ETH zürich

Forced Component Estimation Statistical Method Intercomparison Project (ForceSMIP): First Results

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], Charlotte Connolly[8], Shiheng Duan[4], Homer Durand[7], Martin Fernandez[8], Guillaume Gastineau[6], Emily Gordon[8], Moritz Günther[9], Maren Höver[1], Yan-Ning Kuo[10], Justin Lien[11], Gavin Madakumbra[3], Nathan Mankovic[7], Jamin Rader[8], Jia-Rui Shi[12], 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]MPI-Meteorology, [10]Cornell, [11]Tohoku University, [12]WHOI

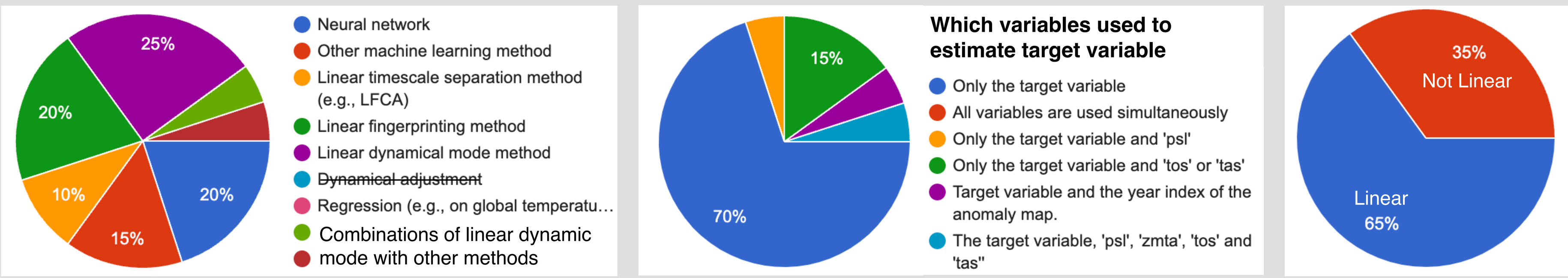
1 Motivation, Protocol, Methods

Goal: Separating the forced response from internal variability can be addressed in climate models by taking the average over a large ensemble. However, there is only one realization of the real world, making it a major challenge to isolate the forced response in observations, as is needed for accurate attribution of historical climate changes, for characterizing and understanding observed internal variability, and for ***confronting climate model trends with observations***. In ForceSMIP, contributors utilized existing and newly developed statistical and machine learning methods to estimate the forced response during the historical period within individual ensemble members and observations. We can evaluate how well the methods performed in the large ensemble testbed before applying them to observations.

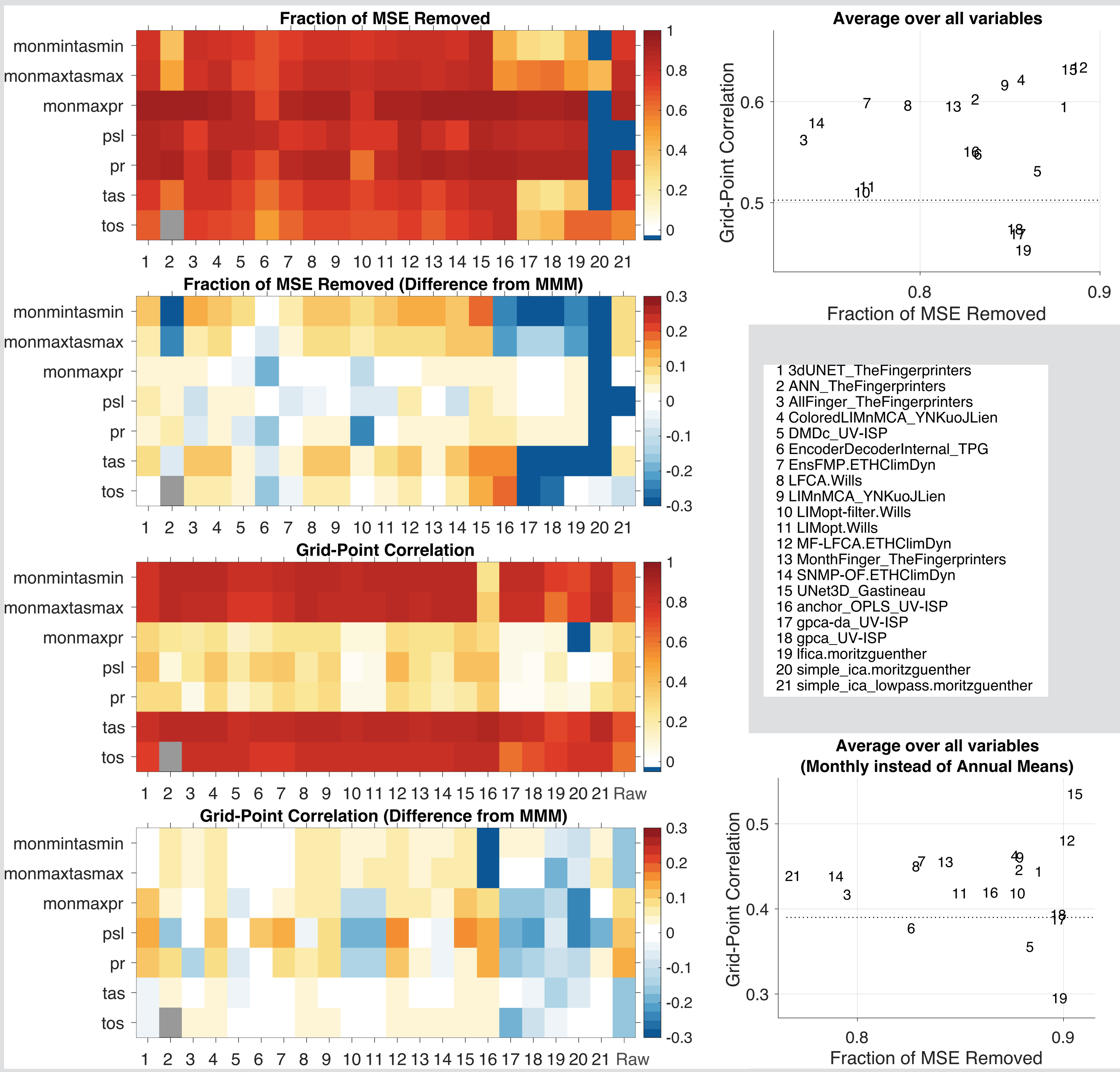
Protocol in brief: • All participants were given access to 5 LEs (CanESM2, CESM2, MIROC6, MIROC-ES2L, MPI-ESM1-2-LR, all 1880-2100) on which to train methods

• The task was then to use any method to estimate the spatiotemporally evolving (monthly resolution) forced response in 8 fields (SST, surface air temperature, precip, SLP, monthly max. and min. temperature, monthly max. daily precip, zonal-mean air temperature) over 1950-2022 (later stages will consider other fields over 1900-2022 and 1979-2022) in 10 evaluation members (5 from unseen LEs, 4 from the training LEs, and 1 from observations)

Contributed methods: The charts below summarize different choices across the 21 submitted methods



3 Skill for (Annual) Spatiotemporal Evolution



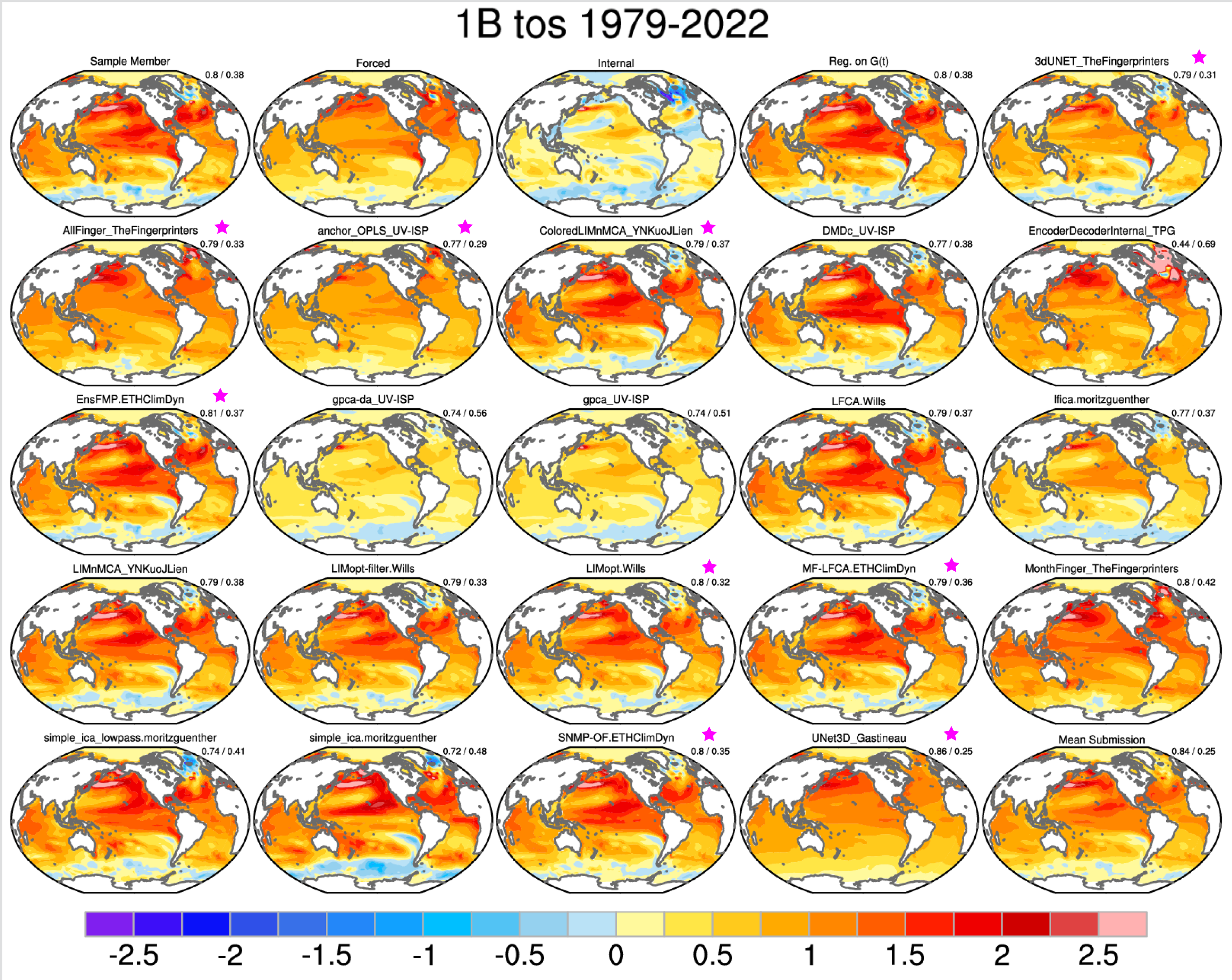
4 Discussion and Conclusions

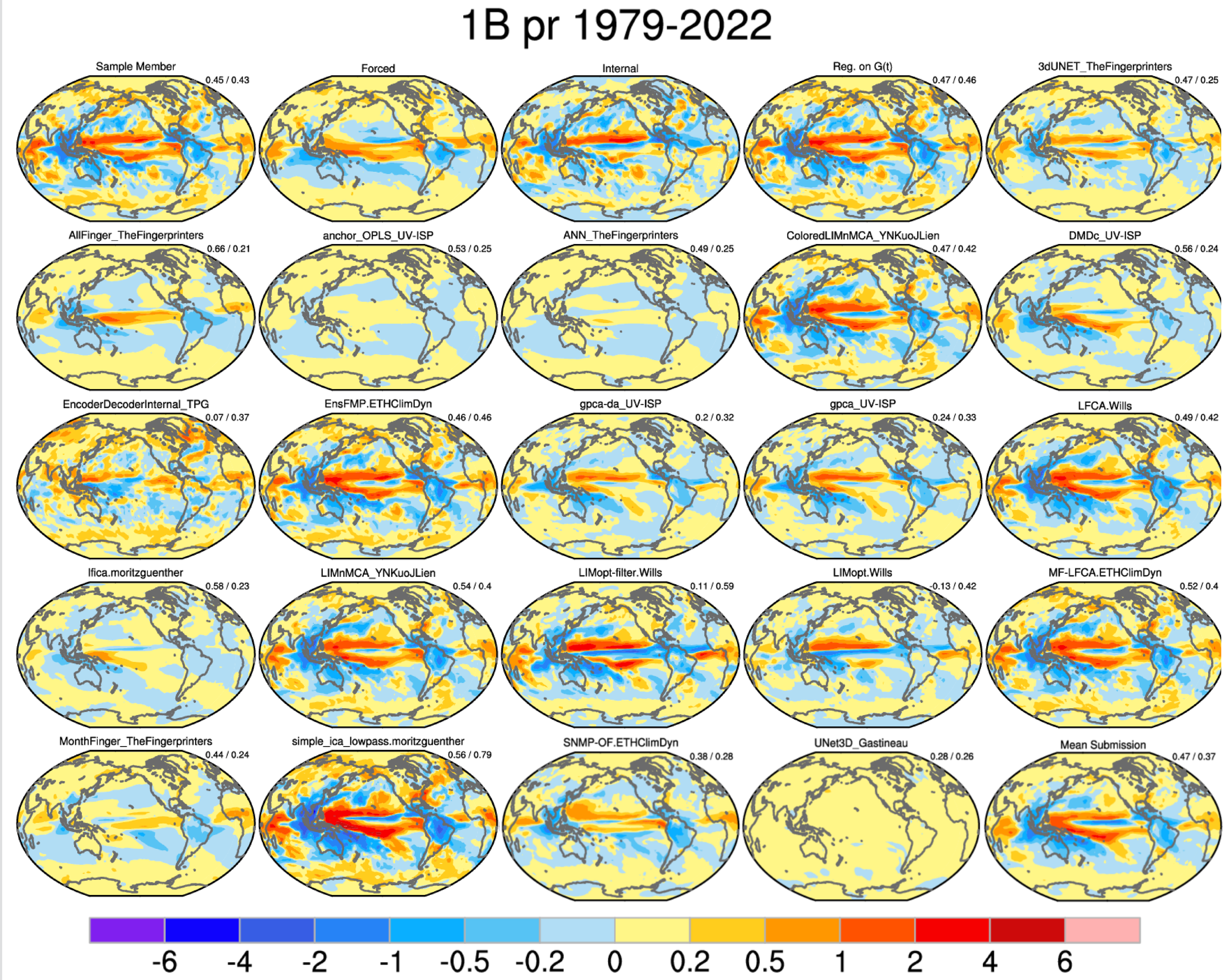
• There is no one best method for estimating the forced response. It depends on which metric you are interested in. The best option is to average over multiple methods.

• ML methods (e.g., CNNs, ANNs) perform well, but only marginally better than linear methods (e.g., variants of LFCA, LIM, and linear regression), which have far few free parameters (as few as 2 vs. as many as several million) and are less likely to overfit to the training data. However, ML methods are newer and may have more room for improvement.

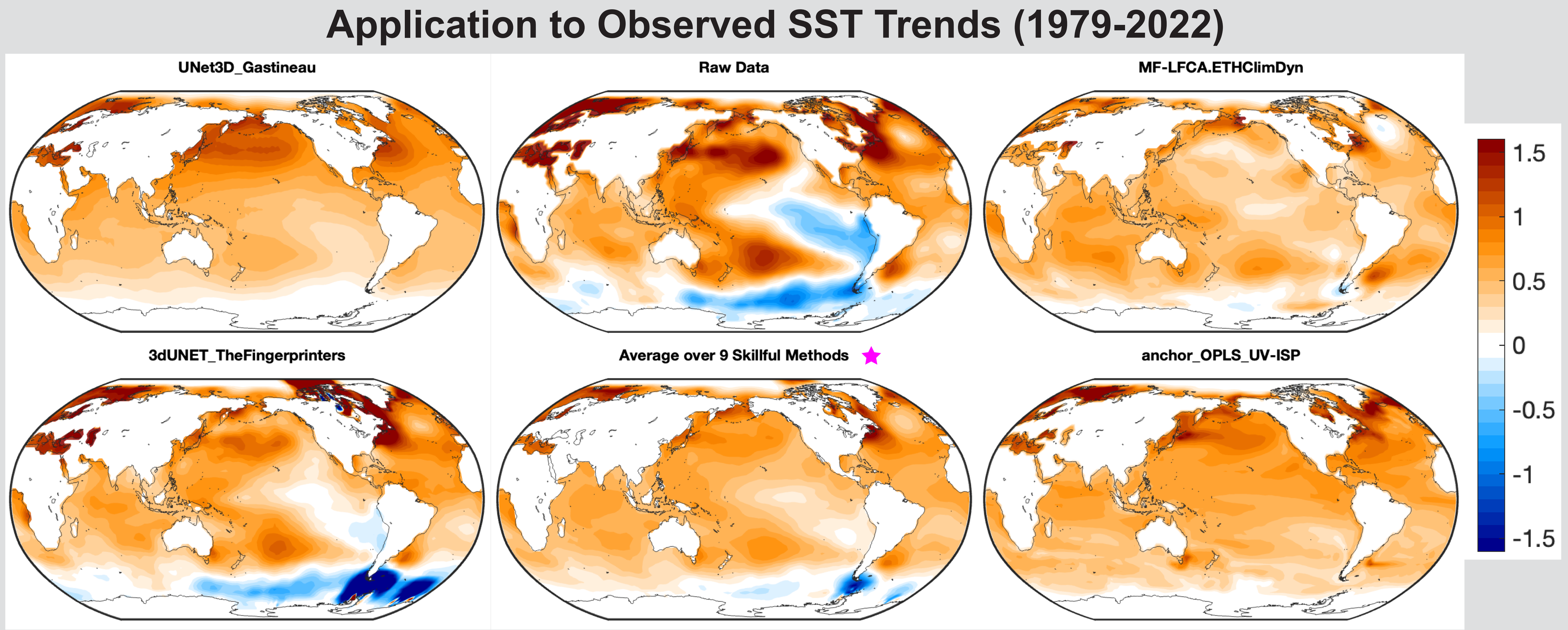
• Methods with similar skill in the model testbed (evaluation data) give very different estimates of the forced response in observations. There is substantial epistemic uncertainty in forced response estimates, and ForceSMIP helps to characterize it for the first time

2 Estimating the Forced Trend Pattern





Application to Observed SST Trends (1979-2022)



• This work is in preparation for 2 publications, one focusing on evaluating statistical methods to estimate the forced response and one presenting a best estimate of the forced response in observations (1950-2022) for all 8 variables

• **There is still chance to contribute** to tiers 2 and 3 (deadline Aug. 1, 2024). See sites.google.com/ethz.ch/forcesmip/ or write me at r.jnglinwills@usys.ethz.ch for more information

Funding/Partners: Swiss National Science Foundation, NSF, NCAR, Lawrence Livermore National Laboratory, CLIVAR, UCLA

Thanks to all participants in the ForceSMIP Hackathon (Aug 29-31, 2023 at NCAR and ETH Zürich)