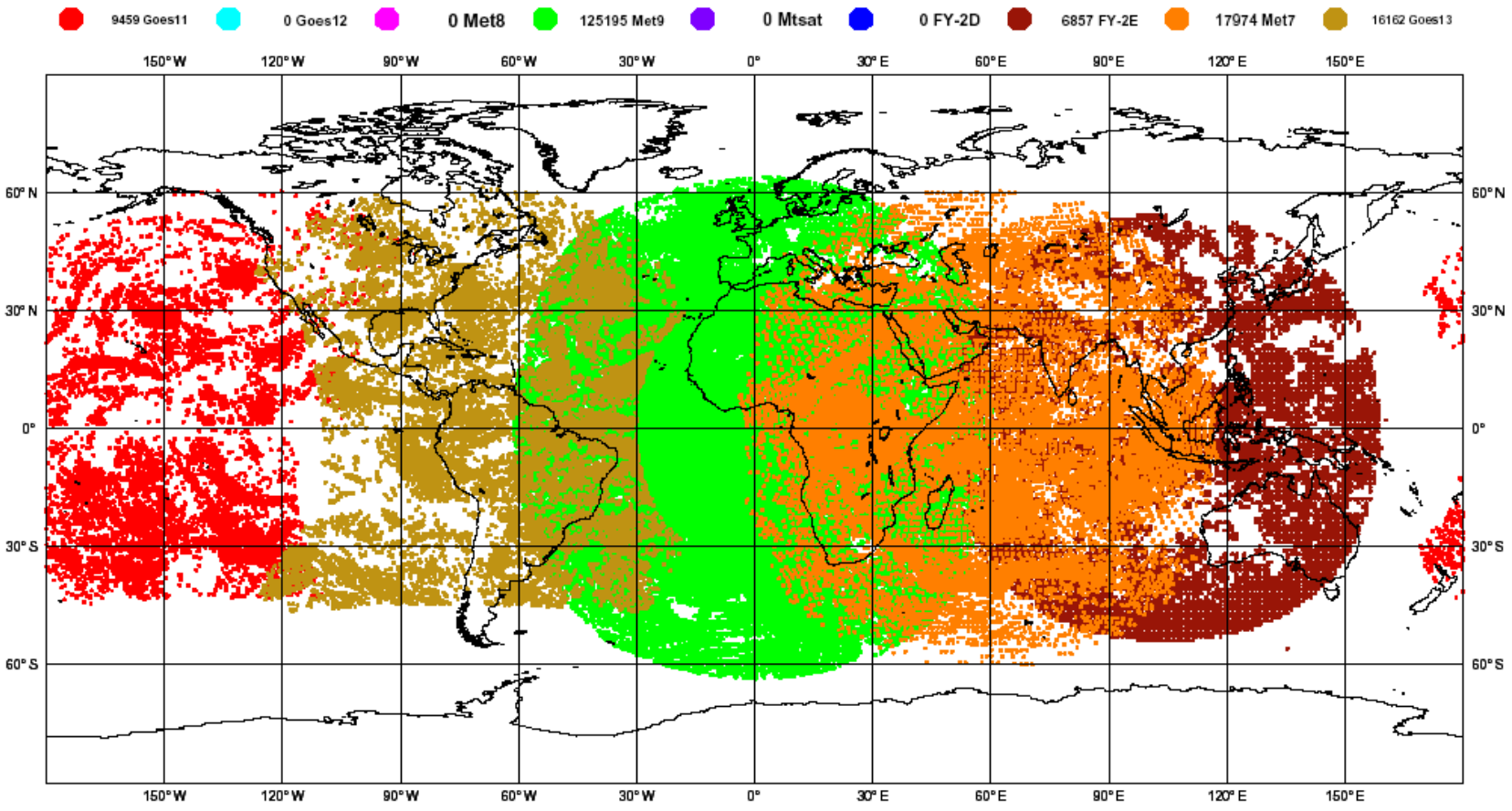


ECMWF Data Coverage (All obs DA) - AMSU-A
13/Nov/2011; 00 UTC
Total number of obs = 607377



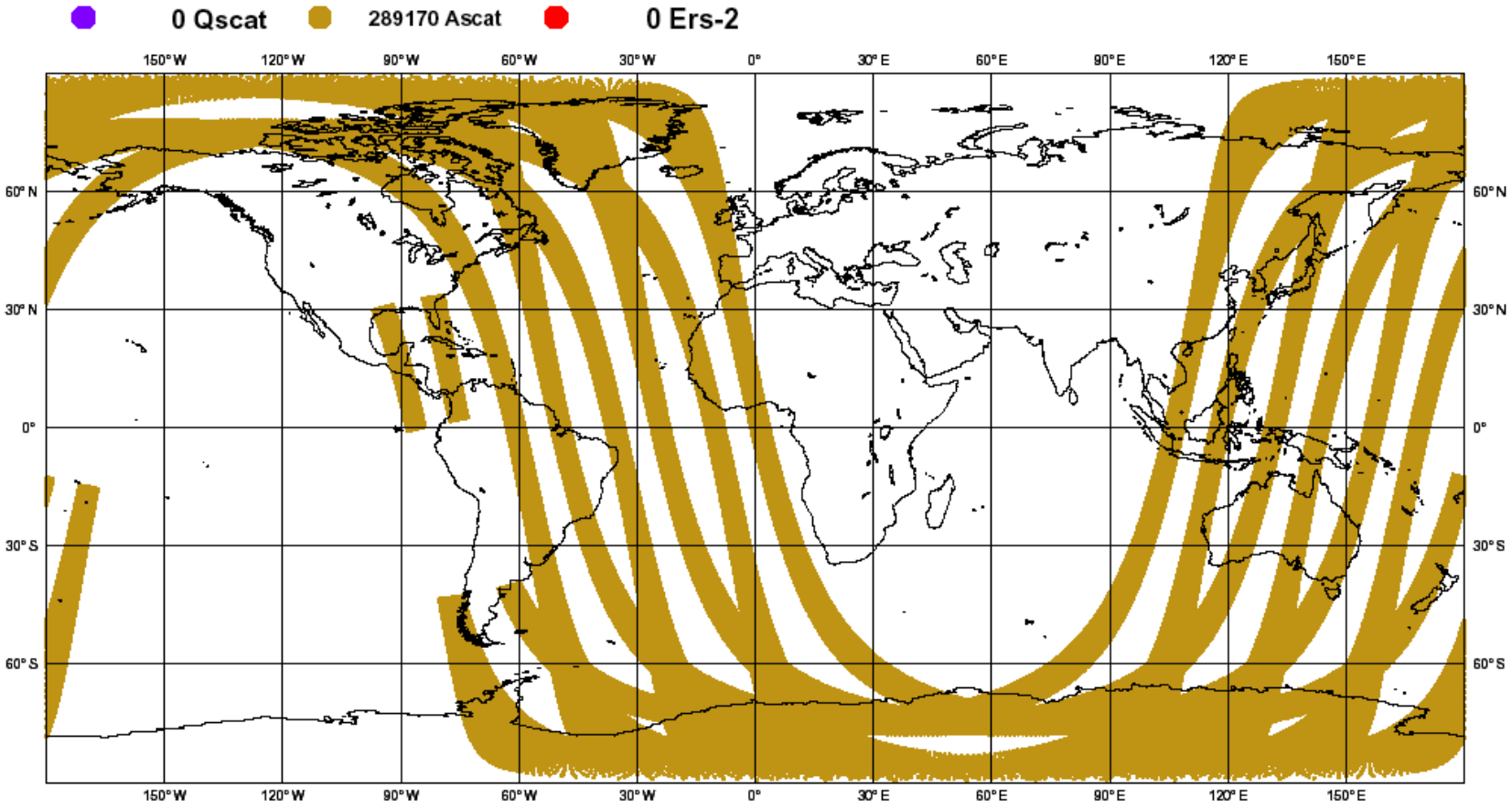
ECMWF

ECMWF Data Coverage (All obs DA) - AMV WV
13/Nov/2011; 00 UTC
Total number of obs = 175647



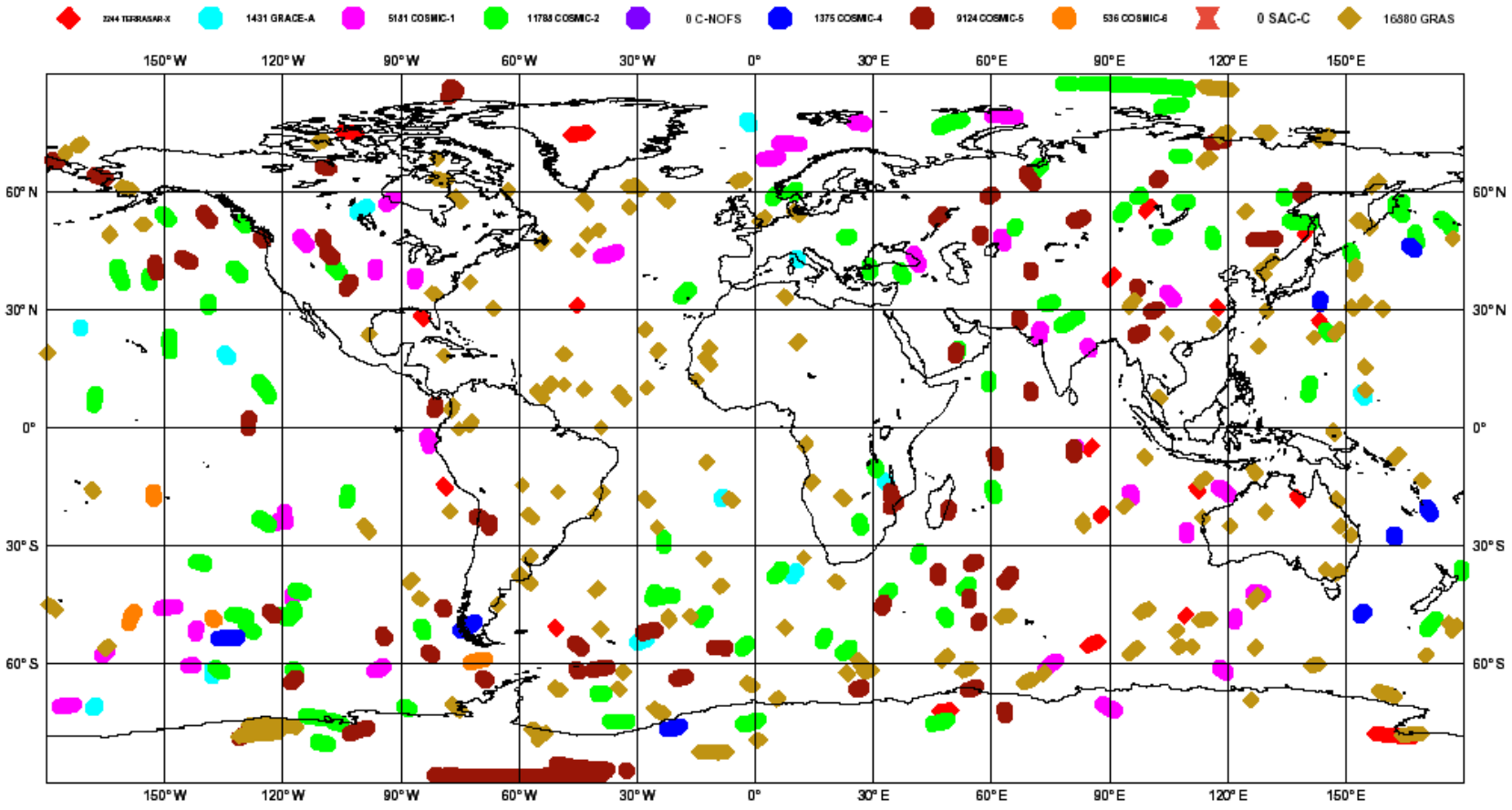
ECMWF

ECMWF Data Coverage (All obs DA) - SCAT
13/Nov/2011; 00 UTC
Total number of obs = 289170



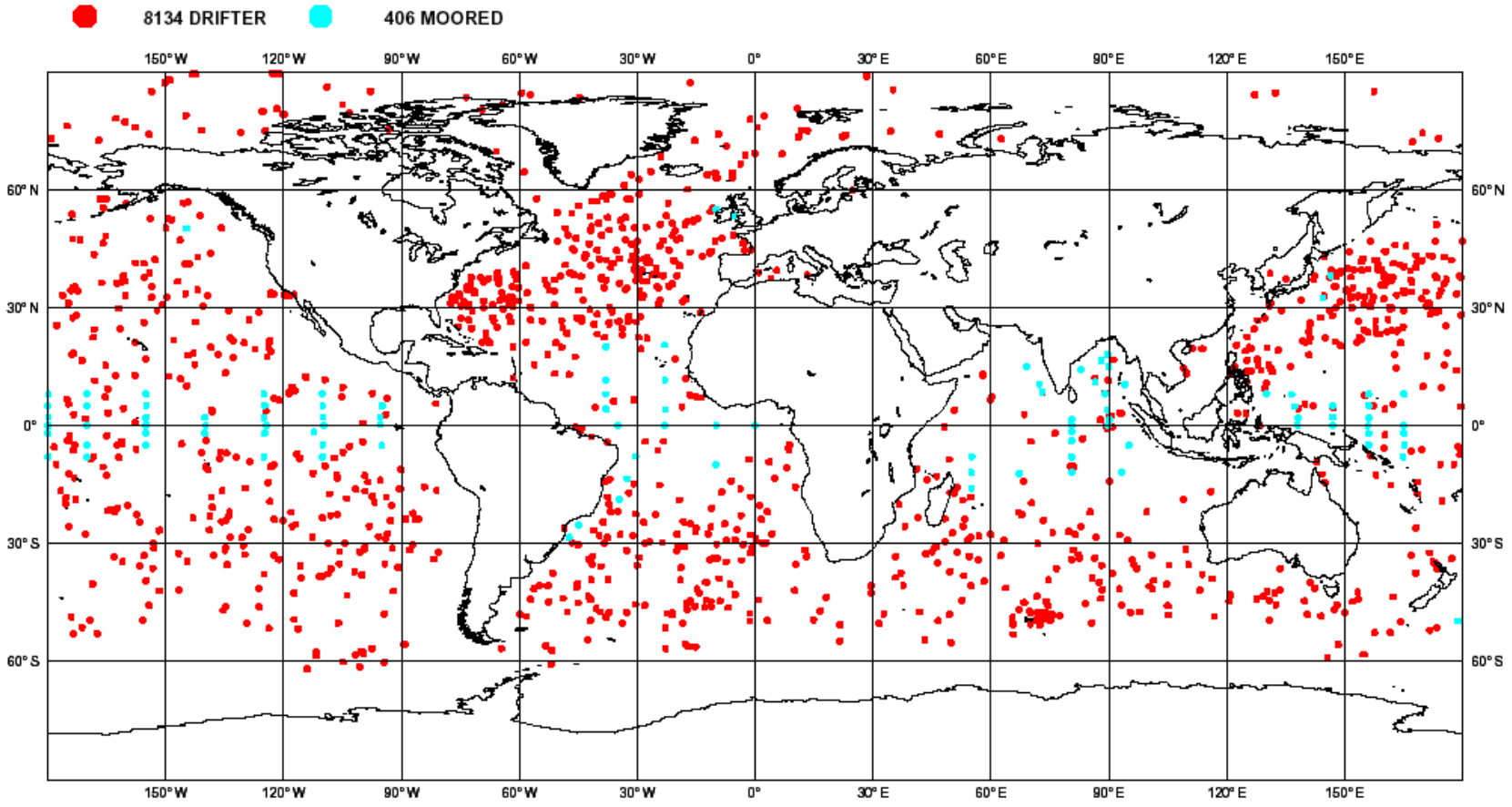
ECMWF

ECMWF Data Coverage (All obs DA) - GPSRO
13/Nov/2011; 00 UTC
Total number of obs = 48559

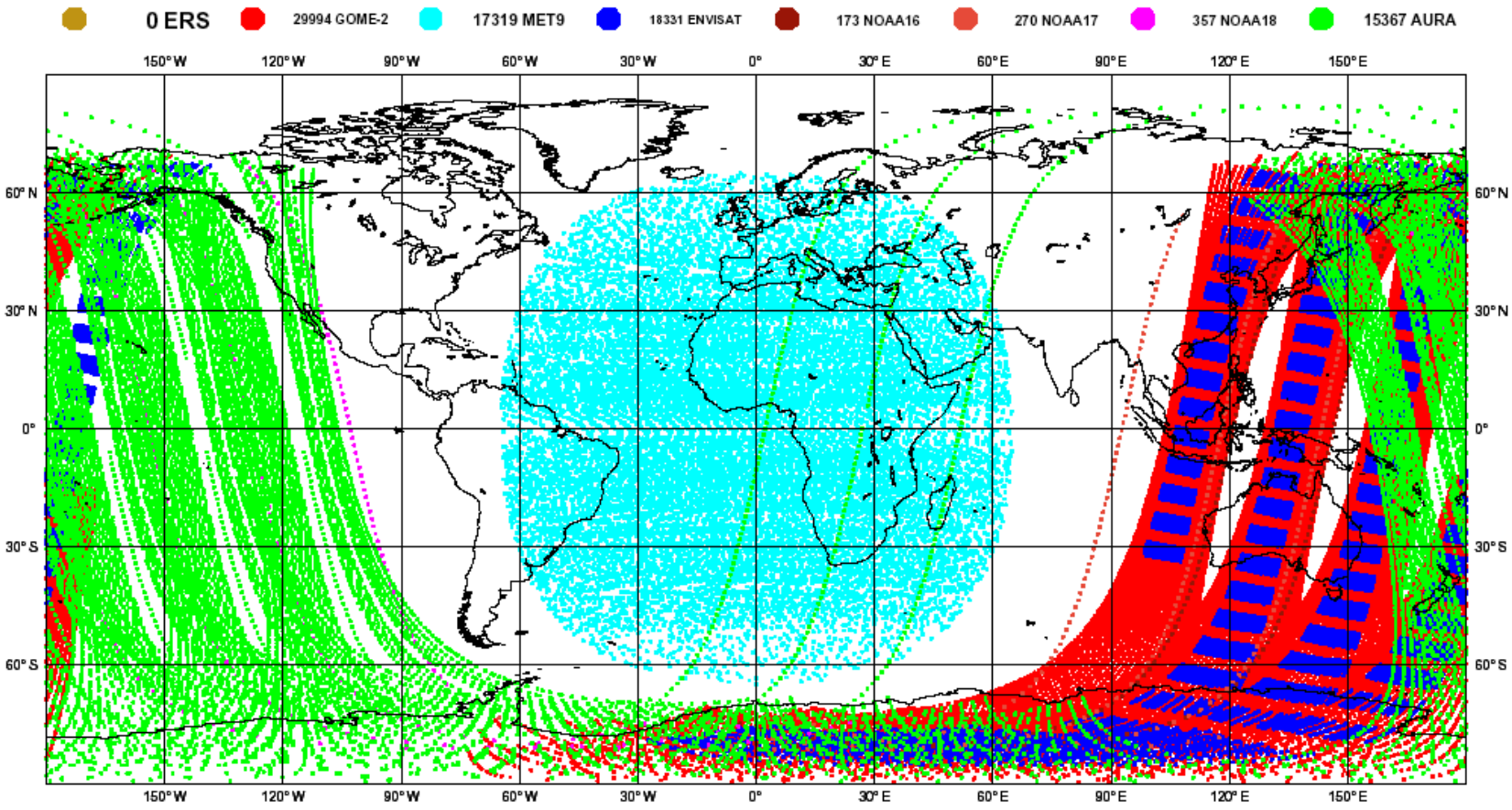


ECMWF

ECMWF Data Coverage (All obs DA) - Buoy
13/Nov/2011; 00 UTC
Total number of obs = 8540

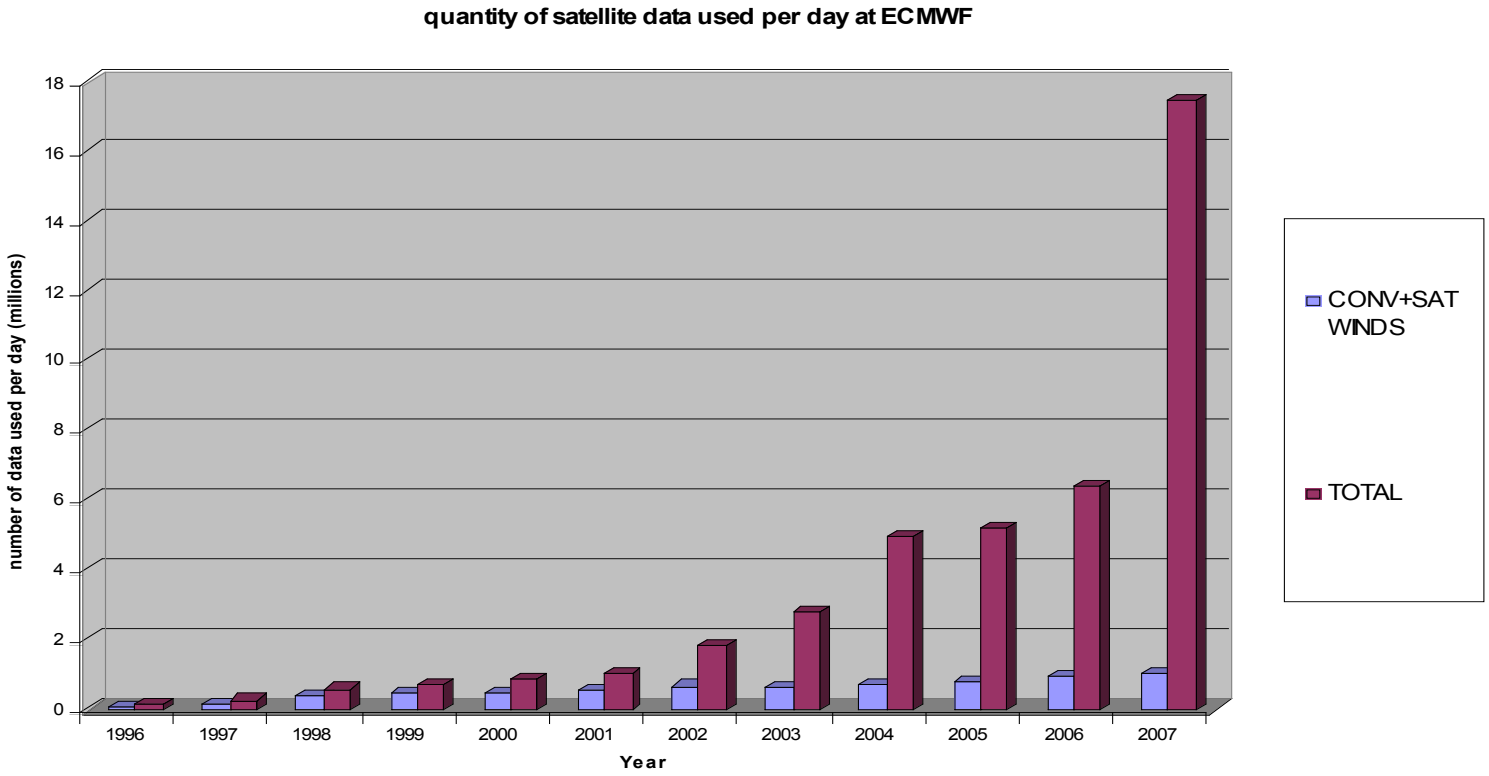


ECMWF Data Coverage (All obs DA) - OZONE
13/Nov/2011; 00 UTC
Total number of obs = 81811



ECMWF

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



Value as of early 2013 : around 25 millions per day

- *Synoptic* observations (ground observations, radiosonde observations), performed simultaneously, by international agreement, in all meteorological stations around the world (00:00, 06:00, 12:00, 18:00 TU), and are in practice concentrated over continents.

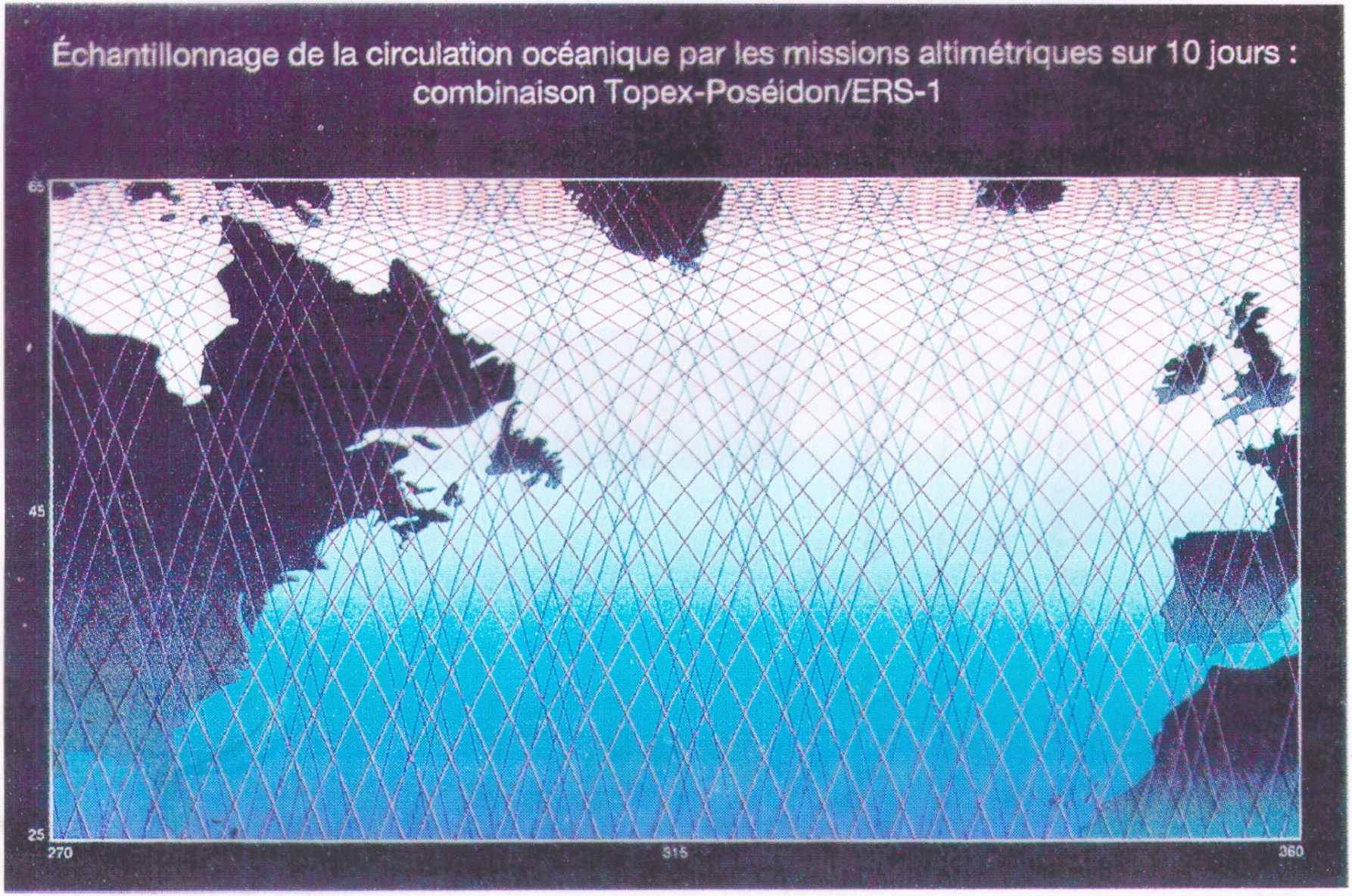
- *Asynoptic* observations (satellites, aircraft), performed more or less continuously in time.

- *Direct* observations (temperature, pressure, horizontal components of the wind, moisture), which are local and bear on the variables used for for describing the flow in numerical models.

- *Indirect* observations (radiometric observations, …), which bear on some more or less complex combination (most often, a one-dimensional spatial integral) of variables used for for describing the flow

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

H : *observation operator (*for instance, radiative transfer equation)



S. Louvel, Doctoral Dissertation, 1999

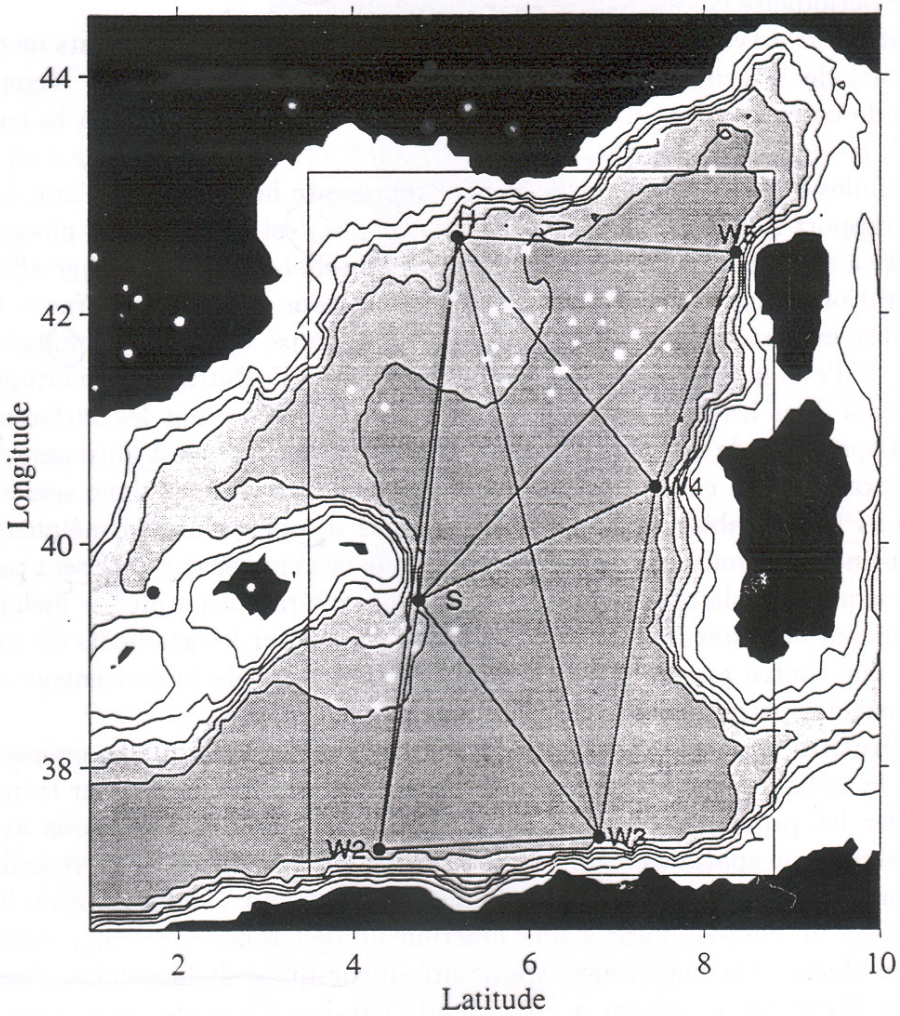


FIG. 1 – *Bassin méditerranéen occidental: réseau d'observation tomographique de l'expérience Thétis 2 et limites du domaine spatial utilisé pour les expériences numériques d'assimilation.*

E. Rémy, Doctoral Dissertation, 1999

Physical laws governing the flow

- Conservation of mass

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

- Conservation of energy

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

- Conservation of momentum

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

- Equation of state

  $f(p, \rho, e) = 0$ (for a perfect gas $p/\rho = rT$, $e = C_v T$)

- Conservation of mass of secondary components (water in the atmosphere, salt in the ocean, chemical species, …)

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

These physical laws must be expressed in practice in discretized (and necessarily imperfect) form, both in space and time ⇒ *numerical model*

Parlance of the trade :

- Adiabatic and inviscid, and therefore thermodynamically reversible, processes (everything except Q, $\underline{F}$ and S) make up *'dynamics'*

- Processes described by terms Q, $\underline{F}$ and S make up *'physics'*

All presently existing numerical models are built on simplified forms of the general physical laws. Global numerical models, used either for large-scale meteorological prediction or for climate simulation, are at present built on the so-called *primitive equations*. Those equations rely on several approximations, the most important of which being the *hydrostatic approximation*, which expresses balance, in the vertical direction, of the gravity and pressure gradient forces. This forbids explicit description of thermal convection, which must be parameterized in some appropriate way.

More and more *limited-area models* have been progressively developed. They require appropriate definition of lateral boundary conditions (not a simple problem). Most of them are non-hydrostatic, and therefore allow description of convection.

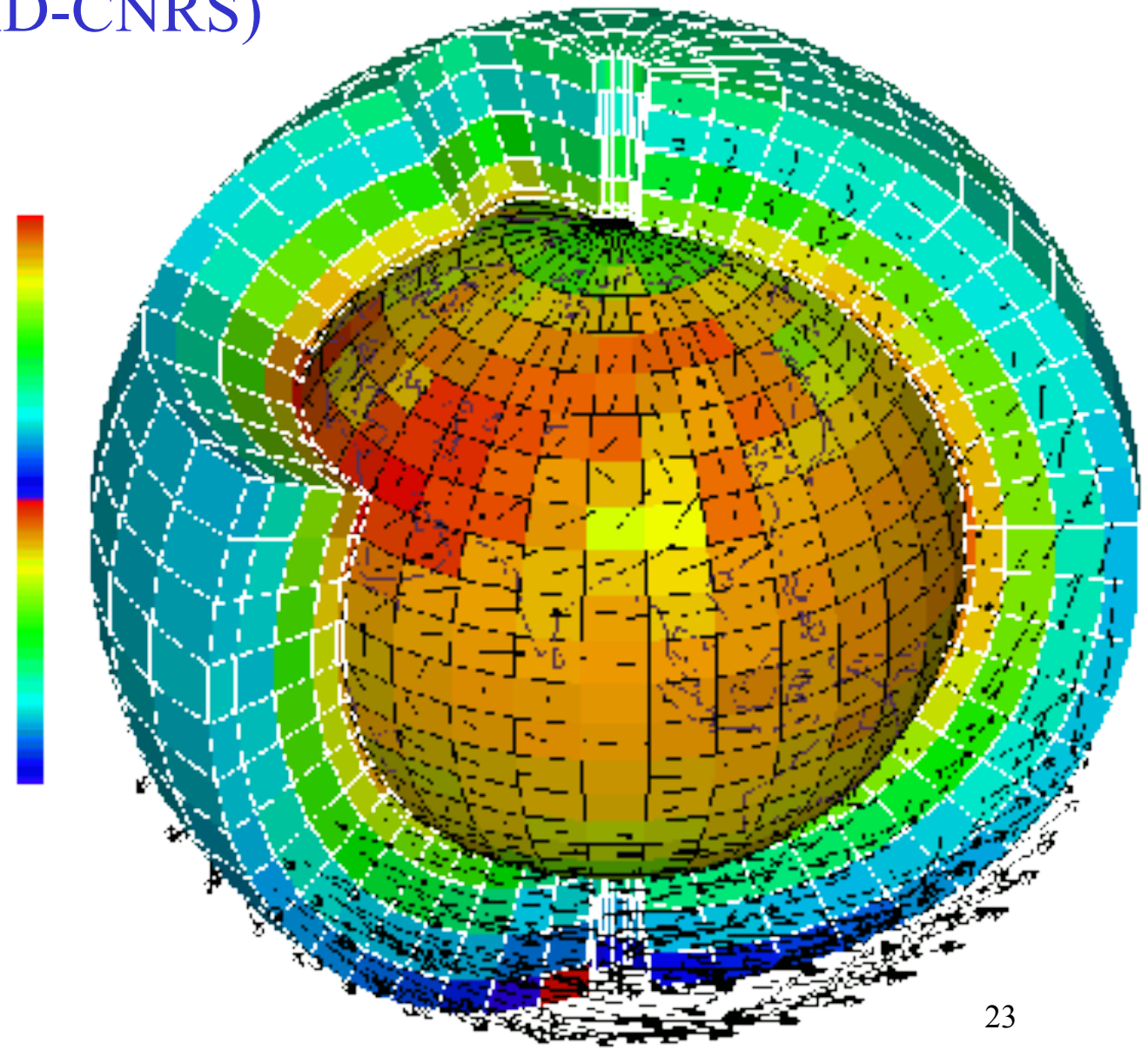
There exist at present two forms of discretization

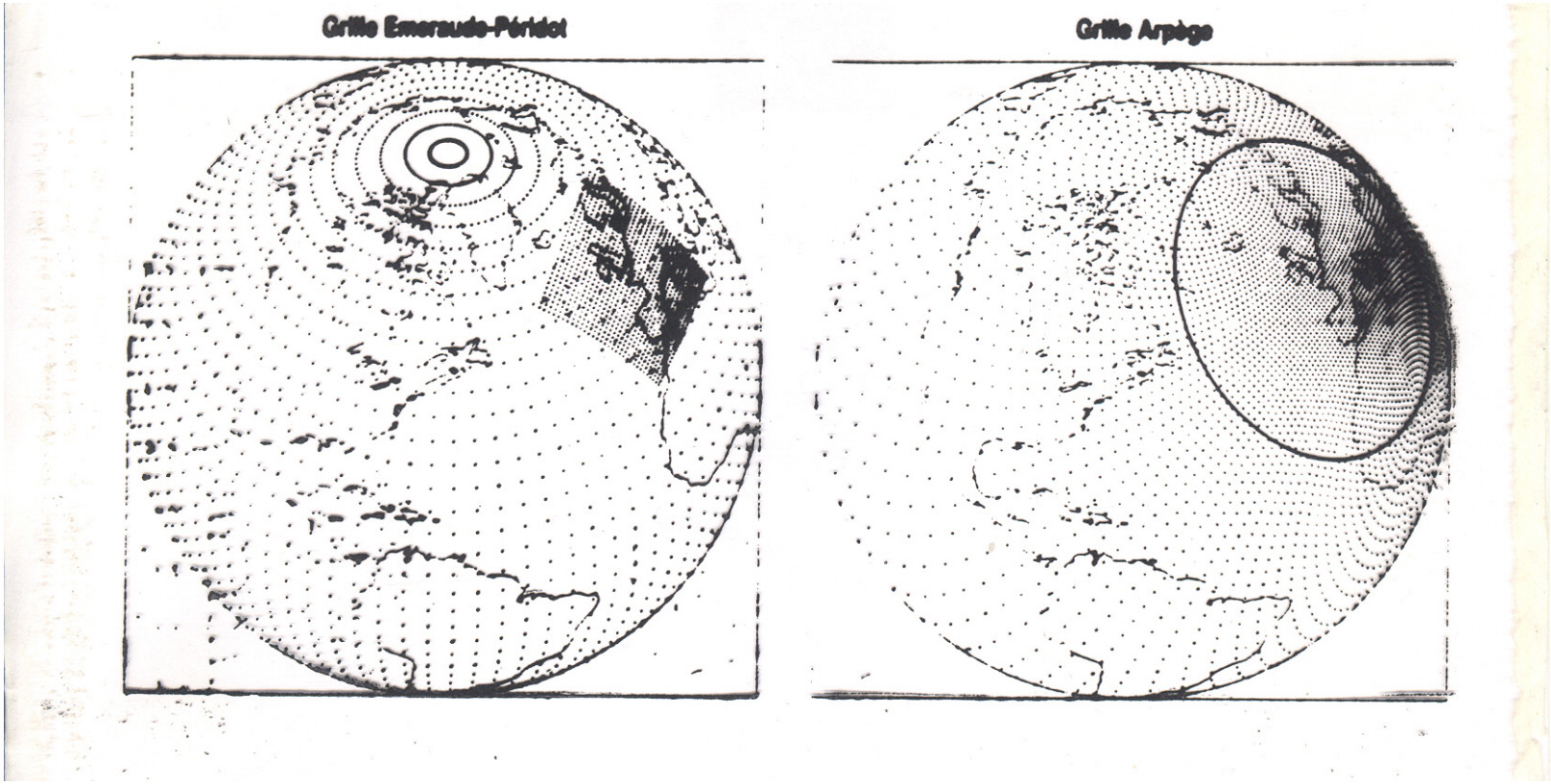
- Gridpoint discretization

- (Semi-)spectral discretization (mostly for global models, and most often only in the horizontal direction)

Finite element discretization, which is very common in many forms of numerical modelling, is rarely used for modelling of the atmosphere. It is more frequently used for oceanic modelling, where it allows to take into account the complicated geometry of coast-lines.

Schematic of a gridpoint atmospheric model (L. Fairhead /LMD-CNRS)



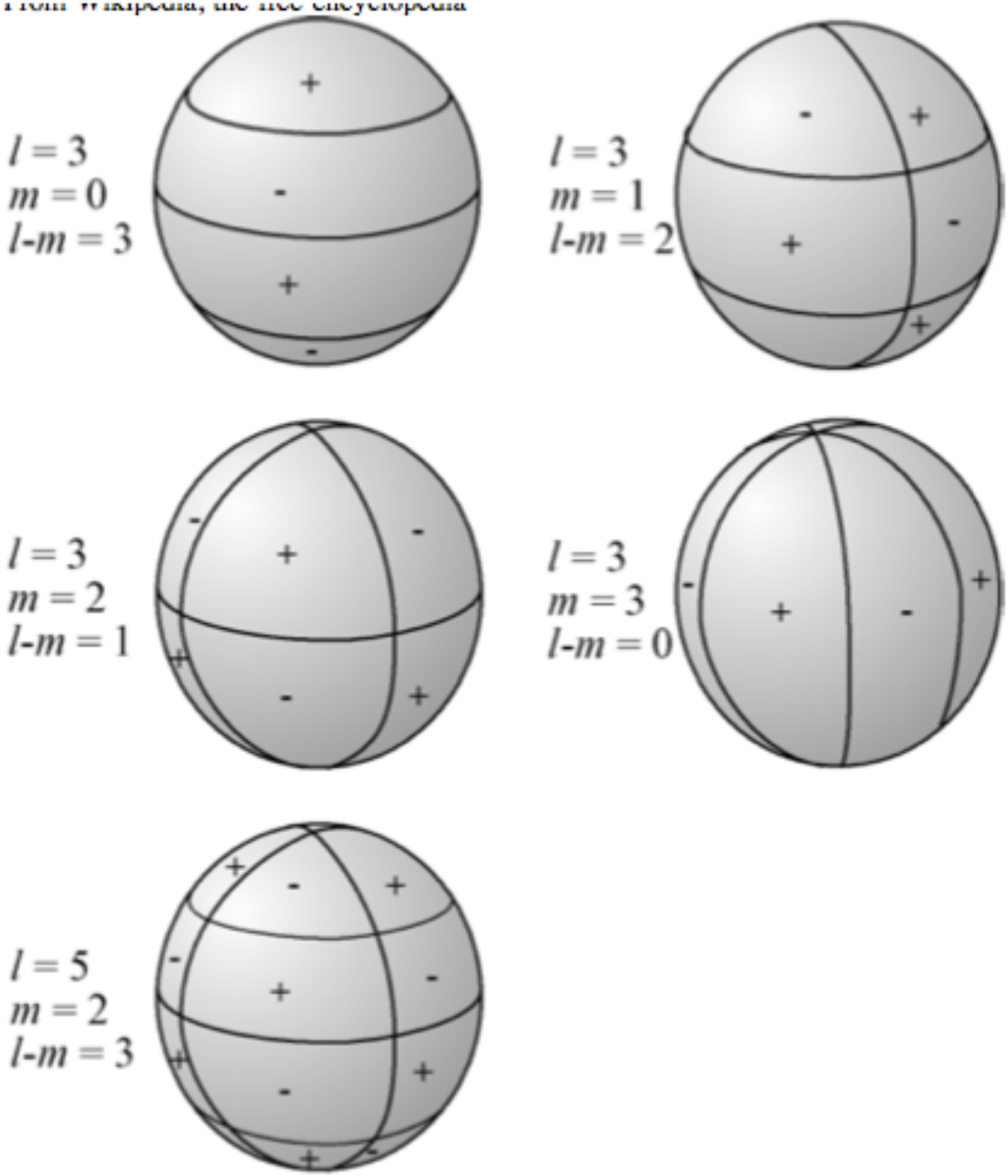


The grids of two of the models of Météo-France *(La Météorologie)*

In gridpoint models, meteorological fields are defined by values at the nodes of a the grid. Spatial and temporal derivatives are expressed by finite differences.

In spectral models, fields are defined by the coefficients of their expansion along a prescribed set of basic functions. In the case of global meteorological models, those basic functions are the spherical harmonics (eigenfunctions of the laplacian at the surface of the sphere).

From Wikipedia, the free encyclopedia



Linear operations, and in particular differentiation, are performed in spectral space, while nonlinear operations and ‘physical’ computations (advection, diabatic heating and cooling, …) are performed in gridpoint physical space. This requires constant transformations from one space to the other, which are made possible at an acceptable cost through the systematic use of Fast Fourier Transforms.

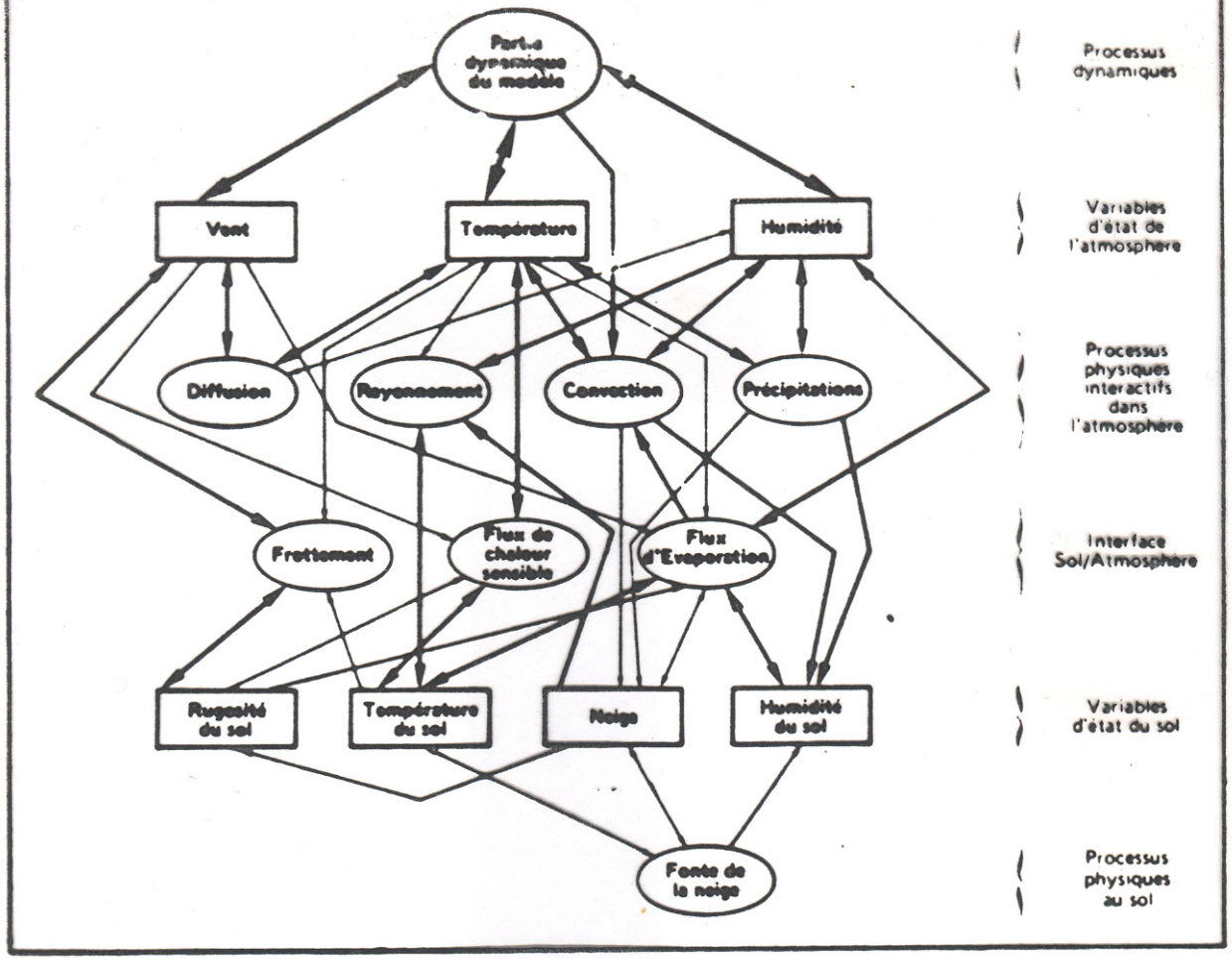
For that reason, those models are called *semi-spectral*.

In the parlance of the trade, one distinguishes two different parts in models. The ‘dynamics’ deals with the physically reversible processes (pressure forces, Coriolis force, advection, …), while the ‘physics’ deals with physically irreversible processes, in particular the diabatic heating term Q in the energy equation, and also the parameterization of subgrid scales effects.

Numerical schemes have been progressively developed and validated for the ‘dynamics’ component of models, which are by and large considered now to work satisfactorily (although regular improvements are still being made).

The situation is different as concerns ‘physics’, where many problems remain (as concerns for instance subgrid scales parameterization, the water cycle and the associated exchanges of energy, or the exchanges between the atmosphere and the underlying medium). ‘Physics’ as a whole remains the weaker point of models, and is still the object of active research.

5 - SCHEMA DES INTERACTIONS PHYSIQUES DANS LE MODELE



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

Since 26 January 2010

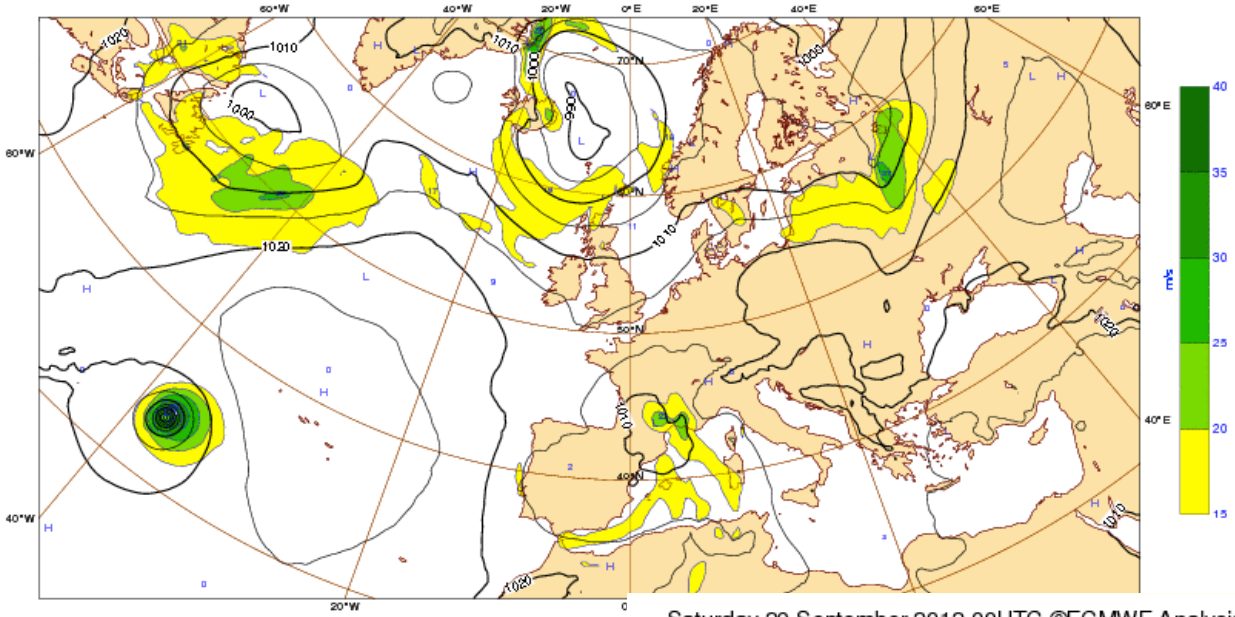
Horizontal spherical harmonics triangular truncation T1279 (horizontal resolution ≈ 16 kilometres, but still hydrostatic)

91 levels on the vertical (0 - 80 km)

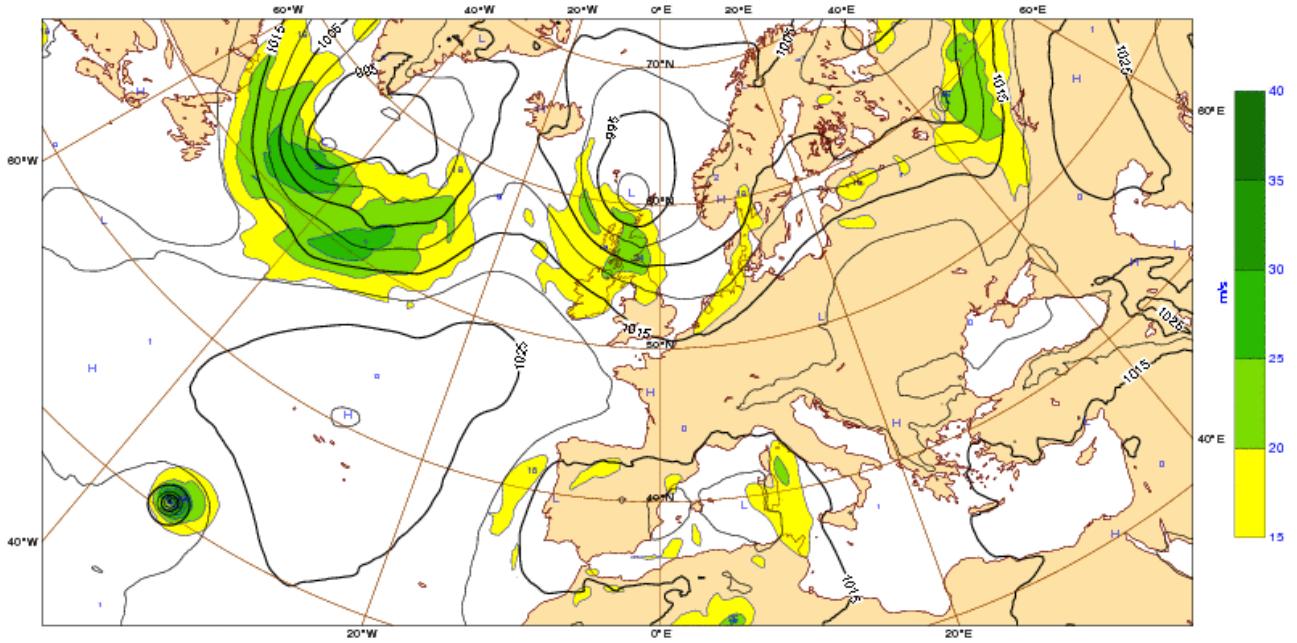
Dimension of state vector $n \approx 1.5\ 10^9$

Timestep = 10 minutes

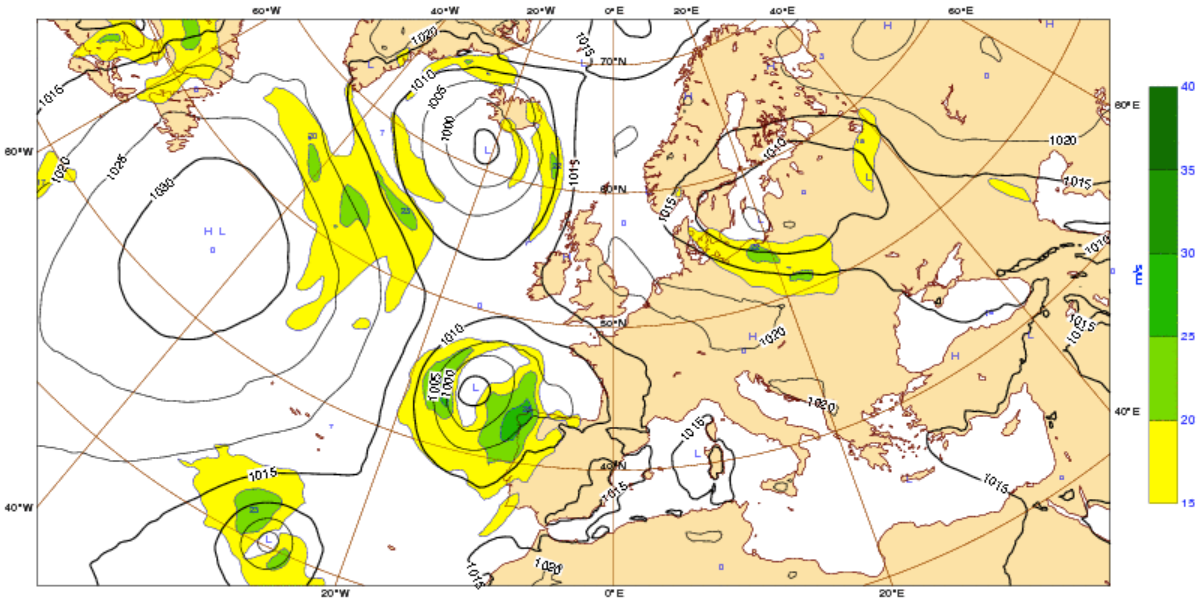
Sunday 23 September 2012 00UTC ©ECMWF Forecast t+144 VT: Saturday 29 September 2012 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



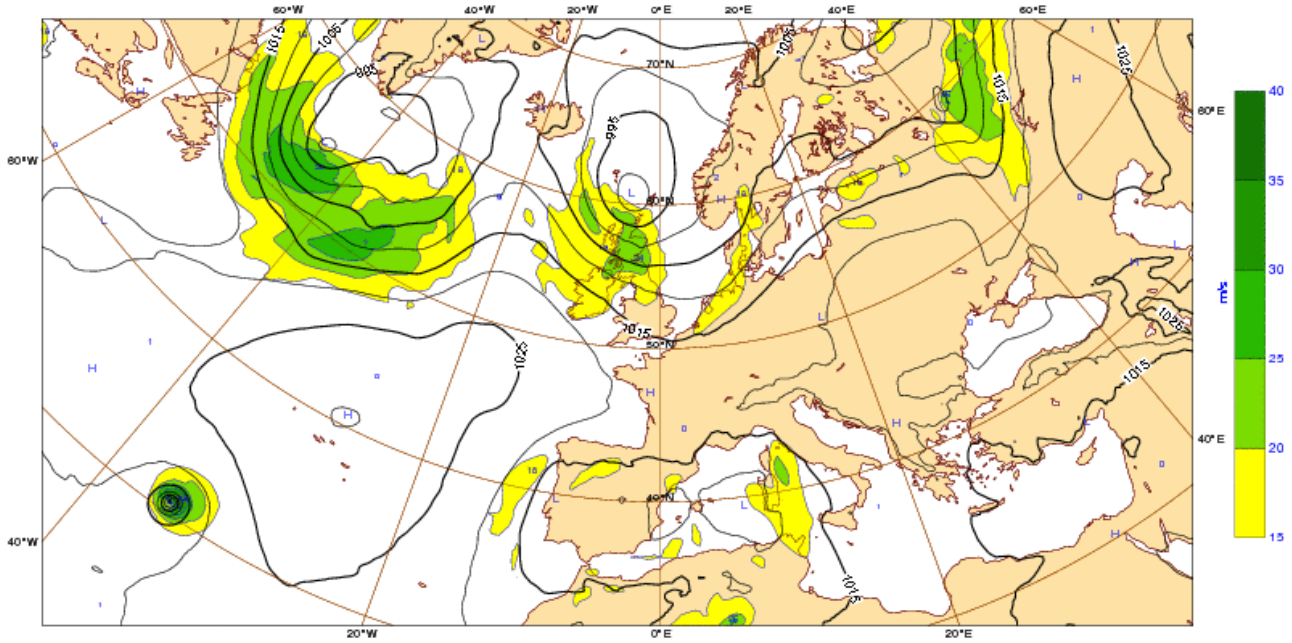
Saturday 29 September 2012 00UTC ©ECMWF Analysis t+000 VT: Saturday 29 September 2012 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



Sunday 23 September 2012 00UTC ©ECMWF Analysis t+000 VT: Sunday 23 September 2012 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



Saturday 29 September 2012 00UTC ©ECMWF Analysis t+000 VT: Saturday 29 September 2012 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



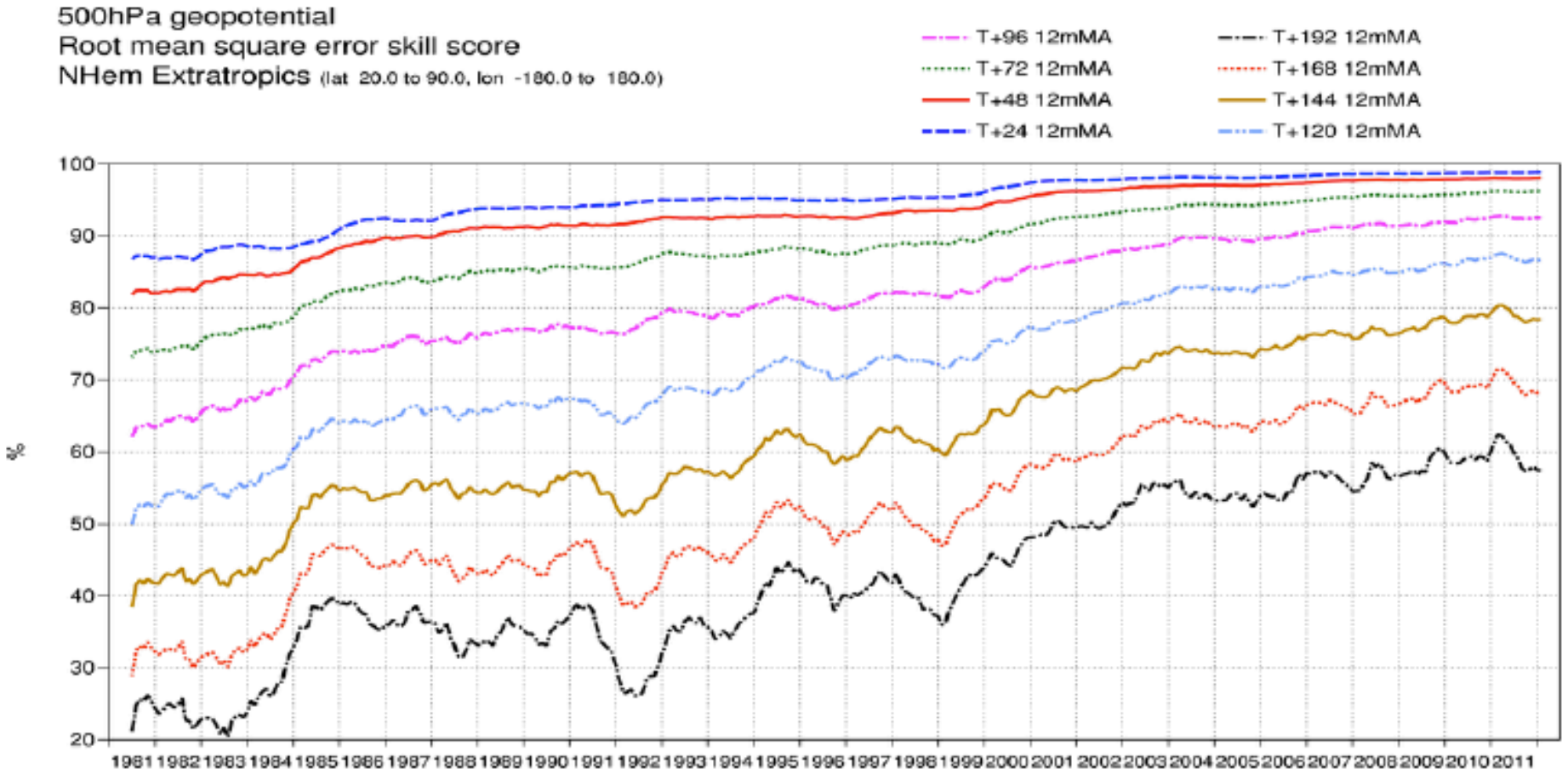
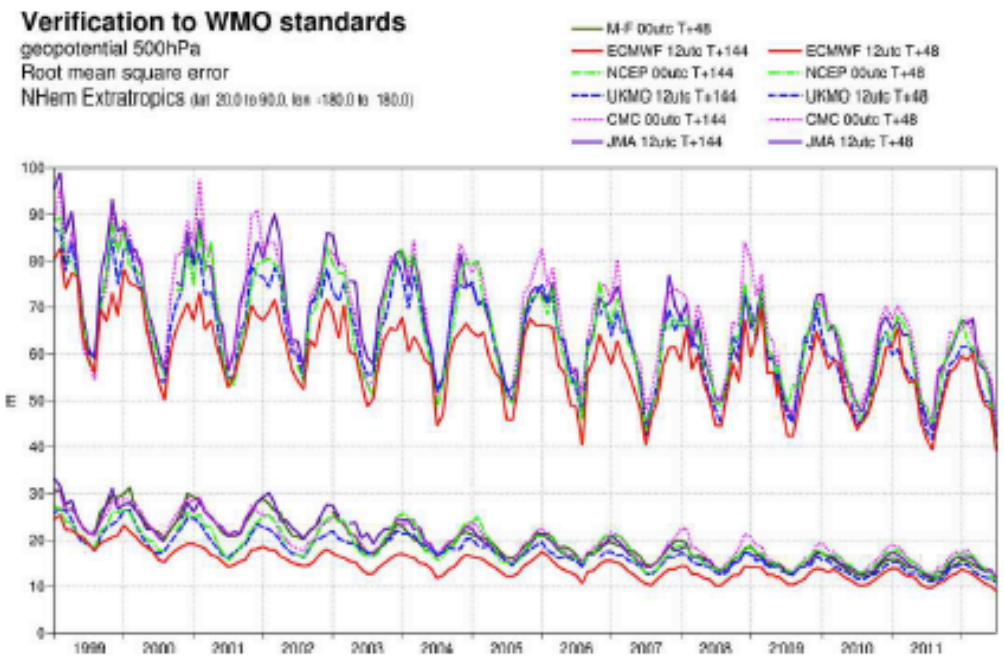


Figure 2: 500 hPa geopotential height skill score for Europe (top) and the northern hemisphere extratropics (bottom), showing 12-month moving averages for forecast ranges from 24 to 192 hours. The last point on each curve is for the 12-month period August 2011–July 2012.

Persistence = 0 ; climatology = 50 at long range

http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/601-700/tm688.pdf

http://www.ecmwf.int/
publications/library/
ecpublications/_pdf/
tm/601-700/tm688.pdf



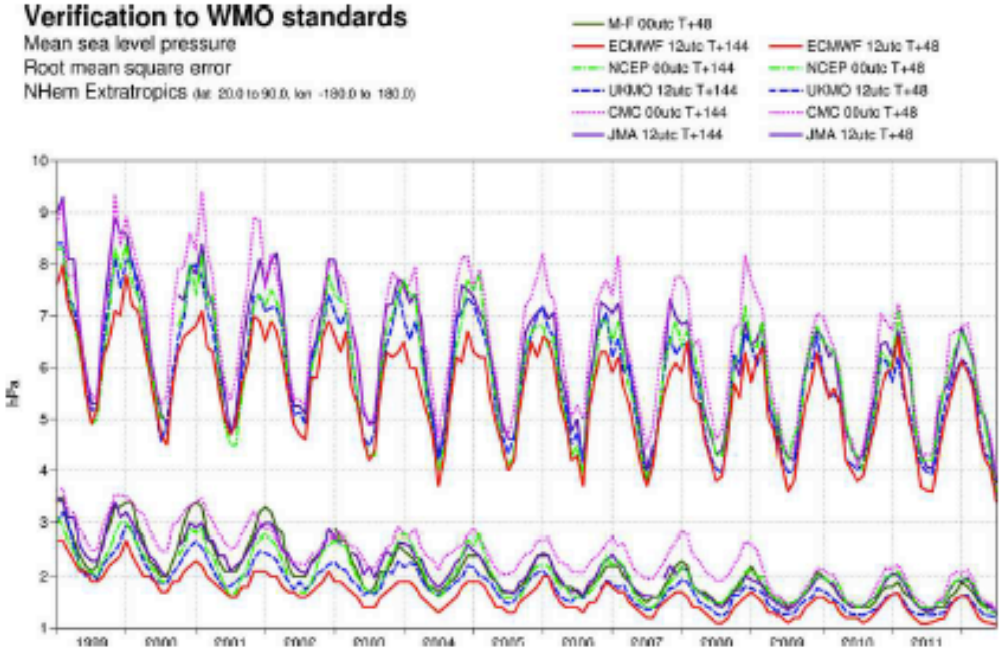
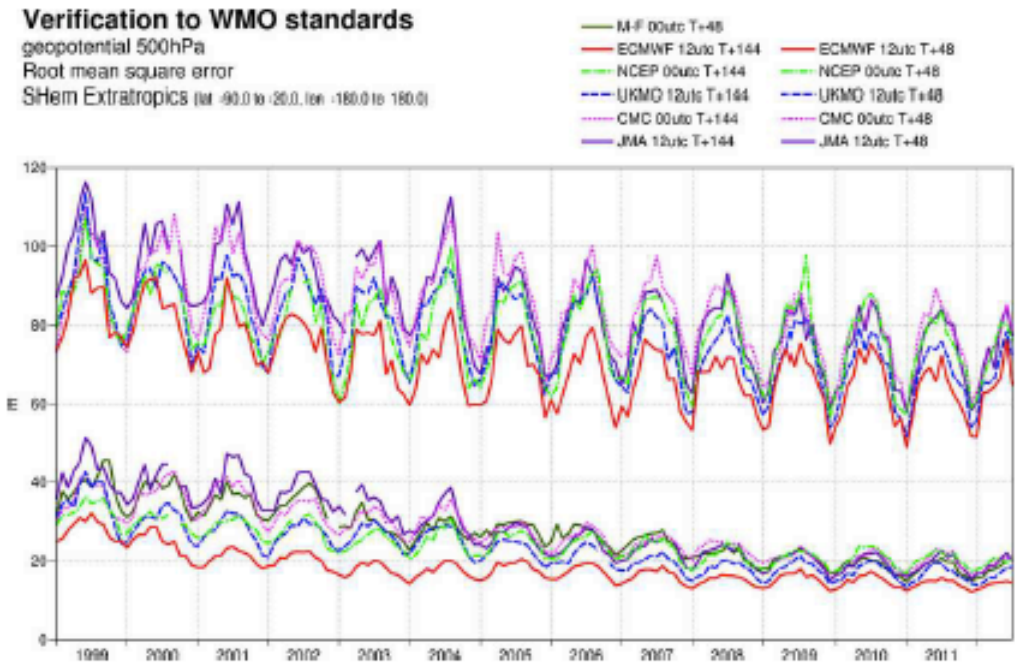


Figure 11: WMO-exchanged scores from global forecast centres. RMS error over northern extratropics for 500 hPa geopotential height (top) and mean sea level pressure (bottom). In each



http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/601-700/tm688.pdf

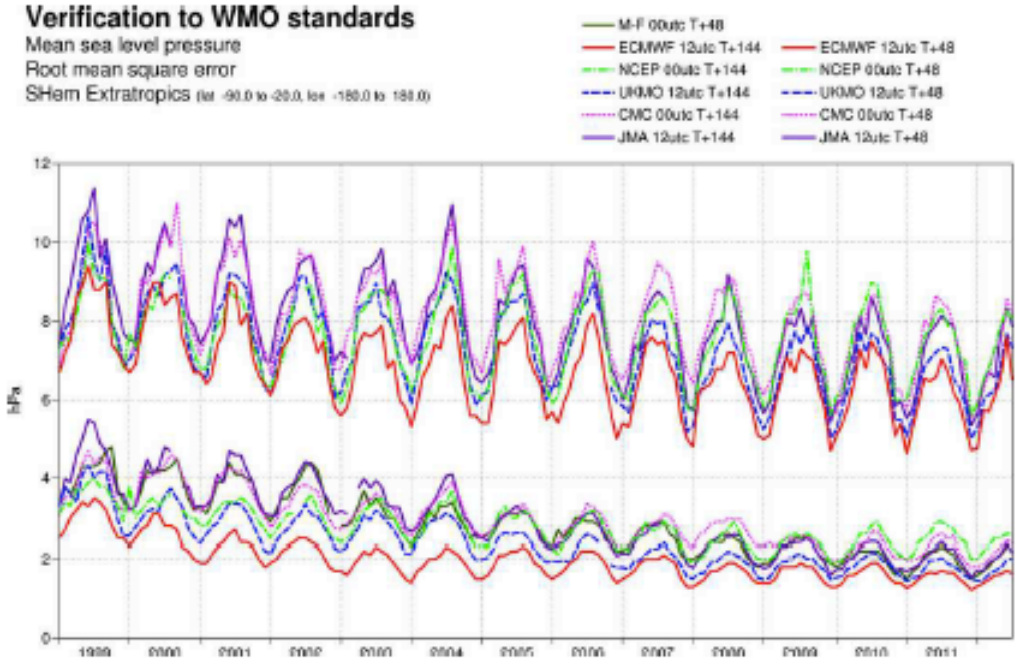
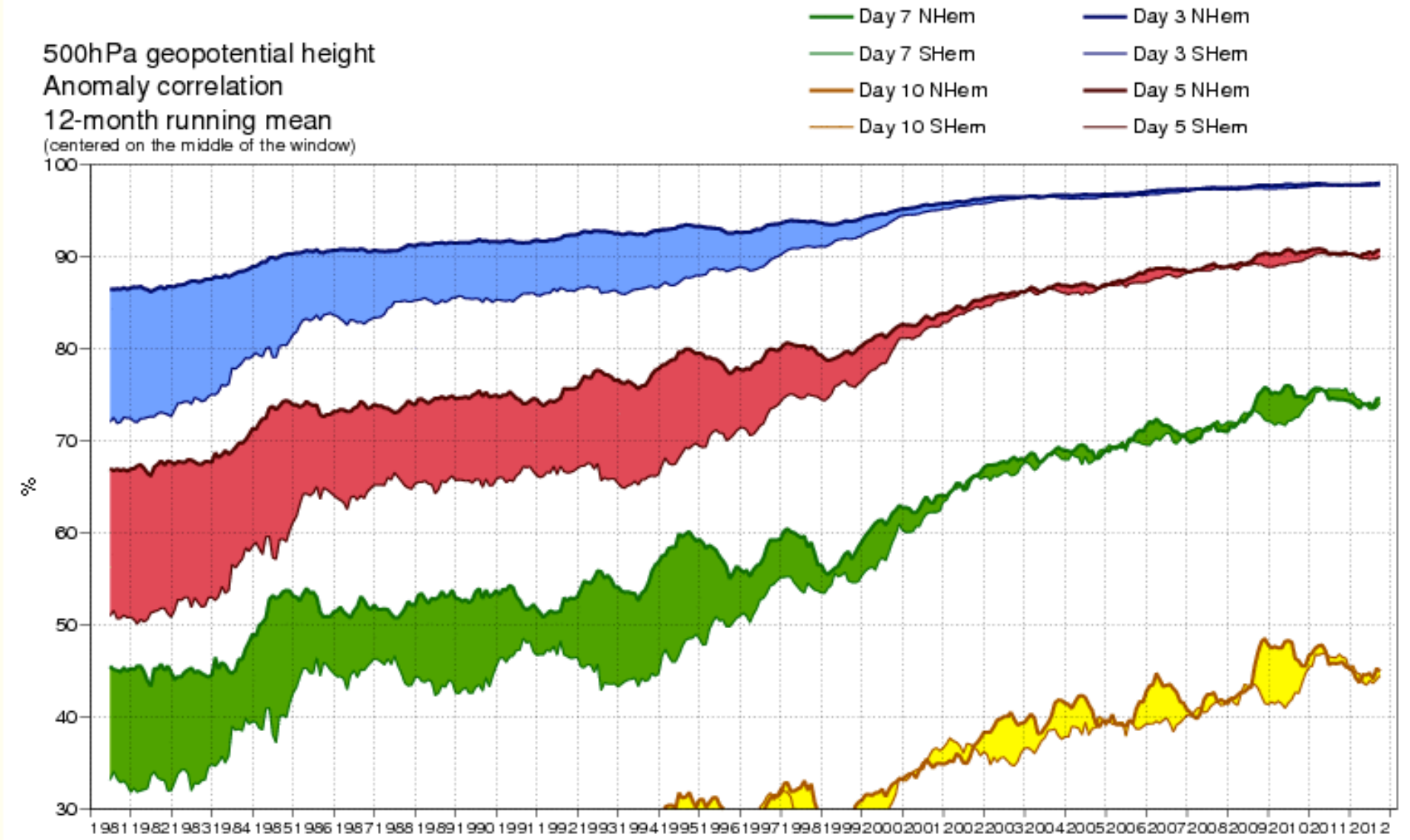


Figure 12: As Figure 11 for the southern hemisphere.

Anomaly correlation of ECMWF 500hPa height forecasts



http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/601-700/tm688.pdf

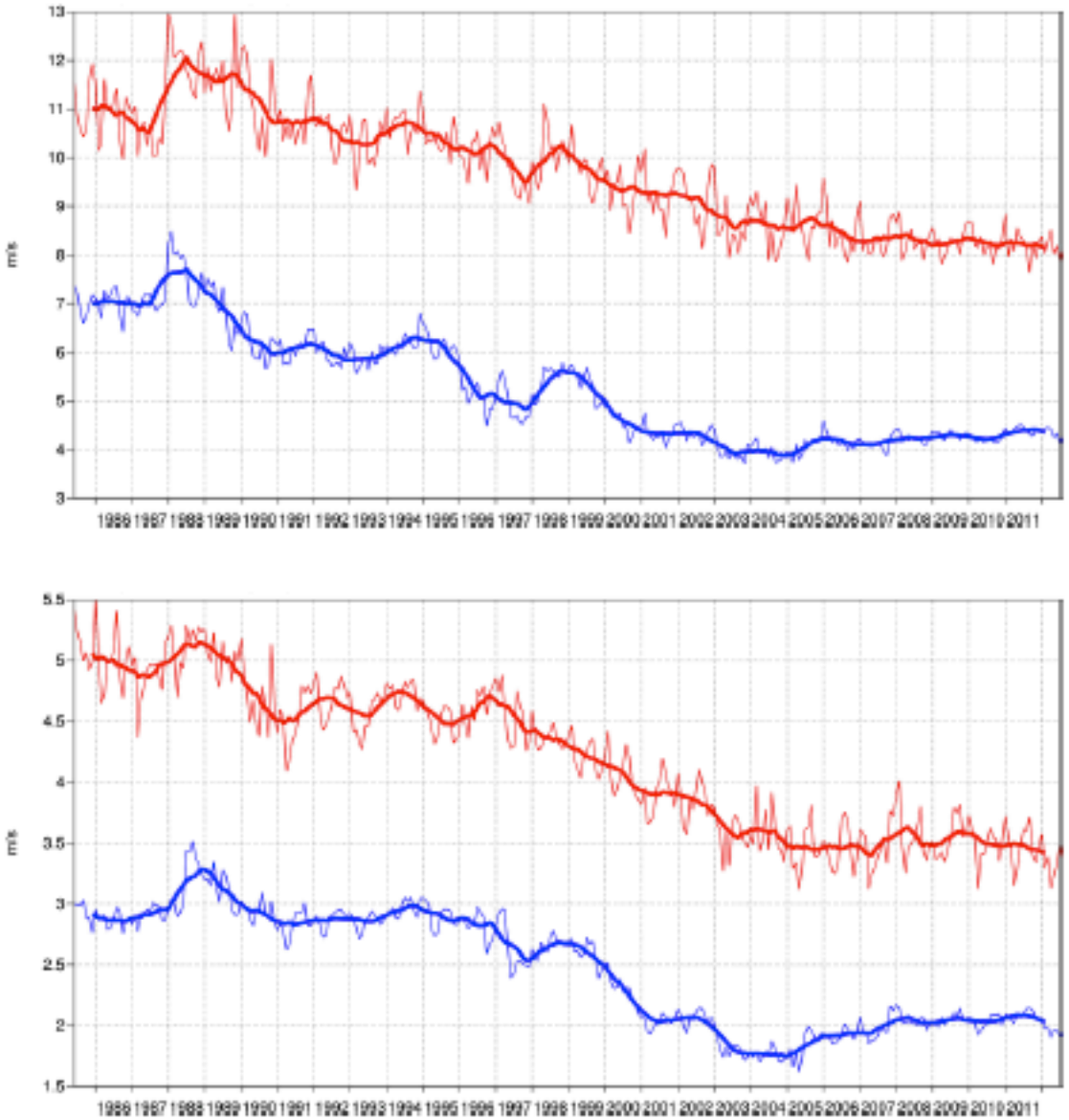


Figure 10: Forecast performance in the tropics. Curves show the monthly average RMS vector wind errors at 200 hPa (top) and 850 hPa (bottom) for one-day (blue) and five-day (red) forecasts. 12-month moving average scores are also shown (in bold).

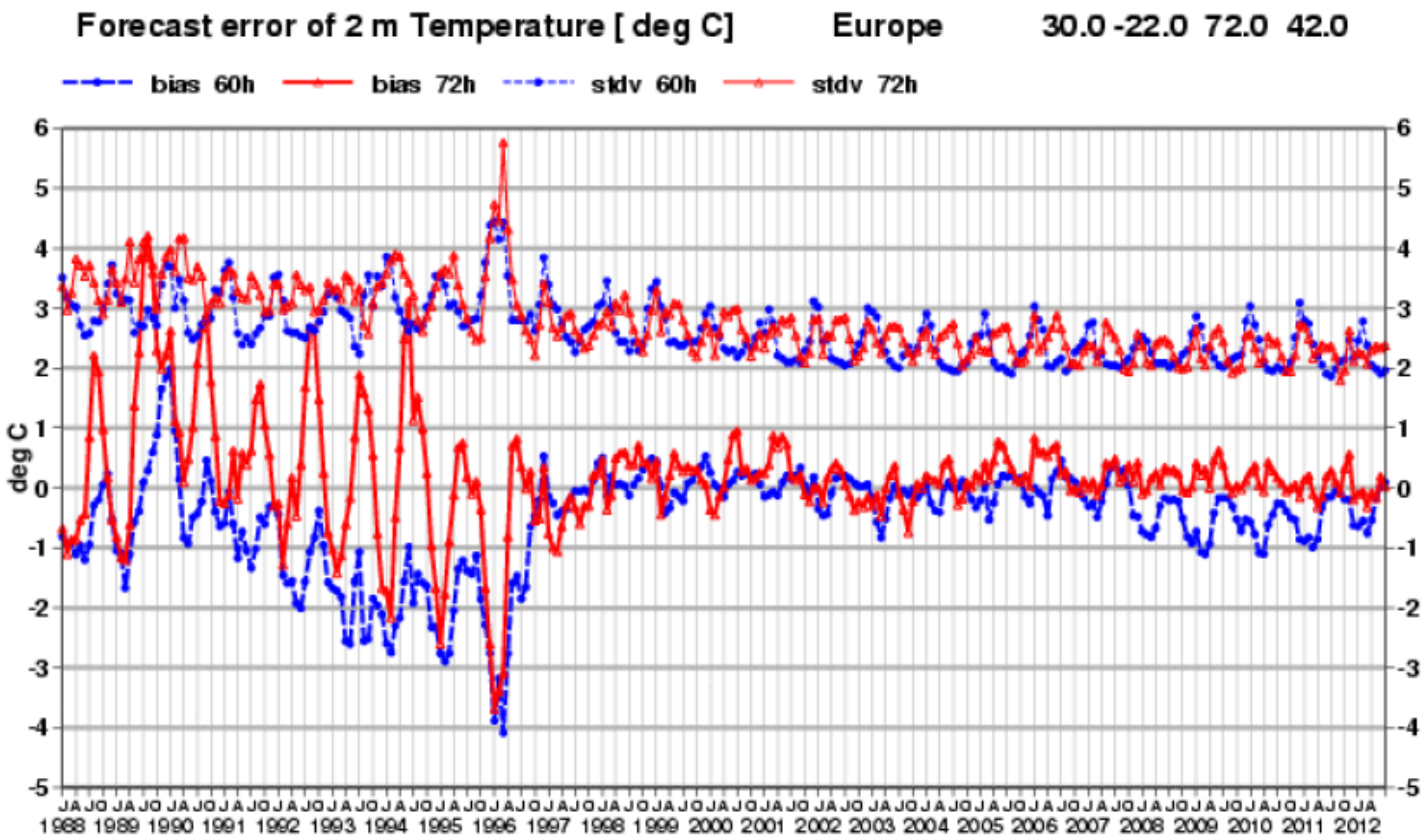


Figure 19: Verification of 2 m temperature forecasts against European SYNOP data on the GTS for 60-hour (night-time) and 72-hour (daytime) forecasts. Lower pair of curves shows bias, upper curves are standard deviation of error.

http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/601-700/tm688.pdf

http://www.ecmwf.int/
publications/library/
ecpublications/_pdf/
tm/601-700/tm688.pdf

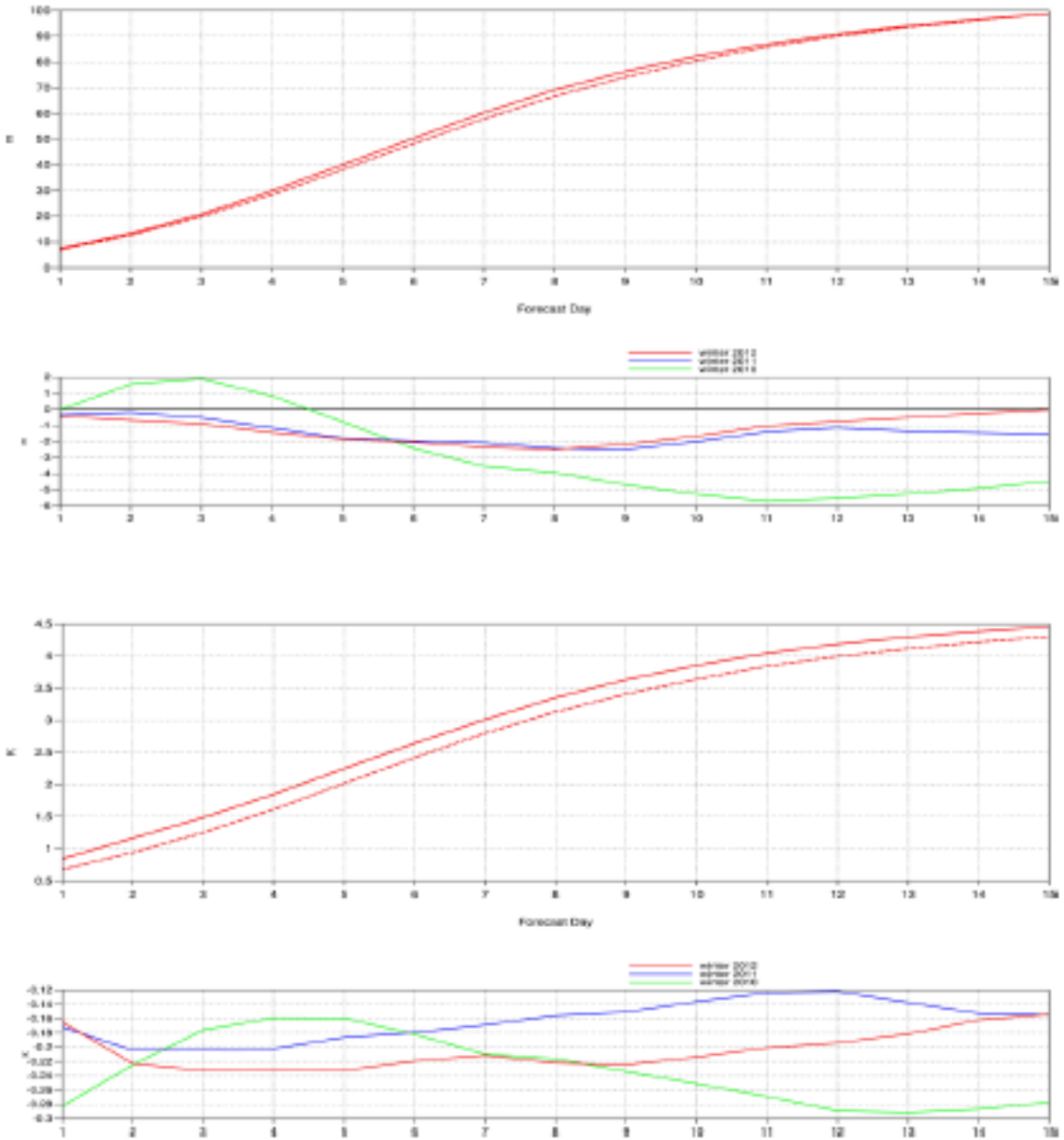
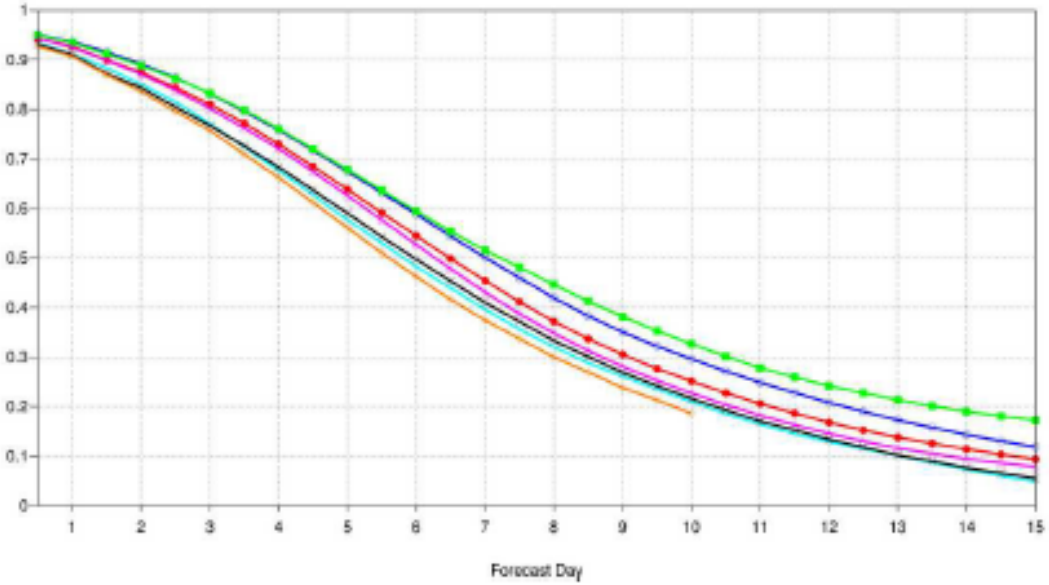
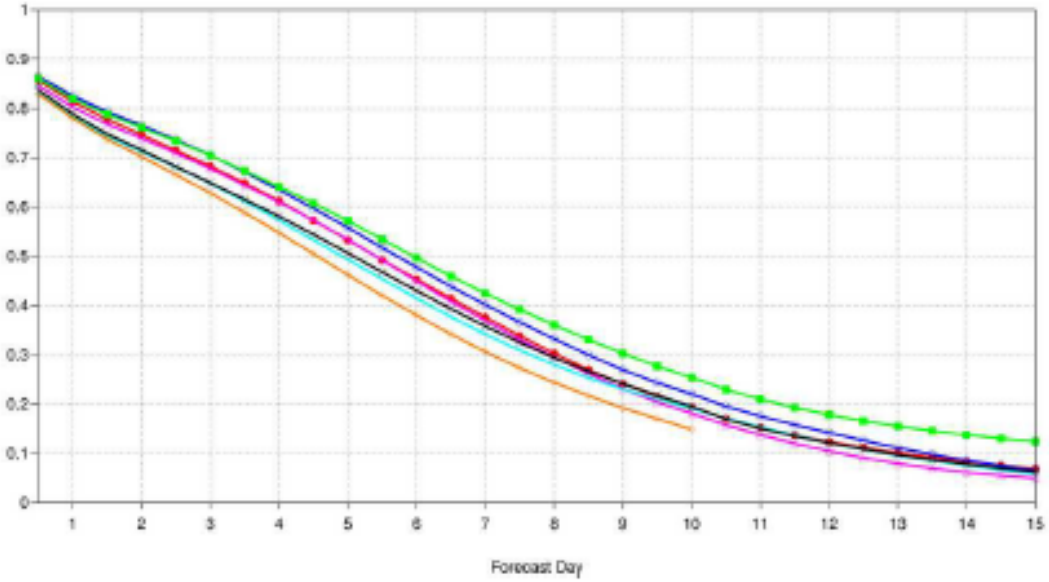


Figure 8: Ensemble spread (standard deviation, dashed lines) and RMS error of ensemble-mean (solid lines) for winter 2011–2012 (upper figure in each panel), and differences of ensemble spread and RMS error of ensemble mean for last three winter seasons (lower figure in each panel, negative values indicate spread is too small); plots are for 500 hPa geopotential (top) and 850 hPa temperature (bottom) over the extratropical northern hemisphere for forecast days 1 to 15.





http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/601-700/tm688.pdf

Figure 9: CPRSS for 500 hPa height (top) and 850 hPa temperature (bottom) ensemble forecasts for winter (December–February) over the extratropical northern hemisphere. Skill from the ensemble day 1–15 forecasts is shown for winters 2011–12 (red), 2010–11 (blue), 2009–10 (green), 2008–09 (magenta) and 2007–08 (cyan). The ensemble only ran to ten days in 2005–06 (orange).

ECMWF

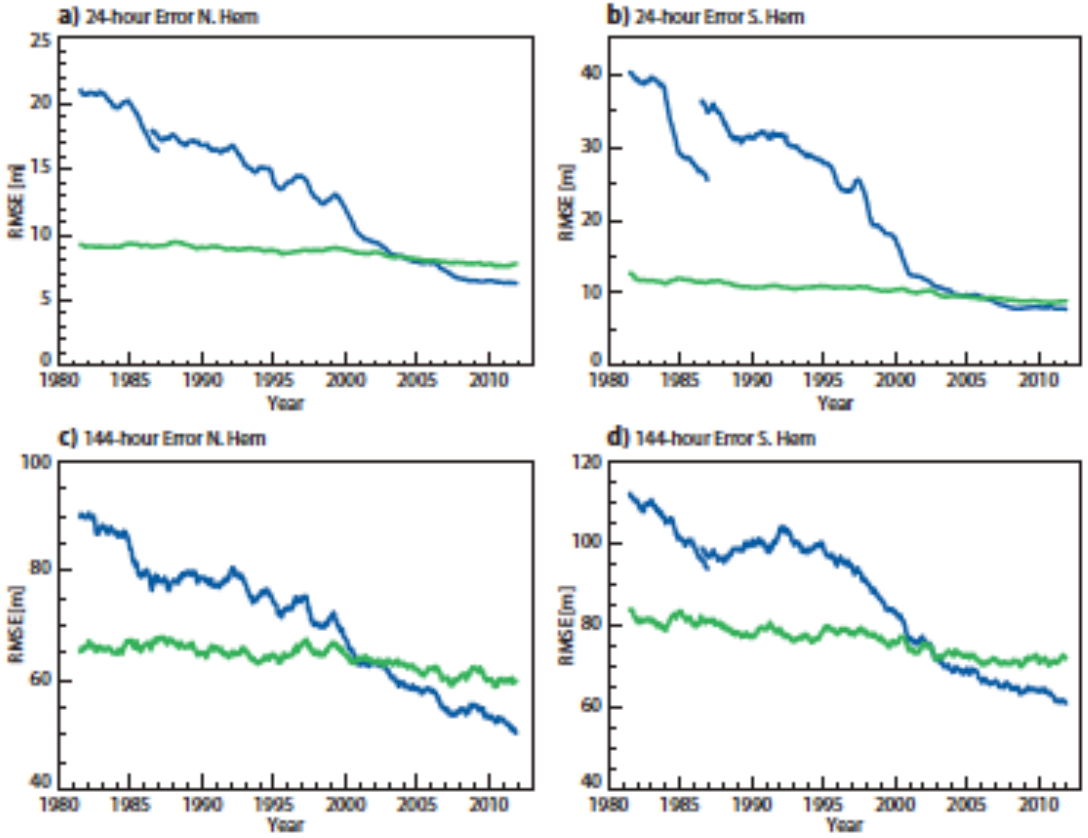


FIG. 3. Evolution of forecast errors from 1981 to 2012 for N.Hem (a and c) and S.Hem (b and d). Operational forecasts (blue) and ERA Interim (green). Note that before 1986 the operational analysis is used to verify the operational forecasts, after 1986 ERA Interim is used for the verification (with an overlap of 6 months present).

Magnusson and Källén, *Mon. Wea. Rev.*, in press

Remaining Problems

Mostly in the 'physics' of models (*Q* and *F* terms in basic equations)

- Water cycle (evaporation, condensation, influence on radiation absorbed or emitted by the atmosphere)

- Exchanges with ocean or continental surface (heat, water, momentum, …)

- …

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

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, ….)
- …

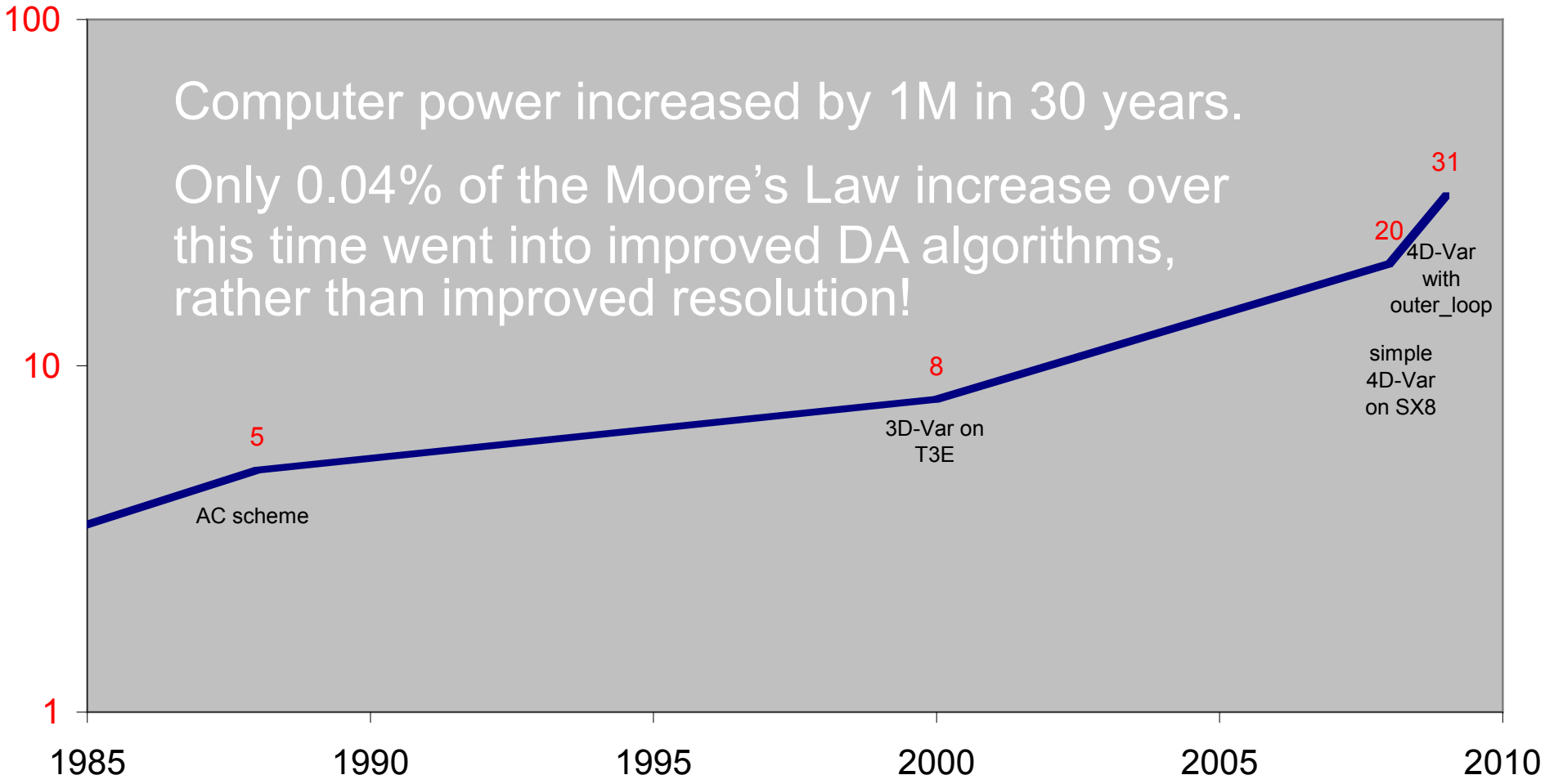
Solution most often (if not always) based on Bayesian, or probabilistic, estimation. 'Equations' are fundamentally the same.

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

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



Courtesy A. Lorenc

Relative cost of the various components of the operational prediction suite at ECMWF (september 2011, J.-N. Thépaut) :

4DVAR: 17%

Ensemble Data Assimilation (EDA) : 15%

Deterministic model : 13%

Ensemble Prediction System (EPS) : 53%

others : 2%

EDA produces both the background error covariances for 4D-Var and the initial perturbations (in addition to Singular Vectors) for EPS.

$z_1 = x + \zeta_1$ 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

$P(x = \xi \mid z_1, z_2)$?

$x = \xi \;\Leftrightarrow\; \zeta_1 = z_1 - \xi$ 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

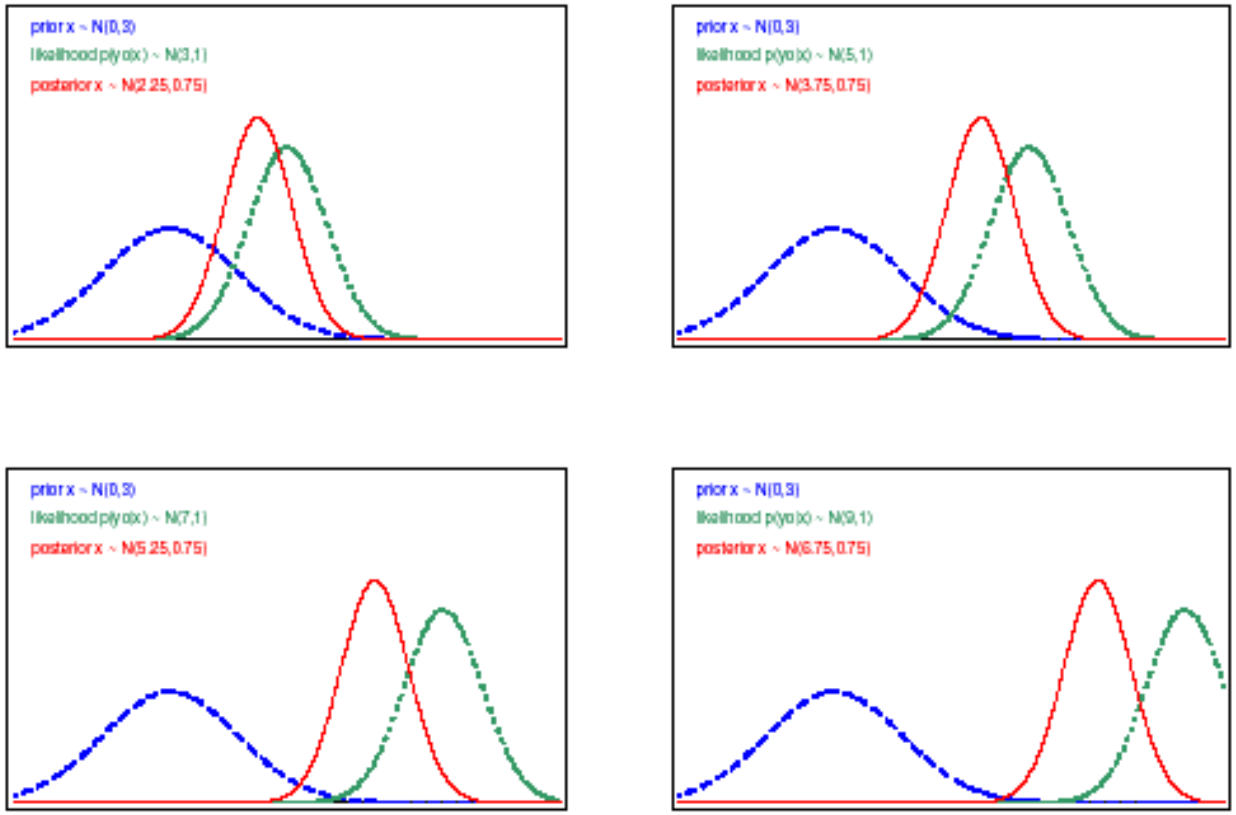


Fig. 1.1: Prior pdf $p(x)$ (dashed line), posterior pdf $p(x|y^o)$ (solid line), and Gaussian likelihood of observation $p(y^o|x)$ (dotted line), plotted against x for various values of y^o. (Adapted from Lorenc and Hammon 1988.)

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

with

Same as before, but ζ_1 and ζ_2 are now distributed according to exponential law 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.

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

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.

$\Leftrightarrow$ 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 $n \approx 10^{6\text{-}9}$ 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 $N \approx O(10\text{-}100)$).

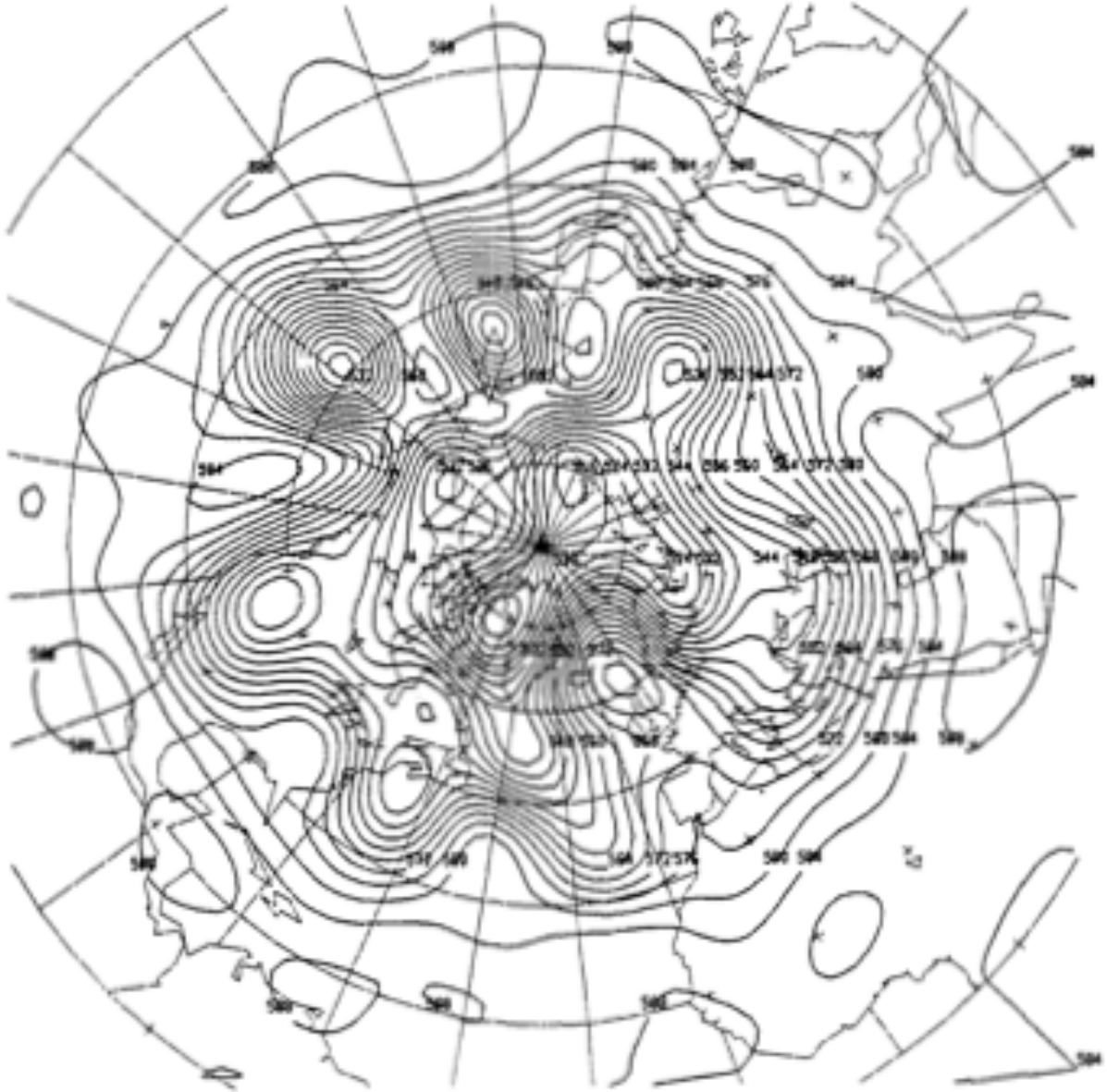


Figure 2. 500 mb height field produced by the operational analysis procedure of Direction de la Météorologie for 00 GMT, 26 April 1984. Units: dam, contour interval: 4 dam. The field has been truncated to the truncation of the model used for the experiments described in the article.

Courtier and Talagrand, *QJRMS*, 1987

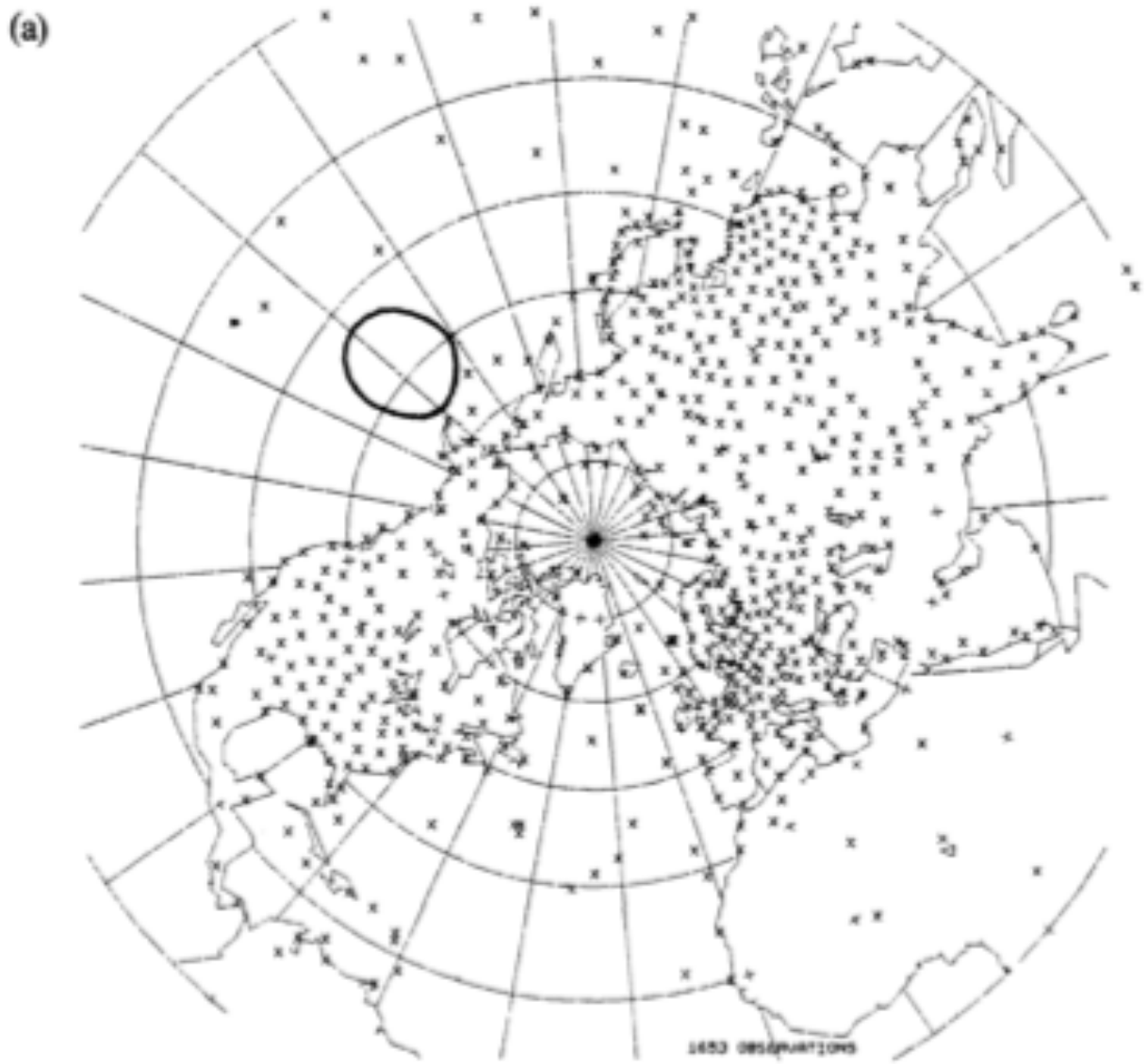


Figure 1. Geographical distribution of the observations used for the assimilation experiments. (a): geopotential observations; (b): wind observations. At most of the points plotted, several observations were made at successive synoptic hours. On each of the two charts, the heavy line delineates the Aleutian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

Courtier and Talagrand, *QJRMS*, 1987

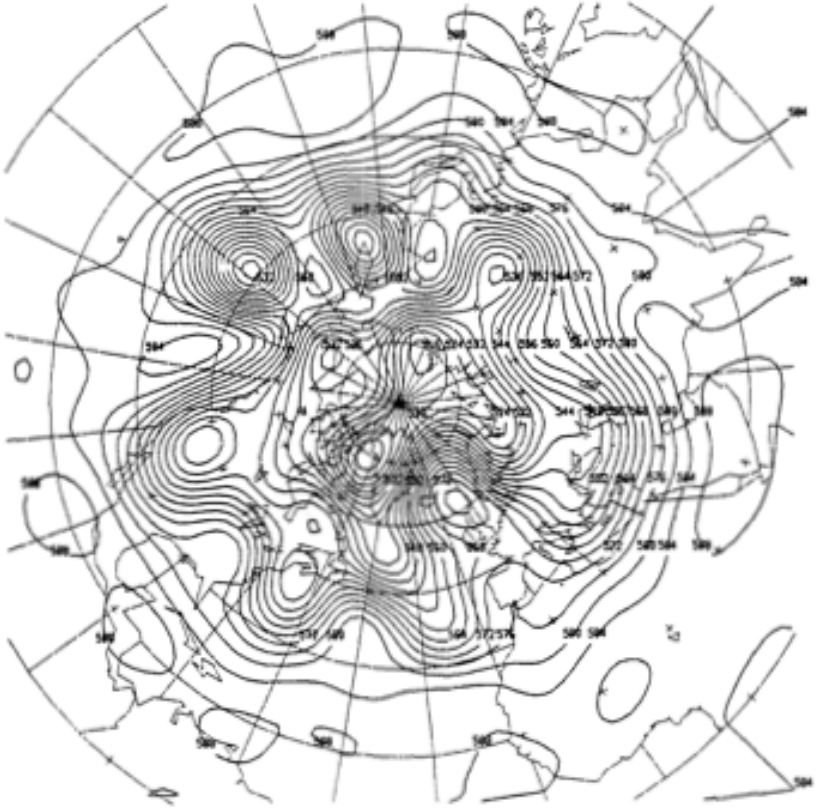


Figure 2. 500 mb height field produced by the operational analysis procedure of Direction de la Météorologie for 00 GMT, 26 April 1984. Units: dam, contour interval: 4 dam. The field has been truncated to the truncation of the model used for the experiments described in the article.

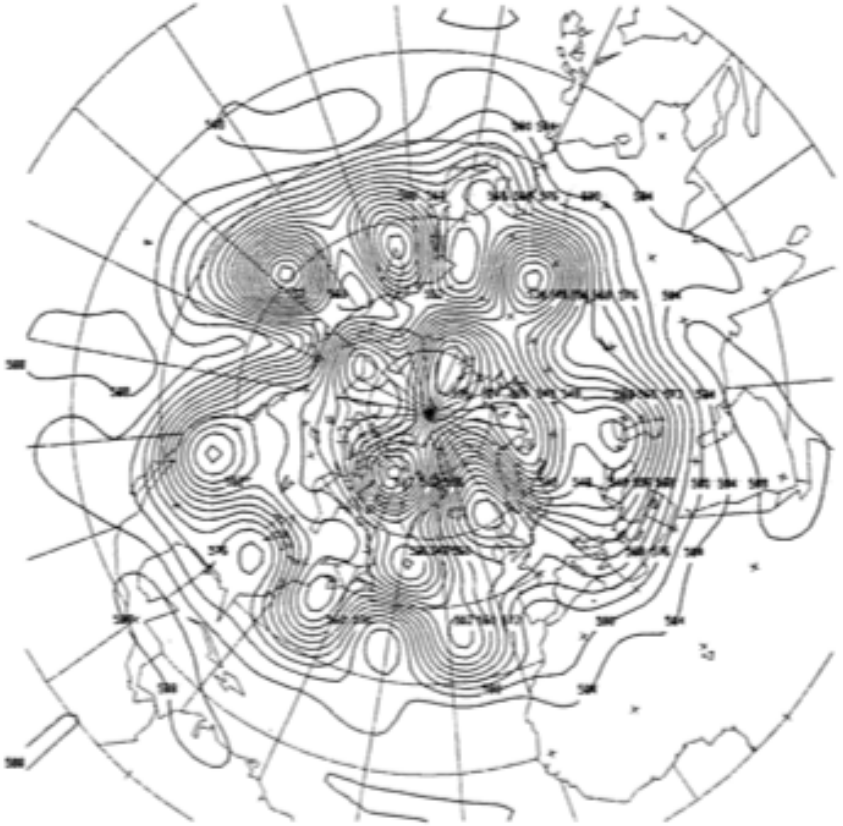


Figure 3. 500 mb height field produced for 00 GMT, 26 April 1984, by the variational analysis minimizing the distance function defined by Eqs. (1)–(2) over a 24-hour period. Units: dam; contour interval: 4 dam.

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)

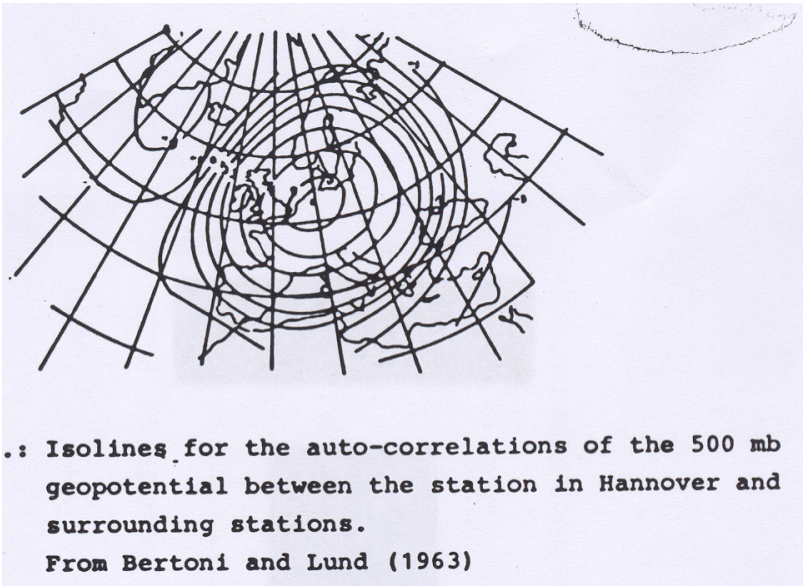
- Expectation $E[\Phi(\xi)]$; $\Phi'(\xi) \equiv \Phi(\xi) - E[\Phi(\xi)]$
- Variance $Var[\varphi(\xi)] = E\{[\varphi'(\xi)]^2\}$

- Covariance function

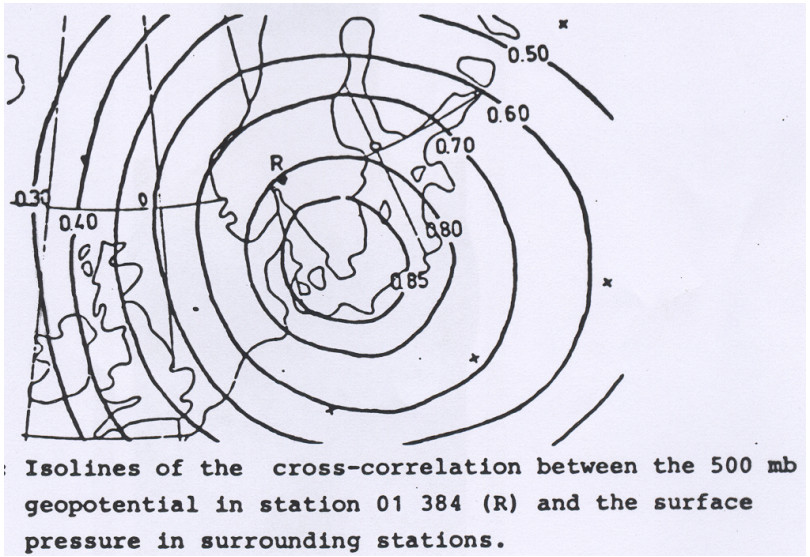
$$(\xi_1, \xi_2) \rightarrow C_\Phi(\xi_1, \xi_2) \equiv E[\Phi'(\xi_1)\, \Phi'(\xi_2)]$$

- 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

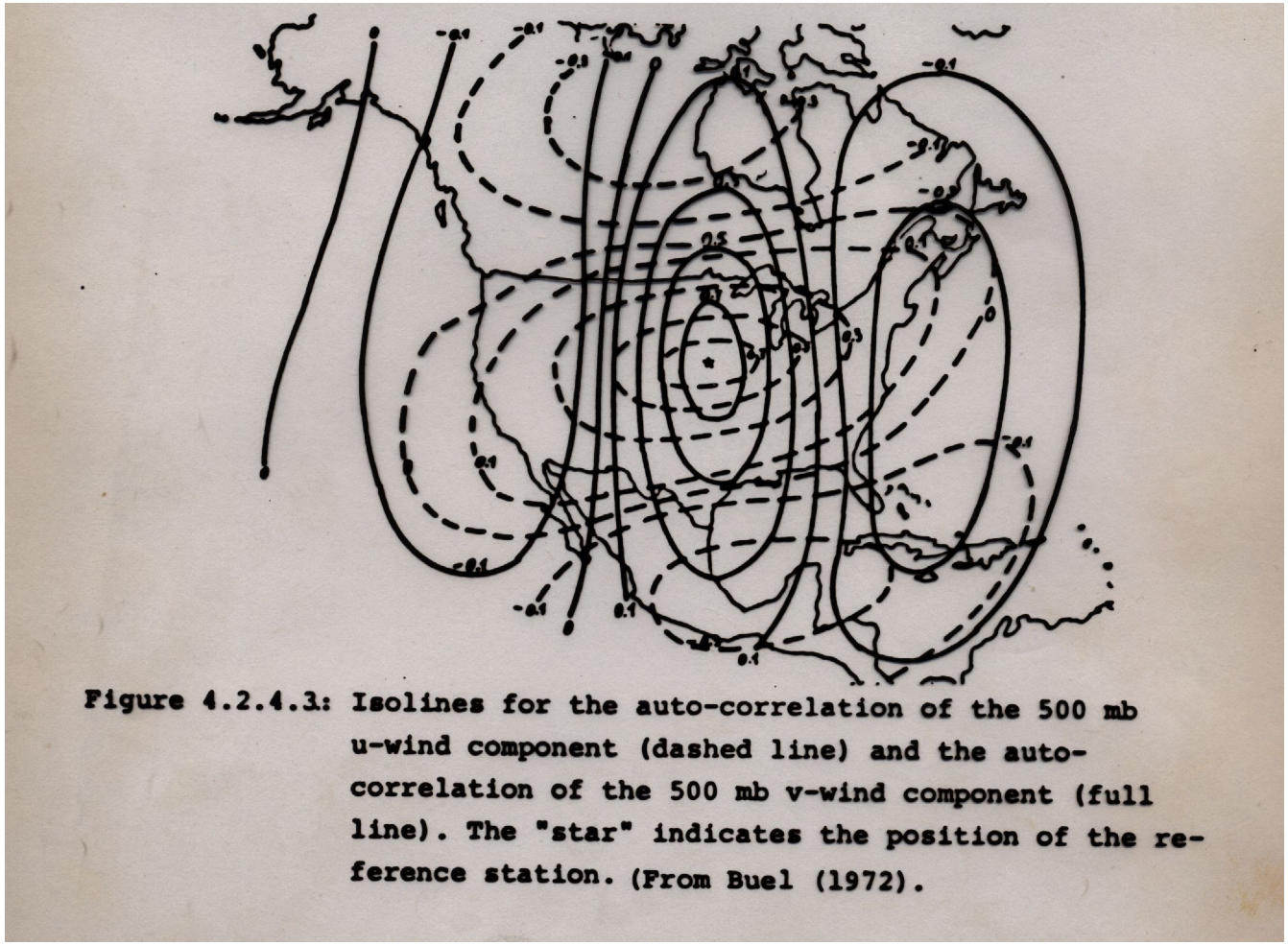


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson

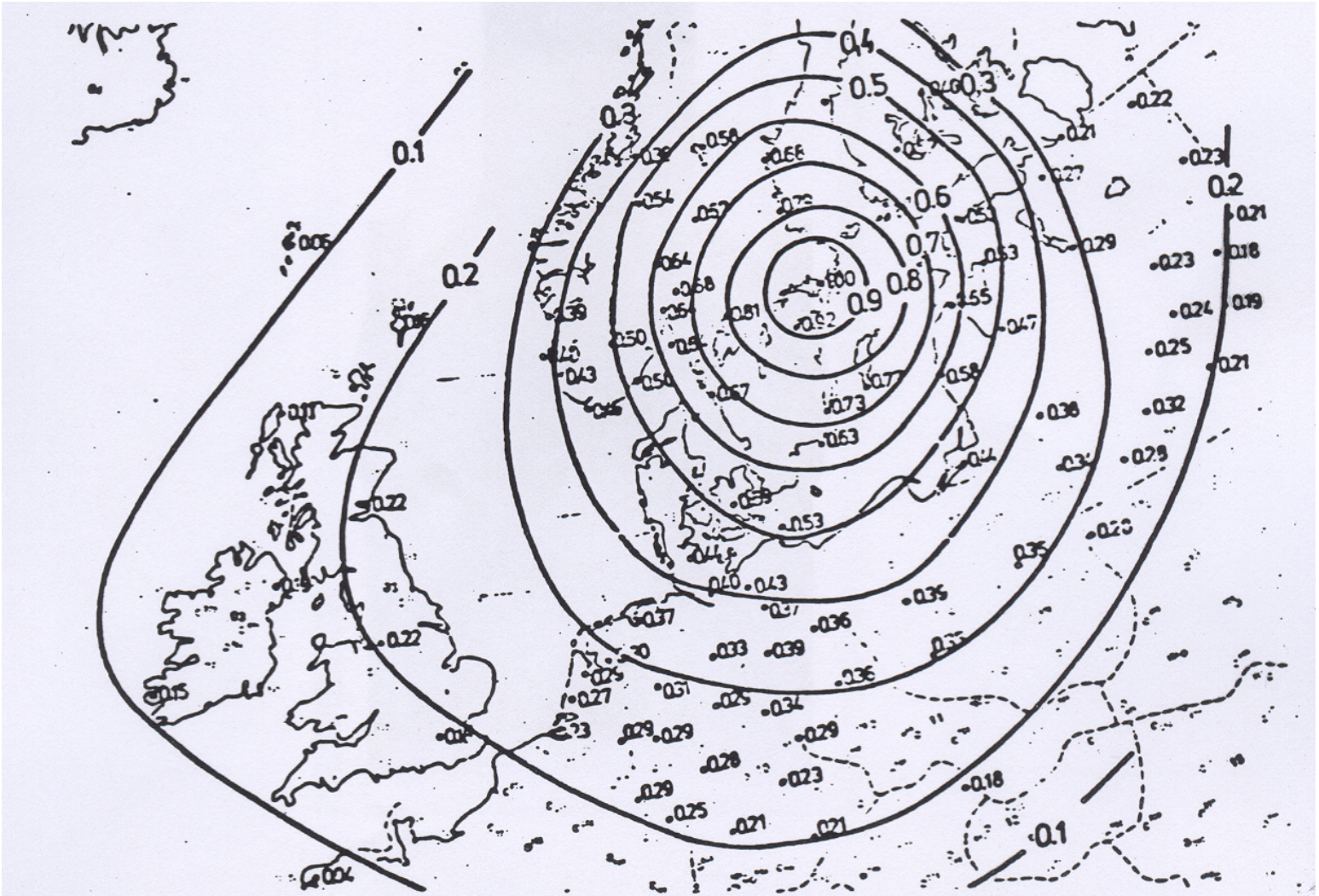


Figure 5.1.1.4.1 Auto-correlation of errors in 12h numerical forecasts of surface pressure in a reference station (Stockholm) and other stations.

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

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

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

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

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

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

i. e.,

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

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

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

Minimized quadratic estimation error

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

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

Optimal Interpolation (continued 2)

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

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

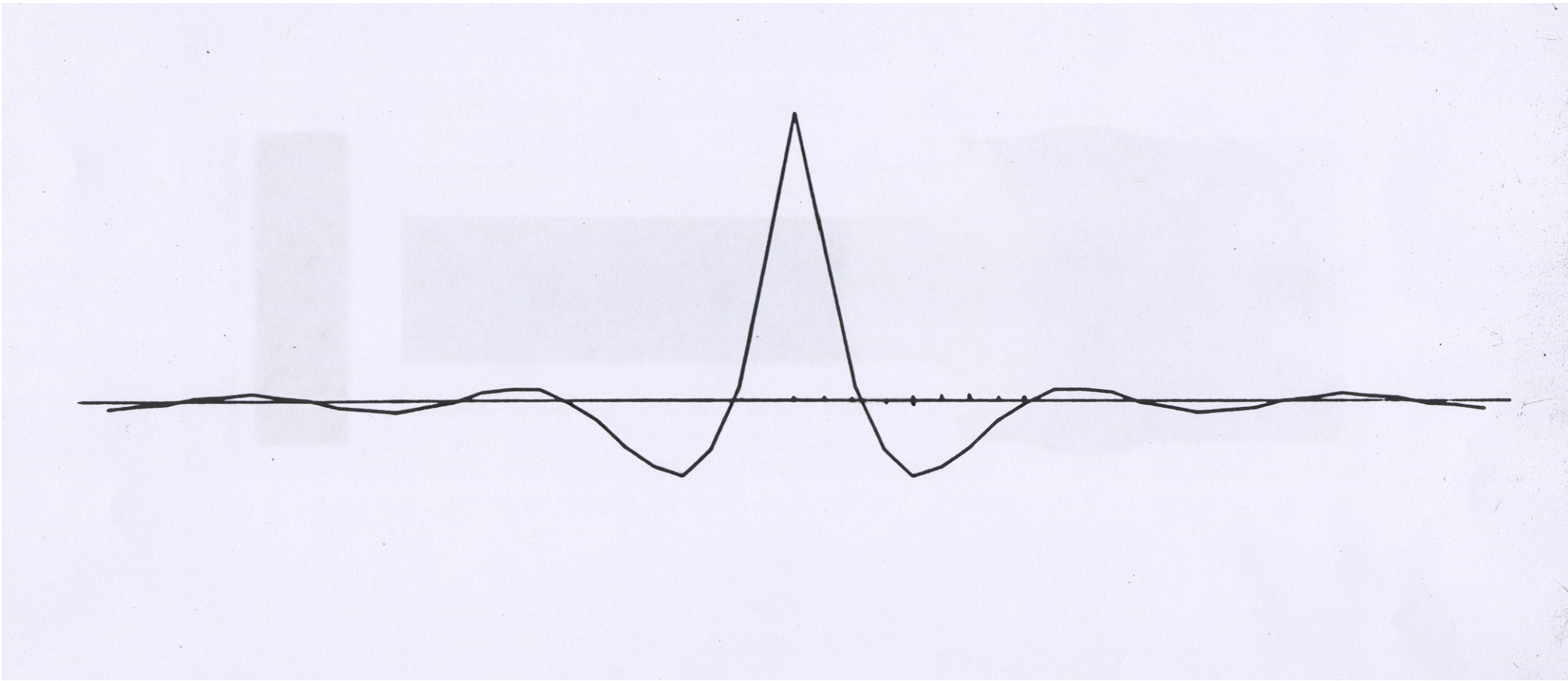
$$E(y_j' y_k') = E[\boldsymbol{\Phi}'(\xi_j) + \varepsilon_j'][\boldsymbol{\Phi}'(\xi_k) + \varepsilon_k']$$

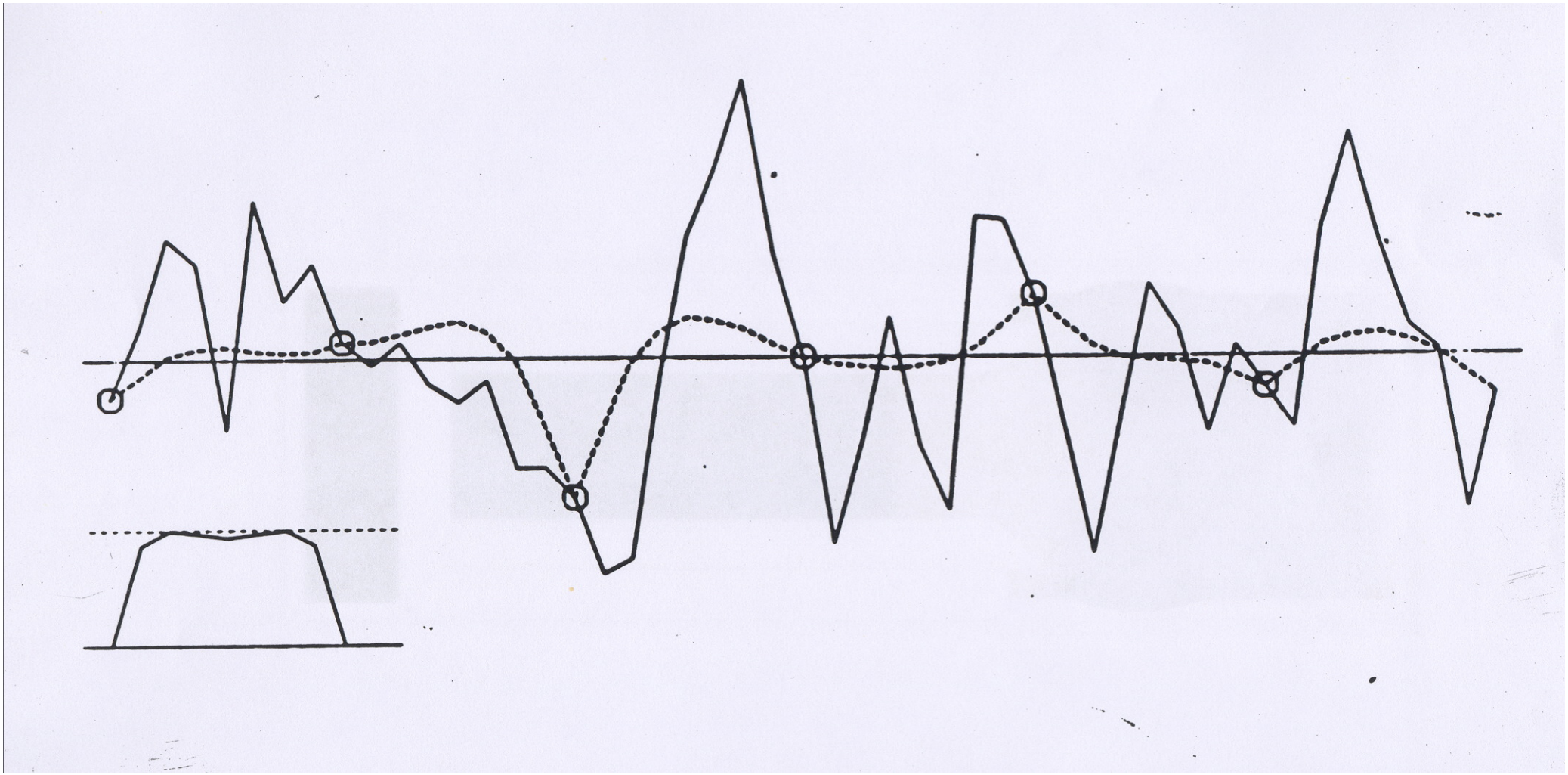
If observation errors ε_j are mutually uncorrelated, have common variance s, and are uncorrelated with field $\boldsymbol{\Phi}$, then

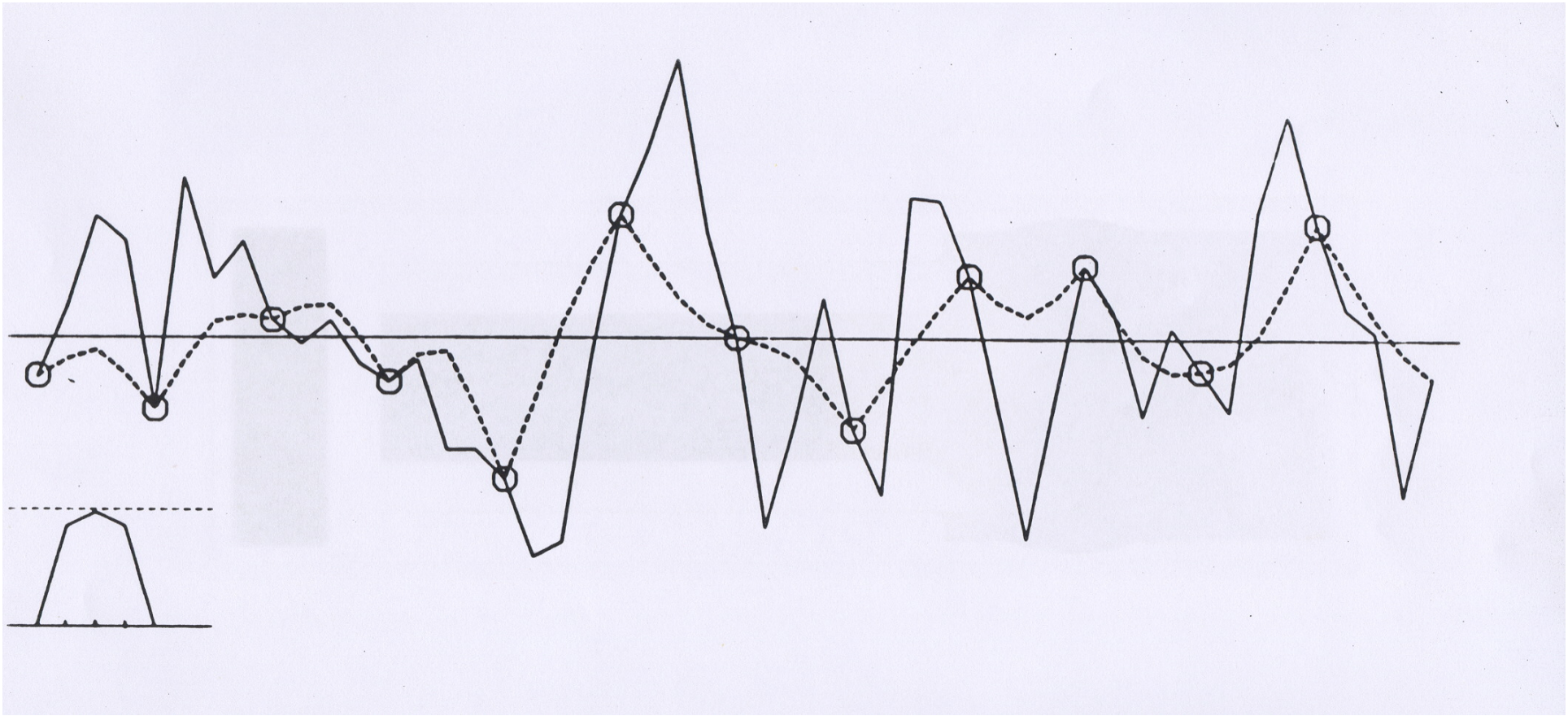
$$E(y_j' y_k') = C_{\boldsymbol{\Phi}}(\xi_j, \xi_k) + s\delta_{jk}$$

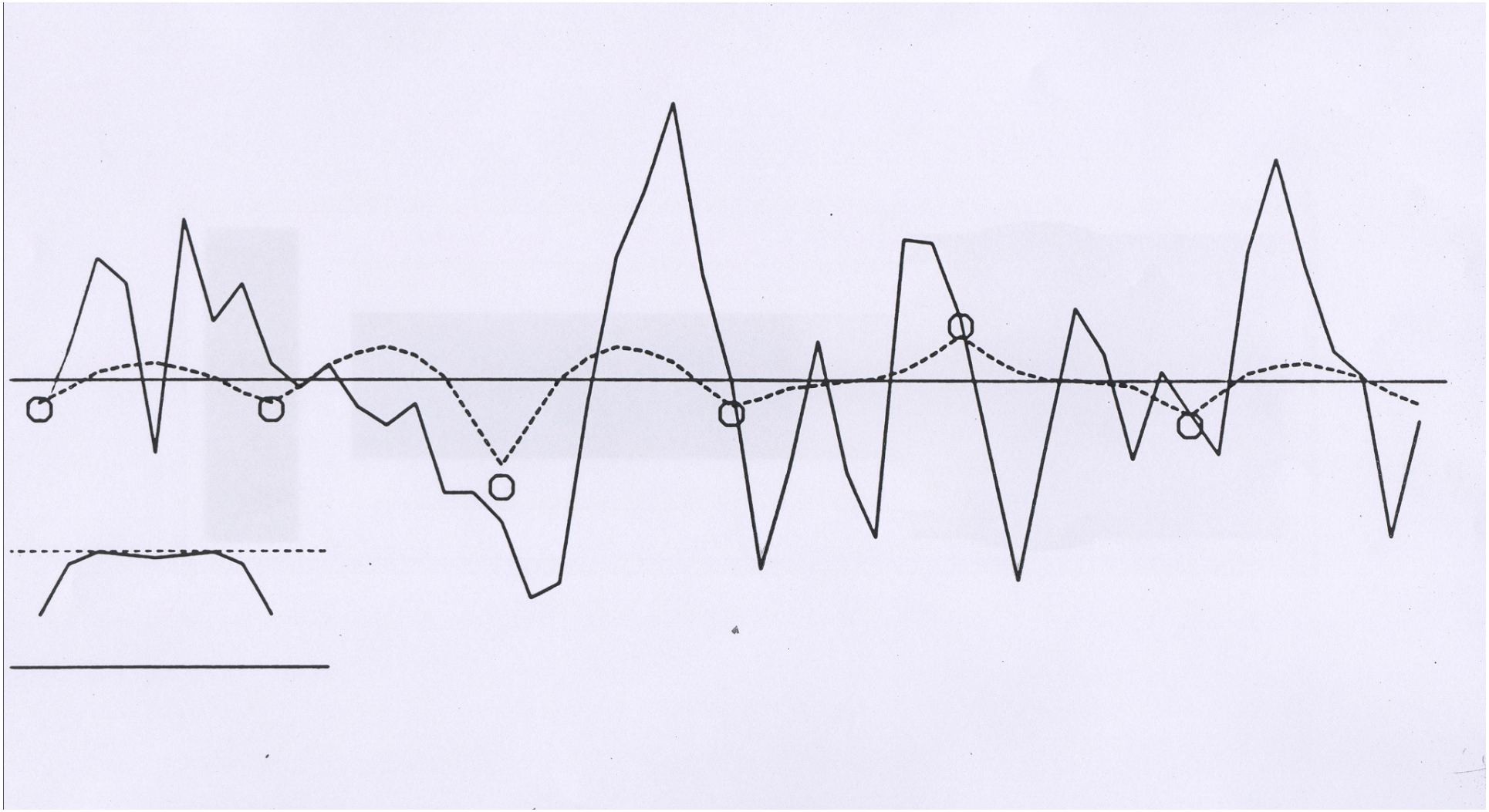
and

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









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

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

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

Correction made on background expectation is a linear combination of the p functions

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

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.

Schlatter's (1975) multivariate covariances

Specified as multivariate 2-point functions.

Not easy to ensure that specified functions are actually valid covariances.

Used in OI and related observation-space methods.

Courtesy A. Lorenc

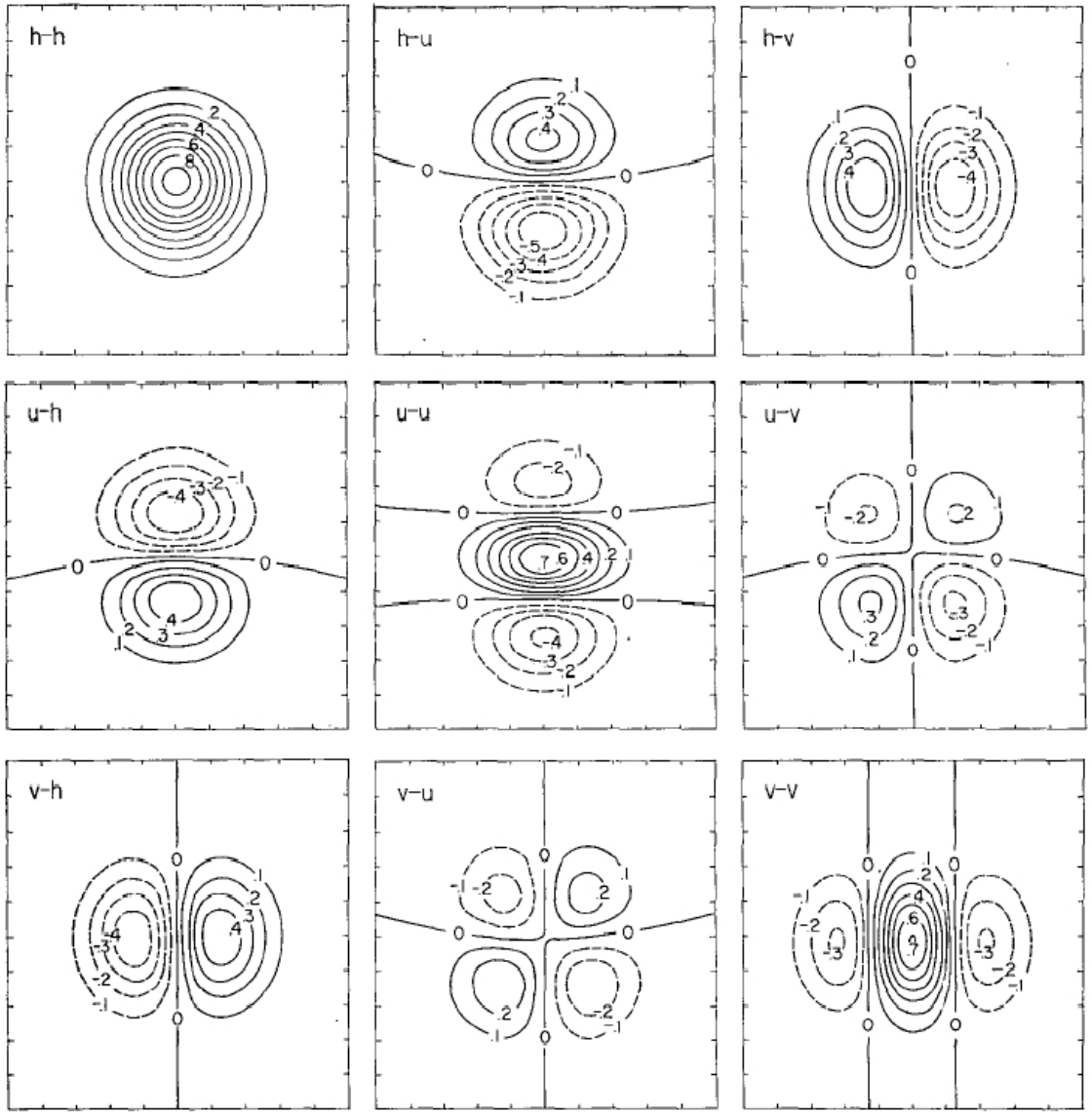


FIG. 3. Correlations among the variables h, u, and v based upon the expression $\mu = 0.95 \exp(-1.24s^2)$ for height-height correlation and the geostrophic relations. Diagrams centered at 110°W, 35°N. Tick marks 500 km apart.

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)

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

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

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

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

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

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

Best Linear Unbiased Estimate (continuation 2)

Apply formulæ for Optimal Interpolation

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

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

Equivalent set of formulæ

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

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

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

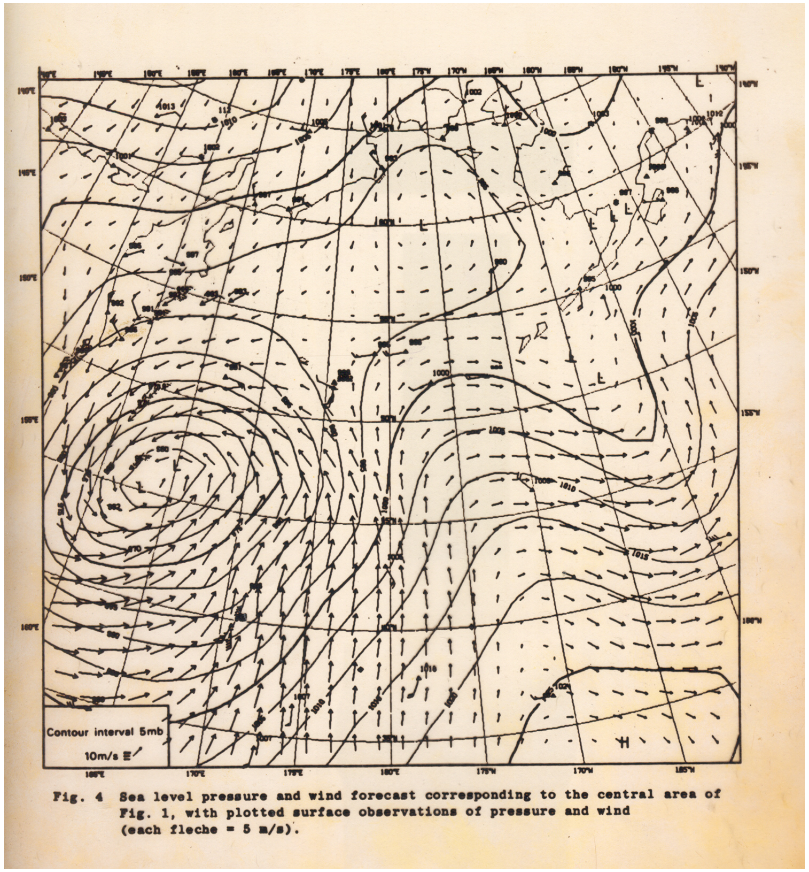


Fig. 4 Sea level pressure and wind forecast corresponding to the central area of Fig. 1, with plotted surface observations of pressure and wind (each fleche = 5 m/s).

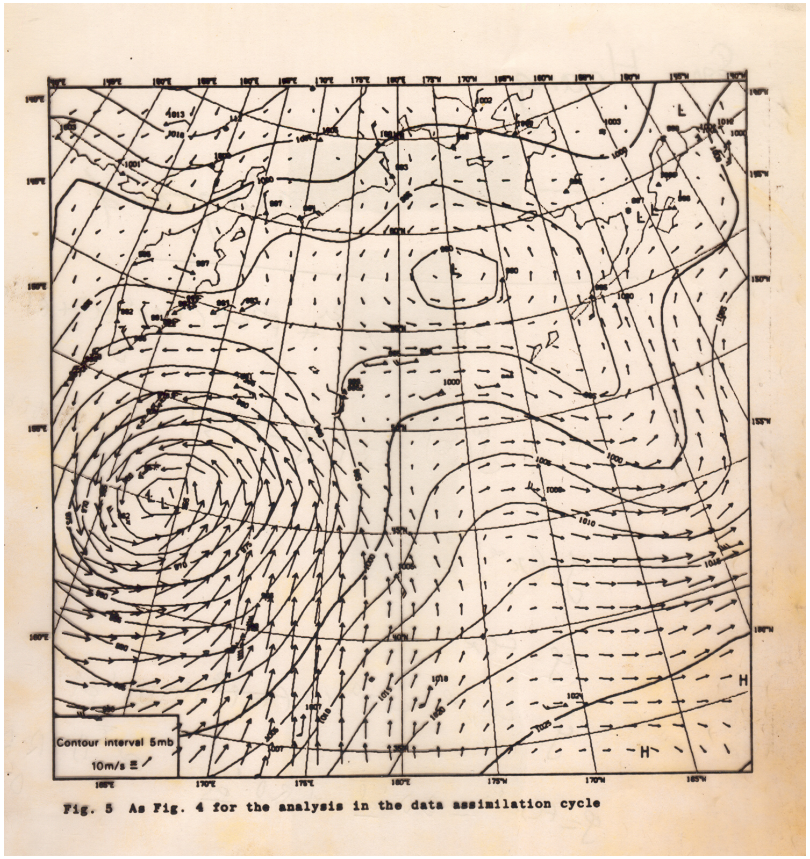


Fig. 5 As Fig. 4 for the analysis in the data assimilation cycle

After A. Lorenc

Best Linear Unbiased Estimate (continuation 4)

Variational form of the *BLUE*

BLUE x^a minimizes following scalar *objective function*, defined on state space

$\xi \in \mathcal{S} \rightarrow$

- $\mathcal{J}(\xi) = (1/2)\,(x^b - \xi)^{\mathrm{T}}\,[P^b]^{-1}\,(x^b - \xi) + (1/2)\,(y - H\xi)^{\mathrm{T}}\,R^{-1}\,(y - H\xi)$

$= \mathcal{J}_b + \mathcal{J}_o$

'*3D-Var*'

Can easily, and heuristically, be extended to the case of a nonlinear observation operator H.

Used operationally in USA, Australia, China, …

Question. How to introduce temporal dimension in estimation process ?

- Logic of Optimal Interpolation can be extended to time dimension.

- But we know much more than just temporal correlations. We know explicit dynamics.

Real (unknown) state vector at time k (in format of assimilating model) x_k. Belongs to state space $\mathcal{S}$ ($\dim \mathcal{S} = n$)

Evolution equation

$$x_{k+1} = M_k(x_k) + \eta_k$$

M_k is (known) model, η_k is (unknown) model error

Sequential Assimilation

- Assimilating model is integrated over period of time over which observations are available. Whenever model time reaches an instant at which observations are available, state predicted by the model is updated with new observations.

Variational Assimilation

- Assimilating model is globally adjusted to observations distributed over observation period. Achieved by minimization of an appropriate scalar *objective function* measuring misfit between data and sequence of model states to be estimated.

- Observation vector at time k

$$y_k = H_k x_k + \varepsilon_k \qquad k = 0, \ldots, K$$

$$E(\varepsilon_k) = 0 \quad ; \quad E(\varepsilon_k \varepsilon_j^{\mathrm{T}}) = R_k \, \delta_{kj}$$

H_k linear

- Evolution equation

$$x_{k+1} = M_k x_k + \eta_k \qquad k = 0, \ldots, K\text{-}1$$

$$E(\eta_k) = 0 \quad ; \quad E(\eta_k \eta_j^{\mathrm{T}}) = Q_k \, \delta_{kj}$$

M_k linear

- $E(\eta_k \varepsilon_j^{\mathrm{T}}) = 0$ (errors uncorrelated in time)

At time k, background x^b_k and associated error covariance matrix P^b_k known

- Analysis step

$$x^a_k = x^b_k + P^b_k H_k^{\mathrm{T}} [H_k P^b_k H_k^{\mathrm{T}} + R_k]^{-1} (y_k - H_k x^b_k)$$
$$P^a_k = P^b_k - P^b_k H_k^{\mathrm{T}} [H_k P^b_k H_k^{\mathrm{T}} + R_k]^{-1} H_k P^b_k$$

- Forecast step

$$x^b_{k+1} = M_k x^a_k$$
$$P^b_{k+1} = E[(x^b_{k+1} - x_{k+1})(x^b_{k+1} - x_{k+1})^{\mathrm{T}}] = E[(M_k x^a_k - M_k x_k - \eta_k)(M_k x^a_k - M_k x_k - \eta_k)^{\mathrm{T}}]$$
$$= M_k E[(x^a_k - x_k)(x^a_k - x_k)^{\mathrm{T}}] M_k^{\mathrm{T}} - E[\eta_k (x^a_k - x_k)^{\mathrm{T}}] - E[(x^a_k - x_k)\eta_k^{\mathrm{T}}] + E[\eta_k \eta_k^{\mathrm{T}}]$$
$$= M_k P^a_k M_k^{\mathrm{T}} + Q_k$$

At time k, background x^b_k and associated error covariance matrix P^b_k known

- Analysis step

$$x^a_k = x^b_k + P^b_k H_k^{\mathrm{T}} [H_k P^b_k H_k^{\mathrm{T}} + R_k]^{-1} (y_k - H_k x^b_k)$$
$$P^a_k = P^b_k - P^b_k H_k^{\mathrm{T}} [H_k P^b_k H_k^{\mathrm{T}} + R_k]^{-1} H_k P^b_k$$

- Forecast step

$$x^b_{k+1} = M_k x^a_k$$
$$P^b_{k+1} = M_k P^a_k M_k^{\mathrm{T}} + Q_k$$

Kalman filter (KF, Kalman, 1960)

Must be started from some initial estimate (x^b_0, P^b_0)

If all operators are linear, and if errors are uncorrelated in time, Kalman filter produces at time k the *BLUE* ${x^b}_k$ (resp. ${x^a}_k$) of the real state x_k from all data prior to (resp. up to) time k, plus the associated estimation error covariance matrix ${P^b}_k$ (resp. ${P^a}_k$).

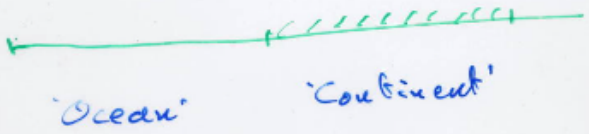
If in addition errors are gaussian, the corresponding conditional probability distributions are the respective gaussian distributions $\mathcal{N}[{x^b}_k, {P^b}_k]$ and $\mathcal{N}[{x^a}_k, {P^a}_k]$.

A didactic example (Ghil et al.)

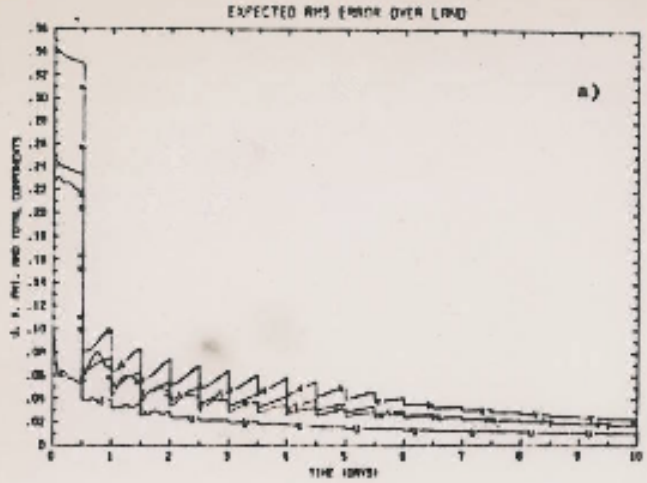
Barotropic model

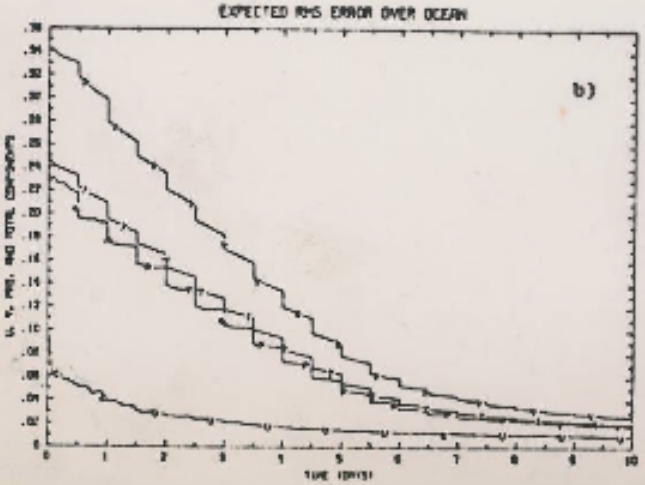
$$\begin{cases} \dfrac{\partial \varphi}{\partial t} + \operatorname{div}(\varphi \underline{U}) = 0 \\ \dfrac{\partial \underline{U}}{\partial t} + \underline{\operatorname{grad}}\left(\varphi + \tfrac{1}{2} U^2\right) + \underline{k} \times (f + \xi)\, \underline{U} = 0 \end{cases}$$

One dimension, periodic



Linearized (conserves energy)





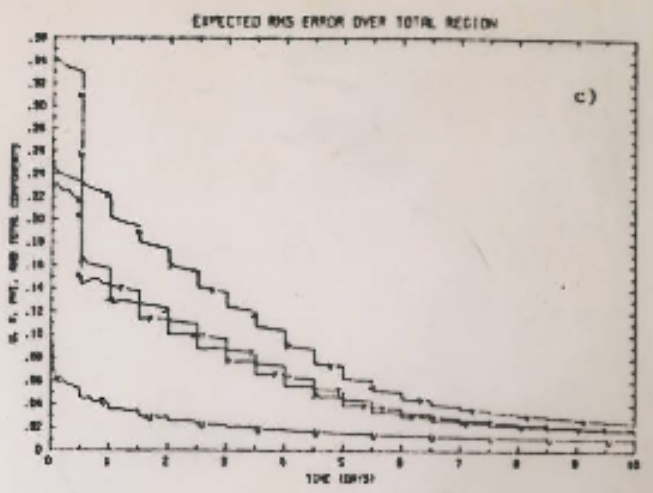


Fig. 2

The components of the total expected rms error (Erms), (trace: $P_k)^{1/2}$, in the estimation of solutions to the stochastic-dynamic system $(\bar{Y},H)$, with $\bar{Y}$ given by (3.6) and H = (I 0). System noise is absent, Q = 0. The filter used is the standard K-B filter (2.11) for the model.

a) Erms over land; b) Erms over the ocean; c) Erms over the entire L-domain

In each one of the figures, each curve represents one component of the total Erms error. The curves labelled U, V, and P represent the u component, v component and ϕ component, respectively. They are found by summing the diagonal elements of P_k which correspond to u, v, and ϕ, respectively, dividing by the number of terms in the sum, and then taking the square root. In a) the summation extends over land points only, in b) over ocean points only, and in c) over the entire L-domain. The vertical axis is scaled in such a way that 1.0 corresponds to an Erms error of v_{max} for the U and V curves, and of ϕ_0 for the P curve. The observational error level is 0.089 for the U and V curves, and 0.080 for the P curve. The curves labelled T represent the total Erms error over each region. Each T curve is a weighted average of the corresponding U, V, and P curves, with the weights chosen in such a way that the T curve measures the error in the total energy $u^2 + v^2 + \phi^2/\gamma$, conserved by the system (3.1). The observational noise level for the T curve is then 0.088. Notice the immediate error decrease over land and the gradual decrease over the ocean. The total estimation error tends to zero.

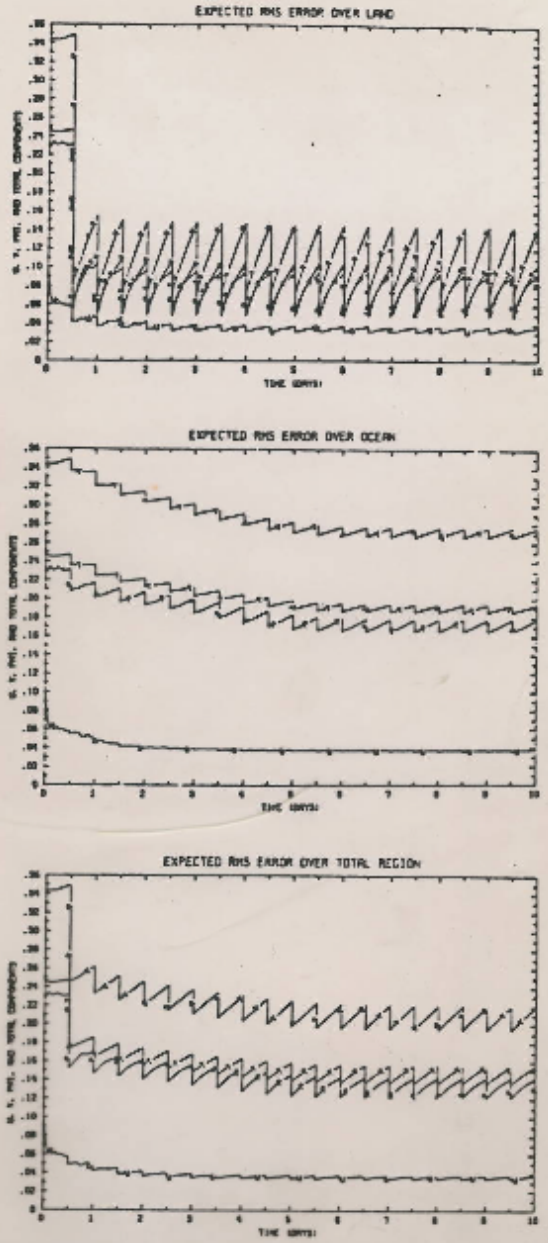


Fig. 6 This figure and the following ones show the properties of the estimated algorithm (2.11) in the presence of system noise, $Q \neq 0$. This figure gives the Erms estimation error, and is homologous to Fig. 2. Notice the sharper increase of error over land between synoptic times, and the convergence of each curve to a periodic, nonzero function.

Nonlinearities ?

Model is usually nonlinear, and observation operators (satellite observations) tend more and more to be nonlinear.

- Analysis step

$$x^a_k = x^b_k + P^b_k H_k{}'^{\mathrm{T}} [H_k' P^b_k H_k{}'^{\mathrm{T}} + R_k]^{-1} [y_k - H_k(x^b_k)]$$
$$P^a_k = P^b_k - P^b_k H_k{}'^{\mathrm{T}} [H_k' P^b_k H_k{}'^{\mathrm{T}} + R_k]^{-1} H_k' P^b_k$$

- Forecast step

$$x^b_{k+1} = M_k(x^a_k)$$
$$P^b_{k+1} = M_k' P^a_k M_k{}'^{\mathrm{T}} + Q_k$$

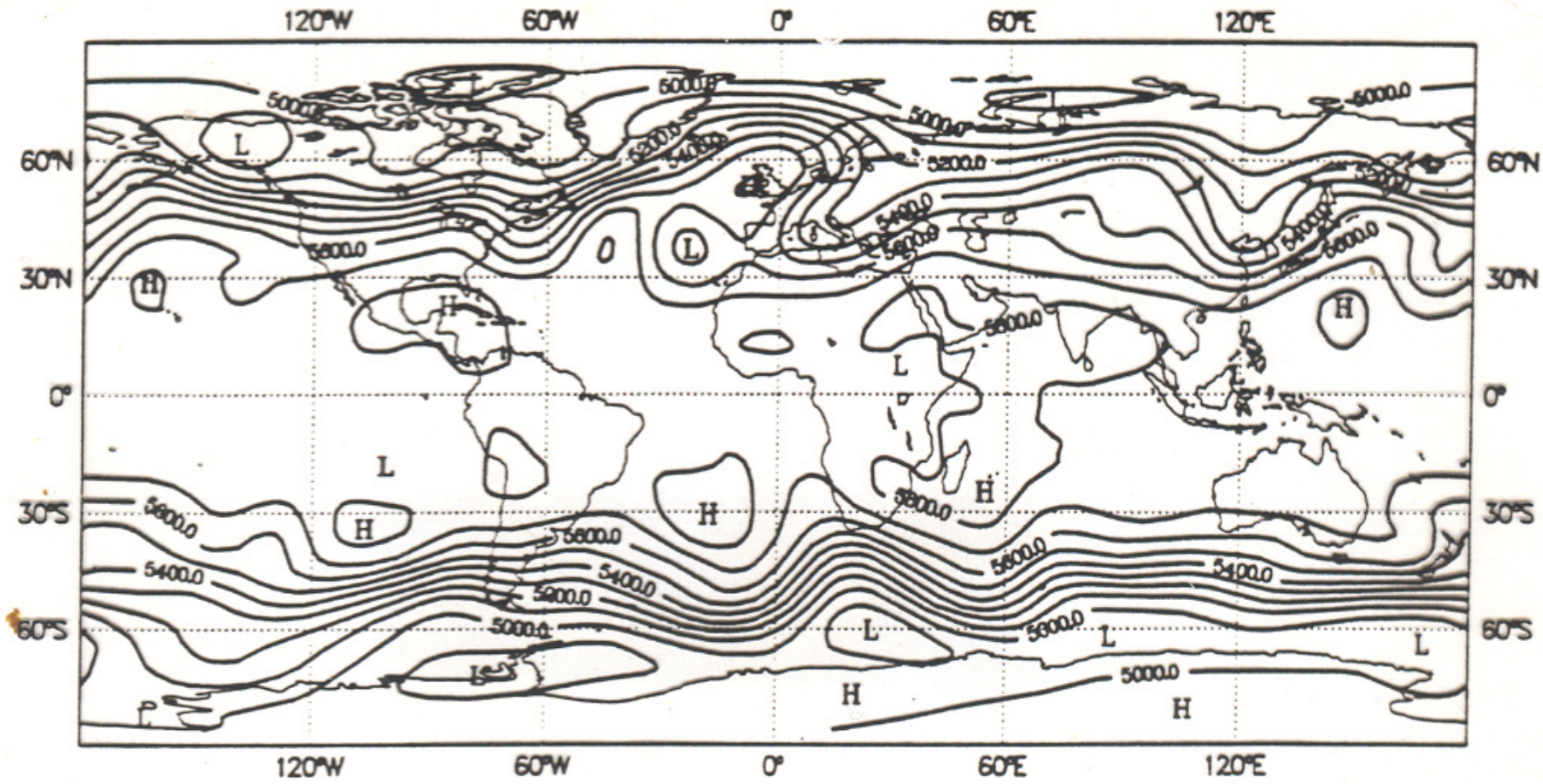
Extended Kalman Filter (EKF, heuristic !)

Costliest part of computation

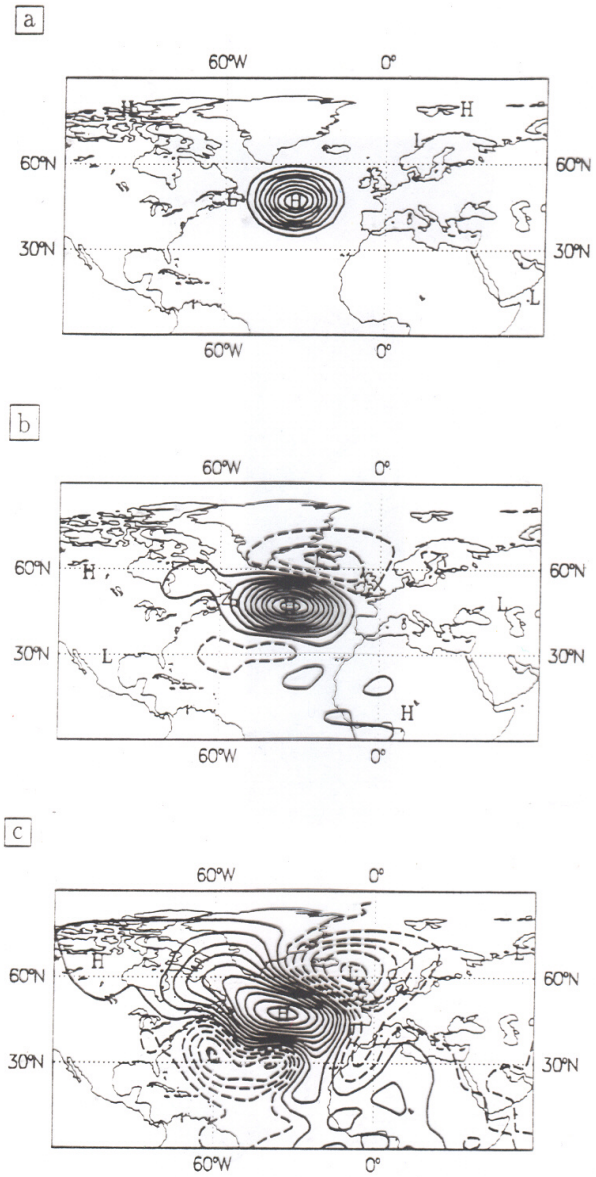
$$P^{b}_{k+1} = M_k P^{a}_{k} M_k^{\mathrm{T}} + Q_k$$

Multiplication by M_k = one integration of the model between times k and $k+1$.
Computation of $M_k P^{a}_{k} M_k^{\mathrm{T}} \approx 2n$ integrations of the model

Need for determining the temporal evolution of the uncertainty on the state of the system is the major difficulty in assimilation of meteorological and oceanographical observations



Analysis of 500-hPa geopotential for 1 December 1989, 00:00 UTC
(ECMWF, spectral truncation T21, unit *m*. After F. Bouttier)



Temporal evolution of the 500-hPa geopotential autocorrelation with respect to point located at 45N, 35W. From top to bottom: initial time, 6- and 24-hour range. Contour interval 0.1. After F. Bouttier.

Two solutions :

- Low-rank filters (Heemink, Pham, …)

  Reduced Rank Square Root Filters, Singular Evolutive Extended Kalman Filter, ….

- Ensemble filters (Evensen, Anderson, Kalnay, …)

  Uncertainty is represented, not by a covariance matrix, but by an ensemble of point estimates in state space which are meant to sample the conditional probability distribution for the state of the system (dimension $N \approx O(10\text{-}100)$).

  Ensemble is evolved in time through the full model, which eliminates any need for linear hypothesis as to the temporal evolution.

How to update predicted ensemble with new observations ?

Predicted ensemble at time t : $\{x^b{}_i\}$, $\quad i = 1, \ldots, N$
Observation vector at same time : $y = Hx + \varepsilon$

- Gaussian approach

  Produce sample of probability distribution for real observed quantity Hx
  $y_i = y - \varepsilon_i$
  where ε_i is distributed according to probability distribution for observation error ε.

  Then use Kalman formula to produce sample of 'analysed' states

  $$x^a{}_i = x^b{}_i + P^b H^{\mathrm{T}} [HP^b H^{\mathrm{T}} + R]^{-1} (y_i - Hx^b{}_i) , \qquad i = 1, \ldots, N \qquad (2)$$

  where P^b is covariance matrix of predicted ensemble $\{x^b{}_i\}$.

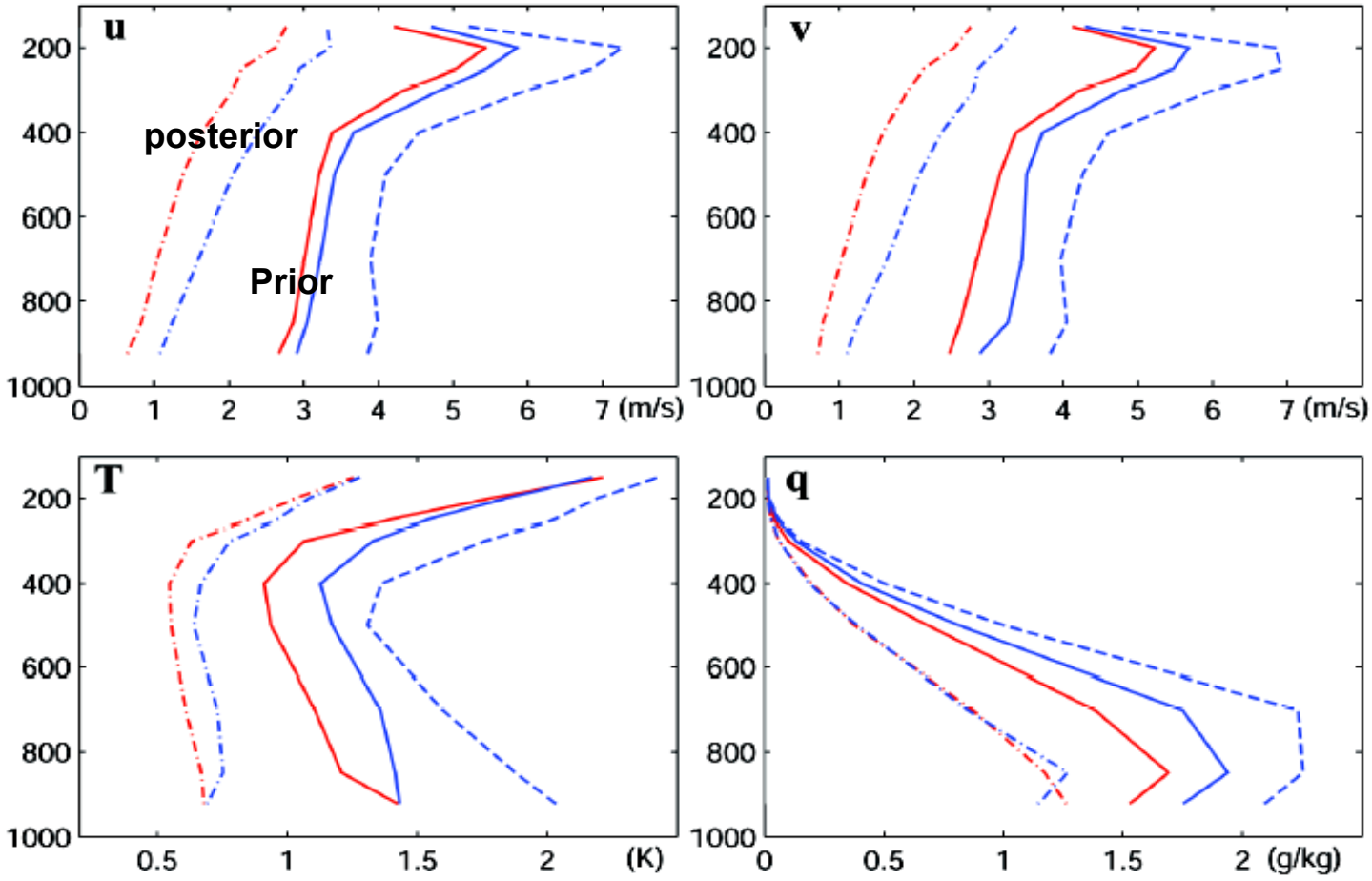
  Remark. If P^b was exact covariance matrix of background error, (2) would achieve Bayesian estimation, in the sense that $\{x^a{}_i\}$ would be a sample of conditional probability distribution for x, given all data up to time t.

Called *Ensemble Kalman Filter*. Has now become one of the two major powerful algorithms for assimilation of meteorological and oceanographical observations.

Local Ensemble Transform Kalman Filter (*LETKF,* Kalnay and colleagues)

Month-long Performance of EnKF vs. 3Dvar with WRF

— EnKF —3DVar (prior, solid; posterior, dotted)



Better performance of EnKF than 3DVar also seen in both 12-h forecast and posterior analysis in terms of root-mean square difference averaged over the entire month

(Meng and Zhang 2007c, MWR, in review)

Variational Assimilation

- Observation vector at time k

$$y_k = H_k x_k + \varepsilon_k \qquad k = 0, \ldots, K$$
$$E(\varepsilon_k) = 0 \;\; ; \;\; E(\varepsilon_k \varepsilon_j^{\mathrm{T}}) \equiv R_k \, \delta_{kj}$$

- Evolution equation

$$x_{k+1} = M_k x_k + \eta_k \qquad k = 0, \ldots, K\text{-}1$$
$$E(\eta_k) = 0 \;\; ; \;\; E(\eta_k \eta_j^{\mathrm{T}}) \equiv Q_k \, \delta_{kj}$$
$$E(\eta_k \varepsilon_j^{\mathrm{T}}) = 0$$

- Background estimate at time 0

$$x^b{}_0 = x_0 + \zeta^b{}_0$$
$$E(\zeta^b{}_0) = 0 \;\; ; \;\; E(\zeta^b{}_0 \, \zeta^b{}_0{}^{\mathrm{T}}) \equiv P^b{}_0$$
$$E(\zeta^b{}_0 \varepsilon_k^{\mathrm{T}}) = 0 \;\; ; \;\; E(\zeta^b{}_0 \eta_k^{\mathrm{T}}) = 0$$

- Errors uncorrelated in time

Variational assimilation leads to the following *weak constraint* objective function

$(\xi_0, \xi_1, .., \xi_K) \rightarrow$

- $\mathcal{J}(\xi_0, \xi_1, ..., \xi_K)$

$$= (1/2)\,(x^b_0 - \xi_0)^{\mathrm{T}}\,[P^b_0]^{-1}\,(x^b_0 - \xi_0)$$

$$+ (1/2)\,\Sigma_{k=0,\ldots,K}\,(y_k - H_k\xi_k)^{\mathrm{T}}\,R_k^{-1}\,(y_k - H_k\xi_k)$$

$$+ (1/2)\,\Sigma_{k=0,\ldots,K-1}\,(\xi_{k+1} - M_k\xi_k)^{\mathrm{T}}\,Q_k^{-1}\,(\xi_{k+1} - M_k\xi_k)$$

If model error is ignored (Q_k=0), problem reduces to minimizing

- $\xi_0 \rightarrow \mathcal{J}(\xi_0) = (1/2)\,(x^b_0 - \xi_0)^{\mathrm{T}}\,[P^b_0]^{-1}\,(x^b_0 - \xi_0)$
  $+ (1/2)\,\Sigma_{k=0,\ldots,K}\,(y_k - H_k\xi_k)^{\mathrm{T}}\,R_k^{-1}\,(y_k - H_k\xi_k)$

  subject to

  $\xi_{k+1} = M_k\xi_k \qquad , \qquad k = 0, \ldots, K\text{-}1$

  Strong constraint four-dimensional variational assimilation, or strong constraint *4D-Var*

  Used operationally in several meteorological centres (Météo-France, UK Meteorological Office, Canadian Meteorological Centre (maybe not any more ?), Japan Meteorological Agency, …) and, until recently, at ECMWF. The latter now has a ‘weak constraint’ component in its operational system.

$$\mathcal{J}(\xi_0) \;=\; (1/2)\,(x_0{}^b - \xi_0)^{\mathrm{T}}\,[P_0{}^b]^{-1}\,(x_0{}^b - \xi_0) + (1/2)\,\Sigma_k[y_k - H_k\xi_k]^{\mathrm{T}}\,R_k^{-1}\,[y_k - H_k\xi_k]$$

Background is not necessary, if observations are in sufficient number to overdetermine the problem. Nor is strict linearity.

Minimization achieved by iterative algorithm, each step of which requires the explicit knowledge of the local gradient $\nabla_u \mathcal{J} \equiv (\partial\mathcal{J}/\partial u_i)$ of $\mathcal{J}$ with respect to u.

Gradient computed by *adjoint method*, which proceeds, in the space of partial derivatives, in reverse order of direct computations.

How to numerically compute the gradient $\nabla_u \mathcal{J}$?

Direct perturbation, in order to obtain partial derivatives $\partial \mathcal{J} / \partial u_i$ by finite differences ? That would require as many explicit computations of the objective function $\mathcal{J}$ as there are components in u. Practically impossible.

Adjoint Method

Input vector $\boldsymbol{u} = (u_i)$, $\dim \boldsymbol{u} = n$

Numerical process, implemented on computer (*e. g.* integration of numerical model)

$$\boldsymbol{u} \rightarrow \boldsymbol{v} = \boldsymbol{G}(\boldsymbol{u})$$

- $\boldsymbol{v} = (v_j)$ is *output vector* , $\dim \boldsymbol{v} = m$
- Perturbation $\delta\boldsymbol{u} = (\delta u_i)$ of input. Resulting first-order perturbation on $\boldsymbol{v}$
  - $\delta v_j = \Sigma_i (\partial v_j / \partial u_i)\ \delta u_i$
- or, in matrix form
  - $\delta \boldsymbol{v} = \boldsymbol{G}' \delta \boldsymbol{u}$
- where $\boldsymbol{G}' \equiv (\partial v_j / \partial u_i)$ is local matrix of partial derivatives, or jacobian matrix, of $\boldsymbol{G}$.

Adjoint Method (continued 1)

$$\delta \boldsymbol{v} \;=\; \boldsymbol{G}'\,\delta \boldsymbol{u} \qquad \text{(D)}$$

- Scalar function of output

$$\mathcal{J}(\boldsymbol{v}) \;=\; \mathcal{J}[\boldsymbol{G}(\boldsymbol{u})]$$

Gradient $\nabla_{\boldsymbol{u}} \mathcal{J}$ of $\mathcal{J}$ with respect to input $\boldsymbol{u}$?

'Chain rule'

$$\partial \mathcal{J}/\partial u_i = \Sigma_j \, \partial \mathcal{J}/\partial v_j \, (\partial v_j/\partial u_i)$$

or

- $$\nabla_{\boldsymbol{u}} \mathcal{J} \;=\; \boldsymbol{G}'^{\mathrm{T}} \nabla_{\boldsymbol{v}} \mathcal{J} \qquad \text{(A)}$$

Adjoint Method (continued 2)

- $\boldsymbol{G}$ is the composition of a number of successive steps

$$\boldsymbol{G} = \boldsymbol{G}_N \circ \ldots \circ \boldsymbol{G}_2 \circ \boldsymbol{G}_1$$

'Chain rule'

$$\boldsymbol{G}' = \boldsymbol{G}_N' \ldots \boldsymbol{G}_2' \boldsymbol{G}_1'$$

Transpose

- $$\boldsymbol{G}'^{\mathrm{T}} = \boldsymbol{G}_1'^{\mathrm{T}} \boldsymbol{G}_2'^{\mathrm{T}} \ldots \boldsymbol{G}_N'^{\mathrm{T}}$$

Transpose, or *adjoint*, computations are performed in reversed order of direct computations.

If $\boldsymbol{G}$ is nonlinear, local jacobian $\boldsymbol{G}'$ depends on local value of input $\boldsymbol{u}$. Any quantity which is an argument of a nonlinear operation in the direct computation will be used again in the adjoint computation. It must be kept in memory from the direct computation (or else be recomputed again in the course of the adjoint computation).

If everything is kept in memory, total operation count of adjoint computation is at most 4 times operation count of direct computation (in practice about 2).

Adjoint Approach

$$\mathcal{J}(\xi_0) = (1/2)\,(x_0^b - \xi_0)^{\mathrm{T}}\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_k [y_k - H_k\xi_k]^{\mathrm{T}}\,R_k^{-1}\,[y_k - H_k\xi_k]$$

subject to $\xi_{k+1} = M_k\xi_k\,, \qquad k = 0, \ldots, K\text{-}1$

Control variable $\xi_0 = \boldsymbol{u}$

Adjoint equation

$$\lambda_K = H_K^{\mathrm{T}}\,R_K^{-1}\,[H_K\,\xi_K - y_K]$$

$$\lambda_k = M_k^{\mathrm{T}}\lambda_{k+1} + H_k^{\mathrm{T}}\,R_k^{-1}\,[H_k\,\xi_k - y_k] \qquad k = K\text{-}1, \ldots, 1$$

$$\lambda_0 = M_0^{\mathrm{T}}\lambda_1 + H_0^{\mathrm{T}}\,R_0^{-1}\,[H_0\,\xi_0 - y_0] + [P_0^b]^{-1}\,(\xi_0 - x_0^b)$$

$$\nabla_u \mathcal{J} = \lambda_0$$

Result of direct integration (ξ_k), which appears in quadratic terms in expression of objective function, must be kept in memory from direct integration.

Adjoint Approach (continued 2)

Nonlinearities ?

$$\mathcal{J}(\xi_0) = (1/2)\,(x_0^b - \xi_0)^{\mathrm{T}}\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_k [y_k - H_k(\xi_k)]^{\mathrm{T}}\,R_k^{-1}\,[y_k - H_k(\xi_k)]$$

subject to $\xi_{k+1} = M_k(\xi_k)\,, \qquad k = 0, \ldots, K\text{-}1$

Control variable $\xi_0 = \boldsymbol{u}$

Adjoint equation

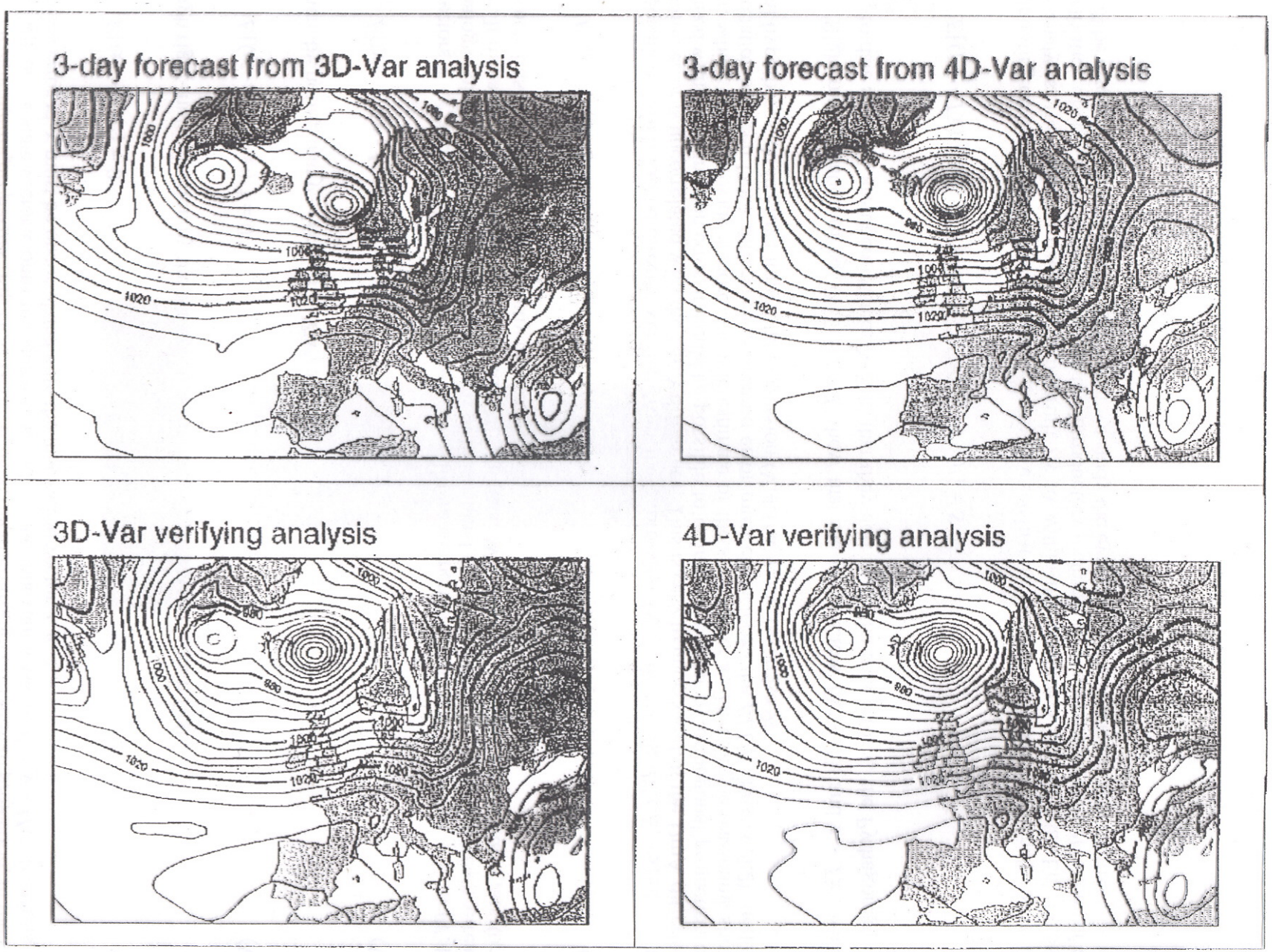
$$\lambda_K = H_K{}'^{\mathrm{T}}\,R_K^{-1}\,[H_K(\xi_K) - y_K]$$

$$\lambda_k = M_k{}'^{\mathrm{T}}\lambda_{k+1} + H_k{}'^{\mathrm{T}}\,R_k^{-1}\,[H_k(\xi_k) - y_k] \qquad k = K\text{-}1, \ldots, 1$$

$$\lambda_0 = M_0{}'^{\mathrm{T}}\lambda_1 + H_0{}'^{\mathrm{T}}\,R_0^{-1}\,[H_0(\xi_0) - y_0] + [P_0^b]^{-1}\,(\xi_0 - x_0^b)$$

$$\nabla_u \mathcal{J} = \lambda_0$$

Not heuristic (it gives the exact gradient $\nabla_u \mathcal{J}$), and really used as described here.



ECMWF, Results on one FASTEX case (1997)

Buehner (2008)

For the same numerical cost, and in meteorologically realistic situations, Ensemble Kalman Filter and Variational Assimilation produce results of similar quality.

Conclusions

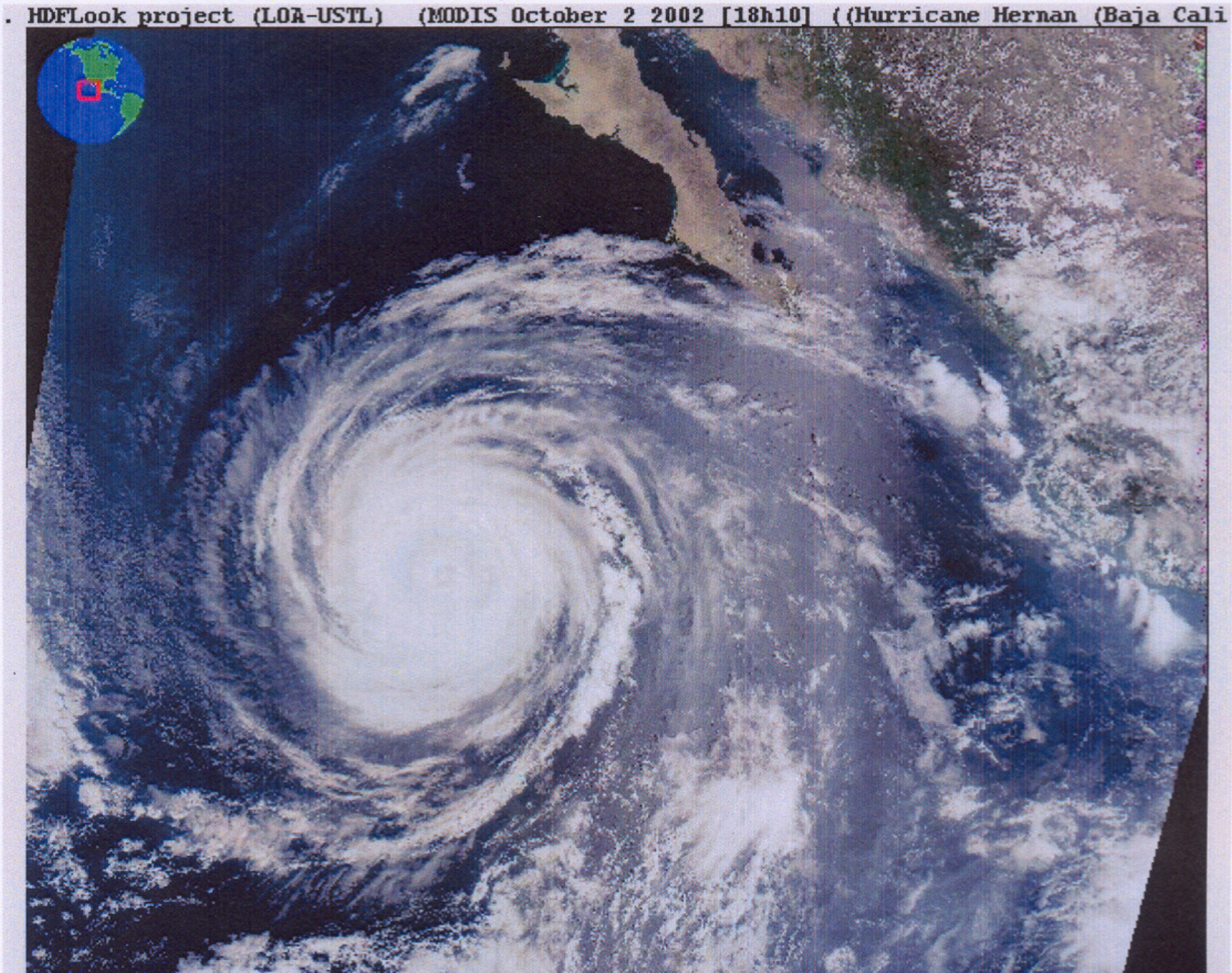
Assimilation, which originated from the need of defining initial conditions for numerical weather forecasts, has progressively extended to many diverse applications

- Oceanography
- Atmospheric chemistry (both troposphere and stratosphere)
- Oceanic biogeochemistry
- Ground hydrology
- Terrestrial biosphere and vegetation cover
- Glaciology
- Magnetism (both planetary and stellar)
- Plate tectonics
- Planetary atmospheres (Mars, …)
- Reassimilation of past observations (mostly for climatological purposes, ECMWF, NCEP/NCAR)
- Identification of source of tracers
- Parameter identification
- *A priori* evaluation of anticipated new instruments
- Definition of observing systems *(Observing Systems Simulation Experiments)*
- Validation of models
- Sensitivity studies (adjoints)
- …

It has now become a major tool of numerical environmental science

Assimilation is related to

- Estimation theory
- Probability theory
- Atmospheric and oceanic dynamics
- Atmospheric and oceanic predictability
- Instrumental physics
- Optimisation theory
- Control theory
- Algorithmics and computer science
- …


. HDFLook project (LOA-USTL) (MODIS October 2 2002 [18h10] ((Hurricane Hernan (Baja Cali