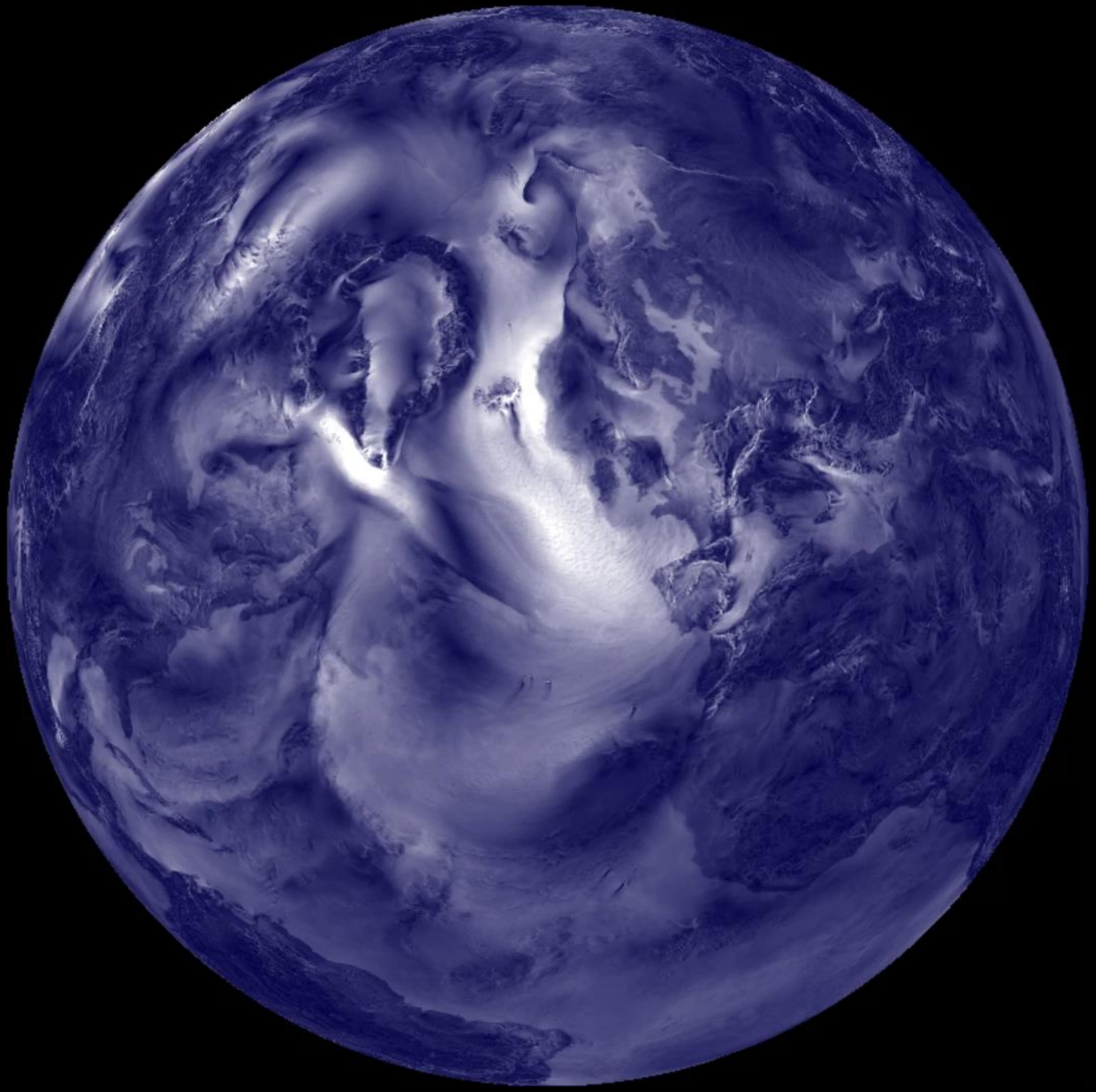
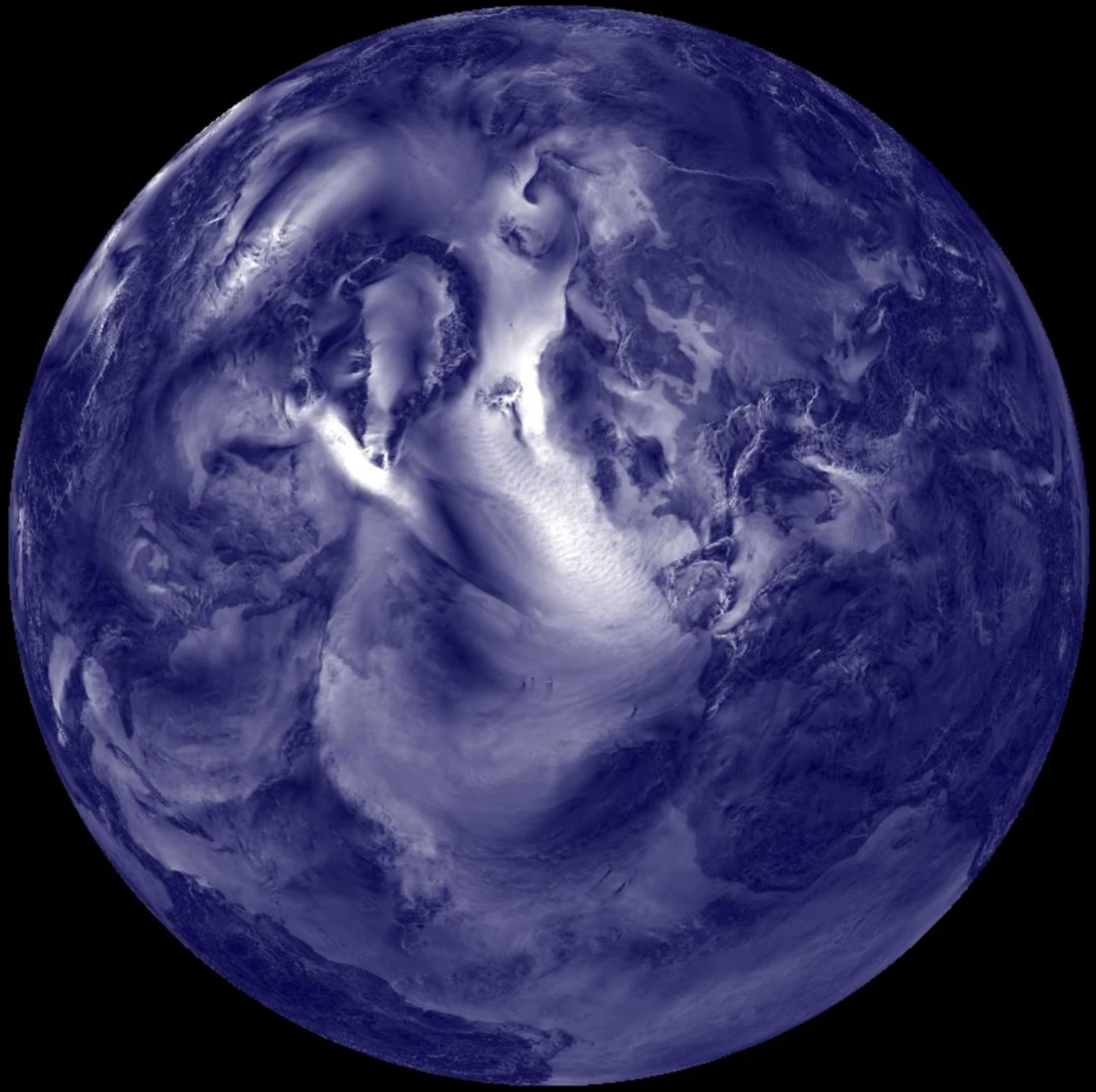
Ensemble Forecasts
Initial Perturbation

Simon Lang

IFS 10m wind gusts, 2020-12-04 00 UTC 720h forecasts, 9 km spatial resolution



Control Member



Perturbed member 1

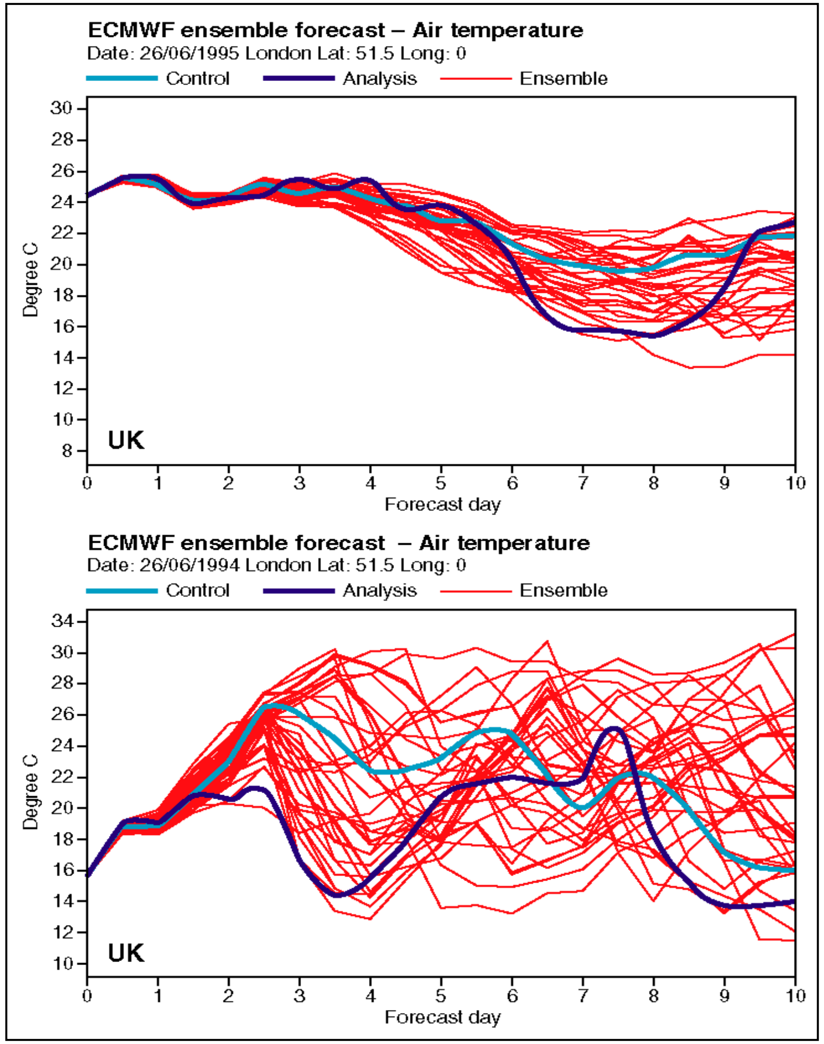
Chaos and weather prediction

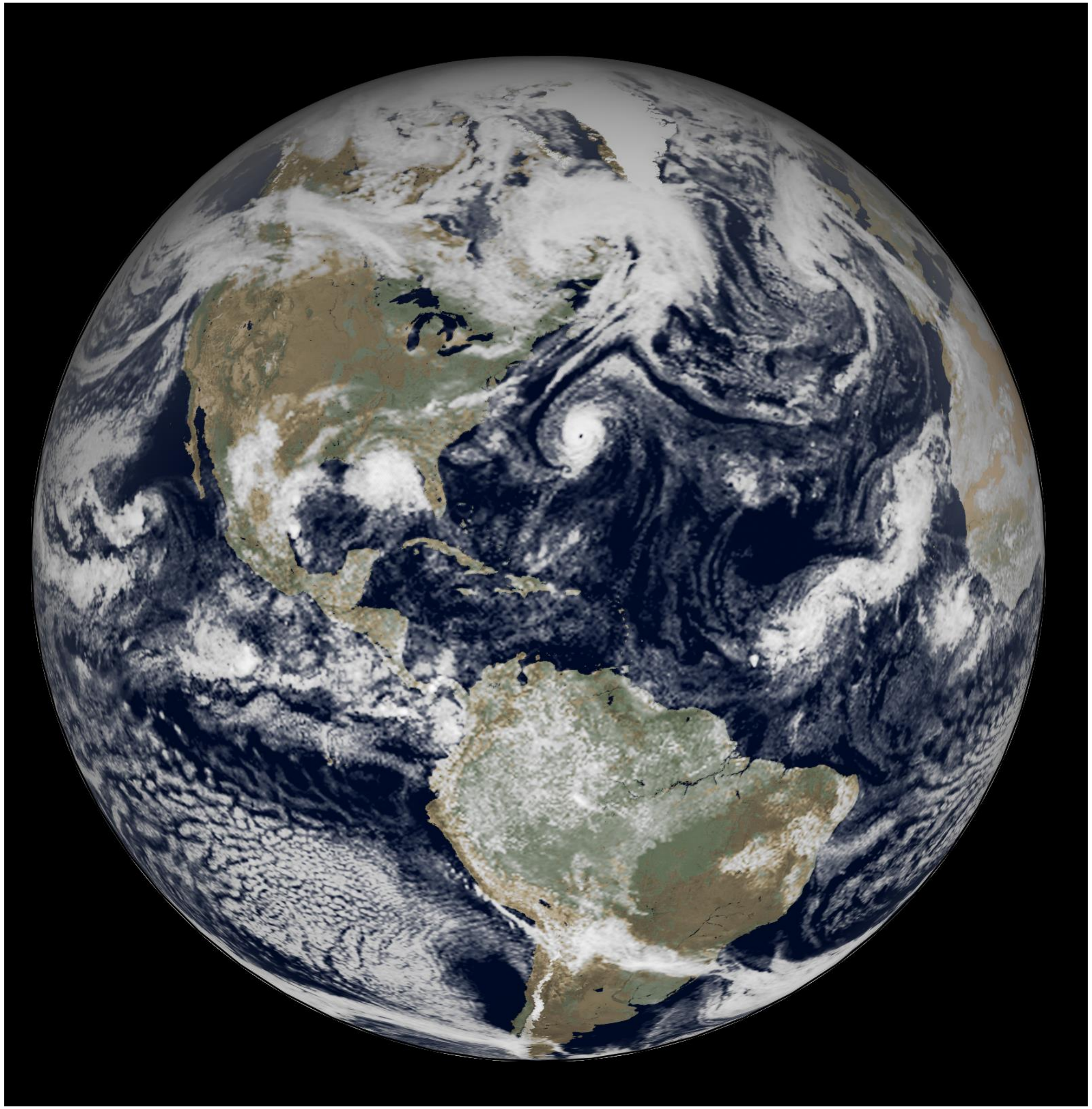
The atmosphere is a chaotic system

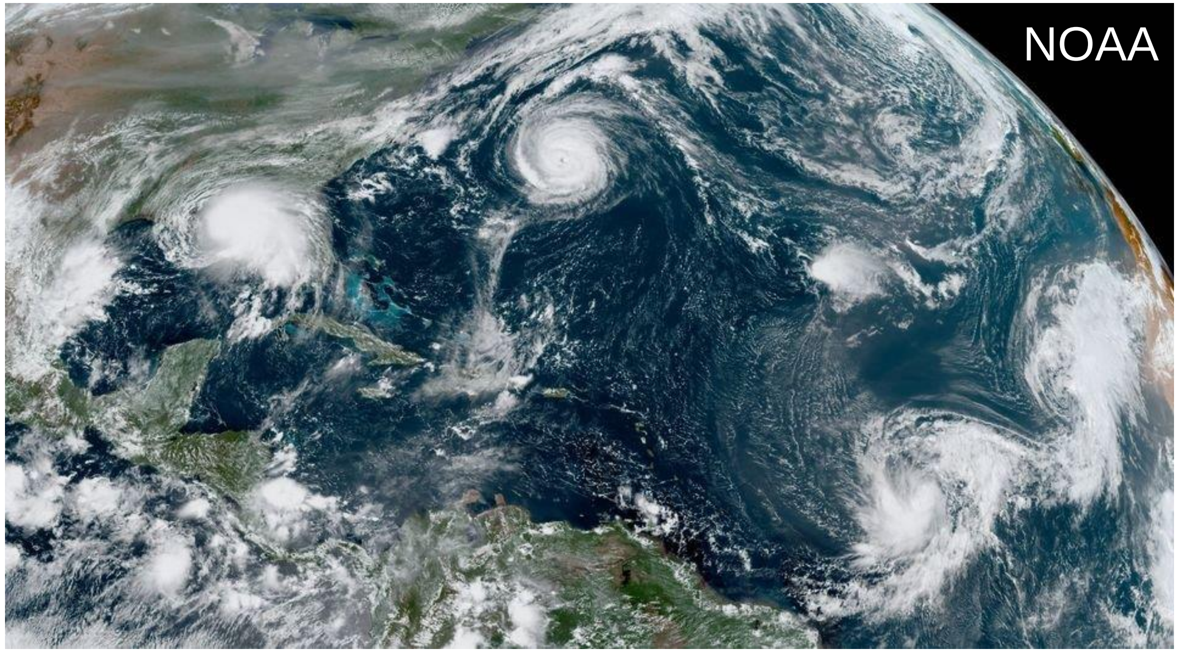
- Small errors can grow to have major impact
- We can never perfectly measure the current state of the whole atmosphere

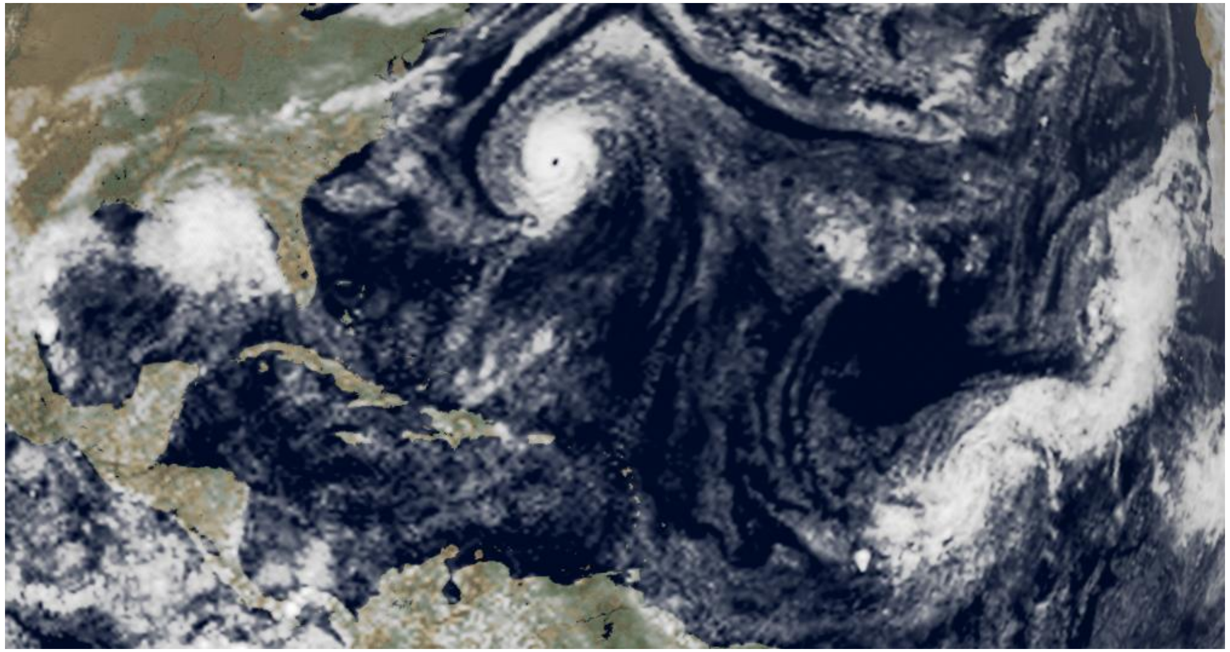
Ensemble Forecasts

- Parallel set of forecasts from very slightly different initial conditions and model formulation
- Assess uncertainty of today's forecast

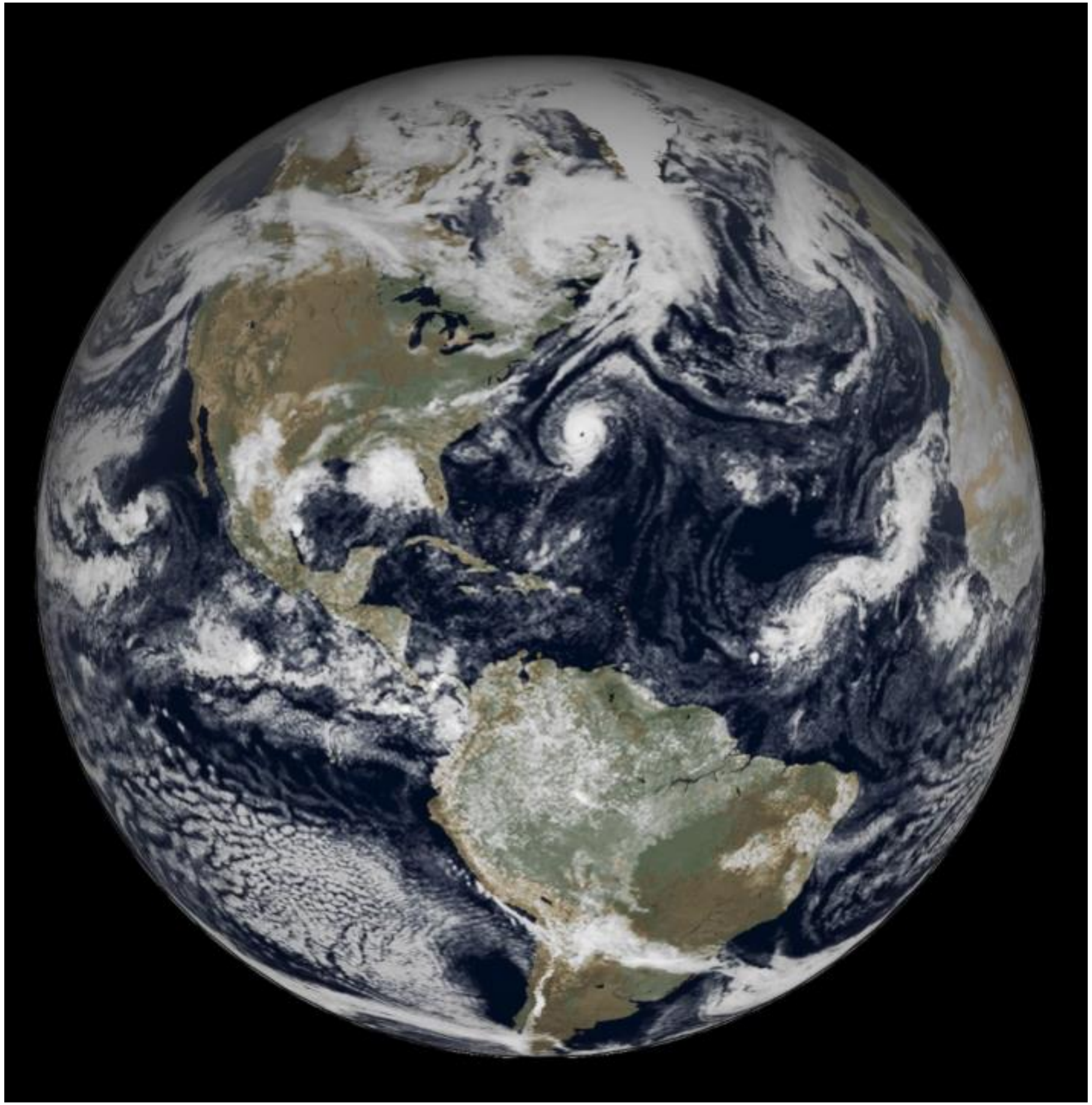


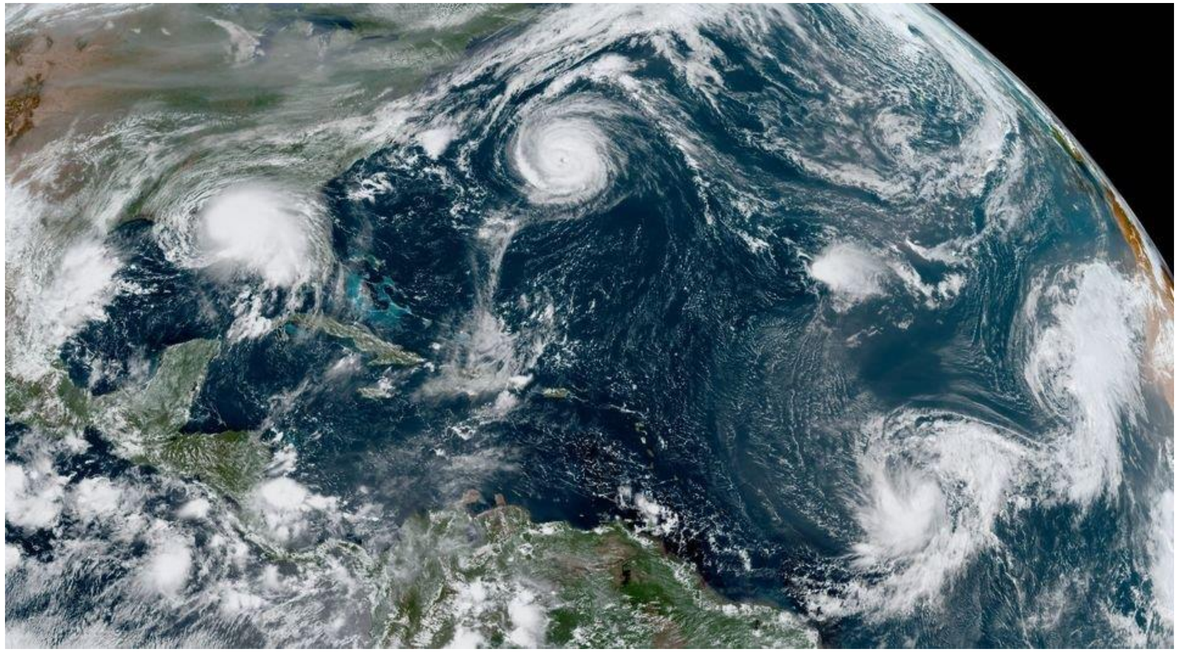






TCo1279L137 ENS, Control Member,
20200913 00 UTC + 41 h





NOAA

TCo1279L137 ENS, 51 Members,
20200913 00 UTC + 41 h

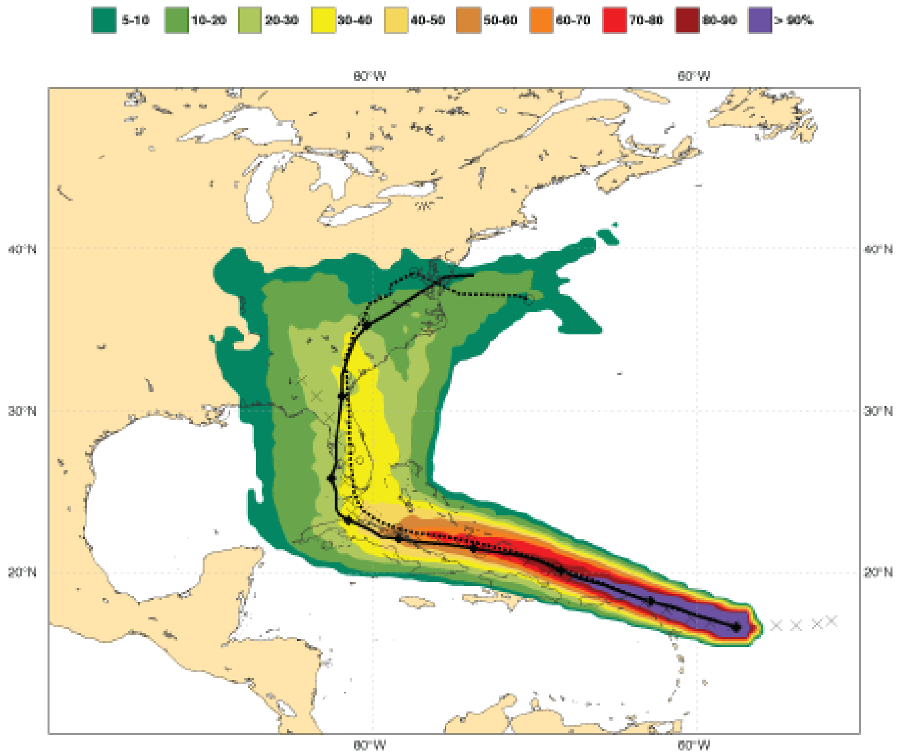
- 51 Members (50 perturbed + control member without perturbations), CY48R1 -> TCo1279 (~ 9 km) to day 15, extended-range, 100 member TCo319, ~ 36 km (see Lang. et al., 2023 for model cycle description).

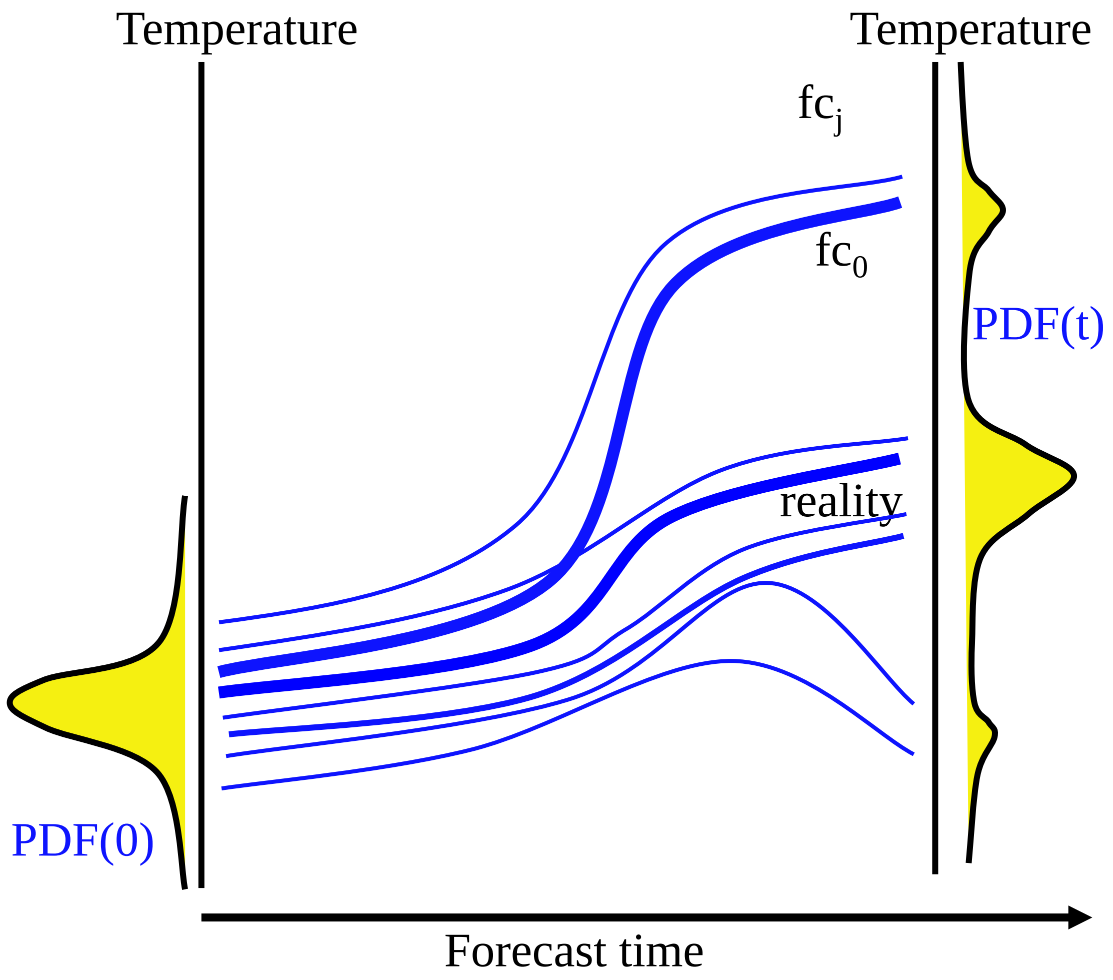
- 137 vertical levels

- Coupled to NEMO ocean model (1/4 degree), ecWAM wave model and LIM2 ice model

- Initial perturbation via an ensemble of data assimilations and singular vectors, 5 member ocean data assimilation

- Model error representation – currently SPPT





from R. Buizza

Sources of Uncertainty:

- Initial Conditions
- Model Formulation

Perturbations to the initial conditions:

Methods that rely on the dynamics only, e.g.:

- bred vectors
- singular vectors

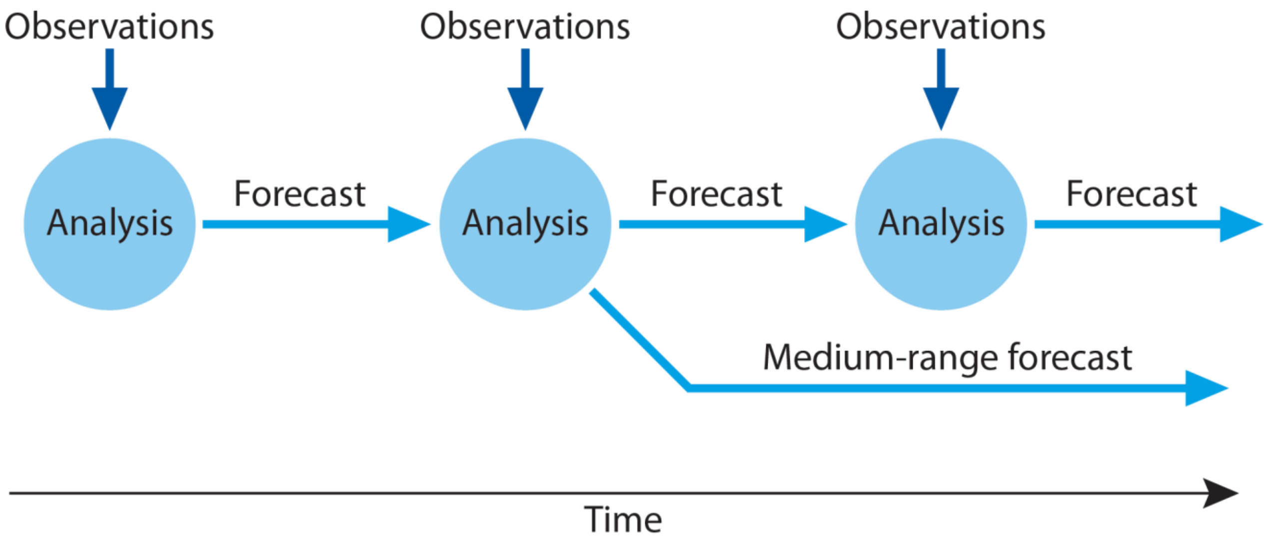
Ensemble data assimilation methods, e.g.:

- Ensemble of 4D-Var data assimilations (EDA)
- Ensemble Kalman Filter

ECMWF: combination of EDA and singular vectors
-> data assimilation methods know about obs coverage etc

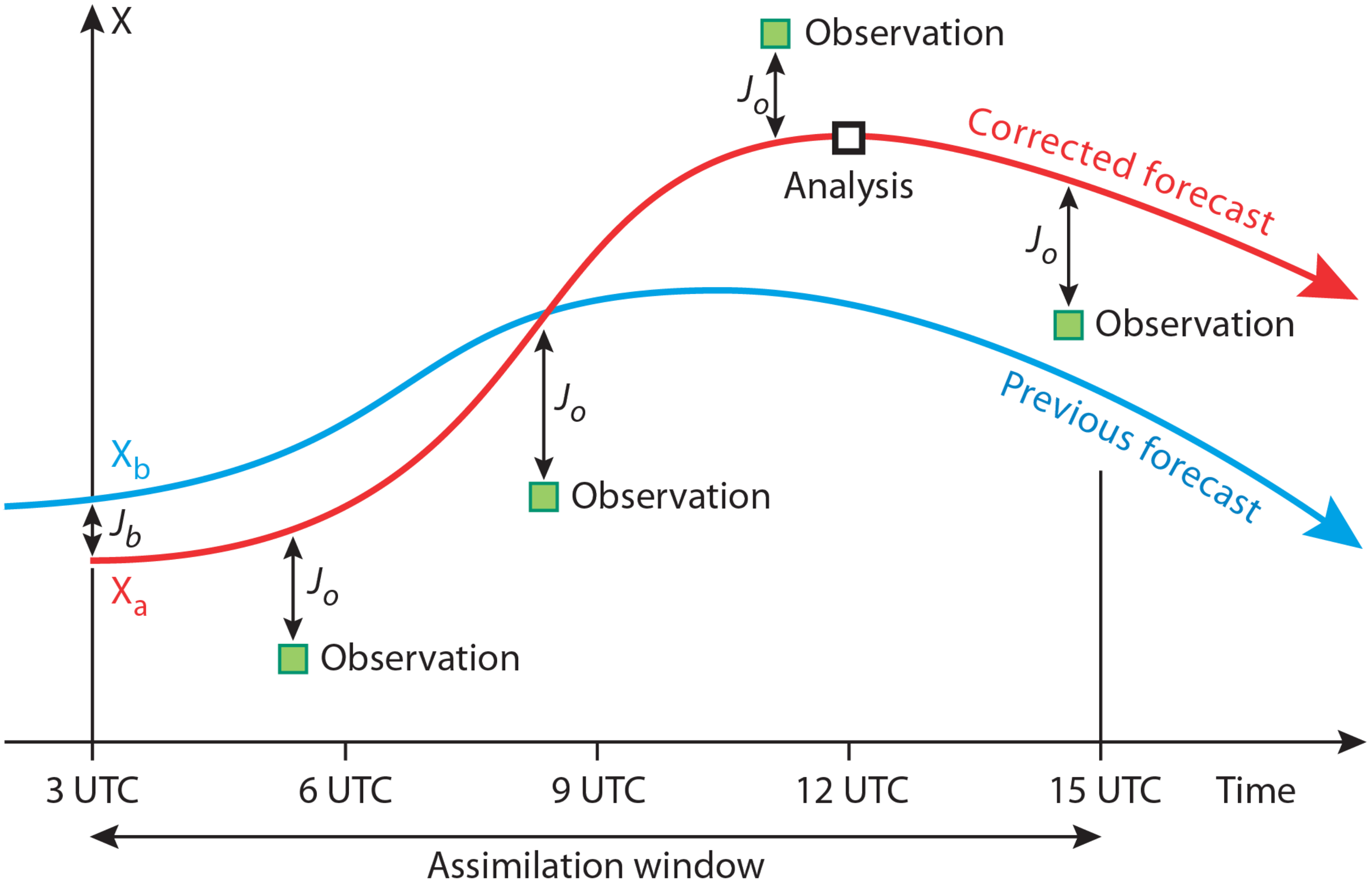
Starting the Medium-Range Forecast – the ‘Analysis’

Analysis: 3 dimensional virtual image of the atmosphere at a given time.



- The short range forecast from the previous analysis is our ‘first estimate’ of the current state of the atmosphere.

4D-Var assimilation

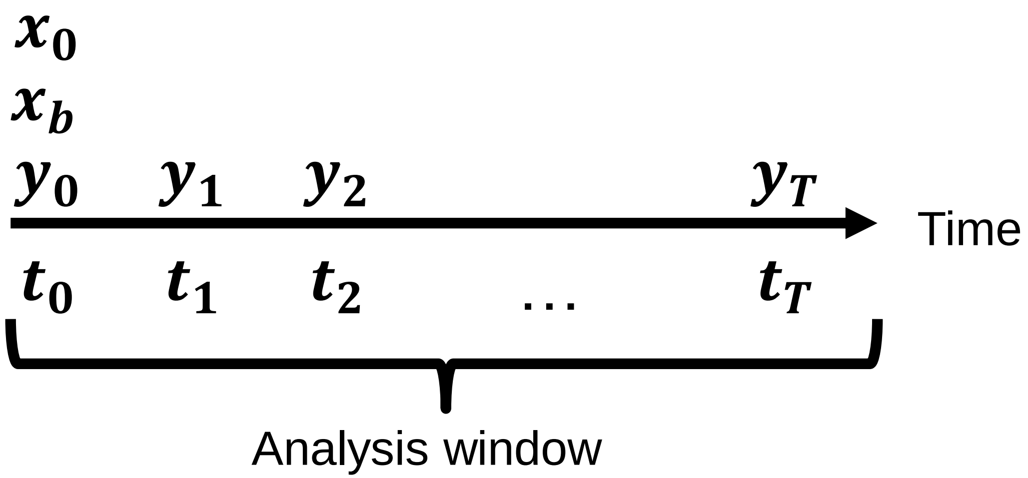


ECMWF

4D-Var assimilation

To find model trajectory that best fits the observations over an assimilation interval (t=0,1,…,T) - > finding the minimum of the 4DVar cost function:

$$J(\mathbf{x}_0) = (\mathbf{x}_b - \mathbf{x}_o)^T (\mathbf{P}^b)^{-1} (\mathbf{x}_b - \mathbf{x}_o) + \sum_{t=0}^{T} (\mathbf{y}_t - H_t M_{0\to t}(\mathbf{x}_0))^T \mathbf{R}_t^{-1} (\mathbf{y}_t - H_t M_{0\to t}(\mathbf{x}_0))$$



See lectures in DA Training

ECMWF

Ensemble of 4D-Var data assimilations (EDA)

- 50 perturbed ensemble members + 1 control:
  TCo639 outer loops (~ 18 km), 137 levels, TL191/TL191 inner loops.
  (HRES DA: TCo1279 outer loops (~ 9 km), TL255/TL319/TL399/TL511 inner loops).

- Observations randomly perturbed according to their estimated error covariances (R)

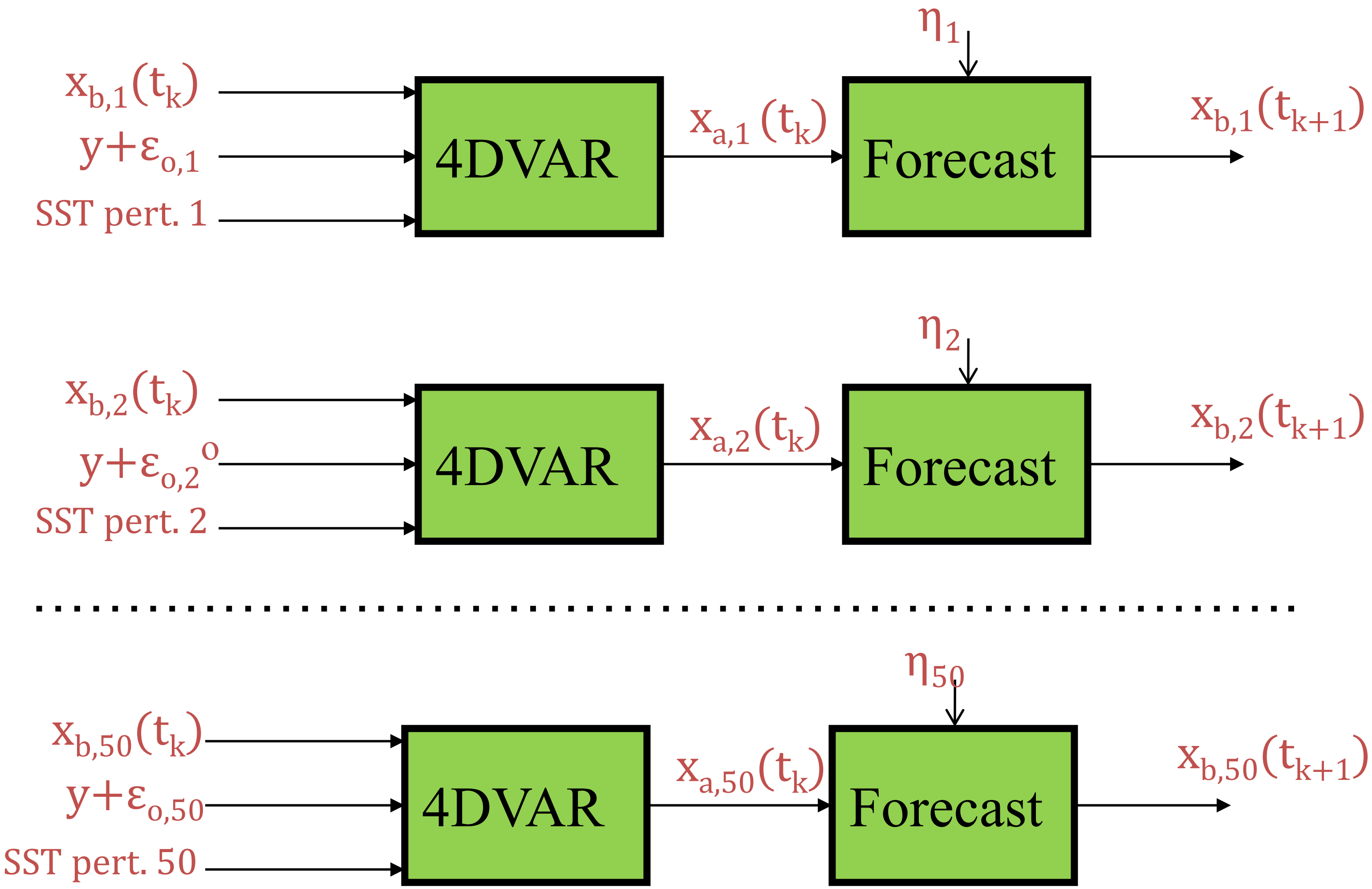
- SST perturbed with climatological error structures

- Model error representation via Stochastically Perturbed Parametrization Tendencies (SPPT)

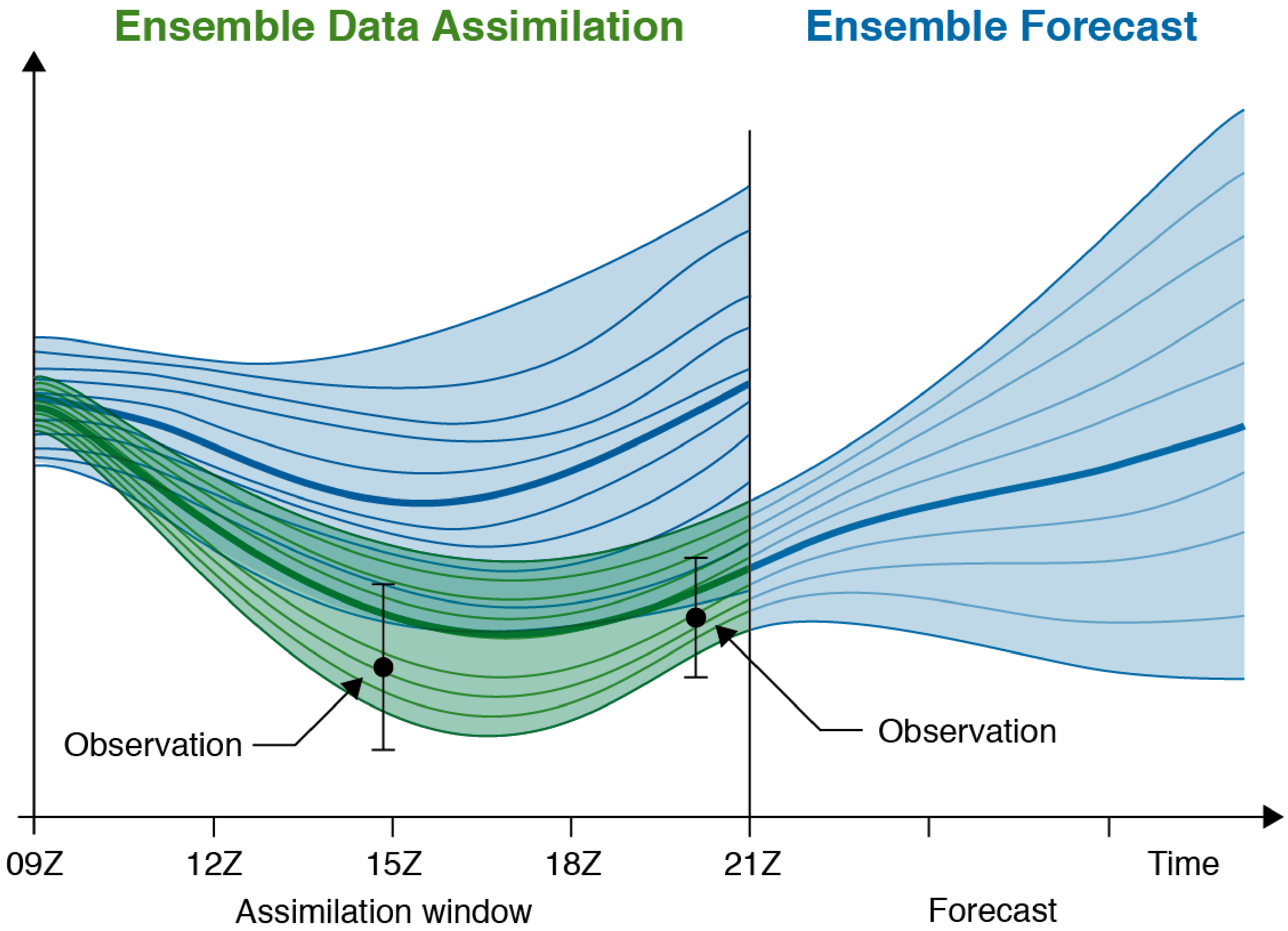
<u>The EDA simulates the error evolution of the 4DVar analysis cycle:</u>

→ uncertainty estimates to initialize ensemble forecasts

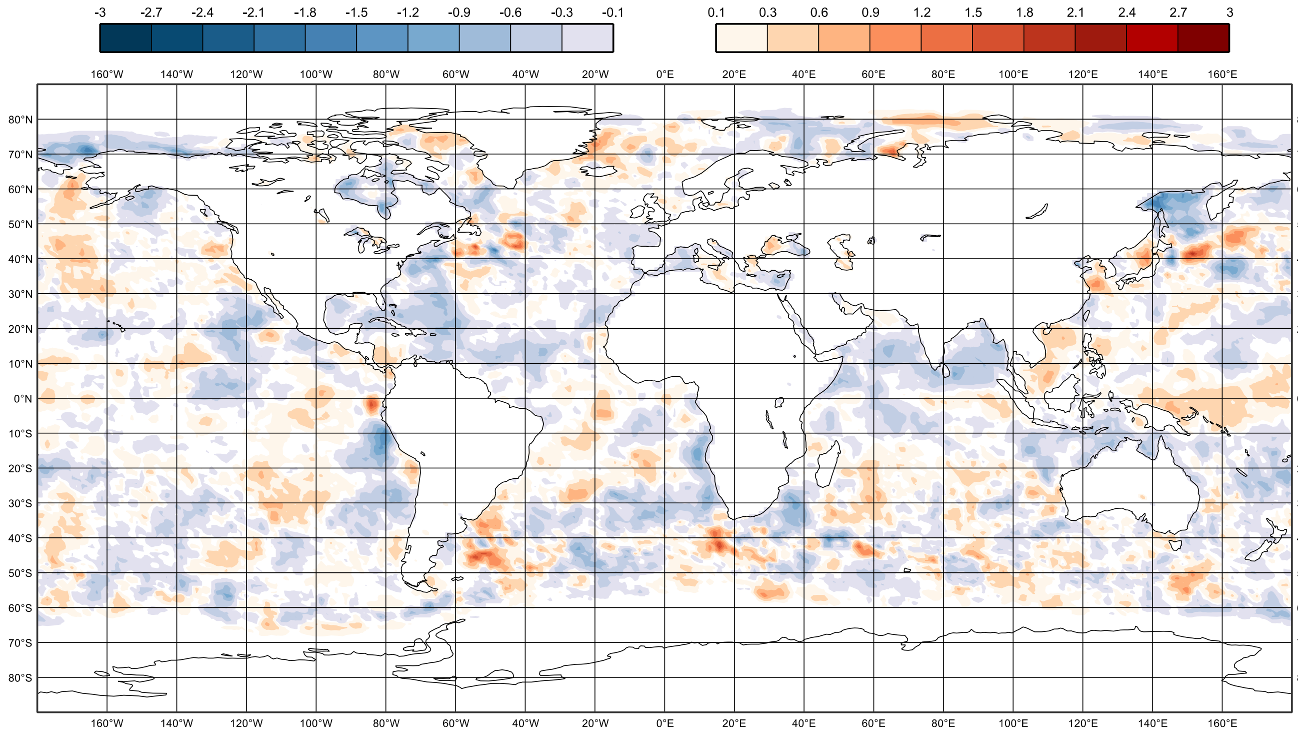
→ Flow dependent estimates of background error covariances for use in 4D-Var

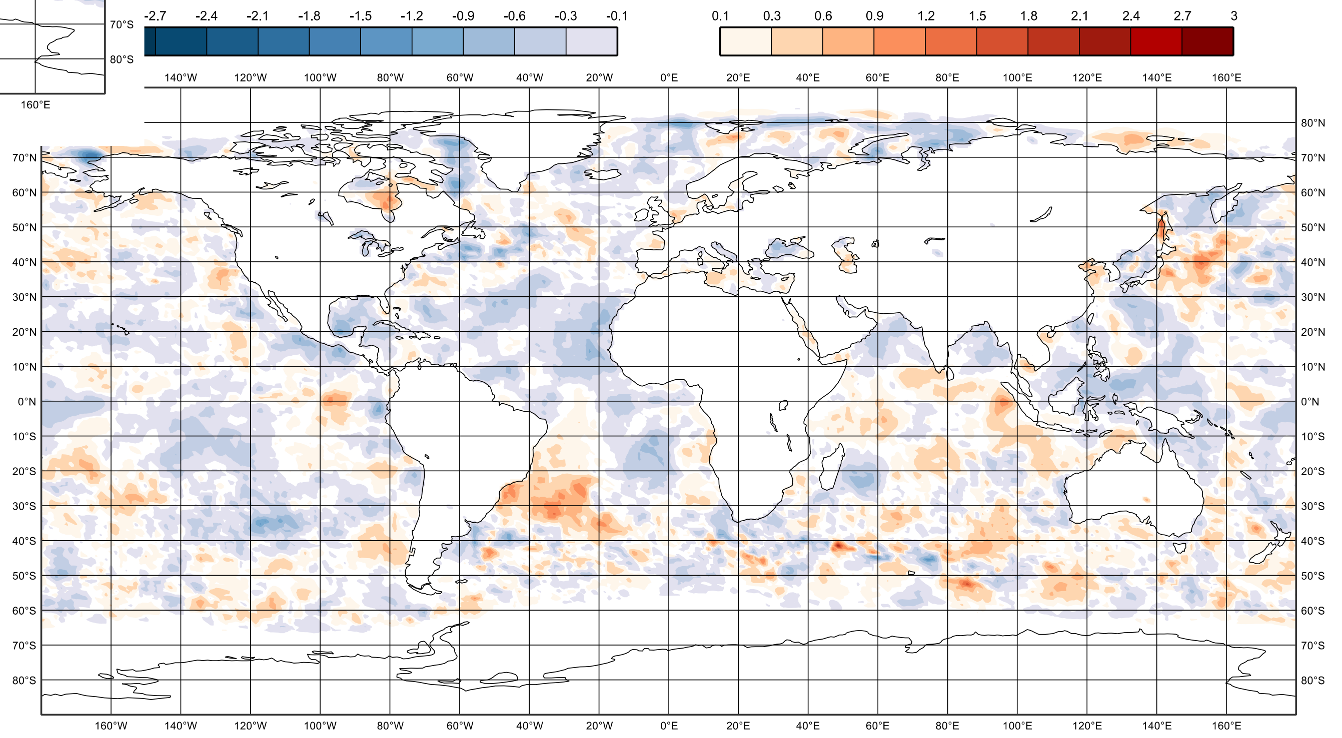


See also Massimo Bonavita's Talk in DA Training


Ensemble Data Assimilation
Ensemble Forecast
Observation
Observation
09Z
12Z
15Z
18Z
21Z
Time
Assimilation window
Forecast

Example: SST perturbations





Current Model Error Representation: SPPT

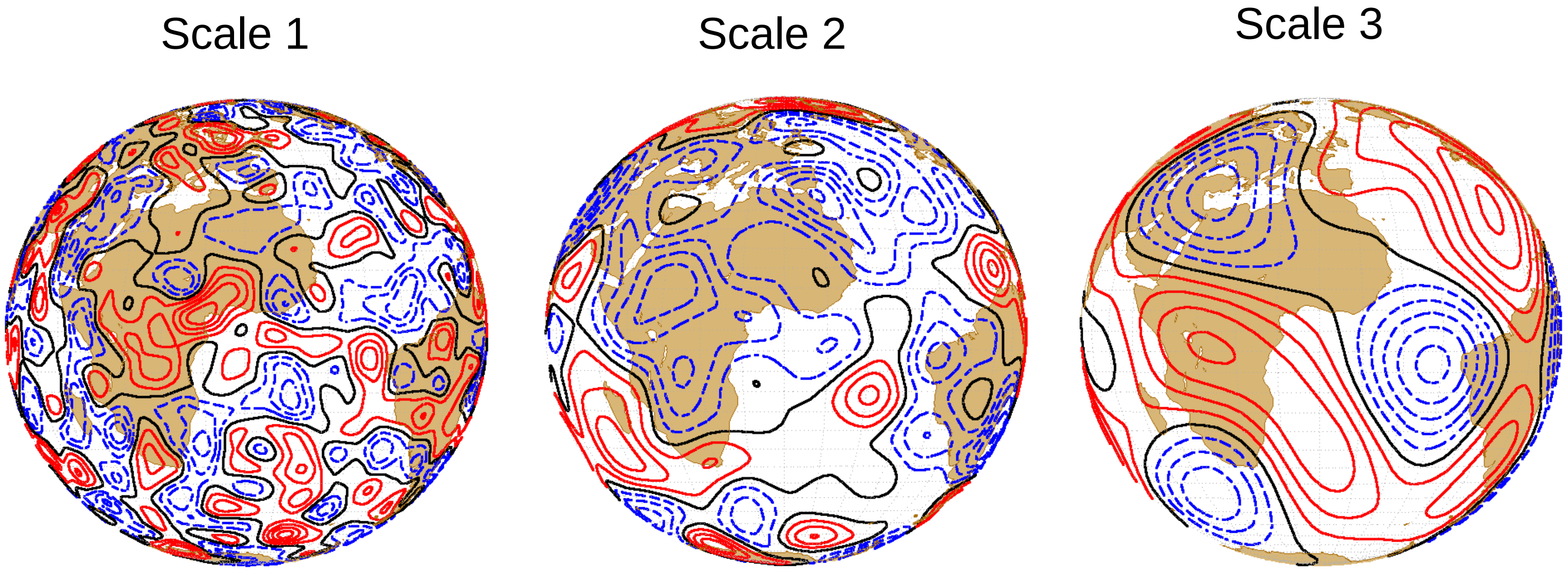
See Leutbecher et al., 2017 and Lock et. al, 2019 for details

Perturb model tendencies during the forecast:

$$\boldsymbol{x_p} = \boldsymbol{x} + \boldsymbol{\alpha x}$$

$\boldsymbol{x}$ sum of tendencies from parametrization schemes (convection, radiation, cloud etc.)

$\boldsymbol{\alpha}$ includes random time and space correlations, provided by a pattern generator



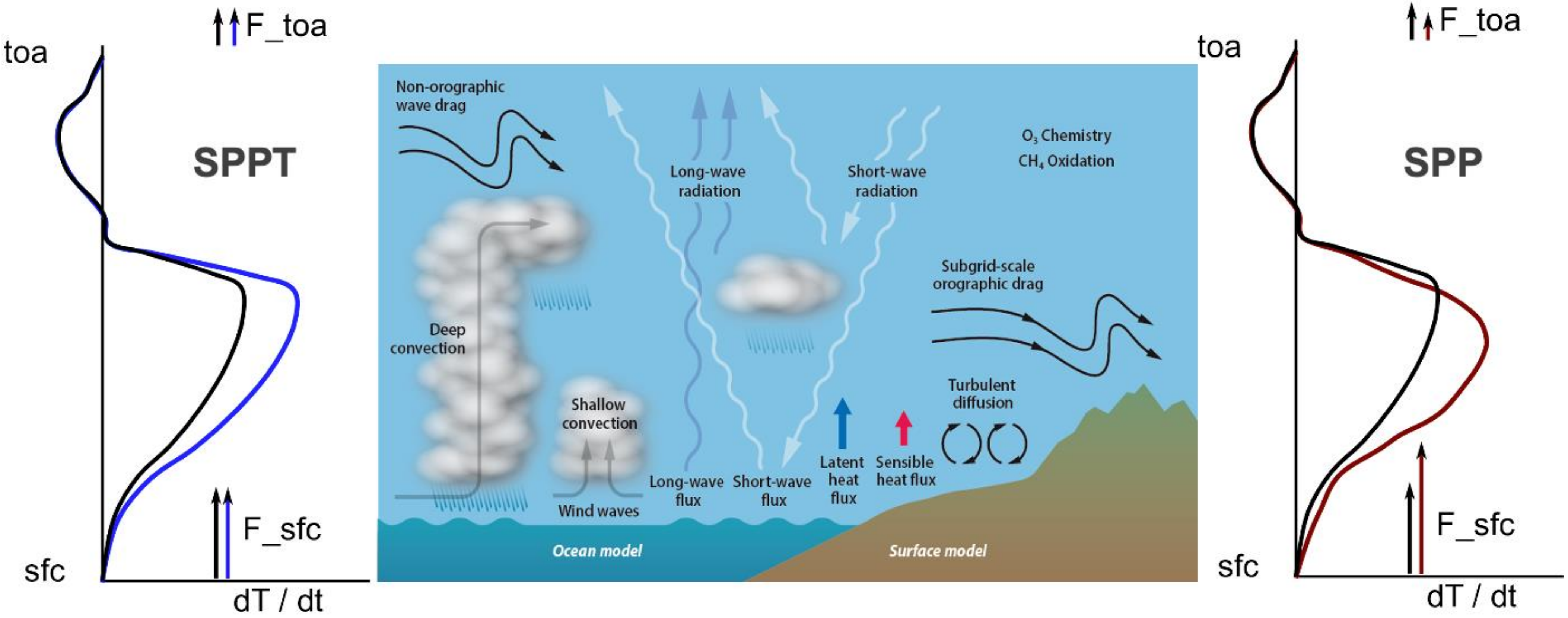
Same model uncertainty representation in ensemble forecasts and ensemble data assimilation

Future Model Error representation: SPP

Ollinaho et al (2017), https://doi.org/10.1002/qj.2931
Leutbecher et al (2017), https://doi.org/10.1002/qj.3094
Lang et al (2021), https://doi.org/10.1002/qj.3978

Key differences between SPPT and SPP:

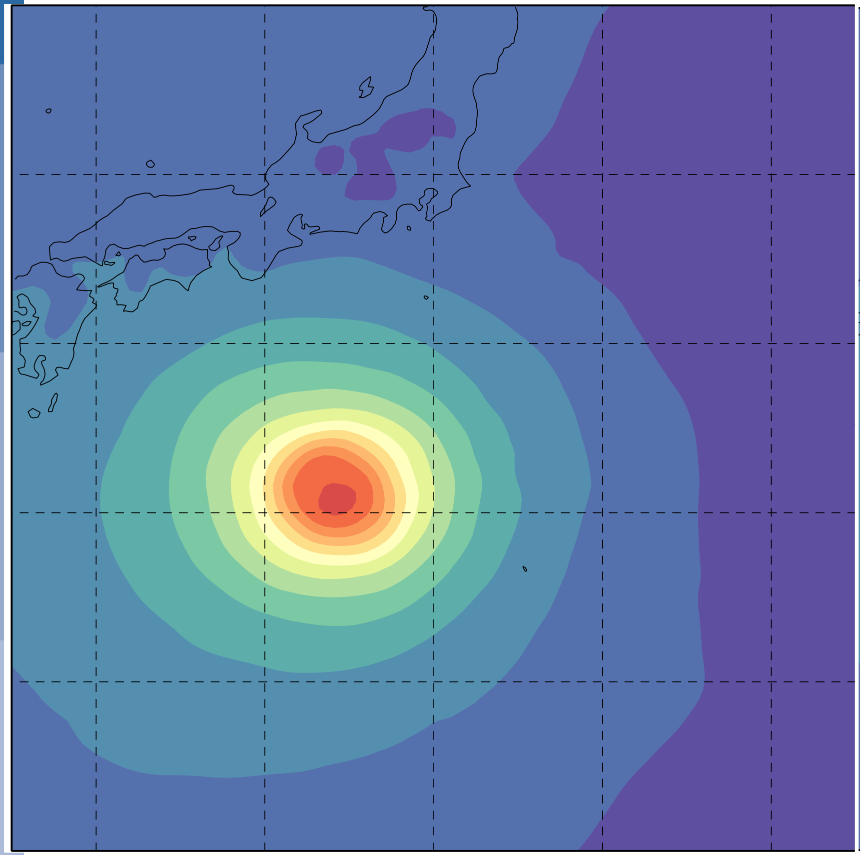
- SPP represents model uncertainties closer to the assumed sources of the errors
- SPP better maintains physical consistency: e.g. local budgets and flux perturbations
- SPPT only represents amplitude errors while SPP can also represent errors in the shape of a heating profile



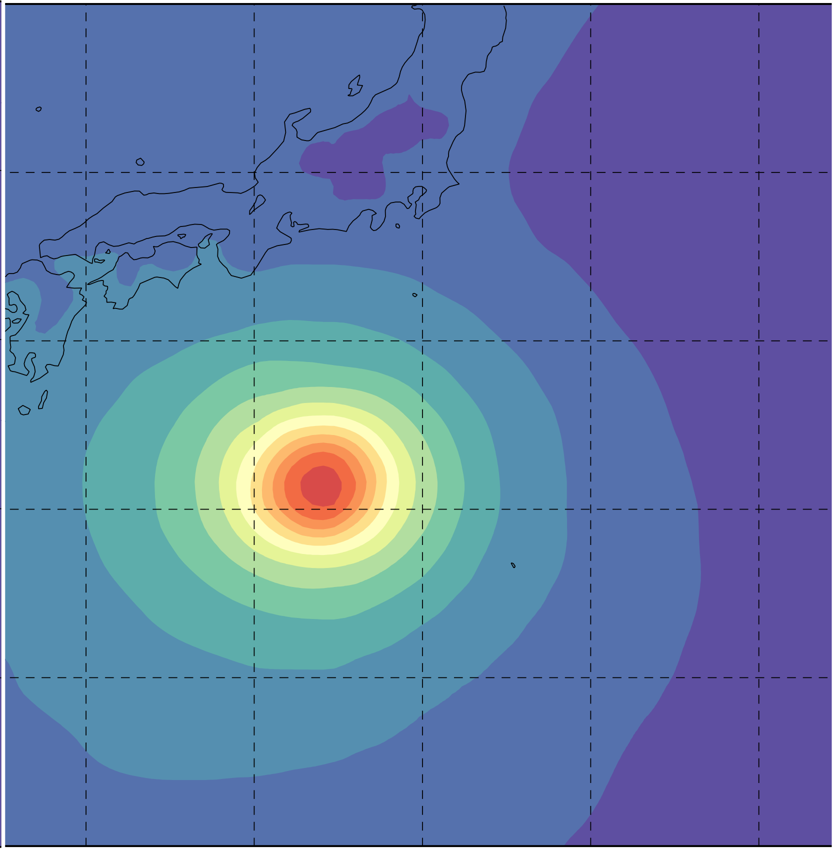
see Sarah-Jane's talk

MSLP

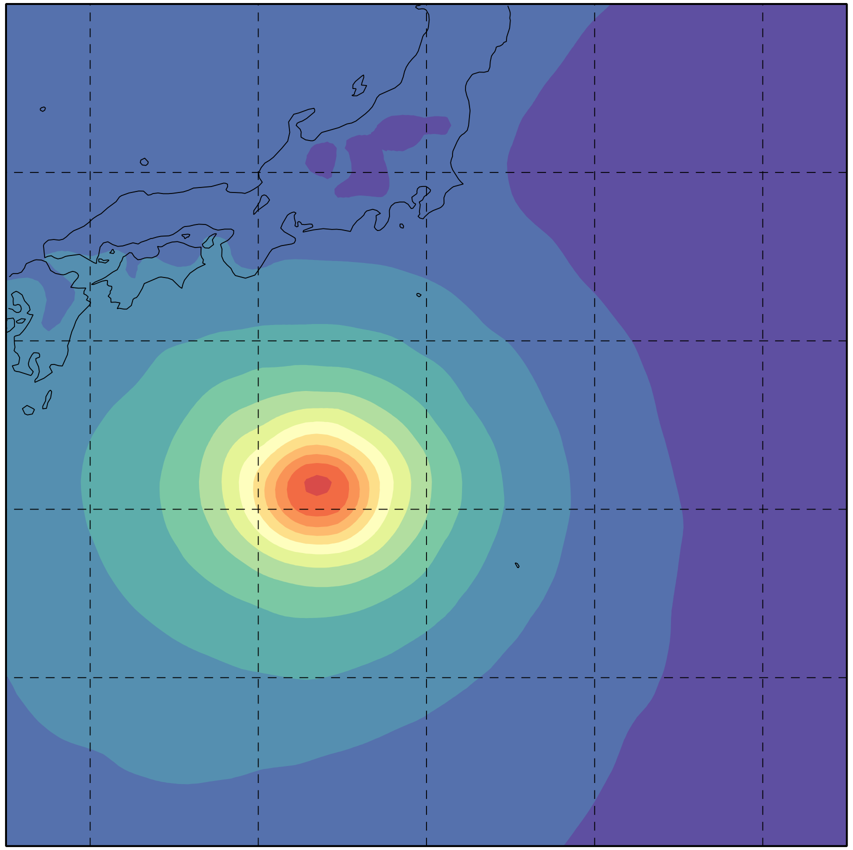
EDA member 2



EDA member 4

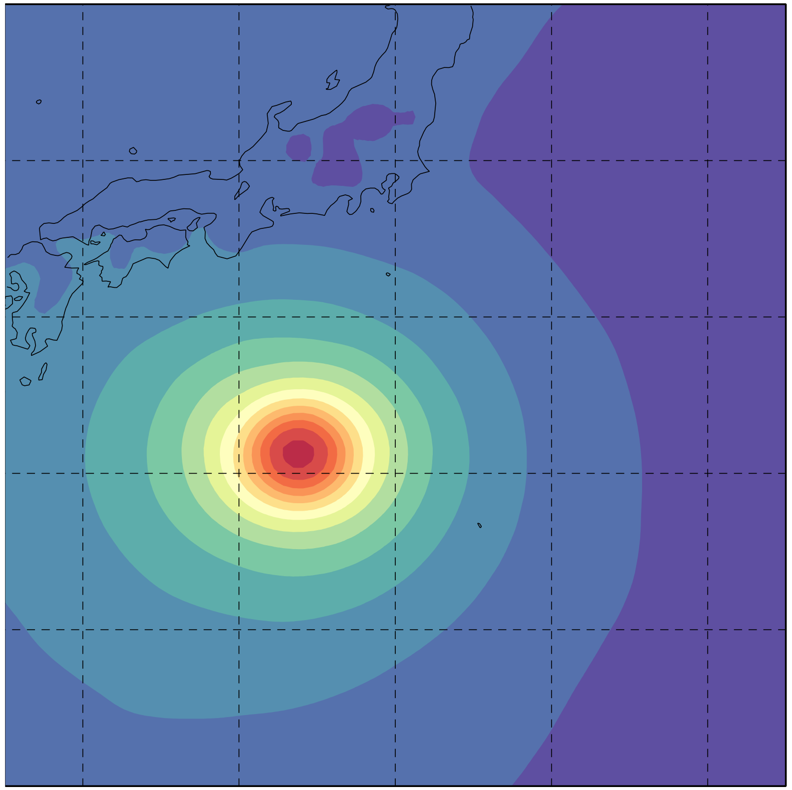


EDA member 22

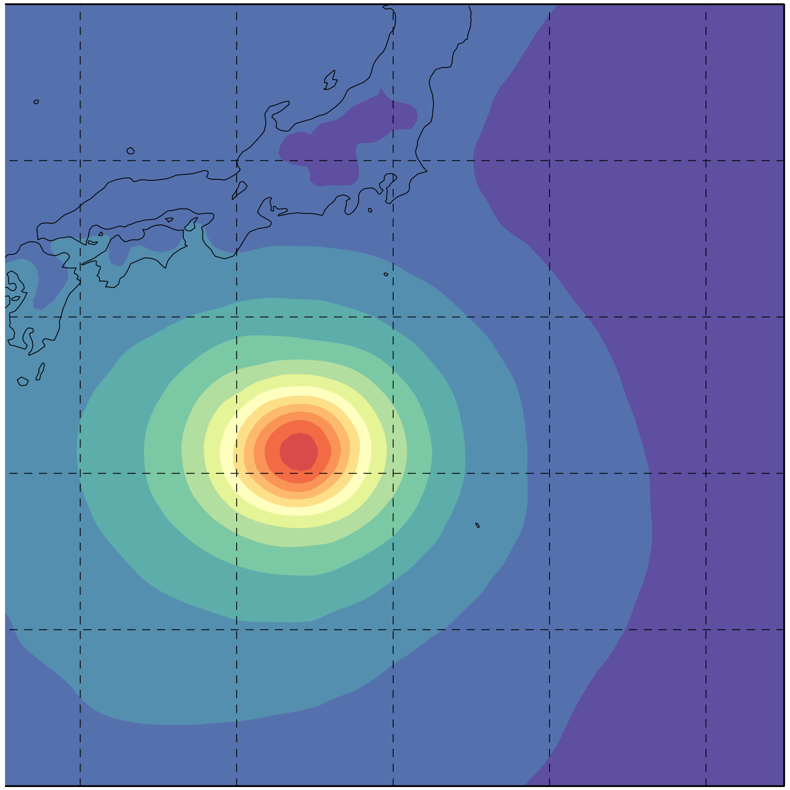


MSLP

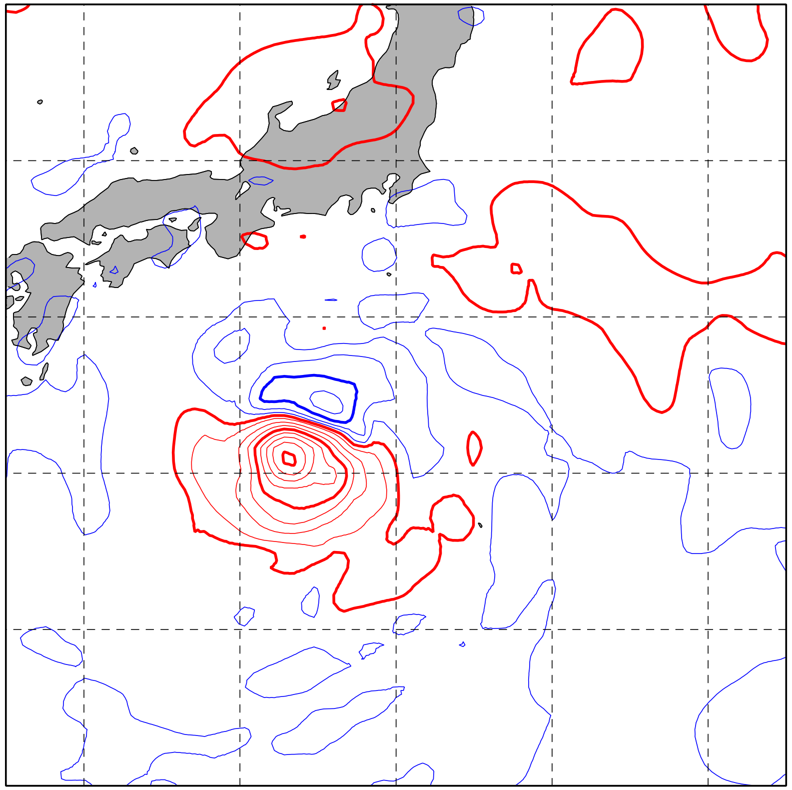
EDA mean



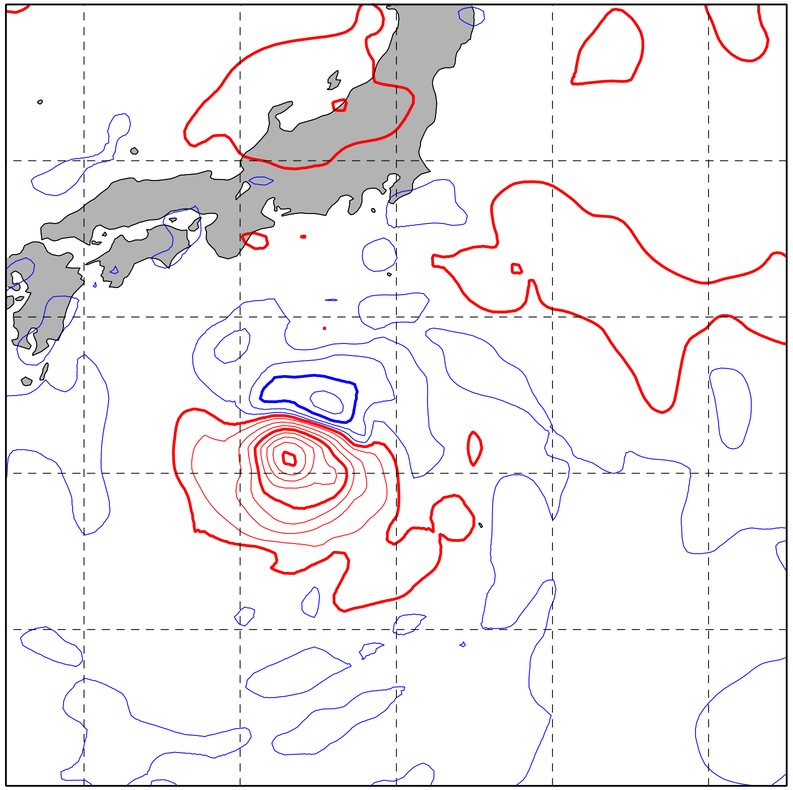
EDA member 2



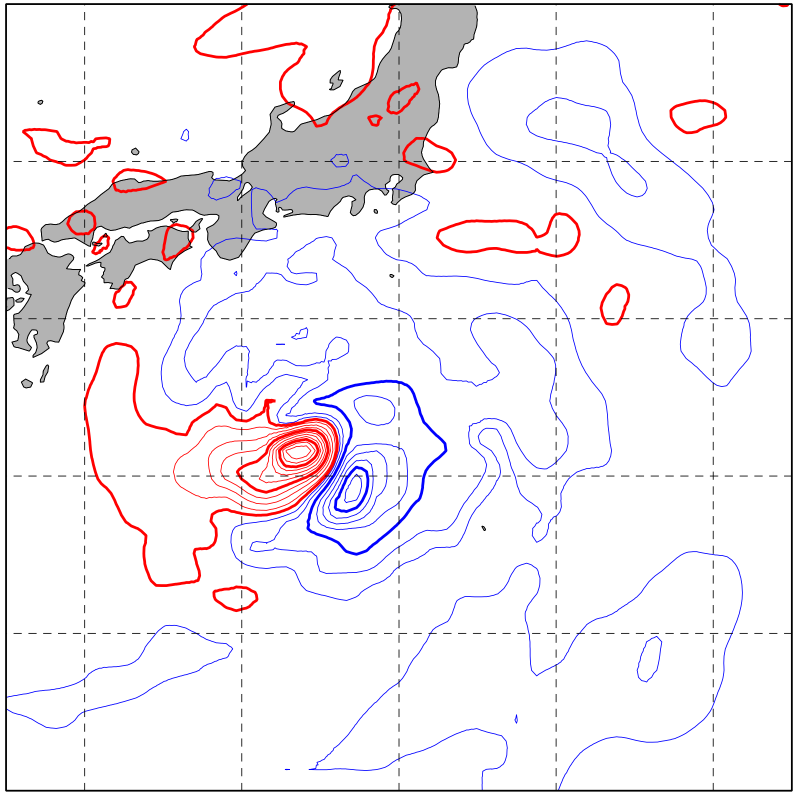
EDA member 2 – EDA mean



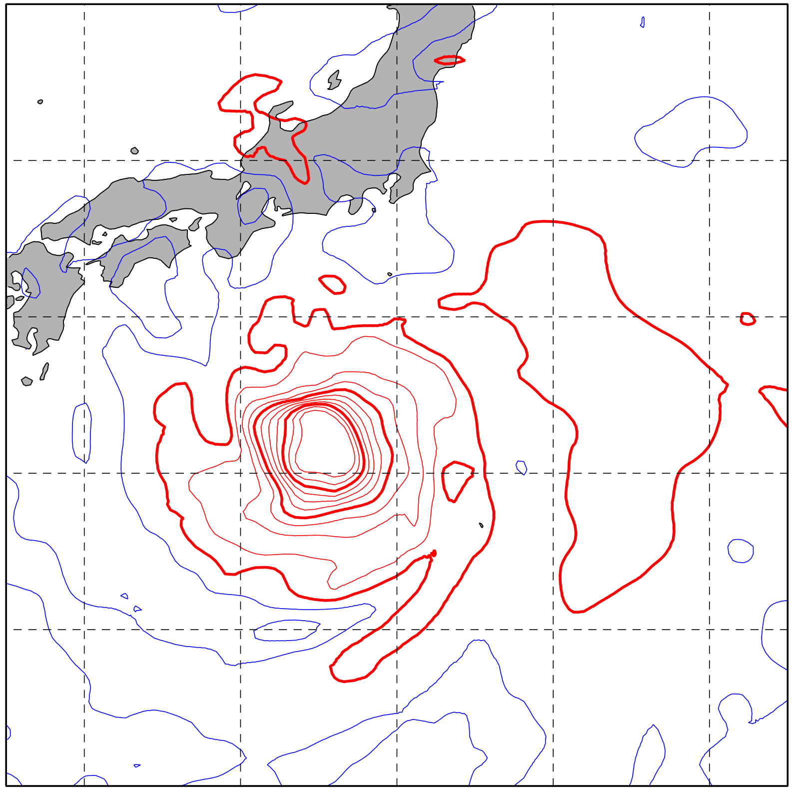
EDA member 2 – EDA mean



EDA member 4 – EDA mean

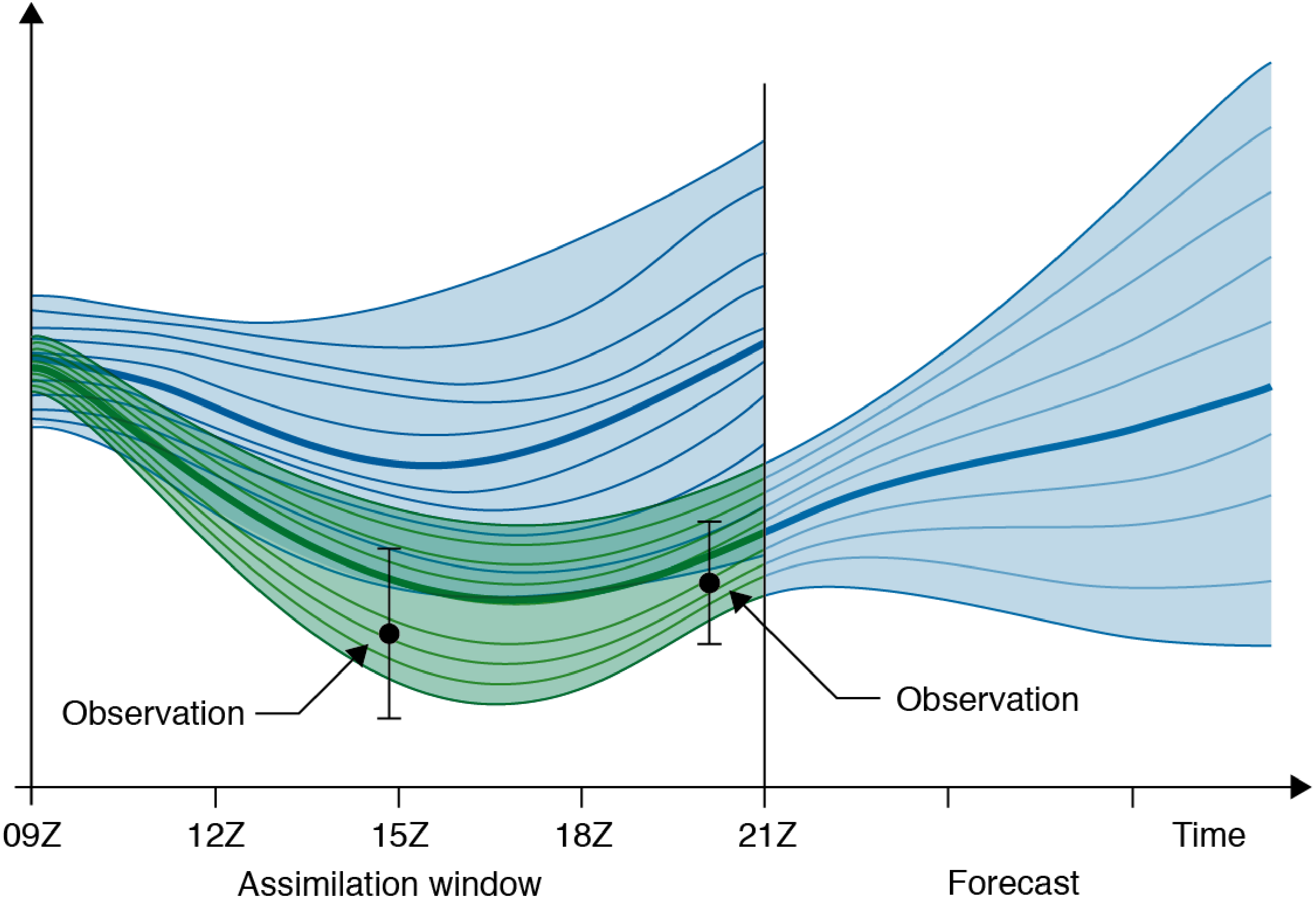


EDA member 22 – EDA mean



Ensemble Data Assimilation

Ensemble Forecast

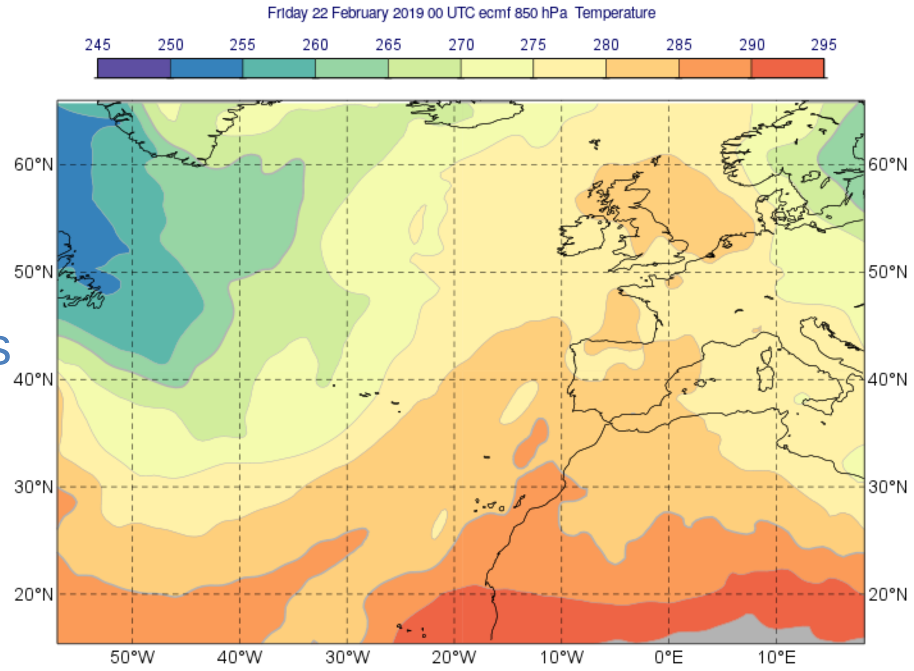


ECMWF

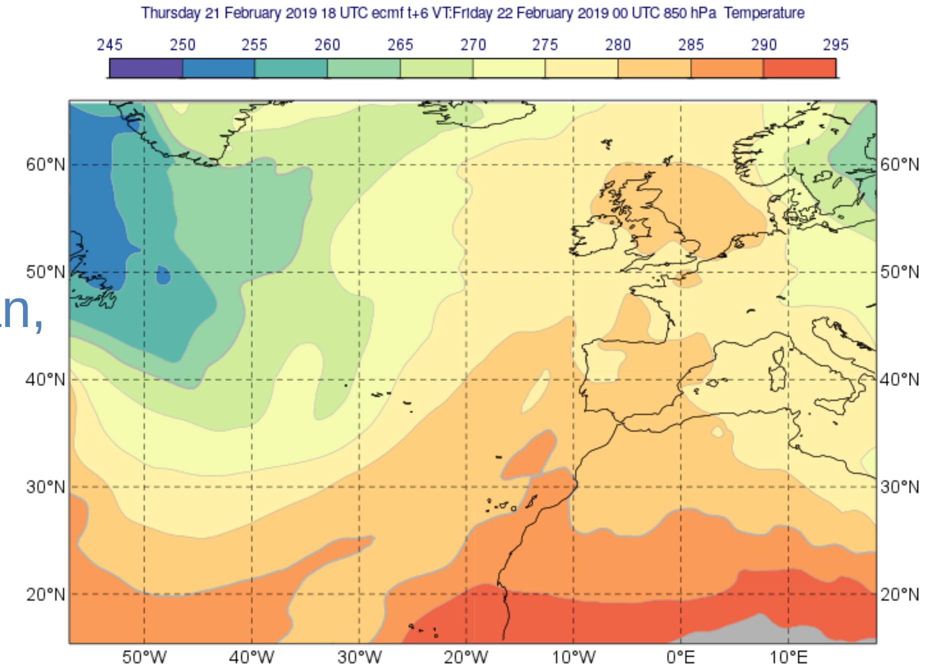
T850hPa

What is available when we start the 00 UTC ensemble forecasts?

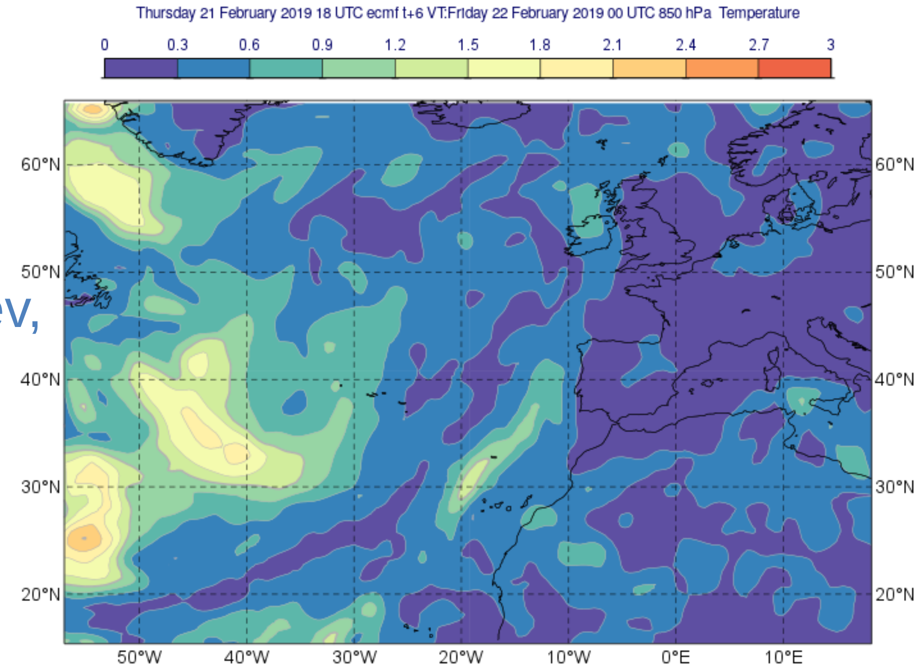
HRES Analysis 00 UTC



EDA Mean, 18 UTC, 6h Fcsts

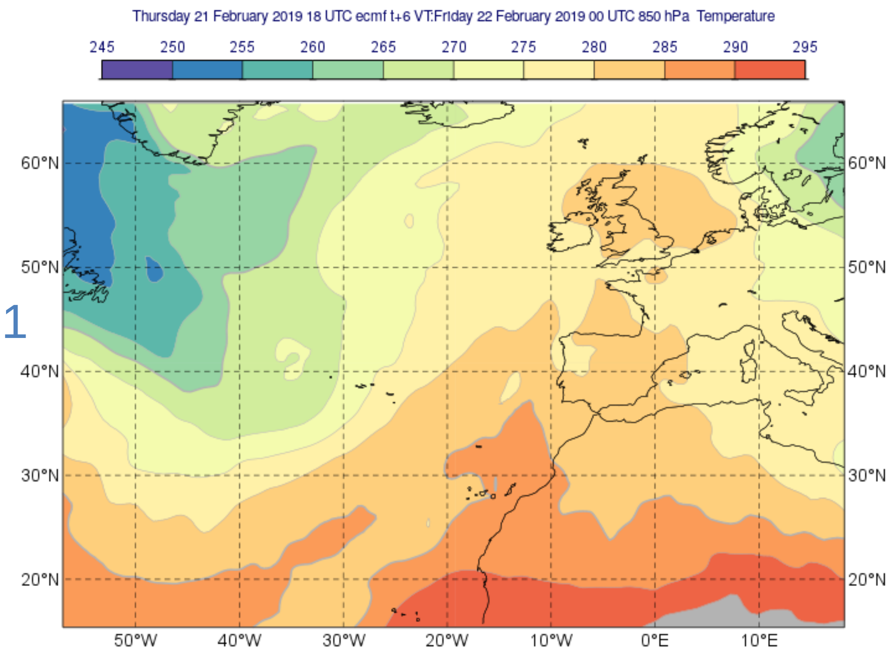


EDA StDev, 18 UTC, 6h Fcsts

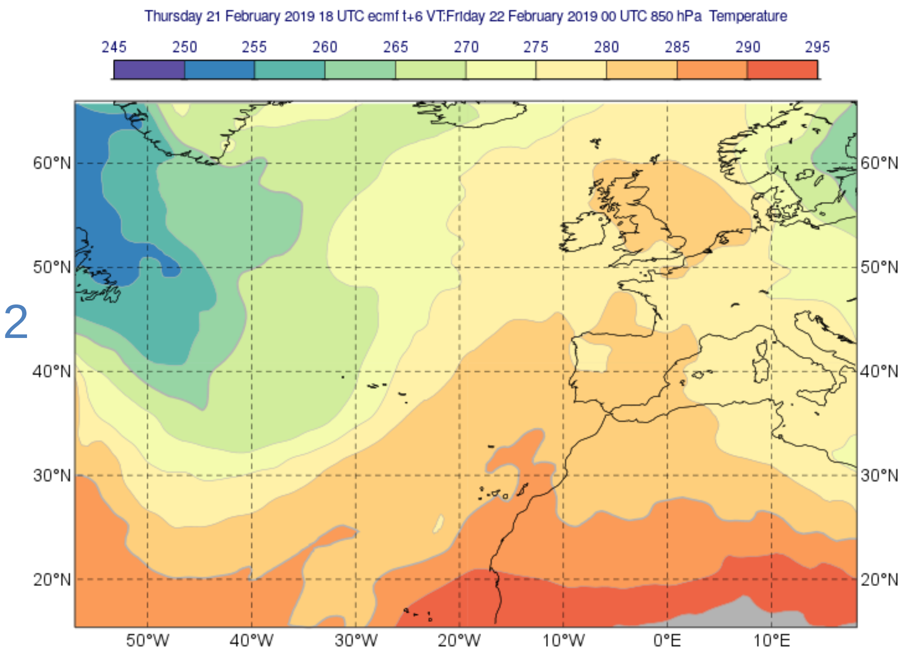


EDA FC, 6h, T850hPa

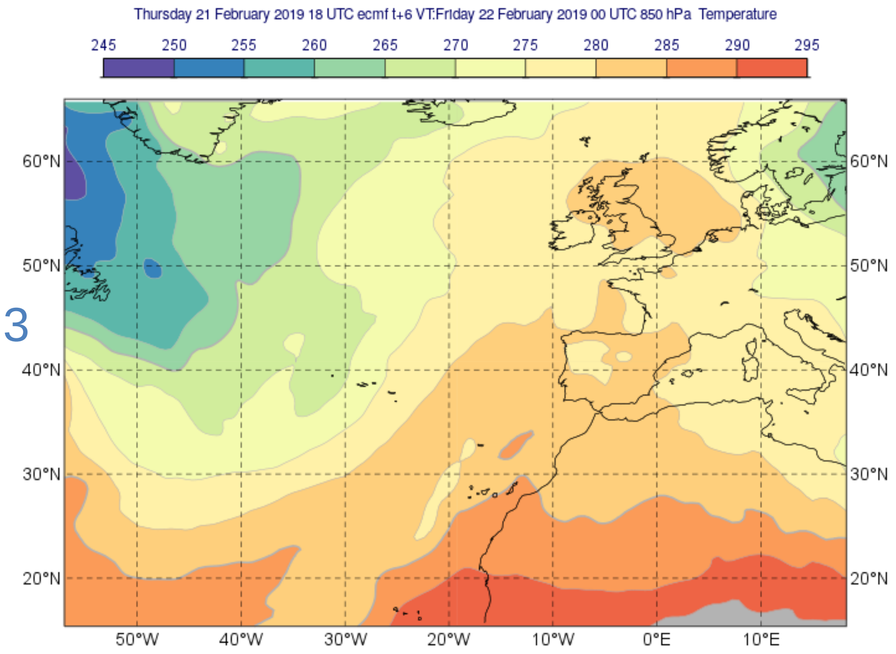
EDA Member 1



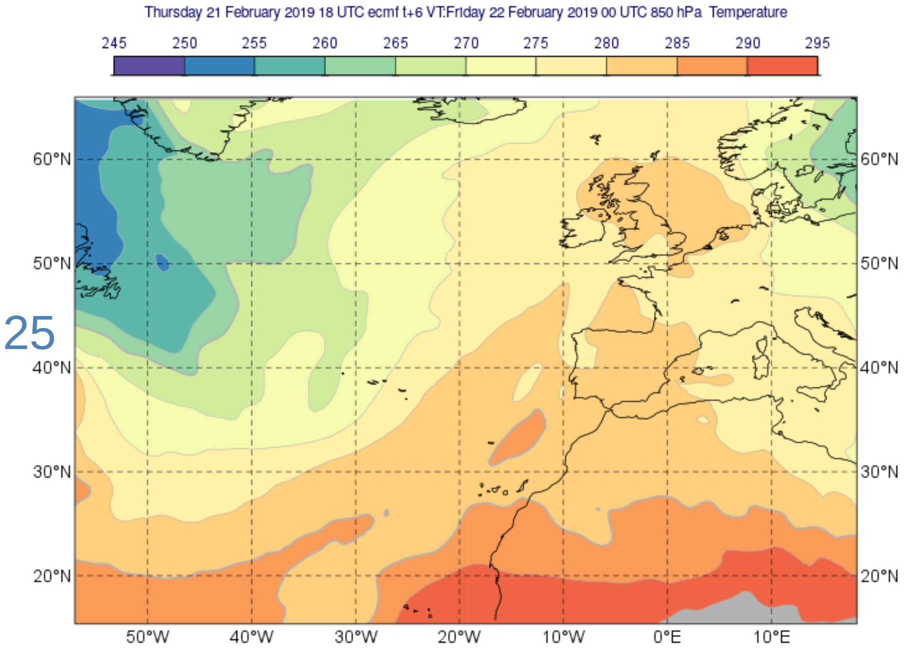
EDA Member 2



EDA Member 3



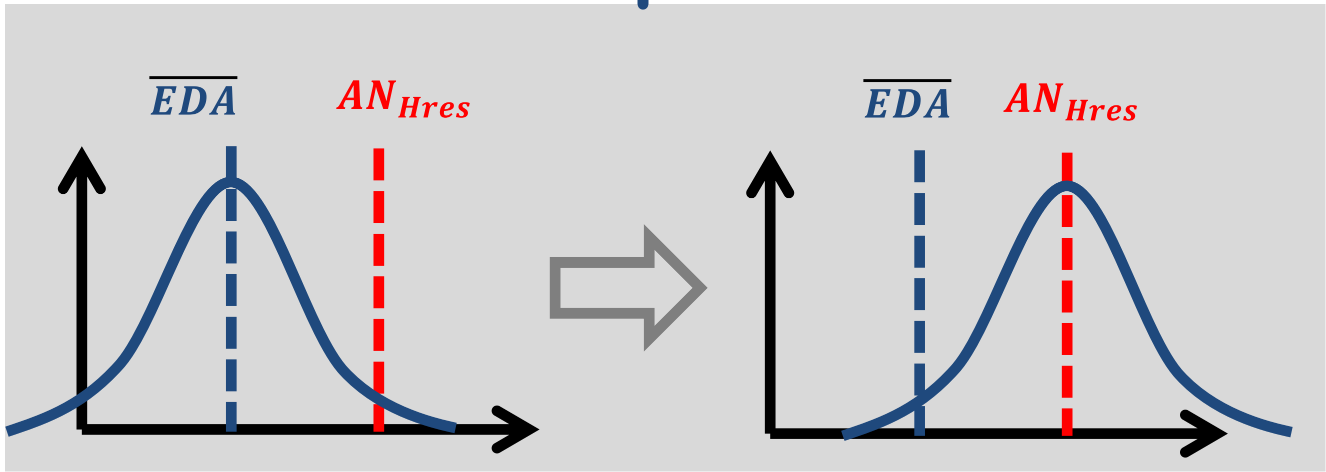
EDA Member 25



Generation of initial conditions for the ensemble forecasts:

$$AN_{pf} = AN_{Hres} + \left(EDA_i - \overline{EDA}\right) + SVPERT_j$$

$i = 1..50$
$j = 1..50$



Re-centre EDA-Distribution on Hres-Analysis

EDA : 6h
Forecasts

$$SVPERT_j = \sum_{l}^{NSET} \sum_{k}^{NSV_l} \alpha_{lk} \, SV_{lk}$$

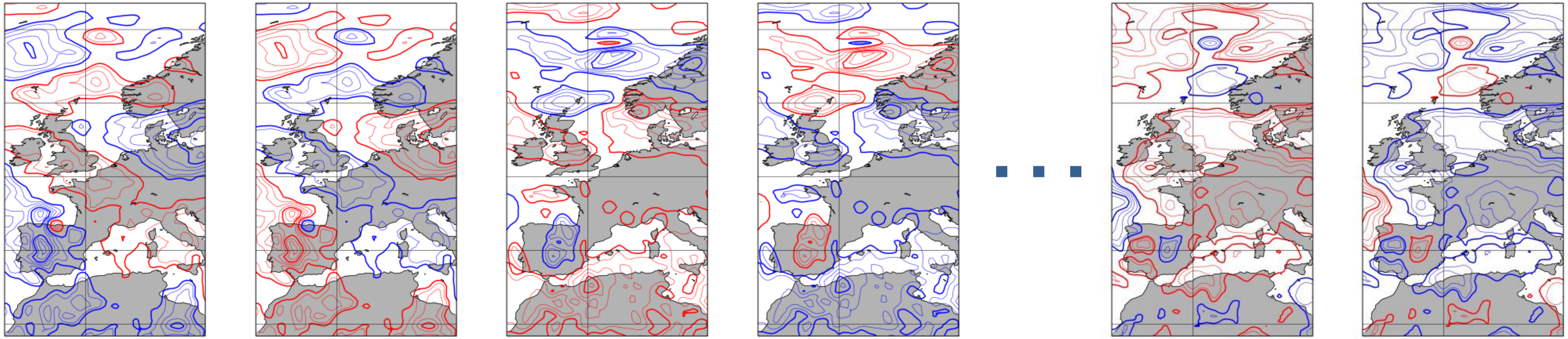
α random number drawn from Truncated gaussian

NSET : nhem, shem, TCs1-6
NSV : 50 for nhem and shem, 5 for TCs

New way to perturb the ensemble initial conditions for 50 Ensemble Members

Old:
Plus-Minus Symmetry with Perturbations from 25-Member EDA

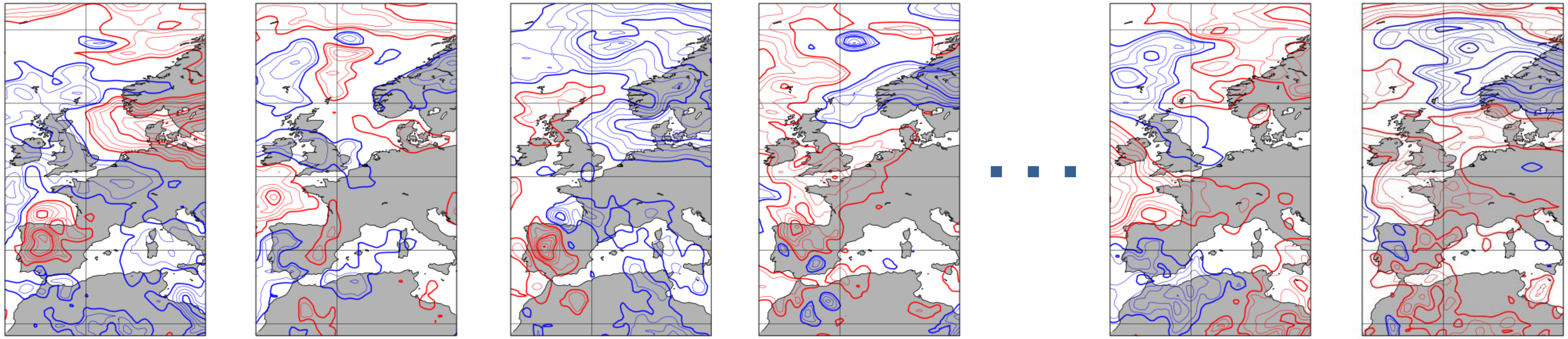
Member 1	Member 2	Member 3	Member 4		Member 49	Member 50
Pert. 1	- Pert. 1	Pert. 2	- Pert. 2		Pert. 25	- Pert. 25



z500hPa

New:
Perturbations from new 50-Member EDA

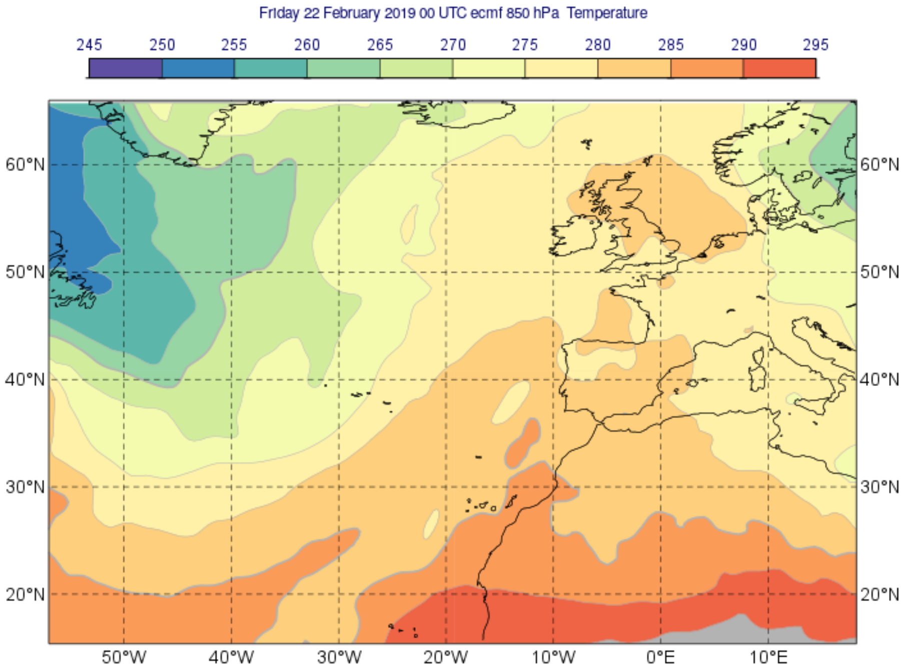
Member 1	Member 2	Member 3	Member 4		Member 49	Member 50
Pert. 1	Pert. 2	Pert. 3	Pert. 4		Pert. 49	Pert. 50



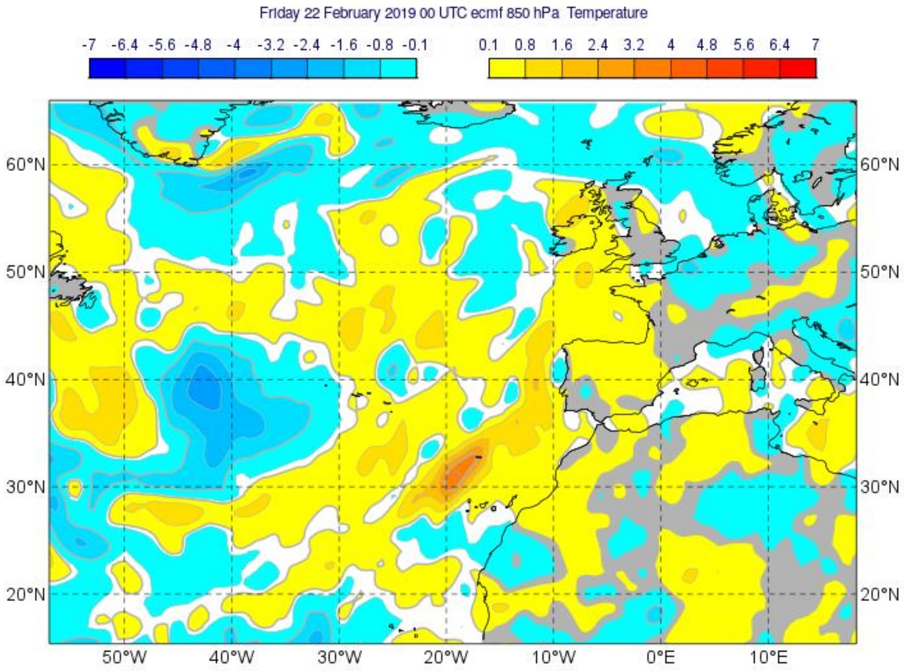
See Lang et al. 2019, ECMWF Newsletter No. 158

T850hPa

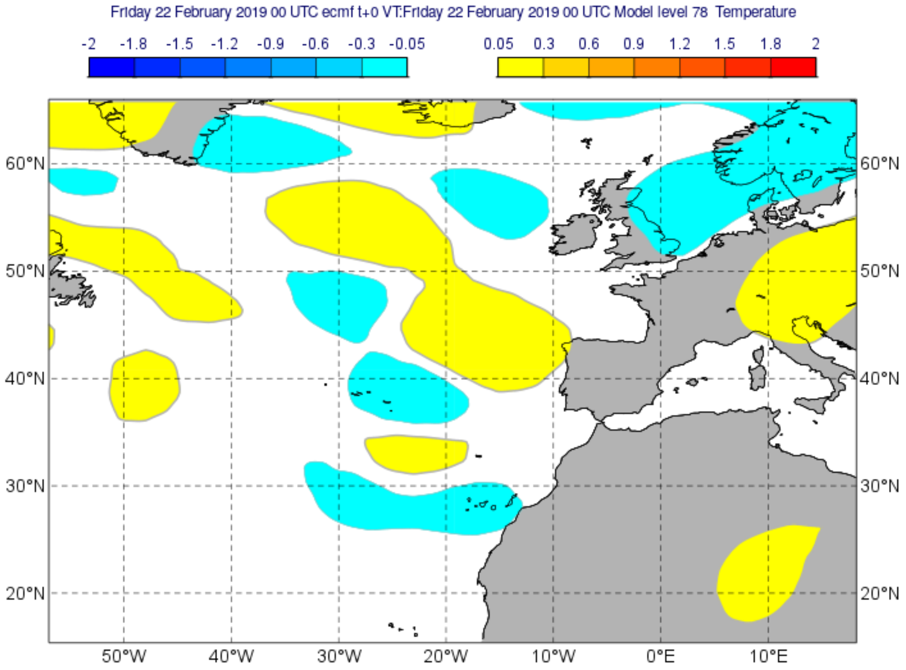
HRES Analysis 00 UTC



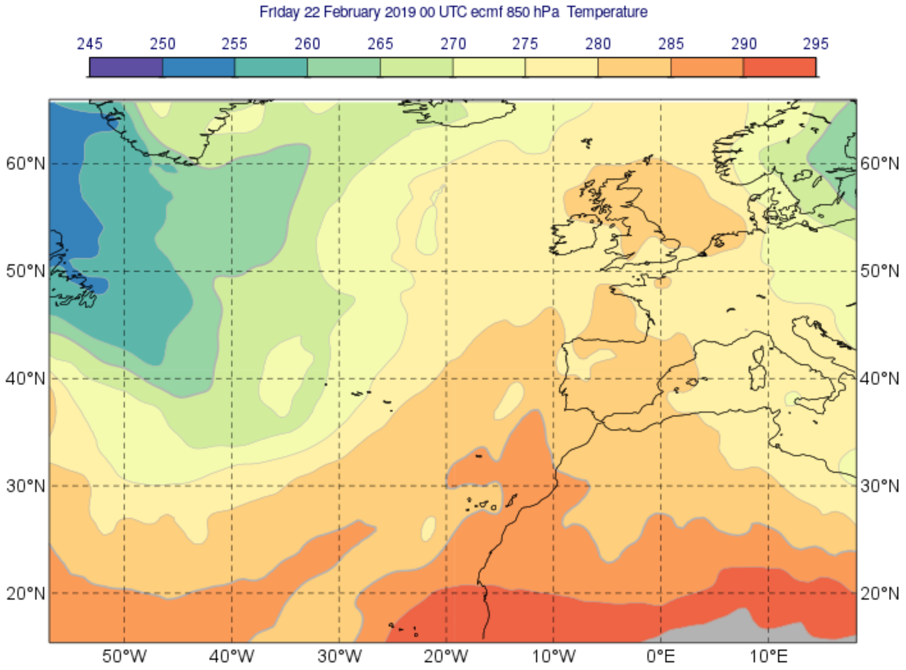
+ EDA-Pert 1



+ SV-Pert 1

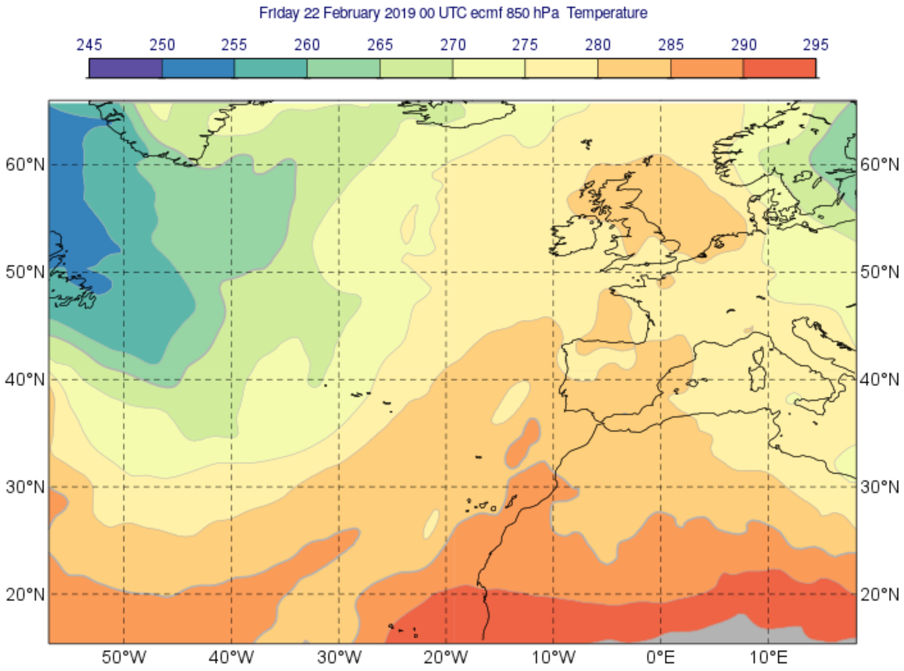


= Initial conditions for ENS member 1



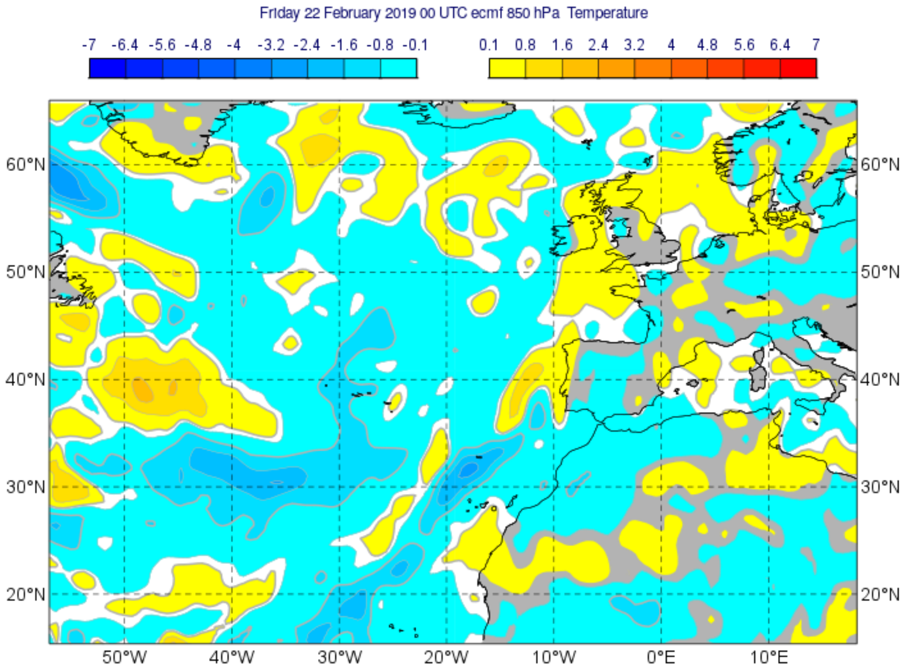
T850hPa

HRES Analysis 00 UTC



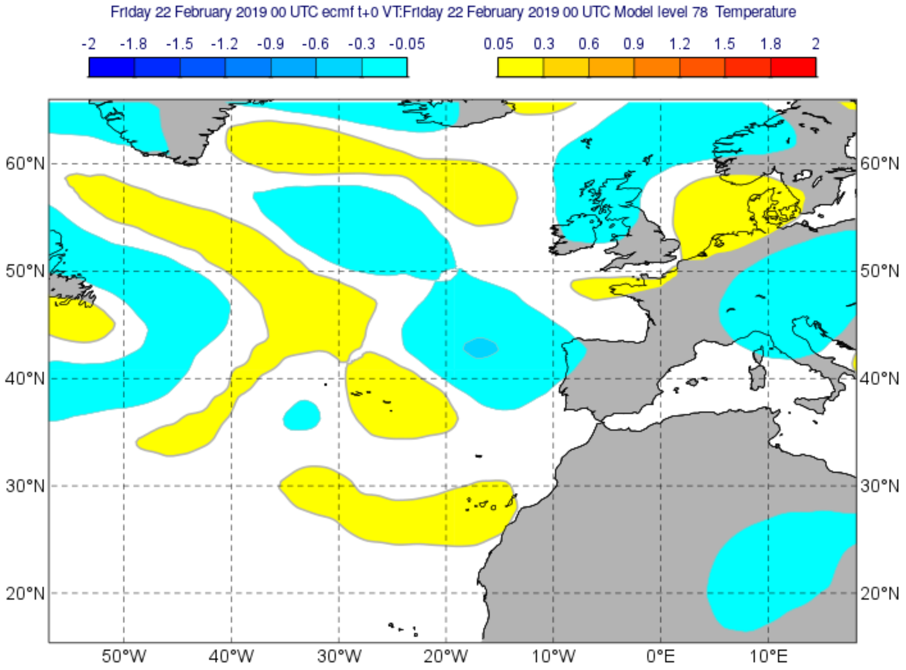
+

EDA-Pert 3



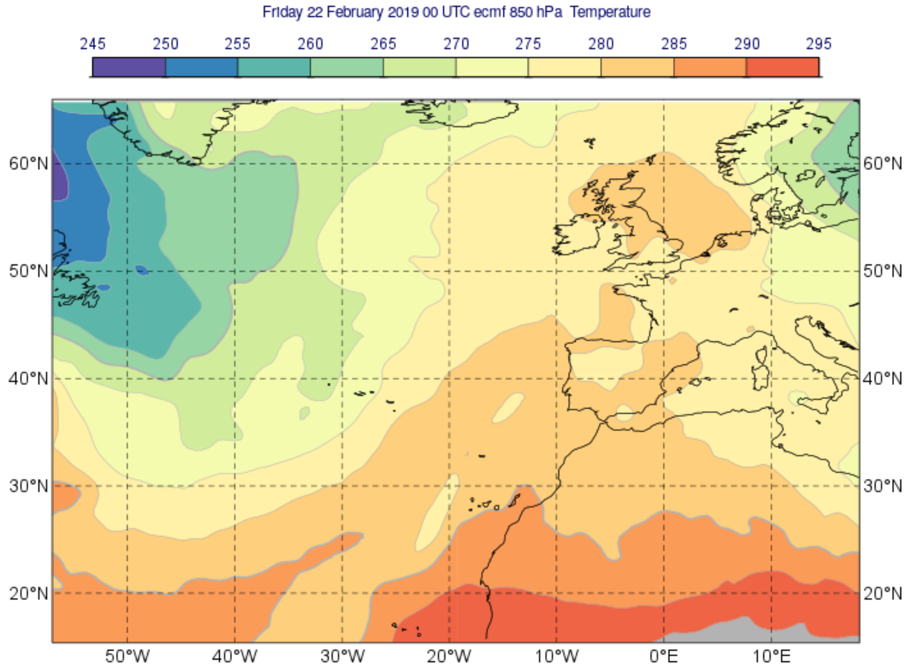
+

SV-Pert 3



=

Initial conditions for ENS member 3

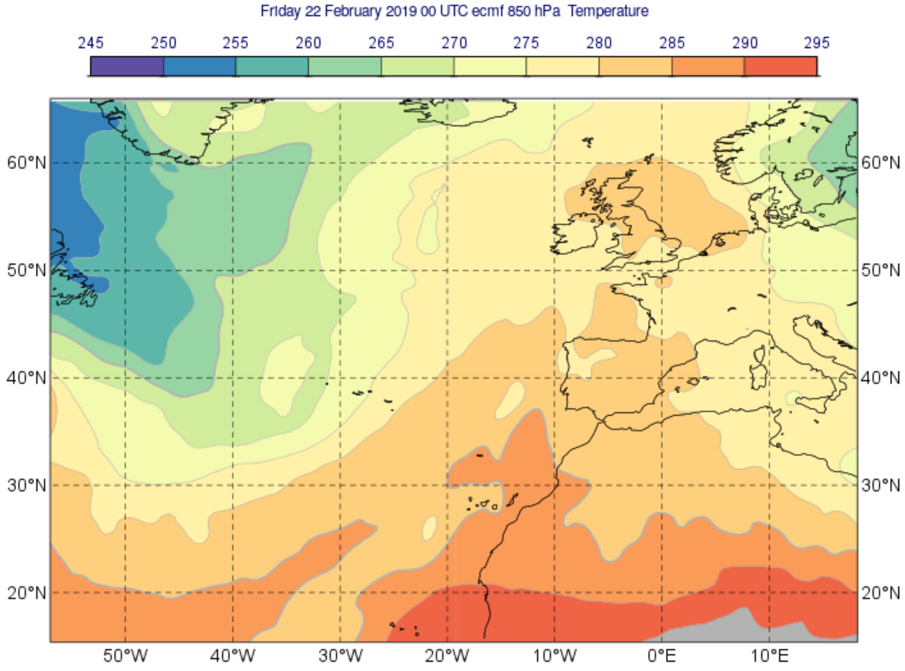


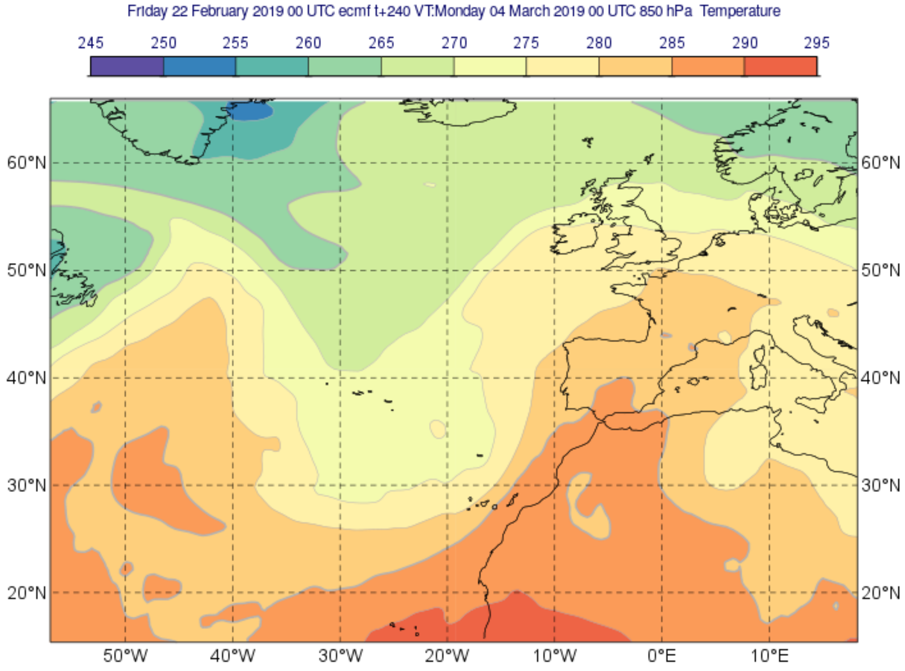
T850hPa

T+0h

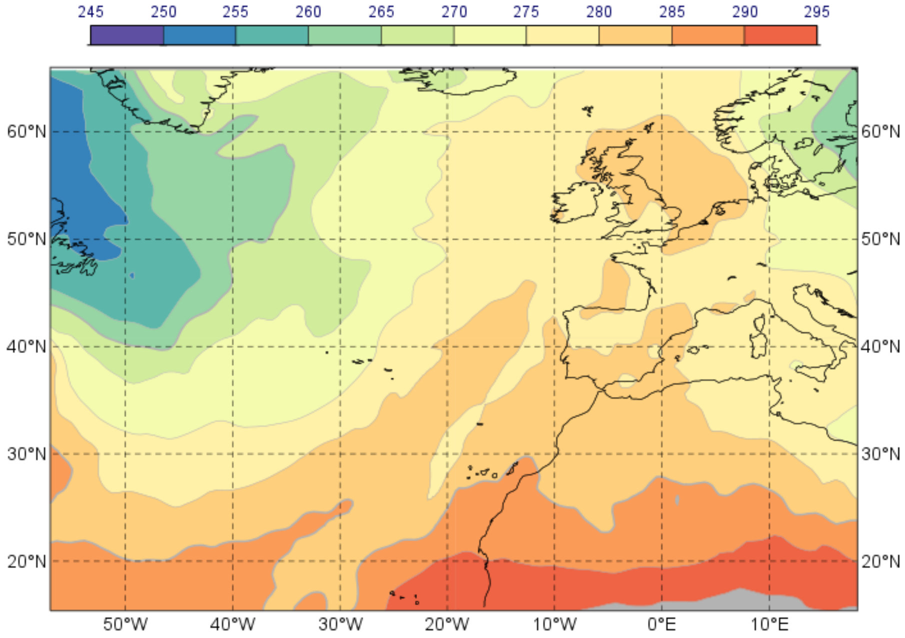
T+240h

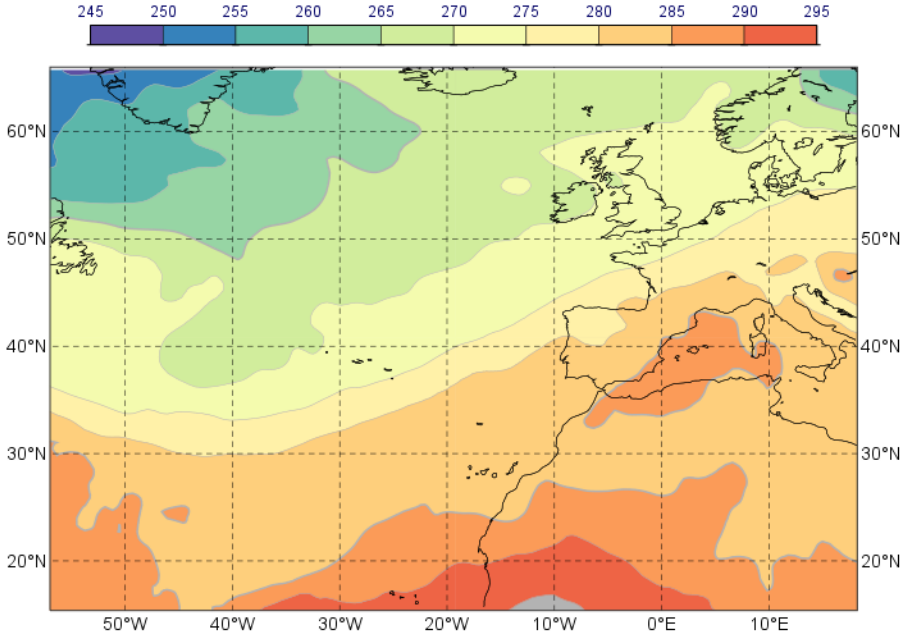
Member 1:





Member 2:





Ocean initial state:

50 Members + 1 Control, 5 Ocean analyses

Member	Ocean analysis
Control	1
Member 1	2
Member 2	3
Member 3	4
Member 5	5
Member 6	1
Member 7	2
…	
Member 50	1

Reliability of the ensemble spread

- Consider ensemble variance ("spread") for an M-member ensemble

$$\frac{1}{M}\sum_{j=1}^{M}(x_j - \overline{x})^2$$

and the squared error of the ensemble mean

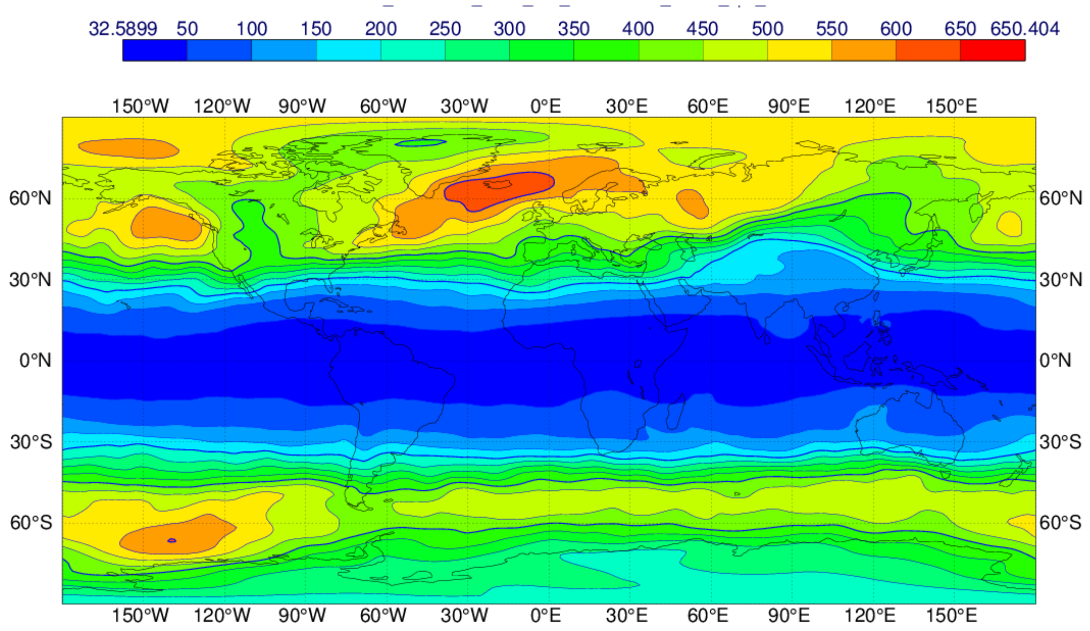
$$(\overline{x} - y)^2$$

- Average the two quantities for many locations and/or start times.
- The averaged quantities have to match for a reliable ensemble (within sampling uncertainty).

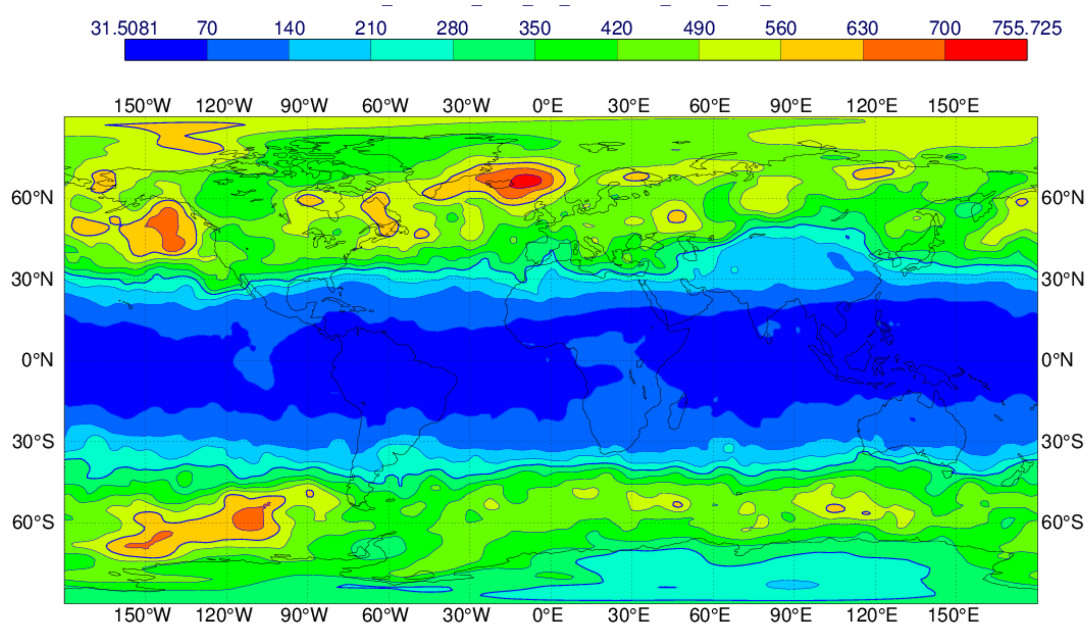
From Martin Leutbecher's lecture "Ensemble Verification 1"

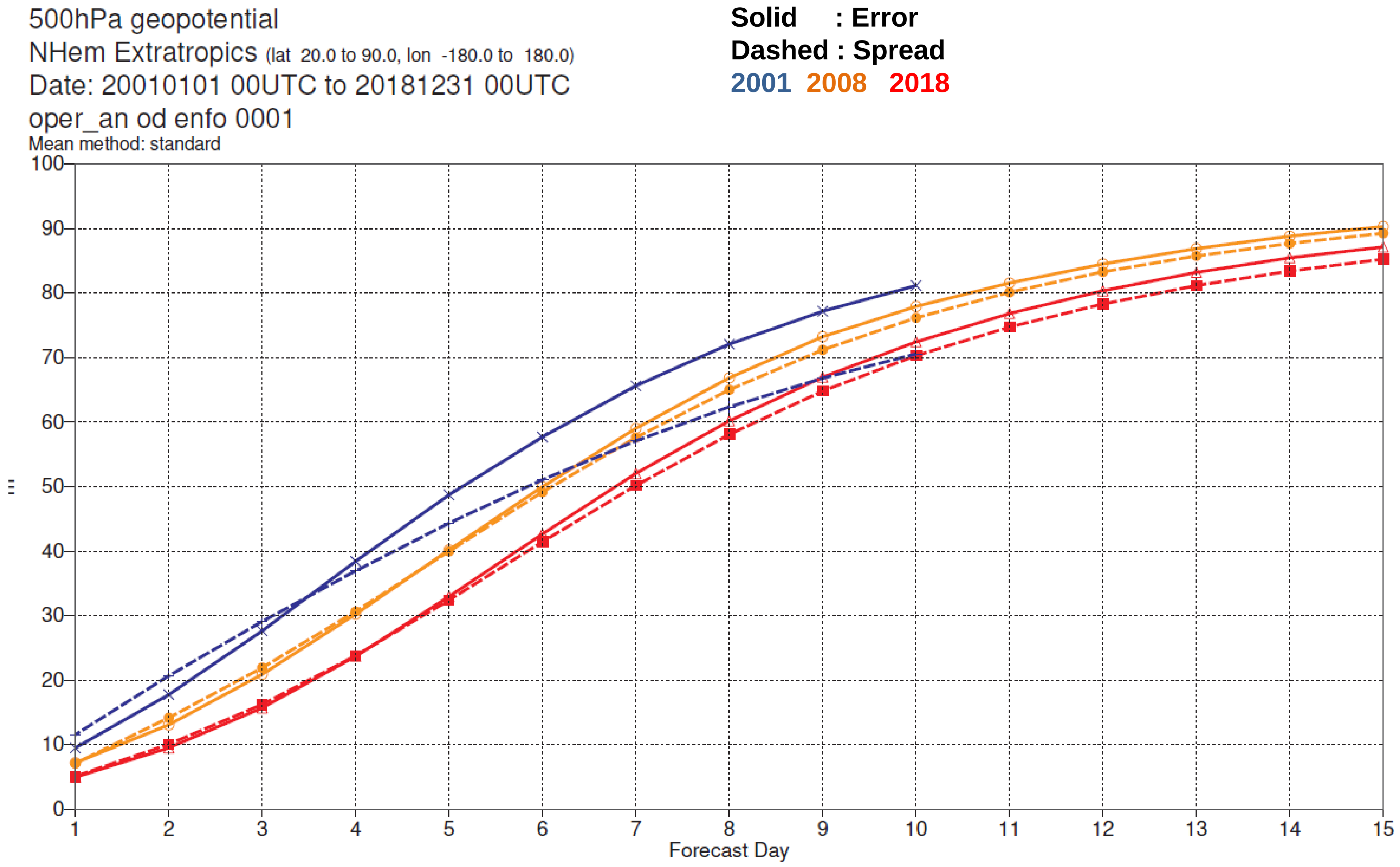
Z500hPa
Ensemble StDev and Ensemble Mean RMSE, averaged 2016112200 – 2017021300
00 UTC Run

StDev T+120h



RMSE T+120h

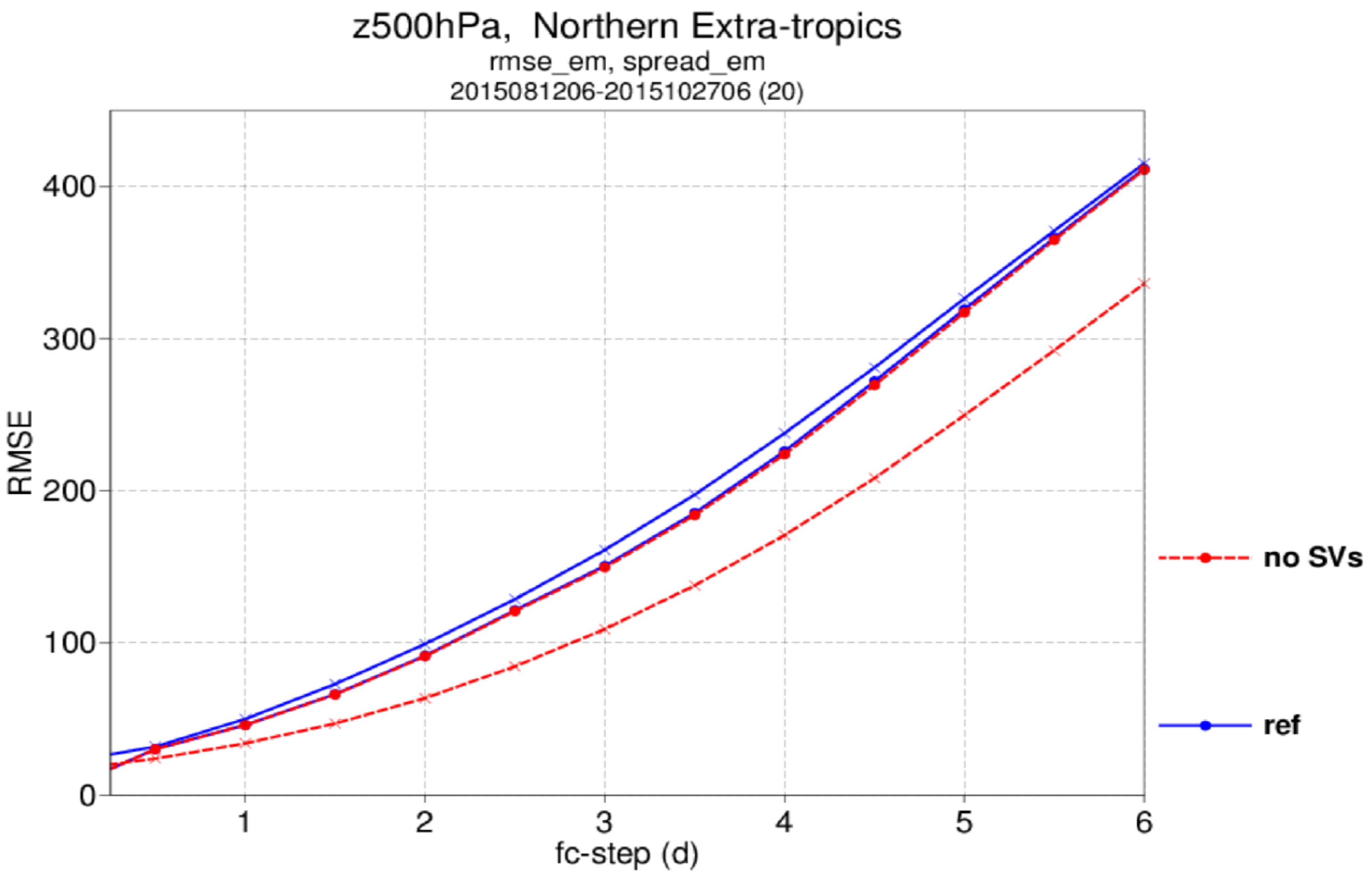




ECMWF EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

Why SVs?

Impact of SVs on ENS

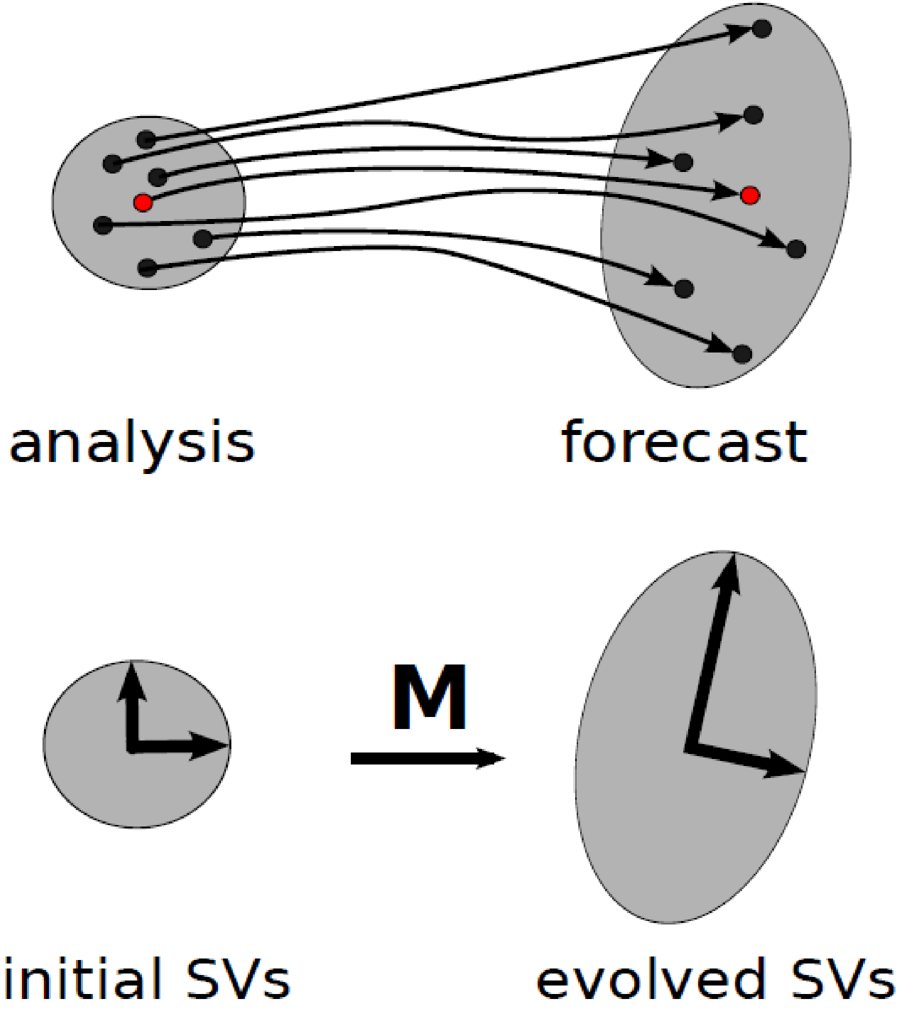


Oper like setup, TCo399, 20 Initial dates

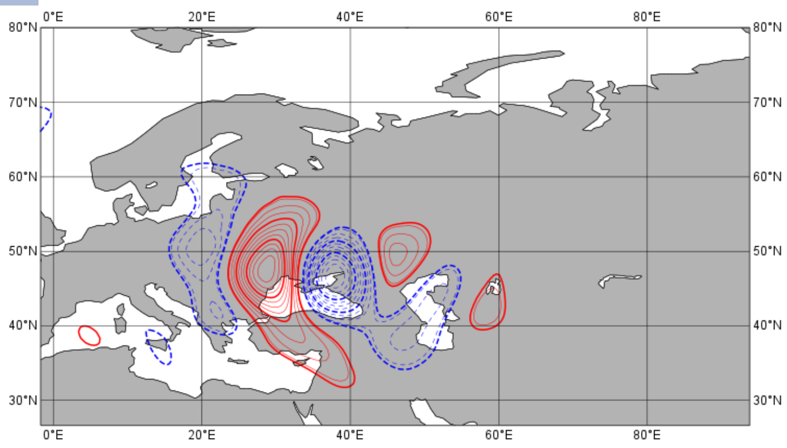
Singular Vector Perturbations

Directions of fastest growth over a finite time interval (optimisation interval)

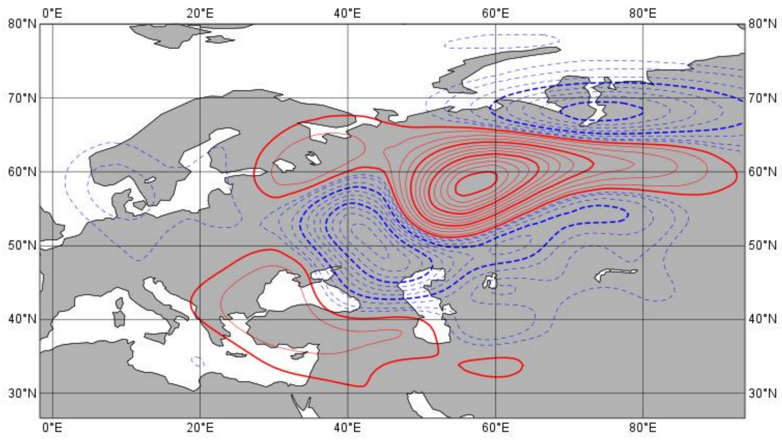
Justification: EDA + Model Uncertainty representation produce substantial spread in the directions of the leading SVs but ensemble still under dispersive (Leutbecher and Lang, 2014, QJRM)



Initial SV, T mlevel 68



evolved SV, T mlevel 49



Singular vectors are computed by solving an eigenvalue problem (e.g. Leutbecher and Palmer, 2008):

$$\boldsymbol{C_0^{-1/2} M^* P^* C_1 P M C_0^{-1/2} v = \sigma^2 v}$$

- C_0 and C_1 initial and final time metrics
- $M(0,t)$ linear propagator from time 0 to t and its adjoint M^*
- P and P^* local projection operator and its adjoint

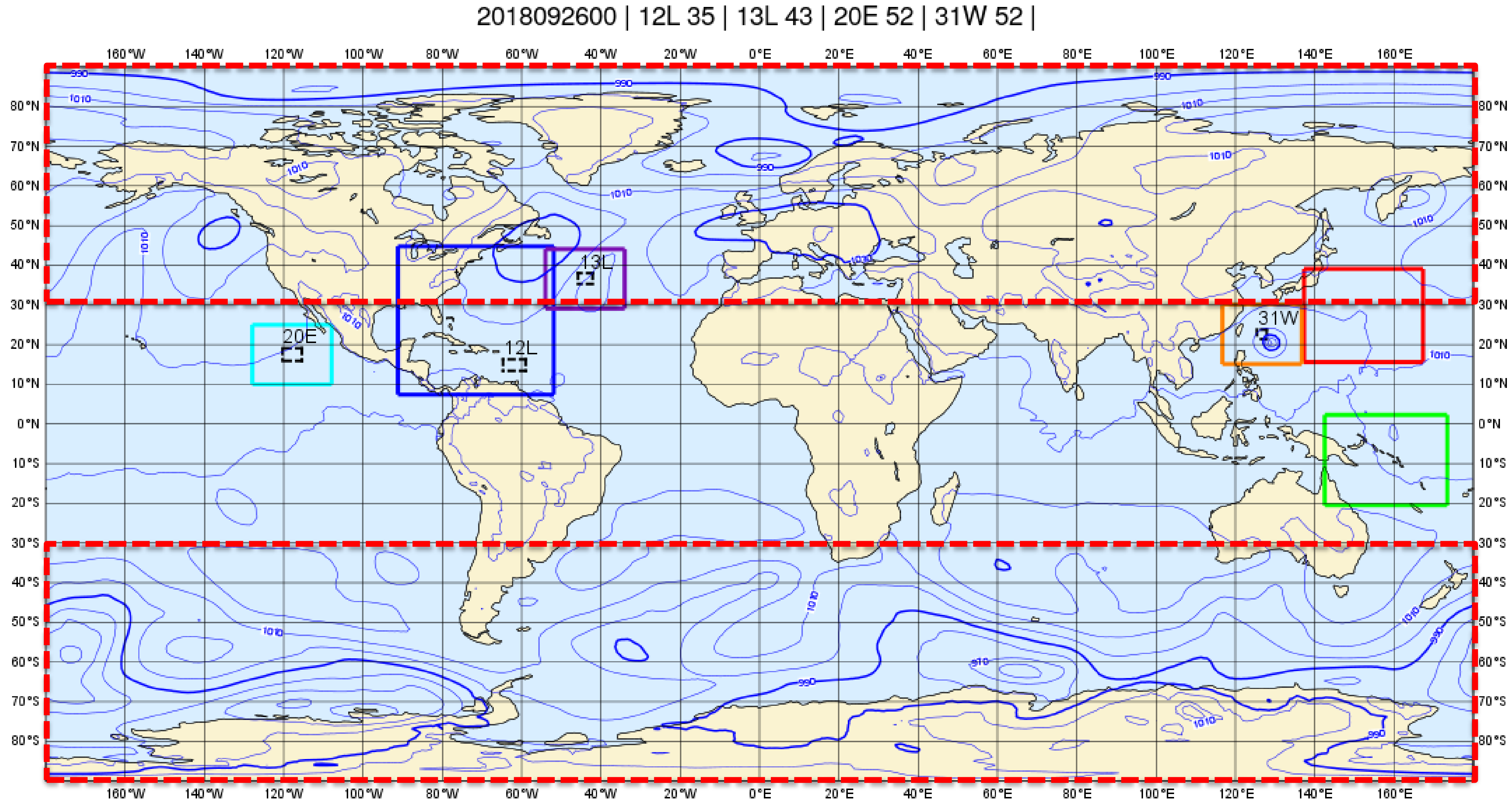
$$\frac{1}{2}\int_{p_0}^{p_1}\int_S \left(u^2+v^2+\frac{c_p}{T_r}T^2\right)\mathrm{d}p\,\mathrm{d}s+\frac{1}{2}R_d T_r p_r \int_S (\ln p_{\mathrm{sfc}})^2\,\mathrm{d}s$$

Singular vectors in the operational EPS

- $t_{\text{opt}} \equiv t_1 - t_0 = 48$ h
- resolution: T42L137
- **Extra-tropics:** 50 SVs for N.-Hem. (30°N–90°N) + 50 SVs for S.-Hem.(30°S–90°S). Tangent-linear model with vertical diffusion and surface friction only.
- **Tropical cyclones:** 5 singular vectors per region targeted on active tropical depressions/cyclones. Up to 6 such regions. Tangent-linear model with representation of diabatic processes (large-scale condensation, convection, radiation, gravity-wave drag, vert. diff. and surface friction).
- **Localisation** is required to avoid that too many leading singular vectors are located in the dynamically more active winter hemisphere (Buizza 1994). Also required to obtain (more slowly growing) perturbations associated with tropical cyclones (Puri et al. 2001). In order to optimise perturbations for a specific region simply replace the propagator **M** in the equations by **PM**, where **P** denotes the projection operator which sets the state vector ($T, u, v, \ln p_{\text{sfc}}$ in grid-point space) to zero outside the region of interest and is the identity inside it.

Now up to 12 target regions!

SV Target Areas

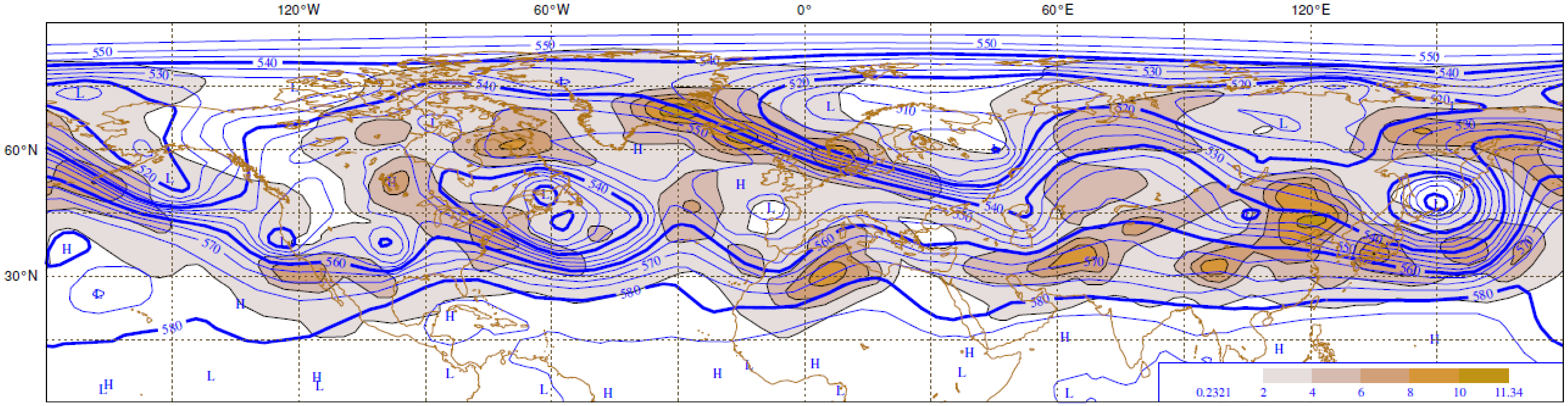


Regional distribution of Northern Hem. SVs

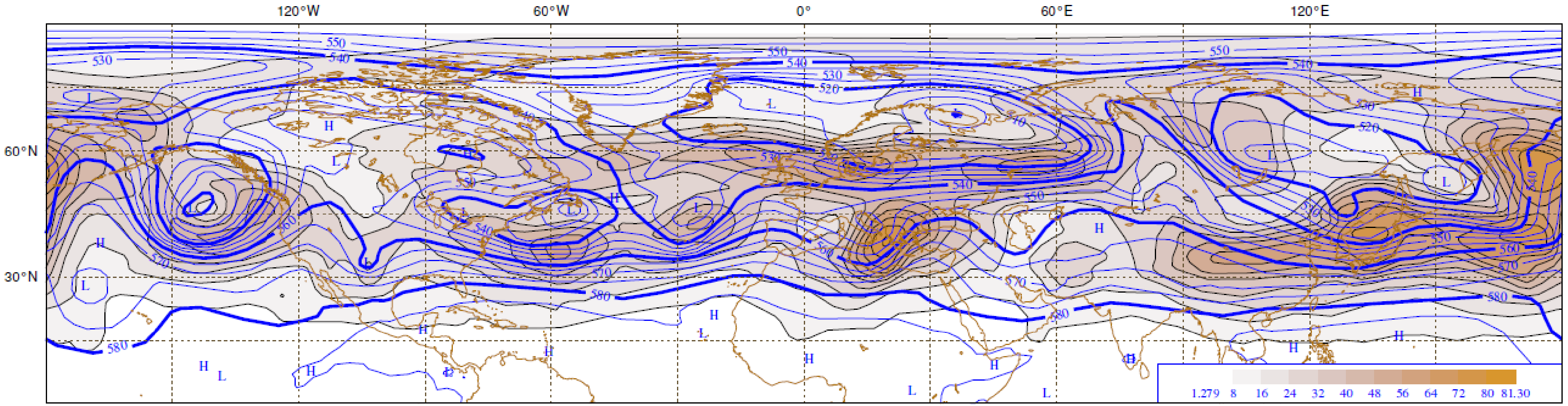
square root of vertically integrated total energy of SV 1–50 (shading)

500 hPa geopotential (contours)

initial singular vectors, 21 March 2006, 00 UTC



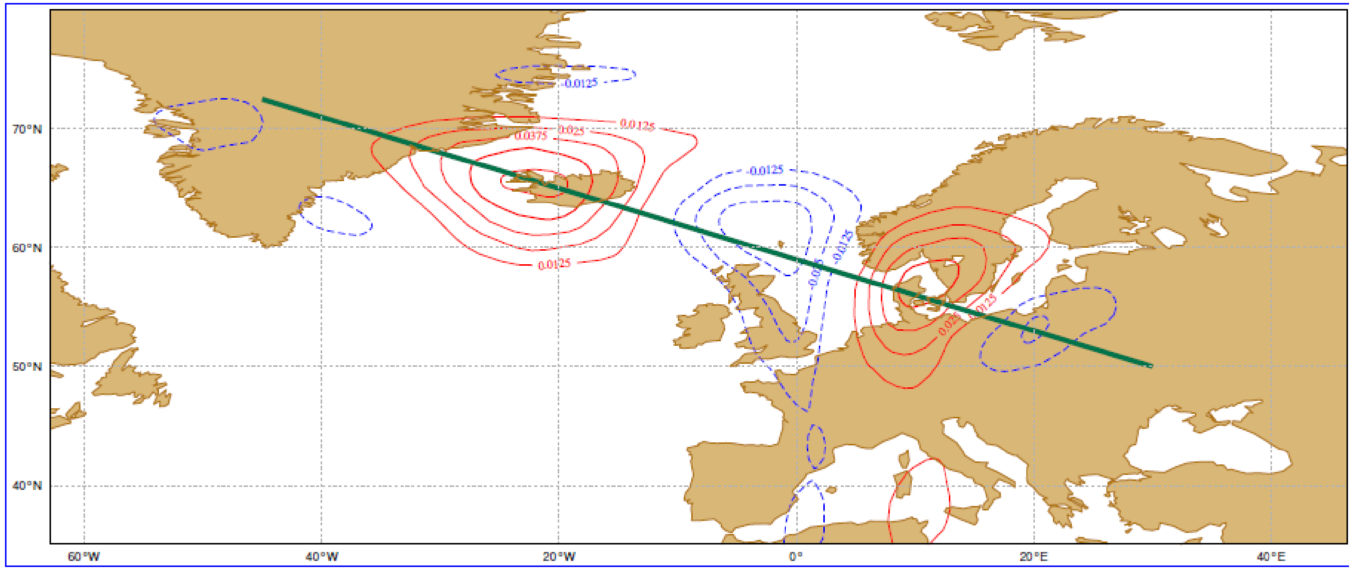
evolved singular vectors, 23 March 2006, 00 UTC



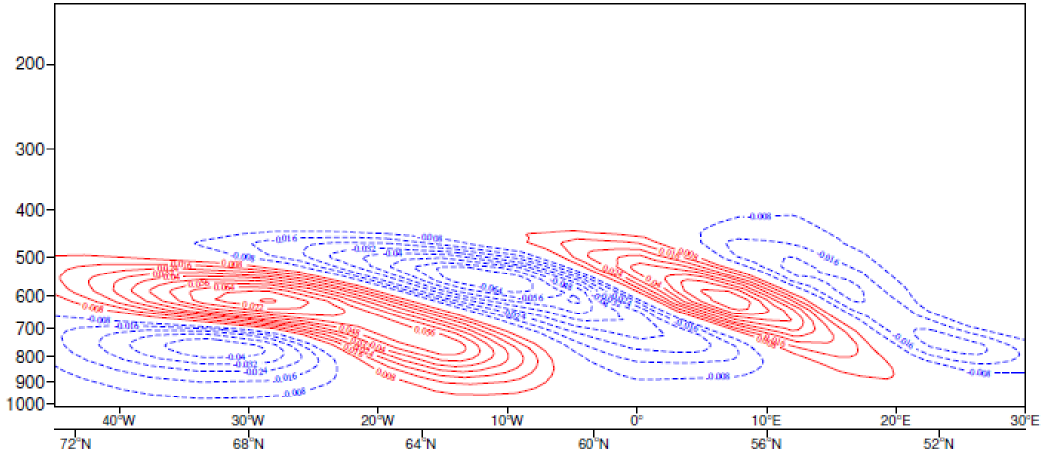
Singular vector 5: initial time

21 March 2006, 00 UTC

Temperature at ≈ 700 hPa



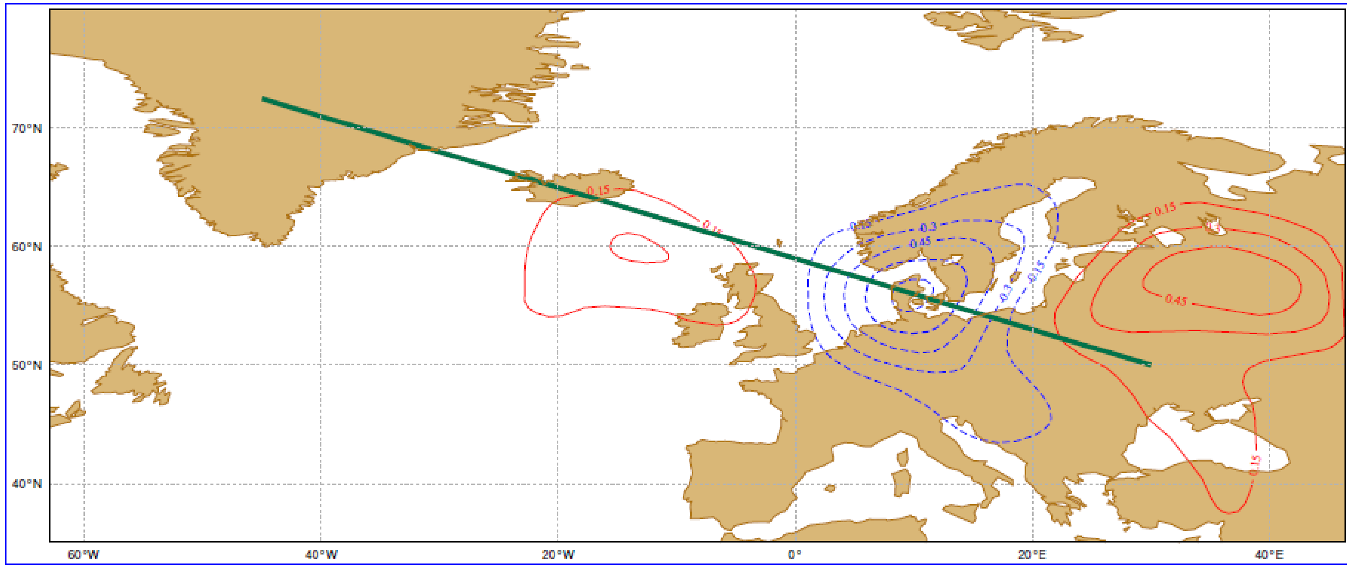
Cross section of temp 20060321 00 step 0 Expver 0001



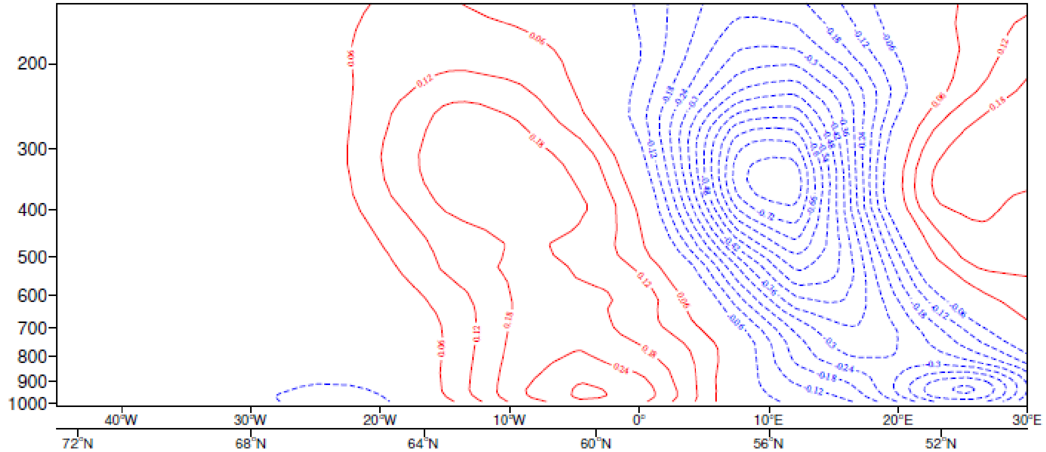
Singular vector 5: final time

23 March 2006, 00 UTC

meridional wind component at ≈ 300 hPa



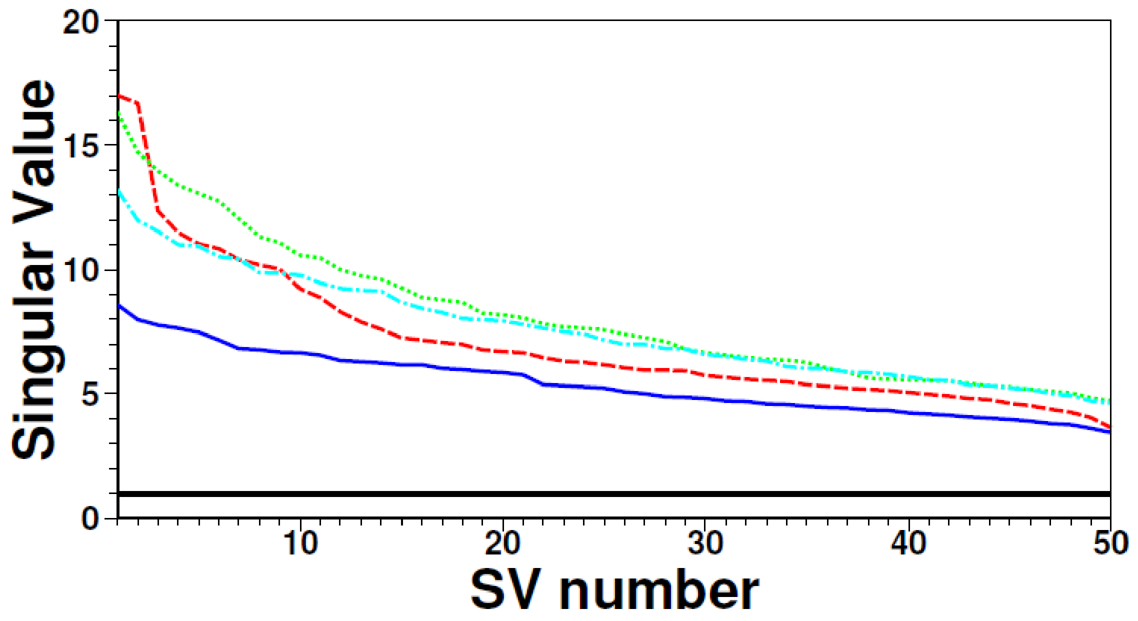
Cross section of v-vel 20060321 00 step 48 Expver 0001



Singular values σ_j — extra-tropics

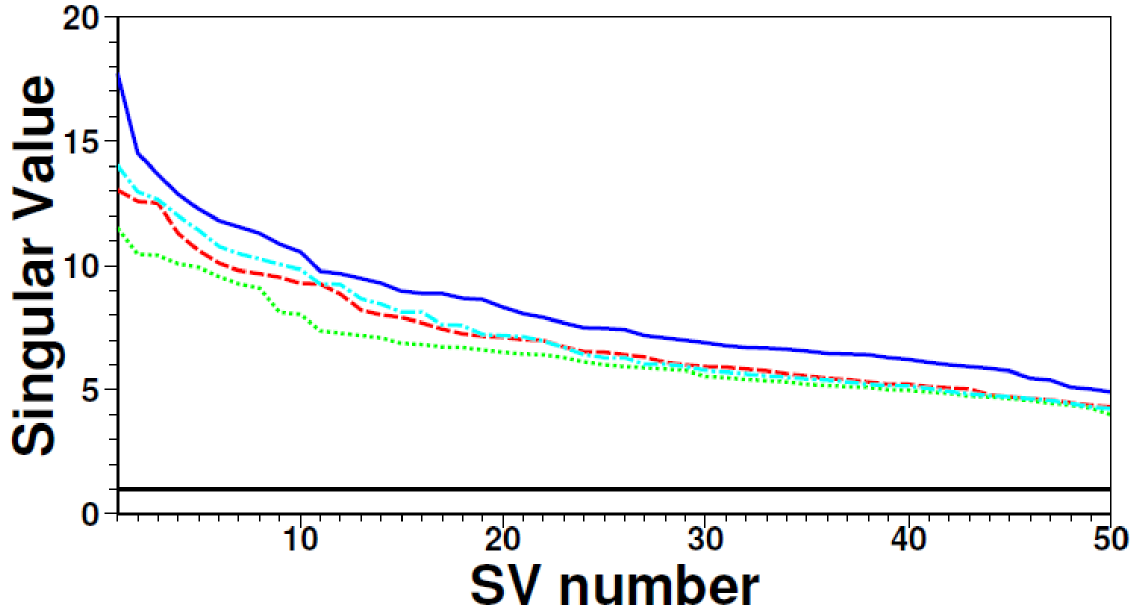
Northern Hem.

solid:	2005070100
dashed:	2005092100
dotted:	2005122100
chain-dashed:	2006032100



Southern Hem.

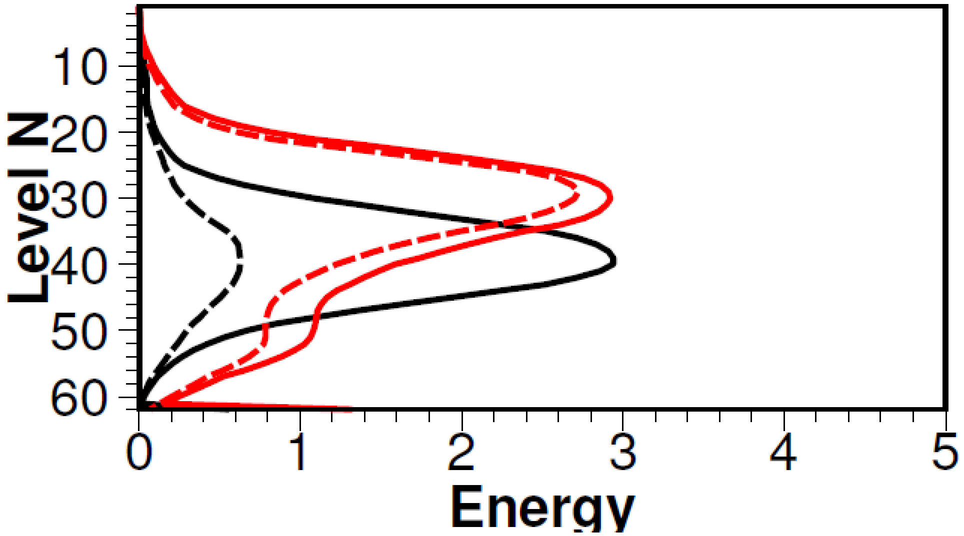
solid:	2005070100
dashed:	2005092100
dotted:	2005122100
chain-dashed:	2006032100



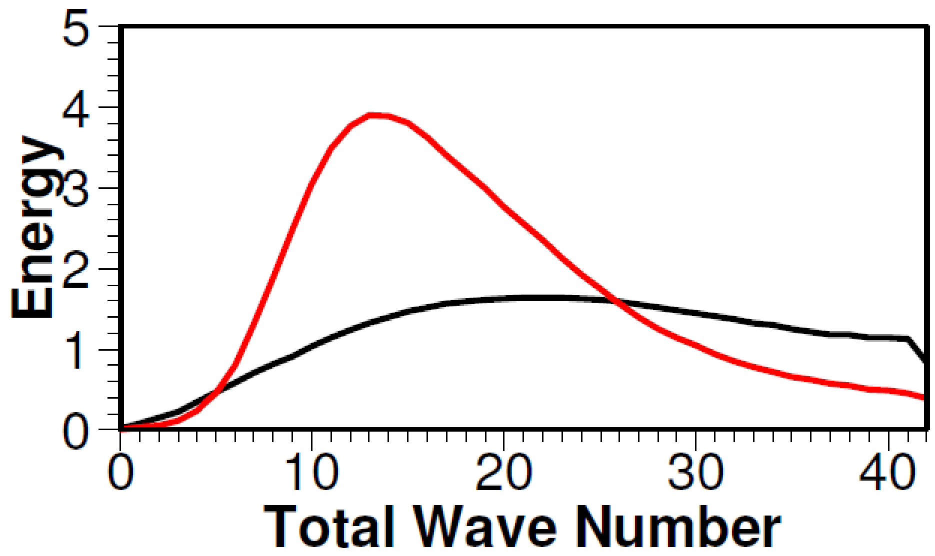
Singular vector growth characteristics

average energy of the leading 50 singular vectors
initial time ($\times$ 50), **final time t =48 h ($\times$ 1)**
▬▬ : total energy; ▬ ▬ : kinetic energy
Northern hemisphere extra-tropics, 2006032100

vertical profile



spectrum



200 hPa ↔ level 20   300 hPa ↔ level 27
500 hPa ↔ level 35   700 hPa ↔ level 42
850 hPa ↔ level 48   925 hPa ↔ level 52

wave number	wave length
5	8000 km
10	4000 km
20	2000 km
40	1000 km

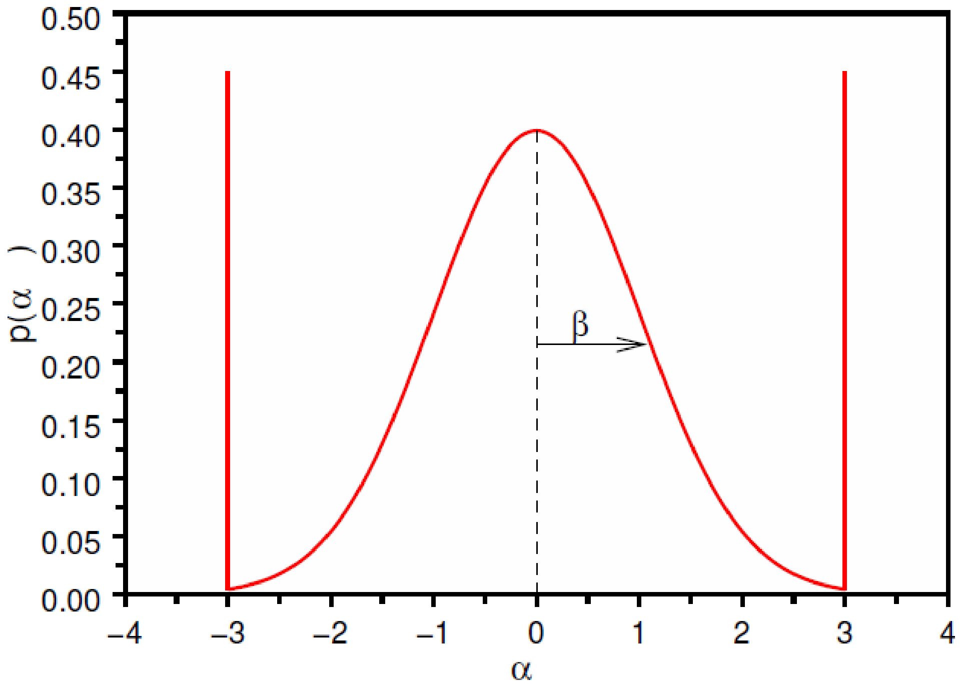
Initial condition perturbations

- Initial condition uncertainty is represented by a (multi-variate) Gaussian distribution in the space spanned by the leading singular vectors
- The perturbations based on a set of singular vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are of the form

$$\mathbf{x}_j = \sum_{k=1}^{m} \alpha_{jk} \mathbf{v}_k \tag{5}$$

- The α_{jk} are independent draws from a truncated **Gaussian distribution**.

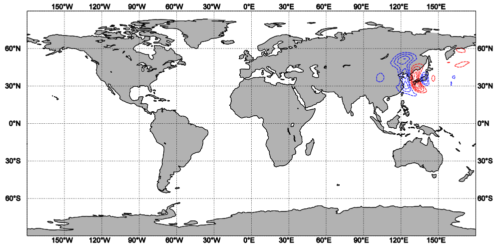
- The Gaussian is truncated at ± 3 standard deviations to avoid numerical instabilities for extreme values.
- The width of the distribution is set so that the spread of the ensemble matches the root-mean square error in an average over many cases ($\beta \approx 10$).



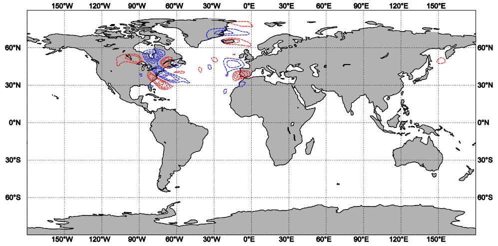
Combine SVs to construct Perturbations:

SV_1 $\qquad$ SV_2 $\qquad$ SV_{50}

$\boldsymbol{\alpha_1} \times$



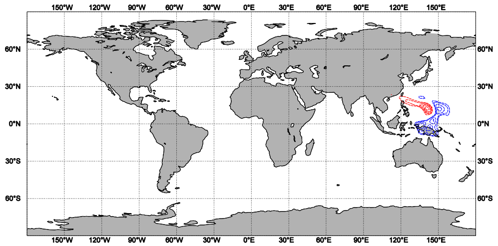
$+ \boldsymbol{\alpha_2} \times$



$+ \cdots$

SV_{TC1} $\qquad$ SV_{TCn}

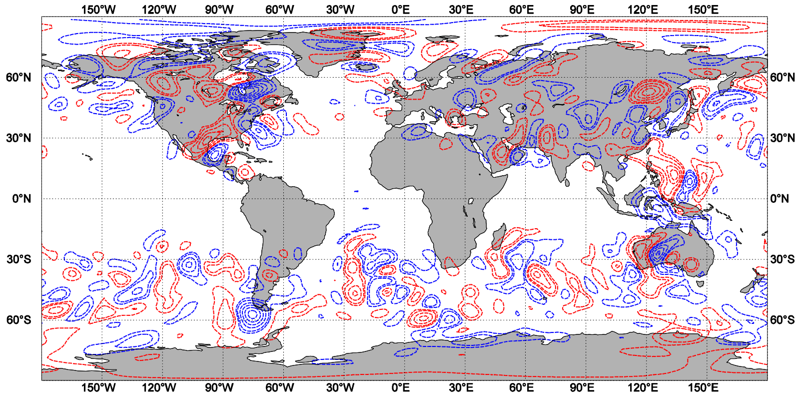
$+ \boldsymbol{\alpha_{TC1}} \times$



$+ \cdots$

$=$

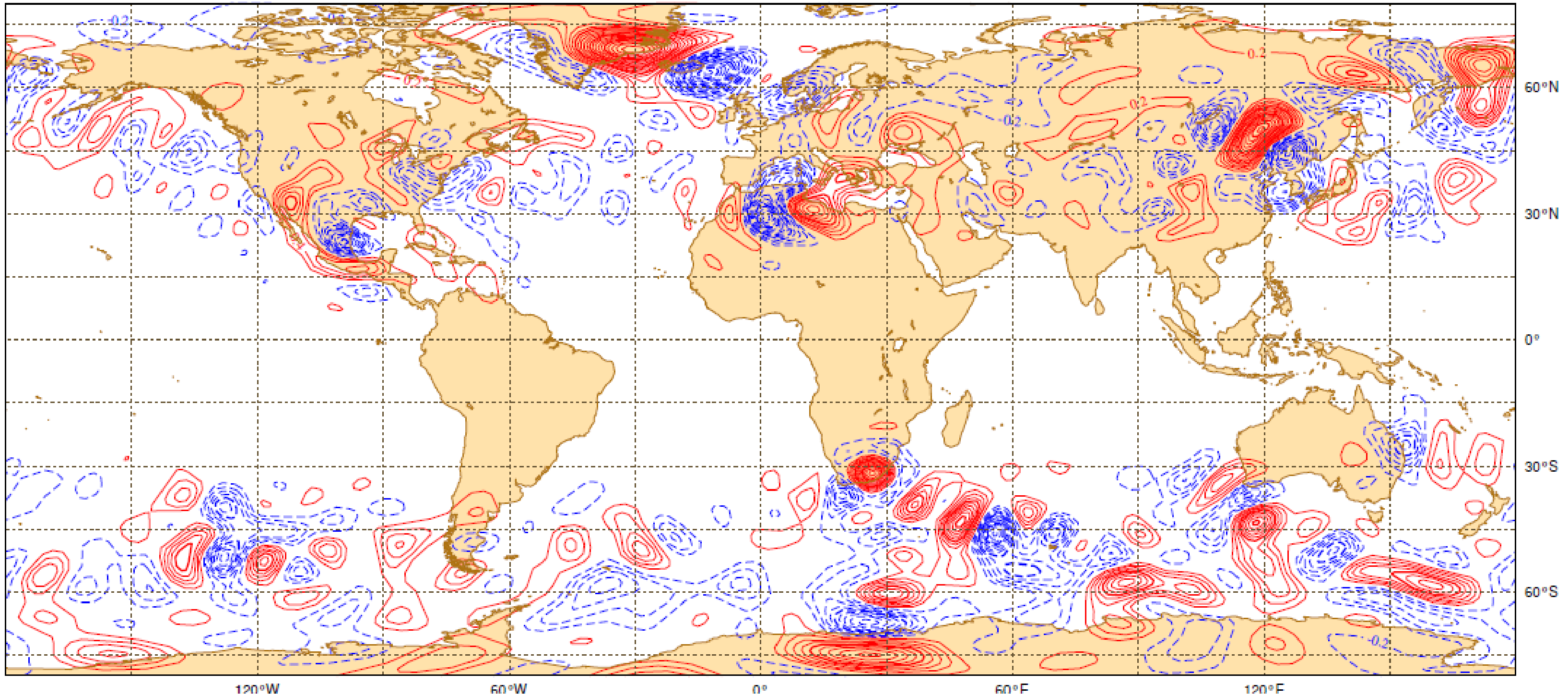
SVPERT 1



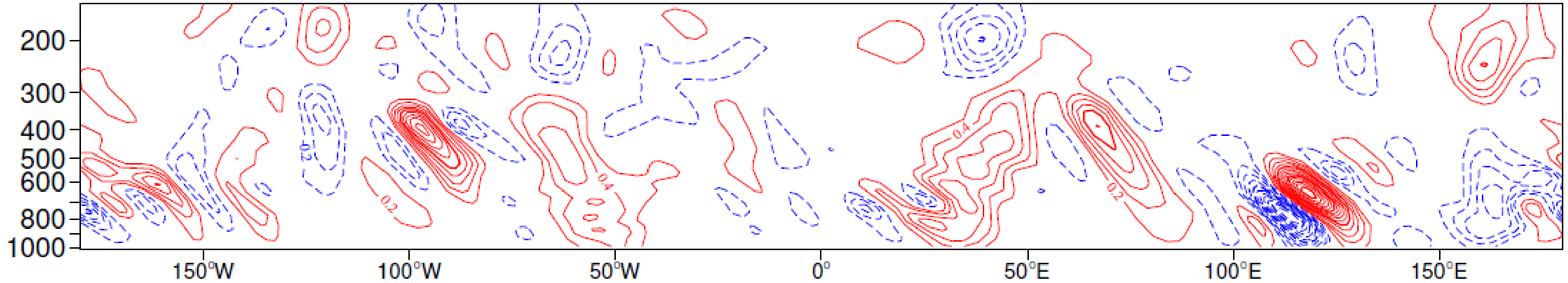
Initial condition perturbation for member 1

Temperature (every 0.2 K); 21 March 2006, 00 UTC

at ≈ 700 hPa



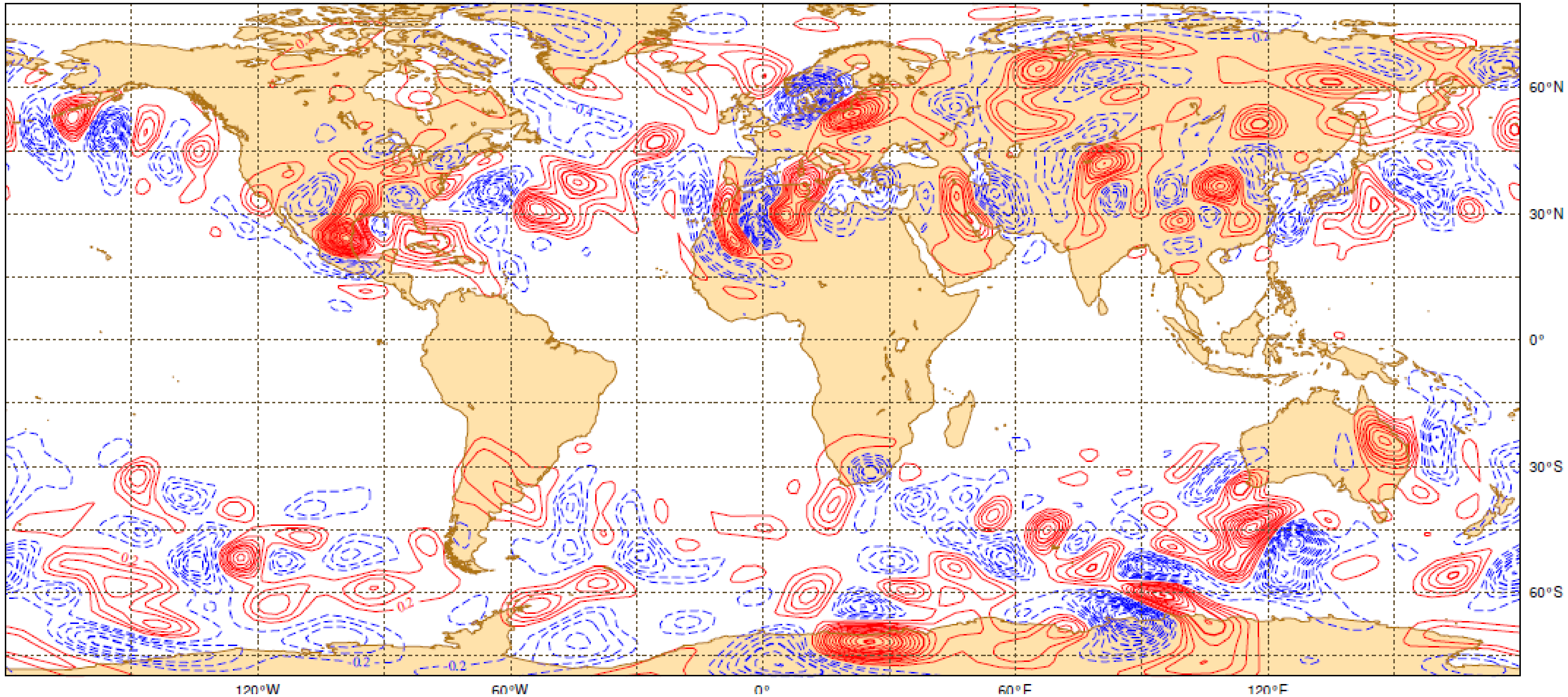
at 50°N



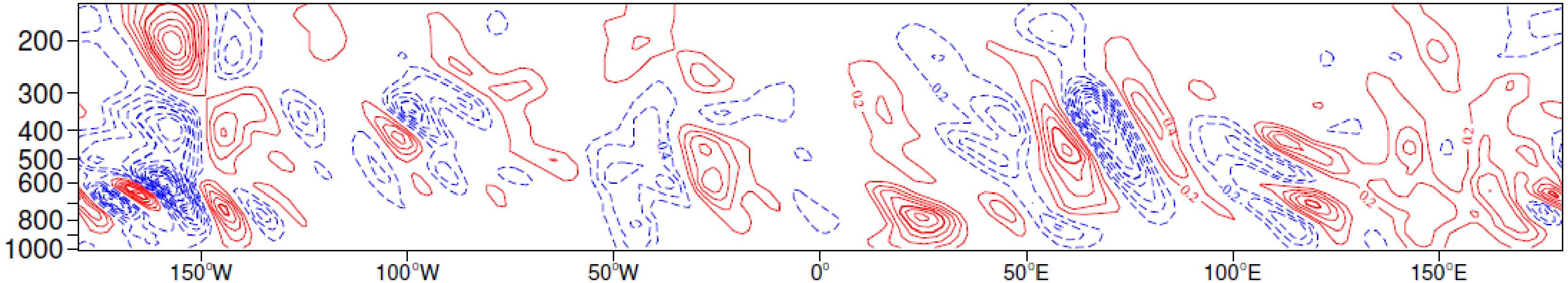
Initial condition perturbation for member 5

Temperature (every 0.2 K); 21 March 2006, 00 UTC

at ≈ 700 hPa



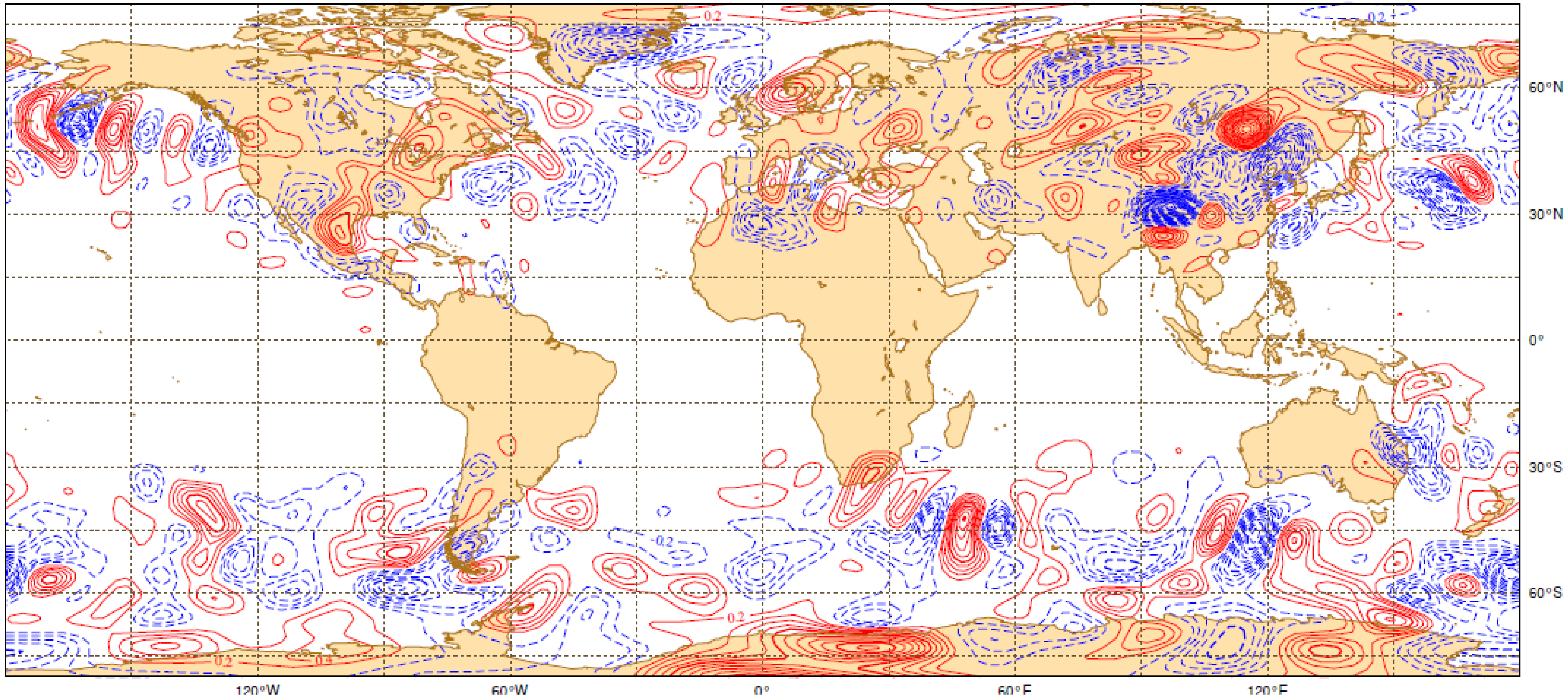
at 50°N



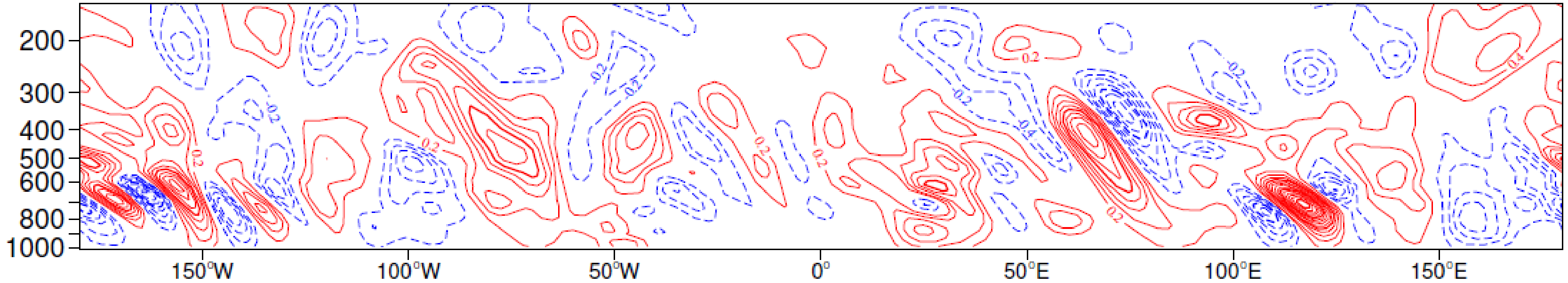
Initial condition perturbation for member 50

Temperature (every 0.2 K);   21 March 2006, 00 UTC

at ≈ 700 hPa

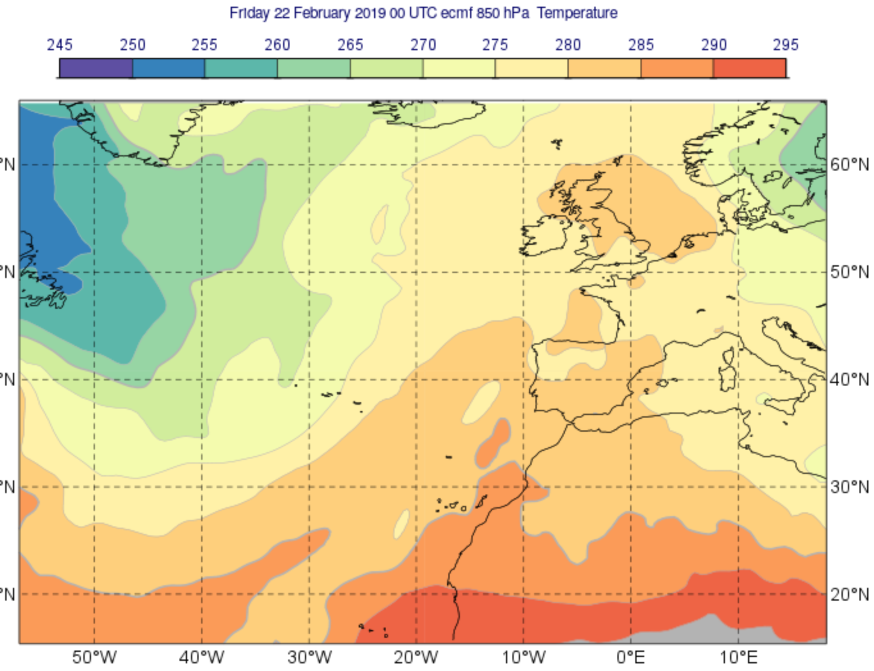


at 50°N



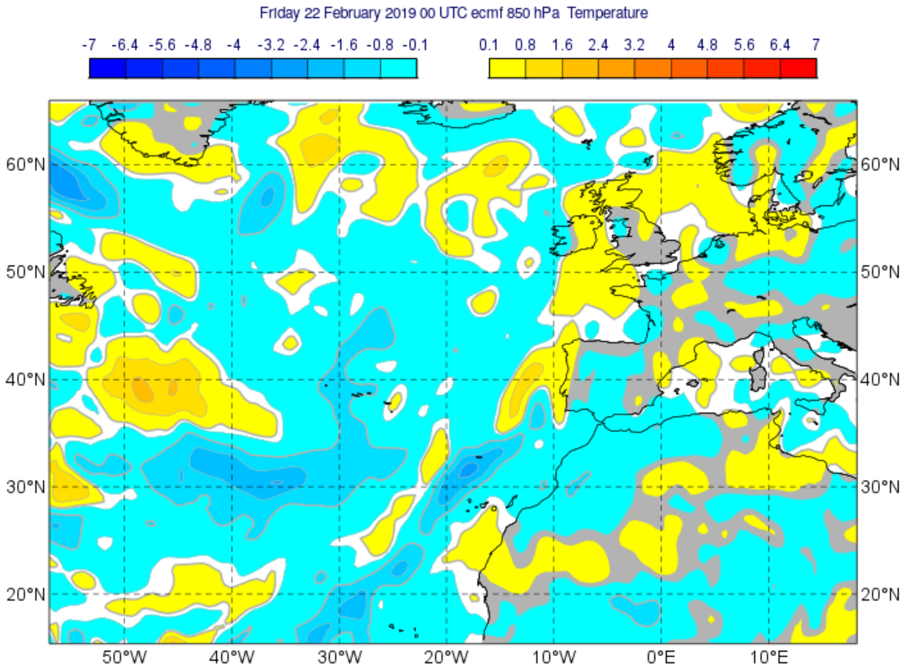
T850hPa

HRES Analysis 00 UTC



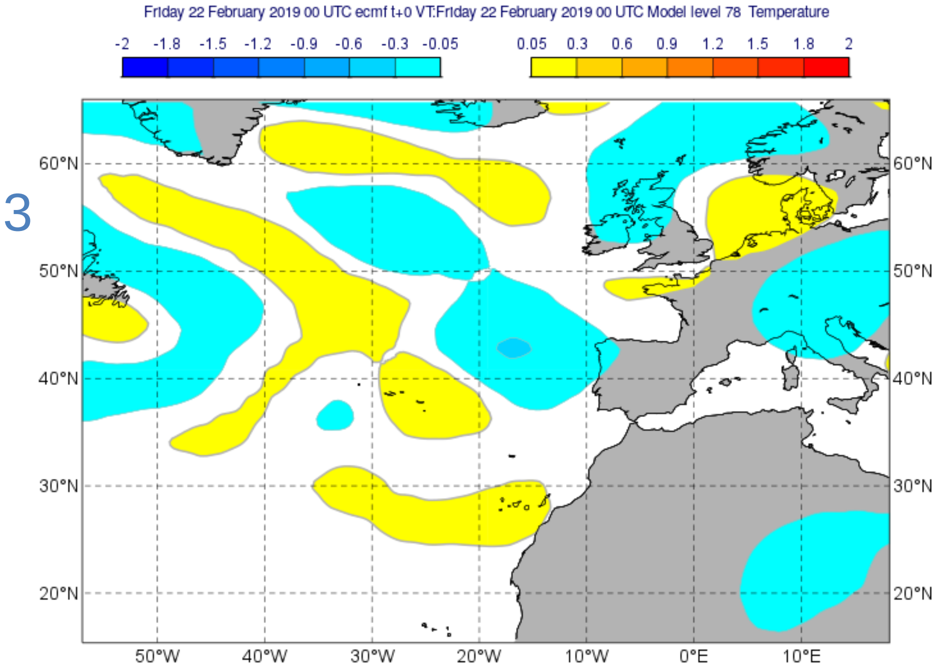
EDA-Pert 3

+



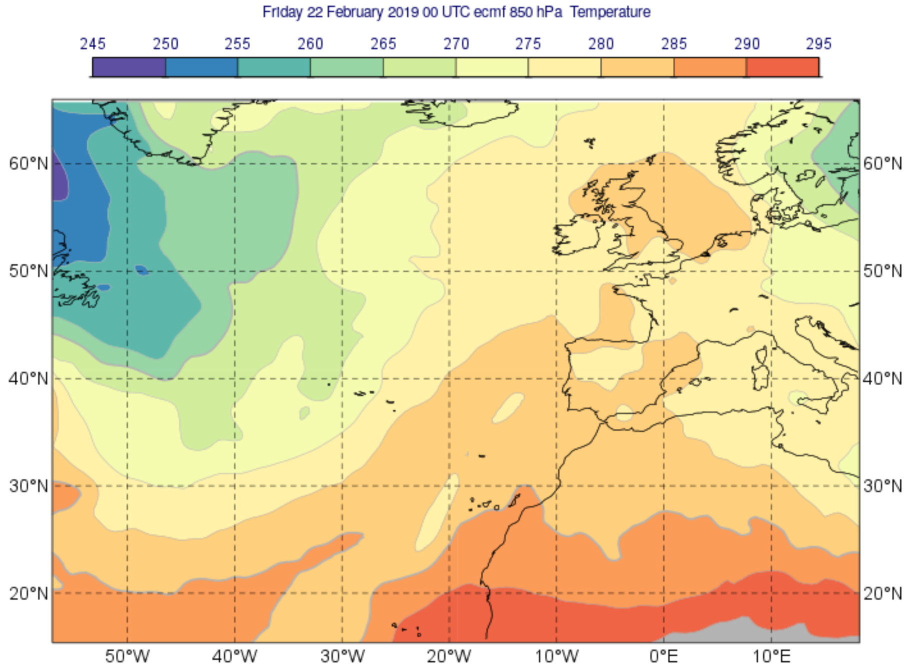
SV-Pert 3

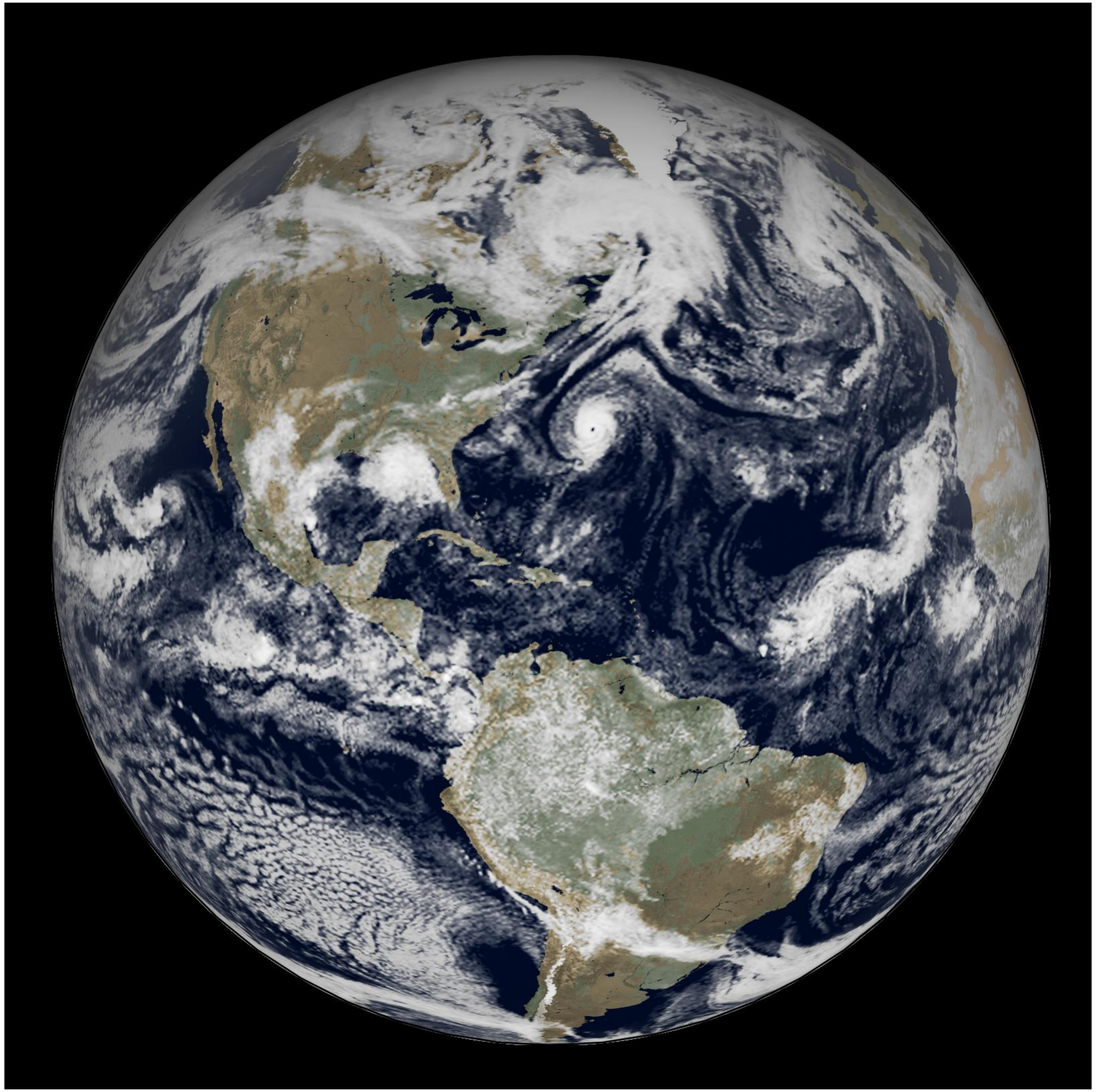
+



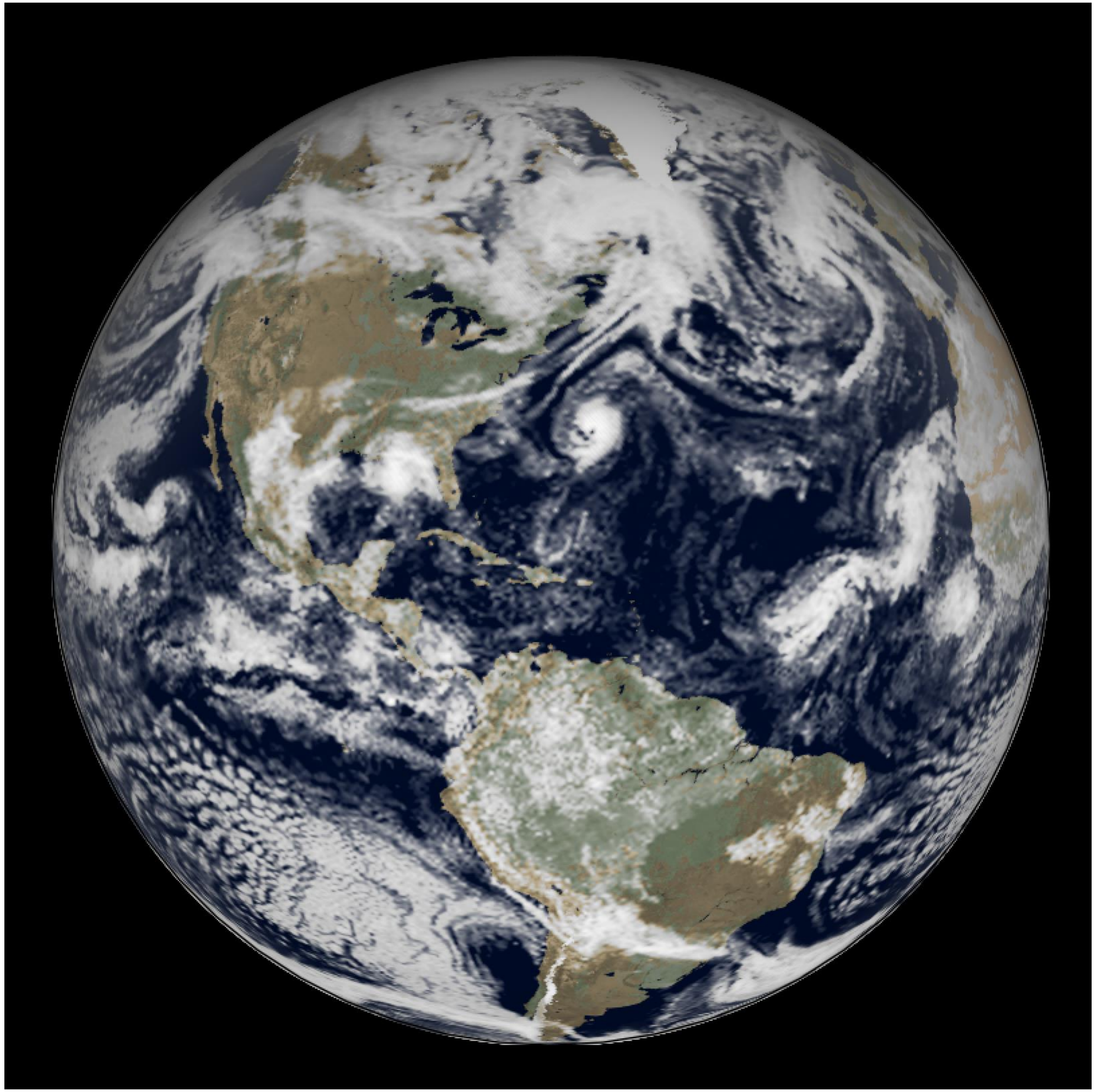
=

Initial conditions for ENS member 3





TCo1279L137

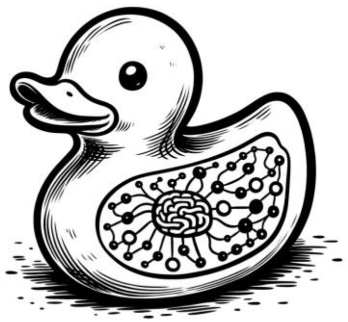


TCo639L91

TCo1279L137

TCo639L91

AIFS - Artificial Intelligence Forecasting System

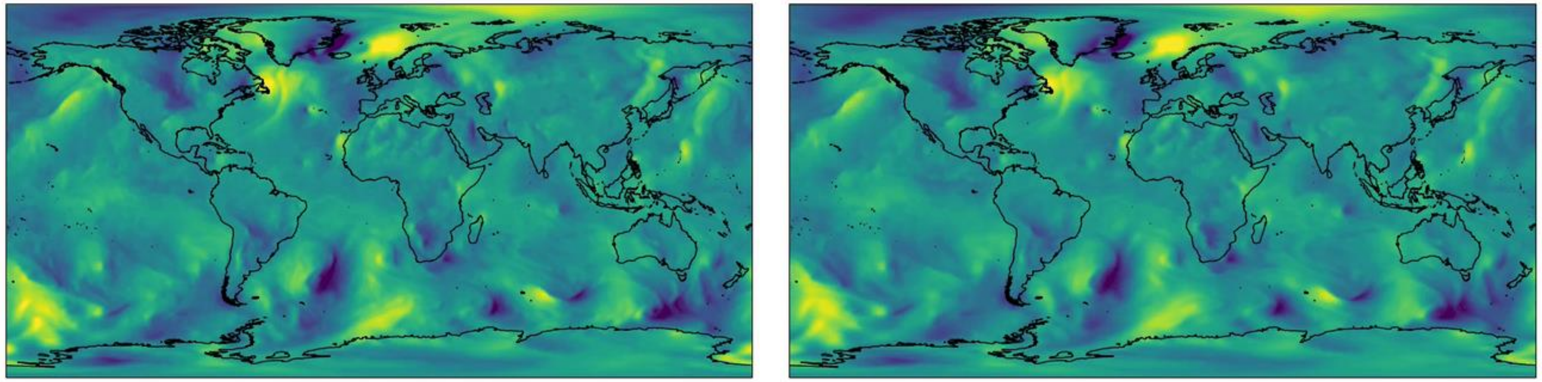


First implementation (~ 1deg resolution) in 2023, following Keisler 2022 and Lam et. al 2022:

- GNN architecture: Interaction Networks (Battaglia et. al 2016)
- Graph representation, hidden multi-scale mesh, edge features
- Scales to > 1000s of GPUs ; tensor parallel implementation, split model across multiple GPUs

Update beginning of 2024, update to ~ 0.25 deg:

- Attention / Transformer based GNN for encoder, decoder (Shi et al., 2021)
- Transformer backbone in processor (with sliding window, e.g. Child et al. 2019, Jiang et al. 2023)
- Trained on 64 GPUs ~ 1 Week

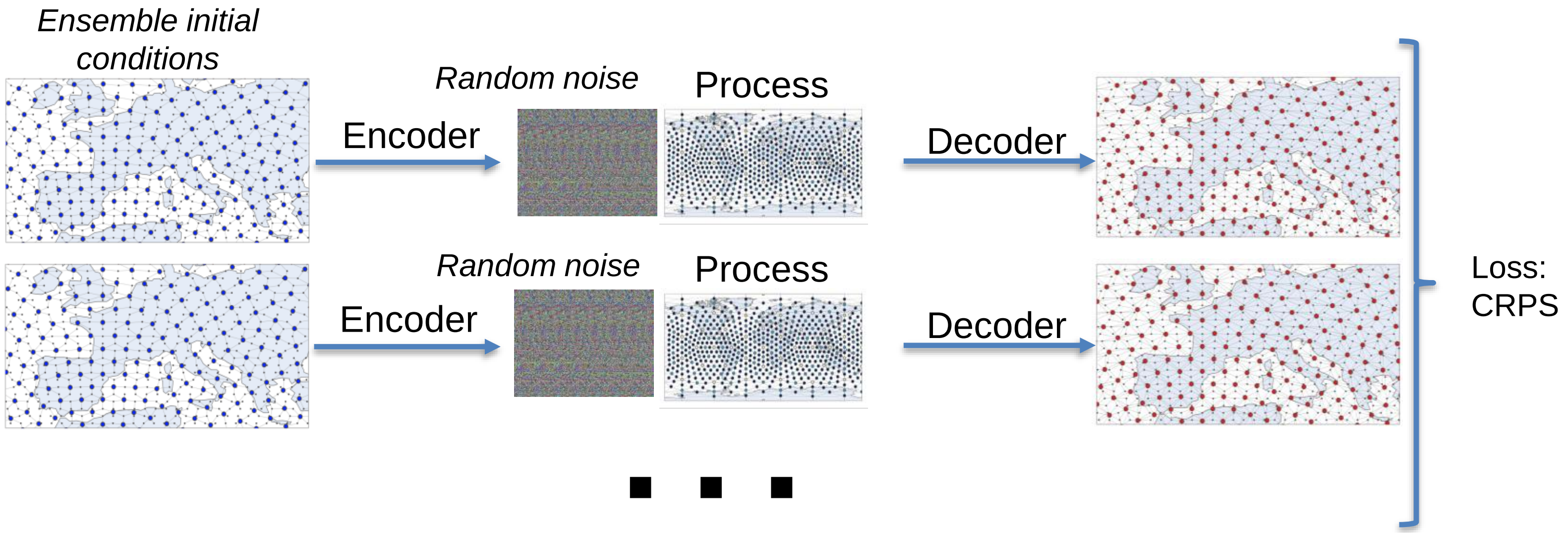


Why GNN: Encoder / Decoder: can handle arbitrary input / output grids, local and ad hoc grid refinement, changing grids etc. ; attractive for use in earth system science

See Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M. C. A., Lessig, C., Maier-Gerber, M., Magnusson, L., Bouallègue, Z. B., Nemesio, A. P., Dueben, P. D., Brown, A., Pappenberger, F., Rabier, F. (2024). AIFS - ECMWF’s data-driven forecasting system. arXiv. https://doi.org/10.48550/ARXIV.2406.01465

Ensemble forecasts …

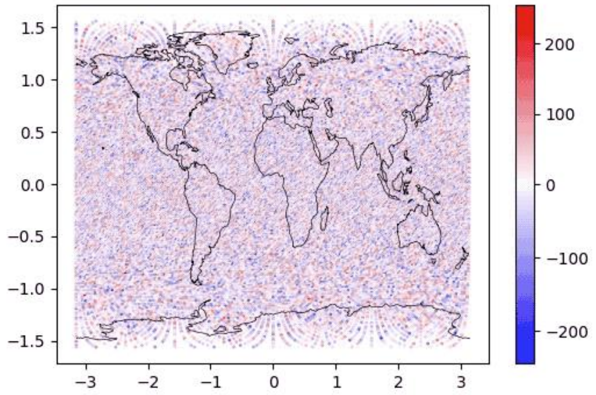
a) Instead of a MSE loss, learn an ensemble via optimizing probabilistic scores

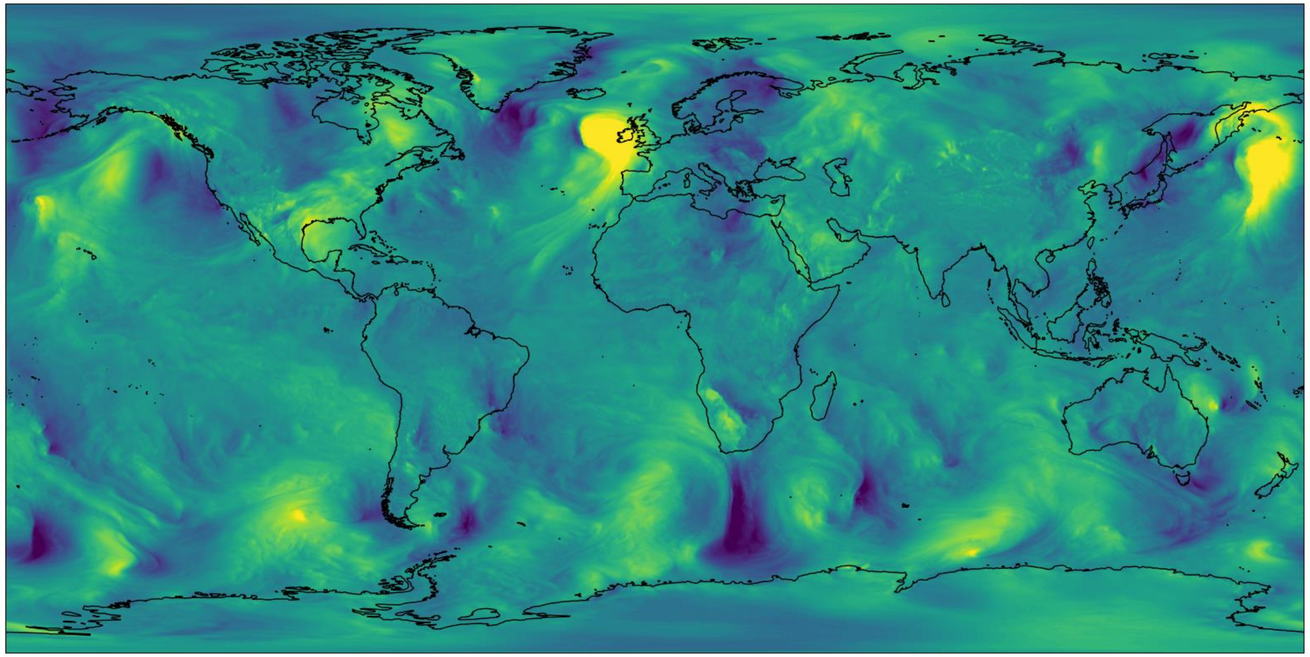


b) Create a forecast as de-noising task (diffusion training)

for example:

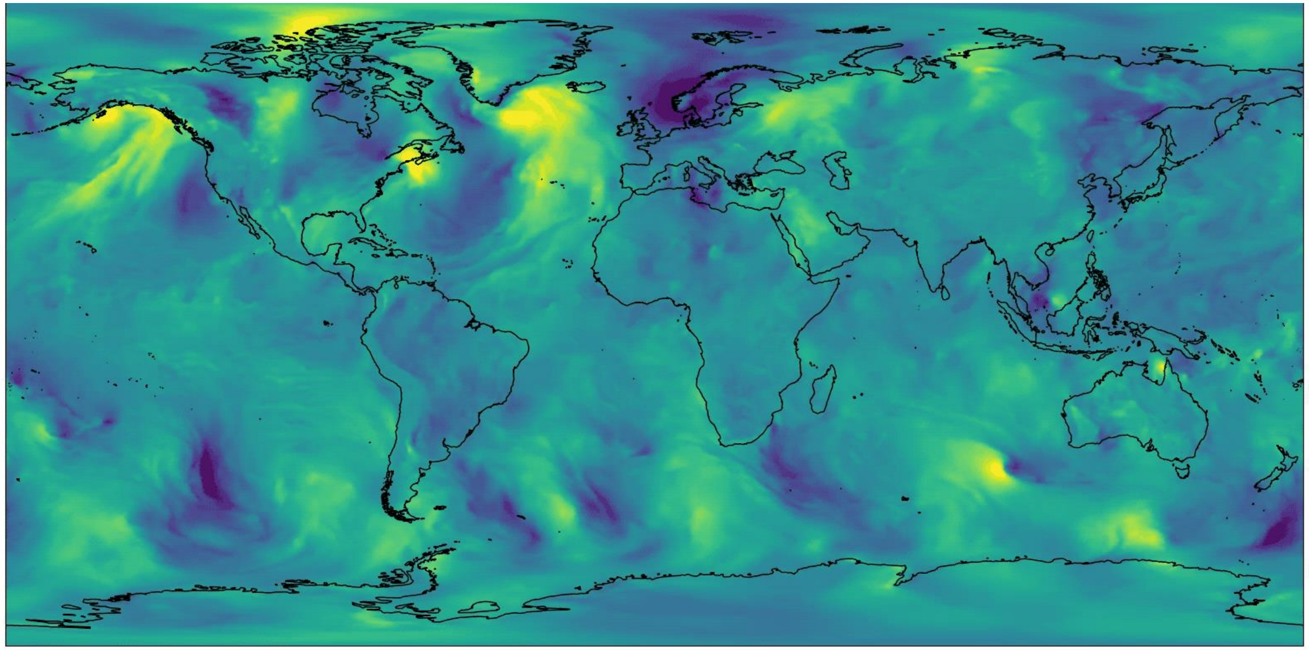
- Stable diffusion -> Images
- Sora -> Video
- Gencast, AIFS-Diffusion -> Weather





AIFS, ~ 0.25deg

Smoothing with lead time



AIFS Diffusion ~ 0.25deg

Stable statistics of forecast fields

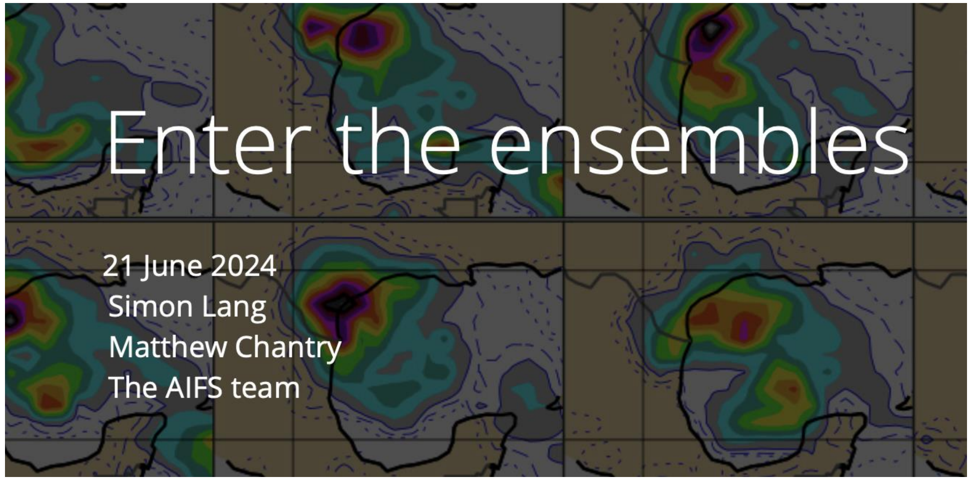
Towards operational data-driven ensemble forecasting:

First test system running now 2x daily,

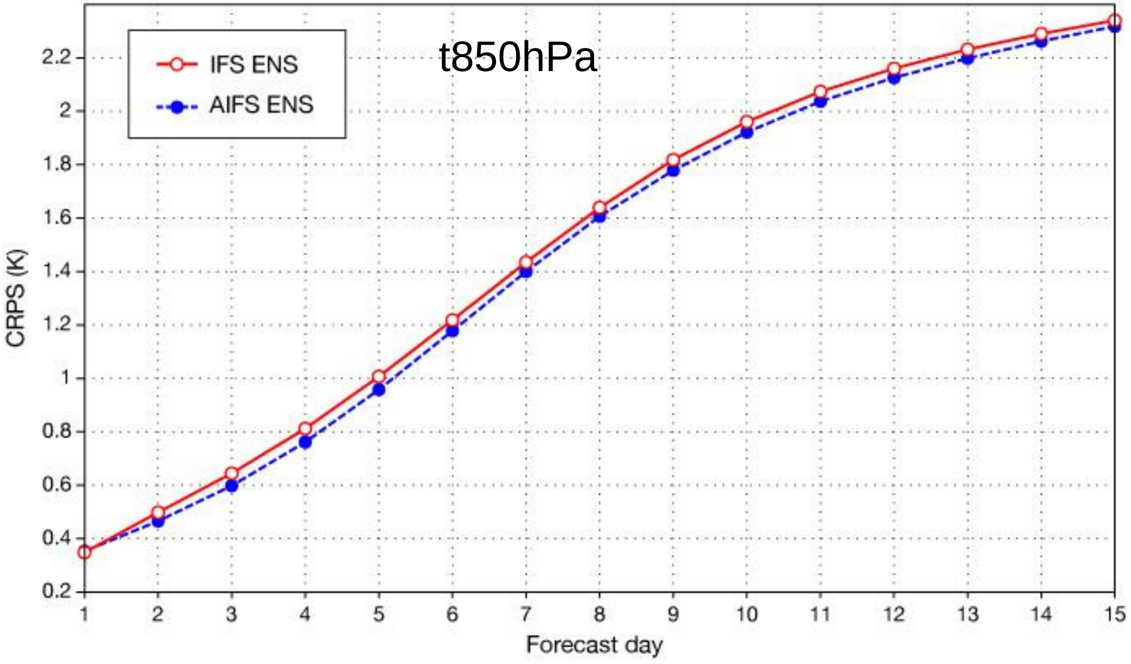
based on AIFS diffusion flavor,

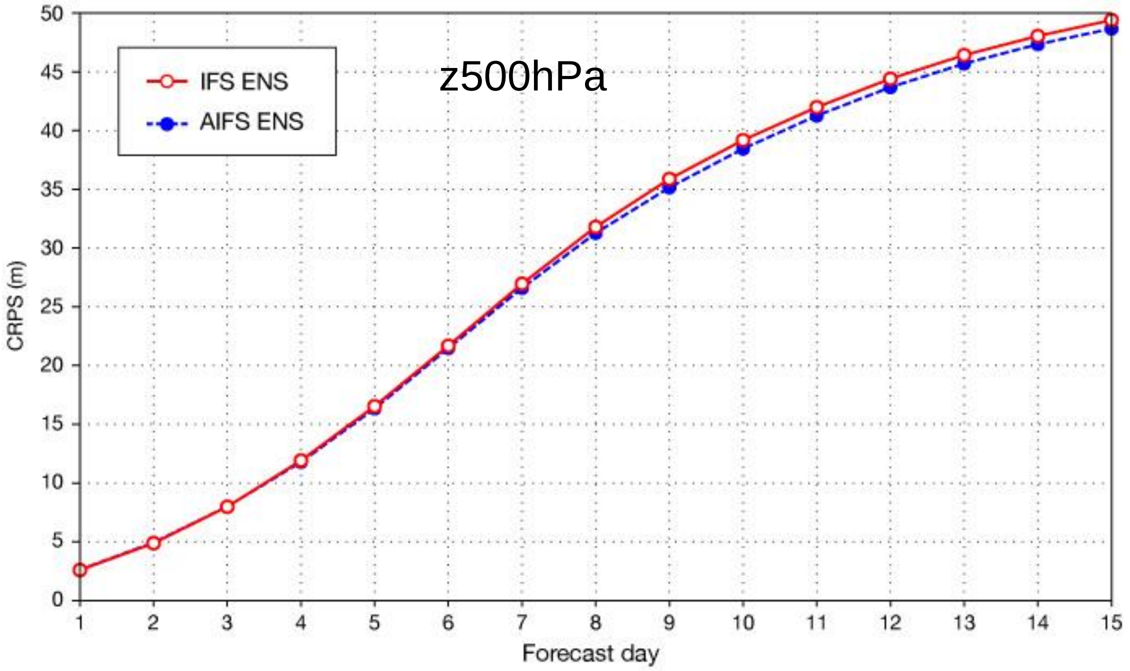
Initial perturbations form operational IFS ensemble

~ 1 deg resolution, fine-tuned on oper analyses

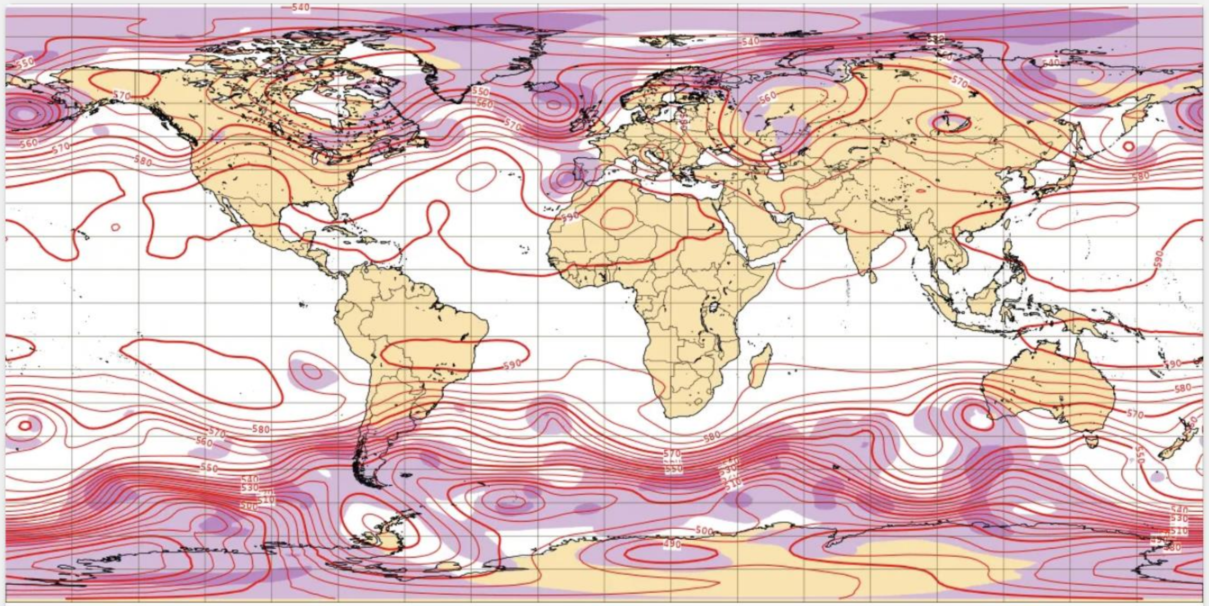


https://www.ecmwf.int/en/about/media-centre/aifs-blog/2024/enter-ensembles

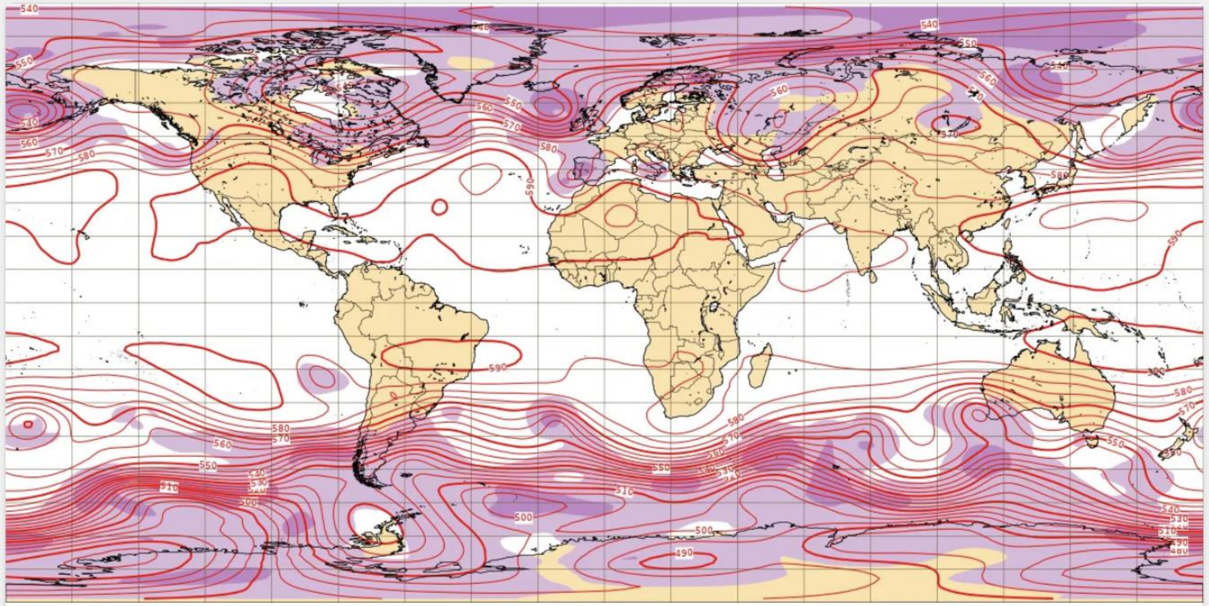




AIFS ENS Products: https://charts.ecmwf.int, Meteograms available

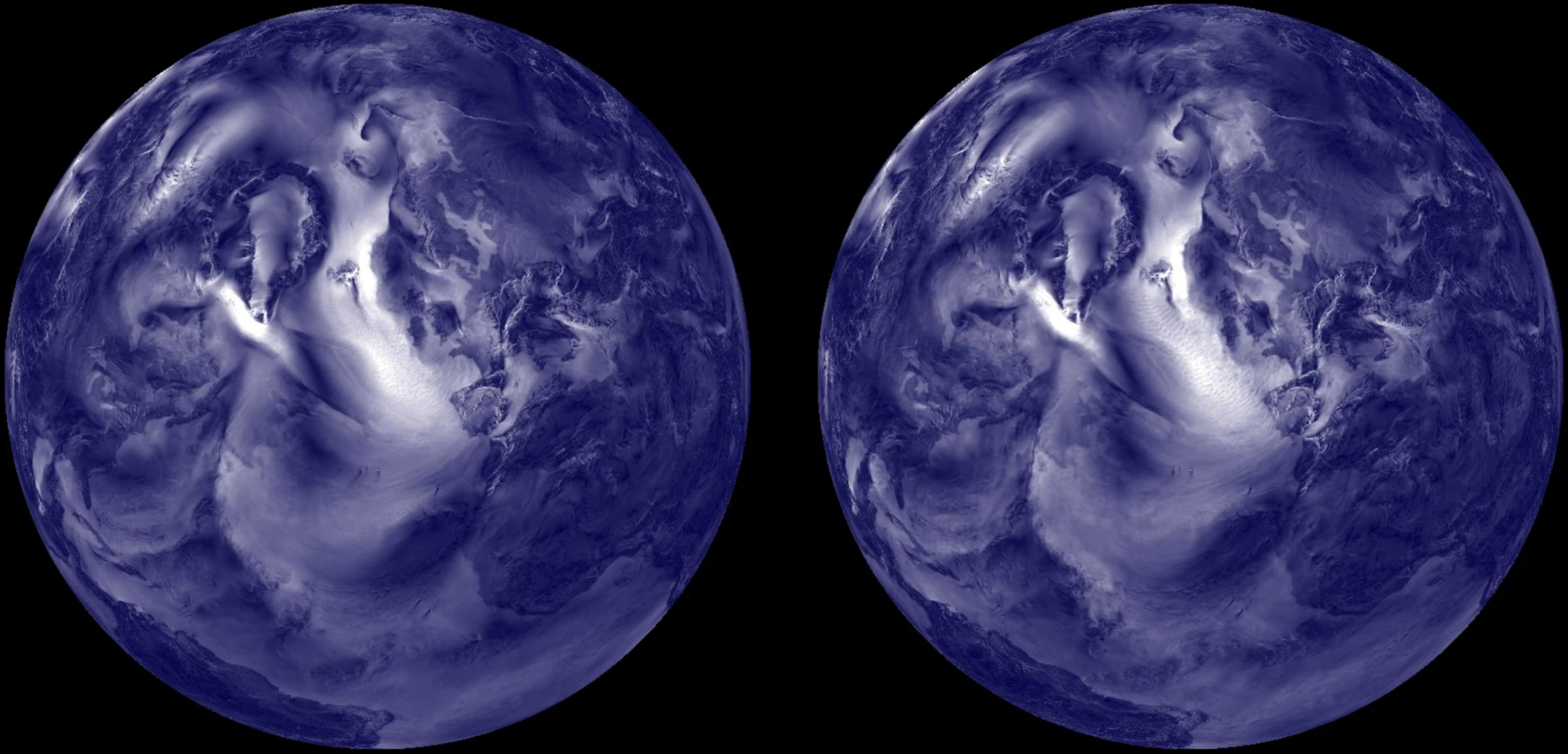


Ensemble mean and spread: 500 hPa geopotential height



Experimental: AIFS ENS (ECMWF) ML model: Ensemble mean and spread: 500 hPa geopotential height

IFS 10m wind gusts, 2020-12-04 00 UTC 720h forecasts, 9 km spatial resolution



Control Member

Perturbed member 1

Some References:

Buizza, R. , Leutbecher, M. and Isaksen, L. (2008), Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. Q.J.R. Meteorol. Soc., 134: 2051-2066.

Lang, S. T., Bonavita, M. and Leutbecher, M. (2015), On the impact of re-centring initial conditions for ensemble forecasts. Q.J.R. Meteorol. Soc., 141: 2571-2581.

Leutbecher, Martin and Tim N. Palmer. “Ensemble forecasting.” J. Comput. Physics 227 (2008): 3515-3539.

Leutbecher, M. and Lang, S. T. (2014), On the reliability of ensemble variance in subspaces defined by singular vectors. Q.J.R. Meteorol. Soc., 140: 1453-1466.

Leutbecher, M. , Lock, S. , Ollinaho, P. , Lang, S. T., Balsamo, G. , Bechtold, P. , Bonavita, M. , Christensen, H. M., Diamantakis, M. , Dutra, E. , English, S. , Fisher, M. , Forbes, R. M., Goddard, J. , Haiden, T. , Hogan, R. J., Juricke, S. , Lawrence, H. , MacLeod, D. , Magnusson, L. , Malardel, S. , Massart, S. , Sandu, I. , Smolarkiewicz, P. K., Subramanian, A. , Vitart, F. , Wedi, N. and Weisheimer, A. (2017), Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. Q.J.R. Meteorol. Soc, 143: 2315-2339.

Lang, S., Hólm, E., Bonavita, M., Trémolet, Y., A 50-member Ensemble of Data Assimilations, ECMWF Newsletter No. 158 - Winter 2018/19
Lang, S.T.K., Dawson, A., Diamantakis, M., Dueben, P., Hatfield, S., Leutbecher, M., Palmer, T., Prates, F., Roberts, C.D., Sandu, I. and N. Wedi (2021). More accuracy with less precision. Q J R Meteorol Soc, 147(741, 4358– 4370. https://doi.org/10.1002/qj.4181

Lopez, P. (2020). Forecasting the Past: Views of Earth from the Moon and Beyond, Bulletin of the American Meteorological Society, 101(7), E1190-E1200.

Lam, R, Sanchez-Gonzalez A, Willson M, Wirnsberger P, Fortunato M, Pritzel A, Ravuri S, Ewalds T, Alet F, Eaton-Rosen Z and Hu W. GraphCast: Learning skillful medium-range global weather forecasting. arXiv preprint arXiv:2212.12794. 2022 Dec 24.

Lang, S., M. Rodwell, D. Schepers, 2023: IFS upgrade brings many improvements and unifies medium-range resolutions. ECMWF Newsletter 176. https://doi.org/10.21957/slk503fs2i

Nonlinear Model

Consider the spatially discretised equations describing the atmospheric dynamics and physics written in this form

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = F(\mathbf{x}), \quad \text{with} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad F = \begin{bmatrix} F_1 \\ \vdots \\ F_n \end{bmatrix}$$

where $\mathbf{x} \in \mathbb{R}^N$ denotes the N-dimensional state vector and $F(\mathbf{x}) \in \mathbb{R}^N$ its tendency. Integrate from t_0 to t gives the nonlinear model:

$$\mathbf{x}(t) = \mathcal{F}(\mathbf{x}(t_0))$$

Tangent-linear system

Let $\mathbf{x}_r(t)$ be a solution of

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = F(\mathbf{x}), \tag{1}$$

Then the *tangent-linear system* is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = \mathbf{A}(\mathbf{x}_r(t))\,\mathbf{x}, \tag{2}$$

where $[\mathbf{A}(\mathbf{x})]_{jk} = (\partial F_j/\partial x_k)(\mathbf{x})$ denotes the Jacobi matrix of F.

For any solution $\mathbf{x}$ of (2), $\mathbf{x}_r + \varepsilon\mathbf{x}$ approximates a solution of (1) starting at $\mathbf{x}_r(t_0) + \varepsilon\mathbf{x}(t_0)$ to first order in ε.

The *(tangent-linear) propagator* from t_0 to t_1 is the matrix $\mathbf{M}_{[t_0,t]}$ such that $\mathbf{M}_{[t_0,t]}\mathbf{x}_0$ is a solution of (2) for any initial perturbation $\mathbf{x}_0$ and where $\mathbf{M}_{[t_0,t_0]} = \mathbf{I}$.

Perturbation Growth

Perturbation growth is defined as:

$$\sigma^2 = \frac{\langle \mathbf{x}(t), \mathbf{x}(t) \rangle}{\langle \mathbf{x}(t_0), \mathbf{x}(t_0) \rangle} = \frac{\langle \mathbf{M}_{[t_0,t]}\mathbf{x}(t_0), \mathbf{M}_{[t_0,t]}\mathbf{x}(t_0) \rangle}{\langle \mathbf{x}(t_0), \mathbf{x}(t_0) \rangle} = \frac{\langle \mathbf{M}^T_{[t_0,t]}\mathbf{M}_{[t_0,t]}\mathbf{x}(t_0), \mathbf{x}(t_0) \rangle}{\langle \mathbf{x}(t_0), \mathbf{x}(t_0) \rangle}$$

with inner product $\langle \cdot, \cdot \rangle$ and growth factor σ^2.

$\Rightarrow$ Largest growth is associated with eigenvectors of $\mathbf{M}^T_{[t_0,t]}\mathbf{M}_{[t_0,t]}$. These eigenvectors are determined by a singular value decomposition of $\mathbf{M}_{[t_0,t]}$.

singular value decomposition of a matrix

Consider a matrix
$$\mathbf{Q} = \begin{pmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & & \vdots \\ q_{m1} & \cdots & q_{mn} \end{pmatrix} \in \mathbb{R}^{m\times n}$$

Its singular value decomposition is defined as

$$\mathbf{Q} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathrm{T}}, \tag{3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal m-by-m and n-by-n matrices.
Matrix $\mathbf{S}$ is a diagonal m-by-n matrix ($s_{ij} = 0$ if $i \neq j$, $s_{jj} \equiv \sigma_j$). The values σ_j on the diagonal of $\mathbf{S}$ are called *singular values*.
The columns $\mathbf{u}_j$ of $\mathbf{U}$ are referred to as *left singular vectors* and the columns $\mathbf{v}_j$ of $\mathbf{V}$ are referred to as *right singular vectors*.
Eq. (3) implies that
$$\mathbf{Q}\,\mathbf{v}_j = \sigma_j \mathbf{u}_j$$

One can show that the $\mathbf{v}_j$ are the eigenvectors of $\mathbf{Q}^T\mathbf{Q}$!

see Golub and Van Loan: *Matrix Computations* for further details

singular value decomposition of the propagator

$$\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} \quad \rightarrow \quad \mathbf{M}\mathbf{v}_j = \sigma_j \mathbf{u}_j$$

with the (initial) singular vectors $\mathbf{v}_j$ being the eigenvectors and the squared singular values σ_j^2 being the eigenvalues of $\mathbf{M}^T\mathbf{M}$. The $\mathbf{u}_j$ are called the evolved singular vectors.

Singular vectors are optimal perturbations in the following sense.

- the ratio of the final time norm to the initial time norm is given by the singular value:
$$\frac{\|\mathbf{M}\mathbf{v}_j\|_f}{\|\mathbf{v}_j\|_i} = \sigma_j \qquad (4)$$

- Singular vector j is the direction in phase space that maximises the ratio of norms in the subspace orthogonal (with respect to $\mathbf{C}_0^{-1}$) to the space spanned by singular vectors $1 \ldots j-1$.

Norms

- The definition of singular vectors in the context of ensemble prediction involves norms (based on an inner product or metric). These are required to measure the amplitude of perturbations.

  $$\langle \mathbf{x}, \mathbf{x} \rangle_C = \mathbf{x}^T \mathbf{C} \mathbf{x}$$

  where $\mathbf{C}$ is symmetric ($\mathbf{C}^{\mathrm{T}} = \mathbf{C}$) and positive definite ($\mathbf{x}^{\mathrm{T}} \mathbf{C} \mathbf{x} > 0$ for $\mathbf{x} \neq 0$).

- For predictability applications, the appropriate choice for the **initial time norm** is the analysis error covariance metric, *i.e.* the norm that is based on the inverse of the initial error covariance matrix (or some estimate thereof).

  $$\|\mathbf{x}\|_i^2 = \mathbf{x}^{\mathrm{T}} \mathbf{C}_0^{-1} \mathbf{x}$$

- The **final time norm** $\|\mathbf{x}\|_f$ is a convenient RMS measure of forecast error.
- **Total energy norm** is used both at initial and final time for the operational singular vector computations at ECMWF:

  $$\|\mathbf{x}\|_{\mathbf{E}}^2 = \mathbf{x}^{\mathrm{T}} \mathbf{E} \mathbf{x} = \frac{1}{2} \int_{p_0}^{p_1} \int_S \left(u^2 + v^2 + \frac{c_p}{T_r} T^2 \right) \mathrm{d}p \, \mathrm{d}s + \frac{1}{2} R_d T_r p_r \int_S (\ln p_{\mathrm{sfc}})^2 \mathrm{d}s$$

On the choice of the initial time norm

- The structure of singular vectors depends on the choice of the norm, in particular the initial time norm.
- An enstrophy norm at initial time penalises perturbations with small spatial scales, the initial SVs are planetary-scale structures.
- A streamfunction variance norm at initial time penalises the large scales and favours sub-synoptic scale perturbations.
- With a total energy norm at initial time, the energy spectrum of the initial SVs is “white” and best matches the spectrum of analysis error estimates from analyses differences (Palmer et al. 1998)

The tangent-linear model and its adjoint

- For a numerical model with $\sim 10^5 - 10^8$ variables it is not possible to obtain the propagator $\mathbf{M}$ as a matrix.
- Instead *algorithmic differentiation* is used to obtain the first derivative of the numerical algorithm that represents the forecast model.

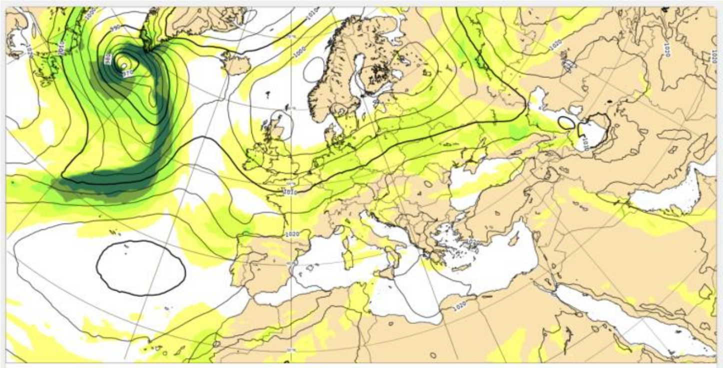
For any initial perturbation $\mathbf{x}$, the evolved perturbation $\mathbf{Mx}$ is obtained via an integration of the *tangent-linear model*.

- Then, the numerical algorithm representing $\mathbf{M}^{\mathrm{T}}$ the *adjoint* (transpose) of the propagator is constructed. The adjoint model is integrated backward from t_1 to t_0.
- The reference solution $\mathbf{x}_r(t)$ about which the equations are linearised is referred to as *trajectory*.
- The time interval the SVs are calculated for is called the optimization interval.

Weather Forecasts – NWP? Data Driven?

Traditionally weather forecasts are generated by running NWP model – computer code that has been designed to represent the physical processes governing the evolution of the atmosphere. But can you produce a forecast without a NWP model?

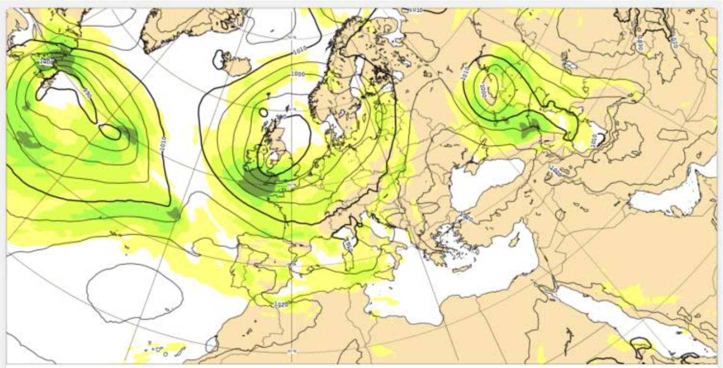
Analysis



Fusion of short-range forecast with latest observations

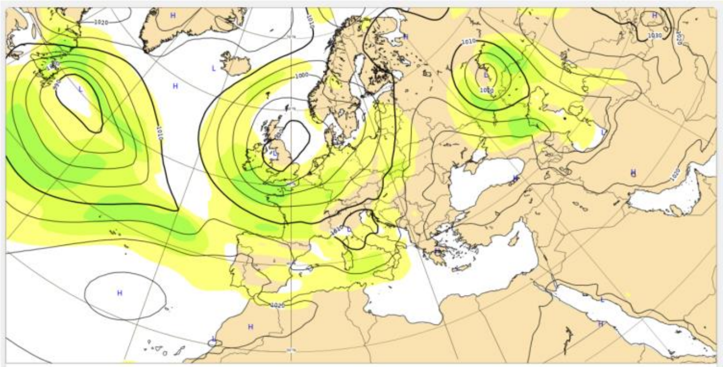
NWP Model

Forecast



Data Driven Model

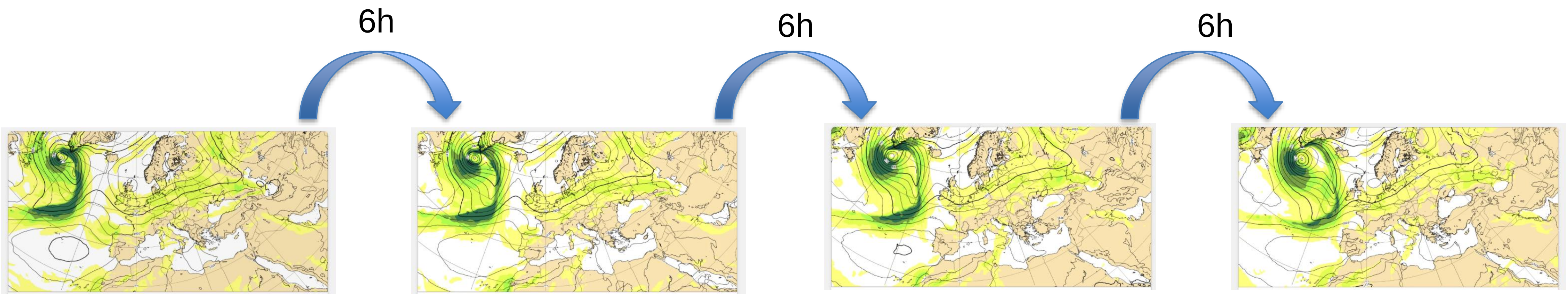
Learned from 40 years of analyses



Weather Forecasts – NWP? Data Driven?

Recent advances by tech companies and individuals show that this is possible (e.g. NVIDIA, Deepmind, Huawei, … and others)

Here, the models learn from ca. 40 years of ERA5 re-analysis data, stepping e.g. 6h from analysis to analysis



The forecast is then autoregressively stepping 6h into the future $x_n = f(x_{n-1})$ …