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

Année Universitaire 2013-2014

# Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 3

18 Avril 2014

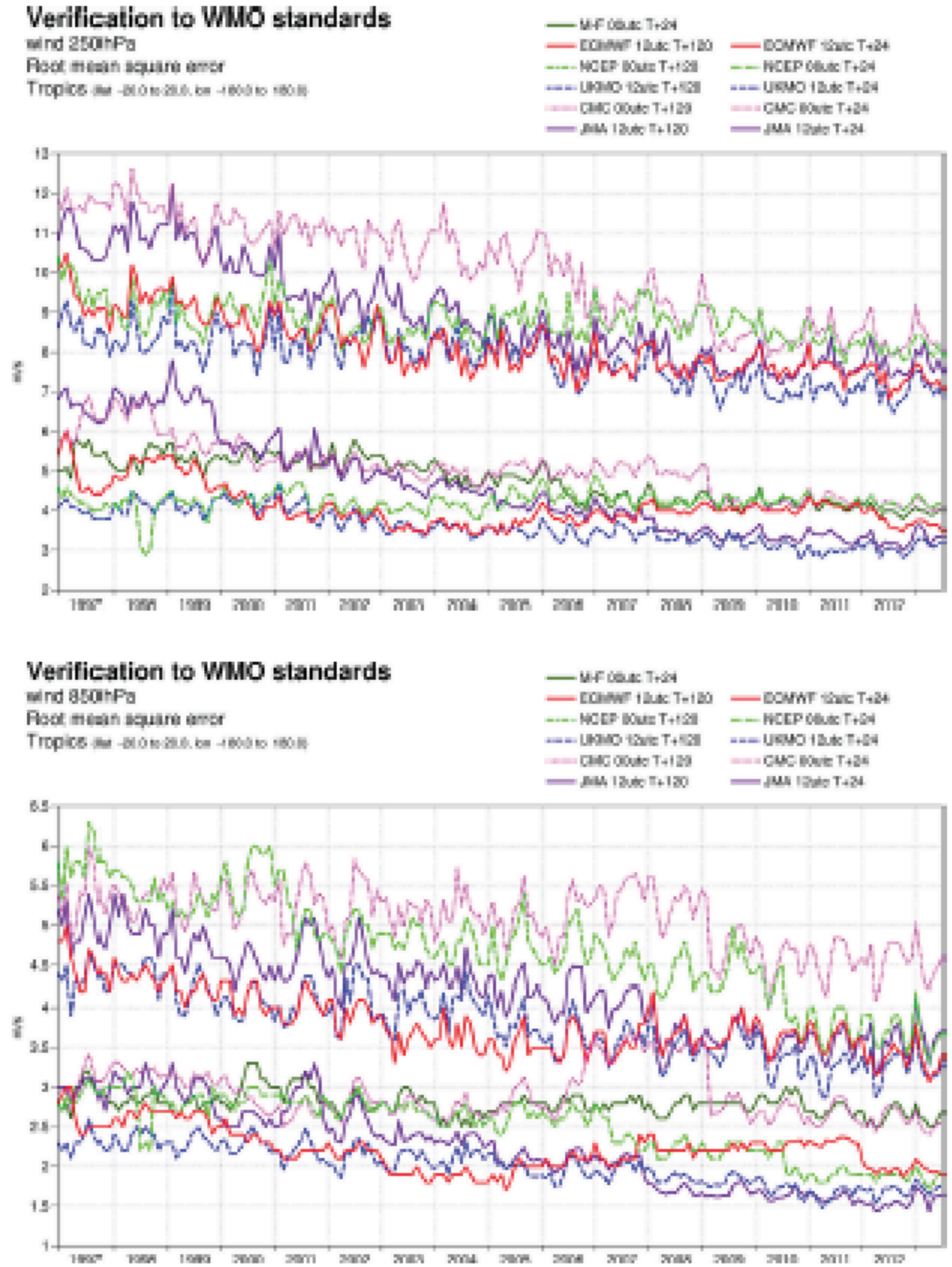



*Figure 16: WMO-exchanged scores from global forecast centres. RMS vector wind error over tropics at 250 hPa (top) and 850 hPa (bottom). In each panel the upper curves show the five-day forecast error and the lower curves show the one-day forecast error. Each model is verified against its own analysis.*

ECMWF

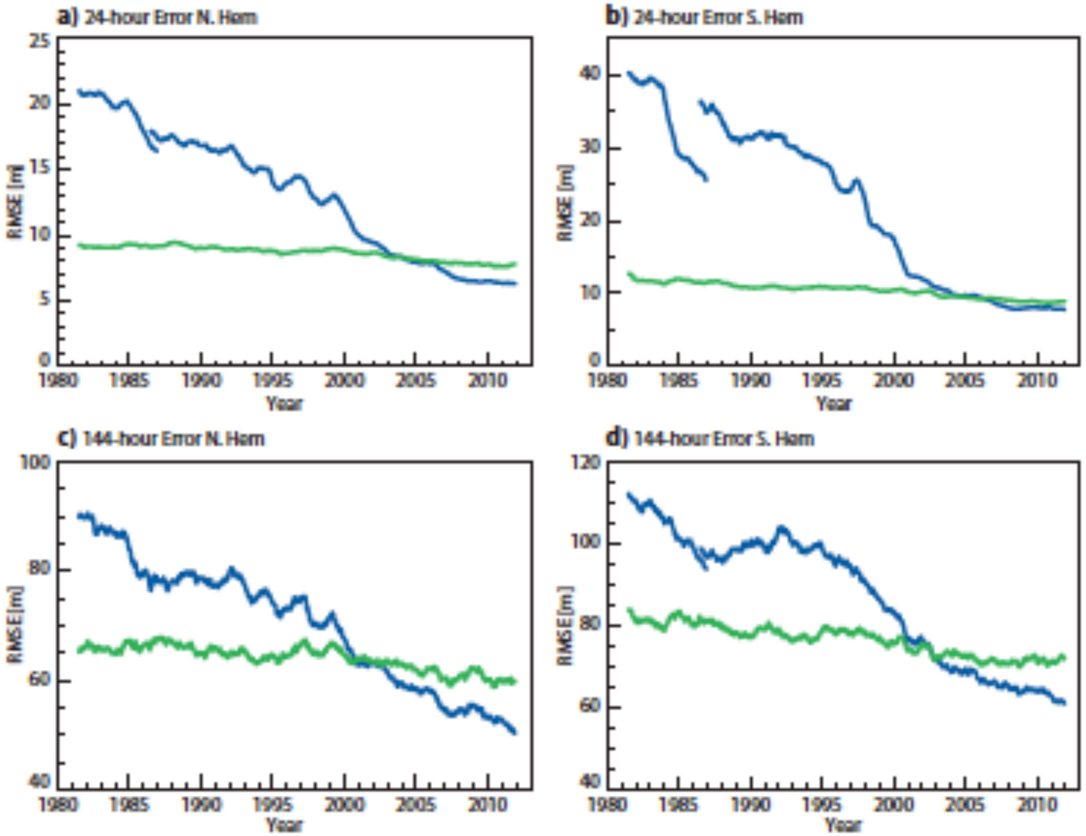



FIG. 3. Evolution of forecast errors from 1981 to 2012 for N.Hem (a and c) and S.Hem (b and d). Operational forecasts (blue) and ERA Interim (green). Note that before 1986 the operational analysis is used to verify the operational forecasts, after 1986 ERA Interim is used for the verification (with an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

Magnusson and Källén, 2013, *Mon. Wea. Rev.*, **141**, 3142–3153


**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:* *1$^{st}$ panel: MSLP analysis for 12 UTC of 29 Aug*
*2$^{nd}$ panel: MSLP t+84h $T_L$511L60 forecast started at 00 UTC of 26 Aug*
*3$^{rd}$ panel: MSLP t+84h EPS-control $T_L$255L40 forecast started at 00 UTC of 26 Aug*
*Other rows: 50 EPS-perturbed $T_L$255L40 forecast started at 00 UTC of 26 Aug.*

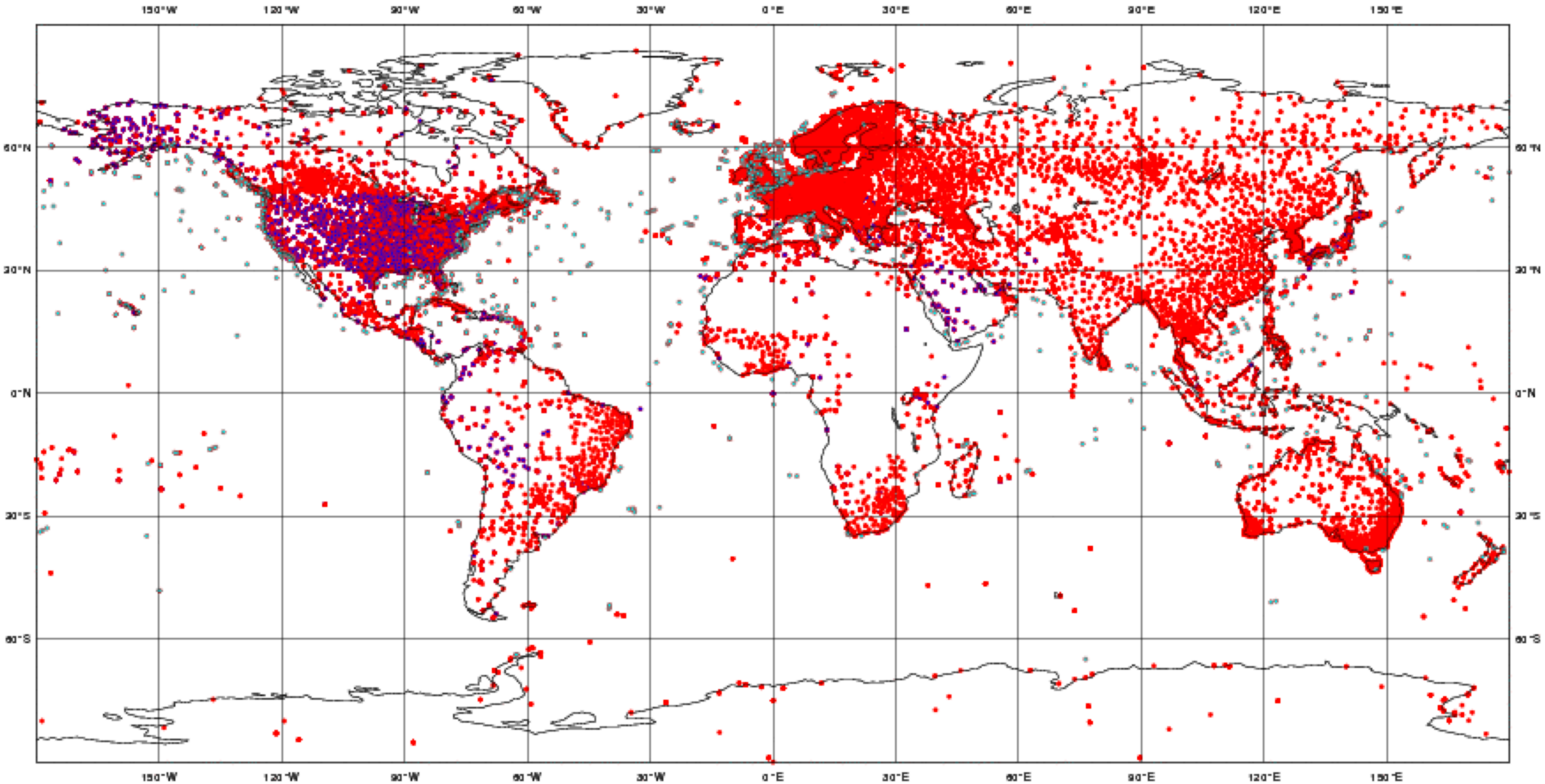
*The contour interval is 5 hPa, with shading patters for MSLP values lower than 990 hPa.*

*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.*

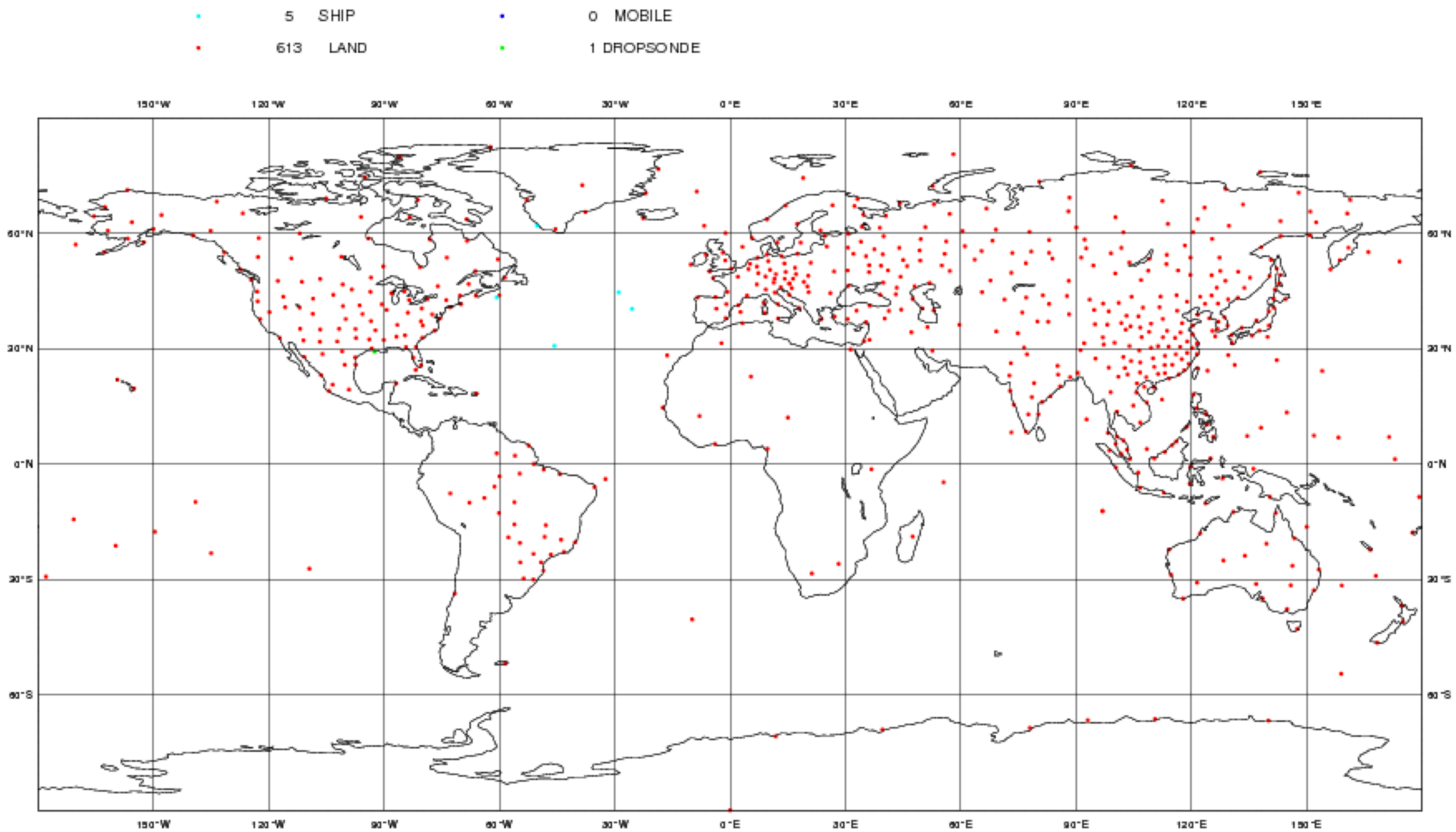
H. Poincaré, *Science et Méthode*, Paris, 1908

# ECMWF Data Coverage (All obs DA) - Synop-Ship-Metar
16/Apr/2014; 00 UTC
Total number of obs = 68742

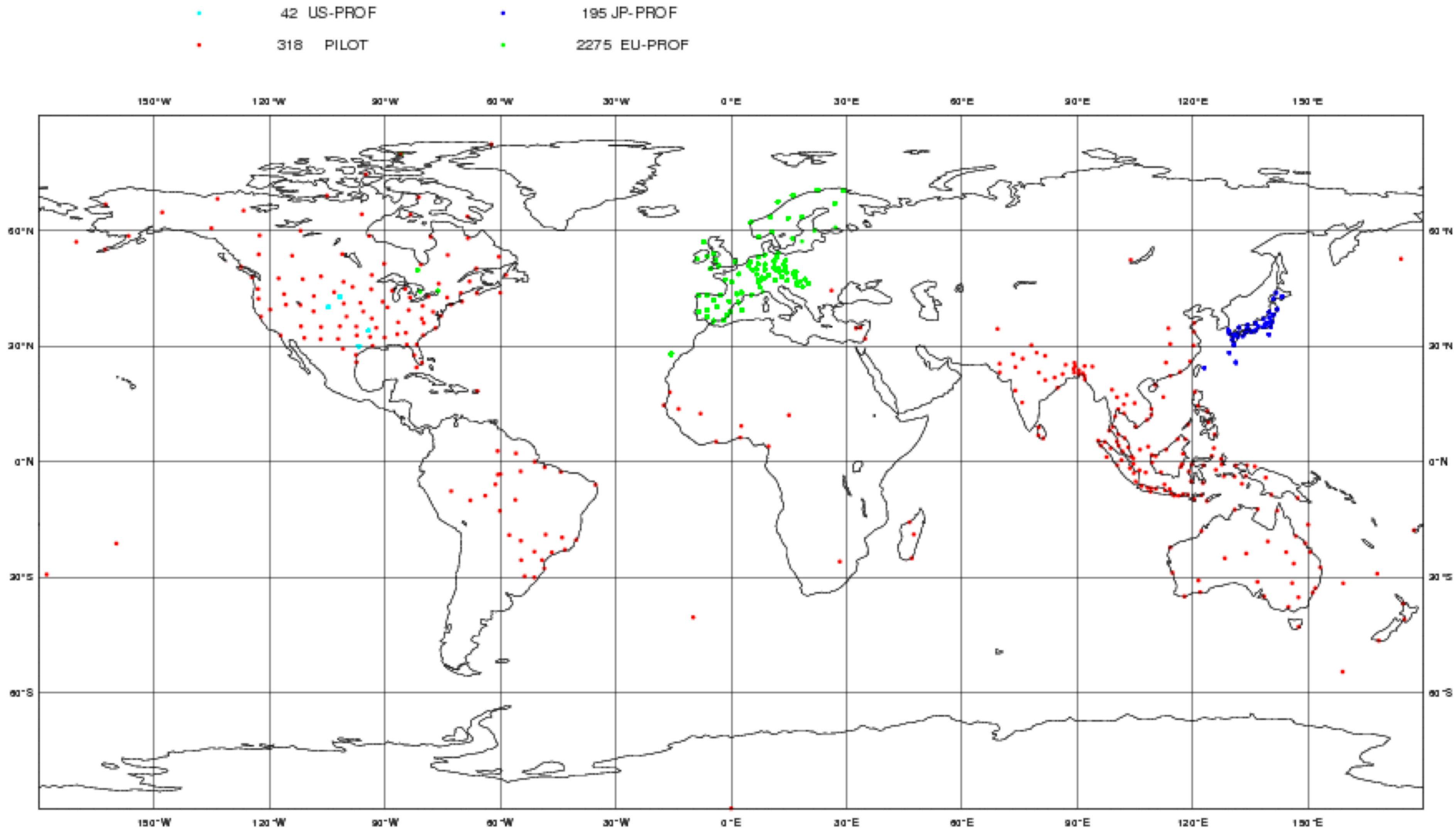
20467 SYNOP · 3163 SHIP · 45112 METAR



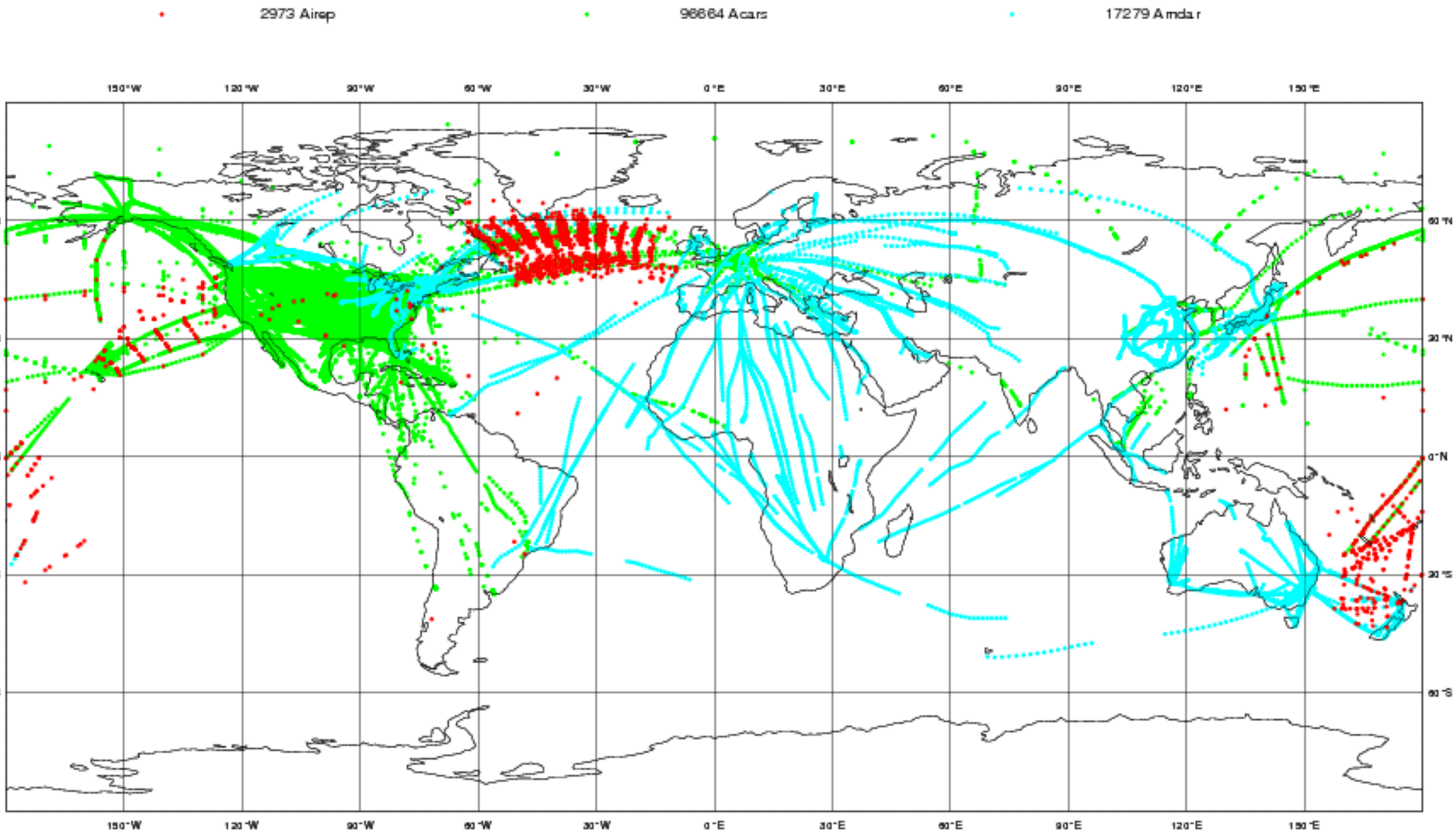
# ECMWF Data Coverage (All obs DA) - Temp
16/Apr/2014; 00 UTC
Total number of obs = 619



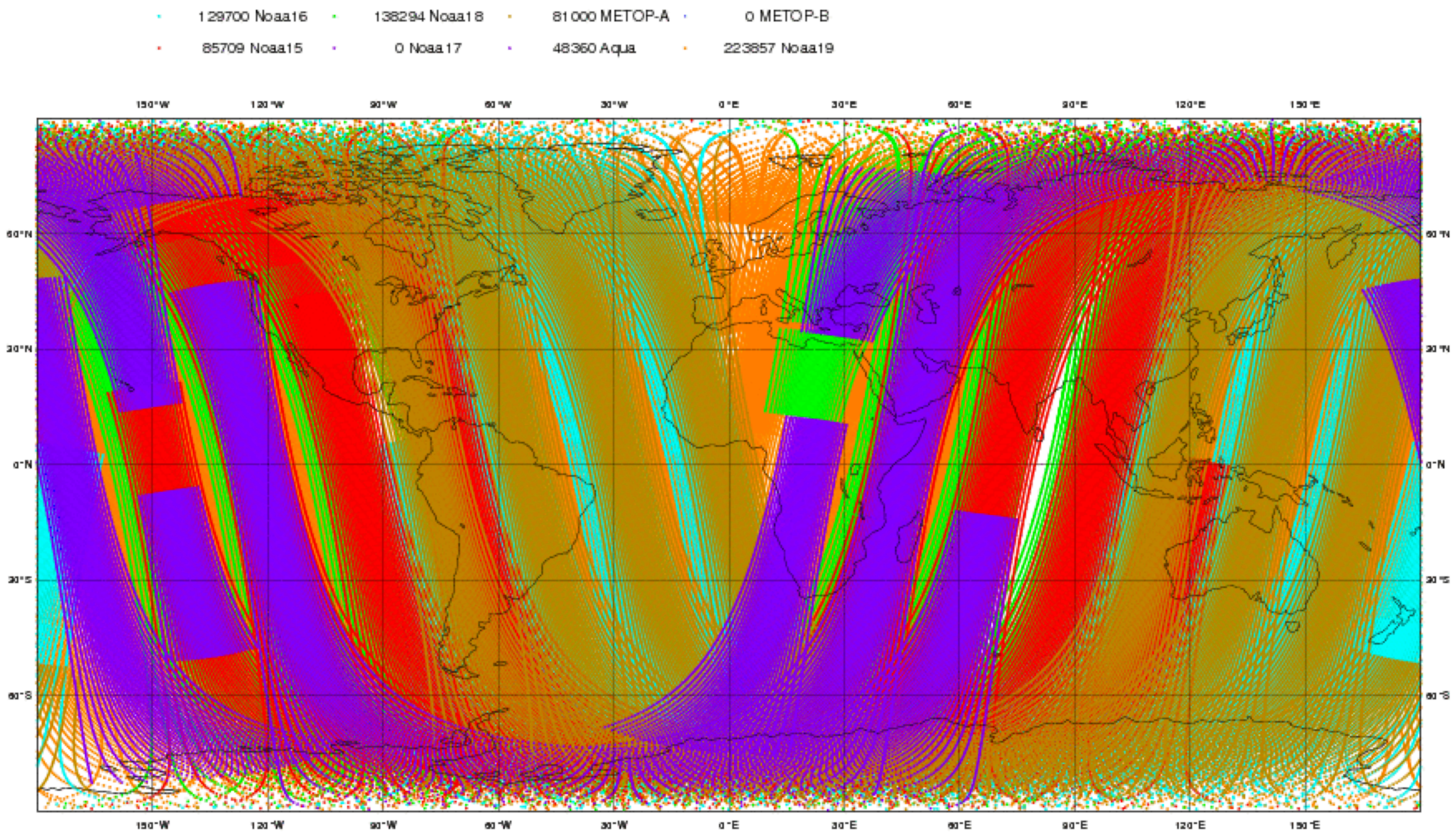
# ECMWF Data Coverage (All obs DA) - Pilot-Profiler
16/Apr/2014; 00 UTC
Total number of obs = 2830



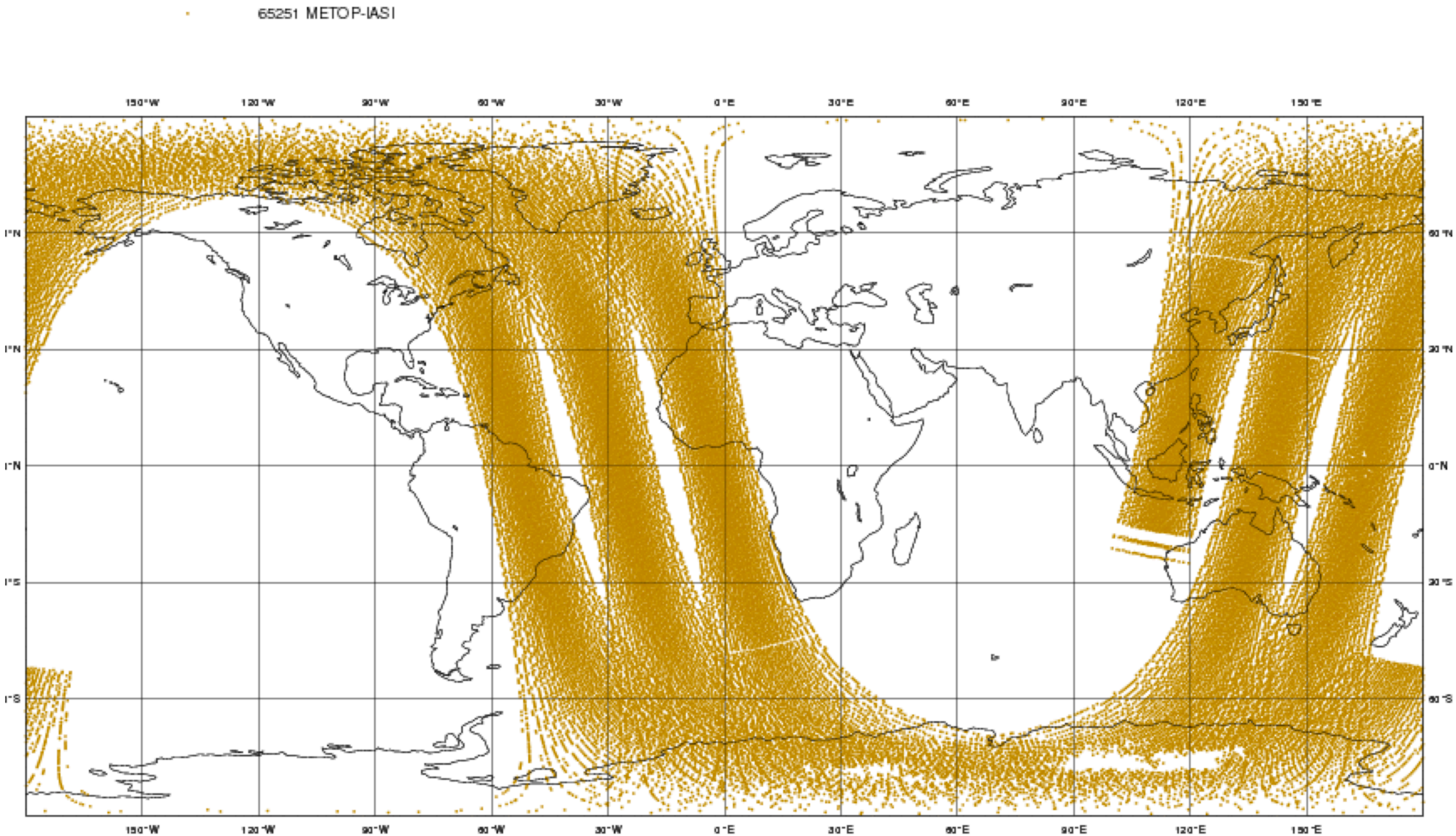
ECMWF Data Coverage (All obs DA) - Aircraft
16/Apr/2014; 00 UTC
Total number of obs = 116916



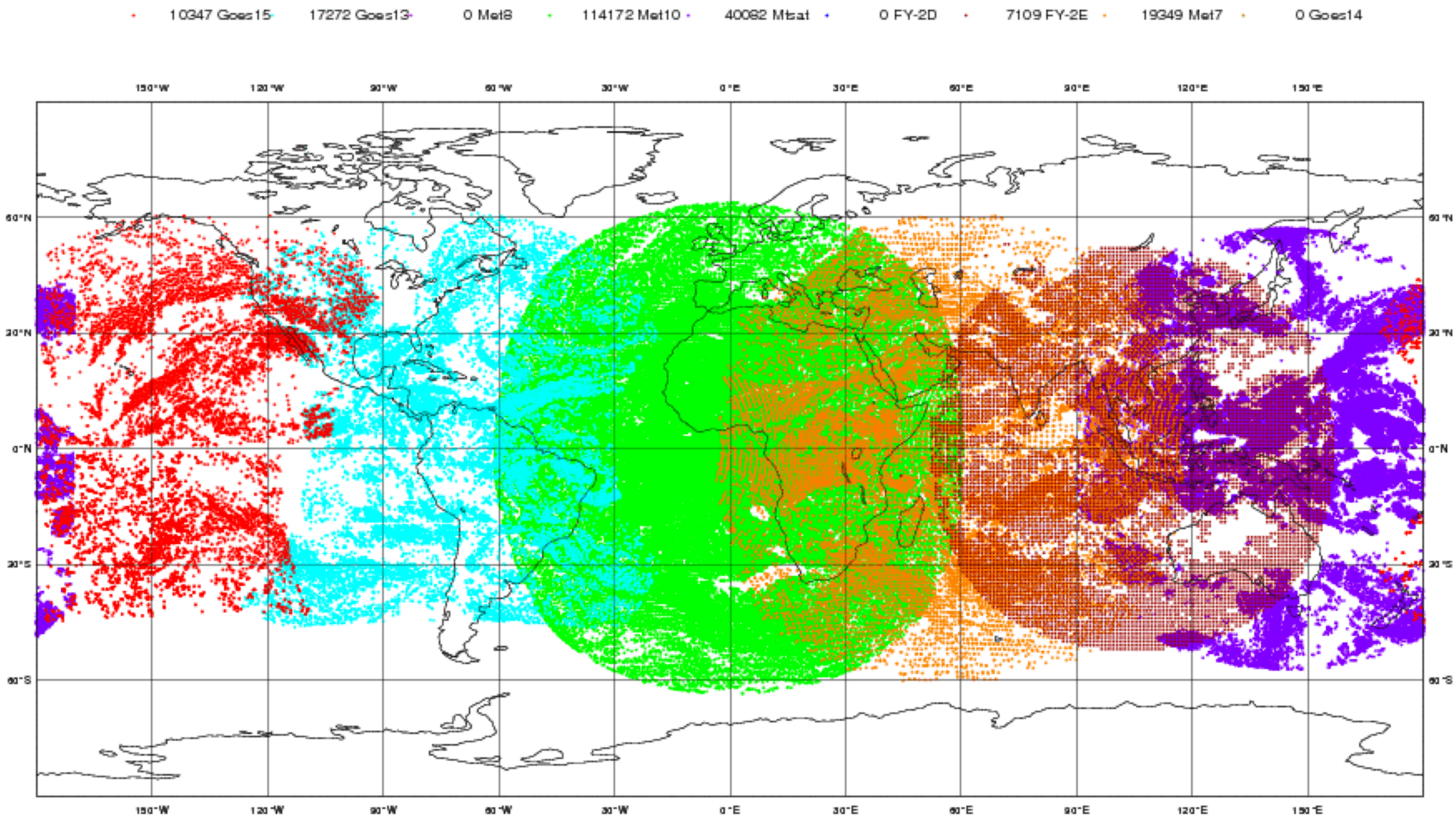
# ECMWF Data Coverage (All obs DA) - AMSU-A
16/Apr/2014; 00 UTC
Total number of obs = 706920



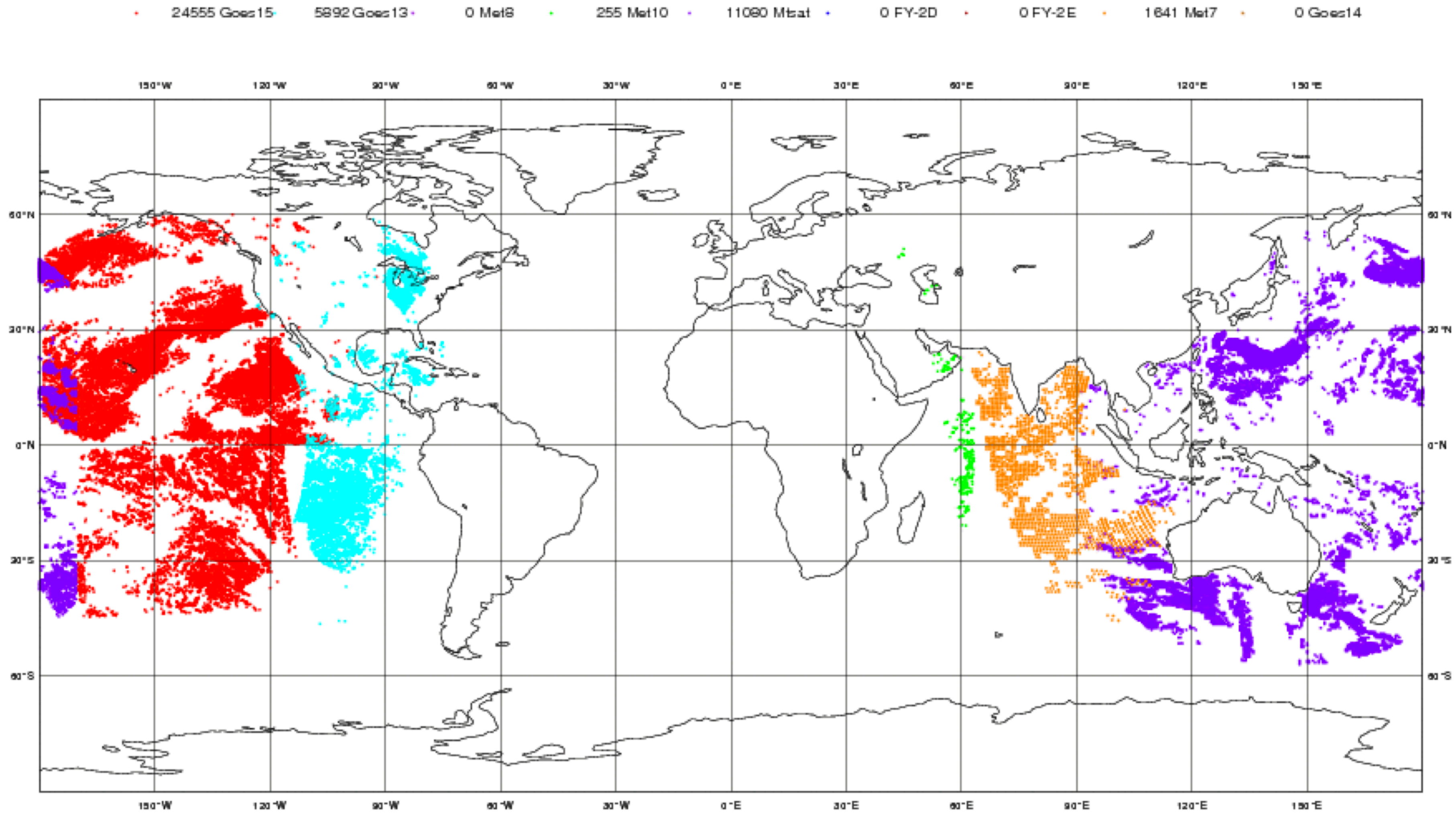
# ECMWF Data Coverage (All obs DA) - IASI
16/Apr/2014; 00 UTC
Total number of obs = 65251



# ECMWF Data Coverage (All obs DA) - AMV WV
16/Apr/2014; 00 UTC
Total number of obs = 208331



# ECMWF Data Coverage (All obs DA) - AMV VIS
16/Apr/2014; 00 UTC
Total number of obs = 43423

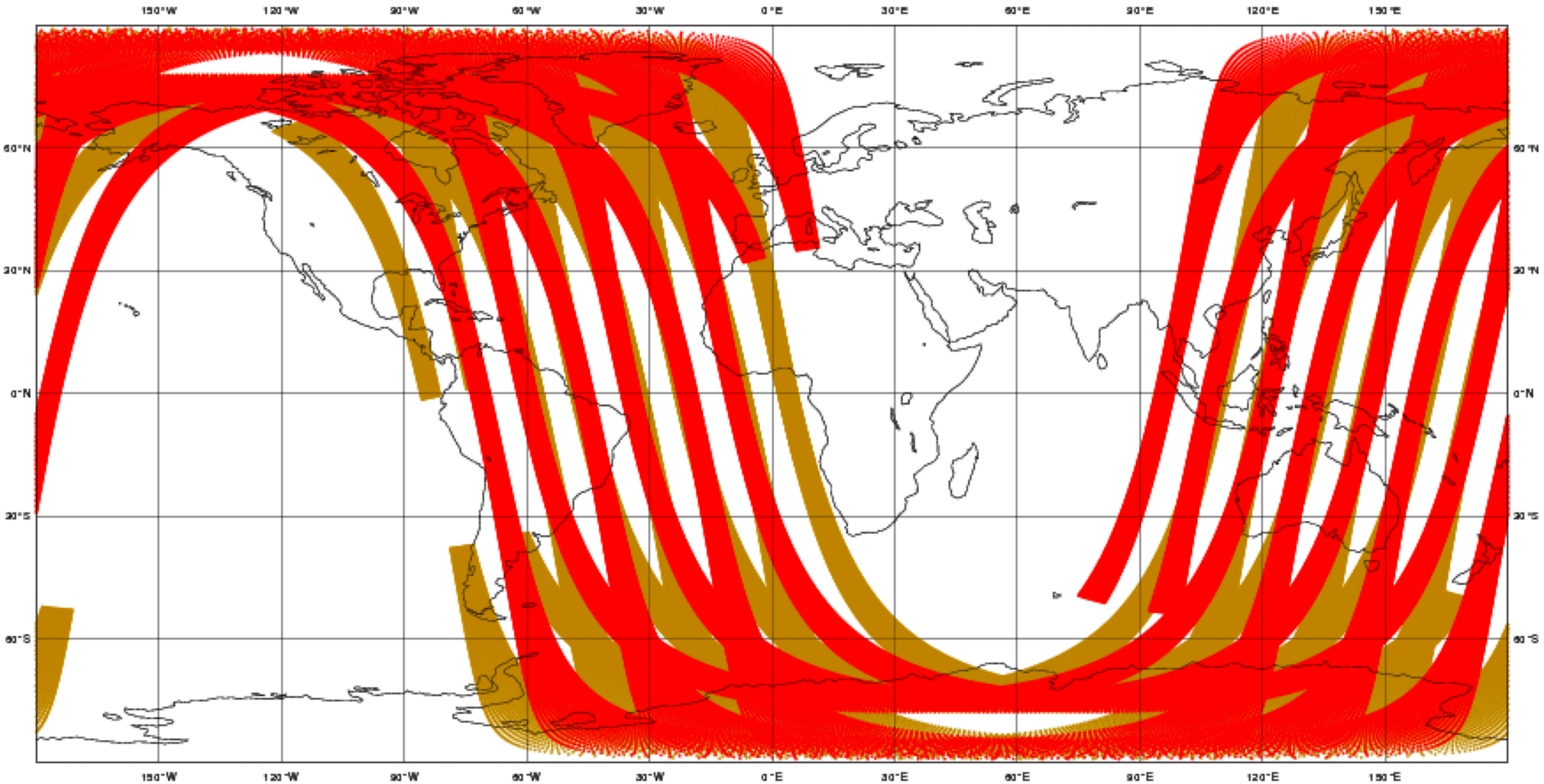


# ECMWF Data Coverage (All obs DA) - SCAT

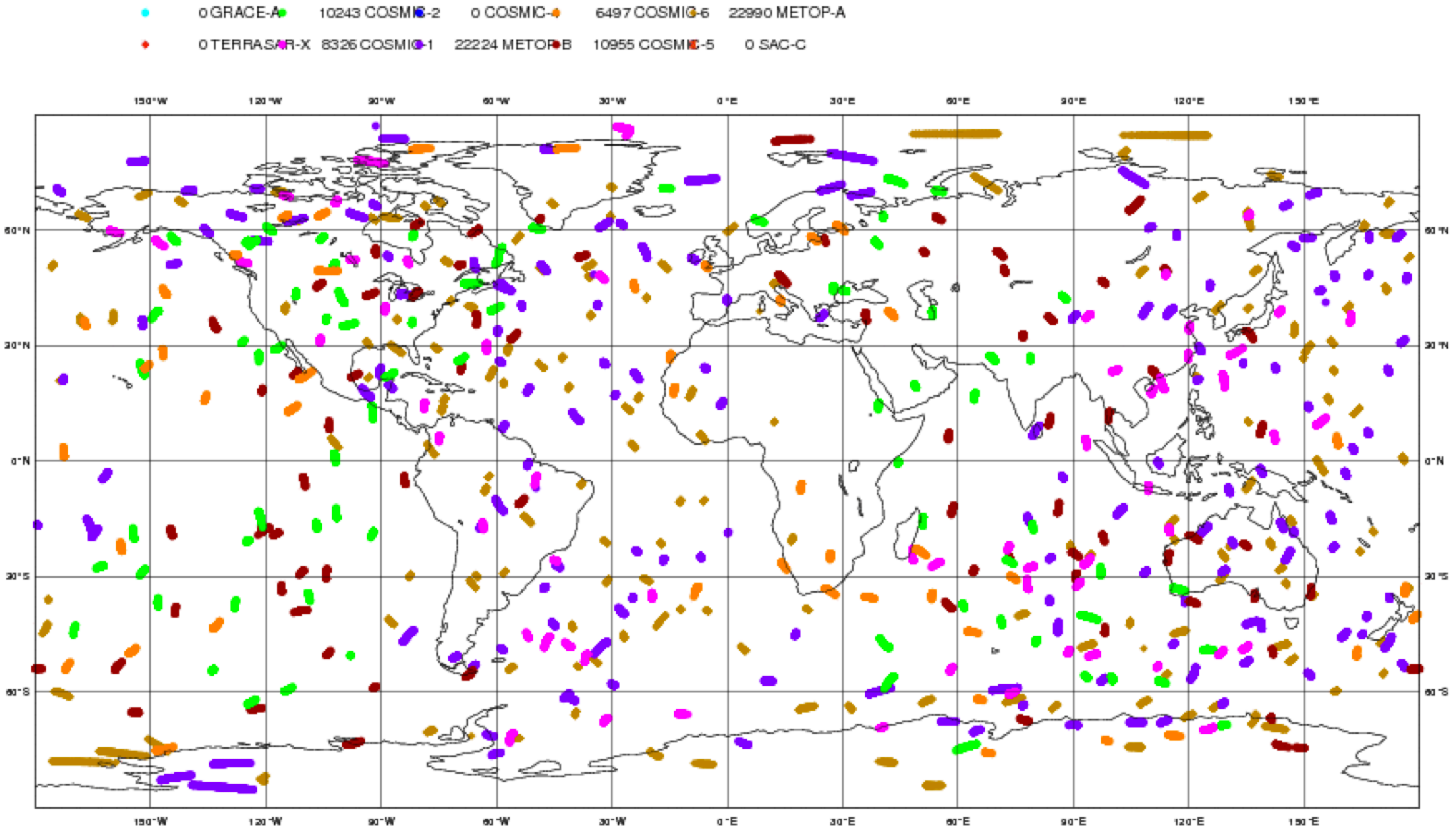
16/Apr/2014; 00 UTC

Total number of obs = 533106

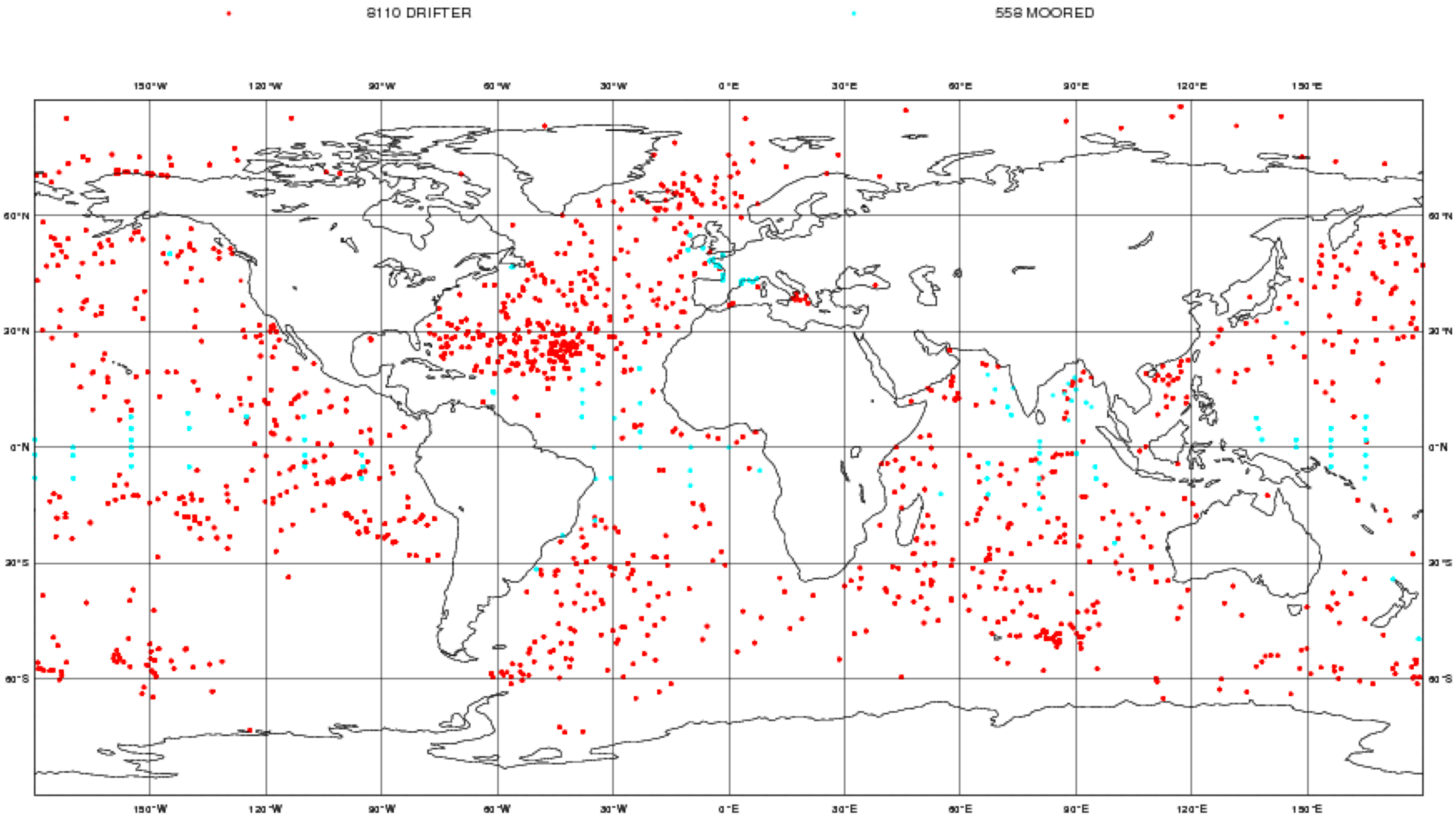
0 OSCAT · 291270 MetopAASCAT · 241836 MetopBASCAT



# ECMWF Data Coverage (All obs DA) - GPSRO
16/Apr/2014; 00 UTC
Total number of obs = 81235



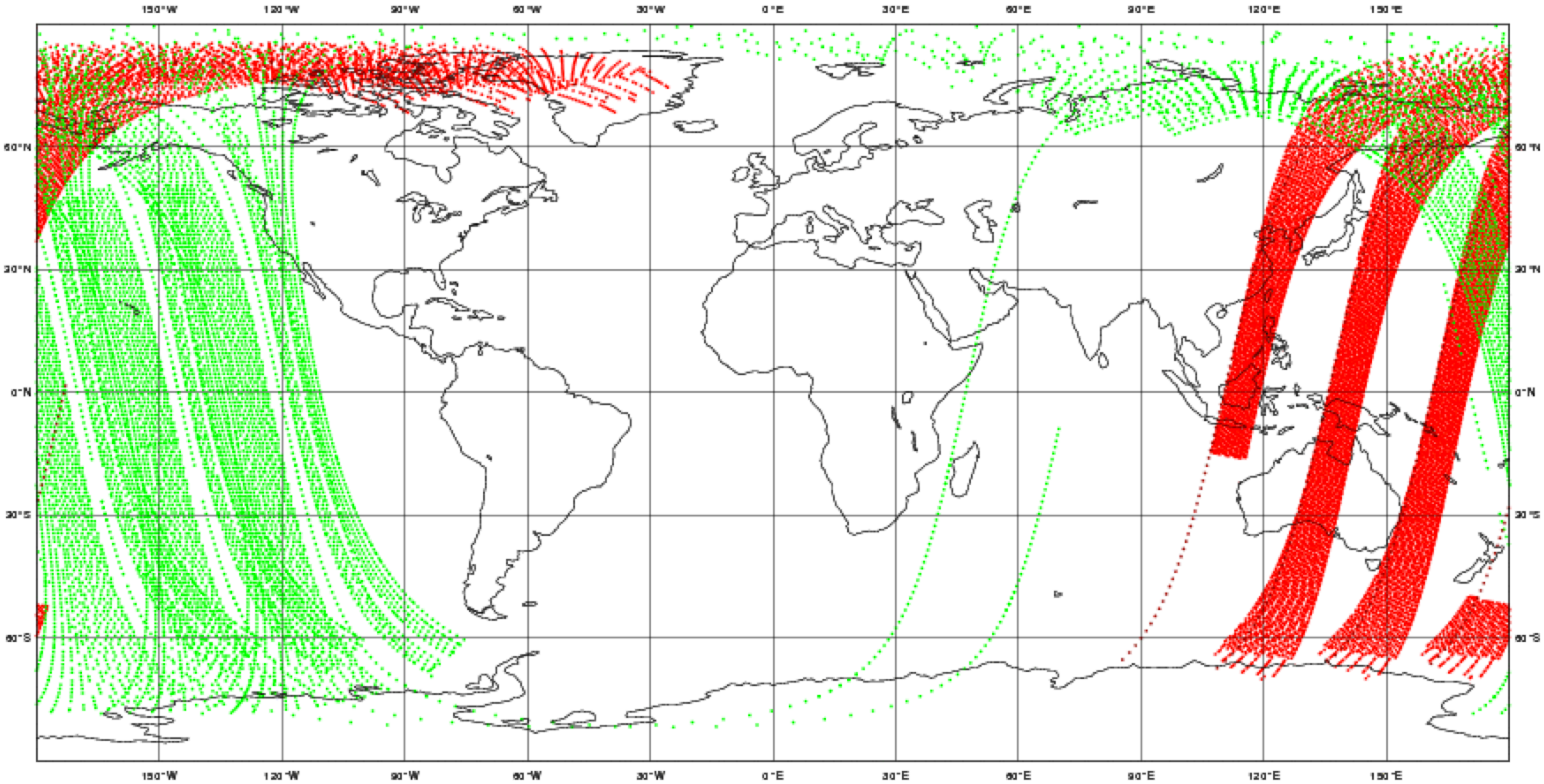
# ECMWF Data Coverage (All obs DA) - Buoy
16/Apr/2014; 00 UTC
Total number of obs = 8668



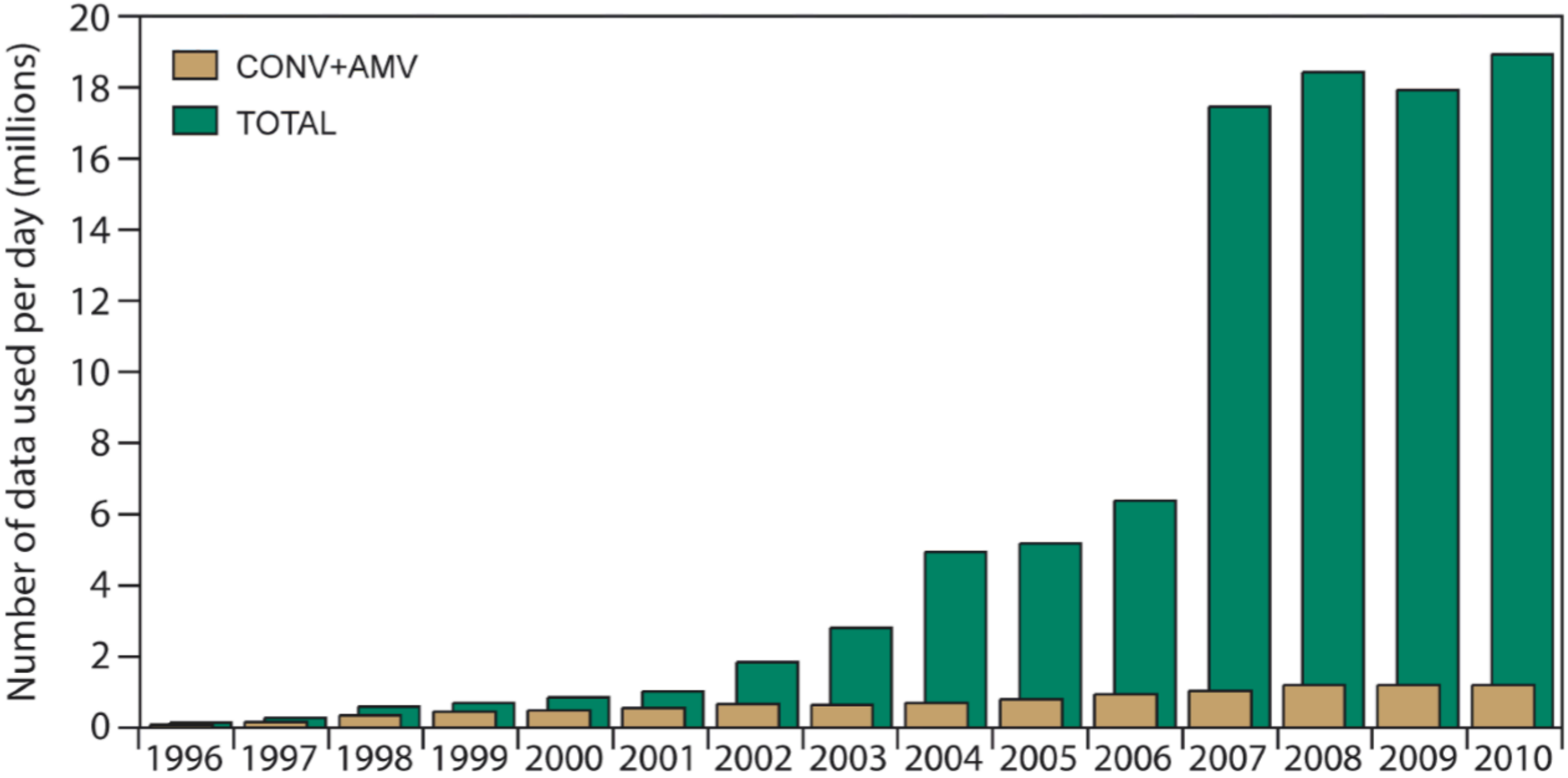
# ECMWF Data Coverage (All obs DA) - OZONE
## 16/Apr/2014; 00 UTC
## Total number of obs = 28045

16752 GOME-2 · 0 ENVISAT · 0 NOAA17 · 10975 AURA

0 ERS · 0 MET10 · 318 NOAA16 · 0 NOAA18 · 0 NOAA19



ECMWF




Value as of early 2013 : around 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 UTC)

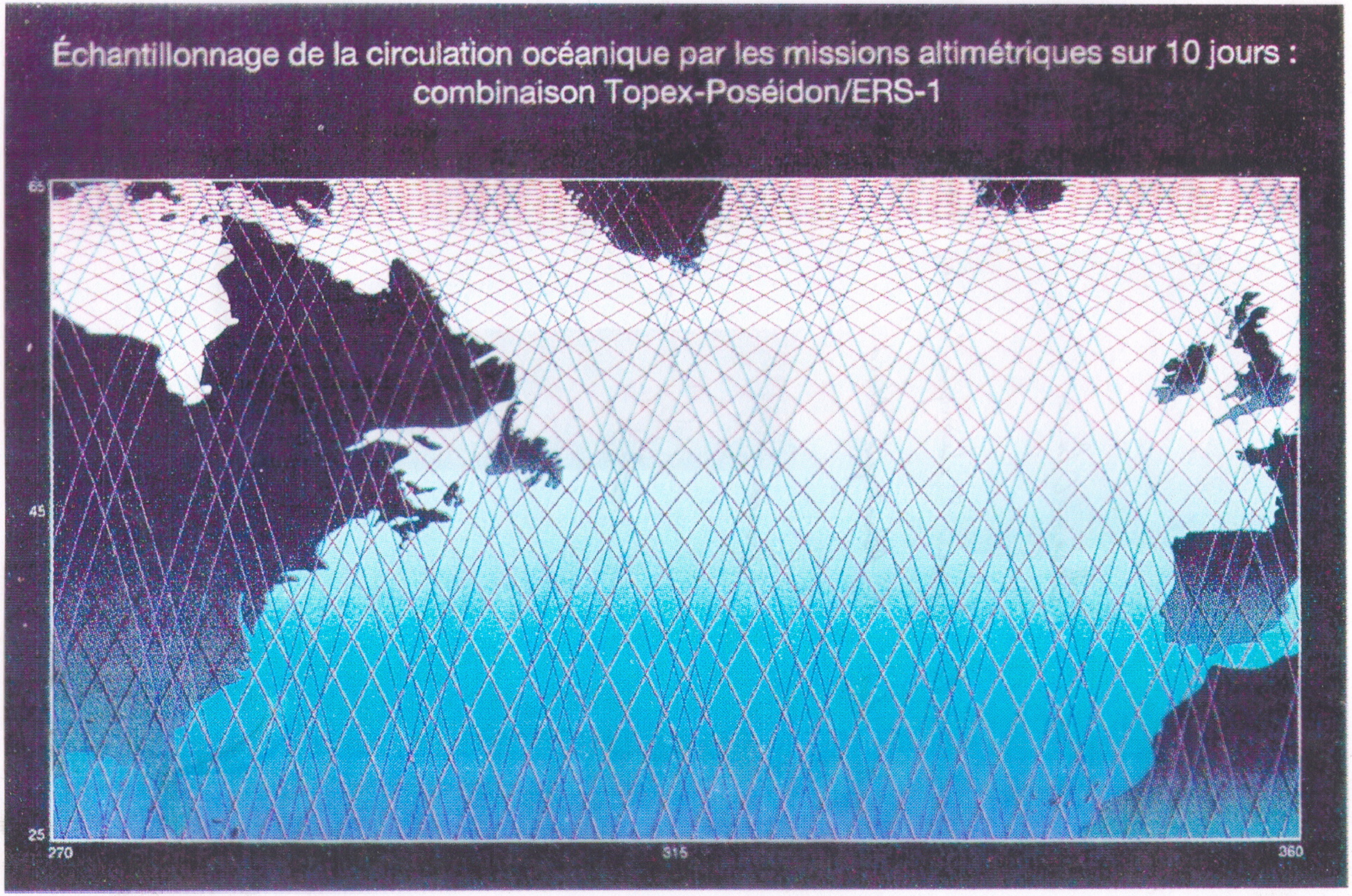
- Observations *asynoptiques* (satellites, avions), effectuées plus ou moins continûment dans le temps.

- 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

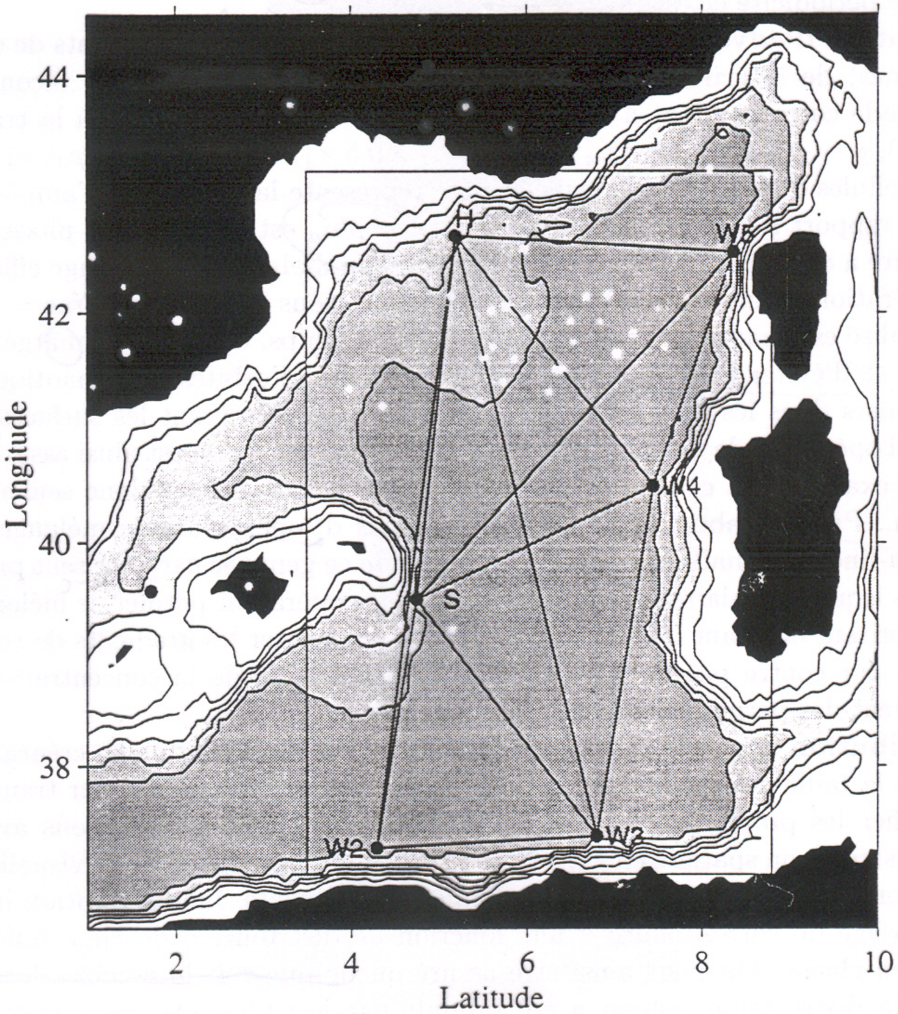



FIG. 1 – *Bassin méditerranéen occidental: réseau d'observation tomographique de l'expérience Thétis 2 et limites du domaine spatial utilisé pour les expériences numériques d'assimilation.*

E. Rémy, Doctoral Dissertation, 1999

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, $p \approx$ 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

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 2011, J.-N. Thépaut) :

4DVAR: 17%

EDA: 15%

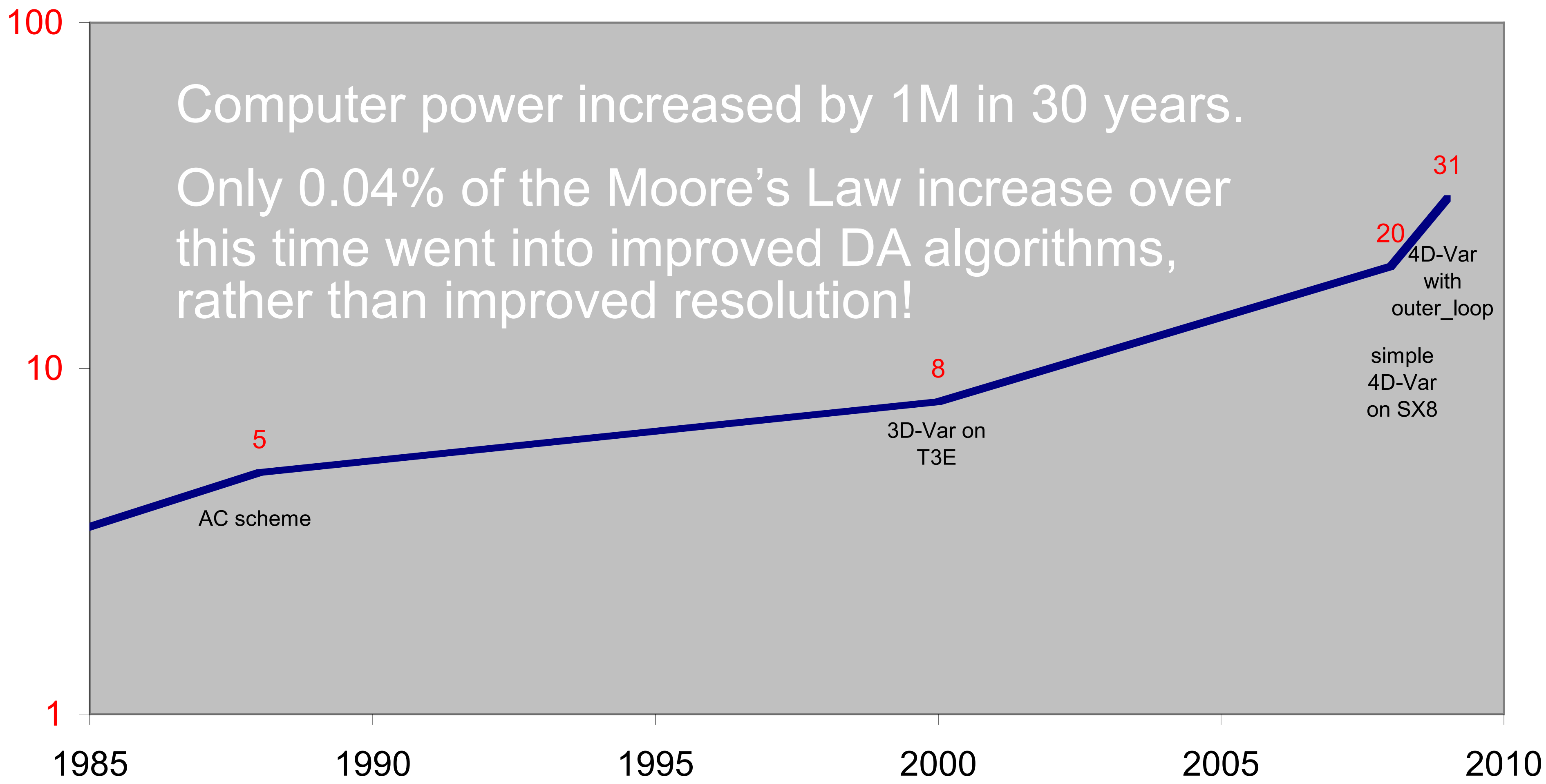
Modèle déterministe: 13%

EPS: 53%

autre: 2%

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

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

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

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

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

- $\zeta_1$ and $\zeta_2$ mutually independent

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

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

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

$\zeta_1$ and $\zeta_2$ mutually independent

$x = \xi \;\Leftrightarrow\; \zeta_1 = z_1 - \xi$ and $\zeta_2 = z_2 - \xi$

- $P(x = \xi \mid z_1, z_2) \propto p_1(z_1 - \xi)\, p_2(z_2 - \xi)$

$$\propto \exp[ - (\xi - x^a)^2/2p^a]$$

where $1/p^a = 1/s_1 + 1/s_2$ , $x^a = p^a (z_1/s_1 + z_2/s_2)$

Conditional probability distribution of $x$, given $z_1$ and $z_2$ : $\mathcal{N}[x^a, p^a]$

$p^a < (s_1, s_2)$ independent of $z_1$ and $z_2$

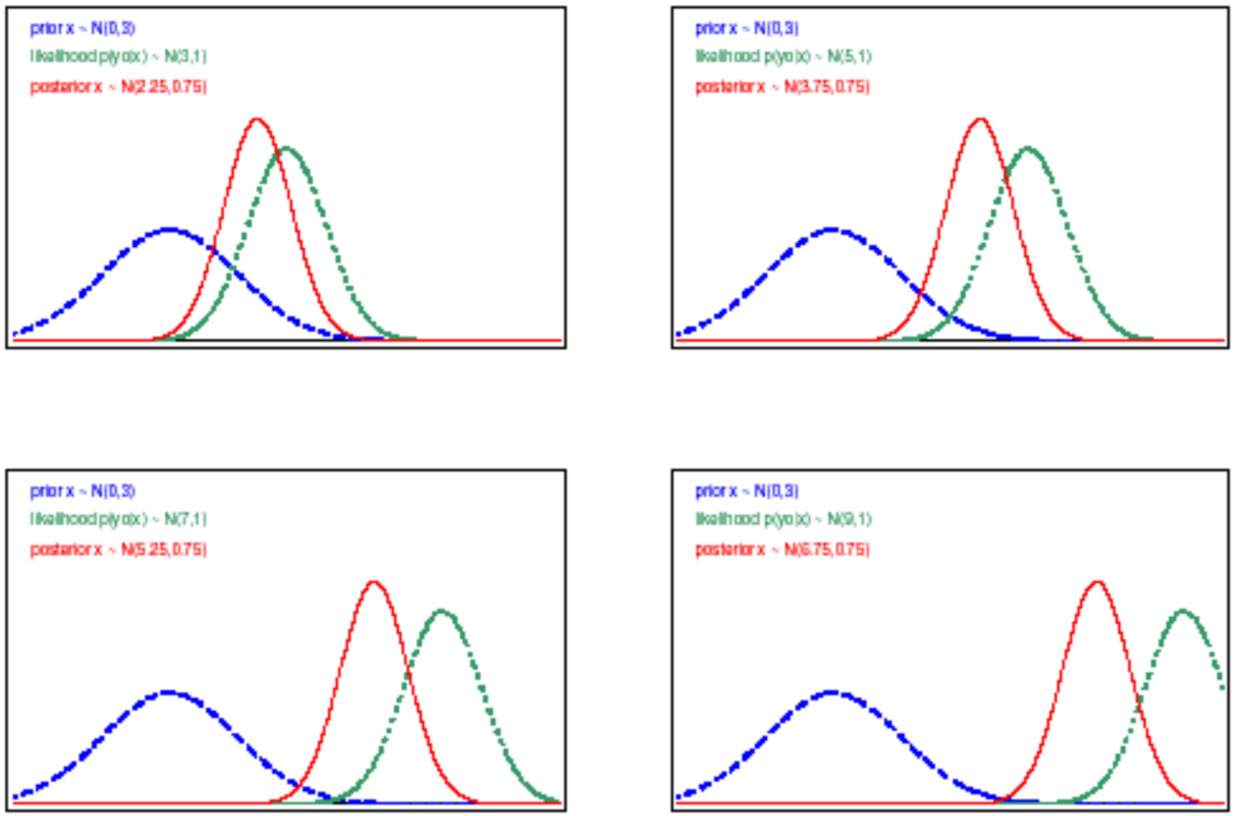



Fig. 1.1: Prior pdf $p(x)$ (dashed line), posterior pdf $p(x|y^o)$ (solid line), and Gaussian likelihood of observation $p(y^o|x)$ (dotted line), plotted against $x$ for various values of $y^o$. (Adapted from Lorenc and Hammon 1988.)

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

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

In addition

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

Conditional probability distribution in Gaussian case

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

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

Same as before, but $\zeta_1$ and $\zeta_2$ 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 $\infty$ as $\delta$ increases from $0$ to $\infty$. 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 $\zeta$ 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 $\xi$ ?

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

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

Unambiguously defined iff, for any $\zeta$, 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 $N \approx O(10\text{-}100)$).