Data assimilation and ensembles: two invaluable tools to increase predictability and quantify uncertainty

Roberto Buizza
Scuola Universitaria Superiore Sant'Anna Pisa

Outline

1. Having an accurate Data Assimilation system is essential
2. ECMWF has been moving towards ensembles of DAs and FCs
3. Coupling to relevant processes has led to improved forecasts
4. Re-forecasts are essential to extract predictable signals
5. Moving towards a seamless ensemble of coupled DAs and FCs
6. Conclusions

Weather prediction is an initial-value problem

The j-th forecast starting from data/time (d,0), is given by the time integration

$$e_j(d,T) = e_j(d,0) + \int_0^T [A(e_j,t) + P(e_j,t) + \delta P_j(e_j,t)]dt$$

of the model equations starting from the j-th initial conditions

$$e_j(d,0) = e_0(d,0) + de_j(d,0)$$

$$de_j(d,0) = \sum_{area}\sum_{k=1}^{N_{SV}}[\alpha_{j,k} \cdot SV_k(d,0) + \beta_{j,k} \cdot SV_k(d-2,+2d)]$$

The perturbed model tendency is defined at each grid point by

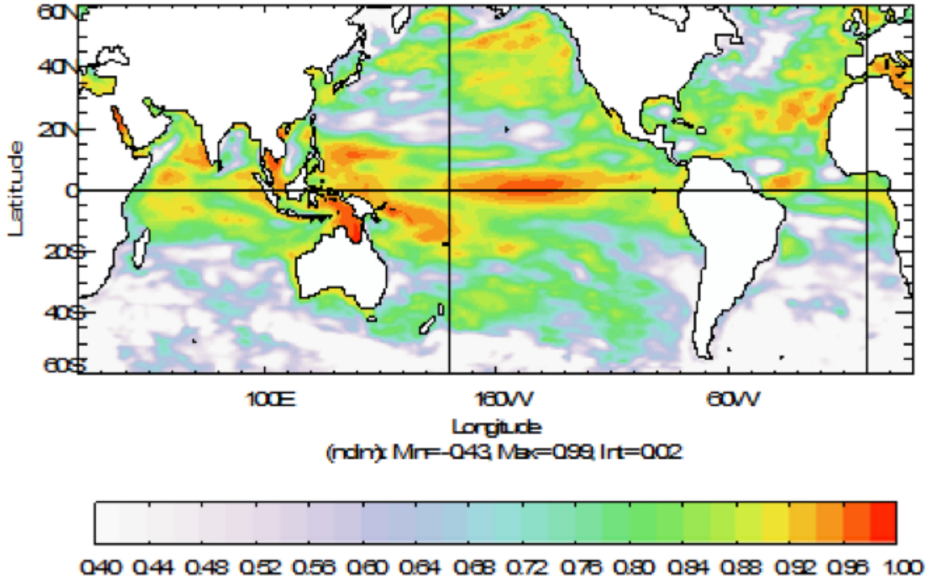
$$\delta P_j(e_j,t;\lambda,\phi,p) = r_j(t;\lambda,\phi)P_j(t;\lambda,\phi,p)$$

where $r_j(t;\Phi,\lambda)$ is a random number.

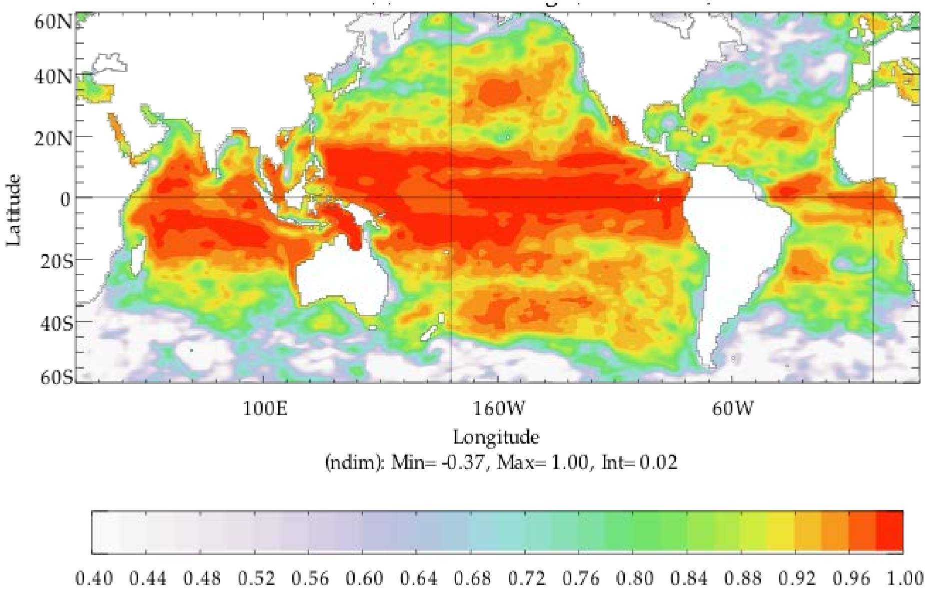
The impact of observations on seasonal fcs: an example

DA in the ocean (rhs) leads to better +1m fcs (higher ACC).

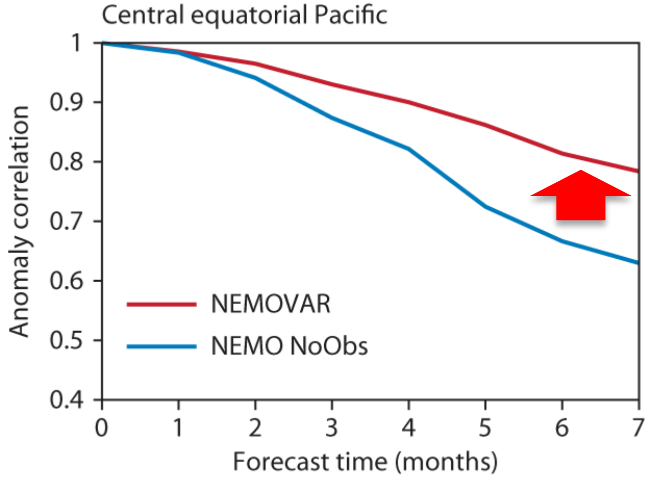
NEMO (no obs)

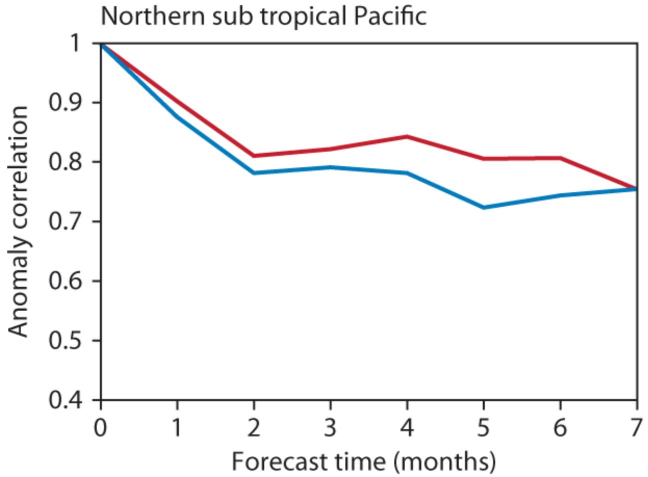


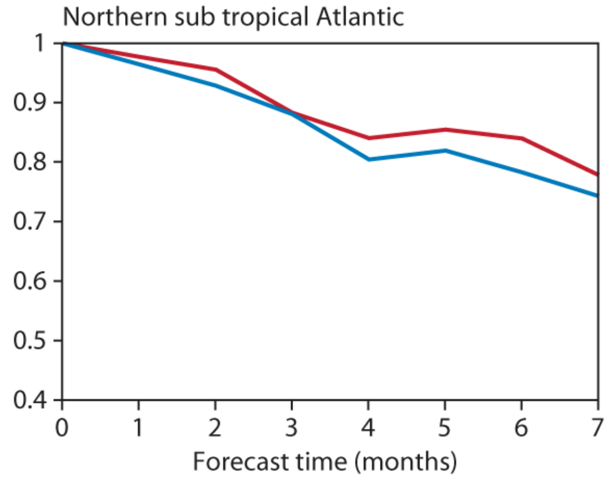
NEMOVAR: T+S+Alti



DA-initialized coupled models give better SST fcs.





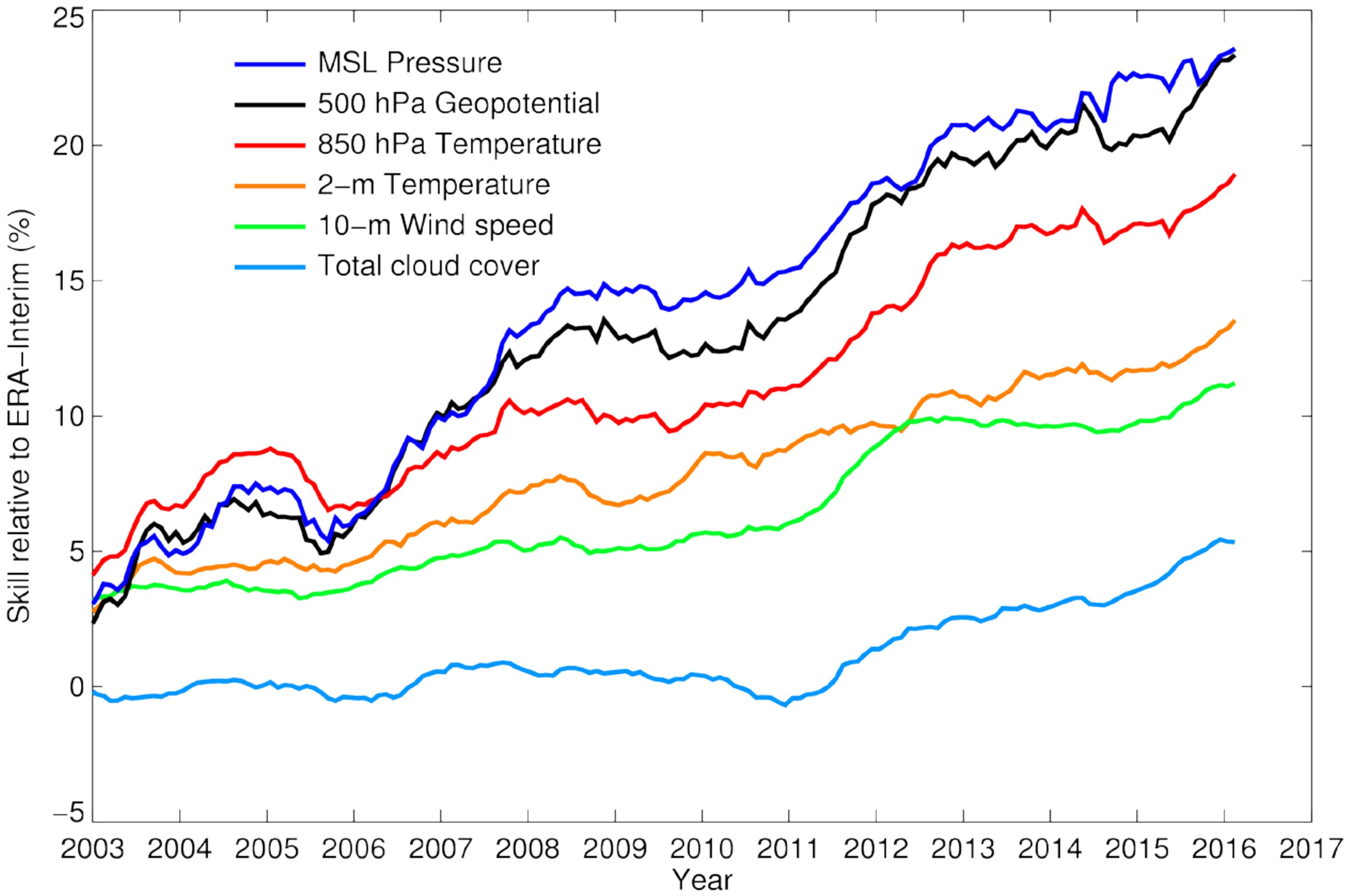


(F Vitart, M Alonso-Balmaseda)

Model, resolution, DA: what is more important?

It is difficult to identify the impact of model (**M**, including resolution) and **DA** improvements.

Improvements are due to a mix of upgrades in model, DA, use of observations (**O**) and resolution.

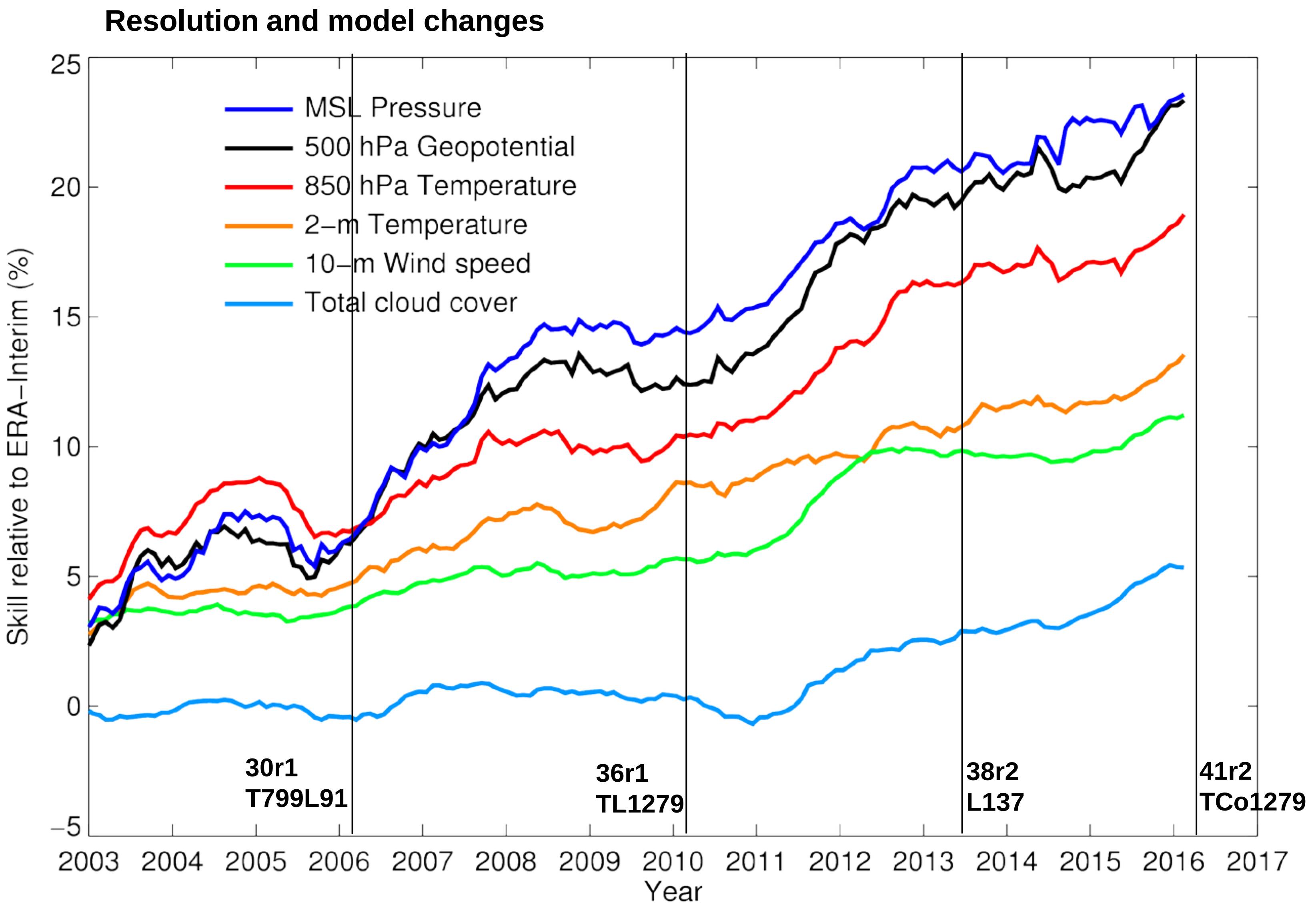


Sant'Anna School of Advanced Studies – Pisa

Some evidence: impact of res+model changes on fc skill

Resolution matters, although it could take time before investments pay off.

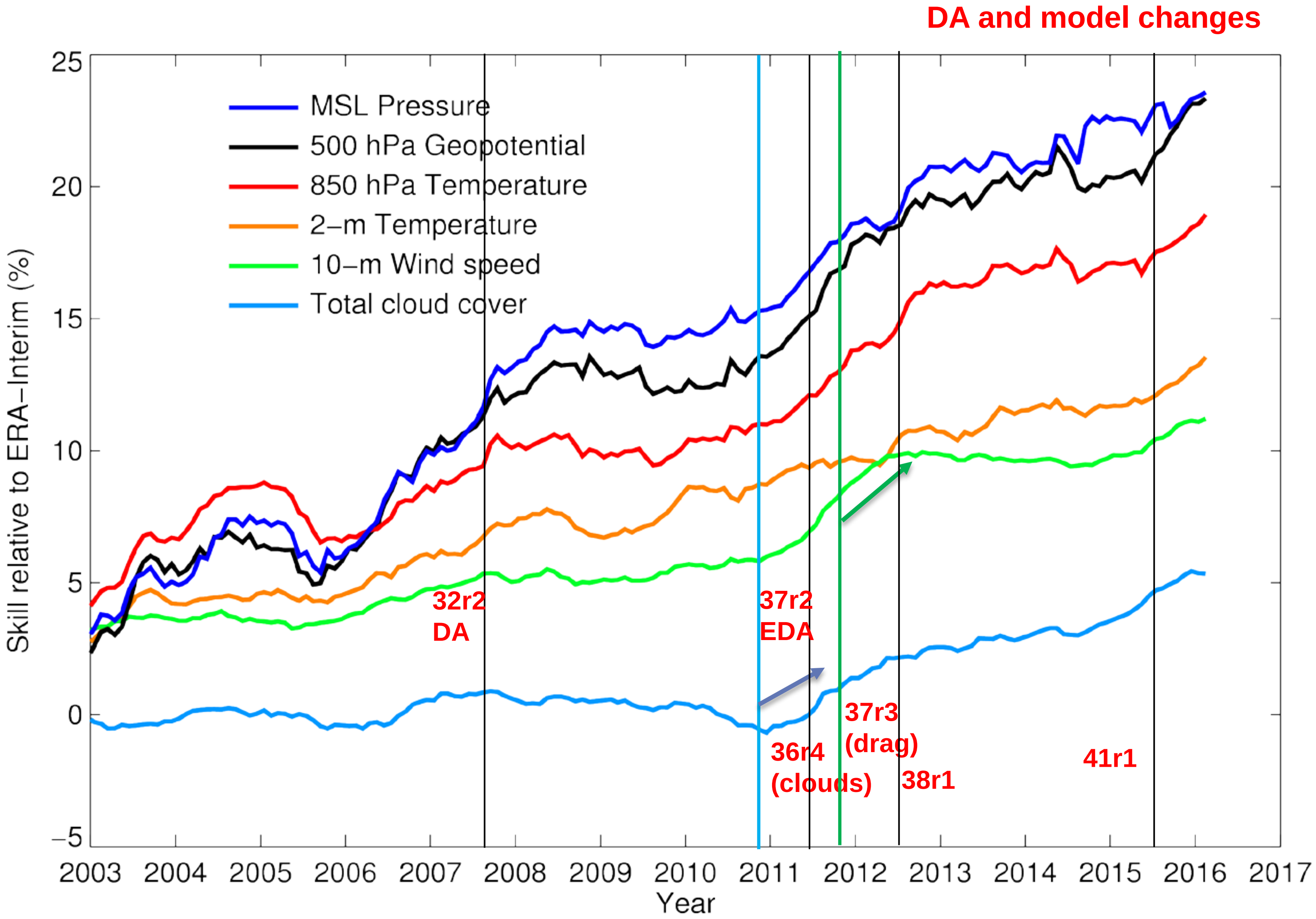
The model might need several cycles' upgrades of diagnostics (**D**) and retuning to improve forecast skill (**SK**).



Some evidence: impact of DA+model changes on fc skill

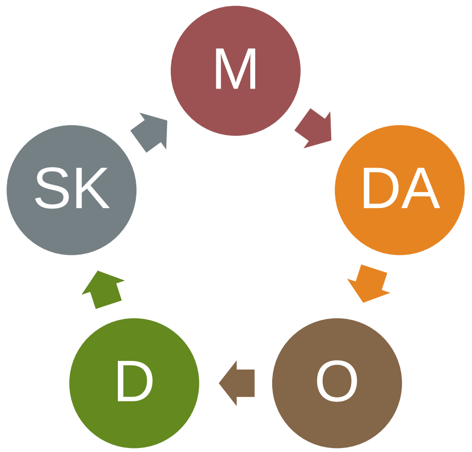
Resolution matters, although it could take time before investments pay off. The model needs to be retuned.

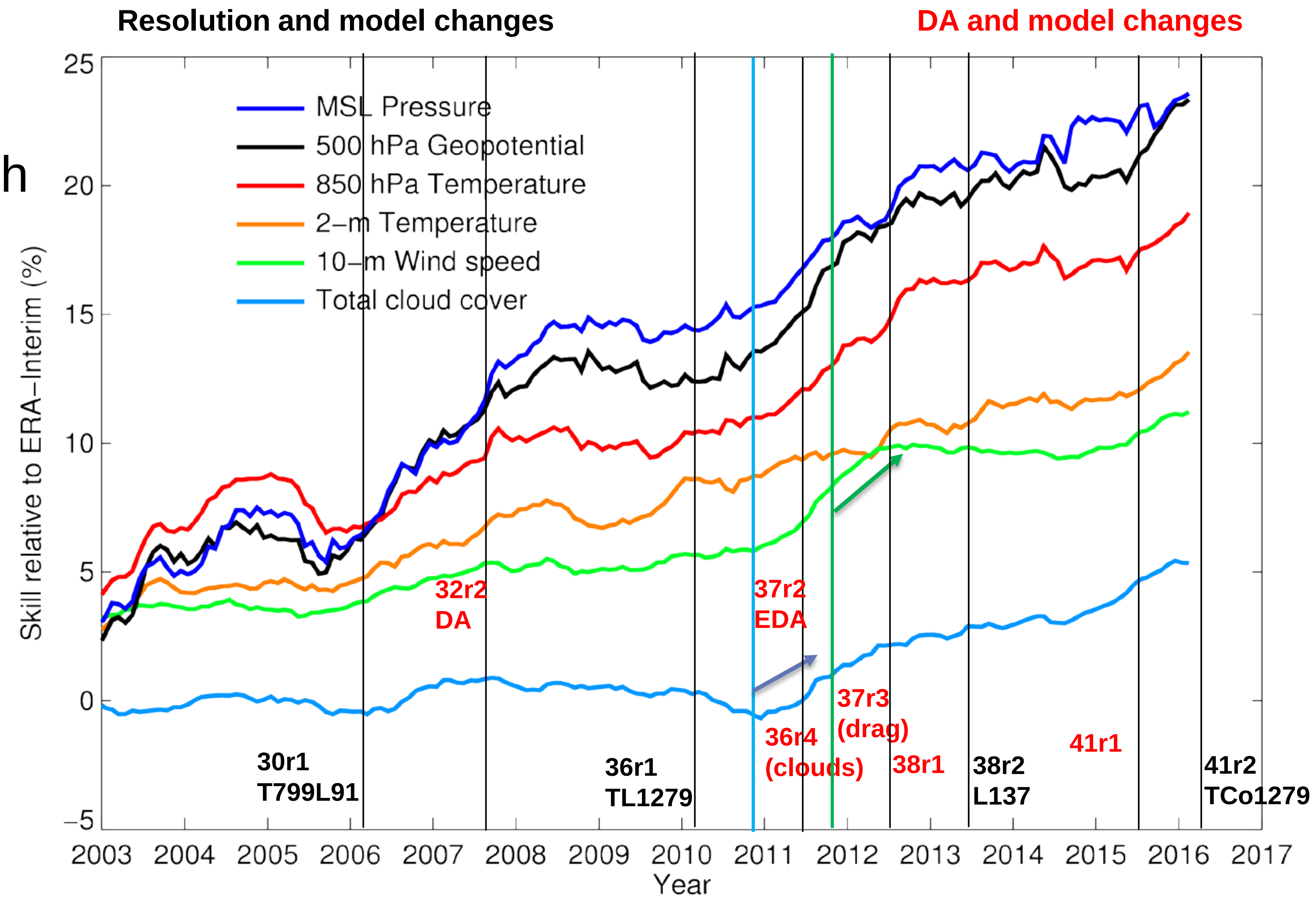
DA matters, although to fully exploit upgrades and the use of more/new obs, the model needs to be upgraded as well.



Some evidence: impact of all changes on fc skill

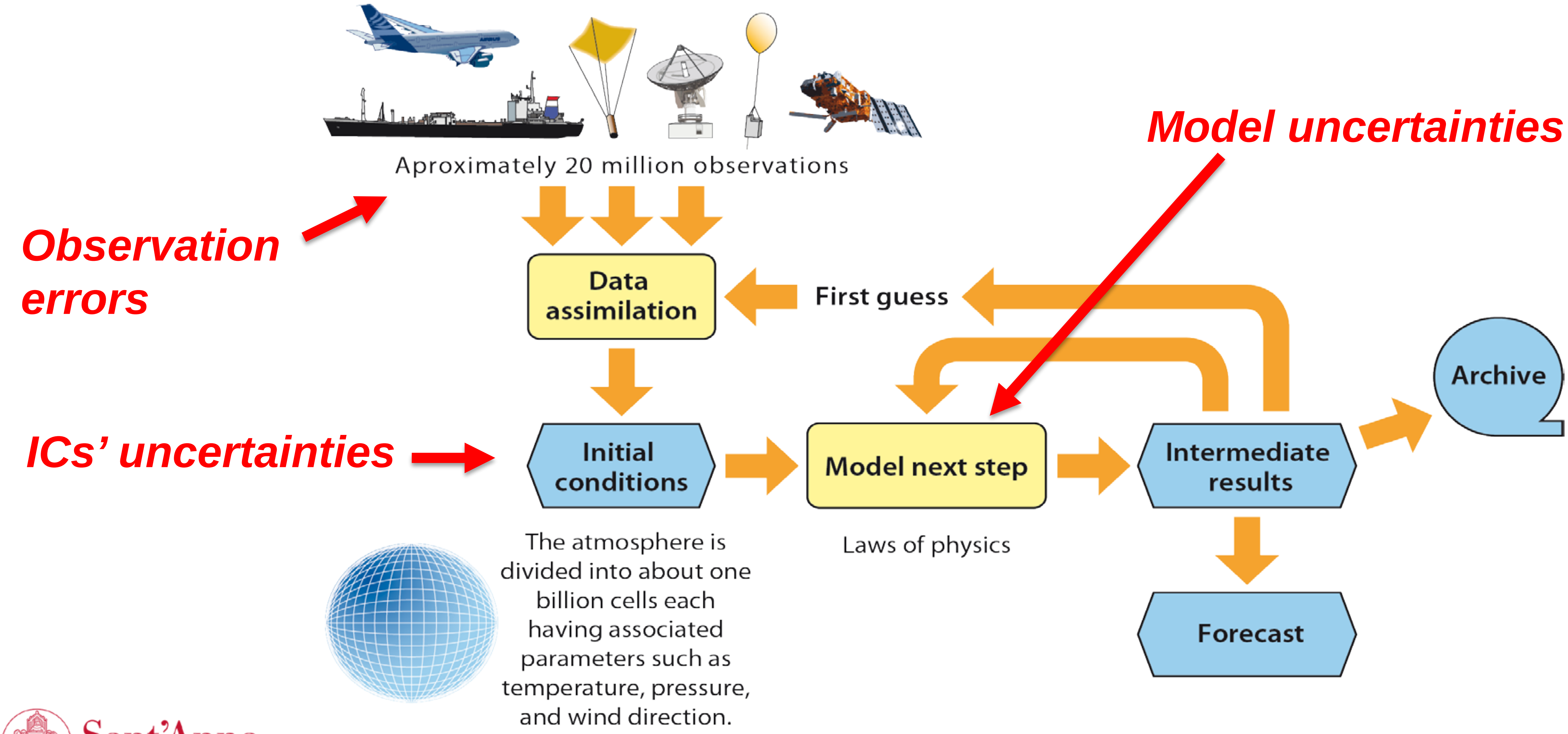
DA upgrades could lead to the use of more/better obs (**O**): this allows more thorough model (**M**) diagnostics (**D**), that leads to model improvements. DA+O+M+D leads to better skill (**SK**).





How can we build reliable ensembles?

The ECMWF ensembles aim to simulate all sources of errors.



Outline

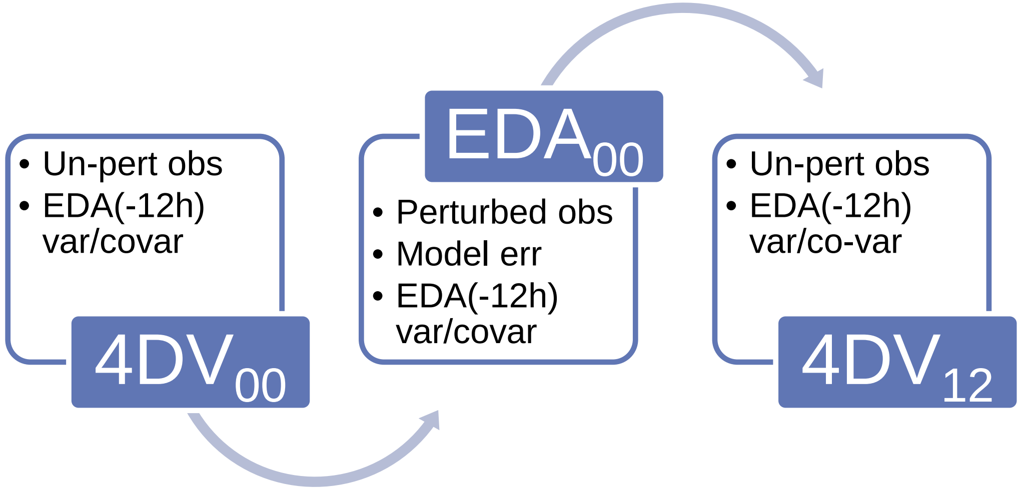
1. Having an accurate Data Assimilation system is essential
2. ECMWF has been moving towards ensembles of DAs and FCs
3. Coupling to relevant processes has led to improved forecasts
4. Re-forecasts are essential to extract predictable signals
5. Moving towards a seamless ensemble of coupled DAs and FCs
6. Conclusions

Key characteristics of the ECMWF ensembles

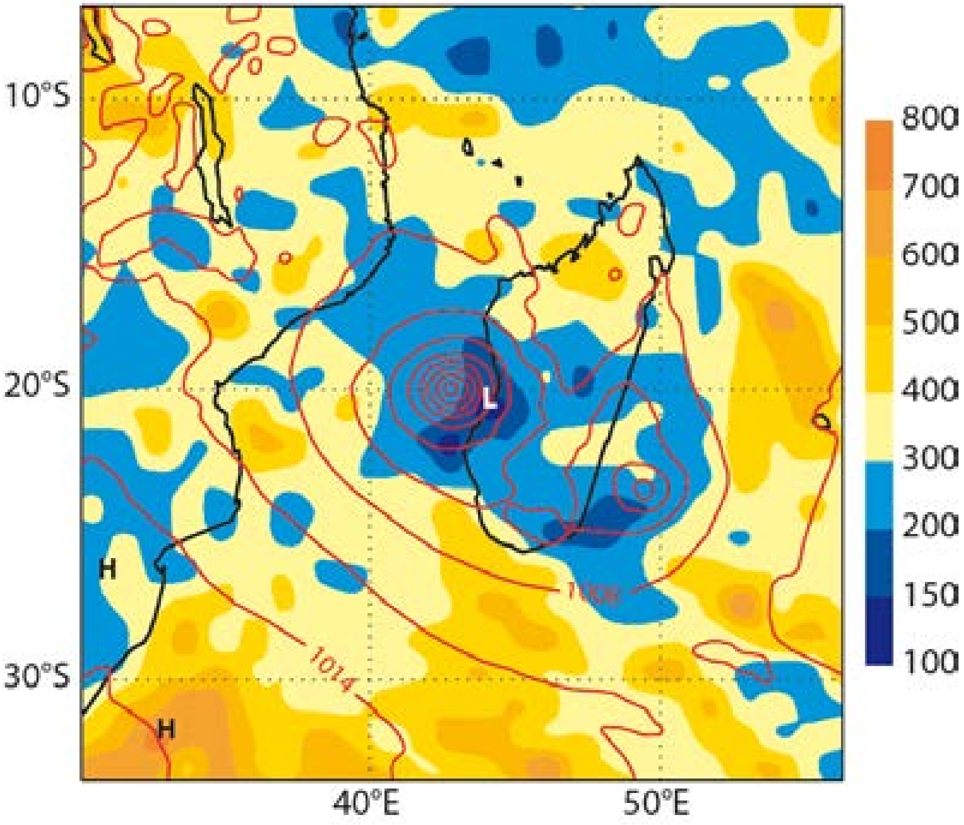
	Ensemble	#	IFS Resolution	Ocean Resolution	Sea-ice	OBS/IC unc	Model unc	Frequency
Analysis	**4DVAR**	1	Tco1279L137 (9km)	--	--	--	--	4 x day
	EDA	51	Tco639L137 (18km)	--	--	pOBS	SPPT	2 x day
	OCEAN5	5	--	0.25deg-z75 (25km)	LIM2	pOBS	--	1 x day
Forecasts	**HRES (d0-10)**	1	Tco1279L137 (9km)	0.25deg-z75 (25km)	LIM2	--	--	4 x day
	ENS (d0-15)	51	Tco639L91 (18km)	0.25deg-z75 (25km)	LIM2	SV+EDA	SPPT	2 x day (4 x day to d6.5)
	ENS (d15-46)	51	Tco319L91 (36km)	0.25deg-z75 (25km)	LIM2	SV+EDA	SPPT	2 x week
	SEAS5 (m0-7/13)	51	Tco319L91 (36km)	0.25deg-z75 (25km)	LIM2	SV	SPPT + SKEB	1 x month (4 x year to 13m)

The EDA is used to estimate background error statistics

The 51-member EDA is used to compute flow dependent background error statistics, and thus define the relative weight to give to the obs and the model first guess.



Background error correlation length scale for long(p_{msl}) and p_{msl}



(M Bonavita - ECMWF)

Outline

1. Having an accurate Data Assimilation system is essential
2. ECMWF has been moving towards ensembles of DAs and FCs
3. Coupling to relevant processes has led to improved forecasts
4. Re-forecasts are essential to extract predictable signals
5. Moving towards a seamless ensemble of coupled DAs and FCs
6. Conclusions

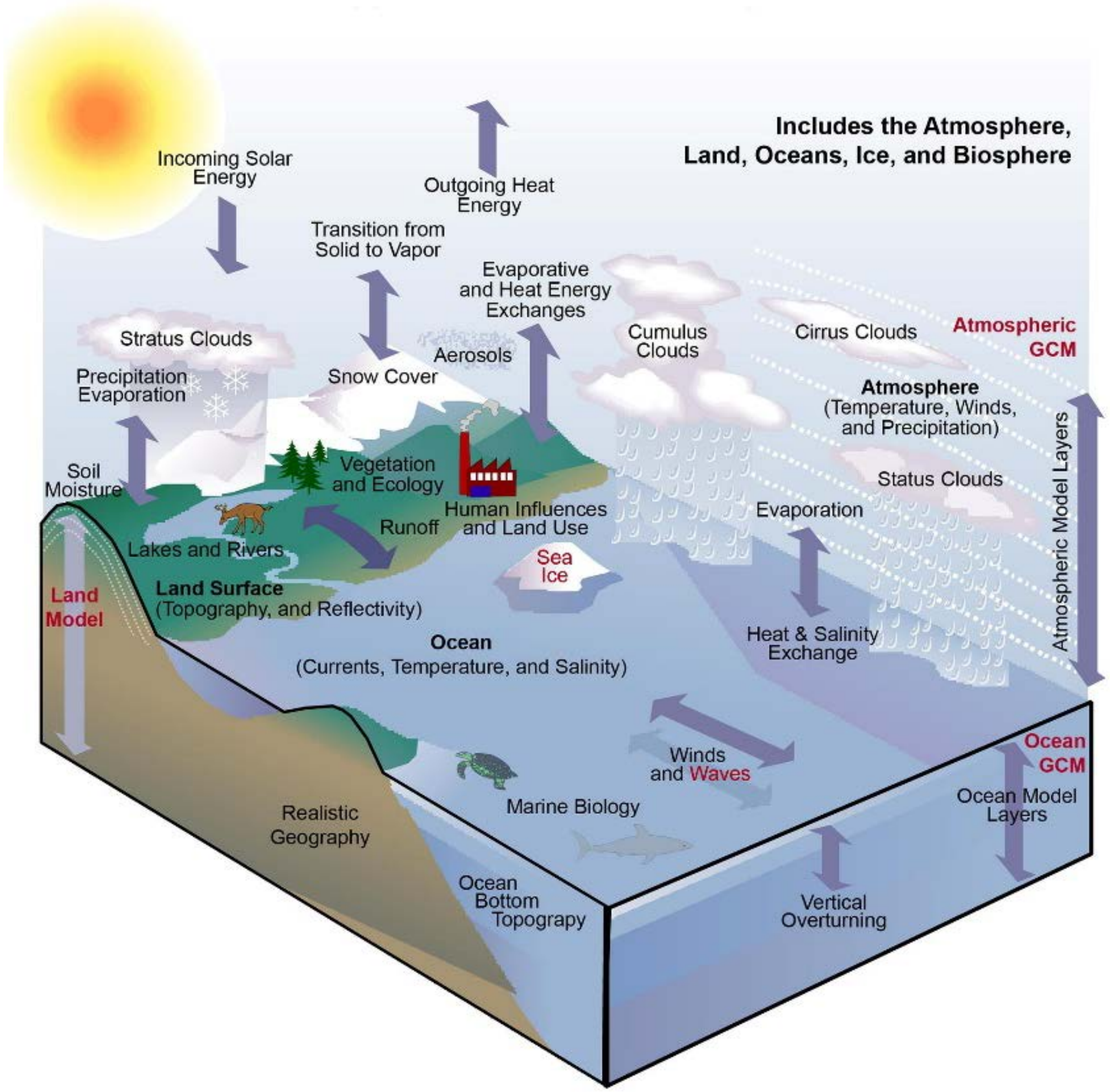
Ensemble design: towards coupled Earth-systems

- *Why?* To be able extract predictable signals from all key relevant processes
- *How?* By modelling them accurately and initialising them using all available observations

The ECMWF coupled model

ATMOSPH | 3D-OCEAN

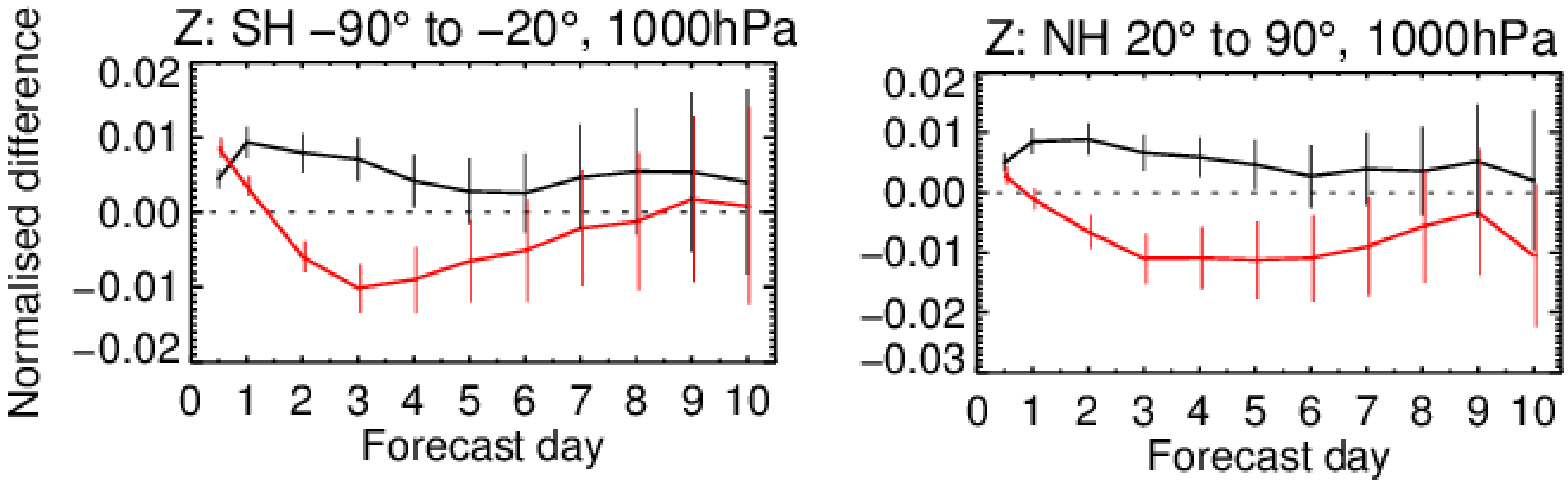
LAND | WAVES | SEA-ICE



Impact of coupling to a wave model

ECMWF has been using a 2-way coupled wave model since the 1990s:

- Removing the coupling to the wave model has a negative impact on the lower-level atmosphere (black line);
- Improving the wave model (heat-fluxes) can bring further improvements.

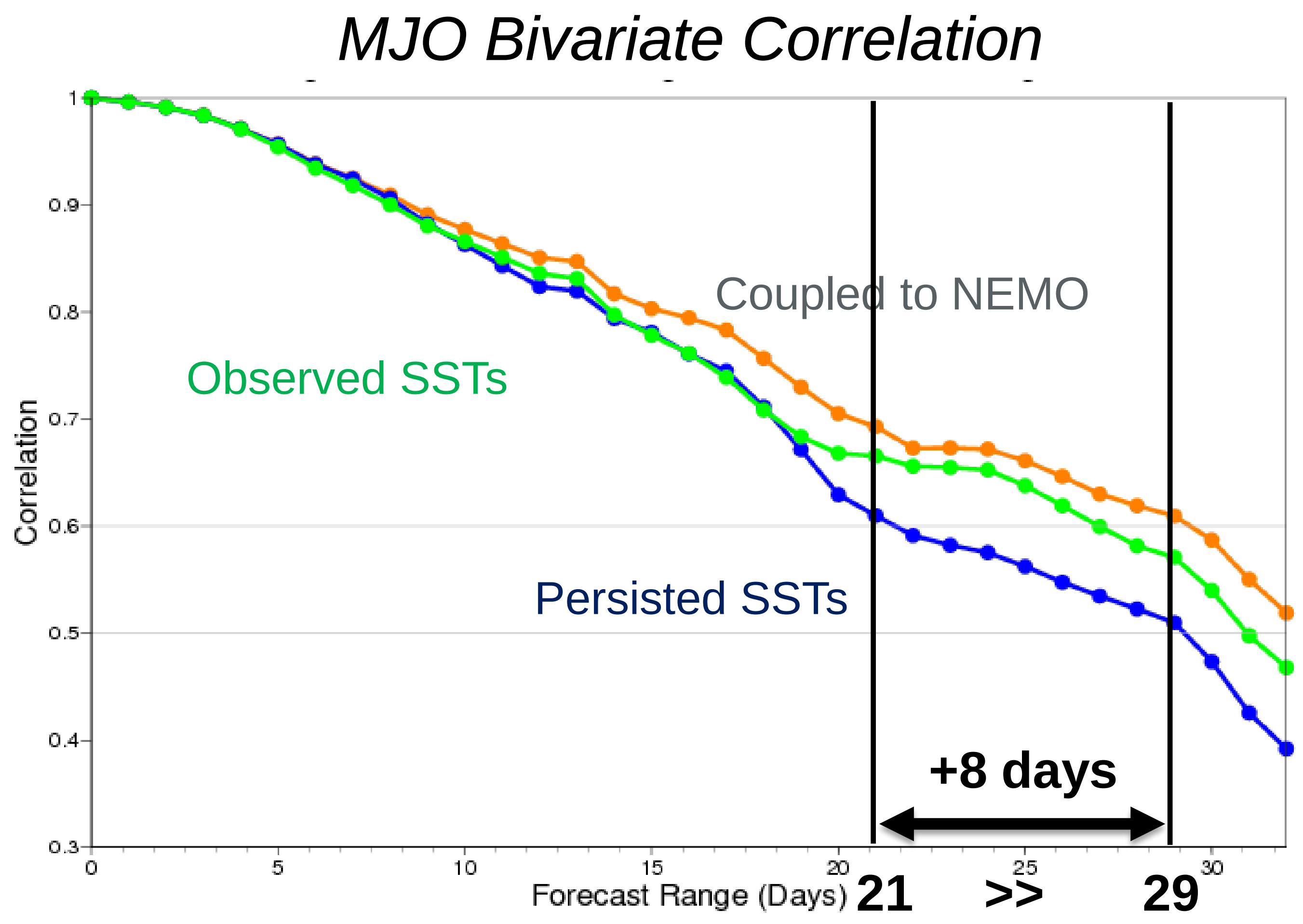


(J Bidlot - ECMWF)

Impact of coupling to a dynamic 3D-ocean

ECMWF started coupling its seasonal and monthly ensembles in the 2000s.

There is clear evidence that coupling improves the predictions of phenomena like the Madden Julian Oscillation.



(F Vitart - ECMWF)

Outline

1. Having an accurate Data Assimilation system is essential
2. ECMWF has been moving towards ensembles of DAs and FCs
3. Coupling to relevant processes has led to improved forecasts
4. Re-forecasts are essential to extract predictable signals
5. Moving towards a seamless ensemble of coupled DAs and FCs
6. Conclusions

Extremes populate the tails of the distributions

Extremes populate the tails of the distributions, and are rare. This implies:

- Prediction is difficult, since ensembles have **low resolution in probability space** (e.g. in weather, 50 members can resolve dp=2%);
- **Diagnostic and verification is also difficult** since they are rare, and each is different: building a solid verification/diagnostic statistics is impossible.

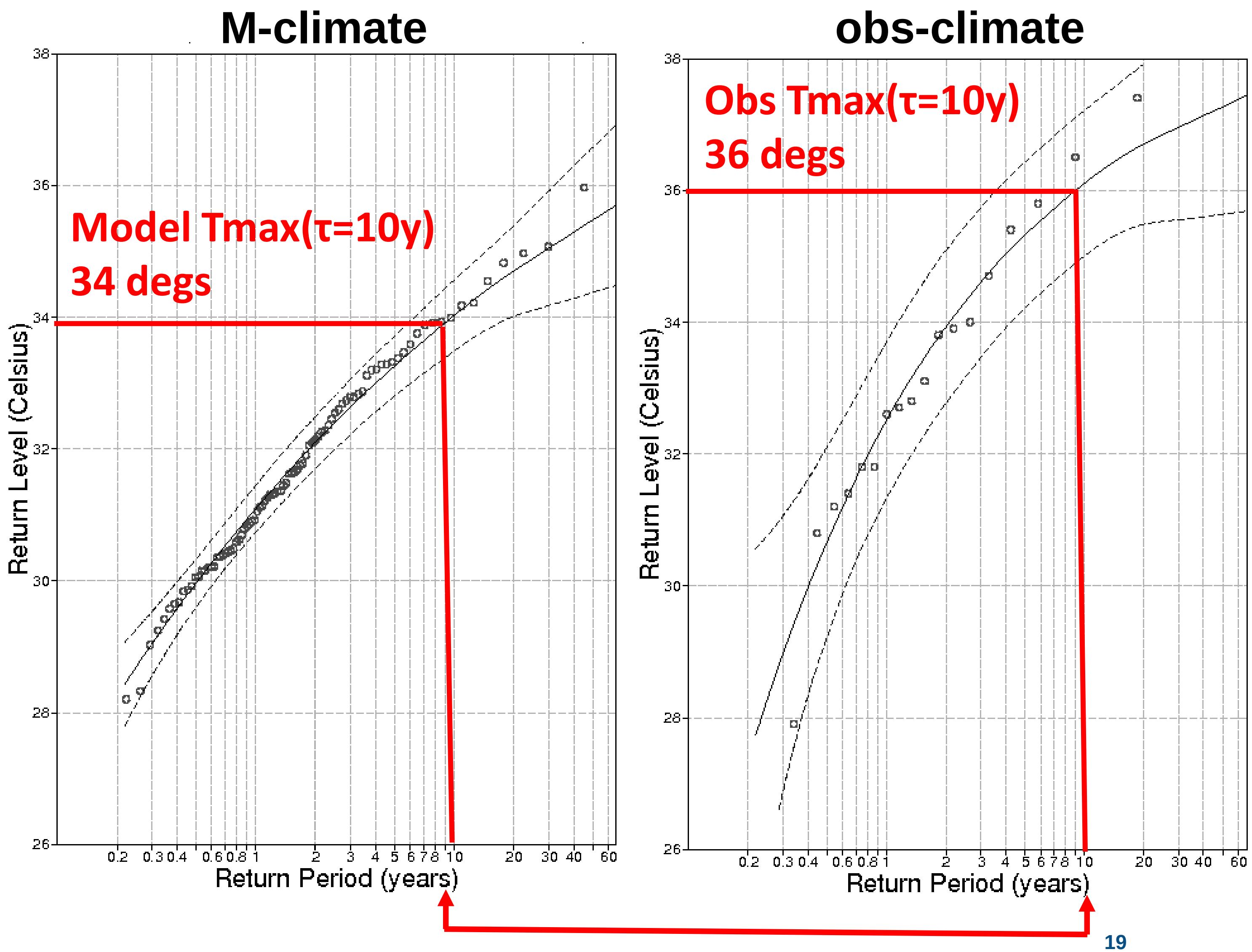
How can we extract signals? With extreme-value theory

‘Extreme-Value theory’ can be used to calibrated probabilistic forecasts (by mapping from/to the model to/from the observation spaces).

EVT is applied to compute the ‘Extreme Forecast Index’ (EFI) product.

(Prates and Buizza, QJRMS 2011)

TMAX for Hannover

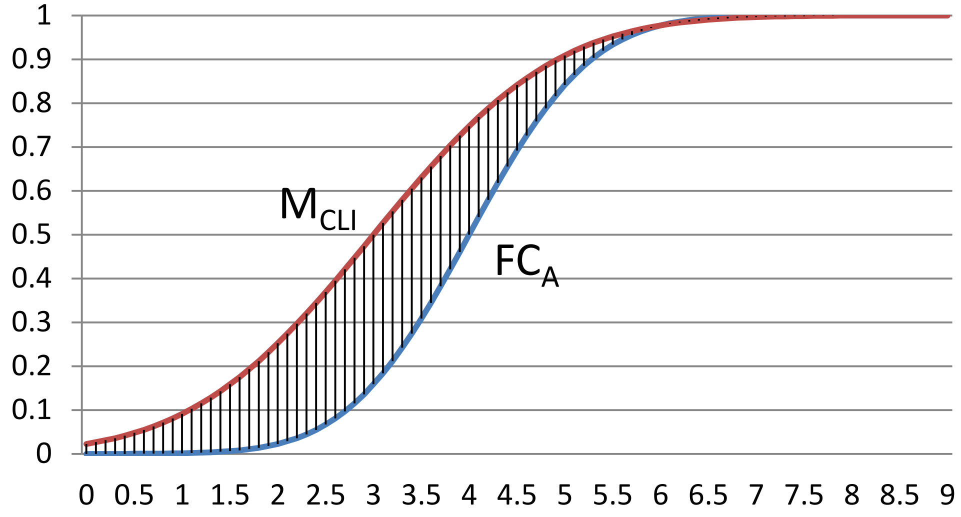


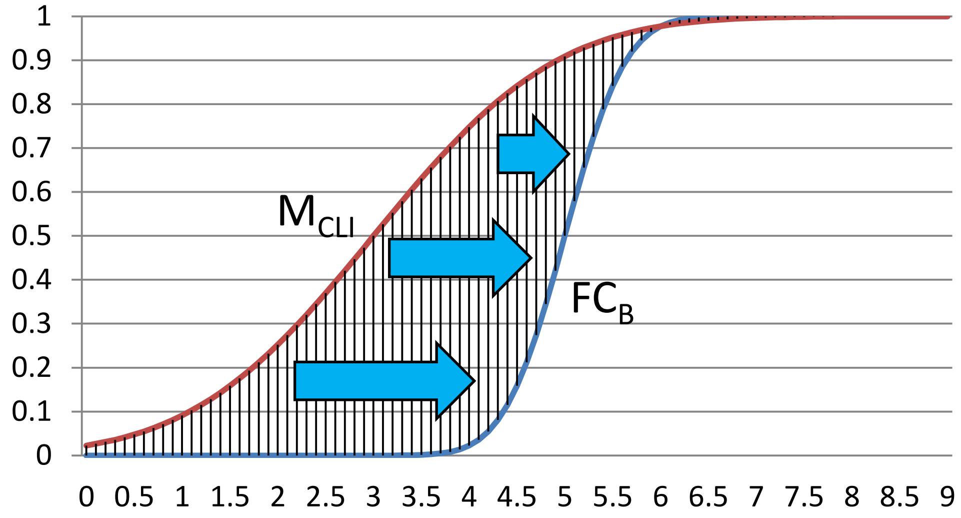
EFI: compares the CDFs of the fc and the model climate

Extreme Forecast Index (EFI):

- Compute the model climate M_{CLI} CDF and the latest available forecast CDF
- Compute the area between the two curves
- If the normalized area is close to 1, there is a high probability of an extreme event

$$EFI = \frac{2}{\pi}\int_0^1 \frac{p - CDF(p)}{\sqrt{p(1-p)}}\,dp$$

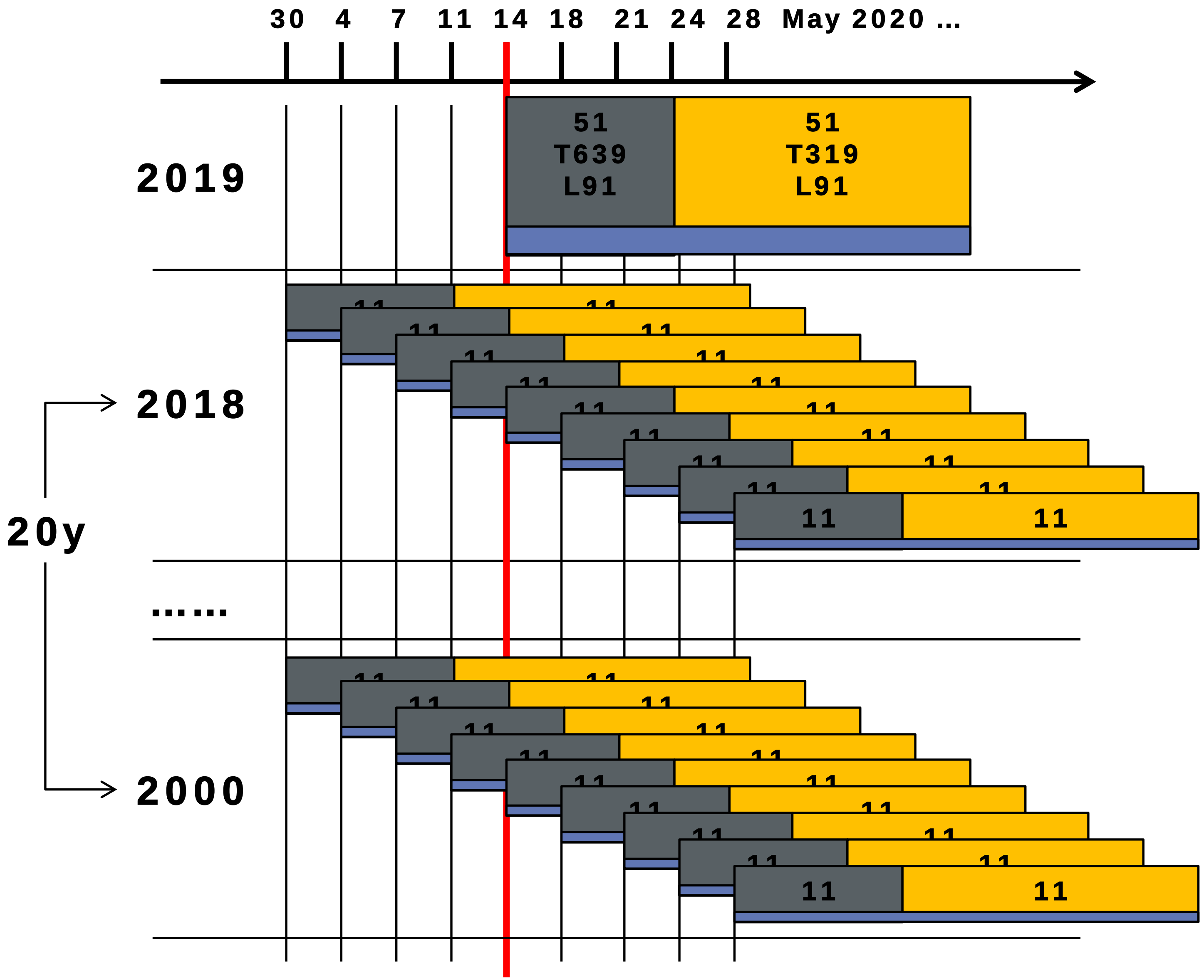




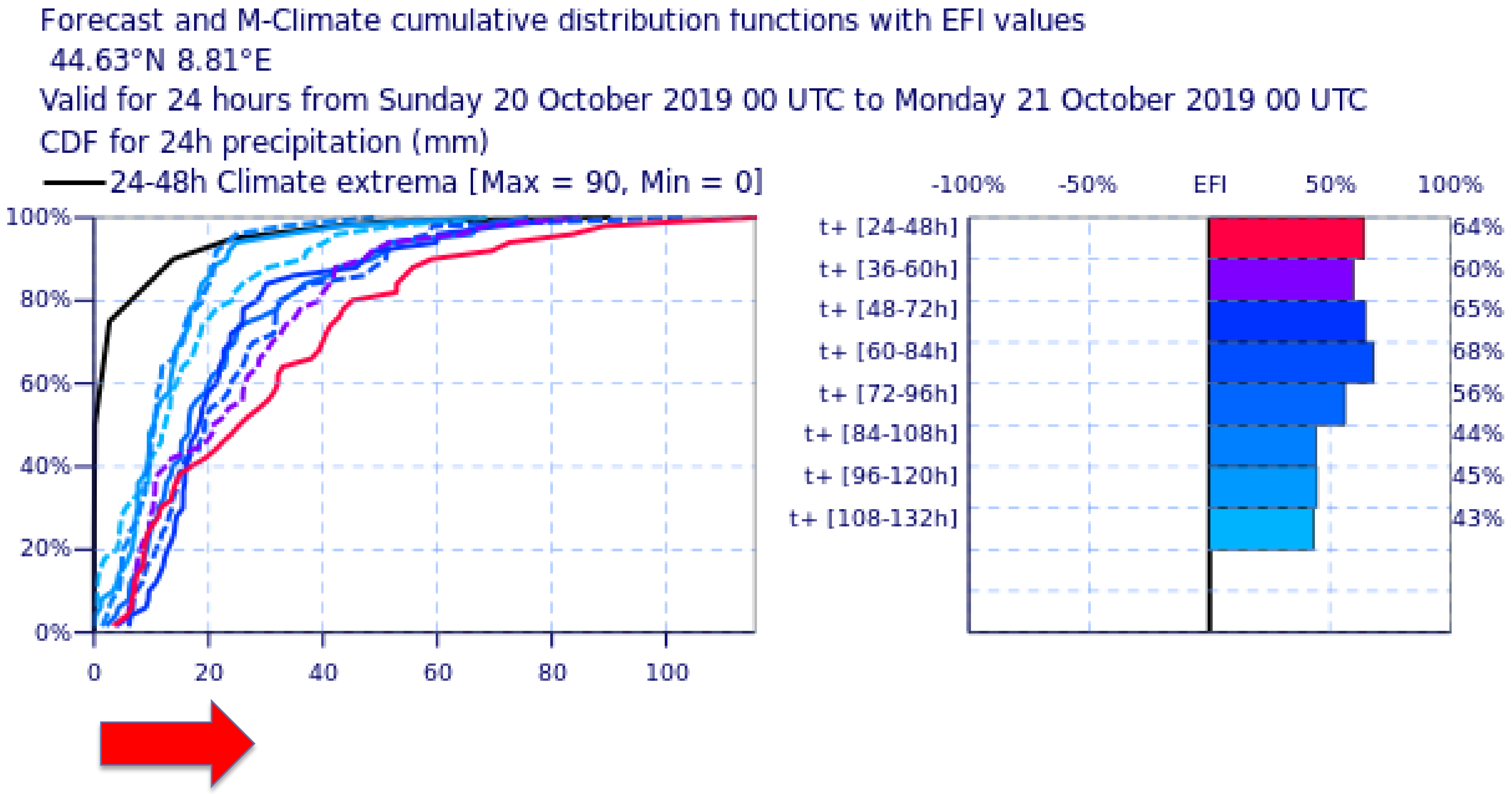
How can we estimate the model climate? With re-fcs

Reforecasts are smaller-size ensembles, run twice-a-week, for the past 20 years.

Suppose today is 14 May 2020. The model climate M_{CLI} is computed from the five 11-member ensembles with a starting calendar date closest to the current one (14 May) for the past 20 years (i.e. 1100 members).

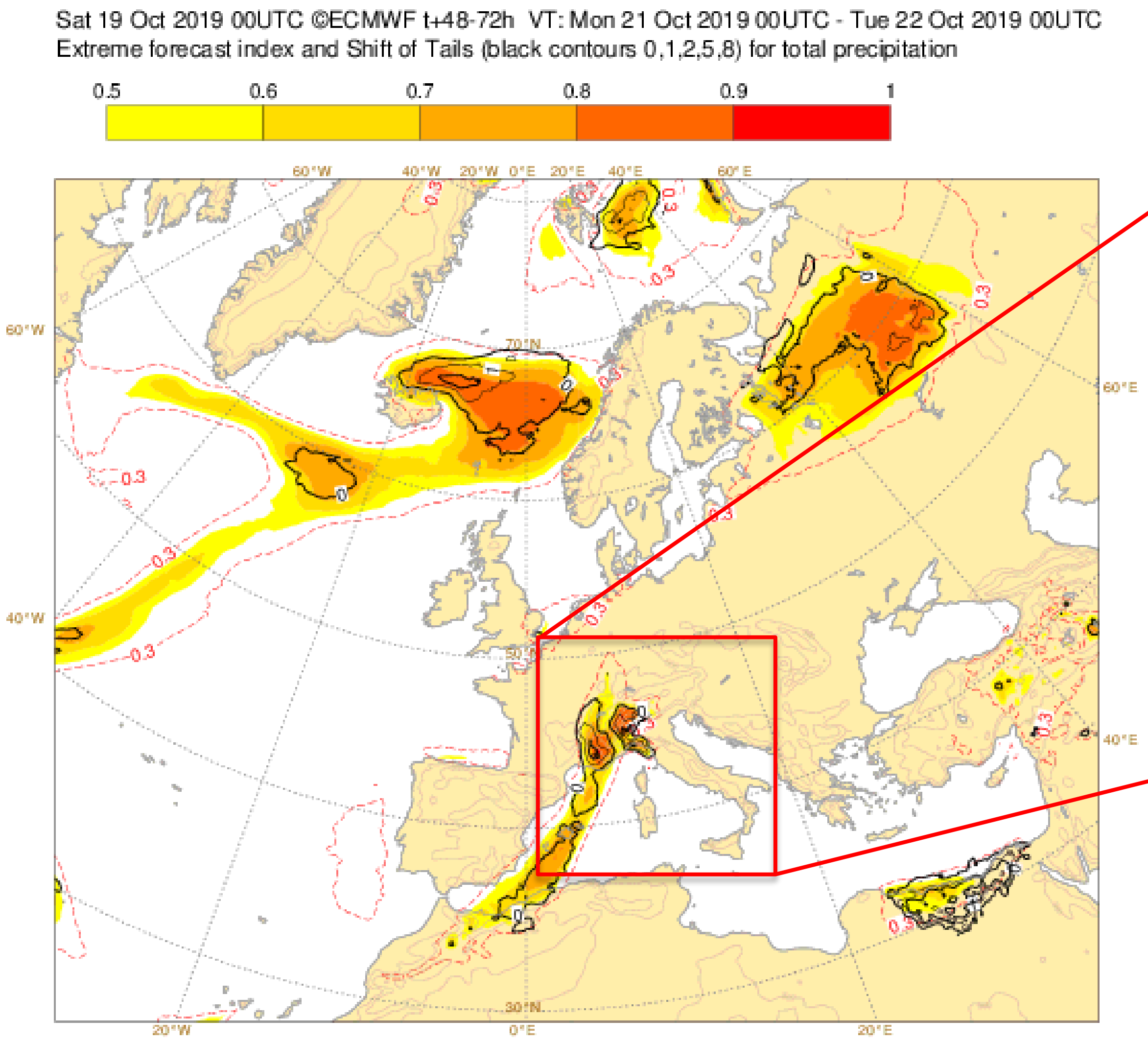


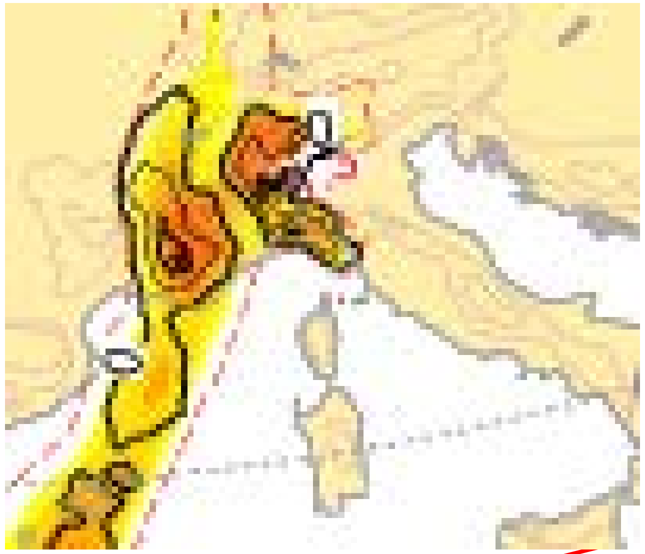
An example: EFI precipitation for Gavi Italy (21/10/19)



EFI maps can be used to identify risk areas

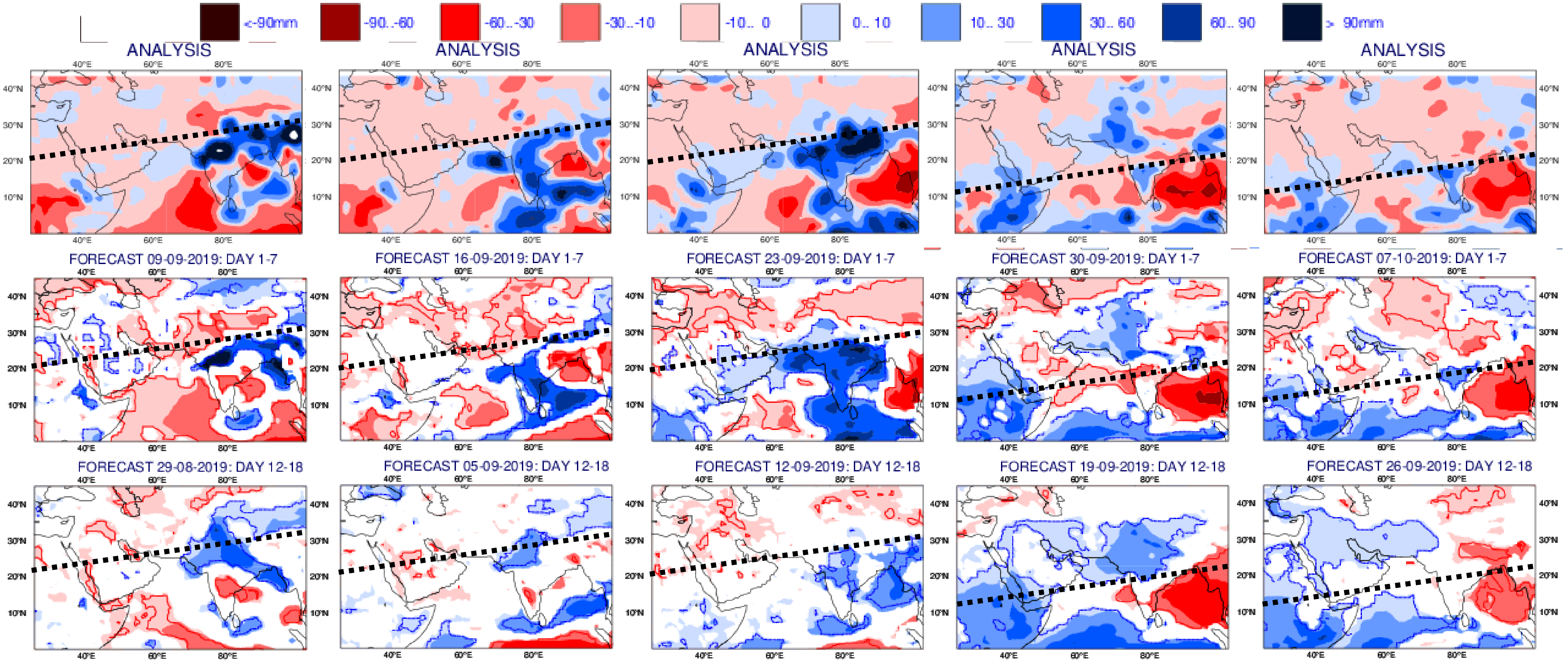
ECMWF ENS Extreme Forecast Index – 19/10@00+48-72h





ENS weekly-average precipitation +d1-7 & +d12-18

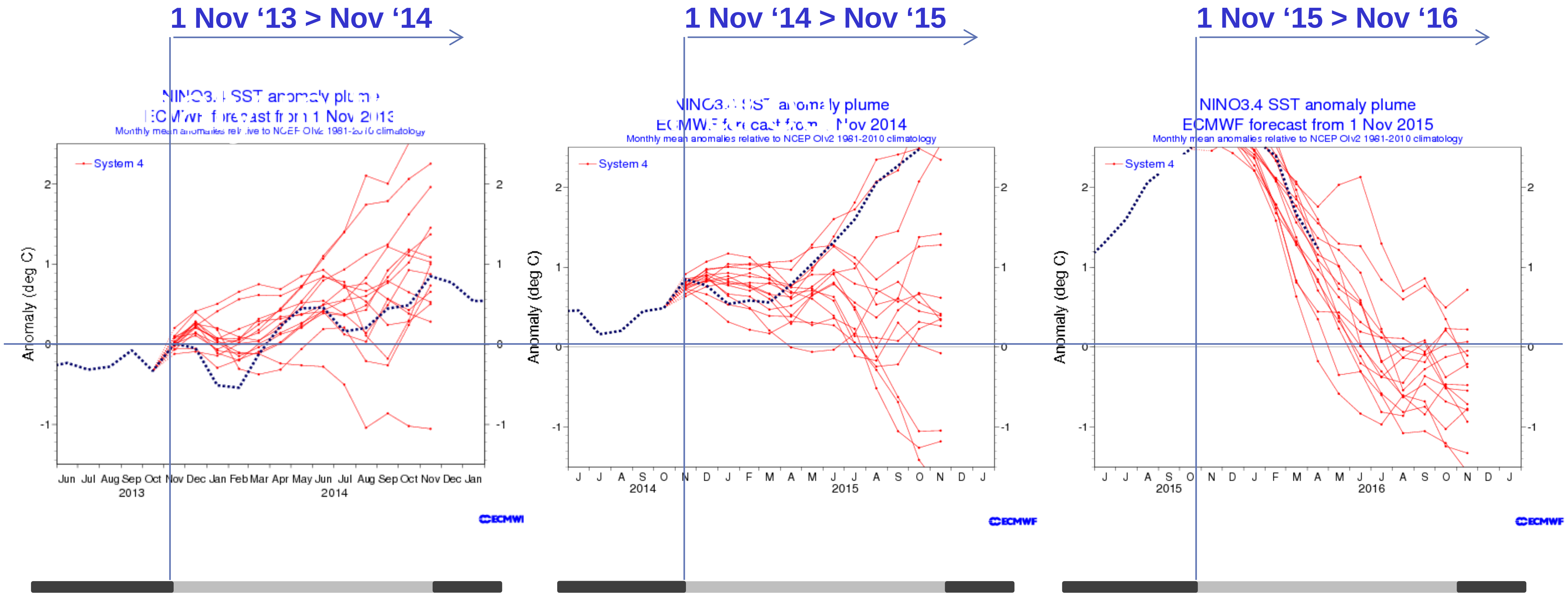
20190909 to 0915 - **0916 to 0922** **0923 to 0929** **0930 to 1006** **1007 to 1013**



Sant'Anna School of Advanced Studies – Pisa

Seasonal range: ENS +1y fc of El Nino

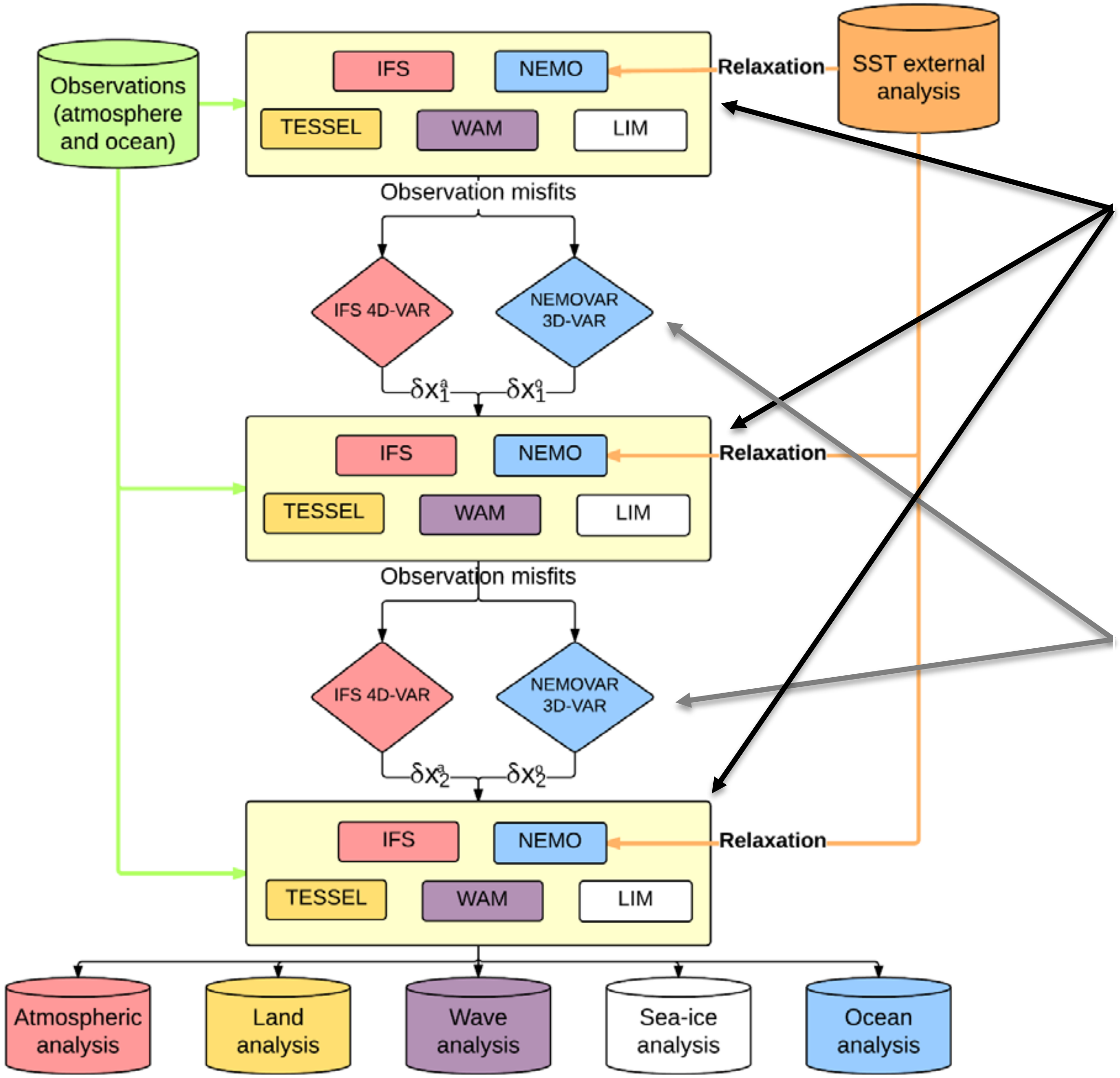
The tropics is the area where seasonal fcs have the highest skill, as indicated e.g. by the accuracy of 1-year forecasts of SST anomaly in the Nino3.4 area.



Outline

1. Having an accurate Data Assimilation system is essential
2. ECMWF has been moving towards ensembles of DAs and FCs
3. Coupling to relevant processes has led to improved forecasts
4. Re-forecasts are essential to extract predictable signals
5. Moving towards a seamless ensemble of coupled DAs and FCs
6. Conclusions

The design of CERA, the ECMWF coupled DA



The coupled model computes observation misfits in each outer iteration

The SST computed in NEMO and constrained by relaxation

The atmospheric and ocean increments are computed in parallel to correct the initial state

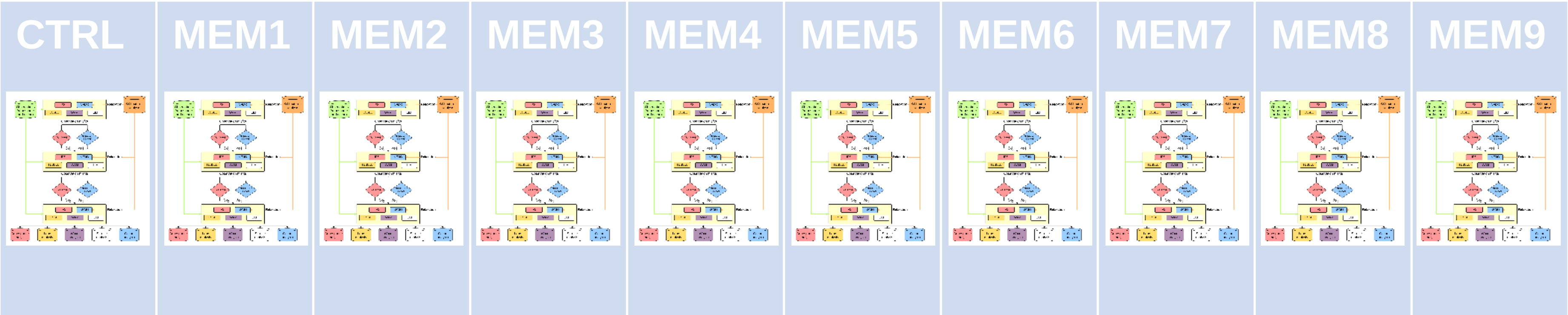
The analysis is dynamically consistent with the coupled model

(P Laloyaux - ECMWF)

CERA20C: an ensemble of coupled climate re-analyses

CERA-20C has been generated with a 10-member EDA using only surface p and ocean surface W from conventional instruments. The ensemble provides a measurement of uncertainty in the climate reconstruction, and flow-dependent background error statistics for the atmosphere assimilation. Each member is generated by a 4D-var cycle that simulates 3 sources of uncertainties:

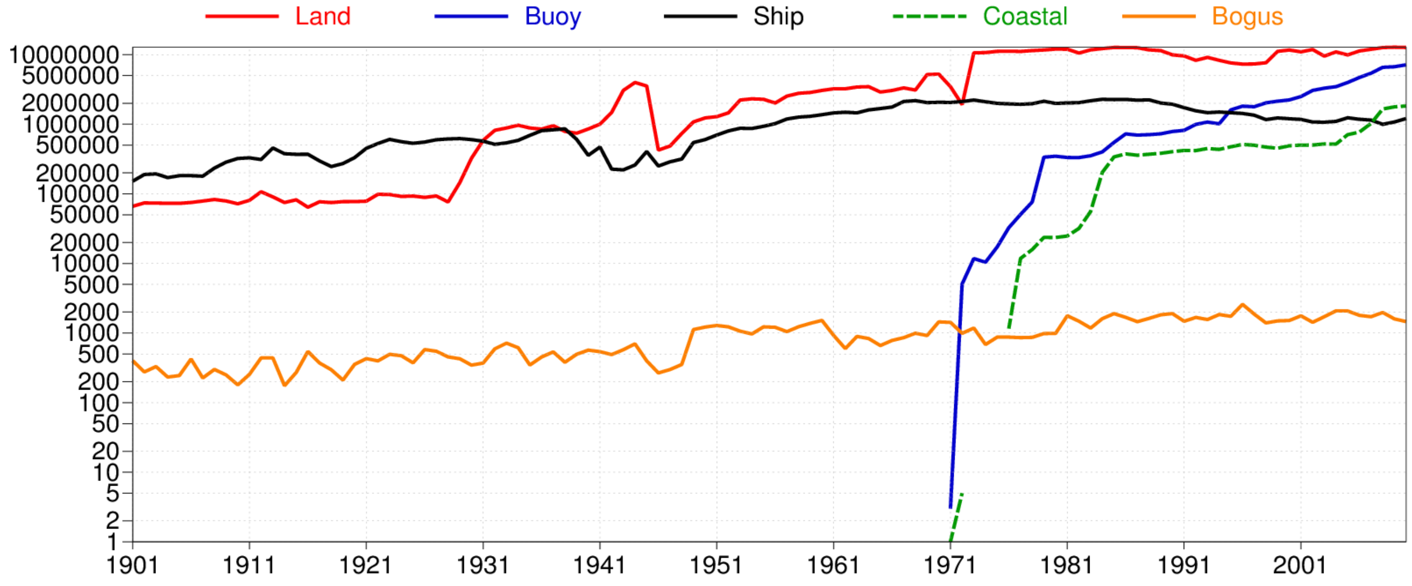
Observation uncertainties + **Model uncertainties** + **Model forcing uncertainties** = **Reanalysis uncertainties**



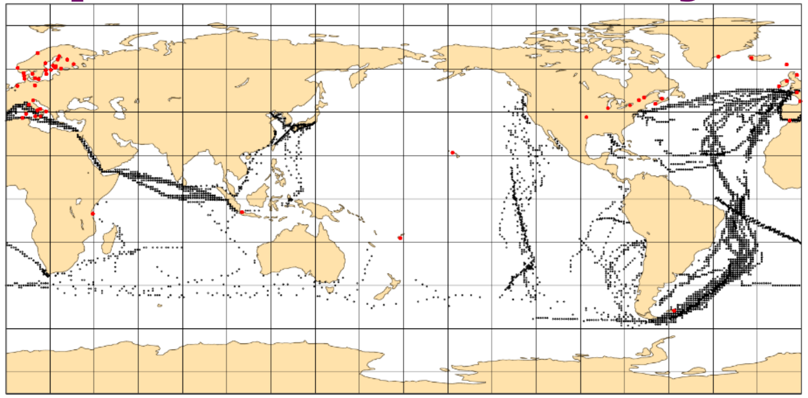
(P Laloyaux - ECMWF)

The ensemble can help estimating the past uncertainty

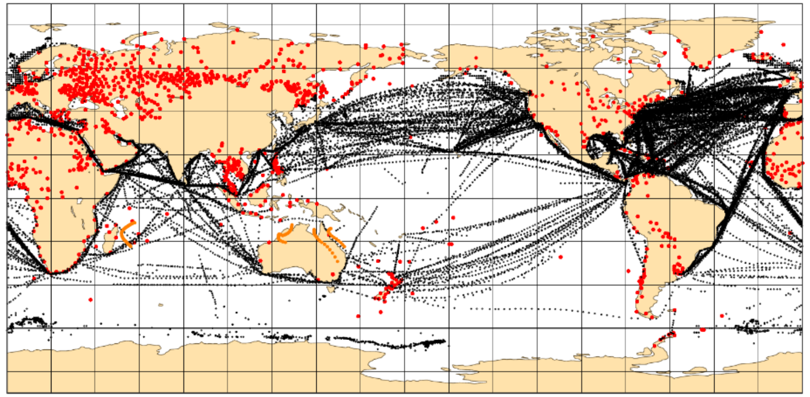
Yearly count of MSLP assimilated obs (log. scale)



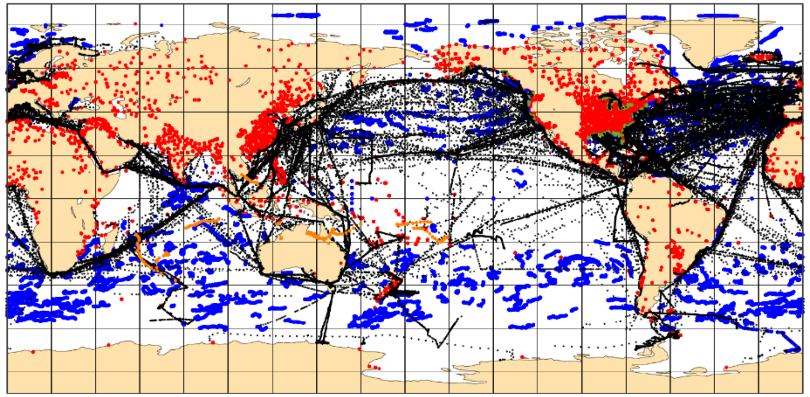
1901



1950



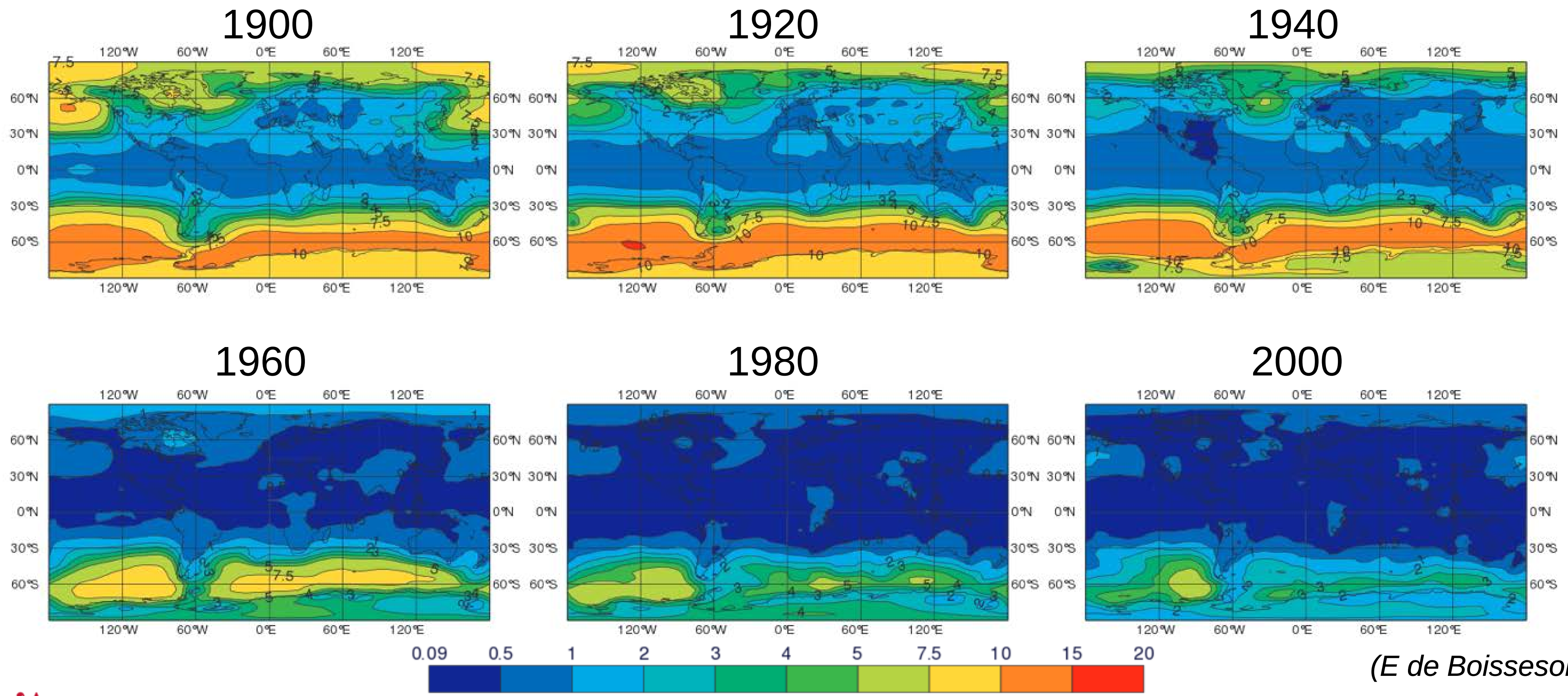
2010



(P Dahlgren - ECMWF)

CERA20C: ENS spread as a confidence measure

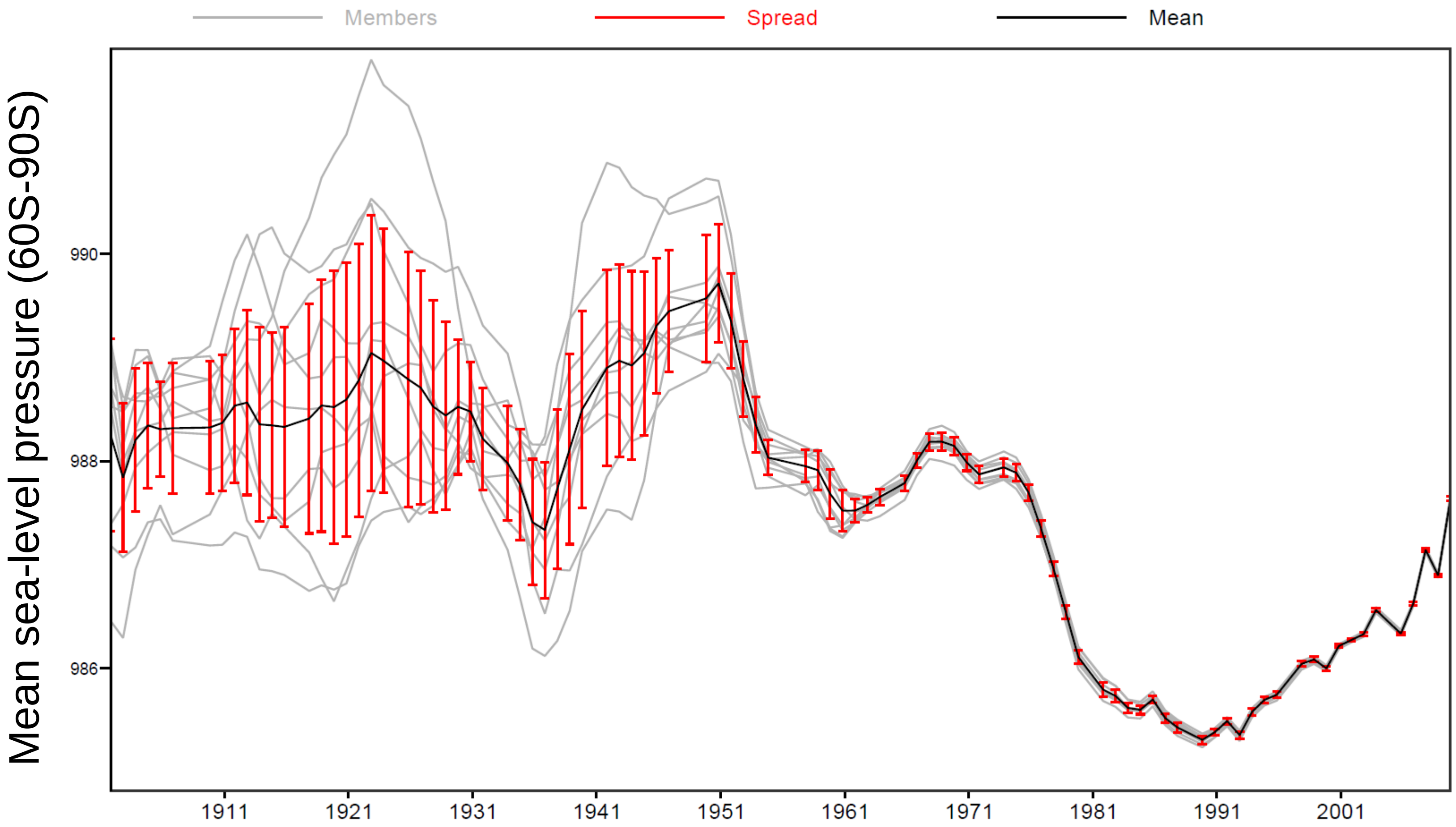
As the system ingests more obs, it is learning from them and the backgrounds become more accurate. At the beginning of the century the background error ranges between 1-15 hPa, while at the end of the century it is between 1-3 hPa.



(E de Boisseson - ECMWF)

CERA20C: 110y of ensemble coupled reanalyses

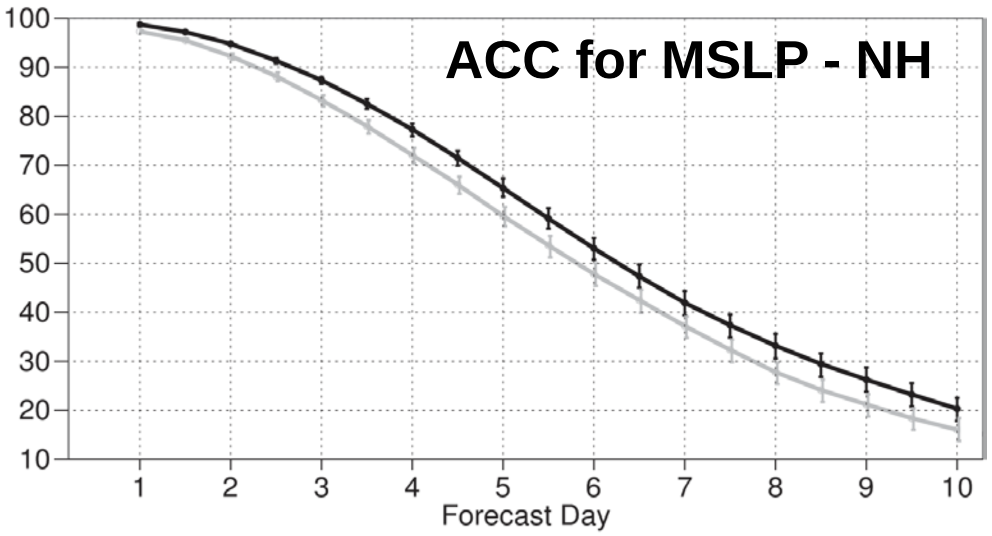
CERA-20C (EU-FP7 ERACLIM2 proj): a 10-member ensemble of coupled reanalyses spanning 110 years (1901-2010). CERA20C gives an estimate of the uncertainty of the past climate (more/better obs lead to smaller spread).

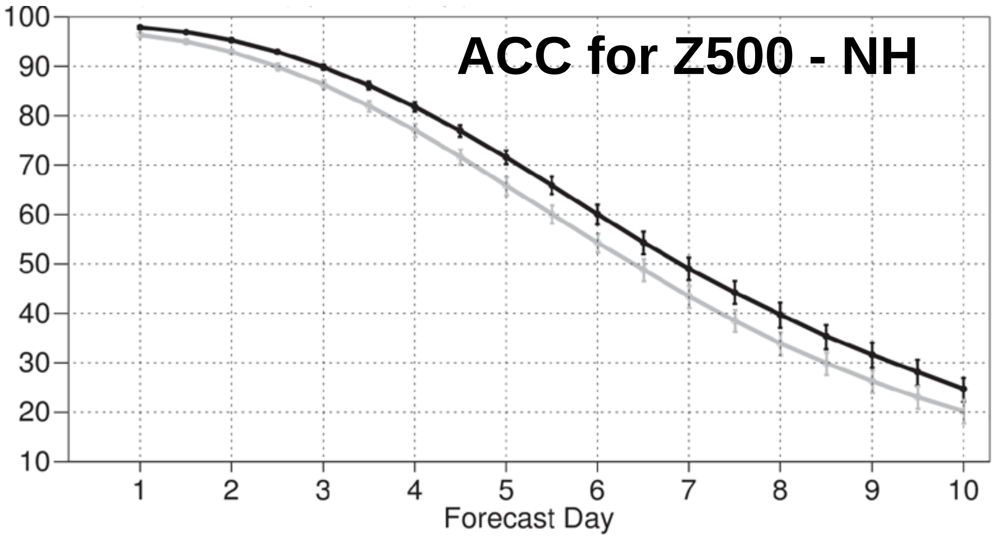


(P Dahlgren - ECMWF)

An example: does coupling have any impact?

CERA-20C and **ERA-20C** forecast scores with respect to ERA-Interim (year 2004) improved by about 0.5 day in CERA-20C (surface and troposphere).





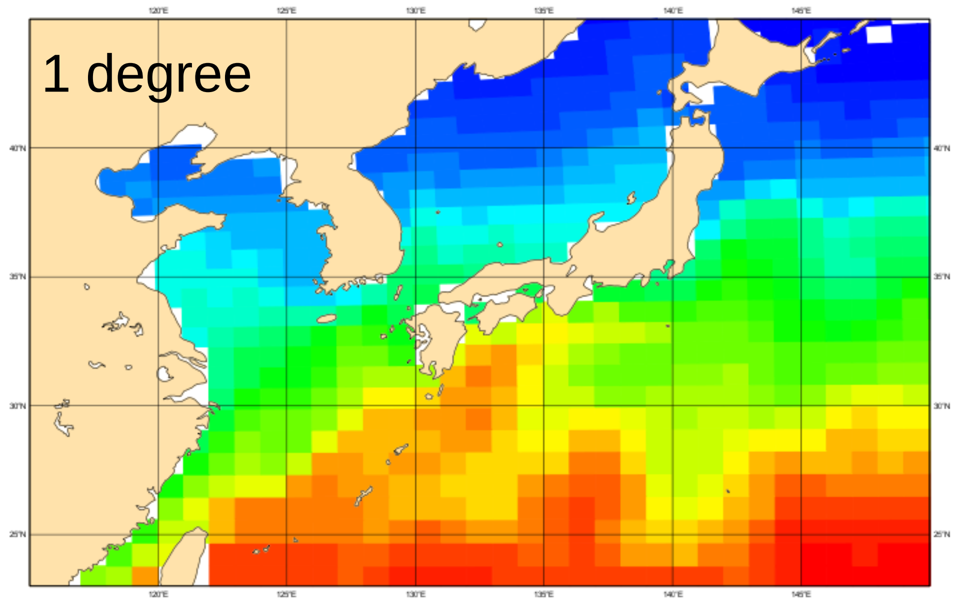
Reasons of the improvement	ERA-20C	CERA-20C
background error	static from operations	10-member EDA
forecast model	IFS CY38R1	IFS-NEMO CY41R2
assimilation method	uncoupled	coupled (CERA)

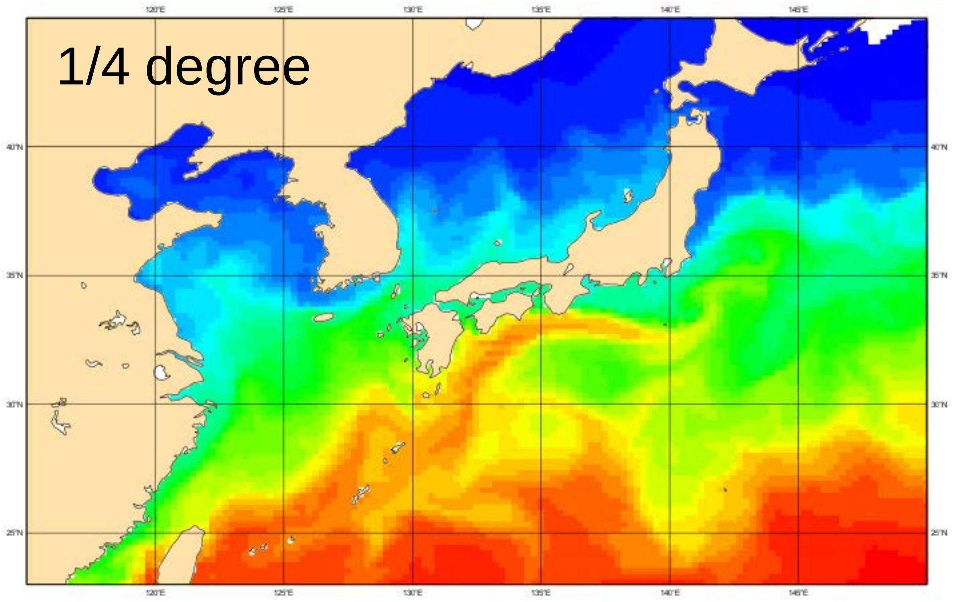
(D Schepers - ECMWF)

CERA-SAT: the prototype operational coupled-EDA

The CERA DA system has been used to generate CERA-SAT, a coupled ensemble reanalysis of 10 years of the satellite era (2008-to-2018). It used the same resolution as ERA5:

- Atmosphere (IFS): 65km
- Ocean (NEMO): ¼ degree (75 levels)





(D Schepers - ECMWF)

Outline

1. Having an accurate Data Assimilation system is essential
2. ECMWF has been moving towards ensembles of DAs and FCs
3. Coupling to relevant processes has led to improved forecasts
4. Re-forecasts are essential to extract predictable signals
5. Moving towards a seamless ensemble of coupled DAs and FCs
6. Conclusions

Areas of active developments

There is still a lot of work to do to be able to predict precipitation few weeks and months ahead! Areas of active developments include:

1. Improving the **ensemble approaches** (the ‘ensemble model’, i.e. the processes and the simulation of the model uncertainties)
2. Improving the estimation of the **initial PDF** (better assimilation systems and coupled DA, including the assimilation of more observations and a better simulation of the obs uncertainties)
3. Building **more consistent ensembles of analyses and forecasts**
4. Increasing **ensembles’ resolution** (to resolve better the small scales, the fast processes, and thus improve the prediction of the extremes)
5. **Understanding predictability** (how can we extract predictable signals?)

Conclusions
Always remember: an ensemble is better than one, even if it is special!!