Revisiting the reanalysis-model discrepancy in Southern Hemisphere winter storm track trends

Joonsuk Kang[1], Joonsuk M Kang[2], Tiffany A Shaw[2], Sarah M Kang[3], Isla R Simpson[4], and Yue Yu[5]

[1]Affiliation not available
[2]Department of the Geophysical Sciences, The University of Chicago
[3]Max Planck Institute of Meteorology
[4]Climate and Global Dynamics Laboratory, NSF National Center of Atmospheric Research
[5]State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources

April 04, 2024

Revisiting the reanalysis-model discrepancy in Southern Hemisphere winter storm track trends

Joonsuk M. Kang,[a] Tiffany A. Shaw,[a] Sarah M. Kang,[b] Isla R. Simpson,[c] Yue Yu,[d]

[a] *Department of the Geophysical Sciences, The University of Chicago, Chicago, IL*

[b] *Max Planck Institute of Meteorology, Hamburg, Germany*

[c] *Climate and Global Dynamics Laboratory, NSF National Center of Atmospheric Research, Boulder, CO*

[d] *State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China*

Corresponding author: Joonsuk M. Kang, jmkang@uchicago.edu

ABSTRACT: Reanalysis datasets show wintertime storminess in the Southern Hemisphere (SH) has significantly increased since 1979. Previous work reported a reanalysis-model discrepancy whereby coupled and prescribed sea surface temperature (SST) models in CMIP6 were unable to reproduce the trend. Here we revisit the reanalysis-model trend discrepancy in SH winter storminess focusing on the impact of observational uncertainty, model ensemble size, a like-for-like comparison, and mechanisms underlying the discrepancy. A large spread is found across available reanalyses indicating observational uncertainty. When the storminess trends in reanalysis and model datasets are quantified on the same time and spatial grids, the reanalysis trends decrease, and a discrepancy between reanalyses and prescribed SST models is unlikely. However, a discrepancy between reanalyses and coupled models is still likely, particularly in the South Pacific. We test the importance of SST trend discrepancies in coupled models using Southern Ocean and tropical Pacific pacemaker simulations. Under Southern Ocean pacemaking, a zonal-mean trend discrepancy between reanalyses and coupled models is unlikely and the improvement is due to the coupled models capturing surface energy flux trends. However, a discrepancy is still likely in the South Pacific. Under tropical Pacific pacemaking, a trend discrepancy between reanalyses and coupled models in the South Pacific is unlikely due to the coupled models capturing the La Nina-like teleconnection trend. Our results show that reanalysis-model trend comparisons should involve all reanalysis and model datasets and like-for-like calculations. Furthermore, regional SST trend discrepancies can lead to non-local reanalysis-model circulation trend discrepancies.

1. Introduction

The extratropical circulation in the Southern Hemisphere (SH) is characterized by a strong storm track related to tracks of cyclones and anticyclones (Hoskins and Hodges 2005; Shaw et al. 2016). The intensity of the storm track, hereafter the storminess, is tightly connected to surface weather in the SH (Pfahl and Wernli 2012; Pepler 2020), and understanding how storminess will change in the future is important (Shaw et al. 2016). Climate models project that the SH storm track will intensify by the end of the 21st century under climate change, (O'Gorman 2010; Chang et al. 2012; Shaw et al. 2016, 2018), bringing increased precipitation (Yettella and Kay 2017) and stronger surface winds (Chang 2017).

Recent work has shown that SH storminess has increased significantly in the satellite era in the reanalysis data from 1979 to present (Chemke et al. 2022; Shaw et al. 2022). However, the trends in reanalysis data were 2–3 times larger than the multi-model mean trends from models participating in the Coupled Model Intercomparison Project Phase 5 and Phase 6 (CMIP5 and CMIP6; Taylor et al. 2012; Eyring et al. 2016). Thus, recent work concluded that climate models significantly underestimate the storminess trend in the reanalysis datasets during the observed period. This reanalysis-model trend discrepancy in the SH winter storm track calls into question the ability of climate models to predict future weather in the SH.

A discrepancy between the climate model and observed trends can have multiple causes that can be categorized into three factors (Schmidt 2013): (I) The observations are in error, (II) The observation-model comparison is flawed, (III) The models are deficient.

For (I), the trends can differ substantially across observational datasets (Deser et al. 2010) and lead to observational uncertainty. The use of up-to-date observational data can be important in reconciling observation-model trend discrepancies (Santer et al. 2008; Grise et al. 2019). For storminess, the observed trend is quantified using reanalysis datasets, which involve uncertainties arising from data assimilation techniques, physical parameterizations, and evolution of observational systems (Bengtsson et al. 2004; Fujiwara et al. 2017). The reanalysis uncertainty can be particularly important in the SH where ground-based observations are limited and thus reanalysis trends can exhibit considerable spread (Guo and Chang 2008; Guo et al. 2009; Martineau et al. 2024). To account for this, previous work quantified trends across multiple reanalyses (Manney and Hegglin 2018; Grise et al. 2019; Dong et al. 2022a; Martineau et al. 2024).

For (II), there are two important aspects to consider. First, reanalysis trends involve a single realization of internal variability whereas model simulations reflect a distribution of realizations. Thus, it is important to properly sample the internal variability by using a large number of model simulations (Deser et al. 2020; Jain et al. 2023). Second, a like-for-like comparison whereby the observations and climate models are compared with the same temporal and spatial sampling has been important for reconciling previous discrepancies (Po-Chedley et al. 2015; Santer et al. 2017). A like-for-like comparison is especially important for storm tracks, which sample specific time and spatial scales (Chang et al. 2002). Finally, it is important to note that there is currently no agreed-upon method for comparing observed and modeled trends. Previous work, for example, used a rank metric (Suarez-Gutierrez et al. 2021) or similarly evaluated the percentile of reanalysis trend in the model trend distribution (Grise et al. 2019).

For (III), the models can be deficient in either the forced response or the internal variability because they are incapable of simulating the physical mechanism responsible for the observed trend. For example, CMIP6 models fail to simulate recent sea surface temperature (SST) trends in the tropical Pacific and Southern Ocean (Wills et al. 2022; Lee et al. 2022; Seager et al. 2022; Kang et al. 2023a). The tropical SST trends in CMIP6 models are characterized by an El Nino-like trend in the tropical Pacific, as opposed to a La Nina-like trend in the observations (Seager et al. 2022; Wills et al. 2022; Lee et al. 2022). The observed cooling trend in the Southern Ocean is also not well captured by the CMIP6 models (Wills et al. 2022) and it has been suggested that this is also related to the SST trend difference in the tropical Pacific (Dong et al. 2022b; Kang et al. 2023a). Previous work concluded that coupled models exhibit a systematic bias in the representation of SST trends and that differences between observed and modeled trends are very unlikely to occur due to internal variability (Wills et al. 2022). Many mechanisms have been proposed to explain the observation-model SST trend discrepancy (Lee et al. 2022; Seager et al. 2022), and it is not fully understood how these SST trend discrepancies impact the storminess trend during the SH winter. The impact of the SST trend discrepancy on storminess trends has not been quantified and should be investigated further given that SST trends are related to other large-scale circulation trends in the SH (Purich et al. 2016; Wills et al. 2022; Cox et al. 2024).

Here we revisit the reanalysis-model discrepancy in the Southern Hemisphere winter storm track trends and examine the impact of (I)–(III). We begin by outlining the data and methods in section 2.

For (I), we quantify the impact of doubling the number of reanalysis datasets compared to previous work in section 3. For (II), we quantify the impact of expanding model ensemble size and like-for-like comparison in section 3. For (III), we quantify the impact of the SST trend discrepancy on storm track trends, including the mechanisms connecting them, using the pacemaker simulations in section 4. We provide summary and discussions in section 5.

2. Data and Methods

a. Methods

We quantify storminess in the SH winter (June–August) using vertically integrated eddy kinetic energy (hereafter EKE), which is defined as

$$EKE = \frac{1}{g}\int_{p_t}^{p_s} u'^2 + v'^2 dp, \tag{1}$$

where g is the gravitational acceleration, p_s is the surface pressure, p_t is the pressure at the highest vertical level (Table 1), and u and v are zonal and meridional winds, respectively. Here, the primes denote 2.5–6 day bandpass-filtered anomalies. To produce the bandpass-filtered anomalies, timeseries of u and v with 92 days of SH winter padded with 10 days at both ends are first created. This equals 112 and 448 data points for daily and six-hourly data, respectively. We then apply a first-order Butterworth filter to the time series to obtain 2.5–6 day bandpass-filtered anomalies. We use p_s data that has the same time frequency as u and v for most datasets, but monthly-mean p_s when high-frequency data is not available.

After quantifying storminess each year, the long-term trends are calculated using the least-squares linear regression. The statistical significance of the trend is evaluated as the 95% confidence level using a two-sided t-test.

b. Reanalysis datasets

Storminess trends are quantified in six reanalysis datasets (observation-based products) that span the time period from 1979 to 2018: NCEP2 (Kanamitsu et al. 2002), ERA-Interim (Dee et al. 2011), JRA-55 (Kobayashi et al. 2015), CFSR/CFSv2 (Saha et al. 2010, 2014), MERRA2 (Gelaro et al. 2017) and ERA5 (Hersbach et al. 2020). Only the first three reanalysis products were used in Chemke et al. (2022). We focus on the 40-year time period following previous work. We use six-hourly instantaneous variables, which is the highest frequency common to all reanalysis datasets, although ERA5 and MERRA2 data are available at higher frequency. The CFSR trend is obtained by merging CFSR (1979–2010) and CFSv2 (2011–2018) datasets. MERRA2 starts in 1980, so its trends are calculated from 1980.

c. CMIP6 and AMIP6 simulations

Storminess trends are quantified in 26 CMIP6 model simulations (Eyring et al. 2016) using the historical (1979 to 2014) and SSP5-8.5 (2015 to 2018) scenarios (Table 1). We use the SSP5-8.5 scenario following previous work (Chemke et al. 2022). Scenario uncertainty is a small contributor since the scenarios begin in 2015. In addition, we quantify storminess in 32 AMIP6 model simulations (Table 1) with observed SSTs prescribed from 1979 to 2014. We refer to the CMIP6 and AMIP6 model simulations as multi-model ensembles. The difference between the CMIP6 and AMIP6 multi-model ensembles quantifies the impact of discrepancies in SST trends in the CMIP6 models (Lee et al. 2022; Seager et al. 2022) on storminess trends. We quantify the statistical significance of the difference between trend distributions in the multi-model ensembles using the Mann-Whitney U test (hereafter MW test; Mann and Whitney 1947) at the 95% level (p-value < 0.05), which is a non-parametric statistical test. The number of models used in each ensemble is based on the availability of daily-mean zonal and meridional wind data on pressure levels. We use the 'r1i1p1f1' ensemble member for all models to equally weight the structural uncertainty across different models.

d. Like-for-like comparison

For reanalyses, CMIP6, and AMIP6 models, we calculate the EKE on the unprocessed spatial and time grids similar to previous work (Chemke et al. 2022). The unprocessed time grid refers to daily-mean for the models and six-hourly instantaneous for the reanalysis. The unprocessed spatial grid refers to different horizontal and vertical grids listed in Table 1. We also perform a like-for-like comparison. This is needed because the time frequency and spatial grids (Table 1) are very different across the different datasets. In order to perform a like-for-like comparison, we time-average six-hourly reanalysis u and v data into daily-mean. Next, we linearly interpolate both

TABLE 1. Summary of unprocessed time frequency and spatial grids of the datasets used in the study. For datasets with unevenly spaced horizontal grids, the horizontal resolution represents an average grid spacing in the longitude and latitude directions. The models that are only included in the CMIP6 and AMIP6 multi model ensemble are superscripted in c or a, respectively.

Dataset	Time frequency	Number of pressure levels (p_t)	Horizontal resolution [longitude×latitude]
Reanalysis			
ERA-Interim	Six-hourly	37 (1 hPa)	$1.5^o \times 1.5^o$
JRA-55	Six-hourly	37 (1 hPa)	$1.25^o \times 1.25^o$
NCEP2	Six-hourly	17 (10 hPa)	$2.5^o \times 2.5^o$
MERRA2	Six-hourly	42 (0.1 hPa)	$0.625^o \times 0.5^o$
ERA5	Six-hourly	37 (1 hPa)	$0.25^o \times 0.25^o$
CFSR/CFSv2	Six-hourly	37 (1 hPa)	$0.5^o \times 0.5^o$
CMIP6 and AMIP6 models			
ACCESS-CM2	Daily-mean	8 (10 hPa)	$1.875^o \times 1.25^o$
ACCESS-ESM1-5	Daily-mean	8 (10 hPa)	$1.875^o \times 1.25^o$
BCC-CSM2-MR	Daily-mean	8 (10 hPa)	$1.125^o \times 1.125^o$
CAMS-CSM1-0^a	Daily-mean	8 (10 hPa)	$1.125^o \times 1.125^o$
CESM2^a	Daily-mean	8 (10 hPa)	$1.25^o \times 0.938^o$
CESM2-FV2^a	Daily-mean	8 (10 hPa)	$2.5^o \times 1.875^o$
CESM2-WACCM	Daily-mean	8 (10 hPa)	$1.25^o \times 0.938^o$
CESM2-WACCM-FV2^a	Daily-mean	8 (10 hPa)	$2.5^o \times 1.875^o$
CMCC-CM2-HR4^a	Daily-mean	8 (10 hPa)	$1.25^o \times 0.938^o$
CMCC-CM2-SR5	Daily-mean	8 (10 hPa)	$1.25^o \times 0.938^o$
CMCC-ESM2^c	Daily-mean	8 (10 hPa)	$1.25^o \times 0.938^o$
CanESM5	Daily-mean	8 (10 hPa)	$2.812^o \times 2.812^o$
EC-Earth3	Daily-mean	8 (10 hPa)	$0.703^o \times 0.703^o$
EC-Earth3-CC	Daily-mean	8 (10 hPa)	$0.703^o \times 0.703^o$
EC-Earth3-AerChem^a	Daily-mean	8 (10 hPa)	$0.703^o \times 0.703^o$
EC-Earth3-Veg	Daily-mean	8 (10 hPa)	$0.703^o \times 0.703^o$
EC-Earth3-Veg-LR	Daily-mean	8 (10 hPa)	$1.125^o \times 1.125^o$
FGOALS-f3-L^a	Daily-mean	8 (10 hPa)	$1.25^o \times 1.0^o$
FGOALS-g3	Daily-mean	8 (10 hPa)	$2.0^o \times 2.25^o$
GFDL-CM4	Daily-mean	8 (10 hPa)	$2.5^o \times 2.0^o$
IITM-ESM	Daily-mean	8 (10 hPa)	$1.875^o \times 1.915^o$
INM-CM4-8	Daily-mean	8 (10 hPa)	$2.0^o \times 1.5^o$
INM-CM5-0	Daily-mean	8 (10 hPa)	$2.0^o \times 1.5^o$
IPSL-CM6A-LR	Daily-mean	8 (10 hPa)	$2.5^o \times 1.268^o$
KACE-1-0-G^c	Daily-mean	8 (10 hPa)	$1.875^o \times 1.25^o$
MIROC6	Daily-mean	8 (10 hPa)	$1.406^o \times 1.406^o$
MPI-ESM-1-2-HAM^a	Daily-mean	8 (10 hPa)	$1.875^o \times 1.875^o$
MPI-ESM1-2-HR	Daily-mean	8 (10 hPa)	$0.938^o \times 0.938^o$
MPI-ESM1-2-LR	Daily-mean	8 (10 hPa)	$1.875^o \times 1.875^o$
MRI-ESM2-0	Daily-mean	8 (10 hPa)	$1.125^o \times 1.125^o$
NESM3	Daily-mean	8 (10 hPa)	$1.875^o \times 1.875^o$
NorESM2-LM	Daily-mean	8 (10 hPa)	$2.5^o \times 1.875^o$
NorESM2-MM^c	Daily-mean	8 (10 hPa)	$1.25^o \times 0.938^o$
SAM0-UNICON^a	Daily-mean	8 (10 hPa)	$1.25^o \times 0.938^o$
TaiESM1	Daily-mean	8 (10 hPa)	$1.25^o \times 0.938^o$

reanalysis and climate model data onto a common 1.5$^{o}\times$1.5^{o} grid. Then, vertical integration is performed over 8 pressure levels (1000, 850, 700, 500, 250, 100, 50, and 10 hPa), which are the standard model output levels for CMIP6 daily data. Note that for reanalysis data, we subsample the vertical grid by extracting the 8 CMIP6 pressure levels.

e. CESM2 large ensemble and pacemaker simulations

In order to quantify the impact of internal variability on the SH winter storminess trend discrepancy, we use the Community Earth System Model version 2 Large Ensemble (CESM2-LE) simulations (Danabasoglu et al. 2020; Rodgers et al. 2021). The CESM2-LE simulations are an initial condition ensemble with a nominal 1-degree spatial resolution in both atmosphere and ocean. We use the first 50 simulations from this ensemble that are forced with historical radiative forcing and standard biomass burning from 1850 to 2014 consistent with CMIP6 simulations. The SST trends in the CESM2-LE simulations during this period fail to capture the observed SST trends in the tropical Pacific and Southern Ocean (Wills et al. 2022; Kang et al. 2023a) consistent with the CMIP6 multi-model ensemble.

To investigate the impact of SST trend discrepancies on the reanalysis-model trend discrepancy in the SH winter storm track, we also use the Southern Ocean pacemaker simulations (hereafter called SOPACE; Kang et al. 2023a) and Pacific pacemaker simulations (hereafter called PacPACE, see `https://www.cesm.ucar.edu/working-groups/climate/simulations/cesm2-pacific-pacemaker` for details). The SOPACE and PacPACE simulations have 21 and 10 ensemble members, respectively. The same CMIP6 historical forcing is used for SOPACE (1979–2013) and PacPACE simulations (1880–2014). They have the same horizontal resolution as CESM2-LE. They are fully coupled except in the regions where SST anomalies (relative to observed 1880–2019 climatology) are nudged to observed SST anomalies from ERSSTv5 (Huang et al. 2017). More specifically, in SOPACE, SST anomalies are nudged to observations poleward of 40^{o}S. In PacPACE, SST anomalies are nudged to observation within a wedge-shaped area of 20^{o}S–20^{o}N from the American coast to the western Pacific. We quantify the impact of pacemaking on the simulated trends as

$$\begin{aligned} \Delta_{SO} &= [\text{SOPACE}] - [\text{CESM2-LE}], \\ \Delta_{Pac} &= [\text{PacPACE}] - [\text{CESM2-LE}], \end{aligned} \tag{2}$$

where the squared brackets denote the ensemble mean (Kang et al. 2023a).

Finally, we utilize AMIP-style CESM2 simulations, namely Global Ocean Global Atmosphere (GOGA) simulations, with 10 members. The GOGA simulations are forced with the same CMIP6 historical forcing from 1880 to 2014 and take observed SSTs from ERSSTv5 as boundary conditions. We quantify the impact of SST trend discrepancy by comparing the trend distributions in CESM2-LE and GOGA simulations using the MW test.

f. Comparing storminess trends in reanalysis and models

As mentioned in the introduction, our goal is to revisit the reanalysis-model discrepancy of SH winter storminess trends. Our starting point is to use the same reanalyses and models used in previous work (Chemke et al. 2022). We then perform the following steps. First, we double the number of reanalysis datasets in order to quantify the impact of observational uncertainty. Second, we expand the model ensemble size to include a broader range of internal variability and structural uncertainty. Third, we calculate reanalysis and model storminess trends on the same time and spatial grids, to ensure a like-for-like comparison.

At each step, we quantitatively compare reanalysis and model trends and evaluate the likeliness of a discrepancy using a rank metric (e.g., Hamill 2001; Suarez-Gutierrez et al. 2021). The rank metric assesses the ranking of reanalyses within the model distribution and the null hypothesis is that the reanalysis is interchangeable with the model simulations and represents a random draw of a single realization from the model distribution. If the models correctly represent the forced trend and the range of internal variability, we expect the reanalysis to have a rank that sits squarely within the model distribution. The rank quantifies the probability of sampling a storminess trend as large as found in reanalysis from the model distribution. For example, a rank of 20% would indicate that there is only a 20% chance of obtaining storminess trends larger than the reanalysis trend. The closer the rank is to 0%, the more the models underestimate the reanalysis trend, indicating a reanalysis-model trend discrepancy. The rank method is consistent with evaluating whether reanalysis trends fall within certain percentiles of the model trend distributions (e.g. Grise et al. 2019).

To evaluate where trends of reanalyses sit on average within the model trend distribution, metrics such as the average of reanalysis ranks and rank of average reanalysis trends can be considered. While both are useful, we focus on the average of reanalysis ranks (hereafter average rank), a non-parametric approach, to prevent an overly strong influence from outliers. This is particularly important for the SH where reanalysis trends exhibit a large spread.

In order to summarize our reanalysis-model trend comparison, we use the following verbal expressions that loosely follow the IPCC language (Mastrandrea et al. 2010). If the average rank of reanalysis trends is $< 5\%$ (or $> 95\%$), a discrepancy is "very likely". If the average rank is $\geq 5\%$ and $< 20\%$ (or $> 80\%$ and $\leq 95\%$), a discrepancy is "likely". If neither conditions are met (i.e., average rank is between 20% and 80%), then a discrepancy is "unlikely".

Table 2. Rank of individual reanalysis in the multi-model or large ensemble simulations in percentage. The rightmost column shows the likeliness of a discrepancy expressed loosely following IPCC language, according to the average rank. See text for details.

Rank (%)	ERAI	JRA55	NCEP2	ERA5	MERRA2	CFSR	Average	Discrepancy
All reanalysis (Fig. 1a)								
CMIP6	0.0	0.0	0.0	0.0	6.3	12.5	3.1	very likely
AMIP6	23.1	0.0	0.0	7.7	46.2	53.8	21.8	unlikely
All models (Fig. 1b)								
CMIP6	3.8	0.0	0.0	3.8	15.4	23.1	7.7	likely
AMIP6	18.8	0.0	0.0	6.3	43.8	46.9	19.3	unlikely
All datasets & Like-for-like (Fig. 1c)								
CMIP6	3.8	3.8	0.0	3.8	38.5	23.1	12.2	likely
AMIP6	25.0	0.0	0.0	18.8	75.0	53.1	28.6	unlikely
South Pacific (Fig. 4)								
CMIP6	11.5	0.0	0.0	7.7	19.2	46.2	14.1	likely
AMIP6	53.1	12.5	0.0	50.0	68.8	90.6	45.8	unlikely
CESM2 zonal mean (Figs. 5a and 9a)								
CESM2-LE	2.0	0.0	0.0	0.0	56.0	56.0	19.0	likely
GOGA	20.0	0.0	0.0	10.0	80.0	80.0	31.7	unlikely
SOPACE	14.3	0.0	0.0	9.5	71.4	71.4	27.7	unlikely
PacPACE	10.0	0.0	0.0	10.0	60.0	60.0	23.3	unlikely
SUM	20.0	0.0	0.0	12.0	86.0	86.0	34.0	unlikely
CESM2 South Pacific (Figs. 5b and 9b)								
CESM2-LE	10.0	2.0	0.0	8.0	26.0	62.0	18.0	likely
GOGA	70.0	20.0	0.0	50.0	100.0	100.0	56.6	unlikely
SOPACE	0.0	0.0	0.0	0.0	23.8	85.7	18.3	likely
PacPACE	40.0	10.0	0.0	30.0	80.0	90.0	41.7	unlikely
SUM	48.0	20.0	0.0	40.0	72.0	96.0	46.0	unlikely

3. Impact of observation uncertainty, model ensemble size, and like-for-like comparison on the storminess trend discrepancy

We quantify the impact of observation uncertainty, model ensemble size, and like-for-like comparison on the SH winter storminess trend discrepancy. We start by comparing reanalysis and CMIP6 model trends, focusing on the impact of these three factors, and then assess the AMIP6 trends.

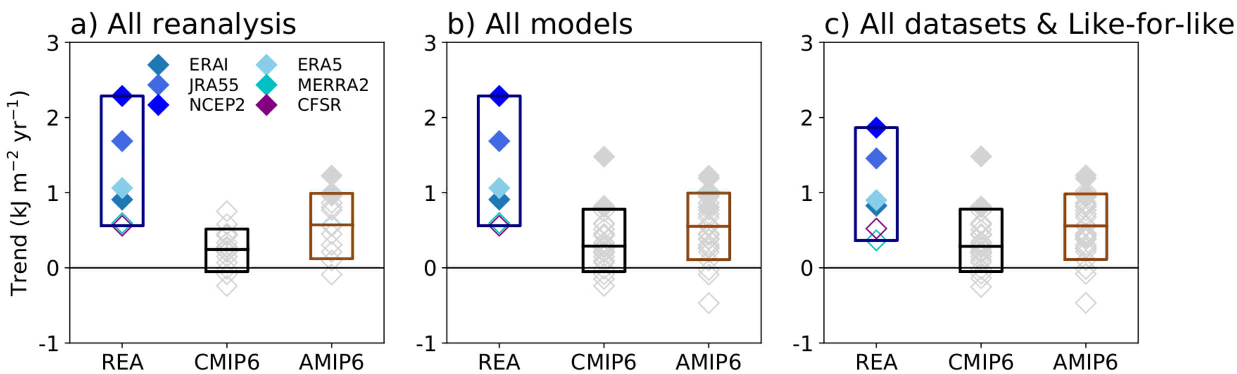

FIG. 1. (a) Linear trends of SH JJA EKE (40–70^{o}S) in 6 reanalysis datasets (blue colors, 1979–2018) and 16 CMIP6 (1979–2018) and 13 AMIP6 (1979–2014) model simulations (diamonds). Statistically significant trends at the 95% confidence level are filled. The box represents the full spread of reanalysis trends and the 10–90% percentile of model ensemble trends. The horizontal line inside the box shows the median trend in the model ensemble. (b) Similar results to (a), but for 26 CMIP6 and 32 AMIP6 models. (c) Similar results to (b), but after performing a like-for-like calculation.

a. Reanalyses-CMIP6 comparison

We start with 16 CMIP6 models and 3 reanalysis datasets used in previous work (Chemke et al. 2022) but add 3 more modern reanalysis datasets (CFSR, ERA5, and MERRA2), which extend to 2018 and have been used in other previous work (Manney and Hegglin 2018; Martineau et al. 2024; Cox et al. 2024). The EKE is calculated on the unprocessed time and spatial grids for each dataset. The trends are calculated from 1979 to 2018 in the reanalysis datasets (except for MERRA2) and CMIP6 models following previous work (Chemke et al. 2022). The storminess trends in the reanalysis datasets show a large spread from 0.56 kJ m^{-2} yr^{-1} to 2.29 kJ m^{-2} yr^{-1}, and not all trends are statistically significant (MERRA2 and CFSR, Fig. 1a). For the 16 CMIP6 model ensemble, 4 reanalysis trends have zero ranks, and MERRA2 and CFSR trends have small non-zero ranks (Table 2). According to the average rank (3.1%), a reanalysis-model discrepancy is very likely for the CMIP6 ensemble after accounting for observational uncertainty.

It is important to note previous work documented a climatological bias in SH storminess in NCEP2 (Fig. 2, see also Guo and Chang 2008; Guo et al. 2009; Martineau et al. 2024). If NCEP2 is excluded for this reason, we obtain similar results (average rank is 3.8%). If ERA-Interim is also excluded because it is a direct predecessor of ERA5, we also get similar results (average rank is 4.7%). Thus, we proceed with using all 6 reanalysis datasets and include their ranks in Table 2.

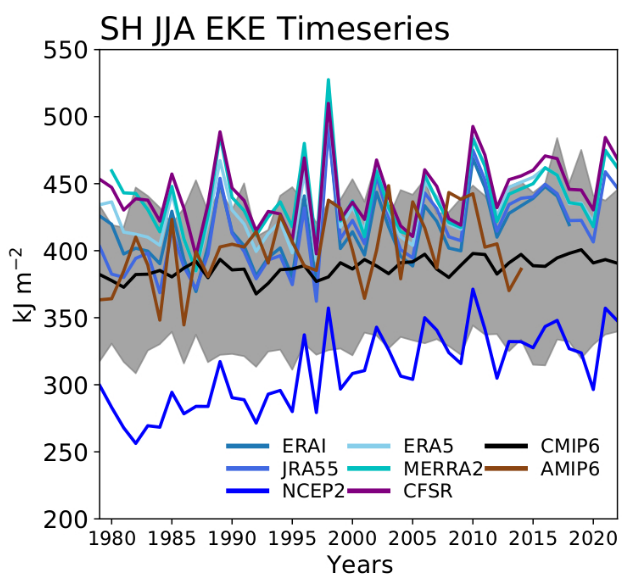

FIG. 2. (a) Time series of SH JJA EKE (40–70°S) for 6 reanalyses datasets (blue colors), CMIP6 (black) and AMIP6 (brown) multi-model mean. The 10–90% percentile of the 26 CMIP6 models is shown in gray shading. Note that EKE is calculated using the like-for-like method.

Next, we assess the impact of expanding the model ensemble size from 16 to 26 CMIP6 models (Fig. 1b). Increasing the number of models in the ensemble increases the ranks of all reanalysis trends, and the resulting average rank is 7.7% (Table 2). Some newly added CMIP6 models show a statistically significant trend, with one model (EC-Earth3) having a greater trend than four reanalysis datasets (Fig. 1b). However, further examination of the storminess trends in the 14 ensemble members of the EC-Earth3 model family reveal the trend is an outlier within the model family (not shown). Thus, according to the average rank (7.7%), a reanalysis-model discrepancy is likely after accounting for the model ensemble size.

Lastly, we examine the impact of the like-for-like comparison by calculating EKE trends in the 6 reanalyses and 26 CMIP6 models in the same time and spatial grids (see section 2d). The like-for-like reanalysis trends exhibit a noticeable decrease in amplitude (Fig. 1c), which is mostly due to calculating EKE using daily-mean instead of six-hourly data (not shown). Interestingly, there still exists a significant spread across reanalysis trends (0.36–1.86 kJ m^{-2} yr^{-1}), which is larger than the model ensemble spread (compare boxes in Fig. 1c). After accounting for the like-for-like trend comparison, according to the average rank (12.2%, Table 2), a reanalysis-model discrepancy is likely for the CMIP6 ensemble.

The results show that accounting for observation uncertainty, model ensemble size, and like-for-like comparison reduces the discrepancy between reanalysis and CMIP6 model trends. However, a reanalysis-coupled model trend discrepancy is still likely after accounting for these factors. In general, the CMIP6 model trends are a combination of internal variability and forced response. We find using CMIP6 Detection and Attribution Model Intercomparison Project simulations (DAMIP, Gillett et al. 2016) that the forced response to greenhouse gas emissions (hist-GHG) dominates the trends (Fig. S1).

b. Reanalyses-AMIP6 comparison

We start by comparing trends in 13 AMIP6 models (1979–2014) used in the previous study (Chemke et al. 2022) and 6 reanalyses (Fig. 1a). The EKE is calculated on the unprocessed time and spatial grids. According to the average rank (21.8%, Table 2), a reanalysis-model trend discrepancy is unlikely after accounting for observational uncertainty.

Next, we quantify the impact of increasing the AMIP6 model ensemble size to 32. The reanalysis ranks decrease slightly (Table 2) as a consequence of some newly added models showing negative trends (compare Figs. 1a and b). According to the average rank (19.3%), a reanalysis-model discrepancy is likely for the expanded AMIP6 ensemble. The different result from the case with 13 AMIP6 models reveals the sensitivity of reanalysis-model discrepancy to the model ensemble size.

Lastly, we quantify the impact of a like-for-like comparison by calculating EKE trends in the 6 reanalyses and 32 AMIP6 models on the same time and spatial grids (Fig. 1c). After a like-for-like calculation, the ranks increase for all reanalyses, resulting in an average rank of 28.6% (Table 2).

According to the average rank (28.6%), a reanalysis-model trend discrepancy is unlikely for the AMIP6 ensemble. This highlights the importance of like-for-like comparison when evaluating reanalysis and model trends.

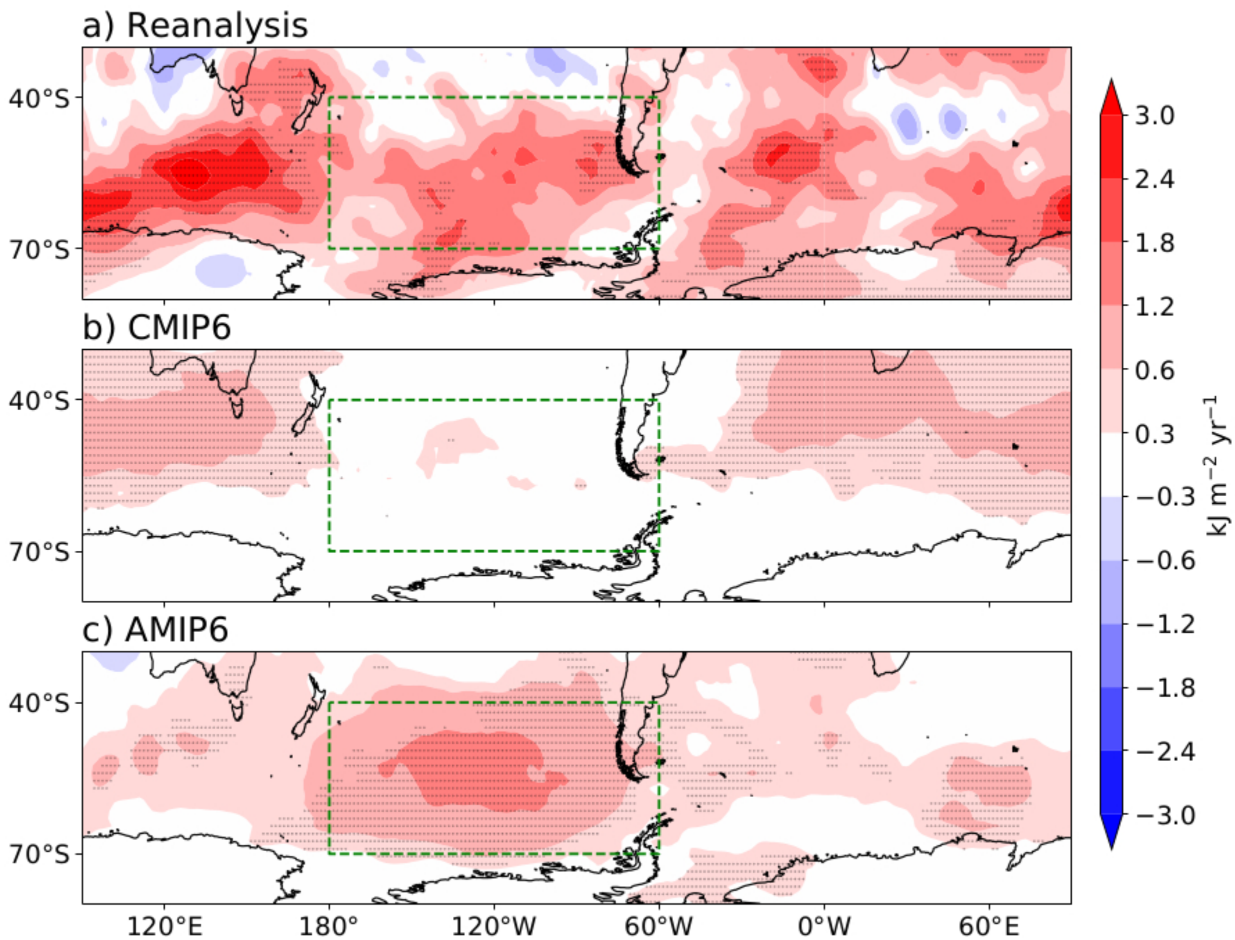

FIG. 3. Spatial pattern of SH JJA EKE trend during 1979–2014 for (a) reanalysis mean (CFSR, ERAI, ERA5, JRA55, MERRA2, NCEP2), (b) CMIP6 and (c) AMIP6 multi-model mean. Stipples indicate where reanalysis-mean or multi-model mean trends are significant at the 95% level. The green dashed lines indicate the South Pacific sector (40–70°S, 180–60°W).

4. Impact of SST trend discrepancies on storminess trends

The results thus far show observational uncertainty, model ensemble size, and a like-for-like comparison significantly impact the reanalysis-model trend discrepancy in the SH winter storminess. After accounting for these three factors, a reanalysis-model trend discrepancy is unlikely for

AMIP6 but likely for CMIP6, according to the average rank. Consistently, the AMIP6 trends are significantly larger than the CMIP6 trends according to the MW test (p-value = 0.01). Our EKE trend results are consistent with Cox et al. (2024), who showed trends in annual-mean atmospheric energy transport from transient eddies in coupled models did not agree with reanalyses (see their Fig. 2).

We further examined the sensitivity of our results to different start and end years. When the trend calculation is repeated for different start years from 1979 to 1985, we find the results are robust (Fig. S2a). When we only consider reanalysis trends from 1979 and 2014 to match those in AMIP6 models (additional factor for like-for-like comparison), the average rank increases to 38.0%, further supporting the conclusion that a reanalysis-AMIP6 trend discrepancy is unlikely. The results are also robust to extending the time series from 2018 to 2022 (Fig. S2b). Overall, our results are not sensitive to the specific start and end years used to calculate storminess trends.

The spatial pattern of the storminess trends from 1979 to 2014 (common period for reanalyses and CMIP6 and AMIP6 models) provides additional insights into understanding the different results for CMIP6 and AMIP6 (Fig 3). The storminess trend in reanalysis is significant across all longitudes including high latitudes of the Indian Ocean, South Pacific, and South Atlantic (Fig 3a and Fig. S3). However, the CMIP6 multi-model mean storminess trends in the South Pacific are negligible (Fig. 3b). The AMIP6 multi-model mean better captures the reanalysis trend, especially in the South Pacific, where CMIP6 models show no trends (compare Figs. 3b and 3c). More quantitatively, in the South Pacific, according to the average rank (14.1%, Table 2), a reanalysis-model discrepancy is likely for the CMIP6 ensemble (Fig. 4). In contrast, according to the average rank (45.8%, Table 2), a reanalysis-model trend discrepancy is unlikely for the AMIP6 ensemble (Fig. 4).

Since the reanalysis-model trend discrepancy in the SH winter storminess is now strictly only for coupled models and not for prescribed-SST models, we hypothesize it is related to SST trend discrepancies (Fig. S4). The SST trend can be connected to the storminess trend through different mechanisms. Shaw et al. (2022) suggested the SH storminess trends in CMIP6 are weaker than reanalyses because CMIP6 models do not capture surface energy flux trends across the SH which is related to SST trends across the Southern Ocean (Armour et al. 2016). Furthermore, the SST trend discrepancy in the tropical Pacific likely impacts the SH through teleconnections. More specifically, a La Nina-like SST trend would be expected to strengthen the South Pacific storminess (Seager et al. 2003; Nakamura et al. 2004; Ashok et al. 2007).

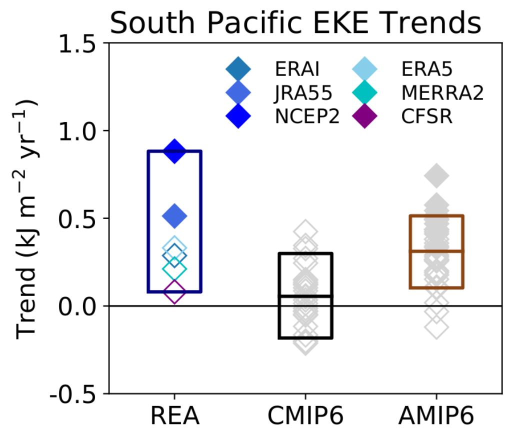

FIG. 4. Similar results to Fig. 1c, but for the South Pacific (40–70°S, 180–60°W). The trends are calculated from 1979 to 2014.

In order to test the hypothesis that SST trend discrepancies contribute to the reanalysis-CMIP6 model SH winter storminess trend discrepancy, we utilize the Southern Ocean (SOPACE) and tropical Pacific (PacPACE) pacemaker simulations. The pacemaker simulations allow us to quantify how SH storminess trends in the coupled simulations would change if the coupled models simulated the observed SST trend. To connect the CESM2 pacemaker simulations to the CMIP6 model ensemble, we use the CESM2-LE simulations that are forced with the same radiative forcing as CMIP6 models. Similar to the CMIP6 and AMIP6 model ensemble comparison, we also compare CESM2-LE and GOGA simulations.

For the CESM2 models, EKE is calculated using the monthly-mean kinetic energy output due to data availability (e.g., Kang et al. 2023b):

$$EKE = \frac{1}{g}\int_{p_t}^{p_s}\left(\overline{u^2}+\overline{v^2}-\overline{u}^2-\overline{v}^2\right)dp, \tag{3}$$

where the $\overline{u^2}$ and $\overline{v^2}$ are the monthly averages of the square of u and v at each model time step (every 30 minutes). As such, this EKE represents the kinetic energy due to sub-monthly variations. For most reanalysis datasets except ERA5, $\overline{u^2}$ and $\overline{v^2}$ at model time step is not provided and it has to be calculated from six-hourly data. However, for ERA5, we find that the difference of calculating

$\overline{u^2}+\overline{v^2}$ at model time step versus six-hourly time step is negligible (about 0.1%, Fig. S5). As in the previous section, we extract the 8 pressure levels from all reanalysis datasets. The CESM2 data are interpolated from model levels to the 8 pressure levels. Then, both reanalysis and CESM2 data are linearly interpolated onto a common 1.5°×1.5° grid and vertically integrated over the 8 pressure levels (p_t = 10 hPa). Here we focus on the trend from 1979 to 2013, which is the common period for the CESM2 simulations.

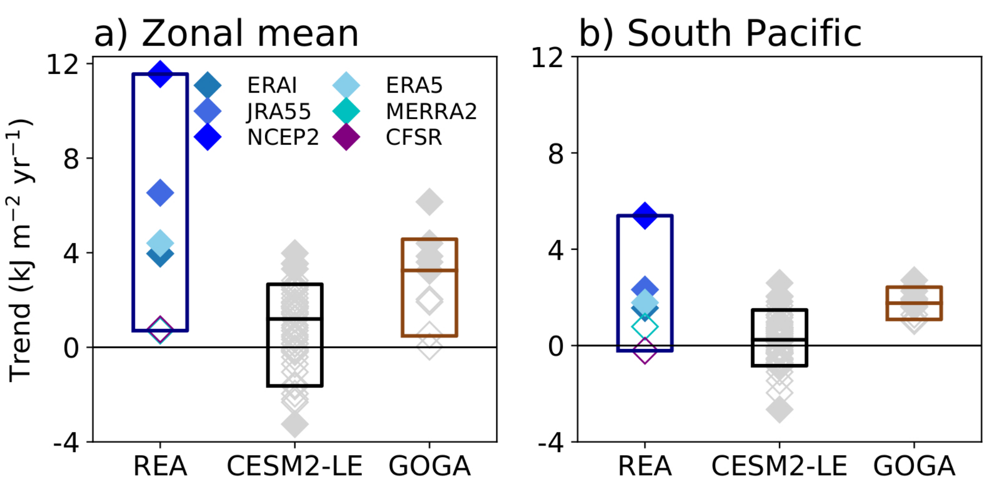

FIG. 5. (a) Linear trends of zonal-mean SH JJA EKE (40–70°S) in 6 reanalysis datasets and CESM2-LE and GOGA simulations (1979–2013, diamonds). Statistically significant trends at the 95% confidence level are filled. The box represents the full spread of reanalysis trends and the 10–90% percentile of model ensemble trends. The horizontal line inside the box shows the median trend in the model ensemble. (b) Similar results to (a), but for the South Pacific (40–70°S, 180–60°W).

When zonal-mean storminess trends in the CESM2-LE simulations are quantified and compared to those in reanalyses (Fig. 5a), a reanalysis-model discrepancy is likely, according to the average rank (19.0%, Table 2). Note that while the average rank is close to 20%, 4 reanalyses have ranks smaller than 5%. The CESM2-LE simulations also show negligible ensemble-mean storminess trends across the South Pacific similar to the CMIP6 models (compare Figs. 3a and b, Figs. 6a and b). More quantitatively, a reanalysis-model discrepancy is likely according to the average rank (18.0%, Table 2) in the South Pacific (Fig. 5b). Additionally, the SST trend discrepancies in the CESM2-LE simulations are similar to those in the CMIP6 models (compare Figs. 7a and b and Figs. S4a and b). Thus, trends in the 50-member CESM2-LE simulations indicate internal variability is unlikely to be the reason for the reanalysis-model trend discrepancy.

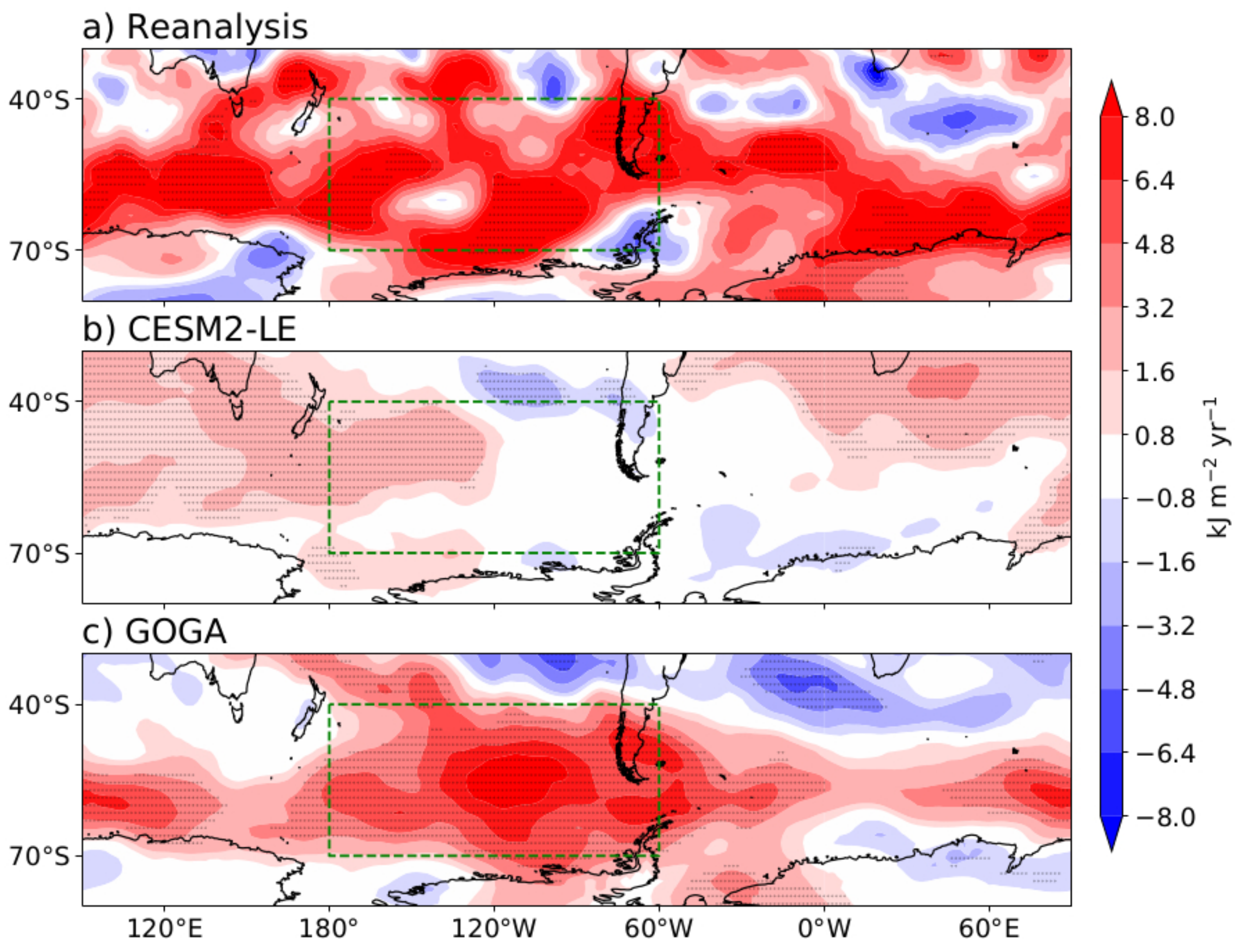

FIG. 6. Spatial pattern of SH JJA EKE trend during 1979–2013 for (a) reanalysis mean (CFSR, ERAI, ERA5, JRA55, MERRA2, NCEP2), (b) CESM2-LE and (c) GOGA ensemble mean. Stipples indicate where reanalysis-mean or ensemble-mean trends are significant at the 95% level. The green dashed lines indicate the South Pacific sector (40–70^{o}S, 180–60^{o}W). Note the EKE is defined differently from Fig. 3.

When zonal-mean storminess trends are compared between GOGA simulations and reanalyses (Fig. 5a), a reanalysis-model discrepancy is unlikely, according to the average rank (31.7% Table 2). The GOGA simulations show significant ensemble-mean storminess trends in the South Pacific (Fig. 6c). Consistently, a reanalysis-model trend discrepancy is unlikely for the South Pacific (Fig. 5b) according to the average rank (56.6%, Table 2). Moreover, trends in the GOGA simulations are significantly larger than trends in the CESM2-LE simulations in the zonal mean (MW test p-value= 0.00) and South Pacific (MW test p-value= 0.00).

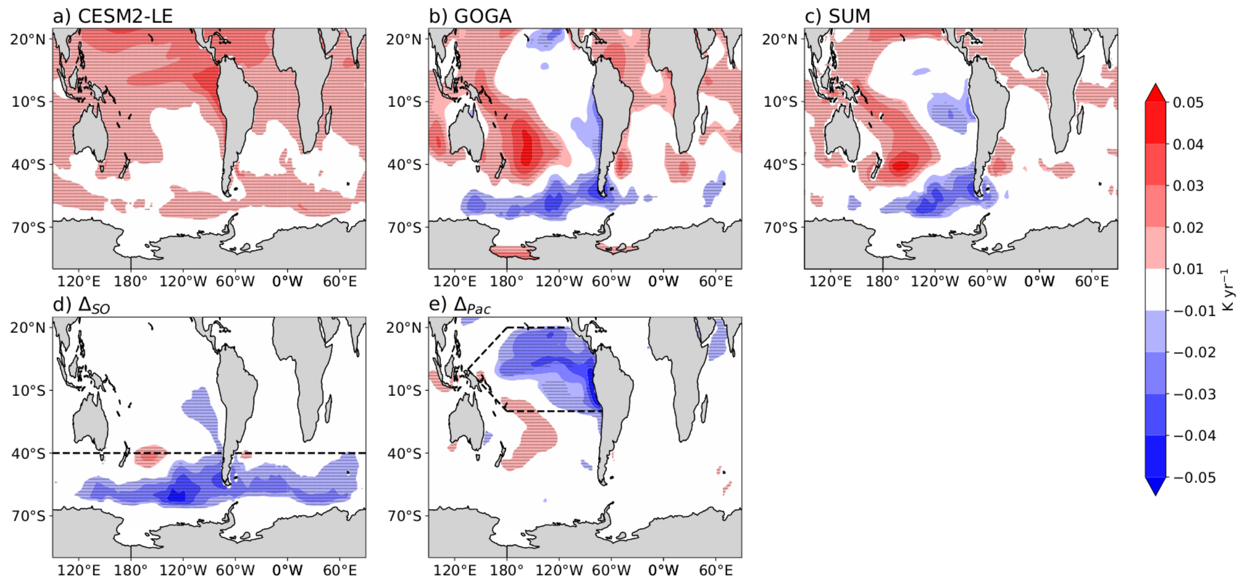

Fig. 7. Spatial pattern of ensemble-mean JJA SST trend from 1979 to 2013 for (a) CESM2-LE, (b) GOGA, (c) SUM= CESM2-LE + Δ_{Pac} + Δ_{SO}, (d) Δ_{SO} = [SOPACE] − [CESM2-LE], and (e) Δ_{Pac} = [PacPACE] − [CESM2-LE] simulations. Stipples indicate where ensemble-mean trends are significant at the 95% level. In (d) and (e), the dashed black lines represent where the SST anomalies are nudged to observation.

The similarity of CESM2-LE and GOGA simulations to CMIP6 and AMIP6 models justifies the use of CESM2 pacemaker simulations to further quantify the impact of SST trend discrepancy on the reanalysis-coupled model discrepancy found more generally in CMIP6 models. The pacemaker simulations can also be used to test the hypotheses discussed above regarding mechanisms connecting SST trend discrepancies to storminess trends. In the following, we separately investigate the impacts of the Southern Ocean and tropical Pacific SST trend discrepancies on the storminess trend discrepancy in the zonal mean and South Pacific using the pacemaker simulations.

a. Impact of Southern Ocean SST trend discrepancy on storminess trends

When CESM2 simulations are forced with historical forcings and SST anomalies are nudged to observations in the Southern Ocean, there is a significant storminess intensification in all longitudes in the SH (Fig. 8a). In particular, the storminess trend is larger in the Southern Ocean compared to CESM2-LE simulations (Δ_{SO}, Fig. 8b). The zonal-mean storminess trends in the SOPACE simulations (green, Fig. 9a) are significantly larger than those in the CESM2-LE simulations (MW-test p-value = 0.02). According to the average rank (27.8%, Table 2), a discrepancy is unlikely for the SOPACE simulations.

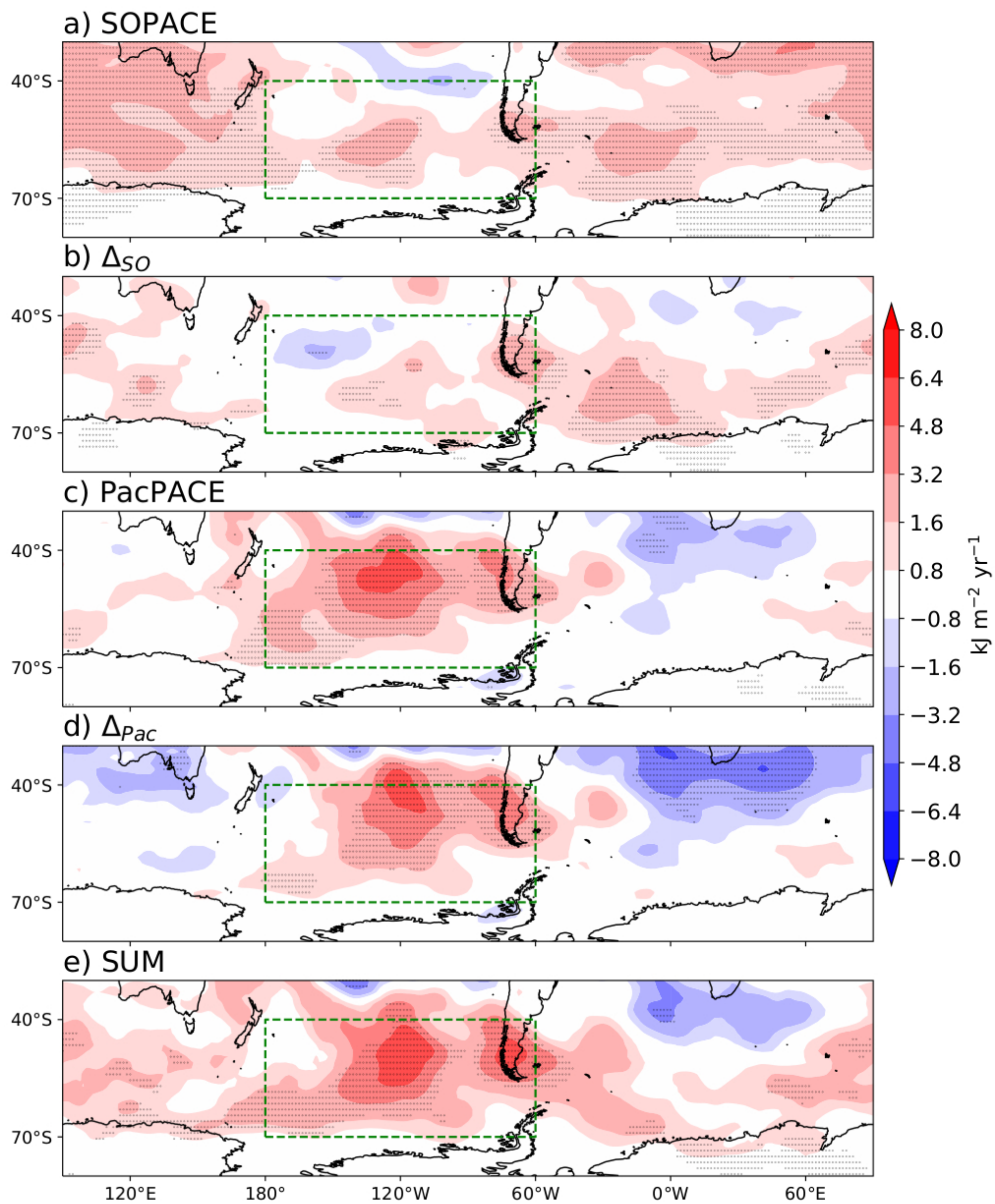

FIG. 8. Spatial pattern of ensemble-mean SH JJA EKE trend during 1979–2013 for (a) SOPACE (b) Δ_{SO}, (c) PacPACE, (d) Δ_{Pac} and (e) SUM simulations. Stipples indicate where ensemble-mean trends are significant at the 95% level.

Shaw et al. (2022) hypothesized that the reanalysis-CMIP6 zonal-mean SH storminess trend discrepancy was due to an underestimated surface energy flux trends in models across the SH, which is connected to Southern Ocean SST trends. The connection between storminess and surface energy flux is made through the moist static energy budget with the atmospheric energy transport implied from surface energy flux. They are related as follows:

$$\nabla \cdot F_{SFC} = S \tag{4}$$

(Kang et al. 2008; Shaw et al. 2018, 2022), where S is the zonal-mean surface energy flux (in W m^{-2}) with the global average removed (defined as positive downward), and $2\pi a \cos\phi F_{SFC}$ (in PW), where a is the Earth's radius, represents the atmospheric energy flux induced by surface energy flux gradient at latitude ϕ. The surface energy flux in ERA5 is obtained by subtracting mass-consistent atmospheric total energy flux divergence and energy tendency from the top-of-atmosphere radiation (Mayer et al. 2021). Note that other reanalyses do not have mass-consistent energy flux datasets available for this calculation. The surface energy flux can be directly obtained in the CESM2-LE and SOPACE simulations. The surface energy flux trend ($2\pi a \cos\phi F_{SFC}$) in the SOPACE simulations is significantly larger than those in the CESM2-LE simulations (MW test p-value = 0.00). In addition, the SOPACE surface energy flux trends are closer to that in ERA5 (Fig. 10). The ERA5 rank is 0.0% in the CESM2-LE and 28.6% in the SOPACE simulations. Thus, the storminess trends are larger in SOPACE consistent with a larger surface energy flux trends that better capture reanalysis trends.

While nudging SST anomalies in the Southern Ocean indicates that reanalysis-model storminess trend discrepancy in the zonal mean is unlikely, a discrepancy in the South Pacific is still likely according to the average rank of 18.3% (Table 2). Moreover, SOPACE and CESM2-LE trends in the South Pacific (Fig. 9b) are not significantly different according to the MW test (p-value = 0.26).

b. Impact of tropical Pacific SST trend discrepancy on storminess trends

When CESM2 simulations are forced with historical forcings and SST anomalies are nudged to observations in the tropical Pacific, there is a significant storminess trend in the South Pacific (180°–60°W, Fig. 8c). Nudging tropical Pacific SST anomalies to observations (Δ_{Pac}) increases

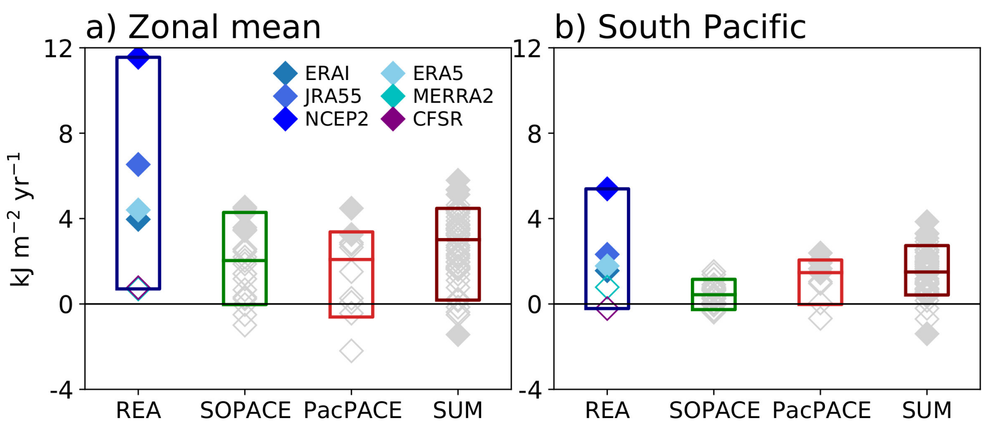

FIG. 9. Same as Fig. 5, but for SOPACE (green), PacPACE (red), and SUM (maroon) simulations in the (a) zonal mean and (b) South Pacific. Note that reanalysis trends are repeated from Fig. 5.

the storminess trend in the South Pacific but weakens it elsewhere (Fig. 8d). The South Pacific storminess trends in PacPACE simulations (red, Fig. 9b) are significantly larger than those in the CESM2-LE simulations (MW-test p-value = 0.01). According to the average rank (41.6%, Table 2), a reanalysis-trend discrepancy in the South Pacific is unlikely for the PacPACE simulations.

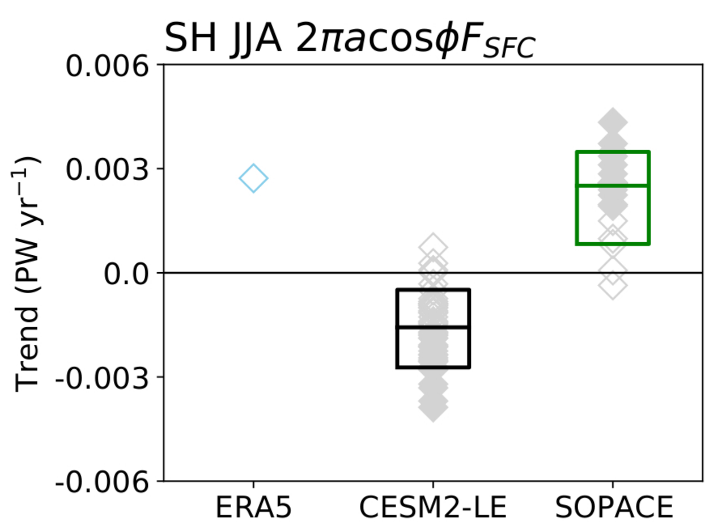

FIG. 10. Linear trends of SH JJA $2\pi a \cos\phi F_{SFC}$ (40–70°S) in ERA5 and CESM2-LE and SOPACE simulations (1979–2013, diamonds). Statistically significant trends at the 95% confidence level are filled. The box represents the 10–90% percentile of model ensemble trends. The horizontal line inside the box shows the median trend in the model ensemble.

We hypothesized the La Nina-like SST trend in the tropical Pacific induces a Rossby wave teleconnection trend to the South Pacific, characterized by weaker subtropical jet and strengthened storminess in the South Pacific consistent with previous work (Seager et al. 2003; Nakamura et al. 2004; Ashok et al. 2007). The 200-hPa zonal wind and eddy geopotential height trends in the PacPACE simulations show a clear La Nina-like teleconnection pattern that is absent in the CESM2-LE simulations (Fig. 11). In particular, the South Pacific subtropical jet trends (averaged over 15°–30°S, 180°–60°W: magenta box in Fig. 11) are significantly different between PacPACE and CESM2-LE simulations (Fig. 12) according to the MW test (p-value= 0.00). According to the average rank (41.7%), a reanalysis-model subtropical jet trend discrepancy is unlikely for the PacPACE simulations. In contrast, for CESM2-LE simulations, the average rank is 98.3%. According to this, a discrepancy is very likely. This confirms that PacPACE simulations exhibit stronger South Pacific storminess by capturing La Nina-like teleconnection trends in reanalysis.

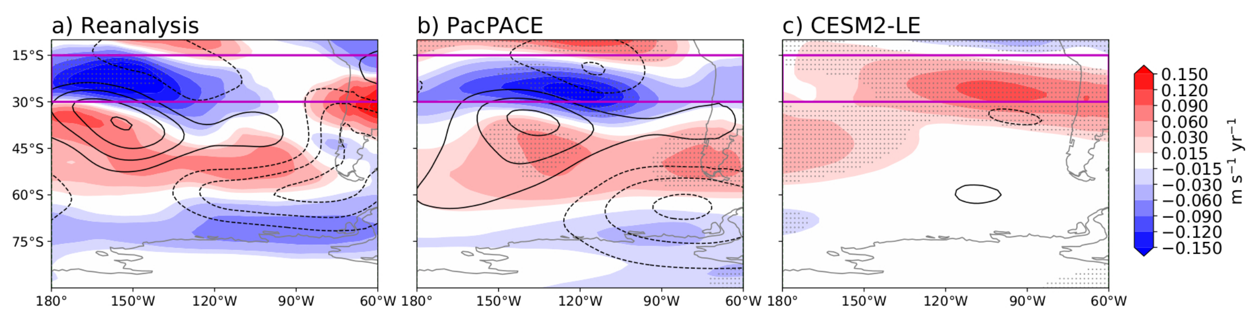

FIG. 11. Spatial pattern of South Pacific JJA 200-hPa zonal wind (shading) and eddy geopotential height (contours, deviation from zonal mean) trends during 1979–2013 for (a) reanalysis mean (CFSR, ERAI, ERA5, JRA55, MERRA2, NCEP2), (b) PacPACE and (c) CESM2-LE ensemble mean. The positive and negative eddy geopotential height trends are respectively depicted in solid and dashed contours in 0.3 m yr^{-1} intervals (zero contour is suppressed). Stipples indicate where reanalysis-mean or ensemble-mean trends are significant at the 95% level. The magenta box (15–30°S, 180–60°W) indicates the domain where the South Pacific subtropical jet is quantified.

c. Combined impact of tropical Pacific and Southern Ocean SST trend discrepancies on storminess trends

The results above suggest that simulating the observed SST trends both in the Southern Ocean and tropical Pacific is necessary to capture the reanalysis SH storminess trend and its spatial structure.

To investigate the combined impact of both pacemakers ($\Delta_{Pac}+\Delta_{SO}$) on the coupled simulations, we create a synthetic large ensemble named SUM with 50 members, which is defined as:

$$\mathrm{SUM} = \text{CESM2-LE} + \Delta_{Pac} + \Delta_{SO} \tag{5}$$

Note that we are adding ensemble-mean impacts ($\Delta_{Pac}+\Delta_{SO}$) to individual ensemble members of CESM2-LE. This synthetic large ensemble is meant to estimate the results for ensemble simulations that nudge the Southern Ocean and tropical Pacific SST simultaneously. It assumes the ensemble-mean impacts of pacemaker simulations are combined with the forced response and internal variability in the CESM2-LE simulations. A similar approach was taken in Kang et al. (2023a) to create synthetic ensemble simulations using SOPACE simulations.

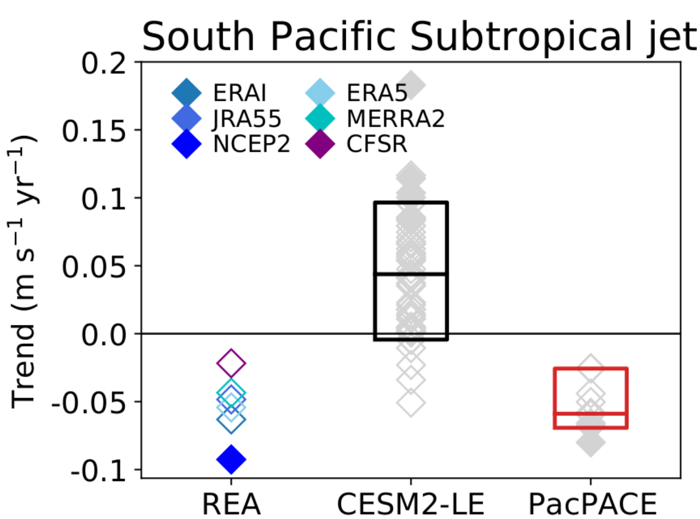

Fig. 12. Linear trends of JJA South Pacific subtropical jet (200-hPa zonal wind averaged over 15–30°S, 180–60°W) in reanalysis datasets and CESM2-LE and PacPACE simulations (1979–2013, diamonds). Statistically significant trends at the 95% confidence level are filled. The box represents the 10–90% percentile of model ensemble trends. The horizontal line inside the box shows the median trend in the model ensemble.

The SUM ensemble provides valuable insights since it captures the observed SST trend in the ensemble mean (compare Figs. 7b and c). This is due to the remote impacts of the pacemaker simulations on SST trends outside the nudged area (see dashed lines in Figs. 7d and e). The SOPACE simulations affect the SST trend in the Southeast Pacific and around Antarctica (Fig. 7d, see also Kang et al. 2023a) while the PacPACE simulations reverse the trend in the tropical Pacific and enhance the warming in the Southwest Pacific (Fig. 7e).

The SUM ensemble shows significant storminess trends across the SH (Fig. 8e), and the trends are larger than individual pacemaker simulations both in the zonal mean and South Pacific (Fig. 9). According to the average rank for both the zonal mean (34.0%, Table 2) and South Pacific (46.0%, Table 2) trends, a reanalysis-model trend discrepancy is unlikely. This confirms that the SST trend discrepancies impact the reanalysis-coupled model storminess trend discrepancy.

5. Summary and Discussion

a. Summary

Our study revisits the reanalysis-model SH winter storminess trend discrepancy and examines the impact of observational uncertainty, model ensemble size, and a like-for-like comparison. We address these aspects by doubling the number of reanalysis datasets to 6, using all available model simulations, and calculating storminess on the same time and spatial grids. The assessment of observational uncertainty reveals substantial spread in reanalysis trends of SH winter storminess. When accounting for model ensemble size and a like-for-like comparison, storminess trends in reanalysis are reduced in amplitude, and a discrepancy is unlikely between reanalyses and prescribed-SST (AMIP) model trends, according to the average rank of reanalysis trends. Even after accounting for observational uncertainty, model ensemble size, and a like-for-like comparison, a discrepancy between reanalyses and coupled (CMIP) models is likely, especially in the South Pacific. The comparison between CMIP and AMIP simulations suggests well-known observation-model SST trend discrepancies across the Southern Ocean and tropical Pacific may impact the storminess trend.

We use Southern Ocean and tropical Pacific pacemaker simulations to test the hypothesis that the reanalysis-coupled model storminess trend discrepancy is connected to the SST trend discrepancies. When the SST anomalies in the Southern Ocean are nudged toward observation, the reanalysis-coupled model storminess trend discrepancy in the zonal mean becomes unlikely, but it is still likely in the South Pacific. Consistent with our hypothesis, the improvement of zonal-mean storminess trends involves simulating surface energy flux trends closer to reanalysis, which are consistent with increased storminess. When the SST anomalies in the tropical Pacific are nudged toward observation, the reanalysis-coupled model storminess trend discrepancy in the South Pacific becomes unlikely. As hypothesized, the improvement of South Pacific storminess trends is a result of capturing trends in teleconnections to the South Pacific induced by a La Nina-like SST trend, consistent with previous work (Seager et al. 2003; Nakamura et al. 2004; Ashok et al. 2007). Thus, the pacemaker simulations show when SST trend discrepancies are removed, the reanalysis-coupled model storminess trend discrepancy becomes unlikely. This confirms the importance of SST trend discrepancies on the reanalysis-coupled model discrepancy in the SH winter storminess trends.

b. Discussion

Our results emphasize that it is important to address observational uncertainty, model ensemble size, and a like-for-like comparison when comparing trends in reanalysis and models. By addressing these aspects, we arrived at a conclusion that a reanalysis-model trend discrepancy is unlikely for AMIP6 models. Since the SH exhibits a significant observation uncertainty (large spread in reanalyses trends), it is important to use all available reanalysis data as in previous work (Manney and Hegglin 2018; Grise et al. 2019; Martineau et al. 2024). The large spread in reanalyses trends, which is comparable to that in the large ensemble simulations, also poses a challenge for reanalysis-model comparison in the SH. Efforts that can try to rule out or verify the fidelity of individual reanalysis trends could be beneficial for similar future work.

The pacemaker simulations show that when the SST trend discrepancy is removed the reanalysis-coupled model trend discrepancy becomes unlikely. It is thus important to understand the SST trend discrepancy and its underlying mechanisms. For the tropical Pacific, many mechanisms have been proposed (Lee et al. 2022; Seager et al. 2022). For the Southern Ocean, one proposed mechanism involving Antarctic meltwater which does not flux to the Southern Ocean in the coupled models seems to be important (Bronselaer et al. 2018; Roach et al. 2023). Understanding the mechanisms underlying the emergent responses is important for having confidence in climate model projections and future work should test the proposed mechanisms (Shaw 2019).

Model resolution is another factor that can impact the fidelity of the climate model trends. In particular, recent work shows that there is an improvement in the high-resolution (0.25° in atmosphere and 0.1° in ocean) CESM1 simulations in simulating observed SST trends in the tropical Pacific and the Southern Ocean (Yeager et al. 2023, DiNezio et al., personal communication). An examination of the SH storminess trends in this three-member high-resolution simulations shows they underestimate the reanalysis trends and simulate trends similar to low-resolution (1° in both atmosphere and ocean) CESM1 simulations (Fig. A1). This may suggest that the improvement of SST trends in the high-resolution simulations is not sufficient for capturing observed SH storminess trends. However, the ensemble size of the high-resolution simulations is small, and thus future work should further investigate the impact of model resolution on reanalysis-model SH storminess trend discrepancy.

Acknowledgments. JMK and TAS are supported by the National Oceanic and Atmospheric Administration award NA23OAR4310597. IRS is supported by the National Center of Atmospheric Research, which is a major facility sponsored by the National Science Foundation under the Cooperative Agreement 1852977. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We also acknowledge the Community Earth System Model, version 1, Large Ensemble Community Project; the Community Earth System Model, version 2, Large Ensemble Community Project; and supercomputing resources provided by the IBS Center for Climate Physics in South Korea (`https://doi.org/10.5194/esd-2021-5`). We acknowledge the Climate Variability and Change Working Group at the National Center of Atmospheric Research for running the Pacific pacemaker simulations and making them available. The Southern Ocean pacemaker simulations are supported by the high-performance computing cluster of State Key Laboratory of Satellite Ocean Environment Dynamics. This research is completed through the International Laboratory for High Resolution Earth System Prediction—a collaboration among the Qingdao National Laboratory for Marine Science and Technology, Texas A&M University, and the U.S. National Center for Atmospheric Research.

Data availability statement. The CFSR reanalysis data are available at `https://www.ncei.noaa.gov/data/climate-forecast-system/access/reanalysis/` and `https://www.ncei.noaa.gov/data/climate-forecast-system/access/operational-analysis/`. The ERA5 reanalysis data are available at `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form`. ERA-Interim reanalysis data are available at `https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-interim`. JRA-55 reanalysis data can be downloaded from `https://rda.ucar.edu/datasets/ds628.0/`. MERRA-2 reanalysis data can be downloaded from `https://disc.gsfc.nasa.gov/datasets?project=MERRA-2`. The NCEP2 reanalysis data is obtained from `https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html`. The CMIP6 and AMIP6 model data are downloadable from the CMIP6 data search interface `https://esgf-node.llnl.gov/search/cmip6/`. The CESM2-LE simulations are accessible online at https://www.cesm.ucar.edu/community-projects/lens2. The CESM1-LE simulations are available at https://www.cesm.ucar.edu/community-projects/lens. The GOGA and PacPACE simulations are available at https://www.cesm.ucar.edu/working-groups/climate. The SOPACE simulation data are archived at https://github.com/yuyuyaoyao/CESM2_SOPACE. The IHESP simulations are obtained from https://ihesp.github.io/archive/products/ds_archive/Datasets.html#global-datasets.

APPENDIX A

Impact of model resolution on the storminess trends

To evaluate the impact of model resolution on the reanalysis-coupled model storminess trend discrepancy, we use the high-resolution CESM version 1 simulations from the International Laboratory for High-Resolution Earth System Prediction (Chang et al. 2020, hereafter called IHESP simulations). These simulations are compared with the CESM version 1 Large Ensemble simulations (Kay et al. 2015, hereafter called CESM1-LE) with lower resolutions. The IHESP simulations have nominal 0.25^{o} and 0.1^{o} resolution in the atmosphere and ocean, respectively. The CESM1-LE simulations, in contrast, have a nominal 1^{o} resolution in both atmosphere and ocean. The IHESP and CESM1-LE simulations have 3 and 40 ensemble members, respectively. We analyze the time period from 1979 to 2013 in both simulations, during which is forced by historical (1979–2005) and RCP8.5 (2006-2013) forcing following CMIP5 protocol. We quantify the storminess trends in the CESM1-LE and IHESP simulations in the same way as CESM2 simulations in section 4 using Eq. (3). These trends are compared with reanalysis trends shown in Fig. 5.

The CESM1-LE simulations, which feature observation-model SST trend discrepancy in the tropical Pacific and the Southern Ocean (Wills et al. 2022), underestimate the storminess trends in the reanalysis similar to the CESM2-LE simulations (compare Fig. 5 and Fig. A1). The average rank is 11.2% in the CESM1-LE simulations suggesting that a discrepancy is likely.

The three members of IHESP simulations also underestimate the reanalysis storminess trend, although they simulate SST trends closer to observations (DiNezio et al., personal communication). Only one member has a trend (1.68 kJ m^{-2} yr^{-1}, Fig. A1) larger than the smallest reanalysis trend (MERRA2, 0.71 kJ m^{-2} yr^{-1}, Fig. 5a). Moreover, the trends in IHESP simulations are not statistically different from CESM1-LE trends (MW test p-value$= 0.84$).

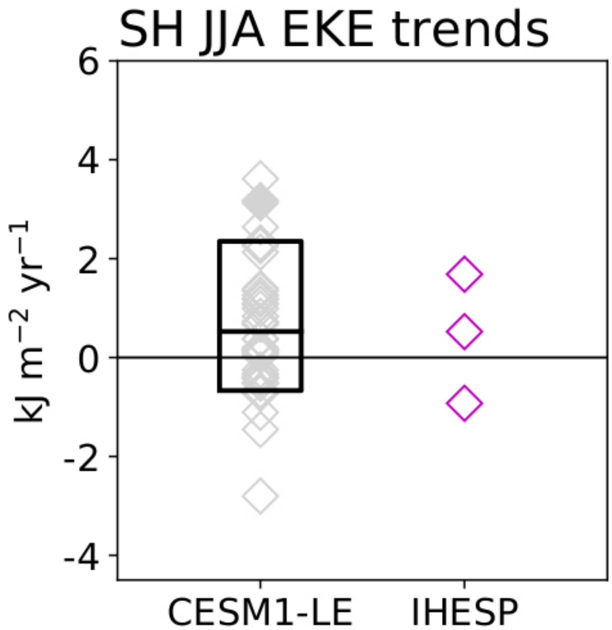

Fig. A1. Linear trends of zonal-mean SH JJA EKE (40–70°S) in CESM1-LE and IHESP simulations (1979–2013, diamonds). Statistically significant trends at the 95% confidence level are filled. The box represents 10–90% percentile of CESM1-LE simulation trends. The horizontal line inside the box shows the median trend in the model ensemble.

References

Armour, K. C., J. Marshall, J. R. Scott, A. Donohoe, and E. R. Newsom, 2016: Southern Ocean warming delayed by circumpolar upwelling and equatorward transport. *Nature Geoscience*, **9 (7)**, 549–554.

Ashok, K., H. Nakamura, and T. Yamagata, 2007: Impacts of ENSO and Indian Ocean dipole events on the Southern Hemisphere storm-track activity during austral winter. *Journal of Climate*, **20 (13)**, 3147–3163.

Bengtsson, L., S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *Journal of Geophysical Research: Atmospheres*, **109 (D11)**.

Bronselaer, B., M. Winton, S. M. Griffies, W. J. Hurlin, K. B. Rodgers, O. V. Sergienko, R. J. Stouffer, and J. L. Russell, 2018: Change in future climate due to Antarctic meltwater. *Nature*, **564 (7734)**, 53–58.

Chang, E. K., 2017: Projected significant increase in the number of extreme extratropical cyclones in the Southern Hemisphere. *Journal of Climate*, **30 (13)**, 4915–4935.

Chang, E. K., Y. Guo, and X. Xia, 2012: CMIP5 multimodel ensemble projection of storm track change under global warming. *Journal of Geophysical Research: Atmospheres*, **117 (D23)**.

Chang, E. K., S. Lee, and K. L. Swanson, 2002: Storm track dynamics. *Journal of climate*, **15 (16)**, 2163–2183.

Chang, P., and Coauthors, 2020: An unprecedented set of high-resolution earth system simulations for understanding multiscale interactions in climate variability and change. *Journal of Advances in Modeling Earth Systems*, **12 (12)**, e2020MS002 298.

Chemke, R., Y. Ming, and J. Yuval, 2022: The intensification of winter mid-latitude storm tracks in the Southern Hemisphere. *Nature climate change*, **12 (6)**, 553–557.

Cox, T., A. Donohoe, K. C. Armour, D. M. Frierson, and G. H. Roe, 2024: Trends in atmospheric heat transport since 1980. *Journal of Climate*, **37 (5)**, 1539–1550.

Danabasoglu, G., and Coauthors, 2020: The community earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, **12 (2)**, e2019MS001 916.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, **137 (656)**, 553–597.

Deser, C., A. S. Phillips, and M. A. Alexander, 2010: Twentieth century tropical sea surface temperature trends revisited. *Geophysical Research Letters*, **37 (10)**.

Deser, C., and Coauthors, 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, **10 (4)**, 277–286.

Dong, B., R. T. Sutton, L. Shaffrey, and B. Harvey, 2022a: Recent decadal weakening of the summer Eurasian westerly jet attributable to anthropogenic aerosol emissions. *Nature communications*, **13 (1)**, 1–10.

Dong, Y., K. C. Armour, D. S. Battisti, and E. Blanchard-Wrigglesworth, 2022b: Two-way teleconnections between the Southern Ocean and the tropical Pacific via a dynamic feedback. *Journal of Climate*, **35 (19)**, 6267–6282.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, **9 (5)**, 1937–1958.

Fujiwara, M., and Coauthors, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmospheric Chemistry and Physics*, **17 (2)**, 1417–1452.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *Journal of Climate*, **30 (14)**, 5419–5454.

Gillett, N. P., and Coauthors, 2016: The detection and attribution model intercomparison project (DAMIP v1. 0) contribution to CMIP6. *Geoscientific Model Development*, **9 (10)**, 3685–3697.

Grise, K. M., and Coauthors, 2019: Recent tropical expansion: Natural variability or forced response? *Journal of Climate*, **32 (5)**, 1551–1571.

Guo, Y., and E. K. Chang, 2008: Impacts of assimilation of satellite and rawinsonde observations on Southern Hemisphere baroclinic wave activity in the NCEP–NCAR reanalysis. *Journal of climate*, **21 (13)**, 3290–3309.

Guo, Y., E. K. Chang, and S. S. Leroy, 2009: How strong are the Southern Hemisphere storm tracks? *Geophysical research letters*, **36 (22)**.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, **129 (3)**, 550–560.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049.

Hoskins, B. J., and K. I. Hodges, 2005: A new perspective on Southern Hemisphere storm tracks. *Journal of Climate*, **18 (20)**, 4108–4129.

Huang, B., and Coauthors, 2017: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *Journal of Climate*, **30 (20)**, 8179–8205.

Jain, S., A. A. Scaife, T. G. Shepherd, C. Deser, N. Dunstone, G. A. Schmidt, K. E. Trenberth, and T. Turkington, 2023: Importance of internal variability for climate model assessment. *npj Climate and Atmospheric Science*, **6 (1)**, 68.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. Hnilo, M. Fiorino, and G. Potter, 2002: Ncep–doe amip-ii reanalysis (r-2). *Bulletin of the American Meteorological Society*, **83 (11)**, 1631–1644.

Kang, J. M., T. A. Shaw, and L. Sun, 2023b: Arctic sea ice loss weakens Northern Hemisphere summertime storminess but not until the late 21st century. *Geophysical Research Letters*, **50 (9)**, e2022GL102 301.

Kang, S. M., I. M. Held, D. M. Frierson, and M. Zhao, 2008: The response of the ITCZ to extratropical thermal forcing: Idealized slab-ocean experiments with a GCM. *Journal of Climate*, **21 (14)**, 3521–3532.

Kang, S. M., Y. Yu, C. Deser, X. Zhang, I.-S. Kang, S.-S. Lee, K. B. Rodgers, and P. Ceppi, 2023a: Global impacts of recent Southern Ocean cooling. *Proceedings of the National Academy of Sciences*, **120 (30)**, e2300881 120.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, **96 (8)**, 1333–1349.

Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, **93 (1)**, 5–48.

Lee, S., M. L'Heureux, A. T. Wittenberg, R. Seager, P. A. O'Gorman, and N. C. Johnson, 2022: On the future zonal contrasts of equatorial Pacific climate: Perspectives from Observations, Simulations, and Theories. *npj Climate and Atmospheric Science*, **5 (1)**, 82.

Mann, H. B., and D. R. Whitney, 1947: On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50–60.

Manney, G. L., and M. I. Hegglin, 2018: Seasonal and regional variations of long-term changes in upper-tropospheric jets from reanalyses. *Journal of Climate*, **31 (1)**, 423–448.

Martineau, P., S. K. Behera, M. Nonaka, H. Nakamura, and Y. Kosaka, 2024: Seasonally dependent increases in subweekly temperature variability over Southern Hemisphere landmasses detected in multiple reanalyses. *Weather and Climate Dynamics*, **5 (1)**, 1–15.

Mastrandrea, M. D., and Coauthors, 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties.

Mayer, J., M. Mayer, and L. Haimberger, 2021: Consistency and homogeneity of atmospheric energy, moisture, and mass budgets in ERA5. *Journal of Climate*, **34 (10)**, 3955–3974.

Nakamura, H., T. Sampe, Y. Tanimoto, and A. Shimpo, 2004: Observed associations among storm tracks, jet streams and midlatitude oceanic fronts. *Earth's Climate: The Ocean–Atmosphere Interaction, Geophys. Monogr*, **147**, 329–345.

O'Gorman, P. A., 2010: Understanding the varied response of the extratropical storm tracks to climate change. *Proceedings of the National Academy of Sciences*, **107 (45)**, 19 176–19 180.

Pepler, A., 2020: Record lack of cyclones in southern Australia during 2019. *Geophysical Research Letters*, **47 (13)**, e2020GL088 488.

Pfahl, S., and H. Wernli, 2012: Quantifying the relevance of cyclones for precipitation extremes. *Journal of Climate*, **25 (19)**, 6770–6780.

Po-Chedley, S., T. J. Thorsen, and Q. Fu, 2015: Removing diurnal cycle contamination in satellite-derived tropospheric temperatures: Understanding tropical tropospheric trend discrepancies. *Journal of Climate*, **28 (6)**, 2274–2290.

Purich, A., W. Cai, M. H. England, and T. Cowan, 2016: Evidence for link between modelled trends in Antarctic sea ice and underestimated westerly wind changes. *Nature communications*, **7 (1)**, 10 409.

Roach, L. A., K. D. Mankoff, A. Romanou, E. Blanchard-Wrigglesworth, T. W. Haine, and G. A. Schmidt, 2023: Winds and meltwater together lead to Southern Ocean surface cooling and sea ice expansion. *Geophysical Research Letters*, **50 (24)**, e2023GL105 948.

Rodgers, K. B., and Coauthors, 2021: Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*, **12 (4)**, 1393–1411.

Saha, S., and Coauthors, 2010: The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, **91 (8)**, 1015–1058.

Saha, S., and Coauthors, 2014: The NCEP climate forecast system version 2. *Journal of climate*, **27 (6)**, 2185–2208.

Santer, B. D., and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **28 (13)**, 1703–1722.

Santer, B. D., and Coauthors, 2017: Comparing tropospheric warming in climate models and satellite data. *Journal of Climate*, **30 (1)**, 373–392.

Schmidt, G., 2013: On mismatches between models and observations. Accessed: 04 September 2023, https://www.realclimate.org/index.php/archives/2013/09/on-mismatches-between-models-and-observations/.

Seager, R., N. Harnik, Y. Kushnir, W. Robinson, and J. Miller, 2003: Mechanisms of hemispherically symmetric climate variability. *Journal of Climate*, **16 (18)**, 2960–2978.

Seager, R., N. Henderson, and M. Cane, 2022: Persistent discrepancies between observed and modeled trends in the tropical Pacific Ocean. *Journal of Climate*, **35 (14)**, 4571–4584.

Shaw, T., and Coauthors, 2016: Storm track processes and the opposing influences of climate change. *Nature Geoscience*, **9 (9)**, 656–664.

Shaw, T. A., 2019: Mechanisms of future predicted changes in the zonal mean mid-latitude circulation. *Current Climate Change Reports*, **5 (4)**, 345–357.

Shaw, T. A., P. Barpanda, and A. Donohoe, 2018: A moist static energy framework for zonal-mean storm-track intensity. *Journal of the Atmospheric Sciences*, **75 (6)**, 1979–1994.

Shaw, T. A., O. Miyawaki, and A. Donohoe, 2022: Stormier Southern Hemisphere induced by topography and ocean circulation. *Proceedings of the National Academy of Sciences*, **119 (50)**, e2123512 119.

Suarez-Gutierrez, L., S. Milinski, and N. Maher, 2021: Exploiting large ensembles for a better yet simpler climate model evaluation. *Climate Dynamics*, **57 (9-10)**, 2557–2580.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bulletin of the American meteorological Society*, **93 (4)**, 485–498.

Wills, R. C., Y. Dong, C. Proistosecu, K. C. Armour, and D. S. Battisti, 2022: Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophysical Research Letters*, **49 (17)**, e2022GL100 011.

Yeager, S. G., and Coauthors, 2023: Reduced Southern Ocean warming enhances global skill and signal-to-noise in an eddy-resolving decadal prediction system. *npj Climate and Atmospheric Science*, **6 (1)**, 107.

Yettella, V., and J. E. Kay, 2017: How will precipitation change in extratropical cyclones as the planet warms? Insights from a large initial condition climate model ensemble. *Climate Dynamics*, **49 (5-6)**, 1765–1781.