

ARTICLE OPEN



Increased wintertime European atmospheric blocking frequencies in General Circulation Models with an eddy-permitting ocean

Simon L. L. Michel¹✉, Anna S. von der Heydt¹, René M. van Westen¹, Michiel L. J. Baatsen¹ and Henk A. Dijkstra¹

Midlatitude atmospheric blocking events are important drivers of long-lasting extreme weather conditions at regional to continental scales. However, modern climate models consistently underestimate their frequency of occurrence compared to observations, casting doubt on future projections of climate extremes. Using the prominent and largely underestimated winter blocking events in Europe as a test case, this study first introduces a spatio-temporal approach to study blocking activity based on a clustering technique, allowing to assess models' ability to simulate both realistic frequencies and locations of blocking events. A sensitivity analysis from an ensemble of 49 simulations from 24 coupled climate models shows that the presence of a mesoscale eddy-permitting ocean model increases the realism of simulated blocking events for almost all types of patterns clustered from observations. This finding is further explained and supported by concomitant reductions in well-documented biases in Gulf Stream and North Atlantic Current positions, as well as in the midlatitude jet stream variability.

npj Climate and Atmospheric Science (2023)6:50; <https://doi.org/10.1038/s41612-023-00372-9>

INTRODUCTION

Extreme weather events are expected to increase in the coming decades as a result of anthropogenic climate change^{1–3}, but the extent of this increase remains highly uncertain⁴. Midlatitude atmospheric blocking is one of the most important phenomena contributing to the development of extreme weather events. Blocking is characterized by abnormally high and long-lasting (i.e., several days) pressure anomalies in the atmosphere at mid-latitudes. When such an event occurs, the anticyclonic high-pressure system disrupts the usual eastward propagating synoptic flows, resulting in deflected jet streams along the blocking's edges. These large changes in jet stream pathways, as major drivers of storm tracks and weather systems, have significant meteorological consequences for a wide range of regions^{5–8}.

There are numerous examples of past episodes that have left an imprint on collective memory as a result of the impacts of atmospheric blocking: the August 2003 heatwave in Europe, which resulted in an excess of 30,000 deaths⁹, the unusually large number of extreme cold spells that occurred in Europe in winter 2009/2010¹⁰, or the recent exceptional heatwave that occurred in Canada in early summer 2021. The latter event resulted in a temperature record high in British Columbia (up to 49 °C), causing increased mortality and economic damages¹¹. Atmospheric blocking events with less severe societal and ecological impacts occur quite frequently, but their presence varies greatly across regions and seasons¹². However, they drive a large proportion of long-lasting and large-scale precipitation and temperature anomalies, explaining significant portions of weekly to intraseasonal climate variability across a wide range of regions¹³.

Thus, in the context of climate change, understanding future atmospheric blocking locations, fingerprints, longevities, and frequencies is crucial for developing accurate adaptation plans in a wide range of societal and economic fields, such as agriculture, energy, tourism, and urbanism. However, the main tool we can rely on to study future blocking activity, namely

General Circulation Models (GCMs) of climate, mostly fail at simulating the observed atmospheric blocking characteristics over the historical period, particularly in terms of frequency of occurrence^{8,14}. Indeed, it has been demonstrated that none of the GCMs used in the last three Coupled Model Intercomparison Projects (CMIP3, CMIP5, and CMIP6) can reproduce the observed blocking frequencies from reanalysis data¹⁴. This is particularly true in areas and during seasons where atmospheric blocking events are more prominent and have a greater societal impact. Winter European Atmospheric Blocking events (WEABs), for example, are among the most prevalent blocking types, with CMIP3 to CMIP6 GCMs all highly underestimating their frequencies^{14,15}. This raises legitimate concerns that GCM projections in such areas may be significantly untrustworthy, as major extreme events caused by WEABs would not be accounted for.

Several studies have attempted to identify sources of improvement for increasing the realism of WEABs simulated by GCMs. For instance, a study¹⁶ illustrated that Sea Surface Temperature (SST) biases related with a model-wise consistent southward-biased position of the North Atlantic Current¹⁷ found in many GCMs, can account for the lack of blocking occurrence over the North Atlantic Ocean and Europe. Indeed, they showed that an atmosphere-only model can accurately reproduce blocking frequencies detected from the ERA40 reanalysis, while its coupling with an ocean model with large biases in North Atlantic SSTs leads to strongly reduced blocking events¹⁶. Furthermore, when using the same coupled GCM and applying a correction for the mean SST bias, this study¹⁶ produced much more accurate blocking frequencies. Similarly, it was shown that WEAB frequencies are largely influenced by SSTs of the Gulf Stream region¹⁸, which are poorly simulated by most GCMs¹⁹. Increased horizontal resolution of the atmospheric model has also been shown to better depict WEAB frequencies^{20–22}, whereas increased orographic resolution for a given atmospheric resolution results in improved blocking frequencies²⁰. Finer atmospheric grids, however, do not guarantee improvement, as

¹Institute for Marine and Atmospheric research Utrecht (IMAU), Department of Physics, Utrecht University, Utrecht, The Netherlands. ✉email: s.l.l.michel@uu.nl

decreases in performance have been observed in several cases²¹. The improvement was also shown to be model-dependent in a set of SST-forced atmospheric global circulation models²². Other biases in GCMs, for instance in tropical convection²³ and the resulting precipitation distribution²⁴, also account for a significant underestimation of WEABs. In summary, it appears that detected and undetected model biases^{16,18,23} all contribute together to the underestimation of WEAB frequencies by coupled GCMs^{8,14}. Therefore, this issue constitute a major challenge in coupled climate modeling, and the identification of new sources of WEAB frequency bias reduction is essential^{8,14}.

Multiple studies showed that several GCM biases, notably in SST patterns, are reduced when using strongly eddying ocean models compared to the standard low-resolution ones^{25–28}, which are still highly prominent in the current CMIP6 GCM generation²⁹. In this respect, refining the horizontal grid spacing of ocean models to less than 0.25° allows the simulation of mesoscale eddies in the ocean. Therefore, ocean models fine enough to resolve the Rossby radius and thus to simulate the largest eddies at low to mid latitudes need a nominal resolution of at least 25 km³⁰. Mesoscale eddies typically span 20 to 100 km and contribute significantly to both horizontal and vertical heat and salt exchange³¹. At the surface, the horizontal distribution of heat driven by mesoscale eddies importantly contribute to sharp SST gradients, thus acting on the atmospheric boundary layer and higher atmosphere through variations in surface stability on momentum transfer, cloud cover and pressure gradients³². They have notably been shown to contribute significantly to the realism of air-sea interactions in GCMs, as well as large-scale oceanic features^{33–35}.

In this study, we analyze WEABs in a set of 49 model simulations from 24 GCMs, each with its own ocean and atmospheric model resolutions (Table 1). Of these simulations, 16 resulted from GCMs with an eddy-permitting (EP) ocean resolution and 33 come from GCMs with no-eddy (NE) ocean resolution. As observations for GCMs evaluation, we use daily data from the ERA5 reanalysis dataset³⁶ for the 36-year-long period 1979–2014. Subsequent years are not considered because the historical model simulations, with which we will compare the reanalysis data, are not run over this period²⁹. We also do not use the ERA5 extension from 1950 to 1978 to avoid adding uncertainties, because this extension is based on less assimilated data³⁶. In the following, the study area for WEABs is restricted to the Euro-Atlantic sector (50°W–50°E, 30°N–75°N) and for winter (December–January–February, DJF) days only, whose number varies slightly depending on the model calendar (Table 1). For blocking detection, a widely used two-dimensional blocking index matrix is first computed from geopotential heights at 500 hPa^{14,15,37,38} (Z500) and only blocking situations of 4 (consecutive) days or more are accounted to only consider large-scale, long-term, and impactful WEAB occurrence as usually done^{14,15} (see Methods). Considering the spatial distribution of ERA5 blocking frequencies from 1979 to 2014, we find that most blocking occurrences occur north of 40°N and are centered over western and northern Europe, as well as Greenland, which is consistent with previous research^{8,14,15,21} (Fig. 1a). When investigating biases for both EP and NE GCM simulations which will be used in the following, we notice at first glance that larger biases appear in areas with large blocking occurrences: western Europe, Nordic and Baltic seas, as well as Greenland (Fig. 1b, c). A large number of grid points where blocking frequencies biases are significantly different (at the 90% confidence level) between the two groups is found, most notably at the advantage of EP models in areas of large blocking occurrence (Fig. 1d).

This first result is in line with results from a very recent study that found significantly larger spatially-averaged instantaneous blocking frequencies over the Euro-Atlantic sector north of 40°N for GCMs with EP ocean components in a set of 17 coupled simulations³⁹. They explained this improvement by reduced biases in meridional SST gradients over the North Atlantic, which are

strongly related to the large cold SST bias in the central North Atlantic found in most GCMs and formerly explained by a southward-biased position of the North Atlantic Current path^{16,40}. As hypothesized by former studies^{16,40,41}, such biases in SST patterns lead to biased low-level baroclinicity and vertical air motions, thus acting on the eddy-driven jet variability and blocking. However, although this recent study³⁹ investigated several biases in jet-stream-related atmospheric circulation features and air-sea interactions, their investigation concerning WEAB remains limited. First, they consider instantaneous blocking frequencies, meaning that noisy and short (less than 4 days) blocking situations are accounted for, which can be problematic as a limited area being blocked for one or two days may not result in a significant and statistically detectable weather extreme³⁹. Second, their study is based on single realizations from 17 GCMs, but this model ensemble is, at the end, composed of different versions from 7 GCMs, which differ in their oceanic and atmospheric models' resolutions³⁹. Such an experimental setup, however, led the authors to conclude that when considering different versions of a given GCM, the shift from NE to EP ocean model resolution consistently results in increased instantaneous WEAB frequencies. However, they also found no consistence throughout their relatively small model ensemble and concluded that the latter result is strongly model-dependent³⁹. In addition, as GCMs with EP ocean model resolutions have generally higher atmosphere model resolutions than NE ones, and as increased atmospheric resolutions generally leads to improved simulated WEABs²¹, it is difficult to disentangle the effects of increased ocean and atmosphere model resolutions on their result. Here, we investigate whether the EP ocean grid effect on simulated WEABs³⁹ can be generalized by: i) looking at a much larger set of coupled simulations from a wider range of climate modeling centers (7 against 17 here, Table 1), and ii) comparing the results to simulated blocking frequencies from another relatively large set of 31 atmospheric model simulations with prescribed SSTs based on atmospheric models of a subset of the coupled GCMs investigated (Table 2). The purpose of such an extensive analysis is to study the effect of the EP ocean resolution without accounting for the effect of atmospheric grid resolutions. Last but not least, the latter study focused on instantaneous blocking frequencies spatially averaged over a large area. Therefore, they did not check whether GCMs simulate long-standing and spatially-consistent blocking events at right locations and with accurate frequencies compared to reanalysis data. In summary, while this recent research suggests that mesoscale-resolving ocean resolutions may aid in decreasing biases in simulated WEABs for a given GCM, a clearer picture of the exact effect of EP ocean resolution beyond the role of atmospheric model resolutions in simulating large-scale and impactful WEAB events is still required.

To provide a more comprehensive sensitivity analysis than previously³⁹, we first introduce an approach to study GCM ability in simulating WEABs, both in terms of frequency and location^{42–45}, based on a *k-means* clustering on the commonly used two-dimensional blocking index matrix depicted above (see Methods). This clustering allows for the separation of detected blocking situations into distinct groups of spatially coherent patterns, with only clustered blocking situations being accounted for. Subsequently, sensitivity analyses for each blocking pattern are carried out and we show that total blocking frequencies through all clusters are significantly improved in EP GCMs compared to NE ones. We further show that EP GCMs outperform NE GCMs in reproducing the wintertime Gulf Stream and North Atlantic Current positions and mean SST distributions, as well as in resolving patterns of variability in the high troposphere geopotential heights. Therefore, based on earlier work^{16,18,19,39,40} and our 49 GCM simulations ensemble, we hypothesize that GCM bias reductions induced by the presence of mesoscale eddies act as a strong source of improvement for simulating more accurate WEAB

Table 1. Description of the General Circulation Models (GCMs) simulations used for this study.

Model name	Research center (Country)	Number of grid points Atmosphere (lon × lat)	Number of grid points Ocean (lon × lat)	Nominal resolution in km (Atmosphere, Ocean)	Eddy presence	Calendar	Experiment	Members
BCC-CSM2-HR	BCC (China)	800 × 400	1441 × 686	(50, 50)	NE	365 days	hist-1950	r1i1p1f1
CanESM5	CCCma (Canada)	128 × 64	361 × 290	(500, 100)	NE	365 days	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1
CMCC-CM2-HR4	CMCC (Italy)	288 × 192	1442 × 1051	(100, 25)	EP	365 days	hist-1950	r1i1p1f1
CMCC-CM2-VHR4	CMCC (Italy)	1152 × 768	1442 × 1051	(25, 25)	EP	365 days	hist-1950	r1i1p1f1
CNRM-CM6-1	CNRM (France)	256 × 128	362 × 294	(250, 100)	NE	Gregorian	hist-1950	r1i1p1f2
CNRM-CM6-1-HR	CNRM (France)	720 × 360	1442 × 1050	(100, 25)	EP	Gregorian	historical	r1i1p1f2
EC-Earth3P	EC-Earth consortium (EU)	512 × 256	362 × 292	(100, 100)	NE	Gregorian	hist-1950	r1i1p2f1, r2i1p2f1, r3i1p2f1
EC-Earth3P-HR	EC-Earth consortium (EU)	1024 × 512	1442 × 1021	(50, 25)	EP	Gregorian	hist-1950	r1i1p2f1, r2i1p2f1, r3i1p2f1
ECMWF-IFS-HR	ECMWF (UK)	1600 × 800	1442 × 1021	(25, 25)	EP	Gregorian	hist-1950	r1i1p1f1, r2i1p1f1, r3i1p1f1
ECMWF-IFS-LR	ECMWF (UK)	800 × 400	362 × 192	(50, 100)	NE	Gregorian	hist-1950	r1i1p1f1, r2i1p1f1, r3i1p1f1
ECMWF-IFS-MR	ECMWF (UK)	800 × 400	1442 × 1021	(50, 25)	EP	Gregorian	hist-1950	r1i1p1f1, r2i1p1f1, r3i1p1f1
FGOALS-g3	CAS (China)	180 × 80	360 × 218	(250, 100)	NE	365 days	historical	r1i1p1f1, r3i1p1f1, r4i1p1f1
GFDL-CM4	GFDL (UK)	360 × 180	1440 × 1080	(100, 25)	EP	365 days	historical	r1i1p1f1
GISS-E2-1-G	NASA GISS (USA)	144 × 90	360 × 180	(250, 100)	NE	365 days	historical	r1i1p1f1, r1i1p1f2
HadGEM3-GC31-MM	MOHC (UK)	432 × 324	1440 × 1205	(100, 25)	EP	Gregorian	historical	r1i1p1f3, r2i1p1f3, r3i1p1f3
IITM-ESM	CCCR-IITM (India)	192 × 94	360 × 200	(250, 100)	NE	360 days	historical	r1i1p1f1
INM-CM5-0	INM (Russia)	384 × 192	720 × 720	(100, 50)	NE	365 days	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1
IPSL-CM6A-LR	IPSL (France)	144 × 143	362 × 332	(250, 100)	NE	Gregorian	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1
KACE-1-0-G	NIMS-KMA (Korea)	192 × 144	360 × 200	(250, 100)	NE	360 days	historical	r1i1p1f1
MPI-ESM1-2-HR	MPI (Germany)	384 × 192	802 × 404	(100, 50)	NE	Gregorian	hist-1950	r1i1p1f1
MPI-ESM1-2-XR	MPI (Germany)	768 × 384	802 × 04	(50, 50)	NE	Gregorian	hist-1950	r1i1p1f1
NorESM2-MM	NCC (Norway)	288 × 192	360 × 384	(100, 100)	NE	365 days	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1
TaiESM1	AS-RCES (Taiwan)	288 × 192	320 × 384	(100, 100)	NE	365 days	historical	r1i1p1f1
UKESM1-0-LL	MOHC (UK)	192 × 144	360 × 330	(250, 100)	NE	360 days	historical	r1i1p1f2, r2i1p1f2, r3i1p1f2

For oceanic and atmospheric grid resolutions, the dimensions are given as the number of grid points and nominal horizontal resolutions. For eddy presence, “EP” indicates GCMs with an eddy-permitting ocean resolution and “NE” indicates GCM with a no-eddy ocean resolution. For calendars, “Gregorian” indicates GCMs with a regular calendar, “365 days” indicates GCMs having the same calendar as “Gregorian” but without 29th of February, “360 days” indicates GCMs with a calendar with 30 days for each month of the year. The “hist-1950” experiment indicates that the GCM simulations were taken from CMIP6 whereas “historical” indicates that the GCM simulations were taken from HighResMIP.

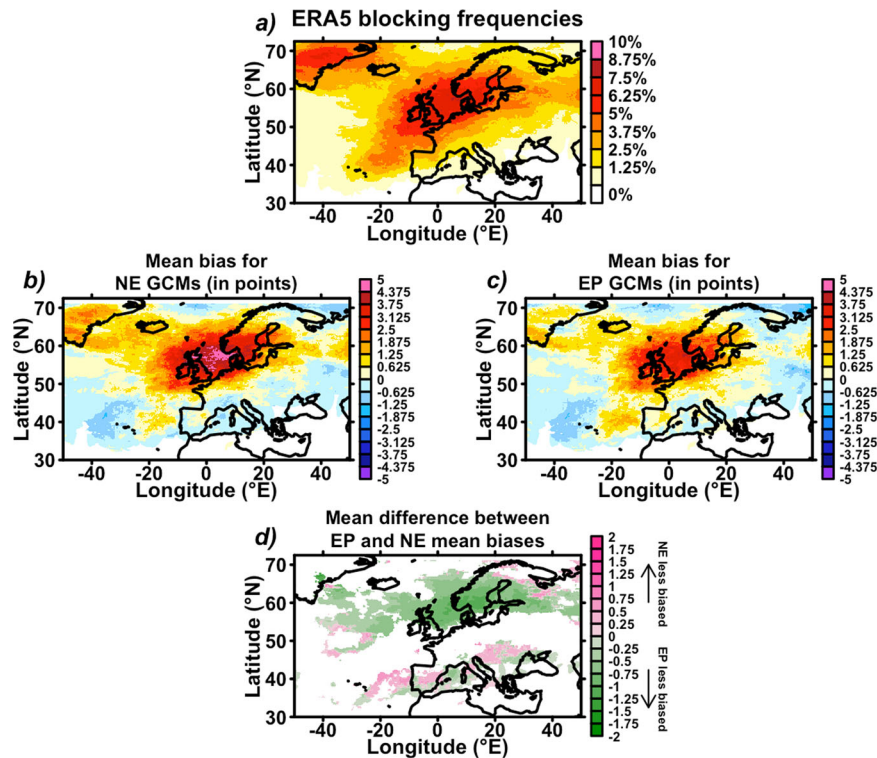


Fig. 1 Spatially distributed blocking frequencies and biases in reanalyses and coupled model data. **a** Blocking frequencies from ERA5 for blocking situations of 4 days or more. White grid points indicate blocking frequencies of 0%. **b, c** Mean biases of blocking frequencies (in points) for the 33 no-eddy (NE) and the 16 eddy-permitting (EP) General Circulation Model (GCM) simulations, respectively. White colors indicate observed blocking frequencies (**a**) of 0%. **d** Difference in mean biases for GCMs with EP (**c**) and NE (**b**) ocean model resolutions. For grid points, green colors indicate that EP GCMs have a significantly lower mean bias than NE GCMs at the 90% confidence level whereas pink colors indicate that NE GCMs have a significantly lower mean bias than EP GCMs at the 90% confidence level. White colors indicate observed blocking frequencies (**a**) of 0% or no significant mean bias differences using a two-tailed Student's *t* test with a 90% confidence level.

frequencies and locations. This finding implies that greater confidence can be attributed to paleoclimate simulations or future projections of WEAB activity and the underlying occurrence of climate extremes when using coupled GCMs with an EP ocean resolution. As a result, our findings strongly encourage climate modeling centers to shift toward coupled models with an EP ocean component in future CMIP updates to accurately study the future evolution of WEAB phenomena and related extremes in the context of anthropogenic climate change, which will greatly assist policymakers in designing effective climate adaptation strategies.

RESULTS

Clustering of blocking events in reanalysis data

We first compute a *k-means* clustering⁴⁶ from the time-varying 2-D blocking index to classify the different DJF days based on the spatial distribution of their blocking situations (Methods). The data used here for the clustering differ from earlier studies in that they either used geopotential height^{43,44} or sea level pressure^{42,45} fields. A major difference with these studies is that we here make a classification of DJF days where only spatially-consistent patterns of blocked grid points for four consecutive days or more are accounted for. The parameter *k* is tuned using Monte-Carlo sampling, which begins at *k* = 2 and ends when it is determined that there is no gain in terms of newly detected blocking type after increasing *k* by one (Methods).

For ERA5, we find a number of *k* = 6 clusters (Methods). The first cluster, consisting of 71.87% of the DJF days from 1979 to 2014 (i.e., 2335 of the 3249 days), is, as expected, the “non-blocking” situation. For non-blocking type situations, no consistent blocking footprint is found, where the highest blocking frequency among

the cluster only reaches 5.3% across all grid points within the study area (Supplementary Fig. 1). The second group is composed of days when a blocking located over Western Europe (WE) occurred (Fig. 2a). This group accounts for 7.63% of all DJF days for the period 1979–2014. The third group consists of days when blocking occurs over Greenland (Gr.) and comprises 6.4% of DJF days for the period 1979–2014 (Fig. 2b). The remaining blocking types are located in the North Sea (NS, 5.79%), Baltic Sea (BS, 4.68%), and Scandinavia (Sc., 3.63%), respectively (Fig. 2c–e). Overall, the clear blocking footprint patterns obtained from the *k-means* confirm the relevance of its application.

In addition, the daily-averaged anomalies (relative to the daily climatologies) within days clustered for each blocking type are used to compute blocking fingerprints (Fig. 2). Each WEAB type has a distinct pattern in 2-meter temperature, precipitation, and sea level pressure anomalies but some patterns share some similarities (e.g., NS and BS, Fig. 2). In contrast, some blocking types have nearly opposite impacts on temperature and precipitation, with some associated with negative (Gr.) or positive (WE) North Atlantic Oscillation patterns. As a further confirmation of the blocking clustering, we also observe surface pressure anomalies that coincides with the blocking footprints (Fig. 2a–e). Finally, we find a slightly significant pattern of positive North Atlantic Oscillation fingerprints for surface temperatures and precipitation in non-blocking situations, with much lower amplitudes of averaged anomalies than in blocking situations (Supplementary Fig. 1). This finding is consistent with an overall higher probability of WEAB occurrence during negative North Atlantic Oscillation episodes¹⁰, as recently demonstrated for the CERA-20C reanalysis covering 110 years⁴⁷. When comparing our clustering result to those from previous

Table 2. Description of the atmospheric model simulations forced by Sea Surface Temperatures (SSTs) used for this study.

Model name	Research center (Country)	Number of grid points Atmosphere (lon × lat)	Number of grid points Ocean (lon × lat)	Nominal resolution in km (Atmosphere, Ocean)	Eddy presence	Calendar	Experiment	Members
CMCC-CM2-HR4	CMCC (Italy)	288 × 192	1442 × 1051	(100, 25)	EP	365 days	highresSST-present	r11ip1f1
CMCC-CM2-VHR4	CMCC (Italy)	1152 × 768	1442 × 1051	(25, 25)	EP	365 days	highresSST-present	r11ip1f1
CNRM-CM6-1	CNRM (France)	256 × 128	362 × 294	(250, 100)	NE	Gregorian	highresSST-present	r11ip1f2, r31ip1f2, r41ip1f2
CNRM-CM6-1-HR	CNRM (France)	720 × 360	1442 × 1050	(100, 25)	EP	Gregorian	highresSST-present	r11ip1f2
EC-Earth3P	EC-Earth consortium (EU)	512 × 256	362 × 292	(100, 100)	NE	Gregorian	highresSST-present	r11ip1f1, r21ip1f1, r31ip1f1
EC-Earth3P-HR	EC-Earth consortium (EU)	1024 × 512	1442 × 1021	(50, 25)	EP	Gregorian	highresSST-present	r11ip1f1, r21ip1f1, r31ip1f1
ECMWF-IFS-HR	ECMWF (UK)	1600 × 800	1442 × 1021	(25, 25)	EP	Gregorian	highresSST-present	r11ip1f1, r21ip1f1, r31ip1f1
ECMWF-IFS-LR	ECMWF (UK)	800 × 400	362 × 192	(50, 100)	NE	Gregorian	highresSST-present	r11ip1f1, r21ip1f1, r31ip1f1
GFDL-CM4	GFDL (UK)	360 × 180	1440 × 1080	(100, 25)	EP	365 days	amip	r11ip1f1
HadGEM3-GC31-MM	MOHC (UK)	432 × 324	1440 × 1205	(100, 25)	EP	Gregorian	amip	r11ip1f3
IITM-ESM	CCCR-IITM (India)	192 × 94	360 × 200	(250, 100)	NE	360 days	amip-hist	r11ip1f1
INM-CM5-0	INM (Russia)	384 × 192	720 × 720	(100, 50)	NE	365 days	amip	r11ip1f1
IPSL-CM6A-LR	IPSL (France)	144 × 143	362 × 332	(250, 100)	NE	Gregorian	amip	r11ip1f1, r31ip1f1, r71ip1f1
MPI-ESM1-2-HR	MPI (Germany)	384 × 192	802 × 404	(100, 50)	NE	Gregorian	amip	r11ip1f1, r21ip1f1, r31ip1f1
MPI-ESM1-2-XR	MPI (Germany)	768 × 384	802 × 404	(50, 50)	NE	Gregorian	highresSST-present	r11ip1f1
TaiESM1	AS-RCES (Taiwan)	288 × 192	320 × 384	(100, 100)	NE	365 days	amip-hist	r21ip1f1
UKESM1-0-LL	MOHC (UK)	192 × 144	360 × 330	(250, 100)	NE	360 days	amip	r11ip1f4

For oceanic and atmospheric grid resolutions of the coupled versions of AGCMs, the dimensions are given as the number of grid points and nominal horizontal resolutions. For eddy presence in the coupled version, "EP" indicates GCMs with an eddy-permitting ocean resolution and "NE" indicates GCM with a no-eddy ocean resolution. For calendars, "Gregorian" indicates GCMs with a regular calendar, "365 days" indicates GCMs having the same calendar as "Gregorian" but without 29th of February, "360 days" indicates GCMs with a calendar with 30 days for each month of the year. The "amip" and "amip-hist" experiments indicate that the simulations were taken from CMIP6 whereas "highresSST-present" indicates that the simulations were taken from HighResMIP.

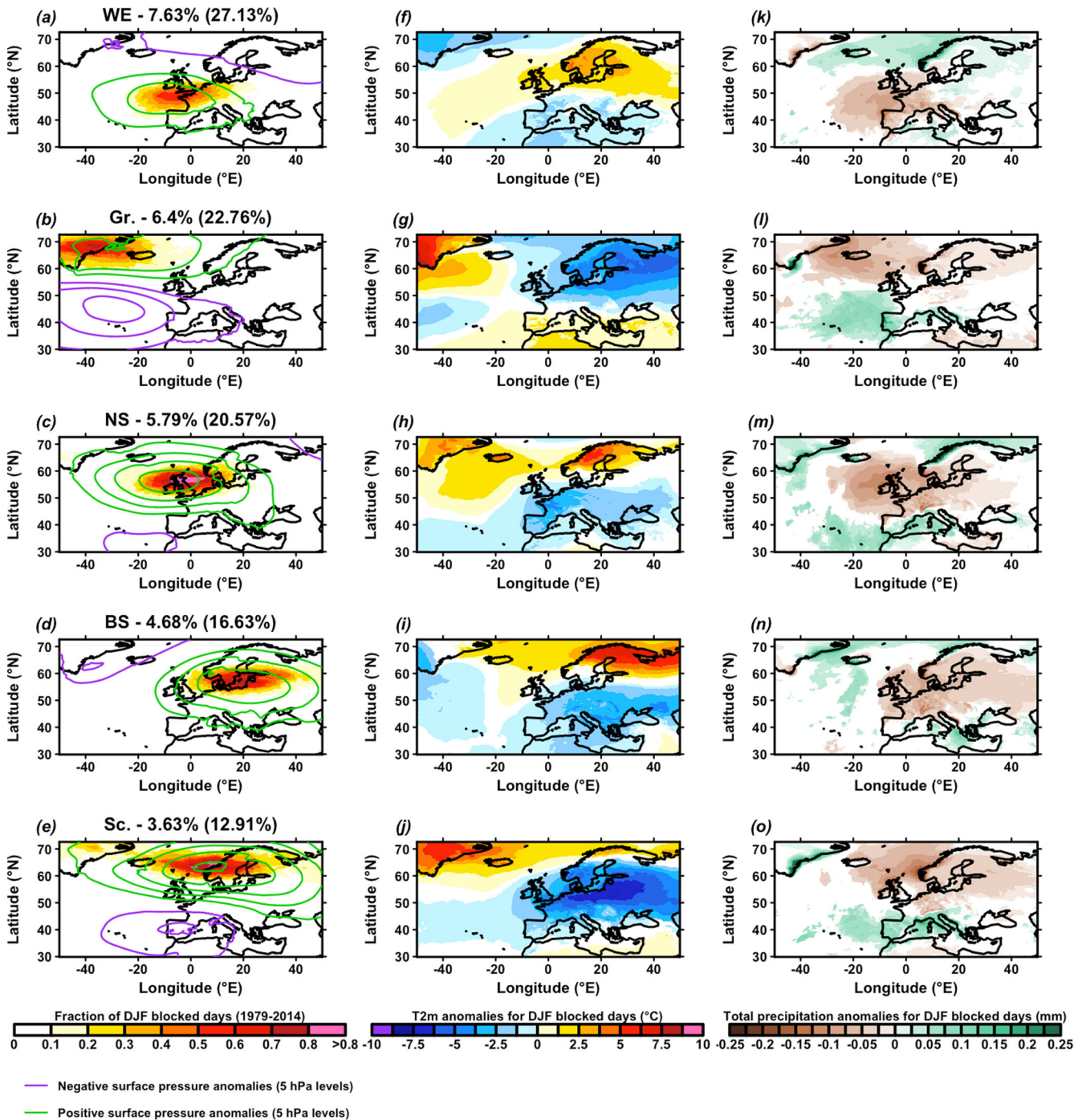


Fig. 2 Frequencies of clustered blocking types and associated climate fingerprints. Clustered Western European Atmospheric Blocking (WEAB) types by *k-means* (a–e) and associated averaged daily anomalies (according to daily climatologies) of sea level pressure (a–e, contours), temperature at 2 meters (T2m) (f–j), and precipitation (k–o) from ERA5 reanalysis data for the period 1979–2014. Blocking footprints in (a–e) are calculated by summing the two-dimensional blocking index matrix composed of ones and zeros (Methods) over the days clustered for each blocking type. The fraction of DJF days clustered is indicated at the top of blocking footprint maps for each blocking type: Western Europe (WE), Greenland (Gr.), North Sea (NS), Baltic Sea (BS), and Scandinavia (Sc.). Numbers between brackets indicate the blocking fractions obtained when excluding the non-blocking days (i.e., 71.87% of the total). Green (respectively purple) contour lines in (a–e) indicate 5 hPa levels of positive (respectively negative) sea level pressure daily anomalies (according to daily climatologies) for days of each cluster. For climate fingerprints (f–o), white colored grid points indicate no significance at the 95% confidence level from a two-tailed Student's *t* test.

studies^{42,45}, we find similar patterns, though with some notable differences. One main difference is that our clustering discriminates Gr. and Sc. WEAB types, whereas they were clustered together in these studies. The differences found may be due to the different experimental setups. Indeed, they^{42,45} used

different study periods (i.e., 1949–2001 and 1950–2000, respectively), different datasets (i.e., NCEP reanalysis and ARPEGE atmosphere GCM, respectively), different study areas (larger than the present one), and another climate variable (i.e. sea level pressure for both).

Increased blocking frequencies in EP GCMs

We use daily data of the 49 simulations from the 24 CMIP6 GCMs²⁹ presented in Table 1 and Fig. 1, on which 11 (20 simulations of 49) participate to the HighResMIP activity⁴⁸ for the same period as ERA5 (1979–2014). The *hist-1950* experiment is used for the GCMs participating to HighResMIP while the CMIP6 *historical* experiment is taken for the remaining ones. Although HighResMIP runs were made with a relatively high number of EP GCMs compared to classical CMIP6 runs, some EP GCM simulations can be found for the *historical* experiment and NE GCM simulations are present for the *hist-1950* one (Table 1). It must be noted that the aerosol forcings from *hist-1950* and *historical* experiments differ⁴⁸, which may impact the sensitivity analyses to ocean grid resolutions carried out below. In addition, simulations for these two experiments start at different years (1950 and 1870, respectively), so they were not initialized in the same way. Consequently, an additional sensitivity analysis for the difference in GCM experimental setups will be proposed in the following.

For each GCM, up to 3 realizations are used for the study, when available (Table 1). In the end, the model ensemble is composed of 16 simulations from 8 EP GCMs and 33 simulations from 16 NE GCMs (Table 1). These data were downloaded from the ESGF portal, where daily Z500 data for HighResMIP models are not yet exhaustively provided for all existing runs. Therefore, the High-ResMIP models we have used for this study are those for which Z500 data were made available. For a complete list of existing HighResMIP simulations, the reader may refer to the related publication⁴⁸.

For each GCM, the same clustering approach as for reanalysis data with the same $k=6$ parameter is computed. Using spatial correlations, a given blocking type clustered from a GCM output is compared to those clustered from ERA5. If at least one of these spatial correlations is higher than 0.5, the GCM cluster is assigned to the ERA5 blocking type with the highest spatial correlation (Methods). This is carried out for all clusters derived from the 49 GCM simulations.

Considering the blocking type assignment for GCM simulations described in Methods and the above paragraph, Fig. 3a shows the blocking frequencies for each blocking type, for each of the GCMs listed in Table 1. The average frequencies for all blocking types are significantly underestimated for the whole GCM ensemble, where none of the ERA5 blocking frequencies fall within the 95% confidence levels determined from the 49 GCM simulations (Fig. 3a). It is important to note that there is a significant spread between the different GCM simulations for each blocking type, meaning that each GCM has its own ability to simulate certain types of blocking and that GCMs' initialization is also important in this respect. Two blocking types (NS and Sc.) are relatively poorly captured, where they can be detected in only 30 and 26 of the 49 GCM simulations considered, respectively (Methods, Fig. 3a). For the other three blocking types, some of the GCMs do not detect them either, but to a lesser extent than for the NS and Sc. types (Fig. 3a).

When separating the GCM simulations into those made with EP ocean resolution and those without, we notice a significant improvement in the frequency of WEABs overall (Fig. 3a, b). For EP GCMs, the improvement in terms of mean blocking frequencies is significant at the 90% confidence level for three of the most prevalent blocking types (WE, Gr., and BS). Based on one-tailed Student's t test, it reaches $+0.012$ ($+26.7%$, $p < 0.05$) for WE, $+0.009$ ($+20.7%$, $p < 0.1$) for Gr., $+0.008$ ($+31.9%$, $p < 0.2$) for NS, $+0.016$ ($+61%$, $p < 0.01$) for BS, and $+0.008$ ($+48.8%$, $0.1 < p < 0.11$) for Sc. (Fig. 3a). However, important outliers in EP GCMs from the same modeling center (CMCC, Table 1) may strongly conceal analysis results. Indeed, CMCC-CM2-HR4 and CMCC-CM2-VHR4, two versions of the same GCM with different atmospheric model resolutions (Table 1), simulate WEABs much less accurately

than for realizations from the 6 other EP GCMs (Fig. 3). Indeed, these two GCMs provide the two realizations with lower WEAB frequencies among the 16 EP GCM simulations investigated (Fig. 3), which is very unlikely to be obtained randomly. In addition, simulations from these two EP GCMs are the only with a total WEAB frequency lower than the average obtained for the 33 NE GCM simulations. When these two GCMs are removed, the improvement in WEAB frequencies reaches the 95% confidence level for all blocking types (Supplementary Fig. 2). However, for the sake of transparency, we continue to consider these two outliers, whereas in Supplementary Fig. 2, we provided a similar figure as Fig. 3 without them.

The WEAB frequencies detected in the reanalysis for the Gr. and BS types are within the 95% confidence interval of blocking frequencies found in EP GCMs (Fig. 3a). In comparison, none of the blocking types detected in NE GCMs have an average frequency that contains ERA5 statistics in its 95% confidence interval (Fig. 3a). Overall, the averaged total WEAB frequency reaches 20.6% for EP GCMs (22.35% without CMCC models) and 15.4% for NE GCMs, which is 33.8% higher for EP GCMs (45.1% without CMCC models), and significant at the 99% confidence level. As stated previously, it must be noted that two strong outliers appear in the EP GCM simulation ensemble (Fig. 3) while they both are two different versions of the same model (Table 1). In any case, since both were included in the analysis, this supports the result that EP GCMs perform better since their legitimate exclusion would have strongly increased the significances described above (Supplementary Fig. 2).

We note that the spatial correlation threshold of 0.5 is set arbitrarily to some extent but was preferred to a classical correlation test as a large number of spatial grid points leads to a poorly restrictive test and hence to the detection of significant spatial correlations for poorly matching WEAB patterns. To assess the effect of this spatial correlation threshold discussed in the Methods section, we show a similar analysis as for Fig. 3 for three other threshold values: (i) using a spatial correlation test, and for thresholds of (ii) 0.2 and (iii) 0.8 rather than 0.5 (Supplementary Figs. 3–5, respectively). As similar results are found in the three cases (see panel b from Supplementary Figs. 3–5), the following will focus on the WEAB attribution results shown in Fig. 3.

To further check the significance of the EP grid effect on the simulated WEAB frequencies, a Monte Carlo sampling was performed on the basis of GCMs with NE ocean grids. The same number of EP GCM simulations (16) is sampled 5000 times from the set of 33 NE GCM simulations and the total frequency of the clustered WEABs are quantified and compared to the total WEAB frequencies of the group of 16 EP GCM simulations. From this Monte-Carlo sampling, we find that the total WEAB frequencies for EP GCMs is higher than the 99th percentile of total WEAB frequencies obtained over the 5000 sampled groups of NE GCM simulations, indicating a very high level of significance.

A similar significance test was performed by distinguishing GCMs according to their atmospheric grid resolution without considering their ocean grid resolution (Fig. 3c). For this, GCMs were separated into those with a zonal atmospheric grid spacing at the equator greater than 0.75° (i.e., the low-resolution atmosphere hereafter) and those with a grid spacing smaller than 0.75° (i.e., the high-resolution atmosphere hereafter). This threshold for the atmospheric grid separates the GCM simulations into two groups: a first group of 19 GCM simulations with high-resolution atmosphere and a group of 30 GCM simulations with low-resolution atmosphere (Fig. 2c). The two sets are different from the NE/EP oceanic discrimination since 8 of the 33 NE GCM simulations are among the 19 with a high-resolution atmosphere while the other 25 have a low-resolution atmosphere. Similarly to the ocean grid resolution, we find that GCMs with a zonal atmospheric grid resolution at the equator greater than 0.75° also significantly outperform GCMs with a low-resolution atmosphere,

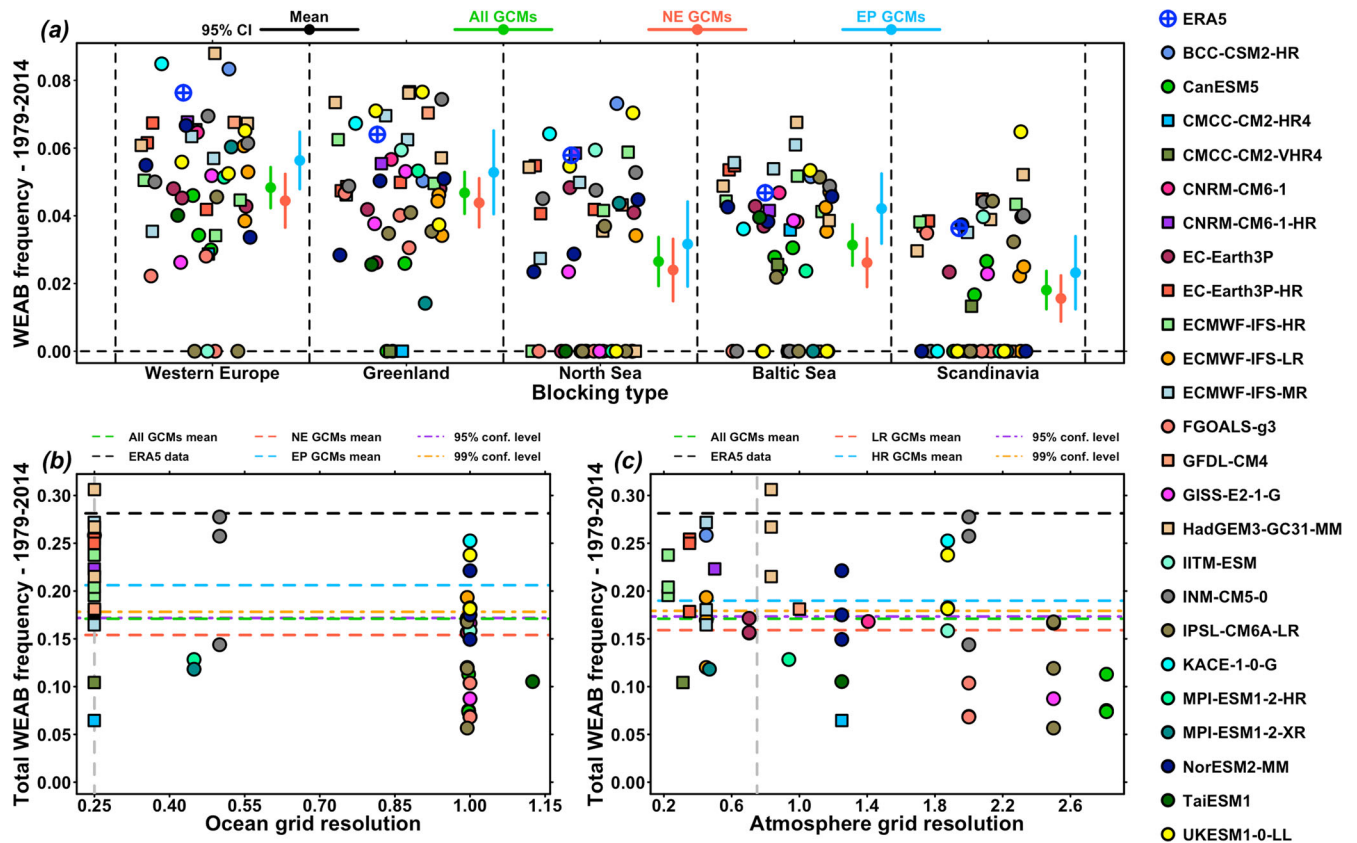


Fig. 3 Sensitivity analysis of Western European Atmospheric Blocking (WEAB) frequencies to ocean and atmospheric grid resolutions in General Circulation Models (GCMs). **a** WEAB frequencies calculated from ERA5 reanalysis data (blue target) and 49 historical simulations from 24 CMIP6 GCMs (other symbols) for each of the 5 main blocking types clustered by the *k-means* algorithm. Green, red, and blue dots on the right side indicate the average of the GCM ensemble, GCMs with no-eddy (NE) ocean grids, and GCMs with eddy-premitting (EP) ocean grids, respectively. Each line crossing these points on the right side give the respective 95% confidence intervals. The same black-colored ensemble of symbols shown at the top-left schematizes the statistics they give (Mean and 95% confidence interval, CI). **b** Total frequencies of clustered WEABs plotted against zonal ocean grid resolution at the Equator for the 49 simulations from the 24 GCMs. Vertical dashed gray line indicates the resolution at which GCMs have EP ocean grid resolutions. Horizontal black dashed line indicates the total WEAB frequencies for ERA5. Horizontal blue (respectively red) dashed line indicates the average of total WEAB frequencies for GCMs with EP (respectively NE) ocean grid resolutions. Horizontal green dashed line indicates the average of total WEAB frequencies for all GCMs. Horizontal purple and orange dashed lines indicate 95% and 99% confidence levels calculated as the percentiles of total blocking frequencies obtained from 5000 random Monte-Carlo samples of 16 NE GCMs, respectively. **c** Total frequencies of clustered WEABs plotted against zonal atmospheric grid resolution at the Equator for the 49 simulations from the 24 GCMs. Vertical gray dashed line indicates the resolution at which GCMs have high-resolution atmosphere (0.75°). Horizontal black dashed line indicates the total WEAB frequencies for ERA5. Horizontal blue (respectively red) dashed line indicates the average of total WEAB frequencies for GCMs with high-resolution (respectively low-resolution) atmospheric grids. Horizontal green dashed line indicates the average of total WEAB frequencies for all GCMs. Horizontal purple and orange dashed lines indicate 95% and 99% confidence levels calculated as the percentiles of total blocking frequencies obtained from 5000 random Monte-Carlo samples of 30 GCMs with low-resolution (LR) atmosphere, respectively. In (a–c), circles indicate GCMs with NE ocean resolution whereas squares indicate GCMs with EP ocean resolution.

with higher WEAB frequencies at the 99% confidence level. This result has also been found by previous studies^{20–22} and was explained by increased transient eddy forcing with higher atmospheric resolution.

Sensitivity to atmospheric resolution and experimental setups. Since GCM simulations with EP ocean resolution generally tend to be accompanied with higher-than-usual atmospheric resolution⁴⁸, it is difficult to tell if the improvement seen in Fig. 3c for finer atmospheric grids is independent from the one obtained by the use of EP ocean model resolutions (Fig. 3b). For this reason, we also computed linear regression slopes between total WEAB frequencies for all simulations, as well as for EP and NE GCM simulations separately. First, a significant linear slope is found between the atmospheric grid resolutions of the GCMs and the total WEAB frequencies (Fig. 3c, $p < 0.01$), which is line with findings from previous studies^{20–22}. For NE GCMs, this linear

relationship is significant ($p < 0.1$) whereas it is not for EP GCMs ($p > 0.1$). Indeed, it appears that GCMs with EP ocean grids have almost systematically high total WEAB frequencies (relative to the whole ensemble, Fig. 3b) regardless of atmospheric resolutions, although there is two EP GCMs from the same modeling center with relatively poorly resolved WEAB frequencies, as mentioned above (Fig. 3a–c). However, since the number of GCMs with an EP grid resolution is relatively low (8) it is difficult to conclude robustly that increasing their atmospheric resolution is not improving WEAB frequencies since this feature has previously been shown to be model-dependent²². To investigate the effect of the EP ocean resolution on simulating WEABs without accounting for atmospheric resolution, we compared our set of 49 historical simulations to another set of 31 historical simulations in which only the atmospheric model is used with prescribed SST. In this case, regardless of ocean resolution (EP or NE), the prescribed SST fields originate from the governing physics of the real ocean.

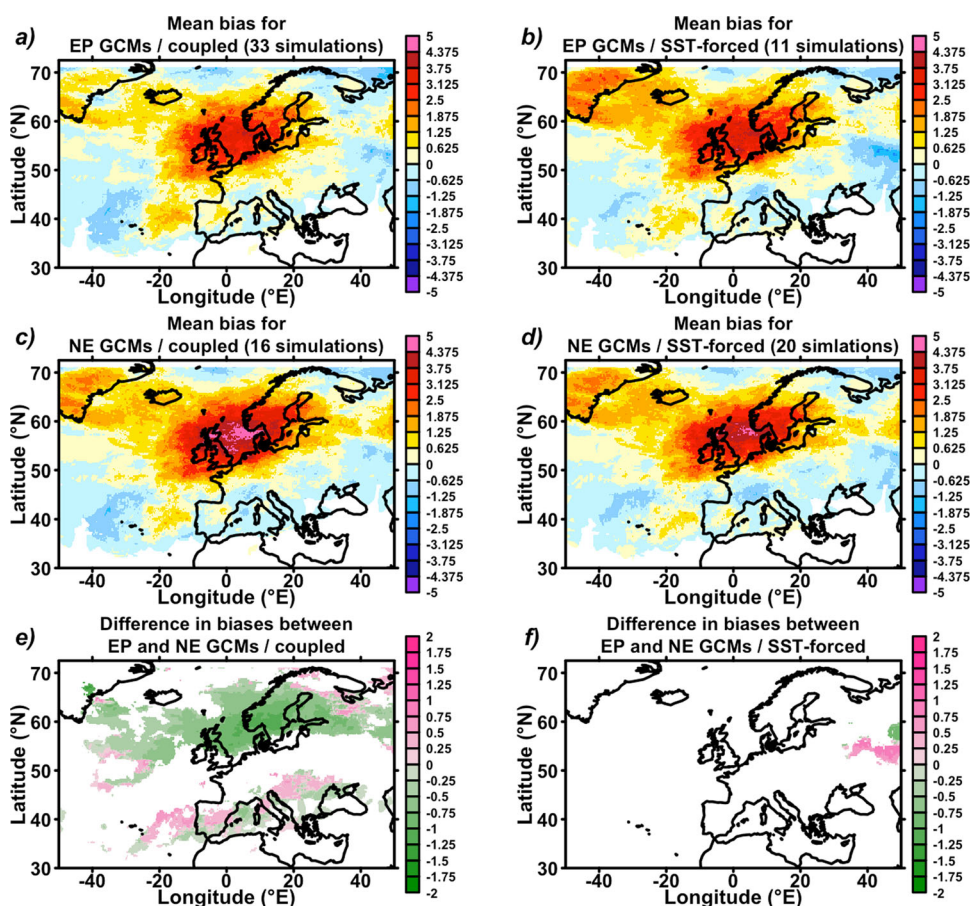


Fig. 4 Sensitivity analysis to atmospheric model resolutions. **a** Mean blocking frequencies biases for 16 coupled simulations of General Circulation Models (GCMs) with an eddy-permitting (EP) ocean model resolution. **b** Mean blocking frequencies biases for 11 atmospheric simulations with prescribed observed Sea Surface Temperatures (SSTs) for models with an EP ocean model resolution in their coupled versions. **c** Mean blocking frequencies biases for 33 coupled simulations of General Circulation Models (GCMs) with a no-eddy (NE) ocean model resolution. **d** Mean blocking frequencies biases for 20 atmospheric simulations with prescribed observed SSTs for models with a NE ocean model resolution in their coupled versions. For **a–d** biases are calculated with respect to ERA5 reanalyses data. **e** Differences in mean biases for coupled EP (**a**) and NE (**c**) GCM simulations. **f** Differences in mean biases for atmospheric simulations with prescribed SSTs for models with EP (**a**) and NE (**c**) ocean resolutions in their coupled versions. For (**e, f**), green colors indicate that EP GCMs have a significantly lower mean bias than NE GCMs at the 90% whereas pink colors indicate that NE GCMs have a significantly lower mean bias than EP GCMs at the 95% confidence level. White colors indicate observed blocking frequencies of 0% or no significant difference mean bias differences using a two-tailed Student's *t* test with a 90% confidence level.

Because EP GCMs frequently but not always have relatively high atmospheric model resolution (Table 1), and assuming that atmospheric model resolution is the primary reason for the improvement found for the set of coupled climate simulations, the significant difference in simulated blocking frequencies in EP and NE GCMs should also be seen in atmospheric-only simulations with prescribed SSTs. The 31 atmospheric-only simulations with prescribed SSTs were generated by 17 of the 24 coupled GCMs initially investigated (Table 2). Up to 11 of these simulations were obtained using the atmospheric component of GCMs with EP ocean grid resolution, whereas the remaining 20 were obtained using the atmospheric component of GCMs with NE ocean grid resolution (Table 2). For this analysis, we used the *amip* experiment simulations for GCMs from CMIP6 and *highresSST-present* experiment simulations for GCMs from HighResMIP (Table 1). First, the spatial distribution of biases for EP and NE sets of GCMs was computed and averaged for both coupled and atmosphere-only simulations (Fig. 4a–d). Then, for both coupled and atmosphere-only simulation sets, we calculated the difference in mean biases between EP and NE GCMs (Fig. 4e, f). We find a large number of grid points with a significant improvement at the 90% confidence level for GCMs with a coupled EP ocean model in the

set of 49 coupled GCM simulations (Fig. 4e). This is most noticeable north of 40°N, where the majority of blocking events occur (Fig. 1). In contrast, when computing the same results for the 31 atmosphere-only simulations with prescribed SST fields (Table 2), we clearly find no significant difference between models that have EP and NE ocean resolutions in their coupled version. An exception is the western Ural area, where the accuracy of simulated WEABs by models is known to be related to orographic model resolution²⁰. We thus conclude from the latter analysis that the improvement in simulated WEAB frequencies found for EP coupled GCMs is effectively due to the EP resolution of the ocean model rather than the generally higher atmospheric resolutions of these models.

Finally, and for the sake of robustness, since *historical* CMIP6 (29 simulations, Table 1) and *hist-1950* HighResMIP (20 simulations, Table 1) runs have slightly different forcings and initialization as explained at the beginning of the section, we also computed and compared the distributions of total WEAB frequencies for each of these experiments (Supplementary Fig. 6). We do not find any significant difference between WEAB frequencies obtained from simulations for the two experiments, where the obtained *p-value* is higher than 0.2. Therefore, since EP GCM simulations were

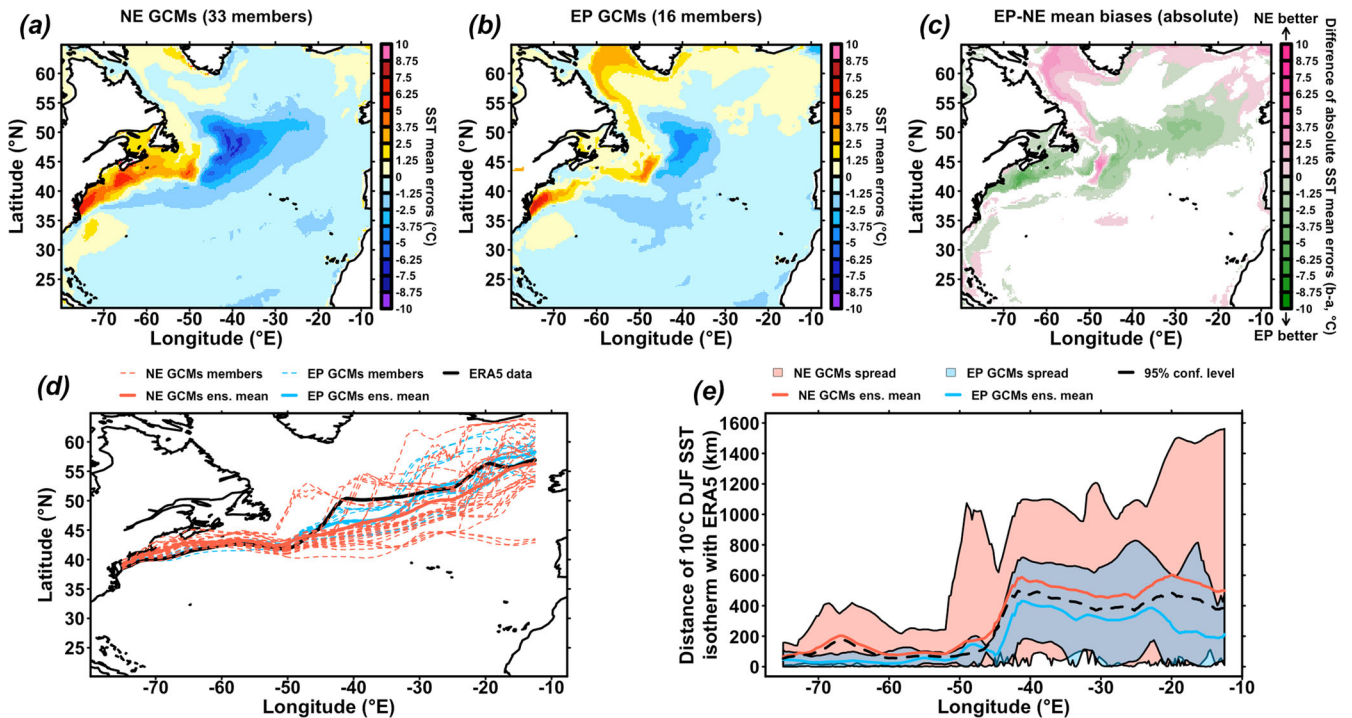


Fig. 5 Comparison of mean biases in North Atlantic sea surface temperature (SST) fields and North Atlantic Current position for no-eddy (NE) and eddy-permitting (EP) General Circulation Models (GCMs). **a, b** Mean December–January–February (DJF) SST biases from GCMs with NE and EP ocean grid resolutions with respect to ERA5 time-mean SSTs over the period 1979–2014, respectively. The GCM data were all interpolated on the ERA5 grid ($0.25^\circ \times 0.25^\circ$). **c** Absolute difference of EP and NE GCM biases. Positive values (pink) indicate that NE GCMs are less biased whereas negative values (green) indicate that EP GCMs are less biased. Uncolored grid points indicate no significant difference between EP and NE GCM biases using two-tailed Student’s *t* tests. **d** North Atlantic Current position estimated as the 10°C DJF SST isotherm from ERA5 reanalysis data (black solid line). Red (respectively blue) dashed lines give the estimated North Atlantic Current positions for NE (respectively EP) GCM simulations. Solid red (respectively blue) line gives the ensemble mean of estimated North Atlantic Current positions from NE (respectively EP) GCM simulations. **e** Distances between estimated North Atlantic Current positions from GCM simulations and ERA5 data, given in kilometers. The red (respectively blue) shaded area give the ensemble spread for NE (respectively EP) GCM simulations. The red (respectively blue) solid line gives the ensemble spread for NE (respectively EP) GCM simulations. The black dashed line gives the 95% confidence level estimated for EP GCM simulations calculated from a 5000 Monte Carlo sampling of 16 NE GCM simulations (same as the number of EP GCM simulations). White colored grid points indicate no significant difference between EP and NE GCM biases using two-tailed Student’s *t* tests.

included in both of the *historical* and *hist-1950* experiments (Table 1), this supports the significant improvement we obtained from considering GCM simulations made with EP ocean resolutions.

Analysis of oceanic and atmospheric bias reductions

In previous sections, we showed from a large ensemble of 49 GCM simulations that a significant improvement in simulated WEAB frequencies and locations is obtained for GCMs with an EP ocean model component (Figs. 1, 3). We also showed that it is not sensitive to the slightly different boundary conditions investigated (Table 1, Supplementary Fig. 6) or to the different resolutions of their atmospheric model component (Fig. 4). Given this significant improvement, we further explore the effect of eddies in simulating wintertime positions of the Gulf Stream and the North Atlantic Current in our GCM ensemble. Both were shown to be strongly biased in previous generations of GCMs^{17,40} and to contribute significantly to biasing North Atlantic SST mean states and hence the formation of WEABs^{16,18,39}.

The time-mean SST biases of NE and EP GCM simulations with respect to ERA5 reanalysis data are shown in Fig. 5a and b, respectively. Both the Gulf Stream and central North Atlantic areas have large SST biases in the two GCM ensembles, which is consistent with previous studies^{16,19}. However, we find significantly lower mean SST biases in GCMs with EP ocean resolution

in most areas, with a few exceptions (e.g., Labrador Sea, Fig. 5a–c), as noticed by a recent study based on a lower set of simulations³⁹. The observed and modeled trajectories of the North Atlantic Current are calculated as the 10°C SST isotherm in DJF¹⁶ and presented in Fig. 5d. It is shown that GCMs with EP ocean grids have a smaller deviation and an ensemble mean closer to the North Atlantic Current path from the reanalysis data than NE GCMs, as suggested by the biases in the mean SST (Fig. 5a–c). The distances between the simulated and ERA5 the North Atlantic Current positions are shown in Fig. 5e. The magnitude of the North Atlantic Current position bias reduction is estimated using Monte Carlo sampling of 5000 groups of 16 NE GCM simulations (i.e., with the same number as the EP GCM simulations). The Monte Carlo experiment confirms that the bias of large ocean currents’ positions in EP GCMs is significantly lower at the 95% confidence level in most regions of the North Atlantic, and most notably in the Gulf Stream extension as well as through the North Atlantic Current path (Fig. 5e). For more details, individual mean SST biases are presented in Supplementary Fig. 7 for EP GCM simulations and Supplementary Figs. 8, 9 for NE GCM simulations.

This finding confirms results from a recent study³⁹ that showed different reduced biases in atmospheric circulation due to reduced biases in SST patterns of central North Atlantic through more accuracy in simulated surface baroclinicity and diabatic heating. To investigate the link between the latter result and less biased

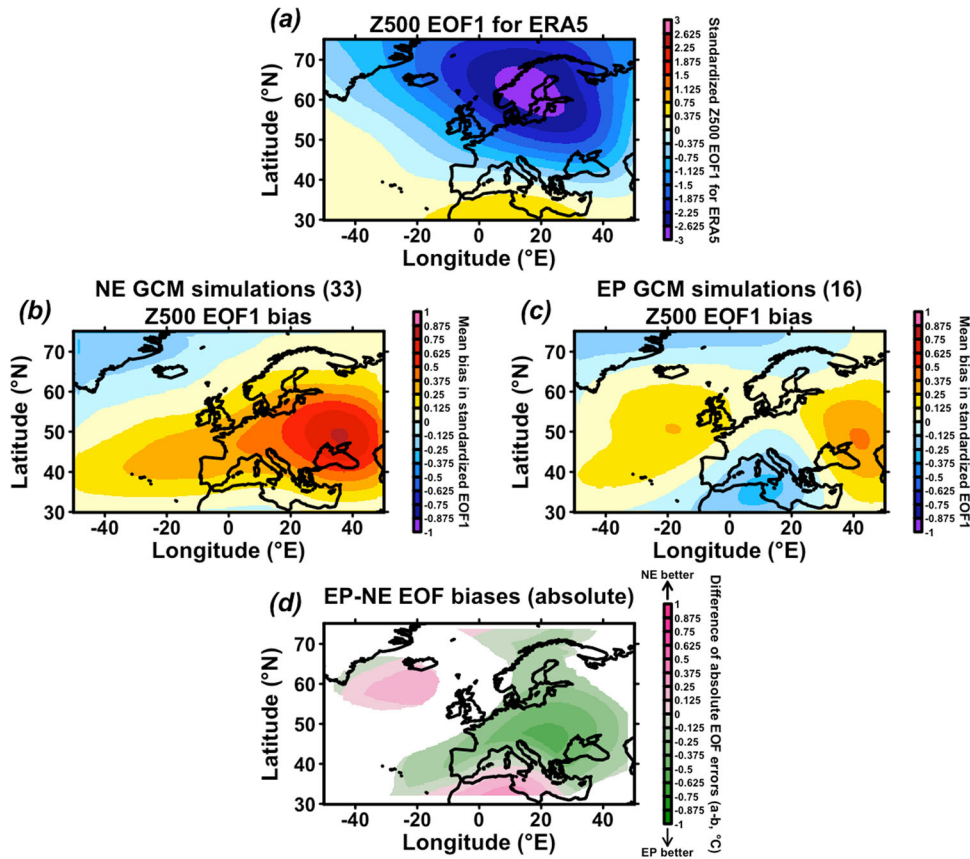


Fig. 6 Comparison of mean biases in 500 hPa geopotential heights (Z500) first Empirical Orthogonal Function (EOF1) for no-eddy (NE) and eddy permitting (EP) General Circulation Models (GCMs). **a** Standardized December–January–February (DJF) Z500 EOF1 for ERA5. **b, c** Mean biases of standardized ensemble mean Z500 EOF1 with respect to ERA5 standardized Z500 EOF1 over the period 1979–2014, for NE and EP GCM simulations, respectively. **d** Absolute difference of EP and NE GCM biases. Positive values (pink) indicate that NE GCMs are less biased whereas negative values (green) indicate that EP GCMs are less biased. White colored grid points indicate no significant difference between EP and NE GCM biases using two-tailed Student’s *t* tests.

high-tropospheric variability within the ensemble of GCMs studied here, we present an additional sensitivity analysis on the ability of EP and NE GCMs to accurately resolve patterns of variability in Z500 fields, which are used to detect atmospheric blocking events in most studies^{12,14,15} including the present one (Methods). This sensitivity analysis is led using an empirical orthogonal function (EOF) analysis⁴⁹. When considering the GCM biases for EP and NE ensembles (with respect to Z500 EOF1 from ERA5, Fig. 6a), we find significant improvements for EP GCMs in resolving Z500 fields (Fig. 6b, c) as NE GCMs have a particularly strong bias across Europe. The decrease in mean Z500 EOF1 biases when switching from NE to EP ocean grids is significant at the 95% confidence level for the majority of Europe (Fig. 6d), particularly over regions where improvement in WEAB frequencies was found in Figs. 1, 3, namely Western Europe and Baltic Sea. For more details, individual Z500 EOF1 are presented in Supplementary Fig. 10 for EP GCM simulations and Supplementary Figs. 11, 12 for NE GCMs simulations.

This analysis illustrates another strong bias reduction for high troposphere patterns of variability brought about by the use of EP ocean grids. This type of bias reduction could here be explained by better representations of the Gulf Stream and the North Atlantic Current (Fig. 5). Indeed, both the North Atlantic Current and the Gulf Stream were shown to influence wintertime weather patterns and storms over Europe^{19,50}, notably by forcing Rossby waves⁵¹.

DISCUSSION

The current study used an ensemble of 49 simulations from 24 GCMs and a clustering method to illustrate that EP GCMs outperform NE GCMs in terms of frequency and spatial extent of simulated WEABs (Fig. 3a–b), whereas increased atmospheric resolution also results in improvement^{21,22} (Fig. 3c). Further analyses of the GCM ensemble show that this gain in WEAB frequencies for EP GCMs is observed concomitantly with bias reduction in SST patterns related to better resolved Gulf Stream and North Atlantic Current wintertime positions (Fig. 5), as well as variability in high-tropospheric geopotential heights (Fig. 6) that is closely related to the activity of the midlatitude jet stream in winter. This finding may have an explanation in the fact that patterns of ocean surface flux anomalies towards the atmosphere are closely related to SSTs described by oceanic fronts⁵². Indeed, it has been demonstrated that in winter, the sharp SST gradient formed by the Gulf Stream and North Atlantic Current fronts contributes significantly to the release of latent and sensible heat fluxes to the atmospheric boundary layer⁵³. The heating and moistening of the atmosphere’s boundary layer increases buoyancy and low-level baroclinicity of air and causes a vertical transfer of heat via convection that remotely affect the jet stream’s mean state^{39,41}. As a result, the presence of mesoscale eddies in our ensemble of coupled climate simulations may explain the better resolving of WEABs and jet stream variability (Figs. 1, 3, 6). Indeed, the contribution of mesoscale eddies to surface temperature gradients and heat fluxes^{28,32}, as well as in better resolving Gulf

Stream and North Atlantic Current wintertime positions and related SST extrema (Fig. 5), may provide a consistent explanation for the current findings. In addition, it was recently shown that the speed and meridional position of the jet stream also leads SST patterns and heat fluxes by a few years⁵⁴. As a result, it may be possible that the global increase in accuracy of both simulated large-scale ocean and atmospheric features may have not been achieved independently but may result from generally better represented interactions between them. The accuracy of simulation of the above processes and interactions in the well-performing EP GCMs from this study compared to NE ones would provide a good framework for improving process-understanding of blocking formation, which is still a strong need⁸.

Despite the global lack of knowledge on blocking events, their frequencies are expected to decrease in the future because of Arctic amplification and subsequent reduced thermal advection, as well as reduced temperature contrasts between oceanic and continental areas^{8,55,56}. In contrast, a recent GCM ensemble analysis found no significant trends in WEAB frequencies for the SSP2-4.5 and SSP5-8.5 CMIP6 scenarios⁵⁷. However, the GCMs' poor ability to simulate blocking events¹⁴ remains a serious limitation in reconciling the various future evolution scenarios of atmospheric blocking events. In this respect, the findings of this work show that EP GCMs represent WEAB frequencies with substantially greater accuracy, and may thus be able to provide better weather extreme projections, which currently are subject to a substantial level of uncertainty, notably in terms of frequency of occurrence⁴.

Other types of prominent atmospheric blocking, such as summer blocking in western North America or Europe, could also benefit from the presented clustering-based investigation of their representativeness in GCMs. Similarly, possible improvements brought on by EP ocean resolutions could help in the investigation and adaptation to future blocking-related heatwave scenarios such as the recent exceptional ones that occurred in western Canada during summer 2021 and Europe during summer 2022.

Another major outlook from this research is the potential benefit of employing GCMs with an eddy-rich ocean resolution, which can be obtained for nominal horizontal resolutions of $0.1^\circ \times 0.1^\circ$ or higher^{30,33,34}. Unlike EP ocean grids, which only include mesoscale eddies with broader horizontal extents (100 km), eddy-rich ocean grids incorporate smaller-scale eddies (20–100 km), resulting in a significantly higher number of these mesoscale eddies and a better represented mean climate^{19,30,33–35,58}. The use of an eddy-rich grid was notably shown to further improve most benefits and bias reductions seen in EP GCMs, notably in terms of SST mean biases³³. Therefore, more progress in high-resolution coupled climate modeling with EP or eddy-rich ocean grid resolutions is crucial and strongly encouraged to better characterize blocking's past, current, and future activities.

METHODS

Detection of blocking situations

For the detection of atmospheric blockings, we use a 2D index based on geopotential height fields at 500 hPa (Z500) in a similar approach to previous studies^{12,14,37}. The Z500-based 2D index consists of determining for a given point on the atmospheric grid whether the meridional geopotential height gradient is reversed, indicating the mid-latitude intrusion of a large-scale high-pressure anomaly that the jet stream-transported air masses will bypass^{14,37}. This 2-D blocking index was originally proposed³⁷ as an extension of a previously published 1D index³⁸.

A large enough reversal in the meridional Z500 gradient for the grid point with longitude λ_0 and latitude ϕ_0 , is detected if the

following conditions are met:

$$C_1 : \text{GHGS}(\lambda_0, \phi_0) > 0 \quad (1)$$

$$C_2 : \text{GHGN}(\lambda_0, \phi_0) < -10 \quad (2)$$

$$C_3 : \text{GHGS}_2(\lambda_0, \phi_0) < -5 \quad (3)$$

Where GHGS, GHGN, and GHGS₂, expressed in meters per latitudinal degree, are given by

$$\text{GHGS}(\lambda_0, \phi_0) = \frac{Z500(\lambda_0, \phi_0) - Z500(\lambda_0, \phi_S)}{\phi_0 - \phi_S} \quad (4)$$

$$\text{GHGN}(\lambda_0, \phi_0) = \frac{Z500(\lambda_0, \phi_N) - Z500(\lambda_0, \phi_0)}{\phi_N - \phi_0} \quad (5)$$

$$\text{GHGS}_2(\lambda_0, \phi_0) = \frac{Z500(\lambda_0, \phi_S) - Z500(\lambda_0, \phi_{S_2})}{\phi_S - \phi_{S_2}} \quad (6)$$

With $\phi_S = \phi_0 - 15^\circ$, $\phi_N = \phi_0 + 15^\circ$, and $\phi_{S_2} = \phi_0 - 30^\circ$. The whole 2D index matrix (filled of 0 and 1 values) is obtained by checking C_1 , C_2 and C_3 for λ_0 varying over the longitude range covering the Euro-Atlantic sector (i.e. 50°W to 50°E), and ϕ_0 ranging from 30°N to 75°N.

Consistent with previous studies^{12,14}, we further consider a grid point to be effectively blocked if the blocking condition as calculated above lasts for at least 4 days. This avoids spurious detection of short-term and small-scale blocking situations that are not expected to have a significant impact on sub-seasonal climate variability. Here, the 2D blocking index is computed for the entire northern hemisphere.

Atmospheric blocking classification

A serious limitation of blocking indices as described in the above Methods section is that they do not account for the propagation and spatial structure of blocking events, since blocking situations are independently calculated for each grid point. Here, we use a *k-means* algorithm⁴⁶ to cluster the different DJF days of a given dataset, according to their similarities in terms of spatial structure. For this, we now consider $\mathbf{M} = \mathbf{M}(\phi, \lambda, t)$, $\phi \in \Phi$, $\lambda \in \Lambda$, $t = 1, \dots, n$ a matrix of detected blocking situations for a given Z500 field, as described in section 2.2. $\{\Phi, \Lambda\}$ describes the discretized horizontal spatial field of the climate grid, which is observed for a series of n regularly separated timesteps. Since we focus on the European region, we set that $\Phi \subset [-50^\circ\text{W}, 50^\circ\text{E}]$, $\Lambda \subset [30^\circ\text{N}, 75^\circ\text{N}]$, and denote $\#\{\Phi\} = q$ and $\#\{\Lambda\} = r$, where $\#\{\cdot\}$ is the cardinal operator. Therefore, n is the number of DJF days in each dataset, with $n = 3249$ for the period covered by the reanalysis data, whereas n varies slightly across GCMs depending on their simulation calendar (Table 1).

According to 2.2, we thus have that

$$\mathbf{M}(\phi, \lambda, t) = \begin{cases} 1 & \text{if } (\phi, \lambda) \text{ is blocked at time } t \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

For the *k-means* to be applied, the three-dimensional matrix \mathbf{M} is transformed into a two-dimensional matrix denoted $\mathbf{X} = (x_{ij})_{i,j} \in \mathbb{R}^{n \times p}$, with $p = qr$. This means that the columns of \mathbf{X} describe the ensemble of (ϕ, λ) pairs described by the climate grid. Thus, each column of \mathbf{X} is a time series of 1 and 0 values, describing the occurrence or not of a blocking situation for a given grid point (ϕ, λ) . In the same way, each row of \mathbf{X} describes a map of blocking situations over the $\{\Phi, \Lambda\}$ grid at a given timestep.

For the *k-means* method, k needs to be pre-determined at the beginning of the algorithm. The approach to select the optimal k parameter will be detailed in the next section. Considering a given k parameter, the *k-means* algorithms works as follows:

1. Draw k random initial position of the cluster centers denoted $\mu_1^{(1)}, \dots, \mu_k^{(1)}$.
2. Repeat until convergence (i.e., when clusters stop changing):
 - Assign all points from the X matrix to a cluster $c_j, j = 1, \dots, k$, whose center is the closest.
 - Update cluster centers:

$$\mu_j^{(t+1)} = \frac{1}{\#\{c_j^{(t)}\}} \sum_{i \in c_j^{(t)}} x_i, j = 1, \dots, k \quad (8)$$

Since the randomly drawn k -means initialization can affect the final clusters obtained by the algorithm, the above procedure is repeated 200 times and the one with the overall lowest distances to the clusters' centers is kept.

After the algorithm has run, and for simplification, final clusters are ordered according to their length (i.e., when related blocking events are the most occurring over the period of study), such that $\#\{c_1\} < \dots < \#\{c_k\}$.

Choice of k -means k parameter

Since we do not a priori know how many types of blocking events to classify, we use a Monte Carlo approach to estimate the best value for k . We start with $k = 2$ and increase it until the new type of blocking detected does not significantly differ from the blocking types previously clustered. Recursively, we check at step k whether the k^{th} blocking cluster were already detected at step $k-1$.

Here, we want to determine whether two given cluster centers $\mu_j \in \mathbb{R}^p$ and $\mu_k \in \mathbb{R}^p, j \neq k$, are significantly different, and thus c_j and c_k are considered as two clusters of different types of blocking events in terms of spatial structure. Thus, the null hypothesis we want to test is formulated, for a given α risk, as

$$H_0: \mu_j = \mu_k \quad (9)$$

$$H_1: \mu_j \neq \mu_k \quad (10)$$

For this, we create two ensembles of $R = 10,000$ vectors $\mu_j^{(r)}, \mu_k^{(r)} \in \mathbb{R}^p, r = 1, \dots, R$. $\mu_j^{(r)}$ and $\mu_k^{(r)}$ are vectors where values of μ_j and μ_k , respectively, are randomly shuffled. We will thus consider that we accept H_0 if the Euclidian distances between μ_j and μ_k is too small as compared to those calculated from the random pairs $\{\mu_j^{(r)}, \mu_k^{(r)}\}, 1 \leq r \leq R$. We denote distances as

$$d = \|\mu_j - \mu_k\|_2^2 \quad (11)$$

$$d^{(r)} = \|\mu_j^{(r)} - \mu_k^{(r)}\|_2^2, r = 1, \dots, R \quad (12)$$

The distances $(d^{(r)})_{1 \leq r \leq R}$ are then sorted and denoted $d_{(1)} \leq \dots \leq d_{(R)}$. In the general case, we consider that H_0 is accepted, i.e. $\mu_j \leq \mu_k$ at the α risk, if $\mathbb{P}_{H_0}[H_1] \leq \alpha$. Using the Monte Carlo approach, this condition is thus estimated to be verified if $d \geq d_{(1-R \times (1-\alpha))}$, where $[\cdot]$ is the integer part operator.

Attribution of models' blocking events

The same k -means procedure with the same number of clusters as determined for ERA5 (i.e. $k = 6$) is applied for all model outputs from Table 1. Then, spatial correlations between each group from the ERA5 reanalysis are computed with the clusters determined from each GCM. We here set that a blocking detected in a GCM is of same type than one from the reanalysis data if its spatial correlation (i.e., the correlation between the centers of the clusters) is higher than 0.5. If this threshold is reached for two blocking types of the reanalysis for the same cluster detected from the GCM, the one with the highest correlation is chosen. If for a given cluster detected from a GCM, the spatial correlation

threshold of 0.5 is not reached with any of the clusters from the reanalysis, the blocking frequency of the corresponding blocking type is set to 0. This threshold of 0.5 was set somewhat arbitrarily as using a regular correlation test is poorly restrictive and accept significantly matching patterns for much lower levels of spatial correlations due to large degrees of freedom. The sensitivity to this choice is addressed in Supplementary Figs. 3–5 where we found similar conclusions for different decision criteria.

Statistical Information

- Figure 2: For (f–o), a two tailed Student's t test is applied for each composite anomalies associated with the days clustered for each blocking type. Therefore, the degree of freedom used over the whole space is given by $n - 1$, where n is the size of clusters for each cluster: 247 for WE (f,k), 207 for Gr. (g,i), 187 for NS (h,m), 151 for BS (i,n), 118 for Sc. (j,o).
- Figure 3: For (a) the 95% confidence intervals are calculated using two-tailed Student's t test statistics. The degrees of freedom are $df = 48, df = 32$, and $df = 15$ for all GCMs (green), NE GCMs (red), and EP GCMs (blue), respectively. For each blocking type the Student's statistics are: $t = 16.1, t = 11.45$, and $t = 14.24$ for WE; $t = 15.23, t = 12.36$, and $t = 9.13$ for Gr.; $t = 7.38, t = 5.33$, and $t = 5.38$ for NS; $t = 10.39, t = 7.43$, and $t = 8.7$ for BS; $t = 6.45, t = 4.68$, and $t = 4.59$ for Sc. (for all, NE, and EP GCMs, respectively for each). For (b) and (c), the Monte Carlo experiments used are detailed in the main text. For the overall 33.8% higher WEAB frequencies for EP GCMs given in the main text, the applied test is a one-tailed Student's t test static for mean comparison of two samples with different lengths. The null hypothesis tests whether the increase in blocking frequencies (seen for all types) not significantly higher than 0. Student's statistics is $t = 3.2$ and the degree of freedom is $df = 160.23$. For the same comparisons but by blocking type in the main text, statistics are given by: $t = 2.15$ and $df = 40.3$ for WE, $t = 1.33$ and $df = 26.62$ for Gr., $t = 1.03$ and $df = 32.51$ for NS, $t = 2.66$ and $df = 30.99$ for BS, and $t = 1.25$ and $df = 28.35$ for Sc.
- Figure 5: For (a) and (b), the average of differences of mean DJF SSTs between ERA5 and each GCM for NE and EP GCMs, respectively. For (c), the applied test is a two-tailed Student's t test for the comparison of means from two samples of different sizes ($n_1 = 33$ and $n_2 = 16$) respectively. Therefore, the degree of freedom used over the whole space depends on empirical variances of GCM ensembles' means at each grid point (denoted s_1^2 for EP and s_2^2 for NE hereafter). These degrees of freedom are given for each grid point by $df = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{s_1^2/n_1}{n_1-1} + \frac{s_2^2/n_2}{n_2-1}}$. The Monte Carlo experiment used in (e) is detailed in the main text.
- Figure 6: (a) and (b) are obtained with a similar approach than Fig. 5a, b but considering the standardized Z500 EOF1 for ERA5 and GCMs. Similarly, (c) is obtained as in Fig. 5c but considering the standardized Z500 EOF1 for ERA5 and GCMs. Degrees of freedom are, therefore, the same as in Fig. 5c.

DATA AVAILABILITY

Initial data, pre-processed data, and post-processed data of the study have been made publicly available on the following Zenodo link: <https://zenodo.org/record/6755295#Yrta1y8Rrs0>. DOI: <https://doi.org/10.5281/zenodo.6755295>.

CODE AVAILABILITY

R and bash codes to produce the analysis and figures of the study have been made publicly available on the following Zenodo link: <https://zenodo.org/record/6755295#Yrta1y8Rrs0>. DOI: <https://doi.org/10.5281/zenodo.6755295>.

Received: 30 June 2022; Accepted: 16 May 2023;
Published online: 27 May 2023

REFERENCES

- Balting, D. F., AghaKouchak, A., Lohmann, G. & Ionita, M. Northern hemisphere drought risk in a warming climate. *npj Clim. Atmos. Sci.* **4**, 61 (2021).
- Cai, W. et al. Increased variability of eastern Pacific el niño under greenhouse warming. *Nature* **564**, 201–206 (2018).
- Hazeleger, W. et al. Tales of future weather. *Nat. Clim. Change* **5**, 107–113 (2015).
- Zhang, S. & Chen, J. Uncertainty in projection of climate extremes: a comparison of cmip5 and cmip6. *J. Meteorol. Res.* **35**, 646–662 (2021).
- Kautz, L.-A. et al. Atmospheric blocking and weather extremes over the Euro-Atlantic sector—a review. *Weather Clim. Dyn.* **3**, 305–336 (2022).
- Masato, G., Woolings, T. & Hoskins, B. J. Structure and impact of atmospheric blocking over the euro-atlantic region in present-day and future simulations. *Geophys. Res. Lett.* **41**, 1051–1058 (2014).
- Park, Y.-J. & Ahn, J.-B. Characteristics of atmospheric circulation over East Asia associated with summer blocking. *J. Geophys. Res.: Atmos.* **119**, 726–738 (2014).
- Woolings, T. et al. Blocking and its response to climate change. *Curr. Clim. Change Rep.* **4**, 287–300 (2018).
- Trigo, R. M., Garcia-Herrera, R. & Diaz, J. How exceptional was the early August 2003 heatwave in France? *Geophys. Res. Lett.* **47**, 2005GL022410 (2005).
- Cattiaux, J. et al. Winter 2010 in Europe: a cold extreme in a warming climate. *Geophys. Res. Lett.* **37**, 2010GL044613 (2010).
- Overland, J. Causes of the record-breaking pacific northwest heatwave, late June 2021. *Atmosphere* **12**, 1434 (2021).
- Davini, P., Cagnazzo, C., Gualdi, S. & Navarra, A. Bidimensional diagnostics, variability, and trends of northern hemisphere blocking. *J. Clim.* **25**, 6496–6509 (2012).
- Williams, P. D. et al. A census of atmospheric variability from seconds to decades. *Geophys. Res. Lett.* **44**, 11201–11211 (2017).
- 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).
- Davini, P. & D'Andrea, F. Northern hemisphere atmospheric blocking in global climate models: twenty years of improvements? *J. Clim.* **29**, 8823–8840 (2016).
- Scaife, A. et al. Improved Atlantic winter blocking in a climate model. *Geophys. Res. Lett.* **38**, L23703 (2011).
- Danabasoglu, G., Large, W. G. & Briebler, B. P. Climate impacts of parametrized Nordic sea overflows. *J. Geophys. Res. Oceans* **115**, 2010JC006243 (2010).
- O'Reilly, C. H., Minobe, S. & Kuwano-Yoshida, A. The influence of the gulf stream on wintertime European blocking. *Clim. Dyn.* **47**, 1545–1567 (2016).
- Moreno-Chamarro, E., Caron, L.-P., Ortega, P., Tomas, S. L. & Roberts, M. J. Can we trust cmip5/6 future projections of European winter precipitation? *Environ. Res. Lett.* **16**, 054063 (2021).
- Berckmans, J., Woolings, T., Demory, M.-E., Vidale, P.-L. & Roberts, M. Atmospheric blocking in a high resolution climate model: influences of mean state, orography and eddy forcing. *Atmos. Sci. Lett.* **14**, 34–40 (2013).
- Davini, P., Corti, S., D'Andrea, F., Rivière, G. & Hardenberg, J. Improved winter European atmospheric blocking frequencies in high-resolution global climate simulations. *J. Adv. Modeling Earth Syst.* **9**, 2615–2634 (2017).
- 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).
- Gollan, G., Bastin, S. & Greatbatch, R. J. Tropical precipitation influencing boreal winter midlatitude blocking. *Atmos. Sci. Lett.* **20**, 5 (2019).
- Flato, G. et al. *Evaluation of Climate Models*, Book Section 9, 741–866 (Cambridge University Press, 2013). www.climatechange2013.org.
- Delworth, T. L. et al. Simulated climate and climate change in the GFDL cm2.5 high-resolution coupled climate model. *J. Clim.* **25**, 2755–2781 (2012).
- Roberts, M. J. et al. Impact of resolution on the tropical Pacific circulation in a matrix of coupled models. *J. Clim.* **22**, 2541–2556 (2009).
- Shaffrey, L. C. et al. U.K. HiGEM: The new U.K. high-resolution global environment model-model description and basic evaluation. *J. Clim.* **22**, 1861–1896 (2009).
- Small, R. J. et al. A new synoptic scale resolving global climate simulation using the community earth system model. *J. Adv. Modeling Earth Syst.* **6**, 1065–1094 (2014).
- Eyring, V. et al. Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
- Hewitt, H. T. et al. Resolving and parameterising the ocean mesoscale in earth system models. *Curr. Clim. Change Rep.* **6**, 137–152 (2020).
- Dong, C., McWilliams, J. C., Liu, Y. & Chen, D. Global heat and salt transports by eddy movement. *Nat. Commun.* **5**, 3294 (2014).
- Small, R. J. et al. Air-sea interaction over ocean fronts and eddies. *Dyn. Atmos. Oceans* **45**, 274–319 (2008).
- Hewitt, H. T. et al. The impact of resolving the Rossby radius at mid-latitudes in the ocean: results from a high-resolution version of the met office GC2 coupled model. *Geosci. Model Dev.* **9**, 3655–3670 (2016).
- Jüling, A., von der Heydt, A. & Dijkstra, H. A. Effects of strongly eddying oceans on multidecadal climate variability in the community earth system model. *Ocean Sci.* **17**, 1251–1271 (2021).
- Moreno-Chamarro, E. et al. Impact of increased resolution on long-standing biases in highresmp-primavera climate models. *Geosci. Model Dev.* **15**, 268–289 (2022).
- Hersbach, H. et al. The era5 global reanalysis. *Q. J. Roy. Meteorol. Soc.* **146**, 1999–2049 (2020).
- Sherrer, S. C., Croci-Maspoli, M., Schwierz, C. & Appenzeller, C. Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in Euro-Atlantic region. *Int. J. Climatol.* **26**, 233–249 (2006).
- Tibaldi, S. & Molteni, F. On the operational predictability of blocking. *Tellus A* **42**, 343–365 (1990).
- Athanasiadis, P. J. et al. Mitigating climate biases in the midlatitude north Atlantic by increasing model resolution: Sst gradients and their relation to blocking and the jet. *J. Clim.* **35**, 3385–3406 (2022).
- Keeley, S. P. E., Sutton, R. T. & Shaffrey, L. C. The impact of north Atlantic sea surface temperature errors on the simulation of north Atlantic european region climate. *Q. J. Roy. Meteorol. Soc.* **138**, 1774–1783 (2012).
- Cronin, M. F. et al. Air-sea fluxes with a focus on heat and momentum. *Front. Mar. Sci.* **6**, 430 (2019).
- Cassou, C., Terray, L., Hurrell, J. W. & Deser, C. North Atlantic winter climate regimes. *J. Clim.* **17**, 1055–1068 (2004).
- Fabiano, F., Meccia, V. L., Davini, P., Ghinassi, P. & Corti, S. A regime view of future atmospheric circulation changes in northern mid-latitudes. *Weather Clim. Dyn.* **2**, 163–180 (2021).
- Michelangeli, P.-A., Vautard, R. & Legras, V. Weather regimes: recurrence and quasi stationarity. *J. Atmos. Sci.* **52**, 1237–1256 (1995).
- Terray, L., Demory, M.-E., Déqué, M., de Coetlogon, G. & Maisonnavé, E. Simulation of late-twenty-first-century changes in wintertime atmospheric circulation over Europe due to anthropogenic causes. *J. Clim.* **17**, 4630–4635 (2004).
- MacQueen, J. B. in *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability* (eds Le Cam, L. M. & Neyman, J.) 281–297 (University of California Press, 1967).
- Wazneh, H., Gachon, P., Laprise, R., de Varnal, A. & Tremblay, B. Atmospheric blocking events in the north Atlantic: trends and links to climate anomalies and teleconnections. *Clim. Dyn.* **56**, 2199–2221 (2021).
- Haarsma, R. J. et al. High resolution model intercomparison project (highresmp v1.0) for CMIP6. *Geophysical Model Development* **9**, 4185–4208 (2016).
- Preisendorfer, R. W. in *Principal Components Analysis in Meteorology and Oceanography* (ed Mobley, C. D.) Ch. 2, 11–81 (Elsevier, 1988).
- de Vries, H., Scher, S., Haarsma, R., Dijkstra, H. & van Delden, A. How gulf-stream sst-fronts influence atlantic winter storms. *Clim. Dyn.* **52**, 5899–5909 (2019).
- Minobe, S., Kuwano-Yoshida, A., Komori, N., Xie, S.-P. & Small, R. J. Influence of the gulf stream on the troposphere. *Nature* **452**, 206–209 (2008).
- Brachet, S. et al. Atmospheric circulations induced by a midlatitude sst front: a GVM study. *Journal of Climate* **25**, 1847–1853 (2012).
- O'Reilly, C. H., Minobe, S., Kuwano-Yoshida, A. & Woolings, T. The gulf stream influence on wintertime north atlantic jet variability. *Q. J. Roy. Meteorol. Soc.* **143**, 173–183 (2017).
- Ma, L., Woolings, T., Williams, R. G., Smith, D. & Dunstone, N. How does the winter jet stream affect surface temperature, heat flux, and sea ice in the north Atlantic? *J. Clim.* **33**, 3711–3730 (2020).
- Matsueda, M. & Endo, H. The robustness of future changes in northern hemisphere blocking: A large ensemble projection with multiple sea surface temperature patterns. *Geophys. Res. Lett.* **44**, 5158–5166 (2017).
- Matsueda, M., Mizuta, R. & Kusunoki, S. Future change in wintertime atmospheric blocking simulated using a 20-km-mesh atmospheric global circulation model. *J. Geophys. Res.—Atmos.* **D12**, D12114 (2009).
- Bacer, S. et al. Impact of climate change on wintertime European atmospheric blocking. *Weather Clim. Dyn.* **3**, 377–389 (2022).
- Roberts, M. J. et al. The benefits of global high resolution for climate simulation: process understanding and the enabling of stakeholder decisions at the regional scale. *Bull. Am. Meteorol. Soc.* **99**, 2341–2359 (2018).

ACKNOWLEDGEMENTS

This work was sponsored by NWO Exact and Natural Sciences for the use of SurfSARA supercomputer facilities under project 2020.022. This article is TiPES contribution No. 163; the TiPES (Tipping Points in the Earth System) project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 820970. We acknowledge the World Climate Research Program for coordinating and promoting CMIP6, including HighResMIP. We thank the different climate modeling groups for producing and making available their model output, and the Earth System Grid Federation (ESGF) for archiving the dataset and providing access through DRKZ and IPSL nodes (<https://esgf-data.dkrz.de/projects/esgf-dkrz/>, <http://esgf-node.ipsl.upmc.fr>). We thank ECMWF for providing the ERA5 dataset in Copernicus Climate Change Service (C3S) Climate Data Store (<https://cds.climate.copernicus.eu/cdsapp#!/home>).

AUTHOR CONTRIBUTIONS

S.M., H.D., and A.vdH. have designed the study. S.M. performed the different analysis and figures. S.M. has written the manuscript with large contributions from H.D. A.vdH. and R.vW. A.vdH., H.D., R.vW., and M.B. made suggestions to set up the manuscript's guiding thread, and contributed to assess the results.

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/s41612-023-00372-9>.

Correspondence and requests for materials should be addressed to Simon L. L. Michel.

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023