

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

Année Universitaire 2016-2017

Modélisation Numérique
de l'Écoulement Atmosphérique
et Assimilation de Données

Olivier Talagrand

Cours 3

20 Avril 2017

Physical laws governing the flow

- Conservation of mass

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

- Conservation of energy

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

- Conservation of momentum

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

- Equation of state

$$f(p, \rho, e) = 0 \qquad (p/\rho = rT, e = C_v T)$$

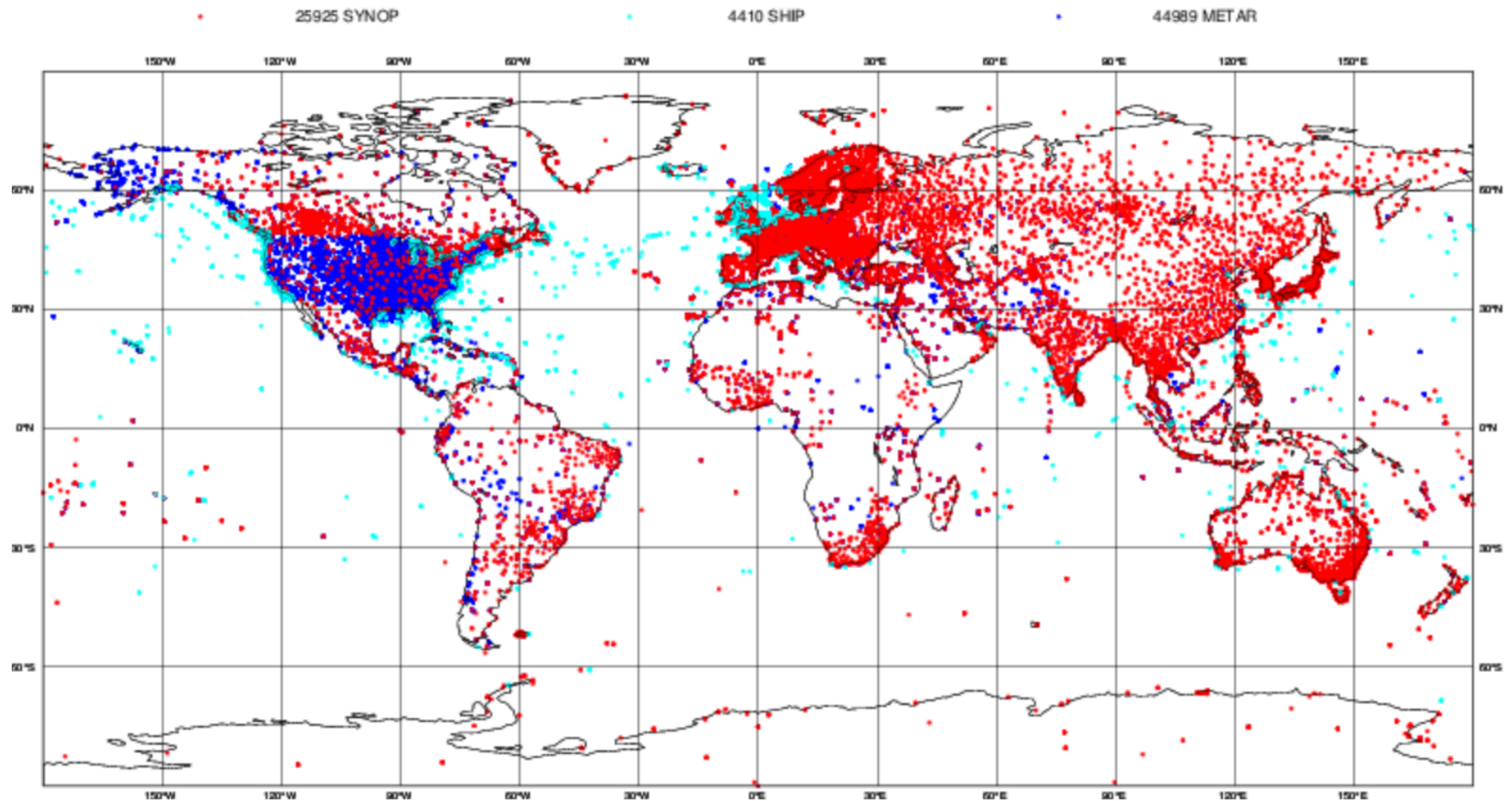
- Conservation of mass of secondary components (water in the atmosphere, salt in the ocean, chemical species, ...)

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

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

- Le système d'observation météorologique
- Assimilation. Les bases de l'estimation statistique

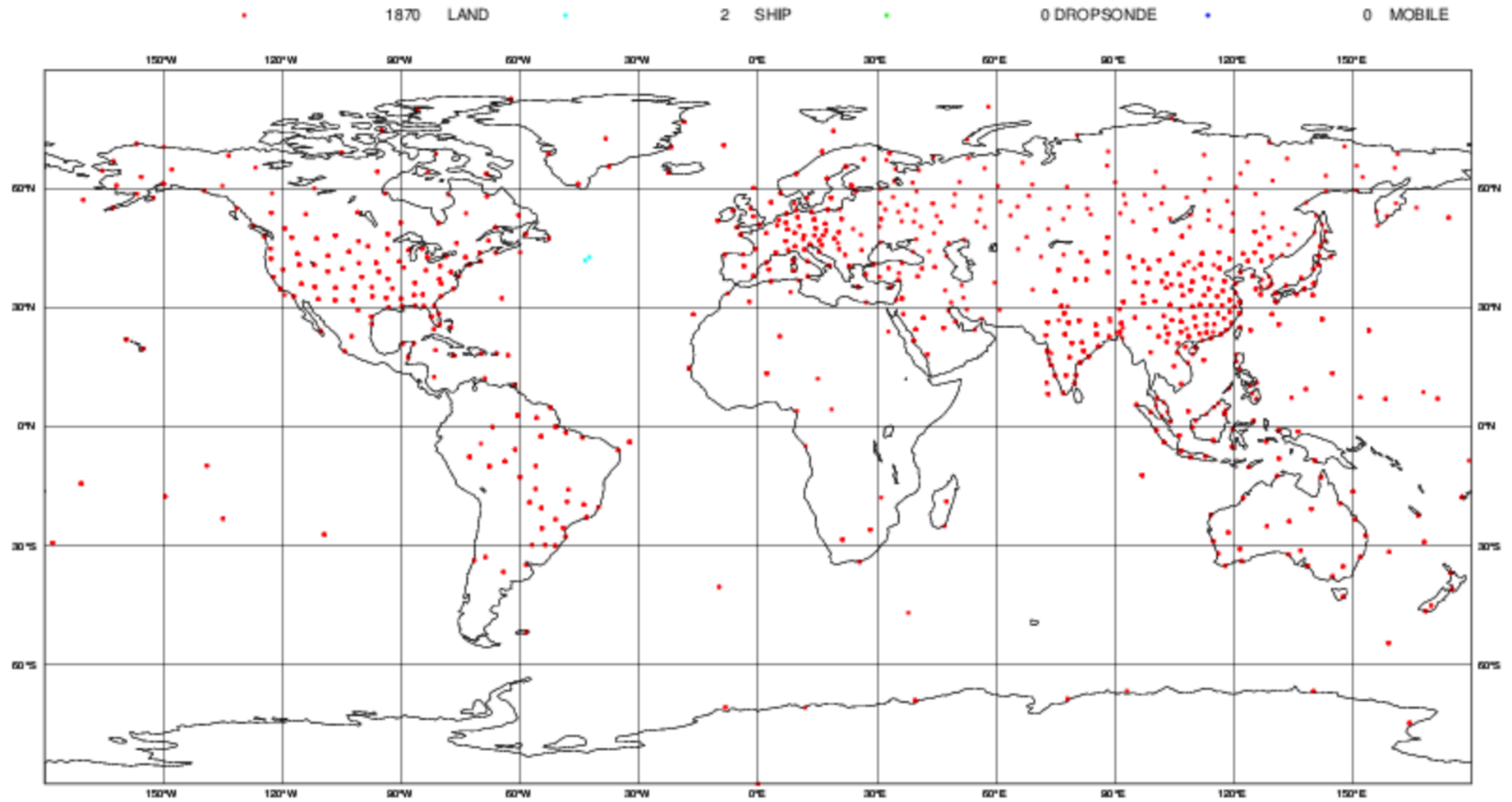
ECMWF Data Coverage (All obs DA) - Synop-Ship-Metar
19/Apr/2017; 00 UTC
Total number of obs = 75324



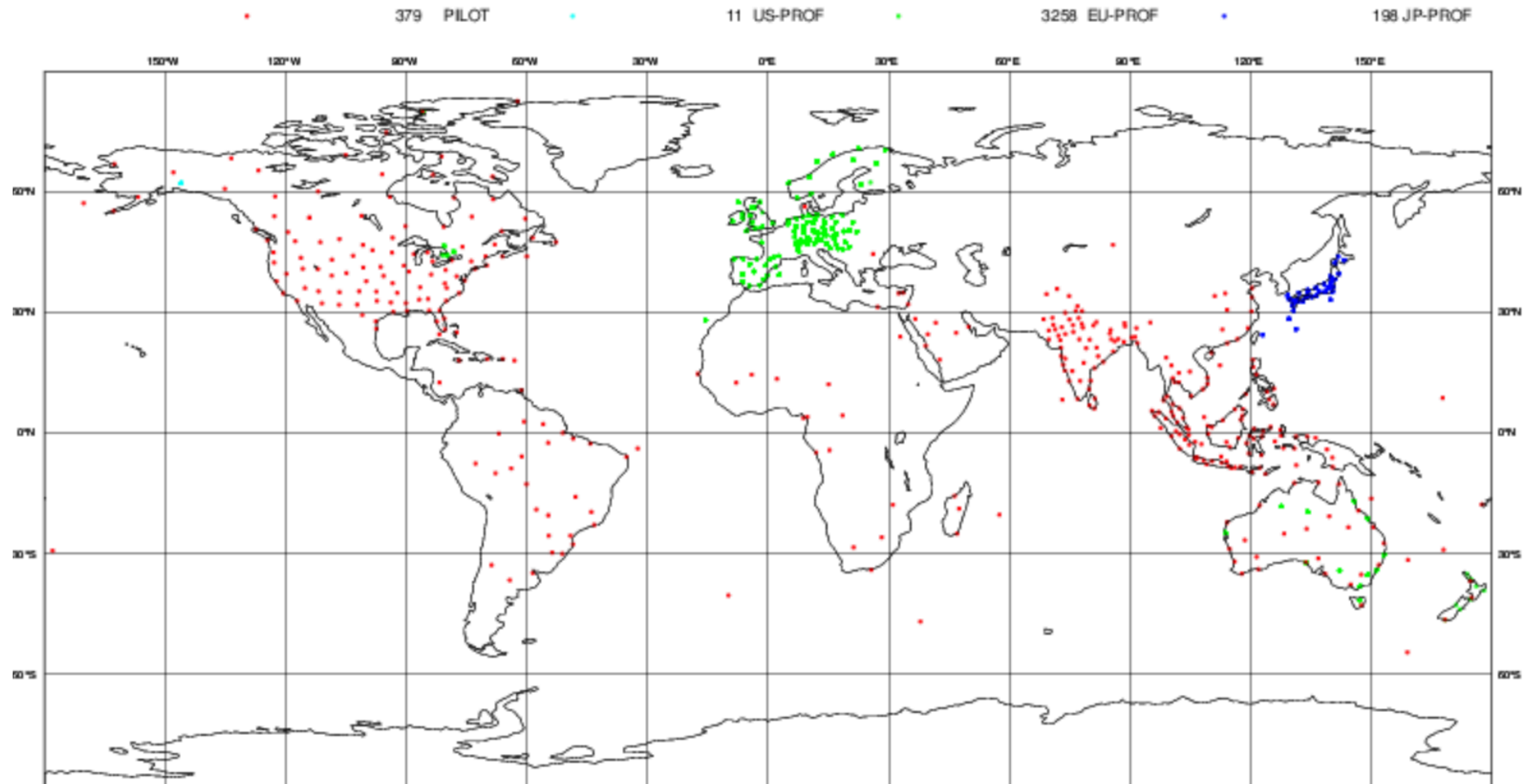
ECMWF Data Coverage (All obs DA) - Temp

19/Apr/2017; 00 UTC

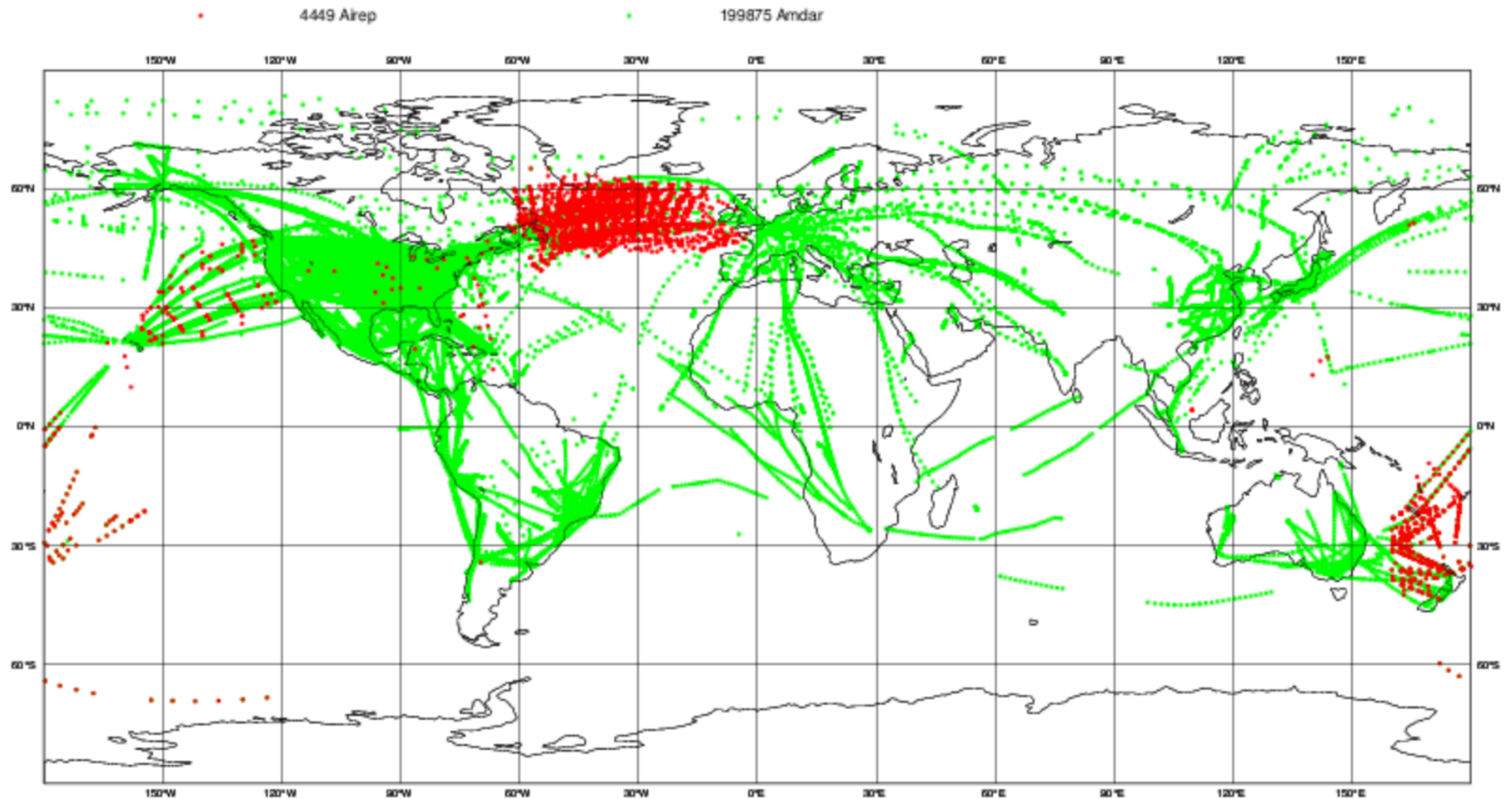
Total number of obs = 1872



ECMWF Data Coverage (All obs DA) - Pilot-Profiler
19/Apr/2017; 00 UTC
Total number of obs = 3846



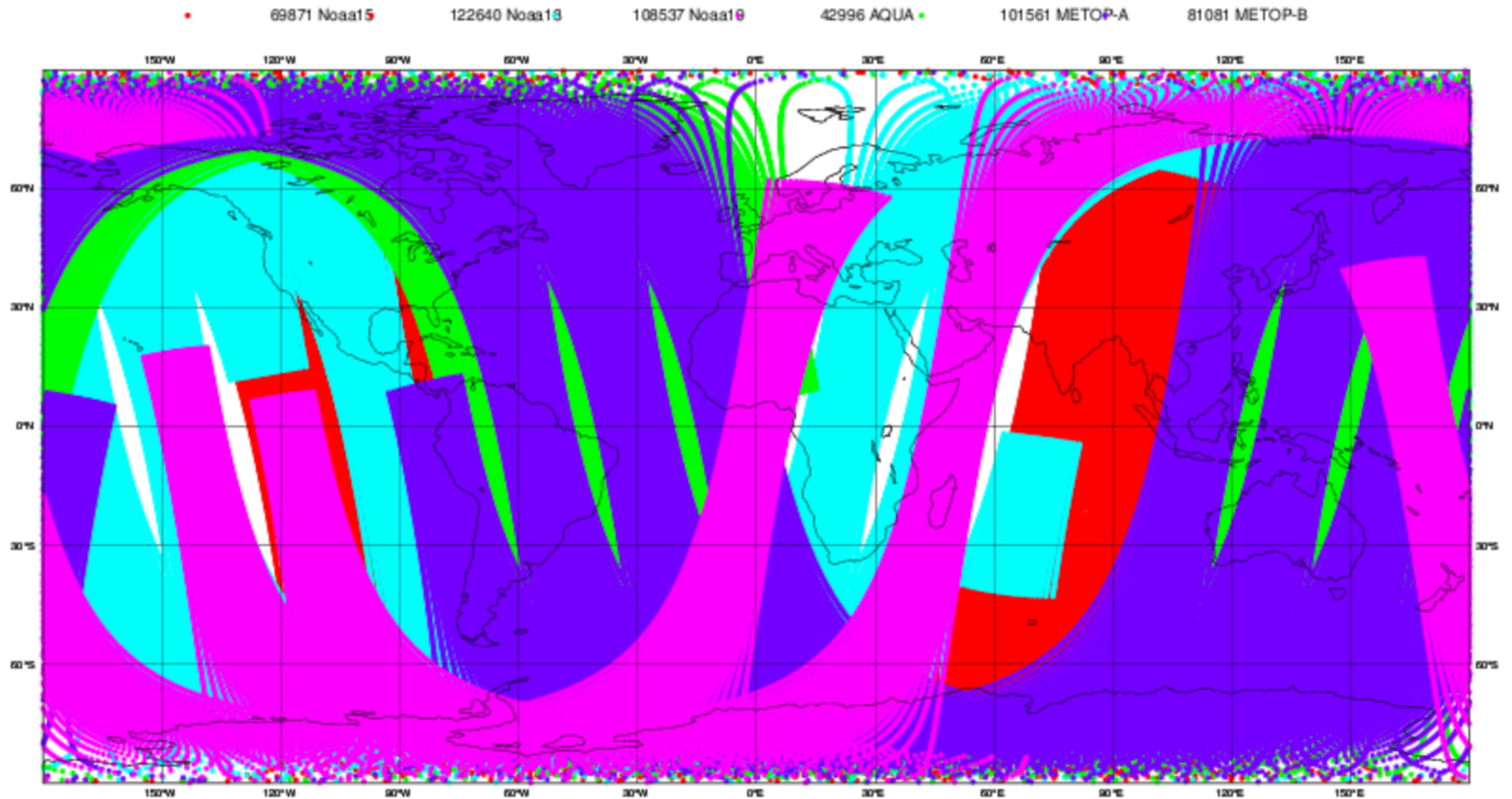
ECMWF Data Coverage (All obs DA) - Aircraft
19/Apr/2017; 00 UTC
Total number of obs = 204324



ECMWF Data Coverage (All obs DA) - AMSU-A

19/Apr/2017; 00 UTC

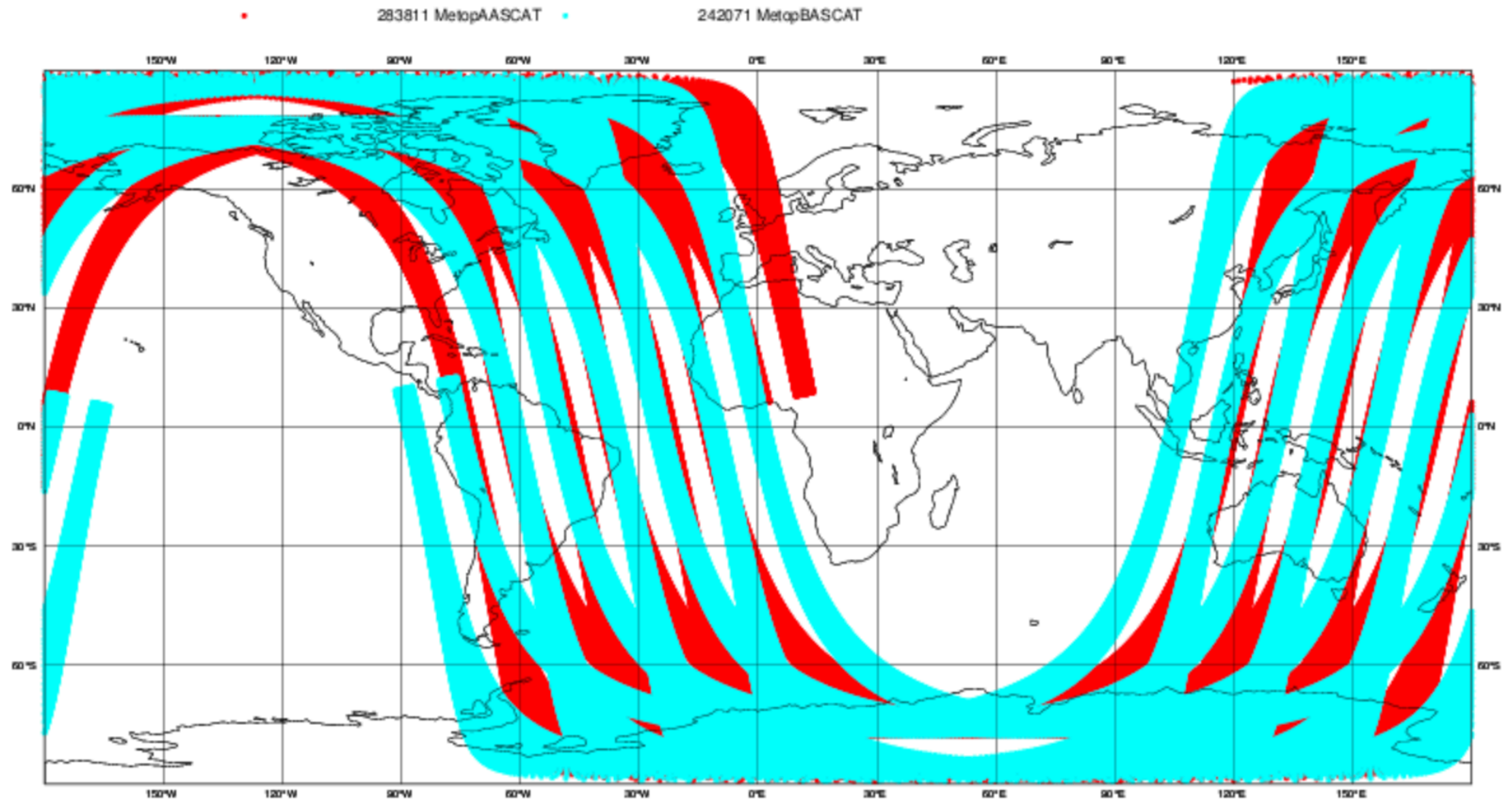
Total number of obs = 526686



ECMWF Data Coverage (All obs DA) - SCAT

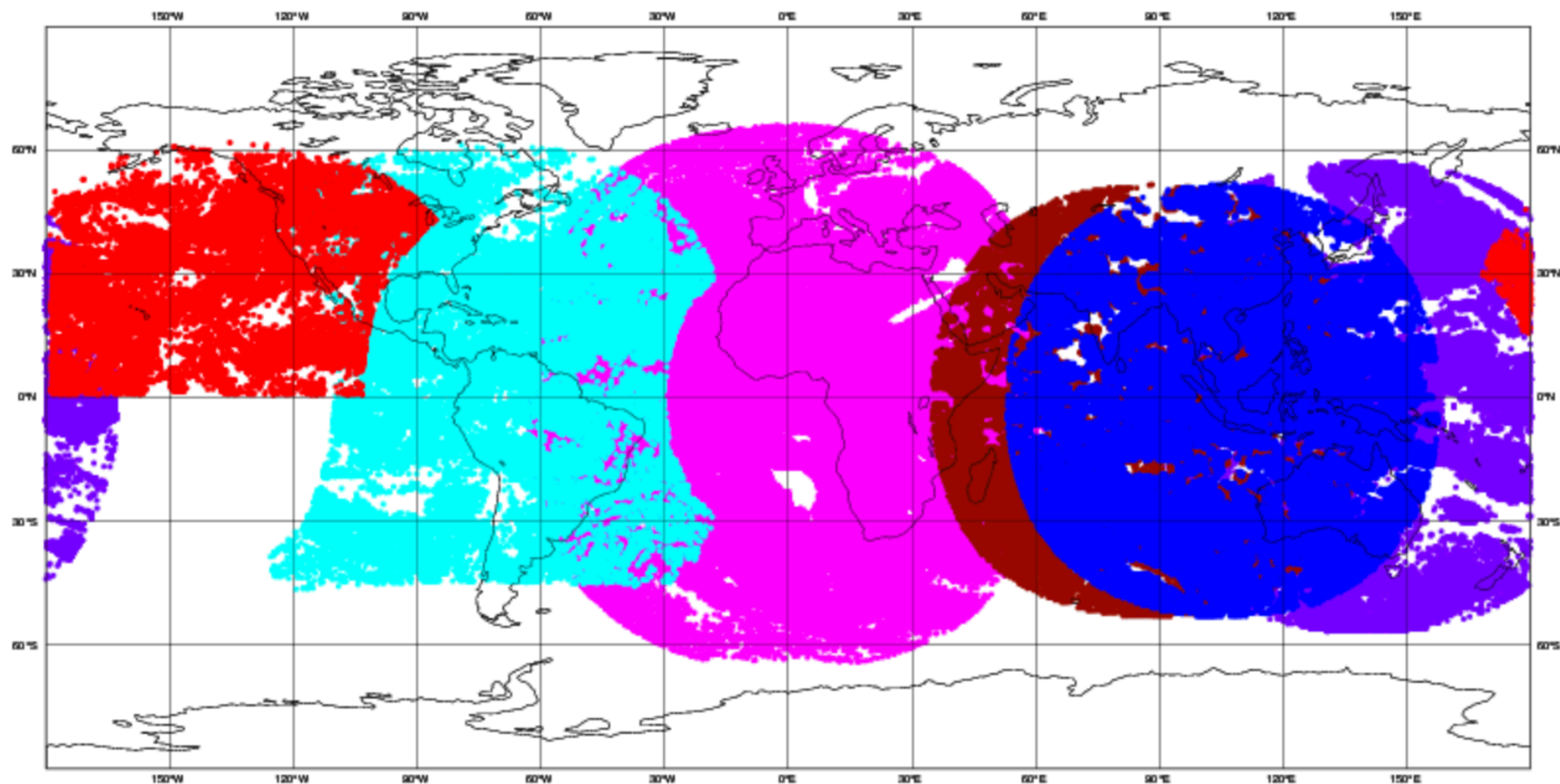
19/Apr/2017; 00 UTC

Total number of obs = 525882



ECMWF Data Coverage (All obs DA) - AMV WV
19/Apr/2017; 00 UTC
Total number of obs = 451877

• 25400 Goes15 • 51256 Goes13 • 122071 Met10 • 0 MTSAT-2 • 188231 HIMAWARIB32519 FY-2E • 32400 FY-2G • 0 Met7 • 0 Goes14

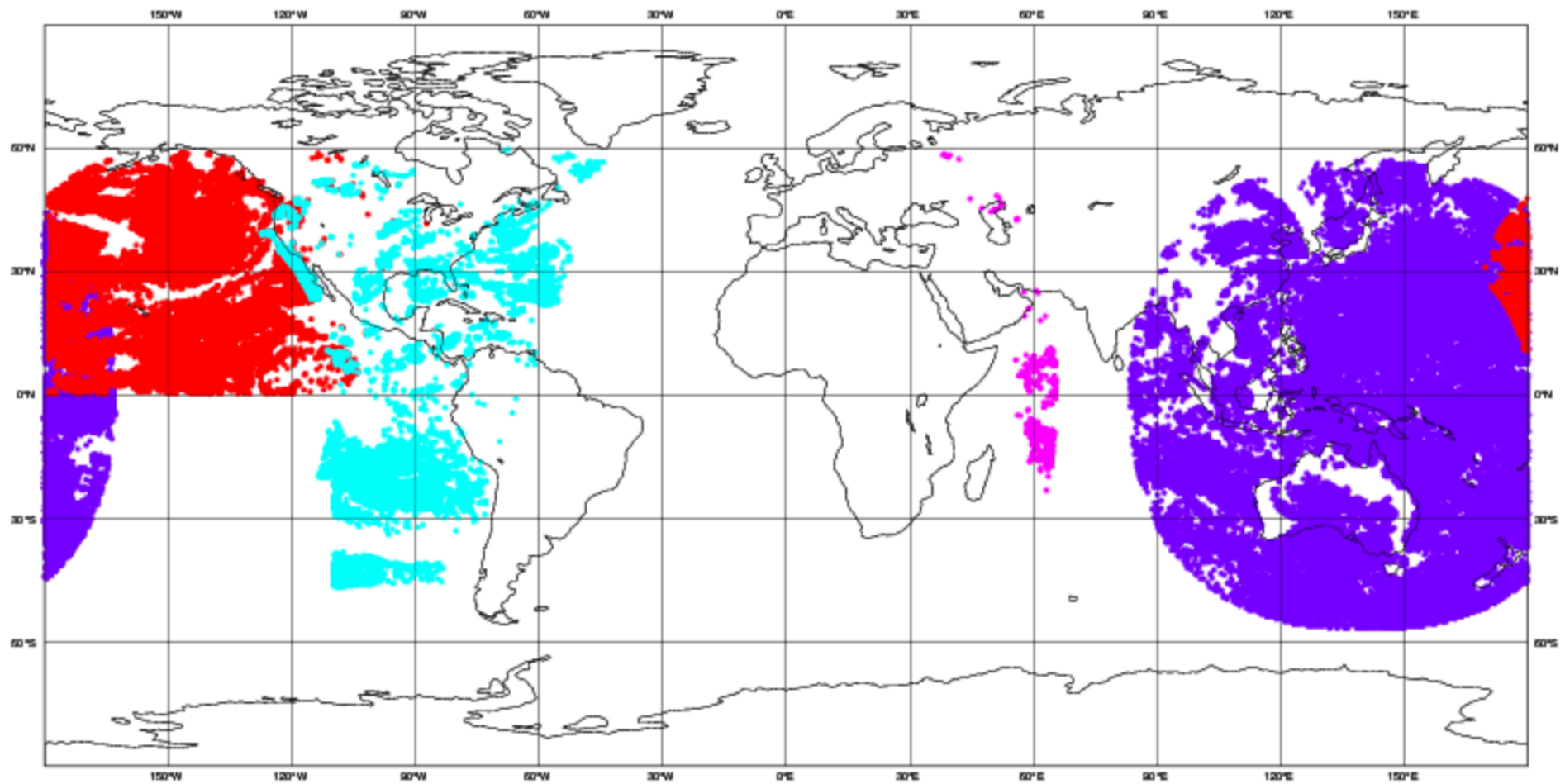


ECMWF Data Coverage (All obs DA) - AMV VIS

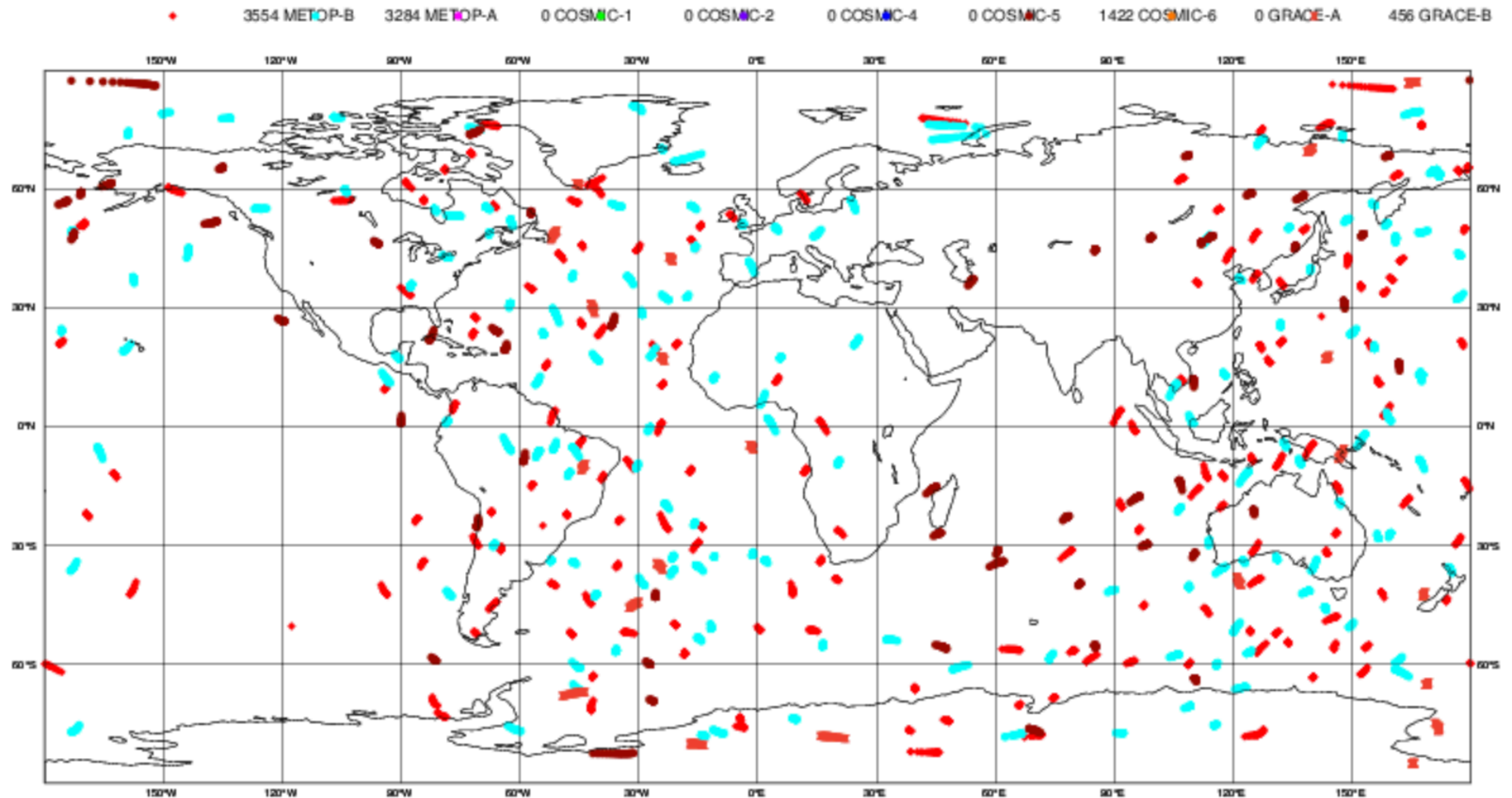
19/Apr/2017; 00 UTC

Total number of obs = 129800

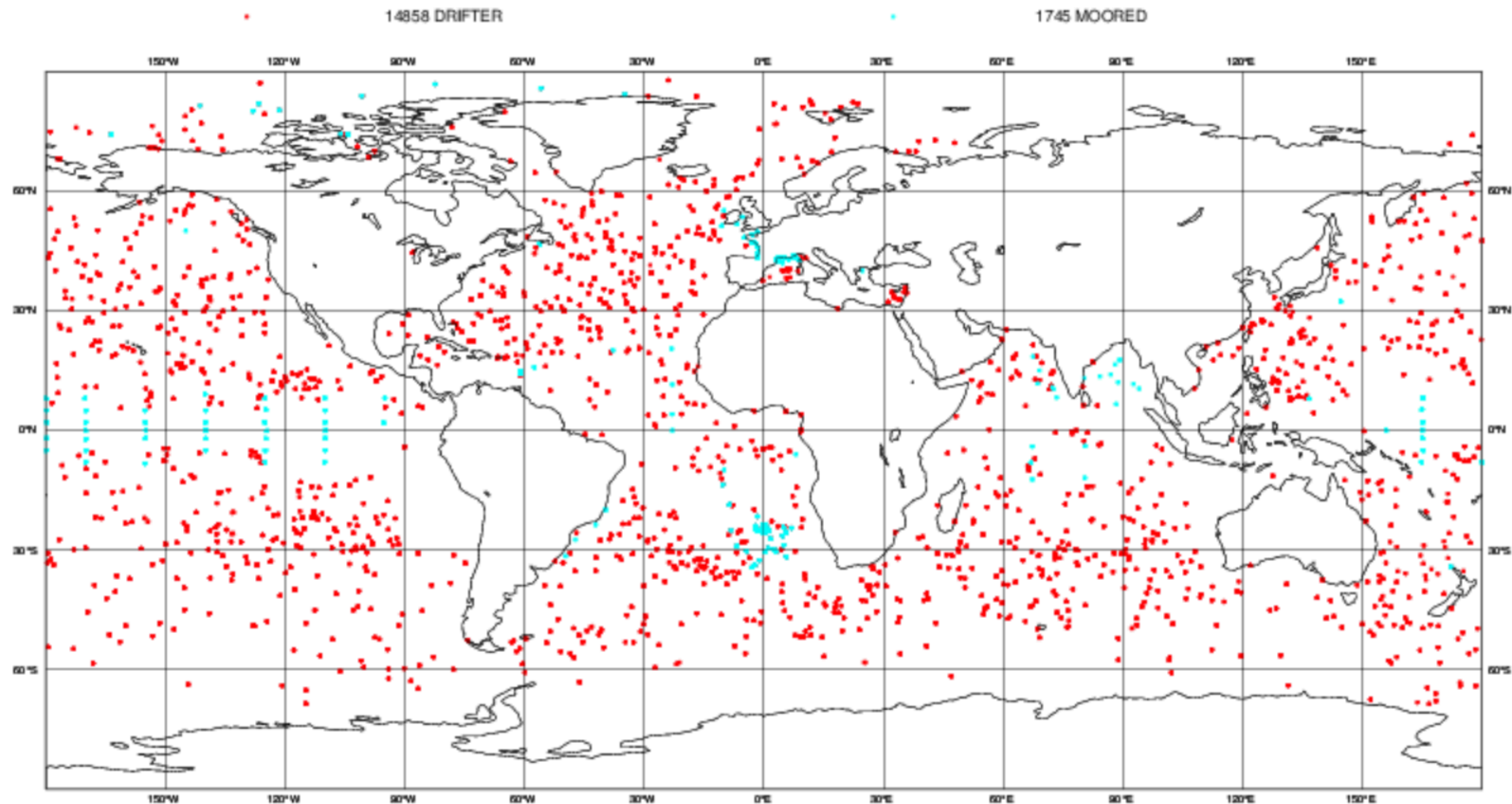
• 33726 Goes15 • 8622 Goes13 • 360 Met10 • 0 MTSAT-2 • 87092 HIMAWARI8 • 0 FY-2E • 0 FY-2G • 0 Met7 • 0 Goes14



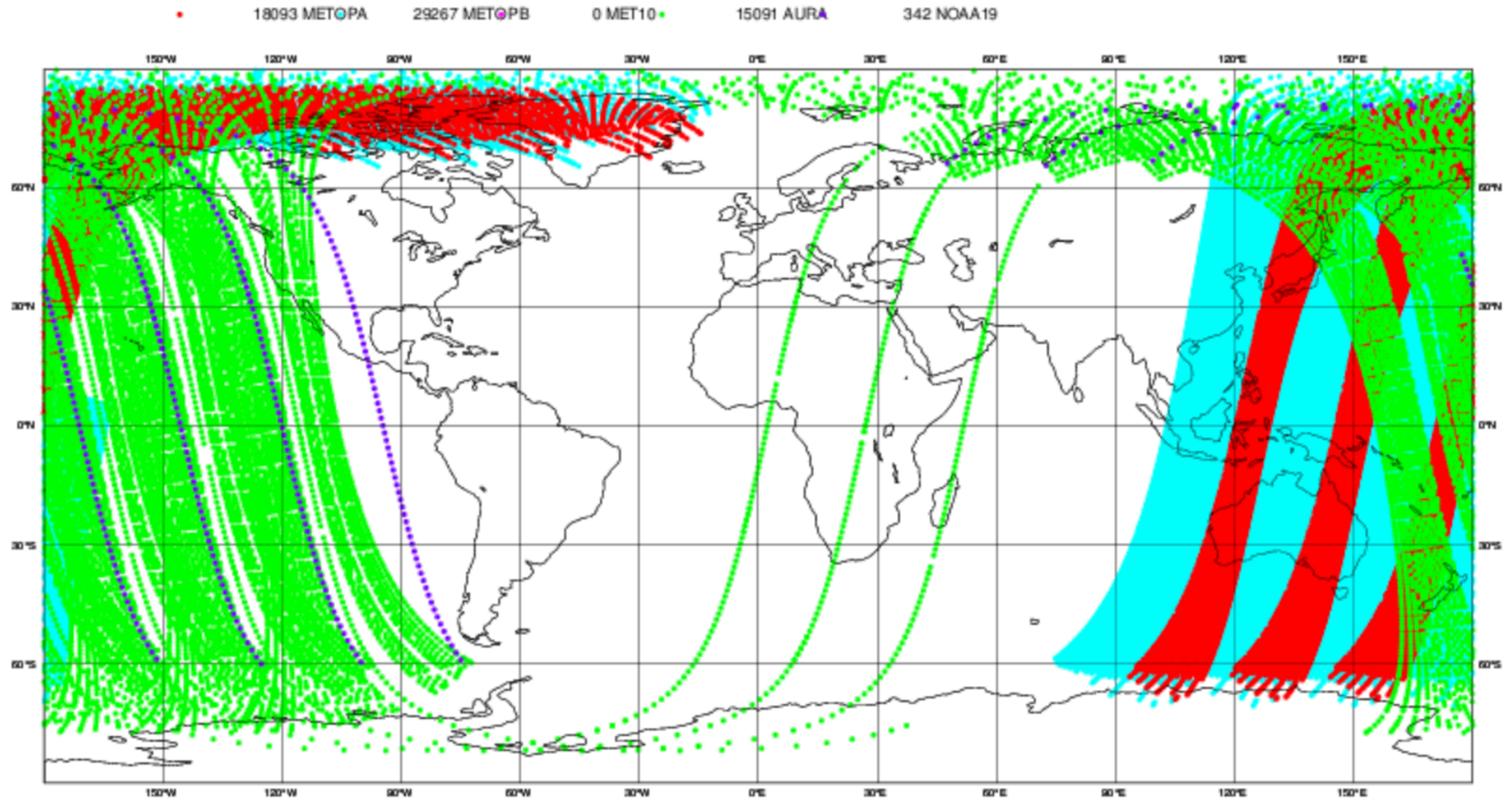
ECMWF Data Coverage (All obs DA) - GPSRO
19/Apr/2017; 00 UTC
Total number of obs = 8716



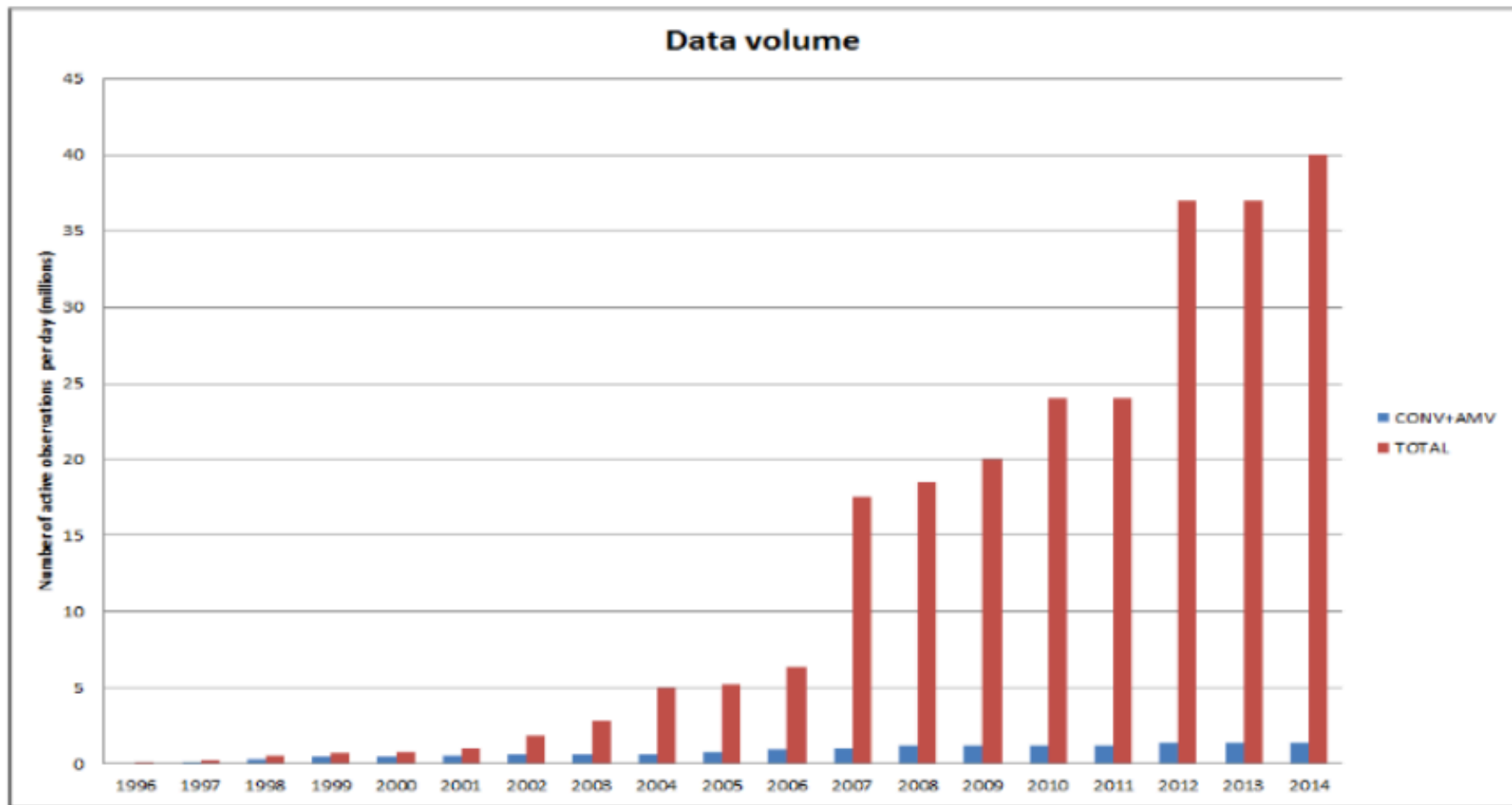
ECMWF Data Coverage (All obs DA) - Buoy
19/Apr/2017; 00 UTC
Total number of obs = 16603



ECMWF Data Coverage (All obs DA) - OZONE
19/Apr/2017; 00 UTC
Total number of obs = 62793



ECMWF



- *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 UTC)
- *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 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 describing the flow

$$y = H(x)$$

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

Échantillonnage de la circulation océanique par les missions altimétriques sur 10 jours :
combinaison Topex-Poséidon/ERS-1



S. Louvel, Doctoral Dissertation, 1999

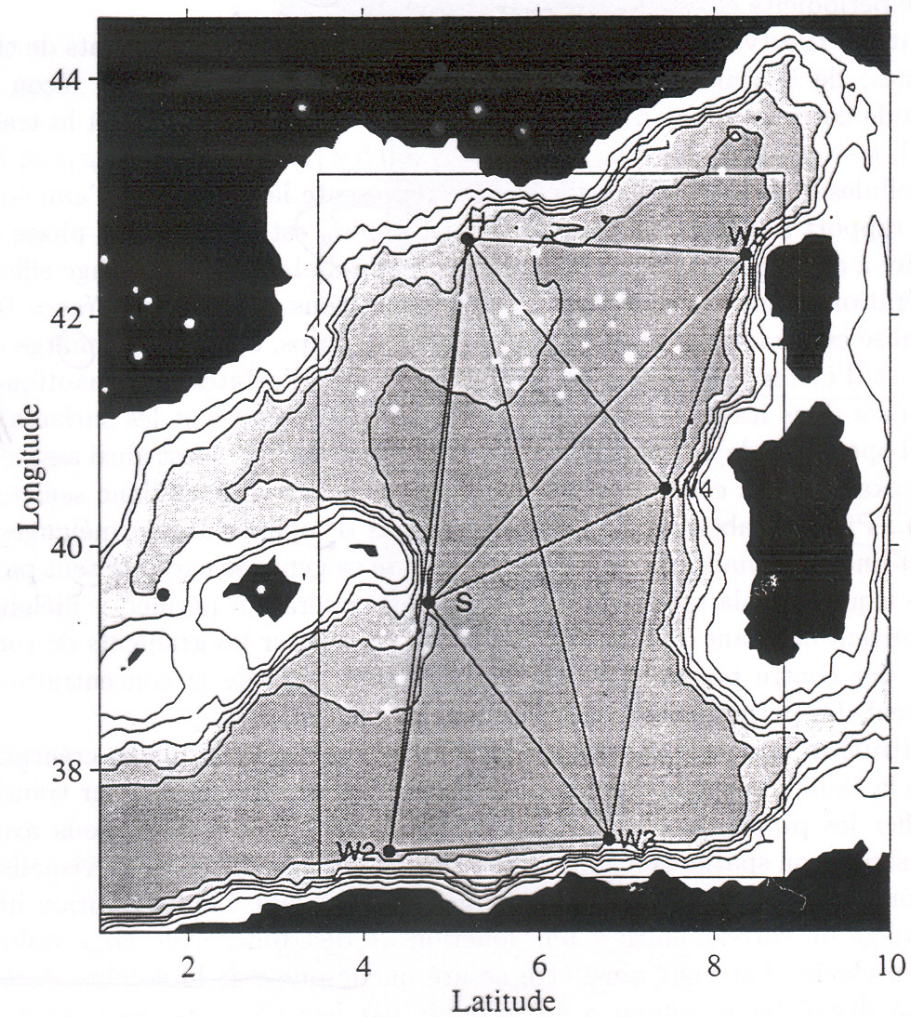


FIG. 1 - Bassin méditerranéen occidental: réseau d'observation tomographique de l'expérience Thétis 2 et limites du domaine spatial utilisé pour les expériences numériques d'assimilation.

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.

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, $p \approx 4$ - $5 \cdot 10^7$ 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

Both observations and ‘model’ are affected with some uncertainty \Rightarrow uncertainty on the estimate.

For some reason, uncertainty is conveniently described by probability distributions (don’t know too well why, but it works; see, *e.g.* Jaynes, 2007, *Probability Theory: The Logic of Science*, Cambridge University Press).

[Assimilation is a problem in bayesian estimation.](#)

Determine the conditional probability distribution for the state of the system, knowing everything we know (see Tarantola, A., 2005, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM).

Coût des différentes composantes de la chaîne de prévision opérationnelle du CEPMMT (septembre 2015, J.-N. Thépaut) :

4DVAR: 9.5%

HRES FC: 4.5%

EDA: 30%

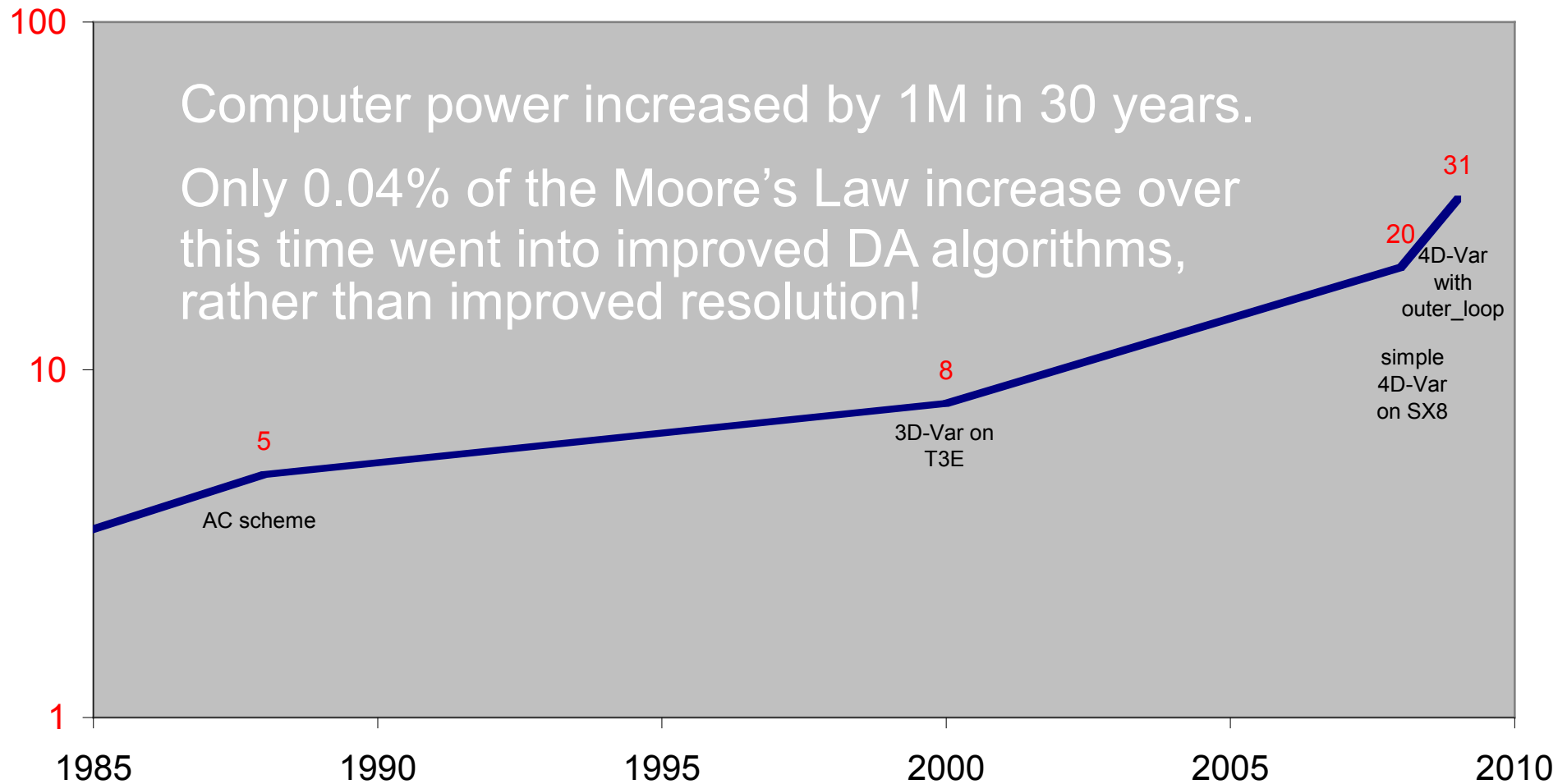
ENS: 22%

ENS: hindcasts 14%

Other: 20% of which BC AN: 3.5% BC FC: 4% BC ENS: 9.5%

L'EDA fournit à la fois les variances d'erreur d'ébauche du 4D-Var, et les perturbations initiales (en complément des vecteurs singuliers) de l'EPS.

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



Courtesy A. Lorenc

Bayesian Estimation

Determine conditional probability distribution of the state of the system, given the probability distribution of the uncertainty on the data

$$z_1 = x + \zeta_1 \quad \zeta_1 = \mathcal{N}[0, s_1]$$

density function $p_1(\zeta) \propto \exp[-(\zeta^2)/2s_1]$

$$z_2 = x + \zeta_2 \quad \zeta_2 = \mathcal{N}[0, s_2]$$

density function $p_2(\zeta) \propto \exp[-(\zeta^2)/2s_2]$

- ζ_1 and ζ_2 mutually independent

What is the conditional probability $P(x = \xi | z_1, z_2)$ that x be equal to some value ξ ?

$$\begin{array}{ll}
z_1 = x + \zeta_1 & \text{density function } p_1(\zeta) \propto \exp[-(\zeta^2)/2s_1] \\
z_2 = x + \zeta_2 & \text{density function } p_2(\zeta) \propto \exp[-(\zeta^2)/2s_2] \\
& \zeta_1 \text{ and } \zeta_2 \text{ mutually independent}
\end{array}$$

$$x = \xi \Leftrightarrow \zeta_1 = z_1 - \xi \text{ and } \zeta_2 = z_2 - \xi$$

- $$\begin{aligned}
P(x = \xi | z_1, z_2) &\propto p_1(z_1 - \xi) p_2(z_2 - \xi) \\
&\propto \exp[-(\xi - x^a)^2 / 2p^a]
\end{aligned}$$

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

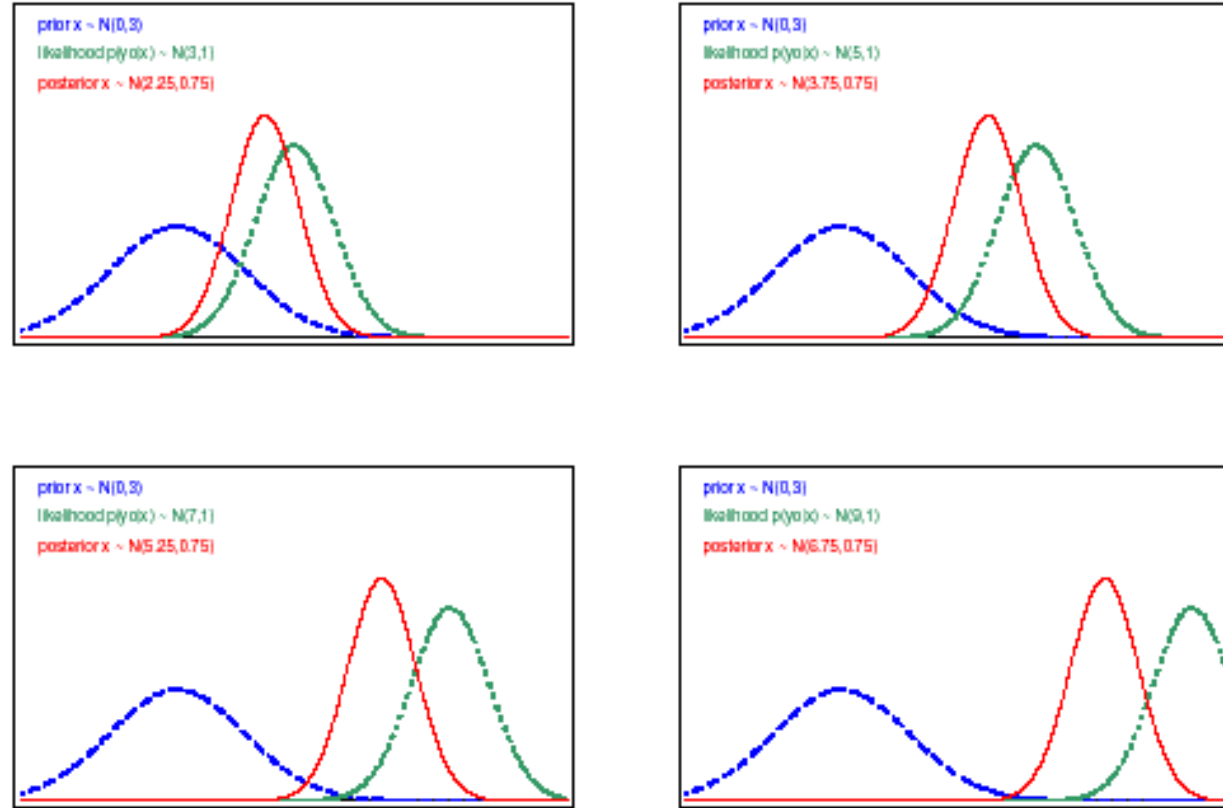


Fig. 1.1: Prior pdf $p(x)$ (dashed line), posterior pdf $p(x|y^o)$ (solid line), and Gaussian likelihood of observation $p(y^o|x)$ (dotted line), plotted against x for various values of y^o . (Adapted from Lorenc and Hammon 1988.)

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

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

In addition

$$p^a = 1 / J''(x^a)$$

Conditional probability distribution in Gaussian case

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

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)

$$z_1 = x + \xi_1$$

$$z_2 = x + \xi_2$$

Same as before, but ξ_1 and ξ_2 are now distributed according to exponential law with parameter a , *i. e.*

$$p(\xi) \propto \exp[-|\xi|/a] \quad ; \quad \text{Var}(\xi) = 2a^2$$

Conditional probability density function is now uniform over interval $[z_1, z_2]$, exponential with parameter $a/2$ outside that interval

$$E(x | z_1, z_2) = (z_1 + z_2)/2$$

$$\text{Var}(x | z_1, z_2) = a^2 (2\delta^3/3 + \delta^2 + \delta + 1/2) / (1 + 2\delta), \text{ with } \delta = |z_1 - z_2| / (2a)$$

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.

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

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 ?

$$x = \xi \Rightarrow z = F(\xi, \zeta)$$

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

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 $N \approx O(10-100)$).

Random vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^T = (x_i)$ (e. g. pressure, temperature, abundance of given chemical compound at n grid-points of a numerical model)

- Expectation $E(\mathbf{x}) \equiv [E(x_i)]$; centred vector $\mathbf{x}' \equiv \mathbf{x} - E(\mathbf{x})$
- Covariance matrix

$$E(\mathbf{x}'\mathbf{x}'^T) = [E(x_i'x_j')]$$

dimension $n \times n$, symmetric non-negative (strictly definite positive except if linear relationship holds between the x_i' 's with probability 1).

- Two random vectors
 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$
 $\mathbf{y} = (y_1, y_2, \dots, y_p)^T$

$$E(\mathbf{x}'\mathbf{y}'^T) = E(x_i'y_j')$$

dimension $n \times p$

Covariance matrices will be denoted

$$C_{xx} \equiv E(\mathbf{x}'\mathbf{x}'^T)$$

$$C_{xy} \equiv E(\mathbf{x}'\mathbf{y}'^T)$$

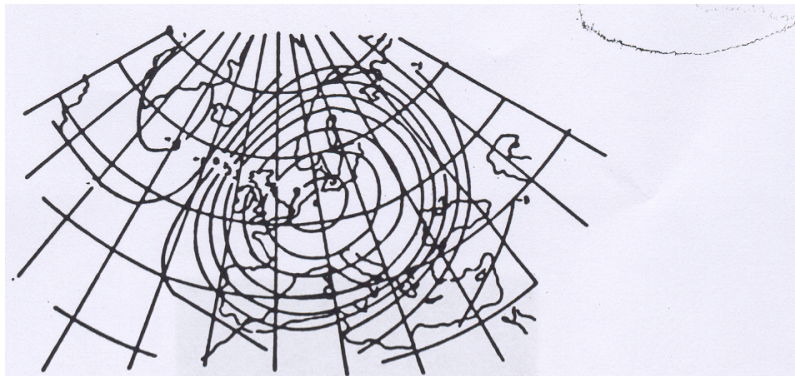
Random function $\Phi(\xi)$ (field of pressure, temperature, abundance of given chemical compound, ... ; ξ is now spatial and/or temporal coordinate)

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

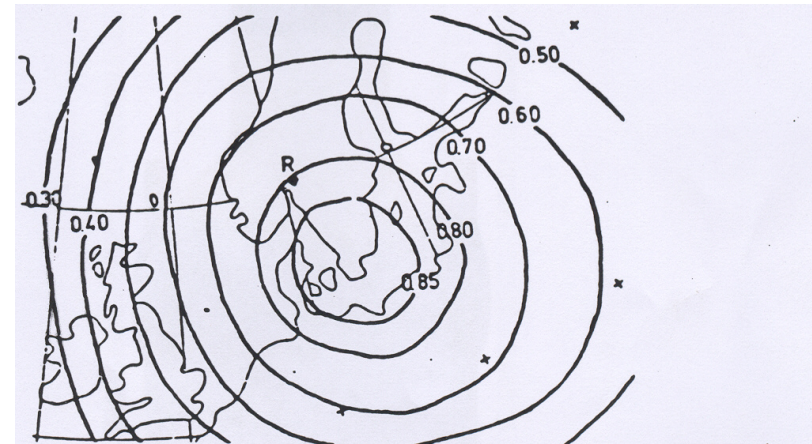
$$(\xi_1, \xi_2) \rightarrow C_\phi(\xi_1, \xi_2) \equiv E[\Phi'(\xi_1) \Phi'(\xi_2)]$$

- 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

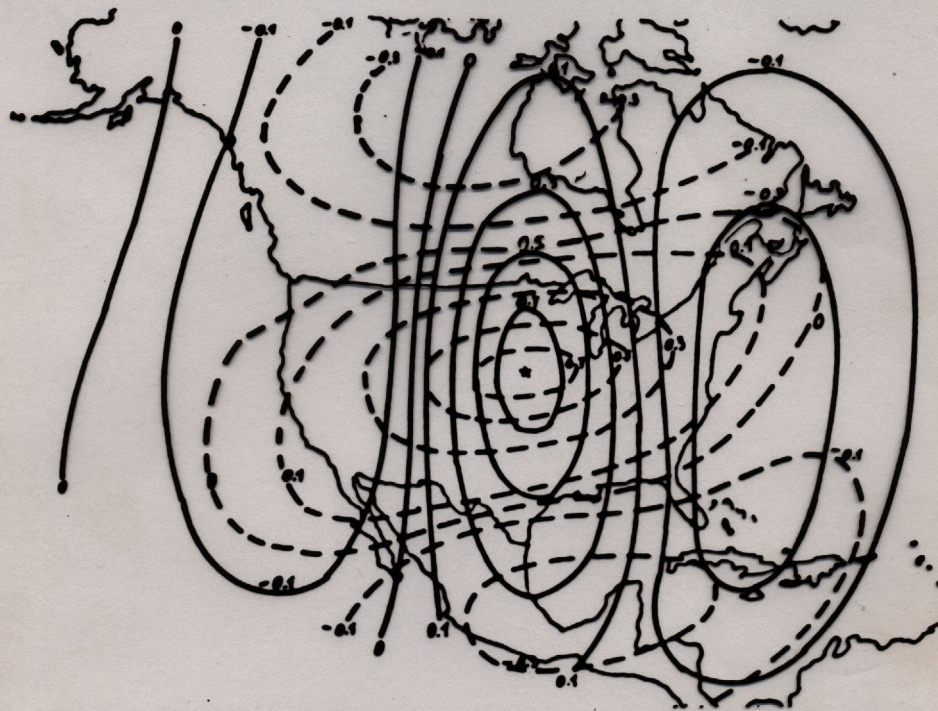
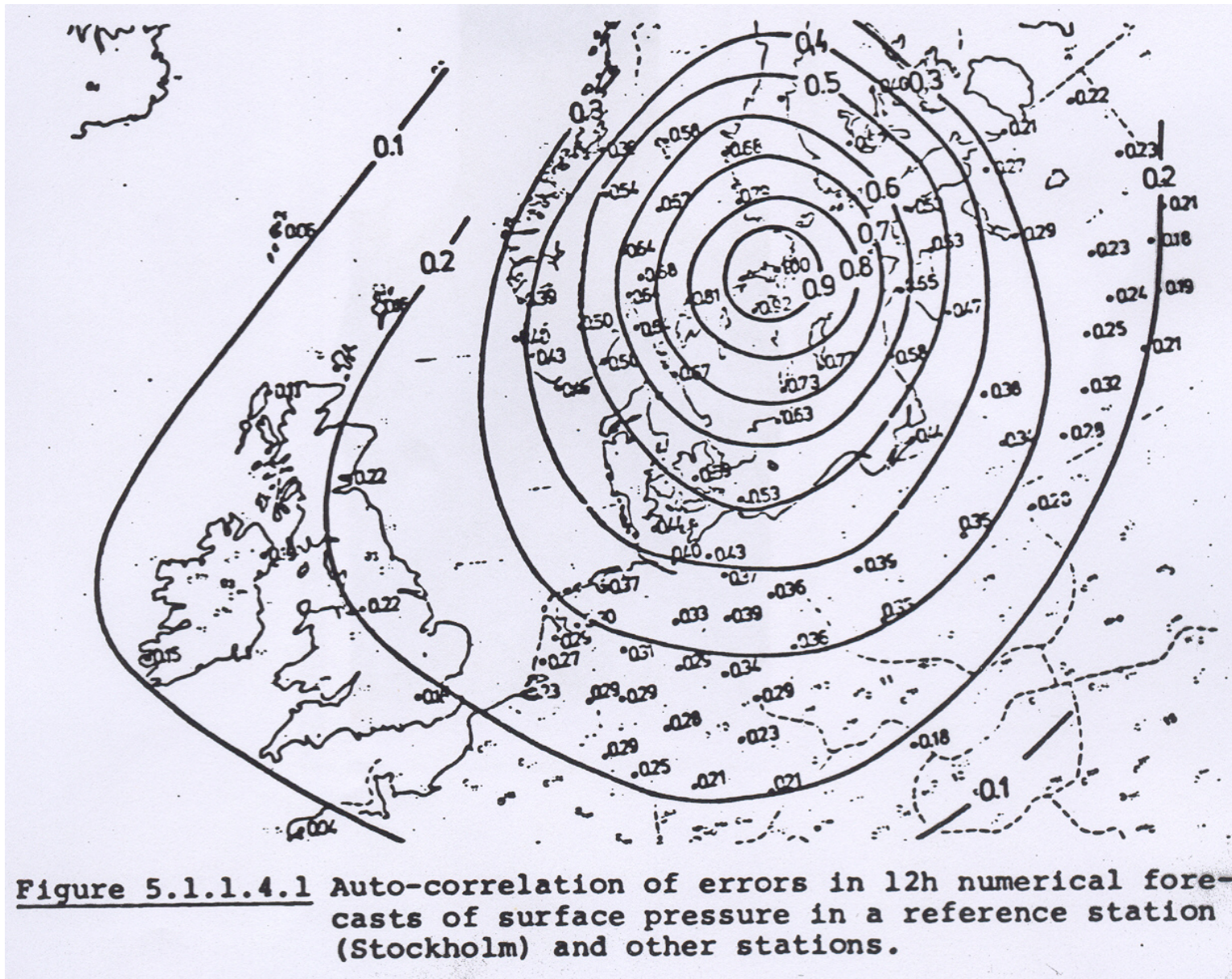


Figure 4.2.4.3: Isolines for the auto-correlation of the 500 mb u-wind component (dashed line) and the auto-correlation of the 500 mb v-wind component (full line). The "star" indicates the position of the reference station. (From Buel (1972).

After N. Gustafsson



After N. Gustafsson

Cours à venir

~~Jeudi 6 avril~~

~~Jeudi 13 avril~~

~~Jeudi 20 avril~~

Jeudi 11 mai

Lundi 29 mai

Jeudi 1 juin

Jeudi 15 juin

Jeudi 22 juin

De 10h00 à 12h30, Salle de la Serre, 5ième étage,
Département de Géosciences, École Normale Supérieure,
24, rue Lhomond, Paris 5