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

Année Universitaire 2019-2020

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 2

26 Mars 2020

- Additional material on numerical modelling of the atmospheric circulation. Temporal discretization.

- Numerical Weather Prediction. Present performance (mostly ECMWF)

- The meteorological observation system

- Assimilation. Basics of statistical estimation.

- Discretization grid. Arakawa C-grid

- Convective adjusment

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

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

- 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

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'*

Integrate equation

$$dx / dt = F(x)$$

(x state vector of the model).

Timestep Δt.

Computed solution at time $n\Delta t$ x_n

Forward (Euler) scheme

$$(x_{n+1} - x_n)/\Delta t = F(x_n)$$

$$x_{n+1} = x_n + \Delta t\, F(x_n)$$

Implemented on equation

$$dx / dt = i\alpha x \quad , \qquad \alpha \text{ real} \qquad (1)$$

Exact solution $x(t) = x(0) \exp(i\alpha t)$

Modulus $|x(t)|$ conserved in time

Discretized solution $x_{n+1} = (1 + i\alpha\Delta t)\, x_n$

Modulus $|x_{n+1}| = \sqrt{(1 + \alpha^2\Delta t^2)}\, |x_n|$

increases exponentially with time.

Forward scheme is *unconditionally unstable* for Eq. (1)

Leapfrog scheme

$$(x_{n+1} - x_{n-1})/2\Delta t = F(x_n)$$
$$x_{n+1} = x_{n-1} + 2\Delta t\, F(x_n)$$

Stable for equation (1) above (*i.e.* modulus remains constant in time) provided

$$\alpha\Delta t < 1$$

Courant-Friedrichs-Lewy (CFL) condition

In a multidimensional system, the largest α will be the highest frequency that is present in the system. In a discretized system of travelling waves, the highest frequency will correspond to the fastest wave that the discretization can explicitly resolve. It will proportional to $c/\Delta x$, where c is the phase velocity of the fastest waves in the system, and Δx the mesh-size of the discretization

$$\alpha = (1/\beta)\, c/\Delta x$$

where β is an $O(1)$ numerical coefficient depending on the particular discretization scheme under consideration.

CFL condition then becomes

$$\Delta t / \Delta x < \beta / c$$

Significance : numerical propagation of signal must be at least as fast as physical propagation.

CFL condition generally applies to explicit schemes of temporal discretization

In hydrostatic atmosphere, fastest propagating wave : gravity wave with largest scale height, $c = \sqrt{(rT)} \approx 300$ $m.s^{-1}$.

$$\Delta x = 30 \text{ km} \quad \Rightarrow \Delta t = 100 \text{ s}$$

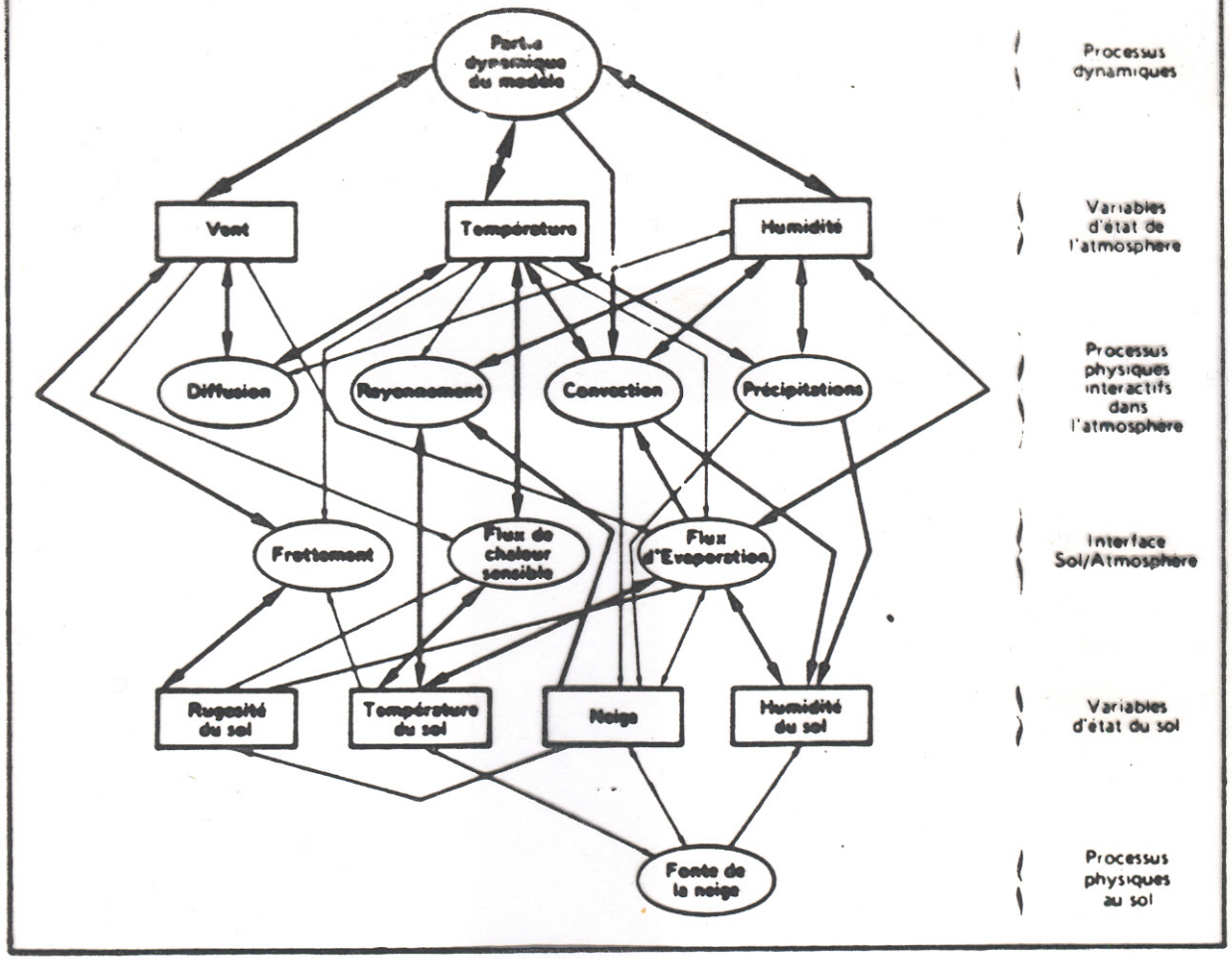
The use of *semi-implicit* schemes allows to get rid of the CFL condition, and to use longer timesteps.

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 gradually 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; project *DYNAMICO*, *Dynamical Core on Icosahedral Grid*, Th. Dubos, IPSL).

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 that take place in the boundary layer 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



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

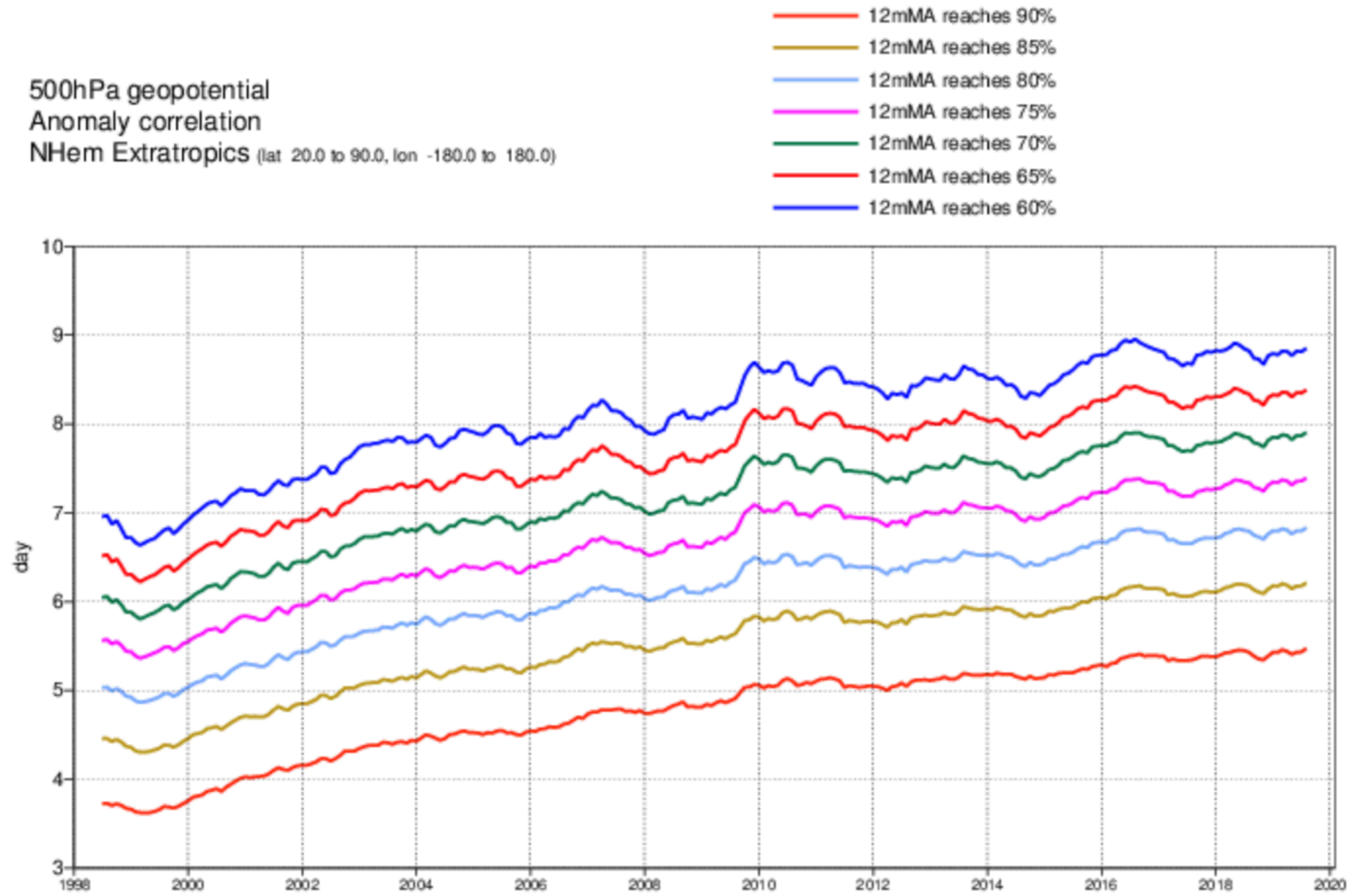
Results extracted from :

T. Haiden *et al.*, 2019, *Evaluation of ECMWF forecasts, including the 2019 upgrade*, Technical Memorandum 853, ECMWF, Reading, UK.

Available at the address :

https://www.ecmwf.int/sites/default/files/elibrary/2019/19277-evaluation-ecmwf-forecasts-including-2019-upgrade.pdf

(see also site of ECMWF)



Lead time ACC reaching thresholds

Spatial correlation between anomalies from climatology of forecast and verifying analysis

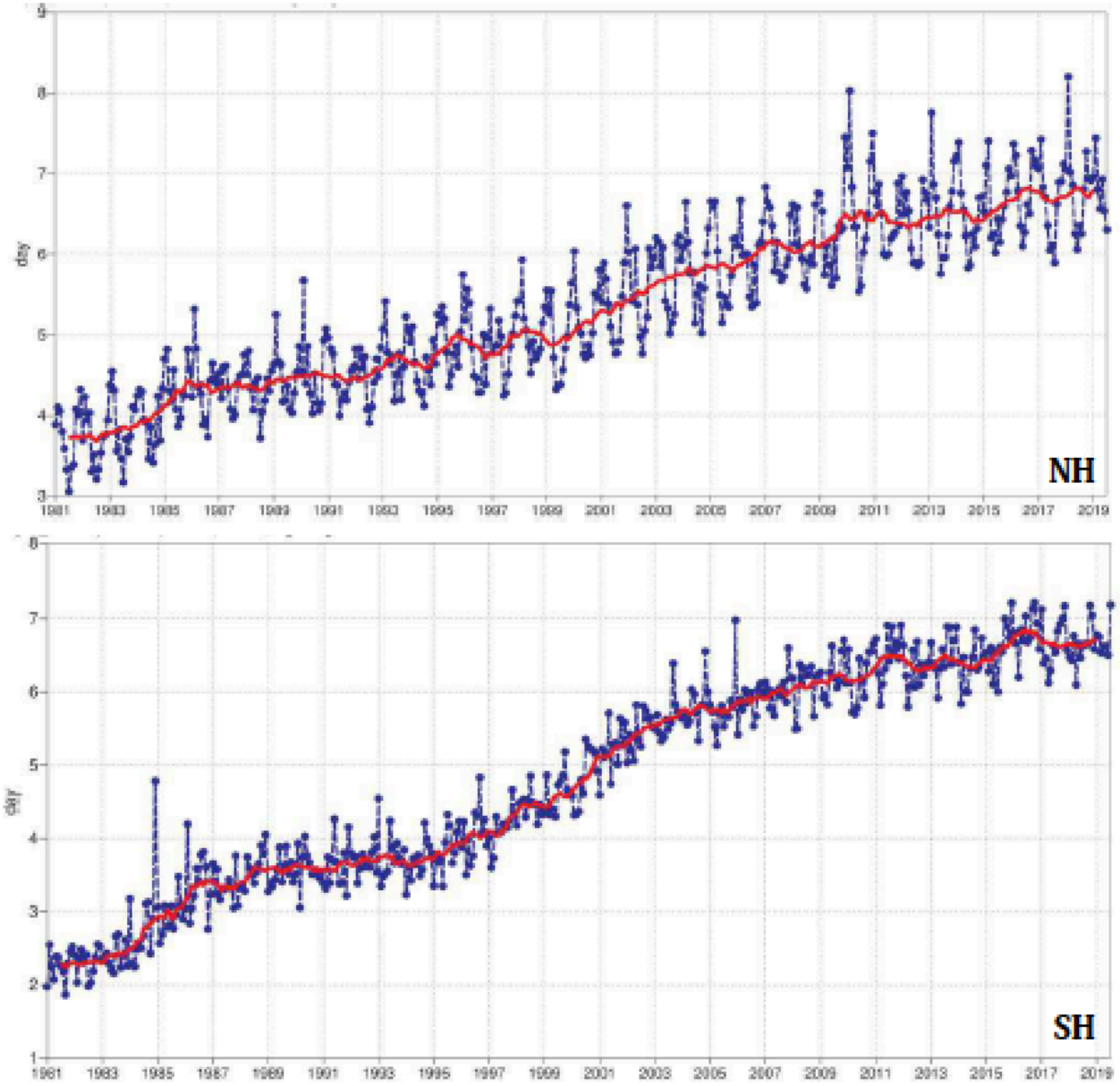


Figure 4: Primary headline score for the high-resolution forecasts. Evolution with time of the 500 hPa geopotential height forecast performance – each point on the curves is the forecast range at which the monthly mean (blue lines) or 12-month mean centred on that month (red line) of the forecast anomaly correlation (ACC) with the verifying analysis falls below 80% for Europe (top), northern hemisphere extratropics (centre) and southern hemisphere extratropics (bottom).

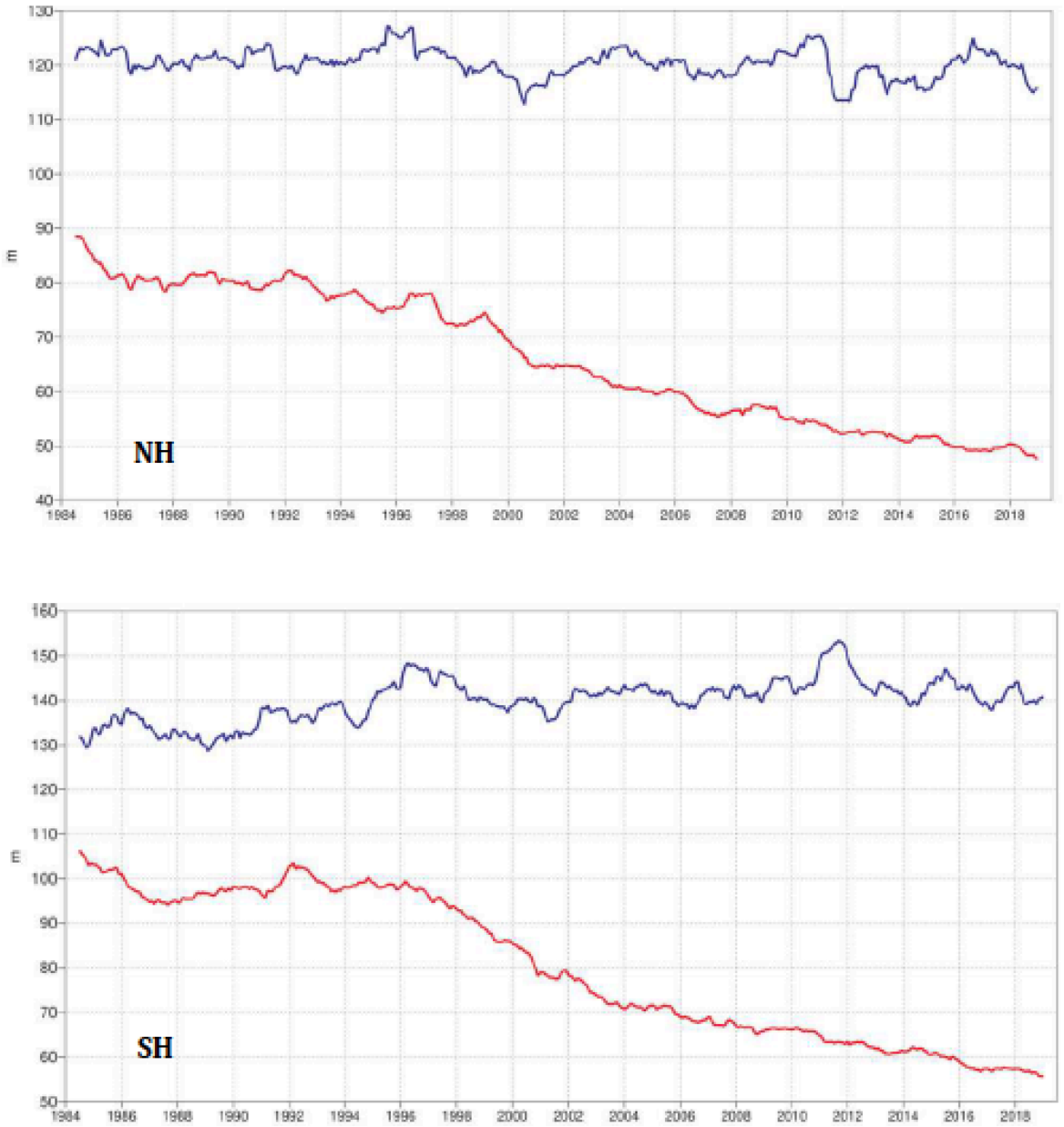
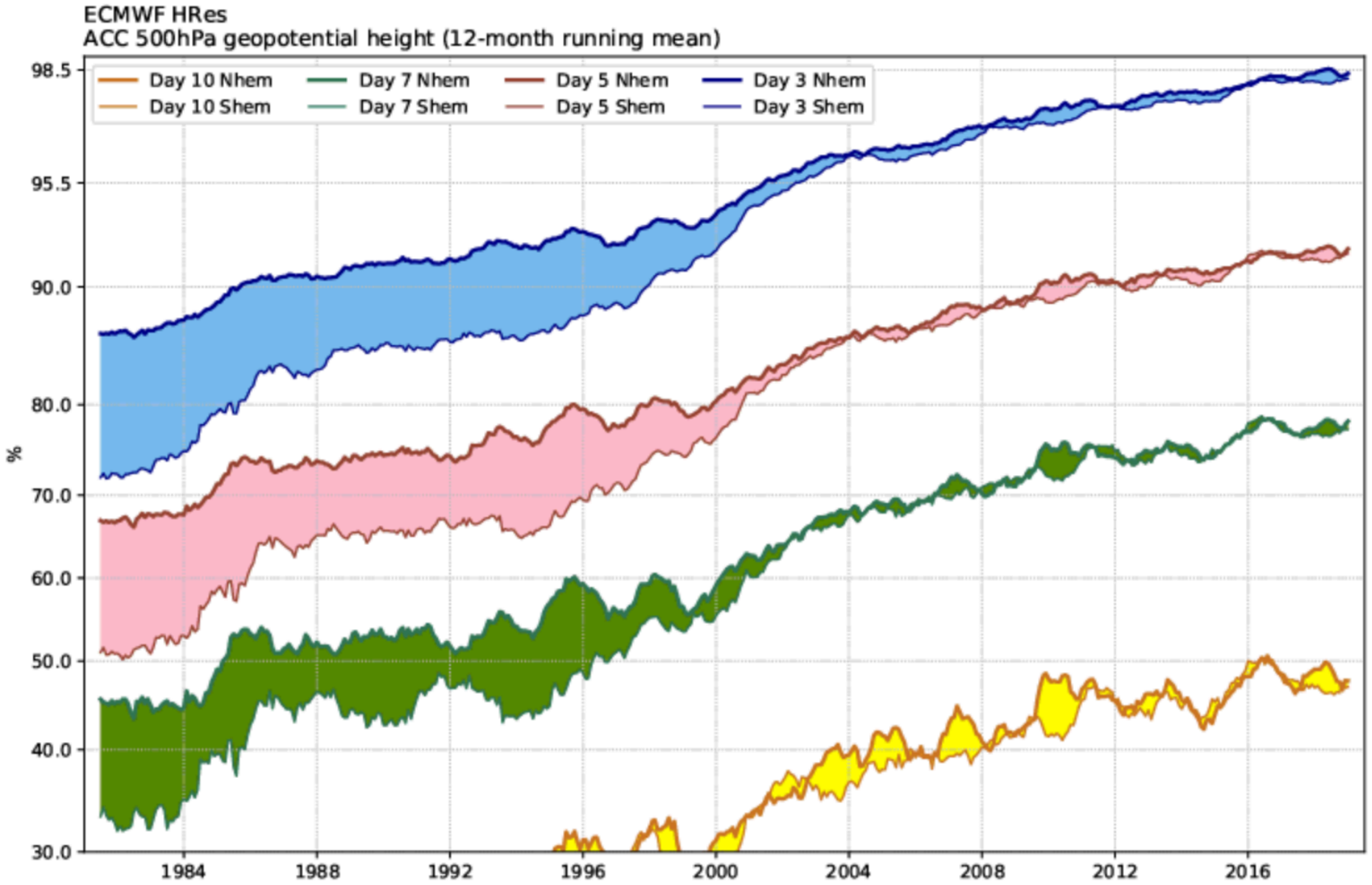


Figure 5: Root mean square (RMS) error of forecasts of 500 hPa geopotential height (m) at day 6 (red), verified against analysis. For comparison, a reference forecast made by persisting the analysis over 6 days is shown (blue). Plotted values are 12-month moving averages; the last point on the curves is for the 12-month period August 2018–July 2019. Results are shown for the northern extra-tropics (top), and the southern extra-tropics (bottom).


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
1984
1988
1992
1996
2000
2004
2008
2012
2016

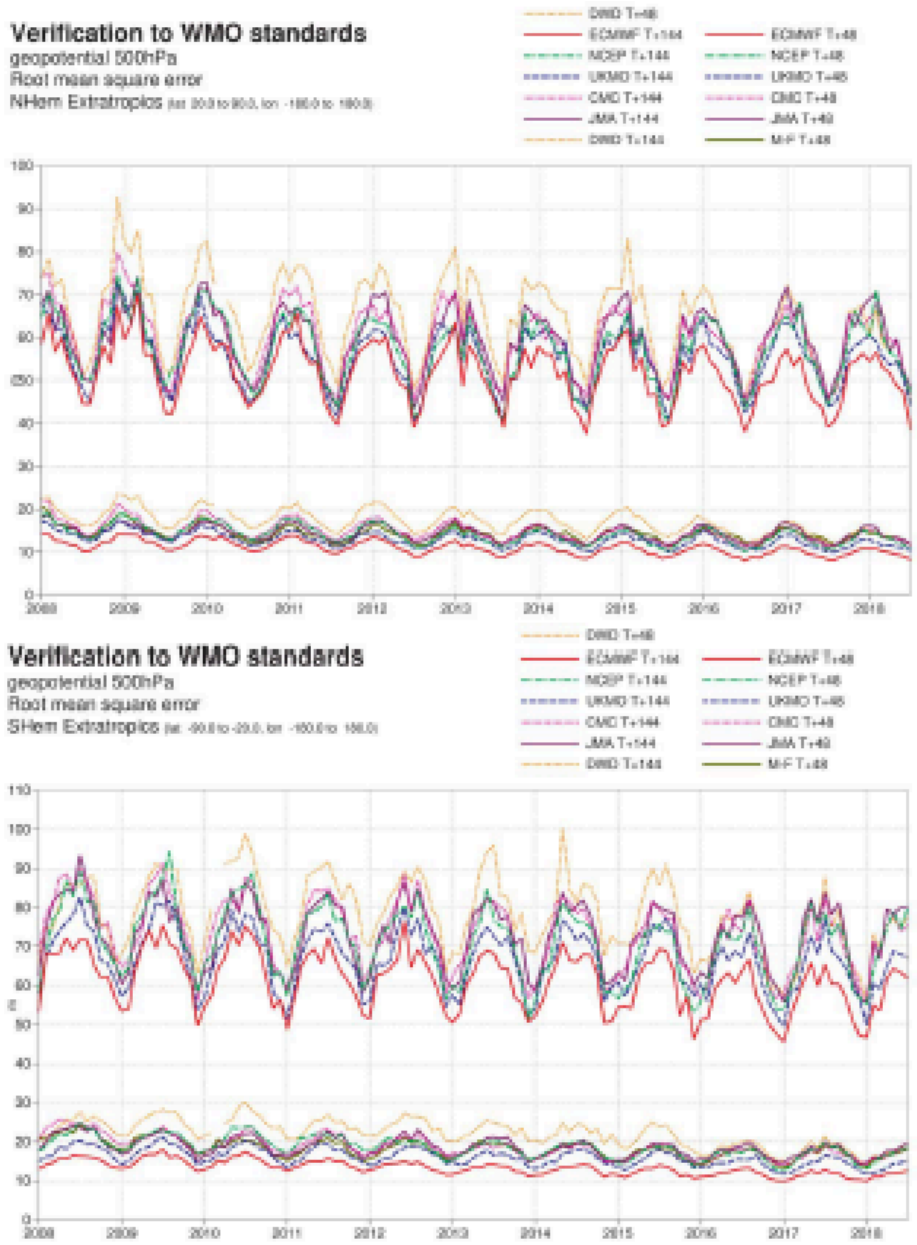


Figure 14: WMO-exchanged scores from global forecast centres. RMS error of 500 hPa geopotential height over northern (top) and southern (bottom) extratropics. In each panel, the upper curves show the six-day forecast error and the lower curves show the two-day forecast error of model runs initiated at 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.

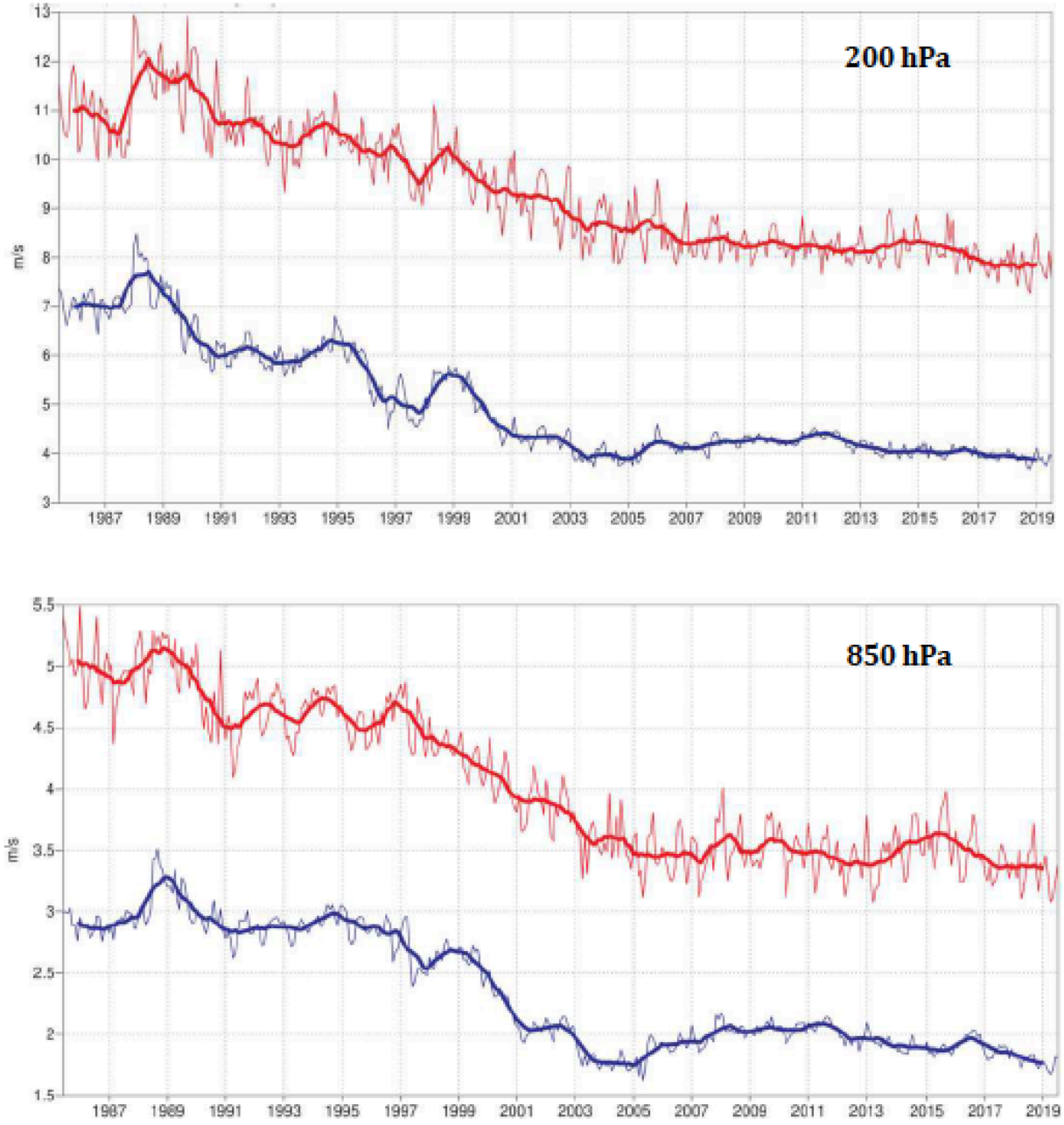


Figure 13: Forecast performance in the tropics. Curves show the monthly average RMS vector wind errors at 200 hPa (top) and 850 hPa (bottom) for one-day (blue) and five-day (red) forecasts, verified against analysis. 12-month moving average scores are also shown (in bold).

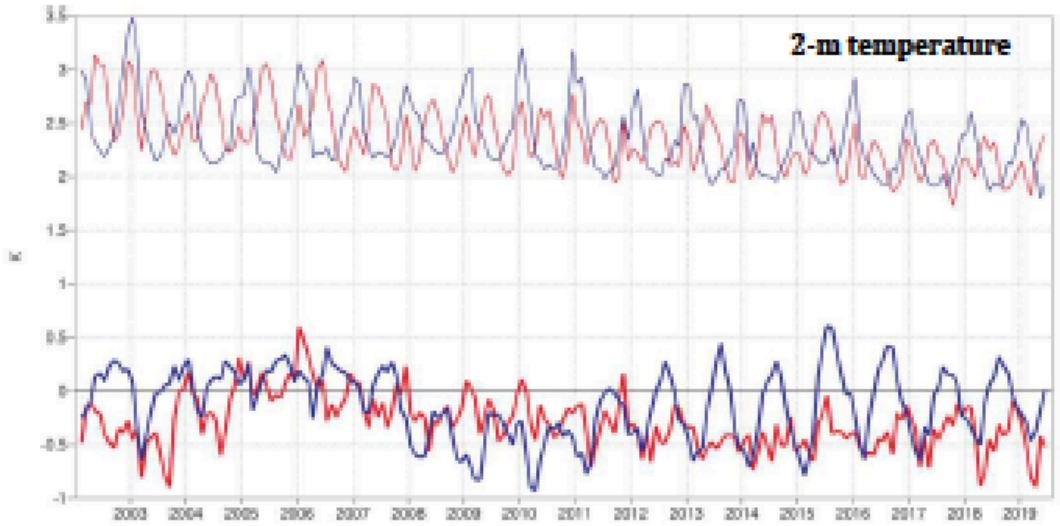


Figure 21: Verification of 2 m temperature forecasts against European SYNOP data on the GTS for 60-hour (night-time, blue) and 72-hour (daytime, red) forecasts. Lower pair of curves shows bias, upper curves are standard deviation of error.

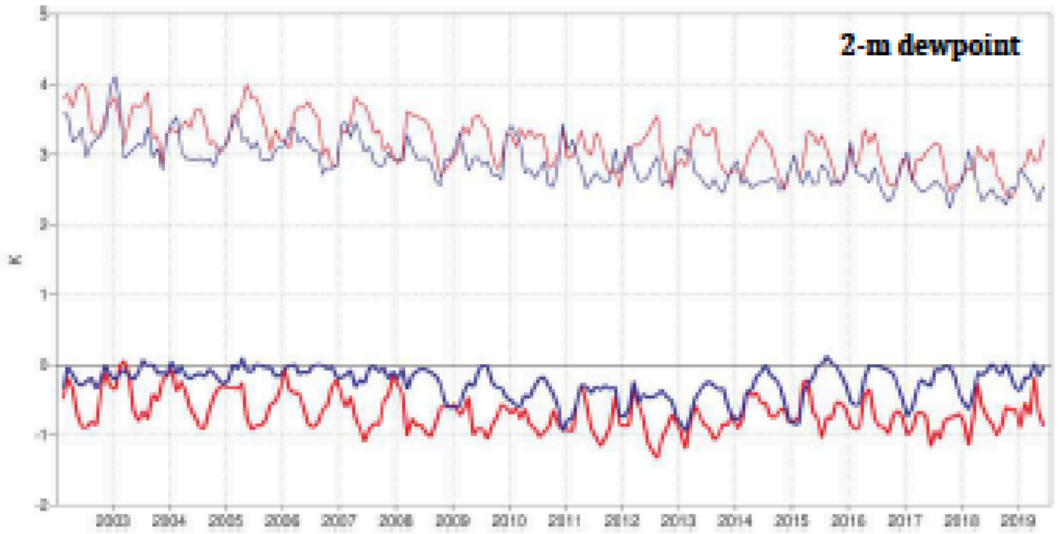


Figure 22: Verification of 2 m dew point forecasts against European SYNOP data on the Global Telecommunication System (GTS) for 60-hour (night-time, blue) and 72-hour (daytime, red) forecasts. Lower pair of curves shows bias, upper curves show standard deviation of error.

Night time: blue curves
Day time: red curves

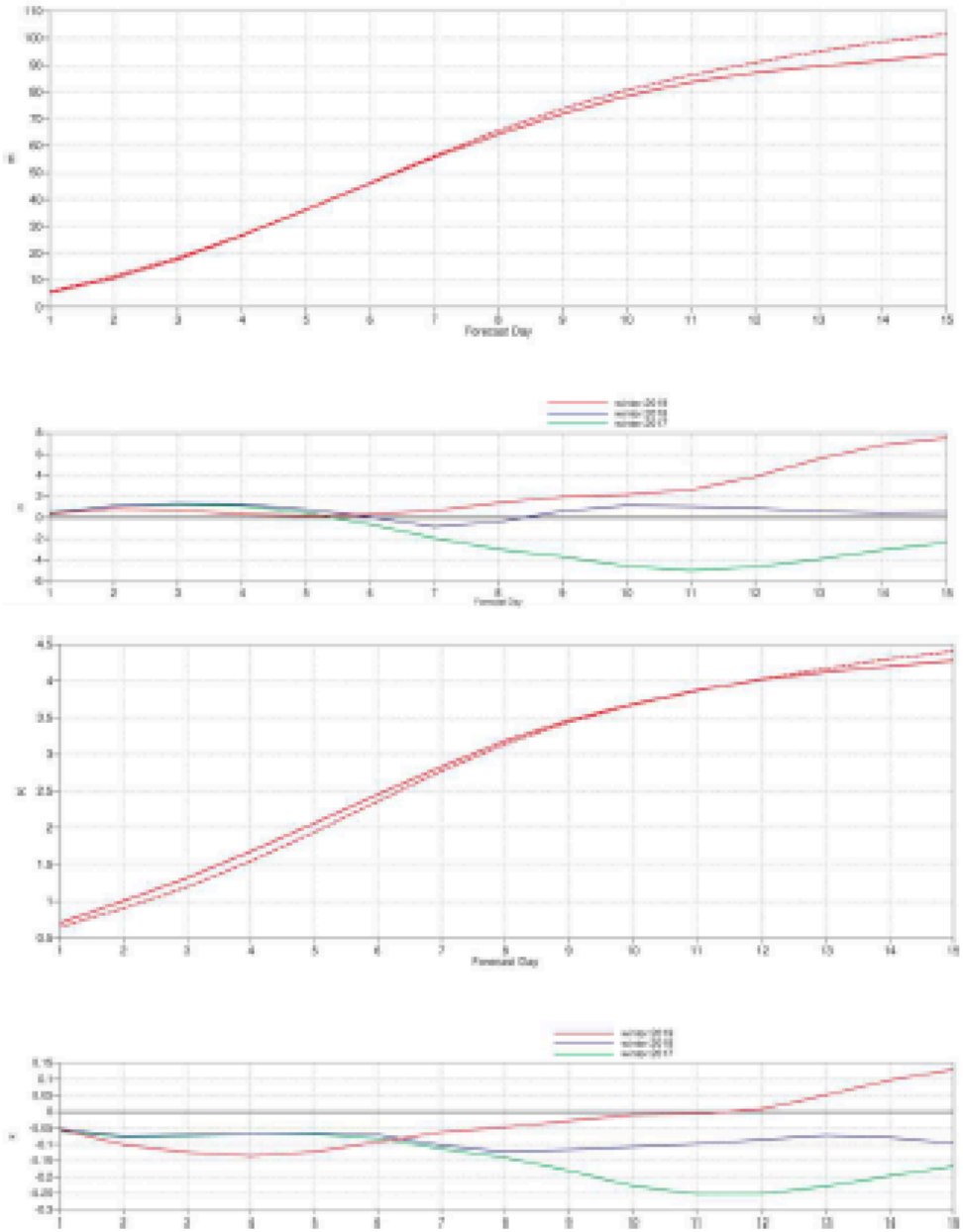


Figure 9: Ensemble spread (standard deviation, dashed lines) and RMS error of ensemble-mean (solid lines) for winter 2018–2019 (upper figure in each panel), and differences of ensemble spread and RMS error of ensemble mean for last three winter seasons (lower figure in each panel, negative values indicate 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.

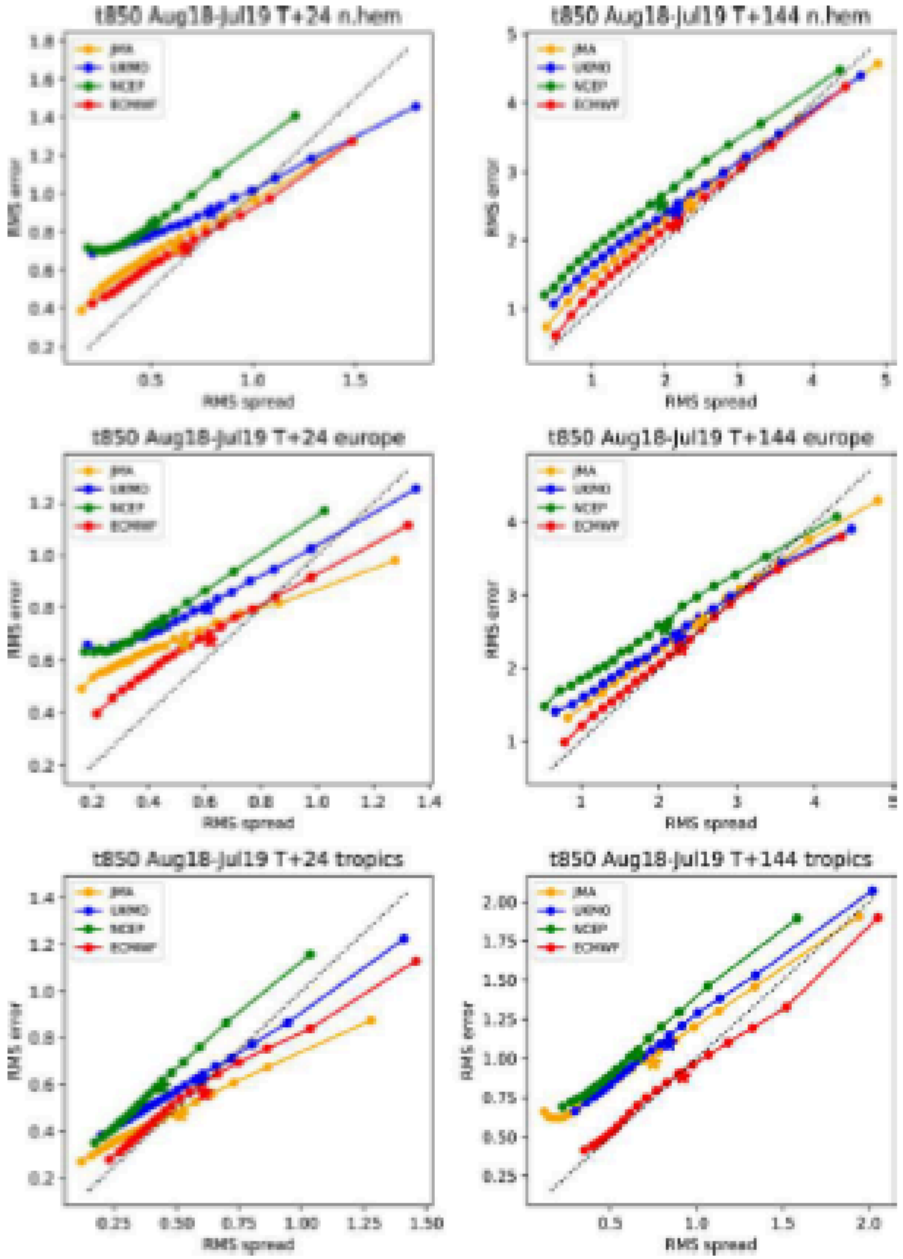
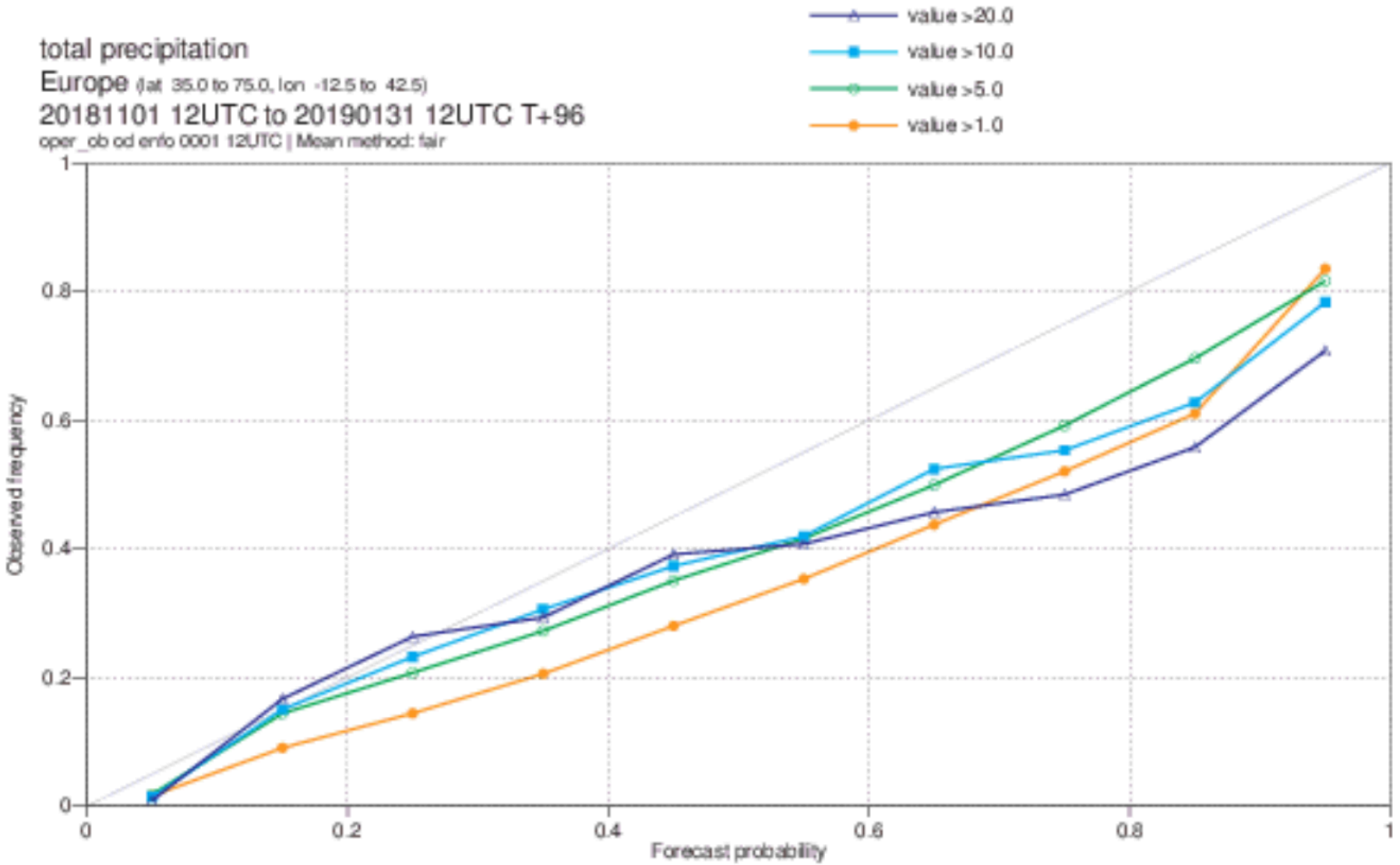


Figure 11: Ensemble spread reliability of different global models for 850 hPa temperature for the period August 2018–July 2019 in the northern hemisphere extra-tropics (top), Europe (centre), and the tropics (bottom) for day 1 (left) and day 6 (right), verified against analysis. Circles show error for different values of spread, stars show average error-spread relationship. Due to random outages in the data supply, NCEP curves are based on a reduced data set (60%).


total precipitation
Europe (lat 35.0 to 75.0, lon -12.5 to 42.5)
20181101 12UTC to 20190131 12UTC T+96
oper_ob od enfo 0001 12UTC | Mean method: fair
value >20.0
value >10.0
value >5.0
value >1.0
Observed frequency
Forecast probability
0
0.2
0.4
0.6
0.8
1

ECMWF

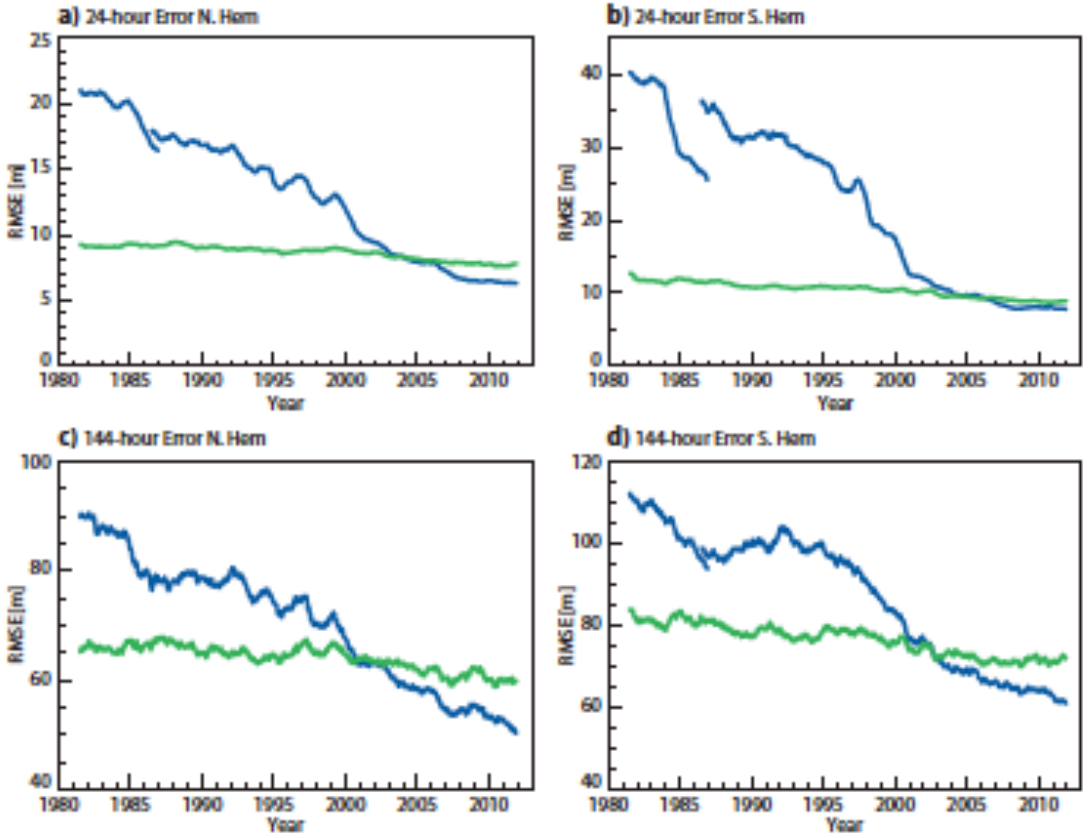


FIG. 3. Evolution of forecast errors from 1981 to 2012 for N.Hem (a and c) and S.Hem (b and d). Operational forecasts (blue) and ERA Interim (green). Note that before 1986 the operational analysis is used to verify the operational forecasts, after 1986 ERA Interim is used for the verification (with an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

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

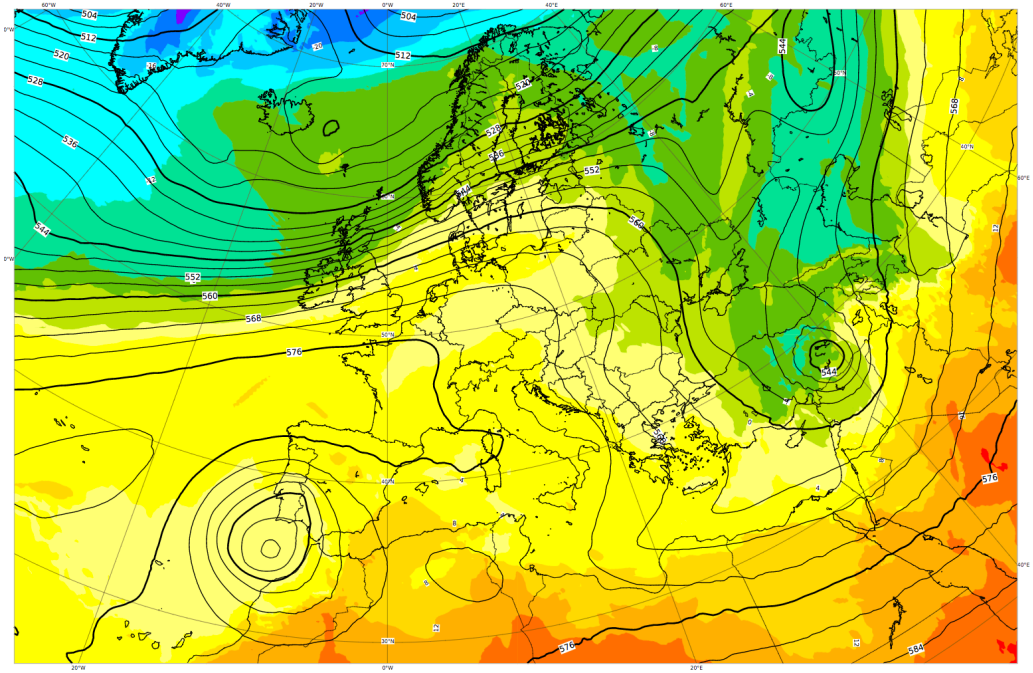
Remaining problems

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

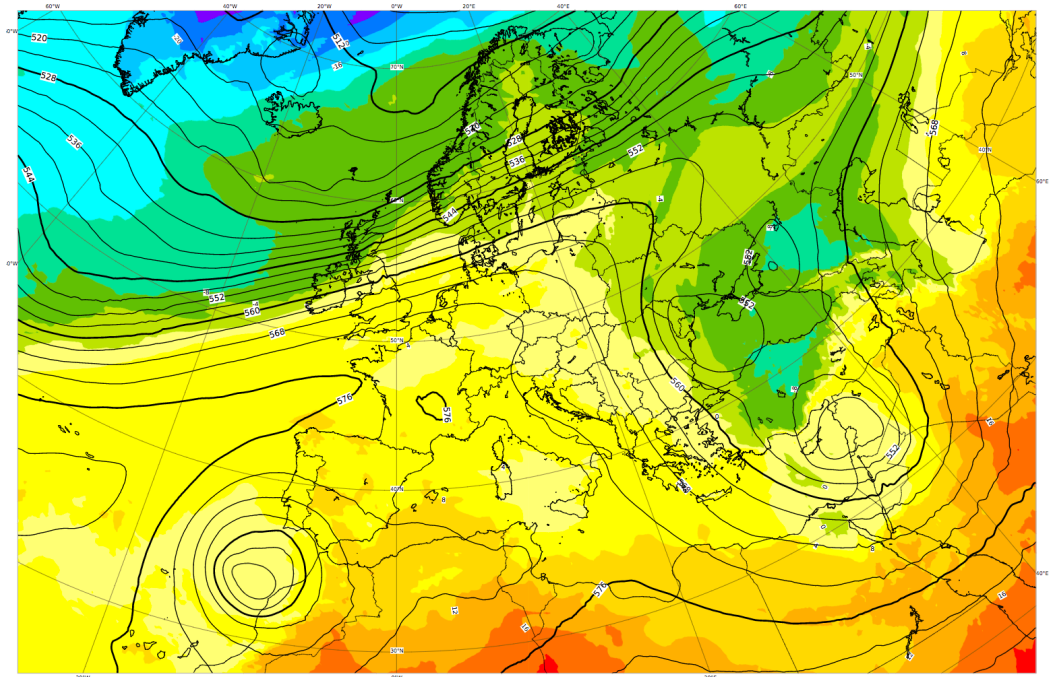
- Exchanges with ocean or continental surface (heat, water, momentum, …)

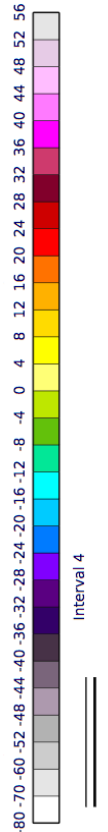
- …

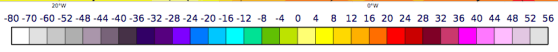
Geopotential 500 hPa and temperature at 850 hPa - Friday 13 Mar 2020, 00 UTC VT Wednesday 18 Mar 2020, 00 UTC Step 120
© ECMWF 2020



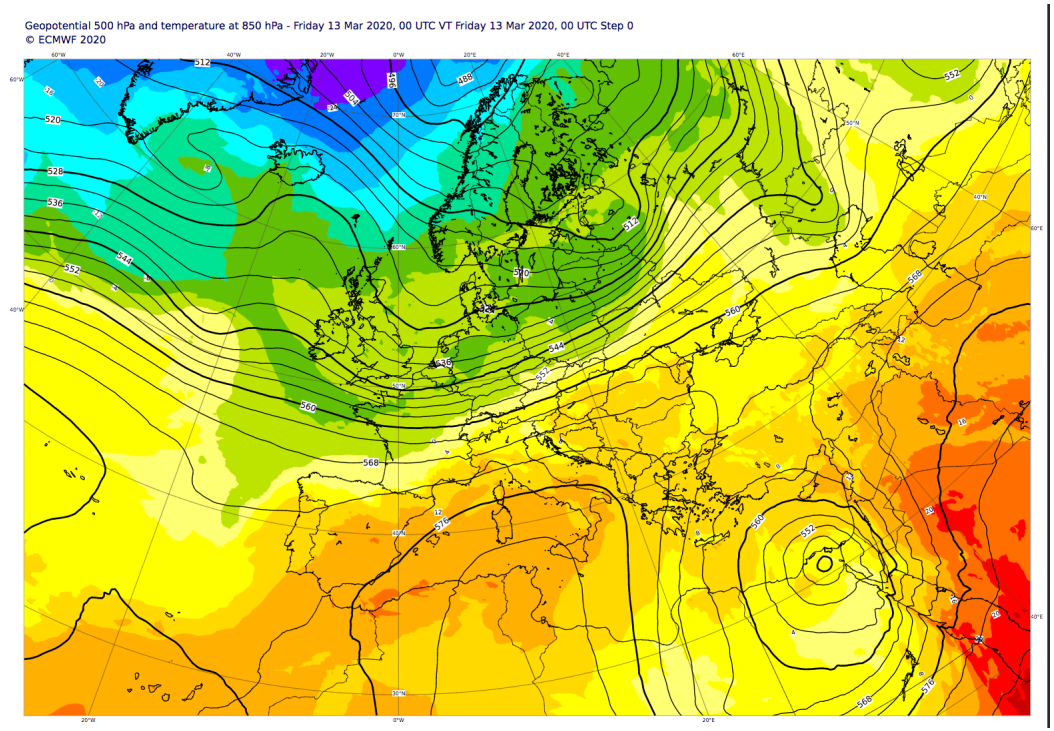
Geopotential 500 hPa and temperature at 850 hPa - Wednesday 18 Mar 2020, 00 UTC VT Wednesday 18 Mar 2020, 00 UTC Step 0
© ECMWF 2020

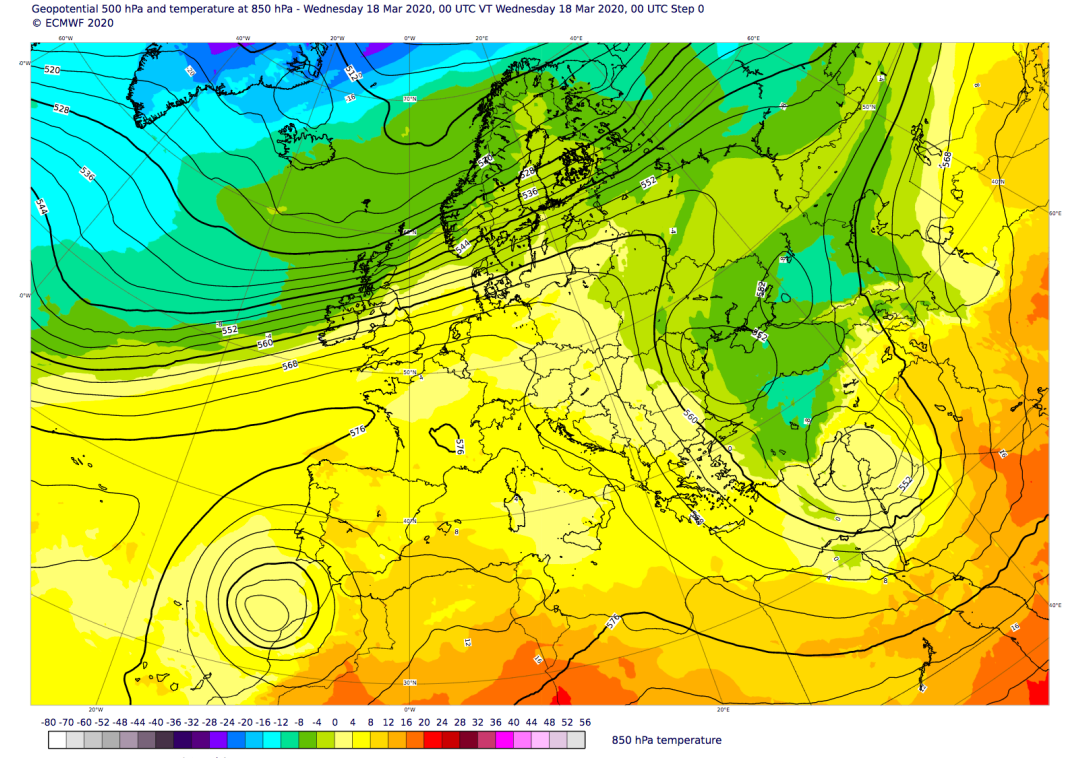


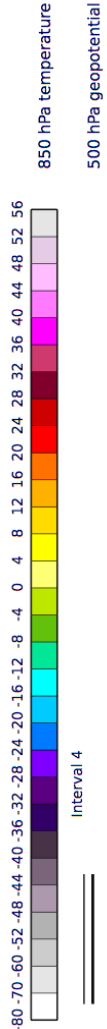




2020

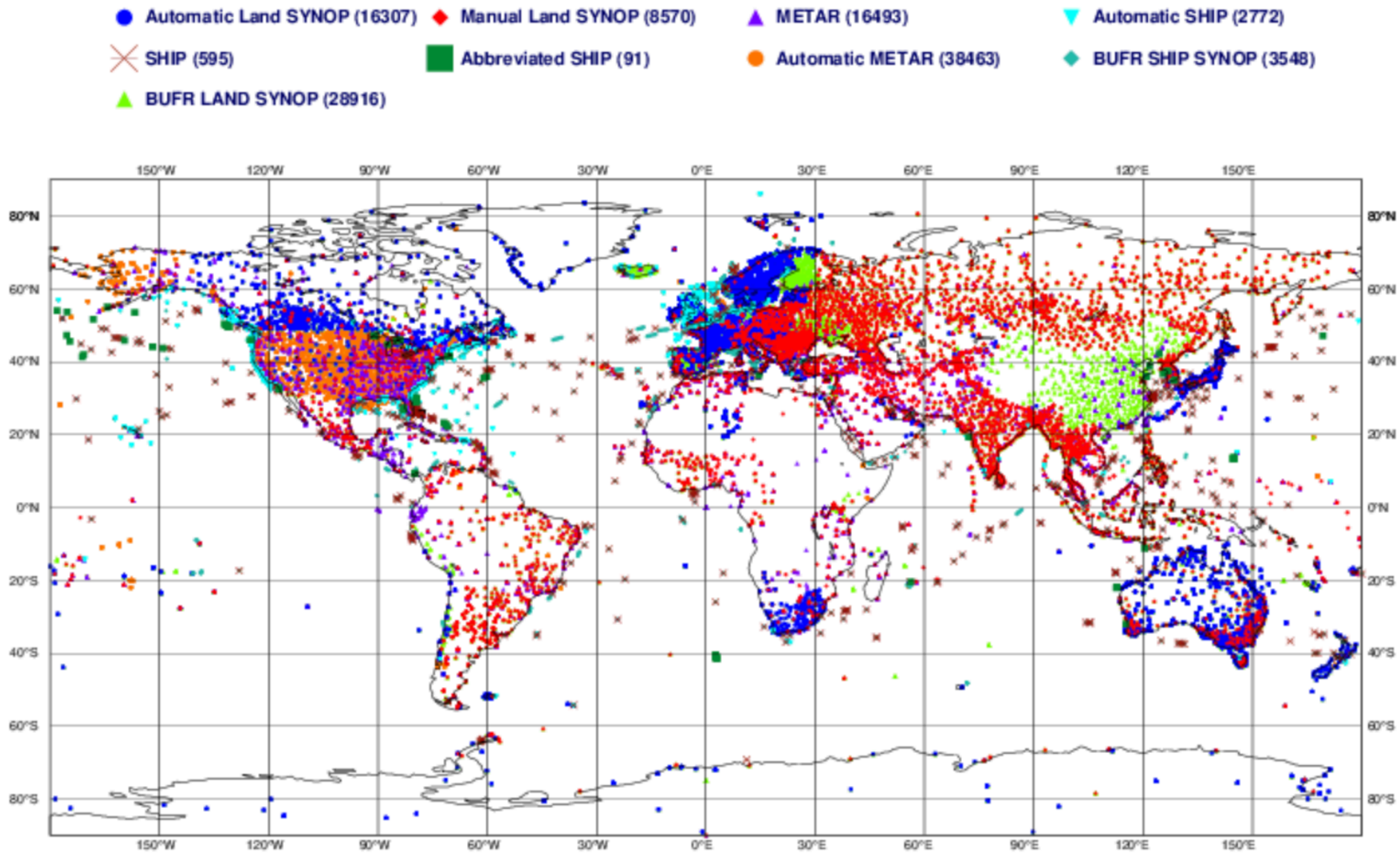

Geopotential 500 hPa and temperature at 850 hPa - Friday 13 Mar 2020, 00 UTC VT Friday 13 Mar 2020, 00 UTC Step 0
© ECMWF 2020


Geopotential 500 hPa and temperature at 850 hPa - Wednesday 18 Mar 2020, 00 UTC VT Wednesday 18 Mar 2020, 00 UTC Step 0
© ECMWF 2020
-80 -70 -60 -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 56
850 hPa temperature

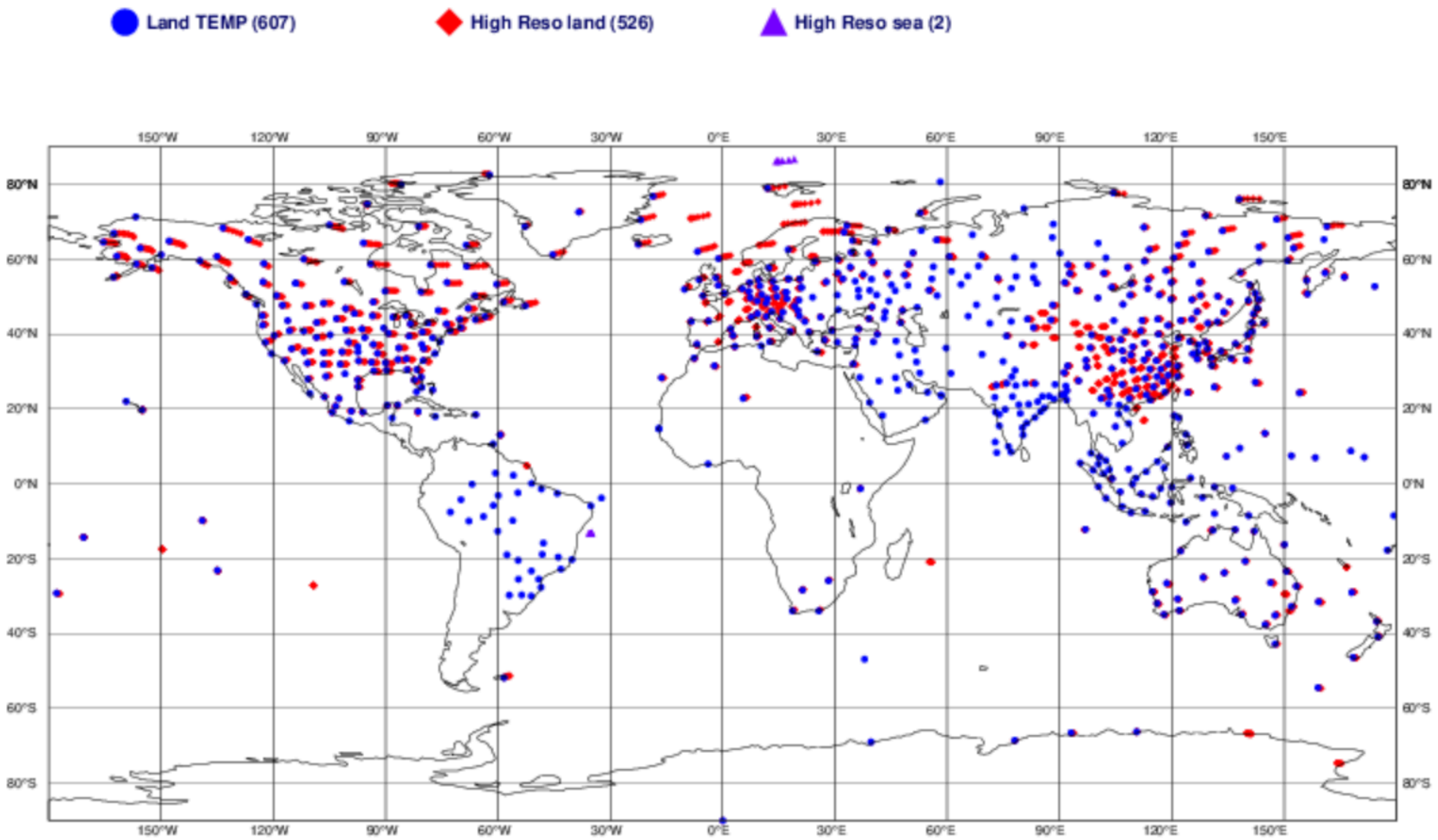

850 hPa temperature
500 hPa geopotential
-80 -70 -60 -52 -48 -44 -40 -36 -32 -28 -24 -20 -16 -12 -8 -4 0 4 8 12 16 20 24 28 32 36 40 44 48 52 56
Interval 4

2020

ECMWF data coverage (all observations) - SYNOP-SHIP-METAR
25/03/2020 00
Total number of obs = 115755

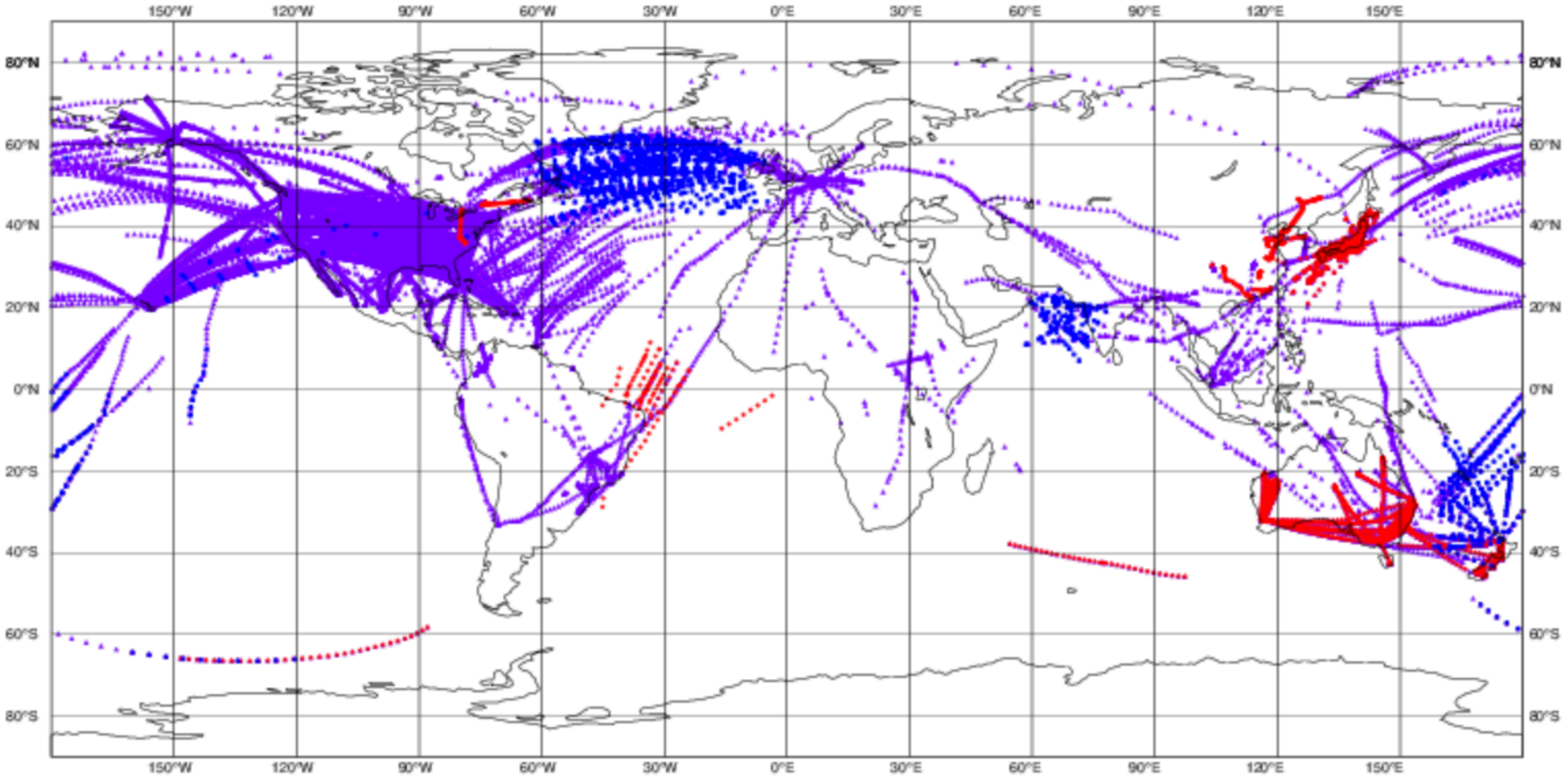


ECMWF data coverage (all observations) - RADIOSONDE
25/03/2020 00
Total number of obs = 1135

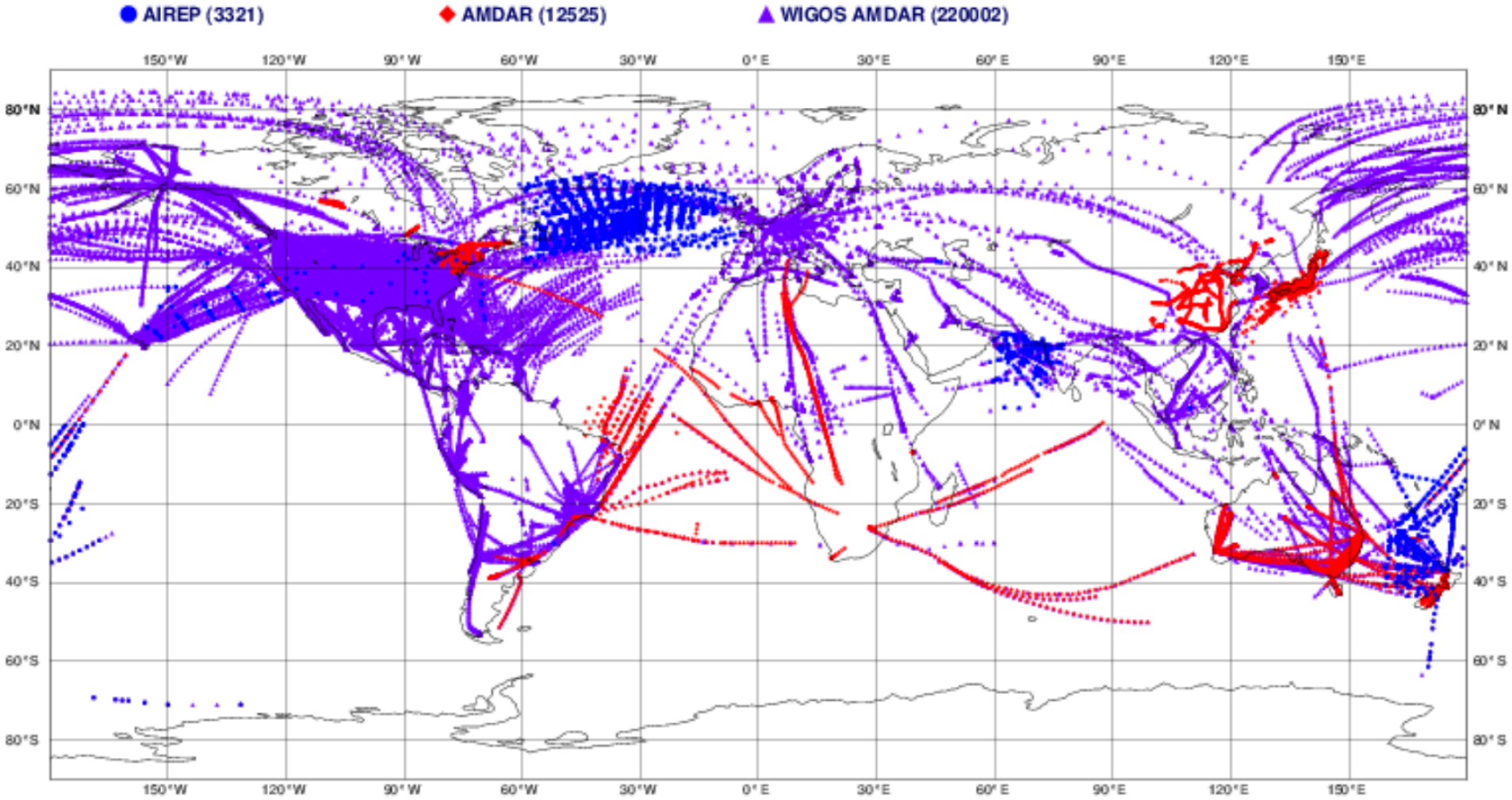


ECMWF data coverage (all observations) - AIRCRAFT
25/03/2020 00
Total number of obs = 191183

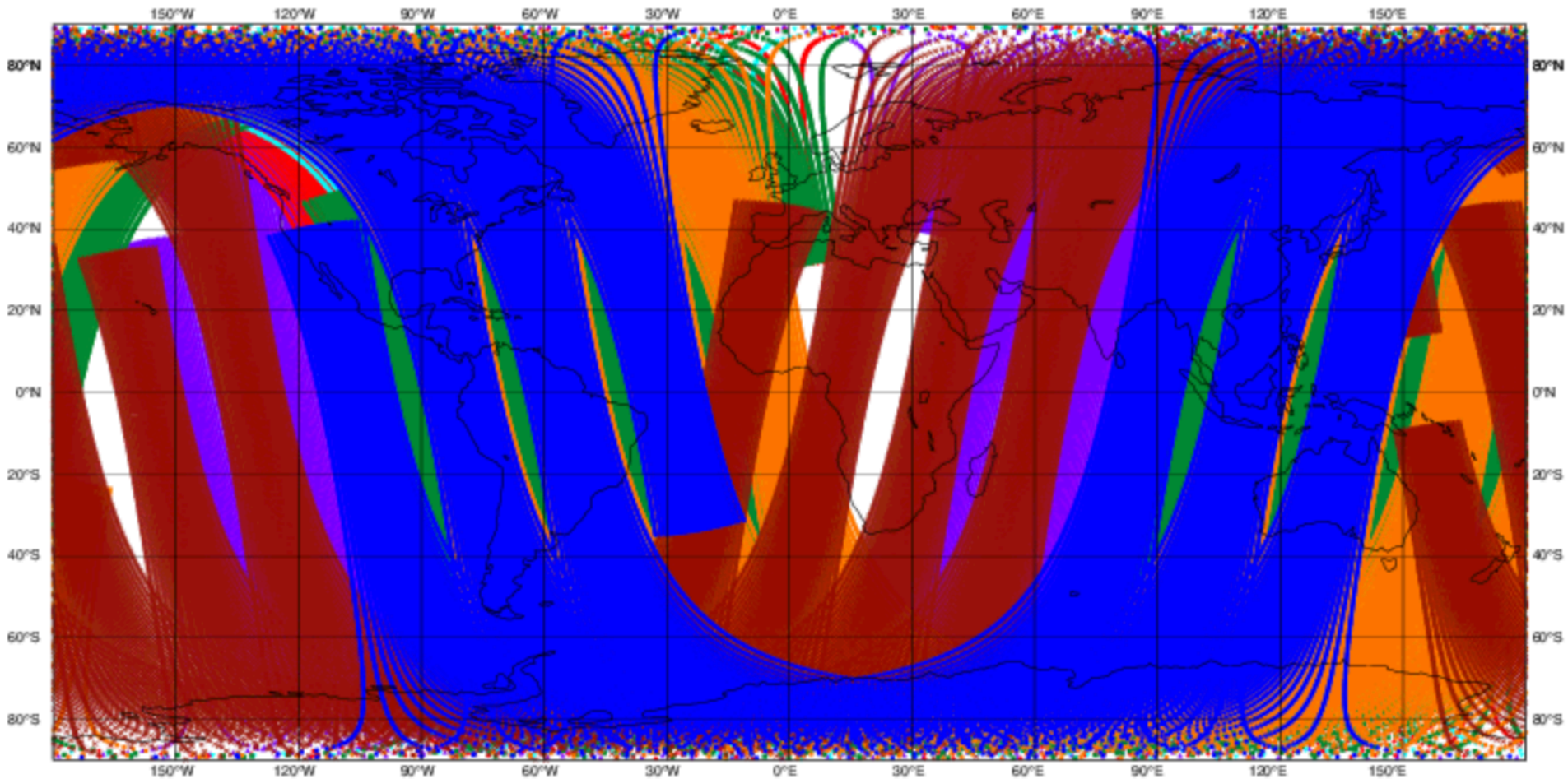
● AIREP (2357) ◆ AMDAR (7216) ▲ WIGOS AMDAR (181610)



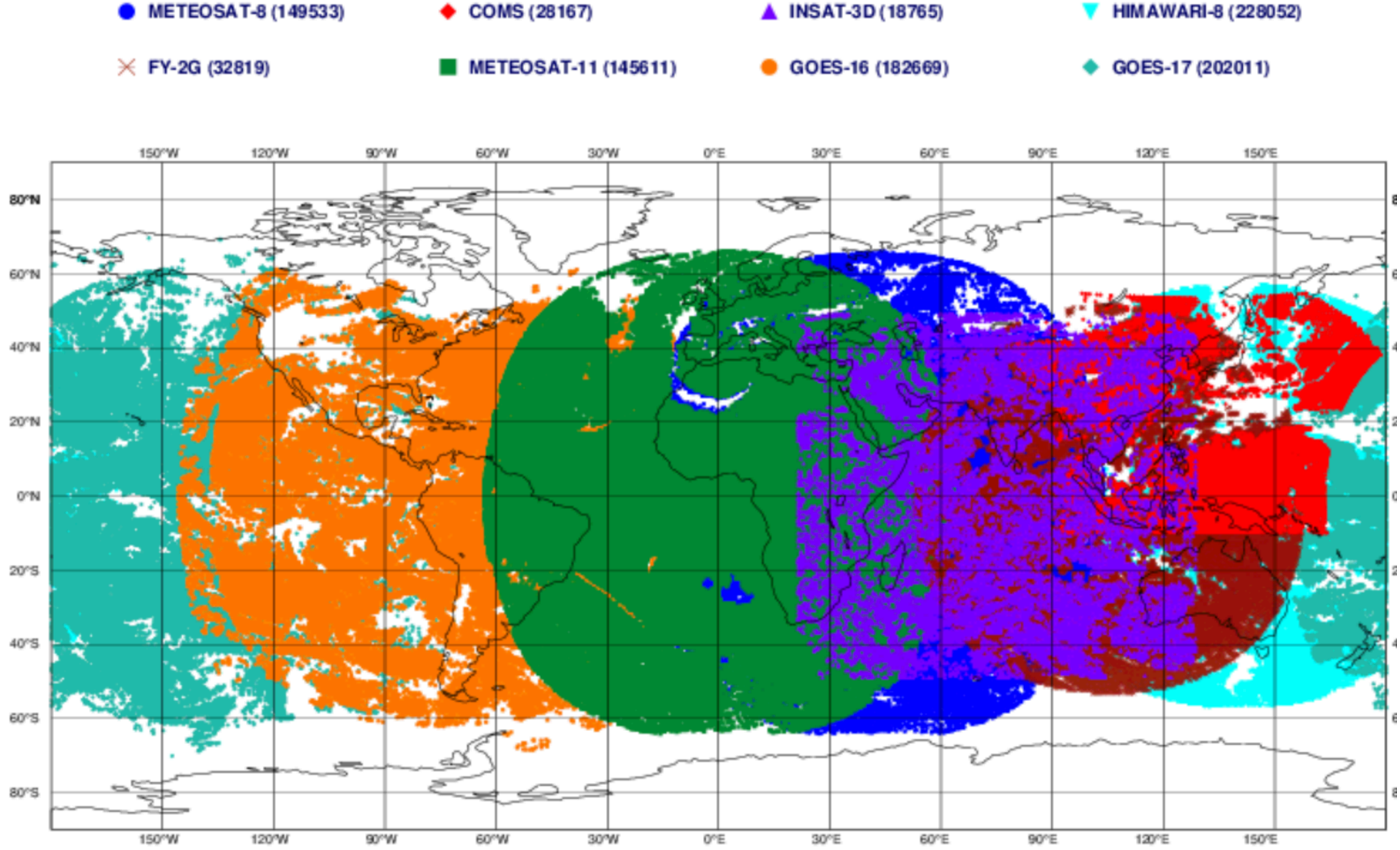
ECMWF data coverage (all observations) - AIRCRAFT
19/02/2019 00
Total number of obs = 235848



ECMWF data coverage (all observations) - AMSUA
25/03/2020 00
Total number of obs = 701479



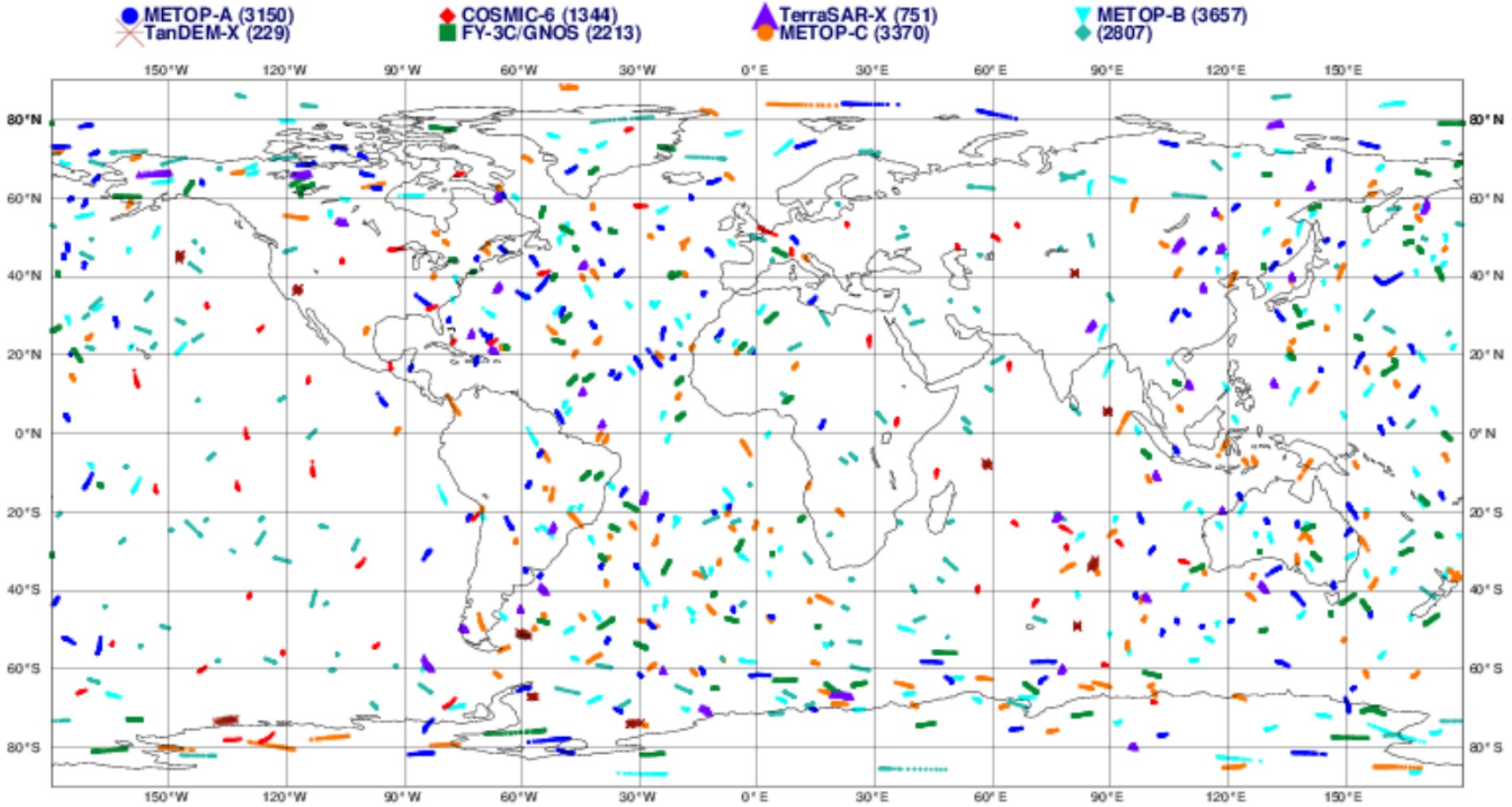
ECMWF data coverage (all observations) - AMV WV
25/03/2020 00
Total number of obs = 987627



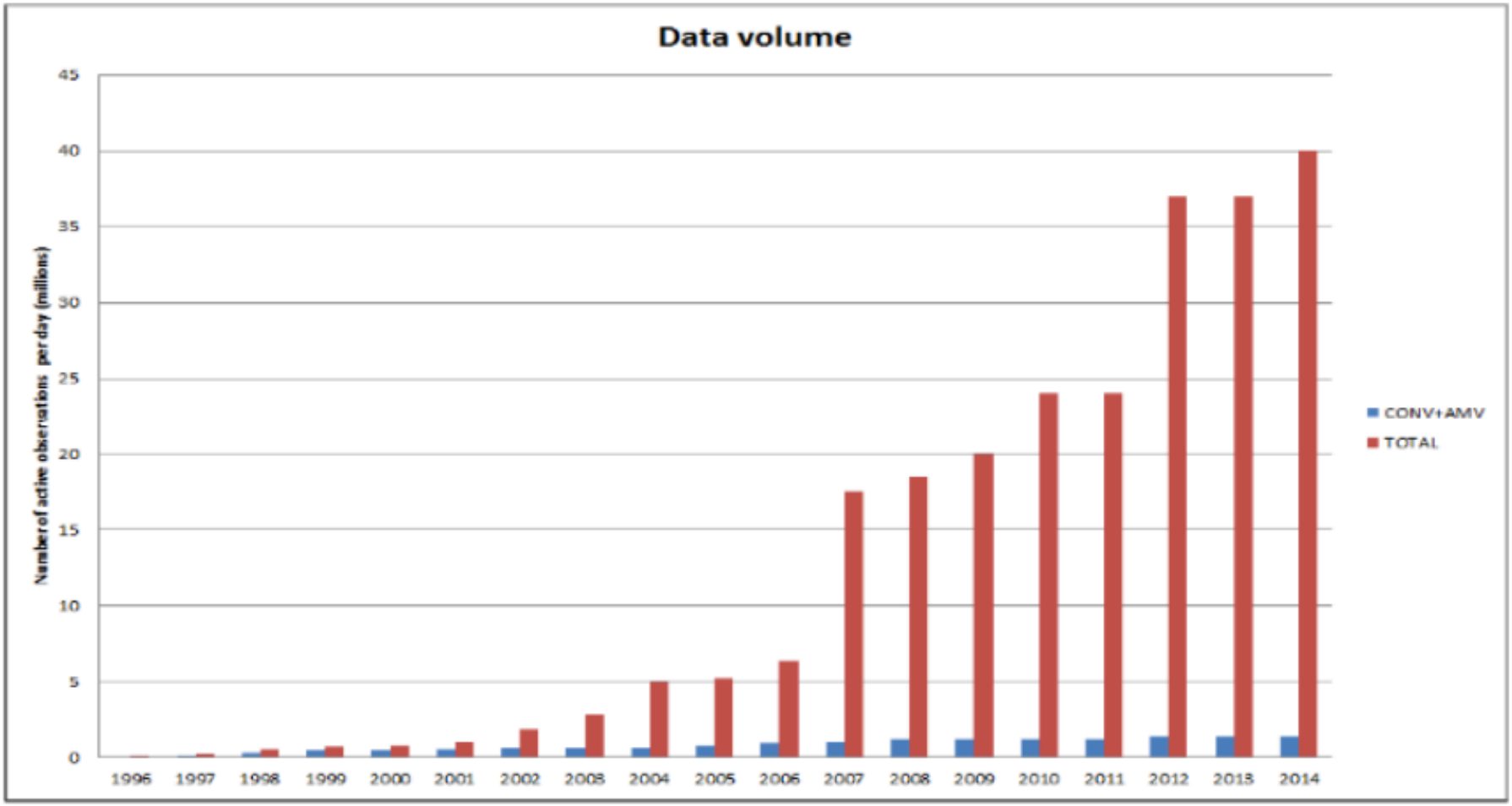
ECMWF data coverage (all observations) - GPSRO

19/02/2019 00

Total number of obs = 17521



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

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. Preliminary tests have shown that these new observations have a positive impact on the quality of the previsions, especially in the tropics and in the Southern Hemisphere.

- *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

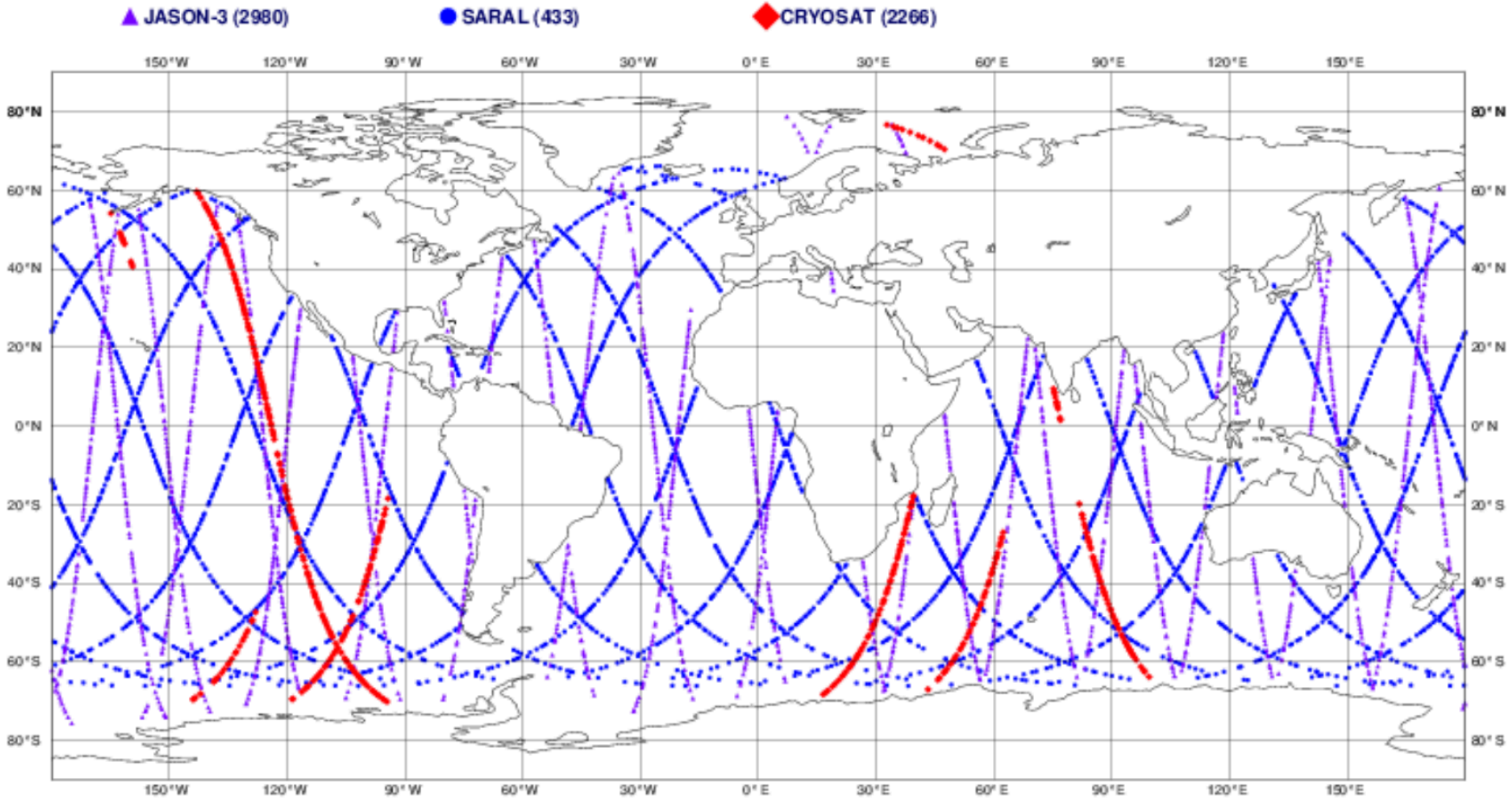
$$y = H(\boldsymbol{x})$$

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

ECMWF data coverage (all observations) - SEA LEVEL ANOMALY

20190216 00

Total number of obs = 5679



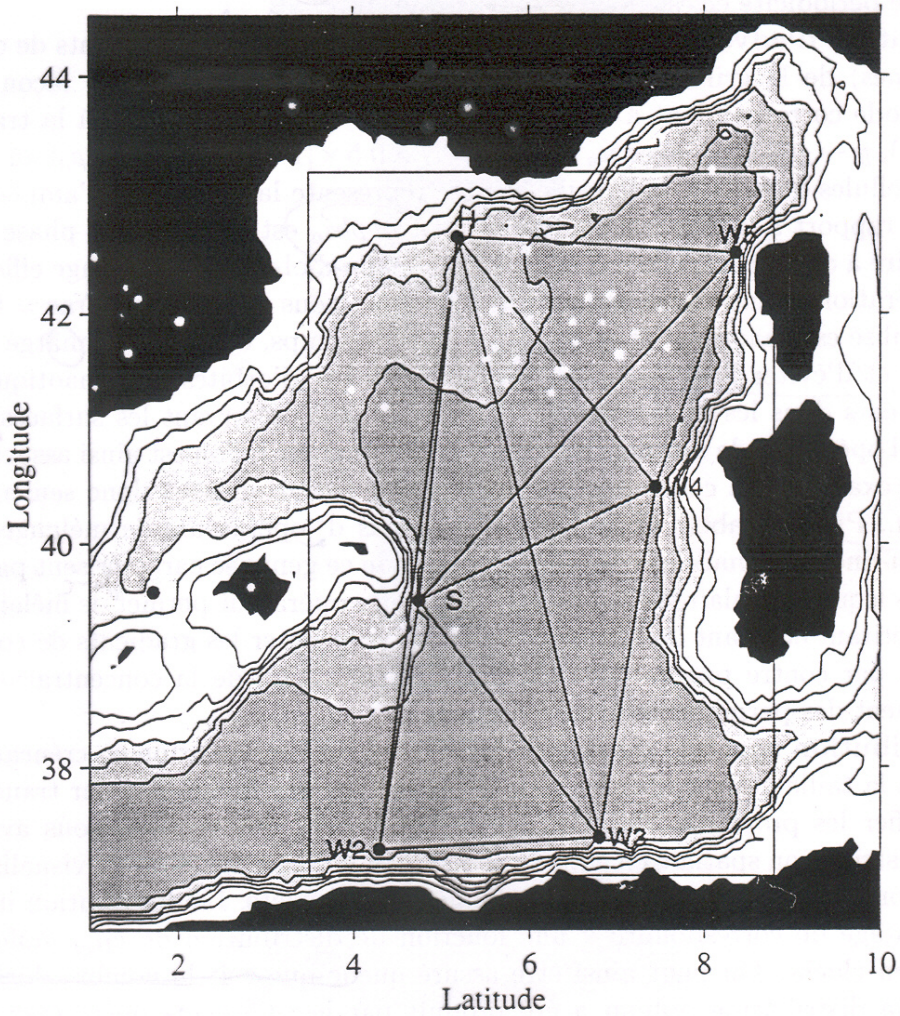


FIG. 1 – *Bassin méditerranéen occidental: réseau d'observation tomographique de l'expérience Thétis 2 et limites du domaine spatial utilisé pour les expériences numériques d'assimilation.*

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, $p \approx 4$-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

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

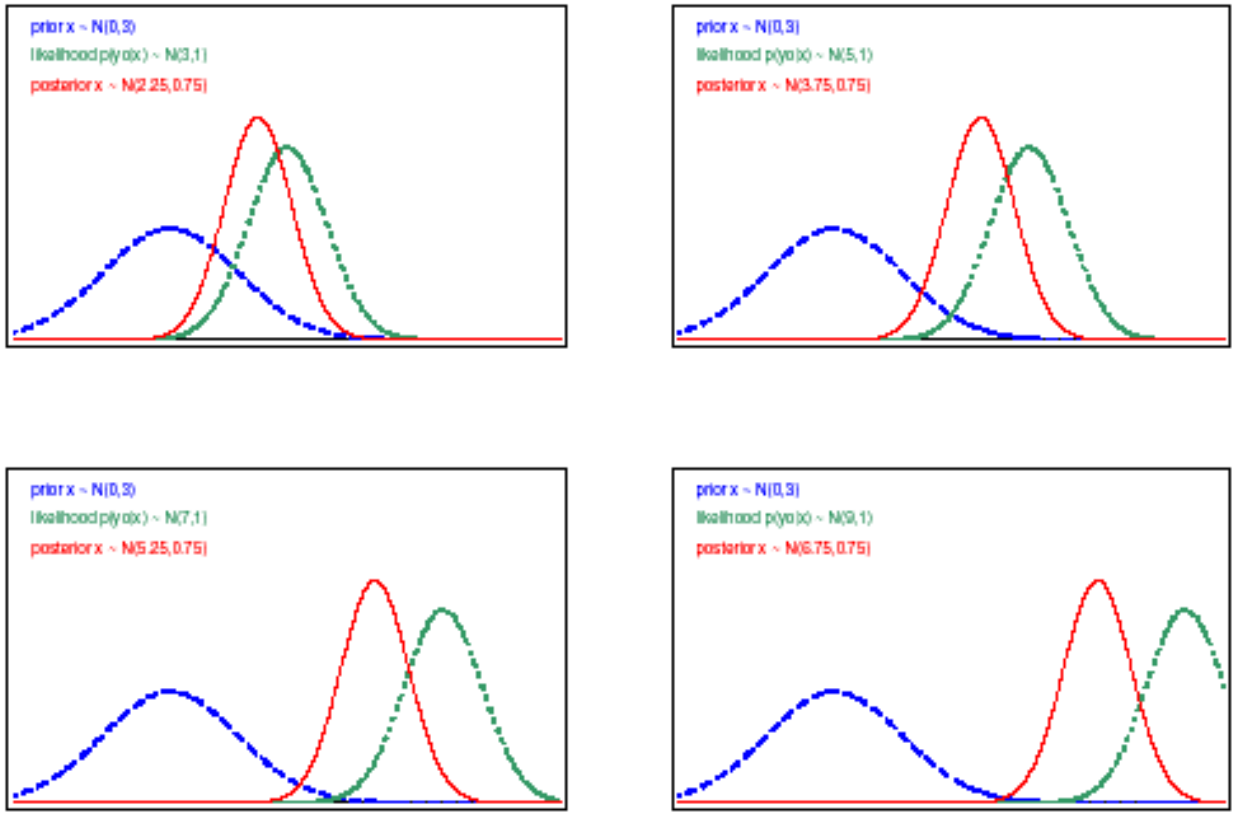


Fig. 1.1: Prior pdf $p(x)$ (dashed line), posterior pdf $p(x|y^o)$ (solid line), and Gaussian likelihood of observation $p(y^o|x)$ (dotted line), plotted against x for various values of y^o. (Adapted from Lorenc and Hammon 1988.)

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

Cours à venir

~~Jeudi 19 Mars~~
~~Jeudi 26 mars~~
Jeudi 02 avril
Jeudi 09 avril
Jeudi 16 avril
Jeudi 23 avril
Jeudi 30 avril
Jeudi 14 mai