University of Colorado
Boulder

The Signal-to-Noise Error in Decadal Climate Modes

Jeremy Klavans

Pedro DiNezio, Amy Clement, Mark Cane, Chengfei He, Lisa Murphy Goes, Clara Deser, Tim Shanahan, Victoria Todd

Confronting Earth System Model Trends with Observations: The Good, the Bad, and the Ugly

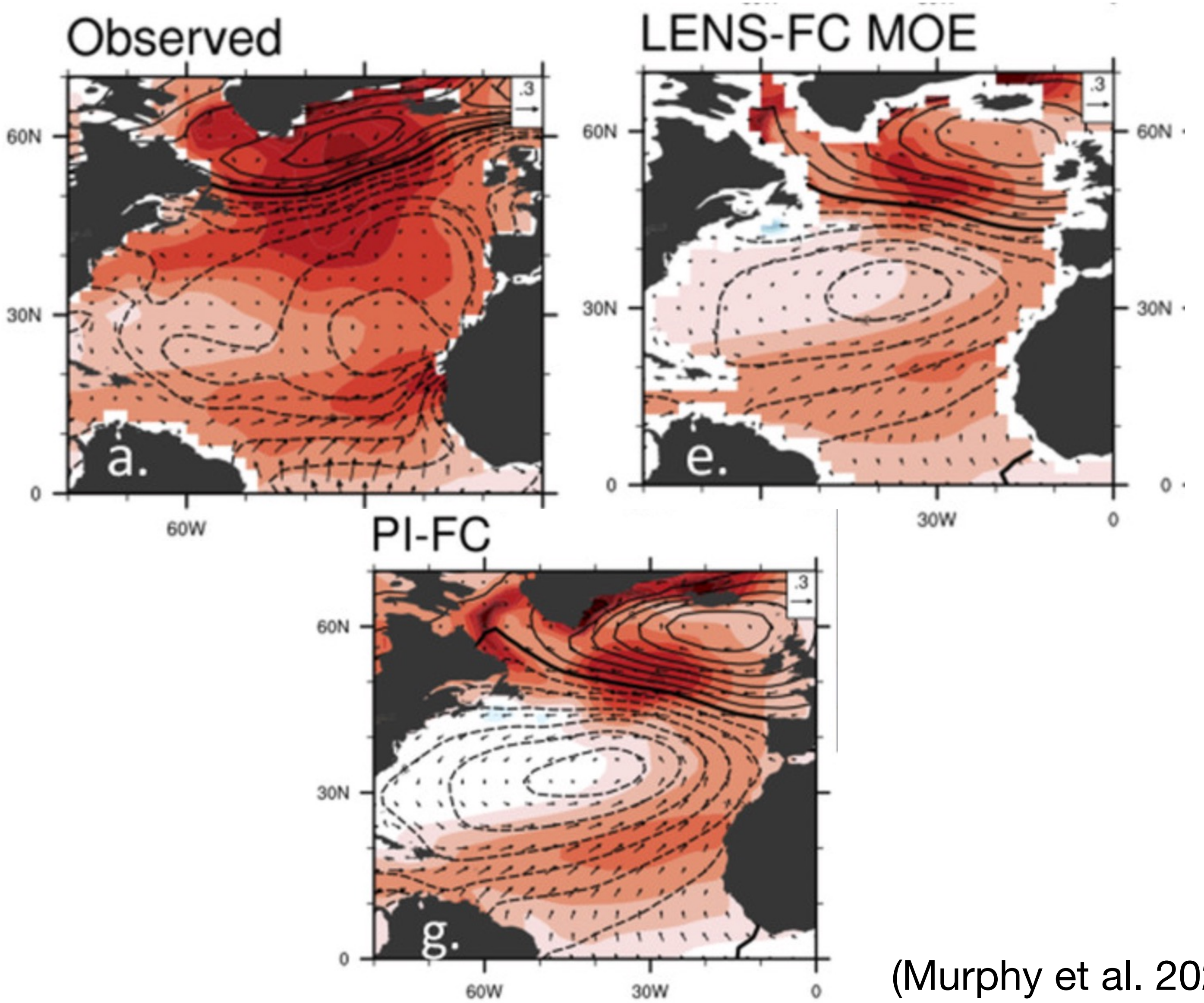
March 14, 2024

Outline

1. What are models doing well?
   - Models produce multidecadal modes with reasonably realistic spatial patterns and impacts
2. Where can models improve?
   - Observed variance is an outlier, relative to model ensemble spread
   - This is an error, caused by a S/N ratio that is too low in models
3. How can the S/N ratio be improved? What could rectifying this error teach us?
   - Potential sources of the S/N error
   - Implications of the S/N error

Models simulate internal patterns of variability

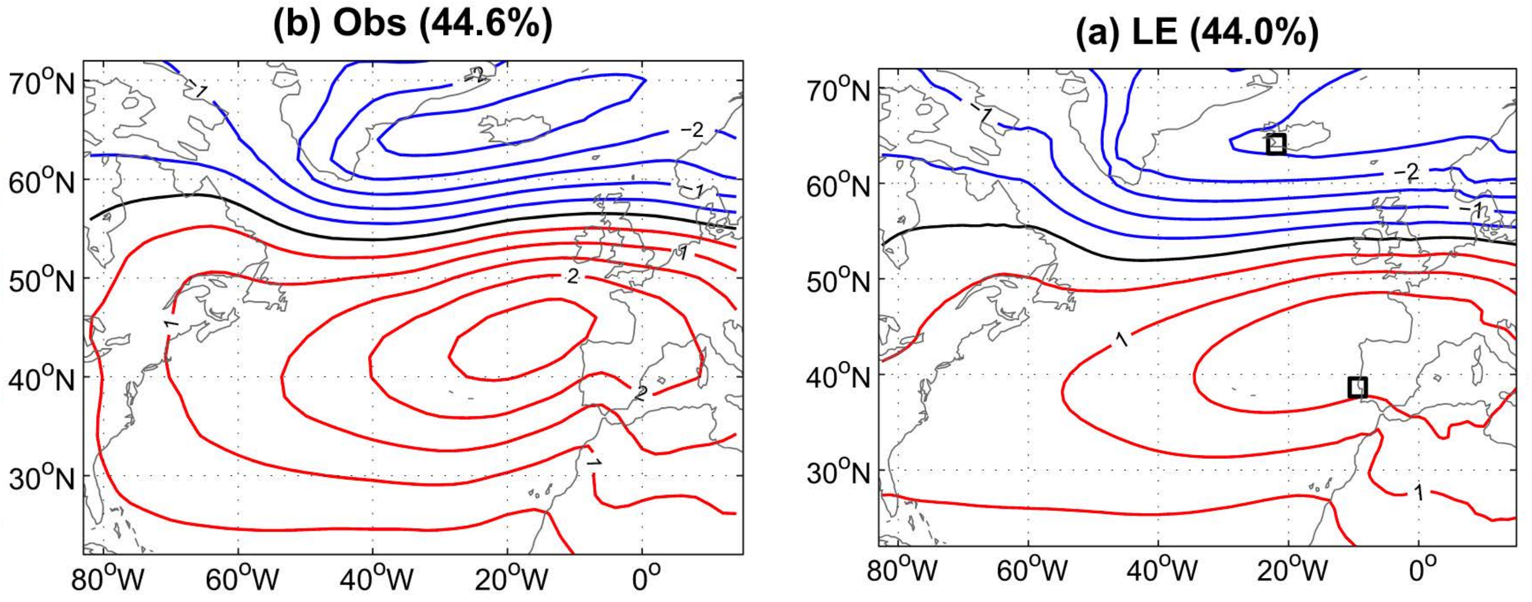
- Models produce multidecadal modes with reasonably realistic spatial patterns and impacts
  - **AMV**
  - NAO
  - PDO
- These internal modes produce have impacts similar to observations
  - MDR VWS; Sahel precipitation
  - Euro. Precipitation
  - SW US Precipitation



(Murphy et al. 2021)

Models simulate internal patterns of variability

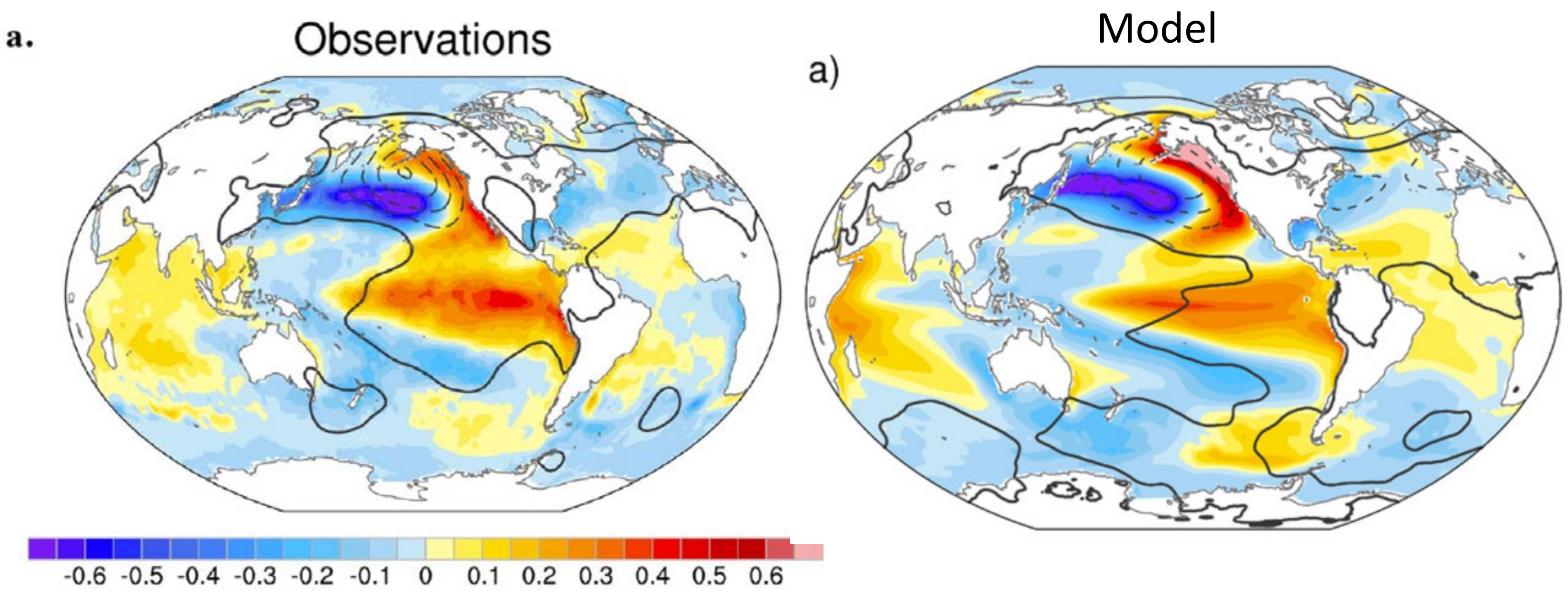
- Models produce multidecadal modes with reasonably realistic spatial patterns and impacts
  - AMV
  - **NAO**
  - PDO
- These internal modes produce have impacts similar to observations
  - MDR VWS; Sahel precipitation
  - Euro. Precipitation
  - SW US Precipitation



(Kim et al. 2018)

Models simulate internal patterns of variability

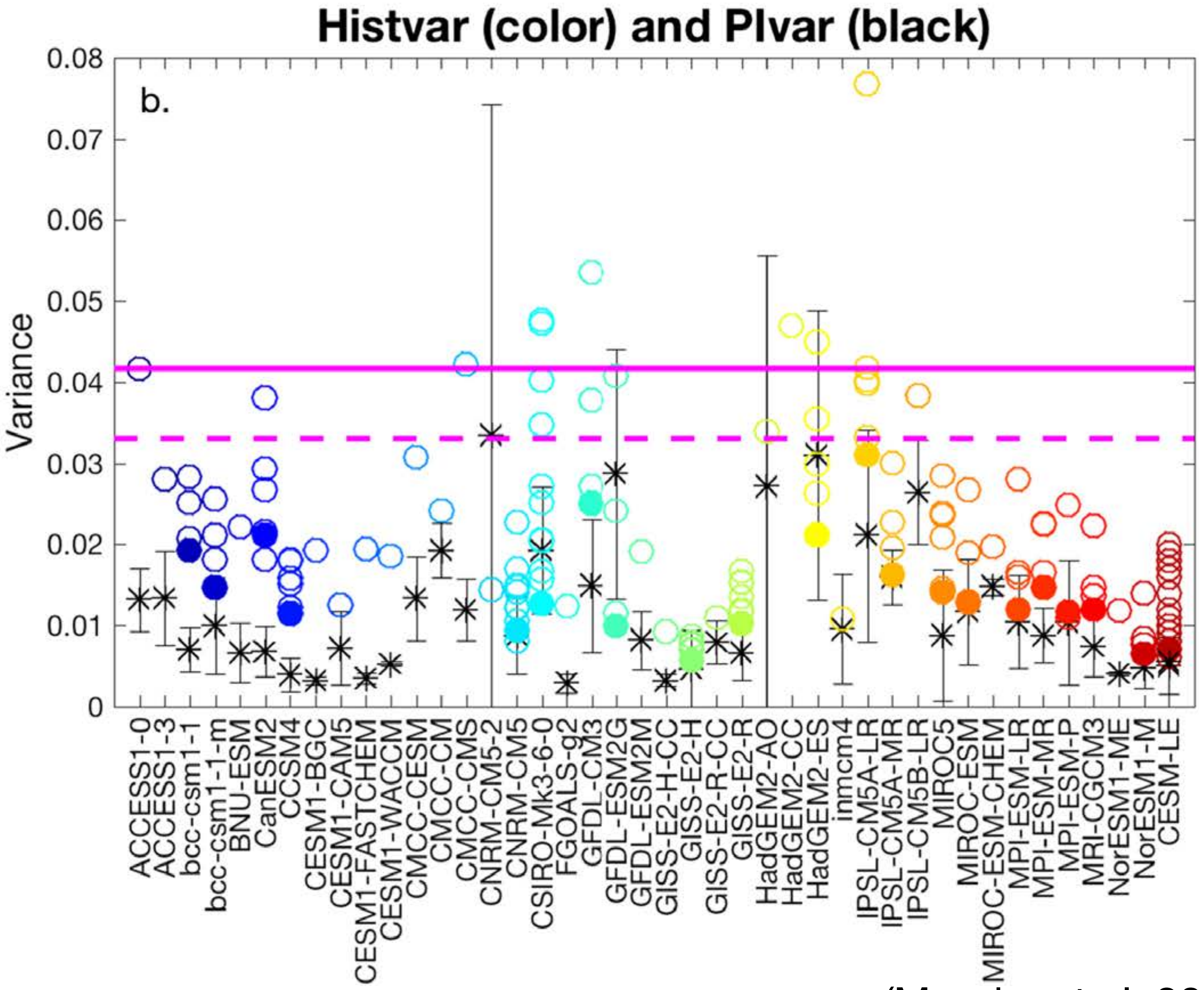
- Models produce multidecadal modes with reasonably realistic spatial patterns and impacts
  - AMV
  - NAO
  - **PDO**
- These internal modes produce have impacts similar to observations
  - MDR VWS; Sahel precipitation
  - Euro. Precipitation
  - SW US Precipitation



(Newman et al. 2016)

Models underestimate the amplitude of variability

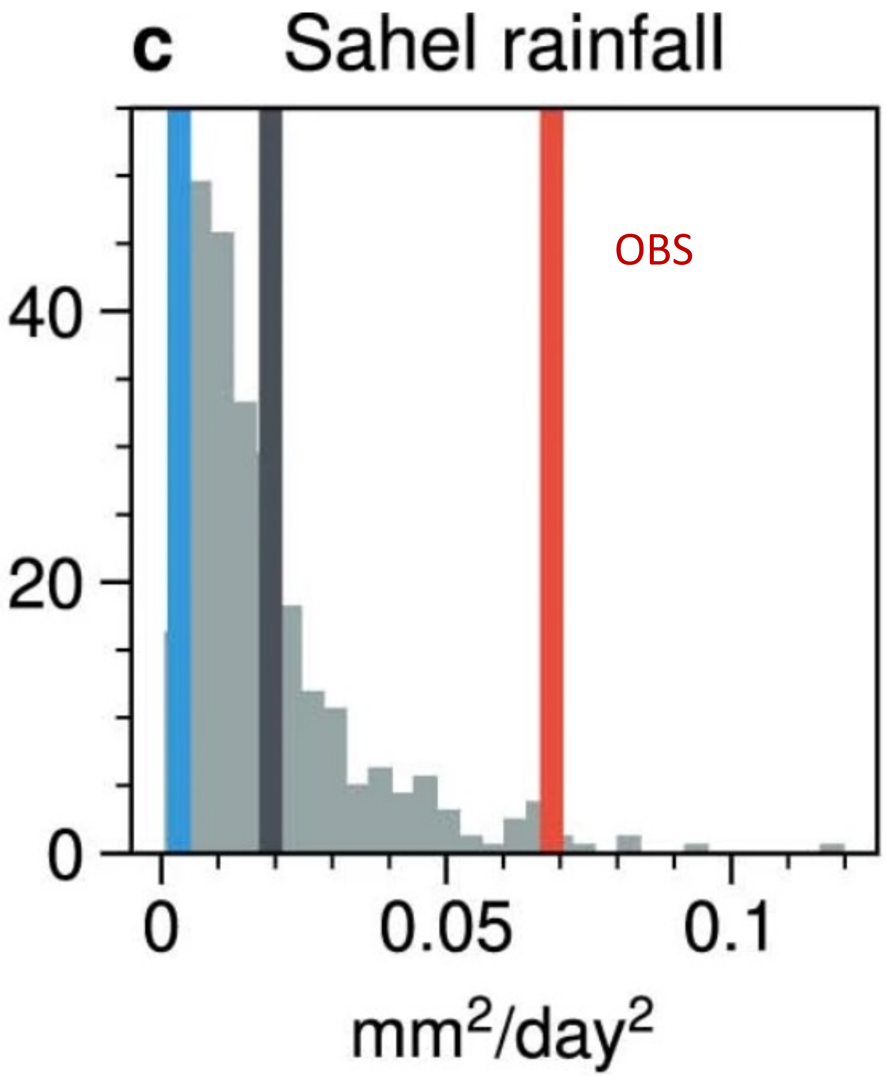
- Observed decadal variances are consistently on the edge/outside of ensemble spread
- If independent, this is an unlikely result
- Examples:
  - **AMV**
  - Sahel precipitation
  - Atlantic vertical wind shear
  - NAO/N. Euro. precip.
  - Western US precipitation



(Murphy et al. 2017)

Models underestimate the amplitude of variability

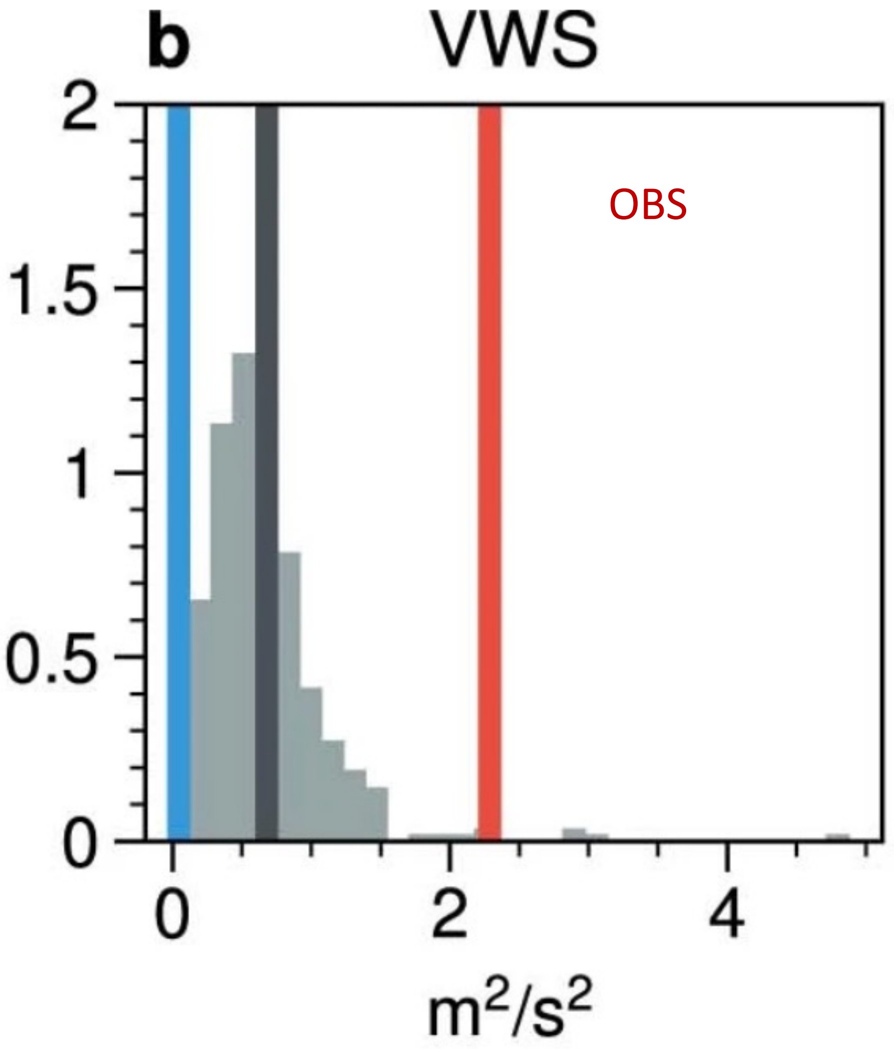
- Observed decadal variances are consistently on the edge/outside of ensemble spread
- If independent, this is an unlikely result
- Examples:
  - AMV
  - **Sahel precipitation**
  - Atlantic vertical wind shear
  - NAO/N. Euro. precip.
  - Western US precipitation



(He et al. 2023)

Models underestimate the amplitude of variability

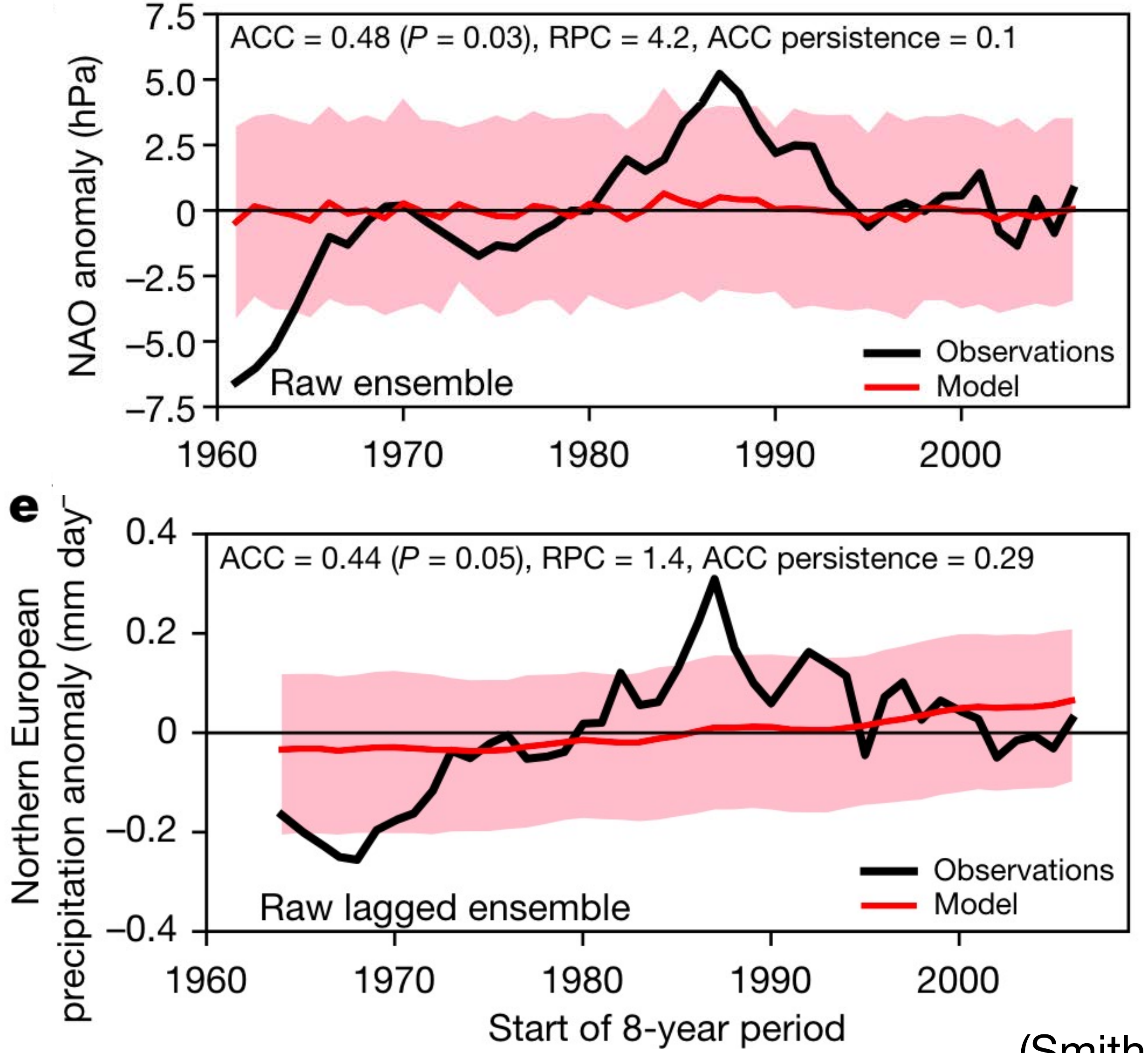
- Observed decadal variances are consistently on the edge/outside of ensemble spread
- If independent, this is an unlikely result
- Examples:
  - AMV
  - Sahel precipitation
  - **Atlantic vertical wind shear**
  - NAO/N. Euro. precip.
  - Western US precipitation



(He et al. 2023)

Models underestimate the amplitude of variability

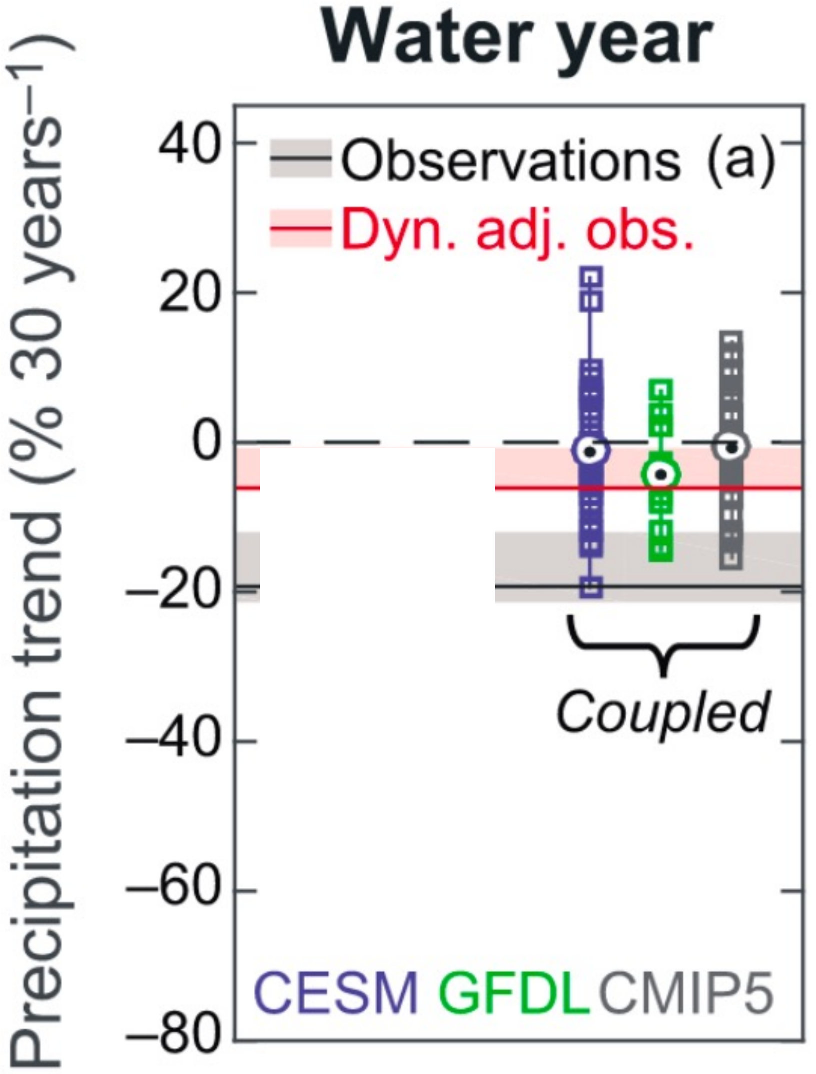
- Observed decadal variances are consistently on the edge/outside of ensemble spread
- If independent, this is an unlikely result
- Examples:
    - AMV
    - Sahel precipitation
    - Atlantic vertical wind shear
    - **NAO/N. Euro. precip.**
    - Western US precipitation



(Smith et al. 2020)

Models underestimate the amplitude of variability

- Observed decadal variances are consistently on the edge/outside of ensemble spread
- If independent, this is an unlikely result
- Examples:
    - AMV
    - Sahel precipitation
    - Atlantic vertical wind shear
    - NAO/N. Euro. precip.
    - **Western US precipitation**



(Lehner et al. 2018)

Large ensembles of climate models

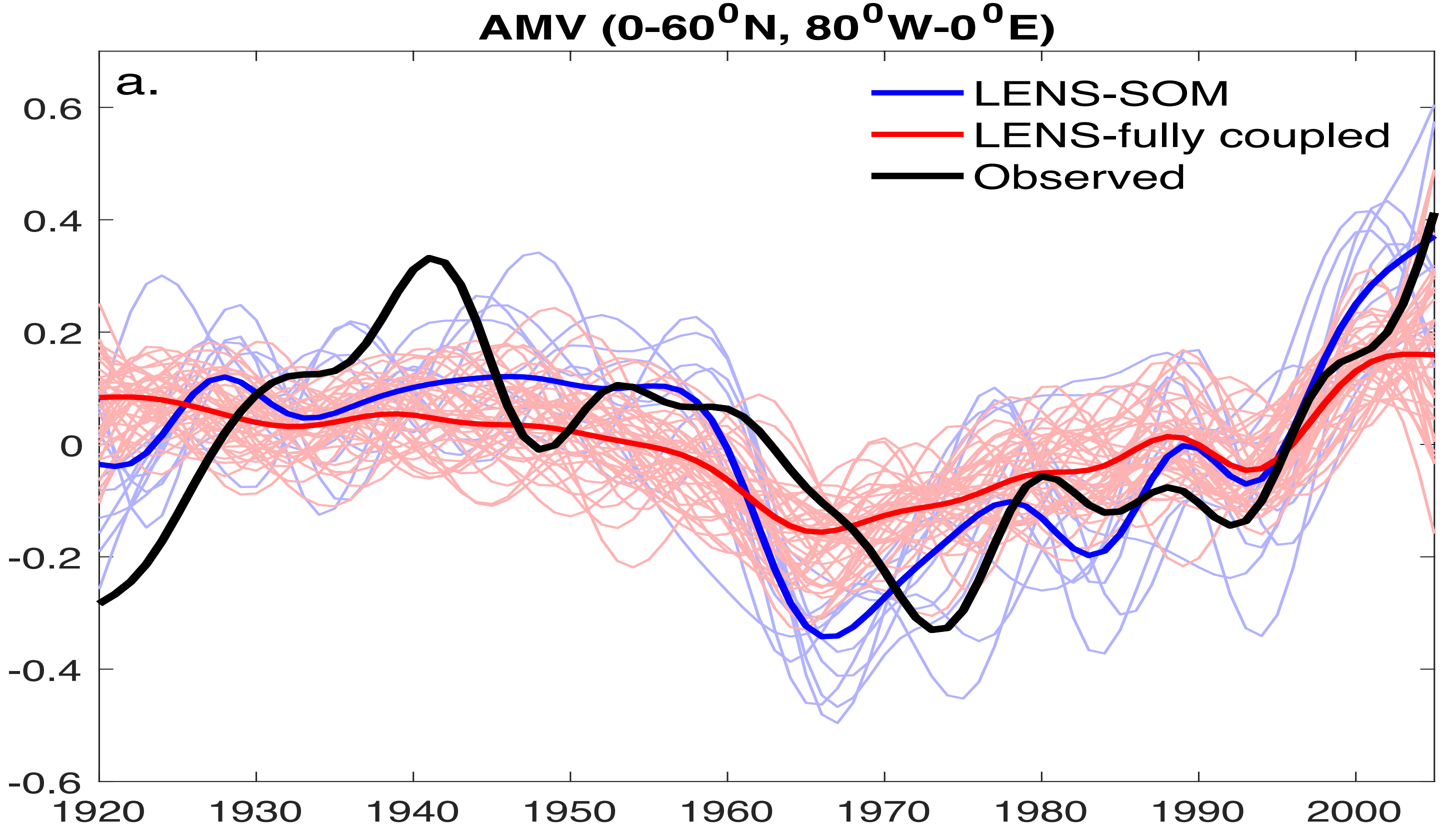
- Either we live at the edge of many pseudo-independent distributions OR there is a problem with models
- If there is a problem in models, either:

  1. Models underestimate interval variability (ensemble spread is too small)
  2. Models underestimate the response to external forcing (ensemble mean too weak)

- Use large ensembles of climate models to evaluate these possibilities

Ensemble means highly correlated with obs.

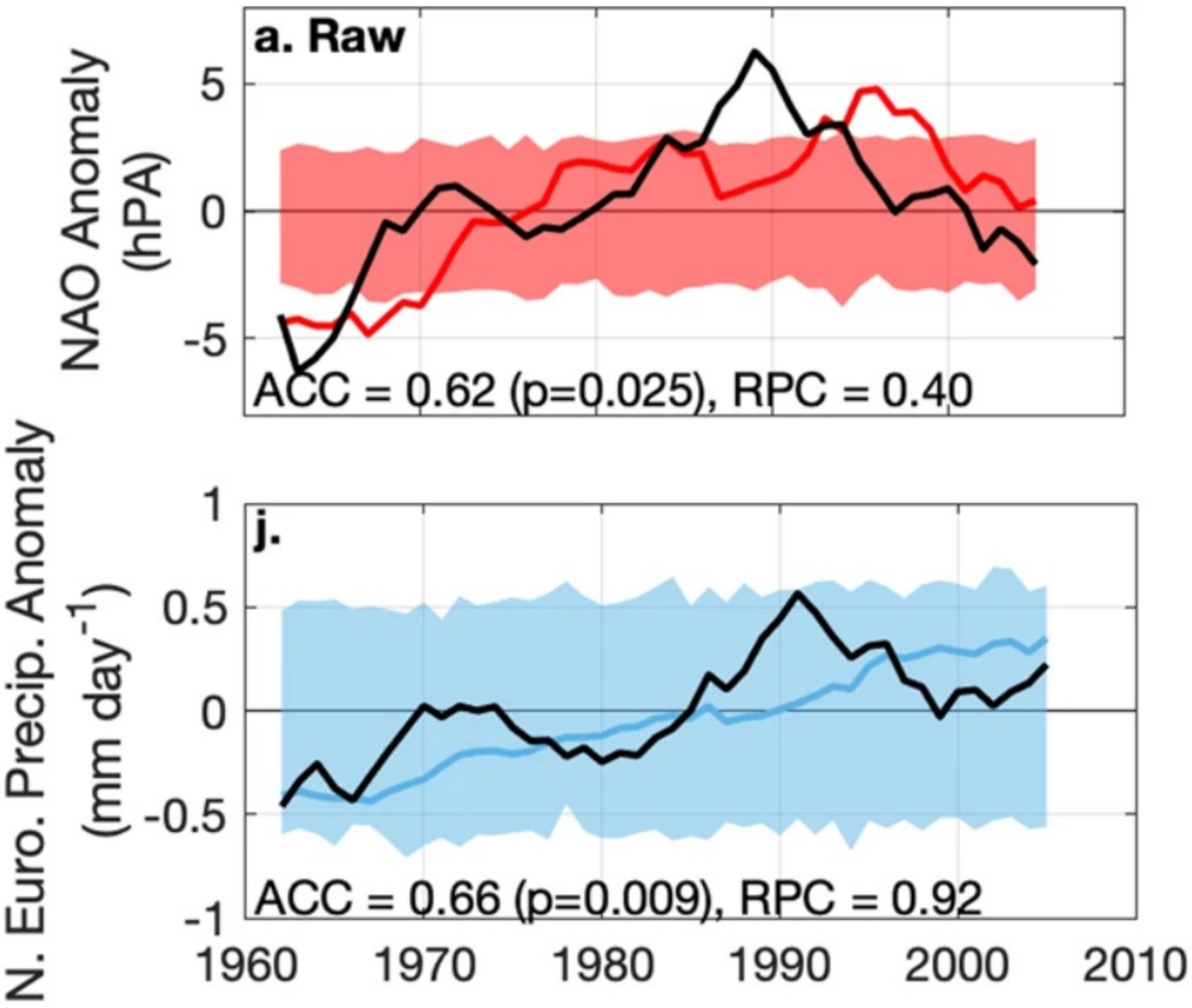
- Ensemble mean indices are surprisingly highly correlated with observations
  - **AMV ($R^2 \approx 0.75$)**
  - NAO ($R^2 \approx 0.60$)
  - PDO ($R^2 \approx 0.50$)
  - Still a role for internal variability!
- And these high correlations are unlikely from internal variability alone



(Murphy et al. 2021)

Ensemble means highly correlated with obs.

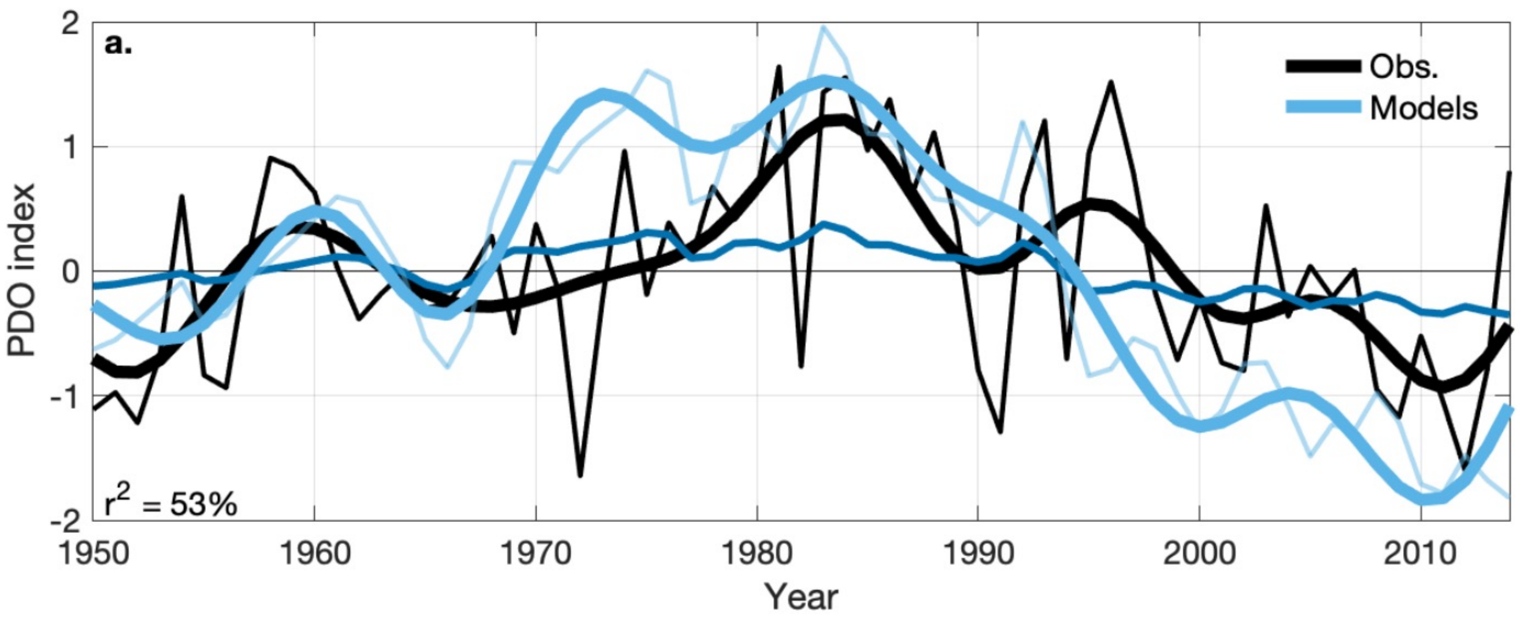
- Ensemble mean indices are surprisingly highly correlated with observations
  - AMV ($R^2 \approx 0.75$)
  - **NAO ($R^2 \approx 0.60$)**
  - PDO ($R^2 \approx 0.50$)
  - Still a role for internal variability!
- And these high correlations are unlikely from internal variability alone



(Klavans et al. 2021)

Ensemble means highly correlated with obs.

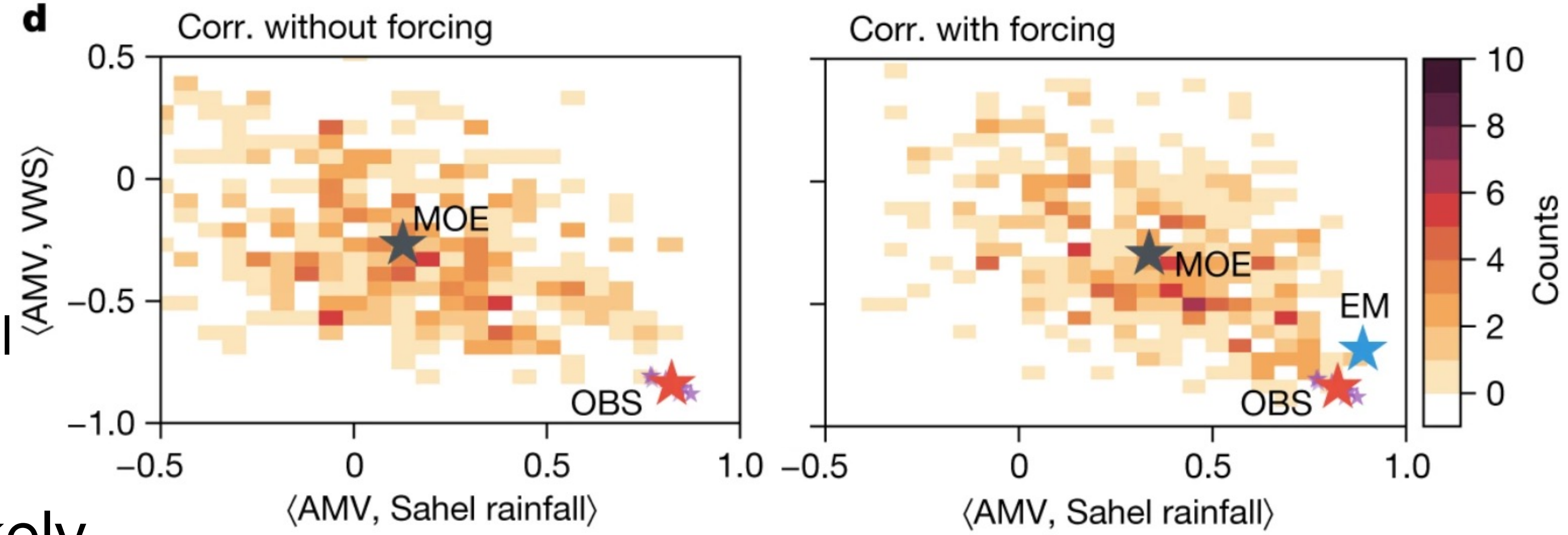
- Ensemble mean indices are surprisingly highly correlated with observations
  - AMV ($R^2 \approx 0.75$)
  - NAO ($R^2 \approx 0.60$)
  - **PDO ($R^2 \approx 0.50$)**
  - Still a role for internal variability!
- And these high correlations are unlikely from internal variability alone



(Klavans et al. *In Review*)

Correlations unlikely from internal variability alone

- Ensemble mean indices are surprisingly highly correlated with observations
  - AMV ($R^2 \approx 0.75$)
  - NAO ($R^2 \approx 0.60$)
  - PDO ($R^2 \approx 0.50$)
  - Still a role for internal variability!
- And these high correlations are unlikely from internal variability alone

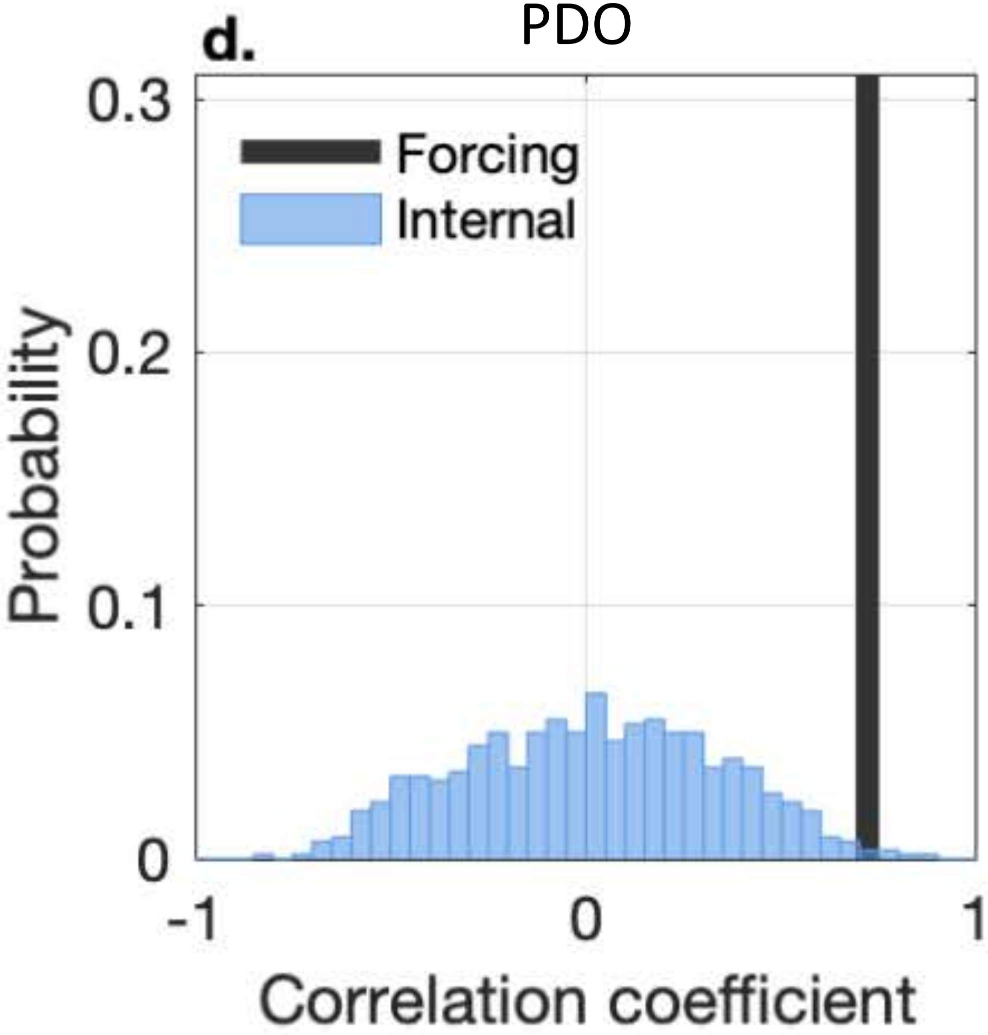


(He et al. 2023)

Correlations unlikely from internal variability alone

- Ensemble mean indices are surprisingly highly correlated with observations
  - AMV ($R^2 \approx 0.75$)
  - NAO ($R^2 \approx 0.60$)
  - **PDO ($R^2 \approx 0.50$)**
  - Still a role for internal variability!
- And these high correlations are unlikely from internal variability alone

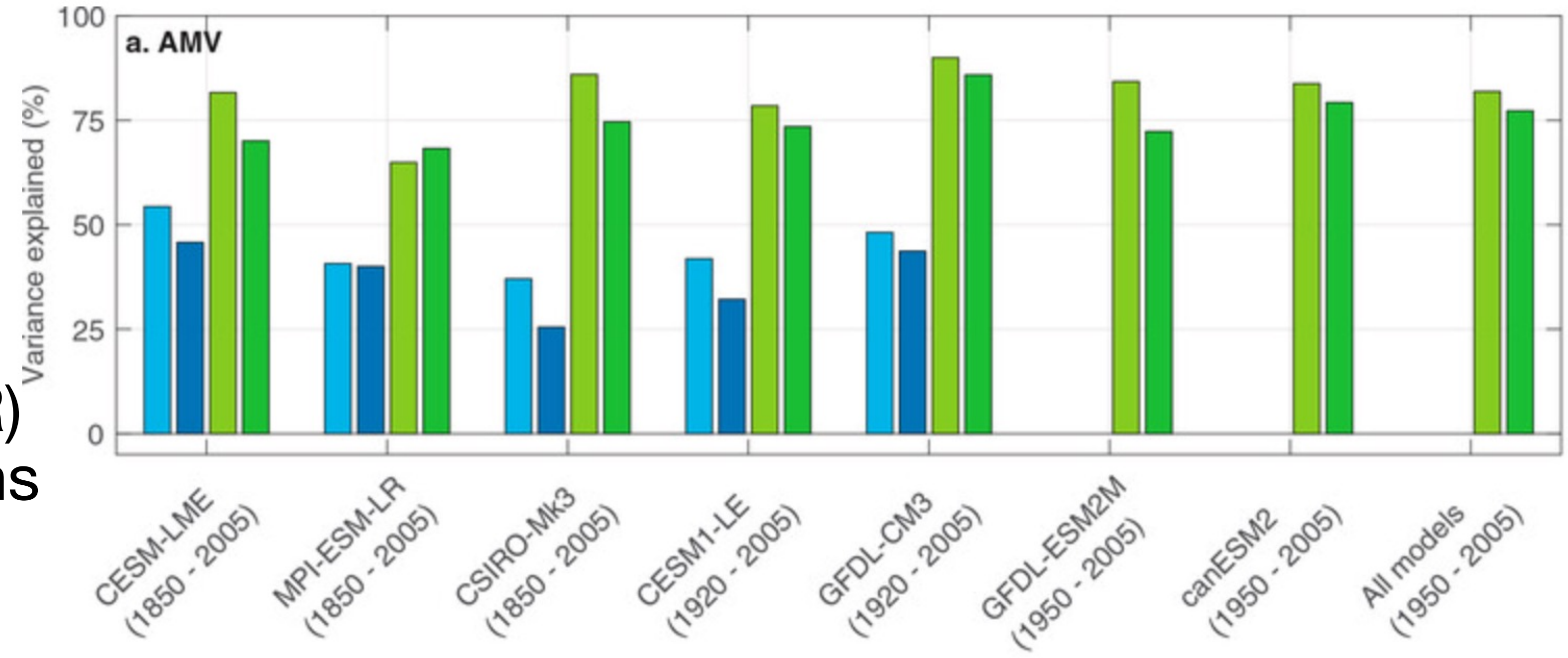
PDO



(Klavans et al. *In Review*)

Assorted evidence for the role of forcing

- Role of forcing increases as anthropogenic forcing increases (AMV, PDO, NAO)
  - Correlation
  - Variance
- Single-forcing (GHG, AER) runs have high correlations during appropriate time periods
- Ensemble mean spatial patterns bear strong similarity to observations (not always perfect)



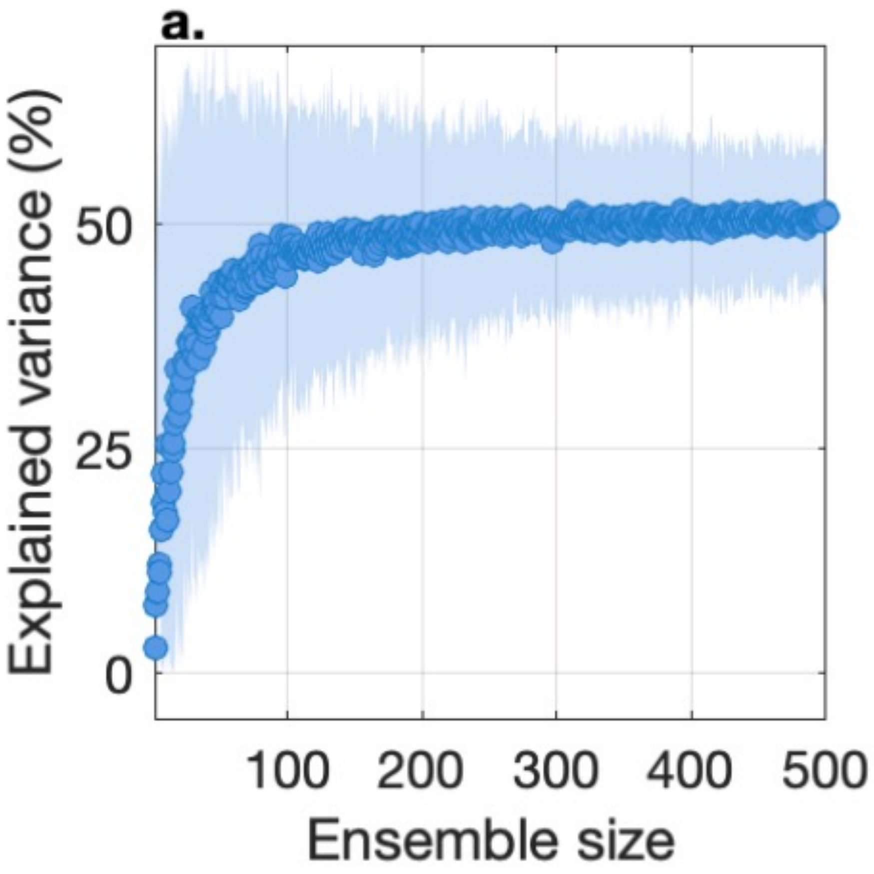
1850/1920 – 2005

1950 - 2005

(Klavans et al. 2022)

The need for large ensembles

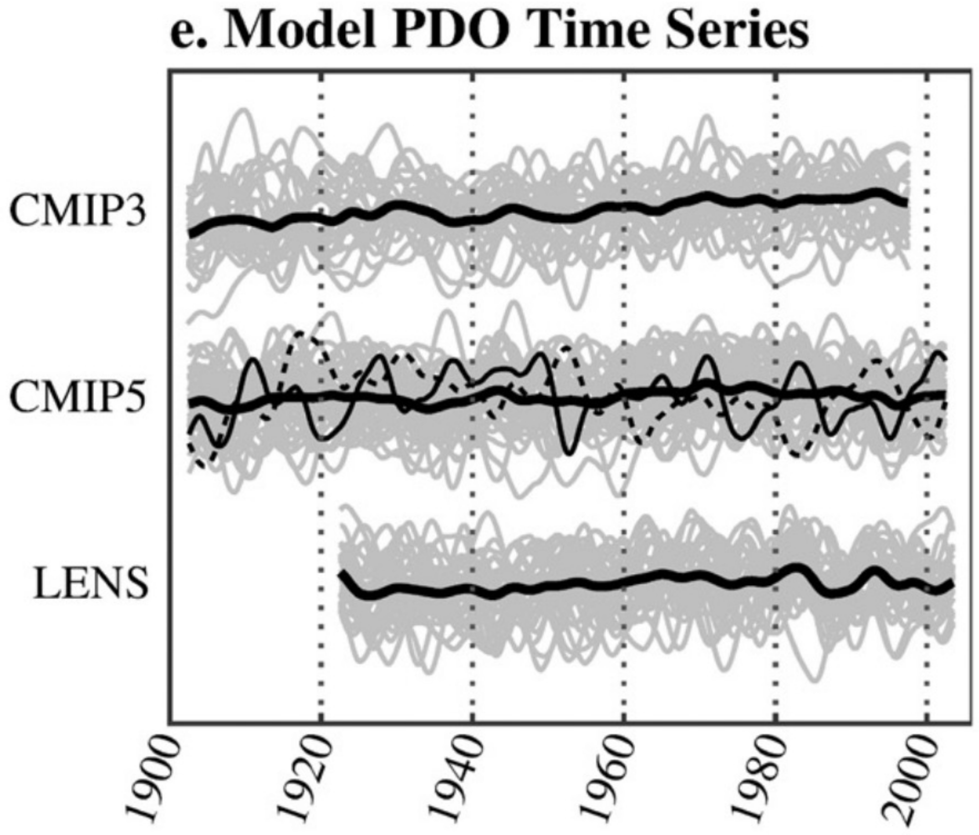
- These high correlations could only be unearthed with very large ensembles of climate models
  - **PDO**, NAO, AMV, impacts
  - More members to isolate NAO, fewer to isolate AMV
- Amplitude of the forced signal is too small, overwhelmed by internal noise



(Klavans et al. *In Review*)

The need for large ensembles

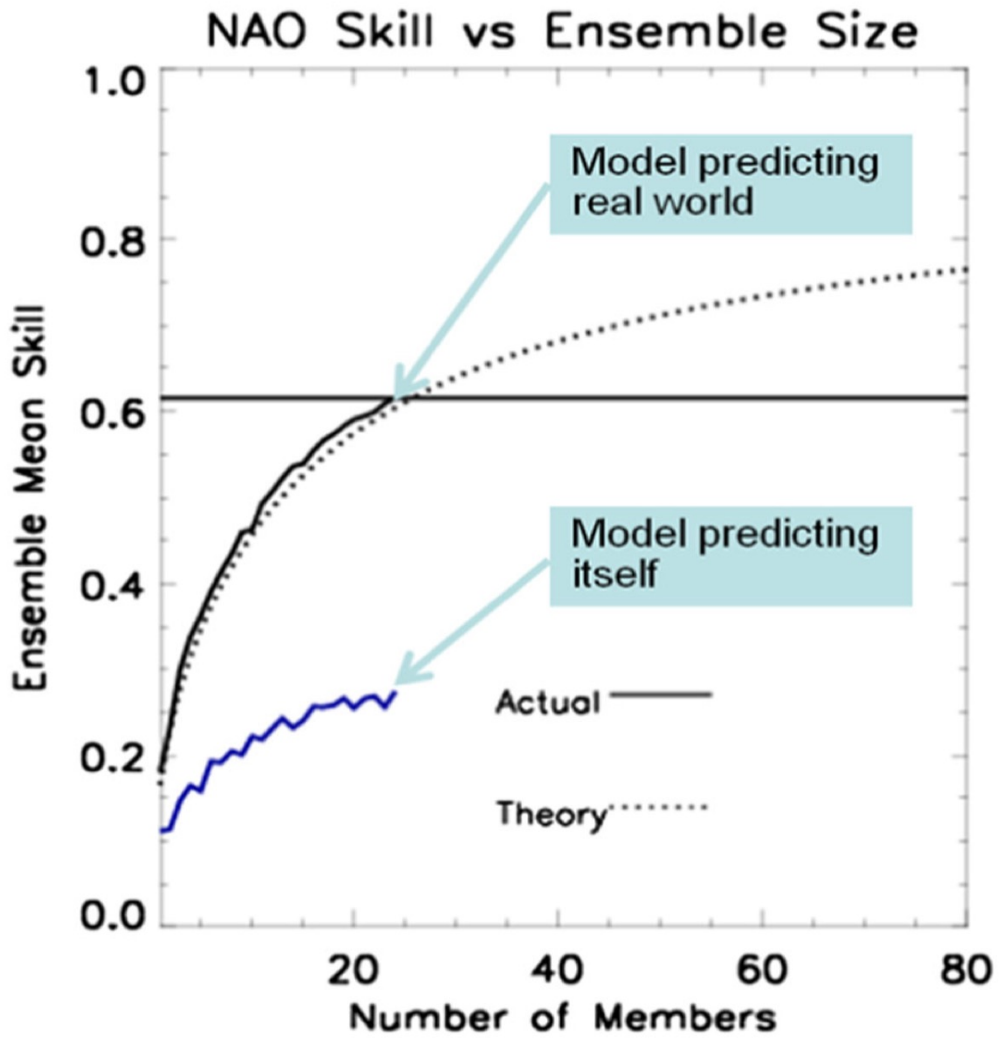
- Forced amplitude is so weak it was very easy to overlook
- Example using the PDO index:
  - In a 40-member ensemble, the R^2 from 1920 – 2005: 5%
  - In a 472-member ensemble, the R^2 from 1950 – 2005: 53%
- The externally forced signal to internally generated noise ratios in models is too weak



(Newman et al. 2016)

Aside: the signal-to-noise paradox

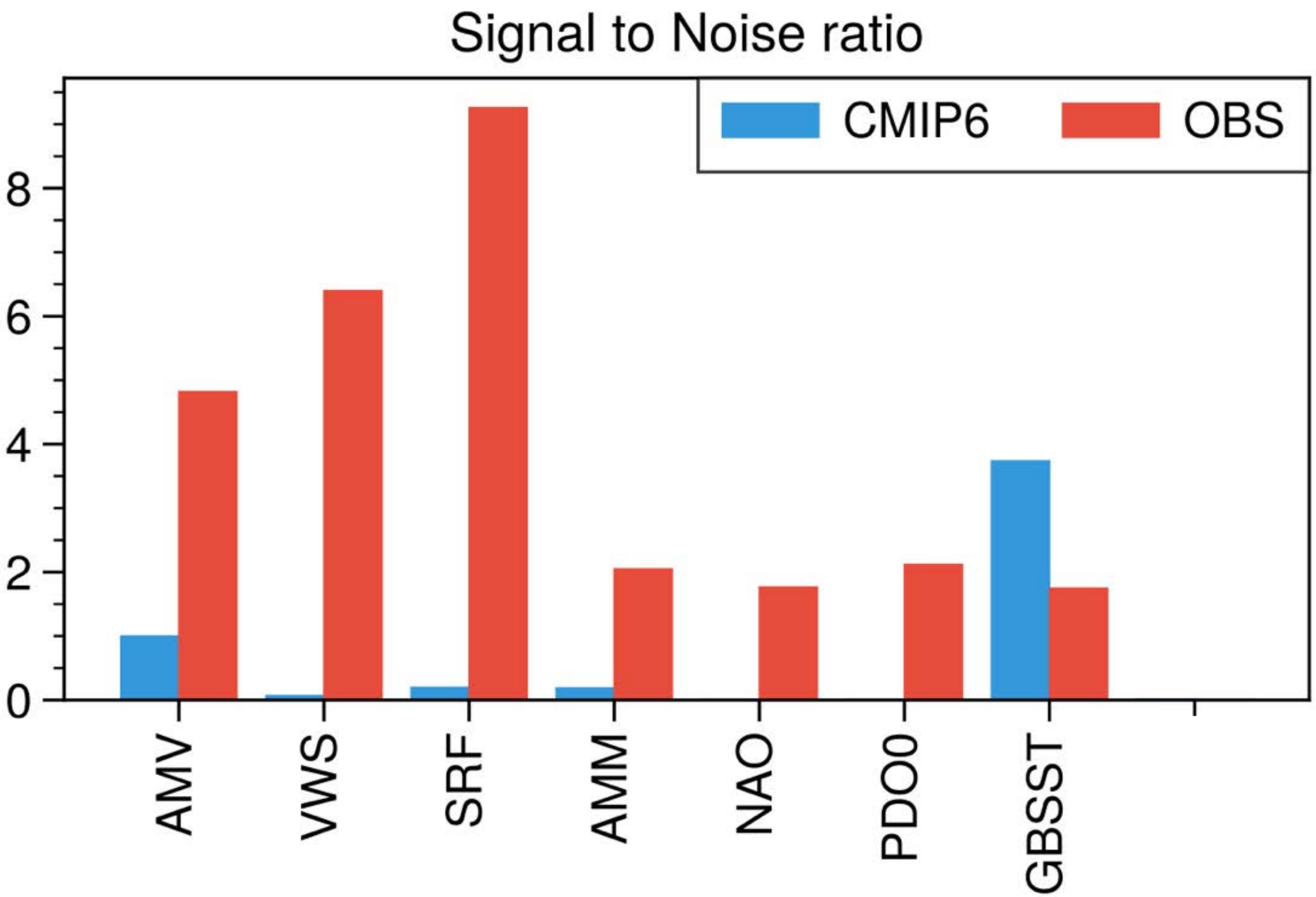
- What we've presented so far is a narrowing of the signal-to-noise paradox (Eade et al. 2014; Dunstone et al. 2016; Smith et al. 2020; and many others)
- Comparing initialized and uninitialized ensembles shows the error at decadal signal primarily associated with the forced signal (Klavans et al. 2022)



(Scaife and Smith 2018)

The signal-to-noise ratio in models is too low

- We can estimate the signal-to-noise ratio in models and observations
- Large ensembles
  - Ensemble mean / internal variability
- Observations
  - $OBS = \beta_1 EM + \varepsilon$
- Could observed internal variability be correlated with the forced signal by chance? Unlikely across many pseudo-independent modes



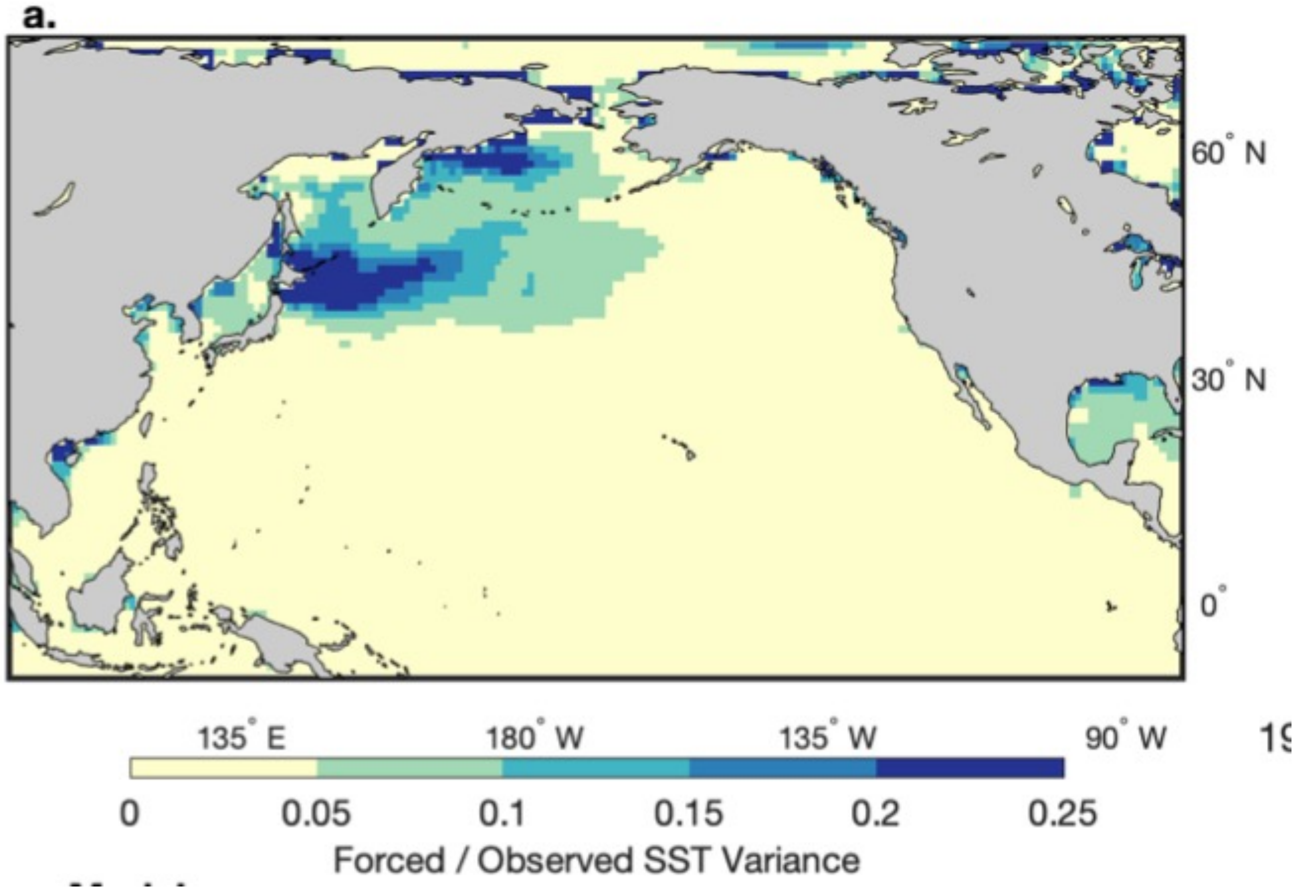
(Chengfei He - does good work, go to his poster!)

Summary

1. What are models doing well?
   - Models produce multidecadal modes with reasonably realistic spatial patterns and impacts
2. Where can models improve?
   - Observed variance is an outlier, relative to model ensemble spread
   - Ensemble mean is highly correlated with observations – but it's amplitude is too weak
   - The forced signal to internally generated noise ratio is too low in models
3. How can the S/N ratio be improved? What could rectifying this error teach us?

How can the S/N ratio be improved?

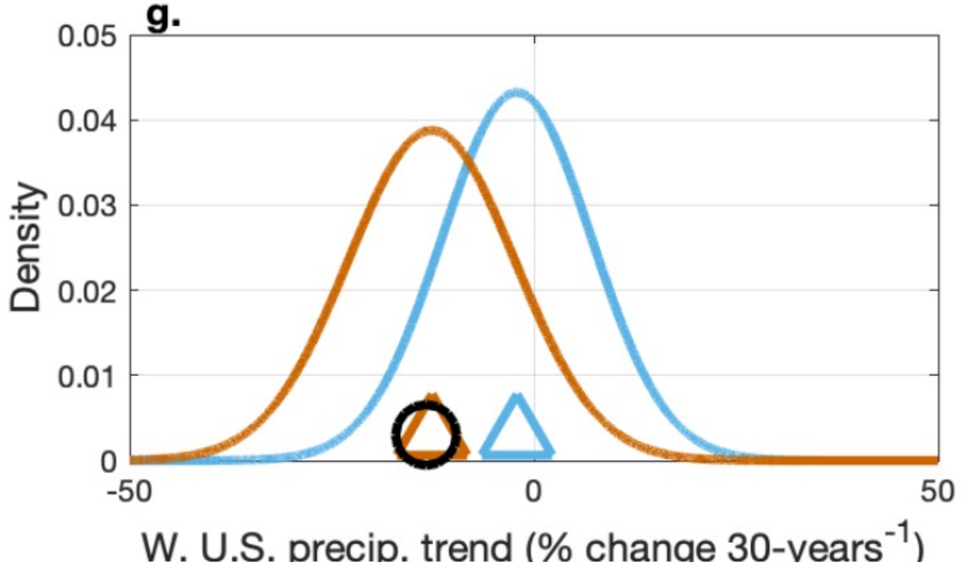
- Some proposed causes of the S/N error:
  - Air-sea coupling (Smirnov et al. 2015; Kim et al. 2018)
  - Upper ocean damping (Murphy et al. 2021)
  - Model resolution (Scaife et al. 2019)
  - Ocean front resolution (Kirtman et al. 2017)

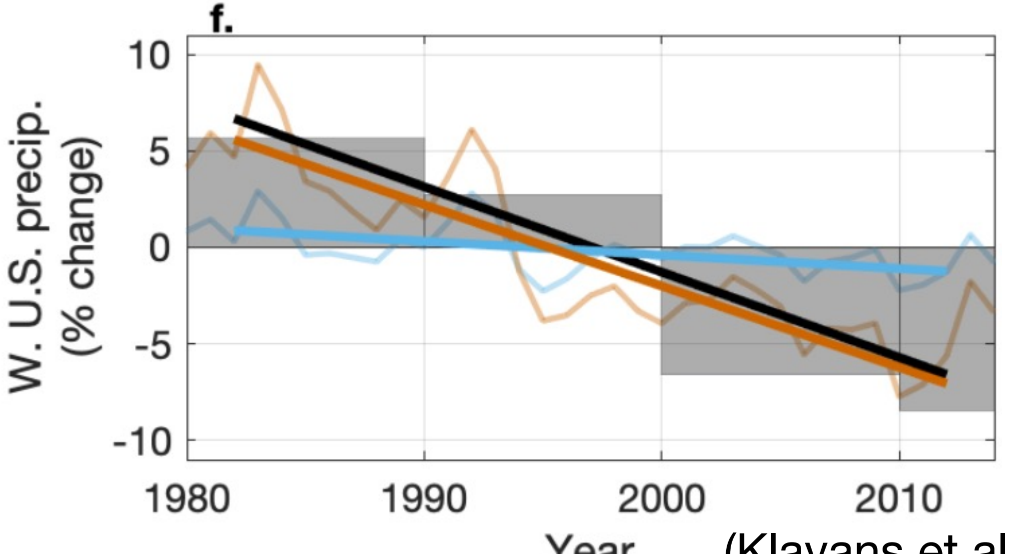


(Klavans et al. *In Review*)

What could rectifying this error teach us?

- Models may be underestimating climate risk while overestimating uncertainty
- Increasing the S/N ratio to match observations suggests that observations are an average response
- External forcing is predictable in the near-term

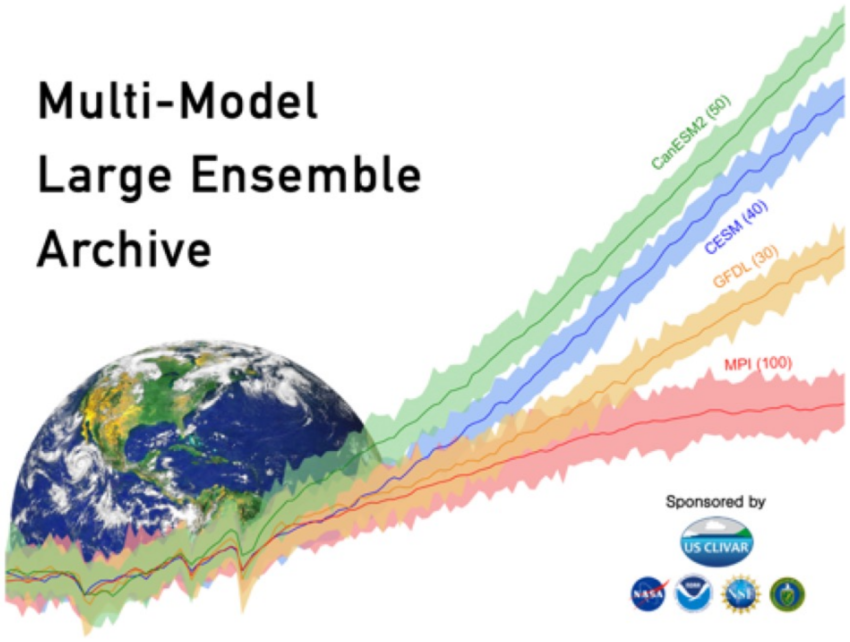




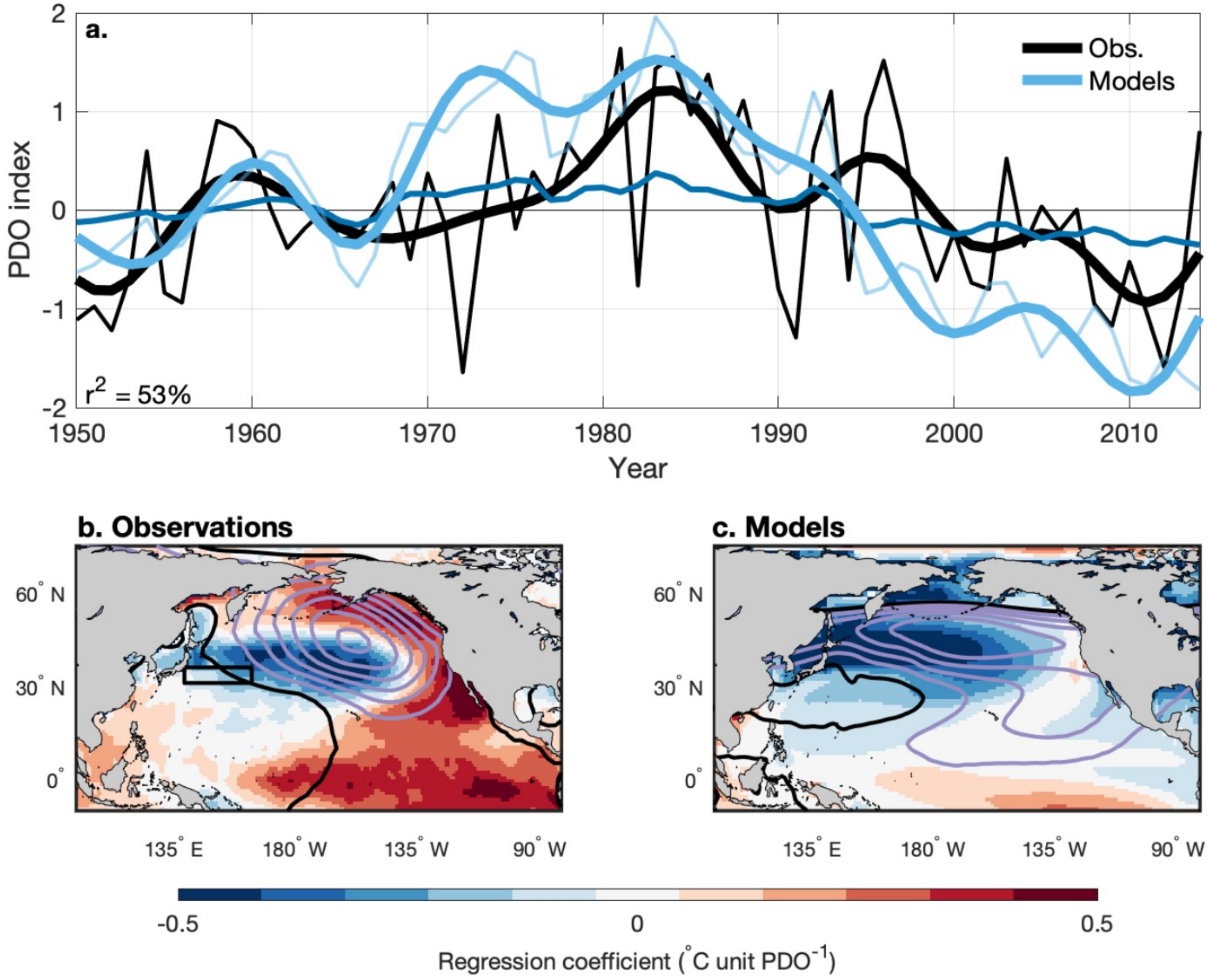
(Klavans et al. *In Review*)

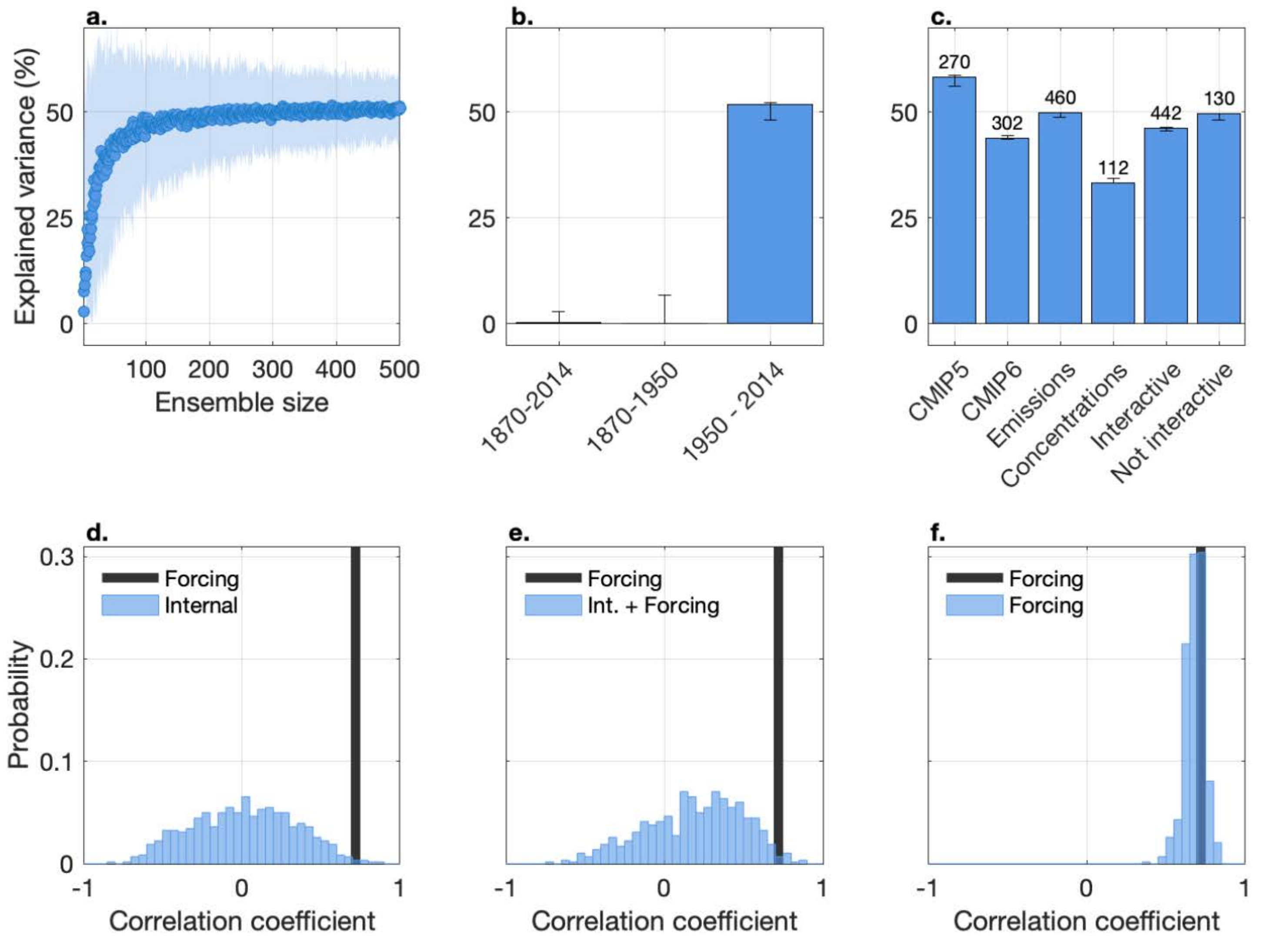
What could rectifying this error teach us?

- Paleoclimate: Is climate model response to solar/orbital forcing too weak? (Victoria Todd and Tim Shanahan)
- Internal variability: how does internal noise change when the signal increases? Some evidence signal and noise are additive
- Until the S/N error in models is fixed, large ensembles are a required tool for understanding regional climate change

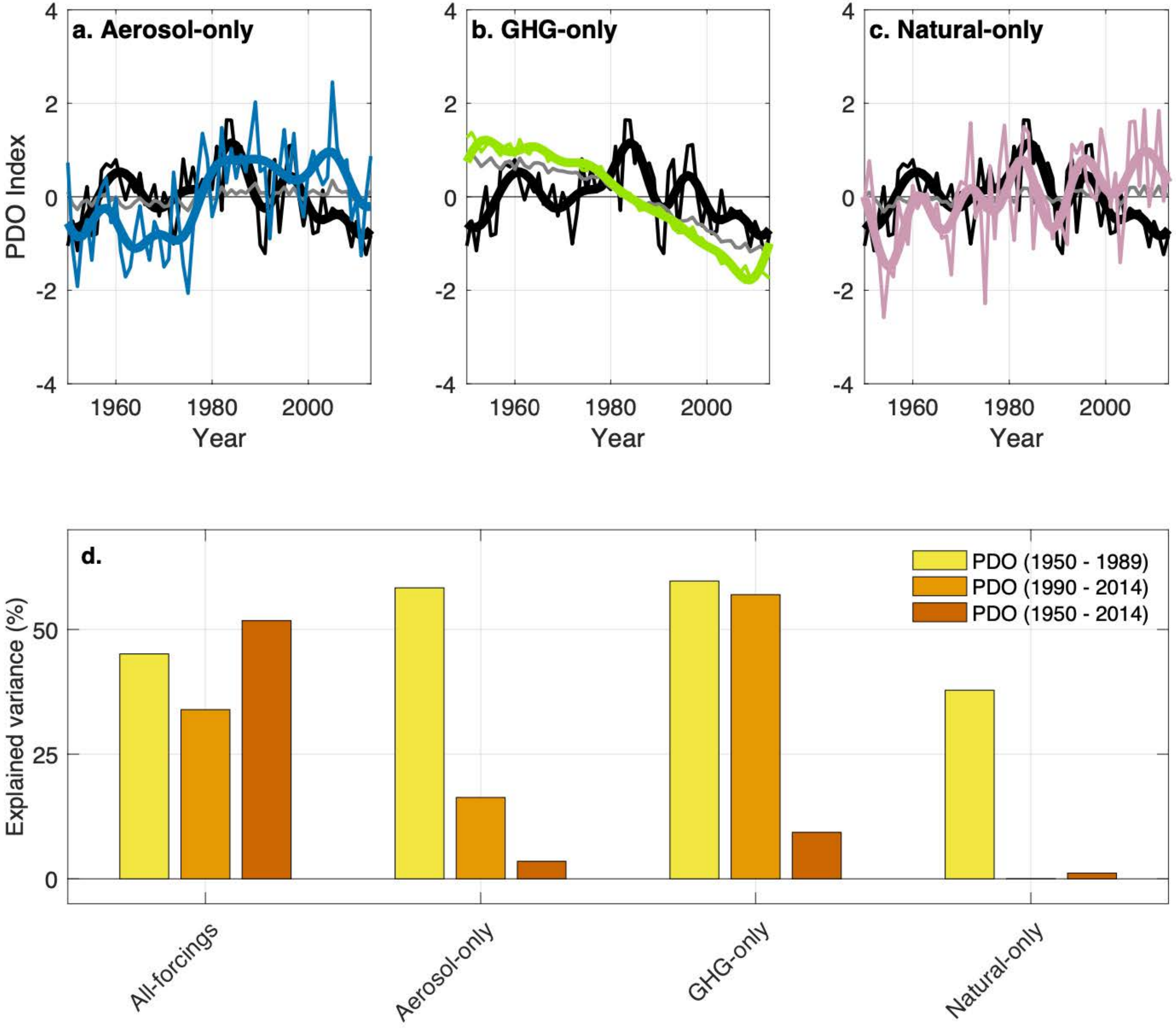


More slides!


a.
Obs.
Models
PDO index
Year
r² = 53%
2
1
0
-1
-2
1950
1960
1970
1980
1990
2000
2010
b. Observations
c. Models
60° N
30° N
0°
135° E
180° W
135° W
90° W
-0.5
0
0.5
Regression coefficient (°C unit PDO⁻¹)

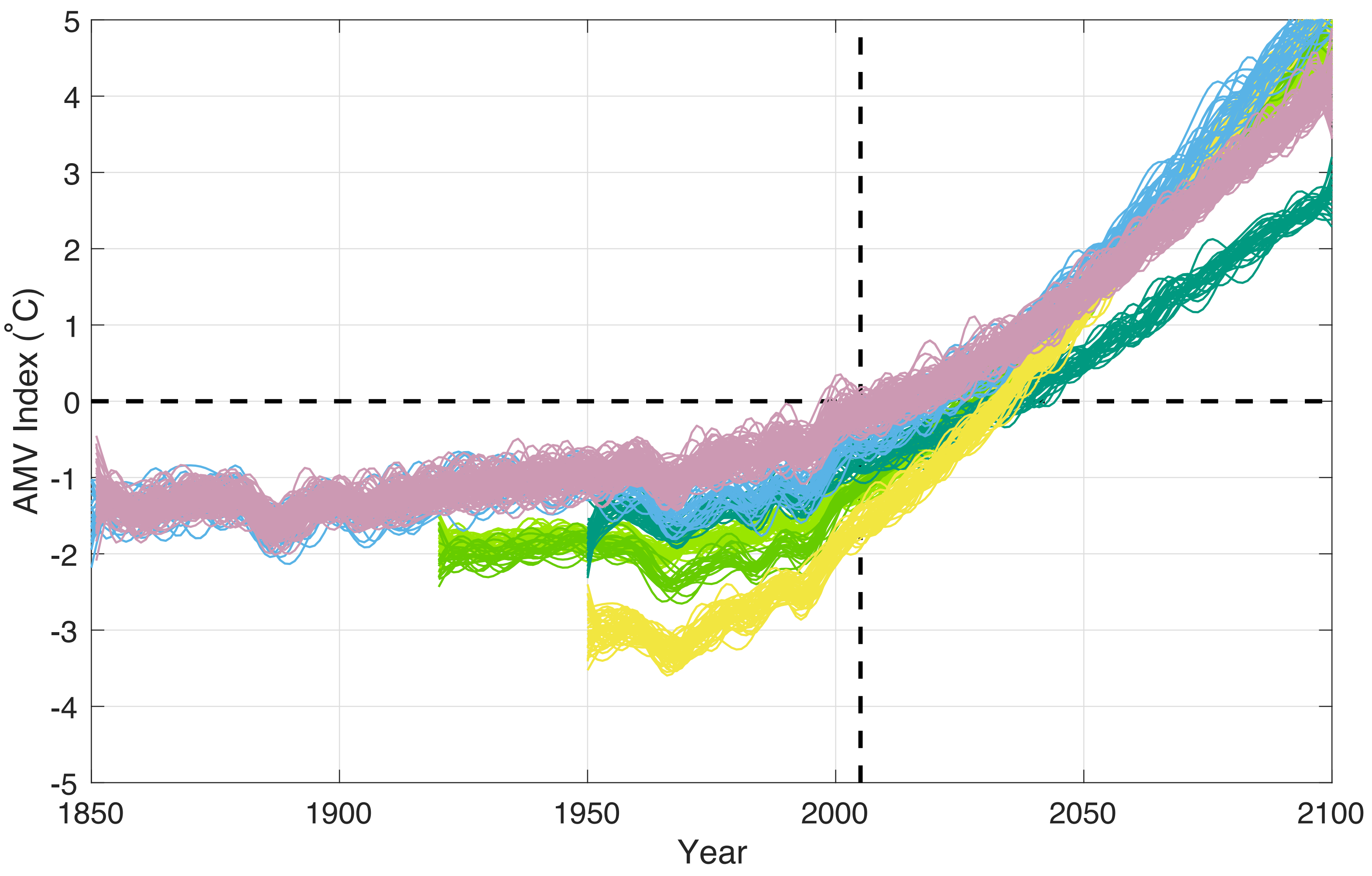


a.
Explained variance (%)
Ensemble size
b.
1870-2014
1870-1950
1950 - 2014
c.
270
302
460
112
442
130
CMIP5
CMIP6
Emissions
Concentrations
Interactive
Not interactive
d.
Forcing
Internal
Probability
Correlation coefficient
e.
Forcing
Int. + Forcing
Correlation coefficient
f.
Forcing
Forcing
Correlation coefficient

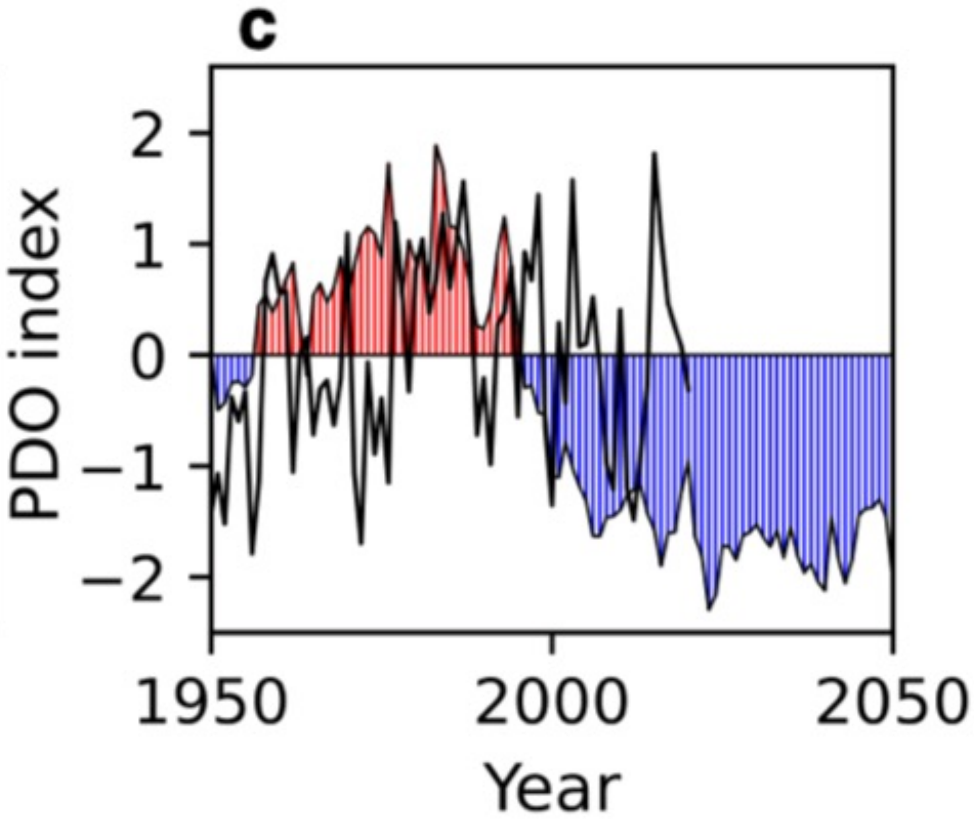


a. Aerosol-only
b. GHG-only
c. Natural-only
PDO Index
4
2
0
-2
-4
1960
1980
2000
Year
d.
PDO (1950 - 1989)
PDO (1990 - 2014)
PDO (1950 - 2014)
Explained variance (%)
50
25
0
All-forcings
Aerosol-only
GHG-only
Natural-only

Future AMV

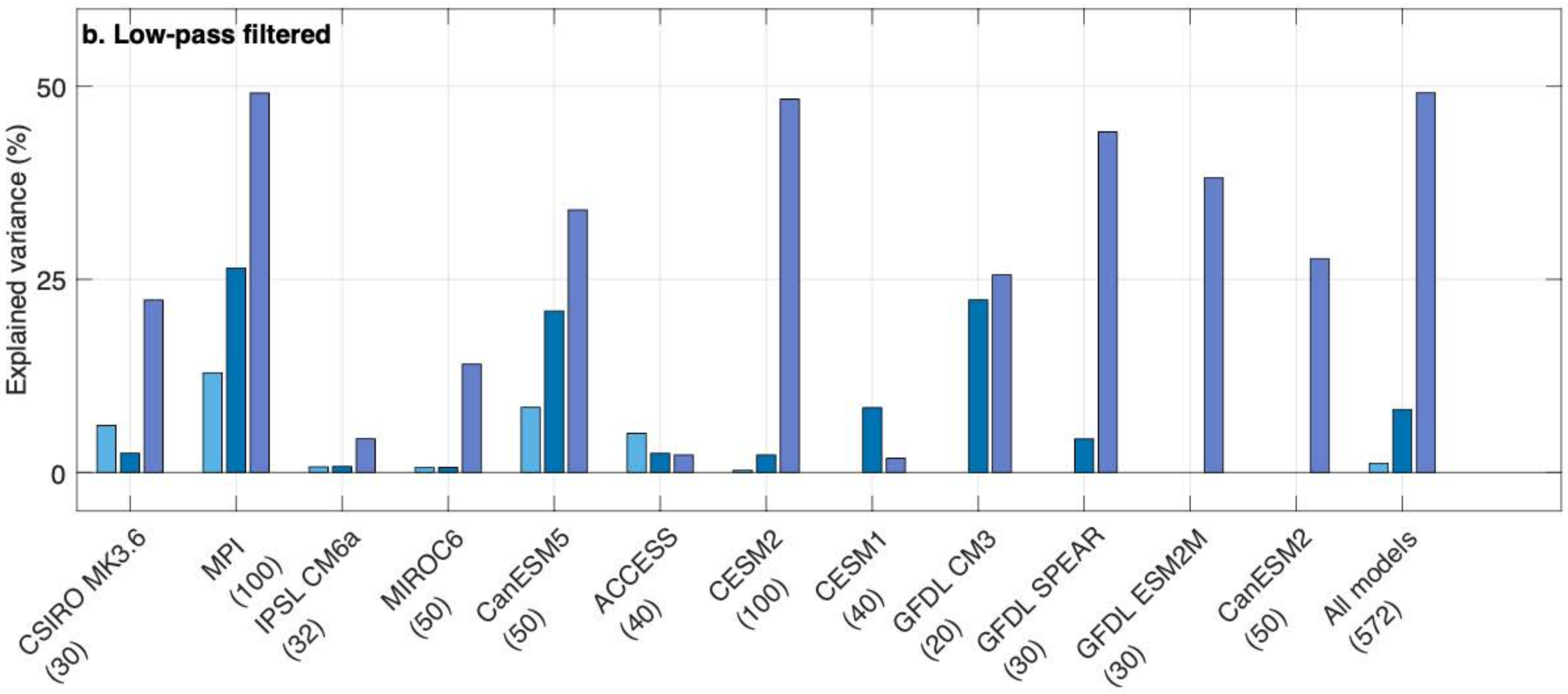


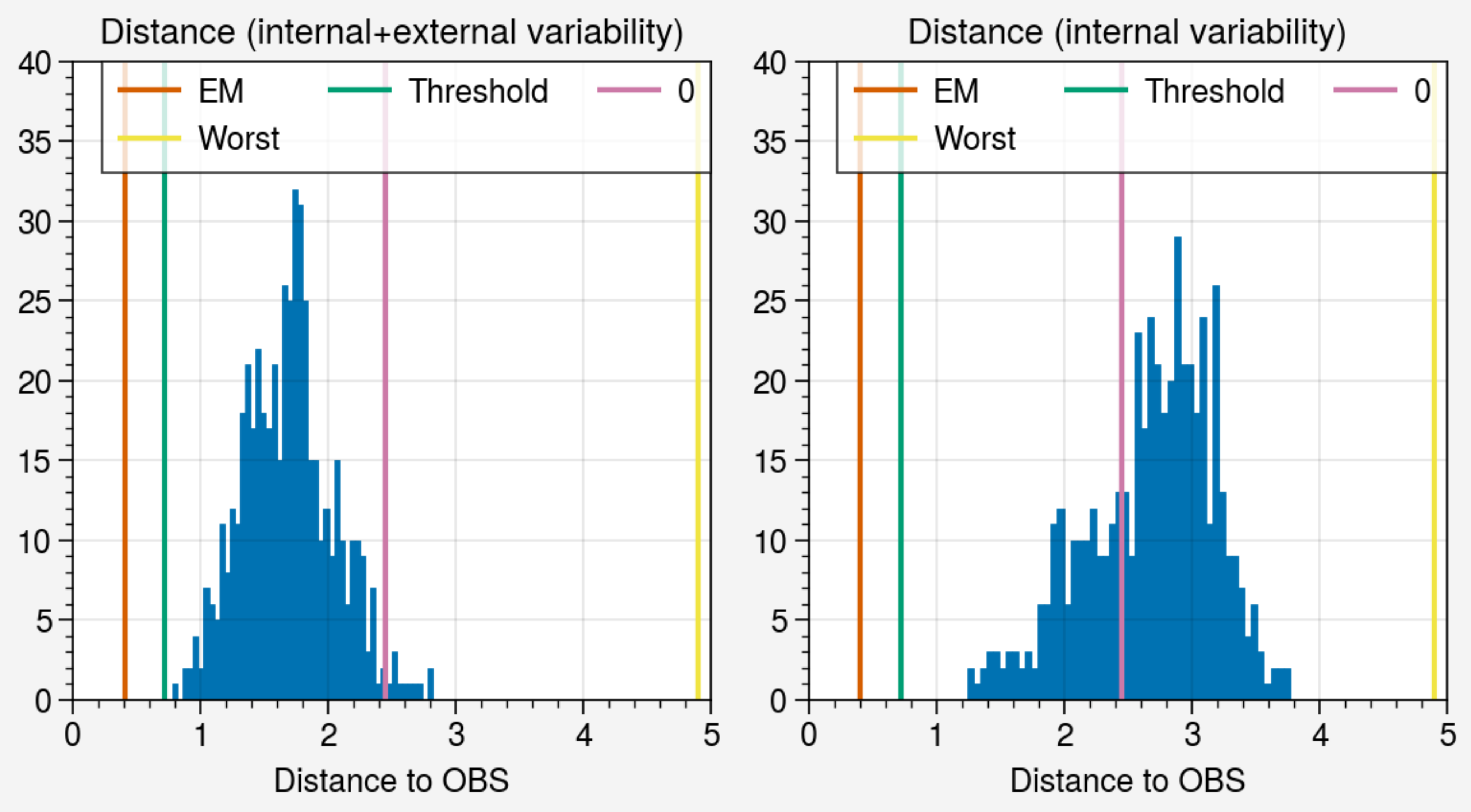
Future PDO



PDO Exp. Var

1850 – 2014, 1920 – 2014, 1950 - 2014




Distance (internal+external variability)
EM
Threshold
0
Worst
Distance to OBS
Distance (internal variability)
EM
Threshold
0
Worst
Distance to OBS