Impact of model uncertainty on seasonal forecast quality: the ENSEMBLES project

F. J. Doblas-Reyes

Catalan Institute of Climate Sciences (IC3), Barcelona, Spain

with contributions from, among others, A. Weisheimer, T. Palmer, M. Balmaseda (ECMWF), J. Murphy, D. Smith (Met Office), N. Philippon (Univ. of Bourgogne), C.A.S. Coelho (CPTEC)

The ENSEMBLES project

- ENSEMBLES is a EU-funded project active from 2004 to 2009, with ~70 partners, that worked on the development of an EPS across space and time scales.
- The main objective was to carry out global and regional dynamical simulations across time scales (seasonal, decadal and longer), with the aim of providing high-resolution climate information to specific applications.
- An innovative and comprehensive seasonal-to-decadal (s2d) experiment was one of the main deliverables.

ENSEMBLES s2d experiment

- Model uncertainty is a major source of forecast error. Three approaches to deal with the impact of model uncertainty on forecast error were investigated in ENSEMBLES: multi-model (ECMWF, GloSea, DePreSys, Météo-France, IfM-Kiel, CERFACS, INGV), stochastic physics (ECMWF) and perturbed parameters (DePreSys_PP).
- Hindcasts in two streams:
  - Stream 1: hindcast period 1991-2001, seasonal (7 months, May and November start date), annual (14 months, November start date), 9-member ensembles, ERA40 initialization in most cases; DePreSys (IC and PP ensembles) 10-year runs in every instance.
  - Stream 2: As in Stream 1 but over 1960-2005, with 4 start dates for seasonal hindcasts (Feb, May, Aug and Nov start dates), at least 1 for annual (Nov start date) and at least one 3-member decadal hindcast every 5 years; DePreSys_PP 10-year runs once a year.

ENSEMBLES s2d experiment

partner	Atmospheric model; resolution	Ocean model; resolution	initialization		external forcing	Additional components, comments	references
			atmosphere and land	ocean			
ECMWF	IFS CY31R1; T159/L62	HOPE; 0.3º-1.4º/L29	ERA-40/oper. analysis, atmospheric singular vectors	wind stress perturbations to generate ensemble of ocean reanalyses; SST perturbations at initial time	observed global well-mixed GHGs and sulphate aerosol and A1B from 2000, observed solar activity, no volcanic aerosol nor ozone	Operational Seasonal Forecasting system S3	*Stockdale et al.* (2010)*; Balmaseda et al.* (2008)
	IFS CY33R1; T159/L62	HOPE; 0.3º-1.4º/L29	ERA-40/oper. analysis, atmospheric singular vectors	wind stress perturbations to generate ensemble of ocean reanalyses; SST perturbations at initial time	“	used for the decadal hindcasts only	*Bechtold et al.* (*2008*)
	IFS CY35R2; T159/L62	HOPE; 0.3º-1.4º/L29	ERA-40/oper. analysis, atmospheric singular vectors	wind stress perturbations to generate ensemble of ocean reanalyses; SST perturbations at initial time	“	used with the stochastic physics approach	*Palmer et al.* (*2009*)
UKMO	HadGEM2-A; N96/L38	HadGEM2-O; 0.33º-1º/L20	ERA-40/oper. analysis, anomaly assimilation for soil moisture	wind stress perturbations to generate ensemble of ocean reanalyses; SST perturbations at initial time	observed global well-mixed GHGs, ozone and sulphate aerosol emissions and A1B from 2000, persisted solar activity and volcanic aerosol	fully interactive sea ice module	*Collins et al.* (2008)
	HadAM3; 3.75x2.5°	HadOM	anomaly assimilation of ERA-40/oper. analysis	anomaly assimilation of an ocean reanalysis	“	perturbed-parameter ensemble	*Smith et al.* (2007)
MF	ARPEGE4.6; T63	OPA8.2; 2°/L31	ERA-40/oper. analysis	wind stress, SST and water flux perturbations to generate ensemble of ocean reanalyses	observed global well-mixed GHGs and sulphate aerosol and A1B from 2000, no solar activity nor volcanic aerosol, dynamical ozone	GELATO sea ice model	*Daget et al.* (2009)*; Salas y Melia* (2002)
IFM-GEOMAR	ECHAM5; T63/L31	MPI-OM1; 1.5º/L40	initial condition permutations of three coupled climate simulations from 1950 to 2005 with SSTs restored to observations		observed global well-mixed GHGs, ozone and sulphate aerosol emissions and A1B from 2000, persisted solar activity and volcanic aerosol	-	*Keenlyside et al.* (2005)*; Jungclaus et al.* (2006)
CMCC-INGV	ECHAM5; T63/L19	OPA8.2; 2°/L31	AMIP-type simulations with forced SSTs	wind stress perturbations to generate ensemble of ocean reanalyses, SST perturbations at initial time	observed global well-mixed GHGs and sulphate aerosol and A1B from 2000, no volcanic aerosol nor ozone	dynamical snow-sea ice model and land-surface model	*Weisheimer et al.* (2009); *Alessandri et al.* (2010)

Error and accuracy

- Some definitions:
  - o Forecast system: the forecast model (e.g., a coupled ocean/atmosphere model), the initialization (and ensemble generation) method and the statistical model that creates the forecast probabilities (an ensemble is not a probability forecast).
  - o Forecast quality: Statistical description of how well the forecasts match the observations; it has multiple attributes.
- Important in probability forecast systems:
  - o Systematic error: difference in the statistical properties of the forecast and the reference distribution functions.
  - o Accuracy: some sort of distance between forecasts and observations. It is used as an indication of the association, the discrimination or resolution of the forecasts.
  - o Reliability: A measure of trustworthiness, not accuracy, that is another component of the systematic error of the forecast system.

INSTITUT CATALÀ DE CIÈNCIES DEL CLIMA

Dealing with systematic errors

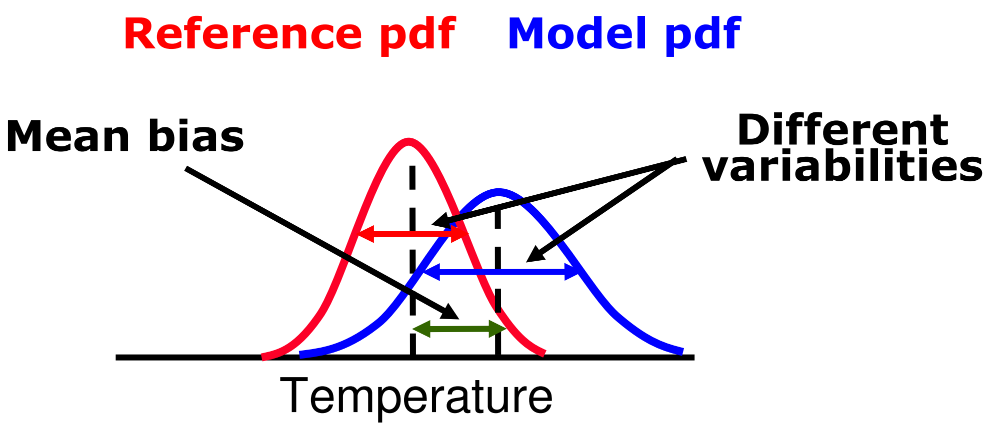
- Model drift is comparable in size to the predicted signal
  - Both in ocean and atmosphere fields.
- Predictions are made *relative* to past model integrations
  - Anomalies computed in one-year out cross-validation.
  - Model climate estimated from all available years and all ensemble members, performed separately for each single-model or model version.
  - Model climate is a function of start date and lead time.
- Implicit assumption of linearity
  - We implicitly assume that a shift in the model forecast relative to the model climate corresponds to the expected shift in a true forecast relative to the true climate, despite differences between model and true climate.
  - Most of the time, the assumption seems to work pretty well. But it’s still a strong assumption.

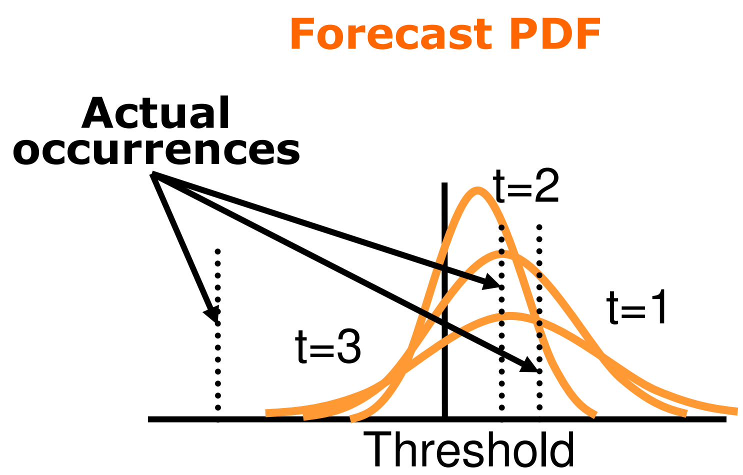
Systematic errors in ensemble forecasts

Main systematic errors in dynamical climate forecasts:

- Differences between the model climatological pdf (computed for a lead time from all start dates and ensemble members) and the reference climatological pdf (for the corresponding times of the reference dataset): systematic errors in mean and variability.
- Conditional biases in the forecast pdf: errors in conditional probabilities implying that probability forecasts are not trustworthy. This type of systematic error is best assessed using the reliability diagram.

Differences in climatological pdfs





Systematic errors in ensemble forecasts

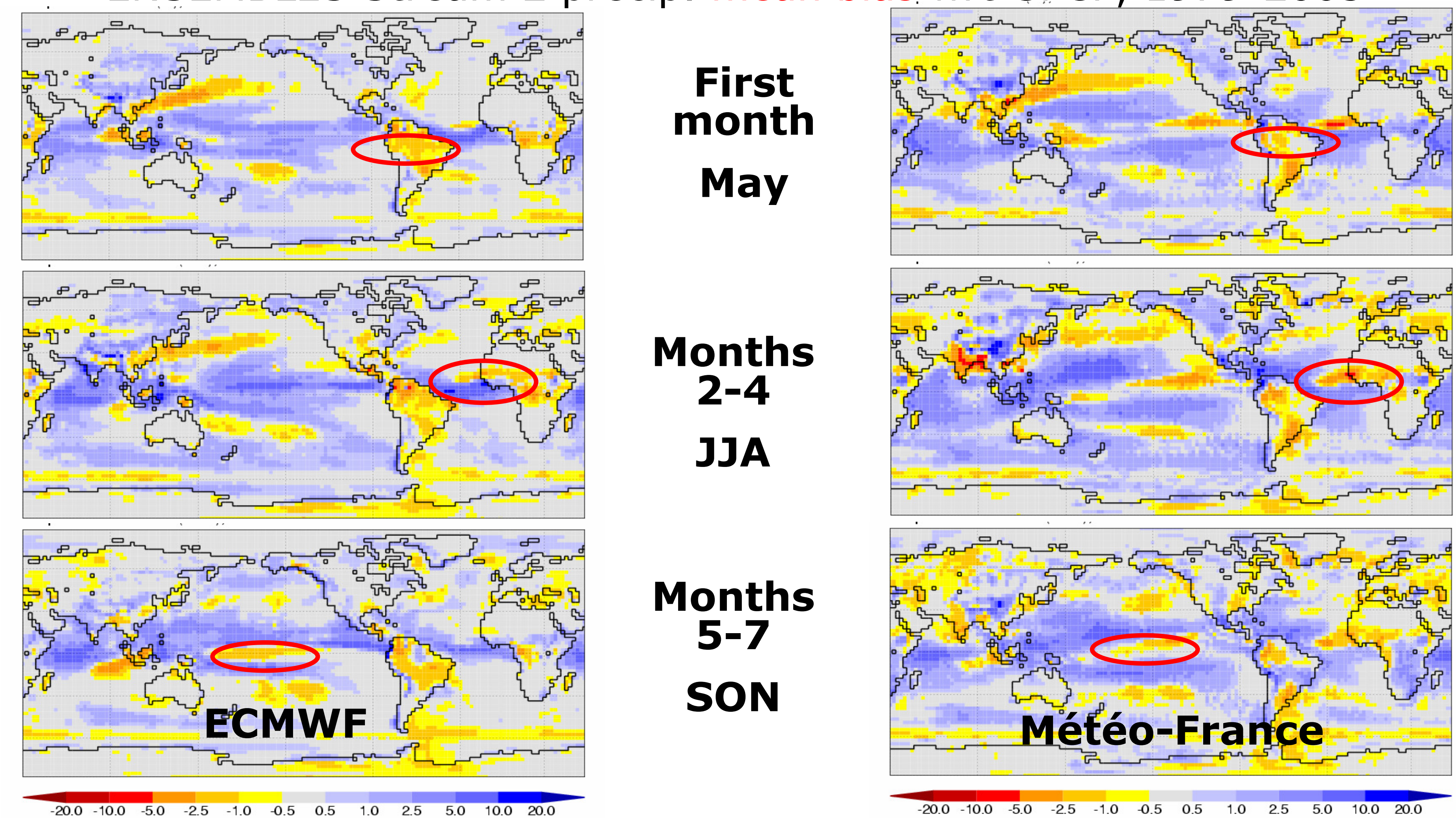
Main systematic errors in dynamical climate forecasts:

- Differences between the model climatological pdf (computed for a lead time from all start dates and ensemble members) and the reference climatological pdf (for the corresponding times of the reference dataset): systematic errors in mean and variability.
- Conditional biases in the forecast pdf: errors in conditional probabilities implying that probability forecasts are not trustworthy. This type of systematic error is best assessed using the reliability diagram.

Systematic errors in seasonal forecasts

INSTITUT CATALÀ DE CIÈNCIES DEL CLIMA

ENSEMBLES Stream 2 precip. mean bias wrt GPCP, 1979-2005



INSTITUT CATALÀ DE CIÈNCIES DEL CLIMA

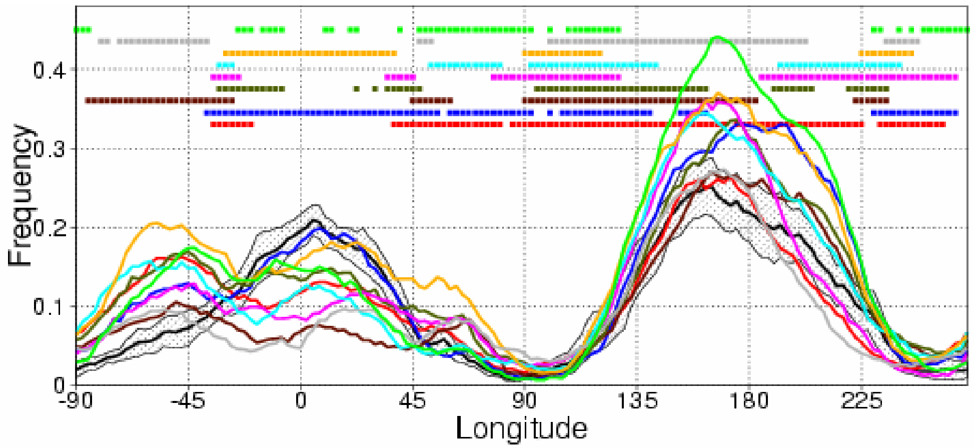
Systematic errors: Blocking frequency

Northern Hemisphere DJF blocking frequency for 1-month lead seasonal hindcasts for the 9 versions of DePreSys_PP (left) and ENSEMBLES Stream 1 (right) over the period 1991-2001. Results are for the Tibaldi and Molteni index (reversal of the meridional gradient of Z500).

The dots on top of each panel show the longitudes where the mean frequency of an experiment is significantly equal with 95% confidence to the mean frequency in ERA40

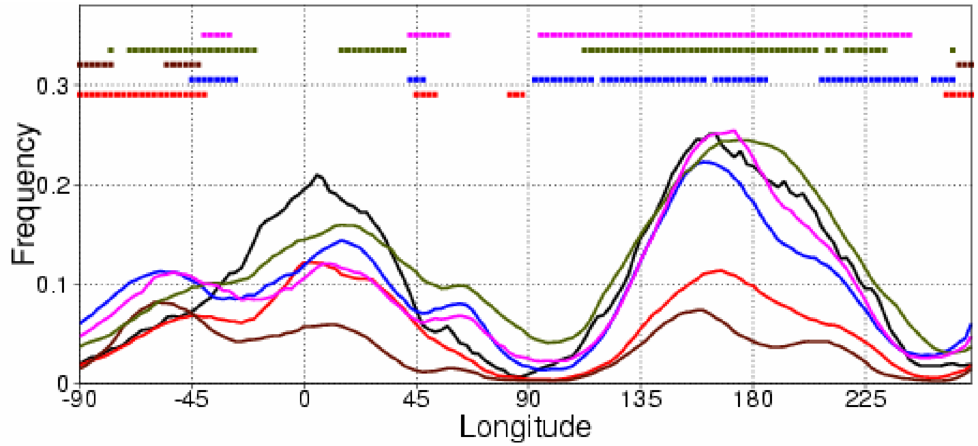
Perturbed parameter ensemble

ERA40, DePreSys and eight perturbed versions



Multi-model ensemble

ERA40 ECMWF Glosea CNRM IfM DePreSys

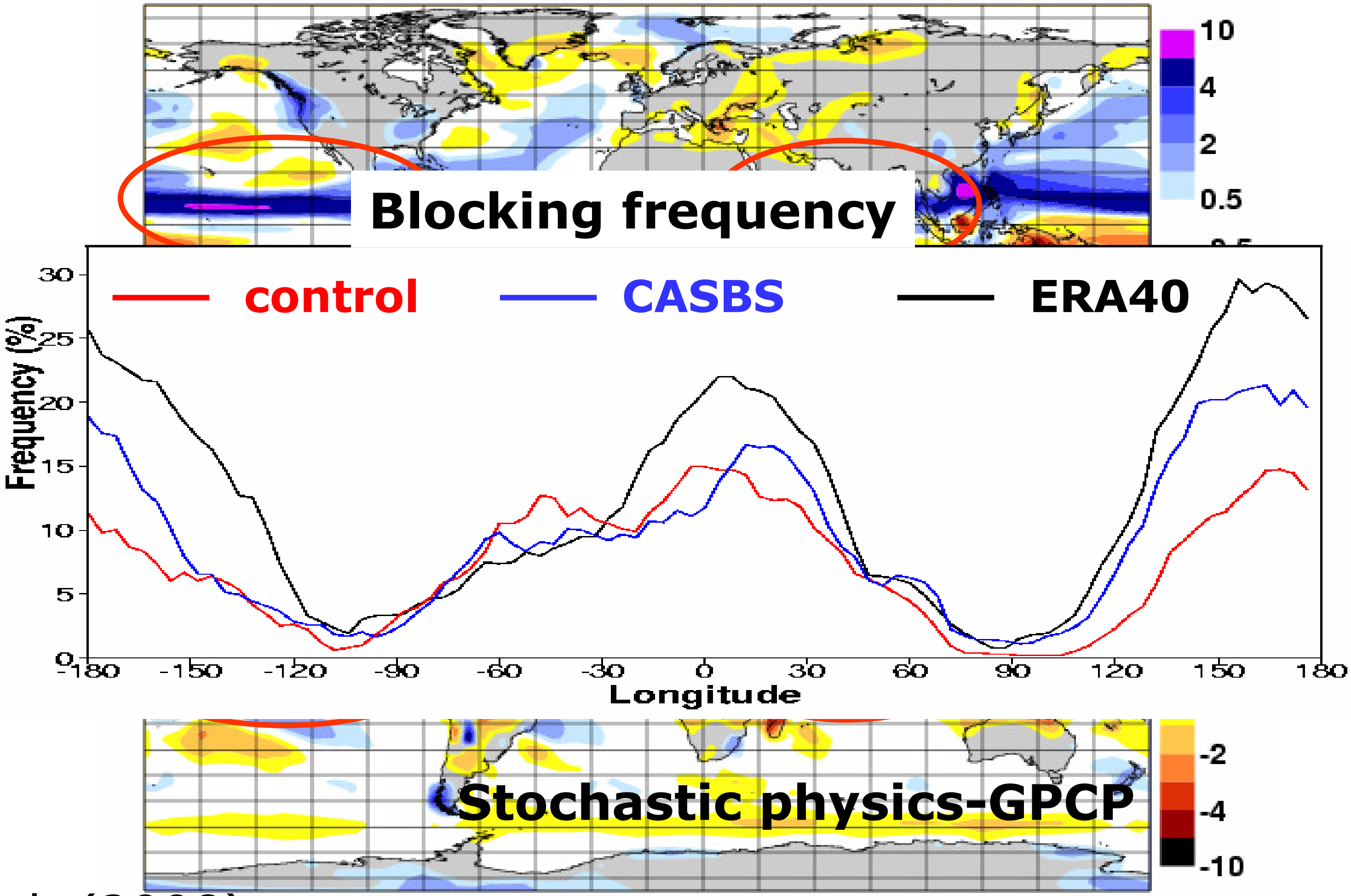


INSTITUT CATALÀ DE CIÈNCIES DEL CLIMA

Error reduction: stochastic physics

Precipitation bias (DJF, 1-month lead, 1991-2001, CY29R2)

CASBS reduces the tropical and blocking frequency biases



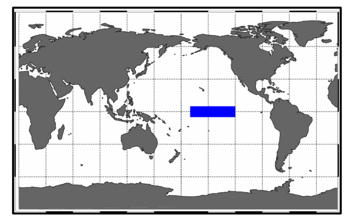
Berner et al. (2008)

INSTITUT CATALÀ DE CIÈNCIES DEL CLIMA

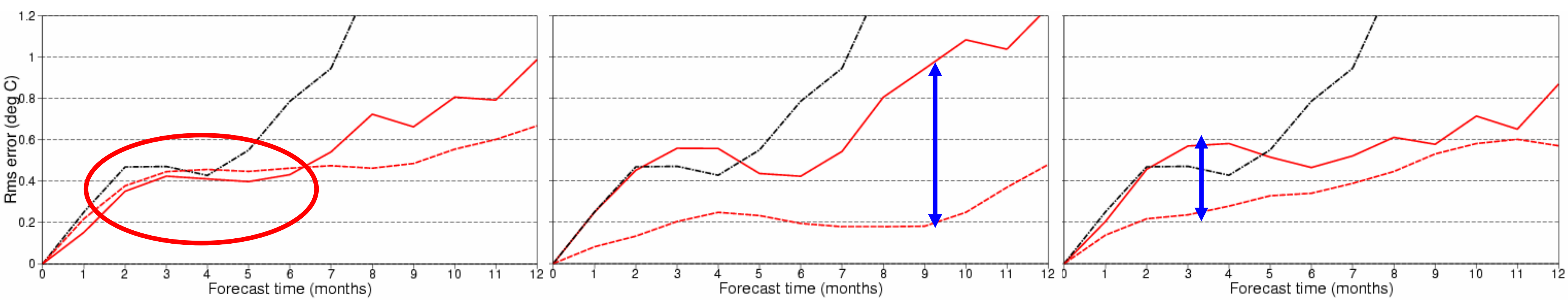
Reliability: Systematic error in spread

Sea surface temperature RMSE (solid) and spread (dashed) averaged over the Niño3.4 region for the ENSEMBLES Stream 1hindcasts of the 1st November start dates over the period 1991-2001. Persistence RMSE in dashed black.

All forecast systems beat simple persistence. The multi-model is the most skilful system, with highest deterministic reliability (RMSE~spread), in the first 6 months, while perturbed parameters is as good for longer lead times.



Multi-model (5 models, 45 members) | **ECMWF Stochastic Physics (9 members)** | **Perturbed Parameters (9 members)**

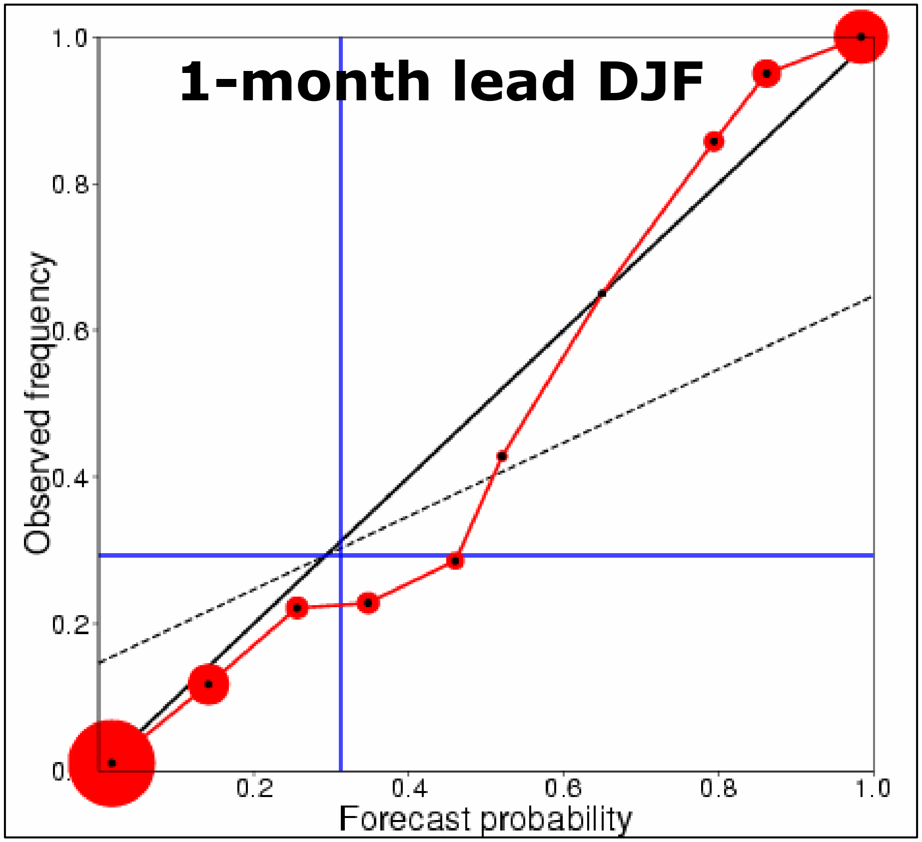


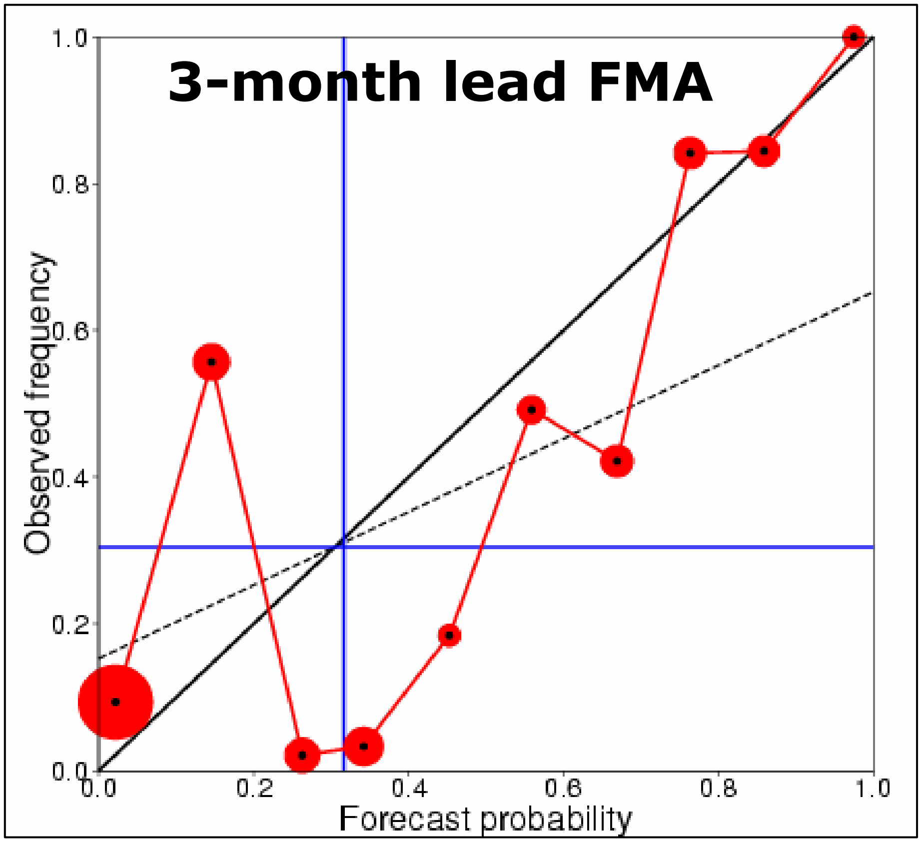
INSTITUT CATALÀ DE CIÈNCIES DEL CLIMA

RMSE-spread and probabilistic reliability

Attributes diagrams for near-surface temperature in Niño3.4 of ENSEMBLES Stream 1 multi-model over 1991-2001.

Multi-model matches RMSE and spread, but gives unreliable probabilities: matching RMSE and spread does not imply reliable probability forecasts.

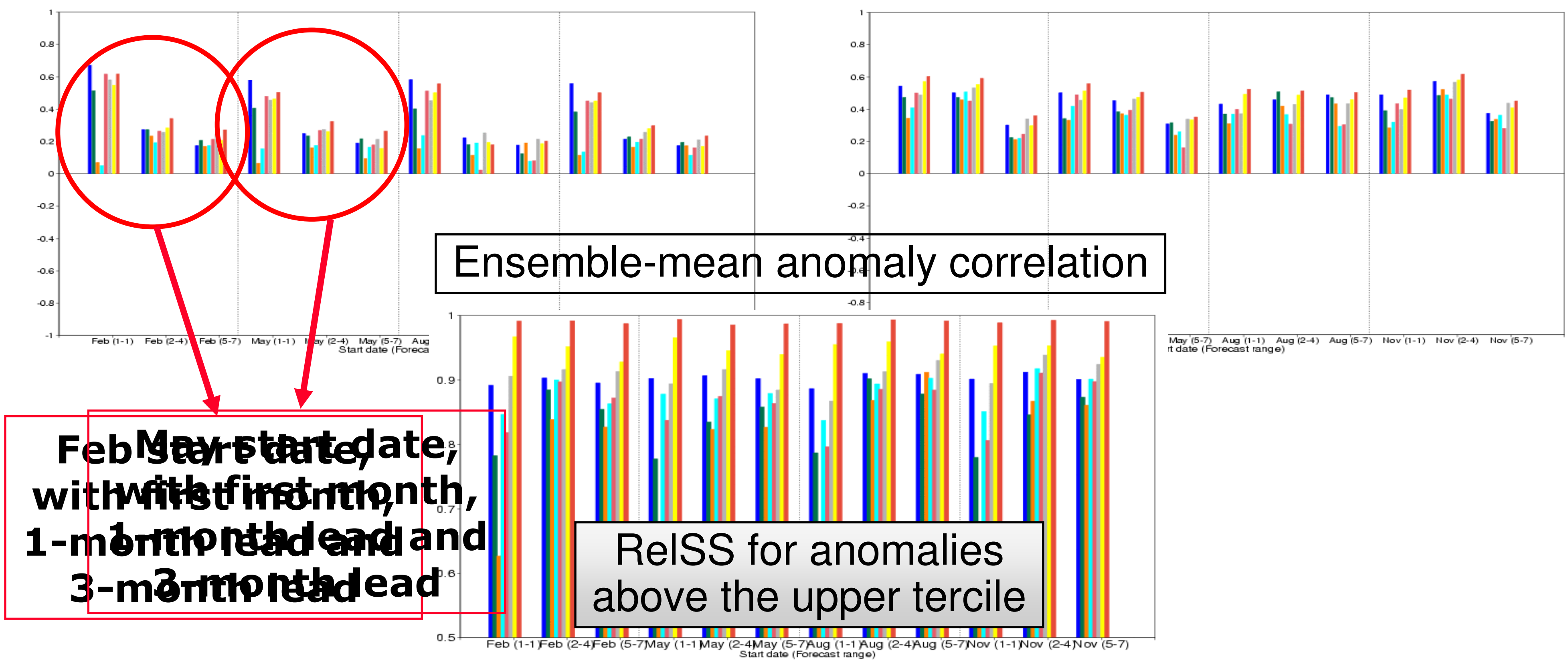




Forecast quality comparison

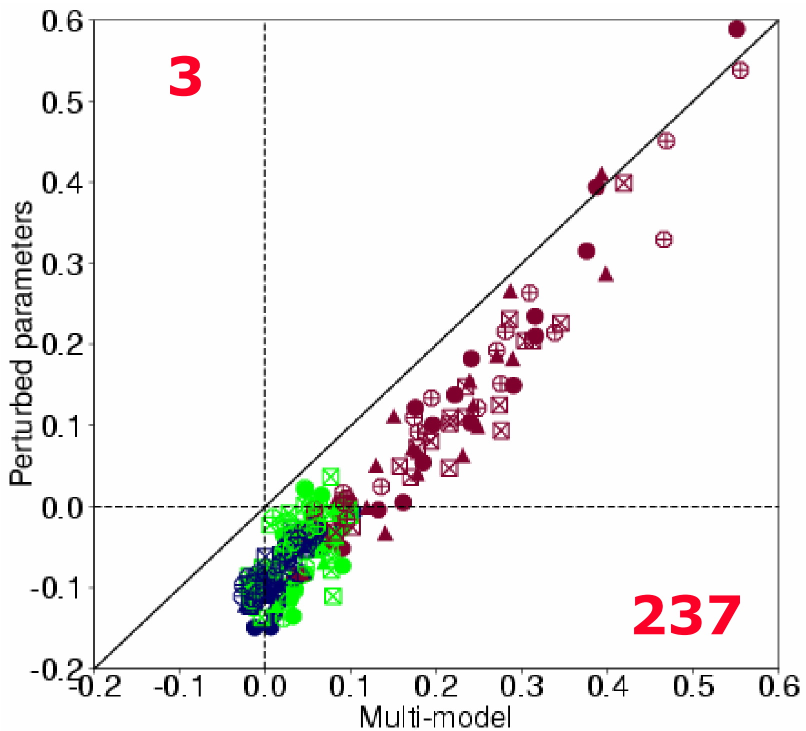
INSTITUT CATALÀ DE CIÈNCIES DEL CLIMA

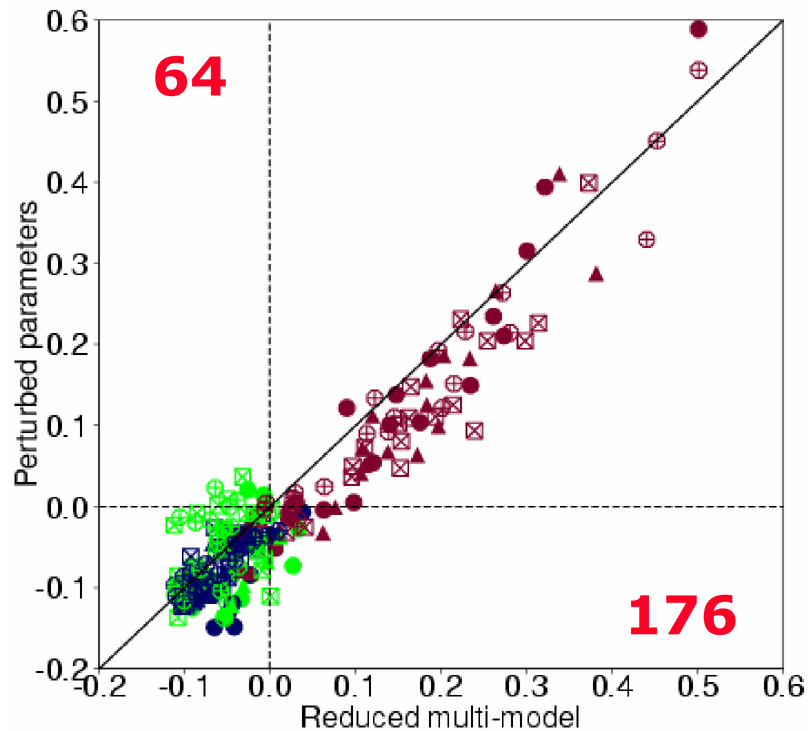
Scores for Northern extratropics T2m (left) and tropical band precipitation (right) from ECMWF, Météo-France, INGV, IfM, Met Office, Perturbed Parameters, reduced multi-model (nine-member ensembles) and multi-model (5 models, 45 members). Temperature verified with ERA40/ERA-Int over 1960-2005 and precipitation verified against GPCP over 1980-2005.



Probabilistic predictions

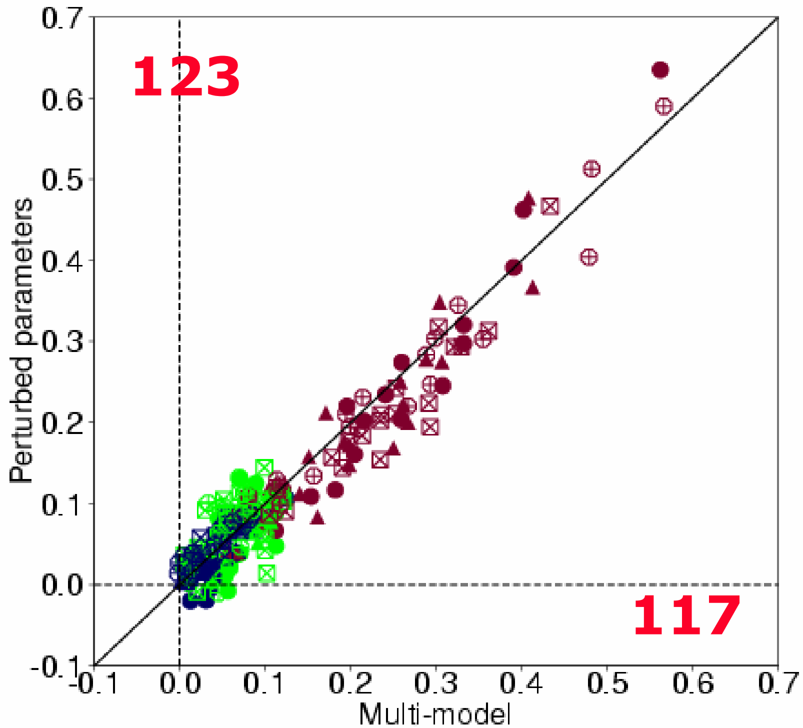
Brier skill score for several regions (Northern Hemisphere, Tropics, Southern Hemisphere), events (anomalies above/below the upper/lower tercile), lead times (2-4, 5-7 months), start dates (Feb, May, Aug and Nov) and variables (near-surface temperature, precipitation, Z500, T850 and MSLP) computed over the period 1960-2005. The inset numbers indicate the number of cases where a system is superior.

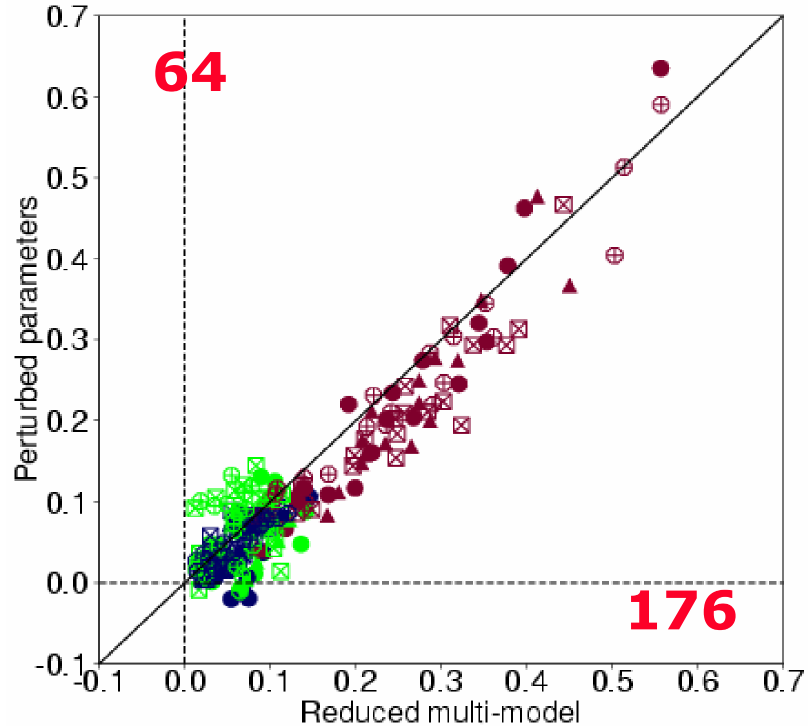




Stream 2 seasonal hindcasts

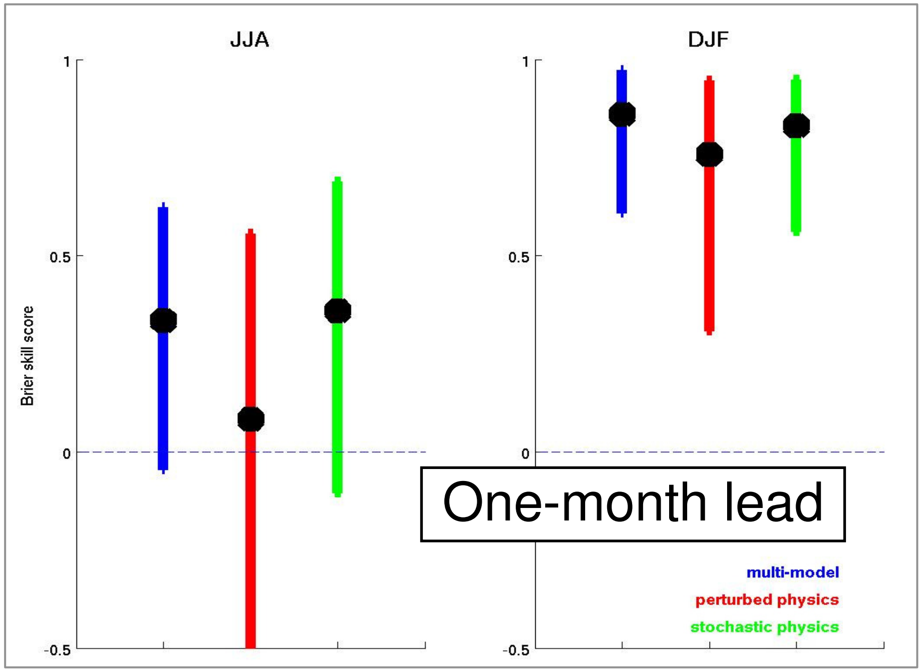
Debiased Brier skill score for several regions (Northern Hemisphere, Tropics, Southern Hemisphere), events (anomalies above/below the upper/lower tercile), lead times (2-4, 5-7 months), start dates (Feb, May, Aug and Nov) and variables (near-surface temperature, precip, Z500, T850 and MSLP) computed over the period 1960-2005. The inset numbers indicate the number of cases where a system is superior.

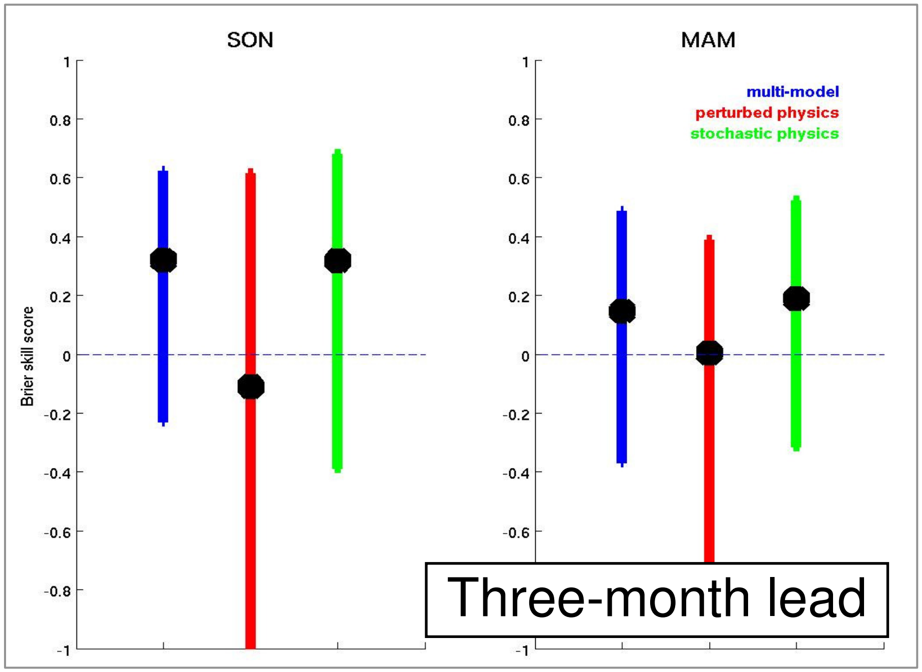




Forecast quality comparison

Debiased (i.e., taking into account the different ensemble size) Brier skill score of one-month lead (left) and three-month lead (right) predictions of Niño 3.4 SST anomalies above the upper tercile for Multi-model (45 members), Perturbed parameters (9 members) and Stochastic physics (9 members) over 1991-2005. The first (second) set of bars in each panel are for the May (November) start dates. Bars represent the 95% confidence interval of the skill score.





Forecast quality comparison

Debiased Brier skill score of one-month lead predictions of land temperature over the Giorgi regions for Multi-model (45 members, left columns), Perturbed parameters (9 members, central columns) and Stochastic physics (9 members, right members) over 1991-2005. Significantly positive or negative scores are in bold.

	NEAR-SURFACE TEMPERATURE											
	JJA		DJF		JJA		DJF		JJA		DJF	
	cold	warm	cold	warm	cold	warm	cold	warm	cold	warm	cold	warm
Australia	**11.5**	**13.9**	3.2	6.7	-0.3	**11.0**	0.5	5.2	7.0	**17.3**	**11.8**	8.0
Amazon Basin	0.2	17.1	4.5	**23.4**	-13.7	2.8	-6.3	11.2	3.9	14.7	2.6	16.9
Southern South America	**9.2**	**9.0**	1.8	**9.9**	-2.8	7.2	2.9	**14.7**	**16.9**	8.8	4.5	**9.3**
Central America	5.9	**11.6**	-2.6	4.5	2.4	5.5	-3.9	3.3	1.2	-0.3	0.2	-3.7
Western North America	10.2	**12.2**	6.3	**12.5**	6.7	-1.2	3.3	8.9	2.8	8.0	6.4	4.7
Central North America	-0.2	**-7.3**	-3.3	10.4	-8.5	**-12.7**	7.2	13.8	**-21.4**	**-20.3**	-2.6	8.8
Eastern North America	4.1	-7.0	-4.5	10.1	-9.9	-14.7	**32.2**	8.2	-13.4	**-10.9**	-11.3	4.0
Alaska	-0.8	-0.9	-0.6	0.6	-0.4	-2.9	6.5	4.8	0.5	**12.2**	**-20.3**	-1.0
Greenland	**15.1**	**8.7**	**13.2**	**12.3**	12.7	-1.5	**17.3**	**15.1**	3.2	3.1	**12.3**	**16.3**
Mediterranean	**18.0**	**12.8**	5.8	4.3	**18.3**	**15.5**	**-17.5**	**-14.5**	**22.7**	**12.2**	6.2	2.6
Northern Europe	-3.3	0.2	4.9	0.5	1.1	4.6	-0.6	-4.0	4.6	6.3	1.5	5.2
Western Africa	7.9	7.0	7.3	**20.5**	-14.8	3.6	3.6	10.6	7.8	-2.0	10.9	**15.8**
Eastern Africa	**9.4**	**7.3**	-7.7	0.9	**-19.5**	-7.1	-3.9	-5.4	**-9.7**	-3.1	-3.7	8.2
Southern Africa	**14.0**	4.7	1.7	**10.6**	-3.2	10.2	-1.7	2.7	0.0	7.7	6.0	**13.6**
Sahel	**12.9**	7.2	**11.5**	**15.4**	**9.9**	**13.1**	6.6	**15.7**	**16.3**	**10.1**	**13.9**	**14.7**
South East Asia	**8.6**	**12.4**	**11.6**	13.4	-9.3	4.2	13.9	6.1	-0.6	9.6	3.8	1.6
East Asia	10.6	**10.2**	0.3	5.8	8.3	**10.5**	-4.2	10.1	6.4	**14.1**	3.1	-0.4
South Asia	**8.7**	**13.3**	**14.4**	**10.6**	4.3	9.2	0.1	9.3	12.9	**15.7**	**13.8**	**18.1**
Central Asia	**14.3**	**8.2**	-2.4	7.1	**14.1**	**11.8**	-2.0	**19.1**	**21.1**	**10.1**	-8.5	6.5
Tibet	**16.9**	**16.1**	-0.1	4.1	7.8	7.2	**-10.4**	3.8	8.3	**15.7**	5.6	7.6
North Asia	**7.3**	3.9	4.2	**8.5**	6.2	**8.4**	-1.5	**12.6**	4.2	1.6	-1.9	1.2
	multi-model				perturbed parameters				stochastic physics			

Forecast quality comparison

Debiased Brier skill score of one-month lead predictions of land temperature over the Giorgi regions for Multi-model (45 members, left columns), Perturbed parameters (9 members, central columns) and Stochastic physics (9 members, right members) over 1991-2005. Significantly positive or negative scores are in bold.

PRECIPITATION

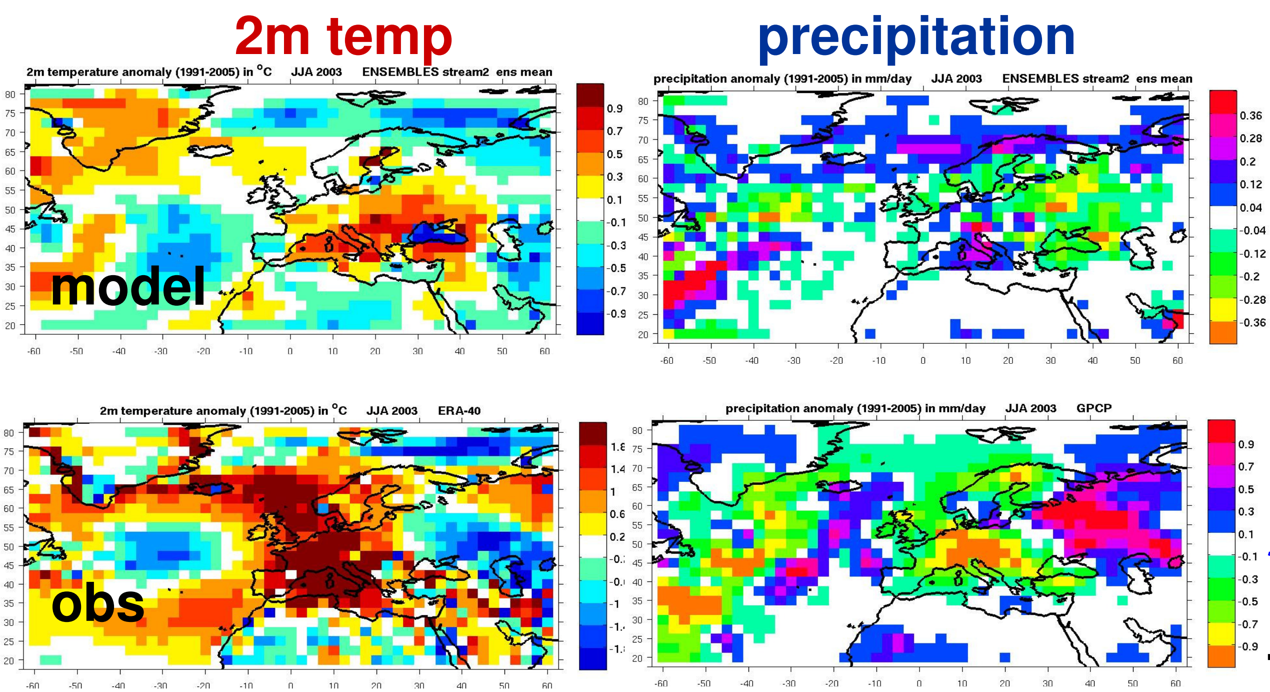
	multi-model				perturbed parameters				stochastic physics			
	JJA		DJF		JJA		DJF		JJA		DJF	
	dry	wet	dry	wet	dry	wet	dry	wet	dry	wet	dry	wet
Australia	7.6	7.0	0.9	3.0	5.1	8.0	**12.4**	5.2	2.5	5.0	**10.5**	6.6
Amazon Basin	**10.3**	**10.3**	**16.0**	14.3	**8.8**	5.4	3.4	0.5	**12.2**	**11.4**	**16.1**	16.8
Southern South America	6.2	7.1	4.6	6.0	1.3	1.6	-4.5	-1.7	3.3	**9.0**	-4.7	0.2
Central America	**9.2**	**7.8**	**23.4**	**18.9**	**12.9**	5.2	**23.3**	**25.9**	**10.6**	7.7	**24.9**	**23.7**
Western North America	2.4	**8.1**	**7.2**	**7.8**	4.5	**7.5**	4.5	4.9	**9.1**	**8.4**	5.7	5.3
Central North America	0.6	2.2	7.7	**10.4**	-3.5	-5.7	**10.0**	**10.4**	1.7	3.0	2.1	5.5
Eastern North America	-1.9	-1.1	**8.3**	**10.6**	**-9.6**	**-11.1**	9.7	**13.2**	**-15.0**	-6.8	7.5	2.1
Alaska	-1.3	0.0	4.0	-2.2	-2.3	-1.0	**11.3**	3.7	-4.3	-0.7	0.2	-2.5
Greenland	2.6	**2.8**	**-3.7**	**-3.0**	1.4	0.2	**7.5**	-1.7	**-6.8**	-2.6	-2.2	-2.1
Mediterranean	-1.2	1.2	-1.0	-1.3	-6.1	-4.4	-3.0	0.1	-0.9	0.1	**11.5**	10.7
Northern Europe	2.3	2.1	-3.1	**-4.7**	**7.7**	**11.5**	-1.8	-1.6	**8.2**	6.0	6.6	1.6
Western Africa	-1.5	-0.1	-0.5	1.3	**-10.9**	-3.8	4.8	-1.6	-4.8	2.4	**-13.7**	-0.1
Eastern Africa	-2.8	1.8	3.9	2.5	-7.0	**-7.6**	**14.4**	**13.2**	-1.5	3.4	0.9	5.7
Southern Africa	3.5	1.0	5.7	**9.5**	**7.2**	4.7	**6.0**	**11.3**	7.8	9.2	7.7	8.9
Sahel	**-4.6**	**-3.6**	**-3.2**	-1.5	**-9.2**	**-6.7**	-2.7	-2.4	**-10.0**	-1.0	**-8.2**	-3.6
South East Asia	**14.3**	9.7	**8.8**	**8.3**	5.5	4.8	5.6	8.3	10.3	1.1	9.6	**12.5**
East Asia	0.5	-0.5	4.7	**4.6**	**5.6**	1.4	8.9	3.6	2.8	0.6	8.9	**15.7**
South Asia	0.2	0.9	**6.5**	**7.4**	0.6	-2.7	7.0	9.4	2.7	1.9	5.5	**10.2**
Central Asia	-0.8	0.2	7.4	5.7	0.8	-3.1	**10.3**	**8.4**	-1.5	0.2	2.9	1.6
Tibet	5.5	3.5	**6.5**	5.4	-1.4	-0.9	1.2	7.8	4.2	**6.4**	**10.7**	**10.0**
North Asia	**2.4**	**2.6**	3.1	0.6	3.3	2.9	2.1	-1.0	1.0	0.6	2.5	-1.9

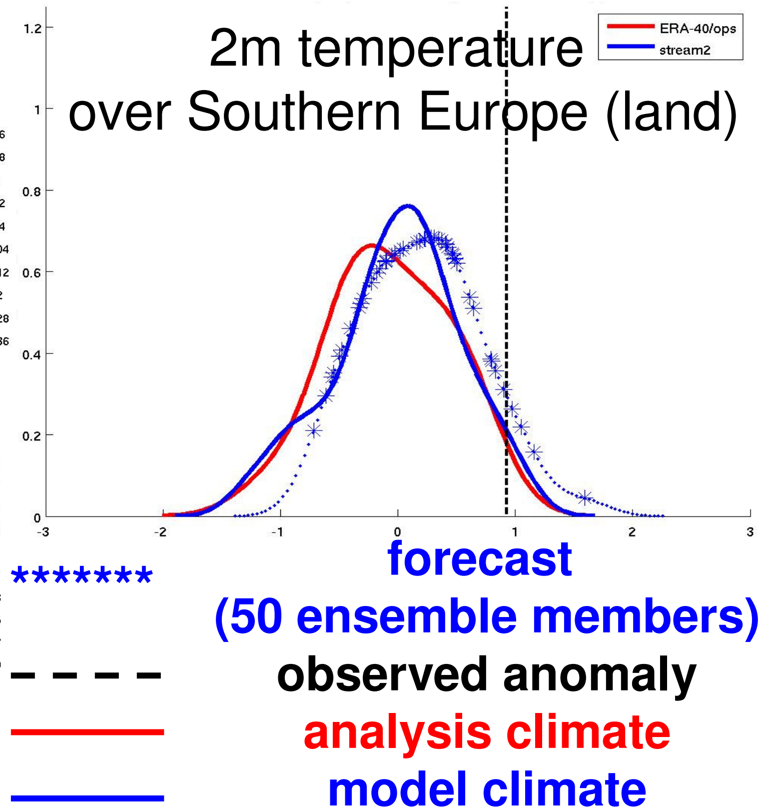
Reducing further the uncertainty

- Improving the initial conditions and the single-model ensemble generation (ocean, soil moisture, snow and sea-ice perturbations).
- All these methods shouldn’t prevent the most required improvement of the models.
- Forecast postprocessing.

Improving the model

- ECMWF ENSEMBLES seasonal prediction for summer 2003 with start date on the 1st of May. Anomalies wrt period 1991-2005.



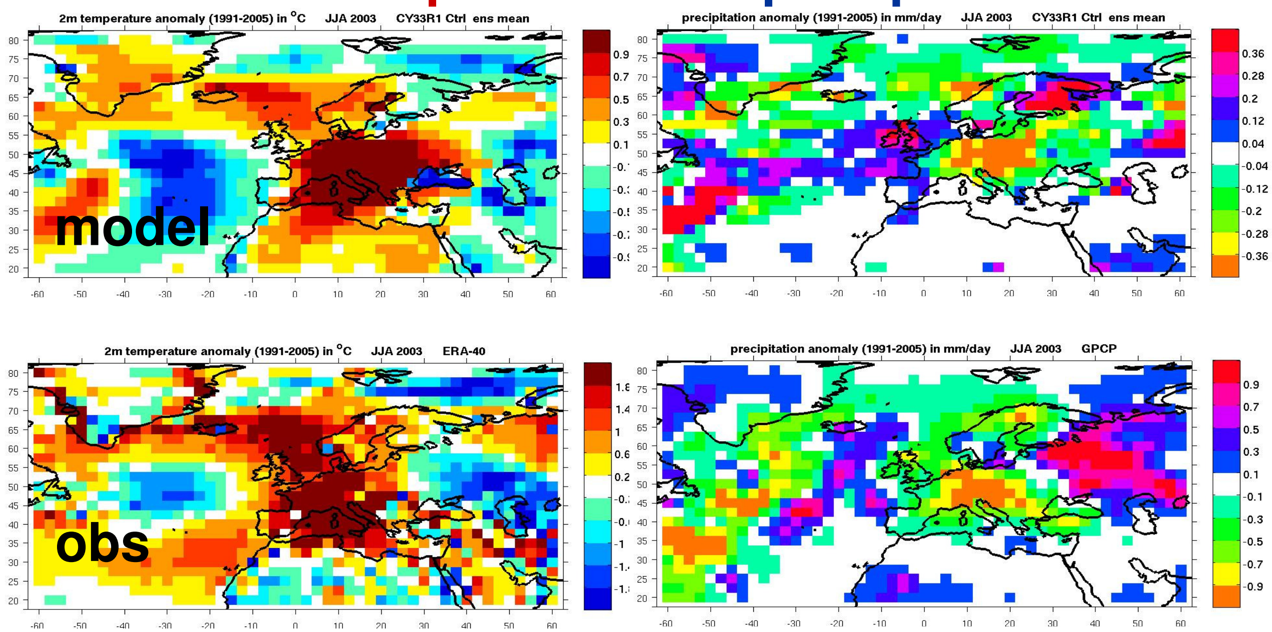


Antje Weisheimer

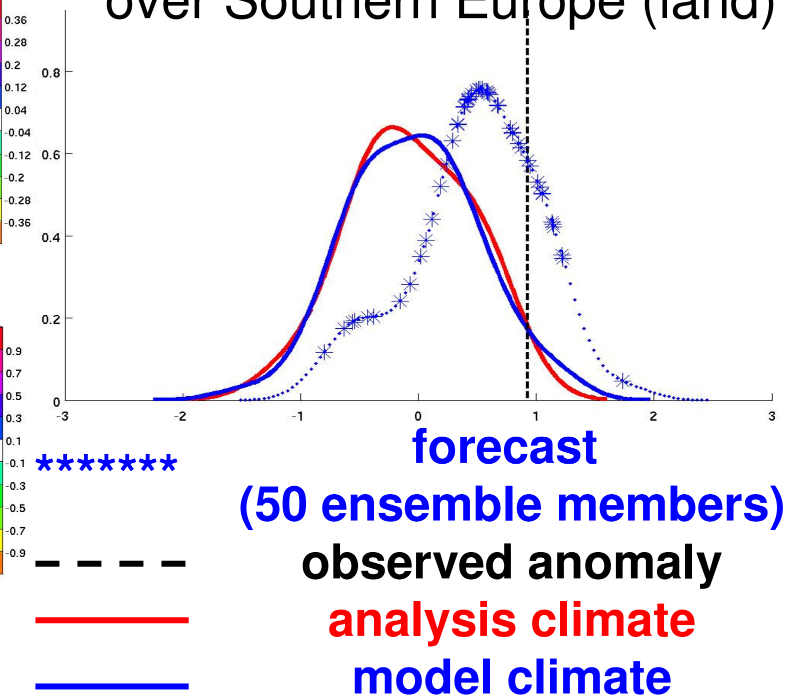
Improving the model

- Seasonal prediction with improved ECMWF system (changes in radiation, soil scheme and convection) for summer 2003 with start date on the 1st of May. Anomalies wrt period 1991-2005.

2m temp **precipitation**



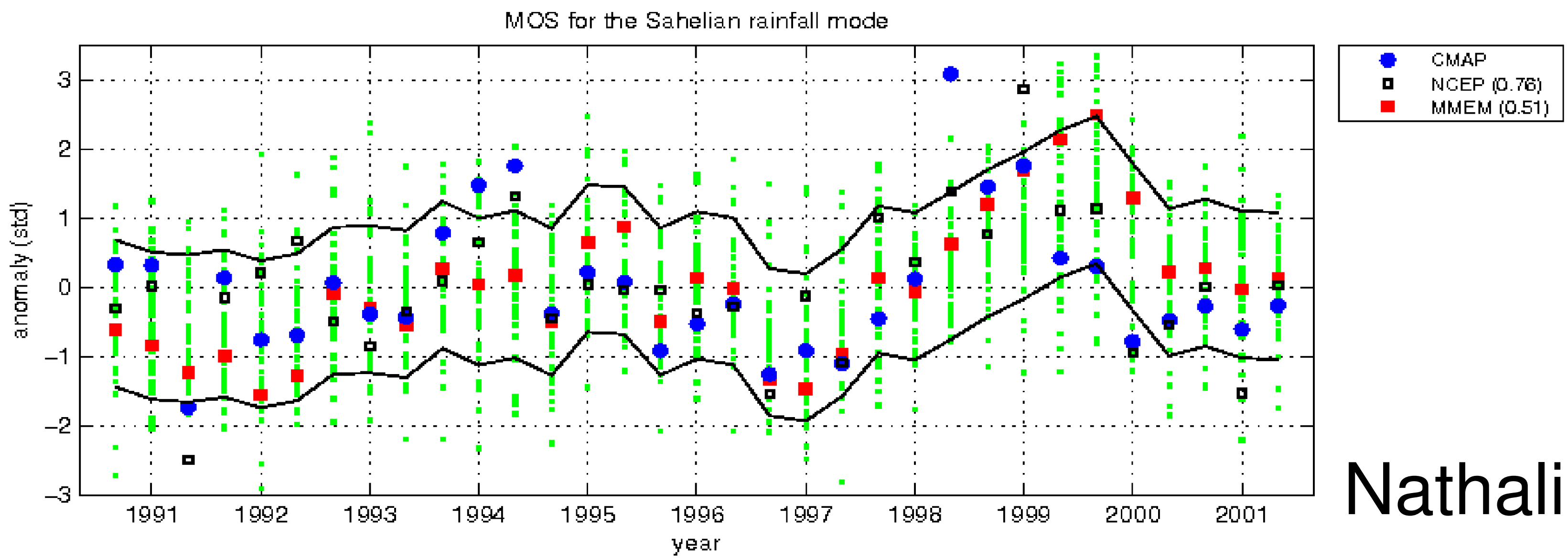
2m temperature
over Southern Europe (land)



Antje Weisheimer

Post-processing: WAM

- Seasonal predictions of averaged Sahelian summer precipitation with the ENSEMBLES multi-model.
- Predictors: five leading modes of the simulated wind/T.
- Method: MLR with stepwise selection of significant predictors (10%) and leave-3-out cross-validation.
- Higher skill (R=0.51) than with direct output (R=0.15).



Nathalie Philippon

Summary

- Substantial systematic error, including lack of reliability, is still a fundamental problem in dynamical forecasting and forces *a posteriori* corrections to obtain useful predictions.
- Comprehensive assessments of the forecast quality measures (including estimates of their standard error) are indispensable in forecast system comparisons.
- Stochastic physics schemes can reduce systematic error without affecting forecast quality (unlike post-processing).
- Perturbed-parameter ensembles are competitive with multi-model ensembles, with gains both in accuracy and reliability with respect to their reference system.
- Comparison of the three methods does not allow for the possibility to unequivocally single out the best one.