Sea ice in the Earth system: a multidisciplinary perspective
Brest, 4-6 June 2019

Challenges in the evaluation of large-scale sea ice models

François Massonnet

francois.massonnet@uclouvain.be

www.climate.be/u/fmasson

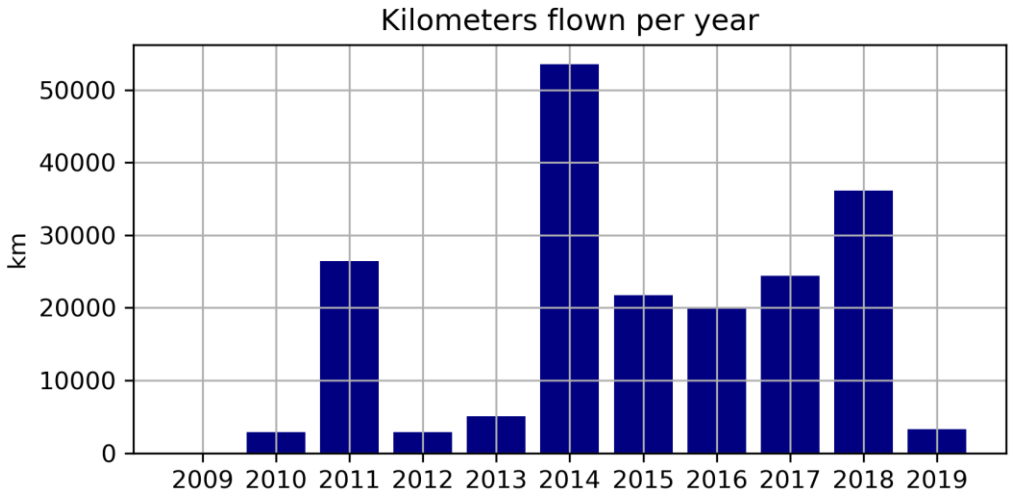
@FMassonnet

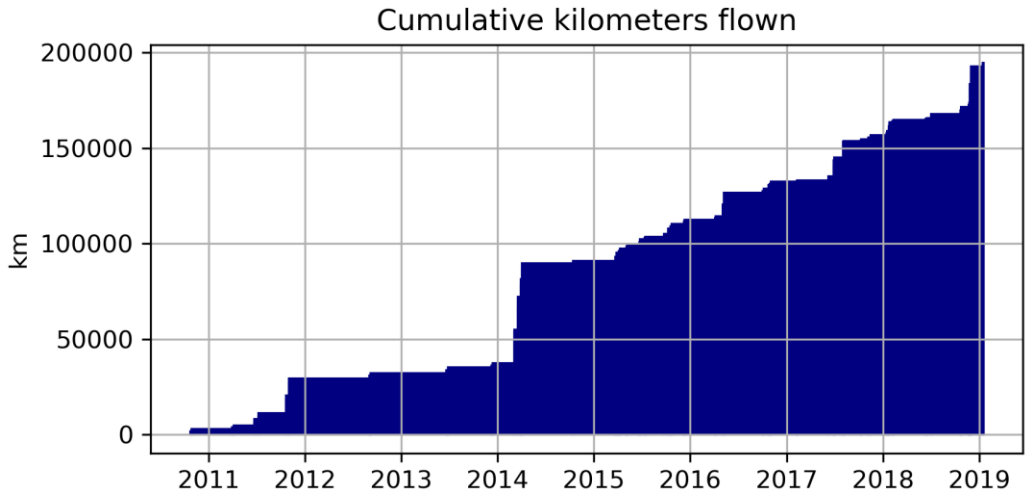
UCLouvain

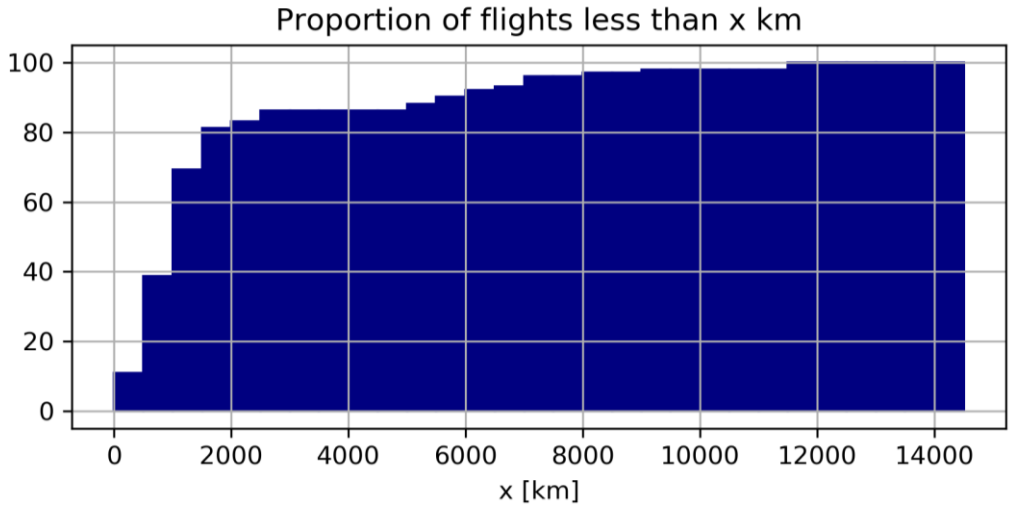
EARTH & LIFE INSTITUTE

fnrs LA LIBERTÉ DE CHERCHER

Flying for research:
The dilemma of climate scientists







Sea ice in the Earth system: a multidisciplinary perspective
Brest, 4-6 June 2019

Challenges in the evaluation of large-scale sea ice models

François Massonnet

francois.massonnet@uclouvain.be

www.climate.be/u/fmasson

@FMassonnet

UCLouvain

EARTH & LIFE INSTITUTE

fnrs LA LIBERTÉ DE CHERCHER

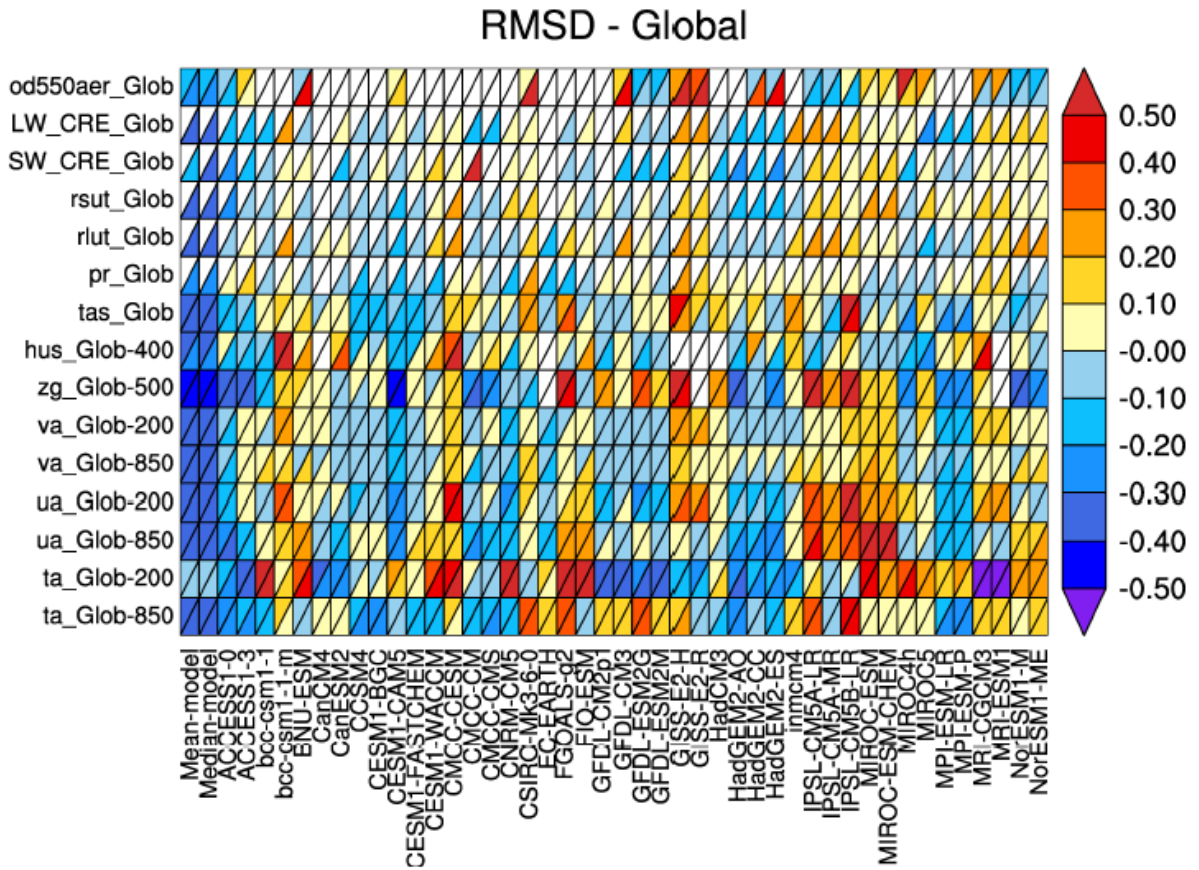
ESMValTool – the right way forward to evaluate our climate models?

Geosci. Model Dev., 9, 1747–1802, 2016
www.geosci-model-dev.net/9/1747/2016/
doi:10.5194/gmd-9-1747-2016

Geoscientific Model Development
Open Access
EGU

ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP

Veronika Eyring[1], Mattia Righi[1], Axel Lauer[1], Martin Evaldsson[2], Sabrina Wenzel[1], Colin Jones[3,4], Alessandro Anav[5], Oliver Andrews[6], Irene Cionni[7], Edouard L. Davin[8], Clara Deser[9], Carsten Ehbrecht[10], Pierre Friedlingstein[5], Peter Gleckler[11], Klaus-Dirk Gottschaldt[1], Stefan Hagemann[12], Martin Juckes[13], Stephan Kindermann[10], John Krasting[14], Dominik Kunert[1], Richard Levine[4], Alexander Loew[15,12], Jarmo Mäkelä[16], Gill Martin[4], Erik Mason[14,17], Adam S. Phillips[9], Simon Read[18], Catherine Rio[19], Romain Roehrig[20], Daniel Senftleben[1], Andreas Sterl[21], Lambertus H. van Ulft[21], Jeremy Walton[4], Shiyu Wang[2], and Keith D. Williams[4]



Eyring et al., Geosci. Mod. Dev., 2016

Today's questions

What are the purposes of
evaluating sea ice models?

What should we pay attention to,
when designing an evaluation protocol?

Is sea ice model evaluation settled?
What are the ways forward?

APPLICATE.eu
Advanced prediction in polar regions and beyond

PRIMAVERA

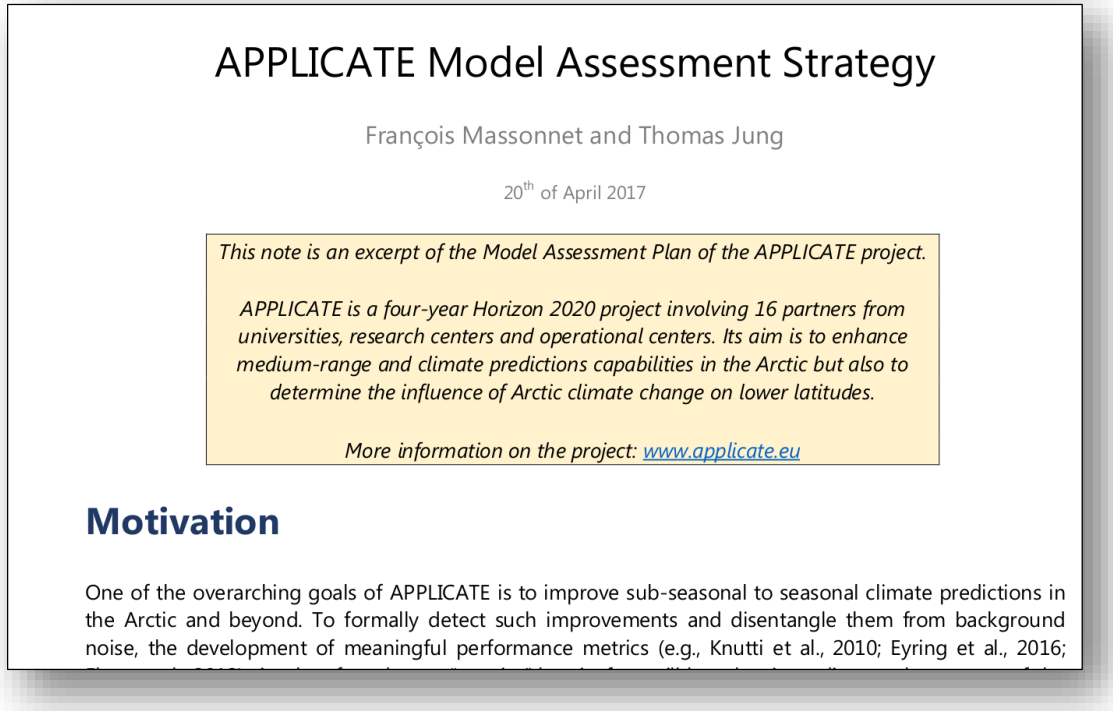
Copernicus
Europe's eyes on Earth

MERCATOR OCEAN INTERNATIONAL

COPERNICUS MARINE ENVIRONMENT MONITORING SERVICE (CMEMS)

EVOLUTION AND OPTIMISATION OF THE NEMO CODE USED FOR THE MFC-GLO IN CMEMS

Sea Ice modelling Integrated Initiative



APPLICATE Model Assessment Strategy

François Massonnet and Thomas Jung

20th of April 2017

This note is an excerpt of the Model Assessment Plan of the APPLICATE project.

APPLICATE is a four-year Horizon 2020 project involving 16 partners from universities, research centers and operational centers. Its aim is to enhance medium-range and climate predictions capabilities in the Arctic but also to determine the influence of Arctic climate change on lower latitudes.

More information on the project: www.applicate.eu

Motivation

One of the overarching goals of APPLICATE is to improve sub-seasonal to seasonal climate predictions in the Arctic and beyond. To formally detect such improvements and disentangle them from background noise, the development of meaningful performance metrics (e.g., Knutti et al., 2010; Eyring et al., 2016;

http://applicate.eu/images/APPLICATE_metrics_final.pdf

Today's questions

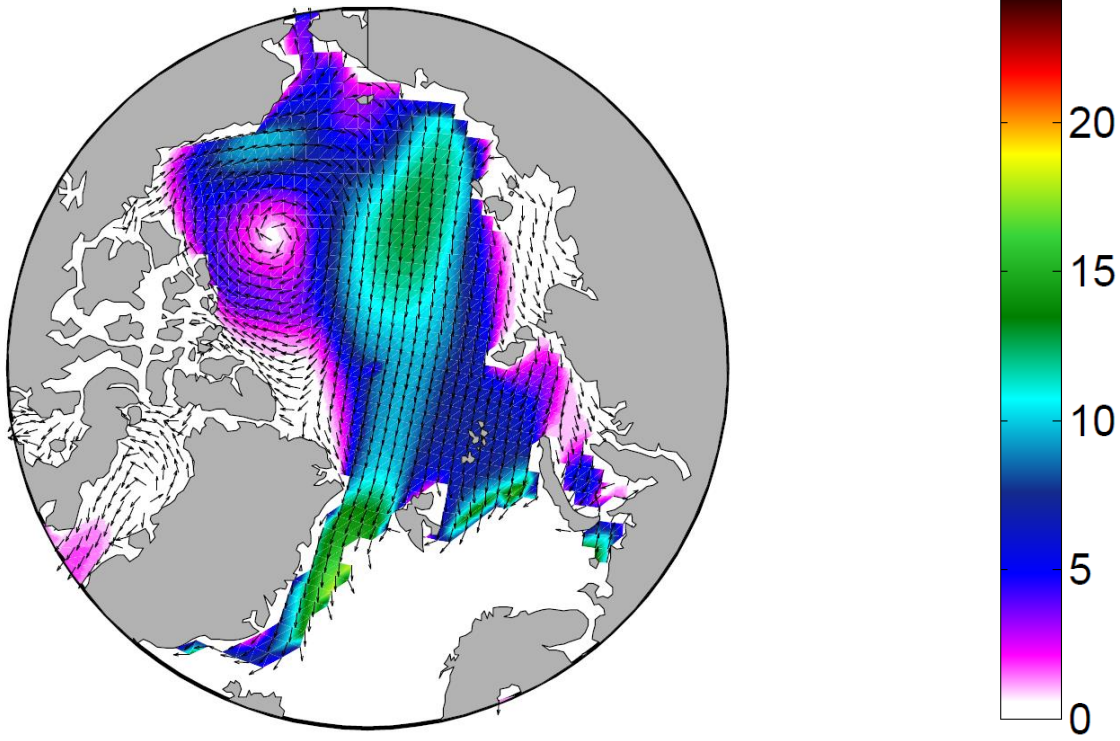
What are the purposes of evaluating sea ice models?

What should we pay attention to, when designing an evaluation protocol?

Is sea ice model evaluation settled?
What are the ways forward?

Diagnostics *vs.* metrics

12-14 April 2012 mean sea ice velocity from **model** (NEMO+LIM3)+atmospheric forcing



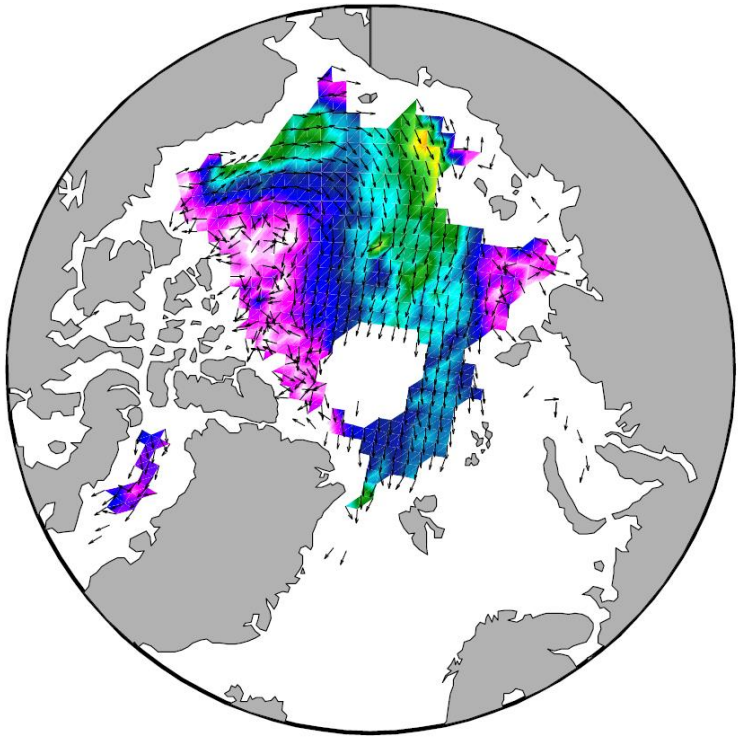
« Diagnostic »

$$\mathbf{u}_{mod}(x,y)$$

low-dimensional object that summarizes a high-dimensional geophysical dataset

Diagnostics *vs.* metrics

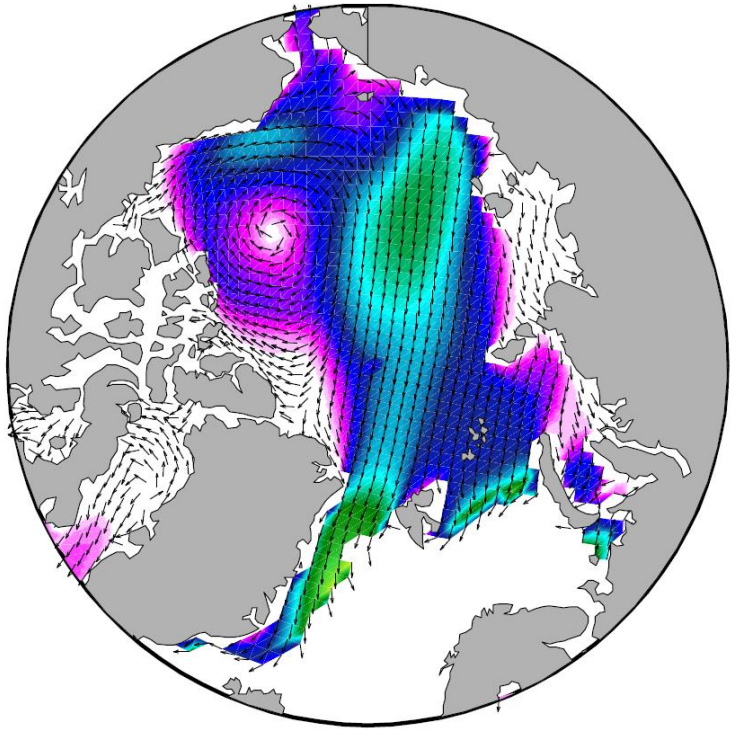
12-14 April 2012 mean sea ice velocity from **observations** (Lavergne et al., J. Geophys. Res., 2010)

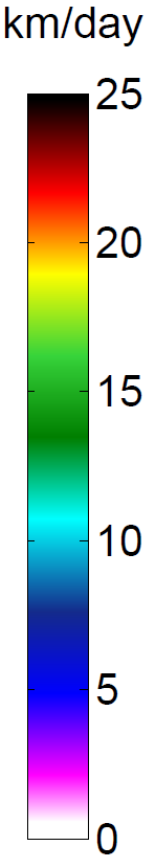


« Diagnostic »

$\mathbf{u}_{obs}(x,y)$

12-14 April 2012 mean sea ice velocity from **model** (NEMO+LIM3)+atmospheric forcing



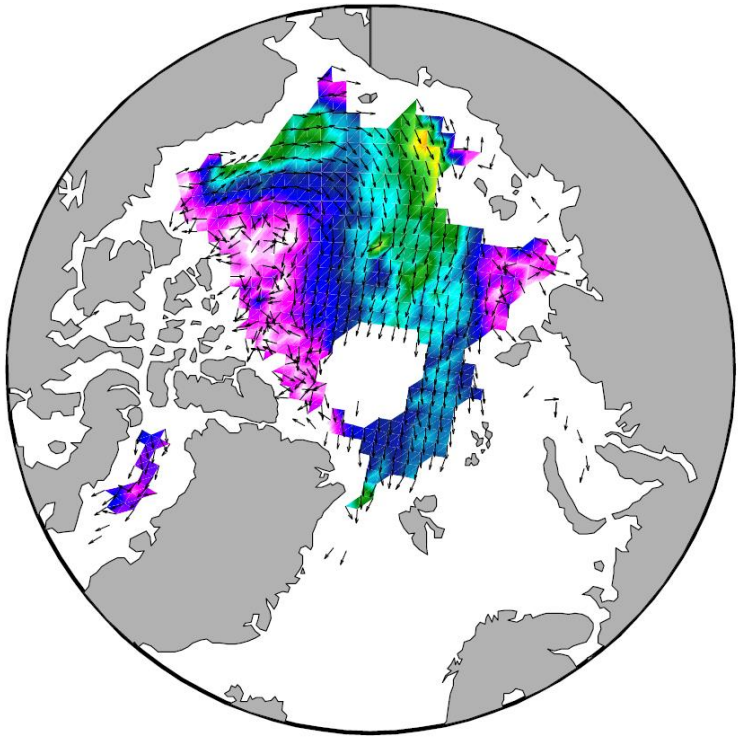


« Diagnostic »

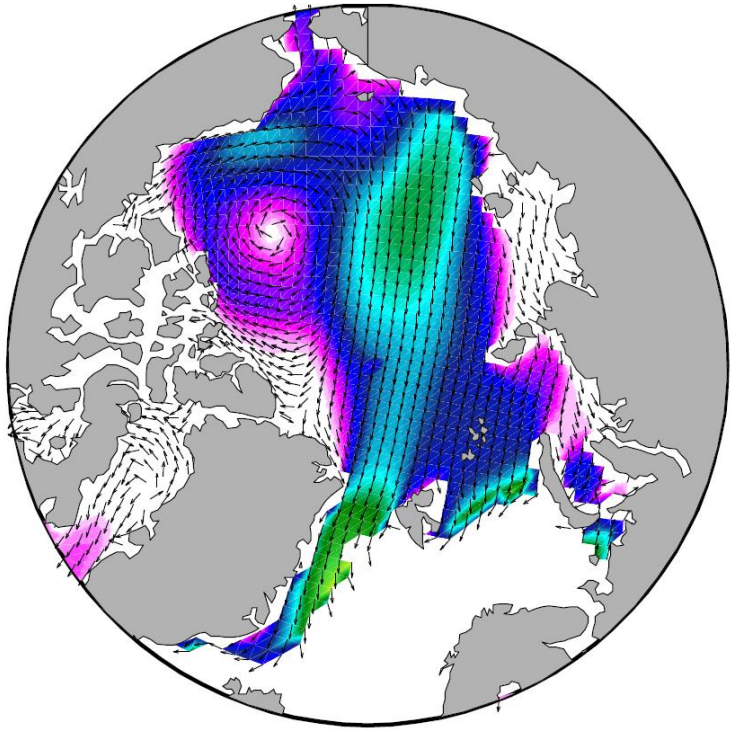
$\mathbf{u}_{mod}(x,y)$

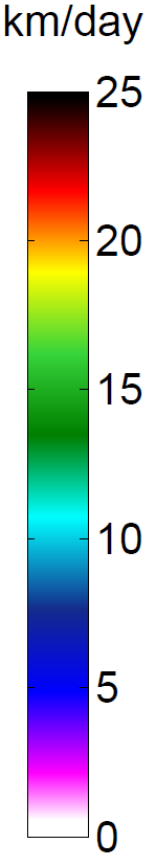
Diagnostics *vs.* metrics

12-14 April 2012 mean sea ice velocity from **observations** (Lavergne et al., J. Geophys. Res., 2010)



12-14 April 2012 mean sea ice velocity from **model** (NEMO+LIM3)+atmospheric forcing





« Diagnostic »
$\mathbf{u}_{obs}(x,y)$

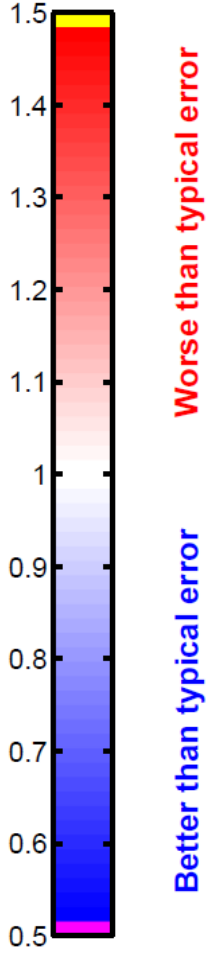
$$\int_{\mathcal{D}} ||\mathbf{u}_{obs}(x,y) - \mathbf{u}_{mod}(x,y)||.A(x,y).dS$$

« Diagnostic »
$\mathbf{u}_{mod}(x,y)$

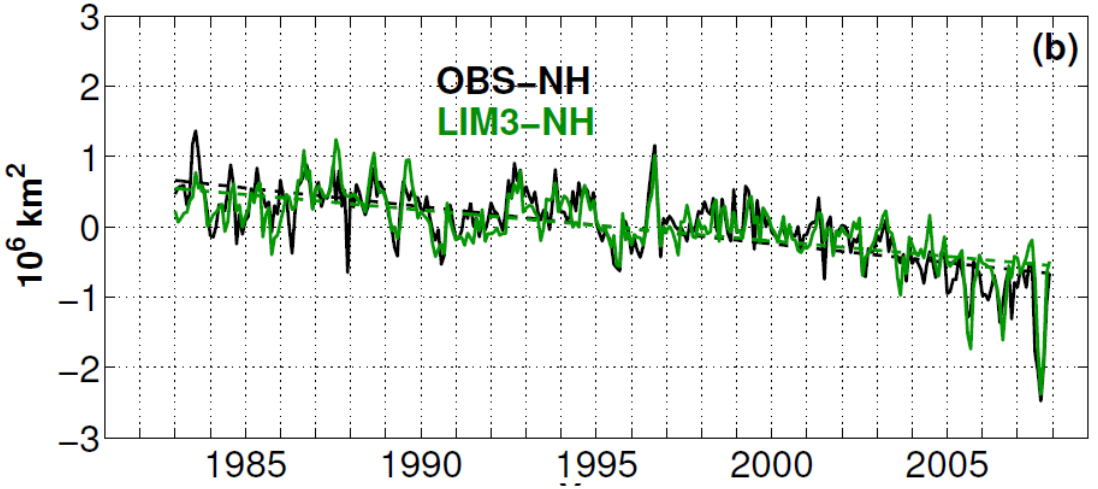
« Metric »: scalar number derived from identical diagnostics processed on two different datasets

1. Standard error metrics
Purpose: tracking model performance

	Metrics		Typical error
1. Mean Conc. North	0.97	0.79	0.15
2. Std Ano Conc. North	1.03	0.77	0.05
3. Trend Ano Conc. North	1.03	0.78	0.02 /decade
4. Mean Conc. South	1.07	1.12	0.15
5. Std Ano Conc. South	0.8	0.71	0.05
6. Trend Ano Conc. South	0.92	0.94	0.02 /decade
7. Mean Ext. North	1.33	0.43	0.8 1e6 km2
8. Std Ano Ext. North	1.22	0.61	0.1 1e6 km2
9. Trend Ano Ext. North	0.23	0.46	0.2 1e6 km2/10y
10. Mean Ext. South	3.58	1.17	0.8 1e6 km2
11. Std Ano Ext. South	0.48	1.1	0.1 1e6 km2
12. Trend Ano Ext. South	0.9	0.52	0.2 1e6 km2/10y
13. Draft North	0.94	0.67	1 m
14. Trend Draft North	0.72	0.32	0.1
15. Thickness South	3.22	2.45	0.15 m
16. Mean Kin. En. North	0.39	0.61	0.0004 J/kg
17. Circulation North	0.86	0.76	0.5
18. Mean Kin. En. South	1.3	1.4	0.0004 J/kg
19. Circulation South	1.26	1.26	0.5
20. Fram Areal Mean Cycle	0.44	0.7	20 1e3 km2
21. Fram Areal Std Ano	0.34	0.9	10 1e3 km2
22. Fram Vol. Mean Cycle	1.14	0.82	50 km3
23. Fram Vol. Std Ano	0.09	0.8	20 km3
	LIM2	LIM3	

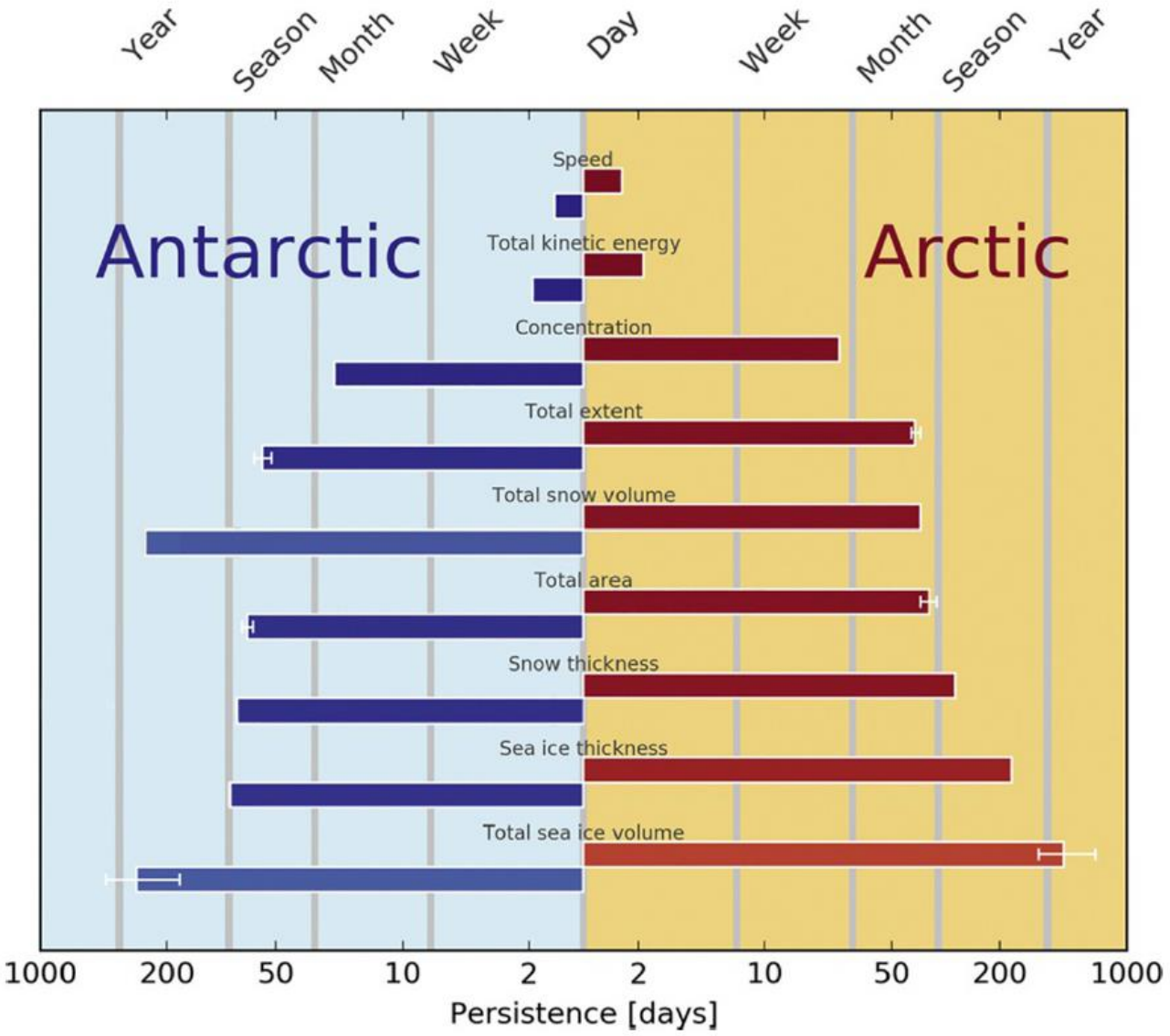


Monthly anomalies of Arctic sea ice extent

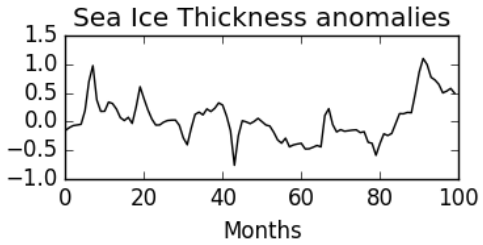


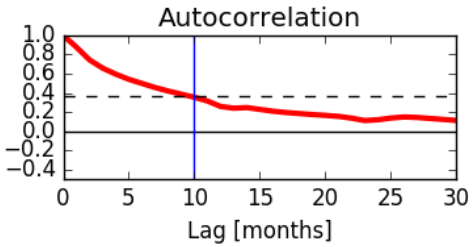
Massonnet et al., *Cryosphere*, 2011

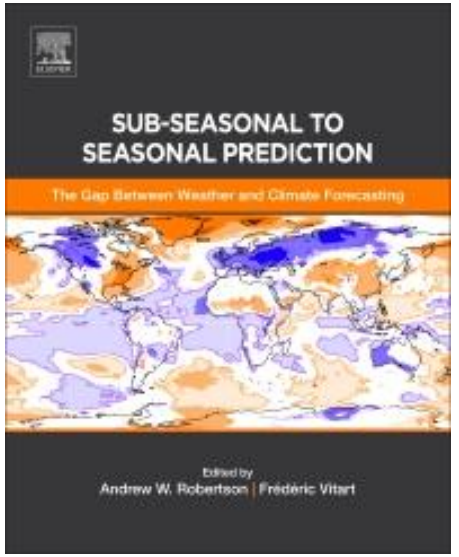
2. Predictability metrics
Purpose: quantifying limits of predictability



Evaluating persistence

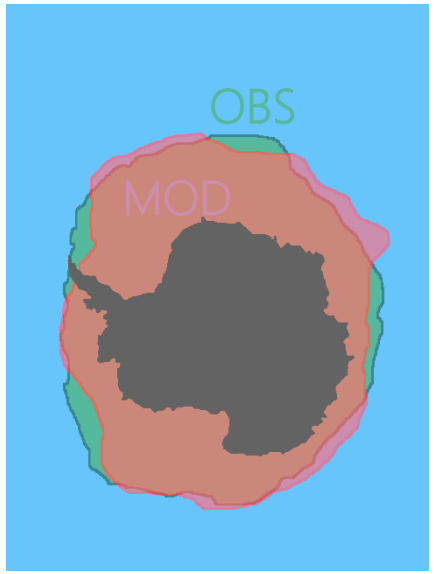




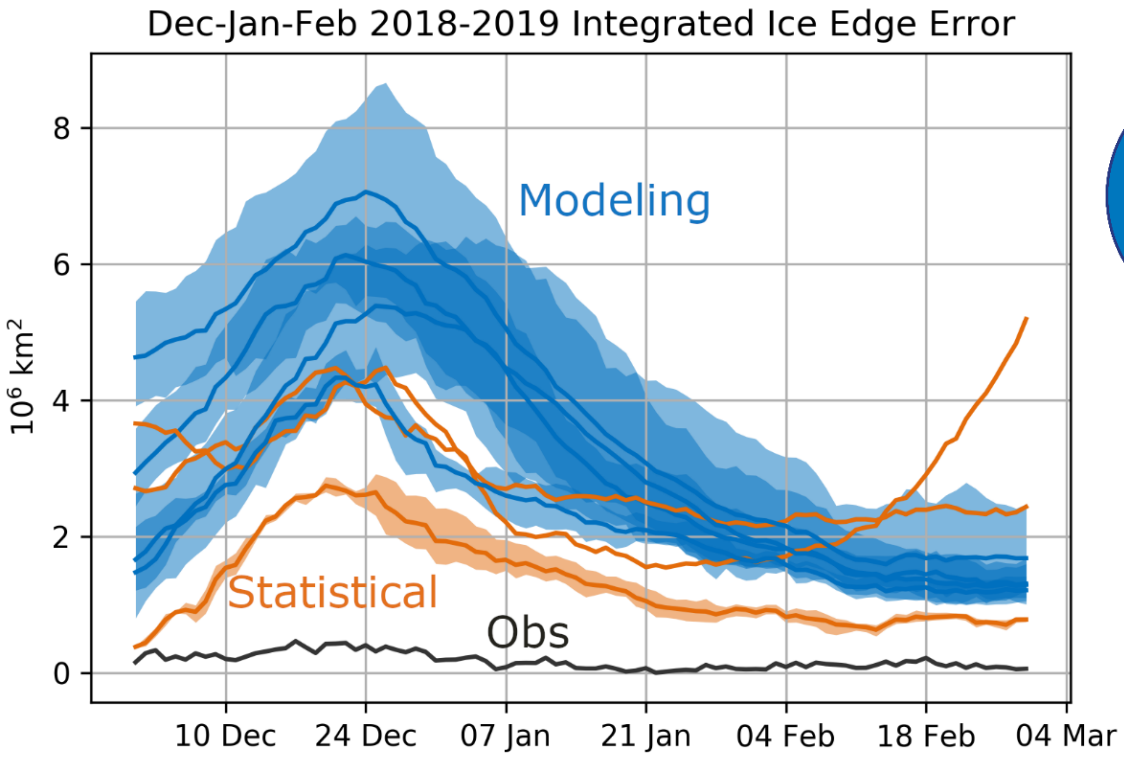


Chevallier et al., S2S book, 2018

3. Forecast error metrics
<u>Purpose</u>: testing the skill of prediction systems



Integrated Ice Edge Error =
Area of overestimation
+
Area of underestimation

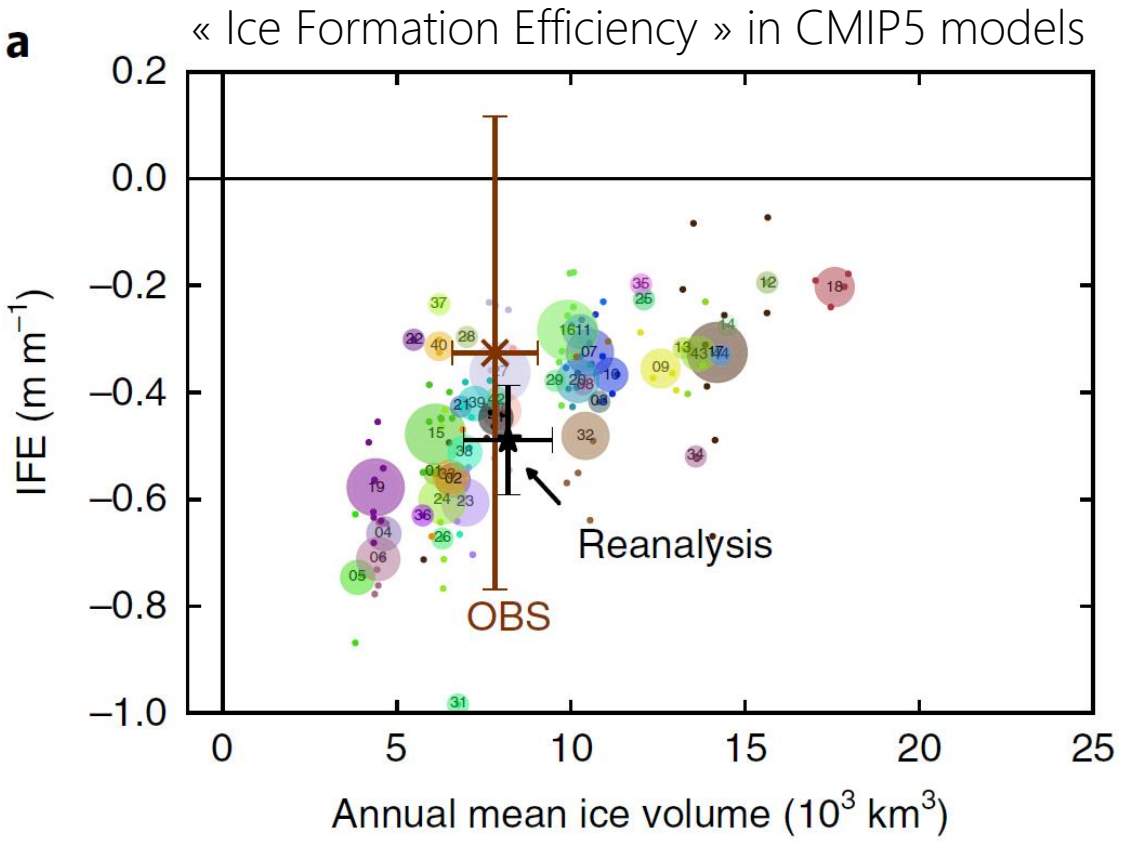


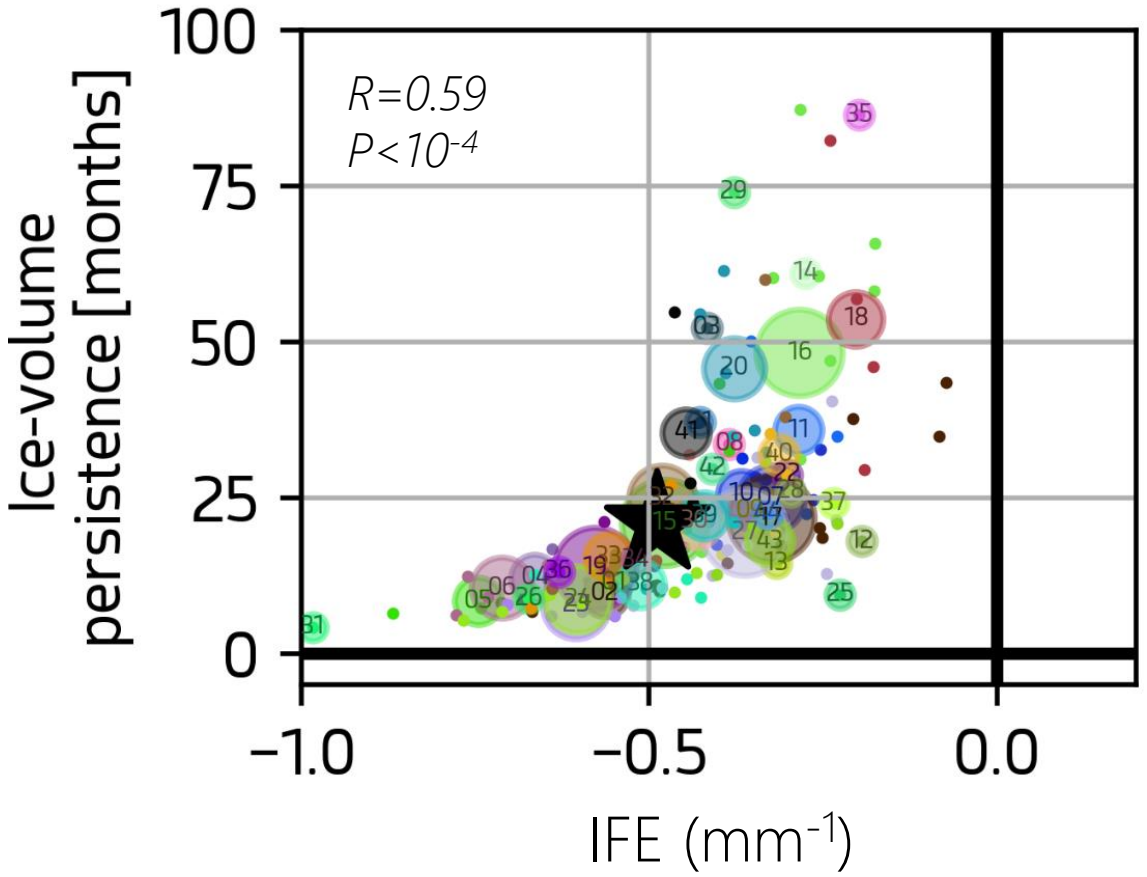
Goessling et al., *Geophys. Res. Lett.*, 2016

Massonnet et al., SIPN South post-season report, 2019

4. Process-based diagnostics

<u>Purpose</u>: measuring the ability of a model to simulate a process or a feedback





Massonnet et al., *Nature Clim. Change*, 2018

Today's questions

What are the purposes of
evaluating sea ice models?

**What should we pay attention to,
when designing an evaluation protocol?**

Is sea ice model evaluation settled?
What are the ways forward?

Requirements for a good set of diagnostics and metrics

Requirements for a good set of diagnostics and metrics

Completeness

Rationale

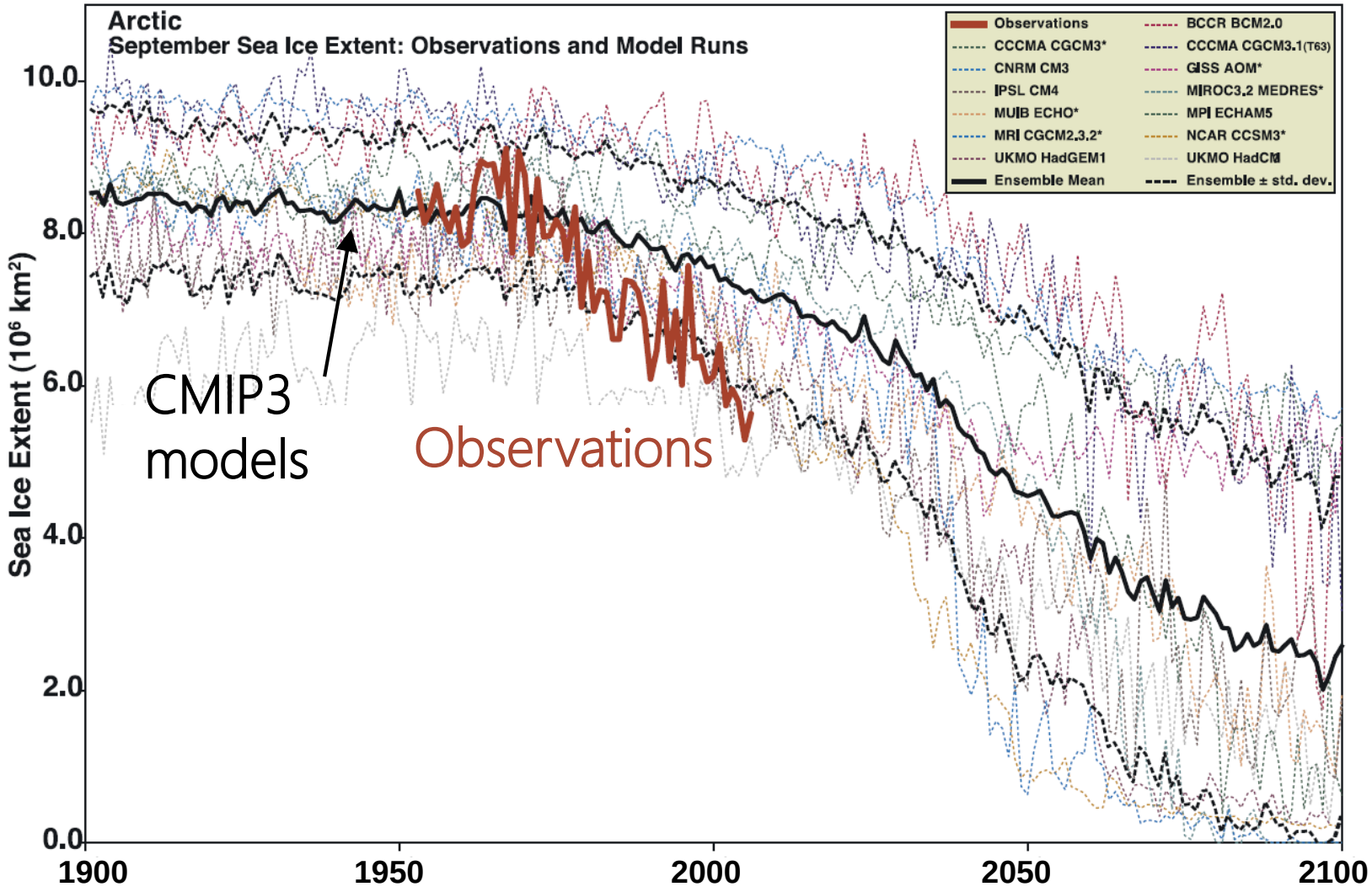
Interpretability

Stability

Transparency

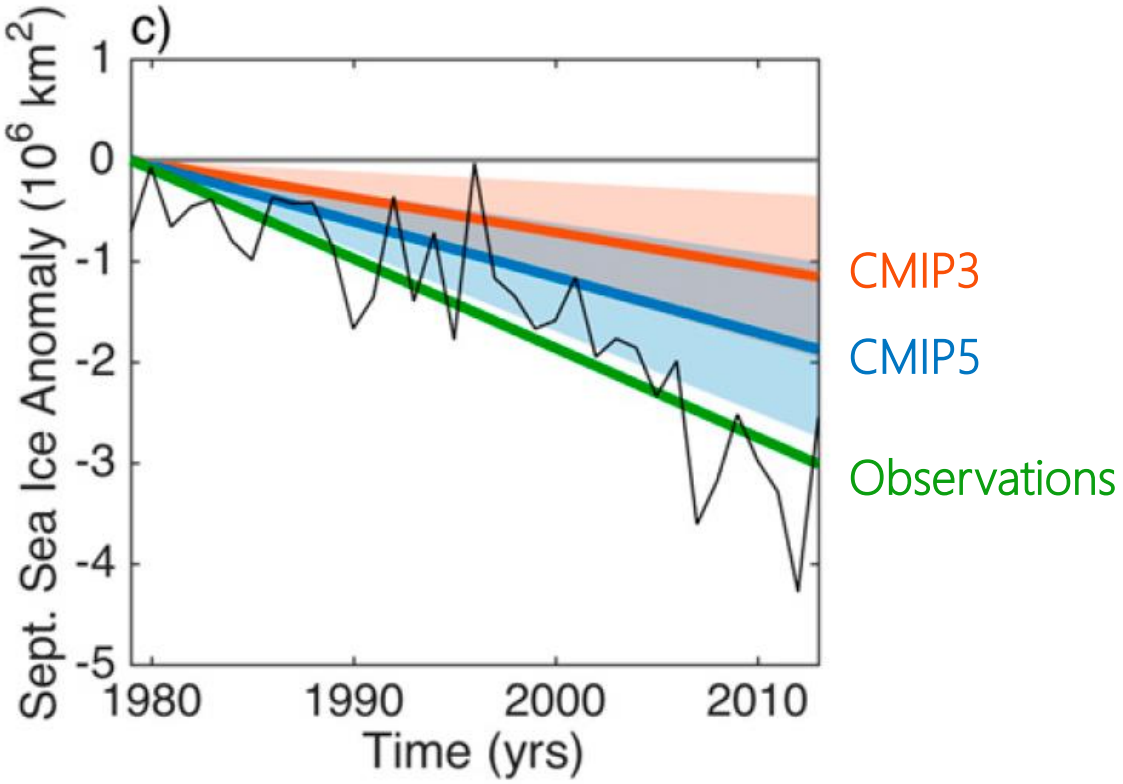
Observability

2007: Arctic sea ice is declining faster than anticipated



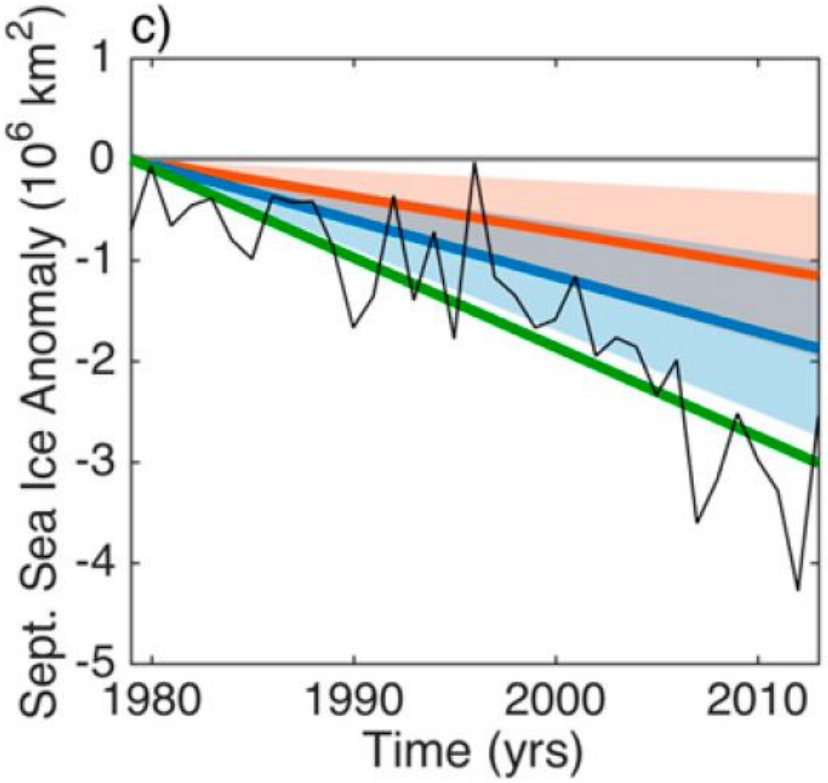
Stroeve et al., *Geophys. Res. Lett.* 2007

2013: Climate models get sea ice extent trends closer to observations



Rosenblum and Eisenman, *J. Clim.*, 2016

2013: Climate models get sea ice extent trends closer to observations



CMIP3

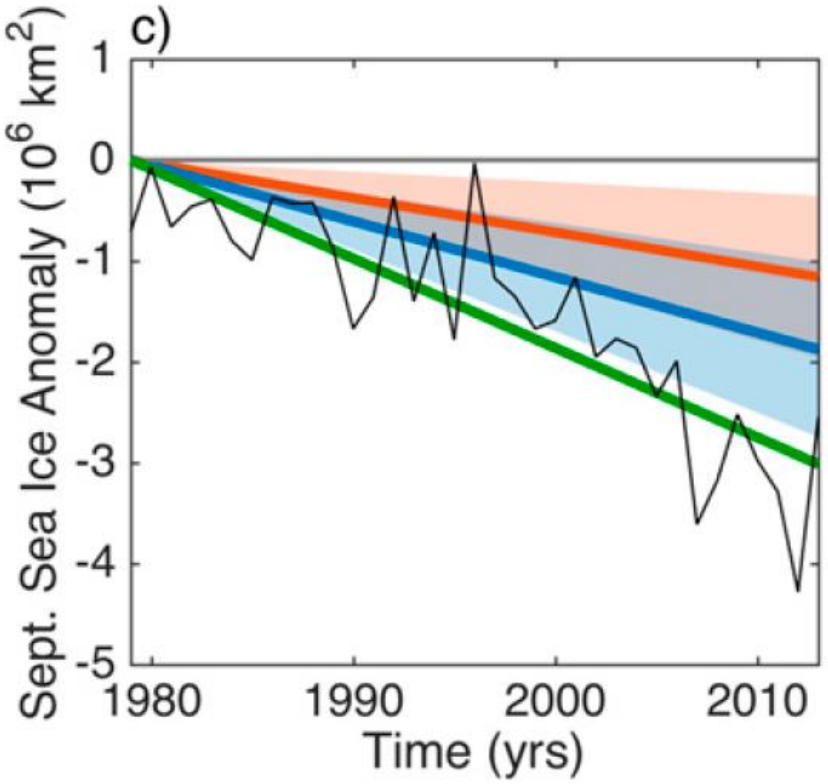
CMIP5

Observations

Model improvement pays off!

Rosenblum and Eisenman, *J. Clim.*, 2016

2013: Climate models get sea ice extent trends closer to observations (for wrong reasons)



CMIP3

CMIP5

Observations

Faster Arctic Sea Ice Retreat in CMIP5 than in CMIP3 due to Volcanoes

ERICA ROSENBLUM AND IAN EISENMAN

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California

(Manuscript received 24 May 2016, in final form 4 October 2016)

ABSTRACT

Rosenblum and Eisenman, *J. Clim.*, 2016

Requirements for a good set of diagnostics and metrics

Completeness

Rationale

Interpretability

Stability

Transparency

Observability

Never look at numbers in isolation, or you may well be fooled

Requirements for a good set of diagnostics and metrics

Completeness

Rationale

Interpretability

Stability

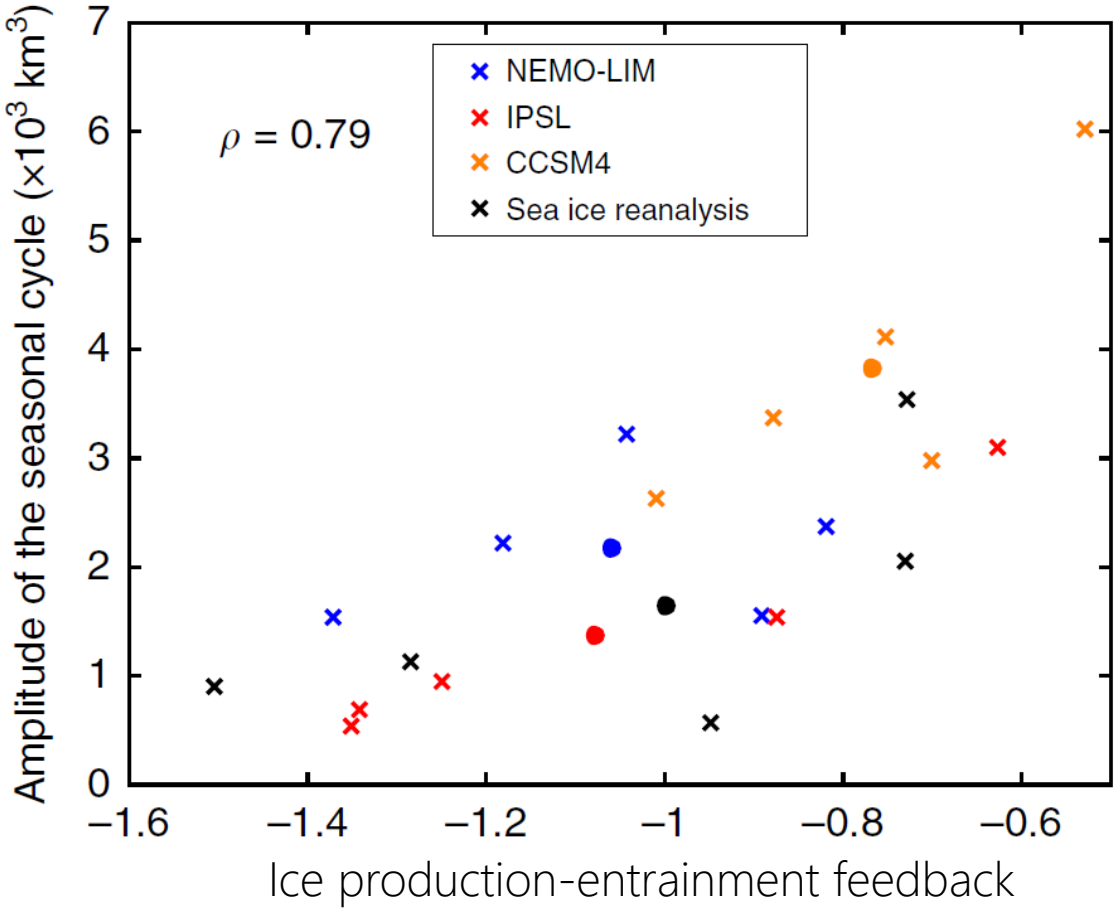
Transparency

Observability

Requirements for a good set of diagnostics and metrics

- Completeness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability

What are the origins of Antarctic sea ice model biases?



Goosse et al., *Nature Commun.*, 2018

Requirements for a good set of diagnostics and metrics

Completeness

Rationale

Interpretability

Stability

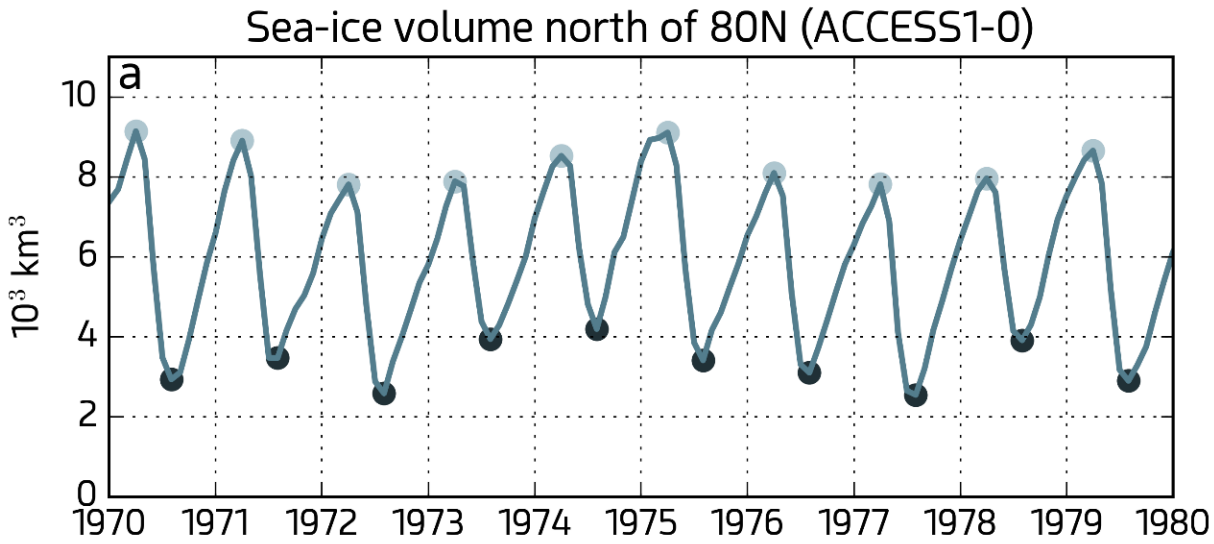
Transparency

Observability

Always design metrics and diagnostics with a scientific question in mind

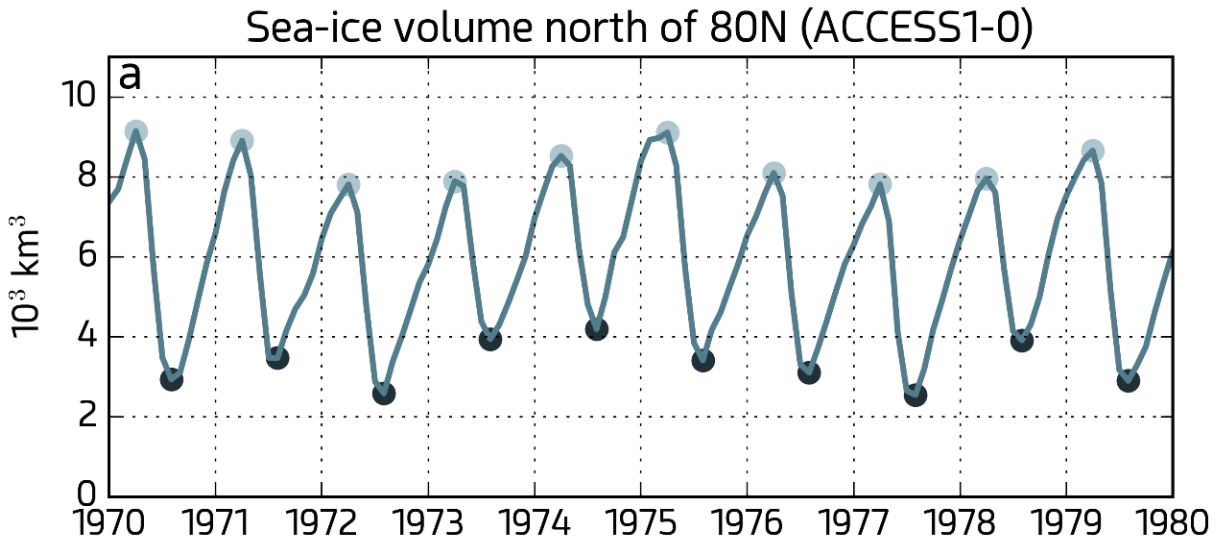
Requirements for a good set of diagnostics and metrics

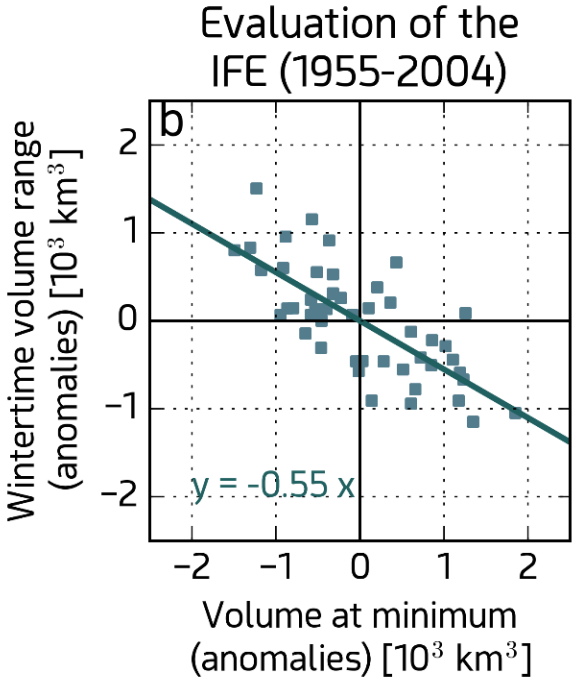
- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability



Requirements for a good set of diagnostics and metrics

- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability





Requirements for a good set of diagnostics and metrics

- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability

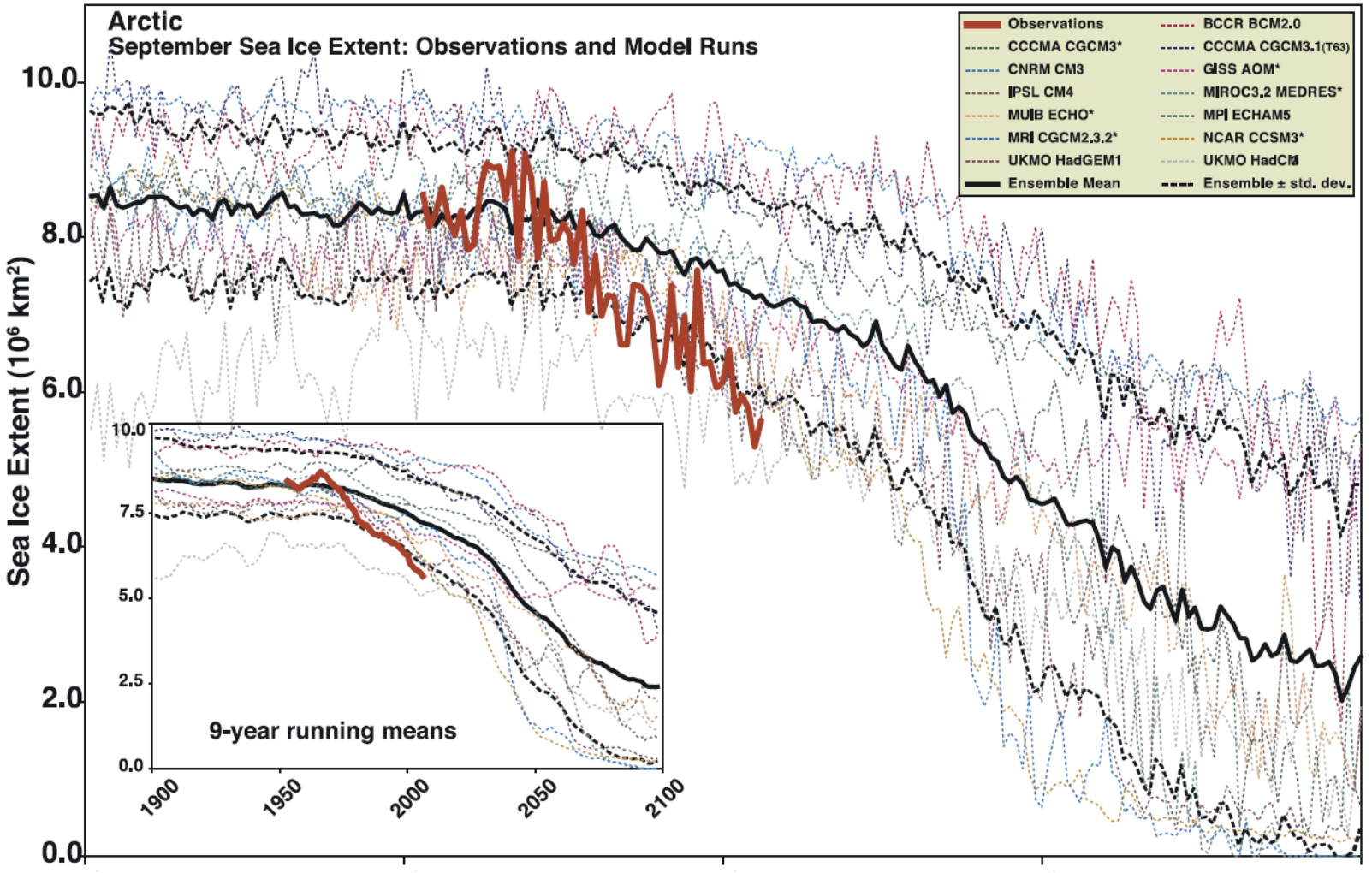
Give the others the chance yo understand what you have done

Requirements for a good set of diagnostics and metrics

- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability

Requirements for a good set of diagnostics and metrics

- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability

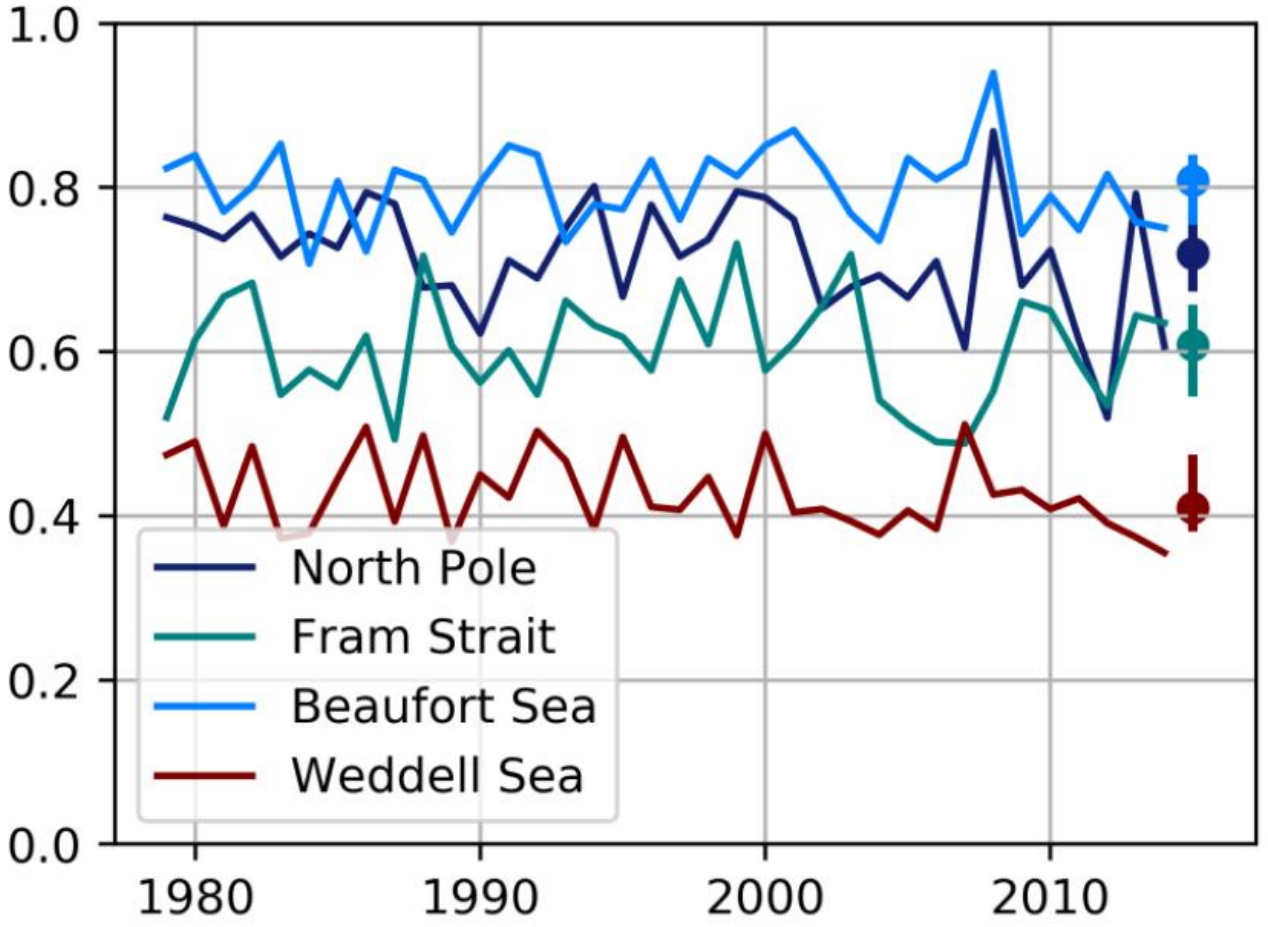


Requirements for a good set of diagnostics and metrics

$$HCI = \frac{k_s h_i}{k_s h_i + k_i h_s}$$

Vancop's « Heat Conduction Index »

- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability



Requirements for a good set of diagnostics and metrics

Completeness

Rationale

Interpretability

Stability

Transparency

Observability

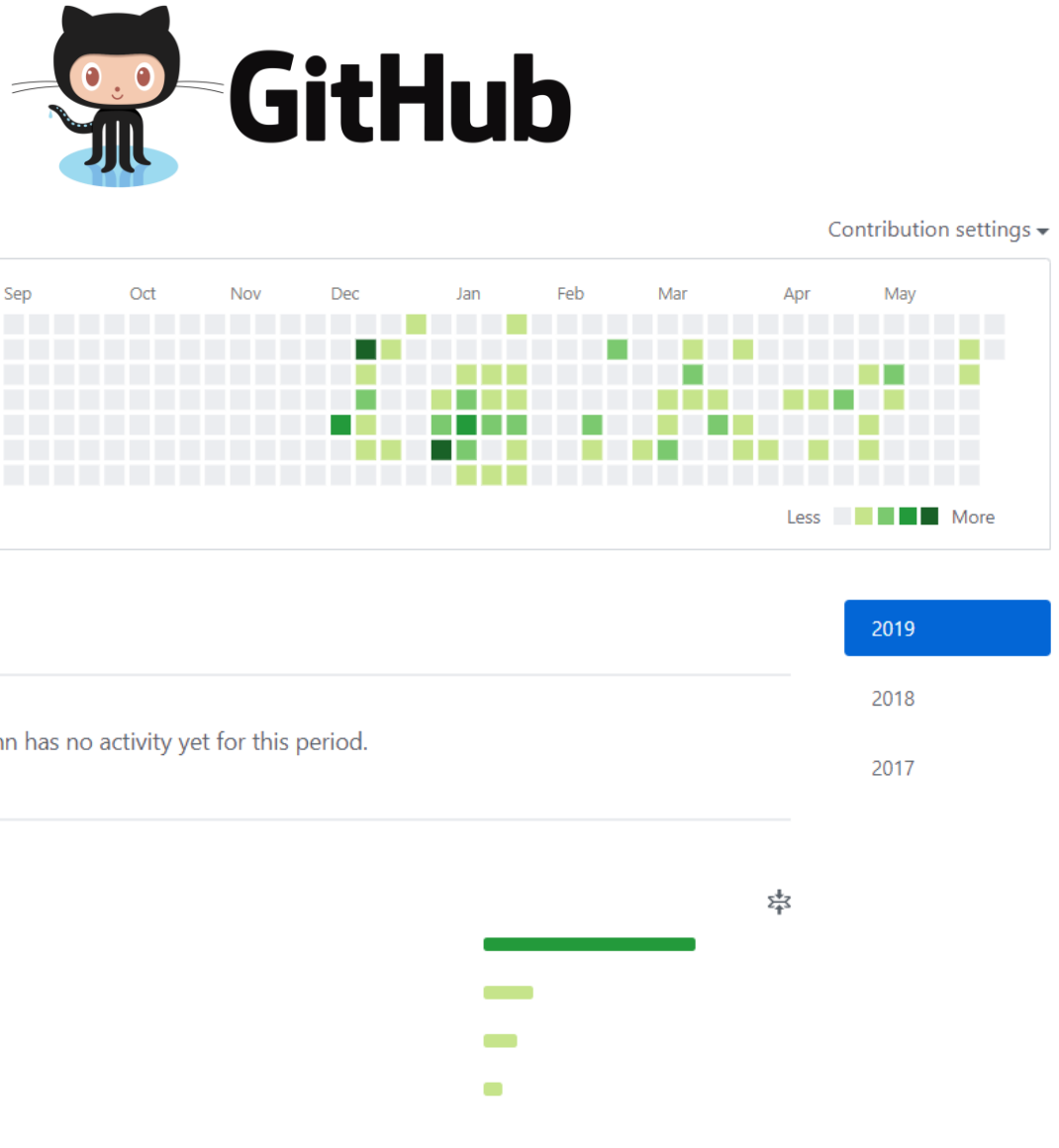
Metrics and diagnostics should be insensitive to interannual and internal climate variability

Requirements for a good set of diagnostics and metrics

- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability

Requirements for a good set of diagnostics and metrics

- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability



Requirements for a good set of diagnostics and metrics

Completeness

Rationale

Interpretability

Stability

Transparency

Observability

Open your model to the scrutiny of other researchers

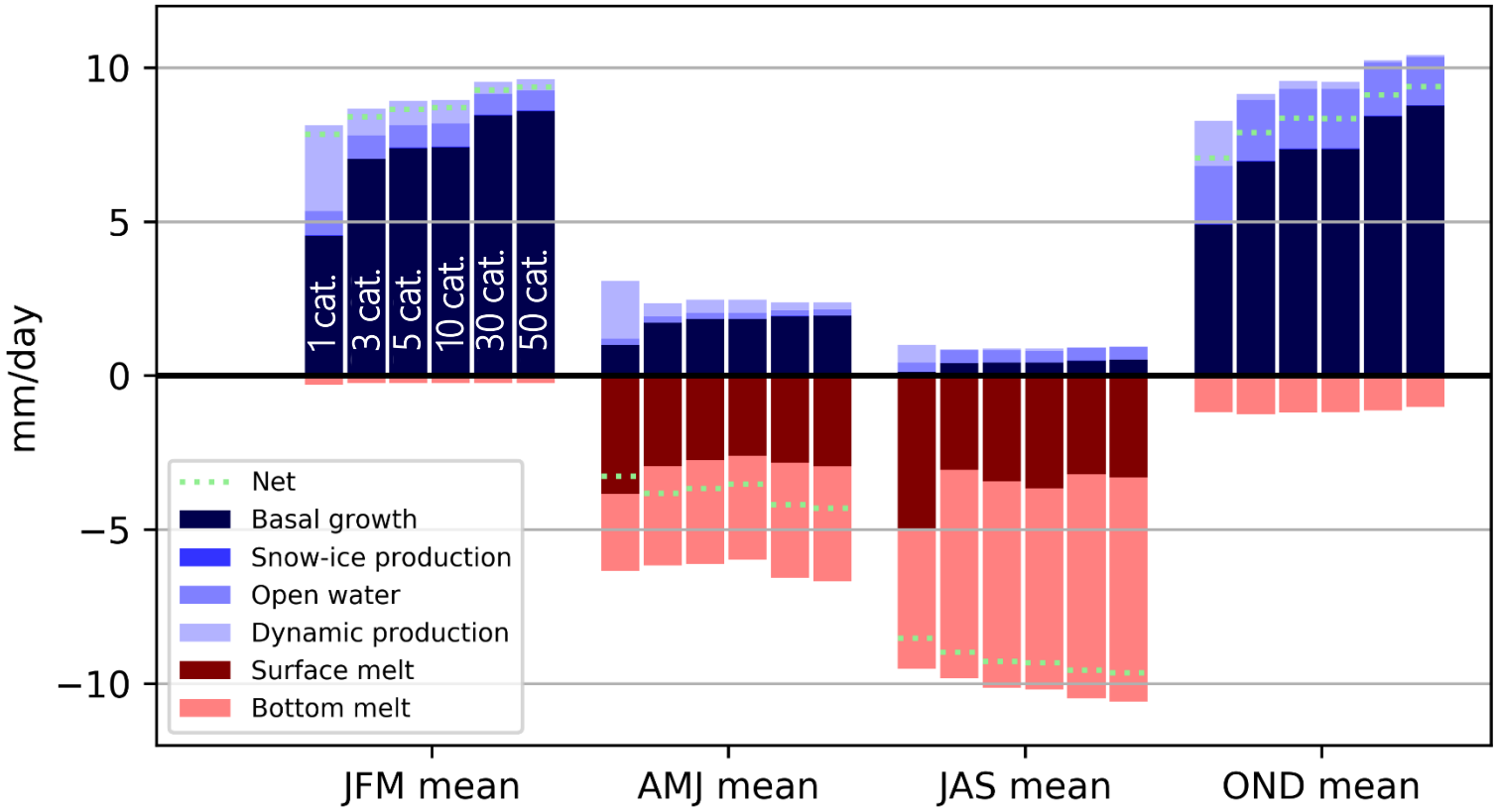
Requirements for a good set of diagnostics and metrics

- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability

Requirements for a good set of diagnostics and metrics

- **C**ompleteness
- **R**ationale
- **I**nterpretability
- **S**tability
- **T**ransparency
- **O**bservability

Simulated Arctic sea ice mass balance when increasing number of thickness categories (NEMO3.6-LIM3)



Massonnet et al., *Geosci. Mod. Dev. Discuss.*, 2019

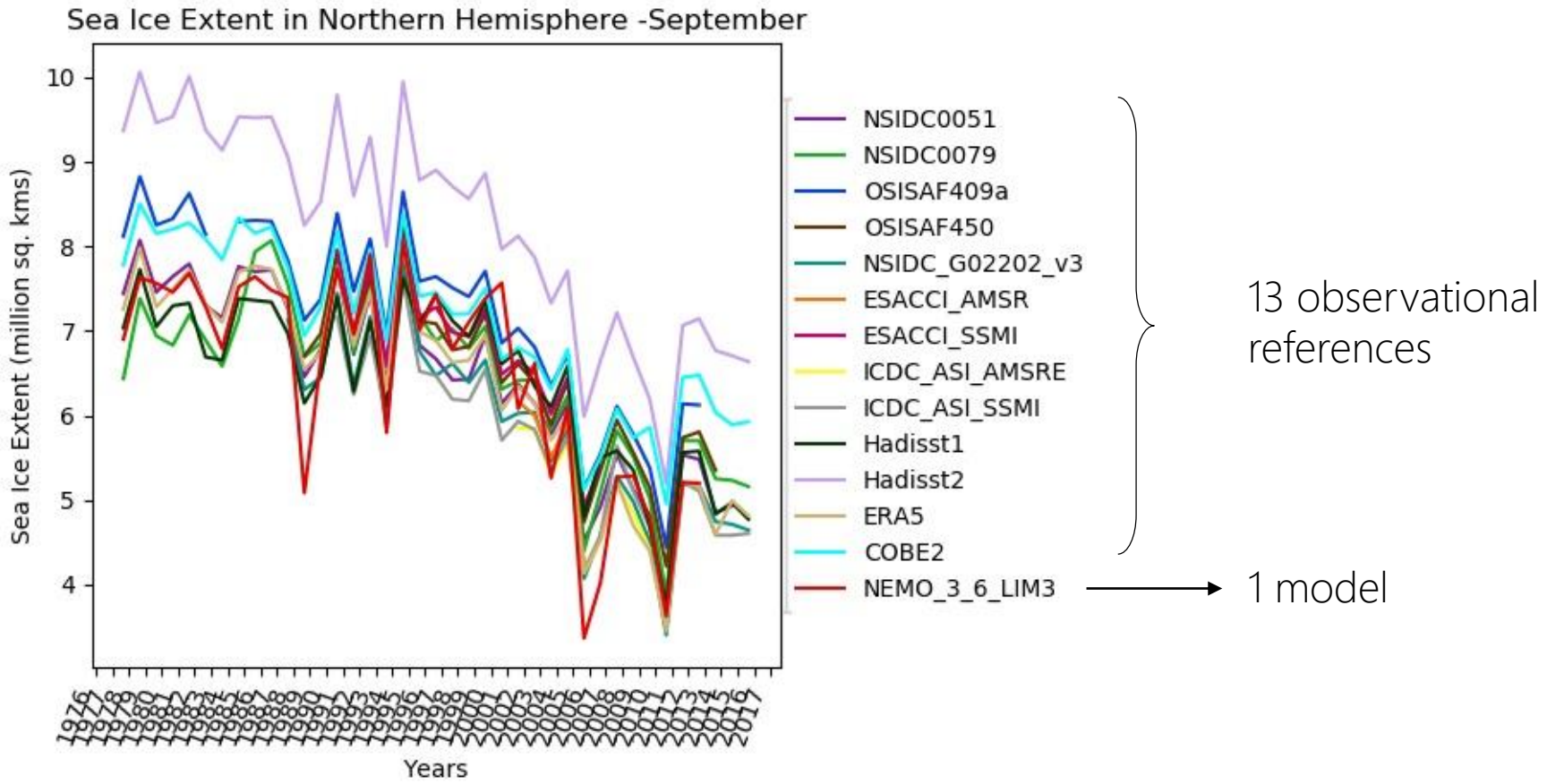
Today's questions

What are the purposes of
evaluating sea ice models?

What should we pay attention to,
when designing an evaluation protocol?

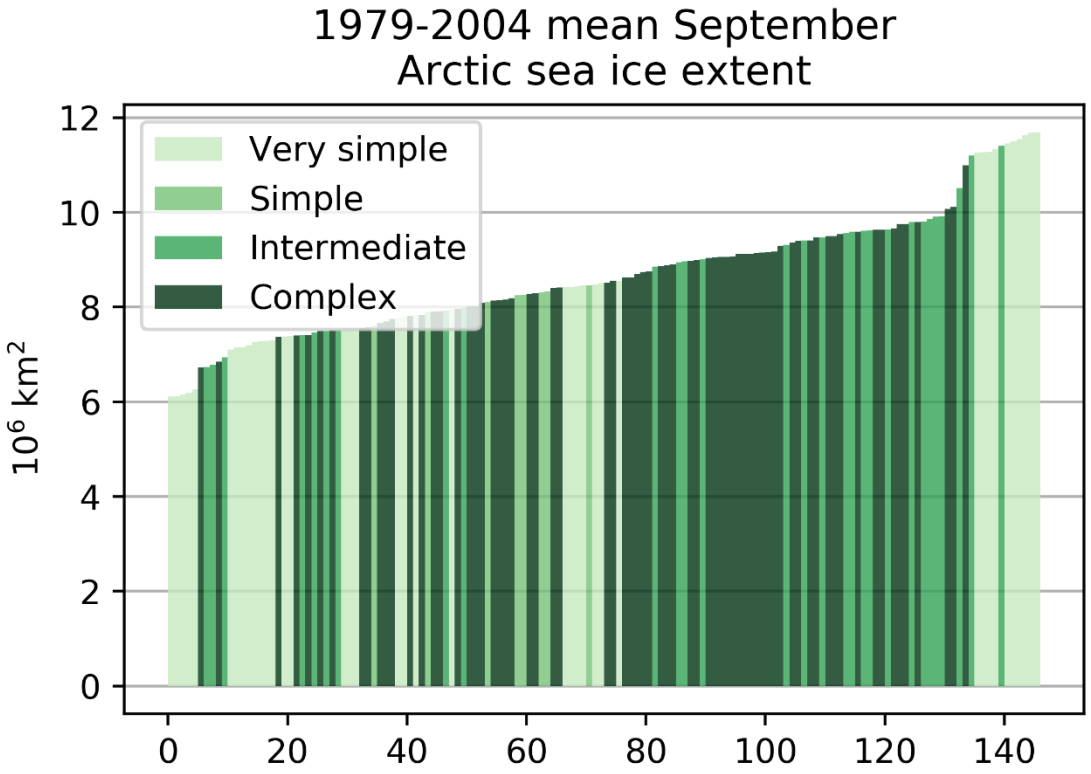
Is sea ice model evaluation settled?
What are the ways forward?

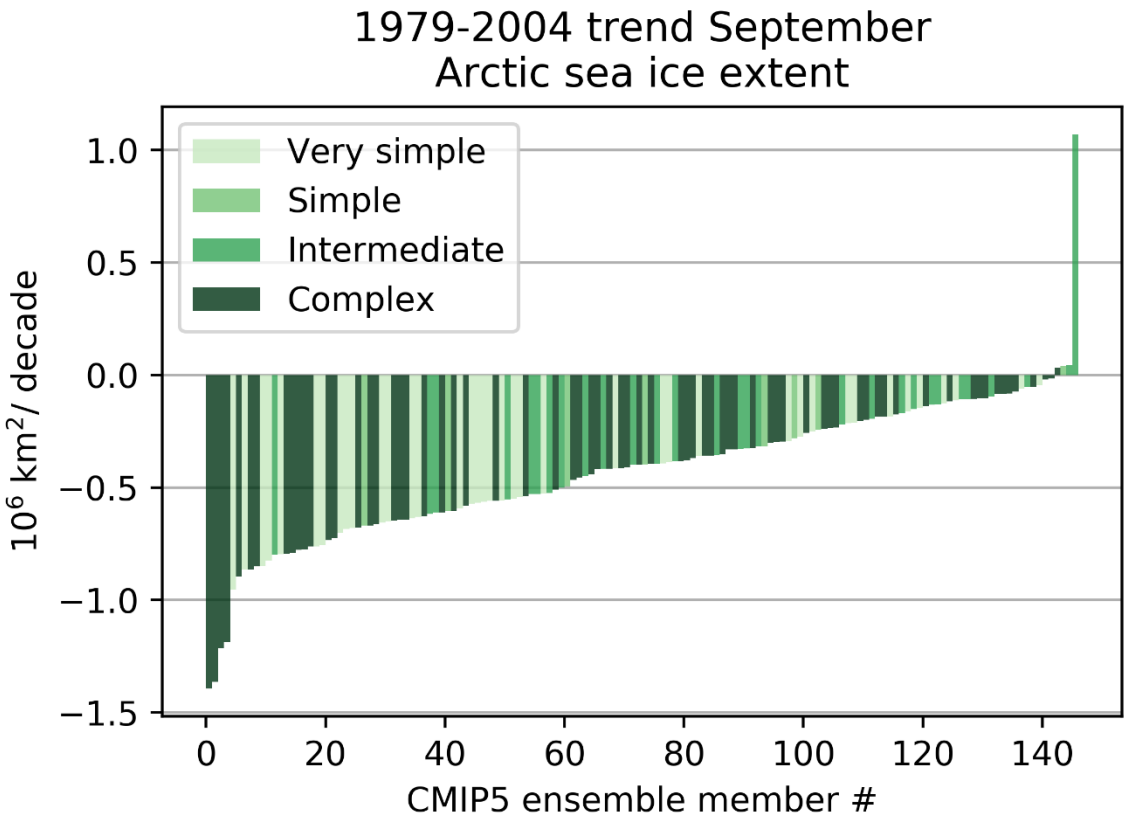
Challenge #1: Introducing observational uncertainty in diagnostics and metrics



Courtesy A. Rana

Challenge #2: Designing metrics that can segregate simple models from complex ones – retrospectively





Notz, Hunke, Massonnet, Vancoppenolle, submitted

Conclusions

- No metric or diagnostic is all-purpose but we can still agree on a number of minimal requirements that they should satisfy.

- Despite apparences, model evaluation is very subjective! Never infer model performance from numbers alone.

- Model evaluation is a statistical inference process, and therefore has to be communicated with uncertainties