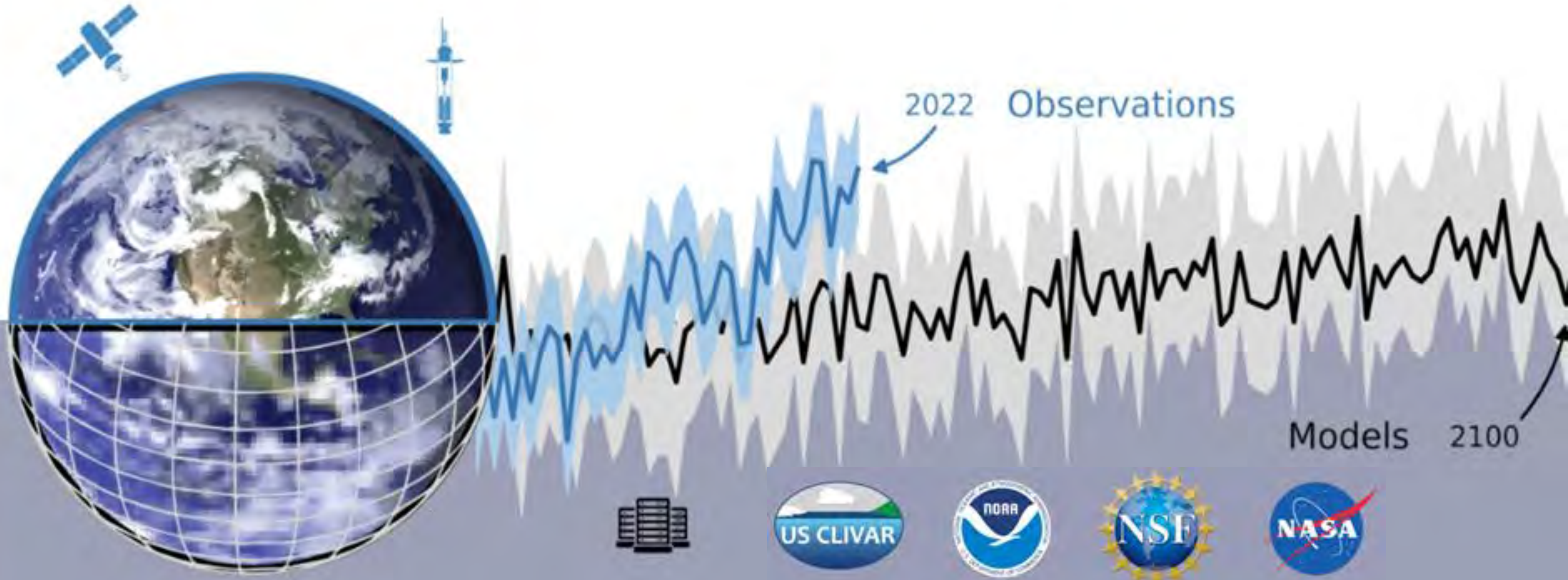
Confronting Earth System Model Trends with Observations



Tiffany Shaw[1], Isla Simpson, Paulo Ceppi, Amy Clement, Erich Fischer, Kevin Grise, Angie Pendergrass, James Screen, Robb Jnglin Wills, Tim Woollings
[1]The University of Chicago

REVIEW | CLIMATOLOGY

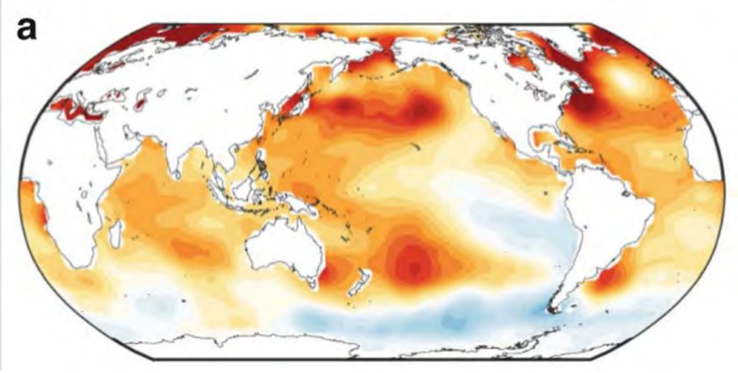
Confronting Earth System Model trends with observations

ISLA R. SIMPSON, TIFFANY A. SHAW, PAULO CEPPI, AMY C. CLEMENT, ERICH FISCHER, KEVIN M. GRISE, ANGELINE G. PENDERGRASS, JAMES A. SCREEN, ROBERT C. J. WILLS, [...], AND STEPHEN PO-CHEDLEY +3 authors Authors Info & Affiliations

SCIENCE ADVANCES • 12 Mar 2025 • Vol 11, Issue 11 • DOI: 10.1126/sciadv.adt8035

As the Earth warms more and more signals are emerging and being compared to Earth System Models

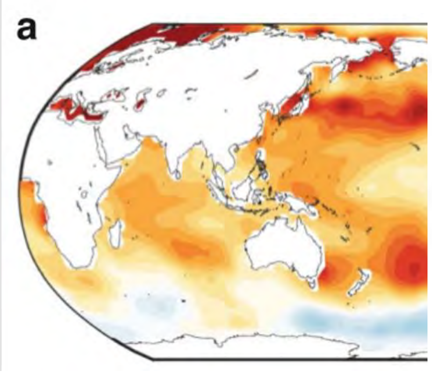
SST trend 1979-2020



Wills et al. (2022, GRL)

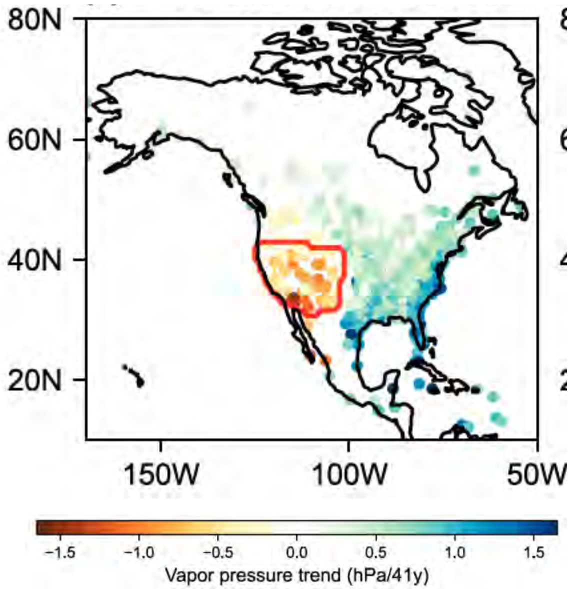
As the Earth warms more and more signals are emerging and being compared to Earth System Models

SST trend 1979-2020



Wills et al. (2022, GRL)

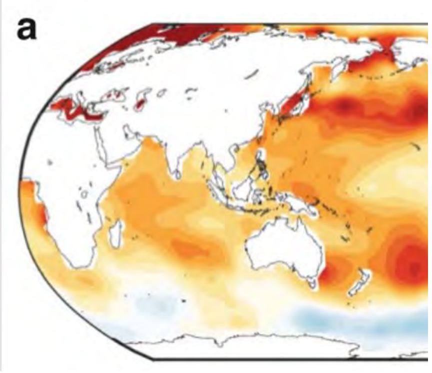
Vapor pressure trend 1979-2020



Simpson et al. (2023, PNAS)

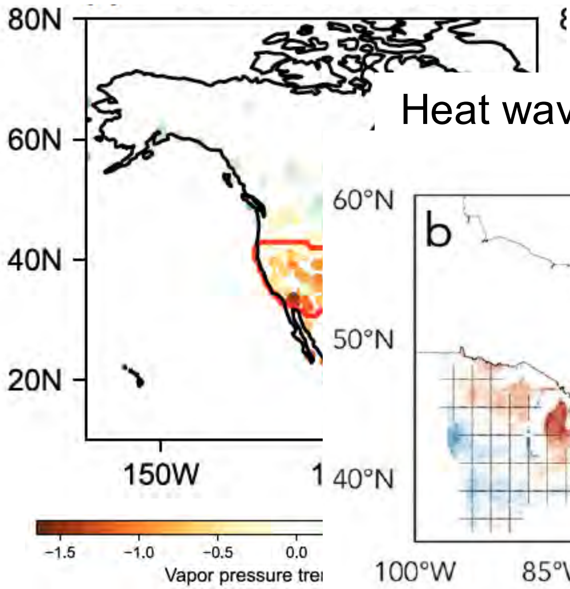
As the Earth warms more and more signals are emerging and being compared to Earth System Models

SST trend 1979-2020



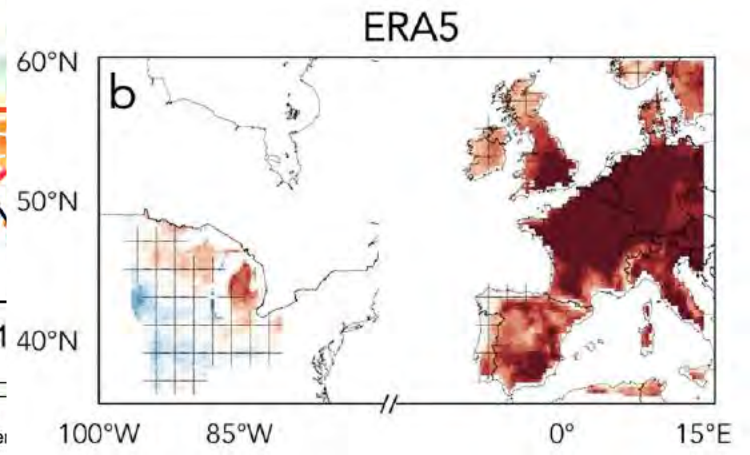
Wills et al. (2022, GRL)

Vapor pressure trend 1979-2020



Simpson et al. (2023, PNAS)

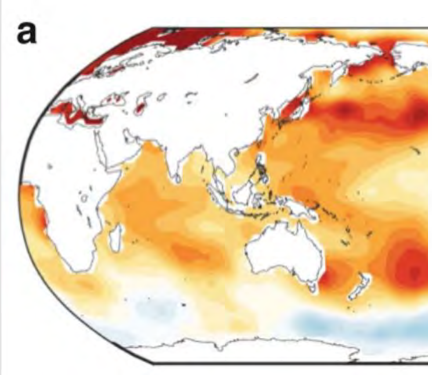
Heat wave trend 1979-2022



Singh et al. (2023, GRL)

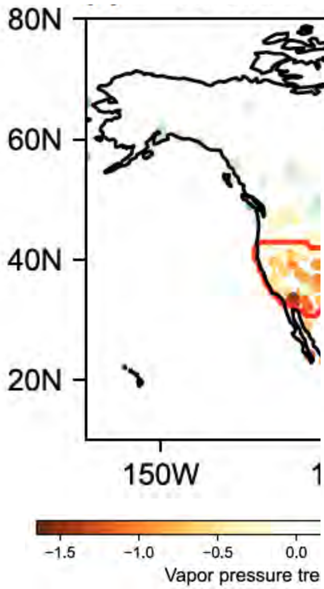
As the Earth warms more and more signals are emerging and being compared to Earth System Models

SST trend 1979-2020



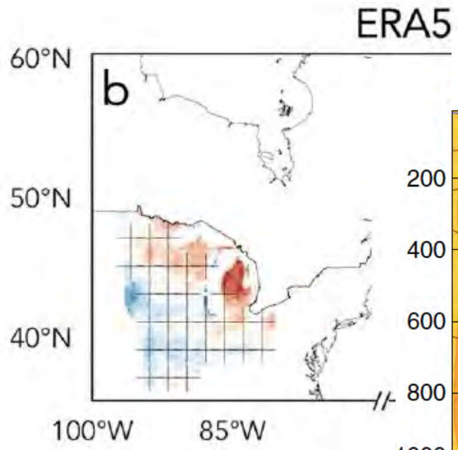
Wills et al. (2022, GRL)

Vapor pressure trend 1979-2020



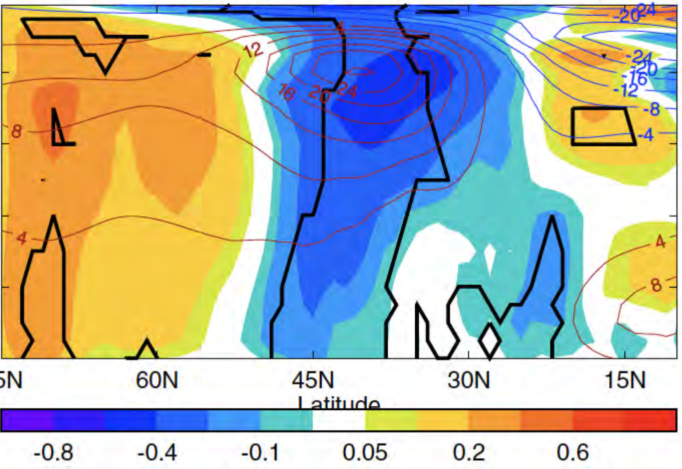
Simpson et al. (2023, PNAS)

Heat wave trend 1979-2022



Singh

East Asian jet trend 1979-2020



Dong et al. (2022, Nature Comm.)

Comparison to Earth System Model predictions reveals Successes, Partial Successes, Discrepancies, and Uncertain Situations

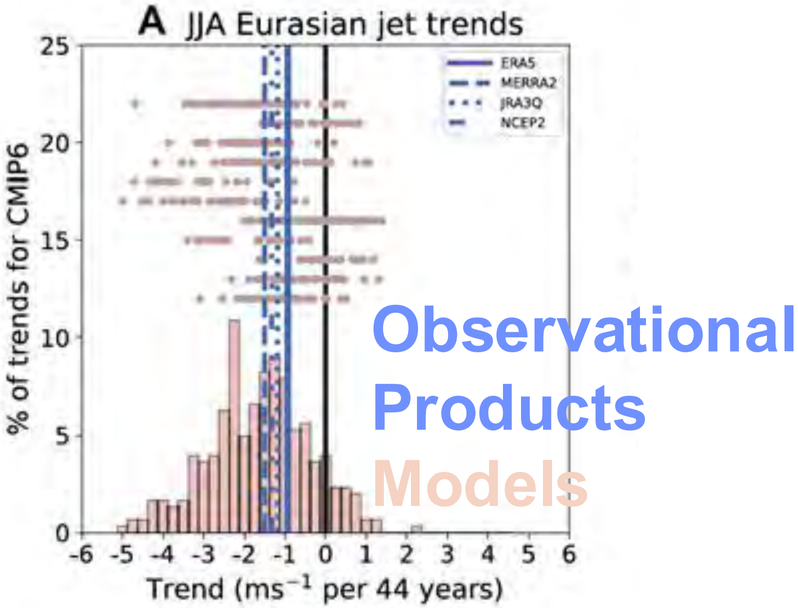
Table 1. Summary of literature comparing modeled and observed historical trends. The first column lists the field, the second column summarizes the conclusions that have been drawn for this field, the third column lists some relevant references, and the fourth column lists whether the ability of models to represent trends in this field represents a success (S), a partial success (PS), a discrepancy (D), or whether the situation is uncertain (U).

Quantity	Summary	References	
Global mean temperature	Most models accurately represent long-term historical global mean temperature rise over the instrumental record, although they tend to do so with greater warming trends in the tropical Pacific than observed (Fig. 2).	(19) and figure 1.9 of (21)	S
Global column water vapor	Models accurately represent the historical rise in globally averaged column integrated water vapor.	(31, 210)	S
NH summer jet stream and storm track trends	Models accurately represent the observed weakening of the NH summer jet stream (Fig. 1A). The CMIP6 models also capture the weakening of the NH summer storm tracks but CMIP5 did not.	(66, 72, 214)	S
Marine heat waves	Models capture the increasing probability of marine heatwaves over the satellite era.	(44, 45)	S
Amplitude of the SST seasonal cycle in the Northern Hemisphere	Models capture the observed increase in the amplitude of the seasonal cycle of SSTs in the Northern Hemisphere.	(19)	S
Increasing intensity of extreme precipitation events (global)	Models and observations broadly agree on increasing trends in extreme precipitation intensity when aggregated globally. Uncertainties are larger for regional patterns, and whether the globally aggregated trends in the magnitude of the increase in intensity is a success can depend on the metric used.	(61–64)	S
Pause of SH circulation trends as ozone recovers	Models capture a poleward shift of the SH mid-latitude jet in response to ozone depletion and also represent a pause in that shift as ozone starts to recover—both of which are also seen in observations.	(68)	S
Hadley cell extent	The expansion of the tropics as measured by the Hadley cell edge lies within the modeled distribution of trends, as long as it is calculated using well-constrained surface metrics in modern reanalyses. Previously documented discrepancies were resolved by using newer-generation reanalyses, considering surface metrics of the Hadley cell edge, and accounting for internal variability.	(58)	S
Wintertime cold extremes in the NH	Earlier studies argued that an increase in observed cold extremes was different from model behavior. But updated analysis accounting for temporal variations in observational coverage now indicates that models and observations agree on a decline of NH cold extremes during winter (Fig. 3).	(47)	S
Arctic warming	The observed warming of the Arctic during the satellite era lies within the modeled distribution of trends.	(20)	S
Tropical overturning circulation	Both models and observations exhibit a weakening of the global overturning circulation over the historical record, but the magnitude of this weakening in the tropics is overestimated in models. There are also local discrepancies such as in the tropical Pacific.	(52–54)	PS
Contrast between tropical dry and wet regions	Precipitation contrasts between dry and wet regions in the tropics have increased. Models represent this, but the magnitude of the observed change is larger than most model simulations.	(55)	PS
Increased precipitation variability	Models show an increase in precipitation variability, which has now been observed, although there may be discrepancies in magnitude in some regions.	(67)	PS
Tropical tropospheric temperature	Models and observations agree on historical warming of the tropical troposphere, but the warming in most model simulations is too large. Recent studies suggest a likely role for the combined influence of internal variability, discrepancies in the tropical warming pattern, issues with the forcings provided to models, too-large climate sensitivity in some models, and observational biases.	(31, 36, 37, 215)	PS
TOA radiative imbalance	Models and observations both exhibit an increasing trend in TOA radiative imbalance, but the magnitude of the trend since 2001 is underestimated in models compared to observations.	(25, 26)	PS
Arctic amplification	Models robustly predict amplified warming of the Arctic compared to elsewhere, which has been observed. Models also seem to capture the magnitude of the warming of the Arctic (see success above), but there are concerns that models may be underestimating the magnitude of this amplified warming relative to warming in the rest of the planet, particularly in recent decades when internal variability is thought to have enhanced Arctic amplification trends in observations.	(23, 24, 176, 216, 217)	PS
Arctic sea ice	Models capture the observed declining trend, but internal variability leads to a large uncertainty and is thought to have contributed to the magnitude of the observed decline. There are indications that models may be capturing sea ice trends for the wrong reasons.	(22, 218–220)	PS

Quantity	Summary	References	
Tropical SST pattern	Most model ensemble members fail to capture the observed strengthening of the tropical Pacific SST gradient and instead predict a weakening (Fig. 1B). This is true both in a narrow band at the equator and for the broader tropical pattern.	(53, 73–75, 126)	D
Wintertime North Atlantic jet	Models fail to capture the observed strengthening of the North Atlantic jet and associated impacts on European precipitation since 1951.	(88)	D
JJA Greenland blocking	The recent increase in Greenland blocking events seen in observations is not captured in model simulations.	(89, 90)	D
Exacerbated summer warming in western-central Europe	Western-central Europe has seen exacerbated warming and drying compared to the global mean and a substantial rise in heat extremes that is not well captured by models.	(91–94)	D
Arid region near-surface specific humidity	Models suggest that near-surface specific humidity in arid regions should have risen over the historical record. A rise has not been observed.	(85–87)	D
Southern Ocean SSTs and sea ice extent	It is rare for model ensemble members to reproduce the observed slight decline in Southern Ocean SSTs and increase in Southern Ocean sea ice extent since 1979, although rapid declines in Southern Ocean sea ice have been observed in recent years.	(53, 79)	D
Winter Eurasian cooling/warming hole	The observed winter cooling or suppressed warming over central Eurasia is within the range of modeled internal variability, but it has also been argued that the forced response in models could be too weak.	(98–102)	U
Hadley circulation strength	Reanalyses exhibit a strengthening, while climate models exhibit a weakening, but there are indications that the reanalyses are in error.	(59, 60, 221)	U
SH storm track	Chemke *et al.* (103) showed that the SH storm track strengthening in certain reanalyses is greater than in models, but using a wider array of reanalyses and like-for-like comparison, Kang *et al.* (84) demonstrated a large observational uncertainty (Fig. 1C) and that the discrepancy may be smaller than originally thought and likely influenced by discrepant tropical Pacific SST trends.	(84, 103, 222)	U
Zonal mean jet stream	There are indications that models are adequately capturing the poleward shift of the jet streams but that they may not have the correct relationship between upper tropospheric warming and this poleward shift, so this may be for the wrong reasons.	(104)	U

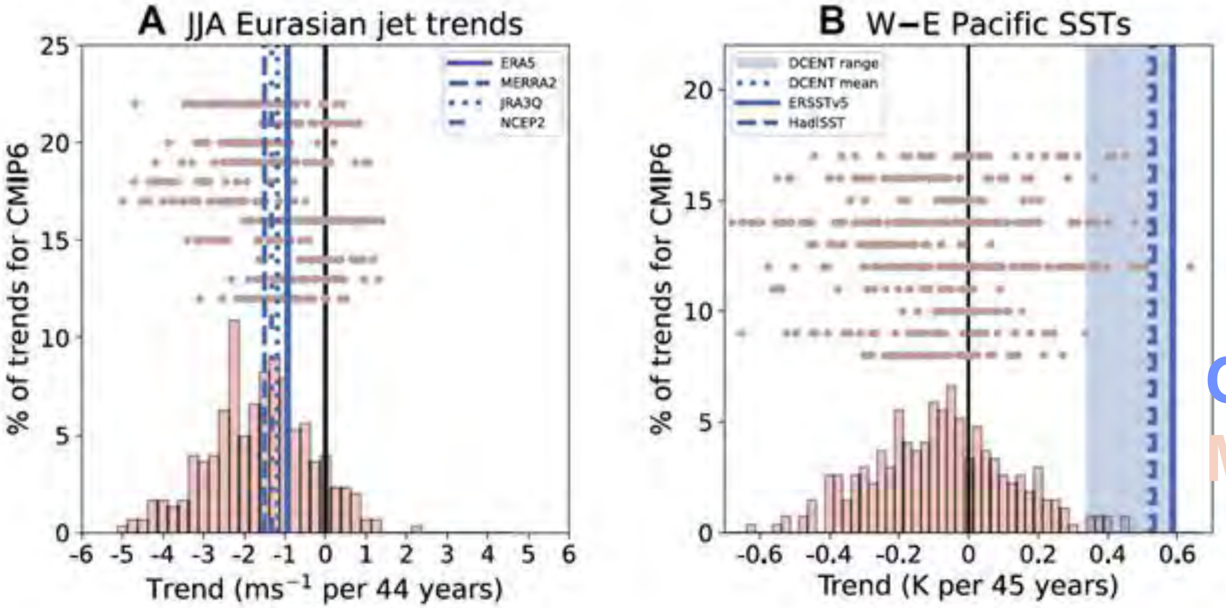
Simpson, Shaw et al. (2025, Sci. Adv.)

Success



Simpson, Shaw et al. (2025, Sci. Adv.)

Success **Discrepancy**

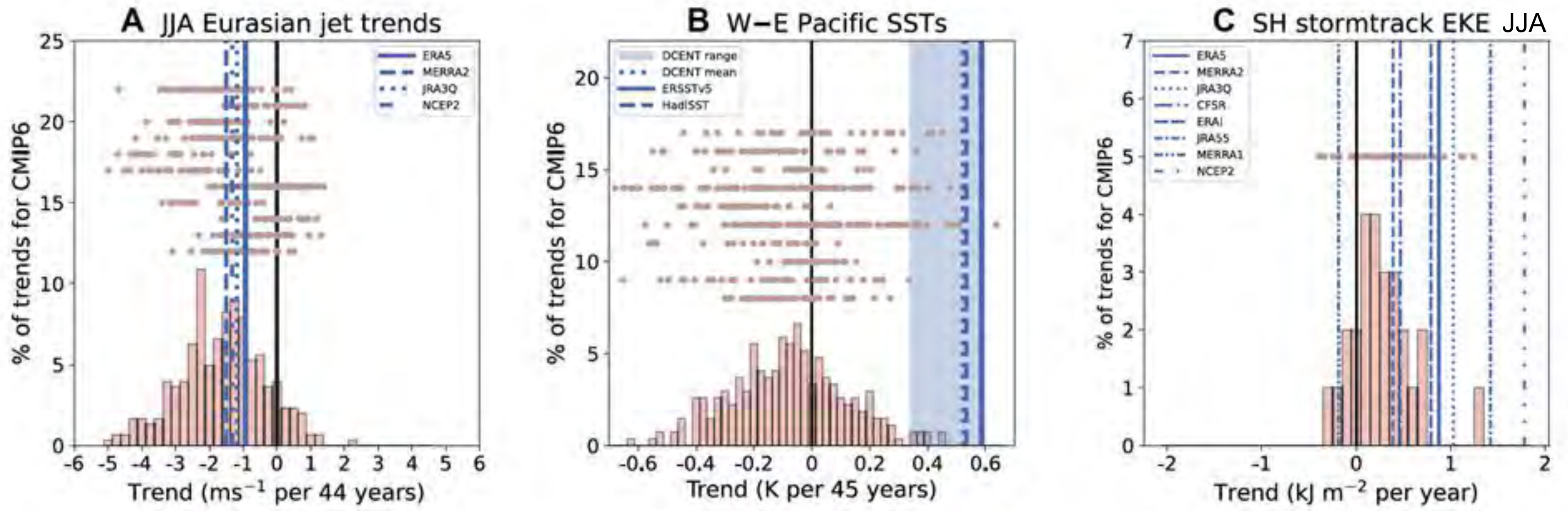


Observations

Models

Simpson, Shaw et al. (2025, Sci. Adv.)

Success | **Discrepancy** | **Uncertain Situation**



Simpson, Shaw et al. (2025, Sci. Adv.)

I. Ensure a “like-for-like” comparison

Best practices for comparing model trends with observations

Simpson, Shaw et al. (2025, Sci. Adv.), Schmidt (2013)

I. Ensure a "like-for-like" comparison

Best practices for comparing model trends with observations

II. Use large number of observations

Simpson, Shaw et al. (2025, Sci. Adv.), Schmidt (2013)

Best practices for comparing model trends with observations

I. Ensure a “like-for-like” comparison

II. Use large number of observations

III. Compare individual model runs to ensemble mean

Simpson, Shaw et al. (2025, Sci. Adv.), Schmidt (2013)

Best practices for comparing model trends with observations

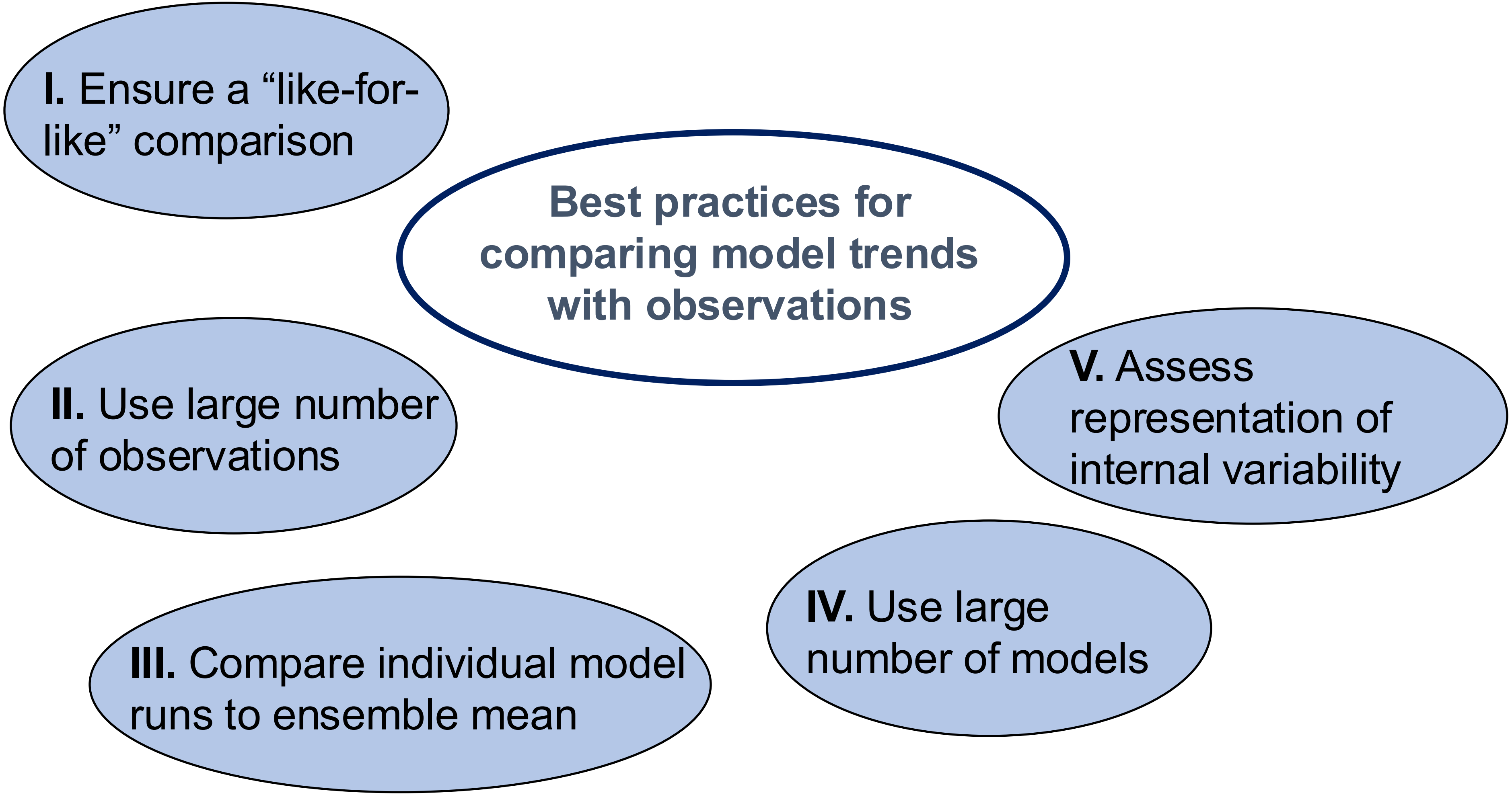
I. Ensure a "like-for-like" comparison

II. Use large number of observations

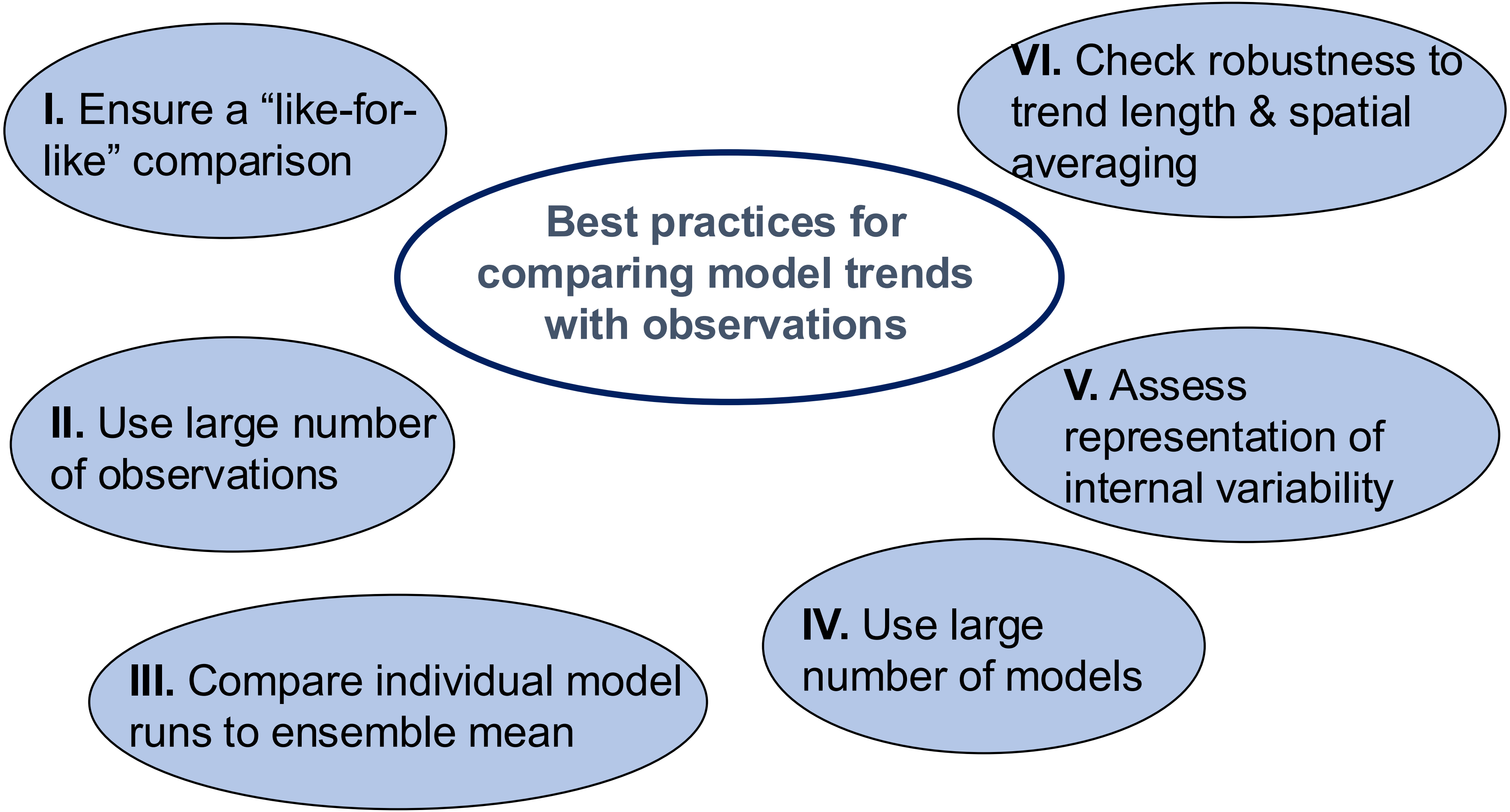
III. Compare individual model runs to ensemble mean

IV. Use large number of models

Simpson, Shaw et al. (2025, Sci. Adv.), Schmidt (2013)

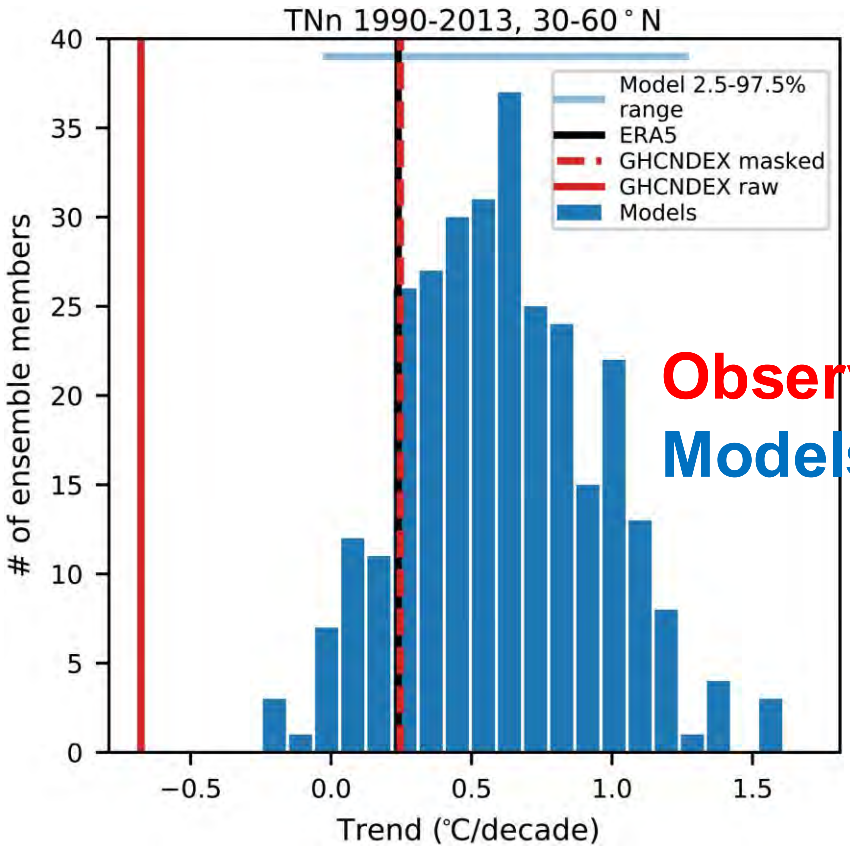


Simpson, Shaw et al. (2025, Sci. Adv.), Schmidt (2013)



Simpson, Shaw et al. (2025, Sci. Adv.), Schmidt (2013)

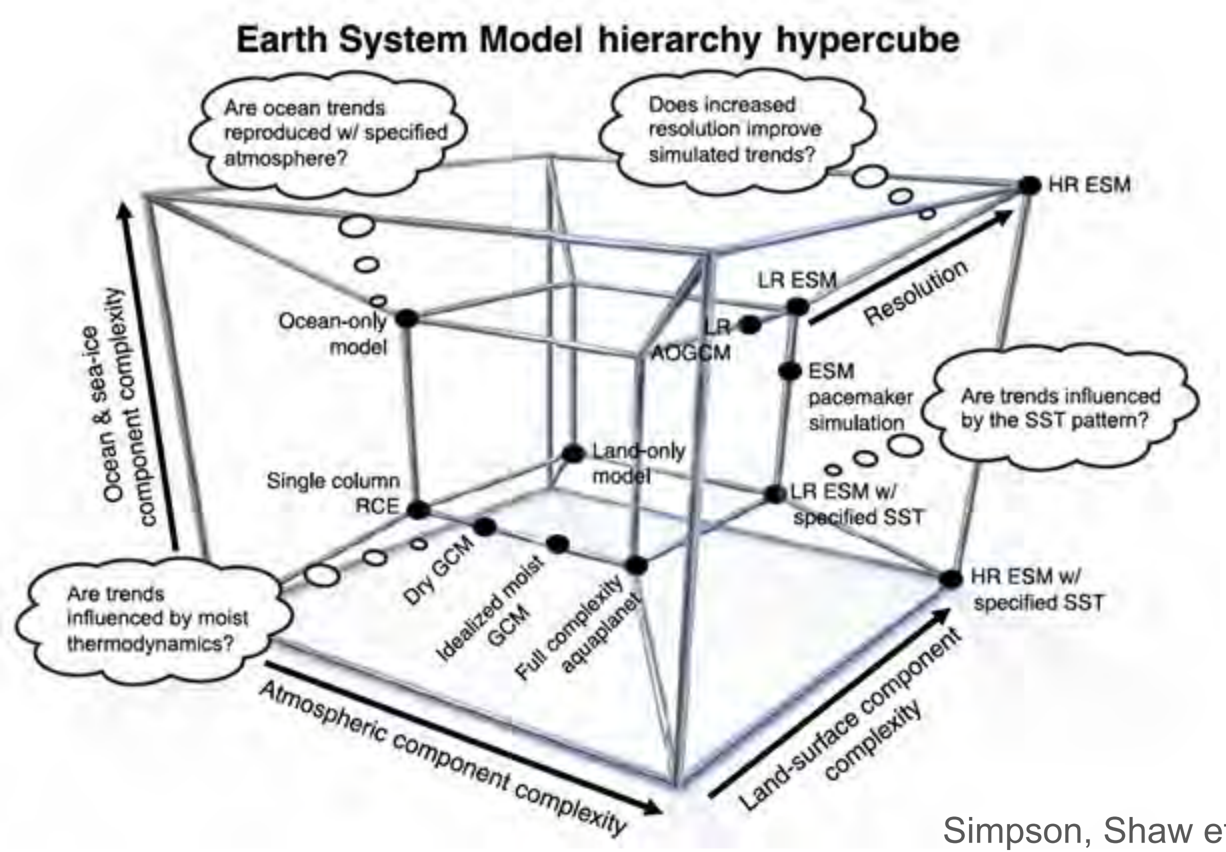
Cold daily minimum temperature trend is sensitive to observational coverage over time



Observations
Models

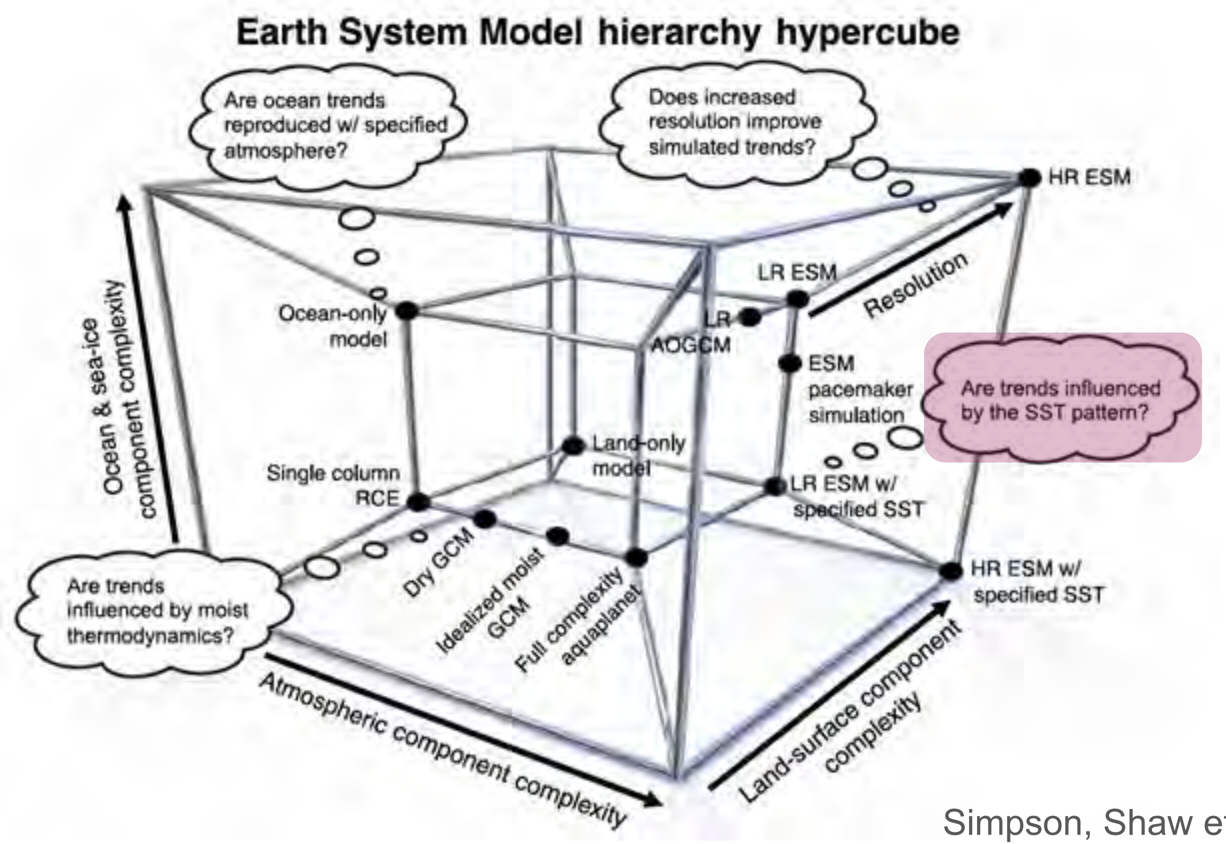
Blackport et al. (2024, Sci. Adv.)

Model Hierarchy can help us move beyond quantification into understanding



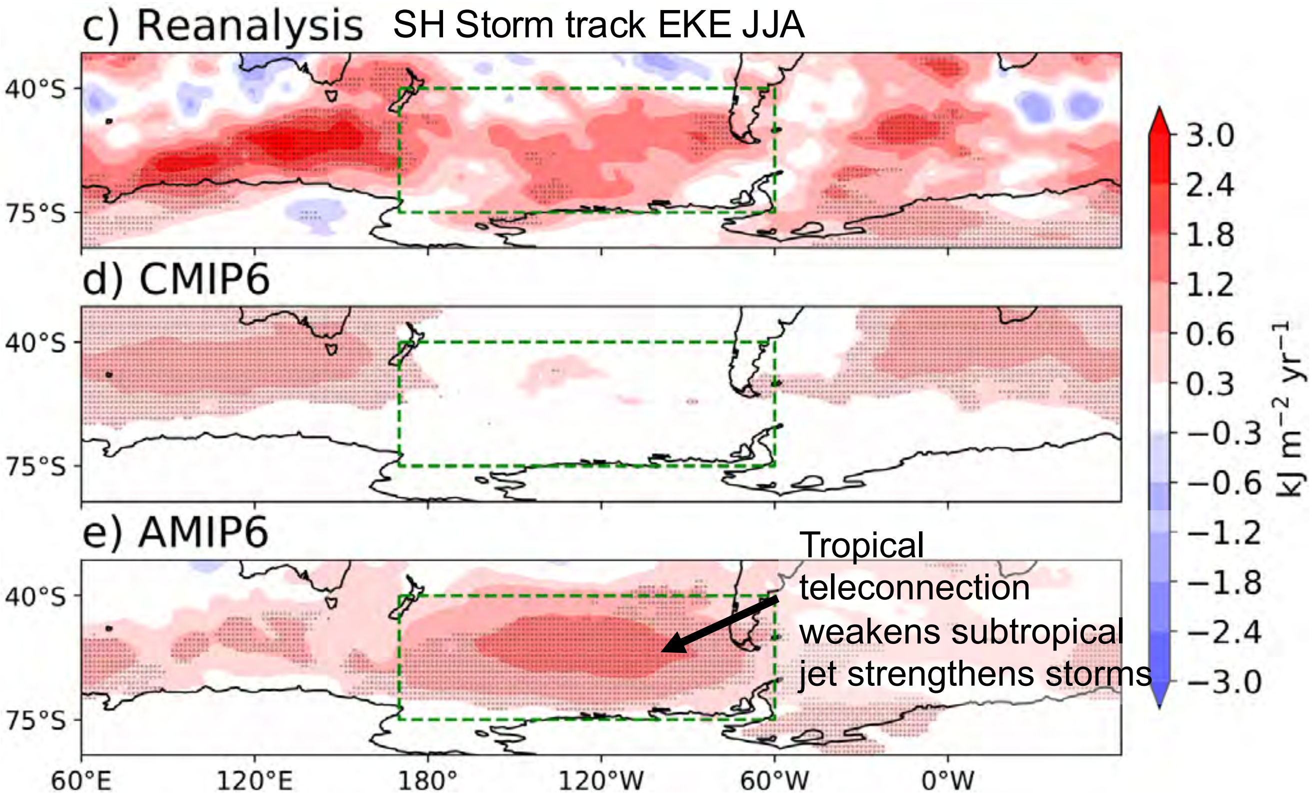
Simpson, Shaw et al. (2025, Sci. Adv.)

Model Hierarchy can help us move beyond quantification into understanding



Simpson, Shaw et al. (2025, Sci. Adv.)

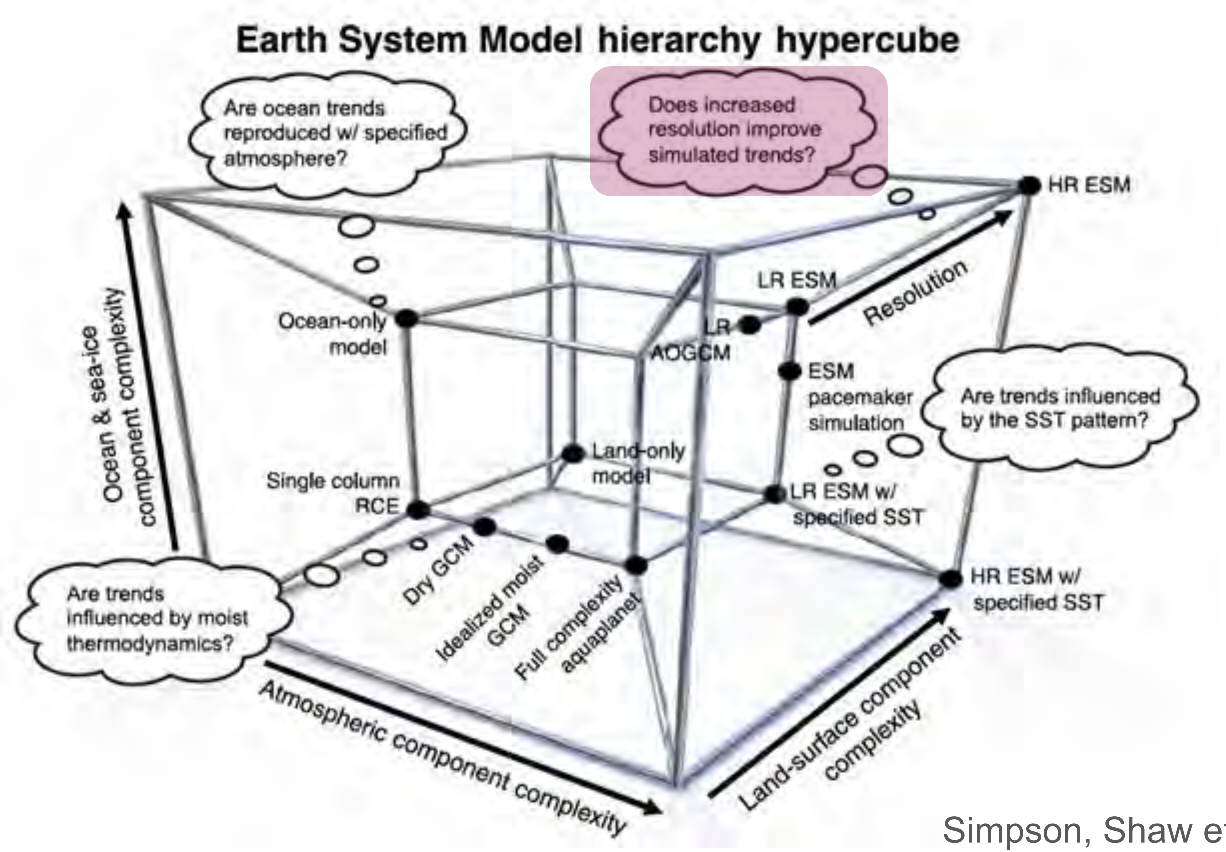
Southern Hemisphere storm tracks trends influenced by SST pattern



Results consistent with Pacemaker and Hindcast simulations

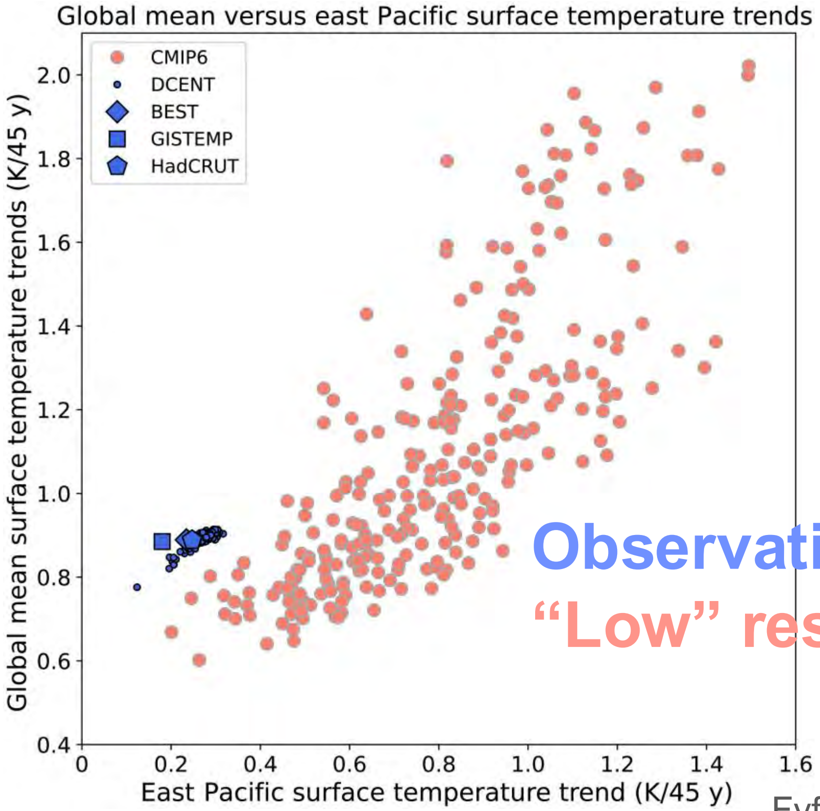
Kang et al. (2024, npj Climate & Atmos. Sci.)

Model Hierarchy can help us move beyond quantification into understanding



Simpson, Shaw et al. (2025, Sci. Adv.)

High resolution simulations should be put in the context of model ensemble distribution

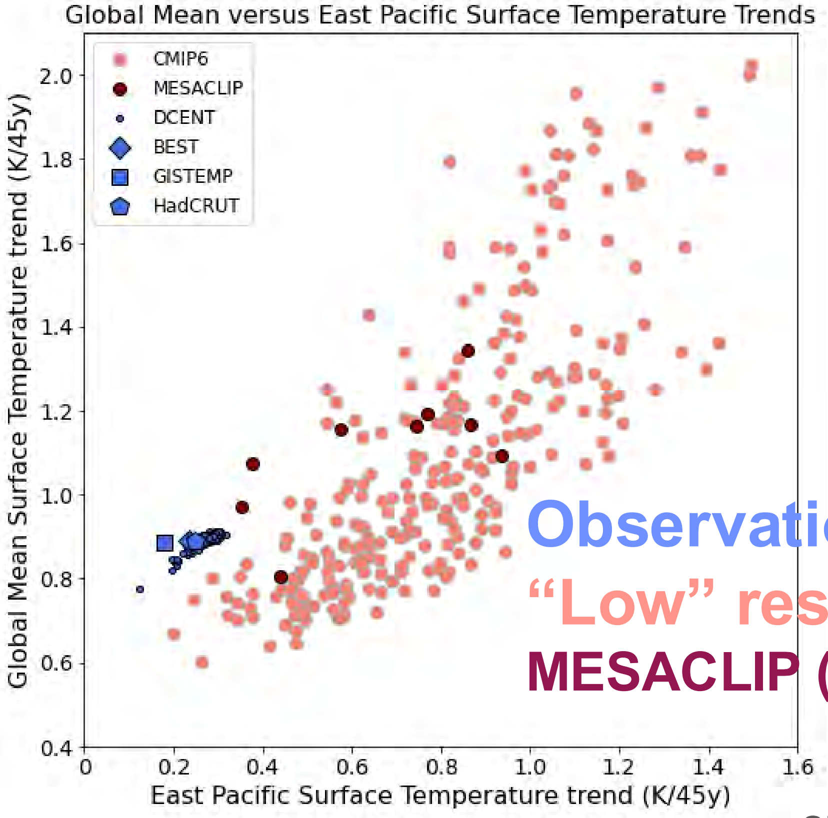


Observations

“Low” res. CMIP6 Models

Fyfe & Gillett (2011, Nat. Clim. Change)
Simpson, Shaw et al. (2025, Sci. Adv.)

High resolution simulations should be put in the context of model ensemble distribution



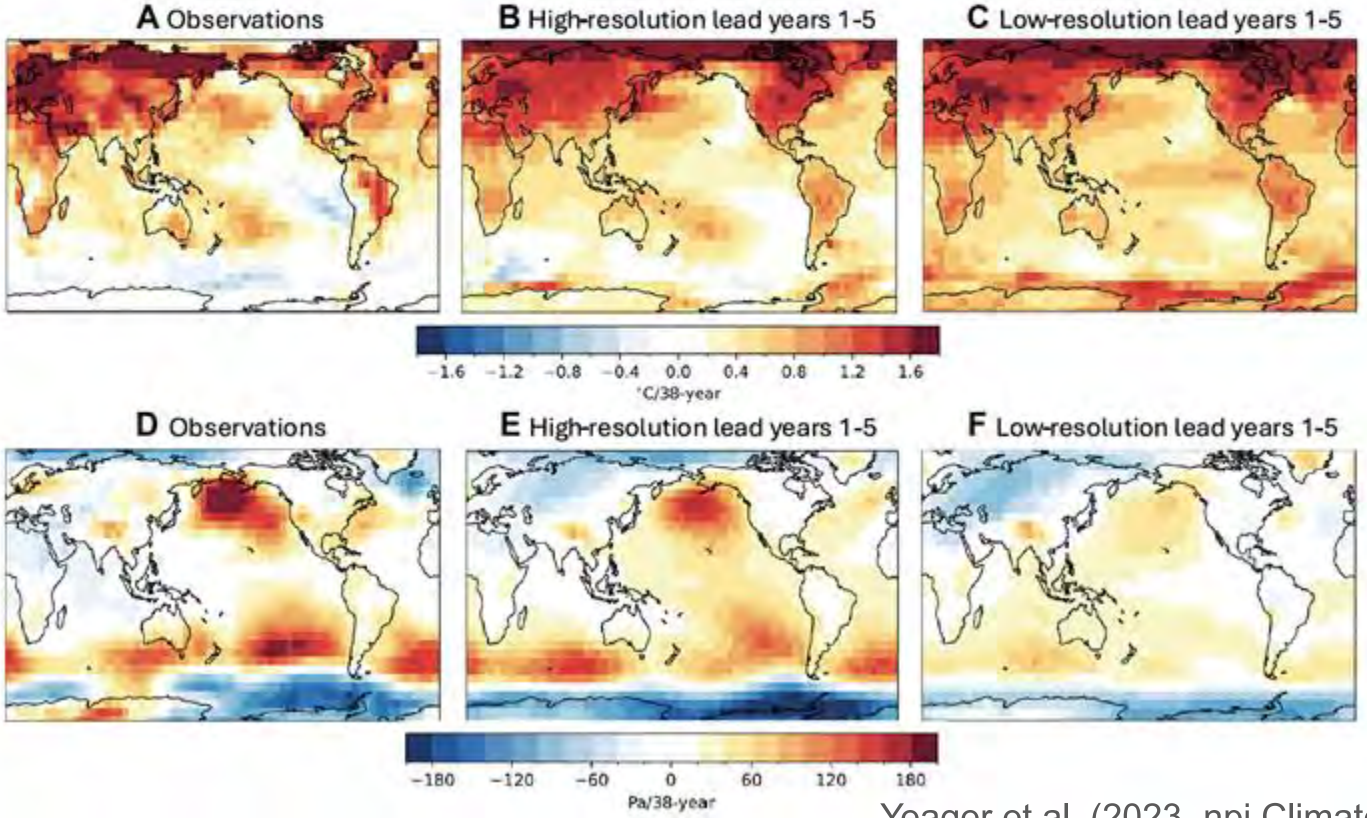
Observations

"Low" res. CMIP6 Models

MESACLIP (0.25◦ atm, 0.1◦ ocn)

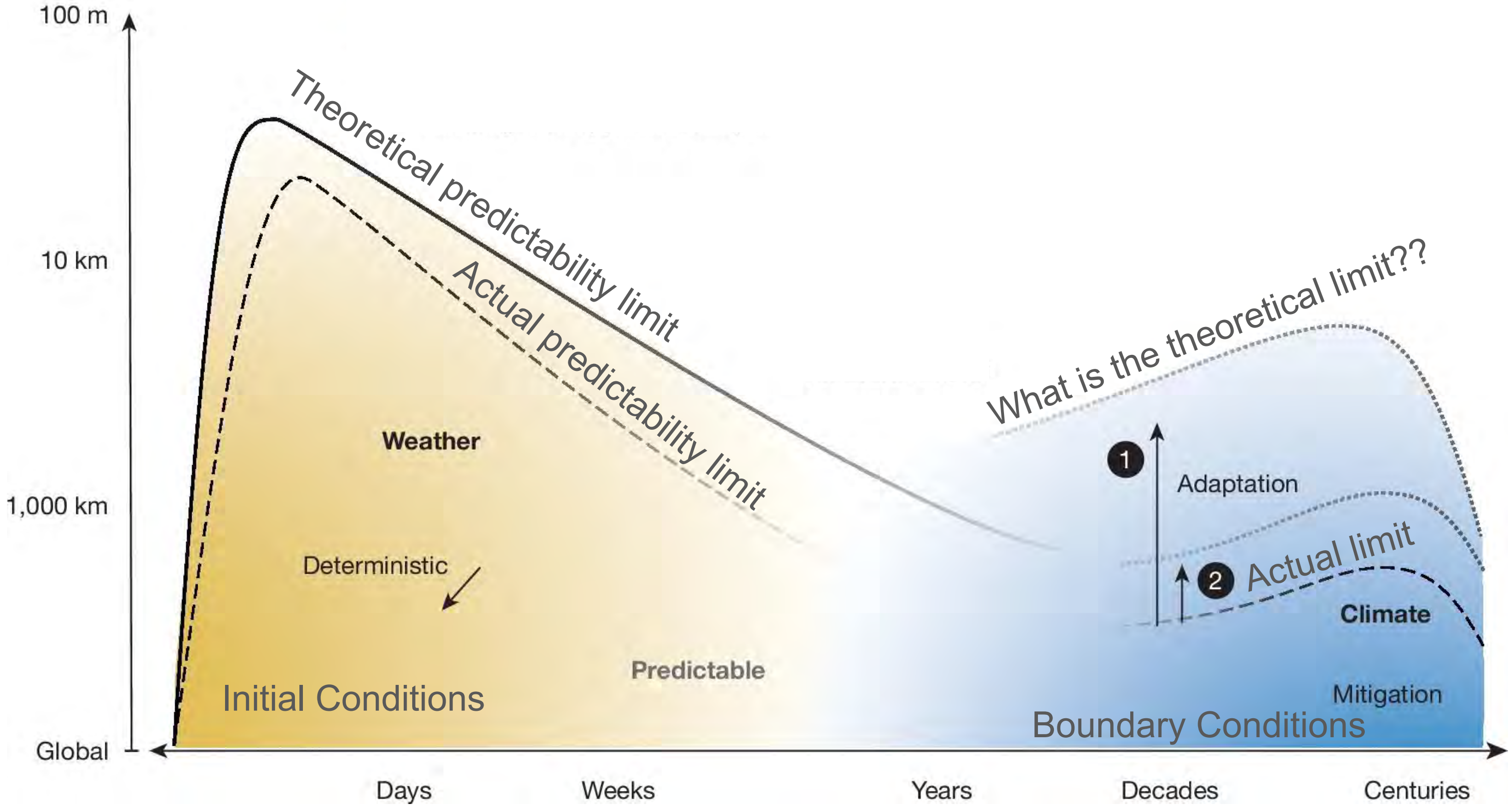
Simpson, Shaw et al. (2025, Sci. Adv.)

Initialized predictions help to reveal impact of model resolution and initialization on trend discrepancies



Yeager et al. (2023, npj Climate & Atmos. Sci.)

Confronting Earth System model trends with observations brings with it a better understanding of the climate system



Shaw & Stevens (2025, Nature)

AGU25 WHERE SCIENCE CONNECTS US

GC036 - Confronting Earth System Model Trends with Observations

Submit an Abstract to this Session

Deadline July 30th!

CMIP26 Session Gallery

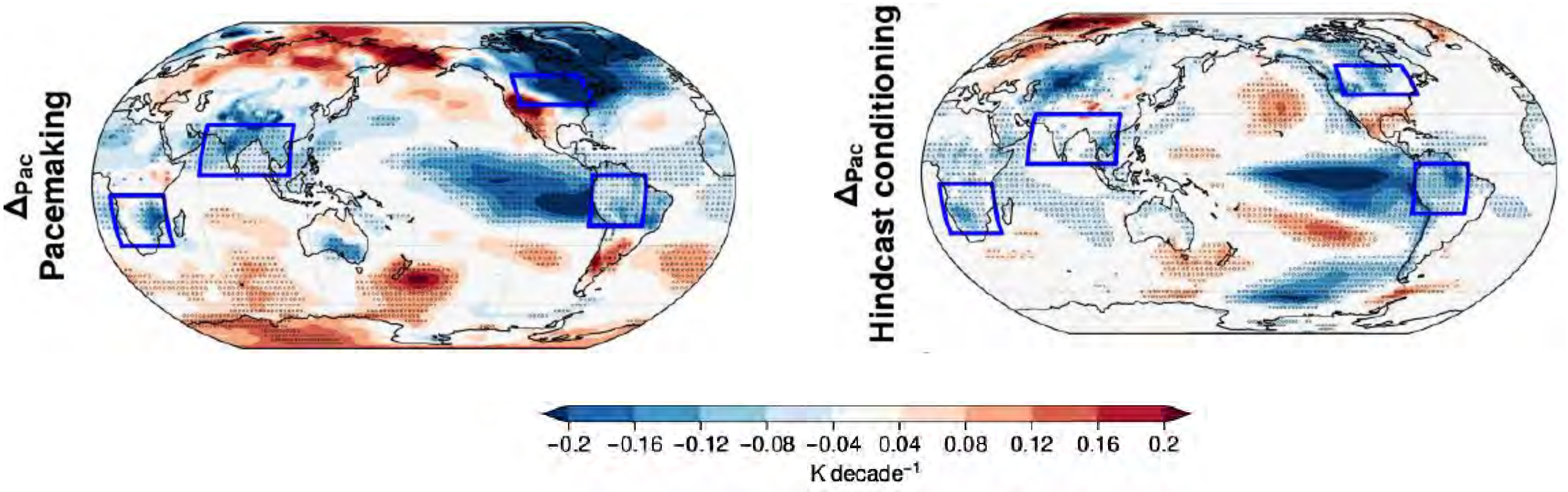
Click on a card to see the full session title and description

Filter Sort ···

1	2	3
Session title Earth System Forcings: CMIP7 and beyond	Session title Confronting Earth System Model Trends w...	Session title Understanding reactive gases, aerosols, a...
Assigned theme T1: Progress in understanding historical...	Assigned theme T1: Progress in understanding historical...	Assigned theme T1: Progress in understanding historical...
Session description Earth system variability and change arise from shifts in atmospheric composition that affect the top-of-atmosphere energy balance, from land use changes that alter ...	Session description Anthropogenically forced climate change signals are emerging from the noise of internal variability in observations, and the impacts on society are growing. For ...	Session description The second phase of the Aerosol and Chemistry Model Intercomparison Project (AerChemMIP2), part of the Coupled Model Intercomparison Project phase 7 (CMIP7),...

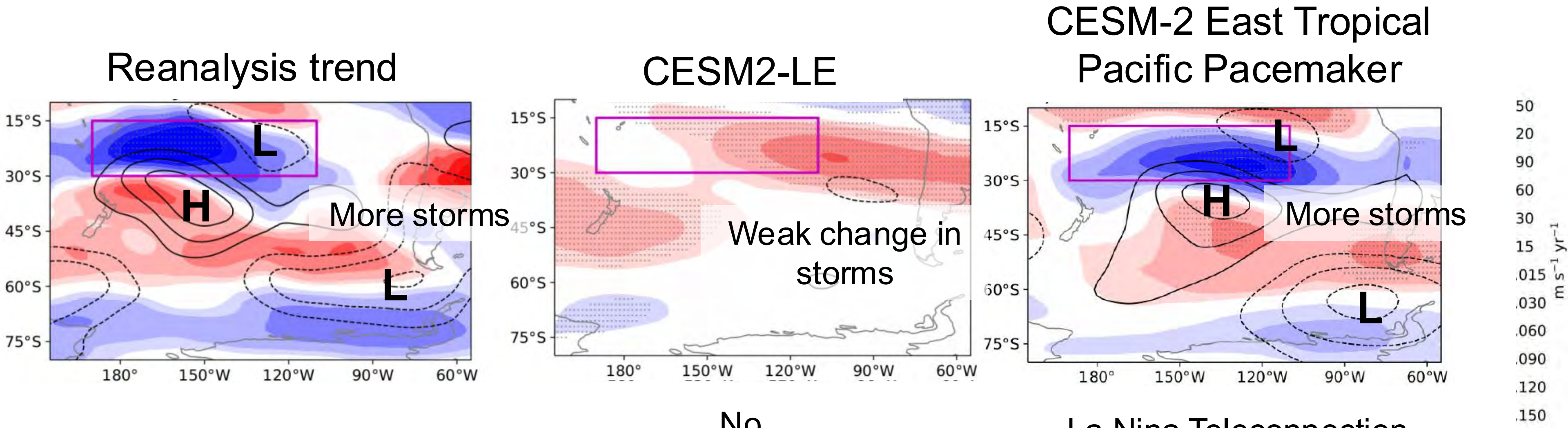
ility in observations,
ɛen predicting how
confront the signals,
t take stock of the
ghlighting successes
yond quantification into
tting-edge methods
cies and separating

Regional temperature trends influenced by SST pattern



Kang et al. (2025, npj Climate & Atmos. Sci.)

Discrepancy in tropical SST trends leads to teleconnection discrepancy



La Nina teleconnection

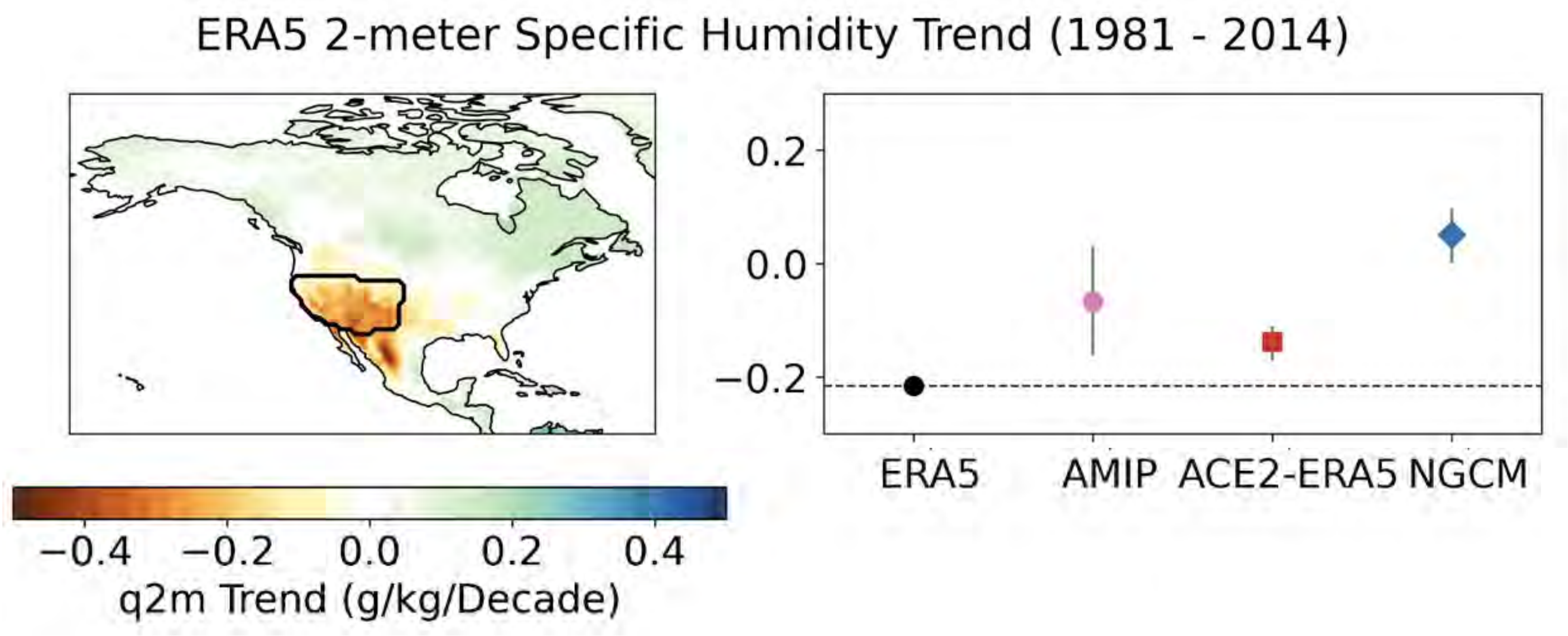
No teleconnection

La Nina Teleconnection emerges with corrected SSTs

Similar story for Northern Hemisphere

Kang et al. (2024, npj Climate & Atmos Sci)

AI emulators also need to be benchmarked in the context of model ensemble distribution

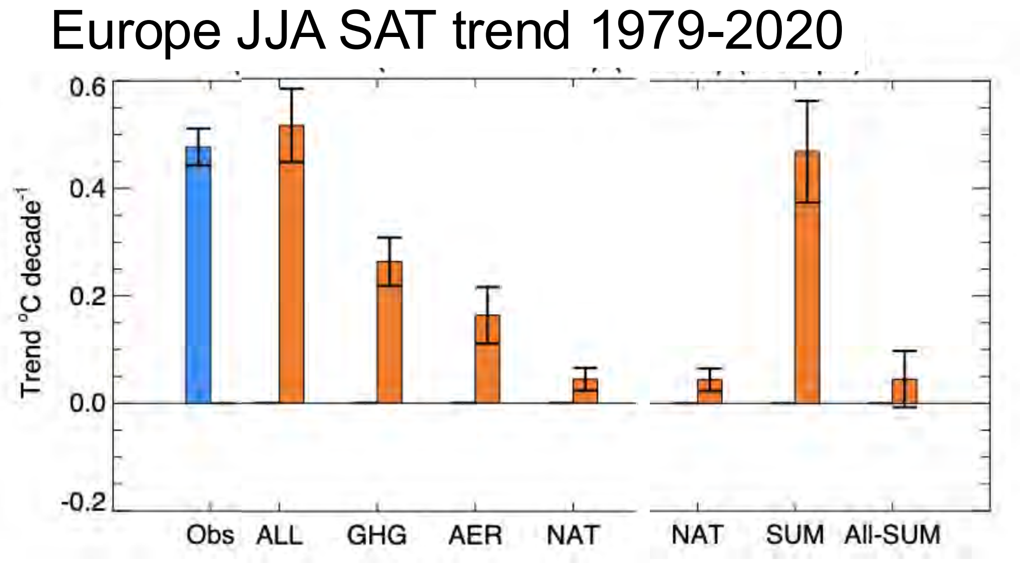


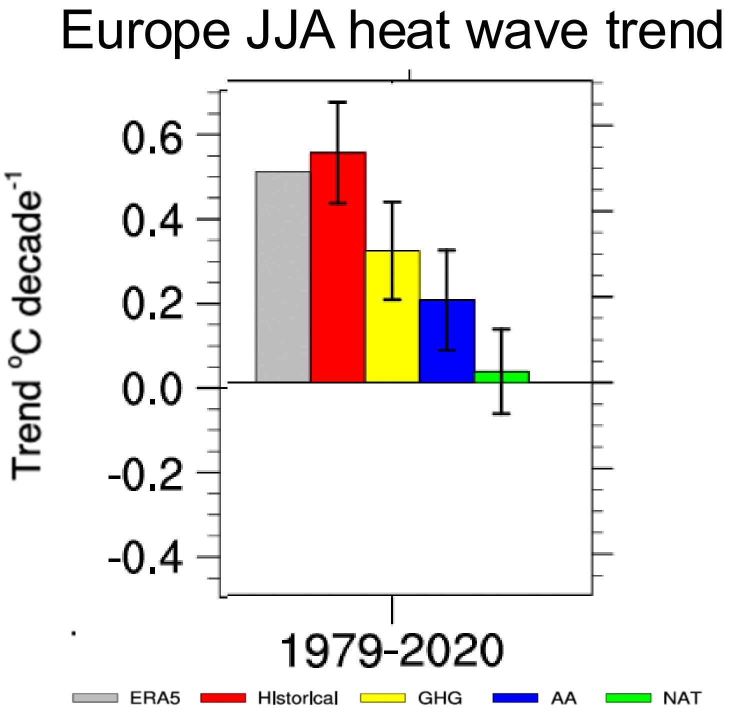
Rucker et al. (2025, in prep.)

Confronting Earth System model trends with observations is a new era

1) As more Earth system processes become incorporated into models, new opportunities will present themselves for confronting trends in these processes with observations
2) “State-of-the-signal” summary assessment on the current knowledge of historical trends and the ability of climate models to reproduce them (build on the current “State of the Climate”).
3) Advocating for more observations.
4) Another key opportunity for the climate science community will be to improve the existing model-observational comparison cycle (CMIP/IPCC).
5) Reliable real-time information on trends similar to climate reanalyzer
6) AGU & CMIP session proposal
7) CLIVAR Research Foci Proposal

Successes in capturing JJA heat wave trends over Europe

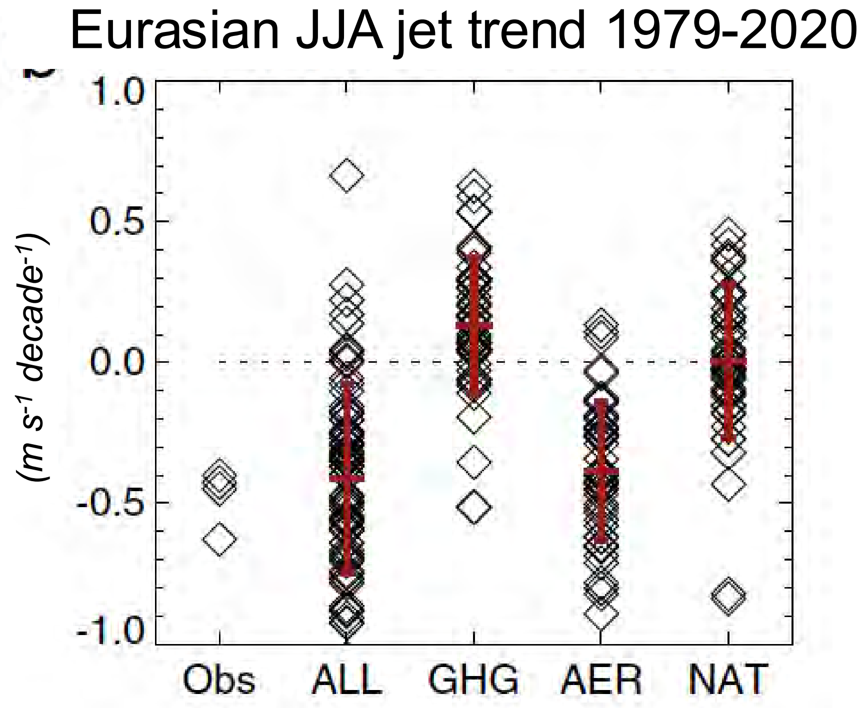




Dong & Sutton (2025, npj Climate & Atmos. Sci.)

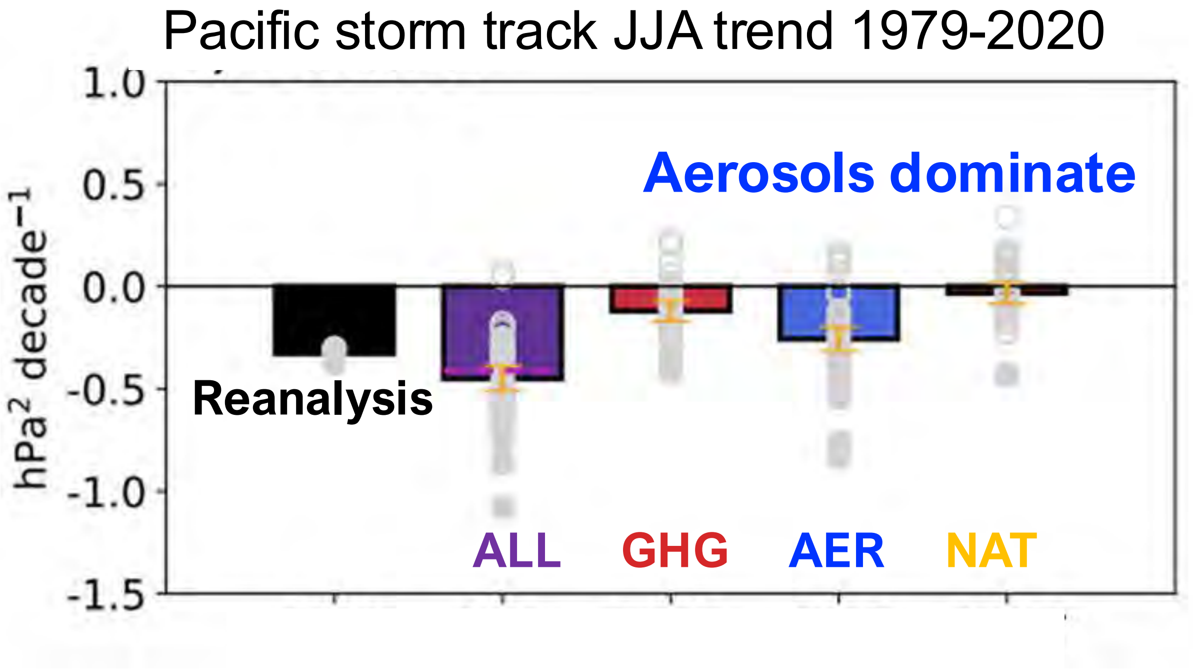
Yin et al. (2024, GRL)

Successes in capturing Eurasian jet stream weakening trend



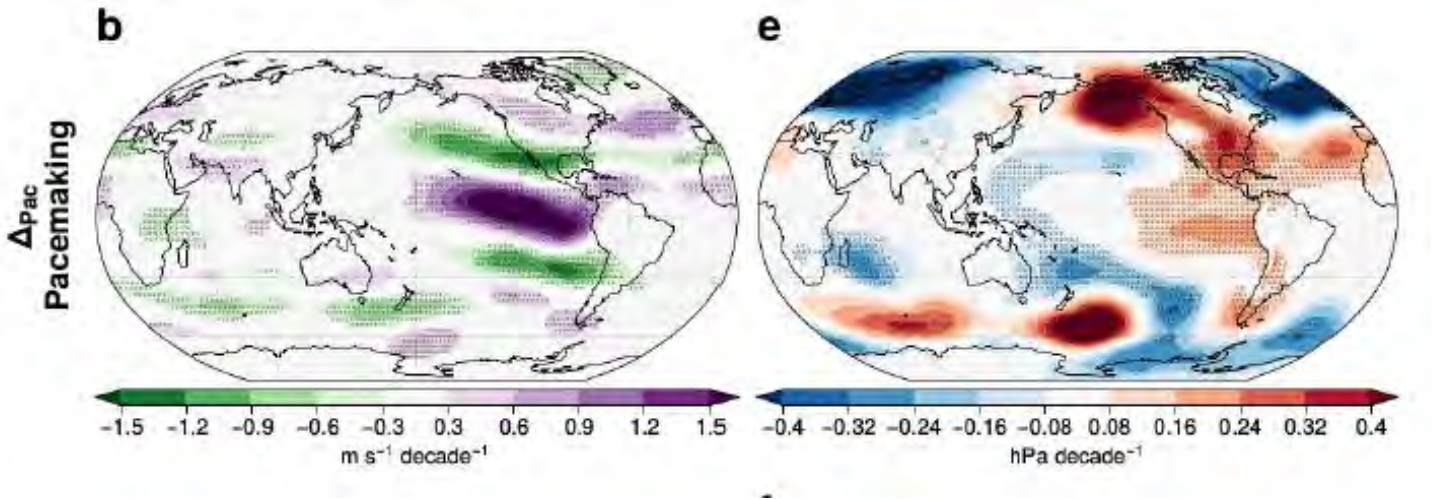
Dong et al. (2022, Nature Comm.)

Success in capturing Pacific storm track weakening trend



Kang et al. (2024, AGU Advances)

Discrepancy in tropical SST trends leads to teleconnection discrepancy



Kang et al. (2025, npj Climate & Atmos Sci)