npj | climate and atmospheric science

www.nature.com/npjclimatsci

ARTICLE OPEN

Check for updates

Reduced Southern Ocean warming enhances global skill and signal-to-noise in an eddy-resolving decadal prediction system

Stephen G. Yeager[1]✉, Ping Chang[2], Gokhan Danabasoglu[1], Nan Rosenbloom[1], Qiuying Zhang[2], Fred S. Castruccio[1], Abishek Gopal[2], M. Cameron Rencurrel[2] and Isla R. Simpson[1]

The impact of increased model horizontal resolution on climate prediction performance is examined by comparing results from low-resolution (LR) and high-resolution (HR) decadal prediction simulations conducted with the Community Earth System Model (CESM). There is general improvement in global skill and signal-to-noise characteristics, with particularly noteworthy improvements in the eastern tropical Pacific, when resolution is increased from order 1° in all components to order 0.1°/0.25° in the ocean/atmosphere. A key advance in the ocean eddy-resolving HR system is the reduction of unrealistic warming in the Southern Ocean (SO) which we hypothesize has global ramifications through its impacts on tropical Pacific multidecadal variability. The results suggest that accurate representation of SO processes is critical for improving decadal climate predictions globally and for addressing longstanding issues with coupled climate model simulations of recent Earth system change.

npj Climate and Atmospheric Science (2023)6:107 ; https://doi.org/10.1038/s41612-023-00434-y

INTRODUCTION

The capacity to predict seasonal to decadal Earth system variability and change has grown rapidly in recent years using initialized coupled climate models that incorporate slowly evolving components such as the ocean, land, and sea ice[1,2]. Initialization from observation-based states improves forecast skill even for decadal outlooks that are strongly influenced by externally forced warming trends[3]. Large ensembles (≥20 members) have revealed the promising potential to predict multiyear to decadal variations in regional surface air temperature (SAT), precipitation (PRE), and atmospheric circulation[4–9]. While decadal climate prediction science has advanced to the point that operational decadal forecasts are now feasible[10], model fidelity remains a key obstacle hindering progress and limiting the utility of real-time forecasts[2]. Reliable operational decadal predictions are greatly needed to inform actionable planning and mitigation measures to avoid the worst impacts of near-term regional climate change.

Initialized prediction permits a more direct evaluation of coupled climate model realism than is possible using traditional ("uninitialized") historical simulations whose internal variations are not synchronized with observations. The co-occurrence of low ensemble mean (signal) variance in seasonal to decadal climate prediction systems with a high correlation between that signal and observations suggests that there are fundamental deficiencies in simulating the predictable component of Earth system variability in the current generation of coupled climate models[3–5,7,11–13]. The explanation of this signal-to-noise paradox, wherein the model ensemble mean correlation with observations is higher than that with individual ensemble members, remains an open research question, but it appears to be intrinsic to model structure rather than initialization[14]. The use of large ensembles allows weak but skillful signals to be extracted from noisy systems, circumventing this issue to some extent but leaving the fundamental problem unresolved. This is a critical problem to address given the role of these models in uninitialized climate projections, where increasing ensemble size does not solve the problem. It has been suggested that weak signals could be related to deficient model response to external forcings such as volcanic eruptions[15], an idea bolstered by recent work showing that the tropical Pacific response to historical volcanic aerosol forcing may be flawed in some decadal prediction systems[16]. A leading hypothesis is that low signal-to-noise in prediction systems (and by extension, in the underlying models) is fundamentally related to their coarse spatial resolution[17–20], but tests of this hypothesis in an initialized prediction framework have been lacking due to the enormous computational expense of running large sets of ensemble climate hindcast simulations as well as technical issues associated with observation-based initialization of high-resolution models.

Most simulations contributed to the Coupled Model Intercomparison Project (CMIP) use component models at roughly 1° (~100 km) horizontal grid resolution, and this is true of all recent contributions to the Decadal Climate Prediction Project[21] (DCPP) of CMIP phase 6. There is mounting evidence that finer model resolution (atmosphere, ocean, or both) reduces mean biases and improves the fidelity of a wide variety of processes relevant to global climate prediction[22–29]. Specifically, the use of coupled models with eddy-resolving ocean components that permit explicit representation of ocean mesoscale turbulence and associated small-scale air-sea interactions is a technically challenging but promising growth area that could deliver more accurate projections of future climate change[26,29,30]. High-resolution atmospheric model simulations (0.5° or finer in the atmosphere) forced with and without ocean mesoscale features show an upscaling impact of ocean eddies and fronts on large-scale atmospheric circulations above the planetary boundary layer[31–34]. Stronger air-sea coupling at the ocean mesoscale has been shown to yield higher signal-to-noise for atmospheric fields[18], and recent studies of decadal potential predictability using high-resolution coupled models have lent support to the hypothesis that explicit representation of ocean eddies and associated atmospheric impacts can enhance climate prediction skill and help to resolve the signal-to-noise paradox[17,20].

[1]National Center for Atmospheric Research, Boulder, CO, USA. [2]Department of Oceanography, Texas A&M University, College Station, TX, USA. ✉email: yeager@ucar.edu

In this study, we explore the benefits of substantially increased horizontal grid resolution in the context of initialized climate prediction by comparing two decadal prediction systems using the Community Earth System Model (CESM): (1) the CESM1.1 Decadal Prediction Large Ensemble (DPLE[5]) that was submitted to the CMIP6 DCPP, and (2) the CESM1.3 High Resolution Decadal Prediction (HRDP) system. The CESM1.3 model in HR configuration (0.25° atmosphere and land coupled to 0.1° ocean and sea ice) exhibits generally improved realism compared to its LR (1° in all components) counterpart[23,26], including reduced biases in ocean deep convection[27] and sea surface temperature (SST)[28], more realistic air-sea coupling[35], and improved eastern boundary upwelling[29]. This decadal prediction comparison builds on these recent resolution sensitivity studies using CESM and reveals significantly enhanced skill for a variety of fields of interest together with improved signal-to-noise characteristics. To first order, performance improvements in HRDP appear to be related to more realistic SO variability and associated improvements in multidecadal trends in the tropical Pacific. This has potential implications for the ongoing debate regarding the mismatch between historical and modeled trends in the tropical Pacific.

RESULTS

Skill for surface fields

A key difference between the two prediction systems is the horizontal resolution of the ocean and atmosphere component models (Table 1 and Methods). While similar techniques are used to generate historical ocean and sea ice initial conditions, HRDP is initialized from a 0.1° ocean and sea ice state reconstruction that is generally more realistic than the 1° reconstruction used in DPLE[24]. It is therefore likely that initial condition quality contributes to skill differences (see discussion below), although we cannot quantify this effect and leave it for future work. The more realistic atmosphere and land initialization in HRDP is unlikely to be a factor contributing to the pentadal skill differences of interest here because neither the atmosphere nor the land is expected to contribute to long timescale memory. Other model differences (atmosphere dynamical core, CESM version, miscellaneous tuning and parameter differences) are likewise believed to have much less impact on the results than differences in model resolution and related model physics representations. The comparison below therefore represents a holistic assessment of prediction system sensitivity to resolution, where the system encompasses both initial condition generation and coupled ensemble hindcasts.

Table 1. Decadal prediction systems.

	HRDP	DPLE
Model	CESM1.3	CESM1.1
Ocean	POP2 (0.1°, 62L)	POP2 (1°, 60L)
Atmosphere	CAM5-SE (0.25°, 30L)	CAM5-FV (1°, 30L)
Land	CLM4 (0.25°)	CLM4 (1°)
Sea ice	CICE4 (0.1°)	CICE4 (1°)
Forcing	CMIP5	CMIP5
Scenario	RCP8.5	RCP8.5
Initialization	Full field	Full field
Ocean	FOSI (0.1°, OMIP2)	FOSI (1°, OMIP1)
Atmosphere	JRA55 reanalysis	N/A
Land	HighResMIP Tier 1	N/A
Sea ice	FOSI (0.1°, OMIP2)	FOSI (1°, OMIP1)
Hindcasts	$N = 21$	$N = 64$
Start date	November 1	November 1
Start year	1976, 1978, ..,2016	1954–2017
Simulation length	62 months	122 months
Ensemble size	10	40
Total simulation years	~1100	~26,000
Normalized computation cost	~100	1

Subsampling DPLE to match the hindcast set and ensemble size available from HRDP permits a direct comparison of system skill. Anomaly correlation coefficient (ACC) maps for forecast years 1–5 (FY1–5) annual mean SAT, PRE, and sea level pressure (SLP) reveal similar overall skill patterns in the two systems but also notable improvements in HRDP (Fig. 1). Previous work has established that external forcing contributes substantially to decadal skill for these fields[3,5], but the primary focus here is not on the source of skill but rather the difference in skill attributable to system resolution. HRDP shows significantly higher SAT skill than DPLE (Fig. 1c) in the subtropical eastern Pacific (SEP) and throughout much of the Southern Ocean (SO). Surface temperature variability in the eastern Pacific sector of the SO (EPSO) stands out as much better represented in HRDP—a critical improvement that will be discussed further below. Some SAT skill degradation in HRDP is seen in some regions, most notably in the Agulhas retroflection region south of Cape of Good Hope—a region of strong ocean eddy activity where the high-resolution ocean might have been expected to deliver improved realism. A quantitative comparison of significant ACC differences between 80°S and 80°N confirms the visual impression that skill for annual SAT is overall better in HRDP than DPLE (Supplementary Table 1).

PRE skill is also generally higher in HRDP, although there are many scattered regions of skill degradation as well (Fig. 1f). Significantly increased ACC for PRE in the eastern tropical Pacific is likely related to the increased tropical SAT skill, and it suggests that improved large-scale teleconnections associated with tropical Pacific convective heating anomalies might explain the widespread skill enhancements seen in both PRE (Fig. 1f) and SLP (Fig. 1i). A promising result that merits a focused study in the future is the improved PRE skill over North America, especially along the US west coast. Skill in this region is negligible in DPLE (even when ensemble size is increased; see Supplementary Fig. 1), whereas the high skill in HRDP implies good potential to forewarn this vulnerable region of impending hydroclimate variability or change. This improvement is likely related to the much-improved tropical Pacific skill given the established links between western US hydroclimate and tropical Pacific variability[36,37], but it could also be associated with improved representation of Kuroshio eddy influence on the Pacific storm track and atmospheric rivers[31,32,34]. Both systems exhibit high (but roughly equivalent) skill for PRE in the Sahel region of Africa, which is a common finding in decadal predictions that appears to be at least partly related to initialization[3,5]. In addition to parts of North and South America (in particular, Chile), regionally enhanced PRE skill over land is evident in a band stretching from central North Africa through the Middle East into southeast Asia, as well as over eastern Eurasia, the Maritime Continent, and the Caribbean. There are also many regions of degraded PRE skill in HRDP, but the percentage area of significant skill increase (25%) is greater than that of significant skill decrease (17%), and furthermore, skill increases tend to be larger in magnitude than skill decreases (Supplementary Table 1).

Large areas of skill improvement are seen for annual SLP (Fig. 1i), corresponding to broad regions where HRDP shows high ACC (exceeding 0.6) in the tropical eastern Pacific, the extratropics of all ocean basins, and the Norwegian Sea (Fig. 1g). Skill is also significantly higher in HRDP over the Indian Ocean and surrounding land masses. Both systems show low SLP skill in the Atlantic sector, and here HRDP shows degradation compared to DPLE in the subtropical North Atlantic and over the Euro-Mediterranean region. However, improvements at high and midlatitudes in the North Atlantic suggests that skill for the North Atlantic Oscillation (NAO) is higher in HRDP. This is confirmed by an examination of annual and winter NAO index timeseries that show higher ACC for

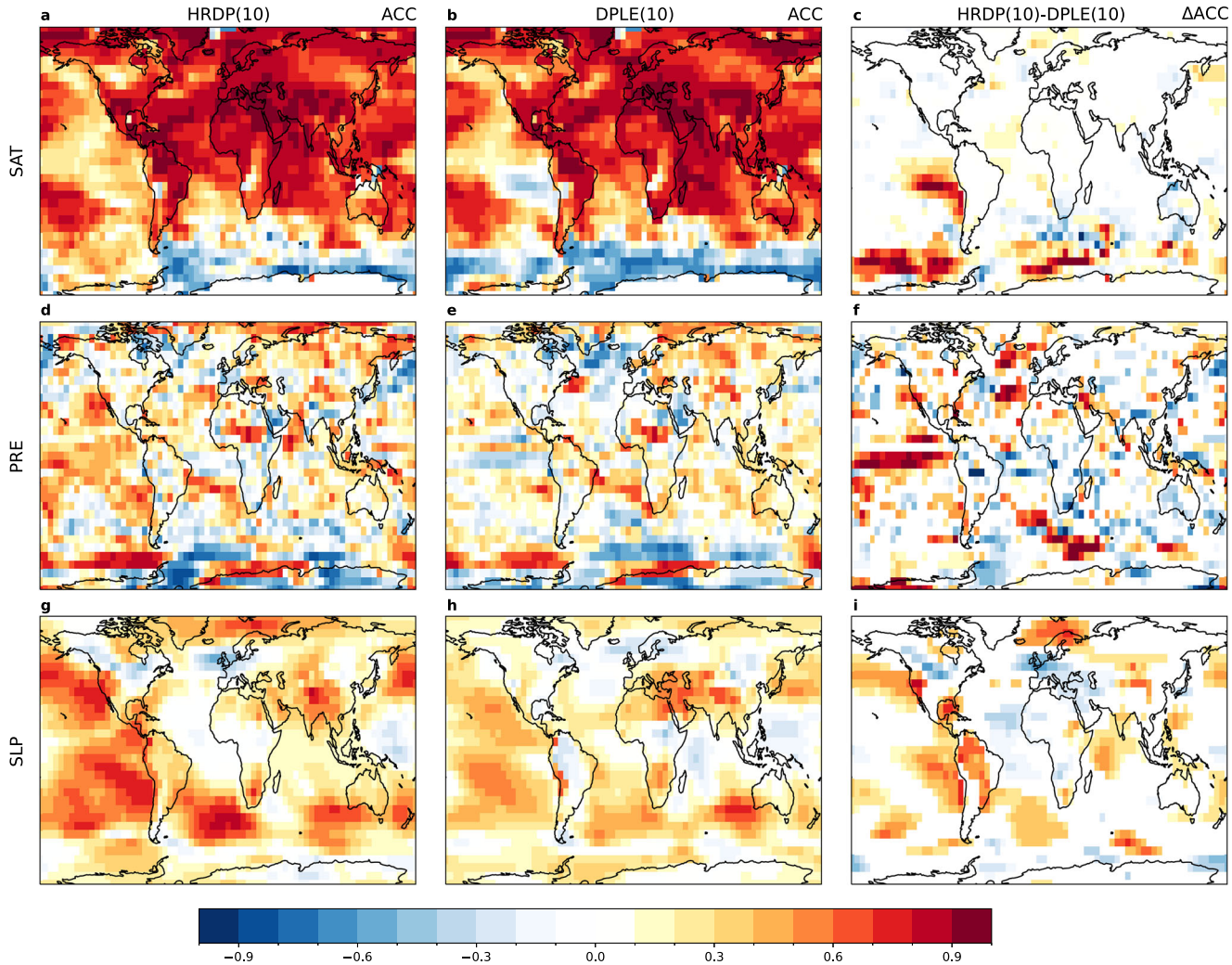

Fig. 1 Correlation comparison for annual fields. ACC skill for 10-member HRDP (**a**, **d**, **g**), 10-member DPLE (**b**, **e**, **h**), and their difference (**c**, **f**, **i**) for years 1–5 forecasts of annual surface air temperature (SAT; top), precipitation (PRE; middle), and sea level pressure (SLP; bottom). Difference values in (**c**, **f**, **i**) are non-zero only where HRDP skill is significantly higher or lower than DPLE skill at the 90% confidence level (see Methods). See Supplementary Table 1 for a quantitative summary of skill differences plotted here.

HRDP than DPLE ($r = 0.28$ compared to $r = -0.17$ for 5-year-averaged annual NAO; see Supplementary Fig. 13).

Mean square skill score (MSSS) maps for annual fields show that the overall ACC skill enhancement in HRDP is accompanied by a general improvement in the magnitude of predicted anomalies (Fig. 2). There is a close correspondence in the skill difference patterns from Figs. 1 and 2 (compare panels c, f, i), but MSSS reveals more widespread improvements in SAT. Significantly enhanced MSSS for PRE and SLP indicates that pentadal atmospheric signals in HRDP, attributable to external forcing and/or ocean SST forcing, are considerably stronger in that system. Increasing the ensemble size from 10 to 40 in DPLE tends to increase ACC magnitude while maintaining the general patterns of positive/negative skill (Supplementary Fig. 1), and MSSS shows a more uniform increase for fields like PRE and SLP (Supplementary Fig. 2). However, the skill differences between 10-member and 40-member DPLE are generally not significant at the grid scale (in contrast with large-scale metrics where significant benefits of a larger ensemble size have been identified[5,6]). While HRDP is much more expensive than quadrupling the DPLE ensemble size (factor of 100 versus 4; Table 1), it yields qualitatively different and overall better skill results that help to shed light on prediction system process representation.

Attribution of HRDP prediction skill to contributions from external forcing and initialization using conventional techniques[3,5] is not possible due to the lack of a large ensemble of uninitialized historical simulations using HR CESM1.3. However, skill for detrended fields offers a rough indication of where initialized internal variability may be an important factor, and furthermore, linear detrending is a method that can be applied consistently to both systems (Fig. 3). The North Atlantic stands out as a region of high detrended SAT skill, with both systems exhibiting a horseshoe-like pattern of elevated ACC reminiscent of the SST loading pattern of Atlantic multidecadal variability[38]. HRDP skill is higher throughout the tropical Atlantic but particularly so in the northern tropical Atlantic (Fig. 3c), a region that has historically shown low skill improvement due to initialization in decadal prediction systems[39]. There are also related SAT skill increases over land in HRDP over North Africa and southern Europe. On the other hand, HRDP shows lower skill in the subpolar North Atlantic (SPNA), which is often highlighted as a key region benefitting from initialization[3,5,40]. Lower SPNA skill in HRDP implies that this region does not account for the higher northern tropical Atlantic skill in HRDP, as might have been expected from previous work highlighting an extratropical-tropical connection in the North Atlantic[41]. The reasons for skill degradation in the SPNA remain

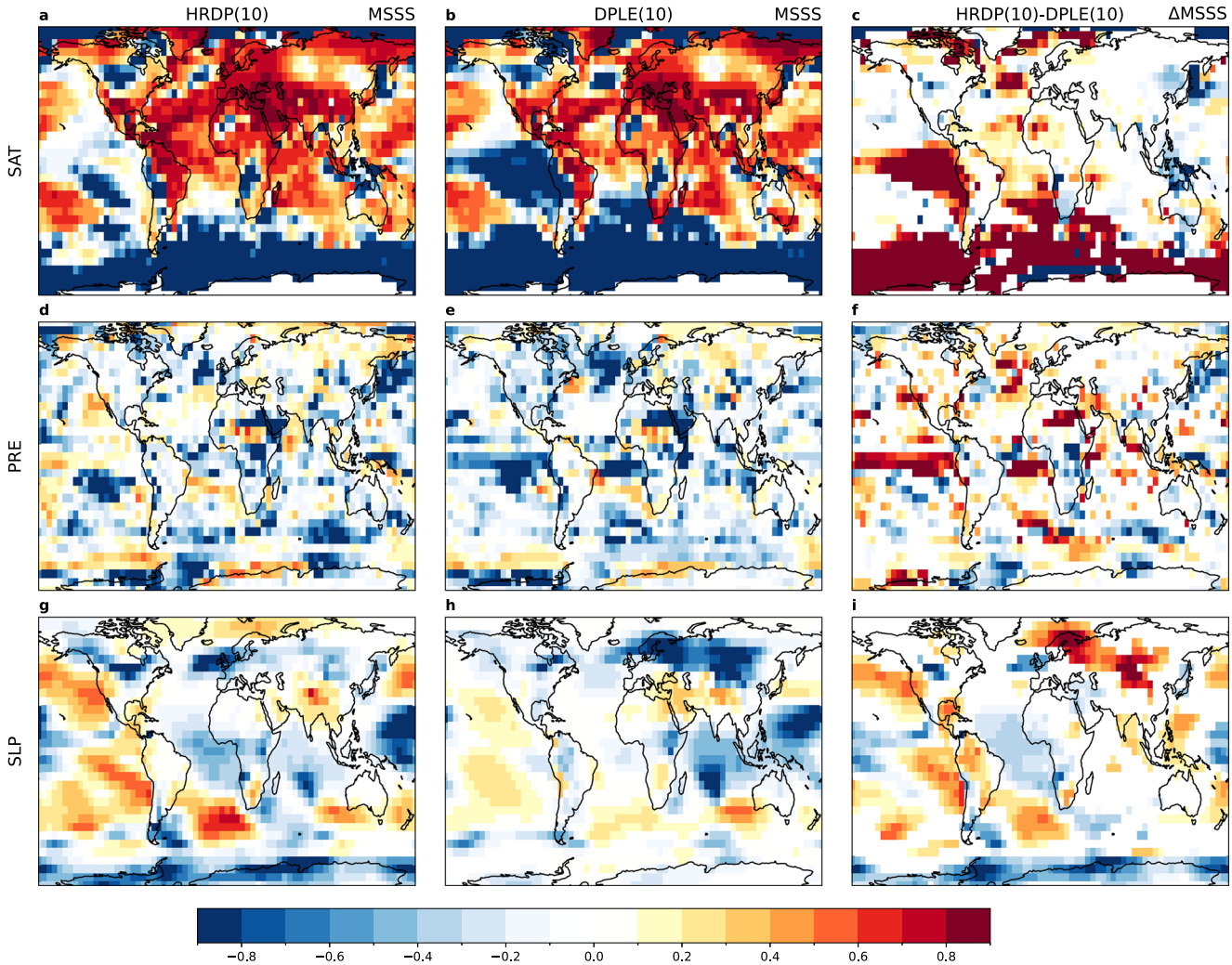

Fig. 2 Mean square skill score comparison for annual fields. MSSS skill for 10-member HRDP (**a**, **d**, **g**), 10-member DPLE (**b**, **e**, **h**), and their difference (**c**, **f**, **i**) for years 1–5 forecasts of annual SAT (top), PRE (middle), and SLP (bottom). Difference values in (**c**, **f**, **i**) are non-zero only where HRDP skill is significantly higher or lower than DPLE skill at the 90% confidence level (see Methods). See Supplementary Fig. 12 for additional perspective on off-scale MSSS values for SAT. See Supplementary Table 1 for a quantitative summary of skill differences plotted here.

under investigation, but it appears to be related to a less accurate reproduction of historical, observed variability in this region in the HR-FOSI simulation used to initialize HRDP compared to the LR-FOSI used for DPLE (Table 1). The SEP and EPSO regions still show large skill improvement ($\Delta ACC > 0.5$) in HRDP after detrending (Fig. 3c), implying that skillful prediction of non-trend (possibly internal) variability contributes to the overall improvement in raw SAT skill (Fig. 1c). The resilience to detrending of the general features of skill difference for annual PRE and SLP (compare panels f and i of Figs. 1 and 3) bolsters the conclusion that the HRDP-DPLE skill comparison reflects important differences in the representation of non-trend variance (see also Supplementary Fig. 3 for detrended MSSS maps).

Skill comparisons for seasonal fields show some noteworthy differences from the annual mean comparisons (see Supplementary Figs. 4–7), but the primary conclusion that high resolution generally enhances skill remains robust. In boreal winter (DJFM), the increased HRDP skill for PRE over the western US becomes even more pronounced while SLP shows more significant improvement ($\Delta ACC > 0.7$) around the coast of the continental US and Mexico (Supplementary Figs. 4 and 5). In boreal summer (JJAS), the enhanced skill for PRE in the tropical Pacific shifts further west and the SLP skill increase in the SEP and EPSO regions becomes more prominent (Supplementary Figs. 6 and 7). However, JJAS SLP skill degradation becomes worse over the North Atlantic and surrounding continents, resulting in overall skill degradation in HRDP compared to DPLE for this field in boreal summer (Supplementary Table 1). The detailed mechanisms that explain diverse regional skill differences and their seasonality will be examined in future work.

Signal-to-noise characteristics

The HRDP-DPLE comparison lends support to the hypothesis that higher model resolution can ameliorate the signal-to-noise paradox. We focus here on the SLP field, following previous work that has highlighted signal-to-noise issues in SLP predictions, often but not exclusively in the context of NAO predictions using large ensemble systems[3,4,6,7,11,13–15,19,42,43]. The model-world predictability of annual SLP (quantified as the square root of the signal-to-total variance fraction, or S2T; see Methods) is higher in HRDP than DPLE over much of the Southern Hemisphere (SH) and into the subtropics of the Northern Hemisphere, but lower over Eurasia and high northern latitudes (Fig. 4d–f). The ratio of predictable components (RPC = ACC/S2T; see Methods) is generally less than 1 over regions of positive skill in DPLE, implying

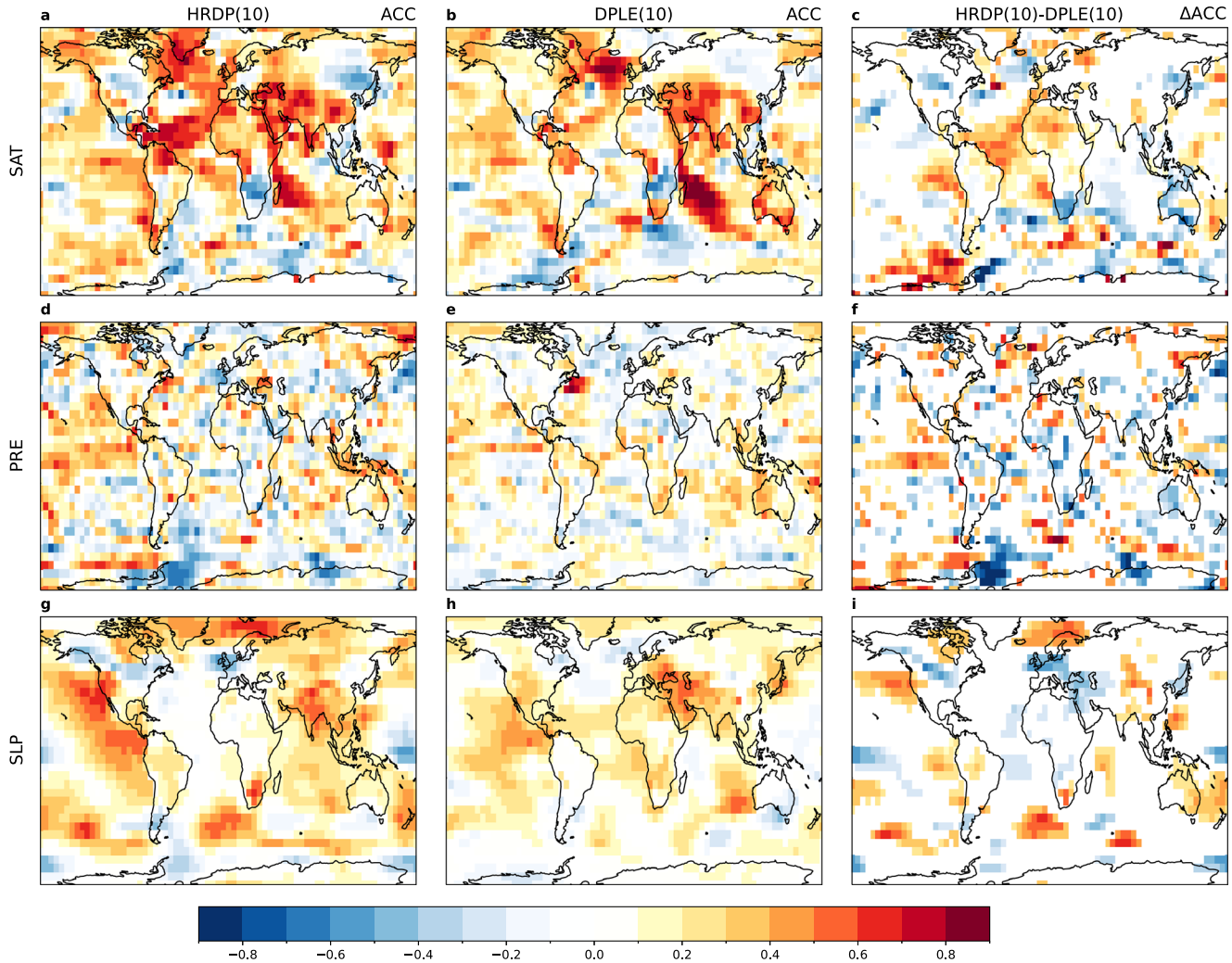

Fig. 3 Detrended correlation comparison for annual fields. ACC skill for 10-member HRDP (**a**, **d**, **g**), 10-member DPLE (**b**, **e**, **h**), and their difference (**c**, **f**, **i**) for years 1–5 forecasts of annual SAT (top), PRE (middle), and SLP (bottom). Difference values in (**c**, **f**, **i**) are non-zero only where HRDP skill is significantly higher or lower than DPLE skill at the 90% confidence level (see Methods). All timeseries were linearly detrended prior to ACC computation. See Supplementary Table 1 for a quantitative summary of skill differences plotted here.

that DPLE is overconfident for SLP predictions that achieve only modest actual skill (Fig. 4b, e, h). In contrast, HRDP ACC exceeds 0.5 over much more of the globe and these high-skill regions generally coincide with RPC values near 1, reflecting commensurate increases in both ACC and S2T over DPLE (Fig. 4g–i). A clear exception is over the Norwegian Sea and surrounding regions where RPC approaches 2 in HRDP, indicating prediction system underconfidence (signal-to-noise paradox). The large increase in the fraction of atmospheric signal variance in HRDP (Fig. 4f) is an intriguing result that implies that the atmosphere is more responsive to forcing (either external or natural variability from the ocean) in the high-resolution system. Furthermore, the combination of high signal variance and high skill in HRDP result in RPC values closer to 1 (indicating the system is more realistic; i.e., skillful without being overconfident or underconfident) over large areas of the Pacific, Southern Ocean, Indian Ocean, and Eurasia (Fig. 4a, d, g). However, the signal-to-noise characteristics in the Atlantic sector exhibit some degradation in HRDP. In addition to the significantly higher RPC in the polar Atlantic region noted above, the northeastern subtropical Atlantic is characterized by higher S2T and lower ACC in HRDP, yielding low RPC (system overconfidence) that is significantly worse than in DPLE.

The improvements in signal-to-noise characteristics in HRDP are quite distinct from the differences obtained by increasing ensemble size in DPLE (Supplementary Fig. 8). As noted above, increasing DPLE ensemble size from 10 to 40 yields slight increases in ACC magnitude that in general cannot be distinguished from 10-member ACC uncertainty at the grid scale. However, the larger ensemble size results in statistically significant reductions in S2T across the globe and especially so in the extratropics. The reduction in S2T with increased ensemble size is due to a reduction in extratropical signal variance, not an increase in total variance, because the overestimation of signal variance due to noise contamination decreases with ensemble size[13] while the estimation of the total variance is relatively insensitive to ensemble size (Supplementary Fig. 9). This combination (ACC increase is less than S2T decrease) results in an RPC field for 40-member DPLE showing a proliferation of signal-to-noise paradox regions (RPC > 1). The complex, regionally-dependent relationships between skill and signal-to-noise are summarized in joint probability distributions that relate ACC, RPC, and S2T from the different systems (Fig. 5). Increasing the DPLE ensemble size from 10 to 40 results in more than a doubling of area showing high skill (here, defined as ACC > 0.5) for annual SLP, but this skill increase comes at the expense of lower S2T and higher RPC (Fig. 5b–d).

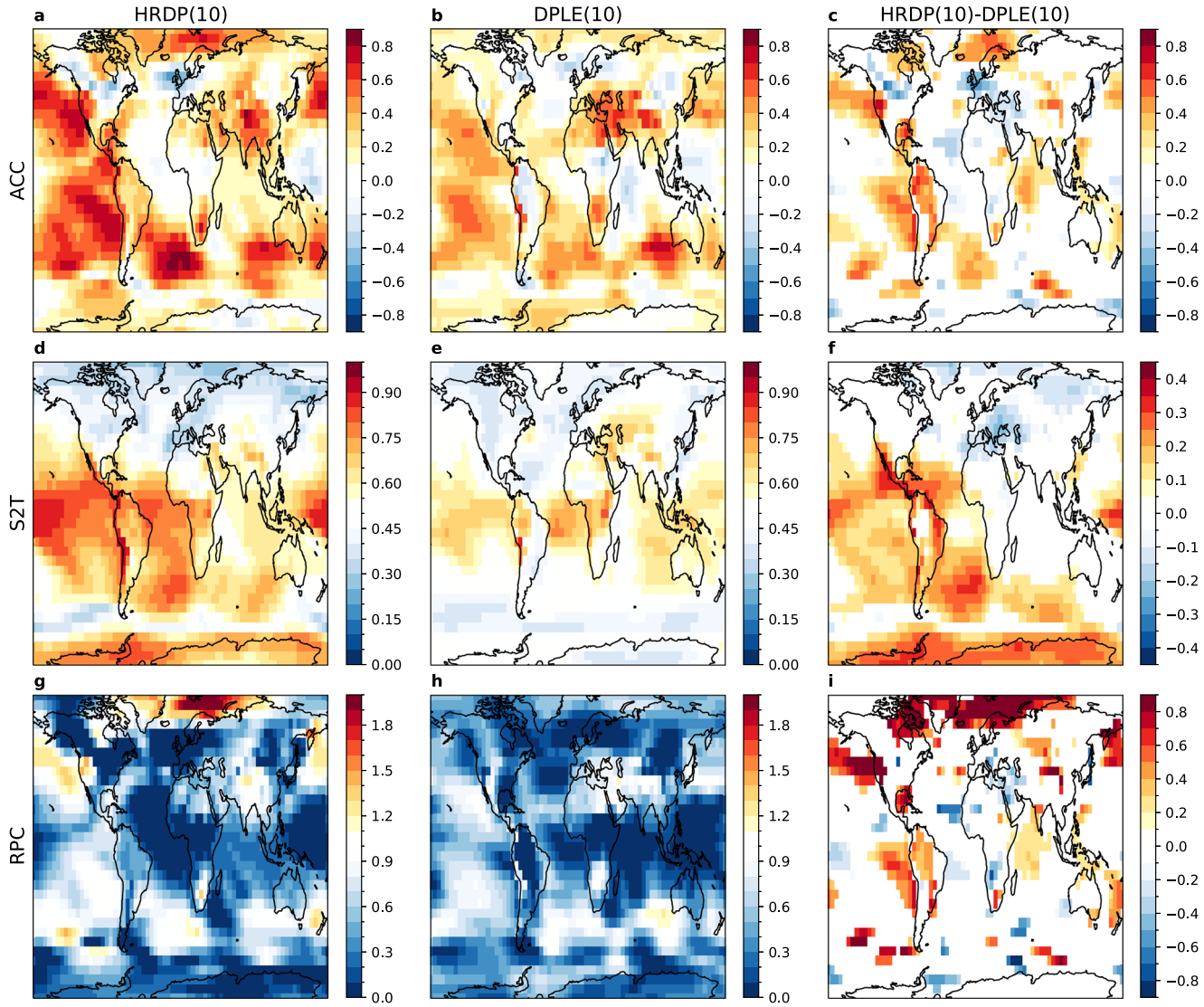

Fig. 4 Signal-to-noise metrics for annual SLP. ACC, S2T, RPC for 10-member HRDP (**a**, **d**, **g**), 10-member DPLE (**b**, **e**, **h**), and their difference (**c**, **f**, **i**) for years 1–5 forecasts of annual SLP. Difference values in (**c**, **f**, **i**) are non-zero only where HRDP values are significantly higher or lower than DPLE values at the 90% confidence level (see Methods). Note that panels (**a**–**c**) replicate Fig. 1g–i to facilitate comparison.

HRDP yields a fourfold increase in high skill area compared to 10-member DPLE for SLP, and the high skill is paired with high S2T such that RPC stays close to 1 (Fig. 5a, d). Metrics for annual PRE yield similar results, although there is less increase in high skill area in HRDP compared to DPLE (slightly more than double given the same ensemble size; Fig. 5e–h). We conclude that HR produces not only higher skill, but more realistic signal-to-noise properties for atmospheric fields compared to a LR large ensemble system due to higher signal variance fractions in most, but not all, high-skill regions.

Large-scale climate trends

These results may also help to shed light on the potential failure of CMIP-class models to reproduce observed large-scale trends in surface temperature and pressure, particularly in the tropical Pacific. Most climate models indicate that the response of the tropical Pacific to rising greenhouse gases is El Nino-like with a relatively greater warming of the tropical Pacific in the east compared to the west and, consequently, a reduction in the west-to-east warm-to-cool SST gradient. This is in stark contrast to the strengthening of the west-to-east SST gradient that has been observed in recent decades and, even when accounting for the role of internal variability, the observed trends lie very close to the edge of the modeled distribution[44–48]. This raises the concern that there may be something fundamentally wrong with the modeled representation of the response to forcing or internal variability in this region, but the possibility remains that the observed trends have just been a very unlikely occurrence. Two potential model errors have been invoked to explain this issue. Some have argued that the origins of the problem lie in the tropical Pacific itself, with models incorrectly simulating the ocean dynamical thermostat mechanism, whereby anomalous oceanic upwelling in response to warming, modulates the temperature rise in the eastern tropical Pacific[45,46,49]. Others have argued that the tropical Pacific issues are actually a result of teleconnections from the SO where models seem to poorly represent SO cooling (and Antarctic sea ice expansion)[48,50–54]. Wills et al.[48] show that the leading signal-to-noise maximizing pattern of model-observations difference based on CMIP historical ensembles highlights deficiencies in simulating EPSO and eastern tropical Pacific cooling together with SLP increase in the extratropical Pacific, South Atlantic, and South Indian sectors (their Fig. 3).

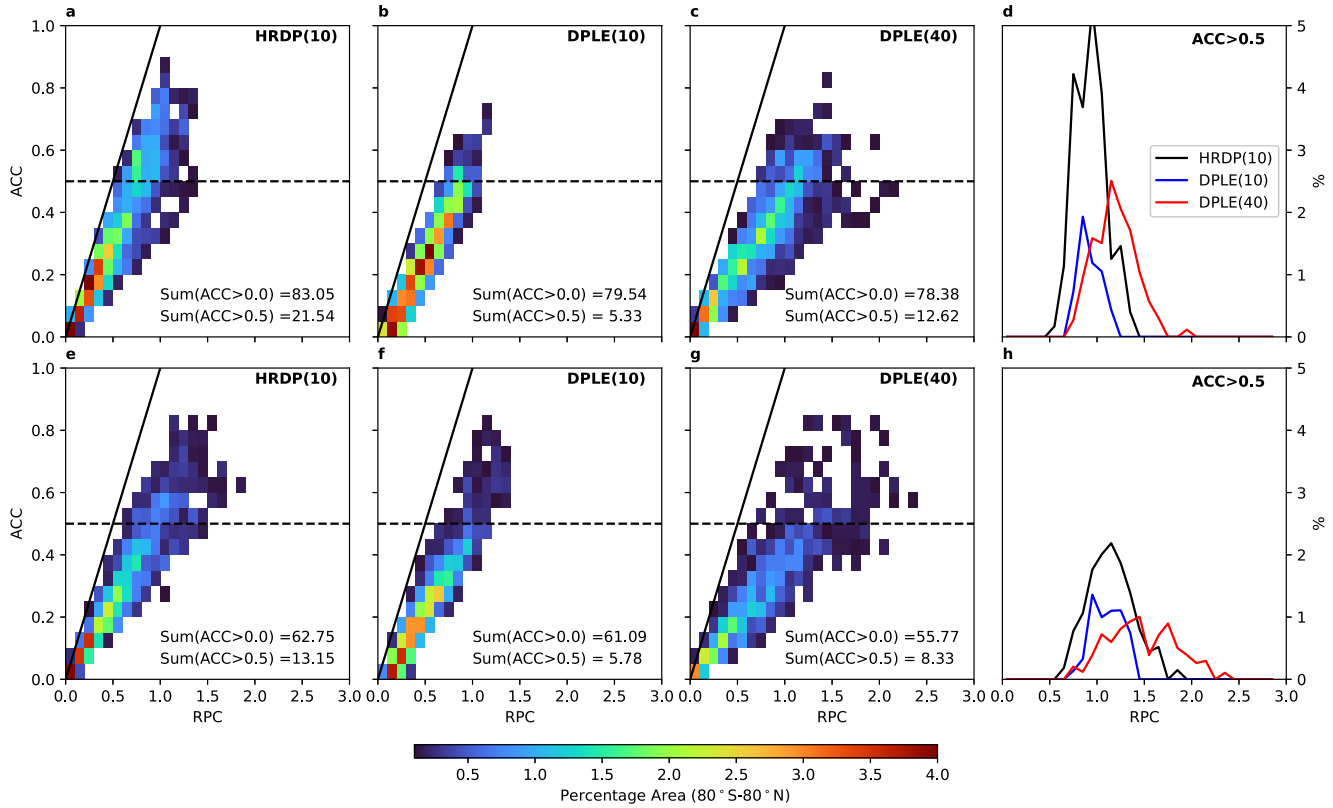

Fig. 5 Global skill and signal-to-noise for annual SLP and PRE. For annual SLP (**a–c**) and PRE (**e–g**), global (80°S–80°N) joint probability distributions summarizing the paired relationship between ACC and RPC scores from 10-member HRDP (**a**, **e**), 10-member DPLE (**b**, **f**), and 40-member DPLE (**c**, **g**). Color fill shows the percent of surface area for each (ACC,RPC) pair, and text indicates total percent area characterized by positive (ACC > 0) and high (ACC > 0.5; horizontal dashed line) skill. Note that the slope of the line between the origin and a point in (ACC, RPC)-space reflects the S2T value for that point (solid line shows S2T = 1). Right panels show the integrated areas (in %) for high-skill regions for annual SLP (**d**) and PRE (**h**) as a function of RPC.

HRDP pentadal forecasts show much more realistic multidecadal trend patterns for SAT, PRE, and SLP than DPLE (Fig. 6). In particular, the lack of warming in the SO and eastern tropical Pacific, and the positive SLP trends in the midlatitude oceans, are well-captured with patterns that project strongly onto the leading bias pattern identified in CMIP models[48]. The spatial structure of HRDP PRE trends is also more in line with observations. The realism of HRDP trends is highest for FY1 and degrades with lead time, but even FY5 trends show muted warming over the SO and SEP together with SLP increase over SH midlatitudes (Supplementary Fig. 10). This implies that initialization and/or how the model propagates the signals introduced at initialization contributes to realistic multidecadal trends in HRDP, not just improved response to external forcing. The realism of DPLE trends is also highest for FY1 but degrades rapidly such that large trend biases are apparent by FY2 (Supplementary Fig. 11). The better trend representation in HRDP corresponds to improved FY1–5 skill metrics for several key climate indices such as the tropical Pacific zonal SST gradient, SO surface temperature, and Walker Circulation strength (Supplementary Figs. 12 and 13). While HRDP forecasts do show a warming trend for SO SAT averaged over all longitudes (resulting in a negative correlation with observations), the SO warming in HRDP is much less than in DPLE and SAT in the Pacific sector of the SO is quite well represented in HRDP (Supplementary Fig. 12).

The results indicate that resolution alters the representation of processes such that trends in the SO and tropical Pacific in these decadal predictions are brought more in line with observations. This strongly suggests that it is unlikely that LR CMIP-class models are all behaving correctly and the observed trends have been a statistically unlikely occurrence, but rather that such models may be mis-representing relevant processes and that higher resolution in the ocean and/or atmosphere may help remedy this. The improvement in skill in the tropical Pacific is also closely linked to the improvement in skill in the SO, adding to the growing body of evidence (cited above) that the SO may act as a key pacemaker for observed multidecadal variability in the tropics, and suggesting that the origins of the improved tropical Pacific and global skill in HRDP may lie in the improved representation of processes in the EPSO. While there may be a role for an improved representation of the ocean dynamical thermostat mechanism and/or improved extratropical-tropical feedbacks, SAT skill improvement in HRDP is significant in the western and central Pacific SO at early leads (FY1 and FY2) when there is no clear evidence of improved tropical Pacific skill (Fig. 7). SEP skill improvement becomes significant only when EPSO SAT skill degrades in DPLE but not in HRDP, starting at FY3. This suggests that SEP skill improvements derive from SO improvements, and not the other way around.

DISCUSSION

We have directly compared two initialized hindcast sets using CESM that permit an assessment of decadal prediction sensitivity to model horizontal resolution (Table 1), with a focus on pentadal timescales. The HRDP system significantly outperforms DPLE, the CESM contribution to DCPP of CMIP6, at predicting multiyear variations in SAT, PRE, and SLP after accounting for sample and ensemble size differences. Skill improvements are evident in both the phasing and magnitude of forecast anomalies, although there are local regions of skill degradation as well. The most prominent

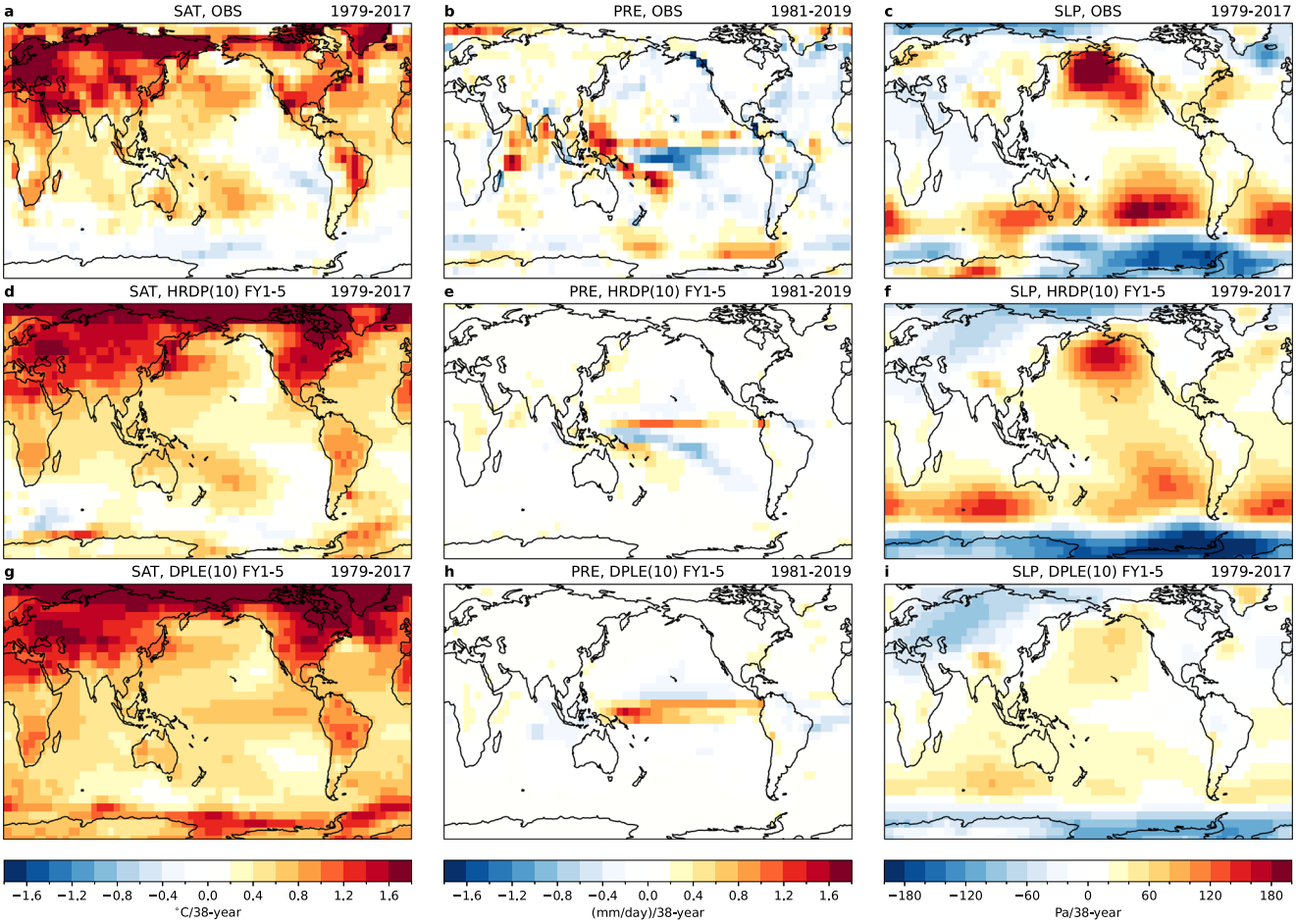

Fig. 6 Global trends for annual SAT, PRE, and SLP. Multidecadal trends for annual SAT (**a**, **d**, **g**), PRE (**b**, **e**, **h**), and SLP (**c**, **f**, **i**) from observations (**a**–**c**), 10-member HRDP (**d**–**f**), and 10-member DPLE (**g**–**i**). HRDP and DPLE trends are computed from FY1–5 predicted anomalies, while a 5-year running smoother is applied to observed timeseries before trend computation. The years included in the trend computation reflect the intersection of available years from HRDP and observations.

skill differences are resilient to detrending, and other regions of skill increase emerge after detrending (e.g., SAT in the tropical Atlantic), suggesting a role for internal variability in explaining performance differences. We showed that unrealistic signal-to-noise ratios in the low-resolution system, most evident when skill is enhanced with a large ensemble, are significantly ameliorated (although not eliminated) in the high-resolution system. The improved signal-to-noise characteristics were related to the significantly larger fraction of signal variance (relative to total variance) in the high-resolution system, particularly for SLP in the SH in the Pacific and Atlantic sectors. The comparison highlights prediction system performance enhancements associated with increased system resolution that are qualitatively different from (and for many regions and fields, better than) those obtained by increasing ensemble size.

The complexity of decadal prediction experiments makes it challenging to attribute skill differences to the improved or degraded representation of specific physical mechanisms, and it is hoped that this global assessment helps to guide future efforts to better understand diverse regional and seasonal sensitivities. The global skill overview reveals a significant improvement in pentadal SAT skill in the southeastern tropical Pacific (Fig. 1c), a region that exerts a strong influence on global climate and hence global climate prediction skill through its impact on the zonal temperature gradient in the tropical Pacific (Supplementary Fig. 12), Walker Circulation strength (Supplementary Fig. 13), and the strength and location of tropical convective heating anomalies (Figs. 1f and 2f). The SEP improvement appears to relate to the reduction of a spurious warming trend in that region present in DPLE as well as better representation of non-trend decadal fluctuations (Supplementary Fig. 12d). It seems likely that SEP improvements in HRDP are fundamentally related to improved predictions in the Pacific sector of the SO (Fig. 1c and Supplementary Fig. 12), a region where HRDP also shows much less spurious warming and better representation of non-trend variance. Our interpretation is that HRDP skill enhancements emanate from the EPSO to the SEP and the rest of the globe by correcting spurious climate trends in DPLE (Fig. 6), in line with and adding supporting evidence for the hypothesis that excessive SO warming is a critical bias that explains the ubiquitous misrepresentation of observed tropical climate trends in LR models[48,50–54].

The results raise questions that will require substantial future work (likely involving dedicated sensitivity experiments) to definitively answer. A key outstanding question is: what specific aspects of the HRDP system are responsible for improved skill and signal-to-noise characteristics? The first-order EPSO and SEP improvements could be related to a more realistic simulation of low cloud-SST feedbacks in the HR atmosphere[51,54] and/or to improved two-way feedbacks between the EPSO and the tropical Pacific[52]. Reduced tropical Pacific SST bias in the HR coupled model[26,28] could contribute to an improved representation of the tropical dynamical thermostat mechanism[45,46,49] which could initiate and sustain two-way EPSO-SEP coupling more realistically than in LR. The eddy-resolving ocean model could also be playing

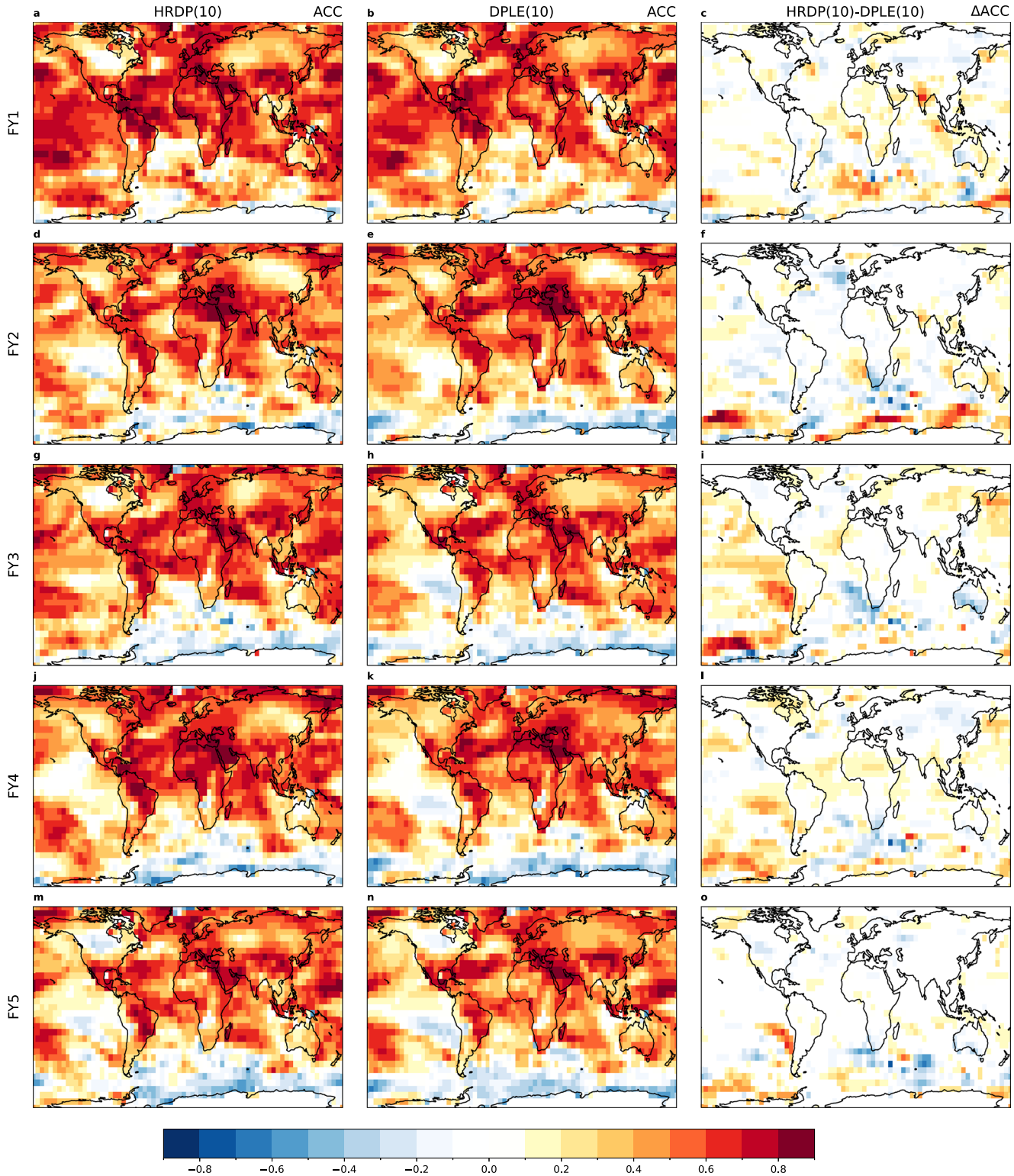

Fig. 7 Correlation comparison for annual SAT by forecast year. ACC skill for 10-member HRDP (**a**, **d**, **g**, **j**, **m**), 10-member DPLE (**b**, **e**, **h**, **k**, **n**), and their difference (**c**, **f**, **i**, **l**, **o**) for annual SAT for forecast years 1 through 5 (top to bottom rows). Difference values are non-zero only where HRDP skill is significantly higher or lower than DPLE skill at the 90% confidence level (see Methods).

a role in amplifying the extratropical-tropical teleconnection through stronger oceanic forcing of the atmosphere via mesoscale air-sea interactions that are present in HR but absent in LR[18,31–35,55,56], and this small-scale exchange could be a factor contributing to higher signal variance in HRDP atmospheric fields (Figs. 4 and 5).

Another possibility is that the use of an eddy-resolving ocean component in HRDP accounts for much of the improved performance in HRDP due to a more realistic representation of ocean and sea ice evolution in the SO. The residual mean overturning circulations in the SO in the FOSI simulations used to initialize HRDP and DPLE exhibit large differences in terms of both

time-mean amplitude and multidecadal trends in response to historical forcing (Supplementary Fig. 14). The strengthening and poleward expansion of the clockwise primary overturning cell and weakening and contraction of the lower (anticlockwise) overturning cell seen in HR-FOSI over 1979–2017 are absent in LR-FOSI, which instead shows a strengthening of the lower cell. The HR-FOSI trend is more in line with an observation-based estimate of recent SO overturning change[57]. The net poleward heat transport trends, which reflect the ocean model response to the historical strengthening of the Southern Annular Mode (SAM) due to greenhouse gas and ozone forcings, are also very different, with LR-FOSI showing increasing poleward heat transport while HR-FOSI shows decreasing poleward heat transport (Supplementary Fig. 14). The latter is more in line with the hypothesized role for the ocean in delaying SO warming[58], and it yields a better reproduction of observed Antarctic sea ice extent change (Supplementary Fig. 15). These results appear consistent with recent work arguing that eddy-resolving ocean models may be needed to improve simulations of SO response to forcing[59,60]. The HRDP system, thus, likely benefits from improved ocean and sea ice initial conditions in the SO that could enhance predictability of internal SO modes of variability[61] while the HR ocean component also yields an improved SO response to predicted SAM increase (Fig. 6 and Supplementary Figs. 12–15).

We hypothesize that a confluence of resolution-related improvements in process representation accounts for enhanced performance in a HR decadal prediction system compared to its LR counterpart using CESM. The SO stands out in this preliminary analysis as a critical region contributing to global skill sensitivity to model resolution due to its apparent influence on global low-frequency climate variability. These prediction results add to a growing body of work highlighting the role of the SO as a global climate pacemaker, and they suggest that increased resolution may deliver not only improved climate predictions but also more realistic historical and projection simulations by reducing spurious SO warming. The high cost of such configurations for climate applications remains a significant barrier, but to the extent that HR improvements reflect better process representation, the reward could be transformative breakthroughs in our understanding of Earth system predictability and coupled model fidelity.

METHODS

Prediction systems

The CESM DPLE system is an ensemble forecast set using the low-resolution component models (nominal 1°) of CESM1.1[62] that was submitted to the DCPP[21] of the CMIP6. An overview paper[5] provides extensive detail about the DPLE prediction system, and important design choices are summarized in Table 1. Historical state information is incorporated in DPLE by initializing the ocean and sea ice component models from a forced ocean−sea ice (FOSI) simulation conducted using reanalysis-derived surface atmospheric boundary conditions based on the version 1 protocol of the Ocean Model Intercomparison Project (OMIP1)[5,63]. The atmosphere and land component models are initialized from a single member of the CESM1 Large Ensemble[62] and as such are not strongly constrained to match historical conditions at the time of initialization. The DPLE system is comprised of 40-member forecasts initialized (full field) each November 1 from 1954 to 2017 (for a total of 64 start dates) and each integrated for 122 months. Ensemble generation is accomplished through round-off level perturbations to the atmospheric potential temperature initial state.

The CESM HRDP system uses a slightly different version of the model (CESM1.3[26,64]) in a high-resolution configuration (~0.1° in the ocean and sea ice; ~0.25° in the atmosphere and land). The 0.1° ocean grid is paired with a 62-level vertical grid (62L) that is identical to that used with the 1° ocean grid apart from the addition of 2 abyssal levels to yield a maximum depth of 6000 m. HRDP uses a spectral element dynamical core in the atmosphere component[64,65], rather than the finite volume dynamical core used in DPLE. Ocean and sea ice are initialized from a FOSI simulation as in DPLE, but HR-FOSI differs from LR-FOSI as follows: (1) it uses the 0.1° ocean and sea ice component models of CESM2[66]; and (2) it uses OMIP2[63] forcing derived from the Japanese 55-year Reanalysis (JRA55-do[67]). Initialization of atmosphere and land components is somewhat more sophisticated in HRDP than in DPLE. The atmospheric initial conditions are regridded JRA55 fields, and the land initial conditions come from a high-resolution atmosphere−land simulation forced with observed SSTs (the CESM contribution to HighResMIP[68]). As in DPLE, full field initialization is used for all components. The HRDP forecast set is necessarily smaller than DPLE (due to computational expense), comprising 10-member ensembles initialized every other November 1 between 1976 and 2016 and each integrated for 62 months. As in DPLE, ensemble generation is accomplished in HRDP through round-off level perturbation of the atmospheric potential temperature initial condition.

Of all the differences in system design between DPLE and HRDP (Table 1), changes in model horizontal resolution and ocean and sea ice initialization are likely to account for most of the differences in system performance on decadal timescales. Investigating the relative impacts of specific changes in system design will be pursued in future work.

Skill metrics

Prediction skill is assessed by comparing ensemble mean forecast anomalies, $\acute{f}_i$, to observed anomalies, $\acute{o}_i$, for each of the forecast annual means available from HRDP (i.e., $i = \{1977, 1979, \ldots, 2017\}$ for FY1). Forecast anomalies are computed relative to model climatology that varies with FY (τ) as follows:

$$\acute{f}_{i\tau} = f_{i\tau} - \bar{f}_{\tau} = f_{i\tau} - \frac{1}{N}\sum_{i=1}^{N} f_{i\tau} \quad (1)$$

where the sum includes the N forecasts $f_{i\tau}$ that verify within the climatology window of 1981–2017 (note that $N = 19$ for $\tau = \{1, 2, 3, 4, 5\}$). Observed anomalies are also computed relative to 1981–2017 climatology. Skill metrics are computed for a particular FY (or FY average), and this lead time dependence is implicit in what follows. The Pearson ACC is given by:

$$\mathrm{ACC} = \frac{\sum_{i=1}^{21} \acute{f}_i \acute{o}_i}{\sum_{i=1}^{21} \acute{f}_i^{\,2} \sum_{i=1}^{21} \acute{o}_i^{\,2}} \quad (2)$$

The MSSS is computed as follows:

$$\mathrm{MSSS} = 1 - \frac{\sum_{i=1}^{21} \left(\acute{f}_i - \acute{o}_i\right)^2}{\sum_{i=1}^{21} \acute{o}_i^{\,2}} \quad (3)$$

Note that the summation over 21 samples corresponds to the number of hindcasts available from HRDP, but that the temporal sample size can be lower than 21 if corresponding observations are not available. ACC and MSSS are both deterministic skill metrics that quantify the accuracy of a prediction[69]. The ACC measures the linear association between forecasts and observations, and as it is insensitive to the magnitude of anomalies, it primarily reflects the correct phasing of variability[7]. The MSSS is a summary metric that combines the correlation with the conditional bias, and since it reflects the ratio of forecast error variance relative to observed variance, it can be viewed as a measure of the accuracy of forecast anomaly magnitudes[69]. For a perfect forecast, both ACC and MSSS are equal to 1.

The ratio of predictable components is defined as in previous studies[7,12,13] as:

$$\mathrm{RPC} = \frac{\mathrm{ACC}}{\sigma^{\mathrm{f}}_{\mathrm{sig}}/\sigma^{\mathrm{f}}_{\mathrm{tot}}} = \frac{\mathrm{ACC}}{\mathrm{S2T}} \tag{4}$$

where the denominator is the signal-to-total (S2T) variability ratio of the forecasts (f), and the numerator is a lower bound approximation of the S2T variability ratio of the real world. As has been discussed in prior work[70], the RPC metric as computed above is only an estimate of the true RPC, and its accuracy and relevance are questionable when ACC is low. In this paper, RPC values are set to zero where ACC < 0. In a perfect prediction system, RPC is equal to 1, and RPC values below and above 1 reflect overconfident and underconfident predictions, respectively[42].

Signal and total variability from the ensemble forecasts is decomposed as follows:

$$\sigma^{\mathrm{f}}_{\mathrm{sig}} = \sqrt{\frac{1}{21}\sum_{i=1}^{21}\left(\frac{1}{M}\sum_{m=1}^{M}\acute{f}_{im}\right)^2} \quad \sigma^{\mathrm{f}}_{\mathrm{tot}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{21}\sum_{i=1}^{21}\left(\acute{f}_{im_{ni}}\right)^2} \tag{5}$$

Here, $\sigma^{\mathrm{f}}_{\mathrm{sig}}$ is the maximum likelihood estimate of the temporal standard deviation of the ensemble mean forecast (the signal). It is computed by first averaging forecast anomalies over member index m for a given ensemble size M, taking the square, averaging over the forecast sample index i, and taking the square root. The total variability of the forecasts, $\sigma^{\mathrm{f}}_{\mathrm{tot}}$, is computed similarly but from individual-member timeseries constructed by randomly selecting one ensemble member (m_{ni}) for each sample (i) and for each iteration (n). For both DPLE and HRDP, the $\sigma^{\mathrm{f}}_{\mathrm{tot}}$ estimate is the average of $N = 100$ iterations. To account for differences in ensemble size between DPLE and HRDP, skill scores from DPLE are computed by randomly sampling a 10-member ensemble from the 40-member pool (for each sample i), computing a skill metric as outlined above, then repeating 100 times to yield a distribution of skill scores that can be compared to the single 10-member skill score available from HRDP.

Verification

Prediction skill is verified against the following observational datasets: CRU-TS version 4.05[71] for SAT over land and HadISST1[72] for ocean surface temperature; GPCP version 2.3[73] for PRE; ERA5[74] reanalysis for SLP; and the Sea Ice Index[75] version 3.0 from the National Snow and Ice Data Center for Antarctic sea ice extent (Supplementary Fig. 15). All model and observed fields are conservatively mapped to a regular 5° × 5° grid prior to skill assessment. The merged SAT and ERA5 SLP fields used here are available for all years through 2020; the PRE field extends from 1979 to 2021. This paper focuses primarily on skill at predicting pentadal anomalies corresponding to FY1–5. Forecast year averages correspond to calendar year averages (January to December), such that the forecast year 1 average includes forecast months 3–14 for predictions initialized in November.

Significance testing

Significance testing focuses on whether skill differences between the two systems are significant. To test whether HRDP skill scores are significantly different from DPLE skill scores, the single realization of 10-member HRDP skill is compared to a bootstrapped distribution ($N = 100$) of 10-member DPLE skill scores (see above). Skill decrease/increase is deemed significant if HRDP scores fall outside of the 0.05–0.95 quantile values of the DPLE distribution.

DATA AVAILABILITY

The full DPLE dataset is available from NCAR's Climate Data Gateway at https://www.earthsystemgrid.org/dataset/ucar.cgd.ccsm4.CESM1-CAM5-DP.html. The regridded (5° × 5°) HRDP data used for the analyses presented in this paper are archived at the NCAR Geoscience Data Exchange (GDEX) at https://doi.org/10.5065/9t56-sm14.

CODE AVAILABILITY

The CESM1.1 code used to generate DPLE is available at https://www.cesm.ucar.edu/models/cesm1.1/index.html. The CESM1.3 code used to generate HRDP is available at https://github.com/ihesp/cesm/tree/ihesp-hires-master. The Python code used to generate manuscript figures is available at the NCAR Geoscience Data Exchange (GDEX) at https://doi.org/10.5065/9t56-sm14.

Received: 12 October 2022; Accepted: 19 July 2023;
Published online: 31 July 2023

REFERENCES

1. Merryfield, W. J. et al. Current and emerging developments in subseasonal to decadal prediction. *B. Am. Meteorol. Soc.* **101**, E869–E896 (2020).
2. Meehl, G. A. et al. Initialized Earth System prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ.* **2**, 340–357 (2021).
3. Smith, D. M. et al. Robust skill of decadal climate predictions. *npj Clim. Atm. Sci.* **2**, 13 (2019).
4. Dunstone, N. et al. Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nat. Geosci.* **9**, 809–814 (2016).
5. Yeager, S. G. et al. Predicting near-term changes in the Earth System: a large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *B. Am. Meteorol. Soc.* **99**, 1867–1886 (2018).
6. Athanasiadis, P. J. et al. Decadal predictability of North Atlantic blocking and the NAO. *npj Clim. Atm. Sci.* **3**, 20 (2020).
7. Smith, D. M. et al. North Atlantic climate far more predictable than models imply. *Nature* **583**, 796–800 (2020).
8. Dunstone, N. J. et al. Skilful interannual climate prediction from two large initialised model ensembles. *Environ. Res. Lett.* **15**, 094083 (2020).
9. Yeager, S. G. et al. The Seasonal-to-Multiyear Large Ensemble (SMYLE) prediction system using the Community Earth System Model version 2. *Geosci. Model Dev.* **15**, 6451–6493 (2022).
10. Hermanson, L. et al. WMO global annual to decadal climate update: a prediction for 2021–25. *Bull. Amer. Meteor. Soc.* **103**, E1117–E1129 (2022).
11. Scaife, A. A. et al. Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.* **41**, 2514–2519 (2014).
12. Eade, R. et al. Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.* **41**, 5620–5628 (2014).
13. Scaife, A. A. & Smith, D. A signal-to-noise paradox in climate science. *npj Clim. Atm. Sci.* **1**, 28 (2018).
14. Zhang, W. & Kirtman, B. Understanding the signal-to-noise paradox with a simple Markov model. *Geophys. Res. Lett.* **46**, 308–13,317 (2019).
15. Klavans, J. M., Cane, M. A., Clement, A. C. & Murphy, L. N. NAO predictability from external forcing in the late 20th century. *npj Clim. Atm. Sci.* **4**, 22 (2021).
16. Wu, X., Yeager, S. G., Deser, C., Rosenbloom, N. & Meehl, G. Volcanic forcing degrades multiyear-to-decadal prediction skill in the tropical Pacific. *Sci. Adv.* **9**, eadd9364 (2023).
17. Siqueira, L. & Kirtman, B. P. Atlantic near-term climate variability and the role of a resolved Gulf Stream. *Geophys. Res. Lett.* **43**, 964–3,972 (2016).
18. Kirtman, B. P., Perlin, N. & Siqueira, L. Ocean eddies and climate predictability. *Chaos* **27**, 126902 (2017).
19. Scaife, A. A. et al. Does increased atmospheric resolution improve seasonal climate predictions? *Atmos. Sci. Lett.* **20**, e922 (2019).
20. Zhang, W., Kirtman, B., Siqueira, L., Clement, A. & Xia, J. Understanding the signal-to-noise paradox in decadal climate predictability from CMIP5 and an eddying global coupled model. *Clim. Dyn.* **56**, 2895–2913 (2021).
21. Boer, G. J. et al. The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev.* **9**, 3751–3777 (2016).
22. Gent, P. R., Yeager, S. G., Neale, R. B., Levis, S. & Bailey, D. A. Improvements in a half degree atmosphere/land version of the CCSM. *Clim. Dyn.* **34**, 819–833 (2010).
23. Small, R. J. et al. A new synoptic scale resolving global climate simulation using the Community Earth System Model. *J. Adv. Model. Earth Syst.* **6**, 1065–1094 (2014).
24. Chassignet, E. P. et al. Impact of horizontal resolution on global ocean–sea ice model simulations based on the experimental protocols of the Ocean Model

Intercomparison Project phase 2 (OMIP-2). *Geosci. Model Dev.* **13**, 4595–4637 (2020).
25. Roberts, M. J. et al. Project future changes in tropical cyclones using the CMIP6 HighResMIP multimodel ensemble. *Geophys. Res. Lett.* **47**, e2020GL088662 (2020).
26. Chang, P. et al. An unprecedented set of high-resolution earth system simulations for understanding multiscale interactions in climate variability and change. *J. Adv. Model. Earth Syst.* **12**, e2020MS002298 (2020).
27. Yeager, S. G. et al. An outsized role for the Labrador Sea in the multidecadal variability of the Atlantic overturning circulation. *Sci. Adv.* **7**, eabh3592 (2021).
28. Xu, G. et al. Impacts of model horizontal resolution on mean sea-surface temperature biases in the Community Earth System Model. *J. Geophys. Res. Oceans* **127**, e2022JC019065 (2022).
29. Chang, P. et al. Uncertain future of sustainable fisheries environment in eastern boundary upwelling zones under climate change. *Commun. Earth Environ.* **4**, 19 (2023).
30. Li, D. et al. The impact of horizontal resolution on projected sea-level rise along US east continental shelf with the Community Earth System Model. *J. Adv. Model. Earth Syst.* **14**, e2021MS002868 (2022).
31. Ma, X. et al. Distant influence of Kuroshio Eddies on North Pacific weather patterns. *Sci. Rep.* **5**, 17785 (2015).
32. Ma, X. et al. Importance of resolving Kuroshio front and eddy influence in simulating the North Pacific Storm Track. *J. Clim.* **30**, 1861–1880 (2017).
33. Foussard, A., Lapeyre, G. & Plougonven, R. Storm track response to oceanic eddies in idealized atmospheric simulations. *J. Clim.* **32**, 445–463 (2019).
34. Liu, X. et al. Ocean fronts and eddies force atmospheric rivers and heavy precipitation in western North America. *Nat. Commun.* **12**, 1268 (2021).
35. Laurindo, C. L. et al. Role of ocean and atmosphere variability in scale-dependent thermodynamic air-sea interactions. *J. Geophys. Res. Oceans* **127**, e2021JC018340 (2022).
36. Deser, C., Simpson, I. R., Phillips, A. S. & McKinnon, K. A. How well do we know ENSO's climate impacts over North America, and how do we evaluate models accordingly? *J. Clim.* **31**, 4991–5014 (2018).
37. Lehner, F., Deser, C., Simpson, I. R. & Terray, L. Attributing the U.S. Southwest's recent shift into drier conditions. *Geophys. Res. Lett.* **45**, 6251–6261 (2018).
38. Zhang, R. et al. A review of the role of the atlantic meridional overturning circulation in Atlantic multidecadal variability and associated climate impacts. *Rev. Geophys.* **57**, 316–375 (2019).
39. Yeager, S. G. & Robson, J. I. Recent progress in understanding and predicting Atlantic decadal climate variability. *Curr. Clim. Change Rep.* **3**, 112–127 (2017).
40. Yeager, S. G. The abyssal origins of North Atlantic decadal predictability. *Clim. Dyn.* **55**, 2253–2271 (2020).
41. Dunstone, N. J., Smith, D. M. & Eade, R. Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean. *Geophys. Res. Lett.* **38**, L14701 (2011).
42. Siegert, S. et al. A Bayesian framework for verification and recalibration of ensemble forecasts: how uncertain is NAO predictability? *J. Clim.* **29**, 995–1012 (2016).
43. Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A. & Scaife, A. A. An intercomparison of skill and overconfidence/underconfidence of the wintertime North Atlantic oscillation in multimodel seasonal forecasts. *Geophys. Res. Lett.* **45**, 7808–7817 (2018).
44. Coats, S. & Karnauskas, K. B. Are simulated and observed twentieth century tropical Pacific sea surface temperature trends significant relative to internal variability? *Geophys. Res. Lett.* **44**, 9928–9937 (2017).
45. Seager, R. et al. Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nat. Clim. Chang.* **9**, 517–522 (2019).
46. Seager, R., Henderson, N. & Cane, M. Persistent discrepancies between observed and modeled trends in the tropical Pacific Ocean. *J. Clim.* **35**, 4571–4584 (2022).
47. Lee, S. et al. On the future zonal contrasts of equatorial Pacific climate: perspectives from observations, simulations, and theories. *npj Clim. Atm. Sci.* **5**, 82 (2022).
48. Wills, R. C. J., Dong, Y., Proistosecu, C., Armour, K. C. & Battisti, D. S. Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophys. Res. Lett.* **49**, e2022GL100011 (2022).
49. Clement, A. C., Seager, R., Cane, M. A. & Zebiak, S. E. An ocean dynamical thermostat. *J. Clim.* **9**, 2190–2196 (1996).
50. Hwang, Y.-T., Xie, S.-P., Deser, C. & Kang, S. M. Connecting tropical climate change with Southern Ocean heat uptake. *Geophys. Res. Lett.* **44**, 9449–9457 (2017).
51. Kim, H., Kang, S., Kay, J. E. & Xie, S.-P. Subtropical clouds key to Southern Ocean teleconnections to the tropical Pacific. *Proc. Natl. Acad. Sci. USA.* **119**, e2200514119 (2022).
52. Dong, Y., Armour, K. C., Battisti, D. S. & Blanchard-Wrigglesworth, E. Two-way teleconnections between the Southern Ocean and the Tropical Pacific via a dynamic feedback. *J. Clim.* **35**, 6267–6282 (2022).
53. Zhang, X., Deser, C. & Sun, L. Is there a tropical response to recent observed Southern Ocean cooling? *Geophys. Res. Lett.* **48**, e2020GL091235 (2021).
54. Kang, S. M., Yu, Y., Deser, C. & Ceppi, P. Global impacts of recent Southern Ocean cooling. *Proc. Natl. Acad. Sci. USA.* **120**, e2300881120 (2023).
55. Ma, X. et al. Western boundary currents regulated by interaction between ocean eddies and the atmosphere. *Nature* **535**, 533–537 (2016).
56. Small, R. J., Bryan, F. O., Bishop, S. P. & Tomas, R. A. Air-sea turbulent heat fluxes in climate models and observational analyses: what drives their variability? *J. Clim.* **32**, 2397–2421 (2019).
57. Lee, S.-K. et al. Human-induced changes in the global meridional overturning circulation are emerging from the Southern Ocean. *Commun. Earth Environ.* **4**, 69. https://doi.org/10.1038/s43247-023-00727-3 (2023).
58. Armour, K. C., Marshall, J., Scott, J. R., Donohoe, A. & Newsom, E. R. Southern Ocean warming delayed by circumpolar upwelling and equatorward transport. *Nat. Geosci.* **9**, 549–554 (2016).
59. Bilgen, S. I. & Kirtman, B. P. Impact of ocean model resolution on understanding the delayed warming of the Southern Ocean. *Environ. Res. Lett.* **15**, 114012 (2020).
60. Rackow, T. et al. Delayed Antarctic sea-ice decline in high-resolution climate change simulations. *Nat. Commun.* **13**, 637 (2022).
61. Zhang, L., Delworth, T. L., Cooke, W. & Yang, X. Natural variability of Southern Ocean convection as a driver of observed climate trends. *Nature Clim. Change* **9**, 59–65 (2019).
62. Kay, J. E. et al. The Community Earth System Model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. *B. Am. Meteorol. Soc.* **96**, 1333–1349 (2015).
63. Griffies, S. M. et al. OMIP contribution to CMIP6: experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geosci. Model Dev.* **9**, 3231–3296 (2016).
64. Meehl, G. A. et al. Effects of model resolution, physics, and coupling on Southern Hemisphere storm tracks in CESM1.3. *Geophys. Res. Lett.* **46**, 408–12,416 (2019).
65. Dennis, J. M. et al. CAM-SE: a scalable spectral element dynamical core for the community atmosphere model. *Int J High Perform Comput Appl* **26**, 74–89 (2012).
66. Danabasoglu, G. et al. The Community Earth System Model Version 2 (CESM2). *J. Adv. Model. Earth Syst.* **12**, e2019MS001916 (2020).
67. Tsujino, H. et al. JRA-55 based surface dataset for driving ocean—sea-ice models (JRA55-do). *Ocean Model.* **130**, 79–139 (2018).
68. Haarsma, R. J. et al. High resolution model intercomparison project (HighResMIP v1. 0) for CMIP6. *Geosci. Mod. Dev.* **9**, 4185–4208 (2016).
69. Goddard, L. et al. A verification framework for interannual-to-decadal predictions experiments. *Clim. Dyn.* **40**, 245–272 (2013).
70. Strommen, K. & Palmer, T. N. Signal and noise in regime systems: a hypothesis on the predictability of the North Atlantic Oscillation. *Q. J. R. Meteorol. Soc.* **145**, 147–163 (2019).
71. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **7**, 109 (2020).
72. Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.* **108**, 4407 (2003).
73. Adler, R. F. et al. The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeor.* **4**, 1147–1167 (2003).
74. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
75. Fetterer, F., Knowles, K., Meier, W. N., Savoie, M. & Windnagel, A. K. Sea Ice Index, Version 3. Distributed by National Snow and Ice Data Center, Boulder, Colorado, USA (accessed 2 June 2023); https://doi.org/10.7265/N5K072F8 (2017).

ACKNOWLEDGEMENTS

The HRDP experiment was completed by the International Laboratory for High Resolution Earth System Prediction (iHESP)—a collaboration between the Qingdao National Laboratory for Marine Science and Technology (QNLM), Texas A&M University (TAMU), and the National Center for Atmospheric Research (NCAR). NCAR is a major facility sponsored by the US National Science Foundation (NSF) under Cooperative Agreement 1852977. The bulk of this work was funded by iHESP. Additional support was provided by US Department of Commerce grant NA20OAR4310408, the US National Academies of Science and Engineering (NASEM) Gulf Research Program grant 2000013283, and US NSF grants AGS-1462127 and AGS-2231237. The HRDP simulations were performed on Frontera at the Texas Advanced Computing Center (TACC) at the University of Texas using project code ATM20005. Special thanks go to Dr Lixin Wu of Ocean University of China for his leadership of the iHESP collaboration. Jim Edwards provided assistance in optimizing CESM code on Frontera, and Alper Altuntas assisted in executing the HR-FOSI simulation.

AUTHOR CONTRIBUTIONS

The ideation of HRDP is attributed to iHESP principal investigators P.C. and G.D. who also led the funding acquisition. The creation of HRDP was jointly led by P.C., G.D., and S.G.Y., with computing resource acquisition led by P.C. Project management for the HRDP experiment was led by S.G.Y., with simulations conducted by N.R., Q.Z., F.S.C., A.G., and M.C.R. The conceptualization, analysis, and writing of this paper are attributed to S.G.Y. All authors contributed to reviewing and editing the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-023-00434-y.

Correspondence and requests for materials should be addressed to Stephen G. Yeager.