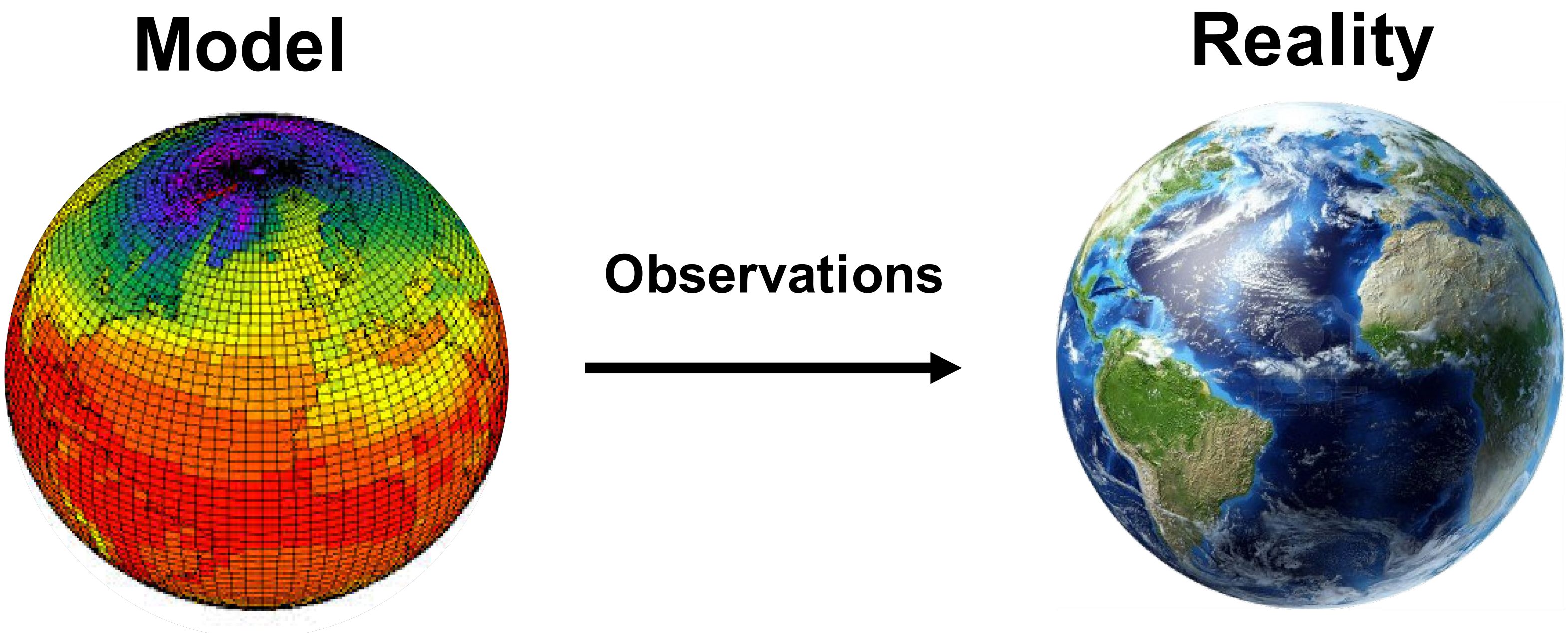
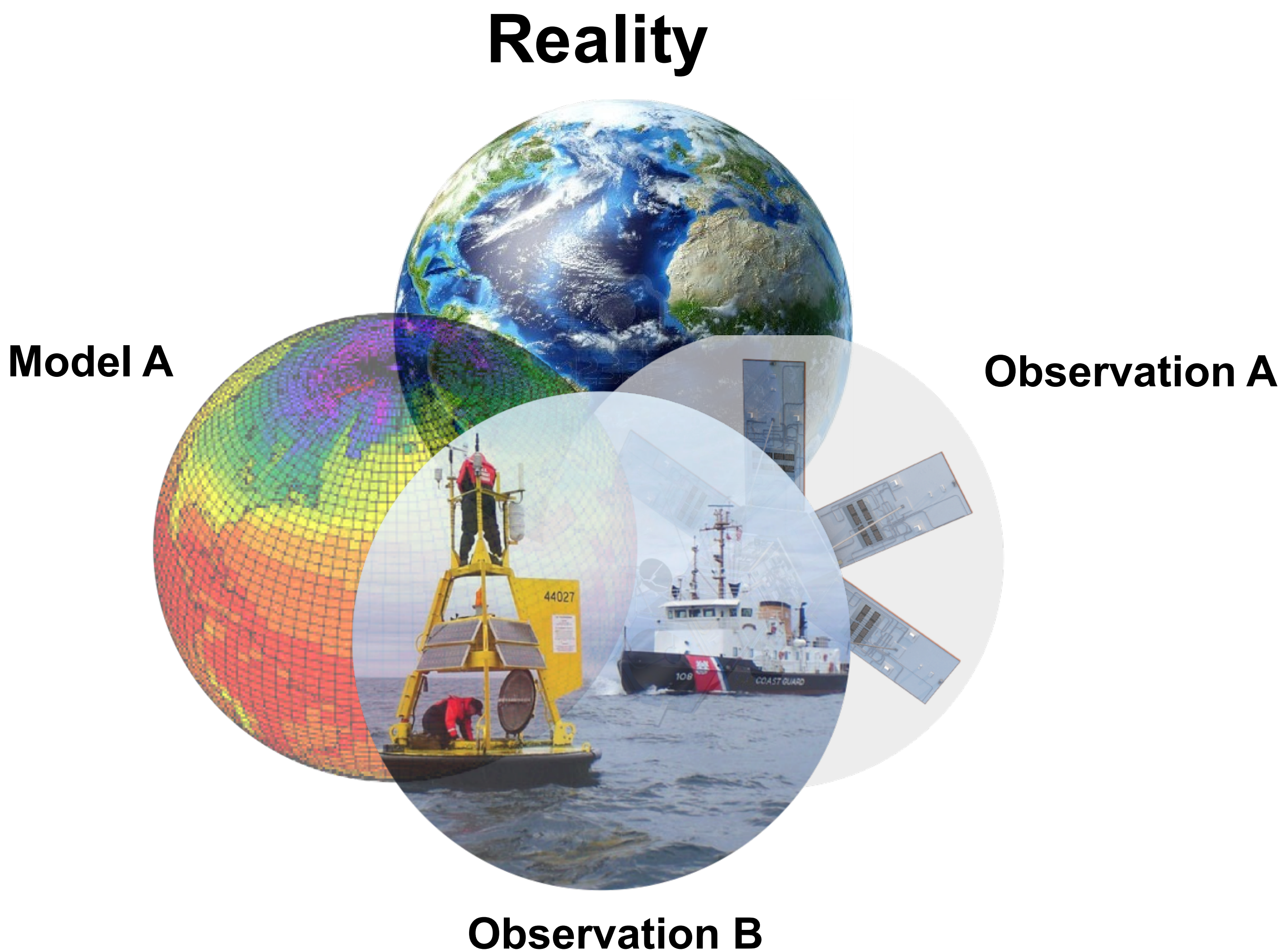
Traditional evaluation perspective



A shift in the paradigm



A traditional verification question

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

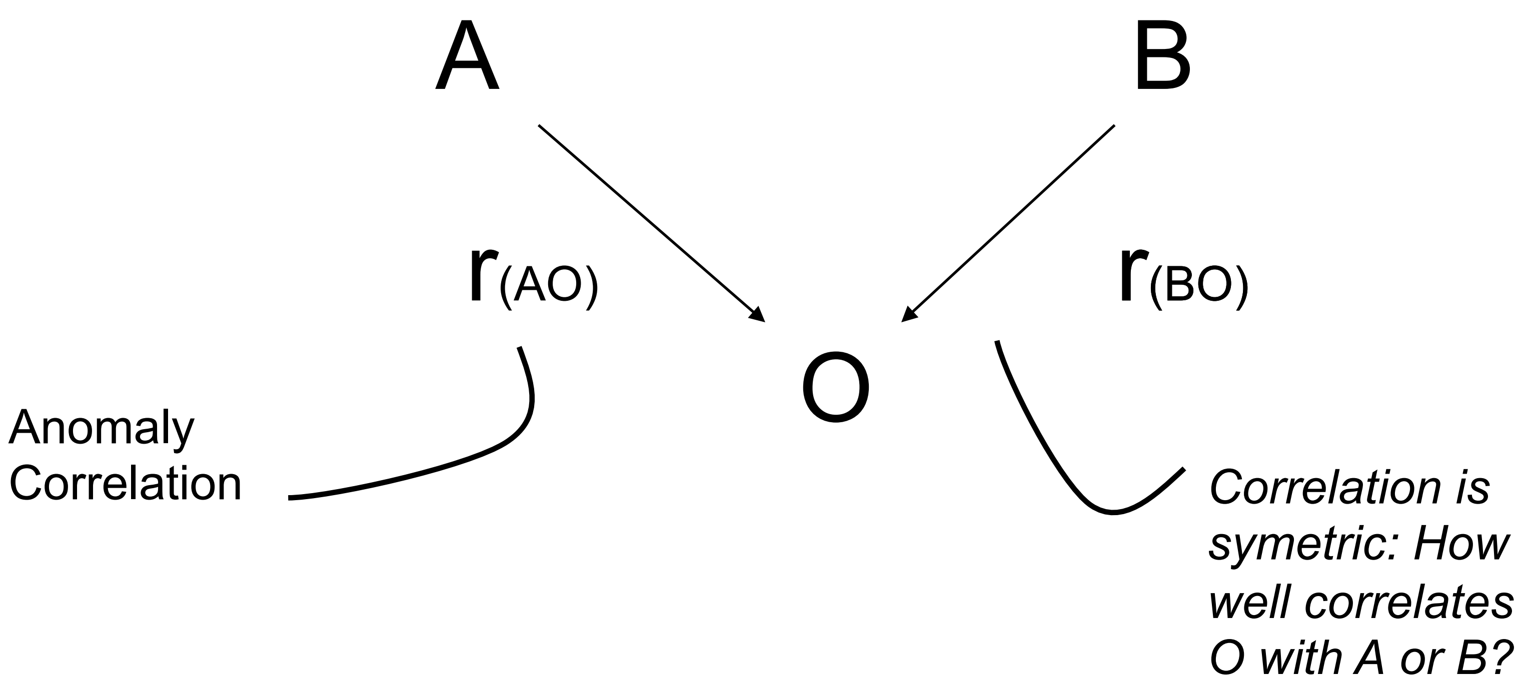
EXCELENCIA SEVERO OCHOA

Is model system B superior to model system A?

A	B
Low horizontal resolution	High horizontal resolution

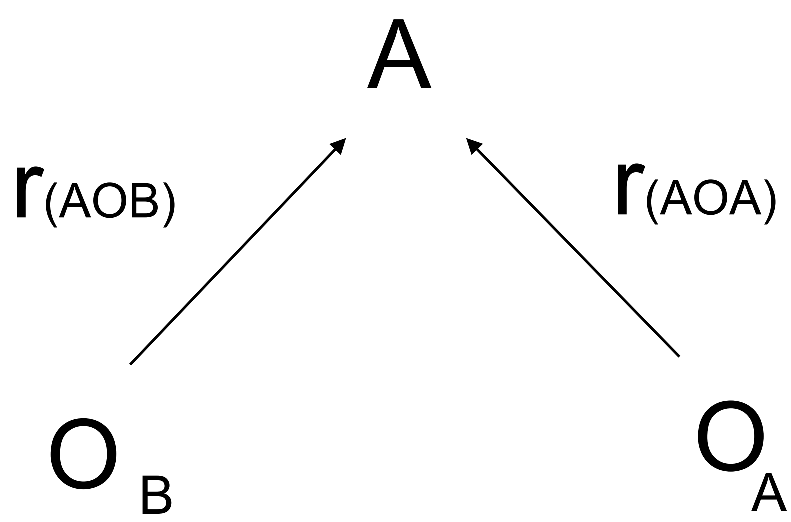
Comparing climate forecasts

Compare hindcast skill with an observation

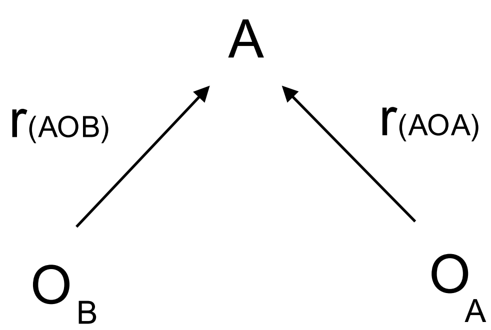


Reversing the question

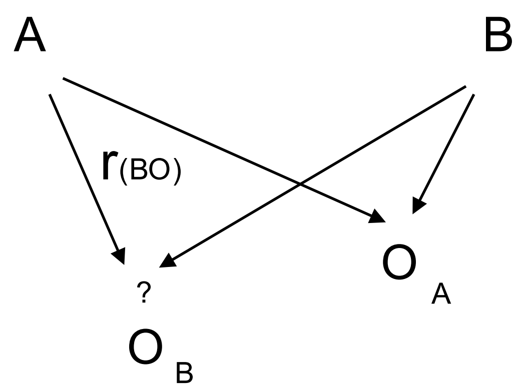
Which observation is better? A useful question?



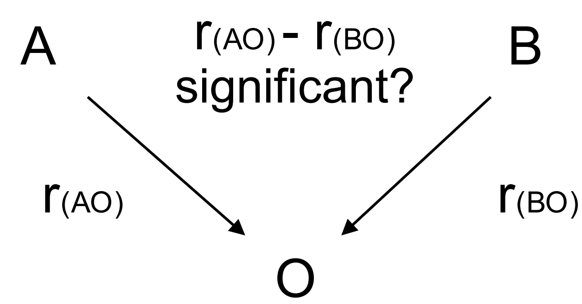
Overview this presentation



(a) Which observations has the smallest error

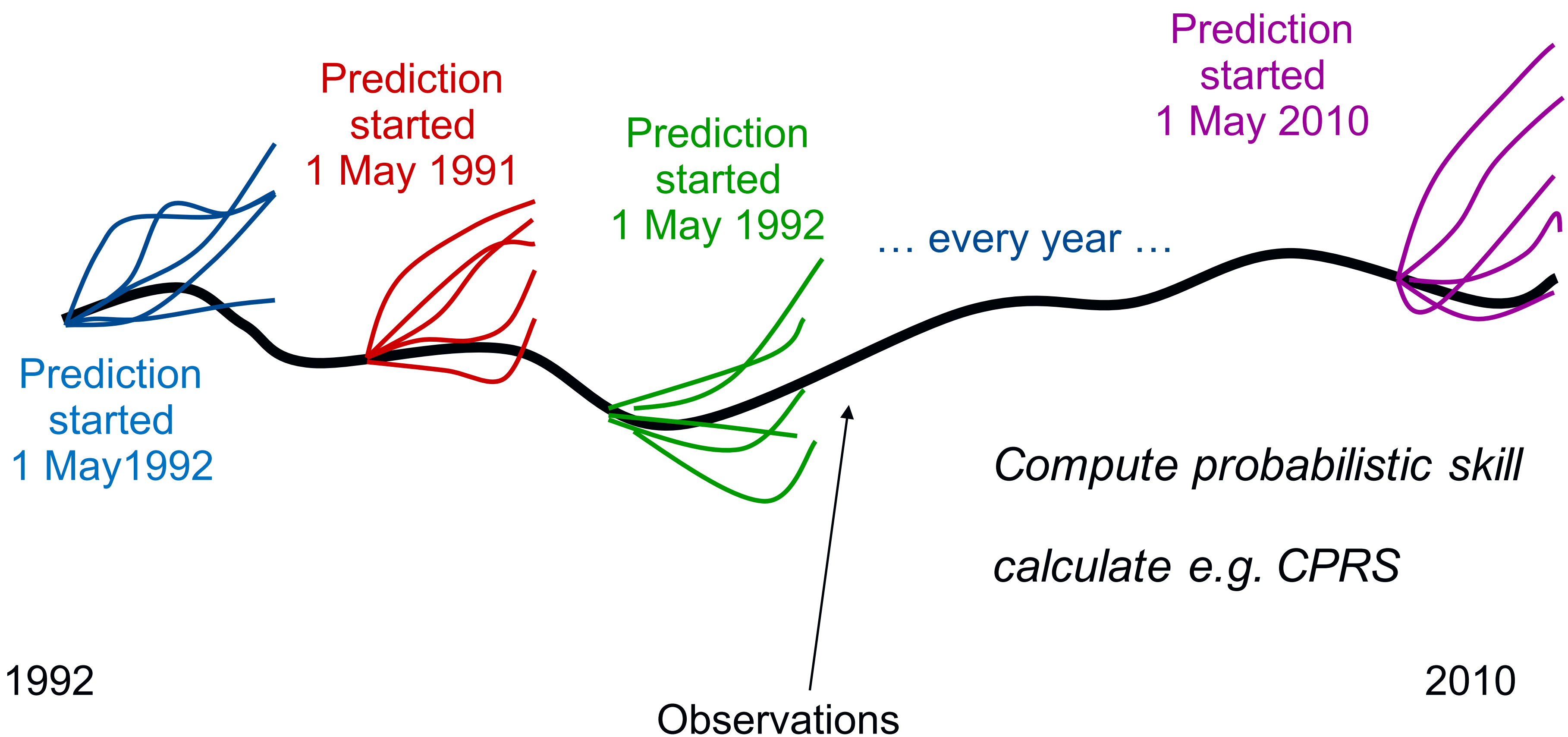


(b) How important is observational uncertainty



(c) How to detect improvements in models or observations

Seasonal forecast skill

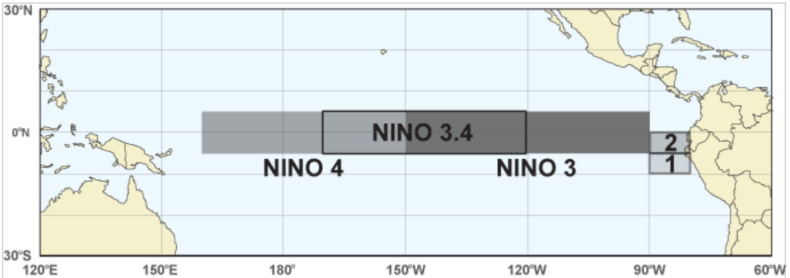


Compute probabilistic skill

calculate e.g. CPRS

Seasonal forecast skill

Target: Global and Niño3.4 SST



Observations: 4 Sea-surface Temperature (SST) observations: ESA-CCI, HadISST, ERSST4, ERA-Interim

Prediction started 1 May1992

Prediction started 1 May 1991

Prediction started 1 May 1992

… every year …

Prediction started 1 May 2010

Compute ensemble-mean to distill climate signal

calculate e.g. anomaly correlation

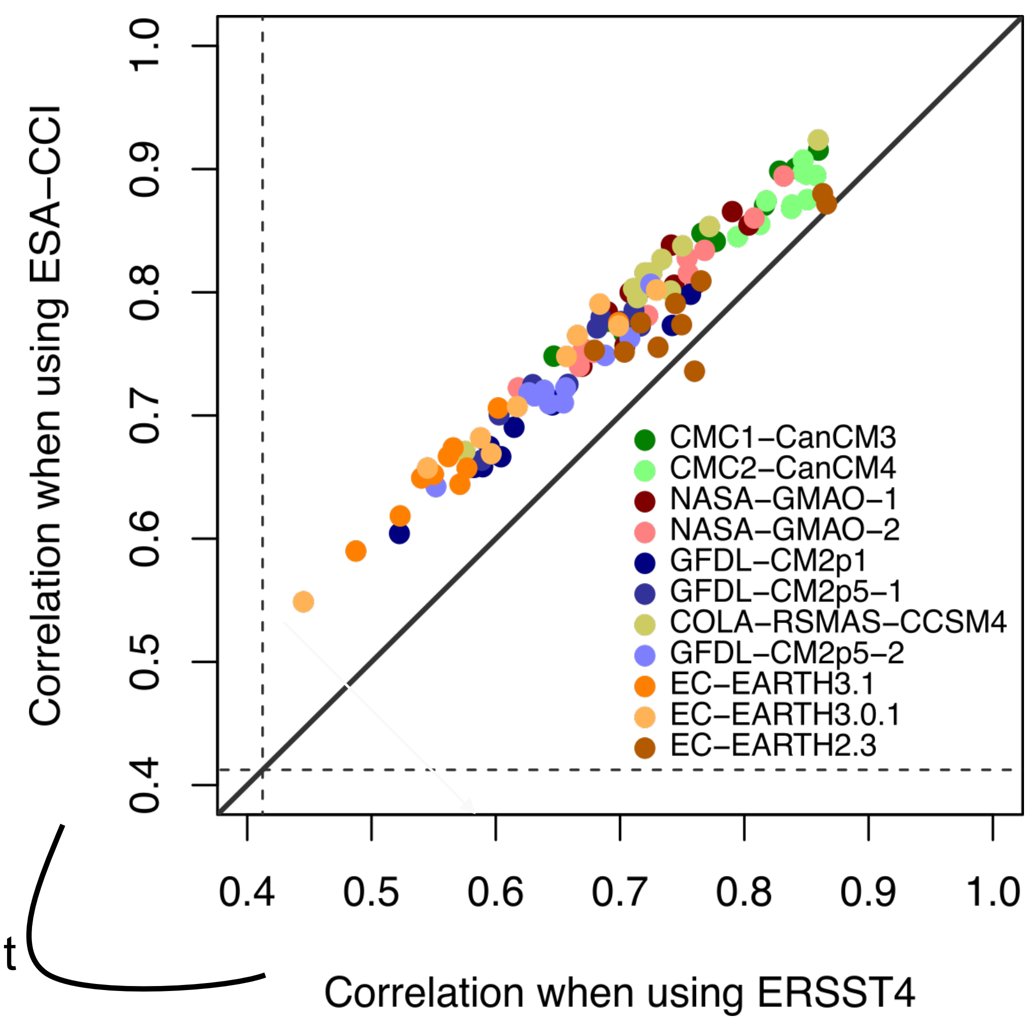
Models: EC-Earth (3 versions), ECMWF S4, North American Multi-Model Ensemble (NMME, 7 models)
10 – member forecast each

Reversing verification question

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

CCI SST yields systematic higher correlation skill across many models

Independent from the models

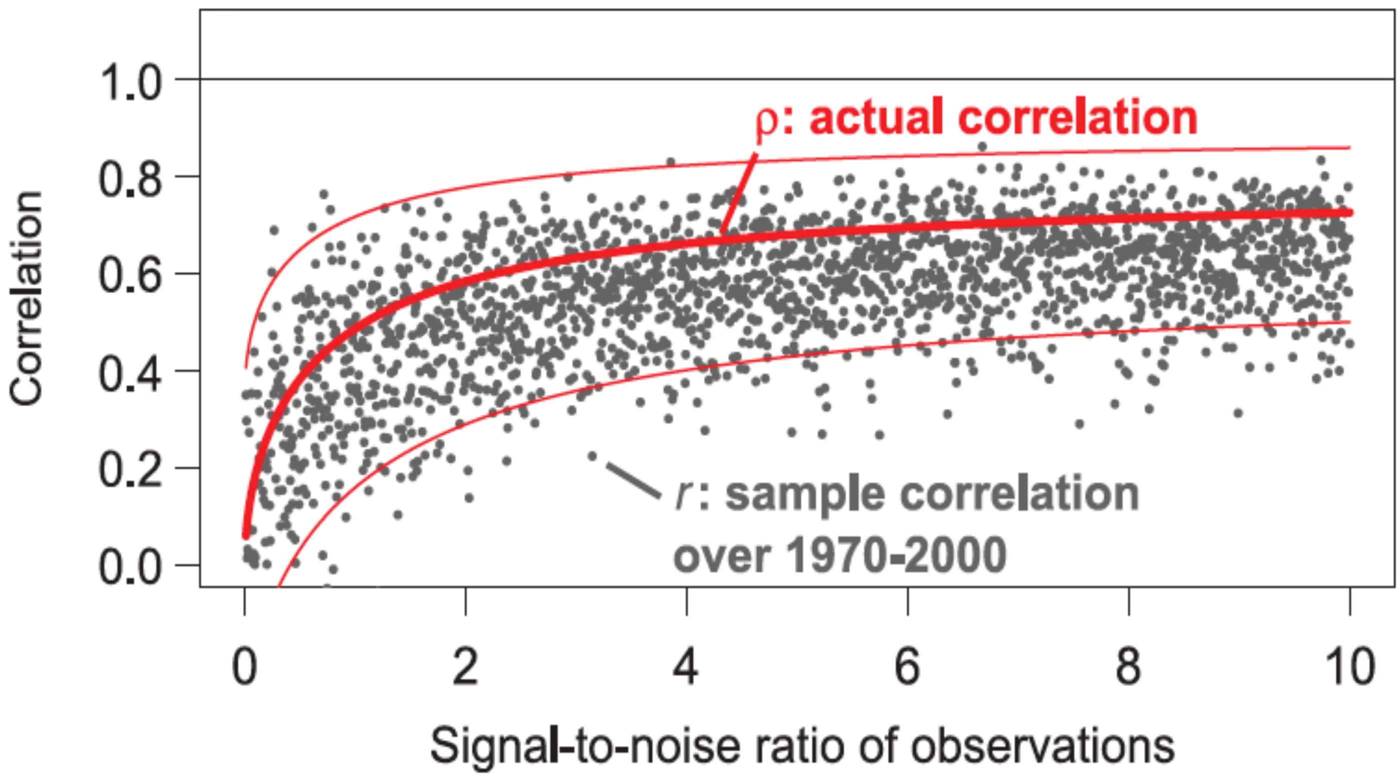


Fully independent datasets

Massonnet et al. (2016) Science

Effect observational uncertainty

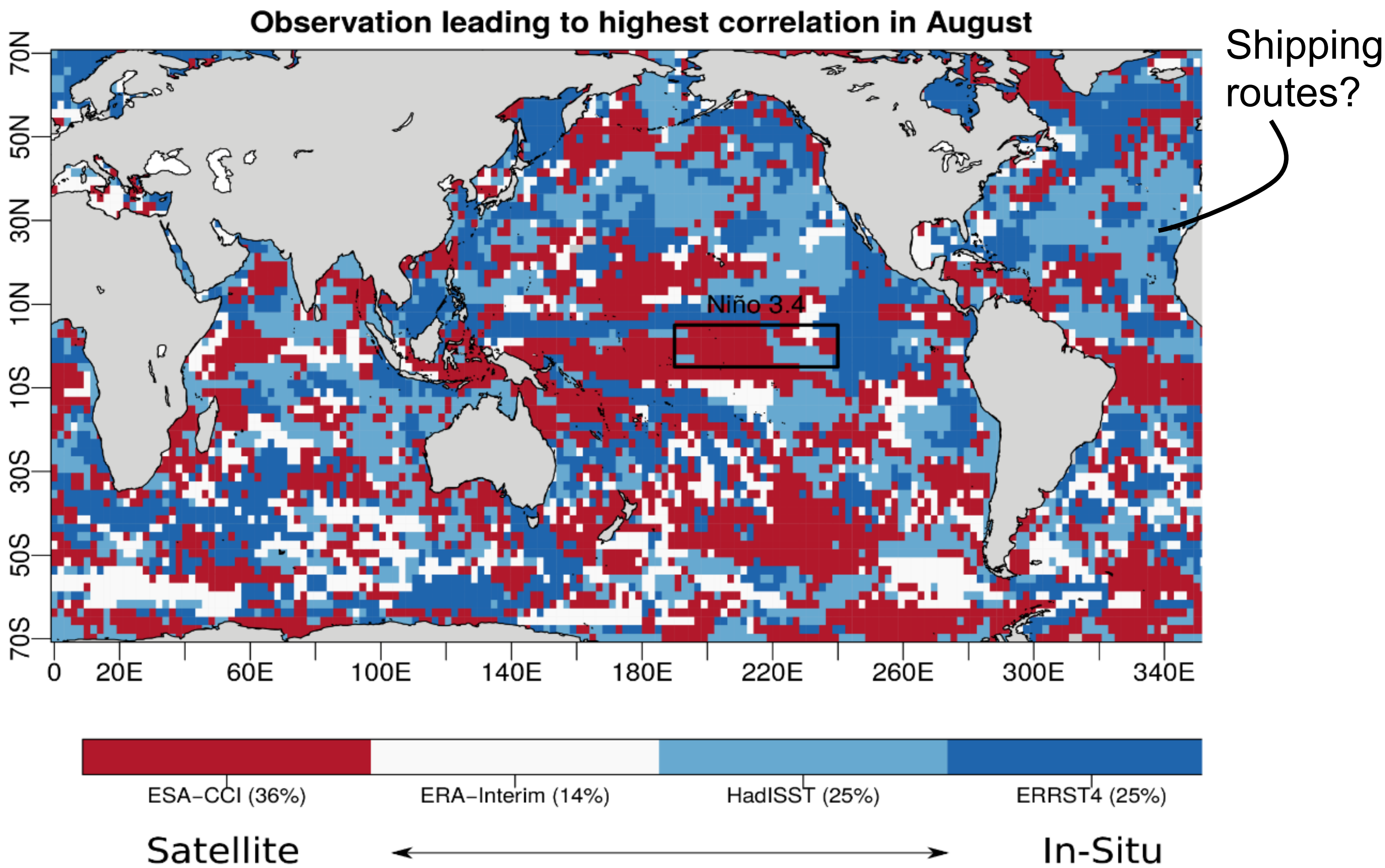
Correlation reduces with noise either co-variates: observational uncertainty reduces forecast skill



Massonnet et al. (2016) Science

Reversing verification question

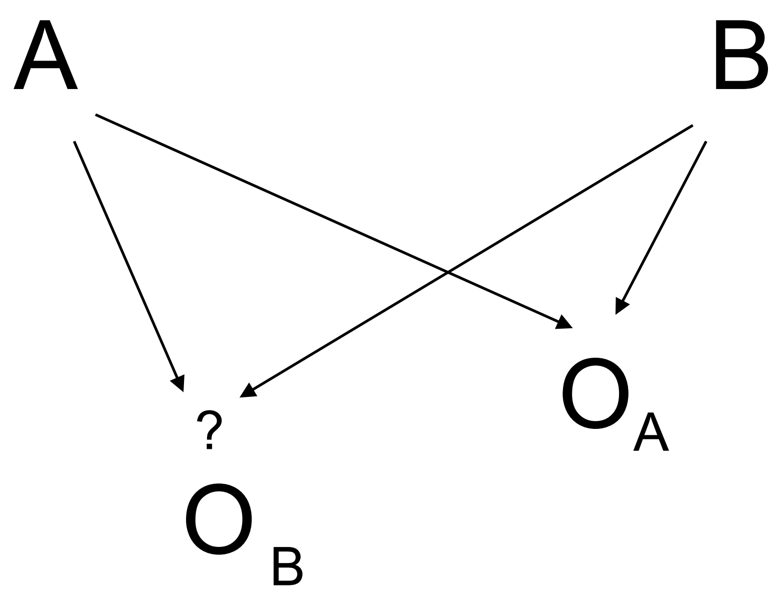
Choice of observation may differ on the location, overall CCI best



Massonnet et al. (2016) Science

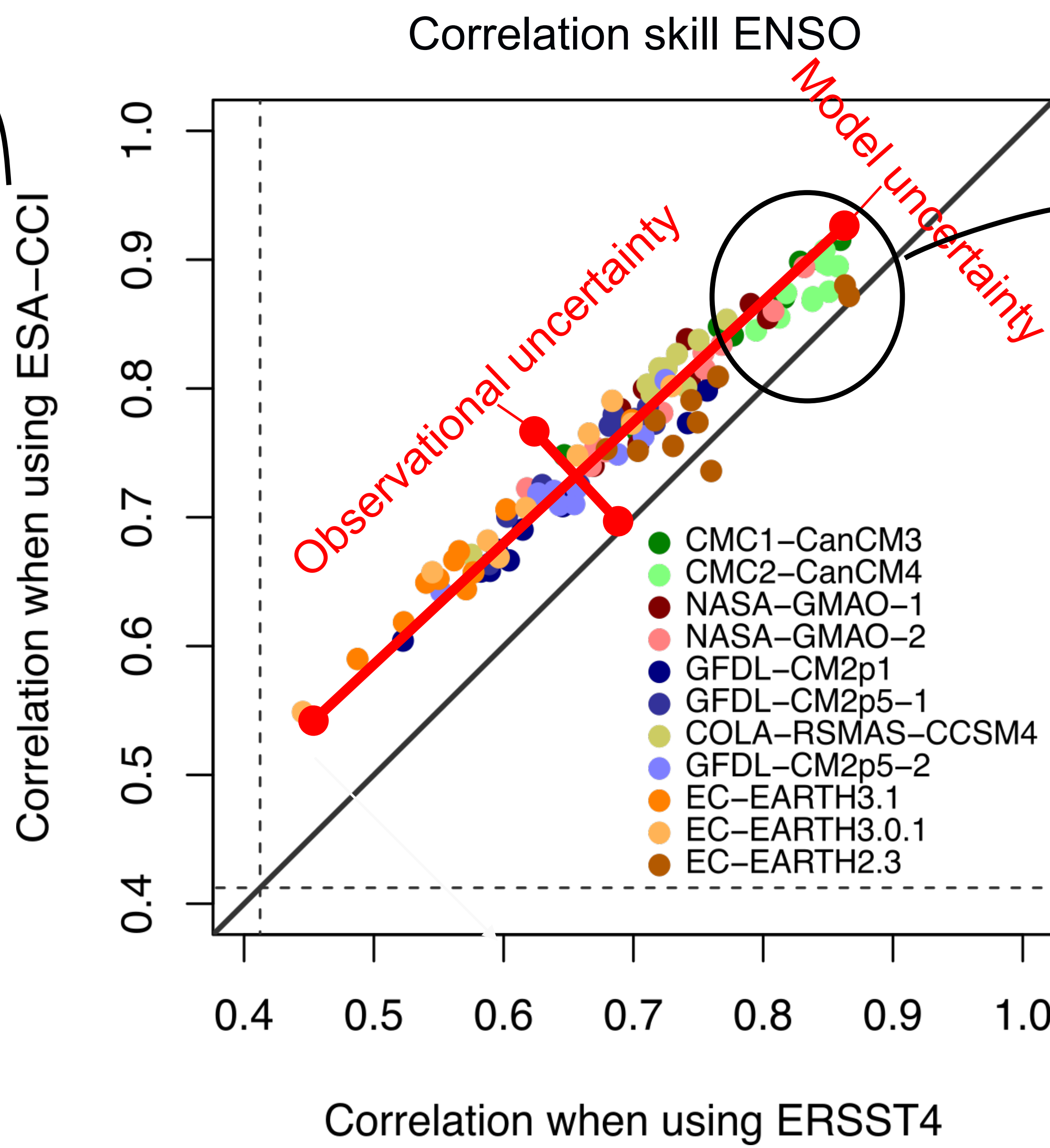
Observational uncertainty

How important is the observational uncertainty?



Acknowledging joint uncertainty

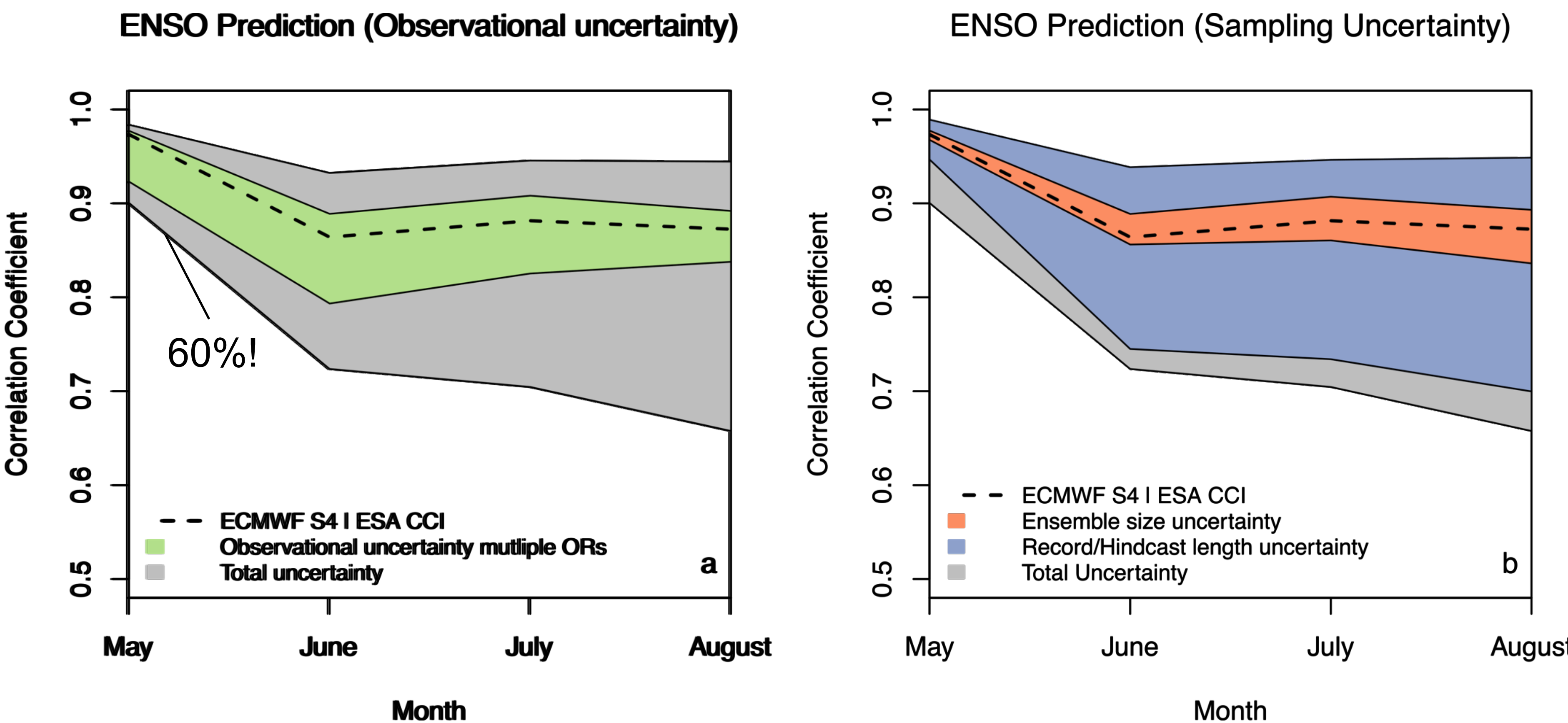
Independent from the models



How important for individual system?

Decomposition of uncertainties

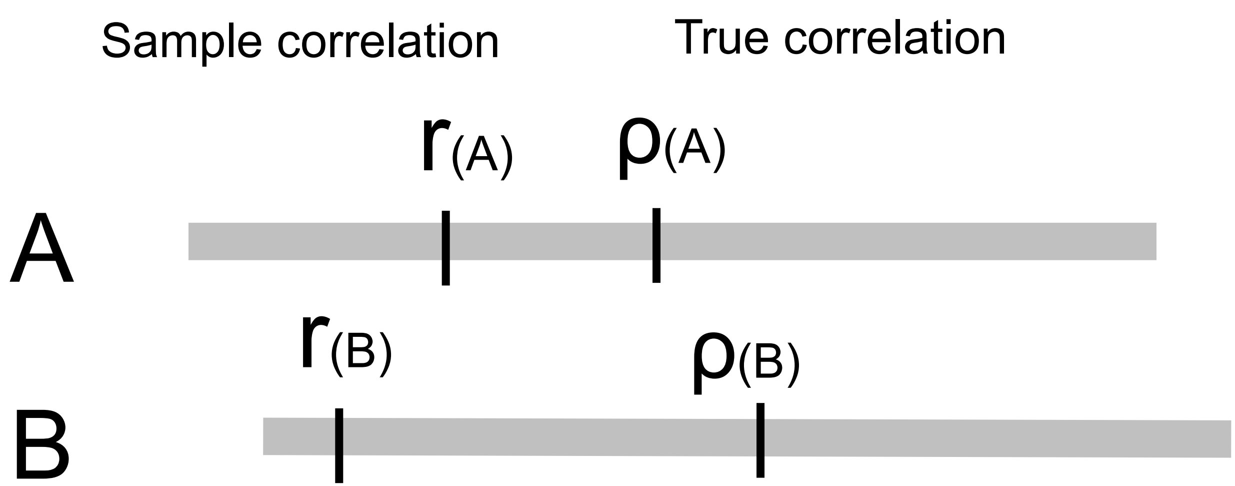
Comparison to sample uncertainties: observational uncertainty is an important source of verification uncertainty for ENSO



Bellprat et al. (2017) Remote Sensing of the Environment

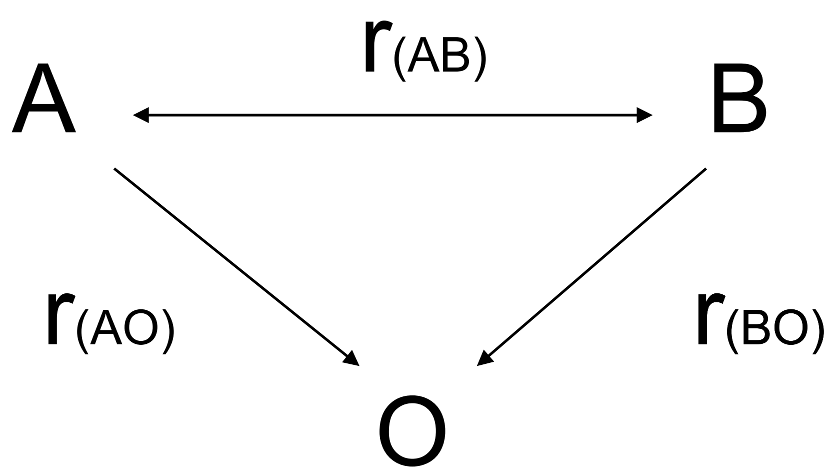
Detecting statistical differences

Are the differences in performance of models or observations significant $r_{(CCI)} > r_{(ERSST)}$?



Dependence

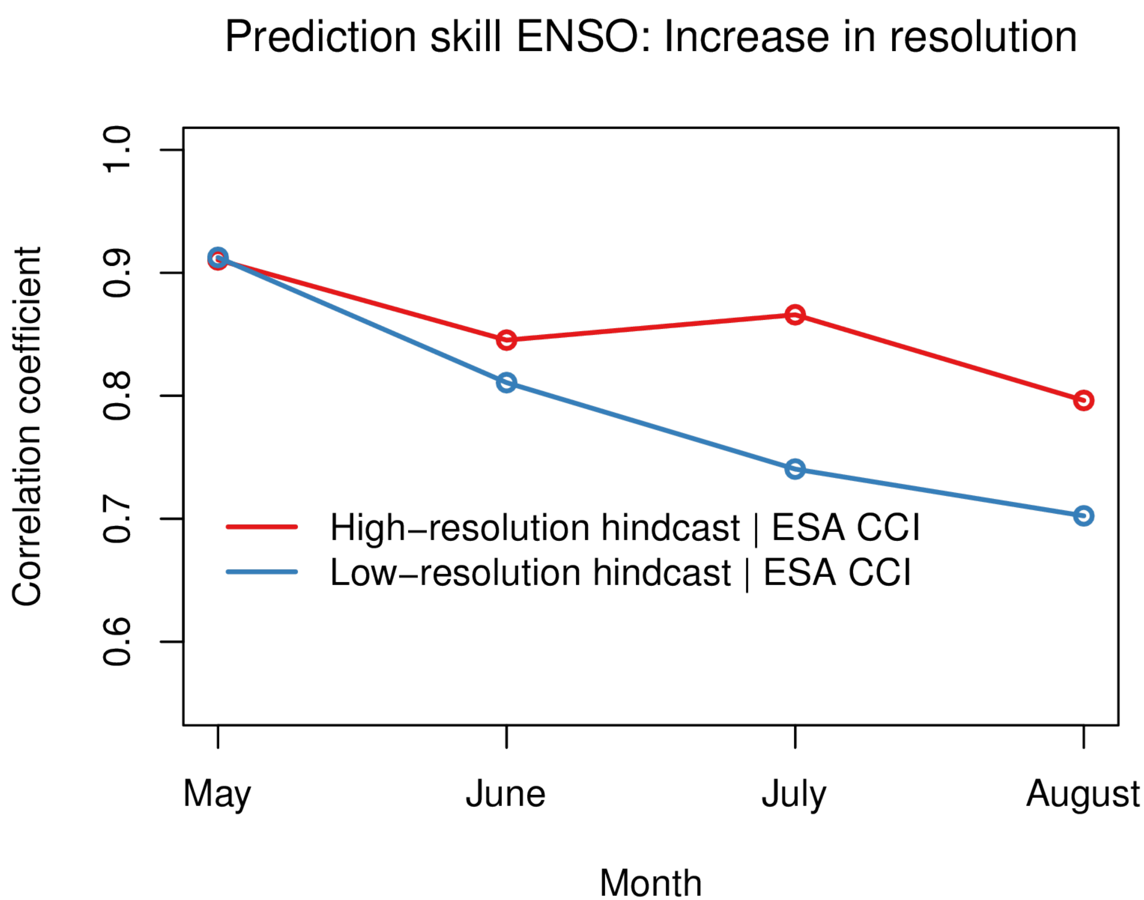
Models and observations are statistically dependent!



Fisher-test (common test in community) assumes independence while newer tests (Steiger, 1980, Zou, 2007) don't.

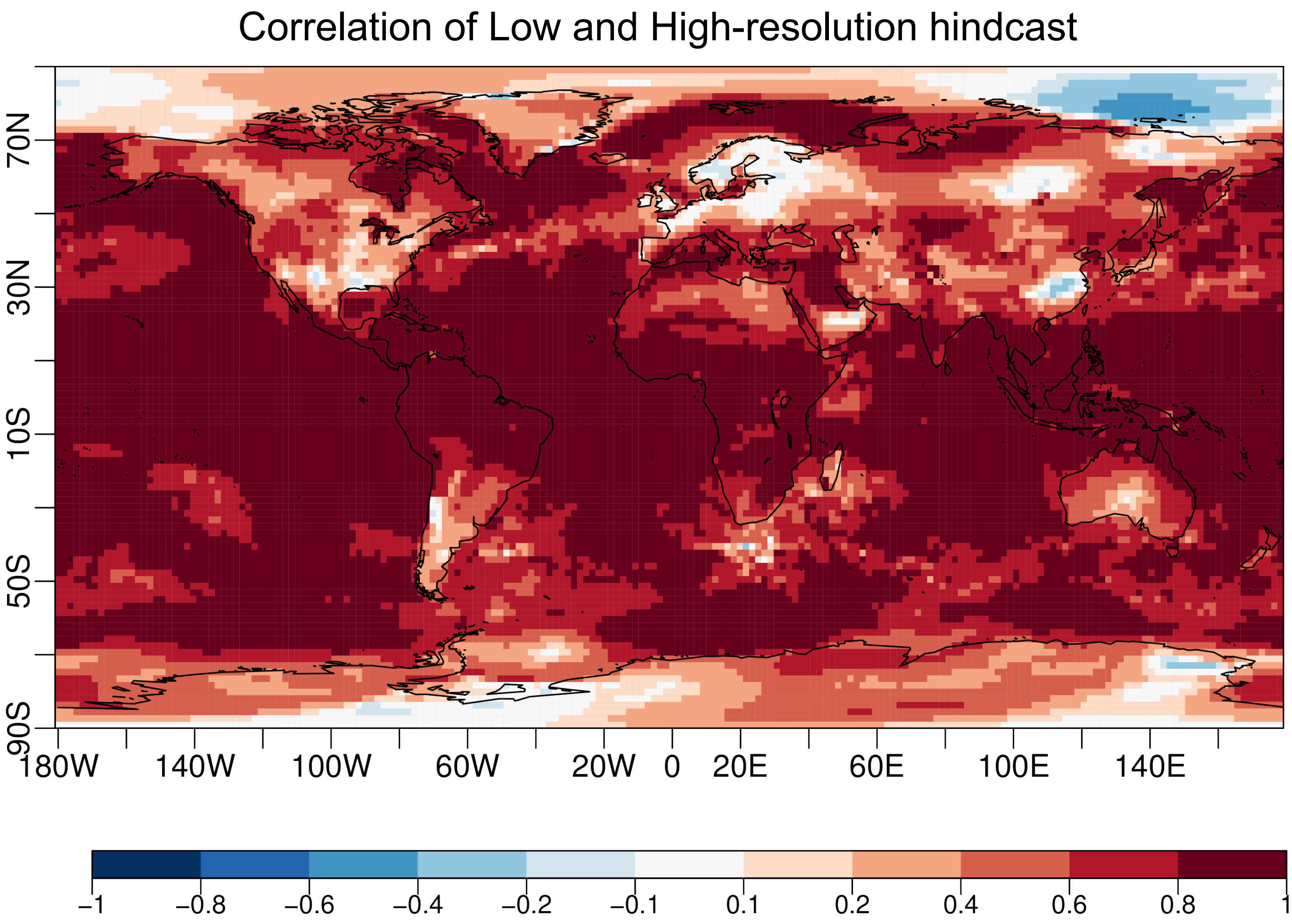
Example ENSO and resolution

High-resolution hindcasts improves El Niño Southern Oscillation (ENSO) predictions, but change not significant at 5% (Fisher-test)

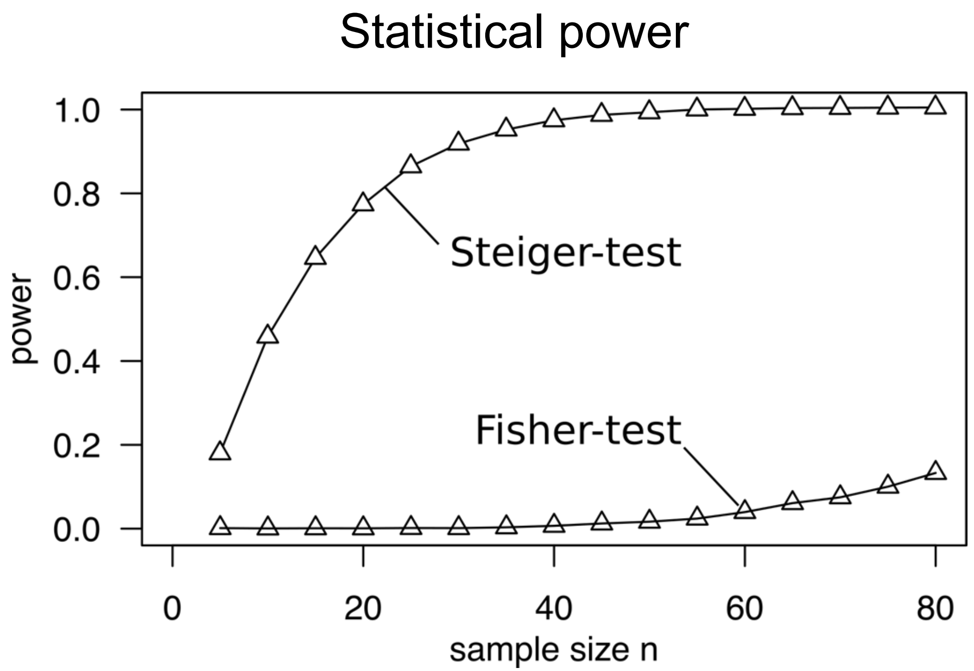


Bellprat et al. (2015) Tech. Rep. BSC 002

Model dependence



Power to detect a difference between increases dramatically. Improvement now statistical significant at 1% level.



In medicinal science only studies with power > 80% are accepted, a guideline for forecasting?

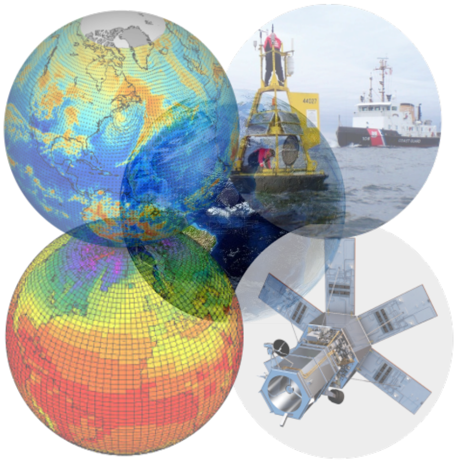
Siegert et al. (2017) Monthly Weather Review

A joint uncertainty perspective

Models and observations are both approximations of the truth and uncertainty in both sources can be important.

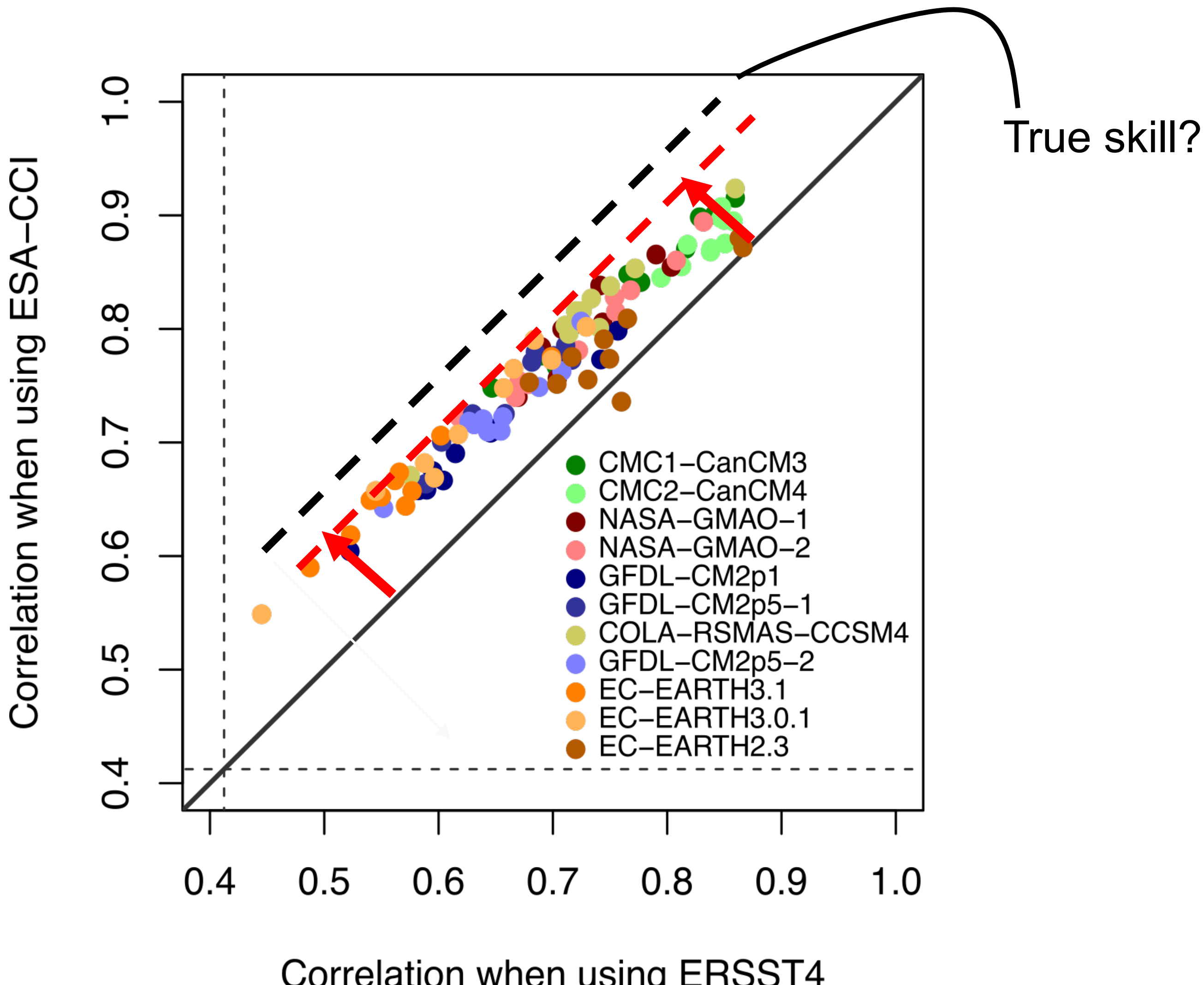
Models can be valuable in assessing observational quality and thus guide a more objective dataset selection

Testing improvements in models and observations requires the consideration of dependence between all source of information



Outlook: What is the “true” skill ?

True climate predictions skill is systematically underestimated due to uncertainties in the observations

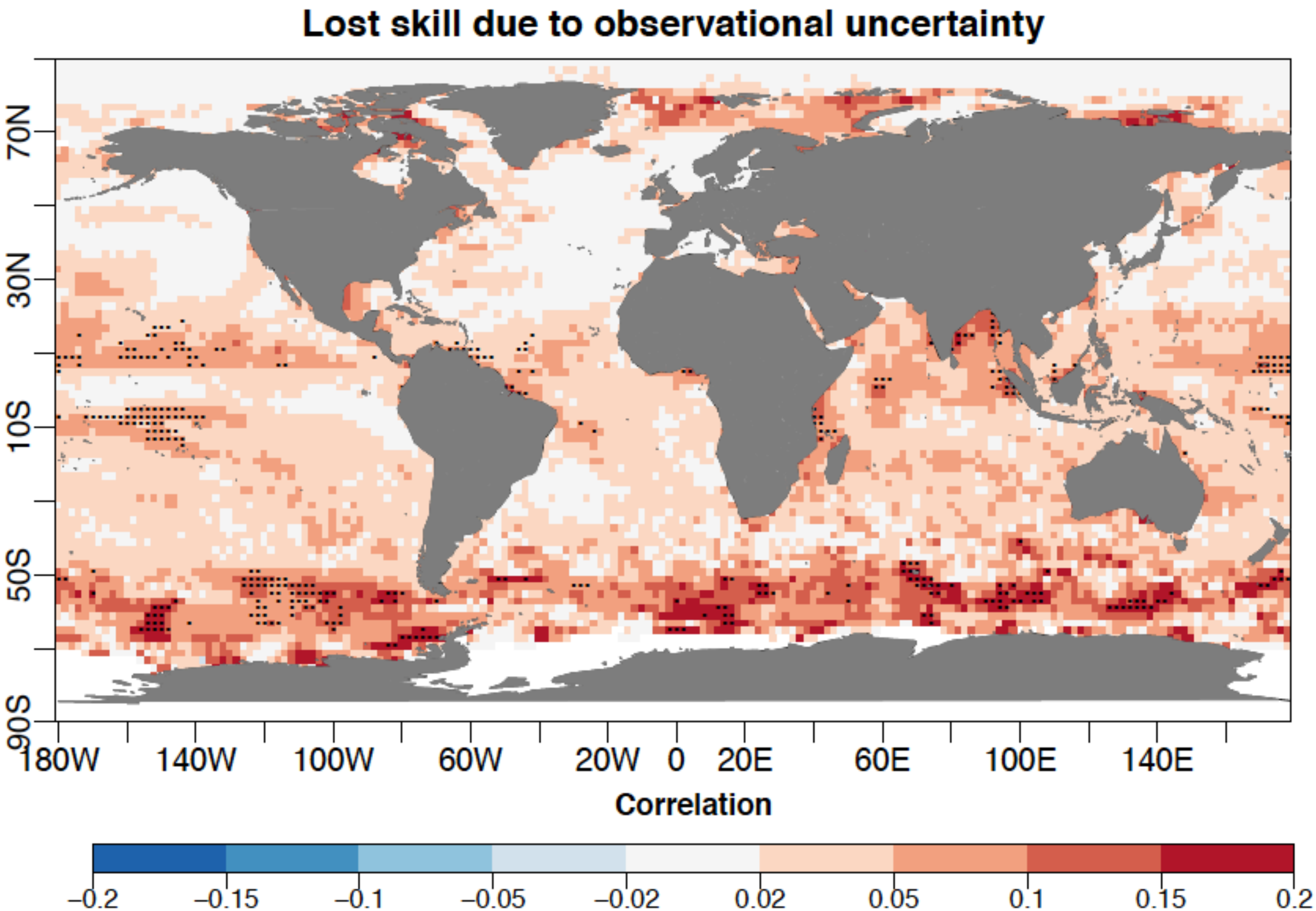


Underestimation of “true” skill

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

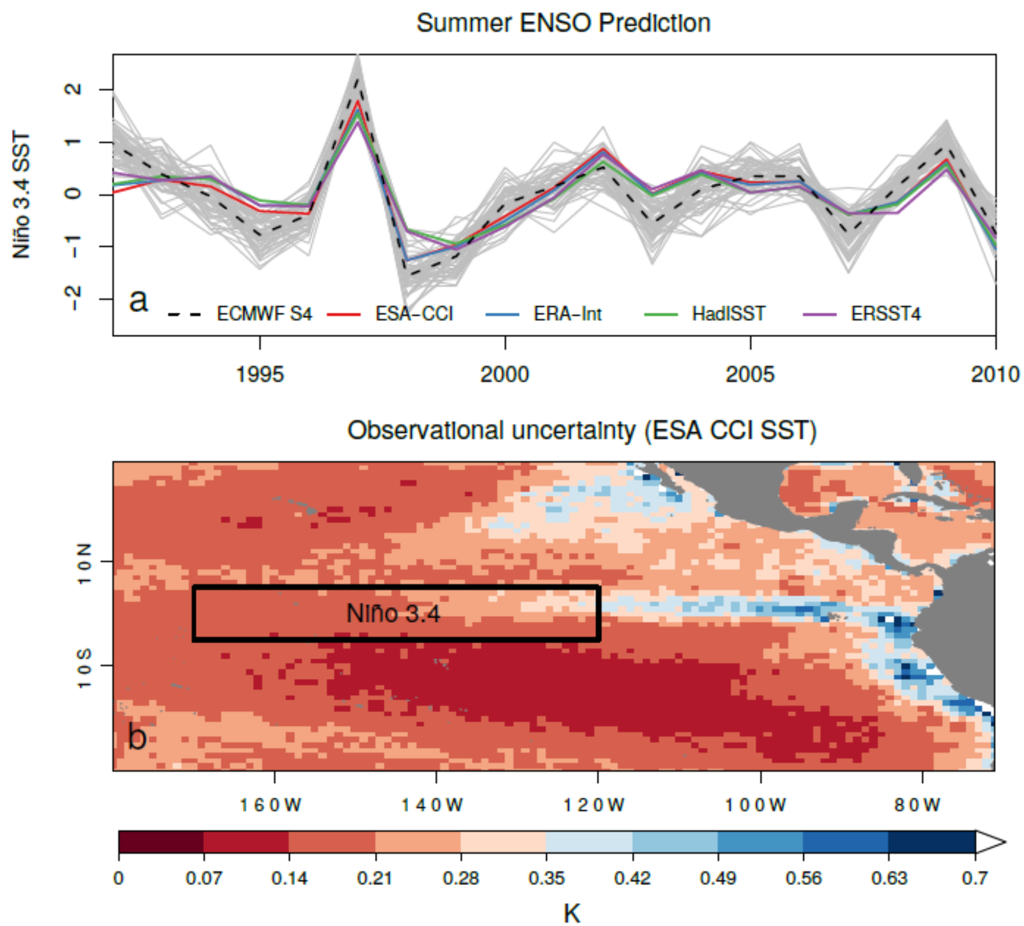
EXCELENCIA SEVERO OCHOA

Seasonal SST forecast skill is underestimated up to 0.2 correlation



Observational uncertainty CCI

Model evaluation often requires spatial and temporal averaging, requires the consideration of error correlation scales



Monthly Niño3.4

Uncorrelated uncertainty reduces by $1/\sqrt{N}$ but errors are not uncorrelated!

Daily, 4 km

Bellprat et al. (2017) Remote Sensing of the Environment

Outlook

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Quantifying observational uncertainty is a challenge and propagation scales represented to by the models is a big gap – A stronger interaction is required with the observational data community

ESA *Climate Modelling User Group (CMUG)* is going to explore observational uncertainty in model – observation inter-comparison strongly in the future

Metrics and new statistical tests are required that can make use of the observational uncertainty data that future data sets are going to provide

www.bsc.es

Thank you!

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

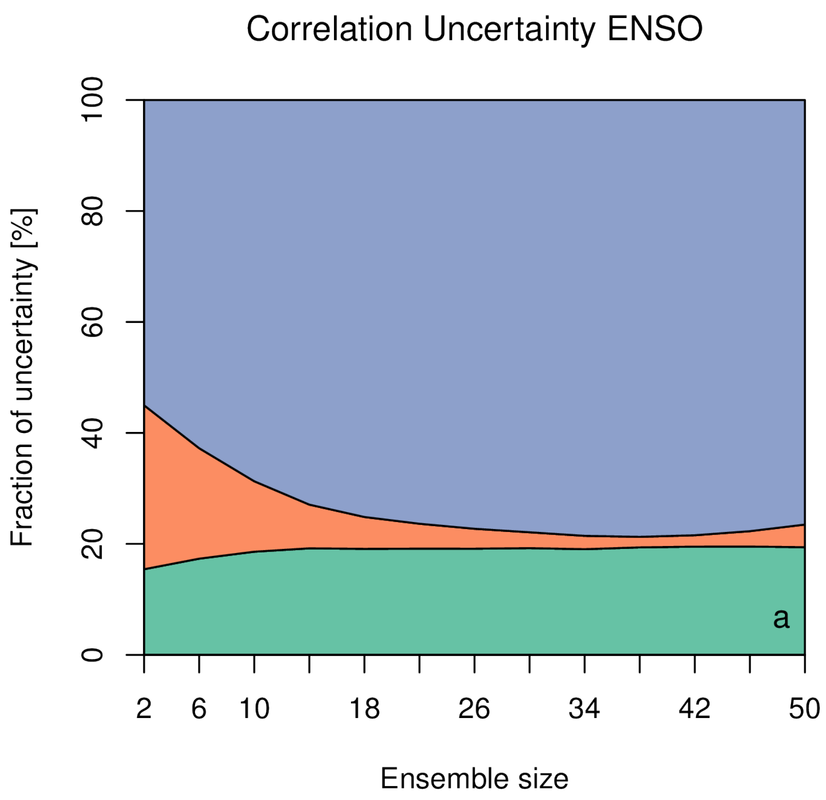
Massonnet, F., Bellprat, O., Guemas, V., Doblas-Reyes, F. J., (2016). Using climate models to estimate the quality of global observational data sets, *Science (AAAS)*

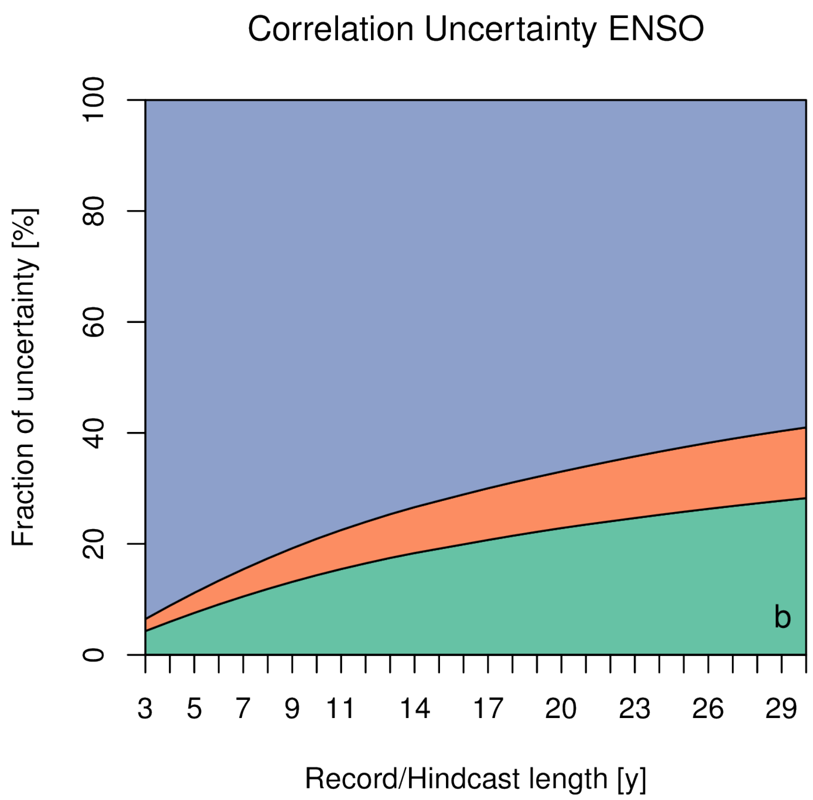
Bellprat, O., Massonnet, F., Siegert, S., Guemas, V., Doblas-Reyes, F. J. (2017). Exploring observational uncertainty in verification of climate model predictions, *Remote Sensing of the Environment (RSE), in review*

Siegert, S., Bellprat, O., Menegoz, M., Stephenson, D., Doblas-Reyes, F. (2016). Detecting improvements in forecast correlation skill: Statistical testing and power analysis. *Monthly Weather Review*

Extra Slides

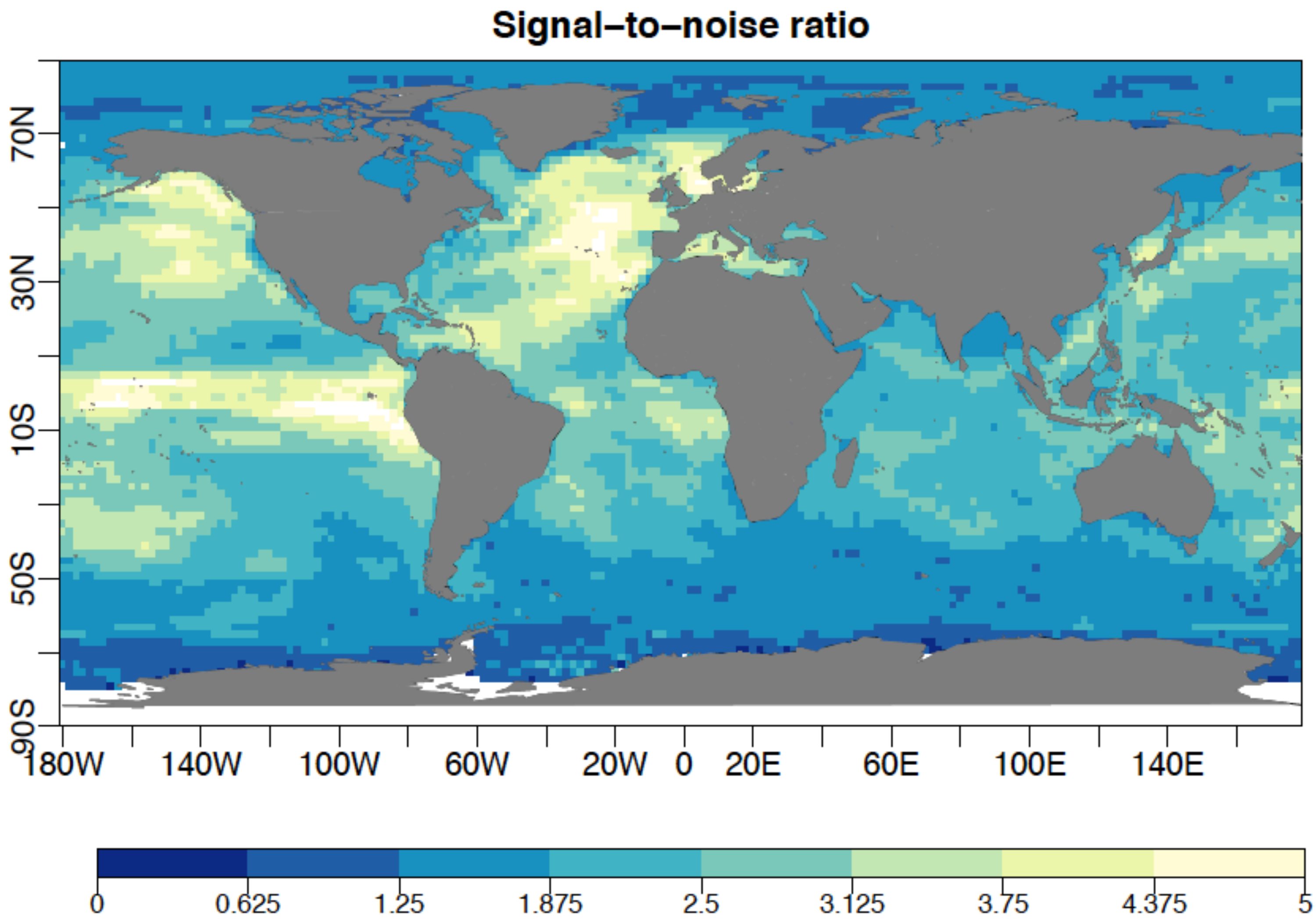
Sensitivity to sample





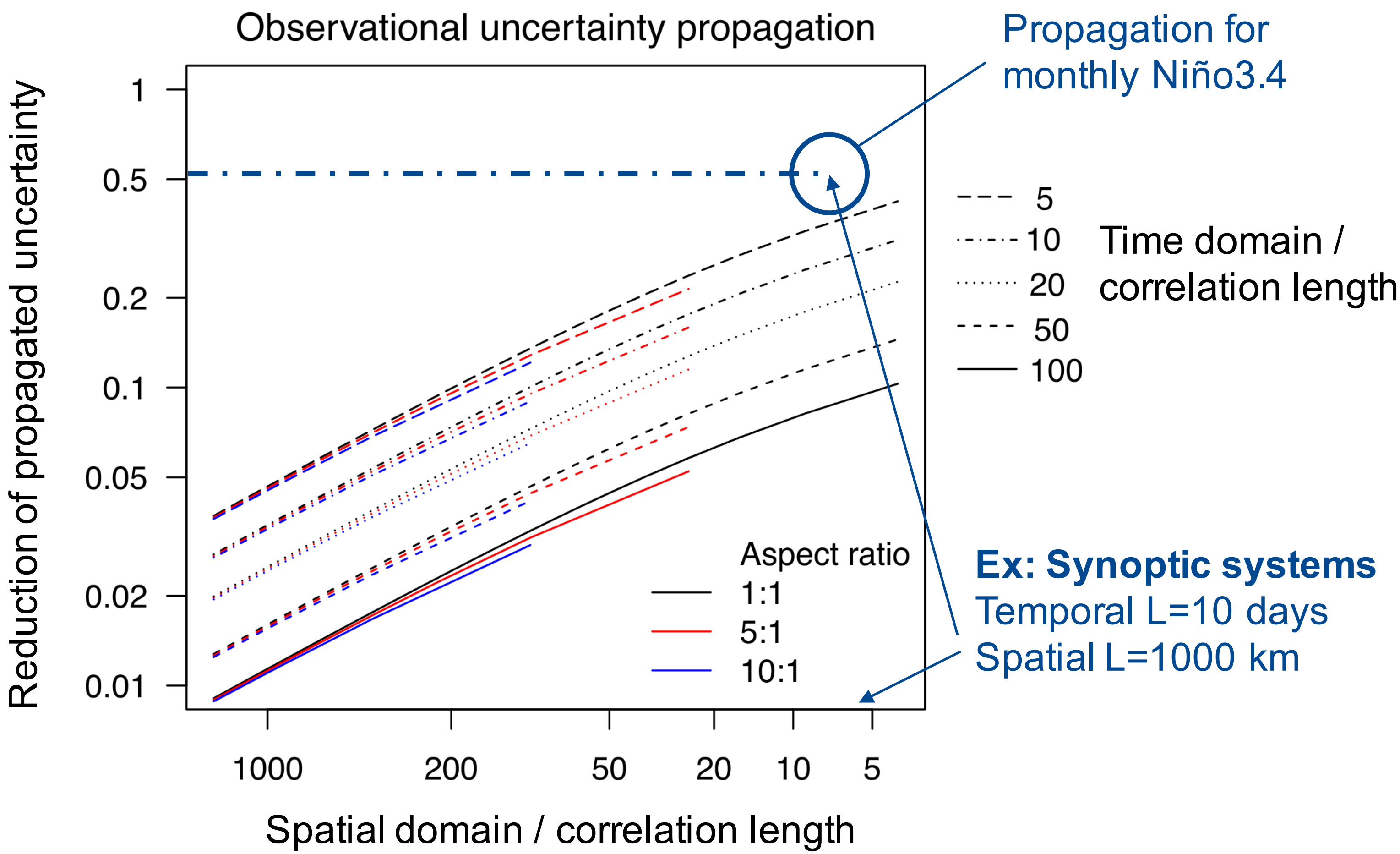
Signal-to-noise ratio SSTs

Signal (inter-annual variability) versus observational uncertainty (noise)



A “look-up” propagation figure

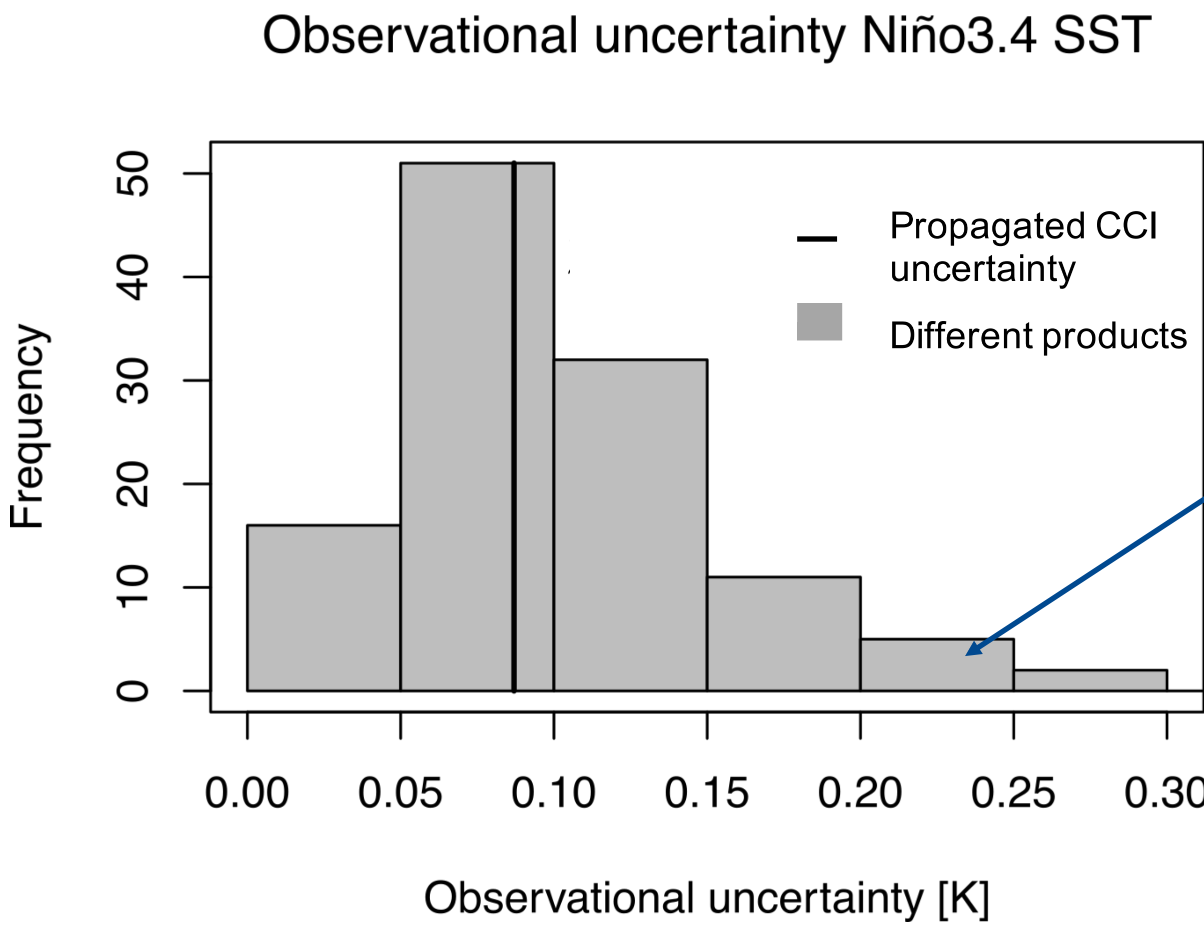
Use of error correlation scales: analytical solution that allows to look-up propagation factors



Bellprat et al. (2017) *RSE (in review)*

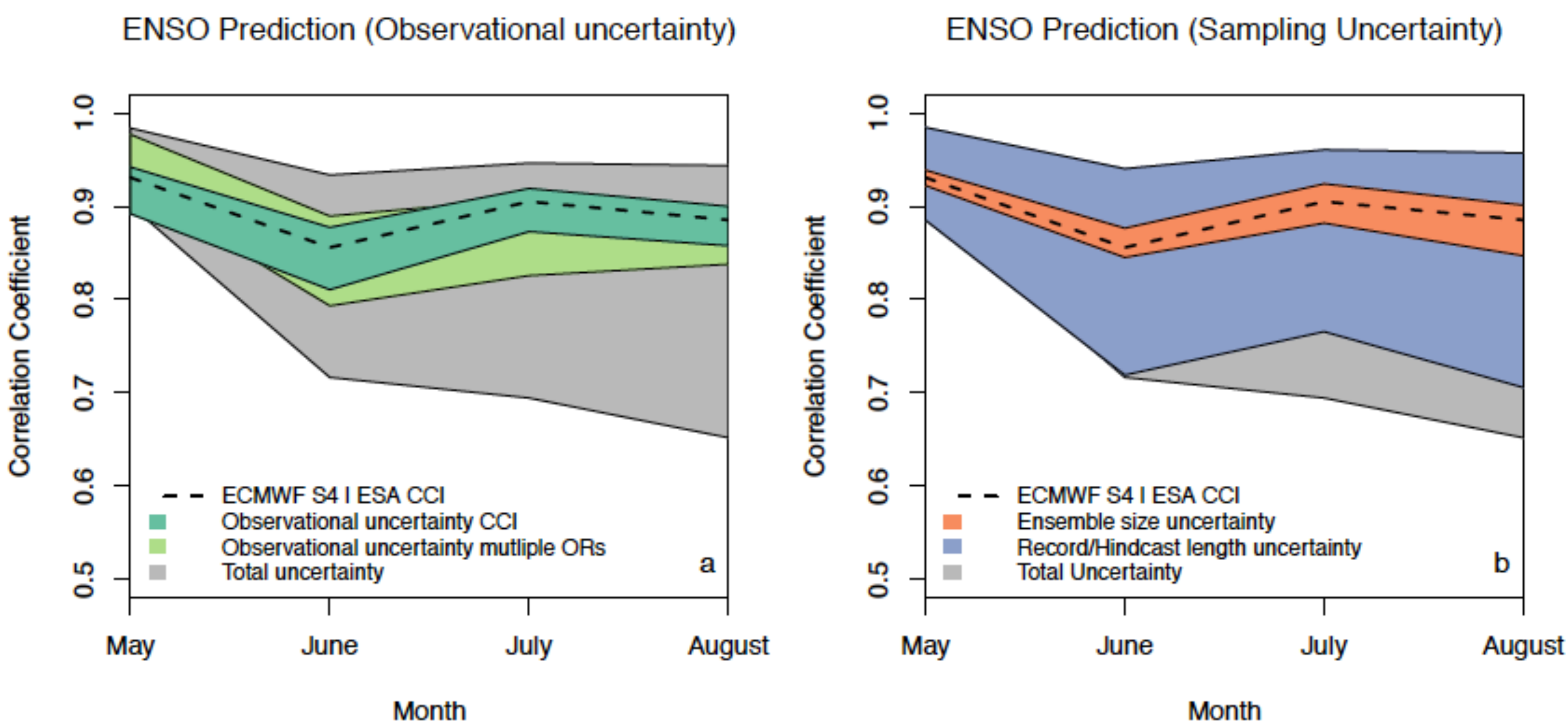
Comparison to multi-observations

Propagation assuming synoptic scales (1000 km, 10 days) of weather systems agrees well with deviations between existing products



Assumes all observations are equally “likely”

ESA CCI Uncertainty estimate



Relative contributions