Sorbonne Université, Université Paris-Saclay

Master 2 Sciences de l'Océan, de l'Atmosphère et du Climat (SOAC)

Parcours Météorologie, Océanographie, Climat et Ingénierie pour les Observations Spatiales (MOCIS)

Institut polytechnique de Paris

Master 2 Water, Air, Pollution and Energy at local and regional scales (WAPE)

Année 2023-2024

Course *Introduction to data assimilation* (*ADOMO*)

From numerical modelling to data assimilation

Olivier Talagrand

9 January 2024

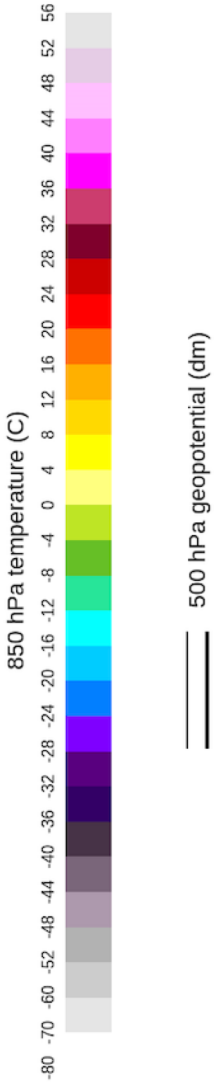
- What is assimilation ?
  - *Numerical weather prediction. Principles and performances*
  - *Definition of initial conditions*

- Bayesian Estimation

- One first step towards assimilation : ‘Optimal Interpolation’

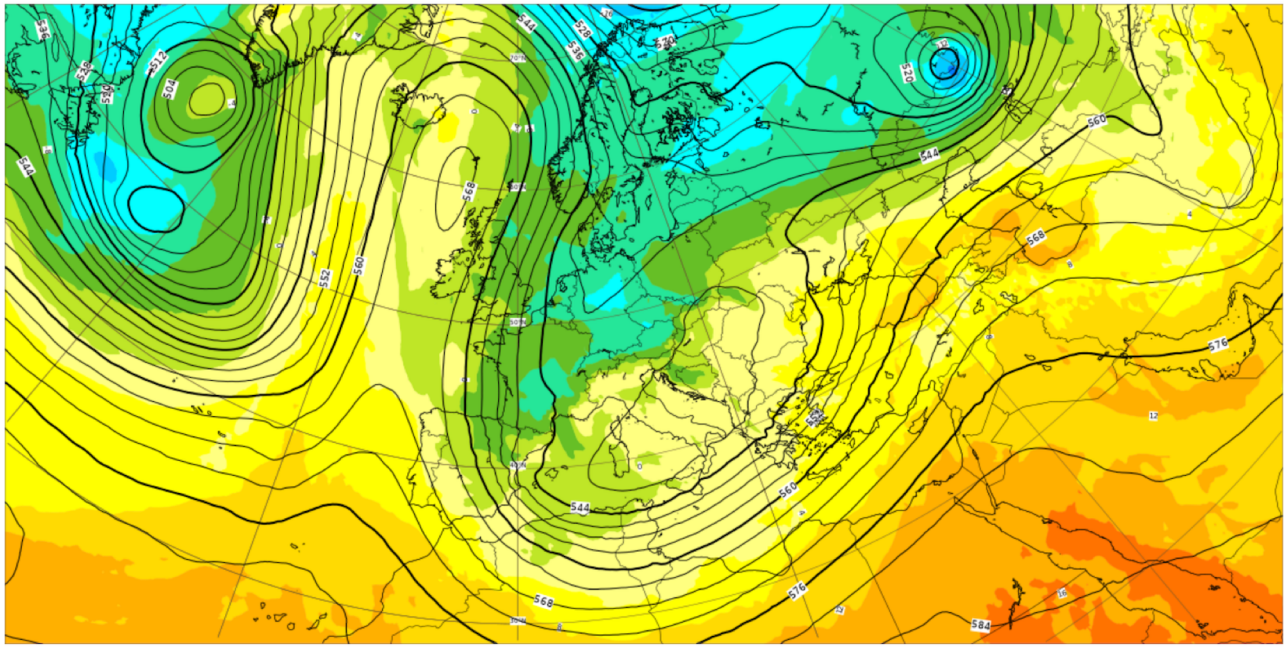
- The temporal dimension : Kalman Filter and Variational Assimilation

ECMWF

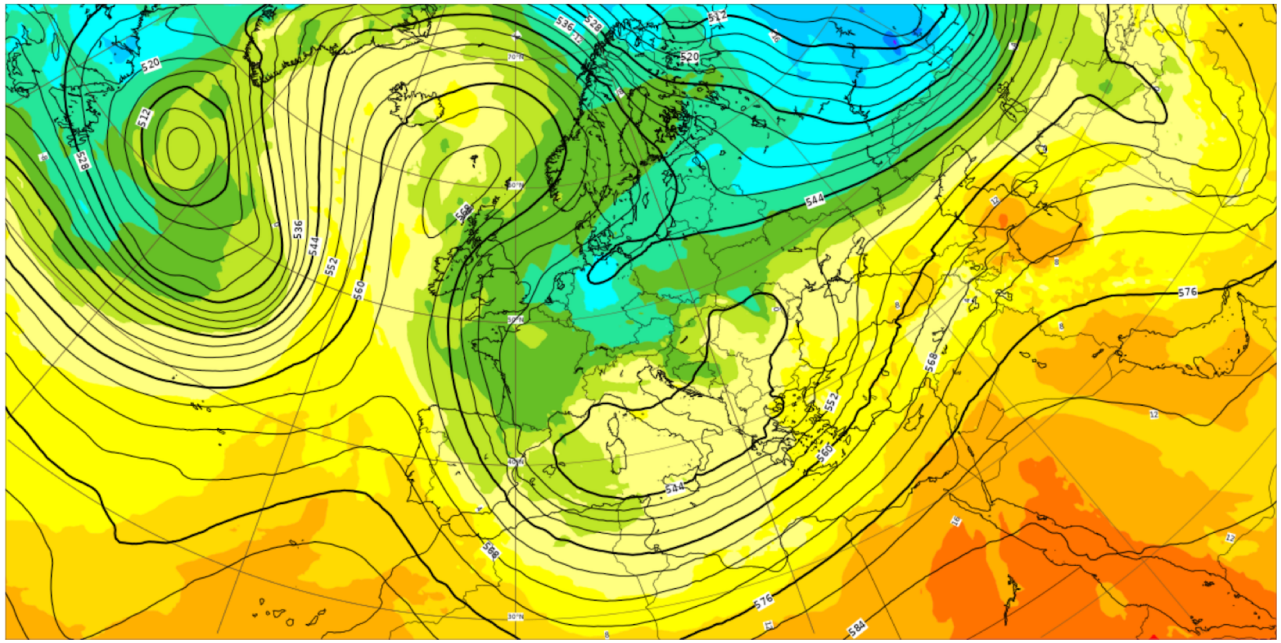


500 hPa geopotential (dm)

Base time: Tue 02 Jan 2024 00 UTC Valid time: Mon 08 Jan 2024 00 UTC (+144h) Area : Europe

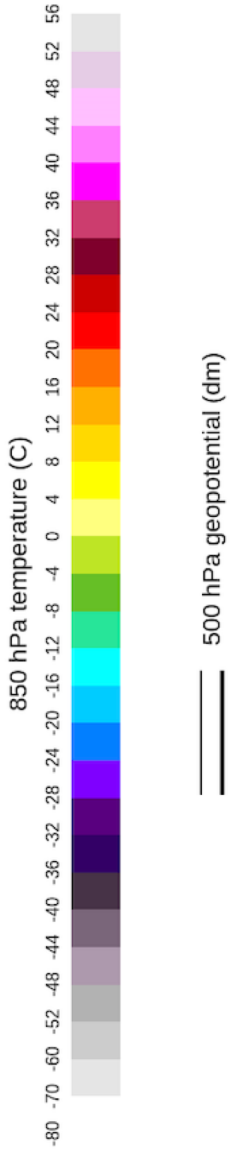


Base time: Mon 08 Jan 2024 00 UTC Valid time: Mon 08 Jan 2024 00 UTC (+0h) Area : Europe



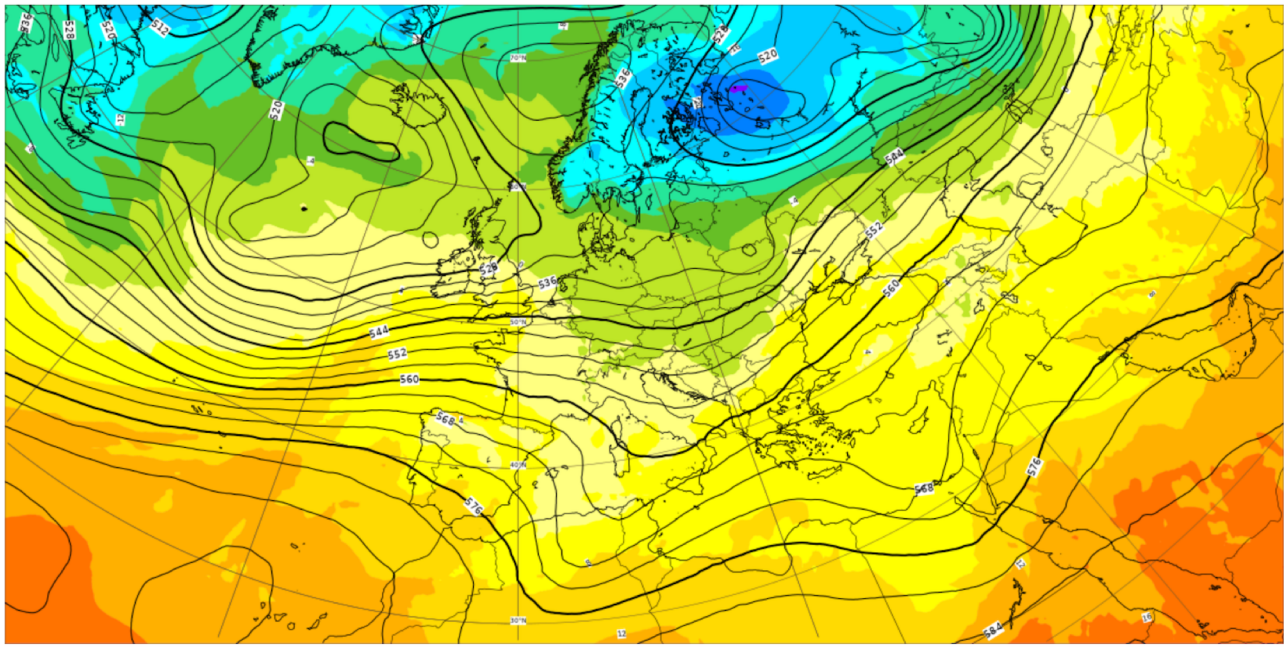
HRES

ECMWF

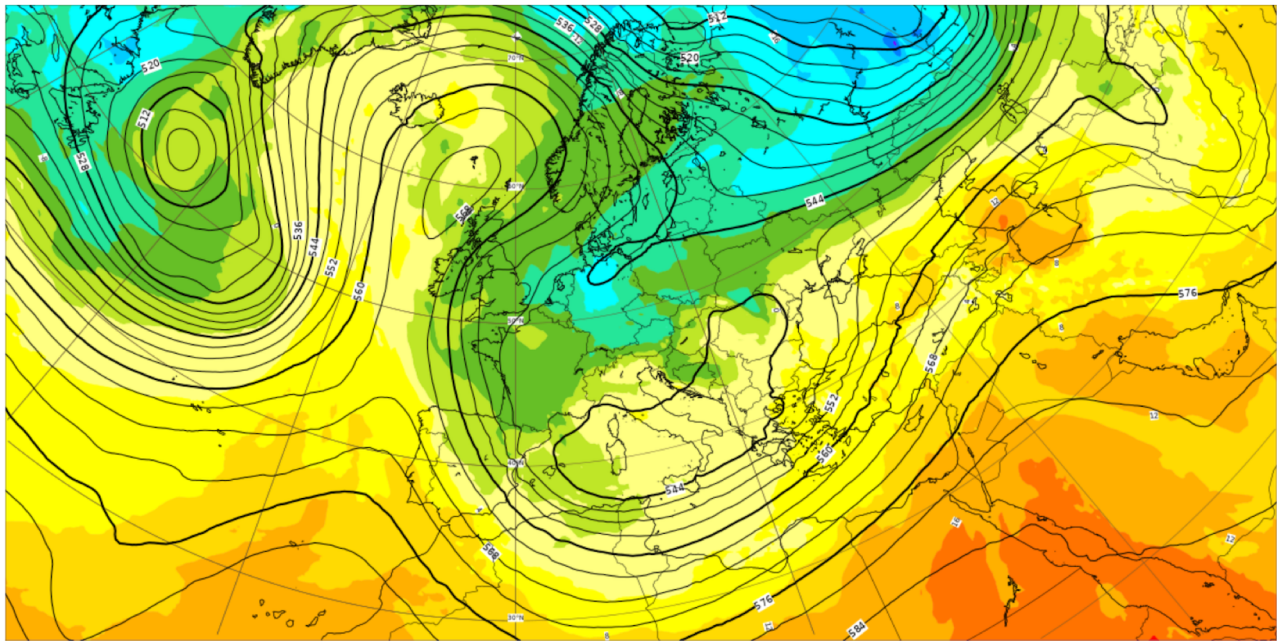


500 hPa geopotential (dm)

Base time: Tue 02 Jan 2024 00 UTC Valid time: Tue 02 Jan 2024 00 UTC (+0h) Area : Europe

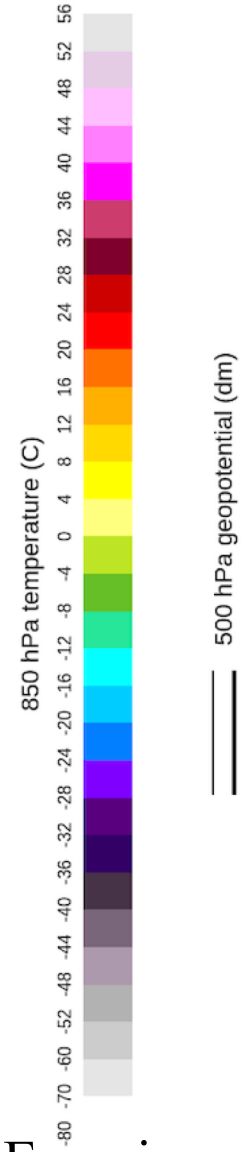


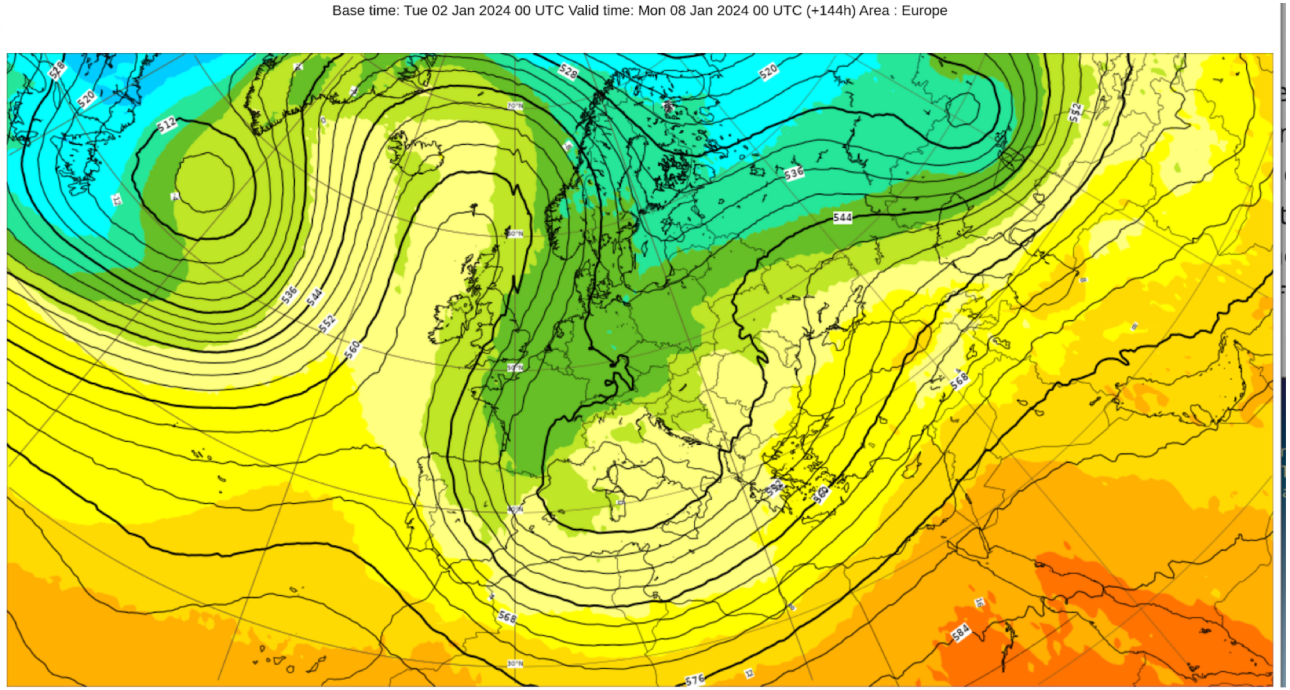
Base time: Mon 08 Jan 2024 00 UTC Valid time: Mon 08 Jan 2024 00 UTC (+0h) Area : Europe



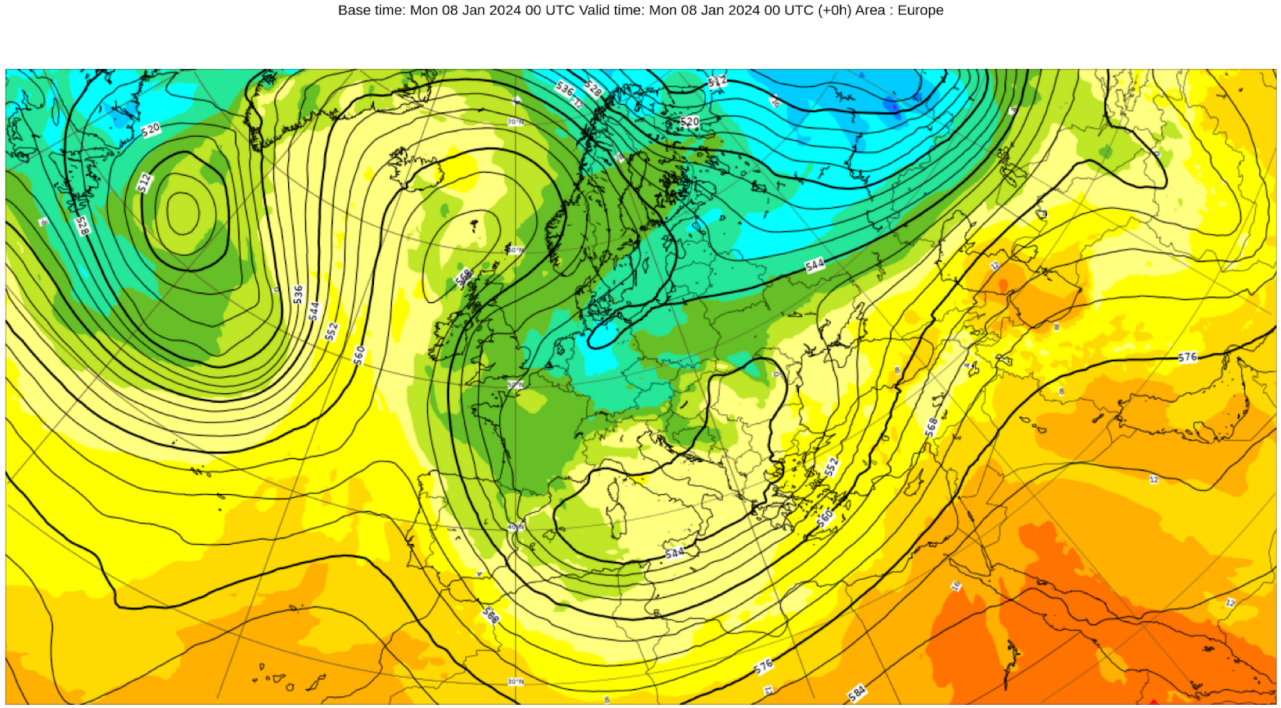
HRES

ECMWF

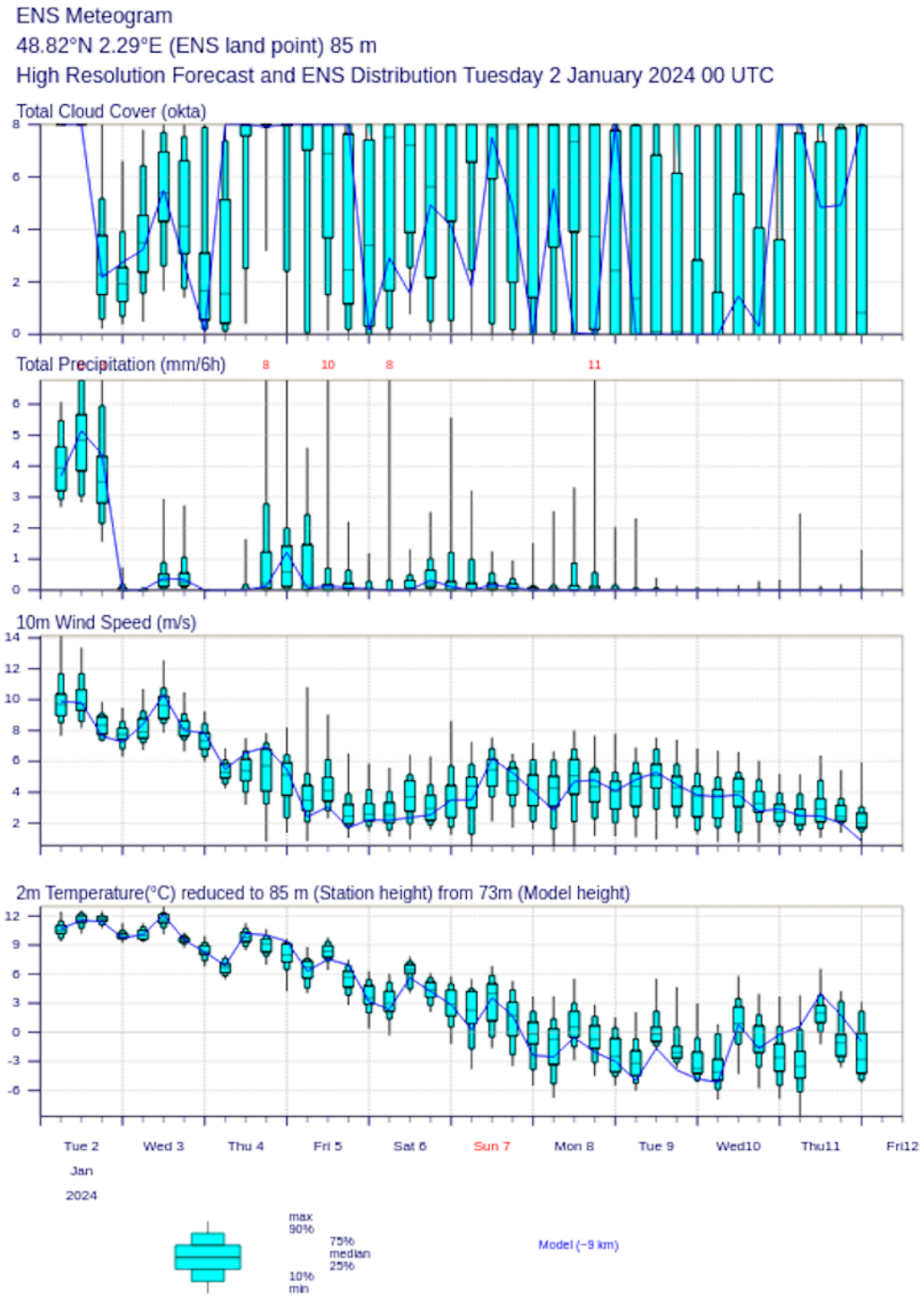




Experimental:
FuXi ML model

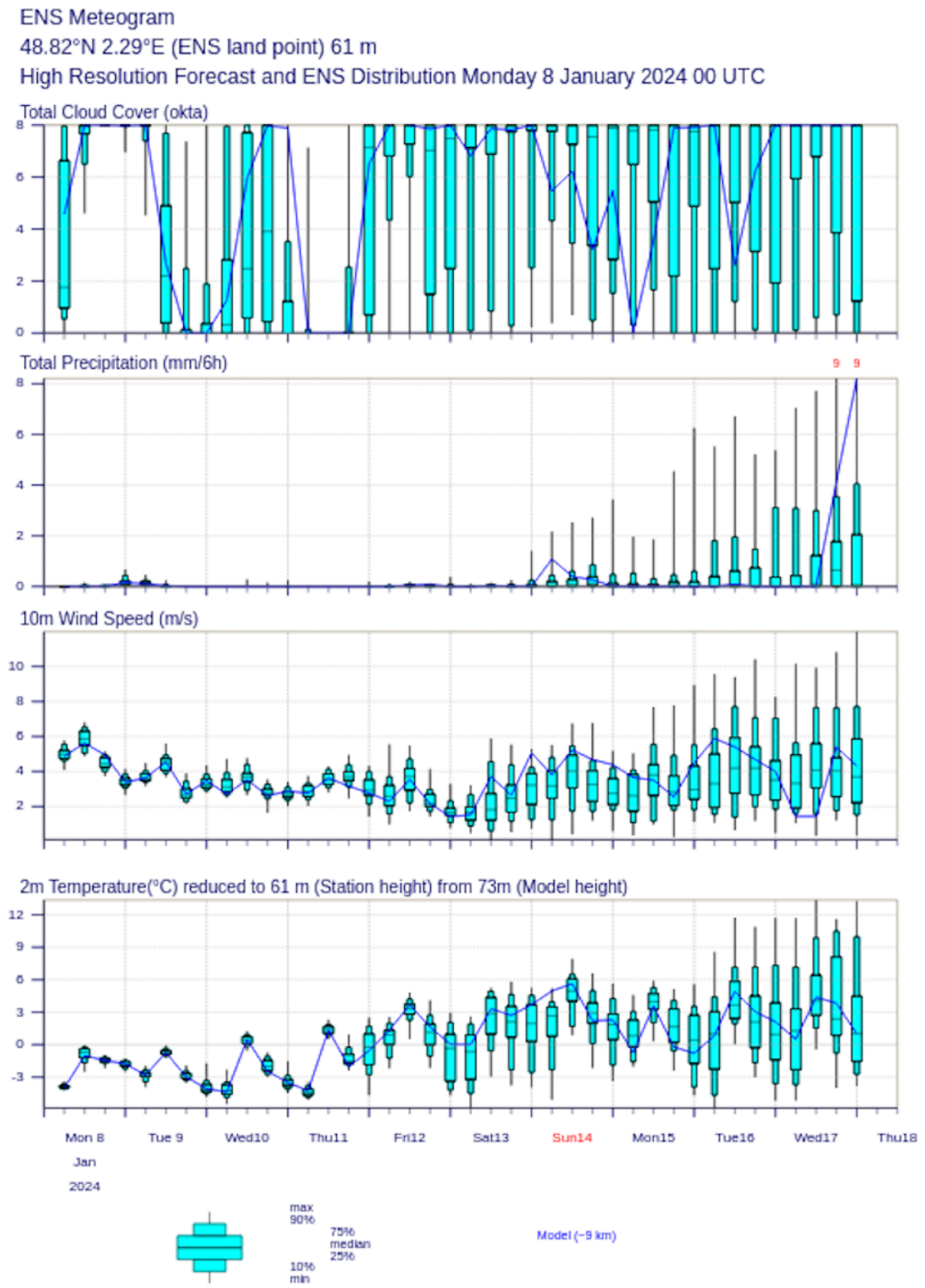


ECMWF


ENS Meteogram
48.82°N 2.29°E (ENS land point) 85 m
High Resolution Forecast and ENS Distribution Tuesday 2 January 2024 00 UTC
Total Cloud Cover (okta)
Total Precipitation (mm/6h)
8
10
8
11
10m Wind Speed (m/s)
2m Temperature(°C) reduced to 85 m (Station height) from 73m (Model height)
Tue 2
Jan
2024
Wed 3
Thu 4
Fri 5
Sat 6
Sun 7
Mon 8
Tue 9
Wed10
Thu11
Fri12
max
90%
75%
median
25%
10%
min
Model (~9 km)

ENS

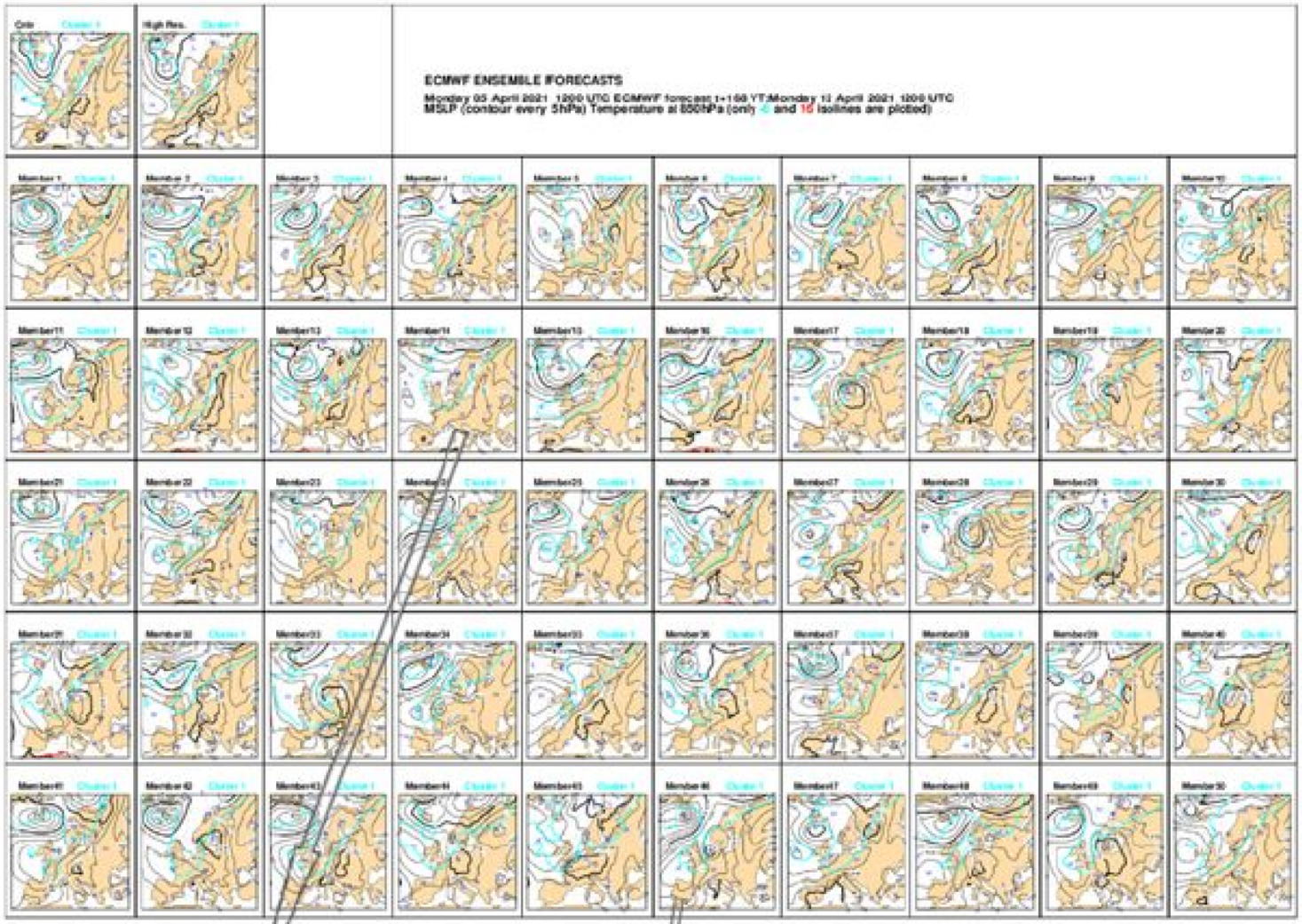
ECMWF

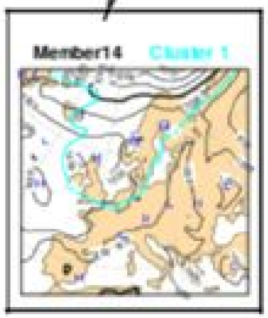

ENS Meteogram
48.82°N 2.29°E (ENS land point) 61 m
High Resolution Forecast and ENS Distribution Monday 8 January 2024 00 UTC
Total Cloud Cover (okta)
Total Precipitation (mm/6h)
10m Wind Speed (m/s)
2m Temperature(°C) reduced to 61 m (Station height) from 73m (Model height)
Mon 8 Jan 2024
Tue 9
Wed10
Thu11
Fri12
Sat13
Sun14
Mon15
Tue16
Wed17
Thu18
max
90%
75%
median
25%
10%
min
Model (~9 km)

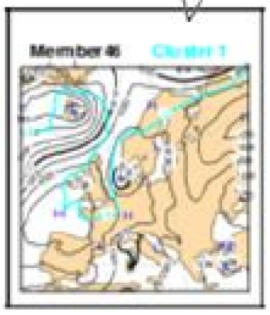
ENS

Fig. 1: Members of day 7 forecast of 500 hPa geopotential height for the ensemble originated from 25 January 1993.

Base time: Mon 05 Apr 2021 12 UTC, Valid time: Mon 12 Apr 2021 12 UTC, - T+168 h, Step : 168







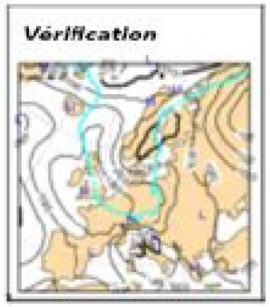


Figure 6 Hurricane Katrina mean-sea-level-pressure (MSLP) analysis for 12 UTC of 29 August 2005 and t+84h high-resolution and EPS forecasts started at 00 UTC of 26 August:

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.

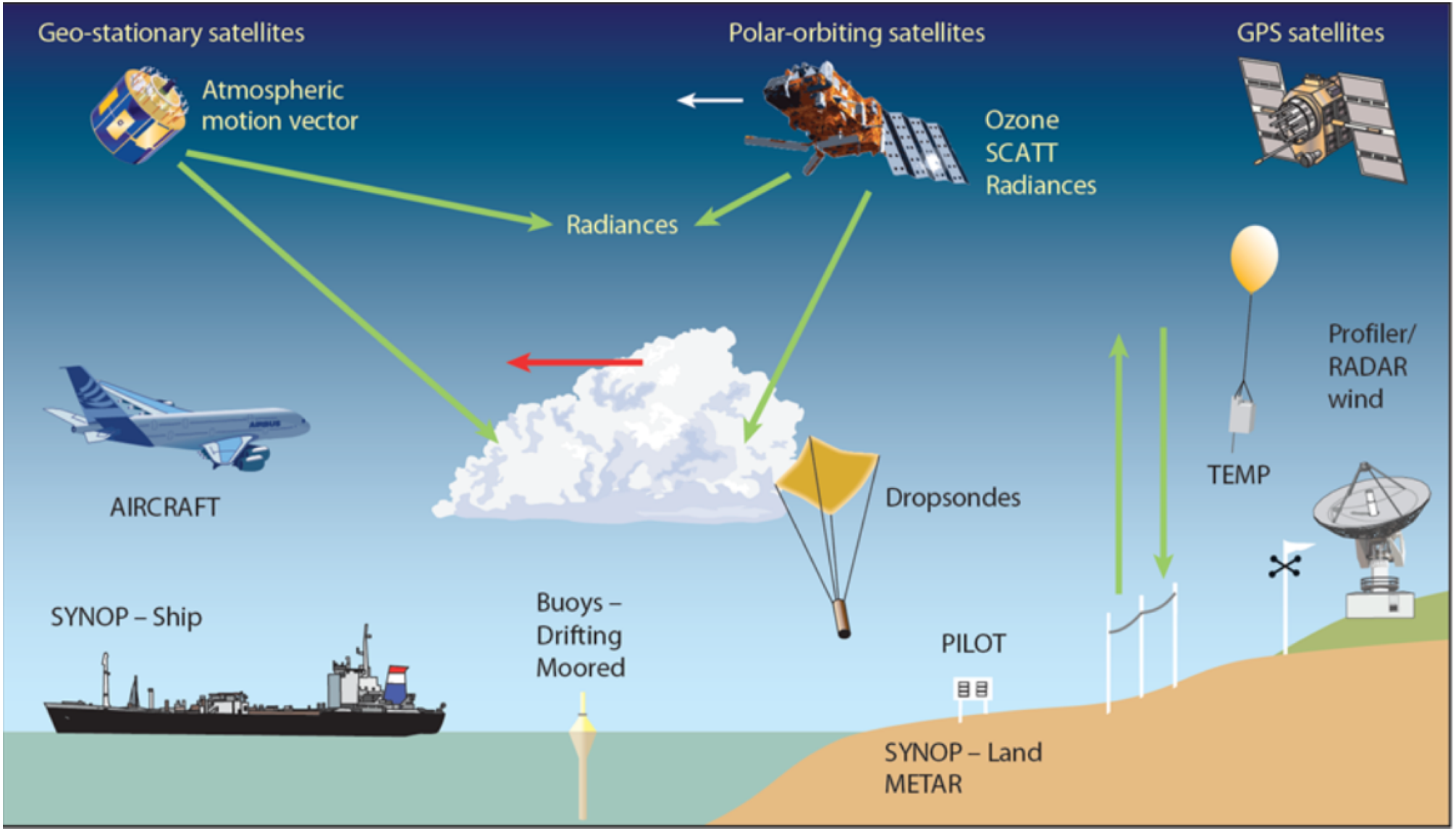
Pourquoi les météorologistes ont-ils tant de peine à prédire le temps avec quelque certitude ? Pourquoi les chutes de pluie, les tempêtes elles-mêmes nous semblent-elles arriver au hasard, de sorte que bien des gens trouvent tout naturel de prier pour avoir la pluie ou le beau temps, alors qu'ils jugeraient ridicule de demander une éclipse par une prière ? Nous voyons que les grandes perturbations se produisent généralement dans les régions où l'atmosphère est en équilibre instable. Les météorologistes voient bien que cet équilibre est instable, qu'un cyclone va naître quelque part ; mais où, ils sont hors d'état de le dire ; 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

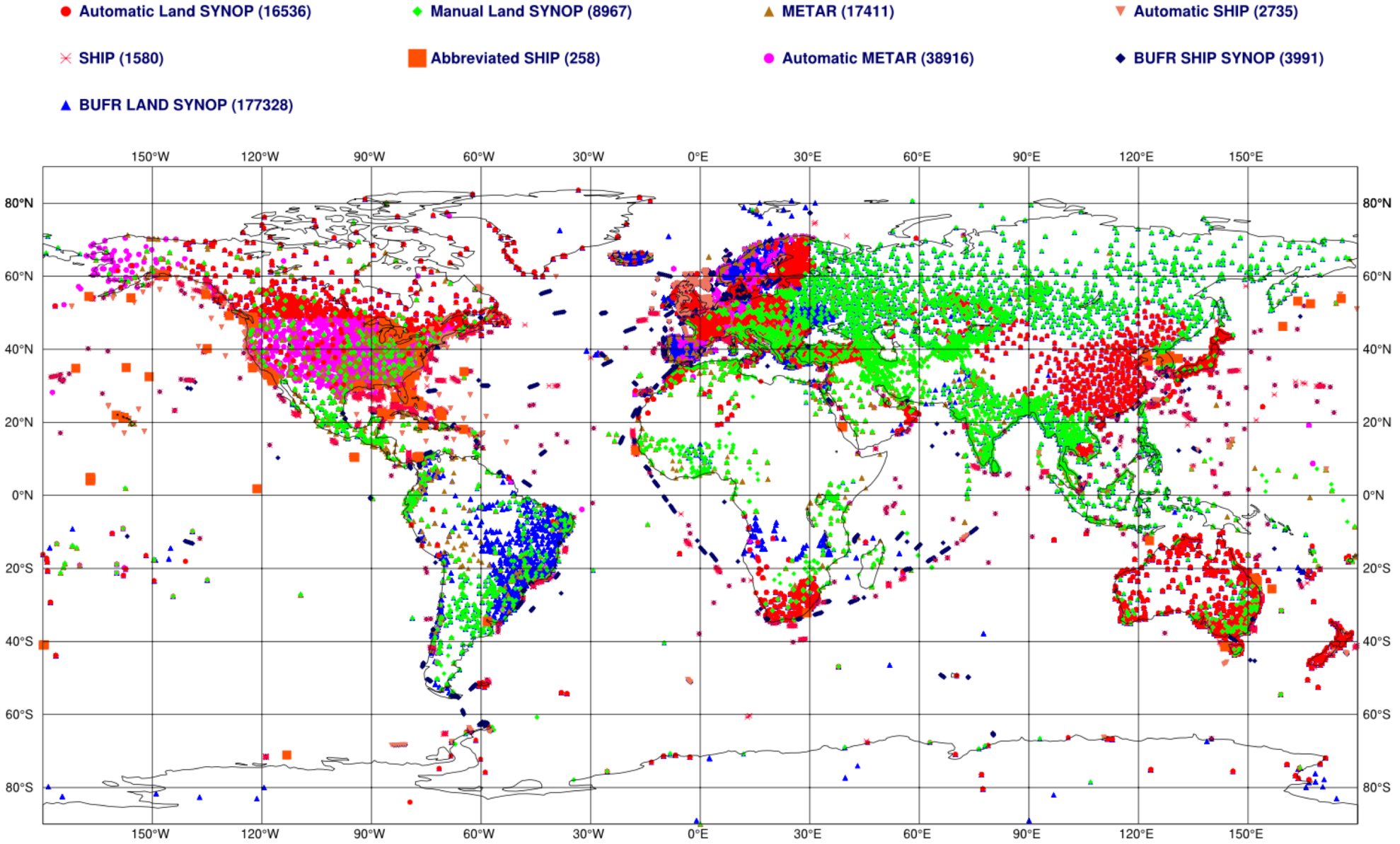
Why have meteorologists such difficulty in predicting the weather with any certainty? Why is it that showers and even storms seem to come by chance, so that many people think it quite natural to pray for rain or fine weather, though they would consider it ridiculous to ask for an eclipse by prayer? We see that great disturbances are generally produced in regions where the atmosphere is in unstable equilibrium. The meteorologists see very well that the equilibrium is unstable, that a cyclone will be formed somewhere, but exactly where they are not in a position to say; 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 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

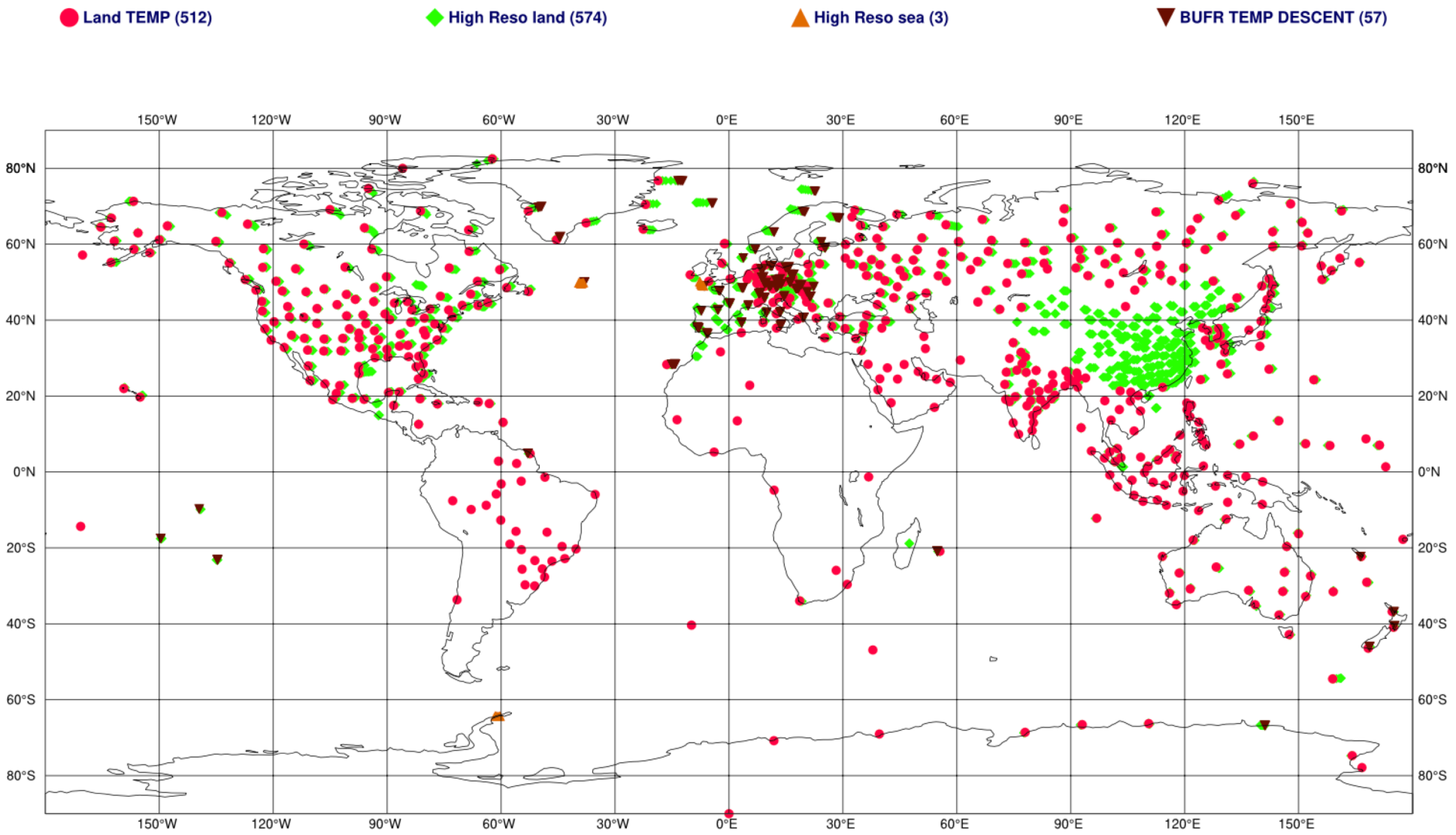
(English transl. by F. Maitland, *Science and Method*,
T. Nelson and Sons, London, 1914)


Geo-stationary satellites
Atmospheric motion vector
Polar-orbiting satellites
Ozone SCATT Radiances
GPS satellites
Radiances
Profiler/ RADAR wind
TEMP
AIRCRAFT
Dropsondes
SYNOP – Ship
Buoys – Drifting Moored
PILOT
SYNOP – Land METAR

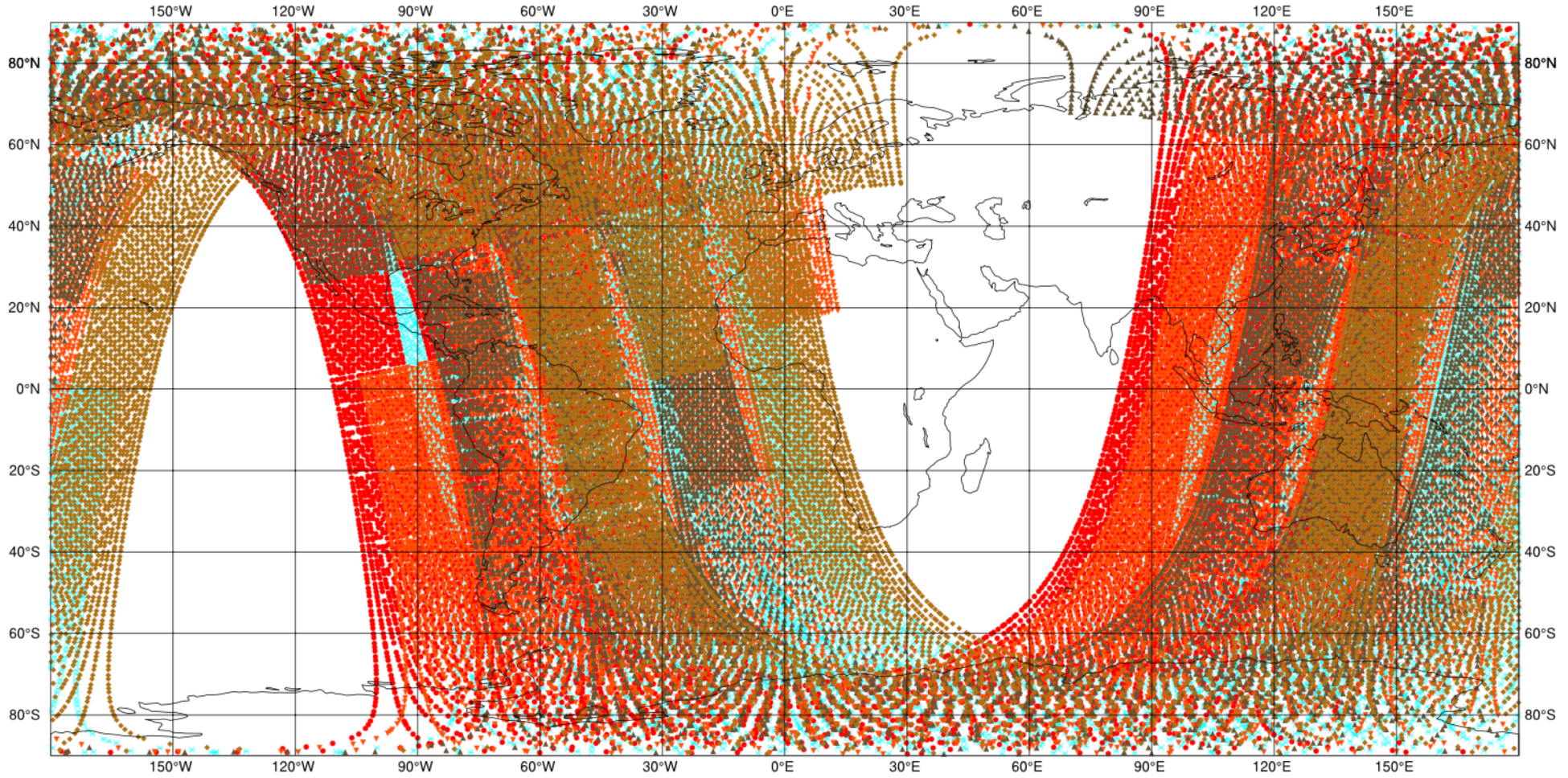
ECMWF data coverage (all observations) - SYNOP-SHIP-METAR
2024010521 to 2024010603
Total number of obs = 267722



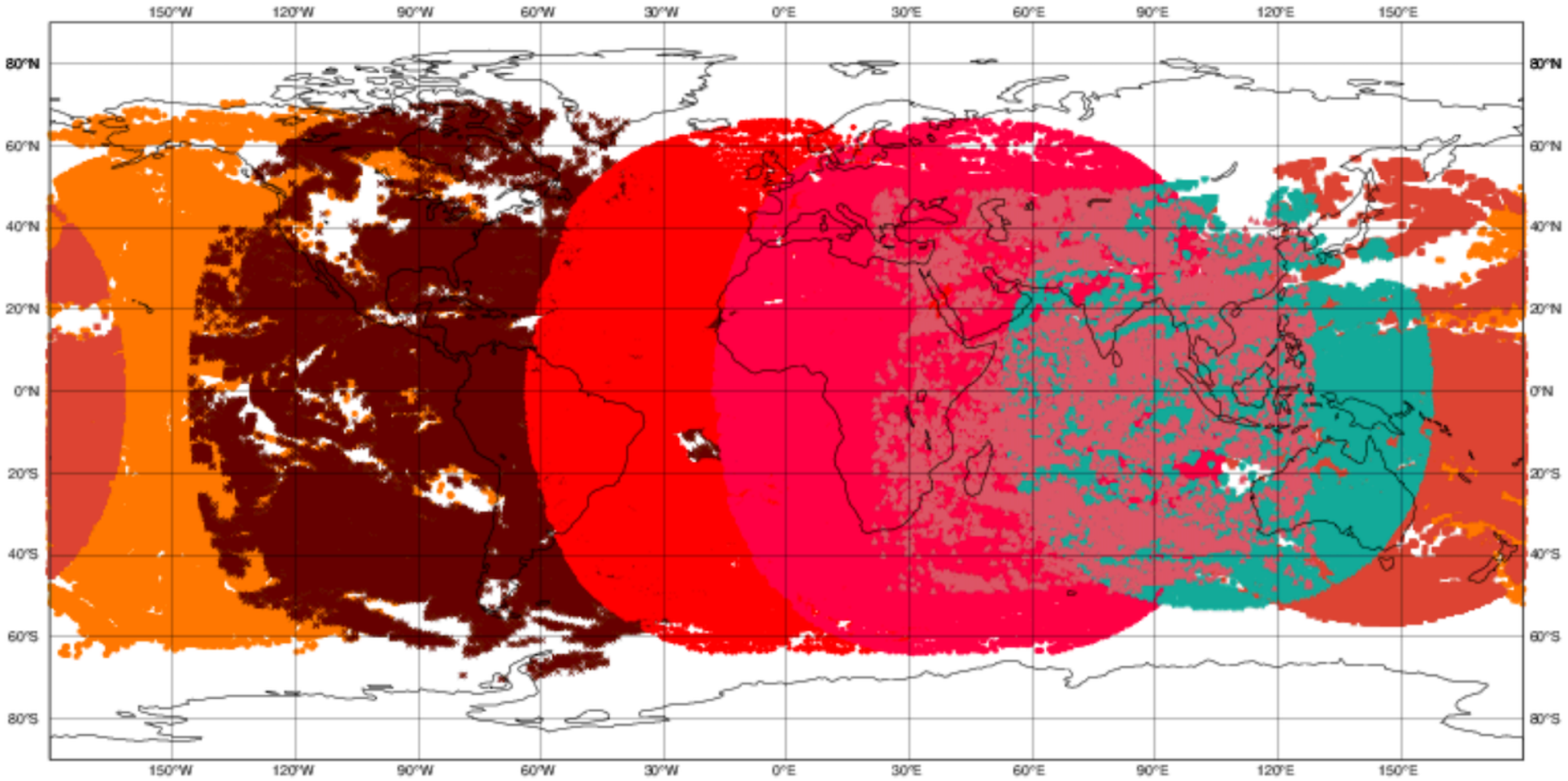
ECMWF data coverage (all observations) - RADIOSONDE
2024010521 to 2024010603
Total number of obs = 1146



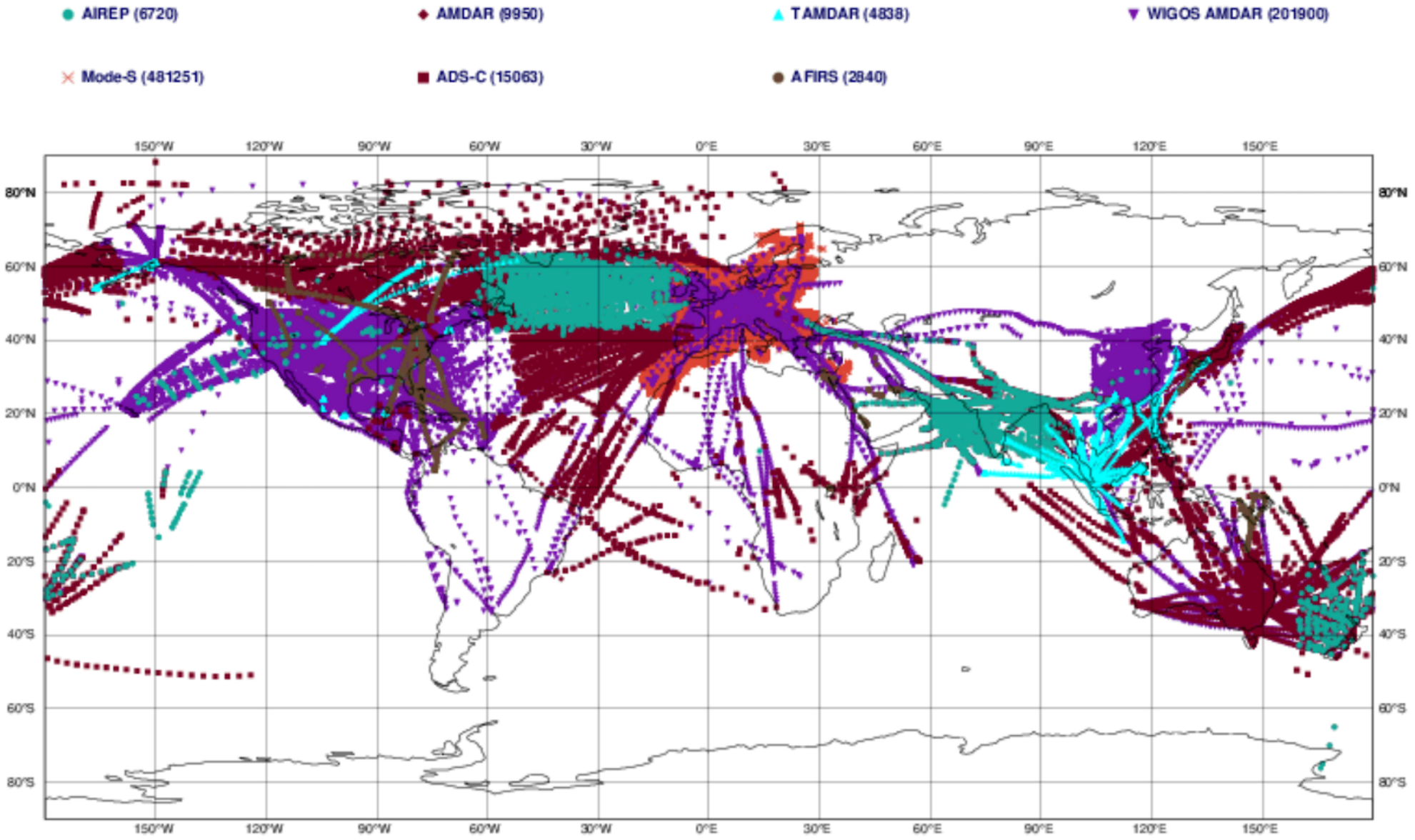
ECMWF data coverage (all observations) - AMSUA
2024010521 to 2024010603
Total number of obs = 83564



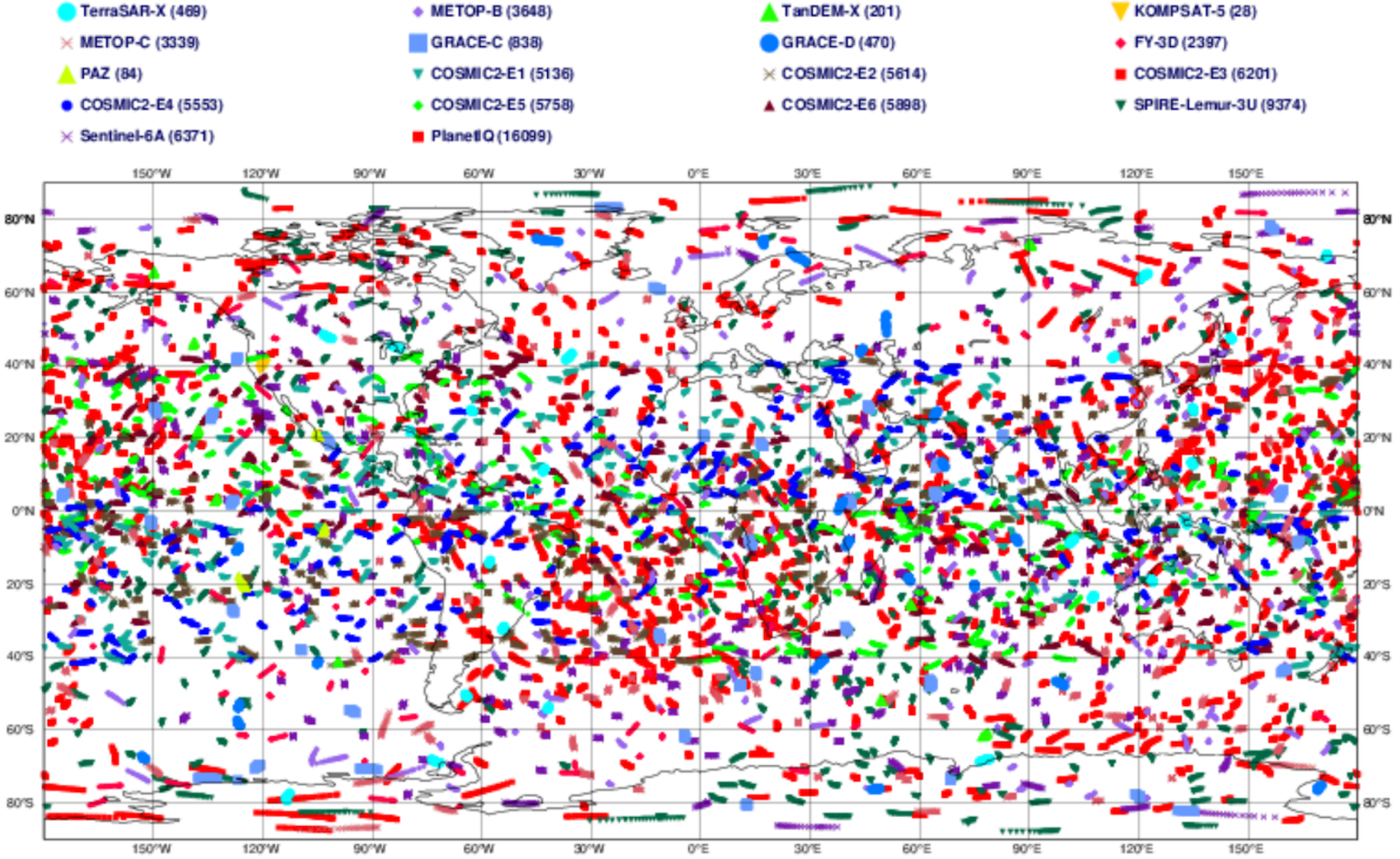
ECMWF data coverage (all observations) - AMV WV
2024010521 to 2024010603
Total number of obs = 989690



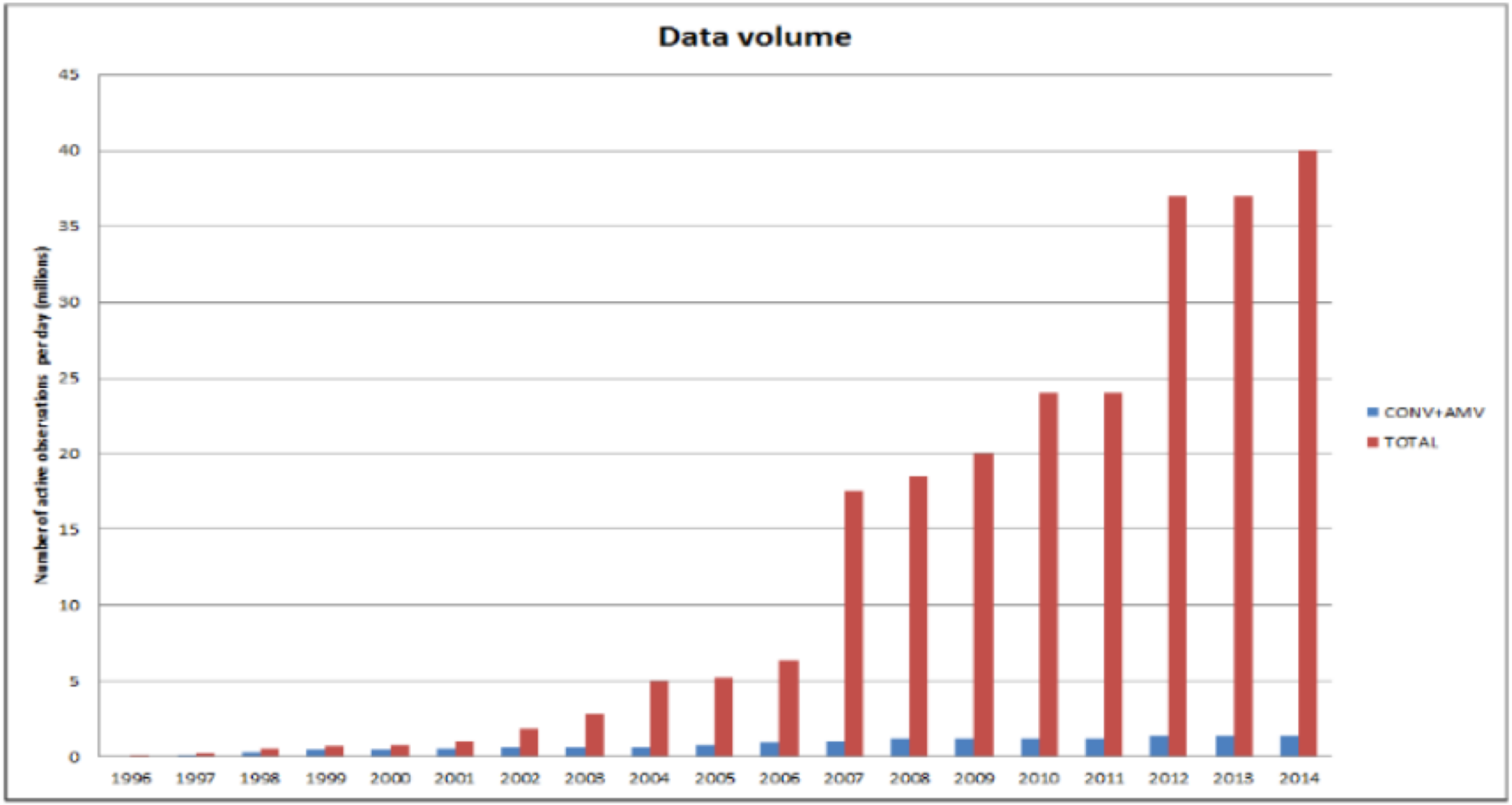
ECMWF data coverage (all observations) - AIRCRAFT
2024010521 to 2024010603
Total number of obs = 722562



ECMWF data coverage (all observations) - GPSRO
2024010521 to 2024010603
Total number of obs = 77478



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

As of 2023

We receive 800 million observations daily, and 60 million quality-controlled observations are available daily for use in the Integrated Forecasting System (IFS); the vast majority of these are satellite measurements, but ECMWF also benefits from all available observations from non-satellite sources, including surface-based and aircraft reports.

- *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), and are in practice concentrated over continents.

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

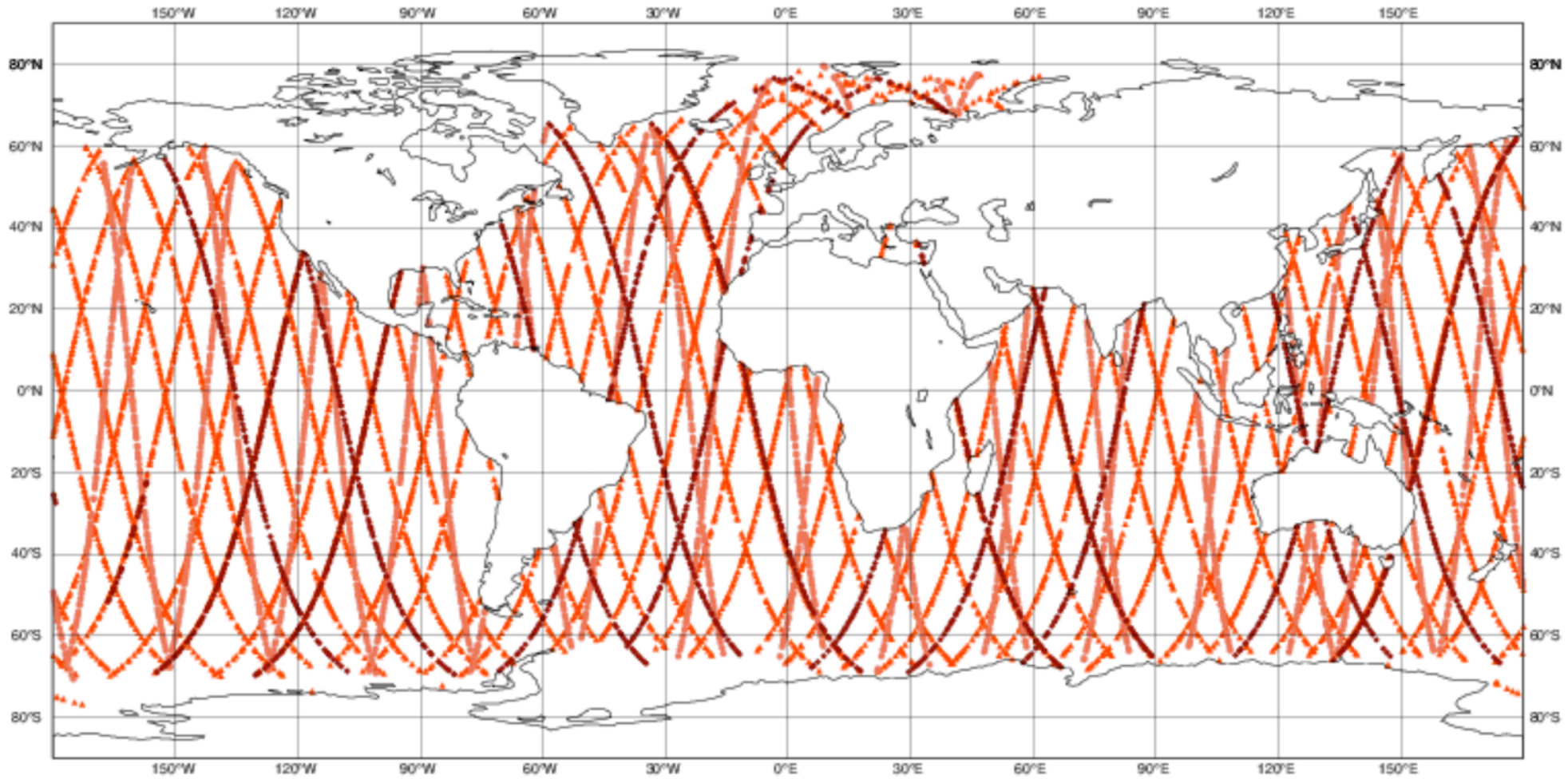
- *Direct* observations (temperature, pressure, horizontal components of the wind, moisture), which are local and bear on the variables used for describing the flow in numerical models.

- *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
20240105 00
Total number of obs = 4914



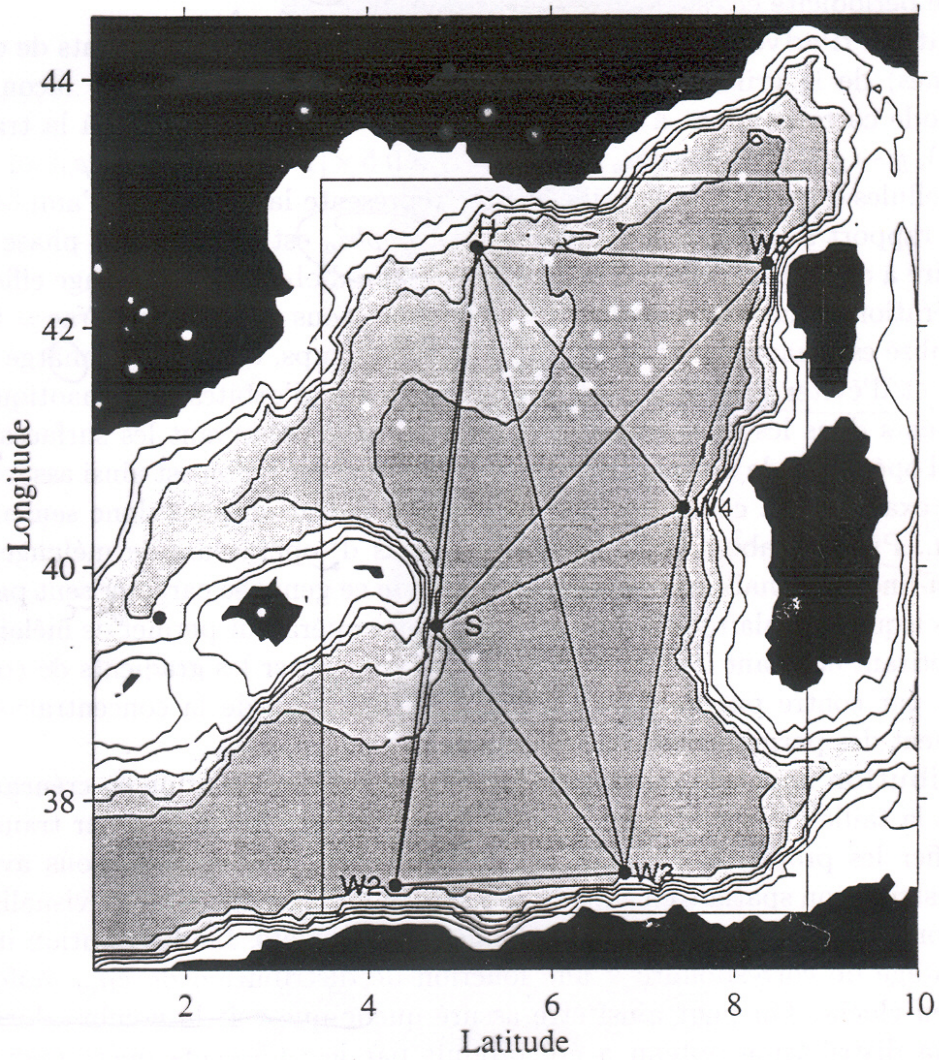


FIG. 1 – *Bassin méditerranéen occidental: réseau d'observation tomographique de l'expérience Thétis 2 et limites du domaine spatial utilisé pour les expériences numériques d'assimilation.*

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

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

- 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 $\Rightarrow$ *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 developed over time. They require appropriate definition of lateral boundary conditions (not a simple problem). Most of them are non-hydrostatic, and therefore allow description of convection.

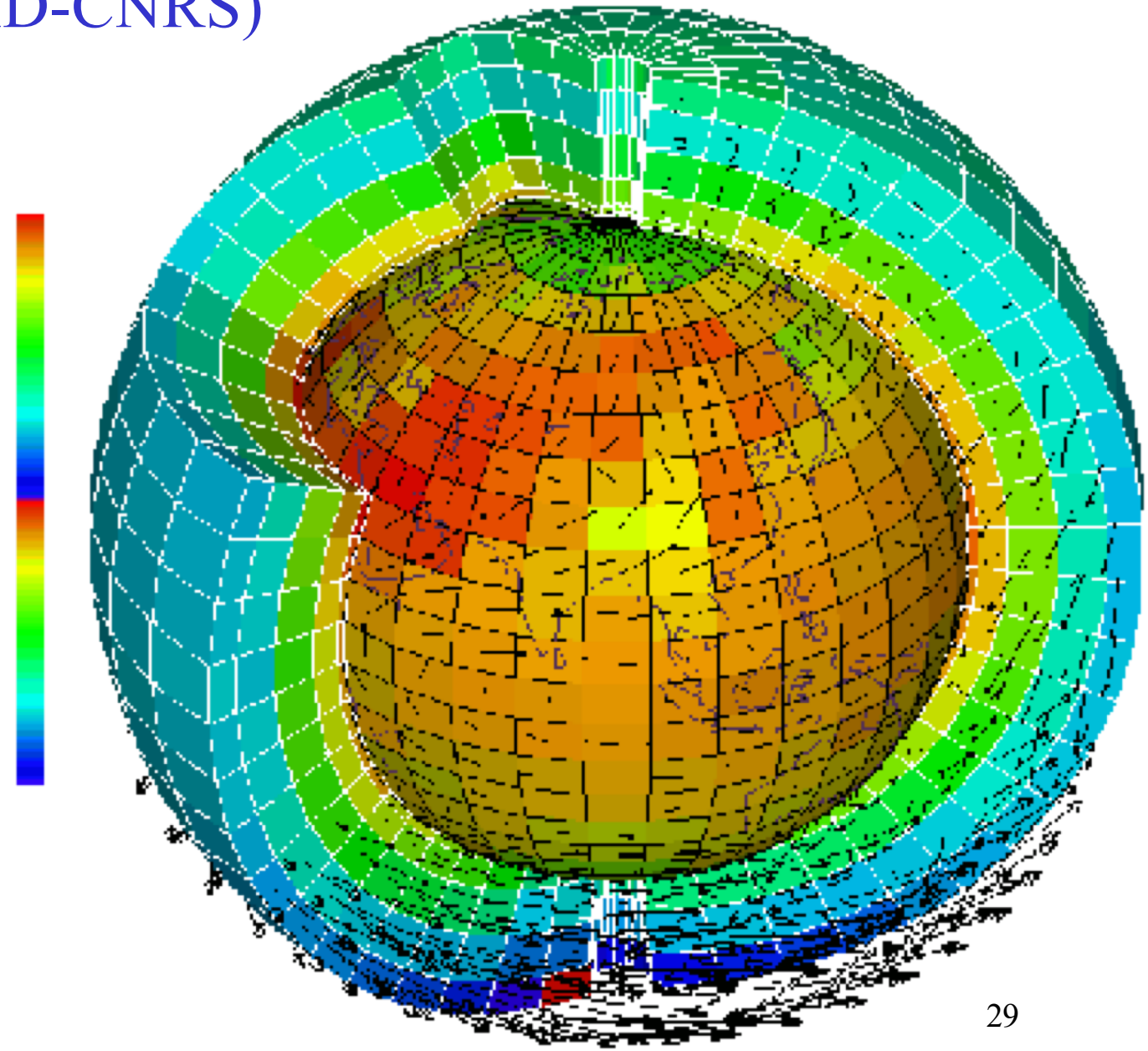
There exist at present two forms of horizontal spatial discretization

- 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 account of the complicated geometry of coast-lines.

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



In gridpoint models, meteorological fields are defined by values at the nodes of 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).

Modèles (semi-)spectraux

$$T(\mu=\sin(\text{latitude}),\ \lambda=\text{longitude}) = \sum_{\substack{0\le n<\infty \\ -n\le m\le n}} T_n^m Y_n^m(\mu,\lambda)$$

où les $Y_n^m(\mu,\lambda)$ sont les *harmoniques sphériques*

$$Y_n^m(\mu,\lambda) \propto P_n^m(\mu)\exp(im\lambda)$$

$P_n^m(\mu)$ est la *fonction de Legendre* de deuxième espèce

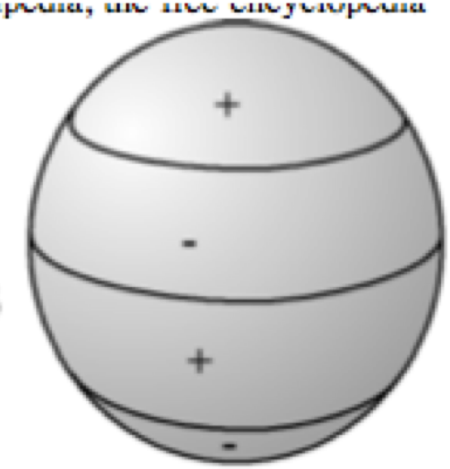
$$P_n^m(\mu) \propto (1-\mu^2)^{\frac{m}{2}} \frac{d^{n+m}}{d\mu^{n+m}}(\mu^2-1)^n$$

n et m sont respectivement le *degré* et l'*ordre* de l'harmonique $Y_n^m(\mu,\lambda)$

$n = 0, 1, \ldots$ $\quad - n \le m \le n$

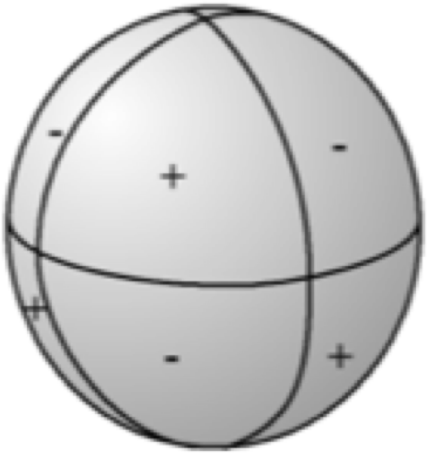
From Wikipedia, the free encyclopedia

$l = 3$
$m = 0$
$l\text{-}m = 3$

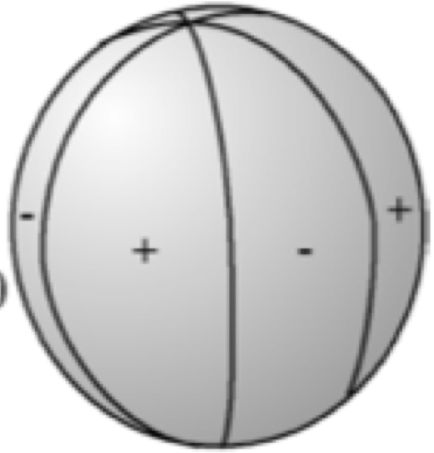


$l = 3$
$m = 1$
$l\text{-}m = 2$

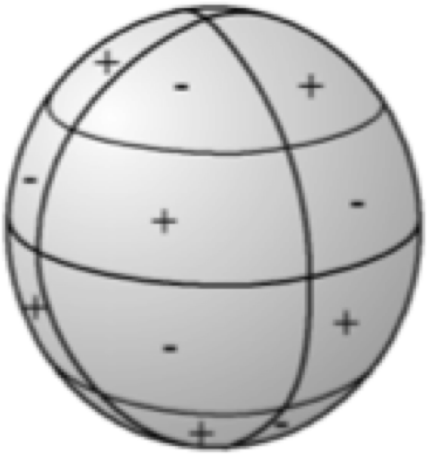
$l = 3$
$m = 2$
$l\text{-}m = 1$



$l = 3$
$m = 3$
$l\text{-}m = 0$



$l = 5$
$m = 2$
$l\text{-}m = 3$



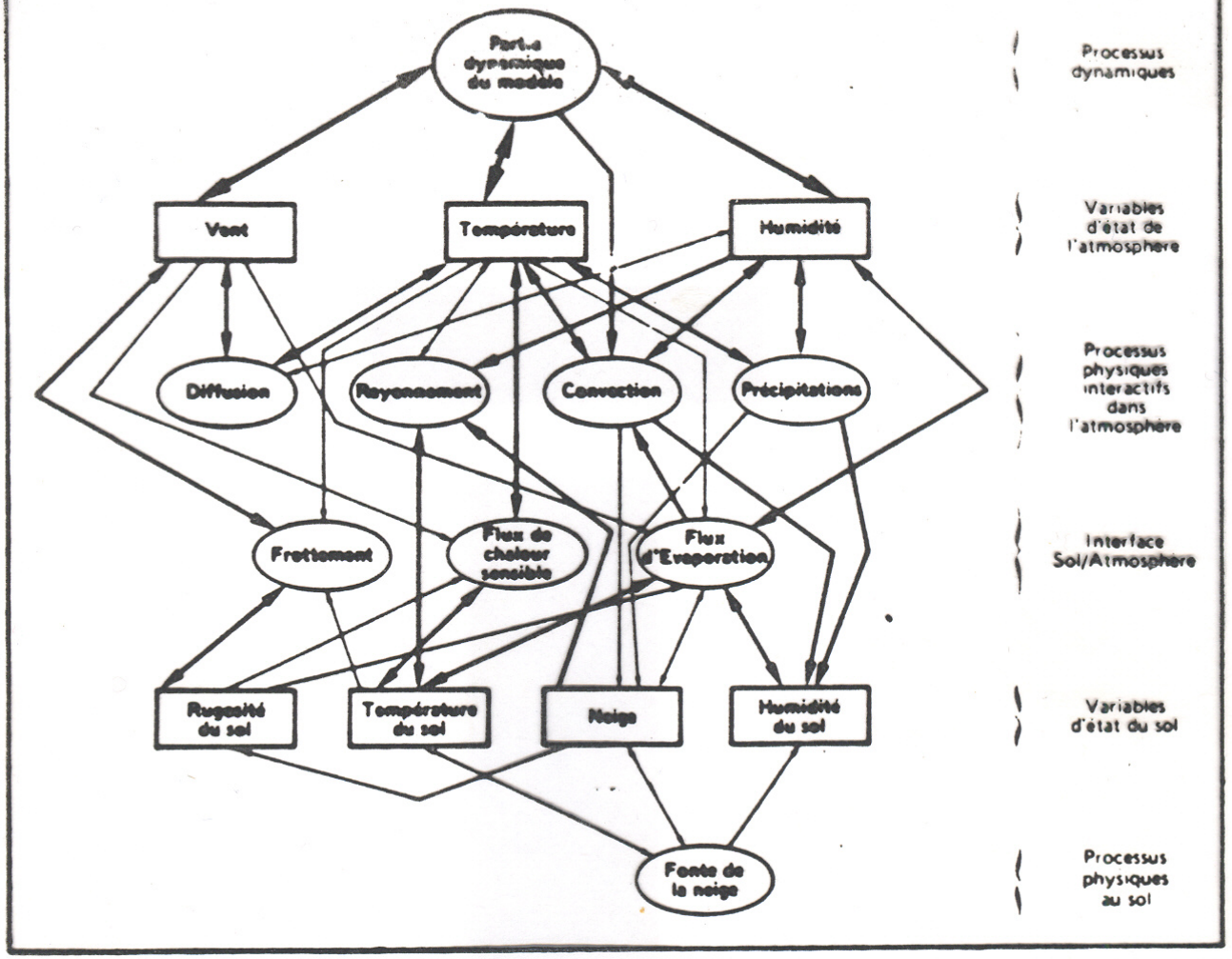
Linear operations, and in particular differentiation with respect to spatial variables, are performed in spectral space, while nonlinear operations and 'physical' computations (advection by the motion, diabatic heating and cooling, …) are performed in gridpoint physical space. This requires constant 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*.

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).

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, GB, Bologna, Italy, Bonn, Germany)

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

June 2023 High-resolution (HRES) model

Triangular semi-spectral truncation TCO1279 / O1280
(horizontal resolution ≈ 9 kilometres)

Hydrostatic primitive equations. 137 vertical levels (0 - 80 km)

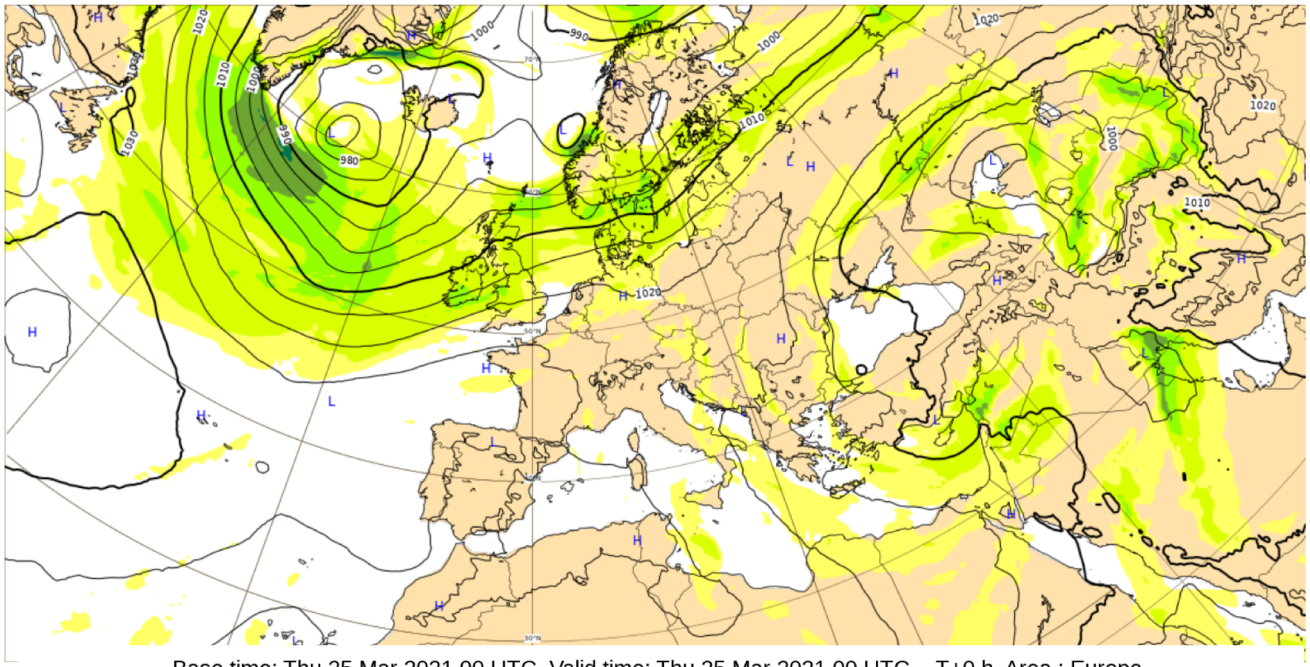
Finite-element vertical discretisation (hybrid coordinate)

Dimension of corresponding state vector > 10^9

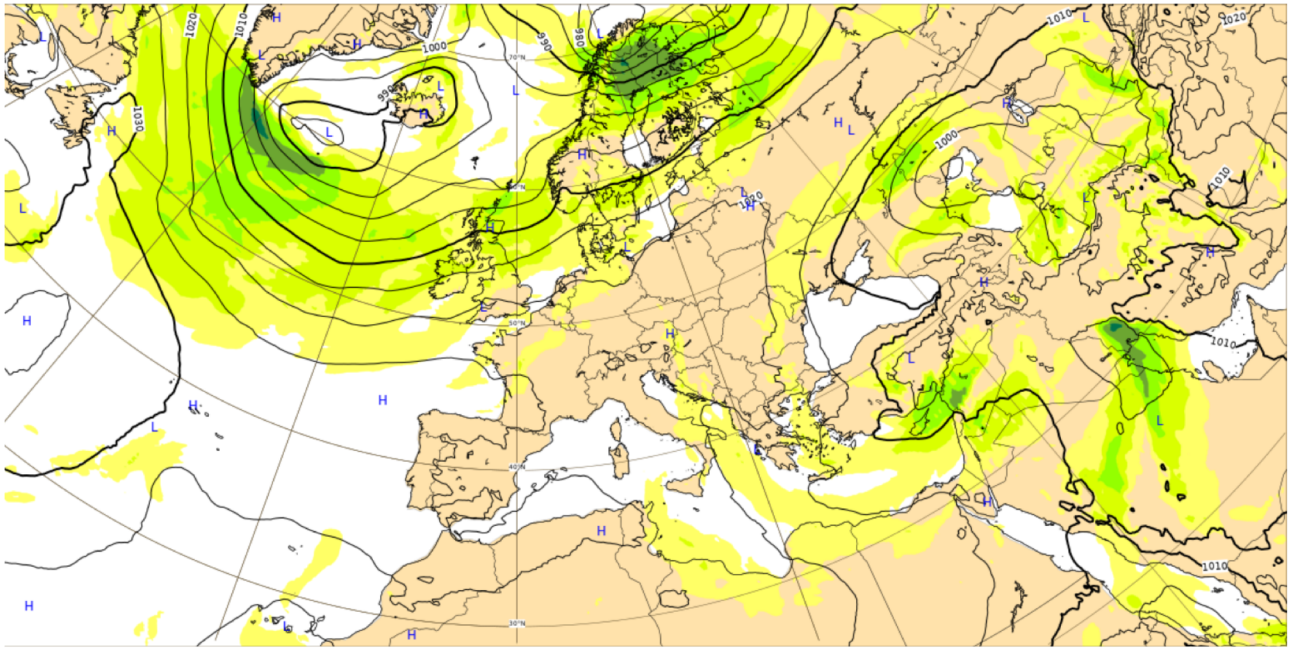
Integration timestep (semi-Lagrangian semi-implicit scheme): 450 seconds

Integrated four times a day (from 00, 06, 12 and 18 UTC) to 10-day range

Base time: Sat 20 Mar 2021 00 UTC, Valid time: Thu 25 Mar 2021 00 UTC, - T+120 h, Area : Europe



Base time: Thu 25 Mar 2021 00 UTC, Valid time: Thu 25 Mar 2021 00 UTC, - T+0 h, Area : Europe



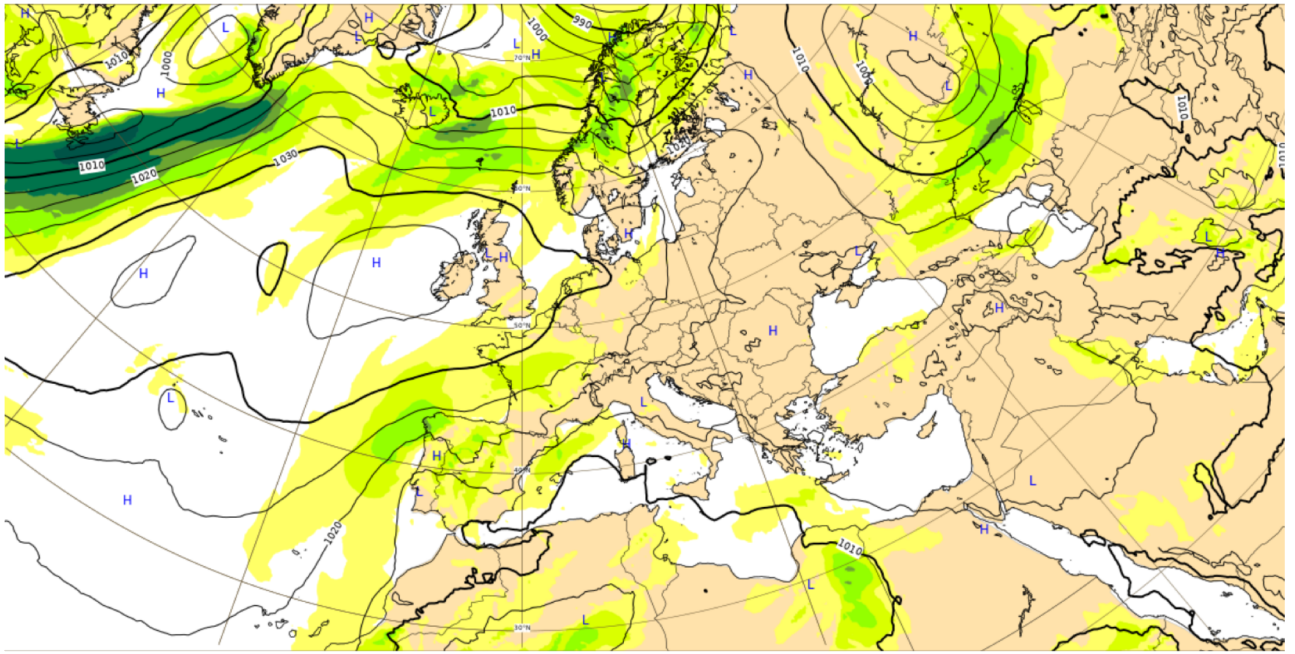
850 hPa wind speed (ms**-1)

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

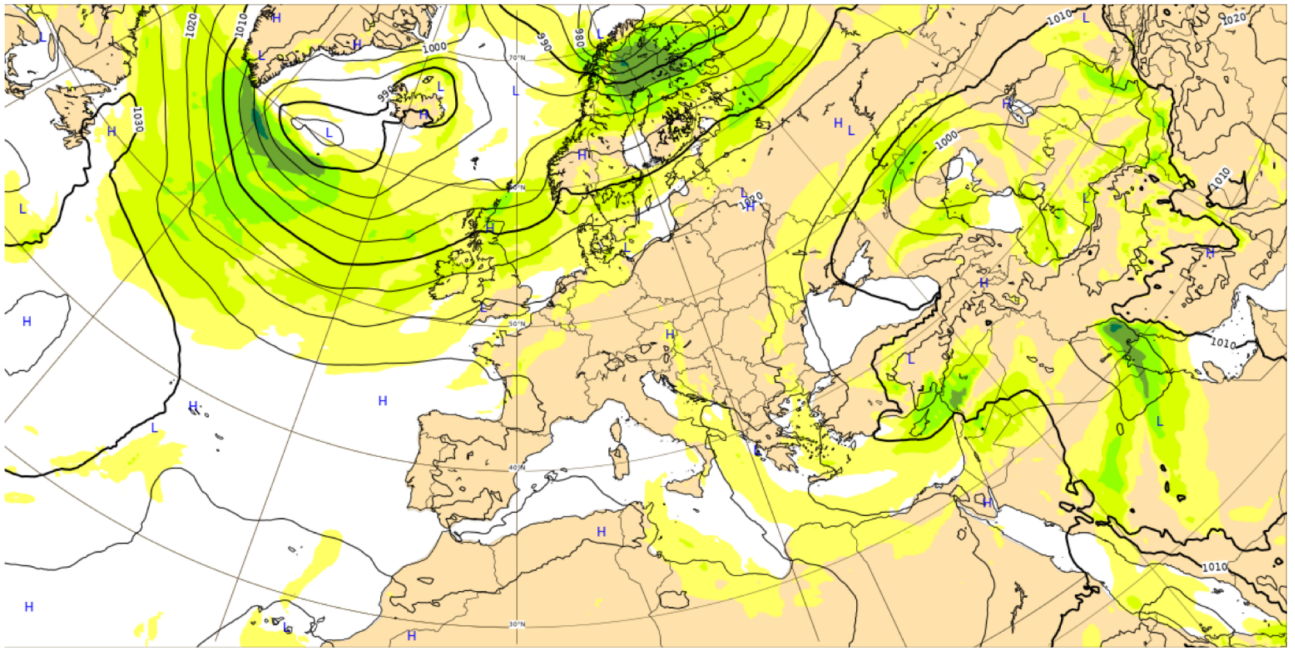
Mean sea level pressure (hPa)

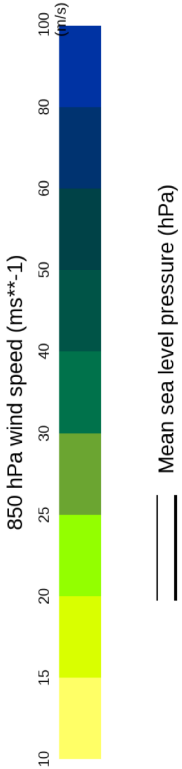
2021

Base time: Sat 20 Mar 2021 00 UTC, Valid time: Sat 20 Mar 2021 00 UTC, - T+0 h, Area : Europe



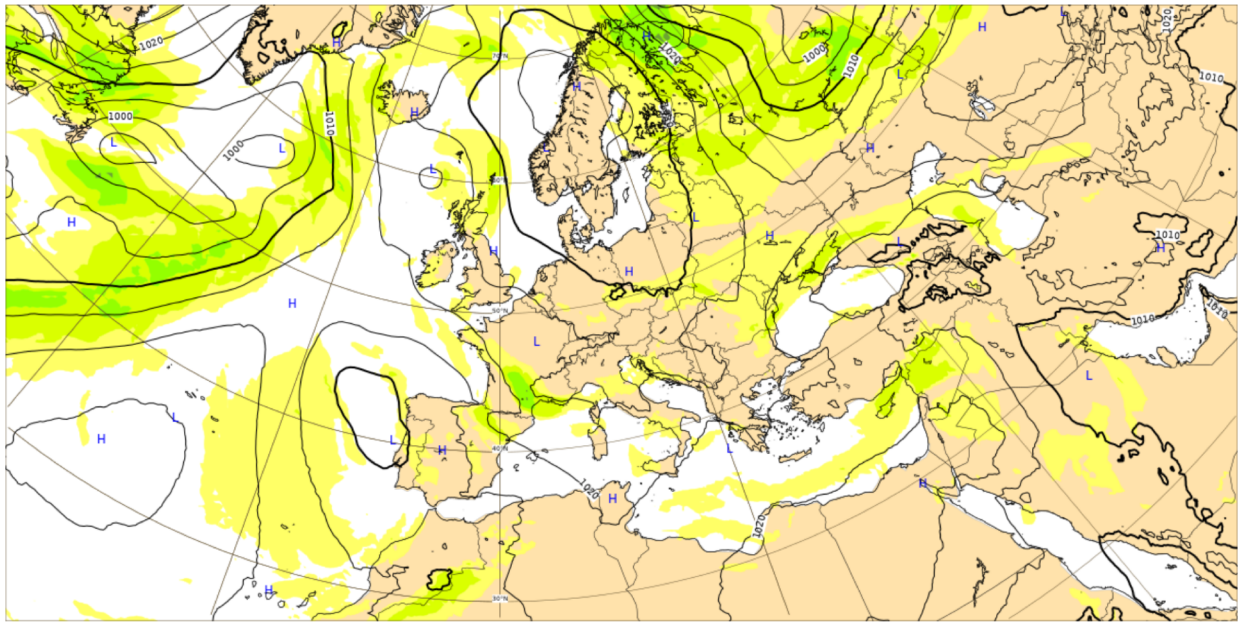
Base time: Thu 25 Mar 2021 00 UTC, Valid time: Thu 25 Mar 2021 00 UTC, - T+0 h, Area : Europe



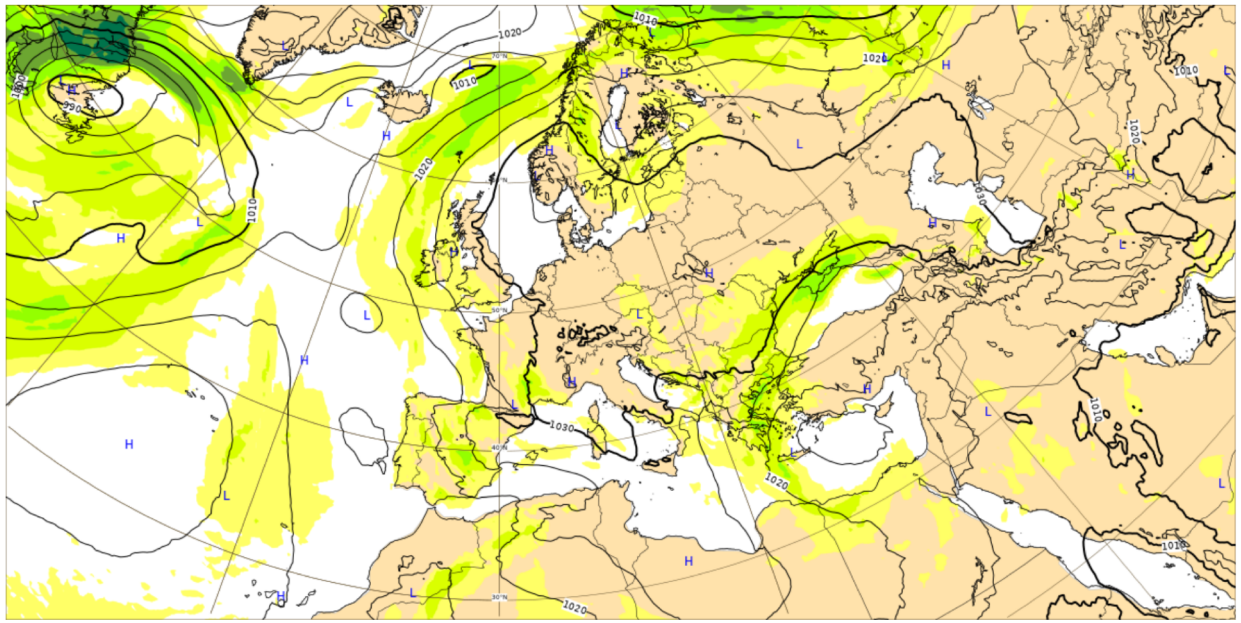


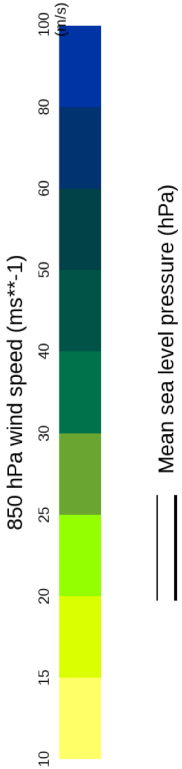
2021

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

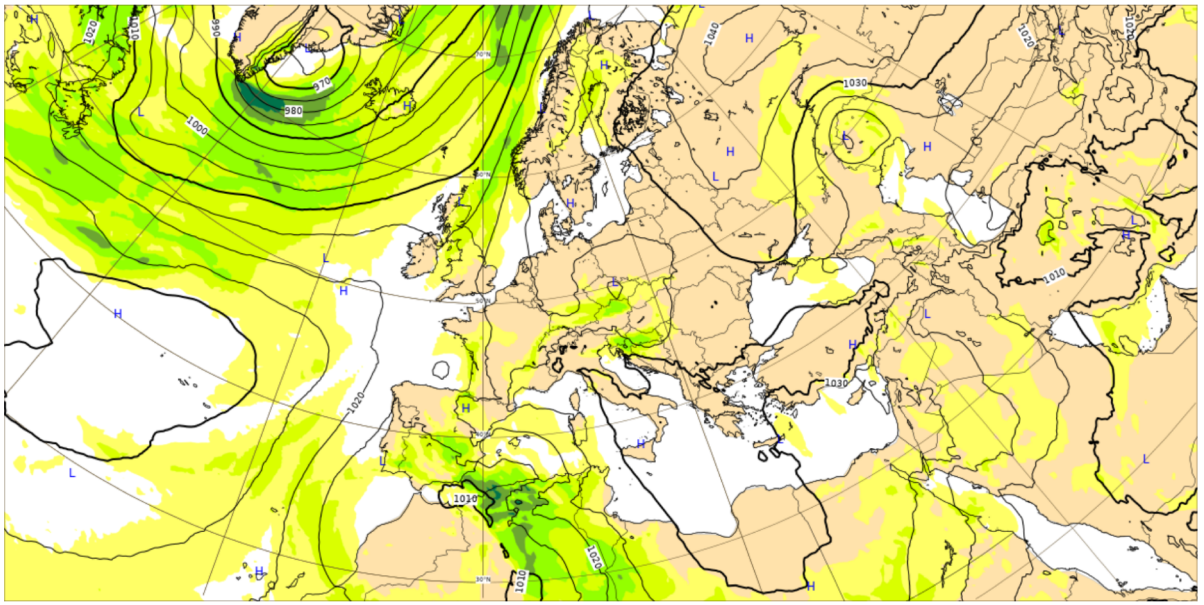


Base time: Wed 23 Mar 2022 00 UTC Valid time: Wed 23 Mar 2022 00 UTC (+0h) Area : Europe

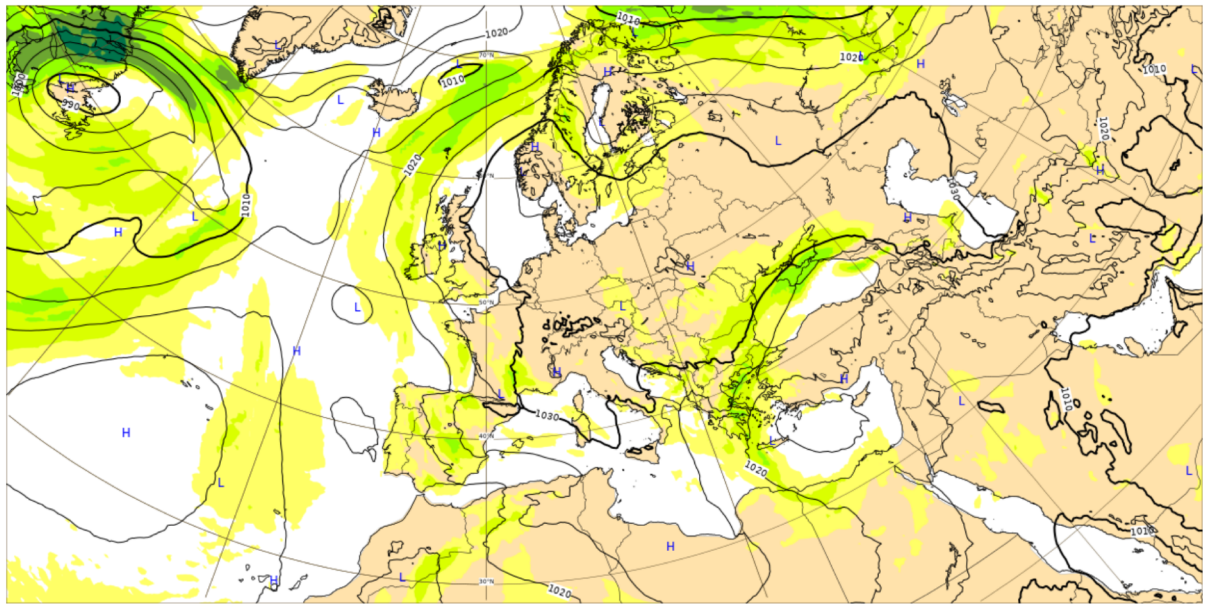




Base time: Wed 16 Mar 2022 00 UTC Valid time: Wed 16 Mar 2022 00 UTC (+0h) Area : Europe



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)

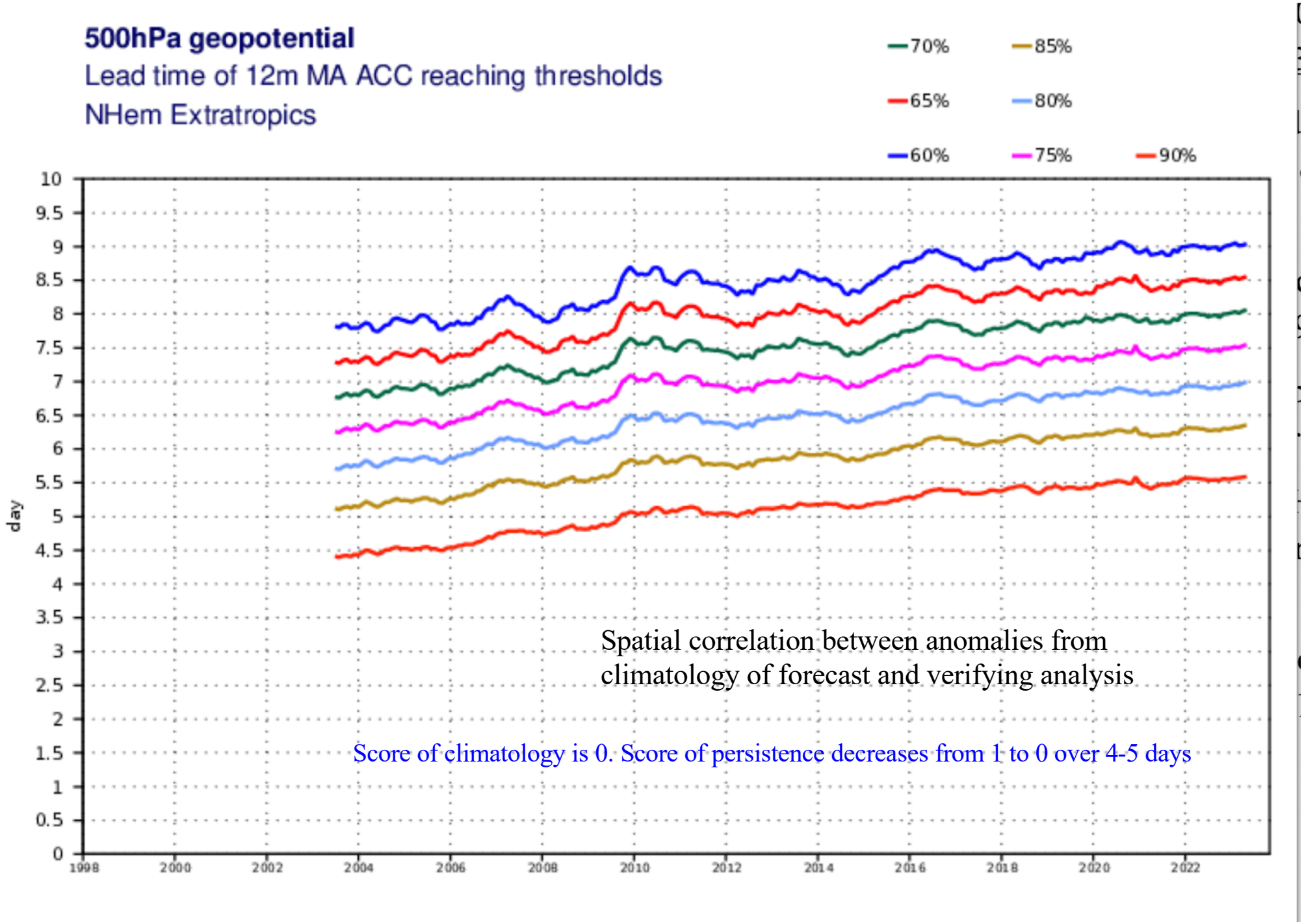
Results on site of ECMWF *www.ecmwf.int*

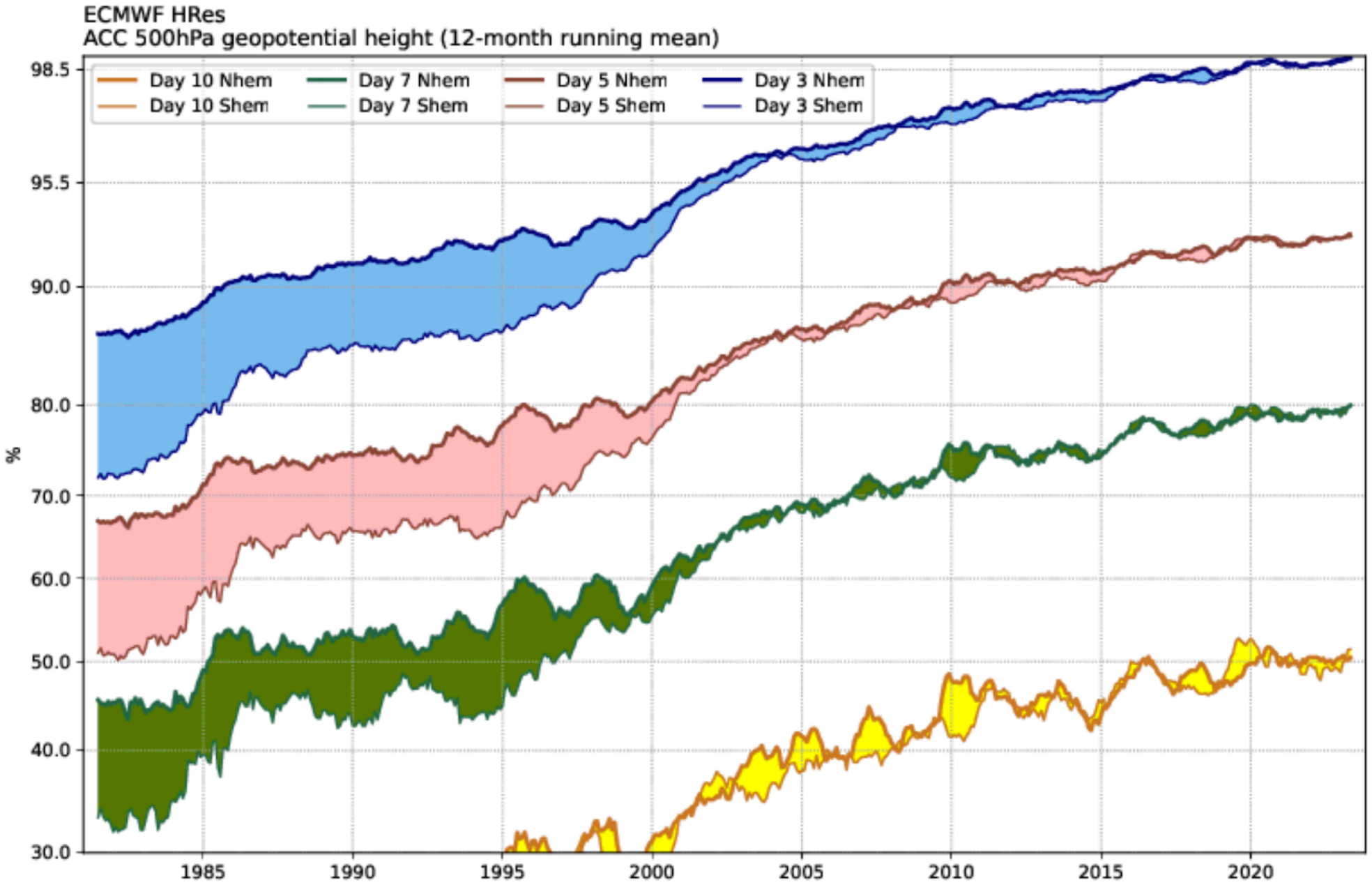
In particular

T. Haiden *et al., Evaluation of ECMWF forecasts, including the 2023 upgrade*, Technical Memorandum 911, September 2023, ECMWF, Reading, UK.

Available at the address :

https://www.ecmwf.int/en/elibrary/81389-evaluation-ecmwf-forecasts-including-2023-upgrade


500hPa geopotential
Lead time of 12m MA ACC reaching thresholds
NHem Extratropics
70%
85%
65%
80%
60%
75%
90%
day
10
9.5
9
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Spatial correlation between anomalies from climatology of forecast and verifying analysis
Score of climatology is 0. Score of persistence decreases from 1 to 0 over 4-5 days


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

2020

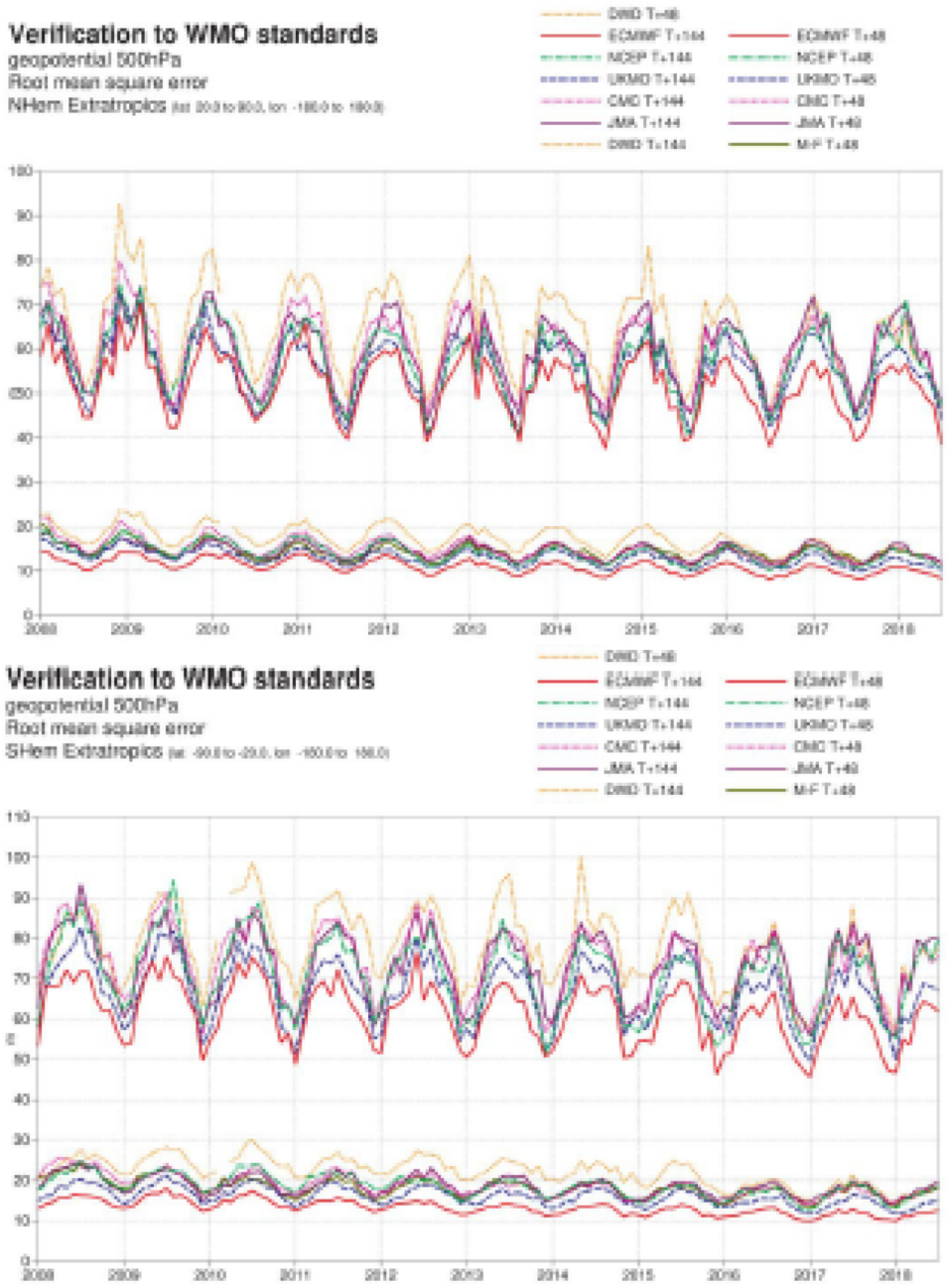


Figure 14: WMO-exchanged scores from global forecast centres. RMS error of 500 hPa geopotential height over northern (top) and southern (bottom) extratropics. In each panel, the upper curves show the six-day forecast error and the lower curves show the two-day forecast error of model runs initiated at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

September 2022 – August 2023

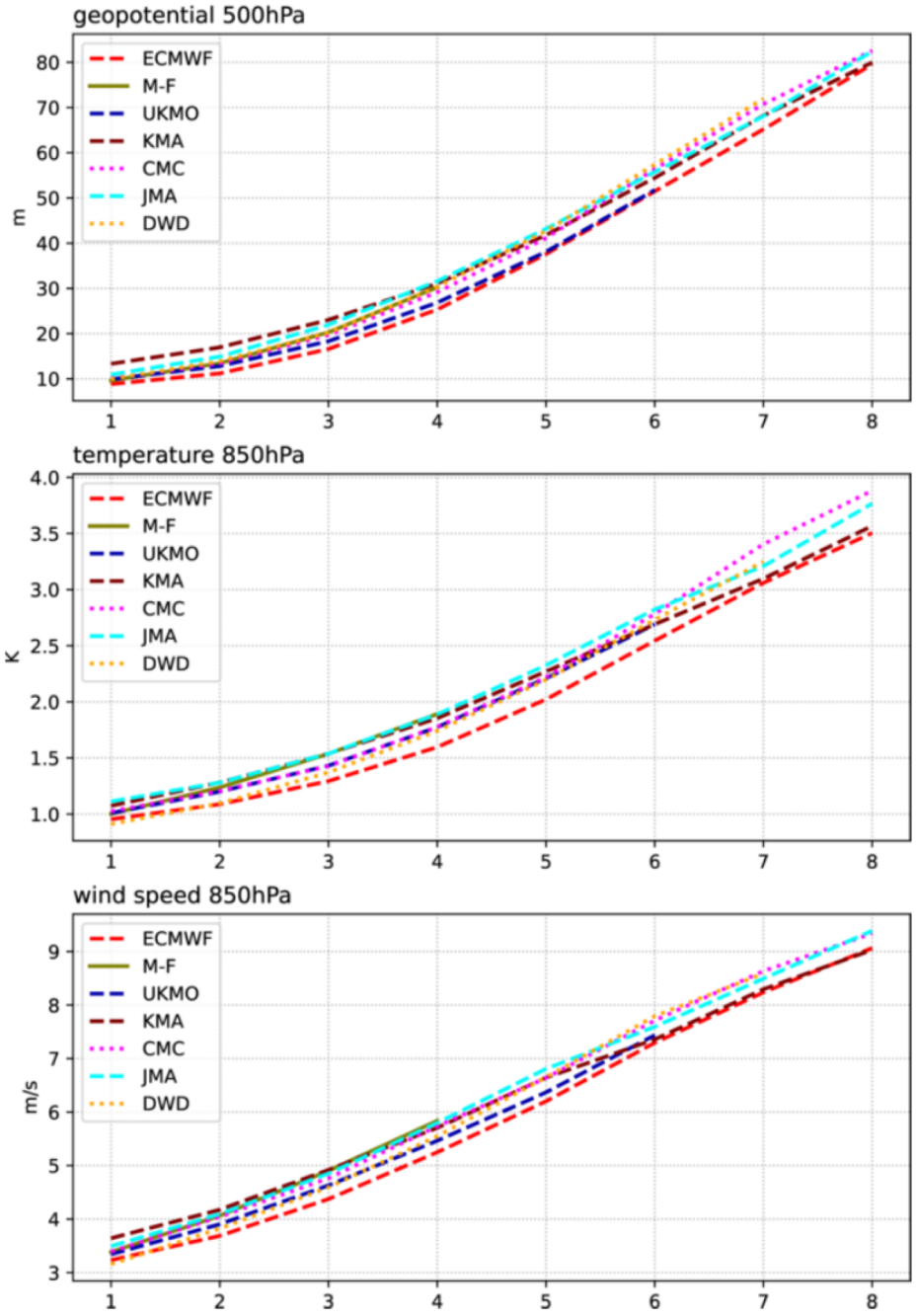


Figure 18: WMO-exchanged scores for verification against radiosondes: 500 hPa height (top), 850 hPa temperature (middle), and 850 hPa wind (bottom) RMS error over Europe and North Africa (annual mean August 2022–July 2023) of forecast runs initialized at 12 UTC. M-F = Météo-France, JMA = Japan Meteorological Agency, CMC = Canadian Meteorological Centre, UKMO = the UK Met Office, KMA = Korea Meteorological Administration, DWD = Deutscher Wetterdienst.

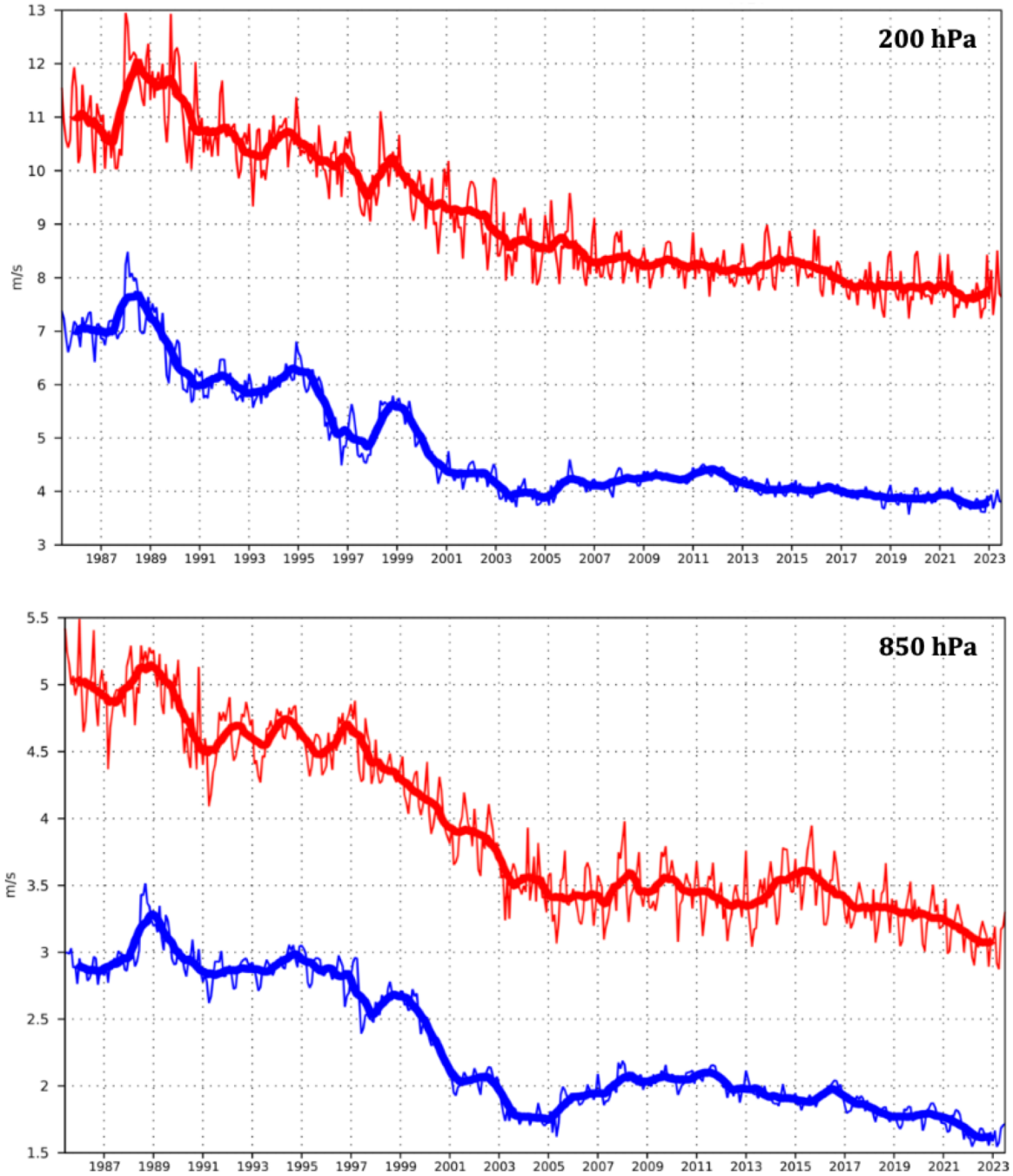


Figure 16: Forecast performance in the tropics. Curves show the monthly average RMS vector wind errors at 200 hPa (top) and 850 hPa (bottom) for one-day (blue) and five-day (red) forecasts, verified against analysis. 12-month moving average scores are also shown (in bold).

ECMWF

500 hPa geopotential

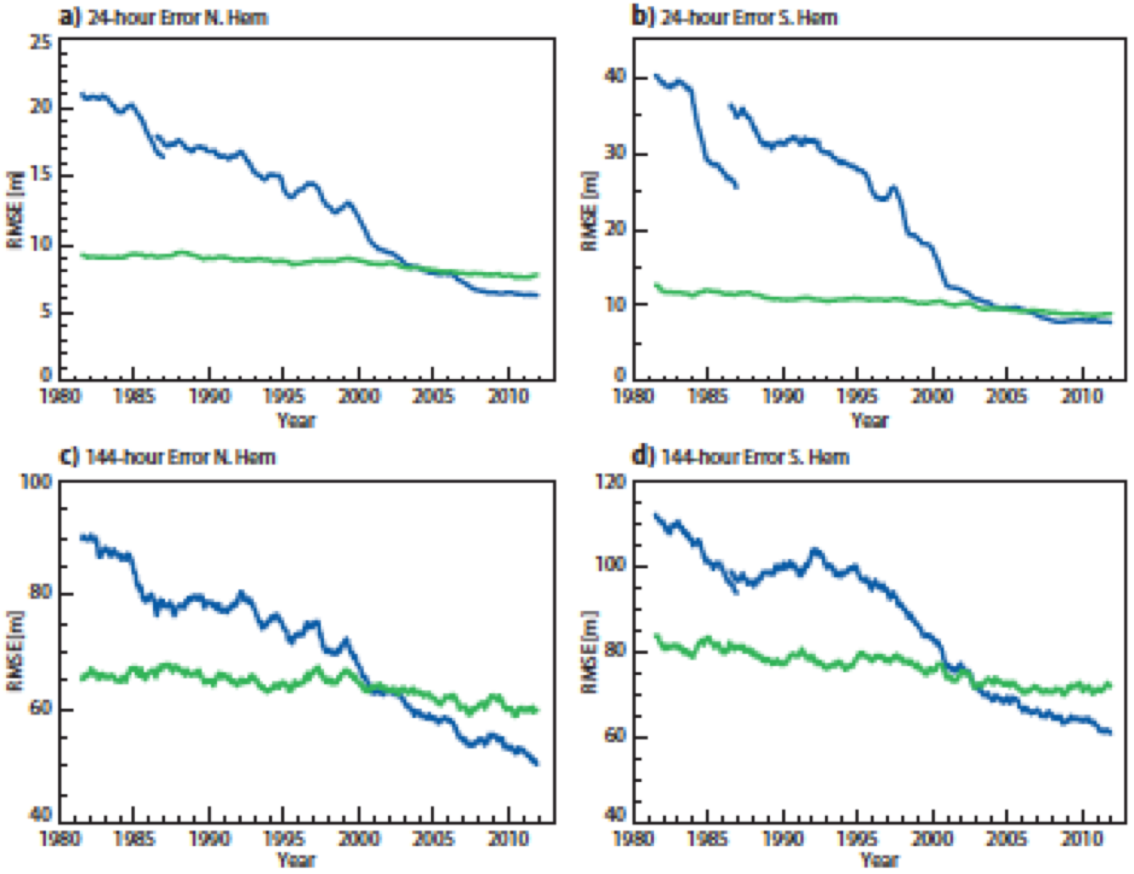


FIG. 3. Evolution of forecast errors from 1981 to 2012 for N.Hem (a and c) and S.Hem (b and d). Operational forecasts (blue) and ERA Interim (green). Note that before 1986 the operational analysis is used to verify the operational forecasts, after 1986 ERA Interim is used for the verification (with an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

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

Remaining Problems

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, …)

- …

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, …)

- …

Alternative Approach to Numerical Weather Prediction

Machine Learning (aka ***Artificial Intelligence***)

Set of empirical data

$$(x_i\,,\,y_i)\;,\;i=1,\,N$$

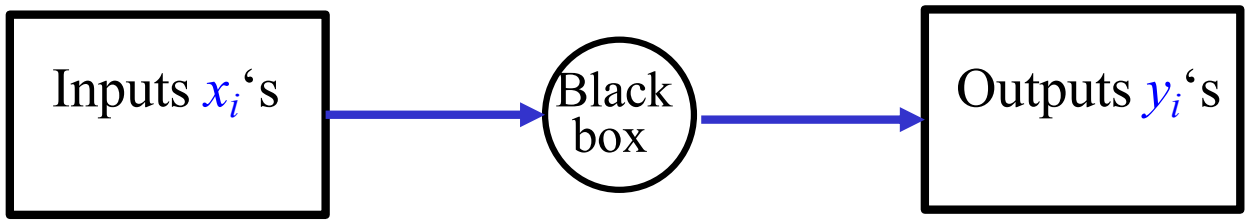
with no *a prori* explicitly known relationship between the inputs x_i's and the outputs y_i's.

Look for an explicit relationship of the form

$$y \approx f(x)$$

at least over a practically useful domain of variation of x.

Machine Learning (2)



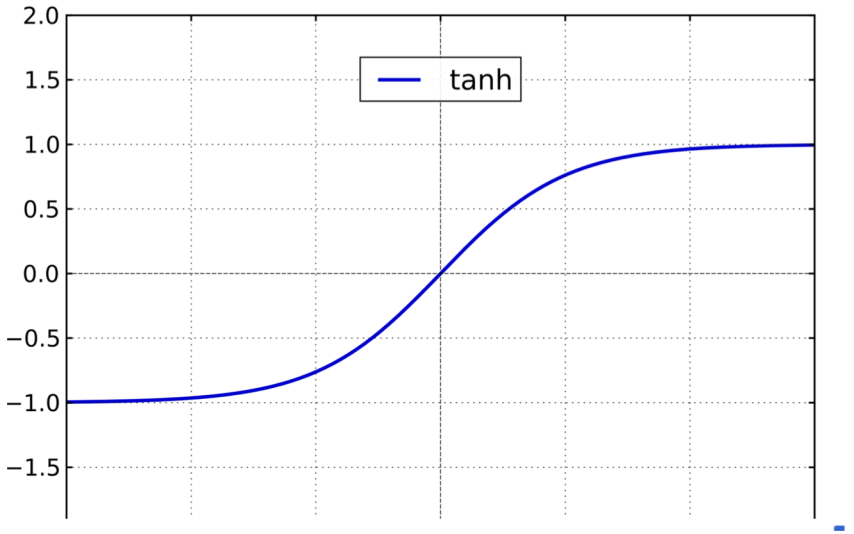
Replace black box with (possibly approximate) function $y \approx f(x)$

Neural networks define the function f as a composition of basic 'simple' functions. Sigmoid functions, *e.g.* the *hyperbolic tangent* function $\tanh(x)$, are very commonly used.

Machine Learning (3)

Neural networks

$\tanh(x) = (e^{2x} - 1) / (e^{2x} + 1)$



Affine change of coordinates. Four degrees of freedom : two for the coordinates of the central point, and one for the range of variation in either direction.

Machine Learning (4)

Neural networks

The initial empirical dataset (x_i , y_i) is typically divided into two sets

- A *training set* over which the composition of basic functions is defined. This usually involves several layers of '*neurons*', the neurons in each layer being compositions of neurons in previous layers. The optimal combination is obtained by minimizing the misfit between the original and computed outputs, often on the basis of a least-squares criterion.
- A *validating set* used to estimate the quality of the adjustment obtained from the training set.

Machine Learning (5)

Neural networks

This approach, with many variants, has proved to be extremely efficient, and is now used for innumerable applications in many different domains.

It has been applied to numerical weather prediction. A number of recently developed softwares are

GraphCast

Pangu-Weather

FourCastNet

FuXi

Machine Learning (5). *Neural networks*

These software pieces have been trained on the ECMWF Reanalysis v5 (ERA5).

ERA5, which covers the period from January 1940 to present, provides hourly estimates of a large number of atmospheric, land and oceanic climate variables. The data cover the Earth on a 30km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80km. ERA5 includes information about uncertainties for all variables. It is produced using 4D-Var data assimilation over 12-hour assimilation windows.

These neural networks trained on ERA5 produce forecasts of quality similar to, or better than, the ECMWF operational forecasts at a much lower numerical cost. Their results are accessible online on the Website of ECMWF.

Machine Learning (6). *Neural networks*

FuXi (伏羲) has been trained on 39 years of ERA5. It has a spatial resolution of 0.25° (28 km, against 9 km for ECMWF HRES) and produces forecasts for a number of meteorological variables

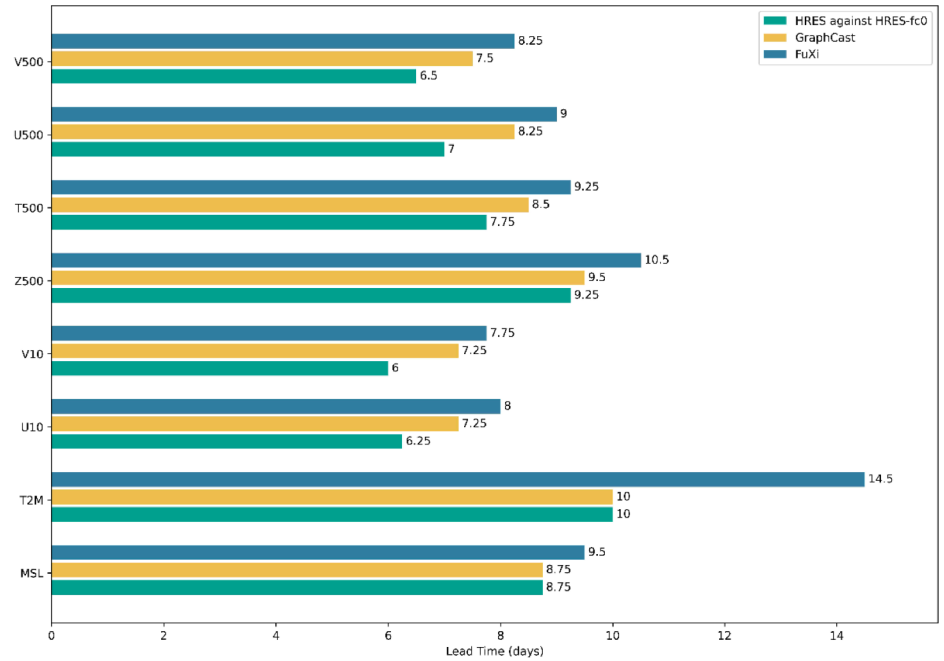
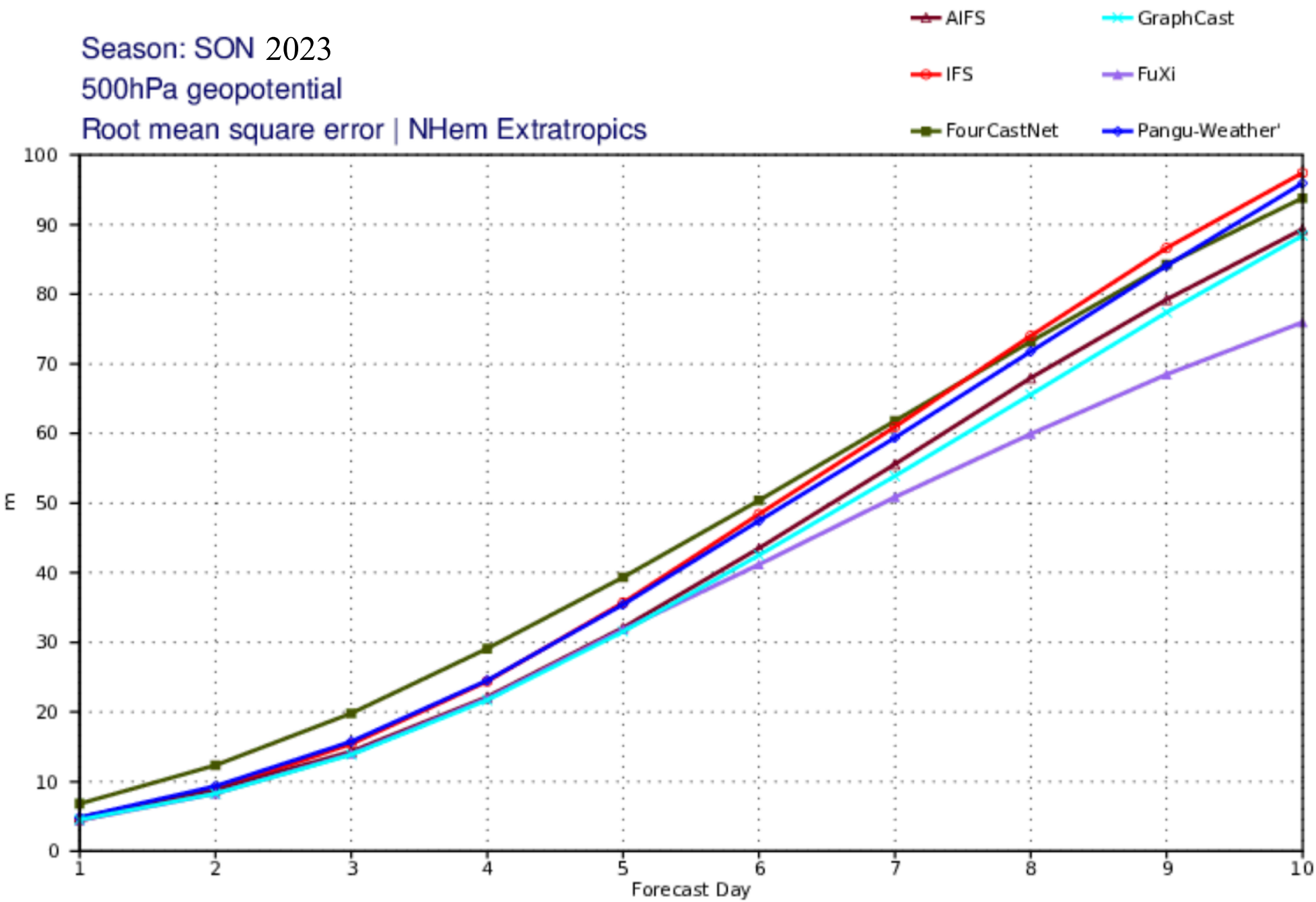


Fig. B.2: Skillful forecast lead times (ACC >0.6) of ECMWF HRES, Graphcast, and FuXi for 4 surface variables (MSL, $T2M$, $U10$, and $V10$) and 4 upper-air variables ($Z500$, $T500$, $U500$, and $V500$) at 500 hPa pressure level.



RMS error of forecasts by experimental machine learning models

Machine Learning (7).

- Machine Learning forecasts are more accurate in terms of RMS error and correlation coefficients, but are also spatially much smoother.

- Machine learning can be implemented for estimating errors in deterministic forecasts

- But, at this stage, machine learning still depends totally on the availability of a training set produced by well-established and thoroughly validated means. How will these be updated ?

Machine Learning (8).

- Machine Learning can be expected to have a significant impact on the way weather prediction is going to be performed, as well as on the quality of the corresponding forecasts.

- But what will precisely be that impact ? It is simply too soon to tell.

- What is assimilation ?

  - *Definition of initial conditions*

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 $\Rightarrow$ 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).

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, $p \approx 10^7$-10^8 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

Proportion of computing resources devoted to assimilation of observations in the whole process of Numerical Weather Prediction has gradually increased over time.

Definition of initial conditions originally required a simple interpolation from observation stations to model gridpoints, with negligible cost. As of now, assimilation over 24 hours of observations requires about the same amount of resources as a 10-day forecast, including probabilistic forecast.

$z_1 = x + \zeta_1$ 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

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

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

$$\begin{aligned} P(x = \xi \mid z_1, z_2) &\propto p_1(z_1 - \xi)\, p_2(z_2 - \xi) \\ &\propto \exp[-(z_1 - \xi)^2/2s_1] \exp[-(z_2 - \xi)^2/2s_2] \\ &= \exp[-A/2] \end{aligned}$$

with $A = (z_1 - \xi)^2/s_1 + (z_2 - \xi)^2/s_2$

$= (\xi - x^a)^2/p^a$ + *terms independent of* ξ

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

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

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

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

$1/p^a = 1/s_1 + 1/s_2$

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

$x^a = p^a\,(z_1/s_1 + z_2/s_2)$ is weighted average of z_1 and z_2, with respective weights $1/s_1$ and $1/s_2$. Larger weight is given to more accurate piece of data.

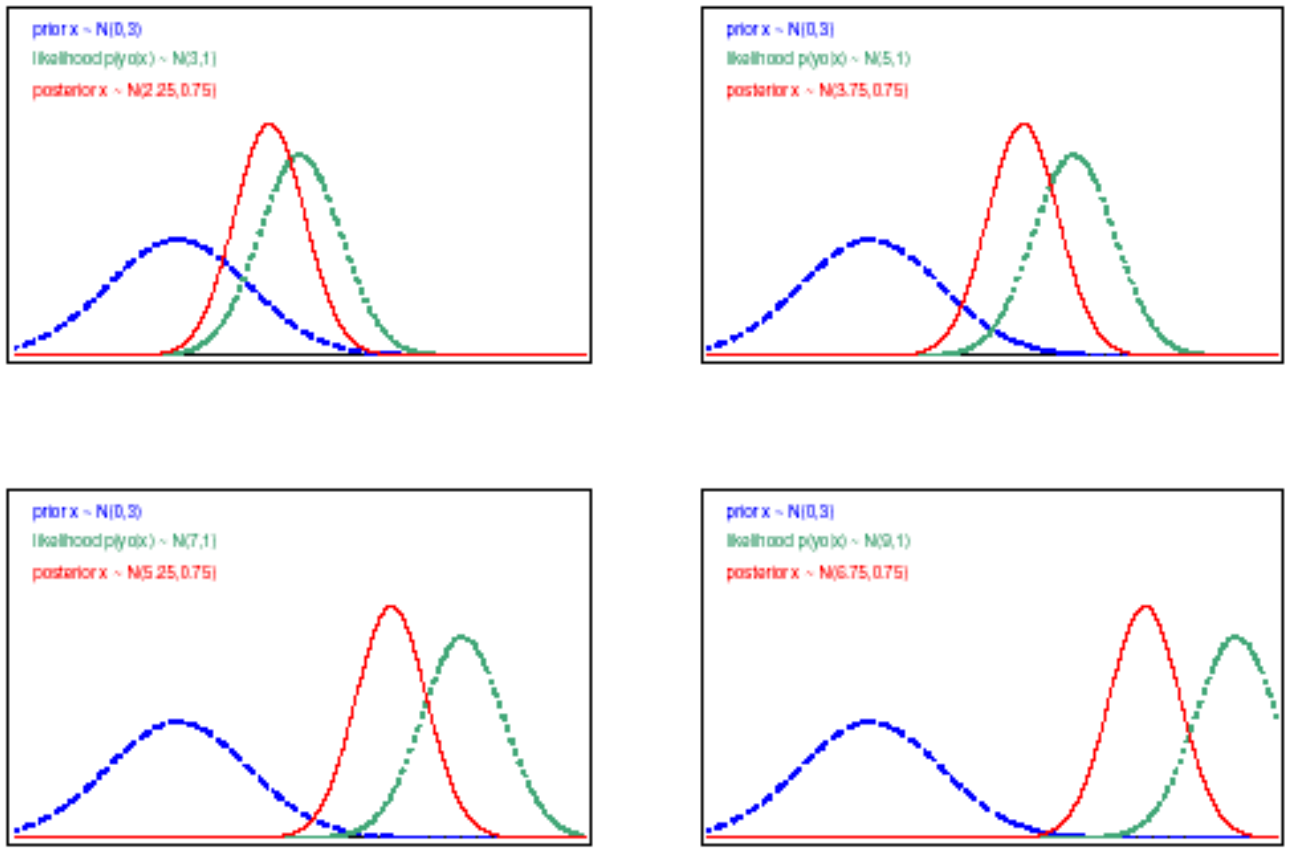


Fig. 1.1: Prior pdf $p(x)$ (dashed line), posterior pdf $p(x|y^o)$ (solid line), and Gaussian likelihood of observation $p(y^o|x)$ (dotted line), plotted against x for various values of y^o. (Adapted from Lorenc and Hammon 1988.)

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*)

International Symposium on Data Assimilation - Online (ISDA-Online)

"*Machine Learning in Data Assimilation*"

Alan Geer I ECMWF, UK
Stephan Rasp | Google Research, US
Ronan Fablet I IMT Atlantique, France

Friday, January 12, 2024 | 15:00 - 17:00 UTC
(04 - 06 pm CET Berlin / 10 - 12 am EST New York)
Visit our website: **https://isda-online.univie.ac.at/**

Please use the following link to connect to the Webex Webinar on January 12:
https://awi.webex.com/awi-en/j.php?MTID=m6ff41a0d091766f4c01561e971bb7a22
Webinar password: qxJYgUHJ383
(79594845 from phones and video systems)

International Symposium on Data Assimilation - Online (ISDA-Online)

"***Advancements in Variational Data Assimilation***"

Friday, February 2, 2024 from 08-10 UTC

Visit our website: **https://isda-online.univie.ac.at/**