ARTICLE   OPEN

Check for updates

Seasonal predictability of baroclinic wave activity

Gan Zhang[1,2,6✉], Hiroyuki Murakami[2,3], William F. Cooke[2,3], Zhuo Wang[4], Liwei Jia[2,3], Feiyu Lu[1,2], Xiaosong Yang[2], Thomas L. Delworth[2], Andrew T. Wittenberg[2], Matthew J. Harrison[2], Mitchell Bushuk[2,3], Colleen McHugh[2,5], Nathaniel C. Johnson[2], Sarah B. Kapnick[2,7], Kai-Chih Tseng[1,2] and Liping Zhang[2,3]

Midlatitude baroclinic waves drive extratropical weather and climate variations, but their predictability beyond 2 weeks has been deemed low. Here we analyze a large ensemble of climate simulations forced by observed sea surface temperatures (SSTs) and demonstrate that seasonal variations of baroclinic wave activity (BWA) are potentially predictable. This potential seasonal predictability is denoted by robust BWA responses to SST forcings. To probe regional sources of the potential predictability, a regression analysis is applied to the SST-forced large ensemble simulations. By filtering out variability internal to the atmosphere and land, this analysis identifies both well-known and unfamiliar BWA responses to SST forcings across latitudes. Finally, we confirm the model-indicated predictability by showing that an operational seasonal prediction system can leverage some of the identified SST-BWA relationships to achieve skillful predictions of BWA. Our findings help to extend long-range predictions of the statistics of extratropical weather events and their impacts.

INTRODUCTION

Baroclinic waves in the atmosphere and their interaction with the mean circulation drive everyday weather in the extratropics. Well-developed baroclinic waves generate surface extratropical cyclones and upper-tropospheric perturbations. Some most extreme upper-tropospheric perturbations occur in the form of Rossby wave breaking (RWB)[1], which naturally characterizes the life cycle of baroclinic waves (LC1 and LC2[2,3]). Similar to the breaking of ocean surface gravity waves, RWB involves nonlinear wave deformation and drives irreversible transport of momentum and substances (e.g., moisture and ozone). When viewed on weather maps (Fig. 1), LC1-type RWB features an anticyclonic overturning of potential vorticity (PV) and a progressively narrowing upper-tropospheric trough, while LC2-type RWB involves a cyclonic overturning of PV and a broadening upper-tropospheric trough[2,3]. Anticyclonic LC1-type and cyclonic LC2-type RWB can occur separately or concurrently (Fig. 1). In the midlatitudes and adjacent subtropical/subpolar regions, LC1 and LC2 events contribute to various high-impact weather events, such as atmospheric blocking[4,5], polar extremes[6,7], sudden stratospheric warming[8], atmospheric rivers[9,10], and continental extreme precipitation[11]. Moreover, LC1 and LC2 events also interact with tropical cyclones and may affect the genesis, tracks, and intensity of a tropical cyclone[12–16]. Collectively, those studies suggest that baroclinic waves and associated high-impact weather events have substantial spatial-temporal variations and differ greatly between LC1 and LC2 events. Although the understanding of these connections is still evolving, such associations are evident in recent high-impact weather events and have attracted intense scientific and public interest (Fig. 1 and Supplementary Table 1 and Note 1).

While skillful long-range probabilistic predictions of tropical weather events (e.g., hurricane frequency[17]) are now routinely issued to assist decision-making, such predictions are largely elusive for extratropical weather events. The recent exploration of the subseasonal-to-seasonal predictability of extratropical atmosphere has led to a rapidly growing body of literature. The relatively high prediction skills reported so far are related to the large-scale climate modes[18–20] (e.g., the North Atlantic Oscillation) and/or the Eulerian statistics (e.g., storm tracks indicated by sea-level pressure variance)[21–23]. Although the predictions of large-scale climate modes and Eulerian statistics are useful, they do not fully characterize weather events experienced by societies and ecosystems. When weather events (e.g., extratropical cyclones) are explicitly considered, the reported prediction skills of event statistics often focus on specific seasons or regions and are generally low to moderate[24–26]. Perhaps surprisingly, the fundamental driver of extratropical weather events—baroclinic waves and related wave-circulation interaction—has received little attention in the existing studies of climate modeling and predictability. The lack of such investigation makes it difficult to address important questions such as whether any potentially predictable signals remain undiscovered.

In the weather and climate research communities, a widely shared notion is that long-range predictions (>2-week lead time) of individual baroclinic waves are not particularly promising. Baroclinic waves develop from potential energy conversion and are governed by the nonlinear dynamics of the midlatitude atmosphere. As illustrated by Lorenz[27], the evolution of a nonlinear dynamical system is sensitive to initial conditions, and this sensitivity limits the deterministic predictability of midlatitude weather to ~2 weeks[28]. In comparison, climate predictions of weather events are intrinsically probabilistic and mainly rely on slowly varying boundary forcings. These boundary forcings—such as sea surface temperatures (SSTs)—are predictable beyond two weeks and strongly regulate the tropical atmosphere[27]. Although such forcings do not strongly constrain the extratropical atmosphere[29,30], observational studies showed that predictable climate

---

[1]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA. [2]Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, Princeton, NJ, USA. [3]University Corporation for Atmospheric Research, Boulder, CO, USA. [4]Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [5]SAIC, Science Applications International Corporation, Reston, VA, USA. [6]Present address: Citadel Enterprise Americas, LLC, Chicago, IL, USA. [7]Present address: J.P. Morgan Chase & Co., New York, NY, USA. ✉email: zhganwing@gmail.com
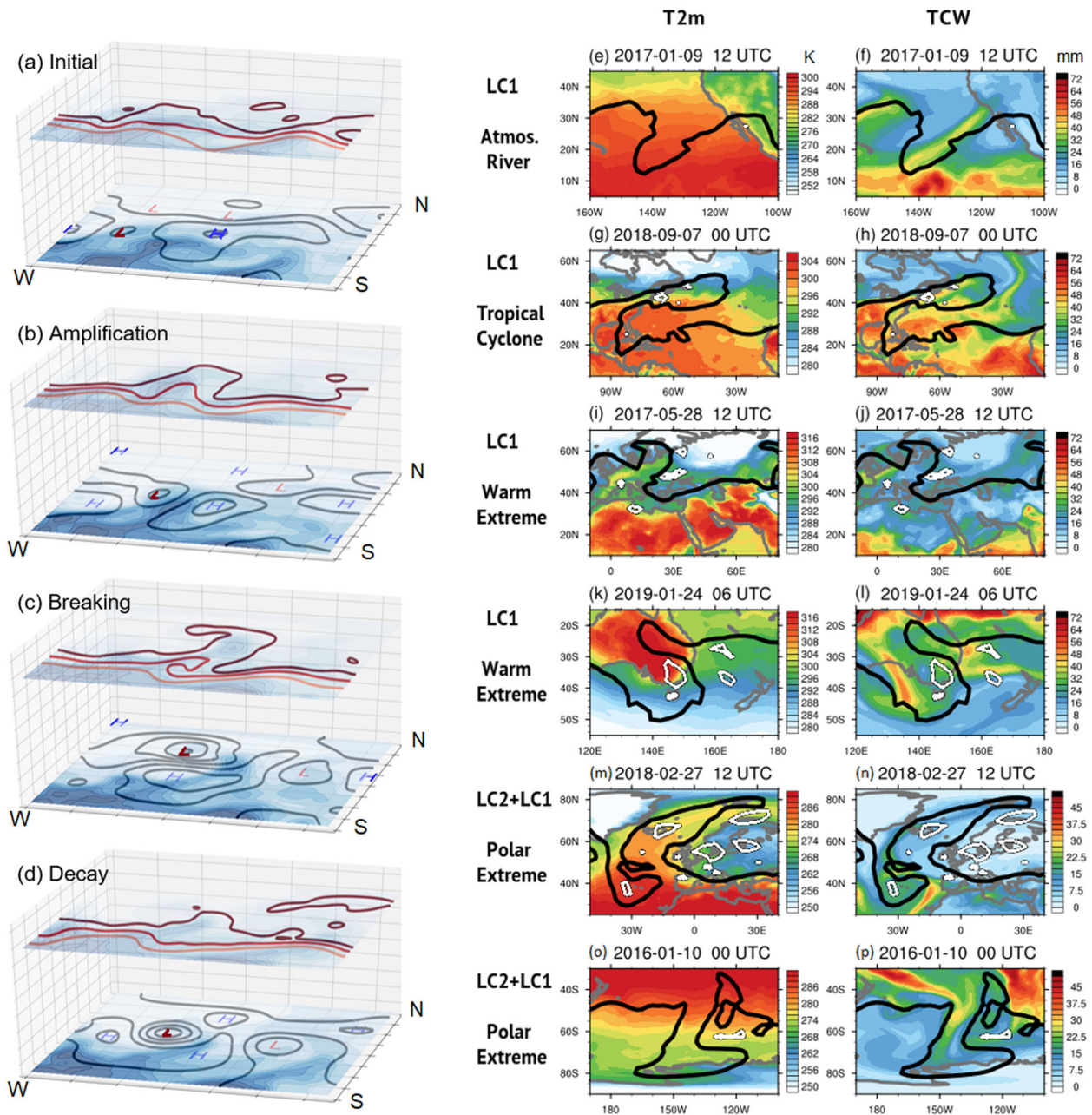
Fig. 1 Observed LC1- and LC2-type life cycle of baroclinic waves. The left column **a–d** shows the schematics derived from a real-world baroclinic wave. The upper-level contours show 350-K potential vorticity, the surface contours show sea-level pressure, and the blue shading shows column integrated water vapor. The wave is associated with a surface low (L) that develops (**a**, **b**) and moves poleward (**c**). The upper-level flow shows in LC2-type RWB in the Northwest and LC1-type RWB in the Southeast (**c**). This event contributed to an atmospheric river (dark blue plume) (**c**), a strong surface cyclone (**c**), and an upper-level blocking pattern (detached dark red feature) (**d**). The right two columns provide additional wave cases. Color shading shows 2-m air temperature (unit: K) and total column water (unit: mm), respectively. Black contours show 350-K potential vorticity (PV), while white contours denote small-scale PV features. The PV contours are ±1.5 PVU in subtropical cases and ±7 PVU in mid-/high-latitude cases. **e**, **f** LC1 event contributing to an atmospheric river landfall that ended a multi-year drought in California. **g**, **h** LC1 event steering weakened Hurricane Florence to the U.S. **i**, **j** LC1 event contributing to record-breaking heatwaves in Central Europe and Southwest Asia. **k**, **l** LC1 event contributing to a record-breaking heatwave in Southern Australia. **m**, **n** LC2 event (near 30°W) and LC1 event (near 15°E) contribute to extreme Arctic warming and a cold air outbreak in Europe. **o**, **p** LC2 event (near 160°W) and LC1 event (near 130°W) contribute to an Antarctic warm extreme event and ice melting.

modes (e.g., the El Niño–Southern Oscillation) modulate the distribution and occurrence probability of baroclinic waves[9,31–35]. Idealized baroclinic wave simulations also established that wave life cycles are sensitive to the background environment (e.g., wind shear[2,3] and stratospheric conditions[36]). As large-scale environmental factors are partly predictable beyond 2 weeks, the existing evidence thus implies long-range predictability of the statistics of

LC1 and LC2 events (hereafter baroclinic wave activity, BWA)[31]. However, it remains unclear under what circumstances such predictability exists and to what extent it can be capitalized by climate models in the form of useful predictions.

A long-standing obstacle to exploring the predictability of BWA is the weak responses of the extratropical atmosphere to climate forcings (e.g., SST forcings)[29,30]. This obstacle manifests as low

signal-to-noise (S/N) ratios of forced atmospheric responses (signals) and unforced atmospheric variability (noise). For statistical predictions based on empirical relations, a low S/N ratio makes it difficult to capture the physical and dynamical structures in observational data, especially when the observation records are limited. For dynamic predictions based on global climate models, a low S/N ratio necessitates the use of resource-intensive ensemble simulations to extract seasonally predictable signals in the chaotic climate system and to adequately sample the uncertainties associated with unforced climate variability. Yet current climate models sometimes suffer from artificially low S/N ratios of potentially predictable signals[19,37], so simulations with an unusually large ensemble size are likely essential for exploring the predictability of BWA.

An emerging tool for differentiating forced climate responses from unforced variability is large ensemble simulations with perturbed initial conditions[38]. Here, we explore the potential seasonal predictability of BWA using observational data (see "Observational data") and large ensemble simulations with prescribed SST forcings (see "SST-forced large ensemble simulations"). The prescribed SST forcings closely follow the historical SST observations but include small deviations that account for observational uncertainties and ocean mesoscale eddies. This prescribed-SST approach differs from the other contemporary large ensemble simulations that use free-running coupled climate models. At the cost of more realistic atmosphere-ocean coupling, the prescribed-SST approach has two advantages for studying SST-related predictability[29]. First, coupled models generate a substantial dispersion of oceanic states[38] that deviate from the historical observations within the first few simulation years. This dispersion makes it difficult to validate the free-running climate simulations against the observed variations of BWA. Second, the prescribed-SST approach greatly reduces the computational resources needed for ocean simulations and enables an exceptionally large ensemble of high-resolution atmospheric simulations. This feature is useful for simulating atmospheric dynamics and revealing SST-forced responses of BWA—even if their S/N ratios are low.

This study leverages these advantages of the aforementioned SST-forced simulations to probe the potential seasonal predictability of BWA. With a focus on the S/N ratio of the regional counts of LC1 and LC2 events, we estimate the potential BWA predictability when SSTs are assumed to be perfectly 'predicted' (i.e., prescribed with observed values). This estimate of the potential BWA predictability is conducted throughout the year and around the globe. Even though this examined climate model lacks the capability of delivering operational forecasts, we manage to confirm the predictability of BWA using an independent operational prediction system that can skillfully predict SSTs. Compared to previous operating systems developed by the U.S. National Oceanic and Atmospheric Administration (NOAA), this prediction system has greatly reduced oceanic biases[39] and simulated tropical–extratropical teleconnections with better fidelity[40]. Our analyses of this system suggest useful and sometimes exceptional seasonal predictability of BWA. The findings establish a scientific basis for using climate models to conduct long-range predictions of the statistics of extratropical weather events and to develop an event-based understanding of extratropical variability.

RESULTS

Large ensemble simulations indicate potential seasonal predictability

The large ensemble simulations (d4PDF) incorporate climate boundary forcings (e.g., SST and greenhouse gases) and produce realistic atmospheric circulation and weather events[41–44]. With respect to the distributions of baroclinic waves (see "Classification

of baroclinic wave life cycle"), a comparison between the observation (Fig. 2a–d) and model climatologies (contours in Fig. 2e–l) suggests that the simulation generates too few LC2 events at high latitudes (e.g., the North Atlantic during January–March). The LC2 underestimate appears consistent with the fact that climate models commonly underestimate the frequency of atmospheric blocking[45–47] (which are often LC2-related). Nonetheless, the seasonal cycle and spatial distribution of LC1 and LC2 events in the simulation are realistic overall. For example, LC1 events mostly occur at lower latitudes relative to LC2 events, especially near strong midlatitude westerlies. Additional composite analysis of individual events also suggests that key wave breaking features are well simulated (Supplementary Fig. 1 and Supplementary Note 2). The model fidelity establishes the foundation for the ensuing predictability analysis, which will use the frequency of LC1 and LC2 features to characterize BWA. In comparison with the commonly used Eulerian statistics[22,23], this approach helps to highlight dynamic features involved in weather events.

A necessary condition for skillful seasonal predictions is that forced responses of a predictant have high S/N ratios, especially in the real-world dynamical system[29]. We characterize the S/N ratio in our model simulation and the observation using the predictable component (PC) metric[19,20] (see "Predictability analysis"). The essence of this PC metric is the square root of the predictable fraction of the total variance. This metric is defined differently for ensemble simulations and observation. For a model system, PC_{mod} is approximated using the following expressions

$$PC_{mod} = \sqrt{\sigma_{sig}^2/\sigma_{tot}^2} \qquad (1)$$

where σ_{sig}^2 is the temporal variance of the model ensemble mean, and σ_{tot}^2 is the average of the temporal variance of individual ensemble members. For observations, PC_{obs} is approximated as the temporal Pearson correlation (r) between the observation and the ensemble mean. A comparison of PC_{obs} and PC_{mod} can distinguish whether a model is overconfident ($PC_{obs} < PC_{mod}$) or underconfident ($PC_{obs} > PC_{mod}$) in its predictions.

Before proceeding to the PC results, we make several clarifications that help with the interpretation. To reduce possible confusion between PC_{obs} and PC_{mod}, we refer to PC_{obs} mainly as the 'skill' of simulating or predicting year-to-year variations and reserve the term PC for PC_{mod}. Also, the ensuing analysis of these two metrics uses three-month periods starting in January, April, July, and October instead of the standard seasons. This choice is made based on a data availability issue related to System for Prediction and EArth System Research (SPEAR) (see "SPEAR seasonal prediction experiments"), and we have verified that this choice does not qualitatively affect the findings of PCs. Finally, the PC analysis of the SST-forced simulations seeks to quantify the fraction of BWA variance that can be explained by SST variability. Although SST variability can be skillfully (though not perfectly) predicted on the seasonal scale and tends to be the leading source of atmospheric seasonal predictability[29], the SST forcings here are not model predictions but prescribed observations. Therefore, PC_{mod} of this set of simulations only helps to infer the SST-related "potential predictability".

Figure 2e–l shows the PCs of the year-to-year variations of LC1 and LC2 frequency, respectively. High values of the PC metric appear in regions with BWA but do not follow the maxima of BWA closely. The PC metric of LC1 events shows high values in the subtropics throughout the annual cycle (Fig. 2e–h). These high values suggest year-to-year variations of these LC1 events are strongly regulated by SSTs and other prescribed climate forcings. For LC2 events, high PC values appear in the deep tropics (Fig. 2i–l), but these high PCs are related to PV features of simulated tropical convection instead of mid-latitude baroclinic waves. At higher latitudes, the PC values of LC2 events show a strong seasonal and regional
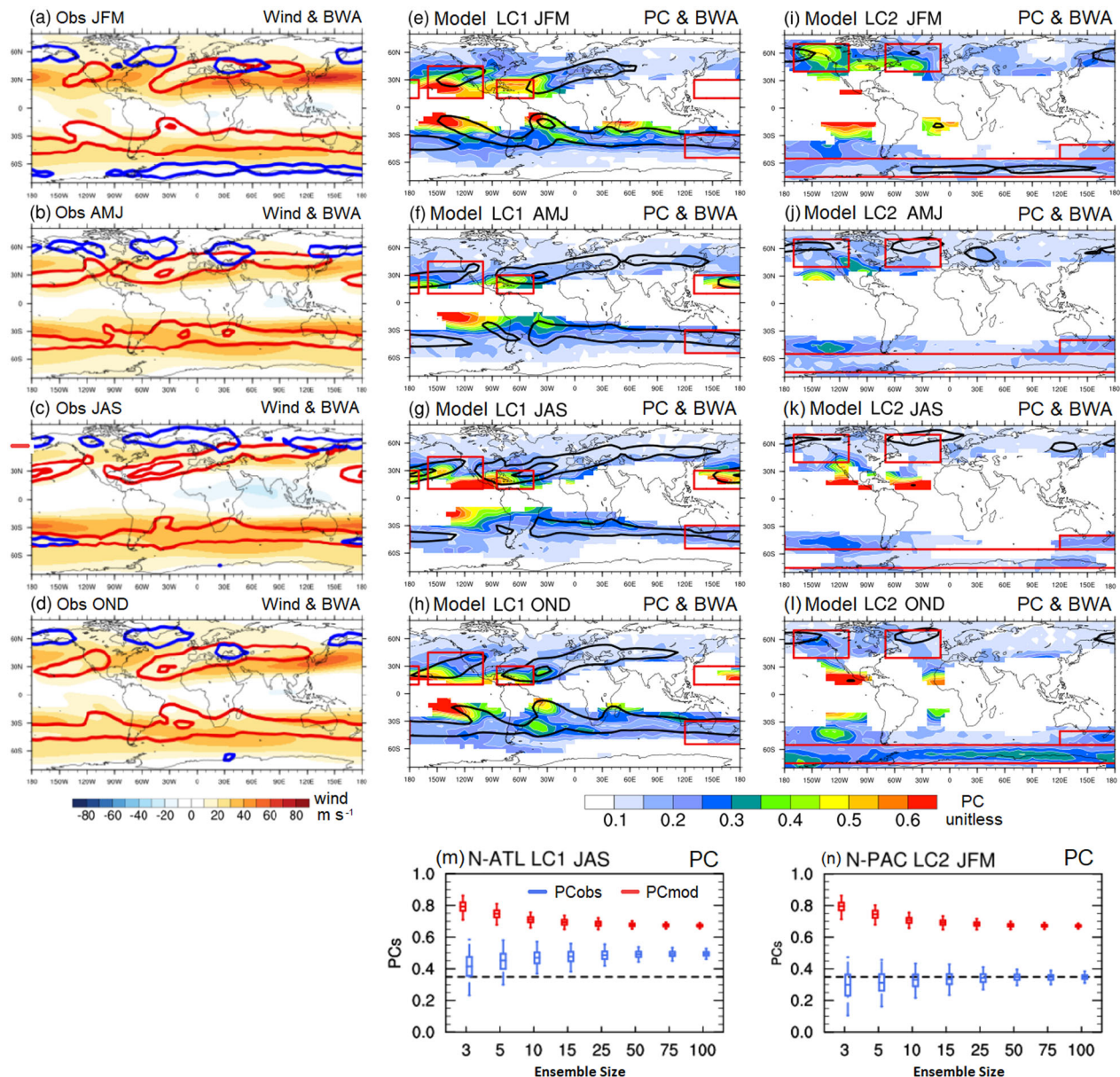
Fig. 2 Climatology and predictability metrics of baroclinic waves in d4PDF simulation. a–d Climatology of 200-hPa zonal wind (m s^{-1}; shading) and the observed spatial distributions of LC1 (red contour) and LC2 (blue contour) events during 1979–2010. The contoured distributions are BWA as represented by the counts of PV features in gridded boxes (see "Classification of baroclinic wave life cycle"). These counts are sums during **a** January–March, **b** April–June, **c** July–September, and **d** October–December. The contours levels are 4 and 10 events season^{-1}, and the latter only shows up with LC1 events as small and enclosed regions. **e–h** Climatology (black contours) and predictable components (PCs from Method **e**) (unitless; shading) of LC1 events as simulated by the SST-forced experiments. **i, l** Same as (**e–h**), but for LC2 events. The climatology of BWA in (**e–l**) shares the same contour levels as (**a–d**). The PC values are omitted from shading in regions with low BWA (<0.5 events season^{-1}). The red boxes indicate the individual regions examined separately later. **m** PCs of the July–September event count of N-ATL LC1 (10–30°N, 45–85°W) as a function of the ensemble size. Red shows the model PCs estimated with the SST-forced large ensemble, and blue shows the correlation skills (i.e., PC$_{obs}$) of the same simulation. The boxplots indicate 2.5th, 25th, 50th, 75th, and 97.5th percentiles of metrics, determined from the bootstrapping technique. The black dashed line indicates the 5% significance level of the correlation coefficient. **n** is the same as (**m**), but for the PCs of the January–March event count of N-PAC LC2 (40–70°N, 110–170°W).

dependence. While the PC values are generally low, a notable high-PC region appears during January–March extending from North America and the North Atlantic (Fig. 2i). Interestingly, this region partially overlaps with the regions where the Pacific-North America pattern and the North Atlantic Oscillation are active[33,48,49], suggesting potential predictability of related wave-circulation interactions.

Motivated by the high-impact events in Fig. 1 and their spatial-temporal distributions (Fig. 2a–d), we further analyze the potential

seasonal predictability of BWA in selected regions (red boxes in Fig. 2e–l). While the highest PC$_{mod}$ and best simulation skills are generally associated with subtropical LC1 events, the discussion here covers both warm-season LC1 events in a low-latitude region (N-ATL LC1) and cold-season LC2 events in a high-latitude region (N-PAC LC2). Our focus on these events prioritizes a balanced evaluation of LC1 and LC2 events that affect North America in different seasons. This choice also helps to show model strengths and weaknesses, as will be discussed later. Readers interested in
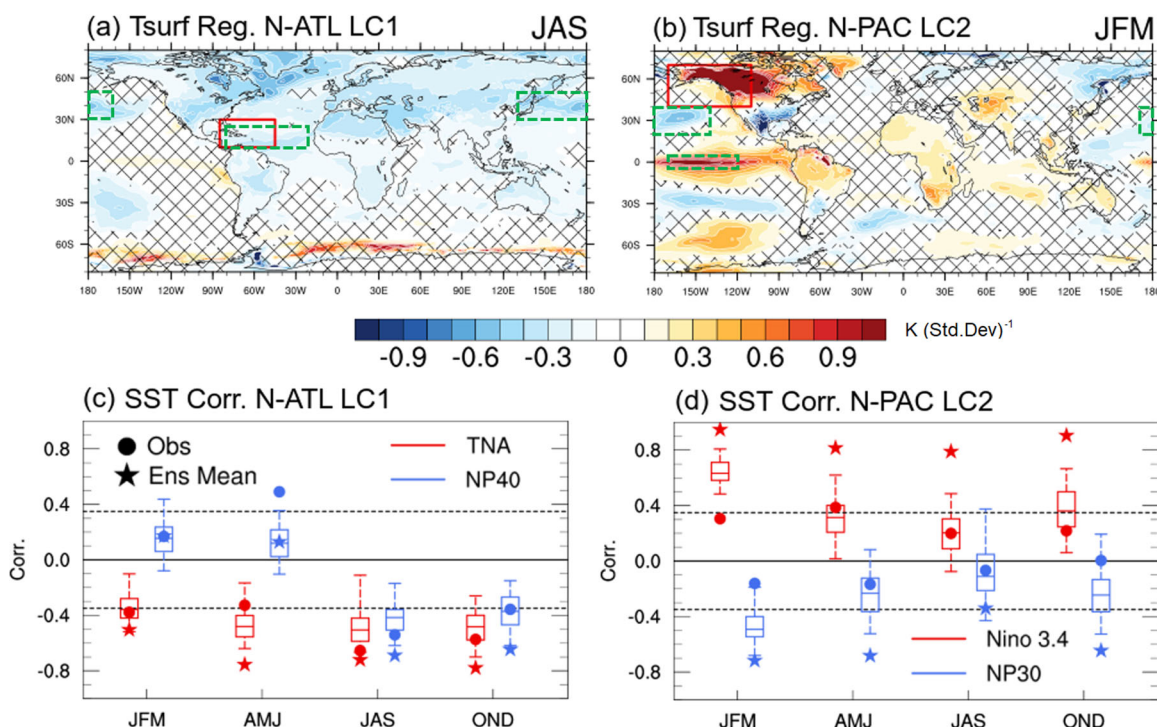
Fig. 3 Relationships between baroclinic waves and surface temperature in d4PDF simulation. a, b Linear regression (color shading) of ensemble-mean surface temperature (K) onto ensemble-mean BWA (standardized) during 1979–2010. The ensemble-mean BWA of **a** N-ATL LC1 and **b** N-PAC LC1 (red boxes) are event counts that are standardized based on year-to-year variations. Hatching in **a, b** indicates the linear relationship is below 95% confidence determined from a non-directional t-statistics. **c, d** Linear correlations between regional surface temperature and BWA. The surface temperature metrics are SSTs of the tropical North Atlantic (TNA; 10–25°N, 20–80°W) and the midlatitude North Pacific (NP40; 30–50°N, 130°E–160°W) for **c** N-ATL LC1 events; and the SST of Niño 3.4 region (120–170°W, 5°S–5°N) and the midlatitude North Pacific (NP30; 20°–40°N, 170°E–140°W) for N-PAC LC2 events. The boxplots indicate 2.5th, 25th, 50th, 75th, and 97.5th percentiles of correlations in individual ensemble members. Dots and stars show the correlations of the ERA-Interim reanalysis and the ensemble means, respectively. The black dotted line shows the 5% significance level determined from the t-statistics. The horizontal axis denotes results from January–March, April–June, July–September, and October–December.

the other individual regions can refer to Supplementary Notes 3 and 4 for additional results.

The PC estimates are sensitive to unforced noises within ensemble members[19,50], so we use bootstrapping to evaluate the metrics' sensitivity to the ensemble size (see "Predictability analysis"). Figure 2m, n shows PC_{mod} and the simulation skill (i.e., PC_{obs}) as a function of the ensemble size. A comparison of the estimated PC and prediction skill yields interesting insights. For example, the estimated PC and prediction skill show an opposite dependence on the ensemble size. This dependence can be explained as follows: an insufficient ensemble size allows more noise to enter the ensemble mean terms in PC_{mod} and PC_{obs}, resulting in an overestimate of the potential predictability (noise being misinterpreted as signals) and an underestimate of simulation skill (signals being contaminated by noise). This dependence suggests seasonal predictions of BWA need relatively large ensembles, which have been also shown helpful in the PC analyses[19] and seasonal predictions of other phenomena[20,21,51,52]. The ensemble size required to reliably extract the SST-forced signal varies across regions and types of baroclinic waves (Supplementary Figs. 2 and 3). In most of the examined circumstances, the skill dependence on the ensemble size does not weaken notably until the ensemble size increases to 15–25, indicating that an ensemble with at least 15 members is likely necessary for investigating BWA predictability.

Another interesting feature in Fig. 2m–n is that the simulation skill is overall below the potential predictability suggested by the PC values. This behavior is not limited to the N-ATL LC1 and N-PAC-LC2 (Supplementary Figs. 2 and 3). It suggests this climate model is generally overconfident in simulating year-to-year

variations of BWA, which contrasts with the underconfident behavior of climate models in predicting certain large-scale Eulerian means[19,37]. We speculate that this overconfident issue in our analyses could arise from model biases or missing physical processes. For example, prescribing SSTs distorts air-sea fluxes that enter the atmosphere and can affect the development of baroclinic waves (e.g., location and strength) in subtle ways. Moreover, the ensemble SSTs closely follow the observation (Supplementary Fig. 4) and their spread does not account for unforced variability of a coupled climate system. Such a small SST spread could pose an additional and unrealistic constraint on the variability of air-sea fluxes. All other conditions being equal, these issues likely lead to an under-dispersive[53] and somewhat biased[54] simulation of the atmosphere. Meanwhile, the climate model examined here––if run with the atmosphere-ocean coupling––could still suffer from a deficiency in simulating realistic S/N ratios of large-scale Eulerian means as other models[19,55]. Despite these limitations, the analysis here suggests that SST forcings can help to simulate realistic year-to-year variations of BWA.

Oceanic contributors to potential seasonal predictability

The year-to-year variations of SST forcings and BWA are overall realistic in the SST-forced large ensemble simulations (e.g., Supplementary Fig. 4). Since averaging over a large ensemble helps to filter out unforced variability, a simple linear regression between BWA metrics and SST forcings can offer insights into their connections (Fig. 3a, b). For N-ATL LC1 and N-PAC LC2 events, the model-based results from linear regression are consistent with

some observed relationships between BWA and SST forcings. For example, the N-ATL LC1 events occur more frequently with cold SSTs in the tropical North Atlantic[35] and correspond to the cold phase of the Atlantic Multidecadal Variability[16]. Meanwhile, the N-PAC LC2 events occur more frequently during the El Niño years[9,31,33]. While these BWA-SST connections and teleconnection mechanisms have been established by earlier observational studies, the analyses here (and in the next subsection) present evidence suggesting that climate models can simulate these connections with some fidelity.

The linear regression also suggests interesting but less familiar connections. For example, the N-ATL LC1 events are negatively correlated with the global skin temperature in July–December, suggesting a potential connection between climate warming and the N-ATL LC1 events. The regression map also suggests that the N-ATL LC1 events are negatively correlated with SST near the oceanic frontal zone in the midlatitude North Pacific with a high statistical significance ($p < 0.05$). This appears consistent with observational evidence that suggests the North Pacific weather patterns modulate the N-ATL LC1 events[35]. Although currently, we cannot rule out that this significant correlation might arise from co-existing tropical SST forcings (e.g., the Indian Ocean and the Western Pacific; Fig. 3a), recent studies with high-resolution atmospheric models do suggest that SST variability at the oceanic front of the Kuroshio Current can influence the downstream extratropical atmosphere[56,57], which has implications for both North America and Europe. Interestingly, the existing studies focused on the wintertime and mechanisms involving changes in the atmospheric baroclinic zone[56] or flux exchange from oceanic mesoscale eddies[57]. However, whether and how such mechanisms might operate in the summertime received little attention. Such unfamiliar relationships are abundant (Supplementary Figs. 5 and 6) and involve factors including the mid-latitude SSTs and high-latitude land temperatures. They present opportunities for future modeling and mechanism research, which will likely complement existing studies on the relationship between baroclinic waves and large-scale oceanic modes[33,35,58,59].

To validate the above SST-BWA relationships, Fig. 3c, d further examines these selected linear relationships throughout the year. Despite the strong seasonality of these relationships, the signs of the observed relationships were correctly represented by the large ensemble simulations. In thirteen out of sixteen boxplots (Fig. 3c, d), the observed correlation coefficients fall within the 95% confidence range of the large ensemble simulations. This consistency between the SST-forced simulations and the observation suggests that the impacts of SSTs on N-ATL LC1 and N-PAC LC2 events are generally credible. Meanwhile, two of the out-range cases happen with the N-PAC LC2 events during JFM, with the model overestimating the strength of correlations. Relationship distortions like this could arise from model deficiency or the experiment design (e.g., prescribing SST) (see "SST-forced large ensemble simulations"). This issue and low S/N ratios (as suggested by PC_{mod}) likely make it more challenging to simulate and predict the seasonal counts of LC2 events (Supplementary Fig. 3).

Interestingly, Fig. 3c, d also suggests that averaging over a large ensemble might clarify the connections between SST forcings and atmospheric responses. The correlations of ensemble means are generally stronger than those in individual ensemble members. Since some SST forcings and large-scale atmospheric responses can be skillfully predicted on the seasonal scale[60], these connections suggest that BWA may be predictable in some regions and seasons. Nonetheless, one should recognize that the strong correlations exhibited by the ensemble means of SST-forced simulations do not necessarily imply high skill of real-world predictions, in which the SSTs need to be predicted rather than 'prescribed'. Such predictions are far from perfect since observed

quantities inevitably contain unforced variability. Moreover, prediction models contain various biases and errors.

Prediction skill in an operational prediction system

Encouraged by the findings from the SST-forced large ensemble, we continue to explore whether skillful predictions of BWA are feasible in real-world operation. It would be ideal to proceed with a prediction system built on the identical atmospheric model. However, such a prediction system is currently unavailable. For individual climate modeling centers, the resource and expertise needed to run both large-ensemble simulations and operational seasonal predictions are still prohibitively high. Therefore, we instead explore the real-world prediction skill using the SPEAR[39,40], which simulates a realistic climatology and skillfully predict variations of the ocean and atmosphere[39,40]. The SPEAR can serve as an independent test of the potential seasonal predictability of BWA. More details on SPEAR and its performance are available in "SPEAR seasonal prediction experiments" and references therein. Due to computational resource constraints, initial retrospective prediction experiments with SPEAR use a medium ensemble size ($N = 15$). An ensemble size of this magnitude likely can extract some SST-forced responses of BWA (Fig. 2). The ensemble experiments cover a recent period (1995–2018), which partially overlaps with the SST-forced large ensemble simulations. The following analysis focuses on the prediction of the first three months after model initializations, for which the SST predictions are skillful in the tropics (Supplementary Fig. 7) due to the large thermal inertia of the ocean and the relatively realistic model representation of physical processes[39,40].

Besides producing relatively realistic climatologies (Fig. 4a–h) and dynamic structures (Supplementary Fig. 1) of LC1 and LC2 events, the PCs of SPEAR also show the potential seasonal predictability of BWA (Fig. 4a–h). The PC patterns in SPEAR are overall similar to those of the SST-forced large ensemble (e.g., relatively high PC values of subtropical LC1 events). Although SPEAR might appear to have higher PC estimates (cf. Figs. 2 and 4), this PC difference should be interpreted with caution. A large portion of this difference can be attributed to the PC estimates' dependence on ensemble size (Fig. 2m, n) and the non-stationarity of the climate system's predictability[61,62]. After the differences in ensemble size and analysis period are controlled, the PC differences between SPEAR and the SST-forced ensemble reduce substantially (Supplementary Figs. 8 and 9). The remaining differences can be partly attributed to SPEAR's observation-based initial conditions of land/atmosphere components, which help to constrain the Month 1–3 predictions. For the Months 1–3 period, we also speculate that SPEAR's atmosphere-ocean coupling (see "SPEAR seasonal prediction experiments") provides useful constraints on ensemble spreads. For example, atmosphere-ocean coupling prevents the unbounded air-sea heat fluxes in SST-forced simulations and might reduce the error growth associated with atmospheric eddies (e.g., near oceanic western boundary currents). Although these model differences complicate the inter-model comparison, the basic comparison here helps to qualitatively probe the robustness and uncertainties of our findings of the potential seasonal predictability of BWA.

Figure 4i, j shows that SPEAR's high PC values can translate to skillful predictions of BWA. For example, the predictions of N-ATL LC1 events are skillful throughout the year, including the Atlantic hurricane season; and the predictions of N-PAC LC2 events are highly skillful for January–March when the LC2-related weather patterns can strongly affect Alaskan rainfall and North American temperature. Additionally, highly skillful predictions are available for LC1 events over the subtropical Pacific (Supplementary Figs. 10 and 11), which are important processes for activities of atmospheric rivers[10] and tropical cyclones[16]. LC1 and LC2 events are also linked to wintertime atmospheric blocking[4,5], which can be
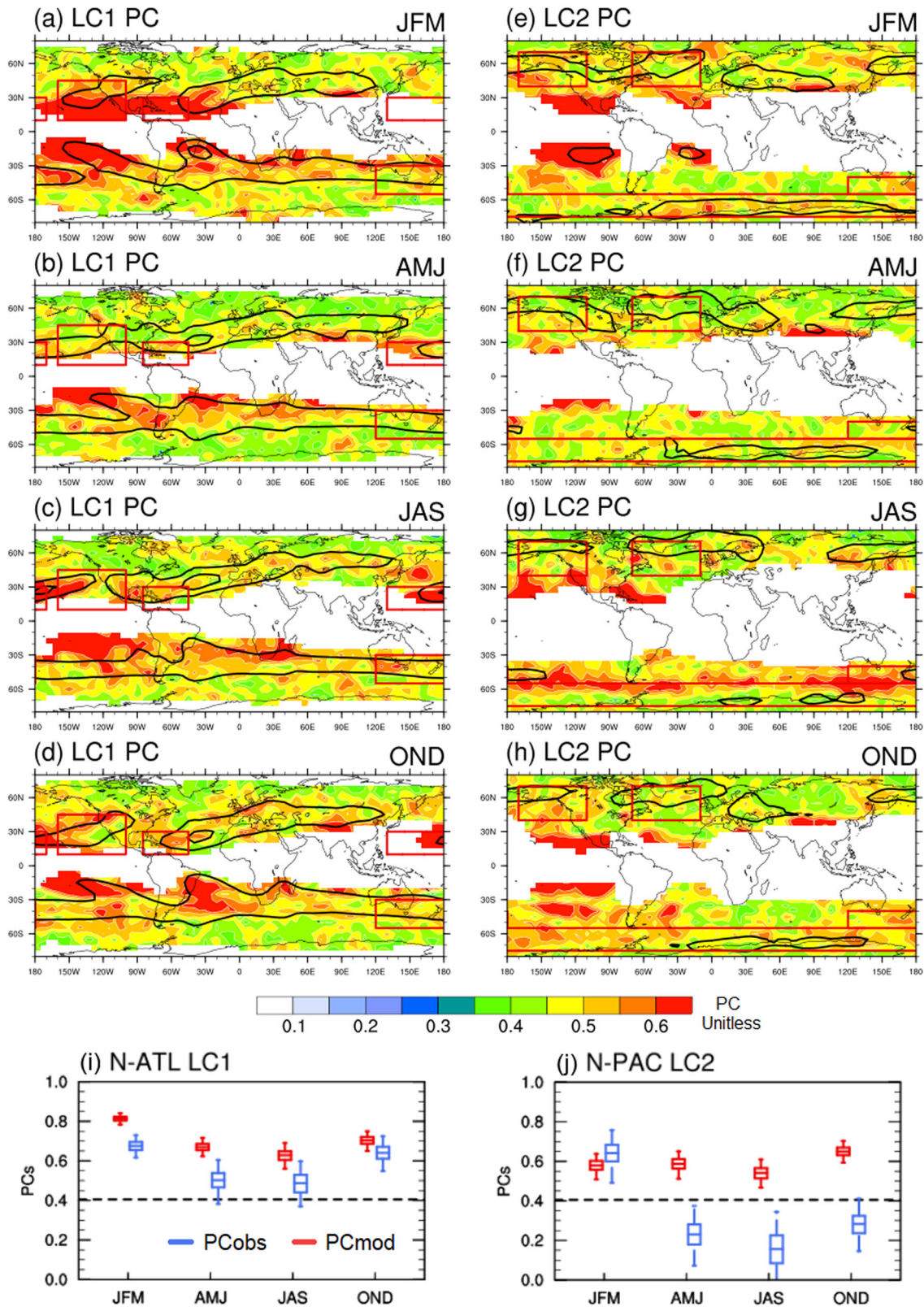
Fig. 4 Predictable components (PCs) and prediction skills of baroclinic waves in SPEAR prediction. a–d PCs of LC1 events in three-month predictions that are initialized on the 1st of January, April, July, and October of 1995–2018. **e–h** Same as (**a–d**), but for LC2 events. **i** PCs (red) and prediction correlation skills (blue) of the N-ATL LC1 events for the four above initialization months. **j** Same as (**i**), but for N-PAC LC2 events. The other plot settings are the same as in Fig. 2.

interpreted as successive or persistent wave-breaking events. SPEAR's prediction skills for LC1 and LC2 events appear consistent with a current European system's prediction skill for wintertime blocking events[63], both showing relatively high skills at lower latitudes and relatively poor skills at higher latitudes. We also note that SPEAR's prediction skills for BWA averaged over large regions are generally higher than the skills for small regions. This might be because the S/N ratio is lower on smaller scales (cf. PC values), and small-scale signals are also more sensitive to model biases (e.g., displaced midlatitude jets).

Lastly, we briefly discuss factors that contribute to SPEAR's skills in predicting BWA. Besides skillfully predicting SSTs in Months 1–3 (Supplementary Fig. 7), SPEAR is also skillful in predicting the 500-hPa geopotential height (Z500) in the tropics and some mid-/high-latitude regions (Supplementary Fig. 12). SPEAR's PC_{mod} and correlation skills of Z500 also show a latitudinal dependence, which is overall consistent with other prediction systems[64]. This latitudinal dependence is similar to BWA, suggesting that SPEAR's prediction skills of LC1 events partly arise from the low-latitude atmospheric circulation. However, the prediction skills of Z500 and BWA are not always aligned. For example, the prediction skill of January–March Z500 is close to zero near the Gulf of Alaska (Supplementary Fig. 12), while the predictions for N-PAC LC2 events are moderately skillful (Fig. 4j). Notably, another state-of-the-art model shows little skills in predicting wintertime atmospheric blocking in the roughly same region[63]. SPEAR's skill in predicting BWA appears to mostly arise from its prediction of tropical SSTs (esp. the ENSO; Supplementary Figs. 13 and 14). The relationship between BWA, SSTs, and the large-scale circulation warrants future investigation (Supplementary Note 5). Lastly, SPEAR's prediction skills for Months 1–3 likely have benefited from the land-atmosphere initial conditions. Preliminary analyses suggest that SPEAR's skills exist in longer-lead predictions but degrade relatively fast for LC2 events, indicating the skills' sensitivity to initial conditions.

DISCUSSION

Despite a common expectation of relatively low skill of climate models in long-range predictions of extratropical weather statistics[25,29,30], this study demonstrates that promising seasonal predictability of event-based BWA may exist in certain regions and seasons—and that such predictability can be (partly) capitalized with a state-of-the-art prediction system. Though the discussion in the main text focuses on BWA in two regions, seasonal predictability and skillful predictions are present in various seasons and regions, especially for LC1 events that are important for high-impact weather at the tropical–extratropical interface. Further work is needed to better quantify the prediction skill with different time leads and in a non-stationary climate system.

Meanwhile, the climate models examined here suggest some potential gaps between potential predictability (PC_{mod}) and actual prediction skills (PC_{obs}). Understanding these potential gaps and their underlying physical processes should be priorities for future research. For example, the model skills for simulating and predicting LC2 events are notably poorer than LC1 events, but PC_{mod} hints at a possibility that the potential predictability of both LC1 and LC2 events has not been fully capitalized. Existing studies suggest these events (and associated blocking events) involve baroclinic development[2] and contributions from diabatic processes[65,66]. Future process-based analyses built on this study might provide additional insights. The prediction of these events might also be improved via a continuing commitment to improving climate models (e.g., refining model initialization, resolution, and physics[63]).

Our analysis also highlights the benefits of using a large ensemble to filter out unforced variability in climate simulations.

This helps to identify some unfamiliar relations between extra-tropical SSTs and BWA, which point to opportunities for future research. The findings suggest that an inadequate ensemble size tends to overestimate PCs and underestimate prediction skill (e.g., Fig. 2m, n), which might hinder the exploration of extratropical predictability. The need for a large ensemble is more pressing if a climate prediction system underestimates the S/N ratio[19,20]. For example, recent studies suggest an underestimation of the S/N ratio might be common among climate models, at least for a key climate mode (i.e., the North Atlantic Oscillation)[37,55].

Finally, the findings of the BWA predictability have important implications for predicting statistical aspects of extratropical weather events beyond the 2-week timescale. Our analyses support the use of climate models in studying prediction problems of this nature. By delineating the seasons and regions with high S/N ratios of BWA, our results provide key information that guides future explorations of predicting extratropical weather events and their impacts. Successful examples that involve the prediction of atmospheric rivers and heat waves will be provided in upcoming studies. We hope that our findings will motivate improvements in extratropical predictions and applications of climate models.

METHODS

Observational data

The primary use of observational data in this study is simulation verification. We use monthly and 6-hourly data from the ERA-Interim reanalysis produced by European Center for Medium-Range Weather Forecasts[67]. To facilitate data processing and model verification, we coarsen the 6-hourly data to a 2.5° horizontal grid. The coarsened data is fed to an automatic detection algorithm to identify LC1 and LC2 events (see "Classification of baroclinic wave life cycle"). We analyze the data from 1979–2018 and examine its subsets when model simulations cover only part of this period. Another observational dataset is the Hadley Center Global Sea Ice and SST (HadISST v 1.1)[68]. The HadISST dataset is used to verify the relationship between SST forcings and the BWA simulated by models. Several other observational datasets were also utilized in the model simulations and will be briefly introduced in the following sections.

SST-forced large ensemble simulations

The SST-forced large ensemble simulations (d4PDF)[69] are conducted using Meteorological Research Institute (MRI) atmospheric general circulation model (AGCM) version 3.2H[70], which has ~60 km horizontal grid spacing and includes 64 vertical levels. This study uses the historical subset of the large ensemble simulations. In the historical simulation, the AGCM is integrated from 1951 to 2010 with the observed forcings, including anthropogenic forcings (e.g., aerosols and greenhouse gases) and time-varying monthly SST and sea ice information from the Centennial in-situ Observation Based Estimates of SST version 2 dataset (COBE-SST2[71]). The SSTs in COBE-SST2 and HadISST are mostly consistent despite some differences (Supplementary Note 4 and Fig. 4). To generate a large number of ensemble simulations, the initial conditions of the climate system were randomly perturbed when the model is initialized around 1951. Furthermore, the SST forcings include random but autocorrelated perturbations of up to 30% of the observed interannual variability throughout the integration[69]. The magnitudes of these perturbations vary across regions (Supplementary Fig. 4) and intend to account for observational uncertainties. The simulation consists of 100 ensemble members, greater than most existing large ensemble simulations[38]. Other model settings are documented in detail by Mizuta et al.[69].

Besides the prescribed SSTs and unusually large ensemble size, our choice of d4PDF also considers its relatively high spatial resolution and high-frequency model outputs that are uncommon for large ensemble simulations. A lack of realistic atmosphere-ocean coupling prevents this SST-forced large ensemble simulations from fully sampling unforced variability in the coupled climate system. Nonetheless, this SST-forced approach helps to isolate the atmospheric response to SST forcings via the ensemble mean, while deviations from the ensemble means can serve as a proxy for the unforced variability of the atmosphere[38]. With the assumption that this simulation reasonably characterizes the variability of historical climate, the observations should be approximately equivalent

to a climate realization of the large ensemble simulations. Therefore, the observed quantities and relationships should generally fall within the envelope of the ensemble members (as illustrated by Fig. 3c, d and Supplementary Fig. 4).

SPEAR seasonal prediction experiments

The Seamless SPEAR is the next-generation coupled modeling system developed by Geophysical Fluid Dynamics Laboratory (GFDL) for seasonal to multi-decadal prediction and projection. The atmosphere and land components are identical to those in the GFDL AM4-LM4 model[72], while the ocean and sea ice components are the MOM6 and SIS2 models, respectively[73]. SPEAR is closely related to the CM4 coupled GCM, GFDL's contribution to Coupled Model Intercomparison Project phase 6 (CMIP6), although different resolutions are used in the various model components. The present study uses the medium-resolution SPEAR version (SPEAR_MED) that has an atmospheric and land horizontal grid spacing of ~50 km, with 33 atmospheric levels. The ocean and sea ice components use a tripolar grid with a nominal horizontal grid spacing of 1° with meridional refinement to 1/3° in the tropics. The SPEAR components and the performance of its control, historical, and scenario experiments are documented in detail by Delworth et al.[40].

The SPEAR seasonal predictions are initialized from a combination of ocean data assimilation and atmospheric nudging experiments. Besides helping to initialize the model atmosphere, the atmospheric nudging also provides observational constraints for land and sea ice via the land/ice–atmosphere coupling. We analyze the SPEAR retrospective predictions initialized on the first day of January, April, July, and October during 1995–2018. Each prediction consists of 15 ensemble members and covers 12 months after model initialization. The retrospective prediction experiments are being extended in preparation for routine forecast implementation at the U.S. NOAA. Preliminary analysis of the extended experiments suggests the model performance is consistent with the experiments analyzed in the present study. The climatological SST bias of the SPEAR predictions is small, due to improved model physics and newly adopted oceanic bias corrections[39]. The bias corrections were conducted online during model integration by introducing three-dimensional and annually varying terms of ocean tendency adjustment (OTA). These terms were estimated by ocean data assimilation experiments in a coupled model with a free-running atmosphere. The OTA helps to substantially reduce SPEAR's SST biases relative to the previous GFDL climate models. Moreover, SPEAR's atmosphere resolution is relatively high compared to the current global climate models (50-km vs. 100–200 km grid spacing). The relatively low model biases and high atmosphere resolution are beneficial for dynamical simulations of weather events. Details of the initialization and performance of the SPEAR seasonal predictions are documented in Lu et al.[39].

Classification of baroclinic wave life cycle

The classification of baroclinic wave life cycles emphasizes RWB features, which correspond to the mature and nonlinear life stage of baroclinic waves. During the wave development, RWB features are dynamically coupled with near-surface features (e.g., extratropical cyclones and frontal systems) and are characterized by meridional reversals of PV contours[2]. The meridional overturning is associated with air masses that rotate anticyclonically for LC1 events or cyclonically for LC2 events[2]. We use the RWB identification algorithm described by Strong and Magnusdottir[49] but implement the code in the open-source Python language using a more efficient numeric search method. The updates help to process an extremely large volume of simulation data (>100 TB).

We analyze the PV on the 350-K isentropic surface that approximately corresponds to the upper troposphere and/or the lower stratosphere (depending on the latitude), where Rossby wave activity is relatively strong. The choice of analyzing the 350-K PV is consistent with earlier studies[10,13,49,58], and we acknowledge that RWB distributions depend on vertical levels and vary across isentropic surfaces. The search for RWB is conducted on the PV contours ranging from 1 to 30 PVU for the Northern Hemisphere (or −1 to −30 PVU for the Southern Hemisphere) at 0.5-PVU intervals (1 PVU = 10^{-6} K kg^{-1} m^2 s^{-1}). When features from adjacent PV contours appear in each other's vicinity, the search algorithm assumes that the features are related to the same breaking wave and only retains the feature with the largest spatial extent. The algorithm also discards small and isolated PV features, such as those denoted with white contours in Fig. 1. While these variable and parameter choices likely affect some

aspects (e.g., climatology numbers) of the tracked LC1 and LC2 features, we expect their year-to-year variations to be less sensitive to the algorithm design.

Despite the noisy nature of the observational data, the composites of identified RWB (Supplementary Fig. 1) resemble the RWB features (e.g., high-PV and low-PV tongues) described in previous studies of baroclinic wave life cycle[2,3]. Although BWA can be characterized using Eulerian-based metrics, this study uses the frequency of LC1 and LC2 events to focus on the dynamical features that often drive weather extremes. We count the centroids of these PV features (which correspond to the coordinate centers in Supplementary Fig. 1) in 5° grid boxes within 3-month periods. To reduce the counting complexity, PV features in adjacent time steps are treated as individual features. The analyzed PV outputs of the d4PDF large ensemble are 12-hourly, while the outputs of ERA-Interim and SPEAR are 6-hourly. For the consistency of results, the counts of LC1 and LC2 events in ERA-Interim and SPEAR are scaled to a 12-hourly basis. The accumulated three-month counts can be considered as the histogram of LC1-type and LC2-type PV features in the latitude-longitude space. We use this output to analyze the spatial-temporal distribution of LC1 and LC2 features.

Predictability analysis

We estimate the potential seasonal predictability of BWA using the PC[19]. The main text has provided a brief introduction of PC$_{obs}$ and PC$_{mod}$. Interested readers can refer to Eade et al.[19] for a detailed introduction of the PC metrics. The PC calculation in this study uses the three-month frequency of LC1 and LC2 events at latitude-longitude grid points (e.g., Figs. 2e–l) or within specific regions (e.g., Fig. 2m, n). The prediction skill estimates and the PC calculation for each season are conducted using the model outputs that are resampled with replacement ("bootstrapping"). Three-month simulations of BWA are randomly drawn from available ensemble simulations of each year. Repeating the drawing N times for every year helps to generate artificial prediction ensembles with N members. This N value is generally equal to the available ensemble size unless otherwise specified (e.g., Fig. 2m, n). In all the analyses, we repeat this bootstrapping procedure 1000 times to better characterize uncertainties related to the unforced noise in the ensemble simulations. In the predictability estimates, the resampling helps to quantify uncertainties related to the unforced variability in the dynamical system. This resampling, however, does not address the sampling uncertainty that can affect PC calculations in a non-stationary climate system[64].

DATA AVAILABILITY

The d4PDF simulation data that support the findings of this study are publicly available on the Data Integration and Analysis System website (http://www.diasjp.net/en/). Restrictions apply to the availability of certain SPEAR experiment data, which are under U.S. government regulation. A subset of SPEAR data has been made available via the North American Multi-Model Ensemble Project (https://www.cpc.ncep.noaa.gov/products/NMME/). The BWA data are available from the authors upon reasonable request and with permission of NOAA.

CODE AVAILABILITY

The RWB identification code was adapted from the code provided by Dr. Gudrun Magnusdottir; the code can be made available upon reasonable request with permission from Dr. Magnusdottir. The other analysis code is available from G.Z. upon reasonable request and with permission of NOAA.

REFERENCES

1. Appenzeller, C. & Davies, H. C. Structure of stratospheric intrusions into the troposphere. *Nature* **358**, 570–572 (1992).
2. Thorncroft, C. D., Hoskins, B. J. & McIntyre, M. E. Two paradigms of baroclinic-wave life-cycle behaviour. *Q. J. R. Meteorol. Soc.* **119**, 17–55 (1993).
3. Polvani, L. M. & Esler, J. G. Transport and mixing of chemical air masses in idealized baroclinic life cycles. *J. Geophys. Res.* **112**, D23102 (2007).
4. Tyrlis, E. & Hoskins, B. J. The morphology of northern hemisphere blocking. *J. Atmos. Sci.* **65**, 1653–1665 (2008).

5. Masato, G., Hoskins, B. J. & Woollings, T. J. Wave-breaking characteristics of midlatitude blocking. *Q. J. R. Meteorol. Soc.* **138**, 1285–1296 (2012).

6. Liu, C. & Barnes, E. A. Extreme moisture transport into the Arctic linked to Rossby wave breaking: extreme moisture flux and wave breaking. *J. Geophys. Res.* **120**, 3774–3788 (2015).

7. Wang, Z., Walsh, J., Szymborski, S. & Peng, M. Rapid Arctic sea ice loss on the synoptic time scale and related atmospheric circulation anomalies. *J. Clim.* **33**, 1597–1617 (2020).

8. Lee, S. H., Charlton-Perez, A. J., Furtado, J. C. & Woolnough, S. J. Abrupt stratospheric vortex weakening associated with North Atlantic anticyclonic wave breaking. *J. Geophys. Res. Atmos.* **124**, 8563–8575 (2019).

9. Ryoo, J.-M. et al. Impact of Rossby wave breaking on U.S. West Coast winter precipitation during ENSO events. *J. Clim.* **26**, 6360–6382 (2013).

10. Hu, H. et al. Linking atmospheric river hydrological impacts on the U.S. West Coast to Rossby wave breaking. *J. Clim.* **30**, 3381–3399 (2017).

11. Moore, B. J., Keyser, D. & Bosart, L. F. Linkages between extreme precipitation events in the Central and Eastern United States and Rossby wave breaking. *Mon. Weather Rev.* **147**, 3327–3349 (2019).

12. Davis, C. A. & Bosart, L. F. The TT problem: forecasting the tropical transition of cyclones. *Bull. Am. Meteorol. Soc.* **85**, 1657–1662 (2004).

13. Zhang, G., Wang, Z., Peng, M. S. & Magnusdottir, G. Characteristics and impacts of extratropical Rossby wave breaking during the Atlantic Hurricane Season. *J. Clim.* **30**, 2363–2379 (2017).

14. Manganello, J. V., Cash, B. A., Swenson, E. T. & Kinter, J. L. III Assessment of climatology and predictability of mid-Atlantic tropical cyclone landfalls in a high-atmospheric-resolution seasonal prediction system. *Mon. Weather Rev.* **147**, 2901–2917 (2019).

15. Zhang, G., Wang, Z., Dunkerton, T. J., Peng, M. S. & Magnusdottir, G. Extratropical impacts on Atlantic tropical cyclone activity. *J. Atmos. Sci.* **73**, 1401–1418 (2016).

16. Wang, Z., Zhang, G., Dunkerton, T. J. & Jin, F.-F. Summertime stationary waves integrate tropical and extratropical impacts on tropical cyclone activity. *Proc. Natl Acad. Sci. USA* **117**, 22720–22726 (2020).

17. Murakami, H., Levin, E., Delworth, T. L., Gudgel, R. & Hsu, P.-C. Dominant effect of relative tropical Atlantic warming on major hurricane occurrence. *Science* **362**, 794–799 (2018).

18. Barnston, A. G., Tippett, M. K., Ranganathan, M. & L'Heureux, M. L. Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Clim. Dyn.* **53**, 7215–7234 (2019).

19. Eade, R. et al. Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.* **41**, 5620–5628 (2014).

20. Dunstone, N. et al. Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nat. Geosci.* **9**, 809–814 (2016).

21. Scaife, A. A. et al. Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.* **41**, 2514–2519 (2014).

22. Yang, X. et al. Seasonal predictability of extratropical storm tracks in GFDL's high-resolution climate prediction model. *J. Clim.* **28**, 3592–3611 (2015).

23. Zheng, C., Chang, E. K., Kim, H., Zhang, M. & Wang, W. Subseasonal to seasonal prediction of wintertime Northern Hemisphere extratropical cyclone activity by S2S and NMME models. *J. Geophys. Res. Atmos.* **124**, 12057–12077 (2019).

24. Athanasiadis, P. J. et al. The representation of atmospheric blocking and the associated low-frequency variability in two Seasonal prediction systems. *J. Clim.* **27**, 9082–9100 (2014).

25. Befort, D. J. et al. Seasonal forecast skill for extratropical cyclones and windstorms. *Q. J. R. Meteorol. Soc.* **145**, 92–104 (2018).

26. DeFlorio, M. J. et al. Experimental subseasonal-to-seasonal (S2S) forecasting of atmospheric rivers over the Western United States. *J. Geophys. Res. Atmos.* **124**, 11242–11265 (2019).

27. Lorenz, E. N. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963).

28. Zhang, F. et al. What is the predictability limit of midlatitude weather? *J. Atmos. Sci.* **76**, 1077–1091 (2019).

29. Shukla, J. Predictability in the midst of chaos: a scientific basis for climate forecasting. *Science* **282**, 728–731 (1998).

30. Kushnir, Y. et al. Atmospheric GCM response to extratropical SST anomalies: synthesis and evaluation. *J. Clim.* **15**, 2233–2256 (2002).

31. Shapiro, M. A., Wernli, H., Bond, N. A. & Langland, R. The influence of the 1997–99 El Niño Southern Oscillation on extratropical baroclinic life cycles over the eastern North Pacific. *Q. J. R. Meteorol. Soc.* **127**, 331–342 (2001).

32. Orlanski, I. A new look at the pacific storm track variability: sensitivity to tropical SSTs and to upstream seeding. *J. Atmos. Sci.* **62**, 1367–1390 (2005).

33. Martius, O., Schwierz, C. & Davies, H. C. Breaking waves at the Tropopause in the Wintertime Northern Hemisphere: climatological analyses of the orientation and the theoretical LC1/2 classification. *J. Atmos. Sci.* **64**, 2576–2592 (2007).

34. Drouard, M., Rivière, G. & Arbogast, P. The link between the North Pacific Climate Variability and the North Atlantic Oscillation via downstream propagation of synoptic waves. *J. Clim.* **28**, 3957–3976 (2015).

35. Zhang, G. & Wang, Z. North Atlantic Rossby wave breaking during the hurricane season: association with tropical and extratropical variability. *J. Clim.* **32**, 3777–3801 (2019).

36. Wittman, M. A. H., Charlton, A. J. & Polvani, L. M. The effect of lower stratospheric shear on baroclinic instability. *J. Atmos. Sci.* **64**, 479–496 (2007).

37. Scaife, A. A. & Smith, D. A signal-to-noise paradox in climate science. *npj Clim. Atmos. Sci.* **1**, 28 (2018).

38. Deser, C. et al. Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* **10**, 277–286 (2020).

39. Lu, F. et al. GFDL's SPEAR seasonal prediction system: ocean data assimilation (ODA), ocean tendency adjustment (OTA) and coupled initialization. *J. Adv. Model. Earth Syst.* **12**, e2020MS002149 (2020).

40. Delworth, T. L. et al. SPEAR: the next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *J. Adv. Model. Earth Syst.* **12**, e2019MS001895 (2020).

41. Matsueda, M. & Endo, H. The robustness of future changes in Northern Hemisphere blocking: A large ensemble projection with multiple sea surface temperature patterns: Robustness of Future Changes in Blocking. *Geophys. Res. Lett.* **44**, 5158–5166 (2017).

42. Yoshida, K., Sugi, M., Mizuta, R., Murakami, H. & Ishii, M. Future changes in tropical cyclone activity in high-resolution large-ensemble simulations: future TC in large-ensemble simulations. *Geophys. Res. Lett.* **44**, 9910–9917 (2017).

43. Kamae, Y. et al. Forced response and internal variability of summer climate over western North America. *Clim. Dyn.* **49**, 403–417 (2017).

44. Zhang, G., Murakami, H., Knutson, T. R., Mizuta, R. & Yoshida, K. Tropical cyclone motion in a changing climate. *Sci. Adv.* **6**, eaaz7610 (2020).

45. Davini, P. & D'Andrea, F. From CMIP3 to CMIP6: Northern Hemisphere atmospheric blocking simulation in present and future climate. *J. Clim.* **33**, 10021–10038 (2020).

46. Woollings, T. et al. Blocking and its response to climate change. *Curr. Clim. Change Rep.* **4**, 287–300 (2018).

47. Schiemann, R. et al. The resolution sensitivity of Northern Hemisphere blocking in four 25-km atmospheric global circulation models. *J. Clim.* **30**, 337–358 (2017).

48. Barnston, A. G. & Livezey, R. E. ClassificatioN, Seasonality and Persistence of Low-frequency Atmospheric Circulation Patterns. *Mon. Weather Rev.* **115**, 1083–1126 (1987).

49. Strong, C. & Magnusdottir, G. Tropospheric Rossby wave breaking and the nAO/NAM. *J. Atmos. Sci.* **65**, 2861–2876 (2008).

50. Zhang, G., Murakami, H., Gudgel, R. & Yang, X. Dynamical seasonal prediction of tropical cyclone activity: robust assessment of prediction skill and predictability. *Geophys. Res. Lett.* **46**, 5506–5515 (2019).

51. Athanasiadis, P. J. et al. A multisystem view of wintertime NAO seasonal predictions. *J. Clim.* **30**, 1461–1475 (2017).

52. Vecchi, G. A. et al. On the seasonal forecasting of regional tropical cyclone activity. *J. Clim.* **27**, 7994–8016 (2014).

53. Wittenberg, A. T. & Anderson, J. L. Dynamical implications of prescribing part of a coupled system: results from a low-order model. *Nonlinear Process. Geophys.* **5**, 167–179 (1998).

54. Sutton, R. & Mathieu, P.-P. Response of the atmosphere–ocean mixed-layer system to anomalous ocean heat-flux convergence. *Q. J. R. Meteorol. Soc.* **128**, 1259–1275 (2002).

55. Zhang, W. & Kirtman, B. Understanding the signal-to-noise paradox with a simple Markov model. *Geophys. Res. Lett.* **46**, 13308–13317 (2019).

56. O'Reilly, C. H. & Czaja, A. The response of the Pacific storm track and atmospheric circulation to Kuroshio Extension variability: response of Storm Tracks to Kuroshio Extension Variability. *Q. J. R. Meteorol. Soc.* **141**, 52–66 (2015).

57. Liu, X. et al. Ocean fronts and eddies force atmospheric rivers and heavy precipitation in western North America. *Nat. Commun.* **12**, 1268 (2021).

58. Strong, C. & Magnusdottir, G. The role of tropospheric Rossby wave breaking in the pacific decadal oscillation. *J. Clim.* **22**, 1819–1833 (2009).

59. Zavadoff, B. L. & Kirtman, B. P. North Atlantic Summertime Anticyclonic Rossby wave breaking: climatology, impacts, and connections to the Pacific decadal oscillation. *J. Clim.* **32**, 485–500 (2019).

60. Kirtman, B. P. et al. The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Am. Meteorol. Soc.* **95**, 585–601 (2014).

61. Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A. & Palmer, T. Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution: Atmospheric Seasonal Forecasts of the Twentieth Century. *Q. J. R. Meteorol. Soc.* **143**, 917–926 (2017).

62. O'Reilly, C. H. et al. Variability in seasonal forecast skill of Northern Hemisphere winters over the twentieth century: variability in seasonal forecast skill. *Geophys. Res. Lett.* **44**, 5729–5738 (2017).

63. Davini, P. et al. The representation of winter Northern Hemisphere atmospheric blocking in ECMWF seasonal prediction systems. *Q. J. R. Meteorol. Soc.* **147**, 1344–1363 (2021).

64. Weisheimer, A. et al. How confident are predictability estimates of the winter North Atlantic Oscillation? *Q. J. R. Meteorol. Soc.* **145**, 140–159 (2019).

65. Pfahl, S., Schwierz, C., Croci-Maspoli, M., Grams, C. M. & Wernli, H. Importance of latent heat release in ascending air streams for atmospheric blocking. *Nat. Geosci.* **8**, 610–614 (2015).

66. Zhang, G. & Wang, Z. North Atlantic Extratropical Rossby wave breaking during the warm season: wave life cycle and role of diabatic heating. *Mon. Weather Rev.* **146**, 695–712 (2018).

67. Dee, D. P. et al. The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).

68. Rayner, N. A. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407–4436 (2003).

69. Mizuta, R. et al. Over 5,000 years of ensemble future climate simulations by 60-km global and 20-km regional atmospheric models. *Bull. Am. Meteorol. Soc.* **98**, 1383–1398 (2017).

70. Mizuta, R. et al. Climate simulations using MRI-AGCM3.2 with 20-km grid. *J. Meteorol. Soc. Jpn.* **90A**, 233–258 (2012).

71. Hirahara, S., Ishii, M. & Fukuda, Y. Centennial-scale sea surface temperature analysis and its uncertainty. *J. Clim.* **27**, 57–75 (2014).

72. Zhao, M. et al. The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *J. Adv. Model. Earth Syst.* **10**, 735–769 (2018).

73. Adcroft, A. et al. The GFDL global ocean and sea ice model OM4.0: model description and simulation features. *J. Adv. Model. Earth Syst.* **11**, 3167–3211 (2019).

AUTHOR CONTRIBUTIONS

G.Z. conceived of the study, implemented the RWB identification, and conducted the analysis with the help of H.M. and W.F. C.G.Z., H.M., and Z.W. contributed to the interpretation of the results. L.-W.J., F.-Y.L., X.-S.Y., and T.L.D. contributed to the retrospective seasonal prediction experiments using the SPEAR system. G.Z. wrote the paper with input from all authors.

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/s41612-021-00209-3.

Correspondence and requests for materials should be addressed to Gan Zhang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.