The representation of Northern Hemisphere blocking in current global climate models

Reinhard Schiemann[1], Panos Athanasiadis[2], David Barriopedro[3], Francisco Doblas-Reyes[4], Katja Lohmann[5], Malcolm J. Roberts[6], Dmitry Sein[7], Christopher D. Roberts[8], Laurent Terray[9], and Pier Luigi Vidale[1]

[1]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, United Kingdom
[2]Euro-Mediterranean Centre for Climate Change (CMCC), Bologna, Italy
[3]Department of Physics of the Earth, Astronomy and Astrophysics, Complutense University of Madrid, Madrid, Spain
[4]Earth Sciences Departement, Barcelona Supercomputing Center (BSC), Barcelona, Spain
[5]Max Planck Institute for Meteorology, Hamburg, Germany
[6]Met Office Hadley Centre, Exeter, United Kingdom
[7]Alfred Wegener Institute, Bremerhaven, Germany
[8]European Centre for Medium Range Weather Forecasting (ECMWF), Reading, United Kingdom
[9]Climat, Environnement, Couplages, Incertitudes, CECI, Université de Toulouse, CNRS, Cerfacs, Toulouse, France

Correspondence: Reinhard Schiemann (r.k.schiemann@reading.ac.uk)

Abstract. Global Climate Models (GCMs) are known to suffer from biases in the simulation of atmospheric blocking, and this study provides an assessment of how blocking is represented by the latest generation of GCMs. It is evaluated (i) how historical CMIP6 (Climate Model Intercomparison Project Phase 6) simulations perform compared to CMIP5 simulations, and (ii) how horizontal model resolution affects the simulation of blocking in the CMIP6-HighResMIP (PRIMAVERA) model ensemble, which is designed to address this type of question. Two blocking indices are used to evaluate the simulated mean blocking

5   frequency and blocking persistence for the Euro-Atlantic and Pacific regions in winter and summer against the corresponding estimates from atmospheric reanalysis data. There is robust evidence that CMIP6 models simulate blocking frequency and persistence better than CMIP5 models in the Atlantic and Pacific and in winter and summer. This improvement is sizeable so that, for example, winter blocking frequency in the median CMIP5 model in a large Euro-Atlantic domain is underestimated

10   by 32% using the absolute geopotential height (AGP) blocking index, whereas the same number is 19% for the median CMIP6 model. As for the sensitivity of simulated blocking to resolution, it is found that the resolution increase, from typically 100 km to 20 km grid spacing, in the PRIMAVERA models, which are not re-tuned at the higher resolutions, benefits the mean blocking frequency in the Atlantic in winter and summer, and in the Pacific in summer. Simulated blocking persistence, however, is not seen to improve with resolution. Our results are consistent with previous studies suggesting that resolution is one of a number

15   of interacting factors necessary for an adequate simulation of blocking in GCMs. The improvements reported in this study hold promise for further reductions in blocking biases as model development continues.

1 Introduction

Atmospheric blocking refers to the occurrence of quasi-stationary high pressure systems in the middle and high latitudes. Blocking highs persist for several days to weeks and divert cyclones pole- or equatorward. Preferred regions of blocking
20   occurrence are the eastern sides of the Atlantic and Pacific basins, and blocking events occur throughout the year. Blocking is associated with anomalous surface weather conditions such as cold spells in winter and heat waves in summer, and these surface impacts can be hazardous especially for long-lasting blocking events (e.g., Barriopedro et al., 2010; Rex, 1950; Woollings, 2010; Cattiaux et al., 2010; Barriopedro et al., 2011; Matsueda, 2011; Otto et al., 2012).

     Global climate models (GCMs) tend to underestimate blocking, and these biases have been long standing and documented
25   for different phases of the Atmospheric/Coupled Model Intercomparison Project (AMIP/CMIP) efforts (D'Andrea et al., 1998; Masato et al., 2013; Anstey et al., 2013). Too coarse horizontal resolution of the atmospheric grid has been put forward as one of the factors limiting the accurate represenation of blocking in GCMs and several studies show how an increase in resolution benefits the simulation of blocking. At the same time, the magnitude of the improvements seen at higher resolution varies considerably between different studies due to issues including the GCM(s) considered, their resolutions, and blocking indices
30   used for evaluation as well as the ensemble size employed and simulation period covered to sample internal variability of blocking. For example, Matsueda et al. (2009) report a dramatic improvement in the simulated Euro-Atlantic blocking as resolution is increased from 180 km to 20 km in an atmospheric GCM (AGCM), whereas Anstey et al. (2013) and Schiemann et al. (2017) show more modest improvements with resolution across the CMIP5 simulations, and in a 4-model ensemble of AMIP simulations with grid spacings down to 20 km, respectively. Also, Davini et al. (2017) show that in their AGCM a
35   very good representation of Euro-Atlantic blocking at grid spacings of 40 km or smaller is the result of compensating biases between too strong eddies at upper and too weak eddies at lower levels. In addition to horizontal resolution, a number of other factors have been shown to be important for blocking simulation including sea surface temperature and associated mean-state biases, vertical resolution, orographic boundary conditions, physcial parameterisations, and the numerical scheme employed in the model's dynamical core (Woollings et al. (2018) and references therein including Scaife et al. (2011), Anstey et al. (2013),
40   Berckmans et al. (2013), Jung et al. (2012), Pithan et al. (2016), Williams et al. (2018)).

     The aim of this paper is twofold: First, recently available model simulations following the CMIP6-HighResMIP protocol, delivered by the EU Horizon 2020 PRIMAVERA project, offer a new set of controlled experiments designed to address the sensitivity to model resolution in a coordinated multi-model ensemble (see Sect. 2.1). Second, we adopt a slightly longer-term view and compare blocking biases in the latest generation of high-resolution HighResMIP GCMs with those seen in historical
45   CMIP6 and CMIP5 simulations, which serves to put the HighResMIP results in context and appears timely given both the very recent availability of CMIP6-HighResMIP simulations and the potential need for up-to-date model evaluations underpinning, for example, the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report. We conduct all evaluations for two different blocking indices so as to assess the robustness of our results with respect to this choice. In addition to evaluating the mean blocking climatology, we also evaluate the representation of blocking persistence, which appears to be missing from
50   the recent literature.

The structure of this paper is as follows: In Sect. 2, we introduce the multi-model ensembles that are evaluated in this study and the reference reanalysis blocking climatology. We also introduce the two blocking indices and describe how the persistence analysis is conducted for each of these indices. The three following sections report on the results of our evaluations, namely the spatial distribution of simulated blocking and biases in Sect. 3, the quantification of domain-mean blocking biases and how they depend on model resolution in Sect. 4, and the evaluation of blocking persistence in Sect. 5. The paper is concluded in Sect. 6.

2 Data and methods

2.1 Multi-model ensembles and experiments

We use the simulations delivered by the PRIMAVERA (PRocess-based climate sIMulation: AdVances in high resolution modelling and European climate Risk Assessment) project to assess the sensitivity of simulated blocking to model resolution. These simulations follow the CMIP6-HighResMIP protocol (Haarsma et al., 2016) and are designed to test how the simulation of a range of phenomena in the climate system depends on model resolution in the atmosphere and ocean. The high-resolution versions of the PRIMAVERA models have therefore been re-tuned as little as possible with respect to their low-resolution counterparts. This is in line with the HighResMIP philosophy that prioritises being able to attribute any changes in model performance to the direct effect of resolution change over building optimally tuned models with as small as possible biases. For the evaluation of how simulated blocking is sensitive to resolution presented here, this implies that our results should be considered to be conservative as mean-state circulation biases are known to statistically explain a large part of, yet not all, blocking biases seen in climate models (Scaife et al., 2010; Schiemann et al., 2017).

The PRIMAVERA models and simulations used in this study are overviewed in Table 1. The simulation period is 1950–2014 and we evaluate historical simulations driven by observed greenhouse-gas and aerosol concentrations both in a coupled ocean-atmosphere-land-sea ice setup (HighResMIP *hist-1950* experiment) and in an AMIP-style setup driven by historically observed sea-surface temperature and sea-ice concentrations (HighResMIP *highresSST-present* experiment). Further details of the experimental setup and a baseline evaluation focusing on coupled aspects of the HadGEM3-GC3.1 model are provided by Roberts et al. (2019).

In addition to the PRIMAVERA simulations, we evaluate the representation of blocking in one historical simulation with each of 29 CMIP5 models for the period 1950–2005 (Taylor et al., 2012) covering the period 1950–2005, and for one historical simulation with each of 13 CMIP6 models for the period 1950–2014 (Eyring et al., 2016).

2.2 Observed blocking

The reference data for evaluating model-simulated blocking is based on both the ERA-40 (Uppala et al., 2005) and ERA-Interim reanalyses (Dee et al., 2011). Following Schiemann et al. (2017), we concatenate data from these two reanalyses to obtain a 50-year reference climatology covering the period 1962–2011 so as to reduce the impact of blocking internal variability

Table 1. PRIMAVERA models and simulations. Columns detail the model name, the atmosphere grid spacing at 50°N, nominal ocean grid spacing, a sub-ensemble indicator ('LF' — low-resolution forced (AMIP), 'LC' — low-resolution coupled, 'HF' — high-resolution forced, and 'HC' — high-resolution coupled), and the number of ensemble members used in this study.

No.	Model	Atm. grid (km)	Ocean grid (km)	Sub-ensemble	Members
1	AWI-CM-1-1-LR	129	50	LC	1
2	AWI-CM-1-1-HR	67	25	HC	1
3	CMCC-CM2-HR4	64	-	LF	1
4	CMCC-CM2-VHR4	18	-	HF	1
5	CMCC-CM2-HR4	64	25	LC	1
6	CMCC-CM2-VHR4	18	25	HC	1
7	CNRM-CM6-1	142	-	LF	1
8	CNRM-CM6-1-HR	50	-	HF	1
9	CNRM-CM6-1	142	100	LC	1
10	CNRM-CM6-1-HR	50	25	HC	1
11	EC-Earth3P	71	-	LF	1
12	EC-Earth3P-HR	36	-	HF	2
13	EC-Earth3P	71	100	LC	3
14	EC-Earth3P-HR	36	25	HC	3
15	ECMWF-IFS-LR	50	-	LF	8
16	ECMWF-IFS-HR	25	-	HF	6
17	ECMWF-IFS-LR	50	100	LC	8
18	ECMWF-IFS-MR	36	25	HC	3
19	ECMWF-IFS-HR	25	25	HC	6
20	HadGEM3-GC31-LM	135	-	LF	5
21	HadGEM3-GC31-MM	60	-	HF	3
22	HadGEM3-GC31-HM	25	-	HF	3
23	HadGEM3-GC31-LL	135	100	LC	8
24	HadGEM3-GC31-MM	60	25	HC	1
25	HadGEM3-GC31-HM	25	25	HC	3
26	HadGEM3-GC31-HH	25	8	HC	1
27	MPI-ESM1-2-HR	67	-	LF	1
28	MPI-ESM1-2-XR	34	-	HF	1
29	MPI-ESM1-2-HR	67	40	LC	1
30	MPI-ESM1-2-XR	34	40	HC	1

on our results. Schiemann et al. (2017) also show that these two reanalyses, as well as the MERRA reanalysis (Rienecker et al., 2011), agree very well with each other on the mean and interannual variability of blocking over different domains implying that for the purposes of this paper reanalysis uncertainty can be considered to be small compared to internal variability.

2.3 Blocking indices

A considerable number of blocking indices have been employed by different authors, and these indices emphasise different aspects of the blocking phenomenon and use different meteorological variables (see, e.g., Barriopedro et al., 2010, for an overview) so that it is advantageous to use more than one blocking index to assess the robustness of model evaluation results to this choice (e.g., Woollings et al., 2018). One fundamental distinction is between blocking indices based on the exceedance of an absolute (fixed) threshold of a meteorological variable and indices based on the detection of anomalies (departures) of a meteorological variable from a climatological mean. Here, we use one index from each of these two groups, namely the so-called absolute geopotential height (AGP) index described in Sect. 2.3.1 and the anomaly index (ANOM) described in Sect. 2.3.2. We calculate both of these indices from daily-mean 500hPa geopotential height data for the simulations and reanalysis data introduced in Sections 2.1 and 2.2.

2.3.1 Absolute Geopotential Height index

The AGP index is a generalisation of the one-dimensional index by Tibaldi and Molteni (1990) to two dimensions (Scherrer et al., 2006). According to the AGP index, three conditions need to be fulfilled for a point at latitude ϕ_0 to be identified as blocked. The first condition is a reversal of the climatological equator-to-pole gradient of the 500hPa geopotential height Z to the south of ϕ_0:

$$\frac{Z(\phi_0) - Z(\phi_\mathrm{S})}{\phi_0 - \phi_\mathrm{S}} > 0 \,, \tag{1}$$

where ϕ_S is 15° south of ϕ_0. The second condition requires westerlies to the north of ϕ_0:

$$\frac{Z(\phi_\mathrm{N}) - Z(\phi_0)}{\phi_\mathrm{N} - \phi_0} < -10\,\mathrm{m}\,(°\mathrm{latitude})^{-1} \,, \tag{2}$$

where ϕ_N is 15° north of ϕ_0. The third condition is that the point is only considered blocked if the first two conditions are met for five consecutive days or more. All model and reanalysis fields are regridded to a common $1.875° \times 1.25°$ grid before the blocking identification is applied, and we calculate the blocking index for all grid boxes between 35°N and 75°N. This index has been used in previous evaluations of blocking in multi-model ensembles (Anstey et al., 2013; Schiemann et al., 2017).

2.3.2 Anomaly index

The ANOM index (following Woollings et al. (2018) and similar to Sausen et al. (1995) and Schwierz et al. (2004) but using 500hPa geopotential height Z500) is based on tracking geopotential height anomalies. The following steps are carried out in its calculation:

1. Daily Z500 data are regridded to a common 2.5° grid and a 31-day running mean is calculated through a baseline period (1981–2010) and a daily Z500 climatology is obtained by taking the mean over the baseline period for each day.

2. A daily anomaly is calculated, separately for each month, by taking the difference between the original Z500 data and the climatology from step 1 for the corresponding day. A monthly anomaly threshold is then obtained by calculating the

115   90th percentile of these differences throughout 50–80°N. The monthly anomaly threshold is smoothed further with a 3-month rolling mean.

3. For each day, potential blocking events are identified as contiguous areas of at least 10^6 km^2 where the Z500 anomaly (as in step 2) exceeds the monthly anomaly threshold (also as in step 2).

4. The candidate events from step 3 are further screened by requiring a spatial overlap of at least 50% between consecutive

120   days (quasi-stationarity) for at least 5 days (minimum persistence).

2.4   Spatial aggregation and persistence analysis

Spatially averaged metrics of blocking performance are calculated for domains centered over the North Atlantic (ATL) and North Pacific (PAC), respectively. The definition of these domains reflects the spatial patterns of AGP and ANOM climatologies and biases (see Sect. 3) and we choose about the same ATL domain for both indices (-90–90°E, 50–75°N for AGP and -90–

125   90°E, 50–90°N for ANOM) but somewhat different domains for PAC (90–270°E, 50–75°N for AGP and 120–240°E, 40–90°N for ANOM).

For the analysis of blocking persistence (Sect. 5), both blocking indices are used as described in Sections 2.3.1 and 2.3.2 but the persistence criterion of 5 days is relaxed so that blocking events of any persistence, including so-called instantaneous blocking events on a single day, are included in the analysis.

130   For the locally defined AGP index, defining the persistence of blocking at the grid-box scale is not meaningful and spatial aggregation is necessary before the persistence analysis is carried out. We aggregate here over 12 sectors of 30° longitude in 50–75°N and a sector is said to be blocked on a given day if at least 10% of the sector area is blocked according to the AGP index. Persistence analysis results shown for the ATL and PAC domains are then average results for the corresponding sectors. For the ANOM index based on tracking spatially extended geopotential height anomalies no such spatial aggregation

135   is necessary and the persistence analysis is carried out using the persistence of these anomalies directly.

For both indices, we determine the empirical survival function ESF(t), i.e. the probability of a blocking event to persist for at least t days. Quantiles of ESF(t) are estimated using the nonparametric Kaplan-Meier estimator for the AGP index, whereas for the ANOM index a parametric exponential fit was found to work well. See, e.g., Tableman et al. (2003), for details, noting that our application is much simpler than a typical survival analysis as there are no censored observations.

140   # 3   Geographical distribution of blocking occurrence

Maps of time mean winter (DJF) AGP blocking frequency are shown in Fig. 1. For an overview of the effect of resolution and coupling, the blocking climatology is shown separately for four PRIMAVERA sub-ensembles in Fig. 1a–d. In the sub-ensembles, simulations have been grouped into (i) low-resolution forced ('LF'), (ii) low-resolution coupled ('LC'), (iii) high-resolution forced ('HF'), and (iv) high-resolution coupled ('HC') simulations as shown in Table 1. Blocking climatologies for

145   the CMIP5 and CMIP6 ensembles are shown in Fig. 1e,f.

All of the model ensembles considered show widespread underestimation of the climatological blocking frequency in the Pacific and especially in the Euro-Atlantic region. These biases are pervasive across the different ensembles as indicated by the stippling showing agreement among models on the sign of the blocking frequency biases. Closer inspection also shows differences between model ensembles. CMIP5 biases are spatially similar to those of the CMIP6 models, yet the multi-model

150 mean CMIP6 bias is smaller than the CMIP5 bias throughout the Euro-Atlantic region, whereas this difference is smaller for the Pacific (Fig. 2e,f). Multimodel-mean biases for the high-resolution PRIMAVERA models are smaller than for the low-resolution PRIMAVERA models (Fig. 1a,b vs. Fig. 1c,d). This improvement with resolution is seen for both AMIP and coupled simulations over the Euro-Atlantic region, and also for the Pacific in the coupled simulations.

In summer, blocking is observed throughout a wide high-latitude region ranging from Greenland across northern Eurasia

155 to Alaska (Fig. 2) so that the distinction between Atlantic and Pacific blocking is not as clear as in winter. As in winter, all model ensembles underestimate the blocking frequency everywhere and this bias is pervasive across models. There are small improvements in this bias in CMIP6 over CMIP5 in the Baltic region and over Siberia, with little change elsewhere (Fig. 2e,f). When comparing the different PRIMAVERA sub-ensembles, blocking is again seen to improve at the higher resolution, and this improvement is seen more clearly and includes the Pacific region in the coupled simulations (Fig. 2a–d).

160 Repeating the same analyses with the ANOM blocking index (Figures S1 and S2 in the supplement) shows results that largely agree with those based on the AGP index. All model ensembles are found to underestimate the occurrence of blocking both in the Euro-Atlantic and Pacific regions, and both in winter and in summer. There are also small improvements from CMIP5 to CMIP6 over the Atlantic in winter and summer, whereas the difference between the CMIP5 and CMIP6 biases is small for the Pacific. Small improvements with resolution are also seen in the PRIMAVERA ensemble, yet these appear

165 smaller, as a fraction of total blocking frequency, than for the AGP index and are not seen for the Pacific.

4 Sensitivity to model resolution

We proceed in this section with a quantitative evaluation of how different metrics of blocking performance depend on atmospheric model resolution. Given the comparatively large biases seen in the Euro-Atlantic sector, we focus on the ATL domain and use the AGP index in the main manuscript (Figures 3 and 4) but also include results for the PAC domain and the ANOM

170 index in the supplement (Figures S3–S8).

4.1 Atlantic

The evaluation for the ATL domain in winter using the AGP index is shown in Fig. 3. The six panels show different metrics of blocking performance (domain-mean blocking frequency — top row, spatial correlation with the reanalysis climatology across the domain — middle row, root-mean-square error with respect to the reanalysis climatology — bottom row), for PRIMAVERA

175 coupled (left-hand panels) and AMIP simulations (right-hand panels). Each panel is divided into two parts; on the left the blocking metric is plotted vs. resolution for each of the PRIMAVERA models, using the ensemble-mean metric if more than one simulation is available for a given model/resolution (Table 1). On the right, the distribution of the same metric is shown

in terms of a boxplot for each of the CMIP5 and CMIP6 ensembles, and the reanalysis estimate is shown by the '*' symbol labelled 'ERA/IV' on the x-axis. All estimates of metrics of blocking performance shown are subject to internal (sampling) variability and a model-based estimate of this internal variability is shown by the boxplot labelled 'ERA/IV'. These boxplot statistics are obtained by forming 15 pairs of simulations from the 6-member ensembles available for the ECMWF-IFS-HR model (No. 16 and 19 in Table 1) and quantifying the agreement on the blocking metric in these pairs.

Atlantic winter blocking is seen to be systematically underestimated by nearly all of the CMIP5, CMIP6, and PRIMAVERA simulations analysed (Fig. 3). For the ATL domain and AGP index chosen, this bias ranges from less than half of the observed blocking frequency for models simulating very little blocking to an underestimation of around 10-20% for models simulating more frequent blocking. There is a systematic shift from CMIP5 to CMIP6 models showing a better agreement of the CMIP6 models with reanalysis data for all three metrics considered. With the exception of one or two models, there is also a general tendency of the coupled PRIMAVERA simulations for an improved simulation, i.e. more frequent blocking, a higher spatial correlation, and a smaller root-mean-squared error, at the higher resolutions. This systematic improvement with resolution is not seen in the AMIP simulations for which the sensitivity of blocking performance to resolution as well as the variation in blocking performance between different models is smaller than for the coupled simulations. The AMIP simulations show similar or slightly better blocking performance than the corresponding coupled simulations for most models with the notable exception of the Hadley Centre model (HadGEM3-GC3.1). Repeating the same analysis with the ANOM index corroborates the pervasive underestimation of ATL winter blocking, by typically 40-10% with this index (Fig. S3). The improvement from CMIP5 to CMIP6 is also seen with the ANOM index. The improvement with resolution is not as clear as for the AGP index, though the RMSE is seen to reduce at higher resolution for 5 out of 7 models (Fig S3e). This difference between the ANOM and AGP results appears plausible when considering how these two indices are defined (Sections 2.3.1 and 2.3.2). While the AGP index identifies blocked situations in terms of the exceedance of fixed thresholds of absolute Z500 (gradients), the ANOM index identifies blocked situations through the exceedance of thresholds defined as quantiles of the model's own large-scale Z500 variability about the model's Z500 mean. In this way, model biases in Z500 mean and variability are partly excluded as a potential source of blocking bias in the ANOM index, and likewise any improvement/deterioration of Z500 mean and variability with resolution will not be fully reflected in an improvement/deterioration in the ANOM blocking index.

The evaluation for the ATL in summer yields similar results to those obtained in winter (Fig. 4). The vast majority of models underestimates blocking, and there is an improvement from CMIP5 to CMIP6. All but one of the PRIMAVERA models are seen to have a smaller mean blocking bias at the higher resolution and with the exception of one model the blocking frequency and spatial correlation also show improvements at higher resolution. As in winter, the sensitivity to resolution is small in the AMIP simulations. The performance of most coupled and corresponding AMIP simulations is similar for most models. The evaluation using the ANOM index (Fig. S4) confirms these results, yet as in winter the sensitivity to resolution is smaller than for the AGP index.

4.2 Pacific

The evaluation for the PAC domain and the AGP index (Fig. S5) shows that the domain-mean winter blocking frequency is similar in models and reanalysis, though this is partly due to error compensation within the domain (see Fig. 1). As in the ATL domain, there an improvement is seen from the CMIP5 to the CMIP6 models for all blocking performance metrics. There is no robust improvement with resolution across the PRIMAVERA ensemble, yet 5 of 7 coupled models do show a decrease in RMSE as the resolution is increased. The corresponding evaluation using the ANOM index agrees with these results, i.e. a small improvement from CMIP5 to CMIP6 and no sensitivity of blocking performance to resolution (Fig. S7).

Turning finally to the evaluation for the PAC domain in summer, we find using the AGP index that there is an improvement in simulated blocking in CMIP6 over CMIP5, and there is alo some suggestion of an improvement with resolution in most of the models (Fig. S6). The ANOM index confirms the improvement in CMIP6 over CMIP5. Interestingly, according to this index we also find a clearer improvement at higher resolution than seen with the AGP index (see especially Fig. S8c,e). Referring to the maps of blocking bias shown in Fig. S2b,d, this may be due in part to an improvement at very high latitudes where the AGP index is not defined.

5 Blocking persistence

In this section we assess how the persistence of blocking is simulated. This section is organised in a similar way to Sect. 4 with a focus on the ATL domain and the AGP index in the main manuscript (Figures 5 and 6) but results for the PAC domain and using the ANOM index available in the supplement (Figures S9–S14).

5.1 Atlantic

The evaluation of winter blocking persistence for the ATL domain and using the AGP index is shown in Fig 5. This figure is organised similarly to the figures in Sect. 4, but here two metrics for blocking persistence, the median persistence of blocking events (top row) and the 90th percentile of blocking persistence (bottom row) are shown.

Models in all of the CMIP5, CMIP6, and PRIMAVERA ensembles tend to underestimate blocking persistence both in coupled and AMIP experiments, by typically 10-15% both for the median and 90th percentile. An improvement towards longer blocking events is seen in the CMIP6 ensemble over CMIP5, though this improvement is not as large, compared to the ensemble spread, as was found in Sect. 4 in the evaluation of mean blocking frequency (Fig. 3). There is a small increase in the 90th percentile of simulated blocking persistence for most of the PRIMAVERA models as resolution is increased, yet no systematic sensitivity to resolution is seen for the median persistence (for which also interval variability is comparatively large) nor for the AMIP simulations, as already seen for mean blocking sensitivity to resolution. These findings are corroborated in terms of the corresponding analysis with ANOM index, except for the fact that there is no systematic sensitivity to resolution in blocking persistence when using this index.

240   Moving on to summer (Fig. 4), we see that models both underestimate and overestimate blocking persistence, with a general
tendency of no or a small positive bias for the median, and a small negative bias for the 90th percentile, potentially indicating
a different shape of the simulated blocking survival functions, namely a slightly faster decrease with surival time (persistence)
in the simulations than in the reanalysis. A small improvement, mainly in the sense of a smaller ensembe spread, can be seen
in CMIP6 over CMIP5. There is no systematic sensitivity of simulated blocking persistence to resolution in the PRIMAVERA
245   ensemble. The analyses in terms of the ANOM index are consistent with these results (Fig. S10), noting that here the plots for
the median and the 90th percentile are just scaled versions of one another due to the choice of the exponential fit.

5.2   Pacific

Pacific winter blocking persistence tends to be overestimated by the coupled PRIMAVERA, CMIP5, and CMIP6 models,
whereas the PRIMAVERA AMIP simulations scatter around the reanalysis estimate (Fig. S11). There is an improvement
250   from CMIP6 to CMIP5 towards shorter blocking events, but not evidence for sensitivity of simulated blocking persistence to
resolution in the PRIMAVERA models. These results are corroborated when using the ANOM index (Fig. S13).

In summer, blocking in the PAC domain is slightly underestimated by most models according to the AGP index (Fig. S12).
No sensitivity to resolution is seen for the PRIMAVERA simulations. The spread of the CMIP6 ensemble about the reanalysis
estimates is smaller than for CMIP5, which constitutes and improvement. In the case of Pacific summer blocking, the ANOM
255   analysis does not confirm the results obtained with the AGP index (Fig. S14). There is a small overestimation of blocking
persistence in most models according to this index, and no systematic difference between the CMIP5 and CMIP6 ensembles.
Interestingly, the PRIMAVERA AMIP simulations show a reduction of blocking persistence at higher resolution. This effect
is small, but seen in all models and constitutes an improvement for most models. As already argued in Sect. 4, the apparent
inconsistency between the AGP and ANOM index may be due to the inclusion of very high-latitude areas in the ANOM
260   index. Furthermore, the interpretation of the ANOM analysis in particular is complicated as it appears to be affected by error
cancellation within the PAC domain with fairly small net results (see also Fig. S2).

6   Summary and conclusions

Climate model simulations suffer from long-standing biases in the representation of atmospheric blocking hampering applica-
tions of these simulations in assessing present and future climate impacts associated with blocking such as winter cold spells
265   and summer heat waves. In this study, we revisit the ability of state-of-the-art climate models to represent atmospheric blocking.
This analysis is timely due to the recent availability of CMIP6 simulations, including those following the CMIP6-HighResMIP
protocol designed to assess the role of model resolution. Our aims are to (i) compare the performance of blocking simulation
in CMIP6 and CMIP5 models assessing the net effect of model development between these two generations of multi-model
ensembles, and (ii) to assess the sensitivity of simulated blocking to model resolution specifically, using the models/simulations
270   developed in the PRIMAVERA project following the HighResMIP protocol.

Concerning our first aim, we find a very clear improvement in simulated blocking in the CMIP6 model ensemble over the CMIP5 ensemble. This improvement is seen robustly for different metrics of mean blocking frequency and blocking persistence, for the Euro-Atlantic and Pacific regions, for winter and summer, and using two different blocking indices (AGP and ANOM) to identify blocking events. The magnitude of the improvement seen depends on the region, season, and blocking

275 index/metric considered — as does the magnitude of the bias itself — yet it is sizeable when compared to the spread of the multi-model ensembles and the total magnitude of the bias. Over a large Euro-Atlantic domain, for example, winter blocking frequency according to the AGP index is seen to be underestimated by 32% for the median CMIP5 model, whereas the same number is 19% for the median CMIP6 model.

We have addressed our second aim using the CMIP6-HighResMIP RIMAVERA simulations to assess the sensitivity of

280 simulated blocking frequency and persistence to resolution. The PRIMAVERA simulations have been designed to assess the role of model resolution specifically by conducting simulations with the same model at both low and high atmosphere resolution (and ocean resolution in coupled setups) without re-tuning the high-resolution version of the model. We find that higher-resolution PRIMAVERA models represent the mean blocking frequency better than the low-resolution models, for the Euro-Atlantic region during winter and summer, and for the Pacific in summer with no sensitivity to resolution seen in winter. This

285 improvement in mean blocking frequency is especially clear for the Euro-Atlantic region and the spatial correlation of the blocking frequency field suggesting that higher-resolution models tend to better simulate the location of blocking occurrences, arguably due to an improvement in the mean circulation over the North Atlantic, which is consistent with previous studies (e.g., Zappa et al., 2013). While the improvement with resolution in the simulated mean blocking frequency is clear, our analysis does not provide robust evidence for a systematic improvement in the simulated blocking persistence.

290 Our results are consistent with previous findings that the successful simulation of blocking in climate models depends delicately on a range of factors and their interactions, including horizontal and vertical model resolution, orographic boundary conditions, physical parameterisations, and the numerical scheme (Woollings et al., 2018). We corroborate here that horizontal resolution in the atmosphere when increased from, broadly, 100 km to 20 km, is one of these factors and benefits the simulation of blocking. We note that our results regarding model resolution should be considered conservative, as PRIMAVERA models

295 have not been retuned at the higher resolutions. At the same time, we show that an increase in resolution in and of itself will not fully remedy blocking biases in models, notably in the persistence of blocking events. We also find that the most recent generation of GCMs continues to be affected by long-standing blocking biases (Fig 1 and 2), albeit at a smaller magnitude in CMIP6 models than in CMIP5 models. This implies that, overall, CMIP6 models strike a better balance of the different factors affecting blocking simulation mentioned above. Our findings show that current approaches to model evaluation and

300 development pursued by different modelling centres, based on ideas such as physical process-based evaluation/development and seamlessness, hold promise to further reduce blocking biases in future.

One question that deserves further attention in future work is the role of the ocean resolution. We find in the PRIMAVERA simulations that the sensitivity to resolution is generally larger in the coupled than in the AMIP simulations. This raises the question if it is not only the better sea-surface temperature mean state that benefits the atmosphere mean state and blocking

305 in higher-resolution models (cf., Scaife et al., 2011) but also the simulation of air-sea-interactions themselves. PRIMAVERA

simulations are not designed to answer this question as atmosphere and ocean resolution are increased simultaneously, so that this will have to be addressed by means of future process studies and coordinated experiments increasing ocean and atmosphere resolution separately.

Data availability. CMIP5 historical, CMIP6 historical, and PRIMAVERA (through CMIP6-HighResMIP) simulations used in this study are
310   available from the Earth System Grid Federation (ESGF). ERA-40 and ERA-Interim reanalysis data are available from the European Centre for Medium Range Weather Forecasting (ECMWF).

Author contributions. RS conducted all of the data analyses and visualisations, and wrote the manuscript. PA, FDR, KJ, MJR, DS, CDR, LT, PLV delivered the PRIMAVERA model simulations. DB and RS developed the ANOM blocking index code. All authors commented on the manuscript.

315   *Competing interests.* The authors declare that there are no competing interests.

Acknowledgements. RS, PA, FDR, KL, MJR, DS, CDR, LT, and PLV acknowledge PRIMAVERA funding received from the European Commission under Grant Agreement 641727 of the Horizon 2020 research programme. Use of the UK NERC (Natural Environment Research Council) CEDA-JASMIN facility is acknowledged for data storage and analysis.

References

320 Anstey, J. A., Davini, P., Gray, L. J., Woollings, T. J., Butchart, N., Cagnazzo, C., Christiansen, B., Hardiman, S. C., Osprey, S. M., and Yang, S.: Multi-model analysis of Northern Hemisphere winter blocking: Model biases and the role of resolution, Journal of Geophysical Research: Atmospheres, 118, 3956–3971, https://doi.org/10.1002/jgrd.50231, http://onlinelibrary.wiley.com/doi/10.1002/jgrd.50231/abstracthttp://doi.wiley.com/10.1002/jgrd.50231, 2013.

Barriopedro, D., García-Herrera, R., and Trigo, R. M.: Application of blocking diagnosis methods to General Circulation Models. Part I:
325 a novel detection scheme, Climate Dynamics, 35, 1373–1391, https://doi.org/10.1007/s00382-010-0767-5, http://link.springer.com/10.1007/s00382-010-0767-5, 2010.

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and García-Herrera, R.: The hot summer of 2010: redrawing the temperature record map of Europe., Science (New York, N.Y.), 332, 220–224, https://doi.org/10.1126/science.1201224, 2011.

Berckmans, J., Woollings, T., Demory, M.-E., Vidale, P.-L., and Roberts, M.: Atmospheric blocking in a high resolution climate model:
330 influences of mean state, orography and eddy forcing, Atmospheric Science Letters, 14, 34–40, https://doi.org/10.1002/asl2.412, http://doi.wiley.com/10.1002/asl2.412, 2013.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., and Codron, F.: Winter 2010 in Europe: A cold extreme in a warming climate, Geophysical Research Letters, 37, 1–6, https://doi.org/10.1029/2010GL044613, 2010.

D'Andrea, F., Tibaldi, S., Blackburn, M., Boer, G., Déqué, M., Dix, M. R., Dugas, B., Ferranti, L., Iwasaki, T., Kitoh, a.,
335 Pope, V., Randall, D., Roeckner, E., Strauss, D., Stern, W., Van den Dool, H., and Williamson, D.: Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979-1988, Climate Dynamics, 14, 385–407, https://doi.org/10.1007/s003820050230, http://www.springerlink.com/openurl.asp?genre=article{&}id=doi:10.1007/s003820050230, 1998.

Davini, P., Corti, S., D'Andrea, F., Rivière, G., and von Hardenberg, J.: Improved Winter European Atmospheric Blocking Fre-
340 quencies in High-Resolution Global Climate Simulations, Journal of Advances in Modeling Earth Systems, 9, 2615–2634, https://doi.org/10.1002/2017MS001082, http://doi.wiley.com/10.1002/2017MS001082, 2017.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz,
345 B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, http://doi.wiley.com/10.1002/qj.828, 2011.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958,
350 https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Catherine, A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., Von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J. J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and Von Storch, J. S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, www.geosci-model-dev.
355 net/9/4185/2016, 2016.

Jung, T., Miller, M. J., Palmer, T. N., Towers, P., Wedi, N., Achuthavarier, D., Adams, J. M., Altshuler, E. L., Cash, B. A., Kinter, J. L., Marx, L., Stan, C., and Hodges, K. I.: High-Resolution Global Climate Simulations with the ECMWF Model in Project Athena: Experimental Design, Model Climate, and Seasonal Forecast Skill, Journal of Climate, 25, 3155–3172, https://doi.org/10.1175/JCLI-D-11-00265.1, http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00265.1, 2012.

360 Masato, G., Hoskins, B. J., and Woollings, T.: Winter and Summer Northern Hemisphere Blocking in CMIP5 Models, Journal of Climate, 26, 7044–7059, https://doi.org/10.1175/JCLI-D-12-00466.1, http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-12-00466.1, 2013.

Matsueda, M.: Predictability of Euro-Russian blocking in summer of 2010, Geophysical Research Letters, 38, 1–6, https://doi.org/10.1029/2010GL046557, 2011.

Matsueda, M., Mizuta, R., and Kusunoki, S.: Future change in wintertime atmospheric blocking simulated using a 20-km-mesh atmospheric
365 global circulation model, Journal of Geophysical Research, 114, D12 114, https://doi.org/10.1029/2009JD011919, http://www.agu.org/pubs/crossref/2009/2009JD011919.shtml, 2009.

Otto, F. E. L., Massey, N., Van Oldenborgh, G. J., Jones, R. G., and Allen, M. R.: Reconciling two approaches to attribution of the 2010 Russian heat wave, Geophysical Research Letters, 39, 1–5, https://doi.org/10.1029/2011GL050422, 2012.

Pithan, F., Shepherd, T. G., Zappa, G., and Sandu, I.: Climate model biases in jet streams, blocking and storm tracks resulting from missing
370 orographic drag, Geophysical Research Letters, 43, 7231–7240, https://doi.org/10.1002/2016GL069551, 2016.

Rex, D. F.: Blocking Action in the Middle Troposphere and its Effect upon Regional Climate, Tellus, 2, 275–301, https://doi.org/10.1111/j.2153-3490.1950.tb00339.x, http://tellusa.net/index.php/tellusa/article/view/8603, 1950.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S.,
375 Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, Journal of Climate, 24, 3624–3648, https://doi.org/10.1175/JCLI-D-11-00015.1, http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00015.1, 2011.

Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., Jackson, L. C., Kuhlbrodt, T., Mathiot, P., Roberts, C. D., Schiemann, R., Seddon, J., Vannière, B., and Vidale, P. L.: Description of the resolution hierarchy of the global coupled
380 HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments, Geoscientific Model Development Discussions, pp. 1–47, https://doi.org/10.5194/gmd-2019-148, 2019.

Sausen, R., König, W., and Sielmann, F.: Analysis of blocking events from observations and ECHAM model simulations, Tellus A, 47, 421–438, https://doi.org/10.1034/j.1600-0870.1995.t01-3-00003.x, http://tellusa.net/index.php/tellusa/article/view/11526, 1995.

Scaife, A. A., Woollings, T., Knight, J., Martin, G., and Hinton, T.: Atmospheric Blocking and Mean Biases in Climate Models, Journal of
385 Climate, 23, 6143–6152, https://doi.org/10.1175/2010JCLI3728.1, http://journals.ametsoc.org/doi/abs/10.1175/2010JCLI3728.1, 2010.

Scaife, A. A., Copsey, D., Gordon, C., Harris, C., Hinton, T., Keeley, S., O'Neill, A., Roberts, M., and Williams, K.: Improved Atlantic winter blocking in a climate model, Geophysical Research Letters, 38, L23 703, https://doi.org/10.1029/2011GL049573, http://doi.wiley.com/10.1029/2011GL049573, 2011.

Scherrer, S. C., Croci-Maspoli, M., Schwierz, C., and Appenzeller, C.: Two-dimensional indices of atmospheric blocking and their
390 statistical relationship with winter climate patterns in the Euro-Atlantic region, International Journal of Climatology, 26, 233–249, https://doi.org/10.1002/joc.1250, http://doi.wiley.com/10.1002/joc.1250, 2006.

Schiemann, R., Demory, M.-E., Shaffrey, L. C., Strachan, J., Vidale, P. L., Mizielinski, M. S., Roberts, M. J., Matsueda, M., Wehner, M. F., and Jung, T.: The Resolution Sensitivity of Northern Hemisphere Blocking in Four 25-km Atmospheric Global Circulation Models,

Journal of Climate, 30, 337–358, https://doi.org/10.1175/JCLI-D-16-0100.1, http://journals.ametsoc.org/doi/10.1175/JCLI-D-16-0100.1, 2017.

Schwierz, C., Croci-Maspoli, M., and Davies, H. C.: Perspicacious indicators of atmospheric blocking, Geophysical Research Letters, 31, L06 125, https://doi.org/10.1029/2003GL019341, http://www.agu.org/pubs/crossref/2004/2003GL019341.shtml, 2004.

Tableman, M., Kim, J. S., and Portnoy, S.: Survival Analysis Using S: Analysis of Time-to-Event Data, Taylor & Francis, Boac Raton, 2003.

Taylor, K. E., Stouffer, R. J., and Meehl, G. a.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-11-00094.1, 2012.

Tibaldi, S. and Molteni, F.: On the operational predictability of blocking, Tellus A, 42, 343–365, https://doi.org/10.3402/tellusa.v42i3.11882, http://www.blackwell-synergy.com/links/doi/10.1034/j.1600-0870.1990.t01-2-00003.xhttp://tellusa.net/index.php/tellusa/article/view/11882http://journals.sfu.ca/coaction/index.php/tellusa/article/view/11882, 1990.

Uppala, S. M., KÅllberg, P. W., Simmons, a. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. a., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. a., Beljaars, a. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. a. E. M., Jenne, R., Mcnally, a. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. a., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961–3012, https://doi.org/10.1256/qj.04.176, http://doi.wiley.com/10.1256/qj.04.176, 2005.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations, Journal of Advances in Modeling Earth Systems, 10, 357–380, https://doi.org/10.1002/2017MS001115, http://doi.wiley.com/10.1002/2017MS001115https://onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001115, 2018.

Woollings, T.: Dynamical influences on European climate: an uncertain future., Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 368, 3733–3756, https://doi.org/10.1098/rsta.2010.0040, 2010.

Woollings, T., Barriopedro, D., Methven, J., Son, S.-W., Martius, O., Harvey, B., Sillmann, J., Lupo, A. R., and Seneviratne, S.: Blocking and its Response to Climate Change, Current Climate Change Reports, pp. 287–300, https://doi.org/10.1007/s40641-018-0108-z, http://link.springer.com/10.1007/s40641-018-0108-z, 2018.

Zappa, G., Shaffrey, L. C., and Hodges, K. I.: The Ability of CMIP5 Models to Simulate North Atlantic Extratropical Cyclones*, Journal of Climate, 26, 5379–5396, https://doi.org/10.1175/JCLI-D-12-00501.1, http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-12-00501.1, 2013.
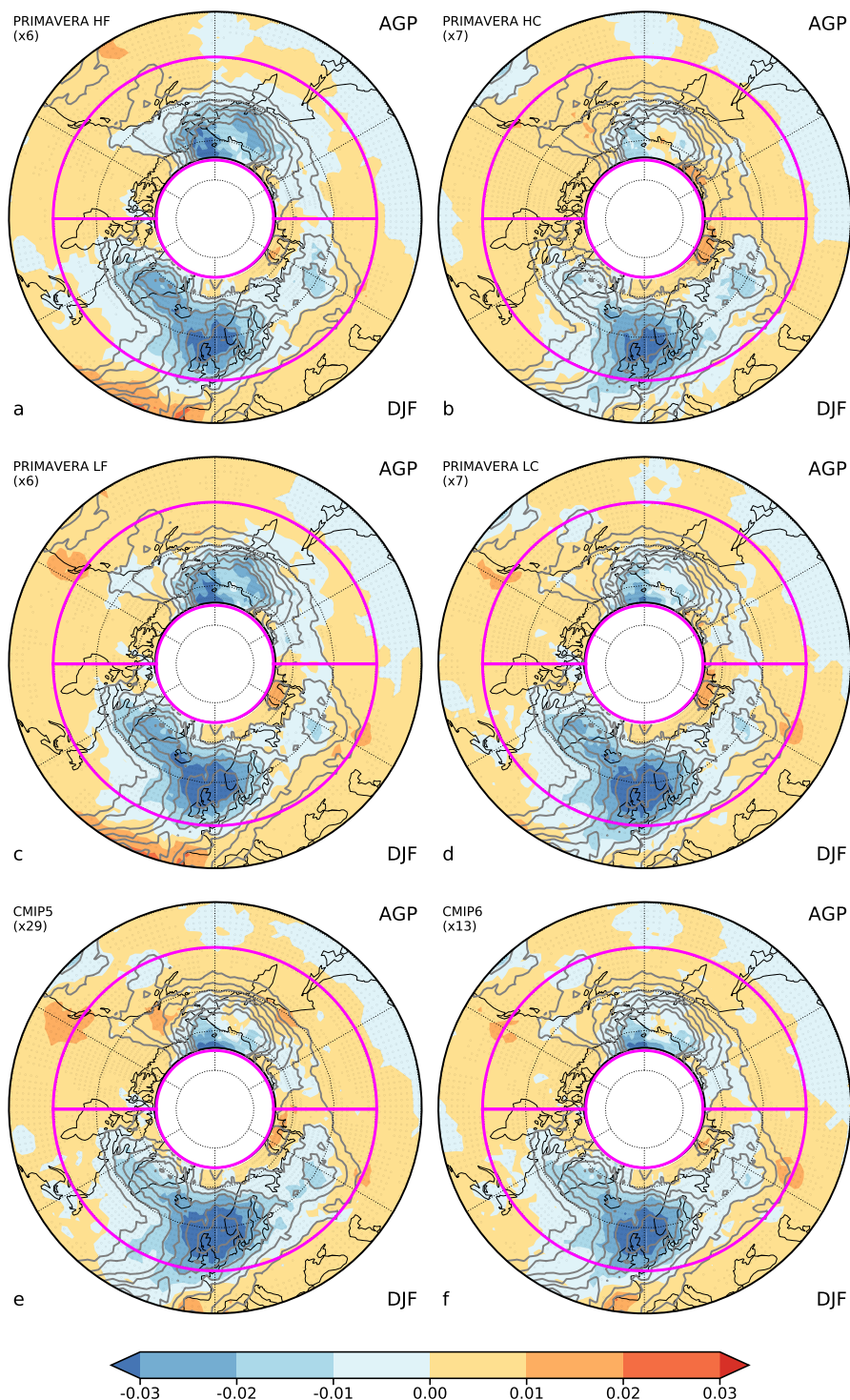
Figure 1. Bias in the frequency of blocked days for the AGP index, boreal winter, and (a) high-resolution forced, (b) high-resolution coupled, (c) low-resolution forced, (d) low-resolution coupled PRIMAVERA simulations, and (e) CMIP5, (f) CMIP6 simulations. Stippling shows agreement on the sign of the bias by at least (a,c) 6 of 6, (b,d) 6 of 7, (e) 19 of 29, and (f) 10 of 13 simulations. Grey contour lines show the reanalysis blocking frequency, at contour intervals of 0.01 and starting from 0.01. ATL and PAC evaluation domains are shown by magenta lines.
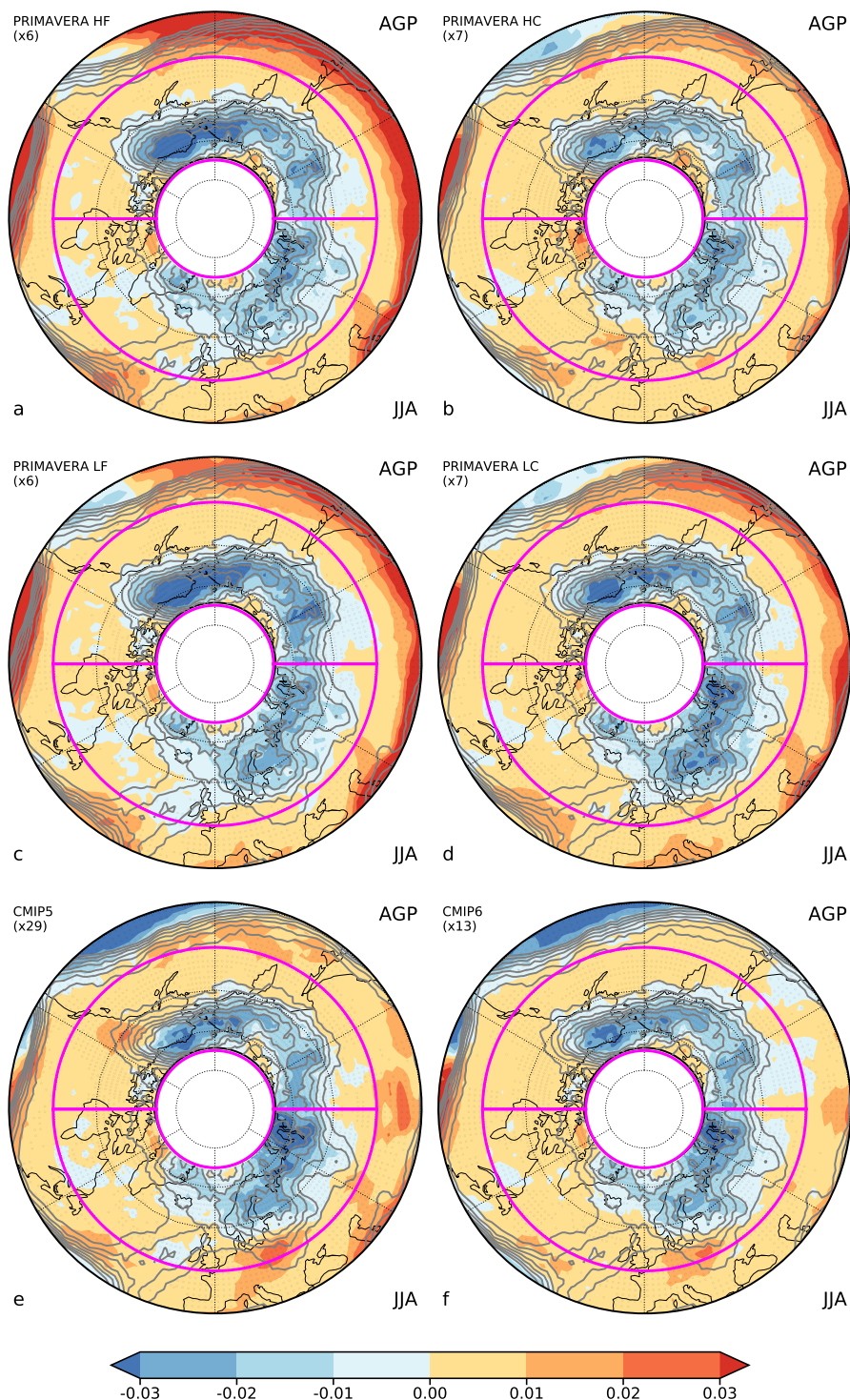
Figure 2. Blocking frequency bias for the AGP index, boreal summer, and (a) high-resolution forced, (b) high-resolution coupled, (c) low-resolution forced, (d) low-resolution coupled PRIMAVERA simulations, and (e) CMIP5, (f) CMIP6 simulations. Stippling shows agreement on the sign of the bias by at least (a,c) 6 of 6, (b,d) 6 of 7, (e) 19 of 29 and (f) 10 of 13 simulations. Grey contour lines show the reanalysis blocking frequency, at contour intervals of 0.01 and starting from 0.01. ATL and PAC evaluation domains are shown by magenta lines.
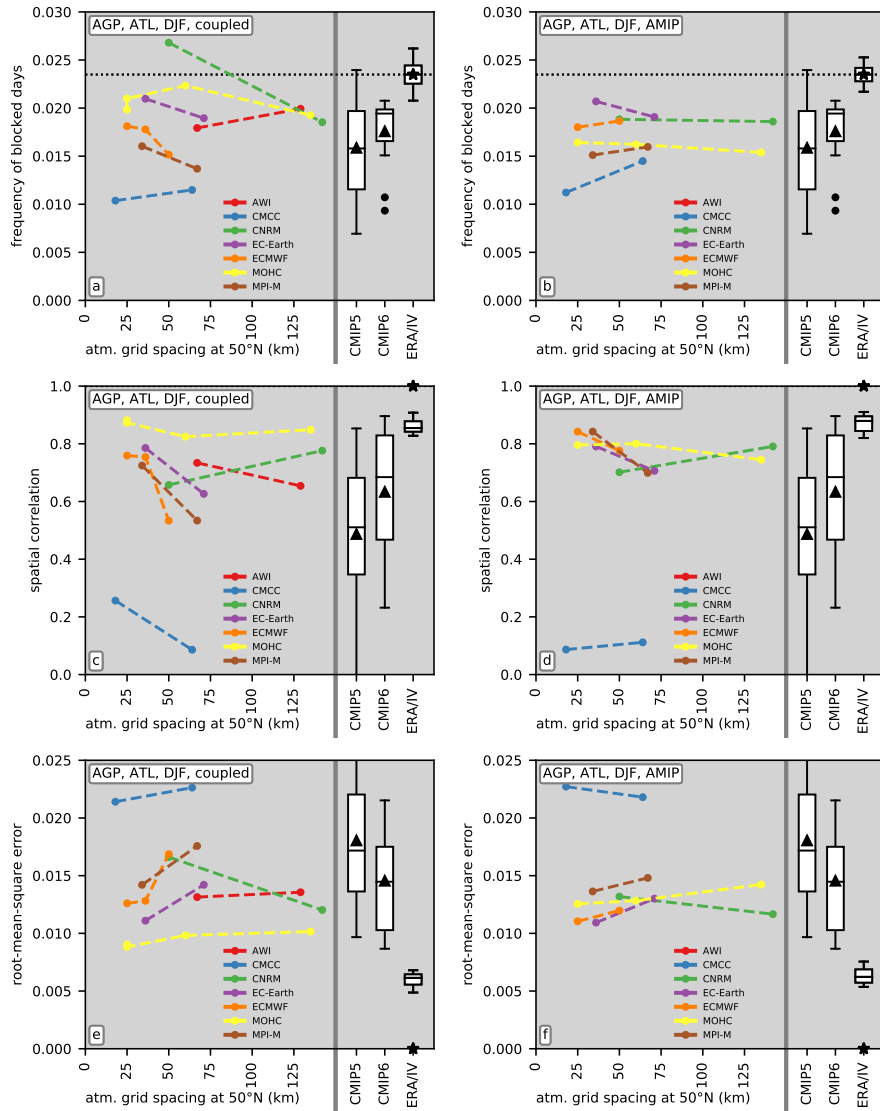
Figure 3. Metrics of blocking performance (a,b - blocking frequency, c,d - spatial correlation, e,f - root-mean-square error) for the AGP index and boreal winter, for the ATL domain (-90-90E, 50-75N). The left-hand side of each panel shows metrics for PRIMAVERA simulations at different grid spacings (resolutions). Boxplots on the righ-hand side show distributions of the metric across CMIP5 and CMIP6 simulations. The '*' symbol in the column 'ERA/IV' shows the reanalysis estimate and the boxplot is an estimate of the expected agreement given internal variability (see text).

Figure 4. Metrics of blocking performance (a,b - blocking frequency, c,d - spatial correlation, e,f - root-mean-square error) for the AGP index and boreal summer, for the ATL domain (-90-90E, 50-75N). The left-hand side of each panel shows metrics for PRIMAVERA simulations at different grid spacings (resolutions). Boxplots on the right-hand side show distributions of the metric across CMIP5 and CMIP6 simulations. The '*' symbol in the column 'ERA/IV' shows the reanalysis estimate and the boxplot is an estimate of the expected agreement given internal variability (see text).
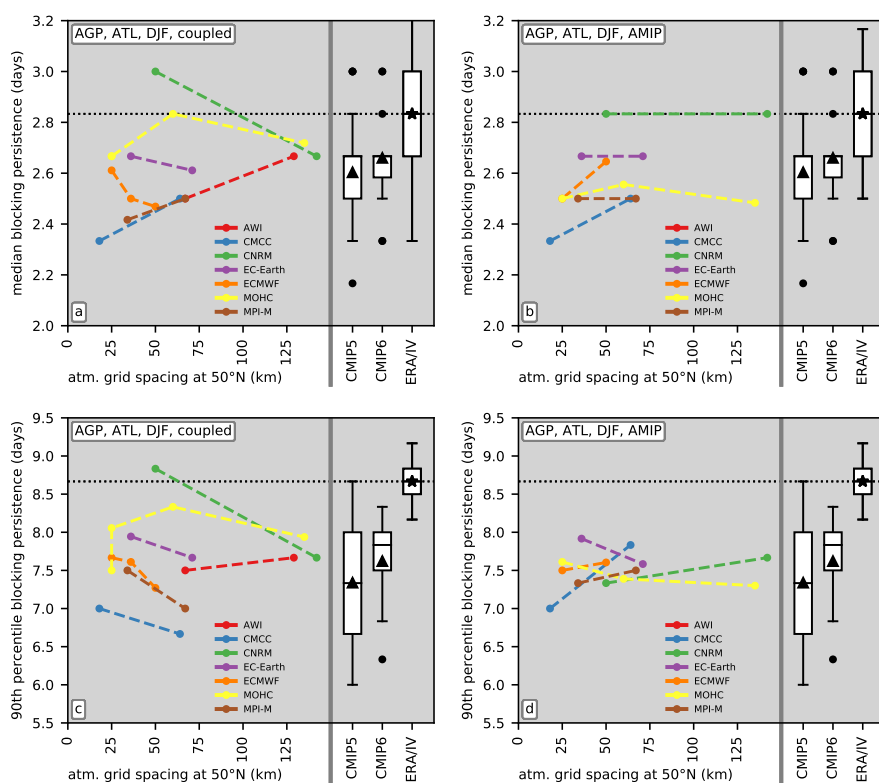
Figure 5. Persistence of blocking events (a,b - median, c,d - 90th percentile) for the AGP index and boreal winter, for the ATL domain (-90-90E, 50-75N). The left-hand side of each panel shows metrics for PRIMAVERA simulations at different grid spacings (resolutions). Boxplots on the righ-hand side show distributions of the persistence metric across CMIP5 and CMIP6 simulations. The '*' symbol in the column 'ERA/IV' shows the reanalysis estimate and the boxplot is an estimate of the expected agreement given internal variability (see text).
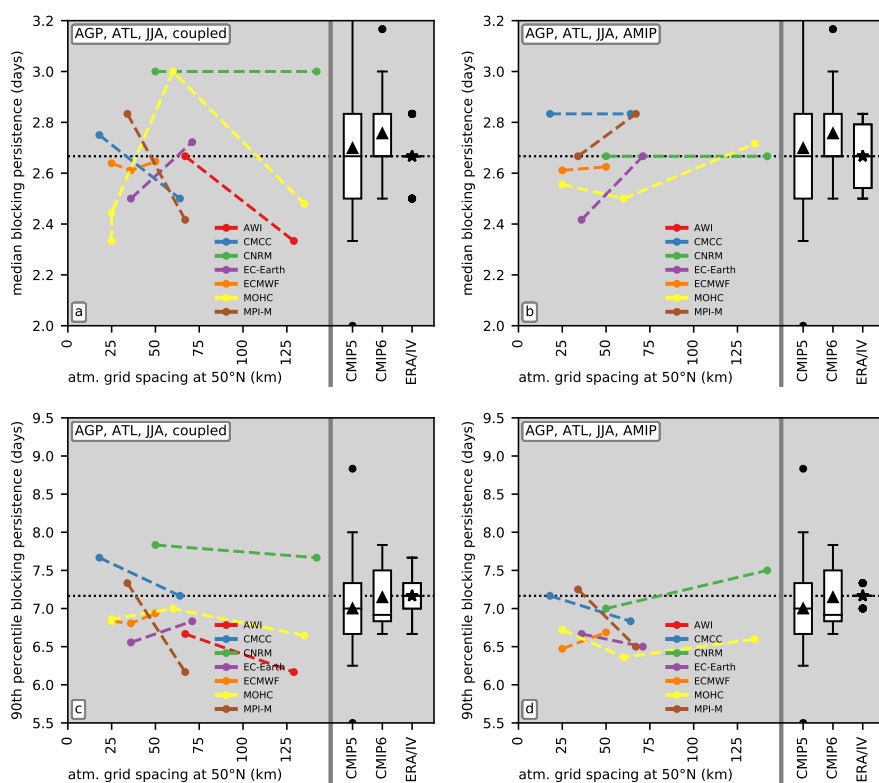
Figure 6. Persistence of blocking events (a,b - median, c,d - 90th percentile) for the AGP index and boreal summer, for the ATL domain (-90-90E, 50-75N). The left-hand side of each panel shows metrics for PRIMAVERA simulations at different grid spacings (resolutions). Boxplots on the righ-hand side show distributions of the persistence metric across CMIP5 and CMIP6 simulations. The '*' symbol in the column 'ERA/IV' shows the reanalysis estimate and the boxplot is an estimate of the expected agreement given internal variability (see text).