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

Année Universitaire 2022-2023

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 2

28 Mars 2023

- Numerical Weather Prediction. Present performance (mostly ECMWF)

- The meteorological observation system

- The problem of ‘Assimilation’

- Inverse Problems. Bayesian Estimation

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

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

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

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

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

Large-scale Numerical Weather Prediction is based on the *primitive equations*, themselves based on a number of simplifications, and particularly the hydrostatic approximation

Climatic simulations are also built on primitive equations, and contain a much more detailed description of the oceanic circulation.

More costly nonhydrostatic models are used for small scale meteorology, and are being developed for global modeling.

- European Centre for Medium-Range Weather Forecasts (ECMWF)

- Centre européen pour les prévisions météorologiques à moyen terme (CEPMMT)

- Europäisches Zentrum für mittelfristige Wettervorhersage (EZMW)

As of 2023, 23 member states, 12 co-operating states

ECMWF established in 1975. Has produced daily forecasts since 1980

Headquarters in Reading (UK), Data Centre in Bologna (Italy)

ECMWF hosts part EU's Earth Observation *Copernicus* programme. Moved to Bonn (Germany)

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

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

Depuis mars 2016 :

Troncature triangulaire TCO1279 / O1280 (résolution horizontale ≈ 9 kilomètres)

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

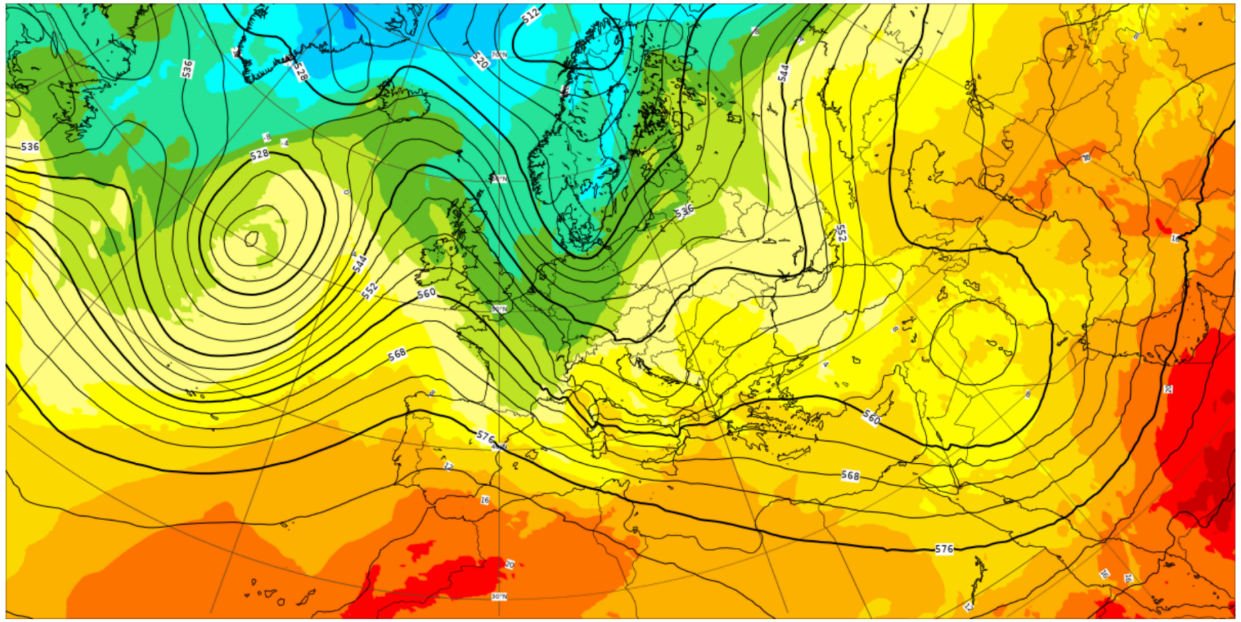
Discrétisation en éléments finis dans la direction verticale (coordonnée hybride)

Dimension du vecteur d'état correspondant > 10^9

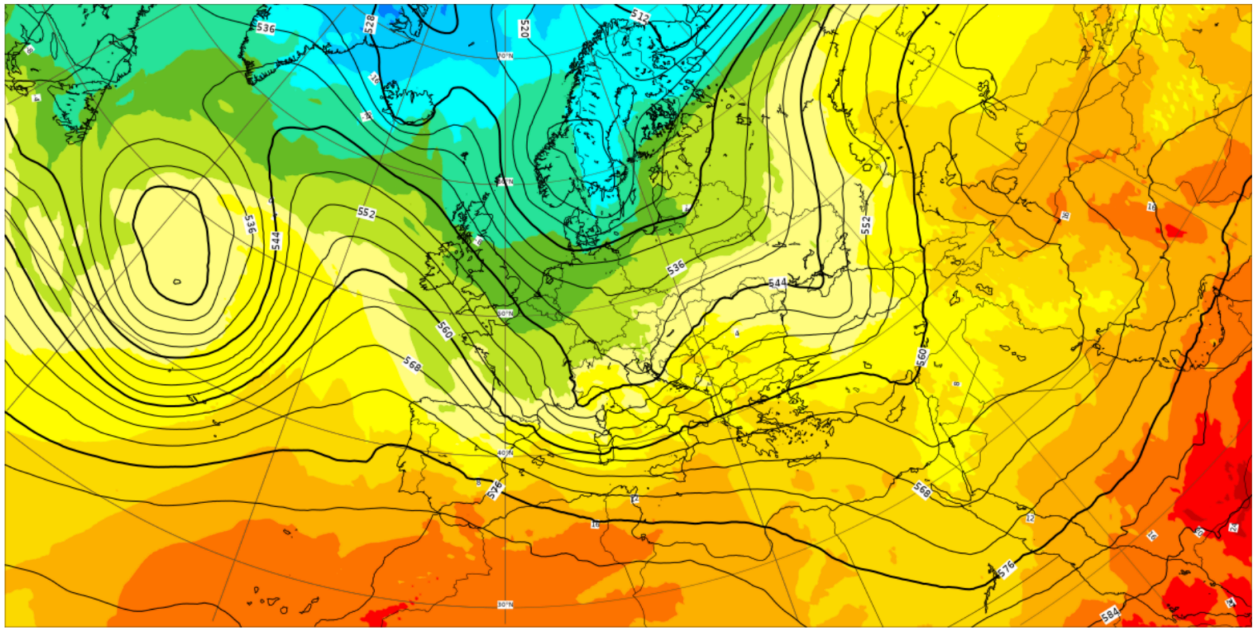
Pas de discrétisation temporelle (schéma semi-Lagrangien semi-implicite): 450 secondes

Intégré 2 fois par jour (00 et 12 UTC) à une échéance de 10 jours

Base time: Mon 20 Mar 2023 00 UTC Valid time: Mon 27 Mar 2023 00 UTC (+168h) Area : Europe



Base time: Mon 27 Mar 2023 00 UTC Valid time: Mon 27 Mar 2023 00 UTC (+0h) Area : Europe

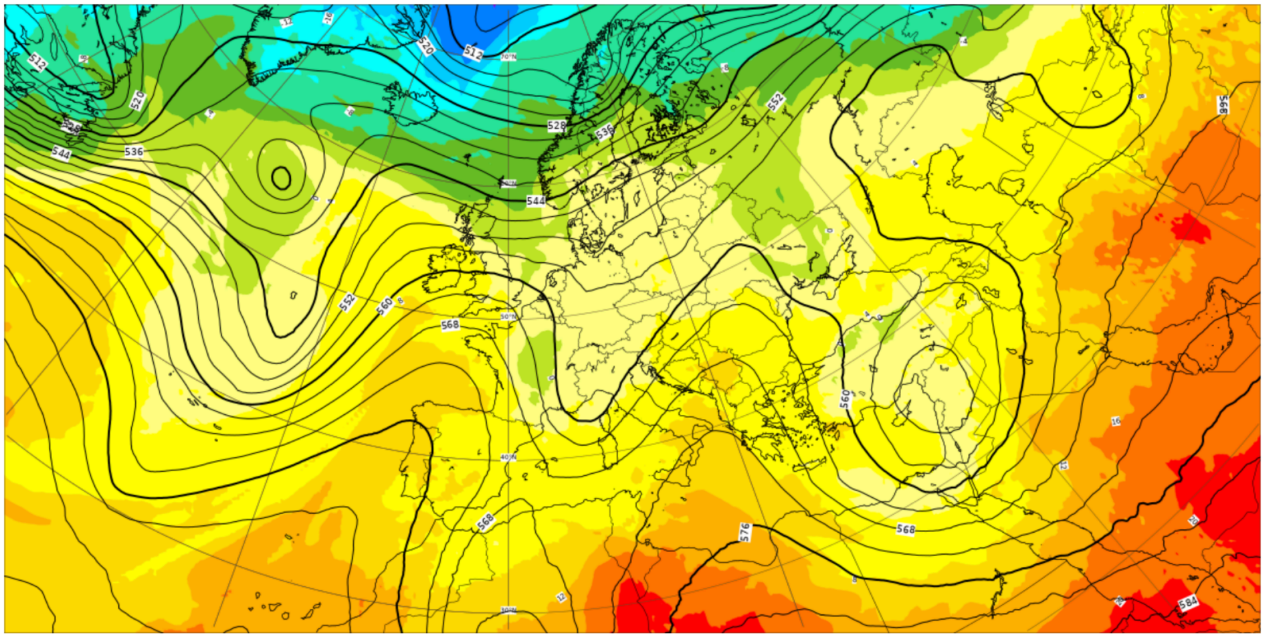


850 hPa temperature (C)

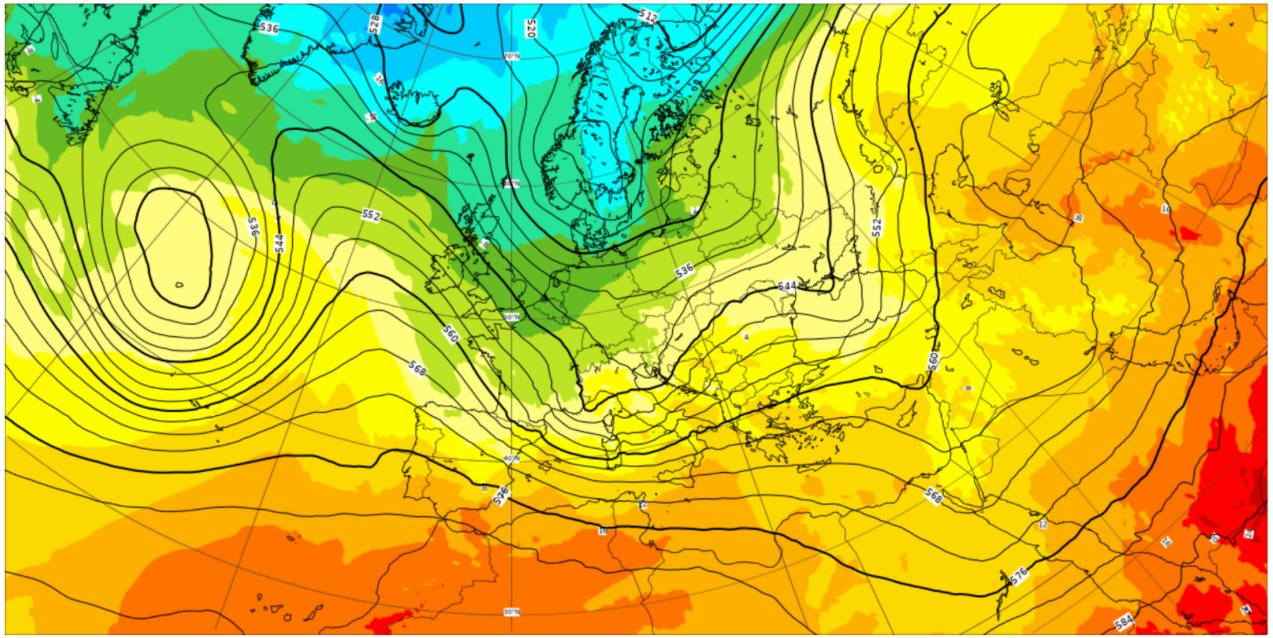
56 52 48 44 40 36 32 28 24 20 16 12 8 4 0 -4 -8 -12 -16 -20 -24 -28 -32 -36 -40 -44 -48 -52 -60 -70 -80

500 hPa geopotential (dm)

Base time: Mon 20 Mar 2023 00 UTC Valid time: Mon 20 Mar 2023 00 UTC (+0h) Area : Europe



Base time: Mon 27 Mar 2023 00 UTC Valid time: Mon 27 Mar 2023 00 UTC (+0h) Area : Europe

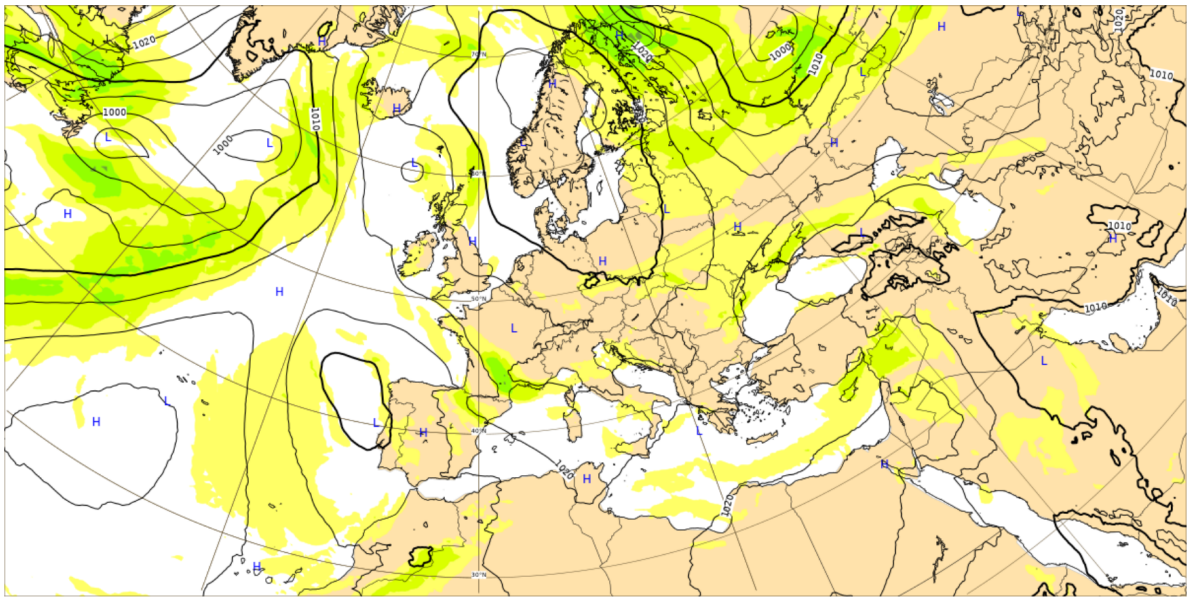


850 hPa temperature (C)

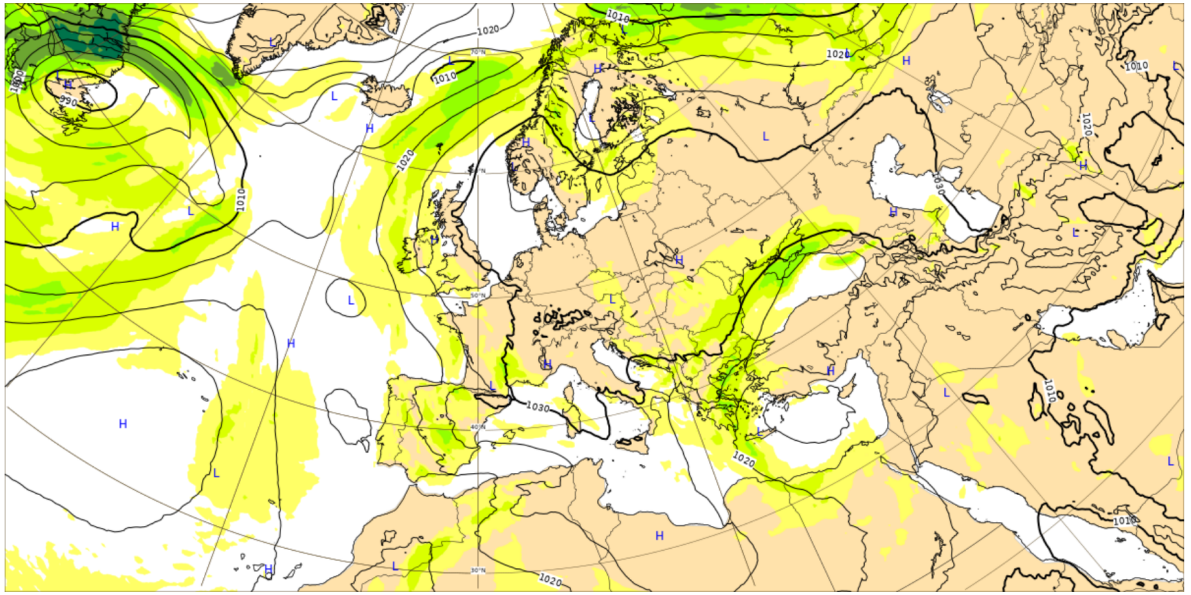
56 52 48 44 40 36 32 28 24 20 16 12 8 4 0 -4 -8 -12 -16 -20 -24 -28 -32 -36 -40 -44 -48 -52 -60 -70 -80

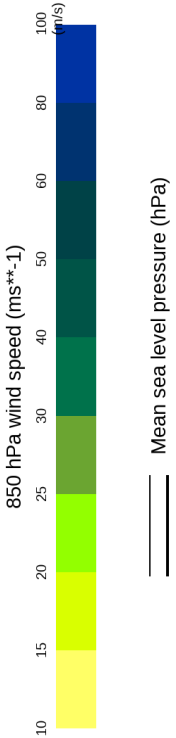
500 hPa geopotential (dm)

Base time: Wed 16 Mar 2022 00 UTC Valid time: Wed 23 Mar 2022 00 UTC (+168h) Area : Europe

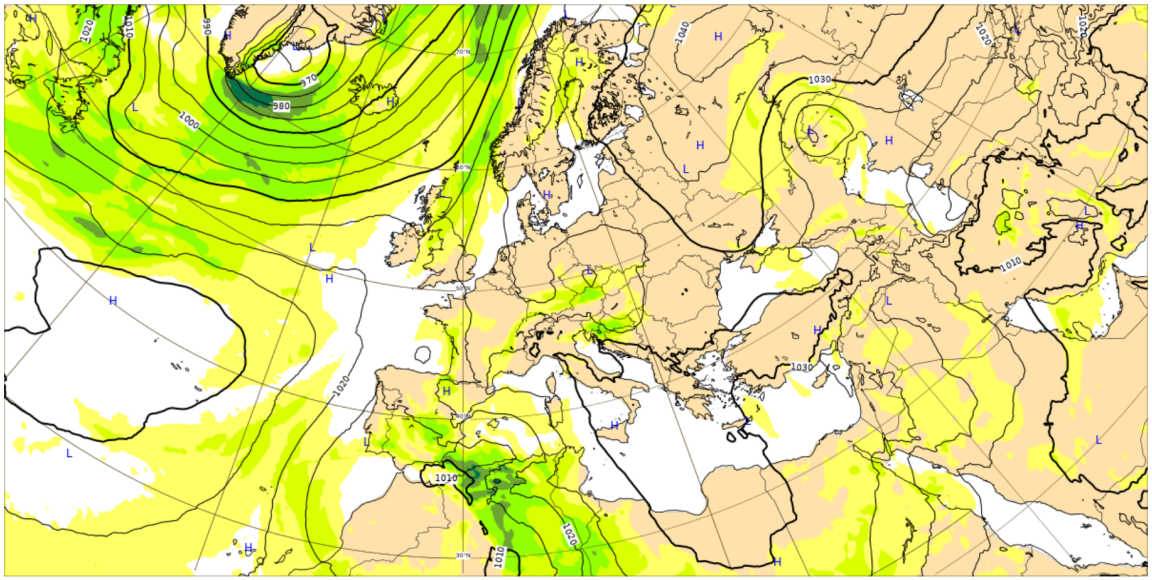


Base time: Wed 23 Mar 2022 00 UTC Valid time: Wed 23 Mar 2022 00 UTC (+0h) Area : Europe

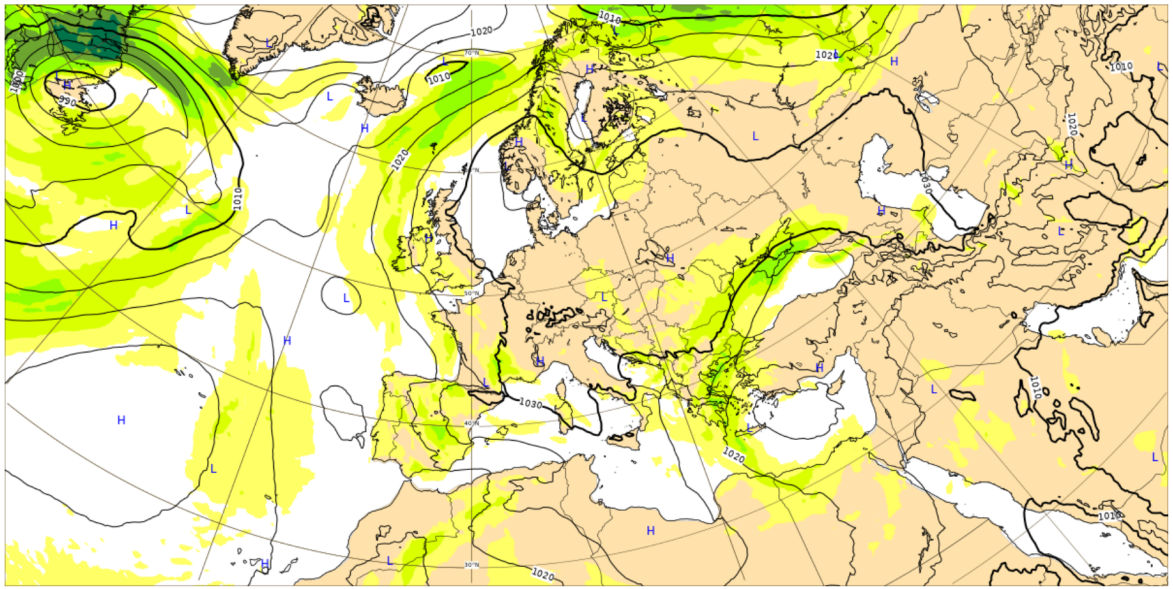




Base time: Wed 16 Mar 2022 00 UTC Valid time: Wed 16 Mar 2022 00 UTC (+0h) Area : Europe



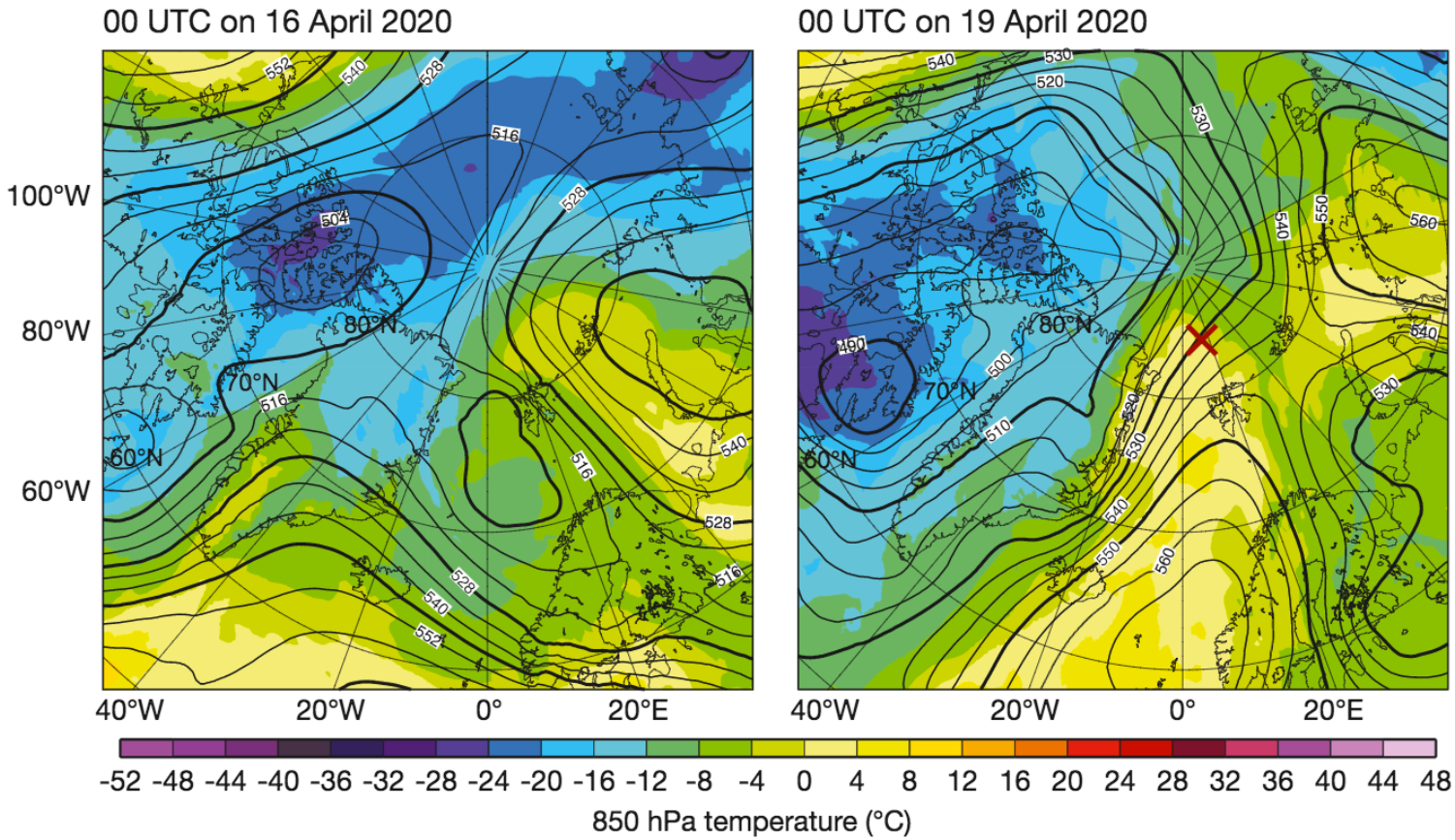
Base time: Wed 23 Mar 2022 00 UTC Valid time: Wed 23 Mar 2022 00 UTC (+0h) Area : Europe



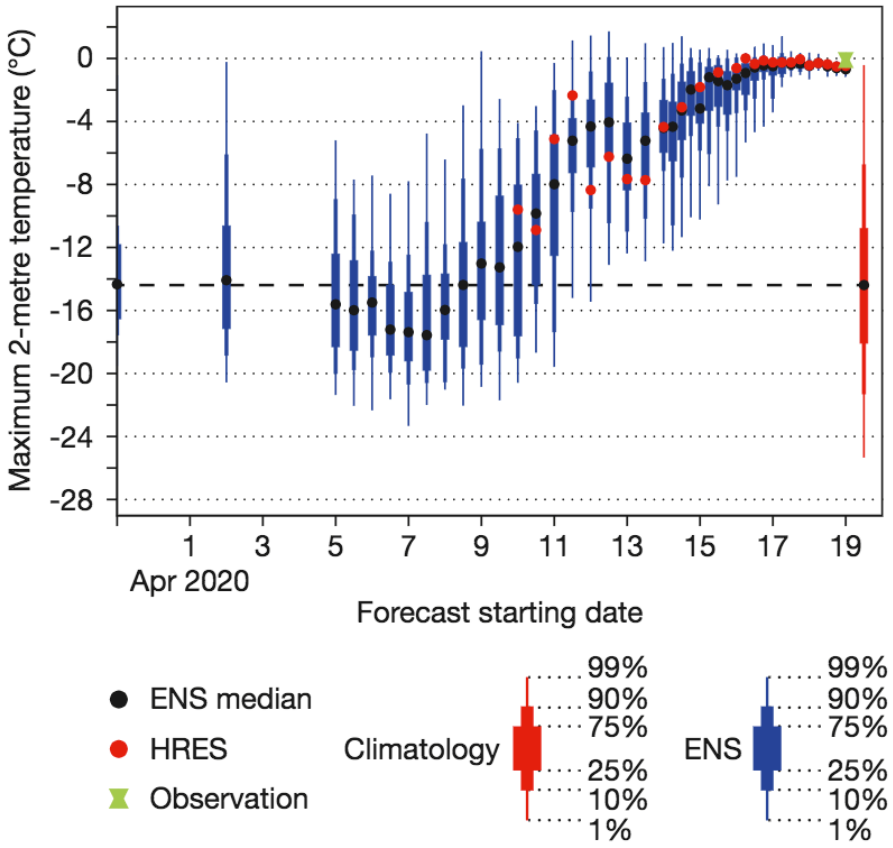
850 hPa wind speed (ms**-1)

10 15 20 25 30 40 50 60 80 100 (m/s)

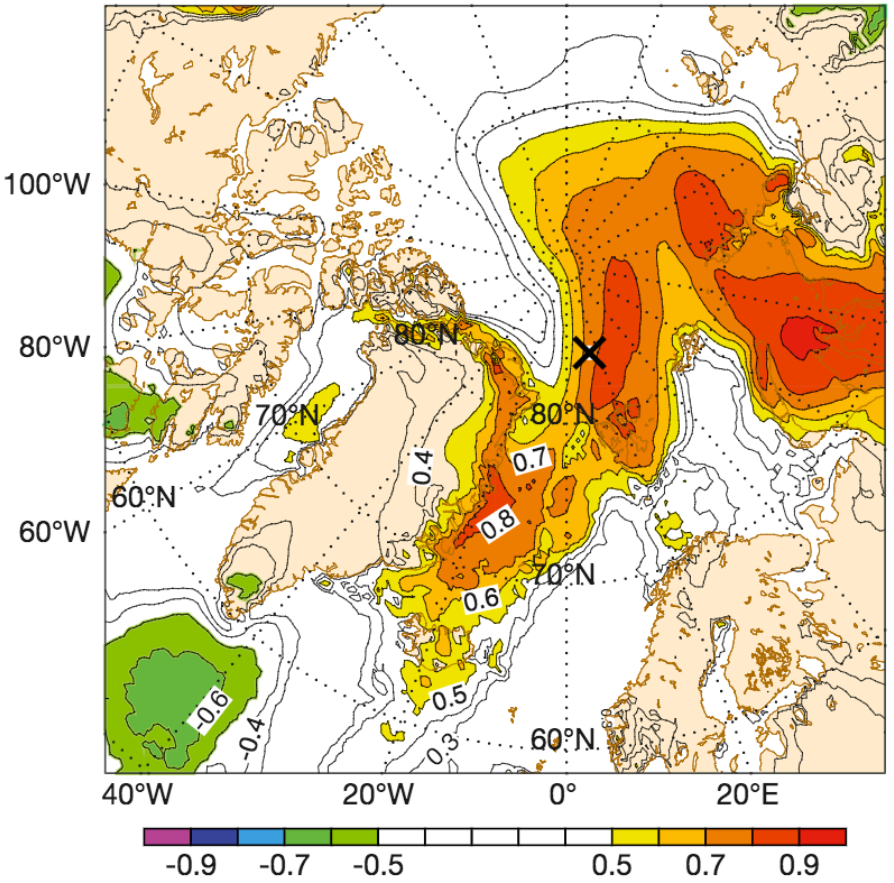
Mean sea level pressure (hPa)



Synoptic situation on 16 and 19 April. Analysis of geopotential height at 500 hPa (contours) and temperature at 850 hPa (shading) for 00 UTC on 16 April 2020 (left) and 00 UTC on 19 April (right). The cross shows the approximate location of Polarstern on 19 April.



Evolution of forecasts for the 19 April warm air intrusion. The plot shows ensemble forecasts with different starting times for maximum 2-metre temperature for the Polarstern location on 19 April.



Two-metre maximum temperature EFI. The chart shows the 5-day forecast from 00 UTC on 15 April 2020 of the EFI for 2-metre maximum temperature on 19 April.

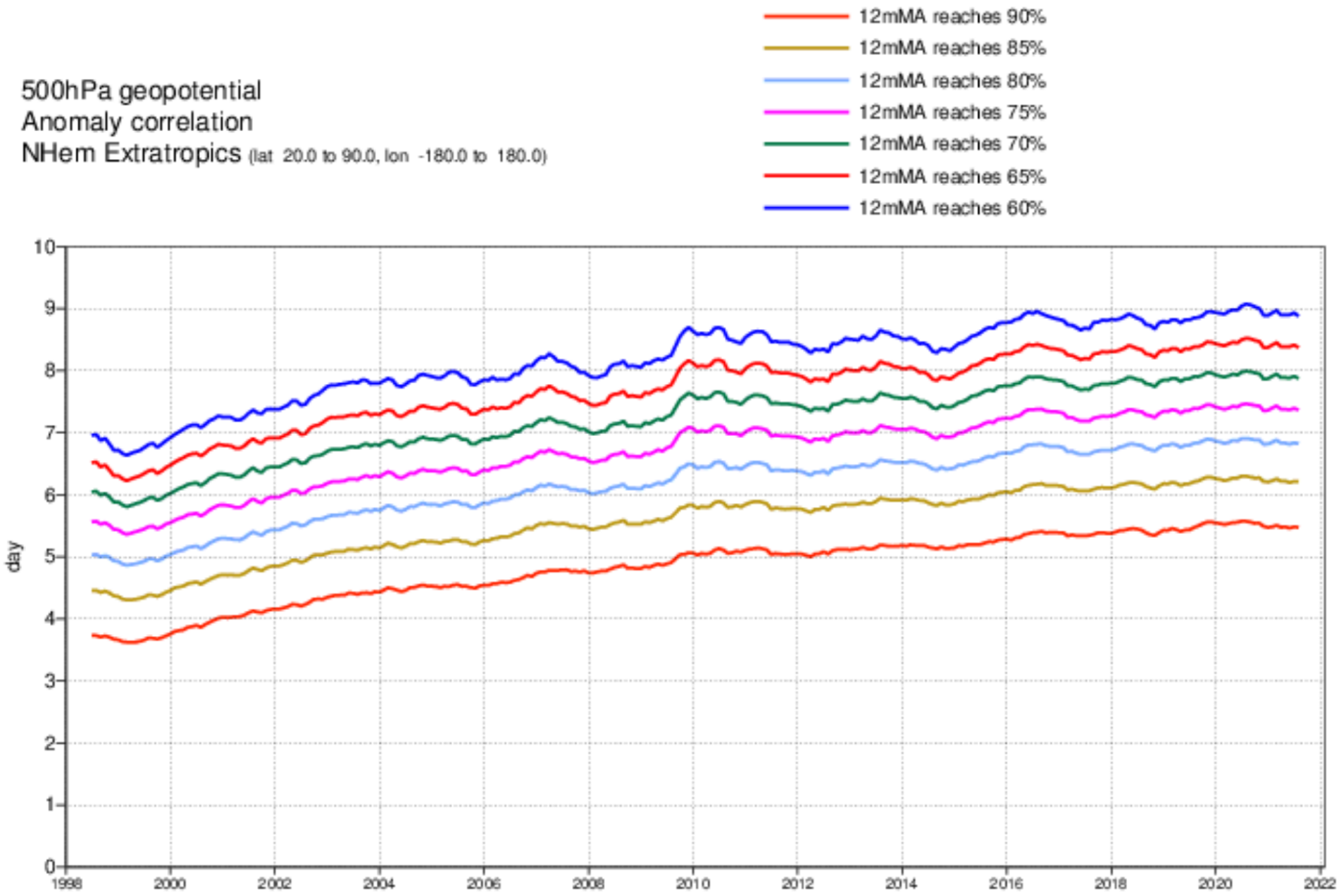
Results on site of ECMWF

In particular

- 09/2022. T. Haiden *et al.*, 2021, *Evaluation of ECMWF forecasts, including the 2021 upgrade*, Technical Memorandum 902, ECMWF, Reading, UK.

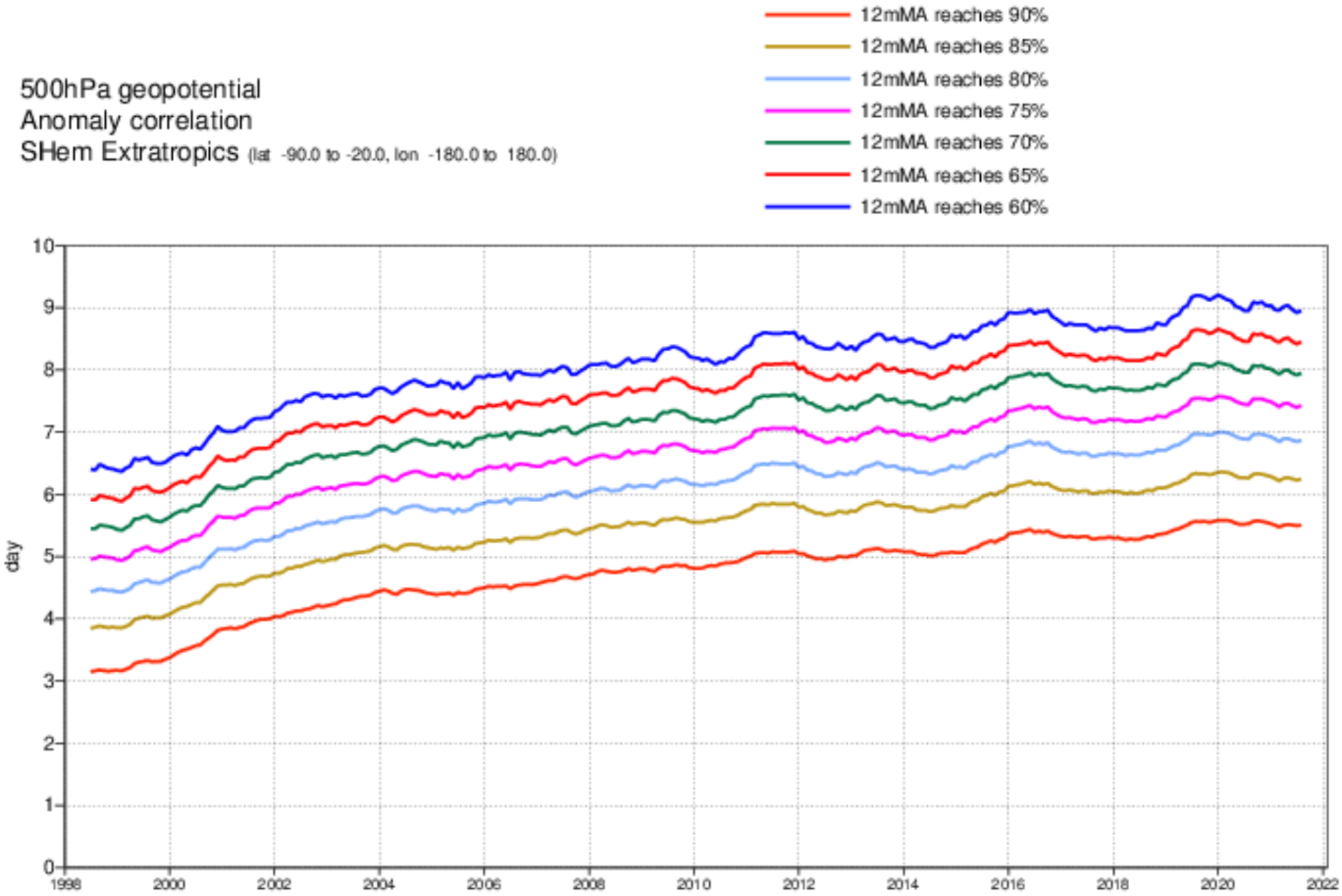
Available at the address :

https://www.ecmwf.int/en/elibrary/81235-evaluation-ecmwf-forecasts-including-2021-upgrade



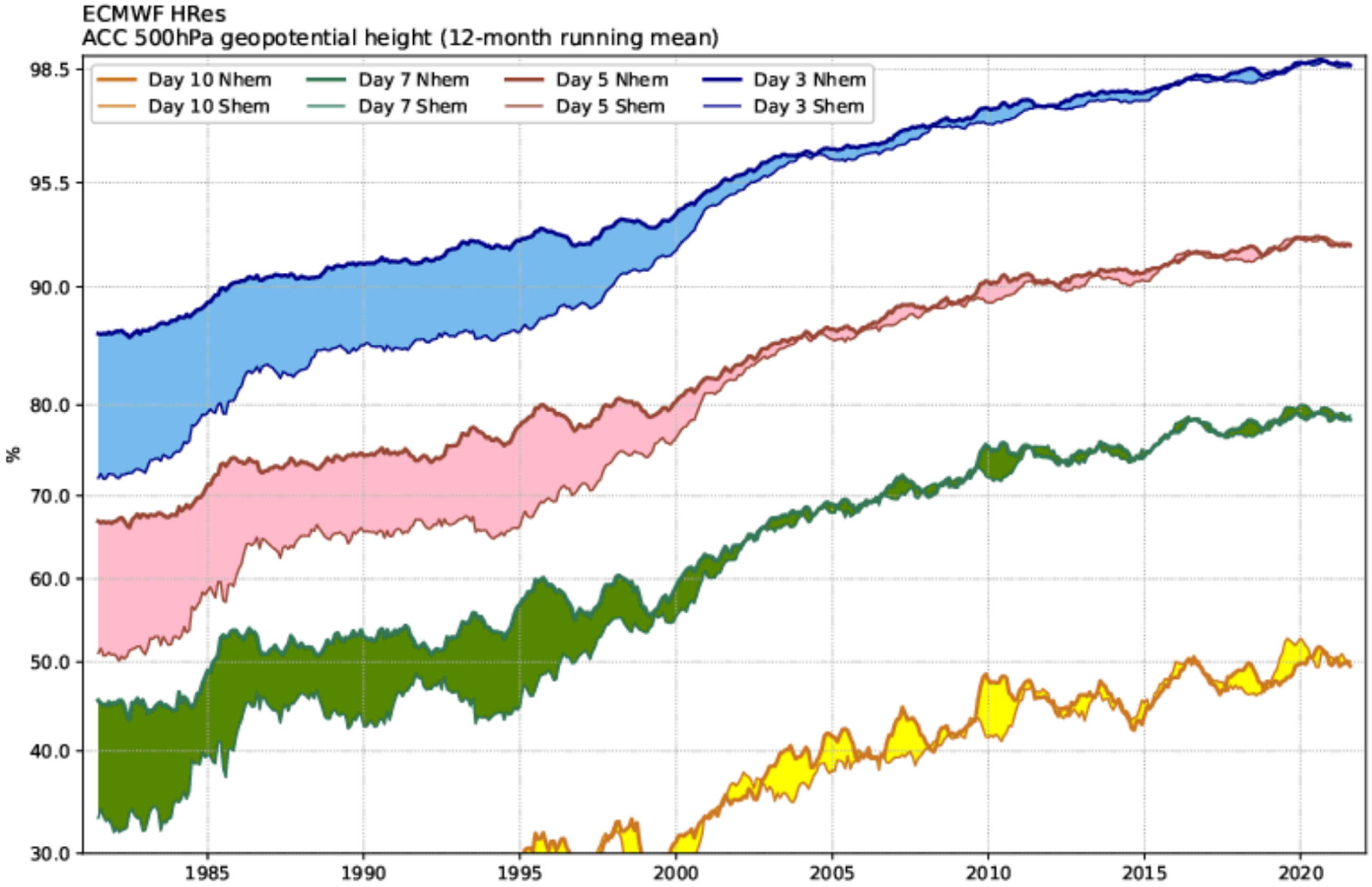
Lead time ACC reaching thresholds

Spatial correlation between anomalies from climatology of forecast and verifying analysis



Lead time ACC reaching thresholds

Spatial correlation between anomalies from climatology of forecast and verifying analysis


ECMWF HRes
ACC 500hPa geopotential height (12-month running mean)
Day 10 Nhem
Day 10 Shem
Day 7 Nhem
Day 7 Shem
Day 5 Nhem
Day 5 Shem
Day 3 Nhem
Day 3 Shem
%
98.5
95.5
90.0
80.0
70.0
60.0
50.0
40.0
30.0
1985
1990
1995
2000
2005
2010
2015
2020

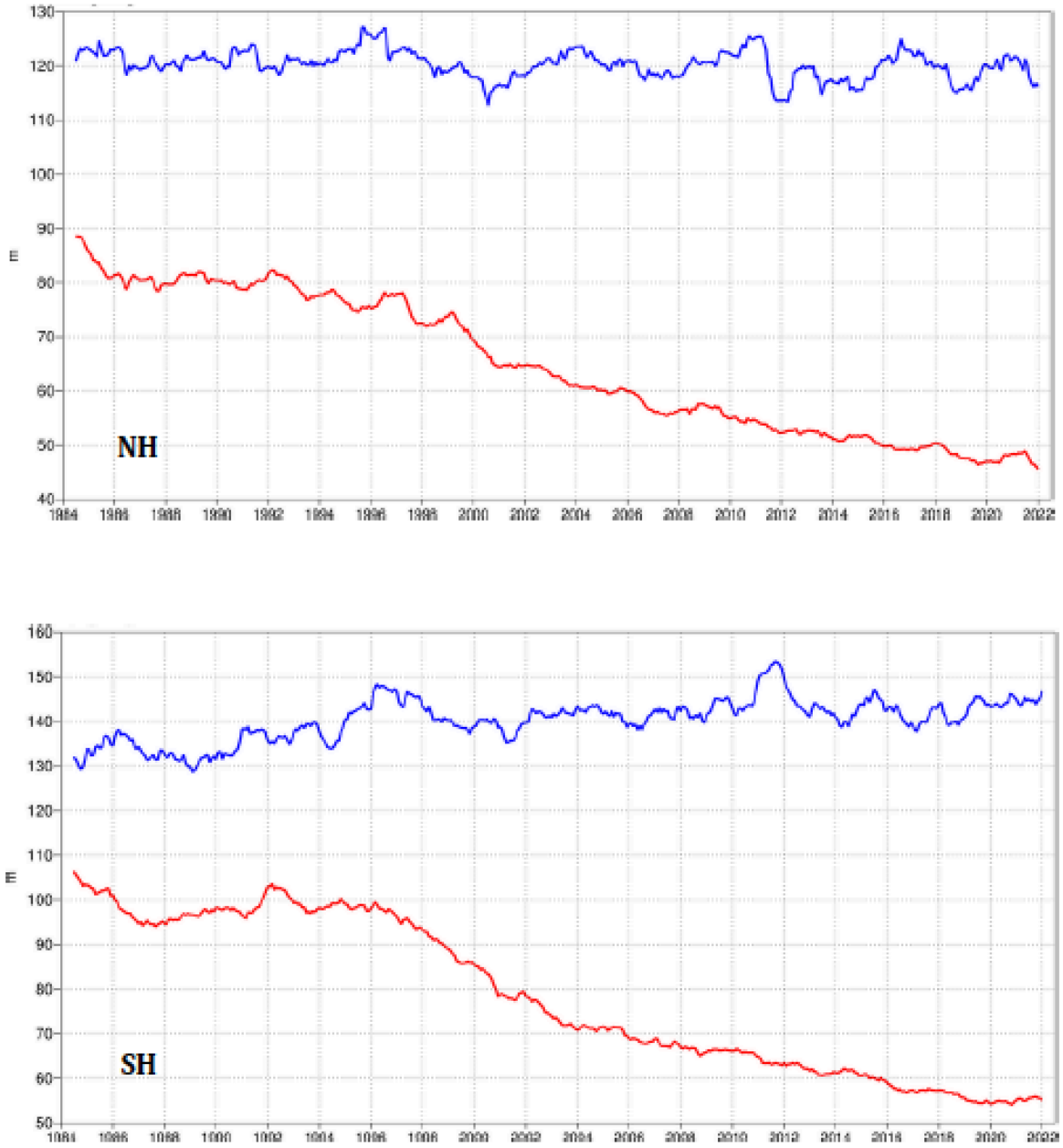
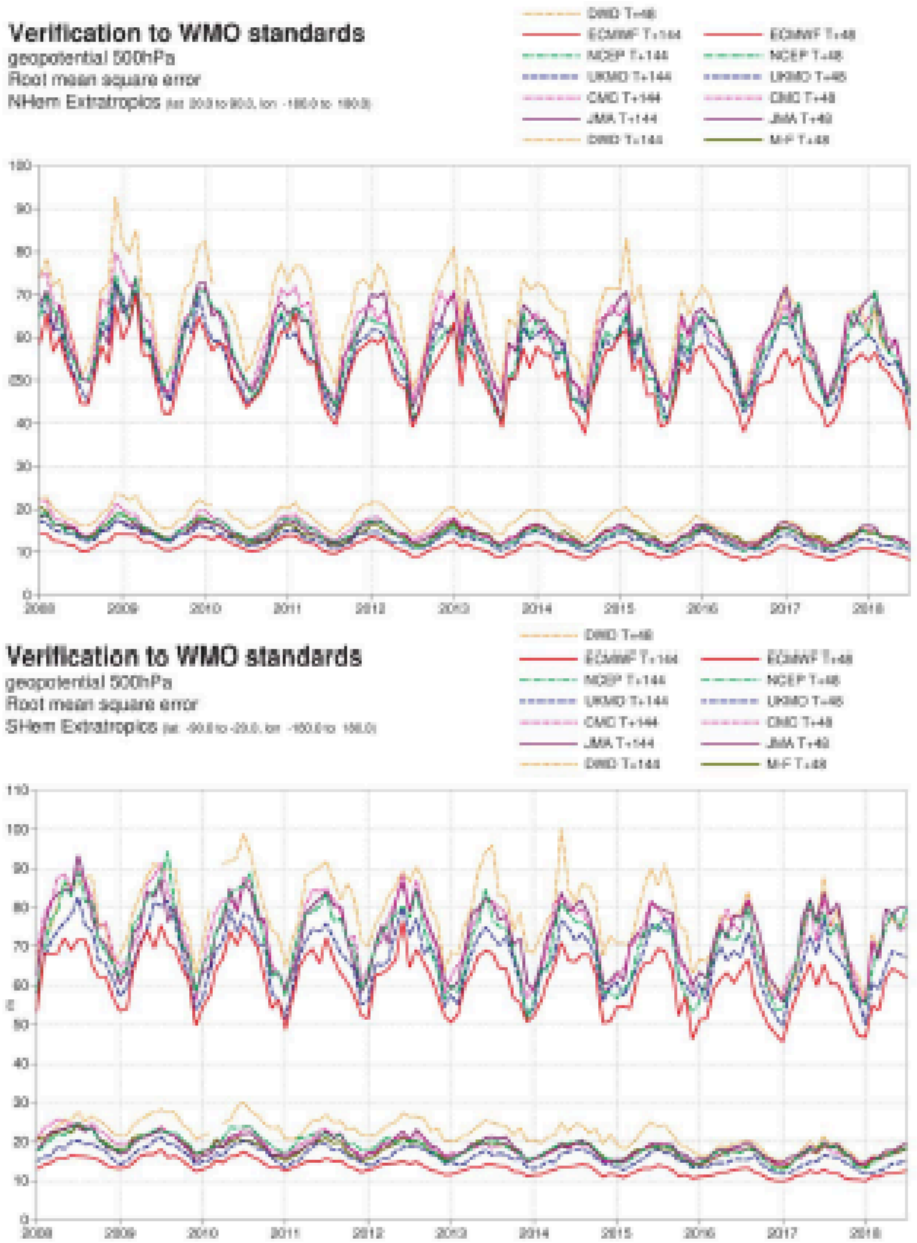
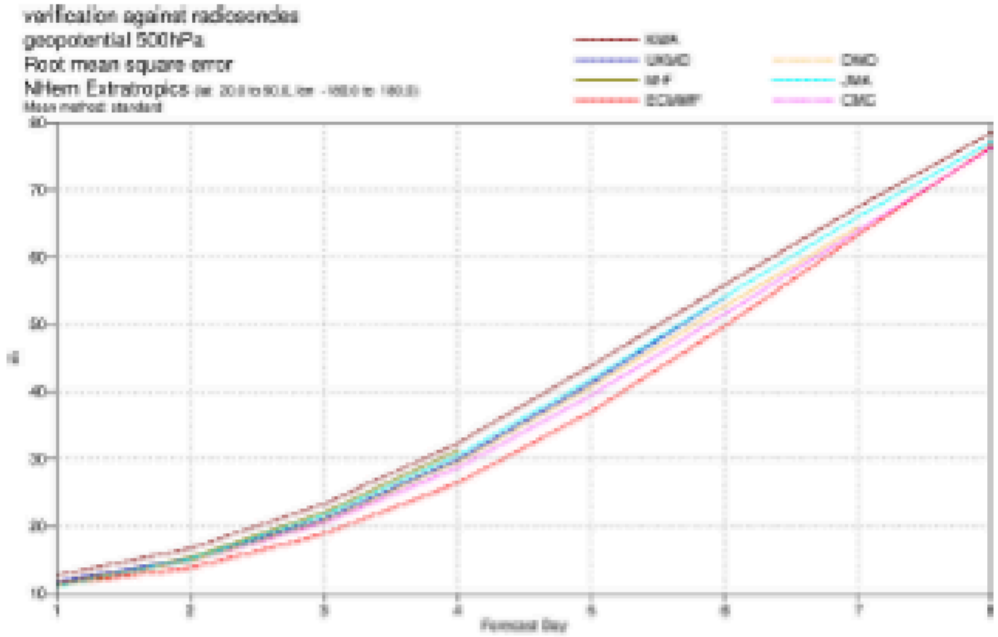


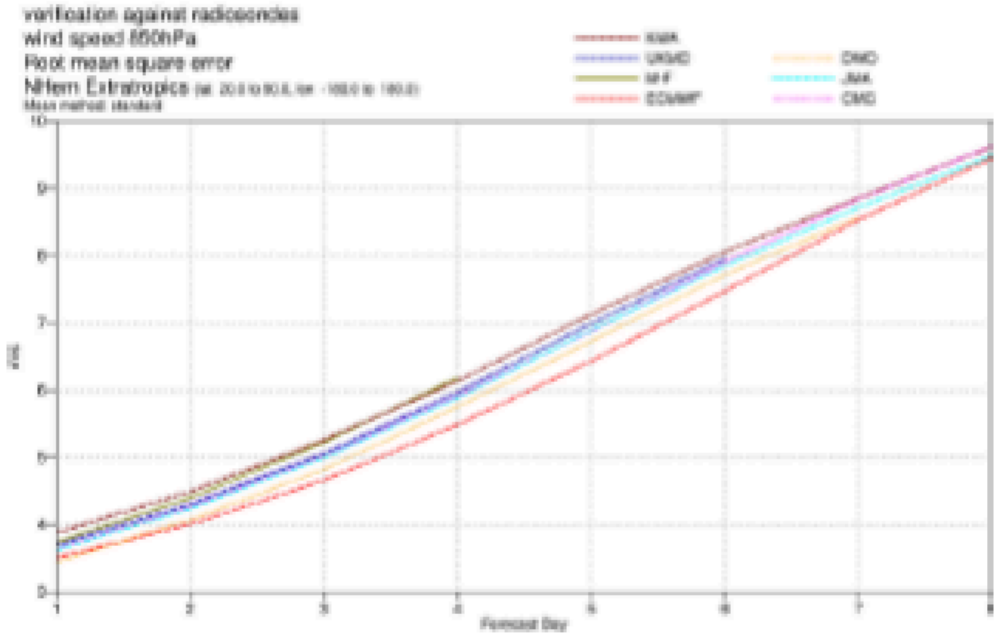
Figure 5: Root mean square (RMS) error of forecasts of 500 hPa geopotential height (m) at day 6 (red), verified against analysis. For comparison, a reference forecast made by persisting the analysis over 6 days is shown (blue). Plotted values are 12-month moving averages; the last point on the curves is for the 12-month period August 2021–July 2022. Results are shown for the northern extra-tropics (top), and the southern extra-tropics (bottom).

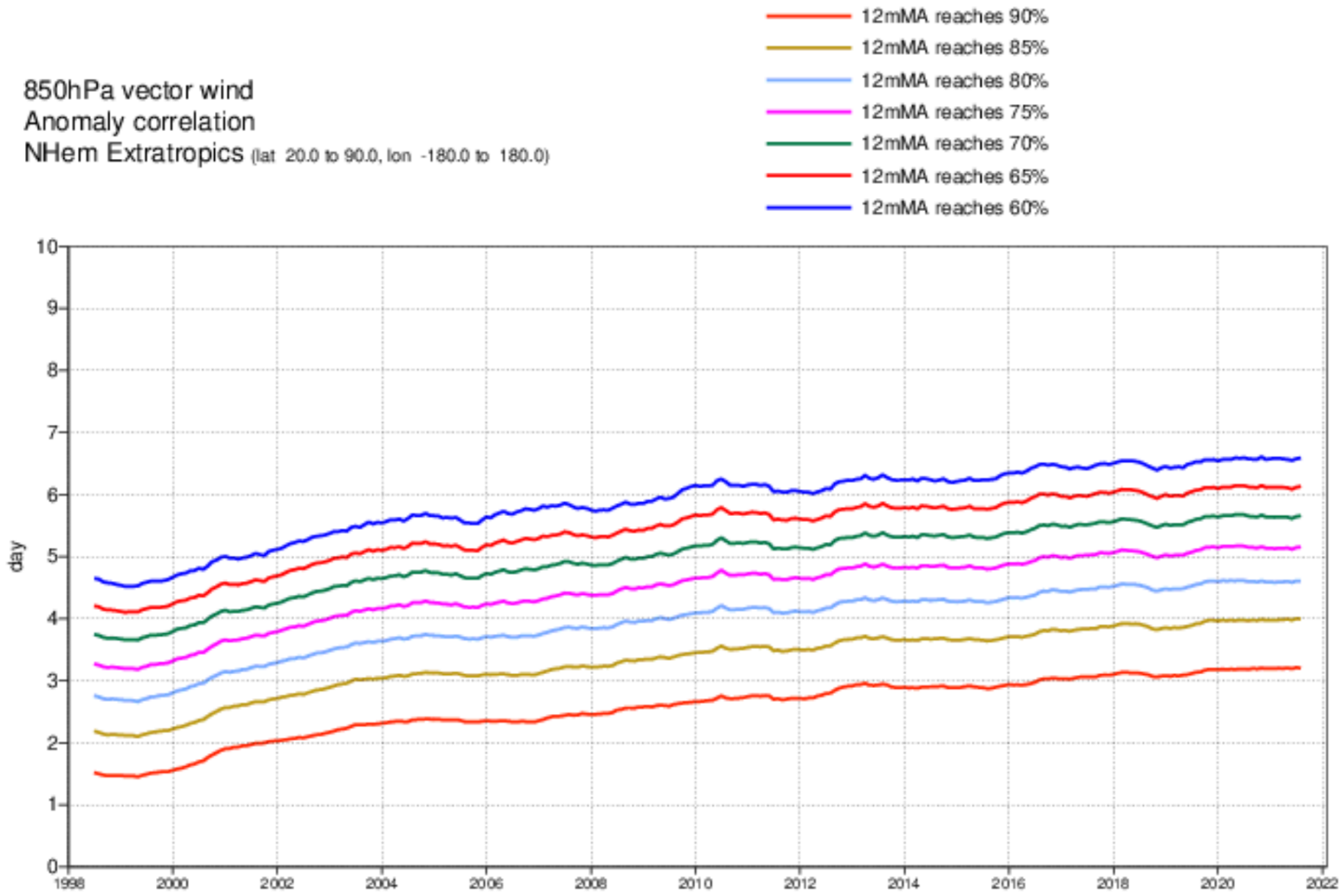


2020

Figure 14: WMO-exchanged scores from global forecast centres. RMS error of 500 hPa geopotential height over northern (top) and southern (bottom) extratropics. In each panel, the upper curves show the six-day forecast error and the lower curves show the two-day forecast error of model runs initiated at 12 UTC. Each model is verified against its own analysis. JMA = Japan Meteorological Agency, CMC = Canadian Meteorological Centre, UKMO = the UK Met Office, NCEP = U.S. National Centers for Environmental Prediction, M-F = Météo France, DWD = Deutscher Wetterdienst.

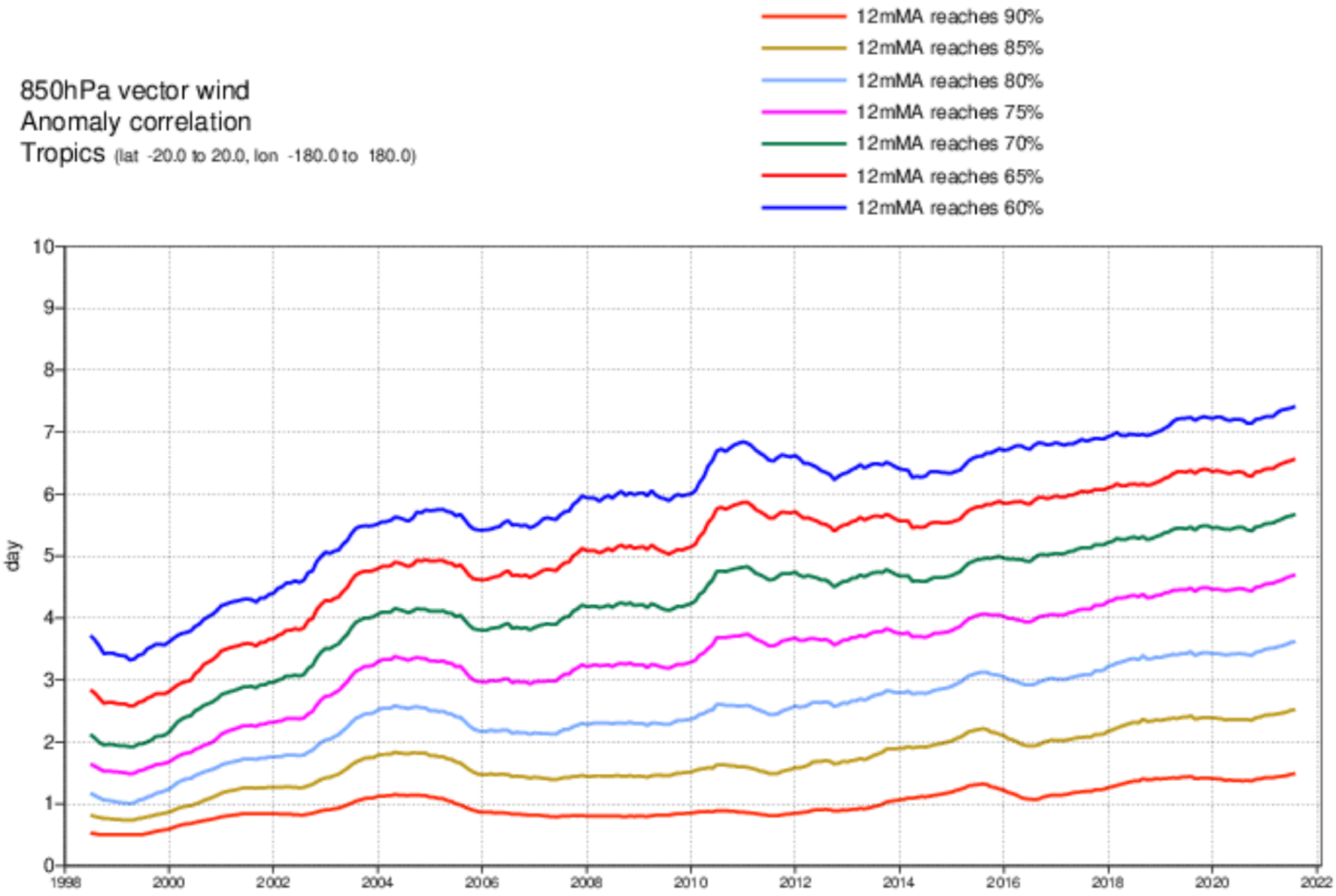

verification against radiosondes
geopotential 500hPa
Root mean square error
NHem Extratropics
Mean method: standard
KMA
UKMO
DWD
MF
JMA
ECMWF
CMC
m
Forecast Day


verification against radiosondes
wind speed 850hPa
Root mean square error
NHem Extratropics
Mean method: standard
KMA
UKMO
DWD
MF
JMA
ECMWF
CMC
m/s
Forecast Day



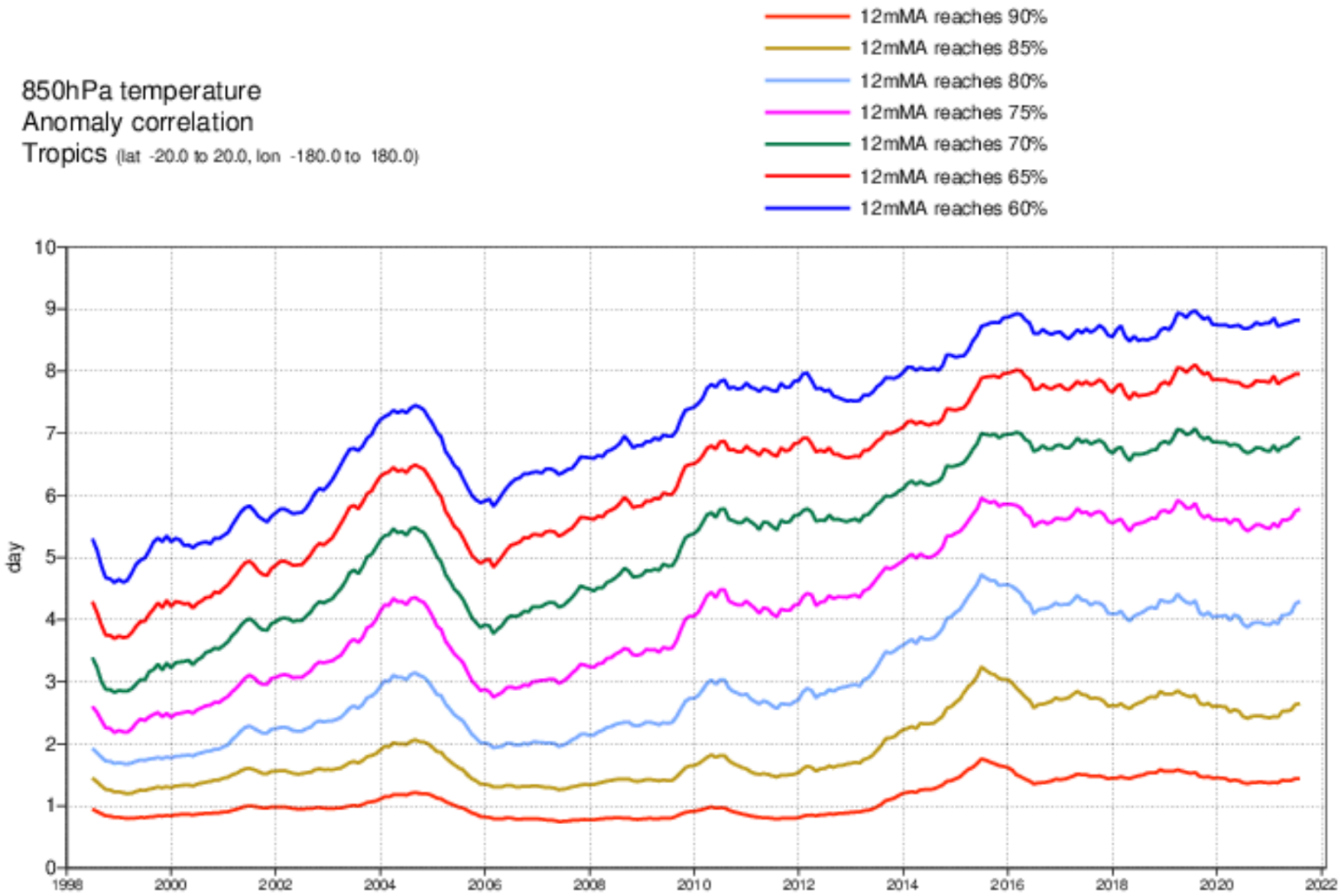
Lead time ACC reaching thresholds

Spatial correlation between anomalies from climatology of forecast and verifying analysis



Lead time ACC reaching thresholds

Spatial correlation between anomalies from climatology of forecast and verifying analysis



Lead time ACC reaching thresholds

Spatial correlation between anomalies from climatology of forecast and verifying analysis

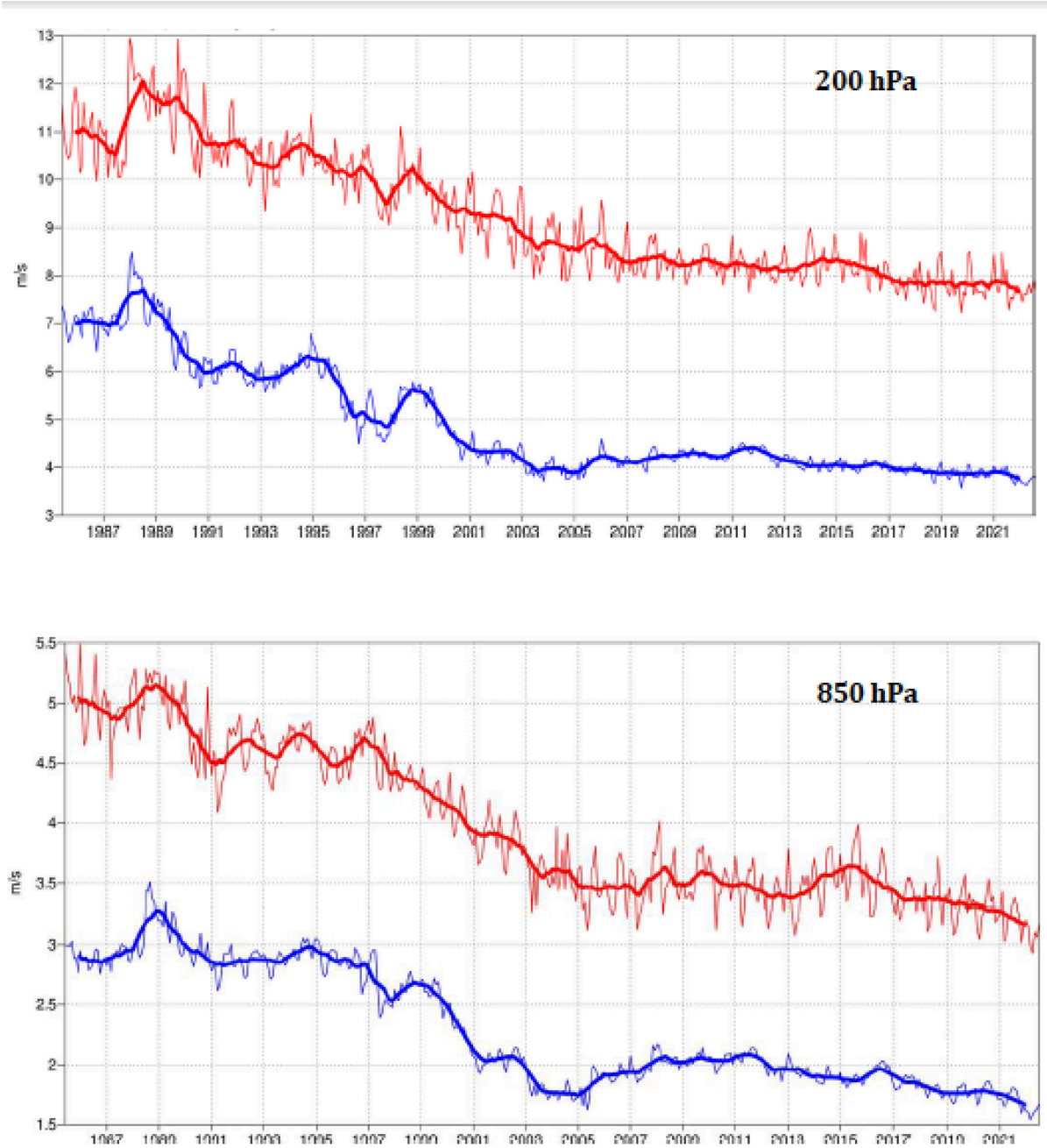


Figure 14: Forecast performance in the tropics. Curves show the monthly average RMS vector wind errors at 200 hPa (top) and 850 hPa (bottom) for one-day (blue) and five-day (red) forecasts, verified against analysis. 12-month moving average scores are also shown (in bold).

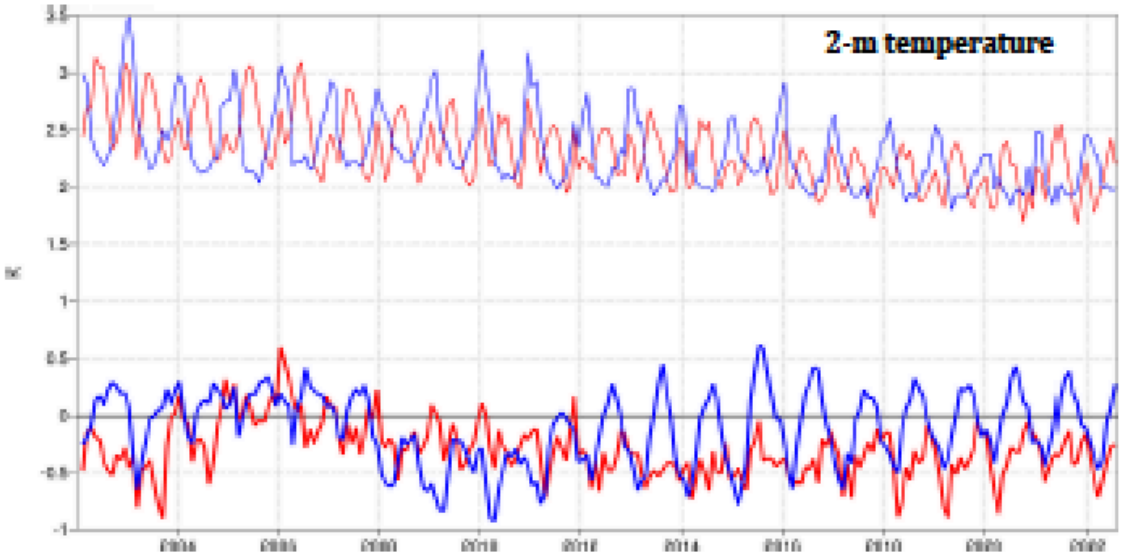


Figure 23: Verification of 2 m temperature forecasts against European SYNOP data on the GTS for 60-hour (night-time, blue) and 72-hour (daytime, red) forecasts. Lower pair of curves shows bias, upper curves are standard deviation of error.

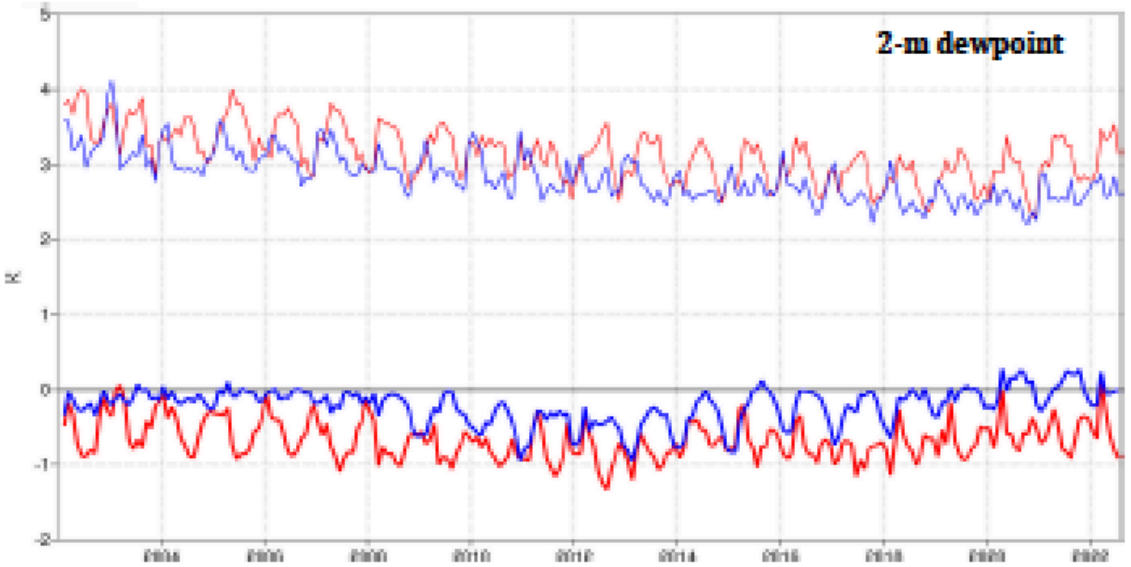


Figure 24: Verification of 2 m dew point forecasts against European SYNOP data on the Global Telecommunication System (GTS) for 60-hour (night-time, blue) and 72-hour (daytime, red) forecasts. Lower pair of curves shows bias, upper curves show standard deviation of error.

Europe

Night time: blue curves
Day time: red curves

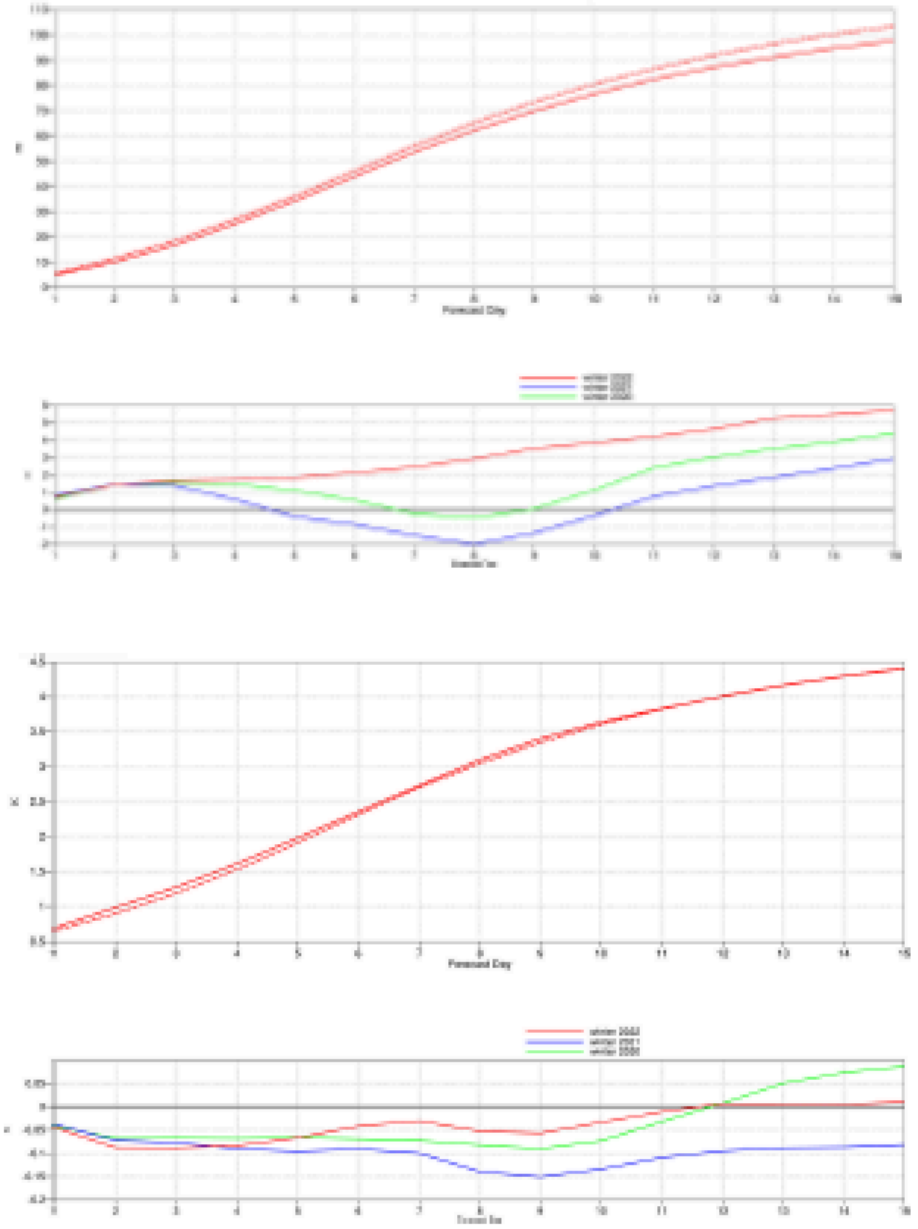


Figure 10: Ensemble spread (standard deviation, dashed lines) and RMS error of ensemble-mean (solid lines) for winter 2021–2022 (upper figure in each panel), and differences of ensemble spread and RMS error of ensemble mean for last three winter seasons (lower figure in each panel, negative values indicate spread is too small); verification is against analysis, plots are for 500 hPa geopotential (top) and 850 hPa temperature (bottom) over the extratropical northern hemisphere for forecast days 1 to 15.

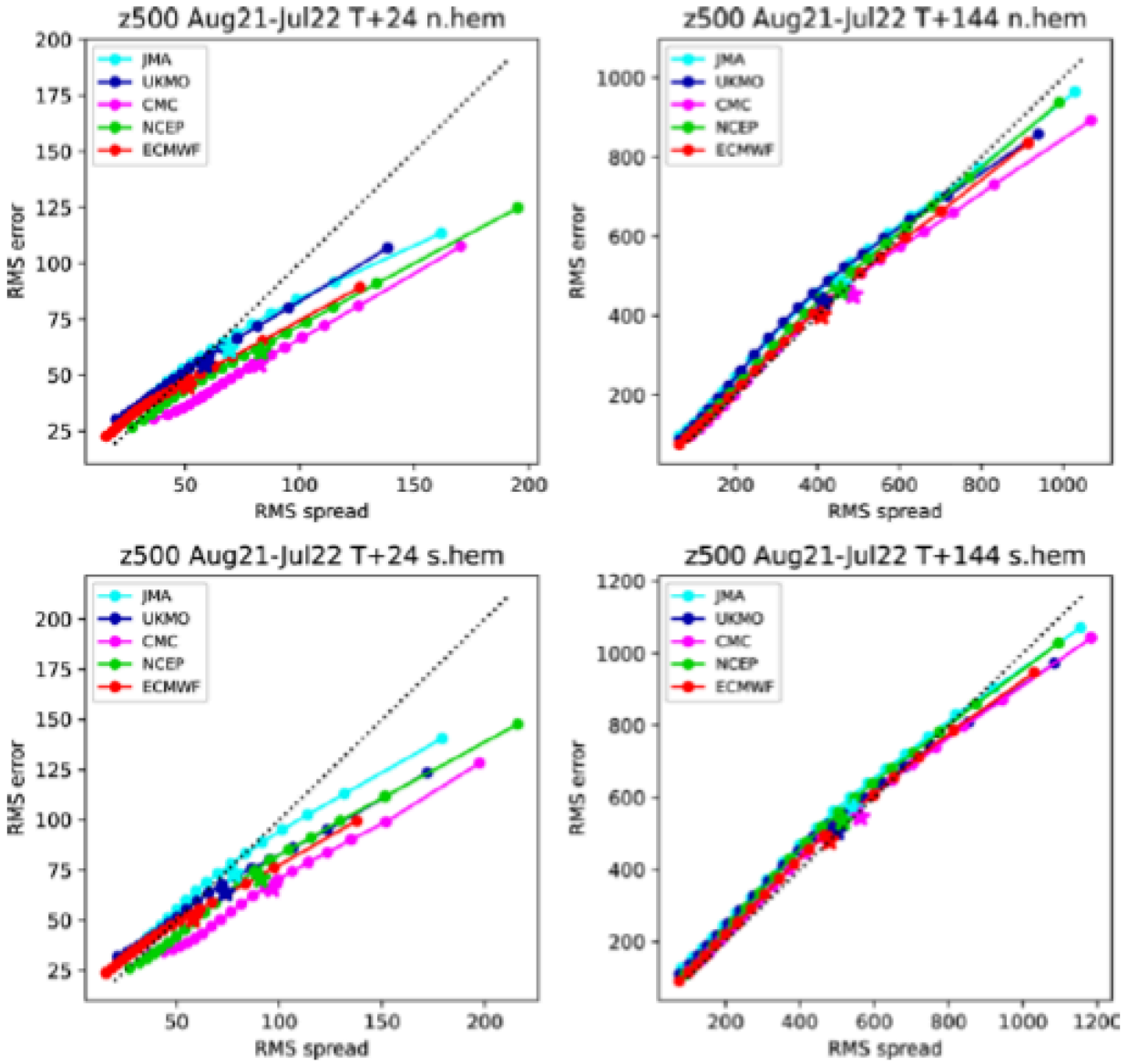


Figure 11: Ensemble spread reliability of different global models for 500 hPa geopotential for the period August 2021–July 2022 in the northern (top) and southern (bottom) hemisphere extra-tropics for day 1 (left) and day 6 (right), verified against analysis. Circles show error for different values of spread, stars show average error-spread relationship.

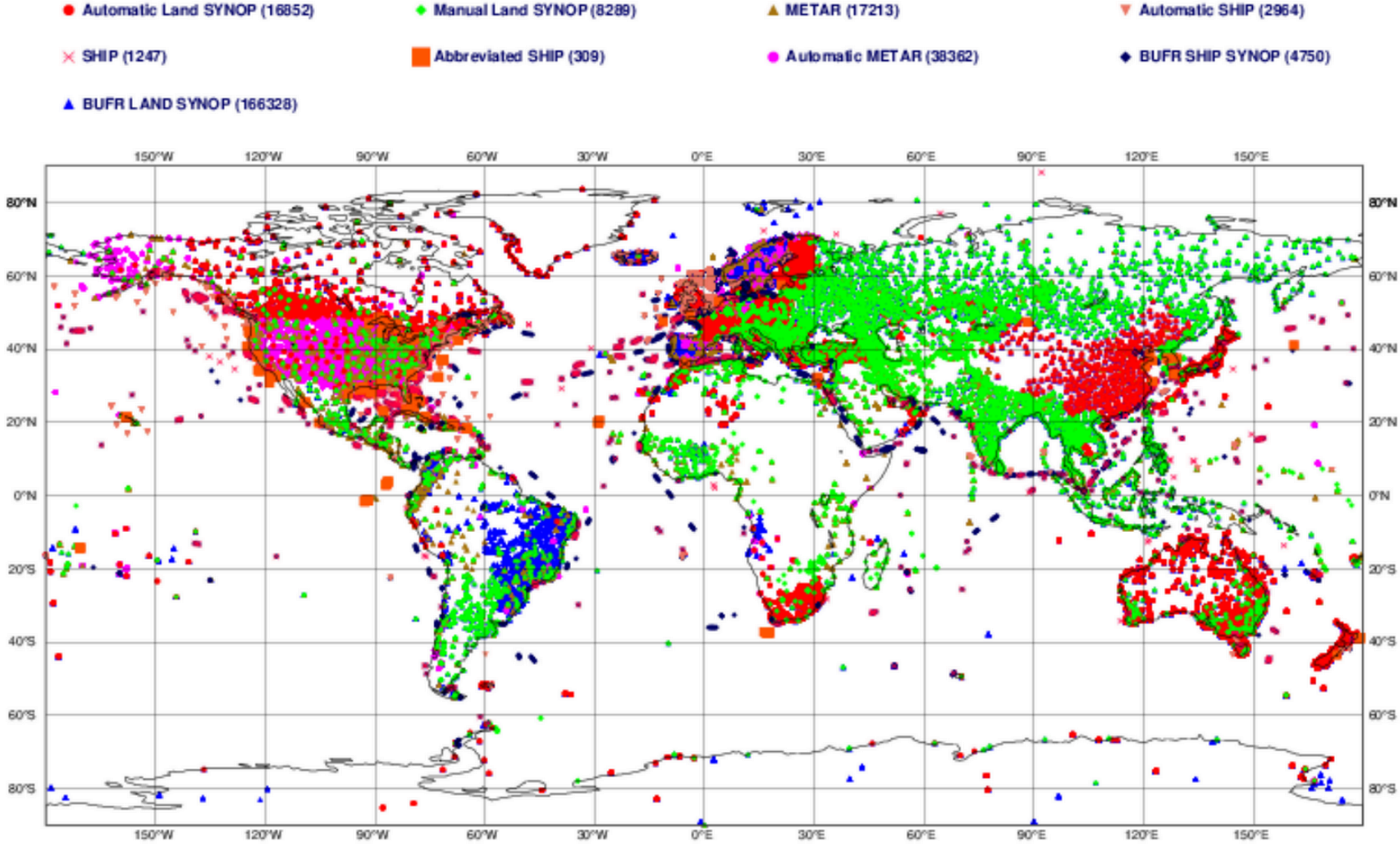
Remaining problems

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

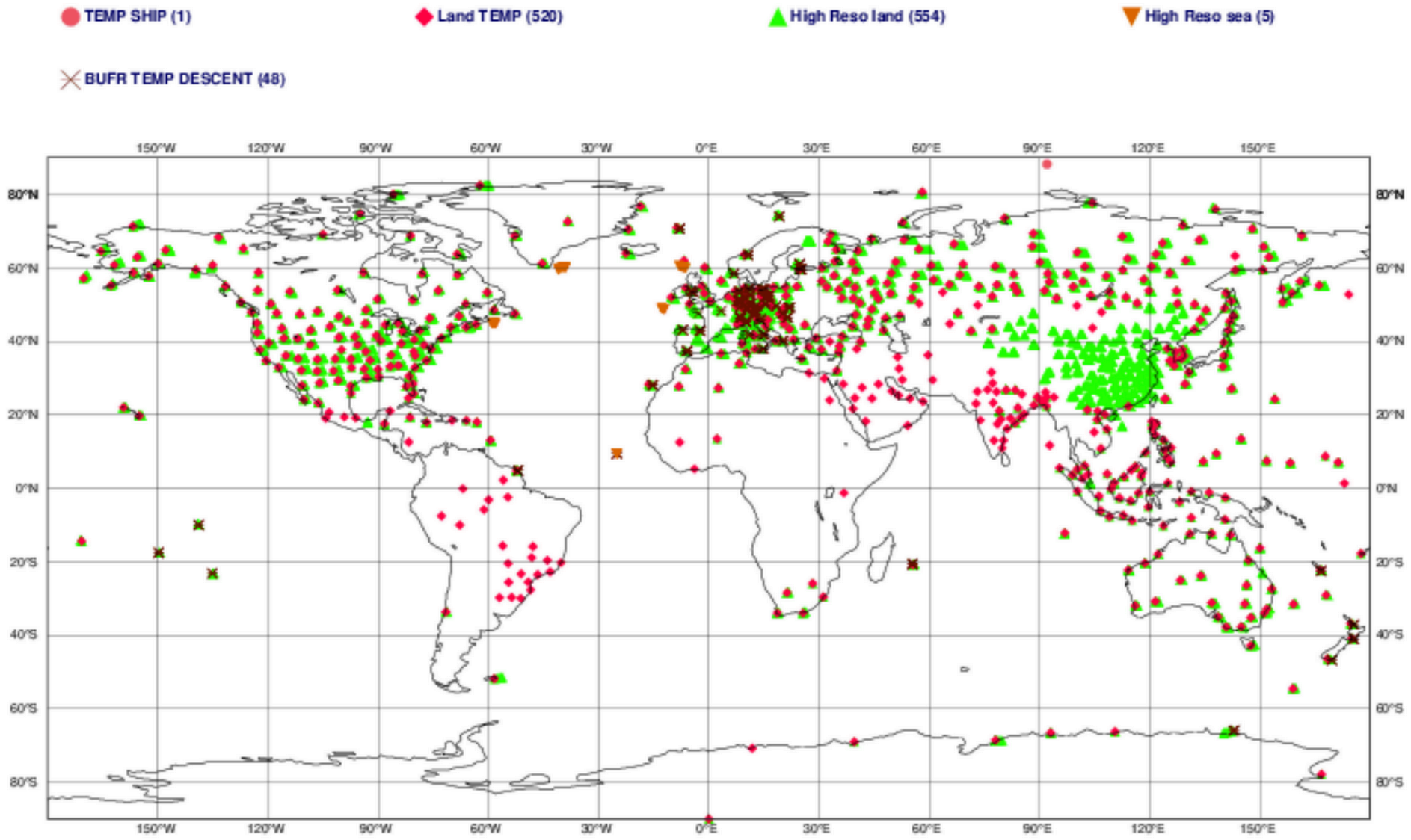
- Exchanges with ocean or continental surface (heat, water, momentum, …)

- …

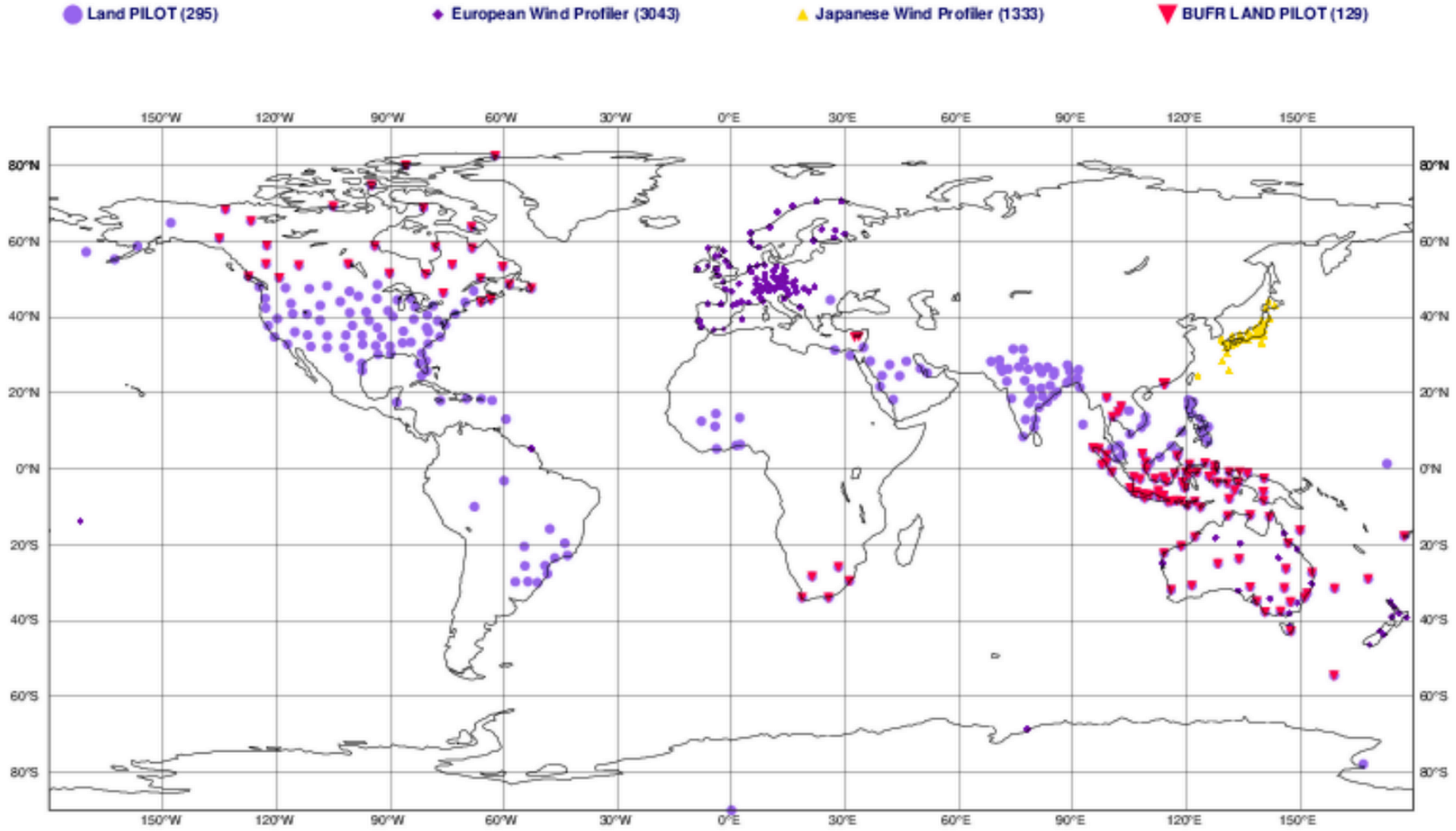
ECMWF data coverage (all observations) - SYNOP-SHIP-METAR
2023032621 to 2023032703
Total number of obs = 256314



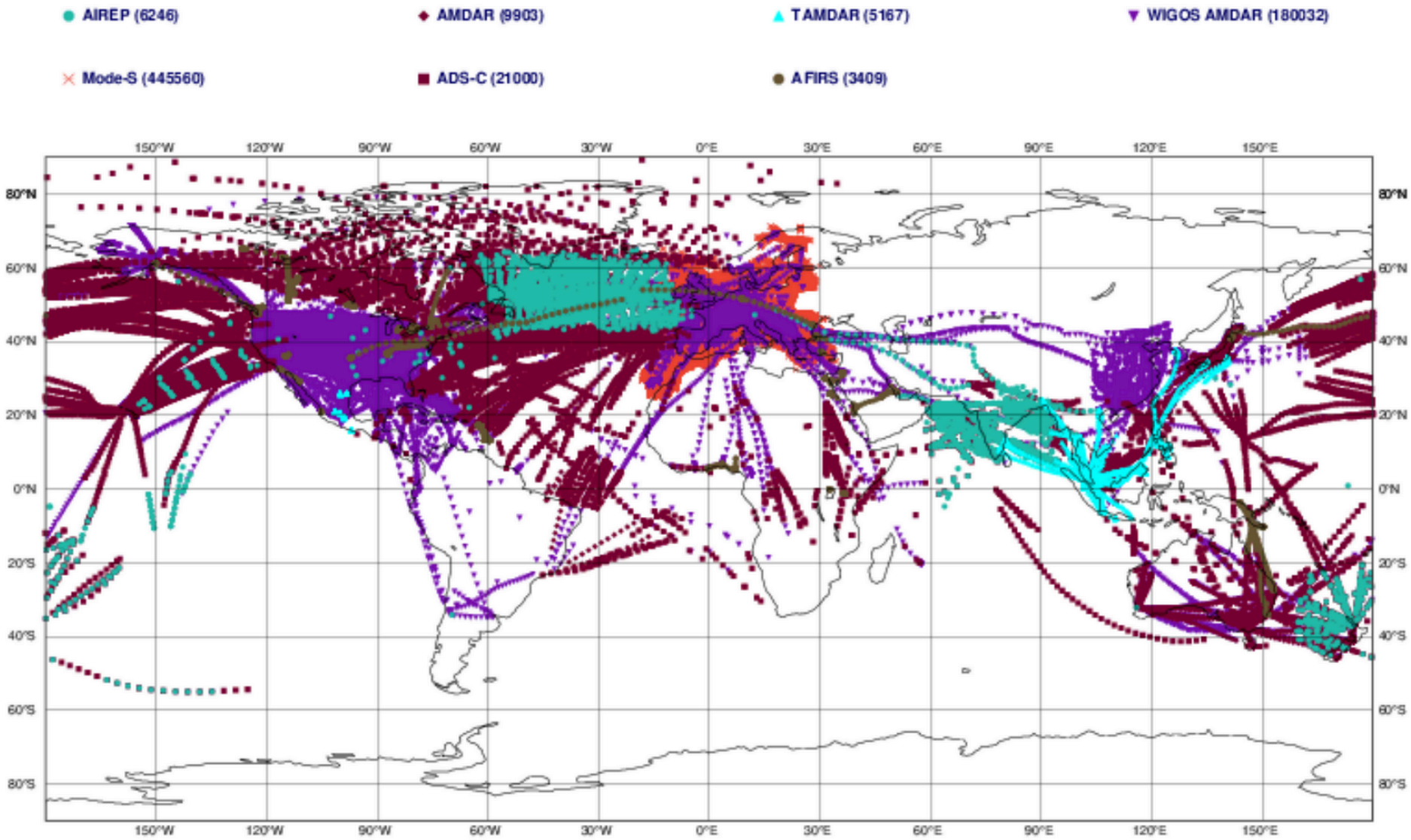
ECMWF data coverage (all observations) - RADIOSONDE
2023032621 to 2023032703
Total number of obs = 1128



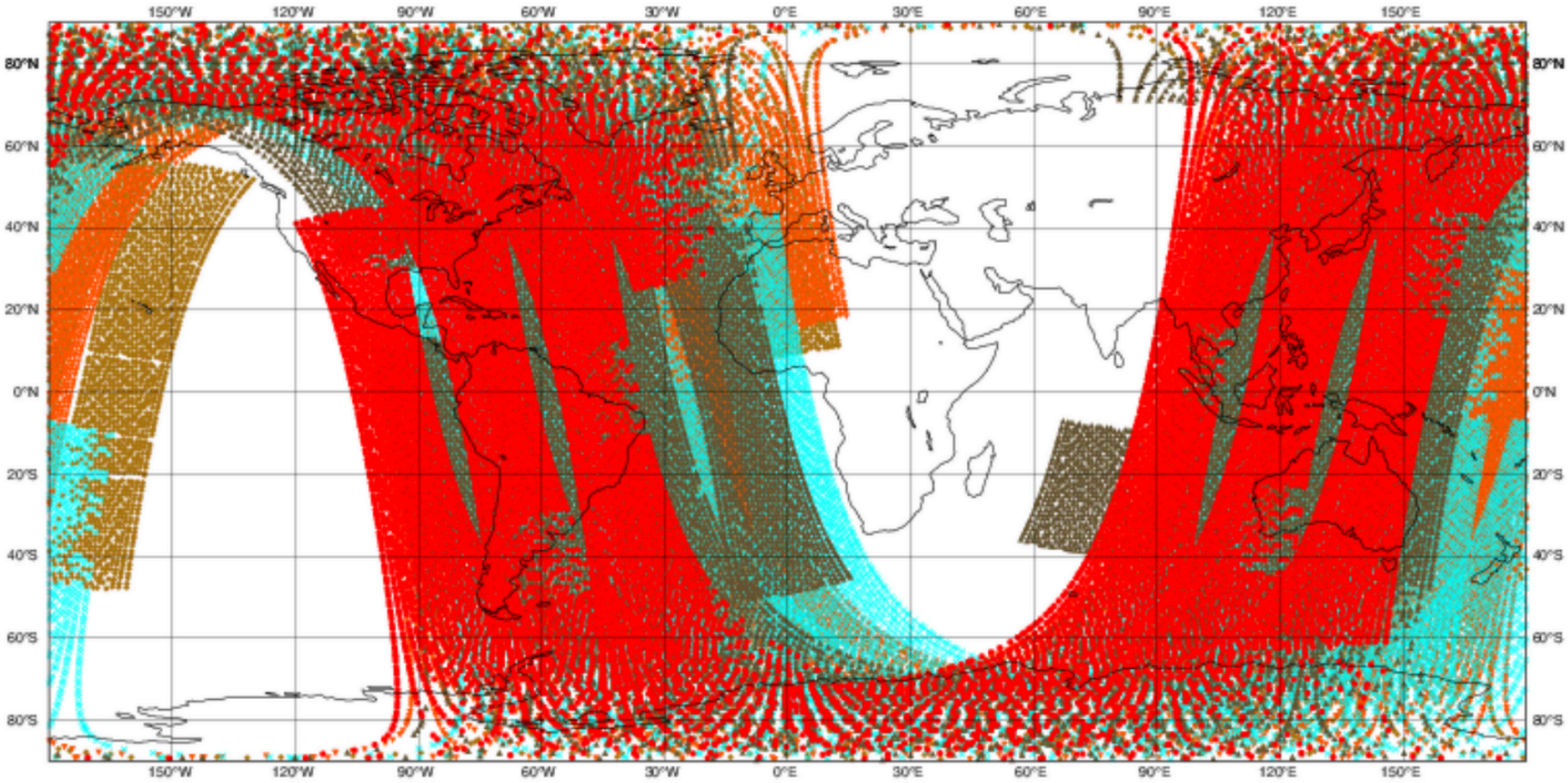
ECMWF data coverage (all observations) - PILOT
2023032621 to 2023032703
Total number of obs = 4800



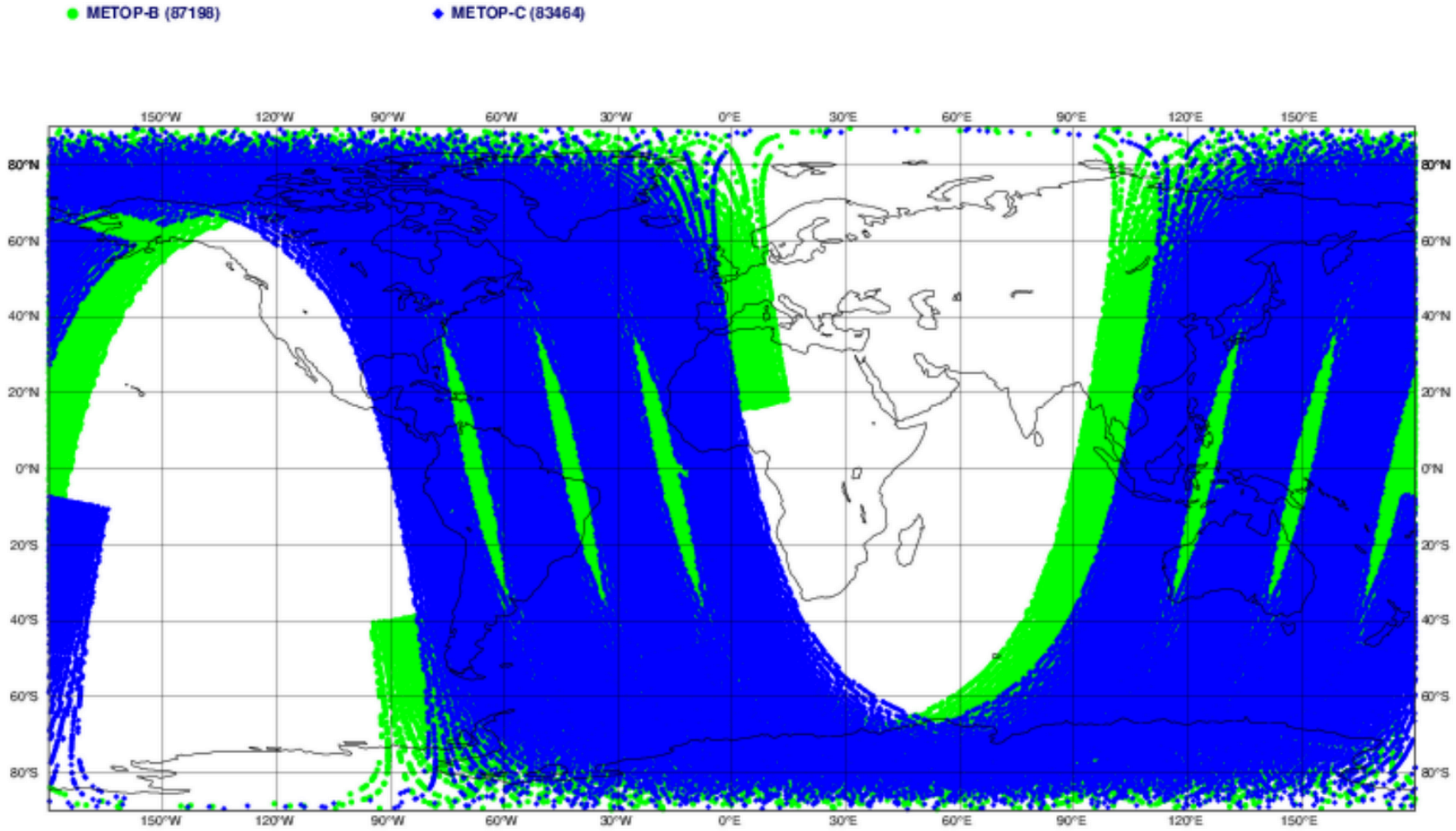
ECMWF data coverage (all observations) - AIRCRAFT
2023032621 to 2023032703
Total number of obs = 671317



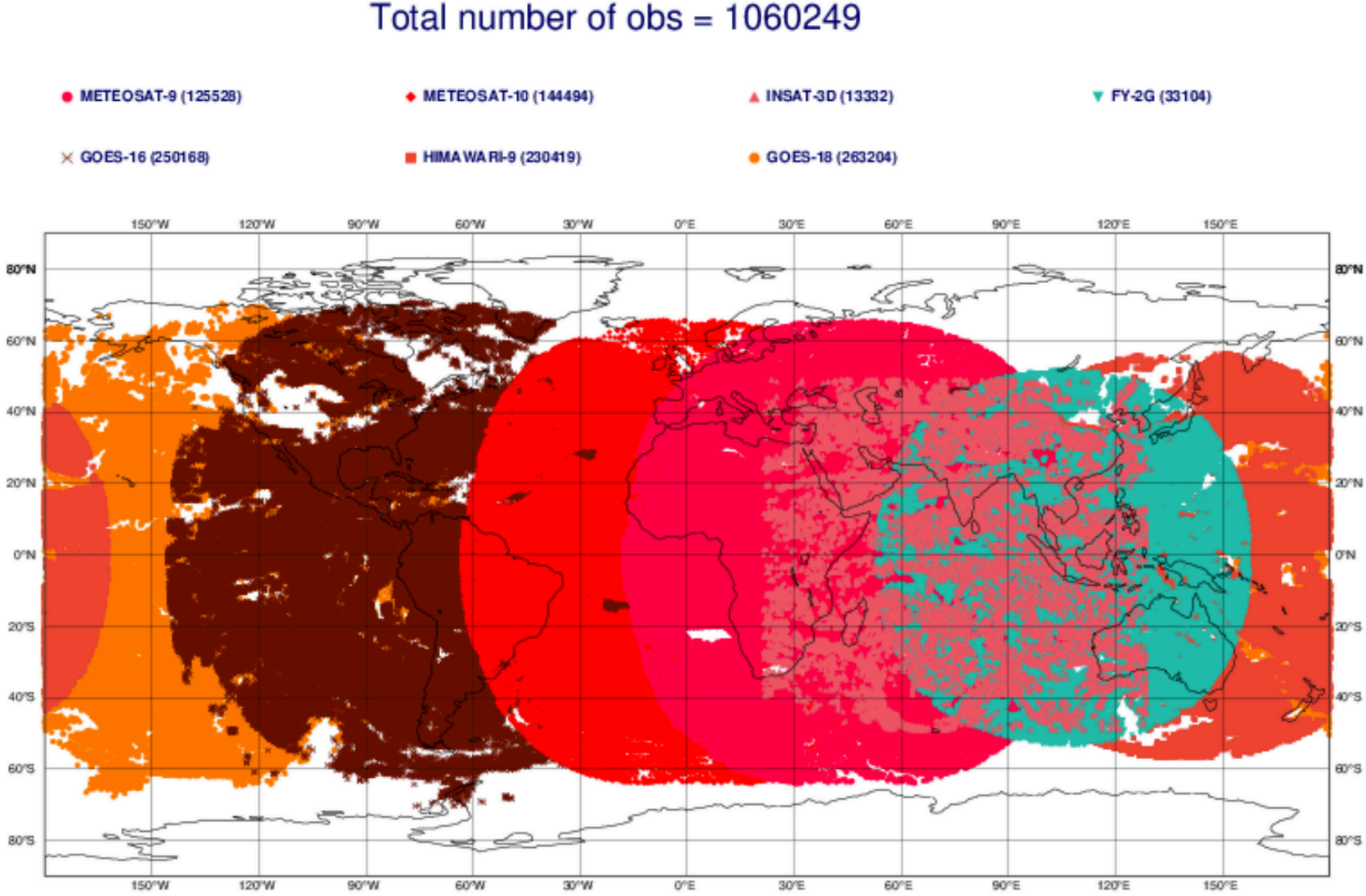
ECMWF data coverage (all observations) - AMSUA
2023032621 to 2023032703
Total number of obs = 102504



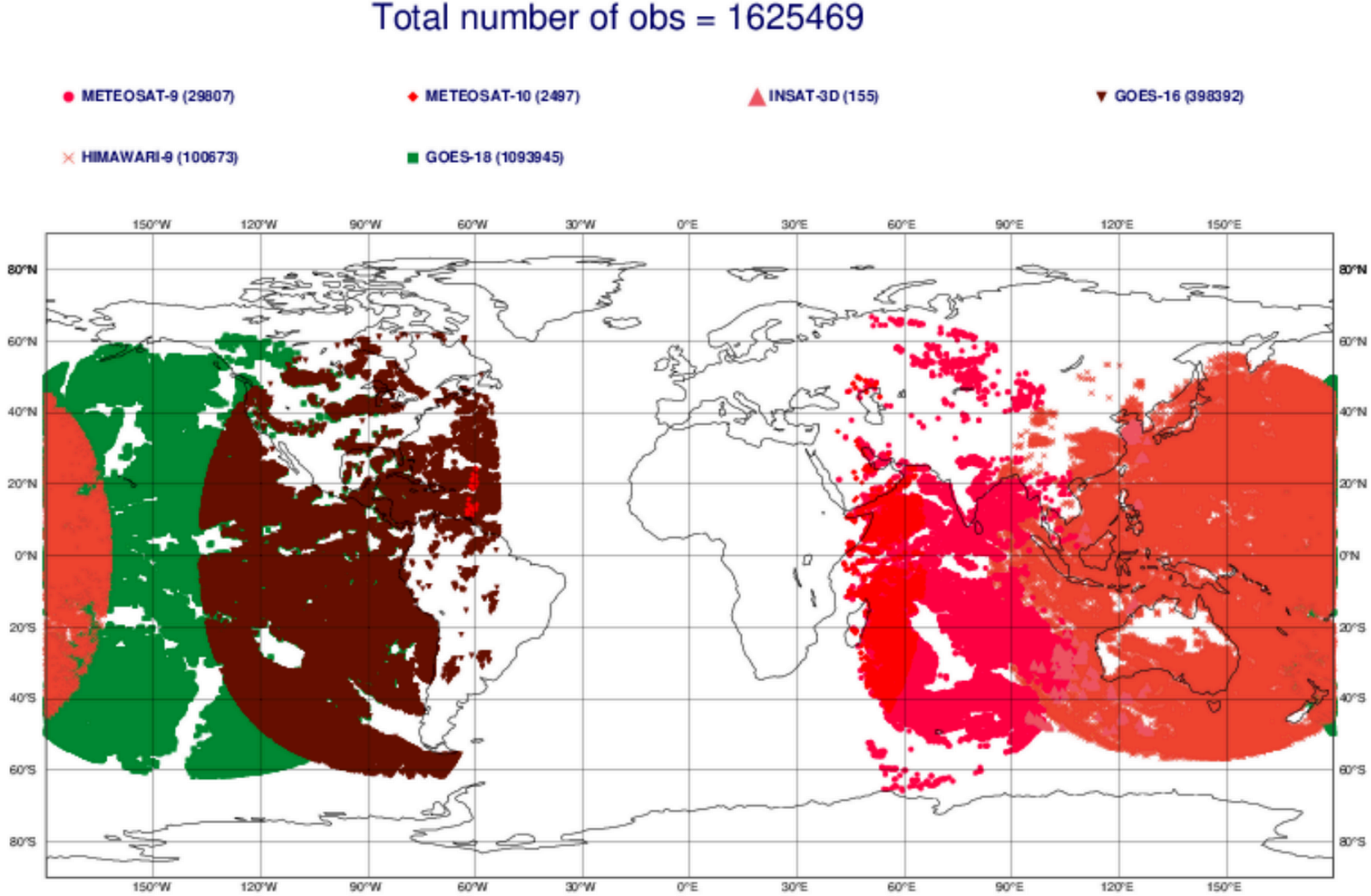
ECMWF data coverage (all observations) - IASI
2023032621 to 2023032703
Total number of obs = 170662



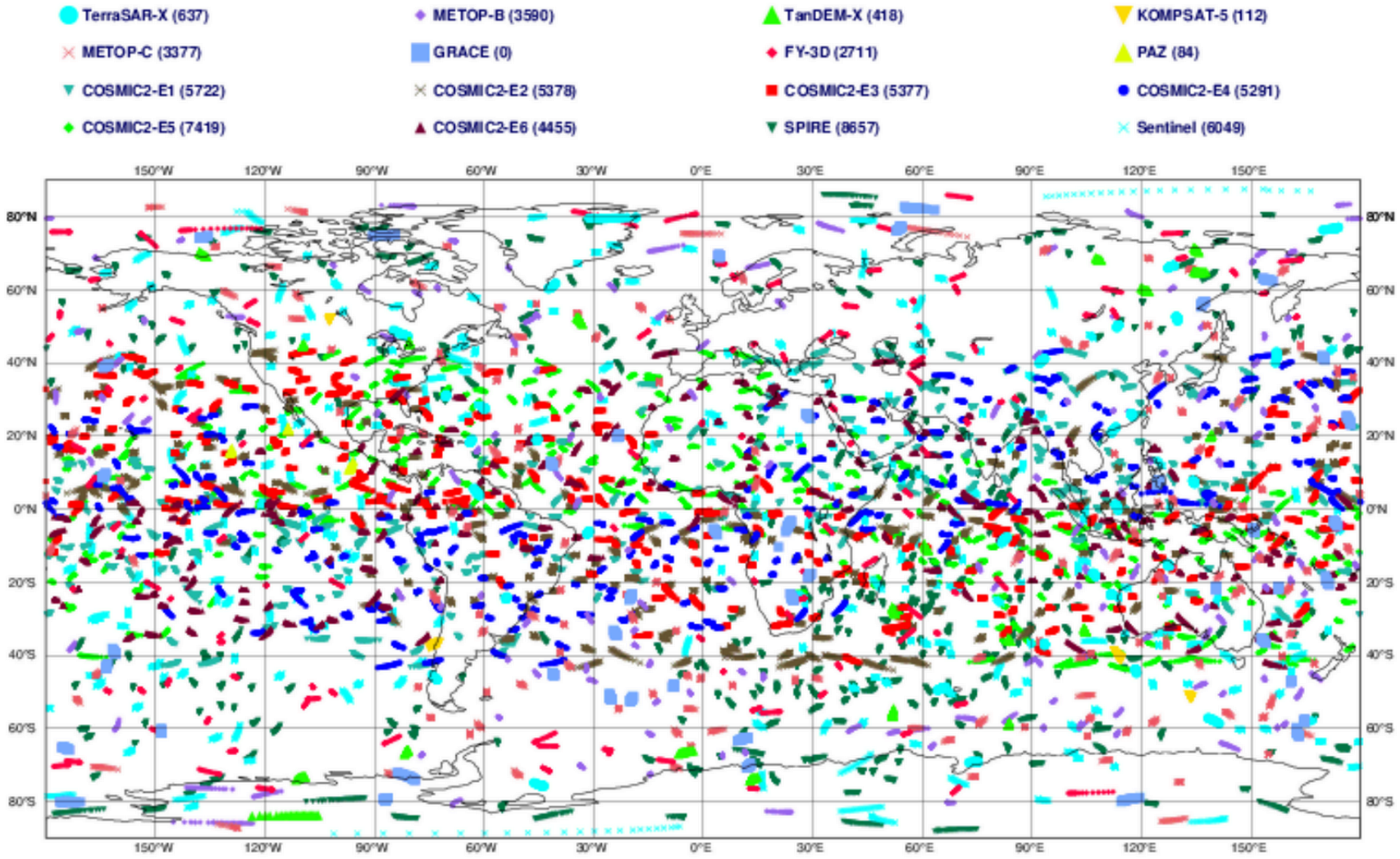
ECMWF data coverage (all observations) - AMV WV
2023032621 to 2023032703
Total number of obs = 1060249



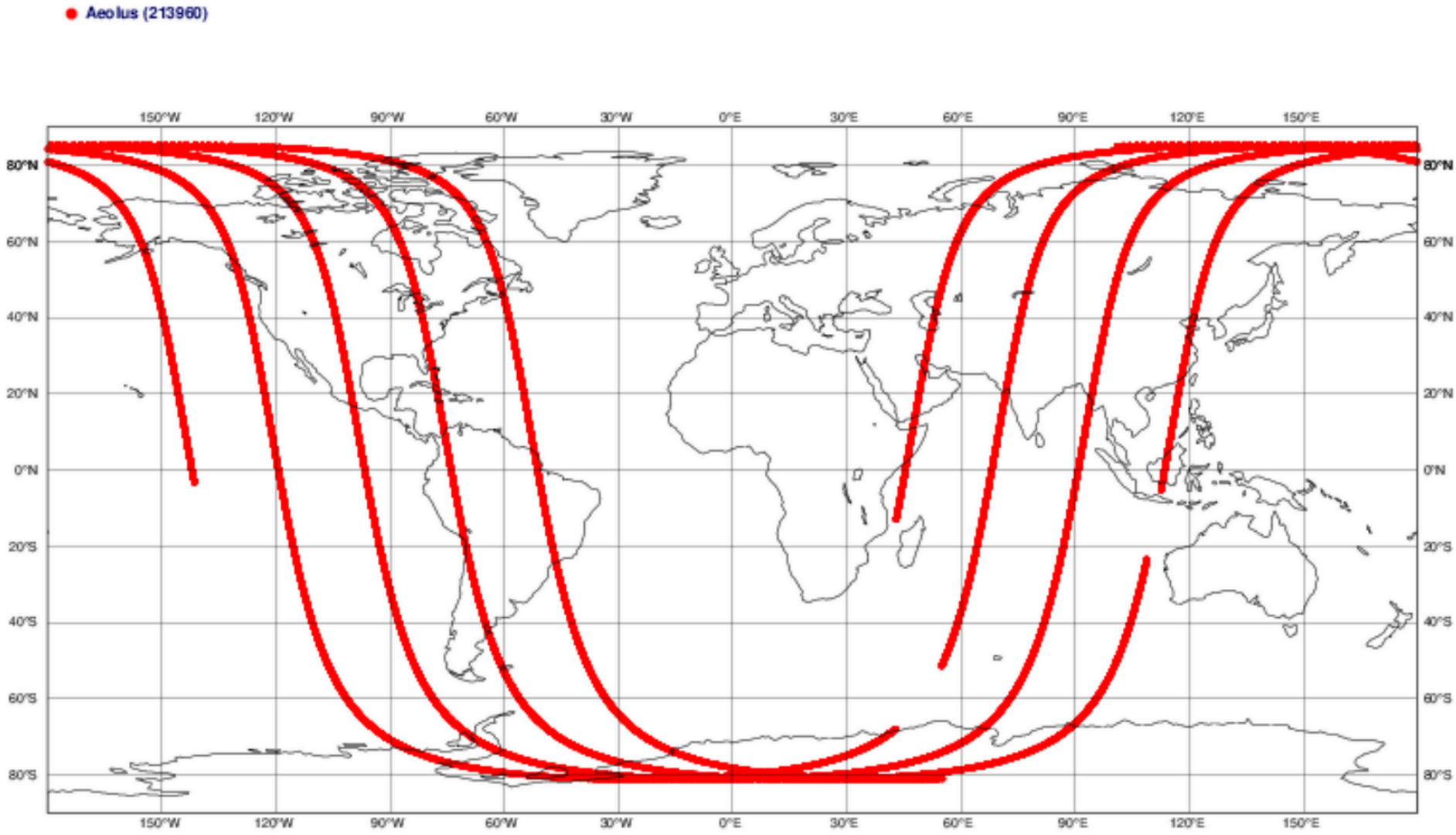
ECMWF data coverage (all observations) - AMV VIS
2023032621 to 2023032703
Total number of obs = 1625469



ECMWF data coverage (all observations) - GPSRO
2023032621 to 2023032703
Total number of obs = 59277



ECMWF data coverage (all observations) - AEOLUS
20230325 00
Total number of obs = 213960



Satellite **ADM-Aeolus** was launched on August 22 2018. It carries a Lidar-Doppler instrument, called Aladin (Atmospheric LAser Doppler Instrument), that makes side measurements of wind in the volume of the atmosphere. These new observations have been shown to have a positive impact on the quality of the previsions, especially in the tropics and in the Southern Hemisphere.

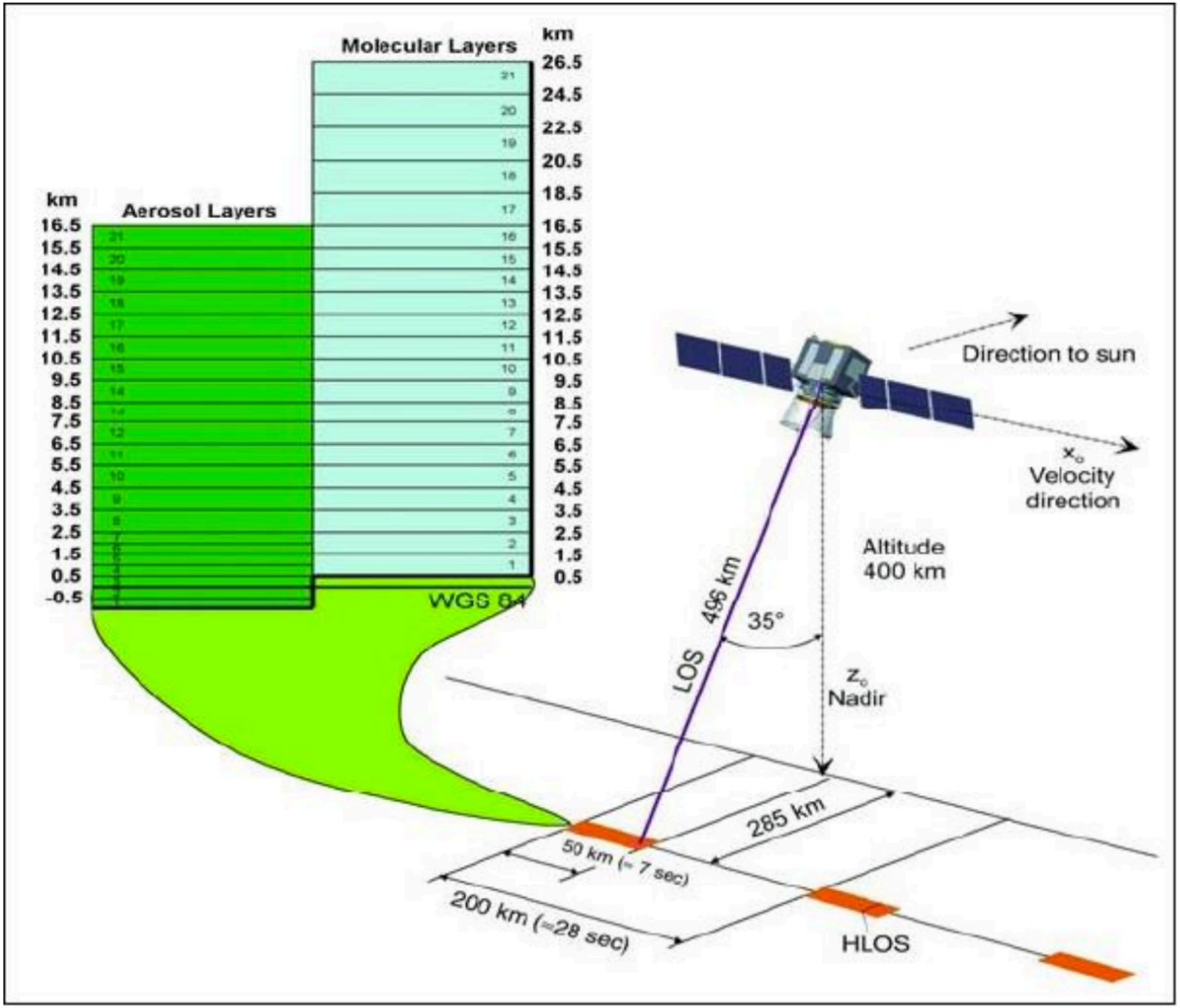
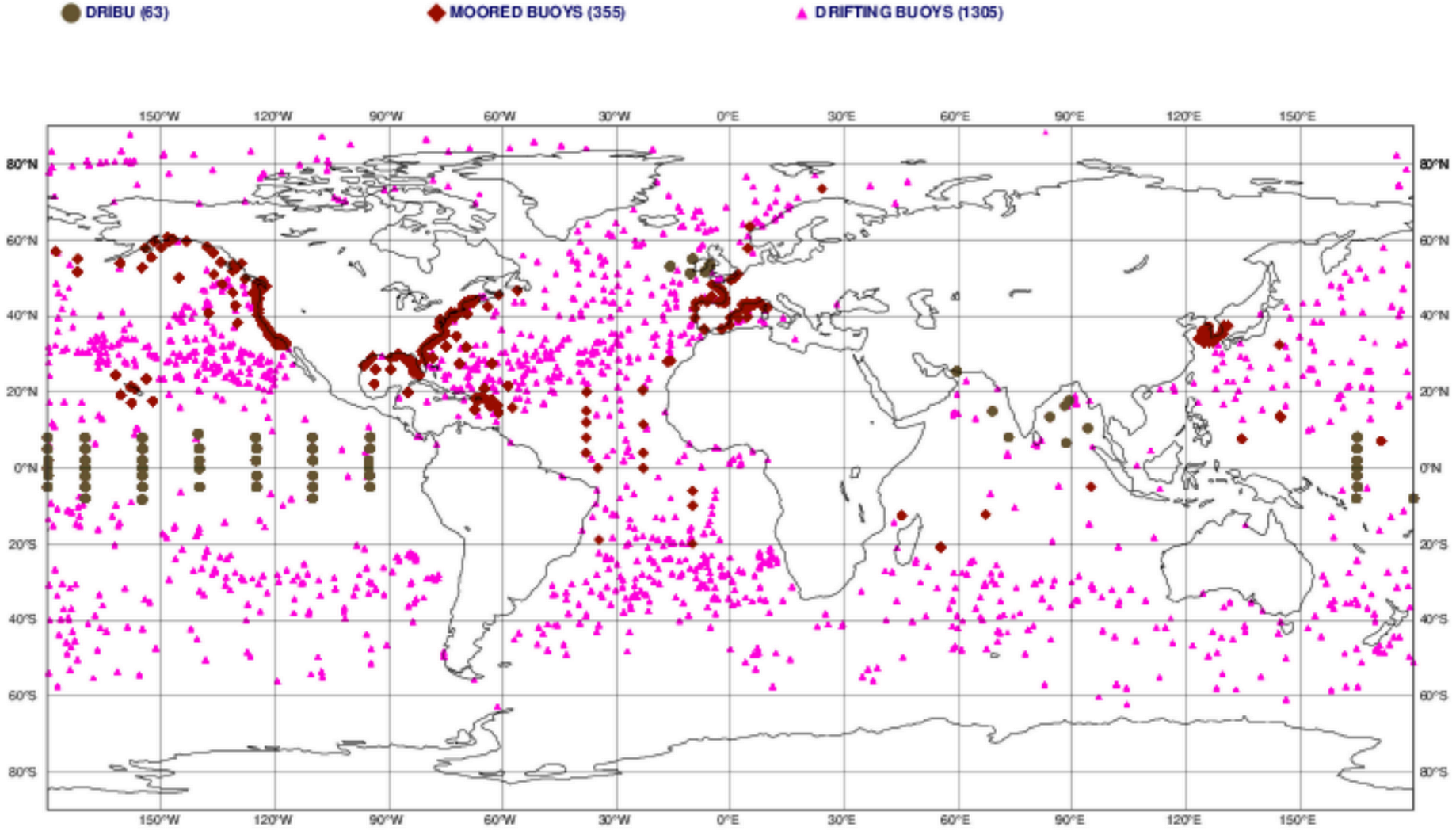
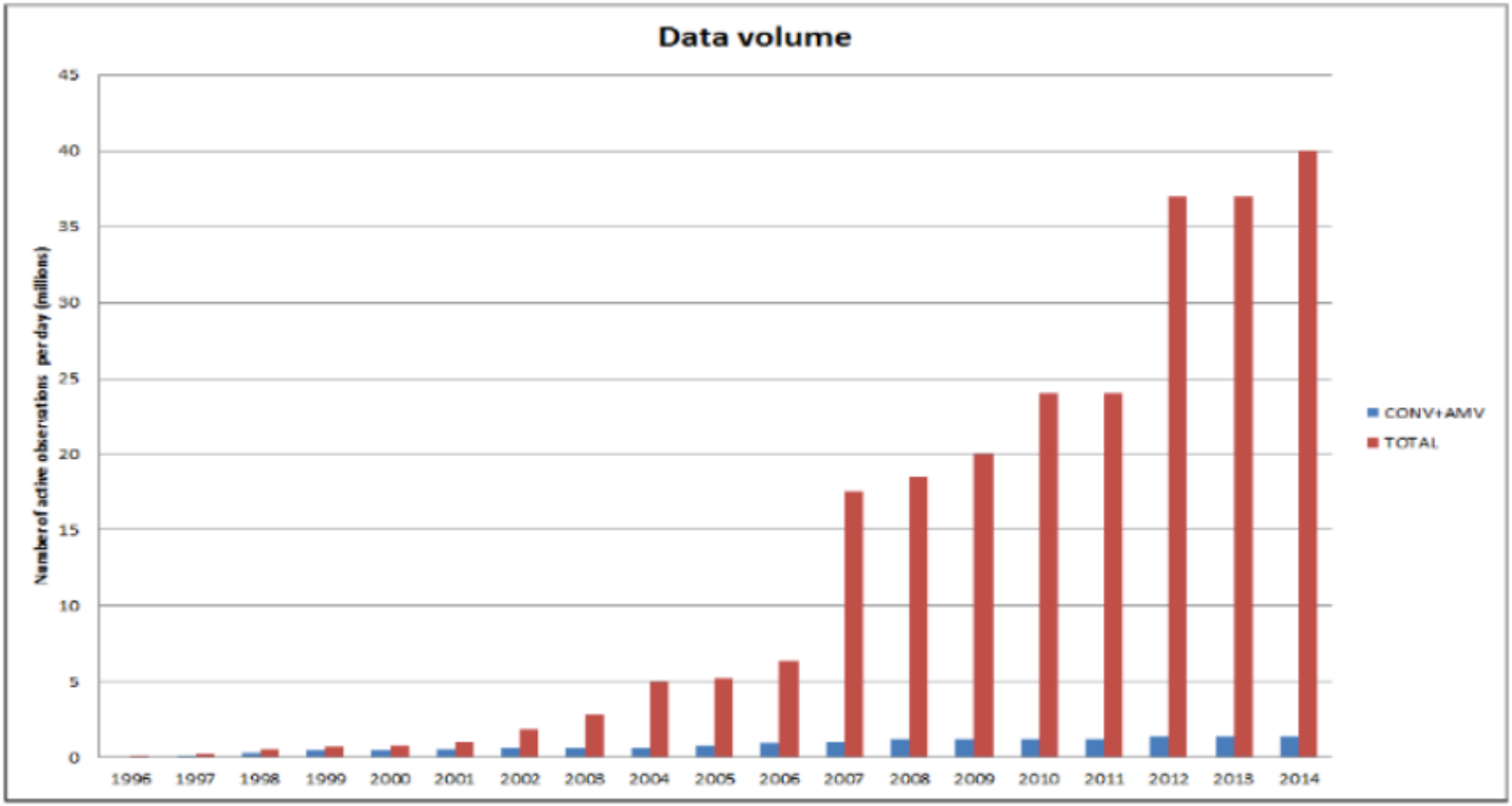


Figure 21: The ADM-Aeolus measurement and sampling concept (image credit: ESA)

ECMWF data coverage (all observations) - BUOY
2023032621 to 2023032703
Total number of obs = 1723



ECMWF


Data volume
Number of active observations per day (millions)
45
40
35
30
25
20
15
10
5
0
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
CONV+AMV
TOTAL

- *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.

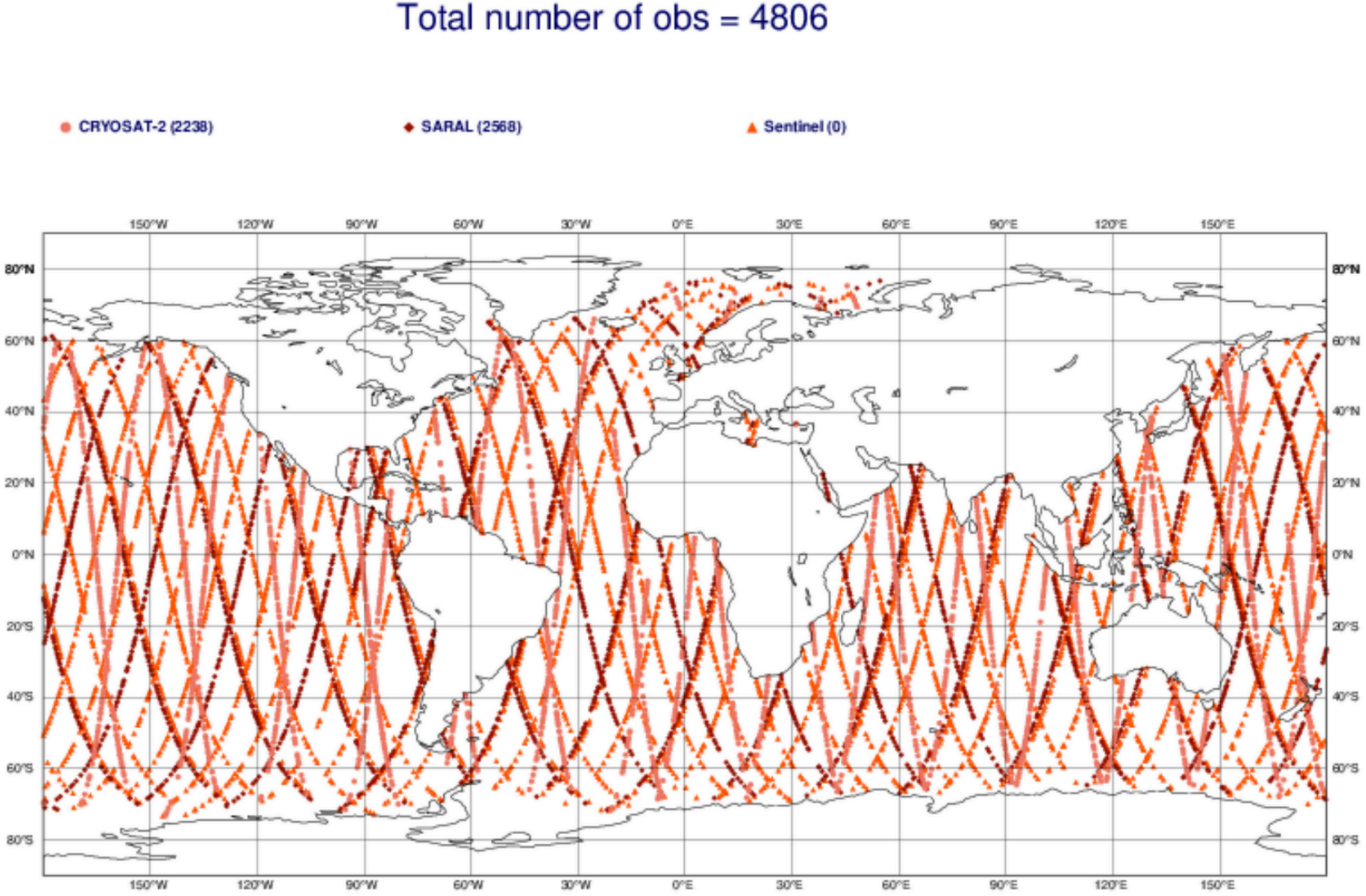
- *Direct* observations (temperature, pressure, horizontal components of the wind, moisture), which are local and bear on the variables used for describing the flow in numerical models.

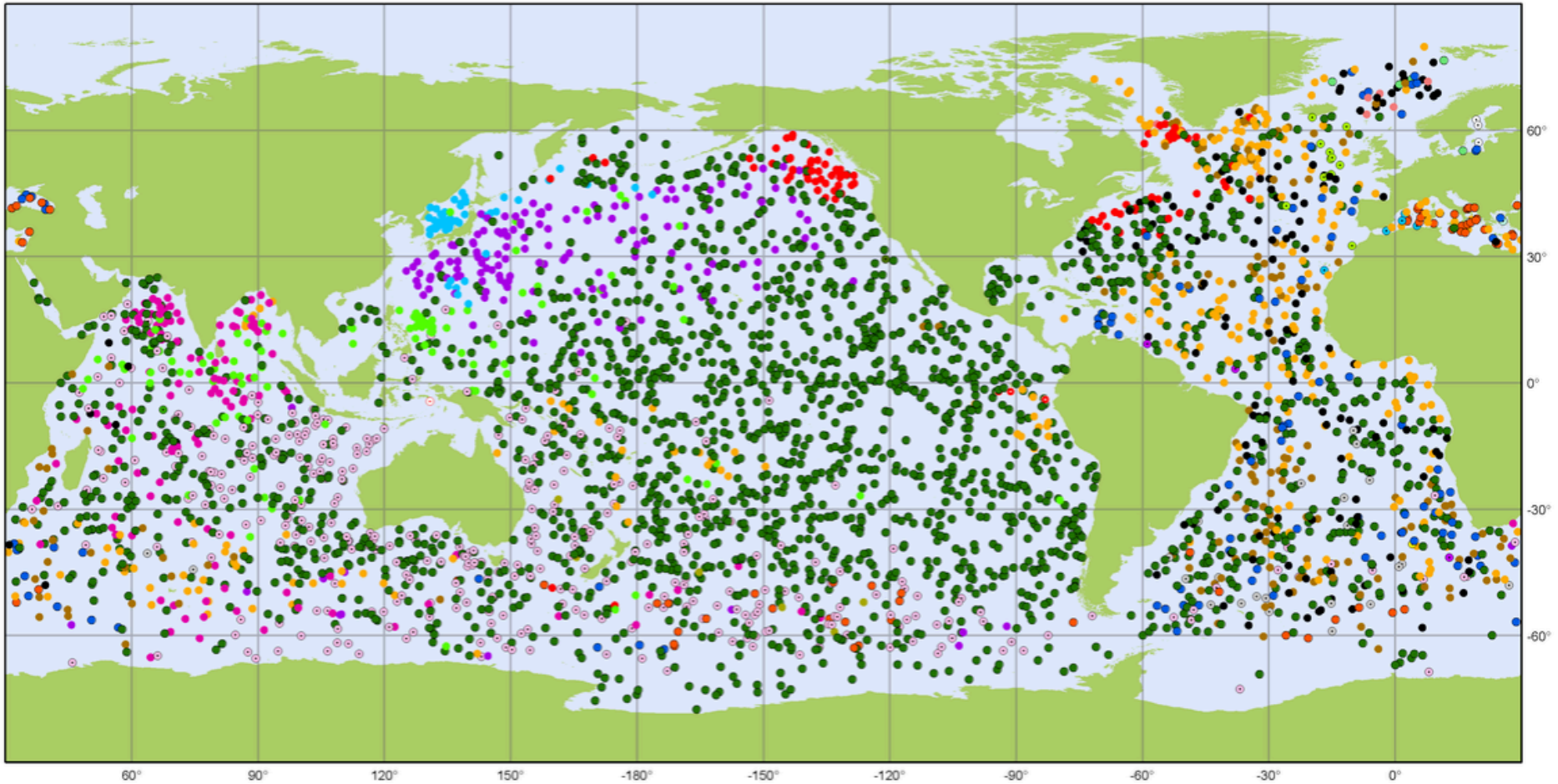
- *Indirect* observations (radiometric observations, …), which bear on some more or less complex combination (most often, a one-dimensional spatial integral) of variables used for for describing the flow

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

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

ECMWF data coverage (all observations) - SEA LEVEL ANOMALY
20230325 00
Total number of obs = 4806





Argo **National contributions - 3881 Operational Floats** **February 2018**

Latest location of operational floats (data distributed within the last 30 days)

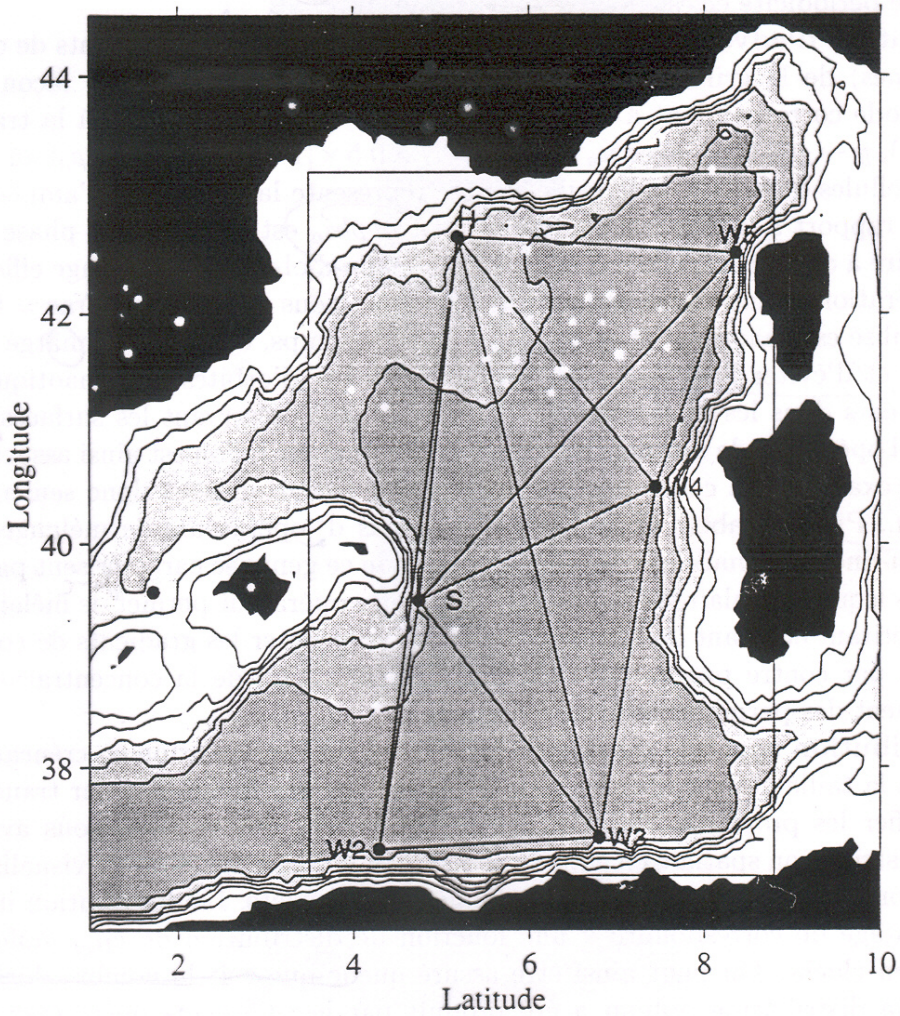


FIG. 1 – *Bassin méditerranéen occidental: réseau d'observation tomographique de l'expérience Thétis 2 et limites du domaine spatial utilisé pour les expériences numériques d'assimilation.*

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.

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; 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).

Assimilation is one of many '*inverse problems*' encountered in many fields of science and technology

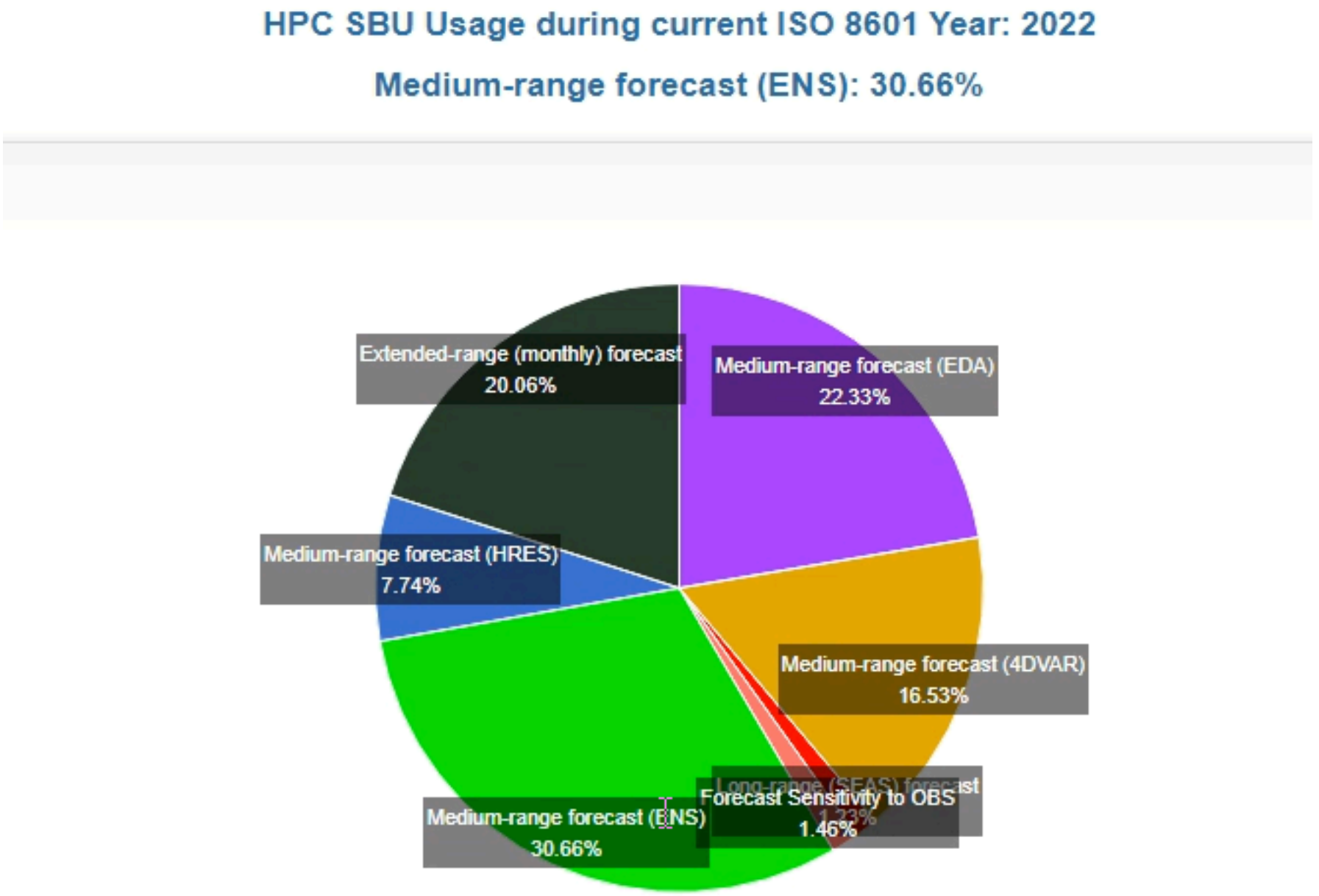
- solid Earth geophysics
- plasma physics
- 'nondestructive' probing
- navigation (spacecraft, aircraft, ….)
- …

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, $p \approx 4$-5.10^7 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



Courtesy J.-N. Thépaut

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

- ζ_1 and ζ_2 mutually independent

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

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

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

ζ_1 and ζ_2 mutually independent

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

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

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

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

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

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

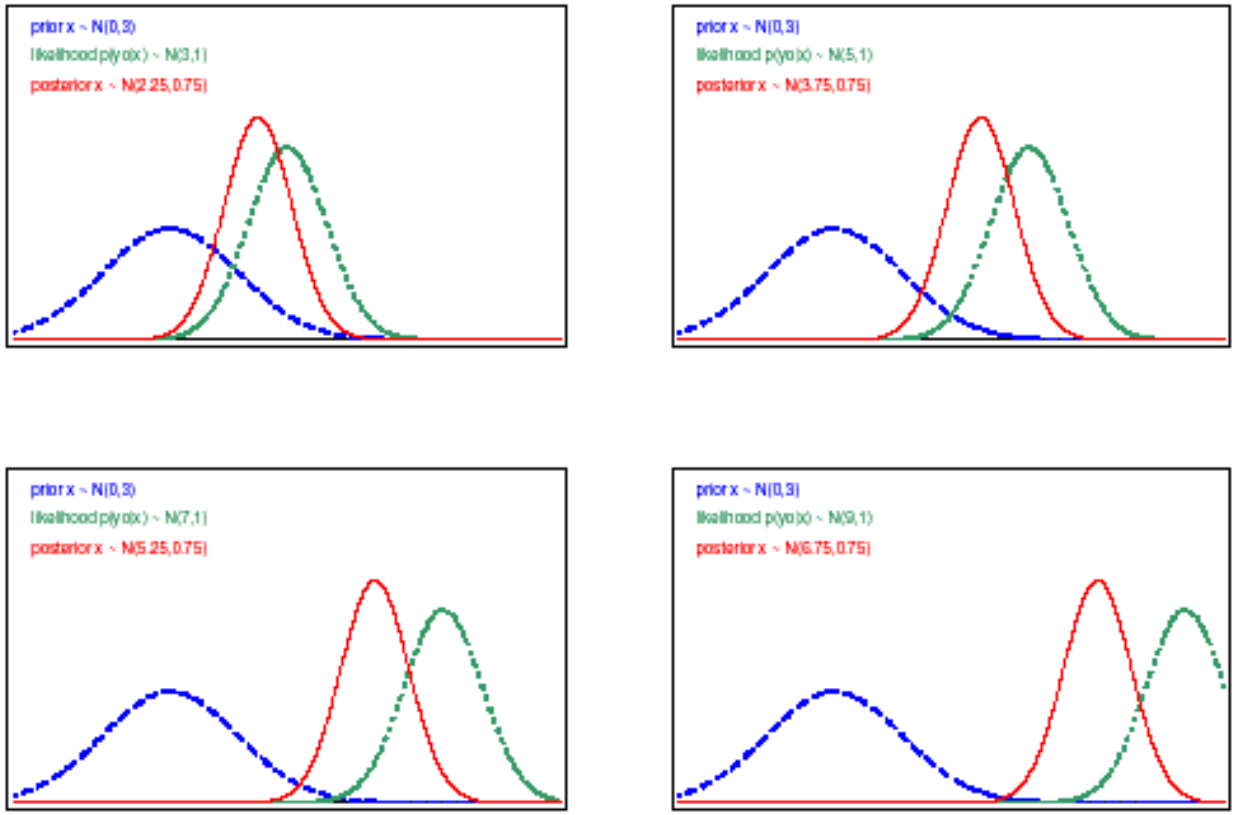


Fig. 1.1: Prior pdf $p(x)$ (dashed line), posterior pdf $p(x|y^o)$ (solid line), and Gaussian likelihood of observation $p(y^o|x)$ (dotted line), plotted against x for various values of y^o. (Adapted from Lorenc and Hammon 1988.)

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

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

In addition

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

Conditional probability distribution in Gaussian case

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

$$\mathcal{J}(\xi) + Cst$$

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 also be obtained, independently of any Gaussian hypothesis, as simply corresponding to the linear combination of z_1 and z_2 that minimizes the error $E\,[(x^a - x)^2]$

Best Linear Unbiased Estimator (BLUE)

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

Same as before, but ζ_1 and ζ_2 are now distributed according to exponential law with parameter a, *i. e.*

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

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

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

$\mathrm{Var}(x \mid z_1, z_2) = a^2 (2\delta^3/3 + \delta^2 + \delta + 1/2) / (1 + 2\delta)$, with $\delta = |z_1 - z_2| / (2a)$
Increases from $a^2/2$ to ∞ as δ increases from 0 to ∞. Can be larger than variance $2a^2$ of original errors (probability 0.08)

Bayesian estimation

State vector $\boldsymbol{x}$, belonging to *state space* $\mathcal{S}$ ($\dim \mathcal{S} = n$), to be estimated.

Data vector $\boldsymbol{z}$, belonging to *data space* $\mathcal{D}$ ($\dim \mathcal{D} = m$), available.

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

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

$$\boldsymbol{z} = \boldsymbol{\Gamma x} + \zeta$$

Bayesian estimation (continued)

Probability that $\boldsymbol{x} = \xi$ for given ξ ?

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

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

Unambiguously defined iff, for any ζ, there is at most one $\boldsymbol{x}$ such that (1) is verified.

$\Leftrightarrow$ data contain information, either directly or indirectly, on any component of $\boldsymbol{x}$. *Determinacy* condition. Implies $m \geq n$.

Bayesian estimation is actually 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 (the *curse of dimensionality)*.

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

- Reminder on elementary probability theory. Random vectors and covariance matrices, random functions and covariance functions

- *Optimal Interpolation*. Principle, simple examples, basic properties.

- *Best Linear Unbiased Estimate* (*BLUE*)

Cours à venir

~~Mardi 21 mars~~
~~Mardi 28 mars~~
Mardi 4 avril
Mardi 11 avril
Mardi 2 mai
Mardi 9 mai
Mardi 23 mai
Mardi 30 mai