

www.bsc.es

Polar Prediction Workshop

4-6th of May, 2016

Lamont-Doherty Earth Observatory –
Columbia University, NY



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

A common framework for evaluation of sea ice forecasts and verification products

F. Massonnet

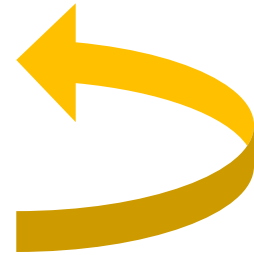
O. Bellprat, V. Guemas, F. J. Doblas-Reyes

Observational
reference



Sea ice
forecast

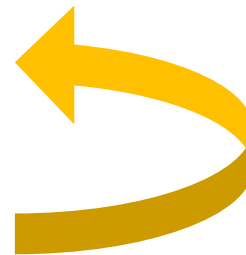
Observational
references



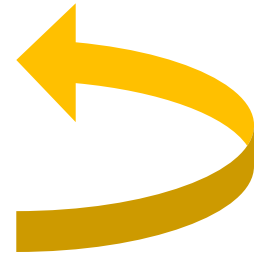
Part 1



Sea ice
forecasts

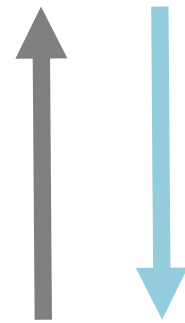


Observational
references

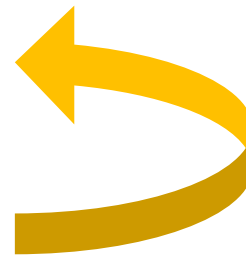


Part 1

Part 2



Sea ice
forecasts



PART 1

From
1-to-1 evaluation
to
N-to-M evaluation

Seasonal forecasts of summer Arctic sea ice extent

CanCM3/4, MPI-ESM-L/M/HR, CNRM, EC-EARTH (3 versions)

9 models



10 members



4-month forecasts initialized in May

360 forecasts (not independent from each other)

Period for evaluation: 1993-2008



ESA-CCI



OSI-SAF



HadISST



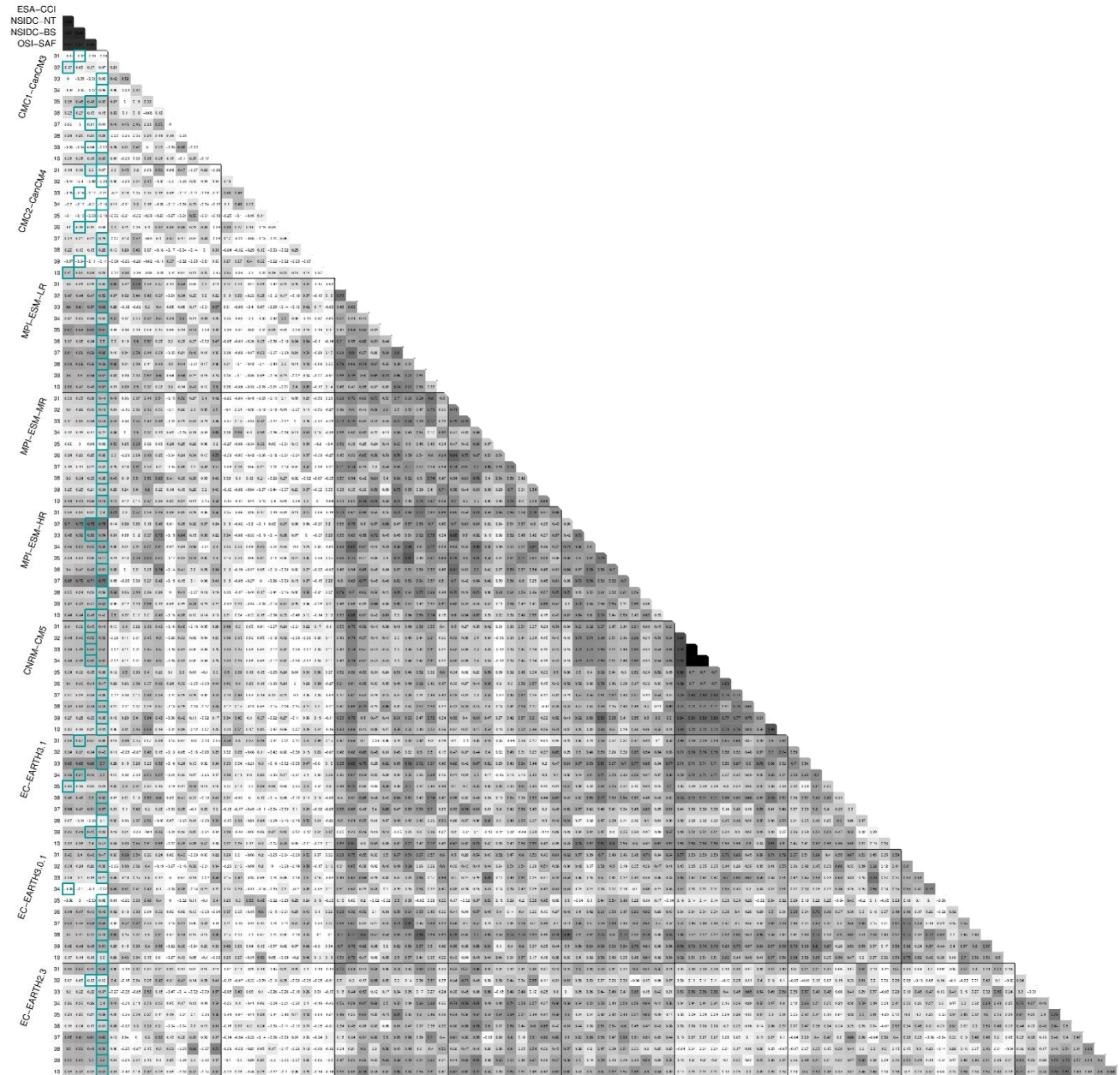
NSIDC



4 observational datasets for verification

(not independent from each other either)

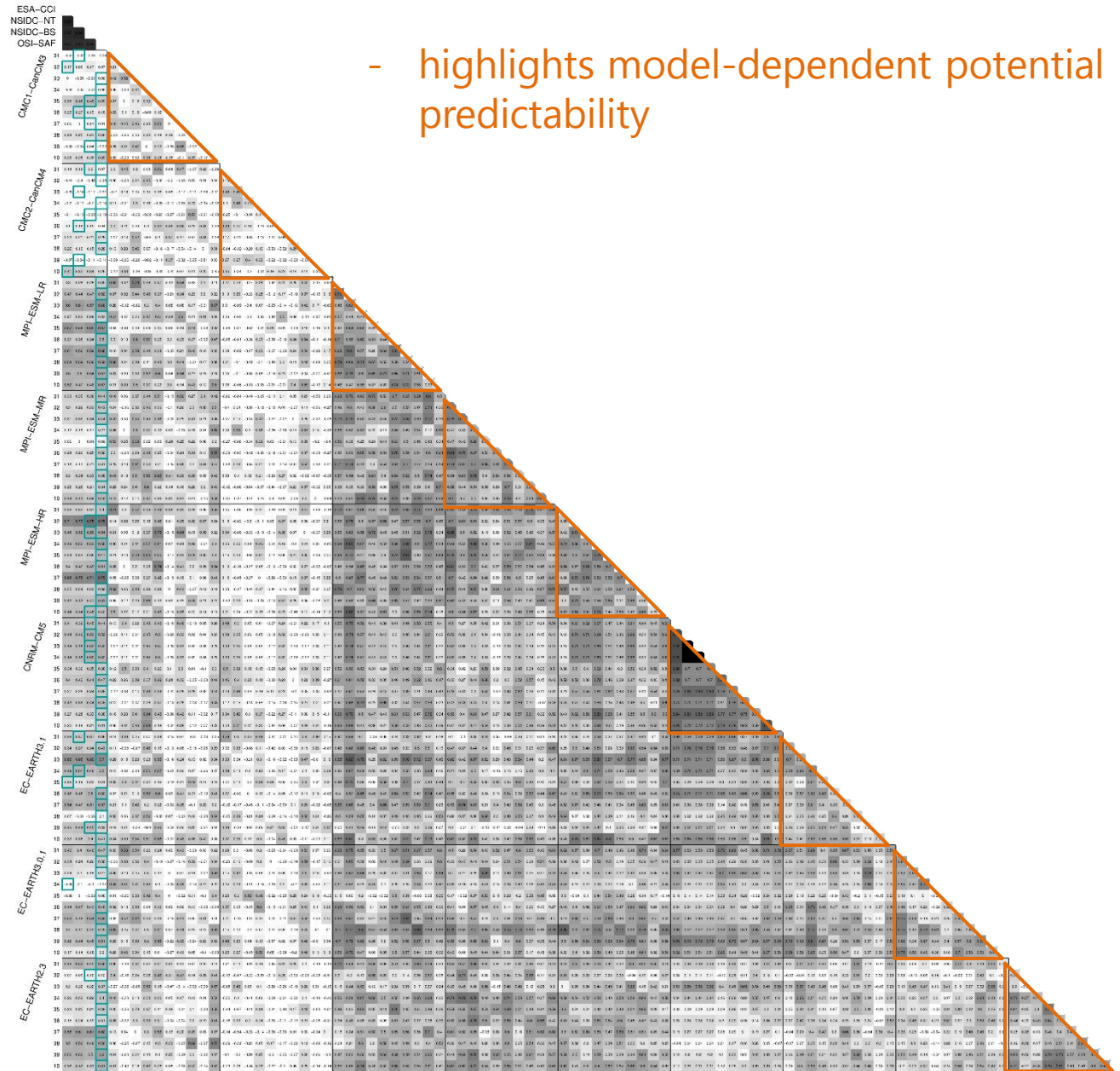
The joint obs-model correlation matrix



Correlation of
August sea ice
extent (1993-2008)

The joint obs-model correlation matrix

- highlights model-dependent potential predictability

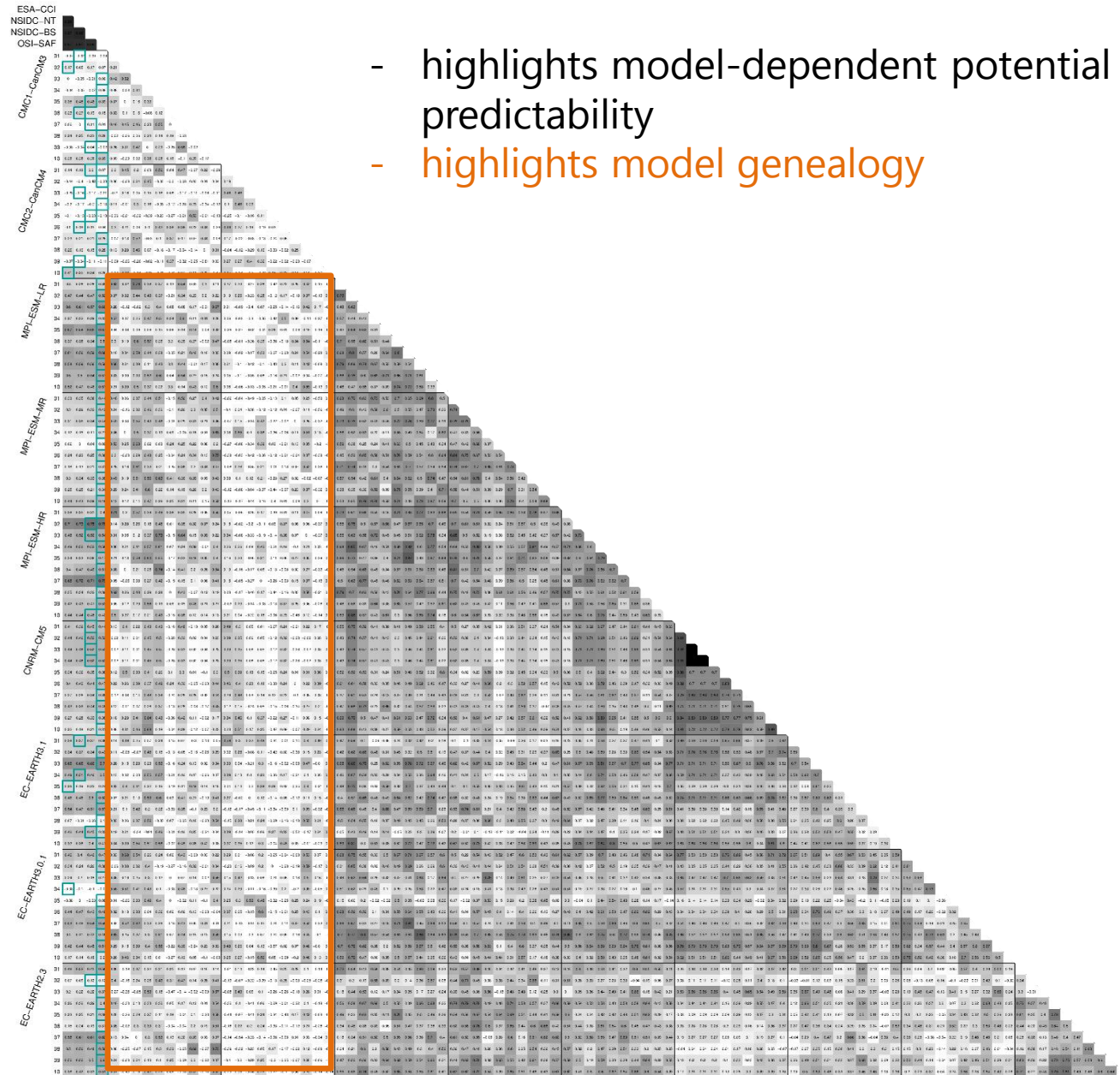


Correlation of
August sea ice
extent (1993-2008)

The joint obs-model correlation matrix

- highlights model-dependent potential predictability
- highlights model genealogy

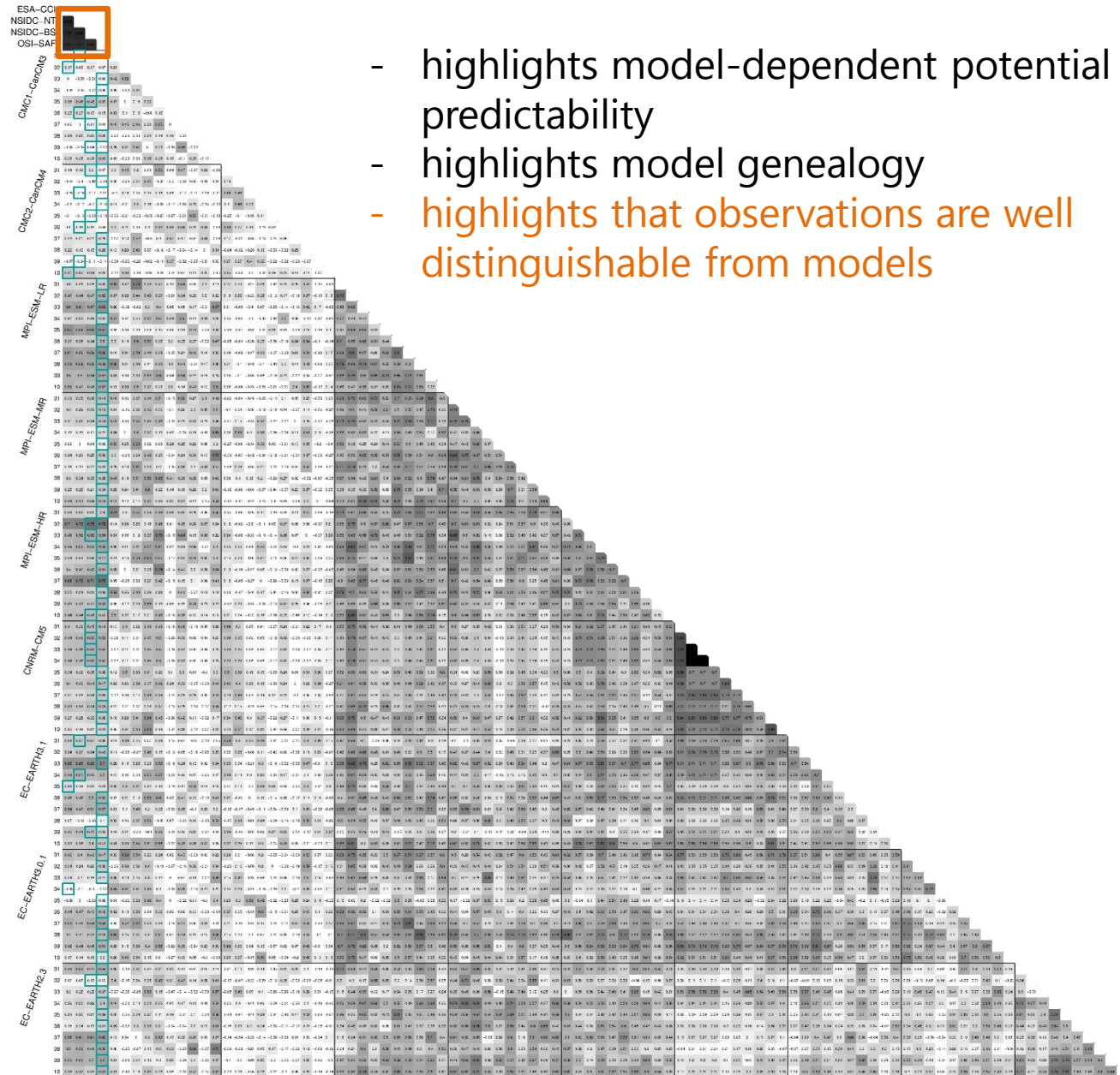
Correlation of
August sea ice
extent (1993-2008)



The joint obs-model correlation matrix

- highlights model-dependent potential predictability
- highlights model genealogy
- highlights that observations are well distinguishable from models

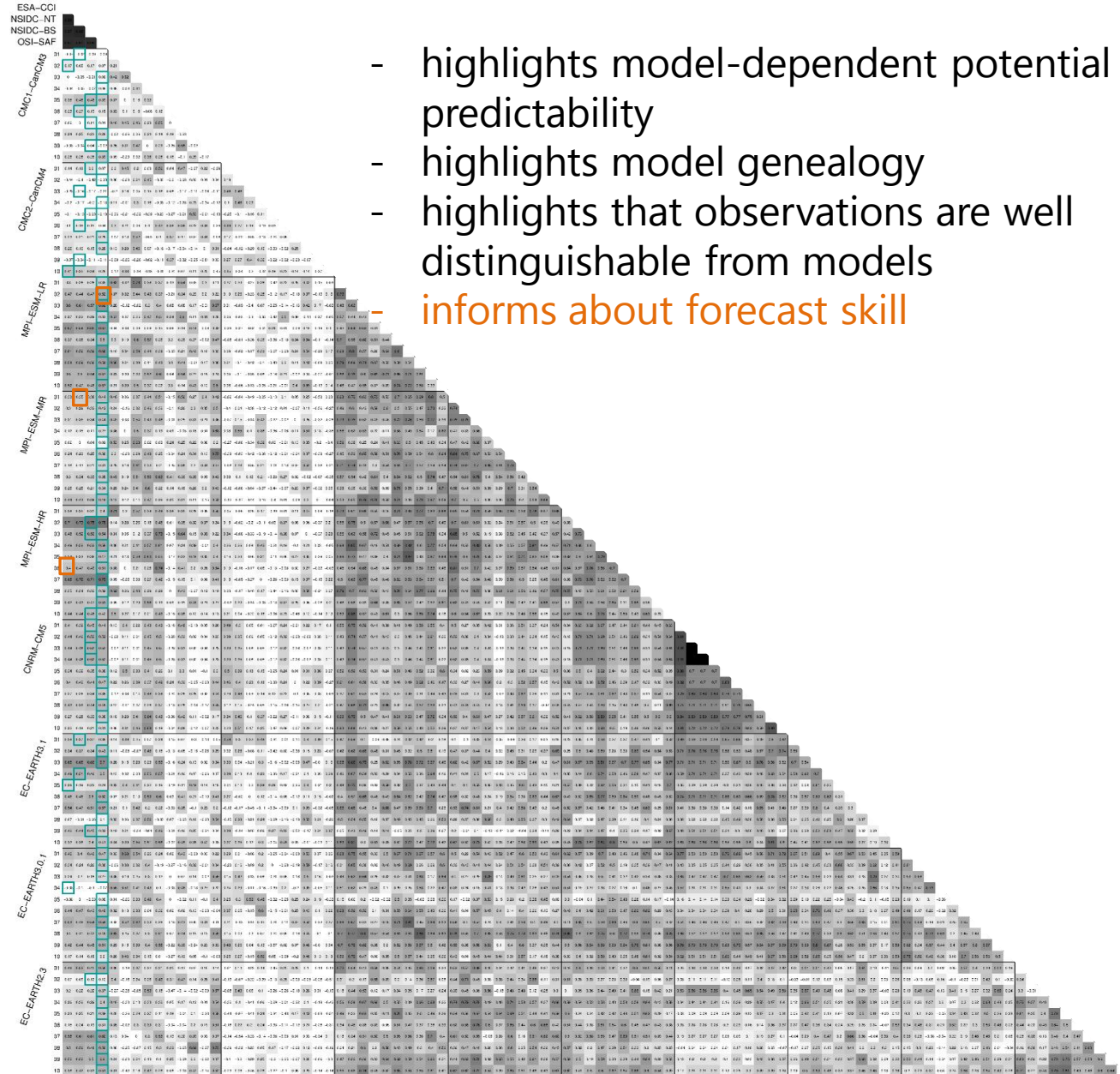
Correlation of August sea ice extent (1993-2008)



The joint obs-model correlation matrix

- highlights model-dependent potential predictability
- highlights model genealogy
- highlights that observations are well distinguishable from models
- informs about forecast skill

Correlation of
August sea ice
extent (1993-2008)



PART 1

From
1-to-1 evaluation
to
N-to-M evaluation

***Much can be learnt from the joint observation-model correlation matrix.
I'm contemplating the idea of applying that to the Sea Ice Outlook. But
for that we need individual members, not just statistics!***

PART 2

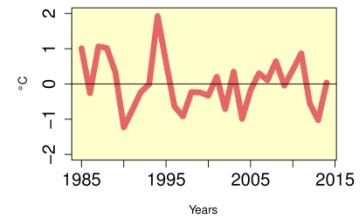
Better observations yield
better verification scores

A signal-plus-noise toy model

TRUTH $X_t = \epsilon$

**Interannual
variability**

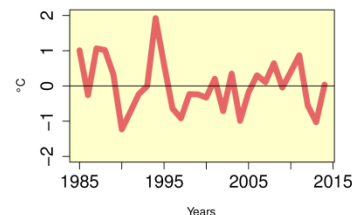
$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon)$$



A signal-plus-noise toy model

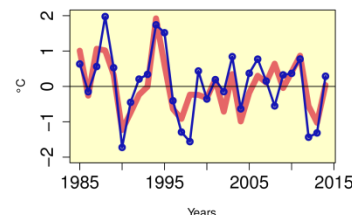
TRUTH $X_t = \boxed{\epsilon}$ **Interannual variability**

$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon)$$



OBS $X_o = \epsilon + \boxed{\eta_o}$ **Measurement and representativity error**

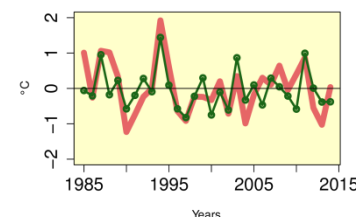
$$\eta_o \sim \mathcal{N}(0, \sigma_o)$$



MODEL $X_f = \alpha\epsilon + \boxed{\eta_f + \eta_m}$ **Model forecast error (physics, initial conditions, resolution) + irreducible error (atmosphere)**

$$\eta_f \sim \mathcal{N}(0, \sigma_f), \quad \eta_m \sim \mathcal{N}(0, \sigma_m)$$

All error terms are assumed to be uncorrelated



In this very simple paradigm, observational error is also a source of low skill

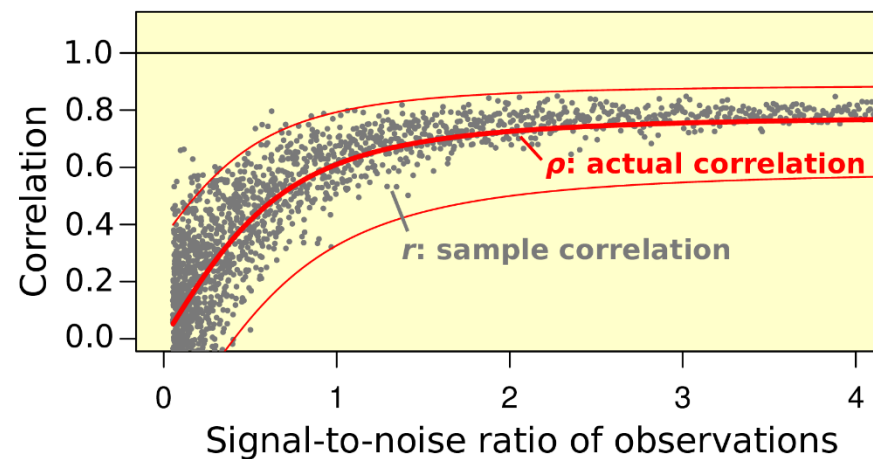
$$\rho(X_o, X_f) = \frac{1}{\sqrt{\left(1 + \frac{\sigma_o^2}{\sigma_\epsilon^2}\right) \cdot \left(1 + \frac{(\sigma_f^2 + \sigma_m^2)/\alpha^2}{\sigma_\epsilon^2}\right)}}$$

Correlation **increases** when

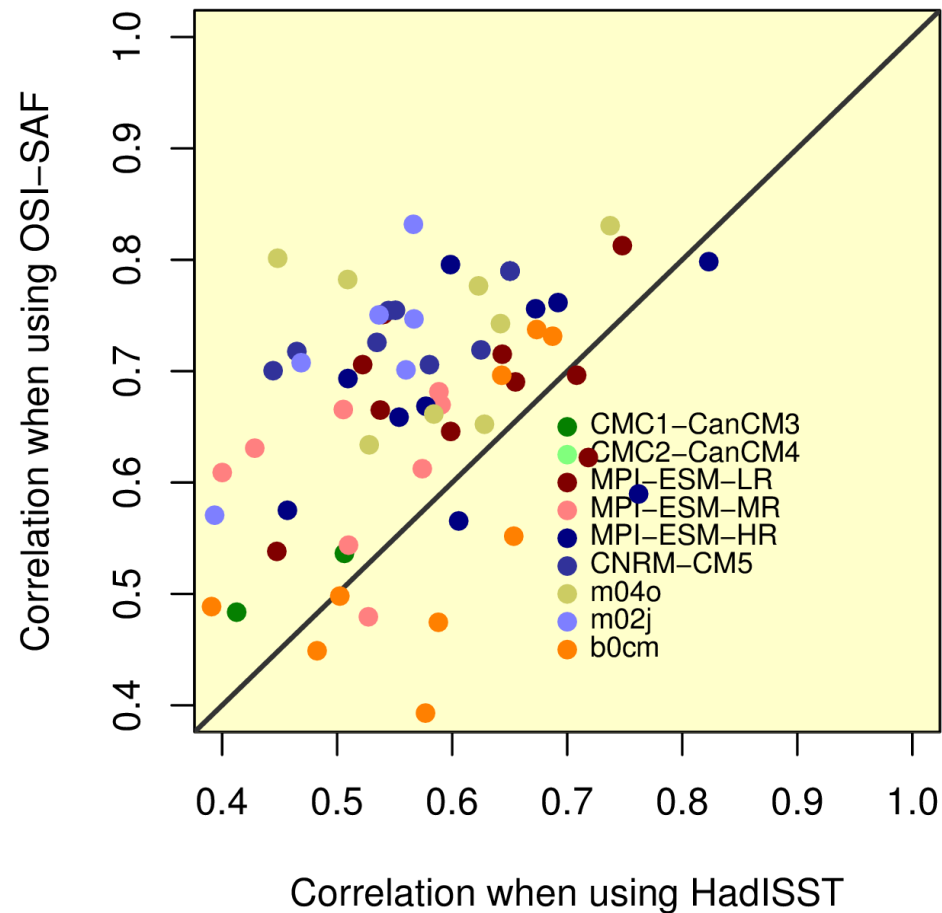
- Model explains more variability,
- Model error decreases,
- Climate signal is stronger,
- Observational error decreases.



If error statistics are known, the dependence can be predicted

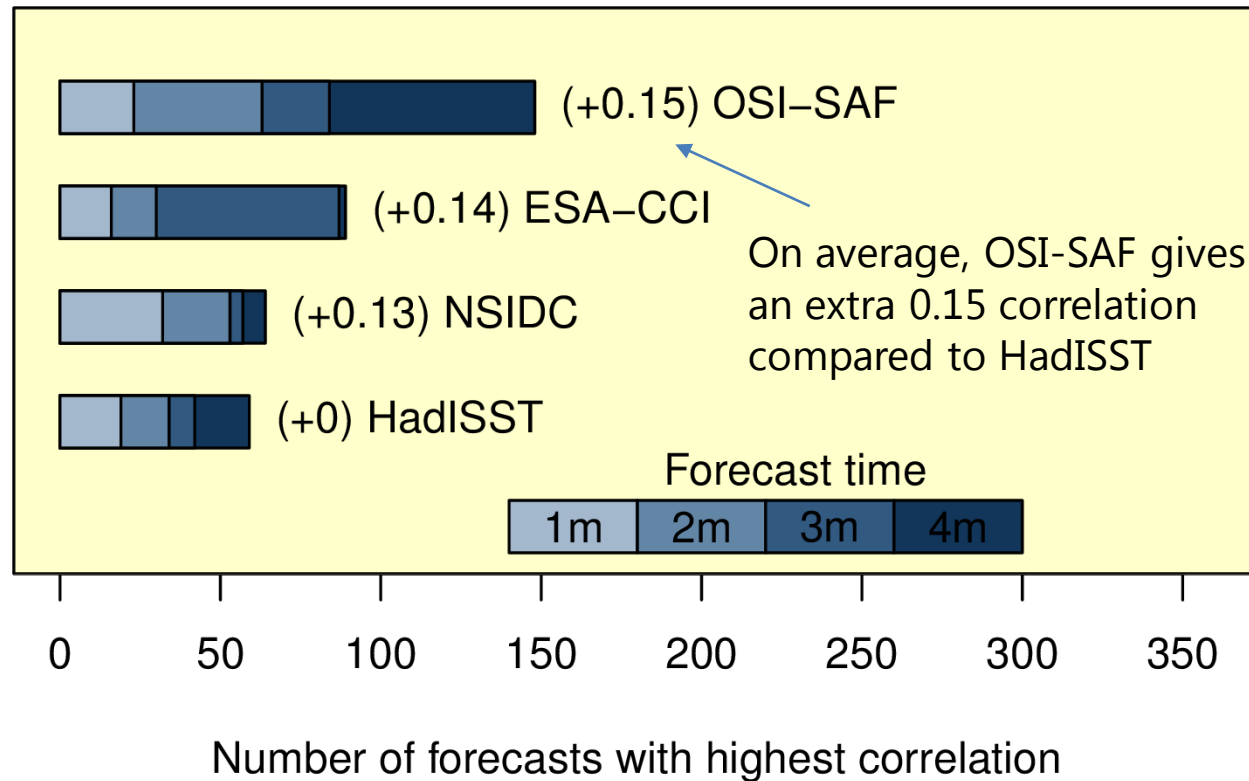


Correlations of summer Arctic sea ice extent from 90 forecasts



Systematic dependence of skill score on the choice of verification product

360 seasonal forecasts of summer Arctic sea ice extent



Such an extreme result is unlikely to have occurred by chance

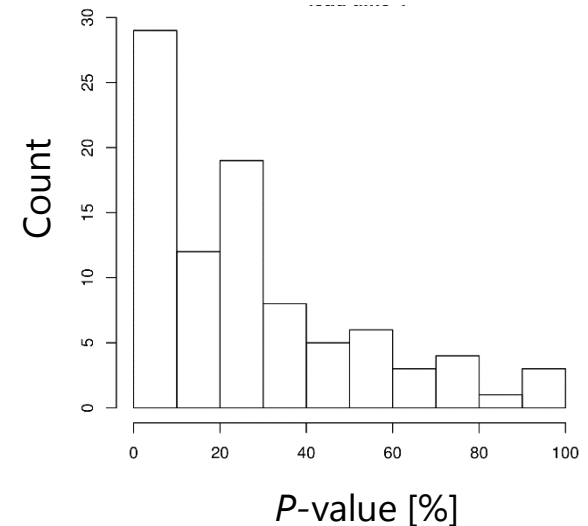
1. Bootstrapping

- Synthetic data is generated from the known sample covariance matrix of the data that we modify so that, for each forecast, correlations are set the same for all observations (our null hypothesis)
- With 10,000 trials, a result as extreme as the one we have happens **~0.2%** of the time

2. Parametric test

[Steiger et al., 1981]
(the test detects changes in correlation in presence of non-independent samples)

p-value of 110 Steiger tests to detect increase of correlation from HadISST (lowest) to OSI-SAF (highest)



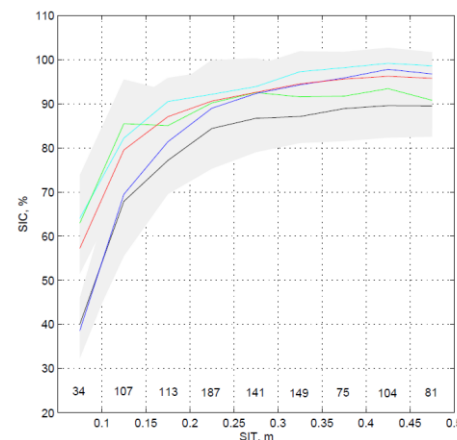
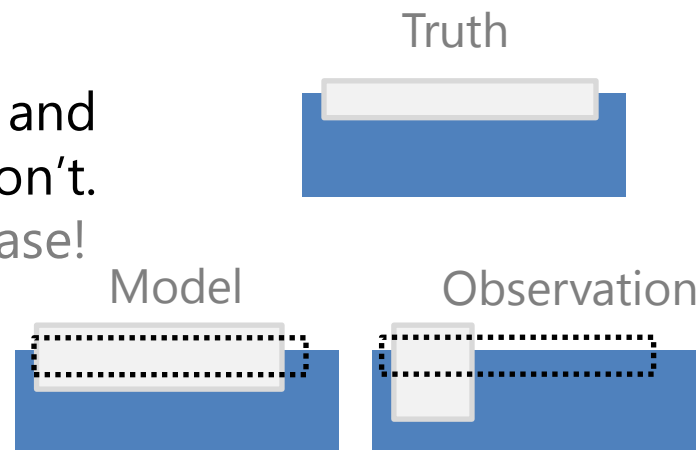
Why do models score better for the two most advanced and recent products?

Models simulate directly sea ice concentration and output it as a physical variable; observations don't. Models can be really good references in that case!

Observations have deficiencies that models don't have e.g. concentration of thin ice

According to the toy model results, ESA-CCI and OSI-SAF should have lower errors (but only these two provide errors)

Note: remarkably, the models are also the most independent w.r.t. OSI-SAF and ESA-CCI



PART 2

Better observations yield
better verification scores

Correlation increases whenever noise decreases (in the models and in the observations). Conventional skill scores can therefore be used to reveal the quality of observational datasets.

Conclusions & Outlooks

Summary | We expand the 1-1 forecast verification problem into an exercise of N-M verification. The joint correlation matrix is the diagnostic that encapsulates all relevant information we need.

Interpretation | These results are best understood if observations and models are considered at the same level (i.e., observations are not superior to models). Observational errors will systematically lower actual forecast skill, in the same way that model errors systematically lower forecast skill.

Recommendations | Models should not always be blamed for low performance. Observations can do this job easily! This overlooked source of error can give differences as large as differences from one model version to another. Modellers should be careful in picking their observational product, or at least use several of them.

Outlooks | Quantifying observational error propagation over time-averaged periods, space-averaged domains, is key to introduce observational uncertainty in current metrics of performance. Yet these error statistics depend on many unknowns, such as decorrelation time- and space-scales between grid-point, daily error statistics.

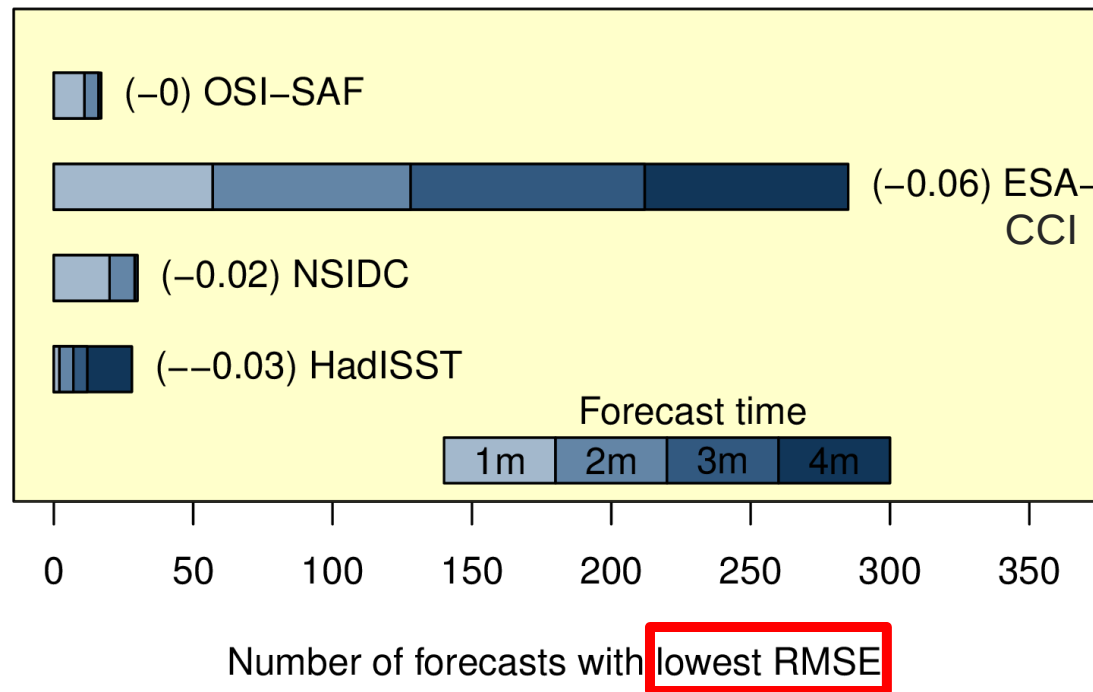
Thank you!

francois.massonnet@bsc.es



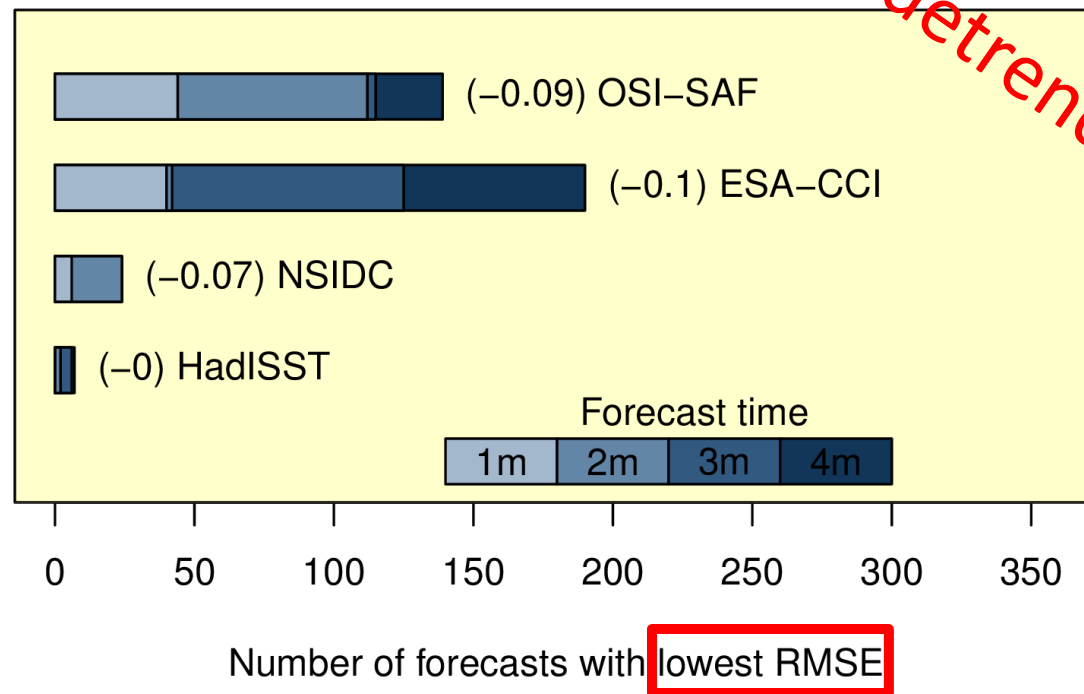
Systematic dependence of skill scores on verification product

360 seasonal forecasts of summer Arctic sea ice extent

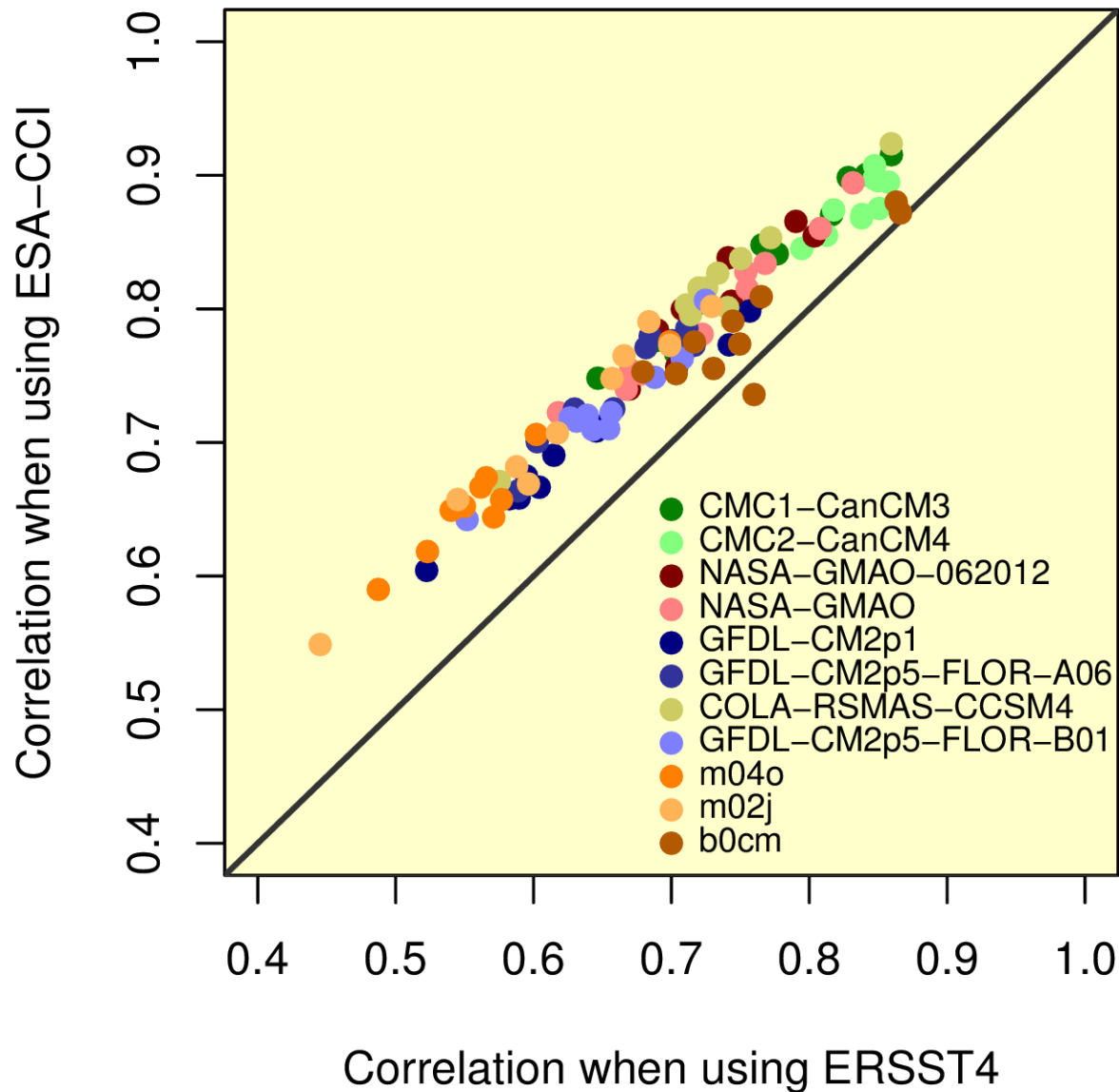


Systematic dependence of skill scores on verification product

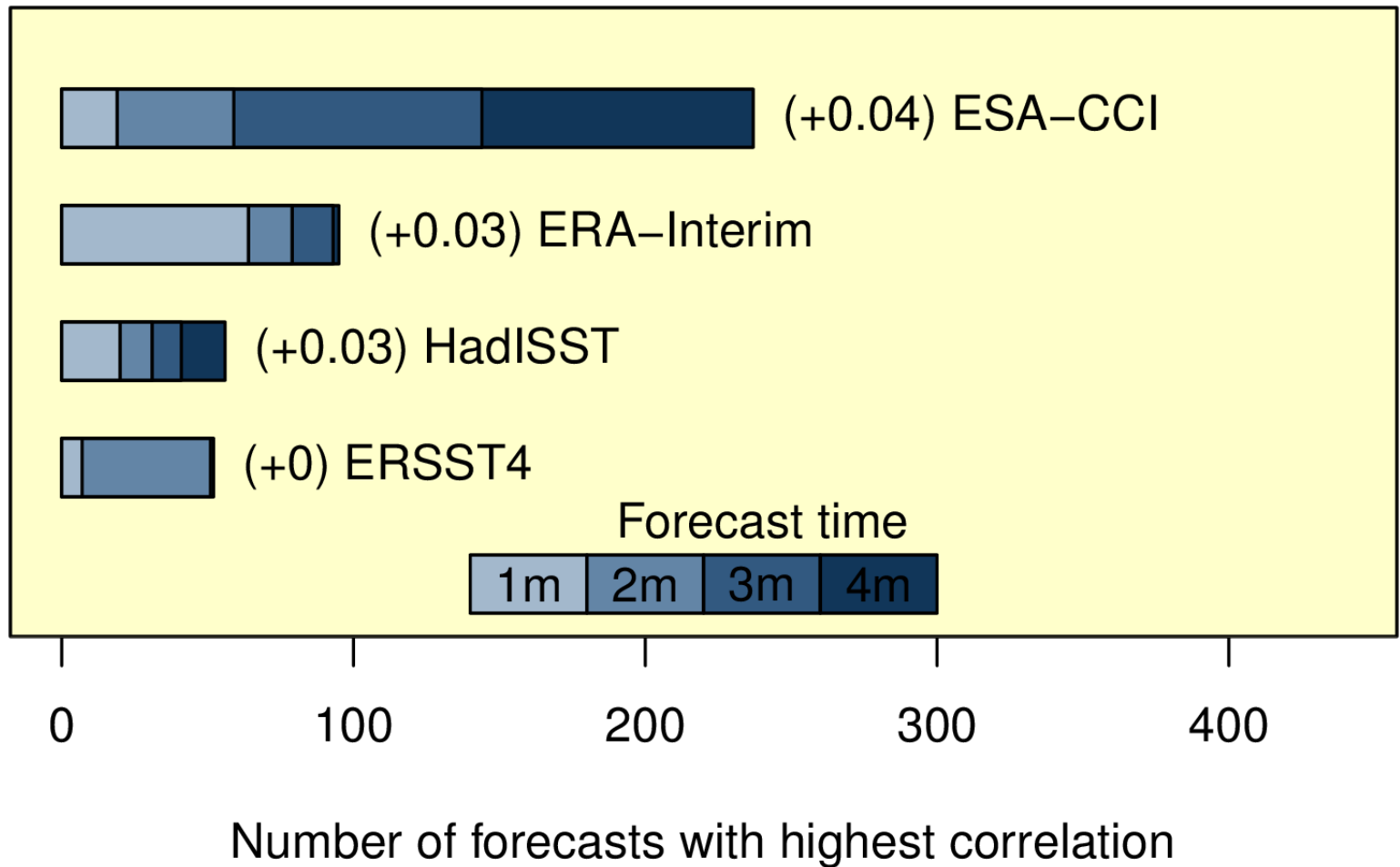
360 seasonal forecasts of summer Arctic sea ice extent



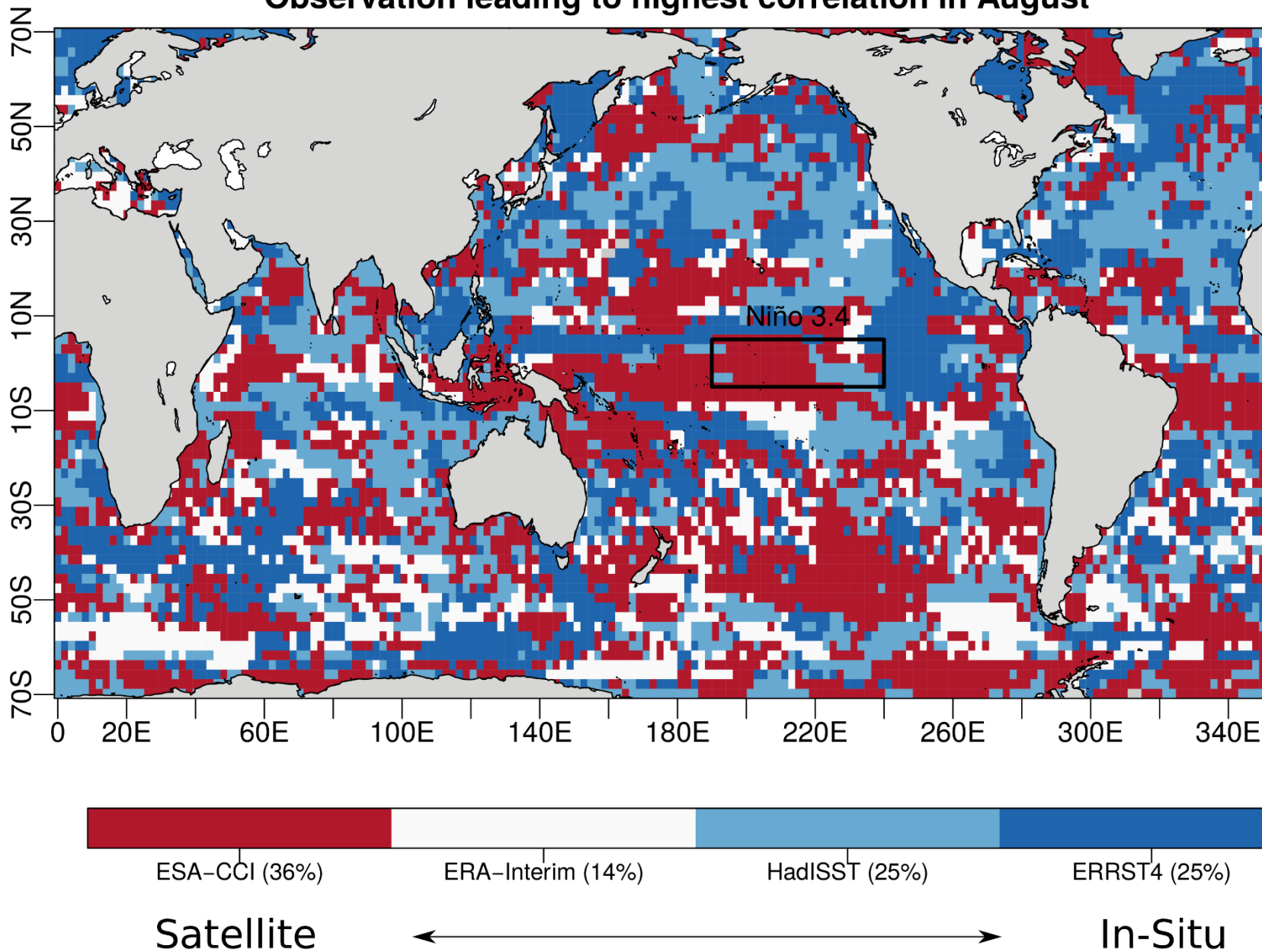
Correlations of Niño3.4 SST from 110 forecasts



440 seasonal forecasts of Niño3.4 SST

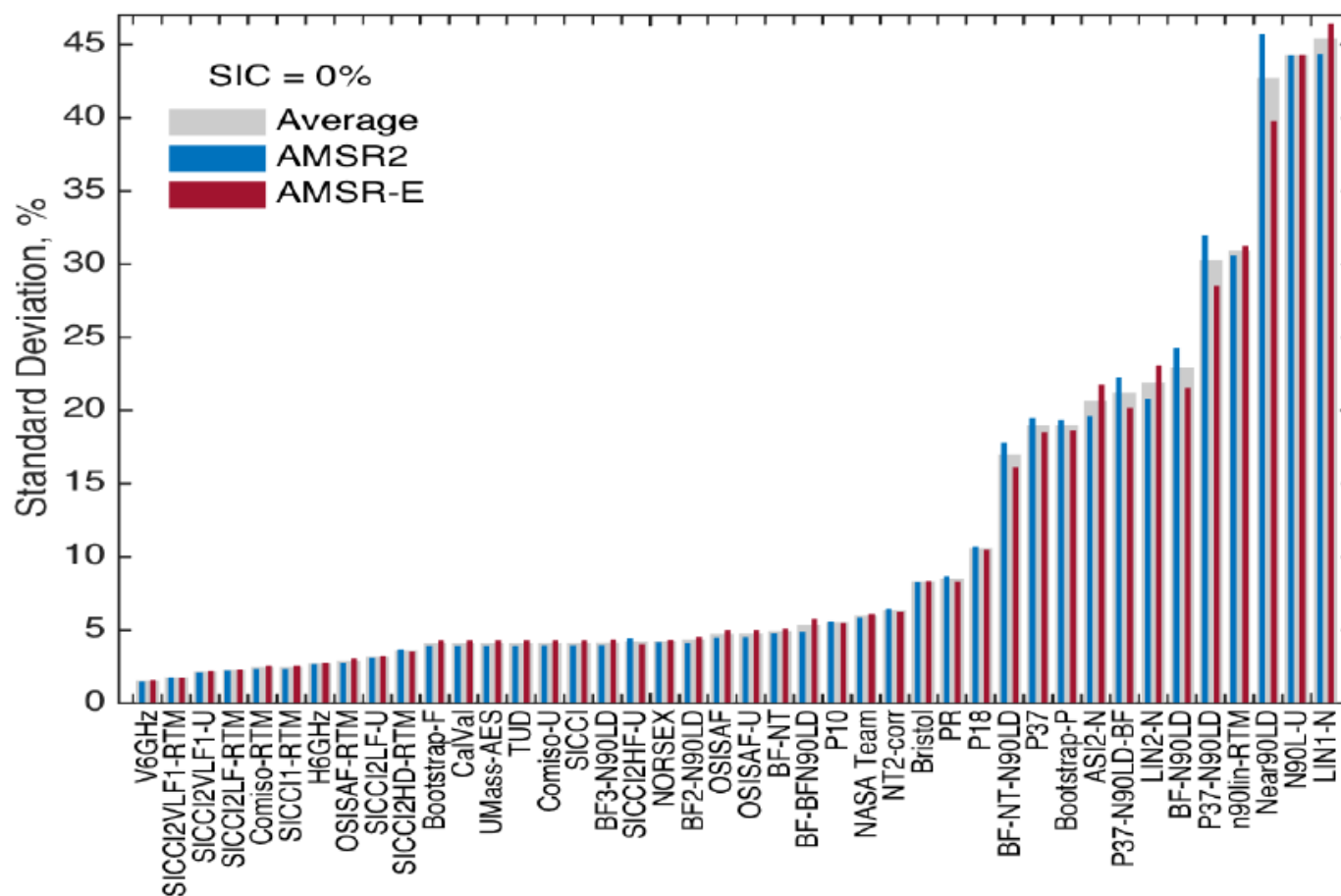


Observation leading to highest correlation in August



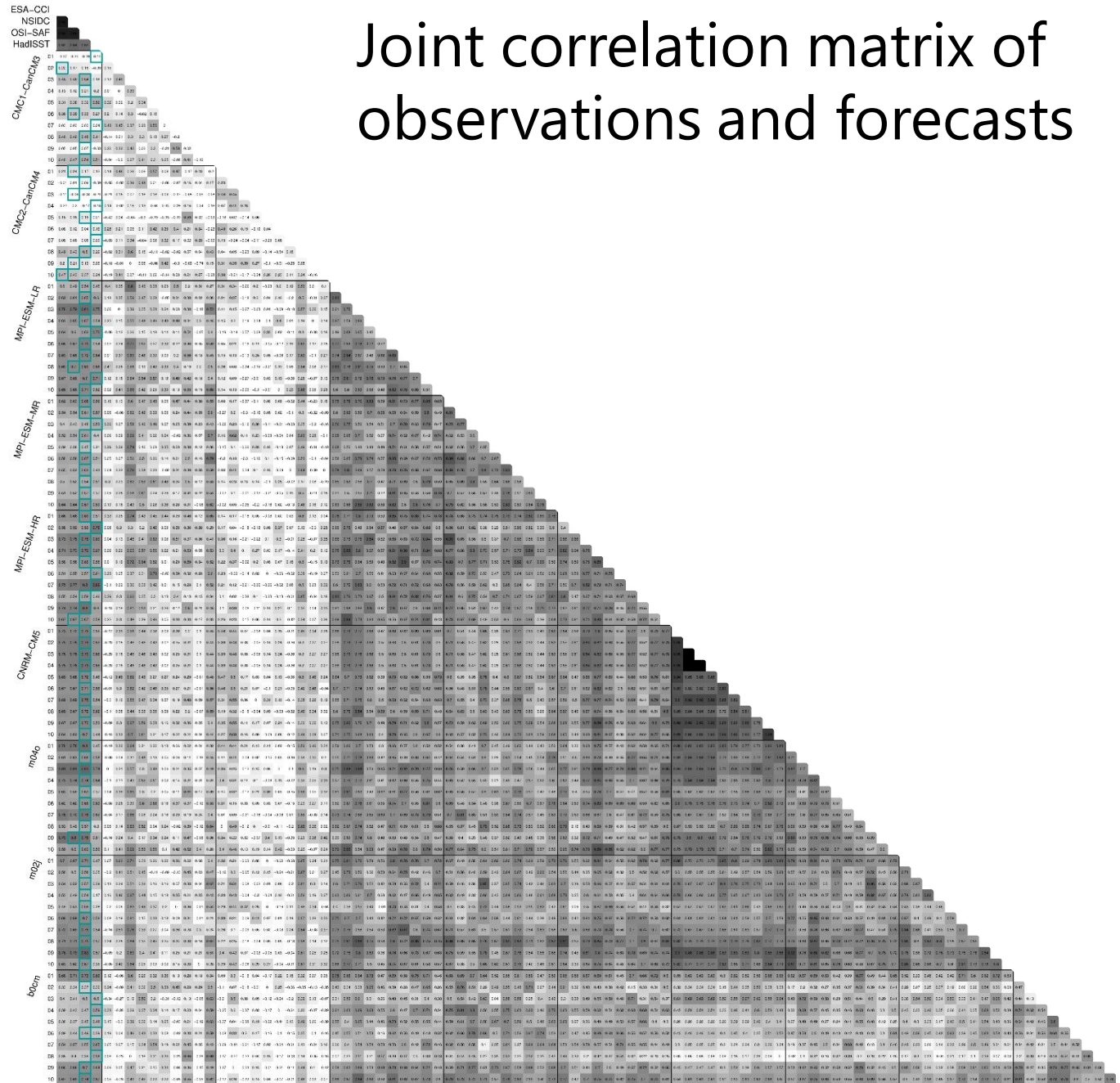
Sea ice concentration observational references	ESA-CCI	OSI-SAF	HadISST	NSIDC
Reference	Ivanova et al., 2015 (35)	Eastwood et al., 2015 (36)	Rayner et al., 2003 (33)	Fetterer and Knowles, 2004 (37)
Institution	European Space Agency	EUMETSAT	MetOffice (UK)	National Snow and Ice Data Center
Webpage	http://esa-cci.nersc.no/	http://osisaf.met.no/p/ice/ice_conc_reprocessed.html	http://www.metoffice.gov.uk/hadobs/hadisst/	https://nsidc.org/data/docs/noaa/g02135_seaice_index/
Period of product availability	1993-2008	1978-2015	1871-present	1978-present
Grid resolution	25 km x 25 km	10 km x 10 km	1.0 x1.0 (~110 km x 110 km near equator)	25 km x 25 km
Grid type	Equal-Area Scalable Earth	Polar Stereographic	Regular	Polar Stereographic
Primary product and technology on which the analysis is based	Passive microwave satellite data: SSM/I (1992-2008)	Passive microwave satellite data: SMMR and SSM/I-SSMIS.	Passive microwave satellite data: SMMR and SSM/I-SSMIS (1978-1996) + NCEP operational dataset (1997-present)	Passive microwave satellite data: SMMR and SSM/I-SSMIS.
Algorithm of processing	Same as OSI-SAF algorithm, but improved so that better performance is achieved over thin ice. Atmospheric filter was applied to brightness temperature directly, and not on the spectral gradient ratio as this latter approach was found to eliminate low ice concentrations.	Hybrid: Bootstrap and Bristol. Dynamical tie-points are used (calibration parameters are time-dependent). Weather filter was improved from NASA Team algorithm, as it was found that this correction tended to eliminate low ice concentrations.	NASA Team (see NSIDC column)	NASA Team: Static tie-points are used, but different datasets for Northern and Southern Hemispheres. Weather filter was used: SIC set to zero when the spectral gradient ratio (GR) is > 0.07
Other comments	Only ESA-CCI phase 1 products are considered. Phase 2 products will be developed jointly with EUMETSAT OSI-SAF.	Version 1.2 of the product was used, i.e. without input data from ESA. The next release of OSI-SAF reprocessed sea ice concentration will include results from the ESA-CCI research project	The HadISST product is updated with NCEP operational analyses from 1997 onwards.	We take as monthly sea ice extent the value already processed by NSIDC.

ICE CONCENTRATION ALGORITHM PERFORMANCE (SIC=0)



Courtesy L. T. Pedersen,
N. Ivanova and co-authors

Joint correlation matrix of observations and forecasts



Original covariance matrix

	OBS 1	OBS 2	FORECAST 1	FORECAST 2	FORECAST 3
OBS 1	σ_{o1}^2				
OBS 2	$r_{o1,o2} \cdot s_{o1} \cdot s_{o2}$	σ_{o2}^2			
FORECAST 1	$r_{o1,f1} \cdot \sigma_{o1} \cdot \sigma_{f1}$	$r_{o2,f1} \cdot \sigma_{o2} \cdot \sigma_{f1}$	σ_{f1}^2		
FORECAST 2	$r_{o1,f2} \cdot \sigma_{o1} \cdot \sigma_{f2}$	$r_{o2,f2} \cdot \sigma_{o2} \cdot \sigma_{f2}$	$r_{f1,f2} \cdot \sigma_{f1} \cdot \sigma_{f2}$	σ_{f2}^2	
FORECAST 3	$r_{o1,f3} \cdot \sigma_{o1} \cdot \sigma_{f3}$	$r_{o2,f3} \cdot \sigma_{o2} \cdot \sigma_{f3}$	$r_{f1,f3} \cdot \sigma_{f1} \cdot \sigma_{f3}$	$r_{f2,f3} \cdot \sigma_{f2} \cdot \sigma_{f3}$	σ_{f3}^2

Modified covariance matrix

	OBS 1	OBS 2	FORECAST 1	FORECAST 2	FORECAST 3
OBS 1	σ_{o1}^2				
OBS 2	$r_{o1,o2} \cdot \sigma_{o1} \cdot \sigma_{o2}$	σ_{o2}^2			
FORECAST 1	$r_1 \cdot \sigma_{o1} \cdot \sigma_{f1}$	$r_1 \cdot \sigma_{o2} \cdot \sigma_{f1}$	σ_{f1}^2		
FORECAST 2	$r_2 \cdot \sigma_{o1} \cdot \sigma_{f2}$	$r_2 \cdot \sigma_{o2} \cdot \sigma_{f2}$	$r_{f1,f2} \cdot \sigma_{f1} \cdot \sigma_{f2}$	σ_{f2}^2	
FORECAST 3	$r_3 \cdot \sigma_{o1} \cdot \sigma_{f3}$	$r_3 \cdot \sigma_{o2} \cdot \sigma_{f3}$	$r_{f1,f3} \cdot \sigma_{f1} \cdot \sigma_{f3}$	$r_{f2,f3} \cdot \sigma_{f2} \cdot \sigma_{f3}$	σ_{f3}^2