Challenges in comparing observed and model-simulated trends on regional scales

Clara Deser, NCAR

US CLIVAR Workshop “Confronting Earth System Model Trends with Observations” March 13-15, 2024.

Challenges in comparing observed and model-simulated trends on regional scales

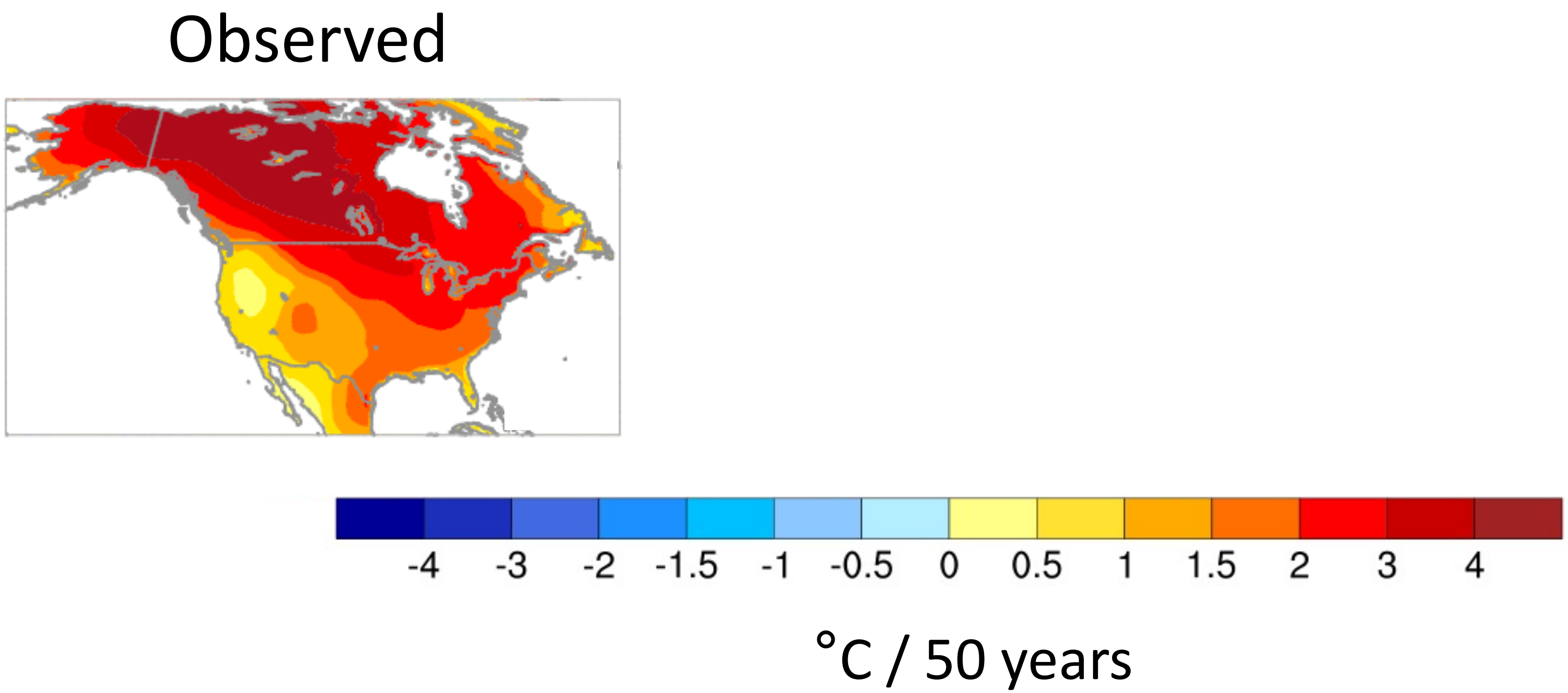
Clara Deser, NCAR

Assessment of model realism needed for credible predictions and projections, and for informing understanding of observed trends.

US CLIVAR Workshop “Confronting Earth System Model Trends with Observations” March 13-15, 2024.

In the old days ...

Winter Air Temperature Trends (1963-2012)

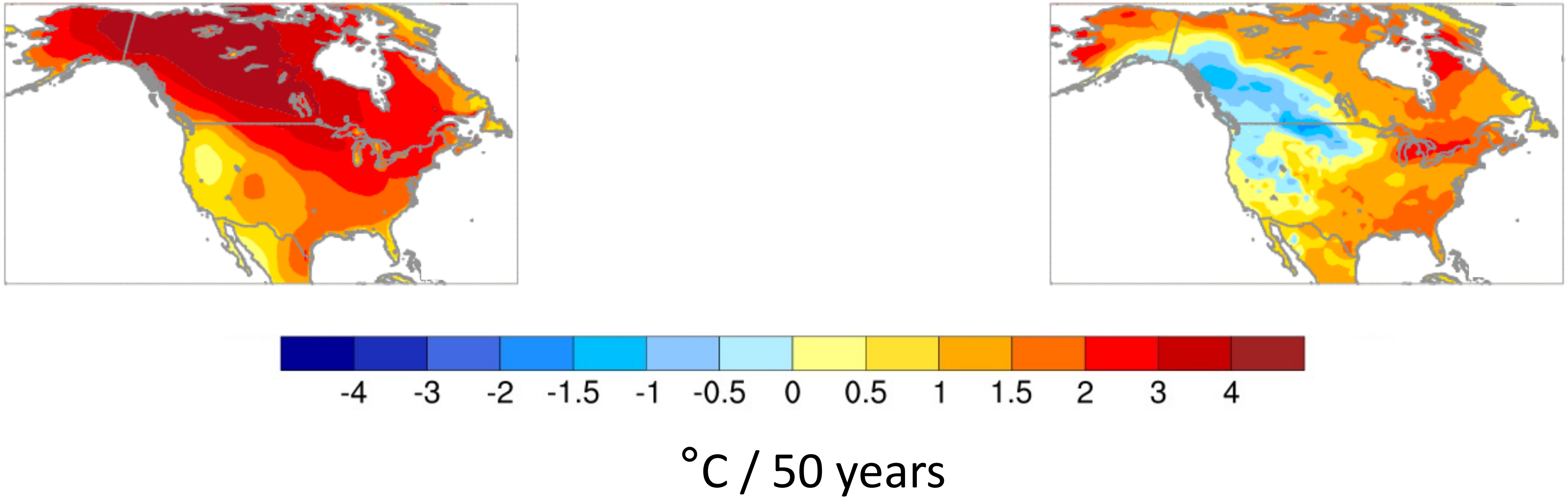


Deser et al., J. Climate 2016

Winter Air Temperature Trends (1963-2012)

Observed

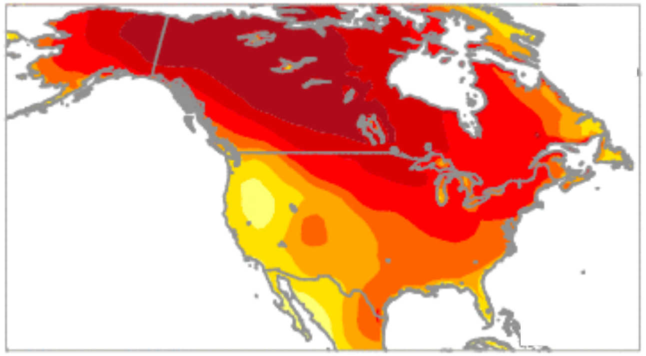
Model simulation



Deser et al., J. Climate 2016

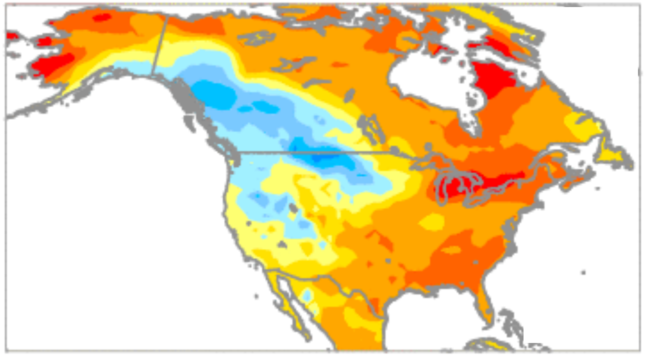
Winter Air Temperature Trends (1963-2012)

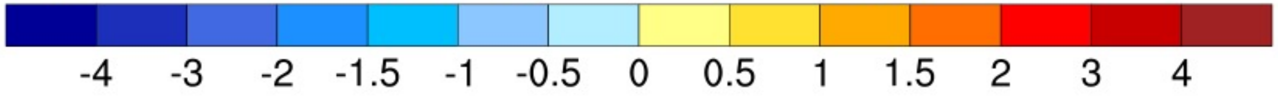
Observed



- **Model and/or forcing is erroneous.**

Model simulation





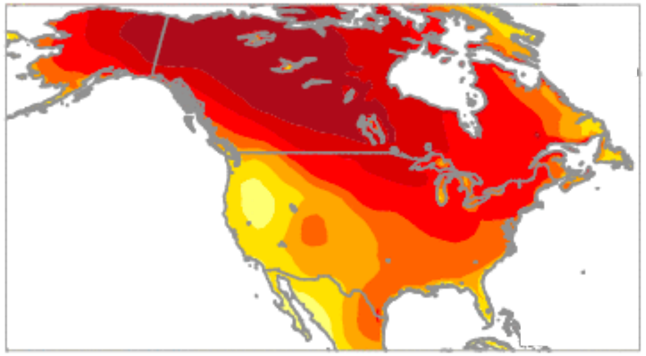
°C / 50 years

Deser et al., J. Climate 2016

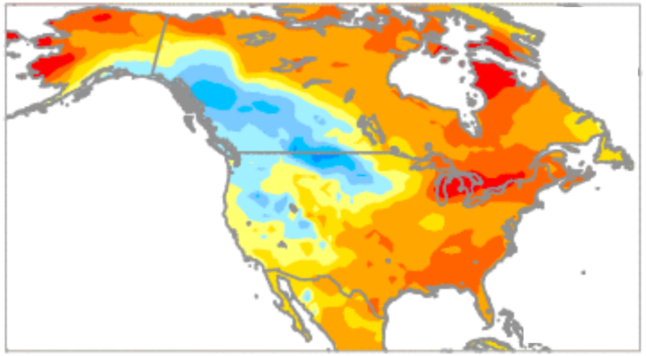
Nowadays ...

Winter Air Temperature Trends (1963-2012)

Observed



Model simulation

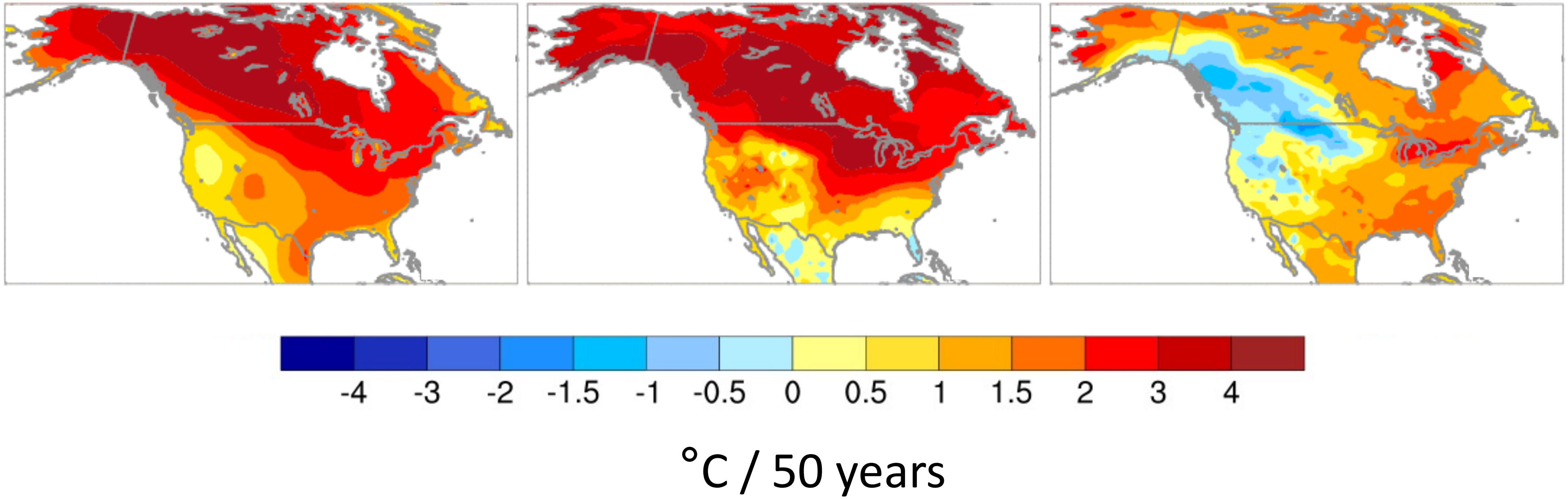


°C / 50 years

Deser et al., J. Climate 2016

Winter Air Temperature Trends (1963-2012)

Observed | Another simulation of the same model | Model simulation



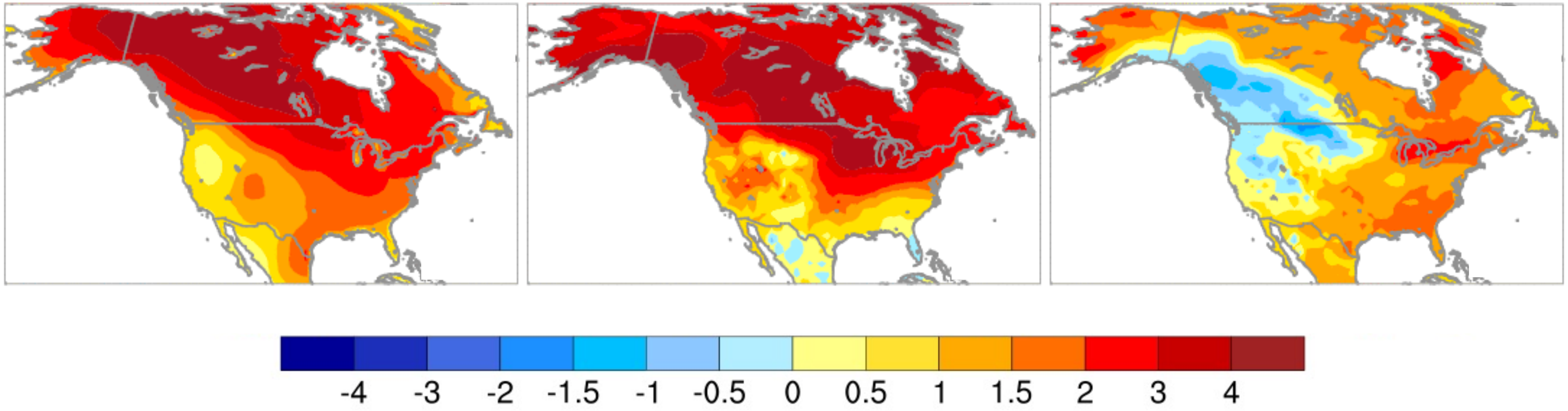
Deser et al., J. Climate 2016

Winter Air Temperature Trends (1963-2012)

Observed

Another simulation of the same model

Model simulation



°C / 50 years

Internal

+

Forced

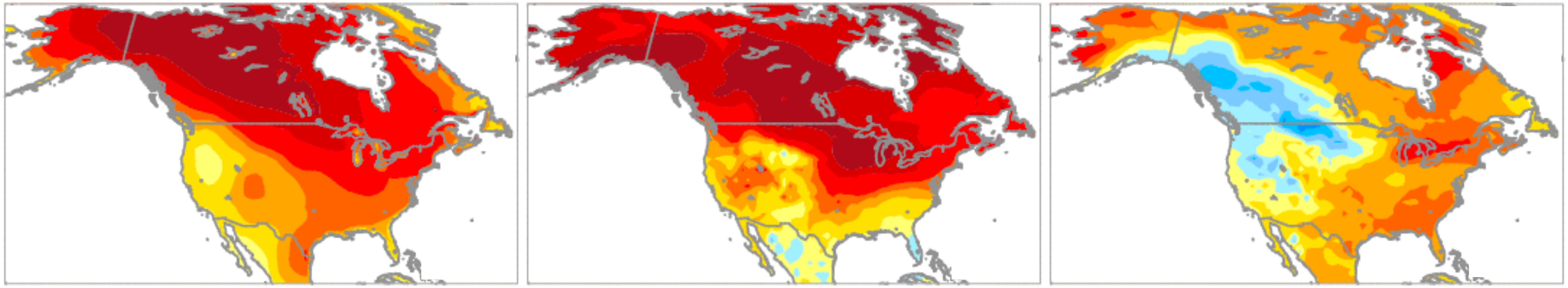
Deser et al., J. Climate 2016

Winter Air Temperature Trends (1963-2012)

Observed

Another simulation of the same model

Model simulation



- **Is the model range plausible?**
- **How should we evaluate models given a single observed outcome?**
- **Could another reality have been possible?**

Internal

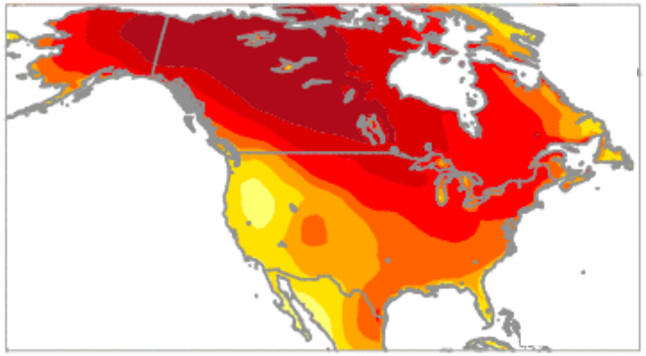
+

Forced

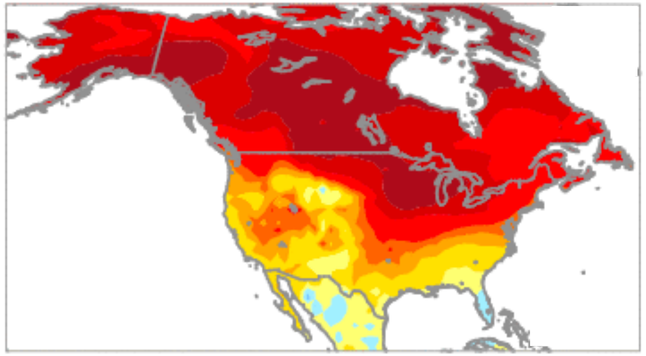
Deser et al., J. Climate 2016

Winter Air Temperature Trends (1963-2012)

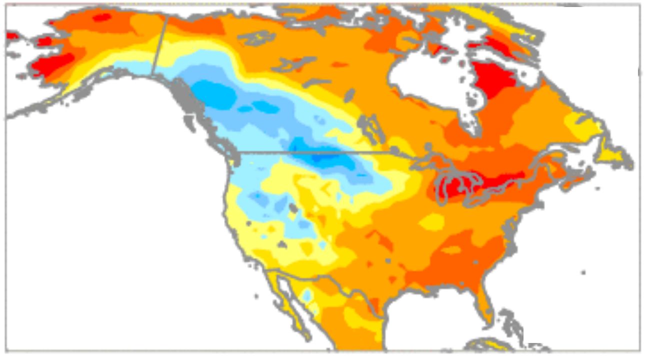
Observed



Another simulation of the same model



Model simulation



- **Is the model range plausible?**
- **How should we evaluate models given a single observed outcome?**
- **Could another reality have been possible?**

➢ **New approaches and continued challenges.**

Internal

+

Forced

Deser et al., J. Climate 2016

Is the model range plausible?

Requires assessing both the model's forced response and its internal variability.

Is the model range plausible?

Requires assessing both the model's forced response and its internal variability.

- **Model's forced response:** average across all members of a "Single Model Initial-Condition Large Ensemble" (SMILE).
- **Model's internal variability:** residual from the forced response in each member.

Is the model range plausible?

Requires assessing both the model's forced response and its internal variability.

- **Model's forced response:** average across all members of a "Single Model Initial-Condition Large Ensemble" (SMILE).
- **Model's internal variability:** residual from the forced response in each member.

- **Observed forced response:** statistical & dynamical methods.

Is the model range plausible?

Requires assessing both the model's forced response and its internal variability.

- **Model's forced response:** average across all members of a "Single Model Initial-Condition Large Ensemble" (SMILE).
- **Model's internal variability:** residual from the forced response in each member.

- **Observed forced response:** statistical & dynamical methods.
  ForceSMIP workshop & hackathon *(Aug 2023, NCAR and ETH)*.
  Extensive comparison of existing methods and development of new ones, using SMILEs as a testbed; double-blind test of the methods, with application to observations.
  (See Robb Wills' poster on Friday.)

Is the model range plausible?

Requires assessing both the model's forced response and its internal variability.

- **Model's forced response:** average across all members of a "Single Model Initial-Condition Large Ensemble" (SMILE).
- **Model's internal variability:** residual from the forced response in each member.

- **Observed forced response:** statistical & dynamical methods. ***ForceSMIP workshop & hackathon*** (Aug 2023, NCAR and ETH).
- **Observed internal variability:** residual from forced response. *Limited to one realization: we don't know where the observed trend lies in the distribution of possible realities.*

How should we evaluate model trends given a single observed outcome?

How should we evaluate model trends given a single observed outcome?

- Determine whether the observed trend lies within the spread of a given SMILE. However, this could be for the "wrong reason": e.g., spread due to internal variability might be too large. Relatedly, the forced response and associated signal-to-noise ratio might be too low.

How should we evaluate model trends given a single observed outcome?

- Determine whether the observed trend lies within the spread of a given SMILE. However, this could be for the "wrong reason": e.g., spread due to internal variability might be too large. Relatedly, the forced response and associated signal-to-noise ratio might be too low.
- How do we evaluate ensemble spread in a SMILE? Need to construct plausible alternative outcomes for observed trends.

Constructing plausible alternative realities (“Observational” Large Ensembles)

Constructing plausible alternative realities ("Observational" Large Ensembles)

- Infer statistical characteristics of trends of any length from the amplitude and autocorrelation properties of interannual variability (Thompson et al. 2015). Works well, but does not provide information on spatial patterns.

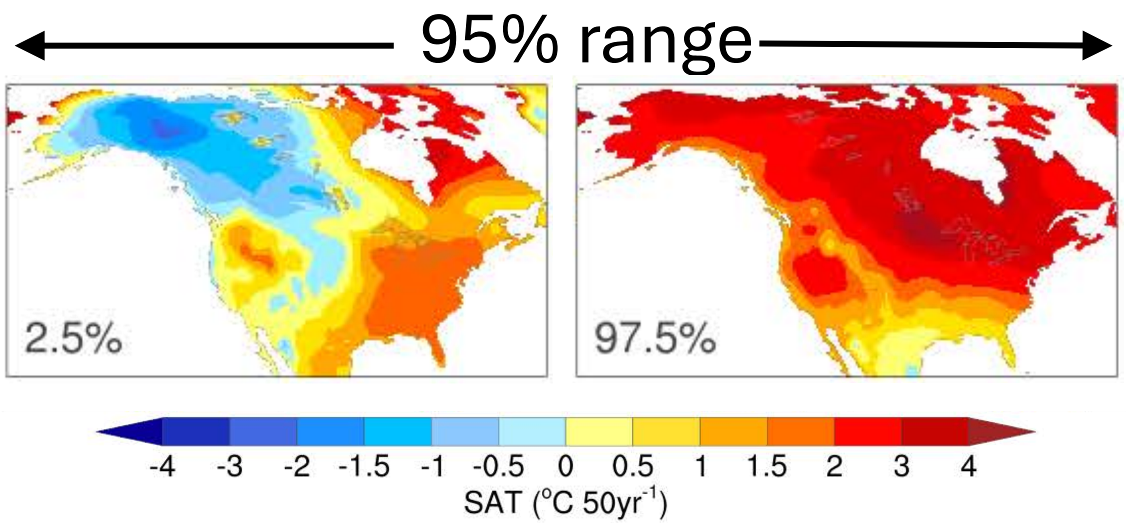
Constructing plausible alternative realities ("Observational" Large Ensembles)

- Infer statistical characteristics of trends of any length from the amplitude and autocorrelation properties of interannual variability (Thompson et al. 2015). Works well, but does not provide information on spatial patterns.
- Statistical regression onto leading modes of variability (ENSO, PDV, AMV) with shuffled phases plus block bootstrapping of the residuals (McKinnon et al.). Preliminary step requires removal of the estimated forced response (in this case, via regression onto a SMILE's forced GMST(t)).

Alternate Realities: Statistical Approach (O-LENS)

Winter Air Temperature Trends (1963-2012)

Internal component from the 1000-member “Obs Large Ensemble” added to forced component from CESM1 SMILE.

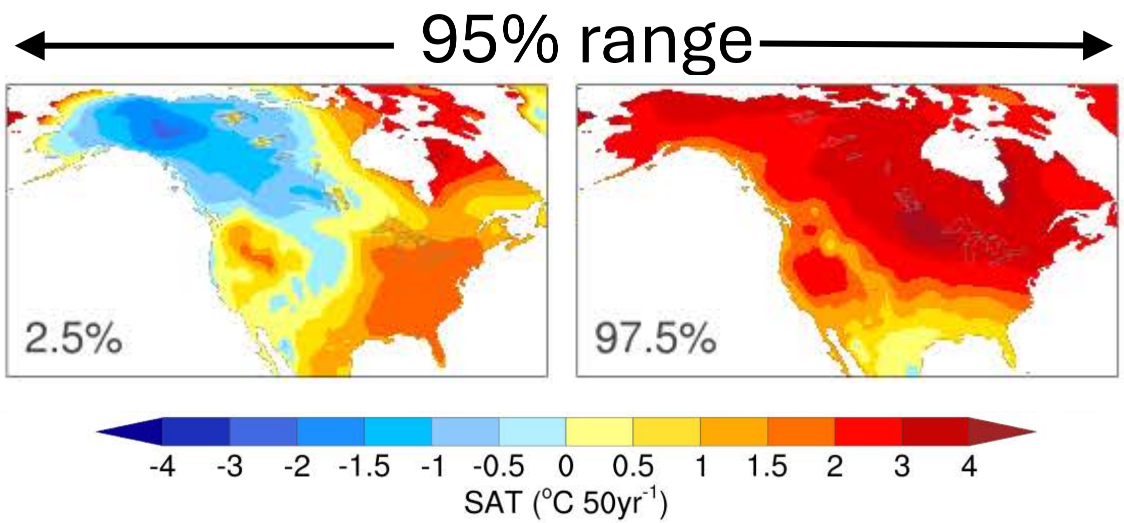


McKinnon and Deser (2018)

Alternate Realities: Statistical Approach (O-LENS)

Winter Air Temperature Trends (1963-2012)

Internal component from the 1000-member “Obs Large Ensemble” added to forced component from CESM1 SMILE.

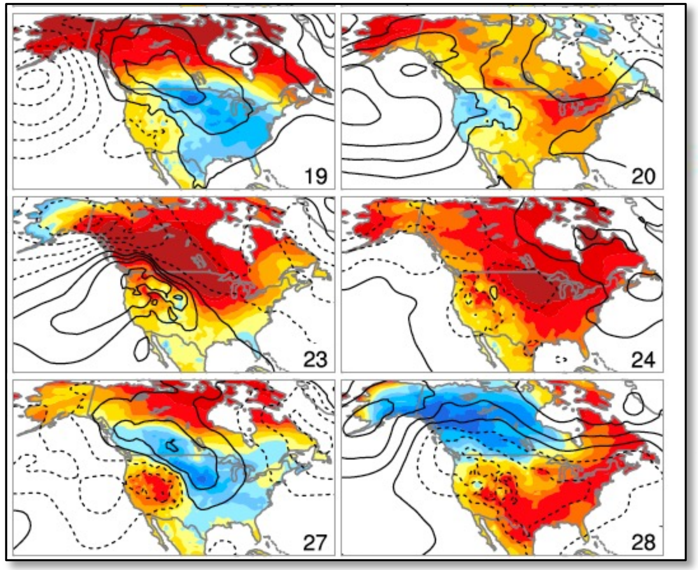


McKinnon and Deser (2018)

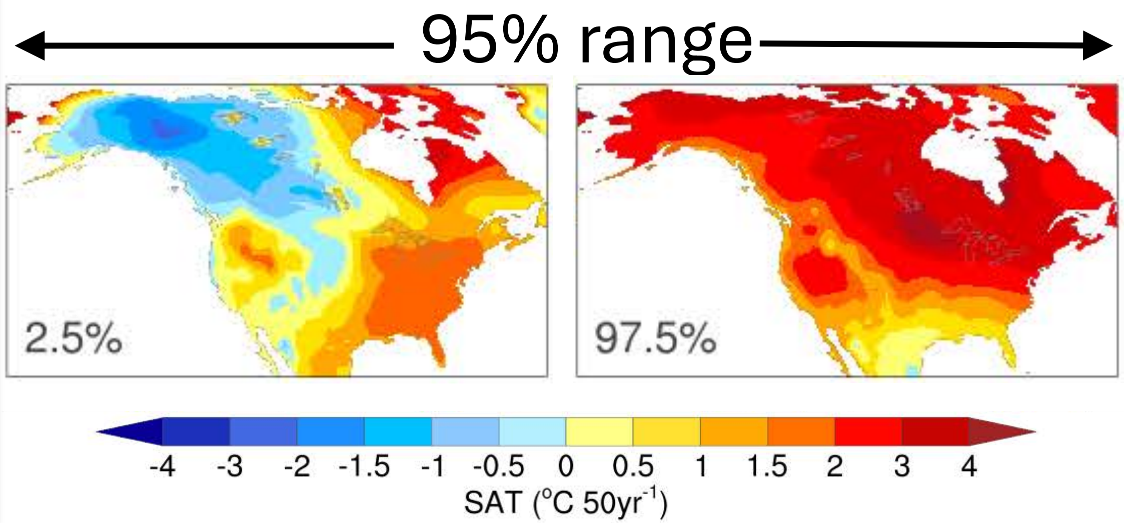
Range is very similar to that of the CESM1 SMILE.

Alternate Realities: Statistical Approach (O-LENS)

Winter Air Temperature Trends (1963-2012)



CESM1 SMILE

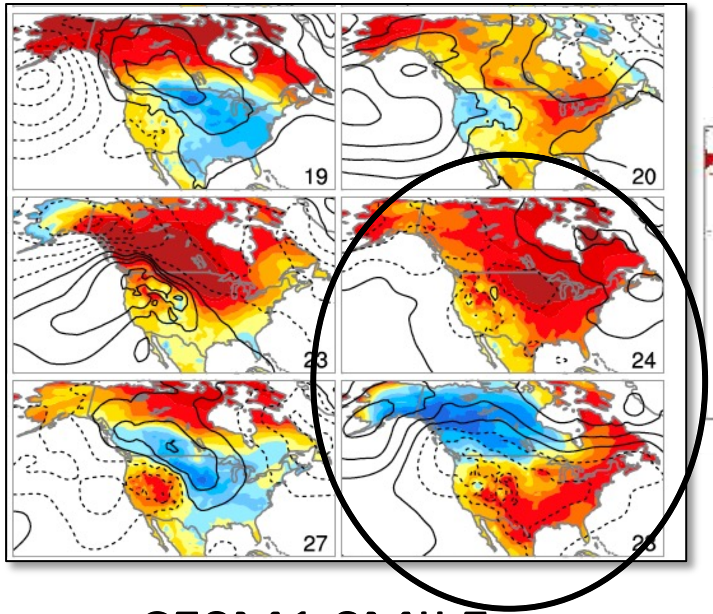


McKinnon and Deser (2018)

Range is very similar to that of the CESM1 SMILE.

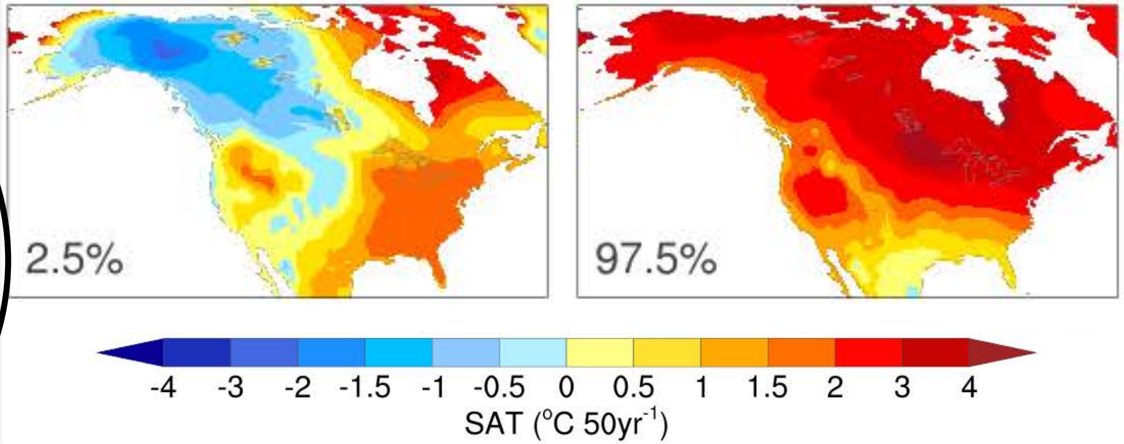
Alternate Realities: Statistical Approach (O-LENS)

Winter Air Temperature Trends (1963-2012)



CESM1 SMILE

95% range



McKinnon and Deser (2018)

Range is very similar to that of the CESM1 SMILE.

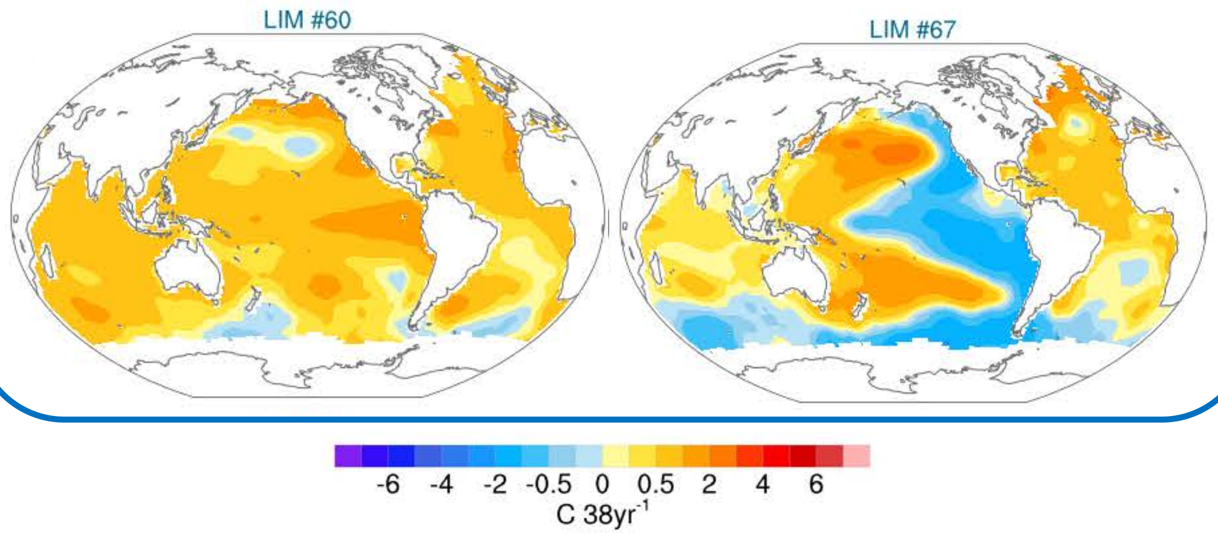
Constructing plausible alternative realities (“Observational” Large Ensembles)

- Infer statistical characteristics of trends of any length from the amplitude and autocorrelation properties of interannual variability (Thompson et al. 2015). Works well, but does not provide information on spatial patterns.
- Statistical regression onto leading modes of variability (ENSO, PDV, AMV) with shuffled phases plus block bootstrapping of the residuals (McKinnon et al.). Preliminary step requires removal of the estimated forced response (in this case, via regression onto a SMILE’s forced GMST(t)).
- Dynamical Linear Inverse Modeling (LIM) of global SSTs (Newman et al.). Yields estimated forced response (gravest modes) and internal variability.

Alternate Realities: Dynamical Approach (LIM)

Observed SST (ERSSTv5) Trends (1980-2017)

Two LIM realizations

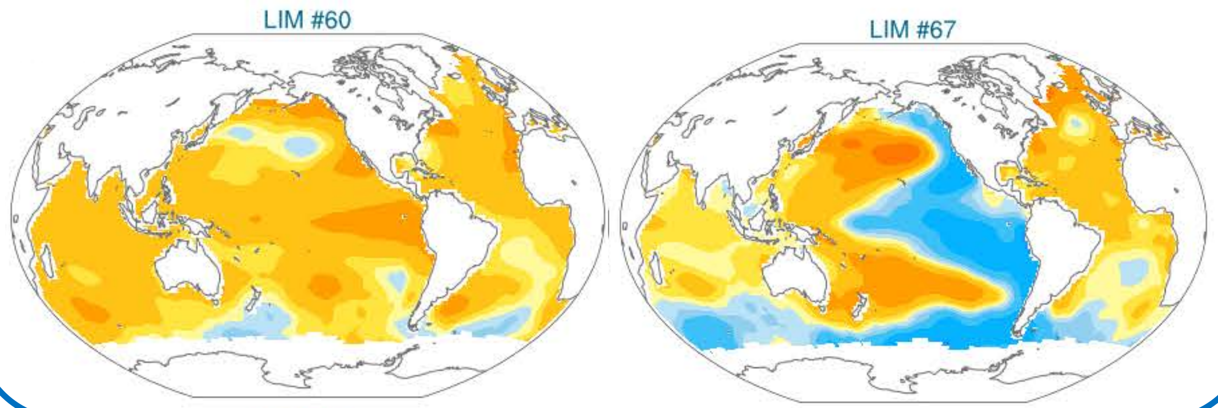


Courtesy of Matt Newman

Alternate Realities: Dynamical Approach (LIM)

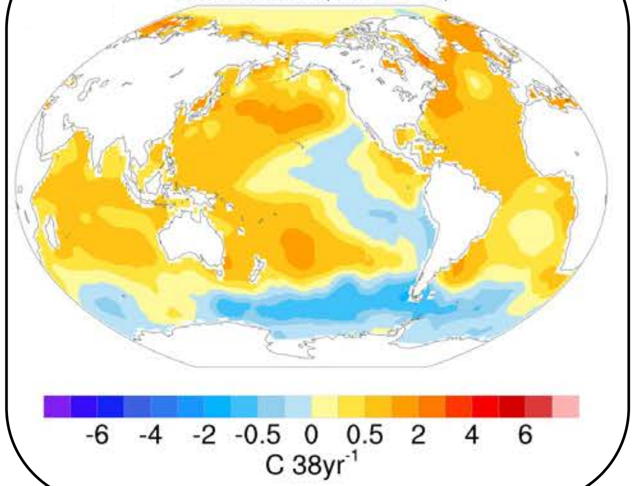
Observed SST (ERSSTv5) Trends (1980-2017)

Two LIM realizations



Courtesy of Matt Newman

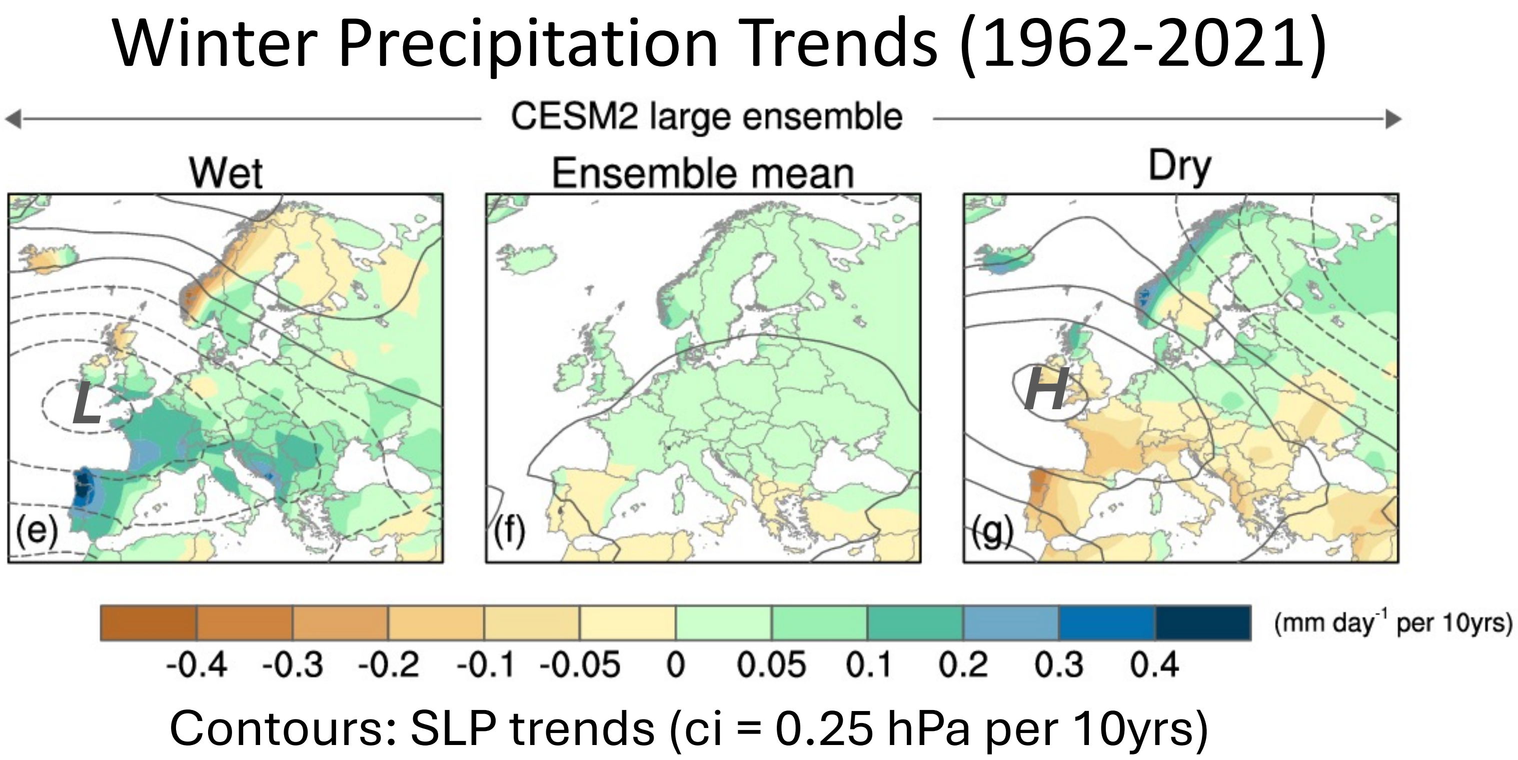
Actual



Alternate Realities: The role of the atmospheric circulation

Alternate Realities: The role of the atmospheric circulation

Internal SLP trends are driving most of the spread in precipitation trends.



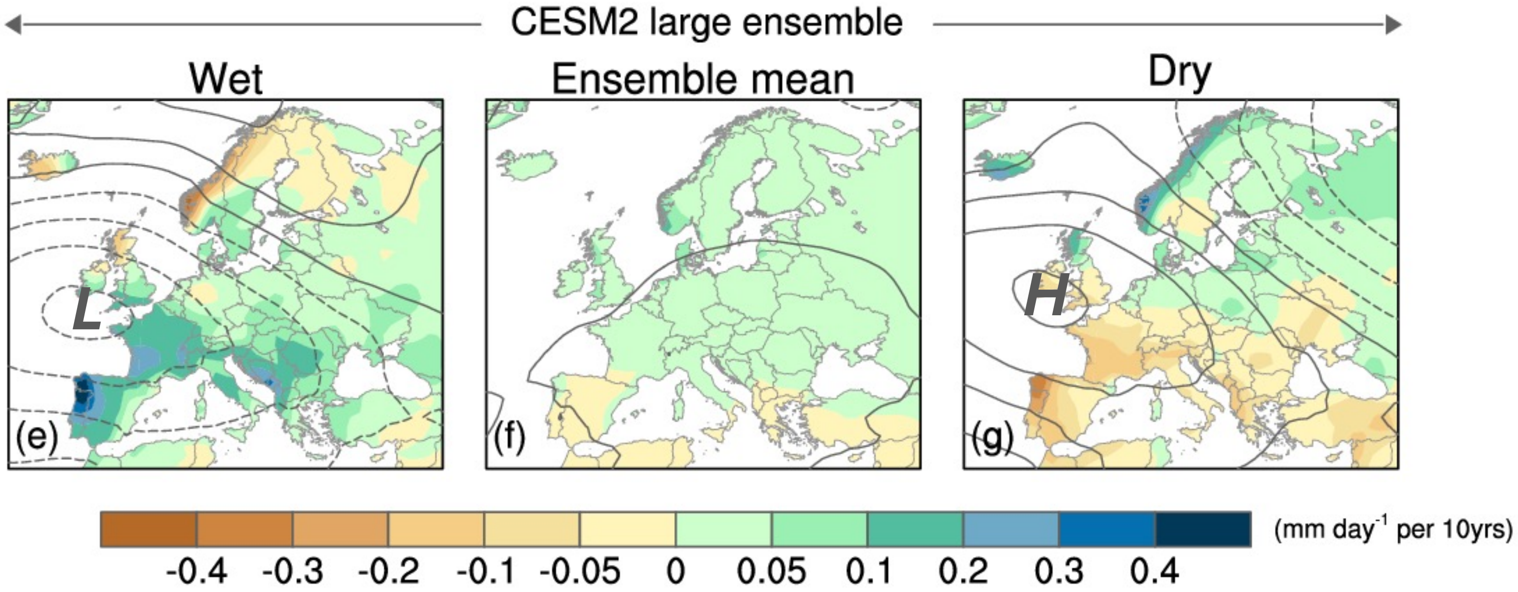
Deser and Phillips (2023)

Alternate Realities: The role of the atmospheric circulation

Internal SLP trends are driving most of the spread in precipitation trends.

Can we use this fact to estimate the forced component of observed precipitation trends?

Winter Precipitation Trends (1962-2021)

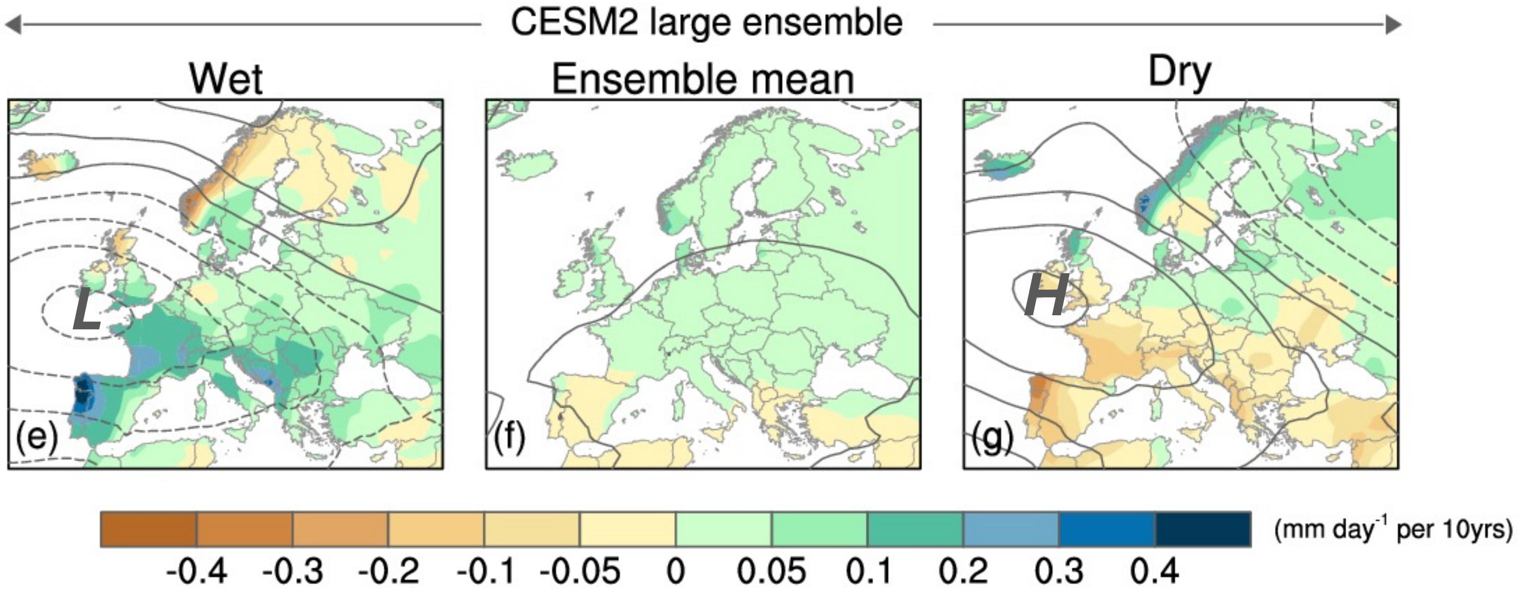


Contours: SLP trends (ci = 0.25 hPa per 10yrs)

Deser and Phillips (2023)

Alternate Realities: The role of the atmospheric circulation

Winter Precipitation Trends (1962-2021)



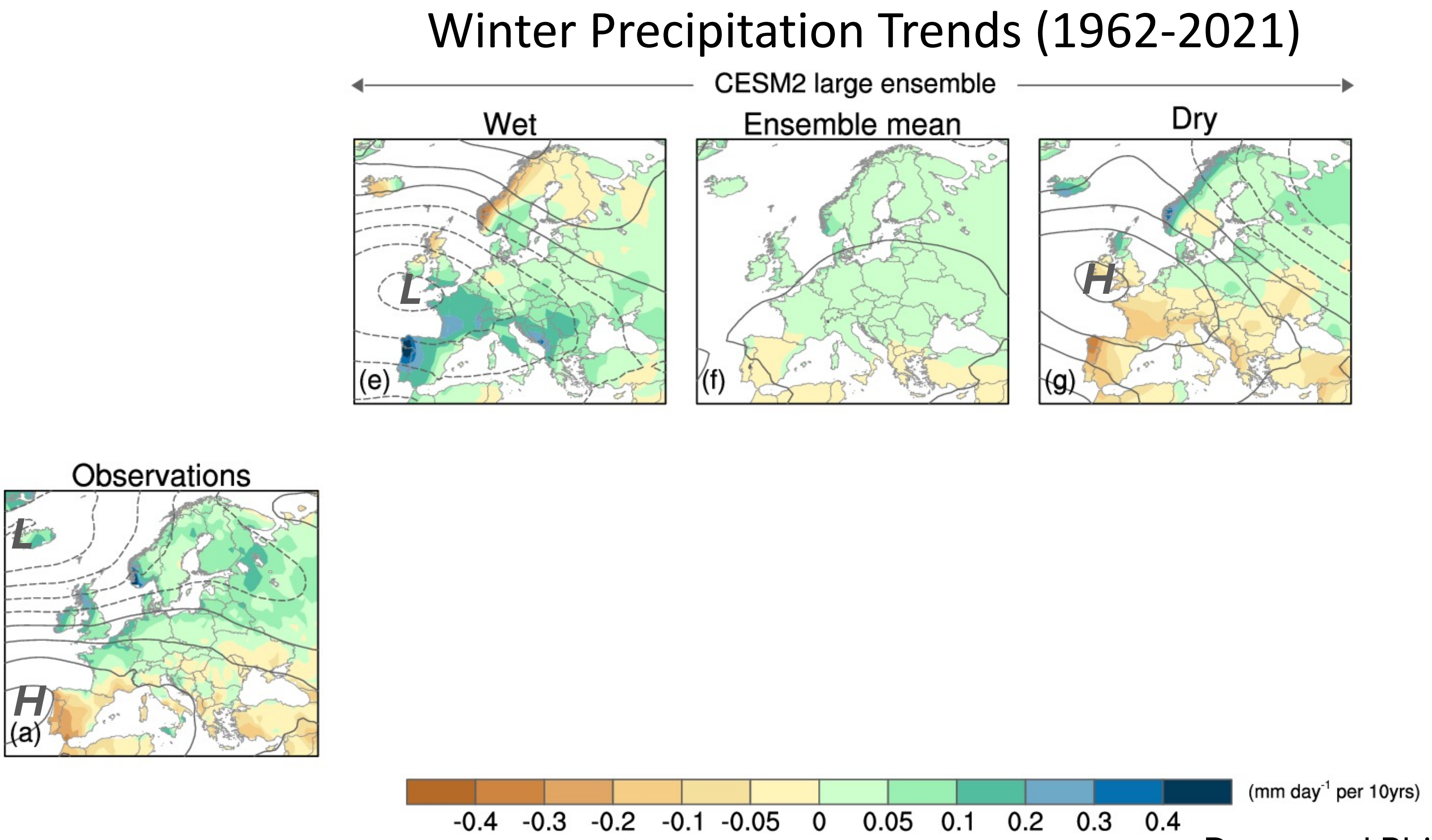
Contours: SLP trends (ci = 0.25 hPa per 10yrs)

Internal SLP trends are driving most of the spread in precipitation trends.

Can we use this fact to estimate the forced component of observed precipitation trends?

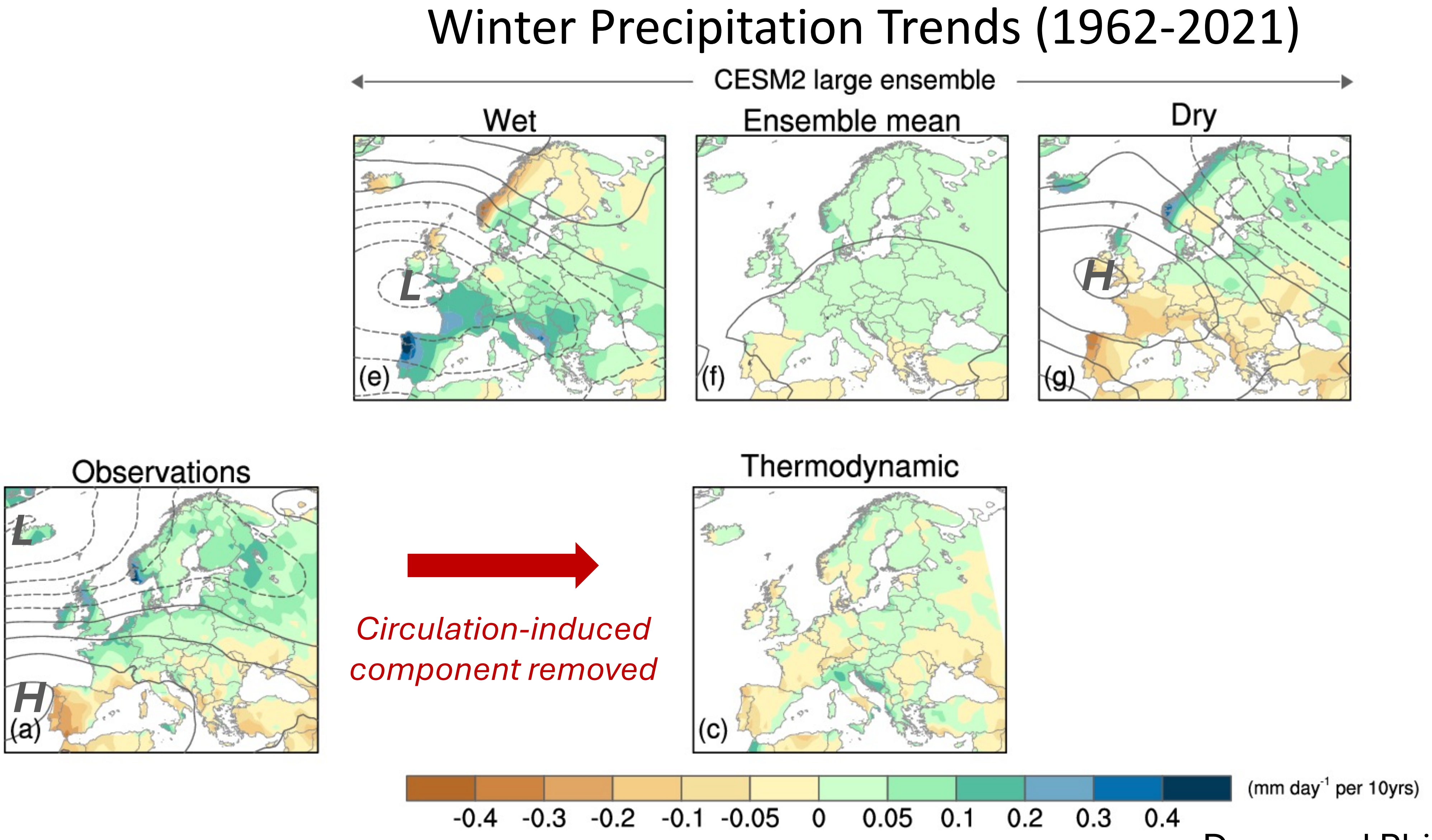
Remove an empirical estimate of the dynamically-induced component to obtain the forced component (thermodynamic) as a residual. "Dynamical Adjustment" procedure based on constructed circulation analogs using observed interannual SLP and precipitation relationships (Deser et al. 2018; Wallace et al. 2012).

Alternate Realities: The role of the atmospheric circulation



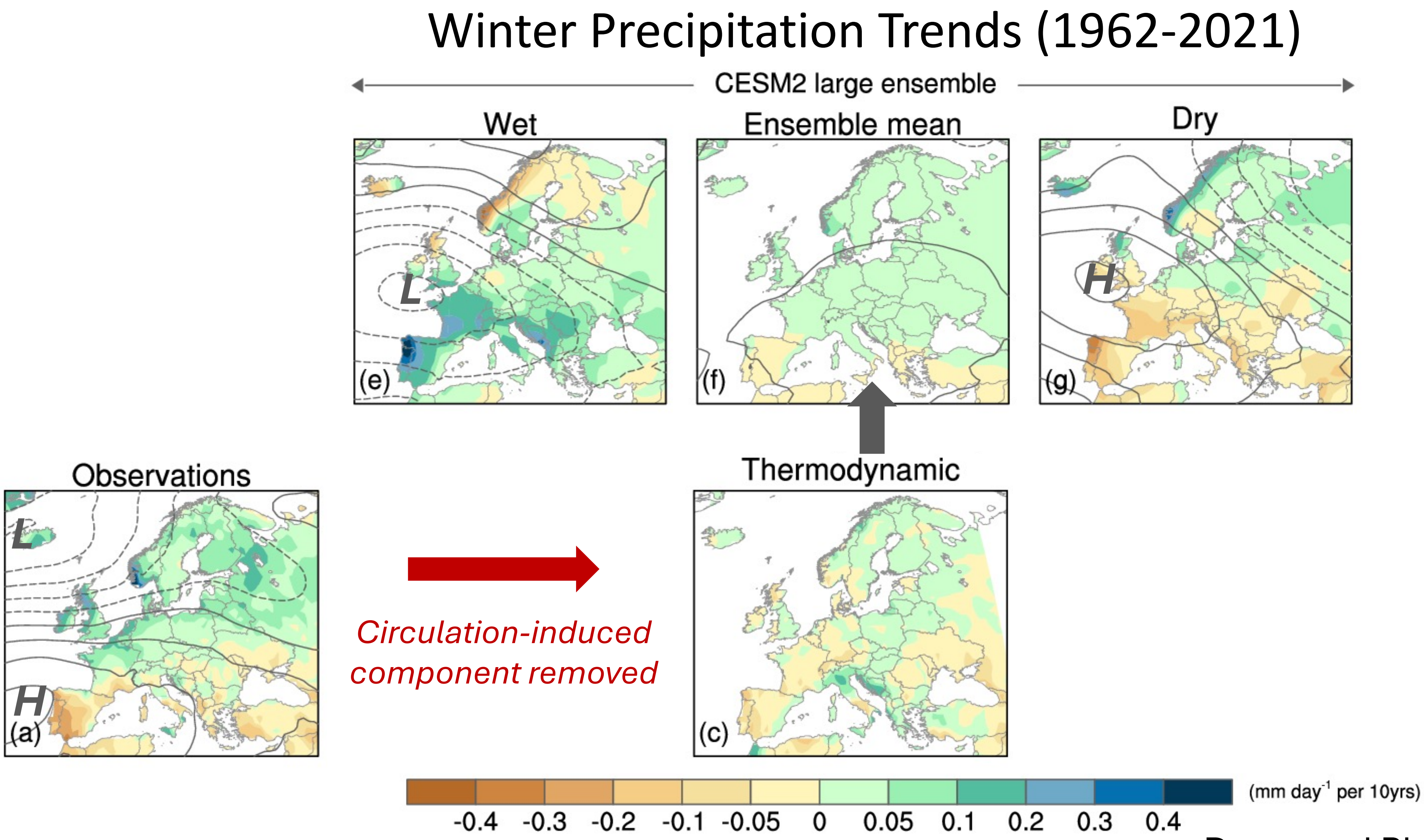
Deser and Phillips (2023)

Alternate Realities: The role of the atmospheric circulation



Deser and Phillips (2023)

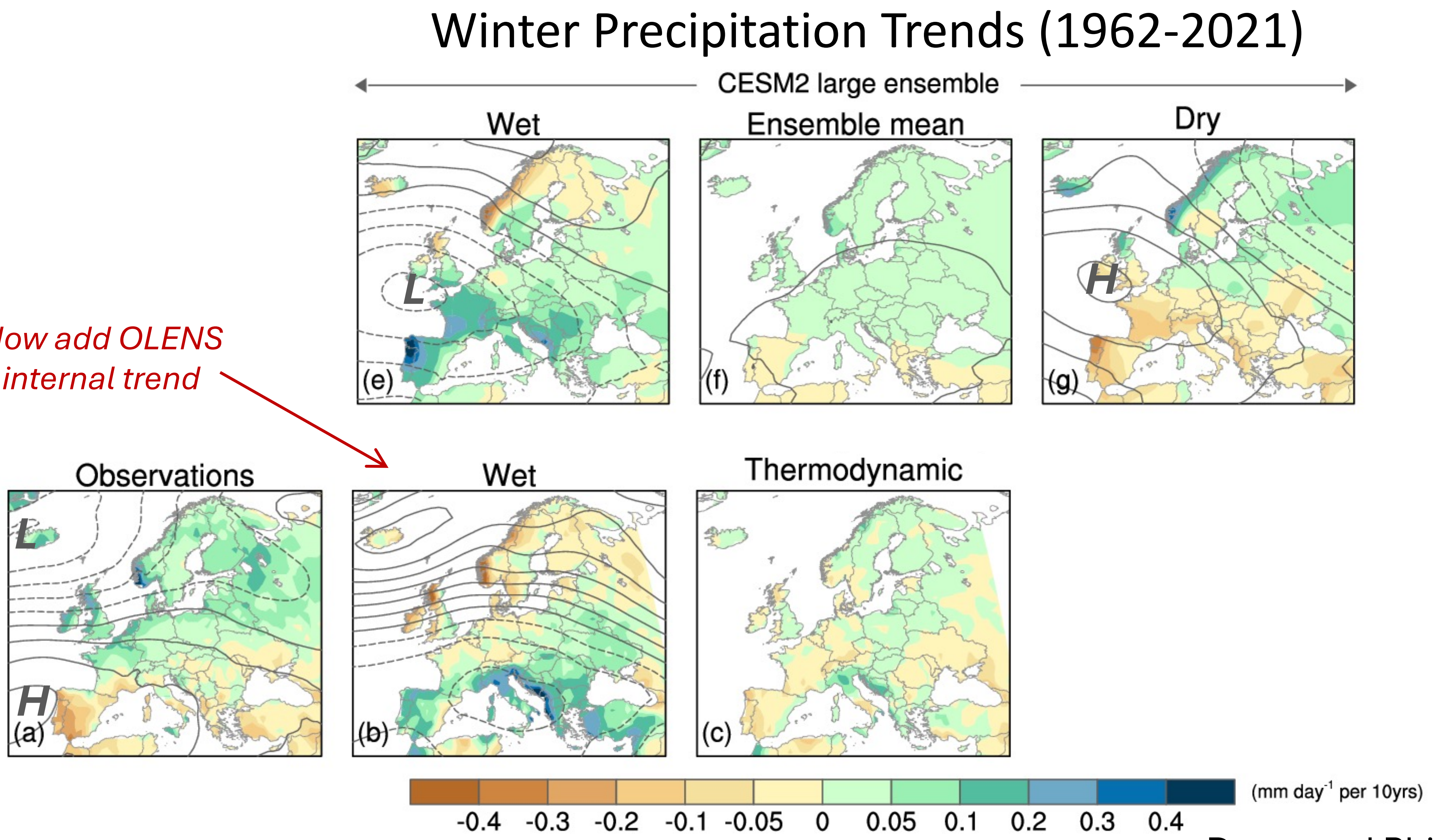
Alternate Realities: The role of the atmospheric circulation



Deser and Phillips (2023)

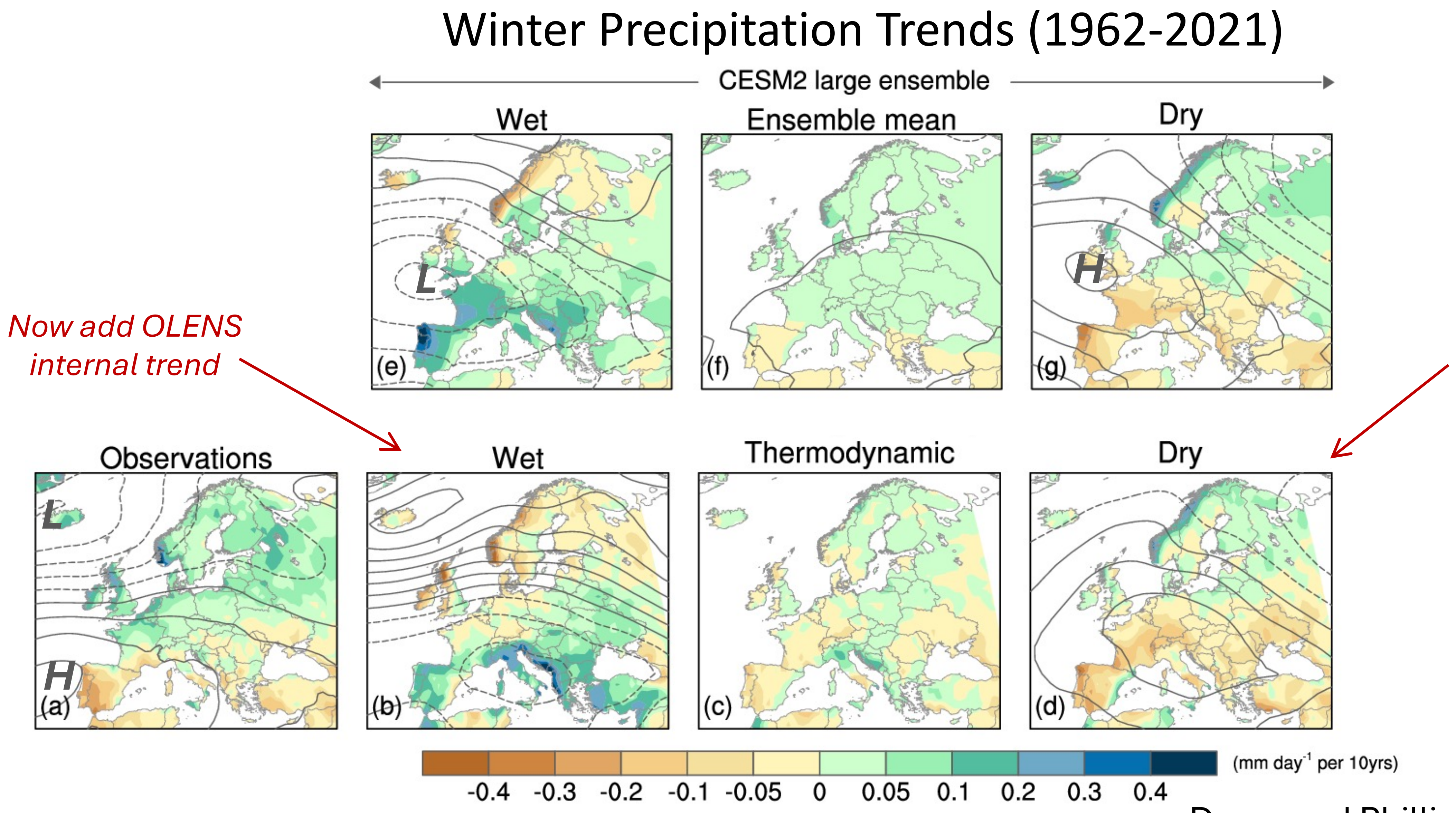
Alternate Realities: The role of the atmospheric circulation

Now add OLENS internal trend



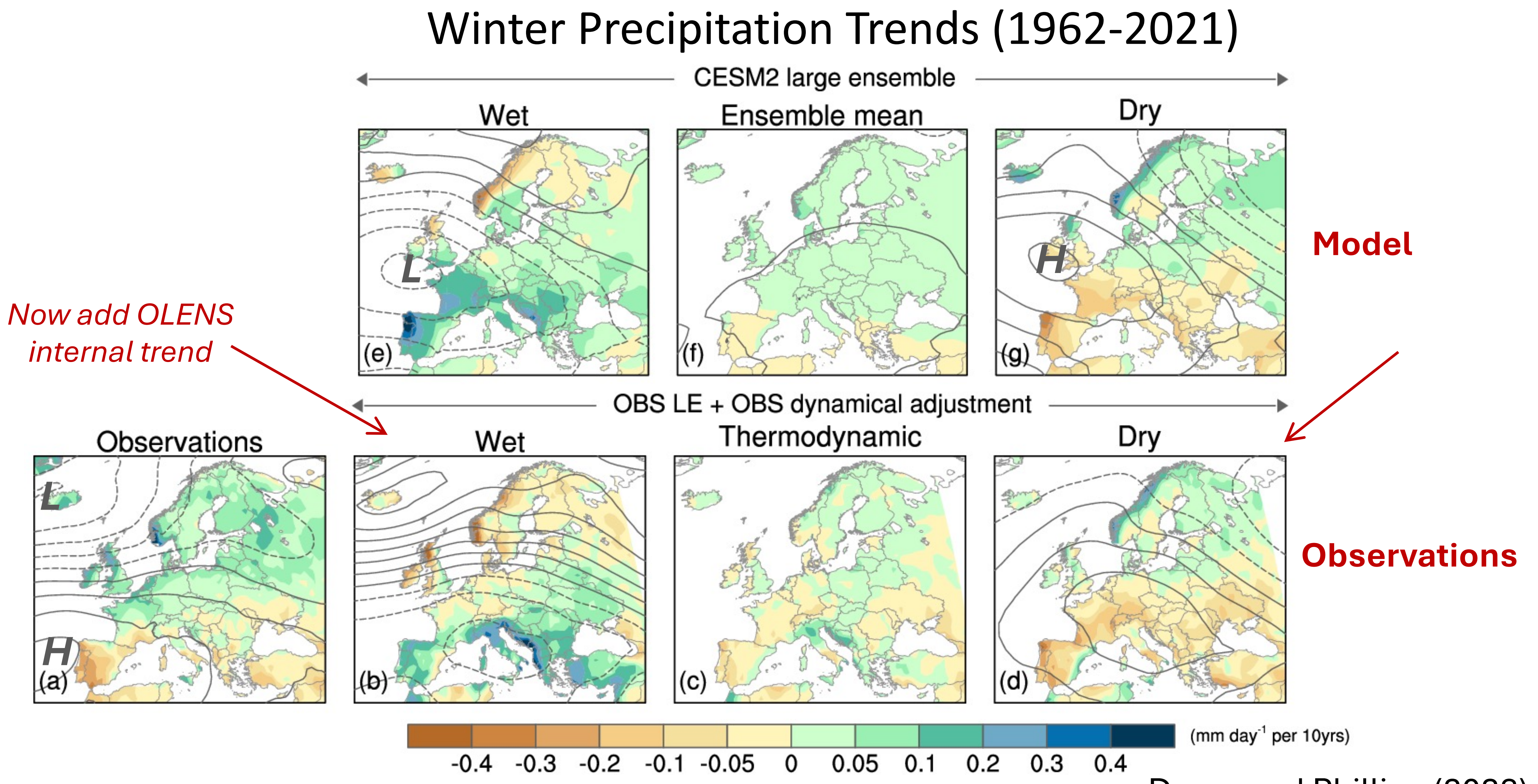
Deser and Phillips (2023)

Alternate Realities: The role of the atmospheric circulation



Deser and Phillips (2023)

Alternate Realities: The role of the atmospheric circulation

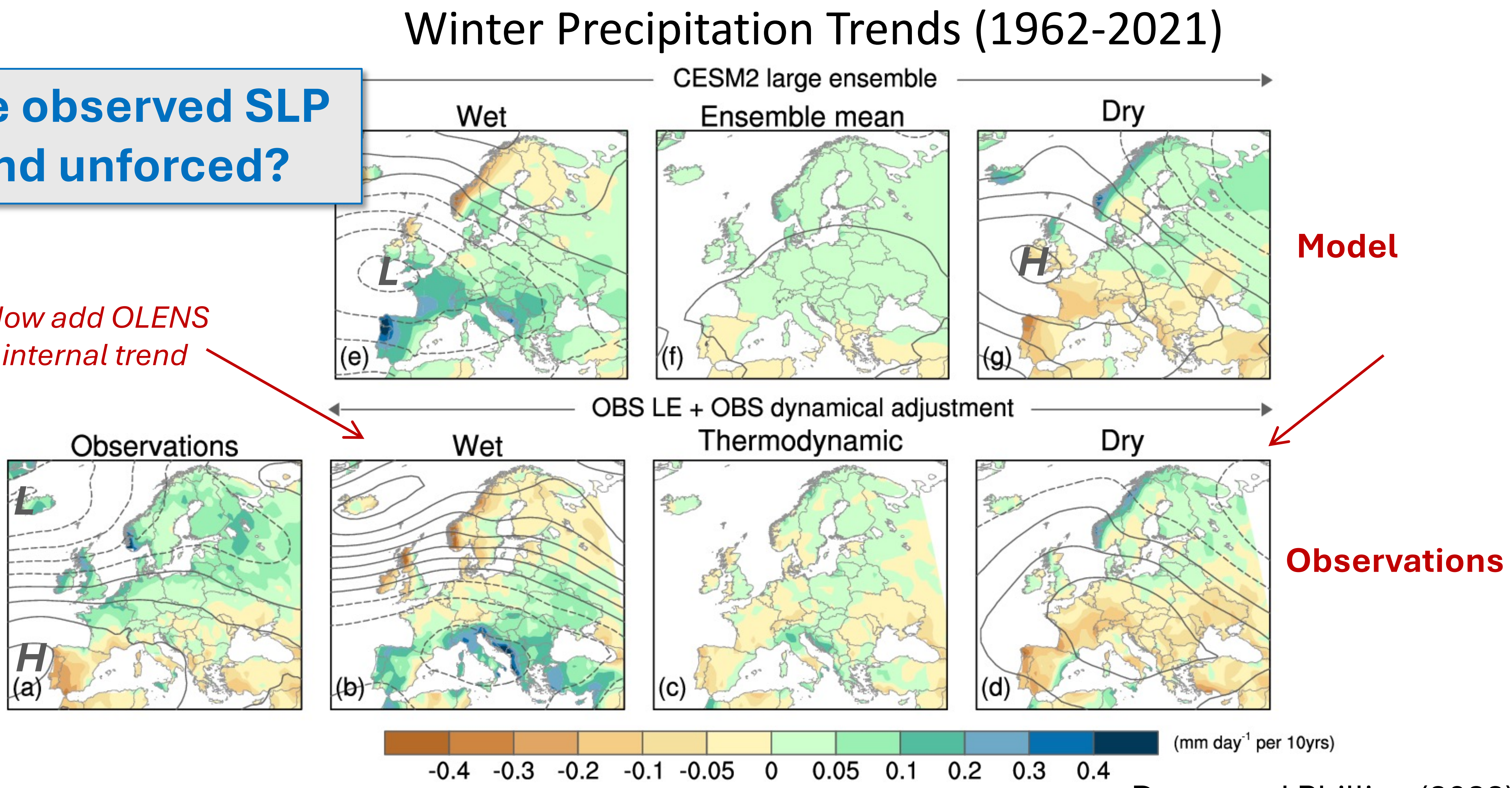


Deser and Phillips (2023)

Alternate Realities: The role of the atmospheric circulation

Is the observed SLP trend unforced?

Now add OLENS internal trend



Deser and Phillips (2023)

Additional Challenges and Implications

- The presence of forced trends may corrupt our empirical definitions of modes of internal variability (e.g., AMV, PDV, ENSO …).
- The limited length of the instrumental record restricts our knowledge of the true range of internal variability.
- SST trends may modulate impacts from the 2023/24 El Nino.
- A cautionary note on comparing CMIP5 & CMIP6 models.

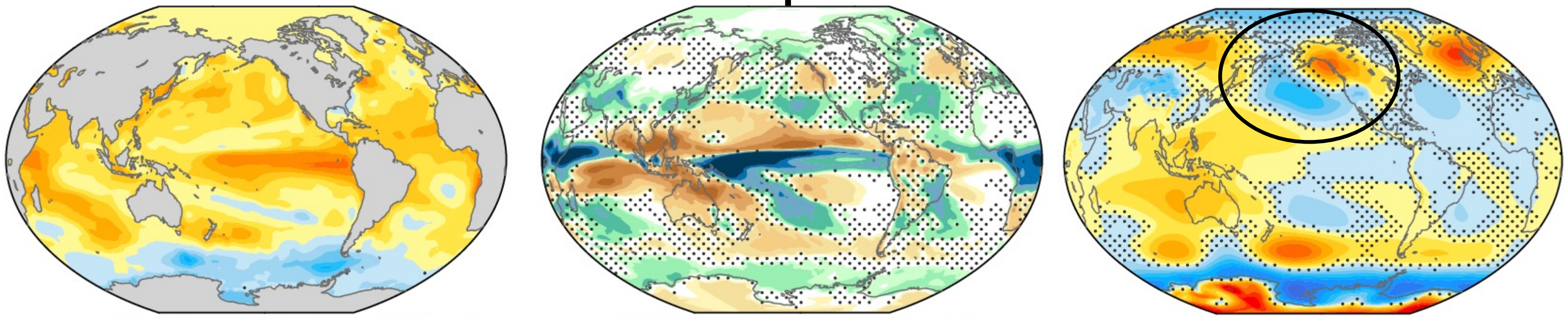
Additional Challenges and Implications

- The presence of forced trends may corrupt our empirical definitions of modes of internal variability (e.g., AMV, PDV, ENSO ...).
- The limited length of the instrumental record restricts our knowledge of the true range of internal variability.
- SST trends may modulate impacts from the 2023/24 El Nino.
- A cautionary note on comparing CMIP5 & CMIP6 models.

50-member AMIP responses

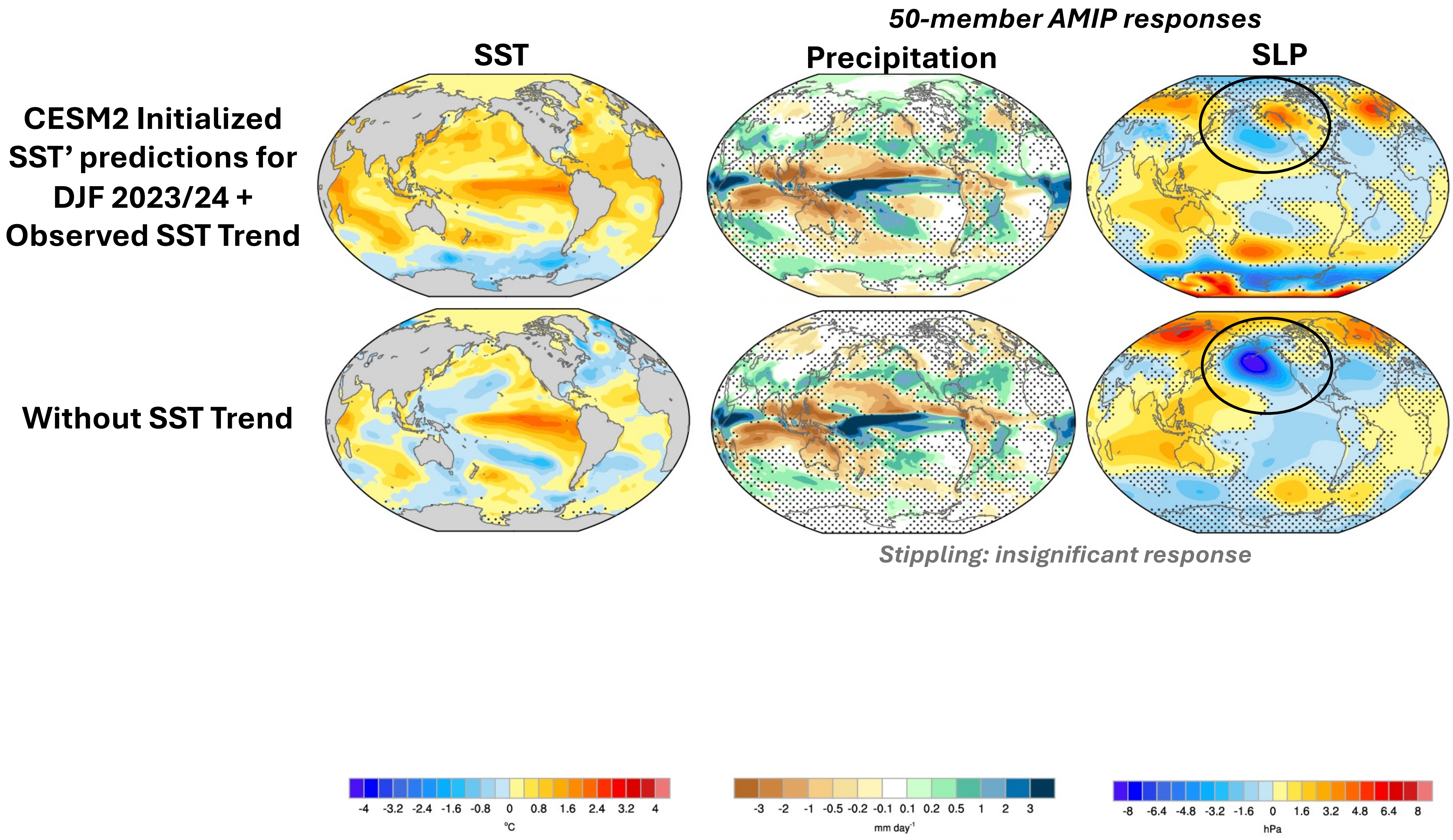
CESM2 Initialized SST' predictions for DJF 2023/24 + Observed SST Trend

SST **Precipitation** **SLP**

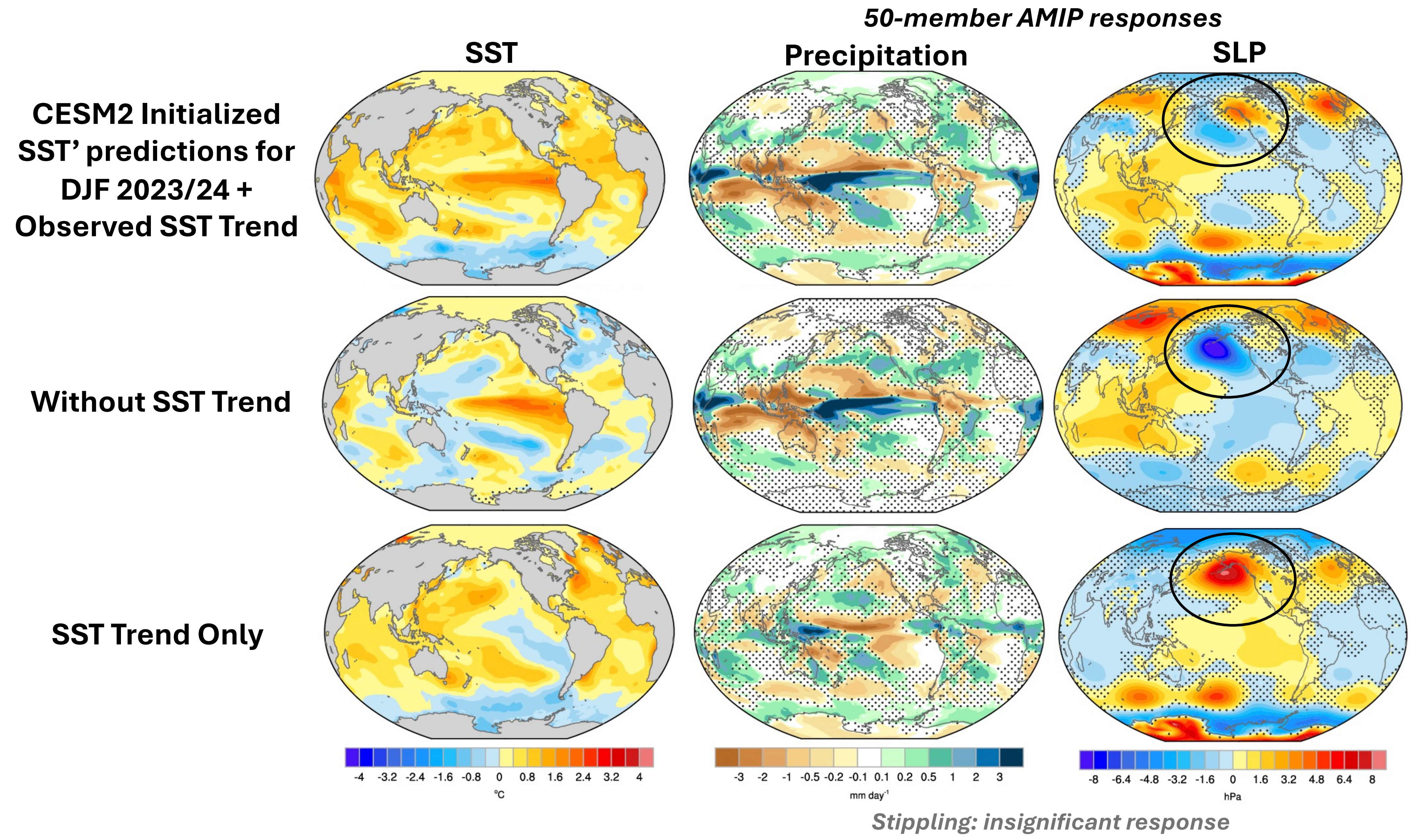


Stippling: insignificant response

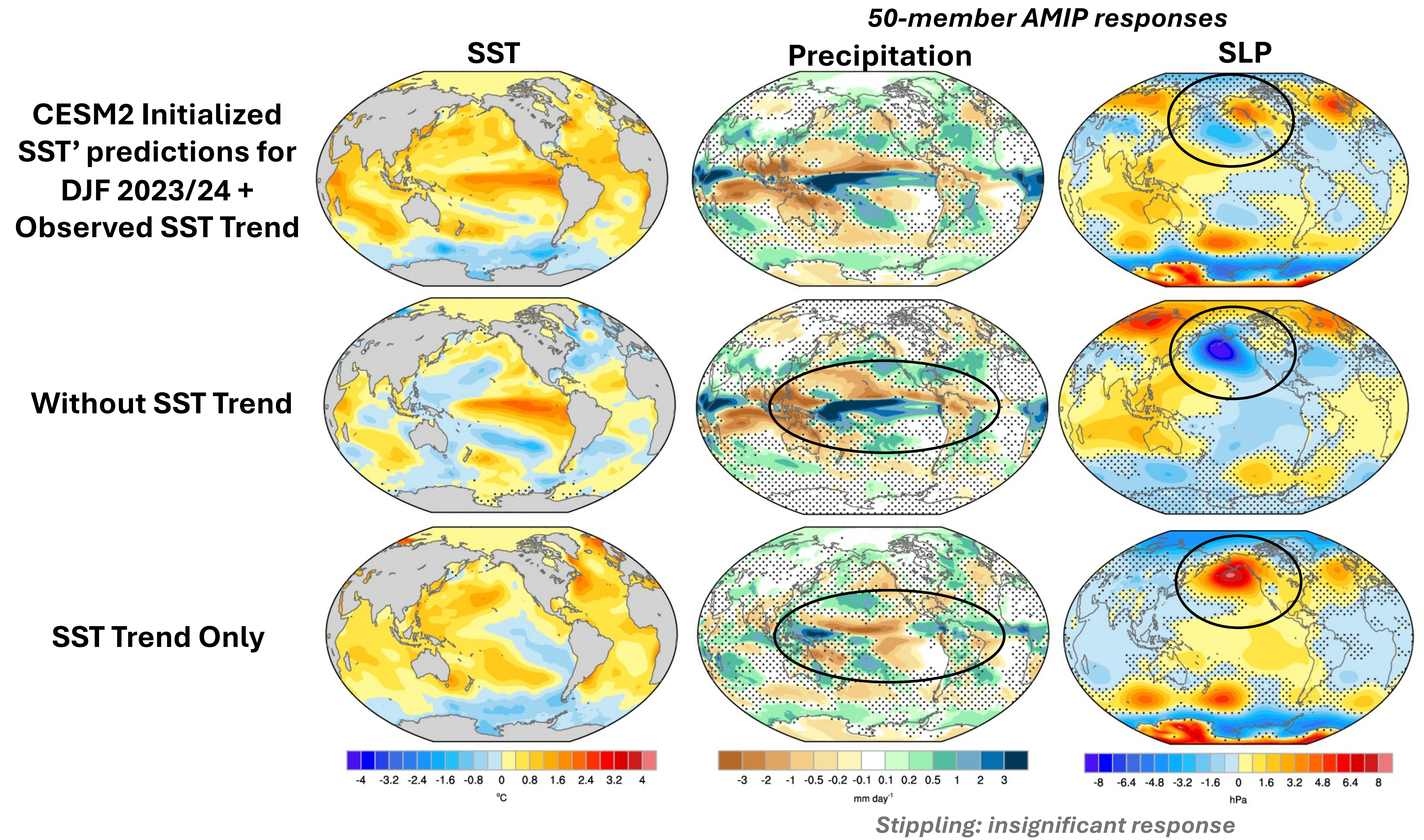
Deser et al. in preparation



Deser et al. in preparation

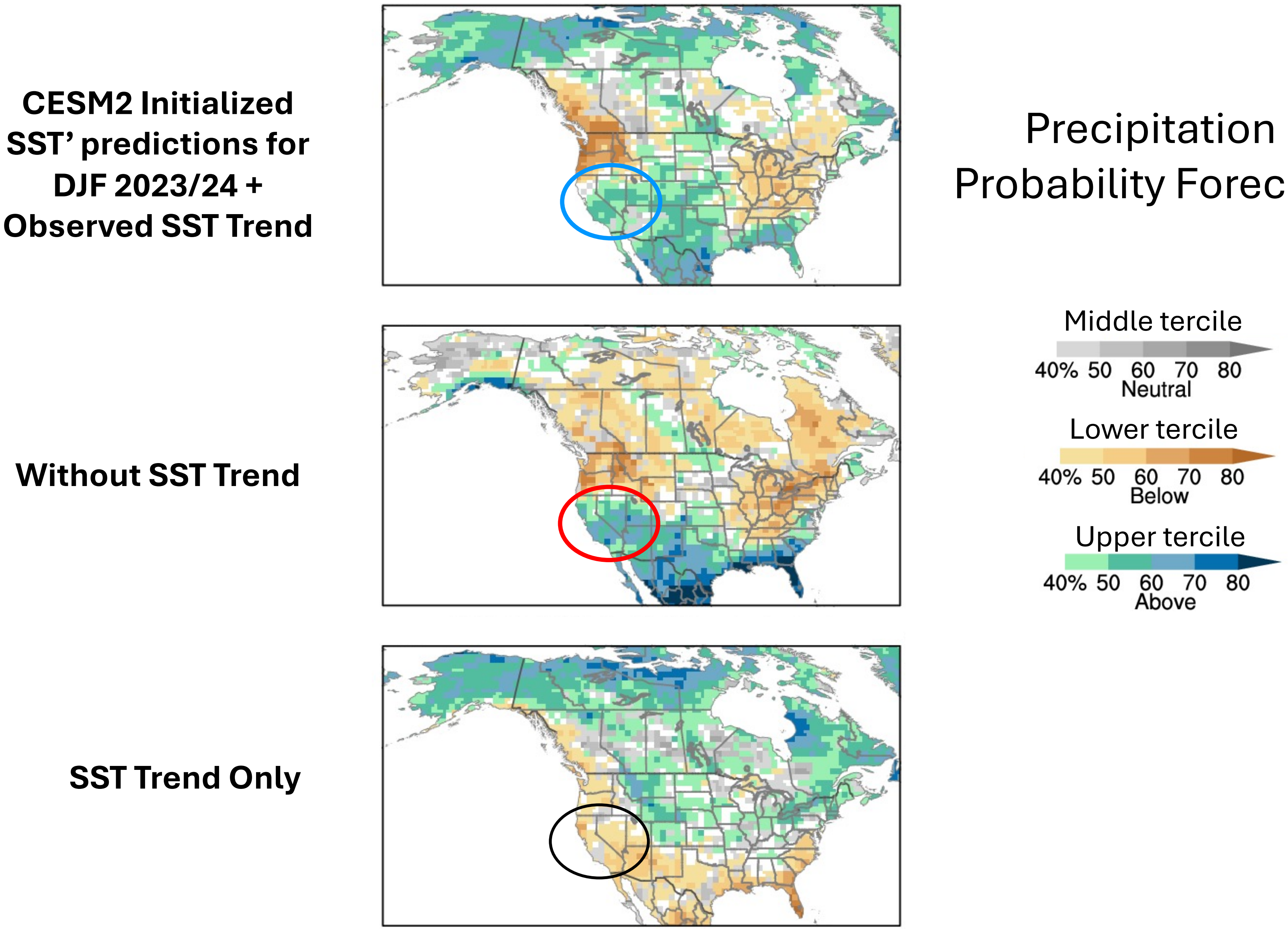


Deser et al. in preparation



Deser et al. in preparation

Precipitation Probability Forecast



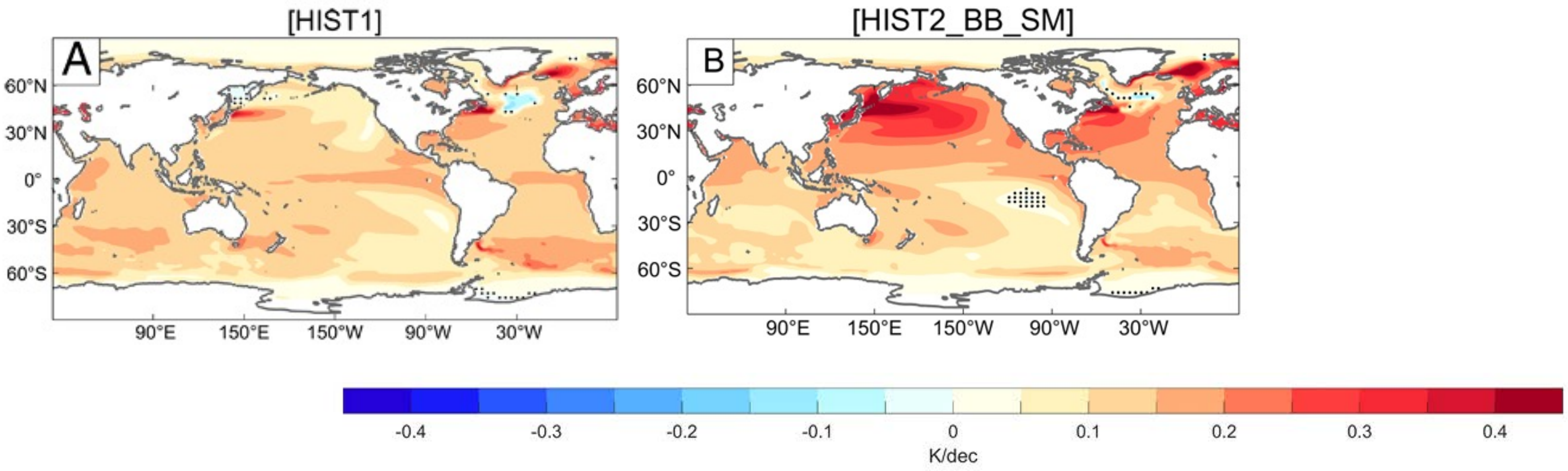
A cautionary note on comparing CMIP5 and CMIP6 models

A cautionary note on comparing CMIP5 and CMIP6 models

Ensemble Mean SST Trends 1979-2013

CESM1 (40)

CESM2 (50)



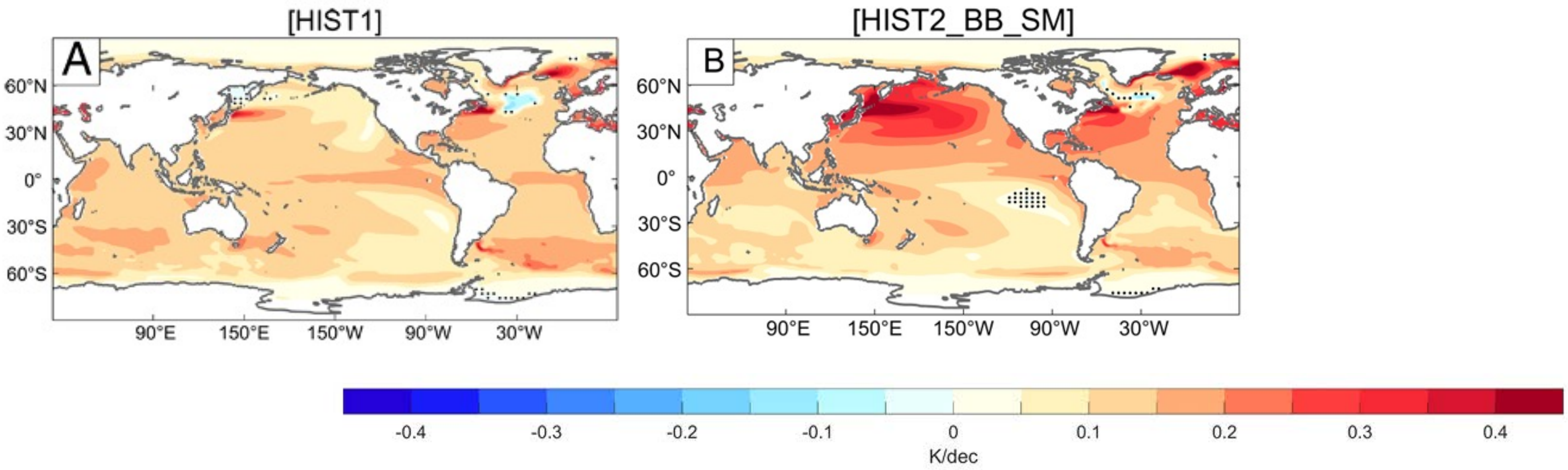
Kang et al. 2023

A cautionary note on comparing CMIP5 and CMIP6 models

Ensemble Mean SST Trends 1979-2013

CESM1 (40) CMIP5 forcing

CESM2 (50) CMIP6 forcing



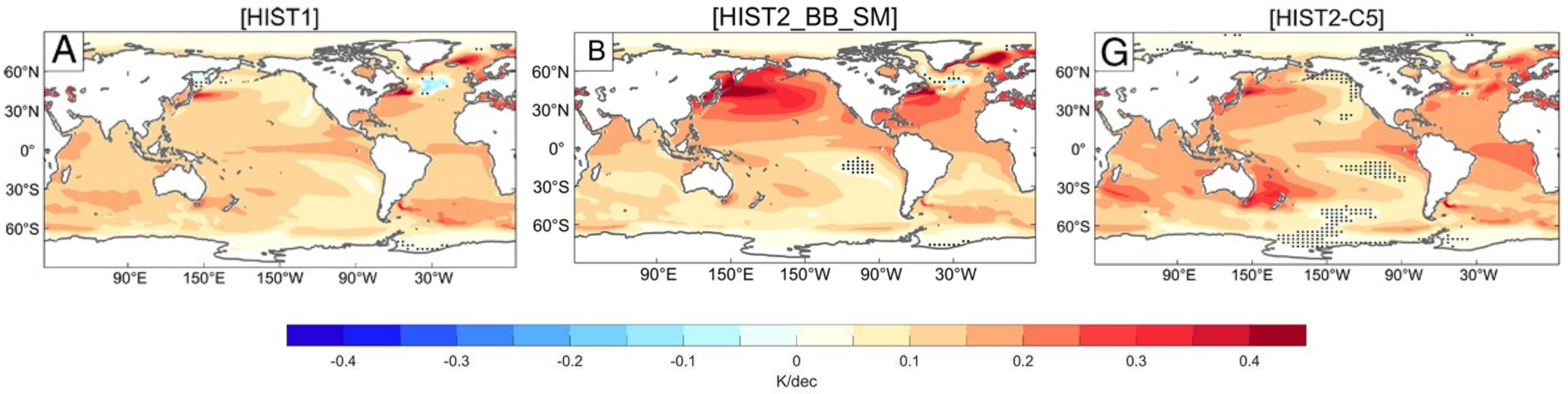
Kang et al. 2023

A cautionary note on comparing CMIP5 and CMIP6 models

Ensemble Mean SST Trends 1979-2013

Holland et al. 2024

CESM1 (40) CMIP5 forcing | **CESM2** (50) CMIP6 forcing | **CESM2 CMIP5** forcing (15)



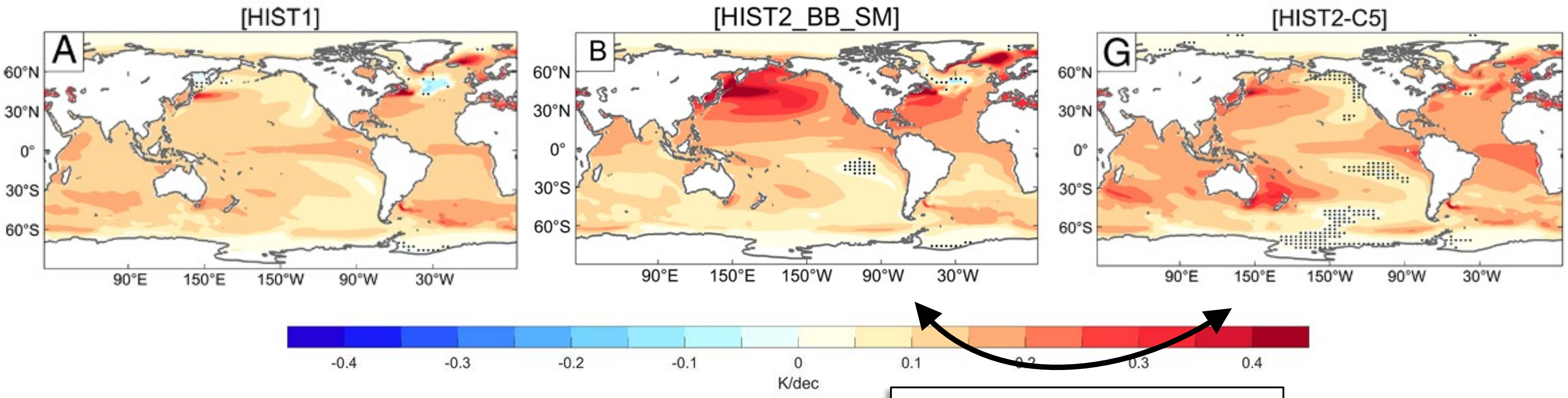
Kang et al. 2023

A cautionary note on comparing CMIP5 and CMIP6 models

Ensemble Mean SST Trends 1979-2013

Holland et al. 2024

CESM1 (40) CMIP5 forcing | **CESM2** (50) CMIP6 forcing | **CESM2 CMIP5** forcing (15)



Kang et al. 202

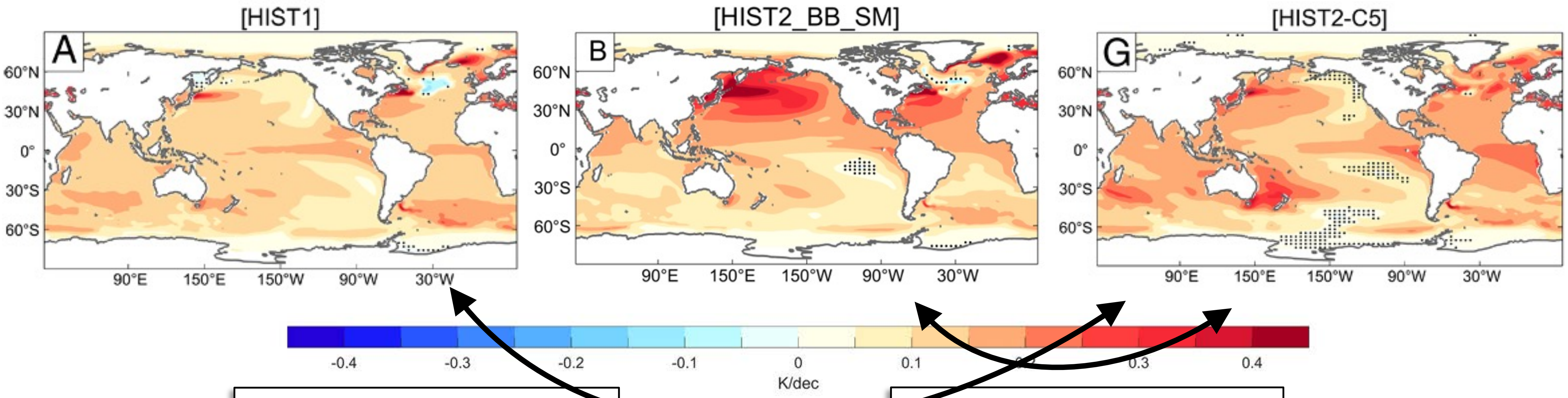
Same model, different forcings.

A cautionary note on comparing CMIP5 and CMIP6 models

Ensemble Mean SST Trends 1979-2013

Holland et al. 2024

CESM1 (40) CMIP5 forcing | **CESM2** (50) CMIP6 forcing | **CESM2 CMIP5** forcing (15)



Different model, same forcings.

Kang et al. 2023

Same model, different forcings.

NCAR
UCAR

CGD's Climate Analysis Section

Climate Variability Diagnostics Package

for Large Ensembles

https://www.cesm.ucar.edu/projects/cvdp-le

An automated analysis tool and data repository for exploring forced and internal components of climate variability and change.

NCAR UCAR

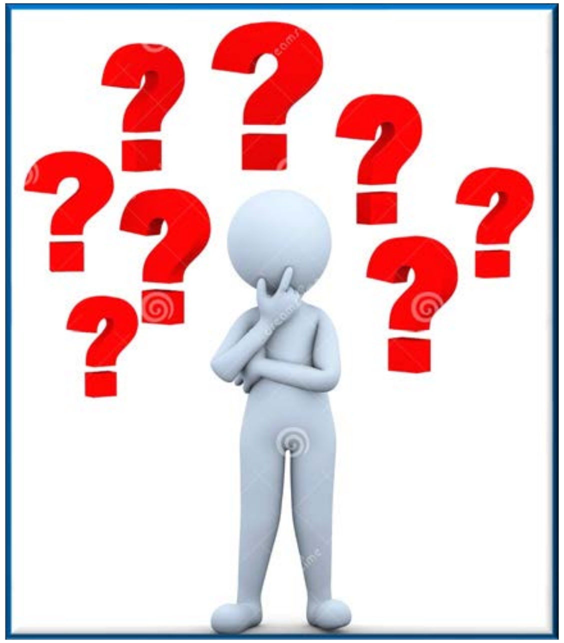
CGD's Climate Analysis Section

Climate Variability Diagnostics Package

for Large Ensembles

https://www.cesm.ucar.edu/projects/cvdp-le

- How well does a given model simulate the mean state, long-term trends, and modes of variability such as ENSO, NAO, AMV, PDV?
- How do models compare with each other? Are there true structural differences?
- How does climate change affect internal variability?
- What are the relative contributions of internal variability and forced climate change to long-term trends?



NCAR UCAR

CGD's Climate Analysis Section

Climate Variability Diagnostics Package

for Large Ensembles

https://www.cesm.ucar.edu/projects/cvdp-le

- How well does a given model simulate the mean state, long-term trends, and modes of variability such as ENSO, NAO, AMV, PDV?
- How do models compare with each other? Are there true structural differences?
- How does climate change affect internal variability?
- What are the relative contributions of internal variability and forced climate change to long-term trends?

Null hypothesis for any apparent model bias, inter-model difference, and projected change in variability should be "sampling fluctuations".

Some Tools and Resources

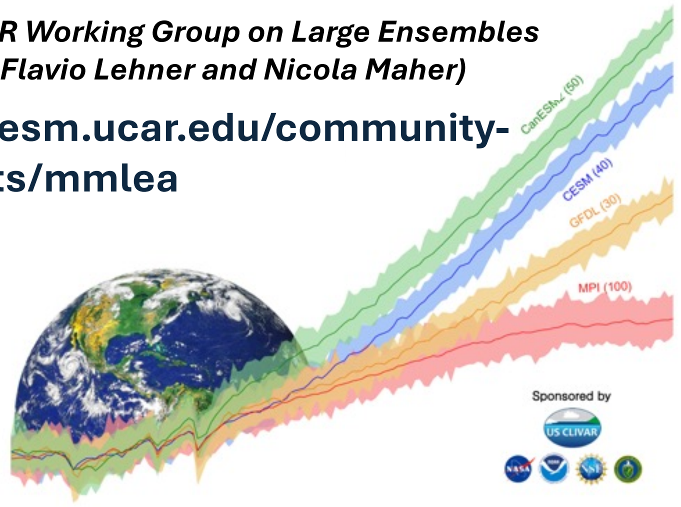
NCAR UCAR
CGD's Climate Analysis Section
Climate Variability Diagnostics Package
for Large Ensembles

http://www.cesm.ucar.edu/working_groups/CVC/cvdp/

MULTI-MODEL LARGE ENSEMBLE ARCHIVE

US CLIVAR Working Group on Large Ensembles
(credit to Flavio Lehner and Nicola Maher)

www.cesm.ucar.edu/community-projects/mmlea



NCAR UCAR ClimateDataGuide
inform • compare • discover

CLIMATE DATA | ANALYSIS TOOLS | MODEL EVALUATION | EXPERT CONTRIBUTORS | ABOUT | Site-wide Search >>

Concise and reliable expert guidance on the strengths, limitations and applications of climate data…

https://climatedataguide.ucar.edu

National Science Foundation

Swiss National Science Foundation

Aug. 29-31, 2023

Forced Component Estimation Statistical Method Intercomparison Project (ForceSMIP)

Organizers: Robb Wills, Clara Deser, Karen McKinnon, Adam Phillips, Stephen Po-Chedley, Sebastian Sippel

Aug. 29-31, 2023

Project (ForceSMIP)

Linear Trends (1950-2022)

Precipitation

SST

National Science Foundation

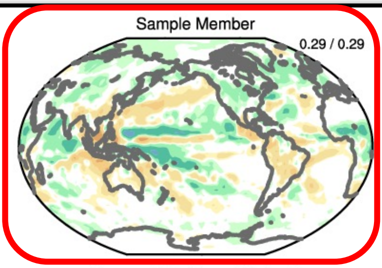
Swiss National Science Foundation

Aug. 29-31, 2023

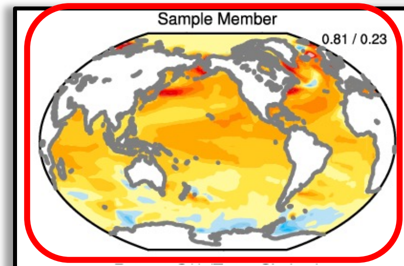
Project (ForceSMIP)

Linear Trends (1950-2022)

Precipitation



SST



National Science Foundation

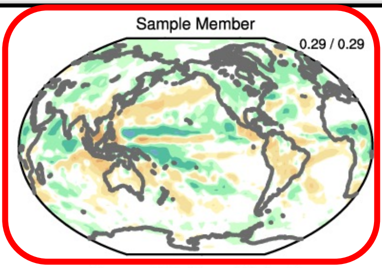
Swiss National Science Foundation

Aug. 29-31, 2023

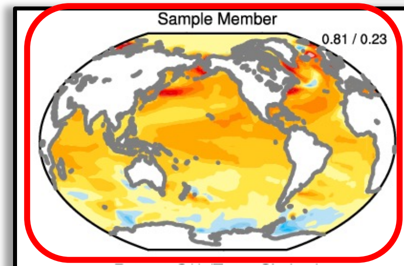
Linear Trends (1950-2022)

Project (ForceSMIP)

Precipitation



SST



Each participant was asked to determine the forced trend component in the sample member, with no knowledge of the truth.

Methods were developed on a set of "training ensembles" and evaluated using a different set of ensembles. Model identification was removed and all models were put on a common grid.

National Science Foundation

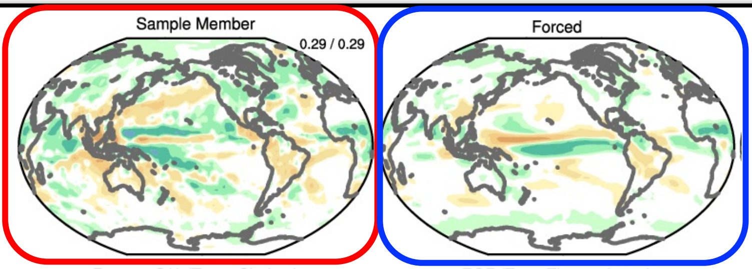
Swiss National Science Foundation

Aug. 29-31, 2023

Linear Trends (1950-2022)

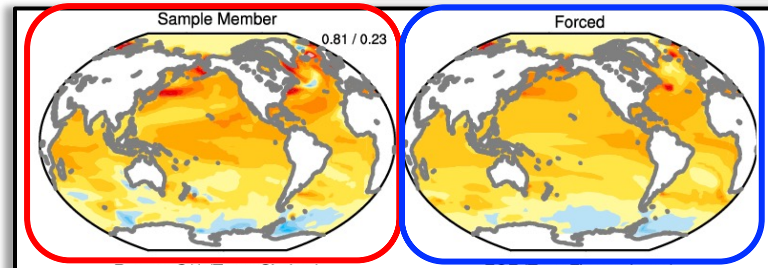
Project (ForceSMIP)

Precipitation



Unknown to participants

SST



Unknown to participants

Each participant was asked to determine the forced trend component in the sample member, with no knowledge of the truth.

Methods were developed on a set of "training ensembles" and evaluated using a different set of ensembles. Model identification was removed and all models were put on a common grid.

National Science Foundation

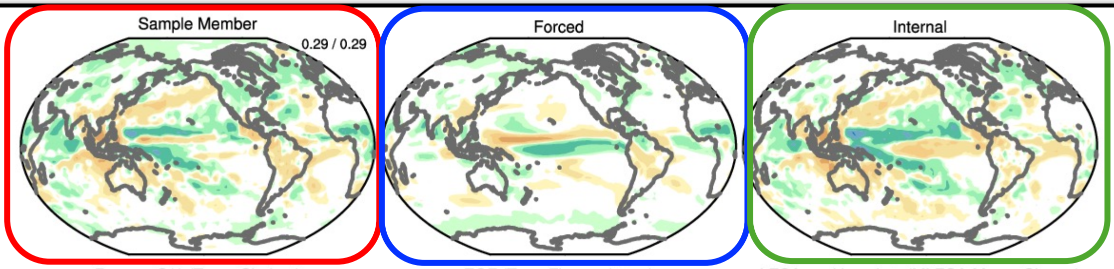
Swiss National Science Foundation

Aug. 29-31, 2023

Linear Trends (1950-2022)

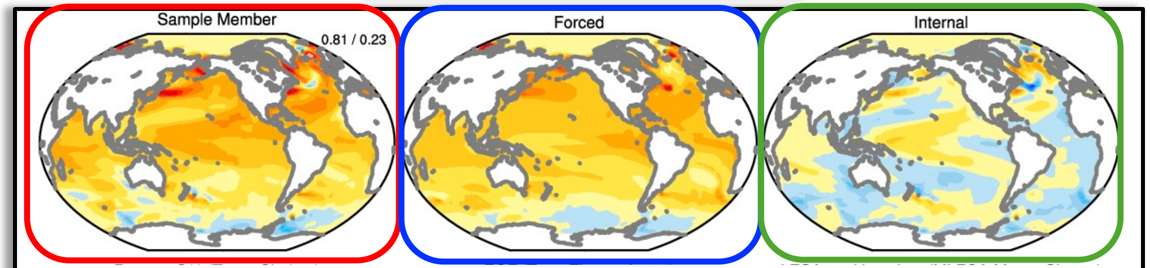
Project (ForceSMIP)

Precipitation



Unknown to participants

SST



Unknown to participants

Each participant was asked to determine the forced trend component in the sample member, with no knowledge of the truth.

Methods were developed on a set of "training ensembles" and evaluated using a different set of ensembles. Model identification was removed and all models were put on a common grid.

National Science Foundation

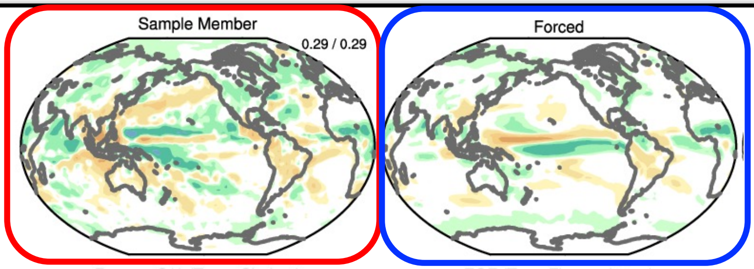
Swiss National Science Foundation

Aug. 29-31, 2023

Linear Trends (1950-2022)

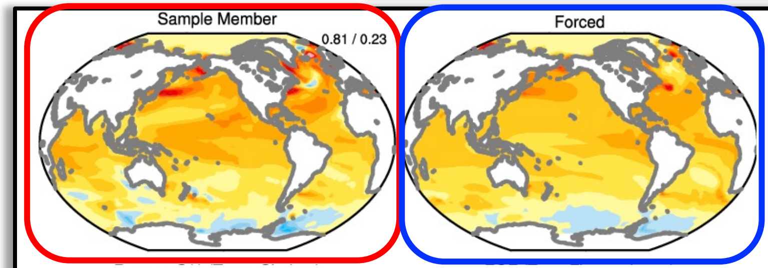
Project (ForceSMIP)

Precipitation



Unknown to participants

SST



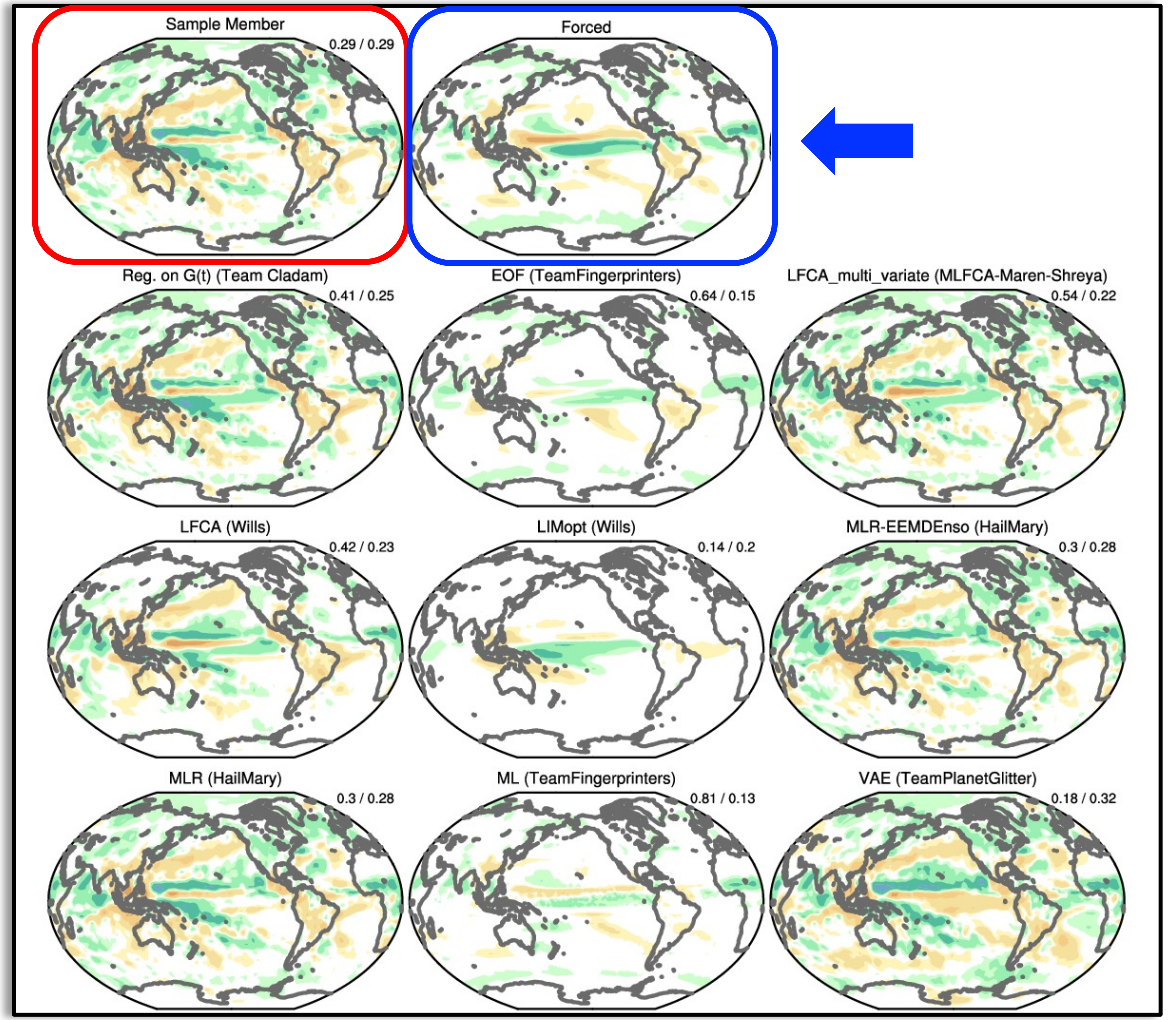
Unknown to participants

Aug. 29-31, 2023

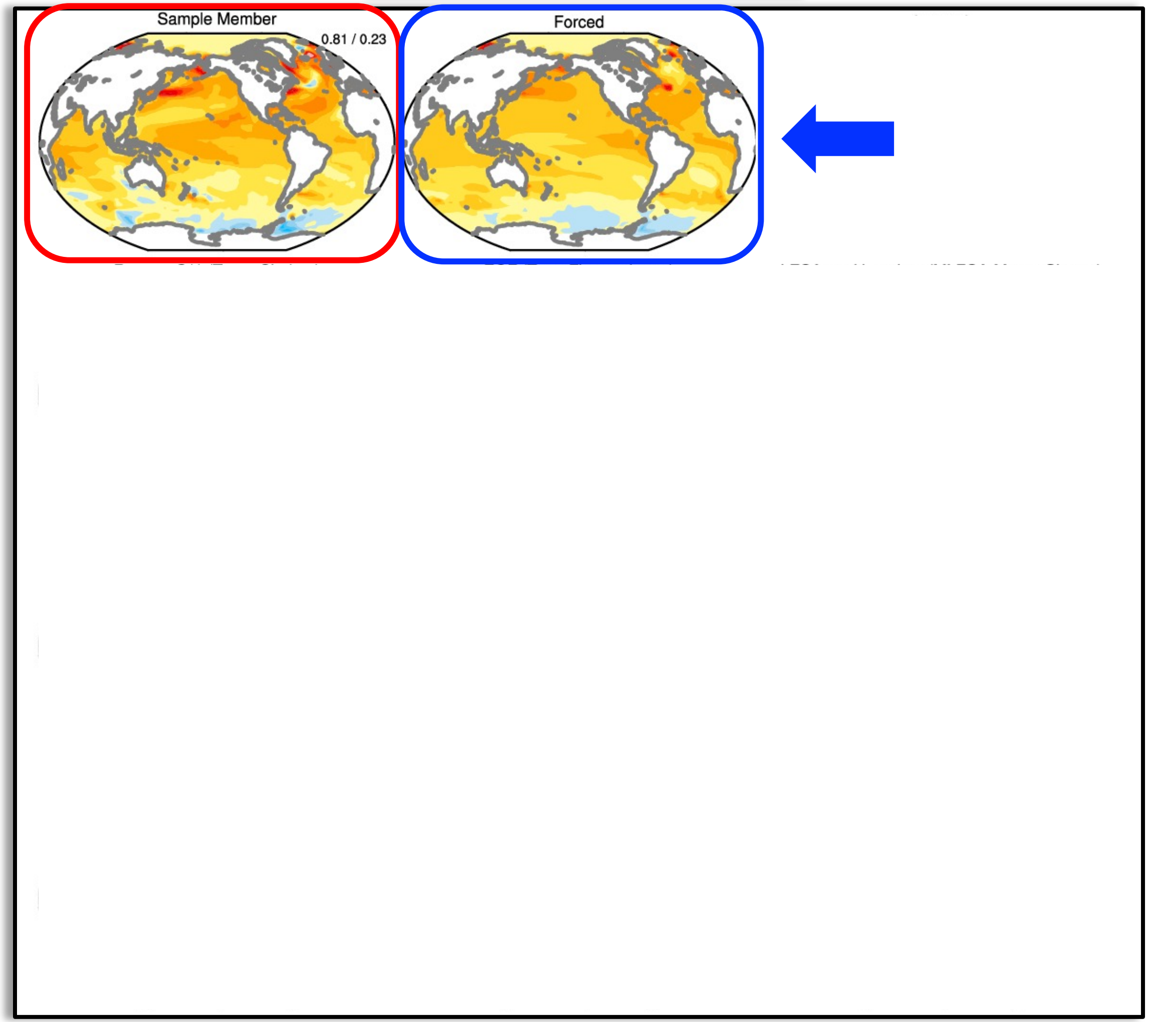
Project (ForceSMIP)

Linear Trends (1950-2022)

Precipitation



SST



National Science Foundation

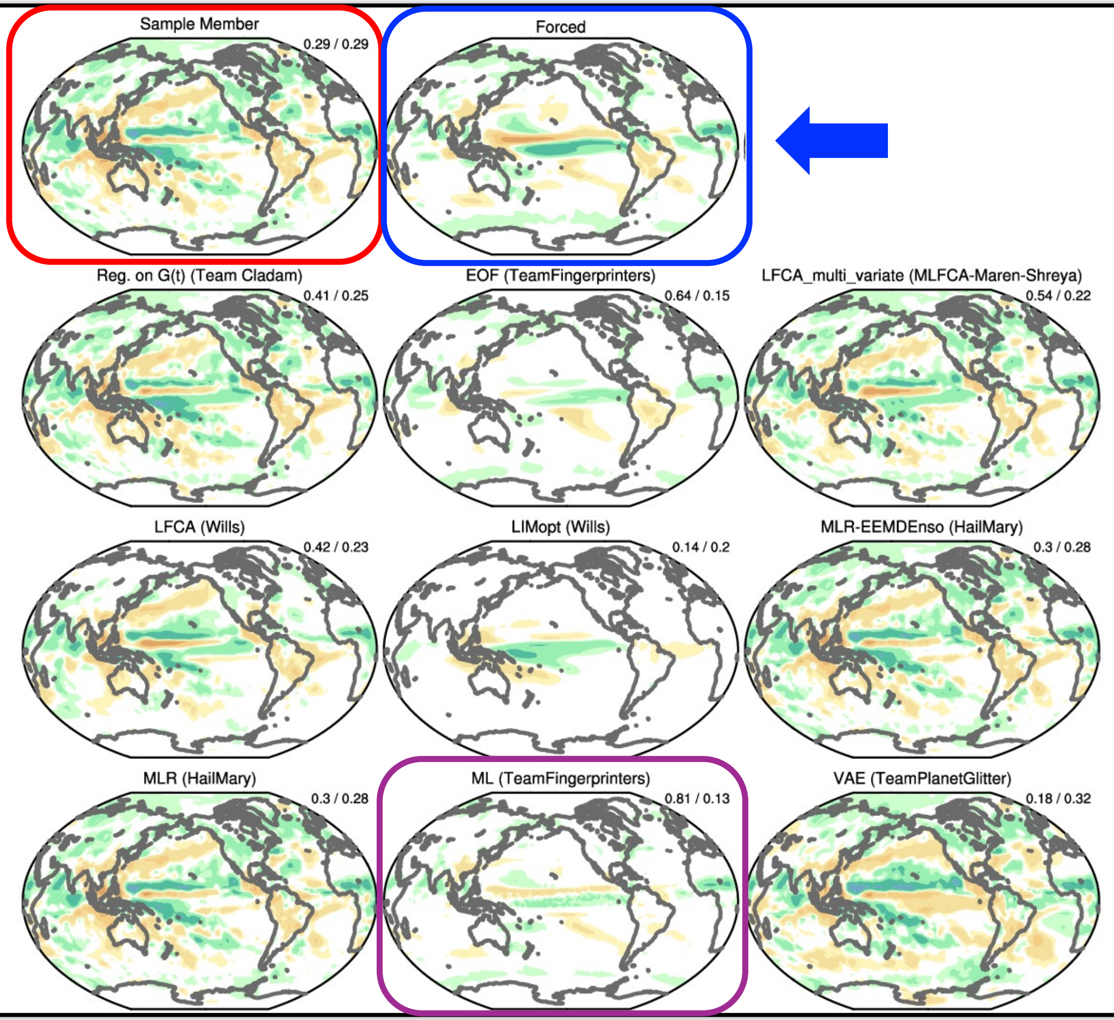
Swiss National Science Foundation

Aug. 29-31, 2023

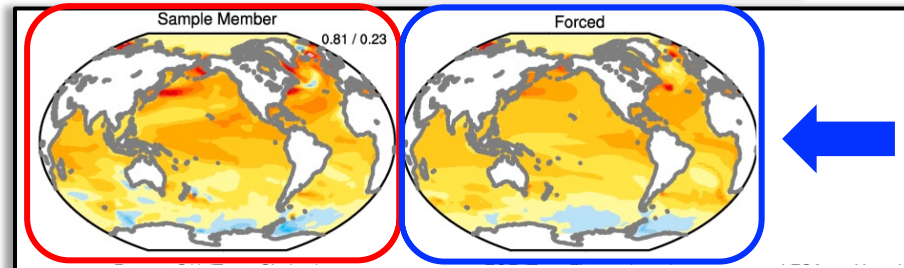
Linear Trends (1950-2022)

Project (ForceSMIP)

Precipitation



SST

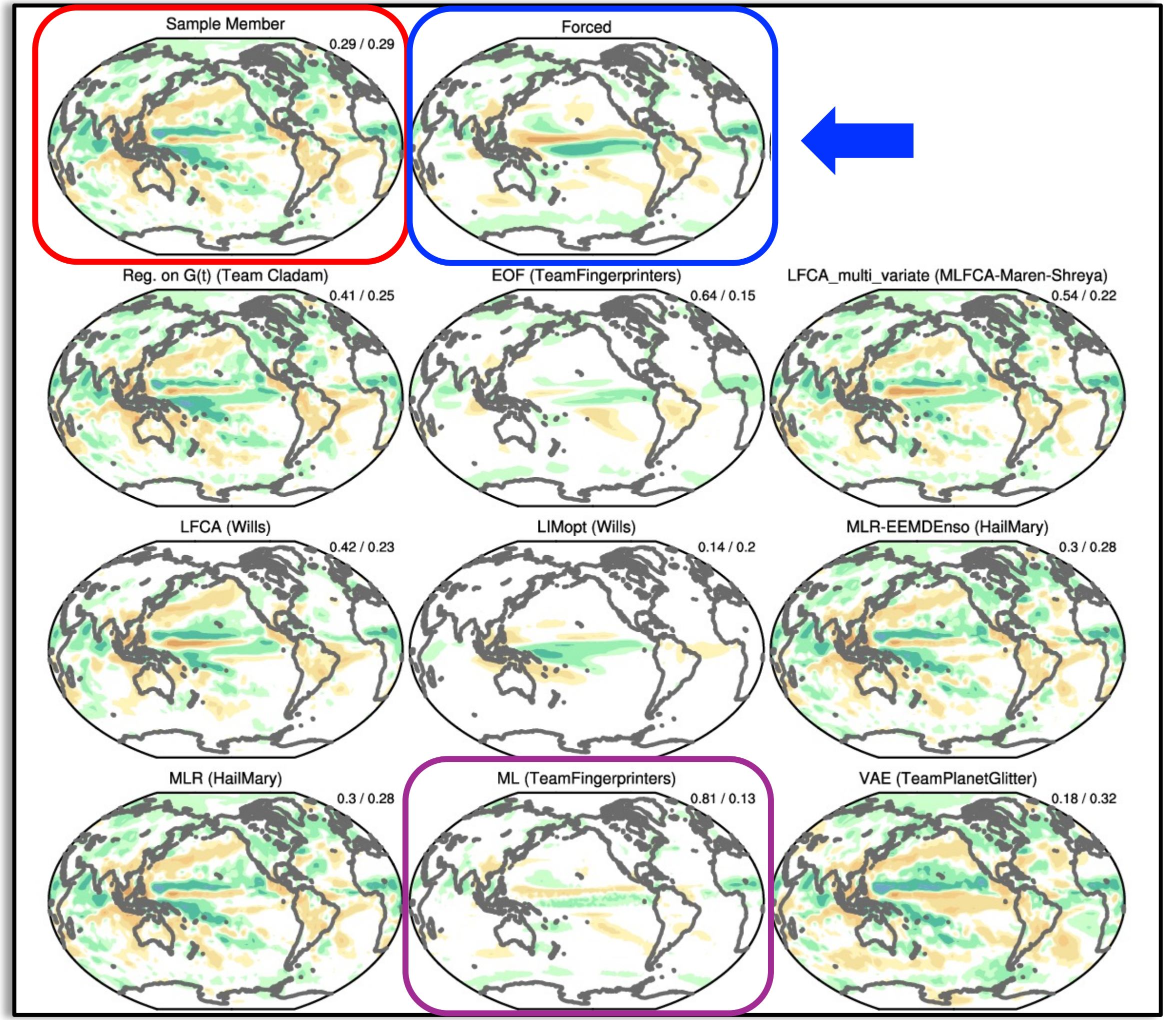


Aug. 29-31, 2023

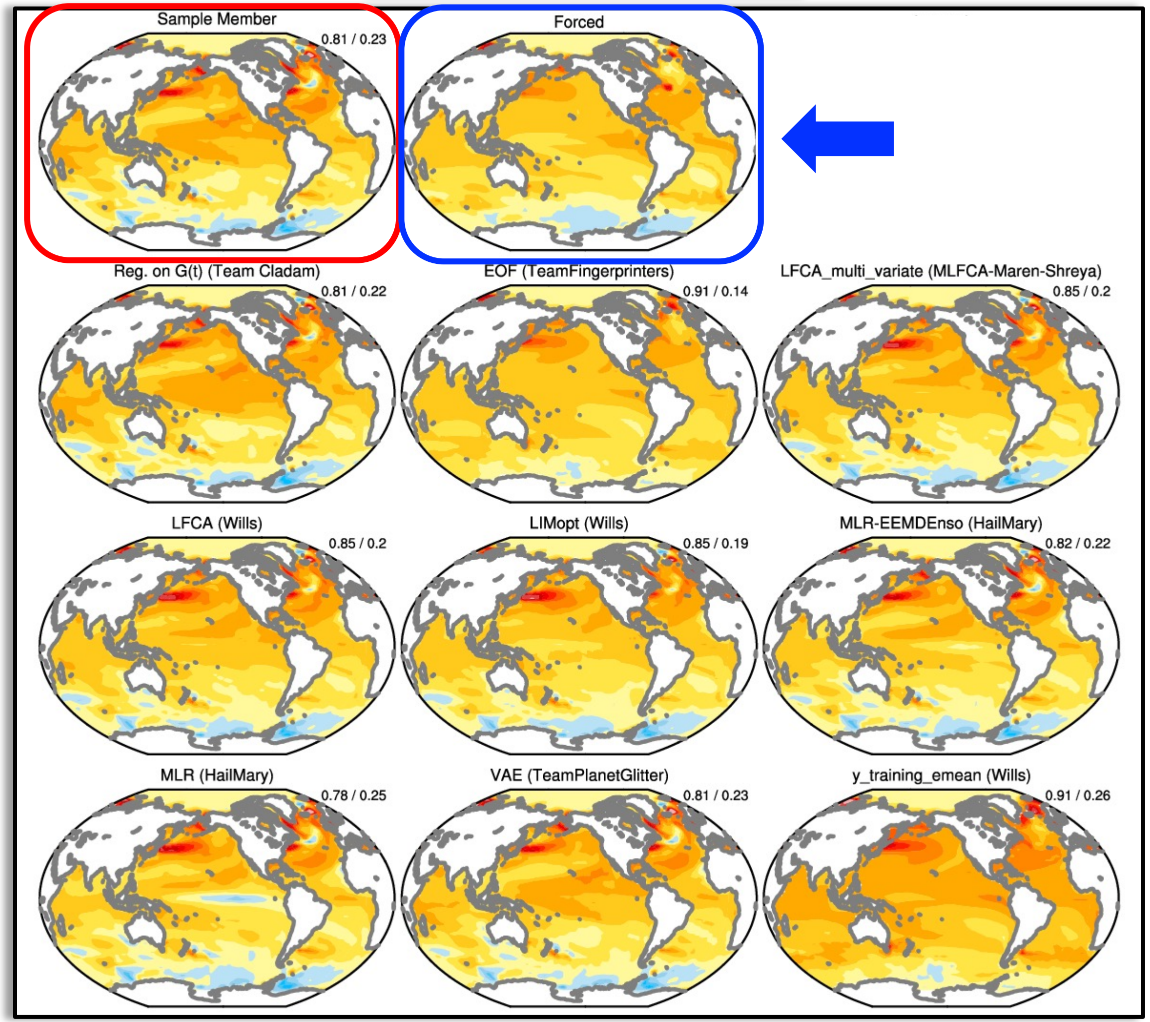
Linear Trends (1950-2022)

Project (ForceSMIP)

Precipitation



SST



National Science Foundation

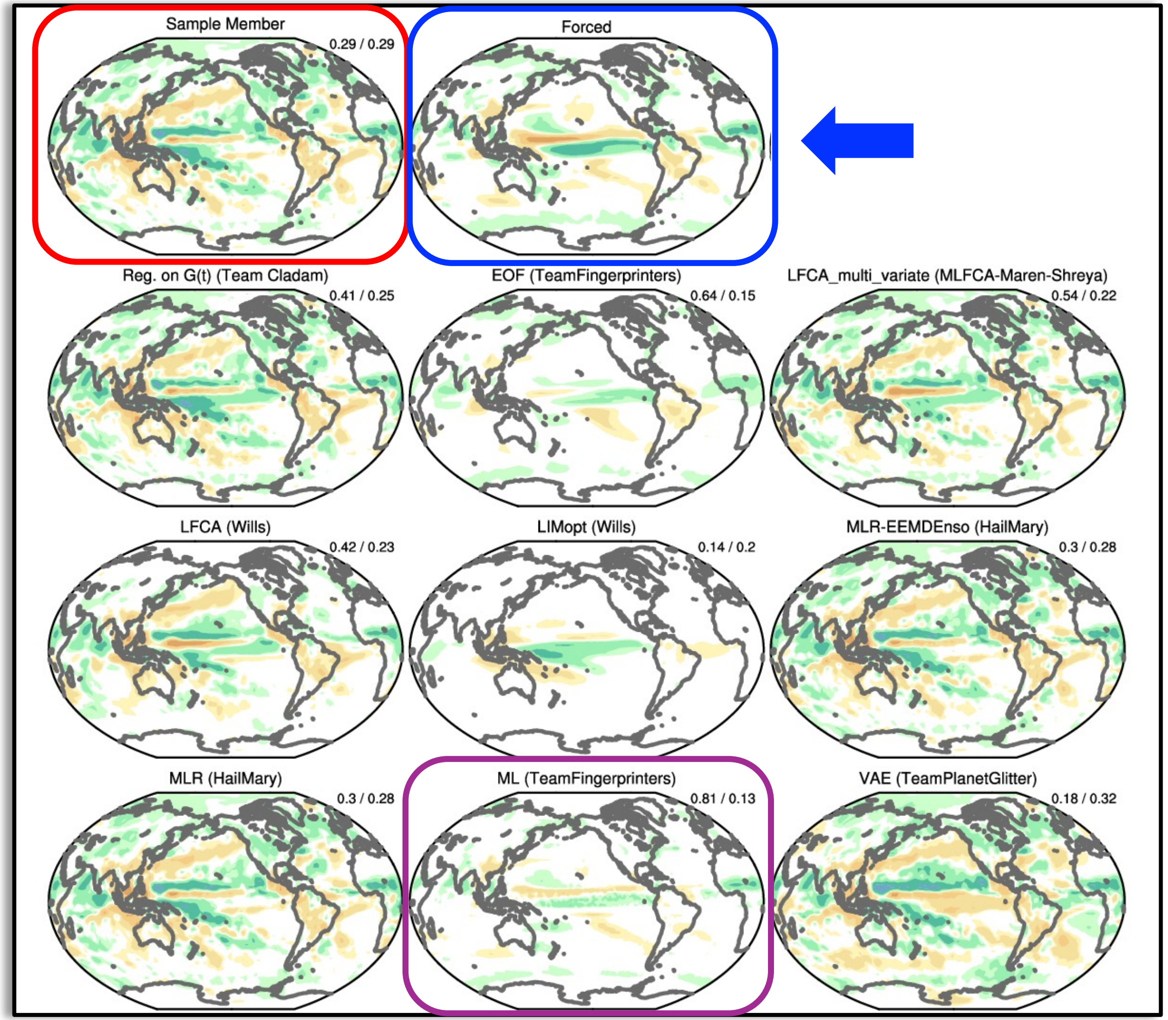
Swiss National Science Foundation

Aug. 29-31, 2023

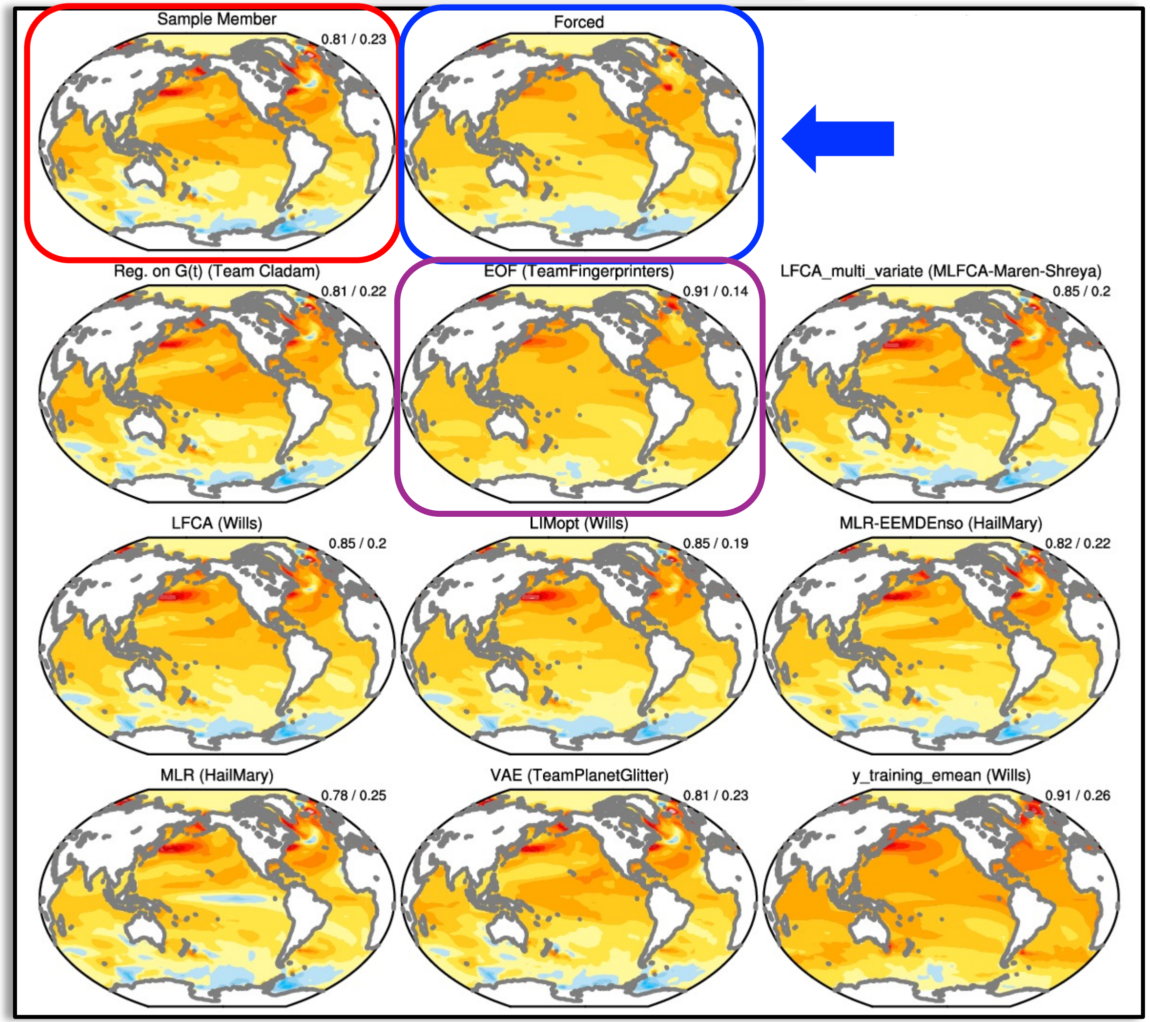
Linear Trends (1950-2022)

Project (ForceSMIP)

Precipitation



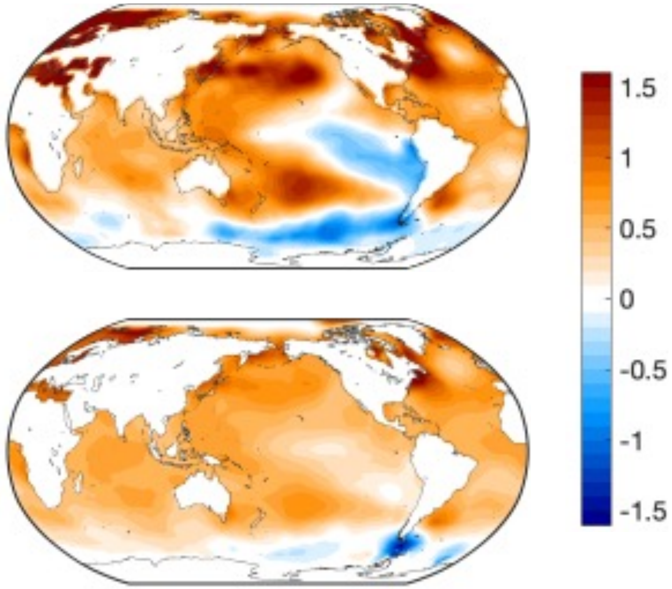
SST



Application to Observed SST Trends (1979-2022)

Raw

Estimated forced response (9 methods)



See Robb Wills' Poster on Friday for further details.

Challenges in comparing observed and model-simulated trends on regional scales

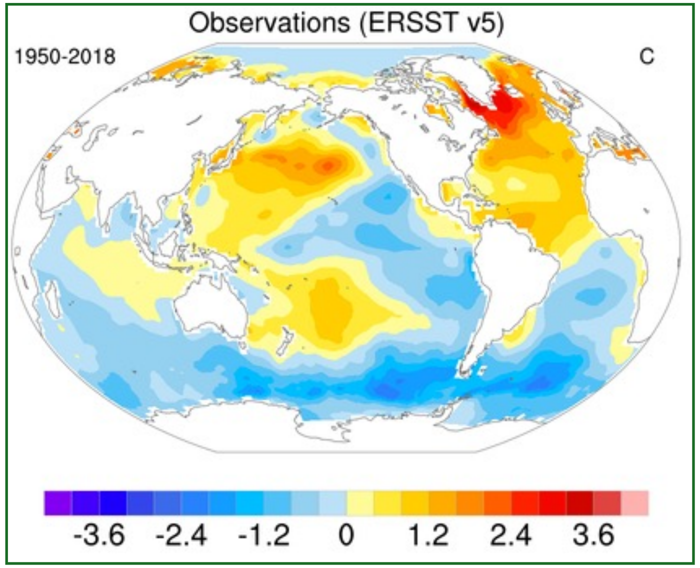
Clara Deser, NCAR

"The lifetime of a trend is the time it takes to be recognized" – *Eugene Rasmussen*

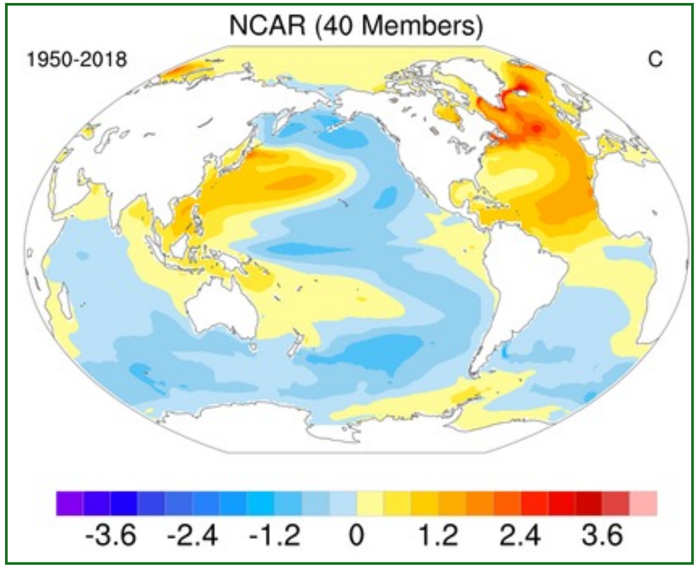
US CLIVAR Workshop "Confronting Earth System Model Trends with Observations" March 13-15, 2024.

Extra Slides

The Atlantic Multidecadal Oscillation (AMO)
Index: 10-year low-pass filtered
North Atlantic SST – Global mean SST (Trenberth and Shea, 2015)



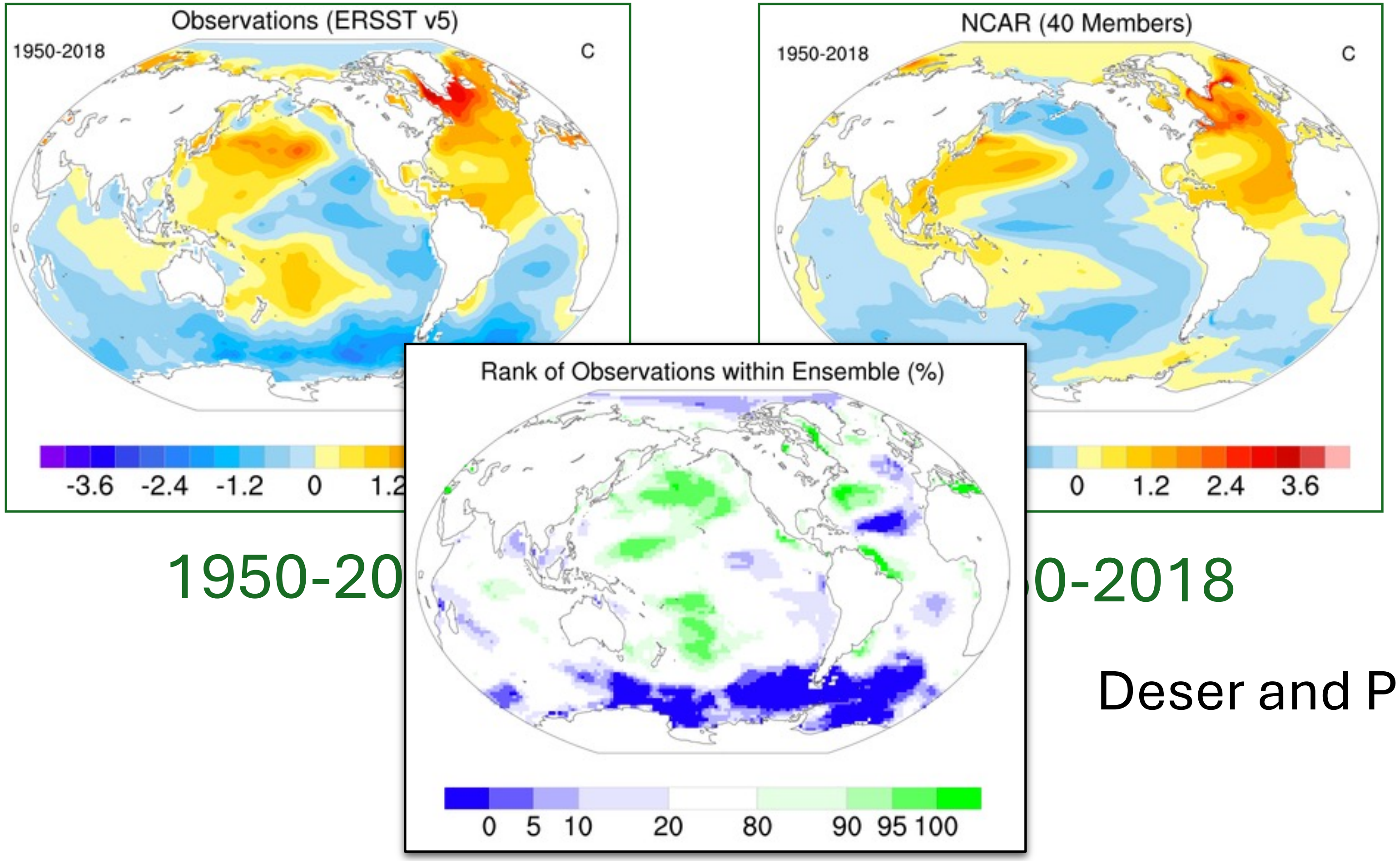
1950-2018



1950-2018

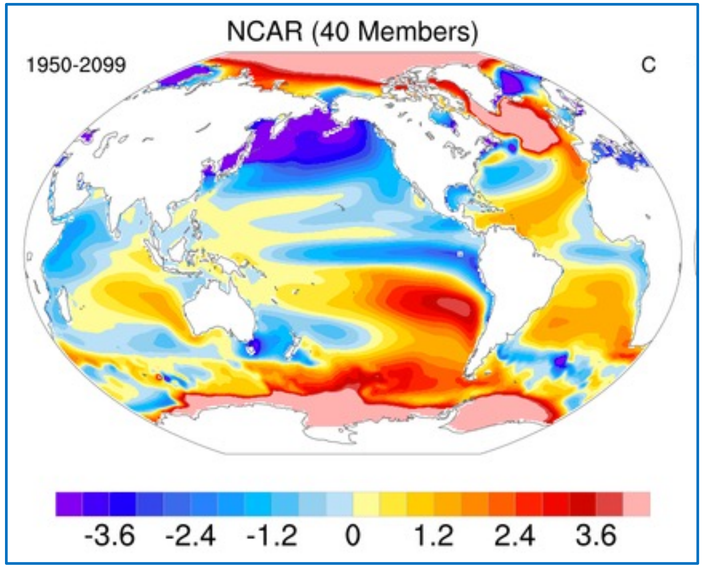
Deser and Phillips (2021)

The Atlantic Multidecadal Oscillation (AMO)
Index: 10-year low-pass filtered
North Atlantic SST – Global mean SST (Trenberth and Shea, 2015)

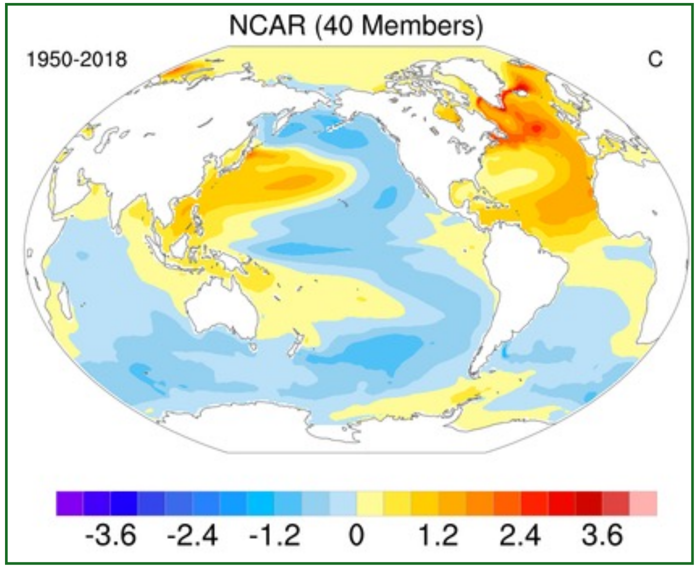


Deser and Phillips (2021)

The Atlantic Multidecadal Oscillation (AMO)
Index: 10-year low-pass filtered
North Atlantic SST – Global mean SST (Trenberth and Shea, 2015)



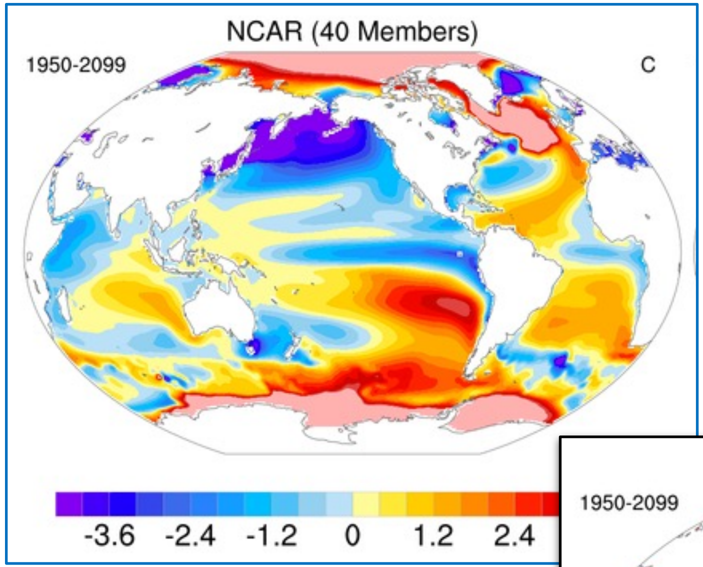
1950-2099



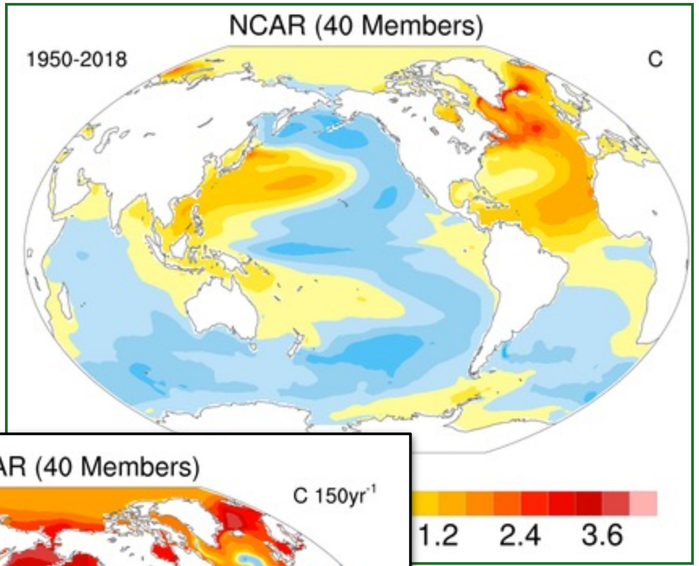
1950-2018

Deser and Phillips (2021)

The Atlantic Multidecadal Oscillation (AMO)
Index: 10-year low-pass filtered
North Atlantic SST – Global mean SST (Trenberth and Shea, 2015)

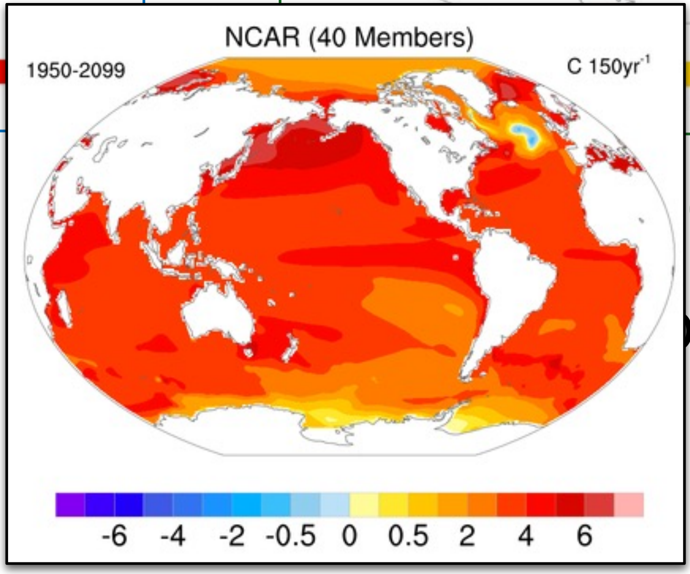


1950-2099



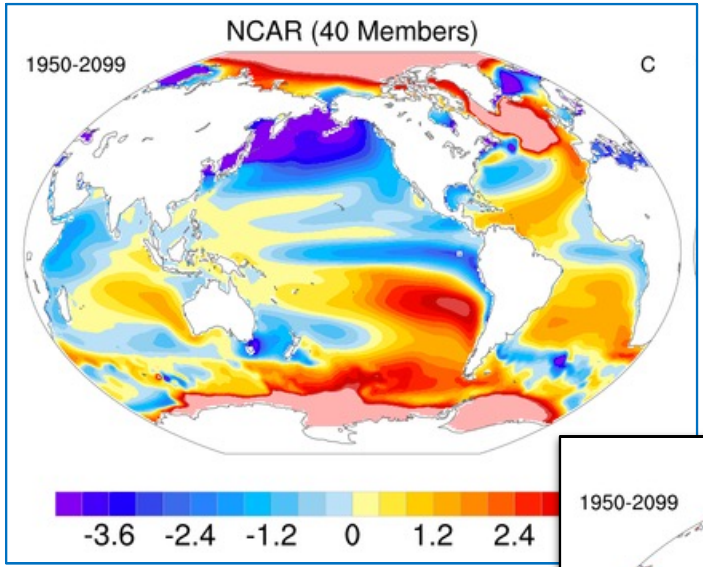
018

eser and Phillips (2021)

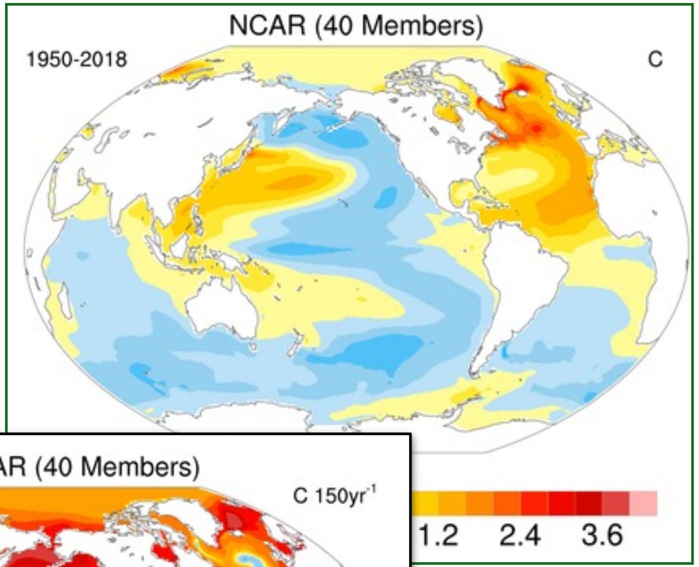


Linear Trend

The Atlantic Multidecadal Oscillation (AMO)
Index: 10-year low-pass filtered
North Atlantic SST – Global Mean SST (Trenberth and Shea, 2015)



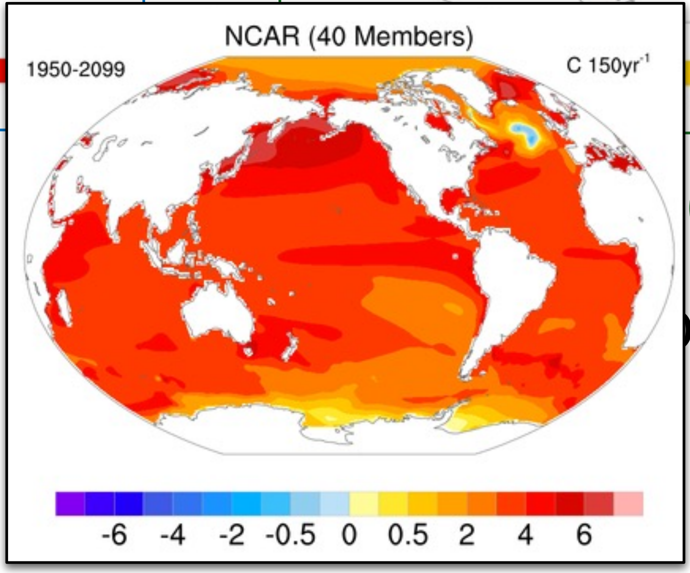
1950-2099



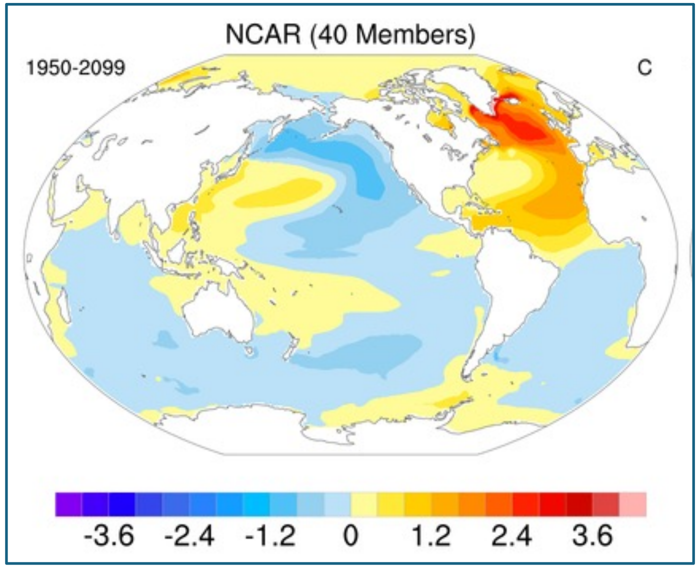
018

eser and Phillips (2021)

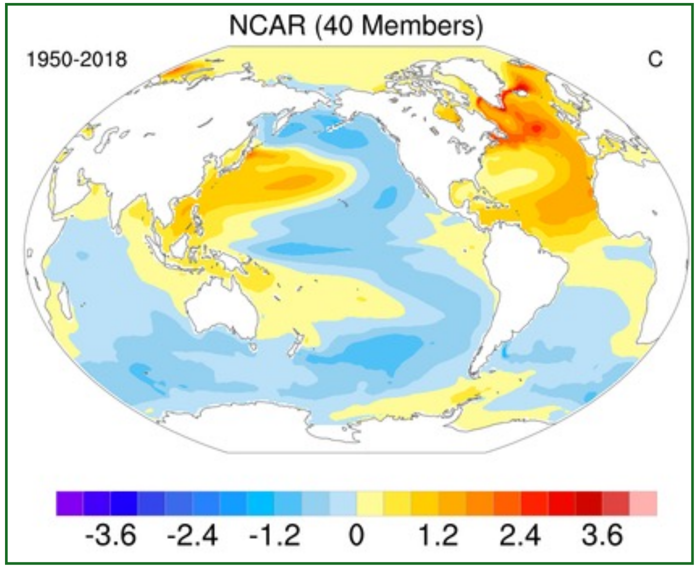
Linear Trend



The **Internal** Atlantic Multidecadal Oscillation (AMO)
Index: 10-year low-pass filtered
North Atlantic SST' (ensemble-mean removed at each location)



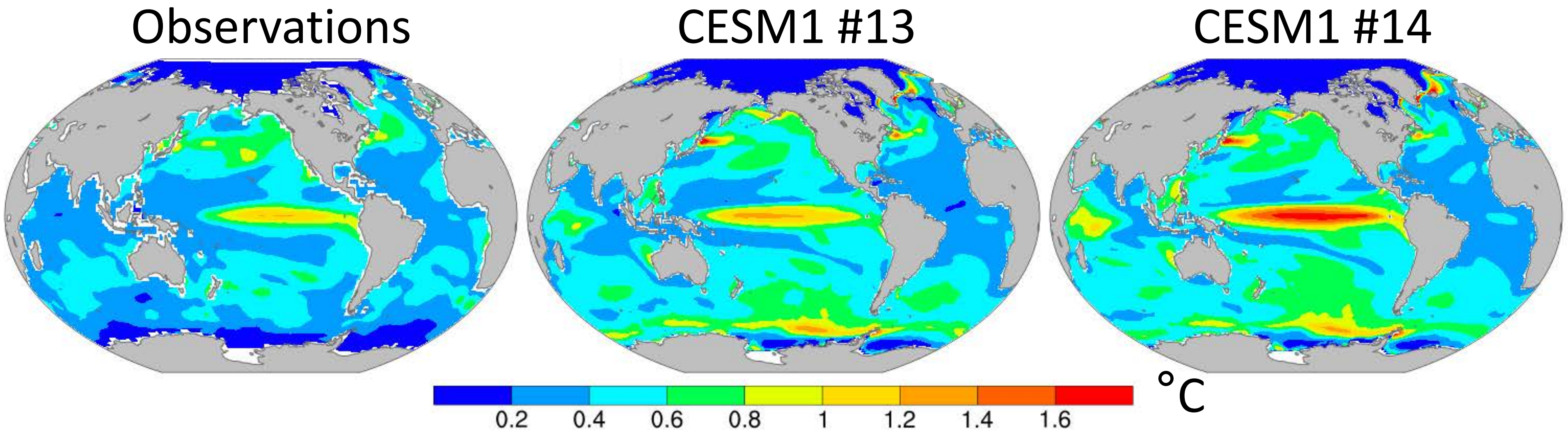
1950-2099



1950-2018

Deser and Phillips (2021)

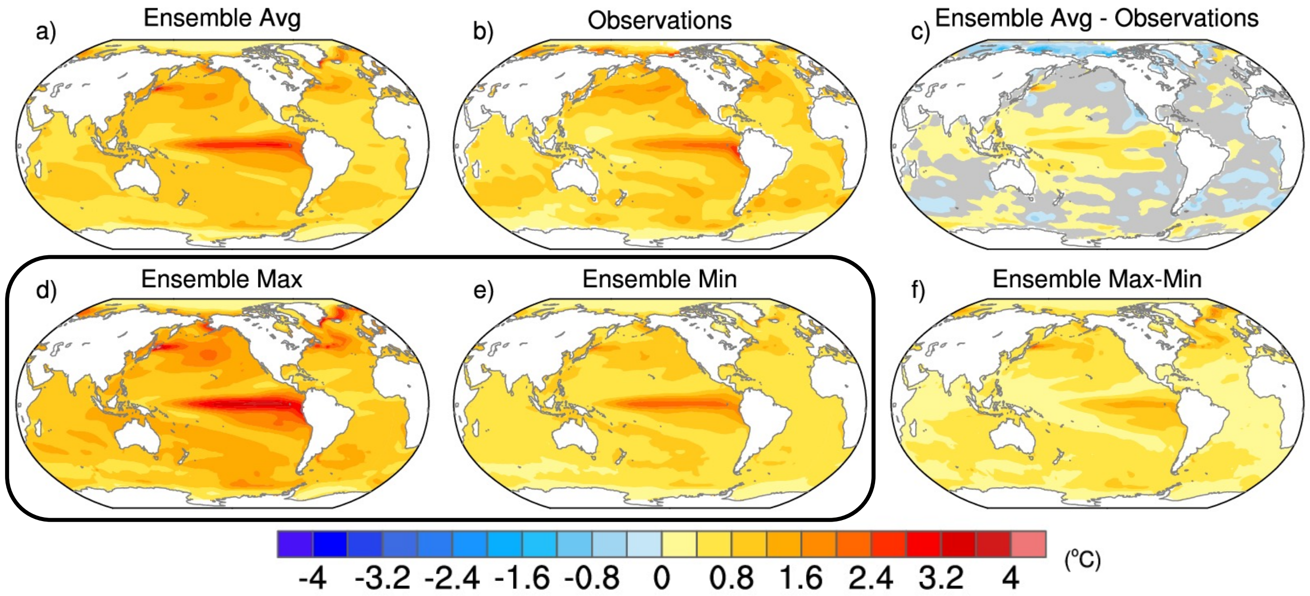
Detrended SST anomaly standard deviation (1920-2015)



Wittenberg et al., 2009; Newman et al., 2011; Deser et al., 2012

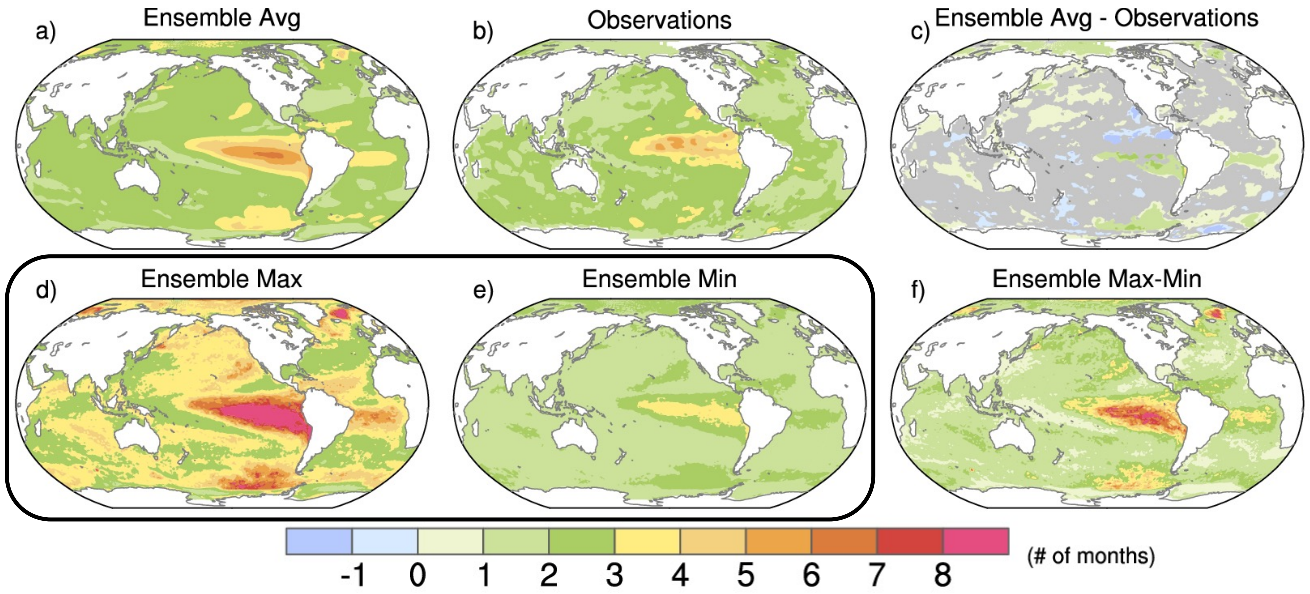
Marine Heatwaves (> 90th percentile SST extremes, 1950-2022)

CESM2
(100-member SMILE)



Composite
Intensity

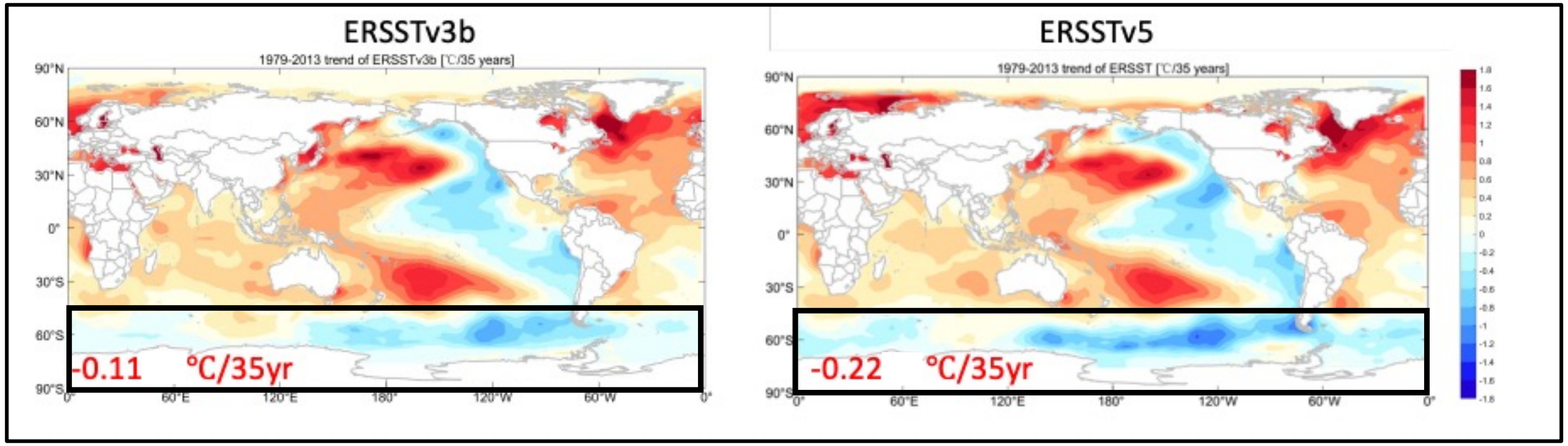
CESM2
(100-member SMILE)



Composite
Duration

Deser et al. (2024)

SST Trends 1979-2013



Kang et al. 2023