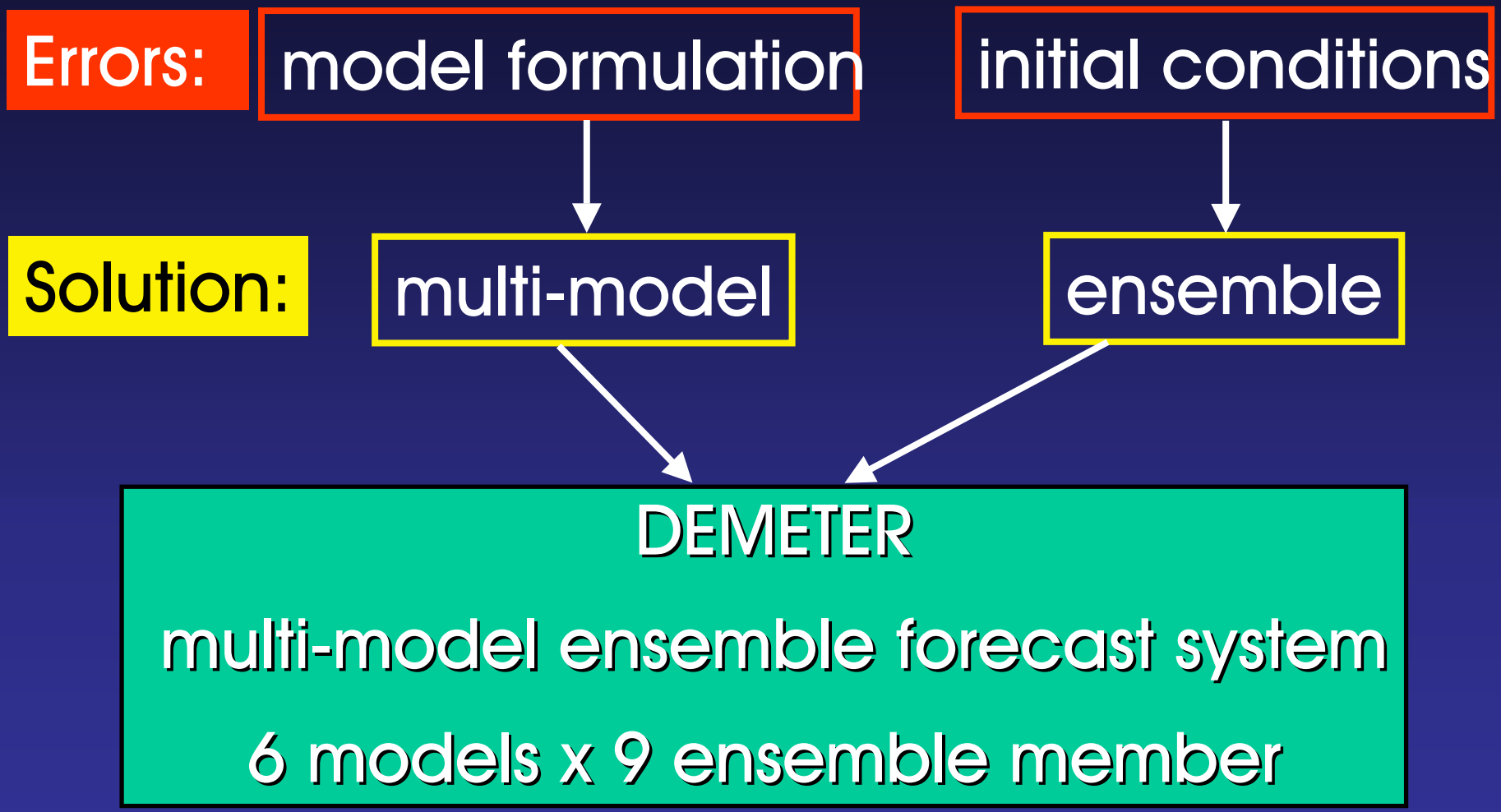
Multi-model hindcasts: latest results

Francisco J. Doblas-Reyes
Renate Hagedorn
Tim Palmer

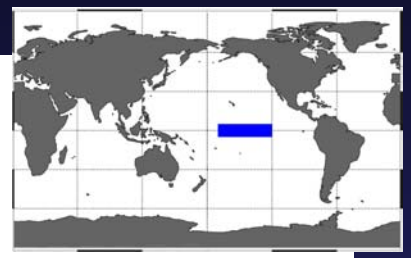
Outline

- Multi-model general results
  - Multi-model vs single models skill
  - Impact of the ensemble size
- Model combination

Multi-Model Ensemble Approach

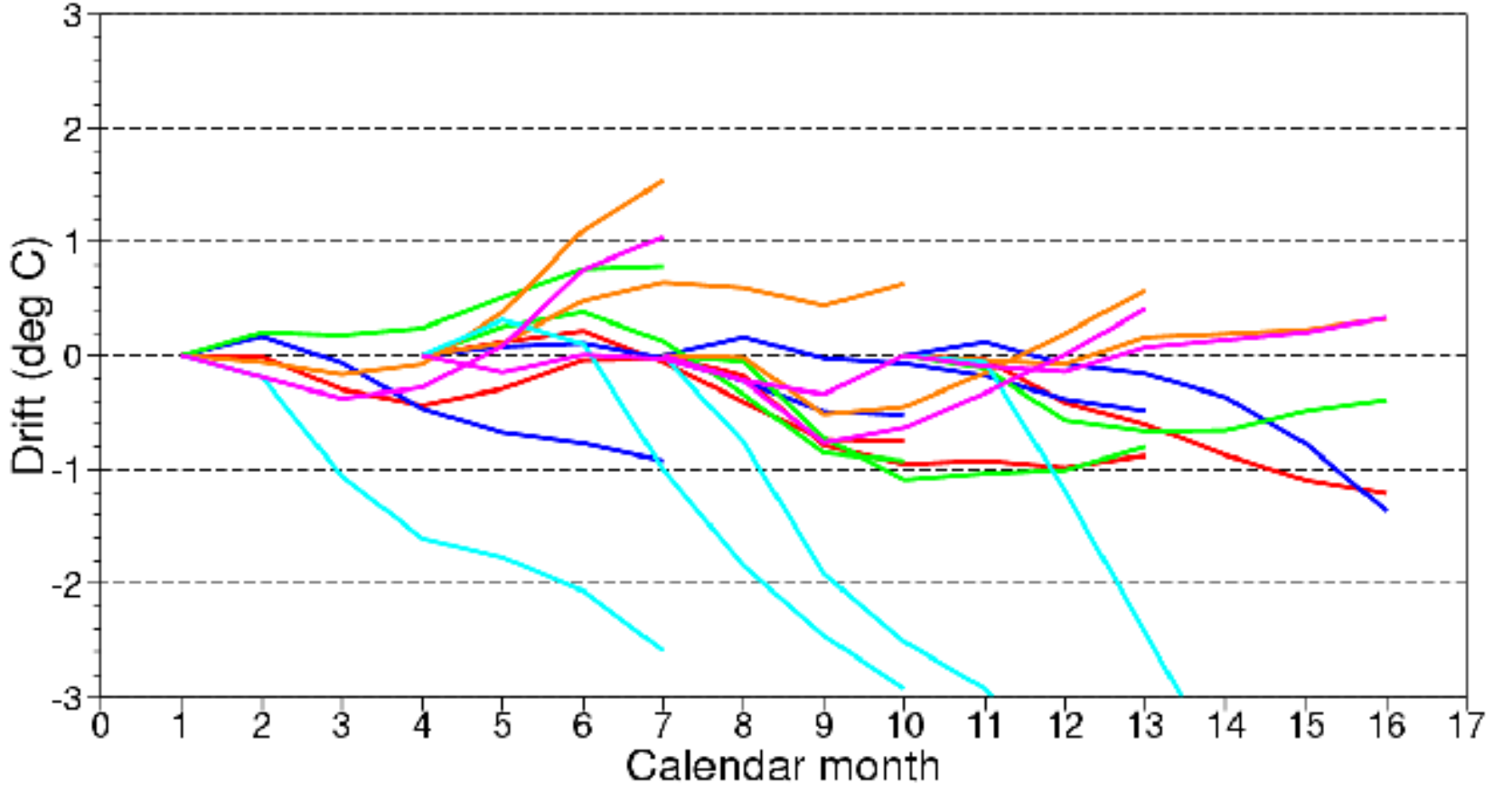


Niño-3.4 SST Hindcasts: Drift

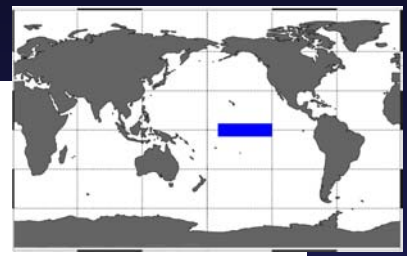


Drift: 1987 – 1998 (48 start dates)

DEMETER ECMWF CNRM UKMO MPI CERFACS

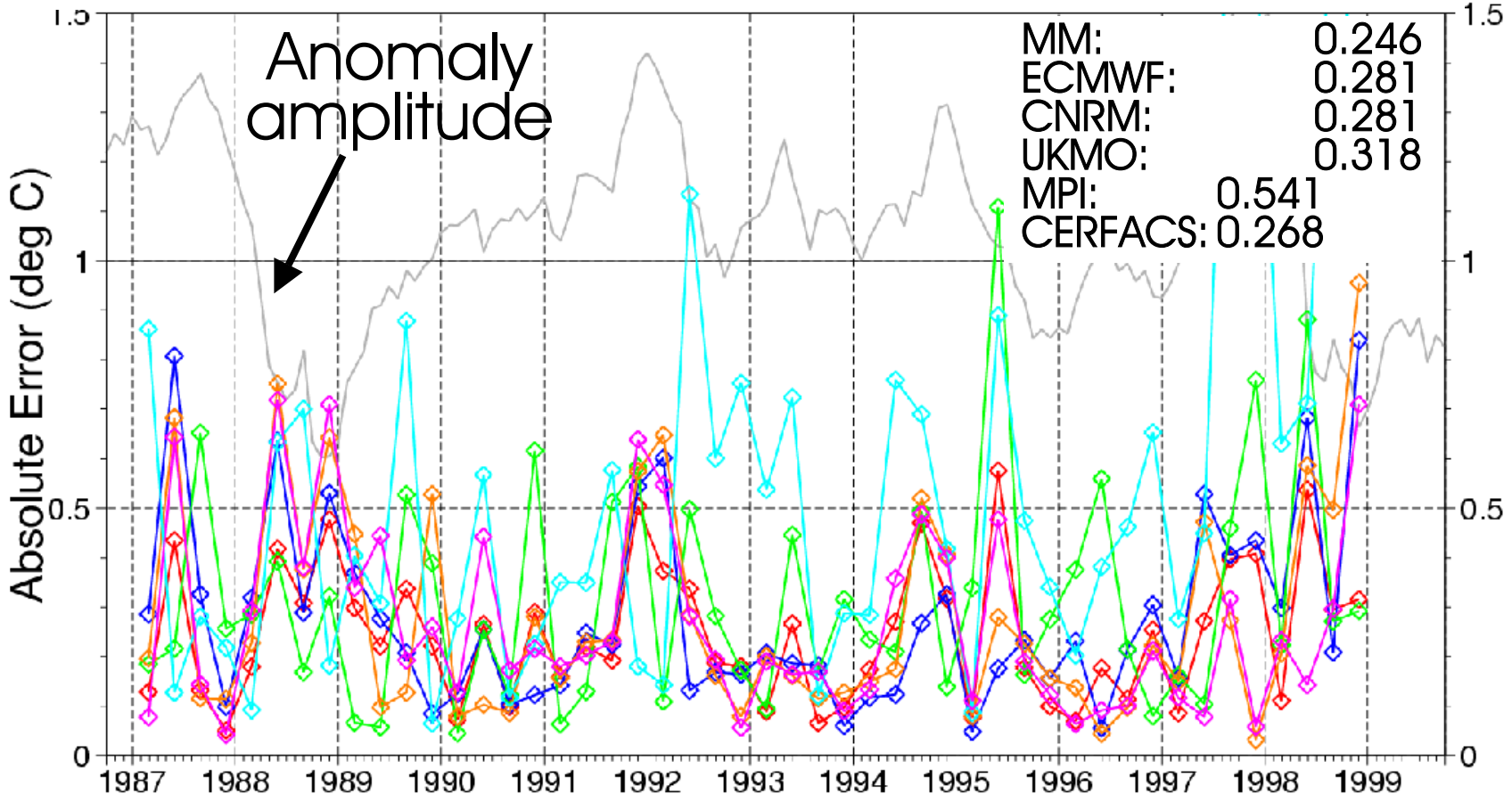


Niño-3.4 SST Hindcasts: MAE

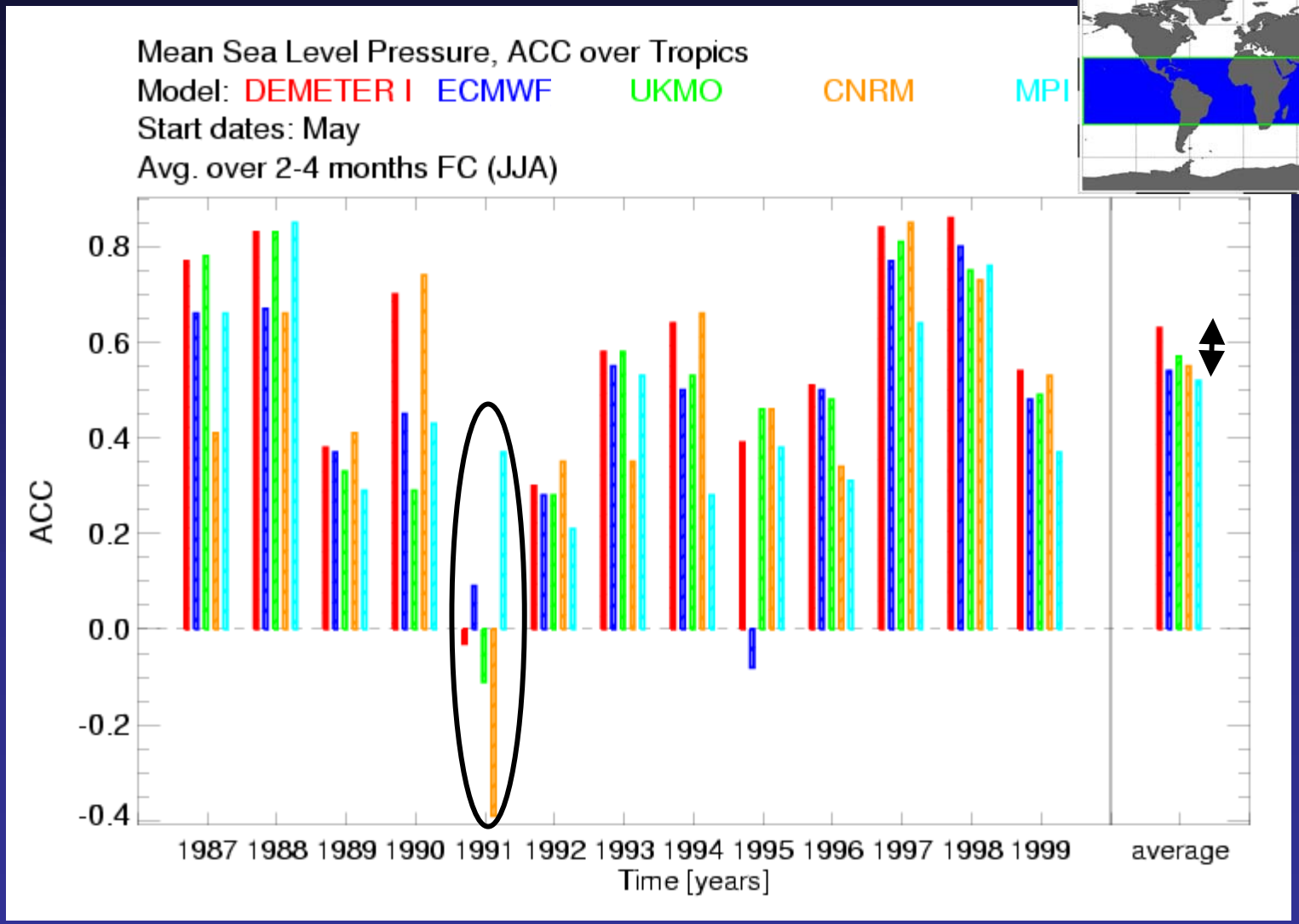


Mean Absolute Error: 1987 – 1998

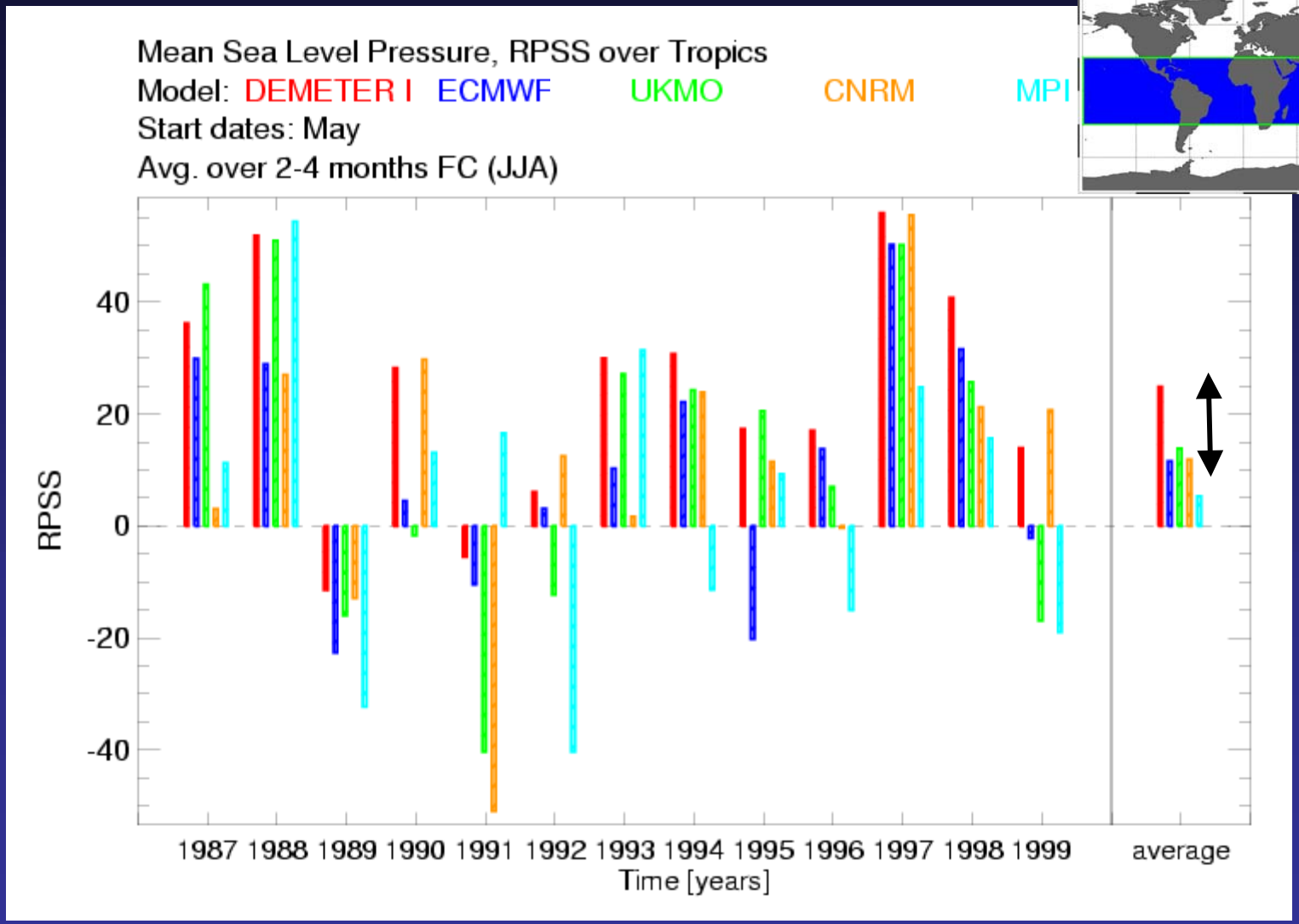
DEMETER ECMWF CNRM UKMO MPI CERFACS



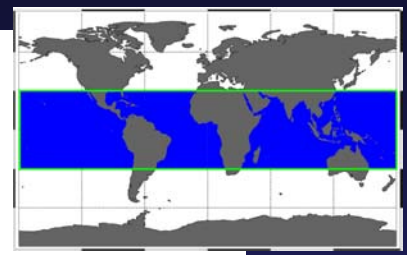
ACC: MSLP Tropics



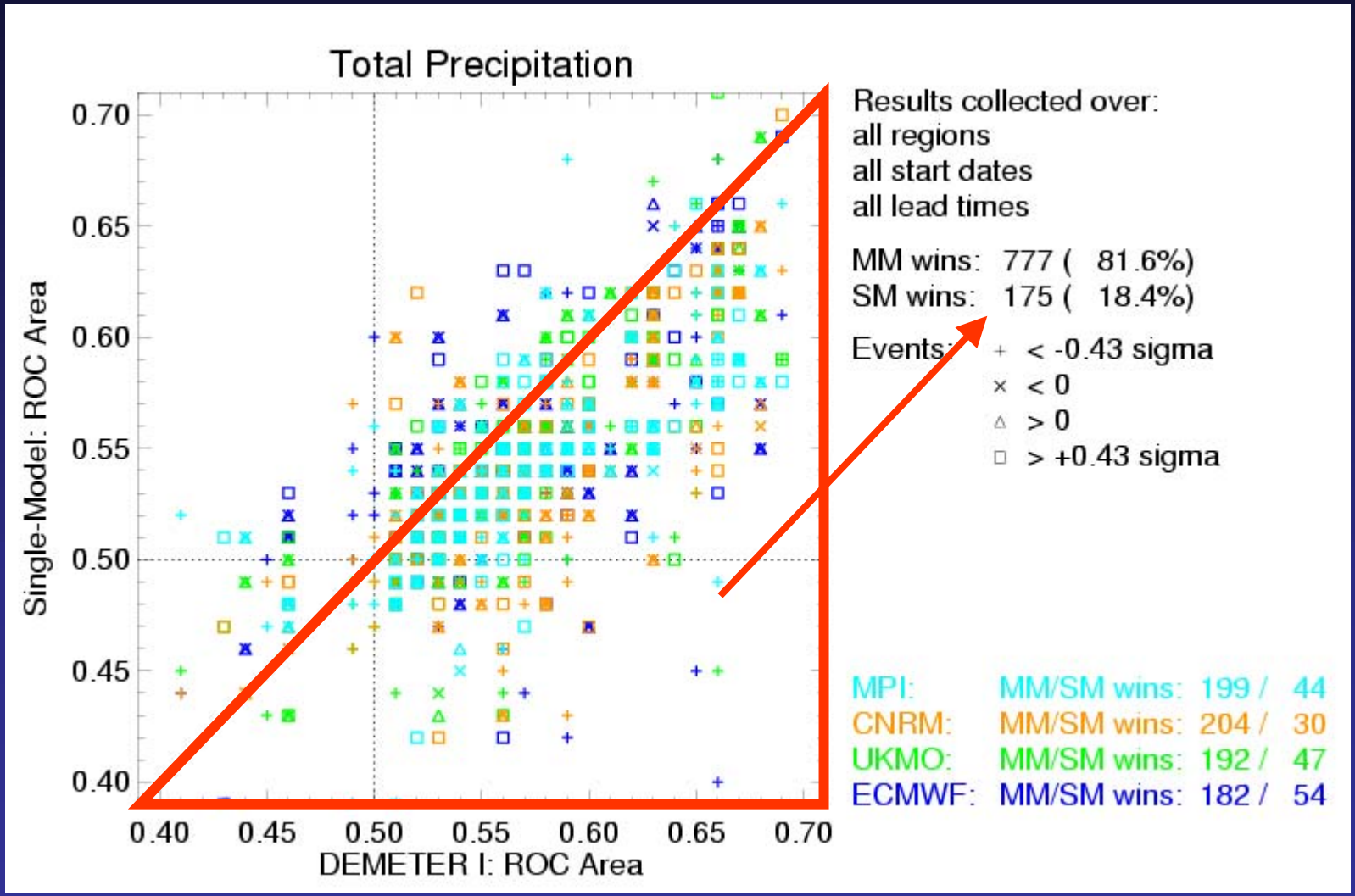
RPSS: MSLP Tropics



PDF of Grid-Point RPSS: MSLP Tropics

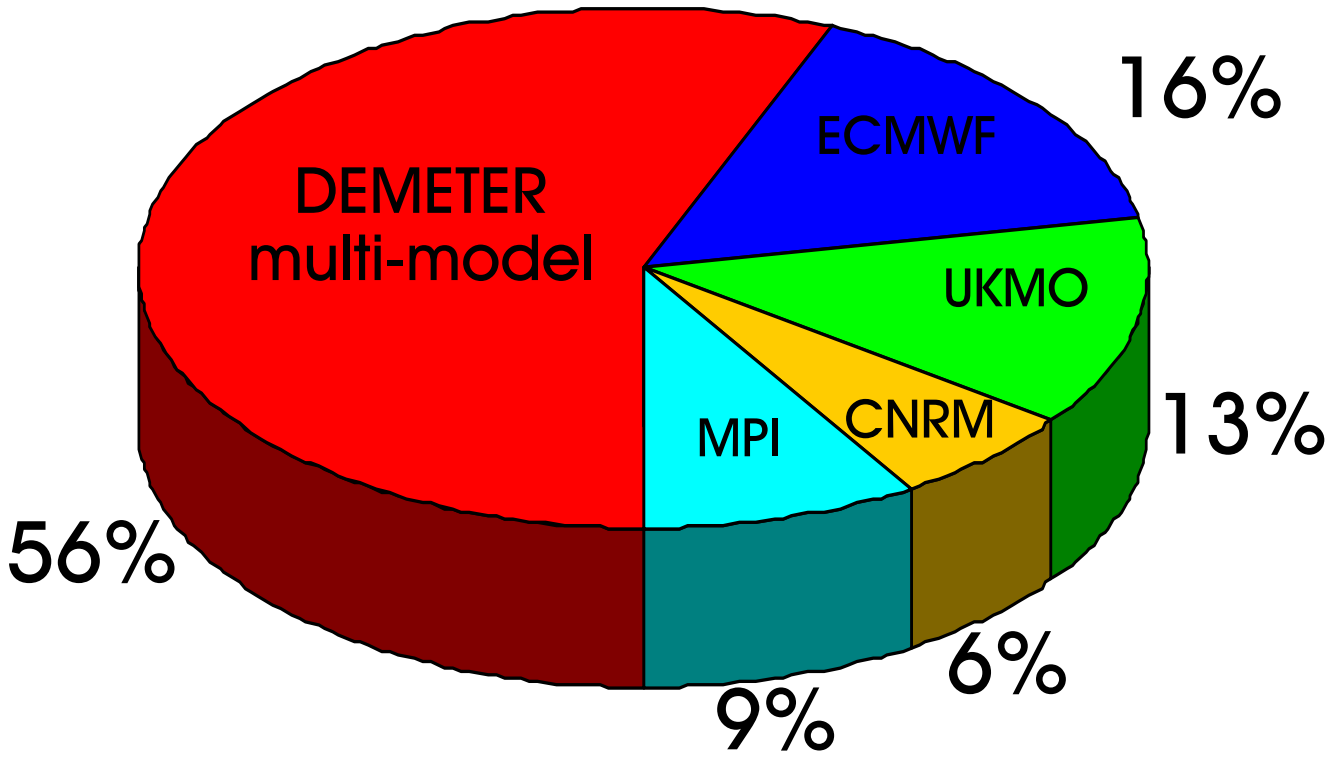


Single vs. Multi-model: ROC



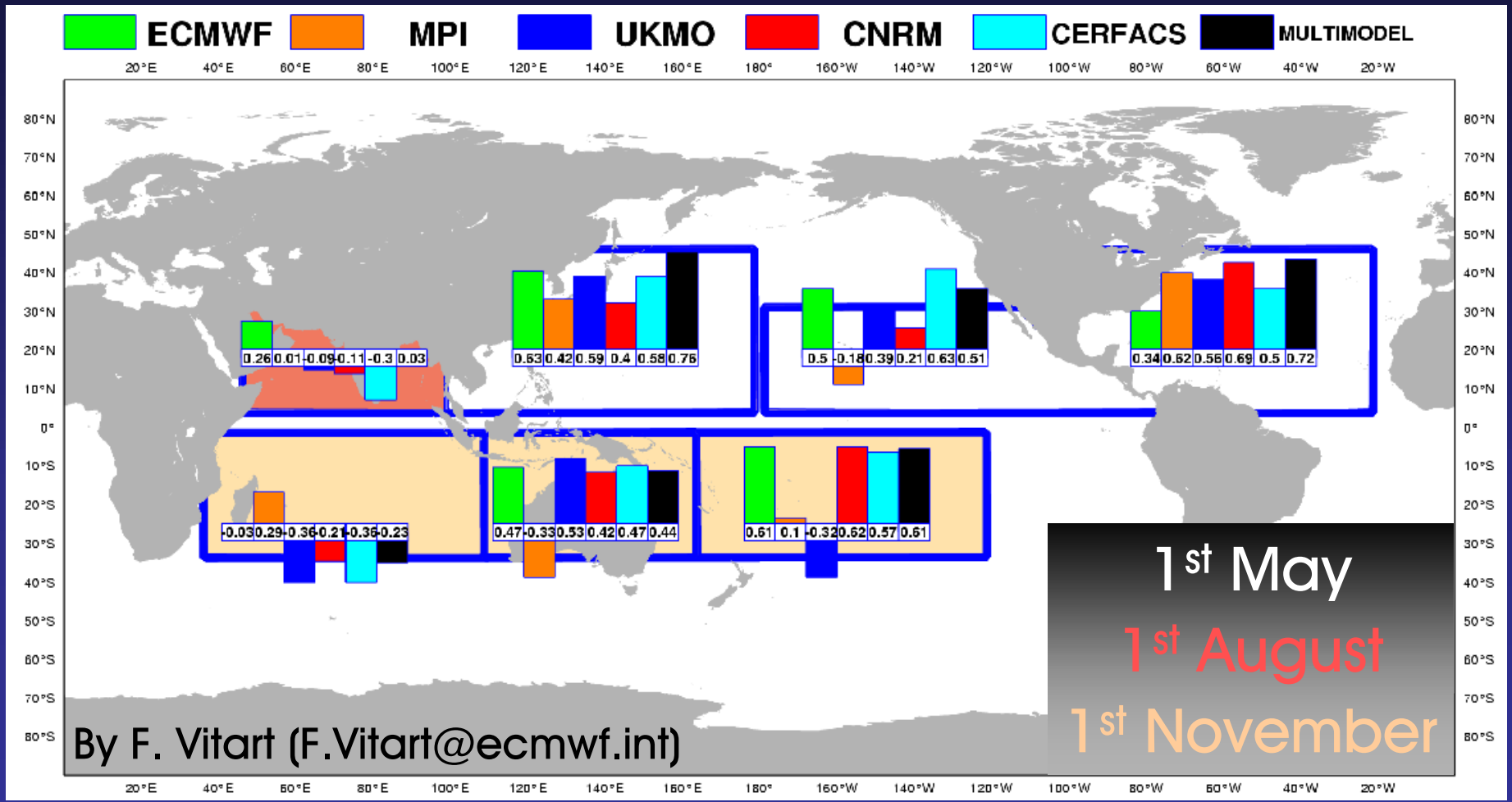
Best-Model ROC Score

Test of individual and multi-model ROC skill (area-averaged) over 8 regions, 4 seasons, 4 events, 2 lead times



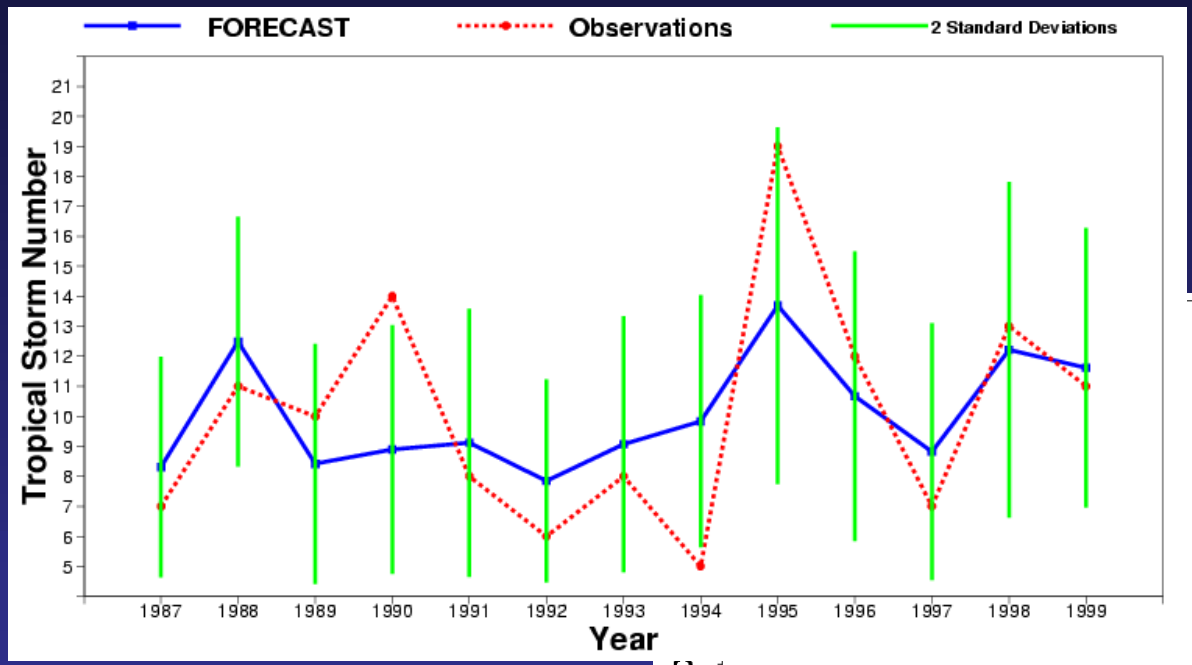
Tropical Cyclone Frequency

Linear correlation of the tropical cyclone frequency



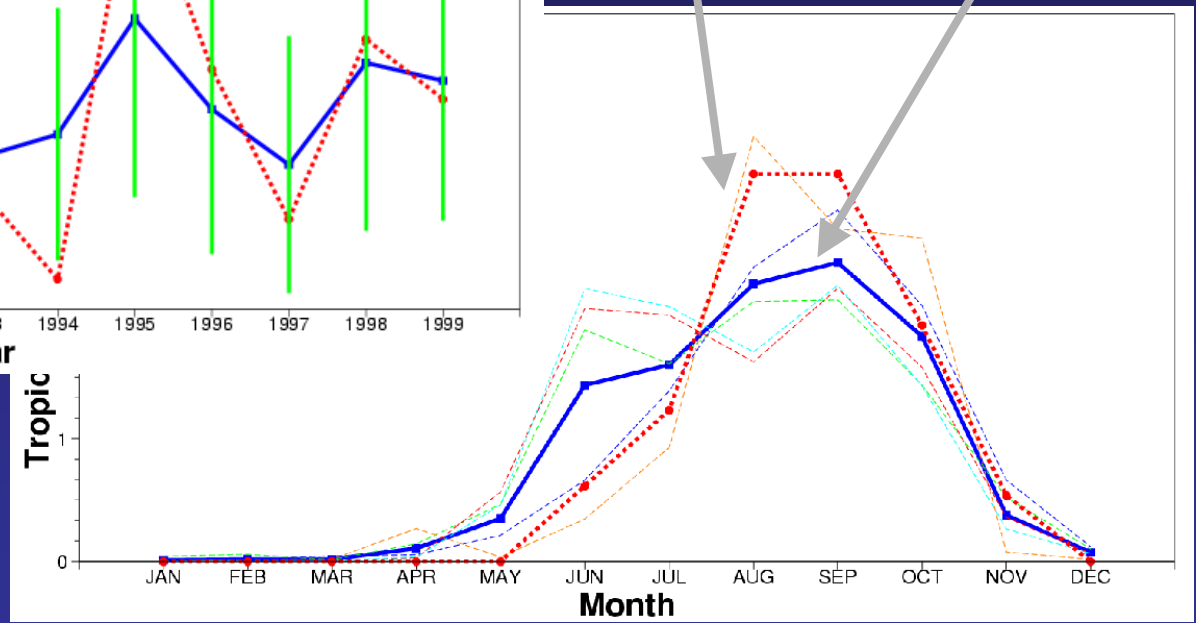
Tropical Cyclone Frequency

Tropical storm number for the tropical Atlantic (JJASO)



Verification

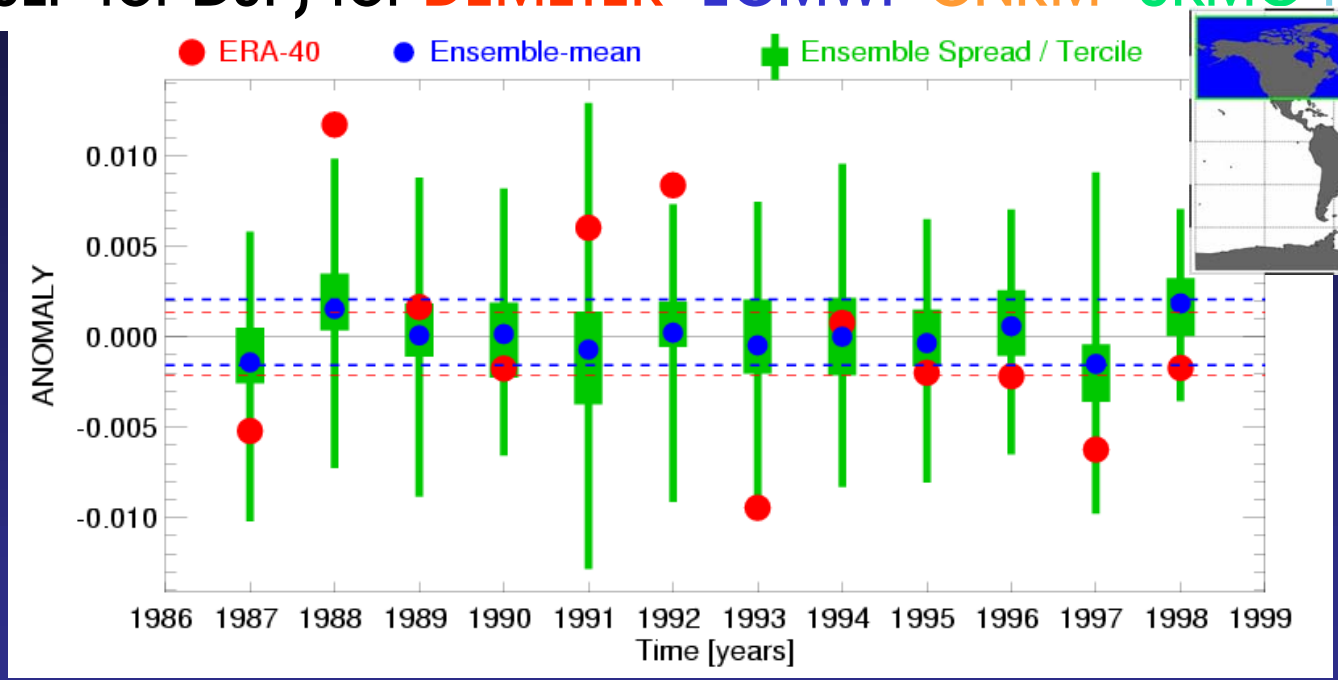
Multi-model

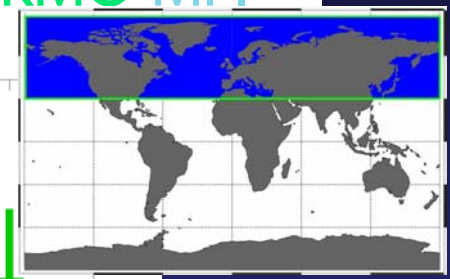


By F. Vitart (F.Vitart@ecmwf.int)

Multi-model Indices

Arctic Oscillation index (leading principal component of SLP for DJF) for DEMETER ECMWF CNRM UKMO MPI



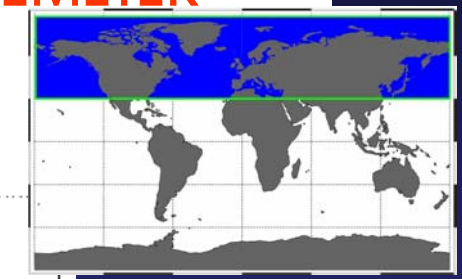


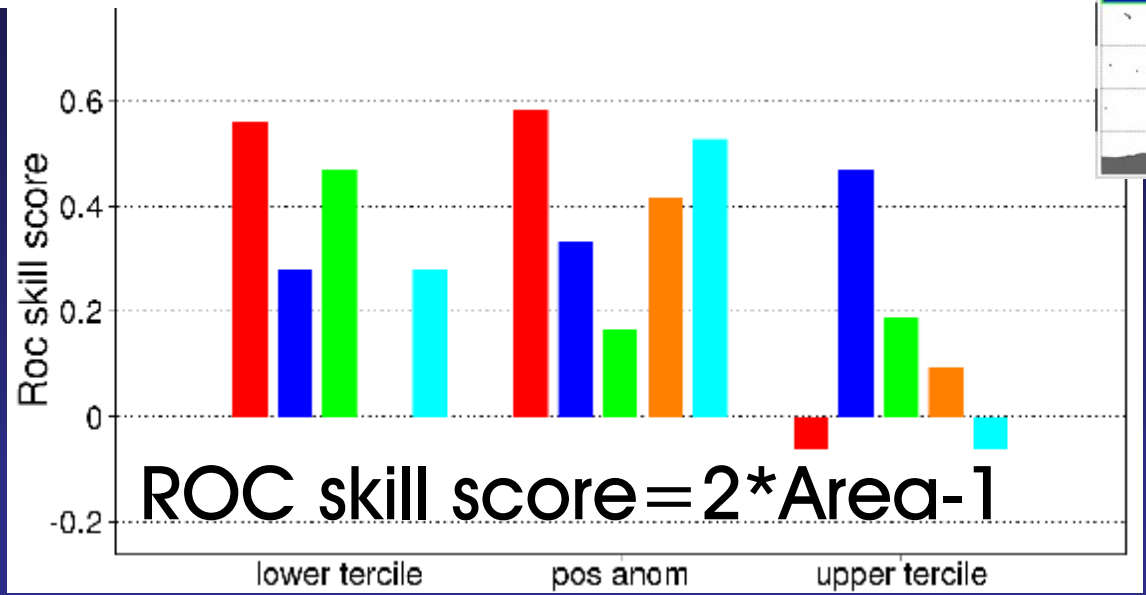
	Multi-Model	ECWMF	CNRM	UKMO	MPI
Correlation	0.48	0.26	0.40	0.28	0.15
RPSS	8.7	9.3	-8.8	9.3	6.9

Multi-model Indices

ROC skill score for Arctic Oscillation index (leading principal component of SLP for DJF) for DEMETER

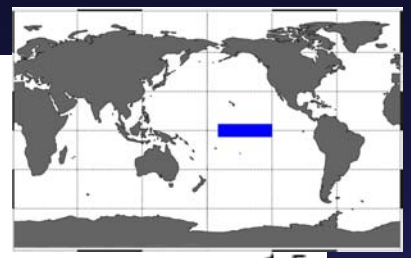
ECMWF CNRM UKMO MPI





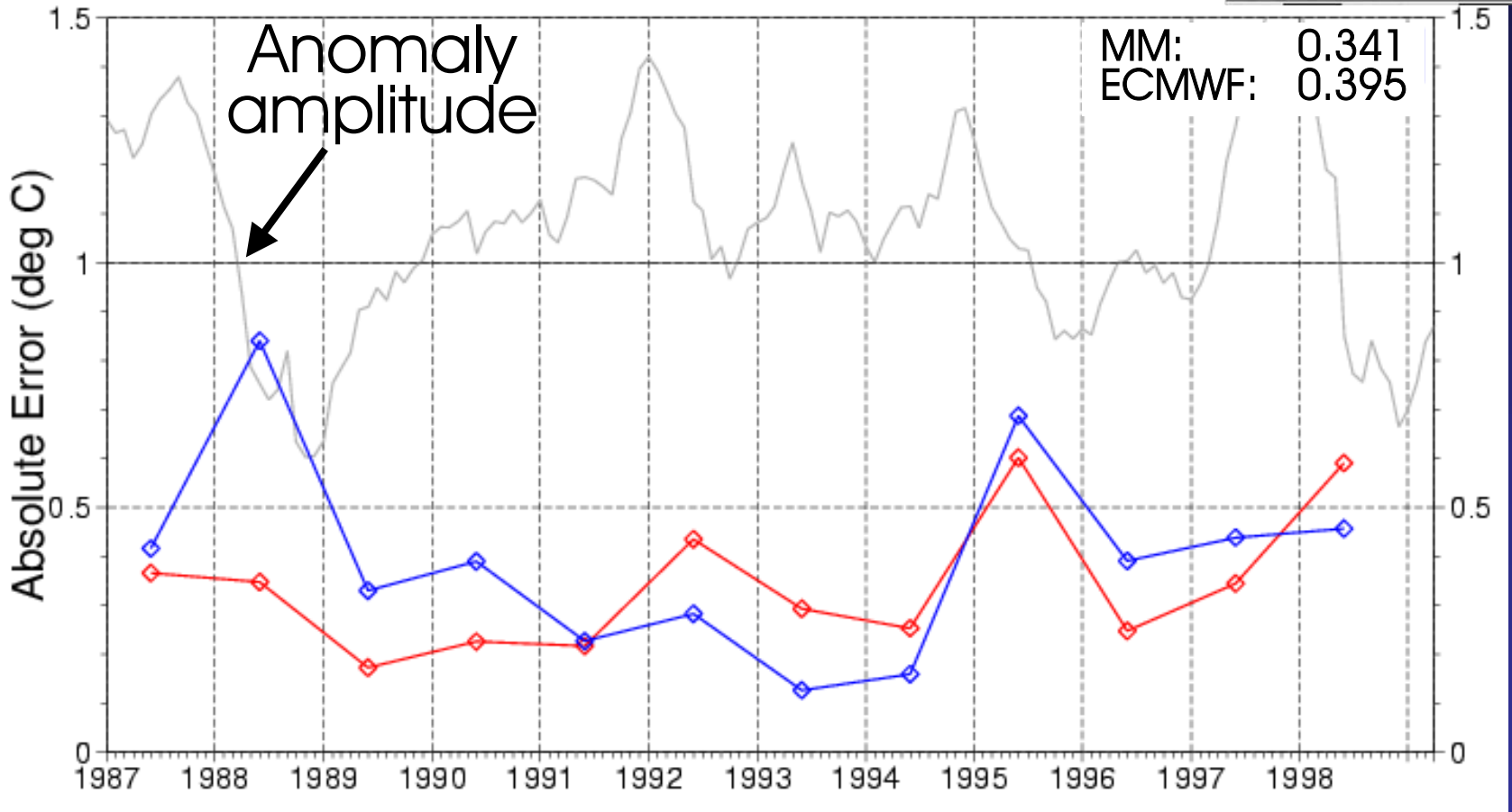
	Multi-Model	ECWMF	CNRM	UKMO	MPI
Correlation	0.48	0.26	0.40	0.28	0.15
RPSS	8.7	9.3	-8.8	9.3	6.9

Niño-3.4 MAE vs. Ensemble Size

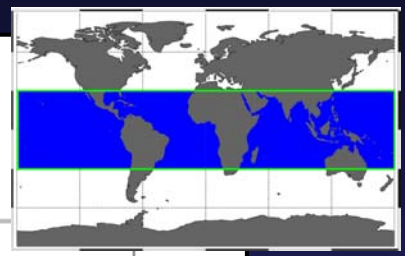


Mean Absolute Error: 1987 – 1998, May starts

DEMETER ECMWF (36 members)



Impact of Ensemble Size



Model Combination

- Multi-model approach, in the end, looks for consensus. This can be achieved by:
  - Merging models with equal weights
  - Estimating optimal weights for each model
- Weight estimation is based on past model performance
- Main difficulty: Finding robust weights

Model Combinations

- Different approaches are available:
  - Robust regression (ensemble mean)
  - Bayesian combination (parametric)

$$\frac{\hat{\mu}_t}{\hat{\sigma}_t^{\,2}} = \frac{\hat{\mu}_{0t}}{\hat{\sigma}_{0t}^{\,2}} + \frac{\beta^2}{\gamma V_t}\left(\frac{\overline{X}_t - \alpha}{\beta}\right)$$

  - Probabilistic approach (minimize Brier-score)

$$p_c = \sum_{i=1}^{M} a_i p_i \qquad a_i -> Min\ BS$$

1) Niño-3.4 Bayesian Combination

5-month lead monthly hindcast (In collaboration with Univ. of Reading)

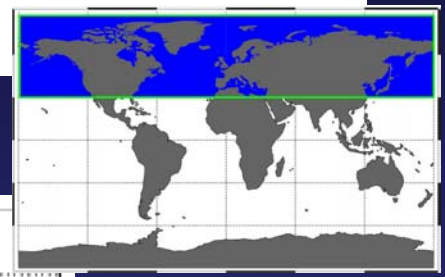
Hindcast	MAE (°C)	Skill Score	Uncertainty
Climatology	1.16	0%	1.19
Empirical	0.53	55%	0.61
Ensemble	0.37	68%	0.39
Combined	0.31	74%	0.32

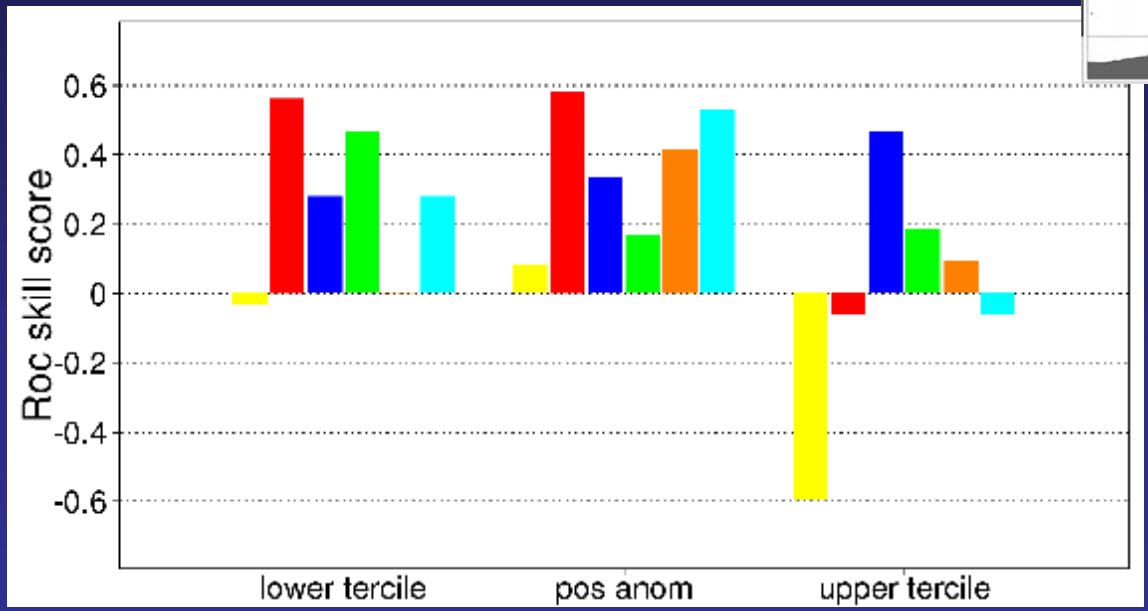
The combined hindcast has:

- A large increase in MAE (and MSE) forecast skill
- A realistic uncertainty estimate

2) Brier Score Minimization: Results

ROC skill score for Arctic Oscillation index (leading principal component of SLP for DJF) for Comb DEMETER ECMWF CNRM UKMO MPI

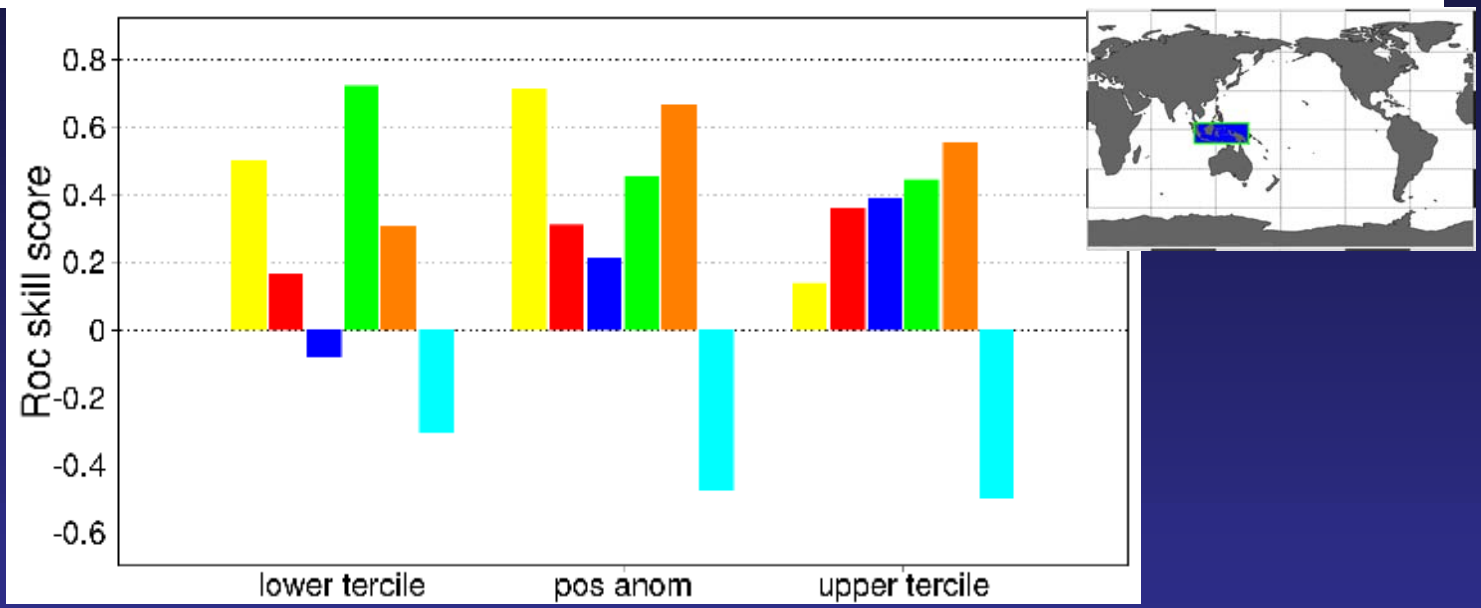




2) Brier Score Minimization: Results

ROC skill score for Indonesian precipitation (MAM) for

Comb DEMETER ECMWF CNRM UKMO MPI



Proportion of times an index performance are better than a given model (based on ROC):

Comb: 48%

DEMETER: 60%

Conclusions

- The multi-model gives, overall, the best results, increasing both the predictability and the reliability of the hindcasts. This behaviour is not just a result of increasing the ensemble size
- Preliminary combination tests provide some positive results, though it does not beat the simple average multi-model