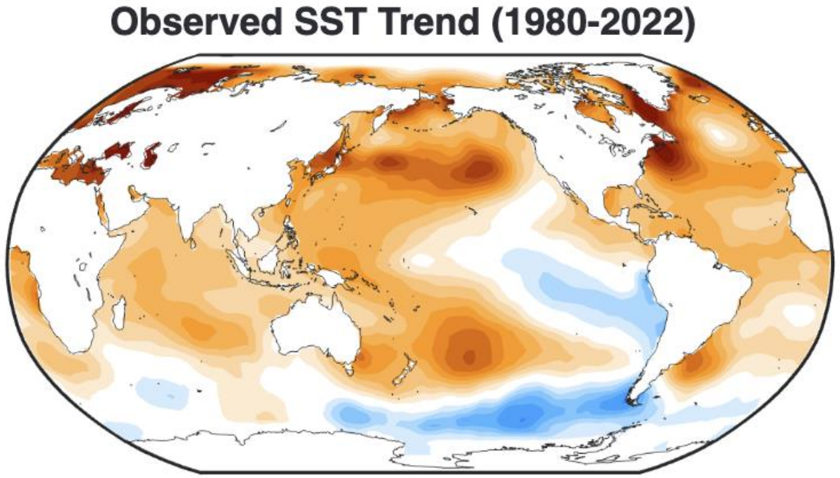
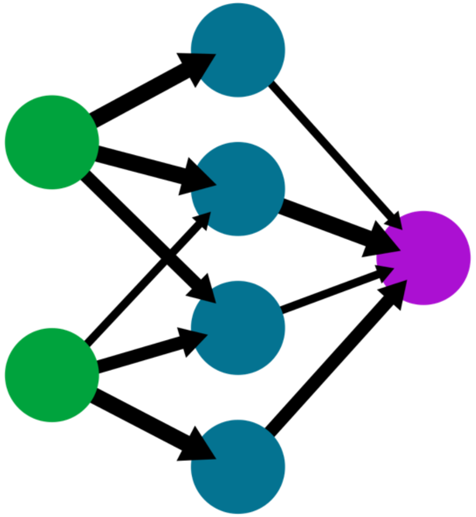
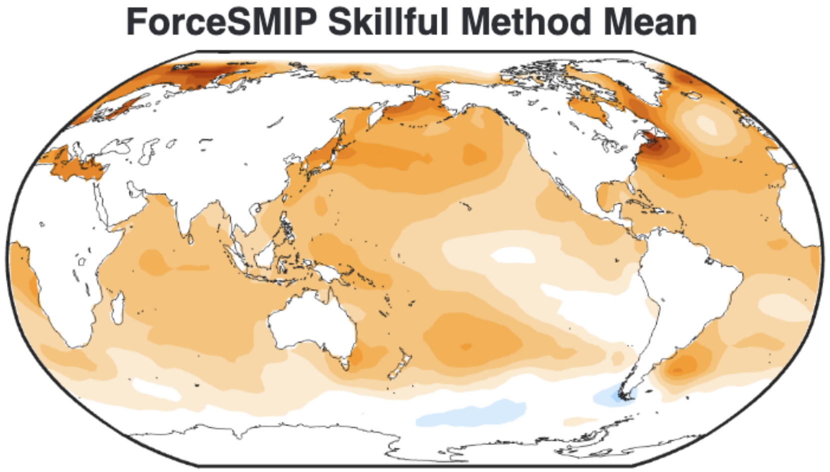
Forced Component Estimation Statistical Method Intercomparison Project (ForceSMIP)







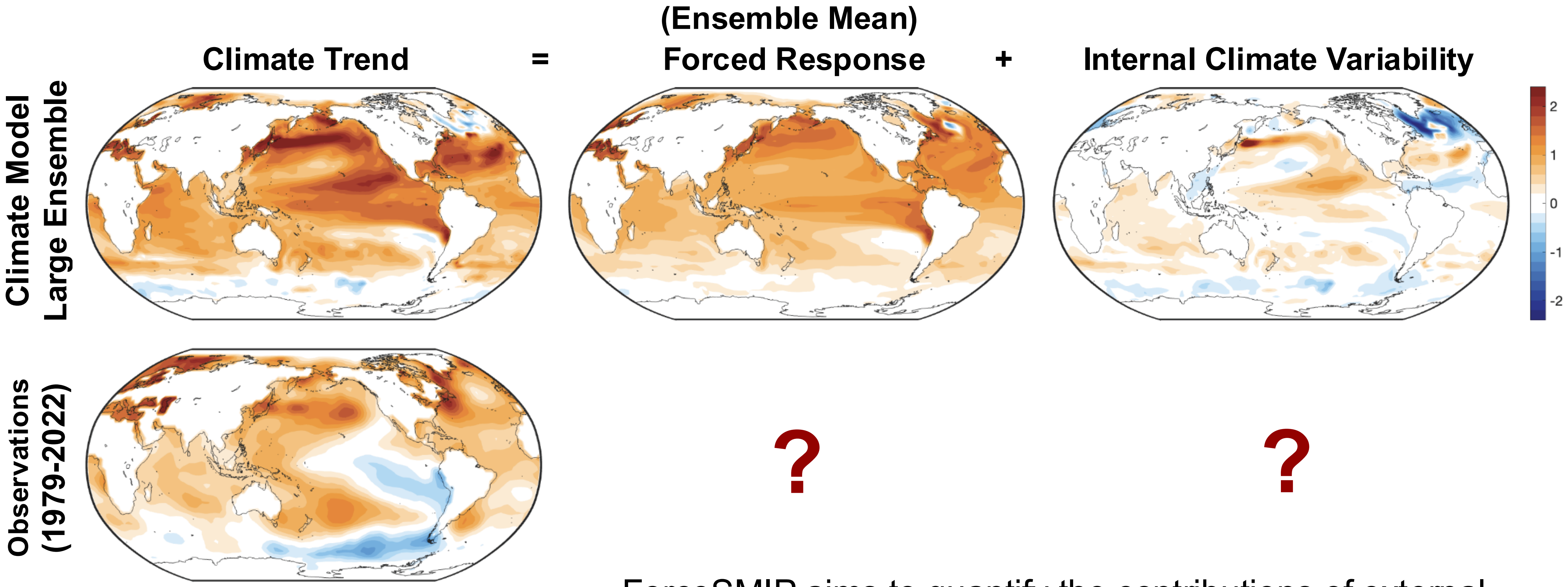
Robb Jnglin Wills[1], Clara Deser[2], Karen McKinnon[3], Adam Phillips[2], Stephen Po-Chedley[4], Sebastian Sippel[5], Anna Merrifield[1], and ForceSMIP contributors

[1]ETH Zurich, [2]NCAR, [3]UCLA, [4]LLNL, [5]University of Leipzig

EGU 2025, CL4.10, 30th April 2025

Sharing is encouraged

How can we best estimate the forced and unforced components of observed climate changes, as is done with large ensembles?



ForceSMIP aims to quantify the contributions of external forcing and internal variability in observations, using statistical and machine learning methods

Wills et al. 2025, preprint coming soon

ForceSMIP Project Hackathon and Contributors

Organizers: Robert Jnglin Wills[1], Clara Deser[2], Karen McKinnon[3], Adam Phillips[2], Stephen Po-Chedley[4], Sebastian Sippel[5]

Contributors: Constantin Bône[6], Céline Bonfils[4], Gustau Camps-Valls[7], Stephen Cropper[3], Charlotte Connolly[8], Shiheng Duan[4], Homer Durand[7], Alexander Feigin[9], Martin Fernandez[8], Guillaume Gastineau[6], Andrey Gavrilov[7,9], Emily Gordon[8], Moritz Günther[10], Maren Höver[11,1], Sergey Kravtsov[11], Yan-Ning Kuo[12], Justin Lien[13], Gavin Madakumbra[3], Anna Merrifield[1], Nathan Mankovich[7], Matt Newman[14], Jamin Rader[8], Jia-Rui Shi[15], Sang-Ik Shin[14,18], Gherardo Varando[7], Tristan Williams[7]

[1]ETH Zürich, [2]NCAR, [3]UCLA, [4]LLNL, [5]University of Leipzig, [6]LOCEAN, [7]University of Valencia, [8]Colorado State University, [9]Institute of Applied Physics, RAS, [10]Stanford, [11]Oxford, [12]MPI-Meteorology, [13]University of Wisconsin Milwaukee, [14]Cornell, [15]Tohoku University, [16]NOAA PSL, [17]WHOI, [18]CIRES, University of Colorado

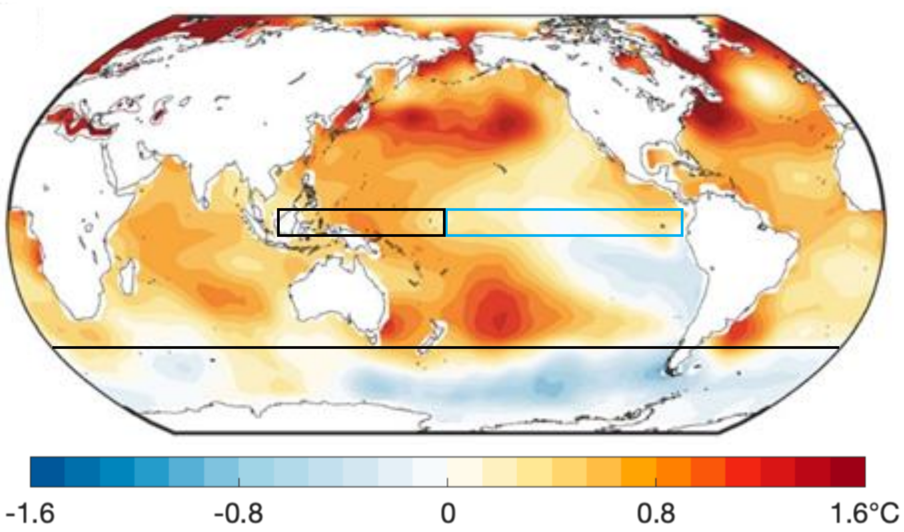
Swiss National Science Foundation · NSF · NCAR · Department of Energy · Lawrence Livermore National Laboratory · CLIVAR · UCLA

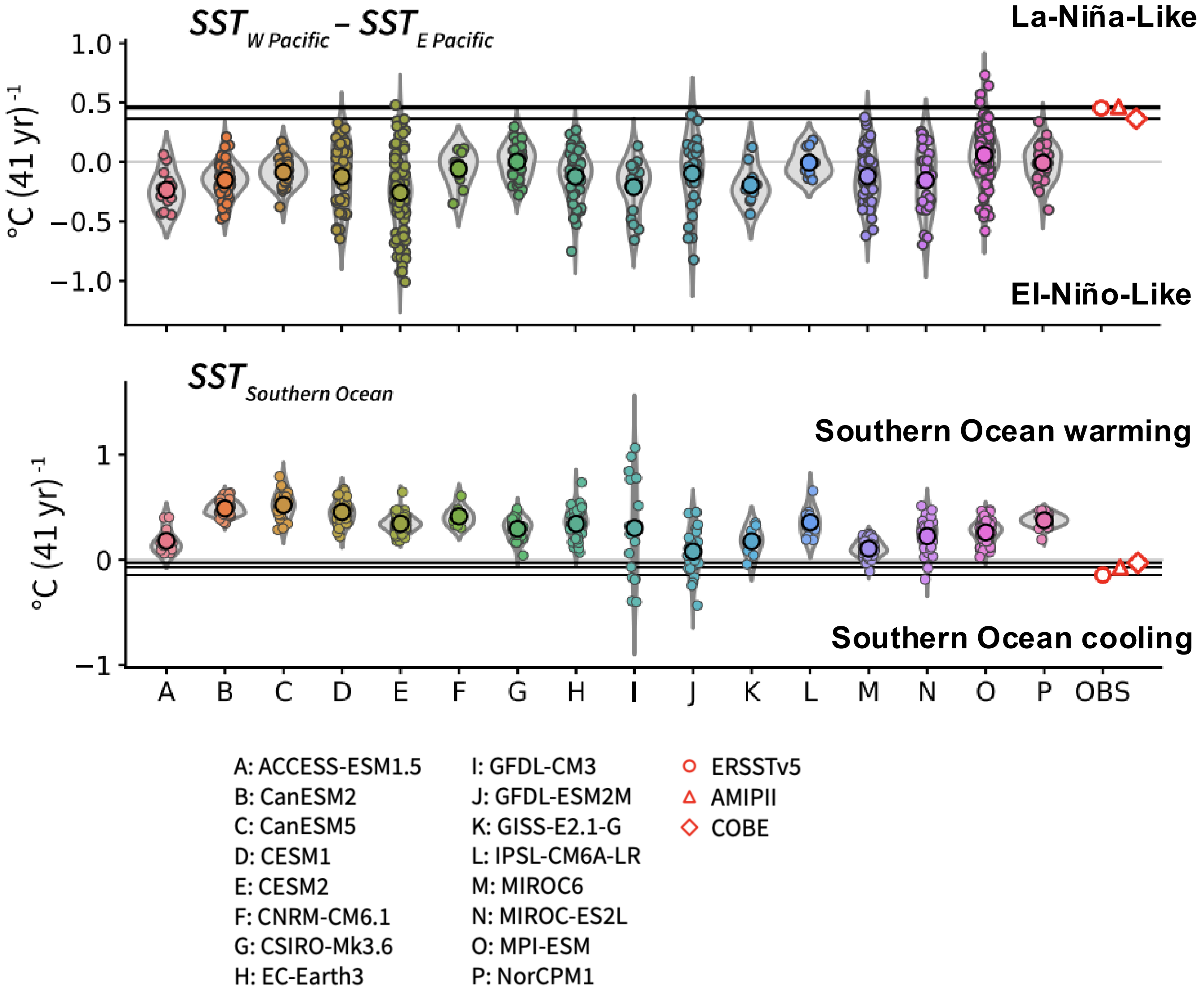
ForceSMIP Hackathon
Aug. 29-31, 2023
NCAR and ETH Zurich

First, some motivation: Open questions related to separating the forced response from internal variability

Are systematic discrepancies between observed and modeled SST trends forced or unforced?

Observed SST Change (1979-2020)

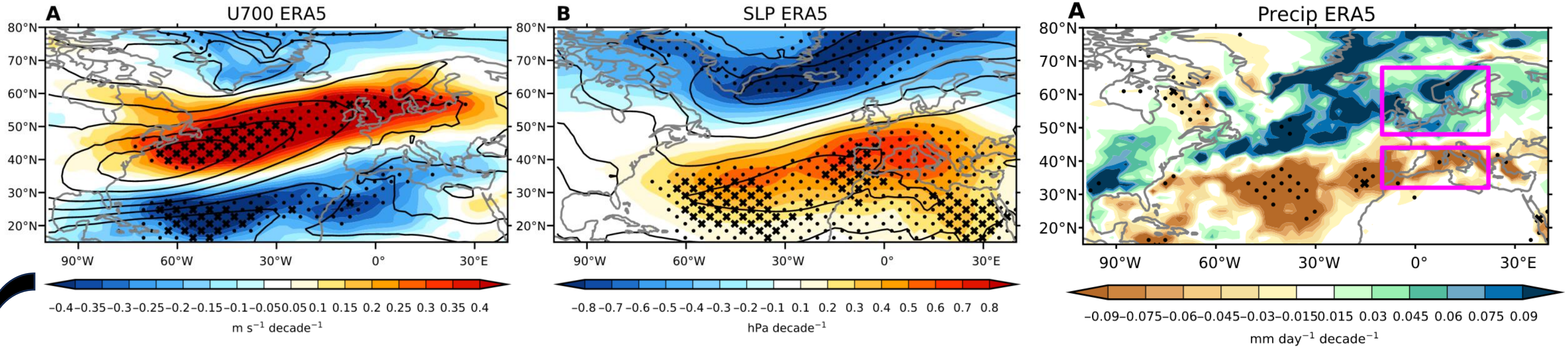


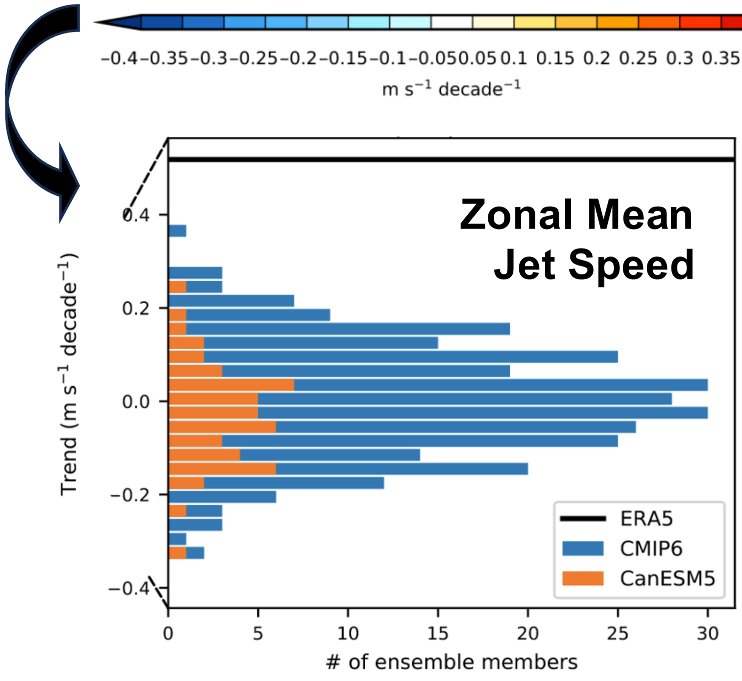


- No member out of 16 large ensembles simulates both the observed tropical Pacific SST gradient strengthening and the Southern Ocean cooling
- The discrepancy could be a bias in the forced response, a bias in the amplitude of multi-decadal variability, or both

Wills et al. 2022

Observed jet speed, SLP, and precip trends (1951-2020) are outside the model distribution. Forced or Unforced?





Reanalysis

- CMIP6 models cannot reproduce the observed strengthening of the North Atlantic jet and related SLP and precip trends
- Knowing whether models have biases in their forced responses or internal variability is critical information for near-future water resource planning, e.g., in Southern Europe

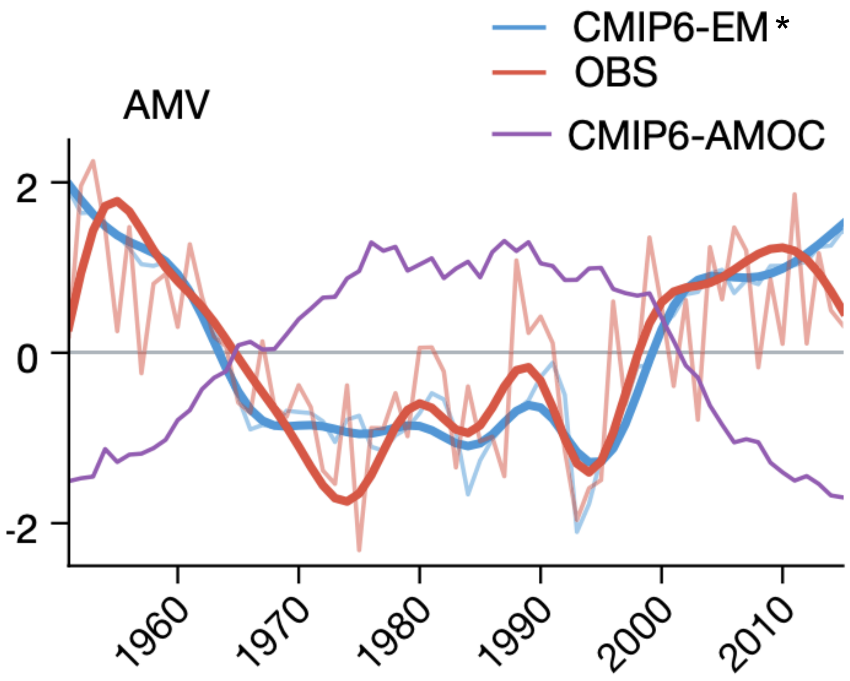
Blackport & Fyfe 2022; see also Simpson, Shaw et al. 2025

Is observed Atlantic multi-decadal variability forced or unforced? Recent studies have come to opposite conclusions

Article

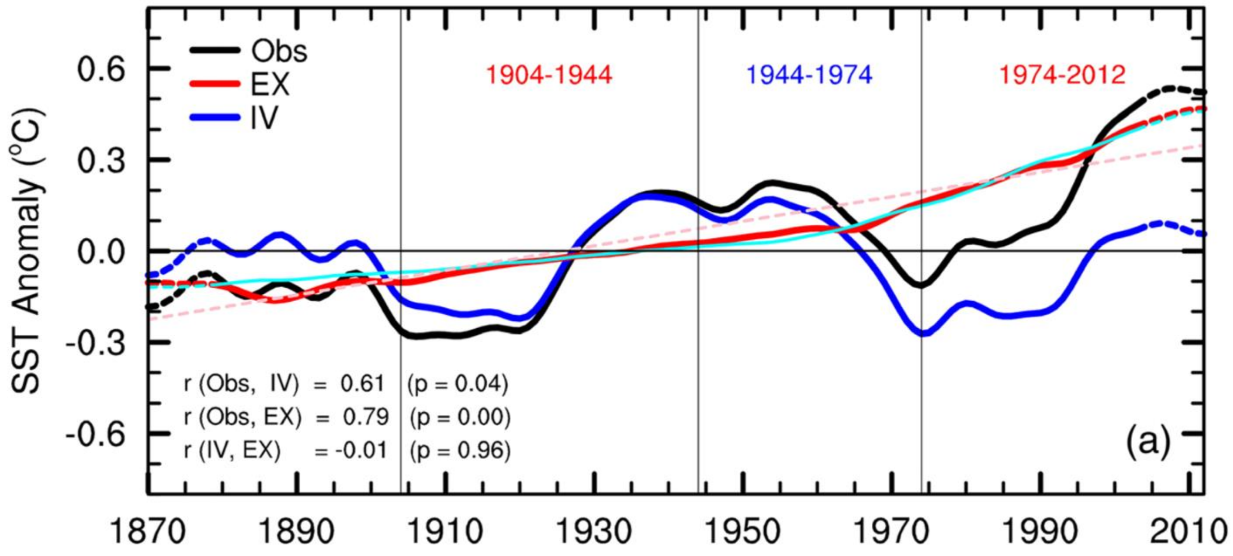
Tropical Atlantic multidecadal variability is dominated by external forcing

Chengfei He[1]✉, Amy C. Clement[1], Sydney M. Kramer[2], Mark A. Cane[3], Jeremy M. Klavans[2], Tyler M. Fenske[1] & Lisa N. Murphy[1]



Quantifying Contributions of Internal Variability and External Forcing to Atlantic Multidecadal Variability Since 1870

Minhua Qin[1], Aiguo Dai[2], and Wenjian Hua[1]

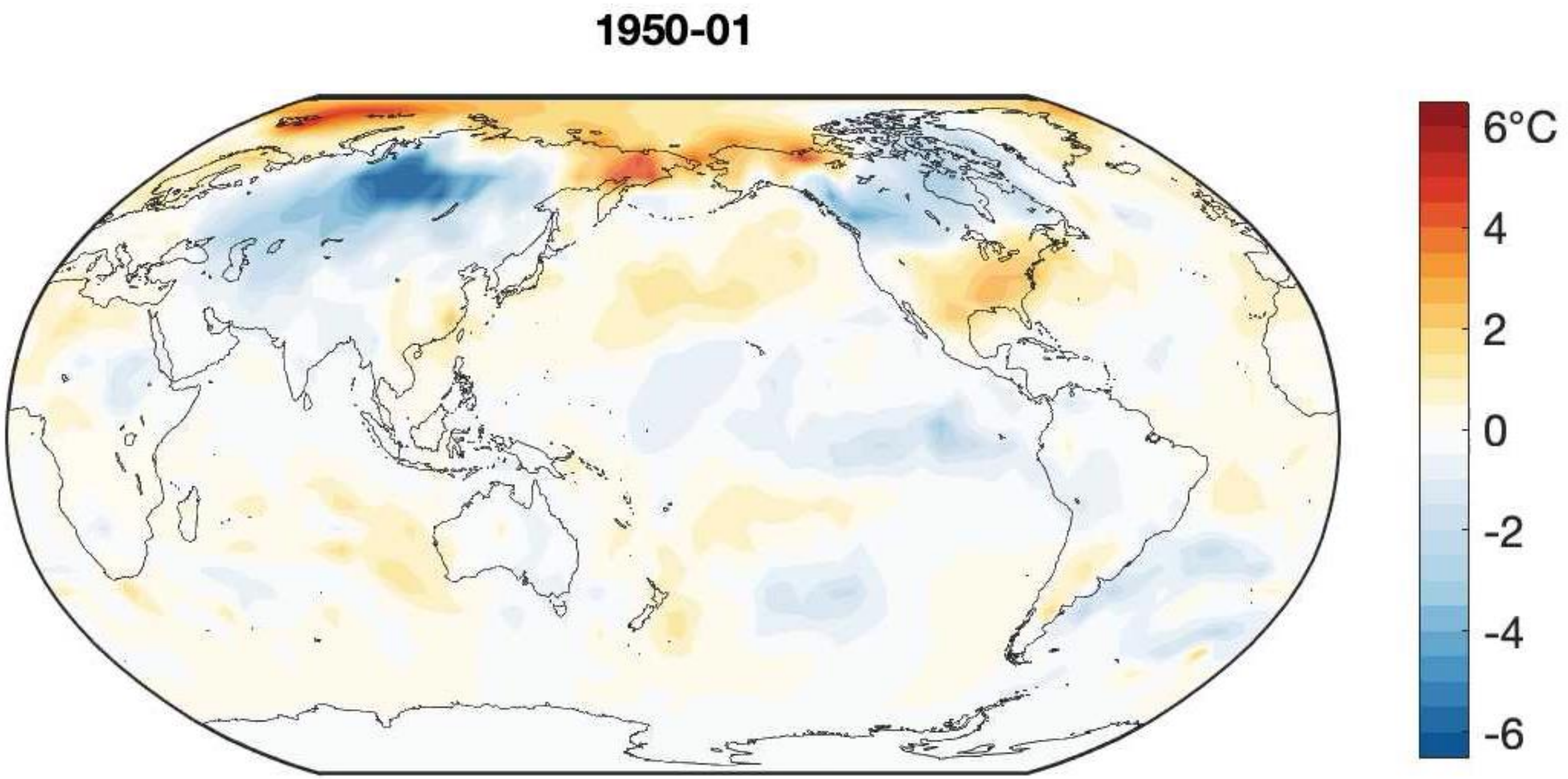


* multiplied by 4.82 to correct for apparent signal-to-noise problems

He et al. 2023; Qin et al. 2020

ForceSMIP: Framework, Datasets, and Statistical Methods

The ForceSMIP Challenge: Take a single realization of the climate system and estimate the forced response



Observed (HadCRUT4) 3-month running mean surface temperature anomalies

Forced Response: Defined here as the spatiotemporally evolving anomalies (from a reference period) due to all anthropogenic and natural (volcanic, solar) external forcing

The ForceSMIP Challenge: Take a single realization of the climate system and estimate the forced response

Training data: 5 large ensembles

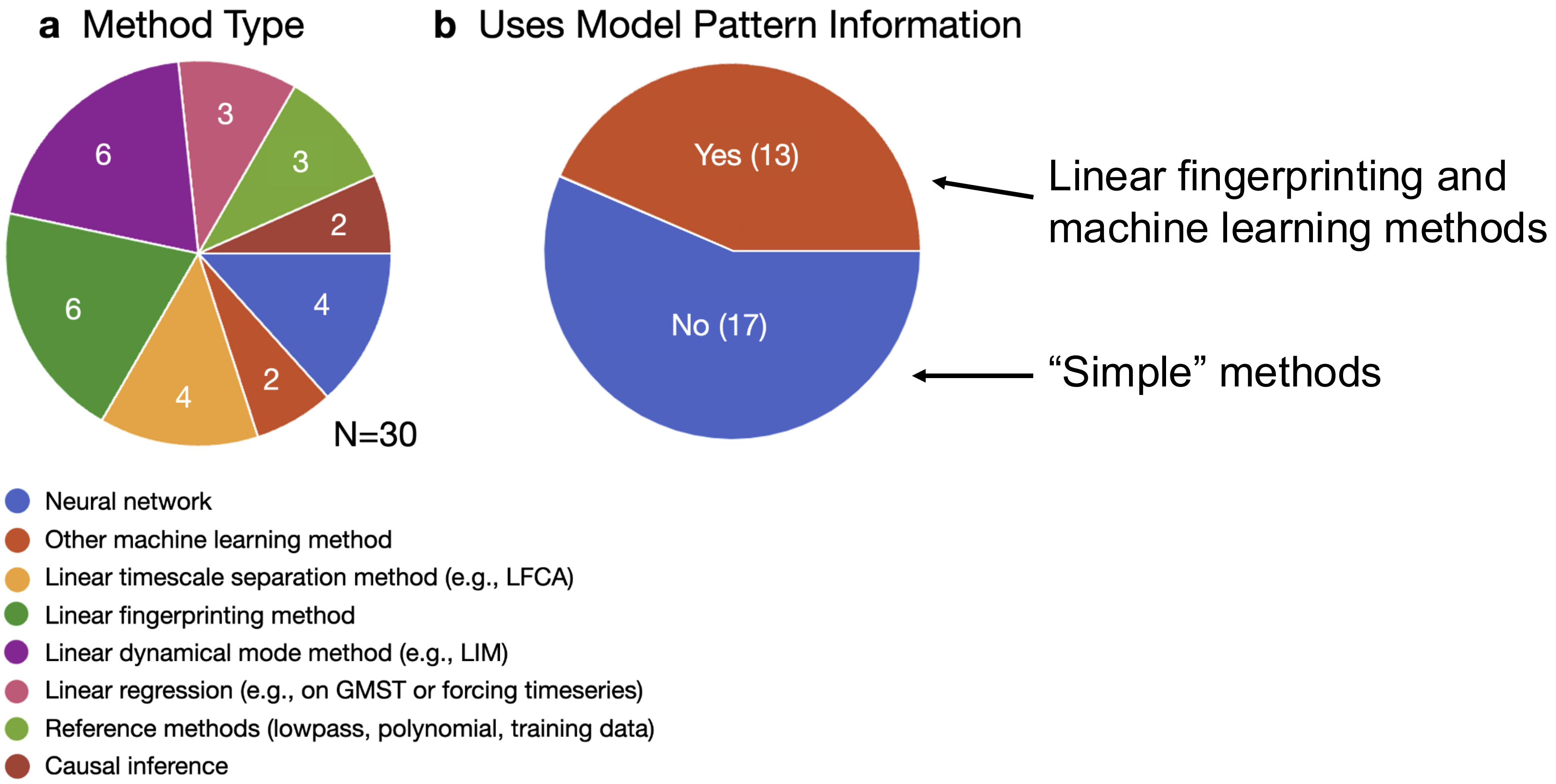
Evaluation data: Unlabeled individual realizations from 9 climate models (including 5 not part of the training data) and 1 observational product

- 8 field variables at monthly resolution, 1950-2022

How it works:

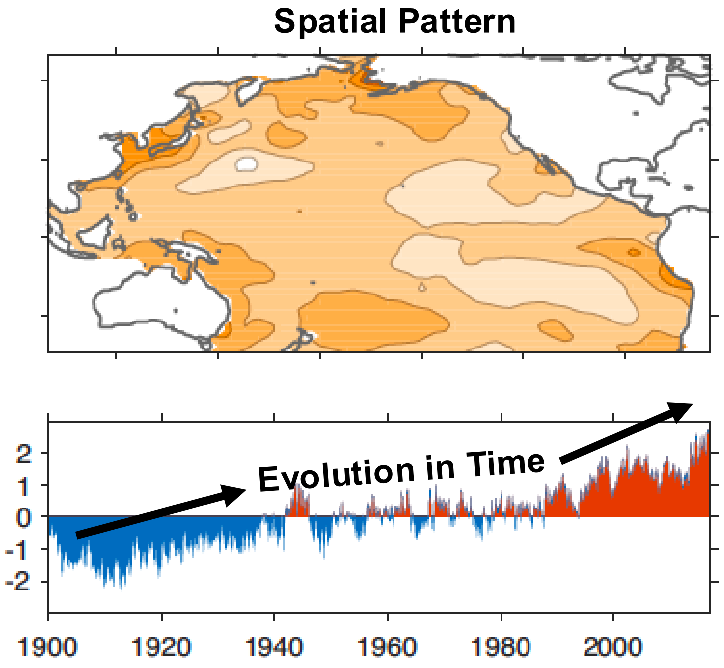
1. Participants **train** statistical methods to estimate the forced response from single realizations and apply them to the evaluation data
2. We **evaluate** how well these estimates reproduced the corresponding ensemble means (which were initially hidden from participants)
3. We examine the **observational forced response estimates** from skillful methods

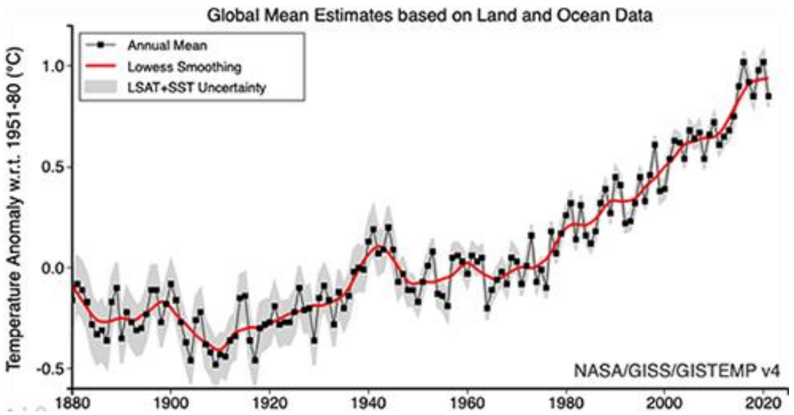
How to estimate the forced response from a single realization: Methods submitted to ForceSMIP



Wills et al. 2025, preprint coming soon

How to estimate the forced response from a single realization: Linear methods with few tunable parameters



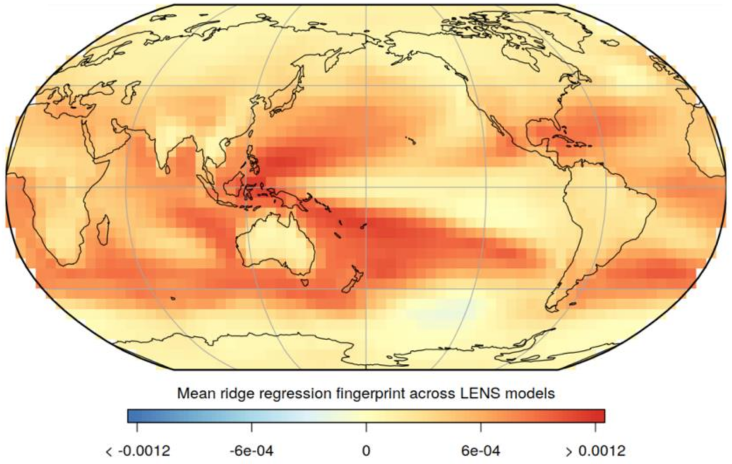


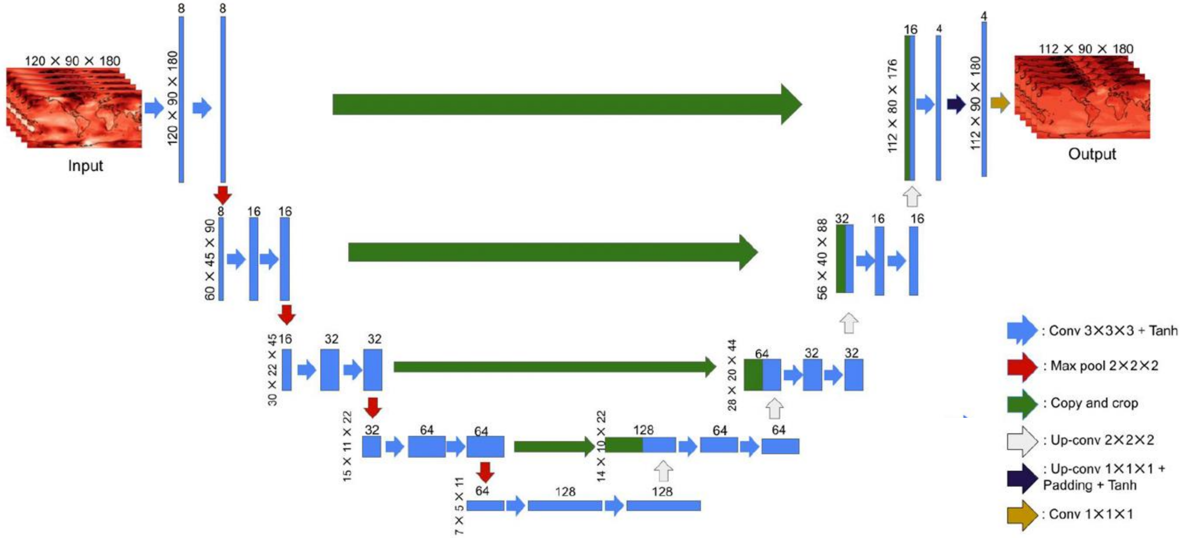
Low-frequency component analysis (LFCA): EOF-based method that estimates the forced response as the patterns evolving on the longest timescale (Wills et al. 2020)

Linear inverse models: Creates a statistical-dynamical model based on lead-lag information (Penland & Sardeshmukh 1995) and estimates the forced response as the pattern that evolves on the longest timescale (e.g., Frankignoul et al. 2017)

Linear regression on global mean or forcing timeseries: Estimates the forced response as anomalies covarying with a forcing or global-mean surface temperature timeseries

How to estimate the forced response from a single realization: Fingerprinting and machine learning methods





Linear fingerprinting methods: Determine the extent to which model-based forced response patterns show up in observations, with less weight put on noisy regions

Machine learning methods: Learns from the training model data what a forced response pattern and/or internal variability pattern looks like and applies this information to observations (or other evaluation members), e.g., Bône et al. 2024

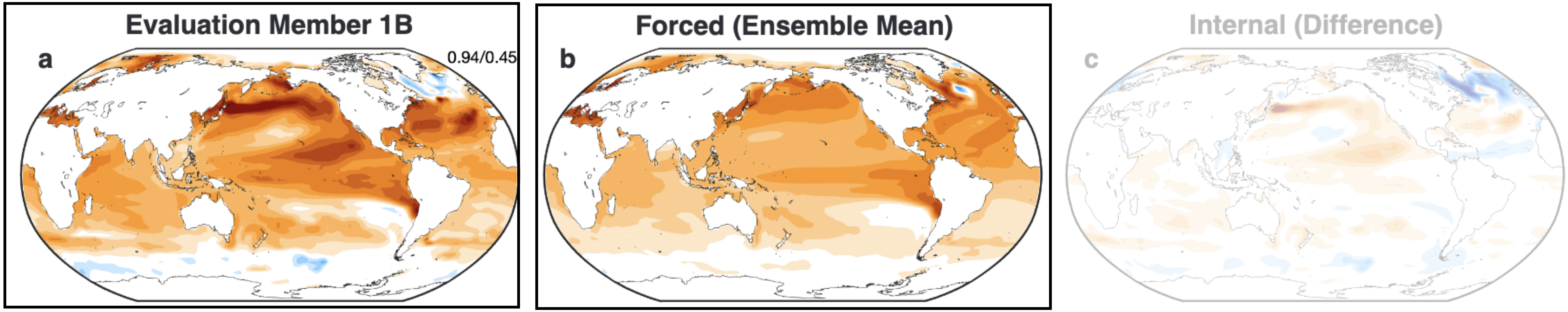
Wills, Sippel, Barnes 2020; Bône et al. 2024

Evaluating the ForceSMIP methods using large ensemble evaluation data

Evaluating estimated trends (e.g., 1980-2022 SST trends)

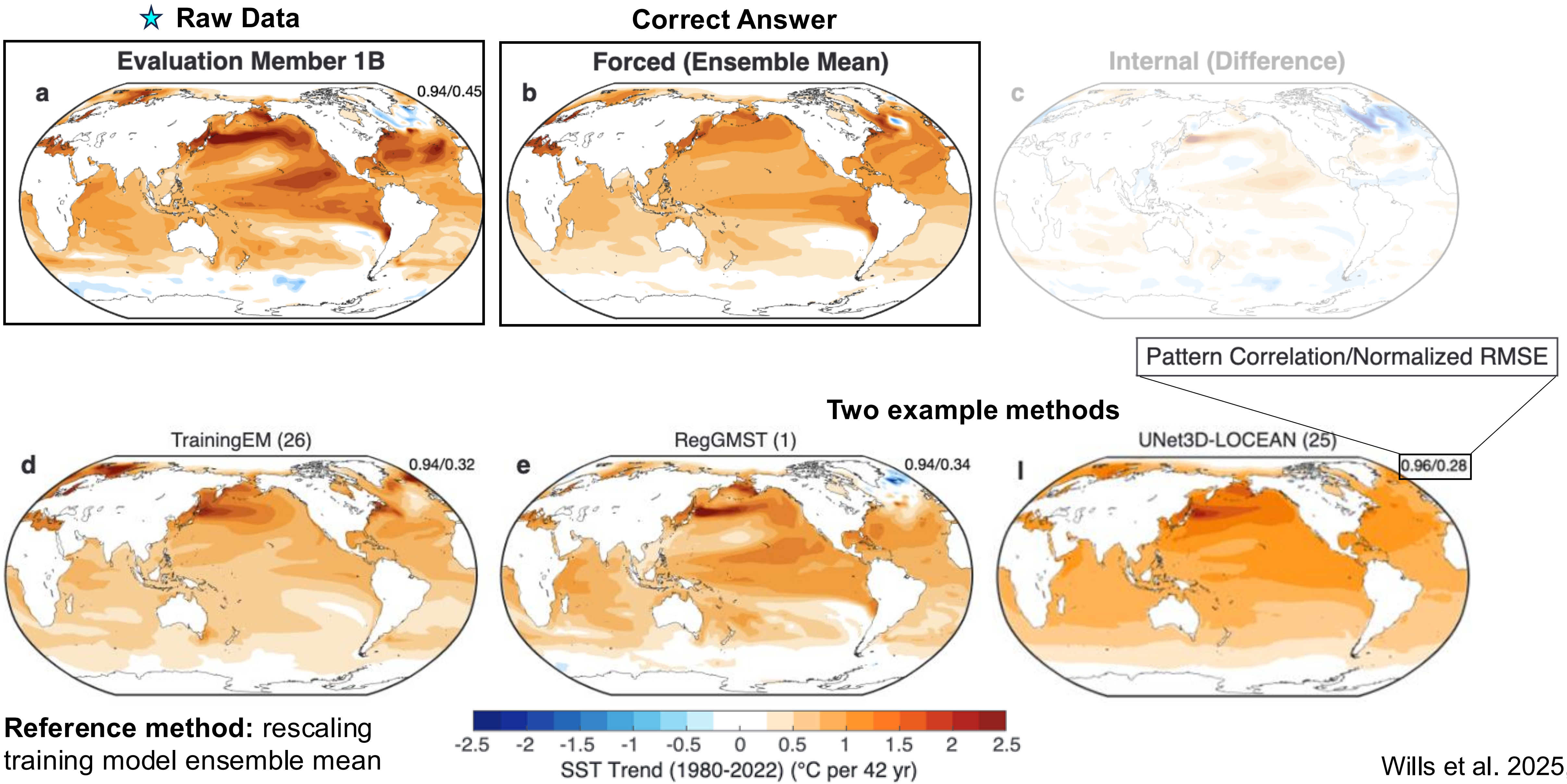
☆ Raw Data

Correct Answer



Wills et al. 2025,
preprint coming soon

Evaluating estimated trends (e.g., 1980-2022 SST trends)

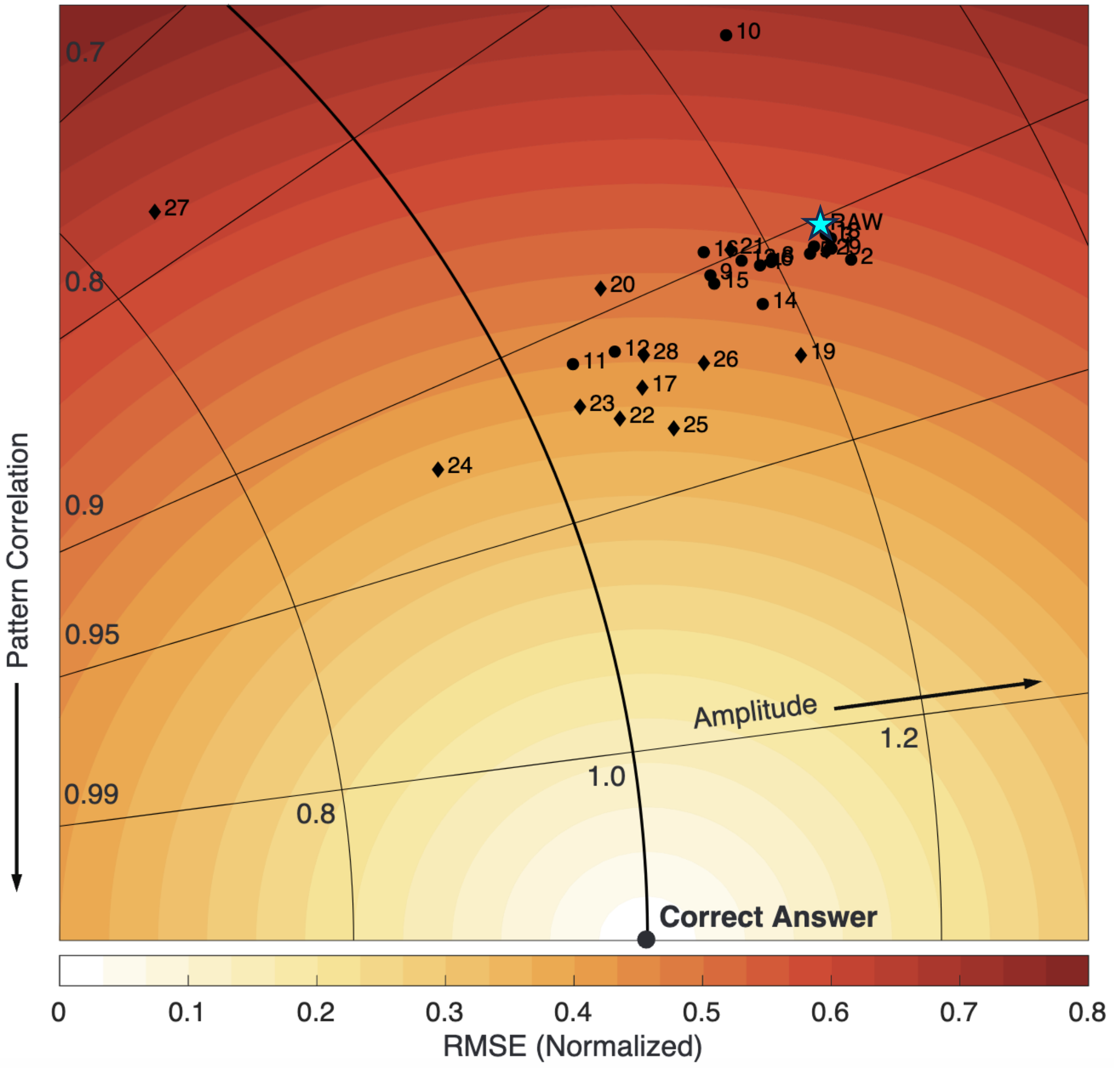


Reference method: rescaling training model ensemble mean

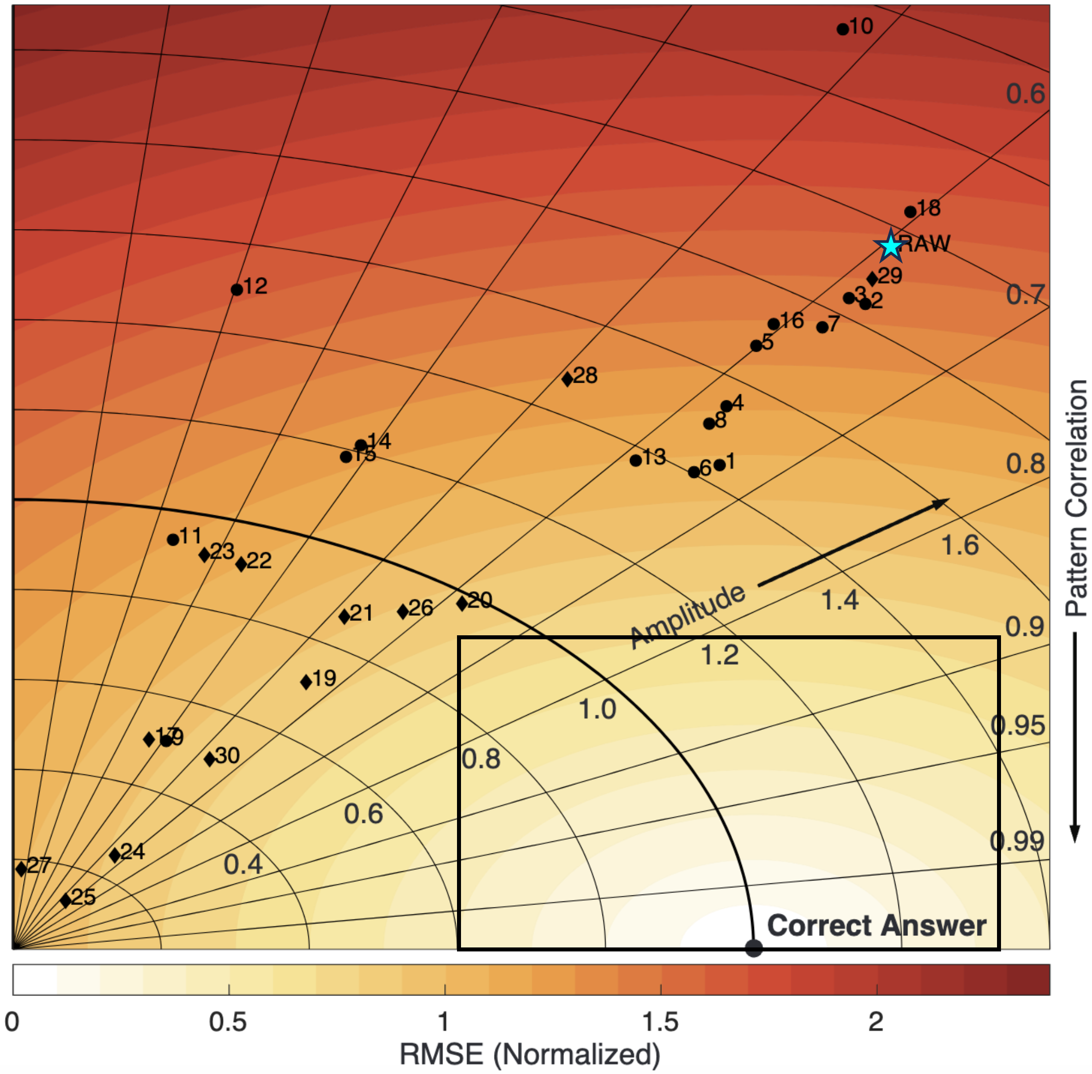
Wills et al. 2025, preprint coming soon

Evaluating estimated trends with Taylor diagrams

SST Trends (1980-2022)



Precipitation Trends (1980-2022)



Note: Left plot is zoomed into the black box region of the right plot, since SST trends are more skillful

Wills et al. 2025,
preprint coming soon

Evaluating the ForceSMIP methods: Findings

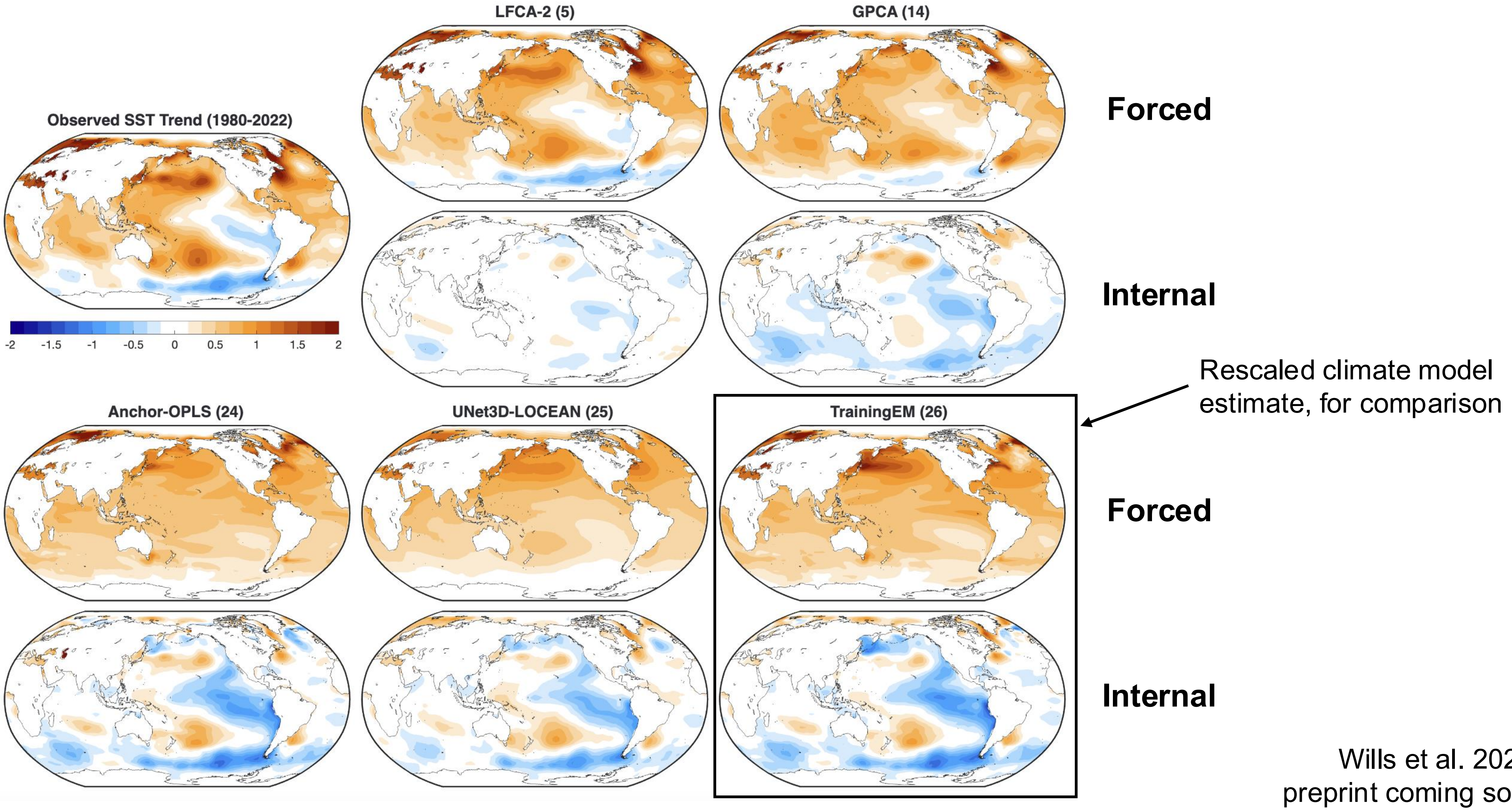
Summarizing the findings (details in the paper):

1. Most methods reduce RMSE in the forced response estimate compared to the raw data (for long-term and short-term trends, spatiotemporal variability, and large-scale indices)
2. This sometimes come at the expense of reducing the pattern correlation and/or making the amplitude to week, especially for precipitation and SLP
3. The relative skill of different methods varies between variables

Wills et al. 2025,
preprint coming soon

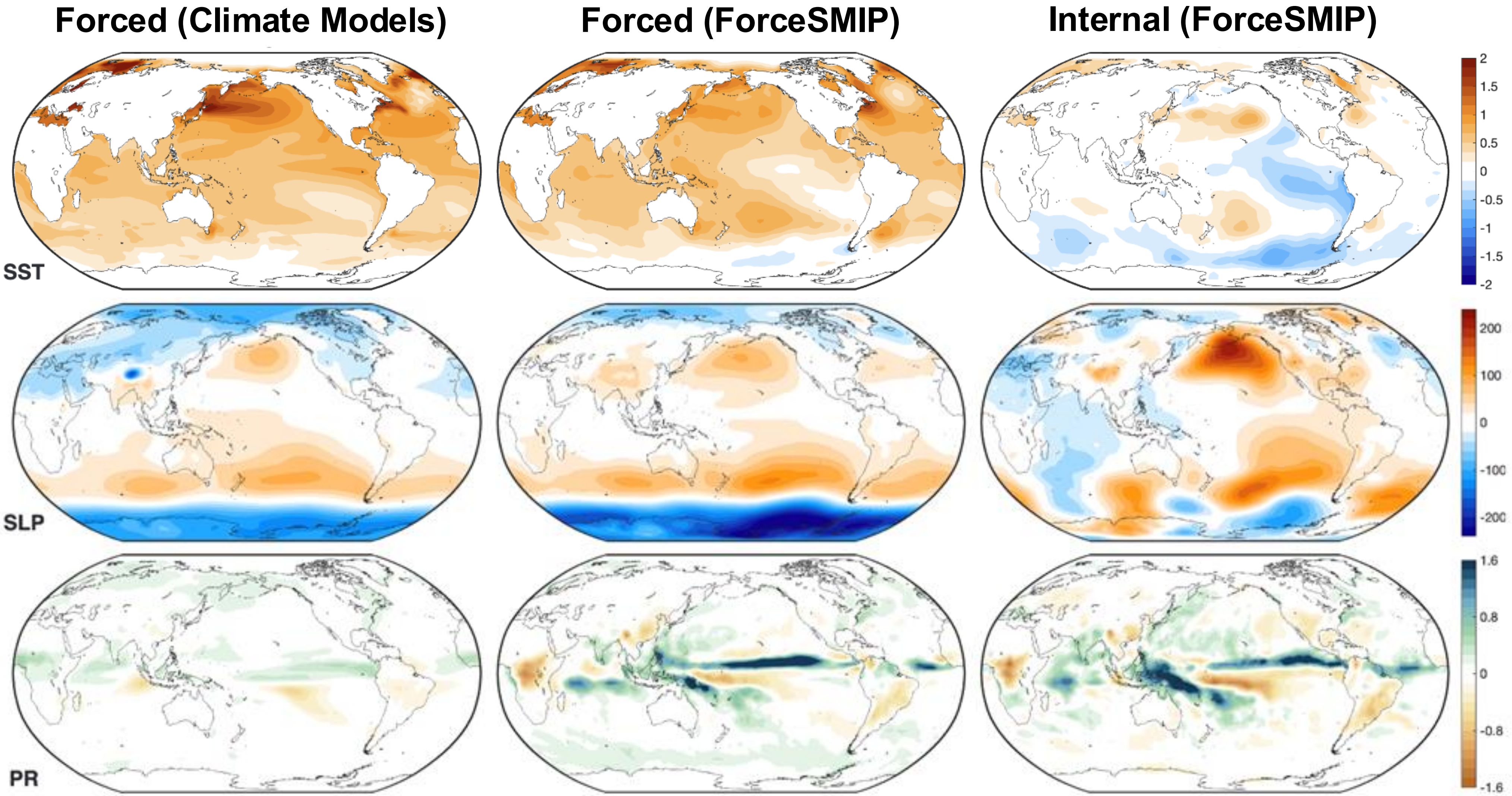
Examining the ForceSMIP estimated forced response in observations

Wide spread of forced vs. internal attribution, even amongst methods determined to be skillful in models



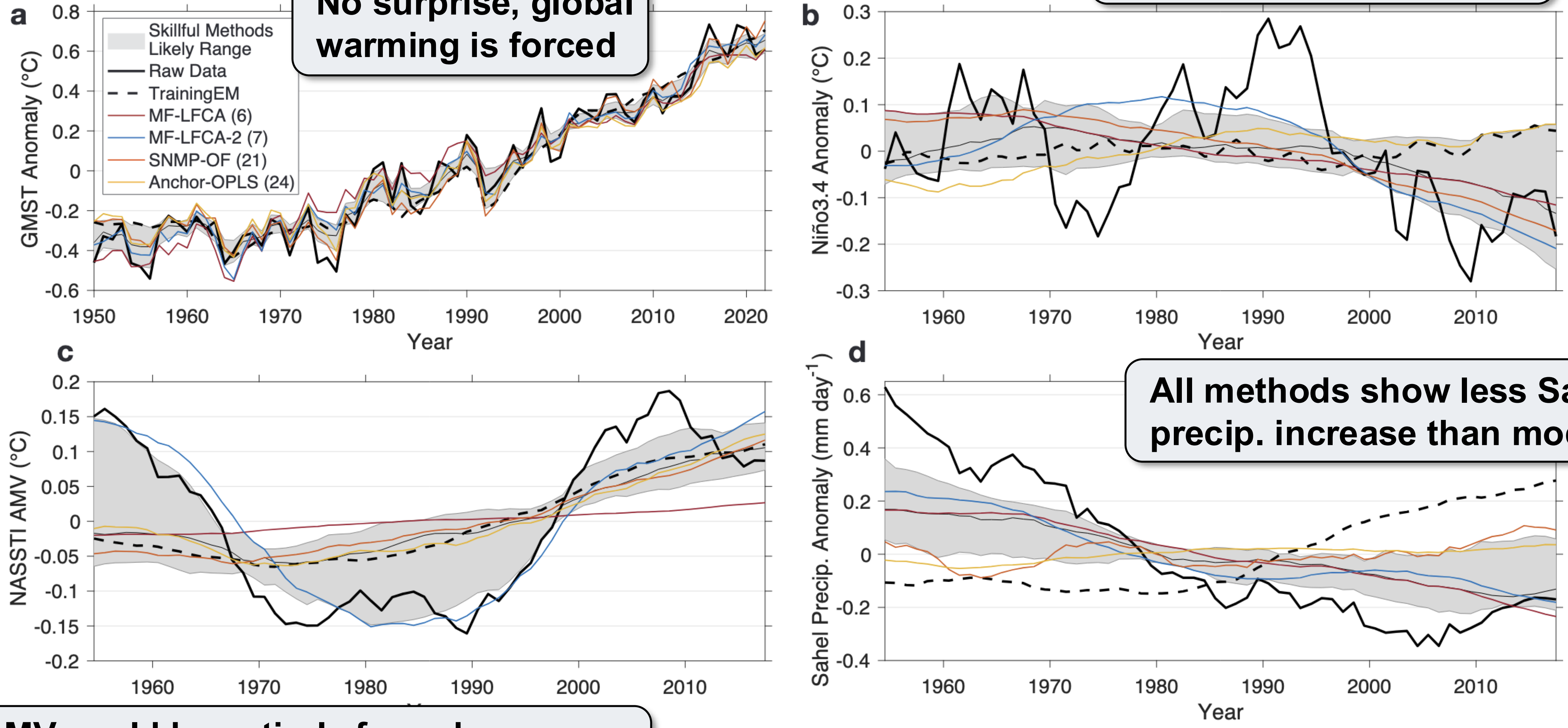
Wills et al. 2025, preprint coming soon

ForceSMIP-mean forced response estimate differs from the forced response in climate models



Wills et al. 2025,
preprint coming soon

Wide spread of ForceSMIP estimated forced responses in large-scale indices



Methods mostly more La-Niña-like than models

No surprise, global warming is forced

All methods show less Sahel precip. increase than models

AMV could be entirely forced or entirely unforced. Anything is possible!

Wills et al. 2025, preprint coming soon

Conclusions and Outlook

- Many types of statistical and machine learning methods exhibit skill at estimating the forced response (removing internal variability) in single realizations
- Methods with comparable skill in the model evaluation dataset show a wide spread of forced response estimates in observations, illustrating the epistemic uncertainty in forced response estimation
  - For example, the AMV could be almost entirely forced or almost entirely unforced
- Nevertheless, ForceSMIP shows systematic differences in estimated forced responses compared to climate models
- We will release the raw ForceSMIP dataset with the forthcoming paper *(Wills et al. 2025, preprint coming soon)*, and a skill-weighted observational forced response will follow after that *(Merrifield et al., in prep.)*
  - Potential applications to model evaluation, near-term climate prediction, observational large ensembles, climate variability analysis